Gerben Bruinsma
David Weisburd
Editors

Encyclopedia of Criminology and Criminal Justice

SpringerReference

Encyclopedia of Criminology and Criminal Justice

Gerben Bruinsma
David Weisburd
Editors

Encyclopedia of Criminology and Criminal Justice

A–B

With 311 Figures and 150 Tables

Springer Reference

Editors
Gerben Bruinsma
Netherlands Institute for the Study of Crime and Law Enforcement (NSCR)
Amsterdam, The Netherlands

VU University Amsterdam
Amsterdam, The Netherlands

David Weisburd
Department of Criminology, Law and Society
George Mason University
Fairfax, VA, USA

Faculty of Law
The Hebrew University
Mt. Scopus, Jerusalem, Israel

ISBN 978-1-4614-5689-6 ISBN 978-1-4614-5690-2 (eBook)
ISBN 978-1-4614-5691-9 (Print and electronic bundle)
DOI 10.1007/978-1-4614-5690-2
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013946647

© Springer Science+Business Media New York 2014
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Introduction

When we began our effort to outline and organize the *Encyclopedia of Criminology and Criminal Justice* with our distinguished Associate Editors, we set before ourselves a major task that would be different from that of most other encyclopedias that are produced today in various fields. We looked back in history to find our task. We sought to define the parameters of the discipline of criminology and criminal justice, in the spirit of the encyclopedias first developed in the 18th century. This Encyclopedia would not be a dictionary of the field, but a cutting edge statement of knowledge in the field at this time.

Crime and criminal justice are major dynamic issues in contemporary societies. Every society is confronted with variations in crime rates and has to deal with large groups of crime places, offenders, and victims. Crimes appear in changing images in everyday life and criminal justice agencies have to find solutions for these societal phenomena. This means that the science of criminology is important not just for scientists and scholars who want to develop basic knowledge and understand the causes of crime, but also to policy makers and practitioners. Criminology as a discipline spans both basic and applied research questions, and criminology has been enriched by its ability to both provide scientific knowledge to practice and to raise critical questions about the basic principles and effects of policy and practice. There is an international community of active producers of criminological knowledge in every continent at universities, research institutes, and governmental offices. The globalization of criminological knowledge has increased the dissemination of criminological research and theory to all parts of the world. This Encyclopedia seeks to summarize that broad array of knowledge.

Undergraduate and graduate students in criminology and criminal justice are taught using general text books with brief sections on relevant topics. Further information is explored on the Internet, using Wikipedia texts for additional knowledge or for writing papers. These texts have not been reviewed by scholars in the field. The print and online versions of this Encyclopedia give students, researchers, policy makers, and practitioners direct access to a reliable body of knowledge on topics that have been written by experts in the discipline. Indeed, in this Encyclopedia it is often the originators of theories, practices, or methods that are writing the entries. These entries will give students and scholars efficient and solid insights into the knowledge they need for their courses.

Goals

This *Encyclopedia of Criminology and Criminal Justice* presents the current state of knowledge of this discipline. Like science in general, criminology expanded during the last two centuries, starting with the first geographic studies on the distribution of crime rates until new neurological knowledge of offenders and innovative forensic techniques to detect perpetrators, and to evidence-based treatment programs for serious or frequent offenders or local governmental to private interventions in crime situations. We needed ten volumes with 579 entries to display today's criminology and criminal justice knowledge base. It represents our efforts to provide the reader with current state-of-the-art knowledge in criminology and criminal justice.

The aim of the *Encyclopedia of Criminology and Criminal Justice* is to supply a comprehensive reference tool for the field of criminology and criminal justice that is both cutting edge and of very high scientific quality. This ten volume work provides a complete and systematic coverage of this growing field that is unprecedented, as it is truly international and includes fields related to criminology, such as police science, forensics, and certain areas of psychology. Each entry offers an extended description of the topic, relevant literature, current base of knowledge, and ideas about what needs to be studied in the future. The on-line version will be updated continuously after the first publication serving millions of students, researchers, policy makers, and practitioners of all continents.

Defining the Discipline

The goal of this Encyclopedia is to provide a comprehensive reference work for the field of criminology and criminal justice. We worked hard, meeting multiple times with our Associate Editors, to identify what the critical areas were, and what research existed that we could include in the Encyclopedia. The Encyclopedia in this sense "defines the field" through its choice of organization and entries. We aimed at identifying emerging ideas and trends, so that the work will be timely at publication and afterwards. This Encyclopedia covers the field broadly, and internationally, and attempts to be up to date as to recent developments in research and practice in the field. This Encyclopedia is not a dictionary, nor a kind of Wikipedia without any quality assessment, and aims at comprehensive and cutting edge knowledge that defines the contours of the field of criminology and criminal justice. That is why we included the fast developing new fields of forensics, psychology of law, and investigative psychology. The ten volumes cover the following broad fields of criminology and criminal justice (listed with the Associate Editors responsible):

Corrections and Criminal Justice Supervision in the Community (Doris MacKenzie)
Courts, Sentencing, and the Judicial System (Leslie Sebba)
Police and Law Enforcement (Stephen Mastrofski)
Crimes, Criminals, and Victims (Alex Piquero)

Crime Places and Situations (Cynthia Lum)
Explanations for Criminal Behavior (Sally Simpson)
History of Criminology (all Associate Editors and Editors-in-Chief)
Data, Methods, and Statistics (Arjan Blokland and Dan Nagin)
Social Interventions and Prevention (Gwladys Gilliéron & Martin Killias)
Forensics and Forensic Science Investigative Psychology (Peter Neyroud)
Psychology of Law (Karen Amendola)

Every section defines a number of areas of research (a total of 111), and consequently a number of relevant topics or articles that are included as entries (579). The entries are arranged in alphabetical order. Each entry provides an overview of what is covered, followed by a short list of suggested readings and references. All entries have been reviewed, first by the Area Editors, then by the Associate Editors, and finally by the Editors-in-Chief. Not all of the articles submitted could be accepted into the Encyclopedia, as is the case with other academic work that is reviewed. We very much appreciate the efforts of all of our authors, who have truly produced an important work in criminology and criminal justice.

Acknowledgements

As the reader can imagine, this ten volume *Encyclopedia of Criminology and Criminal Justice* is a work that involves the collaborative efforts of many people. In the 18th century, Dennis Diderot and Jean le Rond d'Alembert published between 1751 and 1772 the first encyclopedia on science with the assistance of hundreds of scholars. It consisted of 28 volumes with 60,000 articles on all imaginable topics of the natural sciences. The editors themselves contributed 6,000 articles. This Encyclopedia is also a joint effort of many people.

When David Weisburd was approached by Springer, he invited Gerben Bruinsma to join him as Editor-in-Chief. We contributed equally to the Encyclopedia, explaining the alphabetical order of our names. We had the magnificent support of a group of 12 Associate Editors, all renowned experts in their fields. They organized their section by defining the areas of research and used their extended professional network to contact the 111 Area Editors. We owe them very much for their efforts: Karen Amendola (Police Foundation), Arjan Blokland (NSCR), Gwladys Gilliéron (Distance Learning University Switzerland, Brig, Switzerland), Martin Killias (KRC Killias Research & Consulting SA), Cynthia Lum (George Mason University), Doris MacKenzie (Penn State University), Stephen Mastrofski (George Mason University), Daniel Nagin (Carnegie Mellon University), Peter Neyroud (Cambridge University), Alex Piquero (The University of Texas at Dallas), Leslie Sebba (The Hebrew University), and Sally Simpson (University of Maryland). We also would like to thank the 111 Area Editors who selected the authors and reviewed the papers in their areas. Their names are listed separately in this Encyclopedia. Lastly, we are grateful to all our authors who contributed with scholarly entries to the Encyclopedia. It is their collective effort and expertise that has made the publication of Springer's *Encyclopedia of Criminology and Criminal Justice* a successful enterprise.

A special word of gratitude is for the staff of the publisher, Springer. Many people have contributed to the Encyclopedia from its inception through its production. We thank them all for their wonderful work. But some Springer staff deserve special mention. Welmoed Spahr, Executive Editor, Social and Behavioral Sciences, and Katie Chabalko, Criminology Editor, supported us from the beginning to the end of the project often going much beyond their obligations to ensure the academic quality of the work. Without Saskia Ellis, the project manager, this project would never have been finished, or at least would not have come out in a timely manner. She was a power house in moving the Encyclopedia along, and worked closely with us in bringing this major project to completion. To the literally hundreds of people who worked on this project we thank you all, and hope that you are very pleased with the scope and quality of Springer's *Encyclopedia of Criminology and Criminal Justice*.

May 2013

Gerben Bruinsma
Netherlands Institute for the Study of
Crime and Law Enforcement (NSCR)
(Amsterdam, The Netherlands) & VU
University Amsterdam (Amsterdam, The Netherlands)

David Weisburd
Department of Criminology, Law and Society,
George Mason University (Fairfax, VA, USA) &
Faculty of Law, The Hebrew University (Jerusalem, Israel)

Editors-in-Chief

Gerben Bruinsma Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, The Netherlands

VU University Amsterdam, Amsterdam, The Netherlands

Professor Dr. Gerben Bruinsma is, since 1999, director of the Netherlands Institute for the Study of Crime and Law Enforcement (NSCR) in Amsterdam, a national research institute of the National Organization for Scientific Research (NWO). He is also, since 2009, professor of environmental criminology at VU University Amsterdam. In the past he has held positions as professor of criminology at Twente University and Leiden University. He studied sociology and criminology at Utrecht University and finished his doctoral dissertation *Crime as a Social Process: A Test of the Differential Association Theory in the Version of K-D- Opp* at the Radboud University Nijmegen. In the 1990s he co-founded the International Police Institute at Twente University. He was president of the Dutch Society of Criminology and one of the founding fathers of the European Society of Criminology, editor or editorial board member of various journals, and has held a great number of advisory and board positions in the field of police and criminal justice. In 2009, he received the Freda Adler Distinguished International Scholar Award of the Division of International Criminology of the American Society of Criminology (ASC). He has researched and published on juvenile delinquency, organised crime, spatial distributions of crime, police theory, and methodological issues. He has published more than 100 articles, 30 books, and 60 book chapters. His current interests are geographical, theoretical, and historical criminology.

David Weisburd Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA

Faculty of Law, The Hebrew University, Mt. Scopus, Jerusalem, Israel

David Weisburd is Distinguished Professor of Criminology, Law and Society at George Mason University (and Director of its Center for Evidence Based Crime Policy) and Walter E. Meyer Professor of Law and Criminal Justice at The Hebrew University, Faculty of Law in Jerusalem. He also serves as Senior Fellow at the Police Foundation in Washington, DC and is Chair of its Research Advisory Committee. Professor Weisburd is an elected Fellow of the American Society of Criminology and of the Academy of Experimental Criminology. He is a member of the Science Advisory Board of the Office of Justice Programs (USA), the Steering Committee of the Campbell Crime and Justice Group, the Harvard Executive Session in Policing, and the Committee on Law and Justice of the National Research Council (National Academy of Sciences, USA). Professor Weisburd is one of the leading international researchers in crime and justice. He is author or editor of more than 20 books and more than 100 scientific articles that cover a wide range of criminal justice research topics, including crime at place, violent crime, white collar crime, policing, illicit markets, criminal justice statistics, and social deviance. Professor Weisburd is editor of the *Journal of Experimental Criminology* and sits on many journal editorial boards, including *Evaluation Review*, the *Journal of Research in Crime and Delinquency*, and the *Journal of Quantitative Criminology*. He is the 2008 recipient of the Joan McCord Award from the Academy of Experimental Criminology, the 2010 recipient of the Stockholm Prize in Criminology, and the 2011 recipient of the Klachky Prize for the Advancement of the Frontiers of Science. In 2011, he was chosen by a committee of the Israeli Academy of the Sciences and Humanities as a recipient of the Minister's Prize for Outstanding Immigrant Scientists.

Associate Editors

Karen L. Amendola Police Foundation, Washington, DC, USA
Subject Area: Psychology of Law

Arjan A. J. Blokland Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, The Netherlands

Department of Criminology and Criminal Law, Leiden University, Leiden, The Netherlands
Subject Area: Data, Methods, and Statistics

Gwladys Gilliéron Distance Learning University Switzerland, Brig, Switzerland
Subject Area: Social Interventions and Prevention

Martin Killias KRC Killias Research & Consulting SA, Lenzburg, Switzerland
Subject Area: Social Interventions and Prevention

Cynthia Lum Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA
Subject Area: Crime Places and Situations

Doris L. MacKenzie The Pennsylvania State University, University Park, PA, USA
Subject Area: Corrections and Criminal Justice Supervision in the Community

Stephen D. Mastrofski George Mason University, Manassas, VA, USA
Subject Area: Police and Law Enforcement

Daniel S. Nagin Carnegie Mellon University, Pittsburgh, PA, USA
Subject Area: Data, Methods and Statistics

Peter Neyroud University of Cambridge, Cambridge, UK
Subject Area: Forensic and Investigative Psychology

Alex R. Piquero Department of Criminology, The University of Texas at Dallas, Richardson, TX, USA
Subject Area: Crimes, Criminals and Victims

Leslie Sebba Institute of Criminology, Faculty of Law, The Hebrew University, Mt. Scopus, Jerusalem, Israel
Subject Area: Courts, Sentencing and the Judicial System

Sally S. Simpson Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA
Subject Area: Explanations for Criminal Behavior

Area Editors

Michael G. Aamodt DCI Consulting Group, Inc., Washington, DC, USA
Subject Area: Psychology and Policing

Robert Agnew Department of Sociology, Emory University, Atlanta, GA, USA
Subject Area: Strain Theories

Mimi Ajzenstadt Rothberg International School, The Hebrew University, Mt. Scopus, Jerusalem, Israel
Subject Area: Archival and Historical Methods

Karen L. Amendola Police Foundation, Washington, DC, USA
Subject Area: Detection of Deception

Martin A. Andresen School of Criminology, Institute for Canadian Urban Research Studies, Simon Fraser University, Burnaby, BC, Canada
Subject Area: Predictive Models and Geographic Profiling

Robert Apel School of Criminal Justice, Rutgers University, Newark, NJ, USA
Subject Area: Quasi-experimental Methods

Gaylene S. Armstrong College of Criminal Justice, Correctional Management Institute of Texas Sam Houston State University, Huntsville, TX, USA
Subject Area: Prisons and Jails Including Correctional Management

Nachman Ben-Yehuda Department of Sociology and Anthropology, The Hebrew University, Mt. Scopus, Jerusalem, Israel
Subject Area: Labeling Theories

Michael L. Benson School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA
Subject Area: White Collar

Catrien Bijleveld Netherlands Institute for the Study of Crime and Law Enforcement, NSCR, Amsterdam, The Netherlands
Subject Area: Multivariate Analyses

Ingrid A. Binswanger Division of General Internal Medicine and Division of Substance Dependence, University of Colorado School of Medicine, Aurora, CO, USA

Denver Health Medical Center, Denver, CO, USA
Subject Area: Health and Corrections/Medical Issues

Daniel Birks ARC Center of Excellence in Policing and Security, CEPS, Griffith University, Brisbane, Australia
Subject Area: Simulation Models

Donna M. Bishop School of Criminology and Criminal Justice, Northeastern University, Boston, MA, USA
Subject Area: Juvenile Crime

Erhard Blankenburg Emeritus Professor, Faculty of Law, VU University, Amsterdam, The Netherlands
Subject Area: Legal Change, Regulation, and Compliance

Kate J. Bowers Department of Security and Crime Science, Jill Dando Institute of Crime Science, University College London, London, UK
Subject Area: Crime Specialization and Concentrations

Robert Brame Department of Criminal Justice and Criminology, University of North Carolina at Charlotte, Charlotte, NC, USA
Subject Area: Criminal Careers and Selection Models

Marjie Britz College of Business and Behavioural Science, Clemson University, Clemson, SC, USA
Subject Area: New Horizons for Forensics

Gerben Bruinsma Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, The Netherlands

VU University Amsterdam, Amsterdam, The Netherlands
Subject Area: Organized Crime

Catherine Burton Department of Criminal Justice, The Citadel, Charleston, SC, USA
Subject Area: New Horizons for Forensics

Julia L. Carrano School of Public Affairs, American University, Washington, DC, USA
Subject Area: Wrongful Convictions

Janet Chan Faculty of Law, University of New South Wales, Sydney, NSW, Australia
Subject Area: Technology and Policing

Steven M. Chermak School of Criminal Justice, Michigan State University, East Lansing, MI, USA

National Consortium for the Studies of Terrorism and Responses to Terrorism (START)
Subject Area: Media

Heith Copes University of Alabama at Birmingham, Birmingham, AL, USA
Subject Area: Property Crime

Gary Cordner Department of Criminal Justice, Kutztown University, Kutztown, PA, USA
Subject Area: Specialized Police Functions

Robert D. Crutchfield University of Washington, Seattle, WA, USA
Subject Area: Disorganization

Michael R. Davis Centre for Forensic Behavioural Science, Monash University, Melbourne, VIC, Australia
Subject Area: Forensic Psychology

Matthew DeMichelle Justice Center for Research, Department of Sociology, The Pennsylvania State University, University Park, PA, USA
Subject Area: Comparative Corrections

James M. Doyle Independent Researcher, Carney & Bassil, Boston, MA, USA
Subject Area: History of Psychology and Law

Frieder Dünkel Faculty of Law, University of Greifswald, Greifswald, Germany
Subject Area: Alternative/Community Sanctions

Laura Dugan Department of Criminology, University of Maryland, College Park, MD, USA
Subject Area: History of Methods and Statistics

Jennifer Dysart Department of Psychology, John Jay College of Criminal Justice, New York, NY, USA
Subject Area: Eyewitnesses

Henk Elffers Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, The Netherlands
Subject Area: Simulation Models

Lee Ellis Department of Anthropology and Sociology, University of Malaya, Kuala Lumpur, Malaysia, Malaysia
Subject Area: Biosocial Theoretical

Robin Engel School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA
Subject Area: Police Discretion and its Control

Horst Entorf Department of Economics and Business Administration, Goethe University Frankfurt, Frankfurt, Germany
Subject Area: General Deterrence & General Prevention

David P. Farrington Institute of Criminology, University of Cambridge, Cambridge, UK
Subject Area: Psychological Theories

James O. Finckenauer School of Criminal Justice, Rutgers University, Newark, NJ, USA
Subject Area: Comparative Definitions

James Frank School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA
Subject Area: Criminal Investigations

David O. Friedrichs Department of Sociology/Criminal Justice, University of Scranton, Scranton, PA, USA
Subject Area: Political-State Crime

Lior Gideon Department of Law, Police Science and Criminal Justice Administration, John Jay College of Criminal Justice, New York, NY, USA
Subject Area: History of Corrections and Punishment

Charlotte E. Gill Center for Evidence-Based Crime Policy, George Mason University, Fairfax, VA, USA
Subject Area: Probation and Parole

Denise C. Gottfredson Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA
Subject Area: School-based Prevention

Jack R. Greene School of Criminology and Criminal Justice, Northeastern University, Boston, MA, USA
Subject Area: Police Management and Leadership

Matthew Hall School of Law, University of Sheffield, Sheffield, UK
Subject Area: Victims in Criminal Procedure

Roger A. Hanson National Center for State Courts, Williamsburg, VA, USA
Subject Area: Culture and Organisation of the Courts

Nathan Harris Regulatory Institutions Network, Australian National University, Acton, ACT, Australia
Subject Area: Shaming, Defiance, Procedural Justice

Badi Hasisi Institute of Criminology, Faculty of Law, The Hebrew University, Mt. Scopus, Jerusalem, Israel
Subject Area: Policing Terrorism and Homeland Security

Julie Hibdon Department of Criminology and Criminal Justice, Southern Illinois University, Carbondale, IL, USA
Subject Area: Geographic Study of Crime

Joshua C. Hinkle Department of Criminal Justice and Criminology, Georgia State University, Atlanta, GA, USA
Subject Area: Social and Physical Disorder

Tony Hoskin Criminology Program, The University of Texas of the Permian Basin, Odessa, TX, USA
Subject Area: Biosocial Theoretical

Shane D. Johnson Department of Security and Crime Science, University College London, London, UK
Subject Area: Crime Specialization and Concentrations

Gerry Johnstone Law School, University of Hull, Hull, UK
Subject Area: Restorative Justice

Darrick Jolliffe School of Law, University of Greenwich, London, UK
Subject Area: Early Prevention and Interventions in Families

Martin Killias KRC Killias Research & Consulting SA, Lenzburg, Switzerland
Subject Area: Evidence Based Policy

David Klinger Department of Criminology and Criminal Justice, University of Missouri-St. Louis, St. Louis, MO, USA
Subject Area: Police Organization and Culture

Christopher S. Koper Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA
Subject Area: Police Strategies and Practices

Marvin Krohn Department of Sociology and Criminology and Law, University of Florida, Gainesville, FL, USA
Subject Area: Self-control Theories

Candace Kruttschnitt Centre for Criminology and Sociolegal Studies, University of Toronto, Toronto, ON, Canada
Subject Area: Feminist Theories

Sanja Kutnjak Ivkovic School of Criminal Justice, Michigan State University, East Lansing, MI, USA
Subject Area: Police Abuse of Authority

Gary LaFree National Consortium for the Study of Terrorism and Responses to Terrorism (START), University of Maryland, College Park, MD, USA
Subject Area: Terrorism

Tapio Lappi-Seppälä National Research Institute of Legal Policy, Helsinki, Finland
Subject Area: Incapacitation

Brian Lawton Criminology, Law and Society, College of Humanities and Social Sciences, George Mason University, Fairfax, VA, USA
Subject Area: Criminology of Crime Places/History and Context of Crime and Specific Crimes, Specific Places

Michael Levi Centre for Crime, Law and Justice at Cardiff School of Social Sciences, Cardiff University, Cardiff, Wales, UK
Subject Area: Explanations for White-collar and Corporate Crime

Rolf Loeber Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA
Subject Area: Longitudinal Designs

Cynthia Lum Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA
Subject Area: Units of Analysis

Patrick Lussier Laval University, Quebec City, QC, Canada
Subject Area: Sex Offenders

Ineke Haen Marshall Department of Sociology & Anthropology and School of Criminology and Criminal Justice, Northeastern University, Boston, MA, USA
Subject Area: Self-report Methodology

Matt Matravers Morrell Centre for Toleration, University of York, York, UK
Subject Area: Sentencing Philosophy

Rob I. Mawby University of South Wales, Newport, South Wales, UK

Department of Natural Sciences, University of Gloucestershire, Cheltenham, UK
Subject Area: Comparative and International Policing

Lorraine Mazerolle The Institute for Social Science Research, University of Queensland, St Lucia, QLD, Australia
Subject Area: Third Party Policing

Paul Mazerolle Arts, Education, and Law, Griffith University, Griffith, QLD, Australia
Subject Area: Specialization

Jean McGloin Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA
Subject Area: Co-offending

Tracey L. Meares Yale Law School, Yale University, New Haven, CT, USA
Subject Area: Law and the Police

Guus Meershoek Department of Management and Governance, University of Twente, Enschede, The Netherlands

Police Academy, Apeldoorn, The Netherlands
Subject Area: Police History

Robert Meier School of Criminology and Criminal Justice, University of Nebraska at Omaha, Omaha, NE, USA
Subject Area: Moral Crimes

Ruth M. Morgan Security and Crime Science, University College London, London, UK
Subject Area: Identification

Daniel S. Nagin Carnegie Mellon University, Pittsburgh, PA, USA
Subject Area: Systematic Observation and Visual Representation

Roger B. Parks Indiana University Bloomington, Bloomington, IN, USA
Subject Area: Police Performance

Stephan Parmentier Leuven Institute of Criminology, University of Leuven, Leuven, Belgium
Subject Area: International Criminal Justice

Raymond Paternoster Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA
Subject Area: Rational Choice/Deterrance and Developmental Statistical Methods

Ken Pease Jill Dando Institute, University College London, London, UK

Loughborough University, Loughborough, UK
Subject Area: History of Social Interventions and Prevention

Anthony Petrosino Learning Innovations, WestEd, Woburn, MA, USA
Subject Area: Experimental Methods

Alex R. Piquero Department of Criminology, The University of Texas at Dallas, Richardson, TX, USA
Subject Area: Violence

Paul Ponsaers Faculty of Law, Research Group SVA, Ghent University, Ghent, Belgium
Subject Area: Police and Community

Paul Roberts School of Law, University of Nottingham, Nottingham, UK
Subject Area: Forensics and the Criminal Trial Process

Dennis Rosenbaum Department of Criminal Justice, University of Illinois at Chicago, Chicago, IL, USA
Subject Area: Police Legitimacy

Dawn L. Rothe Department of Sociology and Criminal Justice, Old Dominion University, Norfolk, VA, USA
Subject Area: Political-state Crime

David B. Rottman National Center for State Courts, Williamsburg, VA, USA
Subject Area: Alternative Courts and Procedures

Wenona Rymond-Richmond Department of Sociology, University of Massachusetts-Amherst, Amherst, MA, USA
Subject Area: Violations Human Rights

Martin D. Schwartz George Washington University, Washington, DC, USA
Subject Area: Critical Criminology

Leslie Sebba Institute of Criminology, Faculty of Law, The Hebrew University, Mt. Scopus, Jerusalem, Israel
Subject Area: Sentencing and the Judicial System - Miscellaneous topics

Sally S. Simpson Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA
Subject Area: Social Learning and History of Criminological Theory and Routine Activity Theory

Alette Smeulers Department of Criminal Law, Tilburg University, Tilburg, The Netherlands
Subject Area: International Criminal Justice

Nigel South Department of Sociology, University of Essex, Colchester, Essex, UK
Subject Area: Environmental-green

Cyrus Tata Law School, Strathclyde University, Glasgow, Scotland, UK
Subject Area: Sentencing Processes

Faye S. Taxman Criminology, Law and Society, George Mason University, Fairfax, VA, USA
Subject Area: Rehabilitation, Recidivism, and "What Works" and Health and Corrections/Medical Issues

Terrance J. Taylor Department of Criminology and Criminal Justice, University of Missouri-St. Louis, St. Louis, MO, USA
Subject Area: Juvenile Gangs

William J. Tilstone Centre for Forensic Science, University of Strathclyde, Glasgow, Scotland, UK
Subject Area: History and Evolution and Forensic Science

George Tita Department of Criminology, Law and Society, University of California, Irvine, Irvine, CA, USA
Subject Area: Network Analysis

Michael Townsley School of Criminology and Criminal Justice, Griffith University, Brisbane, QLD, Australia
Subject Area: Geographical Information System and Spatial Statistics

Andromachi Tseloni School of Social Sciences, Nottingham Trent University, Nottingham, UK
Subject Area: Victimization

Jan J. Van Dijk Human Security and Safety International Victimology Institute, Tilburg University, Tilburg, The Netherlands
Subject Area: Victimization Survey

Vere M. van Koppen VU University Amsterdam, Amsterdam, The Netherlands
Subject Area: Organized Crime

Antoinette Verhage Faculty of Law, Research Group SVA, Ghent University, Ghent, Belgium
Subject Area: Police and Community

David Weisburd Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA

Faculty of Law, The Hebrew University, Mt. Scopus, Jerusalem, Israel
Subject Area: Experimental Methods

Celine Weyermann Institut de Police Scientifique (IPS), University of Lausanne, Lausanne, Switzerland
Subject Area: Scientific Quality and Forensic Science

Per-Olof H. Wikström Institute of Criminology, University of Cambridge, Cambridge, UK
Subject Area: Developmental Theories

James J. Willis Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA
Subject Area: Political Systems and Policing

David B. Wilson Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA
Subject Area: Systematic Reviews and Meta-analyses

Richard Wortley Jill Dando Institute of Security and Crime Science, University College London, London, UK
Subject Area: Situational Crime Prevention

Sue-Ming Yang Department of Criminology, National Chung Cheng University, Min-Hsiung, Chia-Yi, Taiwan
Subject Area: Criminal Careers of Places

Contributors

Michael G. Aamodt DCI Consulting Group, Inc., Washington, DC, USA

Howard Abadinsky St. John's University, New York, USA

Lynn A. Addington Department of Justice, Law, and Society, American University, Washington, DC, USA

Marcelo F. Aebi School of Criminal Sciences, Institute of Criminology and Criminal Law, University of Lausanne, Lausanne, Switzerland

Robert Agnew Department of Sociology, Emory University, Atlanta, GA, USA

Gracie Aguilar Spring, TX, USA

Cyrus Ahalt Division of Geriatrics, University of California San Francisco, San Francisco, CA, USA

Melissa Alexander US District Court, North Carolina Middle Probation Office, USA

Tracy Alexander Cold Case Investigation, LGC Forensics, Abingdon, UK

Laurence Alison University of Liverpool, Liverpool, UK

Rob Allen Justice and Prisons

Tim Allen Department International development, London School of Economics and Political Science, London, UK

Paul Almond School of Law, University of Reading, UK

Anna Alvazzi del Frate Small Arms Survey, Geneva, Switzerland

Kai Ambos Department of Foreign and International Criminal Law, Institute for Criminal Law and Justice, Georg-August-Universität Göttingen, Göttingen, Germany

Karen L. Amendola Police Foundation, Washington, DC, USA

Gail S. Anderson School of Criminology, Simon Fraser University, Burnaby, BC, Canada

Martin A. Andresen School of Criminology, Institute for Canadian Urban Research Studies, Simon Fraser University, Burnaby, Canada

Donald L. Anthony Program in Community Research and Action, Peabody College, Vanderbilt University, Nashville, TN, USA

Robert Apel School of Criminal Justice, Rutgers University, Newark, NJ, USA

Carol A. Archbold Department of Criminal Justice and Political Science, North Dakota State University, Fargo, ND, USA

Fabio Arcila Jr. Touro Law Center, Central Islip, NY, USA

Barak Ariel Faculty of Law, The Hebrew University, Jerusalem, Israel

Institute of Criminology, University of Cambridge, Cambridge, UK

Shawn Armbrust Mid-Atlantic Innocence Project, American University, Washington, DC, USA

Rachel Armitage Applied Criminology Centre, University of Huddersfield, Huddersfield, UK

Gaylene S. Armstrong College of Criminal Justice, Correctional Management Institute of Texas, Sam Houston State University, Huntsville, TX, USA

Louise Aronson Division of Geriatrics, University of California San Francisco, San Francisco, CA, USA

Bruce A. Arrigo Department of Criminal Justice and Criminology, University of North Carolina at Charlotte, Charlotte, NC, USA

Cassandra A. Atkin-Plunk Sam Houston State University, Huntsville, TX, USA

Leena Augimeri Centre for Children Commiting Offenses, Child Development Institute, Toronto, ON, Canada

Barbara A. Babb Center for Families, Children and the Courts, University of Baltimore School of Law, Baltimore, MD, USA

Tom Barker Department of Criminal Justice, College of Justice and Safety, Eastern Kentucky University, Richmond, KY, USA

Stephen W. Baron Department of Sociology, Queen’s University, Kingston, ON, Canada

Kelle Barrick RTI International, Research Triangle Park, NC, USA

David G. Barrie Discipline Group of History, The University of Western Australia, Crawley, WA, Australia

Jennifer L. Beaudry Swinburne University of Technology, Melbourne, Australia

Eric Beauregard School of Criminology, Simon Fraser University, Burnaby, BC, Canada

Kevin M. Beaver School of Criminology and Criminal Justice, Florida State University, Tallahassee, FL, USA

Center for Social and Humanities Research, King Abdulaziz University, Jeddah, Saudi Arabia

Ryan Bell Australian Centre for Arson Research and Treatment, Bond University, Robina, QLD, Australia

Roberta Belli UNDSS Somalia SIOC, Springette, Spring Valley, Nairobi, Kenya

Jyoti Belur University College London, London, UK

Craig Bennell Department of Psychology, Carleton University, Ottawa, Canada

Christopher Bennett Department of Philosophy, University of Sheffield, Sheffield, UK

Jessica Bennett JB National Investigations, Fairmont, WV, USA

Sarah Bennett Australian Research Council Centre of Excellence in Policing and Security (ARC), Institute for Social Science Research (ISSR), The University of Queensland, St. Lucia, QLD, Australia

Michael L. Benson School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA

Richard A. Berk Department of Criminology, University of Pennsylvania, Philadelphia, PA, USA

Giulia Berlusconi Università Cattolica del Sacro Cuore and Transcrime, Milan, Italy

Wim Bernasco Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, The Netherlands

Department of Spatial Economics, VU University Amsterdam, Amsterdam, The Netherlands

Jón Gunnar Bernburg Faculty of the Social and Human Sciences, University of Iceland, Reykjavík, Iceland

Heather Y. Bersot Department of Criminal Justice and Criminology, University of North Carolina at Charlotte, Charlotte, NC, USA

Alex Biedermann School of Criminal Justice, Institute of Forensic Science, University of Lausanne, Lausanne, Switzerland

Ingrid A. Binswanger Division of General Internal Medicine and Division of Substance Dependence, University of Colorado School of Medicine, Aurora, CO, USA

Denver Health Medical Center, Denver, CO, USA

Ivan Birch The University of Texas at Dallas, Richardson, TX, USA

Daniel Birks ARC Center of Excellence in Policing and Security, CEPS, Griffith University, Brisbane, Australia

Donna M. Bishop School of Criminology and Criminal Justice, Northeastern University, Boston, MA, USA

Beth Bjerregaard Department of Criminal Justice and Criminology, University of North Carolina at Charlotte, Charlotte, NC, USA

Marc Le Blanc School of Criminology and School of Psychoeducation, Université de Montréal and Boscoville 2000, Montreal, QC, Canada

Emmanuel Blanchard Centre de Recherches Sociologiques sur le Droit et les Institutions Pénales (CESDIP), Université de Versailles, Guyancourt, France

Stephen Bleay Centre for Applied Science and Technology, Home Office Science, Sandridge, Hertfordshire, UK

Marieke Bloembergen KITLV, Leiden, The Netherlands

Arjan A. J. Blokland Department of Criminology and Criminal Law, Leiden University, Leiden, The Netherlands

The Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, The Netherlands

Ulf Böckenholt Northwestern University, Evanston, IL, USA

Dankmar Böhning Statistical Sciences Research Institute, University of Southampton, Southampton, UK

Danielle Boisvert College of Criminal Justice, Sam Houston State University, Huntsville, TX, USA

Christian L. Bolden Criminal Justice, Loyola University New Orleans, New Orleans, LA, USA

Robert Boruch Graduate School of Education, University of Pennsylvania, Philadelphia, PA, USA

Ilaria Bottigliero International Development Law Organization (IDLO), Rome, Italy

Anthony Bottoms University of Cambridge, Cambridge, UK

University of Sheffield, Sheffield, UK

Leana A. Bouffard College of Criminal Justice, Sam Houston State University, Huntsville, TX, USA

Kate J. Bowers Department of Security and Crime Science, University College London, London, UK

Jordan Bowman William & Mary School of Law, Williamsburg, VA, USA

Ben Bradford Centre for Criminology, University of Oxford, Oxford, UK

Catherine P. Bradshaw Johns Hopkins Center for the Prevention of Youth Violence, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Elizabeth A. Bradshaw Department of Sociology, Anthropology and Social Work, Central Michigan University, Mt. Pleasant, MI, USA

Anthony A. Braga School of Criminal Justice, Rutgers University, Newark, NJ, USA

Kennedy School of Government, Harvard University, Cambridge, MA, USA

Valerie Braithwaite Regulatory Institutions Network (RegNet), The Australian National University, Acton, ACT, Australia

Robert Brame Department of Criminal Justice and Criminology, University of North Carolina at Charlotte, Charlotte, NC, USA

Steven G. Brandl Department of Criminal Justice, University of Wisconsin-Milwaukee, Milwaukee, WI, USA

Kathleen F. Brickey Washington University School of Law, Washington University, St. Louis, MO, USA

Avi Brisman School of Justice Studies, Eastern Kentucky University, Richmond, KY, USA

Chester L. Britt School of Criminology and Criminal Justice, Northeastern University, Boston, MA, USA

Ton Broeders Faculty of Law, Institute for Criminal Law and Criminology, Leiden University, Leiden, The Netherlands

Maastricht University, Maastricht, The Netherlands

Eileen M. Brooks College of Criminal Justice, Northeastern University, Boston, MA, USA

Jennifer Brown Mannheim Centre, Department of Social Policy, London School of Economics, London, UK

Kenly Brown Department of Sociology, University of California, Santa Barbara, CA, USA

Michelle Brown Department of Sociology, University of Tennessee, Knoxville, TN, USA

Angela Browne Vera Institute of Justice, Washington, DC, USA

Christopher R. Browning Department of Sociology, The Ohio State University, Columbus, OH, USA

Daniel Brown-Kenyon Institute for Social Science Research, Policing and Security, The University of Queensland, St. Lucia, QLD, Australia

Henry H. Brownstein NORC at The University of Chicago, Bethesda, MD, USA

Gerben Bruinsma Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, The Netherlands

VU University Amsterdam, Amsterdam, The Netherlands

Mark Brunger Department of Law and Criminal Justice, Canterbury Christ Church University, Canterbury, Kent, UK

Vaughn M. Bryant Department of Anthropology, Texas A&M University, College Station, TX, USA

Michael E. Buerger Bowling Green State University, Bowling Green, OH, USA

Peter A. Bull Oxford University Centre for the Environment, University of Oxford, Oxford, UK

Christina L. Bullard University of South Carolina Beaufort, Bluffton, SC, USA

Amanda Burgess-Proctor Criminal Justice, Oakland University, Rochester, MI, USA

Alafair S. Burke Maurice A. Deane School of Law, Hofstra University, Hempstead, NY, USA

William D. Burrell Lawrenceville, NJ, USA

Shawn D. Bushway School of Criminal Justice, University at Albany, State University of New York, Albany, NY, USA

Mark Button University of Portsmouth, Portsmouth, UK

Kathleen A. Cagney Departments of Sociology and Health Studies, The University of Chicago, Chicago, IL, USA

Francesco Calderoni Transcrime, Università Cattolica del Sacro Cuore, Milan, Italy

Jesse Cale Key Centre for Ethics, Law, Justice and Governance, Griffith University, Brisbane, QLD, Australia

Rob Canton Community and Criminal Justice, Health and Life Sciences, De Montfort University, Leicester, UK

I. Bennett Capers Brooklyn Law School, Brooklyn, NY, USA

Joel M. Caplan School of Criminal Justice, Rutgers University, Newark, NJ, USA

Elaine Carey Department of History, St. John’s University, Queens, NY, USA

Peter J. Carrington Department of Sociology and Legal Studies, University of Waterloo, Waterloo, ON, Canada

Richard F. Catalano Social Development Research Group, University of Washington, Seattle, WA, USA

Michael Caudy Criminology, Law and Society, George Mason University, Fairfax, VA, USA

Breanne Cave George Mason University, Fairfax, VA, USA

Javier A. Cepeda Center for Interdisciplinary Research on AIDS and Department of Epidemiology and Public Health, Yale School of Public Health, Yale University, New Haven, CT, USA

Spencer Chainey Jill Dando Institute of Security and Crime Science, University College London, London, UK

William J. Chambliss George Washington University, Washington, DC, USA

Steven L. Chanenson School of Law, Villanova University, Villanova, PA, USA

David Charlton Surrey and Sussex Police Forensic Services, School of Applied Sciences, Bournemouth University, Bournemouth, UK

Ali Chaudhary Department of Sociology, University of California Davis, Davis, CA, USA

Jorge M. Chavez Department of Sociology, Bowling Green State University, Bowling Green, OH, USA

Jehanzeb Cheema University of Illinois at Urbana-Champaign, Urbana, IL, USA

Steven M. Chermak School of Criminal Justice, Michigan State University, East Lansing, MI, USA

National Consortium for the Studies of Terrorism and Responses to Terrorism (START), University of Maryland, College Park, MD, USA

Stephen Chicoine Department of Sociology, University of South Carolina, Columbia, SC, USA

Yi-Ning Chiu School of Criminology and Criminal Justice, Griffith University, Brisbane, QLD, Australia

Brendan C. Clark Department of Psychiatry and Behavioral Neurobiology, University of Alabama at Birmingham, Birmingham, AL, USA

Shawna Cleary School of Criminal Justice, University of Central Oklahoma, Edmond, OK, USA

Stephen Clipper The University of Texas at Dallas, Richardson, TX, USA

Alison K. Cohen Health and Human Development Program, WestEd, Oakland, CA, USA

Simon A. Cole Department of Criminology, Law & Society, University of California, Irvine, Irvine, CA, USA

Steven Le Comber School of Biological and Chemical Sciences, Queen Mary, University of London, London, UK

Eric J. Connolly Florida State University, Tallahassee, FL, USA

David P. Connor Justice Administration, University of Louisville, Louisville, KY, USA

Philip J. Cook Sanford School of Public Policy, Duke University, Durham, NC, USA

NBER, Cambridge, MA, USA

David J. Cooke Glasgow Caledonian University, Glasgow, UK

University of Bergen, Bergen, Norway

Barry S. Cooper The Forensic Alliance, The University of British Columbia, Simon Fraser University and The Forensic Psychiatric Services Commission, Vancouver, BC, Canada

Heith Copes University of Alabama at Birmingham, Birmingham, AL, USA

Gary Cordner Department of Criminal Justice, Kutztown University, Kutztown, PA, USA

Nicholas Corsaro School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA

Marina L. Costanzo University of Colorado Colorado Springs, Colorado Springs, CO, USA

Mark A. Costanzo Department of Psychology, Claremont McKenna College, Claremont, CA, USA

Stacy De Coster Department of Sociology and Anthropology, North Carolina State University, Raleigh, NC, USA

Hazel Croall School of Business and Society, Glasgow Caledonian University, Glasgow, UK

Karen L. Cropsey Department of Psychiatry and Behavioral Neurobiology, University of Alabama at Birmingham, Birmingham, AL, USA

Robert D. Crutchfield University of Washington, Seattle, WA, USA

Maarten Cruyff Department of Methodology and Statistics, Utrecht University, Utrecht, The Netherlands

Kelly R. Damphousse University of Oklahoma, Norman, OK, USA

Andrew L. B. Davies School of Criminal Justice, University at Albany, State University of New York, Albany, NY, USA

Jacqueline Davis Australian Research Council Centre of Excellence in Policing and Security (ARC), Institute for Social Science Research (ISSR), The University of Queensland, St. Lucia, QLD, Australia

Jason Davis Criminal Justice, Clayton State University, Morrow, GA, USA

Michael R. Davis Centre for Forensic Behavioural Science, Monash University, Melbourne, VIC, Australia

Robert C. Davis RAND Corporation, Arlington, VA, USA

Pascal Décarpes University of Greifswald, Greifswald, Germany

Mathieu Deflem Department of Sociology, University of South Carolina, Columbia, SC, USA

Walter S. DeKeseredy Institute of Technology, University of Ontario, ON, Canada

Matt DeLisi Criminal Justice Studies, Iowa State University, Ames, IA, USA

Olivier Delémont Ecole des Sciences Criminelles, University of Lausanne, Lausanne, Switzerland

Quentin Deluermoz Quentin Deluermoz, Université Paris 13, Sorbonne Paris Cité, CRESC/Pleiade, Paris, France

Matthew DeMichele Justice Center for Research, Department of Sociology, Penn State University, University Park, PA

Roger L. Depue FBI Retired

James H. Derzon Battelle Memorial Institute, Centers for Public Health Research and Evaluation, Arlington, VA, USA

Rashaan A. DeShay School of Economic, Political, and Policy Sciences, The University of Texas at Dallas, Richardson, TX, USA

Jan Van Dijk Human Security and Safety International Victimology Institute, Tilburg University, Tilburg, The Netherlands

Anja J. E. Dirkzwager Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, The Netherlands

Sandra Ferrari Disner Department of Linguistics, University of Southern California, Los Angeles, CA, USA

David Dixon Faculty of Law, University of New South Wales, Sydney, NSW, Australia

Jo Dixon Department of Sociology, New York University, New York, NY, USA

Jonathan Doak Durham Law School, Durham University, Durham, UK

Rebecca Emerson Dobash School of Law, University of Manchester, Manchester, UK

School of Social Justice and Inquiry, Arizona State University, Glendale, AZ, USA

Centre for Law, Crime and Justice, School of Law, Strathclyde University, Glasgow, Scotland

Russell P. Dobash School of Law, University of Manchester, Manchester, UK

School of Social Justice and Inquiry, Arizona State University, Glendale, AZ, USA

Centre for Law, Crime and Justice, School of Law, Strathclyde University, Glasgow, Scotland

Mary Dodge School of Public Affairs, University of Colorado Denver, Denver, CO, USA

Alan Doig International Development Department, The University of Birmingham, Edgbaston, Birmingham, UK

Rebekah Doley Australian Centre for Arson Research and Treatment, Bond University, Robina, QLD, Australia

Jennifer R. Dolin University of South Carolina Beaufort, Bluffton, SC, USA

Alfonso Donoso Instituto de Ciencia Política, Pontificia Universidad Católica de Chile, Santiago Metropolitan Region, Chile

James M. Doyle Independent Researcher, Carney & Bassil, Boston, MA, USA

Francesco Drago University Federico II Naples and CSEF, Napoli, Italy

Kevin M. Drakulich School of Criminology and Criminal Justice, Northeastern University, Boston, MA, USA

Kirstin Drenkhahn Department of Law, Freie Universitaet Berlin, Berlin, Germany

Itiel E. Dror University College London and Cognitive Consultants International, London, UK

Laura Dugan Department of Criminology, University of Maryland, College Park, MD, USA

Dora M. Dumont The Center for Prisoner Health and Human Rights, The Miriam Hospital/Brown University, Providence, RI, USA

Frieder Dünkel Faculty of Law, University of Greifswald, Greifswald, Germany

Ioan Durnescu Faculty of Sociology and Social Work, University of Bucharest, Bucharest, Romania

Jennifer Dysart Department of Psychology, John Jay College of Criminal Justice, New York, NY, USA

Marleen Easton Faculty of Economics and Business Administration, Ghent University, Ghent, Belgium

Ronald J. Eckert Department of Sociology and Criminal Justice, Villanova University, Villanova, PA, USA

Gary Edmond Network for Interdisciplinary Studies of Law (NISL), School of Law, The University of New South Wales, Sydney, NSW, Australia

Nicole Egli Anthonioz School of Criminal Sciences, University of Lausanne, Lausanne, Switzerland

Tomer Einat Bar-Ilan University, Ramat-Gan, Israel

Bryan D. Eisentraut School of Criminology and Criminal Justice, Arizona State University, Phoenix, AZ, USA

Manuel Eisner Institute of Criminology, University of Cambridge, Cambridge, UK

Paul Ekblom Design Against Crime Research Centre, Central Saint Martins College of Arts and Design, University of the Arts London, London, UK

Gloria D. Eldridge Center for Behavioral Health Research and Services, University of Alaska at Anchorage, Anchorage, AK, USA

Henk Elffers Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, The Netherlands

Anthony Ellis Department of Philosophy, Virginia Commonwealth University, Richmond, VA, USA

Lee Ellis Department of Anthropology and Sociology, University of Malaya, Kuala Lumpur, Malaysia, Malaysia

Graham Ellison Institute of Criminology and Criminal Justice, School of Law, Queen’s University Belfast, Belfast, UK

V. H. Elmer College of Criminal Justice, Northeastern University, Boston, MA, USA

Katja Eman Faculty of Criminal Justice and Security, University of Maribor, Maribor, Slovenia

Traqina Q. Emeka University of Houston-Downtown, Houston, TX, USA

Clive Emsley Department of History, The Open University, Milton Keynes, UK

Horst Entorf Department of Economics and Business Administration, Goethe University Frankfurt, Frankfurt, Germany

Edna Erez Department of Criminology, Law, and Justice, University of Illinois at Chicago, Chicago, IL, USA

Judith van Erp Erasmus School of Law, Criminology Department, Rotterdam, The Netherlands

Alison Evans Cuellar George Mason University, Fairfax, VA, USA

Andrew Evans School of Geography, University of Leeds, Leeds, UK

Jacqueline R. Evans The University of Texas at Tyler, Tyler, TX, USA

Martin Paul Evison Northumbria University Centre for Forensic Science, Northumbria University, Newcastle upon Tyne, UK

Anne R. Van Ewijk Department of Political and Social Sciences, GRITIM (Interdisciplinary Research Group on Immigration), Universitat Pompeu Fabra, Barcelona, Spain

Samhoud, Utrecht, The Netherlands

M. Lyn Exum Department of Criminal Justice and Criminology, University of North Carolina at Charlotte, Charlotte, NC, USA

Lorna Fadden Department of Linguistics, Simon Fraser University, Burnaby, BC, Canada

Susan Fahey Criminal Justice Program, Richard Stockton College of New Jersey, Galloway, NJ, USA

Seth W. Fallik College of Criminal Justice, Sam Houston State University, Huntsville, TX, USA

Graham Farrell Institute for Canadian Urban Research Studies and School of Criminology, Simon Fraser University, Burnaby, BC, Canada

Stephen Farrall University of Sheffield, Sheffield, UK

David P. Farrington Institute of Criminology, University of Cambridge, Cambridge, UK

Katie Feavel School of Criminology and Criminal Justice, Arizona State University, Huntsville, TX, USA

Lynette Feder College of Health and Public Affairs, University of Central Florida, Orlando, FL, USA

Barry C. Feld University of Minnesota Law School, Minneapolis, MN, USA

Marcus Felson School of Criminal Justice, Texas State University-San Marcos, San Marcos, TX, USA

Roxana Ferllini Institute of Archaeology, University College London, London, UK

April Fernandes University of Washington, Seattle, WA, USA

Jeff Ferrell Texas Christian University, Fort Worth, TX, USA

University of Kent, Canterbury, UK

Cassi L. Fields Fields Consulting Group, Inc., McLean, VA, USA

LeT Corporation

James O. Finckenauer School of Criminal Justice, Rutgers University, Newark, NJ, USA

Bonnie S. Fisher School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA

Victor Eugene Flango National Center for State Courts, Williamsburg, VA, USA

John Flatley Crime Statistics and Analysis Division, Office for National Statistics, London, UK

James Alan Fox School of Criminology and Criminal Justice, Northeastern University, Boston, MA, USA

Brian Francis Department of Maths and Statistics, Lancaster University, Lancaster, UK

Joshua D. Freilich Jay College of Criminal Justice, New York, NY, USA

National Consortium for the Studies of Terrorism and Responses to Terrorism (START)

Steven J. Frenda University of California, Irvine, Irvine, CA, USA

Adrienne Freng Department of Criminal Justice, University of Wyoming, Laramie, WY, USA

Kathleen Frey Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA

Susan Friedman Mid-Atlantic Innocence Project, American University, Washington, DC, USA

David O. Friedrichs Department of Sociology/Criminal Justice, University of Scranton, Scranton, PA, USA

Katarina Fritzon Australian Centre for Arson Research and Treatment, Bond University, Robina, QLD, Australia

Trevor Fronius Learning Innovations, WestEd, Woburn, MA, USA

Adrian Furnham Division of Psychology and Language Sciences, Faculty of Brain Sciences, University College London, London, UK

Vasco Furtado Graduate Program in Applied Informatics, University of Fortaleza (Unifor), Fortaleza, Ceará, Brazil

Robert E. Gaensslen Forensic Science Group, Department of Biopharmaceutical Sciences, College of Pharmacy, University of Illinois at Chicago, Chicago, IL, USA

Roberto Galbiati Department of Economics, CNRS-OSC and Sciences Po, Paris, France

Shaun M. Gann School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA

Nuno Garoupa College of Law, University of Illinois at Urbana-Champaign, Champaign, IL, USA

Jacinta M. Gau Department of Criminal Justice, University of Central Florida, Orlando, FL, USA

Julie Gawrylowicz Department of Psychology, Royal Holloway, University of London, Egham, UK

Gilbert Geis School of Social Ecology, University of California, Irvine, Irvine, CA, USA

Karen Gelb Sentencing Advisory Council, Melbourne, Victoria, Australia

G. Geltner University of Amsterdam, Amsterdam, The Netherlands

Charlotte Gerritsen Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, The Netherlands

Jennifer Gibbs West Chester University at Pennsylvania, West Chester, PA, USA

Chris L. Gibson Department of Sociology and Criminology & Law, University of Florida, Gainesville, FL, USA

Charlotte E. Gill Center for Evidence-Based Crime Policy, George Mason University, Fairfax, VA, USA

Gwladys Gilliéron Distance Learning University Switzerland, Brig, Switzerland

Svenja Göbbels Maastricht University, Maastricht, The Netherlands

Toby S. Goldbach Cornell University Law School, Ithaca, NY, USA

Erich Goode Sociology Professor Emeritus, Stony Brook University, Stony Brook, NY, USA

Alasdair M. Goodwill Department of Psychology, Ryerson University, Toronto, BC, Canada

Denise C. Gottfredson Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA

Heather R. Gough School of Social Transformation, Arizona State University, Tempe, AZ, USA

Ioanna Gouseti Department of Methodology, London School of Economics, London, UK

Rick Grannis Department of Sociology, University of California, Irvine, Irvine, CA, USA

Sarah Greathouse RAND Corporation, Arlington, VA, USA

Jack R. Greene School of Criminology and Criminal Justice, Northeastern University, Boston, MA, USA

B. K. Greener Politics Programme, School of People, Environment, and Planning, Massey University, Palmerston North, New Zealand

Chris Greer Department of Sociology, City University London, London, UK

Adam Gregory Serious Organised Crime Agency, London, UK

Elizabeth Griffiths School of Criminal Justice, Rutgers University, Newark, NJ, USA

Maria S. Grigoryeva Department of Sociology, University of Washington, Seattle, WA, USA

Elizabeth R. Groff Temple University, Philadelphia, PA, USA

Garrett Grothoff School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA

Jeff Gruenewald Department of Sociology and Criminal Justice, University of Arkansas, Fayetteville, AR, USA

Ben Grunwald Department of Criminology, University of Pennsylvania, Philadelphia, PA, USA

Rob T. Guerette School of Criminal Justice, Florida International University, Miami, FL, USA

Fiona Haines School of Social and Political Sciences, University of Melbourne, VIC, Australia

Camilla Hald Danish National Police, Research & Development, Copenhagen, Denmark

Matthew Hall School of Law, University of Sheffield, Sheffield, UK

Mark Hamm Department of Criminology and Criminal Justice, Indiana State University, Terre Haute, IN, USA

Paula Hannaford-Agor Center for Jury Studies, National Center for State Courts, Williamsburg, VA, USA

Valerie P. Hans Cornell University Law School, Ithaca, NY, USA

Roger A. Hanson National Center for State Courts, Williamsburg, VA, USA

Bernard E. Harcourt Law School, The University of Chicago, Chicago, IL, USA

Alon Harel Faculty of Law, The Hebrew University of Jerusalem, Jerusalem, Israel

Rachel Harmon University of Virginia School of Law, Charlottesville, VA, USA

Danielle Arlanda Harris Justice Studies, San Jose State University, San Jose, CA, USA

Nathan Harris Regulatory Institutions Network, Australian National University, Acton, ACT, Australia

Rita Haverkamp Max Planck Institute for Foreign and International Criminal Law, Freibur, Germany

J. David Hawkins Social Development Research Group, University of Washington, Seattle, WA, USA

Dana L. Haynie Department of Sociology, The Ohio State University, Columbus, OH, USA

Đurđica Hazard School of Criminal Justice, University of Lausanne, Lausanne, Switzerland

Bill Hebenton Centre for Criminology and Criminal Justice, School of Law, University of Manchester, Manchester, UK

Jon Heidt School of Criminology and Criminal Justice, University of the Fraser Valley, Abbotsford, BC, Canada

Peter G. M. van der Heijden Department of Methodology and Statistics, Utrecht University, Utrecht, The Netherlands

University of Southampton, Southampton, UK

Karen Heimer Department of Sociology, The University of Iowa, Iowa City, IA, USA

Alana Henninger John Jay College of Criminal Justice, New York, NY, USA

Bryn Ann Herrschaft Department of Research, Center for Court Innovation, New York, NY, USA

Hugues F. Hervé The Forensic Alliance, Port Moody, BC, Canada

Joshua C. Hinkle Department of Criminal Justice and Criminology, Georgia State University, Atlanta, GA, USA

Andreas von Hirsch Cambridge University, Cambridge, UK

Goethe University Frankfurt, Frankfurt, Germany

Dick Hobbs University of Essex, Essex, UK

Andrew Hochstetler Department of Sociology, Iowa State University, Ames, IA, USA

Barbora Hola Department of Criminology, VU University Amsterdam, Amsterdam, The Netherlands

Simon Holdaway School of Law, Centre for Criminological Research, The University of Sheffield, Sheffield, UK

Meghan Hollis-Peel Northeastern University and University of Massachusetts, Boston, MA, USA

Kristy Holtfreter School of Criminology and Criminal Justice, Arizona State University, Phoenix, AZ, USA

Robert Hopper JB National Investigations, Fairmont, WV, USA

Jonathan C. Hoskin Clemson Computing and Information Technology, Clemson University, Clemson, SC, USA

Lawrence Houston The Pennsylvania State University, University Park, PA, USA

L. Rowell Huesmann Institute for Social Research, University of Michigan, Ann Arbor, MI, USA

Wim Huisman School of Criminology, VU University Amsterdam, Amsterdam, The Netherlands

Jennifer C. Hunt Montclair State University, Montclair, NJ, USA

Neil Hutton Centre for Law, Crime & Justice, School of Law, Strathclyde University, Glasgow, Scotland

Jonathan Jackson Department of Methodology and Mannheim Centre for Criminology, London School of Economics, London, UK

John D. Jackson School of Law, University of Nottingham, Nottingham, UK

Rick Jacobs Department of Psychology, The Pennsylvania State University and EB Jacobs, University Park, PA, USA

Sonia Jain Health and Human Development Program, WestEd, Oakland, CA, USA

Jesse Jannetta Urban Institute, Washington, DC, USA

Fabien Jobard Centre de Recherches Sociologiques sur le Droit et les Institutions Pénales (CESDIP), CNRS, Guyancourt, France

Brian Johnson Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA

Shane D. Johnson Department of Security and Crime Science, University College London, London, UK

Darrick Jolliffe School of Law, University of Greenwich, London, UK

Tal Jonathan-Zamir Institute of Criminology, Faculty of Law, The Hebrew University of Jerusalem, Jerusalem, Israel

Susyan Jou Graduate School of Criminology, National Taipei University, Taipei, Taiwan

Roberta Julian Tasmanian Institute of Law Enforcement Studies, University of Tasmania, Tasmania, Australia

Marianne Junger Twente University, Enschede, The Netherlands

Robert J. Kane Program in Criminal Justice, Drexel University, Philadelphia, PA, USA

Kristiina Kangaspunta United Nations Office on Drugs and Crime (UNODC), Vienna, Austria

Takhmina Karimova Geneva Academy of International Humanitarian Law and Human Rights, Geneva, Switzerland

David Kauzlarich Department of Sociology and Criminal Justice Studies, Southern Illinois University Edwardsville, Edwardsville, IL, USA

Lila Kazemian John Jay College of Criminal Justice, New York, NY, USA

Ingo Keilitz CourtMetrics, Williamsburg, VA, USA

Gregory F. Kellermeyer Denver Health Medical Center and Division of Forensic Psychiatry, Department of Psychiatry, University of Colorado Denver School of Medicine, Denver, CO, USA

George Kelling Rutgers University, Newark, NJ, USA

Christopher E. Kelly University at Albany, State University of New York, Albany, NY, USA

Sally Kelty Tasmanian Institute of Law Enforcement Studies, University of Tasmania, Tasmania, Australia

Jasper J. van der Kemp Department of Criminal Law and Criminology, VU University Amsterdam, Amsterdam, The Netherlands

Leslie W. Kennedy School of Criminal Justice, Rutgers University, Newark, NJ, USA

Sesha Kethineni Department of Criminal Justice Sciences, Illinois State University, IL, USA

Stuart A. Kinner Melbourne School of Population and Global Health, The University of Melbourne and School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

School of Medicine, The University of Queensland, Brisbane, Australia

Janne Kivivuori Criminological Unit, National Research Institute of Legal Policy, Helsinki, Finland

Gary Kleck College of Criminology and Criminal Justice, Florida State University, Tallahassee, FL, USA

Edward R. Kleemans VU University Amsterdam, Amsterdam, The Netherlands

Matthew Kleiman National Center for State Courts, Williamsburg, VA, USA

Joanne Klein Department of History, Boise State University, Boise, ID, USA

Michel C. A. Klein Department of Computer Science, VU University Amsterdam, Amsterdam, The Netherlands

Ellen Klinkers KITLV, Leiden, The Netherlands

Melanie Klinkner Business School, Bournemouth University, Bournemouth, Dorset, UK

Paul Knepper School of Law, University of Sheffield, Sheffield, UK

Tammy Rinehart Kochel Southern Illinois University, Carbondale, IL, USA

Christopher S. Koper Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA

Vere M. van Koppen VU University Amsterdam, Amsterdam, The Netherlands

Margo De Koster Vrije Universiteit Brussel, VU University, Brussels, Belgium

VU University Amsterdam, Amsterdam, The Netherlands

Ronald C. Kramer Department of Sociology, Western Michigan University, Kalamazoo, MI, USA

Derek A. Kreager Department of Sociology, The Pennsylvania State University, University Park, PA, USA

Jonathan Allen Kringen School of Criminal Justice, Texas State University-San Marcos, San Marcos, TX, USA

Charis E. Kubrin Criminology, Law and Society, University of California, Irvine, Irvine, CA, USA

Landon B. Kuester Centre for Research on Drugs and Health Behavior, London School of Hygiene and Tropical Medicine, London, UK

Annette Kuhlmann University of Wisconsin, Baraboo, WI, USA

Margaret R. Kuklinski Social Development Research Group, University of Washington, Seattle, WA, USA

Megan C. Kurlychek School of Criminal Justice, University at Albany, State University of New York, Albany, NY, USA

Helmut Kury University of Freiburg, Freiburg, Germany

Sanja Kutnjak Ivković School of Criminal Justice, Michigan State University, East Lansing, MI, USA

Sharon Ingrid Kwok City University of Hong Kong, Kowloon, Hong Kong

Peter van der Laan Netherlands Institute for the Study of Crime and Law Enforcement (NSCR) and Faculty of Law, VU University Amsterdam, Amsterdam, The Netherlands

Nicola Lacey All Souls College Oxford, University of Oxford, Oxford, UK

Tapio Lappi-Seppälä National Research Institute of Legal Policy, Helsinki, Finland

A. Karl Larsen Forensic Science Group, Department of Biopharmaceutical Sciences, College of Pharmacy, University of Illinois at Chicago, Chicago, IL, USA

Edward J. Latessa School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA

Pamela K. Lattimore Crime, Violence, and Justice, RTI International, Research Triangle Park, NC, USA

Andrew R. Laurence Department of Anthropology, Texas A&M University, College Station, TX, USA

Julia Lavenberg University of Pennsylvania Medical Center, Philadelphia, Pennsylvania, USA

Christopher Lawless University of Edinburgh, Edinburgh, UK

Paul Lawrence Department of History, The Open University, Milton Keynes, UK

Victoria Z. Lawson Department of Psychology, John Jay College of Criminal Justice and The Graduate Center, The City University of New York, New York, NY, USA

William B. Lawson Department of Psychiatry and Behavioral Sciences, Howard University Health Sciences, Washington, DC, USA

Gloria Laycock Jill Dando Institute of Security and Crime Science, University College London, London, UK

Benoit Leclerc School of Criminology and Criminal Justice, Griffith University, Brisbane, QLD, Australia

Cynthia G. Lee National Center for State Courts, Williamsburg, VA, USA

Daniel R. Lee Department of Criminology, Indiana University of Pennsylvania, Indiana, PA, USA

Michael J. Leiber Department of Criminology, University of South Florida, Tampa, FL, USA

Richard A. Leo University of San Francisco School of Law, San Francisco, CA, USA

Chrysanthi Leon Sociology and Criminal Justice, University of Delaware, Newark, DE, USA

Joanne van der Leun Department of Criminal Law and Criminology, Faculty of Law, Leiden University, Leiden, The Netherlands

S. C. Leung Scientific Consultancy Limited, Hong Kong, China

Kathy Levene Child Development Institute, Toronto, ON, Canada

Michael Levi Centre for Crime, Law and Justice at Cardiff School of Social Sciences, Cardiff University, Cardiff, Wales, UK

Kay Levine Emory University School of Law, Atlanta, GA, USA

Marissa P. Levy School of Social and Behavioral Sciences, Richard Stockton College of New Jersey, Galloway, NJ, USA

Claire Lightowler Institute for Research and Innovation in Social Services, Glasgow, UK

Antonia Linde School of Criminal Sciences, Institute of Criminology and Criminal Law, University of Lausanne, Lausanne, Switzerland

Open University of Catalonia, Barcelona, Spain

Christine H. Lindquist Crime, Violence, and Justice, RTI International, Research Triangle Park, NC, USA

Weiwei Liu NORC at The University of Chicago, Bethesda, MD, USA

Caleb D. Lloyd Department of Psychology, Carleton University, Ottawa, ON, Canada

Shawnta L. Lloyd Rollins School of Public Health, Emory University, Atlanta, GA, USA

T. Wing Lo City University of Hong Kong, Kowloon, Hong Kong

Eric Lock Brigade de Police Technique et Scientifique, Police Cantonale, Geneva, Switzerland

Rolf Loeber Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA

Elizabeth F. Loftus University of California, Irvine, Irvine, CA, USA

Ross London Berkeley College, New York, NY, USA

D. Kall Loper Department of Computer Science and Engineering, Lyle School of Engineering, Southern Methodist University, Dallas, TX, USA

Nicholas Lord School of Social Sciences, Cardiff University, Cardiff, Wales, UK

Friedrich Lösel Institute of Criminology, University of Cambridge, Cambridge, UK

Institute of Psychology, University of Erlangen–Nuremberg, Erlangen, Germany

Jennifer Eno Louden The University of Texas at El Paso, El Paso, TX, USA

Thomas A. Loughran Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA

Barry Loveday Institute of Criminal Justice Studies, University of Portsmouth, Portsmouth, UK

Brian Lovins School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA

Christopher T. Lowenkamp Administrative Office of the U.S. Courts, Washington, DC, USA

Hong Lu University of Nevada, Las Vegas, Las Vegas, NV, USA

Douglas M. Lucas The Centre of Forensic Sciences Ontario, Burlington, ON, Canada

Cynthia Lum Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA

Erik Luna Washington and Lee University School of Law, Washington and Lee University, Lexington, VA, USA

Arthur J. Lurigio Department of Psychology, Loyola University Chicago, Chicago, IL, USA

Patrick Lussier Laval University, Quebec City, QC, Canada

Stephanie Maass Criminology, Law and Society, George Mason University, Fairfax, VA, USA

Anna Macdonald Department of War Studies, King's College London, London, UK

John M. MacDonald Department of Criminology, University of Pennsylvania, Philadelphia, PA, USA

Doris Layton MacKenzie Justice Center for Research, The Pennsylvania State University, University Park, PA, USA

David Maimon Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA

Nick Malleson School of Geography, University of Leeds, Leeds, UK

Michael D. Maltz Department of Sociology, Criminal Justice Research Center, The Ohio State University, Columbus, OH, USA

Emeritus, Department of Criminal Justice and Department of Information and Decision Sciences, University of Illinois at Chicago, Chicago, IL, USA

Sarah M. Manchak School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA

Matthew Manning School of Criminology and Criminal Justice, Griffith University, Mt. Gravatt, QLD, Australia

Peter K. Manning College of Criminal Justice, Northeastern University, Boston, MA, USA

Otwin Marenin Department of Criminal Justice and Criminology, Washington State University, Pullman, WA, USA

Pierre Margot School of Criminal Justice, Institute of Forensic Science, University of Lausanne, Lausanne, Switzerland

Ineke Haen Marshall Department of Sociology and Anthropology and School of Criminology and Criminal Justice, Northeastern University, Boston, MA, USA

Shadd Maruna Queen's University Belfast, Belfast, UK

Gary T. Marx Massachusetts Institute of Technology, Cambridge, MA, USA

Brent D. Mast Office of Policy Development and Research, U.S. Department of Housing and Urban Development, Washington, DC, USA

Sarah van Mastrigt Department of Psychology and Behavioral Sciences, University of Aarhus, Aarhus, Denmark

Giovanni Mastrobuoni Collegio Carlo Alberto, IZA, Netspar, Moncalieri, Italy

Stephen D. Mastrofski George Mason University, Manassas, VA, USA

Katherine E. Masyn Harvard Graduate School of Education, Cambridge, MA, USA

Matt Matravers Morrell Centre for Toleration, University of York, York, UK

Kristy N. Matsuda Criminology and Criminal Justice, University of Missouri-St. Louis, St. Louis, MO, USA

Ross L. Matsueda Department of Sociology, University of Washington, Seattle, WA, USA

Rick A. Matthews Carthage College, Kenosha, WI, USA

Roger Matthews University of Kent, Canterbury, Kent, UK

Rob I. Mawby University of South Wales, Newport, South Wales, UK

Department of Natural Sciences, University of Gloucestershire, Cheltenham, UK

David May Mississippi State University, MS, USA

Pat Mayhew Crime and Justice Research Centre, Victoria University of Wellingon, Wellington, New Zealand

Lorraine Mazerolle Institute for Social Science Research, The University of Queensland, St. Lucia, QLD, Australia

Paul Mazerolle Arts, Education, and Law, Griffith University, Griffith, QLD, Australia

Logan M. McBride The Graduate Center, The City University of New York, New York, NY, USA

James T. McCafferty Department of Sociology and Criminal Justice, Kennesaw State University, Kennesaw, GA, USA

Bill McCarthy Department of Sociology, University of California, Davis, Davis, CA, USA

Ron McCarthy Director Emeritus, National Tactical Officers Association, Los Angeles Police Department (Retired), Doylestown, PA, USA

Carole McCartney School of Law, University of Leeds, Leeds, UK

Richard McCleary School of Social Ecology, University of California, Irvine, Irvine, CA, USA

John D. McCluskey Department of Criminal Justice, Rochester Institute of Technology, Rochester, NY, USA

Jude McCulloch Arts Faculty, Monash University, Clayton, Victoria, Australia

David McDowall School of Criminal Justice, University at Albany, State University of New York, Albany, NY, USA

Jenny McEwan School of Law, Exeter University, Exeter, Devon, UK

Troy McEwan Centre for Forensic Behavioural Science, Monash University and the Victorian Institute of Forensic Mental Health, Melbourne, Australia

Edmund F. McGarrell School of Criminal Justice, Michigan State University, East Lansing, MI, USA

James McGuire Institute of Psychology, Health, and Society, Division of Clinical Psychology, University of Liverpool, Liverpool, UK

Sarah J. McLean John F. Finn Institute for Public Safety, Inc., Albany, NY, USA

Fergus McNeill Scottish Centre for Crime and Justice Research, University of Glasgow, Glasgow, UK

Tracey L. Meares Yale Law School, Yale University, New Haven, CT, USA

Robert T. Meesig Melbourne, FL, USA

Robert Meier School of Criminology and Criminal Justice, University of Nebraska Omaha, Omaha, NE, USA

Christian A. Meissner The University of Texas at El Paso, El Paso, TX, USA

Amina Memon Department of Psychology, Royal Holloway, University of London, Egham, UK

Guadalupe Mendiola-Washington University of Houston-Downtown, Houston, TX, USA

Wyatt Merritt University of Nevada, Las Vegas, Las Vegas, NV, USA

Gorazd Meško Faculty of Criminal Justice and Security, University of Maribor, Maribor, Slovenia

Lauren C. Messina Rollins School of Public Health, Emory University, Atlanta, GA, USA

Steven F. Messner Department of Sociology, Arts, and Sciences, University at Albany, State University of New York, Albany, NY, USA

Silke Meyer Institute for Social Science Research, The University of Queensland, St. Lucia, QLD, Australia

Tom Mijares Texas State University-San Marcos, San Marcos, TX, USA

Erin Miller University of Maryland/START Consortium, College Park, MD, USA

Andrew Millie Department of Law and Criminology, Edge Hill University, Ormskirk, Lancashire, UK

Ojmarrh Mitchell Department of Criminology, University of South Florida, Tampa, FL, USA

Terrie E. Moffitt Departments of Psychology and Neuroscience, Psychiatry, and Behavioral Sciences, Institute for Genome Sciences and Policy, Duke University, Durham, NC, USA

J. Monckton-Smith University of Gloucestershire, Cheltenham, UK

George T. Monteleone Center for Cognitive and Social Neuroscience, The University of Chicago, Chicago, IL, USA

Kathryn D. Morgan Department of Justice Sciences, The University of Alabama at Birmingham, Birmingham, AL, USA

Ruth M. Morgan Security and Crime Science, University College London, London, UK

Christine Morgenstern Ernst Moritz Arndt Universität Greifswald, Greifswald, Germany

Katherine Morris Harvard University, Cambridge, MA, USA

Robert G. Morris The University of Texas at Dallas, Richardson, TX, USA

Christian Mouhanna Centre de Recherches Sociologiques sur le Droit et les Institutions Pénales (CESDIP), CNRS, Guyancourt, France

Giulia Mugellini Institute of Criminology, University of Zurich, Zurich, Switzerland

Rebecca Mugford Department of Psychology, Carleton University, Ottawa, ON, Canada

Kristina Murphy Griffith University, Brisbane, QLD, Australia

Chongmin Na School of Human Science and Humanities, University of Houston-Clear Lake, Houston, TX, USA

Daniel S. Nagin Carnegie Mellon University, Pittsburgh, PA, USA

Joseph L. Nedelec School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA

David Nelken Cardiff University Law School, Cardiff, UK

Karen A. Newirth The Innocence Project, New York, NY, USA

Eryn J. Newman Victoria University of Wellington, Wellington, New Zealand

Graeme R. Newman School of Criminal Justice, University at Albany, State University of New York, Albany, NY, USA

Jamie Newsome Department of Criminal Justice, The University of Texas at San Antonio, San Antonio, TX, USA

Andrew Newton Applied Criminology Centre, University of Huddersfield, Huddersfield, UK

Holly Nguyen Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA

Amanda S. Nicholson Department of Psychology, John Jay College of Criminal Justice, The City University of New York, New York, NY, USA

Paul Nieuwbeerta Department of Criminology and Criminal Law, Leiden University, Leiden, The Netherlands

Utrecht University, Utrecht, The Netherlands

Faye Nitschke Australian Research Council Centre of Excellence in Policing and Security (ARC), Institute for Social Science Research (ISSR), The University of Queensland, St. Lucia, QLD, Australia

Kenneth J. Novak Department of Criminal Justice and Criminology, University of Missouri, Kansas City, MO, USA

William O'Brian School of Law, University of Warwick, Coventry, UK

James R. P. Ogloff Centre for Forensic Behavioural Science, Monash University, Melbourne, VIC, Australia

Daniel Ohana The Hebrew University, Jerusalem, Israel

Tel Aviv University, Tel Aviv, Israel

Michael M. O'Hear Marquette University Law School, Milwaukee, WI, USA

David E. Olson Department of Criminal Justice and Criminology, Loyola University Chicago, Chicago, IL, USA

David O'Mahony Durham University, Durham, UK

Pat O'Malley Sydney Law School, The University of Sydney, NSW, Australia

Brian J. Ostrom National Center for State Courts, Williamsburg, VA, USA

Nicola Padfield Faculty of Law, University of Cambridge, Cambridge, UK

Letizia Paoli Institute of Criminology, University of Leuven, Leuven, Belgium

Eugene A. Paoline III Department of Criminal Justice, University of Central Florida, Orlando, FL, USA

Andrew V. Papachristos Department of Sociology, Yale University, New Haven, CT, USA

Karen F. Parker Department of Sociology and Criminal Justice, University of Delaware, Newark, DE, USA

Robert Nash Parker University of California, Riverside, Riverside, CA, USA

William Parkin Seattle University, Seattle, WA, USA

Roger B. Parks Indiana University Bloomington, Bloomington, IN, USA

Stephan Parmentier Institute of Criminology, University of Leuven, Leuven, Belgium

Raymond Paternoster Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA

Michele Pathé Queensland Forensic Mental Health Service and Griffith University, Brisbane, Australia

Irene Pavesi Small Arms Survey, Geneva, Switzerland

Allison Ann Payne Department of Sociology and Criminal Justice, Villanova University, Villanova, PA, USA

Kenneth J. Peak Department of Criminal Justice, University of Nevada, Reno, Reno, NV, USA

Ken Pease Jill Dando Institute, University College London, London, UK

Loughborough University, Loughborough, UK

Jennifer H. Peck Department of Criminology, University of South Florida, Tampa, FL, USA

Steven D. Penrod Department of Psychology, John Jay College of Criminal Justice, The City University of New York, New York, NY, USA

Jennifer T. Perillo John Jay College of Criminal Justice and The Graduate Center, The City University of New York, New York, NY, USA

Douglas D. Perkins Program in Community Research and Action, Peabody College, Vanderbilt University, Nashville, TN, USA

Barbara Perry Faculty of Social Science and Humanities, University of Ontario Institute of Technology, North Oshawa, ON, Canada

Simon Perry Institute of Criminology, Faculty of Law, The Hebrew University, Jerusalem, Israel

Thomas Søbirk Petersen Department of Philosophy and Science Studies, Roskilde University, Roskilde, Denmark

Joan Petersilia Stanford Law School, Stanford University, Stanford, CA, USA

Hanno Petras Research and Development, JBS International, North Bethesda, MD, USA

K. V. Petrides Research Department of Clinical, Educational, and Health Psychology, Educational Psychology Group, University College London, London, UK

Anthony Petrosino Learning Innovations, WestEd, Woburn, MA, USA

Carolyn Petrosino Bridgewater State University, Bridgewater, MA, USA

Nickie D. Phillips Department of Sociology and Criminal Justice, St. Francis College, Brooklyn, NY, USA

Nathan W. Pino Department of Sociology, Texas State University-San Marcos, San Marcos, TX, USA

Paolo Pinotti Università Bocconi, Milan, Italy

Alex R. Piquero Department of Criminology, The University of Texas at Dallas, Richardson, TX, USA

Nicole Leeper Piquero The University of Texas at Dallas, Richardson, TX, USA

Jenny C. Piquette Department of Sociology, The Ohio State University, Columbus, OH, USA

Greg Pogarsky School of Criminal Justice, University at Albany, State University of New York, Albany, NY, USA

Paul Ponsaers Faculty of Law, Research Group SVA, Ghent University, Ghent, Belgium

Chad Posick School of Criminology and Criminal Justice, Northeastern University, Boston, MA, USA

Hillary Potter Department of Sociology, University of Colorado Boulder, Boulder, CO, USA

Travis C. Pratt School of Criminology and Criminal Justice, Arizona State University, Phoenix, AZ, USA

Tim Prenzler Australian Research Council Centre of Excellence in Policing and Security, School of Criminology and Criminal Justice, Griffith University, Brisbane, Australia

Ineke Pruin Department of Criminology, University of Greifswald, Greifswald, Germany

Lesleigh Pullman University of Ottawa, Ottawa, ON, Canada

David Pyrooz College of Criminal Justice, Sam Houston State University, Phoenix, AZ, USA

Steven M. Radil Department of Geography, Ball State University, Muncie, IN, USA

Myrna Raeder Southwestern Law School, Los Angeles, CA, USA

Lee Rainbow Serious Organised Crime Agency, London, UK

Erin Rajca Child Development Institute, Toronto, ON, Canada

Marieke van de Rakt Avans University of Applied Sciences, Hertogenbosch, The Netherlands

Kevin Ralston Sociology and Criminal Justice, University of Delaware, Newark, DE, USA

Suzanna Ramirez The University of Queensland, St. Lucia, Australia

Janet Ransley School of Criminology and Criminal Justice, Griffith University, Mt. Gravatt, QLD, Australia

Jerry H. Ratcliffe Temple University, Philadephia, PA, USA

Allison D. Redlich University at Albany, State University of New York, Albany, NY, USA

Michael H. Reggio

Robert Reiner Department of Law, London School of Economics and Political Science, London, UK

Herbert Reinke Bergische Universität Wuppertal, Wuppertal, Germany

Michael Rempel Drug Courts, Criminal Justice Intervention, Center for Court Innovation, New York, NY, USA

Callie Marie Rennison School of Public Affairs, University of Colorado Denver, Denver, CO, USA

Danielle M. Reynald School of Criminology and Criminal Justice, Griffith University, Griffith, Australia

Anne Giurianna Rhodes Department of Criminology, Law and Society, George Mason University, Richmond, VA, USA

William Rhodes Abt Associates Inc., Cambridge, MA, USA

Olivier Ribaux Ecole des Sciences Criminelles, University of Lausanne, Lausanne, Switzerland

Michele Riccardi Transcrime, Joint Research Centre on Transnational Crime, Università Cattolica del Sacro Cuore di Milano, Milan, Italy

Josiah D. Rich The Center for Prisoner Health and Human Rights, The Miriam Hospital/Brown University, Providence, RI, USA

Jason Roach Crime and Policing Group, University of Huddersfield, Huddersfield, West Yorkshire, UK

Kent Roach Faculty of Law, University of Toronto, Toronto, ON, Canada

Julian V. Roberts Faculty of Law, Centre for Criminology, University of Oxford, Oxford, UK

Paul Roberts School of Law, University of Nottingham, Nottingham, UK

James Robertson National Centre for Forensic Studies (NCFS), Faculty of Education, Science, Engineering, Technology and Maths, University of Canberra, Canberra, ACT, Australia

Charles Robinson Administrative Office of the U.S. Courts, Washington, DC, USA

Gwen Robinson Reader in Criminal Justice, School of Law, University of Sheffield, Sheffield, UK

Michael Rocque Department of Sociology, University of Maine, Orono, ME, USA

Richard Rosenfeld University of Missouri-St. Louis, St. Louis, MO, USA

Jacqueline E. Ross University of Illinois College of Law, Champaign, IL, USA

Jeffrey Ian Ross School of Criminal Justice, College of Public Affairs, University of Baltimore, Baltimore, MD, USA

D. Kim Rossmo School of Criminal Justice, Texas State University-San Marcos, San Marcos, TX, USA

Thomas Roth NS-Dokumentationszentrum der Stadt Köln, Köln, Germany

Dawn L. Rothe Department of Sociology and Criminal Justice, Old Dominion University, Norfolk, VA, USA

David B. Rottman National Center for State Courts, Williamsburg, VA, USA

Claude Roux University of Technology, Sydney, Centre for Forensic Science, Broadway, NSW, Australia

Danielle S. Rudes George Mason University, Fairfax, VA, USA

Vincenzo Ruggiero School of Law, Middlesex University, London, UK

Stijn Ruiter Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, The Netherlands

Amanda L. Russell The University of Texas at Dallas, Richardson, TX, USA

Jesper Ryberg Department of Philosophy and Science Studies, Roskilde University, Roskilde, Denmark

Wenona Rymond-Richmond Department of Sociology, University of Massachusetts Amherst, Amherst, MA, USA

Mari Sakiyama University of Nevada, Las Vegas, Las Vegas, NV, USA

Gerhard Sälter Documentation and Research, Berlin Wall Memorial and UHK/BND, Philipps-Universität Marburg, Berlin, Germany

Rachel Boba Santos School of Criminology and Criminal Justice, Florida Atlantic University, Boca Raton, FL, USA

Roberto Santos Professional Standards Division, Port St. Lucie, Police Department, Port St. Lucie, FL, USA

Maria Sapouna School of Social Sciences, University of the West of Scotland, Hamilton, UK

Ryan Sarasin School of Justice Studies, Roger Williams University, Bristol, RI, USA

Federica Sarno Transcrime, Joint Research Centre on Transnational Crime, Università Cattolica del Sacro Cuore di Milano and Università degli Studi di Trento, Milan, Italy

Joachim J. Savelsberg Department of Sociology, University of Minnesota, Minneapolis, MN, USA

Ernesto U. Savona Transcrime, Joint Research Centre on Transnational Crime, Università Cattolica del Sacro Cuore di Milano and Università degli Studi di Trento, Milan, Italy

Joseph A. Schafer Department of Criminology and Criminal Justice, Southern Illinois University, Carbondale, IL, USA

Barry C. Scheck The Innocence Project, New York, NY, USA

Conrad Schetter Bonn International Center for Conversion (BICC), Bonn, Germany

Martin Schmucker Institute of Psychology, University of Erlangen–Nuremberg, Erlangen, Germany

Christopher J. Schreck Department of Criminal Justice, Rochester Institute of Technology, Rochester, NY, USA

Anja Schröder Office of the Prosecutor, International Criminal Court, The Hague, The Netherlands

Dorothy Moses Schulz John Jay College of Criminal Justice, The City University of New York, New York, NY, USA

Joseph A. Schwartz School of Criminology and Criminal Justice, Florida State University, Tallahassee, FL, USA

Martin D. Schwartz George Washington University, Washington, DC, USA

Michael S. Scott Law School, University of Wisconsin, Madison, WI, USA

Leslie Sebba Institute of Criminology, Faculty of Law, The Hebrew University, Mt. Scopus, Jerusalem, Israel

Revital Sela-Shayovitz David Yellin Academic College and The Hebrew University, Jerusalem, Israel

Ralph C. Serin Department of Psychology, Carleton University, Ottawa, ON, Canada

Michael C. Seto Integrated Forensic Program, Royal Ottawa Health Care Group, Ottawa, ON, Canada

Karen Sewell Child Development Institute, Toronto, ON, Canada

Joanna Shapland School of Law, University of Sheffield, Sheffield, UK

Lawrence W. Sherman Institute of Criminology, Cambridge University, Cambridge, UK

University of Maryland, Baltimore, MD, USA

Christopher Shields Terrorism Research Center in Fulbright College, University of Arkansas, Fayetteville, AR, USA

Haia Shpayer-Makov Department of History, University of Haifa, Haifa, Israel

Rashi K. Shukla School of Criminal Justice, University of Central Oklahoma, Edmond, OK, USA

Aiden Sidebottom Department of Security and Crime Science, University College London, London, UK

Sally S. Simpson Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA

Megan Sims Department of Criminal Justice and Criminology, University of North Carolina at Charlotte, Charlotte, NC, USA

Simon I. Singer School of Criminology and Criminal Justice, Northeastern University, Boston, MA, USA

Jennifer L. Skeem University of California, Irvine, Irvine, CA, USA

John J. Sloan III Department of Justice Sciences, The University of Alabama at Birmingham, Birmingham, AL, USA

Lee Ann Slocum Department of Criminology and Criminal Justice, University of Missouri – St. Louis, St. Louis, MO, USA

Michel De Smedt Office of the Prosecutor, International Criminal Court, The Hague, The Netherlands

Monika Smit Administration of Justice, Legislations, and International and Aliens Affairs (RWI), Research and Documentation Centre (WODC) of the Dutch Ministry of Security and Justice

Paul R. Smit Ministry of Safety and Justice, Research and Documentation Centre, The Hague, The Netherlands

Brent L. Smith Department of Sociology and Criminal Justice, University of Arkansas, Fayetteville, AR, USA

Chris M. Smith Department of Sociology, University of Massachusetts Amherst, Amherst, MA, USA

Gavin J. D. Smith Senior Lecturer in Sociology, The Australian National University, Canberra, ACT, Australia

Martha J. Smith School of Community Affairs, Wichita State University, Wichita, KS, USA

Brian Soller Department of Sociology, University of New Mexico, Albuquerque, NM, USA

Ragnhild Sollund Department of Criminology and Sociology of Law, Faculty of Law, University of Oslo, Oslo, Norway

Keith Soothill Department of Maths and Statistics, Lancaster University, Lancaster, UK

William Sousa University of Nevada, Las Vegas, Las Vegas, NV, USA

Nigel South Department of Sociology, University of Essex, Colchester, Essex, UK

Toine Spapens Tilburg University, Tilburg, The Netherlands

Anne C. Spaulding Rollins School of Public Health, Emory University, Atlanta, GA, USA

Sandra A. Springer Center for Interdisciplinary Research on AIDS and Yale School of Medicine, Yale University, New Haven, CT, USA

Elizabeth Spruin International Research Centre for Investigative Psychology (IRCIP), University of Huddersfield, Huddersfield, UK

Janet P. Stamatel Department of Sociology, University of Kentucky, Lexington, KY, USA

Elizabeth Stanley School of Social and Cultural Studies, Victoria University of Wellington, Wellington, New Zealand

Richard Stansfield Department of Sociology and Criminal Justice, University of Delaware, Newark, DE, USA

Richard Staring Department of Criminology, Faculty of Law, Erasmus University, Rotterdam, The Netherlands

Carl S. Stauffer Center for Justice and Peacebuilding, Eastern Mennonite University, Harrisonburg, VA, USA

Eric Stauffer School of Criminal Justice, University of Lausanne, Lausanne, Switzerland

Crime Scene Unit and Crime Laboratory, Fribourg State Police, Fribourg, Switzerland

Mark Stevenson School of Biological and Chemical Sciences, Queen Mary, University of London, London, UK

Anna Stewart School of Criminology and Criminal Justice, Griffith University, Brisbane, Australia

Reinoud D. Stoel Netherlands Forensic Institute, The Hague, The Netherlands

Bas van Stokkom Radboud University Nijmegen, Nijmegen, The Netherlands

VU University Amsterdam, Amsterdam, The Netherlands

Magda Stouthamer-Loeber Department of Psychiatry, Medical School, University of Pittsburgh, Pittsburgh, PA, USA

Kevin J. Strom RTI International, Research Triangle Park, NC, USA

Karen Heetderks Strong

Jeanne Subjack College of Criminal Justice, Sam Houston State University, Huntsville, TX, USA

Christopher J. Sullivan School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA

Lucia Summers School of Criminal Justice, Texas State University-San Marcos, San Marcos, TX, USA

Hung-En Sung Department of Criminal Justice, John Jay College of Criminal Justice, New York, NY, USA

Mike Sutton School of Social Sciences, Nottingham Trent University, Nottingham, UK

Rachel Swaner Center for Court Innovation, New York, NY, USA

Gary Sweeten Arizona State University, Phoenix, AZ, USA

Justice Tankebe Institute of Criminology, University of Cambridge, Cambridge, UK

Lusine Tarkhanyan Department of Security and Crime Science, University College London, London, UK

Franco Taroni School of Criminal Justice, Institute of Forensic Science, University of Lausanne, Lausanne, Switzerland

Faye S. Taxman Criminology, Law and Society, George Mason University, Fairfax, VA, USA

Cody W. Telep School of Criminology and Criminal Justice, Arizona State University, Phoenix, AZ, USA

Jan Terpstra Institute of Criminology and Criminal Justice, Radboud University Nijmegen, Nijmegen, The Netherlands

Richard Tewksbury Justice Administration, University of Louisville, Louisville, KY, USA

Jo Thakker School of Psychology, University of Waikato, Hamilton, New Zealand

Delphine Theobald Institute of Psychiatry, King's College, London, UK

Institute of Criminology, University of Cambridge, Cambridge, UK

George C. Thomas III Rutgers University School of Law, Newark, New Jersey, NJ, USA

Kyle J. Thomas Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA

Terence P. Thornberry University of Maryland, College Park, MD, USA

Stephen G. Tibbetts Department of Criminal Justice, California State University, San Bernardino, San Bernardino, CA, USA

Nick Tilley Jill Dando Institute of Security and Crime Science, University College London, London, UK

William J. Tilstone Centre for Forensic Science, University of Strathclyde, Glasgow, Scotland, UK

Victoria B. Titterington College of Criminal Justice, Sam Houston State University, Huntsville, TX, USA

Michael Townsley School of Criminology and Criminal Justice, Griffith University, Brisbane, QLD, Australia

Lawrence F. Travis III School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA

Christian Traxler Department of Economics, University of Marburg, Marburg, Germany

Max Planck Institute for Research on Collective Goods, Bonn, Germany

Ruth Triplett Department of Sociology and Criminal Justice, Old Dominion University, Norfolk, VA, USA

Andromachi Tseloni School of Social Sciences, Nottingham Trent University, Nottingham, UK

Maria M. Ttofi Institute of Criminology, University of Cambridge, Cambridge, UK

Jillian J. Turanovic School of Criminology and Criminal Justice, Arizona State University, Phoenix, AZ, USA

Justin Turner Old Dominion University, Norfolk, VA, USA

Michael G. Turner Department of Criminal Justice and Criminology, University of North Carolina at Charlotte, Charlotte, NC, USA

Tom R. Tyler Yale Law School, New Haven, CT, USA

Craig D. Uchida Justice and Security Strategies, Inc., Silver Spring, MD, USA

Ambros Uchtenhagen Research Institute for Public Health and Addiction, University of Zurich Medical School, Zurich, Switzerland

Jeffery T. Ulmer The Pennsylvania State University, University Park, PA, USA

Brigitte Unger Utrecht University School of Economics, Utrecht, The Netherlands

Abby Vandenberg School of Criminology and Criminal Justice, University of Nebraska at Omaha, Omaha, NE, USA

Sean P. Varano School of Justice Studies, Roger Williams University, Bristol, RI, USA

Jennifer Varriale Carson Department of Criminal Justice, University of Central Missouri, Warrensburg, MO, USA

María B. Vélez Department of Sociology, University of New Mexico, Albuquerque, NM, USA

Antoinette Verhage Faculty of Law, Research Group SVA, Ghent University, Ghent, Belgium

Lynne Vieraitis The University of Texas at Dallas, Richardson, TX, USA

Jill Viglione George Mason University, Fairfax, VA, USA

Kees van der Vijver University of Twente, Enschede, The Netherlands

William J. Vizzard Division of Criminal Justice, California State University, Sacramento, Sacramento, CA, USA

Alison Wakefield University of Portsmouth, Portsmouth, UK

Sandra Walklate Department of Sociology, Social Policy and Criminology, University of Liverpool, Liverpool, UK

Danielle Wallace School of Criminology and Criminal Justice, Arizona State University, Phoenix, AZ, USA

Roy Walmsley World Prison Brief, The International Centre for Prison Studies, London, UK

Margaret Walsh Child Development Institute, Toronto, ON, Canada

Glenn D. Walters Kutztown University of Pennsylvania, Kutztown, PA, USA

Reece Walters School of Justice, Faculty of Law, Queensland University of Technology, Brisbane, Australia

Rasmus H. Wandall Faculty of Law, University of Bergen, Bergen, Norway

Emily A. Wang Yale School of Medicine, New Haven, CT, USA

Tony Ward School of Psychology, Victoria University of Wellington, Wellington, New Zealand

Nicole L. Waters National Center for State Courts, Williamsburg, VA, USA

Frank M. Weerman Netherlands Institute for the Study of Crime and Law Enforcement (NCSR), Amsterdam, The Netherlands

David Weisburd Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA

Faculty of Law, The Hebrew University, Mt. Scopus, Jerusalem, Israel

Douglas B. Weiss University of Maryland, College Park, MD, USA

Elmar G. M. Weitekamp University of Tuebingen, Tuebingen, Germany

Ronald Weitzer Department of Sociology, George Washington University, Washington, DC, USA

Brandon C. Welsh School of Criminology and Criminal Justice, Northeastern University, Boston, MA, USA

Netherlands Institute for the Study of Crime and Law Enforcement (NCSR), Amsterdam, The Netherlands

Hilde Wermink Department of Criminology and Criminal Law, Leiden University, Leiden, The Netherlands

David B. Wexler International Network on Therapeutic Jurisprudence, University of Puerto Rico School of Law, San Juan, Puerto Rico

University of Arizona, Tucson, AZ, USA

Rob White School of Social Sciences, University of Tasmania, Hobart, TAS, Australia

Matthias Wienroth Centre for Forensic Science, Northumbria University, Newcastle Upon Tyne, UK

Per-Olof H. Wikström Institute of Criminology, University of Cambridge, Cambridge, UK

Brie Williams Division of Geriatrics, University of California San Francisco, San Francisco, CA, USA

Kirk R. Williams Department of Sociology and Criminal Justice, University of Delaware, Newark, DE, USA

Robin Williams Centre for Forensic Science, Northumbria University, Newcastle Upon Tyne, UK

Gwenda M. Willis University of Auckland, School of Psychology, Auckland, New Zealand

James J. Willis Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA

David B. Wilson Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA

Dean Wilson Criminology and Criminal Justice, School of Law, Plymouth University, Plymouth, UK

Ronald E. Wilson Office of Policy Development and Research, U.S. Department of Housing and Urban Development, Washington, DC, USA

Sandra Jo Wilson Peabody Research Institute, Vanderbilt University, Nashville, TN, USA

Andreas Wimmer Department of Sociology, Princeton University, Princeton, NJ, USA

Jan M. Winter National Police Services Agency, Zoetermeer, The Netherlands

Jessica Woodhams School of Psychology, University of Birmingham, Birmingham, UK

Alissa Pollitz Worden School of Criminal Justice, University at Albany, State University of New York, Albany, NY, USA

Robert E. Worden School of Criminal Justice, University at Albany, State University of New York, Albany, NY, USA

John F. Finn Institute for Public Safety, Inc., Albany, NY, USA

John L. Worrall School of Economic, Political, and Policy Sciences, The University of Texas at Dallas, Richardson, TX, USA

Richard Wortley Jill Dando Institute of Security and Crime Science, University College London, London, UK

Hal S. Wortzel Division of Forensic Psychiatry, Department of Psychiatry, University of Colorado Denver School of Medicine and VISN–19 MIRECC, Denver Veterans Affairs Medical Center, Denver, CO, USA

John Paul Wright School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA

King Abdulaziz University, Jeddah, Saudi Arabia

Ronald F. Wright Wake Forest University, Winston-Salem, NC, USA

Timbre Wulf-Ludden School of Criminology and Criminal Justice, University of Nebraska Omaha, Omaha, NE, USA

Sue-Ming Yang Department of Criminology, National Chung-Cheng University, Min-Hsiung, Chia-Yi, Taiwan

Angela M. Yarbrough Department of Psychology, John Jay College of Criminal Justice, The City University of New York, New York, NY, USA

Donna Youngs International Research Centre for Investigative Psychology (IRCIP), University of Huddersfield, Huddersfield, UK

John C. Yuille The Forensic Alliance, The University of British Columbia, Salt Spring Island, BC, Canada

Marvin Zalman Department of Criminal Justice, Wayne State University, Detroit, MI, USA

Marjorie S. Zatz School of Social Transformation, Arizona State University, Tempe, AZ, USA

Don Zettlemoyer Penn State Justice and Safety Institute, The Pennsylvania State University, University Park, PA, USA

Gregory M. Zimmerman School of Criminology and Criminal Justice, Northeastern University, Boston, MA, USA

Dirk van Zyl Smit University of Nottingham, Nottingham, UK

A

1%ers

► Motorcycle Clubs or Gangs?

ABM

► Agent-Based Assessments of Criminological Theory

Accomplice Crime

► Co-offending and Offender Attributes

Actualizing Risk-Need-Responsivity

Faye S. Taxman, Michael Caudy and
Stephanie Maass
Criminology, Law and Society, George Mason
University, Fairfax, VA, USA

Synonyms

Principles of effective correctional intervention; Principles of effective intervention; RNR; RNR model; Treatment matching

Overview

The Risk-Need-Responsivity (RNR) model for correctional programming has served as a framework for promoting the use of evidence-based correctional strategies. The RNR framework has three core principles: (1) identify and target the static risk level of the individual, (2) identify the dynamic risk factors (needs) that are associated with offending behavior and that affect psychosocial functioning, and (3) identify appropriate correctional interventions that are suitable to address the risk-need interaction. Adherence to all three of these principles promises a greater impact on recidivism and other individual offender outcomes. This essay provides a brief overview of (1) risk assessment, (2) dynamic offender needs, and (3) responsivity.

The Risk-Need-Responsivity Model

The core components of the RNR Model are:

- **Risk**: Match the intensity of service to the offender's static risk level. Target moderate and higher-risk cases for more intensive treatments and controls. Generally, avoid creating interactions of low-risk cases with higher-risk cases.
- **Need**: Target criminogenic needs (needs that are directly related to recidivism) predominately. Define offender needs both in terms of deficits but also in terms of strengths to improve overall psychosocial functioning.

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

- **General Responsivity**: Employ evidence-based behavioral, social learning, and cognitive-behavioral treatment strategies.
- **Specific Responsivity**: Adapt the style and model of service according to the setting of the service and to relevant characteristics of individual offenders, such as their strengths, motivations, preferences, personality, age, gender, ethnicity, cultural identifications, and other factors.

The RNR framework essentially focuses on reducing recidivism through providing effective interventions that are targeted to the risk and needs of people involved in the justice system. According to the model, the focus of interventions should be on individual-level factors (needs) that directly contribute to criminal behavior. The model also stresses that certain types of interventions are more likely to have recidivism-reducing effects than others. Application of the Risk-Need-Responsivity framework requires that criminal justice agencies (1) use a validated risk assessment instrument to identify static risk and prioritize justice-involved individuals for more intensive services and controls, (2) identify dynamic offender needs (e.g., antisocial cognitions, antisocial peers, and substance abuse) and target rehabilitative interventions to specific offender needs that are related to recidivism, and (3) utilize evidence-based treatment practices grounded in behavioral and cognitive-behavioral models. The model stresses the importance of taking a human services approach to correctional rehabilitation. It is grounded in an extensive empirical literature base that indicates that adherence to the three core principles of the model is related to improved offender outcomes (see Andrews and Bonta 2010 for a review).

Criminal Justice Risk and Risk Assessment

The first ingredient in the Risk-Need-Responsivity model is static risk level. Risk in this context refers specifically to an offender's likelihood of reoffending and is operationalized by indicators of prior criminal involvement. Since the 1920s, there has been a growing awareness of the factors that predict the likelihood of reoffending and the usefulness of these factors in identifying of risk for recidivism. Risk assessment tools are designed to measure the degree to which the individual is likely to have negative outcomes (e.g., more recidivism) during or after experience with the justice system. There are a number of standard risk assessment tools such as the Wisconsin Risk and Need Instrument, the Level of Service Inventory-Revised (LSI-R), the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), the Ohio Risk Assessment System (ORAS), and a myriad of other tools, some in the public domain and others proprietary. Each of these tools has some standard reference to historical involvement with the justice system (i.e., takes into account criminal history). The use of standardized risk assessment tools is an evidence-based best practice in the field of corrections and is a central component of the Risk-Need-Responsivity framework.

Risk assessment is not a new concept in the field of corrections. The use of interviews and assessments that collect data on individuals and predict how they will do during or after involvement with the justice system has been commonplace within the criminal justice system for nearly 100 years. When reviewing what is known about risk prediction in the justice system, it is typical to trace the development of risk assessments through four generations. The four generations of risk assessment are briefly reviewed here:

- **First Generation**: **Clinical Assessment**. First-generation risk assessments rely solely on professional judgment. Depending on the skill of the interviewer, the nature of the decision being made, and the degree to which the offender is forthcoming in the interview, the clinical interview can generate valid information about an individual's risk to reoffend. But, the interview process has been critiqued due to the potential for bias related to subjective factors (e.g., age, race, gender, and other demographics of the offender) that may affect the decision of the interviewer and/or system. This criticism has led to concerns about introducing systematic bias into the risk assessment process. First-generation clinical assessments are based on criminal justice history (generally referred to arrest history or rap sheet) along with unspecified characteristics of interest of interviewers. The risk of bias in these professional judgments limits their utility for risk prediction. Accordingly, actuarial measures of risk are now preferred over professional judgments and have become fairly standard within justice settings.

- **Second Generation**: **Actuarial Risk Assessments**. The second stage in the development of risk assessment techniques involves an actuarial (or statistical) risk calculation. Beginning with parole hearing officers in the 1920s, there was a growing realization that prior involvement with the justice system can be a strong predictor of future involvement with the justice system (recidivism). Key criminal history risk factors such as number of prior arrests, number of incarcerations, attempts to escape, behavior in the institution, prior probation violations, and age of first arrest have been found over the past 90 years to be consistent predictors of continued involvement in the justice system. Interviewer or system biases regarding the demographic characteristics of the offender are minimized by the actuarial approach to risk assessment. Since the statistical assessment is generally demographically blind – it does not include gender, age, or race factors – it minimizes the potential that justice workers will allow bias regarding these factors to impact decisions. Actuarial risk assessments can also be used to improve resource allocation. Consistent with the risk principle of the RNR model, risk assessment tools can be used to identify an offender's risk level, and higher-risk offenders can be placed in more intensive and more control-oriented correctional programs. This resource allocation model represents a practical use of the risk assessment tool and makes up a central component of the RNR model. Risk assessment can have an impact on recidivism through the use of assessment tools to assign higher-risk offenders to more intensive programming.
- **Third Generation**: **Actuarial Risk and Dynamic Risk** (**Needs**). The third generation of risk assessment tools added in dynamic risk factors to improve the alignment between correctional programming and offender needs. Dynamic risk factors are offender characteristics that are amenable to change. In the Risk-Need-Responsivity model, when dynamic risk factors are directly related to recidivism, they are labeled "criminogenic needs." Beginning with the Wisconsin Risk and Needs Assessment Tool, this third generation of risk assessment instruments added needs or psychosocial factors into the equation of risk assessment. Dynamic risk factors may include antisocial attitudes and orientation, employment, substance abuse, living arrangements, mental health status, leisure time activities, associations with criminal peers, and other areas that have a hypothesized correlation with offending. Third-generation risk assessment instruments have two means by which to assess outcomes: historical risk factors (static risk) and dynamic offender needs (dynamic risk). By including dynamic factors in the risk assessment equation, third-generation tools are better equipped to identify targets (treatment needs) for correctional programming. Identifying dynamic needs is an essential component of both the need and responsivity principles of the RNR model.
- **Fourth Generation**: **Case Management Through Risk and Need Assessment**. Fourth-generation risk and need assessment tools are an extension of third-generation tools with a focus on treatment matching/ case management. In these instruments, risk and need factors are used to identify appropriate services (treatments and controls) to reduce the risk of recidivism. These tools focus on addressing offender needs by creating service matching algorithms based on the risk and need profile of the offender. Fourth-generation tools are designed to improve the quality of assessments and to make them more applicable to offender case management and resource allocation by identifying individual-specific behavioral targets for intervention. Fourth-generation tools identify specific offender needs and making treatment placement recommendations.

The use of a validated risk-need assessment is an essential piece of the Risk-Need-Responsivity framework. Identifying static risk is the first step in applying the framework and should occur as early as possible after an offender comes in contact with the justice system. Static risk is generally the strongest predictor of recidivism; however, this construct is limited from a case management perspective because it does not identify factors that are amenable to change. While second-generation risk assessment instruments are sufficient for identifying static risk and subsequently predicting recidivism, they fail to consider dynamic offender needs that can be addressed through programming. Accordingly, the use of a third- or fourth-generation risk and need assessment is recommended by the RNR model.

Dynamic Risk and Criminogenic Needs

In the Risk-Need-Responsivity framework, there is recognition that an individual's current situation – dynamic factors – also influences his or her involvement in criminal behavior. For the most part, these factors, such as substance abuse, mental health, employment retention, pro-social values, friends and families, and criminal thinking, are amenable to change. Because these dynamic factors are amenable to change, they are identified by the model as the targets for rehabilitative correctional interventions. The questions are: which of these needs are criminogenic (needs that are directly related to offending behavior and recidivism), and which are more related to psychosocial functioning as opposed to future offending? Criminogenic needs have a direct link to recidivism, whereas other types of psychosocial functioning needs have an indirect link. The need principle of the RNR model stresses the importance of triaging dynamic offender needs to focus correctional rehabilitation on those needs that are most directly related to recidivism. This process in intended to save resources by focusing interventions on the treatable characteristics of the offender that have an impact on recidivism and can be changed through programming.

The Risk-Need-Responsivity model identifies eight key dynamic needs: history of antisocial behavior, antisocial personality pattern, antisocial attitudes/thinking, antisocial associates, family/marital problems, low levels of education or financial achievement, lack of pro-social leisure activities, and substance abuse. These factors are labeled as the "Central Eight" dynamic risk factors for continued involvement in offending behaviors. The first four needs – history of antisocial behavior, antisocial attitudes, antisocial peers, and criminal personality – are considered to be more predictive of recidivism outcomes than the remaining factors. These four "criminogenic" needs have been labeled as the "Big Four" by the creators of the RNR model. According to the model, these are the dynamic factors that should be the primary target of correctional treatments and interventions. The need principle of the RNR model stresses the importance of using screening and assessment instruments to identify dynamic offender needs to prioritize targets for correctional treatment and interventions.

The following discussion briefly describes each dynamic need included in the RNR model as well as a discussion of a few other key dynamic risk factors that were omitted from the model in its original statement. The goal is to distinguish between criminogenic and non-criminogenic needs, both important facets in a human service-driven model of recidivism reduction. This discussion extends the RNR model beyond its original conceptualization and takes into account empirical evidence regarding the link between dynamic risk factors and recidivism outcomes.

Substance Abuse. The RNR framework identifies substance abuse as a dynamic, but not "Big Four," criminogenic need. The framework generally neglects the fact that there are different types of drug users, and the linkage to criminal conduct varies considerably by substance use disorder problem severity and by drug of choice. Some substance abuse behaviors drive criminal behavior while other use behaviors have negligible or no impact on offending behavior. Most criminal justice risk assessments include substance abuse related questions but fail to distinguish between substance abuse and dependency.

For example, in the Wisconsin Risk-Need Instrument and the Level of Service Inventory-Revised (LSI-R) (among others), the substance abuse questions generally ask whether the individual has ever had a problem with substance abuse. In the Ohio Risk Assessment System (ORAS), substance use is measured with items that reflect lifetime use, longest period of abstinence, and employment-related problems stemming from drug use. All of these assessments are limited in their capacity to identify current substance use disorders and current substance use treatment needs.

Validated substance use screening and assessment tools, such as the Addiction Severity Index, the TCU Drug Screen, and others, tend to garner a better understanding of the offender's pattern of drug use and misuse. These instruments include more detailed information about lifetime use of drugs, drug of choice, mode of delivery (e.g., smoking, injection), longest periods of abstinence, and whether drug use affects employment or is related to legal problems. In general, the risk-need screening instruments used in the justice system tend to cast a large net to describe an array of substance use behaviors, some of which are pertinent to recent use and predicting recidivism, some of which are not.

The drug use and crime relationship (the drug-crime nexus) is complicated. In general, empirical research suggests that there is a link between drug use and offending. The strength of this relationship is conditioned by both the type of drugs being used and the severity of the substance use disorder. Because of this varying relationship between drug use and offending, it is important to consider both drug of choice and clinical diagnosis/level of treatment need when factoring substance use into the Risk-Need-Responsivity framework.

The drug-crime nexus theory suggests that involvement in criminal behavior will be disrupted by providing substance abuse treatment in the justice system especially for those users for whom drug use directly affects criminal behavior. The RNR framework does not differentiate on severity of substance use or on drug of choice. A revisited model of RNR would prioritize substance use treatment needs based on substance use disorder (SUD) severity and drug of choice since there are evidence-based interventions available to address this dynamic need. There is a clear link between dependency on hard drugs (e.g., cocaine, opiates, and amphetamines) and offending which makes hard drug dependence a criminogenic need that should be prioritized for treatment in the justice system.

Based on the extant empirical research, the following refinements to the consideration of substance use in a RNR framework are suggested:

1. Some types of drug use do not impact criminal behavior. The literature illustrates that opiates and cocaine use are more likely to influence criminal conduct through the need to acquire money, to participate in the drug trade to support one's own habit, and the decisions made by an individual in the inebriated state. There is increasing evidence that amphetamine use falls into this category as well.
2. Marijuana use does not appear to follow the same trend as other "hard drugs" (i.e., cocaine, opiates, amphetamines), and therefore marijuana abuse should not be considered a criminogenic need. Marijuana users with criminal thinking patterns should be prioritized for interventions that can address their antisocial thinking needs rather than their substance use issues.
3. The offense should not drive a determination of a substance abuse treatment need since many offenders are arrested on charges that are not related to substance abuse, and many times substance abuse is a secondary driver of criminal behavior. Additionally, many offenders arrested for drug-related offenses do not have clinical SUDs that necessitate intensive treatments and therefore are unlikely to benefit from substance abuse treatment programs.
4. Offenders should be divided into four categories regarding substance abuse: (a) nonusers, (b) used drugs in the past with current use of soft drugs, (c) used hard drugs within a month of the arrest, and (d) used hard drugs prior to the arrest. This classification is suggested

because most studies and assessment tools examine these issues, but they do not include factors that are used in clinical assessments (such as the DSM-IV). This pattern is also more similar to proposed changes in the DSM-V which focuses on dependence disorders and tolerance to substances.

Antisocial Cognitions/Criminal Thinking. Antisocial cognitions are the means by which people can rationalize their deviant behavior to neutralize or reduce the negative consequences resulting from offending. Typical thinking errors include dominance, entitlement, self-justification, displacing blame, optimistic perceptions of realities, and "victim stance" (e.g., blaming society because they are considered outcasts). Offenders tend to exhibit more of these thinking errors than members of the general population. Criminal thinking or antisocial cognitions are included in many criminological theories including subcultural, anomie, differential association, control, labeling, and self-control theory. More importantly, antisocial cognitions contribute to continued involvement in criminal behavior and are identified as a "Big Four" criminogenic need by the RNR model. These distorted thinking patterns are a primary target of the behavioral and cognitive-behavioral treatment approaches that are recommended under the general responsivity principle of the RNR framework. As a criminogenic need, criminal thinking/antisocial cognitions should be prioritized for treatment during correctional supervision. Over the last two decades, a number of instruments have been developed to measure criminal thinking. These instruments may include the Psychological Inventory of Criminal Thinking Styles (PICTS), the Criminal Sentiments Scale, the Measure of Offender Thinking Styles, the Criminogenic Thinking Profile, and the Criminal Cognitions Scale.

Antisocial Personality Pattern. The RNR framework also identifies an antisocial personality pattern as one of the "Big Four" criminogenic needs. This criminogenic need factor is generally characterized by proneness for adventurous pleasure-seeking behavior, low self-control, and/or aggressiveness. The DSM-IV criteria for an antisocial personality disorder (ASPD) is characterized by a callous disregard for the feelings of others; gross or persistent attitude of irresponsibility; disregard for social norms, rules, or obligations; incapacity to maintain enduring relationships; low tolerance; frustration and use of aggression or violence; incapacity to experience guilt or to profit from experience; or marked proneness to blame others for the behavior that the person exhibits. DSM-IV criteria for ASPD offer one possible operationalization of this antisocial personality criminogenic need factor. As with any personality trait, it is important to recognize that all human beings exhibit some aspect of this but the criminal offender tends to have sustained characteristics. Measures of ASPD are as close to a diagnostic criterion for criminal personality as exists. This conceptualization recognizes that such behaviors are common and that the issue is whether the individual consistently displays the traits.

Another possible indicator of an antisocial personality pattern that has been operationalized is the construct of low self-control. Impulsive and risk taking behavior is a dynamic characteristic often noted among offenders. The general premise is that low self-control does not define criminal behavior; instead it provides a context for criminal acts depending on opportunities and other motivating factors. Low self-control is exhibited by the offender being easily persuaded by situational and environmental factors, and without attachments, there is little to constrain the individual. The Risk-Need-Responsivity framework holds that an antisocial personality pattern is in fact amenable to change and can be addressed through cognitive restructuring interventions (e.g., cognitive-behavioral therapy (CBT)).

Antisocial Associates and Social Supports. Consistent with its grounding in social learning theory, the RNR framework stresses antisocial associates (family and peers) as a criminogenic need. At the same time, the model identifies pro-social family and peers as potential protective factors against continued involvement in offending. Pro-social family relations are important for reducing offending behaviors. Family ties can provide emotional support and facilitate the offender change process. But, histories of family involvement in crime and prior incarcerations are often precursors to learned behaviors regarding criminal behavior and drug use. The role of family in the RNR framework is complex. Having an antisocial family is criminogenic while having a pro-social family is identified as key facilitators of desistence from offending.

Relations with others, including antisocial peers and families, can affect three aspects of future offending involvement. As noted by Huebner and Berg (2011), three mechanisms of family and peer associations (referred to as social ties) may foster desistance are (1) ties with pro-social individuals control offender behavior, (2) ties can provide emotional support, and (3) ties can help offenders facilitate a transformation in identifying from offender to citizen, or reinforce an offender identity. While adult pro-social bonds to family and spouse can serve as a catalyst for desistance, a strong empirical link has been established between associations with antisocial peers and offending. Accordingly, the RNR framework stresses identifying family and friends as either supportive of or detrimental to a crime-free lifestyle and prioritizing antisocial associates as a criminogenic need that needs to be addressed in order to improve the likelihood of desistence from crime and/or drug use.

Employment and Educational Attainment. Offenders tend to have lower educational attainment than the general population including a higher percentage of offenders that do not graduate from high school or receive a General Educational Development (GED) degree. But, not completing high school is not criminogenic in that it does not "cause" a person to commit crimes. Not graduating high school however may affect verbal intelligence or literacy levels, which may affect the ability to obtain employment. Prior research has found that individuals who do not graduate high school are more likely to recidivate than those who do graduate high school. Much of the impact of education on recidivism outcomes may operate through its impact on employment which has generally received more empirical support as a correlate of recidivism outcomes. Because it is not directly related to recidivism, education is considered a non-criminogenic dynamic need that requires less treatment attention than criminogenic needs.

Offenders tend to have lower steady employment than non-justice-involved individuals. Again, employment-related issues are not criminogenic given that there is no causal linkage between employment and recidivism. But, employment has an indirect relationship with recidivism and therefore may be targeted through correctional programming when other more direct needs are not present. Employment after release from prison is shown to foster better outcomes, but the overall pattern is unclear regarding the relationship between employment and recidivism. Further research is needed to (1) clarify the mechanisms through which employment and education affect recidivism outcomes and (2) clarify the recidivism reduction potential of educational and employment-related treatment programs.

Mental Health Status. Mental health functioning is a complex need given the prevalence and diversity of mental health conditions in the justice system and their varying impact on offender functioning. Despite the high prevalence of mental health and co-occurring disorders (CODs) in the justice system, few empirical studies have found that the presence of a mental health condition is a direct predictor of criminal conduct. While it does not appear that mental health status is directly related to recidivism, mental health conditions may negatively impact the performance of offenders in programming and can increase risk for technical violations related to failure to complete mandated treatment. While there is no direct link between mental health issues and recidivism, failure to provide mental health services to those with mental illness can negatively impact the outcomes of offenders and lead to increased recidivism rates.

Housing. Housing is another dynamic offender need that warrants consideration in the RNR framework. Although housing status does not directly predict recidivism outcomes, instability in housing indirectly affects outcomes because it impacts the overall stability of the offender in the community. While the relationship between housing stability and recidivism remains empirically unclear, addressing housing issues can provide stability in an offender's life and potentially have an indirect impact on recidivism outcomes. Addressing housing needs may also improve offender performance on community supervision and within community-based treatment. Like other non-criminogenic needs, housing is treated as a tertiary dynamic risk factor within the RNR framework.

Risk of Recidivism and Location. A number of recent research studies have emphasized the importance of examining the influence of both individual and community risk factors. There is much more research available about individual risk factors than about community-level risk factors that can be directly linked to recidivism. This paucity of empirical research on the community context of recidivism is a major impediment to efforts to improve correctional performance. Offender's risk of recidivism is influenced by both individual- and community-level risk factors. The location where the offender resides makes a difference in the likelihood of recidivism.

The notion that offenders with similar individual risk profiles (based on such factors as prior offense history, prior incarceration, history of substance abuse, and employment/education deficits) are more likely to fail if they are released to a small number of identifiable high-risk communities is a factor that has emerged in the recidivism literature in recent years. To the extent that high-risk communities are also resource-poor communities, it seems logical to suggest that as a general principle, you cannot change offenders unless you also change the communities in which offenders reside. Accordingly, when implementing the Risk-Need-Responsivity framework in practice, it is essential to consider the community context in which it is being applied. The RNR framework is unlikely to be successful in improving offender outcomes if programming is not available to address dynamic offender needs. The availability of programming is a central component of the third and final RNR principle: responsivity.

Identifying Targets for Programming

Many individuals have more than one of the "Big Four" or "Central Eight" needs. By examining the number and type of criminogenic needs an offender presents, it is possible to identify offenders who are more ingrained in a criminal lifestyle and therefore should be targeted for more intensive services designed to affect criminal thinking and potentially reduce recidivism. Additionally, screening and assessment of needs provides guidance regarding which needs are present and which needs should be prioritized in interventions.

Within the Risk-Need-Responsivity framework, the goal is to identify dynamic factors – amendable to change – that contribute either directly or indirectly to criminal behavior. Identifying dynamic needs that affect both the opportunity to participate in and the desire to commit criminal acts is important. Similarly, it is important to give attention to factors that either can serve to protect the individual from criminal involvement (stabilizers) or those that may contribute to the opportunity or desire to commit crimes (destabilizers). Included in this framework are factors that tend to have an indirect relationship with recidivism including mental health status, educational attainment, employment history and options, housing, location of residence, and other factors frequently discussed in the realm of dynamic offender needs. The inclusion of these clinically relevant factors is grounded in the notion that these non-criminogenic factors impact the decisions and choices made by offenders and therefore need to be included in decisions regarding treatment matching and amenability to change. And, to some degree, these factors may also indicate severity of problem behavior warranting attention.

The following conceptualization of need severity can be used to augment the original Andrews and Bonta Risk-Need-Responsivity framework and foster more efficient correctional resource allocation:

1. Primary criminogenic needs should fall into three categories: substance dependence (on a criminogenic drug), criminal lifestyle (made up of criminal thinking and other Big Four criminogenic needs), and specific offender type (e.g., sex offender, domestic violence offender, or drunk driver). This priority is based on magnitude of relationship to recidivism and knowledge of effective interventions and the

availability of programming to target these specific criminogenic factors.

2. Offenders who have three or more criminogenic needs should be prioritized for treatment. Offenders with substance dependence on opioid, cocaine/crack, methamphetamines or chronic drunk drivers with alcohol disorders should prioritized for treatment. Research has shown that the number of needs affects recidivism (see Andrews and Bonta 2006), and therefore offenders who present with higher scores on criminal attitudes and values, antisocial peers, family criminal networks, and history of antisocial behaviors need to be prioritized for more intensive and structured programming.
3. Core demographics are important in guiding assignments to programming. Gender, age, and location of residence warrant consideration in the treatment matching process.
4. Substance abuse and mental illness, although not directly related to recidivism, should be prioritized among non-criminogenic needs as intervention targets because they are clinically relevant and treatment programs are available that target these specific offender needs.
5. Given their generally weaker correlations with recidivism, the remaining dynamic need factors should be considered as either stabilizers (protective factors) or destabilizers (negative indicators). This distinction provides for a greater opportunity to consider factors that should be used to adjust programming given the number and type of stabilizers and destabilizers in a person's life.

Responsivity: An Overview

The responsivity principle of the RNR framework has two components: general responsivity and specific responsivity. The general responsivity principle suggests that correctional interventions will have the greatest impact on recidivism when they are evidence-based and employ treatment techniques that have been empirically linked to improved offender outcomes. The principle identifies behavioral and cognitive-behavioral approaches as those that are most likely to result in reduced recidivism. Adherence to the general responsivity principle requires that all offender treatment programs be theory-based and employ either behavioral or cognitive-behavioral offender change approaches.

The RNR framework conceptualizes specific responsivity in terms of the personal characteristics of the offender that can affect their engagement in treatment or prevent them from being successful in programming. These personal characteristics generally included things like mental and emotional problems, cognitive functioning, and the individual's level of motivation and readiness to change. The specific responsivity principle has expanded and also includes personal characteristics such as age, gender, cultural background, personality, and past treatment experiences among other factors. Under the specific responsivity principle, these client characteristics are considered when identifying the type and level of correctional intervention that will be most "responsive" to the treatment needs of the individual offender and therefore most likely to result in improved outcomes.

The term responsivity has evolved into more of a general principle for matching the appropriate type of programming based on the risk and need factors of offenders – recognizing that these risk/need factors should drive the level and type of programming. This conceptualization mirrors some other efforts to develop treatment placement criteria (e.g., American Society of Addiction Medicine (ASAM)) but recognizes that the setting and the type of intervention are important in affecting outcomes. Embedded in this conceptualization of responsivity are a number of program- or intervention-level factors that affect the quality and potential impact of the programming that should also be considered when determining the most appropriate level of correctional intervention for a justice-involved individual. These additional program features include the dosage, the content, the fidelity or adherence to the program model, and the quality of the program resources. It is beyond this essay to review all of these components. But, consistent with the evidence-based practices literature, we are aware that:

- Dosage or the number of hours that a person is involved in the correctional program should vary by risk level and that higher-risk offenders should be involved in larger quantities of programming.
- Content or the theoretical orientation of the intervention should include cognitive-behavioral therapy, therapeutic communities, or integrated service models like drug treatment courts. The nature of the treatment programming has an impact on results.
- Fidelity, adherence to core correctional practices, or the degree to which the program is implemented as planned is directly related to program effectiveness. This construct includes the dosage, content, type of staff, curriculum, and other key functional components of the program and is essential for bringing about offender change.
- Program resources include facilities, resources, and staffing available to the program. These should be aligned with the intended goals of the intervention. Program resources are important to ensure the clinical or responsivity components are adequately resourced to be effective.

While the original Risk-Need-Responsivity framework stands, an expanded empirical literature base has assisted in reconfiguring some of the decision criteria for responsivity to include:

1. Static risk level should drive dosage and type of programming. Since static risk accounts for the majority of the variance in explaining recidivism, it appears that placing moderate to high-risk offenders, with various criminogenic needs, into more intensive programming should be the priority.
2. The concept of problem severity should be integrated into the RNR framework. While the original RNR model separated by priority (Big Four) and lesser priority (non-criminogenic) needs, it appears that this distinction does not separate out those that have more severe behavioral issues from those with less severe issues. The responsivity model should consider all dynamic offender needs, including the host of clinically relevant, non-criminogenic factors in identifying problem severity.
3. Substance use dependence and criminal lifestyle (a composite measure of the "big eight") should be prioritized for more structured, intensive programming. These severe behavioral issues generally require more intensive services, particularly for moderate to high-risk offenders.
4. Age and gender are important responsivity factors that should adjust the content of the program/services.
5. Programming content should be categorized into four levels: cognitive and behavioral, interpersonal and social skills, lifestyle skills, and punishment (no programming). Cognitive and behavioral programming should be the primary intervention type for individuals who are moderate to high risk with criminogenic needs.

This expanded responsivity conceptualization can be used to match justice-involved individuals to levels of care in adherence with the Risk-Need-Responsivity framework. The steps of responsivity are conceptually clear: (1) identify static risk and prioritize high and moderate risk offenders for more intensive treatments and interventions; (2) identify criminogenic needs and target specific criminogenic needs during treatment; (3) use cognitive-behavioral, behavioral, and other evidence-based interventions to address criminogenic needs; (4) take into consideration non-criminogenic needs, stabilizers, and destabilizers in determining the appropriate level of care for a justice-involved individual; and (5) focus on client strengths to refine and improve the treatment match.

Conclusion

The original RNR framework has had a positive impact on the growth of evidence-based programming and treatments in the field of corrections. The model has identified the importance of individual-level factors in making determinations about the appropriate level of programming that is needed. However, a review of the literature on recidivism found that the original positioning of various criminogenic needs in the RNR framework may need to be altered. This is especially true when examining how these individual factors affect the results from participation in different type of programming or the type of programming that is likely to reduce recidivism for specific profiles of offenders. The quagmire of recidivism reduction is that (1) there is great variability across studies regarding the size and (sometimes) the direction of the effect between an individual-level factor and recidivism; (2) oftentimes an individual-level factor is indirectly related to recidivism, usually when other individual-level variables are present; and (3) program-level effects may vary from merely examining the impact of various individual factors on offender outcomes. That being said, these inconsistencies in the literature do not challenge the RNR framework but merely suggest a slightly different ordering of importance of different factors, and the inclusion of clinically relevant factors such as mental health, housing stability, and substance abuse. And, it means that there should be certain drivers of who needs more intensive programming – namely, individuals with moderate to higher risk or individuals with more severe constellations of criminogenic needs.

Adhering to the principles of the Risk-Need-Responsivity framework at a system level can lead to reduced recidivism rates and improved cost-effectiveness of the US correctional system. The RNR model combines several evidence-based correctional practices into one, versatile framework. A particularly valuable component of the RNR framework is its potential to guide resource allocation and improve the fit between correctional interventions and individual offender needs. The framework stresses the importance of building a human services culture within the correctional system; a goal that empirical evidence has shown is more effective at reducing recidivism than incarceration or sanctioning alone. In a time when correctional resources are sparse, the Risk-Need-Responsivity framework offers a model to efficiently and cost-effectively respond to the needs of the offender population. The framework offers not only improved cost-effectiveness but also improved public safety returns.

Related Entries

- ▶ Health Issues in Prison Reentry Models
- ▶ Offender Change in Treatment
- ▶ Rehabilitation
- ▶ Risk Assessment, Classification, and Prediction
- ▶ Risk Factors for Prison Recidivism

Recommended Reading and References

Andrews DA, Bonta J (2006) The psychology of criminal conduct, 5th edn. Anderson, Cincinnati

Andrews DA, Bonta J (2010) The psychology of criminal conduct, 5th edn. Anderson, Cincinnati

Andrews DA, Bonta J, Wormith JS (2006) The recent past and near future of risk and/or need assessment. Crime Delinq 52(1):7–27

Austin J (2006) How much risk can we take? The misuse of risk assessment in corrections. Fed Probat 70(2):58–63

Austin J (2009) The limits of prison based treatment. Vict Offenders 4:311–320

Bonta J, Andrews DA (2007) Risk-need-responsivity model for offender assessment and treatment (user report 2007–06). Public Safety Canada, Ottawa

Bourgon G, Armstrong B (2005) Transferring the principles of effective treatment into a "real world" prison setting. Crim Justice Behav 32(1):3–25

Byrne J (2009) Maximum impact: targeting supervision on higher-risk people, places and times. Public safety performance project. The PEW Center on the States

Friedmann PD, Taxman FS, Henderson CE (2007) Evidence-based treatment practices for drug-involved adults in the criminal justice system. J Subst Abus Treat 32(3):267–277

Huebner BM, Berg M (2011) Examining the sources of variation in risk for recidivism. Justice Q 28(1):146–173

Latessa EJ, Lovins B (2010) The role of offender risk assessment: a policy maker guide. Vict Offender 5(3):203

Lipsey MW, Cullen FT (2007) The effectiveness of correctional rehabilitation: a review of systematic reviews. Annu Rev Law Soc Sci 3:297–320

Taxman FS, Belenko S (2012) Implementing evidence-based practices in community corrections and addiction treatment. Springer, New York

Taxman FS, Cropsey KL, Young D, Wexler H (2007a) Screening, assessment, and referral practices in adult correctional settings: a national perspective. Crim Justice Behav 34(9):1216–1234

Taxman FS, Perdoni M, Harrison LD (2007b) Drug treatment services for adult offenders: the state of the state. J Subst Abus Treat 32:239–254

Adaptation

▶ Innovation and Crime Prevention

Administrative Criminology

▶ So What Criminology?

Admissibility

▶ Identification Technologies in Policing and Proof

Adolescence-Limited Offending

▶ Onset of Offending

Adolescent Contexts

▶ Network Analysis in Criminology

Affordable Care Act

▶ Health Care Reform and Criminal Offenders (US)

Age-Crime Curve

Rolf Loeber[1] and David P. Farrington[2]
[1]Department of Psychiatry, University
of Pittsburgh, Pittsburgh, PA, USA
[2]Institute of Criminology, University
of Cambridge, Cambridge, UK

Overview

One of the most consistent findings across studies on offending in different countries is the age-crime curve (Farrington 1986; Tremblay and Nagin 2005). The relationship between age and crime is of an asymmetrical bell shape, showing that the *prevalence* of offending (the percentage of offenders in a population) tends to increase from late childhood, peaks in the teenage years (around ages 15–19), and then declines from the early 20s, often with a long tail (Fig. 1). The sharp increase during adolescence in the curve reflects an increase in new delinquency recruits during that period, and the rate of recruitment tends to slow down subsequently (Smith et al. 2002). The age-crime differs from developmental trajectories (qv) of offending in that the age-crime curve indicates the prevalence of offending by age of populations of individuals, whereas developmental trajectories make distinctions between subgroups of offenders, such as life-course persistent offenders and adolescent-limited offenders.

Key Issues

Although universal, the age-crime curve is not invariant but differs depending on characteristics of offenses and offenders (Farrington 1986). For example, the age-crime curve for violence tends to peak later than that for property crime (Piquero et al. in press). Studies also show that the age-crime curve for girls peaks earlier than for boys (Farrington 1986; Elliott et al. 2004). The curve is also higher and wider for young males (especially those of a minority status) growing up in the most disadvantaged compared to advantaged neighborhoods (Fabio et al. in press). There are similarities between the age-crime curve and the age-co-offending curve (Piquero et al. in press) and between the age-crime curve and the victim age curve (Stolzenberg and D’Alessio 2008). Since most of the violence is directed at same-age victims, it is not surprising that the age period 16–24 is also a high-risk period for violent victimization (e.g., Kershaw et al. 2008).

In addition, the criminal propensity of successive age cohorts is not necessarily the same, with more individuals in some age cohorts being

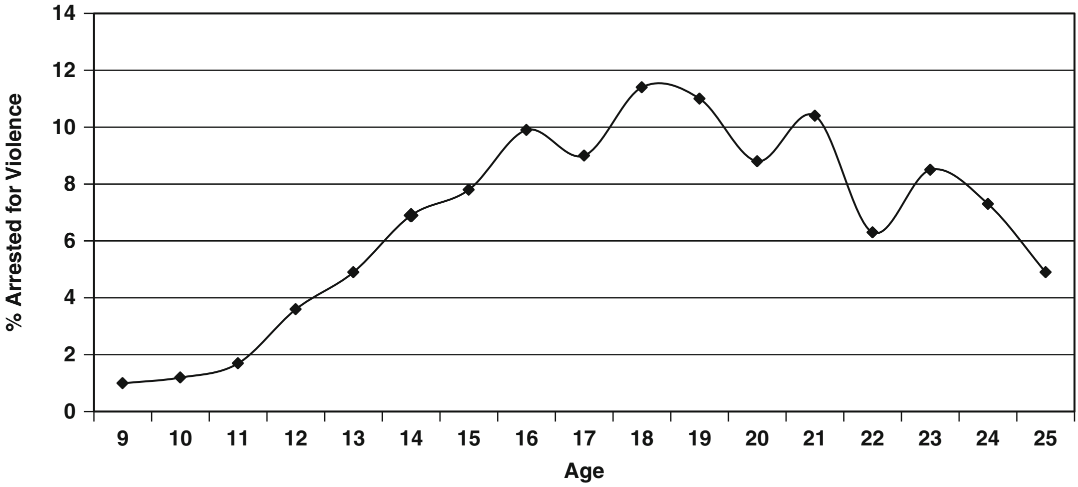

Age-Crime Curve, Fig. 1 Age-crime curve, based on longitudinal data from the Pittsburgh Youth Study (oldest cohort) and using self-reported delinquency and official records of offending

delinquent and over a larger number of years (i.e., with a higher duration) than individuals in other age cohorts (e.g., Farrington 1986; Cook and Laub 2002). The majority of the published age-crime curves are based on cross-sectional data, which usually means that they are based on multiple age cohorts and for that reason are usually different from curves based on longitudinal data. For these data, it is often hard to disentangle cohort, age, and period effects (i.e., changes in historic time). Therefore, preferred are age-crime curves based on self-reports and official records that are usually based on a single cohort (e.g., Loeber et al. 2008) as preferred, because they avoid the confounding influence of multiple cohorts.

Most of the studies on the age-crime curve have been based on official records, which often underestimate the number of delinquent individuals and, hence, the height and probably also the shape of the age-crime curve. Therefore, it is important to examine the age-crime curve based on more comprehensive information, such as that based on self-reports of offending and complemented by reports by parents and teachers (Loeber et al. 2008). There are several other caveats about the age-crime curve that may influence conclusions drawn from it. For example, self-reported delinquency shows an earlier peak than official records (Piquero et al. in press). This may reflect the fact that juvenile offending at a young age (as evident from self-reports) goes unrecorded by the police or there may be a delay in arrest, and as a consequence these offenses remain undetected in official records until a later age.

Researchers have not been in agreement about whether the prevalence of offending is correlated over age with the frequency in offending (e.g., Blumstein et al. 1986) and whether as the prevalence increases so does the frequency (i.e., in the upswing of the age-crime curve). Conversely, when the prevalence of offending decreases, does the frequency of offending also decrease (in the downslope of the age-crime curve)? Yet, research on the frequency of offending at different ages has been scarce (e.g., Barnett et al. 1987; Loeber and Snyder 1990; Piquero et al. in press).

It is not well known that, although an early age of onset, compared to a later age of onset, is associated with a longer criminal career, the highest concentration of desistance takes place during late adolescence and early adulthood *irrespective* of age of onset. This corresponds with the downslope of the age-crime curve. In fact the decrease in prevalence in the downslope of the age-crime curve is very substantial.

In some data sets, it goes down from about 50 % to about 10 % of all persons (e.g., Loeber et al. 2008). Important for our understanding of the transition between adolescence and adulthood is the right-hand tail of the age-crime curve. The higher and longer that tail, the more this indicates that there is a population of youth who may not have outgrown delinquency or who may have started offending during adulthood (see below). In summary, the downslope of the age-crime curve varies between different populations of young people, but often extends from adolescence into adulthood. It is important to realize that the majority of serious forms of crime – including violence – take place in the period of downslope of the age-crime curve.

Individual Differences, Offending Careers, and the Age-Crime Curve

Some criminologists have assumed that individual differences in the propensity of offending are stable over time, including the age-crime curve (e.g., Gottfredson and Hirschi 1990). However, a recent reformulation that corresponds more with the age-crime curve is that the magnitude of the individual differences in offending is modest in the early part of the age-crime curve, increases toward the peak of the curve, and then decreases in the downslope of the curve (Loeber et al. 2012). It still can be argued that what remains stable is the rank-ordering of individuals' propensity of committing delinquent acts.

Factors Influencing the Height and Width of the Age-Crime Curve

Loeber and Farrington (2012) postulated ten processes that might explain the downslope of the age-crime:

1. Individual differences in self-control
2. Brain maturation
3. Cognitive changes (e.g., decision making to change behavior)
4. Behavioral risk factors (disruptive behavior and delinquency) and behavioral protective factors (nervousness and social isolation)
5. Social risk and protective factors (family, peers, school)
6. Mental illnesses and substance use/abuse
7. Life circumstances (e.g., getting married, becoming employed)
8. The situational context of specific criminal events, including crime places and routine activities
9. Neighborhood (e.g., living in a disadvantaged neighborhood and the concentration of impulsive and delinquent individuals in disadvantaged neighborhoods)
10. The justice response (e.g., transfer to adult court, longer sentences)

Of the above processes, several are thought to also explain the upslope of the age-crime curve (Nos. 1, 4, 5, and 9). For example, parents' influence over their children's behavior tends to decrease, and the influence of peers tends to increase during the upslope of the age-crime curve.

The critical question is: *Does each process explain the onset, persistence, and desistance in offending from childhood into early adulthood?* The explanatory processes reviewed by Loeber and Farrington (2012) tend to take place at different age periods from childhood to early adulthood and, consequently, may influence different but interrelated outcomes. For example, early individual differences become manifest after birth and evolve subsequently; exposure to new risk factors increases from childhood through adolescence, while changes in desirable life circumstances – such as marriage and employment – typically accelerate from late adolescence into early adulthood. Thus, the explanatory processes tend to operate at different ages, some early and some later in offending careers.

We will highlight only a few of the ten processes and concentrate on individual differences and brain maturation. Many scholars focused on different underlying constructs thought to be central to individual differences. However, there is little agreement on whether poor self-control, reckless behavior, impulsivity, poor executive functioning, or sensation seeking is the important construct (see also Jolliffe and Farrington 2009), and there is even less agreement on how best to quantify them independent of disruptive child behavior and delinquency. Also, the assumed stability of the underlying constructs across the life-course is often untested and is rarely based on a follow-up of the same individuals over time (but see Piquero et al. 2009). Another possible flaw is the fact that individual differences, in for example, self-control, do not easily map on the age-crime curve, especially on the upslope of the curve representing the increase of offending from childhood (ages 7–12) to middle adolescence (ages 13–16) *and* the downslope of the age-crime curve representing the decrease of criminal activities from late adolescence (ages 17–19) into early adulthood (ages 20–25). Further, the individual differences approach usually does not focus on brain maturation from childhood to early adulthood (but see Moffitt 1993).

Loeber et al. (2008), who examined a large array of protective factors, concluded that in the Pittsburgh Youth Study, desistance in the downslope of the age-crime curve was more difficult to predict than desistance at earlier ages. Specifically, *none* of the protective factors measured either in early adolescence (ages 13–16) or late adolescence (ages 17–19) predicted desistance in early adulthood (ages 20–25), but not all research is in agreement on this point (Stouthamer-Loeber et al. 2008).

Some researchers have proposed that changes in life circumstances, such as marriage and employment, also impact the timing of the decrease in the age-crime curve (e.g., Theobald and Farrington 2009). It should be noted that these changes in life circumstances often occur late into the downslope of the age-crime curve, where they could still cause a decrease in offending.

A critical issue is whether mental maturation and the associated emergence of internal controls are related to brain maturation and whether both influence the downslope of the age-crime curve. In brief, an improved understanding of individual differences would suggest something like the following. Young people differ in the speed of brain maturation, and this brain maturation, accompanied by social learning, causes improved internal controls. These improved controls become evident through a decrease in risk-taking behaviors, reckless behavior, and sensation seeking and improvements in problem solving, future orientation, and decision making. Improved controls are thought to reduce deviant behavior, including the commission of delinquent acts, and increase prosocial behaviors.

There are several pieces of research that fit these propositions. There is increasing evidence that the brain continues to develop during childhood into early adolescence, including ongoing myelination (e.g., Giedd et al. 1996; Sowell et al. 2001), and into adulthood when white matter increases and synapses are pruned. Research also shows that “the dorsal lateral prefrontal cortex, important for controlling impulses, is among the latest brain regions to mature without reaching adult dimensions until the early 20s” (Giedd 2004, p. 77). The importance of white matter is underscored by the finding that decreased white matter is significantly more common among boys with psychopathic tendencies compared to controls (De Brito et al. 2009). Biological changes in the prefrontal cortex during adolescence and the early twenties lead to improvements in executive functioning, including reasoning, abstract thinking, planning, anticipating consequences, and impulse control (Sowell et al. 2001).

The data on brain development, although mostly cross-sectional, is much in line with behavioral evidence that reckless behavior tends to decrease as late as early adulthood, as is evident from data on the incidence of car accidents, even when controlling for the number of miles traveled (Foss 2002). This is widely recognized by insurance companies whose premiums for car insurance for young drivers up to about age 25 are dramatically higher than for older drivers. The idea of improved behavioral controls that emerge between late adolescence and early adulthood is also evident from psychological research. For example, Steinberg et al. (2009) investigated time perspective, planning ahead, and anticipation of consequences among individuals between ages of 10 and 30. Although this was a cross-sectional study, the results suggest that the ability to plan ahead improves dramatically between early adolescence and the early 20s along with the anticipation of consequences, while time perspective improves slightly less in that same period. In addition, a recent meta-analysis concluded that children's and adolescents' coping repertoires increase with age, including planful problem solving, more diverse distraction techniques, positive self-talk, and intentional self-regulation of emotions (Zimmer-Gembeck and Skinner 2011).

A key piece of evidence, however, derives from longitudinal research in which the same individuals are followed up over time. In general longitudinal research shows that childhood impulsivity measured by ratings made by adults (as early as age 5) is significantly predictive of violence (Jolliffe and Farrington 2009). Corresponding to the thesis that young individuals differ not just in their stable traits over time but also in their developmental maturation, research based on the Pittsburgh Youth Study (Loeber et al. 2011) shows that individual differences in cognitive impulsivity, as measured by young males' performance on several psychometric tasks at age 13, map onto individual differences in the age-crime curve. Individuals scoring lowest on cognitive impulsivity at age 13 on average had the highest prevalence of arrest for delinquent acts throughout adolescence into early adulthood (ages 12–26), but from age 27 to 28 onward, their prevalence of arrest became very similar to those with lower cognitive impulsivity. Second, *all* groups that differed in their level of cognitive impulsivity (from highest to lowest) showed the age-crime curve, meaning that *each* group showed a decline in the 20s in the prevalence of convictions for delinquent acts. This is contrary to the notion that highly impulsive individuals or individuals with little self-control on average remain highly delinquent over time (e.g., Gottfredson and Hirschi 1990). Instead, the results show that on average, even highly impulsive individuals decrease their level of delinquency with age, although they took longer to outgrow delinquency than less impulsive individuals.

Yet another piece of evidence contradicts the assumed stability of underlying traits. Piquero et al. (2009) reviewed the evidence that several intervention programs improved self-control and reduced delinquency in young people. The programs reviewed targeted those up to age 10 and showed that juveniles' self-control can be changed as a result of systematic interventions. Less well documented is how well interventions promoting self-control influence older populations.

In summary, the central issue is not whether there are stable individual differences between individuals in their cognitive control and their delinquent acts. Instead, the key question is: To what extent *individual differences increase over time and then decrease?* Thus, the age-crime curve with its increase, peak, and then decrease in prevalence *and*, possibly, severity over time should be related to within-individual differences in brain and cognitive development. However, brain and cognitive development has not been related to adult-onset offending.

Lowering the Age-Crime Curve

One of the limitations of current evaluation studies is the absence of yearly long follow-ups of treated and nontreated individuals during adolescence and early adulthood. For that reason, published evaluation studies have not been able to show the degree to which interventions could have lowered the age-crime curve. To resolve this, longitudinal data from the Pittsburgh Youth Study was used to simulate the impact of an intervention on offending by at-risk youths (Loeber and Stallings 2011). Simulation can be compared with asking the question "What if…" we changed one aspect of the data set, what impact would that have on future offending? To address this, the authors modeled first the efficacy to screen for high-risk individuals and then to apply a treatment that "knocked out" youth who otherwise would have been at risk of committing serious delinquency. To identify high-risk boys, a screening risk score was computed from data collected at the first assessment for each of the three cohorts at ages 7, 10, and 13, for the youngest, middle, and oldest cohorts, respectively (Loeber et al. 1998). This screening score included conduct problems and delinquency as reported by the boys and their parents and teachers. The intervention modeled was inspired by the best intervention programs available (Lipsey and Wilson 1998) evaluated by the comparison of an experimental sample receiving the intervention with a control sample not receiving the program (or receiving an alternative program). Loeber and Stallings (2011) took as a conservative criterion a success rate for the intervention of 30 %.

The intervention was associated with a substantial reduction in serious delinquents by one fifth to one quarter (21.5 % and 27.6 % for the youngest and oldest cohort, respectively), especially from ages 11 to 21. Other results indicate that the intervention also was associated with a decrease of 20 % in the prevalence of arrest by the police, a 35 % lower prevalence of homicide offenders and homicide victims, and a 29 % reduction in the weeks of incarceration. Thus, the modeled intervention had a substantial benefit of lowering the age-crime curve by reducing the prevalence of self-reported serious offenders, officially recorded homicide offenders and homicide victims during adolescence and early adulthood, and had benefits for the justice system by greatly reducing arrests and convictions.

In conclusion, there are many intervention programs available outside of the justice system that reduce recidivism and prevent persistence of offending from adolescence into early adulthood. Preliminary results indicate that such interventions may lower the age-crime curve for a population of youth.

Policy Implications

All available age-crime curves show that the legal age of adulthood at 18 (or for that matter ages 16 or 17 in some states) is not characterized by a sharp change (or decrease) in offending at exactly that age and has no significant relevance to the downslope of the age-crime curve. Serious offenses (such as rape, robbery, homicide, and fraud) tend to emerge after the emergence of less serious offenses of the late adolescence-early adulthood window. Even for serious offenses, however, there is no clear dividing line at age 18. Steinberg et al. (2009) concluded that "The notion that a single line can be drawn between adolescence and adulthood for different purposes under the law is at odds with developmental science" (p. 583). Piquero et al. (in press) emphasized that knowledge about the age-crime curve during the transition from adolescence to adulthood potentially offers valuable insights for incapacitation and reentry policies. Finally, lengthier sanctions administered in the adult court aimed at young offenders, based on our understanding of normative outgrowing of delinquency in young populations, will have no bearing on lowering the age-crime curve. In summary, if lowering of the age of adulthood is advocated, legislators need to take into consideration the age-crime curve, costs and benefits, and the need to screen vulnerable populations who take longer than others in their brain maturation and their outgrowing of the age-crime curve. The litmus test is whether concerted preventive and remedial interventions inside and outside of the justice system can lower and shrink the age-crime curve for future generations of young people.

Related Entries

- ▶ Career Criminals and Criminological Theory
- ▶ Cognitive/Information Processing Theories of Aggression and Crime
- ▶ Desistance from Crime
- ▶ Establishing Causes of Offending in Longitudinal and Experimental Studies
- ▶ Labeling Theory
- ▶ Longitudinal Studies in Criminology
- ▶ Moffitt's Developmental Taxonomy of Antisocial Behavior
- ▶ Offender Decision Making and Behavioral Economics
- ▶ Onset of Offending
- ▶ Social Learning Theory

Recommended Reading and References

Barnett A, Blumstein A, Farrington DP (1987) Probabilistic models of youthful criminal careers. Criminology 25:83–108

Blumstein A, Cohen J, Roth JA, Visher CA (1986) Criminal careers and "career criminals.". National Academy Press, Washington, DC

Cook PJ, Laub JH (2002) After the epidemic: recent trends in youth violence in the United States. In: Tonry M (ed) Crime and justice: a review of research. University of Chicago Press, Chicago, pp 117–153

De Brito SA, Mechelli A, Wilke M, Laurens KR, Jones AP, Barker GJ, Hodgins S, Viding E (2009) Size matters: increased grey matter in boys with conduct problems and callous-unemotional traits. Brain: J Neurol 132:843–852

Fabio A, Cohen J, Loeber R (in press) Neighborhood socioeconomic disadvantage and the shape of the age-crime curve. American Journal of Public Health

Farrington DP (1986) Age and crime. In: Tonry M, Morris N (eds) Crime and justice: an annual review of research, vol 7. Chicago University Press, Chicago, pp 189–250

Foss R (2001) Licensing. Translating injury prevention research into action: A strategic workshop, Dallas, TX (February 1)

Giedd JN (2004) Structural magnetic resonance imaging of the adolescent brain. Ann N Y Acad Sci 1021:77–85

Giedd JN, Rumsey JM, Castellanos FX, Rajapakse JC, Kaysen D, Vaituzis AC, Vauss YC, Hamburger SD, Rapoport JL (1996) A quantitative MRI study of the corpus callosum in children and adolescents. Dev Brain Res 91:274–280

Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Jolliffe D, Farrington DP (2009) A systematic review of the relationship between childhood impulsiveness and later violence. In: McMurran M, Howard R (eds) Personality, personality disorder, and violence. Wiley, New York, pp 41–62

Kershaw C, Nicholas S, Walker A (2008) Crime in England and Wales 2007/08. London: Home Office. www.homeoffice.gov.uk/rds/pdfs08/hosb0708.pdf

Lipsey MW, Wilson DB (1998) Effective intervention for serious juvenile offenders: a synthesis of research. In: Loeber R, Farrington DP (eds) Serious and violent juvenile offenders: risk factors and successful interventions. Sage, Thousand Oaks, pp 313–345

Loeber R, Farrington DP (2012) From juvenile delinquency to adult crime: criminal careers, justice policy and prevention. Oxford University Press, New York

Loeber R, Snyder HN (1990) Rate of offending in juvenile careers: findings of constancy and change in lambda. Criminology 28:97–109

Loeber R, Stallings R (2011) Modeling the impact of interventions on local indicators of offending, victimization, and incarceration. In: Loeber R, Farrington DP (eds) Young homicide offenders and victims: risk factors, prediction, and prevention from childhood. Springer, New York, pp 137–152

Loeber R, Farrington DP, Stouthamer-Loeber M, Van Kammen WB (1998) Antisocial behavior and mental health problems: explanatory factors in childhood and adolescence. Lawrence Erlbaum, Mahwah

Loeber R, Farrington DP, Stouthamer-Loeber M, White HR (2008) Violence and serious theft: development and prediction from childhood to adulthood. Lawrence Erlbaum, Mahwah

Loeber R, Farrington DP, Stallings R (2011) The Pittsburgh youth study. In: Loeber R, Farrington DP (eds) Young homicide offenders and victims: risk factors, prediction, and prevention from childhood. Springer, New York, pp 19–36

Loeber R, Hoeve M, Slot NW, van der Laan P (in press) Persisters and desisters in crime from adolescence into adulthood: explanation, prevention, and punishment. Aldershot, UK: Ashgate

Moffitt TE (1993) Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. Psychol Rev 100:674–701

Piquero AR, Jennings WG, Farrington DP (2009) Effectiveness of programmes designed to improve self-control. Stockholm: National Council for Crime Prevention. Available from www.bra.se (choose English language pages)

Piquero AR, Hawkins JD, Kazemian L (in press) Criminal career patterns. In: Loeber R, Farrington DP (eds) From juvenile delinquency to adult crime: criminal careers, justice policy and prevention. New York, NY: Oxford University Press

Smith P, Goggin C, Gendreau P (2002) The effects of prison sentences and intermediate sanctions on recidivism: general effects and individual differences. Solicitor General of Canada, Ottawa

Sowell ER, Thompson PM, Tessner KD, Toga AW (2001) Mapping continued brain growth and gray matter density reduction in dorsal frontal cortex: inverse relationships during postadolescent brain maturation. J Neurosci 21:8819–8829

Steinberg L, Cauffman E, Woolard J, Graham S, Banich M (2009) Are adolescents less mature than adults? Minors' access to abortion, the juvenile death penalty, and the alleged APA "flip-flop". Am Psychol 64:583–594

Stolzenberg L, D'Alessio SJ (2008) Co-offending and the age-crime curve. J Res Crime Delinq 45:65–86

Stouthamer-Loeber M, Loeber R, Stallings R, Lacourse E (2008) Desistance from and persistence in offending. In: Loeber R, Farrington DP, Stouthamer-Loeber M, White HR (eds) Violence and serious theft: risk and promotive factors from childhood to early adulthood. Erlbaum, Mahwah, pp 269–308

Theobald D, Farrington DP (2009) Effects of getting married on offending: results from a prospective longitudinal survey of males. Eur J Criminol 6:496–516

Tremblay RE, Nagin DS (2005) The developmental origins of physical aggression in humans. In: Tremblay RE, Hartup WH, Archer J (eds) Developmental origins of aggression. Guilford Press, New York, pp 83–106

Zimmer-Gembeck MJ, Skinner EA (2011) Review: the development of coping across childhood and adolescence: an integrative review and critique of research. Int J Behav Dev 35:1–17

Agent-Based Assessments of Criminological Theory

Daniel Birks[1] and Henk Elffers[2]
[1]ARC Center of Excellence in Policing and Security, CEPS, Griffith University, Brisbane, Australia
[2]Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, The Netherlands

Synonyms

ABM; Agent-based modelling; Computational criminology; Individual-based modelling; Simulation

Overview

This entry discusses the application of computational agent-based models (ABM) in exploring the ramifications of crime event mechanisms proposed by criminological theory. In particular, it describes how simulation experiments can be used to systematically assess the plausibility of crime event theories in explaining commonly observed patterns of crime and how the use and development of ABM may offer a viable method of theoretical prototyping free from traditional empirical constraints which in turn encourages theorists to explicitly specify the mechanisms by which they propose observed crime phenomena come about.

Introduction

A wide range of criminological theories provide individual-level depictions of the crime event. Such theories outline hypothesized mechanisms of cognition and action for both potential victims and offenders and, in turn, the influences that the local environment place upon them in situating victimization. Criminologists hypothesize that these mechanisms are operating, interacting, combining, exciting, inhibiting, and subsuming one another in complex and dynamic ways which influence the spatiotemporal distribution of criminal opportunities and in turn crime. Unfortunately, key problems associated with observation and experimentation endemic throughout the social sciences dictate that we often struggle to directly observe how, and in what ways, such interactions take place and instead are left to observe only their output – the patterning of crime. As a result, a divide exists between observed aggregate crime patterns and proposed individual-level crime theory. This divide dictates that many theoretical propositions relating to the proximal causes of crime can be difficult to empirically verify to the degree that would be desirable.

The past decade has seen a number of criminology scholars begin to explore potential applications of computational models in a hope of better characterizing and understanding these complex interactions of the crime event (Birks et al. 2008, 2012; Bosse et al. 2010; Brantingham and Brantingham 2004; Brantingham et al. 2008; Brantingham and Tita 2008; Eck and Liu 2008; Groff 2007; Johnson 2008; Liu et al. 2005; Malleson et al. 2009; Van Baal 2004; Wang et al. 2008). Such approaches aim to provide support to existing theoretical and empirical efforts by allowing researchers to generate artificial worlds free from logistical, monetary, and ethical constraints in which the likely downstream impacts of theoretical propositions can be systematically estimated through simulation experimentation.

Drawing heavily on Epstein’s (1999) generative approach, this entry provides an overview of one particular type of computational model: the agent-based model, advocating its use in exploring the generative sufficiency of individual-level criminological theory, that is, the capacity of proposed individual-level mechanisms in explaining commonly observed aggregate patterns of victimization. The entry begins with an introduction to the agent-based modelling methodology, highlighting a number of its characteristics that are of considerable use in exploring complex social systems. Subsequently, Epstein’s (1999) notion of the generative explanation as afforded by ABM is set out, and a discussion relating to its application in the study of criminological theory provided. In concluding several strengths, weaknesses, and requirements for a generative approach to the study of criminological theory are discussed.

Computational Agent-Based Modelling in the Social Sciences

By definition, the social sciences deal with the functions and interactions of society; thus, complexity is inherent in almost all phenomena they aim to study. In relation to this issue, Herbert Simon called for the so-called "soft" social sciences to be relabelled the hard sciences due to a range of issues that encumbered investigation. Simon asserted that the laboratory conditions used by those who deal with the natural sciences permit a much clearer observation of cause and effect than is ever possible within the social sciences. Similarly, Epstein and Axtell (1996, p. 1) discuss a number of methodological difficulties that have to date limited the productivity of traditional equation-based models commonly applied within the social sciences for the purposes of theory testing. A summary of these observations follows:

- Social systems are rarely characterized by discrete, easily decomposable subprocesses; instead most social phenomena encompass numerous mechanisms which all act in unison – spatial, cultural, economic, demographic, and so on.
- Controlled experiments that aim to test hypotheses within the social sciences are often very difficult to perform due to a number of ethical and logistical constraints.
- Traditional social science models often assume that entities are perfect rational actors who have access to perfect information.
- In order to manage computational requirements, typical social science models suppress unit heterogeneity through the use of "representative agent" methods.
- Such heterogeneity is inherent in social systems; however, within the social sciences "there has been no natural methodology for systematically studying highly heterogeneous populations" (p. 2).
- Models within the social sciences often assume social systems can be characterized as static equilibria and as such ignore the importance of temporal dynamics.

In the attempt to overcome some of these problems, research within the social sciences over the last two decades has begun to embrace the application of computational modelling techniques drawn from a number of disciplines in exploring complex social systems. Agent-based models (ABMs) are one particular technique that offer considerable promise to those who aim to build explanatory models of complex social systems. ABM is a well-established method of computer simulation with many real and promising applications across a range of disciplines. Although technically straightforward, its concept is considered profound (Bonabeau 2002).

Rudimentarily, agent-based computational models simulate the interactions that occur between multiple autonomous entities with the aim of analyzing how the decentralized behavior of individuals impacts on the behavior of the system as a whole. ABMs have been applied in a vast range of applications that include investigating the dynamics of pedestrian flow in emergency situations, exploring youth subcultures, the study of financial markets, and understanding consumer purchasing behavior.

ABM allows social science researchers to create virtual societies and inhabit them with simulated populations of heterogeneous autonomous actors. Using these models, the societal level impacts of differing individual-level behaviors can be examined (Epstein and Axtell 1996). In this way, ABM provides a platform to explore how the decisions people make on a day-to-day basis translate into observable phenomena. Advocates of the approach suggest that it is this ability to capture the links between micro-action and macro-outcome that place ABM as a "natural" methodology for the study of human systems (Axelrod 1997; Bonabeau 2002; Epstein 1999; Hedström 2005). The observed behavior of complex societal systems often arises from the interactions of relatively simple microlevel behaviors (Schelling 1978), and it is in exploring this micro-macro divide among complex social systems that ABM offers considerable promise. By manipulating the initial conditions of the ABM and scrutinizing data collected about virtual populations, researchers gain insight into the likely dynamics of certain societal configurations. Fundamentally, ABMs are made up of two key components: a population of agents and a simulated environment in which they are situated.

Simulation Agents

In an ABM each member of the virtual population is represented by an autonomous decision-making entity, commonly referred to as an agent. Just like the members of a real population, agents can exhibit individual characteristics, preferences, and behaviors; for example, each agent might have an age, home location and preferred social group, etc. Within an ABM agents execute a variety of behaviors that govern how they perceive, reason, and act in particular circumstances. These behaviors define how agents interact with one another, how they observe and analyze their surroundings, and how they might alter that environment or their internal state by performing certain actions. Such agent behavior is commonly defined by a series of condition-action rules (this is a gross, but necessary, oversimplification – for a review of agent behavior architectures, readers are directed to Gilbert and Troitzsch (2005)). In the case of explanatory ABM this decision calculus is often inspired by the formalization of theory, such that algorithms and heuristics are developed to reflect the mechanisms theory suggests are operating at the individual level. By exploring the macro-level ramifications of particular individual-level theoretical constructs and comparing them to known aggregate characteristics of the real-world system, social scientists can use ABM to interrogate the validity of social science theories.

Simulation Environment

Agents within ABM are situated within an environment. This environment may take on a wide variety of forms depending upon the purpose of the model being built. For instance, a model environment may represent some form of abstract physical or social space, where proximity relates to the spatial convergence of entities or, alternatively, their ideals, opinions, and connectivity. Conversely, model environments may also be developed to closely mirror real environments, be they individual floor plans, neighborhoods, or cities. The significant notion here is that agents are situated in an explicit space, abstract, or otherwise, thus allowing the concept of localized interaction to be appropriately modelled. The level of realism suitable for particular simulation environments, and moreover ABM in general, is subject to much debate (Elffers and Van Baal 2008).

The statement that simulation agents should be situated in an environment that is as close to reality as can be managed may sound like a truism. For example, if one aims to study patterns of burglary by examining simulated burglar agents who roam some environment, applying some hypothesized choice mechanism, it may sound desirable to situate these agents within an environment that closely mirrors a map of say Amsterdam, representing all Amsterdam homes as targets and in turn their associated characteristics – i.e., value, visibility, and population density. However, on second thought, such environmental realism may not be such a good idea after all. If, for instance, this simulation does show the emergence of burglary patterns congruent with empirical study, how does one recognize whether such results emerge as a result of the specified choice mechanism of interest or, alternatively, the peculiarities of the Amsterdam housing constellation? If, as is usually the case in simulation research, and as argued here, one tries to establish whether the specification of the agents' choice mechanism is sufficient to generate the pattern of interest, this environmental setup only serves to obfuscate this assessment. As such, the application of simulation seeks to establish the sufficiency of the mechanism irrespective of the specific environment in which it operates. Thus, regularities should occur not only in Amsterdam but also in Brisbane, Cancun, Durban, etc. One, rather cumbersome, solution to this problem may be to devise simulations in all of the above realistic environments. An alternative is to attempt to reduce the environment to its bare essence and have the simulation work in an abstract environment, specifying the minimally necessary characteristics of the geography in which the agents operate (Elffers and Van Baal 2008).

In considering model complexity in general, Gilbert (2004, p. 9) succinctly states that " the art of modelling is to simplify as much as possible, but not to oversimplify to the point where the interesting characteristics of the phenomenon are lost." It may of course be the case that the actual spatial configuration is indeed part of the essence of a simulation study. In studying whether the Amsterdam canal configuration hampers burglars in their movements, one may well classify the Amsterdam geography to be an essential part of the agents' decision rules. In such cases, a realistic background (with respect to those characteristics, that is) is indeed necessary, as leaving them out would be a case of Gilbertian oversimplification in which essential characteristics are lost.

Agent-Based Interactions

Given these two key components, a number of interactions may be modelled through ABM. *Agent-agent* interactions are those where agents receive information or resources from one another, develop social links with other agents, or compete for some entity within the simulation, be it territory, resources, or status. *Agent-environment* interactions are those where agents draw information or resources from the world in which they are situated. *Environment-agent* interactions characterize those situations where the environment influences an agent, perhaps by constraining movement or defining the locations in which certain activities can and cannot take place. *Environment-environment* interactions may also occur, for instance, a resource within some environment may spread over time from its initial location to those adjacent. In reality, the types and ways in which elements of ABMs may interact are almost limitless, and it is this level expressiveness that allows ABMs to be used to represent an unprecedented number of systems, processes, and mechanisms.

The Temporal Dynamics of ABM

Another important feature of the ABM is its inherent ability to capture the temporal dynamics of a system. ABMs simulate the progression of time via discrete increments, often referred to as cycles. During each cycle agents within the environment perceive, reason, and act based upon their specified behaviors that in turn draw on an agent's local circumstances and individual characteristics. Many thousands of these cycles may occur as a simulation progresses. Thus, ABM is performed in an iterative fashion, permitting the longitudinal examination of time-dependent phenomena and the formation, interaction, and separation of system elements over time. Such temporal dynamics are especially important for the modelling of phenomena such as tipping points, where the accumulation of individual action over time can lead to rapid and significant diversions in system behavior.

Common Characteristics of ABM

Considering the utility of ABM within the social sciences, Epstein and Axtell (1996) and Epstein (1999) succinctly highlight a number of key characteristics that, while not requisite, are often exhibited by ABM. These characteristics, they suggest, make ABM ideal tools for the study of complex social systems and provide a number of strengths that overcome some of the weaknesses associated with more traditional attempts to understand complex dynamic social systems. A summary of these characteristics follows:

- *Autonomy*: ABMs are devoid of overarching top-down control mechanisms. Rather, each agent within the simulation perceives, reasons, and acts individually. While exchanges of information between agents may occur directly or indirectly through the environment, no centralized controller regulates behavior.
- *Heterogeneity*: ABMs often simulate large numbers of entities as agents, which differ both within and between groups. For example, agents may operate using different decision-making strategies: for instance, probabilistic vs. deterministic reasoning. Agents may also differ by characteristics, with all agents utilizing the same decision calculus but drawing on

different internal characteristics. This ability to capture unit heterogeneity is of great importance, especially when attempting to investigate real-world phenomena where unit homogeneity is rare.

- *Explicit Space*: ABMs represent entities embodied in some abstract or realistic space, allowing the concept of local interaction to be well formed.
- *Local Interactions*: Equation-based models often assume system entities possess complete knowledge of both the world they inhabit and the other entities within it. This is often an unrealistic assumption. ABMs, on the other hand, predominantly deal with localized interactions occurring between entities that are spatially or socially proximate within the simulation environment.
- *Bounded Rationality*: While agents are often bestowed with rational decision-making behaviors, these behaviors can be developed to draw only from localized, limited information. Thus, rationality is limited by the information available at the time a decision is made. In addition, agent behaviors can be designed to utilize limited processing power – *bounded computation* – and therefore do not exhaustively search all possible actions in order to determine an optimal solution. Such representations of rationality are much closer to those employed by human actors in the real world.

In addition to these inherent strengths in modelling social systems, the agent-based methodology also confers a number of distinct epistemological benefits:

- *Accessibility*: Agent-based approaches are often very effective at demonstrating complex concepts to both researchers and a wider audience. ABM elements are most commonly specified at the individual level – that is, as the decision-making strategies of individuals. Specifying model concepts at this level means that model assumptions and propositions can be much more easily understood than models which require complex mathematical abstractions (Bonabeau 2002). This dictates that the ABM audience need not be highly skilled in the field of ABM in order to both inform and interpret ABM. As a result, ABM can often be interrogated by domain experts who may well ask more pertinent questions of models than those whose interests lie predominantly in their development (Gilbert and Troitzsch 2005).
- *Aiding Scientific Discourse*: ABM's intuitive depiction of complex phenomena also dictates that they can often better scientific debate concerning the target system they examine. The process of theory formalization whereby theoretical constructs are formalized for inclusion within an ABM forces researchers to be explicit about the mechanisms their theories propose. This process can often highlight logical inconsistencies which are not readily apparent when theories are represented by traditional verbal or written models. Furthermore, the results of ABM can lead to the development of new questions and the generation of novel hypotheses, some of which may have seemed counterintuitive prior to the observation of ABM.
- *Simulation Experimentation*: ABMs allow for experiments to be performed that would otherwise be impossible due to monetary, ethical, or logistical constraints (Gilbert and Troitzsch 2005). Furthermore, simulation experiments can be performed en masse easily and quickly. Once a model is built, minor adjustments are simple to perform (Gilbert and Troitzsch 2005). In essence, the number of experiments that can be performed is only limited by the computing power and the time available to researchers.
- *Absolute Control*: ABM offers a novel analogue to controlled experiments for examining social phenomena. Researchers can manipulate any number of influencing factors otherwise outside their control in traditional experimentation, thus allowing the exploration of dose-response relationships in endless configurations.
- *Absolute Observation*: ABMs provide synthesis of real-world systems in which perfect observation and measurement can occur. Many social science problems deal with

phenomena for which real-world data cannot be obtained. By contrast, in simulation data sets can be collected to describe every action undertaken by members of a virtual population and furthermore the internal calculus employed by each member in undertaking such actions.

- In Situ *and In Silico Experimentation*: ABM may also identify potential new lines of empirical enquiry. In turn, this external experimentation may lead to the development of better ABM through parameterization and/or validation. This iterative interaction of in situ and in silico experimentation can offer substantial advances in the way phenomena are investigated.

Agent-Based Modelling of Social Systems and the Generative Explanation

Drawing heavily on the inspiring work of Thomas Schelling, recent advances in the application of ABM within the social sciences have seen authors propose that ABM permits a new "third way of doing science" alternate to traditional forms of inductive and deductive reasoning (Axelrod 2005). This approach sees the agent-based computational model or artificial society, as it has become known (Epstein and Axtell 1996), as a new scientific instrument that permits a unique experimental method through which social macrostructures of interest can be "computed" (Epstein 1999, 2006; Hedström 2005). As a branch of analytical sociology, this approach focuses on mechanisms as explanations, and in doing so its advocates suggest that artificial societies allow researchers to systematically establish which microlevel mechanisms *can* and *cannot* be viable explanations for observed macroscopic phenomena: "Agent-based models provide computational demonstrations that a given micro specification is in fact sufficient to generate a macrostructure of interest" (Epstein 1999, p. 42).

This approach views society itself as a form of distributed computational device through which macrostructures are computed from micro-action. The generativist proposes that complex social science phenomena can be understood through the synthesis of their emergence from lower-order action and interaction. The central premise of generative social science is as follows: "If you didn't grow it, you didn't explain its emergence" (Epstein 2006, p. 6). Consequently, the generativist approaches observed macro-phenomena by attempting to identify what combination of micro-conditions is capable of generating them. Or more specifically, "How could the decentralised local interactions of heterogeneous autonomous agents generate the given regularity?" (Epstein 2006, p. 5).

In attempting to answer this question, the ABM is employed. In addressing the above question, the following course of action is proposed: "Situate an initial population of autonomous heterogeneous agents in a relevant spatial environment; allow them to interact according to simple rules and thereby generate – or "grow" – the macroscopic regularity from the bottom up" (Epstein 1999, p. 42). Following this approach it is suggested that theory describing observed social science phenomena can be tested by building ABMs of the proposed microlevel mechanisms of a system and testing if these mechanism are sufficient to produce observed macroscopic regularities of the target. Such macroscopic regularities describe those salient macro-level patterns that are consistently observed in the empirical study of the target system. Examples of such regularities might include right skewed wealth distributions, price equilibria, segregation patterns, or in the case of criminology the spatial clustering of crime.

Whereas traditional statistical or equation-based explanations operate in a *top-down* manner, with the identification of associations between aggregate observations used to make inferences about underlying mechanisms, generative social science operates from the *bottom-up*, identifying *generative explanations* as those hypothesized microlevel mechanisms that produce macro-level output patterns consistent with observed regularities of the target (Epstein 1999). Applying ABM as its principal scientific instrument, generative social science assesses the generative sufficiency of theory. Generatively sufficient mechanisms are those that when employed by an agent population are sufficient to generate macro-level patterns congruent with the target. The more regularities a mechanism is capable of generating, the greater the confidence one can have in its validity. Importantly, if a mechanism cannot produce such regularities, confidence in its validity is reduced, and it may be eliminated as a potential explanation of the target phenomena. Thus, the use of ABM provides a method through which theory can be falsified, a principal requirement of any scientific proposition.

An important observation regarding this form of investigation is that while generative sufficiency is a prerequisite of causal explanation, the converse is not necessarily the case. While mechanisms may be deemed sufficient to bring about the observed effect, they are not necessarily necessary. Thus, while ABMs may identify hypotheses that are generatively sufficient, they cannot be used to infer causal explanation (Epstein 1999; Hedström 2005). This is obviously the case. It is highly probable that a number of different microlevel mechanisms may produce output phenomena consistent with that observed of the target. This however is no different from the scientific discipline as a whole – there is no finite limit of the number of potential explanations one can conjure up for a given phenomena. Generative social science aims to use ABM to identify which of those potential explanations are viable candidates. It is only through further empirical experimentation that the most tenable sufficient mechanisms can be identified (Epstein 1999; Hedström 2005).

The initial task of the generativist, then, is to eliminate those theories that are generatively insufficient, leaving only generatively sufficient candidate explanations. This ability to identify insufficient hypothetical constructs is where the strength of the generative approach lies. Having identified a number of candidate explanations, each should be examined in further detail, considering its plausibility and identifying other potential metrics that may be used to test for the presence of the mechanisms purported using empirical experimentation. As such, the development of generative models guides further empirical observation of the target phenomena, which in turn may identify further potential explanations that can subsequently be assessed for generative sufficiency (Epstein 1999). While this process is unlikely to produce a single viable explanation of a phenomenon, it has eliminated those that are insufficient and implausible or have been falsified through empirical observation, in effect prioritizing theory in terms of its plausibility.

In application, generative social science has demonstrated its utility in exploring a number of social science phenomena. Epstein (2006) provides a review of a number of these endeavors, which include exploring the dynamics of civil violence, potential strategies for controlling epidemics, and the impacts of cultural change within indigenous communities. Note that the choice to treat generative simulation is indeed a choice. Other uses of simulation methodology are available, but are not discussed here.

Linking Microcrime Theory and Macrocrime Patterns: Generative Agent-Based Models of the Crime Event

Having outlined the ABM approach and its potential application in exploring generative explanations of observed phenomena, the following sections now discuss how the ABM can be used to explore propositions derived from criminological theory. As previously discussed this is certainly not the first time the ABM methodology has been suggested or undertaken within criminology. Here, however, this entry draws heavily on Epstein's principle of generative social science advocating the application of ABM in systematically exploring the robustness of criminological theory.

To undertake the generative approach in examining the interactions of crime, two distinct constructs are required: (1) micro-specifications of the crime event and (2) empirically derived macroscopic regularities of crime against which the explanatory sufficiency of such micro-specifications can be assessed. Luckily within criminology both abound.

Micro-specifications of the Crime Event

For a theory to be explored in the ways proposed here, there is a sole requirement. Theories must provide individual-level mechanisms describing how they propose observed outcomes come about. As such, researchers are forced to specify mechanisms not correlates. A number of criminological theories provide such individual-level depictions of the crime event. This section does not attempt to sketch these well-known theories in detail, but instead aims to highlight a few that might be explored through simulation, hinting which agents and rules would be pertinent in such research:

1. Routine activity approach and pattern theories of crime outline how the characteristics and interactions of individual-level activities dictate the spatial and temporal distribution of offending and victimization. Agents may be the potential offenders, potential victims, and potential guardians, moving around in a geographical space. Decision rules specify how the various agents react on each other's presence.
2. The rational choice perspective offers a tangible offender decision calculus that relies on boundedly rational assessments of the perceived risks, rewards, and efforts encountered in potential pre-crime situations. Agents are individual would-be offenders, reacting on various opportunities that may have rewards and efforts needed to victimize targets and associated risks, partly in terms of social (dis)approval) by other agents that form a social network, having certain relations among each other.
3. Social learning theory provides hypotheses concerning mechanisms of offender reinforcement. Here, agents may be potential offenders who repeatedly get the opportunity to misbehave, and their decision calculus is taking stock of previous decisions to offend and their success or failure.
4. Network theory of peer association. Agents may be individuals, having certain peer relations among each other. Decision rules to engage in crime as well as decisions to (dis)continue being a member of a peer network are specified in terms of number of peer agents that commit simulated crimes themselves.
5. Social disorganization theory proposes communications of risk and crime control among peer (neighborhood) groups. Agents will be individuals in a social environments, modelled by social connection and disconnection between various agents. Agents may engage in crime and may exercise crime control against other agents; decision rules will be in terms of risk communication between agents and the likelihood that agents will engage in crime control against others.

All of these examples provide theoretical depictions of autonomous, heterogeneous actors whose interactions are well suited for formalization in computational agent-based models. While deriving formalized mechanisms will likely be more difficult for some theories than others, undertaking such formalization is of obvious utility. That is to say, formalized depictions of theoretical constructs are desirable irrespective of modelling, computational, or otherwise.

Assuming then that agents and their associated behaviors can be specified to reflect these theories, the plausibility of proposed mechanisms can then be evaluated by assessing their ability to generate known macro-patterns of crime – several of which are now discussed.

Macrostructures of Crime

Empirical research in the field of criminology has highlighted a number of macroscopic regularities of crime consistently observed across a wide range of crime types, contexts, localities, and measurement instruments (Birks et al. 2012). These regularities are the emerged trace effects of whatever mechanisms are indeed operating at the microlevel and thus are those that many individual-level theories of crime attempt to explain. To illustrate, five regularities are briefly summarized. While by no means exhaustive, this list alone demonstrates well that individual crime variability can give rise to macro-level crime pattern predictability. Again, this section refrains from paying detailed and due attention to the literature, instead just hinting at what is out there:

1. *Spatial and Temporal Clustering* – Research consistently shows that the spatial patterning of crime conforms to a Pareto law of concentration,

with the majority of crime concentrating in a minority of geographic areas often referred to as crime hot spots or hot places. Similarly, crime is not uniformly distributed over time. Very few locations experience constant levels of victimization. Instead, crime hot-times where the level of victimization is disproportionate relative to other times are often observed, be they by hour of the day, day of the week, or month of the year.

2. *Repeat and Near-Repeat Victimization* – Crime not only concentrates spatially and temporally but also with respect to a small number of persons/places/targets. These individuals commonly referred to as "repeat victims" experience disproportionate levels of victimization. Recent research has also consistently demonstrated spatiotemporally clustered near-repeat victimization, i.e., victimization of other targets close by earlier victimized targets.
3. *Repeat Offending* – While research has demonstrated that victimization concentrates in specific locations, at specific times, and against specific targets, research also consistently shows that the incidence of offending is also disproportionally distributed among offenders. As such, a small proportion of offenders are commonly responsible for a large proportion of offending.
4. *Journeys to Crime* – Numerous studies examining the spatial characteristics of crime trips have demonstrated the presence of a distance decay function where the majority of crime trips lie within a short distance of offenders' homes. Such observations have been made across a range of different crime types including residential burglary, robbery, and rape.
5. *The Age-Crime Curve* – Research consistently shows that younger offenders are more likely to commit crime than their older counterparts. The oft-cited age-crime curve follows a steady decay after initially ramping up through adolescence and peaking in the late teens/early twenties.

Computational Experiments

Given proposed theories of the crime event and the observed macro-regularities of crime, the generative ABM is ideally placed to explore if the former is indeed, as theorists would suggest, capable of generating the latter. If propositions provided by theory do reflect proximal mechanisms that are significant in the commission of crime, one would expect a virtual population of offenders operating according to them to generate patterns of crime which exhibit similar characteristics to those observed in empirical research.

In testing these hypotheses simulation experiments are performed whereby the mechanisms under which virtual populations operate are systematically manipulated. Subsequently, simulated crime patterns produced by virtual offenders are compared to suitable macroscopic regularities of crime. From this analysis conclusions are drawn about the generative sufficiency of the mechanisms formalized and moreover the validity of the theories from which they are derived.

Adapting from Schlesinger (1980), Fig. 1 depicts conceptually how such efforts can be undertaken. First, a theoretical model is selected for study and formalized as a conceptual model, through which key propositions of theory are defined. Model qualification then assesses through observation if this conceptual model is an adequate representation of the original, and often "fuzzy," theoretical model, posing the question: *do these conceptual formalizations and the proposed interactions amongst them adequately capture the theoretical propositions they aim to explore*? Where discrepancy or ambiguity exist, conceptual formalizations are refined to reflect more accurately the theoretical constructs of interest. Once qualified, a computational model representing the conceptual model is created, such computational models commonly define agents within the model, their characteristics, behaviors, and the environment in which they are situated. Model verification then assesses whether these computational constructs sufficiently represent the conceptual model from which they are derived. Verification often takes place by examining the output of specific algorithms in relation to particular scenarios. While the emergent properties of simulation are often beyond the scope of model verification (as they are often unexpected), model verification allows

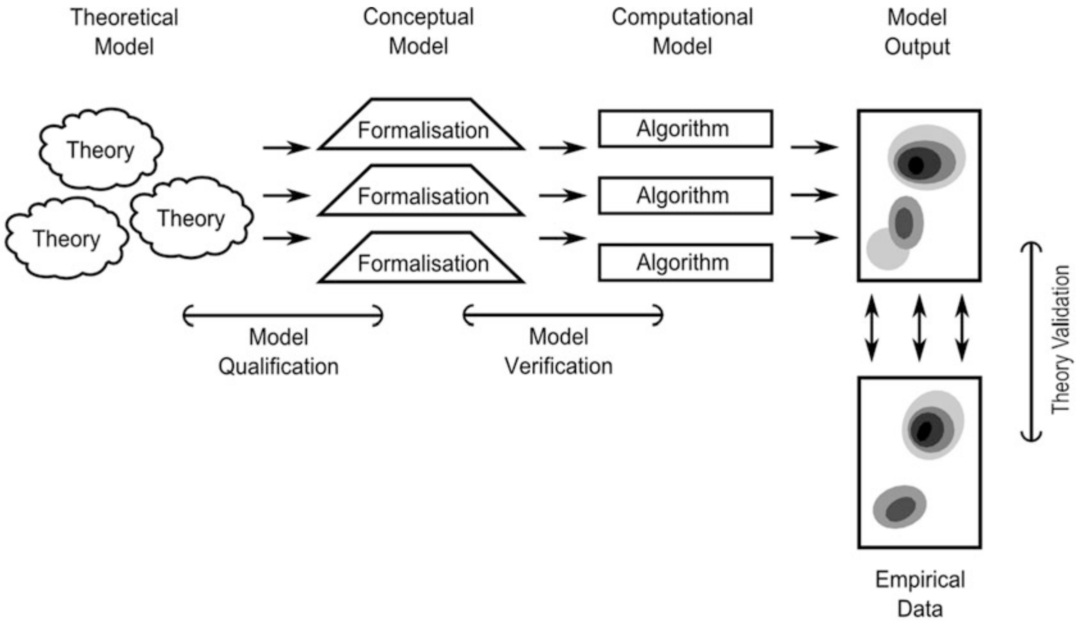

Agent-Based Assessments of Criminological Theory, Fig. 1 Generative ABM development, validation, and experimentation

computational constructs to be examined individually with respect to the results expected by theory. Given a sufficient number of test cases where model outputs are consistent with the results expected by both theoretical and conceptual models, the computational model may be considered appropriately verified. After both conceptual and computational models have been assessed, model outcome validation assesses the sufficiency of theoretically defied computational constructs in capturing the output behavior of the target system as a whole. This is done by comparing characteristics of both simulated data and empirical data. In undertaking such endeavors, systematic controlled simulation experimentation is not just sufficient, but necessary. The following section outlines several characteristics critical in the design of implementation of such simulation experiments:

- *Address Theoretical Research Question* – Just like traditional experimentation, simulation experiments must be designed to address a specific theoretical research question. While simulation can be a useful tool, it can also be unwieldy when no clear goals of experimentation are specified. While perhaps obvious, this requirement is of considerable importance.
- *Systematic Control of Experimental Conditions* – Further mirroring traditional experimentation, in silico experiments should follow strict design principles. Appropriate counterfactual simulation states can and should be developed through which the impact of particular model manipulations can be systematically assessed. To illustrate, when exploring the impact of proposed offender behaviors, equivalent agent behaviors should be developed to reflect both the absence (a behavioral control condition) and presence (a behavioral experimental condition) of a proposed mechanism (see Birks et al. 2012). Subsequently simulations should be run with agents operating under both experimental and control conditions, and resulting outputs analyzed.
- *Simulation Replication* – While the results of a single simulation may be interesting, the stochastic nature of ABM dictates that experiments must be reproducible in order for

them to be deemed valid (Axelrod 1997; Townsley and Birks 2008; Townsley and Johnson 2008). Simulation replication, just like its empirical counterpart, aims to highlight implementation-specific factors which may influence observed results. Indeed, simulation models may be more susceptible to these factors due to their relative complexity and the skill set required in examining them (Axelrod 2005). As a result, without systematic and rigorous methods for scrutinizing simulation, factors unbeknownst to onlookers, or even the investigator, may influence simulation outcomes and, hence, the inferences being drawn from them. *Within-model* replication commonly averages the results of specific simulation configurations over numerous "runs." Such replications provide an important method through which the range and consistency of possible simulation outputs can be explored, assessing simulation statistical conclusion validity (Townsley and Johnson 2008). Thankfully, the nature of development environments used in simulation development dictates that within-model replication is relatively easy to perform, requiring little investment other than time and computing power. The main effort, however, is making sense of the multiplicity of generated results.

- *Robustness Testing* – Researchers commonly select model parameters that dictate the initial conditions of a given simulation. Such parameters might describe the size of the simulation environment, agent populations, or specific thresholds at which agents undertake certain actions. Such parameters may be derived from empirical research or, where appropriate, common sense estimations. Importantly, this mapping between input parameters and output behavior should be scrutinized (Axelrod 1997; Gilbert and Troitzsch 2005). Model robustness tests seek to examine the influence that changes in model parameters have on outcome patterns, thus ensuring that observed results are not unique to specific parameters. Akin to sensitivity analysis performed in a number of statistical models, simulation robustness tests involve systematically manipulating model parameters (ideally reflecting both model initial conditions and behavioral parameters (Fung and Vemuri 2003)), rerunning simulations, and examining changes in model outcomes. While it is unlikely that model outputs from such robustness tests will not differ somewhat, when variations in outcomes are observed, the evaluation of robustness testing assesses distributional equivalence. That is, *do the same types of macro-patterns remain given changes in initial parameters*?. Such tests ensure that model plausibility does not break down when seemingly innocuous changes in model parameters are performed (Fung and Vemuri 2003). When this does occur, underlying model structures should be scrutinized (returning to the processes of qualification and verification), ensuring that observed results are indeed indicative of unforeseen but plausible interactions, rather than an indication of underlying errors in the model formalization. While it is often impractical to sweep an entire range of possible input parameters, it is reasonable to explore a number of key model parameters within the computational constraints of a proposed study. In addition, models that incorporate stochastic elements should also be subjected to robustness testing with respect to selected random number seeds (Axelrod 1997).
- *Levels of* In Situ/In Silico *Equivalence* – In comparing empirical and simulated data, it is obvious that various levels of equivalence can be observed. Axtell and Epstein (1994) propose multiple cumulative levels of in situ and in silico system equivalence, beginning at qualitative macro equivalence (that is, where simulated outcomes offer a macro-level caricature of the system of interest) to both micro and macro quantitative equivalence – where both individual- and aggregate-level regularities can be shown to be statistically similar to empirically observed regularities. With respect to the study of crime, it is suggested that any model providing some level of in situ/in silico equivalence should warrant interest from scholars. Importantly, in this regard such demands for quantitative replication of

"real world results" are dependent on the often not warranted hypothesis that there is a clear and concise picture of the totality of real-world data. For example, while "distance decay" is a generally recognized characteristic in the journey-to-crime literature, there is not even a beginning of a commonly supported body of knowledge about the exact, let alone parameterized, form of the distance decay curve's functional form. Thus, while abstract models may well not provide quantitative equivalence, they do permit estimations of generative sufficiency and in doing so offer insight into the likely underlying dynamics of those processes which are difficult to characterize.

- *Multiple Output Measures* – A basic premise of the scientific enterprise is to identify parsimonious depictions of a system which are compatible with the widest variety of observable phenomena. Thus, when assessing in situ/in silico equivalence, the use of multiple distinct outcome measures is desirable. In this sense, while a proposed mechanism may generate highly plausible patterns of one macroscopic regularity, it is important to examine other related regularities. For instance, if a model of target selection produces spatially concentrated crime similar to that observed in empirical research but analysis of repeat offending highlights that only a single offender is responsible for all victimization throughout a virtual society, confidence in the validity of such a mechanism should be reduced. Here, it is reasonable to assume that some proposed mechanisms may only have explanatory capacity with respect to one regularity. However, the most tenable (and most useful) are likely those which are sufficient to explain multiple regularities of the target system.

Conclusion

This entry has discussed how computational agent-based models can be used to gain insight into the potential ramifications of individual-level crime event mechanisms proposed by theory and, in particular, assess their plausibility in explaining patterns of crime that are commonly observed in empirical studies. Given the numerous constraints placed upon observation and experimentation which face those who aim to study the crime event, the computational agent-based modelling approach provides criminology scholars with a unique and significant method that is complimentary to existing theoretical and empirical efforts. Contrasting itself against the commonly applied statistical explanation where associations are commonly observed, quantified, and mechanisms through which they might come about inferred, ABM provides a bottom-up approach to understanding crime patterns where the ability of hypothesized mechanisms to generate known associations is assessed through simulation experimentation.

This approach allows researchers to explore complex, dynamic, spatially situated interactions between boundedly rational, heterogeneous actors without the need to suppress important complexity and, in doing so, provides an analogue to the petri dish for criminology scholars. By undertaking controlled simulation experimentation using ABM, the underlying dynamics of proposed mechanisms can be examined and the viability of theories from which they are derived better explored, in ways that are often beyond the scope of traditional empirical activity alone. The purpose of the generative agent-based model is not to mimic reality, but instead to provide a tool through which the content, extent, and ramifications of criminological theories can be systematically explored. Not predicting, but understanding, complex behavior is the ultimate goal of generative simulation research.

Yet, while there are many advantages to a generative study of crime events, there are also clearly many considerations that must be undertaken in order that the utility of simulation be capitalized upon. In summary, simulation experiments must be systematic, rigorous, have clear goals, and be subject to replication and robustness testing.

While one cannot overstate the importance of theoretical and empirical efforts, the computational agent-based model and the generative approach it affords may well provide a complimentary method through which the consistency and plausibility of theory can be assessed without need for significant ethical, logistical, and monetary investment. Identifying those mechanisms that offer viable generative explanations for observed crime patterns both increases our ability to select those theories whose propositions are most likely to reflect real-world mechanisms and diminishes the likelihood of pursuing mechanisms that lack explanatory capacity.

Furthermore, the principle requirement of the generative approach – an adequate micro-specification of the crime event from which to derive computational constructs – is of considerable use in highlighting those theories that lack adequate specification of the mechanisms through which they propose observed outcomes come about. Without revision, such theories are likely of little use in developing either simulation or, more importantly, crime prevention intervention. Looking forward, it is hoped that an adaptation of the crime script methodology (Cornish 1994) will bear fruit in this regard, given the obvious parallels between both approaches.

While assessing the generative sufficiency of theory is but one potential application of simulation, it is one that we believe (1) is of considerable utility and (2) can be realized.

Related Entries

- ▶ Agent-Based Modeling for Understanding Patterns of Crime
- ▶ Agent-Based Models to Predict Crime at Places
- ▶ Dynamical Simulation in Criminology
- ▶ Simulation as a Tool for Police Planning

Recommended Reading and References

Axelrod R (1997) The complexity of cooperation: agent-based models of competition and collaboration. Princeton University Press, Princeton

Axelrod R (2005) Advancing the art of simulation in the social sciences. Idea Group, Hersey

Axtell R, Epstein J (1994) Agent-based modeling: understanding our creations, vol. 9. Bulletin of the Santa Fe Institute

Birks D, Donkin S, Wellsmith M (2008). Synthesis over Analysis: Towards an Ontology for Volume Crime Simulation. In: Liu L, Eck J (eds) Artificial crime analysis systems: using computer simulations and geographical systems. Hershey, PA: IGI Global, pp 160–191

Birks D, Townsley M, Stewart A (2012) Generative models of crime: using simulation to test criminological theory. Criminology 50(1):221–254

Bonabeau E (2002) Agent-based modeling: methods and techniques for simulating human systems. Proc Nat Acad Sci U S A (PNAS) 99(3):7280–7287

Bosse T, Elffers H, Gerritsen C (2010) Simulating the dynamical interaction of offenders, targets and guardians. Crime Patterns Anal 3(1):51–66

Brantingham P, Brantingham P (2004) Computer simulation as a tool for environmental criminologists. Secur J 17(1):21–30

Brantingham P, Tita G (2008) Offender mobility and crime pattern formation from first principles. Idea Group. In: Liu L, Eck J (eds) Artificial crime analysis systems: using computer simulations and geographical systems. Hershey, PA: IGI Global, pp 193–208

Brantingham P, Glasser U, Jackson P, Kinney B, Vajihollahi M (2008) Mastermind: computational modelling and spatiotemporal aspects of crime in urban environments. In: Liu L, Eck J (eds) Artificial crime analysis systems: using computer simulations and geographical systems. Hershey, PA: IGI Global, pp 252–280

Cornish D (1994) The procedural analysis of offending and its relevance for situational prevention, vol 3, Crime prevention studies. Criminal Justice Press, Monsey

Eck J, Liu L (2008) Contrasting simulated and empirical experiments in crime prevention. J Exp Criminol 4:195–213

Elffers H, Van Baal P (2008) Realistic spatial backcloth is not that important in agent based simulation research. An illustration from simulating perceptual deterrence. In: Liu L, Eck J (eds) Artificial crime analysis systems: using computer simulations and geographic information systems. Hershey, PA: IGI Global, pp 19–34

Epstein J (1999) Agent-based computational models and generative social science. Complexity 4(5):41–60

Epstein J (2006) Generative social science: studies in agent-based computational modeling. Princeton University Press, Princeton

Epstein J, Axtell R (1996) Growing artificial societies: social science from the bottom up. MIT Press, Cambridge

Fung K, Vemuri S (2003) The significance of initial conditions in simulations. J Artif Soc Soc Simul 6(3)

Gilbert N (2004) Agent-based social simulation: dealing with complexity. http://cress.soc.surrey.ac.uk/resources/ABSS

Gilbert N, Troitzsch K (2005) Simulation for the social scientist, 2nd edn. Open University Press, London

Groff E (2007) Simulation for theory testing and experimentation: an example using routine activity theory and street robbery. J Quant Criminol 23(2):75–103

Hedström P (2005) Dissecting the social: on principles of analytical sociology. Cambridge University Press, Cambridge

Johnson S (2008) Repeat burglary victimisation: a tale of two theories. J Exp Criminol 4:215–240

Liu L, Wang X, Eck J, Liang J (2005) Simulating crime events and crime patterns in a RA/CA model. Idea Publishing, Reading

Malleson N, Evans A, Jenkins T (2009) An agent-based model of burglary. Environ Plan B Plan Des 36: 1103–1123

Schelling T (1978) Micromotives and macrobehaviour. W.W. Norton & Company, New York

Schlesinger S (1980) Terminology for model credibility. Simulation 34:101–105

Townsley M, Birks D (2008) Building better crime simulations: systematic replication and the introduction of incremental complexity. J Exp Criminol 4(3):309–333

Townsley M, Johnson S (2008) The need for systematic replication and tests of validity in simulation. In: Liu L, Eck J (eds) Artificial crime analysis systems: using computer simulations and geographical systems. Hershey, PA: IGI Global, pp 1–18

Van Baal P (2004) Computer simulations of criminal deterrence: from public policy to local interaction to individual behaviour. Boom Juridische uitgevers, The Hague

Wang X, Liu L, Eck J (2008) Crime simulation using GIS and artificial intelligent agents. In Liu L, Eck J (eds) Artificial crime analysis systems: using computer simulations and geographical gystems. Hershey, PA: IGI Global, pp 209–225

Agent-Based Modeling for Understanding Patterns of Crime

Elizabeth R. Groff
Temple University, Philadelphia, PA, USA

Synonyms

Individual-based simulation modeling

Overview

Many questions in criminology focus on dynamic processes and individual decision-making. For example, crime events represent the end result of a multitude of decisions made by a variety of different people in the context of specific situations. Potential offenders, potential victims, police, and other informal guardians of places make choices that bring them together at the same place and time. However, data describing any of those individual decisions prior to, during, and after the crime event are only rarely available and never complete. Once at a place, potential victims may take actions that make them a more suitable target. Potential offenders observe these actions and reevaluate the likelihood of getting away with a crime. At the same time, likely offenders also notice the actions of potential guardians and make adjustments. Even if data were collected to document the actions of individuals, the dynamic and nonlinear nature of human interaction and decision-making makes them difficult to study using traditional statistical techniques.

Agent-based modeling is a type of computer simulation modeling that allows the study of dynamic processes and the outcomes that emerge from individual decisions. An artificial world, akin to a scientific version of a video game, is created. This model is a simplified version of "real life" that contains only the most important aspects of the behavior being simulated. The computer program is initiated and the agents (i.e., representations of individuals) in the model interact. The patterns of outcomes produced by those interactions, for example, a distribution of crime events, is then compared with the pattern theory would predict or empirical data indicates. Since, we would expect that crime would concentrate across space, so geographic patterns of crime from a model should evidence this characteristic. Agent-based models (ABMs) allow group-level outcomes to emerge from individual decision-making. They enable researchers to conduct "what if" experiments because the outcome under alternative scenarios is known (i.e., both the control condition and the treatment condition can be simulated). The bottom-up approach, rather than the top-down of traditional statistical models, combined with the ability to systematically change the rules governing individual decision-making, creates a powerful modeling environment for exploring and strengthening theory as well as uncovering new insights.

Fundamentals of Agent-Based Modeling

Some Background and Basic Definitions

Computer simulation is a broad area of study. Flight simulators are a familiar example of a computer simulation that allows the user to practice flying an airplane. Agent-based modeling is one type of simulation modeling within the general category of computer simulation modeling "that enables a researcher to create, analyze, and experiment with models composed of agents that interact within an environment" (Abdou et al. 2012, p. 141). *Models* are simplified versions of the real world and have long been used to understand both structure and process. *Simulation models* are simply models that are programmed to run inside a computer and as such they share many characteristics with computer games but their purpose is scientific rather than recreational. An *agent-based model* (ABM), then, is a simplified representation of a real-world process that is implemented in the form of a computer program. ABMs employ a bottom-up approach in which agents are imbued with unique characteristics and general behavioral rules (Epstein and Axtell 1996; Gilbert and Troitzsch 2005). The *agents* in an ABM are typically individuals but can also represent collective entities such as neighborhoods or schools as well as companies or organizations. The fact that the agents are autonomous and make decisions based on the changing circumstances that occur during the running of the model is what makes ABMs unique. Because the decisions of individual agents are at the heart of ABMs, they are said to take a bottom-up approach to modeling processes. The individual decisions made by agents are what generate the outcome. This characteristic is what makes ABMs so valuable for modeling social processes as outcomes of individual decisions. It is also what has led some to call it a *generative social science* (Epstein and Axtell 1996). A core characteristic of agent-based models (shared by cellular automata) is the capacity to produce unexpected results. The term *emergence* is used to describe outcomes from a model that were not anticipated.

Simulation models, in general, have a variety of uses (Gilbert and Troitzsch 2005). They can be used to improve our *understanding* of processes that are hard to measure directly. Offender decision-making is an example of one potentially fruitful area of inquiry because we can look inside the criminal decision-making process and examine the relative importance of each piece of information considered. Another use of ABMs is for *prediction*. If a model is built that can faithfully reproduce a dynamic process, then it can be used to predict an outcome based on different inputs or different behavioral rules. One example of prediction using ABMs are the models used to predict changes in demographic structure based on decisions related to decisions regarding age of first childbirth, family size, and other related factors. Not all ABMs are suitable for prediction; specifically in the case of micro-level ABMs that test nonlinear theories, prediction is not an appropriate standard because they are inherently unpredictable (Gilbert and Troitzsch 2005). ABMs are of particular interest to scientists because of their utility for *discovery* and *formalization*. A simple model can be created that embodies a theory about how a process works. The simulation can be run in the computer and the outcomes tested to discover whether the predicted outcomes occur. In order to create such a model, the scientist must formalize critical aspects of the theory in order to program it. The process of formalization is important because it forces scientists to increase their precision as compared to verbal or statistical realizations of a theory. In this tradition, simulation allows for the exploration and elaboration of theory (Dowling 1999). Some have proposed that ABM can be used to investigate theoretical mechanisms and provide a way to eliminate potential explanations that cannot be grown from individual behavior (Eck and Liu 2008a; Epstein and Axtell 1996).

Components of an ABM

An ABM has two main components, the agents and the environment in which the agents interact. While there is no standard definition, an agent can be thought of as an "autonomous goal-directed software entity" (O'Sullivan and Haklay 2000, p. 1410). Agents have characteristics and behaviors which are modeled after their real counterparts. The environment in which they interact also can have characteristics and rules.

Agents and Their Characteristics

Agents can represent people such as potential offenders or police officers. Even groups of people with a collective identity such as neighborhoods, gangs, businesses, or city departments can be agents.

Agents share several general characteristics, only a few of which are highlighted here (Crooks and Heppenstall 2012). First, they are autonomous; they make decisions for themselves. Second, agents are heterogeneous. They have a set of characteristics that are unique to them. For example, an agent representing a police officer can have a race, a gender, an age, a number of years on the job, and any other characteristic pertinent to the goal of the model. More importantly, each police officer agent can have their own values for each of those characteristics. This means the modeler can have an agent's characteristics factor into her own decisions as well as the decisions made by other agents.

Third, agents are proactive and have goals that guide their actions. They are interacting and changing over time. They can interact with other agents and perceive and react to their environment. In the case of individuals, the actions they undertake often require movement. For example, a police officer agent may begin patrolling using a random pattern within her assigned beat. She is constantly scanning her immediate environment with the goal of identifying crime events and stopping them. A potential offender may proceed through a number of routine activities. During the course of those activities he may notice the quality of opportunities to commit a crime he encounters. Whether or not he takes advantage of those opportunities may depend on whether it will make him late for his legitimate job, whether a police officer is at the same place and his perceived chance of success as compared to potential reward. But organizations, such as businesses, can also be dynamic and take purposive action. For example, they can choose to reinvest in the community in reaction to public pressure or continue to prioritize maximizing profits.

Each type of agent in a model is given a set of rules that guide their behavior and interactions with other agents and their environment (Crooks and Heppenstall 2012). For example, an offender agent would have a different rule set than a police officer.

Agents are typically created within an object-oriented programming environment (Abdou et al. 2012). Such programs contain collections of *classes* and *objects*. Classes describe groups of agents that share the types of actions they can undertake. For example, citizens would be in a different class from police officers since only police officers can make arrests. Each class has a set of shared *attributes* and *methods*. *Attributes* are the characteristics of the agents. For example, attributes of age, offending propensity, and income are possible if the agents represent people. Characteristics are unique to each agent. *Methods* are the actions members of a class can undertake (e.g., movement, making an arrest, committing a crime). During the model run, classes are *instantiated* in the form of objects representing individual agents with unique characteristics. For example, the first agent might be 24 years old, have an offending propensity of .49, and an income of $129 per week. The next agent might be 62 years old, have an offending propensity of .05, and income of $1,129 per week. Both agents have the ability to move and to commit crime because these methods belong to each agent in that class.

Environments in ABM

All agents exist in an environment and this is what allows them to interact with each other and with their surroundings (Crooks and Heppenstall 2012; Gilbert and Troitzsch 2005). However, there are several types of environments that can be used, and the type of environment chosen is dependent on the phenomena being modeled (Gilbert and Troitzsch 2005). In some models, there is no need for movement. For example, deterrence theory has been tested by modeling an agent's likelihood of engaging in income tax fraud based on their own and their neighbor's experience with being audited and the perceived rewards of cheating (van Baal 2004). Environments can also provide a spatial context for agents to interact.

Spatial context can be derived from the position in geometric space or geographic space. Geometric spaces can be represented as grids (similar to the squares on a checkerboard) or networks. Each agent is located within the geometric space. Some models have agents on a grid that can move from one adjacent grid cell to another. Other models have agents connected to one another via a social network (with the agents as nodes and the connections as lines). Networks in models can represent connections that are physical (e.g., a transportation network) or relational (e.g., friendships or trading relationships).

Models that use geographic space are enhanced by situating agents in "real" environments. These "real" environments can be raster surfaces describing land use or the street network from a city. Such simulations use data created by a geographic information system as the environment in the simulation model but do not actively use the topological properties of the data during the model run.

There are also software packages that integrate GIS and ABM to provide a platform for the dynamic modeling of individuals across space and time. One example is Agent Analyst, which follows the middleware approach in which the temporal relationships are handled by the ABM software and the topological relationships are managed by the GIS (Brown et al. 2005). Agent Analyst combines two of the most popular packages for ABM and GIS, the Recursive Porous Agent Simulation Toolkit (Repast) and ArcGIS. To make the software easier to use, Agent Analyst is built using the rapid development version of Repast called Repast for Python Scripting (RepastPy) which has a graphical user interface that automates much of the programming to create the framework of a model (see Johnston, 2012 for an introduction to Agent Analyst). Agent Analyst is designed to be added into ArcGIS as a toolbox. Once the toolbox is added in ArcGIS, models can access shapefiles allowing (1) individual agents to become spatially aware and (2) the visualization of agent movement and decision outcomes (e.g., patterns of crime events).

The integration of ABM and GIS leverages the temporal capabilities of ABMs and the spatial capabilities of GIS. ABMs permit the researcher to (1) collect data about the characteristics of each individual present during an interaction, (2) randomly assign characteristics to agents greatly reducing the possibility of systematic bias, (3) have agents make independent decisions within behavioral guidelines, and (4) systematically vary one attribute while holding all others constant to undertake controlled, repeatable experiments (Epstein and Axtell 1996). GIS makes it possible to take into account how the characteristics of the real environment (e.g., transportation networks and land use) impact the activities of spatially aware agents.

ABM as Methodology

ABM is a research methodology in its own right (Axelrod 2007). Some have suggested it is a third type of scientific inquiry (Axelrod 2007) in addition to deduction and induction. Like deduction, the beginning point is a set of theoretically based assumptions, but instead of proving theorems, ABM produces data which is analyzed inductively. But these data have not been measured empirically (in the real world); they have been generated from agent interactions in a virtual world. It lends itself to conducting thought experiments, but in a virtual world.

Other scholars have labeled ABM a third symbol system in addition to natural language and mathematics (Ostrom 1988). ABM is capable of incorporating both natural language descriptions of behavior and relationships as mathematical ones but is not limited to one or the other. ABM allows the researcher to examine "bottom-up" processes that involve the interactions of heterogeneous individuals with each other and with their environment. Thus, ABM can be used to systematically examine complex and dynamic relationships at the individual level.

The guiding principle of ABM is "simpler is better." Modelers want to build a model that is as simple as possible while maintaining the most important aspects of the target phenomena. This is primarily because complex virtual behavior often develops from relatively simple rule sets. In order to recognize when surprising (i.e., emergent) phenomena occur and to be able to explain what is happening in the model, the modeler must be clear about every aspect of the model (Axelrod 2007). It is easier to be clear when the model is simple.

This section focuses on ABM as a methodology and details the process of conducting such investigations. In other words, it provides the series of steps that are followed to construct an ABM. Although presented in a sequence here, these steps are often iterative with the results of one step causing a reevaluation of an earlier step.

Building an ABM

The initial task is to identify a problem that needs solved, a theory that needs tested, or a question that needs answered (Abdou et al. 2012; Gilbert and Troitzsch 2005). For example, a researcher might want to test a theory of how gangs form. ABMs are also frequently constructed to investigate regularities in patterns of behavior observed at societal or macro levels (Abdou et al. 2012; Gilbert and Troitzsch 2005). One example involves the clustering evident in the spatial distributions of crime events. Several authors have used the ability of their models to replicate clustering as a by-product of the individual decisions of agents (Birks et al. 2012; Groff 2008).

The next task is to examine existing theories which are important to explaining the phenomena of interest. When building an ABM, it is necessary to systematically examine the specific components of theory that may be relevant and to attempt to define them as explicitly as possible based on existing research (Birks et al. 2008, 2012; Groff 2008). While models vary in how faithfully they represent reality, they typically operate on the principle that “simpler is better”; thus a primary goal of modelers is to try to assemble the most parsimonious model to answer a question. The degree to which the theory is represented in the model represents the structural validity of the model (An et al. 2005). Simulation models, in particular, start with simple models and then systematically add complexity to ensure that the dynamics are well understood before continuing (Gilbert and Troitzsch 2005). These two tasks provide the basis for the more programming-oriented ABM building steps.

At this point, the modeler is ready to create a conceptual model. This is usually a visual diagram that captures both the essential constructs and how they are related to one another. The types of agents to be included in the model are specified, as well as the environment. A model exploring routine activity theory would, at a minimum, need to have potential offender agents, target agents, and guardian agents (Groff 2007). One examining crime pattern theory would need to include activity spaces of offenders (Brantingham and Brantingham 2004).

Once a model has been specified, it is programmed. The constructs of the theory are formalized at this point so they can be coded in the computer program. In some cases, the constructs are formalized as clearly stated verbal guidelines that underlie the behavior of agents, their interactions with other agents, and their interaction with the environment. For example, a potential offender will not commit a crime if a police officer agent is present. In other cases, the definition of these constructs takes the form of mathematical equations for evaluation of specific situations an agent encounters during the course of a simulation. Where theory is not detailed enough for implementation or does not address an issue, empirical research is used to enhance the representation of behavior within the model. The programming to implement the model is typically done via a software package.

Random numbers are an important component of ABMs. They are used to represent uncertainty in the model. Random number generators (RNGs) are used to provide numbers that fit a statistical distribution (e.g., normal, Poisson, or uniform). The modeler chooses the distribution that reflects the assumption of the model. The seed, or beginning number for the RNG, produces a set of random numbers. Each time the same seed number is used, the RNG produces the same set of numbers. This critical property of RNGs is what enables experiments to be run in ABMs. If an assumption of the model is changed but the same seed or series of seeds is used in the experiment, everything else is held constant from one experiment to the next. Thus, the researcher knows any changes in outcome are due to the change in the model assumptions. Seeds can be systematically varied across a set of runs and the results from those runs averaged before being presented as model results. This ensures the outcomes are not dependent upon the numerical value of the seed.

RNGs are also used to compensate for the fuzzier areas of what is known about how a phenomena "works." Returning to the example of exploring guardianship, we know bystanders in a situation act as informal guardians but we do know exactly how the presence of those individuals translates into potential risk as perceived by an offender, and we are not likely to be able to collect data on that aspect of offender decision-making. If we assume it is equally likely that each agent present in a situation represents one additional unit of guardianship, as it is that they represent five units of guardianship, we can use a uniform random number generator that ranges from 1 to 5. Since the distribution chosen is uniform, each time the RNG generates a new number it has an equal probability of being a 1, 2, 3, 4, or 5.

Verification, Validation, and Sensitivity Analysis

Three types of model testing are important to understanding the quality of the model results, verification, validation, and sensitivity analysis. "*Model Verification* is substantiating that the model is transformed from one form into another, as intended, with sufficient accuracy. Model verification deals with building the model *right*." (Balci 1994, p. 121). This usually takes the form of debugging and logic-testing both during the programming of the model and during its testing to ensure the interactions produced by the code are as the theory intended. When ABMs include random numbers, without specifying the seed, no two model runs are alike and the only point of comparison is the distribution of results that the theory would suggest (Gilbert and Troitzsch 2005).

In contrast, "*Model Validation* is substantiating that the model, within its domain of applicability, behaves with satisfactory accuracy consistent with the study objectives. Model validation deals with building the *right* model." (Balci 1994, p. 123). This aspect of model testing answers the question of how well the model represents the "target." Model validation is analogous to the criteria of external validity used in traditional modeling (Gilbert and Troitzsch 2005). Validity in an ABM context is not an all or nothing proposition (Law and Kelton 1991). Rather, a model can have varying levels of validity relating to different properties of the target. There are several challenges in evaluating model validity of ABMs (Gilbert and Troitzsch 2005). First, both the target and the model have random components. Thus, the outcomes will vary across model runs. The critical question concerns how much the model varies from the expected statistical distribution of the outcomes. Second, many models are sensitive to the starting values of parameters (e.g., the ratio of motivated offenders to suitable targets in the model). Third, there may be problems with the "real" data rather than with the model results. This is an important issue especially when data sources are nonexistent (e.g., threshold of risk versus reward when deciding to commit a crime) or when they are unreliable (e.g., reported crime data only contains events reported by the public and those deemed to be crimes by the police officer taking the report) (Eck and Liu 2008a; Groff and Birks 2008).

Some researchers have compared the results of their simulations to widely recognized regularities in crime patterns, also known as "stylized facts" (Birks et al. 2012; Groff 2008). For example, crime patterns should exhibit (1) a high degree of clustering, (2) concentration of crime in relatively few places, (3) relatively few offenders responsible for most of the crime, (4) rather few victims accounting for most of the victimization, and (5) non-static patterns of crime over time. When results from a model share characteristics with empirical ones, their credibility increases (Eck and Liu 2004).

Sensitivity analysis addresses whether the parameter values chosen to represent the assumptions of the model affect the outcome of the model. For example, in a model of guardianship, if the number of people necessary to establish guardianship in a potential offender's view is set to one person, will the model results change significantly if the parameter is changed to require two people, or three or four? Sensitivity issues can be tested by varying the parameter values and looking whether the overall results change. Given the large number of parameter values that are typically used in a model, only the parameters which are most likely to affect the results are usually investigated (Gilbert and Troitzsch 2005). Sensitivity to initial parameter settings and agent interaction rules are widely recognized limitations of ABMs that can only be partially mitigated through sensitivity testing (Couclelis 2002).

In sum, the strength of a model increases when rigorous verification is implemented throughout the programming process. It is improved further when validation tests reveal the model-produced findings share stylized characteristics with empirical findings. For example, since crime is clustered, crime events produced by a model should also be clustered. However, matching distributions is not a sufficient criterion for validation since a different model could also produce comparable patterns investigated (Gilbert and Troitzsch 2005). Establishing model credibility is an incremental process that involves multiple comparisons and is not an exact science.

Communication for Replication and Evaluation

One of the most challenging yet important stages of ABM is sharing the model with other researchers. This requires in-depth and complete description of the model which is difficult to accomplish in the space available in the typical journal. A template called the Overview, Design concepts, and Details (ODD) protocol has been developed for communicating models to facilitate replication and evaluation (Grimm and Railsback 2012). The first section of the ODD describes the focus of the model. The next section describes how the model implements ten core design elements including emergence, adaptation, objectives, learning, prediction, sensing, interaction, stochasticity, collectives, and observation. The final section contains all the information necessary to replicate the model. This protocol is becoming more widely used which should facilitate replication and evaluation of ABMs (Grimm and Railsback 2012).

When to Use an ABM

There are several circumstances in which ABMs are an appropriate method to use. When existing theories describe the mechanisms involved in criminal behavior (Eck and Liu 2008a), ABMs can be used to test the mechanisms in silica. For example, rational choice perspective (Clarke and Cornish 1985) describes how offenders make decisions whether to commit a crime based on situational characteristics. Agents can be created that use the logic of rational choice perspective to evaluate a situation. The modeler can examine what combination of situational characteristics translates into a decision to commit a crime.

When there are no data available to test a theory or describe the phenomena of interest (Eck and Liu 2008a), an ABM may be the only methodology available. For example, convergence is a core mechanism for crime events in routine activity theory (Cohen and Felson 1979). Data describing the movements of individuals prior to their involvement in a crime event are rarely available now and that is not likely to change (O’Sullivan 2004). The space-time movements of agents can be collected within an ABM and used to explore explanations for how offenders and victims converge in space-time.

In the same vein, ABMs can be used to explore the decision-making process related to crime. Crime events are the end result of decisions made by both offenders and victims that bring them together at the same place and time. However, data describing individual decisions prior to, during, and after the crime event are not available. Even if the data were collected, the dynamic, nonlinear quality of crime events makes them difficult to study using traditional statistical techniques. Agent-based modeling can accommodate individual agents with unique characteristics and decision-making capability.

When opportunities to conduct field experiments are challenging or even impossible, ABMs can be used as exploratory devices (Groff and Mazzerole 2008), for example, when ethical concerns preclude random assignment of people to treatment and control conditions or when it is not practical to vary the attribute of interest, such as changing the configuration of streets, which would be prohibitively expensive to do in a field trial but is easily done in an ABM. Although ABMs are not inherently experimental, they can be designed to systematically manipulate or randomly allocate a condition. The counterfactual or "null model" is simply the model without the "treatment" or behavior of interest turned on. The outcome from the null model can be compared to systematic manipulations of some aspect of agent behavior or the environment while literally holding all other aspects of the model constant. This provides a level of control nearly impossible to attain in field experiments. "While there is no substitute for field experimentation, simulation may be able to play a significant role in vetting and/or strengthening programs prior to their empirical testing." (Groff and Mazzerole 2008, p. 189).

Along those lines, ABMs can be used to vet potentially expensive or invasive crime-prevention strategies prior to implementation (Groff and Mazzerole 2008). The combination of heterogeneous agents and complete control allows testing of a variety of crime-prevention programs and evaluation of outcomes for minimal cost as compared to field experiments. This makes them an excellent alternative when field experimentation would be prohibitively expensive or ethically difficult to mount (Eck and Liu 2008a). ABMs generally represent a complimentary research method to existing empirical ones.

Modeling the Process and Structure of Crime

ABMs have been used to study a variety of crime types including residential burglary, commercial robbery, street robbery, fraud, heroin use, drug markets, and crime in general (see both Eck and Liu 2008b; Groff and Birks 2008 for an overview). Most persuasively, these examples have illustrated how ABMs are able to model both the process and structure of events. In the case of crime events, ABMs can take into account the process that brings together the actors involved in crime. They can also represent the interactions that occur among those actors plus the decision-making process as the situation unfolds. ABMs can include structure in the sociological sense in terms of social influences on agent behavior. This is analogous to modeling the effects of unemployment, poverty, and education on crime. They can also incorporate the physical structure within which agents interact in terms of, for example, transportation and land use.

Future Directions and Challenges

ABM offers a promising alternative method for exploring how individual/micro-level actions over time produce group/macro-level phenomena. Although ABM has been around for over 30 years (40 years if theoretical/pen-and-paper models are included), significant challenges to using the methodology remain. Firstly, and most practically, the use of ABM as a methodology continues to be hampered by the steep learning curve required. Skills involved in the development of conceptual models for the most part follow standard model-building practice. Of course, modelers do have to learn to think using a bottom-up rather than a top-down paradigm, which can be difficult. But the greater challenge remains the need for programming skills to implement models in current software. Given the lack of experience of most social scientists with computer programming, this represents a significant hurdle to using ABM.

Secondly, validation of ABMs generally is challenging. Many of these challenges are widely recognized such as the need for standardized techniques for model building and for analyzing models and routine replication of models (Gilbert and Troitzsch 2005), as well as the fact that ABMs using different mechanisms can produce similar results.

In addition to these intrinsic challenges, the use of ABMs to examine crime has to depend on outcome data (i.e., crime statistics) with widely recognized shortcomings. Since crime data only reflect a subset of the crime that is committed, comparing the crime patterns produced by an ABM to empirical crime data is, in effect, comparing them to the subset of crime that is both reported to the police and for which the police take a report. Thus, it is difficult to tell whether the ABM's crime pattern is incorrect or whether it actually reflects the "real" distribution (i.e., the full set) of crime committed (Eck and Liu 2008a; Groff and Birks 2008).

Despite these challenges, ABM's use to investigate the potential crime prevention value of situational crime prevention techniques has great potential. The necessary foundation for agent-based modeling to contribute in evaluating crime-prevention strategies is the development of models that can produce realistic crime patterns. ABMs have traditionally emphasized offender behavior but there are other actors who have important roles in whether a crime occurs when the necessary elements converge in space and time (Eck and Liu 2008a). Decision-making by non-offenders such as potential victims, intimate handlers, and place managers are critical to understanding why crime occurs in one situation and not another (see Groff and Birks 2008 for additional suggestions). Given sufficiently robust models of why a particular crime occurs where it does, the possibilities for testing crime-prevention strategies in a relatively low-cost environment like an ABM would be limited only by the imagination of the researchers.

Related Entries

- ▶ Crime Prevention Through Environmental Design
- ▶ Integrating Rational Choice and Other Theories
- ▶ Randomized Experiments in Criminology and Criminal Justice
- ▶ Rational Choice Theory
- ▶ Routine Activities Approach
- ▶ Theories for Situational and Environmental Crime Prevention

Recommended Reading and References

Abdou M, Hamill L, Gilbert N (2012) Introduction to agent-based modelling. In: Heppenstall AJ, Crooks AT, See LM, Batty M (eds) Agent-based models of geographical systems. Springer, London, pp 141–166

An L, Linderman M, Qi J, Shortridge A, Liu J (2005) Exploring complexity in a human-environment system: an agent-based spatial model for multidisciplinary and multiscale integration. Ann Assoc Am Geogr 95(1):54–79

Axelrod R (2007) Advancing the art of simulation in the social sciences. In: Rennard J-P (ed) Handbook of research on nature inspired computing for economy and management. Idea Group, Hershey, pp 90–100

Balci O (1994) Principles of simulation model validation, verification and testing. Technical report #TR-94-24. Virginia Polytechnic Institute and State University, Blacksburg

Birks D, Donkin S, Wellsmith M (2008) Synthesis over analysis: towards an ontology for volume crime simulation. In: Eck JE, Liu L (eds) Artificial crime analysis systems: using computer simulations and geographic information systems. Idea Group, Hershey, pp 160–192

Birks D, Townsley M, Stewart A (2012) Generative explanations of crime: using simulation to test criminological theory. Criminology 50(1):221–254

Brantingham PL, Brantingham PJ (2004) Computer simulation as a tool for environmental criminologists. Secur J 17(1):21–30

Brown DG, Riolo R, Robinson DT, North M, Rand W (2005) Spatial process and data models: toward integration of agent-based models and GIS. J Geogr Syst 7:25–47

Clarke RV, Cornish DB (1985) Modeling offender's decisions: a framework for research and policy. In: Tonry M, Morris N (eds) Crime and justice: an annual review of research, vol 6. University of Chicago Press, Chicago, pp 23–42

Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44:588–608

Couclelis H (2002) Modelling frameworks, paradigms, and approaches. In: Clarke KC, Parks EB, Crane MP (eds) Geographic information systems and environmental modelling. Prentice Hall, London, pp 36–50

Crooks AT, Heppenstall AJ (2012) Introduction to agent-based modelling. In: Heppenstall AJ, Crooks AT, See LM, Batty M (eds) Agent-based models of geographical systems. Springer, London, pp 85–108

Dowling D (1999) Experimenting on theories. Sci Context 12(2):261–273

Eck JE, Liu L (2004) Routine activity theory in a RA/CA crime simulation. Paper presented at the American Society of Criminology, Nashville

Eck JE, Liu L (2008a) Varieties of artificial crime analysis: purpose, structure, and evidence in crime simulations. In: Eck JE, Liu L (eds) Artificial crime analysis systems: using computer simulations and geographic information systems. IGI Global, Hershey, pp 413–432

Eck JE, Liu L (eds) (2008b) Artificial crime analysis systems: using computer simulations and geographic information systems. IGI Global, Hershey

Epstein JM, Axtell R (1996) Growing artificial societies. Brookings Institution Press, Washington, DC

Gilbert N, Troitzsch KG (2005) Simulation for the social scientist, 2nd edn. Open University Press, Buckingham

Grimm V, Railsback SF (2012) Designing, formulating, and communicating agent-based models. In: Heppenstall AJ, Crooks AT, See LM, Batty M (eds) Agent-based models of geographical systems. Springer, London, pp 361–377

Groff ER (2007) Simulation for theory testing and experimentation: an example using routine activity theory and street robbery. J Quant Criminol 23(2):75–103

Groff ER (2008) Spatio-temporal aspects of routine activities and the distribution of street robbery. In: Liu L, Eck JE (eds) Artificial crime analysis systems: using computer simulations and geographic information systems. Idea Group, Hershey, pp 226–251

Groff ER, Birks D (2008) Simulating crime prevention strategies: a look at the possibilities. Polic: J Policy Pract 1:1–10

Groff ER, Mazzerole L (2008) Simulated experiments and their potential role in criminology and criminal justice. J Exp Criminol 4(3):187–193

Johnston KM (ed) (2012) Agent analyst: agent-based modeling in ArcGIS. Esri, Redlands

Law AM, Kelton D (1991) Simulation modelling and analysis. McGraw-Hill, New York

O'Sullivan D (2004) Too much of the wrong kind of data: implications for the practice of micro-scale modeling. In: Goodchild MF, Janelle DG (eds) Spatially integrated social science. Oxford University Press, New York, pp 95–107

O'Sullivan D, Haklay M (2000) Agent-based models and individualism: is the world agent-based? Environ Plan A 32(8):1409–1425

Ostrom TM (1988) Computer simulation: the third symbol system. J Exp Psychol 24:381–392

van Baal P (2004) Computer simulations of criminal deterrence: from public policy to local interaction to individual behavior. BJU Boom Juridische uitgevers, Netherlands

Agent-Based Modelling

► Agent-Based Assessments of Criminological Theory

Agent-Based Models to Predict Crime at Places

Nick Malleson and Andrew Evans
School of Geography, University of Leeds,
Leeds, UK

Overview

Crime occurrences are driven by a complicated mix of distinct influences, including those of the environment, the surrounding social context, and personal behavior/psychology of the people who could influence a crime event. Agent-based modelling is a methodology used in computer simulation that concentrates on individual-level behaviors and is ideally suited to modelling crime. This is particularly true of crimes such as burglary or street crime, which are heavily influenced by environmental factors and by the behavior of individual people. In an agent-based crime model, virtual "agents" are placed in an environment that allows them to travel through space and time, behaving as they would do in the real world. This entry will discuss why the crime system is such an ideal candidate for agent-based models and will review a number of crime models that have recently arisen.

Introduction

Individual crime occurrences are caused by a complicated mix of factors, including – but not limited to – the surrounding physical environment, the local social context, the presence or absence of crime reduction programs, and the behavior, psychology, and interactions between those people who might be able to influence a crime event (including victims, offenders, and guardians). Traditional mathematical models of crime can face difficulties with modelling systems which are inherently nonlinear (Eck and Liu 2008), and therefore, methods which are better suited to capturing the dynamics of complex, nonlinear systems are becoming popular.

Agent-based modelling (ABM) is a methodology used in computer simulation that concentrates on individual-level behaviors and is ideally suited to modelling crime. This is particularly true of crimes, such as burglary or street crime, which are heavily influenced by environmental factors and by the behavior of individual people. In an agent-based crime model, virtual "agents" are placed in an environment that allows them to travel through space and time, behaving as they would do in the real world. The environment can be as simple or detailed as the researcher chooses so the methodology can thus be used both to explore criminology theory in the absence of confounding factors or to make realistic predictions in a real-world virtual environment.

This entry will review the technique of agent-based modelling and discuss what advantages it offers to the field of environmental criminology. It begins an overview of the "crime system" and discusses what makes it such an ideal candidate for agent-based modelling, followed in Section 3 by a discussion of the technique itself. Section 4 continues by reviewing a small number of the most recent and advanced agent-based crime models. Finally Sections 5 and 6 review some of the controversies in the literature and propose open questions for future research.

Background: Crime Is a Complex System

Social systems, including that of crime, belong to a class of system known as "complex systems." These are systems that consist of large numbers of diverse components, varied and complex interactions between components, and emergent properties – in this case city – or neighborhood-wide crime rates that cannot be attributed to any individual part of the system. Coupled with this inherent system complexity are "human" factors, such as a complex psychology, that further compound the rules that drive the individuals (Bonabeau 2002). In addition, the crime system is made even more complicated because as crime at places (Eck and Weisburd 1995) and situational crime prevention (Clarke 1995) research have demonstrated, the geography of the system itself introduces additional complexity. However, understanding the processes and drivers that characterize the crime system is the key to crime prevention and policy development.

Acquisitive crimes, such as burglary and street robbery, are important exemplars of the manner in which environmental, social, and individual drivers combine to create a set of complex processes. Because occurrences are dependent on a number of interacting factors (motivation, opportunity, recidivism, etc.), it is difficult to predict crime hotspots in advance, and therefore, many reduction schemes react to a crime hotspot once it has formed. Ultimately, acquisitive crimes are committed by individuals in a local environment and a particular time. Modern criminology highlights the importance of "micro-places" that act as the environment for a specific crime (Eck and Weisburd 1995), and research at scales larger than houses and streets hides key crime dynamics. The same scale issues apply to the individuals involved in a crime event (be they perpetrators, victims, or bystanders) whereby aggregate treatment of individual people is likely to miss the key dynamics associated with individuals and their daily lives. As Brantingham and Brantingham (1993) predicted:

> Potentially, the most productive model in environmental criminology is one that places both the actual criminal events at a specific site, situation and time and the individual committing the crime while in a specific motivational state on (or in) an environmental backcloth, that may itself be mostly stable, regular and predictable or may instead be irregular, rapidly changing and unpredictable.

With this in mind, agent-based modelling is a technique that is being shown to hold considerable promise as it represents a shift away from aggregate models towards those that work at the level of the individual. An agent-based computer model is comprised of autonomous entities called "agents," who have the ability to make decisions and interact with each other and their environment. As the model iterates, each agent has the ability to assess its circumstances and, based on a set of probabilistic rules or more advanced decision-making algorithm, make an informed/educated decision about its future course of action. Through this mechanism, it is possible to incorporate realistic human behavior into computational models. With respect to modelling crime, an agent-based model can be built to directly simulate the behavior of offenders/victims/guardians as they travel around their environment on typical routine activities and predict, as a result of these individual behavior patterns, when a crime occurrence is likely to occur. Building a model in this manner – *from the bottom up* – is a much more natural way of describing a complex system than by formulating rules to drive the system from an aggregate level

(Bonabeau 2002). Section 3 will now discuss the technique of agent-based modelling in more detail.

A Description of Agent-Based Modelling

The range of general agent-based theory is extensive, and so this entry will not attempt to provide a full account of all concepts, practices, and applications. For this, the reader is directed to Wooldridge (2009). Instead, it will provide a brief introduction to the methodology and focus on its application to crime. Described as a “breakthrough in computational modelling in the social sciences” (Gilbert and Terna 2000, p. 60) and “one of the most exciting practical developments in modelling since the invention of the relational database” (Macal and North 2005, p. 2), ABM is a reasonably new method of modelling systems. An agent-based model is comprised of virtual “agents” who are able to behave autonomously (i.e., without a central controller). They exist in a virtual environment which is often spatial, they can navigate around their environment, and they are able to make decisions about what they would like to do in a given situation. This approach is particularly relevant to criminology because using ABM, it becomes possible to use models to carry out experiments that would be impossible or unethical to perform otherwise.

There are many definitions of the term “agent” but, from a crime modelling perspective, the following are consistently applied:

- Autonomy: an agent should be free to control its own state, interact with other agents and its environment, and make decisions without direct control from some central source. This seems to be an ideal mechanism for modelling people, including offenders, victims, or other necessary individuals (guardians, managers, passersby, etc.).
- Heterogeneity: agents need not be identical. Offender agents can thus be created to reflect the variety of different offending behaviors that have been exhibited, allowing for the incorporation of qualitatively obtained data and theories.
- Reactivity: agents should be able to respond to changes in their environment, and the response should be proactive, indicating goal-directed behavior (Wooldridge 2009). This is particularly useful for a crime model because the environment will change as a result of crime which will in turn influence the future behavior of the agents.
- Bounded rationality: particularly with modelling in the social sciences, it is important that agents do not always act perfectly rationally. Agents can be programmed with “bounded” rationality by limiting their knowledge of the world so that choices are not always perfectly optimal (Castle and Crooks 2006).

Perhaps the most significant advantage of ABM is the “natural description” of a system which it provides. Complex systems, whose behavior is characterized by the behavior/interactions of its individual components, cannot usually be described by mathematical equations. Although mathematics provides a good basis for describing unexplained phenomena in the natural sciences, this experience is not echoed in the social sciences (Moss and Edmonds 2005). Often simplified assumptions are required if mathematical models become too complicated and these assumptions are often implausible or reduce the realism of nonlinear systems (Evans 2011). To understand geographical human systems, it is necessary to understand the reasoning behind individual decisions and modelling individuals directly is more natural than trying to build aggregate equations to control them (Bonabeau 2002). For example, with acquisitive types of crime, an individual’s cognitive understanding of a local area can be as important as the physical characteristics of the area, as offenders commonly commit crimes within their routine activity spaces. These cognitive representations would be very difficult to incorporate into models which do not characterize individuals directly, whereas in an agent-based model they can be built directly up as the virtual agents navigate around their environment. The same can be said of environment: it has been shown that spatial aggregation hides important patterns in crime (Andresen and Malleson 2011) and ABM is

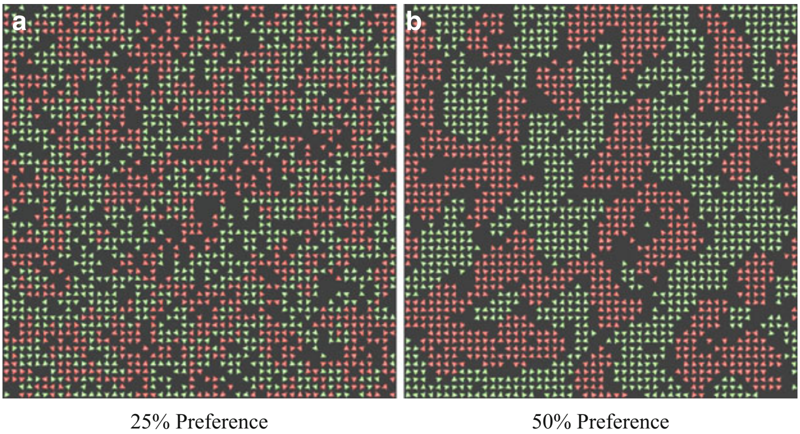

Agent-Based Models to Predict Crime at Places, Fig. 1 An example of an agent-based model of segregation, based on Schelling (1969). Figures show the percentage of similar racial type that each household unit wants to live next to

ideally suited to modelling at resolutions of the individual street or household.

Arguably the first published agent-based model in the social sciences was Schelling's hand-developed model of residential stability (Schelling 1969). Although extremely simple, the model goes a long way to illustrate how useful ABM can be for modelling social systems. The model consists of a one-dimensional environment (a line of spaces, some empty) populated by households that belong to one of two types of racial group. A global parameter dictates the percentage of the same group that each household wants to live next to. Households are able to move to an empty cell if they are unsatisfied (i.e., they live near to too many of the opposite racial type). The novel research finding was that, even with a relatively low preference for the same racial type, the environment can become highly segregated. It is relatively simple to develop a two-dimensional version of this now, and such a model, which is included as a demonstrator in the agent-based modelling software "NetLogo," is illustrated in Fig. 1. Here a preference to live next to only 50 % of the same racial group leads to clear segregation. Methodologically the Schelling results are interesting because they would not have been predicted by examining the rules of the individual in isolation, and from a practical position, it provides an insight into the dynamics of human residential segregation, suggesting that extreme segregation can emerge from individuals who actually have low preferences for segregation.

The finding that segregation arises even though no individual wants to live in a completely homogeneous area relates to the concept of *emergence*, exploring which is a key advantage of ABM. An emergent phenomenon is one that comes about as a natural (and often unintended) consequence of the behaviors and interactions of a system's constituent components. A city-wide crime rate is an example of an emergent phenomena and one which cannot be attributed to any part of the system itself – there is no single person who is responsible for, or even attempts to generate, the observed crime patterns. Rather than attempting to find global rules that determine aggregate crime rates, the agent-based crime modeller describes the individual components of the system and then tries to "grow" the observed crime patterns from the "bottom up." This also has the advantage of being a much more naturalistic way of describing a system. Social networks, human psychology, and detailed physical environments are essential parts of the crime system, and these are relatively simple to incorporate by using ABM. A related advantage, which is particularly relevant in environmental criminology, is that with ABM it is possible to treat offenders in a similar manner to non-offenders and explore the effects that noncriminal activities will have on crime. In this manner, the "natural variety" of cities becomes part of the model, rather than being smoothed out by aggregate methods (Brantingham and Brantingham 2004).

There are, of course, certain drawbacks to ABM which must be addressed. The advantage of being able to describe the behavior of individual people directly is tempered by the difficulty in modelling the “soft factors” exhibited by humans – such as seemingly irrational behaviors and complex psychology (Bonabeau 2002) – which can be very difficult to actually implement in a computer model. These factors must be defined explicitly in models which work at the microlevel, leading to a strong commitment to minimal behavioral complexity (O’Sullivan and Haklay 2000). This is compounded in crime modelling research by problems with the availability and accuracy of crime data to build up an accurate picture of offending behavior. Although there is considerable qualitative information available, quantitative data is sparse, and even when data are available, it is unclear whether or not they are an accurate reflection of true offending patterns because of potential reporting biases. For more information the interested reader could refer to Chainey and Ratcliffe (2005) for a comprehensive assessment of the issues surrounding the use of crime data.

There are also major difficulties that relate to the implementation of the model. Firstly, although there are numerous ABM tools that can assist with the development of models (NetLogo is a good starting point for those new to the technique), it is likely that a researcher will need a relatively high level of computer programming experience. Agent-based models are also extremely processor and storage heavy, not least because most are run multiple times with varying parameters to give probabilistic results. This means that the most advanced models often need to be distributed across large collections of computers to obtain the required processing and storage power. Such models present considerable debugging issues, not least because small pieces of computer code can form integral parts of hundreds of agents, with small errors in the logic of the code having huge effects on the outcome of the model.

Nevertheless, ABM is becoming an increasingly popular tool to aid crime analysis. The following section summarizes some of the more recent and advanced approaches.

State of the Art

The last decade has seen a gradual rise in agent-based crime simulation work. The earliest models are generally abstract and take place either in geographical social spaces or on abstract plains (much like the segregation model outlined in section “A Description of Agent-Based Modelling”), whereas more recent models include realistic representations of the physical environment and comprehensive agent decision-making. Most models consider some subset of the core elements of the crime system, such as the physical environment, the social context, victims, offender motives, and offender behaviors. However, the system is clearly extremely complicated, and it should come as no surprise to find out that most models concentrate on one particular aspect of the system. To provide an overview of the current research, this entry will briefly summarize a small number of applications which have very different foci. For a much more comprehensive review, the reader is directed to the book entitled *Artificial Crime Analysis Systems* (Liu and Eck 2008), a special issue of the *Journal of Experimental Criminology* entitled “Simulated experiments in criminology and criminal justice” (Groff and Mazerolle 2008) or the review paper by Malleson et al. (2011).

Offender Behavior

One of the obstacles with agent-based crime modelling is that it becomes necessary to explicitly define how agents in the model will behave. However, modelling human behavior is an extremely challenging endeavor. Fortunately there are numerous cognitive architectures that have been designed to model human behavior, and these can be used in crime models. Malleson et al. (2010) provide an example of such a model applied to residential burglary. The authors use the PECS (physical conditions, emotional states, cognitive capabilities, and social status) model of human behavior which controls agents by comparing the sizes of different *motives*. At any given time, the motive which is the strongest determines what the agent’s current behavior will be such that the agent tries to satisfy that motive. Some motives can be very simple to satisfy (such as travelling home to sleep), but others might require considerable planning and reevaluation before they can be accomplished (such as making money through the commission of burglary). By varying the ways that different motives affect the agents' behavior, it is possible to create different types of offender agents to reflect current criminological thinking (e.g., the difference between a "professional" burglar and an "opportunist"). By placing the virtual offenders in an environment that closely reflects that of the real world, the authors are able to explore the potential real-world crime patterns that might emerge under different scenario conditions.

Street Networks

An example of research which takes advantage of ABM's ability to incorporate a highly detailed physical environment is that of Groff (2007). Due to the lack of dynamic, individual-level interactions, the authors note that previous studies had failed to effectively test routine activity theory. The model by Groff addresses this through the incorporation of accurate street-level data in order to test the applicability of routine activity theory to street robbery. The model contains two types of agent: citizens (offenders, victims, and guardians) and police. The citizen agents are randomly assigned a particular home location, and in the model's most advanced form, they spend time away from home by visiting randomly assigned work and activity nodes following predefined routes. The offenders' decision to offend is stochastic and based on levels of guardianship and the wealth of the potential target at their current location. The model found that the number of street robberies increased with the amount of time spent away from home because citizens had the chance to meet more potential offenders. Interestingly, some street intersections exhibited significant clusters of events even though the travel patterns of the agents were random. This provides an insight into how the urban configuration of streets can influence locations of street crime with direct reference to the roads in a real city.

Social Networks and Cohesion

Plainly the social structure of areas and criminals will play a big part in developing larger patterns of offending. Hayslett-McCall et al. (2008) include a number of social-cohesion and guardianship indicators in their household burglary model. Homes are given attractiveness values based on variables including the socioeconomic status of the area and the degree of guardianship, as well as the ethnicity and income levels of the inhabitants – with agents preferring areas that are similar to their own. In terms of social interaction, Dray et al. (2008) include the relationships between local actors, including law enforcement agencies, social services, and drug users and dealers in an ABM of the Melbourne drug market. The model also includes larger-scale variables such as the strength of the national drug supply chains. Using this model, they showed that small-scale interventions were generally more successful at disrupting drug markets than national initiatives.

Abstract Theory

The models introduced thus far both utilize realistic virtual environments in order to place crime in an accurate environmental context. However, accurately predicting spatiotemporal crime patterns does not need to be the aim of the research – indeed, as the following section will discuss, a realistic virtual environment is seen by some as a drawback. An advantage of ABM is that the model can act as a "virtual laboratory," allowing researchers to explore the individual-level dynamics that emerge as a result of different crime theories in an environment that is free from the usual complexity of the real world. Brantingham and Brantingham (2004) have developed a model of crime in which they utilize an abstract state machine to provide a precise logical/mathematical foundation to an agent-based model. In the model, agents can move through time and space, interacting with each other and the environment. They are also able to learn, and they have "preferences" which translate to forms of behavior. Along with experimenting with criminology theory, the resulting simulation can be used as an interdisciplinary tool to assist criminologists in investigating the dynamics of urban crime.

Controversies in the Literature

In general, agent-based models tend to either be pessimistically abstract or optimistically realistic. The Schelling segregation model and the model by Brantingham and Brantingham (2004) both fall at the more abstract end of the scale. With these types of models, the difficulty of capturing the total behavior of such large, open, and nonlinear systems is accepted, and modellers concentrate on building simple models based on limited subsets of the rules and environmental conditions that may be operating. These models are then treated as thought experiments and can be manipulated to see what effect varying our broad ideas about the systems has on the patterns it generates. In the field of crime modelling, Elffers and van Baal (2008) advocate the notion that such simple models can be a powerful explanatory force, particularly when the aim of the research is to explore theory rather than make accurate predictions. One always has to be aware, however, that patterns may match models even though explanations differ (the identifiability problem), and it is often difficult to verify that human systems would respond as these simple models do.

The second point of view holds that it is possible to accurately model the world with a finite ruleset and list of environmental variables, but that the more realism the model encompasses, the more constrained to the real world the model will be. In some cases, such as the models by Groff (2007) and Malleson et al. (2010), geographical realism is added by combining the model with a geographical information system (GIS). The hope with more realistic models is that the multiple patterns and data streams used to constrain the model will resolve the identifiability problem and make comparisons with reality more reliable. However, it is usual to calibrate such models using real data, and the large numbers of variables involved may mean that the models are over-flexible, essentially fitting any data – again raising the issue of identifiability. Rigorous retesting with new data is often difficult for models that study crime; even though more accurate data is being made available to researchers all the time, it is still at a premium and it is usually unclear how well the data reflects real-world crime patterns in the first place. In general, models are best built up gradually, verifying that simple behaviors with simple elements of the system work as expected to start with and then, if needed, build up the complexity to something that more closely matches reality in order to make more reality-aiming predictions. However, as Edmonds and Moss (2005) note, a phrase such as “for the sake of simplicity” is not well founded, and simplicity should only be a target if this is justified by the underlying system. The crime system is not, usually, simple.

Open Questions

Two significant and related directions seem of immediate future significance in the ABM of crime (Malleson et al. 2011). The first is the more detailed simulation of social networks, particularly within criminal organizations. There has been some work in this area; however the potential for spatially enabled social network modelling is increasing as law enforcement agencies gain access to mobile phone and internet records. The second direction is more contentious – the modelling of specific individuals within ABM. To our knowledge there are not, as yet, individual-level models that predict the actions of real individuals. However, there is an increasing use of real crime data to seed models of abstract individuals, and increasing amounts of personal information about all citizens, potential victims, and potential offenders are stored electronically on a daily basis. It seems likely, therefore, that the modelling community will need to consider the ethical issues of the Minority Report option within the very near future: is it appropriate to model real individuals in a predictive sense, and what use should such predictions be put to once made?

Related Entries

- ▶ Agent-Based Modeling for Understanding Patterns of Crime
- ▶ Spatial Models and Network Analysis

Recommended Reading and References

Andresen M, Malleson N (2011) Testing the stability of crime patterns: implications for theory and policy. J Res Crime Del 48:58–82

Bonabeau E (2002) Agent-based modeling: methods and techniques for simulating human systems. Proc Natl Acad Sci USA 99(90003):7280–7287

Brantingham PL, Brantingham PJ (1993) Nodes, paths and edges: considerations on the complexity of crime and the physical environment. J Environ Psychol 13(1):3–28

Brantingham P, Brantingham PJ (2004) Computer simulation as a tool for environmental criminologists. Secur J 17:21–30

Castle C, Crooks A (2006) Principles and concepts of agent-based modelling for developing geospatial simulations, UCL working paper series. Centre for Advanced Spatial Analysis, London

Chainey S, Ratcliffe J (eds) (2005) GIS and crime mapping, 1st edn. Wiley, Chichester

Clarke RV (1995) Situational crime prevention. Crime Justice 19:91–150

Dray A, Mazerolle L, Perez P, Ritter A (2008) Policing Australia's 'heroin drought': using an agent-based model to simulate alternative outcomes. J Exp Criminol 4(3):267–287

Eck JE, Liu L (2008) Contrasting simulated and empirical experiments in crime prevention. J Exp Criminol 4(3):195–213

Eck JE, Weisburd D (1995) Crime and place, volume 4 of crime prevention studies. Criminal Justice Press, Monsey

Edmonds B, Moss S (2005) From KISS to KIDS: an 'anti-simplistic' modelling approach. In: Davidsson P, Logan B, Takadama K (eds) Multi agent based simulation. Springer, Heidelberg, pp 130–144

Elffers H, van Baal P (2008) Realistic spatial backcloth is not that important in agent based simulation research: an illustration from simulating perceptual deterrence. In: Eck JE, Liu L (eds) Artificial Crime Analysis Systems: Using Computer Simulations and Geographic Information Systems. Information Science Reference, Hershey

Evans A (2011) Uncertainty and error. In: Heppenstall AJ, Crooks A, Batty M (eds) Agent-based models for geographical systems. Springer, Dordrecht

Gilbert N, Terna P (2000) How to build and use agent-based models in social science. Mind Society 1(1):57–72

Groff ER (2007) Simulation for theory testing and experimentation: an example using routine activity theory and street robbery. J Quant Criminol 23(2): 75–103

Groff E, Mazerolle L (2008) Simulated experiments and their potential role in criminology and criminal justice. J Exp Criminol 4(Special Issue):187–193

Hayslett-McCall K, Qui F, Curtin KM, Chastain B, Schubert J, Carver V (2008) The simulation of the journey to residential burglary. In: Liu L, Eck J (eds) Artificial crime analysis systems: using computer simulations and geographic information systems. Information Science Reference, Hershey

Liu L, Eck J (2008) Artificial crime analysis systems: using computer simulations and geographic information systems. Information Science Reference, Hershey

Macal CM, North MJ (2005) Tutorial on agent-based modelling and simulation. In: Kuhl ME, Steiger NM, Armstrong FB, Joines JA (eds) Proceedings of the 2005 winter simulation conference. Institute of Electrical and Electronics Engineers, Piscataway, pp 2–15

Malleson N, See L, Evans A, Heppenstall A (2010) Implementing comprehensive offender behaviour in a realistic agent-based model of burglary. SIMULATION. Published online before print: http://sim.sagepub.com/content/early/2010/10/20/0037549710384124

Malleson N, Evans A, Heppenstall A, See L (2011) Crime from the ground-up: agent-based models of burglary. Geogr Compass (forthcoming)

Moss S, Edmonds B (2005) Towards good social science. J Artif Soc Soc Simulat 8

O'Sullivan D, Haklay M (2000) Agent-based models and individualism: is the world agent-based? Environ Plann A 32(8):1409–1425

Schelling T (1969) Models of segregation. Am Econ Rev 59:488–493

Wooldridge MJ (2009) An introduction to multiagent systems. Wiley, Chichester

Aging Correctional Populations

Brie Williams, Cyrus Ahalt and Louise Aronson
Division of Geriatrics, University of California
San Francisco, San Francisco, CA, USA

Overview

The American prison population is rapidly aging. Yet relatively little is known about the unique challenges facing older adults in the criminal justice system. This knowledge gap has significant implications for the understanding of the health and health-care needs of a growing, medically vulnerable population. This entry describes the demographics of aging in corrections, focusing on the dramatic growth in the geriatric prison population that has taken place since the early 1990s; introduces the comprehensive, integrative framework of geriatric medicine, the medical specialty focused on the care of older adults; and reviews what is known about the health of older prisoners through the lens of geriatric medicine, examining biomedical, psychosocial, and environmental contributors to health and physical function. Age-related risk factors that may result in obstacles to successful reentry for older adults following incarceration are also discussed throughout. This entry concludes with a proposed agenda for future policy and research based on the geriatrics framework.

Demographics: The Aging of the Criminal Justice System

Over the past several decades, a dramatic increase in the overall population of US prisoners has included staggering growth in the number of incarcerated older adults. At its zenith in 2008, the criminal justice system incarcerated 1 in every 100 US adults, and 1 in 31 was a ward of the criminal justice system in some way, via jail, prison, parole, or probation (Pew Center on the States 2009). In the two decades from 1990 to 2010, there was a 100 % increase in total state and federal prisoners. During the same period, the number of prisoners 55 years of age or older grew by over 300 % to nearly one in ten of all prisoners. Although the total prison population decreased 0.5 % from 2009 to 2010, the number of prisoners aged 55 or older increased by 57.3 % (Human Rights Watch 2012). Absent significant policy change this exponential growth in the number of older prisoners is likely to continue as more than 450,000 prisoners, nearly one in three, are between the ages of 40 and 54 (Bureau of Justice Statistic [BJS], 2011).

There are many reasons for the disproportionate growth in older adult incarcerations: increased arrests and sentencing of older adults, increased use of life sentences, the discontinuation of discretionary parole, tougher drug laws, and mandatory sentencing practices (Aday 2003; Williams and Abraldes 2007; Human Rights Watch 2012). As a result, older prisoners represent a range of criminal justice dispositions. According to a 2012 report from Human Rights Watch, approximately 45 % of older state prisoners (in this case, prisoners aged 51 or older) are serving sentences of less than 10 years, 35 % are serving sentences of more than 10 years but not a life sentence, and 20 % are serving a sentence of life or greater. While 65 % of older prisoners in 2009 were serving sentences for violent crimes, 35 % were in prison for property, drug, or other nonviolent crimes. In 26 states, so-called "Three Strikes" laws, which confer extended sentences for some repeat offenders, likely contribute to the increasing number of older prisoners serving long sentences for nonviolent crimes. In California, for example, 43,500 prisoners were serving enhanced sentences under Three Strikes Laws in 2010, prompting the state auditor to release a report pointing to $19.2 billion in additional anticipated costs associated with these prisoners as a significant source of strain on the state budget (California State Auditor 2010).

Indeed, the aging correctional population is increasingly cited as one of the main contributors to the rising cost of incarceration nationwide, primarily due to the greater health-care usage of older prisoners. In 1976, the Supreme Court's ruling in *Estelle v. Gamble* guaranteed prisoners' rights to "community standard" health care. Failure to provide such care, the court ruled, would be considered a violation the 8th Amendment's prohibition on cruel and unusual punishment. Since the *Estelle v. Gamble* ruling, however, a host of litigation targeting inadequate prison health care suggests that many prison systems are struggling to meet the appropriate standard of care (Wool 2007). The majority of states are also experiencing significant budget constraints alongside widespread prison overcrowding; as of 2005, 239 prisons (13 %) across 32 states were under court order to improve conditions (Pew Center on the States 2009; Damberg et al. 2011). It is in this challenging context that the number of older prisoners, many who would be considered medically vulnerable even outside the prison setting, continues to grow at a faster rate than any other demographic groups.

The Field of Geriatrics

Geriatrics is the field of medicine specializing in the care of older adults, defined in the community as people aged 65 years and older. The goal of geriatric medicine is to increase the health, independence, and quality of life of older adults by providing high-quality, patient-centered, interdisciplinary care. At the same time, the field recognizes that life circumstances may create a mismatch between physiologic and chronologic age. Thus, while the age of 65 was an appropriate demarcation between middle and older age at the inception of Medicare and Social Security in the 1950s, subsequent increases in lifespan and decreases in years of disability for certain populations (e.g., whites and upper income groups) have meant that some Americans remain physiologically middle aged well into their 70s. Conversely, a condition known as *accelerated aging* is often evident in vulnerable aging adults and may apply to groups like the homeless, refugees, the very poor, prisoners, and individuals with profound, lifelong, or chronic mental or physical illness (Aday 2003; Wahidin 2006). Because many incarcerated adults fall into several of categories of increased vulnerability, most criminal justice systems consider prisoners to be "older," or geriatric, by age 50 or 55 (Aday 2003; Baillargeon et al. 2000).

Most medical specialties in the USA approach health and illness using the biomedical model as their core theoretical paradigm, focusing on anatomy, physiology, disease, and organ systems. Geriatrics, by contrast, is based on the integrative biopsychosocial model. The biopsychosocial model recognizes many aspects of a person's well-being – biological, psychological-behavioral, and sociocultural – as integral to their health. Geriatric medicine takes this approach one step further by placing the biopsychosocial paradigm within a person's larger environmental context – in this case, prison – and by making explicit associations between functional status, goals of care, and health in the context of one's environment. This integrative conceptual framework accounts for the many interrelated influences that affect the health and functional status of older adults. In turn, the health and functional status of older prisoners is best understood by considering the relevant biomedical, psychological, social, and environmental factors that underlie health outcomes for the aging correctional population.

Biomedical Factors

Older adults in America have higher rates than younger adults of nearly all chronic medical conditions including hypertension, arthritis, heart disease, chronic obstructive pulmonary disease (COPD), and cancer. Indeed, a strong association between age and disease burden in general is well established, though some notable exceptions, like HIV and hepatitis B and C, are more common in younger adults. Additionally, for most medical conditions, and many disabilities, prisoners have higher rates than non-prisoners (Aday 2003; Baillargeon et al. 2000; Williams et al. 2006). Studies further suggest that older prisoners have a higher burden of medical conditions than their younger counterparts (Binswanger et al. 2009; Baillargeon et al. 2004).

Despite robust evidence of increased disease burden, many questions regarding the health of aging prisoners remain unanswered. Most relevant studies are small and limited to one geographic area. It should also be noted that the health of older prisoners serving long sentences may differ from the health of prisoners who were first arrested or incarcerated as older adults as the concept of incarceration as a potential accelerator of poor health is itself controversial. For example, some authors note that older prisoners often experience poor health and inadequate care throughout their lives. As a result, they suggest, incarceration may lead to improvements in health due to access to meals, shelter, and health care, even though prison-based care is diminished in quality relative to the standard of care generally available outside of prison (Aday 2003).

This section considers the biomedical aspects of health for older prisoners by examining, in turn, the current base of knowledge on chronic medical illnesses, geriatric syndromes, infectious diseases, death, and health services and utilization patterns.

Chronic Medical Illness

In the prison setting, as on the outside, the prevalence of chronic medical conditions increases dramatically with age. Data from the Texas Department of Criminal Justice (TDCJ) shows that 85 % of prisoners aged 50 or older have one or more chronic medical condition and 61 % have two or more. By comparison, 37 % of prisoners age 30–49, and just 16 % of prisoners under age 30, have more than one such condition (Baillargeon et al. 2000). Moreover, three of the most common chronic medical conditions found in the US population overall – hypertension, arthritis, and diabetes – appeared twice as frequently in the TDCJ prisoner population aged 50 or older. This trend of increasing illness burden continues as older prisoners are divided into subgroups by age. A study in Iowa compared prisoners in their 50s to prisoners over age 60 and found increases in not just hypertension, arthritis, and diabetes but also in ulcers, prostate disease, myocardial infarction, emphysema, and stroke (Colsher et al. 1992). Among female prisoners, another growing demographic, a study of 120 women aged 55 and older in California prisons found similarly high rates of chronic medical conditions: hypertension (64 %), diabetes (16 %), COPD/asthma (43 %), arthritis (73 %), stroke (13 %), depression (36 %), cancer (12 %), and heart disease or heart attack (31 %). In this study, 33 % of self-reporting respondents said they had three or more of these conditions (Williams et al. 2006). Though few nationally representative studies of older prisoner health have been conducted, one such study using BJS data from 2004 found that adults aged 50 or older in prisons and jails are far more likely than their age-matched counterparts in the community to suffer from chronic illnesses including hypertension, diabetes, asthma, arthritis, and liver disease (Binswanger et al. 2009).

Geriatric Syndromes

Geriatric syndromes are defined as conditions that have multifactorial etiologies, significant morbidity, adverse effects on quality of life, and are more common in older adults (Inouye et al. 2007). The term is often used for clinical conditions such as frailty, falls, dementia, incontinence, and sensory impairment that do not fit easily in traditional disease categories. A person may develop such a syndrome in myriad ways, and treatment often requires a multifocal intervention to minimize the impact of the syndrome on the person's health, functional status, and well-being.

In the community, geriatric syndromes have been shown to be more important than medical conditions in predicting older adults' quality of life, morbidity, mortality, ability to remain independent, and health-care utilization (Landefeld et al. 2004). For this reason, geriatricians and other health-care providers who specialize in the care of older adults focus as much on assessing and addressing geriatric syndromes as on the diagnosis and management of medical conditions like heart disease or diabetes. In prison, geriatric syndromes are similarly important, afflicting many older prisoners and increasing their risk for adverse events (Aday 2003; Hill et al. 2006; Williams et al. 2006).

Functional Status

Often, geriatric syndromes have an impact on functional status, which is defined as a person's ability to be independent in Activities of Daily Living (ADLs) which include bathing, dressing, eating, transferring, and toileting. Dependence in these and higher-level or *instrumental* ADLs ("IADLs," e.g., managing medications and finances, transportation, or shopping) increases with advancing age and is associated with higher health-care costs, further decline, and morbidity (Covinsky et al. 1997). One study of prisoners found that 20 % of men over age 50 were dependent in instrumental activities and 11 % required assistance in some ADLs (Colsher et al. 1992). Such figures, however, may underestimate the true effect of limited functional status on health and well-being for older adults living in the unique prison environment. In a cross-sectional study of women prisoners in California, for example, 16 % needed help with one or more ADL, but 69 % reported great difficulty in performing at least one Prison Activity of Daily Living (ADL-P) (Williams et al. 2006). In this case, ADL-P included dropping to the floor for alarms, standing for head count, getting to the dining hall for meals, hearing orders from staff, and climbing onto and off of the top bunk. Not surprisingly, prisoners with worse functional status reported more chronic medical conditions, lower self-rated health, and higher rates of falls and depression (Williams et al. 2006). Using the ADL-P to measure functional independence in the prison environment, the study's authors showed that many older prisoners who would not be disabled in the community are disabled in prison because of the unique physical demands of the prison environment (Williams et al. 2006).

Falls

In the USA, approximately one third of adults aged 65 or older fall each year, and falls are a leading cause of injury and injury-related deaths in older adults. Many older people who fall develop a fear of falling that may limit their activities, sometimes resulting in social isolation, depression, increased fall risk, and a further decline in functional status (Landefeld et al. 2004). In the California study, 55 % of prisoners over age 55 reported a fall in the previous year (Williams et al. 2006). In addition to accelerated aging and increased morbidity, prisoners must also negotiate prison settings often marked by poor lighting, a lack of handrails, crowds of younger prisoners, and strenuous work assignments (Hill et al. 2006).

Dementia

Dementia is defined as the impairment of memory and at least one other cognitive area interfering with daily function. Dementia prevalence doubles every 5 years from age 60 to 80 and, overall, dementia afflicts one third to one half of people over age 80 (Landefeld et al. 2004). Dementia risk is increased for populations that are also at risk for incarceration, including those with less education and racial minorities such as African Americans and Latinos (Alzheimer's Association [AA] 2010). Court liaison referrals show rates of dementia in older adults of 19–30 %, and in prison, correctional officers identify cognitively impaired prisoners at nearly five times the rate of prison health-care officials (Aday 2003; Williams et al. 2009). Adding to the sizable personal costs associated with cognitive and behavioral symptoms, dementia presents a huge cost to society. One 2010 report cited $172 billion as the annual cost of care in the USA related to dementia (Alzheimer's Association [AA] 2010). The implications of these findings for older prisoners, and for the correctional system itself, are significant. Undiagnosed older prisoners with dementia may be at risk for poor self-care, abuse or manipulation by other prisoners, citation for behaviors beyond their control, and failure to meet parole requirements leading to reincarceration for reasons having to do with an undiagnosed or inappropriately managed medical condition rather than a criminal offense. These prisoners are also likely to increase financial burdens on an already strained correctional system at a point in their lives when they may be less and less capable of committing crimes or posing a threat to society.

Incontinence

Incontinence also increases with age (Landefeld et al. 2004). Still, community-based studies consistently find that incontinence is underdiagnosed and that older patients are unlikely to raise the issue unless specifically asked by their provider. Very little data describing incontinence in prisoners exists; however, one study of prisoners with data collected in 1989 found incontinence in 13.9 % of inmates in their 50s and in nearly 40 % of those aged 60 and older (Colsher et al. 1992). In another study of female prisoners aged 55 and older, 22 % of participants reported incontinence (Williams et al. 2006). Although incontinence can be reversible, and a majority of cases will improve with treatment, a study of California prisons found no evidence of nursing or medical staff investigating the causes of incontinence (Hill et al. 2006). The same study found that in many prisons incontinence supplies were either unavailable or required a co-pay. The observed lack of either provider training or readily available supplies in many prisons could result in social isolation, depression, decreased functional status, and potential ridicule or violence (Williams et al. 2006).

Sensory Impairment

Of similar concern, sensory impairment increases markedly with age and affects a majority of older adults (Landefeld et al. 2004). In the community and in prison, hearing and vision impairment are common with age and are associated with increased rates of balance impairment, social isolation, and disability (Hill et al. 2006; Williams et al. 2006). The risks associated with sensory impairment while in prison are significant as older prisoners may have trouble negotiating unseen obstacles and can be unable to hear orders or accused of disrespecting other inmates whose comments they have not heard (Williams and Abraldes 2007).

Infectious Diseases

Prisoners of all ages tend to have higher rates of hepatitis B and C, HIV/AIDS, and sexually transmitted diseases (STDs) than do non-incarcerated persons. Among older prisoners, both acute and chronic infectious diseases account for many hospitalizations (Glaser et al. 1990). Several studies from the Texas Department of Criminal Justice found that prevalence rates of tuberculosis, hepatitis B and C, methicillin-resistant Staphylococcus aureus (MRSA), syphilis, and pneumonia were all substantially higher in prisoners 50 years of age or older than in young or middle-aged prisoners. It should also be noted that while HIV/AIDS prevalence was highest among prisoners aged 30–49 in these studies, 10 % of all new AIDS diagnoses in the USA outside of prison were in older adults aged 65 or older (Baillargeon et al. 2000, 2004). High rates of infectious disease among older prisoners may increase their risk of adverse outcomes following release while presenting an added public health challenge.

Death

Both inside and outside prison, death rates increase with advancing age. The death rate for prisoners over age 55 is three times that for prisoners aged 45–54, and two thirds of all state prison deaths occur in people aged 45 or older (Bureau of Justice Statistics 2007). While prisoners 65 and older make up just 1 % of the prison population, they account for 15 % of deaths. For the 55–64 age group, prison death rates are 56 % higher than for community populations of the same age (Bureau of Justice Statistics 2007). During incarceration, the three leading causes of death are heart disease, cancer, and liver disease (Bureau of Justice Statistics 2007). For female prisoners, breast, ovarian, cervical, and uterine cancers together accounted for 24 % of all cancer deaths in prison (Bureau of Justice Statistics 2007). In addition, during the first 2 weeks following release from prison, older former prisoners face a relative risk for death that is three times the community norm (Binswanger et al. 2007).

Older prisoners with compromised health or functional status worry about dying while incarcerated (Aday 2003). Such worries appear well founded given current sentencing practices and the lack of control prisoners often feel over their lives and health-care options (Linder et al. 2002). Prisoners' limited understanding of the terminology related to end-of-life care, and their desire to be dependent on family or friends for care rather than on institutional staff, may contribute to such feelings. Further, services for chronic medical conditions, age-related functional disabilities, palliative care, and death are far from optimal in prisons (Kerbs and Jolley 2009). There is, however, some evidence of increased attention to end-of-life services in prison, including an increase in the number of prisons providing hospice care (Linder et al. 2002).

Psychosocial Factors

The approach of geriatric medicine is to understand how psychosocial factors in an older adult's life interact with biologic attributes and the environment to accelerate or decelerate health problems and functional ability over time. While many individual psychological and sociocultural factors exist for older prisoners, this section highlights several that are particularly salient in understanding the health and function of older adults in the criminal justice context. Specifically, this section discusses five interrelated "psychosocial" factors that are of particular importance in understanding and contextualizing the health of older prisoners: sociodemographics and identity, educational attainment and health literacy, mental health problems, substance abuse, and social support and isolation.

Sociodemographics and Identity

The racial and ethnic composition of America's older population is changing, trends that are mirrored in US prisons. According to the US Census Bureau, the percentage of Americans aged 55 or older who are nonwhite increased from 12.1 % in 1990 to 14.7 % in 2010. Estimates from the American Geriatric Society suggest that the proportion of nonwhites in America's older population will increase in coming decades. In prisons, while people from racial or ethnic minorities make up a disproportionately large percentage of the overall population, this is less so in the older prisoner age groups. Black men, for example, comprise 39.0 % of the total state and federal male prison population, yet they account for only 24.8 % of male prisoners aged 50 years or older (BJS 2011). Yet, as many prisoners in the 40–55-year-old age group have long, indeterminate, or life sentences, a significant demographic shift towards a higher percentage of older, nonwhite prisoners is anticipated.

As a result, health conditions that disproportionately affect nonwhite prisoners are likely to become more common in prison, as may racial and ethnic variations in disease presentation and response to medications. Unfortunately, the precise ways in which the health of older prisoners differs according to race/ethnicity is understudied and can only be extrapolated from what is known about racial/ethnic differences in health among older Americans in general. For example, black Americans have a higher mortality and morbidity than white Americans, and numerous racial/ethnic health disparities exist in areas like cognitive impairment, functional impairment, chronic medical conditions, and nursing home placement (Gornick et al. 1996). More research is needed, however, to understand how these health-related epidemiologic differences may play out in the criminal justice population as demographic shifts occur.

Educational Attainment

Low educational attainment and poor health literacy are strongly associated with increased morbidity and mortality. The 1992 National Adult Literacy Survey revealed that 60 % of all prisoners performed at the lowest literacy levels, placing prisoners among the least literate populations in the country (Kirsch et al. 1993). Moreover, approximately half of those over age 60 in the USA read at the lowest reading level (Sudore et al. 2006), suggesting that older prisoners may have even lower literacy rates than the overall correctional population. Educational attainment can also affect adjustment to prison life, whereby older prisoners who are able to read and can participate in prison activities requiring some educational skills are found to be more likely to make a positive adjustment to prison life (Kratcoski and Babb 1990). In the period following release from prison, attaining a job and secure housing may also be more difficult for older adults with limited literacy.

Mental Health

Diagnoses of mental health disorders have reached epidemic proportions in US prisons. Many experts ascribe the high rates of mental health problems to the "deinstitutionalization movement" of the 1970s which sought to treat persons with serious mental illness in community mental health facilities rather than in psychiatric hospitals (Baillargeon et al. 2009). Unfortunately, the policy change was not accompanied by a corresponding increase in the number of mental health facilities, leaving a significant gap between communities' needs and their capacity to meet them. As a result, prisons and jail have increasingly assumed the burden of care for psychiatric patients. The Bureau of Justice Statistics now reports that half of prisoners have at least one mental health condition (James and Glaze 2006).

The precise prevalence of mental health diagnoses in older prisoners is not known. One study by the Department of Justice found that 39.6 % of state prisoners aged 55 or older had mental illness (Haugebrook et al. 2010). Other studies have reported rates of "serious mental illness" among older adults in prison similar to rates in non-incarcerated older adults. Such mental health conditions generally include depressive disorders, bipolar disorders, and schizophrenia or related psychotic disorders, and it is important to note that some mental health problems, such as posttraumatic stress disorder (PTSD), are often not captured in the definition of "serious mental illness." Reported rates vary from state to state, from Texas (11.0 %) and Utah (13.6 %) to Tennessee (16.0 %). In comparison, the prevalence of serious mental illness in non-incarcerated older adults is approximately 15–20 % (Caverley 2006).

Prison mental health systems have been criticized for focusing on the management of mental illness-related crises at the expense of facilitating recovery and promoting coping skills. A BJS report found that although more than 50 % of inmates exhibited symptoms of mental illness, only 22 % of state prisoners and 7 % of jail inmates received treatment (James and Glaze 2006). Indeed, a review of the existing health-related research on older prisoners found that psychiatric conditions are the most commonly untreated health problem (Loeb and AbuDagga 2006). Even in the absence of a diagnosis of serious mental illness, the psychological stress and acutely depressed moods often associated with incarceration may have an additional negative impact on older prisoners' overall health status (Aday 2003).

Substance Abuse

Substance use and abuse in prisoners is both common and often related to their incarceration. Over 80 % of state prisoners in 2004 reported past drug use, and over 30 % said they were using drugs at the time of their offense (Mumola and Karberg 2006). Over half of all state prisoners (53 %) met the *DSM-IV* criteria for drug dependence or abuse (Mumola and Karberg 2006). In addition, a study using similar data found that 20 % of state prisoners had a history of injection-drug use (James and Glaze 2006), a practice associated with increased risk for a number of infectious diseases. Unfortunately, there are limited substance abuse treatment programs in the US criminal justice system. Despite the high rate of prior substance use among prisoners, only 40 % reported participating in any type of drug or alcohol treatment or related program during their incarceration (Mumola and Karberg 2006). Little is known about the prevalence of substance abuse in older prisoners or effective methods of treatment in this population.

Social Support and Social Isolation

The social science literature suggests that, with increasing age, adults constrict their web of social interactions. The same effect appears to occur as a result of imprisonment, where Bond et al. (2005) showed a general decrease in the size of social networks for older and younger prisoners alike. This study also noted that social visits from outside prison were less common for older prisoners (Bond et al. 2005). Similarly, Kratcoski and Babb (1990) found that nearly one in three older prisoners never had visitors. Those who did, the study reported, had them less frequently than younger prisoners. This was explained in part by prisoners' preexisting social networks and in part by factors within the correctional system. Over 50 % of older prisoners, for example, were separated, divorced, or single at the time of their incarceration. Older prisoners were also less likely to interact with their fellow prisoners, with less than half of study participants reporting involvement in educational programming and fewer than a quarter involved in self-help groups (Kratcoski and Babb 1990). Such isolation may lead to depression and other mental health problems, a fact underscored by Bond et al.'s (2005) finding that older prisoners tend to place greater emotional value on their diminishing social networks than do younger prisoners.

Problems with social support and social isolation may also lead to restricted prisoner participation in programming with the potential to improve physical health and functioning, such as recreational activities. Declining physical health in older prisoners can itself limit the range and number of activities in which they participate, thus contributing to their social isolation and completing a cycle of worsening overall health outcomes (Kratcoski and Babb 1990). For this reason, and also because some older adults report being fearful of younger prisoners, prison experts have debated the benefits of age-segregated housing for older prisoners

(Kerbs and Jolley 2009; Williams and Abraldes 2007). In this debate, the potential to decrease victimization and improve the physical environment for older adults (e.g., more lower bunks, ramps, and wheelchair accessibility) is balanced against the potential pitfalls of segregation, which may include decreased access to programming and informal caregiver support from younger inmates as well as increased social isolation and boredom (Kerbs and Jolley 2009; Williams and Abraldes 2007).

Environmental Factors

Both functional ability and one's environment have an impact on older prisoners' health because of the interrelation of health and function in older persons and also because the unique environment often mediates the impact of medical, psychological, or social factors on an older person's health and functional status.

Most jails and prisons were built and intended for younger persons. An older prisoner with poor vision and arthritis who is placed with a cell-mate willing to help with his personal care might be able to function adequately despite his limitations. Still, facilities that are generally geared to younger, healthier prisoners are likely to adversely affect an older prisoner's health due to environmental factors that disregard older prisoners' needs such as inappropriately high calorie meals or psychological services which fail to assess cognition. Prisons are also frequently located in remote rural areas. This basic geographic characteristic can decrease prisoners' access to ADA-type resources (Kerbs and Jolley 2009) and/or decrease their access to outside health-care services, such as physical therapy, occupational therapy, and speech therapy. In addition, remote prison locations may lead to prohibitive expenses associated with family visitation or duress on the part of similarly aged spouses or siblings, contributing to the social isolation that is a common feature of incarceration.

Geriatric medicine distinguishes between *disability*, defined as the physical inability to perform a task, and *handicap*, a functional deficit in a social role or context that may result from a disability. Given skilled medical assessment and adequate environmental accommodations, disabilities do not necessarily need to lead to handicaps. Assistive devices, social supports, and environmental modifications can each prove essential in effectively managing disabilities. As such, older prisoners often require modified routines and environments to meet common health and functional challenges (Aday 2003; Wahidin 2006). This is of particular import as correctional facilities, when compared to the community, typically require a higher level of physical functioning for older adults. Common daily activities that might be difficult for older prisoners to perform include, for example, climbing onto a top bunk and dropping to the floor for alarms (Williams et al. 2006). In some instances, specialized housing units have been developed within prisons to provide specialized services for older prisoners with functional limitations (Aday 2003). In other jurisdictions, younger prisoners are identified to assist older prisoners with the daily tasks of prison life (Hill et al. 2006). At a minimum, provision of adequate heating and cooling, signage, lighting, and accessible bathrooms and sleeping spaces are necessary to decrease the functional threats to safety that can be present for older, often frail prisoners (Hill et al. 2006).

In many prisons, accessing medical care can paradoxically pose disproportionate challenges for ill and functionally challenged prisoners. The report (Hill et al. 2006) on California prisons described four "unnecessary bureaucratic and financial obstacles" associated with some prisoners' access to health-care services. First, prisoners were required to make a written case for their need before they could schedule nonemergency appointments. As described above, older prisoners are more likely than their younger counterparts to have low literacy levels and poor writing skills. Second, to receive same-day care, prisoners had to sign in and wait outside, regardless of weather and often in a standing position. Older and ill prisoners are both less likely to be able to stand for hours at a time and are often more vulnerable to the effects of extremes in temperature. Third, prisoners in need of emergent care were required to convince untrained guards of the legitimacy of their needs.

Here, language barriers, guards' inflated perceptions of malingering, and illness itself can interfere with older prisoner's abilities to effectively plead their case. Finally, the report noted that most medical services required a co-pay of between 2 and 10 dollars. Prisoners could accrue money for co-payments from family or through work during incarceration. Yet, older prisoners are more likely to lack social support from outside prison and may be unable to work due to functional or health-related limitations, possibly leading to delayed care. Further, in such instances where co-payments work as a disincentive to seeking preventive or routine care, overall health-care expenditure may go up as older prisoners are only treated at the acute stages of illness. While some of these issues may be unique to the California prison system, most are paradigmatic of barriers found in prisons nationwide.

An Agenda for Future Policy and Research on Older Prisoners

Given the many potential biopsychosocial challenges that exist for older prisoners, a research agenda focused on identifying, assessing, and improving the myriad threats to optimal health is needed. Ideally, such an agenda would reflect existing opportunities for policy change. For example, increased attention is being paid to periods of transition: from one prison to another and from incarceration to the community. Here, improved policies to minimize the adverse health events that occur at the time of release (such as homelessness, lack of access to medications or care, recidivism, and emergency services utilization) could begin to address the elevated death rate for older prisoners during community reentry.

In addition, compassionate (or "medical") release, which allows eligible prisoners with serious medical illness to die outside of prison before sentence completion, is being increasingly examined by policymakers. These programs are being assessed for their potential to optimize the care of prisoners with life-limiting illnesses while unburdening prison health-care systems and lowering state correctional expenditures. Increased leadership and research from the medical and public health communities is needed to develop compassionate release programs with transparent, evidence-based medical eligibility criteria. These criteria should move beyond the common criteria of having a prognosis of 6–12 months and instead reflect the many ways that people experience serious medical illness and death. For example, progressive frailty and dementia may lead to profound impairment lasting far beyond 12 months – such scenarios are not addressed in many compassionate release policies. The application process for compassionate release should also be designed so that it can be navigated by seriously ill prisoners and their advocates (Williams et al. 2011). These policy challenges, and others related to the rapid growth of older prisoners, require a more substantial research base than currently exists to describe this growing population.

The integrative biopsychosocial geriatrics model presented in this entry presents a framework through which to fashion a research and policy agenda. In many states, however, health-related data collection carried out by state correctional systems is not differentiated by age groups. Moreover, the age-related statistics that are collected are not always divided into those subgroups with the greatest clinical utility. Unfortunately, the existing knowledge gap only widens when considering how to optimize care for older prisoners with complex medical conditions such as dementia and functional impairment (Williams et al. 2012a). Future policy-oriented research should specifically focus on caregiver and prison-worker education and on the development and provision of strategies to screen older prisoners for cognitive dysfunction, fall risk, sensory deficits, and functional impairment. In addition, because many older adults will develop a serious medical illness and die in prison but will not qualify for early release, enhancement of prison palliative care services is greatly needed. In the psychosocial arena, it is important to systematically and scientifically explore the utility and acceptability of prison-based volunteer programs designed to address the needs of older adults. Also important is the need to develop and test treatment programs that can address the complex impacts of substance abuse and mental health disorders on optimal health status in older prisoners. From an environmental standpoint, the academic debate about clustering or age segregation of prisoners is sure to continue, with future cost-effectiveness and quality research likely playing a central role (Williams et al. 2012b). Other potential strategies, however, such as designated "older prisoner" yard time, which might serve as an effective compromise between total segregation and total immersion, should also be explored.

As the population of older and frail prisoners increases, it is critical that health and social science researchers work together with clinicians, criminal justice health-care and policy experts, and legislators. This would enable, for example, more analysis of current and future unintended public health consequences of criminal justice policies and sentencing laws. Enhanced collaboration between the health sciences, law, bioethics, and criminology will help to ensure socially acceptable, legally appropriate, and medically justifiable medical programs that pose minimum risk to public safety and span from the time of arrest through successful reentry into the community (Williams et al. 2011). Overall, the "aging crisis" in corrections requires medical, legal, social policy and political input to identify humane and cost-effective ways of promoting safety and optimizing health and function for older adults in the criminal justice system. Analyzing the aging crisis through a multidisciplinary geriatrics lens is a good first step.

Acknowledgements Funding Support: Brie Williams received funding from the National Institute of Aging (K23AG033102) and the John A. Hartford Foundation's Center of Excellence in Geriatric Medicine and Training. Dr. Williams is an employee of the Department of Veterans Affairs. The opinions expressed in this entry may not represent those of the VA. Potential Conflicts of Interest: Dr. Williams has been a consultant in legal cases about prison conditions of confinement.

Related Entries

- ▶ Age-Crime Curve
- ▶ Aging Prison Population: Factors to Consider
- ▶ Drug Abuse and Alcohol Dependence Among Inmates
- ▶ Early Release from Prison
- ▶ European Perspectives on Long-Term Imprisonment
- ▶ Health Issues in Prison Reentry Models
- ▶ Imprisonment and Cardiac Risk
- ▶ Mortality After Release from Prison
- ▶ Parole and Prisoner Reentry in the United States
- ▶ Treating Mentally Ill Offenders
- ▶ Unintended Effects of Imprisonment

Recommended Reading and References

Aday R (2003) Aging prisoners: crisis in American corrections. Praeger, Westport

Alzheimer's Association (2010) 2010 Alzheimer's disease facts and figures. Retrieved from http://www.alz.org/documents_custom/report_alzfactsfigures2010.pdf

Baillargeon J, Black SA, Pulvino J, Dunn K (2000) The disease profile of Texas prison inmates. Ann Epidemiol 10(2):74–80. doi:10.1016/S1047-2797(99)000332

Baillargeon J, Black SA, Leach CT, Jenson H, Pulvino J, Bradshaw P, Murray O (2004) The infectious disease profile of Texas prison inmates. Prev Med 28:607–612. doi:10.1016/j.ypmed.2003.11.020

Baillargeon J, Binswanger I, Penn J, Williams B, Murray O, Raimer B (2009) The revolving prison door: psychiatric disorders and repeat incarcerations. Am J Psychiatry 166:103–109. doi:10.1176/appi/ajp.2008.08030416

Binswanger I, Stern MF, Deyo RA, Heagerty PJ, Cheadle A, Elmore JG, Koepsell TD (2007) Release from prison – a high risk for death for former inmates. N Engl J Med 356:157–165. doi:10.1056/NEJMsa064115

Binswanger I, Krueger P, Steiner J (2009) Prevalence of chronic medical conditions among jail and prison inmates in the USA compared with the general population. J Epidemiol Community Health 63(11):912–919. doi:10.1136/jech.2009.090662

Bond GD, Thompson LA, Malloy DM (2005) Lifespan differences in the social networks of prison inmates. Int J Aging Hum Dev 61:161–178. doi:10.0010/1602835400013268546

Bureau of Justice Statistics (2007) Data brief: medical causes of death in state prisons, 2001–2004 [Data file]. Retrieved fromhttp://www.ojp.usdoj.gov/bjs/abstract/mcdsp04.htm

Bureau of Justice Statistics (2011) Prisoners series: 1990–2010. Department of Justice, Washington, DC. Retrieved from http://bjs.ojp.usdoj.gov/index.cfm?ty=pbse&sid=40. Accessed 1 Nov 2011

California State Auditor (2010) Inmates sentenced under the three strikes law and a small number of inmates receiving specialty health care represent significant costs. Bureau of State Audits, Sacramento, Report 2009-107.2

Caverley SJ (2006) Older mentally ill inmates: a descriptive study. J Correct Health Care 12:262–268. doi:10.1177/1078345806295546

Colsher PL, Wallace RB, Loeffelholz PL, Sales M (1992) Health status of older male prisoners: a comprehensive survey. Am J Public Health 82:881–884, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1694178/

Covinsky KE, Justice AC, Rosenthal GE et al (1997) Measuring prognosis and case-mix in hospitalized elders: the importance of functional status. J Gen Intern Med 12:203–208

Damberg C, Shaw R, Teleki S, Hiatt L, Asch S (2011) A review of quality measures used by state and federal prisons. J Correct Health Care 17(2):122–137. doi:10.1177/1078345810397605

Glaser JB, Warchol A, D'Angelo D, Guterman H (1990) Infectious diseases of geriatric inmates. Rev Infect Dis 12:683–692, http://www.jstor.org/stable/4455607

Gornick ME, Eggers PW, Reilly TW, Mentnech RM, Fitterman LK, Kucken LE, Vladeck BC (1996) Effects of race and income on mortality and use of services among medicare beneficiaries. N Engl J Med 335:791–799, http://www.nejm.org/doi/pdf/10.1056/NEJM199609123351106

Haugebrook S, Zgoba KM, Maschi T, Morgen K, Brown D (2010) Trauma, stress, health, and mental health issues among ethnically diverse older adult prisoners. J Correct Health Care 16:220–229. doi:10.1177/1078345810367482

Hill T, Williams BA, Cobe G, Lindquist KJ (2006) Aging inmates: challenges for healthcare and custody. The California Department of Corrections and Rehabilitation, San Francisco

Human Rights Watch (2012) Old behind bars: the aging prison population in the United States. Human Rights Watch Report, New York

Inouye SK, Studenski S, Tinetti ME, Kuchel GA (2007) Geriatric syndromes: clinical, research and policy implications of a core geriatric concept. J Am Geriatr Soc 55:780–791. doi:10.1111/j.1532-5415.2007.01156.x

James DJ, Glaze LE (2006) Mental health problems of prison and jail inmates [Bureau of Justice Statistics Special Report, NCJ 213600]. US Department of Justice, Washington, DC, http://www.ojp.usdoj.gov/bjs/pub/pdf/mhppji.pdf

Kerbs JJ, Jolley JM (2009) A commentary on age segregation for older prisoners: philosophical and pragmatic considerations for correctional systems. Crim Justice Rev 34:119–139. doi:10.1177/0734016808324245

Kirsch IS, Jungeblut A, Jenkins L, Kolstad A (1993) Adult literacy in America: a first look at the findings of the national adult literacy survey. National Center for Educational Statistics, U.S. Department of Education, Washington, DC

Kratcoski P, Babb S (1990) Adjustment of older inmates: an analysis by institutional structure and gender. J Contemp Crim Justice 6:264–281. doi:10.1177/104398629000600407

Landefeld CS, Palmer RM, Johnson MA, Johnston CB, Lyons WL (2004) Current geriatric diagnosis and treatment. McGraw-Hill, New York

Linder JF, Enders SR, Craig E, Richardson J, Meyers FJ (2002) Hospice care for the incarcerated in the United States: an introduction. J Palliat Med 5:549–552. doi:10.1089/109662102760269788

Loeb SJ, AbuDagga A (2006) Health-related research on older inmates: an integrative review. Res Nurs Health 29:556–565. doi:10.1002/nur.20177

Mumola CJ, Karberg JC (2006) Drug use and dependence, state and federal prisoners, 2004. U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics, Washington, DC

Pew Center on the States (2009) One in one hundred: behind bars in America in 2008. The Pew Charitable Trust, Washington, DC

Sudore RL, Yaffe K, Satterfield S, Harris TB, Mehta KM, Simonsick EM, Schillinger D (2006) Limited literacy and mortality in the elderly: the health, aging, and body composition study. J Gen Intern Med 21:806–812. doi:10.1111/j.1525-1497.2006.00539.x

Wahidin A (2006) 'No problems – old and quiet': imprisonment in later life. In: Wahidin A, Cain M (eds) Ageing, crime and society. Willan Publishing, Portland, pp 171–192

Williams B, Abraldes R (2007) Growing older: challenges of prison and re-entry for the elderly. In: Greifinger R (ed) Public health behind bars: from prisons to communities. Springer, New York, pp 56–72

Williams B, Lindquist K, Sudore RL, Strupp HM, Willmott DJ, Walter LC (2006) Being old and doing time: functional impairment and adverse experiences of geriatric female prisoners. J Am Geriatr Soc 54:702–707. doi:10.1111/j.1532-5415.2006.00662.x

Williams B, Lindquist K, Hill T, Baillargeon J, Mellow J, Greifinger R, Walter L (2009) Caregiving behind bars: correctional officer reports of disability in geriatric prisoners. J Am Geriatr Soc 57:1286–1292. doi:10.1111/j.1532-5415.2009.02286.x

Williams B, Sudore R, Greifinger R, Morrison RS (2011) Balancing punishment and compassion for seriously ill prisoners. Ann Intern Med 155(2):122–127. doi:10.1059/0003-4819-155-2-201107190

Williams BA, Stern MF, Mellow J, Safer M, Greifinger RB (2012a) Aging in correctional custody: setting a policy agenda for older prisoner health care. Am J Publ Health 102(8):1475–1481

Williams B, Goodwin J, Baillargeon J, Ahalt C, Walter L (2012b) Addressing the aging crisis in U.S. Criminal justice healthcare. J Am Geriatr Soc 60(6):1150–1156. doi:10.1111/j.1532-5415.2012.03962.x

Wool J (2007) Litigating for better medical care. In: Greifinger R (ed) Public health behind bars: from prisons to communities. Springer, New York, pp 25–41

Aging Prison Population: Factors to Consider

Jeanne Subjack[1], Gracie Aguilar[2] and Victoria B. Titterington[1]
[1]College of Criminal Justice, Sam Houston State University, Huntsville, TX, USA
[2]Spring, TX, USA

Overview

Elderly inmates are considered the fastest-growing segment of the incarcerated population. In turn, they pose unique challenges for corrections management, particularly as it relates to their health. This review includes a description of the historic growth of an older prison population, the current health-care demands being placed upon prison services and budgets, and the attendant controversies about how to best address these demands. This entry concludes with the open questions that must necessarily be answered in moving forward with the effective care and management of an ever-growing group of older, and often infirm, inmates.

Introduction

There is unprecedented growth in the incarcerated elderly population. These prisoners are now the fastest-growing group of inmates, increasing at a rate three times that of the general prison population. Thousands of inmates are baby boomers, and as they age, prisons find themselves having to plan to care for an aging population (Gater 2010). Among reasons for this growth are changes in sentencing and parole practices, recidivism rates, and increasing incarceration rates in several major offense categories (Reimer 2008). For example, in the past decade, the war on drugs and tough mandatory sentencing laws have doubled the number of incarcerated persons. Inmates sentenced to life, or practical life, represents the future of the already growing population of older inmates. This population has an increased potential for medical problems and emergencies, and these circumstances often develop sooner in prison populations.

Understanding and addressing prisoner health care is a relatively recent phenomenon, primarily in the past 30 years. Prior to the 1970s, little was known about how inmates were cared for in prison. What really put the health-care issue on the forefront for change were lawsuits filed by inmates against prisons. As a result, the US Supreme Court became more interested in changing health-care practices. Professional standards of care were implemented, beginning with the American Public Health Association and followed by the American Medical Association, American Correctional Association, and the National Commission on Correctional Health Care.

Older Prison Inmates Today

Defining the Elder Inmate

It is difficult to determine how states classify their elderly inmates and what differences exist among prisons. In a statement made in a report written about California, regarding the differing ages, “it is virtually impossible to determine how many elderly are incarcerated nationwide because scholars and correction officials differ as to what age is an appropriate cut off to label someone as elderly” (Gubler and Petersilia 2006, p. 10). Using the National Institute of Corrections as the benchmark, the elderly incarcerated are defined as any inmates above the age of 50 years, because, morphologically, the average prisoner has a reduced health status approximating the health condition of a non-incarcerated person who is 10–15 years older. In turn, research indicates that since 1995, the total population of older prisoners has grown by about 10,000 per year (Aday 2003). Prison administrators have found that the “graying” of America is evident in the prison population as well.

Due to high-risk lifestyles and the lack of regular health care before their incarceration, many inmates age faster than their contemporaries on the outside. So, for example, a 50-year-old inmate may be considered a “senior citizen” and would need to take advantage of services and facilities for the aging (Gater 2010). The inherent stress of prison life, along with the declining health of many inmates, also contributes to them being elderly before their time. As a result, they have become the major consumers of corrections services (Aday 2003). Collectively, Beckett et al. (2003) found that incarceration results in aging in place, challenges to self-care abilities, and greater likelihood of parolees that leave prison with serious health problems.

Reasons for Their Imprisonment

Tougher sentencing accounts for much of the increase in the proportion of correctional populations who are considered old. In the 1980s, many states abolished parole and passed "three strikes and you're out" laws. The intention of three strikes laws has been to ensure longer prison sentences and greater punishment for those who commit a felony and have been previously convicted of serious and/or violent felony offenses (California Penal Code, Section 667.5). Though the average age of persons nationally who fall under three strikes legislation is 31 (Reimer 2008), individuals convicted of second or third strikes may not be granted probation and must serve their sentences in state prisons. Unintended consequences of this legislation include an overcrowding crisis and an aging prison population.

Using California as a major example, at the end of 2004, there were almost 43,000 inmates serving time in prison under the three strikes law, making up about 26 % of the total prison population. Of the striker population, more than 35,000 were second strikers, and about 7,500 were third strikers. The most common offenses for which strikers were currently serving time in prison were robbery, burglary, assault, and possession of drugs (California Legislative Analyst's Office 2005, pp. 15–16). But critics of these laws seriously question their cost-effectiveness.For example, Greenwood (1994) and colleagues constructed a mathematical model that predicted the benefits and costs of the three strikes law in California and found a possible increase to the costs of that state's criminal justice system of an average of $5.5 billion annually.

At the federal level, Congress imposed mandatory sentencing guidelines, such as determinate terms, requiring the application of maximum sentences for crimes of violence and the doubling of maximum penalties for all felony classes (Reimer 2008). Truth-in-sentencing laws (requiring the convicted criminal to serve 85 % of the time imposed before qualifying for release) will push these numbers even higher. More inmates will remain in prison for longer periods of time. And more inmates will grow old in prison.

Recidivism is another factor accounting for an ever-older inmate population nationally. Reimer (2008) defines recidivism as a new criminal activity or technical violation of parole, probation, or nondepartmental community placement within 3 years of release. Colorado is used as an example to show the high frequency of recidivism. The average recidivism for males is 53.15 % and for females is 51.1 % (Bureau of Justice Statistics 2006). This amounts to one person for every two that are released from prison returning to prison within 36 months (Reimer 2008).

Factors Negatively Impacting Elder Inmates and Their Health

The increasing numbers of older offenders behind bars have resulted in greater numbers of persons needing specialized medical treatment, meaning they will consume a disproportionately large share of health-care resources. This reality of a larger portion of health care being consumed by a smaller subgroup of prisoners parallels the phenomenon in the general US population: approximately one-fifth of all persons consume about 80 % of all medical resources. Offenders are more likely to require hospitalization than younger offenders and account for a large proportion of hospital and specialty services costs to prisons. Much like the free world populations, the aging offender population has a high incidence of hypertension, diabetes, atherosclerotic heart disease, and other medical conditions that require long-term and oftentimes expensive treatment. Issues concerning diet also arise, especially if inmates require liquid diets or restricted diets due to an illness or other condition (Caverley 2006).

Stress is yet another factor contributing to the acceleration of the aging process in prison.

For example, there is an increasing concern about the growing number of older prisoners and their safety (Aday 2003), including a fear of victimization by younger, stronger inmates.

Similar to older males, older female inmates housed in the general prison population often express a need for greater privacy, and studies report that many older inmates prefer to live with people of their own age (Walsh 1992). Research indicates that these older females are much more likely to see other inmates as aggressive and violent than do older male inmates (Krabill and Aday 2005). In one study, a significant number (30 %) of the older female inmates stated that they were either occasionally, frequently, or always afraid (Kratcoski and Babb 1990). Additionally, older women are significantly less likely to be involved in various forms of sporting and recreational activities (Krabill and Aday 2005), thereby contributing to the possibility of victimization due to their relatively poorer health.

Along with an increase in medical issues that may typically accompany the aging process is the effect that poor health-care services and prison conditions may have on inmates' health. It is believed that the reality of being imprisoned as well as the conditions therein may adversely impact older inmate's health (Aday 2003; Haugebrook et al. 2010). Other issues exist as well, such as confrontations among older and younger inmates. Assaults are common among prisoners, especially among those inmates other prisoners view as weak. This may further contribute to the weakened state of elderly inmates and therefore increase their need for medical attention (Oklahoma Department of Corrections 2008). Many older inmates have to address the fact that they may die while incarcerated, and the social support to deal with such issues is lessened for them while in prison (Haugebrook et al. 2010).

Prisoners' Rights to Health Care

Although prisoners have diminished rights compared to those in the free world, health care is one right to which they are entitled. They are given a basic right to health care, because they are unable to care for themselves and once they are imprisoned, they are the responsibility of the state. The landmark case that addressed prisoners' right to health care was *Estelle v. Gamble*, in which the US Supreme Court ruled that deliberate indifference to a prisoner's medical needs is considered cruel and unusual punishment under the Eighth Amendment (*Estelle v. Gamble, 429 U.S. 97* 1976). This established the minimum standards for prisoner health care, which instructed prison officials that a basic level of health care was required for prisoners and intentionally overlooking their needs would be unacceptable. Although the case did not create specific guidelines for prison officials to follow when implementing health-care policies, it still indicated that prison officials can be held accountable for the mistreatment of their prisoners' medical needs and rights. *Estelle*, however, was not the only case to address prisoner health care.

West v. Atkins declared that the state has an obligation under the Eighth Amendment to provide proper medical care to an incarcerated inmate (*West v. Atkins, 487 U.S. 42* 1988). This case was significant because it stated that even state prisons that contracted physicians to care for inmates could still be held accountable for the care provided to its inmates. Therefore, the physician has a duty to provide the necessary medical procedures to ensure the well-being of the inmate. Under the decision of *Fernandez v. United States* in Florida, it was decided that prisons must provide "services at a level reasonably commensurate with modern medical science and of a quality acceptable within prudent professional standards" (*Fernandez v. United States and 500 U.S. 948, at 10* 1991, p. 10). Finally, *Tillery v. Owens* stated that all prisoners are to receive a "level of health services reasonably designed to meet routine and emergency medical, dental and psychological or psychiatric care" (Tillery v. Owens, 907 F23d 418 1989, p. 1301). This decision was decided in Pennsylvania and upheld by the State Supreme Court as well, making it a requirement for all prisons in the state. Combining the declarations of the landmark cases on the issue, all inmates have three basic rights to health care. They are the right of access to care, the right to the care that is ordered by officials, and the right to a medical judgment by a professional in the field (Oklahoma Department of Corrections 2008). Prisoners are the only persons in the USA who have a constitutionally protected right to health care, and the courts show no sign of extinguishing that right.

Considering that elderly prisoners may face more serious illnesses, and more frequent illnesses, what level of care is considered appropriate and satisfactory? Things to consider for elderly prisoners include kidney dialysis, routine exams such as colonoscopies, glasses, canes, walkers, and other tools to help them function, as well as cancer treatment.

Current Controversies

Service Needs

It has long been believed that aged inmates should receive special attention reflecting the physiological, psychological, and sociological effects of aging. Yet in a prison system assessment conducted by the Criminal Justice Institute (2001), only 15 of the 49 state prison systems had housing areas designated for elderly inmates. Of these 15, seven were only available for elderly inmates who may have had special medical needs or were eligible for hospice care. Thus, an overarching theme to the current challenges of providing for health-care and related needs of older prisoners is that current services and programs are inadequate. For example, elderly inmates need additional preventive care, orderly conditions, safety, and emotional support. Geriatricians also recommend the immediate step of increasing preventive care and educating older prisoners in strategies for maintaining and monitoring their own health (Aday 2003).

The mental health needs of inmates, including dementia among older persons, currently represent another major challenge to correctional programs. Sufficient community-based treatment resources are presently inadequate to support and stabilize mentally ill persons in their communities, with the result that prisons and jails have become the principal institutions for housing mentally ill persons. Also, according to Reimer (2008), conditions of poverty, isolation, and abandonment may contribute to anger and depression in the aging incarcerated population. Those who are depressed and suicidal must be taught what symptoms may be part of a condition that is treatable (Goldney et al. 2002). Even though older prisoners still encompass a relatively small percentage of the inmate population, they represent significant managerial challenges in relation to adapting prison programs to meet their developmentally specific needs for counseling, health care, education, recreation, and vocational training (Aday 2003).

Staffing Challenges

Typically, recruiting prison physicians trained in specialties has been difficult. A common solution is to hire specialists on a consulting basis to conduct periodic “clinics” at the prisons. Some prisons, however, have determined that it is cost-effective to buy equipment and to build suites for certain frequently used technologies, such as x-rays. Further, some prisons have implemented medical copayments as well as educational programs as means of controlling the use of medical services and their attendant costs.

Another barrier in responding fully to the special needs of the aging inmate is a deficit of the necessary staff aptitudes or essential skills to manage elderly people. This lack of gerontological training for correctional officers suggests that they may fail to properly supervise and safeguard an aging prison population (Aday 2003). Experts believe that corrections staff who work with the elderly should receive specialized training to effectively provide care for the physical and mental health of these inmates (Reimer 2008). In addition, there is much debate about inmates receiving advanced medical care that is not equally available to poorer, uninsured people in the free world.

Prison- and Community-Based Options

In 1992, Flynn recommended that corrections officials pursue a number of strategies to address the health care of elderly inmates, including modify existing work and education programs to include health-care education, preventive medicine, and counseling of the elderly. Staff should also be encouraged to seek inmates who have no close ties to the free world and encourage them to become involved in organized educational, recreational, and vocational activities (Krabill et al. 2005). Corrections professionals must find a way to balance public safety concerns and public demands for a government that operates with economy and efficiency, while responding appropriately to legitimate correctional interests.

Health Care and Other Costs of Older Inmates

Increasing costs are arguably the major issue associated with this rapid increase in older inmates. Studies indicate that inmates 55 and older suffer from an average of three chronic illnesses at a time, and they may experience these illnesses at earlier ages than the general population due to conditions of imprisonment and difficult lives prior to incarceration (Aday 2003). Part of the recent increased cost of elderly inmates includes the mental health services they require, as elderly inmates are more likely to suffer from mental illnesses than other inmates (Caverley 2006). As a result, it is estimated that average annual costs nationally for inmates over age 50 is $70,000, three times the cost of younger inmates (Davidson 2009; Rikard 2007). It is expected that these costs will continue to increase rapidly, especially if the cost of health care in the free world increases.

Current Examples of Meeting the Challenges

A recent edition of the *Correctional Health Care Report* (National Institute of Corrections 2004) highlights some examples of how the program, housing, and treatment considerations of elder inmates are being accommodated. For example, in Fishkill, New York, there is a 30-bed center for inmates who are cognitively impaired. This facility provides for the treatment of inmates with dementia-related conditions, which is something that typically worsens with age. Within the unit, the staff is specially trained to handle inmates with dementia and other related conditions. Although the inmates may commit acts while imprisoned that would typically result in punishment, they often times may not understand the implications of their actions because they lack the cognitive ability (New York Department of Corrections 2008).

Some other examples of special facilities prisons have implemented include Texas, which has a Type I Geriatric Facility that is specifically designed for geriatric offenders and eight Type II Geriatric Facilities to house portions of geriatric offenders. In Michigan, there is a men's geriatric unit and a long-term care unit, while Rhode Island has special housing for wheelchair-bound male inmates. Nevada has one geriatric prison in the entire state, while Alabama also has one facility for elderly and medically disabled inmates. In Florida, there are two facilities for elderly inmates and dorms at two other facilities for elderly inmates within those facilities (Price 2006).

The State of Washington (2010) has an extensive health-care plan that is listed on their Department of Corrections website. Inmates are eligible for infirmary and hospital care, medical and surgical services, maternity services, chemical dependency treatment, mental health services, dental care, emergency care, skilled nursing care and hospice, preventive care, access to pharmacies, durable medical equipment, optical care, and hearing care. While this represents the treatments and options available for all prisoners, there are certainly many treatments that are applicable to elderly inmates, such as hearing, optical, nursing care, and hospice. Although all of the treatments are restricted by necessity, approval, and other means of regulation, the State of Washington clearly has an expansive array of services that it provides to its inmates.

Oklahoma also has health assessments for all prisoners. According to the state's procedure on assessments, "periodic health assessments are to be done every 3 years for offender's age 18–39; every 1–3 years for offenders age 40–64; and annually for offenders 65 and older" (Oklahoma Department of Corrections, p. 7). The increased frequency for elderly prisoners over the age of 65 indicates the importance of monitoring inmates more frequently as they age in order to better assess their needs and conditions as time progresses.

North Carolina conducted an extensive survey of elderly inmates and their health conditions while in prison. The state experienced a 61 % increase in its elderly inmates in a 5-year span, while the overall inmate population only increased by 16 % (Price 2006). Concerning the health status of elderly inmates, 20 % of the surveyed inmates were on a special diet due to various health issues. Furthermore, 28 % of those surveyed required some type of walking assistance, whether it was a cane, brace, or wheelchair. Finally, the cost of housing inmates 50 and older was almost four times the cost of housing younger inmates (Price 2006). Taking all of the costs into consideration, it is understandable why officials are concerned with the growth in the elderly prison population.

As indicated by the above examples, some state prisons provide thorough care of elderly inmates. This increased concern and care, however, is not without its problems. Victims and their families are upset over the care given to the offenders who have negatively impacted their lives so significantly. For example, a proposal in Illinois which would have provided for a path to parole for inmates over 50 years of age who had served more than 25 years in prison has been voted down three times (Davidson 2009). Families of victims advocated against the bill because they believed the prisoners should serve their full sentences.

A major concern when discussing the early release of older prisoners is research indicates that many elderly inmates are in prison due to violent or sexual offenses (Abner 2006). The facility in New York that was described above, however, took the risk offenders may present into consideration when designing the facility. As a result, the facility is analogous to a maximum security prison and has the ability to house all levels of offenders. Facilities with this security level may be more satisfactory to those concerned with the risk presented by violent offenders, regardless of their age.

Compassionate Release

The cost of treating sick, elderly inmates as compared to early release is another issue that brings about several concerns among policy makers and the general population alike. A specific issue that has been considered for elderly prisoners is compassionate release and whether it should be an option for some low-status offenders. Compassionate release programs are those which "call for the early release of prisoners with terminal illnesses that are expected to die within 6 months and whose release poses no risk to society" (del Carmen et al. 2008, p. 15). However, this too continues to be a controversial means of addressing the ever-growing proportion of elder inmates.

Budgetary concerns will continue to increase as the population increases, thus perpetuating the argument for alternative options. Related to early release for low-status elderly prisoners is the idea of transferring some low-risk inmates to nursing homes. However, there are still several issues that exist with this alternative, including the fact that the burden of paying for their care shifts from prisons to other government agencies. Furthermore, statutes that require nursing homes to make public the criminal records of their residents may cause serious public backlash. Although the inmates who were transferred may be too unhealthy or physically unable to present any harm to the other residents, the public may still resent the idea of having convicted offenders residing with other elderly people who have no criminal record.

End-of-Life Issues

Yet another unprecedented challenge for corrections management of older prisoners is the provision of end-of-life services. Action needs to be taken to ensure that elderly inmates are adequately provided for while incarcerated, including those inmates who will pass away during the span of their sentence. Presently, the California Department of Corrections and Rehabilitation (CDCR) operates a licensed hospice for terminally ill inmates. In New York, terminally ill inmates are not dealt with according to age per se, but rather their ability to take care of themselves. Some 70-year-old inmates are in very good shape while some in their 50s are not (Gater 2010). In Virginia, end-of-life or palliative care approaches are provided in-house, and inmates are directly involved in delivery of hospice services to fellow inmates in Angola, Louisiana. Family involvement, including liberal visitation, is also a vital feature to this process. The balance is between addressing public safety concerns while providing for the unique end-of-life care requirements of the elderly (Gater 2010).

Elderly prisoners will continue to serve prison sentences within facilities, whether it is offenders who age while incarcerated or those who are imprisoned at an older age. As a result, it will be interesting to see if, or when, the Supreme Court rules on a case dealing with the compassionate release of an elderly, low-status offender who is in fragile health. This is something to consider for the future, knowing that the elderly prison population continues to increase. Related to this is the question of what is considered deliberate indifference, and if these standards will also evolve over time or if detaining inmates with chronic illnesses who are almost to the point of death will ever be viewed as cruel and unusual punishment. The important question to ask concerning all of these possibilities is if the country will ever be at a place to allow such a release. As it stands now, the general population may view compassionate release as a policy that is too lenient for prisoners, regardless of the crime they committed and their current age and health condition.

Conclusions and Future Research

States are struggling with the best models of both facilities and programs to deal with an aging inmate population (Aday 2003). Also, the diversity of the growing number of older offenders should be recognized and incorporated into rehabilitative programs. In some cases, instead of preparing the inmate for reentry as a productive member of society, wellness programs which aim to keep the individual alert and active are needed. It is also the case that, in order to transfer elderly offenders back to the community, housing and financial assistance must usually be secured for inmates who have been imprisoned for long terms and who have lost all contacts in the community.

There are several ways this current research could be expanded upon in order to increase our understanding of the health care of elderly inmates. One interesting approach would be to study the different needs that exist among male and female elderly inmates, since men and women suffer from different illnesses and chronic diseases throughout the life course. Furthermore, some states have specialized facilities for male elderly inmates, but few have such facilities for female elderly inmates.

Another area for more in-depth research will be to further investigate different chronic illnesses and ailments from which elderly inmates suffer. Instead of looking at the health care of elderly inmates in general, focusing on specific diseases such as cancer and examining how prisons respond to these inmates would be beneficial. Related to this, some research indicates that those inmates who have aged in prison may be healthier than individuals who were older at the time of their incarceration (Caverley 2006). Among other things, this may give us a better assessment of the health services provided in prison.

At this practical level, Aday (2003) points out that corrections administrators and policy makers still want empirical answers to some fundamental questions. These include what may be categorized as the relative costs (savings) to relative benefits of long-term incarceration versus early parole or extended medical furlough for infirm prisoners. Administrators and policy makers also need more information on aging lifers without parole and the effects of long-term institutionalization, as well as the impact of sentencing law changes on the size of the long-term inmate populations. At both programmatic and policy levels, there remain many open questions and much work to be done as it relates to the graying of the US prison population.

Related Entries

- ▶ Aging Correctional Populations
- ▶ Early Release from Prison

Recommended Reading and References

Abner C (2006) Graying prisons: states face challenges of an aging inmate population. State News 49(4). The Council of State Governments. Retrieved from http://www.csg.org

Aday RH (2003) Aging prisoners: crisis in American corrections. Praeger, Westport

Becerra H, Blankstein A (2008) Panel denies release for dying Manson follower Susan Atkins. The LA Times, 16 July 2008. http://articles.latimes.com/2008/jul/16/local/me-atkins16

Beckett J, Peternelj-Taylor C, Johnson RL (2003) Growing old in the correctional system. J Psychosoc Nurs Ment Health Serv 41(9):12–19

Bureau of Justice Statistics (2006a) Medical problems of jail inmates. U.S. Department of Justice, Washington, DC

Bureau of Justice Statistics (2006b) Mental health problems of prison and jail inmates. U.S. Department of Justice, Washington, DC

Bureau of Justice Statistics (2006c) Prisoners in 2005. U.S. Department of Justice, Washington, DC

California Legislative Analyst's Office (1999) Retrieved from http://www.lao.ca.gov/laoapp/main.aspx

California Penal Code (1999) Title 16, Section 667

Caverley SJ (2006) Older mentally ill inmates: a descriptive study. J Correct Health Care 12:262–268

Criminal Justice Institute (2001) Managing the needs of aging inmates and inmates with chronic and terminal illnesses. Criminal Justice Institute, Middletown

Davidson A (2009) The aging prison population. http://news.change.org/stories/the-aging-prison-population

del Carmen RV, Ritter SE, Witt BA (2008) Briefs of leading cases in corrections, 5th edn. LexisNexis, Newark

Estelle v. Gamble, 429 U.S. 97 (1976)

Fernandez v. United States, 500 U.S. 948, at 10 (1991)

Gater L (2010) Aging & chronic care. Correct Forum 19(3):18–22

Goldney RD, Fisher LJ, Wilson DH, Cheok F (2002) Mental health literacy of those with major depression and suicidal ideation: an impediment to help seeking. Suicide Life Threat Behav 32(4):394–403

Gubler T, Petersilia J (2006) California prison reform. Elderly prisoners are literally dying for reform. Stanford University Criminal Justice Center, Palo Alto

Haugebrook S, Zgoba KM, Maschi T, Morgen K, Brown D (2010) Trauma, stress, health, and mental health issues among ethnically diverse older adult prisoners. J Correct Health Care 16:220–229

Krabill JJ, Aday RH (2005) Exploring the social world of aging female prisoners. Women Crim Justice 17(1):27–53

Kratcoski PC, Babb S (1990) Adjustment of older inmates: analysis by institutional structure and gender. J Contemp Crim Justice 6:139–156

National Institute of Corrections Information Center (1997) Prison medical care: special needs populations and cost control. LIS, Longmont

Neeley C, Addison L, Craig-Moreland D (1997a) Addressing the needs of elderly offenders. Correct Today 59(5):120–123

Neeley C, Addison L, Craig-Moreland D (1997b) Addressing the needs of elderly offenders. Correct Today 59(5):120–123

New York Department of Corrections (2008) Prison Gray; New York tackles inmate dementia with special-needs facility. http://www.correctionalnews.com

Oklahoma Department of Corrections (2008) Managing increasing inmate populations. Oklahoma Department of Corrections, Oklahoma City

Price CA (2006) Aging inmate population study. North Carolina Department of Corrections Division of Prisons, Raleigh

Reimer G (2008) The graying of the U.S. prisoner population. J Correct Health Care 14(3):202–208

Rikard RV, Rosenberg E (2007) Aging inmates: a convergence of trends in the American criminal justice system. J Correct Health Care 13:150–162

State of Washington (2010) Department of Corrections Offender Health Plan. http://www.doc.wa.gov/family/offenderlife/docs/OffenderHealthPlan.pdf

Tillery v. Owens, 907 F23d 418, at 1301 (1989)

Walsh C (1992) Correctional theory and practice. Aging inmate offenders: another perspective. Nelson-Hall, Chicago

West v. Atkins, 487 U.S. 42 (1988)

AIDS

► Drug Abuse and Alcohol Dependence Among Inmates

Alcohol Dependence

► Drug Abuse and Alcohol Dependence Among Inmates

Alternative Courts

► Problem-Solving Courts

Alternatives to Pre-trial Detention

Christine Morgenstern
Ernst Moritz Arndt Universität Greifswald,
Greifswald, Germany

Overview

All over the world, unconvicted prisoners form a large part of the prison population; in some countries they even outnumber sentenced prisoners. This is contrary to the fact that, from the fundamental right to liberty and the presumption of innocence, deprivation of liberty must only be applied when less severe mechanisms are insufficient to control the suspect and to guarantee his or her presence at trial. Generally, the principle of necessity or subsidiarity must be applied in a way that the suspect or accused may await the procedure in unrestricted liberty or, in justified cases, under specified conditions. The authorities may only detain a suspect if no noncustodial options are adequate to meet the established aims. The available alternatives to pretrial detention (conditional bail) and their use vary considerably throughout the world. They may take less invasive forms, such as release on recognizance with some obligations attached or reporting regularly to the authorities. But they also may take the form of curfew, electronically monitored curfew, or drug treatment. A problem that arises from the use of alternatives to pretrial detention are the potential for net-widening. Additionally, there may be a hidden agenda of punitive or coercive, and sometimes also rehabilitative or therapeutic, aims of alternative measures that go beyond the legitimate goals of ensuring that the suspect stands trial and does not continue offending. Further problems are the exclusion of foreigners or the poor from such alternatives. Other questions relate to the consequences of breach of conditions and how periods spent or efforts made under bail conditions may be credited in the sentencing decision.

Fundamentals on Pretrial Detention and Its Alternatives

Statistical Background

According to the World Pre-trial/Remand Imprisonment List (Walmsley 2008), two and a quarter million people were known to be held in pretrial detention and other forms of remand imprisonment throughout the world in 2008, and a further estimated 250,000 were held in countries on which such information was not available. At midyear 2010, about 460,000 persons were in pretrial detention in the USA (U.S. Bureau of Justice Statistics 2011b). In the 47 Member States of the Council of Europe, the number of remand detainees amounted to 370,000; 136,000 of them in the 27 Member States of the European Union (Council of Europe Annual Penal Statistics 2011, data for the year 2009). Globally, one out of every three detainees has not been found guilty of a crime (Open Society Justice Initiative 2011, on the Basis of the World Prison Population List, Walmsley 2007). In Europe, between 11 % (Poland, Germany) and almost 60 % (Turkey) of all prisoners are remand detainees (Morgenstern 2013). In the USA, the percentage was 28 % in 2010 (U.S. Bureau of Justice Statistics 2011a and 2011b), in Canada it was 37 % (Statistics Canada 2011, for 2009/2010). It is important to note that in countries heavily burdened with high prison population rates, remand imprisonment does not always contribute greatly to this burden. On the other hand, some countries seem to manage to contain their overall prison population but have a high number of remand detainees, so that their share is considerable. This is a problem not only for those affected but also for the prison authorities, who must deal with these “preliminary prisoners.” Some of the obvious differences are explained by different scopes and notions of remand detention (see below), but the data nevertheless show that, in many countries, the remand population impacts significantly on the overall prison population.

Statistical data on the use of alternatives to pretrial detention are often hard to obtain and even harder to compare. What can be seen, however, is that the use of precautionary measures during pretrial procedures varies greatly. In England and Wales, 9 % of all persons charged with an offense were remanded in custody in the year 2009. While it is not exactly known how many of the other 91 % were subject to bail conditions ("alternatives to pretrial detention" in the sense used here), research studies suggest that this was the case for over half of them (Hucklesby 2011; Hucklesby et al. 2007). In Ireland, bail is granted in about 25 % of all murder cases and 80 % or more in cases involving less serious offenses (Mellett 2012). In New Zealand, 13 % of all persons charged with an offense were remanded in custody and 87 % were remanded at large or on bail with or without conditions, most typically the latter (Young 2012, for the year 2008). In the USA, a study showed that, in the year 2004, more than 40 % of defendants who had state felony charges were detained in custody. Released defendants were most likely to be released on commercial surety or their own recognizance (Viano 2012). In Germany, in a sample of 1,000 recent convicts, 30–40 had been remanded in custody before. No statistics on alternatives exist, but studies suggest that, additionally, between four and ten of those 1,000 convicts had been under pretrial supervision in the community (depending, for example, on their age; Morgenstern 2009, with further references). In other states, remand detention still clearly outnumbers unconditional or conditional release. In China, for example, almost 80 % of all suspects were detained on remand, more than 16 % were released under supervision, and only 4 % were released without conditions (Zhang 2012).

Scope and Definition of Pretrial Detention

"Pretrial" and "remand" detention are used synonymously here. It must be recognized, however, that depending on the criminal justice system, different scopes and definitions of remand detention and consequently of its alternatives exist. Remand detention may include pretrial detention in a strict sense (police custody and remand custody ordered by a judicial authority), trial pre-conviction detention, and trial post-conviction detention. The latter may include the period between conviction and sentence and detention during an appeal or cassation procedure (Morgenstern 2013). In all these stages, alternatives may apply.

As indicated above, remand detention primarily serves procedural purposes, that is, to prevent a suspect from absconding or from tampering with evidence. The justice system needs him or her to be present and at its disposal – the etymology of the word *remand* (from the Latin *re*- and *mandare*, literally "order back") points to this original meaning. For this period of custody, the term "pretrial" detention is also often used, but it is obvious that remand detention may last longer as it may extend beyond the beginning of the trial. The legal status of remand prisoners is shaped by two factors: the first has to do with criminal procedure (i.e., the need for certain precautionary measures), while the second relates to the presumption of innocence, which implies that an unconvicted person must be treated as such and may not be restricted in his or her rights and liberties any further than absolutely necessary. Additionally, it implies that a remand prisoner must be treated differently from a sentenced prisoner. This latter principle plays a central role in the theoretical underpinning of remand detention as well as of the criminal justice system as a whole. The role and scope of the presumption are not entirely clear, not within certain jurisdictions and even less in a comparative perspective. They vary from a mere rule for the burden of proof to an overarching principle to guarantee procedural rights of the suspect or accused (Ashworth and Redmayne 2010; Stuckenberg 1998, with further references). One difference relates to the period during which the presumption has an effect on those detained during a criminal process. In most European countries, persons are presumed innocent until there is the last and final sentence and, therefore, are (or should be) kept in institutions other than ordinary prisons. In others, however, namely in England and Wales (Morgan 1994), this is not (necessarily) the case. Here, the presumption of innocence expires with the conviction of first instance. From a European perspective, both concepts seem acceptable as the European Court of Human Rights (ECtHR) explicitly acknowledged that the English practice meets the criteria of Article 6 (2) and Article 5 (1) of the European Convention of Human Rights (ECHR; Monell and Morris v United Kingdom, application no. 9562/81 and 9818/82, judgment of 2 March 1987 and Wemhoff v. Germany, application no. 2122/64, judgment of June 27, 1968, Series A No. 7). These conceptual differences explain some of the above-mentioned disparities in the size and proportion of the remand prison population and also have implications for the understanding of alternatives.

Avoiding Pretrial Detention: Concepts and Forms of Alternatives

The concepts and forms of avoiding remand detention are diverse and relate to different systems of imposing detention. Generally, however, judges all over the world find themselves in similar situations when they have to choose between three options for the suspect: The first option is unrestricted liberty (release on recognizance) and the sincere hope that the suspect will show up once the trial starts. The second is conditional bail or a similar concept with some safeguards put in place by imposing certain obligations on the suspect. The third option is remand custody. To avoid remand custody, two models may be distinguished. For example, the English (or other Common Law) along with the Polish and French practices can be described as choosing from a continuum – from no restrictions (unconditional bail) to more restrictions (conditional bail) to a full restriction of freedom (custody) (van Kalmthout et al. 2009a). The German model is different: the legal hurdle to order remand custody has to be taken (implying certain thresholds, excluding minor offences and a certain degree of risk of absconding or collusion). Only if this can be done may the judge choose less restrictions; that is, release the suspect or accused under certain conditions (conditional suspension of the arrest warrant). Here, noncustodial measures are meant solely as substitutes for detention and should be able to avoid net-widening effects. More often than not, however, they are merely used to shorten the term of remand custody and not as a substitute from the outset (Morgenstern 2009). Other examples for this substitution model are found in the Netherlands (van Kalmthout 2009) and Denmark (Rentzmann 2012).

Alternatives to pretrial detention may take various forms, usually with obligations imposed on the offender such as reporting to the police; depositing of passports or other identity papers; observing of orders relating to residence, work, spare time, association with certain persons; making oneself available for inquiries and reports; submitting to supervision (involving adherence to the instructions of the probation service); residing in a special hostel or other institution; undergoing psychiatric or addiction treatment; participating in a day training program; monetary bail; and curfew/house arrest with or without electronic monitoring (this is by no means an exhaustive list; examples are taken from country reports in van Kalmthout et al. (2009b) and van Kempen 2012). In these reports as well as in this essay, the concept of "bail" is frequently mentioned: it should be clarified that this term does not only cover property or money deposited or pledged to a court on the understanding that the suspect will return for trial or forfeit the bail. This sometimes is misunderstood when the word "bail" is translated (e.g., into the German word *Kaution*, *caution* in French or *fianza* in Spanish, which all exclusively refer to a financial surety). It rather must be understood in a much wider sense as a status in which the suspect is obliged to turn up before court or before the police (often with conditions attached to secure this aim). The measures listed above stand alone or are conditions of bail or of the suspension of detention that can be combined. An accumulation of several bail conditions may ultimately lead to severe restrictions of personal liberties with a serious risk of breach.

Further practical and ethical grounds exist for why unrestricted bail or alternative measures should be applied instead of remand custody. Usually, remand prisons are in far worse condition than ordinary prisons (Hucklesby 2002, for England/Wales; van Kalmthout and Knapen 2012, with reference to the work of the European

Committee for the Prevention of Torture, CPT and examples from all over Europe; Maruna et al. 2012, with examples from the USA). They are often subject, far more than regular prisons, to overcrowding (France and other countries, van Kalmthout et al. 2009b) and they often are not able to provide any meaningful activities for the inmates, leaving them "languishing for weeks, possibly months, locked up in their cells" (European Committee for the Prevention of Torture and Inhuman or Degrading Treatment or Punishment 2010).

The Use of (Conditional) Bail and Other Alternatives to Pretrial Detention

Grounds and Limitations

The legitimate aim of remand regulations, custodial or noncustodial, is securing the attendance of the defendant at the trial. Reasonable suspicion of the committal of an offense is an undisputed requirement. The traditional grounds to order pretrial detention that are applied all over the world are flight or risk of absconding, tampering with evidence, interfering with witnesses, or otherwise obstructing the course of justice (van Kalmthout et al. 2009a). Increasingly, however, preventive purposes play a role and are accepted as legitimate aims as well (Ashworth and Redmayne 2010, for England and Wales; Morgenstern 2009, for Germany; Maruna et al. 2012, for the USA). The latter is acknowledged also in the European Convention on Human Rights (ECHR) as long as the danger of re-offending is a concrete and palpable one. Art. 5 ECHR generally secures the right to liberty of everyone by restricting the grounds for detention: "[. . .]No one shall be deprived of his liberty save in the following cases and in accordance with a procedure prescribed by law." The normative justification of pretrial detention follows more specifically from Art. 5 (1) lit c ECHR, which allows "the lawful arrest or detention of a person effected for the purpose of bringing him before the competent legal authority on reasonable suspicion of having committed an offence or when it is reasonably considered necessary to prevent his committing an offence or fleeing after having done so."

This must be read in conjunction with Article 5 (3) of the Convention, which incorporates a number of essential guarantees in order to make deprivation of liberty an exception to the rule of liberty and to ensure that judicial supervision is in place. The European Court of Human Rights has, on several occasions, stated that pretrial detention shall be used only as a measure of last resort. This, on the one hand refers, to the shortest possible time spent in pretrial detention (e.g., Wemhoff v. Germany, application no. 2122/64, judgment of June 27, 1968, Series A No. 7) and, on the other hand, to the use of alternatives from the outset (e.g., Neumeister v. Austria, judgment of 27 June 1968, Series A No. 8; Jabłoński v. Poland, application no. 33492/96, judgment of 12 December 2000; see also van Kalmthout et al. 2009b, with further references). Art. 9 of the International Covenant on Civil and Political Rights (ICCPR) equally protects the liberty of every person against arbitrary detention without, however, specifying legitimate grounds for pretrial detention. Art. 9 (3) ICCPR states, somewhat more weakly than the ECHR, that "[i]t shall not be general the rule that persons awaiting trial shall be detained in custody, but release may be subject to guarantees to appear for trial, at any other stage of the judicial proceedings, and, should the occasion arise for execution of the judgment." In addition to that guarantee of personal liberty, and as mentioned already above, it has to be taken into account that "[e]veryone charged with a criminal offence shall have the right to be presumed innocent until proved guilty according to law.' (Art. 14 (2) ICCPR; Art. 6 (2) ECHR)."

Three consequences follow from these provisions with regard to alternatives to pretrial detention. Firstly, there is a presumption in favor of liberty, which must be interpreted as a right to bail, or more precisely a right to have bail considered in every individual case (see, for example, ECtHR, Caballero v. the United Kingdom, application no. 32819/96, judgment of 8 February 2000: The applicant was arrested for attempted rape and there had been a violation of Article 5

(3) ECHR because the court refused the applicant bail pursuant to a law that precluded – without exception – all persons charged with, or earlier convicted of, murder, manslaughter and rape from being granted bail. This prevented the courts from considering the particular circumstances of the case). This means, secondly, that pretrial supervision measures generally are restricted to the same aims and objectives as pretrial detention. Thirdly, the same human rights guarantees – safeguarding the right to liberty, the presumption of innocence, and the principle of minimum intervention – must be applied to supervision measures that, albeit less severely, restrict personal liberties and freedoms within the community. Equally, due process safeguards must be provided for, in particular the right to judicial review and a proportionate approach to breach of condition.

Further restrictions to impose pretrial detention and a larger scale of alternatives are provided in the area of juvenile justice (see Dünkel et al. 2011; in particular, the chapter of Dünkel, Dorenburg and Grzywa, 1747 ff), however, the practice does not always consider them sufficiently, partly because of a lack of resources in the juvenile welfare agencies at the community level.

Problems

Problems relating to the imposition and execution of pretrial detention are widely acknowledged. This is far less the case for its alternatives because they operate below what can be called a custody threshold and are, therefore (seemingly) less troublesome. First of all, practical problems relate to the availability of suitable and proportionate alternatives to supervise suspects in the community; either because of the lack of the necessary infrastructure and human resources (in particular, probation or other services able to supervise efficiently) or because legislation does not foresee enough noncustodial options (see country reports in van Kempen 2012, and the case studies in the volume compiled by the Open Society Justice Initiative 2008). Particular problems arise for suspects who are not citizens or residents of the state they are charged in. The risk to abscond is easily assumed and noncustodial options to supervise them are either not deemed sufficient or are not be applied for practical reasons such as language difficulties, imminent deportation, or the like (van Kalmthout et al. 2009a, with further references from the country reports in that volume).

Where a range of alternatives exist, it is often the neglect of human rights of a suspect, mentioned above, that may lead to serious problems. There are, for example, indications that bail conditions – unintended or intentionally – often serve coercive, therapeutic, or punitive purposes (Hucklesby 2011). The latter can be seen, for example, when judges, particularly for juveniles, want to start a meaningful program run by the probation service without losing time during the trial (Morgenstern 2009) or try to compensate for far-reaching possibilities to suspend prison sentences (Tange 2011). Whenever pretrial measures serve purposes other than securing the due course of justice, because they are designed as rehabilitative measures or therapy, it is important to note that they depend on the consent of the offender at that stage. Criminal Justice authorities may also intend a smooth move from unconvicted to convicted status for offenders supervised in the community that, even if well-meant, equally neglects that the person affected is still presumed innocent. Even if programs are intended to address social needs, negative impact can be seen when they do not have "a consistent exit strategy" (Hucklesby 2011, in an evaluation of the "Effective Bail Scheme" in England/Wales). First, these programs may ignore the fact that a significant number of those suspects involved are acquitted; here, social support instead of contact with the criminal justice system would be the only adequate way of assisting people. Second, even when they are not acquitted but receive another noncustodial sentence, they often may fall out of the scheme immediately, which then makes the support during the bail period meaningless.

Another issue connected to the use of alternatives to pretrial detention is their potential for net-widening. Although difficult to prove, some indications can be found for this matter. It is noteworthy that, in several of the countries that have very high rates of pretrial prisoners, for example, the USA and South Africa (van Kempen 2012; see also Morgenstern 2013, on different rates in Europe), alternatives – particularly conditional bail – are applied on a large scale, while they are hardly ever applied in other countries with moderate or low rates (e.g., Germany or Norway). In an illustrative country case study for Belgium, it is shown that the introduction of alternatives was not able to reduce the use of pretrial detention, albeit crime rates only increased to a very small degree. Instead, the number of conditionally suspended warrants now is almost as high as the number of arrest warrants (Jonckheere and Maes 2011). On the other hand, we find examples of countries that have a low rate of remand detainees making use of bail to a high degree (Ireland; see Mellett 2012), which shows that alternatives may be of great value when they are carefully monitored. This refers both to individual cases where it is particularly necessary to adapt the measure to the suspect and also to react adequately to breach of condition, and to the legal and practical provisions of a criminal justice system more generally.

Both aspects are equally important with regard to electronically monitored curfew as bail condition. Electronic monitoring in various forms of pretrial supervision is used in an increasing number of jurisdictions (e.g., Australia, the USA, New Zealand, England/Wales, Scotland, and Belgium) and tested in various others (e.g., Germany, France, and Portugal). It is generally acknowledged that it reduces the risk to public safety from offenders living in the community as it reduces the likelihood of individuals committing new offenses and acts as deterrence in relation to absconding from a curfew. On the other hand, electronic monitoring often is experienced as particularly intrusive and also impacts on other persons living in the same household and, therefore, may be regarded as disproportionate in many cases (Nellis et al. 2012; see also the entry on electronic monitoring in this encyclopedia).

A further unresolved question is how periods under conditional bail can be credited, particularly when electronic monitoring and other pretrial supervision measures are involved that significantly restrict the rights and liberties of the suspect. While at least in most European states periods spent in pretrial detention must be deducted from the prison sentence, usually a day for a day (van Kalmthout et al. 2009b), normally, no legal provisions exist for these periods under conditional bail. It is then left to the discretion of sentencing judges whether and how they are credited. This problem is aggravated by the fact that cases where suspects are supervised in the community procedures may take longer than in cases of pretrial detention because the courts feel less committed to ensure a speedy trial (Morgenstern 2009).

The risk of discriminating against foreigners has been mentioned above. Serious disadvantages for another major group of suspects can be seen where bail is mainly understood as providing financial surety (bail bonds). In the USA, where most states use a system of commercial bail bonding, many suspects cannot afford to post bail. This has been heavily criticized for several reasons, among them the far-reaching and insufficiently controlled power of the commercial bail bonding entities and their extended arms, the bounty hunters. Most prominently, however, it has been seen as clearly discriminating against the poor (see Maruna et al. 2012, for a recent account).

Future Developments

With relatively little attention paid by scholars, conditional bail and other pretrial supervision measures in some parts of the world have developed in a way that fits well with policy trends toward a harsh response to (alleged) criminals, a punitive turn and the “new penology” framework. Indicators are the accumulation of intrusive bail conditions, certain net-widening effects, privatization, and commercialization, often along with the neglect of the presumption of innocence and due process requirements (such as a review of the measure imposed, the monitoring of those implementing it, etc.). In other countries, however, both unconditional bail and pretrial supervision measures in the community obviously are widely used, having a beneficial impact on the remand practice and remand prison population.

In recent years, not only pretrial detention but also the aforementioned problems and the merits of its alternatives have increasingly aroused scholarly interest from a comparative perspective. Several volumes compiling case studies, thematic essays, and country reports (Open Justice Initiative 2008 and 2011; van Kalmthout et al. 2009b; van Kempen 2012) provide insight in current developments throughout the world. Nevertheless, sound criminological (empirical) research is still needed in many countries, for example, with regard to net-widening issues and more generally the decision-making process of the competent authorities, and also with regard to discriminatory effects and implementation practices.

In Europe, pretrial supervision more recently became a matter of discussion for pan-European criminal policy as the European Union adopted a Framework Decision in this regard. It seeks to avoid pretrial detention for EU citizens who are not resident in the state where they are charged with an offense (Council framework decision 2009/829/JHA of 23 October 2009 on the application, between Member States of the European Union, of the principle of mutual recognition to decisions on supervision measures as an alternative to provisional detention, OJ L294/20). By applying the principle of mutual recognition of judicial decisions – the decision in this case being the order to supervise the suspect – they shall be able to return to their home under certain conditions instead of being detained in the state where they are prosecuted. Several issues are discussed regarding this Framework Decision, namely whether there is a real need for it and whether it will have net-widening effects. Doubts are particularly expressed as to whether it can really function in the face of diverse legal and practical provisions as well as different legal traditions and cultures in the EU Member States. Also, the Council of Europe (2006) adopted a recommendation that contains safeguards for the rights of pretrial detainees and also seeks to encourage the use of alternatives to pretrial detention (Morgenstern 2013). Follow-up is needed for both initiatives; whereas the recommendation by the Council of Europe is soft law, the Framework Decision must be implemented by the Member States of the EU. Pretrial supervision, therefore, will stay (at least for now) on the European agenda.

Recommended Reading and References

Ashworth A, Redmayne M (2010) The Criminal Process, 4th edn. Oxford University Press, Oxford

Council of Europe (2006) Recommendation (2006) 13 of the Committee of Ministers concerning the use of remand in custody, the conditions in which it takes place and the provision of safeguards against abuse adopted by the Committee of Ministers on 27 September 2006 at the 974th meeting of the Ministers' Deputies

Council of Europe Annual Penal Statistics (2011) Survey 2009 by Aebi M, Delgrande N. Council of Europe PC-CP (2011) 3

Council of the European Union (2009) Council framework decision 2009/829/JHA of 23 October 2009 on the application, between Member States of the European Union, of the principle of mutual recognition to decisions on supervision measures as an alternative to provisional detention, Official Journal L294/20

Dünkel F, Grzywa J, Horsfield P, Pruin I (eds) (2011) Juvenile Justice Systems in Europe – Current Situation and Reform Developments, 2nd edn. Forum Verlag Godesberg, Mönchengladbach

European Committee for the Prevention of Torture and Inhuman or Degrading Treatment or Punishment (CPT) (2010) The CPT standards – 'substantive' sections of the CPT's General Reports, CPT/Inf/E (2002) 1, Rev. 2010 Council of Europe, Strasbourg

Hucklesby A (2002) Bail in criminal cases. In: McConville M, Wilson G (eds) Handbook of the criminal justice process. Open University Press, Buckingham

Hucklesby A (2011) Bail support schemes for adults. The Policy Press, Bristol

Hucklesby A, Eastwood C, Seddon T, Spriggs A (2007) The evaluation of restrictions on bail pilots: Final report, RDS-On-line report 06/07. Home Office, London

Jonckheere A, Maes E (2011) À la recherche de mesures juridiques susceptibles de réduire la détention préventive. In: Jonckheere A, Maes E (eds) La detention preventive et ses alternatives. Academia Press, Gent

Maruna S, Dabney D, Topalli V (2012) Putting a price on prisoner release: The history of bail and a possible future of parole. Punishm Soc 14:315

Mellett T (2012) Pre-trial detention in Ireland. In: van Kempen P (ed) Pre-trial Detention. Human Rights, criminal procedural law and penitentiary law, comparative law. Intersentia, Cambridge

Morgan R (1994) England/Wales. In: Dünkel F, Vagg J (eds) Waiting for Trial. International Perspectives on the Use of Pre-Trial Detention and the Rights and Living Conditions of Prisoners Waiting for Trial. Max-Planck-Institut für ausländisches und internationales Strafrecht, Freiburg

Morgenstern C (2009) Germany. In: van Kalmthout A, Knapen M, Morgenstern C (eds) Pre-trial Detention in Europe. Wolf Legal Publishers, Nijmegen

Morgenstern C (2013) Remand detention in Europe: Comparative and pan-European aspects as elements of a wider European penology. In: Daems T, Snacken S, van Zyl Smit D (eds) European Penology? Oñati International Series in Law and Society. Hart Publishing, Oxford

Nellis M, Beyens K, Kaminski D (2012) Electronically Monitored Punishment: International and Critical Perspectives. Routledge, Oxford

Open Society Justice Initiative (ed) (2008) Justice initiatives: pretrial detention. Open Society Foundations, New York

Open Society Justice Initiative (2011) The Socioeconomic Impact of Pretrial Detention. Open Society Foundations, New York

Rentzmann W (2012) Pre-trial detention in Denmark. In: van Kempen P (ed) Pre-trial Detention. Human Rights, criminal procedural law and penitentiary law, comparative law. Intersentia, Cambridge

Shaw M (2008) Reducing the excessive use of pretrial detention. In: Open Society Justice Initiative (ed.) Pretrial detention. Open Society Foundations, New York (online: www.soros.org/publications/justice-initiatives-pretrial-detention, last retrieved 10 Sept 2012)

Statistics Canada (2011): Trends in the use of remand in Canada. http://www.statcan.gc.ca/pub/85-002-x/index-eng.htm (last retrieved 10 Sept 2012)

Stuckenberg CF (1998) Untersuchungen zur Unschuldsvermutung. de Gruyter, Berlin

Tange C (2011) La détention préventive: pis-aller du système penal? In: Jonckheere A, Maes E (eds) La détention préventive et ses alternatives. Academia Press, Gent

U.S. Bureau of Justice Statistics (2011a) Prisoners in 2010. http://bjs.ojp.usdoj.gov/ (last retrieved 12 Sept 2012)

U.S. Bureau of Justice Statistics (2011b) Jail inmates at midyear 2010. http://bjs.ojp.usdoj.gov/(last retrieved 12 Sept 2012)

van Kalmthout A (2009) Netherlands. In: van Kalmthout A, Knapen M, Morgenstern C (eds) Pre-trial Detention in Europe. Wolf Legal Publishers, Nijmegen

van Kalmthout A, Knapen M (2012) The CPT and pre-trial detention in Europe. In: van Kempen P (ed) Pre-trial Detention. Human Rights, criminal procedural law and penitentiary law, comparative law. Intersentia, Cambridge

van Kalmthout A, Knapen M, Morgenstern C (2009a) Introductory summary. In: van Kalmthout A, Knapen M, Morgenstern C (eds) Pre-trial Detention in Europe. Wolf Legal Publishers, Nijmegen

van Kalmthout A, Knapen M, Morgenstern C (eds) (2009b) Pre-trial Detention in Europe. Wolf Legal Publishers, Nijmegen

van Kempen P (ed) (2012) Pre-trial Detention. Human Rights, criminal procedural law and penitentiary law, comparative law. Intersentia, Cambridge

Viano EC (2012) Pre-trial detention in the United States of America: a multiple problem crisis. In: van Kempen P (ed) Pre-trial Detention. Human Rights, criminal procedural law and penitentiary law, comparative law. Intersentia, Cambridge

Walmsley R (2007) World Prison Population List, 6th edn. International Centre forPrison Studies, London

Walmsley R (2008) World pre-trail/remand imprisonment list. International Centre for Prison Studies, London

Young W (2012) Pre-trial detention in New Zealand. In: van Kempen P (ed) Pre-trial Detention. Human Rights, criminal procedural law and penitentiary law, comparative law. Intersentia, Cambridge

Zhang X (2012) Pre-trial detention in Ireland. In: van Kempen P (ed) Pre-trial Detention. Human Rights, criminal procedural law and penitentiary law, comparative law. Intersentia, Cambridge

American Dream

▶ Institutional Anomie Theory

AML

▶ Money Laundering

Analytical Criminology

▶ Situational Action Theory

Anomie

▶ Institutional Anomie Theory
▶ Social Disorganization and Terrorism

Anomie and Crime

Jón Gunnar Bernburg
Faculty of the Social and Human Sciences,
University of Iceland, Reykjavík, Iceland

Synonyms

Anomie theory (strain theory)

Overview

For over a century, anomie theory has had a profound impact on the direction of sociological criminology. Originally emerging in classical social thought as an analytical tool to study how broadly defined social conditions influence normative regulation and rates of deviant behavior, anomie theory has been applied and extended in different directions, guiding macrolevel research on societal crime rates as well as microlevel research on individual differences in crime. This entry reviews the theoretical and empirical work on anomie and crime, starting with the classic theoretical propositions and then moving on to contemporary extensions and applications.

Classic Anomie Theory

Anomie is introduced as an analytical tool for studying the social causes of deviance in the writings of the French sociologist Émile Durkheim. Focusing on the structural bases for social organization, Durkheim ([1893] 1984) first uses the anomie concept in his classic work *The Division of Labor in Society* to refer to situations when society fails to provide the appropriate cultural regulatory codes – normative values and beliefs – that are necessary to regulate the cooperation of individuals in complex societies. "Equilibrium" between different parts of the social system, Durkheim (p. 301) writes, requires "predetermined rules of conduct" entailing "rights and duties" that are "obligatory" and applicable in "most common circumstances." In a state of anomie, that is, when normative values are weak, irrelevant, or absent, cooperation becomes unstable and rife with conflict.

In *Suicide*, Durkheim ([1897] 1951) extends anomie theory to study the consequences of social disruption for well-being and deviant behavior. Emphasizing that human aspirations lack biological limits, Durkheim argues that normative values are necessary to define and restrain the desires of individuals so that they can become achievable by the means available to them. "A regulative force must play the same role for moral needs which the organism plays for physical needs" (p. 248). Thus, Durkheim assumes that under "normal" conditions, social values are in place that convince most individuals that they are getting what they should and that they are not justified in wanting more. These include, for example, commonly accepted standards that members of different social classes can legitimately aspire to as well as rules specifying how they can legitimately attain their social positions (pp. 250–251). In a state of anomie, society fails to produce normative values to restrain the aspirations of many individuals who may experience limitless and unattainable aspirations, resulting in disappointment and frustration that increases suicide and deviant behavior (p. 293).

Durkheim goes on to discuss the social conditions causing anomie in modern society, all of which he regards as temporary by-products of modernization. First, the weakening of traditional institutions such as religion and marriage constitutes a fertile ground for anomie, since the main function of such institutions is to provide moral regulation of goals and behavior. Thus, for example, the institution of marriage regulates desire for love and sex, but a high divorce rate weakens this regulatory power, as it makes married individuals aware of other possibilities (Durkheim [1897] 1951, p. 271). Linked to the weakening of traditional institutions is the increasing dominance of the market economy in the society, which Durkheim sees as a major source of anomie. Increasingly unregulated by religion and other traditional institutions, the sphere of "trade and industry" promotes a cultural orientation of materialism that not only encourages overweening ambition but represents it "as a mark of moral distinction" (p. 257). For Durkheim, the disposition of excessive materialism and unlimited want emerging in the economic sphere extends to other parts of society, further undermining the moral authority of traditional institutions.

Second, Durkheim ([1893] 1984, pp. 310–322; [1897] 1951, pp. 250–251) provides a link between anomie and distributive justice, arguing that since unequal opportunity contradicts the cultural ideals justifying inequality in modern society, it may cause anomie. Again, cultural restraint on ambition is efficient insofar that it makes individuals believe that they are not justified in asking for more. Since equal opportunity is widely accepted as the just rule for the allocation of social status in modern society, social practices contradicting this principle – privilege, inheritance, and the like, which Durkheim views as remnants of traditional society – undermine people's sense of distributive justice. Thus, when inequality is rooted in unequal opportunity, many individuals see their social positions as unjustly "forced" on them, and they thus think that they deserve something better – another case of society's failure to place moral restraint on ambition.

Finally, Durkheim argues that abrupt changes may create a temporary state of anomie. Normative values change only slowly because they emerge only in repeated social interaction. Thus, slowly changing normative values will fail to limit aspirations if they do not correspond with a rapidly changing social reality. For example, in modern society, economic fluctuations routinely disrupt the connection between social reality and normative values. Thus, a sudden depression casts many individuals into worse conditions than they have internalized as just, while sudden growth raises prosperity so rapidly that existing standards can become irrelevant but "a new scale cannot be immediately improvised" ([1897] 1951, p. 253). In both cases, normative regulation of aspirations is temporarily weakened due to rapid change, resulting in widespread disappointment and frustration.

Merton's Anomie Theory

Durkheim's discussion of anomie has influenced criminological research, particularly through the extension of Robert K. Merton (1938). While Durkheim sees anomie as the consequence of social change, Merton uses the anomie concept to explain how stable features of contemporary societies influence criminal behavior. Merton focuses on a malintegration *within* and *between* two fundamental components of society: *cultural structure* and *social structure*. Cultural structure comprises two distinct sets of commonly held normative values governing the behavior of individuals: values regulating goals, and values regulating the means of achieving goals. Social structure by contrast refers to ongoing patterns of social relationships and positions. Merton assumes that in a well-integrated society, the cultural structure strikes a rough balance between its emphasis on goals and means, that is, various social mechanisms are in place that sanction individuals roughly equally for striving for normative goals and for using normative means to do so. Moreover, a well-integrated society comprises a social structure providing avenues for most individuals to achieve valued goals via normative means.

However, disjunctions can emerge among these elements: On the cultural level, a disjunction may emerge between the relative emphasis placed on goals versus means. Moreover, the social structure may not offer opportunities for many individuals to achieve normative goals using normative means. Merton focuses on the United States as a prime example of a society characterized by such disjunctions. On the one hand, its cultural structure overemphasizes goal achievement, placing a pressure on all individuals to strive for economic success but placing comparatively little pressure on individuals to use legitimate means to achieve such success. Thus, there are more social rewards for achieving economic success than for playing by the rules, creating an imbalance that often encourages individuals to bypass the legitimate means and “innovate,” that is, to use the most efficient means to achieve economic goals, including crime. Thus, support for the normative means is weak and only personal interest and fear of punishment are left to control behavior, which therefore increases crime and deviance.

For Merton, this cultural imbalance is manifested in the American Dream ideal that places an exaggerated value on the single goal of monetary success as a common goal for all citizens to strive for, regardless of their social background or position. The social rewards for economic success are high almost regardless of how it is attained, which leads to attenuation of the normative values regulating means, that is, to a situation of anomie, or normlessness.

On the other hand, Merton points out that the social structure of American society is characterized by economic and social inequality, and hence, the cultural mandate to strive for economic success creates strain for a large part of the population that in fact has limited access to legitimate means to achieve economic success. The reality of inequality thus creates a lack of legitimate avenues for many individuals to strive for the valued goals, creating a strain toward normlessness and crime. Individuals in the lower social strata have the most limited access to the legitimate means, and hence feel this strain most acutely. The commonly valued goal of economic success is most likely to undermine commitment to normative means among disadvantaged groups and individuals. But, in an open society, affluent members of the society provide an influential reference point for economic success, and thus, strain toward anomie and crime is also felt in the middle and even the upper middle classes where individuals may feel deprived relative to those who are more affluent than they are.

Contemporary Applications and Extensions

The statements of Durkheim and Merton have been applied and extended in several directions in crime causation research, both in macrolevel and microlevel research. First, theorists have elaborated and extended the propositions of Durkheim and Merton regarding the macrosocial conditions conducive to anomie, focusing in particular on inequality, materialism, and social change. Second, scholars have applied and revised anomie theory on the microlevel, focusing on the interrelations among socioeconomic status, goal-means disjunction, frustration, normlessness, and delinquency.

Macrolevel Applications

Inequality

Based on anomie theory, particularly the extension of Blau and Blau (1982), macrolevel research has examined the association between inequality and crime, particularly violent crime. This work has interpreted anomie theory to imply that extensive economic and social inequality may increase crime in democratic societies. Again Merton emphasizes that economic inequality blocks access to legitimate opportunities for a large part of the population, producing strain as well as feelings of injustice. Moreover, extending Durkheim’s idea that unequal opportunity may lead to anomie, Blau and Blau (1982) argue that excessive inequality in democratic societies may produce widespread normlessness, resentment, and frustration, especially if it is rooted in unequal opportunity.

Thus, anomie theory implies that the effect of economic inequality on crime should be more pronounced in cultures emphasizing economic goals and equal opportunity. In such contexts, excessive inequality is most likely to create normlessness and resentment, due to widespread upward social comparison and relative deprivation (Stack 1984). Few studies have tested this proposition, with the exception of Krahn et al. (1986) who finds support for a more pronounced effect of inequality on homicide rates in more democratic nations, and Stack (1984) who does not find support for such a pattern in his cross-national study of property crime. More research is needed to examine how democratization contextualizes the impact of inequality on crime rates.

Research has examined whether inequality rooted in unequal opportunity increases violent crime. Again, as Durkheim implies, social practices inconsistent with equal opportunity may create a sense of distributive injustice, hence undermining normative regulation and creating widespread resentment, frustration, and conflict in the society (Blau and Blau 1982; see Messner 1989). Blau and Blau (1982, p. 119) examined metropolitan areas in the United States and separated the statistical effects of economic inequality and black-white inequality on violent crime rates. Blau and Blau argued that while economic inequality should generally increase normlessness and frustration, racially based inequality should do so in particular, as socioeconomic inequalities associated with ascribed positions are condemned as illegitimate in democratic societies. In support, the study found that economic inequality and black-white inequality both influence violent crime rates. Messner (1989) has provided cross-national evidence on this point, showing that nations with a high level of economic discrimination against social groups (e.g., ethnic minorities) have higher homicide rates, net of economic inequality. Finally, Chamlin and Cochran (2006) studied homicide rates among 33 nations and found that a measure of citizens' perception of the illegitimacy of social stratification has a positive effect on the homicide rate, but the effect held only for highly modernized societies.

Materialism and Social Institutions

Messner and Rosenfeld's (2007) institutional anomie theory elaborates and extends the classic idea that excessive materialism breeds anomie. Institutional anomie theory situates anomie in the context of the broad institutional framework of society, seeing it resulting from (as well as exacerbating) an imbalance among the major social institutions: economy, polity, education, and family. These major social institutions entail different orientations and sometimes competing claims. The key argument is that when the institutional balance of power is tilted toward the economy, that is, when the values of the economy are given higher priority than noneconomic values, crime rates increase (American society is discussed as a case in point).

The reason is that the dominance of economic values undermines normative regulation (that is, increases anomie) as well as informal social control. First, the cultural dominance of the value-orientations of the market economy, namely, pursuit of self-interest, attraction to monetary rewards, and competition, breeds the type of cultural disjuncture that Merton sees in the American Dream, namely, the cultural overemphasis on economic goals coupled with relatively weak emphasis on normative means. Second, the dominance of the economy weakens the moral authority of noneconomic institutions. For example, individuals receive more social rewards, such as prestige, for performing economic roles than noneconomic roles, and they frequently feel a pressure to give priority to economic roles (e.g., to skip family dinner so that they can work overtime). The regulatory functions of noneconomic institutions are thus weakened, reducing their crucial function in providing informal social control through socialization and social sanctions.

In recent years, institutional anomie theory has inspired macrolevel research on whether crime rates are influenced by the relative strength of economic and noneconomic institutions. The theory has been tested in cross-national research by treating low welfare state spending as an indicator of economic dominance in the society. Consistent with the theory, such research finds that welfare state spending is associated with less violent crime (Messner and Rosenfeld 1997), and it reduces the positive effect of economic inequality on violent crime (Savolainen 2000).

Most tests of institutional anomie theory have examined crime rates across states and counties within a single society, mostly within the United States. Measuring the strength of noneconomic institutions with data on divorce, voter turnout, civic engagement, educational and welfare expenditure, and the like, this research generally finds lower crime rates in places characterized by stronger noneconomic institutions (Maume and Lee 2003). Moreover, strong noneconomic institutions appear to decrease the crime-enhancing effects of poverty (Chamlin and Cochran 1995).

Social Change

Durkheim's idea that social change can cause anomie has prompted cross-national research to examine the association between modernization, social change, and crime. For example, researchers have recently used the opportunity to test anomie theory by focusing on the crime-enhancing effect of the social transition from communism to capitalism in Russia (Pridemore et al. 2007), Eastern Europe (Zhao and Cao 2010), and China (Liu 2005).

For example, Pridemore et al. (2007) examined changes in crime rates in transitional Russia, noting that Russia in the 1990s and 2000s provides a unique large-scale natural experiment to test Durkheim's thesis. Following the dissolution of the Soviet Union in the beginning of the 1990s, the society was suddenly converted from communism into a market society, celebrating ideals of open competition and individual achievement. The sudden political and economic changes were not accompanied by comparable institutional changes, thus the "legal, political, regulatory, and social institutions necessary for a properly functioning market economy were and continue to be underdeveloped in the country" (p. 275). As Durkheim's theory would predict, Pridemore et al. find huge increases in rates of crime, suicide, and alcohol-related deaths in Russia during this period.

The Missing Element of Social Values

An important limitation associated with macrolevel research on anomie theory is the common failure to include direct measures of normative values. None of the aforementioned studies directly show that the effects of structural factors (inequality, economic dominance, social change) on crime are actually mediated by normative deregulation. One way to overcome this problem is to combine the use of survey data and official statistics, but such research has been rare. Baumer and Gustafson (2007) examined county-level crime rates in the United States, using aggregated survey responses to examine normative values in combination with social structural factors. Supporting anomie theory, the study found higher crime rates in counties characterized by the particular cultural disjuncture discussed by Merton, namely, strong average commitment to economic success *and* weak average commitment to legitimate means. Moreover, consistent with Messner and Rosenfeld, the tendency for goals-means disjuncture to increase crime was less pronounced in counties characterized by more family socializing and more welfare assistance. Finally, consistent with Merton, commitment to economic goals had a more pronounced effect on crime in counties where legitimate means are more limited and unequal (indicated by low levels of educational and economic attainment and high levels of inequality).

Multilevel Approaches

Baumer (2007) has argued that since anomie theory implies links among sociocultural elements, individual attitudes and experiences, and criminal behavior, a comprehensive test of it requires a multilevel research strategy connecting group-level and individual-level data. Multilevel tests of anomie theory have been rare, however, with the exception of a few recent studies. Zhao and Cao (2010) combined survey data and official data and examined the effect of national residence on the experience of normlessness. Normlessness was measured with survey items asking respondents what they felt about the legitimacy of several instrumental crime-related scenarios. The study found citizens of Eastern European countries, which at the time were undergoing rapid democratic transition, to exhibit more normlessness than citizens of other countries (also, see Bjarnason 2009). Finally, a multilevel study by Cao et al. (2009) found more normlessness in countries with the most rapid population growth in recent decades, again supporting the idea that swift social change is conducive to anomie.

Microlevel Applications

Although classic anomie theory defines anomie as a societal-level condition and emphasizes the effects of macrosocial conditions on crime rates, it has implications for microlevel research on individual engagement in crime. Merton's discussion of how a cultural emphasis on economic goals creates strain for individuals has been applied, extended, and revised as microlevel strain theory. Also, anomie theory has guided research on the role of value commitments (and lack thereof) in crime and delinquency.

Classic Strain and Subculture Formation

Early extensions of classic strain theory attempted to integrate it with subculture theories. Cohen (1955) argued that the emergence of delinquent subcultures among low class youths can be seen as a collective response, a type of subcultural adaptation, to structural strain. That is, strained youths collectively adapt to the perception of blocked opportunities for conventional success by developing an alternative status system revolving around the subversion of conventional norms. Cohen's idea of subcultural adaptation to strain is not easily studied with quantitative data, but the notion that youths adapt to blocked opportunities by collectively creating a status system that defies common norms is consistent with qualitative studies on troubled working class kids (Willis 1977).

Cloward (1959) provides a different link between anomie and subculture theory by emphasizing that access to illegitimate opportunities is unevenly distributed in the society just like access to legitimate opportunities. Criminal behavior often requires and is greatly enhanced by illegitimate opportunities, that is, exposure to criminal networks and groups that entail social learning mechanisms (techniques, attitudes, etc.) as well as access to social networks that allow individuals to profit from crime. Association with criminal groups may thus explain why some individuals adapt to structural strain by engaging in crime. In subsequent work, Cloward and Ohlin (1964) argued that low class youths are differently exposed to criminal opportunity structures in their neighborhoods, and hence adapt differently to strain. Some low class youths are raised in neighborhoods characterized by a well-developed illegitimate opportunity structure, a tradition of criminal behavior that integrates criminal activity and conventional business activity. These youths are exposed to opportunities to learn and use illegal means for their advantage, and hence are more likely to adapt to strain with utilitarian criminal behavior. Conversely, low class youths in neighborhoods with no legitimate or illegitimate opportunities tend to adapt to strain by forming subcultures of violence and vandalism.

Limited research has examined how illegitimate opportunities may specify the effect of strain on crime. Hoffman and Ireland (2004) examined whether the influence of strain on adolescent delinquency was contingent on school-level opportunity structure (as indicated by the school-levels of delinquency, delinquent values, school problems, and school quality). The study found that classic strain (measured as the disjunction between personal economic goals and perceived educational opportunities) was positively associated with delinquency, but it did not support Cloward and Ohlin as the effect of strain was not contingent on school-level opportunity structure.

Classic Strain and Individual Criminal Behavior

Most research on strain theory has focused more narrowly on the individual-level effect of strain on criminal behavior, usually examining juvenile delinquency. The research literature is characterized by different methods of operationalizing classic strain. A common strategy has been to assume that individuals are differently committed to the goal of economic success, and will therefore respond differently to blocked opportunities for success. Thus, having high success aspirations and low expectations for success should increase engagement in crime and delinquency more than, say, having low aspirations and low expectations. But, support for this hypothesis is weak and appears to depend on the way in which aspirations and expectations are operationalized. Thus, studies have found little support for the hypothesis that juveniles with high educational aspirations but low educational expectations are more delinquent than others (Liska 1971). But, as Farnworth and Leiber (1989) argue, examining the dysjunction between *economic aspirations* and *educational expectations* may be more consistent with Merton's emphasis on the strain stemming from the cultural emphasis on economic success. Their study and others (Hoffman and Ireland 2004) find that having both high economic goals and low educational expectations is associated with increased adolescent delinquency.

Other studies have operationalized classic strain by focusing on expectations of success, assuming that most individuals value the goal of economic success. This work generally finds a positive effect of low expectations on delinquency (Menard 1995), but the association may be spurious due to missing variables from social control and social learning theories (Burton et al. 1994).

Agnew et al. (1996) propose a different approach to test classic strain theory. Instead of focusing only on aspirations and expectations, Agnew et al. argue that classic strain may be directly operationalized by measuring the individual's subjective dissatisfaction with his or her monetary status. Agnew et al. found that low socioeconomic status is in fact positively associated with dissatisfaction with monetary status. Furthermore, net of socioeconomic status, both high economic goals and low expectations about future economic success, was positively associated with monetary dissatisfaction, net of socioeconomic status. In turn, dissatisfaction with monetary status was associated with income-generating crime and drug use. Finally, the effect of dissatisfaction on crime was stronger among those having criminal friends and those having beliefs conducive to crime, lending support for the role of illegitimate opportunity in enhancing the effect of classic strain on delinquency.

Menard (1995) has tested classic strain theory by including a measure of normlessness, or "anomia," which is measured with questions about the individual's expectancy that socially unapproved behaviors are necessary to achieve one's goals. Among other findings, Menard found low socioeconomic status to positively influence adolescent anomia, in part through the negative effects of socioeconomic status on educational performance and expectations of success. In turn, anomia was associated with delinquency. Menard's findings support classic strain theory, indicating that limited access to legitimate means is conducive to anomia and involvement in crime.

Finally, researchers have operationalized classic strain by focusing on relative deprivation as a key social-psychological mechanism through which the cultural-structural discrepancies, described by Merton, produce anger, normlessness, and crime. As Passas (1997) has argued, in societies characterized by egalitarianism and economic materialism, reference group comparisons tend to be focused upward, with many individuals in the lower and middle classes experiencing anger and injustice due to unfavorable comparisons to more affluent members of the society. Guided by this view, researchers have found that measures of subjective relative deprivation, that is, the individual's frustration about other people being more economically successful than they are, positively influences delinquency, regardless of socioeconomic position (Agnew et al. 1996; Baron 2004; Stiles et al. 2000). In a multilevel study of adolescents, Bernburg et al. (2009) argued that economic deprivation should have a more pronounced positive effect on adolescent anger, normlessness, delinquency, and violent behavior in affluent neighborhoods (where comparisons of poor youths to other residents are highly unfavorable) than in impoverished affluent neighborhoods (where comparisons of poor youths to other residents are more favorable). The study found evidence of such a pattern, thus supporting the criminogenic role of relative deprivation.

General Strain Theory

In recent years, the revisionary work of Robert Agnew has led research on the strain-crime link to move away from anomie/classic strain theory to focusing more generally on the criminogenic effects of all categories of social strain. Agnew's (1992) general strain theory broadens the focus of strain theory at the social-psychological level and attempts to incorporate all major types of potentially criminogenic social strains. Specifically, the theory focuses on three major categories of social strain. First, building on classic strain theory, it emphasizes that strain may stem from actual or anticipated failure to achieve positively valued goals. This type of strain includes the failure to achieve culturally valued economic goals (classic strain), but it also includes strain stemming from a failure to achieve various other valued goals, such as being popular or getting good grades. Second, actual or anticipated removal of positively valued stimuli may cause strain, for example, loss of a parent or a friend. The third type of strain is the presence of noxious stimuli, such as being the victim of violence or sexual abuse.

Each of these strains may trigger negative emotions, including disappointment, fear, and anger. The theory focuses on anger in particular. Anger presumably increases the individual's level of felt injury, creates a desire for retaliation, energizes the individual for action, and lowers inhibitions, and hence increases the likelihood of delinquent behavior. In short, the theory argues that the three major types of social strains increase the likelihood of delinquency because strain often triggers anger and frustration. Finally, the theory argues that various social and personal factors are likely to determine whether strain leads individuals to engage in delinquency, for example, value commitments, social bonding, and involvement in delinquent associations or networks.

Due to the social-psychological focus of general strain theory, research on it rarely attempts to link strain with macrosociological issues, but its broader view of social strain has added to our understanding of the role of social strain in delinquency. Studies show that various types of social strain increase juvenile delinquency, including negative life events, conflict with parents, parental fighting, and that such effects are conditioned by factors such as association with delinquent peers and having values that are conducive to delinquency (Baron 2004; Brezina 1996).

Microanomie

While classic anomie theory implies that strain and frustration plays a role in crime and delinquency, it also implicates personal commitment (or lack of commitment) to social values as an important factor in deviance (see Bernard 1987). Again anomie refers to situations where social values do not provide sufficient control of the thoughts and behaviors of individuals. Nevertheless, microlevel applications and extensions of anomie theory have not focused much on commitment to social values (Bernard 1987).

But there are noteworthy exceptions. Hagan et al. (1998) argue that the core values of the market economy, namely, the emphasis on competition and individual achievement, can promote an excessive emphasis on individual self-interest that in turn is conducive to anomic amorality (lack of personal commitment to normative means) and delinquency. Building on this idea, Konty (2005) has termed the concept of "microanomie." Konty argues that individuals may be imbalanced in terms of their commitment to values that promote social-interest (self-transcendent values) and values that promote self-interest (self-enhancing values). When individuals are more strongly committed to self-enhancing values than to self-transcending values, they are in a state of microanomie. Microanomie is criminogenic in that it encourages individuals to pursue self-interested goals by using any means necessary. Konty implies that microanomie should be more prevalent in societies characterized by the dominance of the economy in the institutional balance of power, as Messner and Rosenfeld's institutional anomie theory suggests. Although providing no evidence for this macro–micro link, Konty's study confirms a positive association between microanomie and delinquency.

It may be noted that research based on social control theories has also attempted to measure individual normlessness. But in contrast to anomie/strain theory that sees normlessness resulting from blocked goals, social control theory argues that weak social bonds undermine commitment to normative values because they constitute channels for communicating and reinforcing normative values as well as entailing social sanctions for norm violations (Bernburg and Thorlindsson 2007). Research has underscored the role of weak social bonds and neighborhood networks in adolescent normlessness and delinquency (Bernburg and Thorlindsson 2007). Also, neighborhood research has relied on the normlessness concept arguing that the concentration of disadvantage in the neighborhood can lead to the "attenuation" of normative values. As Warner (2003, p. 76) has argued, disadvantaged residents are less likely to "live out and thereby reinforce within their communities many of society's common values," rendering cultural values less present in their everyday lives. Attenuation of culture decreases residents' willingness to exert informal control of places and property, due to mistrust stemming from the perception of low normative commitment of others.

White-Collar and Business Crime

Criminologists have yet to make full use of anomie theory in studying white-collar and business crime, but this perspective appears to be well suited for studying these topics. Building on Durkheim's idea that modern economic life promotes an exaggerated materialism leading to limitless monetary aspirations that breed anomie, anomie scholars have emphasized that strain toward anomie and crime exists in all social strata (Passas 1990, 1997). Passas (1990) has argued that in capitalist societies, egalitarian values and materialism jointly promote upward social comparisons and feelings of relative deprivation in all social classes: "the meaning and content of success goals vary from one part of the social structure to another, similar difficulties in attaining diversely defined goals may be faced by people in the upper social reaches too; they are, therefore, far from immune to pressures toward deviance" (p. 159). Indeed, the cultural emphasis on economic success is particularly salient in the corporate world. Scholars have suggested that the overemphasis on economic goals may have important effects on the ethical climate within business organizations, for example, by promoting organizational cultures that rationalize and neutralize crime and unethical conduct.

There is limited research that has directly applied anomie theory to white-collar, corporate, and business crime, but noteworthy exceptions (e.g., Vaughan 1997) indicate the usefulness of this approach. Criminological research needs to move beyond its usual focus on delinquency and street crime in order to examine the full range of the explanatory power of anomie theory. Perhaps events such as the global financial crisis of 2008 will prompt criminologists to devote more attention to developing research strategies to study these issues.

Conclusion

Anomie theory brings attention to the ways in which macrosocial forces influence the power of social values and norms in constraining the behavior of individuals, thus highlighting macrolevel as well as microlevel mechanisms in criminal behavior. Consistent with anomie theory, macrolevel research finds support for the role of social and economic inequality, economic dominance, and abrupt social change in societal crime rates. However, this research is limited in that it has usually failed to examine the role of social values (or lack of values) in mediating the effects of these features on crime rates. Recent attempts to use survey data to measure social values are promising, and will hopefully prompt future macrolevel research on anomie and crime to incorporate measures of social values.

Anomie theory has inspired different methodological and theoretical strategies to capture the individual-level manifestation of anomie. Scholars have focused on the role of opportunities and subculture in specifying how individuals adapt to blocked opportunities, and they have argued that anomie results in individual strain that can be measured by focusing on goals-means dysjunction, perception of limited life chances, relative deprivation, and subjective dissatisfaction with monetary status. Finally, researchers have examined how lacking normative restraint – anomia, normlessness, microanomie – influences criminal behavior. The variety of approaches at the individual-level bears witness to the fact that the broad social forces emphasized by anomie theory can be thought to impact crime through different social and psychological mechanisms. As each of these methods provides at least some support for the validity of these mechanisms, strain and anomie theories will continue to guide research on individual differences in crime. More research is needed on how broad social forces influence these microlevel mechanisms. Multilevel research linking the macro- and microlevels is likely to be especially valuable in providing full tests of anomie theory.

Related Entries

- ▶ General Strain Theory
- ▶ History of Criminological Theories: Causes of Crime
- ▶ Poverty, Inequality, and Area Differences in Crime

Recommended Reading and References

Agnew R (1992) Foundation for a general strain theory of crime and delinquency. Criminology 30:47–87

Agnew R, Cullen FT, Burton VS, Evans TD, Dunaway RG (1996) A new test of classic strain theory. Justice Q 13:681–704

Baron SW (2004) General strain, street youth and crime: a test of Agnew's revised theory. Criminology 42:457–483

Baumer EP (2007) Untangling research puzzles in Merton's multilevel anomie theory. Theor Criminol 11:63–93

Baumer EP, Gustafson R (2007) Social organization and instrumental crime: assessing the empirical validity of classic and contemporary anomie theories. Criminology 45:617–647

Bernard TJ (1987) Testing structural strain theories. J Res Crime Delinq 24:262–280

Bernburg JG, Thorlindsson T (2007) Community structure and adolescent delinquency in Iceland: a contextual analysis. Criminology 45:415–444

Bernburg JG, Thorlindsson T, Sigfusdottir ID (2009) Relative deprivation and adolescent outcomes in Iceland: a multilevel test. Soc Forces 87:1223–1250

Bjarnason T (2009) Anomie among European adolescents: conceptual and empirical clarification of a multilevel sociological concept. Sociol Forum 24:135–161

Blau JR, Blau PM (1982) The cost of inequality: metropolitan structure and violent crime. Am Sociol Rev 47:114–129

Brezina T (1996) Adapting to strain: an examination of delinquent coping responses. Criminology 34: 39–60

Burton VS, Cullen FT, Evans TD, Dunaway GR (1994) Reconsidering strain theory: operationalization, rival theories, and adult criminality. J Quant Criminol 10:213–239

Cao L, Zhao R, Ren L, Zhao JS (2009) Social support and anomie: a multilevel analysis of anomie in Europe and North America. Int J Offender Ther Comp Criminol 54:625–639

Chamlin MB, Cochran JK (1995) Assessing Messner and Rosenfeld's institutional anomie theory: a partial test. Criminology 33:411–429

Chamlin MB, Cochran JK (2006) Economic inequality, legitimacy, and cross-national homicide rates. Homicide Stud 10:231–252

Cloward RA (1959) Illegitimate means, anomie, and deviant behavior. Am Sociol Rev 24:164–176

Cloward RA, Ohlin L (1964) Delinquency and opportunity. Free Press, New York

Cohen AK (1955) Delinquent boys: the culture of the gang. Free Press, Glencoe

Durkheim É ([1893] 1984) The division of labor in society. The Free Press, New York

Durkheim É ([1897] 1951) Suicide: a study in sociology. The Free Press, New York

Farnworth M, Leiber MJ (1989) Strain theory revisited: economic goals, educational means, and delinquency. Am Sociol Rev 54:263–274

Hagan J, Hefler G, Classen G, Bohenke K, Merkens H (1998) Subterranean sources of subcultural delinquency: beyond the American Dream. Criminology 36:309–342

Hoffman JP, Ireland T (2004) Strain and opportunity structures. J Quant Criminol 20:263–292

Konty M (2005) Microanomie: the cognitive foundations of the relationship between anomie and deviance. Criminology 43:107–131

Krahn H, Hartnagel TF, Gartrett JW (1986) Income inequaliity and homicide rates: cross-national data and criminological theories. Criminology 24: 269–295

Liska AE (1971) Aspirations, expectations, and delinquency: Stress and additive models. Sociological Quarterly 12:99–107

Liu J (2005) Crime patterns during the market transition in China. Brit J Criminol 45:613–633

Maume MO, Lee MR (2003) Social institutions and violence: a sub-national test of institutional anomie theory. Criminology 41:1137–1172

Menard S (1995) A developmental test of mertonian anomie theory. J Res Crime Delinq 32:136–174

Merton RK (1938) Social structure and anomie. Am Sociol Rev 3:572–682

Messner SF (1989) Economic discrimination and societal homicide rates: further evidence on the cost of inequality. Am Sociol Rev 54:597–611

Messner SF, Rosenfeld R (1997) Political restraint of the market and levels of criminal homicide: a cross national application of institutional-anomie theory. Soc Forces 75:1393–1416

Messner SF, Rosenfeld R (2007) Crime and the American Dream, 4th edn. Wadsworth, Belmont

Passas N (1990) Anomie and corporate deviance. Contemp Crises 14:157–178

Passas N (1997) Anomie, reference groups, and relative deprivation. In: Passas N, Agnew R (eds) The future of anomie theory. Northeastern University Press, Boston, pp 62–94
Pridemore WA, Chamlin MB, Cochran JK (2007) An interrupted time-series analysis of Durkheim's social deregulation thesis: the case of the Russian Federation. Justice Q 24:271–290
Savolainen J (2000) Inequality, welfare state, and homicide: further support for the Institutional Anomie Theory. Criminology 38:1021–1042
Stack S (1984) Income inequality and property crime. Criminology 22:229–257
Stiles BL, Liu X, Kaplan HB (2000) Relative deprivation and deviant adaptations: the mediating effects of negative self-feelings. J Res Crime Delinq 37:64–90
Vaughan D (1997) Anomie theory and organizations: culture and the normalization of deviance at NASA. In: Passas N, Agnew R (eds) The future of anomie theory. Northeastern University Press, Boston, pp 95–123
Warner BD (2003) The role of attenuated culture in social disorganization theory. Criminology 41:73–97
Willis P (1977) Learning to labour: how working class kids get working class jobs. Saxon House, Farnborough
Zhao R, Cao L (2010) Social change and anomie: a cross-national study. Soc Forces 88:1209–1230

Anomie Theory (Strain Theory)

▶ Anomie and Crime

Antisocial Personality Disorder

▶ Psychopathy and Offending

Antisocial Potential

▶ Integrated Cognitive Antisocial Potential Theory

Anxieties About Crime

▶ Fear of Crime and the Psychology of Risk

Applied Geographical Profiling

Alasdair M. Goodwill[1], Jasper J. van der Kemp[2] and Jan M. Winter[3]
[1]Department of Psychology, Ryerson University, Toronto, BC, Canada
[2]Department of Criminal Law and Criminology, VU University Amsterdam, Amsterdam, The Netherlands
[3]National Police Services Agency, Zoetermeer, The Netherlands

Synonyms

Geo-forensic Analysis; Geographical Offender Profiling; Journey to Crime

Overview

Geographical offender profiling (GP) is the practical application of various geographical, criminological, and psychological principles to typically estimate the most likely area of an offender's home base (e.g., the offender's residence) or anchor point (e.g., place of work or frequent activity) based on the location of their criminal activity (e.g., crime sites) (Rossmo 2000). The output of a GP analysis can inform ongoing investigative strategy in several ways (Rossmo 2000): suspect prioritization, patrol saturation, neighborhood canvasses, police information systems, and DNA searches, to name but a few (for further strategies used, see Knabe-Nicol and Alison 2011). This entry aims to introduce the reader to the environmental criminological and psychological basis of geographical profiling, discuss state-of-the-art journey to crime (JTC) research, and advocate the further integration of routine activity theory into a geographical offender profiling framework.

Although GP is not firmly grounded in empirically tested theories (Levine 2005), it does utilize several theoretical approaches from the broader framework of environmental criminology: namely *Crime Pattern Theory* (Brantingham and Brantingham 1984), the *Routine Activity Approach* (Cohen and Felson 1979), and the *Rational Choice Perspective* (Cornish and Clarke 1986). A number of researchers have (partially) borrowed from these three theories to develop mathematical and computerized models to (typically) predict the location of an offender's home (see Canter and Youngs 2008a, for review). However, to date, very few studies (see section " Journey to Crime (JTC)" for a review) have tested theories and/or provided empirical explanations integrating the routine activity of offenders or the socio-geographical features of the areas surrounding crime locations (e.g., topography, transportation routes, social nodes, etc.).

It is perhaps an obvious fact that to commit an offense, an offender must, in all but the rarest of cases, be at least for some time period in the area of the crime location. Even so, an offender may still live far away from that area and/or they may only be in that area for a short period of time. Nevertheless, the offender *must* be in that area at some point and thus the area *must* form part of geographic area they (now) know of – their *activity* space. Further, several researchers (e.g., Canter and Larkin 1993) have suggested that offenders, just like any other individual, will *routinely* visit the same areas (e.g., activity space) and are less likely to venture into unknown areas of equal opportunities. This suggests that offenders not only have areas of *activity* but also areas of *routine activity* – places or area they routinely travel to or through. According to routine activity theory (discussed in detail in section " Routine Activity Approach" below), crimes (and their associated geographical locations) occur during this routine, normal activity of everyday life. As such, offense locations may give insight into that particular offender's *routine activity* – which may give rise to their identification (Canter and Gregory 1994).

Thus, it is vitally important to know who is routinely using the area in which a crime is committed and for what reason because the offender will theoretically be among this, albeit potentially large, list of individuals. In spite of this, the majority of GP approaches and computerized systems do not prioritize lists of people known to use the area around a crime location but instead typically prioritize areas to investigate, which in turn, may generate a list of actual suspects. However, even in cases where a GP prediction of an area is so precise that it pinpoints a single dwelling (e.g., a GPS coordinate or prioritized area), the *identity* of the offender is still not predicted, only that dwelling is the area most "associated" with the crime locations (e.g., may indicate a psychological anchor point or place of importance to the offender, such as their home). Arguably, this can still be vitally important to an investigation if accurate, but ultimately, incorporating information on who (e.g., individual identities) routinely uses that space (among other types of information) may elicit a more comprehensive view of the crime location area, potentially identifying the actual perpetrator.

In the latter half of this entry, the current authors call for a greater integration of routine activity theory in GP research and theory, highlighting the current practices of geo-profilers and future directions of geographical profiling. However, first, the underlying theoretical aspects of GP, the state-of-the-art research in JTC analyses, and the prominent debates within the literature will be highlighted. Our aim is to state the importance of moving away from a largely mathematical and computational approach focused on predicting an offender's home base (or anchor points) using *only* x/y coordinates to a more substantive and comprehensive psychological sociological and criminological GP model.

Describing Offender Movements: Environmental Criminological Theory

There has been a long history of describing the geographical or spatial movements of offenders (see Canter and Youngs 2008b for review). Research has developed largely from a criminological perspective, describing the movement in relation to differences in offender characteristics, victim types, and cultures and most commonly in relation to sex offending (e.g., Rossmo 2009, for review). The basis of the environmental criminological perspective will now be discussed.

Crime Pattern Theory and Awareness Space

In their seminal *Crime pattern theory*, Brantingham and Brantingham (1984) consider how offenders move about in time and space. Building on the work of urban planners, they suggest that the spatial occurrence of crimes is not random, but directly related to the immediate physical and environmental circumstances. Brantingham and Brantingham (1984) propose that offending has a rational basis as offenders move predominantly within familiar spaces in which they identify potential targets. Therefore, each person's own *activity* space is confined by their *awareness* space, which is reflected in each individual's own "*mental map*." The latter is a unique cognitive representation that explains and defines each individual's *awareness* space, where the *activity* space represents the *habitual* area that the offender uses for the majority of his non-criminal activity. However, the nonuniform distribution of targets (e.g., victims) causes criminals to offend in a subset of their *awareness* space, termed the *opportunity* space. *Activity* space and *opportunity* space will most likely overlap completely (Brantingham and Brantingham 1984). Hence, the *opportunity* space (e.g., criminal area) of an offender can also reveal the non-criminal *activity* space (e.g., routine activity space) of the same individual.

Routine Activity Approach

Routine activities approach posits that offender and victims (or targets) usually meet during daily, non-criminal activities or routines (Cohen and Felson 1979). In other words, crime originates in the context of the normal everyday routines as three spatiotemporal factors converge: a *motivated offender*, a victim or *potential target*, and the *absence of a capable guardian* (Cohen and Felson 1979). This is important for practical purposes since it suggests that offenses committed in a rather spontaneous fashion are likely to be in an area familiar or habitual to the offender. Thus, it may be hypothesized that a rapist that attacked his victim in an impulsive or opportunistic way was in the area for non-criminal reasons (e.g., on his way home from his local bar). In other words, the area is likely a part of the offender's routine activity space and, as will be discussed in later sections of this entry, there is possible information (e.g., eyewitness, archival records, parking tickets, etc.) relating this particular offender to that particular activity space, potentially leading to his identification.

Rational Choice Perspective

The *rational choice perspective* suggests that crime is committed in terms of a multi-level, goal-oriented, purposeful, and crime-specific decision-making process, especially concerned with the analysis of costs versus benefits within a constrained situation (Cornish and Clarke 1986). However, determining the proportion of offenders using an implicitly unconscious cognitive or consciously reasoned cost-benefit analysis to commit crime is difficult. On the other hand, the concept of "satisficing" could account for a rather intuitive approach of offender decision making (Canter and Shalev 2000). *Satisficing*, derived from "satisfy" and "suffice," refers to the acquirement of just enough information to make an acceptable but non-optimal decision. This is also supported by recent research investigating the heuristics burglars use to determine whether or not a house is occupied (Snook et al. 2010). Regardless of the underlying decision process, Cohen and Felson (1979) propose that the rational choice perspective better accounts for the *content* of the decisions, while routine activities approach better accounts for *why* decisions are made.

Theory Summation

These three general environmental criminological theories converge in the following way: Offenders are relatively more compelled to commit an offense (than not) if the situation is attractive (*routine activity approach*), if it occurs in a familiar environment (*crime pattern theory*) and yields a desired reward (*rational choice perspective*). Such an environmental criminology perspective emphasizes the complex interplay of the offender motivation, situational factors, geographical location, and the victim or target in the genesis of a criminal event.

Geographical Offender Profiling (GP) Theory

In order to understand, and ultimately predict, the geographical mobility of offenders, one must take certain psychological principles into consideration, hence the emergence of the application of psychological principles to geo- and environmental criminology, namely, *geographical offender profiling*.

The Circle Hypothesis: Commuters and Marauders

It has been postulated that an offender's home is a psychological anchor point that influences their geo-spatial movements and criminal activity (Canter et al. 2000; Gabor and Gottheil 1984; Rossmo 2000). In fact, over 80 % of the homes of serial rapists (Canter and Larkin 1993) and arsonists (Kocsis and Irwin 1997) were located in the center of a circle whose diameter was defined by the two most distant crime locations in their series.

This phenomenon, now known as the "circle hypothesis" (Canter and Larkin 1993), led to the bifurcation of offenders into *marauders* and *commuters*. The former live somewhere within the area circumscribed by their offenses, with their home base as the nexus of their criminal activity (e.g., they offend outwardly from the center of the circle). Conversely, commuters live outside the circle, which may be due to more optimal offending opportunities elsewhere from the area around their home. However, the ratio of marauders to commuters found initially, roughly 80–20 % respectively, has not been consistently observed. For instance, looking at international studies, the percentage of serial raping marauders varies from 60 % to 71 % in Germany (Janke and Henningsen, 1995 in Mokros and Schinke 2006 and Snook et al. 2005a, respectively), 71–93 % in Australia (Kocsis and Irwin 1997 and Meaney 2004, respectively), while research into serial property offenders has found an approximately 50/50 split between commuters and marauders (Van der Kemp et al. 2005).

Even considering a liberal proportion of commuters (20 %), one in five offenders typically live outside the area circumscribed by their offenses, making any GP predictions of where the offender lives based on the circle hypothesis alone inherently flawed. Therefore, before embarking on GP analysis, Rossmo (2000) suggests that it is imperative that the geo-profiler establishes whether the offender is a commuter or marauder (or "*poacher*" or "*local hunter*" in Rossmo terms, respectively). Although Rossmo does not elucidate on how this distinction is to be accomplished (Van der Kemp et al. 2005), some researchers have suggested a framework to predict whether an offender is a commuter or marauder (Meaney 2004), albeit with limited success (Paulsen 2007).

Thus, it seems that although utilizing the circle-hypothesis principle to predict the home location of an offender is unlikely to be *enough* to produce accurate results in all cases, it is not without its theoretical wisdom. Conceptually, the circumscribed area is identical to that of Brantingham and Brantingham's (1984) criminal space, a subset of the offender's overall routine activity space. Therefore, the circle hypothesis may be a useful "rule-of-thumb" to begin conceptualizing the routine activity space of the unknown offender – the area the offender may commonly frequent in their everyday normal life as well. Accordingly, one can begin to investigate what features that space has, how that space was used in the criminal activity, and importantly, why the offender chose that particular space in relation to their criminal activity. A careful examination of the nature of the space and its relationship to the routine activity space of the individuals within it may help to develop a richer geo-profile of the crime or crime series. In a routine activities approach to GP, what topographic features, transport routes, sociogeographical aspects, etc., of that area are likely influencing the offender and how does that inform predictions or prioritization of the offender?

Geo-spatial Decision Making: Distance-Decay Theory

The *least-effort principle* (Zipf 1965) offers a very plausible explanation of an offender's decision-making process in selecting criminal activity targets by postulating that an individual faced with alternatives will most likely choose the easiest course of action. This appears to be supported by a wealth of research on *distance-decay theory* (e.g., Snook et al. 2004) which states that the probability of offending decreases with increasing distance from the offender's home. The generalizability of distance-decay as a function of various types of crimes and individual offenders has been both supported (Rengert et al. 1999) and contested (Townsley and Sidebottom 2010; Van Koppen and De Keijser 1997). For example, first Van Koppen and De Keijser (1997) and later Townsley and Sidebottom (2010) demonstrated that the phenomenon of distance-decay may be due to the *aggregation* of the individual variations in distances traveled by offenders, not by the individual distances themselves. However, a recent study by Bichler, Christie-Merrall, and Sechrest (2011) supports the robust findings of distance-decay advocated by Rengert et al. (1999), suggesting that the *ecological fallacy* argument in distance-decay research is perhaps unwarranted.

Nevertheless, on an individual level, the least-effort principle, routine activity approach, and rational choice perspective offer explanations of offender movements and geographically related offending decisions. Pragmatically, it takes time, resources, and effort to overcome distance. Thus, close targets and locations – especially those involved in daily routines – are preferable to those further away if the subjective benefit remains the same. Additionally, it is more challenging to function effectively in an environment one knows little about which makes familiar areas more attractive (Bernasco 2007). The inherent nonuniformity in target (or area) attractiveness and familiarity in offenders' routine activity spaces offers rational choices to be made – *if (or when) I offend where should I do it* and *how close is too close to home*?

Accordingly, distance-decay is thought to be modified by two additional rational choice perspectives: *specialist selectivity* and *buffer zones*. Pettiway (1982) found that specialist selectivity is best reflected in the amount of effort that an offender puts into planning and selecting a crime. Criminals that were more specialized and selective in their choice of targets did not follow normal distance-decay patterns, but offended further away from home. This appears to apply interrelatedly across crime types as well, with the more impulsive crime of rape occurring closer to home than carefully planned armed robberies (Pyle 1974). Furthermore, the risk to the offender of being identified may result in *less* criminal activity around the offender's home.

This area, of relatively higher risk to the offender of identification, the so-called *buffer zone* (Brantingham and Brantingham 1984), exists as a sort of bulwark between the offender's *activity* space (e.g., home) and their criminal *opportunity* space. Research findings are mixed with regard to the existence of a buffer zone, some studies show confirming evidence (e.g., Canter and Larkin 1993; Dern et al. 2005; Rossmo et al. 2004) while other studies refute it (e.g., Levine 2005). However, it seems plausible that a buffer zone may exist and that it may be closely related to population density such that offenders operating within areas of high population concentration (e.g., urban areas) are less likely to have "buffer zones" as these areas offer greater anonymity than less densely populated areas. Additionally, Goodwill and Alison (2005) argue that, particularly for violent and sexual crime series, the buffer zone may expand from one offense to the next as the offender's perception of risk and consequence of detection increases. At the least, it is conceivable that an offender's perception, his psychological understanding, of the area that he offends influences whether he feels there is an increased or decreased risk of identification due to the familiarity or anonymity the areas provides. Applying this to geo-OP would require the reverse logic, potentially making an inference about how much risk the offender perceived to predict how familiar they might be to that area.

Journey to Crime (JTC)

As discussed in section " The Circle Hypothesis: Commuters and Marauders," the "psychological center" of most individuals lives is their home (Gabor and Gottheil 1984), this is also literally the case in strictly geographical terms (Canter and Larkin 1993). Research has consistently found that most offenders, across many types of crimes, commit their offenses in areas that form a routine part of their lives (e.g., Cornish and Clarke 1986; Rengert et al. 1999) and thus the central point of their offending is hypothesized to be their home (Canter and Gregory 1994; Canter and Larkin 1993; Snook et al. 2005b). Further, it has been theorized that offenders travel outward from their homes to commit offenses based on various psychological, geographical, and situational aspects (Goodwill and Alison 2005). Researchers have utilized this theoretical approach by considering that the *distance* an offender travels to their crimes from their home (journey to crime, or JTC) may be related to various aspects of their character or situation.

Distances and Offense Characteristics

Overall, the JTC distance an offender travels has been found to vary between crime types (Goodwill and Alison 2005; van Koppen and Jansen 1998) and across individuals (e.g., van der Kemp and van Koppen 2007). Goodwill and Alison (2005) illustrated that serial murderers traveled in a far more distributed pattern than serial burglars and rapists, based on sequential angulation scores and pattern-analysis techniques. However, the displacement *within* a sexual offense (e.g., contact location compared to the offense location compared to the release/disposal location) has evidenced no clear patterns (Dern et al. 2005; Rossmo et al. 2004). Across individuals, research (see Wiles and Costello 2000) has indicated that the more person-targeted, expressive offenses (e.g., rape, murder) tend to occur much closer to the offender's anchor point or home base (e.g., shorter JTC distances) than object-targeted, instrumental offenses (e.g., burglary, armed-robbery, etc.). For example, Van Koppen and Jansen (1998) found that bank robbers will travel farther, typically over 10 km, for greater rewards, whereas rapists appear to travel on average just 2.5 km (Rossmo 2009). Dern et al. (2005) illustrated that planning rapists had significantly larger JTC distances than those that acted spontaneously. Further, Santtila et al. (2007) and Rossmo et al. (2004) found sexual offenders that used vehicles and exhibited planning in their offenses traveled farther than sexual offenders who did not. Clearly the access to a vehicle enables an offender to cover greater distances and has thus been found to be positively correlated to distance traveled in German sexual murderer body disposal sites (Snook et al. 2005a) and Dutch robbery crime locations (van Koppen and Jansen 1998).

Temporal aspects of an offense are also posited to relate to JTC distance (Ratcliffe 2006), in which "weekend" rapists in the UK traveled longer distances than "work-week" offenders (Canter and Gregory 1994), as did day-time as opposed to night-time rapists (Santtila et al. 2007). Mokros and Schinke (2006) offer a hypothetical explanation of disparate JTC distances based on temporal aspects by suggesting that offenders who work during the week may be more socially integrated into their surroundings and thus may feel the need to move either farther away from their home (typically on weekends) or to attack at night to avoid being identified by potential witnesses. Similarly, Rossmo (2000) suggests that the opportunity for night-time offenses is spatially biased toward certain activity areas: A rapist is much more likely to find a suitable victim in an urban night-time entertainment area or arterial routes of a city than in a rural area in general. In fact, utilizing the temporal aspects of a crime series to predict the home base of an offender was introduced by Kind (1987) in the "Yorkshire Ripper" investigation in the UK. Kind (1987) combined the time when murders were committed with assumptions about the offender's routine activities, concluding crimes later in the day would be closer to the offender's home base.

Distances and Offender Characteristics

Research has consistently shown that many offender characteristics are related to JTC distances and the overall spatial behavior of offenders. For example, JTC distance varies with gender (Wiles and Costello 2000), race (Canter and Gregory 1994) intelligence (Pettiway 1982), and arguably age. Some research has found age and JTC distance in rape offenses to be positively correlated (Dern et al. 2005) while other studies have found no significant age-related trends (Rossmo et al. 2004; Wiles and Costello 2000). However, age-related studies may be confounded by the fact that younger offenders may commit crimes closer to home due to lack of vehicle access (Dern et al. 2005).

Other relationships between offender characteristics and crime distances have also received mixed support (Mokros and Schinke 2006). For example, across several disparate crime types, offenders with previous convictions appeared to travel longer distances (Gabor and Gottheil 1984; Rossmo et al. 2004), whereas others have failed to find such a relationship (Dern et al. 2005).

The inconsistency of findings relating JTC distances to offender characteristics is perhaps not surprising; offenders, particularly across crime types, will make risk versus reward and cost versus effort judgments disparately (Goodwill and Alison 2005). This has an effect on comparative distance-decay readings which will inevitably reveal that different types of offenders may form subgroups where interrelated JTC distances are significantly different than intra-subgroup or aggregated group-level distances. This has led researchers (see, e.g., Santtila et al. 2003) to suggest that JTC research, based on distance-decay modeling to predict offender movements, is limited in use if not calibrated on the subgroup (e.g., offense type) and area (e.g., city or neighborhood) under scrutiny. Moreover, as JTC studies are not typically focused on *why* the particular pattern or distance-decay model exists for a given group of offenders, there is little prospect of applying these findings in real-world investigations.

Distance-Decay Models and Computerized Geographical Profiling (CGP) Systems

GP has seen rapid growth both in academic interest and practical use. As crime analysis in general has increased, policing agencies around the world have begun to increasingly focus on the operational use of geographic profiling in case investigations (Van Schaaik and van der Kemp 2009). Although in recent years, the breadth of operational issues and support a geographical profiler may be involved in has increased (see section "Beyond the Pins: Revealing How UK Geo-profilers Operate" for discussion), the primary role of geographic profilers is still often to predict offender anchor points (e.g., their home, place of work, etc.) through the computerized analysis of crime locations (e.g., approach, abduction, offense, release sites, etc.). There are several computerized geographical profiling (CGP) systems used by police investigators and geographical profilers and these mainly differ in how they apply various forms of the distance-decay function to produce probabilistic predictions of offender anchor points.

Dragnet (Canter et al. 2000), *Rigel* (Rossmo 2000) and *CrimeStat III* (Levine 2010) are the most common CGP systems. *Dragnet* utilizes a negative exponential function for the computation of distance-decay so that as an offender moves away from their home base, the probability of offending decreases exponentially. *Rigel* utilizes a distance-decay function with two parts, the first being a positive linear exponential function and the second a negative exponential function. Rossmo has not stated explicitly in his published work on Rigel, but it can be hypothesized that the first positive linear function reflects the assumption of a "buffer zone." *CrimeStat* III can utilize a number of JTC distance-decay functions, including a Bayesian JTC module, or calibrated functions based on an analysts' or researchers' own data. Similar to Dragnet and Rigel, it creates a probability grid in which each observation falling into an area (or "bin") is replaced by a density function (or "kernel"). The sums of those probabilities across the search area indicate the most likely areas that an offender may live Santtila et al. 2007). Density functions are used to enable a "continuous distribution" without probability "borders" giving a more realistic interpretation of the probability distribution for offending (Santtila et al. 2007, p. 5).

Although these approaches are useful tools for mathematically predicting likely areas of the offenders' anchor points (e.g., home, work, etc.), they are limited by the fact that they all assume a uniform distribution of targets across the search area by default (Kent and Leitner 2009). However, the latest version of CrimeStat (ver. 3.3) can utilize a Bayesian origin-destination function that can model previous crime activity and offender home-base backcloths (Levine 2010). A study by Van Koppen, Elffers, and Ruiter (2011) sought to develop an *ex ante* test based on the distribution of targets to determine if geo-OP was possible for a particular series. Although the results proved accurate in determining if a crime series followed a specific distance-decay pattern, it is still unclear what influence that may have on the ability to provide useful geo-OP advice. Nevertheless, if target backcloth is ignored, much of the information needed to make an accurate probabilistic estimation of the offender's anchor points cannot be made (Bernasco 2007). Next to the distribution of targets, the offenders' activity space is also influenced by environmental factors such as lakes, rivers, highways (Kent and Leitner 2009), and also by social boundaries, such as neighborhoods (de Poot et al. 2005), which are again, not considered. The saturation of targets is itself related to the environmental characteristics of an area; high-rise buildings might give less opportunity to commit burglaries, for instance. The creators of Dragnet, Rigel, and CrimeStat all purport that these factors should be taken into consideration when "applying their tools" by "trained geo-profilers"; however, how this is to be done has not been explicitly stated or published (Snook et al. 2004; Van der Kemp and Van Koppen 2007).

Finally, most CGP systems, and the vast majority of the GP and JTC research, rely on using straight-line (e.g., Euclidian) distance to compute JTC and home-base predictions (Kent et al. 2006). As Kent et al. (2006) point out this means that current profiling models assume an isotropic surface, where impedance is uniform in every direction. As a result, contemporary GP techniques do not accommodate the inherent variations of an area, such as the influence of a particular transportation network, landscape features, land-use policies, physical and psychological boundaries, etc. However, on a different note, Kent et al. (2006) suggest that a more realistic way to model the geographic space offenders are utilizing is to use a street-grid system (e.g., Manhattan distance) that better captures the potential travel patterns of offenders. Although Manhattan distance may be a more appropriate measure for North American cities, that typically have a grid-style pattern of roads and transportation networks, many cities in Europe have circular-style grid pattern in which a Euclidian approach may be more effective (Canter and Youngs 2008a). In this regard, a novel method proposed by Trotta, Bidaine, and Donnay (2011) does take account of the road networks surrounding crime locations investigating the effect of driving time and speed. Clearly, such an approach has a greater potential benefit for real-world investigative applications than elaborating on distance-decay functions that ignore important environmental aspects.

Computerized Geographical Profiling Systems in Practice

In terms of geographical profiling success, van der Kemp and van Koppen (2007) summate that geographical profiling "*. . .has the ring of being very successful, but no study to date demonstrates that it is successful in helping police investigations in more than a small percentage of cases.*" As only a few cases seem to reach the threshold criteria (Rossmo 2000) for applying geographical profiling analysis, this may not be that surprising. For example, Rossmo argues that in order to establish a pattern, Rigel requires at least five related or linked crimes (Rossmo 2000). Research by Snook, Zito, Bennell, and Taylor (2005b) contest Rossmo's assertion, citing a lack of substantial empirical underpinning of the criterion. However, more recent research by Leitner, Kent, Oldfield, and Swoope (2007), replicating the work of Newton (1988) on serial murder, has shown that predictions of an offender's "haven" (e.g., a home base or anchor point) becomes successively more accurate after the fifth offense in a series in urban burglary data. Nevertheless, in reality, the point may be moot as recent research by Knabe-Nicol and Alison (2011) report that UK geo-profilers nevertheless provide investigative advice in a substantial amount of single offenses (39 % of cases from 2002 to 2007) often in which only one crime location is known.

The Man (e.g., heuristics) Versus Machine (CGP Systems) Debate

A publication by Snook, Canter, and Bennell (2002; later followed by Snook et al. 2004) called into questions the benefit of computerized geographical profiling (CGP) systems, such as Rigel, Dragnet, and CrimeStat. An academic debate emerged between Rossmo and Snook and colleagues, the latter presenting empirical evidence that students using simple heuristics could predict an offender's home base as accurately as a CGP system (Rossmo's Rigel CGP) using simple geo-profiling heuristics. Although this debate led to an important increase in research empirically scrutinizing CGP systems, it also spawned numerous research studies that sought to produce the most optimal distance-decay algorithm for use in CGP systems using a multitude of methodologies (see Kent et al. 2006, for review). However, it is suggested that the search for the most optimal method or algorithm for determining the home base of an offender from *only* x/y coordinates on a map perhaps misses the overall point of the GP – to examine the geographical movements and decision making of an individual offender for the purposes of identifying or prioritizing areas of interest or, even better, offenders. In other words, the pragmatic utility of using *only* a CGP system or *only* a heuristic approach and not considering the target type, distribution, or attractiveness (Bernasco 2007) is questionable on a number of compelling grounds (for further review, see Stangeland 2005). By the same token, the results of Paulsen (2006) are quite telling, no matter whether anchor point predictions were based on human judgment, geometric principles, or distance-decay functions, they were all still off the mark.

Pragmatic Use of Geographical Offender Profiling (Geo-OP)

Why Are Those Pins, in That Pattern on That Particular Map?

For some time, the act of (literally) "sticking pins on map" to indicate the x/y, northing-easting's, or GPS coordinates of crimes and crime series by police and researchers alike has been used to display and analyze geographical crime patterns of individuals and groups (see van Schaaik and van der Kemp 2009, for review). With regard to GP, much of the research focus has been on developing mathematical and computational models of serial offenders' "pins on the map" (e.g., offenses) to predict an offender's home base or anchor point (s) (Canter and Youngs 2008a). Yet, a major limitation of GP research has been that the crime location "map" (e.g., topography, transport routes, social nodes, etc.) has largely been neglected (Kent and Leitner 2009).

Further and perhaps of even greater relevance to GP is the fact that only a handful of studies have attempted to understand the psychological aspects of *why* the pins are where they are (e.g., Goodwill and Alison 2005; Rossmo 2000; Canter and Shalev 2000). It should be apparent from the discussion thus far that GP is a complex process and cannot be reliably reduced to the prediction of an offender's home base from x/y coordinates, irrespective of considering the influence of the geographic area, or the routine activity of the probable offender (e.g., *why* that "space"? *why* then? *why* that route? etc.). As Van der Kemp and Van Koppen (2007) postulate, using that information is potentially the best way to fine-tune geographical profiling. Therefore, in light of the limited research exploring the complexities and reliability of GP methods, one might question what it is that geo-profilers actually advise on, what that advice is, and how it is derived?

Beyond the Pins: Revealing How UK Geo-profilers Operate

Knabe-Nicol and Alison (2011) carried out the first detailed, qualitative analysis to explore and explain the various stages and demands of the decision-making processes of geographical profilers in the UK. Using Applied Cognitive Task Analysis (ACTA), Knabe-Nicol and Alison (2011) were able to explicate the different decision-making stages, their degree of difficulty, the most common errors that can be made, and the cue and strategies to counteract possible pitfalls in the geographical profiling process undergone by UK geo-profilers.

Importantly, Knabe-Nicol and Alison's (2011) research revealed that the analysis of spatial information by UK geo-profilers was not limited to simple x/y map location(s) analysis but more often involved identifying the activity space and routes of victims, offenders, and the general public. In other words, geo-profilers did seem to be more concerned with how individuals *use* the geographical space, what *influences* that use, and what that can tell us about a number of aspects of the offense or offender. Accordingly, they found that a fundamental part of the process used by geo-profilers was to examine offenses from an environmental criminology viewpoint: Did the offender meet his victim while carrying out a routine activity (e.g., the attack was somewhat opportunistic) or does he have intimate knowledge of that area from previous exposure (for whatever reason) to it and has returned to it to offend (e.g., there was an element of planning involved in the attack)? Geo-profilers would also scrutinize the various aspects of crime scene information, the locations of the offenses, and the behaviors observed based on the least-effort principle (as discussed in section "Geo-spatial Decision Making: Distance-Decay Theory"), such as why did the offender not take the shortest escape route?

This approach, far removed from the mathematical process of pinpointing the offender's home base based on x/y crime location coordinates, enables geo-profilers to offer advice on any offense or series of offenses, even those in which the offender may be a nonlocal "commuting" or "poaching" offender. However, it is still difficult to estimate the reliability or validity of their methods when one recognizes that there is a dearth of empirical research investigating the relationship between routine activity theory and GP. Again, this begs the question: Just what are they using to make their investigative recommendations? Further, are the UK geo-profilers unique in their approach or do geo-profilers from other countries use similar strategies? Although this initial research by Knabe-Nicol and Alison is compelling and tantalizingly insightful, it still leaves many questions unanswered and requires further investigation.

Offender "Hunting Style": The Potential for Behavioral Integration in Geo-OP

Although the research into the integration of routine activities and GP is still developing, a tentative theory on how sexual offenders may make geo-spatial decisions to locate and attack victims has been offered by Rossmo (2000). Rossmo proposed that offenders employ various "hunting patterns" based on their geo-spatial movement and behavioral characteristics – acknowledging that geo-spatial movement is a dynamic decision-making process. Rossmo delineates four types of hunting patterns the: hunter (a.k.a. marauder), poacher (a.k.a. commuter), troller, and trapper. Hunters, as the name implies, are offenders who actively seek out their victim. They tend to use their place of residence as an anchor point before engaging in crime. Poachers are also active in their pursuit; however, they tend to travel further distances, even to other cities, in order to find their victims. Trollers offend opportunistically, coming across their victims while they are engaged in non-predatory activities, and seizing the chance when they become aware of it. Trappers choose to situate themselves in positions that provide them with the largest accessibility to a particular victim type. These offenders choose to work in professions such as nursing or find means to lure victims into their homes.

It is important to note that although these hunting types are theoretically possible and may make some intuitive sense, they are still being empirically validated (Beauregard et al. 2010). As Beauregard, Rossmo, and Proulx (2007) point out: No explanation is provided as to why offenders choose one method over another. Nevertheless, as stated by van der Kemp and van Koppen (2007), the process of relating behavioral inferences to geo-spatial movement patterns will no doubt result in a more fine-tuned and accurate GP approach.

Geographical Offender Profiling Validity

As with most investigative methods, it is quite difficult to assess to what extent geographical profiles have actually helped in the investigation. For example, a geo-profile may help to define the area for door-to-door enquiries, it may indicate a potential linked crime to a series in question, and/or it may pinpoint an offender's residence or simply estimate an offender's activity space, among other geographical aspects and investigative uses (Rossmo 2000). However, some aspects of a report may help while other aspects may not and some may even hinder an investigation. Unless a geo-profile specifically names the offender, it is difficult to ascertain the extent to which it helped (or hindered) the investigation – and this is rarely the case.

In practice, a number of problems arise in estimating the accuracy of geo-profiles (see van der Kemp and van Koppen 2007, for a detailed discussion). For example, how to assess the accuracy of a geo-profile, that is succinctly and comprehensively based on compelling (e.g., state of the art) theoretical and empirical research when in actuality the offender deviates from the "norm"? Is a well-informed well-reasoned geo-profile "incorrect" if the offender lives outside the predicted area? Is it possible to be "correct" based on our current knowledge (e.g., the predictions are theoretically and empirically sound), yet, still, "inaccurate"?

Integrating Routine Activity Space in Geographic Offender Profiling

As discussed previously, research integrating routine activity theory into geo-profiling is sparse. Instead much debate has surrounded on how one should go about determining the likely home base of an offender based on the x/y coordinates of their crimes (see sections "Routine Activity Approach," "Rational Choice Perspective," "Theory Summation"). This line of research is likely to be futile in real pragmatic terms of aiding police investigations. For example, even if the "perfect" algorithm existed to pinpoint exactly where, theoretically, an offender should live to the utmost precision, it still may be inaccurate. The offender may live, work, play, somewhere else. The point is clear that crime locations offer investigators and researchers alike a chance to know one particular thing about the offender: The offender has been there, for some reason, and has offended. Thus, in line with Brantingham and Brantingham's (1984) crime pattern theory, we may now know something about the offender's *activity* and *opportunity* space.

Recently, several researchers have supported this view suggesting that GP analyses must be combined with other types of information, such as who utilizes the "space" around crime locations, to increase accuracy and reliability, and/or to ultimately *identify* an offender (Bichler et al. 2011; Goodwill and Alison 2006; Stangeland 2005; van der Kemp and van Koppen 2007). Importantly, this "other" information need not only be additional crime scene information but archival information in general: phone records, land registry ownership, hospital admissions, parking infractions, etc. Historically, police investigators combined geo-profiling home-base predictions with police records of previous convictions, using this information to focus their searches and enquiries around the "usual suspects" – those already known to the police. In addition, police would prioritize search areas for the assailant within the immediate vicinity of the crime(s) to conduct door-to-door interviews, pamphlet drops, and potentially DNA swabbing (Rossmo 2000). However, as Stangeland (2005) points out:

> An intelligent criminal, of middle-class background, is less likely to be found in police files. However, the possibility of tracking him down through other kinds of records increases. Precisely because he lives a normal life, he is more likely to be a homeowner, registered on the local census rolls, to have a mobile phone, credit cards, and a car in his name. The possibility to combine two or more known data on the person and perform a computer search in public or private registers can reduce the circle of suspects. (p.467)

The current authors suggest that it is precisely this combination of various types of information databases, most readily available and in the public or semipublic domain, that will enable investigators to identify which individuals are utilizing a specific area (around a crime location) or the entire area defined by a crime series. In other words, whose *routine activity space* do these crime locations represent?

Investigative Strategies: Developing and Prioritizing Suspect Lists Based on a Routine Activity Approach

As discussed throughout this entry, the advancement of GP will be as a result of the integration of information beyond x/y crime coordinates and other mere mathematical calculations (e.g., distances between crime sites, JTC distances, etc.). It is suggested that crime locations offer insight into the activity space of an offender and these are likely areas that the offender routinely visits. Recently, researchers have begun to recognize that identifying the routine activity space of offenders is an important aspect of geo-spatial analyses. Bichler, Christie-Merrall, and Sechrest (2011) suggest that activity space, specifically points of "social" focus (e.g., gathering points), may influence offender distance-decay functions and in turn GP and CGP systems.

Within a GP paradigm,with the aim to predict or prioritize *who* has committed an offense, it seems of great importance then to determine a) *who* uses that area and b) *what* that area looks like. In terms of *what* an area looks like, geographers and criminologists have a long history of investigating the topographic, social, and transportation features associated with that area, yet arguably those have not, as of yet, been integrated into mainstream GP research or CGP systems. In fact, in relation to GP research, the topographic characteristics of the area of the crimes are seldom considered, giving rise to the chance that home-base predictions will be in a lake or an airport runway (Santtila 2010). Importantly, recent JTC research has begun to integrate topographic features of the area under study to improve JTC estimates (Kent and Leitner 2009); however, improvements have been cited as "inconsistent."

In determining *who* utilizes the activity space surrounding a crime location, the obvious first step is to identify who lives (or perhaps works) in the area. In other words, who "belongs" or has reason to be in that area. Identifying who lives and works in the area can be achieved through various archival and public records as well as through door-to-door enquiries. Other information such as traffic violations, store receipts, hospital admissions, even information on who are walking dogs, or regular joggers in the area may help to determine who *routinely* uses that area. Archival records such as census enrolment, tax records, library card ownership, General Practitioners registrations, school enrolment and attendance, and even birth records (e.g., home-towns and links to the community) could establish links between the area(s) and who routinely is in it.

In an ideal situation, one could generate lists of individuals who are routinely using the activity space around each crime location of a linked series and cross-reference the lists to prioritize individuals. This is an admittedly lofty goal and in some cases may not generate or prioritize the individual responsible, but in arguably the majority of cases, the offender will be somewhere on those lists. It is then up to researchers and investigative expertise to develop methods to identify or prioritize the offender responsible.

Concluding Comments on the Need for a Routine Activity Integrative Approach to GP

As discussed throughout this chapter, current GP research and supporting CGP systems may help us to make estimates and judgments about the activity space (e.g., home base and/or anchor points) of an offender. However, greater GP accuracy and precision in identifying the unknown offender will undoubtedly come from research that integrates analysis of crime locations (e.g., x/y coordinates), features of the geographic location (e.g., topography, routes, and social nodes), and the routine activity of offenders. Pragmatically, geo-profilers must utilize all manner of other types of information relating to the activity space surrounding crime locations to identify links to other crimes or series and to identify who is utilizing that particular activity space. In conclusion, the current authors advocate that the integration of environmental criminological theory, particularly that of routine activities, into current GP research and CGP systems is a necessary next step in advancing geographical offender profiling.

Related Entries

- ▶ Bayesian Updating and Crime
- ▶ Crime Location Choice

- ▶ Crime Mapping
- ▶ Criminal Investigative Analysis
- ▶ Criminal Profiling
- ▶ History of Geographic Criminology Part I: Nineteenth Century
- ▶ Hot Spots and Place-Based Policing
- ▶ Integrating Rational Choice and Other Theories
- ▶ Investigative Psychology
- ▶ Linkage Analysis for Crime
- ▶ Near Repeats and Crime Forecasting
- ▶ Offender Decision Making and Behavioral Economics
- ▶ Outdoor Serious Violence: The Role of Place
- ▶ Profiling Arson
- ▶ Rational Choice and Prospect Theory
- ▶ Rational Choice, Deterrence, and Crime: Sociological Contributions
- ▶ Rational Choice Theory
- ▶ Residential Burglary
- ▶ Sex Offending and Criminal Mobility

Recommended Reading and References

Beauregard E, Rossmo D, Proulx J (2007) A descriptive model of the hunting process of serial sex offenders: a rational choice perspective. J Family Violence 22:449–463

Beauregard E, Rebocho M, Rossmo DK (2010) Target selection patterns in rape. J Investig Psychol Offender Profil 7:137–152

Bernasco W (2007) The usefulness of measuring spatial opportunity structures for tracking down offenders: a theoretical analysis of geographic offender profiling using simulation studies. Psychol Crime Law 13:155–171

Bichler G, Christie-Merrall J, Sechrest D (2011) Examining juvenile delinquency within activity space: building a context for offender travel patterns. J Res Crime Delinq 48:472–506

Brantingham PJ, Brantingham PL (1984) Patterns in crime. Macmillan, New York

Canter DV, Gregory A (1994) Identifying the residential location of rapists. J Forensic Sci Soc 34: 169–175

Canter DV, Larkin P (1993) The environmental range of serial rapists. J Environ Psychol 13:63–69

Canter DV, Shalev K (2000) Putting crime in its place: Psychological process in crime site selection, Wheredunit? Investigating the role of place in crime and criminality. Crime Mapping Research Center of the National Institute of Justice, San Diego

Canter DV, Youngs D (eds) (2008a) Applications of geographical offender profiling. Ashgate, Aldershot

Canter DV, Youngs D (eds) (2008b) Principles of geographical offender profiling. Ashgate, Aldershot

Canter DV, Coffey T, Huntley M, Missen C (2000) Predicting serial killers' home base using a decision support system. J Quant Criminol 16(4):457–477

Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44:588–608

Cornish D, Clarke R (1986) Rationale choice approaches to crime. In: Cornish D, Clarke R (eds) The reasoning criminal: rational choice perspectives on offending. Springer, New York, pp 1–6

de Poot CJ, Luykx F, Elffers H, Dudink CGD (2005) Living here and perpetrating there? Social boundaries and location choice. Tijdschr Criminol 47:255–268

Dern H, Frönd R, Straub U, Vick J, Witt R (2005) Geographical behaviour of stranger offenders in violent sexual crimes. BKA, Wiesbaden

Gabor T, Gottheil E (1984) Offender characteristics and spatial mobility: an empirical study and some policy implications. Can J Criminol 26:267–281

Goodwill AM, Alison LJ (2005) Sequential angulation, spatial dispersion and consistency of distance attack patterns from home in serial murder, rape and burglary. Psychol Crime Law 11(2):161–176

Goodwill AM, Alison LJ (2006) The development of a filter model for prioritising suspects in burglary offences. Psychol Crime Law 12:395–416

Kent J, Leitner M (2009) Utilizing land cover characteristics to enhance journey-to-crime estimation models. Crime Mapping: J Res Pract 1:33–54

Kent J, Leitner M, Curtis A (2006) Evaluating the usefulness of functional distance measures when calibrating Journey-to-Crime distance decay algorithms. Comput Environ Urban Syst 30:181–200

Kind SS (1987) Navigational ideas and the Yorkshire Ripper investigation. J Navig 40:385–393

Knabe-Nicol S, Alison L (2011) The cognitive expertise of geographic profilers. In: Alison L, Rainbow L (eds) Professionalizing offender profiling: forensic and investigative psychology in practice, 1st edn. Routledge, London, pp 126–159

Kocsis R, Irwin H (1997) An analysis of spatial patterns in serial rape, arson, and burglary: the utility of the circle theory of environmental range for psychological profiling. Psychiatr Psychol Law 4(2):195–206

Leitner M, Kent J, Oldfield I, Swoope E (2007) Geoforensic analysis revisited—the application of Newton's geographic profiling method to serial burglaries in London, UK. Polic Pract Res 8:359–370

Levine N (2005) The evaluation of geographic profiling software: response to Kim Rossmo's critique of the NIJ methodology. Retrieved October 5, 2005, from Ned Levine & Associates web site at: http://www.nedlevine.com/Response%20to%20Kim%20Rossmo%20Critique%20of%20the%20GP%20Evaluation%20Methodology.May%208%202005.doc

Levine N (2010) CrimeStat: a spatial statistics program for the analysis of crime incident locations (v 3.3). Ned Levine & Associates/the National Institute of Justice, Houston/Washington, DC

Meaney R (2004) Commuters and marauders: an examination of the spatial behaviour of serial criminals. J Investig Psychol Offender Profil 1:121–137

Mokros A, Schinke D (2006) Geografische Fallanalyse [Geographical case-analysis]. In: Musolff C, Hoffmann J (eds) Täterprofile bei Gewaltverbrechen: Mythos, Theorie, Praxis und forensische Anwendung des Profilings, 2nd edn. Springer, Berlin, pp 207–37

Newton MB (1988) Geographical discovery of the residence of an unknown dispersing localized serial murder. Unpublished manuscript, Louisiana State University, Baton Rouge, LA

Paulsen D (2006) Human versus machine: a comparison of the accuracy of geographic profiling methods. J Investig Psychol Offender Profil 3(2):77–89

Paulsen D (2007) Improving geographic profiling through commuter/marauder prediction. Polic Pract Res 8: 347–357

Pettiway LE (1982) Mobility of burglary offenders: ghetto and non-ghetto spaces. J Urban Aff Quart 18:255–270

Pyle G (1974) The spatial dynamics of crime (Research Paper No. 159). University of Chicago, Chicago

Ratcliffe J (2006) A temporal constraint theory to explain opportunity-based spatial offending patterns. J Res Crime Delinq 43:261–291

Rengert G, Piquero A, Jones P (1999) Distance decay re-examined. Criminology 37(2):427–445

Rossmo DK (2000) Geographic profiling. CRC Press, Boca Raton

Rossmo DK (2009) Geographic profiling in serial rape investigations. In: Hazelwood RR, Wolbert-Burgess A (eds) Practical aspects of rape investigation: a multidisciplinary approach, 4th edn. CRC Press, Boca Raton, pp 139–169

Rossmo DK, Davies A, Patrick M (2004) Exploring the geo-demographic and distance relationships between stranger rapists and their offences (Special Interest Series: Paper 16). Research, Development and Statistics Directorate, Home Office, London

Santtila P (2010) Offender profiling. In: Granhag PA (ed) Forensic psychology in context: Nordic and international approaches. Willan Publishing, Cullompton, pp 154–172

Santtila P, Zappala A, Laukkanen M, Picozzi M (2003) Testing the utility of a geographic profiling approach in three rape series of a single offender: a case study. Forensic Sci Int 131:42–52

Santtila P, Laukkanen M, Zappalà A (2007) Crime behaviours and distance travelled in homicides and rapes. J Investig Psychol Offender Profil 4:1–15. doi:10.1002/jip. 56

Snook B, Canter D, Bennell C (2002) Predicting the home location of serial offenders: a preliminary comparison of the accuracy of human judges with a geographic profiling system. Behav Sci Law 20(1–2):109–118

Snook B, Taylor P, Bennell C (2004) Geographic profiling: the fast, frugal, and accurate way. Appl Cogn Psychol 18:105–121

Snook B, Cullen R, Mokros A, Harbort H (2005a) Serial murderers' spatial decisions: factors that influence crime location choice. J Investig Psychol Offender Profil 2:147–164

Snook B, Zito M, Bennell C, Taylor PJ (2005b) On the complexity and accuracy of geographic profiling strategies. J Quant Criminol 21:1–26

Snook B, Dhami MK, Kavanagh JM (2010) Simply criminal: predicting burglars' occupancy decisions with a simple heuristic. Law Human Behav 35(4):316–326. Springer. doi:10.1007/s10979-010-9238-0

Stangeland P (2005) Catching a serial rapist: hits and misses in criminal profiling. Police Practice and Research 6:453–469

Townsley M, Sidebottom A (2010) All offenders are equal, but some are more equal than others: variation in journeys to crime between offenders. Criminology 48(3):897–917

Trotta M, Bidaine B, Donnay JP (2011) Determining the geographical origin of a serial offender considering the temporal uncertainty of the recorded crime data. Proceedings of the 3rd international conference on advanced geographic information systems, applications, and services. Gosier, Guadeloupe, France

Van der Kemp JJ, van Koppen PJ (2007) Fine-tuning geographical profiling. In: Kocsis RN (ed) Criminal profiling: international theory, research, and practice. Humana Press, Totowa

Van der Kemp JJ, Pulinckx S, van Beijsterveldt K (2005) Moving in different directions: commuters and marauders. In: 15th European conference on psychology and law, Vilnius

Van Koppen PJ, de Keijser J (1997) Desisting distance decay: on the aggregation of individual crime trips. Br J Criminol 35(3):505–515

Van Koppen PJ, Jansen RWJ (1998) The road to the robbery: travel pattern in commercial robberies. Br J Criminol 38:230–246

Van Koppen MV, Elffers H, Ruiter (2011). When to refrain from using likelihood surface methods for geographic offender profiling: An ex ante test of assumptions. J Investig Psychol Offender Profil, DOI: 10.1002/jip.141

Van Schaaik JGJ, van der Kemp JJ (2009) Real crimes on virtual maps, the application of geography and GIS in criminology. In: Scholten HJ, van de Velde RJ, van Manen N (eds) Geospatial technology and the role of location in science (GeoJournal Library, 96). Springer, New York, pp 217–237

Wiles P, Costello A (2000) The 'road to nowhere': the evidence for travelling criminals (No. 207). Home Office, London

Zipf GK (1965) Human behavior and the principle of least effort: an introduction to human ecology. Hafner, New York

Arms Races

▶ Innovation and Crime Prevention

Associations

▶ History of Police Unions

Auto Theft

▶ Motor Vehicle Theft

Automated

▶ Automated and Manual Forensic Examinations

Automated and Manual Forensic Examinations

Jessica Bennett and Robert Hopper
JB National Investigations, Fairmont, WV, USA

Synonyms

Automated; Computer forensics; Manual

Overview

Constant changes in the technology of computers and small-scale digital devices create a number of challenges for computer forensic examiners and the law enforcement community. Two of these challenges are the identification of devices and media and the retrieval of the data from these devices coupled with having the proper training in forensic techniques to recover digital data properly. Automated tools and command-line-driven techniques, known as manual forensics, are the two methods used to retrieve digital evidence from these devices. Today's automated tools are often a more efficient method for forensic examiners to collect and analyze digital evidence than command-line procedures, which gain the same result. Depending on the situation and the type of device and media, different forensic methodologies are used. The computer forensic process is the identification, collection, preservation, analysis, and reporting of evidence recovered from computers and small-scale digital devices. This entry looks at collection of digital data in three categories: (1) on-scene collection of digital devices and media, (2) on-scene collection of digital evidence, and (3) in-lab collection of digital evidence. Standards and Best Practices require examiners to perform analysis from a bit-by-bit copy of the original data to ensure data integrity. This bit-by-bit copy is commonly referred to as a forensic image. Depending on the circumstances an examiner is confronted with, forensic analysis can be performed at the logical or physical level of the device or media. Traditional command-line techniques and automated tools each have their own purpose during forensic examinations. Time constraints, type of examination, and exigent circumstances are some examples of why a particular methodology might be appropriate to use during an examination.

Introduction

Computer forensics is the identification, collection, preservation, analysis, and reporting of evidence found on computer hard disk drives (HDD) and other small-scale digital devices and various storage media. This type of evidence is commonly referred to as digital evidence (Department of Justice 2004). Digital evidence is not only recovered from HDD but also CDs, DVDs, USB drives, SIM cards, etc. Small-scale digital devices such as cell phones, readers, IDevices (IPod®, IPhone®, IPad®), game boxes, GPSs, and even vehicle computers are commonly examined to recover digital evidence. The driving force for the evolution of computer forensics is the investigative needs of law enforcement and the rapid growth of the microcomputer field (Dixon 2005). Digital evidence was being collected and analyzed as early as the mid-1980s by the Federal Bureau of Investigation and certain military units within the Department of Defense. Law enforcement at the state and local level began to develop the expertise in the late 1980s and early 1990s.

In its infancy, computer forensics was accomplished using manual command-line-driven techniques, which evolved into suites of automated tools that are commonly used today. Microsoft's DOS™ operating system provided some basic tools for early practitioners. For example, examiners used list, dir, chkdsk (check disk), file copy, CD, etc., with switching to identify and copy files. Examiners created batch files in DOS to execute multiple commands to accomplish steps in a forensic examination. Simple commands using DOS could accomplish many of the functions needed to retrieve digital evidence during an examination. However, as technology became more sophisticated, forensic examinations needed to evolve on pace with the technology. This evolution to automated tools happened when software developers began creating suites of automated tools to interpret data and perform the forensic exam (Dixon 2005). Even though some comparatively sophisticated automated computer forensic tools are now commonly used, practitioners often still rely on manual methods using Linux to perform many forensic functions.

Fundamentals

Identification of Digital Evidence

In the 1980s and1990s, identification of devices that actually stored digital data was very simple. With the exception of mainframe computers – which used magnetic tape as a storage media – storage media was only in one form, floppy disks. Personal and business computers used floppy disks to access their operating systems as well as application and data storage. As technology advanced, the operating systems, applications, and user-generated data could be stored in two places: the HDD or on some configuration of floppy disk, that is, 5.25 in., 8 in., or 3.5 in. (Department of Justice 2004). As a result, law enforcement needed to know what an HDD and floppy disks of various sizes looked like. For those who were already working in the field of computer forensics, this was simple and straightforward because they were using HDDs, various kinds of floppy disks, and, although not as often, tape drives. However, very quickly, this simple task became much more challenging. Manufacturers began producing many different types (a veritable cornucopia) of storage media, Zip® disks, Bernoulli® disks, Jazz® disks, etc., which were capable of storing much more data. These new storage media were relatively expensive, and law enforcement often could not afford them. Another difficulty was that examiners might not have had the right equipment to read and image these media when the digital evidence was seized for examination in the lab.

Today's USB devices are a great example of storage media that have drastically changed. These devices are now made in hundreds of configurations from Santa Claus statutes to sushi, making it sometimes difficult for the untrained to recognize them as potential repositories of digital evidence. Add to this the number of new small-scale digital devices, cell phones, GPSs, gaming stations, cameras, readers, IDevices, etc. that might also contain digital evidence, and training in identification becomes even more critical. These new sources for digital evidence have also dramatically increased the level of training required by examiners as well as increased the caseload for computer forensics' laboratories. Additionally, storage schema and proprietary formatting continue to make forensic imaging of media a challenge and require specific training for examiners. Cell phone, GPS, and reader devices are other excellent examples of the diverse number of operating systems and storage schema being used by manufacturers, which now requires not only new tools and techniques to collect digital evidence but again specialized training for examiners.

All levels of law enforcement, from patrol officers to detectives and, of course, computer forensics examiners, are now expected to be able to identify devices and media that could potentially contain digital evidence; specific training in identification and seizure of digital media is now often a basic academy curriculum. This training is becoming standardized by organizations such as the National White Collar Crime Center (NW3C); however, these standards are not mandated. Further, although new automated tools and techniques have been and are being developed to keep pace with these challenges, training in their proper use is also required. Departments are always financially challenged to find the funding and time to keep officers properly trained in the various disciplines within the law enforcement field; computer forensics training at all levels is no less a challenge. (Refer to Open Questions for further discussion.)

Collection and Preservation of Digital Evidence

Digital evidence must be collected in a fashion that both preserves and protects the device, the media, or the record(s) so that the original evidence is preserved in its exact original state. Additionally, practitioners must be able to demonstrate that the copy used for examination is an exact copy of the original. Preservation begins with the first contact with the evidence, which may be as simple as maintaining the chain of custody, and in the case of an HDD, it is packaged in non-static cushioned wrap and care taken not to expose the HDD to extremes of heat or cold. The National Institute of Justice, in partnership with the National Institute for Standards and Technology (NIST), authored the *Electronic Crime Scene Investigation: A Guide for First Responders* (2008). This publication is still the definitive guide for law enforcement's collection of digital evidence today. These guidelines provide law enforcement with standards, guidelines, best practices, and strategies for collecting and preserving digital evidence.

For ease of discussion, collection is explained in three phases: (1) on-scene collection of physical devices and media, (2) on-scene collection of digital evidence, and (3) in-lab collection of digital evidence.

On-scene collection of devices and media simply means the proper marking and packaging and scrupulous documentation of chain of custody of the physical evidence. In the case of a computer that is connected to peripheral devices, such as a computer connected to a printer, router, mouse, and monitor, photographs and diagrams of all connected devices should be taken. These should be in such detail that the "system" could be reconfigured in court exactly as it was at the time of seizure (Department of Justice 2008; NW3C 2011). Digital evidence response training is now broken into three tiers: tiers I, II, and III. Generally, training required for identification and physical collection of digital evidence is tier I level training.

The second area is on-scene collection of digital evidence; tier II level training and expertise is required. There are a number of reasons that digital evidence might be collected on scene rather than in a computer forensics lab. Often the court requires that if a search warrant is being served at a business, for example, officers are required to collect or image the original media on scene and leave the business system up and running; this requirement usually depends on the size of the business. If the business is a relatively small operation, officers can use a forensic copy of the original HDD(s) to leave behind for the business to continue operations because an exact bit-by-bit image is made of the original media on scene. In other circumstances, when a large business operation is involved, examiners may be required to obtain forensic copies of files and folders (Schweha and Inch 2008).

On-scene collection of evidence may require that imaging be accomplished at a logical level as opposed to a physical bit-by-bit image level. A couple of examples that would require a logical level collection of digital evidence would be when encryption is encountered or exigent circumstances require immediate access to information because life or harm is eminent.

Data that is encrypted is not accessible once a computer is powered down without the password or phrase that allows re-accessing the encrypted volume, that is, file, folder, or disk. Normally, a computer that is being seized as evidence is powered down by removing the power source – power cord or battery – rather than using the operating system to turn the computer off (polite or soft shutdown) (Department of Justice 2004; SWGDE 2008). Powering down is not an option if encryption is running on a computer unless passwords or phrases are known. This circumstance requires either live imaging of the system using automated tools or copying files and folders manually from the computer.

A logical image is not a bit-by-bit image of the drive, but an image of files and folders on the drive. A logical image collects data based on the installed operating system, file system, and applications (Craiger 2005). A logical image does not copy unallocated or slack space, boot partitions, partition tables, etc. Automated tools will produce a forensically sound logical image because they perform hashes of the data at the logical level as they are being obtained during the imaging process. These hashes, which will be discussed later, ensure that the data copied from logical volumes are an exact copy used during on-scene or forensic lab examination and analysis of the evidence. In the case of exigent circumstances, a logical image may be the only method of obtaining the evidence. There are at least two software tools that allow forensic-acceptable logical imaging of on-scene digital evidence: EnCase® (EnCase Portable) and Access Data® (Live Response®). These two tools are commonly used for on-scene, live acquisition of digital evidence. If encryption is present, manual methods of accessing data cannot be used because they require that a system be booted from a powered down state before commands can be executed.

Forensic examiners work from a forensic image and use various algorithms such as MD5 or SHA to verify that the two pieces of evidence, the original and the forensic image, are mirror images of each other (Allen 2006). Hashing is referred to as the process of assigning “a mathematical algorithm against data to produce a numeric value that is representative of that data” (Department of Justice 2004, p. 40). Hash values change if the evidence is changed or altered in any way, thus assuring the integrity of the data (SWGDE 2006b). Hashing of the evidence is accomplished in two stages: when the original data is hashed and when the copy is created and hashed. These two hashes must match in order to confirm that an exact copy of the original data has been obtained for examination. If a bit is changed in the process of the hashing, hashes will not match. If hashes do not match, the examiner must be able to explain what happened that caused the alteration of the evidence, and the change must have been a viable change (Bell and Boddington 2010). Automated forensic tools do this seamlessly.

Using manual command-line techniques, examiners are able to copy bit-by-bit copies of files; however, the hash function is missing from the procedure. If the examiner knows that the DOS command-line “copy” command will result in a bit-by-bit copy of a file, a hash is only necessary to verify that an automated tool actually did what it was suppose to do – make a bit-by-bit image. A file could be hashed on the original drive in the exact location it is stored. The file then copies to another forensically clean media, and the file is again hashed. Those two hashes are compared for an exact match. This would confirm that the file on the original media was copied as an exact, mirrored, image to the copy media (Craiger 2005).

An important concept in the forensics process is the difference between a forensic examination and device interrogation. These two processes are performed on various devices and involve two different methodologies. Standard forensic examination of digital evidence requires that an exact, uncorrupted, bit-by-bit image of the data stored on a media be obtained and an exact copy is used to analyze the data to obtain evidence (Department of Justice 2004; Lewis 2008). Although examination of small-scale digital devices such as GPS devices and cell phones are often called cell phone forensics or GPS forensics, the methodologies used to collect data from these devices often cannot be done to forensic standards. These types of examinations are more accurately defined as device interrogations not forensic examinations. The proprietary nature of the operating systems and storage schema of handheld devices often negates the ability for a true forensics copy to be made of stored data as well as the operating and file system. Because many of these devices do not allow a forensic copy to be made, methods such as photographing individual screens and logical acquisition of stored data must be performed (SWGDE 2006a). Because of these limitations, often, the only methodology for collection of digital evidence is capturing photographic images (screenshots) and un-hashable data dumps of stored data on these devices, that is, contact lists, text messages, recent calls, calendars, and photos.

Another methodology for on-scene collection of digital evidence is triage. Triage tools access the computer from a boot disk and parse through files and folders at the logical level. There are a number of tools that use CD, thumb drives, and other bootable methods along with triage application that are specifically designed to assist the investigator in quickly and efficiently determining if there is obvious evidence on the computer. One of the most used triage tools is TUX4N6™ developed and distributed free to law enforcement by NW3C (NW3C 2010). Some of the other commercial triage tools are Blade™, EnCase Portable™, Drive Prophet™, LiveResponse™.

Prior to performing any triage examination on a computer, one key element is to recognize what files can be collected during the process. Because triage is done on the logical level, files and folders are accessible, not slack space, partition tables, and other areas of the media. Knowing what files to look at is key. Relevant locations to search commonly include My Documents, Desktop, Recents, Recycle Bin, Temp, Downloads, Internet Explorer®, or Firefox®, and Flash® usage histories, the "low hanging fruit" (NW3C 2010). This listing is certainly not all-inclusive, but simply some examples of the data locations that can be quickly searched for digital evidence. An advantage of using modern triage tools is that examiners can do both automated and manual searches of the media for evidence. Digital evidence collected using the triage application with automated or manual search techniques is forensically sound and can often be submitted in court without further forensic examination in the lab. On-scene as well as in-lab triage of computers has been a great help in decreasing the case backlog in many computer forensic labs around the United States.

The third area for discussion is examination and analysis in the forensic lab environment. Any evidence seized from the scene of the crime should be imaged, if possible, and the examination performed on a forensic image of the evidence. According to the working group that set forth the NIJ Analysis Guidelines (Department of Justice 2004), an image is "an accurate digital representation of all data contained on a digital storage device (e.g., hard drive, CD-ROM, flash memory, floppy disk, Zip®, Jazz®) which maintains contents and attributes but may include metadata such as CRCs, hash value, and audit information" (p. 40).

Working from an image is important for several different reasons. First, it is important that there are no modifications accidentally performed to the original evidence. This enables the evidence to stand up to authenticity if it is taken to court (Lewis 2008). Second, all tests performed on the evidence must be able to be replicated. This means that the defense counsel, or an outside examiner, must be able to reproduce all tests performed on the evidence in an attempt to reproduce the same results that the examiner received during the examination (Crozby 2001). Lastly, working from an image is always a fail-safe method in case the image becomes damaged for any reason in the process of examination. In any case, the examiner always has the original evidence to rely on once again to make another image (Department of Justice 2008; SWGDE 2006a).

Analysis of Digital Evidence

The analysis must conform to the scope of the search warrant (Department of Justice 2009). By example, a search warrant authorizing a search for images, that is, child pornography, may very well prohibit the examination and analysis of documents that may reside on the same hard drive.

Prior to performing the analysis on any evidence, the examiner must become familiar with the operating system and file structure of the device in which is being examined. Operating systems vary depending on manufacturer, and although the types of evidence will always be the same, the location of where it is on the machine or device will change. The Microsoft® operating system is updated often; this requires new training and expertise by examiners. Microsoft® is entirely different from a Linux or Macintosh® operating system. This is also true with different file systems (Britz 2009). While testifying in court, the examiner must be able to state which operating system was used and where that piece of evidence was found in the file structure. Changes made to operating systems and file systems require examiners to have updated training. Updated training for automated tools is expensive, and it is often very difficult for examiners to afford this training. The time away from their caseload for training, not to mention the time required in gaining expertise using the new tool capability, is another burden on examiners and their agency.

Solid-State Drives (see discussion in “Key Issues/Controversies”)

Performing keyword searching, data carving, and email examination are some of the data analysis functions that can be performed at the physical level both manually and by using automated tools. Keyword searches allow the examiner to look for items not related to the operating or file system (Department of Justice 2004). A keyword may be used to search for specific items that pertain to things such as images, documents, spreadsheets, and databases. An examiner can also search by file extension. Email searching falls into two categories: application-based and web-based email. Common email applications are those that are installed when Microsoft® is installed. Yahoo®, Gmail®, and Hotmail® are a few web-based email services that examiners will find in the temporary Internet folders. Data carving looking for deleted files in unallocated and slack space is another important element in physical analysis (Craiger 2005).

Reporting of Digital Evidence

Automated analysis tools have built-in reporting capabilities that establish an audit trail of most of the steps of an examination; this is critical for court presentation of digital evidence gleaned from examinations. In addition to these auto-generated report capabilities, when performing manual analysis or automated forensic examinations, it is also necessary to document other steps in the analysis process. In addition to automated reporting, standard policy requires that a separate examination report is prepared that documents steps taken that are not part of application-generated reports, that is, system documentation, peer review of analysis results, and chain of custody. An important element of reporting is to document that standards, best practices, and department policy were followed.

State of the Art

Various automated tools assist with computer forensic examinations during the imaging and analysis stages. Essentially, all these automated tools do for forensic examiners is to run multiple command-line functions on the computers and media looking for specific evidence. In comparison, this is what a manual examination is; however, it takes a much longer process, and the examiner has to execute one command at a time or create batch files to perform various forensic functions. Law enforcement must be able to prove that the automated tool or manual methods used in the examination process were forensically sound and did not jeopardize the integrity of the evidence that was collected during the examination. The National Institute of Standards and Technology’s (NIST) Computer Forensics Tool Testing project (2001) created validation and testing guidelines that tool developers can use in order to measure assurance for forensic requirements in judicial proceedings.

The automated tools available to law enforcement and forensic examiners can essentially be broken into three categories: imaging tools, triage tools, and analysis tools. Some of the more popular automated forensic tools used by law enforcement forensic labs include WinHex™, EnCase®, Access Data® Forensic Toolkit™, and various Linux tools; however, the availability of these tools relies on the budgets of the agencies (Allen 2006; Dixon 2005). Automated tools have even been developed to perform a variety of forensic functions such as triage, live system acquisition, and small-scale device interrogation. These tools are constantly updated and new tools developed because of the ever-changing varieties of cell phones and other small-scale or handheld digital devices.

Key Issues/Controversies

Changes in technology, such as data storage devices, new proprietary operating and file systems, and other advances, will cause forensic standards to always evolve. Funding for research, testing, and validation to organizations such as NIST is critical for law enforcement to have validated tools available for forensics examinations of these new and changing technologies.

A current example of evolving technology forcing a paradigm shift in computer forensics is solid-state drives (SSD). Files are written differently on HDD and solid-state drives. Some of the differences between the two types of drives are that SSD drives are faster and allow access to parts of data. Bell and Boddington in their research state, "A paradigm shift has taken place in technology storage and complex, transistor-based devices for primary storage are now increasingly common" (2010, p. 5). As a result of their research, the authors identify twenty-one critical areas that examiners must be aware of in their "Recommendations and Guidance" section. This research could indeed change the state of computer forensics as it applies to SSD. Further research will no doubt be forthcoming in this specific area of computer forensics.

Future Direction

There are several areas that are open for discussion in the field of computer forensics, including certification and accreditation, standards for forensic training for practitioners, and law and policy.

Practitioner Certification and Forensic Lab Accreditation

- As the Internet makes connectivity ubiquitous, law enforcement will be challenged in a number of areas. During a criminal investigation, how can law enforcement locate and collect the data generated by hundreds of Wi-Fi networks that will be commonly pushing communication to handheld devices such as cell phones, readers, and pads?
- Software tools used in EDiscovery are beginning to be used by law enforcement to assist in parsing large datasets, that is, email servers. Although these EDiscovery can be of great assistance, testing and validation of forensic soundness of the tools are not commonly done.
- Should live system acquisition and device interrogation be treated as the same type of forensic examination?
- The number of handheld devices has now exceeded the number of computers being purchased. Handheld devices are evolving to include much more capability and sophistication. The research and development required to develop training that keeps law enforcement abreast of these new technologies is a challenge now because of lack of congressional funding, and without long-term solutions, computer forensics will not keep pace with the quantum leaps that technology is making in the twenty-first century.

Standards for Forensic Training for Practitioners

Are colleges and universities keeping pace in their curriculum in the area of computer forensics to match the needs of the law enforcement, the intelligence community, private sector business, and industry? Although there are some excellent undergraduate- and graduate-level programs, private sector business and government are still challenged to find enough qualified graduates to fill their needs. Techno-based business often expresses a need for university curriculum focused on knowledge and skill sets that result in graduates being able to enter the field having the practical skills instead of having a degree based on theory and not practice.

Law and Policy

The list of legal issues surrounding computer forensics is endless, from ensuring that privacy is protected to outdated laws and regulations that were enacted decades ago for the telecommunications system that simply do not apply to today's Internet-driven society. Many heated debates have been raging for decades over jurisdictional issues as they apply to search warrants, subpoenas, preservation orders, etc. Is a digital search warrant that is accepted in all jurisdictions the answer? The Internet has no state, county, city, or even country jurisdictional boundaries, but the laws governing these issues certainly do. Legislation must be as nimble as the technology and the Internet. Crafting cogent regulations, laws, and policy that will live to the future is beyond challenging.

Although more and more potential evidence is now held by third parties such as Internet service providers and cloud services, the laws governing law enforcements' legal access to potential evidence are not keeping pace with these changes. How are law, policy, and regulation going to change to solve these issues?

Conclusions

Computer forensics is evolving rapidly for law enforcement due to the ever-changing growth of technology. Forensic examiners rely on two methods for analyzing data: automated tools and manual forensics. Automated tools evolved from command-line methods, producing a more efficient method of collecting data. Depending on the circumstances surrounding the forensic examination, that is, whether it is an on-scene device collection, on-scene data collection, live system acquisition, device interrogation, or lab examination, the most important concept during the examination is to maintain the integrity of the evidence. Identifying, collecting, preserving, analyzing, and reporting of digital data found on hard disk drives and small-scale digital devices are known as the computer forensic process. Automated and manual forensic exam principles are introduced at the collection stage and can be used throughout the remaining forensic examination process. Analysis can take place at both the logical and physical levels of a device. The goal of the computer forensic exam is to document the process (no matter whether automatic or manual methods were used), ensure evidence integrity, and maintain proper chain of custody in order to present the evidence in court.

Related Entries

▶ Computer Forensics

Recommended Reading and References

Allen W (2006) Computer forensics. IEEE Secur Priv 59–62. Accessed 6 Oct 2011. http://dsonline.computer.org

Bell G, Boddington R (2010) Solid state drives: the beginning of the end for current practice in digital forensic recovery? J Digit Foren Secur Law 5(3):1–20

Britz M (2009) Computer forensics and cyber crime: an introduction, 2nd edn. Prentice Hall, Columbus

Craiger JP (2005) Computer forensics procedures and methods. In: Bidgoli H (ed) Handbook of information security, vol 3. Wiley, Hoboken

Crozby P (2001) Methods in behavioral research, 10th edn. McGraw Hill, Boston

Dixon P (2005) An overview of computer forensics. IEEE Potent 24:1–10

Lewis P (2008) Understanding the basics of computer forensics. EDPACS 37:3

National Institute of Standards and Technology (2001) Computer forensics tool testing program. Accessed 6 Oct 2011. http://www.cftt.nist.gov/

National White Collar Crime Center (NW3C) (2010) "TUX4N6™ user guide." Accessed 6 Oct 2011. www.nw3c.org

National White Collar Crime Center (NW3C) (2011) Identifying and seizing electronic evidence training. www.nw3c.org

Schweha J, Inch S (2008) Remote forensics may bring the next sea change in e-discovery: are all networked computers now readily accessible under the revised federal rules of civil procedure? J Digit Foren Secur Law 3(3):5–28

Scientific Working Group on Digital Evidence (SWGDE) (2006) Best practices for computer forensics. Accessed 6 Oct 2011. http://www.swgde.org/

Scientific Working Group on Digital Evidence (SWGDE) (2006) Data integrity within computer forensics. Accessed 20 Oct 2011. http://www.swgde.org/

Scientific Working Group on Digital Evidence (SWGDE) (2008) Special considerations when dealing with cellular phones. Accessed 24 Oct 2011. http://www.swgde.org/

U. S. Department of Justice (2009) Searching and seizing computers and obtaining electronic evidence in criminal investigations. Office of Legal Education Executive Office for United States Attorneys

U.S. Department of Justice (2004) Forensic examination of digital evidence: a guide for law enforcement. NIJ Special Report

U.S. Department of Justice (2008) Electronic crime scene investigation: a guide for first responders, 2nd edn. NIJ Special Report

Babies Behind Bars

Leana A. Bouffard
College of Criminal Justice, Sam Houston
State University, Huntsville, TX, USA

Overview

Although there has been massive growth in the number of women in prison in the USA and other countries, women remain a significant minority within correctional systems. In the United States, the number of women in prison increased by more than 800 % in the 30-year period between 1977 and 2007. In the United Kingdom, the female prison population nearly tripled in just a 12-year period from 1992 to 2004. Despite these increases, in most countries, women still constitute less than 10 % of the prison population (Alejos 2005). There have long been concerns that the correctional system, which was built around a need to house large numbers of male offenders, ignores, minimizes, and marginalizes female inmates and their specific needs, including medical needs.

In particular, statistics find that a large proportion of female inmates are mothers. One recent study in the United Kingdom reported that more than half of women in prison have children under the age of 16 and one-third of women have children under 5 years old (Edge 2006). Similarly, in the United States, more than half of state and federal inmates have an estimated 1.7 million minor children (Glaze and Maruschak 2008). In most cases, the women are the primary caregivers and had been living with their children prior to incarceration. The children of men who are imprisoned typically stay with their mothers during the father's incarceration; however, children whose mothers are incarcerated are much less likely to remain in their family home, often being placed with extended family. Additionally, incarcerated mothers are five times more likely than incarcerated fathers to report that their children were placed in foster care or other institutions during their incarceration (Glaze and Maruschak 2008). While the number of women who are pregnant, give birth, or have new infants while in prison is much smaller than the number who had children prior to incarceration, all of these women and their infants and children have specific needs that are often not addressed by prison policy, practice, or administration.

Legal Issues with Mothers and Children in Prison

In 1990, the United Nations Congress on the Prevention of Crime and Treatment of Offenders issued a resolution that the imprisonment of pregnant women should be avoided, stating that "the use of imprisonment for certain categories of offenders, such as pregnant women or mothers with infants or small children, should be restricted and a special effort made to avoid the

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

extended use of imprisonment as a sanction for these categories" (Alejos 2005, p. 13). There are a number of issues that arise in dealing with convicted women who are either pregnant or are mothers, especially to young children. For women (and men) who have children, applicable legal considerations include the Convention on the Rights of the Child, which specifies that children should not be separated from their parents against their will except in a case where it is in the best interest of the child (Alejos 2005, p. 14). The African Union's Charter on the Rights of Children, for example, specifies that noncustodial sentences should be considered first in cases of pregnant women and mothers of infants and young children and that the purpose of punishment should be reform and reintegration with family (Alejos 2005). In fact, the best interests of the child are indicated as the primary guiding factor, which raises the question of whether it is in a child's best interests to remain with their mother in a prison or to be separated from their mother in order to remain in the community.

In terms of pregnancy and childbirth in prison, the United Nations standard minimum rules for the treatment of prisoners specify that women should receive all necessary prenatal and postnatal care, that birth should occur in a hospital outside of the institution, and that a child's birth certificate should not list a prison as the place of birth. Additional criteria require that efforts should be made to maintain and improve relationships between prisoners and their family, which would presumably include children. While many international legal standards suggest that prison should only be used for pregnant women in extreme circumstances, for those women who are incarcerated while pregnant, additional legal considerations address the necessity of appropriate prenatal care, including nutrition, medical care, a healthy environment and exercise, and attention to any cultural or religious issues surrounding childbirth. Physical restraint of pregnant women should not be used except in extreme circumstances, and numerous international standards call for abolishing the practice of shackling or restraining women during labor and childbirth.

Children living with a parent in prison also present special legal challenges, and the management of these children varies widely by nation (Bastick and Townhead 2008). General international standards specify the need for monitoring mechanisms and the involvement of child welfare agencies in institutional decision-making that affects the child. Adequate accommodations should be available, including separate mother-child units that are removed from the general prison population, the provision of a safe and healthy environment, as well as appropriate facilities to house the children at times when their parent is participating in activities not appropriate for children. Underlying these general standards is a fundamental assumption that the child is not a prisoner and should never be treated as one.

Pregnant in Prison

Data compiled by the Bureau of Justice Statistics indicated that in 2004 in the USA, 4 % of female state inmates and 3 % of federal inmates were pregnant when they were admitted to prison. International standards require that pregnant inmates receive all appropriate medical care, nutrition, and assistance. In the USA, a number of medical organizations as well as the National Commission on Correctional Health Care have issued policy statements on appropriate health care for inmates in general and for pregnant inmates more specifically. These recommended guidelines include "timely and appropriate prenatal care, specialized obstetrical services when indicated, and postpartum care" (NCCHC). Specific issues that should be addressed among pregnant inmates include prenatal medical exams, nutrition, and counseling, along with diagnostic screening for high-risk pregnancies, including HIV testing. In addition, the NCCHC recommends that institutions should have written agreements with a local medical facility for delivery. Despite these policy guidelines and standards, research finds that pregnant women in prison do not receive adequate prenatal care and health screening (Edge 2006). Research with pregnant inmates in England finds that these women report being constantly hungry, being physically restrained with shackles and belly chains, and invasions of privacy with male guards present during medical exams and delivery.

In 2010, the Rebecca Project for Human Rights in conjunction with the National Women's Law Center conducted a state-by-state analysis of policies and practices for dealing with pregnant and parenting women in prison in the USA. According to this study, three-quarters of states do not have adequate (or any) policies in place for providing prenatal care for women who are pregnant during their incarceration. This may be especially problematic because female inmates typically are less likely than other women to have had routine health care prior to their incarceration. These women may have undiagnosed health conditions as well as higher levels of drug use, sexually transmitted diseases, and risk factors for HIV, which may lead to a higher proportion of high-risk pregnancies in this population. Incarcerated women are also more likely to have histories of neglect and victimization, including sexual assault. Some research has noted that since many women in prison receive short sentences, the rapid turnover of the female prison population may interfere with the ability of institutions to ensure adequate prenatal care and aftercare (Edge 2006). The failure to provide a high level of prenatal care may have implications for the pregnancy as well as short-term effects on the infant, including low birth weight or other birth complications, and long-term consequences for the health of both mother and baby.

Despite a great deal of increased attention to the issue recently, the practice of shackling female inmates during pregnancy, labor, and delivery remains an area of concern. The National Commission on Correctional Health Care (NCCHC), the American Congress of Obstetricians and Gynecologists (ACOG), the American Medication Association (AMA), and the Association of Women's Health, Obstetrics, and Neonatal Nurses (AWHONN) have all issued position statements arguing that the use of restraints for pregnant women should be avoided completely except in cases of extreme risk, and when restraints are judged to be necessary, the least restrictive method should be used. For example, the AWHONN (2011, p. 817) "opposes the practice of shackling incarcerated pregnant women...[which] should only take place if prison officials reasonably believe, after an impartial and thorough evaluation, that a particular individual may attempt to harm herself or others or presents a legitimate flight risk." In January 2012, the American Correctional Association (ACA) also issued a policy statement that pregnant women should be restrained in the least restrictive method possible, that waist and electronic restraints should never be used during pregnancy, and that leg restraints should never be used during labor and delivery. While efforts to restrict the use of physical restraints on pregnant and laboring women in prison are gaining momentum, more than half of the states do not have policies that comprehensively limit the use of restraints (The Rebecca Project 2010). As of 2012, only 17 states have adopted laws that limit the use of physical restraints for pregnant inmates, with four states addressing the issue in policy.

International Examples of Prison-Based and Community-Based Programs

While it is more common outside of the USA for babies born in prison as well as other children to reside in prison with their mothers, there is little information available and even less consistency in how institutions and nations deal with mothers and their children. For example, in Australia, a mother or primary caregiver may submit a request to the superintendent of the institution that her baby live in the facility with her for up to 12 months. Requests may also be made for older children to stay overnight. In France, however, the decision to keep a child in prison is left up to the mother with no approval necessary except for the agreement of the father, and babies may stay in the prison until they are 18 months old (Alejos 2005). In Finland, legislation allows both mothers and fathers to bring their children into prison with them. While there is no specific age limit, this typically involves children younger than 2 or 3 years old. Some countries, like

Germany and Spain, allow older children (in some cases, up to age 6) to live in prison with their mothers. In other countries, like Norway, children are not allowed in prisons at all.

There are a number of examples of the various ways that programs throughout the world manage babies, children, and their mothers in prisons and community-based alternatives. In England, the prison service has designated Mother-Baby Units (MBUs), which are described as separate living arrangements within a women's prison where women and their children may live together during her incarceration (Edge 2006). These units are designed with the rights of the child as the primary consideration. Children can stay with their mothers in prison up to a maximum of 18 months of age. Women must apply to participate in the program, and admission is determined based on what is in the best interests of the child. Mothers must be and remain drug-free, must be willing to take primary parental care of the child, and must have no physical and/or mental health or other concerns that would interfere with their ability to care for the child. Applications are taken system-wide, and women may be transferred to another facility if they are accepted. Because of the geographic distribution and isolation of facilities, placement in a Mother-Baby Unit may mean that a mother and baby are separated from the rest of their family, including any older children she may already have.

Canada also has an Institutional Mother-Child Program (Alejos 2005). Mothers who are categorized as either minimum or medium security and who are housed in institutions that provide the program are eligible to participate as long as the offense did not involve a child. In general, participants may request that their young children (up to 4 years old) live with them full time and that older children (up to 12 years old) may have part-time residence. Canadian policy specifies that the primary consideration in allowing women to participate should be what is in the best interests of the child, and the goal of allowing children to reside with their mothers is to maintain and support stable mother-child relationships. Interestingly, those reviewing applications for participation in the program are also directed, when feasible, to consider the wishes of the child in determining who will be accepted. If accepted into the program, mothers sign a parenting agreement, which may include parenting training, health-care plans, and other criteria. Program staff monitor a mother's compliance with program requirements and review cases after the first month and then every 6 months.

While women have long been allowed to keep young children in prison with them in Spain, Feintuch (2010) describes an institutional program implemented in 2004 with the stated goal of removing all children from prisons by 2012. These "external mother units" are designed for those women with longer sentences, who committed a nonviolent offense. In these units, women are housed in separate apartments with their children, and the women are responsible for the feeding and care of their children. Efforts are made to maintain a homelike environment, including the language used to describe the units, referring to a woman's "apartment" and "home." During the day, the children attend a community preschool, while the women participate in programs focusing on parenting classes, education, and job training. The women may also work in the community as long as the job does not interfere with their child's schedule. Cases are reviewed frequently to ensure the health and safety of the children, and as a child approaches 3 years of age, the mother's case is reviewed to determine whether she could complete her sentence in the community. The units have an age limit of 3 for the children, so older children are not eligible to live with their mothers in the units. There are only a few facilities in the country, which means that mothers may be located in a facility some distance from their other family, and this may also interfere with the involvement of fathers. Interestingly, Spain does have one family unit where families can live if both parents are incarcerated. Finally, the external mother unit program is also heavily dependent on help from nongovernmental organizations, so continued funding may be an increasing concern. Community-based programs are also available for convicted parenting women and their children. In addition to the "external mother units" described, Spain is increasing the use of community-based alternatives for nonviolent, parenting women with shorter sentences.

Children in Prison: US Examples

In contrast to the numerous examples from the international perspective, it is uncommon for imprisoned mothers in the United States to keep their children with them while serving their sentence. In the USA, it is common for mothers who give birth while incarcerated to be separated from their babies within a few days. While the mother returns to the institution to finish serving her prison term, the baby is placed in the community with family or social services. Mothers who have children prior to incarceration simply leave those children behind when they begin serving their sentence. Support for the use of prison nursery programs in the United States has varied over the years, and few states operate programs. There are, however, a few notable examples of programs ranging from prison nurseries to community-based sanctions. There are generally two goals associated with these types of programs: that allowing mothers and babies to stay together will foster positive attachment and bonding and, relatedly, that this environment will be rehabilitative for the mother and will reduce her chances of recidivism.

Within the past two decades, a minority of states have developed prison nursery programs, which allow children born during their mother's imprisonment to remain with the mother for a period of time following birth. The oldest program, operating since 1901, is located in New York. Other programs have been developed in Nebraska, Washington, Massachusetts, Illinois, Indiana, Ohio, California, West Virginia, and South Dakota. The Federal Bureau of Prisons also operates a program, Mothers and Infants Nurturing Together (MINT). Prison nursery programs in the USA are generally designed for women who give birth during their incarceration, who were convicted of a nonviolent offense, and who have no history of child abuse or neglect. The length of time that babies can stay with their mothers in the program varies from a low of 30 days in South Dakota to a maximum of 3 years in Washington. Most of the prison nursery programs in operation are housed in an area of the prison separated from the general population and incorporate parenting skills training.

The Bedford Hills Correctional Facility for Women in New York houses the nation's oldest prison nursery program, operating since 1901. Women who are pregnant when admitted to prison and who will give birth in custody are eligible for the program. Selection of participants is determined by a number of factors, including a consideration of who will have custody of the child, the length of the mother's sentence, and the type of crime she committed. If the mother will be discharged from prison within 18 months following the birth of her child, the babies can stay with their mothers the entire time. Otherwise, the maximum stay is 12 months. Mothers receive weekly nurse visits, and the program includes specialized children's activities, daycare, parenting education, family counseling, and assistance with child placement.

In 1994, Nebraska expanded an existing program for mothers and children into a prison nursery program modeled after the Bedford Hills New York program (Carlson 2001). Mothers in the women's prison in Nebraska give birth at a local hospital. For program participants, both mother and baby return to the nursery facility after discharge from the hospital. To be eligible for the nursery program, women must have less than 18 months left on their sentence following the birth of their child, must have no prior convictions for child abuse, and must sign a parenting program agreement. As part of that agreement, prenatal classes are required, in addition to parenting classes, educational programs, and employment. In its early stages, participating mothers reported strong support for the program, feeling that they have a better relationship with their child and that they are better mothers as a result of their participation and involvement in parenting classes (Carlson 2001).

More recently, the Washington Correctional Center for Women opened the Residential Parenting Program in 1999 for minimum security women who had committed a nonviolent offense and would have no more than 3 years remaining on their sentence following the birth of their child (Women's Prison Association 2009). In a separate unit within the facility, mothers in the program each have a private room with a bed for her child. The unit also includes indoor and outdoor children's play areas. Monthly pediatrician visits are provided to monitor the health and well-being of the children. Because children may stay with their mothers up to 3 years, the facility partners with a local organization to provide an early head start program for the children, which incorporates activities for the children, nutrition counseling, information about child development, and maternal depression screenings.

Alternatives to Incarceration in the United States

There is also a growing effort in the USA to develop alternatives to incarceration for convicted women with young children. More than half of the states have some form of family-based treatment as an alternative to prison. Summit House in North Carolina was one example of a model residential alternative to incarceration program for women with young children who were convicted of a nonviolent offense (Women's Prison Association 2009). Participation in the program for 12–24 months was a court-ordered condition of probation. With the goal of rehabilitating women while maintaining family bonds, the program included counseling, life and job skills training, substance abuse counseling, supportive housing, and parenting education. Like many alternative community-based programs, Summit House is a nonprofit organization with funding dependent on the current economic climate. Despite being recognized as a model program and evidence of substantial cost savings in terms of both reduced recidivism among the participants and reduced social service costs of dealing with the children, funding was cut substantially, and the program was forced to close in June 2011.

Another example is Drew House, described by Goshin and Byrne (2011), a newly designed program in New York that provides supportive housing to women charged with felony offenses and their children. The women in the program are typically charged with nonviolent offenses, although some women with violent felonies may be eligible if there was no serious injury and the victim agrees to the placement. The program uses a gender-responsive, relational model that promotes independence, and court-mandated conditions typically involve drug testing, educational and vocational training, efforts to find employment, as well as participation in parenting classes. Participation in the program typically lasts between 12 and 24 months. During program participation, the women and up to three children live in their own apartments, paying some or all of the rent. As with the Summit House program, availability of funding is also an issue with Drew House.

Results of Prison-Based and Community-Based Programs

Research on the effectiveness of prison-based and community-based programs for parenting women and their children is very limited. While knowledge about the consequences of the separation of mothers and their children due to incarceration is also limited, research does suggest that both the mother and their children can be adversely affected. For example, children separated from their mothers may experience attachment disorders, mental health problems, and behavioral problems. Others report academic failure and increased levels of criminal involvement among children of incarcerated women (Byrne et al. 2012). Thus, it is important to consider the effectiveness of these types of programs in terms of their impact on the mothers who are participating but also in looking at the long-term impacts on the children.

When sufficient resources are dedicated to prison nursery programs designed for convicted women and their children, results are generally positive, providing an environment that facilitates appropriate child development and allows mothers and their children to develop strong relational bonds. For example, studies in the UK indicate that pregnant inmates are more likely to reduce their levels of smoking, drinking, and drug use when presented with information about healthy behaviors during pregnancy (Edge 2006). Longitudinal research also suggests that mothers are likely to retain custody of their children following participation, and the women also demonstrate reduced recidivism (Byrne et al. 2012). Community-based alternatives to incarceration also appear to produce positive results related to mother-child attachment and recidivism rates (Campbell and Carlson 2012). Additionally, the children who participate with their mothers avoid placement in the foster care system, maternal separation, and the related negative repercussions.

Future Directions

With the increasing attention being paid to pregnant and parenting women in prisons, it has become apparent that there is very little information maintained to track the number of pregnant women, the outcome of those pregnancies, and the numbers of children born or housed within prisons throughout the world. Poso et al. (2010) refer to the "institutional invisibility" of these children and the policies and practices related to children and their parents. In the USA, research has found that correctional administrators are generally unfamiliar with prison nursery programs (Campbell and Carlson 2012). While these administrators expressed some interest in learning more about this type of program, they expressed reservations that would ever be implemented in their state or facility. At a minimum, facilities should consider focusing on ways to facilitate relationships between incarcerated mothers and their children, including creating visitation areas for children that provide a more homelike setting or allowing mothers to create audio-recordings of bedtime stories that they could send to their children (Bastick and Townhead 2008). Other community-based programs face issues with maintaining sufficient resources to serve the women and children who need them.

Most of the parenting women in prison-based programs are low risk and could reasonably serve their sentences in the community. The Women's Prison Association suggests increasing the use of community corrections and alternatives to imprisonment for parenting women. Whether in a prison-based or in a community-based program, these programs should address the needs of both the woman and her family and should offer educational and vocational services as well as education on parenting skills. More generally, there is a great need for additional research and evaluation to assess these programs, including the components that produce the greatest benefit to both the women and their children.

Related Entries

- ▶ Extent of Imprisonment: Global View
- ▶ Health Issues in Prison Reentry Models
- ▶ HIV in the Correctional System
- ▶ Legal Status of Abortion
- ▶ Unintended Effects of Imprisonment

Recommended Reading and References

Alejos M (2005) Babies and small children residing in prisons. Quaker United Nations Office, Geneva (http://www.quno.org)

American Civil Liberties Union (No Date) State standards for pregnancy-related health care and abortion for women in prison. http://www.aclu.org/state-standards-pregnancy-related-health-care-and-abortion-women-prison-map

Association of Women's Health, Obstetric, and Neonatal Nursing (2011) Shackling incarcerated pregnant women. J Obstet Gynecol Neonatal Nurs 40:817–818

Bastick M, Townhead L (2008) Women in prison: a commentary on the UN standard minimum rules for the treatment of prisoners. The Quaker United Nations Office, Geneva

Byrne MW, Goshin L, Blanchard-Lewis B (2012) Maternal separations during the reentry years for 100 infants raised in a prison nursery. Family Court Rev 50:77–90

Campbell J, Carlson JR (2012) Correctional administrators' perceptions of prison nurseries. Crim Justice Behav 39:1063–1074

Carlson JR (2001) Prison nursery 2000: a five-year review of the prison nursery at the Nebraska Correctional Center for Women. J Offender Rehabil 33:75–97

Edge D (2006) Perinatal healthcare in prison: a scoping review of policy and provision. The Prison Health Research Network, Manchester

Feintuch S (2010) New Spanish practice aims to break the cycle among mothers and children. Corrections Today, December. https://www.aca.org/fileupload/177/ahaidar/Feintuch.pdf

Glaze LE, Maruschak LM (2008) Parents in prison and their minor children (NCJ 222984). Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice, Washington, DC

Goshin LS, Byrne MW (2011) Evaluation of a supportive housing alternative to incarceration for women charged with a felony and their children. Report to the Kings County District Attorney

Poso T, Enroos R, Vierula T (2010) Children residing in prison with their parents: an example of institutional invisibility. Prison J 90:516–533

The Rebecca Project for Human Rights (2010) Mothers behind bars: a state-by-state report card and analysis of federal policies on conditions of confinement for pregnant and parenting women and the effect on their children. The Rebecca Project for Human Rights and the National Women's Law Center, Washington, DC

Women's Prison Association (2009) Mothers, infants and imprisonment: a national look at prison nurseries and community-based alternatives. Women's Prison Association, New York

Bayesian Learning

► Bayesian Updating and Crime

Bayesian Updating and Crime

Derek A. Kreager[1] and Ross L. Matsueda[2]
[1]Department of Sociology, Pennsylvania State University, University Park, PA, USA
[2]Department of Sociology, University of Washington, Seattle, USA

Synonyms

Bayesian learning; Cognitive change; Experiential effects; Perceptual updating

Overview

Bayesian updating – the principle that individuals update prior beliefs in light of observed data according to probability rules – has important substantive implications for criminology. Theoretically, this principle may help formalize key causal mechanisms of deterrence, rational choice, social learning, symbolic interactionist, and developmental perspectives of crime. Empirically, recent research linking individuals' perceptions of punishment risk to the objective certainty of arrest has developed formal models drawn from Bayesian updating. Such models help link macrolevel research on aggregate crime rates to microlevel research on individual risk perceptions and self-reported crime. This essay reviews empirical work on Bayesian updating of risk perceptions, points to theoretical and methodological challenges in this area, and outlines future research opportunities for perceptual dynamics and crime.

Introduction

Bayesian updating, or Bayesian learning, has become an increasingly important principle for specifying how human beings change their beliefs in light of new evidence. It has been applied to a variety of substantive topics, including machine learning, language acquisition, artificial intelligence, and dynamic systems. In criminal justice research, Bayesian inference has been applied to jury decision-making, as a rational way of accumulating evidence to reach a verdict (e.g., Robertson and Vignaux 1995). And in criminology, Bayesian updating has primarily been approached from a deterrence perspective, where individuals are argued to follow Bayesian processes when updating their perceptions of formal sanction risk in light of new evidence (Nagin 1998). Indeed, the link between Bayesian updating and deterrence theory is a useful place to begin the current essay.

Bayesian Updating, Deterrence, and Rational Choice

The deterrence doctrine is rooted in a rational decision-making framework. In his seminal work, *Essay on Crimes and Punishment*, the Italian Enlightenment scholar Cesare Beccaria ([1775] 1983:44) presented a utilitarian philosophy of criminal punishment that assumed actors weight pleasures and pains associated with behavior and seek to maximize pleasure and minimize pain. It follows that threatening citizens with punishments that are certain, swift, and proportional to the severity of crime would deter the public from violating the terms of the social contract. Beccaria argued that deterrence requires that punishment must be known in advance by all citizens, and therefore, written laws must clearly stipulate proscribed behaviors and unequivocally designate penalties for transgressors. Beccaria ([1775] 1983:44) further argued that formal sanction by the state is only effective insofar as citizens accurately perceive the cost of crime and apply this information to future offending decisions: punishments "ought to be chosen, as will make the strongest and most lasting impressions on the minds of others, with the least torment."

The deterrence doctrine of the classical school was later formalized by neoclassical economists, who assume that actors maximize expected utility subject to constraints. Drawing on von Neumann and Morgenstern's (1944) expected utility theory of risky decisions under uncertainty, Becker (1968:177) specified a utility function for criminal behavior that included the deterrent effect of punishment:

$$E(U) = (1 - p)\ U(Y) + p\ U(Y - F) \quad (1)$$

where $E(U)$ refers to expected utility, p is the probability of getting caught as perceived by the criminal, $(1 - p)$ is the perceived probability of getting away with crime, Y is returns to crime (both monetary and psychic), and F is the penalty. This utility function describes two states: getting caught or getting away with crime. When $p = 1$, the criminal expects to get caught with certainty and, therefore, $E(U) = U(Y - F)$; that is, the expected utility of crime is equal to the utility of the perceived returns to crime minus the punitive sanction (assuming that the criminal keeps her booty when caught). When $p = 0$, the criminal expects to get away with certainty and, therefore, $E(U) = U(Y)$; that is, the expected utility of crime is equal to the utility of the returns to crime. A person is assumed to commit crime when the expected utility of crime is higher than the expected utility of alternative legal pursuits (Taylor 1978). Moreover, Eq. (1) implies that, all else being equal, an increase in p, the perceived certainty of punishment will reduce the utility of crime, and thereby the probability of crime.

Both the classical and neoclassical models' emphases on the actors' perceived probabilities of punishment underscores the importance of information for deterrence and for decision-making in general. Because individuals' perceptions of sanction risk are not exogenously determined but rather are endogenously produced through social interaction, a rational choice theory of deterrence and crime requires a theory of information. Such a theory would specify how information about sanction risks are communicated and disseminated to individuals. Here, Bayesian updating can provide a mechanism for risk communication which is consistent with rational choice theory. Understanding Bayesian updating first requires an understanding of the basics of Bayesian statistical inference.

Bayesian Inference and Updating

Based on the probability theorem posthumously published by Thomas Bayes (1701–61), Bayesian updating refers to the general principle that subjective beliefs should change given exposure to new evidence (Bayes and Price 1763). Bayesian updating provides a rational and principled way of combining prior beliefs with new evidence using Bayesian inference. It begins with two assumptions: (1) Subjective hypotheses about the world can be expressed as degrees of belief, which in turn, can be expressed in terms of probabilities ranging from 0 to 1. (2) Human beings are able to use probability distributions to represent uncertainty in inference. Given these assumptions, actors can use probability theory to compute the degree of belief of a hypothesis, h_i, given some observed data d, where h_i is a member of the set of mutually exclusive and exhaustive hypotheses $\mathcal{H}$. Belief in h_i prior to observing the data d is defined as the *prior probability*, denoted as $P(h_i)$. The probability of observing datum d given that h_i is true is the *likelihood*, denoted $P(d|h_j)$. Bayes' rule can then be used to derive our belief in h_i after observing the data, which is the posterior probability denoted $P(h_i|d)$:

$$P(h_i|d) = \frac{P(d|h_i)P(h_i)}{\sum_{h_j} P(d|h_j)P(h_j)} \quad (2)$$

where $h_j \in \mathcal{H}$. The denominator is simply the sum of all possible hypotheses under consideration

Bayesian Updating and Crime, Table 1 Hypothetical Burglar Arrest Perceptions Given Observed Arrests

Prior probability $P(h_i)$	Likelihood $P(d/h_i)$	Posterior probability $P(h_i/d)$
P(certain) = .30	P(.80|certain) = .80	P(certain|.50) = .522
P(50/50) = .30	P(.80|50/50) = .60	P(50/50|.50) = .391
P(get away) = .40	P(.80|get away) = .10	P(get away|.50) = .087

which ensures that the posterior probabilities of all hypotheses sum to one. This equation describes a rational updating process in which new evidence is combined with prior beliefs to yield a new subjective belief. The posterior probability $P(h_i|d)$ is equal to the likelihood of the data given h_i is true $P(d|h_j)$ times the prior probability $P(h_i)$.

A simple example helps to illustrate how this equation produces Bayesian learning (see Table 1). For simplicity, assume there are only three prior hypotheses about the risk of arrest for burglary: certain arrest ($P = 1.0$), 50/50 ($P = .50$) and certain to get away ($P = 0$). The prior probability is .30 for certainty, .30 for 50/50, and .40 for getting away. Thus, given the opportunity, an actor would be expected to engage in burglary, since the highest probability is associated with getting away with the crime. The actor then observes new information or data, in which four of five burglars are arrested for the burglaries. Given the new data (that 80 % of burglars are arrested), the probabilities for our three hypotheses are as follows: .80 for certain arrest, .60 for 50/50, and .10 for getting away.

The updated probabilities become .522 for certain, .391 for 50/50, and .087 for get away (Table 1, column 3). Note that the highest subjective probability for the posterior is now getting caught (.552). Thus, all else being equal, after updating, the actor would be expected to refrain from burglary.

Bayesian inference assumes that the observed evidence or data are generated by some underlying process or mechanism, which has crucial implications for making inferences. The likelihood is based on a probability model of the mechanism by which the data were generated. In this way, Bayesian learning is a way of evaluating different hypotheses about the underlying process generating the data and making predictions about the most likely ones. For example, if the data on arrested burglars were generated from a random sample of the population of all burglars, and one views oneself as an average burglar, applications of Bayesian inference would be straightforward. However, if the data on arrested burglars were generated from a sample of very unskilled novice burglars, one would draw a different inference. Although the observed data are identical, this difference in the generation of the data will produce distinct likelihoods, altered posterior distributions, and different inferences. Thus, in applying Bayesian learning to substantive applications, careful attention must be paid to the generative process producing the data, a point returned to later in the essay.

The Heuristic Critique

Bayesian updating, as well as other rational principles of learning, has been subject to theoretical and empirical critique. Indeed, the emerging discipline of behavioral economics has as its focus the study of systematic ways human beings depart from rationality. Much of this research derives from the important work of Tversky and Kahneman (1974), who conducted a series of ingenious experiments that showed actors departing from rational updating in systematic ways, which they termed "cognitive heuristics" or "cognitive shortcuts." The assumption is that human beings have a limited ability to process information cognitively, and therefore, must rely on cognitive shortcuts. Based on results of their social experiments, Tversky and Kahneman (1974) outline four heuristic rules that individuals use to form perceived risks, and which could bring about departures from a Bayesian learning process. The first, representativeness, refers to a tendency to rely on stereotypes, while ignoring information on population distributions. For example, people are likely to overestimate the probability a mother is black when told that she is a teenage mother, thereby forgetting or ignoring the extent to which whites are disproportionately represented in the population. In the case of certainty of sanction, individuals are likely to rely on stereotypes depicted in the media, in which criminals are caught and arrested. Research suggests that naïve individuals with no experience with the criminal justice system tend to overestimate the likelihood that they will be arrested if they commit crimes. Tittle (1980, p. 67) termed this "the shell of illusion."

A second heuristic, availability, refers to the tendency to update based only on information that is easily or quickly retrieved from memory (Tversky and Kahneman 1974). Rare and mundane events are less likely to be recalled than common and vivid events. Moreover, the two could interact: events that are vivid, salient, and dramatic – as well as rare – could be brought to mind quicker than other events. The result can be bias due to differences in ease of retrieval, as vivid experiences or events trump other sources of information. For example, a dramatic event, such as being arrested for a crime, may swamp other sources of information in an individual's estimate of rearrest. A third heuristic, anchoring, refers to a failure to adjust initial probability estimates sufficiently in light of new information. For example, when individuals are given an initial probability estimate that is arbitrary or even randomly assigned, followed by additional accurate information with which to update, their new estimates are consistently biased in the direction of the initial estimate (Tversky and Kahneman 1974). The estimates are anchored at the initial value, rather than adjusted properly in light of the new information. Applied to updating perceived risk of formal sanction, anchoring could lead to an effect opposite from that of availability. Individuals may fail to adjust risk estimates appropriately in light of new information and instead anchor on their baseline estimates. Tversky and Kahneman (1974) mention a fourth departure from Bayesian learning: the gambler's fallacy. Stated simply, the gambler's fallacy occurs when one assumes that a departure from what happens in the long run will be corrected in the short run. For example, if seven coin flips in a row have come up tails, one might think one is "due" for a heads. Applied to updating perceived risk, this might cause individuals who continuously get away with crime to think they are "due" for an arrest, or those who experience a string of arrests think they are "due" to get away with crime (Pogarsky and Piquero 2003).

Although experimental evidence suggests that actors do depart systematically from rational updating, some scholars argue that the departures are relatively small in magnitude, given the overall decision-making process. Thus, Bayesian updating of perceived sanction risk may be present net of cognitive heuristics. The next section summarizes related research on deterrence and updating risk perceptions.

Research on Deterrence and Perceptual Updating

Early empirical tests of Becker's model used statistical models of aggregate crime rates, focusing on the deterrent effects of *objective* risk of punishment, using for example, risk of imprisonment (measured by imprisonment per capita) or risk of arrest (measured by arrests per crimes reported to police). Most notably, Ehrlich (1973) found deterrent effects of risk of imprisonment, but scholars criticized his simultaneous equation models for using implausible solutions to the identification problem – the problem of finding good instrumental variables to identify reciprocal effects between rates of imprisonment and rates of crime – such as assuming population age, socioeconomic status, and region have zero direct effects on crime (Nagin, 1978). Later work using aggregate data includes more plausible instrumental variables to address the problem of reverse causality and found deterrent effects. For example, Levitt (1997) employed the timing of mayoral elections as an instrument for the number of police per capita, under the assumption that such elections should have a direct effect on investment in the police force (as newly elected mayors seek to crack down on crime), but only an indirect effect on crime. For a review of aggregate deterrence research, see Nagin (1998) and Durlauf and Nagin (2011).

Tests of the deterrence hypothesis using aggregate data assume that aggregate clearance rates are good proxies for individuals' perceptions of formal sanction risk, which is the key explanatory variable. A few economists remain uninterested in directly measuring individual risk perceptions and instead assume that the models need not describe perceptual or cognitive processes so long as actors behave "as if they are rational" and the models make good predictions. By contrast, most scholars view the measurement issue as an empirical question and welcome research on the relationship between aggregate rates of objective certainty of punishment and perceptions of the risk of sanction. Subjective expected utility models replace the objective certainty of sanction with a probability distribution of subjective probabilities. Such models are still rational models because the statistical mean of the subjective probability distribution is assumed to fall on the value of the objective probability (Nagin 1998). Empirical research from a subjective expected utility framework uses survey methods to measure perceived risk of punishment directly from respondents, rather than inferring it from behavior through the method of revealed preferences (e.g., Kahneman et al. 1997). These studies of individuals have the potential of linking subjective risk of punishment measured with survey data and objective risk of punishment measured with police clearance rates.

Early perceptual deterrence research by sociologists used cross-sectional data, eliciting self-reports of delinquent behavior and perceptions of risk of arrest in the same questionnaire or interview. These studies generally found small but significant deterrent effects for certainty but not for severity. That is, youth who perceive a high probability of arrest for minor offenses (like marijuana use and petty theft) tend to report fewer acts of delinquency. Such research was immediately criticized for using cross-sectional data in which past delinquency is regressed on present perceived risk, resulting in the causal ordering of the variables contradicting their temporal order of measurement (Paternoster 1987). These criticisms led to panel studies in which individuals are followed over short periods of time and both risk perceptions and self-reported crime are remeasured repeatedly.

The initial panel studies surveyed two waves of undergraduate students and estimated cross-lagged panel models. Here, self-reported delinquency is regressed on lagged delinquency plus lagged perceived risk and then perceived risk is regressed on lagged risk plus lagged delinquency. These studies found that both perceived risk and delinquency were fairly stable over a 6-month or 12-month period. Moreover, they found little evidence of a deterrent effect of the certainty of sanctions: net of lagged delinquency, lagged perceived risk of punishment was not significantly related to delinquent behavior. They did find support for the opposite effect: net of lagged risk, lagged delinquency exerted significant effects on perceived risk. Paternoster et al. (1982) called this an "experiential effect," because it suggested that youth who experienced getting away with crime – arrest is fairly rare for the nonserious self-reported delinquent acts measured – reported lower risk of arrest. These findings were replicated on other two-wave panels of students. Furthermore, the results – a strong experiential effect and weak deterrent effect – were replicated in samples of disadvantaged adults in several cities at risk of serious crimes (Piliavin et al. 1986). The experiential effect was the first important empirical finding about the formation of individual risk perceptions. An experiential effect is consistent with a Bayesian updating model insofar as respondents have not been arrested between waves. If they had been arrested, that information must be included in any updating model.

Paternoster et al. (1985) found, for their two-wave panel of undergraduate students, experiential effects for minor property offender. They also found that students who were arrested between waves had higher perceptions of arrest risk. Horney and Marshall (1992) interviewed incarcerated felons and obtained their retrospective self-reported arrests and offenses. They computed, for a variety of offenses, the ratio of arrests to offenses and found that arrest ratios strongly related to perceived risk of punishment. This finding is consistent with a model of Bayesian updating. Subsequent research used prospective longitudinal designs to examine Bayesian learning of perceived risk. Pogarsky, Piquero, and Paternoster (2004) focused on changes in risk perceptions among a sample of high school students surveyed at the 10th and 11th grade. They found that students who experienced an arrest between the waves increased their perceptions of arrest certainty. This effect was most pronounced for offenders with an initially low-risk perception, which the authors attribute to these offenders having more room for change (i.e., a floor effect). Additionally, they found that individuals who reported higher peer offending had lower perceptions of risk, presumably because those friends avoided arrest. Stafford and Warr (1993) termed such a process vicarious punishment avoidance. This peer effect was greatest for non-offenders, consistent with the idea that naive individuals have a "shell of illusion" (Tittle 1980:67) regarding police effectiveness that is eroded through vicarious experiences.

Two recent studies incorporate Bayesian updating of perceived risk into deterrence models of subsequent offending. Matsueda, Kreager, and Huizinga (2006) examined changes in risk perceptions and offending with longitudinal data of adolescents from high-risk Denver neighborhoods. They found strong support for a Bayesian learning hypothesis for both property and violent crime: lagged ratios of arrest per offense – which they termed, "experienced certainty" – were monotonically (positively) associated with perceived risk of arrest. In addition, respondents' unsanctioned offenses were monotonically (negatively) related to perceived risk of arrest. They also found that perceptions of peer delinquency were negatively associated with perceived risk (see also Pogarsky et al. 2004). Each of these findings is consistent with Bayesian updating. Finally, Matsueda et al. (2006) also estimated a rational choice model of crime, finding that perceived risk of arrest was significantly associated with subsequent offending. Specifically, they found that, on average, a 10 % increase in perceived certainty of arrest was associated with a 3 % decrease in theft and violence. Lochner (2007) reported similar findings using data from the National Longitudinal Survey of Youth 1997 and National Youth Survey. He found that respondents in both surveys updated their risk perceptions in ways consistent with Bayesian expectations. Offenders who got away with crime reported lower risk of arrest, and those who got arrested reported higher risks. Lochner (2007) also found that naïve non-offenders held the highest perceptions of risk certainty. When comparing perceived certainty to actual offending, he strikingly found almost the exact same pattern as Matsueda et al. (2006): using the NLSY97 data, he found that a 10 % increase in perceived certainty of arrest was associated with a 3 % decrease in theft. The similarity of these results builds confidence in the deterrent effect of perceived certainty of arrest on future offending.

The most recent study of Bayesian updating, conducted by Anwar and Loughran (2011), examined serious juvenile offenders enrolled in the Pathways to Desistance Study. They found that offenders in their sample appeared to update their risk perceptions following a Bayesian model. Offenders who committed crimes and were arrested for them reported an average of 6.3 % higher-risk perceptions than those who committed crime but were not arrested. This is an important finding, as it suggests that risk perceptions remain malleable even among serious offenders, a population often written off as irrational or impulsive and thus outside policy intervention.

The weight of recent evidence thus supports Bayesian expectations when applied to updated risk perceptions and personal experiences of offending and sanctioning. As mentioned above, one might also ask if subjective sanction perceptions are rooted in objective rates of arrest and punishment. For example, does increased police arrest activity alter offenders' subjective arrest perceptions? Interestingly, Lochner (2007) observed that the risk perceptions of his NLSY97 sample were fairly unresponsive to county-level arrest-per-crime rates. Combined with his findings of individual experiential effects, the lack of a contextual effect suggests that proximate conditions are more important determinants of perceptual change than macro-contextual conditions. Kleck et al. (2005), in a phone survey of respondents in 300 counties, also found little correspondence between individuals' estimates of police clearance rates and the actual clearance rates in those counties. However, Apel, Pogarsky, and Bates (2009) did find an association between changes in a school's disciplinary regime and students' perceptions of discipline, suggesting that individuals' risk perceptions are responsive to contextual conditions in at least some instances. The limited and mixed findings in this area suggest that further research is required to connect objective sanction risk in a given geographic area to individual risk perceptions.

Extralegal Benefits and Costs

Research of the linkage between perceptual change and offending should also extend beyond formal sanctions. Theory and qualitative evidence suggest that other costs and rewards are equally, if not more so, related to criminal decision-making. As mentioned previously, Becker's (1968) criminal utility model includes both subjective costs and benefits in criminal decision-making. This utility calculus lies at the heart of rational choice theories of crime (Clarke and Cornish 1985). The relative neglect of crime's perceived returns, and how these perceptions are adjusted over time, is a serious omission both for etiological and policy reasons. Understanding crime's perceived benefits will likely provide valuable insights for understanding criminal motivation, while also pointing to potential interventions that downwardly adjust individuals' positive perceptions of crime over time.

In his phenomenological examination of violent and property crime, Katz (1988) provided perhaps the most detailed account of crime's "seductive" psychic and social rewards. He explored the "sneaky thrills" of shoplifting and the social status associated with the "badass" gang member. Such perceived benefits are intimately linked to the criminal event and may override the certainty, celerity, and severity of perceived punishment. Indeed, Matsueda et al. (2006) found that the perceived excitement and "coolness" of offending were stronger predictors of crime than perceptions of arrest. Missing from their analyses, and generally overlooked in empirical analyses of criminal perceptions, are the origins and dynamics of crime's perceived benefits as predicted by Bayesian learning.

Such analyses would appear particularly relevant for understanding individual trajectories of drug use. The objective risks of apprehension for drugs are low and likely swamped by perceptions of their psychic and social returns. In his classic *Becoming a Marihuana User*, Howard Becker (1953) described the learning process associated with marijuana initiation. In interviews of marijuana-smoking Jazz musicians, he found that users often entered their first marijuana experience uncertain of the drug's effects. Moreover, marijuana's psychopharmacological properties may result in potentially ambiguous physical effects, such as hunger, paranoia, dizziness, and euphoria, or no effects at all. Becker argued that the presence of more knowledgeable peers help the initiate translate these effects into a pleasurable experience worth repeating, or these peers may push the initiate to smoke again if he or she experienced no discernible effects the first time. In these ways, perceptions of fear and uncertainty are updated into fun and excitement. The change in marijuana's perceived rewards upon initiation thus provides a particularly fruitful context for studying Bayesian updating.

Costs other than formal sanctions may also be important for understanding criminal decision-making. McCarthy and Hagan (2005) argued that the proximal fear of physical harm likely overrides perceptions of punishment for offending decisions. In their qualitative and quantitative study of street youth living in Toronto and Vancouver, they found that perceptions of danger deterred youth from theft, drug dealing, and prostitution. Interestingly, they found little evidence for the threat of legal sanctions, but, consistent with Matsueda et al. (2006), they did find that perceived excitement predicted theft and drug dealing. Note, however, that the cross-sectional nature of their data did not allow them to directly address the Bayesian updating hypothesis. More work is required to test if repeated exposure to crime and delinquency increases or decreases perceived danger.

Life-Course Transitions and Cognitive Change

The growth in life-course theories and research has opened up new avenues for understanding Bayesian updating processes. An axiom of life-course perspectives is that life events can meaningfully alter individual behavioral trajectories. Although the dominant explanation for how events become "turning points" in criminal trajectories is through external social control mechanisms (Sampson and Laub 1993), cognitive change is increasingly the focus of life-course criminology. Giordano et al. (2002) provided a symbolic interactionist theory that connects life-course transitions, cognitive transformation processes, and criminal desistance. Their central premise was that desisters are likely to reflect on their past and present circumstances and create new conventional identities. Life-course transitions, such as marriage, incarceration, and parenthood, then become "hooks for change" in this cognitive process. Interpreted from a Bayesian perspective, significant life events should provide new evidence by which prior perceptions are updated to shape future behavior. As Maruna (2001) points out, however, whether such life events positively or negatively affect self-perceptions is extremely difficult to predict. For example, a drug addict may interpret a friend's overdose as (1) the final impetus needed to "get clean," (2) unrelated to their own fate, or (3) a reason to use drugs to manage the resulting grief. The subjectivity of experience, and the meaning of experiences derived in contextualized social interactions, complicates an understanding of perceptual change and use of formal models of Bayesian inference. Identifying the origins of such heterogeneity is challenging but worth future investigation.

Conclusion and Future Directions

Under many guises, Bayesian updating remains an influential concept for criminological theory and research. Within the deterrence literature, researchers have consistently documented the experiential effect of crime on reduced risk perceptions, while also documenting an association between sanction perceptions and future offending. Rational choice studies have extended the study of perceptual dynamics to extralegal domains, including changes in perceived excitement, social status, monetary rewards, and fear of physical danger. More recently, life-course research has explored the impacts of life events on cognitive changes and desistance processes. All of these research strands continue to produce contributions for both criminology and the understanding of individual perceptual dynamics over time.

The obstacles facing perceptual research also provide opportunities for scientific advancement. As mentioned previously, a continuous challenge facing research in this area is disentangling experiential from perceptual effects. For example, an individual's experiences in crime should change his or her perceptions of its costs and benefits, which should then impact his or her probability of future crime. Longitudinal designs are clearly necessary for distinguishing these reciprocal processes. One attractive strategy might be to design an experiment where individuals are randomly exposed to sanction risk information, such as local arrest statistics, to examine if their pretest arrest perceptions and behavioral intentions are updated upon receiving new information. Another potentially fruitful research avenue would focus on perceptual changes surrounding criminal onset. Prior to initiation, individuals must rely on less-than-optimal information sources – such as peers, the media, or experience with related behaviors – to formulate expectations for a novel behavior. But the inadequacy of prior information may result in actors initiating a behavior under extreme uncertainty. The situational contexts of initiation, and the physical experience of that event, are then highly influential in revising perceptions from uncertainty to increased clarity. The large amount of perceptual change potentially associated with initiation makes understanding this event critical for our knowledge of Bayesian updating. Although measuring the perceptions and contexts at the point of initiation would be difficult, it holds tremendous potential for broadening our insights of dynamic risk perceptions.

Future research is also required to order individuals' perceptions of crime's costs and benefits. Prior research has demonstrated that crime's perceived extralegal costs and benefits often exceed the effects of perceived formal sanctions in predicting future criminal behavior. What is needed is research that identifies the relevant subjective perceptions of crime and helps rank order or weight these perceptions while also examining how such orderings may change over time. For example, the perceived social benefits of drug use may be a strong predictor of drug initiation, but upon initiation, such perceptions may be overridden by the perceived fun and excitement associated with getting "high." Further, perceptions of pleasure may fade over time and be replaced by the perceived costs of heavy use. Bouffard (2007) has taken steps toward identifying preference orderings using subject-generated perceptions of criminal consequences, but more work is required to understand the dynamics of such orderings given individual experience.

More broadly, greater theorizing is necessary for the generative process of risk perceptions. Aside from personal exposure and vicarious experiences through peer networks, few studies have focused on the origins of offending perceptions and their changes over time. It is clear that non-offenders' perceptions differ greatly from offenders' perceptions, and that individuals transitioning from the former to latter typically experience substantial perceptual shifts. However, more research should center on the information sources associated with each status and how naïve perceptions are updated or negated with offending decisions or experiences. Such investigations have substantial implications for interventions aimed at manipulating perceptions to prevent initiation (general deterrence), or increasing offenders' perceived risks to prevent future offending (specific deterrence).

Finally, research should also continue to isolate the effects of salient life events (e.g., marriage, parenting, employment, military service, incarceration, etc.) on changing risk perceptions. Symbolic interactionist perspectives of desistance (Maruna 2001; Giordano et al. 2002) have taken strides in presenting theoretical models and qualitative evidence for how life events relate to cognitive transformation and desistance, but further quantitative evidence is needed to validate this line of inquiry and calibrate risk perceptions so that they may be formally analyzed using Bayesian inference. Prospective studies of sanction perceptions and prominent life-course transitions would add to an understanding of desistance while also testing the Bayesian learning hypothesis. Such investigations will provide greater clarity to the cognitive processes and decision-making associated with "knifing off" a criminal past and the construction of a conventional future (Maruna 2001).

Related Entries

- ▶ Certainty, Severity, and Their Deterrent Effects
- ▶ Cognitive Forensics: Human Cognition, Contextual Information, and Bias
- ▶ Cognitive/Information Processing Theories of Aggression and Crime
- ▶ Deterrence: Actual Versus Perceived Risk of Punishment
- ▶ Deterrence
- ▶ Economic Theory of Criminal Behavior
- ▶ Integrating Rational Choice and Other Theories
- ▶ Rational Choice, Deterrence, and Crime: Sociological Contributions
- ▶ Rational Choice Theory

Recommended Reading and References

Anwar S, Loughran TA (2011) Testing a Bayesian learning theory of deterrence among serious Juvenile offenders. Criminology 49:667–698

Apel R, Pogarsky G, Bates L (2009) The sanctions-perceptions link in a model of school-based deterrence. J Quant Criminol 25:201–226

Bayes T, Price R (1763) An essay towards solving a problem in the doctrine of chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. Philosophical Transactions (1683–1775) (53): 370–418

Beccaria C (1764[1985]) Essay on crimes and punishment (trans: Paolucci H). MacMillan, New York

Beccaria C ([1775]1983) An essay on crimes and punishments. Branden Press, Brookline

Becker HS (1953) Becoming a Marihuana user. Am Sociol Rev 59:235–242

Becker GS (1968) Crime and punishment: an economic approach. J Polit Econ 76:169–217

Bouffard JA (2007) Predicting differences in the perceived relevance of crime's costs and benefits in a test of rational choice theory. J Offender Ther Comp Criminol 51:461–485

Clarke RV, Cornish DB (1985) Modeling offenders decisions: a framework for research and policy. In: Tonry M, Morris N (eds) Crime and justice: an annual review of research, vol 6. University of Chicago Press, Chicago, pp 147–185

Durlauf SN, Nagin DS (2011) Imprisonment and crime: can both be reduced? Criminol Public Policy 10:13–54

Ehrlich I (1973) Participation in illegitimate activities: a theoretical and empirical investigation. J Polit Econ 81:521–565

Giordano PC, Cernkovich SA, Rudolph JL (2002) Gender, crime, and desistance: toward a theory of cognitive transformation. Am J Sociol 107(4):990–1064

Horney J, Marshall IH (1992) Risk perceptions among serious offenders: the role of crime and punishment. Criminology 30:575–594

Kahneman D, Wakker PP, Sarin R (1997) Back to Bentham? Explorations of experienced certainty. Quart J Econ 112:375–406

Katz J (1988) Seductions of crime: moral and sensual attractions in doing evil. Basic Books, New York

Kleck G, Sever B, Li S, Gertz M (2005) The missing link in general deterrence research. Criminology 43:623–660

Levitt SD (1997) Using electoral cycles in police hiring to estimate the effect of police on crime. Am Econ Rev 87:270–290

Lochner L (2007) Individual perceptions of the criminal justice system. Am Econ Rev 97:444–460

Maruna S (2001) Making good: how ex-convicts reform and rebuild their lives. American Psychological Association, Washington, DC

Matsueda RL, Kreager DA, Huizinga D (2006) Deterring delinquents: a rational choice model of theft and violence. Am Sociol Rev 71:95–122

McCarthy B, Hagan J (2005) Danger, deterrence, and the decision to offend. Soc Forces 83:1065–1096

Nagin D (1978) Crime rates, sanction levels, and constraints on prison population. Law Soc Rev 12:341–366

Nagin D (1998) Criminal deterrence research at the outset of the twenty-first century. Crime Justice 23:1–42

Paternoster R (1987) The deterrent effect of the perceived certainty and severity of punishment: a review of the evidence and issues. Justice Quart 4:173–217

Paternoster R, Saltzman LE, Waldo GP, Chiricos TG (1982) Perceived risk and deterrence: methodological artifacts in perceptual deterrence research. J Crim L & Criminology 73:1238–1258

Paternoster R, Saltzman LE, Waldo GP, Chiricos TG (1985) Assessments of risk and behavioral experience: an exploratory study of change. Criminology 23:417–436

Piliavin I, Gartner R, Thornton C, Matsueda RL (1986) Crime, deterrence, and rational choice. Am Sociol Rev 51:101–119

Pogarsky G, Piquero AR (2003) Can punishment encourage offending? Investigating the 'resetting' effect. J Res Crime Delinq 40:95–120

Pogarsky G, Piquero AR, Paternoster R (2004) Modeling change in perceptions about sanction threats: the neglected linkage in deterrence theory. J Quant Criminol 20:343–369

Robertson B, Vignaux GA (1995) Interpreting evidence: evaluating forensic science in the courtroom. Wiley, New York

Sampson RJ, Laub JH (1993) Crime in the making: pathways and turning points through life. Harvard University Press, Cambridge

Stafford M, Warr M (1993) A reconceptualization of general and specific deterrence. J Res Crime Delinq 30:123–135

Taylor JB (1978) Economic models of criminal behavior: a review. In: Heineke JM (ed) Economic models of criminal behavior. New Holland, Amsterdam, pp 35–81

Tittle CR (1980) Sanctions and social deviance: the question of deterrence. Praeger, New York

Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. Science 185:1124–1131

Von Neumann J, Morgenstern O (1944) The theory of games and economic behavior. Princeton University Press, Princeton

BCS

▶ British Crime Survey

Behavioral Health Courts

▶ Mental Health Courts

Behavioral Investigative Advice

Lee Rainbow[1], Adam Gregory[1] and
Laurence Alison[2]
[1]Serious Organised Crime Agency, London, UK
[2]University of Liverpool, Liverpool, UK

Synonyms

Crime scene profiling; Criminal investigative profiling; Criminal personality profiling; Criminal profiling; Investigative psychology; Offender profiling; Psychological profiling

Overview

Behavioral Investigative Advice represents the UK approach to the provision of advice and support to major crime investigations from a behavioral science perspective. Although the foundations of Behavioral Investigative Advice are built upon such precursor influences as criminal investigative analysis, offender profiling, and investigative psychology, the emergence of Behavioral Investigative Advice should be recognized more as a qualitatively distinct *profession*, than a variation of technique.

All such behavioral support to UK policing is delivered exclusively by a nationally funded cadre of professional Behavioural Investigative Advisers (BIAs) who have significantly widened the scope from the traditional, specific view of "offender profiling" to a discipline which now encompasses a broad range of scientifically based yet pragmatic activities related to supporting police investigations, based on replicable, transparent, and valid knowledge and research.

While elements of this approach and philosophy have begun to evolve within other (predominantly European) countries, it is emphasized that the formal recognition of Behavioural Investigative Advisers is a UK-specific initiative and should not be misinterpreted as any form of internationally recognized standard.

Brief Historical Overview

The term "offender profiling" was first regularly used by members of the FBI Behavioural Science Unit who defined it as the process of drawing inferences about a suspect's characteristics from details of his or her actions exhibited during the commission of a crime.

Critical review of these initial efforts however drew attention to the lack of systematic basis and empirical evidence supporting the claims made. David Canter, a UK psychologist, was highly influential in advocating a more robust scientific psychological approach to offender profiling, and in recognizing the broader utility of providing investigative support to serious crime enquiries from a behavioral science perspective.

In 1992, the Metropolitan Police service commissioned research into the investigative usefulness of such offender profiling, following concerns that despite over 200 British Police inquiries utilizing such services in the preceding decade, no reliable or valid scientific assessment has been conducted to evaluate its usefulness.

While the explicit focus of this research was directed toward the methodologies employed and the resulting usefulness of profiling, it also provided important signposts toward the challenges facing this emerging discipline within the UK, most significantly in the observation that no governing body for the regulation of professional or ethical standards existed, and that there was no formally recognized program for the training of UK practitioners in profiling techniques.

However, it was not until 2001 that a significant watershed in such endeavors occurred, most conspicuously evidenced by the replacement of the term "offender profiler" with "Behavioural Investigative Adviser" (BIA). A set of common standards and working conditions were introduced, making explicit the responsibilities of BIAs, including: administrative protocols, commitment to producing written reports, agreement to have work annually audited and evaluated, and acceptance that the results of such audits would determine the retention or removal of their authority to provide Behavioral Investigative Advice within the UK.

The introduction of these new working conditions had a twofold effect. In making explicit the need to document all advice provided, not only were quality assurance issues addressed, but the true extent of behavioral science provision to UK investigations could be known for the first time, allowing for detailed qualitative and quantitative analyses of all aspects and supply and demand. The availability of such management information led to a formal strategic review of UK Behavioral Investigative Advice in 2010, which provided a clear overview of the current status of the discipline, clarified current and potential future demand, made explicit the current and anticipated threats to the effective delivery of the service and considered regulatory and governance issues, including minimum performance standards. It concluded with a proposed model of management and delivery of Behavioral Investigative Advice which significantly changed the landscape of such activity within the UK.

Today, all behavioral science contributions to major crime investigations are delivered exclusively by a nationally funded cadre of full-time professional Behavioural Investigative Advisers (BIAs).

International Perspectives

While one of the key successes of the UK approach to Behavioral Investigative Advice is the effective integration of the BIA into the investigative arena, the UK model differs significantly from many of its international contemporaries in the philosophy of individual specialism within a multidisciplinary approach. Whereas in many countries, the "profiler" assumes the role of investigator, whose behavioral science knowledge and expertise represents but one facet to an overall "investigative consultancy" role, in the UK, such expertise is far more discrete. The BIA is concerned only with contributions based on behavioral/psychological principles, and integrates such expertise with other specific experts, including those focused exclusively on investigative doctrine. An additional benefit of such approach is that unburdened by the same training and experiences of police investigators, the BIA has the benefit of a more objective and "different" view of the case. While approaches in which seasoned investigators are "trained" in behavioral science have their own merit, the UK's approach favors the behavioral scientist gaining a sufficient and relevant understanding of the investigative process, but delegating the more "traditional" investigative advice to an "expert" investigator.

Investigative Contribution

Behavioral Investigative Advice has the potential to contribute to many aspects of the investigative process and may take many forms throughout the life of the enquiry. While all of the products and services available offer tactical or strategic solutions in their own right, all are underpinned by a broader philosophy of adding value to the decision-making of the Senior Investigating Officer (SIO), through an enhanced understanding of the offense and offender from a perspective different from that routinely employed within major crime investigation teams. Such differences in perspective can be broadly characterized as evidence (SIO) versus understanding (BIA), although both are directed at supporting the single goal of case resolution. It is this additional perspective and associated expertise which should be recognized as the critical success factor of Behavioral Investigative Advice.

While the prediction of those facets of an unknowns criminal's background which are amenable to investigative action serves as a great asset in major crime enquiries, this does not represent the only contribution of contemporary BIAs. Analysis of the type and scope of reports produced by the national cadre of UK BIAs reveals a steady decline in producing "offender profiles," from over 60 % of all operational activity at the launch of the unit over a decade ago to a little over 10 % in 2009. Such figures do not however imply that such inference generation regarding offenders is no longer undertaken on a routine basis, but rather that such inferences often form the basis of a more bespoke service to SIOs, tailored to their specific investigative needs.

However, the implicit assumption that all the additional services provided by BIAs are dependent upon such prescriptive predictions concerning the unknown offender is also inaccurate. Many of the products and services, as outlined below, are wholly independent of any such speculative inferences regarding the type of unknown person who has committed the offense in question.

Crime Scene Assessment and Hypothesis Generation

Critical to the provision of many BIA products and services is a fundamental understanding of the offense, and hence the offender(s), from a behavioral perspective. This is achieved through crime scene assessment (CSA) and hypothesis generation. This involves a thorough examination of the criminal event and generating hypotheses based upon the available information. Support for or against each of the possible hypotheses is then forwarded with reference to psychological theory, relevant research findings, and experiential knowledge, with information gaps identified that will further enhance the process. The benefits of such an approach are that specific hypotheses regarding the offense can be tested in a systematic, reasoned, and objective fashion, based upon sound supporting rationale. Such a methodology is consistent with, and provides a tangible product of the investigative philosophy promoted within national policing guidance. Recent practice advice advocates the application of scientific principles and methods as core investigative doctrine to be adopted across the UK police service (ACPO 2005).

Offense Linkage Analysis

In the absence of any physical evidence linking a number of crimes, the contribution of a behavioral analysis may be significant. Research into behavior exhibited by offenders during the commission of their crimes has led to a greater understanding in consistency and variability of offenders when committing a series of offenses. Through national mandates to collect and analyze a range of sexually motivated offenses throughout the UK, BIAs have access to the largest collection of data of its kind in the UK and one of the largest in Europe via the ViCLAS (Violent Crime Linkage Analysis System) database (the collection criteria cover homicide, serious sexual assault and abduction offenses, including attempts). The creation of such large datasets allows for validation of initial hypotheses regarding linkage, as well as providing statistics with regard to the frequency of individual behaviors, and more significantly, combinations of behaviors.

Predictive Profiling

Drawing inferences in relation to a particular offender on the basis of a comprehensive crime scene assessment is a process commonly referred to as predictive (or offender) profiling. However, it is important to recognize that in contrast to its media portrayal, the focus of modern day predictive profiling is very much on investigative utility rather than psychological interest. Perhaps the most widely held misconception regarding the role of the contemporary BIA surrounds the insistence that speculative predictions concerning the unknown criminal's *personality* are routinely made. They are not. Not only is such activity lacking scientific reliability and validity, in itself it serves very little, if any purpose in assisting police officers to identify the offender. Despite the tabloid appeal of psychological musings over traits such as narcissism, misogyny, introversion, and the like, the lack of any police database recording such facets of the criminal population makes them somewhat limited in terms of investigative utility.

A BIA will endeavor to make accurate assessments in relation to objective and verifiable elements of an offender's background. Consideration will be given to the likely age of the offender, whether he is likely to have previous police convictions and if so what these may be, and where he may reside or be based. The goal of the BIA in this process is to allow the SIO and the enquiry team to focus on areas of investigation most likely to identify the offender. It is perhaps only in the form of investigative suggestions and interview strategy development where BIAs may on occasion consider more trait-based interpretations and their associated behavioral implications at the investigative level (see below).

Nominal Generation

An extension to predictive profiling, suspect generation may be undertaken by the investigation under the guidance of a BIA. By taking the predictions made in relation to an offender's likely background characteristics, it is possible to utilize local crime and intelligence databases as well as the Police National Computer (PNC) in order to generate pools of potential suspects. Outside of the criminal arena, in some cases, it may also be possible to highlight additional potential suspect pools on the basis of advice offered by a BIA, such as those from housing lists, voters register, employment records, and so on. This will always be evaluated by the BIA on a case-by-case basis.

Prioritization Matrices

Again, an extension of the predictive profiling process, a prioritization matrix takes the individual predictions made in relation to the proposed background of the unknown offender and integrates them in the form of a matrix. Each facet of a potential suspect will be given a numerical value such that nominals within an enquiry can be objectively scored and ranked in terms of how well their background characteristics fit with those proposed for the unknown offender. This process is of particular utility if an investigation wishes to undertake an intelligence-led DNA screen or is seeking to prioritize many hundreds of potential suspects from a cold case enquiry or mass media appeal. Where possible, a suspect prioritization matrix will be developed so as to integrate the behavioral predictions (in relation to an offender's background) with a geographic profile in relation to their most likely area of residence.

Investigative Suggestions

In line with the BIAs intention to make their report as investigatively focused as possible, it is normal practice now for BIAs to offer direct investigative suggestions on the basis of the information supplied to them. It should be recognized that while the BIA role is very much an "advisory" one, they do typically possess significant experience of major criminal investigations (the national cadre of BIAs have an accumulated experience of over 1,000 cases). This experience combined with their ability to draw logical inferences on the basis of an offender's behavior means that they will offer investigative suggestions to the SIO as a routine part of their report. Suggestions are made strictly on a case-by-case basis and should always be accompanied by a clear supporting rationale.

Interview Advice

Contributions from a behavioral perspective can provide a significant enhancement to the development of interview strategies. Such contributions can be classified as either interviewee-specific advice or crime-scene-specific advice. With respect to the former, more traditionally recognized interviewee-specific advice, an identification of salient behavioral characteristics of the individual to be interviewed can inform strategies to maximize interaction and the quantity and quality of information disclosed and minimize confabulation and fabrication. The more contemporary contribution from a crime-scene-specific perspective provides additional guidance to interviewing officers in understanding the offense; will identify any gaps, inconsistencies, and ambiguities in the information; and provide a template against which investigative hypotheses can be systematically tested during interview. It is however essential that interview advice gained from a BIA is complemented with advice from an appropriate in-force or ACPO (Association of Chief Police Officers) Approved Interview Adviser to ensure compliance with relevant legislation and support of the overall investigative aims, objectives, and strategies.

Media Advice

In certain circumstances, it may be advisable to seek opinions from a BIA in relation to utilizing the media in major investigations. This advice is intended to maximize the use of the media when, for example, making appeals to the public or releasing information about an offense, but also to enable SIOs to better understand the potential effects of the media on the behavior of an offender.

Familial DNA Prioritization

The recently introduced technique of Familial DNA (fDNA) searching allows investigations to search for relatives of an unknown offender in cases where a full DNA profile is available at the crime scene that does not match anyone on the National DNA Database (NDNAD). It works on the general principle that people who are related are likely to have more DNA in common than those who are not, and thereby seeks to identify individuals on the NDNAD who have a greater genetic similarity to the unknown offender and hence a greater potential to be related. The BIAs can utilize a sophisticated process that allows the resulting lists from forensic science providers to be re-prioritized with respect to age and geographic association. By adjusting the genetic prioritization to take into account an individual's age and geographic association, those individuals who are more likely to be relatives of the offender should become more readily identifiable from the more general backdrop of the lists, while still preserving the appropriate weight assigned to them through their genetic similarity.

Search Advice

BIAs may also be able to contribute to any search activity within an investigation. An enhanced understanding of the offense and/or likely offender can inform search parameters for forensic evidence, witnesses, CCTV, etc., and in combination with relevant research findings assist in prioritizing potential body deposition sites. Additionally, an enhanced knowledge of specific psychological or criminal dispositions may assist in broader search considerations, such as briefing search officers as to the potential significance (from an intelligence perspective) of items which may be observed within the course of more evidential searches.

Decision Support

As highlighted above, every contribution to the investigation of serious crime from a behavioral science perspective has the single underlying goal of supporting investigative decision-making. While the above provides a more prescriptive summary of the individual products and services that offer explicit tactical and strategic solutions in their own right, the more implicit role of the BIA on the decision-making process is perhaps a less obvious, but nevertheless critical contribution.

The field of decision-making is well established within the psychological literature and offers a multitude of well-established influences on decision-making. The focus of BIA contribution is directed toward those factors which are commonly encountered within major crime investigations, namely, heuristics and biases. It is cognizance of these factors within an investigative environment that underpins an implicit but often overlooked contribution from BIAs and further highlights the distinction between "traditional" paradigms of profilers assisting investigators with "offender profiles" and the contemporary UK approach of discrete, multifaceted, professional psychological expertise.

The ability for investigators to make rational decisions in serious crime enquiries is influenced by the same heuristics and biases that affect all of our day-to-day judgments. Base rate fallacies, representativeness heuristics, illusory correlations, clustering illusions, availability heuristics, anchoring and adjustment, belief persistence, confirmation bias, and selective information search, all have the potential to undermine even the most astute of investigators' decision-making capabilities. While such influences are by no means any more (or indeed less) prevalent among Senior Investigating Officers than among the general population, the consequences of error within a murder enquiry are of arguably greater consequence than the majority of more mundane daily decisions.

It is a critical role of the contemporary BIA to mitigate against such errors of decision-making through the implicit integration of such considerations within all advice and support offered. While such cognizance of these bias and heuristics does not form a "product" or "service" in its own right, it may be argued to represent one of the biggest contributions from BIAs, providing the foundations on which the goal of supporting investigative decision-making can be soundly built.

This requires a sound theoretical understanding of common biases and heuristics, an ability to recognize their presence and potential negative impact within the specific environment of a major incident, the capability to diplomatically address such errors, and an awareness of and access to relevant data and information to optimize subsequent investigative decision-making.

Working Practice and Process

All Behavioral Investigative Advice within the UK is provided upon request, and although advocated on Senior Detective training programs and within national guidance manuals, does not represent a procedural requirement, but rather a carefully considered investigative decision.

Once a BIA has been engaged by the Investigating Officer, a meeting takes place where an exchange of information and views leads to the agreement of explicit terms of reference. These terms are established in writing and clearly articulate what is expected of both parties. This is particularly important with regard to ownership of any material and confidentiality, which is expected in instances where a BIA might have privileged access to sensitive information about crime scenes and/or victims. This should, for example, inhibit disclosure of certain information to the media without the SIO's permission. Terms of reference should also provide similar assurances to the BIA that all relevant case materials will be made available, and any developments which may support, refute, or refine the advice proffered be communicated as soon as reasonably possible.

It is at this stage that the precise nature of any behavioral science support will be discussed and agreed. As has been made explicit above, the potential products and services from contemporary BIAs are far broader and more diverse than either the media or naive academics would proffer, and should not be regarded as the exclusive generation of inferred offender characteristics, either as a product in its own right or as the necessary foundation to other contributions.

In order for the BIA to undertake their behavioral analysis and provide timely advice, a variety of case materials will be required from the investigation. The provision of these assist the BIA to commence their analysis and return a detailed report to the investigation within a timeframe that maximizes the utility of the advice (i.e., as soon as practically possible, but with cognizance to the investigation's overall strategy and resourcing and competing BIA demands from other cases). Although by no means exhaustive, and with cognizance that different enquiries will generate different information and different products and services require different source materials, the following represents a summative overview of the material required and utilized.

- Full verbal case briefing and access to the SIO/ investigation team
- All relevant statements
- Crime report
- Any officers' reports/status reports
- Pathology and forensic reports/findings
- Full set of crime scene and postmortem photographs (where applicable)
- Available analysis (e.g., telephony, palynology, entomology, etc.)
- Relevant maps
- Visit to all relevant scenes

It is worth highlighting that the scene visit represents a critical component of the process as it allows the BIA to gain a fuller understanding of the decision-making process of the offender. Such information is not routinely available from crime reports, statements, or photographs, where often the evidential focus is too restrictive to provide the necessary "behavioral" perspective.

In addition, the scene visit is typically complemented by a visit to the incident room/ enquiry team, allowing the BIA to ask questions concerning the demographics and crime profile (i.e., the type, frequency, patterns, and interpretation of previous criminal activity) of the area, as well as to get a full briefing from the SIO or nominated member of the investigation team. It is therefore of great importance that in addition to making the necessary case documentation available to the BIA, officers who meet with the BIA and who accompany them to the scene(s) should have a good knowledge of the relevant areas and of the offense(s) for which the support has been requested.

Once this information has been collated, the BIA (except in rare circumstances involving critical incidences or "crimes in action") departs the incident room and returns to their place of work to critically review the information and begin report preparation. Establishing direct lines of communication with relevant officers within the enquiry removes any necessity to remain physically located within the incident room, which may be viewed as counterproductive when one considers the resources required by the BIA, the priorities of the investigative team, and the advantages of distanced objectivity. The BIA is reliant upon both the statistical information contained within relevant datasets (which for security reasons are not accessible outside of secure national policing premises) and the extensive academic research libraries they have collated to inform and support inference generation and investigative strategy development. The capacity to reflect and synthesize these resources with the known case material is greatly enhanced by such remote working, eliminating both the distractions and potential biases inherent in major incident rooms.

Once the analysis is complete, a report will be forwarded to the enquiry with the explicit recommendation that the document should not be viewed as the completion of the BIA's input, but rather provide the starting point for a dialogue between the investigator and the BI-A. This is critical to ensure the SIO's understanding of the inferences and recommendations made and to promote understanding that such conclusions should be continually evaluated against additional forthcoming information.

The importance of the BIA becoming part of an advisory team is emphasized by the BIA. This ensures all experts are aware of the findings, opinions, and advice of other members of the team, allowing for hypotheses to be supported, rejected, or refined, and to prompt searches for information in other directions. Discussions with specialists from other disciplines can also act as a useful safeguard against too narrow a focus in an investigation and can encourage officers to continue to look for alternative explanations for events.

Having decided to seek assistance, the SIO should remain critical and reflective of any information or opinion put forward by BIAs. He or she should be ready to explore and challenge assumptions made as well as pointing out any inferences that are inconsistent, contradictory, or logically incoherent. Only by thoroughly discussing the opinion in this way will the SIO be able to gage the validity of the BIA contribution to the investigation. The SIO should ensure that the BIA's opinion is fully explained in order to know what weight to assign to the information provided. Asking the BIA to justify the way in which they have developed their methodology and derived any inferences should ensure their involvement in any investigation is transparent and explicit. SIOs are urged to look beyond the psychological interest factor toward the investigative utility of the advice and are regularly reminded by BIAs to ensure any advice proffered achieves the highest possible standards of quality and effectiveness.

It is of utmost importance that any behavioral advice provided to an investigation is utilized in the manner in which it was intended. In particular, it should be remembered that advice is typically offered on the basis of what is most likely or what should be prioritized. Behavioral Investigative Advice does not deal in absolutes and as such advice from a BIA should always be evaluated carefully by the SIO to ensure that its impact on the investigation is proportionate.

Expert Evidence

Closely aligned to the above discussion, it is made explicit within BIA reports that they do not constitute expert evidence. BIA reports deal in probabilities, not certainties and provide the most likely "type" of individual in order to systematically prioritize lines of enquiry. Due to the probabilistic nature of the findings, while the majority of recommendations will prove effective for the majority of cases, it is to be expected that in a minority of cases, the individual responsible will demonstrate significant variance with the reported probabilities. Even if a significant overlap exists between the "profile" and defendant, the best that could be inferred is that the defendant has the characteristics which have been suggested as most likely from a behavioral analysis of crime scene and related information. This should be deemed insufficiently relevant and reliable and too prejudicial to be received by the court as evidence of the defendant's guilt. It is for this reason that it is made explicit that a BIA's report is not fit for purpose in attempting to provide evidence of an individual's involvement or guilt, or conversely as evidence that a specific individual cannot be responsible for an offense. This is not a failing of BIA advice, rather recognition of its investigative utility as a means of better understanding an event and informing and prioritizing investigative decision-making and actions. As such a "profile" should never be attempted to be used as evidence with regard to identity. However, the wider discipline of Behavioral Investigative Advice may, under specific circumstances, be able to contribute to the court process, although the exact parameters and process for such a contribution are at present unclear and may be found to fall foul of the same obstacles as "profiles" if tested (for a review of the evidential obstacles facing "offender profiling," see Omerod and Sturman 2005).

Conclusions and Future Directions

Behavioral Investigative Advice has developed in the UK over the past decade as a distinct profession which has significantly widened the scope from the traditional, specific view of "offender profiling" to a discipline which now encompasses a broad range of scientifically based yet pragmatic activities related to supporting police investigations, based on replicable, transparent, and valid knowledge and research.

It is acknowledged that while perceptions of such activity may be misguided and outdated, both predecessors and many contemporary practitioners are perhaps still guilty of enthusiasm over professionalism. However, such experiences are perhaps inevitable in any emerging discipline and should not continue to represent the stick that beats contemporary efforts into submission.

Such contemporary optimism is bolstered by many years of frontline experience, hundreds of case consultations, and, significantly, by the policing community. Over the past decade, Senior Investigating Officers have gone from blind acceptance of an unproven innovative technique, through suspicion and virtual dismissal, to the more balanced mid-ground of critically and judiciously evaluating the potential such contributions BIA's make within the overall investigative process.

This contemporary policing environment not only safeguards against the excesses and failures of the past, but demands the very evidence-based practice that contemporary BIAs have recognized as critical to the continuing development and professionalization of the discipline. This cannot be achieved by the practitioners alone. The research community has a key role to play in such endeavors, but activity must be focused on solution-oriented research of direct relevance to the practitioners. Historically, BIAs have had to rely on what has been published within disparate research domains (e.g., psychiatry, criminology, environmental psychology, forensic science, medical science, etc.) and tease out the small morsels of relevance to investigative application. What is required is a much more integrated approach where practitioners and academics better understand one another's aims and objectives and produce findings of direct relevance to real world application. It is acknowledged that many academics will view such bold proclamations as somewhat naive and fanciful, protesting that they have been advocating precisely the same intentions but that such rhetoric is never backed up with access to the masses of relevant data held by the police. This too is changing. The emerging investigative philosophy of a more scientific approach has led to not only the development of larger, centralized, and more sophisticated data collection processes, but also a greater willingness to share such data with the academic community.

In addition to broader knowledge and research strategies aimed at improving the way in which the police service create, assure, share, and use knowledge in support of practice and decision-making, greater access to data held within the various national databases together with clear articulation of specific research questions has began to develop such collaborative activity. It is hoped that such initiatives, viewed in tandem with an enhanced understanding of the contemporary philosophy of Behavioral Investigative Advice, will encourage further solution-oriented research and continue to inform future practice.

Related Entries

▶ Behavioral Science Evidence in Criminal Trials
▶ British Police
▶ Cognitive Forensics: Human Cognition, Contextual Information, and Bias
▶ Criminal Investigative Analysis
▶ Evidence-Based Policing

- ▶ Expert Witnesses: Role, Ethics, and Accountability
- ▶ Investigative Psychology
- ▶ Linkage Analysis for Crime
- ▶ Profiling Arson
- ▶ Psychological Autopsy of Equivocal Deaths

Recommended Reading and References

ACPO (2005) Practice advice on core investigative doctrine. NCPE, Wyboston

Alison L (ed) (2005) The forensic psychologist's casebook: psychological profiling and criminal investigation. Willan, Cullompton

Alison L, Rainbow L (eds) (2011) Professionalizing offender profiling: forensic and investigative psychology in practice. Routledge, Abingdon

Omerod D, Sturman J (2005) Working with the courts: advice for expert witnesses. In: Alison L (ed) The forensic psychologist's casebook: psychological profiling and criminal investigation. Willan, Cullompton

Behavioral Learning

- ▶ Social Learning Theory

Behavioral Linkage Analysis

- ▶ Linkage Analysis for Crime

Behavioral Management in Probation

Faye S. Taxman
Criminology, Law and Society, George Mason University, Fairfax, VA, USA

Synonyms

Case management; Community supervision; Contingency management; Evidence-based supervision; Probation; RNR Supervision

Overview

Behavioral management is the newest framework for community supervision; it is a hybrid combining the traditional theoretical approaches of law enforcement (monitoring and compliance) and social work (resource broker and/or counseling). The social work perspective was more commonplace in the USA until the 1970s and is currently slowly eroding in other countries over the last several decades. The erosion is part of the diminishing support for rehabilitation-type programming for offenders as well as the increased concerns about offender accountability and public safety. Enforcement and compliance management are now the more predominant framework in the USA with other countries emulating this style of supervision given the emphasis on offender accountability. As a hybrid approach, behavioral management arms the probation officer with a new toolkit that recognizes the balance between offender change and accountability, and it offers new strategies for managing the offender in the community that incorporates evidence-based practices and treatments ("what works"). The behavioral management supervision model is based on a theoretical framework that offender change is due to a trusting, caring relationship between the offender and the officer where the emphasis is on structure and accountability. The officer facilitates offender change through therapeutic techniques of individualized case plans, cognitive restructuring, and graduated responses to positive and negative behaviors.

This chapter reviews the literature on the effectiveness of community supervision and then explores the concepts of behavioral management with examples from existing efforts across the globe. Next this chapter discusses the policies, programming, and developmental issues that affect the likelihood of behavioral management becoming predominant in supervision settings. This is critically important since a behavioral management framework requires organizational strategies to emphasize an altered mission and goal for the supervision organizations to encompass structure along with programs and services to alter offender outcomes.

Evolving Nature of Community Supervision Models

Community supervision is the broad category that refers to any correctional control in a community setting. This generally covers pretrial (preadjudication), probation (sentence that involves community control), and parole (post-incarceration) as well as case management of offenders in the community. Probation had its early roots as a humane method of dealing with alcoholics. John Augustus, the grandfather of contemporary probation in the USA, realized that alcohol abuse contributed to workers being unreliable. Probation was conceived as a court service designed to reduce the use of jail sentences for first-time offenders. The early era of probation was focused on alternatives to incarceration and assisting those with needs (mostly alcohol abuse). In the 1970s, more emphasis was placed on oversight (law enforcement) and enforcement of conditions in the USA as support for rehabilitation dwindled (Garland 2001). The current emphasis on enforcement of the conditions of release (i.e., generally reporting, informing the officer of their whereabouts, and other emphasis on controls) dominates probation services.

Probation supervision is built on suppressing criminal behavior through a contact approach. Contacts, or face-to-face meetings between offenders and community supervision officers, are the main tool of overseeing offender behavior. While on probation, the offender has a series of liberty restrictions that define the requirements of supervision. Some of these restrictions are imposed by the court (for probation or pretrial supervision) or the parole board (for parole). Others are part of standard supervision conditions such as limiting activities and curfews, requirements to report to the supervision officer, no possession of guns, and so on. That is, the offenders have a set of rules to guide their behavior. The purpose of supervision monitoring is to improve compliance with these conditions. The notion is that offenders will be deterred from further criminal behavior with the threat of incarceration (due to failure to comply with the conditions of release). The core component of supervision – contacts – has been empirically tested in a limited fashion, primarily with studies of intensive supervision (more contacts), where frequency of supervision has not resulted in improved results (MacKenzie 2006; Petersilia and Turner 1993; Taxman 2002). A recent study of low-risk offenders (i.e., offenders that have a low probability of further criminal justice involvement) found that increasing the number of contacts with supervision officers did not result in reductions in recidivism (Barnes et al. 2012). There have been no experiments or studies on whether being on probation (i.e., having required contacts between the probation officer and offender) or having no oversight has any impact on future offending behavior. This limits the ability to understand whether supervision suppresses offending behavior or how supervision leads to more desirable impacts on offender behavior. The question of what type of supervision is most effective has largely been unanswered except for a recent systematic review that favored the risk-need-responsivity (RNR) (behavioral management) model of supervision (Drake 2011).

A small body of literature has examined the impact of probation on per-person offending rates. MacKenzie and Li (2002) examined the rate of offending of probationers and found that during the period of supervision, their self-reported offending rates decreased slightly. They found that having more conditions of supervision or requirements did not impact offending rates or offending behavior. But there is caution that needs to be acknowledged given that this was a quasi-experimental design with no control group (no one was assigned to not being supervised by an authority of the state) or comparison group; instead the study examined pre-post self-reported behaviors of offending. MacKenzie and Li (2002) also raise questions about whether added supervision conditions are useful given that the additional conditions do not appear to have a deterrent effect.

Supervision studies have primarily focused on contact levels and caseload size. The premise is that the caseload size is used as a proxy for the amount of time that the officer can devote to each offender. Two outstanding questions are: (1) what is the optimal size of the caseload a probation officer can manage to increase the success on probation? and (2) what impact does increasing the contact between the probation officer and the probationer have on contacts? The question about caseload size was addressed in several randomized trials to assess whether a caseload of 25–40 (depending on the study) had any impact on offending behavior. The theory was that having too many probationers to supervise would reduce the effectiveness of supervision by reducing the time that the offender can spend with the officer (Taxman 2002; MacKenzie 2006). Of course, these studies did not consider the difference when caseloads were 100 or more as compared to prior studies when caseloads were under 40 offenders per officer. Most of the caseload size findings from these studies (examining smaller caseloads under 40 offenders) were null, meaning that the caseload size did not affect offender outcomes. Since the concept of size of caseload was not associated with a particular theory of supervision, the studies did not contribute to a better understanding of what the probation officer should do as part of supervision.

Dosage studies (i.e., "frequency of contact") were the next series of studies conducted to detect how much contact is needed to improve offender outcomes. This set of studies was built on similar theoretical deficits in that the studies did not examine different styles of supervision but instead examined how more contact between the officer and offender/probationer affected outcomes. Petersilia and Turner (1993), in the largest multisite randomized trial (13 sites for probation supervision and 2 for parole) in probation supervision, found that intensive supervision services (the amount of required contacts varied by site) served to increase technical violations (i.e., allowed officers to be aware of compliance issues) but had no impact on rearrests. Intensive supervision heightened the number of contacts between the officer and probationer, increased the use of drug testing to detect drug use, and increased conditions of release. Yet, it had no notable impact on offending behaviors of offenders. The increased contact model has been found to be ineffective in reducing criminal behavior, although it is acknowledged that the law enforcement component has the potential to actually increase technical violations (Aos et al. 2006; MacKenzie 2006; Taxman 2002).

The caseload size and dosage studies on supervision have not found any impact on these factors on improvements in supervision outcomes. That is, the law enforcement and monitoring approach does not appear to improve outcomes, but this can only be stated with qualifications given that insufficient studies have examined or compared different types of supervision strategies. A recent systematic review by Drake (2011) found that the RNR model of supervision has the greatest impact on offending, followed by supervision with treatment conditions. Other forms of supervision such as contact only or contacts and drug testing (controls) had no impact on outcomes. The RNR model of supervision allows supervision to be more "client centered," where the goal is to tailor probation conditions and style of supervision to the individual. This draws from the "what works" (evidence-based practices) literature, focusing on reducing the risk of the individual offending by attending to dynamic criminogenic needs through specific case plans and responsive treatment programs. A major component of the model – besides the hard technology of using standardized tools for screening, assessment, and placement in appropriate programming – is the emphasis on the "soft" technology of deportment or the relationship between officer and offender (Taxman 2002, 2008a; Taxman et al. 2004). The RNR model of supervision addresses the tension between enforcement and social work models of supervision by realigning the goals of supervision officers to be behavioral managers. The theoretical framework is that offender change is a function of attention to target, proximal behaviors that are crime reducing, and a trusting relationship between the offender and officer.

The RNR model draws upon the importance of probationers and officers having a working relationship to address the general factors that affect involvement in criminal conduct. Recent studies illustrate how improved working relationships (e.g., empathy, trust, care, voice) have a positive relationship with offender outcomes (Taxman and Ainsworth 2009). Probationers (with mental health conditions) who perceived their officer to be tough had more failures and higher numbers of violations than those that did not (Skeem et al. 2007). Thanner and Taxman (2003, 2004) found that when offenders reported that they had a "voice" (i.e., the probation officer allowed the offender to participate in deciding the type of sanctions for failure to comply with requirements), reductions occurred in arrests and positive drug tests. While recognizing the importance of the person-centered model in supervision, James Bonta and colleagues in Canada studied whether probation officers have the appropriate interpersonal skills, role modeling, and communication skills to work effectively with offenders in an evidence-based practices model of risk, needs, and responsivity. The general findings from their studies are that probation officers do not have these skills and, even if the officers have the skills, they do not use them in supervision of offenders (Bonta et al. 2011).

Within the field, the following principles are considered important to an evidence-based supervision system (Taxman 2008b; Taxman et al. 2004; NIC 2004):

1. *Manage risk levels*. Providing higher-risk offenders with more structured control and treatment services, it is possible to reduce recidivism. Using risk level to articulate a supervision policy based on providing more intensive services by risk level serves to use probation resources wisely and use treatment services in a fashion that generates the best results. A related finding is that low-risk offenders could be supervised with fewer resources, and by providing low-risk offenders with minimal services, one can obtain relatively positive findings. In a recent experiment with low-risk offenders, Barnes and colleagues (2010) found that minimal supervision (average of 2.5 contacts per year) had similar recidivism results as offenders supervised on a more traditional schedule (average 4.5 contacts per year).
2. *Treat dynamic criminogenic needs* (i.e., substance abuse, impulsive behavior, criminal peers, and criminal thinking). Offenders vary considerably in the factors that affect their criminal behavior. While risk level indicates the likelihood of further involvement with the justice system, criminogenic needs identify the dynamic factors that affect involvement with criminal behavior. These dynamic factors can be altered with the use of proper treatment programming and correctional controls. The criminogenic needs should be targeted to improve outcomes.
3. *Manage compliance*. The most difficult part of any behavior change is sustaining the effect. Probation agencies can assist with compliance with judicial or parole board mandates. Compliance is similar to monitoring offenders, but the key is to use short-term goals to allow offenders to achieve positive outcomes; short-term goals are stacked towards longer-term outcomes of reduced drug use and alcohol abuse and continued involvement with the justice system. Retention strategies are important to achieve longer-term impacts. The use of administrative sanctions (matrix of responses to different types of behavior such as drug use, failure to pay fees, etc.) is recommended to provide graduated, structured responses.
4. *Create an environment for offender change*. An important part of the EBP model is the creation of an environment to support offender change. A focus only on revocations is counter to the emphasis on behavioral change, which is long and difficult. More attention is needed to creating an organizational culture that supports trusting relationship on supervision. The working relationship between the offender and the probation officer is an important component of the change process in that it builds support for change (Taxman and Ainsworth 2009).

This constitutes the behavioral management model of supervision where therapeutic strategies are used to facilitate offender change. The toolkit for supervision officers includes motivational enhancements, cognitive restructuring for meeting criminogenic needs and compliance management, and targeted responses to behaviors; all of this must occur in an environment where behavioral change is expected to be slow and steady with some relapses. In fact, relapses are expected to be commonplace and provide the tools to build resiliency to further change.

Maryland Proactive Community Supervision Study

A recent systematic review found that the evidence-based practices model that emphasizes behavioral management reduces recidivism by 16 %, in contrast to enforcement supervision (no impact) and supervision with treatment (10 % effect) (Drake 2011). The question is whether an evidence-based supervision model can be implemented and can deliver the same expected results of reduced recidivism as the "what works." The Maryland Proactive Community Supervision (PCS) experiment was a test of this model (see Taxman 2008a). Taxman and colleagues (2004) translated the risk-need-responsivity (RNR) model into a model for probation supervision given that the original model of "what works" principles was primarily geared towards treatment-related programming. The core components of the probation supervision model were based on the following theoretical approaches that blend accountability with therapeutic interventions, as shown in Table 1. One key component is the role of the supervision officer in helping the offender facilitate behavioral change by using cognitive restructuring strategies into routine practices of risk and need assessment, case planning, and compliance monitoring. That is, officers were trained in using cognitive restructuring and motivational interviewing as part of how they address traditional offender processing procedures (e.g., assessment, case planning, treatment and control monitoring, and determination of progress). The PCS model was designed to address the tension between the law enforcement and social work frameworks, with the emphasis on using quality contacts as a conduit for offender change. The behavioral management components were designed to integrate cognitive restructuring and behavioral modification frameworks within supervision settings by having the officer use these techniques with offenders.

Behavioral Management in Probation, Table 1 Components of the PCS model

Components of the PCS model	Theoretical framework
Risk and need assessment;	Risk assessment
Prioritize moderate- and high-risk offenders in treatment programs	Treatment matching (responsivity)
Case planning and typologies for major criminogenic needs (specifying needs)	Treatment matching (responsivity)
Offender input into case plans	Motivational enhancement; Cognitive restructuring
Deportment (working alliance between officer and offender)	Motivational enhancement
Compliance management including sanctions and rewards	Enforcement; Boundary setting; Behavioral modification
Cognitive restructuring for offenders	Cognitive behavioral interventions
Officers as role models	Motivational enhancement

The PCS offender supervision process model consists of five major components (see Taxman et al. 2004 for more details). A five-item risk screener was used to determine initial assignment. Offenders scoring moderate to high risk are then involved in a lengthier assessment process involving the Level of Service Inventory–Revised, a third-generation risk assessment tool (Andrews et al. 2006). Officers also conduct a thorough review of the person's place of residence and social networks. As part of the process, the offender is asked about their desired goals during supervision. All of this information is used to assess the factors that affect criminal behavior, with attention to the specific typology of criminogenic behavior. The six typologies (and related goals) are disassociated (develop prosocial network), drug addiction (achieve abstinence and recovery), drug-involved entrepreneur (obtain prosocial employment), violent offender (address power and control issues), domestic violence (address relationship and control issues), mental health (continue treatment and address support network), and sexual deviance (achieve control over urges and manage behavior). The typology helps define the components of the case plan, including the appropriate type of treatment programming and controls applied, such as drug testing, curfews, house arrest, or other types of mechanisms to limit liberties. On a monthly basis, the offender and the supervision officer review progress on the case plan (see Taxman 2008a). The case plan is revised monthly based on the offender's progress in meeting goals and the offender's attitude. Offenders are given no more than three goals a month, with participation in treatment services prioritized as one of the key goals.

The process hinges on the offender's engagement to achieve ownership of the case plan through a positive working relationship with the supervising officer. The agency achieves the emphasis on engagement through a series of skill development sessions that emphasize (1) learning to use motivational interviewing strategies, (2) using cognitive restructuring strategies to assist offenders in problem solving to address their criminogenic needs, and (3) using performance measurements (e.g., drug tests, attendance) to assess offenders' progress. The skill development efforts provide officers with the ability to adapt and translate motivational interviewing, behavioral modification, and cognitive restructuring in a community supervision environment (see Taxman et al. 2004) into traditional functions of assessment, case planning, monitoring, and compliance assessments. These skill-building components provide the officers with new behavioral strategies to work with offenders.

The PCS model of skill development and refocus of supervision could not be accomplished without changes in the culture of the probation officer. The agency created a social learning environment (for staff and probationers) with an emphasis on building a probation culture that supported officers' use of the behavioral management strategies instead of traditional law enforcement techniques. The agency adapted core tenets of a social learning environment: greet offenders with proper salutations, shake hands with the offender upon entrance and exit, and maintain eye contact with the offender during all personal conversations. Additionally, the face-to-face contact of supervision is realigned. The traditional face-to-face contacts are altered to bilateral information exchange sessions where the goals of supervision could be assessed, refined, and restated. Assessment and other data collection exercises are shared with offenders to allow them to learn about their own behavior. Finally, performance measures for offenders, supervision staff, and PCS offices are used to reinforce the emphasis on behavioral management strategies such as number of offenders employed, number of goals achieved, and then attention to the accomplishments per typology. Officers and probationers reviewed the target goals in each section as part of the behavioral contract and these accomplishments are used to build future target goals. The emphasis on target goals is to gradually improve the probationers' success on probation. This is the process that defines the behavioral management framework used in a responsive face-to-face contact system.

In a place-based experimental design where four probation/parole offices in Maryland implemented the behavioral management supervision model (PCS) and four similar offices used the traditional law enforcement supervision style, logistic regression results found that, controlling for length of time on supervision and prior criminal history, offenders that were supervised in the behavioral management style were less likely to be rearrested (30 % for the PCS and 42 % of the non-PCS sample; $p < 0.01$) and less likely to have a warrant issued for technical violations (34.7 % of the PCS group and 40 % for the non-PCS group; $p < 0.10$). The study also included a process evaluation to assess the degree to which the key components of the model were implemented. The assessment inventory (LSI-R) was implemented in 70 % of the cases and the typologies were used in 56 % of the cases. The typologies were used as a vehicle to identify short-term goals to improve outcomes, and for those that were assigned typologies, a greater percentage were placed into appropriate programming and controls than in the control group. Some officers did not implement the model as planned which resulted in some cases with no typology assignment. In these cases, fewer services were obtained for the probationer. Several replications of the application of the risk-need-responsivity model of supervision have confirmed reductions in recidivism (Drake 2011).

Further Work on Behavioral Management (RNR) Supervision Models

Trying to disentangle the effective mechanisms of supervision is the subject of a growing body of research. The focus of attention has been on (1) specialized caseloads, (2) different collaboration between probation agencies and treatment providers, (3) the use of graduated sanctions to manage offender behavior, (4) "light" supervision for low-risk cases, and (5) improving the problem-solving skills of officers.

Specialized Caseloads. Very little research has been conducted on the impact of specialized caseloads, or the concentration of offenders with like criminogenic needs with one officer. The available research has been on offenders with mental health disorders or domestic violence. In a national survey of specialized caseloads, special mental health units were distinguished by (1) caseloads consisting of clients with mental disorders, (2) reduced caseload sizes, (3) trained officers in mental health issues and behavioral management problems, (4) integration of probation and community mental health resources, and (5) use of problem-solving strategies to address treatment noncompliance. Three separate evaluations of specialized caseloads for probationers with mental illness have found a tendency towards positive outcomes of reduced technical violations or rearrests, although these studies are not conclusive given their weak research designs and measures (Epperson 2009). Specialized units for domestic violence caseload have shown similar results to intensive supervision programs with increased contacts that have lead to increased violations. The emphasis of most domestic violence research focused on enforcement and control of offenders, which has the same outcomes of other intensive supervision studies – that is, the specialized caseloads led to greater probation violations.

Collaboration Models Between Probation and Treatment Providers. Most community correctional agencies in the USA do not provide treatment services directly by the probation agency. Instead, the offender is referred to treatment programs provided by publically funded treatment programs or the probation agency contracts for services in the community. Increasing the access to treatment services has been a long-standing challenge for probation and community corrections agencies in that the services are seldom designed specifically for an offender population. The specialized case management services, commonly referred to as Treatment Accountability for Safer Communities (TASC), are designed to provide a bridge between justice and treatment agencies. The five-site evaluation of the TASC case management model found that case management did not result in reduced recidivism or technical violations and did not improve access to treatment services. In the sites where positive findings occurred, the service agency also provided treatment services and therefore directly improved access to treatment (Anglin et al. 1999). This is a common theme regarding case management; agencies that provide both case management and treatment are more likely to improve outcomes.

In contrast to a brokerage model of service linkage, Taxman (1998) developed the concept of a policy-driven seamless system of care. The concept was built upon information sharing across the spectrum of service delivery: assessment, drug testing, treatment access and participation, supervision requirements, and outcomes. The goal was to have probation officers and treatment providers jointly share information that would be useful in their efforts to assist offenders. The study results vary considerably, mainly based on the features of the seamless system that are present. For pretrial supervision focusing on testing, treatment, and sanctions, a randomized trial found reductions in rearrest rates for higher-risk offenders (Harrell et al. 1998). In another study of probation that used a person-centered model of probation supervision, the findings are also mixed. An assessor in the probation office would screen for substance abuse disorders and then target offenders to appropriate treatment. Treatment should be 9 months in duration with two levels of care (generally a step down to outpatient but it could be from in-jail treatment to outpatient care, drug testing, and sanctions). The probation/parole officer's role was as a resource broker, working with the treatment counselor on treatment access, monitoring compliance with drug treatment requirements, and responding to positive drug test results. The study resulted in increased access to treatment services and increased days in treatment, but the seamless intervention was more costly to deliver and did not result in reduced recidivism for some higher-risk offenders (Alemi et al. 2006; Taxman and Thanner 2006). For higher-risk offenders, there was a reduction in arrests and opiate drug use (Thanner and Taxman 2003; Taxman and Thanner 2006). A more recent study used a slightly different seamless framework but added a manualized cognitive behavioral therapy treatment sessions offered in the probation office. Again, there were increases in treatment access, and the study found improvements in outcomes for moderate-risk offenders, but not for high-risk offenders (Taxman and Belenko, 2011). The provision of treatment in the probation office did not universally improve outcomes for offenders, but the study found that certain treatments are more effective for some offenders. Overall, the study found that offenders with drug dependence disorders tended to have less positive outcomes than those with abuse disorders. It appears that the outpatient treatment is not as effective for those with dependence disorders.

Another study focused on probation/parole officer and treatment provider working together in a collaborative model of care. The Step 'n Out study, conducted in six sites, required the officer and treatment provider to work together on a collaborative behavioral management (CBM) process of accessing treatment needs, reviewing treatment progress, and using a structured reward schedule to incentivize offenders for positive behavior (Friedmann et al. 2008). CBM is modeled after contingency management, which focuses on providing rewards for positive behaviors (see Stitzer et al. 2011). The study found a treatment effect for offenders with substance use disorders, particularly those that did not use hard-core drugs (e.g., opiates, cocaine, methamphetamine) (Friedmann et al. 2012). In this model, the probation officer and treatment provider have frequent meetings to review the progress of offenders.

Use of Graduated Responses. Supervision officers have tremendous discretion regarding the handling (and managing) of offender outcomes. As drug treatment courts and the seamless system evolved, the emphasis was on using gradual responses to offender noncompliance, commonly referred to as graduated sanctions. The notion is that sanctions need to be swift, certain, and severe (increasing in intensity based on the number and type of violation behavior). Sanctions can be delivered by probation or parole officers, or they can be delivered by judicial or parole board authorities. The former is referred to as administrative sanctions and the latter as sanctions. Swiftness means that the sanction occurs in a period of time close to the noncompliant event. Certainty means that the sanction is delivered as it was indicated. Severity refers to the type of response and whether it involves liberty restrictions such as incarceration. Paternoster (1987), in a study of perceptional deterrence, found that swiftness and certainty are more likely to influence offending behavior. While there has been general support for probation and parole officers using administrative sanctions (a set regulatory procedure) or a matrix of noncompliant behavior and type of responses, there are relatively few empirical studies of the degree to which administrative sanctions reduce offending behaviors. In a non-experimental study examining the use of graduated sanctions on parolees during the first year after release from prison, researchers found that graduated sanctions could alter the offending behaviors of offenders. The use of graduated sanctions in a manner that features swift, certain, and severe can enhance the suppression effect of parole on criminal behavior. Steiner et al. (2012) did not examine the use of any specific sanction schedule and used the sanction mechanism as a proximity for severity of response (i.e., supervising officer gives a privilege restriction, parole board or a hearing officer, placement in a halfway house, and formal revocation hearing where it was possible the offender may have been returned to prison). They found that sanctions were effective regardless of offender risk level, but younger offenders did not respond as well to sanctions.

Another graduated response is the use of rewards. Following the model of contingency management in the substance abuse treatment literature, rewards are to be delivered swiftly, certainly, and severely (progressively increasing in value). Rewards tested in drug treatment programs demonstrate that offenders respond well, with rewards impacting drug use and treatment retention. Marlowe and colleagues (2009) found that rewards were very successful for young offenders in a drug court setting in terms of reducing offending behavior. Rudes and colleagues (2011) outlined how rewards could be adapted in justice-type settings and used as part of a progressive behavioral management scheme. Rewards create an environment for probation that is consistent with the efforts to facilitate offender change. That is, recognizing that offenders respond to acknowledgment about positive steps during probation facilitates positive outcomes.

Low-Risk Offenders. The risk-need-responsivity model focuses attention on moderate-to high-risk offenders. The triage model indicates that low-risk offenders should not have the same oversight or monitoring as other offenders, presuming that their criminogenic needs are minimal (similar to their low criminal justice risk status). This principle is based on the lower recidivism rates of low-risk offenders. Barnes and colleagues (2012) conducted an experiment where low-risk offenders were supervised on administrative caseloads (caseloads averaging 400 offenders per officer with requirements of two in-person contacts and two phone contacts a year) or traditional supervision (monthly in-person contacts). The study found that offenders assigned to traditional supervision are more likely to abscond from probation (partially a function of the traditional supervision defining absconding as missing for 2 months of contacts, whereas the experimental group had to not have contact in 3 months). There were no statistically significant differences in the rearrest rates or drug test positive rates. In examining the comparative effectiveness of the two models for supervising low-risk offenders, the experiment confirms that there is a slight advantage in reducing contact levels in that there are fewer absconders. Similar to intensive supervision, which finds that increased attention is likely to denote more compliance-related issues, this study found that low-risk offenders can be managed with less supervision. These findings are important for policy considerations in that risk level is an important consideration for managing offenders in the community.

Skill building of Officers. The behavioral management approach to supervision has yielded attention to an important part of the equation to achieve more positive offender outcomes: the skills of the officers supervising offenders. Officer skills are important because the model, as described above, relies upon officers to create an environment where change can occur, to engage offenders in the change processes, to use cognitive restructuring techniques in working with offenders, and to use problem solving and other strategies to address compliance-related issues. The notion that officers need to be trained on the basic premises of the behavioral management principles (including RNR) was evident in the MD PCS study (cited above), which recognized that preservice training and ongoing in-service training did not include the basic principles of managing risks and attending to criminogenic needs. The MD PCS project designed and built three training programs: Officer Engagement skills (based on principles of motivational enhancement and interviewing), “Sizing Up” (based on applying the risk and need assessment tools), and in-office refresher training (see Taxman et al. 2004). More recent attention has been given to enhancing the training of officers through similar curriculums that focus on structuring sessions, relationship-building skills, behavioral techniques, cognitive techniques, and effective correctional skills (Bonta et al. 2011; Robinson et al. 2011). The premise is that in order for officers to use the RNR principles, their work flow needs to be adapted to the principles of their work environment, including attention to intake and assessment, monitoring compliance, monitoring treatment compliance, and reinforcing cognitive restructuring.

Without training in applying the RNR principles, it is unlikely that officers will discuss procriminal attitudes or cognitions with offenders and even more unlikely to use cognitive intervention techniques with their clients. Bourgon et al. (2011) found that the Canadian STICS training, which included review tapes, improved officers' discussion about procriminal attitudes and results in increased utilization of cognitive techniques (namely, helping the offender to examine their own behavior and outline options). In a cox regression model, the authors also found that the use of cognitive techniques reduced offending rates. Bonta and colleagues (2011) had similar findings when most probation officer-offender contacts focused on probation conditions and noncriminogenic factors (e.g., work, sports). The increased training improved the focus on procriminal attitudes among the experimental group with more attention to attitudes and employment and educational issues. Offenders supervised by officers using the skills were also significantly less likely to reoffend.

Conclusion

The behavioral management model of supervision is currently being implemented in the USA as well as other countries. The recent systematic review confirmed that the RNR supervision model had better outcomes than other supervision approaches (Drake 2011). However, probation officers are not aware of how the risk-need-responsivity (RNR) principles apply to supervision, and therefore, the successful implementation of this model will depend on the technology transfer strategies that are used. The "what works" model (RNR) has primarily been targeted to treatment programs, but as learned in the MD PCS demonstration, it is possible for supervision officers to use the RNR principles. The challenge is to adapt RNR to align with supervision. RNR relies upon principles of behavioral management in terms of recognizing that there are boundaries that need to be in place (enforcement) but also that therapeutic techniques of individualized care, cognitive restructuring, and graduated response are useful to achieve improvements in supervision outcomes. The advent of various training protocols (i.e., STICS in Canada, STARRS in federal probation, EPICS in other supervision models) illustrates the complexity of the transfer process and the need to import RNR principles into supervision practices. The proliferation of these curricula illustrates that there is a need to improve the skill sets of officers to use the RNR principles, especially given that few officers are trained in social work or counseling skills. This becomes the challenge to implementation in that supervision staff need to develop new skills to work with offenders in a different manner.

The other challenge is the development of policies to support the use of behavioral management strategies. An important part of the model is to use the risk level to allocate resource levels so that moderate- and high-risk offenders receive ongoing (and more intensive) care than low-risk offenders. While this is an intuitive policy, it is actually more challenging to implement since officers tend to find lower-risk offenders easier to supervise (e.g., they tend to be less resistant). But other policies that are important to put in place have to do with collaborations with treatment providers and using graduated responses. The collaboration with treatment providers is an important policy issue. The working relationship between supervision and treatment agencies should be defined at the office level to ensure that treatment providers use evidence-based treatments and that information is shared between the two disciplines. The creation of policy-level agreements will avoid probation officers having different agreements with treatment providers and having different expectations from the treatment providers. The use of graduated responses – both negative and positive – is another policy issue in that the agency needs to determine what can be used as reinforcers both in terms of liberty restrictions (sanctions) and liberty enhancers (rewards).

This discussion of behavioral management supervision has not included another dimension that is being integrated into the workspace of supervision: temporally and spatially based supervision. The increasing awareness that some offenders are concentrated in certain neighborhoods has spurred an interest in community-based probation models that focus on ameliorating the risks of the individual and the place. These models are based on the premise that concentration of offenders into typically resource-deprived communities only serves to undermine supervision; the demands of the community override the requirements of supervision. Byrne (2009) also points out that targeting supervision at high-risk times – generally referring to the early part of the supervision period or the first 90 days – is a sound strategy for reducing recidivism. During this early period, offenders are more likely to fail (MacKenzie and Li 2002; MacKenzie and Brame 2001), and officers have an opportunity to set expectations about the behavior of offenders under supervision. The early period of supervision is when offenders are going to test the conditions of release. Little research has been conducted on different methods for improving compliance during this early period of supervision, but it is an important factor.

Behavioral management offers a solution to the age-old debate about supervision being enforcement vs. social work. Since the behavioral management (RNR) model incorporates a balance between compliance-focused supervision and working on offender change issues, it offers a model of supervision that fulfills the multiple goals of sentencing. Moreover, behavioral management helps assist officers in learning strategies to facilitate offender change. The test over time will be whether behavioral management supervision can be successfully implemented and what the scaled-up version will have an impact on offenders' outcomes.

Related Entries

- ▶ Community Courts
- ▶ Effective Supervision Principles for Probation and Parole
- ▶ Probation and Parole Practices

Recommended Reading and References

Alemi F, Taxman FS, Baghi H, Vang J, Thanner M, Doyon V (2006) Costs and benefits of combining probation and substance abuse treatment. J Ment Health Policy Econ 9(2):57–70

Andrews DA, Bonta J, Wormith JS (2006) The recent past and near future of risk and/or need assessment. Crime Delinq 52(1):7–27

Anglin MD, Longshore D, Turner S (1999) Treatment alternatives to street crime. Crim Justice Behav 26(2):168–195

Aos S, Miller M, Drake E (2006) Evidence-based public policy options to reduce future prison construction, criminal justice costs, and crime rates(. Washington State Institute for Public Policy, Olympia, p 44, No. 06-10-1201

Barnes GC, Ahlman L, Gill C, Sherman LW, Kurtz E, Malvestuto R (2010) Low-intensity community supervision for low-risk offenders: a randomized, controlled trial. J Exp Criminol 6(2):159–189

Barnes GC, Hyatt JM, Ahlman LC, Kent TL (2012) The effects of low-intensity supervision for lower risk probationers: updated results from a randomized controlled trial. J Crime Justice 35:200–220

Bonta J, Bourgon G, Rugge T, Scott T-L, Yessine AK, Gutierrez L, Li J (2011) An experimental demonstration of training probation officers in evidence-based community supervision. Crim Justice Behav 38(11):1127–1148

Bourgon G, Gutierrez L, Ashton J (2011) The evolution of community supervision practice: the transformation from case manager to change agent. Irish Probat J 8:28–48

Byrne J (2009) Maximum impact: targeting supervision on higher-risk people, places and times. Public Safety Performance Project. The PEW Center on the States

Drake EK (2011) "What works" in community supervision: interim report. Washington State Institute for Public Policy, Olympia, Document No. 11-12-1201

Epperson MW (2009) Criminal justice involvement and sexual HIV risk among drug-involved men and their primary female sexual partners: individual and couple-level effects (Dissertation). Columbia University

Friedmann PD, Katz EC, Rhodes AG, Taxman FS, O'Connell DJ, Frisman LK et al (2008) Collaborative behavioral management for drug-involved parolees: rationale and design of the Step'n Out study. J Offender Rehab 47(3):290–318

Friedmann PD, Green TC, Taxman FS, Harrington M, Rhodes AG, Katz E et al (2012) Collaborative behavioral management among parolees: drug use, crime & re-arrest in the Step'n Out randomized trial. Addiction 107:1099–1108. PMCID: PMC3321077

Garland D (2001) The culture of control: crime and social order in contemporary society. Oxford University Press Ill: University of Chicago Press

Harrell A, Cavanagh S, Roman J (1998) Findings from the evaluation of the DC superior court drug intervention program. The Urban Institute, Washington, DC

MacKenzie DL (2006) What works in corrections: reducing the criminal activities of offenders and delinquents. Cambridge University Press, New York: Oxford University Press

Mackenzie DL, Brame R (2001) Community supervision, prosocial activities, and recidivism. Justice Q 18(2): 429–448

MacKenzie DL, Li SD (2002) The impact of formal and informal social controls on the criminal activities of probationers. J Res Crime Delinq 39(3):243

Marlowe DB, Festinger DS, Arabia PL, Dugosh KL, Benasutti KM, Croft JR (2009) Adaptive interventions may optimize outcomes in drug courts: a pilot study. Curr Psychiatry Rep 11(5):370–376. doi:10.1007/s11920-009-0056-3

Paternoster R (1987) The deterrent effect of the perceived certainty and severity of punishment: a review of the evidence and issues. Justice Q 4(2):173

Petersilia J, Turner S (1993) Intensive probation and parole. Crime Justice 17:281–335

Robinson CR, VanBenschoten S, Alexander M, Lowenkamp CT (2011) Random (almost) study of Staff Training Aimed at Reducing Re-Arrest (STARR): reducing recidivism through intentional design. Fed Probat 75(2):57–63

Rudes, Danielle S, Faye S. Taxman, Shannon Portillo, Amy Murphy, Anne Rhodes, Maxine Stitzer, Peter Luongo, & Peter Friedmann. 2011. "Adding Positive Reinforcements in a Criminal Justice Setting: Acceptability and Feasibility." The Journal of Substance Abuse Treatment 42(3): 260-270.

Skeem JL, Louden JE, Polaschek D, Camp J (2007) Assessing relationship quality in mandated community treatment: blending care with control. Psychol Assess 19(4):397–410. doi:10.1037/1040-3590.19.4.397

Steiner B, Makarios MD, Travis LF, Meade B (2011) Examining the effects of community-based sanctions on offender recidivism. Justice Q Short-term effects of sanctioning reform on parole officers' revocation decisions. Law and Society Review, 45(2), 371- 400

Stitzer ML, Jones HE, Tuten M, Wong C (2011) Community reinforcement approach and contingency management interventions for substance abuse. In: Cox WM, Klinger E (eds) Handbook of motivational counseling: goal-based approaches to assessment and intervention with addiction and other problems. Wiley, Chichester, pp 549–569, Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/9780470979952.ch23/summary

Taxman FS (1998) Reducing recidivism through a seamless system of care: components of effective treatment, supervision, and transition services in the community. Office of National Drug Control Policy, Treatment and Criminal Justice System Conference, Washington, DC, Feb 1998. NCJ 171836

Taxman FS (2002) Supervision–exploring the dimensions of effectiveness. Fed Probat 66(2):14–27

Taxman FS (2008a) No illusion, offender and organizational change in Maryland's proactive community supervision model. Criminol Public Policy 7(2):275–302

Taxman FS (2008b) To be or not to be: community supervision déjà vu. J Offender Rehab 47(3):250–260

Taxman FS, Ainsworth S (2009) Correctional milieu: the key to quality outcomes. Vict Offender, 4(4), 334–340

Taxman FS, Belenko S (2011) Implementing evidence-based practices in community corrections and addiction treatment. Springer, New York: Springer-Verlay

Taxman FS, Shepardson E, Byrne JM (2004) Tools of the trade: a guide to implementing science into practice. National Institute of Corrections, Washington, DC

Taxman FS, Thanner M (2004) Probation from a therapeutic perspective: results from the field. Contemp Issues Law 7(1):39–63

Taxman FS, Thanner M (2006) Risk, need, & responsivity: it all depends. Crime Delinq 52(1):28–51, PMID: 18542715

Thanner M, Taxman FS (2003) Responsivity: the value of providing intensive services to high-risk offenders. J Subst Abus Treat 24:137–147

Behavioral Science Evidence in Criminal Trials

Jenny McEwan
School of Law, Exeter University, Exeter, Devon, UK

Overview

Expert witnesses are by no means always scientists, but they must know something the judge or jury do not. The judge will decide whether or not that is the case. For the behavioral sciences, this approach tends to favor the clinical end of the spectrum; even there, however, lawyers have been known to be skeptical of the value of such evidence in a criminal trial. Such reservations are often founded on ignorance of the culture, purpose, and circumstances that form the background against which experts make behavioral assessments. Courts should be aware of the requirements of diagnoses for clinical purposes. The length of observation period normally necessary for diagnostic purposes is unlikely to be available in the context of preparation for trial. A relatively short period of contact possibly makes it easier for a mental illness to be faked. In any event, a retrospective analysis of how an observed condition would have affected past events inevitably involves much speculation.

Behavioral sciences such as psychology, sociology, or anthropology do not fit comfortably within conventional perceptions of expert scientific evidence. There is, within these disciplines, extensive disagreement about questions of central importance, such as methodology. This makes it more difficult to persuade judges that witnesses specializing in these areas of scholarship have an expert opinion of evidential value. Even more problematically, to some extent behavioral sciences represent a challenge to the power of a tribunal of fact, whether judge, juror, or magistrate, to determine the ultimate issue, the matter to be determined in reaching a verdict. Frequently the central or ultimate issue in a criminal trial requires a determination of how an individual, often the defendant or victim, responded to a particular situation. This decision by the trier of fact tends to be based to some extent on intuitive or heuristic judgments about human behavior. Many lawyers would defend such decision-making as an essential element in the role of the lay jury as the representative of popular sentiment, bringing community moral and practical judgments into the criminal court. This role, and the way in which it has traditionally been performed, insofar as it consists of assessments of human reactions and motivations, occupies the very area of expertise claimed by behavioral science. Unlike "hard science," therefore, where courts defer to specialist knowledge, behavioral science is seen as trespassing on the fundamentals of the jury's responsibility and rationale. Lay finders of fact and behavioral experts are effectively pitted in competition against each other in terms of their ability to understand and evaluate human nature.

Admissibility of Behavioral Science Evidence

Where a potential witness can contribute expertise indubitably outside the experience of the tribunal of fact, such as a psychiatrist or clinical psychologist in relation to the defenses of diminished responsibility or insanity, the expert may, even in jurisdictions which prohibit expert opinion on the ultimate issue, give evidence as to the presence or absence and nature of the condition in question and whether it satisfies legal requirements. In most cases, however, behavioral science evidence occupies a problematic middle ground between acknowledged scientific expertise, such as clinical diagnosis, and the kind of "common sense" or everyday perceptions and generalizations of human behavior that are considered properly to be the province of the jury. Given that, essentially, the common law approach to the admissibility of expert opinion evidence is one of laissez-faire insofar as the question is left to the trial judge in the particular proceedings (Roberts and Zuckerman 2010, pp. 482–493), the admissibility of behavioral science evidence is determined on a case-by-case basis. However, courts are jealously protective of the role of the finder of fact as the arbiter of everyday life. In the Court of Appeal of England and Wales, Lord Justice Lawton famously observed:

> The fact that an expert witness has impressive scientific qualifications does not by that fact alone make his opinion on matters of human nature and behaviour within the limits of normality any more helpful than that of the jurors themselves; but there is a danger that they may think that it does....Jurors do not need psychiatrists to tell them how ordinary folk who are not suffering from any mental illness are likely to react to the stresses and strains of life. (*R* v *Turner* [1975] 1 All ER 70, 74)

A concern to protect the role of the jury in interpreting events and applying their own standards of reasonableness may prompt courts to distinguish "normal" behavior, preempting expert opinion evidence, from "abnormal" mental states. This sharp distinction is a legal invention. In psychiatry, the difference is a matter of degree rather than kind, and as medical knowledge advances, the blurry line between them shifts over time. "Medical expertise lies along a line between the predictive technological know-how of the physical science and the loose generalizations of social science practice" (Smith 1985, p. 69). Nevertheless, legal jurisdictions may develop their own typologies of mental states indicating which will be deemed pathological or normal as far as the law of that jurisdiction is concerned. This will determine whether the jury require expert help to decide what the effect of the disorder might have been. And, while psychological research on memory in "normal" adults would be generally excluded from evidence, child development will usually be regarded as an established scientific discipline upon which a specialist could assist the court. Explaining developmental aspects of children's memories and how a particular child compares with the norm does not usurp the function of the jury, as long as the expert does not say that a particular child is telling the truth; that would be for the court to decide.

As "gate-keepers" determining whether or not particular expert evidence is admissible in a given case, judges arbitrate between areas of psychology in which it is possible to draw on specialist expertise over and above the everyday experiences of lay jurors. Obscure or recently identified conditions are more likely than familiar ones to be regarded with hostility. The judge has to accept both that the claimed expertise exists as a discipline and that the particular witness is sufficiently acquainted with it. A second aspect of the "gate-keeper" role is the possibility that the judge will exercise some kind of reliability standard in deciding whether or not the opinion evidence should be admitted. In England and Wales, however, there has been concern that a "culture of acceptance" of purportedly scientific evidence could be creating miscarriages of justice. A generous approach is likely to be founded on the belief that cross-examination by opponents is an effective tool to unmask any bogus science or scientist. True, the anxiety in England and Wales about "junk science" expressed in recent case law, the academic literature, and a significant official report (Law Commission 2011) has arisen in the context of "hard," nonbehavioral scientific evidence on such matters as the reliability of "earprint" evidence (*R* v *Dallagher* [2002] EWCA Crim 1903), or shaken baby syndrome (*R* v *Harris* [2006] 1 Cr App R 5, [2005] EWCA Crim 1980). However, the Law Commission's recommendations for an explicit reliability standard to regulate the admission of all expert opinion evidence in criminal proceedings would, if implemented, apply to "soft" science, too. This approach has caused problems in the United States, where judicial attempts to evolve a workable reliability test have been accused of favoring dated theories and militating against the admission of reliable and pertinent behavioral science evidence.

Defining Reliability Standards

An influential reliability test, deriving from *Frye* v *US*, 293 F 1013 (DC Cir 1923), is that the discipline commands "general acceptability to the scientific community." A general acceptance criterion could be thought to discriminate against any developing science, and is likely to exclude any discipline where there is substantial controversy among practitioners. An Australian test, inquiring "whether the subject matter of the [expert's] opinion forms part of a body of knowledge or experience which is sufficiently organized or recognized to be accepted as a reliable body of knowledge or experience" (*R* v *Bonython* (1984) 38 SASR 45) is along similar lines. The alternative American test, advocated by *Daubert* v *Merrell Dow Pharmaceuticals*, 113 S Ct 2786 (1993), aims to enable the courts to be more responsive to scientific innovation, while at the same time demanding that sufficient reliability is demonstrated. Trial judges are enjoined by *Daubert* to consider factors such as whether the technique in question has been subjected to peer review, whether it has been generally accepted, and whether the theory or technique can be or has been tested. Another factor that may be considered is the falsifiability of the opinion, which means that its plausibility depends upon whether the theory or concept on which it rests is capable of being shown to be false. Any proposition incapable of being verified in this way should, on the *Daubert* test, be treated with great suspicion by the courts.

The difficulty for the social and behavioral sciences, economics, sociology, and psychology, is that they do not fit easily into this portrait of a body of knowledge consisting of a set of propositions which have been demonstrated to hold in all circumstances. The English Law Commission proposes a new test for reliability which would force judges to decide whether or not the purported science is soundly based. Trial judges should consider a number of factors in making this judgment, the most significant for behavioral science being: whether or not any relevant hypothesis has been sufficiently scrutinized, the nature of the data on which the opinion is based, where the expert's opinion relies on the results of any measurement or survey, and whether the opinion takes proper account of matters such as the degree of precision or margin of uncertainty in data collection (Law Commission 2011).

Reliability Standards and Behavioral Science Evidence

Evidence given by experts in behavioral or social sciences is not easily accommodated within a *Daubert*-style test for admissibility. Falsifiability is not an appropriate technique in the social sciences. Controlled experiments are virtually impossible: "we correctly refuse to abuse a child for the sake of research" (Graham 1998). Aware of such practical and ethical constraints, many American judges have, sidestepped *Daubert* on the grounds that the test does not apply to "soft" science. They have proceeded to admit evidence of false confessions and suggestible personalities, post-traumatic stress disorder (PTSD), and repressed memory syndrome (Slovenko 1998). Meanwhile, Canadian courts have allowed expert evidence on "learned helplessness" to be adduced to dispel any jury myths about battered women (see *R* v *Lavallee* (1990) 55 CCC (3d) 97).

The reception of "syndrome evidence" has not met with universal approval. For instance, controversy surrounds the reception in rape trials of evidence of rape trauma syndrome. This is offered to demonstrate that the complainant's behavior is consistent with her allegation that she was raped. Critics suggest that the alleged syndrome is based on inconclusive research and could consequently cause great prejudice to the defendant. It has not been established whether rape trauma syndrome is a subset of post-traumatic stress disorder. It lacks specificity, in that many kinds of behavior appear to fall within its compass; the victim may be afraid to be indoors, or she may be afraid to go out of doors, depending on where the rape took place. Notwithstanding considerable skepticism, evidence of rape trauma syndrome has been admitted in some courts in the United States. Some American judges, however, allow it only as rebuttal evidence after the defense has argued that the complainant's behavior indicates fabrication. Given the adverse effect on the fact finder where a rape complainant displays no emotion, expert evidence to preempt the complainant's superficial serenity from being misconstrued may be advisable (Frazier and Borgida 1992). It appears, however, that expert evidence on general aspects of rape trauma syndrome has less impact on a jury than expert testimony linking rape trauma syndrome to specific features of the case (Brekke and Borgida 1988).

It has been argued that courts are being bombarded with an escalating number of novel afflictions, such as "urban survival syndrome," which could be derided as "trash syndromes." Over-willingness to accept syndrome evidence could lead to miscarriages of justice (Mosteller 1996). Repressed memory syndrome has been a particular target for criticism; it has been said that according to the falsifiability criterion, claims for repressed memory are not based on valid science. Yet there are psychiatric disorders, such as multiple personality disorder and attention-deficit hyperactivity disorder, listed in professional diagnostic manuals and commanding respect as legitimate medicine, which could equally be said to lack scientific validation (Richardson et al. 1995). One of the problems here is that techniques and theory developed for one purpose are being considered for use in a very different context. Conditions such as child sexual abuse accommodation syndrome or theories of repressed memory, developed originally for therapeutic purposes and therefore characteristically evaluated within their scientific communities in terms of alleviation of symptoms, are being scrutinized by judges for a level of objectively established reliability never sought or claimed for them. Courts which understand this, and mindful of the presumed ability of cross-examination to expose any overstated claim, tend to allow such evidence on the basis that *Daubert* permits a flexible approach and does not require general acceptance in the scientific community. To apply the *Daubert* test strictly would deprive fact finders of relevant bodies of knowledge. Yet to abandon any admissibility threshold or critical screening process in relation to expert evidence on behavioral issues would expose courts to the risk that plausible "junk science" would warp decision-making in criminal proceedings.

One might attempt to distinguish evidence presented as scientific knowledge, requiring compliance with *Daubert*, from specialist knowledge based on experience. The "experiential" category would include, for example, lip-reading expertise (*R* v *Luttrell* [2004] 2 Cr App R 31, CA), knowledge of black market drug prices (*R* v *Hodges* [2003] 2 Cr App R 247, CA), and handwriting comparisons. In cases involving testimony regarding battered woman syndrome, the evidence would have the beneficial effect of countering potential jury prejudice, but would not be masquerading as proven science. This is, effectively, the pragmatic strategy of many American courts (Renaker 1996). But if "junk science" is to be excluded, judges are left searching for a true test of genuine experience and knowledge. American and British judges appear to have found no alternative, where "soft science" is concerned, to reliance on their own intuitive assessment of the alleged field of expertise. Thus, while an American court may allow evidence of "psychological autopsies," retrospectively assessing the alleged murder victim's state of mind and tendency to suicide on the basis of medical records and interviews with family and friends (*State* v *Huber*, 62 Ohio Misc.2d 237), English courts remain unpersuaded: *R* v *Gilfoyle* [2001] 2 Cr App R 57, CA.

The work of criminal profilers, although heavily featured in film and television drama and, arguably, of use during police investigation, seems to be another area on which judicial opinion is divided – as is the expert community itself. From an objective point of view, psychological profiling seems to be an area of behavioral science far too undeveloped for use in a criminal trial. Profiling does not enjoy universal support among the ranks of psychologists, some of whom regard it as little more than a professional sideline with dubious scientific credentials (Gudjonsson and Copson 2000). It is likely that decision-makers who became aware of earlier criminal or other discreditable behavior by an accused person will be readier to convict (Tanford and Penrod 1984; Wagenaar et al. 1993), and so such evidence is particularly problematic when offered as an indicator of guilt as part of the prosecution case (Roberts and Zuckerman 2010, Chap. 14). Conversely, defense profile evidence was excluded from a rape trial on the ground that there was no proof that rapists have particular mental characteristics, or that psychiatrists could, by examination of a person, determine the presence or absence of these supposed characteristics: *State* v *Cavallo*, 88 NJ 508; 443 A (2d) 1020, (1982).

Behavioral Science Having Significant Legal Effect

The most well-known area in which the findings of behavioral scientists have influenced criminal trials concerns eyewitness evidence. Where identification is a contested issue in a trial, the ability of an eyewitness to link the accused person to the offense is a potentially invaluable asset to the prosecution. Direct observation by an individual of the perpetrator committing the offense is just about the best evidence that can be presented to the court. The true value of such evidence is, however, in considerable doubt following the findings of psychological experiments testing eyewitness reliability. According to these experimental simulations, accuracy of description varies between 65 % and 75 %, and decreases somewhat over time (Shepherd et al. 1982). Given that most developed legal systems demand prior identification in the relatively clinical conditions furnished by an identification parade before the witness can identify the defendant in court, the more significant data relate to the accuracy of recognition. Here, little reassurance can be offered to criminal lawyers. Although recognizing someone is easier than describing them, the accuracy of recognition is also alarmingly low. Across the multiplicity of studies, it varies between 35 % and 60 %. Although practical difficulties affecting the observation, such as lack of light or a short-lived event affording only a fleeting glimpse, exacerbate the problem, good observation conditions and prolonged observation periods make little difference to overall accuracy. High levels of error in recognition remain.

Identification parades provide only limited safeguards against wrongful identifications. If a parade is held at which the real actor is not present, there is a one-third likelihood of the eyewitness selecting a volunteer (Pigott et al. 1990). It appears that the eyewitness, anxious to choose *someone*, will select the line-up member who, compared with the other members of the line-up, most resembles the culprit (Wells 1993). If there is a familiar face in the line-up, it may be confused for that of the offender, the witness forgetting the context of a previous acquaintance. This is known as the "transference effect." Similar mistakes can be made where the person the witness saw resembled someone familiar, such as a well-known TV personality. There are extra layers of unreliability in identifications across races (Shepherd et al. 1974) and, more debatably, across genders and age groups (Jalbert and Getting 1992). In 1984, a black man was convicted of rape in the United States on the basis of an identification by the white female victim, who picked him out from a line-up. In 1995, DNA evidence established that another man had carried out the attack (Memon and McDaid 2000). The witness had been confident of the accuracy of her recognition throughout. Research indicates that eyewitness confidence is no guarantor of accuracy (Deffenbacher and Loftus 1982). The forensic impact of such confidence may be profound, however. Jurors' common sense presuppositions support the existence of a significant link between witness confidence and accuracy (Wells et al. 2002), and so an eyewitness who insists on the reliability of their memory may exert a very powerful influence on the outcome of the trial.

One way for the problem of eyewitness inaccuracy to be accommodated in a trial setting is to allow an expert witness to explain to the jury that identification evidence may be considerably less reliable than it appears, and to draw attention to any factors in the particular case that would cast doubt on the accuracy of the recognition in question. Uninstructed, jurors seem to be impressed by identification evidence, particularly if there is more than one witness. Judicial warnings apparently do little to moderate juror enthusiasm for eyewitness evidence (Cutler and Penrod 1995), so courts in some US states have attempted to amplify its effect by allowing expert testimony to explain the unreliability of identifications. Whether greater judicial attention should be devoted to the voluminous psychological literature on the accuracy of identification turns partly upon the extent to which it is possible to extrapolate directly from laboratory findings to real-life situations. In one naturalistic study, eyewitnesses interviewed following an armed robbery and fatal shooting in the street achieved strikingly high rates of accuracy. In follow-up interviews, accuracy did not decrease, but new information, also accurate, was offered. The recall of those who suffered the most stress was not adversely affected by their trauma (Yuille and Cutshall 1986). In any event, it would be unwise to dismiss all eyewitness evidence as equally unreliable. Certain individuals may have a good track record for recognition (Cutler and Penrod 1995). It is possibly misleading to claim that eyewitness confidence is no indicator of accuracy whatsoever, since the issue may be more complicated than some research literature allows. Eyewitnesses may legitimately feel confident about the accuracy of some of the details they recall while being less confident about others (Stephenson 1984). Also, there is a significant difference between a confident delivery and actually believing that one's identification is undoubtedly correct (King 1986). However, most of the relevant research focuses on subjects' beliefs in the correctness of their identifications rather than upon the personality or demeanor of particular eyewitnesses.

Another vital evidential issue illuminated by behavioral science research concerns the probative value of admissions made by suspects under criminal investigation. Admissions of guilt are regarded as key items of evidence for the prosecution, but have come under critical scrutiny from researchers. The result is increasing recognition of the extent to which the readiness to confess to a crime may be influenced not only by improper pressure from police officers, but by personality traits peculiar to the individual which may cause him or her to make false admissions. This may occur without any impropriety on the part of investigators. The vulnerability of some suspects, such as those with low or borderline IQ, may be clear to courts and can be taken into account when considering the reliability of their admissions. In some cases, however, factors indicating that a confession should be viewed with caution are less obvious. This does not necessarily mean that courts will routinely accept expert evidence on the reliability of a confession. They tend to be most receptive to such evidence where the defendant has a personality disorder or learning difficulties, that is, at the more overtly scientific end of behavioral science typically involving clinical diagnosis. Yet even in a nonclinical case, there may be powerful personality factors indicating that the suspect's confession should be treated with skepticism.

The psychological characteristics which need to be addressed are interrogative *suggestibility* and *compliance*, which are traits that may cause false admissions to be made once in police custody, even during perfectly proper and legitimate questioning by police officers. Suggestibility is a tendency to accept uncritically information communicated during questions. It is most pronounced in people of low intelligence. Gudjonsson (1992) has devised what he considers to be a reliable scale to assess suggestibility as a continuing condition, while allowing suggestibility may be aggravated by circumstances. Gudjonsson defines compliance as a tendency to go along with requests made by a person perceived to be in authority, even though the subject does not necessarily agree with or wholeheartedly endorse them. Building on these foundational concepts, Gudjonsson (1992) advanced a useful threefold classification of false confessions.

First, "voluntary" confessions are not induced by the police in any way; often there is no interview. The confession may constitute a response to publicity for a serious offense. Certain individuals are well known to their local police station for attempting to take responsibility for a whole series of crimes reported in the media. This kind of behavior may be motivated by a desire for notoriety, or to relieve a general feeling of guilt, or reflect an inability to distinguish fact and fantasy. A second kind of unreliable confession is the "coerced-compliant," a false confession motivated by the desire to escape from a highly stressful situation, such as police custody The immediate gain, frequently nothing more than the need for short-term resolution and predictability of future events, becomes a more powerful influence on the subject's behavior than the more uncertain long-term effects of the confession, even if the allegation concerns a serious offense. The majority of false confessions encountered in practice are coerced-compliant. Wagenaar concurs that common occasions for false confessions include the suspect taking an apparently easy way out, not anticipating the long-term consequences, perhaps thinking it will be possible to retract later, or choosing the only conceivable way out because the suspect is unable to cope with immediate pressures, for example, of confinement and interrogation, or a situation where the suspect is simply outwitted by the questioner – a not inconceivable event in the inevitably oppressive atmosphere of police custody (Wagenaar et al. 1993). A third category of false confession is the "coerced-internalized." People who do not trust their own memory may begin to accept the suggestions of apparently authoritative and knowledgeable individuals, like police investigators. Such suspects may become temporarily persuaded that they might have, or did indeed, commit the crime. Such a confession is more likely to be elicited by gentle, rather than aggressive, interviewing. It may be retracted later on, although the subject is more likely to stick to internalized admissions than is the coerced-compliant confessor who never truly accepted responsibility in the first place. Even if a confession is later withdrawn, the subject's memory may be permanently distorted (Wagenaar et al. 1993) and the confession itself might still be admissible in court.

Some legal jurisdictions recognize the risks involved in attributing significant evidential weight to confession evidence even in the absence of illegitimate pressure from the investigating authorities. Elsewhere, the psychology of false confessions is simply not recognized (see, e.g., Hodgson 2000's depictions of French criminal process). In Scotland and the Netherlands, no one can be convicted on the strength of a confession alone. In some US jurisdictions, a police station confession by the defendant, without corroboration, is not considered sufficient to sustain a conviction. In jurisdictions without a general corroboration requirement, there may be particular protections for vulnerable suspects, as in England and Wales, but they tend to allow weak corroboration, such as the suspect's "special knowledge" of circumstantial details that only the perpetrator of the crime would know, to support the admissibility of the confession. However, the notion that the suspect's guilt is reliably confirmed by his awareness of facts not communicated to the public is a dubious method of corroborating confessions. There are documented instances of suspects who confess in apparent knowledge of such information, only later to be proved innocent in the light of further revelations. They must have become aware of the relevant facts at some point during the period of detention, possibly through careless "confabulation" whereby police interviewers unwittingly feed information to the suspect (Kassin and Wrightsman 1988).

In prosecutions of domestic violence, theories of learned helplessness in consequence of domestic abuse have been influential in developing defenses or partial defenses, for example, in murder cases where prolonged exposure to domestic abuse has apparently caused the victim to respond by killing the abuser. Colloquially known as "battered woman syndrome," this psychological disorder is said to establish a causal relationship between the pattern of abuse suffered by the defendant, her psychological reactions to it, and her perception of her subsequent conduct. There is some evidence that lay persons, and therefore potential jurors, tend to consider that victims of domestic violence should leave home to escape it. This rationalist interpretation of behavior appears to be a powerful influence on jury reasoning, relatively impervious to expert evidence explaining why the obvious escape route is not employed in many such cases (Dodge and Green 1991). It is difficult to communicate to nonspecialists that prolonged experience of a highly controlled situation and exposure to violent sanctions for disobedience can reduce to vanishing point a person's subjective free choice. A battered partner might have genuinely (and sometimes even reasonably) believed that deadly force was the only way to escape from a perilous situation in which his or her own life was in mortal danger (Ewing 1987).

Future Directions

Behavioral science research and expertise have influenced many aspects of criminal investigation and the way evidence is presented to the court. Psychological autopsies and offender profiling have assisted the police in some cases. Sociological and psychological research has established that the emphasis on orality of testimony has had a markedly detrimental effect on a whole range of witnesses who have in the past struggled with the performance elements of the adversarial trial. In many jurisdictions, the evidence of vulnerable witnesses is now facilitated by technological innovations such as "live link" videoconferencing or videorecording (see, e.g., Roberts and Zuckerman 2010, Chap. 10). Criminal courts owe much to researchers in these fields. In addition, some aspects of behavioral science research are gaining ground in terms of providing evidence that may assist in reaching a verdict. Greater use of expert evidence in rape cases, in particular to dispel commonly held assumptions about the likely effect of having been raped upon a person's demeanor and likely response, could be very useful.

However, behavioral science is not yet ready to assist the court with the most problematic and the most crucial of the judgments that the jury has to make, namely, whether someone is speaking the truth or telling lies. Experts are not allowed to testify as to the credibility of a particular witness, although they may be able to explain the effect of a mental disorder on sufferers in general. At the same time, the literature on what, if any, exhibited behaviors constitute evidence of lying is at best inconclusive (Vrij 2000). Some years ago naïve faith in technological development encouraged flirtations with truth drugs and lie detector machines, but experience has since shown that they are not reliable indicators of dishonesty. Simple devices measuring levels of moisture on the palms of the hands may indicate nervousness, and nothing more (Saxe 1991). Polygraphs are used in criminal investigations in Canada, Israel, and Japan, but nowhere more than in the United States; albeit that courts in some US states will still not accept polygraph evidence (Patrick and Iacono 1991). In the United Kingdom, a working group produced such a devastating report on polygraph machines that the Government of the day abandoned any attempt to introduce them as a source of evidence in criminal trials (Working Group of the British Psychological Society on the Use of Polygraphs 1986).

The only conclusions to be drawn from the voluminous psychological literature on lying, credibility, and demeanor is that, first, neither laymen nor machines are accurate in recognizing liars or lies. Secondly, there is no valid scientific reason to believe that there is any observable indicator of lying behavior that an expert could use to identify it on behalf of fact finders in criminal adjudication. If the consequence of successful cross-examination in court is to produce behaviors commonly associated with lying but equally consistent with nervousness or confusion, the only inferential conclusions are negative. Behavioral science research certainly casts doubt on various familiar categories of evidence and challenges many commonly held lay assumptions capable of frustrating the courts' truth-seeking function. Unfortunately, however, it cannot provide much in the way of positive information as to where the truth of the case lies. The credibility of a witness, or the weight to attach to their evidence, lies in the two issues of reliability, depending on memory, and truthfulness. Behavioral science can tell us only of generalities, not specifics. Were it otherwise, jury dominance over the ultimate issue could be challenged. Unless and until our understanding of human conduct and thinking develops into a demonstrably convincing and coherent scientific discipline affording positive, case-specific guidance to fact finders, jurors (and victims and defendants) will be left at the mercy of their "common-sense" perceptions of behavior – prejudices, stereotypes, myths, and all.

Recommended Reading and References

Brekke N, Borgida E (1988) Expert psychological testimony in rape trials: a social cognitive analysis. J Pers Soc Psychol 55:372–386

Commission L (2011) Expert evidence in criminal proceedings in England and Wales (report). The Stationery Office, London

Cutler BL, Penrod SD (1995) Mistaken identification: the eyewitness, psychology and the Law. Cambridge University Press, Cambridge

Deffenbacher KA, Loftus EF (1982) Do jurors share a common understanding concerning eyewitness behavior? Law Hum Behav 6:15–30

Dodge M, Green E (1991) Jurors and expert conceptions of battered women. Victims Viol 6:271–282

Ewing CP (1987) Battered women who kill: psychological self-defence as legal justification. Heath, Lexington

Frazier PA, Borgida E (1992) Rape trauma syndrome; a review of case law and psychological research. Law Hum Behav 16:293–311

Graham MH (1998) The daubert dilemma: at last a viable solution. E P 2:211–223

Gudjonsson GH (1992) The psychology of interrogations, confessions, and testimony. Wiley, Chichester

Gudjonsson GH, Copson G (2000) The role of the expert in criminal investigation. In: Jackson JL, Bekerian DA (eds) Offender profiling: theory, research and practice. Wiley, Chichester

Hodgson J (2000) Comparing legal cultures: the comparativist as a participant observer. In: Nelken D (ed) Contrasting criminal justice. Ashgate, Aldershot

Jalbert NL, Getting J (1992) Race and gender issues in facial recognition. In: Lösel F, Bender D, Bleisener T (eds) Psychology and law: international perspective. De Gruyter, New York

Kassin SM, Wrightsman LS (1988) The American jury on trial. Hemisphere, New York, p 144

King M (1986) Psychology in or out of court. Pergamon, London, p 37

Memon A, Mc Daid AT (2000) Factors influencing witness evidence. In: McGuire J, Mason T, O'Kane A (eds) Behaviour, crime and legal processes. Wiley, Chichester

Mosteller RP (1996) Syndromes and politics in criminal trials and evidence law. Duke Law J 46:461–516

Patrick CJ, Iacono WG (1991) Validity of the control question polygraph test: the problem of sampling bias. J Appl Psychol 76:229–238

Pigott MA, Brigham JC, Bothwell RK (1990) A field study of the relationship between quality of eyewitness description and identification accuracy. J Police Sci Admin 17:84–88

Renaker TS (1996) Evidentiary legerdemain: deciding when daubert should apply to social science evidence. Calif Law Rev 84:1657–1692

Richardson JT, Ginsburg GP, Gatowski S, Dobbin S (1995) The problem of applying daubert to psychological syndrome evidence. Judicature 79: 10–16

Roberts P, Zuckerman A (2010) Criminal evidence, 2nd edn. OUP, Oxford

Saxe L (1991) Science and the GKT polygraph: a theoretical critique. Curr Dir Psychol Sci 3:69–73

Shepherd JW, Deregowski JB, Ellis MD (1974) A cross-cultural study of recognition memory for faces. Int J Psychol 9:205–212

Shepherd JW, Ellis J, Davies G (1982) Identification evidence: a psychological evaluation. Aberdeen University Press, Aberdeen

Slovenko R (1998) The daubert sequelae. E P 2:190–197

Smith R (1985) Expertise and causal attribution in deciding between crime and mental disorder. Soc Stud Sci 15:67–98

Stephenson GM (1984) Accuracy and confidence in testimony: a critical review and some fresh evidence. In: Muller DJ, Blackman AE, Chapman A (eds) Psychology and Law. Wiley, Chichester

Tanford S, Penrod SD (1984) Social inference processes in juror judgments of multiple-offence trials. J Pers Soc Psychol 47:749–765

Vrij A (2000) Detecting lies and deceit. Wiley, Chichester

Wagenaar WA, van Koppen PJ, Crombag HM (1993) Anchored narratives: the psychology of criminal evidence. Harvester Wheatsheaf, Hemel Hempstead

Wells GL (1993) What do you know about eyewitness identification? Am Psychol 48:553–571, at 560

Wells GL, Olson EA, Chapman S (2002) The confidence of eyewitnesses in their identification from lineups. Curr Dir Psychol Sci 1(1):151–154

Working Group of the British Psychological Society on the Use of Polygraphs, Report (1986) Bull Brit Psychol Soc 39: 81–94

Yuille JC, Cutshall JL (1986) A field study of eyewitness memory of a crime. J Appl Psychol 71:291–301

Bias Crime

▶ Hate Crime

Bias in Forensic Science

▶ Cognitive Forensics: Human Cognition, Contextual Information, and Bias

Bias/Hate Motivated Crime

▶ Hate Crime

Biased Policing

Seth W. Fallik[1] and Kenneth J. Novak[2]
[1]College of Criminal Justice, Sam Houston State University, Huntsville, TX, USA
[2]Department of Criminal Justice and Criminology, University of Missouri, Kansas City, MO, USA

Synonyms

Differential enforcement; Police or officer discretionary decision-making; Racial profiling; Selective enforcement

Overview

This entry will address the issues associated with police racial and ethnic bias, including research findings, research limitations, and policy issues related to reducing racial and ethnic disparities.

Introduction

Why are racial and ethnic minority citizens overrepresented at every stage of the criminal justice system? Answers to this reality are not clear or straightforward. While all criminal justice actors enjoy a degree of discretion in their decision-making, it is particularly important to determine whether bias is present in the early stages of the criminal justice process because initial biases can cause reverberations throughout the entire criminal justice process. The differential treatment patterns and practices of police in their decision-making have historically manifested itself in several ways (e.g., sexism, ageism) and garnered widespread attention. More specifically, a well of skepticism exists among citizens, scholars, legislatures, and the judiciary in response to the overrepresentation of racial and ethnic minorities at every stage of the criminal justice process. In response to the appearance of racial and ethnic animus in police decision-making, voluntary, legislative, and judicial data explorations of racial and ethnic biases in police discretionary decision-making have been encouraged.

The following manuscript addresses several issues associated with racial and ethnic bias among police. First, decision-making is contextualized as it relates to bias among policing discretion. Second, the difficulties of measuring bias-based policing are discussed within the context of future research endeavors. Third, research findings across various police decision-making points, including the use of force, the decision to arrest, and automobile stops, are presented. Fourth, commentary on contemporary manifestations of bias-based policing is set against a post-9/11 America. Fifth, bias-based policing results are tempered by some of the acknowledged limitations of prior research. Finally, a policy framework for reducing racial and ethnic disparities among officer discretionary decision-making is advanced.

Discretionary Decision-Making

Discretionary decision-making has been defined as "informal decision-making or [a] judgment by professionals based on unwritten rules, their training and their experience" (Samaha 2005, p. 10). Discretionary decision-making stands in opposition to mandatory acts, such as obeying the law and following departmental policies. Davis (1969) contextualized the relationship between law and discretion as "where law ends, discretion beings" (p. 3). Although this absolute dichotomy is false (see Walker 1993), law and discretion are related to the concept of procedural justice. Procedural justice refers to processes of fairness (Sunshine and Tyler 2003). Police discretionary decision-making is perceived to be fair when it is based on facts. Rarely is the legitimacy of fair outcomes called into question. Alternatively, discretionary decision-making is problematic when the judgments of police are seen as unfair, not based on law, and/or based on personal biases. This lead Walker (1993) to conclude that "[t]he problem is not discretion itself, but its misuse" (p. 4). More specifically to the topic at hand, bias-based policing or the selective/differential enforcement patterns and practices of police on minorities have the appearance of abuse.

Contrary to folklore, research indicates police receive overwhelming support from the public, and this support has been stable over time (Tuch and Weitzer 1997). There is, however, some variability in police support across races and ethnicities. These differences are most pronounced in African American and Hispanic communities (Withrow 2006). African Americans and Hispanics consistently report higher rates of dissatisfaction with police when compared with Whites (Ramirez et al. 2000). Although explaining the satisfaction gap between racial and ethnic groups is beyond the scope of this entry, one source for the disparity may be police differential treatment of minorities during encounters. Minorities believe police are more punitive during their encounters with police (Lundman and Kaufman 2003; Eith and Durose 2011). Beckett and Sasson (2004) noted that perceptions of being singled out impact attitudes of procedural justice and may explain "African Americans' growing alienation from the institutions of criminal justice" (p. 126).

When looking at the prevalence of actual police-citizen encounters, a disconnect between perceptual and actual disparities becomes apparent. The Bureau of Justice Statistics (BJS) has collected and reported on the racial and ethnic composition of police-citizen encounters every triennial since 1999. In the most recent report titled *Contacts between Police and the Public*, Eith and Durose (2011) estimated that 16.9 % of the population, 16 or older, will have some sort of face-to-face interaction with police (p. 1).

Among the 40.0 million citizens who encountered the police through citizen- and police-initiated contacts, 14.2 % were African American and 15.2 % were Hispanic (Eith and Durose 2011, p. 5). These statistics nearly mirror the proportion of racial and ethnic minorities in the general population.

Taken together, it appears as though the quality of interactions between police and racial and ethnic minorities may be the driving force behind differences in support. There are two observations on this issue. First, isolated instances of police misconduct usually have a bigger social impact than expected. For example, media sensationalism surrounding incidents like Rodney King, Abner Louima, and Amadou Diallo has the potential to generalize individual incidents of police misconduct to all law enforcement officials. Despite the isolated nature of events, racially and ethnically motivated events are, to a certain degree, vicariously experienced by all minorities. Second, bias-based policing may be more pronounced than these statistics indicate because contextual nuances are rarely expressed in national surveys. This is explored further in the following section (Walker and Meyers 2000).

Contextualizing Bias-Based Policing

Despite evidence of racial and ethnic neutrality among police, identifying race- and/or ethnicity-based police patterns and practices is a much more difficult task (Eith and Durose 2011). First, race and/or ethnicity are rarely absolute influencers of police behavior. Rather, bias typically manifests itself in other more subtle ways. The importance of race and/or ethnicity varies on a spectrum in which race and/or ethnicity is more or less influential among other exogenous factors (Higgins et al. 2008). Relevant factors have typically been found to fall within the following typologies: environmental, organizational, officer, and situational. Environmental factors suggest that the physical location in which the police-citizen encounter occurs has bearing on police decision-making. Organizational factors prescribe that departmental norms, culture, and structure differentially impact police decision-making. Officer factors charge that the characteristics of the officer, such as the age, race, ethnicity, experience, training, and rank of the officer, may influence decision-making. Situational factors, including citizen demographics and the reason(s) for police interaction, are also said to have pull in police decision-making. Finally, it is important to note that race and/or ethnicity may interact with each of these factors. Research found that minorities were significantly more likely to be arrested than nonminorities (Smith et al. 1984). The impact, however, vanished when the suspects' demeanor was taken into consideration and concluded that while race and/or ethnicity is influential, it does not have a direct impact on police decision-making. Similarly, other research found that police use significantly more force during encounters with minorities, yet the relationship between race and the use of force dissipated after controlling for neighborhood disadvantage and crime rates (Terrill and Reisig 2003). These examples highlight the complex nature of identifying bias within police decision-making.

The use of race and/or ethnicity may also be motivated by legally accepted practices (Schafer et al. 2006). These practices include drug courier profiles and "race out of place" policing. Drug courier profiling is a set of characteristics, including the race and/or ethnicity of the citizen, thought to be typical among persons carrying illegal drugs. Although highly subjective, police may stop, question, and under some situations search citizens fitting the drug courier profile. In choosing to engage citizens, race and/or ethnicity may be one of many factors contributing to police decision-making. Alternatively, "race out of place" policing indicates that race and/or ethnicity is the sole determining factor for police engagement with citizens. "Race out of place" policing refers to the legal practice of police engaging citizens based on a perceived inconsistency between the racial and/or ethnic identity of the citizen and the racial and/or ethnic composition of the ecological or neighborhood context (Ramirez et al. 2000). Each of these instances demonstrates the importance of distinguishing legal and extralegal factors when identifying race- and/or ethnicity-based policing patterns and practices. Given the contextual uniqueness of bias-based policing, the influence of race and/or ethnicity cannot be considered within a vacuum. Rather, many nuanced factors can be said to be important in police decision-making.

Measuring Bias-Based Policing Through Discretionary Decision-Making

Measuring police discretionary decision-making has been a difficult task because racial and ethnic animus is likely to be concealed within police behaviors that have low visibility. Since there is a great deal of difficulty in observing low-visibility police behaviors, the bulk of bias-based policing research has focused on systematic difference in decision-making outcomes. Furthermore, a host of discretionary decision-making outcomes with a variety of analytical approaches have been studied. The three most prominently featured discretionary decision-making points among bias-based policing focus on the use of force, the decision to arrest, and automobile stops. Each of these decision-making points provides measurable aspects to the cognitive processes of police discretionary decision-making.

The capacity to use coercive force is one of the defining features of police work. Racial and ethnic minorities consistently report the use of force against them is excessive. Among the people who contacted the police, only 1.4 % had force threatened or used against them (Eith and Durose 2011). Among threats and the use of force statistics, the most frequently cited type was being "pushed" or "grabbed" (53.5 %) (Eith and Durose 2011, p. 13). Although the threat and use of force is uncommon, African Americans and Hispanics are overrepresented among these instances. In fact, African Americans were more than twice (3.4 %) as likely to have force threatened or used against them than the national average (Eith and Durose 2011, p. 12). African Americans are also overrepresented among deadly force statistics.

Although consequences for the use of force are more immediate to citizens, the consequences of "arrests have a far more pervasive effect on peoples' lives" (Walker 1993, p. 39). An arrest occurs when a person is taken into custody for the purpose of criminal prosecution or interrogation. The evidentiary threshold for determining the appropriateness of an arrest is probable cause. Unfortunately for citizens, probable cause does not provide safeguards against arrest harassment or police failure to make an arrest. Further complicating the issue is the unreliability of arrest data because the point at which an individual is arrested is hard to identify and highly influenced by departmental norms. Some jurisdictions report arrests when citizens are formally restrained. Other jurisdictions, however, report arrests when citizens are formally booked. The difference between each reporting threshold represents a "dark figure" of arrests in which bias-based policing may be concealed (Erez 1984). Despite sporadic arrest reporting, official arrest statistics indicate that African Americans are overrepresented. Tillman (1987) estimated that nearly two-thirds (65.5 %) of African American males and one-third of African American females (29.6 %) will be arrested before the age of 30. This rate is nearly two times the rate of White males (33.9 %) and three times the rate of White females (10.1 %) arrested before the age of 30. While there is some indication that race and ethnicity may be an indirect cause for the decision to arrest, the equitable application of the probable cause standard appears to be a myth.

The final discretionary decision-making point measures automobile stop outcomes. Racial profiling, a phenomenon also known as "driving while Black/Brown," has gained parlance in American vernacular to represent the bias-based treatment patterns and practices of police on minorities during automobile stops (Alpert et al. 2007). Automobile stops present three unique opportunities for measuring the discretionary decision-making of police: stop initiation, searches, and the punitiveness of automobile stop outcomes. Research on automobile stop initiation focuses on the racial and/or ethnic composition of drivers stopped by the police. The bulk of drivers are stopped by the police for traffic enforcement-related issues, but more discretionary stops, such as investigatory stops, are also often captured. Among the citizens that encountered the police during automobile stops, 8.4 % were White, 8.8 % were African American, and 9.1 % were Hispanic (Eith and Durose 2011, p. 7). The relative parity of these statistics does not demonstrate the full dynamics of police-citizen automobile stops. Additionally, some researchers indicate that stop initiation poorly represents instances of bias-based policing because police officers can "only determine the race of the driver prior to the stop approximately 30 % of the time" (Alpert et al. 2007, p. 48). In response to these issues, researchers (Fallik and Novak 2012) have suggested that post-stop decision-making provides better indications of the presence of bias.

The second measured outcome from automobile stops concerns the officers' decision to search the driver, vehicle, passenger(s), or a combination of some or all three entities. African American drivers (12.3 %) were approximately three times more likely to be searched than white drivers (3.9 %) and approximately two times more likely to be searched than Hispanic drivers (5.8 %) (Eith and Durose 2011, p. 10). Despite being searched at a greater rate than Whites, minorities are no more likely to possess illegal contraband (Engel and Calnon 2004; Lundman 2004). Predicting searches becomes even more complicated when race and ethnicity are considered among other exogenous factors. Some researchers have concluded that race and/or ethnicity was one of numerous search predictors (Williams and Stahl 2008). Alternatively, while race remained a consequential predictor for Smith and Petrocelli (2001), Whites were nearly two and half times more likely to be the subject of consent searches. Consent searches are the most discretionary type of search legally permissible. Others found that the influence of race and/or ethnicity is neutralized once search types or typologies are specified (Fallik and Novak 2012; Schafer et al. 2006). There are eight search types permissible by Supreme Court precedent and police procedures: (1) searches incident to arrest, (2) inventory searches after a vehicle has been impounded, (3) searchers based on the presence of an existing search warrant, (4) probable cause searchers, (5) searches where contraband was discovered in plain view, (6) searches following a drug-sniffing dog alert, (7) "Terry" stop or pat-down searches, and (8) searches subsequent to the driver or passenger(s) given consent. When identifying typologies, each of the search types is typically categorized by the level of discretion required to execute the search, such as highly discretionary or non-discretionary. Finally, some researchers have discovered that race and/or ethnicity is not a significant predictor of searches (Higgins et al. 2008).

Similar nuanced inconsistencies can be said to exist for automobile stop punitiveness. Some research indicates that African Americans are more likely to be treated harshly by police sanctions (Engel and Calnon 2004). Alternatively, a more contextual awareness of less punitive outcomes among police-citizen encounters in that African Americans and Hispanics were less likely to receive punitive sanctions from the police but were more likely to be stopped for highly discretionary reasons, such as equipment violations and failure to signal (Novak 2004). This suggests that police may be using minor traffic violations as a pretextual motive for engaging minority drivers in automobile stops. An additional measure for an automobile stops' punitiveness considers how long citizens are detained. This body of research lacks consistency in the nature, strength, and sometimes presence of bias-based policing (Withrow 2006).

Contemporary Examples of Bias-Based Policing

Although most bias-based policing research has primarily focused on African Americans, other racial and ethnic groups have also been the targets of bias-based policing patterns and practices. Arab Americans, persons of Middle Eastern decent, and Muslims became the targets of bias-based policing. After September 11, over 12,000 Arab Americans, persons of Middle Eastern decent, and Muslims were "detained and held indefinitely" on suspicion of terrorism (Nguyen 2005, p. XVII). Furthermore, public opinions for the use of racial and ethnic profiling changed due to the fear society felt toward terrorist and terrorism. Prior to September 11, national opinion polls generally reported opposition to the use of race and/or ethnicity in policing, but after September 11, the majority of society supported the targeting of Arab Americans, persons of Middle Eastern decent, and Muslims in airports (Nguyen 2005) leading one scholar to explain, "after 9/11 the rules changed and everything we had learned about the social costs and ineffectiveness of racial profiling was largely ignored" (Withrow 2006, p. 244).

In response to terrorism threats at home, the US Justice Department began pressuring local and state authorities to enforce and enact immigration laws. Hispanics were particularly victimized by these efforts because of the loose US-Mexican border. Post September 11, political leaders "framed the border as a critical front in the war on terror" (Nguyen 2005, p. 92). For example, Arizona enacted the "Support Our Law Enforcement and Safe Neighborhoods" Act – also known as Senate Bill 1070 – which was designed to discourage illegal immigration in the United States. The bill required that police, during the course of lawful citizen contact, determine the immigration status of the people they encounter when there is reasonable suspicion to believe that the individual is an illegal alien, encouraging police to target Hispanic populations. Although the law never took effect, due to a federal injunction and a Supreme Court decision in 2012, it remains uncertain what the future will hold on this issue.

Research Limitations

Despite researchers' best efforts to explain the etiology of bias-based policing, two methodological controversies have arisen from this body of literature that are important to consider when drawing conclusions regarding the breadth and depth of bias in policing activities. First, there is some concern for officer self-reports measures of bias-based policing. Concern stems from the fact that officers may be aware of how their accounts of situations are being used and they may fear that accurately reporting some or all of their encounters with citizens may reflect poorly on them or the department. In automobile stops, research indicated that this may result in officers "ghosting' their data or recording race and ethnicity incorrectly to create the illusion of equitable stop and search procedures" (Williams and Stahl 2008, p. 231). This type of reactivity or Hawthorne effect greatly threatens the external validity of results. Alternatively, others found few signs of officer disengagement in his time-series analyses of encounters once it was announced that data collection would begin (Novak 2004). Given these inconsistencies, researchers should measure officer reactivity when using self-report data.

Second, measuring bias-based policing patterns and practices requires that benchmarks for performance be set. Benchmarks enable researchers to determine if police behavior matches expected decision-making outcomes. Deciding on an appropriate benchmark can be problematic with illusive and transitory populations. Related to automobile stops, "given a group of citizens stopped by the police (the numerator), what could be used as a denominator to conclusively determine whether certain drivers were stopped at a disproportionate rate?" (Schafer et al. 2006, p. 187). Over time, a variety of benchmarks have evolved, including census or modified census population estimates, information from drivers' licenses, not-at-fault accident data, blind enforcement data, systematic social observations of violator populations (e.g., "rolling surveys"), and internal comparisons. Despite the ecological fallacies found in all of these benchmarks, coverage error – from the benchmark to the expected outcome – is more pronounced under certain research contexts. Bias-based researchers should select the benchmark that minimizes coverage error within the research context while recognizing that all benchmarks have their own limitations.

To address both of these issues, future research must tie theoretical rationales to understandings of police discretionary decision-making by using multiple data sources, ideally coming from a triangulation of sources, including "police-reported, citizen-reported, and observer-reported data" (Lundman 2004, p. 343). Single source data explorations are often riddled with invalidity, inconclusiveness, and, worst of all, biases. A triangulation of data sources has the best potential to address each of these methodological issues. Efforts should be coupled with research that examines equitable and lawful policing that results in compliance by citizens (Piquero 2009). The etiology of police discretionary decision-making is best achieved from these methods.

Policy Implications

The improper use of race and/or ethnicity places the legitimacy of all law enforcement agencies in jeopardy, but efforts to curb the abuse of discretion must weigh individual rights against crime control objectives. To control the use of discretionary decision-making, several evidence-based practices have been proposed and implemented (Davis 1969; Gottfredson and Gottfredson 1988; Walker 1993). Four typologies have emerged among these strategies: structure, confinement, checking, and options.

Structural policies are rules, regulations, and guidelines that identify appropriate behaviors. An example of a structural policy geared toward bias-based policing is mandatory reporting. Mandatory reporting policies require that officers document every encounter they have with citizens. Although some mandatory reporting initiatives have been court ordered, many of these policies are agency initiated. By documenting the nature of encounters, departments have developed early intervention systems (EIS) to address issues of bias-based policing (Walker and Katz 2008). EIS is a tool used to identifying individual officers who disparately engage racial and ethnic minorities. These officers are more commonly known as "bad apples." The use of EIS has been encouraged by the shift toward electronic record keeping.

Confinement policies are rules, regulations, and guidelines that limit police behaviors. Examples of confinement are easily found in Supreme Court decisions. For example, the Supreme Court in *Whren v. United States* ruled that an officers' initial decision to engage citizens during automobile stops has to exceed the reasonable suspicion threshold. This ruling attempted to limit officer racial and ethnic pretextual motivations for stopping citizens. Although *Whren* may have encouraged police to hide behind minor traffic infractions when using race and/or ethnicity inappropriately during automobile stops, it confined police-initiated contacts to the reasonable suspicion standard.

Checking refers to the review of discretionary decision-making which can occur before, during, or after a decision is made. In order to be effective, the person or entity doing the checking must have a direct influence on the decision-making process. An example of a checking policy initiative used to combat bias-based policing is community review boards (CRBs). CRBs vary in the nature of the citizen input, review substance, breadth of the jurisdictional issues addressed, organizational structure, and operating policies. Functionally, CRBs provide external accountability to police-community-related issues. Police departments with CRBs have improved community-police relations, and cities with CRBs have higher rates of reporting. Both of these research findings indicate that communities with CRBs have greater confidence in police and the justice process. In essence, CRBs increase the visibility of citizen-reported police misconduct by providing an external review of instances in which an officers' discretionary decision-making may have gone awry (Walker 2001).

Perhaps the most radical policy suggestion presented in the research literature is the abolition of discretionary. While abolition has typically been focused on court discretionary decision-making, few supporters of abolition contest that police discretionary decision-making should remain unaffected. However, many researchers feel as though abolition is "unrealistic and ill-advised" (Gottfredson and Gottfredson 1988, p. 51). One of the extra option policy initiatives jurisdictions are moving toward would include issuing summons to appear in court, tickets, or fines for possession of person use marijuana.

Since drug crimes disproportionately affect minorities, the extra option – other than arrest – reduces the racial and ethnic disproportionality among arrest statistics (Johnson et al. 2008). Providing extra options may not address bias-based policing directly, but it has the ability to shift the disproportionality of discretionary decision-making to less punitive outcomes, thereby reducing the prevalence of more punitive outcomes. Furthermore, it allows researchers to more easily identify instances of bias in more punitive response outcomes.

Finally, although the ability to use coercive force may be the sine qua non of policing, extralegal police aggression – also known as police brutality – is not without safeguards (Holmes and Smith 2008). Each of the aforementioned strategies for reducing discretionary decision-making (i.e., structure, confinement, checking, and options) can be found in recent use of force policy responses. First, police training began incorporating the use of force continuum into their curriculum. The use of force continuum is a structural tool that teaches officers proper responses to escalating threats of violence. Second, the Supreme Court rulings in *Tennessee v. Garner* (1985) confined the use of deadly force to instances where a reasonable person would have acted on the threat of death or serious physical injury posed by a suspect. The check on the use of force occurs after the fact. When an officer discharges their weapon, departments often require that the officer explain the circumstances that precipitated the discharge of their weapon in a formal report. Supervisors review these reports before determining if it was a "good" or "bad" shot. Finally, police departments nationally began issuing and training officers in the use of nonlethal weapons, such as pepper spray, Tasers, and rubber bullets. Additional options allow police to apply the appropriate amount of force to the level of escalation each situation requires. While each of these typologies uniquely addresses discretionary decision-making, when combined they have reduced the racial and ethnic disparity of persons killed by police (Sherman and Cohn 1986).

Summary

Bias-based policing conjures up raw emotions. It is a complex social question rooted deeply within the historical and structural conditions of America. Acknowledging the existence of bias-based policing potentially compromises the legitimacy of the police in a democratic society. Ignoring or minimizing bias-based policing does not provide relief to targeted populations and does not hold individuals or organizations accountable. Research on the extent and impact of bias-based policing yields inconsistent conclusions. Furthermore, the impact of policies designed to address bias-based policing is difficult to determine. Given the current state of bias-based policing, the most productive conciliatory path is to join others in calling for further examination into the extent to which race and/or ethnicity impacts discretionary decision-making.

Related Entries

- ▶ Control of Police Misconduct
- ▶ Democratic Policing
- ▶ Law of Police Seizures and the Exercise of Discretion
- ▶ Law of Police Use of Force
- ▶ Legal Control of the Police
- ▶ Race and the Likelihood of Arrest

Recommended Reading and References

Alpert G, Dunham R, Smith M (2007) Investigating racial profiling by the Miami-Dade Police Department: a multimethod approach. Criminol Public Policy 6(4):671–678

Beckett K, Sasson T (2004) The politics of injustice: crime and punishment in America, 2nd edn. Sage, Thousand Oaks

Davis K (1969) Discretionary justice: a preliminary inquiry. Louisiana State University Press, Baton Rouge

Eith C, Durose M (2011) Contacts between police and the public, 2008. Office of Justice Programs, United States Department of Justice, Bureau of Justice Statistics, Washington, DC

Engel R, Calnon J (2004) Examining the influence of drivers' characteristics during traffic stops with police: results from a national survey. Justice Q 21(1):49–90

Erez E (1984) On the “dark figure” of arrest. J Police Sci Admin 12(4):431–440

Fallik S, Novak K (2012) The decision to search: is race or ethnicity important. J Contemp Crim Justice 28(2): 146–165

Gottfredson M, Gottfredson D (1988) Decision-making in criminal justice: toward the rational exercise of discretion, 2nd edn. Plenum Press, New York

Higgins G, Vito G, Walsh W (2008) Searches: an understudied are of racial profiling. J Ethn Crim Justice 6(1):23–39

Holmes M, Smith B (2008) Race and police brutality: Roots of an urban dilemma. State University of New York Press, Albany, NY

Johnson B, Golub A, Dunlap E, Sifaneck S (2008) An analysis of alternatives to New York City’s current marijuana arrest and detention policy. Institute for Special Populations Research, National Development and Research Institutes, Bethesda

Lundman R (2004) Driver race, ethnicity, and gender and citizen reports of vehicle searchers by police and vehicle search hits: toward a triangulated scholarly understanding. J Crim Law Criminol 11(2):309–350

Lundman R, Kaufman R (2003) Driving while black: effects of race, ethnicity, and gender on citizen self-reports on traffic stops and police actions. Criminology 41(1):195–220

Nguyen T (2005) We are all suspects now: untold stories from immigrant communities after 9/11. Beacon, Boston

Novak K (2004) Disparity and racial profiling in traffic enforcement. Police Q 7(1):65–96

Piquero A (2009) Finding the right balance between data, research, findings, and policy in racial profiling. Criminol Public Policy 8(2):371–379

Ramirez D, McDevitt J, Farrell A (2000) A resource guide on racial profiling data collection: promising practices and lessons learned. Department of Justice, National Institute of Justice, Washington, DC

Samaha J (2005) Criminal procedures, 6th edn. Thomas Wadsworth Learning, Belmont

Schafer J, Cater D, Katz-Bannister A, Wells W (2006) Decision-making in traffic stop encounters: a multivariate analysis of police behavior. Police Q 9(2):184–209

Sherman L, Cohn E (1986) Citizens killed by big-city police: 1974–1984. Crime Control Institute, Washington, DC

Smith M, Petrocelli M (2001) Racial profiling? A multivariate analysis of Police Traffic Stop Data. Police Q 4(4):4–27

Smith D, Visher C, Davidson L (1984) Equity and discretionary justice: the influence of race on police arrest decisions. J Crim Law Criminol 75:234–249

Sunshine J, Tyler T (2003) The role of procedural justice and legitimacy in shaping public support for policing. Law Soc Rev 37(3):513–547

Tennessee v. Garner, 471 U.S. 1 (1985)

Terrill W, Reisig M (2003) Neighborhood context and police use of force. Crime Delinq 40:291–321

Tillman R (1987) The size of the “criminal population”, the prevalence and incidence of adult arrest. Criminology 25:561–579

Tuch S, Weitzer R (1997) Racial differences in attitudes toward the police. Public Opin Q 4(61): 642–663

Walker S (1993) Taming the system: the control of discretion in criminal justice, 1950–1990. Oxford University Press, New York

Walker S (2001) Police accountability: the role of citizen oversight. Wadsworth, Belmont

Walker S, Katz C (2008) The police in America: an introduction, 6th edn. McGraw Hill, New York

Walker S, Meyers R (2000) Police interactions will racial and ethnic minorities: assessing the evidence and allegations. Police Executive Research Forum (PERF), Washington, DC

Whren v. United States, 517 U.S. 806 (1996)

Williams B, Stahl M (2008) An analysis of police traffic stops and searches in Kentucky: a mixed methods approach offering heuristic and practical implications. Policy Sci 41:221–243

Withrow B (2006) Racial profiling: from rhetoric to reason. Pearson Education, Upper Saddle River

Bicycle Theft

Aiden Sidebottom and Shane D. Johnson
Department of Security and Crime Science,
University College London, London, UK

Overview

Bicycle theft is defined as the unauthorized removal of a nonmotorized pedal cycle. This usually refers to the theft of an entire bicycle but can also include theft of component parts and accessories such as lights, seats, and wheels. Bicycle theft can take many forms, reflecting the range of motivations of bicycle thieves. Some cycles are stolen for transportation purposes, others to joyride or to trade in for cash or drugs; *specific types* of bicycle may be stolen to order; a cycle may be stolen to facilitate another crime – an example of a *crime multiplier*; or a bike may be stolen in response to an offender’s bicycle being taken, thus comprising part of a *crime chain*. Awareness of the heterogeneity of cycle theft is important; different forms of cycle theft are likely to demand different preventive responses.

Several studies suggest that bicycle theft is a common problem. However, compared to other high-volume property crimes, bicycle theft has been the subject of limited academic research, whether in criminology or related fields such as transport and urban planning. This is often attributed to the incompleteness of relevant police-recorded crime data where reporting rates and detection rates are both low. Failures to report may be explained by victim expectations that the chances of either detection or recovery of the bicycle are slim. Most lacking is reliable evaluations of what works to reduce cycle theft.

The spiraling obesity rates experienced in many industrialized regions in recent years (OECD 2011) have, among other things, prompted government agencies to promote cycling as a healthier alternative to motorized transport. Opportunity theories of crime would predict that increases in the population of cyclists will be associated with increases in the opportunities for cycle theft. This is a concern given that research suggests that bicycle ownership (a proxy for opportunities), measured at the country level, is strongly correlated with national levels of cycle theft. Moreover, the risk of cycle theft, real or perceived, is found to be a significant barrier to cycling (Rietveld and Koetse 2003). Cycle theft hence threatens to jeopardize strategies to better integrate cycling as a sustainable, health-promoting form of transport. Enriching our knowledge of cycle theft and determining effective ways to reduce it are therefore of social as well as academic importance.

This entry sets out what we know about cycle theft and the progress made in reducing it. It begins by describing the extent of cycle theft internationally and articulates challenges associated with measuring it. Next, some of the harms caused by cycle theft are outlined, particularly theft as a potential barrier to cycling uptake. Third, the major known correlates of cycle theft victimization are addressed. This is followed by a discussion of efforts to prevent cycle theft, and we conclude by outlining as yet unanswered research questions.

The Extent of Bicycle Theft

Several studies suggest that bicycle theft is a common problem and that levels are highest in countries where cycling is popular. For example, using data from the International Crime Victim Survey (ICVS) and the European Survey on Crime and Safety (EU ICS), Van Dijk et al. (2007) report that bicycle ownership levels are strongly correlated ($r = 0.76$, $n = 30$, $p < 0.05$) with national levels of bicycle theft. They find that the risks of cycle theft are considerably higher in Northern Europe, where cycling is widespread, than in North America and Australia where cycling is less prevalent. To illustrate, the 1-year victimization prevalence rates for 2003/2004 in Holland were 6.6 per 100 population, whereas in Australia, the figure was only 1.2. Despite this variation, across countries cyclists were consistently found to be around three times more likely to have their bike stolen than car owners their car or motorcyclists their motorbike (Van Dijk et al. 2007).

In the USA, the ICVS indicates that levels of cycle theft remained stable between 1998 and 2004, with a mean rate of 2.8 victims per 100 population. In terms of official crime statistics, the FBI classifies bicycle theft as an example of “larceny-theft,” alongside offenses such as shoplifting and pocket-picking. In 2009, bicycle theft constituted 3.4 % of all incidents of larceny, equating to over 200,000 offenses. This likely underestimates the true extent of the problem as many thefts fail to reach the police because of underreporting. Data from the ICVS, for example, suggest that roughly half of all bicycle thefts are reported to the police (Van Kesteren et al. 2000). The US National Crime Victimization Survey showed a fivefold difference between cycle thefts that are reported to the police and those that actually occurred in 2008.

A similar picture is observed in England and Wales. Police-recorded crime data indicate only minor yearly fluctuations in the levels of cycle theft since 1997, with around 100,000 thefts recorded in 2010/2011. This contrasts sharply with estimates from the British Crime Survey (BCS) which suggest that over the same year, around 526,000 incidents occurred. Moreover, while recorded crime figures depict cycle theft as largely static over the past decade, for the same interval of time, estimates from the BCS suggest an upward trend both in the volume of cycle theft and rates per 10,000 bicycle-owning households. The true figures are probably higher still given the BCS excludes those aged under 16 from the survey sampling frame. Furthermore, viewed against a backdrop of general reductions in property crimes (for both data sources) in England and Wales since 2002/2003 (and elsewhere in the world), the increase in bicycle theft relative to changes in other types of acquisitive crime is more dramatic than the absolute change.

Explanations for the "dark figure" of cycle theft are varied. Interviews with cycle theft victims (Bryan-Brown and Saville 1997) reveal that many fail to notify the police because they believe they are unlikely to apprehend offenders or recover stolen cycles. Insofar as available data can be used to examine this, this notion appears to be justified. For example, police detection rates for bicycle theft are consistently low. In Sweden, where cycle theft is common, only 1 % of cycle thefts result in an arrest (Bra 2008). In England and Wales, sanction detection rates in 2008/2009 were marginally better at 5 % (Walker et al. 2009) but still low compared to other types of offenses. Poor police detection rates may partly be explained by the fact that for many bike thefts, there is little relationship between the offender and the victim, and thus identification of suspects is difficult. Another factor is the proof-of-ownership problem. Many cyclists cannot provide adequate proof that they own their bicycle. A study conducted by the London Metropolitan Police Service found that of the 836 cyclists surveyed, over three quarters did not know their cycle frame registration number (see Halliwell and Brown 2011). This frustrates reunification efforts because should stolen cycles be recovered by the police, there is often insufficient evidence to pursue the case and detained offenders will often be released without charge. Even when offenses are reported, Gamman et al. (2004) argue that many fail to be processed further because bicycle theft is considered a low police priority. A similar observation is made in China where Zhang et al. (2007) claim that the police tend to limit their attention to repeat bicycle thieves.

The Harms of Cycle Crime: Theft as a Barrier to Cycling

In many urban settings, current levels of obesity as well as road congestion and traffic-related air and noise pollution are generally considered to be excessive. Various initiatives have been proposed to reduce automobile dependency and encourage the use of bicycles. These include changes to infrastructure (such as the creation of cycling lanes), better integrating cycling with other forms of public transport (such as increasing cycle parking facilities at transport nodes), and increasing the provision of cycles (such as bike-sharing schemes). At this time the results appear mixed. While cycling remains relatively rare in many industrialized nations (compared to motorized transport), other countries have seen encouraging increases involving a small fraction but quite large numbers of individuals. Cycling in London, for example, has increased by 91 % since 2000 (TFL 2008). Similar trends, albeit less dramatic, have also been observed in several European cities (Oja and Vuori 2000) and in parts of the USA (Xing et al. 2008).

Commentators argue that cycle theft poses a threat to attempts to increase cycling rates – one that presently has received inadequate attention (see Sidebottom et al. 2009). Research finds that theft, and the fear of cycle theft, is a barrier to cycle use. Experience of theft can deter individuals from cycling due to an inability to replace the stolen bike or due to (justified) apathy that it may happen again. This is supported by interviews with cycle theft victims that find a sizable portion of cyclists do not replace their bicycle once stolen and/or cycle less frequently (Davies et al. 1998; Mercat and Heran 2003). Direct or vicarious experience of theft may also increase the likelihood of cyclists' purchasing cheap, low-quality cycles and associated protective measures (such as locks) due to an assumption that theft is an inevitable "occupational hazard." Tentative support for this comes from a study by Mercat and Heran (2003) in which half of all cycle theft victims who purchased a replacement bike opted for a second-hand rather than a new cycle.

Reducing cycle theft may also yield wider social benefits. A review of numerous programs to increase physical activity internationally reports strong evidence that as the number of cyclists and pedestrians increases, the frequency of collisions between those groups and motorists *decreases* (Jacobson 2003). The putative mechanism underlying this pattern is that motorists modify their driving behavior when they expect or encounter cyclists and walkers. To slightly overstate the point, effectively reducing the risks of cycle theft may indirectly contribute to improvements in the safety of cyclists.

The Bicycle as Crime Target

Bicycles are attractive targets for theft. They fit Clarke's (1999) CRAVED model which outlines the attributes commonly associated with frequently stolen items. Johnson et al. (2008) suggest that bicycles are:

Concealable – Widespread availability of cycles in many urban settings enables cycle thieves to remain inconspicuous when riding away on a stolen bicycle, rendering the crime concealable.

Removable – If poorly locked, bicycles are easy to remove. With respect to theft *from* cycles, quick-release features (such as wheels or seat posts) that are inadequately secured also require little effort to steal.

Available – More bikes provide more opportunities for cycle theft, as well as a greater demand for bicycles and component parts.

Valuable – While many bicycles are available cheaply, others can cost in advance of $3,000 dollars, thereby constituting a desirable crime target. There is also often an imbalance between the value of the bike and the value of the security measures used to protect it.

Enjoyable – Bicycles offer utility as a form of transport, particularly in regions where public transport is inadequate and/or expensive. From an offender's perspective, a stolen cycle can also facilitate further crimes through increasing the speed and area in which they can forage for crime targets.

Disposable – Thieves can easily sell stolen bikes, either "whole" or "piecemeal." Abundant buyer's markets for stolen bikes, swelled by the internet, may also provide an incentive to steal.

Correlates of Victimization

Research has identified several factors that are reliably associated with the risk of cycle theft. Consistent with the national trends described above, opportunity plays a central role. Cycle theft is found to concentrate at locations where opportunities are plentiful. Common examples include in and around the victim's home, at university campuses, or at transport hubs. Estimates from the BCS suggest that over half of all reported cycle thefts in England and Wales occur immediately outside the victim's home (i.e., bikes stored in sheds or gardens), thereby constituting the most common location for cycle theft. University campuses often house a large proportion of high-performance and expensive bicycles that attract bicycle thieves. A study by Lovejoy and Handy (2011) at the University of California, Davis, shows that 19 % of sampled cyclists report experiencing cycle theft on campus during their tenure, with 9 % suffering a theft in the previous year. A regular supply of opportunities in the form of parked bicycles also explains why railway stations are common locations for theft. Data from the British Transport Police, who are responsible for policing the railway system of England, Wales, and Scotland, show that recorded cases of theft and damage to bicycles increased by 74 % between 2002/2003 (3,277 offenses) and 2006/2007 (5,686 offenses) and have remained at comparatively high levels since (in 2010/2011 there were 5,859 offenses). Other forms of vehicle theft and criminal damage fell considerably over the same period.

The way in which cyclists lock their bicycles is a further factor. Several studies report that stolen bicycles are often locked insecurely (Mercat and Heran 2003; Roe and Olivero 1993). A study comprising over 8,500 observations of cycle locking events in the UK found "secure" locking techniques – therein measured as instances where both wheels and the frame were attached to parking furniture – to be atypical; other (less secure) ways of parking a bicycle were dominant (Sidebottom et al. 2009). Such findings are important in highlighting the potential limitations of schemes that aim to increase the use of robust locks but neglect to provide guidance as to how they should be applied.

Insecure locking behavior is also likely to be a function of what a cycle is locked to. In urban settings many bicycles are *flyparked* (Johnson et al. 2008): fastened to street furniture not designed for that purpose (i.e., railings, parking meters, and trees). Flyparked cycles may be at greater risk of victimization because such furniture affords little scope for secure locking (as defined previously). While this is yet to be systematically tested, recorded crime statistics from the London Borough of Camden, England, indicate that of all bicycles reported stolen between 2004 and 2005, 72 % were flyparked. From a town planning perspective, high levels of flyparking may indicate insufficient or inappropriate (poorly positioned) formal bike parking opportunities.

In one of the few empirical studies not to employ Anglo-American data, Zhang et al. (2007) assessed household- and area-level determinants of residential cycle theft victimization in the city of Tianjin, China, using data collected as part of a household survey. Informed by the routine activity approach, they find, among other things, that the risk of cycle theft is positively associated with neighborhood-level crime and deviance, therein taken to be a proxy measure for exposure to offenders. The number of adults per household (a proxy for guardianship) was found to be a significant protective factor against cycle theft.

Research across many crime types demonstrates that previous victimization is a reliable predictor of future victimization. While the evidence for repeat cycle thefts is limited, those studies that are available suggest the same patterns are also apparent. Analyzing survey data from the Netherlands, Wittebrood and Nieuwbeerta (2000) report that prior victimization was significantly associated with the probability of further incidents of bicycle theft. This is consistent with data from Melbourne, Australia, which indicate that under a third of bicycle theft victims accounted for 60 % of all thefts reported (Johnson et al. 2008). Put differently, a small proportion of cycle theft victims disproportionately account for a large number of cycle theft victimizations. Extending this concept further, Johnson et al. (2008) show that the risk of cycle theft victimization displays a contagion-like quality across space and time. Analyzing recorded crime data from Dorset, England, they find that in the wake of a cycle theft and for a period of around 4 weeks, further incidents were more likely to occur at locations nearby and up to a distance of about 450 yards.

Prevention

Johnson et al. (2008) provide a review of various interventions designed to reduce cycle theft and categorize them as follows: (1) interventions which seek to catch bicycle thieves *in the act*, (2) initiatives designed to deter bicycle thieves through focusing on the registration and recovery of bicycles (thereby making stolen bicycles more risky to dispose of), (3) schemes to improve the security of bicycle parking facilities (both cycle parking furniture and parking facilities as a whole), and (4) interventions which try to increase the security of locks and/or the manner in which they are applied.

Despite various interventions being implemented, few robust evaluations of attempts to reduce cycle theft are available. Situational measures designed to prevent (rather than detect) bicycle theft through altering cyclists' behavior in ways that increase the effort and risk associated with stealing cycles have yielded positive results. Two studies evaluating situational measures warrant mention. The first concerns a targeted publicity campaign in which stickers – designed to improve cyclists' locking practice by depicting how to lock a bicycle securely – were attached to a series of bicycle parking stands in on-street public cycle parks in the UK (Sidebottom et al. 2009). Pre- and post-intervention observations found modest albeit statistically significant improvements in the security of cyclist's locking practices at bicycle parking stands where the intervention was fitted compared to the control sites. Importantly this pattern was consistent across two different settings. Sidebottom and colleagues conclude that communication strategies of this kind might constitute a cheap, scalable "quick win" against cycle theft.

The second intervention concerns the design of bicycle parking furniture. As alluded to previously, how secure a cycle is parked is partly a function of what it is locked to. The common Sheffield (or ∩-shape) stand, for example, makes it difficult for cyclists to lock the frame and both wheels. Many only lock the top crossbar of their bicycle as it runs parallel to the horizontal bar of the Sheffield stand. Alternative designs have therefore been developed in an attempt to encourage more secure locking styles. An example is the caMden M-shaped stand, which is designed in such a way so as to remove the opportunity for cyclists to lock the crossbar to the stand, encouraging the cyclist to instead apply a more effective locking practice, such as securing both the wheels and frame. Thorpe et al. (2012) describe a study in which the impact of seven prototype parking stands designed to increase securer locking practices was evaluated (some examples are in Fig. 1). Using data gathered through 3,563 observations, they report significant improvements in locking practice at the prototype bicycle stands compared to the control (Sheffield) stands. They also report considerable variation *between* the prototype stands in terms of their effectiveness and popularity.

A limitation in the above studies is noteworthy. Both evaluations use locking practice as an intermediate outcome measure which is hypothesized to precede the ultimate outcome measure of interest (a reduction in bike theft). In both studies, it was important to conduct an intermediate outcome measure evaluation because if the interventions are not found to change the locking practice of cyclists, then there would be no reason to expect an impact on levels of cycle theft. The reason the study authors did not conduct an ultimate outcome measure evaluation was due to shortcomings in the recorded crime data available, specifically the level of geographical resolution associated with incidents of bicycle theft. Briefly, many recorded bicycle thefts are often reported as having taken place at landmarks, shops, or streets rather than specific locations, and consequently the exact site of the theft

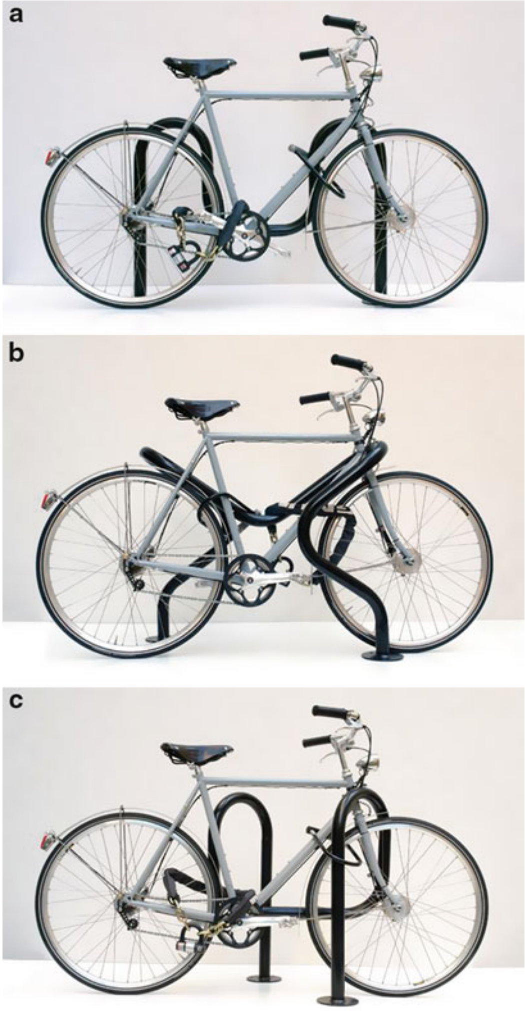

Bicycle Theft, Fig. 1 Examples of the prototype bicycle parking stands (Source: Thorpe et al. 2012) event cannot easily be determined. This lack of information may, in part, reflect the low priority status that some suggest the police attach to cycle theft. This lack of specificity is problematic because the interventions described in the above studies were implemented at *specific* cycle parking sites located on streets where other – nonintervention – cycle parking opportunities are also available. This presents an obvious problem for analysis because it would be unclear whether stolen bicycles were fastened to "treatment" street furniture or elsewhere (be that control parking stands or flyparked). This brings forth two issues. First, such problems are likely generalizable to other settings and should be considered in advance of conducting comparable studies. Second, further research is needed to determine whether, as hypothesized, positive changes in the locking practice of cyclists are associated with reductions in cycle theft.

Future Directions

We began this entry by highlighting the lack of bicycle theft research hitherto. Various areas of relevance to policy and practice remain unexplored. The following are considered to be some of the most important. They are outlined with the intent of stimulating the reader to initiate further research.

The Stolen Bike Market

It is a widely held assumption that many bicycles are stolen with the intention of being resold for profit. Indirect evidence in the form of low recovery rates of stolen cycles lends support to this claim. It follows that disrupting the market for stolen cycles holds much potential for effectively reducing the problem. Evidence on the use of *market reduction approaches* (see Sutton 2005) to reduce other forms of acquisitive crime suggests that such strategies are promising. These schemes employ various tactics to make the buying and selling of stolen goods riskier for offenders. Regrettably, little is known about the market for stolen bicycles, specifically the role of online (legitimate) auctions as a means of disposal. The proof-of-ownership problem implies that few bicycles could easily be identified as stolen, which aids the resale of stolen bikes and reduces the risk of apprehension. Research concerned with better illuminating the dynamics of stolen bike markets is therefore required.

Cycle Hire Schemes: Criminogenic or Criminocclusive?

Cycle hire schemes are increasingly common in many urban settings. The rationale is that increasing the availability of cycles may lead to increases in the number of individuals cycling – the sought-for objective. As already mentioned, increased cycle usage – particularly as an alternative form of urban transport – may generate increases in the levels of cycle theft. Alternatively, cycle hire schemes may lead to reductions in theft by providing secure storage facilities and (as is often the case) by requiring cyclists to register their personal details, including banking information, before they can access cycles thereby reducing anonymity which would otherwise facilitate cycle theft. Furthermore, where financial details are required to use the cycles, this should reduce the rewards associated with stealing hire cycles.

Successful cycle hire schemes could reduce bicycle theft in other ways. For example, if such schemes result in an increase in the number of cyclists, and if the parking facilities for cycle hire schemes are located near to other cycle parking facilities, this may provide additional natural surveillance at such facilities throughout the day. Furthermore, the components used to manufacture cycle hire bikes can often only be used for that type of bicycle. For example, the London cycle hire scheme uses bicycles constructed of oversize, non-common parts which are incompatible with non-hire bikes. An advantage of this is that such cycles are unlikely to be stripped for parts (see Johnson et al. 2008) as they are not typically the sort of bicycle that would be attractive to own, largely because they are designed so as to be clearly identifiable as for-hire bicycles, thereby making them unattractive to thieves who might seek to steal them, or those who might want to purchase a second-hand bike.

To our knowledge, no published research has explored the impact of cycle hire schemes on the levels and patterns of cycle theft. Such research could usefully guide prevention efforts as well as inform the operation of cycle hire schemes.

Residential Cycle Theft

Victimization surveys maintain that the majority of bicycles are stolen from in and around the victim's home. Yet the research which is available on bicycle theft is overwhelmingly concentrated on theft of bicycles from public locations. This asymmetry may relate to the ease with which agencies (and individuals) with the ability to influence the conditions conducive for cycle theft can be mobilized to behave differently; there is arguably greater scope, through levering public agencies responsible for crime prevention and community safety, to manage and reduce opportunities for cycle theft in public places than to persuade individuals in or around their homes to modify their behavior. Either way, there is much to be gained by analyzing the patterns of residential cycle theft with a view to determining practical ways to tackle it.

The Broken Bike Effect

Abandoned bicycles are common. In calculating the volume of cyclists at the University of California, Lovejoy and Handy (2011) estimate that 23 % of total rack capacity was taken up by abandoned cycles. The relationship between abandoned bikes and theft is yet to be empirically examined, but observational research by Gamman et al. (2004) suggests that bicycle parking stands adjacent to damaged or abandoned bicycles are less popular among cyclists, even when located closest to the destination served. Two outcomes are envisaged, both leading to increases in cycle theft. First, they argue that on encountering abandoned bikes, cyclists may be provoked to park their cycles elsewhere, possibly leading to an increase in flyparking (which as described often affords less secure parking opportunities than street furniture which is designed with bike parking in mind). Second, they suggest that cycle parks containing damaged or abandoned bicycles may signal to offenders that this is an area lacking in guardianship thereby constituting a conducive location in which to offend. Gamman et al. (2004) refer to this phenomenon as the "broken bike" effect, invoking the mechanism underpinning the well-known broken windows theory (Kelling and Wilson 1982).

The broken bike effect is yet to be empirically tested, but could be. For example, in a recent study Keizer et al. (2008) examined broken windows theory using an experimental manipulation. Briefly, they observed peoples' behavior in a real-world setting for a range of conditions. In one experiment, leaflets were attached to cyclist's parked bicycles, and for one condition other leaflets were scattered (by the experimenters) on the floor nearby. In the "control" condition, no leaflets were littered in this way. The empirical question was whether the way in which cyclists' would dispose of the leaflets was affected by the experimental condition. That is, would those who discovered the attached leaflets in the presence of litter also be more likely to litter than their counterparts? The results of the study were quite conclusive, clearly demonstrating that where litter was present, cyclists were also more likely to litter than when it was not. While there may be ethical issues associated with conducting an analogous experiment; that is, by seeing if cycles are more likely to be stolen where "broken bikes" are introduced at a parking facility and when they are not, it would be possible to see if the removal of existing damaged or abandoned bikes has an effect on rates of cycle theft.

In closing, it is evident that cycle theft is a considerable problem. Relative to other property that people are likely to own, such as an iPad, they are also valuable. However, they are routinely left unguarded around the home and in public spaces thereby making them attractive targets to thieves. Recent studies have focused on the problem of bicycle theft, but it is quite clear that more research is required if we are to understand how to prevent this type of crime. As has been discussed, if cycle theft continues to increase, it is possible that this will be an impediment to the uptake of cycling and hence the sustainable transport agendas of many countries.

Related Entries

▶ Designing Products Against Crime
▶ Fencing/Receiving Stolen Goods
▶ Motor Vehicle Theft

Recommended Reading and References

Bra The Swedish National Council for Crime Prevention (2008) Bicycle theft. Available online at: www.bra.se/extra/pod/?action=pod_show&id=38&module_instance=11

Bryan-Brown K, Saville T (1997) Cycle theft in Britain. TRL report 284. Transport Research Laboratory, Crowthorne

Clarke RV (1999) Hot products: understanding, anticipating, and reducing demand for stolen goods. Police Research Series Paper 112. Home Office Policing and Reducing Crime Unit, London

Davies D, Emmerson P, Gardner G (1998) Achieving the aims of the national cycling strategy: summary of TRL research. Transport Research Laboratory, Crowthorne

Gamman L, Thorpe A, Willcocks M (2004) Bike off! Tracking the design terrains of cycle parking: reviewing use, misuse, and abuse. Crime Prev Commun Saf An Int J 6(4):19–36

Halliwell T, Brown S (2011) Reducing cycle theft: a partnership approach between transport for London and the metropolitan police service Winner of the Herman Goldstein Award for Excellence in Problem-Oriented Policing. Available online at: http://www.popcenter.org/library/awards/goldstein/2011/11-54%28W%29.pdf

Jacobsen P (2003) Safety in numbers: more walkers and bicyclists, safer walking and bicycling. Inj Prev 9(3):205–209

Johnson SD, Sidebottom A, Thorpe A (2008) Bicycle theft. Problem-oriented guides for police series. Washington, D.C.: U.S. Department of Justice, Office of Community Oriented Policing Services

Keizer K, Lindenberg S, Steg L (2008) The spread of disorder. Science 322(5908):1681–1685

Kelling G, Wilson J (1982) Broken windows: the police and neighborhood safety. Atlantic Mon 249(3):29–38

Lovejoy K, Handy SL (2011) Mixed methods of bike counting for better cycling statistics: example of bicycle use, abandonment, and theft on University of California, Davis, Campus. Transportation Research Board 90th annual meeting, Washington, 23–27 Jan 2011

Mercat N, Heran F (2003) Bicycle theft in France. In: Tolley R (ed) Sustainable transport: planning for walking and cycling in urban environments. CRC Press/Woodhead Publishing, Boca Raton/Cambridge, England

OECD (2011) Health at a glance 2011: OECD indicators. Available online at: http://www.oecd.org/dataoecd/6/28/49105858.pdf

Oja P, Vuori I (2000) Promotion of transport walking and cycling in Europe: strategy directions. The European network for the promotion of health-enhancing physical activity. UKK Institute, Tampere

Rietveld P, Koetse M (2003) Crime and offenses in transport. In: Dullaert W, Jourquin B, Polak JB (eds) Across the border: building upon a quarter century of transport research in the Benelux. De Boeck, Antwerpen

Roe L, Olivero J (1993) Profiles in bicycle theft: there is a pattern. J Secur Adm 16(2):17–24

Sidebottom A, Thorpe A, Johnson SD (2009) Using targeted publicity to reduce opportunities for bicycle theft: a demonstration and replication. Eur J Criminol 6(3):267–286

Sutton M (2005) Complicity, trading dynamics and prevalence in stolen goods markets. In: Tilley N (ed) Handbook of crime prevention and community safety. Willan Publishing, Cullompton

Thorpe A, Johnson SD, Sidebottom A (2012) Designing against bicycle theft. In: Ekblom P (ed) Design against crime: crime proofing everyday products. Crime prevention studies 27. Lynne Rienner Publishers, Boulder, Colo

Transport for London (2008) Cycling in London. Available online at: http://www.tfl.gov.uk/assets/downloads/businessandpartners/cycling-in-london-final-october-2008.pdf

Van Dijk JJM, Van Kesteren J, Smit P (2007) Criminal victimisation in international perspective: key findings from the 2004–2005 ICVS and the EU ICS. Boom Juridische Uitgevers, The Hague

Van Kesteren J, Mayhew P, Nieuwbeerta P (2000) Criminal victimisation in 17 industrialized countries: key findings from the 2000 international crime victims survey. Ministry of Justice, The Hague

Walker A, Flatley J, Kershaw C, Moon D (2009) Crime in England and Wales 2008/09: Volume 1 Findings from the British Crime Survey and police recorded crime. Home Office, London

Wittebrood K, Nieuwbeerta P (2000) Criminal victimization during one's life course: the effects of previous victimization and patterns of routine activities. J Res Crime Delinq 37(1):91–122

Xing Y, Handy SL, Buehler TJ (2008) Factors associated with bicycle ownership and use: a study of 6 small U.S. cities. Submitted to the Committee on Bicycle Transportation. Available online at: http://www.des.ucdavis.edu/faculty/handy/Bike_Draft_11.14.pdf

Zhang L, Messner S, Liu J (2007) Bicycle-theft victimization in contemporary urban China: a multilevel assessment of risk and protective factors. J Res Crime Delinq 44(4):406–426

Biker Clubs/Gangs

▶ Motorcycle Clubs or Gangs?

Biological Geographical Profiling

Mark Stevenson and Steven Le Comber
School of Biological and Chemical Sciences,
Queen Mary, University of London,
London, UK

Overview

Geographic profiling was originally developed as a statistical tool in criminology, where it uses the spatial locations of linked crimes (e.g., murder, rape, and arson) to identify areas that are most likely to include the offender's residence. In criminology, geographic profiling uses these crime locations to create a probability surface that is overlaid on the study area to produce a geoprofile. Geoprofiles do not provide an exact location for the criminal's home, but allow the police to prioritize investigations by systematically checking suspects associated with locations in descending order of the height of these locations on the geoprofile, facilitating an optimal search process based on decreasing probability density. The technique has been extremely successful in this field, and it is now widely used by police forces and investigative agencies around the world. Recently, the same techniques have begun to be applied to biological data, initially in the field of animal foraging and hunting behavior, but also in epidemiology and invasive species biology, where geographic profiling can be used to locate the sources of infectious disease or of invasive plants and animals as a prelude to targeted control efforts. As the technique has now been shown to be useful in such divergent scenarios from those for which it was originally developed, it raises the intriguing possibility that geographic profiling could be a useful general tool for studying spatial patterns in biological data. Here, we review the work in this area, and suggest further avenues for future research. We go on to consider ways in which this highly successful transfer of ideas from criminology to biology might also work in the opposite direction.

Introduction

Geographic profiling is well established in criminology, with a proven track record of success. It has recently begun to be applied to biological data. In this entry, we will first of all outline the basic ideas underlying geographic profiling (section "Introduction"), before reviewing the existing studies of biological data using geographic profiling, and considering how these differ from the model's application in criminology (section "From Criminology to Biology: Applications of Geographic Profiling to Biology"). Having discussed how ideas from geographic profiling have fed into biology, we will suggest ways in which ideas from biology might in turn feed-back into geographic profiling in general, and criminology in particular (section "From Biology to Criminology: Future Developments").

Geographic Profiling

Geographic profiling is a statistical technique used in criminology to prioritize large lists of suspects in cases of serial crime, such as murder, rape, or arson. The need for such a technique arises because criminal investigations often generate too many, rather than too few, suspects; for example, the enquiry into the Yorkshire Ripper murders in the UK between 1975 and 1980 generated 268,000 names (Doney 1990). Obviously, it will frequently be impractical to follow up all but a handful of these names. Clearly, any technique that allows police forces to prioritize such lists of suspects is likely to be of enormous value.

In essence, the model is simple and depends on two concepts: (1) distance decay and (2) the buffer zone (Rossmo 1999; Le Comber et al. 2006). The first concept relies on the fact that, because traveling incurs costs in time, effort, and/or money, most crimes occur relatively close to an "anchor point" (usually the criminal's home or workplace) (for instance, 70 % of arsons occur within two miles of the arsonist's home (Sapp et al. 1994)). However, the anchor point is also typically surrounded by an area (the buffer zone) in which offenses are relatively rare. This buffer zone arises partly because of increases in detection risk related to reduced anonymity within the criminal’s local neighborhood, and partly because the number of criminal opportunities increases geometrically with distance traveled from home. The size of the buffer zone is specific to an individual criminal and location, since it will be affected by the criminal’s willingness and ability to travel, and also the underlying distribution of opportunities for crime (the “target backcloth” (Rossmo 1999)).

Geographic profiling uses the opposing effects of distance decay and the buffer zone to calculate the probability of offender residence for each location within the study area, producing a three-dimensional probability surface (called a jeopardy surface). Locations in which it is more likely that the offender might live are indicated by higher points on the jeopardy surface. Overlaying the 3-D jeopardy surface on to a search area map produces a geoprofile. Hence, geoprofiles do not provide an exact location for the criminal’s home, but they allow the police to prioritize search locations by starting with the highest point on the jeopardy surface. Systematically checking locations in descending order of their height on the geoprofile probability surface describes an optimal search process based on decreasing probability density. Therefore, the better the geographic profiling model performs, the shorter the search before the real location of the offender’s home is found (Rossmo 1999).

Because geographic profiling seeks to describe an optimal search, the model’s performance can be assessed simply by asking how high up the jeopardy surface, the true anchor point lies. This measure, which is expressed in terms of the hit score percentage (HS%), is the proportion of the study area (in criminology, this is usually the area bounding the crimes, plus an additional 10 % of the area surrounding this, to allow offender source locations to come from outside the area of the crime sites) that must be searched before the true anchor point is found. The smaller the HS%, the more accurate is the geoprofile. A hit score of 50 % is what would be expected from a nonprioritized (i.e., random or uniform) search; thus, a hit score of 10 % describes a search which is five times more efficient than a random search (Rossmo 1999).

From Criminology to Biology: Applications of Geographic Profiling to Biology

Although geographic profiling was originally designed to apply to crimes such as murder, rape, and arson, it has had numerous success in other areas, including burglary, counter-insurgency and piracy (see, e.g., Kucera (2005) and Rossmo and Harries (2011)). Based on its applicability to a range of problems in a variety of different fields, its application to biological data was an obvious next step. Given the similarities between criminal hunting behavior and animal behavior, it is perhaps not surprising that the first paper to apply geographic profiling in a biological context looked at animal foraging (Le Comber et al. 2006).

Geographic Profiling and Animal Foraging

Geographic profiling was introduced to biology in a 2006 paper in the *Journal of Theoretical Biology* (Le Comber et al. 2006). In this study, the authors used data from radio-tracking studies of two species of bat, the common and soprano pipistrelles (*Pipistrellus pipistrellus* and *P. pygmaeus*) in north-east Scotland. A previous study had identified both roost sites and foraging sites, and the authors fitted Rossmo’s criminal geographic targeting (CGT) model (Rossmo 1999) for each bat and showed that the fitted model parameters (*B*, *f*, and *g*) could be used to locate roost sites, using foraging sites as input, analogous to crime sites. Interestingly, the fitted values differed between the two species, despite their close genetic relatedness. This probably reflects their different foraging strategies; *P. pygmaeus* forages preferentially in riparian habitats (i.e., along the edges of rivers and lakes) that support higher numbers of insect taxa (Gressit and Gressit 1962; Townes 1962), while *P. pipistrellus* is more generalist. This specialization in *P. pygmaeus* is likely to mean that this species must forage over greater distances to locate sufficient prey items to satisfy its energetic demands. This was an intriguing result, suggesting that when anchor points such as nests, roosts, or dens are known, fitted CGT model parameters could provide a concise way of describing complicated foraging patterns.

The bat study was followed by a second study of animal foraging, this time in bees, but using laboratory data rather than field data (Raine et al. 2009). Bees were allowed to enter a flight arena approximately 1 m square, via a central hole, and allowed to forage on artificial flowers containing a sucrose solution. Again, the CGT algorithm successfully located this entrance. Fitting model parameters in the same way as in the bat study also showed that when the artificial flowers were presented at higher density, the size of the buffer zone decreased. This was of interest because, in criminology, little may be known about the target backcloth, since law enforcement agencies will have information on crimes committed, but not always on potential crimes that were not.

Another interesting extension of this study involved using "virtual" bees, in a similar experimental design to the real bees, using a variety of plausible foraging algorithms (including spiral searches, nearest-neighbor methods, and a variety of others). Just as the fitted model parameters could be used to differentiate between the foraging patterns of the two bat species, they could be used here to distinguish between different foraging rules. Crucially, for biologists, these could also be compared to the behavior of the real bees, allowing the authors to rule out some of the suggested foraging algorithms as inconsistent with the patterns observed in the real bees.

At about the same time, Martin et al. (2009) used geographic profiling to study great white shark predation on seals off the coast of South Africa. Again, much of the interest of this paper derived from aspects tangential to the main purpose of geographic profiling, that is, identifying sources for point pattern data. In this case, the study identified a well-defined search base or anchor point 100 m seaward of the seal's primary island entry-exit point. This is not where the chances of intercepting seals are greatest, and the authors suggested that it represented a balance between prey detection, capture rates, and competition. In addition, the different geoprofiles observed for sharks of different ages showed that smaller sharks exhibited more dispersed search patterns and had lower success rates than larger sharks, suggesting either that hunting success improved with experience or that larger sharks excluded smaller sharks from the most profitable areas.

Geographic Profiling and Epidemiology

As noted above, animal foraging behavior has much in common with criminal hunting behavior. The extension of geographic profiling to epidemiological datasets, however, involves several important differences, notably the increased importance of multiple sources, and the possibility, for some diseases, of secondary sources. These issues are discussed in section "Differences Between Biology and Criminology."

The application of geographic profiling to epidemiological data fills a surprising gap in epidemiology. As Buscema et al. (2009) pointed out, classical epidemiology tends to model the spread of infectious epidemic diseases, and few attempts have been made to identify the origin of the epidemic spread. This is surprising because, as Le Comber et al. (2011) noted, in many diseases, infection sources can be highly clustered: For example, malaria parasite transmission is strongly dependent on the location of vector breeding sites, and most transmission only occurs within short distances of these sites; in Africa, these distances are typically between a few hundred meters and a kilometer, and rarely more than 2–3 km (Carter et al. 2000). Because of this clustering, untargeted control efforts are highly inefficient. Although source reduction of mosquito larval habitats can dramatically mitigate malaria transmission (Yohannes et al. 2005; Gu et al. 2006; Walker and Lynch 2007; Gu and Novak 2009), the transient nature and diversity of potential vector breeding sites makes the identification and control of breeding sites difficult (Carter et al. 2000). As a result, evidence-based targeting of interventions is more efficient, environmentally friendly, and cost-effective than untargeted intervention. This, of course, is exactly why the problem geographic profiling was designed to solve.

The first attempt to apply geographic profiling to epidemiological data was by Buscema et al. (2009). This study examined Chikingunya fever, foot and mouth disease, and cholera, but concluded that geographic profiling was less efficient than the authors' preferred artificial intelligence method, the H-PST (Hidden-Pick and Squash Tracking) Algorithm. However, this study mistakenly used the distance between the peak of the geoprofile and the correct source as a measure of model performance. As Rossmo (1999) was careful to point out, geographic profiling does not attempt to provide a point estimate for the anchor point (here, the infection source), as methods such as spatial mean, spatial median, and center of minimum distance seek to do; rather, it describes an optimal search strategy. Because of the complexity of jeopardy surfaces, the distance from the peak of the geoprofile to the anchor point is irrelevant; what matters is what percentage of points within the study area have higher Z values than the anchor point. In fact, when there are multiple sources of infection (e.g., the malaria cases examined by Le Comber et al. (2011)) this is an important advantage of geographic profiling over the H-PST algorithm, since methods that provide point estimates of sources will typically perform poorly when there is more than one source. When Le Comber et al. (2011) revisited one of the case studies in the Buscema paper (John Snow's data on the 1854 London cholera outbreak (Snow and Frost 1936)), geographic profiling performed extremely well.

Geographic Profiling and Invasive Species Biology

Another promising area for the application of geographic profiling to biological research concerns the spread of invasive species, an area with more in common with epidemiology (e.g., multiple and secondary sources) than with animal foraging. The issue is not trivial; invasive species are now viewed as the second most important driver of world biodiversity loss behind habitat destruction and have been identified as a significant component of global change (Vitousek et al. 1996; Wilcove et al. 1998). The cost of invasive species can run from millions to billions of dollars per occurrence (Mooney and Baker 1986; Pimentel et al. 2001), and invasive species have been shown to affect native species through predation and competition, modify ecosystem functions, alter the abiotic environment, and spread pathogens (Strayer et al. 2006; Ricciardi 2007). In addition, the problem is likely to get worse as climate change and anthropogenic influences lead to increased range shifts (Hulme 2007). For these reasons, prevention and control of invasive species has been identified as a priority for conservation organizations and government wildlife and agriculture ministries globally (Mooney and Baker 1986; Hulme 2006).

Although only one study has looked at invasive species and geographic profiling, the results are promising. Stevenson et al. (2012) analyzed historical data from the Biological Records Centre (BRC: http://www.brc.ac.uk/) for 53 invasive species in Great Britain, ranging from marine invertebrates to woody trees, and from a wide variety of habitats (including littoral habitats, woodland and man-made habitats). For 52 out of these 53 datasets, geographic profiling outperformed spatial mean, spatial median, and center of minimum distance as a search strategy. The study also compared fitted parameter values between different species, groups, and habitat types, with a view to identify general values that might be used for novel invasions where data are lacking, with some success.

Differences Between Biology and Criminology

The first applications of geographic profiling to biology involved fairly straightforward mapping of the basic concepts from criminology: In these studies, animal foraging sites were used to identify animal roosts (or other home locations) in the same way that crime sites are used to identify probable areas of offender residence in criminology. However, later extensions, and most notably studies of invasive species biology and epidemiology, differ in a number of areas.

In criminology, the application of geographic profiling will usually (or at least often) deal with the crimes of single individual with a single anchor point, often (hopefully!) over a short period of time. In contrast, biological data can involve multiple organisms (and hence multiple anchor points), secondary anchor points, and extended time periods.

Multiple Anchor Points

In criminology, although jeopardy surfaces may have several peaks, relating perhaps to the criminal's home, work place, or a relative's home (or, in the case of the Hillside Strangler, the two homes of the two cousins who committed the crimes together; Rossmo (1999)), it is usually assumed that the crimes are linked; that is, they are carried out by a single individual (some applications of geographic profiling to terrorist activities may be an exception). In invasive species biology or epidemiology, it is usually impossible, or at least impractical (e.g., perhaps requiring expensive genetic testing to identify particular strains of virus, or genotypes of individual plants or animals), to link events to individual sources. For example, the malaria cases in Le Comber et al. (2011) were treated as a single group of "crimes," although it is possible that six or more *An. sergentii* breeding sites were involved. In this case, data was simply pooled and the heights of each potential source on the geoprofile examined separately; Stevenson et al. (2012) took a similar approach with invasive species. At this point, no studies have explicitly examined the effect of multiple sources on geographic profiling model performance, although the data in Le Comber et al. (2011) and Stevenson et al. (2012), along with some simulation data (unpublished), suggest that geographic profiling's performance relative to simple measures of spatial center tendency (spatial mean, spatial median, center of minimum distance) will increase as the number of sources increases.

Secondary Anchor Points

Murder victims do not go out and commit murders; victims of arson do not go out and burn down other buildings. Similarly, in the context of animal foraging, seals predated upon by great white sharks do not then predate upon other seals. However, the sites of new biological invasions can go on to act as sources for further waves of invasion; similarly, in many disease systems, infected individuals will go on to affect other individuals. These secondary sources/anchor points may dramatically alter the spatial patterns observed.

Extended Time Periods

In criminal investigations, the persistence of a series of linked crimes over a number of years obviously represents a failure of law enforcement; cases such as Jeffrey Dahmer (1978–1991) or the Yorkshire Ripper (1975–1980) (Rossmo 1999) are, hopefully, an exception. In biology, this need not be the case, and longer-term datasets may in fact be highly desirable. Ecological datasets in particular can span decades or even centuries (Stevenson et al. 2012), and can involve multiple "outbreaks," while criminal cases typically span shorter periods of time. In this sense, biological data may offer a distinct advantage over criminological data. Invasions and disease outbreaks have long histories and repeated outbreaks, so assuming that repeated invasions follow similar histories, previous outbreaks (perhaps with known sources) can be used to validate the geographic profiling model. Future spread can then be predicted using parameters established from the organism's own invasion history.

From Biology to Criminology: Future Developments

Clearly, geographic profiling has already made important and interesting contributions to biology, in fields including animal foraging, invasive species biology, and epidemiology. In this section, we will consider how insights from biology might feedback into geographic profiling theory, and criminology generally. Broadly speaking, these fall into three classes: (1) mathematical developments; (2) spatial methods; and (3) experimental methods.

Mathematical Developments

The underlying mathematics of geographic profiling has recently attracted attention, notably from O'Leary (2009, 2010). Here, we will briefly discuss four areas of interest. These are (1) incorporating a Bayesian framework; (2) fitting model parameters; (3) incorporating explicit models of behavior; and (4) considering different mathematical distributions in addition to the exponential functions used in the Rossmo model (Rossmo 1999).

Bayesian Statistics

Recent papers on the mathematics of geographic profiling, notably those of O'Leary (2009, 2010), highlight two different approaches to the subject. Because of its origins in criminology, one of these approaches is highly practical, concentrating on the model's use as a tool in investigations of serial crimes such as murder, rape, and arson. Most applications of the model within biology to date have taken a similar approach, with the main results of the various studies (Le Comber et al. 2006; Martin et al. 2009; Raine et al. 2009) being to demonstrate the applicability of the model to different types of data. O'Leary's papers take a different tack, considering the underlying mathematics themselves, with less attention paid to the model's practical utility. It might be argued that these are two parallel avenues of research that are unlikely to intersect. However, our view is that it might be possible to bring these two strands together. The acid test, of course, will be whether different mathematical approaches (e.g., embedding the model within a Bayesian framework, or considering other underlying spatial distributions such as the Cauchy distribution (see below)) can improve the model's performance.

Fitting Model Parameters

Rossmo's (2009) model uses three parameters, *B*, *f*, and *g*. *B* is the width of the buffer zone, while *f* and *g* together determine first the increase in the probability of a crime occurring moving outward from the anchor point toward the edge of the buffer zone, and second the decrease moving further beyond this. In criminology, the width of the buffer zone is typically set at half the mean nearest-neighbor distance, with *f* and *g* set at 1.2. To date, studies in biology have either adopted this method (Martin et al. 2009; Le Comber et al. 2011), or attempted to fit spatial data to known anchor points (Le Comber et al. 2006, 2011; Raine et al. 2009). This may matter because multiple sources (see above) could lead to different patterns of "crime sites," depending on their number and proximity to each other; there may also be some use in using fitted model parameters as descriptors of more complex spatial patterns, as in Raine et al. (2009). Thus, the issue of precisely how model parameters are fitted may be come of interest. Although to date published studies have used only simple methods (see, for instance, (Le Comber et al. 2006; Raine et al. 2009)), an obvious approach is to use Markov Chain Monte Carlo (MCMC) methods to explore parameter space. In our view, this is likely to be an interesting area of development, although care will have to be taken to avoid well-known problems of over-fitting (Hawkins 2004).

Incorporating Explicit Models of Behavior

Current geographic profiling models are generic. They express simple geometric patterns such as exponential or normal decay. The Rossmo model incorporates *B*, an explicit parameter for the buffer zone, but *f* and *g* remain as exponential decay functions. Rather than fitting a generic statistical model, the aim of mathematical modeling is to find true relationships that underlie the data. Ideally we should have models that have parameters that explicating relate to some aspect of predator behavior, invasive species dispersal, or epidemic spread.

Distribution Models

Models of offender behavior will obviously depend on the precise spatial distribution of crime sites. Canter and Hammond (Canter 2006) examined logarithmic, exponential, and quadratic functions, while the Rossmo model (Rossmo 1999) uses an exponential function, as does CrimeStat (Levine 2009). O'Leary (2010) introduced the idea of alpha as a single parameter from a normal distribution, suggesting that it could be used as a predictive parameter to represent average offense distance; the optimized parameter *B* in (Stevenson et al., 2012) fulfills similar criteria if *f* and *g* are fixed in the Rossmo model. In a later study, O'Leary (2010) compared single- and two-parameter normal, and single- and two-parameter exponential functions, concluding that there was no difference between normal and exponential in the single-parameter models in Baltimore county burglaries; neither of the two-parameter models performed as well as the single-parameter models.

The distribution models appropriate for offender behavior are likely to be substantially different to models appropriate for disease and animal dispersal. Invasive species dispersal is likely to be strongly nonnormal; dispersal includes both short local dispersers and long escalated movements: for example, Levy flight (see, e.g., (Viswanathan et al. 2000; Viswanathan and Bartumeus 2002; Bartumeus 2009)). This type of dispersal is described by a Cauchy distribution (Viswanathan et al. 2000). The Cauchy distribution is related to normal distributions, in that dividing one normal distribution centered on zero by another centered on zero will yield a Cauchy distribution. In fact, the Cauchy distribution can be made to resemble a smoother version of the Rossmo distribution, but avoiding the latter's sharp peak at the radius of the buffer zone. In addition, Cauchy distributions also allow the inclusion of a width or thickness parameter, gamma, which describes the "fatness" of the distribution's tail, relating to the amount of longer dispersal events. A Cauchy distribution could also be used in criminology, and would relate to the occurrence of two different types of jumps already documented in the criminology literature, prowlers, and commuters (Rossmo 1999).

Running Models Forward in Time

Finally, one intriguing possibility following O'Leary's work (2009, 2010) is that of running models forward in time – rather pleasingly, this would link current models of spatial epidemiology such as risk mapping (Leung et al. 2002) with geographic profiling, which is essentially retrospective in nature. O'Leary suggests using integration to generate a function that can then be used to predict future spread. The possibility of producing risk maps for offenders to predict possible sites of future burglaries or even more serious crimes is a real possibility, even when based on a small number of data points, in contrast to current hot spot mapping.

Spatial Developments

In the same way that mathematical studies are beginning to feed-back into geographic profiling, it seems likely that techniques from spatial epidemiology may offer interesting avenues for research; in fact, as noted above, O'Leary's proposed approach (2009, 2010) may go some way toward allowing techniques from spatial epidemiology and spatial ecology to feed-back into geographic profiling.

One fruitful area is likely to be ecological niche mapping. Ecological niche mapping is used in invasion biology and macroecology to predict species spread based on associated habitat types (see, e.g., Kaschner et al. (2006)). Species found in particular ranges are associated with particular ecological factors; these can then be mapped to new ranges to predict spread based on these niche factors. We suggest that this type of modeling can be used to generate priors or hyperpriors for a Bayesian model of geographic profiling. True prediction requires knowledge of not only where the species is now and where it is dispersing to, but also what habitat is suitable for it to live in. Our prior estimation of where a species is capable of dispersing to will inform the geoprofile to produce a surface that includes both types of information. Obviously, this approach could also be applied to criminology, using information about neighborhood quality, street lighting, open spaces, frequency of police patrols, and so on.

Experimental Developments

One advantage the biologist has over the criminologist is that, in biology, experiments are far easier, and certainly less ethically problematic. For example, Raine et al. (2009) were able to manipulate the target backcloth in a study of bee foraging behavior that would simply not be possible in the context of crime. Thus, biology might be better placed to explicitly test some of the underlying assumptions and methodologies of geographic profiling.

A related point, alluded to above, is that in biology datasets, long-term datasets are more abundant, and replication is more exact: For example, in invasive species biology, it is possible, and even common, to have (1) repeated invasions of the same geographical area by the same species; (2) repeated invasions of different geographical areas by the same species; (3) repeated invasions of the same geographical area by different species. A rigorous comparison of these different cases may help to disentangle those aspects of any observed spatial patterns that are due to the invader's behavior and/or biology, and the habitat's own attributes. The analogous exercise – understanding what aspects of spatial patterns of crimes are due to the criminal's behavior, and what aspects are due to the peculiarities of the target backcloth – is much more difficult.

Conclusions

Although the application of geographic profiling to biological datasets is still relatively new, the early indications are that the method's considerable success in the field of criminology may be replicated in areas as seemingly diverse as animal foraging behavior, epidemiology, and invasive species biology. This is encouraging because, despite the obvious similarities between criminal hunting behavior and animal foraging behavior, there are several key differences that can arise in biology, notably the importance of multiple and secondary anchor points. To date, most studies using geographic profiling in biology have concentrated on demonstrating the method's utility. In contrast, much of the mathematical research has pursued a parallel line of enquiry, ignoring the method's application. Future work may bring these two seemingly distinct approaches closer together, and we suggest that, just as methods from criminology have proved of use in biology, some techniques from biology – for example, niche modeling, Levy flight, and optimization methods – may prove useful in criminology.

Recommended Reading and References

Bartumeus F (2009) Behavioral intermittence, Lévy patterns, and randomness in animal movement. Oikos 118:488–494

Buscema M, Grossi E, Breda M, Jefferson T (2009) Outbreaks source: a new mathematical approach to identify their possible location. Phys Stat Mech Appl 388:4736–4762

Canter D (2006) A comparison of the efficacy of different decay functions in geographical profiling for a sample of US serial killers. J Investig Psychol Offender Profil 3:91–103

Carter R, Mendis K, Roberts D (2000) Spatial targeting of interventions against malaria. Bull World Health Organ 78:1401–1411

Doney R (1990) The aftermath of the Yorkshire Ripper: the response of the United Kingdom police service. In: Egger SA (ed) Serial murder: an elusive phenomenon. Praeger, London

Gressit J, Gressit M (1962) An improved malaise insect trap. Pac Insects 4:87–90

Gu W, Novak RJ (2009) Agent-based modelling of mosquito foraging behaviour for malaria control. Trans R Soc Trop Med Hyg 103:1105–1112

Gu W, Regens J, Beier J, Novak R (2006) Source reduction of mosquito larval habitats has unexpected consequences on malaria transmission. Proc Nat Acad Sci USA 103:17560–17563

Hawkins DM (2004) The problem of overfitting. J Chem Inf Model 44:1–12

Hulme PE (2006) Beyond control: wider implications for the management of biological invasions. J Appl Ecol 43:835–847

Hulme P (2007) Biological invasions in Europe: drivers, pressures, states, impacts and responses. *Biodiversity under threat issues in environmental science and technology* 25:56–80

Kaschner K, Watson R, Trites AW, Pauly D (2006) Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. Mar Ecol Prog Ser 316: 285–310

Kucera H (2005) Hunting insurgents – geographic profiling adds a new weapon. GeoWorld 37:3

Le Comber SC, Nicholls B, Rossmo DK, Racey PA (2006) Geographic profiling and animal foraging. J Theor Biol 240:233–240

Le Comber SC, Rossmo DK, Hassan AN, Fuller DO, Beier JC (2011) Geographic profiling as a novel spatial tool for targeting infectious disease control. Int J Health Geogr 10:35

Leung B, Lodge DM, Finnoff D, Shogren JF, Lewis MA, Lamberti G (2002) An ounce of prevention or a pound of cure: bioeconomic risk analysis of invasive species. Proc Biol Sci 269:2407–2413

Levine N (2009) CrimeStat: a spatial statistics program for the analysis of crime incident locations, vol 3.2 a. Ned Levine & Associates, Houston

Martin RA, Rossmo DK, Hammerschlag N (2009) Hunting patterns and geographic profiling of white shark predation. J Zoology 279:111–118

Mooney H, Baker H (1986) Ecology of biological invasions of North America and Hawaii. Springer, New York

O'Leary M (2009) The mathematics of geographic profiling. J Investig Psychol Offender Profil 6: 253–265

O'Leary M (2010) Multimodel inference and geographic profiling. Crime Mapping 2(1)

Pimentel D, McNair S, Janecka J, Wightman J, Simmonds C, O'Connell C, Wong E, Russel L, Zern J, Aquino T et al (2001) Economic and environmental threats of alien plant, animal, and microbe invasions. Agr Ecosyst Environ 84:1–20
Raine NE, Rossmo DK, Le Comber SC (2009) Geographic profiling applied to testing models of bumble-bee foraging. J R Soc Interface 6:307–319
Ricciardi A (2007) Are modern biological invasions an unprecedented form of global change? Conserv Biol 21:329–336
Rossmo DK (1999) Geographic profiling, 1st edn. CRC Press, Boca Raton
Rossmo DK, Harries K (2011) The geospatial structure of terrorist cells. Justice Quart 28:221–248
Sapp A, Huff T, Gary G, Icove D, Horbert O (1994) A motive-based offender analysis of serial arsonists. NCAVC, Quantico
Snow J, Frost W (1936) Snow on cholera: being a reprint of two papers. Commonwealth, New York
Stevenson MD, Rossmo DK, Knell RJ, Le Comber SC (2012) Geographic profiling as a novel spatial tool for targeting the control of invasive species. Ecography 35:704–715
Strayer DL, Eviner VT, Jeschke JM, Pace ML (2006) Understanding the long-term effects of species invasions. Trends Ecol Evol 21:645–651
Townes H (1962) Design for a malaise trap. Proc Entomol Soc Washington 64(4):253–262
Viswanathan G, Bartumeus F (2002) Lévy flights random searches in biological phenomena. Phys Stat Mech Appl 314:208–213
Viswanathan G, Afanasyev V, Buldyrev S, Havlin S, da Luz M, Raposo E, Stanley H (2000) Levy flights in random searches. Phys Stat Mech Appl 282:1–12
Vitousek PM, D'Antonio CM, Loope LL, & Westbrooks R. (1996) Biological invasions as global environmental change. American Scientist 84(5):468–478
Walker K, Lynch M (2007) Contributions of Anopheles larval control to malaria suppression in tropical Africa: review of achievements and potential. Med Vet Entomol 21(1):2–21
Wilcove DS, Rothstein D, Dubow J, Phillips A, Losos E (1998) Quantifying threats to imperiled species in the United States. Bioscience 48: 607–615
Yohannes M, Haile M, Ghebreyesus T, Witten K, Getachew A, Byass P, Lindsay S (2005) Can source reduction of mosquito larval habitat reduce malaria transmission in Tigray, Ethiopia? Trop Med Int Health 10:1274–1285

Biometrics

▶ Biometrics and Border Control Policing

Biometrics and Border Control Policing

Dean Wilson
Criminology and Criminal Justice, School of Law, Plymouth University, Plymouth, UK

Synonyms

Biometrics; Border control; Migration policing; Technology

Overview

This entry examines the role of biometric technologies and databases in border policing. It commences with a definition of biometric technology and then proceeds to discuss the political economy of the biometric industry since the attacks of 9/11. Drawing upon a range of critical scholarship, a number of theories are discussed which interrogate the changing nature of citizenship and identity in relation to biometric databases. The entry also discusses some of the critiques raised by civil libertarians in relation to the assemblage of large-scale and frequently transnational databases. The discussion concludes by considering the constitutive role of biometric technologies in broader processes of the securitization of mobility.

Key Issues/Controversies

Over the past decade borders and border enforcement have become the subject of intense scholarly interest within criminology, and more generally within the social sciences. In 1996 Saskia Sassen presciently identified border crossing as a "strategic site of inquiry about the limits of the new order" (xvi) under conditions of globalization. The drive – most readily apparent in states of the global north – to secure seamless circuits of capital, trade, and skilled migration, while fortifying national boundaries against perceived threats of organized crime, terrorism, and illegal migration, has resulted in considerable political, technical, and financial investment in border control (Aas 2011). The "informatization" of the border has been a crucial component of such developments – with a proliferation of databases, records, and files on citizens, migrants, business visitors, asylum seekers, and cross-border workers. Biometric identifiers are key elements of these proliferating databases which seek to control flows, identify risks, and banish threats (van der Ploeg 2006). This entry will detail key areas in the study of biometrics and border policing. While there is a technical literature in this field, primarily concerned with implementation (e.g., Woodward et al. 2003), this entry will focus on key issues and controversies that arise from critical scholarship. It will commence with a definition of biometrics, and then discuss the political economy of biometrics in relation to the security industry. The entry will then proceed to position the study of biometrics and border policing within wider debates about the changing nature of borders and mobility under conditions of globalization.

At the outset it is important to provide a brief explanation of what is under discussion in the scholarship surrounding "biometrics." In general, biometric technology involves the collection of digital representations of physiological features unique to an individual, such as fingerprints, pattern of the iris, the retina, or voice patterns. Digital representations are then normally transformed into binary code by means of an algorithm to construct a template. These templates are accumulated in a centralized database that is accessed when, on subsequent occasions, the finger, hand, face, eye, or voice is presented to the system. If a matching template is found, the person is acknowledged and counts as known to the system. In some instances the representation may be stored on a chipcard rather than a template. The user then has to present the chipcard and requested body part to prove they are the legitimate user of the card (van der Ploeg 1999, p. 37).

Biometric technology uses two main methods for identity checks: verification and identification. Verification is essentially confirming people are who they say they are, and is achieved through one-to-one matching. One-to-one matching is used, for example, in the identification of people traveling on false passports, but might also be used to facilitate more rapid immigration procedures for those possessing the designated biometric documents. Biometric technologies are also increasingly utilized in identification (that is determining who the person is, or in some instances is not) through one-to-many matching. The clearest example of one-to-many matching is evident in the use of facial recognition technology matched to CCTV cameras. Air passengers can have their faces scanned and checked against a database of "terrorists" or other "wanted" people (Lyon 2003).

In the wake of September 11, 2001 (hereafter9/11), and the subsequent "war on terror" hi-tech solutions to problems of security — such as facial recognition technology, iris and retina scans, and smart identity cards — have proven incredibly persuasive and ascendant (Van der Ploeg 2003; Zuriek and Hindle 2004). The proliferation of biometric technologies has been most evident at the borders between nation states, where the use of biometric passports matched with surveillance tools such as face and iris recognition technologies is being rapidly deployed at airports and other land and sea border posts. It is important to note that the security industry was already expanding prior to the attacks of 9/11. As Garland (2001) noted, late modern societies had witnessed an expansion of the commercial security sector that frequently fueled public anxieties in relation to crime control while claiming to allay them. Nevertheless, the already expansionist trajectory of the security sector received a massive boost from the climate of heightened anxiety following 9/11, and later compounded by the 7/7 bombings on the London Underground and the Madrid bombings. Mike Davis predicted in 2001 that "the 'Fear Economy,' as the business press has labelled the complex of military and security firms rushing to exploit the national nervous breakdown, will grow fat amidst the general famine" (p. 45). Later commentators have suggested Davis' prediction materialized. Webb (2007) has argued for the emergence of the "corporate-security complex" post 9/11, while Hayes (2010) outlines the emergence of a "security-industrial complex."

Prior to 9/11 biometrics was a fringe element of the security industry widely regarded as being of limited application and subject to significant errors. However, post-9/11 the biometric industry has grown considerably in size (Zuriek and Hindle 2004; Wilson 2006; Webb 2007). Moreover, particularly in the US, government expenditure on security technologies accelerated in the wake of the attacks of 9/11 and the subsequent "war on terror." In 2003 the Bush Administration earmarked $38 billion in new spending for homeland security, while security budgets in Europe and Australia also ballooned (Zuriek and Hindle 2004). In this climate the biometrics industry has been well organized to claim market share and gain "brand recognition" for biometrics as a security technology. Security technology companies were aggressively marketing their wares to airport officials within days of the 9/11 attacks (ACLU 2004, p. 27). Within a fortnight of 9/11 the International Biometrics Industry Association (IBIA), an advocacy organization representing major biometric companies in the US, issued a press release highlighting the role of biometrics in the fight against terrorism. Visionics – a US manufacturer of facial recognition technology – suggested in a white paper entitled *Protecting Civilization from the Faces of Terror* that airport security, in the US the responsibility of the Federal Government, "demands substantial financial resources" to develop "technology that can be implemented to immediately spot terrorists and prevent their actions." Boarding a plane, the white paper suggested, should no longer be considered "a right granted to all, but as a privilege accorded to those who can be cleared of as having no terrorist or dangerous affiliations" (cited in Zuriek and Hindle, p. 122). The political economy of biometric technology is constituted of complex networks of research laboratories, venture capitalists, and transnational corporations both stimulating and responding to expanding security markets that are both public and private (Walters 2010).

Biometric technologies have been heavily promoted as infallible despite their actual efficacy remaining highly questionable. At Palm Beach International Airport in the US, a trial of facial recognition technology failed to recognize participating airport employees 53 % of the time (Wilson 2006). Results of independent evaluations of facial recognition technology in other locations suggest similarly dismal results (Clarke 2003). Accuracy rates for facial recognition and fingerprint scanning remain problematic, and a Japanese researcher was able to fool a system by using a gelatin-based fake finger (Lyon 2003, p. 71; Zuriek and Hindle 2004, p. 120). Even some within the security industry are uneasy about the rapid pace with which untested and poorly researched biometric technologies are being deployed (Wilson 2006, 2007). Nevertheless the appeal of biometric technologies may reside in their symbolic resonance rather than in any technical efficacy, a point to which this entry will later return.

Automated systems, it might be suggested, reduce questions of discretion and discrimination to the abstracted question of whether one is granted or denied access. Technological discourse frequently constructs biometrics as a neutral security tool enacting abstracted processes of authentication and verification. As Muller notes "why biometrics is so successful in concealing its exclusionary and discriminatory character is the way in which it tries to hide the question of identity behind the preoccupation with authenticity" (2004, p. 286). Nevertheless, while biometric technology may appear to do little more than enact automated functions of verification and authentication, critical scholars have argued that it heralds new forms of control and exclusion. Gilles Deleuze (1992), for example, has argued that contemporary societies are "societies of control" where surveillance is detached from discipline and is primarily concerned with the distribution of entitlements on the basis of identity. Nikolas Rose (1999) has argued that contemporary shifts represent the "securitization of identity" whereby the exercise of freedom requires proof of legitimate identity.

The concepts of "societies of control" and the "securitization of identity" suggest that biometric technologies are far from neutral but are rather deeply imbricated within new modalities of power. Digital biometric identifiers become vital tokens of contemporary citizenship, and an integral aspect of what Jones (2001) terms "digital rule." "Digital rule" refers to the capacity of control agents to assign access and authorization. Punishment increasingly takes the form of "withdrawal of access privileges." This is a form of remote rule based upon databases and codes. The capability of remotely granting or denying access through biometric technologies linked to databases thus contains the capacity to deepen and widen social discrimination. Biometric technologies are hastening processes of what Lyon (2003) has termed "social sorting" where populations are digitally categorized as worthy and unworthy, included and excluded.

Similarly, scholars such as Graham and Wood (2003) have argued that these technological systems are far from neutral and abstracted, but are rather actively formed through social practices and decisions. As they note of digital surveillance systems, while they may be characterized by flexibility and ambivalence, and contingent upon judgments of social and economic worth built into their design, they are "likely to be strongly biased by the political, economic and social conditions that shape the principles embedded in their design and implementation" (2003, p. 229). Pugliese (2010) also argues that the categorizations of biometric systems are inherently racialized. Biometric systems, Pugliese suggests, are "calibrated to white," whereby Whiteness is configured as the universal gauge determining the technical limits for image capture. Categories of racialized exclusion are therefore embedded within the technological infrastructure. Several scholars have examined biometrics through the prism of Foucaldian notions of biopower (Pugliese 2010). Alternatively, Didier Bigo (2006), drawing upon Agamben's notion of the "ban" has suggested the concept of the "banopticon." David Lyon neatly summarizes this concept noting that "the majority of mobile citizens of the global north are normalized travelers who accept the need security passes for fast-tracking through the airport as they accept the need for frequent flyer cards to get access to the club lounges, a focused surveillance is reserved for the *sans-papiers*, the potential terrorist, the refugee – those 'trapped in the imperative of mobility'" (2008, p. 44).

Biometric technologies are thus not neutral but both constitutive and symbolic of transformations in power and social classification and exclusion. By their very nature, biometrics form components of technological systems that imply the assembling of digitized databases on an unprecedented scale. Scholars such as Lyon (2008) remain concerned that such databases are being assembled amidst a "state of exception," whereby nation-states increasingly enact decisions based upon political will rather than the constraints of normative law (Agamben 2005). Thus it could be argued that biometrics are deployed within a social context and organizing logic that inevitably heads towards exclusionary rather than inclusionary functioning. The potential for "actuarial justice" where the objective is risk management rather than rehabilitation and where predictors of dangerousness are engaged to preemptively target "problem" populations becomes amplified.

While a substantial scholarship has emerged around the issue of borders, the principal focus here is on how biometric systems connected to databases have facilitated the "delocalized border" (Bigo 2002). Aas (2005, p. 207) has observed that new technologies "seem both to transform the traditional space of government and to disrupt territorial boundaries," as well as appearing to provide the most efficient solution to problems of risk and security. The borders of the global north are fast becoming high-tech borders, "capable of materializing in different sites around the globe" (Walters 2006, p. 197). Actuarial logic requires that containment measures be applied *before* any harm results (Wilson and Weber 2008). Risk thinking in relation to border control therefore leads inexorably to temporal, and hence spatial, displacement of the border. Thus the delocalized, technologically realized border is "performed" rather than drawn as a line on a map, is functional rather than physical (Weber 2006), and significantly reorganizes the spatiality of power (Walters 2006).

The logic of exclusion inherent in border - control, coupled with mentalities of "risk reduction" and corresponding information technologies, produces strategies of "preemptive immobilization" aimed at certain categories of travelers who conform to set criteria. These risk profiles are constructed using aggregate information about suspect populations, and applied through information technologies accessed by immigration authorities and their agents on or before the border. Surveillance technologies therefore expand the potential codes and categories of risk (O'Malley 2006). Nevertheless, border control is underpinned by more than a simple exclusionary logic. States have been seeking migration flows that are simultaneously accelerated for some and fortified against others, most often expressed as a binary opposition between "facilitation" and "security." Scholars have variously described this bifurcation of mobility – broadly equating to the increasing economic chasm between the global south and the global north – as the widening divide between "global tourists" and "global vagabonds," the sorting of mobile bodies into "kinetic elites" and "kinetic underclasses," the "mobility rich" and the "mobility poor," and "two-speed citizenship" (Bauman 1998; Adey 2004; Aas 2011).

For nation-states the engagement of biometric technology at physical borders bears a signification that neatly encapsulates the paradoxes of state sovereignty under conditions of globalization. Biometric technology offers the possibility of rapid flows of capital and global elites through borders while simultaneously fortifying the border against unwanted intrusions from "deviant" outsiders. The technology becomes instrumental in processes of global social classification in which "low risk" travelers are granted unimpeded access and mobility while "high risk" travelers from the Global South, compelled to travel with limited or no documentation, are excluded and blocked. Registered traveler programs highlight this two-tiered system of mobility emerging in the post 9/11 context. Privileged passenger programs facilitate the rapid and unimpeded movement of elite travelers from the global north. Simultaneously those from the global south are increasingly blocked from entry. In a global economy where mobility is equated to participation the consequences of this discrimination are enormously significant. In the US as such these systems of preferential mobility are encouraged in law. The *Aviation and Transportation Security Act 2002* provides that the Under Secretary of Transportation for Security may establish requirements to implement trusted traveler programs. Registered Traveler Programs are intended not only to provide selected travelers with expedited processing at border points, but also to "target security resources to those travelers who might pose greater security risks."

Such systems have however been in place for some time. The US Immigration and Naturalization Service Passenger Accelerated Service System (INSPASS) a hand geometry system used in seven US and two Canadian airports has been in operation since 1993 (GAO 2003). Amsterdam's Schiphol Airport was the world's first airport to employ an automatic border control system using iris recognition technology for travelers. The "Privium" system, installed in October 2001, is intended to fast track passengers carrying an iris data-embedded smart card through passport control (Woodward et al. 2003, pp. 295–296; Lyon 2003, p. 71). Similar projects have also been undertaken in other locations. At Tokyo's Narita airport frequent customers of Japan Airlines are enrolled in a comparable system, has have been selected passengers at London's Heathrow airport and at Toronto and Vancouver (Wilson 2006). Biometric border control systems thus serve to exacerbate the trend identified by Bauman as "the extraterritoriality of the new global elite and the forced territoriality of the rest" (2000, p. 221). Biometric technology is therefore both a powerful signifier and a formative element in the widening chasm between a small global elite possessed of mobility and capital and the many in the global south geographically confined and impoverished.

It has been argued that states with diminished control of economic and social questions within their national boundaries increasingly turn to border security and migration control as key "performances" asserting their continued relevance and strength (Wilson and Weber 2008; Wilson 2006). Political scientist Wendy Brown suggests that border construction projects are "hyperbolic tokens" that attempt to signify resurgent nation-state sovereignty, despite the reality that they are often expressions of the withering of the nation-state and "the waning relevance and cohesiveness of the form" (Brown 2010, p. 24). For the truncated nation-states of late modernity, biometric technologies are potent signifiers of a reinvigorated sovereign power with the capacity to assert impermeable borders. Iris and fingerprint scanners, passport kiosks, and smart cards are, from this perspective, integral props in the theater - of statehood.

Such a theater of statehood is nowhere more in evidence than in the increased securitization of airports. The 9/11 attacks, and earlier instances such as the TWA flight crash in 1996, have been instrumental in facilitating heightened security measures in airports, and in embedding airports in public consciousness as filters of risk (Adey 2004). As border posts airports thus have substantial symbolic capital. As a report on Australian airport security noted "perhaps as important as their contributions to transportation and the wider economy is the symbolic significance attached to airports. They embody the modern world in all its complexity, since few other places bring together our most advanced technological creations and the intricate interconnected systems we have devised to serve both those creations and ourselves" (Wheeler 2005, p. 5). As highly symbolic nodes of modernity, airports become crucial platforms through which the power of the state to include and exclude is enacted. The centrality of securitizing airports in the post 9/11 environment then has powerful signifying functions. Mike Davis suggests that airports may function as microcosmic spaces of wider security and surveillance trends, noting that "the security regime of the airport departure lounges will likely provide a template for the regulation of crowds at malls, shopping concourses, sports events and elsewhere" (2001, p. 45). Biometric identification systems are integral elements of the intensified security regimes evident in airports (Zuriek and Hindle 2004, p. 129).

The signifying and political functions of biometric technology are also evident in its incorporation within the international passport system. Passports are the primary document for identifying, regulating, and tracing mobile individuals. The passport is, as Salter suggests, "a modern heuristic device which serves to link individuals to foreign policy, and according to which government agents classify travellers as safe or dangerous, desirable or undesirable, according to national, social or political narratives" (2004, p. 72). The spread of biometric passports has largely been dictated by US legislation, and offers a tangible expression of US hegemonic power in the post 9/11 global order. Section 302 of the *Enhanced Border Security and Visa Entry Reform Act 2001* requires "a visa waiver country, in order to maintain program participation, to certify by October 26, 2003, that it has a program to issue to its nationals qualifying machine-readable passports that are tamper proof and contain biometric identifiers."

The hegemonic power of the US is therefore instrumental in leading and setting the final standards for the global introduction of biometric passports (Wilson 2006). The International Civil Aviation Organization (ICAO), the organization governing international civil aviation, has been examining the issue of biometric passports since 1995. However, until recently national and regional legislation protecting privacy and civil liberties mitigated against their adoption. In May 2003, the G8 countries (Canada, UK, France, Japan, Italy, Russia, Germany, and the US) entered into an agreement to implement a biometric passport system. Civil libertarians remain troubled that passports will utilize face recognition, not only among the least reliable of biometrics but also one that can be used at a distance to track individuals without their knowledge. Moreover the passports are to include RFID (Radio Frequency Identification) tags. These are tiny computer chips that, when receiving a radio signal from an RFID reader, use that power to transmit the data they store. And in US passports this information would include name, date of birth, place of birth, and a digital photograph and digital face recognition template. Organizations such as the American Civil Liberties Union maintain that this raises substantial privacy and security issues and significantly intensifies the potential for the surveillance and tracking of individuals.

Critics maintain that biometric passports are a proxy for a global database. The International Campaign Against Mass Surveillance, for example, argues that biometric passports represent one avenue of obtaining "nearly universal registration of everyone on the planet" (cited in Wilson 2006). For critics the potentially ominous ramifications of this development reside not in the documents themselves but in the possible databases linked to them and the tracking and classification potential they unleash. The assemblage and mobilization of large-scale government databases in relation to visa applications and the tracking of "non-citizens" was evident in the detention of hundreds of Muslim non-citizens in the wake of September 11. US authorities systematically registered and created dossiers on nearly every male over the age of 15 with origins in a list of countries (mostly Muslim) traveling in the US. Individuals on the list were required to report to the government to be fingerprinted, photographed, and questioned. Those who failed to register risked deportation and criminal penalties. This was carried out under the aegis of a program called the National Security Entry-Exit Registration System (NSEERS) (Wilson 2006).

The registration that occurred under NSEERS has been expanded via a program called US-VIST to most visitors to the US. A similar system in the EU, the Visa Information System (VIS) has been developed in the EU to capture and store information – including biometric information – from all visa applications to EU member states (Wilson 2006). The storage of this data and its linkage across databases will facilitate surveillance and monitoring on an unprecedented scale. The integration of biometric technologies into the international passport and visa system is then both a signification and realization of new forms of power. The technologizing of the border via biometrics betokens a renewed capacity of nation-states to police their boundaries. Technologized means of exerting control over movement form elements of "processes of statecraft" whereby the sovereignty of the state is enacted through processes of inclusion and expulsion. Such expressions of sovereignty have significant repercussions on the ground, as expanding databases facilitate new and powerful forms of cataloging, coding and filing individual identities. Increasingly too, databases are globalized through the incorporation of biometric identifiers in the international passport and visa system, an architecture forged by the US that some have suggested opens up the possibility of a global ID.

Post 9/11 questions of migration and mobility are increasingly interpreted through the prism of security (Guild 2009). States of the global north have assessed the specter of uncontrolled migration as a threat to sovereignty. Within this rhetoric asylum seekers and refugees are increasingly cast as "folk devils" in moral panics fuelled by law and order politics (Wilson 2006). The demonization and criminalization of asylum seekers has been accompanied by coercive measures such as indefinite detention and forced deportation. In the nation states of the global north such coercive measures are elements of more sustained processes of criminalization whereby the act of seeking asylum is itself constructed as a "crime of arrival."

Consequently it is perhaps unremarkable that biometric identification systems have initially been deployed on asylum seeker populations, where they serve as technological signifiers of the securitization of migration. Biometric identification systems involving fingerprinting those seeking asylum have become increasingly common and are in use in the UK, the Netherlands, and Australia (Woodward et al. 2003; Wilson 2006). The United Kingdom's Home Office has adopted a biometric solution in the identification of asylum seekers within the UK. Asylum seekers are issued an Application Registration Card that carries a template of the bearer's fingerprints, as well as their photograph, name, date of birth, and nationality (Home Office UK, 2002). The ARC is designed to facilitate rapid identification of asylum seekers following initial processing at ports of entry or at the Asylum Seekers Unit in Croydon. The objective of the card is to facilitate identification of asylum seekers by authorities. In so doing the ARC firmly inscribes and secures a "non-citizen" identity on individual asylum seekers. The system is designed not so much to facilitate access to services as to deny them, one stated purpose of the system being to counter fraud through the claiming of services under multiple identities (Woodward et al. 2003, p. 290.)

Similar schemes are also initiated in the EU, US, and Canada. The countries of the European Union have established EURODAC, an automated fingerprint identification system. In the Convention of Dublin, June 15, 1990, the member states of the European Union agreed that the first country where the applicant arrives was to be responsible for their application. This was to preclude one country passing asylum seekers on to another, as well as asylum seekers submitting multiple applications in different member states (Van der Ploeg 1999, p. 298). EU ministers responsible for immigration established a Community wide system for the comparison of fingerprints of asylum applicants in 1991, while the agreement to build a central fingerprint database of asylum seekers was established in 1997 by the European Council. EURODAC was tested in 2002, and from January 15, 2003, asylum seekers fingerprints have been taken whenever they seek asylum, whether inside or outside the EU's borders. Fingerprints are then digitally transferred from member states to a central EURODAC unit for comparison against the existing database. Fingerprints of asylum seekers can be stored for up to 10 years, or immediately erased if an applicant is granted nationality in a member state. In addition to asylum seekers, EURODAC is also a biometric database of "persons who have crossed an external frontier of the Community in an irregular manner." These "foreign nationals" have their biometric data retained for 2 years or immediately erased if they receive a residence permit or leave the territory of the EU (Wilson 2006).

EURODAC was justified by the EU Council in terms of the need to accelerate decisions on asylum bids. Another powerful rationale behind the establishment EURODAC was a perceived need to stem "asylum shopping" across member states of the EU and separate those categorized as "legitimate" asylum seekers from other categories of migrant. Across nation-states of the global north biometric technology is being extensively deployed to demarcate and signify identities of "non-citizens." This process is intertwined with the wider securitization of migration as a political issue and the attendant discursive criminalization of migrant identities. International relations scholars have noted the significance of policing borders not only to preserve the integrity of the nation-state but as a marker and arbiter of state sovereignty overall. Increasingly the identity of the asylum seeker has been configured as a site on which state sovereignty is enacted. The capacity to establish legitimacy, confirm identity, and expel those deemed "illegitimate" has thus been a pivotal "performance of sovereignty" (Pickering 2005).

The particular focus upon asylum seekers in biometric deployment is intertwined with the securitization of migration and the criminalization of persons seeking to move from the global south to the global north. The issuing of smart entitlement cards and the targeting of "illegal" migrants with biometric technologies serves to wend discourses of terrorism, organized crime, and migration as a single coherent threat to national sovereignty. The need for biometric data incorporated into travel documents is most often cited as "national security" – meaning security from "illegal" migrants and terrorists. However, in practice such documents inscribe on those seeking asylum the identity of "non-citizen." Fixing these "non-citizen" identities through biometric documentation and "entitlement" cards serves to heighten exclusion and reinforces the notion of migration as a problem of security.

Biometrics has become a pivotal technology in a range of security domains but most evidently in border control. Its ascendancy has been partially due to the expanding security market – both public and private – in the post 9/11 global north. Critical scholars, however, dispute the claims of the technical literature which frequently asserts the neutrality of the technology. It is argued that biometric databases are integral to processes of social sorting that distinguish between the eligible and ineligible, the included and the excluded, the banished and the admitted. Moreover biometric databases have deterritorialized the border, creating a situation in which the border is – quite literally – everywhere. Biometric technologies have also played an instrumental role in facilitating the enhanced mobility of a privileged elite from the global north through trusted traveler programs while impeding the movement of those from the global south through risk profiles and categorizations of dangerousness.

Related Entries

▶ Forced Migration and Human Rights
▶ Surveillance Technology and Policing

Recommended Reading and References

Aas KF (2005) “Getting ahead of the game”: border technologies and the changing space of governance. In: Zureik E, Salter M (eds) Global surveillance and policing: borders, security, identity. Willan, Cullompton, pp 194–214

Aas KF (2011) “Crimmigrant” bodies and bona fide travelers: surveillance, citizenship and global governance. Theor Criminol 15(3):331–346

Adey P (2004) Secured and sorted mobilities: examples from the airport. Surveill Soc 1(4):500–519

Agamben G (2005) State of exception. University of Chicago Press, Chicago

American Civil Liberties Union (ACLU) (2004) The surveillance-industrial complex: how the American government is conscripting businesses and individuals in the construction of a surveillance society. ACLU: New York. http://www.aclu.org

Bauman Z (1998) Globalization: the human consequences. Polity Press, Cambridge

Bauman Z (2000) Social issues of law and order. Brit J Criminol 40(2):205–221

Bigo D (2002) Security and immigration: toward a critique of the governmentality of unease. Alternatives 27:63–92

Bigo D (2006) Security, exception, ban and surveillance. In: Lyon D (ed) Theorizing surveillance: the panopticon and beyond. Willan, Cullompton

Brown W (2010) Walled states, waning sovereignty. Zone, New York

Clarke R (2003) SmartGate: a facial recognition trial at Sydney airport. http://www.anu.edu.au/people/Roger.Clarke/DV/SmartGate.html

Davis M (2001) The flames of New York. New Left Rev 12:34–50

Deleuze G (1992) Postscript on societies of control 59:3–7

Garland D (2001) The culture of control: crime and social order in contemporary society. Oxford University Press, Oxford

Gates K (2011) Our biometric future: facial recognition technology and the culture of surveillance. New York University Press, New York

General Accounting Office (US) (GAO) (2003) Border security: challenges in implementing border technology: statement of Nancy Kingsbury, managing director, applied research and methods, GAO-03-546T. Washington DC

Graham S, Wood D (2003) Digitizing surveillance: categorization, space, inequality. Crit Soc Pol 23(2):227–248

Guild E (2009) Security and migration in the 21st century. Polity Press, London

Hayes B (2010) “Full spectrum dominance” as European security policy: on the trail of the “NeoConOpticon”. In: Haggerty K, Samatas M (eds) Surveillance and democracy. Routledge, London

Jones R (2001) Digital rule: punishment, control and technology. Punishm Soc 2(1):5–22

Lyon D (2003) Surveillance after September 11. Polity Press, London

Lyon D (2008) Filtering flows, friends, and foes: global surveillance. In: Salter M (ed) Politics at the airport. University of Minnesota Press, Minneapolis

Muller B (2004) (Dis)Qualified bodies: securitization, citizenship and “identity management”. Citizsh Stud 8(3):279–294

Muller B (2010) Security, risk and the biometric state: governing borders and bodies. Routledge, London

O’Malley P (2006) Risk, ethics and airport security. Can J Criminol Crim Just 48(3):413–421

Pickering S (2005) Refugees and state crime. Federation Press, Sydney

Pugliese J (2010) Biometrics: bodies, technologies, biopolitics. Routledge, London

Rose N (1999) Powers of freedom: reframing political thought. Cambridge: Cambridge University Press

Salter M (2004) Passports, mobility, and security: how smart can the border be? Int Stud Perspect 5:71–91

Sassen S (1996) Sovereignty in an age of globalization. Columbia University Press, New York

Van der Ploeg I (1999) Written on the body: biometrics and identity. Comput Soc 29(1):37–44

Van der Ploeg I (2003) ‘Biometrics and privacy: a note on the politics of theorizing technology’ information. Communication and Society 6(1):85–104

Van der Ploeg I (2006) Borderline identities: the enrollment of bodies in the technological reconstruction of borders. In: Monahan T (ed) Surveillance and security: technological politics and power in everyday life. Routledge, New York

Walters W (2006) Border/control. Eur J Soc Theor 9(2):187–203

Walters W (2010) Migration and security. In: Peter Burgess J (ed) The Routledge handbook of new security studies. Routledge, London

Webb M (2007) Illusions of security: global surveillance and democracy in the post-9/11 world. City Lights, San Francisco

Weber L (2006) The shifting frontiers of migration control. In: Pickering S, Weber L (eds) Borders, mobilities and technologies of control. Dordrecht, Springer

Wilson D (2006) Biometrics, borders and the ideal suspect. In: Pickering S, Weber L (eds) Borders, mobilities and technologies of control. Springer, Dordrecht

Wilson D (2007) Australian biometrics and global surveillance. Int Crim Just Rev 17(3):207–219

Wilson D, Weber L (2008) Surveillance, risk and preemption on the Australian border. Surveill Soc 5(2):124–141

Wheeler J (2005) An independent review of airport security and policing for the government of Australia, Commonwealth of Australia, Canberra

Woodward J, Orlans N, Higgins P (2003) Biometrics: identity assurance in the information age. McGraw Hill, Berkeley

Zuriek E, Hindle K (2004) Governance, security and technology: the case of biometrics. Stud Polit Econ 73:113–137

Boot Camps

▶ History of Boot Camps

Border Control

▶ Biometrics and Border Control Policing

Brain

▶ Neurology and Neurochemistry of Crime

Breakdown

▶ Social Disorganization and Terrorism

Break-In

▶ Residential Burglary

Breaking and Entering

▶ Residential Burglary

Bribery

▶ Corruption

British Crime Firms

Dick Hobbs
University of Essex, Essex, UK

Overview

Historically organized crime is a relatively new phenomenon in Britain that was only formally recognized by politicians and law enforcement in the 1980s and 1990s and emerged as a social problem that was heavily associated with transnationality (Hobbs 2013). However, there were precursors to this obsession with “transnational-organized crime,” which warrant attention, and they were very much home grown and embedded in the class relations and political economy of industrial society. The criminal firm emerged from the cultures of the urban working class as a form of localized organized crime that was central to the economies and cultures of traditional working class communities and were particularly prominent in London among those communities lacking the strict disciplines of industrial life.

History

Working class London was made up of, “a social patchwork of intensively localist culture and sentiment” (Robson 1997, 7), that lacked the homogeneity of cultures created by a single mode of production (Hobbs 1988, 87). London’s economic base was diffuse, and as a consequence the capital’s proletariat evolved as a complex and highly individualistic amalgam of class specific identities gleaned from localities sharing little in common with their neighbors other than socioeconomic deprivation. Even dock work, upon which vast swathes of south and east London were reliant, created working cultures dependant upon semiautonomous groups of workers hired by the day before shifting to another cargo, possibly on another ship, maybe in a different dock (Hobbs 1988, 109–111). The localism of Londoners manifested itself in a tendency to regard territories beyond their own with deep suspicion, maintaining exposure to those residing in adjacent neighborhoods to a minimum.

The strictly defined territorial parameters of industrial working class communities produced clusters of youthful collaborations who utilized territorial identity as a vehicle for recreational transgression before morphing seamlessly into adult pursuits, and if we take some of the better known criminal brands of the mid-twentieth century, youthful conflict was certainly important in forming identities, creating reputations, and molding communal distinctiveness in the face of opposition. These territorially based groups occupied spaces or interstices between sections of the urban fabric and were frequently situated within "delinquency areas" where youthful deviant collaborations were organized around working class territorial imperatives defined by the experiences of low income, poor housing, and poor health. It is here where criminal careers are nurtured and where youth formulate criminal associations. Then as now neighborhood-based groups abounded in working class London, territorial imperatives featuring individual, local, and family reputations were forged in youthful combat that was ideal preparation for the world of acquisitive gangsterism, where fractious networks came to define much of what came to be regarded as organized crime in the twentieth-century London.

Throughout the industrial era generations of youths forged reputations in defense of their neighborhood, as groups of young men drifted into adjacent territory in order to provoke violence. However, these groups emanated from a domain that was shaped, refined, and restrained by elements of social control that were lodged in the pragmatic realities of the local political economy and its associated cultures and institutions. For most participants these were temporary collaborations before engagement with the local employment market-enabled maturation into an unambiguously proletarian society where crime was normal (Hobbs 2013). However, certain individuals, groups, and families on establishing a monopoly of informal violence would commodify this monopoly and via the finely grained differential between security and extortion, exploit businesses, and individuals within their home territory. While most groups stayed within the cultural and economic frameworks provided by the industrial neighborhood, the most competent and ambitious sought out and colonized adjacent territories where the object was to create "protected enclaves" (Arlacchi 1986, 195), via the maintenance of a viable threat.

Examples of crime firms who succeeded in breaking out of the local neighborhood to occupy adjacent territories can be found in the three following cases.

Three Crime Firms

After the First World War, there was a boom in horse racing and many who wished to gamble were drawn to the opportunities for lawful betting at racecourses. However, groups of men with violent reputations exploited the numerous extortionate opportunities available at the racecourses, and these coalitions of violent men are significant for their willingness to pit local reputations within a national arena. The Sabini family emerged as defenders of Clerkenwell's "Little Italy," resisting incursions from equally deprived areas of the capital, and they soon commodified their growing reputations by organizing a moderately lucrative extortion business. However, the racecourses offered the richest prize to any group who could transcend the restricted opportunities of the working class territory where their reputations were formed and to enable this transcendence the Sabinis formed alliances with Jewish bookmakers and criminals, as well as some significant figures in the police (Hill 1955, 3).

Eventually, against a backdrop of considerable violence, the Sabinis established something approaching a monopoly at most of the major racecourses in England, progressing into the affluent West End of London, and "Bottle parties, clubs, public houses, cafes, restaurants, even ordinary shops had to pay protection money to the Sabini extortionists" (Hill 1955, 5). However, further progress was halted by the outbreak of the Second World War and the interment of many of the "Italian Firm" and the Sabinis never fully recovered their power. The Sabinis were a powerful criminal group who transcended their immediate neighborhood, briefly succeeded in dominating a wide territory that included most major English racecourses, as well as the affluent west end of London, and as a consequence Darby Sabini was "the nearest Britain got to an organising gangster" (Pearson 1973, 30. For further details of the Sabinis, see Hobbs 2013).

The Kray twins who emerged from the East London neighborhood of Bethnal Green during the post-Second World War era also utilized violent reputations forged in adolescence before commodifying this violence and extorting from local businesses. Changes in the British gambling laws opened up new opportunities in adjacent territories, and they were soon competing with neighborhood crime firms from across London for a portion of the lucrative West end market (Pearson 1973, 16–18, 27–28). Mental illness plagued the Kray firm, and delusional fantasies based upon the replication of Gangster scenarios with origins in Hollywood restricted their progress. Eventually reckless violence and murder made the Krays firm difficult to ignore, and by the end of the 1960s the police had imprisoned the entire firm, the twins making more money from nostalgia, bogus charity events, books, and a film than they ever did from crime (Hobbs 2013).

Charlie and Eddie Richardson were born into a traditional south London working class family, and like their East London counterparts the brothers endured a wartime childhood, enjoyed the vibrant streetlife of working class London and developed as talented young boxers and prominent street fighters. However, unlike the Krays, both Charlie and Eddie Richardson were hard workers and made good money from the postwar scrap metal trade. The Richardsons established an extortion business across a wide south London territory and moved into a number of disparate commercial areas, including wholesale chemists and mineral mining. The honey-pot of the West End brought them into limited conflict with the Krays and the Richardons diversified into long firm fraud (Levi 1981, 76–77), gaming machines, pornography, scrap metal yards, a perlite mine in South Africa, and control over car parking at Heathrow Airport. The Richardson firm was eventually closed down by the police after a fatal shooting at a club, a bizarre "torture" trial and Charlie Richardson's dalliance with the South African Secret service (Hobbs 2013).

These three examples indicate the manner in which the British crime firm emerged from working class neighborhoods with commercially viable violent reputations forged in youth. The foundation of the crime firm is extortion or "protection," and the elite were able to apply their reputations to colonizing adjacent territories. However, the local working class neighborhood was the platform that formed the reputations upon which subsequent operations depended, and when this territory was obliterated, the neighborhood firm was forced to adapt.

Deindustrialization eroded the traditional working class milieu from which the traditional crime firm emerged, and the fragmentation of working class neighborhoods created arenas that have made it difficult for family-based crime firms to establish the kind of parochial dominance they once enjoyed in the 1950s and 1960s. The new arenas for criminal activity are entrepreneurially orientated and consist of "…flexible, adaptive networks that readily expand and contract to deal with the uncertainties of the criminal enterprise" (Potter 1994, 12), and as a consequence, the family firm has had to adapt (Hobbs 1997). Although we should not assume that this constitutes redundancy, Lambrianou reminds us, "Brothers were your strength" (1992, 29).

Suburbs

The postindustrial drift of London's traditional working class communities away from its urban heartlands was enabled by both public housing policy and the private housing market and resulted in the colonization of rural space, bloating the periphery with new towns and extending the suburbs with a rash of housing projects. Those involved in criminal enterprise were merely part of this migration, and the key to understanding their change of address rests not with their criminality, but with their status as part of a proletarian diaspora whose trajectories out of the working class city can encompass a wide range of possibilities. Yet the shift out of the old neighborhood is seldom total, for the restructuring of the old industrial neighborhood has created new opportunities for money making, and pubs, clubs, and gyms make good investments that will always benefit from being associated with an old established brand (Hobbs 2013).

As geographical location becomes a somewhat ambiguous concept around which to base an illegal business, criminal labor has taken on increasingly transient, characteristics. The ever-mutating interlocking networks of opportunity created by the late modern city (Ruggiero and South 1997), in common with its legitimate counterpart, features interactions negotiated within networks of small flexible firms that are characterized by short-term contracts and lack of tenure, where the workforce can be "...hired/fired/offshored, depending upon market demand and labor costs" (Castells 1996, 272).

Understanding local housing policies is crucial to reaching an understanding of the distribution of various forms of delinquency, and the emergence of potent postindustrial forces that have "emptied out" marketable labor from communities formed around the assumptions of traditional cultures has relocated criminal coalitions within decentered, unpredictable, and relatively fragile economies. As a result of the sanctioning of the sale of two million homes in the UK, working class council tenants were able to buy their homes at a massive discount before selling at a profit and heading for London's periphery, which combined with deindustrialization, effectively drained working class London of much of its established human capital (Hobbs 2013). Municipal housing policy, along with the London's private housing market combined with deindustrialization to relocate criminal coalitions within decentered, unpredictable, relatively fragile economies, spread over a wide, and at times seemingly inappropriate, terrain. By the late 1970s the traditional neighborhood had been decimated by regeneration, and most of the traditional family firms had become at least partly ensconced within London's periphery, establishing new business ventures, and in particular engaging with the drugs trade (Hobbs 2013).

However, deindustrialization is not a homogenous process, and in an examination of the serious crime community in another part of Britain, housing policy deliberately retained traditional neighborhood structures, and as a consequence key aspects of the locales cultural inheritance such as the neighborhood family firm, along with the cultural territory that spawned them, remained relatively intact. These firms have of course adapted to the new market place, occasionally expanding beyond the old neighborhood, while striving to retain the traditional order through traditional means (Hobbs 2001).

Discussion

Within the new serious crime community, relations between individuals vary according to demographic dispersion, familial composition, ethnic distribution and integration, commercial practice, trading routes and patterns, the economic backcloth of the legitimate culture, and the particular use of space (Soja 1989). As a consequence serious crime networks have been extended across time and space, its members residing upon an anonymous terrain that is apparently indistinguishable from that of the legitimate economic sphere. Yet these networks stand in stark contrast to their predecessors, particularly with regard to their dynamism, and the extent to which they impact globally via their ability to both exploit and undercut traditional cultures. This ambiguity that is located within the new family firm is highly instrumental and typified by "vertically disintegrated networks of small firms engaged in transaction rich linkages of market exchanges" (Lash and Urry 1994, 23). The family firm is historically and cognitively positioned among a habitus incorporating multiple chains of legitimate and illegitimate opportunity networks to exploit these linkages to the full. The scope and extent of their activity depends upon the degree of connectivity established between groups and individuals, and it is this connectedness rather than corporate identity that forms the structural connotation inherent in "organized crime." The at times confusing list of the family firm's contemporary activities, especially compared to the one dimensional nature of their forefather's criminal clout, is indicative of the families' ability to establish interactions that consist of infinite mutations that exploit geographic space without being confined to territorial imperatives.

The same mechanisms that contrive in the destruction of traditional communities are also liable for the creation of new forms of locality and identity, and the neighborhood firm is essentially a constantly evolving social system (Ianni 1971, 35), whose strength and longevity is derived from its continued identification with territory that retains a deep and enduring well of historical sentiment. For even when the material base for such a community is removed, and crime is located within loose collectivities of ad hoc groupings, the family firm survives as a functional touchstones of competence and honor (Hobbs 1997).

The sense of "retrospective unity" (Bauman 1992, 138) that is exploited by the family firm is reliant upon a generalized recipe of locality (Robertson 1995, 26) that enables a highly durable system of transposable dispositions, from which emerge practices that reproduce "the objective structures of which they are the product" (Bourdieu 1977, 72). The flexibility of contemporary serious crime collaborations, licenses engagement with new markets, and particularly when they are at least part populated by powerful individuals and proven practices from previous eras, they become somewhat more difficult to locate and pigeonhole than their predecessors. For while they may be steeped in traditional criminal cultures established and maintained by the precedents of indigenous markets, the contemporary criminal firm is not restrained by the parameters of specific neighborhoods, for they inhabit a universe that prioritizes the ability to perform successfully within a local class milieu that has not disappeared, but realigned according to the dictates of global markets.

Criminal career trajectories are no longer restricted to highly specific and essentially redundant geographic and cultural spaces within an urban "underworld" but are located "…all over the main streets and back alleys of the global economy" (Castells 1996, 168). The result is the creation of criminal firms typified by flexibility and unpredictability, operating within multilayered, networks of opportunity, "small, fragmented, and ephemeral enterprises (which) tend to populate illegal markets-not large corporate syndicates" (Potter 1994, 13). The illegitimate economic sphere now features interlocking networks of small flexible firms, which like all aspects of contemporary capitalism, is typified by versatility, and a degree of flexibility that were inconceivable during the industrial era (Lash and Urry 1987). While this is contrary to the popular notion of transnational crime corporations (Williams 1993), as Robertson (1992, 1995) has indicated, globalization has improved the viability of locality as a generator of a discrete social order. Indeed the dialectic between the local and the global brings to the fore markets where alliances between global and local spaces created enacted environments where "the methodologies for the integration of organised criminals into civil society are established" (Block 1991, 15).

This brings the family firm back into the spotlight, for by perpetually realigning local precedents in the context of global markets, some kind of continuity and long-term viability can be maintained (Hobbs 1995, Ch. 5). Globalization should not be allowed to override the particular (Ferguson 1992), for the result, as we see from any critique of transnational organized crime (Hobbs 2013), is an a-historical theory built upon a rigidly structured view of economic life drained of cultural resonance. Indeed such a concept could not possibly reproduce faithfully the experiences of the family firm, whose enacted environment is a blend of the local and the global, constituting a highly effective vehicle that enable both individuals and groups to move between these two interlocking spheres. For criminal career trajectories are no longer restricted to highly specific and essentially redundant geographic and cultural spaces within an urban "underworld" but are located "…all over the main streets and back alleys of the global economy" (Castells 1996, 168).

Serious crime should be understood in terms of interlocking networks which are metaphors for relationally and manifested as amalgams of family, neighborhood, region, and nationality. Indeed, kinship in itself should be regarded as a highly instrumental trust variable, assuring loyalty by appealing to something other than self-interest (Lupsha 1986, 34). This is particularly relevant to the family firms continued use of violence, which remains a tactic of the increasingly blurred personal and commercial domain, and serves to stress the importance of trust for, “The everyday presence of violence prevents personal and effectively neutral trust relationships. Where violence is paramount, interpersonal ties must necessarily be strong, intense and affectively connoted” (Catanzaro 1994, 273). The continued use of violence to create, protect, and maintain a slice of the market within a densely populated sector inhabited by a multiplicity of networked interfaces (Arlacchi 1986, 195) is an enduring characteristic of the serious crime market place (Arlacchi 1998, 205).

Conclusion

Criminals do not experience the world globally or transnationally, for these are essentially fields devoid of interpersonal relations, and although it has undergone major configurations and relocations during post-industrialism, the firm is far from obsolete. The firm operates both in the “inherited basic architecture” (Castells 1996, 146) of the industrial neighborhood, in their new suburban idyll, and across regional, national, and international boundaries (Hobbs 2013), constituting the embodiment of Rose’s belief that the individual, “can within the limits permitted by the culture define for himself somewhat new patterns suggested by the variation among the old ones” (1962, 14).

Related Entries

- ▶ Careers in Organized Crime
- ▶ Organized Crime

Recommended Reading and References

Arlacchi P (1986) Mafia business: the mafia ethic and the spirit of capitalism. Verso, London

Arlacchi P (1998) Some observations on illegal markets. In: Ruggiero V, South N, Taylor I (eds) The new European criminology. Routledge, London, pp 203–215

Bauman Z (1992) Intimations of postmodernity. Routledge, London

Block A (1991) Masters of paradise. Transaction, New Brunswick

Bourdieu P (1977) Outline of a theory of practice. Cambridge University Press, Cambridge

Castells M (1996) The rise of the network society. Blackwell, Oxford

Catanzaro R (1994) Violent social regulation: organized crime in the Italian South. Soc Leg Stud 3(2):267–270

Ferguson M (1992) The mythology about globalisation. Eur J Commun 7(1):69–93

Hill B (1955) Boss of Britain’s underworld. Naldrett Press, London

Hobbs D (1988) Doing the business: entrepreneurship, detectives and the working class in the east end of London. Clarendon, Oxford

Hobbs D (1995) Bad business: professional criminals in modern Britain. Oxford University Press, Oxford

Hobbs D (1997) Professional crime: change continuity and the enduring myth of the underworld. Sociology 31(1):57–72

Hobbs D (2001) The firm: organizational logic and criminal culture on a shifting terrain. Br J Criminol 41(4):549–560

Hobbs D (2013) Lush life: constructing organised crime in the UK. Oxford University Press, Oxford

Ianni F (1971) Formal and social organisation in an organised crime “family”: a case study. Univ Florida Law Rev 24:31–41

Lambrianou T (1992) Inside the firm. Pan, London

Lash S, Urry J (1987) The end of organised capitalism. Polity, Cambridge

Lash S, Urry J (1994) Economies of signs and space. Sage, London

Levi M (1981) The phantom capitalists. Gower, Aldershot

Lupsha P (1986) Organised crime in the United States. In: Kelly R (ed) Organised crime: a global perspective. Rowman and Littlefield, Totowa

Pearson J (1973) The profession of violence. Granada, London

Potter GW (1994) Criminal organisations. Waveland Press, Illinois

Reuter P (1984) Disorganised crime: illegal markets and the mafia. MIT Press, Cambridge

Robertson R (1992) Globality and modernity. Theory Cult Soc 9(2):153–161

Robertson R (1995) Glocalization: time-space and homogeneity-heterogeneity. In: Featherstone M, Lash S, Robertson R (eds) Global modernities. Sage, London

Robson G (1997) Class, criminality and embodied consciousness: Charlie Richardson and a south east London habitus. Goldsmiths College, London

Rose A (1962) Human behaviour and social processes. Routledge and Kegan Paul, London

Ruggiero V, South N (1997) The late modern city as a bazaar. Br J Sociol 48(1):54–70
Soja E (1989) Postmodern geographies. Verso, London
Williams P (1993) Transnational criminal organisations and national security. Survival 36:96–113

British Crime Survey

John Flatley
Crime Statistics and Analysis Division,
Office for National Statistics, London, UK

Synonyms

British Crime Survey; BCS

Overview

The Crime Survey for England and Wales (CSEW), until 1 April 2012 known as the British Crime Survey (BCS), is a victimization survey of the population resident in households in England and Wales. The primary motive for launching the survey, over 30 years ago, was to assess how much crime went unreported in official police records. Today the survey has a high profile and has become a key social indicator charting trends in crime experienced by the general population. Its quarterly results receive considerable media attention, and politicians debate the implications of its findings.

However, from its inception, the survey was not designed to be merely a social indicator. One of its strengths has been that it has provided a rich source for criminological research and for informing policy development. Analysis of the survey has been instrumental in shedding more light on the nature and circumstances of victimization. The survey has not stood still and has continued to evolve and develop its methodology and content to respond to emerging issues.

The survey was first commissioned by the UK Home Office (the Government department responsible for crime and policing), but responsibility for it has recently transferred to the central Office for National Statistics. This move raises questions about the future role of the survey and whether it can continue to effectively balance its role as a key social indicator on the one hand with a continuing role as an invaluable research tool on the other.

Origins

The Home Office, the Government department responsible for crime and policing in England and Wales, started collecting statistics on crimes recorded by the police over 100 years ago. The numbers of crimes recorded by the police grew inexorably from less than 100,000 crimes per year at the turn of the twentieth century to half a million in the 1950s, a million in the 1960s, and two million in the 1970s. By 1981, over three million crimes a year were recorded by the police. Many criminologists were skeptical that this scale of increase represented a real rise of such magnitude in crime.

The first attempts to derive alternative measures of crime, based on sample surveys of the general population, started to develop in the United States (US) in the late 1960s. The first surveys of this kind were carried out for the US President’s Commission on Crime which in turn led to the establishment of the National Crime Survey (NCS) program and subsequently the National Crime and Victimization Survey (NCVS).

This and the development of small area victimization surveys in the UK inspired analysts working in the Home Office to press for a national victimization survey to be developed. The discussions around the case for such a survey within the Home Office has been outlined by some of those closely involved in the decision back in the late 1970s and early 1980s (Hough et al. 2007). A strong part of the case for the creation of the BCS, as it later became known, was to yield an assessment of what was referred to at the time as the “dark figure” of crime, that is, crimes which escaped official police records. Indeed the first published report of the survey’s findings refers to one of its main aims being to provide an “index of crime as a complement to the figures of (police) recorded offences" (Hough and Mayhew 1983).

The decision to commission a national survey of crime in England and Wales was taken by the Home Office in mid-1981, and the Scottish Home and Health Department (the Government department responsible for crime and policing in Scotland, at the time) decided to fund an extension to cover Scotland shortly thereafter. A contract for data collection was awarded to an independent survey organization, Social and Community Planning Research (SCPR) which has subsequently become the National Centre for Social Research (NCSR).

However, from its beginnings, the survey was conceived as being more than merely an instrument for counting crime. The survey was seen as providing a rich source of data about the characteristics of victims of crime, the nature, and consequences of victimization experiences. This contrasted with the police-recorded statistics which yielded little more than aggregate counts of offences by area. In addition, the survey was able to include questions on public attitudes to crime and crime-related issues, and again, this provided a rich new vehicle for criminological research.

Methodology

An annual technical report covering the survey methodology is produced by the survey contractor each year, and the latest version, at the time of writing, covered the 2010/2011 survey year (Fitzpatrik and Grant 2011). Below, a brief summary of the key aspects of methodology and how it has developed over time can be found.

Population Coverage, Sampling, and Weighting

The BCS was conceived as a survey of the population resident in households and, as such, was never intended to capture the victimization experiences of other groups (e.g., those permanently resident in institutions) or organizations (e.g., commercial victims). The survey was designed to yield results for a nationally representative sample of the adult population (defined as those aged 16 years and over). In more recent years, the sample has been extended to cover the experiences of children aged 10–15 years (see below).

While the survey covered the whole of Britain in its early years, it soon became restricted in its geographical coverage to just England and Wales. Thus, for most of its history, the BCS was something of a misnomer, and following the transfer of responsibility for the survey from the Home Office to the Office for National Statistics (see below), the survey was renamed as the *Crime Survey for England and Wales* on 1 April 2012 to better reflect its geographical coverage.

Since the first survey in 1982, the CSEW has randomly selected one adult per household sampled to take part in the survey. The sampling frame used to draw the original sample of households/adults has evolved over time from the electoral register (a record of those adults entitled to vote) for the first three rounds of the survey to the small-user Postcode Address File (a record of residential addresses maintained by the Royal Mail) ever since. The sample design has been subject to a number of modifications with high-crime areas over-sampled in the early years when the annual sample size typically ranged between 10,000 and 15,000 adults. The 2000 BCS was the only survey to adopt a fully proportional sample (i.e., with every area selected with probability proportional to size).

Following the expansion of the sample size thereafter, and the need for the survey to yield a number of key estimates at police force area level, the survey has moved again to a disproportionate sampling but this time with less populous (and thus lower crime areas) over-sampled. The sample design has been refined in recent years to move to a partially clustered sample design involving an unclustered sample of addresses being drawn in the most densely populated areas, with more clustered designs in the medium population density and low population density areas.

Throughout the history of the survey, appropriate weighting systems have been developed to adjust for the different selection probabilities involved in the respective designs. An independent review was commissioned by the Home Office in 2009 to examine the impact of the changes in sample design on the reliability of BCS crime trend estimates (Tipping et al. 2010). This showed that no bias had been introduced to the survey from changes in the sample design and that all designs examined had generated estimates of victimization with low levels of variance. The authors concluded that changes in the sample design had not affected the ability of the survey to identify trends in victimization reliably.

Crime Reference Period and Victimization Measures

The data collection period for the first BCS took place in the first few months of 1982 and asked respondents about incidents that had happened since 1 January 1981. For the purpose of analysis and reporting, all incidents occurring in 1982 were excluded so that respondents had a common crime reference period of the 1981 calendar year. A change to the crime reference period was required with the move to a continuous survey, with data collection taking place throughout the year, from 2001/02. The survey then moved to asking respondents about crimes that had been experienced in the 12 months prior to interview. This resulted in the development of a rolling reference period with, for example respondents interviewed in January asked about crimes experienced from the previous January to December, those interviewed in February asked about crimes experienced between the previous February and January of the year of interview, and so on. Respondents were additionally asked for information on the month in which an incident took place or, if they could not recall the exact month, the quarter. This allows analysts to retrospectively splice data from multiple years to construct crime rates on the basis of calendar years if they should wish to do so (see (Tipping et al. 2010).

Since the survey began, the basic approach to obtaining information on victimization experiences has not changed. Respondents are asked a series of screener questions designed to ensure that all incidents of crime within the scope of the survey, including relatively minor ones, are mentioned. The screener questions deliberately avoid using legal terms such as “burglary,” “robbery,” or “assault,” all of which have a precise definition that many respondents will not know. The wording of these questions has been kept consistent since the survey began to ensure comparability across years.

Depending upon individual circumstances, a maximum of 25 screener questions are asked covering:

- Vehicle-related crimes (e.g., theft of vehicle, theft from vehicle, damage to vehicle, bicycle theft)
- Property-related crimes (e.g., whether anything was stolen, whether the property was broken into, whether any property was damaged)
- Experience of personal crimes (e.g., whether any personal property was stolen, whether any personal property was damaged, whether they had been a victim of force or violence or threats)

All vehicle-related and property-related crimes are considered to be *household* incidents, and respondents are asked about whether anyone currently residing in their household has experienced any incidents within the reference period. For respondents who have moved within the last 12 months, questions on household crimes are asked both in relation to the property they are now living in as well as other places they have lived in the last 12 months. *Personal* incidents refer to all crimes against the individual and only relate to things that have happened to the respondent personally, but not to other people in the household. Weights are created to allow analysis at either household or personal level.

All incidents identified at the screener questions are then followed through in a series of detailed questions. The first three victimization modules include detailed questions relating to each incident; the last three victim modules are shorter modules, designed to be much quicker to complete to avoid respondent fatigue during the interview. The order in which the victim modules are asked depends on the type of crime – less common crimes are prioritized in order to collect as much detailed information as possible. A total of up to six victimization modules can be completed by each respondent.

Most incidents reported represent one-off crimes or single incidents. However, in a minority of cases, a respondent may have been victimized a number of times in succession. At each screener question where a respondent reported an incident, they are asked how many incidents of the given type had occurred during the reference period. If more than one incident had been reported, the respondent was asked whether they thought that these incidents represented a "series" or not. A series is defined as "the same thing, done under the same circumstances and probably by the same people."

Where this was the case, only one set of detailed victimization questions are asked in relation to the most recent incident in the series. There are two practical advantages to this approach of only asking about the most recent incident where a series of similar incidents has occurred. First, since some (although not all) incidents classified as a series can be petty or minor incidents (e.g., vandalism), it avoids the need to ask the same questions to a respondent several times over and thus reducing respondent burden. Secondly, it avoids "using up" the limit of six sets of detailed victimization questions on incidents which may be less serious.

In subsequent analysis and reporting, the convention has been for single incidents to be counted once and *series* incidents to be given a score equal to the number of incidents in the series occurring in the reference period, with an arbitrary top limit of five. This procedure was introduced in order to ensure that the survey estimates are not affected by a very small number of respondents who report an extremely high number of incidents and which are highly variable between survey years. The inclusion of such victims could undermine the ability to measure trends consistently. This sort of capping is consistent with other surveys of crime and other topics. However, the decision to cap series incidents at five has led to some criticism with it being claimed that this has served to mask the extent of chronic victimization and harms experienced by the most frequently victimized (Farrell and Pease 2007).

The first BCS developed an Offence Coding System to enable incidents reported by respondents to be assigned a criminal offence code (e.g., burglary or robbery) that approximated to the way in which incidents were classified by the police. This offence coding process takes place outside of the interview itself using a team of specially trained coders who determine whether what has been reported represents a crime or not and, if so, what offence code should be assigned to it. Apart from some minor changes, the code frame and the instructions to coders for the core survey have remained stable since the first survey.

Mode of Interview

The 1982 BCS was conducted in respondent's homes with trained interviewers collecting data via a paper-and-pencil-administered interviewing (i.e., interviewers filled in paper questionnaires by hand with extensive post-interview data keying and editing process). In 1994, the BCS moved to computer-assisted personal interviewing (CAPI) which brought improvements to data quality as it was possible to automatically route respondents through the questionnaire, without relying on interviewers following written instructions, and to incorporate logic and consistency checks on data entered in different parts of the questionnaire. CAPI also allows date calculations and text substitutions which help the interviewer ensure the interview flows smoothly.

The move to laptop computers provided further opportunities for innovation with self-completion modules introduced to cover sensitive topics. The use of self-completion on laptops allows respondents to feel more at ease when answering questions on sensitive topics due to increased confidence in the privacy and confidentiality of the survey.

Self-completion modules were first included in the 1996 BCS to obtain improved estimates of domestic violence (Walby and Allen 2004), and a similar module has been included since the 2004/2005 survey. A paper self-completion module on illicit drug use was included in the 1994 BCS though a change in the way in which these questions were asked when it was reintroduced with CASI means that comparable figures are only available from 1996. A separate annual report on the findings from the drugs self-completion continues to be published by the Home Office with the latest one at the time of writing relating to the 2011/2012 survey year (Home 2012a).

The self-completion modules are restricted to those respondents aged 16–59 years (the decision to exclude those aged 60 and over was an economy measure). In the 2008/2009 survey, a 6-month trial was conducted to examine the feasibility of raising the age limit to 69, but this proved unsuccessful (Bolling et al. 2009).

Regularity of Survey Fieldwork

Following the first BCS in 1982, a second survey took place in 1984 with fieldwork taking place early that year and with the preceding calendar year (1983) acting as the reference period. Subsequent surveys were repeated on the same basis in 1988, 1992, 1994, 1996, 1998, 2000, and 2001.

In the late 1990s, the value of the survey for policy purposes was recognized by the Home Office which planned to move the survey onto an annual cycle and to increase substantially its sample size. These moves were designed to as a means of increasing the frequency with which survey-based estimates could be published, while the increase in sample size was motivated by a desire to increase the precision of certain estimates. The latter was related, in part, for the need to use the survey to provide key performance indicators to assess the performance of the police.

The Home Office commissioned an independent review by leading experts in survey methodology and this reported in 2000 (Lynn and Elliot 2000). They recommended a number of major changes to the survey including:

- A move to a continuous fieldwork with the crime reference period being the 12 months ending with the most recent completed calendar month
- Expansion of the sample to obtain a minimum achieved sample of 600 or 700 interviews per year in each police force area
- Introduction of nonresponse weighting to incorporate both sample weighting (which was already being done) and calibration to population totals (i.e., to correct for potential additional nonresponse bias).

These recommendations were accepted, and the change took place with the 2001/2002 survey. There being 43 territorial police forces in England and Wales, this required a large increase in the annual sample size of the survey from around 20,000 in 2000 to around 33,000 in the 2001/2002 survey. Over the next few years, further demands for the BCS to provide key performance indicators led to further boosts to the annual sample size, which increased to over 36,000 in 2002/2003, some 38,000 in 2003/2004, and over 45,000 by 2004/2005.

More recently, it has been announced that due to cuts in government funding, the sample size of the survey will reduce with the 2012/2013 survey cut back to a sample of around 35,000 per year (Home 2012b). Nevertheless, it is a sign of the importance of the survey and the value in which it is held that in a period when other surveys have been stopped, the BCS continues to run with a sample some three times larger than the first survey in 1982.

Extension of Survey to Children

The survey was restricted to the experiences of adults except for a one-off exercise in 1992 when a separate sample of children aged 12–15 were interviewed (Aye Maung 1995). This previous exercise did not attempt to replicate the methodology of the adult survey or to combine estimates from the adult and child surveys.

Following recommendations in two related reviews of crime statistics (Smith 2006; Statistics Commission 2006), the BCS was extended to children aged 10–15 from January 2009.

The Home Office commissioned methodological advice on the feasibility of extending the survey to children (Pickering and Smith 2008), and following an extensive period of development and testing work during 2008, live data collection started in January 2009.

Children aged 10–15 are interviewed in households that have taken part in the main survey; where an eligible child is identified (according to age), one is selected at random to take part. Extending the survey to encompass children's experience of crimes raised some difficult issues with regard to classifying criminal incidents; e.g., minor incidents that are normal within the context of childhood behavior and development can be categorized as criminal when existing legal definitions of offences are applied. Four different methods for counting crime against children were examined and subject to user consultation (Millard and Flatley 2010). Following consultation, two main measures were adopted. The first, called the "broad measure," is one in which all incidents that are in law a crime are counted. This will include minor offences between children and family members that would not normally be treated as criminal matters. The second, referred to as a "preferred measure," is a more focused method which takes into account factors identified as important in determining the severity of an incident (as judged by focus groups with children).

Methodological differences between the adult and children's survey mean that direct comparisons cannot be made between the adult and child victimization data, although these estimates are now presented in the published quarterly crime statistics (Office for National Statistics 2012) to provide a better understanding of victimization experiences among adults and children resident in households.

Methodological Limitations

The CSEW is viewed as a gold-standard survey of its kind in terms of its methodology and response rate (the survey has continued to obtain a response rate of 75 % during the last decade when other voluntary household surveys in the UK have experienced falling response rates). However, as with other sample surveys, there are a number of inherent methodological limitations, some of which have already been touched on above.

The most obvious of these relate to sampling error and the inherent imprecision around survey estimates. There are other sources of survey error, such as related to respondents' ability to recall events and report them to the interviewer accurately. Some incidents may be forgotten or events beyond the reference period brought into scope. Respondents may be embarrassed or unwilling to disclose some information to interviewers. Questions may be misinterpreted by respondent, and interviewers themselves may make mistakes in recording information given to them, and coders may make errors of interpretation when assigning incidents to offence codes. The survey contractors go to great lengths to minimize these errors through careful training of interviewers, testing of questions, and quality assurance of all aspects of the survey process.

The Influence of the Survey on Criminological Thinking

The CSEW is now part of the UK National Statistics system with results from the survey reported on a quarterly basis alongside the police-recorded crime figures to provide a complementary and more comprehensive account of crime in England and Wales (Office for National Statistics 2012). The survey charted a rise in the volume of crimes covered by the survey through the 1980s reaching a peak in the mid-1990s. In line with the experience of many other industrialized economies, the survey described a fall in crime over the next 10–15 years with a generally stable trend in more recent years. However, the survey has had wider influence on criminological thinking, and it is this which is touched on below.

Victimization Risks Including Repeat and Multiple Victimization

The first BCS report highlighted the relative rarity of serious criminal victimization but also the fact that a small segment of the general population was particularly prone to falling victim (Hough and Mayhew 1983). At the time, the fact that it was younger adults, rather than the elderly, who were more likely to be victims of crime was a striking finding.

More detailed analysis of the dimensions of risk followed (Gottfredson 1984) including area-based analysis. For example, secondary analysis of the 1982 BCS revealed significant differences between high- and low-crime areas (Trickett et al. 1992). This showed a strong relationship such that the number of victimizations per victim rose markedly as the area crime rate increased. Further analysis showed similar patterns during the 1980s with a small number of areas hosting a disproportionate amount of crime (Hope and Hough, 1988). Further work examining the extent of repeat and multiple victimization includes one study of the BCS showing that 70 % of all incidents were reported by just 14 % of respondents who were multiple victims (Farrell and Pease 1993).

Such studies were influential in suggesting that a large proportion of all crime might be prevented if repeat or multiple victimization could be reduced and that crime prevention policy should focus on reducing the repeat victimization experienced by the most vulnerable people and in the most vulnerable places.

Sexual and Domestic Violence

The BCS has been groundbreaking in its approach to the measurement of sexual and domestic violence. A self-completion module on domestic violence was included in the 1996 BCS (Mirrlees-Black 1999) and on sexual victimization in both the 1998 and 2000 surveys. The 2001 BCS included a more detailed self-completion questionnaire on both these topics. It was designed to yield national estimates of the extent and nature of domestic violence, sexual assault, and stalking including the first ever such estimates of sexual assault against men (Walby and Allen 2004).

This showed that interpersonal violence affected around one-third of the population at some time in their lives with women suffering higher levels of victimization than men. It also demonstrated that such victims suffered high levels of repeat victimization, in particular of domestic violence. While these findings were not surprising to experts in the field, they quantified a phenomenon which helped to challenge others to seek to tackle these forms of violence.

Fear of Crime and Perceptions of Crime and Antisocial Behavior

Fear of crime has been a topic included in the survey from its very beginnings. It is a concept which is complex and difficult to measure. Over the years, the survey has deployed a range of questions, and researchers have sought to explain the key drivers and relationships (see, e.g., Allen 2006). Other researchers have used BCS questions to explore the frequency and intensity of fear of crime (Farrall and Gadd 2004). It has been suggested that few people experience specific events of worry on a frequent basis and that “worry about crime” is often best seen as a diffuse anxiety about risk, rather than any pattern of everyday concerns over personal safety (Gray et al. 2011).

The survey has also been used to describe two types of “perception gap”: one related to differences between what is happening to crime nationally and in the local area and, the second, the difference between perceptions of crime and actual crime levels. New questions added to the 2008/2009 survey revealed that the perception gap between changes nationally and in the local area were greater for the more serious violent (and therefore rarer) crimes and smaller for acquisitive crimes suggesting personal experience was more likely to play a part in the perceptions of the more common crimes while perceptions of rarer crime types are likely to be influenced by media reporting. In addition, analysis of small area police-recorded crime data showed a clear linear relationship between actual levels of crime and perceptions of the comparative level of crime in the local area (Murphy and Flatley 2009).

Antisocial behavior has become a major focus for government policy and practitioners since the late 1980s. Questions were added to the 1992 BCS to ask people about their perceptions of antisocial behavior in their area, and these have been revised and expanded over time to include personal experiences as well as perceptions and what such perceptions were based on (see, e.g., Upson 2006).

Public Attitudes to the Police and Criminal Justice System

The coverage of policing issues has expanded in the survey over time with specific modules of questions being included on contact with the police and public attitudes to the police (see, e.g., Skogan 1994). Analysis of the experience of ethnic minorities with the police has also been a focus of BCS analysis (see, e.g., Clancy et al. 2001). In more recent years, the focus on neighborhood policing and on police performance has led to questions being more routinely included on police visibility and public attitudes towards the police (see, e.g., Allen et al. 2006; Moon and Flatley 2011).

The BCS included questions on public attitudes to sentencing in the 1982 and in a number of subsequent rounds of the survey which has made an important contribution to the understanding of such issues. The 1996 BCS included a range of specific questions on attitudes to punishment and revealed a rather jaundiced view of sentencers and sentencing thinking them to be too lenient (Hough and Roberts 1998). This analysis showed that public disaffection was partly a result of ignorance of judicial practice and, ironically, when people were asked to prescribe sentences for actual cases, their sentencing tended to be in line with current sentencing practice.

New Types of Crime

Throughout its life course, the survey has sought to adapt to emerging issues, and below are some examples of how the survey has evolved to meet changing information needs.

Following the Stephen Lawrence inquiry in 1999, the survey included questions on racially motivated crimes, and in recent years, the module of questions has been expanded to ask victims of crime if they thought the crime had been religiously motivated. Other strands of "hate crime" have also been added to the survey more recently (see, Lader 2012).

Responding to changes in technology, a mobile phone theft module was added to the 2001/2002 survey to examine levels of such thefts. Subsequent analysis has explored the extent of mobile phone theft which showed the increased risk of such victimization among children and young adults (Hoare 2007).

Similarly, a module covering fraud and technology crimes was first introduced in the 2002/2003 BCS. This covered the extent to which people had been victims of debit and credit card fraud, their computers had been infected with viruses or hacked into, or they had been sent harassing emails. An Identity Theft module was introduced to the BCS in 2005/2006 (Hoare and Wood 2007).

Future Directions

As has been described above, responsibility for the CSEW has recently transferred to the Office for National Statistics. In a piece written to mark the 25th anniversary of the birth of the BCS, the founders of the survey described the tensions within the Home Office between the researchers (who promoted the survey) and the statisticians (who were more skeptical of its value).

As shown in Fig. 1, the survey has charted the rise in crime from the early 1980s through to the mid-1990s and the subsequent crime drop that has also been seen across other developed economies. As the survey has become institutionalised as a key social statistic in the UK, responsibility for it has passed from the researchers to the statisticians (first, following a structural reorganization in the Home Office in April 2008 and more recently with the move to ONS). Perhaps inevitably this has resulted in a subtle shift in priorities. However, up to now the survey has managed to combine its social indicator and research tool functions.

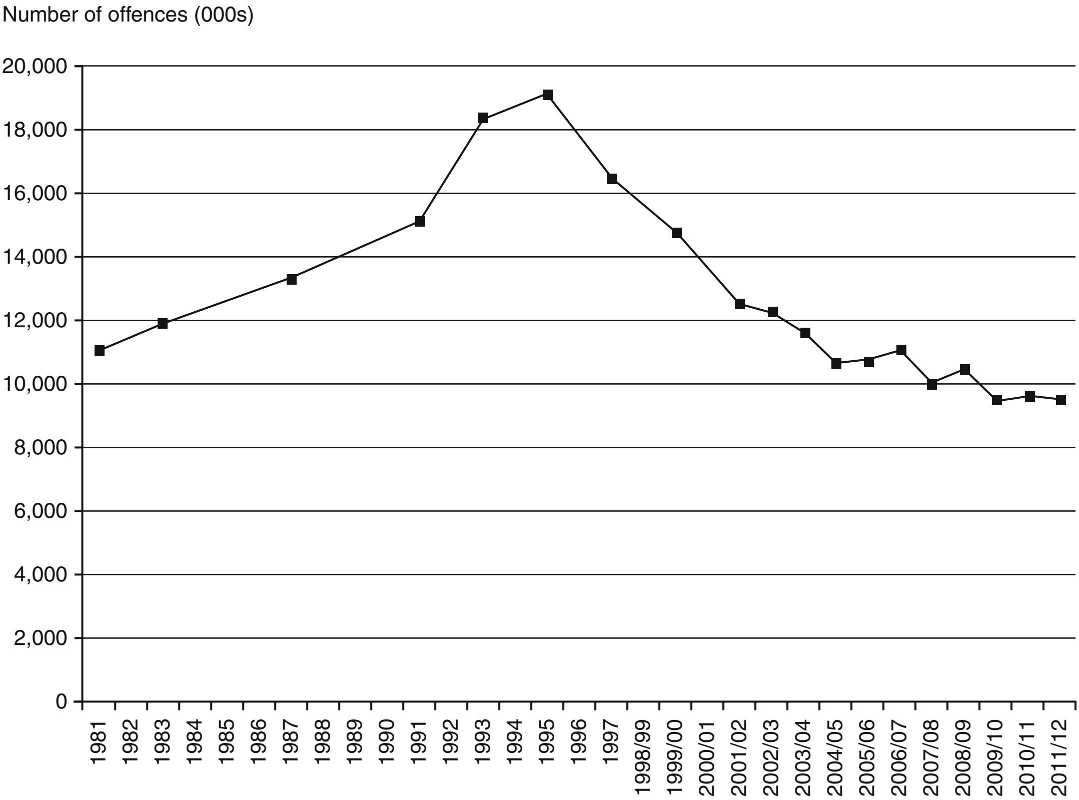

British Crime Survey, Fig. 1 Trends in crime estimated from the Crime Survey for England and Wales

One of the key challenges for the ONS going forward is to ensure the continuation of both these functions to ensure the rich legacy of the survey as key resource for criminological enquiry continues.

Related Entries

▶ History of the Dutch Crime Victimization Survey(s)
▶ International Crime Victimization Survey
▶ National Victimization Surveys
▶ US National Crime Victimization Survey

Recommended Reading and References

Allen J (2006) Worry about crime in England and Wales: findings from the 2003/04 and 2004/05 British Crime Survey, Home Office Online Report, 15 June 2006

Allen J, Patterson A, Edmonds S, Smith D (2006) Policing and the criminal justice system – public confidence and perceptions: findings from the 2004/05 British Crime Survey. Home Office, London, Home Office Online Report 07/06

Aye Maung N (1995) Young people, victimization and the police: British crime survey findings on experiences and attitudes of 12 to 15 years old. Home Office, London, Home Office Research Study 140

Bolling K, Grant C, Donovan J-L (2009) 2008–09 British crime survey (England and Wales) technical report, vol I. Home Office/BMRB Social Research, London

Clancy A, Hough M, Aust R, Kershaw C (2001) Crime, policing and justice: the experience of ethnic minorities. Findings from the 2000 British Crime Survey. Home Office, London, Home Office Research Study 223

Farrall S, Gadd D (2004) The frequency of the fear of crime. Br J Criminol 44(1):127–132

Farrell G (1999) Multiple victimization: its extent and significance. Int Rev Victimol 2:85–102

Farrell G, Pease K (1993) Once Bitten, Twice Bitten: Repeat Victimisation and its Implications for Crime Prevention, Police Research Group, London, Home Office Crime Prevention Unit Paper 46

Farrell G, Pease K (2007) The sting in the tail of the British crime survey: multiple victimisations. In: Hough M, Maxfield M (eds) Surveying crime in the 21st century, crime prevention studies, vol 22. Criminal Justice Press Monsey, New York

Fitzpatrik A, Grant C (2011) The 2010/11 British crime survey technical report volume 1. Home Office/TNS-BMRB, London

Gottfredson M (1984) Victims of crime: the dimensions of risk. HMSO, London, Home Office Research Study No 81

Gray E, Jackson J, Farrall S (2011) Feelings and functions in the fear of crime: applying a new approach to victimization insecurity. Br J Criminol 51(1):75–94

Hoare J (2007) Mobile phones: ownership and theft. In Flatley J (ed) Mobile phone theft, plastic card and identity fraud: findings from the 2005/06 British Crime Survey, Supplementary Volume 2 to Crime in England and Wales 2005/06, Home Office Statistical Bulletin 10/07

Hoare J, Wood C (2007) Plastic card and identity fraud. In Flatley J (ed) Mobile phone theft, plastic card and identity fraud: findings from the 2005/06 British Crime Survey Supplementary Volume 2 to Crime in England and Wales 2005/06, Home Office Statistical Bulletin 10/07 (2007)

Home Office (2012a) Drug misuse declared: findings from the 2011/12 crime survey for England and Wales. Home Office Online Report

Home Office (2012b) Consultation on changes to the BCS sample design, Home Office online report

Hope T, Hough M (1988) Area crime and incivilities: a profile from the British crime survey. In: Hope T, Shaw M (eds) Communities and crime reduction. HMSO, London

Hough M, Mayhew P (1983) The British Crime Survey: first report: Home Office Research Study No. 76. HMSO, London

Hough M, Maxfield M, Morris B, Simmons J (2007) The British Crime Survey over 25 years: progress, problems and prospects. In: Hough M, Maxfield M (eds) Surveying crime in the 21st century, crime prevention studies, vol 22. Criminal Justice Press Monsey, New York

Hough M, Roberts J (1998) Attitudes to punishment: findings from the British Crime Survey. Home Office, London, Home Office Research Study 179

Lader D (2012) The extent of and perceptions towards hate crime. In Smith K (ed) Hate crime, cyber security and the experience of crime among children: findings from the 2010/11 British Crime Survey: supplementary volume 3 to crime in England and Wales 2010/11, Home Office Statistical Bulletin 06/12

Lynn P, Elliot D (2000) The British Crime Survey: a review of methodology. National Centre for Social Research, London

Millard B, Flatley J (ed) (2010) Experimental statistics on victimization of children aged 10 to 15: findings from the British Crime Survey for the year ending December 2009 England and Wales, Home Office Statistical Bulletin 11/10

Mirrlees-Black C (1999) Domestic violence: findings from a new British Crime Survey self-completion questionnaire. Home Office, London, Research Study 191

Moon D, Flatley J (eds) (2011) Perceptions of crime, engagement with the police, authorities dealing with anti-social behaviour and Community Payback: findings from the 2010/11 British Crime Survey, supplementary volume 1 to crime in England and Wales 2010/11, Home Office Statistical Bulletin 18/11

Murphy R, Flatley J (2009) Perceptions of crime. In Moon D, Walker A (eds) Perceptions of crime and anti-social behaviour: findings from the 2008/09 British Crime Survey supplementary volume 1 to crime in England and Wales 2008/09, Home Office Statistical Bulletin 17/09

Office for National Statistics (2012) Crime in England and Wales, quarterly. First Release to March 2012; ONS Statistical Bulletin, London

Pickering K, Smith P (2008) British Crime Survey: options for extending the coverage to children and people living in communal establishments. Home Office, London

Skogan W (1994) Contacts between police and public: findings from the 1992 British Crime Survey. Her Majesty’s Stationery Office, London, Home Office Research Study 134

Smith A (2006) Crime statistics: an independent review. Independent report. Home Office, London

Statistics Commission (2006) Crime statistics: user perspectives. Statistics Commission Report No. 30. Statistics Commission, London

Tipping S, Hussey D, Wood M, Hales J (2010) British Crime Survey: methods review 2009 final report. National Centre for Social Research, London

Trickett A, Osborn D, Seymour J, Pease K (1992) What is different about high crime areas? Br J Criminol 32(1):81–89

Upson A (2006) Perceptions and experience of anti-social behaviour: findings from the 2004/05 British Crime Survey, Home Office Online Report 21/06

Walby S, Allen J (2004) Home Office Research Study 276. Domestic violence, sexual assault and stalking: findings from the British Crime Survey

British Police

Paul Lawrence
Department of History, The Open University,
Milton Keynes, UK

Overview

The development of policing in Britain prior to 1918 can be loosely divided into three phases.

From the early modern period until around 1850, the "old" police, a system of amateur or semiprofessional parish constables and night watchmen, were involved in the enforcement of laws and social norms. They shared this duty with a broad range of privately paid individuals and community groups, but, from c.1750 onwards, the efficacy of these localized and (relatively) informal mechanisms of policing was increasingly called into question. A variety of new initiatives were trialled, and, following on from the establishment of the Metropolitan Police in London in 1829, the so-called "new" police forces of the nineteenth century – uniformed, salaried, and preventive – were developed. These new forces did not initially mark a distinctive break with past practices and personnel, and there was a protracted period of overlap and reform. By about 1870, however, all administrative areas of Britain had instigated new, more "professional" institutions of policing. During the course of the nineteenth century, these new forces developed increasingly sophisticated policing techniques. Yet, while popularly depicted as a "golden age," policing in the later nineteenth and early twentieth century was typified as much by the persistence of police violence, scandal, and discontent as it was by the spread of new standards of order and control. In addition, regional variations across Britain further complicate any simplistic depiction of gradual and consensual reform.

Policing Before the "New Police"

Historical Antecedents

The presence of a uniformed and salaried police service is now so woven into the fabric of daily life that it is easy to forget that this is an institution of recent (in historical terms) advent. In Britain, the role of the central state in everyday life was relatively marginal prior to the nineteenth century, and the role of the individual and the local community in crime control, prosecution, and punishment was correspondingly greater. Indeed, the term "police" was not much used in Britain before 1800 and, even when in use, implied a much broader set of activities than just the prevention of crime and the detection of criminals (Dodsworth 2008).

This does not, however, mean that there was no "policing." During the early modern period, official policing duties in Britain were carried out by parish constables and urban watchmen. Under legislation dating primarily from the thirteenth century, constables were appointed locally, usually for no more than a year, and were charged with maintaining the King's peace in the district and reporting felons, miscreants, and nuisances to the local court. The efficacy of these parish constables was variable. There are some well-documented claims of negligence and inefficiency in apprehending offenders but there is also evidence that some served several terms of office, thereby gaining valuable expertise and authority. However, parish constables were not obliged to do anything which would take them beyond the boundaries of the jurisdiction they served, and there was no expectation that they would investigate all offenses which occurred in their locality (although they were expected to search premises or make arrests in instances where a magistrate's warrant was issued). Watchmen were recruited and funded locally to act as a deterrent to theft, usually in urban areas. They walked established beats at set times (usually at night), checking that doors and windows were locked. Watchmen made occasional arrests if they chanced across offenders, but were also not usually expected to seek out or track offenders after an offense had been committed.

In addition to parish constables and watchmen, there were a range of other, private arrangements made to safeguard property. Game keepers, private armed guards on mail coaches, individuals paid to watch barns and haystacks, and toll-gate keepers all had a role to play in the mixed economy of regulatory activity, and a significant amount of court cases resulted from apprehensions under such arrangements (King 2000, p. 63). In addition to these private policing initiatives, the role of the individual (whether as prosecuting victim or as helpful friend or neighbor) was crucial in early criminal justice processes. Individuals were expected to press (and pay for) charges, if the identity of the offender was known, and would often investigate crimes themselves or offer rewards for the recovery of stolen property.

Activities which would now be termed policing were thus carried out via a mixed regulatory economy, a rudimentary patchwork composed of limited state involvement, private initiative, business and community sanctions, and above all local enforcement of both legislation and social norms. Despite some tensions between the demands of magistrates and the norms of local communities, parish constables and watchmen maintained a broadly effective system of policing up until the end of the seventeenth century (Kent 1986).

The Bow Street Runners

From the 1690s onwards, successive governments attempted to encourage the apprehension of offenders (particularly highwaymen, burglars, and forgers) by expanding the system of rewards offered for successful prosecutions. However, as well as encouraging victims to prosecute, these incentives also served to multiply the number of private, entrepreneurial "thief-takers," many of whom were involved in corrupt practices such as encouraging (or even facilitating) robberies in order to claim the reward for subsequent arrests. With the bill for reward payments rising, the novelist and Bow Street magistrate Henry Fielding was granted a financial subvention in 1749 to enable him to establish a more organized body of men to confront violent offenders on the streets and highways around London. After Henry Fielding's death in 1754, this group was developed over subsequent decades by his blind half-brother John Fielding, one of the most innovate magistrates of the eighteenth century, into the Bow Street Runners. Sir John shaped these individuals into a well-known and stable group of officers with recognized expertise in investigating crimes, tracking and arresting offenders, and presenting evidence in court. While the men had no official title, because their role had no official standing, they were an innovation in policing in being both proactive in seeking out offenders (prompted either by orders from a magistrate or by a fee paid by the victim of a crime) and in gaining a degree of proficiency and continuity of service hitherto unmatched.

As part of their working practices, the Bow Street officers developed one of the first enduring systems of record-keeping relating to criminals and a range of rudimentary forensic techniques (e.g., ballistic and medical analyses). Thus, the "runners" can be thought of as "the first detectives," although the word itself was not used until the nineteenth century (Beattie 2012, p. 2). Indeed, some "principal officers" (a term they employed to differentiate themselves from the more junior patrolmen of the Bow Street magistrate's office) became highly sought-after professional investigators. John Townsend, for example, who had previously worked as a street vendor, gained such notoriety after a string of detective successes that he developed a close friendship with the Prince Regent and remained in popular demand among those who could afford his services.

The work of the principal officers was not, moreover, confined to London. Unlike parish constables, they were not bound to operate in any particular jurisdiction and, from the 1780s onwards, Bow Street officers were increasingly engaged by employers across the country, to the extent that they might be considered in some ways a precursor to a national police force (Cox 2010). The provincial magistracy, in particular, often required Bow Street assistance in the hunt for offenders, and principal officers often traveled significant distances in pursuit of felons. Because it was primarily the wealthy who engaged the services of Bow Street, this did mean that the officers primarily investigated crimes damaging to commercial interests (such as forgery, arson, and fraud) and there is certainly no sense that this was a police force acting to enforce the law uniformly and in the interest of all members of the public. Moreover, while there is little doubt that many of the principal officers were energetic and effective thief-takers, they were less effective as a preventive force and their numbers were arguably insufficient to make a significant impact on daily life in London, let alone further afield.

Debate and Reform from 1750

During the latter half of the eighteenth century, policing was thus by no means wholly unsatisfactory in many urban areas of Britain. The significance of London as a center of population and commerce meant that policing here was the most well developed in the nation. In addition to the Bow Street Runners, many districts operated their own locally controlled police services (Paley 1989), and many parishes also paid for and operated sophisticated watch systems (Reynolds 1998). There has been comparatively little research into the policing of the rural counties, but it is likely that the quality of policing here depended primarily on the personal inclination and ability of local justices of the peace and parish constables. However, from the mid-eighteen century onwards, significant national debates took place about the desirability of developing new modes of policing which might curtail local control and variability in the field of law enforcement. These debates were centered on London, as the most significant (and rapidly developing) urban center in the period, but also generated strong views from local authorities concerned about the implications of new forms of policing for their power to govern.

A number of factors prompted a focus on policing reform from the 1750s onwards. There was a public concern about the increasing incidence of certain types of crime (including theft, highway robbery, and rioting), but this concern has to be seen as part of a broader anxiety about how to cope with a period of rapid social change in which a mobile laboring class was developing. Equally, the expansion of the ambitions of the state during the later decades of the eighteenth century led to tensions over the correct relationship between central and local authority, and debates over policing reflected these tensions. There were also specific debates over the efficacy of existing policing arrangements as Fielding's view that crime was best prevented by an efficient detective force came under attack from others (such as Patrick Colquhoun) who believed that a preventive police (operating primarily via the patrol of predetermined beats) would be a better solution to the problem (Neocleous 2000).

Attempts to reform policing in London began in the 1780s when the City of London (the administrative division at the wealthy heart of the capital) organized a small, regular patrol for the city's streets. This patrol was uniformed from 1791 and gradually increased until it numbered 24 in 1824. In other areas of London, reform attempts began in 1785 when Pitt the Younger introduced the Westminster Police Bill, which was designed to put in place a centrally controlled police force for the whole city. This bill was defeated, but a far more moderate alternative was passed in 1792 and established seven new police offices in the city, each staffed by three stipendiary (salaried) magistrates, with six constables attached to each office. Further reforms followed, with a privately funded police force established by Act in 1798 to protect goods at London's docks. In 1800, this force was publicly funded as the Thames River Police. Private involvement in crime control remained significant. The Bank of England, for example, developed sophisticated nationwide mechanisms for apprehending and prosecuting forgers (McGowen 2005). The situation in London was unique, however, and most of the country, outside the large towns, "relied for their basic law enforcement on unpaid, part-time constables aged 30 or over" (Philips and Storch 1998, p. 5). Thus, while a variety of debates and reforms were already underway before the advent of the "New Police" in 1829, these were very much focussed on London and the larger cities (notably Glasgow), with much of Britain still policed lightly by systems which had been in place for centuries.

The "New Police"

The Metropolitan Police Act 1829

The most obvious outcome of the debates around policing which took place in the period 1780–1820 was the formation of the Metropolitan Police in London in 1829. This force was the most significant element of the so-called "New Police" of the nineteenth century. This shorthand term, which came into regular use in the decade after 1829, is now used as a useful descriptor for what many have seen as a wholesale reorganization of policing in Britain between 1829 and 1870. It refers to any of the new forces set up in response to the Police Acts of 1829, 1839–1840, and 1856 or, collectively, to all of them.

The Metropolitan Police Act was introduced to parliament by Sir Robert Peel in 1829. When Peel took over as Home Secretary in 1822, he argued the need for a new system of police – uniformed and supposedly both more efficient and more professional than their eighteenth-century counterparts and able to control the capital's new suburbs where the old system was weak. Although the validity of the statistics Peel used to "prove" his case in 1829 (by showing a tide of rising criminality) was probably minimal, he handled the political sensitivities of the issue adeptly, concealing the fact that he had already made up his mind what the reformed police should look like (Philips 1980, p. 186). Given an easy ride through parliament because of the attention being directed to the issue of Catholic emancipation, the subsequent act established the Metropolitan Police, neatly sidestepping concerns expressed by the City of London Corporation (which was worried about infringements on its authority) by leaving it outside the jurisdiction of the measure.

Thus, in 1829, the Metropolitan Police Improvement Bill was passed, and the new force of 3,000 uniformed constables took to the streets between September 1829 and May 1830. Throughout the 1830s, there was vociferous criticism of Peel's new force. Many commentators felt the uniforms of the new police betrayed their essentially "military" character, associating them with the gendarmeries of the continent, which were seen to serve political ends. Others complained about the cost of the new system, which was more expensive than the old and often put fewer men on the streets, but which still had to be paid for out of local taxes. In addition, there was considerable overlap with the existent forces policing London. The new police initially operated in tandem with 300 or so constables operating out of Bow Street and other police offices, under the direction of magistrates (loosely supervised by the Home Office). While these constables were primarily a detective force, and the new Metropolitan Police were primarily preventive, there was obviously overlap in their activities. This was investigated by a number of parliamentary select committees during the early 1830s and, in 1836, the Horse Patrol was amalgamated into the Metropolitan Police, with the Bow Street constables being absorbed in 1839. The 1833 Select Committee on the Metropolitan Police did also recommend that the separate jurisdiction of the City of London be abolished, but eventually the city introduced its own police force, via a private bill which precipitated fierce debate (Harris 2004, p. 132–53). As a legacy of this, the City of London still maintains its own police force distinct from that of the rest of metropolitan London.

Thus, the introduction of the Metropolitan Police was the result of a complex process of gradual and contested reform and did not necessarily mark the distinct and sharp break with previous practices that the epithet "New Police" might suggest. Such a large, uniformed force (dressed in top hats and blue tunics with high, stiff collars), firmly based on the preventive principle of beat policing (and initially featuring no detective division), was an innovation. The fact that this new force's commanders (two Commissioners of Police and a Receiver responsible for financial matters) were appointed by the Home Secretary and ultimately responsible to Parliament was also a departure from previous practice. Despite this, the new metropolitan force had considerable continuities of working practice with previous law-enforcement bodies, and there was also a considerable overlap in personnel, with many "old" police constables finding work within the Metropolitan Police.

National Police Reform

Outside of London, by the late 1820s, members of the provincial ruling class were also increasingly discussing the presumed inadequacy of existing policing arrangements (Philips and Storch 1994). This did not mean that reform followed swiftly, however. After the success of his 1829 Act, Peel devised plans for a "national" network of policing managed by stipendiary magistrates controlled by the Home Office. A bill along these lines was nearly introduced in 1832, but plans were shelved when it became apparent that local authorities were primed to resist on grounds of cost and the erosion of local control. Central control was imposed in some areas where fears of public disorder were particularly acute. In 1839, for example, Birmingham, Bolton, and Manchester had the control of policing taken out of local hands due to fears of political agitation. In general, however, the legislation which was successfully passed in the 1830s was primarily permissive – allowing for policing reform under local control but not requiring it. The national Lighting and Watching Act of 1833, for example, paved the way for urban policing reform by providing borough authorities with the means to improve their daytime watches and night patrols. The Municipal Corporations Act of 1835 specified that watch committees be set up to appoint and direct urban police forces. For rural areas, the recommendations of the rather one-sided 1836 Royal Commission on a Constabulary Force led to the Rural Police Act of 1839 (and a further amending act in 1840). This permissive act left the decision on whether to establish a rural police (as well as the control of that force) in the hands of the county magistrates.

While 24 counties had implemented the 1839 Act by 1841, and a further 11 did so in the next 15 years, around a third of counties still had not implemented any reforms by the mid-1850s. For this reason, another Select Committee was appointed in 1853 to consider the expediency of adopting a more uniform system of police. The resultant police bill, introduced in 1854 against a backdrop of extensive public order problems in the north of England, proposed that boroughs with less than 20,000 inhabitants would cede control of their forces to the counties and that there would be greater Home Office interaction with Chief Constables. These proposals again encountered stiff opposition, primarily from boroughs, with strong support from the City of London, and the bill was withdrawn. A subsequent bill introduced in 1856, which left local control of the police intact, was more favorably received, partly because of the ending of transportation to Australia, which elevated public concern over ex-convicts remaining at large in Britain. The resultant County and Borough Police Act of 1856 made the establishment of police forces a requirement at local government level in all counties and boroughs and allowed for part-funding from central government funds via a grant which the force would only receive if it passed an annual inspection (Parris 1961).

By c.1870 almost all provincial authorities had established new police forces but, overall, police reform in the nineteenth century was a complex, evolutionary process. Most counties and boroughs recognized the need for change but feared the loss of their authority to central government. The process of reform was further complicated by the fact that some new forces around the country were actually set up in advance of the national legislation, as a result of local initiative. In Gloucestershire, for example, the parish of Dursley had created its own force in 1814 in a way which prefigured later, national interventions. Thus, the idea that the Metropolitan Police Act served as an example to other areas of the country cannot be wholly sustained. A further layer of complexity is added by the fact that other national legislation passed during the period (such as the Parish Constables Act of 1842 and the Superintending Constables Act of 1850) in fact allowed rural parishes to appoint constables on the model of the "old" system, an allowance that many availed themselves of (Philips and Storch 1998, p. 213–19). Thus, while the police forces introduced as a result of the police acts passed between 1829 and 1856 were "new" insofar as novel legislation compelled the implementation of new structures locally, the debates, delays, and diversions around this implementation meant that the "New Police" did not necessarily represent a sharp break with the past. This statement is equally true in terms of police practice, where continuities with older mechanisms of policing were also in many ways also apparent.

The Practice of Policing

The Process of Professionalization

By the latter quarter of the nineteenth century, recognizably modern structures of policing were in place in England and Wales (with Scotland and Ireland, as outlined below, being policed under separate arrangements). In some senses, these "new" police forces were professional from the outset. In contrast to the "old" police, they were salaried, uniformed, and paid pensions after retirement. However, initially at least, many forces had considerable problems with both the enforcement of discipline and high staff turnover, with many constables fined or dismissed for drunkenness and inefficiency. This was partly due to key differences in the expectations the new forces had of their employees. In particular, the establishment of close supervision of the body of the constable (via drill, beat monitoring, and instruction), combined with the novel forms of time discipline required, meant that many "old" police constables found it hard to make the transition to the newer forces (Williams 2013). There were also periodic petitions in some cities for better pay and conditions. While a Chief Constables Association was formed in 1893, the clandestine Police and Prison Officers Union, which sought to represent the rank and file and organized a police strike in 1918/1919, was swiftly outlawed. British police officers have been barred from joining trade unions ever since.

Gradually, however, police forces matured as employers. The Metropolitan Police was one of the largest work organizations of any kind in nineteenth-century London, employing almost 22,000 men by 1914 (Shpayer-Makov 2002). Yet, from the outset, the Metropolitan Police (as with other British forces) was essentially a meritocracy, with all but the top 11 positions being filled from those who had risen through the ranks. This, combined with good pension provision (unusual for working-class employees during the nineteenth century), meant that the labor force of the new police gradually bedded in and the high turnover rates of the early years eventually subsided. The size, nature, and working practices of the new forces were far from static, however. As police forces matured, so new elements of professional practice were continually added to the policing repertoire, a key element of which was plain clothes detection.

Because of public fears that the new police could easily become a political "spy" police such as that believed to operate within many European countries, there were initially no detective departments or specialized detectives within the new police forces. While there were some officers who worked in plain clothes, it soon became clear that dedicated detection activity was a necessary part of police work. The Metropolitan Police set up a detective division in 1842, and the City of London and some larger cities and towns followed suit. Yet detective activity remained a surprisingly small element of police work. There were only a few hundred police detectives in the mid-nineteenth century, although this number had more than doubled by 1918 (Shpayer-Makov 2011). Britain did also, eventually, develop a "political" police of sorts. The secretive Special Branch, initially the Special Irish Branch, was set up in 1883 to combat Irish terrorism on the British mainland, but eventually extended its activities to encompass the surveillance of a wide range of individuals, including those with known left-wing political views (Porter 1987).

The identification of suspects was another important element of police work which underwent significant development. In an age before routine identification documents such as passports, knowing whether a person held in custody was a repeat offender was a complex problem, and the new police were quick to put emerging technologies to good effect. Photography was in use in British prisons as early as 1847 but was not common practice within police forces until the 1871 Prevention of Crime Act, which set up a central registry held at Scotland Yard (the headquarters of the Metropolitan Police). By 1886, this registry contained c.60,000 photographs of known offenders, showing the police as early and enthusiastic adopters of the technology (Stanford 2002). The crucial aid to the identification of suspects was the fingerprint bureau set up at Scotland Yard in 1901. Fingerprinting rapidly became the default means of identification in use in Britain and was much emulated around the world. The technology was not developed in England, however, but rather in India in the 1890s (by the Inspector General of Police in Bengal, Edward Henry), before exported back to the metropole (Cole 2002).

Novel technologies were certainly useful to the practice of policing but much police work was (and still is) based on routine bureaucracy and this, too, is worthy of consideration. The development of sophisticated systems for sharing information of all kinds marks part of the process of professionalization of the police. Internally, from the outset, the new police forces ran themselves via interlocking systems of reports and returns, which subjected the actions of each level to the surveillance and control of those above them. In relation to crime control, the 1871 Prevention of Crime Act set up a register of "habitual criminals," and from this register, the Home Office prepared an annual register of distinctive marks for circulation to all police forces. In addition, the Metropolitan Police maintained its own sophisticated archives which enabled better tracking and identification of suspects. Other regional forces had similar mechanisms for circulating information about increasingly mobile criminals in an era before widespread telephone communications. Such everyday practices as keeping files and circulating information in a timely and effective manner were the product of considerable thought and design and came to form an integral element of modern policing (Williams 2013).

The Limited Power of the State

Although police forces were larger and better trained than ever by the turn of the twentieth century, there were always limits to the capabilities of the police. Just as the narrative of institutional police reform was a convoluted one, so the story of the development of policing practice is not just one of increasingly sophisticated techniques of control. On the one hand, despite a public rhetoric which presented the British policing as consensual and the individual constable as simply "a citizen in uniform," the new forces undoubtedly acted to some extent to enforce new, middle-class standards of decorum. In some districts, the new constables were termed "a plague of blue locusts" due to their role in regulating drinking, gambling, and other working-class leisure activities (Storch 1975). Street offences (such as public drunkenness, vagrancy, and begging) took up an inordinate share of police time, with a similarly large focus devoted to the prevention and investigation of property crime. White-collar crime (a term coined in 1939 to describe fraud and other financial crimes committed by those in positions of authority) was not something which occupied the police particularly during the nineteenth century. The new police could, therefore, be viewed as "a new engine of power and authority," helping to manufacture an orderly working class suitable to the modern, capitalist economy (Philips 1980).

In reality, however, it was always impossible for the police (as for contemporary forces) to accomplish everything which might be required of them. While the police may have been granted extensive legislative powers and been asked to intervene frequently in the fields of popular morality and public behavior, this does not mean that they always did. Prevalent street culture remained resistant to police control, and the two Metropolitan Police Commissioners (and other chief constables around the country) often found themselves, of necessity, having to tread a path midway between the demands of local authorities and pressure groups, on the one hand, and the practical necessity for good relations with the populations they policed, on the other. Overall, policing practice in the nineteenth century was often designed to establish minimum standards of public order without provoking social conflict by aiming at unattainable standards. The role of discretion on the part of both individual officers (and senior commanders) in deciding what offenses to prioritize was always significant in the operation of policing. While the new police undoubtedly did spend a great deal of their time concerned with minor property crime and street offenses, these duties were often undertaken with sympathy and a degree of understanding (Lawrence 2004), and the police were in some ways a social mechanism which could be engaged by all classes (Davis 1984).

Consideration of the social backgrounds of police officers is a useful key to understanding some elements of policing practice during the period to 1918. English policemen (and they were exclusively male until the First World War) came from a range of backgrounds, but were overwhelmingly drawn from the unskilled and semiskilled working class (Emsley and Clapson 1994). This meant that some officers were at times understandably reluctant to enforce what appeared to many of them to be pointless attempts at social engineering. Moreover, the acquisition of social status through physical strength and aggression remained a key element of working-class culture during this period, and this meant that many police officers strove to be "as hard as the local hard men" (Emsley 2009, p. 149). Violence on the part of the police was a recognized (although not celebrated) facet of policing practice for much of the nineteenth century. Public order policing (particularly the regulation of political demonstrations and trades unions meetings) remained a key part of policing and here, too, violence on part of police also far from unusual.

Overall, it is apparent that the way the police were organized and the duties they were given meant that they were more likely to come into contact with members of the working class than other social groups. It was not until the advent of mass motoring that the middle classes really came into extensive contact with the police. However, any consideration of the operation of the new police in the nineteenth century needs to recognize both the novel demands made of constables and the pragmatism and discretion with which they mediated these demands.

Regional Variations

The policing of the outlying areas of Britain reveals both similarities and differences with England. Wales had the same legal system as England during the nineteenth and early twentieth century, and hence policing developments there followed similar patterns. Initial resistance to reform on the grounds of cost and the erosion of local control was followed by gradual acceptance. While the new police in Wales were involved in the imposition of new standards of social order, the size and composition of forces (particularly in rural areas) very often impeded the efforts of elites to impose certain aspects of social regulation (Jones 1982).

Scotland had its own legal system, yet there has been relatively little research focussed specifically on Scottish policing. What is evident is the significance of the English model, but also the ways in which it was tailored to local conditions (Barrie 2008). Scotland emerged from the age of improvement with a modern, professional police, but constabularies were initially administered under a dual control system whereby elected police commissions shared responsibility for the administration of forces with town councils and magistrates. Scottish police forces, as was the case in England and Wales, customarily adopted a pragmatic approach to the many demands placed upon them. More significantly, as in the English case, some elements of working-class culture, such as heavy drinking, remained more resistant to the new police than both police commissions and town councils would have liked.

Ireland, despite having the same legal framework as England from 1801 to 1922, had particular social, economic, and political circumstances which made issues of law and order (particularly public order policing) especially significant. As in England, an "old" police existed in Ireland at the turn of the nineteenth century. Parochial constables had no uniform or training and were usually only policemen in their spare time, but were tolerably effective in some areas (Malcolm 2006). Reforms introduced by the government in London were implemented in 1822 via the Irish Constabulary Act, which established a new force in each barony with chief constables and inspectors general under the control of Dublin Castle (the seat of British power in Ireland). The duties of these forces were in many ways similar to those of the new constabularies in England but they were paramilitary in nature. Armed from the start and divorced from local civil control, the police in Ireland had a much more prominent role in the maintenance of public order than forces in mainland Britain.

Following the Police Reform Act of 1836 and under the command of James Shaw Kennedy (who had originally been offered the commissionership of the Metropolitan Police by Robert Peel), the Irish Constabulary (which gained the prefix "Royal" in 1867) wore bottle-green uniforms, were armed with short-barreled carbines and sword bayonets, and lived in barracks. Dublin itself was policed by a separate civil force – the unarmed Dublin Metropolitan Police. While the RIC gradually became more representative of the Irish population, largely due to sustained efforts to recruit more Catholics into the rank and file, this process took decades, and the turnover among police personnel remained high. Thus, as with the English case, the nineteenth century saw a gradual evolution of recognizably modern mechanisms of policing in Ireland. Although the public order issues raised by Ireland’s complex relationship with mainland Britain, and the armed and national nature of the RIC, meant that there were distinct differences, it is also obvious that constables in both England and Ireland faced many of the same challenges and also undertook very similar sorts of duties on a daily basis.

Conclusions and Future Research

The advent of the “new police” in nineteenth-century Britain did not represent the sudden or total shift in institutions and practices which historians once assumed. The “old” police was in many ways a tolerably effective system in some areas, and there was significant continuity in both practices and personnel as new forces were set up post-1829. Police reform and changes in operational practice during the nineteenth century were extensive but were also often gradual and cumulative, rather than sudden or decisive. By 1918, policing in Britain and Ireland was very different to the pre-1829 era, but even in the 1880s and 1890s, many remnants of older, private, or informal mechanisms of policing still survived.

Despite 30 years of sustained academic research into British police history, however, there are still numerous unexplored topics. In particular, while significant attention has been lavished on London (rightly so, given its significance in the reform process) and a few other major cities, there is relatively little research which looks in detail at how policing was conducted in the rural areas and smaller towns of England. In addition, Wales, Scotland, and Ireland require considerable research to establish the extent to which developments in England are generalizable to those areas also. The interaction of the new police with the many other interest groups and agencies which were involved in the regulation of social life in the nineteenth century is another area requiring further attention. The significance of policing for any true understanding of the social history of the recent past hopefully means that these areas will not be lacking research for long.

Recommended Reading and References

Barrie D (2008) Police in the age of improvement: police development and the civic tradition in Scotland 1775–1865. Willan, Cullompton

Beattie JM (2012) The first english detectives. The bow street runners and the policing of London, 1750–1840. Oxford University Press, Oxford

Cole S (2002) Suspect identities: a history of fingerprinting and criminal identification. Harvard University Press, Cambridge

Cox DJ (2010) A certain share of low cunning. A history of the bow street runners, 1792–1839. Willan, Cullompton

Davis J (1984) A poor man’s system of justice: the London police courts in the second half of the nineteenth century. The Hist J 27(2):309–335

Dodsworth F (2008) The idea of police in eighteenth-century England: discipline, reformation, superintendence, c.1780–1800. J Hist Ideas 69(4):583–604

Emsley C (1985) The thump of wood on a Swede turnip: police violence in nineteenth-century England. Crim Justice Hist 6:125–149

Emsley C (1996) The English police, 2nd edn. Pearson, Harlow

Emsley C (2009) The great British bobby. Quercus, London

Emsley C, Clapson M (1994) Recruiting the English policeman, c.1840–1940. Polic Soc 3:269–286

Harris A (2004) Policing the city: crime and legal authority in London 1780–1840. Ohio State University Press, Columbus

Jones D (1982) Crime, protest, community and police in nineteenth-century Britain. Routledge and Kegan Paul, London

Kent J (1986) The English village constable 1580–1642. Oxford University Press, Oxford

King P (2000) Crime, justice and discretion in England 1740–1820. Oxford University Press, Oxford

Lawrence P (2004) Policing the poor in England and France, 1850–1900. In: Emsley C, Spierenburg P, Johnson E (eds) Social control in modern Europe 1800–2000. Ohio University Press, Athens

Malcolm E (2006) The Irish policeman 1822–1922: a life. Four Courts Press, Dublin

McGowen R (2005) The bank of England and the policing of forgery, 1797–1821. Past and Present 186(1): 81–116

Neocleous M (2000) Social police and the mechanisms of prevention: Patrick Colquhoun and condition of poverty. Bri J Crim 40:710–726

Paley R (1989) An imperfect, in adequate and wretched system? Policing London before peel. Crim Justice Hist 10:95–130

Parris H (1961) The home office and the provincial police in England and Wales, 1856–1870. Public Law 117(26):230–255

Philips D (1980) A new engine of power and authority: the institutionalization of law-enforcement in England 1780–1830. In: Gatrell VAC, Lenman B, Parker G (eds) Crime and the Law: the Social History of Crime in Western Europe since 1500. Europa, London

Philips D, Storch R (1994) Whigs and coppers: the grey ministry's national police scheme, 1832. Hist Res 67(162):75–90

Philips D, Storch R (1998) Policing provincial England 1829–1856. Leicester University Press, Leicester

Porter B (1987) The origins of the vigilant state. Wiedenfeld & Nicolson, London

Reynolds E (1998) Before the bobbies: the night watch and police reform in metropolitan London, 1720–1830. Macmillan, London

Sharpe JA (1998) Crime in early modern England, 2nd edn

Shpayer-Makov H (2002) The making of a policeman. A social history of a labour force in metropolitan London. Ashgate, Aldershot, pp 1829–1914

Shpayer-Makov H (2011) The ascent of the detective: police Sleuths in Victorian and Edwardian England. Oxford University Press, Oxford

Stanford T (2002) Who are you? We have ways of finding out. Tracing the police development of offender identification techniques in the late nineteenth century. Crimes Misdemeanours 3(1):54–81

Williams CA (2013) Directing Britain's police 1780–1980. Manchester University Press, Manchester

Broken Windows Thesis

Joshua C. Hinkle
Department of Criminal Justice and
Criminology, Georgia State University,
Atlanta, GA, USA

Synonyms

Incivilities thesis; Order maintenance

Overview

Few ideas in criminology have had the type of direct impact on criminal justice policy exhibited by the broken windows thesis. From its inauspicious beginnings in a nine-page article by James Q. Wilson and George Kelling in *The Atlantic Monthly*, the broken windows thesis has impacted policing strategies around the world. From the "quality of life policing" efforts in New York City (Kelling and Sousa 2001) to "Zero Tolerance" policing in England (Dennis and Mallon 1998), police agencies around the world have embraced Wilson and Kelling's idea that focusing on less serious offenses can yield important benefits in terms of community safety and prevention of more serious crime. Despite this broad policy influence, research on the theory itself has been relatively weak and has produced equivocal findings as will be detailed in this entry.

The Broken Windows Model

In a nutshell, the broken windows thesis (Wilson and Kelling 1982) suggests that police could more effectively fight crime by focusing on more minor annoyances which plague communities – hereafter referred to as disorder (some works also label these issues as "incivilities"). Disorder includes both rundown physical conditions in the form of litter, dilapidated buildings, graffiti, etc. as well as social nuisances such as panhandling, loitering, and public drinking. Their idea that crime could be prevented through targeting these issues was based on their thesis that such social ills eventually lead to community decline if left untended.

This process essentially has four key steps and begins with disorder not being dealt with in a timely manner. Trash is not picked up; loiterers are not asked to move on; people are drinking in public without being warned away. In time this invites more trash being thrown in the street, more loiterers to gather, and more people to start drinking in public. As this disorder accumulates, it sends a message to residents that things are getting out of control and that social

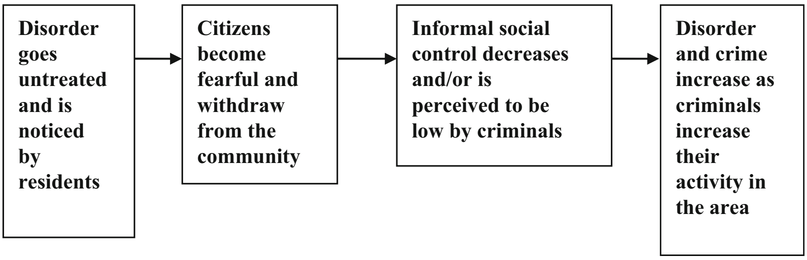
Disorder goes untreated and is noticed by residents
Citizens become fearful and withdraw from the community
Informal social control decreases and/or is perceived to be low by criminals
Disorder and crime increase as criminals increase their activity in the area

Broken Windows Thesis, Fig. 1 The broken windows thesis

controls have failed in their neighborhood. The key here is that residents perceive this untended disorder accumulating. It will likely have little impact if residents are not aware of the disorder in the community. In turn, the next step of the process is that residents who perceive worsening disorder problems eventually become fearful and begin to withdraw from the community. They spend less time outside, become less likely to intervene and ward off disorderly people, and, in the extreme, "good" residents may move away.

This leads to the third step of the process, which is a lowering of informal social controls. The community withdrawal results in fewer watchful eyes on the street, and the area is now less able to regulate behavior through informal social controls. As such disorder and minor crimes continue to flourish in these areas. Finally, this brings about the fourth step, in which criminals take these signs of untended disorder as a cue that such a neighborhood is a good place for them to work with relative impunity as they perceive their chances of being arrested in such areas to be slim. In Wilson and Kelling's terms, such neighborhoods are vulnerable to criminal invasion. It is not inevitable, but such places are much more likely, in their view, to see an increase in crime than neighborhoods which exert control in regulating the occurrence of disorder. Once crime occurs, residents also notice this and the cycle of fear and withdrawal is likely to worsen as the process is essentially a feedback loop. The steps of the community decline cycle outlined by broken windows thesis can be visualized as shown in Fig. 1.

As such, a main thrust of Wilson and Kelling's argument was that police could fight crime more effectively by dealing with disorder. If they stop disorder from accumulating and prevent neighborhoods from reaching the tipping point where they become vulnerable for criminal invasion, they can have a large impact on crime. Wilson and Kelling do not discuss what police may do in neighborhoods already past the tipping point and fully invaded by criminal behavior, but one could imply that cleaning up disorder would still play a role in restoring informal social control in such neighborhoods and helping residents take back the streets. As with any theory about crime, the broken windows thesis can be more readily understood by examining the research that influenced its formation.

Theoretical Development of the Broken Windows Thesis

While Wilson and Kelling are credited with developing the broken windows thesis, they were not the first to examine the role disorder played in communities. In the area of criminology, concern over disorder can partly be traced to early research on fear of crime. One issue that drove interest in the topic was a body of research that consistently found that fear of crime had seemingly little to do with crime. For instance, most studies found that females and the elderly reported the highest levels of fear of crime, yet the National Crime Surveys conducted by the Bureau of Justice Statistics consistently showed young males to have the highest rates of victimization. Thus, fear of crime did not appear to be driven by actual victimization risk, and these findings naturally led criminologists to question what was driving fear of crime if not crime itself.

A number of studies turned to disorder to explain fear of crime. For example, James Q. Wilson first noted in 1975 that people were troubled not only by crime but also by:

> The daily hassles they are confronted with on the street – street people, panhandlers, rowdy youths, or 'hey honey' hassles – and the deteriorated conditions that surround them – trash strewn alleys and vacant lots, graffiti, and deteriorated or abandoned housing – inspire concern. (p. 66)

Similarly, Garofalo and Laub (1978) stated that "...what has been measured in research as the 'fear of crime' is not simply fear of crime" (p. 245) and tied fear to quality of life and concern for the community. Ideas closely related to the broken windows thesis are most clearly seen in work by Hunter (1978). Hunter suggested that disorder affected both fear of crime and actual crime through a process in which disorder signaled to residents that local controls had failed and caused them to become personally at risk of victimization. He suggested that this would increase crime and further increase fear. His work can easily be seen as an early version of the broken windows thesis.

Finally, Wilson and Kelling's ideas were greatly influenced by a social-psychological experiment conducted by Stanford psychologist Philip Zimbardo in 1969, as indicated by the detailed discussion of the experiment in their broken windows article. Zimbardo abandoned a car with its hood up in two places – the Bronx in New York City and on the Stanford Campus in Palo Alto, California. The car in the Bronx was vandalized within 10 min, and within 24 h everything of value was removed. The car in Palo Alto, however, was not touched for more than a week. Zimbardo then smashed the windshield with a sledgehammer, and from that point on, people passing by saw the activity and the damaged car and joined in the destruction. This is where the broken windows metaphor came from for Wilson and Kelling and, combined with the above work on fear of crime and disorder, formed the basis for their ideas that untended disorder is what eventually leads to a neighborhood becoming crime plagued. Just like the broken window on the car in Palo Alto invited more vandalism, untended disorder is a visual cue in a community which invites more disorder and eventually more serious crime.

While the theoretical underpinnings of the broken windows thesis can clearly be seen in this early work on fear of crime and the Zimbardo experiment, the law enforcement portion of Wilson and Kelling's ideas was directly influenced by earlier research they had conducted on policing. Most notably, this is seen in the work of Wilson and Boland (1978) who noted that aggressive policing can reduce crime. Their main point was that the police may reduce crime not by how many are on patrol but rather by *what* they do while on patrol. They suggested that if police are aggressive in arresting criminals, they can have more of an impact on crime. Their study used traffic citations for as a proxy for aggressive policing and found a negative relationship between police aggressiveness and crime rates.

George Kelling was also greatly influenced by his earlier work on policing. In particular, the broken windows article repeatedly makes reference to his experience working on an evaluation of foot patrol in Newark, New Jersey (Kelling et al. 1981). In the Broken Windows article, Kelling elaborates on how his experiences on this study showed him that by being active in the community, the police could maintain order and make residents feel better, which could have positive impacts even if the police strategy was not directly reducing crime.

The above discussion lays out the theoretical foundation for Wilson and Kelling's broken windows thesis. It is clear how work both on causes of fear of crime and studies of specific police practices laid the groundwork for their ideas that police could fight crime by tackling smaller problems – the disorder in a community that made residents fearful and uneasy. In the quarter century since its inception, the broken windows thesis has had a profound impact on policing and continues to be a subject of theoretical debate in scholarly journals. The sections below provide an overview of evidence on the validity of the broken windows thesis itself.

Testing the Broken Windows Thesis

For a theory that has had such a large impact on police practice, the broken windows thesis has received relatively little research attention in terms of testing the theoretical propositions the broken windows thesis itself. As this section will show, there are a number of studies which have examined individual propositions of the broken windows thesis (often with no direct connection to or mention of broken windows), but few studies have set out to explicitly and directly test the entire broken windows thesis.

Looking at the broken windows thesis outlined above, one can see that there are three main theoretical propositions that make up the path from disorder to more serious crime. First, disorder must increase fear of crime among residents. In particular, perceptions of disorder are most relevant here. If residents are unaware of disorder on their block, it is unlikely to impact their levels of fear. Secondly, residents who become fearful must withdraw from the community, thus lowering informal social controls/ collective efficacy. Third, crime must increase in the area in response to this withdrawal as criminals perceive the area to be a place to commit crimes with relative impunity.

Does Disorder Cause Fear of Crime?

Wilson and Kelling (1982) suggested that disorder's first effect was increasing resident fear of crime. A number of studies have explored this issue by testing the connection between disorder and fear of crime, and a selection of these studies are reviewed here. A neighborhood-level study by Covington and Taylor (1991) found that both objectively measured disorder, as well as survey measured perceptions of disorder, were related to fear of crime. Moreover, their perceived disorder measure was the dominant effect in the model. Another study involving a panel interview of residents in Baltimore found that between the two surveys, those whose perceptions of disorder increased reported less satisfaction with the block they lived on, as well as showing greater increases in feelings of vulnerability and fear of crime (Robinson et at. 2003). Also of note, some research has suggested that social disorder has a stronger relationship with fear of crime than physical disorder (LaGrange et al. 1992).

More recent research has also supported the link between disorder and fear. One study tested a variation of the broken windows/decline hypothesis similar to the model in Fig. 1 using a sample of neighborhoods from three waves of the British Crime Survey (Markowitz at al. 2001). Their findings suggested the model was correct. Of particular interest to the current research, they found that "the dominant effect in the cycle is the effect of disorder on fear" (p. 310). The authors thus concluded that their findings were consistent with the broken windows/decline hypothesis as they show that disorder may increase crime indirectly by increasing levels of fear which in turn reduce the level of social cohesion which may then lead to crime.

On the other hand, some research has suggested that the relationship between disorder and fear may not be straightforward. Taylor and Shumaker (1990) found a relationship between disorder and fear of crime, but noted that it may not be linear. Their findings suggested a quadratic relationship where disorder does increase fear, but as disorder gets higher, the strength of the relationship weakens. They note that perhaps people in high disorder areas get somewhat inoculated to disorder, just as people living in areas plagued by natural disasters may fear earthquakes or hurricanes less than people from other areas. In short, some people may be accustomed to disorder and not bothered by it as much. Additionally, a few studies have not found clear evidence of a link between disorder and fear of crime. For instance, Taylor's (2001) well-known study in Baltimore found that incivilities measured in 1981 and 1982 were not related to changes in fear of crime in these neighborhoods between 1982 and 1994.

Overall, the studies reviewed above illustrate that the link between disorder and fear of crime has generally been supported in past research, with a few exceptions. Thus, there is reason to believe that Wilson and Kelling (1982) and those before them (e.g., Garafalo and Laub 1978; Hunter 1978; Wilson 1975) were correct in asserting that signs of disorder make people feel uneasy and even fearful of being victimized by crime in their neighborhood. Moving on to the next proposition, what does the research say about the link between fear of crime and withdrawal from the community?

Does Fear of Crime Lead to Withdrawal?

As with the link between disorder and fear of crime, there has been work examining whether fear of crime leads to community withdrawal. Again, most of this work was done without explicit reference to broken windows, but there are studies of the topic in other areas such as tests of variables related to social control or collective efficacy (Sampson et al. 1997). A review of these studies is relevant to understanding the evidence on the broken windows thesis as informal social control/collective efficacy can be viewed as analogous to Wilson and Kelling's (1982) discussion of the ability (or inability) of a community to regulate behavior.

A study by Garafalo (1981) found that fear of crime was related to a host of social outcomes related to collective efficacy such as heightened interpersonal distrust, withdrawal of support for formal control agencies and decreased social interaction. The Markowitz et al. (2001) study cited above which found a link between disorder and fear also found that fear in turn reduced social cohesion. Thus, as with the link between disorder and fear, a review of the research on the link between fear and withdrawal/collective efficacy (and related constructs) is generally supported in past research. The majority of the studies reviewed find that increases in fear of crime are tied to reductions in things such as neighborhood cohesion, social involvement, and interpersonal trust. Moving to the final theoretical proposition, what does the research say about the impacts of withdrawal and reductions in social control/collective efficacy on crime?

Does Withdrawal/Weakening of Social Controls Lead to Crime?

This link of the broken windows thesis really gets at the notion that reductions in informal social lead to increases in crime. This topic has received a fair amount of research attention in work examining the impact of social control/collective efficacy on crime, as well as work on social disorganization theory. This brief review will focus on research on collective efficacy, as that is the specific theory of informal social control that is at the center of current debates over the broken windows thesis. Collective efficacy refers to the notion that crime can be controlled in a community where there is a high level of social cohesion/social ties and willingness to intervene for the common good (Sampson et al. 1997). Places with higher levels of collective efficacy are more able to effectively regulate behavior and thus have lower crime rates.

In general, the collective efficacy literature has been supportive of this inverse relationship between collective efficacy and crime rates. For instance, in the work first advancing the concept, Sampson et al. (1997) found that collective efficacy was negatively associated with violence after controlling for individual factors and prior violence. A study by Sabol et al. (2004) found that collective efficacy reduced youth violence, child maltreatment, and intimate partner violence. Ford and Beveridge (2004), in looking at illegal drug sales, noted that collective efficacy can prevent drug sales not only by resisting undesirable factors (like drug dealers) but also through increasing the capacity to attract positive factors (i.e., legal businesses).

However, a study by Morenoff et al. (2001) found that collective action for social control was negatively related to homicide, but found little support that social ties mattered. Thus, it appears to be important that communities have shared goals and that residents are willing to intervene for the common good, but that social cohesion/social ties may not necessarily be a key component of collective efficacy. What matters is whether a community can work together to solve crime and disorder problems. Residents may not have to have strong social ties/cohesion to share mutual goals for the community and be willing to intervene to achieve these goals. These findings fit with the broken windows thesis as Wilson and Kelling's (1982) work focused on the notion that disorder led people to become fearful and less willing to intervene in the community. Social cohesion was not explicitly focused on in their work, as they focused more on the notion of people being too afraid to intervene and/or even moving away in response to rising untended disorder in their neighborhood.

In short, work looking at the relationship between collective efficacy and crime has generally been supportive and builds on the social disorganization literature which has a long history of showing that a community's ability to exert informal social control is a powerful predictor of crime rates. As such, the final link of broken windows has been supported with a long history of criminological work which suggests that withdrawal and other factors that reduce informal social control, collective efficacy, etc. are likely to lead to increases in crime.

The Current Empirical Status of the Broken Windows Thesis

The above review shows is that there is a fair amount of support for the individual theoretical propositions behind the broken windows thesis. However, this support comes largely from a body of unrelated studies, few of which examine more than one step of the broken windows model, and even for the individual propositions, there are still mixed findings at times. In turn, there is still a great deal of debate about the validity of the broken windows thesis. While some authors have strongly advocated the model and/or broken windows policing (e.g., Bratton and Kelling 2006; Kelling and Coles 1996; Kelling and Sousa 2001), some recent research has challenged the broken windows thesis by suggesting that the relationship between disorder and crime is spurious and explained by collective efficacy (see Sampson and Raudenbush 1999).

A Direct or Indirect Relationship Between Disorder and Crime?

A study by Sampson and Raudenbush (1999) sparked considerable debate over the broken windows thesis. Sampson and Raudenbush set out to test the broken windows thesis and collective efficacy theory using data collected as part of the Project on Human Development in Chicago Neighborhoods (PHDCN). These data included systematic social observations of physical and social disorder which were collected by videotaping segments with video cameras mounted to a vehicle which drove through study areas at 5 mph recording both sides of the streets. The tapes were then coded by members of the PHDCN research team to record levels of social and physical disorder on the study blocks. Collective efficacy and other individual-level variables were obtained through a resident survey, while crime was measured through official police data (and survey measured victimization in some models).

Sampson and Raudenbush used weighted regression analysis and measured variable path analysis in their study and found that disorder was positively related to crime. However, when they added collective efficacy to the model, they found that the relationship between disorder and crime disappeared – except for robbery where it remained significant. As such they concluded that the broken windows thesis was not supported as disorder and crime were only spuriously related – their results showed both to be a result of low collective efficacy. They argued that disorder and crime were simply different degrees of the same problem with the same underlying cause, rather than being causally related as the broken windows thesis suggests. As such, they concluded "....that neighborhoods high in disorder do not have higher crime rates in general than neighborhoods low on disorder once collective efficacy and structural antecedents are held constant" (p. 638).

Other studies have also examined this notion of a direct link between disorder and crime and produced findings that challenged the broken windows thesis. A study by Taylor (2001) found that disorder measured in 1981 was not strongly related to crime in 1994 after controlling for initial neighborhood structure. A more recent study found that neither collective efficacy nor disorder was a sufficient explanation for crime (St. Jean 2007). While high collective efficacy and low social disorder (physical disorder did not matter) explained low crime rates, places with low levels of collective efficacy and/or high levels of social disorder were found to be about equally likely to have high or low crime rates. Finally, in another study using 16 years of census block-level data collected in Seattle, Yang (2007, 2010) found that the trends between violent crime and disorder were correlated. The direction of causation, however, was opposed to what was suggested by the broken windows thesis. The results from Granger causality tests generally showed no causal relationship between disorder and violence, and in a few places, the causality appeared to run from violent crime to disorder.

On the other hand, a recent social-psychology experiment conducted in the Netherlands did find some support for a direct link between the presence of disorder and minor criminal behavior (Keizer et al. 2008). The study manipulated the levels of disorder in places (graffiti, trash, etc.) and tested whether it had an impact on behavior among passersby. Most of the tests looked at the spread of disorder – for instance, one of the experiments which was part of this study found that the presence of graffiti in a researcher-manipulated area led to a higher likelihood of people littering (throwing a flyer attached to their bicycles on the ground) compared with another area where the researches did not place graffiti (but also attached flyers to bicycles). However, one part other experiment examined minor theft. This experiment involved placing an envelope with a small amount of money clearly visible sticking of a mailbox. In the control setting, there was no graffiti or trash around the mailbox, and in the two experimental conditions, there was graffiti on the mailbox in one setting and trash strewn around it in the other. In both cases, passersby in the conditions with disorder present were significantly more likely to steal the envelope than those in the control condition. Of course, it is impossible to say from this study if the presence of disorder could lead to more serious crimes as the broken windows thesis suggests, or if it is just another example of how disorder spreads (which was the main focus of the study) given the very minor nature of the crime in question.

While the methodologies of these studies are sound, and the negative results of all but the Netherlands study seem to be a large challenge to the tenets of the broken windows thesis, this body of work has been criticized as an unfair test of broken windows. The challenge to this work comes from its assertion that crime and disorder are *directly* related – that disorder directly leads to crime. Some scholars, including Kelling himself, have denied that broken windows ever implied a direct relationship between disorder and crime (Bratton and Kelling 2006; Gault and Silver 2008; Xu et al. 2005), arguing that it has always posited an *indirect* relationship between disorder and crime.

A reading of Wilson and Kelling's (1982) article would seem to support Bratton and Kelling's (2006) assertion that they have always posited an indirect relationship between crime and disorder. For instance, Wilson and Kelling stated "…at the community level, disorder and crime are usually inextricably linked, in a kind of developmental sequence" (p. 31). A developmental sequence does not imply a direct relationship. In fact, going back as far as Zimbardo's 1969 experiment, the broken windows thesis has been a social-psychological theory. It has never suggested that disorder in a community directly causes crime. Rather it has always posited that perceptions of disorder, created through visual cues of untended disorder in a community, increase fear and lead to residential withdrawal which leaves communities vulnerable to criminal invasion. Studies that look for a direct relationship between observed disorder and crime ignore the social-psychological foundation of the broken windows thesis and thus are not complete tests of the theory.

Given this flaw, the impact of Sampson and Raudenbush's study, and other work testing for a direct relationship between disorder and crime, is challenged. If Sampson and Raudenbush had specified their model in accordance with the propositions and social-psychological foundation of the broken windows thesis, their data may have actually supported the broken windows thesis. A correct specification would test whether perceptions of disorder were positively related to levels of fear and then test whether fear in turn was negatively related to collective efficacy (the informal social control portion of the broken windows thesis) and whether collective efficacy was related to crime. Given that they found collective efficacy inversely related to crime, a finding that disorder, through increased fear of crime, reduced collective efficacy would be supportive of the broken windows' notion that disorder erodes informal social controls and leads to increases in crime.

Sampson and Raudenbush (1999) themselves even made some statements that are supportive of a true reading of broken windows. For instance, in their conclusion, they state that "[e]radicating disorder *may* indirectly reduce crime by stabilizing neighborhoods. . . ." (p. 648). This is likely based on an analysis presented in passing on page 636 in which they found that "[t]he results indicated that observed disorder increases perceived disorder, which in turn reduces collective efficacy. The significant reciprocal relationship between violence and collective efficacy nonetheless remained intact. . ...." Thus, if Sampson and Raudenbush had set up their study to test the *indirect* link between perceived disorder and crime as outlined above, their conclusions likely would have been very different, as others have noted (Gault and Silver 2008; Xu et al. 2005).

Some recent research has aimed to address this issue by testing models that are more faithful to Wilson and Kelling's conceptualization of the broken windows thesis. Xu et al. (2005), in their study of community policing, found that perceived disorder had strong direct and indirect impacts on perceived crime after controlling for collective efficacy (Xu et al. 2005) and thus challenge the assertion made by Sampson and Raudenbush. However, Xu et al. also did not model the relations in the specific order suggested by the broken windows thesis and thus limited the ability of their study to test the theoretical propositions behind the broken windows thesis. Specifically, collective efficacy is only included as an exogenous variable in their structural equation models. Collective efficacy (along with community policing variables) is said to affect disorder and crime, which in turn affect fear of crime and perceptions of quality of life which then affect satisfaction with the police. As outlined above, a true test of the broken windows thesis as outlined by Wilson and Kelling (1982) would test the impact of disorder on fear of crime, which should in turn affect collective efficacy and crime.

A recent study aimed to shed light on this issue by using structural equation modeling to test the relationships between perceived disorder, fear of crime, collective efficacy, and perceived crime specified by the broken windows thesis (see Fig. 1 above) using structural equation modeling (Hinkle 2009). This study used perceptual data measured through resident surveys and found support for both direct and indirect pathways between perceived disorder and perceived crime in the best fitting model (see Fig. 2 below). While caution is needed in interpreting these findings due to the study using perceptual measures of disorder and crime from the same survey, the results nonetheless suggest that theories dealing with the relationship between disorder and crime may be overly simplistic if they only consider either direct or indirect pathways between these two variables. Additionally, the final model from this study challenged the centrality of fear of crime to the broken windows thesis. Once disorder was allowed to have direct impacts on collective efficacy and crime, the relationship between fear and collective efficacy hypothesized by the broken windows thesis was no longer statistically significant. The implications of these findings are elaborated on in the concluding section below.

The Future of the Broken Window Thesis

This review of the broken windows thesis essentially points out two key issues about the topic. First, it is an idea that has had a very

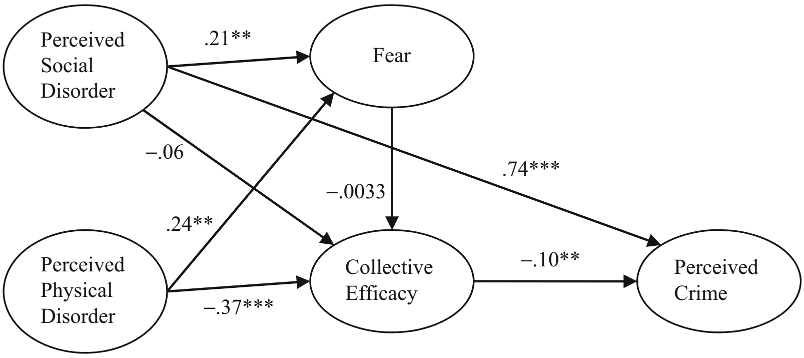

* =p <.05; **p <.01; ***p <.001; n=773; df=405
SRMR=.066; RMSEA=.047; CFI=.97

Broken Windows Thesis, Fig. 2 Testing the broken windows thesis (Final model from Hinkle 2009)

large impact on policing policy and practice. Second, despite this impact, the thesis itself has not received much direct empirical research attention. While there is a good deal of evidence on the individual relationships between disorder, fear of crime, collective efficacy/informal social control, and crime, there are hardly any studies which examine all of these variables simultaneously in one model with the relationships tested in a manner consistent with Wilson and Kelling's (1982) hypotheses.

As such, we still know little about the validity of the idea behind the policing model, and thus, debates over the thesis have been a hot topic in academic circles in recent years as outlined above. Given this dearth of empirical research explicitly testing the broken windows thesis, the future of the idea is very much up in the air. What is clear is that we need a substantial amount of empirical research testing the broken windows thesis. The study by Hinkle (2009) outlined above is one example and can serve as a starting point for a discussion of where future research on the broken windows thesis should begin.

A first takeaway point from that study is that future research should ideally consider both direct and indirect pathways from disorder to crime when testing the broken windows thesis. Future studies would be aided by including hard measures of crime in addition to perceptual measures. A second takeaway point was that fear of crime did not have the expected impact on collective efficacy once disorder was allowed to have direct impacts on efficacy and crime. While this is only one study, and this finding should thus be interpreted with caution, it does challenge the centrality of fear in the broken windows framework and suggests the need for theoretical elaboration on Wilson and Kelling's model.

For instance, while the results from this study challenge the role of fear of crime, it may also be unlikely that disorder and collective efficacy are truly directly related as the model in Fig. 2 shows. It may be more likely that there are other intervening variable besides fear of crime which mediate the impact of disorder on levels of informal social control. Maybe instead of untended disorder increasing fear and scaring off residents who care, it is more a matter of residents getting annoyed or frustrated and thus they give up intervening or even move away. Or perhaps over time some residents simply become used to the disorder and are not bothered by it as Taylor and Shumaker (1990) argued – in which case residents may become apathetic over time and thus stop intervening for the common good. Future theoretical work should consider these types of issues and attempt to elaborate on the broken windows thesis. It is up to such work to put forth more elaborate models based on theoretical propositions which clearly outline the relationships between disorder, crime, and whatever mediating variables are hypothesized to lie between these two constructs. These models can then be tested with SEM and other techniques to gauge whether they are supported empirically.

Finally, a key in sorting out these debates over the validity of the broken windows thesis will be testing the idea with longitudinal data. Most of the studies discussed above have tested the model with cross-sectional data. While such studies offer valuable insight into the broken windows model, it is nonetheless a theory of neighborhood decline that occurs over time. As such, it is best tested using data that covers a number of years. There have been some longitudinal tests of the theory by researchers such as Taylor (2001) and Yang (2007, 2010), and these tests tended to challenge the relationship between disorder and crime. However, these studies only tested for direct relationships between disorder and crime and thus were not full tests of the broken windows thesis.

Related Entries

- ▶ Defining Disorder
- ▶ Discriminant Validity of Disorder and Crime
- ▶ Disorder: Observational and Perceptual Measures
- ▶ Order Maintenance Policing

Recommended Reading and References

Bratton WJ, Kelling GL (2006) There are no cracks in the broken windows. Natl Rev Online, February 28. Retrieved from http://www.nationalreview.com/comment/bratton_kelling200602281015.asp

Covington J, Taylor RB (1991) Fear of crime in urban residential neighborhoods: implication of between- and within-neighborhood sources for current models. Sociol Q 32(2):231–249

Dennis N, Mallon R (1998) Confident policing in Hartlepool. In: Dennis N (ed) Zero tolerance: policing a free society. Institute of Economic Affairs Health and Welfare Unit, London, pp 62–87

Ford JM, Beveridge AA (2004) "Bad" neighborhoods, fast food, "sleazy" businesses, and drug dealers: relationships between the location of licit and illicit businesses in the urban environment. J Drug Issues 34(1):51–76

Garafalo J (1981) The fear of crime: causes and consequences. J Crim Law Criminol 72(2):839–857

Garofalo J, Laub JH (1978) The fear of crime: broadening our perspective. Victimology 3(3–4):242–253

Gault M, Silver E (2008) Spuriousness or mediation? Broken windows according to Sampson and Raudenbush (1999). J Crim Just 36(3):240–243

Hinkle JC (2009) Making sense of broken windows: the relationship between perceptions of disorder, fear of crime, collective efficacy and perceptions of crime. Unpublished doctoral dissertation, University of Maryland, College Park

Hunter A (1978) Symbols of incivility: Social disorder and fear of crime in urban neighborhoods. Paper presented at the annual meeting of the American Society of Criminology, Dallas, TX

Keizer K, Lindenberg S, Steg L (2008) The spreading of disorder. Science 12:1681–1685

Kelling GL, Coles C (1996) Fixing broken windows: restoring order and reducing crime in American cities. Free Press, New York

Kelling GL, Sousa WH (2001) Do police matter? An analysis of the impact of New York City's police reforms (Civic report No., 22). Manhattan Institute for Policy Research, New York

Kelling GL, Pate T, Ferrara A, Utne M, Brown CE (1981) The newark foot patrol experiment. The Police Foundation, Washington, DC

LaGrange RL, Ferraro KP, Supancic M (1992) Perceived risk and fear of crime: role of social and physical incivilities. J Res Crime Delinq 29(3):311–334

Markowitz FE, Bellair PE, Liska AE, Liu J (2001) Extending social disorganization theory: modeling the relationships between cohesion, disorder and fear. Criminology 39(2):293–320

Morenoff JD, Sampson RJ, Raudenbush SW (2001) Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. Criminology 39(3):517–560

Robinson JB, Lawton BA, Taylor RB, Perkins DD (2003) Multilevel longitudinal impacts of incivilities: fear of crime, expected safety, and block satisfaction. J Quant Criminol 19(3):237–274

Sabol WJ, Coulton CJ, Korbin JE (2004) Building community capacity for violence prevention. J Interpers Violence 19(3):322–340

Sampson RJ, Raudenbush SW (1999) Systematic social observation of public spaces: a new look at disorder in urban neighborhoods. Am J Sociol 105(3):603–651

Sampson RJ, Raudenbush SW, Earls F (1997) Neighborhoods and violent crime: a multilevel study of collective efficacy. Science 277(5328):918–924

St. Jean PKB (2007) Pockets of crime: broken windows, collective efficacy, and the criminal point of view. The University of Chicago Press, Chicago

Taylor RB (2001) Breaking away from broken windows: Baltimore neighborhoods and the nationwide fight against crime, grime, fear, and decline. Westview Press, Boulder

Taylor RB, Shumaker SA (1990) Local crime as a natural hazard: implications for understanding the relationship between disorder and fear of crime. Am J Community Psychol 18(5):619–641

Wilson JQ (1975) Thinking about crime. Basic Books, New York

Wilson JQ, Boland B (1978) The effect of the police on crime. Law Soc Rev 12:367–390

Wilson JQ, Kelling GL (1982) Broken windows: the police and neighborhood safety. Atl Mon 211:29–38

Xu Y, Fiedler ML, Flaming KH (2005) Discovering the impact of community policing: the broken windows thesis, collective efficacy and citizens' judgment. J Res Crime Delinq 42(2):147–186

Yang S (2007) Causal or merely co-existing: a longitudinal study of disorder and violence at places. Unpublished doctoral dissertation, University of Maryland, College Park

Yang S (2010) Assessing the spatial-temporal relationship between disorder and violence. J Quant Criminol 26(1):139–163

Zimbardo PG (1969) The human choice: Individuation, reason, and order versus deindividuation, impulse, and chaos. Nebr Sym Motiv 17:237–307

"Broken Windows" Policing

► Order Maintenance Policing

Brutality

► Police and the Excessive Use of Force

Bullying

Maria Sapouna
School of Social Sciences, University of the West of Scotland, Hamilton, UK

Overview

The study of bullying has flourished in recent years, reflecting the growing recognition of the negative long-term effects this type of aggressive behavior can have on victims and bullies. There is now a considerable body of cross-sectional and longitudinal research that has informed our understanding of what bullying is; who it is most likely to affect; what consequences it can have for victims, bullies, and the peer group; and what can be done to reduce its occurrence. This entry attempts to shed some light on these questions by reviewing key empirical research conducted in this field over the last three decades. The piecemeal and theoretically unsystematic way in which much of this evidence has been produced, however, makes it difficult to reach a clear and coherent conclusion on why bullying happens and how it can be effectively prevented. Readers of this review should be aware of these research limitations, which, to some extent, reflect the complexity of the bullying phenomenon that precludes straightforward, one-size-fits-all conclusions.

Bullying Defined

Bullying has received worldwide attention in the last 30 years as a form of aggressive behavior that can have a significant negative impact on the physical, emotional, and academic development of victims. The first major contribution to the academic study of bullying was made by Dan Olweus, who wrote the first scholarly book in English to deal with bullying. The book was written in response to the suicide of three bullied boys in Norway and reported a high prevalence of school bullying (20 % of Norwegian children reported having some involvement) as well as discussed the success of the world's first bullying prevention program (Olweus 1993). Olweus' work opened the way for an explosion of research on bullying, which expanded from an initial interest in schools to include broader contexts such as the workplace, prisons, and sibling relationships. While much of this work is of interest, showing that bullying has the potential to affect a significant proportion of the population, this review focuses on school bullying, as this is the area that has attracted the most research interest to date.

The international literature is repleted with definitions of school bullying, most of which seem to accept that bullying is any type of negative action intended to cause distress or harm that is repeated and targeted against individuals who cannot defend themselves. When research on bullying started in the 1980s, bullying was perceived to comprise only episodes of physical or verbal aggression where the victim was physically attacked or called names. In recent years, the definition of bullying has broadened to include other forms of aggression that are relational in nature and aim to damage the victim's peer relationships and their social status such as spreading of malicious gossip and social exclusion. Fighting between people of approximately equal strength, a one-time attack, or a good-natured teasing and play fighting are not counted as bullying.

The advent and widespread use of electronic means of communication such as mobile phones and the Internet has made it easier to bully anonymously, through the use of pseudonyms and temporary accounts, at any time and in any place involving a wide audience. This development has meant that the definition of bullying has had to be expanded to account for what the literature refers to as "cyber-bullying" or "electronic bullying." A nationally representative survey of 7,508 adolescents in the United States in 2005 found that 8.3 % had bullied others and 9.8 % had been bullied electronically at least once in the last 2 months (Wang et al. 2009). In the same year in England and Wales, a survey of pupils aged 11–16 found that 22 % had been cyber-bullied at least once or twice in the last couple months (Smith et al. 2008). The most common form of cyber-bullying internationally is sending threatening and/or nasty text messages.

Prevalence and Continuity

National Variation

There are large variations across countries in the prevalence of bullying perpetration and victimization. In an international survey of health-related symptoms among school-aged children, the percentage of students who reported being frequently bullied during the current term ranged from a low of 5 % to 10 % in some countries to a high of 40 % in others (Due et al. 2005). The prevalence of bullies in primary school ranges, in most countries, between 7 % and 12 % and remains at those levels in secondary school (around 10 %). It is unclear whether these differences in prevalence reflect genuinely different levels of engagement in bullying among countries or, at least partly, result from different meanings of the term "bullying" in different countries and differences in methodologies and samples used.

An example of why valid comparisons between countries are not possible is Portugal where the bullying rate is high compared to other countries. Berger (2007) in her analysis found that one detail of educational policy in Portugal may account, among other things, for this higher rate of bullying. In Portuguese schools, children are asked to repeat sixth grade unless they pass a rigorous test. This practice results in at least 10 % of all sixth graders (more often boys) to be held back 2 years or more, and these older, bigger children are almost twice as likely to bully compared to the class average. This suggests that the difference in prevalence rates between countries may be, at least partly, accounted for by external factors including national differences in school policies and environments but also differences in the methodologies used (self-reports vs. peer and/or teacher reports), students' differing levels of cognitive ability, cultural differences in reporting, and different meanings of the term "bullying" in different countries.

The Importance of Age

Despite variations in prevalence, it is a universal finding that bullying victimization is more frequent among younger children and steadily declines with age. A range of explanations have been put forward to explain these age differences (Smith et al. 1999a, b). Compared to older children, younger children are less likely to have developed the appropriate skills and coping strategies to deal effectively with bullies and avert further victimization. Younger children are also less likely to refrain from bullying others due to socialization pressure. Finally, there is evidence that younger students adopt a more inclusive definition of bullying when responding to prevalence surveys, and this may, at least partly, account for the higher reported frequency of bullying victimization in primary school. For example, younger pupils might find it more difficult to distinguish between bullying and fighting, broadening the use of the term bullying to include aggressive behaviors that involve no imbalance of power. Within the general trend of decreasing bullying victimization over time, researchers have observed an abrupt increase in bullying during the transition from primary to secondary school which may reflect some students' attempts to establish dominance hierarchies in the new school environment. Relational forms of bullying take precedence over physical modes of attack as children grow older and their social skills improve.

Stability of Bullying Roles

There is some controversy in the literature as to the stability of bullying victimization in primary school. Some studies have reported that bullying victimization is relatively stable over a period of up to 4 years in primary school and often continues in secondary school. Other studies have found that only a relatively small proportion of children (around 4–5 %) are victimized repeatedly over time in primary school.

In secondary school, the stability of both bully and victim roles is considerably higher than in primary school according to teacher, peer, and self-reports. It is estimated that two out of three male bullies remain in their role over a 1-year period. Despite the moderate to high stability of the victim and bully roles in secondary school, prevalence rates are lower than in primary school. This suggests that a small number of victims are targeted consistently and systematically in secondary school.

Stability in bullying victimization has been explained in two ways. Firstly, it has been observed that victims select social environments that reinforce the risk of victimization, for example, they are more likely to have friends who are less accepted by the peer group and often victimized themselves. Secondly, victims often lack the social skills to break through in new environments, and this increases the risk that they are labeled as victims and locked in that role over a long period of time. It is important, therefore, to acknowledge that although for some children bullying victimization will be situational, for others it will develop into a trait.

Gender Differences

The view that males are more likely to bully and be bullied than females has been dismissed in recent years following a better understanding about the different forms aggressive behavior such as bullying can take. Although males are more likely to engage in physical forms of bullying such as pushing and hitting, females are, according to some studies, more adept at employing relational forms of aggression (e.g., social exclusion, spreading of nasty rumors) against their victims especially during adolescence. No consistent gender differences have been identified in the use of verbal bullying (e.g., calling names, nasty teasing). This suggests that overall gender differences are not as pronounced as originally thought and that bullying is not a male problem.

Characteristics of Children and Adolescents Involved in Bullying

The Bully

There is some controversy in the literature about the profile of bullies. Initially, studies described children who bullied others as insecure, anxious individuals who have low self-esteem, are unpopular among their classmates, and use aggressive strategies to resolve conflicts. This stereotype was later disputed by research that suggested bullies are socially competent and have superior theory of mind skills (i.e., awareness of others' mental functions and states) and good levels of social intelligence, knowing how to attain goals without damaging their reputation. Linked to this, there is also debate concerning whether bullies lack empathic skills. Some research suggests that bullies understand the emotions of others but do not share them. The inconsistencies across studies may be, at least partly, due to different definitions of bully status and different methodologies employed. Studies which have distinguished between "pure" bullies and bully/victims have revealed that "pure" bullies have few conduct problems, perform well at school, are popular among their classmates, and do not suffer from physical and psychosomatic health problems.

The Victim

There is more consensus on the profile of "pure" victims. Research has identified that "pure" victims exhibit elevated levels of depression and anxiety, low self-esteem, and poor social skills. Hawker and Boulton's (2000) meta-analysis found that peer victimization is more strongly concurrently associated with depression than with anxiety, loneliness, or self-esteem. Another meta-analysis by Card (2003) found that the strongest correlates of the victimization experience are low self-concept, low physical strength, low school enjoyment, poor social skills, and high internalizing and externalizing problems. It was unclear from these reviews of cross-sectional studies, however, whether internalizing problems lead to victimization or vice versa.

The recent body of longitudinal research on bullying and peer victimization more widely suggests that the relationship between internalizing problems such as depression, anxiety and loneliness, and victimization is more likely to be reciprocal, that is, internalizing problems contribute to victimization and vice versa. A meta-analysis of 18 longitudinal studies examining associations between peer victimization and internalizing problems in children and adolescents concluded that internalizing problems both precede and follow peer victimization experiences (Reijntjes et al. 2011). It is worth noting, however, that the path from psychological maladjustment to victimization has not been replicated in all studies. For instance, Bond et al. (2001) found no support for the hypothesis that emotional maladjustment invites victimization.

Recent work suggests that bullying might arise out of early cognitive deficits, including language problems, imperfect causal understanding, and poor inhibitory control that lead to decreased competence with peers, which over time develops into bullying. Research does not support the assertion that physical appearance (e.g., wearing glasses) is a risk factor for being bullied at school. The only physical characteristic that has been associated with an increased risk of victimization is low physical size and strength. There is less evidence on how equality characteristics influence victimization. There is no consistently robust evidence to suggest that ethnic minority children are more at risk of being bullied at school. Sexual orientation has rarely been investigated in longitudinal studies as a possible risk factor of bullying victimization, but there is some, mainly qualitative, evidence of sexual minorities being targeted in secondary schools. There is stronger evidence that children with disabilities are particularly vulnerable to victimization in mainstream settings, although it might be other characteristics of disabled children that make them more vulnerable to victimization such as lack of friends rather than the disability per se.

The Bully-Victim

Olweus (1993) was the first researcher to identify a small proportion of victims of bullying that he called "provocative victims" or "bully-victims," who bully other children as well as being bullied by them. Research has identified that bully-victims are the most troubled group among children and adolescents involved in bullying incidents. This group displays the highest levels of internalizing problems, including depression, anxiety, low self-esteem, and loneliness. At the same time, they score high on externalizing problems such as aggression, impulsivity, hyperactivity, and conduct problems. Other research has shown that bully-victims display higher levels of neuroticism and psychoticism than either bullies or victims. Bully-victims use aggressive strategies to cope with stressors at school that increase the risk of further victimization and rejection from peers.

The Peer Group

Besides the traditional roles of bully, victim, and bully-victim, research has identified that all students take on a role when bullying episodes emerge. Salmivalli et al. (1996) distinguished between six different roles children can take in bullying situations: the bully (leader), the reinforcer (encourages and provides audience), the assistant (follower/helper, e.g., holds the child down), the defender (helps the victim and/or tells bullies to stop), the outsider (stays away from bullying situations), and the victim. Subsequent research established that the three roles of bully, reinforcer, and assistant are closely correlated with each other and, therefore, cannot usefully discriminate between children. In kindergarten, the three most commonly held roles are those of the bully, the victim, and the defender. Fewer students are defenders by middle school, and the majority becomes witnesses or bystanders when bullying takes place. Such passive behavior, although not directly encouraging of bullying, provides a permissive context for bullies that allows them to continue harassing their victims.

Environmental Influences

Parenting and Home Environment

There is clear evidence that parenting styles are related to bullying behavior. Studies indicate that bullies are more likely to have parents who are authoritarian and punitive, disagree more often, and are less supportive. The parents of bullies are more likely to have been bullies themselves when they were young. Victims, on the other hand, are more likely to have been reared in an overprotective family environment. Bully-victims tend to come from family backgrounds that are exposed to abuse and violence and favor the use of harsh, punitive, and restrictive discipline practices. This group reports little positive warmth in their families and more difficulties in communicating with parents.

Family characteristics are related to bullying victimization in different ways for boys and girls. Boys are more prone to victimization when the father is highly critical or absent in his relationship with his son, thus failing to provide a satisfactory role model. Victimization in boys is also associated with maternal overprotectiveness which may hinder boys' search for autonomy and independence, whereas victimization in girls is more strongly related to maternal hostility which may lead to anxiety and decreased sense of connectedness in relationships.

Very little research has examined longitudinal associations between early home environment and subsequent bullying behavior. The few studies that exist suggest a link between low emotional support and subsequent bullying behavior at school. Parents who are disagreeable, hostile, cold, or rejecting tend to have children who are at risk of becoming aggressive in the future. In a small longitudinal study, Schwartz et al. (1997) found that bully-victims at 10 years were significantly more likely than the other groups to have had experiences with harsh, disorganized, and potentially abusive home environments 5 years earlier. Mother-child interactions at 5 years were characterized by hostile, restrictive, or overly punitive parenting. They were significantly exposed to higher levels of marital conflicts and more likely to come from marginally lower socioeconomic backgrounds. Bullies were found to be exposed to adult aggression and conflicts, but not victimization by adults, and were from lower socioeconomic backgrounds. These findings need to be replicated in larger samples before any safe conclusions can be drawn.

Sibling Relationships

More recently, there has been interest in how sibling relationships affect the development of bullying behavior. There is international evidence that children who are victimized at school are more likely, compared to other groups, to be victimized by their siblings at home. Wolke and Samara (2004) found that more than half of victims of bullying by siblings (50.7 %) were also involved in bullying behavior at school compared to only 12.4 % of those not victimized by siblings, indicating a strong link between intrafamilial and extrafamilial peer relationships. Those who were both victimized at home and at school had the highest behavior problems and were the least prosocial. Similar evidence exists in relation to bullying perpetration, suggesting that those who bully at school tend to exhibit similar behaviors towards their siblings at home.

School Factors

A number of school factors have also been implicated as correlates of bullying behavior. One of the most consistent findings in the international literature is that the number and quality of friends at school is one of the strongest, if not the strongest, protective factor against bullying victimization. Having friends is not sufficient in itself to protect against victimization. For instance, when at-risk children have friends with internalizing problems, who are physically weak or who themselves are victimized, the relation of children's behavioral risk to victimization is exacerbated.

More recent work on the role of class structure and climate on bullying has shown that variations in peer structure and dominance hierarchies influence the stability of bullying victimization. For example, victims in primary school classes with a more pronounced hierarchical structure are less likely to escape their victim role compared to those in classes with less clearly marked hierarchies (Schäfer et al. 2005).

Consequences

There has been a growing interest in recent years to investigate the long-term effects of bullying involvement on children's and adolescents' social, emotional, behavioral, and academic development using longitudinal samples. The results of these studies suggest that victims and bully-victims manifest more adjustment problems than bullies. Victims and, especially, bully-victims are more likely to show elevated levels of depression, anxiety, and loneliness; perform less well academically; and display conduct problems. The only negative long-term outcome that has consistently been reported in the literature for bullies is their involvement in later offending. There is also some initial evidence that bullying perpetration is a significant risk factor of poor academic performance.

Internalizing Problems

Several cross-sectional studies have demonstrated negative associations between peer victimization and a range of internalizing problems, including loneliness and low self-esteem. A meta-analysis of 23 cross-sectional studies of the association between peer victimization and psychological maladjustment found that peer victimization was more strongly concurrently associated with depression than with anxiety, loneliness, or self-esteem (Hawker and Boulton 2000).

Over the last decade, research on bullying is increasingly reliant on longitudinal methodologies to disentangle whether victimization contributes to internalizing problems or vice versa. It has been argued, for example, that children who display internalizing behaviors (e.g., anxiety or shyness) are more at risk of being targeted by peers due to their inability to cope effectively with provocation. The majority of longitudinal studies investigating associations between peer victimization and psychological maladjustment have found evidence for both directions.

Academic Performance

There is some longitudinal evidence that bullying involvement has a negative impact on academic performance, although more studies are needed to reach a definitive conclusion. A US longitudinal study that began in 2002 with a sample of about 1,700 adolescents found that being a bully had a stronger negative effect on self-perceived academic competence over time than being a victim after controlling for demographic background variables and baseline academic competence (Ma et al. 2009). Furthermore, only bully status predicted lower self-reported grades.

Delinquency and Criminality

Despite showing fewer adjustment problems than victims and bully-victims, bullies are at an increased risk of later delinquency and criminal offending. A recent meta-analysis of studies measuring school bullying and later offending found that school bullies were 2.5 times more likely than noninvolved students to engage in offending over an 11-year follow-up period (Ttofi et al. 2011). The risk was lower when major childhood risk factors were controlled for, but remained statistically significant. The effect of bullying on later offending was especially pronounced when bullying was assessed in older children. The longitudinal association between bullying perpetration and later offending has been replicated in many countries, including Australia, Canada, and Europe.

Impact Beyond Victims

Finally, there is evidence that bullying and victimization have a negative impact not only on the individual children involved but also on bystanders. Children who witness bullying incidents report increased anxiety, less satisfaction with school, and lower academic achievement.

There is also evidence that in school classes where a lot of victimization is taking place, school satisfaction among students is low.

Interventions

Following the development of the first anti-bullying program by Dan Olweus in Norway in the 1980s, a considerable number of anti-bullying interventions have flourished around the world to reduce bullying behaviors and protect victims. These fall under four broad categories: curriculum interventions generally designed to promote an anti-bullying attitude within the classroom; whole-school programs that intervene on the school, class, and individual level and address bullying as a systemic problem; social and behavioral skills training; and peer support programs including befriending and peer mediation. A systematic review conducted in 2004 evaluated the strength of scientific evidence in support of anti-bullying programs (Vreeman and Carroll 2007). The review concluded that only a small number of anti-bullying programs have been evaluated rigorously enough to permit strong conclusions about their effectiveness.

Whole-school interventions were found to be more effective in reducing victimization and bullying than interventions that focused only on curriculum changes or social and behavioral skills training. Targeting the whole school involves actions to improve the supervision of the playground, having regular meetings between parents and teachers, setting clear guidelines for dealing with bullying, and using role-playing and other techniques to teach students about bullying. The success of whole-school interventions, relative to other stand-alone approaches, supports the view that bullying is a systemic, sociocultural phenomenon derived from factors operating at the individual, class, school, family, and community level. Hence, interventions that target only one level are unlikely to have a significant impact.

A more recent systematic review of school-based anti-bullying programs found that, overall, these programs are effective in reducing bullying perpetration and victimization by an average of 20–23 % and 17–20 %, respectively (Farrington and Ttofi 2009). The interventions that were found to be most effective were those that incorporated parent training/meetings, disciplinary methods, and videos; targeted older children; and were delivered intensively and for longer. There is less robust evidence on the effectiveness of peer support programs that include activities such as befriending, peer counseling, conflict resolution, or mediation, and a systematic review suggested their use may lead to increases in bullying victimization.

More recently, there has been a growing interest in the use of virtual learning environments to reduce bullying at schools. The basic feature of these programs is a computer-based environment that creates a highly believable learning experience for children who find themselves "present" in the situation that causes emotional distress and, as a result, learn experientially how to deal with school problems. An example of such a program is "FearNot," an intervention that was developed to help victims of bullying explore the success or otherwise of different coping strategies to dealing with bullying victimization through interactions with "virtual" victims of school bullying. The evaluation of this intervention found that the victims that received the intervention were more likely to escape victimization in the short term than victims in control schools who did not interact with the software (Sapouna et al. 2010). These results suggest that the use of virtual environments might be an engaging and useful component of whole-school anti-bullying policies that merits further testing. A key finding that emerged from this research is that interventions are more likely to be successful if they have the support of teachers and other school personnel and there is a strong commitment to reduce bullying in the school community. This is considered to be one of the reasons behind the huge success of the Olweus' prevention program that has not been replicated to date.

Future Directions

Although an abundance of knowledge has emerged in recent years regarding the correlates of bullying behavior, there is still relatively little known about the causal processes and mechanisms associated with the bully and victim status. Longitudinal studies, which track bullies and victims over time, offer one of the best chances of disentangling the antecedents of bullying perpetration and victimization from its consequences, and these should form a key part of future research in this field. Another approach which shows much promise is the cutting-edge attempt to unravel the causes of bullying behavior made by researchers investigating biological and environmental influences and the way these influences interact.

One of these studies, involving 1,116 families with 10-year-old twins, found that the tendency for children to be bullied was largely explained by genetics (73 % of variance) and less so by environmental factors that were unique to each child (Ball et al. 2008). Another study of 506 six-year-old twins found that variance in victimization was accounted for only by shared and non-shared environmental influences (29 % and 71 %, respectively) and was not related to the child's genetic predisposition (Brendgen et al. 2008). These discrepancies might be explained by differences in methodologies used, as studies drew on different informants to assess bullying victimization (mothers and peers, respectively). Although results to date have been contradictory, future breakthroughs in this area have the potential to transform radically the study of bullying.

To understand more fully how bullying behaviors develop, future research will also need to investigate in more depth how individual and classroom level factors interact to cause involvement in bullying. It is not currently understood whether the relationship between risk factors and bullying is the same across different school and class environments or the extent to which consequences of bullying and victimization are dependent on class- and school-level factors.

Finally, another area that would benefit from more attention is the investigation of resilience to bullying. Some initial evidence suggests that maternal warmth has an environmental effect in protecting children from negative outcomes associated with victimization (Bowes et al. 2010). However, we still know relatively little about the factors that promote resilience to bullying and victimization among at-risk children, and also what role bullying has to play in increasing resilience. We also know little about the factors that help victims cope better with the effects of victimization.

To conclude, what the recent flurry of research activity has highlighted is how complex the bullying phenomenon is and that, although much has been learned to date, there is clearly a great need to understand how variables describing the family, school, class, and community environment interact with individual characteristics to determine who gets bullied and who bullies others. Research should neither be blind to nor discouraged by these complexities.

Related Entries

- ► School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis

Recommended Reading and References

Ball HA, Arseneault L, Taylor A, Maughan B, Caspi A, Moffitt TE (2008) Genetic and environmental influences on victims, bullies and bully-victims in childhood. J Child Psychol Psychiatry 49(1):104–112

Berger Stassen K (2007) Update on bullying at school: science forgotten? Dev Rev 21:90–126

Bond L, Carlin J, Thomas L, Rubin K, Patton G (2001) Does bullying cause emotional problems? A prospective study of young teenagers. Br Med J 323:480–484

Bowes L, Maughan B, Caspi A, Moffitt TE, Arseneault L (2010) Families promote emotional and behavioral resilience to bullying: evidence of an environmental effect. J Child Psychol Psychiatry 51(7):809–817

Brendgen M, Boivin M, Vitaro F, Girard A, Dionne G, Perusse D (2008) Dev Psychopathol 20(2):455–471

Card N (2003) Victims of peer aggression: a meta-analytic review. Presented at the biennial meeting of the society for research on child development, Tampa, 24–27 Apr 2003

Due P, Holstein BE, Lynch J, Diderichsen F, Gabhain SN, Scheidt P, Currie C, Health Behavior in School-Aged Children Bullying Working Group (2005) Bullying and symptoms among school-aged children: international comparative cross-sectional study in 28 countries. Eur J Pub Health 15:128–132

Espelage DL, Swearer SM (2003) Bullying in American schools: a social-ecological perspective on prevention and intervention. Lawrence Erlbaum, Mahwah

Farrington DP, Ttofi MM (2009) School-based programs to reduce bullying and victimization. Campbell Syst Rev 2009:6

Hawker D, Boulton M (2000) Twenty years' research on peer victimization and psychosocial maladjustment: a meta-analytic review of cross-sectional studies. J Child Psychol Psy 41:441–455

Juvonen J, Graham S (2001) Peer harassment in school: the plight of the vulnerable and victimized. Guilford Press, New York

Ma L, Phelps E, Lerner JV, Lerner RM (2009) The development of academic competence among adolescents who bully and who are bullied. J Appl Dev Psychol 30(5):628–644

Olweus D (1993) Bullying at school:what we know and what we can do. Blackwell, Cambridge, MA

Olweus D (1994) Annotation: bullying at school: basic facts and effects of a school-based intervention program. J Child Psychol Psychiatry 35:1171–1190

Reijntjes A, Kamphuis JH, Prinzie P, Boelen PA, van der Schoot M, Telch MJ (2011) Prospective linkages between peer victimization and externalizing problems in children: a meta-analysis. Aggress Behav 37(3):215–222

Salmivalli C, Lagerspetz K, Bjorkqvist K, Osterman K, Kaukiainen A (1996) Bullying as a group process: participant roles and their relations to social status within the group. Aggress Behav 22:1–15

Sapouna M, Wolke D, Vannini N, Watson S, Woods S, Schneider W, Enz S, Hall L, Paiva A, Andre E, Dautenhahn K, Aylett R (2010) Virtual learning intervention to reduce bullying victimization in primary school: a controlled trial. J Child Psychol Psychiatry 51(1):104–112

Schäfer M, Korn S, Brodbeck FC, Wolke D, Schulz H (2005) Bullying roles in changing contexts: the stability of victim and bully roles from primary to secondary school. Int J Behav Dev 29:323–335

Schwartz D, Dodge KA, Pettit GS, Bates JE (1997) The early socialization of aggressive victims of bullying. Child Dev 68(4):665–675

Smith PK, Madsen K, Moody J (1999a) What causes the age decline in being bullied at school? Towards a developmental analysis of risks of being bullied. Educ Res 41:267–285

Smith PK, Morita Y, Junger-Tas J, Olweus D, Catalano R, Slee P (eds) (1999b) The nature of school bullying: a cross-national perspective. Routledge, London

Smith PK, Cowie H, Olafsson R, Liefooghe APD (2002) Definitions of bullying: a comparison of terms used, and age and sex differences, in a 14-country international comparison. Child Dev 73:1119–1133

Smith PK, Pepler D, Rigby K (2004) Bullying in schools: how successful can interventions be? Cambridge University Press, New York

Smith PK, Mahdavi J, Carvalho M, Fisher S, Russell S, Tippett N (2008) Cyberbullying: its nature and impact in secondary school pupils. J Child Psychol Psychiatry 49(4):376–385

Ttofi MM, Farrington DP, Losel F, Loeber R (2011) The predictive efficiency of school bullying versus later offending: a systematic/meta-analytic review of longitudinal studies. Crim Beh Ment Health 21:80–89

Vreeman RC, Carroll AE (2007) A systematic review of school-based interventions to prevent bullying. Arch Pediatr Adolesc Med 161(1):78–88

Wang J, Ionnotti RJ, Nansel TR (2009) School bullying among adolescents in the United States: physical, verbal, relational, and cyber. J Adolesc Health 45:368–375

Wolke D, Samara M (2004) Bullied by siblings: association with peer victimization and behavior problems in Israeli lower secondary school children. J Child Psychol Psychiatry 45(5):1015–1029

Bullying Prevention: Assessing Existing Meta-Evaluations

Maria M. Ttofi[1], Manuel Eisner[1] and
Catherine P. Bradshaw[2]
[1]Institute of Criminology, University of
Cambridge, Cambridge, UK
[2]Johns Hopkins Center for the Prevention of
Youth Violence, Johns Hopkins Bloomberg
School of Public Health, Baltimore, MD, USA

Synonyms

Evaluation; Intervention research; Meta-analysis; Prevention; School aggression; School programs

Overview

In this entry contemporary methodological issues in bullying prevention research are considered. The findings of extant systematic reviews and meta-analyses of school-based bullying prevention programs are assessed and integrated, with the aim of drawing clearer and more differentiated conclusions regarding their efficacy. Conclusions are drawn based on six reports, of which two included a systematic review but no meta-analysis, two included a systematic review and a meta-analysis, and two were not based on systematic searches of the literature but included some level of meta-analytic assessment. Based on a careful screening of all available meta-analytic investigations, it is concluded that bullying prevention programs are effective in reducing bullying and victimization. However, research users should be careful in identifying those intervention components and implementation procedures that are associated with a reduction in bullying. The entry concludes by identifying important challenges currently faced in the field of bullying prevention and highlighting areas for future research and implications of this work for psychologists and social scientists in general. The findings from this review of reviews are intended to inform both policy and public health practice related to bullying prevention.

Introduction

Research on school bullying has expanded considerably over the past two decades. It is now acknowledged as an established international research program, with worldwide coordinated efforts in founding a concordant methodological terminology for the explanation of this phenomenon (Smith et al. 2002). To a great extent, the strong scientific interest in bullying has been linked to the detrimental concurrent effects of school bullying for both perpetrators and victims. Notably, previous research has also established the long-term significant association of school bullying with internalizing (e.g., depression) and externalizing (e.g., offending and violence) problems (Ttofi et al. 2011a, b, 2012).

It comes as no surprise that a great deal of research has been invested in intervention efforts targeting the school environment (e.g., Waasdorp et al. 2012). Ideally, at the level of primary research, social scientists and research users should aim to draw conclusions from the most successful intervention studies always in line with explicit methodological quality standards. At the level of secondary research, a number of reviews of bullying prevention programs are also available. Ideally, these reviews should critically assess and synthesize the existing evaluation research, with the final aim of minimizing bias in the conclusions on which policymakers and practitioners are based. However, bias is not only the province of primary studies – it may also exist at the level of a review (Wilson 2009).

This is also true in the case of anti-bullying research (with a marked number of competing bullying prevention programs available), and it lays further weight on the importance of carefully assessing and utilizing *evidence-based* bullying prevention programs. It is for this reason that this entry focuses primarily on summarizing and juxtaposing what is methodologically the most high-quality evaluation research, with the final aim of drawing conclusions about what has been learned and what needs to be done next in the bullying prevention field. These aims will be accomplished by taking into account existing findings at the level of secondary research, namely, via systematic reviews and meta-analyses.

Methods

Bullying prevention could be seen as a form of early crime prevention as well as a form of early health promotion initiatives (Ttofi et al. 2012). This is true, however, only in the case of efficacious interventions. In an attempt to make suggestions about the most *scientific* and *evidence-based* bullying prevention programs, this synthesis of reviews of current evaluation studies is based on explicit inclusion criteria set in advance, namely, (a) reports presenting a systematic review of evaluations of bullying prevention programs aimed to reduce the level of school bullying perpetration and victimization (and not other outcome measures) or, ideally, (b) reports presenting both a systematic and meta-analytic review of the relevant literature. Reports that present some level of meta-analytic synthesis are also reviewed, although they may not be based on systematic searches. Reports that assessed bullying prevention programs based on narrative reviews are excluded, because this type of research carries a high risk of bias. An extensive search was conducted in order to obtain all relevant systematic and meta-analytic reviews.

Results

To date, a growing number of school-based bullying prevention programs have been developed and evaluated, but relatively few attempts have been made to synthesize the relevant rigorous outcome-based research findings (see Table 1). An extensive search of the literature revealed just six reports that met the inclusion criteria. Of the six reports, (a) two included a systematic review but no meta-analysis (i.e., Smith et al. 2004; Vreeman and Carroll 2007), (b) two included a systematic review and a meta-analysis (i.e., Merrell et al. 2008; Farrington and Ttofi 2009; Ttofi and Farrington 2011; references referring to the same project), and (c) two were not based on systematic searches of the literature but included some level of meta-analytic assessment (i.e., Baldry and Farrington 2007; Ferguson et al. 2007; only the latter carried out a full meta-analysis calculating weighted mean effect sizes for bullying perpetration).

Consistent with previous literature (Farrington 2003, 2006; Petrosino 2003), it is clear from the review of Table 1 that there is a marked degree of variation in the criteria employed and, to some extent, the quality of the currently available research reviews. Four of the included reviews of bullying prevention programs were based on systematic searches of the literature, but the intensity of the searches varied considerably. For example, the studies ranged from (a) 1 to 35 journals searched, (b) 1 to 18 databases screened, (c) 2 to 14 keywords for the online searches, and (d) the screening of 321–622 relevant documents. Only the most recent review (Farrington and Ttofi 2009; Ttofi and Farrington 2011) coded all relevant manuscripts based on a "relevance scale" in line with the aims of the systematic review.

Within each review, the timeline for carrying out searches also varied. While some authors set the beginning of searches in the 1980s and after the first Olweus Bullying Prevention Program evaluation (i.e., Farrington and Ttofi 2009; Merrell et al. 2008; Ttofi and Farrington 2011), others set their timeline for beginning searches in the 1960s, perhaps because the authors included outcome measures other than bullying (e.g., "school violence" and "peer aggression" in the Vreeman and Carroll (2007) review). It is plausible that variations in searching strategies are also affected by the type of language restrictions that the reviewers set in advance. Of the four systematic reviews, two were not transparent on this issue (i.e., Smith et al. 2004; Vreeman and Carroll 2007), and one specified language restrictions in obtaining studies written in English (Merrell et al. 2008). Only the Ttofi and Farrington (2011) systematic review was unrestricted and included studies in other languages such as German, Spanish, and Italian.

What reviewers and meta-analysts defined as "systematic" varied greatly in line with the type of inclusion criteria set in advance – and, more importantly, the extent to which these criteria were carefully followed. It is disconcerting that two out of four meta-analytic reviews computed effect sizes from uncontrolled designs despite the fact that their inclusion criteria explicitly defined "either controlled studies or quasi-experimental studies only" (i.e., Baldry and Farrington 2007; Merrell et al. 2008). Another meta-analytic review (i.e., Ferguson et al. 2007) included only evaluations of bullying prevention programs that were implemented using a controlled design. However, since the actual evaluations included in the meta-analysis are not shown in the relevant publication, it is not possible to ensure that all studies met this criterion.

All meta-analytic reviews calculated effect sizes based on evaluations of age-cohort designs, consistent with the approach employed by Olweus (2005). However, only one of these reviews (Farrington and Ttofi 2009; Ttofi and Farrington 2011) presents not only a summary effect size across all studies irrespective of their methodological design but also a summary effect size relating to each of the four types of methodological designs (i.e., an overall summary effect size and also four separate summary effect sizes specific to randomized experiments, before-after experimental-control studies, other experimental-control studies, and age-cohort designs).

What is also apparent from this synthesis of existing evaluation research is that the results vary greatly depending on the *theoretical stance* of the authors. For example, three out of six reviews included "bullying and other related antisocial behaviors" as an outcome measure (Ferguson et al. 2007; Merrell et al. 2008; Vreeman

Bullying Prevention: Assessing Existing Meta-Evaluations, Table 1 Efficacy of bullying prevention programs: A synthesis of reviews

Publication	Smith et al. (2004)	Baldry and Farrington (2007)	Ferguson et al. (2007)	Vreeman and Carroll (2007)	Merrell et al. (2008)	Farrington and Ttofi (2009); Ttofi and Farrington (2011)
Inclusion criteria/ characteristics of the evaluation synthesis	Systematic review (no meta-analysis) including studies if: Evaluation of whole-school bullying prevention programs Reports with quantitative data Studies conducted in more than one classroom (see p. 549)	No systematic review “Meta-analytic” evaluation of programs (effect sizes for each study is given, i.e., percent change, but not a weighted effect size for each study or a weighted summary effect size) Evaluations with at least 200 students included Experimental or quasi-experimental or age-cohort designs Programs specific to school bullying Studies with quantitative data	No systematic review Meta-analytic evaluation of programs aiming to reduce either bullying or violence Only studies with random assignment to experimental and control group Controlled studies only School-based programs Only manuscripts published in journals (see pp. 406–407)	Systematic review (no meta-analysis) including studies if: Controlled studies with follow-up measures School-based programs aiming (i.e., specific)to prevent bullying (see p. 79) The reviewers “extracted data from the selected articles regarding direct outcome measures of bullying, including bullying . . . aggressive behavior, violence . . .” (ibid.)	Systematic review and meta-analysis including studies if (see p. 28): School-based intervention Targeting bullying or peer-related antisocial behavior that could be interpreted as bullying Controlled studies or quasi-experimental studies with some sort of control group Published in English only No restriction to the type of manuscript (e.g., books and theses included) Quantitative data presented	Systematic review and meta-analysis including studies if (see pp. 31–32 from journal paper): Evaluation of a program specifically targeting (kindergarten to secondary) school bullying Clearly stated definition of bullying shown in the publication Measurement of bullying via self-, peer, and teacher reports or observational or other data Evaluation including an experimental and control group with baseline and follow-up measures (including age-cohort designs analyzed separately) Published and unpublished reports of any type (including theses and government reports) Quantitative information that allowed calculation of an effect size
Keywords used (N)	3 (p. 549)	4 (p. 186)	12 (p. 406)	2 (p. 79)	5 (p. 28)	14 (see p. 33)

Language restrictions	Not specified (one German report included)	Yes; only reports in English (p. 185)	Not specified (the names of the actual reports are not indicated)	Not specified (all included reports in English)	Yes; only reports in English (p. 28)	None (e.g., included reports in the English, German, Spanish, and Italian languages)
Databases searched (N)	3 (see p. 549)	1 (p. 186)	Not specified; but not a systematic review	7 (p. 79)	2 (p. 28)	18 (p. 33)
Journals searched (N)	1 (aggressive behavior)	Not specified, but not a systematic review	Not specified; but not a systematic review	Not specified	Not specified	35 (p. 33)
Screened documents (N)	323 (p. 549)	Not specified	Not clear: “42 studies relating to 45 separate observations” (p. 407)	321 (p. 79)	Not clear; but it is indicated that 40 studies were obtained and screened (p. 29)	622 reports screened and coded based on a relevance scale (p. 33 and Table 2, p. 34)
Timeline for beginning searches	Not specified	Not specified	Beginning of 1995	January 1966	Beginning of 1980	Databases searched from the inception; journals searched from beginning of 1983
Timeline for ending searches	End of December 2002	End of 2006	End of 2006	End of August 2004	End of 2004	End of May 2009 (see Campbell Report, Tables 1 and 2)
Coding of program components	Yes; 16 components coded (Tables 1, 2, 3; pp. 551–553)	Yes; general components for each program indicated (Table 1, pp. 187–188)	No	No (Table 2 shows outcome measures targeted by the program only (pp. 83–84)	Yes; general components for each program indicated (Table 3, pp. 32–36)	Yes; 20 (see Fig. 3 from Campbell Report, p. 144) Feedback/confirmation via e-mail about the way components were coded from 40 out of 44 evaluators of studies (see p. 40)
Coding of implementation features	Yes; 5 features coded (Tables 1, 2, 3; pp. 551–553)	Yes; 3 (Table 1, pp. 187–188)	Yes; 3 features (Table 1, p. 408)	No (see above)	Yes; general features of implementation indicated (Table 3, pp. 32–36)	Yes; 12 (see Fig. 2 from Campbell Report, p. 143) Feedback/confirmation via e-mail about the way features were coded from 40 out of 44 evaluators of studies (see p. 40)
Age	Primary and secondary students	Primary and secondary students	Primary and secondary school students	Primary and secondary students	Primary and secondary students	Kindergarten, primary and secondary students

(*continued*)

Bullying Prevention: Assessing Existing Meta-Evaluations, Table 1 (continued)

Publication	Smith et al. (2004)	Baldry and Farrington (2007)	Ferguson et al. (2007)	Vreeman and Carroll (2007)	Merrell et al. (2008)	Farrington and Ttofi (2009); Ttofi and Farrington (2011)
Informants	No restrictions: Student reports Peer reports Parent reports Teacher reports School records (see p. 549 and Table 1, pp. 551–552)	No restrictions: Student reports Peer reports Parent reports Teacher reports School records (see p. 185)	Not specified (see pp. 406–407)	Not specified (see p. 79 and Tables 1 and 2, pp. 80–84)	Outcome measures based only on: Student reports Teacher reports (see Table 1, p. 29)	No restrictions: Student reports Peer reports Parent reports Teacher reports School records (see pp. 31–32)
Presenting a meta-analysis with summary effect size	No; standardized effect sizes within studies and a summary weighted mean effect are not presented	No; percent change for time 1 to time 2 within each program is given, but not: (a) a standardized effect size within each study or (b) a weighted summary effect size (see Table 2, p. 189)	Yes; Fisher's z	No; standardized effect sizes within studies and a summary weighted mean effect are not presented	Yes; Cohen's d is presented (Table 4, p. 37)	Yes; standardized effect sizes given: a) separately for each program, b) for the total number of evaluations, and c) for studies grouped within each of the four methodological designs
Effect size magnitude	Not applicable	Perpetration: Reductions within programs (percent change) from −1 % to −80 % Increases within programs (percent change) from +1 % to +59 % Victimization: Reductions within programs (percent change) from −3 % to −62 % Increases within programs (percent change) from +2 % to +44 %	Perpetration: Fisher's z (bullying) = 0.12 Fisher's z (violence) = 0.13 Victimization: Not applicable (only a summary effect size across studies is presented; standardized effect sizes within each program are not shown)	Not applicable	Perpetration: d = 0.04 Victimization: d = 0.27 (only a summary effect size across studies is presented; standardized effect sizes within each program are not shown)	Perpetration: Summary OR across 44 studies = 1.36 Summary ORs within each methodological design also given individual ORs within studies ranging from 0.68 to 2.56 Victimization: Summary OR across 44 studies = 1.29 Summary ORs within each methodological design also given individual ORs within studies ranging from 0.40 to 3.14 (see Table 3, p. 36)

Reports included (N)	14 (p. 549)	19 (Table 1, pp. 187–188)	Only number of studies specified clearly (p. 407)	26 (Table 1, pp. 80–82)	Only number of studies specified clearly (p. 29; see also Table 3	89 (see Appendix, p. 48)
Studies included	14: 8 controlled designs 4 age-cohort designs 2 uncontrolled designs/no control group (Muthe 1989; Peterson and Rigby 1999)	16: 7 controlled designs 6 age-cohort designs 3 uncontrolled designs/ no control group (Pepler et al. 1994; Peterson and Rigby,1999; Pitts and Smith 1995)	42 studies relating to 45 separate observations (p. 407); the actual reports are not shown, so the actual type of study designs cannot be confirmed; the authors indicate that only controlled studies were included	26, but only 17 with bullying as an outcome (see Table 1, pp. 80–82) Of the 17: 13 controlled designs 1 age-cohort design (Olweus 1994) 2 uncontrolled designs/ no control group (Roland 2000; Tierney and Dowd 2000)	15: 15 studies, but only 13 specific to bullying Of the 13: 7 controlled designs 2 age-cohort designs 4 uncontrolled designs/no control group (Cowie and Olafsson 2000; Mueller and Parisi 2002; Orpinas et al. 2003; Pepler et al. 1994)	53, but only 44 had statistical information that allowed inclusion in the meta-analyses (see Appendix, p. 48) Of the 53: 17 evaluations of randomized experiments 21 evaluations with before-after, intervention-control comparisons 4 other intervention-control comparisons 11 age-cohort designs (see Appendix, pp. 48–50)
Outcome measures	All studies specific to bullying	All studies specific to bullying	23 on nonviolent bullying 22 on violence (see Table 1, p. 408)	17 studies specific to bullying 9 studies on other outcomes (see footnote)	13 studies specific to bullying 2 studies on other outcomes (see footnote)	All studies specific to bullying (perpetration or victimization)
Correlating features to effect sizes	No	No	Yes; moderator analyses based on age, outcome variable (i.e., bullying or violence), and population type (children at risk for violence versus general school population)	No	No	Yes; moderator analyses for all program components and implementation features (see Tables 4 and 5, pp. 41–42) Also weighted multiple regression analyses (Table 6, p. 44)
Conflict of interest (COI) analyses	No mention in the review of the presence of COI analyses at the level of primary studies	No presentation of separate effect sizes between independent evaluations and evaluations conducted by the developer as evaluator in the meta-analysis; this	No presentation of separate effect sizes between independent evaluations and evaluations conducted by the developer as evaluator in the meta-analysis	No mention in the review of the presence of COI analyses at the level of primary studies	No presentation of separate effect sizes between independent evaluations and evaluations conducted by the developer as	No presentation of separate effect sizes between independent evaluations and evaluations conducted by the developer as evaluator in the meta-analysis

(*continued*)

Bullying Prevention: Assessing Existing Meta-Evaluations, Table 1 (continued)

Publication	Smith et al. (2004)	Baldry and Farrington (2007)	Ferguson et al. (2007)	Vreeman and Carroll (2007)	Merrell et al. (2008)	Farrington and Ttofi (2009); Ttofi and Farrington (2011)
		topic is acknowledged (see p. 201)			evaluator in the meta-analysis	
Cost-benefit analyses	Not present at the level of secondary analyses; no mention of cost-benefit analyses at the level of primary research either	Not present at the level of secondary analyses; no mention of cost-benefit analyses at the level of primary research either	Not present at the level of secondary analyses; no mention of cost-benefit analyses at the level of primary research either	Not present at the level of secondary analyses; no mention of cost-benefit analyses at the level of primary research either	Not present at the level of secondary analyses; no mention of cost-benefit analyses at the level of primary research either	Not present at the level of secondary analyses, but the importance of this research is acknowledged (see p. 46); no mention of cost-benefit analyses at the level of primary research either

Note 1: Evaluations with other outcome measures in the Vreeman & Carroll (2007) synthesis include: Elliot & Faupel, 1997 (on children's solutions on hypothetical bullying scenarios); Englert, 1999 (on violence); Fast et al., 2003 (on peer aggression); Metzler et al., 2001 (on perceptions of safety, feelings of being target of harassment, discipline referrals and reports of receiving praise); Mitchell et al., 2000 (on school initiatives promoting health); Teglasi & Rothman, 2001 (on peer aggression); Twemlow et al., 2001 (on school violence); Warden et al., 1997 (on ability to respond to hypothetical dangerous situations, including hypothetical incidences of bullying). Exact references can be found in the Vreeman and Carroll (2007) paper.

Note 2: The authors (Merrell et al., 2008) argue that they have included 16 evaluations in their meta-analyses of various outcome measures (see table 4, p. 37). Effect sizes for each study are not presented and searches/screening of documents is not clearly specified, but the following should be noted: (a) four publications of the same evaluation (i.e. those related to the Flemish Anti-bullying Program by Stevens and colleagues) are presented as separate evaluations. On table 1, we have counted all four publications as one study; (b) the Leadbeater et al. (2003) publication includes relational and physical victimization as outcome measures; it is not specific to bullying victimization; and (c) one evaluation (i.e Newman, 1999) is presented alongside the age-cohort design evaluated in Norway by Olweus. However, the Newman (1999) evaluation was conducted in the States and the outcome measure targeted improving teachers' knowledge of appropriate bullying intervention skills that they should be utilizing; participants were school teachers and not students. Exact references can be found in the Merell et al. (2008) paper.

and Carroll 2007), whereas other reviews focused specifically on bullying (Smith et al. 2004; Baldry and Farrington 2007; Farrington and Ttofi 2009). The *methodological approach* of the reviewers can also affect their analytic strategies. For example, Vreeman and Carroll (2007, p. 79) "… did not exclude or discount studies based on … *retention rates* or program *intensity* because these characteristics are not associated definitely with the strength of treatment effects." In contrast, the authors of the latest systematic review (Farrington and Ttofi 2009; Ttofi and Farrington 2011) coded the *program intensity* and *duration* and correlated these features with the effect sizes in order to examine their association. Ttofi and Farrington (2011) also identified concerns about *retention rates* – and issues with possible *differential attrition* – which are related to the effect size measures; some of the effect sizes for specific evaluations included in their review were not based on the published reports but based on results obtained via e-mail communication with evaluators, in order to avoid biased findings resulting from issues related to differential attrition, multiple imputation methods, etc. (Farrington 2006; Farrington and Ttofi 2009).

As another example, the review by Ferguson et al. (2007), which was not based on systematic searches, primarily aimed at extending previous research by conducting publication bias analyses (see pp. 405–406), but they restricted their searches to journal articles only (see p. 407). In contrast, other reviews (e.g., Farrington and Ttofi 2009; Ttofi and Farrington 2011) did not set language or sample size restrictions, or limit the type of manuscripts to be included, because of concerns about publication bias. While there is some disagreement about this issue in the literature, it is common for researchers to extend searches beyond published reports in evaluation research of primary studies (Wilson 2009).

In conclusion, the existing reviews of the efficacy of bullying prevention programs varied greatly in methodology employed, including their inclusion criteria, depth of searches, and screening of relevant documents. Consequently, the resulting meta-analyses presented a summary effect size obtained from 13 (Merrell et al. 2008), 23 (Ferguson et al. 2007), and 44 (Farrington and Ttofi 2009; Ttofi and Farrington 2011) evaluations accordingly. These differences in methodology and number of studies included could explain the marked discrepancies in the summary effect obtained across reviews, as shown in Table 1. This could also explain why some previous reviews (e.g., Ferguson et al. 2007; Merrell et al. 2008) concluded that bullying prevention programs had little effect in reducing the level of bullying perpetration and victimization, while other reviews (Farrington and Ttofi 2009) were optimistic in their findings. Ideally, these reviews should critically assess and synthesize the existing evaluation research, with the final aim of minimizing bias. However, bias is not only the province of primary studies – it may also exist at the level of a review (Wilson 2009).

Open Questions and Future Research Directions

The rapid growth in the research on school bullying has undoubtedly advanced the scientific knowledge in this issue. However, considerable work is still needed for its successful translation into effective practice and policy (Swearer et al. 2010). For example, systematic reviews in social sciences offer a great promise. Unfortunately, reviews published in social science journals quite often lack a methodological rigor or transparency in the methods. This is evident in the current synthesis of previous meta-evaluation studies of bullying prevention programs. When conducting research trials of prevention programs, researchers in psychology and other social sciences should follow the methodological quality criteria of the Consolidated Standards of Reporting Trials (CONSORT) statement (Altman et al. 2001), or the procedures outlined by the Cochrane Collaboration for secondary research reviews, as is common in public health and medicine. There is a steady trend in the social sciences toward greater transparency in the research process, as illustrated by increased efforts to create an equivalent CONSORT statement for the social sciences (Perry and Johnson 2008) and the creation of the Campbell Collaboration's methodological criteria for secondary analyses (Farrington et al. 2011). These standards for evaluation research can also be used by scholars, policymakers, and the general public to assess the validity of conclusions about the effectiveness of interventions in reducing bullying at school (Ttofi and Farrington 2011). This entry has provided clear examples of how substantial differences in the procedures followed for secondary reviews can result in marked differences in the conclusions drawn across these reviews regarding the effectiveness of bullying prevention programs.

Another important guideline in future meta-evaluation research refers to the issue of conflict of interest. The importance of conducting conflict of interest (COI) analyses is well established in the fields of medicine and public health. Within the field of criminology, recent studies have also shown that the reported effect sizes of prevention and intervention trials are larger when program developers are involved in a study than when trials are conducted by independent researchers; these differences may be due to various different types of biases, including biases resulting from conflict of interest issues (Eisner 2009; Farrington 2006). COI analyses have not been conducted in previous meta-analytic studies of bullying prevention programs, and this is a promising new line of research. It is noteworthy, for example, that the effect sizes obtained from the independent evaluation of a specific bullying prevention program are substantially smaller than those obtained from trials conducted by the developer as evaluator (Eisner 2009), although other explanations are also possible. Eisner and Humphreys (2011) developed an instrument for assessing COI in evaluation research, and initial tests of the scale in the area of family interventions are very promising. Interested scholars in the area of school bullying could begin a new line of evaluation research by testing and perhaps refining the relevant scale.

Another highly neglected topic at the level of meta-analytic research is that of cost-benefit analyses (Farrington and Ttofi 2009). Cost-benefit analyses can be a big "selling point" to policymakers and potential funding agencies, especially given the scarce resources that schools are often faced with. Admittedly, this type of analyses can also be conducted at the level of primary research. It is interesting to indicate, however, that of the 53 different evaluations of bullying prevention programs, only the review by Bagley and Pritchard (1998) included a cost-benefit analysis.

In prevention and intervention research, a distinction is made between efficacy, effectiveness, and dissemination trials (Flay et al. 2005). The importance of this distinction has been highlighted in meta-analyses that have examined factors that affect the size of treatment effects, with larger effect sizes found in efficacy trials (Eisner et al. 2011), most likely due to the increased researcher-imposed support and structure during the implementation process. This distinction is absent in the current meta-analytic studies of bullying prevention programs despite the fact that it may be critical when it comes to translating research findings into policy recommendations. Most evaluations of bullying prevention programs would fall under the category of efficacy trials, with relatively few meeting criteria for effectiveness research (Flay et al. 2005); additional research is needed to examine whether dissemination trials would demonstrate significant effects of bullying prevention programs. Within the area of bullying prevention, there are very few examples of programs being rigorously tested when being brought to scale (e.g., Karna et al. 2011; Waasdorp et al. 2012). Future research should also try to establish factors related to sustainability of the treatment effect when going to scale as well as the issue of sustainability of the treatment effect at longer follow-ups. Notably, the majority of bullying trials have relatively short follow-up periods, which does not allow for the contextual changes, such as in the school climate or school culture, or leadership, which we theorized are important impacts of effective bullying prevention programs (Bradshaw and Waasdorp 2009). Many other recommendations for future research can be made, and for a more detailed discussion, interested readers should seek advice from a paper that is being prepared for a special issue of the *American Psychologist* journal (Bradshaw et al. submitted).

Conclusions

The mixed findings from the extant reviews of bullying prevention approaches have been particularly disconcerting to policymakers, researchers, and practitioners alike (Bradshaw and Waasdorp 2011), as several reviews have provided a rather pessimistic interpretation of the state of the science in bullying prevention programming (e.g., Ferguson et al. 2007; Merrell et al. 2008), whereas others have been more optimistic (e.g., Farrington and Ttofi 2009; Ttofi and Farrington 2011). As a result, many are unclear where the field stands in terms of the evidence base for bullying prevention programming. The current entry aimed to provide greater transparency across the extant reviews so that various constituents can have a better understanding of potential reasons for the discrepant findings. Given the current focus on contrasting the scientific approaches employed by the different reviews of bullying prevention studies and extracting a general conclusion regarding the efficacy of bullying prevention efforts, it is not possible to endorse any specific program or model. Rather, the current study aimed to provide some guidance for researchers, policymakers, and practitioners on the extent to which each of the reviews employed various research-based approaches, as these methodological issues likely impacted the conclusions drawn from the research. Given the significance of bullying prevention for policy and prevention science, we issue a call for more longitudinal randomized controlled trials of promising programs and programs in wide use (particularly in the USA), as well as more rigorous systematic and meta-analytic reviews, in order to strengthen the current evidence base for preventing bullying programming. Many challenges but also many promising avenues lie ahead for future research in the area of bullying prevention and intervention. The time is ripe for a new research agenda in unexplored scientific avenues such as conflict of interest analyses and cost-benefit analyses to mention a few.

Recommended Reading and References

Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gotzsche PC, Lang T, CONSORT GROUP (2001) The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 134(8):663–694

Bagley C, Pritchard C (1998) The reduction of problem behaviours and school exclusion in at-risk youth: an experimental study of school social work with cost-benefit analyses. Child Fam Soc Work 3:219–226

Baldry AC, Farrington DP (2007) Effectiveness of programs to prevent school bullying. Vict Offender 2(2):183–204

Bradshaw, CP, Ttofi MM, Eisner M (submitted) A public health approach to bullying prevention: translating the research to practice. Am Psychol

Bradshaw CP, Waasdorp TE (2009) Measuring and changing a culture of bullying. Sch Psychol Rev 38:356–361

Bradshaw C, Waasdorp T (2011) Effective strategies in combating bullying. White paper prepared for the 2011 white house conference on bullying, Washington, DC. http://www.stopbullying.gov/references/white_house_conference/index.html

Cowie H, Olafsson R (2000) The role of peer support in helping the victims of bullying in a school with high levels of aggression. School Psychol Int 21: 79–94

Eisner M (2009) No effects in independent prevention trials: can we reject the cynical view? J Exp Criminol 5:163–183

Eisner M, Humphreys D (2011) Measuring conflict of interest in prevention and intervention research: a feasibility study. In: Bliesener T, Beelmann A, Stemmler M (eds) Antisocial behavior and crime: contributions of developmental and evaluation research to prevention and intervention. Hogrefe Publishing, Cambridge, MA, pp 165–180

Eisner M, Malti T, Ribeau D (2011) Large-scale criminological field experiments: the Zurich project on the social development of children. In: Gadd D, Karstedt S, Messner SF (eds) Sage handbook of criminological research methods. Sage, London, pp 410–424

Farrington DP (2003) Methodological quality standards for evaluation research. Ann Am Acad Pol Soc Sci 587(1):49–68

Farrington DP (2006) Methodological quality and the evaluation of anti-crime programs. J Exp Criminol 2(3):329–337

Farrington DP, Ttofi MM (2009) School-based programs to reduce bullying and victimization

(Campbell systematic reviews No. 6). Campbell Corporation, Oslo. doi:10.4073/csr.2009.6

Farrington DP, Weisburd DL, Gill CE (2011) The Campbell collaboration crime and justice group: a decade of progress. In: Smith CJ, Zhang SX, Barberet R (eds) Routledge handbook of international criminology. Routledge, New York, pp 53–63

Ferguson CJ, Miguel CS, Kilburn JC, Sanchez P (2007) The effectiveness of school-based anti-bullying programs: a meta-analytic review. Crim Justice Rev 32(4):401–414

Flay BR, Biglan A, Boruch RF, Gonzalez Castro F, Gottfredson D, Kellam S, Moscicki EK, Schinke S, Valentine JC, Ji P (2005) Standards of evidence: criteria for efficacy, effectiveness and dissemination. Prev Sci 6(3):151–175

Karna A, Voeten M, Little TD, Poskiparta E, Alanen E, Salmivalli C (2011) Going to scale: a non-randomized nationwide trial of the KiVa antibullying program for grades 1–9. J Consult Clin Psychol 79(6):796–805

Merrell KW, Gueldner BA, Ross SW, Isava DM (2008) How effective are school bullying intervention programs? A meta-analysis of intervention research. Sch Psychol Q 23(1):26–42

Mueller EE, Parisi MJ (2002) Ways to minimize bullying. Unpublished Master's thesis. Saint Xavier University, Chicago

Munthe E (1989) Bullying in Scandinavia. In: Roland E, Munthe E (eds) Bullying: an international perspective. David Fulton, London, pp 66–78

Olweus D (1994) Bullying at school: basic facts and effects of a school based intervention program. J Child Psychol Psychiatry 35:1171–1190

Olweus D (2005) A useful evaluation design, and effects of the olweus bullying prevention program. Psychol Crime Law 11(4):389–402

Orpinas P, Horne AM, Staniszewsld D (2003) School bullying: changing the problem by changing the school. School Psychol Rev 32:431–444

Pepler DJ, Craig WM, Ziegler S, Charach A (1994) An evaluation of an antibullying intervention in Toronto schools. Can J Commun Ment Health 13:95–110

Perry A, Johnson M (2008) Applying the consolidated standards of reporting trials (CONSORT) to studies of mental health provision for juvenile offenders: a research note. J Exp Criminol 4(2):165–185

Peterson L, Rigby K (1999) Countering bullying at an Australian secondary school. J Adolesc 22:481–492

Petrosino A (2003) Standards for evidence and evidence for standards: the case of school-based drug prevention. Ann Am Acad Pol Soc Sci 587(1):180–207

Pitts J, Smith P (1995) Preventing school bullying. Home Office, London

Roland E (2000) Bullying in school: three national innovations in Norwegian schools in 15 years. Aggress Behav 26:135–143

Smith PK, Cowie H, Olafsson RF, Liefooghe APD (2002) Definitions of bullying: a comparison of terms used, and age and gender differences, in a 14-country international comparison. Child Dev 73(4):1119–1133

Smith JD, Schneider B, Smith PK, Ananiadou K (2004) The effectiveness of whole-school anti-bullying programs: a synthesis of evaluation research. Sch Psychol Rev 33:548–561

Swearer SM, Espelage DL, Vaillancourt T, Hymel S (2010) What can be done about school bullying? Linking research to educational practice. Educ Res 39(1):38–47

Tierney T, Dowd R (2000) The use of social skills groups to support girls with emotional difficulties in secondary schools. Support Learn 15:82–85

Ttofi MM, Farrington DP (2011) Effectiveness of school-based programs to reduce bullying: a systematic meta-analytic review. J Exp Criminol 7:27–56

Ttofi MM, Farrington DP, Lösel F, Loeber R (2011a) Do the victims of school bullies tend to become depressed later in life? A systematic review and meta-analysis of longitudinal studies. J Aggress Confl Peace Res 3(2):63–73

Ttofi MM, Farrington DP, Lösel F, Loeber R (2011b) The predictive efficiency of school bullying versus later offending: a systematic/meta-analytic review of longitudinal studies. Crim Behav Ment Heal 21:80–89

Ttofi MM, Farrington DP, Lösel F (2012) School bullying as a predictor of violence later in life: a systematic review and meta-analysis of prospective longitudinal studies. Aggress Violent Behav 17:405–418

Vreeman RC, Carroll AE (2007) A systematic review of school-based interventions to prevent bullying. Arch Pediatr Adolesc Med 161(1):78–88

Waasdorp TE, Bradshaw CP, Leaf PJ (2012) The impact of School-wide Positive Behavioral Interventions and Supports (SWPBIS) on bullying and peer rejection: a randomized controlled effectiveness trial. Arch Pediatr Adolesc Med 166(2):149–156

Wilson DB (2009) Missing a critical piece of the pie: simple document search strategies inadequate for systematic reviews. J Exp Criminol 5(4):429–440

Buprenorphine

► Drug Abuse and Alcohol Dependence Among Inmates

Burglary at Construction Sites

► Theft at Construction Sites

Calibration of Simulation Models

Nick Malleson
School of Geography, University of Leeds,
Leeds, UK

Overview

Simulation models are becoming increasingly popular in criminology research. In order for researchers to have confidence in the results of simulation studies, it is essential to make sure that the models are properly evaluated. *Calibration* is a major element to this evaluation and refers to the estimation and adjustment of model parameters to improve the agreement between model output and a data set. This entry will discuss some different methods of model calibration as well as related quantitative methods of error assessment.

Introduction

Simulation models are becoming increasingly popular in criminology research, both as tools for exploring real-world crime patterns (explanatory models) and for experimenting with underlying theory (conceptual models). In order for researchers to have confidence in the results of simulation studies, it is essential to make sure that the models are properly evaluated. The process of evaluation is commonly divided into three activities, defined by Rykiel (1996) as follows:

- **Verification** is a demonstration that the modeling formalism is correct.
- **Calibration** is the estimation and adjustment of model parameters and constants to improve the agreement between model output and a data set.
- **Validatio**n is a demonstration that a model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model.

It is possible to calibrate a model quantitatively by assessing error using statistics or qualitatively by manually comparing model results and field data. Qualitative approaches are particularly well suited to spatial models where the researcher is able to compare maps. However, for complicated models with many parameters, a qualitative approach to calibration will probably be very time consuming and unlikely to reveal the optimal model configuration (which is the ultimate aim of calibration). Therefore, this article will focus on quantitative methods of error assessment (rather than relying on human objectivity) and automatic calibration routines that are able to explore a model's parameter space and estimate error without human intervention. Also, in the context of environmental criminology, it is much more important to accurately reflect field conditions when working with *explanatory models* – i.e., those that simulate real-world conditions – so the discussion will focus on these in particular. Conceptual models,

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

on the other hand, will not necessarily attempt to replicate real environmental conditions so automatic calibration methods are less relevant.

The article is organized as follows. Section outlines the main principles of calibration and Section follows with a more in-depth assessment of a number of methods for model calibration. Sections and complete the article by discussing the state of the art, literature controversies, and open questions.

Background

Calibration refers to the process of configuring a model's parameters to match some observed historical data. This usually consists of searching for a combination of parameter values that cause the model to produce data which are similar to that collected from the real system under investigation. In many cases, a single "fitness" value is sought which succinctly summarizes the correspondence between simulated data and field observations.

Figure 1 illustrates the calibration process. The model is repeatedly reconfigured with the aim of reducing the error between the results and the field data. Once a predetermined error level has been reached, the process ends and the model can be considered calibrated, i.e., configured in such a way that it is apparently able to simulate the real system effectively. The level of error deemed acceptable is subjective and depends on the individual study. It is also worth noting that simply matching data is not necessarily a sufficient criteria for establishing model correctness, which is a point Section will address in more detail.

With simple models, the process of configuring parameters and calculating error is usually relatively simple. However, *simulation* models are often extremely complex and can contain a large number of configurable parameters. To confound the situation, model parameters often have nonlinear effects on the model's behavior which makes it difficult to predict how the model will behave under new parameter configurations. Hence, the process of manipulating a model's parameters to match some field conditions is often nontrivial.

To being with, the means of evaluating the degree of similarity between model results and expected data can be troublesome in itself. Once a quantitative measure of similarity has been developed, there are numerous methods that can be used to explore the parameter space of the model in search of the optimal configuration. These are also known as *optimization* methods. The following section will first discuss the

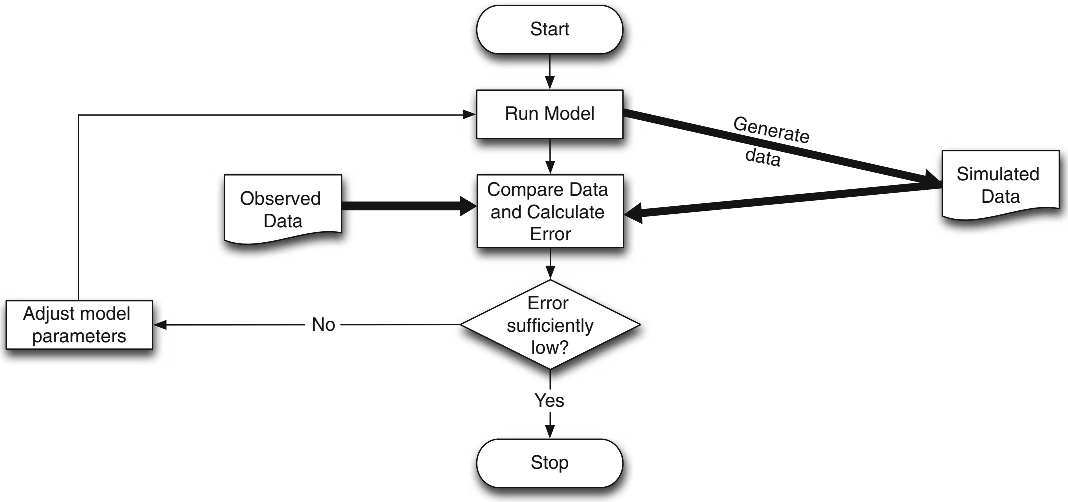

Calibration of Simulation Models, Fig. 1 An overview of the process of calibration different methods that can be used to calculate error and follow with a review of a number of automatic calibration/optimization routines.

Methods for Model Calibration

Assessing Goodness of Fit

In the context of calibration, goodness of fit (GoF) is an important measure. It refers to the error between the results of a model and the data that it is trying to replicate (hereafter referred to as "observed data"). Commonly, GoF statistics are applied to tabular data, such as the number (or rate) of crimes against different types of people, within different spatial areas or during a particular time period. Knudsen and Fotheringham (1986) experimented with a number of goodness-of-fit statistics and found the standardized root mean square error (SRMSE) to be the best performing. A drawback with SRMSE, however, is that the value itself is difficult to understand. For example, with the SRMSE it is not possible to state what percentage of the variation in observed data can be accounted for by a model. An alternative statistic, R^2 solves this problem because it represents the percentage of agreement between the model and the expected data. However, R^2 is insensitive to the overall amount of error, predicting a good fit in some circumstances where the SRMSE would not (Harland 2008).

The SRMSE can be defined as

$$SRMSE = \frac{\sqrt{\left(\Sigma\left(y_i' - y_i\right)^2/n\right)}}{\bar{y}} \tag{1}$$

where y_i' is the predicted value at matrix point i, y_i is the actual value at i, $\bar{y}$ is the mean value of the predicted values (y'), and n is the total number of values. The lower limit of the statistic is 0 which indicates no difference between the predicted values (y_i') and the observed values (y_i). The upper limit is usually 1 (Knudsen and Fotheringham 1986) but can be greater, particularly when matrices are sparse (Harland 2008).

Using the same notation, R^2 can be defined as

$$R^2 = 1 - \frac{\sum_i \left(y_i - y_i'\right)^2}{\sum_i \left(y_i - \bar{y}\right)^2} \tag{2}$$

and a value of 1 indicates identical data sets. The lower limit of the statistic is 0.

There are a variety of alternative GoF statistics that could be used to assess model error and the most appropriate will depend on the data and application area. For more information about the statistics discussed here, the interested reader can refer to the comprehensive assessment in Knudsen and Fotheringham (1986). Alternatively, there are a wide range of textbooks that define methods for both parametric and nonparametric data. In terms of calibration, the most important decision is to choose the GoF statistic that is appropriate to the study and the nature of the data so that a reliable assessment of error can be made. In addition, more than one statistic could be used simultaneously to provide a more comprehensive assessment of model error.

Goodness of Fit for Spatial Data

Many simulation models in environmental criminology explore the spatial distributions of crime (Liu et al. 2005; Groff 2007; Hayslett-McCall et al. 2008; Dray et al. 2008; Birks et al. 2012; Malleson et al. 2013). If the simulations work at an aggregate spatial scale (measuring crimes per area), then the procedure for assessing GoF is the same as that for nonspatial data. However, it is preferable for simulation models to use data on individual crime occurrences and, therefore, generate point pattern data. Common GoF measures cannot be used to compare point patterns directly as the data must be in the form of a table or matrix. One solution to this problem is to first aggregate the point data to commonly used areal boundaries such as the enumeration district or census tract. However, this process is far from ideal. Firstly, such aggregation will expose the results to the modifiable areal unit problem (MAUP: Openshaw 1984). Openshaw found that changing the size and shape of the boundaries themselves can have a dramatic effect on the resulting spatial patterns and subsequent results. Secondly, the process of aggregating is likely to hide important patterns that are present at finer geographies. Andresen and Malleson (2011), for example, showed that there was considerable spatial heterogeneity in crime rates at the street level which would be hidden at larger spatial scales. Hence aggregation will inevitably introduce error.

An alternative approach is to compare the point patterns directly *without aggregating*. Although no commonly accepted methods exist for this purpose, such as the SRMSE or R^2 there are a a number of useful spatial statistics that can describe the distributions of point patterns and can be used to make mathematical comparisons. These could replace, or compliment, traditional GoF statistics used to determine model error during calibration. Table 1 summarizes some of these statistics and illustrates the results of their application to three data sets: two similar point patterns produced by a simulation model ("Model1" and "Model2") and a point pattern produced by a random process ("Random").

The functions outlined in Table 1 provide information about the degree of clustering in

Calibration of Simulation Models, Table 1 A summary of spatial statistics that can be used to describe and compare the spatial structure of point patterns (Malleson 2010)

Statistic	Pros/cons	Usage with example data
Nearest Neighbour Index (NNI) – also known as the Clark and Evans *R* statistic (Clark and Evans 1954) – is the ratio of the minimum nearest-neighbour distance. The nearest-neighbour distance for a point *i* is the distance to the closest neighbouring point. ($d_{\min}$) to the expected minimum distance for a random point pattern ($\bar{\delta}$) $NNI = \frac{\bar{d}_{\min}}{\bar{\delta}} = \frac{\sum_{i=1}^{n} \frac{d_{ij}}{n}}{\frac{1}{2\sqrt{A/n}}}$ (3)	Gives a concise general picture of whether or not clustering is present (compared to random data) Usefil as a preliminary procedure (Bailey and Gatrell 1995) It is difficult to account for edge effects (there are some solutions to edge effect problems, such as circular or rectangular corrections (Levine 2006), but these are not ideal (Chainey and Ratcliffe 2005)) Too simplistic to be really useful	The NNI statistics suggests that all the data used in the following examples (Modell, Model2 and Random) are clustered which is to be expected (Malleson et al. 2010). Althought the statistic is not comprehensive enough to assess error in isolation, it could be useful as a preliminary measure of similarity as part of a larger error assessment during calibration
The *G* function, at a given distance, *d*, is the fraction of points, s_i, whose nearest neighbour is less than *d* away: $G = \frac{\#(d_{\min}(s_i)<d)}{n}$ (4) where # means "the number of" (as in Bailey and Gatrell 1995) **The *F* function** is similar to *G* but uses the distance from a randomly selected map location to the nearest point. As defined bu O'Sullivan and Unwin (2003): if $\{p_1 \ldots p_i \ldots p_m\}$ is a set of *m* randomly selected locations and *S* is the set of all points, then: $F = \frac{\#(d_{\min}(p_i,S)d)}{m}$ (5)	Describe clustering in more detail than the NNI index by providing a measure at different distances Can be used to differentiate between clustered and uniform data (see O'Sullivan and Unwin (2003) for a fuller discussion) Only consider a *single* nearest neighbour distance in their calculations so disregard a considerable amount of information	The following illustrates the *G* functions for the Model1, Model2 and Random data sets (graphs of *F* are similar). The graphs illustrate that, at shorter distances, the model data sets are more clustered than the Random one. Therefore a traditional GoF statistic could subsequently be used to estimate the differences between the *G* of *F* graphs for different data. This would provide a single statistical measure of similarity which could be used during calibration 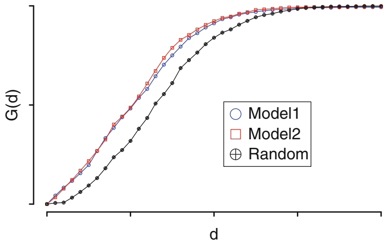

(*continued*)

Calibration of Simulation Models, Table 1 (continued)

Statistic	Pros/cons	Usage with example data
The *L* **function** is a transformation of Ripley's *K* that provides evidence for whether or not clustering is more or less than would be expected under complete spatial randomness (CSR) (O'Sullivan and Unwin 2003): $L(d) = \sqrt{\frac{K(d)}{\pi}} - d \qquad (6)$ Values of $L(d) < 0$ suggest that there are fewer events in the space than would be expected under CSR and that the data are therefore less clustered. The reverse is true for $L(d) > 0$	Takes all the neighbours that are within a given distance into account so is a more descriptive statistic By comparing graphs of $L(d)$ it is possible to determine how similar the clustering of to point patterns is	The following graph illustrates that the Model1 and Model2 data are more clustered than the random dataset (which equates to approximately $L(d) = 0$ for low d values). Above $d \approx 3{,}000$ L begins to fall due to boundary effects (this is because many of the large circles produced by the underlying K function are nearly empty at large distances because they cover areas outside the simulation boundary with no points (O'Sullivan and Unwin 2003). As with the G and F functions, it would be possible to use a traditional GoF statistic to quantitatively compare the similarity of L functions for two point patterns (e.g., simulated data and calibration data) 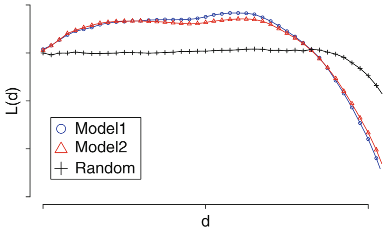

point pattern data. Although these statistics have the benefit that they are not susceptible to the modifiable areal unit problem, they could not be used during calibration, in isolation, because it is possible that two different point patterns nevertheless have the same clustering properties. However, these could be used in conjunction with other methods (such as aggregating the points to area boundaries and applying a GoF test) to provide a more comprehensive error assessment.

As well as comparing mathematical descriptions of clustering in point patterns, it is also possible to generate raster density maps from the point patterns and compare these mathematically. This approach is common in the field of spatial modeling for comparing simulated and real land use. For a review of recent approaches, the reader is directed to Kuhnert et al. (2005). Finally, it is also possible to aggregate the point patterns to a regular grid and then use traditional GoF statistics on the resulting matrix as discussed by Costanza (1989). This approach has the advantage of reducing the effects of the modifiable areal unit problem because numerous regular grids can be applied to the point patterns at the same resolution.

Exploring the Parameter Space

Having determined an effective means of estimating the goodness of fit between simulated and observed data, calibration itself can begin. The task of model calibration is effectively a search through the model parameter space, assessing the accuracy of the model under different combinations of parameters. For simulation models with large numbers of continuous parameters, this search space will be extensive. This is confounded by the fact that simulation models are often nonlinear, so the effects of changing parameter values are not easy to predict.

For example, Fig. 2 presents an example parameter space for a nonlinear model with two continuous parameters (A and B). Varying parameter combinations changes the accuracy of

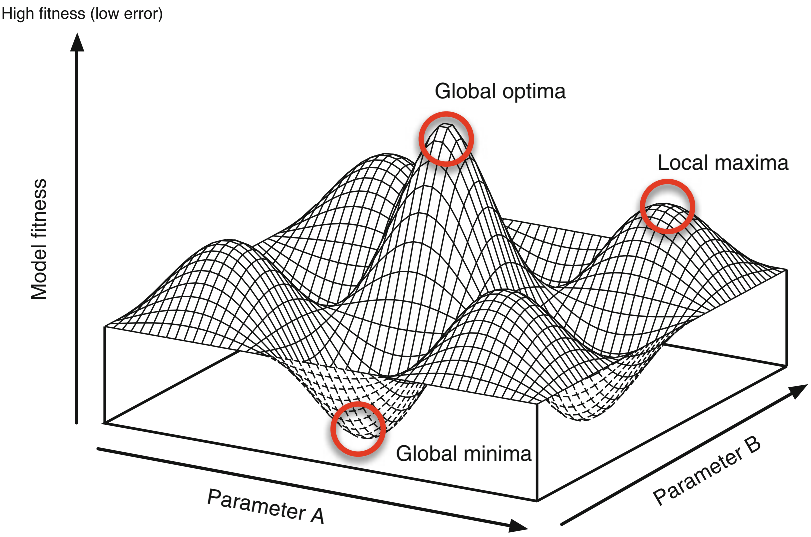

Calibration of Simulation Models, Fig. 2 An example parameter space for a nonlinear model with two parameters

the model results; hence the aim of the calibration process is to find the optimal parameter configurations ("global optima"). However, even for a model with a low number of parameters, it becomes apparent that a manual search of the parameter space is unlikely to reveal global optima. Therefore calibration is often conducted by computer algorithms which are able to intelligently search the vast parameter space.

Parameter Sweeps

A parameter sweep is a simple process of systematically varying the model parameters in sequence so that many possible combinations are explored. Each parameter to be tested has a *start*, *end*, and *increment* value which determines the range of possible values that the parameter will take. Again using Fig. 2 as an example, a parameter sweep with $start = 1$, $end = 10$, and $increment = 1$ for both parameters A and B results in 100 possible parameter combinations.

Although the method is simple, it has a number of drawbacks, namely:

- The number of individual runs can be extensive because it increases exponentially with the number of parameters. For example, to conduct a parameter sweep on a model with five parameters, each of them integers between 1 and 10, a total of $10^5 = 100{,}000$ model runs will be required.
- The sweep explores the entire parameter space equally. More advanced approaches are able to expend greater effort fine tuning already successful configurations and ignore parts of the parameter space that are extremely unlikely to reveal global optima.

Hill Climbing

Hill climbing is a procedure that starts with an arbitrary model configuration and makes a small change to one of the parameters. If the change improves the performance of the model (reduces error), then it is retained. If the change increases error, then it is discarded. This process repeats until there are no possible parameter changes that improve the model. The main drawback with the approach is it is likely to get stuck on suboptimal configurations (e.g., the local maxima in Fig. 2), in which case it would be necessary to temporarily accept changes to the configuration that will actually *increase* the error between the model and the calibration data.

Simulated Annealing

Annealing is the process used in metallurgy in which a material is heated and then cooled in order to alter its properties (strength, brittleness, etc.). The amount of time spent at a high temperature and the rate of cooling influence the resulting properties of the material. The

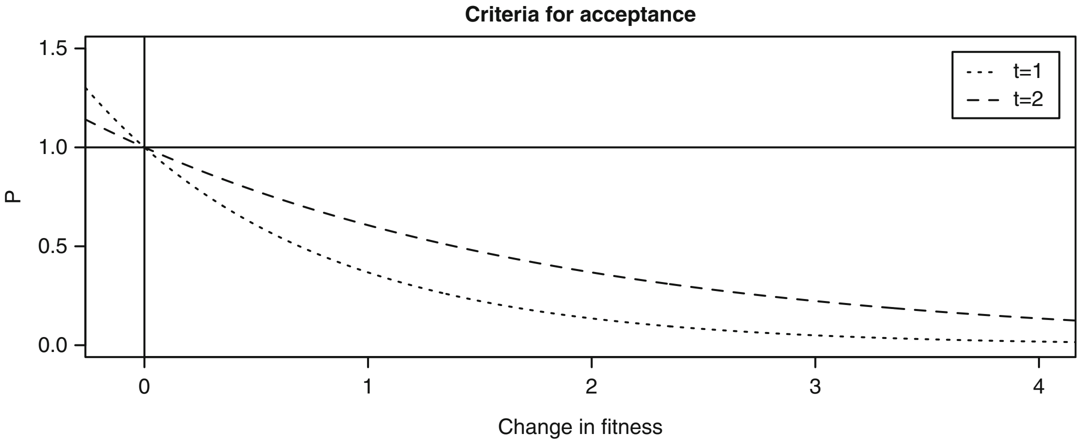

Calibration of Simulation Models, Fig. 3 The effect of temperature on the probability of accepting a regressive move

simulated annealing optimization procedure takes its inspiration from an algorithm that simulates this annealing process (Metropolis et al. 1953). Originally formulated by Kirkpatrick et al. (1983), the procedure improves upon hill climbing algorithms by occasionally allowing choices that *lower* the fitness of the model. Whereas hill climbing will always choose the best move from those available, simulated annealing chooses a random move from the neighborhood. If the move improves the fitness, then it is always accepted, but if it does not, then there is still a *possibility* that it will be accepted. This helps the algorithm to climb out of local maxima.

The criteria for accepting a lower fitness is given by

$$P = e^{-c/t} > R(0,1) \tag{7}$$

where c is the change in fitness (negative for an improvement, positive for deterioration), t is the temperature, and R(0,1) is a random number in the range 0–1. The temperature is used to reduce the chance of accepting poor moves over time so that the algorithm will converge. This is analogous to gradually reducing the temperature in the annealing process. If $t = 0$ then only improvements to the fitness will be accepted which causes the algorithm to behave like a hill climbing procedure. Figure 3 illustrates the change in P for a range of moves under two different temperatures. Note that if the change is positive ($x < 0$), then $P > 1$ so the move will always be accepted. As t decreases, so does the probability of accepting a move.

The rate at which t drops, as well as its initial and final values, will determine how successfully the algorithm will run. Unfortunately there are no standard rules for determining these values, although a number of methods for estimating suitable values have been proposed (Reeves 1995).

Genetic Algorithms

A Genetic Algorithm (GA) is a form of evolutionary algorithm, based on Darwin's theory of natural selection (Darwin 1859). The algorithm works on the premise that small variations in organisms can accumulate if they induce an increase in the overall fitness and this improves the individuals' ability to reproduce (Reeves and Rowe 2003). When describing GAs, the following concepts are important:

- A **gene** is a single model parameter.
- A **chromosome** is a combination of genes, i.e., a unique model configuration.
- The **population** is the current set of chromosomes that the algorithm is using to explore the parameter space.

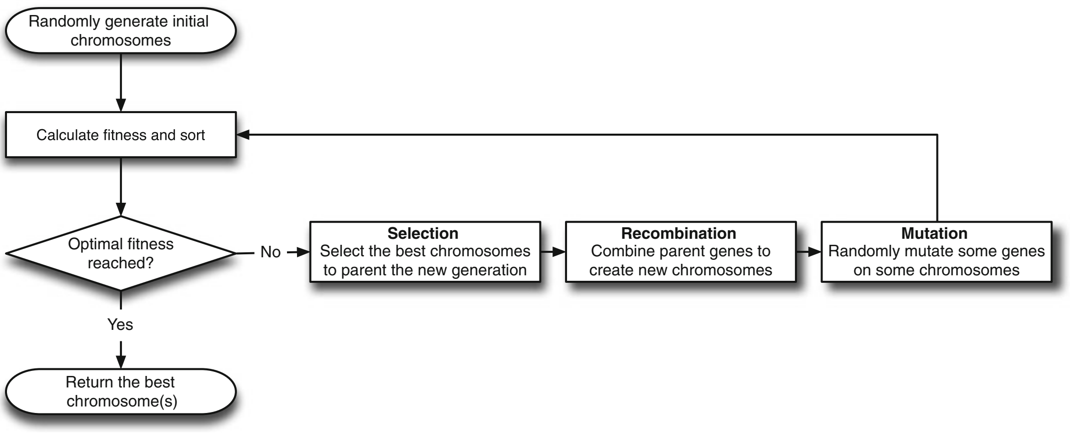
Randomly generate initial chromosomes
Calculate fitness and sort
Optimal fitness reached?
No
Yes
Selection
Select the best chromosomes to parent the new generation
Recombination
Combine parent genes to create new chromosomes
Mutation
Randomly mutate some genes on some chromosomes
Return the best chromosome(s)

Calibration of Simulation Models, Fig. 4 The process of running a genetic algorithm

- **Selection** is the process of choosing the fittest chromosomes.
- **Recombination** is the process of combining the best chromosomes to initialize the new population.
- **Mutation** is a means of exploring a wider area in parameter space by randomly varying some genes in a new population.

The GA is an iterative process, as illustrated by Fig. 4. At each iteration, the algorithm finds the fittest chromosomes (the model configurations with lowest error) and uses these to create the population in the subsequent iteration. The algorithm runs for a set number of iterations or until an acceptable goodness-of-fit value has been found. Although the process of running a GA does not change substantially across implementations, the means of performing selection, recombination, and mutation can be adapted depending on the characteristics of the parameter space. See Reeves and Rowe (2003) for more information.

The main advantage of a GA over the other procedures is that it is able to "home in" on the space with the fittest parameter configurations without devoting effort to exploring suboptimal areas. This makes it much more efficient. Also, mutation allows the algorithm to explore a wider search space which can prevent it from becoming trapped in local maxima.

Calibration of Criminology Models

On the whole, very few simulation models in the field of criminology apply the techniques outlined here. This is partly because simulation modeling has been, until recently, relatively underused in criminology so there is only a limited literature base to begin with. Also, and perhaps more importantly, a large number of studies are *conceptual* rather than *predictive*. Conceptual models do not attempt to replicate real-world crime patterns and instead explore the dynamics of criminology theory in an abstract, artificial environment. Hence it is normally neither necessary nor possible to configure these models to simulate data from the real world. Predictive models, on the other hand, do attempt to replicate real-world patterns and therefore calibration should be an important part of the modeling process.

However, of the limited relevant published studies, very few have applied the formal methods outlined here. For example, Malleson et al. (2013) implement a simulation model of residential burglary and, although automatic calibration is noted as advantageous, the authors result to manual parameter configuration because the extensive run time of the model leads to in insurmountable computational requirements for automated calibration algorithms (Malleson et al. 2012). Similarly, Groff (2007) discusses the advantages of a "common sense" approach to model evaluation by examining factors such as the degree of clustering and the spatial dynamics of hotspots. Again, formal calibration methods such as those outlined here are not applied, although the authors stress that the field is "wide open" (Groff 2007, p. 99). There are numerous other predictive simulation models that would benefit from a rigorous approach to calibration but, for a number of reasons, calibration is not applied with the same rigor as with simulation models in other fields. Partly, this will be because the methods discussed here are not simple and require a degree of computer literacy. But, also, it is due to the novelty of the models to the field and, with time, it is extremely likely that calibration, as well as the other elements to model evaluation, will form a more substantial part of the modeling process.

Controversies and Open Questions

This entry has made reference to a number of calibration methods that can be used to calibrate simulation models. Other methods are available, including simply adjusting parameters manually. But those discussed here are the most suitable for models of complex systems which often have a large number of parameters, behave nonlinearly, and consume/produce large amounts of data. However, unless reliable observed data can be gathered and methods developed to compare the data to simulation results, the process of calibration will not be able to improve model performance (at least in the sense that the model is a good representation of the real world). The first problem is how to obtain reliable, real-world data on which to build a picture of the underlying system. Police recorded crime data is a common source but has a number of drawbacks:

- Discrepancies in police recording practices mean that some crimes, which are heavily underreported, will be misrepresented in police data – although this is somewhat mediated by the assertion that unreported crime clusters near reported crime (Chainey and Ratcliffe 2005).
- The temporal accuracy is often questionable because the actual time of the event is not always known.
- Spatial accuracy can be variable if a location is hard to code accurately (e.g., somewhere in a park) or if human error corrupts the recorded location.
- Offender data, which can be useful for the calibration of the spatial movements of offenders, by definition only provides information about people who have been in contact with the police and therefore misrepresents the actual population of offenders.

Assuming good-quality crime data are available, it is still not certain that those data are the most suitable for use in calibration. Firstly, many models might be able to recreate the observed crime patterns, but this does not guarantee that any of them correctly represent the internal dynamics of the system. This is known as the identifiability problem and one that is common to all modeling approaches – see, for example, the discussion in Windrum et al. (2007). To determine which, if any, are "correct," it might be necessary to simultaneously calibrate against various different data sources that capture elements of the system other than simply the crimes committed. Examples might include the use of social surveys to represent victim behavior, crowd-sourced data to explore "normal" day-to-day behavior patterns or transport data to estimate the routes that people use to navigate cities. The "correct" model will fit the patterns illustrated by these data *and* closely approximate the observed crime data.

Although there is huge scope for improving the calibration of criminology simulation models, relatively little has been done in practice. This is not, however, unexpected. The methods employed are still in their infancy, relative to their traditional mathematical counterparts at least, so it takes some considerable effort to develop a model in the first place. However, as the methods become more widely used and the tools to develop them become easier to manage, there is no reason that standard, widely adopted approaches to calibration cannot emerge.

Related Entries

- ▶ Agent-Based Models to Predict Crime at Places
- ▶ Spatial Models and Network Analysis

Recommended Reading and References

Andresen M, Malleson N (2011) Testing the stability of crime patterns: implications for theory and policy. J Res Crime Delinq 48(1):58–82

Bailey T, Gatrell T (1995) Interactive spatial data analysis. Prentice Hall. http://www.amazon.com/Interactive-Spatial-Analysis-Trevor-Bailey/dp/0582244935

Birks D, Townsley M, Stewart A (2012) Generative explanations of crime: using simulation to test criminological theory. Criminology 50(1):221–254

Chainey S, Ratcliffe J (2005) GIS and crime mapping, 1st edn. Wiley, Chichester

Clark PJ, Evans FC (1954) Distance to nearest neighbour as a measure of spatial relationships in populations. Ecology 35:445–453

Costanza R (1989) Model goodness of fit: a multiple resolution procedure. Ecol Model 47:199–215

Darwin C (1859) On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London

Dray A, Mazerolle L, Perez P, Ritter A (2008) Policing Australia's 'heroin drought': using an agent-based model to simulate alternative outcomes. J Exp Criminol 4(3):267–287

Groff ER (2007) Simulation for theory testing and experimentation: an example using routine activity theory and street robbery. J Quant Criminol 23(2):75–103

Harland K (2008) Journey to learn: geographical mobility and education provision. PhD thesis, School of Geography, University of Leeds, Leeds

Hayslett-McCall K, Qui F, Curtin KM, Chastain B, Schubert J, Carver V (2008) The simulation of the journey to residential burglary. In: Liu L, Eck J (eds) Artificial crime analysis systems: using computer simulations and geographic information systems. Information Science Reference, Hershey

Kirkpatrick SS, Gelatt CDC, Vecchi MPM (1983, May) Optimization by simulated annealing. Science 220(4598):671–680

Knudsen DC, Fotheringham AS (1986) Matrix comparison, goodness-of-fit, and spatial interaction modeling. Int Reg Sci Rev 10:127–147

Kuhnert M, Voinov A, Seppelt R (2005, August) Comparing raster map comparison algorithms for spatial modeling and analysis. Photogramm Eng Remote Sens 71(8):975–984

Levine N (2006) Crime mapping and the crimestat program. Geogr Anal 38:41–56

Liu L, Wang X, Eck J, Liang J (2005) Simulating crime events and crime patterns in RA/CA model. In: Wang F (ed) Geographic information systems and crime analysis. Idea Publishing, Reading, pp 197–213

Malleson N (2010) Agent-based modelling of burglary. PhD thesis, School of Geography, University of Leeds, Leeds

Malleson N, Heppenstall A, Evans A, See L (2010) Evaluating an agent-based model of burglary. Working paper 10/1, School of Geography, University of Leeds, Leeds. http://www.geog.leeds.ac.uk/fileadmin/downloads/school/research/wpapers/10_1.pdf

Malleson N, See L, Evans A, Heppenstall A (2012) Implementing comprehensive offender behaviour in a realistic agent-based model of burglary. Simul Trans Soc Model Simul Int 88(1):50–71

Malleson N, Heppenstall A, See L, Evans A (2013) Using an agent-based crime simulation to predict the effects of urban regeneration on individual householdburglary risk. Environ Plan B Plan Des 1–28. http://www.envplan.com/abstract.cgi?id=b38057

Metropolis N, Rosenbluth AW, Teller E, Rosenbluth MN, Teller AH (1953) Equation of state calculations by fast computing machines. J Chem Physica 21(6):1087–1091

O'Sullivan D, Unwin DJ (2003) Geographic information analysis. Wiley, Hoboken, NJ

Openshaw S (1984) The modifiable areal unit problem. Concepts and techniques in modern geography. Geo Books, Norwich

Reeves C (1995) Modern heuristic techniques for combinatorial problems. McGraw-Hill, London

Reeves CR, Rowe JE (2003) Genetic algorithms – principles and perspectives: a guide to ga theory. Kluwer, Dordrecht

Rykiel EJJ (1996) Testing ecological models: the meaning of validation. Ecol Model 90:229–244

Windrum P, Fagiolo G, Moneta A (2007) Empirical validation of agent-based models: alternatives and prospects. J Artif Soc Soc Simul 10(2)

Cambridge-Somerville Youth Experiment

Sonia Jain and Alison K. Cohen
Health and Human Development Program,
WestEd, Oakland, CA, USA

Synonyms

Cambridge-Somerville youth study

Overview

The Cambridge-Somerville Youth Experiment was a longitudinal study of 650 boys that began in 1935 in Cambridge and Somerville, two neighboring towns in the Boston, Massachusetts, metropolitan area. It was the first documented criminology experiment of that scale (Farrington 2003). A treatment group of at-risk "difficult" boys and control group were created, each with 325 boys assigned based on 160 matching characteristics. The treatment group received social welfare services including family guidance, medical and academic assistance, coordination of community agencies, and recreational services, provided by a friendly counselor sustained for almost 10 years; the control group received no special services. The goal of the study was to examine if delinquency could be prevented and if counseling was an effective approach to preventing delinquency (Powers and Witmer 1951). However, results were mixed, with no crime prevention benefits observed (McCord and McCord 1959), and it has largely been described as both well-intentioned and a failure (Dodge 2001).

Fundamentals

Participants

Boys attending public and parochial schools in Cambridge, Massachusetts, and public schools in Somerville, Massachusetts, who were age six or seven at enrollment were eligible for participation in the study (Powers and Witmer 1951). At the time, Cambridge and Somerville were primarily working-class, factory towns (McCord and McCord 1959). "Difficult" boys were identified by schools and social service agencies in the two towns and referred to the study team (Powers 1949). The selection process was lengthy; it involved interviewing hundreds of boys that had been identified as difficult, conducting medical and psychological exams and home visits, and completing a camping trip in which detailed notes were taken about each boy (Cabot 1940; McCord and McCord 1959). This initial process happened without potential participants' knowledge, but upon being contacted, the majority were willing to participate (Powers 1949). Participants were solicited beginning in 1935, and participants were enrolled between 1937 and 1939 (Powers and Witmer 1951).

Study Design

The study was designed by Dr. Richard C. Cabot and funded by the Ella Lyman Cabot Foundation, which was created in honor of Dr. Cabot's wife. The study was originally designed to be implemented over the course of 10 years, although this plan was curtailed by World War II and lasted for just over 8 years (Powers 1949; Powers and Witmer 1951). Additionally, some boys dropped out sooner, and other boys were intentionally dropped from the sample after 2–3 years if they were definitively classified as nondelinquent (Powers 1949). The study design was a randomized matched study: the two boys closest to each other in terms of approximately 160 background characteristics were paired, and then one of them was assigned to the intervention group, and the other was assigned to the control group on the basis of a coin flip (Cabot 1940; McCord and McCord 1959). These pairs were categorized as likely to be delinquent, not likely to be delinquent, or future delinquency trajectory uncertain (Cabot 1940).

Intervention

The structure of the intervention was to pair each boy in the treatment group with a trained counselor, who would then determine what services were most appropriate for each boy's unique situation with an eye towards positive character development (Cabot 1940). The trained counselor used social welfare practices, including health, educational, family, and community assistance, guidance, and support; boys on average received five such years of counseling (McCord and McCord 1959).

However, the quality of the intervention varied in terms of frequency of contact between the counselor and student, length of time the boy participated in the program, the gender of the counselor, and counselor turnover (McCord and McCord 1959). The most intensive intervention was defined as weekly participation for at least

2 years, in which the counselor and boy had a close relationship that included discussing personal issues; however, only 12 of the boys met these criteria, and so potential analyses were limited but illustrative that this more intense level of treatment may be effective (McCord and McCord 1959). Most boys had done recreational activities with their counselors, approximately half received tutoring, and medical care and summer camps were also common among boys in the intervention group (McCord 2003). Some boys in the intervention group were sent to foster homes and/or private schools per counselors' recommendations (Powers 1949), although this was done after attempting more moderate services (McCord et al. 1960). In comparison, the control group received "the usual services of the community" (McCord and McCord 1959).

Outcomes of Interest

The primary outcome of interest was criminal behavior (as measured by criminal convictions) (McCord and McCord 1959), although additional outcomes were observed in follow-up studies. Other outcomes of interest included mental hospital visits, participation in alcoholic treatment, and vital statistics (e.g., marriage, death) (McCord 1978). Anecdotally, individuals and counselors credited the program with changing boys' developmental trajectories (Powers 1949).

Key Findings

Although boys who participated in the intervention qualitatively credited the program to helping them in their development, there were null and/or harmful effects of the intervention for a range of quantitative outcomes, including criminal and health outcomes (McCord 1978, 2003). For example, there were no statistically significant differences between the intervention and control groups in criminal convictions (both ever being convicted and number of convictions) in their matched analyses and after controlling for additional potential confounders that had been identified more recently (McCord and McCord 1959). After a series of subgroup analyses, though, they found that boys who had met with their adult counselor once a week on average and boys who were younger when they enrolled in the program had statistically significantly lower incidences of criminal convictions compared to their counterparts (McCord and McCord 1959). In later follow-up studies, boys who received the intervention were more likely to have been convicted of "serious street crimes," died at an earlier age, and been diagnosed with a mental illness, compared to the controls (McCord 2003).

Some of the programming characteristics that were later attributed to the study's failure include that the boys were publicly identified as at risk and were given resources without being held accountable, and the boys also participated in activities together (e.g., summer camps) (Dodge 2001). Others noted that adverse effects associated with the intervention were most apparent for boys who participated in the intervention for the longest amount of time, with the most frequent visits, and with family cooperation (McCord 2003).

Additional studies took advantage of the matched pairs to consider subgroups of boys. For example, the boys who were in the intervention group and placed in foster care were compared to their matched pairs, and harmful effects of foster care placement were observed (McCord et al. 1960).

Since the Cambridge-Somerville Youth Experiment, several other programs have been rigorously evaluated, and other delinquency prevention program evaluations have also observed less-than-positive impacts, indicating that more research must be done and published in the academic literature, regardless of positive or negative results (McCord 2003). The Cambridge-Somerville Youth Experiment has also inspired the development of new theories, like deviancy training, to explain what may have occurred then; these theories have since been applied to and tested in new contexts (Gottfredson 2010).

Conclusion

The Cambridge-Somerville Youth Experiment is considered the first major randomized experiment in criminology. It is commonly used to exemplify the importance of rigorous evaluation of well-intentioned programs, including use of evaluation to determine if the positive results desired are actually observed.

Related Entries

► Randomized Experiments in Criminology and Criminal Justice

Recommended Reading and References

Cabot PSD (1940) A long-term study of children: the Cambridge-Somerville youth study. Child Dev 11(2):143–151

Dodge KA (2001) The science of youth violence prevention: progressing from developmental epidemiology to efficacy to effectiveness to public policy. Am J Prev Med 20(1S):63–70

Farrington DP (2003) A short history of randomized experiments in criminology: a meager feast. J Adolesc Res 27(3):218–227. doi:10.1177/0193841X03027003002

Gottfredson DC (2010) Deviancy training: understanding how preventive interventions harm. J Exp Criminol 6(3):229–243. doi:10.1007/s11292-010-9101-9, The Academy of Experimental Criminology 2009 Joan McCord Award Lecture

McCord J (1978) A thirty-year follow-up of treatment effects. Am Psychol 33:284–289

McCord J (2003) Cures that harm: unanticipated outcomes of crime prevention programs. Ann Am Acad Polit Soc Sci 587(1):16–30. doi:10.1177/0002716202250781

McCord J, McCord W (1959) A follow-up report on the Cambridge-Somerville youth study. Ann Am Acad Polit Soc Sci 322(1):89–96. doi:10.1177/000271625932200112

McCord J, McCord W, Thurber E (1960) The effects of foster-home placement in the prevention of adult antisocial behavior. Soc Serv Rev 34(4):415–420

Powers E (1949) An experiment in prevention of delinquency. Ann Am Acad Polit Soc Sci 61:77–88

Powers E, Witmer H (1951) An experiment in the prevention of delinquency: the Cambridge-Somerville youth study. Columbia University Press, New York

Cambridge-Somerville Youth Study

► Cambridge-Somerville Youth Experiment

Camera Surveillance

► CCTV and Crime Prevention

Campus Crime

John J. Sloan III[1] and Bonnie S. Fisher[2]
[1]Department of Justice Sciences, The University of Alabama at Birmingham, Birmingham, AL, USA
[2]School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA

Overview

During the past two decades, campus crime has captured the attention of both the public and researchers who sought to unpack the causes, correlates, and consequences of various types of violent and property crime committed on college and university campuses. Much research to date has focused on identifying how students' lifestyles and daily routines create opportunities for on-campus victimization. Other studies have examined college campuses as places and explored how their features create opportunities for crime. This entry examines both lines of research and synthesizes the extant literature to describe the dynamics of campus crime. The entry also presents some of the controversies within the study of campus crime, including measurement issues, overreliance upon case studies of single campuses, and the problem of underreporting of crimes occurring on college campuses. The entry concludes with a brief discussion of open questions in the study of campus crime, including the question of the extent of campus crime, how to better translate research into policies or programs aimed at reducing campus crime, and the ultimate utility of legislation at the federal and state level designed to reduce campus crime.

Introduction

America's colleges and universities present an apparent anomaly: physically attractive places fostering the intellectual development of young adults that may, in reality, be dangerous places beset by violence, vice, and victimization. More specifically, the seemingly safe appearance of college campuses has been coupled with a conventional view held by many campus administrators, students, and their parents that campus violence, vice, and victimization are merely "youthful indiscretions" resulting from students' newfound personal freedoms, including sexual promiscuity, alcohol consumption, and recreational use of illegal drugs.

These views, however, changed during the 1990s when four groups – a student advocacy group called Security On Campus, campus feminists, campus crime victims and their families, and public health researchers – made claims that crime on US college campuses was not only increasing but becoming more serious and deadly in its consequences. Collectively these groups convinced the public, Congress, and several state legislatures that the "dark side" of the ivory tower was a "dangerous place" rife with "heinous crimes" that not only undermined students' safety and well-being but also threatened the core educational mission of higher education. As a result, during the early 1990s, Congress and multiple states passed statutes designed to address campus crime. For example, Congress passed the *Student Right to Know and Campus Security Act of 1990* (20 USC 1092[f]) (now known as the *Jeanne Clery Disclosure of Campus Security Policy and Campus Crime Statistics* or the *Clery Act*) which mandates that all Title IV schools (those eligible to participate in federal financial aid) not only publicly report their crime statistics in an annual report but also to implement policies and programs designed to address campus crime, in particular sexual assault against college coeds (Sloan and Fisher 2011). Additionally, nearly one-third of the states passed campus crime-related legislation with wide-ranging requirements (Sloan and Shoemaker 2007).

Looking at 2009 campus crime statistics compiled by the United States Department of Education (2011a) in fulfillment of the *Clery Act*, there were over 47,000 violent and property crimes reported on college campuses in the United States. The total included 31 murders and non-negligent manslaughters; over 3,300 forcible and non-forcible sex offenses; more than 4,600 robberies; approximately 4,900 aggravated assaults; more than 26,000 burglaries; over 7,200 motor vehicle thefts; and more than 700 incidents of arson. While the magnitude of these numbers is alarming, to put them into perspective, in 2009, over 20 million students were enrolled at 6,883 Title IV designated post-secondary institutions that were required to file annual campus crime reports with the federal government (United States Department of Education 2011b).

As the public began perceiving campus crime as a "serious" problem, a number of researchers started exploring the patterns, correlates, and causes of campus crime. During the past 20 years, this body of research can be classified as having two different, yet complementary, foci: studies whose primary interest was describing and explaining patterns of victimization relating to *individuals* and studies whose interest was understanding how the *characteristics of college campuses* such as physical layout and geography explained victimization patterns. In the former instance, the studies examined students' victimization patterns by different types of personal and property crime, including recent analyses of cyber-based victimizations suffered by students. Identifying individual-level behavior that apparently increased the risk for victimization led researchers to examine how students' lifestyles and daily routines played a significant role in increasing (decreasing) their risk of victimization. In the latter instance, researchers adopted a perspective which emphasized the college campus as a specific place with distinctive physical and social characteristics that contributed to victimization. Social scientists – including sociologists, criminologists, and geographers – adopting this perspective argued that a *criminology of place* is not only possible, but is a valuable theoretical and practical lens through which researchers and crime prevention practitioners alike can examine and understand campus crime and, more importantly, take steps to prevent its occurrence (Eck and Weisburd 1995).

This entry discusses the interrelationship between the characteristics of colleges and universities and students' lifestyles and daily routines that provide ample opportunities for on-campus personal and property victimization. Campuses are unique places characterized by a distinctive, yet somewhat stereotypical, built environment that helps to structure the daily routines of places and ultimately influences students' daily lifestyles and ensuing crime patterns. As discussed in more detail below, the patterns of the campus environment-student lifestyle nexus enhance opportunities for different types of crimes to occur more frequently than others.

The entry begins by discussing the notion of the college campus as a "place." Relevant literature is then presented that addresses how the characteristics and daily routines of places either may facilitate or retard crimes occurring at or near them. Next, what is known about the relationship between the characteristics of college campuses and crime occurring there is presented. The entry concludes by exploring important controversies and questions that remain concerning campus crime, especially those arising from the perspective of the criminology of place.

Background: The College Campus as a Place

To draw a detailed portrait of offenses occurring on campuses, one must first understand that these events occur at a specific *place*. Further, one must understand that the characteristics and daily routines of places are not simply physical locations where illegal activities occur. Rather, one must appreciate that the characteristics and routines of a place provide important ingredients necessary to create opportunities for victimization.

Geiryn (2000, pp. 464–465) described a place – such as a college campus – as possessing several salient features. First, a place occupies geographic location (a "unique spot in the universe") the size of which can vary tremendously. Many land-grant public universities occupy a dozen or more square miles of space, while so-called virtual universities exist in a single building whose carbon footprint is several 1,000 sq ft. Places, such as college campuses, also possess material form "[a] compilation of things or objects" (Gieryn 2000, p. 465). When people describe a particular college campus to others, they may do so by commenting on the particular architectural features the campus possesses such as a large bell tower atop the student union or in terms of natural features such as being adjacent to mountains or having a stream running through it. Places like college campuses also have *physicality* to them, including geographic markers (e.g., logo signage; staffed entry gates; the hallmark student union) that uniquely characterize a place's physical space and functions, and signal to visitors that they have entered a place separate and distinct from the adjunct community.

The physicality of a college or university campus is important because ". . . social processes happen *through* the material forms" that comprise the place (Gieryn 2000, p. 465). On campuses, daily interactions occurring among students, faculty, staff, and visitors take place both inside and outside of offices, classroom buildings, residence and dining halls, stadia, libraries, pathways, streets, parking structures, and green spaces. These interactions are the result of lifestyles or daily routines that each of these campus community members takes on as part of their respective purpose or role while on campus. These interactions and lifestyles – which are structured by physical spaces and their layout – create opportunities for a variety of violent (e.g., rapes, assaults) and property (e.g., automobile thefts, burglaries) victimizations to occur.

Finally, a college campus is invested with meaning and value. When individuals walk by a set of buildings consisting of a large residence hall, a library, a classroom building, and green space where young adults are throwing Frisbees, reading books and connecting to the internet, lounging in the sun, or playing touch football, most would not think "oh, that is a prison." Some scholars have argued that the meanings created by and resulting from the material form of a place prompt humans to think of these places as *communities*: locations where people with similar interests purposely gather to interact with one another (e.g., Smith 2002). Annually, college campuses draw groups of individuals who share interests. Researchers are drawn to college campuses because they house needed facilities and resources. Faculty members are drawn to the campus as a place where they can teach young adults and not only help shape their minds, but train them to enter an increasingly complex and competitive workforce. Those seeking positions as support staff are drawn for the opportunities the campus presents to find meaningful employment. Ultimately, those drawn to the campus – despite disparate reasons for arriving there – interact with one another and create a community (Bromley 2007).

The Routines of Place and Their Patterns

Places also have identifiable routines or rhythms that vary by time of day, day of the week, and season of the year which are shaped by both the built environment – manmade structures and features – and the natural environment. The routines of place include not only the movement of people into or away from it at regular intervals during the day but also the myriad of interactions that are facilitated by the features of the place. In effect, places have *patterns* associated with them.

Brantingham et al. (1995) have suggested that the patterns or routines that characterize college campuses can be conceptualized as occurring at different levels of spatial analysis. At the *macro level*, campus routines include vehicular and pedestrian traffic coming onto or exiting the campus and students moving *en mass* across the campus as one set of classes end and another begins from early morning to late evening. At the *meso level*, a particular building – for example, the student union – has *its* own routines and patterns which vary by time of day, day of the week, and season of the year. For example, the student union may be overrun with students during meal times each day during an academic term, but relatively quiet at other times of day when students are studying in their residence halls or in the library. Students' presence at the union on weekends may be reduced when visitors come to enjoy concerts, movies, or lectures. On some weekends, especially during final exams or the summer, the union may be essentially vacant of both students and visitors.

Finally, at the *micro level* are the routines associated with a *single room* within a building such as "Dr. Jones' laboratory" or "Ms. Smith's dorm room." Again, variation in the activities occurring in or around the lab or dorm room takes place according to time of day, day of the week, or academic term. The lab might be very busy in the morning but empty at night while just the opposite might be true of the dorm room. At the beginning (or end) of a term, the lab and the dorm room would be a beehive of activity, while during various breaks in the academic year, each may be occupied by only a handful of students.

Conceptualizing the routines of place as occurring at different levels can be useful to understanding the crime patterns of a place, including college campuses. On a college campus, identifying macro level crime patterns helps to understand the "big picture" of crime at that particular place. What is the most commonly occurring offense on the campus? Are there particular areas of the campus where crime tends to be concentrated – so-called hot spots? If so, where are these located? At the meso level, crime patterns can also be identified for a much smaller part of the larger place. How many rapes or burglaries occurred in the undergraduate residence halls during the fall semester? Which parking garage had the most auto break-ins? Finally, at the micro level, one can again assess patterns that may be associated with a single residence hall room, laboratory, office, or classroom. For example, several years ago at UAB, a particular computer laboratory in one of the older buildings on campus was repeatedly broken into over the course of several months and most of the equipment was stolen. By analyzing evidence at the site of the break-ins, the campus police learned the thefts were generally confined to very early morning hours, that homeless people had been seen climbing through unlocked windows into the building, and that the security devices protecting the equipment had been disabled from the inside.

The patterns associated with the routines of place have both theoretical and practical importance for understanding crime and developing sound situational interventions to address it. Research has consistently shown, for example, that crime is *not* randomly distributed either spatially or temporally but rather tends to cluster by place and in time (Eck and Weisburd 1995). When crime clusters at certain locations, these are known as "hot spots" (Sherman 1995). The most basic hot spot is a place or area that has a greater than average number of crime or disorder events, or where targets have higher than average risk of victimization. Places include specific addresses, street corners, or other small locations that can be seen by a person, such as a house, convenience store, or drug dealing location. As a result of the groundbreaking "hot spots" work by Sherman and others over the past several decades, combined with developments in spatial and temporal analysis of crime, criminologists developed a new theoretical orientation that instead of focusing on individuals and their behaviors focuses on places – the "criminology of place" – to underscore that opportunities for crime are shaped by the characteristics of places, individuals' lifestyles and daily routines, and time of day. Further, places – like individual offenders – have "careers" associated with them including onset, continuance, specialization, and desistance (Sherman 1995). A particular hot spot was not always "hot." It develops over time based on how its features – both physical and symbolic – and routines draw prospective offenders and targets to it (onset). The place continues as a hot spot for some period (continuance), which may be years, and may disproportionately involve certain crimes such as armed robbery (specialization). Finally, because of direct police intervention, neighborhood changes, or other factors, the hot spot desists as a place disproportionately responsible for crime (desistance).

On a practical level, understanding the dynamics of place including how they are organized for crime to occur and how the routines of place at the macro, meso, and micro level bring together would-be offenders and potential targets can inform primary prevention measures to reduce the probability of crime occurring. Strategies such as target hardening, changing the physical design features of a place, having individuals change their daily routines (e.g., not carrying large amounts of cash on their person; insuring valued goods stored in a vehicle are not left in the open), or altering their lifestyles (e.g., not getting drunk, not consuming illegal substances) when frequenting certain types of places (e.g., residence halls) can help to change the routines of a place and patterns of behavior and therefore might reduce the opportunity for victimization.

College Campuses and Opportunities for Crime: Institutional, Demographic, and Lifestyle-Routine Characteristics

The institutional, demographic, and lifestyle characteristics of colleges and universities create untold opportunities for violent and property victimization for the nearly 20 million students and the millions of faculty and staff, and countless visitors to college campuses each year. Below, these characteristics and their contribution to crime on college campuses are examined.

Institutional and Demographic Characteristics

Although the more than 6,000 Title IV 2- and 4-year post-secondary institutions differ widely across their institutional characteristics, researchers have identified a set of characteristics strongly correlated with official crime rates – crimes known to and recorded by the campus police or security (Fernandez and Lizotte 1995; Bromley 1994; Sloan 1994). Schools with large enrollments, for example, have higher rates of on-campus rape, assault, robbery, burglary, and larceny than schools with smaller enrollments. Campuses with large student enrollments apparently provide more opportunities for victimization because they: (1) possess more targets for would-be offenders, and (2) allow for greater anonymity among potential offenders within the student body.

Research also shows that campuses with a large percentage of full-time students have higher official rates of assault, robbery, burglary, and motor vehicle theft. One explanation for this is that full-time students tend to spend more time on campus than part-time students which increases their exposure to potential offenders and provides a larger number of targets from which offenders may choose.

Residential campuses – schools that have a large number of students living in on-campus residence halls or apartments – have higher rates of rape and larceny whereas commuter schools have higher rates of motor vehicle theft and robbery. In contrast, rural campuses, on average, have lower rates of all forms of violent (rape, assault, and robbery) and property crime (burglary, motor vehicle theft, and larceny) than do urban campuses.

One of the strongest and most consistent predictor of official rates of campus crime is the cost associated with room and board per academic year. Campuses with higher room and board costs generally report higher rates of crime whose motivation is economic, for example, robbery, burglary, motor vehicle theft, and larceny. Students enrolled at "costly" schools might also have plenty of attractive, high-quality targets (e.g., electronic devices and their accessories; designer clothing) to steal with relatively little risk for the offender. Further, research shows that most students while on campus do *not* routinely engage in proven crime prevention behaviors that reduce their chances of victimization (Fisher et al. 1997).

Research further indicates there is a demonstrable relationship between students' demographic characteristics and campus crime rates (Fernandez and Lizotte 1995; Bromley 1994; Sloan 1994). Since the 1940s, the demographic composition of colleges and universities has changed dramatically, with total student enrolled increasing annually. In particular, the number of women and minorities enrolled at post-secondary degree-granting institutions has seen tremendous increases, especially over the past three decades. In 2009, approximately 55 % of the 20 million students enrolled at all post-secondary institutions were women (United States Department of Education 2011c) while some 32 % of students enrolled at degree-granting post-secondary schools in 2009 were racial and/or ethnic minorities (Rigg 2010).

Several studies find that the percentage of male students enrolled at a college or university consistently is linked with campus crime rates, such that the more males enrolled, the higher the overall crime rates (Fernandez and Lizotte 1995; Bromley 1994; Sloan 1994). Research also shows that colleges and universities with higher percentages of minority students also have higher rates of violence reported to officials. A similar relationship is found between the percentage of African-American students enrolled and rates of campus assault, robbery, and burglary.

Understanding the effects of institutional and students' demographic characteristics allows post-secondary administrators to determine whether campus crime rates at their school are typical for or similar to comparable campuses or due to other factors, perhaps students' lifestyles and routines. For example, some campuses with large student enrollments may have "naturally" high crime rates. A problem arises when campuses with *high* crime rates have institutional and demographic characteristics associated with *low* crime rates. This requires exploring how the physical design features of the campus as a place potentially interact with the lifestyle-routine activities of campus members to see how those interactions provide opportunities for victimization.

Campus Places

College campuses occupy finite geographical areas that vary in size. Campuses typically are divided into smaller parcels designated for various uses including housing, instruction, research, common areas, entertainment and leisure, etc. These spaces also are usually arranged in a way such that the space utilized for one function does not mix with space designated for another. Collegiate residence halls or apartments for married students, for example, would ordinarily not be located directly adjacent to a football stadium or research laboratories where exotic and potentially lethal viruses are being studied. Rather, residence halls tend to be located near one another and away from other space clusters designed for different uses.

Arranging common usage of space on the campus is conducive to creating daily interactions among faculty members, students, staff, and visitors. There may be several such segmented areas on a campus depending on its age and needs, or as new needs arise or priorities change to account for expansion of student and faculty needs. However, for reasons ranging from the utilitarian to the aesthetic, the general pattern is that similarly designated spaces tend to cluster together in space.

This segmentation of a campus into areas of designated usage has implications for understanding the spatial and temporal distribution of campus crime. A college or university with several thousand students living on campus will have many residence halls and student apartments to accommodate housing demand. These, in turn, will likely be clustered together and several such clusters will be located around campus. Buildings that primarily contain classrooms or research laboratories will likewise be clustered together, rather than randomly placed, and again may be found in several locations around campus. Parking lots and decks are typically located on the edges of the campus, especially if the campus is "pedestrian friendly" and vehicular access is restricted to the fringes of the campus.

Clusters of similarly utilized space provide differential opportunities for offenders. For example, a five-story parking deck is a far more likely candidate for automobile theft or automobile break-ins than is a building housing research laboratories where theft of property, such as laboratory equipment, would be likely more common. Offenders may perceive that a cluster of dormitories is a good target for burglary, but less so for crimes against the person such as robbery. A small, isolated parking lot on the fringe of campus may be perceived as a good target for robberies but not for automobile theft, due to the limited number of vehicles actually parked there. Ultimately, spatial distribution and space utilization of a campus has significant implications for target selections by prospective offenders.

Additionally, clusters of similarly utilized spaces have their own routines at various points during the day or evening which likely differs from the routines associated with other clusters. During a typical day, there will likely be specific times when clusters of buildings, parking lots and decks, residence halls, etc. are "busy" with students, faculty, or staff coming and going. There are specific times during the day or at night when such spaces are relatively quiet. These "down-periods" during the 24-h cycle of the day present opportunities to offenders since few people are present to serve as guardians for the places that prospective offenders may perceive as attractive targets. Large parking decks or lots while classes are in session likely mean few people would be present, which increases opportunities for vehicular thefts or break-ins. As a result, because of variation in the routines associated with clusters of differentially utilized spaces, prospective offenders are likely to perceive these clusters – and the individuals typically found there – as differently convenient targets. Thus, not only the spatial distribution of spaces and their usage but also the temporal patterns associated with those spaces matter in terms of understanding campus crime.

Campus Routines and Lifestyles

The routine, everyday activities of students, faculty, staff, and visitors have significant influences on their chances of becoming a victim of crime while on campus. Although its population is moderately transitory, college campuses also have constant and familiar activities and routines.

The lifestyle-routine activity approach to understanding crime victimization describes how one's daily activities and lifestyles converge in time and space with a would-be offender to create opportunities for criminal victimization (Cohen and Felson 1979). Four concepts are central to this approach for explaining how and why victimization can occur: attractiveness of the target; proximity of one or his/her property to a motivated offender; guardianship over a person or his/her property; and exposure of his/her property to a potential offender. Understanding how each of these concepts plays out on campus broadens one's understanding of the mechanisms underlying the occurrence of crime on college campuses.

Students and Victimization. Students' pursuit of education and recreation has them on campus at all hours of the day, every day of the week. They also experience a variety of violent and property victimizations while on campus. Fisher et al. (1998) found that nearly one-fourth of the students in a national sample had been victimized at least once while on campus during the current academic year. Campus victimizations occur in a wide array of situational contexts that are unique to on-campus locations, compared to those that are off campus. For example, Hart and Miethe (2011) have shown that the specific attributes of "minor assaults of male victims in daytime hours" were far more likely to occur in on-campus contexts than off campus.

Students who live on campus spend a considerable amount of time in their residence halls. These provide not only shelter and food, but they are also places where daily relationships and routines are established (and broken). Living in a residence hall, especially those that are high-density with little supervision as to who enters or exits, increases opportunities for victimization by increasing the number of suitable targets and exposure of them to motivated offenders, while reducing the level of guardianship over person and property. Fisher and her colleagues (1997) reported that of the student victims of crime who lived on campus in undergraduate or graduate housing, or in a fraternity or sorority, 86 % of them lived in an undergraduate residence hall. Identifying on-campus locations where incidents had occurred revealed that crimes of violence, vandalism, threats, and harassment occur most frequently in students' living quarters.

Students' lifestyles and routines also lend themselves to opportunities for victimization every day of the week around the clock. Each year students bring attractive, portable, lightweight targets to campus with them: portable computers, electronic pads and tablets, cell and smart phones, calculators, backpacks, books, roller blades, skateboards, and bicycles. Students carry many of these items with them while they attend classes or social events. Many students drive motor vehicles to campus daily or park vehicles on or near campus during the duration of the academic term. Concurrently, they are all too often inattentive guardians of their property (or their person) and leave residence hall, office, or vehicle doors unlocked or propped open; leave bicycles unlocked; or just walk away from their property at the library, student union, or classroom for "just a few minutes." As a result, dormitories, bicycle racks, libraries, student unions, or classrooms can all become "hot spots" of crime because students are typically poor guardians of their property. Moreover, students leave the campus *en mass* at the end of an academic term or for long periods during holiday breaks, which leaves the campus filled with many attractive targets without guardianship.

The college years are also notorious for an active social life for many students. They may attend weekly parties; drink habitually and heavily or experiment with drugs; and attend entertainment or cultural events. All of this activity occurs in places on campuses where students naturally congregate. Such lifestyles and routines have, in turn, been linked to increased probabilities for on-campus victimizations. The Harvard Alcohol Study found that students who had been the victims of crime on campus reported significantly more frequent drug and alcohol use than students who had not been victimized (Dowdall 2009). The number of nights spent on campus partying and the likelihood of regularly taking recreational drugs during the past year also have been reported to be significant predictors of students' on-campus violent victimization (Fisher et al. 1998). Given their 24–7 on-campus lifestyles, it is not surprising that a temporal pattern concerning on-campus student victimization shows that violent crimes were likely to have occurred in the early morning hours, in particular between 12:00 a.m. and 2:00 a.m. (Fisher et al. 1998).

Faculty, Staff, and Victimization. Faculty members, like students, also have routine activities on college campuses that increase their risk for victimization. Faculty members, especially untenured assistant professors, are frequently on campus during the week and on weekends, sometimes spending long hours into the night and early morning working on research, grading papers, or writing manuscripts. Many spend considerable amount of time in their offices or laboratories, or walking to and from classes, meetings, the library, or a seminar. Most faculty members work alone in their office even when support staff have gone home for the day. Many faculty members also teach night courses. Because of their routines, many faculty members are exposed to a large number of students every academic term. Some faculty members even socialize with students outside of class in the student union or other establishments on or near campus.

Faculty members also have valuable, attractive targets such as laptop computers, multimedia equipment, laser printers, laboratory equipment, secured data, books, motor vehicles, and other personal items. Faculty can also be poor guardians of their property. At times, they leave offices or laboratories unlocked or computers unsecured between classes or meetings, and may prop open doors to buildings, offices, or laboratories so as not to have to fetch keys or recall a lock combination. With the routine arrival of new faculty members each academic term, while others are on leave or sabbatical during the school year, access control to buildings and offices may be severely compromised on a daily basis.

Wooldredge, Cullen, and Latessa (1995), in a study of faculty members' victimization experiences at the University of Cincinnati West campus (all colleges except the medical and nursing colleges), reported 27 % of faculty members surveyed reported they had experienced at least one property victimization on campus (burglary, stolen property, or damage to property) during the academic year, while 5 % reported they had experienced at least one personal victimization (robbery, assault, sexual assault, or assault with a deadly weapon). Those whose offices are not within shouting distance of their colleagues; who do not teach in buildings where their office is located; and whose office is in a comparatively non-secure building on campus are more likely to experience an on-campus property crime. Faculty members who spend more time on campus after hours and on Saturdays, who walk alone on campus more frequently (other than going to class), and who socialize with students outside class are also more likely to experience personal victimization while on campus.

Staff members typically work 8-h shifts, primarily Monday–Friday and during daylight hours. Like faculty, they too possess attractive targets. To date, there are no published studies on the extent and nature of victimization among staff working on college campuses. Anecdotal evidence, however, suggests that their patterns of victimization may be similar to those of faculty.

Visitors to Campus and Victimization. The routine activities of visitors to campus are somewhat different from those of students, faculty members, and staff. Every term, prospective students and their parents frequently meander around the campus. Students from other campuses visit with friends; sports fans attend athletic events or hold tailgate gatherings; and cultural buffs frequent movies, concerts, art galleries, and plays. Delivery trucks filled to the brim with products and packages are daily sights on campus. Visitors bring plenty of property with them that is attractive to would-be offenders – cash, portable electronic equipment, expensive jewelry and watches, motor vehicles. Like staff, no studies of campus visitors have been published but it is likely that students who visit friends and enjoy activities where alcohol and drugs are plentiful and visitors who attend sporting events and enjoy pregame, during-the-game, and post-game celebrations might experience a pattern of victimization similar to students who actively engage in an active "party lifestyle."

Tracking the Spatial and Temporal Distribution of Campus Crime

Crime on campus, like crime in other locations, is not randomly distributed in space or time. Robinson and Roh (2007) illustrated this point in an analysis of crime on the campus of Appalachian State University (ASU) for the period 2004–2005 using calls-for-service data stored in campus police crime logs. Analyzing those data revealed that the four most commonly reported offenses on the ASU campus over the 2-year period of study included alcohol violations, theft of property, illegal drug possession, and vandalism (Robinson and Roh 2007, p. 239). More importantly, they determined that hot spots on the ASU campus included locations at or near: (1) student residence halls, (2) high-traffic areas between the main parts of the campus such as near classroom buildings and parking lots near pedestrian tunnels, and (3) specific locations for drug violations, alcohol violations, breaking and entering, and assaults/sexual assaults/forcible fondling. These "crime attractors" or "crime generators" remained relatively stable over the 2 years of study and possessed certain features. For example, one hot spot included two residence halls located within walking distance to the football stadium, its parking lots, and an open field which students use for recreational activities including consuming alcohol, playing sports, and hosting parties. Robinson and Roh also reported that residence halls that were "hot spots" for various crimes tended to be older residences located in the heart of the campus. Compared to these "older" residence halls, newer residence halls' distance from the center of the campus made them both more physically and socially isolated and apparently less vulnerable to opportunities for victimization.

In a similar study, Resler (n.d.) examined the spatial distribution of crime on the campus of the University of Texas at Austin using police calls-for-service data from 1998. Using geographic information system (GIS) crime mapping software and spatial analysis models, Resler (n.d.) found that ". . . concentrations of crime occur[ed] predominantly around the periphery of campus [and] the characteristics of these high crime areas, such as location and density, influence victimization." Resler concluded that use of GIS software to map crime distributions and concentrations was an effective tool, and its use opened new avenues for research.

Brower and Carol (2007) examined the temporal and spatial distribution of reported incidents of liquor law violations, assaults and batteries, vandalism, and noise complaints during 2003 on the campus of the University of Wisconsin-Madison and the City of Madison. They also used GIS crime mapping technology and tracked the movement of different crimes through the city of Madison by time of day, and investigated the relationship between these spatial and temporal movements of crime and the proximities of various student and nonstudent neighborhoods to each other and to high-density bar areas. They found different categories of crime presented different temporal and spatial patterns. For example, serious crimes such as assaults tended to peak between 2:00 a.m. and 3:00 a.m. which coincided with bar closing time. Less serious crimes, such as noise violations, peaked between 11:00 p.m. and 12:00 a.m. Vandalism tended to peak during the late morning and early afternoon. "Hot spots" for the different crimes moved throughout the downtown by time of day. Results of the study were then used by University of Wisconsin and City of Madison officials to implement changes to address high-risk drinking behavior.

Controversies

Because the study of crime on college and university campuses is, relatively speaking, still in its infancy, controversies over study results are bound to arise. Below, a few of the more important questions that have been raised with these studies are presented.

One significant controversy in the literature concerns wide-ranging estimates that have been produced concerning the extent college women experience various forms of sexual violence during their college tenure. Part of the explanation for disparate estimates of college women's victimization experiences has to do with how questions about rape, sexual assault, stalking, and other forms of violence are framed. Generally, the broader the question about sexual victimization asked of college women, the larger are the estimates of sexual violence. Fisher and her colleagues (2000) were among the first to document how specific, behaviorally based questions relating to sexual violence against college women significantly affected estimates for victimization. As a result of their work, studies now are more apt to use narrow, behaviorally based questions in surveys of college women to explore the extent of sexual victimization perpetrated against them.

A second controversy is the predominance of research using case studies of crime occurring on a single campus. Typically, one or more researchers will survey students (or faculty or staff) enrolled at a single school or use official data compiled by a single school to examine the causes, correlates, and consequences of crime occurring on campus. While many of these studies have contributed insight into the dynamics of campus crime, the problem is that results from such case studies cannot be generalized to colleges and universities more generally. The works of Fisher and her colleagues (1998, 2000) are notable exceptions; both projects involved national samples of college students.

A final major controversy revolves around the utility of officially compiled data collected by Title IV eligible post-secondary institutions in compliance with the mandates of legislation such as the federal *Clery Act*. Among other requirements, the legislation mandates that colleges and universities compile statistics on crimes reported to campus authorities for the following offenses: murder and non-negligent homicide; rape; robbery; assault; burglary; automobile theft; and arson. One weakness is that the *Clery Act* fails to include what is generally regarded (based on victimization surveys) as the most commonly occurring crime on campus, namely, theft or larceny. A second problem, again based on victimization surveys conducted on college students, is that large numbers of offenses occurring on campus go unreported to campus law enforcement or other authorities. As a result, those offenses are never included in the crime statistics published as part of *Clery Act* requirements, which means *Clery Act* statistics grossly underestimate the volume of crime occurring on the nation's college campuses (Fisher et al. 1998).

Unless and until these controversies are settled, contributions from studies of campus crime will likely not have the impact they might, were these issues to be addressed. In particular, campus administrators who may be relying upon extant literature to help them craft responses to crime occurring on *their* campuses may not develop as effective a response as they might were the controversies settled.

Open Questions

Given the above controversies, several questions remain open to exploration, comment, and debate. The first is rather basic: How much crime occurs on college campuses each year? Currently, the answer to that question depends on where one looks: case studies (from which national-level extrapolations might be developed); "official" data (such as those compiled in compliance with the *Clery Act* but which, as discussed above omit a major type of crime and suffer from underreporting); or national-level studies (some of which are more than 15 years old).

A second open question is how to better translate the results of academic research on campus crime into practical policies and programs to address the problem and then to evaluate their impact on both crime and fear of crime. Assuming study results used to develop specific programs or policies to reduce campus crime are sound, there is little evidence that programs based on these results are ever evaluated. Further, policies or programs aimed at reducing campus crime too often fail to use sound scientific designs that allow for such necessities as control groups, or designs which rule out rival hypotheses. Criminologists are increasingly advocating the need to use controlled experiments to examine the effectiveness of interventions designed to reduce crime; the same should be true of efforts occurring on college campuses.

Finally, but certainly of no less importance is the ultimate utility of legislation like the *Clery Act* that is designed to reduce/eliminate campus crime. There are actually several problems with campus crime reporting legislation. First, because reliable baseline statistics for the years prior to the passage of *Clery* do not exist, it is difficult to know if the legislation has had a positive, negative, or neutral impact on campus crime. Second, post-secondary institutional compliance with *Clery* has been less than total for a number of reasons, none the least is the amount and type of resources necessary to compile and report the data required by the law. Finally, there is enough evidence to indicate that the very consumers whom the law was passed to serve (particularly prospective college students and their parents) are either unaware of the law and the various reports that are to be generated by colleges and universities in compliance with the law's mandates, or if they are aware of these reports, few actually consult them (Gregory and Janosik 2007).

Conclusions

Campuses are neither hot beds of criminal victimization nor are they ivory towers immune from the grim reality of crime. They are *places* whose built and physical environments, institutional and demographic characteristics, space distributions, functional usage, temporal patterns, lifestyles and routines create opportunities for victimization. By developing an orientation that focuses on the criminology of place and on identifying campus characteristics and lifestyles/routines that facilitate opportunities for crime, researchers now better understand how and why campus crime happens. Through identification of crime patterns, researchers also have informed interventions (e.g., banning alcohol on campus or enhancing campus lighting) intended to disrupt the rhythms and routines associated with campus crime.

This work has clear implications for campus policing and campus security, because it is these agencies that will benefit the most from research focusing on the relationship between campus *places* and campus *crime* (Ratcliffe 2002). It is these agencies that routinely collect data that can then be used to identify hot spots and adapt personnel and tactics according to changing conditions. These agencies, through place-based analyses of campus crime, also can work with campus administrators to develop and better enforce policies relating to parking, consumption of alcohol, noise ordinance violations, and other problems.

There is reason for optimism that as research devoted to college campuses as places matures, nagging questions about campus crime will be answered. There is also reason to hope that the answers uncovered will, in turn, inspire the development and evaluation of evidence-based policies and programs to reduce or even eliminate campus crime.

Related Entries

- ▶ Criminology of Place
- ▶ Hot Spots and Place-based Policing
- ▶ Lifestyle Theory
- ▶ Routine Activities Approach
- ▶ Situational Crime Prevention

Recommended Reading and References

Brantingham P, Brantingham P, Seagrave J (1995) Crime and fear at a Canadian university. In: Fisher BS, Sloan JJ (eds) Campus crime: legal, social and policy perspectives. Charles C. Thomas, Springfield, pp 123–155

Bromley ML (1994) Correlates of campus crime: a nationwide exploratory study of large universities. J Secur Admin 17:37–52

Bromley ML (2007) The evolution of campus policing: different models for different eras. In: Fisher BS, Sloan JJ (eds) Campus crime: legal, social, and policy perspectives, 2nd edn. Charles C. Thomas, Springfield, pp 280–303

Brower AM, Carroll L (2007) Spatial and temporal aspects of alcohol-related crime in a college town. J Am Coll Health 55:267–75

Cohen L, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44:588–605

Dowdall GW (2009) College drinking: reframing a social problem. Praeger, Westport

Eck JE, Weisburd D (1995) Crime places in crime theory. In: Eck JE, David W (eds) Crime and place, vol 4. The Police Executive Research Forum, Washington, DC, pp 1–34

Fernandez A, Lizotte AJ (1995) An analysis of the relationship between campus crime and community crime: reciprocal effects? In: Fisher BS, Sloan JJ (eds) Campus crime: legal, social, and policy perspectives. Charles C. Thomas, Springfield, pp 79–102

Fisher BS, Sloan JJ, Cullen FT, Chunmeng L (1997) The on-campus victimization patterns of students: implications for crime prevention by students and post-secondary institutions. In: Lab SP (ed) Crime prevention at a crossroads. Anderson Publishing Co, Cincinnati, pp 101–126

Fisher BS, Sloan JJ, Cullen FT, Chenmung L (1998) Crime in the ivory tower: the level and sources of student victimization. Criminology 36:671–710

Fisher BS, Cullen FT, Turner MG (2000) The sexual victimization of college women. http://www.ncjrs.gov/pdffiles1/nij/182369.pdf. Accessed 4 Dec 2012

Gieryn TF (2000) A space for place in sociology. Annu Rev Soc 26:463–496

Gregory DE, Janosik SM (2007) Research on the *Clery Act* and its impact on higher education administrative practice. In: Fisher BS, Sloan JJ (eds) Campus crime: legal, social, and policy perspectives, 2nd edn. Charles C. Thomas, Springfield, pp 45–64

Hart TC, Miethe TD (2011) Violence against college students and its situational contexts: prevalence, patterns, and policy implications. Victims Offenders 6:157–180

Mustaine E, Tewksbury R (2007) The routine activities and criminal victimization of students: lifestyle and related factors. In: Fisher BS, Sloan JJ (eds) Campus crime: legal, social, and policy perspectives, 2nd edn. Charles C. Thomas, Springfield, pp 147–166

Ratcliffe J (2002) Damned if you don't and damned if you do: crime mapping and its implications in the real world. Policing Soc 12:211–225

Resler L (n.d) "Campus crime at the university of Texas at Austin: a spatial and organizational approach using GIS." http://www.utexas.edu/depts/grg/hudson/grg394k/studentprojects/resler/resler.html. Accessed 4 Dec 2012

Rigg L (2010) "College student enrollments: facts and figures." http://www.chillicious.com/finance/college-student-enrollment-facts-and-figures/. Accessed 4 Dec 2012

Robinson M, Roh S (2007) Crime on campus: spatial aspects of crime at a regional comprehensive university. In: Fisher BS, Sloan JJ (eds) Campus crime: legal, social, and policy perspectives, 2nd edn. Charles Thomas, Springfield, pp 231–255

Sherman L (1995) Hot spots of crime and criminal careers of places. In: Eck JE, David W (eds) Crime and place. Criminal Justice Press, New York, pp 35–52

Sloan JJ (1994) The correlates of campus crime: an analysis of crimes known to campus police and security. J Crim Just 22:51–61

Sloan JJ, Fisher BS (2011) The dark side of the ivory tower: campus crime as a problem. Cambridge University Press, New York

Sloan JJ, Shoemaker J (2007) State level clery Act initiatives: symbolic politics or substantive policy? In: Fisher BS, Sloan JJ (eds) Campus crime: legal, social, and policy perspectives, 2nd edn. Charles C. Thomas, Springfield, pp 102–124

Smith MK (2002) "Community." http://www.infed.org/community/community.htm. Accessed 4 Dec 2012

United States Department of Education (2011a) http://www2.ed.gov/admins/lead/safety/campus.html#data. Accessed 4 Dec 2012

United States Department of Education (2011b) http://nces.ed.gov/fastfacts/display.asp?id=372. Accessed 4 Dec 2012

United States Department of Education (2011c) http://nces.ed.gov/Programs/digest/d08/tables/dt08_190.asp. Accessed 4 Dec 2012

Wooldredge J, Cullen F, Latessa E (1995) Predicting the likelihood of faculty victimization: Individual demographics and routine activities. In: Fisher BS, Sloan JJ (eds) Campus crime: legal, social, and policy perspectives, Charles C. Thomas, Springfield, pp 122–133

Capture Recapture to Estimate Criminal Populations

Peter G. M. van der Heijden[1,2], Maarten Cruyff[1] and Dankmar Böhning[3]
[1]Department of Methodology and Statistics, Utrecht University, Utrecht, The Netherlands
[2]University of Southampton, Southampton, UK
[3]Statistical Sciences Research Institute, University of Southampton, Southampton, UK

Synonyms

Population size estimation

Overview

Methodology is presented that allows to estimate the size of population from a single register, such as a police register of offenders. A capture-recapture variable is constructed from Dutch police records and is a count of the police contacts for a violation. A population size estimate is derived assuming that each count is a realization of a Poisson distribution and that the Poisson parameters are related to covariates through the truncated Poisson regression model or variants of this model. As an example, estimates for perpetrators of domestic violence are presented. It is concluded that the methodology is useful, provided it is used with care.

Fundamentals of Capture Recapture

Introduction

For many policy reasons, it is important to know the size of specific delinquent populations. One reason is that it provides insight into the threat these populations may pose on society. Another reason is that it gives an estimate of the workload of the police.

However, estimating the size of a delinquent population may be problematic for various reasons. Counting the number of crimes from police records may lead to a dark number problem. It may be that the crime is registered but the offender is not known or, as is often the case with victimless crimes, the crime is not registered at all. In victim surveys, people report the number of times they have been the victim of a particular crime, like robbery or burglary. Based on that information, an estimate can be obtained of the total number of these crimes. However, victim surveys do not provide an estimate of the number of offenders, since they usually are unknown to the victim, nor do they provide insight into victimless offenses. Self-report studies can potentially estimate the size of a delinquent population since people are simply asked whether they are member of this type of population. Problems related to self-report studies are the difficulty of obtaining a representative sample, the risk of socially desirable answers, and the need for large samples if offenses are infrequent. For a more elaborate comparison and an overview of the literature on police registers, victim surveys, and self-report studies, the reader is referred to Wittebrood and Junger (2002).

The methodology presented here makes use of a single register. There is a large literature on capture recapture making use of two registers. Estimation based on a single register has two important advantages: first, it does not require the unverifiable assumption that the two sources are statistically independent, and, second, it does not require the elaborate process of linking databases that is often troubled by privacy regulations. Also, often there are technical problems in making correct linkages and in avoiding incorrect linkages. A single register that contains (re)captures circumvents these problems.

In this entry, a way to estimate the number of offenders from police data is discussed. The data are from the Dutch police register system. Offenses committed by a known offender are registered in this system. So for each specific offense, like, for example, illegally owning a gun, an offender-based data set can be constructed that shows the number of illegal gun owners apprehended once, twice, three times, and so on. Note that illegal gun owners who were not apprehended are not part of this offender-based data set. Yet, if their number could be estimated, this would yield an estimate of the total number of illegal gun owners (compare van der Heijden et al. 2003a).

The aim is to estimate the number of offenders never apprehended, using the data about offenders apprehended at least once. These estimates are derived under two assumptions. First, the number of apprehensions is a realization of a Poisson distribution. Second, the logarithm of the Poisson parameter for an offender is a linear function of his covariates. These assumptions are discussed in greater detail at the end of the introduction.

At this point, it is indicated how, using these assumptions, the size of the population never apprehended can be estimated. Consider an offender with a Poisson parameter that gives him a probability of.25 to be apprehended at least once. Suppose that this offender is indeed apprehended, then there are three other offenders with the same Poisson parameter who have not been apprehended. By performing this trick for every offender who is apprehended and adding up all individual estimates, an estimate is obtained of the number of offenders who are not apprehended, and this solves the problem.

The methods employed in this entry originate from the field of biology, where they are used to estimate animal abundance. In these applications, the data are collected at specific time points, and for each animal that is seen at least once, there is a capture history. For example, if there are five capture times, a history could be 01101 if the animal is seen at captures 2, 3, and 5 and not seen at captures 1 and 4. In the present methodology, however, the data are collected in continuous time, so only the total number that someone is captured is used.

Typically, in the biological application area, covariate information is not available or not used, leading to a basic model in which the Poisson parameters are assumed to be homogenous over the animals. For an overview of this area, the reader is referred to Seber (1982, Chap. 4), Chao (1988), and Zelterman (1988). In the statistical literature, this problem is also known as the estimation of the number of (unseen) species (Bunge and Fitzpatrick 1993).

In criminology, there are some early studies by Greene and Stollmack (1981) who use arrest data to estimate the number of adults committing felonies and misdemeanors in Washington D.C. in 1974/1975; Rossmo and Routledge (1990) who estimate migrating (or fleeing) fugitives in 1984 and prostitutes in 1986/1987, both in Vancouver; and Collins and Wilson (1990) who use arrest data to estimate the number of adult and juvenile car thiefs in the Australian capital territory in 1987. In the field of drug research, a one-source capture-recapture analysis has also been applied to estimate, for example, the size of the marijuana cultivation industry in Quebec (Bouchard 2007) and young drug users in Italy (Mascioli and Rossi 2008). These studies do not devote systematic attention to covariate information on the apprehended individuals.

The methodology reviewed here makes use of covariates that are available in police registers, such as age, gender, and so on. The methodology yields the following results: (1) the hidden number of offenders and a 95 % confidence interval for this hidden number, (2) a distribution of these hidden numbers over covariates, and (3) insight into which part of the hidden number is visible in the register and which part is missed, stratified by the levels of the covariates.

Covariate information is incorporated by using a regression model that makes use of truncated Poisson distributions, such as the truncated Poisson regression model and the truncated negative binomial regression model, that are well known in econometrics (e.g., Cameron and Trivedi 1998, Chap. 4). These models are elaborated so that a frequency can be estimated for the zero count as well as a confidence interval for this point estimate (see Van der Heijden et al. 2003a, 2003b; Cruyff and van der Heijden 2008; Böhning and van der Heijden 2009, that also show examples on undocumented immigrants, illegally owned firearms, and drunk driving). The methodology will be illustrated for perpetrators of domestic violence (van der Heijden et al. 2009).

Assumptions

As the methodology originates from the field of biology and it is used in the field of criminology, the assumptions of the methodology are discussed here in greater detail. Of course, assumptions that are realistic for animals will not always be realistic for human offenders.

The first assumption is that the number of times an individual is apprehended is a realization of a Poisson distribution (equations will be provided later). Johnson et al. (1993) discuss the genesis of the Poisson distribution and state that it was originally derived by Poisson as the limit of a binomial distribution with success probability p and N realizations, where N tends to infinity and p tends to zero, while Np remains finite and equal to λ. It turns out that even for $N = 3$, the Poisson distribution approximates the binomial distribution reasonably well if p is sufficiently small. Johnson et al. (1993) also note that the probability of success p does not have to be constant for the Poisson limit to hold. This implies that an individual's count still follows a Poisson distribution, even if this individual's Poisson parameter has changed during the period of observation. It follows for the type of applications being discussed that individuals do not need to have a constant probability to be apprehended, but it suffices if they can be apprehended a number of times (see van der Heijden et al. 2003a).

It is important to note that the Poisson assumption is only valid if a change in the individual Poisson parameter is unrelated to any prior apprehensions or non-apprehensions. This follows from the independence of subsequent trials in a binomial distribution. In the biostatistical literature, this problem is known as positive contagion (if the probability increases) or negative contagion (if the probability decreases).

Closely related to the contagion issue, there is the problem of an open or closed population. A population is closed if the number of offenders is constant over the period of data collection and is open if offenders may enter or leave the population during this period. Given what has been noted above, it is clear that the population may be open as long as entering or leaving it is not related to apprehension or non-apprehension. For example, detention following an apprehension removes the person from the population and excludes the possibility of any subsequent apprehensions and can therefore be seen as an extreme case of negative contagion.

So far the Poisson assumption pertaining to an individual count is discussed. The second assumption follows from using a regression model, in which the logarithm of the Poisson parameters is a linear function of covariates. In the regression model, the Poisson parameters are still assumed to be homogeneous for individuals with identical values on the covariates, but they are allowed to be heterogeneous for individuals with different values. Since here the differences in Poisson parameters are determined by the observed covariates, this is referred to as observed heterogeneity. If, in addition to observed heterogeneity, there are differences in the Poisson parameters that cannot be explained by the observed covariates, we speak of unobserved heterogeneity. If the Poisson regression model does not fit due to unobserved heterogeneity, this is referred to as overdispersion.

In conclusion, the most important violations of the Poisson assumptions in criminological applications are contagion and overdispersion. The contagion problem may be larger for some offenses than for others, and an indication of its importance can be obtained by studying the behavior of offenders as well as police officers (e.g., by doing qualitative research). If no additional information is available on their behavior, it seems best to interpret the results with caution. Overdispersion can be assessed in the data as a result of the analysis, and this will be discussed below.

Models

An informal definition of a Poisson distribution is as follows. The Poisson distribution is characterized by a Poisson parameter denoted by λ. This parameter λ expresses the probability of a given number of events (i.e., the count) under two assumptions, namely, that events occur:

1. With an average rate in a fixed interval of time
2. Independent of the time since the last event

The probability that the count Y generated by a Poisson distribution with Poisson parameter λ is equal to j is

$$p_j = P(Y = j) = \frac{e^{-\lambda}\lambda j}{j!}. \tag{1}$$

Three models for count data are discussed, namely, (1) the truncated Poisson regression model, (2) the truncated negative binomial regression model, and (3) the Zelterman regression model.

First consider Eq. 1. Two examples of a Poisson distribution are provided in Table 1. In the first example, an individual has a Poisson parameter $\lambda = .5$. Then his or her probability to be seen zero times is .607, the probability to be seen once is .303, twice is .076, and so on. These probabilities add up to 1. In the second example, it is assumed that there is an individual with Poisson parameter 1. His or her probability not

Capture Recapture to Estimate Criminal Populations, Table 1 Two examples of Poisson distributions (line 1 and 2) and their corresponding truncated Poisson distributions.

	0	1	2	3	4	5	6	Total
$\lambda = .5$	.607	.303	.076	.013	.002	.000	.000	1.000
$\lambda = 1$	.368	.368	.184	.061	.015	.003	.001	1.000
$\lambda = .5$	–	.771	.193	.032	.004	.000	.000	1.000
$\lambda = 1$	–	.582	.291	.097	.024	.005	.001	1.000

Notes: The columns show the number of times an individual is seen. The cells give the probabilities

to be seen is .368, the probability to be seen once is .368, twice is .184, and so on. Note that the individual with Poisson parameter $\lambda = 1$ has a larger probability to be observed at least once, namely, (1 − .368) = .632, whereas this probability for the individual Poisson parameter $\lambda = .5$ is only (1 − .607) = .393. It follows that the individual with Poisson parameter $\lambda = 1$ has a larger probability to be seen.

Model 1: The Truncated Poisson Regression Model

In single-register capture-recapture data the observed individuals each have a count larger than zero. Since for the observed individuals the count cannot be zero, these individuals have the so-called truncated Poisson distributions. For the data in the first and second row of Table 1, the truncated distributions for $y > 0$ are obtained by dividing the probabilities by $1 - \mathrm{P}(y = 0|\lambda)$, that is, for the first example we divide by 1 − .607 = .393 and for the second example we divide by 1 − .368 = .632; see rows three and four of Table 1.

As a first step towards developing the truncated Poisson regression model, assume that there are no covariates. This implies the homogeneity assumption, so that only a single Poisson parameter needs to estimated. This Poisson parameter is then used to obtain the probability of not being registered given that the Poisson parameter equals the value λ, as denoted $P(y = 0|\lambda)$. Using $P(y = 0|\lambda)$, the part of the population is estimated that we did not see. A small example will illustrate this. Assume that $n = 250$ individuals are observed with a Poisson parameter for which $P(y = 0|\lambda) = .667$. This would mean that $P(y > 0|\lambda) = .333$, that is, only one-third of the population is observed so our n refers to one-third of the population size. The missed part of the population size is (.667/.333) * 250 = 500. The estimated population size would then be equal to the observed part plus the missed part, that is, $N = 250 + 500 = 750$. In the statistical literature, this is known as the Horvitz-Thompson estimator of the population size (see Van der Heijden et al. 2003b).

Secondly, the Poisson parameter for individual i is related to the covariate values $x_{1i}, x_{2i}, \ldots$ of individual i by a log-linear model, that is,

$$Log\, \lambda_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots \qquad (2)$$

This equation explains the term "regression" in the name "truncated Poisson *regression* model." Once the model is estimated and the parameters are known, the Horvitz-Thompson method can be applied on an individual level (see Van der Heijden et al. 2003b). Using the earlier example, for each of the 250 individuals in the data, there is an estimated λ_i-parameter. For each individual i separately, $P(y = 0|\lambda_i)/P(y > 0|\lambda_i)$ can be calculated, and this yields the number of missed individuals with covariate values identical to individual i. If these numbers of missed individuals over i are summed and 250 is added, the estimated population size is found derived under the truncated Poisson regression model.

Summarizing, the homogeneity assumption is exchanged for a heterogeneity assumption that allows individuals to be different with regard to their covariates. It is important to consider the case where heterogeneity is completely or partly ignored. Van der Heijden et al. (2003a) show that ignoring heterogeneity leads to an estimated population size that is too low. Ignoring heterogeneity may happen if important covariates are not used in Eq. 2.

It is possible to investigate whether there is ignored heterogeneity by using a test presented by Gurmu (1991). If this test is significant, then there is evidence for additional heterogeneity (i.e., additional with regard to the heterogeneity that is already taken into account by the covariates), and the estimated population size is to be interpreted as a lower bound for the true population size. Thus, the Gurmu test can be used to investigate the fit of the model. Another way of investigating the validity of the Poisson assumption is the ratio plot (Böhning and Del Rio Vilas 2008). According to the previous section, $r_j = (j + 1)p_{j+1}/p_j = \lambda$, a constant, if the p_j follows a Poisson distribution. Departure from this constant indicated unobserved heterogeneity.

Model 2: The Truncated Negative Binomial Regression Model

The distinction between the truncated negative binomial regression model and the truncated Poisson regression model is that the truncated negative binomial regression model allows for additional heterogeneity of Poisson parameters after the covariates have been taken into account. Equations for the negative binomial model are complicated; see Hilbe (2011) for details.

Cruyff and Van der Heijden (2008) discuss the application of the truncated negative binomial regression model for population size estimation. It turns out that this model is often difficult to fit numerically, in particular when Poisson parameters are small. For details, see Cruyff and Van der Heijden (2008) where this is shown using a simulation study. Unfortunately there are also numerical problems for the example discussed below. For an example of opiate users where fitting the truncated negative binomial regression truncated Poisson regression model did not lead to numerical problems, see Cruyff and Van der Heijden (2008).

Model 3: The Zelterman Regression Model

The Poisson distribution Eq. 1 has the property that, for any j, $P(j + 1|\lambda)/P(j|\lambda) = \lambda/(j + 1)$. This can be rewritten as $\lambda = (j + 1)\ P(j + 1|\lambda)/P(j|\lambda)$. Zelterman (1988) uses this property to propose local estimators of the Poisson parameter by plugging in observed frequencies f_j of count j for $P(j|\lambda)$ and $P(j + 1|\lambda)$. In particular, if the frequencies f_1 and f_2 are plugged in, an estimate of the Poisson parameter

$$\hat{\lambda} = \frac{2f_2}{f_1}$$

is found, and this estimate of λ can then be used to estimate $P(0) = 1 - \exp(\lambda)$ and hence the Zelterman population size estimator $N_Z = n/(1 - \exp(\lambda))$. This estimator is closely related to Chao's estimator $N_C = n + f_1{}^2/2f_2$, that is used for similar purposes (Böhning 2010).

Zelterman's proposal has the advantage that it is robust against violations of the Poisson assumption such as unobserved heterogeneity. Also, when interest goes out to the frequency of the missed count f_0, then it makes sense to use the information that is the most close to f_0, that is, f_1 and f_2, because individuals that make up f_1 and f_2 will be most similar to the individuals that make up f_0. Probably for this reason and because the estimator is easy to understand, Zelterman's estimator is quite popular in estimates of drug using populations (see, Van Hest et al. 2007).

Recently Böhning and Van der Heijden (2009) generalized the Zelterman estimator so that it can take covariates into account as in Eq. 2. As the Zelterman estimator is a useful competitor of the Poisson estimator if there is heterogeneity of the Poisson parameters, the Zelterman regression model is a useful competitor of the truncated Poisson regression model when the homogeneity assumption of the Poisson parameters (conditional on the covariates) is violated. This follows from the robustness property of the Zelterman estimator.

Model Choice

In the truncated Poisson regression model, the presence of unobserved heterogeneity in the data can be investigated with a test proposed by Gurmu (1991) and ratio plots (Böhning and Del Rio Vilas 2008). If unobserved heterogeneity can be ignored, the truncated Poisson regression model is the model of choice. If unobserved heterogeneity cannot be ignored, the population size estimate given by the truncated Poisson regression model is to be interpreted as a lower bound for the true population size. In this situation, the truncated negative binomial regression model can be tried. However, regularly numerical problems are encountered in the estimation of the model. If there is doubt about

Capture Recapture to Estimate Criminal Populations, Table 2 Observed and fitted frequencies of counts 0, 1, 2, . . . for domestic violence data under truncated Poisson regression model

	0	1	2	3	4	5	>5	Total
Observed	–	15,169	1,957	393	99	28	16	17,662
Estimated	51,629	14,814	2,494	317	33	3	0	69,290[a]

[a] 95 % confidence interval is 66,242–72,338

the Poisson assumption for individuals having counts higher than 2, then the Zelterman regression model provides a robust estimate.

Example

In the context of domestic violence policy, it is important to have reliable estimates of the scale of the phenomenon. In 2009, a study was conducted to supplement a victim study and perpetrator study (van der Heijden et al. 2009). Here the perpetrator study is reported. cthe prevalence of domestic violence, particularly because so few victims are willing to file charges, which leads to underreporting, that is, the dark number in registers. The aim of the capture-recapture methods used in this study is to estimate the size of the underreporting. Adding up the reporting and underreporting then yields an estimate of the total population of offenders. The estimates have been made using data from incidents that were reported and where charges were filed. The data represent the Netherlands except the police region for The Hague.

The estimates presented have been calculated using the Poisson regression model and the Zelterman regression model. The negative binomial regression model had numerical estimation problems. The estimates are presented for a year ranging from mid-2006 to mid-2007. Distinctions were also drawn as regards specific features of the estimated populations, that is, the sex and age of the individual suspect, the type of violence used, the type of victim of domestic violence, and the ethnic background of the suspect.

The first line of Table 2 shows the observed distribution of the counts. A total of 17,662 perpetrators are observed in the year ranging from mid-2006 to mid-2007, of whom 15,169 were observed once, 1,957 twice, and so on. The second line of Table 2 shows the fitted values under the truncated Poisson regression model. The estimated population size is 69,290 (confidence interval is 66,242–72,338) of whom 51,629 perpetrators are not found in the police register. The fit of the model is not good, as is revealed by comparing the observed frequency of the counts and the estimated frequency of the counts: The Gurmu test on overdispersion gives a chi-squared distributed test value of 30.04 for 1 ° of freedom ($p < .0001$), showing that the departure from the truncated Poisson regression model is significant. Similarly, Fig. 1 shows a ratio plot (Böhning and Del Rio Vilas 2008) where these ratios are monotonically increasing, indicating strong unobserved heterogeneity. For this reason, further results for the truncated Poisson regression model are not presented, but the focus will be on the Zelterman regression model.

The Zelterman regression model yields a population size estimate of 84,862 (confidence interval is 80,293–89,431). Columns 1–3 of Table 3 provide parameter estimates, standard errors, and p-values for a Wald test. Gender is significant, and the fitted conditional Poisson parameter for males is exp (.49) = 1.63 times as large as the fitted conditional Poisson parameter for females. For type of domestic violence, sexual and physical assaults have a lower fitted Poisson parameter than threat. In comparison to an (ex-) partner, a family friend and "other" have lower fitted Poisson parameters. Ethnic background (first or second generation) is not significant, except for "not registered" in the official register, which are undocumented foreigners, who have a lower fitted Poisson parameter than the native Dutch population.

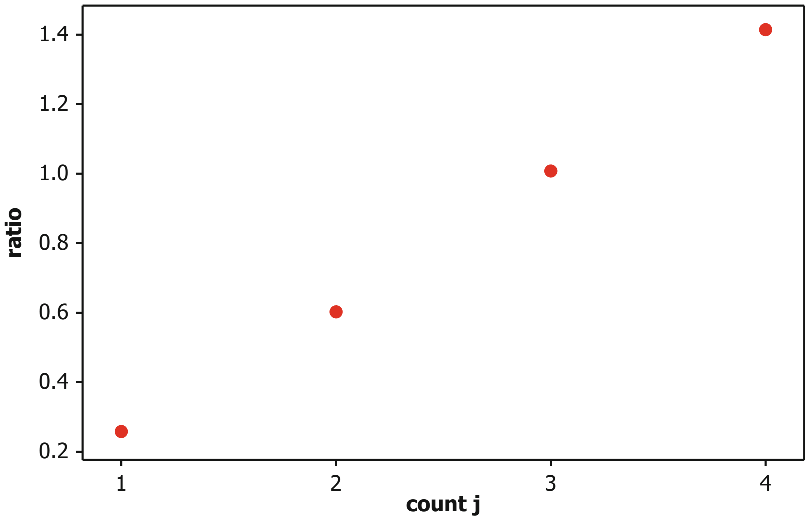

Capture Recapture to Estimate Criminal Populations,
Fig. 1 Ratio plot for the observed frequencies in Table 2

Columns 4 and 5 of Table 3 provide observed univariate frequencies and fitted frequencies. For example, 14,827 males are found in the police register, and the estimate of the number of males is 68,048; thus, it can be concluded that 14,827/68,048 * 100 = 22 % of the male perpetrators is observed. For females, this percentage is 2,653/16,813 * 100 = 16 %. The other percentage can be derived in an identical way.

Future Directions

It is shown how police records can be used to estimate the size of criminal populations. These estimates can be used to evaluate the effectiveness of the police forces and grant insight into differential arrest rates (Collins and Wilson, 1990) for different groups. Even though the definition of the data is straightforward, however, the data are sometimes contaminated with errors. The reason is that the police does not collect these data for the purpose of conducting statistical analyses, they do so to facilitate the process of apprehending offenders. Therefore, the register is not always as careful as it should be. For example, data cleaning was required to minimize the likelihood of incorrect double counts (i.e., the same apprehension appears twice in the system). Clearly, an incorrect double count incorrectly decreases f_k by 1 and increases f_{k+1} by 1, so that there appear to be more recaptures than there actually are. The result is that the estimated zero count is too low. Although careful attention has been devoted to the appearance of incorrect double counts, it is possible that there are still a few in the data.

Conversely, suppose that an individual has been apprehended several times, but this has not been recognized so that this individual has been entered several times as several single persons. This will lead to the result that f_1 is too large whereas f_j with $j > 1$ is too small, leading to an overestimation.

Apart from problems in the correctness of the data, it may also be that data do not follow the assumptions of the model. This will lead to biased estimates. Above we discussed the problem of positive contagion (leading to estimates that are too low) and negative contagion (leading to estimates that are too high). Another problem discussed above is unobserved heterogeneity, leading to estimates that are too low. It may very well be that unobserved heterogeneity always plays a role when the data are derived from a police registration.

As regards the meaning of our population size estimate, one might wonder what it stands for?

Capture Recapture to Estimate Criminal Populations, Table 3 Parameter estimates, observed and estimated subpopulation sizes under Zelterman regression model. Estimated total population size is 84,862 (95 % CI is 80,293–89,431)

	Par	SE	*p*	Observed	Estimated
Constant	−2.07	.13	.00		
Male	.49	.09	.00	14,827	68,048
Female	.00			2,653	16,813
Age/10	−.02	.02	.26		
Threat	.00			3,822	14,442
Waylay	.20	.13	.12	556	1,830
Physical	−.38	.06	.00	11,234	59,046
Other viol.	−.20	.11	.08	1,063	4,859
Psychological	.13	.18	.46	298	1,067
Sexual	−.42	.15	.01	686	3,607
(Ex-)partner	.00			12,130	56,113
Friend family	−.40	.16	.01	616	4,178
Child	.03	.11	.78	1,271	6,366
Elderly	−.05	.26	.84	168	811
Parents	.16	.09	.08	1,301	4,890
Other family	.11	.09	.18	1,611	7,027
Other	−.79	.19	.00	561	5,467
Native Dutch	00			9,642	42,253
Morocco (2)	−.10	.18	.57	345	1,597
Morocco (1)	.03	.21	.89	730	3,267
Turkey (2)	−.03	.18	.89	318	1,451
Turkey (1)	−.25	.22	.26	687	3,811
Suriname (2)	−.05	.15	.72	489	2,297
Suriname (1)	−.16	.19	.40	1,368	7,214
Dutch Antilles (2)	−.64	.39	.10	109	927
Dutch Antilles (1)	.76	.41	.07	735	2,460
Not western (2)	−.28	.34	.41	108	641
Not western (1)	.16	.35	.65	1,146	5,242
Western (2)	−.08	.12	.49	830	3,869
Western (1)	.02	.17	92	593	2,807
Not registered	−.54	.11	.00	1,268	9,325

Basically, the apprehended individuals can only be generalized to similar individuals who are not apprehended (but who are in principle apprehensible and where charges could be filed). Thus, the population size estimate does not stand for the total number of perpetrators of domestic violence; it stands for the apprehensible ones. The population estimate is still useful, since it represents the number of individuals who pose a threat to society and is thus an indication for the police potential workload.

Related Entries

- ► Domestic Violence
- ► Feminist Theory and Domestic Violence
- ► History of the Dutch Crime Victimization Survey(s)
- ► History of the Self-Report Delinquency Surveys
- ► History of the Statistics of Crime and Criminal Justice
- ► International Crime Victimization Survey

- ▶ National Victimization Surveys
- ▶ Police Family Violence Services
- ▶ Repeat Victimization
- ▶ Self-Reported Offending: Reliability and Validity
- ▶ Surveys on Violence Against Women
- ▶ Understanding Victimization Frequency

Recommended Reading and References

Böhning D (2010) Some general comparative points on Chao's and Zelterman's estimators of the population size. Scand J Stat 37:221–236

Böhning D, Del Rio Vilas VJ (2008) Estimating the hidden number of scrapie affected holdings in Great Britain using a simple, truncated count model allowing for heterogeneity. J Agric Biol Environ Stat 13:1–22

Böhning D, van der Heijden PGM (2009) A covariate adjustment for zero-truncated approaches to estimating the size of hidden and elusive populations. Ann Appl Stat 3(2):595–610

Bouchard M (2007) A capture-recapture model to estimate the size of criminal populations and the risks of detection in a Marijuana cultivation industry. J Quant Criminol 23(3):221–241

Bunge J, Fitzpatrick M (1993) Estimating the number of species: a review. J Am Stat Assoc 88:364–373

Cameron AC, Trivedi PK (1998) Regression analysis of count data, vol 36, Econometric society monographs no. 30. Cambridge University Press, Cambridge

Chao A (1988) Estimating animal abundance with capture frequency data. J Wildl Manag 52:295–300

Collins MF, Wilson RM (1990) Automobile theft : estimating the size of the criminal population. J Quant Criminol 6(4):395–409.

Cruyff MJLF, van der Heijden PGM (2008) Point and interval estimation of the population size using a zero-truncated negative Binomial regression model. Biom J 50(6):1035–1050

Greene MA, Stollmack S (1981) Estimating the number of criminals. In: Fox JA (ed) Models in quantitative criminology. Academic, New York, pp 1–24

Gurmu S (1991) Tests for detecting overdispersion in the positive Poisson regression model. J Bus Econ Stat 9:215–222

Hilbe JM (2011) Negative Binomial regression. Cambridge University Press, Cambridge

Johnson NL, Kotz S, Kemp AW (1993) Univariate discrete distributions, 2nd edn. Wiley, New York

Mascioli F, Rossi C (2008) Capture-recapture methods to estimate prevalence indicators for the evaluation of drug policies. Bull Narc LX(1/2):5–25

Rossmo DK, Routledge R (1990) Estimating the size of criminal populations. J Quant Criminol 6:293–314

Seber GAF (1982) The estimation of animal abundance, 2nd edn. Griffin, London

Van der Heijden PGM, Cruyff MJLF, van Houwelingen HC (2003a) Estimating the size of a criminal population from police registrations using the truncated Poisson regression model. Stat Neerl 57:289–304

Van der Heijden PGM, Bustami R, Cruyff MJLF, Engbersen G, van Houwelingen HC (2003b) Point and interval estimation of the truncated Poisson regression model. Stat Model 3:305–322

Van der Heijden PGM, Cruyff MJLF, Van Gils GHC (2009) Omvang van huiselijk geweld in Nederland. Ministerie van Justitie, Den Haag, WODC (33 pagina's)

Van Hest NHA, Grant AD, Smit F, Story A, Richardus JH (2007) Estimating infectious diseases incidence: validity of capture–recapture analysis and truncated models for incomplete count data. Epidemiol Infect 136:14–22

Wittebrood K, Junger M (2002) Trends in violent crime: a comparison between police statistics and victimization surveys. Soc Indic Res 59:153–173

Zelterman D (1988) Robust estimation in truncated discrete distributions with application to capture-recapture experiments. J Stat Plan Inference 18:225–237

Car Theft

▶ Motor Vehicle Theft

Career Criminals and Criminological Theory

Raymond Paternoster
Department of Criminology and
Criminal Justice, University of Maryland,
College Park, MD, USA

Synonyms

Chronic offenders; Persistent offenders

Overview

In 1986, the National Academies Press published a two-volume compendium entitled "Criminal Careers and 'Career Criminals'"

(Blumstein et al. 1986) that gave expression to the growing interest in both the field of criminology and among policy makers in the career criminal. Although there is no exact agreement on what a career criminal is, in the literature it has generally referred to persons who commit many crimes beginning at an early age and who persist in offending over the life course. Among other things, then, the career criminal is a habitual, persistent, or "chronic" criminal offender, committing criminal acts at every stage of the life course. Blumstein and colleagues' report was not, however, the first instance of academic interest in the career criminal. From the 1930s to 1950s, Sheldon and Eleanor Glueck assembled several large longitudinal data sets that tracked offending from a young age to adulthood. Most notably, the Unraveling Delinquency Study (Glueck and Glueck 1950) which included 500 boys selected from two youth correctional centers in Massachusetts – the Lyman School for Boys and the Industrial School for Boys. Initial data were collected when the average age of the boys was 14, and they were followed up at age 25 and again at age 32. The presence of a persistent or career offender could be deduced from the fact that 55 % of these seriously delinquent boys experienced an arrest at some time between the ages 32 and 45 (Sampson and Laub 1993: 30).

Additional interest in the notion of career criminals was sparked by Wolfgang and colleagues' Philadelphia cohort studies. In the 1945 cohort, they classified youth with five or more police contacts as chronic or habitual offenders (Wolfgang et al. 1972). While these chronic offenders comprised only 6 % of the total birth cohort of nearly 10,000, they had 52 % of all police contacts and committed 63 % of all recorded index crimes, 71 % of all homicides, 73 % of all the rapes, and 82 % of the robberies. In a subsequent work, they followed a sample of this cohort up to age 30 and identified what they called "persistent" offenders as those who committed offenses both as juveniles and adults. Of the 975 boys who were followed up, 459 had at least one official arrest, and nearly 39 % of these were persistent offenders who had an arrest as juveniles and adults. Further, these persistent offenders committed both more frequent and more serious offenses than those who committed crimes only as juveniles or only as adults (Wolfgang et al. 1987: 24).

Key Issues and Themes

Interest in the career criminal persists today among scholars and policy makers. With respect to the former, Laub and Sampson (2003) have obtained arrest histories for the Glueck boys up to age 70 and identified a small group (less than 5 % of the total) of chronic offenders who were experiencing arrests into their 60s. In a Dutch sample, Blokland (2005; Blokland and Nieuwbeerta 2005; Blokland et al. 2005) identified a small group of offenders (also less than 5 % of the total) who experienced criminal convictions from their teens to their 60s. Finally, Sohoni et al. (2013) found a group of offenders (about 4 % of the total) in the South London data of Farrington and West who experienced convictions from their teens to age 50. With respect to public policy interests, most states have habitual offender statutes that provide for enhanced penalties for offenders with several prior convictions as well as "three-strike" laws that provide for life imprisonment for those repeatedly convicted of serious felonies.

Given the very strong interest in *identifying* the career criminal, it might perhaps be surprising that an equivalent amount of effort and attention has not been devoted to understanding the cause of persistent offending. The Glueck's, for example, were loath to provide a comprehensive theoretical account either of delinquents in general or those who persisted in crime throughout their lives arguing that it was "premature and misleading to give exclusive or even primary significance to any one of the avenues of interpretation" (Glueck and Glueck 1950: 281). Instead, they put forward an atheoretical multifactored explanation that simply consisted of a list of the risk factors they empirically found to be related to crime. Wolfgang and his colleagues were silent with respect to how the small group of chronic and habitual offenders they identified could be explained. It was really not until the mid-1980s rebirth of empirical interest in the career criminal that some attention was beginning to be devoted to understanding and explaining the high-rate offender who persists over time. In fact, the empirical work of career criminal researchers was in part responsible for the emergence of developmental and life-course theories of offending, theories that among other things were devised to explain why it is that offending takes place throughout the life course.

Theoretical Explanations for the Career Criminal

There really is no theory of the career criminal in the sense of a separate theory of crime to explain this phenomenon. Rather, there are theories of persistent offending, without speaking to the career criminal, and there are general theories of offending that can be adapted to explain the habitual or persistent offender whose long-term pattern of offending can be described as a career in crime.

Terrie Moffitt (1993) has presented a typological theory of criminal offending that provides some theoretical light on the career criminal. She argues for a distinction between two qualitatively distinct types of offenders: the adolescent-limited and the life-course-persistent offender. The life-course-persistent offender would qualify as a career criminal since they are described as committing crimes at a high rate over a protracted period of time: as children, adolescents, and adults. They are a small group, comprising approximately of 5 % of any sample of youth, consistent with the Glueck's high-rate offender and the chronic and habitual offender of Wolfgang et al.

In accounting for the life-course-persistent offender, Moffitt appeals to an integration of constitutional and environmental factors. The life-course-persistent offender suffers initially from neuropsychological deficits that are either prenatal or neonatal in origin and are manifested in profound cognitive and executive function difficulties (impulsivity and poor verbal skills are particularly important). Since infants with these kinds of neuropsychological deficits are frequently born in disadvantaged and criminogenic environments, these deficits frequently go uncorrected and are more often made worse. In a nutshell, a difficult child often overwhelms distressed and poorly resourced parents unable to cope with the challenge who frequently respond with erratic and harsh discipline. Such responses by caregivers only exacerbate the child's initial difficulties, and as a result of this destructive interaction, the child fails to develop behavioral restraints – they are undersocialized. This negative interaction continues with development, and the poorly controlled toddler becomes the poorly controlled child who becomes the poorly controlled adolescent on into adulthood. As described by Moffitt (1993: 12) the life-course-persistent child becomes ensnared in an ongoing sequence of failed development: "personal characteristics such as poor self-control, impulsivity, and inability to delay gratification increase the risk that antisocial youngsters will make irrevocable decisions that close the doors of opportunity."

A related theory of the persistent offender, related because it appeals to the same set of empirical facts and develops a similar theoretical argument, was presented by Patterson and colleagues (1990). Like Moffitt, Patterson et al. define two etiologically distinct trajectories of offending. One starts early (the early-onset offender) and persists over the life course into adulthood in a way that would be harmonious with the idea of a career criminal. The behavioral trajectory of the early-starting offender is similar to Moffitt's life-course-persistent offender. The second starts offending in mid- to late adolescence and ceases offending as they enter adulthood. The behavioral trajectory of the late-starting offender is similar to Moffitt's adolescent-limited offender.

In Patterson's typology, the cause of the early-starting and late-persisting offender is a combination of disruptive family practices (harsh discipline, emotional coldness and little intimacy, little parental involvement with the child, and poor monitoring and supervision). The immediate result of these inept child-rearing practices is that the child learns to be coercive and aggressive both directly from parents (modeling their physical violence) or indirectly (parents tolerate or support the child's violence) which over time culminates in an aggressive poorly controlled child. Although perhaps adaptive in the short term in the home, such verbally and physically aggressive behavior results in both academic and social failure in the school as well as rejection by normal peers and acceptance by equally disruptive peers. Finally, in a process that is reminiscent of Moffitt's "ensnarement," these aggressive adolescents are at a great risk for committing serious delinquent offenses as adolescents and criminal offenses as adults.

There is an important common causal thread that runs through the theories of the persistent or career criminal in both Moffitt and Patterson's work. While Moffitt's theory admittedly has a neuropsychological component that is lacking in Patterson, both appeal to a disruption in family processes, particularly inadequate training or socialization by parents or caregivers. Caregivers who for various reasons are emotionally uninvolved with their children and therefore inadequately monitor and supervise their conduct create children who are impulsive, poorly controlled, and act in ways that may be in their immediate interest (ex: snatching a toy away from another child, stealing another child's bike) but is destructive in the long term. Such poorly socialized children are not easily able to make up for past deficits and in fact misplay prosocial opportunities and experiences they confront. In other words, they never catch up and in fact fall further behind developmentally in diverse areas, with the ultimate outcome that they are at high risk for frequent adult criminal offending and other self-destructive behaviors such as drug use, alcohol abuse, long bouts of unemployment, and unhealthy social relationships.

The preceding paragraphs may be very familiar to many criminologists since it describes a causal explanation for repeated offending that is not tied directly to the career criminal paradigm, in fact one that has been hostile to that paradigm – Gottfredson and Hirschi's (1990) general theory of crime. There are in fact several general theories of crime that provide an adequate theoretical explanation for the persistent offending that characterizes a criminal career.

In 1990, Michael Gottfredson and Travis Hirschi (1990: 117) published what was to be a highly influential book titled *A General Theory of Crime* in which they presented a theory which could explain "all crime, at all times" and that could also explain behaviors that are as self-destructive in the long term as crime. The key theoretical construct they premised their theory on was the notion of low self-control which, while its precise meaning has changed somewhat over time, is manifested in a person's inability to consider both the short- and long-term negative consequences of their actions. The person low in self-control, then, is attracted to behaviors which while they may provide gratification in the short term, such as crime and drug and alcohol abuse, have detrimental long-term consequences, and since the costs of actions are generally delayed relative to their benefits, they play a much less consequential role. Very similar to Moffitt's and Patterson's theories, Gottfredson and Hirschi trace the origins of low self-control to family disruption. In fact, there is a very close correspondence in the causal mechanisms underlying the person with low self-control and the life-course-persistent and early-starting offender. In each theory, an emotional connection or relationship between caregiver and child is absent, and this affective distance results in inept socialization. Caregivers who are not emotionally attached to the child are unlikely to monitor or supervise the child's conduct properly and unlikely to correct inappropriate and reward appropriate behavior.

Another important piece of common ground among the theories is that the initial failure to socialize has reverberations down the life course. Moffitt (1993), for instance, speaks about poorly socialized youth becoming "ensnared in a deviant life style by crime's consequences" because they "will make irrevocable decisions that close the doors of opportunity." She links poor childhood socialization to teenage pregnancy, school dropout, drug addiction, unemployment, and incarceration such that "the behavior of life-course-persistent antisocial persons is increasingly maintained and supported by narrowing options for conventional behavior." Patterson (1990: 266) in a process that parallels the notion of "ensnarement" argues that the child who is poorly socialized becomes aggressive and predatory at an early age and that such behavior leads to a lack of success in school and ostracism from normal peers in a causal spiral that over time leads to "school dropout, uneven employment histories, substance abuse, marital difficulties, multiple offenses, incarceration and institutionalization." Similar to both of these, Gottfredson and Hirschi (1990: 96) account for the persistent offender by noting that because they are relatively unaffected by the negative consequences of their actions, the poorly socialized person with low self-control fails to take advantage of opportunities to change throughout life, and they too become ensnared in a cycle of repeated failure and persistent criminality: "The traits composing low self-control are also not conducive to the achievement of long-term individual goals. On the contrary, they impede educational and occupational achievement, destroy interpersonal relations, and undermine physical health and economic well-being."

There are, however, some important distinctions among the three theories. First, the theory of self-control argues that low self-control and therefore the propensity to become a habitual or career offender lies on a continuum rather than being a typology. They argued (1990: 95) that "though there will be little variability among people in their ability to see the pleasures of crime, there will be *considerable variability* in their ability to calculate potential pains" (emphasis added). Second, unlike Moffitt's theory that relies on antecedent neuropsychological mechanisms or Patterson's that relies on the mediating causal effect of deviant peers, Gottfredson and Hirschi locate the source of poor self-control solely in the family, and involvement with deviant peers has no causal effect on persistent offending but is simply another manifestation of low self-control.

Another general theory of crime that can offer us an explanatory account of the persistent offender is Sampson and Laub's (1993, 2003) age-graded informal theory of social control. In discussing this theory, one cannot help but be struck with the substantial common ingredients in all of these theories of the persistent offender in spite of their supposed, much ballyhooed differences. Acknowledging the existence of a persistent criminal offender whose numbers may be small but whose maladaptive behavior is quite broad to cover marital problems, unemployment, drug and alcohol addiction, and the corporal punishment of children (heterotypic continuity), Sampson and Laub (1993: 123–138) appeal to a causal process that should by now be familiar to the reader. Children who are poorly socialized and antisocial are maintained or ensnared in their behavior through two causal mechanisms. First, antisocial behavior in children calls forth destructive and maladaptive responses by caregivers and others in their social environment, and such responses fuel further antisocial behavior by children, a process termed "interactional continuity" (Sampson and Laub 1993: 124). Second, the antisocial behavior of young children and the delinquent acts of adolescents have "negative structural consequences" for subsequent life chances that "may lead to the 'closing of doors' as far as opportunities go (for example, school failure, unemployment)," in a process termed "cumulative continuity."

In their 2003 book, *Shared Beginnings*, Laub and Sampson (2003: 150–195) discuss the persistent or habitual offender at some length and make a concerted effort to draw a distinction between their theoretical account and Moffitt's. Their view of Moffitt's theory is that it places great weight on early childhood factors (perhaps ignoring Moffit's explanation of being ensnared in a life of crime over time), and they find little evidence either in the 2003 or earlier 1993 book that childhood risk factors have an effect on adult criminality. Instead, Sampson and Laub put the majority of their eggs in the "adult life events" basket, arguing that persistence in adult crime is due to a weakness in adult social bonds. In point of fact, however, in further differentiating their position from so-called propensity theorists like Moffitt, Sampson and Laub also place great weight on a labeling process to account for the persistent or career offender. In *Crime in the Making* (1993: 137), they refer to a "structural version of labeling theory and state dependence" in which early arrest and incarceration experiences cut persons off from conventional opportunities: "the connection between official childhood misbehavior and adult outcomes may be accounted for in part by the structural disadvantages and diminished life chances accorded institutionalized and stigmatized youth." In fact, in an earlier paper, Sampson and Laub (1997) described the criminal career as a "stable pattern of deviant behavior that is sustained by the labeling process." In other words, contact with both the juvenile and adult criminal justice system creates problems of adjustment (Lemert 1967: 63) for people because it "denies them the ordinary means of carrying on the routines of everyday life open to most people" (Becker 1963: 35). In a reversal of the desistance process they describe, those with histories of imprisonment are "knifed off" from normal lives because of weakening social bonds. Sampson and Laub (2005: 171) in fact subsequently concede that "explanations of desistance from crime and persistent offending are two sides of the same coin." Just as a strengthening of adult social bonds may knife off a criminal offender from a previous life of crime and criminal associates, so may criminal justice intervention knife them off from a conventional.

It is important at this time to reiterate the similarities among all these theories of the persistent offender. In Moffitt's, Patterson's, Gottfredson and Hirschi's, and Sampson and Laub's theory, some childhood deviants grow up to be persistent adult offenders because they are all too frequently cut off from many conventional opportunities and fail at virtually every opportunity they are afforded to do good. In addition, events that occur in childhood are important in initiating that causal process. As such, all of the theories we have discussed thus far reflect Hagan's (1993) notion of the consequences of "criminal embeddedness." Hagan (1993: 487) argues that youth crime and contacts with the justice system cut adolescents off from sources of social capital and other social networks that connect to conventional lives. He further asserts that youths in trouble instead acquire "criminal capital," that is, they have social networks that connect them to illegal labor opportunities such as theft and drug dealing which embed them in a criminal lifestyle: "criminal youths are embedded in contexts that isolate them from the likelihood of legitimate adult employment."

Other general theories of crime offer their own theoretical account of the persistent, career criminal that are variations of the ones just discussed (with, of course, their own nuances). Farrington (2005) presented what he called an Integrated Cognitive Antisocial Potential (ICAP) theory that is a very eclectic account of criminal offending over the life course, including both desistance from and persistence in crime. According to ICAP, initiation into offending is due to a mixture of short-term and long-term antisocial potential (AP, which is defined simply as the potential to commit criminal acts) interacting with a person's social environment, while persistence in criminal offending is due to the fact that an initial distribution of long-term AP is relatively stable over time. Reminiscent of Gottfredson and Hirschi, then, persistence in offending in Farrington's theory is brought about because of the relative stability of an initial distribution of a time-stable individual trait, long-term AP, which consists of impulsiveness, the experience of strain, criminal modeling, socialization, and criminogenic life events.

Thornberry (Thornberry and Krohn 2005) applies his general Interactional Theory of crime to account for the persistent, career offender. In this theory, long-term persistence in crime is accounted for by two developmental processes: (1) the causal factors that produce an introduction into crime are themselves stable over time, and (2) there is a constancy in "negative temperamental traits" that are stable from childhood to adulthood. In a quasi-typological approach, Thornberry and Krohn (2005) suggest that for those who commit their first offense at an early age, the developmental deficits (poor parental supervision and monitoring, weak parental attachments, poor school performance and adjustment) are more extreme than for those whose onset into crime comes later, and since they are more extreme, they are also more stable over time. In addition to this, they describe a state-dependence-like process whereby early offending triggers hostile and maladaptive responses from the environment such as caregivers, teachers, and conventional peers. Essentially, early conduct problems ensnare youths by limiting conventional opportunities and social networks (i.e., they become embedded in a criminal lifestyle), restricting the development of social, personal, and human capital. In their own words, but words that seem to just as easily could have come from Moffitt, Patterson, Hagan, or Sampson and Laub, Thornberry and Krohn (2005: 198) posit that: "individuals who initiate antisocial behavior at very young ages are more likely than average to persist because the causal factors are likely to remain in place (both environmental and personality/trait factors) and because early involvement in antisocial behavior generates cumulative and cascading consequences in the person's life course. All of this reduces the formation of social bonds and social capital and increases embeddedness in deviant networks and belief systems, foreclosing conventional lifestyles and entrapping the individual in deviant lifestyles."

This account for persistence in offending for those who onset early is not, however, likely to hold for those whose initiation into offending is more "on time" and occurs during adolescence. Nevertheless, some of these on-time onsetters can also persist in offending in their adult years, and they do so primarily because their involvement with crime puts them at higher risk for school failure, unstable and unsatisfying social relationships, and problems at home and in the workplace and disrupts their successful transition into adult roles. Thornberry and Krohn (2005: 200) place great emphasis (as does Moffitt and Patterson for the adolescent-limited and late-onset offender) on the detrimental consequences of becoming embedded in deviant social networks for the youth who commits his first crime later rather than earlier.

Bouffard and Piquero (2010) recently offered a theory of the career criminal offending that appeals to Sherman's (1993) notion of defiance as an emotional reaction to labeling. Sherman had argued that when a poorly bonded person responds to the imposition of what they perceive to be an unfair sanction with anger rather than shame and an acknowledgement of harm, they become defiant and resentful and as a result are likely to commit further acts of crime rather than be deterred. Bouffard and Piquero (2010: 233) take this general theory of crime and apply it to the study of desistance and persistence arguing with respect to the latter that persistent offending is a "defiant response of a poorly bonded offender who defines their sanction as unfair and stigmatizing and refuses to acknowledge the shame they feel. These individuals may continue or escalate their offending." What happens when a poorly bonded person gets labeled, then, is that they react in such a way (with defiance, resentment, and rejection) that their already weak bonds become even weaker in a process that by now should be very familiar. The ingredient that Bouffard and Piquero add to this mix is the role of emotions, a welcome addition since emotions are a neglected factor in criminological research.

Wikström and Treiber (2009) presented a theory of the persistent or career offender that is a derivative of Wikström's (2005) Situational Action Theory (SAT). According to SAT, all criminal offending is moral rule breaking with the explicit implication that a key explanatory variable is an individual's moral values as well as the interaction between an individual's personal characteristics (such as their moral values) and environmental factors. An individual with a strong moral position against committing a criminal act will not view crime as a viable course of action and will not, therefore, be affected by contingencies in the environment, such as opportunities or the perceived costs and benefits of offending. It is also reasonable to conclude that such individual factors as an individual's propensity to commit crimes as manifested in their self-control are activated only when crime is perceived to be a morally available course of action. In SAT, then, the interaction between a person's moral values and their propensity to rule breaking will strongly influence what courses of action are available in a given situation and, therefore, what choices are available to be selected.

With respect to the career criminal, Wikström and Treiber (2009: 411) make a distinction between perception, choice, and action in a given setting that is "either *habitual* or *deliberate*, depending on the actor's familiarity with the setting" (emphasis is original). The more familiar an actor is to a given setting, the more their perceptions and behavior are likely to be guided by habit rather than deliberation. Actors who repeatedly find themselves in particular environments, such as one of extreme deprivation or abundant criminal opportunities with low moral restraints and little risk, will respond habitually to such contexts, and their behavior will tend to be highly stable and persistent over time. This constancy of the person-environment interaction and resultant habituation led Wikström and Treiber (2009: 411) to conclude that "there may be strong habitual elements in what drives much of persistent offenders' criminality."

Finally, it must be noted that there is more than one biologically based theory of the persistent, career criminal, though the common ground seems to be that they all link habitual offending to an early onset of offending and the interaction between biological and environmental factors (see Savage 2009). One such example is Tibbetts (2009: 183) whose initial premise is that "early onset is the most important factor in predicting which individuals are most likely to become chronic, habitual offenders." The secret to unlocking the origins of the career criminal, therefore, is to understand who onsets their offending early. In explaining the early-onsetting offender, Tibbetts' account relies on the interaction between a cluster of biological risk factors such as pre- or perinatal brain trauma, lower heart rate, arousal, slower brain wave patterns, deficits in pre- and perinatal nutrition, personality attributes (impulsivity), and a cluster of environmental deficits such as poverty, single-parent families, family size, and a criminogenic neighborhood. Whatever their particular composition, Tibbetts' and other biologically inclined theories of persistent offending are clear in stating that it is not biological factors nor environmental factors that are themselves responsible for the persistent offending of the career criminal but the interaction between the two.

Career criminals, offenders with both a high frequency and long duration of offending, have been the topic of theoretical discussion in criminology for over 20 years. One might get the impression from that fact that there is a great diversity of opinion with respect to the causes of career criminal offending. If so, one is in for a disappointment, for while there are numerous attempts to explain the persistent or habitual offender, for the most part the explanatory schemes are surprisingly similar, despite the many disciplines that have participated. While some theorists do place more of their eggs in an initial biological/psychological basket – those who eventually become persistent offenders have various constitutional, neuropsychological, or personality deficits – and others less, the natural progression to persistent offending is oddly very similar across many theories. The explanatory schemes involve notions of a state-dependent or dynamic process whereby initial developmental failures (hostile relationships with inept caregivers, failure at school, social rejection by normal peers) have reverberations that are felt down throughout the life course. It is not an oversimplification to say that toddlers who get into trouble with parents turn into adolescents who get into trouble with teachers and "good peers" who get into trouble with employers and spouses/partners who then get into trouble with the criminal justice system. In many if not all of these theories, there is a strong labeling component whereby initial antisocial behavior that is responded to informally leads to formal criminal justice system intervention which then leads to a “knifing off” of legitimate and conventional routines of life. The repeat adolescent offender finds it hard in Lofland’s (1969) terms to have access to normal places, others, and hardware; as a result of which, they are only able to secure “criminal capital” and become more and more embedded in a criminal lifestyle (Hagan 1993).

While one may lament the absence of theoretical variety in theories of the persistent offender (many in fact argue that persistence is simply the flip side of desistance), the brute fact is we know very little about how adequate these explanations really are. The extensive empirical work to determine what factors best explain persistent offending or the career criminal has hardly begun. While we know something about the process by which offenders drop out of crime, we know far less about the causal mechanisms behind repeated, habitual criminal offending.

Related Entries

- ▶ Age-Crime Curve
- ▶ Criminal Careers
- ▶ Criminal Careers and Public Policy Responses
- ▶ Desistance from Crime
- ▶ General Strain Theory
- ▶ General Theory of Crime
- ▶ History of Criminological Theories: Causes of Crime
- ▶ Individual Characteristics and General Strain Theory
- ▶ Integrating Rational Choice and Other Theories
- ▶ Labeling Theory
- ▶ Labelling, Deviance, and Media
- ▶ Lifestyle Theory
- ▶ Longitudinal Studies in Criminology
- ▶ Monetary Strain and Individual Offending
- ▶ Optimizing Longitudinal Studies in Offending
- ▶ Pathways to Delinquency
- ▶ Psychopathy and Offending
- ▶ Risk Factors for Adolescent Sexual Offending

Recommended Reading and References

Becker HS (1963) Outsiders. Free Press, New York

Blokland A (2005) Crime over the life span: trajectories of criminal behavior in Dutch offenders. Netherlands Institute for the Study of Crime and Law Enforcement, Leiden

Blokland A, Nieuwbeerta P (2005) The effects of life circumstances on longitudinal trajectories of offending. Criminology 43:1203–1240

Blokland A, Nagin D, Nieuwbeerta P (2005) Life span offending trajectories of a Dutch conviction cohort. Criminology 43:919–954

Blumstein A, Cohen J, Roth JA, Visher CA (1986) Criminal careers and career criminals. National Academies Press, Washington, DC

Bouffard LA, Piquero NL (2010) Defiance theory and life course explanations of persistent offending. Crime Delinq 56:227–252

Farrington DP (2005) The integrated cognitive antisocial potential (ICAP) theory. In: Farrington DP (ed) Integrated developmental and life course theories of offending. Transaction, New Brunswick, pp 73–92

Glueck S, Glueck E (1950) Unraveling juvenile delinquency. Commonwealth Fund, New York

Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Hagan J (1993) The social embeddedness of crime and unemployment. Criminology 31:465–491

Laub JH, Sampson RJ (2003) Shared beginnings, divergent lives: delinquent boys to age 70. Harvard University Press, Cambridge

Lemert EM (1967) Human deviance, social problems, and social control. Prentice Hall, Englewood Cliff

Lofland J (1969) Deviance and identity. Prentice-Hall, Englewood Cliffs

Moffitt T (1993) Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. Psychol Rev 100:674–701

Nagin D, Paternoster R (1991) On the relationship of past and future participation in delinquency. Criminology 29:163–190

Sohoni T, Paternoster R, McGloin J, Bachman R (2013) Hen’s teeth and horses’ toes: the adult onsetter in contemporary criminology. Journal of Crime and Justice (forthcoming).

Patterson G, DeBaryshe B, Ramsey E (1990) A developmental perspective on antisocial behavior. Am Psychol 44:329–335

Sampson RJ, Laub JH (1993) Crime in the making: pathways and turning points through life. Harvard University Press, Cambridge

Sampson RH, Laub JH (1997) Al life-course theory of cumulative disadvantage and the stability of delinquency. In: Thornberry T (ed.) Developmental theories of crime and delinquency: Advances in criminological theory. Transaction Publishers, New Brunswick, NJ, pp 133–161

Sampson RJ, Laub JH (2005) A general age-graded theory of crime: lessons learned and the future of life-course criminology. In: Farrington DP (ed) Integrated developmental and life course theories of offending. Transaction, New Brunswick, pp 165–181

Savage J (2009) The development of persistent criminality. Oxford University Press, New York

Sherman LW (1993) Defiance, deterrence, and irrelevance: a theory of the criminal sanction. J Res Crime Delinq 30:445–473

Thornberry TP, Krohn MD (2005) Applying interactional theory to the explanation of continuity and change in antisocial behavior. In: Farrington DP (ed) Integrated developmental and life course theories of offending. Transaction, New Brunswick, pp 183–209

Tibbetts SG (2009) Perinatal and developmental determinants of early onset of offending: a biosocial approach for explaining the two peaks of early antisocial behavior. In: Savage J (ed) The development of persistent criminality. Oxford University Press, pp 179–201

Wikström PH (2005) The social origins of pathways in crime: Towards a developmental ecological action theory of crime involvement and its changes. In: Farrington DP (ed) Integrated developmental and life-course theories of offending, vol 14, Advances in criminological theory. Transaction, New Brunswick, pp 211–246

Wikström P-OH, Treiber K (2009) What drives persistent offending? The neglected and unexplored role of the social environment. In: Savage J (ed) The development of persistent criminality. Oxford University Press, Oxford, pp 389–422

Wolfgang ME, Figlio RM, Sellin T (1972) Delinquency in a birth cohort. University of Chicago Press, Chicago

Wolfgang ME, Figlio RM, Sellin T, Thornberry TP, Figlio RM (1987) From boy to man, from delinquency to crime. University of Chicago Press, Chicago

Careers in Organized Crime

Edward R. Kleemans and Vere M. van Koppen
VU University Amsterdam, Amsterdam,
The Netherlands

Overview

This entry elaborates upon theoretical and empirical insights into criminal careers in organized crime. The first section provides an overview of life-course criminology, older offenders, and adult onset. The second section briefly reviews literature on criminal careers in white-collar crime and concludes that situational context is a key concept in explaining a large proportion of involvement in white-collar crime. The third section explores involvement mechanisms in organized crime: recruitment, social ties, work ties, leisure activities and sidelines, and life events (including financial setbacks). The fourth section reviews major studies on criminal careers in organized crime and describes how the findings of these studies differ from the main findings of developmental and life-course criminology. The last section comprises the main conclusions and suggestions for further research.

Traditional research on criminal careers has mainly focused on juveniles, adolescents, and high-volume crime. Although much empirical progress has been made in developmental and life-course criminology (e.g., Farrington 2005), this research tradition tends to ignore certain kinds of offenders, particularly adult offenders, and certain types of crime, particularly organized crime, white-collar crime, and terrorism (Weisburd and Waring 2001; Leeper Piquero and Benson 2004; Steffensmeier and Ulmer 2005). Mainstream research focuses on property crimes and violent crimes, crimes for which opportunities are ubiquitous and open to anyone. Furthermore, many empirical longitudinal studies focus on a relatively early and short lifespan (Piquero 2008). These studies are therefore unable, almost by definition, to identify offenders who become engaged in crime only later on in life.

There are several reasons why criminal careers in organized crime may be different from high-volume crime. Anybody is able to steal something or commit a violent act, but things are somewhat more complex in organized crime and particularly in cases of cross-border crime or “transit crime.” Nowadays, many activities of organized crime groups boil down to international smuggling activities such as drug trafficking, smuggling illegal immigrants, human trafficking for sexual exploitation, arms trafficking, trafficking in stolen vehicles, and other transnational illegal activities, such as money laundering and evasion of taxes (e.g., cigarette smuggling, value-added tax fraud, and

European Community fraud) (Kleemans 2007). Such activities pose different requirements than traditional high-volume crime (Kleemans and De Poot 2008).

The first distinct feature is the greater importance of social relations in organized crime. It is simply impossible to get started without access to suppliers, clients, co-offenders, and profitable criminal opportunities. Trust is also important, as the financial stakes are high and the rules and mechanisms that make transactions in the legal world so much easier are absent: entering into contracts, paying via the official banking system, and – in case of disagreement – the availability of mediation or the judiciary (Reuter 1983). For this reason, existing social ties are used or illegal business relationships have to be built up. Not everyone has suitable ties that provide access to these profitable illegal opportunities, whereas building up such relationships takes time and energy.

The second distinct feature of organized crime is the transnational character of many of these criminal activities. Many types of organized crime are based on international smuggling activities. Yet many people lack international contacts that may connect them to these illegal opportunities, as social networks are typically constrained by laws of geographical and social distance (e.g., Feld 1981). Thus, not all offenders have access to these transnational contacts and some only later on in life.

A third common feature is that the activities are logistically considerably more complex and call for more co-offenders and specific expertise (e.g., Cornish and Clarke 2002). As more co-offenders are generally required for the successful commission of these crimes, seeking and finding suitable co-offenders is important (see, e.g., Tremblay 1993). Reliance on co-offenders from within one's own social circle is not always sufficient, as they may not possess the necessary capabilities. Contacts with the legal world are also salient for transport, money transactions, and shielding activities from the authorities. Many people lack these contacts and expertise, and some acquire them only later on in life, e.g., through their professional activities and contacts.

This entry elaborates upon theoretical and empirical insights into criminal careers in organized crime. The first section provides an overview of life-course criminology, older offenders, and adult onset. The second section briefly reviews literature on criminal careers in white-collar crime and concludes that situational context is a key concept in explaining a large proportion of involvement in white-collar crime. The third section explores involvement mechanisms in organized crime: recruitment, social ties, work ties, leisure activities and sidelines, and life events (including financial setbacks). The fourth section reviews major studies on criminal careers in organized crime and describes how the findings of these studies differ from the main findings of developmental and life-course criminology. The last section comprises the main conclusions and suggestions for further research.

Life-Course Criminology, Older Offenders, and Adult Onset

Developmental and life-course criminology place a strong emphasis on individual characteristics and long-term risk factors when explaining differences in criminal careers (see, e.g., Moffitt 1993). For example, Moffitt's well-known "dual taxonomy" distinguishes between a large group of people who engage in crime and antisocial behavior mostly during adolescence ("adolescence-limited" offenders) and a small group of people who are antisocial from an early age, engage in crime in adolescence, and remain active in crime and other forms of antisocial behavior throughout their lives ("life-course persistent"). The explanation for "life-course-persistent" criminal behavior is generally sought in fairly stable biological or psychological characteristics – often deficiencies – such as low intelligence, high impulsivity, or low self-control. Moffitt argues that early problem behavior is an indicator that a person belongs to the life-course-persistent group and that antisocial and criminal behavior is likely to continue into later life.

Gottfredson and Hirschi (1990) follow a similar route in their "general theory of crime." Although they deny the very existence of different "types" of offenders, their basic approach is quite comparable: people develop self-control early in life, and those who lack sufficient self-control engage in criminal acts and similar behavior – with short-term benefits and long-term costs such as drinking, smoking, and adultery – at higher rates than others throughout their lives. Many developmental and life-course theories of offending, though often more nuanced and empirically grounded, share this strong emphasis on individual characteristics and long-term risk factors and treat antisocial behavior in children as essentially similar to offending in adults (Farrington 2005).

In this light, many theories assume that onset occurs early in life and that adult offending requires childhood antisocial behavior (see, e.g., Robins 1978). From this perspective, adult-onset offending is a rare phenomenon that needs no serious research effort (e.g., Gottfredson and Hirschi 1990). As a result, adult onset is an anomaly in developmental and life-course criminology and is treated in the same way as anomalies are generally treated in "normal science" (e.g., Kuhn 1962). Anomalies are discovered now and then but are often brushed aside empirically or theoretically (for an overview on adult onset, see Van Koppen et al. 2010).

The general neglect of adult-onset offending can be attributed to two major factors. First, almost all research on criminal careers is concerned with high-volume crime, in which adult-onset offending seems relatively scarce. Second, the majority of trajectory research focuses on a relatively small (and mostly young) age frame and is therefore unable to identify adult-onset offenders by definition. Piquero (2008) provides an overview of all research using trajectory methodology in studying criminal activity over the life course between 1993 and 2006. These 90 studies reflect about 50 different samples. Only half of the samples include individuals after age 18, and only a fifth have data available after age 30. Birth cohort studies that do follow individuals long enough into adulthood systematically identity a significant share of adult-onset offenders. A quarter up to half of the criminal samples of cohorts drawn in Racine, Stockholm, and Philadelphia turn out to start offending in adulthood (e.g., Eggleston and Laub 2002).

Hence, the dominant way in which criminal careers are studied – longitudinally, over a short (adolescent) life-span, using self-reports with limited follow-ups, and mainly focused on high-volume crime and (re)traceable young offenders – reinforces received wisdoms that might be challenged by other types of research, focusing on other types of crime, such as white-collar crime and organized crime.

Criminal Careers in White-Collar Crime

Over the past years, a number of studies have examined criminal careers in white-collar crime. Although there is a difference between white-collar crime and organized crime, both concern more uncommon types of crime, and both may reveal different patterns than the ones discovered in more traditional research. Until 1980, most of what was known concerning white-collar offenders and their crimes was extracted from case studies. From 1980 on, datasets were constructed containing white-collar crimes committed in the United States (Leeper Piquero and Benson 2004; Weisburd and Waring 2001). These new efforts immediately refuted a common belief about white-collar offenders. They are not all one-time offenders, as had been stated before (see, e.g., Edelhertz and Overcast 1982). Almost half of the white-collar offenders have had at least one other official contact with the criminal justice system (Benson and Moore 1992). Furthermore, these criminal careers turn out to differ substantially from careers of high-volume crime offenders on several aspects. Or as noted by Leeper Piquero and Benson (2004), "the typical white-collar offender greatly differs from the typical street offender and does not appear to fit into the proposed explanations of life-course-offending patterns."

Based on their findings on white-collar criminals, Leeper Piquero and Benson (2004) propose a theory on career patterns in white-collar crime called *punctuated situationally dependent offending*. It assumes that white-collar criminals are criminally active during adolescence (just like high-volume crime offenders), then desist for a while, and begin to offend again when they reach their 30s or 40s. External factors are used to explain this revival later in life. One explanation involves a personal or occupational crisis the offender is going through. A second explanation relates to opportunities presented when a certain occupational status has been reached. Situational context is a key concept in explaining a large proportion of involvement in white-collar crime (see for a review, e.g., Benson and Simpson 2009). Results of a study by Weisburd and Waring (2001) corroborate this idea: the white-collar offenders studied differ on various career aspects from offenders of high-volume crime, the age of onset was relatively late, and they were criminally active for a substantially longer period and were arrested for fewer offenses throughout their criminal career. These lines of research show the importance of incorporating new groups of offenders.

Organized Crime: Involvement Mechanisms

Before contemporary research on criminal careers in organized crime will be discussed further below, several involvement mechanisms are presented, explaining processes and situations that provide opportunities for organized crime. How do people get involved in organized crime?

Recruitment

According to the standard idea of recruitment, criminal "organizations" recruit outsiders who may move up in the hierarchy by proving their suitability. They start by doing dirty jobs and are ultimately rewarded. Behind this idea lies the assumption of a more or less static criminal organization (cf., e.g., Cressey 1969), while research shows that, much more often, the world of organized crime consists of criminal networks in which offenders may cooperate with one another in changing constellations (for a review, see, e.g., Morselli 2009). Within these criminal networks, there often are nodal actors, on whom many other offenders depend because of their money, knowledge, and contacts. These nodal actors resurface time and again, in different criminal investigations and different criminal collaborations. Still, other offenders may gradually become less dependent on these nodal actors because they accumulate money, knowledge, and contacts for themselves and subsequently develop increasingly more criminal activities of their own.

Social Ties and the Social Snowball Effect

Social ties are very important in organized crime. Opposite to the standard view of recruitment, social ties create a "social snowball effect": people get involved in organized crime through their family, friends, and acquaintances; as they go along, their dependency on the resources of other people (such as money, knowledge, and contacts) gradually declines; and they generate new criminal collaborations. They involve people from their own environment in their activities, keeping the snowball in motion (Kleemans and Van de Bunt 1999: 29–34). Examples comprise partners, girlfriends, spouses, sons, daughters, brothers, sisters, and half-brothers, who get involved in criminal activities; a father who starts helping his son in drug trafficking; or a smuggled Iranian who ultimately becomes responsible for human smuggling operations himself. Hence, the choices people make are prestructured, and the opportunities that present themselves are not presented to anyone.

Similar prestructured choices relate to three other involvement mechanisms, which are encountered when analyzing starters in organized crime. Within the context of the Dutch Organized Crime Monitor, 120 major investigations into organized crime groups have been analyzed systematically (see for more information: Kleemans 2007). In 2008, the criminal careers of 92 "starters" were analyzed, suspects that did not have any other judicial contacts before and who had not "progressed" from high-volume crime into organized crime (Kleemans and De Poot 2008). Next to deliberate recruitment, emphasis was placed upon social ties and the social snowball effect, work ties, leisure activities and sidelines, and life events (including financial setbacks).

Work Ties

Generally, people associate occupations with white-collar crime, yet organized crime is equally dependent on contacts with the licit world. Therefore, occupations are also important in explaining involvement in organized crime. Just like social ties cross boundaries between legality and illegality, so do some work settings in which people spend a considerable share of their daily life. Several occupational characteristics may help in providing opportunities for organized crime (Kleemans and Van de Bunt 2008). First, the social nature of work activities may create opportunities; individuals who often interact with others can meet with potential co-offenders. Second, international contacts and travel movements may create such opportunities. Contacts with other countries and other social groups enable someone to discover and effectuate certain possibilities to commit (transnational) criminal activities. Examples are occupations involving mobility, transport, and logistics. It is the story of the fruit trader who also uses his contacts and knowledge for the import and export of drugs; airline personnel who travel on a regular basis, and who know how to evade baggage checks in airports; and an international trader in clothes who also engages in human trafficking and underground banking. Third, the individual freedom of action of particular occupations is important, or in other words: autonomy. This explains the involvement of directors of (usually) small businesses, independent professionals, and relatively autonomous persons employed by larger organizations, such as corporations and banks. It also explains the relative absence of professionals in the salaried employment of the government or large organizations. Similar findings on occupations that emphasize the assets of international social contacts and autonomy follow from interviews with incarcerated offenders, most notably highlighting the salience of people working in logistically important sectors such as boating, fishing, import and export, transport, and the airline sector (e.g., Dorn et al. 2005).

Leisure Activities and Sidelines

The chances of getting into contact with organized crime differ from person to person and from occupation to occupation. Yet people from different social worlds, involved in disconnected networks, may also run into each other through leisure activities and sidelines. Local bars, nightlife, parties, after parties, drug joints, brothels, shooting clubs, or motor clubs, these are all places where all sorts and conditions of people can meet. These meeting places are especially important for forging contact between representatives of the “underworld” and the “licit world,” who would not meet each other so easily in everyday life. This is why these places also play a vital role in the coming about of involvement in organized crime.

Life Events and Financial Setbacks

A final involvement mechanism relates to life events. In developmental and life-course criminology, the emphasis is on the positive effects of life events such as getting a permanent job, a stable relationship, or having children (Laub and Sampson 2003). These positive life events make problems like crime disappear like snow in summer. Involvement in organized crime, however, results from the negative sides of life events, particularly among men in middle age: divorce, illness, disability, unemployment, and bankruptcies; these are all moments when people are suddenly confronted with setbacks, socially and financially. They are down and out, bankrupt or deeply in debt, and subsequently some of them weigh pros and cons differently when generous financial investors appear on the scene or lucrative illegal opportunities present themselves.

What these five involvement mechanisms have in common is that the opportunities presenting themselves to people are prestructured in a way which those people do not necessarily have to realize. Options differ from person to person and are different at the beginning of one’s life course compared to later on in life.

This is also one of the reasons why criminal careers in organized crime may evolve differently than traditional high-volume crime careers, as the options are not open to anyone and generally only provide themselves later on in life for a selected group of criminals and non-criminals.

Criminal Careers in Organized Crime

Empirical research into criminal careers in organized crime is scarce, yet recent studies have challenged several basic facts and received wisdoms of developmental and life-course criminology. For one, the age-crime curve describing that in general the prevalence of offending rises steeply in the early teenage years and reaches a peak between the ages of 15 and 17, followed by a decline over the rest of the life course. For another, the dominance of individual differences (risk factors) explaining persistence in offending for a very small group of offenders ("life course persistent offenders") as opposed to the majority of "adolescence-limited offenders" (e.g., Moffitt 1993).

First, juveniles are almost absent in organized crime, and most offenders are quite old, in their 30s, 40s, 50s, or even older. Second, a large share of organized crime offenders get involved in crime only at a later age. Late starters are not exceptional, as in traditional research on criminal careers, yet comprise a substantial group of offenders (Kleemans and De Poot 2008). This finding of a substantial presence of "late-onset" offenders is robust across several different criminal activities (drugs, fraud, and other activities) and different roles in criminal groups (leaders, coordinators, and lower-level suspects) (Van Koppen et al. 2010). These findings contradict the general idea that crimes at a later age should be preceded by crimes in adolescence and that individual characteristics and long-term risk factors are the major explanation for a life in crime, let alone a life in serious and organized crime.

The scarce literature on organized crime offenders often resonates the preoccupation with special characteristics of offenders. For instance, Bovenkerk (2000) states that certain characteristics that are conducive to a career in business (based on the "Big Five" from test psychology) are also conducive to a career in organized crime. Basically, the personality traits of successful legal entrepreneurs would be mirrored in the personality traits of illegal entrepreneurs: *extraversion, controlled impulsiveness, a sense of adventure, megalomania, and narcissistic personality disorder*. This line of reasoning is based upon ideas of "the survival of the fittest" in which only offenders with certain characteristics end up in organized crime or in certain leadership positions.

However, there are important lines of research that raise objections against this traditional focus on individual characteristics. Some offenders of organized crime are quite normal and well adapted in many respects, even though they are involved in serious forms of crime. A significant amount of offenders have legitimate jobs and sources of income. Moreover, not all crimes are the same nor just symptoms of latent characteristics such as low self-control and a variety of people can become involved in organized crime (e.g., Steffensmeier and Ulmer 2005). Next to individual personality differences, these studies draw attention to situational factors that may be important to explain involvement in organized crime.

Most of what is known on organized crime offenders, comes from (auto)biographies and case studies of mafia groups (e.g., Paoli 2003; Steffensmeier and Ulmer 2005). These studies, based on memoirs and testimonies, illustrate the complexity of a criminal life course, including contacts, situational context, and processes such as persistence and desistance. Morselli (2005) describes in detail the careers of Howard Marks and "Sammy The Bull" Gravano. Based upon Burt's work on brokerage (Burt 2005), he has put forward the idea that "brokers do better" and that advancements in criminal careers can be explained by deliberate choices people make in investing in certain relationships. Based upon Marks' autobiography, Morselli focuses on the personal network that explains why Marks was able to be successful in the cannabis trade. Central issues are brokerage (linking disconnected networks), the number of (non-redundant) contacts, and social network dynamics.

One of the main problems of case studies is that one may wonder whether these case studies reflect the average or exceptional offender in organized crime and whether these case studies are able to capture the full range of different pathways in organized crime. Case studies may be biased toward more interesting, more talkative or defective, atypical offenders. Moreover, autobiographies often focus on people with long criminal careers instead of "late-onset" offenders. Two studies, however, comprise a broader, comparative view on several criminal careers in organized crime.

The first study is based upon quantitative and qualitative information about the criminal careers of 979 suspects who were involved in 79 different organized crime cases that have been analyzed in the context of the Dutch Organized Crime Monitor (Kleemans and De Poot 2008). First, key figures of these criminal careers are compared to careers of the total offender population. Offenders in the studied organized crime cases are mostly men between 30 and 50 years old. They are generally older than the total offender population, and the studied group of suspects includes no juveniles at all. The age distribution seems to reflect the fact that it is a highly specific group that does not come within the focus of traditional research into criminal careers. Kleemans and De Poot also analyzed criminal careers of 92 "starters," suspects who did not have any other judicial contacts before and who had not "progressed" from high-volume crime into organized crime. Five involvement mechanisms were described in more detail: deliberate recruitment by criminal groups, social ties and the social snowball effect, work ties, leisure activities and sidelines, and life events (including financial setbacks) (see above). Finally, they analyzed the careers of 66 (ring) leaders and "nodal" offenders in more detail to find out how their criminal careers had developed. Some (ring) leaders had a long criminal career: the versatile but regionally constrained "local hero" (versatility as a result of both opportunities and constraints of the local context); progression through an increase in scale (specialization, particularly in international drug trafficking); progression through capital accumulation (investing in drug transports (while refraining from hands-on activities) and becoming a "background operator"; and expertise, contacts, and network formation (people getting a central position in criminal networks because many offenders and criminal groups need their contacts and expertise). Surprisingly, a significant share of (ring) leaders could be typified as "late-onset" offenders, who had made a switch from a legal occupational background to organized crime. Among these 32 offenders, a distinction could be made between a group of 19 with a background in legal trade (including import and export) and a group of 13 with other types of occupation – people from the business sector, the construction industry, assembly, hotels and catering, financial services, or government. Three different types of "late onset" were described: criminal activities extending from legal activities (opportunities that arise during day-to-day work, particularly in fraud cases); people who obviously switch careers, from legal to illegal commodities, motivated by the big profits that can be obtained in the trade in prohibited commodities such as narcotics; and finally, people seizing upon criminal opportunities only later on in life after specific, significant life events (including financial setbacks and problematic debt situations).

Both the research on "starters" and on "late-onset (ring) leaders" illuminate the mechanisms behind the "anomaly" of "adult onset." In essence, the story is that explaining involvement and success in organized crime requires more than a focus on individual differences and long-term risk factors. Kleemans and De Poot (2008) explain their findings by *social opportunity structure* – social ties providing access to profitable criminal opportunities. Social opportunity structure is unequally distributed across the population and across age. Certain kinds of people have social ties that give them direct access to illegal opportunities. Major examples are migrants, who connect source countries of narcotics to Western consumer markets, and people with occupations that are connected to mobility, transport, and logistics. Furthermore, social opportunity structure is unequally distributed across age. Some offenders lack necessary social ties at a younger age, which explains why some offenders become involved in organized crime only at a later age. It also explains the interesting phenomenon of "late starters" – people without any appreciable criminal history – and legally employed people switching careers.

The second study that takes a comparative view on careers in organized crime was a follow-up study by Van Koppen et al. (2010). They elaborated upon this research and investigated criminal trajectories of 854 suspects, based upon quantitative and qualitative data from the Dutch Organized Crime Monitor. A semiparametric group-model was used to cluster 854 individuals into groups with similar developmental trajectories. The most important finding of this study relates to the substantial group of adult-onset offenders (40 %) and a group without any previous criminal records (19 %), next to a group of early starters (11 %) and a group of persisters (30 %). Up till then, no trajectory study had ever discovered such a vast share of adult-onset offenders. Furthermore, the findings turn out to be quite robust across different kinds of criminal activities (drugs, fraud, and other criminal activities) and different roles in criminal groups (leaders, coordinators, and lower-level suspects). Adult-onset offenders emerge in several kinds of criminal activities and in several kinds of roles in criminal groups. Hence, it is not true that the amount of fraud cases in the sample would explain the substantial share of late-onset offenders; late onset is equally important in other types of criminal activities such as drug trafficking. Neither could roles in criminal groups explain "late onset": late onset seems to be equally present in lower-level suspects as in coordinators and (ring) leaders. The mechanisms described in the study by Kleemans and De Poot (2008) may explain these robust findings, as these mechanisms may equally apply to different kinds of criminal activities and to people with different types of roles in criminal groups.

Conclusion and Challenges for Future Research

Differences between organized crime and high-volume crime may explain several interesting findings that challenge basic findings from developmental and life-course criminology. First, juveniles are almost absent in organized crime, and most offenders are quite old, in their 30s, 40s, 50s, or even older. Second, a large share of offenders get involved in organized crime only at a later age. Late starters are not exceptional, as in traditional research on criminal careers, yet comprise a substantial group of offenders. This finding of a substantial presence of "late-onset" offenders is robust across several different criminal activities (drugs, fraud, and other activities) and different roles in criminal groups (leaders, coordinators, and lower-level suspects). These findings contradict the general idea that crimes at a later age should be preceded by crimes in adolescence and that individual characteristics and long-term risk factors are the major explanation for a life in crime, let alone a life in serious and organized crime. The findings can also be explained by different involvement mechanisms that play a role in organized crime. Next to deliberate recruitment, four other involvement mechanisms play a key role: social ties and the social snowball effect, work ties, leisure activities and sidelines, and life events (including financial setbacks). What these involvement mechanisms have in common is that the opportunities presenting themselves to people are prestructured. Options differ from person to person and are different at the beginning of one's life course compared to later on in life. This is also one of the reasons why criminal careers in organized crime may evolve differently than traditional high-volume crime careers, as the options are not open to anyone and generally only provide themselves later on in life for a selected group of criminals and non-criminals.

Kleemans and De Poot (2008) use the theoretical concept of *social opportunity structure* – social ties providing access to profitable criminal opportunities. Social opportunity structure is unequally distributed across the population and across age. Certain kinds of people have social ties that give them direct access to illegal opportunities. Major examples are migrants, who connect source countries of narcotics to Western consumer markets, and people with occupations that are connected to mobility, transport, and logistics. Furthermore, social opportunity structure is unequally distributed across age. Some offenders lack necessary social ties at a younger age, which explains why some offenders become involved in organized crime only at a later age. It also explains the interesting phenomenon of "late starters" – people without any appreciable criminal history – and legally employed people switching careers.

An important new insight from the literature is that occupations do not only play a major role in white-collar and corporate crime but in organized crime as well. Many studies on organized crime tend to draw a sharp distinction between licit activities and illicit activities and between the licit world and the underworld. Some studies focus on the infiltration or extortion of licit economic sectors by organized crime, while other studies tend to emphasize symbiosis and "interfaces" between legal and illegal actors. Empirical research, however, indicates that both worlds may converge. Occupations, work relations, and work settings may provide the breeding ground for organized crime activities, particularly transit crime. Furthermore, occupations offer work relations and work settings that are directly relevant for providing opportunities to search for suitable co-offenders, to gain trust, and to exercise control. Social ties cross boundaries between legality and illegality, as do some work settings in which people spend a considerable share of their daily life. Empirical research highlights the salience of social relations, work relations, and work settings for criminal activities, also in cases of organized crime. From a theoretical point of view, these findings are important in bridging ideas from social network theory and opportunity theory. Social relations are influenced by "foci of interaction" (Feld 1981). Social relations do not happen at random but often obey the laws of social and geographical distance, producing a certain kind of clustering based on geographical distance, ethnicity, education, age, etc. (e.g., Burt 2005). For many people, work is also an important focus of interaction. Second, social relations and work relations often coincide, either because people find a job through social relations or because working relationships develop into more or less close social relations. Third, the blending of work relations and social relations may also contribute to the effective concealment of illegal activities. For a variety of reasons, potential witnesses may remain silent, look the other way, or feel reluctant to blow the whistle. Effective concealment does not require secret societies. "Walls of silence" may also emerge in licit worlds that are mobile, dynamic, and social (Van de Bunt 2010).

Studies on criminal careers in organized crime pose several challenges for criminal career research. First, these studies might refocus attention to social context and co-offenders promoting certain pathways in crime. Co-offending is still an understudied aspect in criminal career research. Second, these findings challenge the traditional boundaries between legality and illegality. Social ties cross boundaries between legality and illegality, as do some people's activities at particular times in their lives. Third, the studies on organized crime careers suggest that criminal career research should focus not only on adolescence but also on later stages in life, as important changes in criminal careers occur with age. Fourth, studying these stages in life should go beyond a preoccupation with desistance. Steffensmeier and Ulmer (2005) state that desistance cannot be explained by stakes in conformity alone; another important factor is a lack of profitable criminal opportunities or the reverse. One could argue that this is particularly salient during stages in life when making money is more important than in adolescence. It could also provide an answer to the question why successful offenders in organized crime never seem to "retire" and continue along the path of crime. Finally, some people seize upon illegal opportunities, while others do not. Clearly, there is both "agency" and "opportunity" (Laub and Sampson 2003). The question why, when confronted with opportunities, some people turn to crime and others do not, and under which conditions, is still unresolved and needs further exploration.

Related Entries

- ▶ Age-Crime Curve
- ▶ Career Criminals and Criminological Theory
- ▶ Co-offending
- ▶ Criminal Careers
- ▶ Criminal Careers and Public Policy Responses
- ▶ Moffitt's Developmental Taxonomy of Antisocial Behavior
- ▶ Organized Crime, Types of
- ▶ Pathways to Delinquency
- ▶ Social Network Analysis of Organized Criminal Groups

Recommended Reading and References

Benson ML, Moore E (1992) Are white-collar and common offenders the same? An empirical and theoretical critique of a recently proposed general theory of crime. J Res Crime Delinq 29:251–272

Benson M, Simpson S (2009) White collar crime: an opportunity perspective. Routledge, New York

Bovenkerk F (2000) Wanted: Mafia boss: essay on the personology of organized crime. Crime Law Soc Change 33:225–242

Burt RS (2005) Brokerage & closure: an introduction to social capital. Oxford University Press, Oxford

Cornish DB, Clarke RV (2002) Analyzing organized crime. In: Piquero R, Tibbetts SG (eds) Rational choice and criminal behavior: recent research and future challenges. Garland, New York, pp 41–62

Cressey DR (1969) Theft of the nation: the structure and operations of organized crime in America. Harper & Row, New York

Dorn N, Levi M, and King L (2005) Literature review on upper level drug trafficking. Home Office Online Report 22/05

Edelhertz H, Overcast TD (1982) White collar crime: an agenda for research. Lexington Books, Lexington

Eggleston EP, Laub JH (2002) The onset of adult offending: a neglected dimension of the criminal career. J Crime Justice 30:603–622

Farrington DP (2005) Integrated developmental and life-course theories of offending: advances in criminological theory 14. Transaction Publishers, New Brunswick

Feld SL (1981) The focused organization of social ties. Am J Sociol 86(5):1015–1035

Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Kleemans ER (2007) Organized crime, transit crime, and racketeering. In: Tonry M, Bijleveld C (eds) Crime and justice in The Netherlands: crime and justice: a review of research, vol 35. University of Chicago Press, Chicago, pp 163–215

Kleemans ER, De Poot CJ (2008) Criminal careers in organized crime and social opportunity structure. Eur J Criminol 5(1):69–98

Kleemans ER, Van de Bunt HG (1999) The social embeddedness of organized crime. Trans Organized Crime 5:19–36

Kleemans ER, Van de Bunt HG (2008) Organised crime, occupations, and opportunity. Global Crime 9(3):185–197

Kuhn TS (1962) The structure of scientific revolutions. University of Chicago Press, Chicago

Laub JH, Sampson RJ (2003) Shared beginnings, divergent lives: delinquent boys to age 70. Harvard University Press, Cambridge

Leeper Piquero N, Benson ML (2004) White collar crime and criminal careers: specifying a trajectory of punctuated situational offending. J Contemp Crim Justice 20:148–165

Moffitt TE (1993) Adolescence-limited and life-course persistent anti-social behavior: a developmental taxonomy. Psychol Rev 100:674–701

Morselli C (2005) Contacts, opportunities, and criminal enterprise. University of Toronto Press, Toronto

Morselli C (2009) Inside criminal networks. Springer, New York

Paoli L (2003) Mafia brotherhoods: organized crime, Italian style. Oxford University Press, New York

Piquero AR (2008) Taking stock of developmental trajectories of criminal activity over the life course. In: Liberman AM (ed) The long view of crime: a synthesis of longitudinal research. Springer, New York

Reuter P (1983) Disorganized crime: illegal markets and the Mafia. MIT Press, Cambridge

Robins LN (1978) Sturdy childhood predictors of adult antisocial behaviour: replications from longitudinal studies. Psychol Med 8:611–622

Steffensmeier D, Ulmer JT (2005) Confessions of a dying thief: understanding criminal careers and criminal enterprise. Transaction Aldine, New Brunswick

Tremblay P (1993) Searching for suitable co-offenders. In: Clarke RV, Felson M (eds) Routine activity and rational choice: advances in criminological theory, vol 5. Transaction Publishers, New Brunswick, pp 17–36

Van de Bunt H (2010) Walls of secrecy and silence. The Madoff case and cartels in the construction industry. Criminol Public Policy 9(3):435–453

Van Koppen MV, de Poot CJ, Kleemans ER, Nieuwbeerta P (2010) Criminal trajectories in organized crime. Br J Criminol 50(1):102–123

Weisburd D, Waring E (2001) White collar crime and criminal careers. Cambridge University Press, Cambridge

Case Management

▶ Behavioral Management in Probation

Causal Processes

▶ Situational Action Theory

Cause of Police Legitimacy

▶ Democratic Policing

Causes

▶ Establishing Causes of Offending in Longitudinal and Experimental Studies

Causes of Police Legitimacy

Tammy Rinehart Kochel
Southern Illinois University, Carbondale,
IL, USA

Overview

Police legitimacy plays a crucial role in promoting a variety of crucial, pro-social outcomes. Research demonstrates the capacity of police legitimacy to promote law-abiding behaviors, cooperation with police, reporting crime and suspicious behaviors, support for police, and public collaboration with police (Fagan and Tyler 2004; Fagan and Tyler 2005; Gau 2011; Kochel et al. 2011; Reisig et al. 2007; Sunshine and Tyler 2003; Tankebe 2009; Tyler 1990; Tyler 2006; Tyler and Fagan 2008). These outcomes are particularly important to the efficiency and effectiveness of police within democratic contexts. Police rely on public support for resources and authorization, and even more importantly, police are expected to generate *voluntary* compliance to the law and to police directives by a majority of citizens, rather than using force and fear to coerce cooperation (Tyler 2004). Jackson and colleagues (2012) suggest that normative-based compliance (such as that motivated by perceptions of police legitimacy) is more stable than instrumentally motivated compliance. Police cannot be at all places at all times, thus minimizing the potential effectiveness of a deterrence-approach for securing order. Instead, police in democratic contexts rely on a social culture supportive of law-abiding behavior and order. Police legitimacy serves as a basis for this social culture (see "▶ Procedural Justice, Legitimacy, and Policing" and "▶ Police Legitimacy and Police Encounters" in this volume to learn about the consequences of police legitimacy). Views about police legitimacy and other judgments about police tend to be fairly stable in the United States and elsewhere, but scholars have reported declines in police legitimacy in recent decades LaFree 1998; Tyler et al. (1997).

Thus, understanding the causes of or contributors to police legitimacy and whether patterns of police legitimacy and the factors that promote it vary under different contexts or across different characteristics of people has important practical implications for police, the public, and democracies in general. Tyler (1990) suggests that policing style and the process of policing may influence individuals' judgments about police legitimacy. This encyclopedia entry describes, based on past theoretical and empirical research, three factors that may contribute to police legitimacy – procedural justice, crime control effectiveness, and distributive justice. The entry describes the three models, their relative impacts, and whether these effects vary under different conditions.

Background: Defining Legitimacy

Police legitimacy is a normative perspective that police authority is valid and ought to be obeyed and respected. Police have legitimacy when members of the public believe police have a right to their authority and thus that citizens have a moral obligation and civic duty to honor, accept, and adhere to that authority (Tyler and Huo 2002). The concept of institutional legitimacy originates with Max Weber (1947). He first explained that citizens who view an authority as legitimate feel a personal sense of responsibility to act consistently with that authority's expectations, even in the absence of that authority and even when the compliant behavior contradicts self-interests.

Although this definition of legitimacy is widely acknowledged, research examining the correlates and consequences of police legitimacy has suffered from construct validity problems (Gau 2011; Reisig et al. 2007). Scholars have sometimes operationalized legitimacy to include the anticipated consequences of police legitimacy, such as compliance with the law or cooperation with police, and at other times, legitimacy measures incorporate theorized antecedents, especially elements of procedural justice. Additionally, while most studies of police legitimacy incorporate into the measure a perceived obligation to obey, some of Tom Tyler's studies (the present day scholar who refocused the criminologists and criminal justice practitioners on this issue with a series of survey studies) and other studies incorporate moral trust or institutional trustworthiness as a component to legitimacy. This component is not part of Weber's initial construct, and recent psychometric studies have suggested that trust does not combine with views about the obligation to obey to form a legitimacy measure (Gau 2011; Reisig, et al. 2007). Thus, this entry aims to provide the status of current research on police legitimacy conceptualized as the perceived moral obligation to voluntarily obey police, while avoiding some of the construct validity pitfalls experienced by it.

Key Issues

One way individuals form opinions about whether and the degree to which police authority is legitimate is based on personal or vicarious experiences with police. While Tyler (1990) found that only about four percent of the variance of legitimacy in a Chicago sample was explained by individual personal experiences with police, Tyler and Huo (2002) and Tyler (2001) saw a much stronger impact of personal experience on legitimacy judgments, explaining up to 34 % of the variation in legitimacy. Other sources include stories and media coverage about police behaviors and accomplishments, peer assessments of police, generalized understandings about the institution of police, and demographic and cultural characteristics of the individual and context that may color individuals' perceptions. Past scholarship has focused on how three aspects of these experiences and attitudes may contribute to police legitimacy. Procedural justice, distributive justice, and police effectiveness are the three antecedents expected to promote police legitimacy.

The Process-Based Model/Procedural Justice Model

Procedural justice refers to an impartial process that is not affected by race, gender, age, or other demographic characteristics: one in which people have a voice, one where different people in similar situations are treated consistently, a process whereby decisions are based on facts, one in which people are treated with dignity and respect, and a process protected by accountability to higher authorities (Leventhal 1976). The process-based model or procedural justice model purports that when members of the public perceive that police behave fairly and respectfully toward citizens, this generates a perception that the police and the public share similar values and beliefs. Lipset (1959), writing about the time of Weber, suggests that social institutions acquire legitimacy based on whether the institution's values align with its constituents.

Gau (2011) explains that being treated well by police may indicate to people that they are a valued member of society, affirming the individual's status as a member of the group, which includes police. This belief in a common moral perspective with police or common group membership encourages acceptance of a moral obligation to obey police authority and promotes a cultural value supportive of lawfulness (Tyler and Huo 2002). Tyler (1997) links this process back to social identity theory, whereby individuals gain information about their identity and status by how authorities treat them. When individuals feel that they have been treated respectfully by an authority, this action signifies their accepted group status and they respond with deference.

Of course, not all members of the public experience personal contact with the police from which to assess procedural justice. Tyler and others assert that residents can draw on their general perceptions about the character of police to form opinions about procedural justice and thus legitimacy. Additionally, the process-based model does not require an objective assessment of the fairness of police procedures. Rather, what matters are individuals' subjective opinions about police fairness, which derive from a variety of perspectives and experiences.

The types of measures that are typically used to operationalize procedural justice include items asking whether police treat people fairly or with dignity and respect, make decisions based upon facts, explain their actions, listen to people before making decisions, respect people's rights, try to help people solve problems, or have the best interest of a group or area at heart. Some studies also include whether police fairly decide whom to stop and question or arrest or whether police apply too much force under certain circumstances.

The Instrumental/Performance-Based Model

Lipset (1959), who advocated the importance of the process-based model, also contends that social institutions gain legitimacy when they can demonstrate prolonged effectiveness. Lipset suggests that both shared values *and* effectiveness can provide the source of legitimacy, but prolonged effectiveness in the absence of shared values offers only an unstable form of legitimacy for a social institution.

The instrumental or performance-based model of legitimacy suggests that residents accept and conform to police authority when they perceive that doing so will produce beneficial outcomes or avoid negative ones. To gain benefits, residents must see police as a useful resource. Residents who perceive that police are competent and capable of reducing crime, bringing offenders to justice, creating and maintaining order, reducing fear of victimization, and improving neighborhood conditions will grant them legitimacy and subsequently provide their cooperation and adherence to police authority because they believe that doing so will achieve anticipated goals of better safety and less crime. In other words, they see police as an authority with the capacity to secure compliance and thus grant them legitimacy. Thus, when members of the public experience positive interactions with police or perceive these favorable outcomes are being and can be obtained, the performance-based model expects higher police legitimacy. However, when residents doubt the capacity of police to produce order, control crime, and hold criminals accountable, an instrumental model predicts that residents will be less inclined to acknowledge or feel obligated to conform to police authority, because they have little to gain.

The types of questions that are typically used to measure police performance address assessments about response time, the effectiveness of police at addressing social disorder, crime, or citizen fear, as well as general assessments about the safety of a neighborhood and other neighborhood conditions.

The Distributive Justice Model

The distributive justice model suggests that residents are more apt to view police as legitimate when they perceive that police services are equitably distributed and not motivated by characteristics about people or places (e.g., wealth, social class, race/ethnicity, political affiliation, geographic location) that are unrelated to the police mission. This is not to suggest that police attention would have to be evenly or equally allocated, rather that the allocation of resources is perceived as appropriate. The distributive justice model suggests that police will have greater legitimacy when residents perceive everyone is getting suitable attention and services from the police. Judgments about what is equitable are subjective and, like the other models, do not require personal interactions with police in order to form them.

The types of questions that are typically used to measure distributive justice generally ask whether groups are treated equally based on social characteristics such as race or wealth, or whether a certain group is treated better by police or receives better quality outcomes from police.

Weight of the Evidence

Much of the evidence about the antecedents of police legitimacy is based on citizen surveys across cities in the United States. Tom Tyler conducted a series of these studies in Oakland, California, Los Angeles, New York City, and Chicago (Sunshine and Tyler 2003; Tyler 1990; Tyler 2001; Tyler and Huo 2002; Tyler and Fagan 2008). Since this initial contribution, other scholars have also contributed, examining the antecedents of legitimacy in other US cities (Gau 2011) and in other developed and developing countries such as Australia (Hinds 2007; Hinds and Murphy 2007; Murphy et al. 2008), England (Hough et al. 2010; Jackson et al. 2012), Ghana (Tankebe 2008), Israel (Jonathan-Zamir and Weisburd 2011), Jamaica (Reisig and Lloyd 2009), and Trinidad and Tobago (Kochel 2012; Kochel et al. 2011).

Past research has shown some support for each of the three antecedents of legitimacy, suggesting that the pathway to legitimacy is not one dimensional. For example, Papachristos and colleagues (2009) show that a measure combining elements of police performance, procedural justice, and distributive justice was a strong positive predictor of legitimacy. Hinds and Murphy (2007) as well as Murphy et al. (2008) found a positive association among Australians between legitimacy and each of procedural justice, police performance, and distributive fairness. Tyler's (2002) New York sample showed the same result. Generally, however, these predictors do not have *equal* influence.

Past research strongly supports procedural justice as the most consistent and influential antecedent: the primary mechanism for promoting police legitimacy. Studies conducted by Sunshine and Tyler (2003); Fagan and Tyler (2004); Tyler et al. (2010), as well as Van der Toorn et al. (2011) in New York, Tyler (1990) in Chicago, Tyler and Huo (2002) in Los Angeles and Oakland, by Hinds (2007), Hinds and Murphy (2007), and Murphy et al. (2008) in Australia, by Hough and colleagues (2010) as well as Jackson and colleagues (2012) in England and Wales, Kochel (2012) and Kochel et al. (2011) in Trinidad and Tobago, and by Reisig and Lloyd (2009) in Jamaica have all found support that when individuals perceive the police as being more procedurally just, they also view them with higher levels of legitimacy and that procedural justice is a stronger predictor than distributive justice or perceived effectiveness. Kochel (2012) further demonstrates that these relationships carry through to the neighborhood level. In Trinidad and Tobago neighborhoods, when residents perceived higher levels of procedural injustice, a smaller portion of residents reported normative obligations to accept police decisions. Other studies demonstrating the important positive influence of procedural justice on legitimacy, but without accounting for either distributive justice or performance include Fagan and Tyler (2005), Tyler and Fagan (2008), and Gau (2011).

Limited past research also indicates a positive relationship between distributive justice judgments and perceptions about police legitimacy; however, the evidence is not nearly as clear or consistent as that for procedural justice. Tyler and Fagan (2008), Tyler and Huo (2002), Murphy and colleagues (2008), and Hinds and Murphy (2007) show a positive association between distributive justice and police legitimacy. However, Reisig and Lloyd (2009) and Tyler et al. (2007) found no significant effect. No studies of police legitimacy have reported a significant negative effect of distributive justice.

Similarly, past research shows mixed findings for the instrumental model; the influence of performance effectiveness in fighting crime, improving neighborhood conditions, and reducing fear of victimization on police legitimacy is not clear. However, Sunshine and Tyler (2003) (see the pre-crisis sample) and Van der Toorn and colleagues (2011) suggest that while performance effectiveness is less important than procedural justice, it can play a substantive role in forming legitimacy judgments. In particular, Van der Toorn and colleagues (2011) show that perceptions about police performance are nearly as important to their sample as procedural fairness in predicting police legitimacy. (Although Hinds and Murphy (2007) and Murphy and colleagues (2008) also support this association and suggest that in their study in Australia, police performance is nearly as influential as procedural justice on perceptions of police legitimacy, their measure of legitimacy is not well aligned with how it is defined in this encyclopedia entry as it includes components of satisfaction and performance and confidence in police.) Fagan and Tyler (2004), Tyler and Fagan (2008), and Tyler and colleagues (2010) did not show a significant effect of police performance on legitimacy judgments in New York among residents overall, or Muslim-Americans specifically. Hough and colleagues (2010) and Jackson and colleagues (2012) did not find a significant influence of effectiveness on legitimacy in a survey of British residents. No studies of police legitimacy have reported a significant negative effect of performance effectiveness.

Panel studies that have examined views about legitimacy over time suggest that individuals have fairly stable opinions about police legitimacy – their views about police legitimacy at the earlier timeframe are the best predictor of the present views about legitimacy. Often, more recent assessments about the fairness of police procedures also matter (Murphy et al. 2008; Tyler 1990; Tyler and Fagan 2008). In a few of these studies, some of the other models are also supported. Tyler and Fagan (2008) and Murphy and colleagues (2008) show support for performance effectiveness, while Murphy and colleagues also support the role of recent distributive justice assessments. Tyler and Fagan (2008), who conducted a panel study in New York City, reported that legitimacy judgments at time two were influenced by their initial views of police legitimacy at time one, as well as both procedural justice components and distributive justice opinions at time two, but not by police performance. Hinds (2007) conducted a study among Australian youth and found that procedural justice considerations outweighed the other significant predictors including police performance, youth-police relationships, expectations of police, negative contact with police, and fear of property crime.

Among the three models purported to influence police legitimacy, the process-based model is the clear champion. More research is needed to better understand the nature of the relationship between legitimacy and both distributive justice and police performance. In addition to these factors, favorable outcomes and prior positive interactions with police also have positive relationships with legal institution legitimacy (Tyler 1990; Tyler and Huo 2002; Sunshine and Tyler 2003).

Demographic Characteristics and Police Legitimacy

Individuals with different cultural, demographic, and experiential backgrounds often differ in their views about police legitimacy. For example, recent studies by Hough and colleagues (2010) and by Jackson and colleagues (2012) found a significant relationship between London residents’ moral values (a cultural measure of the degree to which the individual viewed certain behaviors as morally wrong) and their sense of duty to obey police. Independent of perceptions of procedural justice or police effectiveness, individuals with a stronger sense of personal morality had a greater sense of obligation to obey police. Other personal characteristics, discussed below, have shown similar independent effects.

Cherney and Murphy (2011) caution, however, that background characteristics may not only be associated with different levels of legitimacy, but in forming legitimacy judgments differently. They assert that when individuals have diminished legitimacy about the law, procedural justice qualities of police may not promote police legitimacy and in fact may have a backfire effect (They base this on their study, reported in Murphy et al. (2009), that found that among ethnic minorities who have reduced legitimacy for the law, procedural justice had a negative relationship on willingness to cooperate with police – an anticipated outcome of legitimacy and often used proxy for legitimacy).

Conversely, Tyler and Huo (2002) have suggested that people form opinions about legitimacy in the same way across different background characteristics. They theorize that even if people vary in the level of legitimacy that they ascribe to police, they tend to follow the same psychological process in coming to that conclusion.

Several studies have found relationships between views about police legitimacy and demographic characteristics. For example, a series of studies have shown a positive link between being Caucasian and higher levels of police legitimacy. Sunshine and Tyler (2003) found that white residents viewed police with more legitimacy than African-American and Hispanic residents in New York. Kirk and Papachristos (2011) found similar results in Chicago. Gau (2011) found that White residents feel a stronger obligation to obey police than Asian individuals. In spite of these differences, Tyler (2001), Tyler and Huo (2002), and Sunshine and Tyler (2003) found evidence that residents of different racial and ethnic backgrounds all rely primarily on procedural justice judgments to inform their legitimacy assessments. In these studies, different ethnic groups relied on similar psychological processes to form legitimacy judgments.

Additionally, older individuals tend to perceive the police with more legitimacy than younger people (Murphy et al. 2008; Papachristos et al. 2009; Tyler 1990; Tyler and Fagan 2008). Yet, Fagan and Tyler (2005) report that the adolescents they sampled relied on procedural justice to form views about legitimacy, as did adults in the many surveys conducted by Tyler across cities in the United States. Individuals of different ages rely on a similar process, even if they grant police different levels of legitimacy.

Past research by Tyler (1988) found that residents in Chicago who had recent experiences with police but with different demographic backgrounds (age, gender, race, income, education, and political ideology) used the same criteria when they assessed the fairness of their experience with police. Also in the New York sample, residents who had citizen-initiated contact with police and residents who had police-initiated contact, both reported that the quality of the treatment they received had a significant influence on their perceptions of police legitimacy, when accounting for other demographic characteristics (Tyler and Fagan 2008). Furthermore, a sample of Muslim-Americans relied on procedural justice judgments when forming opinions about legitimacy, just as has been found in samples that do not distinguish religious affiliation (Tyler et al. 2010). Finally, a sample of law enforcement officers and a sample of soldiers each also relied on the procedural justice model to form their views about legitimacy (Tyler et al. 2007). Thus, based on available research, preliminary indications favor Tyler and Huo (2002), suggesting that while some demographic groups have significantly higher or lower opinions about police legitimacy, different groups form their opinions relying primarily on procedural justice judgments.

Context and Police Legitimacy

Under certain contextual conditions, the strength of the factors predicting legitimacy or what is most important may vary. Prior research addressing a variety of individuals' attitudes toward police (e.g., satisfaction with police, support for police, confidence in police, views about fairness and effectiveness) suggests that neighborhood characteristics such as concentrated disadvantage, community disorder, collective efficacy, and the violent crime rate are important factors affecting individuals' opinions (Bradford and Jackson 2010a; Cao et al. 1996; Dunham and Alpert 1988; and Kirk and Papachristos 2011; Kochel 2012; Reisig and Parks 2000; Sampson and Bartusch 1998; Sun et al. 2004). So, it may be reasonable to expect that perceptions about legitimacy will also be impacted by these contextual factors and the realities and feelings that they generate, and that different antecedents to legitimacy may matter more in one context than another (Hough et al. 2010).

Dunham and Alpert (1988) found that generally, attitudes about police practices were more similar within neighborhoods than across neighborhoods. Kochel and Tankebe (forthcoming) find support that this holds true for police legitimacy. A variety of neighborhood qualities may explain this, including shared experiences with poverty, victimization risk, police contact, and sanctions, which may affect residents' fears about safety, feelings of marginalization or acceptance, desires for police assistance, and other attitudes that may be important in forming legitimacy opinions. Supporting this notion, Kochel (2012) found that neighborhoods with a higher proportion of victims, fewer negative police contacts, as well as fewer extremely wealthy residents had more residents who viewed police with legitimacy. Sun, et al. (2004) reported that police legitimacy is lower in areas of concentrated disadvantage.

Other scholars have looked at the related measure of legal cynicism within a neighborhood context. Sampson and Bartusch (1998) and Kirk and Papachristos (2011) found that people living in neighborhoods with concentrated disadvantage were significantly more likely to express legal cynicism. In Chicago, Kirk and Papachristos (2011) also identified high proportions of youth and residential stability as predictors of legal cynicism.

However, none of these studies examined whether residents of different kinds of neighborhoods or other varied contexts not only hold different views, but also rely on different processes in making legitimacy judgments. Two past studies offer the only indications about this issue. Both studies attempt to determine whether different legitimation processes are used when security is threatened.

Sunshine and Tyler (2003) conducted a survey of New York City residents immediately prior to the terrorist crisis on September 11, 2001, and then shortly after the attacks. They found that under the different security threat circumstances, the relative importance of the three models differed somewhat. The first survey revealed that procedural fairness and to a lesser degree performance effectiveness were the antecedents to police legitimacy. In this survey, the effect of procedural justice was nearly five times stronger than performance. Subsequently, residents who viewed police with more legitimacy also were willing to empower police to undertake a variety of actions, including actions that limit the freedoms of citizens. Shortly after the terrorist attack, a second survey found that the influence of distributive justice on legitimacy was statistically significant and was about half as important as procedural fairness to predict perceptions of legitimacy. Performance effectiveness during this time of crisis was not a significant antecedent of legitimacy; instead, it was directly related to residents' willingness to empower police to perform a variety of actions, some intrusive, including stopping and questioning people on the street and searching homes without permission from a judge. Empowerment was also still predicted by legitimacy, but in this context, performance drove empowerment to action, not perceptions of legitimacy (Similar results are reported in Ghana by Tankebe 2009).

Jonathan-Zamir and Weisburd (2011) studied a community in Israel that was located in close proximity to the Gaza Strip and subject to severe threat for missile attacks compared to other communities in Israel, who lacked similar threats. In both communities, police legitimacy (actually a measure of trust in the police) was positively predicted first by procedural justice and second by performance. The difference was that within non-threatened communities, but not in the community undergoing crisis, negative encounters with police and offenders' risk of getting caught also impacted their legitimacy measure.

Thus, early indications are that the process of forming legitimacy assessments may not drastically vary across social contexts. Procedural justice judgments are extremely important under a variety of situations and appear to be the primary influence on legitimacy. However, more research is needed on the process of forming legitimacy judgments within different contexts.

Conclusions

The important outcomes of police legitimacy substantiate careful consideration of the factors that promote it. Three models dominate theory: the process-based model, the instrumental model, and the distributive justice model. Available research strongly supports the consistent and strong influence of fair procedures in forming positive views about police legitimacy. This model is supported across individuals with a variety of different background characteristics as well as across social contexts, including in developed and developing countries, across a variety of cities in the United States and in places of severe security threats and those lacking such threats. Mixed evidence on the role of police performance effectiveness and distributive justice on legitimacy assessments speak to the need for further research in those areas.

Recommended Reading and References

Bradford B, Jackson J (2010a) Cooperating with the police: social control and the reproduction of police legitimacy. Available at SSRN: http://ssrn.com/abstract=1640958. Retrieved 20 July 2012

Cao L, Frank J, Cullen FT (1996) Race, community context and confidence in the police. Am J Police 15(1):3–22

Cherney A, Murphy K (2011) Understanding the contingency of procedural justice outcomes. Policing 5(3):228–235

Dunham RG, Alpert GP (1988) Neighborhood differences in attitudes toward policing: evidence for a mixed-strategy model of policing in a multi-ethnic setting. Criminology 79(2):504–523

Fagan J, Tyler T (2004) Policing, order maintenance, and legitimacy. In: Mesko G, Pagon M, Dobovsek B (eds) Policing in central and eastern Europe: dilemmas of contemporary criminal justice. Faculty of Criminal Justice, University of Maribor, Ljubljana, Slovenia, pp 38–48

Fagan J, Tyler T (2005) Legal socialization of children and adolescents. Soc Just Res 18(3):217–242

Gau JM (2011) The convergent and discriminant validity of procedural justice and police legitimacy: an empirical test of core theoretical propositions. J Crim Just 39:489–498

Hinds L (2007) Building police-youth relationships: the importance of procedural justice. Youth Justice 7(3):195–209

Hinds L, Murphy K (2007) Public satisfaction with police: using procedural justice to implement police legitimacy. Aust NZ J Crim 40:27–42

Hough M, Jackson J, Bradford B, Myhill A, Quinton P (2010) Procedural justice, trust and institutional legitimacy. Policing 4(3):203–210

Jackson J, Bradford B, Hough M, Myhill A, Quinton P, Tyler TR (2012) Why do people comply with the law? Legitimacy and the Influence of Legal Institutions. Brit J Criminol 9:2012. doi:10.1093/bjc/azs032, Published online: July

Jonathan-Zamir T, Weisburd D (2011) The effects of security threats on antecedents of police legitimacy: findings from a quasi-experiment in Israel. J Res Crime Delinq. doi:10.1177/0022427811418002, Available online DOI

Kirk DS, Papachristos AV (2011) The structural and cultural dynamics of neighborhood violence. U.S. Department of Justice, Washington, DC

Kochel TR (2012) Can police legitimacy promote collective efficacy? Justice Q 29(3):384–419

Kochel TR, Tankebe J (forthcoming) Legitimacy as a neighborhood-level construct

Kochel TR, Parks RB, Mastrofski SM (2011) Examining police effectiveness as a precursor to legitimacy and cooperation with police. Justice Q. doi:10.1080/07418825.2011.633544, Available online November 23, 2011

LaFree GD (1998) Losing legitimacy. Westview Press, Boulder

Leventhal GS (1976) Fairness in social relationships. In: Thibaut J, Spense JT, Carson RC (eds) Contemporary topics in social psychology. General Learning Press, Morristown

Lipset SM (1959) Some social requisites of democracy: economic development and political legitimacy. Am Pol Sci Rev 53(1):69–105

Murphy K, Hinds L, Fleming J (2008) Encouraging public cooperation and support for police. Policing and Society 18(2):136–155

Murphy K, Tyler TR, Curtis A (2009) Nurturing regulatory compliance: is procedural justice effective when people question the legitimacy of the law? Regul Gov 3:1–26

Papachristos AV, Meares TL, Fagan J (2009) Why docriminals obey the law? The influence of legitimacy and social networks on active gun offenders. Yale Law & Economics Research Paper No. 373; Columbia Public Law Research Paper No. 09–199. Available at SSRN: http://ssrn.com/abstract=1326631

Reisig M, Lloyd C (2009) Procedural justice, police legitimacy, and helping the police fight crime: results from a survey of Jamaican adolescents. Police Quart 12:42–62

Reisig M, Parks R (2000) Experience, quality of life, and neighborhood context: a hierarchical analysis of satisfaction with police. Justice Quarterly 7(3):607–630

Reisig MD, Bratton J, Gertz MG (2007) The construct validity and refinement of process-based policing measures. Crim Justice Behav 34:1005–1028

Sampson R, Bartusch D (1998) Legal cynicism and (subcultural?) tolerance of deviance: the neighborhood context of racial differences. Law Soc Rev 32(4):777–804

Sun IY, Triplett RA, Gainey RR (2004) Social disorganization, legitimacy of social institutions and neighborhood crime: an exploratory study of perceptions of the police and local government. J Crim Just 27(1):33–60

Sunshine J, Tyler T (2003) The role of procedural justice and legitimacy in shaping public support for policing. Law Soc Rev 37(3):513–547

Tankebe J (2008) Police effectiveness and police trustworthiness in Ghana: an empirical appraisal. Criminol Crim Just 8:185–202

Tankebe J (2009) Public cooperation with the police in Ghana: does procedural fairness matter. Criminology 47(4):1265–1293

Tyler TR (1988) What is procedural justice? Criteria used by citizens to assess the fairness of legal procedures. Law Soc Rev 22:301–355

Tyler TR (1990) Why people obey the law: procedural justice, legitimacy, and compliance. Yale University Press, New Haven

Tyler TR (1997) The psychology of legitimacy: a relational perspective on voluntary deference to authorities. Pers Soc Psychol Rev 1(4):323–345

Tyler TR (2001) Public trust and confidence in legal authorities: what do majority and minority group members want from law and legal institutions? Behav Sci Law 19:215–235

Tyler TR (2004) Enhancing police legitimacy. Ann Am Acad Pol Soc Sci 593:84–99

Tyler TR (2006) Psychological perspectives on legitimacy and legitimation. Annu Rev Psychol 57:3

Tyler TR, Fagan J (2008) Legitimacy and cooperation: why do people help the police fight crime in their communities? Ohio St J Crim L 6:231–276

Tyler TR, Huo Y (2002) Trust in the law: encouraging public cooperation with the police and courts. Russell-Sage, New York

Tyler TR, Boeckmann RJ, Smith HJ, Huo YJ (1997) Social justice in a diverse society. Westview Press, Boulder

Tyler TR, Callahan PE, Frost J (2007) Armed and dangerous (?): motivating rule adherence among agents of social control. Law Soc Rev 41(2):457–492

Tyler TR, Schulhofer S, Huq AZ (2010) Legitimacy and deterrence effects in counterterrorism policing: a study of Muslim Americans. Law Soc Rev 44(2):365–402

Van der Toorn J, Tyler TR, Jost JT (2011) More than fair: outcome dependence, system justification, and the perceived legitimacy of authority figures. J Exp Psychol 47:127–138

Weber M (1947) Max Weber: the theory of social and economic organization (trans: Henderson AM, Parsons T). New York: The Free Press

Causes of Wrongful Convictions

Shawn Armbrust and Susan Friedman
Mid-Atlantic Innocence Project, American University, Washington, DC, USA

Overview

In any system of justice, there is a risk of convicting innocent people. The United States is no exception, and innocent people have been convicted throughout American history, a phenomenon that was first documented in 1923 (Borchard 1923). In the past two decades, however, more attention has been paid to this problem because, for the first time, post-conviction DNA testing has allowed us to determine with scientific certainty that innocent people have been convicted and often can point to the real perpetrator. The first such exoneration occurred in 1989, and as of 2012, the Innocence Project reports the exoneration of 297 individuals by post-conviction DNA testing. The National Registry of Exonerations identified hundreds more who have been exonerated using nonbiological evidence of innocence. As more of these exonerations occurred, they sparked interest in what can go wrong during an investigation and trial that can lead to such troubling systemic failures. Thus, advocates, academics, and policymakers have invested significant time and resources into identifying isolating the causes of wrongful convictions.

In 2008, in the most thorough study to date, University of Virginia law professor Brandon Garrett exhaustively studied the first 200 DNA exonerations, examining how they were handled from investigation through the trial, appellate, and post-conviction process. Garrett and others have demonstrated that there are six major causes of wrongful convictions: eyewitness testimony, unreliable forensic science evidence, false confessions, incentivized witnesses, government misconduct, and inadequate defense counsel. This entry will summarize how each of these problems can cause wrongful convictions and what can be done to ensure that these problems are addressed so that they do not lead to wrongful convictions in the future.

Eyewitness Identification

To date, eyewitness error is the number one cause of documented wrongful convictions in DNA cases and also is quite prevalent in non-DNA exoneration cases (Garrett 2008; Gross 2005). Professor Garrett found that in 158 of the first 200 DNA exonerations – 79 % – eyewitness error led to the wrongful conviction, and Professor Gross found that 219 of the 340 DNA and non-DNA wrongful convictions between 1989 and 2003 involved eyewitness error. Thus, it is clear that eyewitness fallibility is a significant problem.

The simple reason for this is that human memory for faces is far less accurate than most people – particularly participants in the criminal justice system – have long believed. In addition to being inaccurate, both human memory and confidence about the accuracy of one's memory are quite malleable (Wells et al. 1998). Moreover, even when problems with an eyewitness's memory can be exposed in the courtroom by pointing out discrepancies in the witness's description or problems with their ability to clearly see the perpetrator of a crime, research has shown that if an eyewitness is confident, that is the number one factor considered by jurors and courts in determining the accuracy of an eyewitness (Wells et al. 1998). This is true even though confidence at trial does not have a correlation with accuracy (although confidence at the time of the initial identification has a slightly higher correlation with accuracy) (Wells et al. 1998).

Unfortunately, unintended manipulation of human memory and confidence through the traditional identification process – in which an investigator on the case in question shows the witness several photos or individuals at the same time – exacerbates these problems (Wells et al. 1998). Social psychologists have been studying this issue for decades and have identified several key problems in traditional eyewitness identification procedures that exacerbate the fallibility and malleability of human memory. Those problems include:

- Feedback (often unintentional) to witnesses by law enforcement officers who are implementing a lineup procedure, which can improve a witness's confidence in the accuracy of the identification
- The use of "relative judgment," in which a witness looking at a group of photos or individuals at the same time chooses the individual who looks *most like* the perpetrator, even when the actual perpetrator is not in the lineup
- Repeated identification procedures (photo array, live lineup, preliminary hearing, etc.) that can increase a witness's confidence in the accuracy of an identification (Wells et al. 1998)

The same social scientists have recommended a series of reforms that have been shown to make eyewitness identifications more accurate and capture the confidence of a witness's identification at the time it was made, instead of relying on inflated confidence estimates at the time of trial. Those reforms include:

- Double-blind administration, a procedure in which neither the witness nor the officer administering the photo array or lineup knows who the suspect is
- Confidence statements that would require the witness to immediately make a statement about his or her confidence after making an identification

- Sequential lineups in which photos or individuals are shown to a witness one at a time, requiring the witness to use absolute judgment instead of relative judgment, comparing her memory to one individual in a lineup or one photo at a time
- Videotaping the identification procedures, which will allow defense counsel to identify any potential issues with the identification procedures (Wells et al. 1998)

According to a recent review of reforms that have been implemented (Norris et al. 2010), ten states have taken steps to improve eyewitness identification procedures by either recommending or requiring some or all of these recommendations. Those states are Georgia, Maryland, New Jersey, North Carolina, Ohio, Rhode Island, Vermont, Virginia, West Virginia, and Wisconsin.

Unreliable Forensic Science Evidence

With its purported ability to scientifically link a defendant to a crime scene, forensic science plays a critically important role in criminal trials. For decades, forensic science and forensic scientists were seen as infallible. With the development of DNA testing, however, flaws in many types of forensic science have been exposed. In fact, forensic science evidence is the second most common type of evidence used to secure a wrongful conviction. One hundred thirteen of the first 200 DNA exonerations involved cases in which some kind of forensic science evidence either linked the defendant to the crime scene or placed the defendant within the population of individuals who might have committed the crime (Garrett 2008). In those cases where analysts said that forensic science linked the defendant to the crime, DNA testing proved that this testimony and those sciences were wrong. The most common forensic evidence involved serology (conventional blood typing), followed by hair evidence, soil comparison, DNA tests, bite-mark comparisons, fingerprint evidence, dog scent identification, spectrographic voice evidence, shoe prints, and fiber comparison.

Most forensic science used in courtrooms – both before and after the advent of DNA testing – involves the attempt to determine whether a suspect can be tied to a crime scene. This can happen in several ways, including the following: (1) because a suspect leaves part of himself, such as hair, semen, saliva, sweat, or blood, at a crime scene; (2) because a suspect leaves something else, such as fibers, soil, or footprints, at a crime scene; or (3) because the weapon used to injure the victim has markings that can be compared to an item (usually a gun or knife) owned by the suspect. Oftentimes, forensic scientists testified that a suspect or items in possession of the suspect "matched" evidence found at a crime scene and that this match was conclusive proof that the defendant had, in fact, been at the crime scene. This testimony often went unchallenged by defense lawyers in the era before DNA testing.

However, the explosion of DNA exonerations in cases where at trial other forensic sciences conclusively linked a defendant to a piece of evidence at a crime scene proved that in many cases, what people believed to be infallible science was actually quite fallible and possibly not even science. Thus, in 2009, a committee of the National Academy of Sciences ("NAS") issued a report commissioned by Congress that was designed to evaluate the reliability of forensic science techniques. The report identified several flaws with the way that forensic science has been developed, regulated, and practiced. It expressed particular concern with microscopic hair comparison, fingerprint comparison, gunshot residue analysis, firearm and toolmark analysis, shoe prints, and soil and fiber comparisons. In each of those sciences, the report found that the conclusions reached by experts (that samples from crime scenes could be matched to defendants or items in the possession of defendants) were not grounded in any scientific research and were not validated by any empirical data. In addition, the lack of accreditation and certification among crime laboratories means that there frequently are no standards to determine what constitutes a match – examiners are allowed to determine for themselves, based on their experience, whether items match. Finally, even where there are standards, the lack of proficiency testing means that unqualified examiners are allowed to continue working and testifying without sufficient oversight.

In recent years, states have made some effort to reform crime laboratories and the use of forensic evidence in criminal trials. By statute, 13 states enacted some type of forensic oversight, and in two states, the state's attorney general established oversight entities. In addition to ensuring that crime laboratories are accredited and follow best practices, some oversight commissions are also charged with investigating forensic misconduct. In 2004, Congress passed the Justice for All Act, which included a provision that links the receipt of grants by state crime laboratories under the Paul Coverdell Forensic Science Improvement Grants Program to the establishment of guidelines for handling allegations of negligence or misconduct. In the wake of the NAS report, various advocates also are working to create an entity within the federal government that would foster real scientific research into various areas of forensic science, as well as set standards and licensing requirements that would help create some uniformity within the various forensic science disciplines.

False Confessions

Prior to the advent of DNA testing, it was difficult for most actors in the criminal justice system – including defense lawyers – to believe that an innocent person would confess to a crime that he or she did not commit. Although it was easy to imagine that such confessions would occur in an environment with physical coercion, most American police departments long ago abandoned those tactics and focus instead on persuasive psychological tactics. However, in Professor Garrett's study of the first 200 DNA exonerations, he found that 31 cases involved false confessions. Professors Steve Drizin and Richard Leo documented in a 2004 study published in the University of North Carolina Law Review 125 recent provable false confessions in the United States.

False confessions are particularly problematic because the phenomenon is so hard for most people to understand. Once an individual confesses to a crime, a conviction is almost assured. The police stop investigating, prosecutors ask fewer questions, defense lawyers are more likely to encourage a guilty plea, and juries are unlikely to believe protestations of innocence. This is true regardless of how reliable the confession appears. In many exoneration cases, the confessors were wrong about basic facts like the race of the victim, the location of the crime, and the weapon used, but they were convicted nonetheless. Quite simply, very few people believe that anyone could confess to something they did not do.

Because these false confessions are such a perplexing problem, they have been the subject of significant research by both psychologists and legal scholars who have sought to examine what factors – both individual and systemic – might lead someone to falsely confess to a crime. That research has determined that individuals who are "unusually suggestible and compliant" – such as juveniles and the mentally retarded – are more likely to provide authorities with a false confession (Drizin and Leo 2004). Of the 31 false confessions documented by Professor Garrett, 11 involved mentally retarded defendants, and 12 involved juvenile defendants. Juveniles are likely to falsely confess because they are more likely to make impulsive decisions, engage in risky behavior, and be more susceptible to suggestions. Individuals with mental retardation may be more susceptible to suggestions because they want to please authorities, want others to believe they are competent, often cannot control their impulses, and are willing to accept blame.

The idea that inherently suggestible individuals are more likely to falsely confess to crimes is not terribly surprising; it is far more surprising that a significant number of the false confessions documented by Professor Garrett and by Professors Drizin and Leo involved individuals who are not inherently suggestible. Thus, despite that lack of inherent suggestibility, these individuals made the choice to confess to a crime that they did not commit. Legal scholars and social scientists examining the reasons for this have determined that modern interrogations involve multiple tactics that – particularly when combined – can help cause false confessions. First, the purpose of interrogations is to elicit confessions and to ignore protestations of innocence (Drizin and Leo 2004) ("police are trained to interrogate only those suspects whose guilt they presume or believe they have already established"). Second, recognizing the inherent difficulty of getting anyone – guilty or innocent – to confess to a crime, interrogations are designed to convince the suspect that confessing is the logical choice. To do this, investigators rely upon a technique known as "minimization and maximization" (Drizin and Leo 2004), in which investigators maximize the extent of the evidence against the defendant to intimidate them but then offer them the opportunity to minimize their involvement (e.g., by claiming self-defense in a murder or consent in a rape) and thus avoid harsher criminal penalties. Third, the police are allowed to lie to suspects about the evidence against them (Drizin and Leo 2004). Fourth, during the course of an interrogation, police may – often inadvertently – provide the suspect with information about the crime that can be used to validate a false confession later. And fifth, interrogations often are quite long, which has been shown to increase an individual's likelihood of falsely confessing (Kassin 2010). One study found that in cases where a false confession was made, the interrogation lasted an average of 16.3 h (Kassin 2010). The same study also found that 34 % of interrogations lasted 6–12 h, and 39 % lasted 12–24 h (Kassin 2010).

Solving the problem of false confessions is achievable through reforms that often are beneficial to both law enforcement and to innocent defendants. The most common reform recommended to combat false confessions is the videotaping of all interrogations, from start to finish. Although this cannot prevent false confessions, it can increase the likelihood that false confessions are caught before they lead to conviction. More than 500 jurisdictions now do just that. In addition, some have recommended a reformation of interrogation practices, advocating for investigative interviews of suspects instead of interrogations that are designed only to elicit confessions (Drizin and Leo 2004). Thus, while false confessions are one of the most perplexing causes of wrongful convictions, they also are one of the easiest problems to solve.

Government Informants: Snitches

Investigating crime is difficult for many reasons, and one of those reasons is that relevant witnesses often do not want to cooperate with the police, often because of fear or bias against law enforcement or because it is against their interest. Thus, the government is permitted to incentivize testimony by offering reduced sentences, reduced charges, favorable treatment in prison, or – in some cases – reward money. Such deals can be offered to codefendants, reluctant witnesses, or – in the most dangerous form – jailhouse informants who allegedly heard the defendant confess to the crime in a jail cell. While such deals are a necessary evil, the potential for abuse is obvious.

Wrongful conviction cases prove this point again and again. In 35 of the first 200 DNA exonerations, Professor Garrett found that some type of incentivized testimony was used from an informant, a jailhouse informant, or an alleged co-perpetrator. In 2005, the Center on Wrongful Convictions released a study on the 111 death row exonerations since the death penalty was reinstated in this country in the 1970s. The study found that of those exonerations, 51 individuals were wrongfully convicted based on informant testimony, making incentivized testifying the leading cause of wrongful convictions in capital cases.

The reason for this is that the incentives, combined with the ease of making up seemingly credible information, create a world in which informants are able and willing to lie in ways that are difficult to detect. In the most problematic cases, informants make statements that the defendant has confessed to the informant – either out on the street or in a jail cell. Those statements often contain details that seemingly would be difficult to obtain from anyone other than the true perpetrator of a crime. However, when an informant has access to the Internet, media, friends, and – in some cases – the court file that the defendant has left in his jail cell, he or she can obtain information about the case that is used to create a plausible story that is difficult to either verify or disprove. Indeed, in a 60-min segment, Leslie Vernon White, who admitted to lying on many occasions as a jailhouse informant, demonstrated the ease with which he could obtain information about a crime and make up a confession.

These problems are exacerbated when either the deals provided to an informant or an informant's full history of cooperating with the government is not fully disclosed to the defense. If a jury knows that an informant has received some benefit for his testimony or that he has testified as an informant on multiple occasions, the jury at the very least has the benefit of more balanced information about the informant. In many cases, however, that information is not disclosed, which takes testimony that is inherently suspect and inappropriately bolsters it in the minds of the jury.

Despite its inherent flaws, banning informant testimony is not a tenable solution. Without the ability to incentivize witnesses, the government would have a difficult time convincing witnesses to talk. Thus, any solution to the problem of informants involves both appropriately verifying facts in the informant's testimony and properly disclosing an informant's history and benefits he has received in exchange for his testimony. Recommendations on how to handle incentivized testimony include having informants wear wires, recording informant conversations with authorities and providing the recording to defense counsel, disclosing the benefit that the informant will receive in exchange for testifying, and keeping a database of informants so that law enforcement and the defense know who may be manipulating the system to repeatedly receive benefits from the government. In Illinois, the state legislature requires significant disclosures, pretrial reliability hearings, corroboration, and cautionary jury instructions in any capital case involving a jailhouse informant.

Government Misconduct

In many cases, wrongful convictions are caused by an unfortunate series of honest mistakes – eyewitnesses who have erred, scientific testimony that was unintentionally overstated, and confessions that, while false, were obtained using what police believed to be appropriate tactics. In other cases, however, police and prosecutors who are under pressure to get convictions sometimes lose sight of their obligation to do justice and – whether intentionally or not – commit misconduct by, among other things, failing to turn over exculpatory evidence, improperly characterizing evidence at trial, engaging in improper interrogation tactics, or inappropriately attempting to influence an eyewitness. Such misconduct is far more widespread than anyone might have believed before the advent of DNA testing.

Although it is impossible to quantify exactly how many exonerations involved such misconduct because it is so difficult to uncover, it clearly has been a significant factor in many wrongful conviction cases (Garrett 2008). In 2003, the Center for Public Integrity found that since 1970 prosecutorial misconduct was a factor in dismissed charges, reversed convictions, or reduced sentences in at least 2,012 cases of the 11,452 cases in which claims of prosecutorial misconduct were reviewed by courts (The Justice Project 2009). Unfortunately, studies also have shown that there are few incentives for prosecutors, in particular, to avoid committing misconduct (Ridolfi and Possley 2010). Thus, this is an area that is ripe for reform.

The reforms suggested in the sections on eyewitness error and false confessions also would address government misconduct in these areas, but they would not address the problem of failing to turn over exculpatory evidence. In civil cases, both sides are required to liberally exchange information in lengthy and complicated discovery procedures. In criminal cases, however, discovery is much more limited. The Supreme Court found in *Brady v. Maryland* that the Constitution requires the government to disclose exculpatory and impeachment information to the defense, but the

Constitution requires nothing else. Some states, such as North Carolina, require the government to provide open-file discovery, but that is not the norm. In the federal system and the District of Columbia, for example, criminal defendants are entitled to only what is constitutionally required, their own statements, and expert notice. Unless a defendant has a good investigator, he may not even know who the witnesses against him are until those witnesses testify at trial.

Because so many states have such limited discovery, the onus usually is on the prosecutor to determine what information needs to be disclosed. Prosecutors are trusted to do this because they have two roles in the criminal justice system – to zealously prosecute crimes and to do justice. Unfortunately, the DNA exonerations and other exonerations have shown that prosecutors often fail to disclose information that could have been helpful to the defense. Sometimes, this is deliberate. In other cases, however, prosecutors who are working to zealously prosecute a defendant simply do not see certain information as exculpatory and therefore do not turn it over to the defense.

Various entities looking at this problem have proposed various solutions. Those solutions include open-file discovery and more meaningful sanctions for prosecutors who do break the rules. In addition, the Innocence Project has proposed criminal justice reform commissions that can bring all parties – including police and prosecutors – to the table to come to a consensus on how to handle the actors in the criminal justice system who do not follow the rules.

Inadequate Defense Counsel

The Sixth Amendment of the United States Constitution guarantees criminal defendants the right to counsel, and the Supreme Court in *Gideon v. Wainwright* found that this requires the state to provide indigent defendants with a lawyer who is financed by the state. While *Gideon* was a tremendous milestone for indigent defendants, it did not specify how states were required to make attorneys available to indigent defendants. Thus, states responded to *Gideon* in a variety of ways, including statewide public defender systems, contract systems in which independent agencies bid for court appointments and are awarded contracts based on a variety of factors, contract systems in which contracts go to the lowest bidder, and ad hoc court appointments in which judges appoint lawyers on a case-by-case basis. Predictably, this has resulted in a huge disparity between jurisdictions, with some committing to the provision of high-quality indigent defense lawyers despite the unpopularity of doing so and others providing defendants with inexperienced, poorly paid attorneys who are overburdened with extraordinary caseloads and denied access to investigators and necessary experts.

There is little incentive for states to provide anything other than the bare minimum. Although the Supreme Court found in *Strickland v. Washington* that defendants have a right to effective counsel, a defendant challenging his conviction based upon ineffective assistance of counsel must prove that his attorney was defective and that if his attorney had acted properly, the result of the proceeding would have been different. In any event, challenging a conviction based on ineffective assistance of counsel requires competent counsel, something many defendants simply do not have, particularly after they have been convicted and through the appellate process, when the right to counsel ends.

Unsurprisingly, the lack of quality defense lawyers leads to wrongful convictions. When an innocent defendant lacks an advocate who can investigate and meaningfully challenge the government's case, it is to be expected that a wrongful conviction might occur. According to the Innocence Project, many wrongful convictions in DNA cases involved lawyers who were incompetent, ineffective, overworked, or all three. Among the errors noted in these cases are lawyers who slept in the courtroom during trial, failed to investigate alibis, failed to consult experts or adequately challenge forensic evidence, and failed to show up for hearings.

This problem is difficult to solve largely because it requires the investment of scarce government resources in something that is considered to be unpopular, the representation of accused criminals. However, a commitment of sufficient resources to indigent defense systems would solve most, if not all of the problems that cause wrongful convictions. While it would not directly solve the problem of unethical or lazy attorneys, the proper investment of resources would help decrease caseloads, allow for sufficient investigative funds, provide funding for experts, provide training that is sorely lacking, and help reduce the disparity of resources between indigent defense systems and prosecutors. While such funding might not be popular, most prosecutors and other actors in the criminal justice system recognize that a well-funded indigent defense system helps make the system function properly and can help prevent the conviction of innocent people.

Conclusion

Every wrongful conviction is a tragic example of how the criminal justice system can fail at its most critical function: the reliable conviction of the guilty and exoneration of the innocent. While these cases are tragic, they also provide us with an unprecedented opportunity to learn where our criminal justice system is fallible and how it can be improved. It is known that wrongful convictions have exposed problems with eyewitness testimony, forensic science evidence, false confessions, government informants, government misconduct, and ineffective assistance of counsel. It is important to continue to study wrongful convictions so that best practices can be determined for how to handle these issues. With continued study and reform, steps can be taken to correct problems with these factors in order to prevent the conviction of the innocent in the future.

Recommended Reading and References

Borchard E (1923) Convicting the innocent. Yale University Press, New Haven

Committee on Identifying the Needs of the Forensic Science Community of the National Research Council (2009) Strengthening forensic science in the United States: a path forward. The National Academies Press, Washington, DC

Drizin SA, Leo RA (2004) The problem of false confessions in the post-DNA world. North Carol Law Rev 82:891

Garret B (2011) Convicting the innocent. Harvard Press, Cambridge

Garrett B (2008) Judging innocence. Columbia Law Rev 108:55

Gideon v. Wainwright (1963) 372 U.S. 335

Gross SR et al (2005) Exonerations in the United States 1989 through 2003. J Crim Law Criminol 95:523

InnocenceProject.org

Kassin SM et al (2010) Police-induced confessions: risk factors and recommendations: looking ahead. Law Hum Behav 34:49

Norris R, Bonventre C, Redlich AD, Acker JR (2010/2011) Than that one innocent suffer: evaluating safeguards against wrongful convictions. Albany Law Rev 74:1301

Ridolfi KM, Possley M (2010) Preventable error: a report on prosecutorial misconduct in California 1997–2009

Strickland v. Washington (1984) 466 U.S. 668

The Justice Project (2009) Improving prosecutorial accountability: a policy review. Justice Project, Washington, DC

United States Constitution Amendment VI

Wells G et al (1998) Eyewitness identification procedures: recommendations for lineups and photospreads. Law Hum Behav 22(6):603–647

CCTV

▶ Identification Technologies in Policing and Proof

CCTV and Crime Prevention

Brandon C. Welsh[1,2] and David P. Farrington[3]
[1]School of Criminology and Criminal Justice, Northeastern University, Boston, MA, USA
[2]Netherlands Institute for the Study of Crime and Law Enforcement (NCSR), Amsterdam, The Netherlands
[3]Institute of Criminology, University of Cambridge, Cambridge, UK

Synonyms

Camera surveillance; Closed-circuit television; Video surveillance

Overview

Western countries are experiencing a substantial increase in the use of closed-circuit television (CCTV) surveillance cameras to prevent crime in public places. Amid this expansion and the associated public expenditure, as well as concerns about their effectiveness and social costs, there is an increasing need for an evidence-based approach to inform CCTV policy and practice. This entry reports on an updated Campbell Collaboration systematic review and meta-analysis of the effects of CCTV on crime in public places. The results suggest that CCTV has a modest but significant effect on reducing crime. This overall result was largely driven by the effectiveness of CCTV schemes in car parks, which were targeted at vehicle crimes and included other interventions such as improved lighting, fencing, and security guards. Nonsignificant effects on crime were observed in the other public settings in which CCTV schemes were evaluated: city and town centers, public housing communities, and public transportation facilities.

CCTV has become a widely popular and commonplace technology that is being deployed to help make public (and private) places safer from crime. Across many cities and towns, it is touted as a panacea to crime. But it is not without its controversies, and many questions remain about its ability to reduce crime. This entry reports on the results of a systematic review – incorporating meta-analytic techniques – to assess the scientific evidence on the effectiveness of CCTV to prevent crime in public places.

In recent years, there has been a marked and sustained growth in the use of CCTV surveillance cameras to prevent crime in public places in many Western nations. The United Kingdom in particular is on the cusp of becoming, in the words of some, a "surveillance society," with upward of a million or more cameras in public places (Norris 2007).

There are no national estimates as yet on the number of CCTV cameras in the United States, but local accounts indicate that they are being installed at an unprecedented rate and their popularity is not limited to large urban centers. Some of this increased use in the United States has come about in an effort to aid the police in the detection and prevention of terrorist activities, especially in New York City and other metropolises. However, the prevention of crime remains an important aim of these CCTV systems. Similar claims about the purpose of public CCTV have been made in the United Kingdom. There are also signs that other countries are increasingly experimenting with CCTV to prevent crime in public places. Evaluation studies of public CCTV schemes have been carried out recently in a number of European countries, including Germany, Norway, and Sweden, as well as in Australia, Canada, and Japan. Many of these countries have not previously used CCTV in public places, let alone evaluated its effects on crime.

The growth in CCTV has come with a huge price tag. In the United Kingdom, CCTV has been and continues to be the single most heavily funded crime prevention measure operating outside of the criminal justice system. In the United States, estimates suggest that public expenditure on CCTV may be as much as $100 million each year (Savage 2007).

There has been much debate about the effectiveness of CCTV in preventing crime and hence on the wisdom of spending such large sums of money. A key issue is to what extent funding for CCTV, especially in the United Kingdom and the United States, has been based on high quality scientific evidence demonstrating its effectiveness in preventing crime (Welsh and Farrington 2009a, b).

Methods

The methodology employed in our systematic review follows the conventions set out by the Campbell Collaboration. Evaluations were included in the systematic review if they met a number of criteria, including if CCTV was the main intervention, if there was an outcome measure of crime, and if the evaluation design was of high methodological quality. At a minimum, high quality evaluation designs involve before and after measures of crime in experimental and comparable control areas. Control areas are needed to assess what would have happened in the absence of CCTV and to counter threats to internal validity.

Extensive search strategies were employed to locate studies meeting the criteria for inclusion, including searches of electronic bibliographic databases, searches of literature reviews on the effectiveness of CCTV on crime, and contacts with leading researchers. Altogether, 44 studies were found that met our inclusion criteria.

In order to carry out a meta-analysis, a comparable measure of effect size and an estimate of its variance are needed in each program evaluation (Lipsey and Wilson 2001). In the case of CCTV evaluations, the measure of effect size had to be based on the number of crimes in the experimental and control areas before and after the intervention. This is because this is the only information that was regularly provided in these evaluations.

The "relative effect size" or RES is used to measure effect size. The RES is intuitively meaningful because it indicates the relative change in crimes in the control area compared with the experimental area. For example, RES = 2 indicates that the change in the control areas is twice as large as the experimental areas. This value could be obtained, for example, if crimes doubled in the control area and stayed constant in the experimental area, or if crimes decreased by half in the experimental area and stayed constant in the control area, or in numerous other ways. The RES is computed from the pre- and post-CCTV crime counts in the control and experimental areas. The equation is as follows:

$$\mathrm{RES} = (\mathrm{a} * \mathrm{d})/(\mathrm{b} * \mathrm{c})$$

where a and b are the pre- and post-CCTV crime counts for the experimental areas, respectively, and c and d are the pre- and post-CCTV crime counts for the control areas, respectively.

The meta-analysis was performed on the logged values of RES with the final results converted back into the original scale by taking the antilog. In order to perform the meta-analysis, the variance of each logged RES is needed. If we assume that the crime counts are Poisson distributed, then the variance is 1/a + 1/b + 1/c + 1/d. This assumption is plausible because 30 years of mathematical models of criminal careers have been dominated by the assumption that crimes can be accurately modeled by a Poisson process (Piquero et al. 2007). In a Poisson process, the variance of the number of crimes is the same as the number of crimes (i.e., v = n). However, the large number of changing extraneous factors that influence the number of crimes may cause overdispersion; that is, where the variance of the number of crimes exceeds the number of crimes. Farrington et al. (2007) estimated the overdispersion (D) in crime counts as follows:

$$\mathrm{D} = 0.0008 * \mathrm{N} + 1.2$$

D increased linearly with N and was correlated 0.77 with N. The mean number of crimes in an area in the CCTV studies was about 760, suggesting that the mean value of D was about 2. However, this is an overestimate because the monthly variance is inflated by seasonal variations, which do not apply to N and VAR. Nevertheless, in order to obtain a conservative estimate, V(LRES), which is the variance of LRES (and is distinguished from Va, which is the variance of a), calculated from the usual formula above was multiplied by D (estimated from the above equation) in all cases. Specifically,

$$\mathrm{V(LRES)} = \mathrm{Va}/\mathrm{a}^2 + \mathrm{Vb}/\mathrm{b}^2 + \mathrm{Vc}/\mathrm{c}^2 + \mathrm{Vd}/\mathrm{d}^2$$

where Va/a = 0.0008 * a + 1.2

This is our best available estimate of the degree of overdispersion in area-based crime prevention studies.

Forty-one of the 44 studies could be used in the meta-analysis. RES effect sizes could not be calculated for three studies because numbers of crimes were not reported in the city and town center schemes in Ilford or (for the control area) Sutton or for the public housing scheme in Brooklyn. (Throughout this entry, the included studies are identified by the name of the city or town where they were implemented. Full references can be obtained from the authors' Campbell Collaboration systematic review, which is available at: http://www.campbellcollaboration.org/crime_and_justice/index.php.)

Findings

Setting

City and town centers. Twenty-two evaluations were carried out in city and town centers. Seventeen of these were carried out in the United Kingdom, three in the United States, one in Sweden, and one in Norway. Only some of the studies reported the coverage of the CCTV cameras. For example, in the Newcastle-upon-Tyne and Malmö studies, the camera coverage of the target or experimental area was 100 %. Many more studies reported the number of cameras used and their features (e.g., pan, tilt, zoom). Information on camera coverage is important because if a large enough section of the target area or even high crime locations in the target area are not under surveillance, the impact of CCTV may be reduced.

Most of the evaluations that reported information on the monitoring of the cameras used active monitoring, which means that an operator watched monitors linked to the cameras in real time. Passive monitoring involves watching tape recordings of camera footage at a later time. In some of the schemes, such as Newcastle and Birmingham, police carried out active monitoring. But more often it was carried out by security personnel who had some form of communication link with police (e.g., by a one-way radio or direct line telephone).

On average, the follow-up period in the 22 evaluations was 15 months, ranging from a low of 3 months to a high of 60 months. Six programs included other interventions in addition to the main intervention of CCTV. For example, in the Doncaster program, 47 "help-points" were established within the target area to aid the public in contacting the main CCTV control room. Four other studies used notices of CCTV to inform the public that they were under surveillance, but CCTV notices do not necessarily constitute a secondary intervention. A couple of the evaluations used multiple experimental areas (e.g., police beats), meaning that the CCTV intervention was quite extensive in the city or town center. Multiple control areas (e.g., adjacent police beats, the remainder of the city) were used in many more of the evaluations. The most comparable control areas were chosen for analysis. Where control and adjacent areas were used, control areas were analyzed.

The city and town center CCTV evaluations showed mixed results in their effectiveness in reducing crime. Ten of the 22 evaluations were considered to have a desirable effect on crime, 5 were considered to have an undesirable effect, and 1, the multisite British evaluation by Sivarajasingam et al. (2003), was considered to have both (desirable effects according to emergency department admissions and undesirable effects according to police records). The remaining six evaluations were considered to have a null ($n = 5$) or uncertain ($n = 1$) effect on crime. Schemes usually showed evidence of no crime displacement rather than displacement or diffusion of benefits.

As an example, in the program evaluated by Armitage et al. (1999), an unknown number of cameras were installed in the town center of Burnley, England. The experimental area consisted of police beats in the town center with CCTV coverage. Two control areas were used. The first comprised those police beats that shared a common boundary with the beats covered by CCTV. The second control area consisted of other police beats in the police division. The first control area was more comparable to the experimental area and was the one used in our analysis. After 12 months, the experimental area, compared with the two control areas, showed substantial reductions in violent crime, burglary, vehicle crime, and total crime. For example, total incidents of crime fell by 28 % in the experimental area compared with a slight decline of 1 % in the first control area and an increase of 10 % in the second control area. The authors found evidence of diffusion of benefits for the categories of total crime, violent crime, and vehicle crime, and evidence of territorial displacement for burglary.

In pooling the data from the 20 studies for which effect sizes could be calculated, there was evidence that CCTV led to a small but nonsignificant reduction in crime in city and town centers. The weighted mean effect size was a RES of 1.08 (CI 0.97–1.20), which corresponds to a 7 % reduction in crimes in experimental areas compared with control areas. However, when these 20 studies were disaggregated by country, the 15 British studies showed a slightly larger effect on crime (a 10 % decrease; RES = 1.11, CI 0.98–1.27, ns), while the five others showed no effect on crime.

An analysis of heterogeneity showed that the 20 effect sizes were significantly heterogeneous ($Q = 143.9$, $df = 19$, $p < 0.0001$). This means that the differences were not simply a matter of sampling error. The 15 UK studies were also significantly heterogeneous ($Q = 118.6$, $df = 14$, $p < 0.0001$), as were the five other studies ($Q = 14.02$, $df = 4$, $p = 0.007$). Therefore, random effects models that assume study-level variability were used in calculating weighted mean effect sizes.

Public housing. Nine evaluations were carried out in public housing. Seven were carried out in the United Kingdom and two in the United States. Camera coverage ranged from a low of 9 % (in Dual Estate) to a high of 87 % (in Northern Estate) in the six evaluations that reported this information. Active monitoring was used in all of the schemes, with monitoring in the Brooklyn evaluation conducted by police. In the six British schemes evaluated by Gill and Spriggs (2005), security personnel who monitored the cameras had some form of communication link with police (e.g., a one-way or two-way radio). On average, the follow-up period in the nine evaluations was 12 months, ranging from a low of 3 months to a high of 18 months. Only three schemes included other interventions in addition to the main intervention of CCTV. These involved improved lighting and youth inclusion projects.

The public housing CCTV evaluations showed mixed results in their effectiveness in reducing crime. Three of the nine evaluations were considered to have a desirable effect on crime, two had an undesirable effect, three had an uncertain effect, and one had a null effect. Only five schemes measured diffusion of benefits or crime displacement, and in each case it was reported that displacement and diffusion did not occur.

In pooling the data from the eight studies for which effect sizes could be calculated, there was evidence that CCTV led to a small but nonsignificant reduction in crime in public housing. The weighted mean effect size was a RES of 1.07 (CI 0.83–1.39), which corresponds to a 7 % reduction in crimes in experimental areas compared with control areas. The eight effect sizes were significantly heterogeneous ($Q = 47.94$, $df = 7$, $p < 0.0001$).

The evaluation by Williamson and McLafferty (2000) in Brooklyn, New York, the only one that could not be included in the meta-analysis, is somewhat representative of CCTV's rather negligible effect on crime in public housing. The housing community that received the intervention (Albany project) did not show any change in the total number of police-recorded crimes, either inside the project or inside a 0.1 mile buffer zone (established to measure crime displacement or diffusion of benefits), while total crime in the control community (Roosevelt project) dropped by 5 % inside the project and 4 % inside the 0.1 mile buffer zone. When total crime was disaggregated, a desirable program effect was observed for major felonies in both experimental and control projects. However, the authors note that, "the substantial decrease in major felonies around both public housing projects seems to be part of a larger downward trend that was occurring not only in Brooklyn but across New York City in the late 1990s" (Williamson and McLafferty 2000, p. 7). Furthermore, the authors' investigation of the occurrence of crime displacement or diffusion of benefits concluded that there was "no clear evidence" of either, "as the change in crime around the two housing projects does not vary predictably with distance" (p. 7).

Public transport. Four evaluations were carried out in public transportation systems. All of them were conducted in underground railway systems: three in the London Underground and one in the Montreal Metro. None of the studies reported on the percentage of the target areas covered by the cameras, but most did provide information on the number of cameras used. For example, in the Montreal program a total of 130 cameras (approximately 10 per station) were installed in the experimental stations. Each of the schemes involved active monitoring on the part of police; in the London Underground, this meant the British Transport Police.

With the exception of the Montreal program, each evaluation included other interventions in addition to CCTV. In the first Underground scheme, special police patrols were in operation prior to the installation of CCTV. For the two other Underground schemes, some of the other interventions included passenger alarms, kiosks to monitor CCTV, and mirrors. For each of these three Underground schemes, CCTV was, however, the main intervention. The follow-up periods ranged from a low of 12 months to a high of 32 months.

Overall, CCTV programs in public transportation systems present conflicting evidence of effectiveness: two had a desirable effect, one had no effect, and one had an undesirable effect on crime. However, for the two effective programs in the London Underground (southern sector and northern line), the use of other interventions makes it difficult to say with certainty that it was CCTV that caused the observed crime reductions, although in the first of these programs CCTV was more than likely the cause. Only two of the studies measured diffusion of benefits or crime displacement, with one showing evidence of diffusion and the other showing evidence of displacement.

In pooling the data from the four studies, there was evidence that CCTV led to a sizable but nonsignificant reduction in crime in public transport. The weighted mean effect size was a RES of 1.30 (CI 0.87–1.94), which corresponds to a 23 % reduction in crimes in experimental areas compared with control areas. The substantial reduction in robberies and thefts in the first Underground evaluation (an overall 61 % decrease) was the main reason for this large average effect size over all four studies. The four effect sizes were significantly heterogeneous ($Q = 30.94$, $df = 3$, $p < 0.0001$).

Car parks. Six CCTV evaluations met the criteria for inclusion and were conducted in car parks (parking lots). All of the programs were implemented in the United Kingdom between the early 1980s and early 2000s. Camera coverage was near 100 % in the two schemes that reported on it. All of the schemes, with the exception of one that did not provide data, involved active monitoring on the part of security staff. The large-scale, multisite Hawkeye scheme evaluated by Gill and Spriggs (2005) also included a radio link with the British Transport Police.

Each of the programs supplemented CCTV with other interventions, such as improved lighting, painting, fencing, payment schemes, and security personnel. In Coventry, for example, improved lighting, painting, and fencing were part of the package of measures implemented to reduce vehicle crimes. In each program, however, CCTV was the main intervention. The follow-up periods ranged from a low of 10 months to a high of 24 months.

Five of the car park programs had a desirable effect and one had an undesirable effect on crime, with vehicle crimes being the exclusive focus of five of these evaluations. For example, Tilley (1993) evaluated three CCTV programs in car parks in the following UK cities: Hartlepool, Bradford, and Coventry. Each scheme was part of the British Government's Safer Cities Programme, a large-scale crime prevention initiative that operated from the late 1980s to mid-1990s. In Hartlepool, CCTV cameras were installed in a number of covered car parks and the control area included a number of non-CCTV-covered car parks. Security personnel, notices of CCTV, and payment schemes were also part of the package of measures employed to reduce vehicle crimes. Twenty-four months after the program began thefts of and from vehicles had been substantially reduced in the experimental car parks compared with the control car parks. A 59 % reduction in thefts of vehicles was observed in the experimental car parks compared with a 16 % reduction in the control car parks. Tilley (1993, p. 9) concluded that, “The marked relative advantage of CCTV covered parks in relation to theft of cars clearly declines over time and there are signs that the underlying local trends [an increase in car thefts] begin to be resumed.” The author suggested that the displacement of vehicle thefts from covered to non-covered car parks occurred. Most studies did not measure either diffusion of benefits or crime displacement.

The RESs showed a significant and desirable effect of CCTV for five of the schemes. In the other scheme (Guildford), the effect was undesirable, but the small number of crimes measured in the before and after periods meant that the RES was not significant. When all six effect sizes were combined, the overall RES was 2.03 (CI 1.39–2.96, $p = 0.0003$), meaning that crime decreased by half (51 %) in experimental areas compared with control areas. This indicates a very large and highly significant desirable effect of CCTV on vehicle crimes in car parks. The six effect sizes were significantly heterogeneous ($Q = 31.93$, $df = 5$, $p < 0.0001$).

Other settings. Three of the 44 evaluations took place in other public settings: two in residential areas and one in a hospital. It was considered necessary to categorize these three schemes separately from the others because of the differences in the settings in which these three schemes were implemented as well as their small numbers.

There were some notable differences between the two residential schemes. The City Outskirts scheme was implemented in an economically depressed area on the outskirts of a Midlands city, while the Borough scheme was implemented throughout a southern borough of mixed affluence. Camera coverage was quite good in City Outskirts (68 %), but in Borough it was considered low. Gill and Spriggs (2005) noted that this was due in large measure to the use of redeployable cameras in Borough, while fixed cameras were used in City Outskirts. Other interventions were used in City Outskirts, but not in Borough. Evaluations of the two schemes also found contrasting effects on crime: a significant desirable effect in City Outskirts (a 25 % decrease) and a nearly significant undesirable effect in Borough (a 25 % increase).

The one evaluation of CCTV implemented in a city hospital showed that it produced a desirable but nonsignificant effect on crime (RES = 1.38, CI 0.80–2.40), corresponding to a 28 % decrease in crime in the experimental area compared with the control area. Among some of the scheme’s distinguishing features, camera coverage was high (76 %), active monitoring was used, there was a direct line between the camera operators and police, and other interventions were implemented, including improved lighting and police operations.

Crime Type

The major crime types that were reported were violence (including robbery) and vehicle crimes (including thefts of and from vehicles). Violence was reported in 23 evaluations, but CCTV had a desirable effect in reducing violence in only three cases (Airdrie, Malmö, and Shire Town). Overall, there was no effect of CCTV on violence (RES = 1.03, CI 0.96–1.10, ns). The 23 effect sizes were not significantly heterogeneous ($Q = 30.87$, $df = 22$, n.s.).

Vehicle crimes were reported in 22 evaluations, and CCTV had a desirable effect in reducing them in ten cases: in five of the six car park evaluations (all except Guildford), in three city or town center evaluations (Burnley, Gillingham, and South City), and in City Outskirts and City Hospital. Over all 22 evaluations, CCTV reduced vehicle crimes by 26 % (RES = 1.35, CI 1.10–1.66, $p = 0.004$). The 22 effect sizes were significantly heterogeneous ($Q = 115.1$, $df = 21$, $p < 0.0001$). The greatest effect was in the large-scale, multisite Hawkeye study, but there was a significant effect even if this study was excluded (RES = 1.28, corresponding to a 22 % decrease in crimes).

Country Comparison

Of the 41 evaluations that were included in the meta-analysis, the overwhelming majority of them were carried out in the United Kingdom ($n = 34$). Four were from the United States and one each from Canada, Norway, and Sweden. When the pooled meta-analysis results were disaggregated by country, there was evidence that the use of CCTV to prevent crime was more effective in the United Kingdom than in other countries. In the British studies, CCTV had a significant desirable effect, with an overall 19 % reduction in crime (RES $=$ 1.24, CI 1.10–1.39, $p = 0.0005$). The British studies were significantly heterogeneous ($Q = 350.5$, $df = 33$, $p < 0.0001$). In the other studies, CCTV showed no desirable effect on crime (RES = 0.97, CI 0.86–1.09, ns). The other studies were also significantly heterogeneous ($Q = 14.51$, $df = 6$, $p = 0.024$). Importantly, the significant results for the British studies were largely driven by the effective programs in car parks.

Key Issues/Controversies

Von Hirsch (2000) argues that there are two major issues that confront the "proper uses and limits" of surveillance for crime prevention in public places. The first issue pertains to privacy concerns. This can be expanded to include other social costs that may infringe on public interests or violate legal or constitutional protections. The second issue concerns the matter of the "legitimising role of crime prevention," or as von Hirsch (2000, p. 61) posits, "To what extent does crime prevention legitimise impinging on any interests of privacy or anonymity in public space?"

In car parks, there may be little resistance to the installation of CCTV cameras. In part, this is because the public space is utilized for one rather inconsequential purpose – parking vehicles. It is also the case that a car park is a well-defined and clearly marked physical space, meaning that individuals know that it is a car park and can choose to park their vehicle there or not (providing there are other alternatives). These points stand in sharp contrast to how individuals come into contact with CCTV in other public settings.

CCTV in other public settings such as city and town centers, public housing communities, and transportation facilities evokes more resistance on the basis of threats to privacy and other civil liberties, and is associated with a larger number of social harms, including the reinforcement of the notion of a fortress society and the social exclusion of marginalized populations (Clarke 2000). Indeed, it is often these settings that are at the center of the debate over how best to strike a balance between the potential crime reduction benefits and social costs associated with CCTV.

In these other public settings, CCTV may also result in the social exclusion of vulnerable or marginalized populations such as unemployed youths and the homeless. The fear is that, instead of providing assistance for these groups to get off the street so to speak, CCTV, among other interventions like police and security guards, may push them further away from available services and cause increased harm in the form of crime, victimization, or both. Efforts to reduce any social exclusion effect associated with CCTV in these settings would need to involve other services.

Future Directions

Advancing knowledge about the crime prevention benefits of CCTV schemes should begin with attention to the methodological rigor of the evaluation designs. The use of a reasonably comparable control group by all of the 44 included evaluations went some way toward ruling out some of the major threats to internal validity, such as selection, maturation, history, and instrumentation. The effect of CCTV on crime can also be investigated after controlling (e.g., in a regression equation) not only for prior crime but also for other community-level factors that influence crime, such as neighborhood poverty and poor housing. Another possible research design is to match two areas and then to choose one at random to be the experimental area. Of course, several pairs of areas would be better than only one pair.

Also important is attention to methodological problems or to changes in programs that take place during and after implementation. Some of these implementation issues include: statistical conclusion validity (adequacy of statistical analyses); construct validity (fidelity); and statistical power (to detect change). For some of the included evaluations, small numbers of crimes made it difficult to determine whether or not the program had an effect on crime. It is essential to carry out statistical power analyses before embarking on evaluation studies. Few studies attempted to control for regression to the mean, which happens if an intervention is implemented just after an unusually high crime rate period. A long time series of observations is needed to investigate this. The contamination of control areas (i.e., by the CCTV intervention) was another, albeit less common, problem that faced the evaluations.

There is also the need for longer follow-up periods to see how far the effects persist. Of the 44 included schemes, many were in operation for 12 months or less prior to being evaluated. This is a very short time to assess a program's impact on crime or any other outcome measure, and for these programs the question can be asked: Was the intervention in place long enough to provide an accurate estimate of its observed effects on crime? Ideally, time series designs are needed with a long series of crime rates in experimental and control conditions before and after the introduction of CCTV. In the situational crime prevention literature, brief follow-up periods are the norm, but "it is now recognized that more information is needed about the longer-term effects of situational prevention" (Clarke 2001, p. 29). Ideally, the same time periods should be used in before and after measures of crime.

Research is also needed to help identify the active ingredients of effective CCTV programs and the causal mechanisms linking CCTV to reductions in crime. Forty-three percent (19 out of 44) of the included programs involved interventions in addition to CCTV, and this makes it difficult to isolate the independent effects of the different components, including the unique effect of CCTV, and the interaction effects of CCTV in combination with other measures. Future experiments are needed that attempt to disentangle elements of effective programs. Also, future experiments need to measure the intensity of the CCTV dose and the dose-response relationship, and need to include alternative methods of measuring crime (surveys as well as police records).

In order to investigate displacement of crime and diffusion of crime prevention benefits, the minimum design should involve one experimental area, one adjacent area, and one nonadjacent comparable control area. If crime decreased in the experimental area, increased in the adjacent area, and stayed constant in the control area, this might be evidence of displacement. If crime decreased in the experimental and adjacent areas and stayed constant or increased in the control area, this might be evidence of diffusion of benefits. Unfortunately, few CCTV studies used this minimum design. Instead, most had an adjacent control area and the remainder of the city as another (noncomparable) control area. Because of this, any conclusions about displacement or diffusion effects of CCTV seem premature at this point in time.

This systematic review focused on CCTV's effects on crime, but it would also be desirable to investigate effects on detection, prosecution, and conviction. Exactly what the optimal circumstances are for effective use of CCTV schemes is not entirely clear at present, and this needs to be established by future evaluation research. But it is important to note that the success of the CCTV schemes in car parks was mostly limited to a reduction in vehicle crimes (the only crime type measured in five of the six schemes) and camera coverage was high for those evaluations that reported on it. In the national British evaluation of the effectiveness of CCTV, Farrington et al. (2007) found that effectiveness was significantly correlated with the degree of coverage of the CCTV cameras, which was greatest in car parks. Furthermore, all six car park schemes included other interventions, such as improved lighting and security guards. It is plausible to suggest that CCTV schemes with high coverage and other interventions and targeted on vehicle crimes are effective. Conversely, the evaluations of CCTV schemes in city and town centers and public housing measured a much larger range of crime types and only a small number of studies involved other interventions. These CCTV schemes, as well as those focused on public transport, did not have a significant effect on crime.

Crime policy is rarely formulated on the basis of hard evidence, and the use of CCTV is no exception. Campbell Collaboration systematic reviews, like the one presented here, provide the most rigorous source for scientific evidence on the leading criminological interventions. It is high time that the evidence is put at center stage in political and policy decisions about preventing crime.

Acknowledgments We wish to thank David Wilson for helpful comments on an earlier draft of this entry.

Related Entries

▶ Situational Crime Prevention

Recommended Reading and References

Armitage R, Smyth G, Pease K (1999) Burnley CCTV evaluation. In: Painter K, Tilley N (eds) Surveillance of public space: CCTV, street lighting and crime prevention, vol 10, Crime prevention studies. Criminal Justice Press, Monsey

Clarke RV (2000) Situational prevention, criminology, and social values. In: von Hirsch A, Garland D, Wakefield A (eds) Ethical and social perspectives on situational crime prevention. Hart, Oxford, UK

Clarke RV (2001) Effective crime prevention: keeping pace with new developments. Forum Crime Soc 1(1):17–33

Farrington DP, Gill M, Waples SJ, Argomaniz J (2007) The effects of closed-circuit television on crime: meta-analysis of an English national quasi-experimental multi-site evaluation. J Exp Criminol 3:21–38

Gill M, Spriggs A (2005) Assessing the impact of CCTV, Home Office Research Study No. 292. Home Office, London

Hier SP (2010) Panoptic dreams: streetscape and video surveillance in Canada. University of British Columbia Press, Vancouver

Lipsey MW, Wilson DB (2001) Practical meta-analysis. Sage, Thousand Oaks

Norris C (2007) The intensification and bifurcation of surveillance in British criminal justice policy. Eur J Crim Policy Res 13:139–158

Piquero AR, Farrington DP, Blumstein A (2007) Key issues in criminal career research: new analyses of the Cambridge study in delinquent development. Cambridge University Press, New York

Savage C (2007) US doles out millions for street cameras: local efforts raise privacy concerns. Boston Globe, 12 Aug 2007. www.boston.com

Sivarajasingam V, Shepherd JP, Matthews K (2003) Effect of urban closed circuit television on assault injury and violence detection. Inj Prev 9:312–316

Tilley N (1993) Understanding car parks, crime and CCTV: evaluation lessons from safer cities, Crime prevention unit series paper, No. 42. Home Office, London

von Hirsch A (2000) The ethics of public television surveillance. In: von Hirsch A, Garland D, Wakefield A (eds) Ethical and social perspectives on situational crime prevention. Hart, Oxford, UK

Welsh BC, Farrington DP (2009a) Making public places safer: surveillance and crime prevention. Oxford University Press, New York

Welsh BC, Farrington DP (2009b) Public area CCTV and crime prevention: an updated systematic review and meta-analysis. Justice Q 26:716–745

Williamson D, McLafferty S (2000) The effects of CCTV on crime in public housing: an application of GIS and spatial statistics. Paper presented at the annual meeting of the American Society of Criminology, San Francisco

Centre for Children Committing Offences (CCCO)

▶ Stop Now and Plan (SNAP$^{®}$) Model

Certainty, Severity, and Their Deterrent Effects

Horst Entorf
Department of Economics and Business Administration, Goethe University Frankfurt, Frankfurt, Germany

Overview

Severity of punishment without the credible threat of being detected and convicted "… *is the sound of one hand clapping*" (original quote: "Mediation without the credible threat of judicial determination is the sound of one hand clapping," Genn 2010, p. 125). Theoretical and empirical analyses of general deterrence need to consider both certainty *and* severity of sanctions, that is, the intertwining activities of police, public prosecution, courts, and prison conditions. Many studies only rely on the probability of detection, that is, on police activity or efficiency, when discussing the role of *expected* sanctions. Other studies, in particular articles dealing with US data and thus motivated by the highly persistent upward trend in prisoner population, focus on the imprisonment rate and the severity of sanctions as crucial factors of deterrence (see the survey by Donohue 2009). The (brief) survey at hand summarizes the classical rational-choice fundamentals of modern deterrence theory and covers major theoretical and empirical findings on the interplay of certainty and severity of punishment as well as important underlying methodological problems since the early 1970s of the last century.

Any survey of the economics of crime would be incomplete without reminiscence of recent developments in behavioral economics of crime. The last few years seem to have witnessed a change in mainstream economics of crime. Rational-choice models are often criticized because they ignore that cognitive restrictions and emotional factors such as time pressure, peer group influence, or anger restrict the long-run *optimality* of individual decisions. Simon (1957) was the first to point out that the complexity of situations and limitations of both available information and cognitive capacity would lead to decisions under *bounded rationality*. Several recent papers focus on shortcomings and necessary extensions of the classical notion of certainty and severity of sanctions. This contribution discusses significant insights stemming from behavioral economics such as impulsiveness (myopia), prospect theory, and anger. Moreover, stigma effects (dynamic deterrence) and the distinction between objective and perceived threats of potential punishment will be addressed.

This survey contains the following subchapters: fundamentals of general deterrence theory, theoretical and empirical results concerning (a) the certainty of punishment and (b) the severity of sentences, limitations and extensions of the classical deterrence model, and, finally, the threat of punishment when trust in criminal law is absent. The entry concludes with remarks on future directions of research in deterrence.

Fundamentals of General Deterrence Theory

General deterrence is the avoidance of crime by (deterrable) potential offenders through the credible threat of punishment. This definition entails (a) there is a threat in terms of a judicial system of appropriate sanctions and (b) this threat is credible, that is, there is a legal enforcement system of police, prosecutors, and judges which has the capability to realize a perceptibly high (strictly positive) probability of detection and conviction. The definition also requires that there is a nonempty group of compliers, that is, a deterrable subgroup of the population which can be correctly described as *potential offenders* because otherwise the theory of deterrence would be futile. The identification of *deterrable* and *non-deterrable* subpopulations is important but a rather neglected field in criminology.

The theoretical foundations of general deterrence are usually ascribed to Gary Becker's (1968) seminal article, although the philosophical foundation of deterrence dates back to eighteenth century (Bentham (1781)). In a broader sense, the idea of general deterrence is related to rational-choice theory, which assumes that all individuals, irrespective of being criminal or not, respond to incentives, or as Becker puts it, "*Some people become 'criminals' not because their basic motivation differs from that of other persons, but because their benefits and costs differ*" (Becker 1968, p. 176). Though often disputed and criticized (see, among many others, McAdams and Ulen 2009), the importance of rational choice and general deterrence has also been acknowledged by criminologists and sociologists (as discussed in Rupp 2008, p. 6).

Becker's theory on the supply of offenses is based on the comparison of (uncertain) expected utility from criminal activities to the (relatively certain) utility from not committing crimes. Thus, increasing the individual *costs* of crime by increasing the expected sentence would lead to a reduction of criminal behavior for rational offenders. Assuming additivity, the theory predicts that if the expected utility from committing a crime,

$$E_C = (1 - p)U_C + p(U_C - C_C), \qquad (1)$$

exceeds the expected utility from obeying the societal legal norms, E_{NC}, a crime will be committed; otherwise individuals refrain from wrongdoing. Here p denotes the probability that the illegal act will be detected and punished, $(1 - p)$ is the probability of getting away with it, U_C represents utility from crime, E_{NC} is utility from noncrime, and C_C captures disutility (cost) from being punished. Obviously, to make punishment effective, C_C needs to be higher than U_C. This condition fits Bentham's (1781) "Principles of Morals and Legislation": According to Rule 1 of "Of the Proportion Between Punishments and Offences," "*The value of the punishment must not be less in any case than what is sufficient to outweigh that of the profit of the offence*" (Bentham 1781, p. 141).

Equation 1 can be expressed as the difference between the benefit from committing a crime and the expected sentence in case of detection, that is,

$$E_C = U_C - pC_C. \qquad (2)$$

Equation 2 can be considered as the most parsimonious representation of severity and certainty of conviction. It is obvious that credibility of the threat of sanctions requires the expected sentence, that is, product pC_C, to be positive. As an example of the deterrence rationale, consider the choice between paying a certain amount of money, say, 3 Euros, for a short-term city parking space and not paying the amount and taking the risk of a fine of, say, 20 Euros. If on average every 10th *free parking* is detected, that is, $p = 0.1$, then the utility from the illegal action, U_C, would be 3 Euros, whereas the expected fine was 2 Euros. Hence the risk-neutral rational "*offender*" would decide against putting a coin into the parking meter and in favor of the illegal alternative. Of course, such reasoning may be ambiguous in case of risk preference, and as mentioned above, it requires that a nonempty subgroup is complying according to rational-choice rules, that is, a significant number of people belong neither to the group of always strictly norm-abiding citizens nor to the group of never law-abiding individuals. Thus, summarizing the standard rational-choice reasoning covered in Eq. 2, deterrence would not work either when the certainty of a sanction is zero or when sanctions lack severity. In both cases, the product pC_C would be zero, that is, there would be no credible threat from expected sanctions for wrong doing.

Equation 2 also represents the basic foundation of Becker's so-called supply of offenses, which is grounded on first-order conditions of the difference between illegal and legal sources of expected utility, $E_C - E_{NC}$. Taking derivatives with respect to the crucial impacting factors p, C_C and E_{NC} yield the behavioral equation

$$C = C(\underset{(-)}{p}, \underset{(-)}{C_C}, \underset{(-)}{E_{NC}}, X). \qquad (3)$$

Thus, under standard neoclassical assumptions (see Becker 1968 for details), crime (C) would fall when the probability of detection and punishment increases or when potential offenders face the risk of higher costs from committing crimes. These disincentives to crime change in response to longer or tougher prison sentences, for example, in terms of tough-on-crime policies such as the Californian *three-strikes* law or the application of adult criminal law instead of juvenile criminal law (as discussed in Entorf 2012). Crime would also decrease when utility from noncriminal activities, E_{NC}, increases, for instance, because of lower unemployment risk, or improving current and future legal income opportunities, which include the broad and lasting impact of education on crime. Of course, empirical models on crime would fail if only p, C_C, and E_{NC} were included as explanatory factors of crime. Thus, X represents further factors of which age, gender, alcohol and drug addiction, migration background, race, peer influence, and family background belong to the (incomplete) list of variables often (but not regularly) found in crime studies. An indeterminacy of statistical tests of deterrence models is the rather arbitrary choice of X. Indeed, a meta-study by Dölling et al. (2009, Table 5) reveals that the choice of control variables has a high influence on finding significance or nonsignificance of deterrence indicators in empirical studies: Dependent on the choice of the control variable, reported average t-values on deterrence variables range from −0.5 to −2.3. The problem of ad hoc specifications is discussed in more detail by Durlauf and Nagin (2010).

On the Impact of Certainty

Early Contributions. The first empirical tests of Becker's *supply of offenses* model have been conducted in the late 1960s and early 1970s of the last century. Early studies such as Tittle (1969) have found significant crime reducing effects in response to increasing certainty (*p*). However, these contributions were flawed by endogeneity problems and ignorance of incapacitation effects. Although follow-up papers such as Ehrlich (1973) used two-stage least squares techniques to account for simultaneity problems, identifications strategies such as use of lagged endogenous variables would be considered problematic from the current viewpoint of modern econometrics (see also Durlauf and Nagin 2010). Ehrlich's (1973) results were intensively discussed, and not all studies agreed on the evidence in favor of the deterrence hypothesis. Many studies argue that the model is too simplistic and empirical findings are not reliable enough to draw *any* conclusion in favor or against the deterrence hypothesis. According to Rupp (2008), this criticism seems to be exaggerated, but the ambiguity of Ehrlich's findings was and still is perhaps the main reason for the lasting debate on Ehrlich's problematic contribution to the deterrence literature.

Ehrlich's (1973) article also received prominent attention as a significant contribution to the theoretical literature. He modified Becker's theoretical model by using a time allocation model. This framework widens the perspective because it includes leisure time as a source of utility besides utility from time spent on legal and illegal activities. This makes theoretical predictions less clear than in Becker (1968), where utility is achieved either from illegal or from legal activities. Ehrlich (1973, p. 530, footnote 13) draws attention to the point that the unambiguous negative sign of deterrence effects only holds when individuals are risk neutral or risk averse.

Survey of Surveys and Meta-studies. In the first three decades after Becker's (1968) seminal article, several surveys on the economics of crime have been published. Rupp (2008) summarizes them in more detail. During the last few years, the frequency and number of surveys seems to accelerate. Surveys by Rupp (2008), Donohue (2009), Durlauf and Nagin (2010), and Ritchie (2011), to name only a selection of remarkable new publications, confirm what has been found by earlier studies, that is, that the deterrent effect of the certainty of sanctions far outweighs the severity of punishment.

Eide et al. (1994) is a good starting point in order to provide a sample of typical estimates of the deterrent effect of the certainty of sanctions. The authors summarize 20 international cross-sectional studies based on a variety of model specifications, types of data, and regressions. They find the median value of the 118 elasticity estimates of crime rates with respect to various measures of the probability of punishment to be about −0.7. The median of the somewhat fewer severity elasticities is about −0.4.

Donohue (2009) focuses on the effect of the imprisonment rate in six studies based on aggregate data and finds that most studies show a negative effect of incarceration rates on crime. However, the estimates of the elasticity of crime range considerably between −0.70 for robbery and results on all index crimes which suggest that marginal imprisonments would even increase crime. Donohue (2009, p. 17) admits that his best guess for the elasticity is "*highly uncertain*", most likely being between −0.10 and −0.15 but "*conceivably within the broader interval between −0.05 and −0.40*". Durlauf and Nagin (2010) discuss the papers of Donohue's survey at some length. They criticize the statistical methodology employed in these studies and dismiss Levitt (1996) as the sole author of the survey who convincingly addressed the simultaneous interdependencies between crime and imprisonment rates. A further problem of studies relying on the imprisonment rate as the sole indicator of p is that it covers compound effects which might cause an omitted variable bias (see the subsequent subchapter on *Decomposing the Certainty Effect*).

Durlauf and Nagin (2010) consider studies based on police manpower more persuasive than those based on imprisonment rates. Their sympathy seems to be affected by the highly influential paper by Levitt (1997) and its critics on the choice of meaningful instruments. Durlauf and Nagin (2010) summarize the findings from these studies and further replication studies and mention an average elasticity of −0.3 for most estimates relating total crimes and police presence. However, given the high dependency of results on the type of crime under consideration (see below), such *typical* results covering the whole range of crime categories can hardly be considered useful for practical public policy purposes (which is in line with Donohue's "*uncertainty*" about his best guess).

A recent extensive meta-analysis based on 700 empirical studies with 7,822 estimates of the crime-preventing effect of general deterrence (certainty or severity of sanctions) by Rupp (2008; see also Dölling et al. 2009) reveals a large variation of effects depending on the choice of the deterrence variable. For instance, using the indicator *ratio of convictions to reported crimes* produces a highly statistical evidence in favor of the certainty of sanctions (median t-value = −3.5), whereas using the clearance rate would produce less significant results (median t-value = −1.9; Dölling et al. 2009, Table 3). Results also differ with respect to the crime category under consideration. Rupp (2008), Table 3.43, p. 126 finds that among the offenses which have been in best accordance with the deterrence hypothesis are speeding (median t-value = −2.2), tax evasion (−2.1), severe theft (−2.1), and fraud (−2.0), whereas sexual assault (−0.5), manslaughter (0.0), and drug dealing (0.0) were only rarely found to be consistent with the deterrence hypothesis. Thus, very prominent in the group of consistent types of crimes are nonviolent crimes, while the discordant part seems to incorporate violent crimes and drug-related offenses.

Decomposing the Certainty Effect. Testing the theoretical prediction of p, that is, the certainty of sanction, requires empirical measurement of the joint probability of detection *and* subsequent punishment. Many empirical models are restricted to only one indicator such as clearance rates (Entorf and Spengler 2000) or number of police officers (Levitt 1996). However, variations in clearance rates may not change the threat of expected sentences when prior to judicial decisions cases are dismissed or discharged by the public prosecutor. Moreover, expected sentences differ in response to court decisions, which can be dismissal, unconditional imprisonment, probation, (financial) fine, or other sanctions such as educational measures. Thus, to improve the consistency of empirical models with their theoretical counterparts and in order to analyze the effectiveness of the various sources of deterrence, it is necessary to decompose the certainty effect p into its components. In its simplest way, this can be done by distinguishing between the probability of detection (clearance rate), p_{cl}, and the conditional probability that a suspect is sentenced, $p_{s|cl}$ (probability of a sentence, conditional on detection), that is,

$$p = p_{cl}\, p_{s|cl}. \tag{4}$$

The variation of certainty effects becomes more transparent when changes in clearance or arrest rates are separated from the ones of courts and sentences. To fully understand such variations, changes in $p_{s|cl}$ can be further decomposed into changes of the indictment rate (driven by public prosecution), $p_{court|cl}$ (i.e., the probability that a suspect is brought to court), and those stemming from court decisions, $p_{s|court}$ (i.e., the probability that a suspect is convicted, given he or she is brought to court), eventually leading to $p_{s|cl} = p_{court|cl}\, p_{s|court}$. In both the US and European legal systems, the (high) discretionary power of the prosecutor is to determine which case should be disposed of before trial either by dismissal of the charges or by imposing certain obligations on suspects in exchange for laying the file aside. The picture is still incomplete unless the risk of the most severe outcome conditional on detection has been covered, that is, the certainty of a (unconditional) prison sentence. After including $p_{pris|s}$, that is, the probability of imprisonment given a conviction, the full decomposition of the probability of imprisonment (conditional on detection) looks as follows:

$$p_{pris|cl} = p\, p_{pris|s} = p_{cl}\, p_{court|cl}\, p_{s|court}\, p_{pris|s}. \quad (5)$$

Not surprisingly, also empirical indicators of $p_{pris|cl}$ vary across time and space (see Entorf and Spengler 2013) Thus, to summarize, the probability of a sanction depends on the interplay of many factors, and all of them are far from being constant parameters.

Reasonable public policy analysis and recommendations should take the role of all key players and their decisions, that is, interplay of police, prosecutors, and courts, into account. Only few studies also cover the risk of convictions, for example, by the ratio of convictions to arrests. Cornwell and Trumbull (1994), who were among the first applying panel econometrics for testing general deterrence models, present exceptional work because of their comprehensive list of law enforcement variables containing the probabilities of arrest, conviction (conditional on arrest), and imprisonment (conditional on conviction) as well as the severity of sanctions. However, the overall impression is that only few studies are taking account of a comprehensive list of factors.

In particular, the interplay of conviction rates and sentence lengths, that is, the essential components of certainty and expected severity of punishment, has been neglected. However, it is obvious that components of general deterrence do not work independently of each other. Missing factors of the law enforcement system might cause severe omitted variable biases (Entorf and Spengler 2013).

On the Impact of Severity

Death Penalty. The severity of punishment is the fundamental idea of capital punishment: "*. . . if rational people fear death more than other punishment, the death penalty should have the greatest deterrent effect*" (Ehrlich 1977, quotation found in Rupp 2008, p. 22). Although this statement is questionable, either because the expected sentence would be nil when there would be no (perceived) risk of detection or because "*rational*" suicide bombers expect going to paradise with the infamous 72 virgins, the focus of most studies in the aftermath of Ehrlich (1975) lies on the absolute deterrent effect of capital punishment. Most publications refer to US evidence, as can be seen from the fact that 71 out of 82 studies evaluated by Rupp (2008), p. 22, use US data and 5 are based on Canadian data. Ehrlich's (1975, 1977) articles have fueled an ongoing debate about the effectiveness of the death penalty. His results have been criticized for data errors, misspecification problems, and other methodological problems. Numerous subsequent studies have rejected but also confirmed significance of the death penalty. Donohue and Wolfers (2005) summarize several studies on the death penalty in the United States. They conclude that all outcomes are too fragile and that the number of executions is too low to draw any noteworthy and robust conclusion. This result is in line with the meta-study on the significance of death penalties by Dölling et al. (2009, Table 10), where median t-values of effects range between −0.5 and +0.1, that is, below usual significance levels. This survey is not the right place to summarize all facets of the complex debate on the death penalty which is interesting for its methodological aspects, but most probably the large majority of scientists from industrialized countries outside the USA would not consider it a practical and ethically tolerable alternative to lifelong sentences.

Evidence on the Length of the Prison Sentence and Prison Conditions. According to the aforementioned meta-study (Dölling et al. 2009, Table 3), most results using the average length of served prison sentences as an indicator of the severity of sanctions do not show statistical significance (median t-value = −0.6). This result is in line with the tenor of previous and recent surveys such as Eide et al. (1994), Durlauf and Nagin (2010), or Ritchie (2011). However, Durlauf and Nagin (2010) correctly state that most effects are measured as marginal effects in addition to already existing long sentences. Thus, more reliable evidence would be based on discontinuous jumps of severity. Levitt (1998) has found a significant drop in the offending of young adults when they reach the age of 18, that is, the age of jurisdiction for adult courts in Florida.

Other abrupt and unexpected changes of deterrence come from natural experiments. Maurin and Ouss (2009) and Drago et al. (2009) study the effect of external variations arising in response to collective pardons in France and Italy, respectively. Maurin and Ouss (2009) show that 5 years after release, those who have received a reduced sentence as a consequence of the pardon had a 12 % higher rate of recidivism than those who had received no reduction. In the case of the Italian clemency, inmates received a *conditional* reduction of prison sentences. In case of reoffending, they had to serve the remaining amount of their sentence (in addition to the new sentence). Drago et al. (2009) find that the threat of increased sanction continued after release: For every month the former prisoner would have to serve if convicted, there was a 1.2 % reduction in the propensity to recommit crimes. Both studies touch the problem of *specific deterrence*, that is, whether individuals would avoid future imprisonment because they are deterred by their own previous sentencing experience. However, in particular the study by Maurin and Ouss (2009) also tests the deterrent effects of longer prison sentences, though based on a subgroup of former inmates. In the case of Drago et al. (2009), the effect is a combination of certainty and severity because the threat of serving the remaining sentence also depends on the probability of detection and conviction.

Limitations and Extensions of the Classical Deterrence Model

Predictions of the fundamental model (as presented in Section 2) are based on a static model under regular neoclassical assumptions. The classical rational-choice model needs to be reconsidered in the face of challenging recent findings of behavioral economics. In particular, the (objective) detection probability and the disutility from punishment need to be adjusted given insights from prospect theory, myopia, and dynamic deterrence:

$$C_t = C(\underset{(-)}{\pi(p)},\ \underset{(-)}{\beta_c^{t+\tau} C_{C.t+\tau}},\ \underset{(-)}{\beta_{nc}^{t+\tau} E_{NC.t+\tau}},\ \underset{(-)}{E_{NC.t}}\, X). \tag{6}$$

The new elements of a *behavioral supply of offenses model* (Eq. 6) are to be introduced: Criminological research suggests that offenders have unusually high discount rates, that is, they place high value on immediate utility gains, whereas future events are strongly discounted, that is, they consider them as less important for current decisions than noncriminals. As consequences of criminal activities, if any, would have to be faced at some unknown time in the future, this so-called present bias (also called impulsiveness or myopia) leads to underestimation of expected *future* costs from punishment, $C_{C.t+\tau}$. In particular, adolescence, alcohol, or drug misuse (not to mention mental impairment) may lead to a very small personal weight of future consequences, that is, $\beta_c << 1$.

The classical static deterrence model also neglects future consequences of illegal activities which a rational forward-looking individual would take into account. Given the person will be detected and punished, she would expect diminished reemployment chances, reductions in social capital, relational problems, etc. Thus, rational individuals would also be deterred by the potential future reduction of expected legal utility gains, that is, shrinking $E_{NC,t+\tau}$, in response to the stigma of a criminal record. This potential threat is also called *dynamic deterrence*.

A related but different deviation from classical rational-choice and general deterrence models is the existence of *subjective* probabilities of detection and punishment, $\pi(p)$. These should be more precisely called decision weights because they do not satisfy the classical probability axioms. Allais (1953) was the first who criticized expected utility theory for its inconsiderate use of objective probabilities, that is, $\pi(p) = p$, in models of human behavior. Kahneman and Tversky (1979) developed an alternative model called prospect theory. They emphasized that "… *people are limited in their ability to comprehend and evaluate extreme probabilities, highly unlikely events are either ignored or overweighted […] Consequently, π is not well-behaved near the end-points*" (Kahneman and Tversky 1979, pp. 282–283). Experiments show that individuals overweight low probabilities of gains, that is, $(1 - \pi(p)) > (1 - p)$ in case of criminal gains, which implies $\pi(p) < p$, that is, underweighting the risk of being caught and punished. The overestimation of the criminals' own ability to avoid any apprehension and sanction is also known as *overconfidence* in the behavioral economics literature (McAdams and Ulen 2009).

The perceived risk of detection of re-offenders is of particular importance because of the high share of crimes committed by recidivists. Piquero and Pogarsky (2002) argue that criminals behave according to the belief that "lightning won't strike twice": They underestimate the probability of being detected and convicted because they (erroneously) believe that the risk of being punished again is very low. Some further details and discussions on the *perceived* (instead of objective) risk of being punished can be found in McAdams and Ulen (2009) and Ritchie (2011).

The Threat of Punishment When Trust in Criminal Law Is Absent

One of the stylized facts in empirical criminological research is the individual victim-offender overlap. A prominent explanation of the victim-offender overlap is the subculture-of-violence approach (Singer 1981), according to which individuals who attack others risk retaliations from former victims. Rational-choice theory seems to be generally questioned by such "*irrational*" retaliatory behavior of victims and criminals. However, retaliation can be seen as a rational deterrence strategy in a *prelegal* or *pre-societal* community. Given lacking access or trust in the public institution of criminal law systems in modern societies, victims might be tempted to take the law in their own hands. In particular in disadvantaged neighborhoods and subcultural societies where the retaliatory ethic of the *code of the street* (Anderson 1999) is used in lieu of criminal codes, the credible threat of punishing by strong retaliation might deter potential future perpetrators.

Rational-choice models are often criticized because they ignore that cognitive restrictions and emotional factors such as time pressure, peer group influence, or anger restrict the long-run *optimality* of individual decisions. As regards the overlap of victimization and offending, anger seems to be the major motivation of retaliatory behavior, as also stressed by many criminological and psychological research papers. Anger in response to perpetrated injury, frustration, and unfair treatment can be a triggering event that motivates *striking back*, not necessarily to the perpetrator himself but also to noninvolved bystanders and other available victims, also at some later point in time. Such behavior is often the consequence and origin of norms of honor and respect (or fear of dishonor and shame, respectively), prevailing and potentially escalating in subcultural societies (Anderson 1999) and serving as a deterrent factor when trust in criminal justice systems is absent. Such retaliatory behavior is not restricted to deprived subgroups: As suggested by findings in Fehr and Gächter (2002), punishing unkind behavior of others or *negative reciprocity* seems to be a social norm rooted in general human behavior. Participants in their experiments revealed some *altruistic punishment* behavior, that is, they punished defectors even when they had costly disadvantages from the retaliation. This so-called pro-social behavior has its origin in the notion of fairness as can be seen from the outcome of many ultimatum-game experiments: Responders often destroy any portion of their (guaranteed) gains when they perceive the proposal of the proposer as unfairly low. Thus, anger about unfair treatment is the individual motivation, but its social effect could be deterrence.

Future Directions

Certainty and severity of sanctions form the two cornerstones of general deterrence. This survey reveals that its underlying classical rational-choice theory has been reconsidered by recent research, yielding important insights from bounded rationality and behavioral economics. Future research will continue to analyze the threat of punishment, but the focus will be on the role of hyperbolic discounting, overconfidence, perceived risk, ambiguity, prospect theory, and dynamic models, just to name only a few of many interesting fields. Also the empirical approach has changed dramatically during the last 10 years or so. Modern criminological research is often based on laboratory and (quasi) natural experiments (such as changes in legislation or clemencies) which create convincing external variations of deterrence variables. From a criminal policy point of view, the problem of cost-efficient use of deterrence instruments has a high potential for future research. Given the relatively small effect coming from the severity of sanctions, one might be tempted to shift resources from imprisonment to policing. Though certainly high on the political agenda, one should nevertheless be aware of the limited empirical evidence (mostly coming from the USA and dealing with marginal effects of already long sentences). Moreover, alternative (long-run) cost-benefit considerations such as shifting scarce resources to education and families should be taken into account, since both poor education and family background are identified as the major reasons of lacking legal opportunities over the life cycle and resulting criminal careers.

Related Entries

- ▶ Deterrence of Tax Evasion
- ▶ Deterrent Effect of Imprisonment
- ▶ Economic Theory of Criminal Behavior
- ▶ Rational Choice, Deterrence, and Crime: Sociological Contributions

Recommended Reading and References

Allais M (1953) Le Comportement de l'Homme Rationnel devant le Risque, Critique des Postulats et Axiomes de l'Ecole Americaine. Econometrica 21:503–546

Anderson E (1999) Code of the street. Norton, New York

Becker G (1968) Crime and punishment: An economic analysis. Journal of Political Economy 78(2):169–217

Bentham J (1781) An introduction to the principles of morals and legislation. Batoche Books, Kitchener 2000, http://socserv.mcmaster.ca/econ/ugcm/3ll3/bentham/ morals.pdf

Cornwell C, Trumbull WN (1994) Estimating the economic model of crime with panel data. Review of Economics and Statistics 76(2):360–366

Dölling D, Entorf H, Hermann D, Rupp T (2009) Is deterrence effective? Results of a meta-analysis of punishment. European Journal on Criminal Policy and Research 15:201–224

Donohue J (2009) Assessing the relative benefits of incarceration: The overall change over the previous decades and the benefits on the margin. In: Raphael S, Stoll M (eds) Do prisons make us safer? The benefits and costs of the prison boom. Russell Sage Foundation Publications, New York

Donohue JJ, Wolfers J (2005) Uses and abuses of empirical evidence in the death penalty debate. Stanford Law Review 58(3):791–845

Drago F, Galbiati R, Vertova P (2009) The deterrent effects of prison: Evidence from a natural experiment. Journal of Political Economy 117(2):257–280

Durlauf SN, Nagin DS (2010) The deterrent effect of imprisonment. National Bureau of Economic Research working paper. http://www.nber.org/confer/2010/CRIs10/Durlauf_Nagin. pdf

Ehrlich I (1977) Capital punishment and deterrence: Some further thoughts and additional evidence. Journal of Political Economy 85(4):741–788

Ehrlich I (1975) The deterrent effect of capital punishment: A question of life and death. American Economic Review 65(3):397–417

Ehrlich I (1973) Participation in illegitimate activities: A theoretical and empirical investigation. Journal of Political Economy 81:521–565

Eide E, Aasness J, Skjerpen T (1994) Economics of crime: deterrence and the rational offender. North Holland, Amsterdam

Entorf H (2012) Expected recidivism among young offenders: comparing specific deterrence under juvenile and adult criminal law. Forthcoming in European Journal of Political Economy. Previous version published as IZA Paper No 5434

Entorf H, Spengler H (2013) Crime, prosecutors and the certainty of conviction. Forthcoming in European Journal of Law and Economics

Entorf H, Spengler H (2000) Socioeconomic and demographic factors of crime in Germany: Evidence from panel data of the German states. International Review of Law and Economics 20(1):75–106

Fehr E, Gächter S (2002) Altruistic punishment in humans. Nature 45:137–140

Genn H (2010) Judging civil justice. Cambridge University Press, Cambridge

Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. Econometrica 47(2):263–292

Levitt S (1998) Juvenile crime and punishment. Journal of Political Economy 106(6):1156–1185

Levitt S (1997) Using electoral cycles in police hiring to estimate the effect of police on crime. American Economic Review 87:270–290

Levitt S (1996) The effect of prison population size on crime rates: Evidence from prison overcrowding legislation. Quarterly Journal of Economics 111:319–352

Maurin E, Ouss A (2009) Sentence reductions and recidivism: lessons from the Bastille Day Quasi Experiment. IZA Discussion Paper 3990

McAdams RH, Ulen TS (2009) Behavioral criminal law and economics. In: Garoupa N (ed) Criminal law and economics, 2nd edn, Encyclopedia of law and economics. Edward Elgar, Northampton

Piquero AR, Pogarsky G (2002) Beyond Stafford and Warr's reconceptualization of deterrence: Personal and vicarious experiences, impulsivity, and offending behavior. Journal of Research in Crime and Delinquence 39(2):153–186

Ritchie D (2011) Sentencing matters. Does imprisonment Deter? A review of the evidence. Sentencing Advisory Council, Melbourne. http://www.abc.net.au/mediawatch/transcripts/1128_sac.pdf

Rupp T (2008) Meta analysis of crime and deterrence. A comprehensive review of the literature. Dissertation, Technische Universität Darmstadt. http://tubiblio.ulb.tu-darmstadt.de/38503/

Simon HA (1957) Models of man, social and rational: Mathematical essays on rational human behavior in a social setting. Wiley, New York

Singer S (1981) Homogeneous victim-offender populations: A review and some research implications. Journal of Criminal Law and Criminology 72:779–788

Tittle C (1969) Crime rates and legal sanctions. Social Forces 16:409–423

Child Pornography

Richard Wortley
Jill Dando Institute of Security and Crime Science, University College London,
London, UK

Overview

The pictorial and written portrayal of children as sexual objects has existed through the millennia and across civilizations, with examples dating back to ancient Greece, Rome, China, and India (Bullough 2004; Linz and Imrich 2001). In the modern era, the nature and availability of child pornography has been driven by technological advances, first the invention of the camera in the early nineteenth century and more recently the advent of the Internet and associated digital technologies towards the end of the twentieth century. The Internet, in particular, has transformed the production, dissemination, and consumption of child pornography, increasing exponentially the amount of material available and the number of people accessing that material. Prior to the Internet, images were traded furtively in hard-copy form among relatively small groups of dedicated users. The technical quality of the images was poor, the range limited, and the distribution networks unreliable and highly restricted (Ferraro and Casey 2005). By the early 1980s, law enforcement agencies could justly claim considerable success in stemming the international and domestic trafficking of child pornography. With the Internet and digitization came the ability to access vast quantities of technically high-quality images, cheaply, apparently anonymously, and all from the comfort of home.

It is argued in this chapter that the current child pornography problem is the result of the interaction between two factors. First, child pornography would not exist if there was not a pool of individuals – mostly men – who derive sexual gratification from viewing sexual images of children. This pool comprises those who have a preferential sexual attraction to children, but it also includes many individuals with a latent, casual, or ill-formed potential to become aroused by sexual images of children. Second, the scale of the child pornography problem that confronts society in the twenty-first century is a function of the increased availability and ease of access of child pornography provided by the Internet. The Internet is not simply an alternative delivery mechanism by which dedicated offenders satisfy their sexual cravings, viewing material that they would have otherwise obtained through some other means; the Internet has played an active role in stimulating the growth of child pornography, drawing in many individuals who, in the pre-Internet era, would have had neither the necessary drive nor the know-how to seek out and obtain child pornography.

What Is Child Pornography?

Child pornography is defined as a legal term for the purposes of investigation and prosecution and as a working construct for the purposes of research, prevention, and advocacy. A problem with legal definitions is that the meaning of both "child" and "pornography" varies from jurisdiction to jurisdiction. The age of consent around the world can be as low as 12 years (e.g., some parts of Mexico) and as high as 18 years (some jurisdictions in the United States), with 16 years perhaps the most common in the Western world (e.g., the United Kingdom, the Netherlands, New Zealand, Canada, and most parts of Australia) (e.g., Avert n.d.). Likewise, definitions of pornography also vary, and acts that are illegal in some countries may be legal in others. Generally, however, legal definitions distinguish between obscene material and artistic portrayals of nudity and eroticism, judging what is acceptable or unacceptable against community standards. Typically, more stringent standards are applied to images involving children than to images involving adults; child pornography need not necessarily involve obscene behavior but may include acts that are lascivious or suggestive.

The International Centre for Missing and Exploited Children (2008) conducted an audit of child pornography legislation among the 187 members of Interpol. It found that 36 countries do not specifically criminalize possession of child pornography, while a further 95 countries had no child pornography legislation at all. For countries that do define child pornography, there are variations in the exact content of the legislation. Apart from the age of the child, there is disagreement about whether computer-generated images are legal. For example, in a recent Australian case, a cartoon depicting Bart and Lisa from the Simpsons engaged in sexual acts was found to be pornographic, a judgment that could not occur in the USA.

Nonlegal definitions of child pornography can be more broadly framed than legal definitions and so avoid some of the debates around specific elements. According to the Office of the United Nations High Commissioner for Human Rights (2002), "Child pornography means any representation, by whatever means, of a child engaged in real or simulated explicit sexual activities or any representation of the sexual parts of a child for primarily sexual purposes." Child pornography may include still photographs, videos, audio recordings, and written material. It also may involve a wide range of acts that vary in degree of explicitness. Taylor et al. (2001) developed the COPINE scale that delineated ten levels of child pornography severity, ranging from nonsexualized pictures of children collected from legitimate sources such as magazines to

Child Pornography, Table 1 The COPINE scale

Content	Description
1. Indicative	Nonsexualized pictures collected from legitimate sources (e.g., magazines)
2. Nudist	Naked or semi-naked pictures of children in appropriate settings collected from legitimate sources
3. Erotica	Photographs taken secretly of children in which they reveal varying degrees of nakedness
4. Posing	Posed pictures of children in varying degrees of nakedness
5. Erotic posing	Pictures of children in sexualized poses and in varying degrees of nakedness
6. Explicit erotic posing	Pictures emphasizing the genitals
7. Explicit sexual activity	Record of sexual activity involving a child, but not involving an adult
8. Assault	Record of children subjected to sexual abuse involving digital touching with an adult
9. Gross assault	Record of children subjected to sexual abuse involving penetrative sex, masturbation, or oral sex with an adult
10. Sadistic/ bestiality	Record of a child subjected to pain or engaging in sexual activity with an animal

Source: Adapted from Taylor et al. (2001)

graphic depictions of children engaging in sexual acts with other children, adults, and even animals (Table 1). In a content analysis of child pornography identified by the Internet Watch Foundation (2010), 73 % of victims were assessed as being under 10 years of age, and around two-thirds of the images involved penetrative sexual activity between the victim and an adult, equivalent to levels nine to ten on the COPINE scale.

It should be noted that many researchers and advocacy groups object to the term child pornography, believing that it trivializes the severity of the behavior and may facilitate excuse-making by offenders by linking it to consensual adult erotica. It is argued that every so-called child pornography image is a record of a child's sexual exploitation. Alternative terms include child abuse images, child sexual abuse material, documented child sexual abuse, and child exploitation material.

Child Pornography and the Internet

The Internet and associated technologies have transformed the way that child pornography is produced, how it is distributed to users, and how those users view and retain it. While child pornography may be produced professionally by organized crime groups, the ready availability of inexpensive electronic recording devices (digital cameras and video recorders) permits individual offenders to create "high-quality," homemade images involving records of their own sexual abuse exploits. The Internet then provides an unparalleled distribution network, connecting producers and users from all corners of the world. As with the production phase, the distribution of child pornography may involve organized crime rings, but the Internet also allows offenders to interact directly with one another, exchanging files and providing information and social support. For users, the Internet means that vast quantities of child pornography can be acquired inexpensively and accessed with apparent anonymity at any time or place. The downloaded images do not deteriorate and can be conveniently catalogued, stored, and retrieved.

The Internet provides a number of methods by which child pornography can be distributed and accessed. Child pornography websites may be posted on the World Wide Web. Around three-quarters of these child pornography websites are commercial enterprises (Internet Watch Foundation 2008). In order to thwart detection efforts of authorities, images are often hosted on temporary websites, and customers may need to negotiate numerous redirects before they arrive at the target site. If a particular website is closed down, the images may reappear shortly thereafter at another location. Nevertheless, the open web is a relatively risky place to seek child pornography and tends to attract novice offenders; through the vigilance of the police and Internet service providers, the more dedicated and experienced offenders are increasingly being driven to deeper levels of the Internet (Jenkins 2001). Offenders may communicate directly with one another via chat rooms, newsgroups, bulletin boards, and peer-to-peer networks (Ferraro and Casey 2005; Wortley and Smallbone 2006, 2012). Entry into these offender networks may be highly controlled; potential members may need to gain the trust of existing members, and a password may be required to order to participate. As discussed further on, infiltration of these groups by undercover police is a major priority of agencies tasked with controlling child pornography.

Estimates of the size of the Internet child pornography problem are notoriously rubbery, usually involving little more than someone's best guess. Various indices have been used. The number of offenders worldwide has been estimated in the millions (National Center for Missing and Exploited Children 2005); the number of children estimated to be victims of child pornography is between 10,000 and 100,000 (Maalla 2009); at any one time, there are estimated to be over one million pornographic images of children on the Internet, with 200 new images posted daily (Wellard 2001); offenders have been arrested with as many as a million downloaded child pornography images (NewsFlavor 2011); and child pornography on the Internet is estimated to turn over between $3 billion (Ropelato 2003) and $20 billion (Hoover 2006) per year. Some have cautioned against the uncritical acceptance of such figures and being swept along in a moral panic (Jewkes 2010; Schottenfield 2007). Nevertheless, most commentators agree that child pornography is a serious problem that has escalated dramatically with the growth of the Internet.

Who Are the Offenders?

Perhaps the most remarkable feature of identified child pornography offenders is their general lack of remarkable features. There is no single offender type and nor is there a consistent set of sociodemographic or psychological characteristics that distinguish offenders from nonoffenders. Those arrested for downloading online child pornography have come from all walks of life and have included judges, dentists, soldiers, teachers, rock stars, and police officers. In a survey of 605 individuals arrested for possession of child pornography in 2006, Wolak et al. (2011) found that most were white (89 %), male (99 %), 26 years of age or older (89 %), and in full-time employment (61 %). Thirty-one percent were married or living with a partner, 29 % earned more than $50,000 per year, and 16 % were college graduates. A minority had previous sexual (9 %) or nonsexual (27 %) offence convictions.

Psychological assessments of offenders have produced mixed and sometimes conflicting findings, with few clearly distinguishing features emerging. Few arrestees have been found to be mentally ill (6 %) or sexually disordered (5 %) (Wolak et al. 2011). Laulik et al. (2007), using the Personality Assessment Inventory, found that fewer than 10 % of offenders scored in the clinically significant range on any of the clinical scales, with the exception of depression (30 %), borderline features (17 %), and schizophrenia (13 %). Tomak et al. (2009) compared child pornography offenders with incarcerated contact sexual offenders on the Minnesota Multiphasic Personality Inventory. Mean scores fell below the clinically significant levels for both groups, with the child pornography offenders scoring significantly lower than the other group on the psychopathic deviation and schizophrenia scales. Babchishin et al. (2011) conducted a meta-analysis of studies that have used psychological inventories to compare child pornography offenders with other sex offenders. In general, online offenders had fewer problems with victim empathy, cognitive distortions, and emotional identification with children, and there were no differences in loneliness and self-esteem. However, online offenders showed greater sexual arousal to images of children than did contact offenders, suggesting higher levels of pedophilia. One explanation for this latter finding is that the laboratory conditions under which sexual arousal was measured closely resemble the actual behavior of online offenders. In other words, sexual arousal on online offenders may be specifically triggered by sexual images of children, but this may not necessarily indicate a generalized sexual attraction to children in an offline environment (Wortley and Smallbone 2012).

It should be noted that almost all studies examining the characteristics of child pornography offenders have involved incarcerated samples. This is problematic because police tend to target the most serious offenders for investigation, and so results of research on these offenders are likely to be skewed. In a rare exception, Seigfried et al. (2008) anonymously surveyed 307 general Internet users. They found that 30 respondents admitted to having accessed child pornography online. Most surprisingly, ten of the users were females, representing a rate of offending that far exceeds official arrest figures. Overall 16 % of males and 6 % of females admitted accessing child pornography. There were no differences between users and nonusers on the "big-five" personality dimensions (extraversion, neuroticism, openness, conscientiousness, and agreeableness), but users scored lower on internal moral choice and higher on amoral dishonesty, suggesting that they were more likely to construe their behavior as morally acceptable. Overall, these findings indicate that there is a significant "dark figure" of child pornography use and that differences between offenders and non-offenders shrink when the full range of offenders is taken into consideration.

Perhaps the most contentious issue concerning the characteristics of offenders is the extent to which users of child pornography also engage in contact sexual offending. Conceptually the relationship between online and contact offending may take several forms. Child pornography may be sought out by individuals who have an existing sexual attraction to children; child pornography may have a long-term corrosive effect, and individuals may become increasingly attracted to child pornography and desensitized to the harm experienced by victims; or child pornography may be the sole outlet for an individual's sexual attraction to children and may even help him resist engaging in contact offending (Wortley and Smallbone 2006). In a meta-analytic review of the research, (Seto et al. 2011) found that on average 12 % of child pornography offenders also had official contact sexual offending histories, with the range reported in individual studies between 0 % and 43 %. Examining self-report data, the average prevalence of contact offending was 55 % with a range of 32–85 %. However, the studies examined by Seto et al. did not distinguish between child pornography users and child pornography producers. A great deal of the available child pornography involves offenders recording their own sexual abuse exploits and so necessarily involves contact offending. Again, too, all of the research in this area has been conducted on convicted offenders who, because of the priorities set by police when selecting cases to investigate, may be more likely than undetected offenders to also commit contact offences. It remains unclear how many individuals who view child pornography also commit contact offences.

While the research reviewed in this section generally indicates that on average child pornography offenders exhibit few distinguishing sociodemographic or psychological, some studies have reported considerable within group variation. It may be that averaged findings mask the existence of distinct subtypes of offenders. Cooper et al. (1999) suggested a three-category typology of general Internet pornography users that may be useful when considering the particular case of child pornography users. The categories are:

1. *Recreational users*: They access pornography sites on impulse, out of curiosity, or for short-term entertainment. They are not seen to have long-term problems associated with pornography use.
2. *At-risk users*: They are vulnerable individuals who have developed an interest in pornography, but may not have done so had it not been for the Internet.
3. *Sexual compulsive*: They seek out pornography to satisfy existing pathological sexual interests.

This typology, while not yet empirically verified for child pornography offenders, has good face validity and can help explain the dramatic escalation in the child pornography problem that has accompanied the growth of the Internet. While it may be comforting to divide the world neatly into pedophiles ("them") and nonpedophiles ("us"), in reality the potential to become sexually attracted to children falls along a continuum, and under the "right" conditions many men may be tempted to view child pornography (Wortley and Smallbone 2012). For the individuals in categories 1 and 2 of Cooper et al.'s (1999) typology, the Internet has provided the easy opportunity for them to explore hidden thoughts and desires in order to satisfy their sexual curiosities. Moreover, the virtual world is a disinhibiting environment that encourages individuals to engage in riskier behavior than they would ever consider in the real world (Quayle and Taylor 2001). Had these individuals lived in the pre-Internet era, it is likely that most would not have gone to the effort to seek out hard-copy child pornography. To be sure, there are offenders (those in category 3 of the typology) who are unambiguously attracted to children and who will determinedly and repeatedly seek out opportunities to access child pornography. Even for these offenders, however, had they lived in the pre-Internet era, the task of locating child pornography would have been much more difficult, their choice would have been restricted, and their collection of images would have been much smaller.

Policing the Net

The investigation of child pornography on the Internet presents unique challenges for law enforcement agencies (Wortley and Smallbone 2012). The Internet is complex network of networks with a decentralized structure that makes it difficult to screen content, enforce legislation, and track offenders. Even if one pathway to a child pornography site is blocked, there are many alternative pathways that can be taken to reach the same destination. Moreover, the reach of the Internet is global, and local citizens may access child pornography images that were produced and stored on another continent. Most investigations of child pornography on the Internet, therefore, will cross jurisdictional boundaries and require cooperation among agencies, often at an international level. International cooperation is in turn impeded by a lack of effective legislation in many countries and inconsistencies between countries where legislation does exit (International Centre for Missing and Exploited Children 2008). There is variation, too, in the capacity and commitment of countries to enforce child pornography laws and to act against offenders, either because of cultural reasons, through a lack of resources and know-how, or because of corruption (UNICEF 2009). In addition, rapid technological developments such as peer-to-peer networks, anonymous remailers, and file encryption make it easier for offenders to avoid detection (Ferraro and Casey 2005). Finally, the sheer volume of child pornography on the Internet and the number of offenders accessing that pornography can simply overload police resources. Police cannot hope to arrest every offender of whom they become aware and must prioritizetheir effort, for example, by targeting those offenders who possess "first-generation" (i.e., previously unknown) images that indicate the offender has made a record of his own sexual offending.

In order to meet these challenges, police must acquire new knowledge about the characteristics and modus operandi of child pornography offenders, develop specialized technical expertise in online offending, and forge links with other police agencies, the Internet industry, and nongovernment organizations. Many larger police departments will have dedicated Internet child exploitation teams. These teams may be supported by national agencies that provide specialized expertise and training, maintain centralized databases, track offenders and victims across jurisdictions, and help coordinate investigations. The activities of national agencies may in turn be coordinated through various international agencies and programs. An example is the Virtual Global Taskforce (VGT), an international partnership comprising nine members – national agencies from seven counties (Australia, Canada, Italy, New Zealand, the United Arab Emirates, the United Kingdom, and the United States) and two international agencies (Europol and Interpol). The VGT conducts coordinated international law enforcement operations and helps track down victims and offenders across borders.

Krone (2005) suggests that police investigations can be divided into four main types. The first type of investigation involves the targeting of individual offenders. Investigations may begin with information received from the public, for example, via one of the various child pornography hotlines. Child pornography may also be discovered during the investigation of contact child sexual abuse cases. There are also a number of documented cases in which computer repairers have found child pornography on an offender's hard drive and notified police. The most notorious case this category involved the pop singer Gary Glitter.

The second type of operation targets offenders who are members of covert groups and who operate in closed chat rooms, newsgroups, and peer-to-peer networks. Investigation may require police to operate undercover, infiltrating an offender group in the guise of a prospective member. Once the group has been infiltrated, the arrest of one offender and the interrogation of his Internet records can quickly lead to the arrest of other members of the group. There have been a number of international operations targeting covert groups, including Operation Cathedral, Operation Ore, and Operation Rescue (Wortley and Smallbone 2012). Operation Rescue, for example, involved the investigation of a covert newsgroup that was run from the Netherlands and had 70,000 members across the United Kingdom, the United States, New Zealand, Australia, and Thailand. The group was detected in 2007 by investigators in Australia and the United Kingdom, and an operation was launched, coordinated through the VGT. Officers infiltrated the group and took over the account of the UK administrator, giving them access to the member list. The Dutch founder of the group was jailed and 640 members were identified, 184 of whom were arrested.

The third type of operation targets website subscribers. These are offenders who purchase access to child pornography sites, which are typically set up as commercial enterprises. Because offenders need to pay for images, once accounts of the business are seized, offenders are usually tracked through their credit card details. The operation targeting Landslide Productions (called Operation Avalanche in the United States and Operation Ore in the United Kingdom) is perhaps the best known in this category. Landslide Productions was a child pornography company run by from Ft Worth in Texas and involved a complex network of some 5,700 computer sites around the world. Online customers, of whom there were more than 250,000, provided credit card details to obtain network access. The investigation began in 1999 when the US Postal Inspection Service discovered that Landslide's customers were paying monthly subscription fees into a post office box or via the Internet. A joint investigation between the Postal Inspection Service and the Internet Crimes Against Children Task Force (ICAC) was conducted over 2 years. Landslide's accounts were seized and officers tracked down the owners of credit cards used to pay for site access. There were 120 arrests in the United States, including the ringleaders in Texas and 1,300 arrests in the United Kingdom.

The fourth type of operation involves police stings. Stings may take a number of different forms. Undercover police may enter youth-oriented chat rooms posing as children and engage predatory pedophiles lurking in the group. Police may also set up "honeypot" traps, the best known example of which was Operation Pin. The operation was started in 2003 by West Midlands (UK) police and expanded to include other VGT members (the FBI, the Australian Federal Police, the Royal Canadian Mounties, and Interpol). A bogus website was set up purporting to contain child pornography. Visitors to the site were required to go through a series of web pages and at each point it was reinforced that they were in a child pornography site. If they ultimately tried to access an image, their credit card details, which they were required to provide to log in, were captured. The operation has resulted in numerous arrests although precise numbers are not available.

Prevention

Despite the best efforts of police, arresting offenders is not a viable long-term solution to the problem of Internet child pornography.

Certainly, offenders – and especially those involved in the direct sexual abuse of children – ought to be targeted by law enforcement personnel and arrested where possible. But no more than a few thousand people are arrested annually for child pornography offences in the United States (Wolak et al. 2011) and fewer still in other countries. As has been discussed, even highly successful coordinated international operations typically result in fewer than a couple of hundred arrests worldwide. If we accept that the number of individuals actively accessing child pornography via the Internet is in the millions, these figures represent a drop the bucket. For similar reasons, we cannot rely on the therapeutic interventions with known offenders – the usual response of psychologists – to make any significant dent in the problem. The prevention of Internet child pornography cannot be approached simply as a problem of individual sexual deviancy. The proliferation of child pornography in the Internet age has been a function of the increased opportunity offenders have been given to access images, and tackling the child pornography problem in any serious way must be centered on reducing that opportunity.

The task of reducing the opportunities offered by the Internet can be viewed in terms of the situational crime prevention model (Clarke 2008). This involves increasing the effort, increasing the risks, reducing rewards, removing excuses, and reducing provocations associated with accessing child pornography. The following discussion provides an example of each of these strategies. The list is not intended to be exhaustive (see Wortley and Smallbone 2012, for a more detailed analysis).

Increasing the effort involves making it more difficult for offenders to access images on the Internet. The principal strategy is to reduce the amount of child pornography that is available. When police become aware of child pornography on the Internet in many jurisdictions, they are empowered to issue a notice and take down (NTD) order, requiring Internet service providers (ISPs) to remove or filter (i.e., block the pathway to) the offending images. Arguably there has been some success in reducing the volume of child pornography in open areas of the Internet such as the World Wide Web. The Internet Watch Foundation (2010) reported that they took action against nearly 17,000 child pornography domains in 2010 and that it has on average 500 domains on its blacklist at any one time. One consequence of police activity has been to drive offenders to deeper and more secure levels of the Internet. While this is of concern, it is evident that offenders feel vulnerable in open areas of the Internet. The more hidden child pornography becomes, the more difficult it is for the offenders themselves to access it.

Increasing the risks most obviously involves making it more likely that the offender's online behavior will be detected. As has been discussed, however, even with active policing of the Internet realistically, an individual's chance of being arrested is in fact very small. But the risk of detection and the *perceived* risk of detection are two different things. Police stings described earlier demonstrate the importance of this distinction. It is noteworthy that the launch of Operation Pin was accompanied by widespread media coverage. The police did not hide fact that they were setting up honeypot traps, but rather they gave offenders explicit forewarning. Their aim was not to maximize the number of arrests but to create the uncertainty in the mind of potential offenders about safety of sites that they were considering accessing.

Reducing rewards involves frustrating attempts by offenders to derive benefits from child pornography. This strategy has been used particularly to target the distributers of child pornography who are seeking to benefit themselves financially. The Financial Coalition Against Child Pornography (FCACP) is a grouping of 35 ISPs, credit card companies, banks, and other companies providing online payment services. The goal of FCACP is to undermine the commercial viability of child pornography by tracking the flow of money in child pornography transactions and blocking payments for illegal downloads. The Financial Coalition Against Child Pornography (2011) claim that there has been a 50 % drop in the number of commercial child pornography sites following their activities. There has also been a dramatic increase in the cost of commercial child pornography as a result of the financial squeeze that has been applied, with monthly subscriptions rising from around $30 per month in 2006 to as much as $1,200 per month.

Removing excuses involves permitting offenders fewer opportunities to minimize the seriousness of child pornography offending. Offenders may take comfort in the belief that viewing child pornography does no harm (e.g., "They are just images not real people") and even that it is not illegal. Some countries (e.g., Norway) include an additional element in the NTD procedures designed to challenge these excuses. When access to a child pornography site is blocked, instead of a standard error message, the person trying to access the site receives a "stop" page containing a warning message. The precise wording of the stop page can vary, but the general intention is to make it clear to the potential offender that the behavior he is contemplating is illegal.

Reducing provocations involves removing situational factors that might trigger the impulse to view child pornography. For many offenders involvement with child pornography is an insidious, progressive process (Quayle and Taylor 2003). One pathway to child pornography is through legal adult-focused pornography and child pornography embedded within legal sites may be particularly problematic. Individuals may first encounter child pornography without having to actively seek it out and when they are already sexually aroused. While law enforcement agencies rightly focus resources on the most serious forms of child pornography, which typically appears on dedicated sites, attention should also be given to embedded child pornography.

Conclusion

While child pornography has a long history, the scale of the problem has increased exponentially since the introduction and rapid growth of the Internet. There are undoubtedly many more child pornography images available now, and many more individuals accessing those images, than would have been the case images had the Internet not existed. The Internet has created opportunities for individuals with no more than a casual interest in child pornography to satisfy their sexual curiosities and has allowed dedicated offenders to give full expression to their obsession, amassing collections that may exceed a million images. The Internet is not just an alternative method by which child pornography may be accessed; the Internet is a *cause* of child pornography.

Policing the Internet presents unique challenges for law enforcement. Despite the establishment of dedicated police units to tackle child pornography, and unprecedented international cooperation, the sheer volume of the problem threatens to overwhelm policing resources. Capturing personal details of offenders and launching prosecutions against those individuals is time consuming and resource intensive, and police become aware of many more offenders than they could possibly arrest. As well as prioritizing efforts on the most serious offenders, greater attention must be given to prevention. If the current child pornography problem is a product of the increased opportunities afforded by Internet, then solutions to the child pornography problem must seek to reduce these opportunities. Admittedly the task is not an easy one and it is easy to become pessimistic about the chances of success. But it is undoubtedly true that without the current prevention efforts, the problem would be much worse than it currently is.

Related Entries

- ▶ Criminal Career of Sex Offenders
- ▶ Pornography
- ▶ Situational Crime Prevention

Recommended Reading and References

Avert (n.d.) World wide ages of consent. Available at http://www.avert.org/age-of-consent.htm

Babchishin KM, Hanson RK, Hermann CA (2011) The characteristics of online sex offenders: a meta-analysis. Sex Abuse: J Res Treat 23(1):92–123

Bullough VL (2004) Children and adolescents as sexual beings: a historical overview. Child Adolesc Psychiatr Clin North Am 13:447–459

Clarke RV (2008) Situational crime prevention. In: Wortley R, Mazerolle L (eds) Environmental criminology and crime analyses. Willan, Cullompton, pp 178–194

Cooper A, Putnam DE, Planchon LA, Boies SC (1999) Online sexual compulsivity: getting tangled in the net. Sex Addict Compuls 6(2):79–104

Ferraro MM, Casey E (2005) Investigating child exploitation and pornography: the internet, the law and forensic science. Elsevier Academic Press, London

Financial Coalition Against Child Pornography (2011) Report on trends in online crime and their potential implications in the fight against commercial child pornography. Available at http://www.missingkids.com/missingkids/servlet/PageServlet?LanguageCountry=en_US&PageId=3703

Hoover JN (2006) As child porn industry grows, coalition launches counterattack. Information Week. 17 March. Available at http://www.informationweek.com/news/183700580

International Centre for Missing & Exploited Children (2008) Child pornography: model legislation & global review, 5th edn. Available at http://www.icmec.org/missingkids/servlet/PageServlet?LanguageCountry=en_X1&PageId=4348

Internet Watch Foundation (2008) Annual and charity report. http://www.iwf.org.uk/accountability/annual-reports/2008-annual-report

Internet Watch Foundation (2010) Annual and charity report. Available at http://www.iwf.org.uk/accountability/annual-reports/2010-annual-report

Jenkins P (2001) Beyond tolerance: child pornography on the internet. New York University Press, New York

Jewkes Y (2010) Much Ado about nothing? Representations and realities of online soliciting of children. J Sex Aggress 16(1):5–18

Krone T (2005) International police operations against online child pornography. Trends & issues in crime and criminal justice, No. 296. Canberra, Australia: Australian Institute of Criminology. Available at http://www.aic.gov.au/documents/3/C/E/%7B3CED11B0-F3F4-479C-B417-4669506B3886%7Dtandi296.pdf. Accessed 29 Apr 2010

Laulik S, Allam J, Sheridan L (2007) An investigation into maladaptive personality functioning in internet sex offenders. Psychol Crime Law 13:523–533

Linz D, Imrich D (2001) Child pornography. In: White SO (ed) Handbook of youth and justice. Kluwer Academic/Plenum, New York, pp 79–114

Maalla NM (2009) Promotion and protection of all human rights, civil, political, economic, social and cultural rights, including the right to development. Report of the Special Rapporteur on the sale of children, child prostitution and child pornography. United Nations, General Assembly, Human Rights Council

National Center for Missing and Exploited Children (2005) Child porn among fastest growing internet businesses. Available at http://www.missingkids.com/missingkids/servlet/NewsEventServlet?LanguageCountry=en_US&PageId=2064

NewsFlavor (2011) British man in possession of the largest collection of child pornography in history. NewsFlavor, 16 July. Available at http://newsflavor.com/alternative/british-man-in-possession-of-the-largest-collection-of-child-pornography-in-history/#ixzz1ZRvZUZ6k

Office of the United Nations High Commissioner for Human Rights (2002) Optional protocol to the convention on the rights of the child on the sale of children, child prostitution and child pornography, G.A. Res. 54/263, Annex II, U.N. Doc. A/54/49, Vol. III, art. 2, para. c, entered into force 18 Jan 2002. Available at http://www2.ohchr.org/english/law/crc-sale.htm

Quayle E, Taylor M (2001) Child seduction and self-representation on the internet. Cyberpsychol Behav 4:597–608

Quayle E, Taylor M (2003) Model of problematic internet use in people with a sexual interest in children. Cyberpsychol Behav 6(1):93–106

Ropelato J (2003) Pornography statistics 2003. Available at http://web.archive.org/web/20030621095030/http://www.internetfilterreview.com/internet-pornography-statistics.html

Schotenfield DL (2007) Witches and communists and internet sex offenders, oh my: why it is time to call off the hunt. St Thomas Law Rev 20:359–386

Seigfried KC, Lovely RW, Rogers MK (2008) Self-reported online child pornography behavior: a psychological analysis. Int J Cyber Criminol 2(1):286–297

Seto MC, Hanson RK, Babchishin KM (2011) Contact sexual offending by men with online sexual offenses. Sex Abuse: J Res Treat 23(1):124–145

Taylor M, Quayle E, Holland G (2001) Child pornography, the internet and offending. Can J Policy Res (ISUMA) 2(2):94–100

Tomak S, Weschler FS, Marjan G-H, Thomas V, Nademin ME (2009) An empirical study of the personality characteristics of internet sex offenders. J Sex Aggress 15:139–148

UNICEF (2009) Handbook on the optional protocol on the sale of children, child prostitution and child pornography. UNICEF Innocenti Research Centre, Florence

Wellard S (2001) Cause and effect. Community Care 1364:26–27, 15th–21st March

Wolak J, Finkelhor D, Mitchell KJ (2011) Child pornography possessors: trends in offender and case characteristics. Sex Abuse: J Res Treat 23(1):22–42

Wortley R, Smallbone S (2006) Child pornography on the internet. Problem-oriented guides for police. problem-specific guides series No. 41. U.S. Department of Justice. Available at http://www.popcenter.org/problems/child_pornography/

Wortley R, Smallbone S (2012) Internet child pornography: causes, investigation and prevention. Praeger, Santa Barbara

Choice Theory

▶ Integrating Rational Choice and Other Theories

Chronic Offenders

▶ Career Criminals and Criminological Theory

Chronic Victimization

▶ Repeat Victimization

Chronic Victims

▶ Multiple Victims and Super Targets

City Homicide Rates

▶ Youth Homicide in the United States

Civil Actions

▶ Civil Remedies

Civil Remedies

Martha J. Smith
School of Community Affairs, Wichita State University, Wichita, KS, USA

Synonyms

Civil actions

Overview

Looking at civil remedies through the broad prism of opportunity theory – which includes situational crime prevention (SCP) and its related theoretical and practical approaches – has a number of advantages for those whose primary interests lie either in civil remedies or in SCP. For those interested in civil remedies and third-party policing (TPP), this analysis does the following:

- Provides an integrated set of articulated concepts, with linkages to other similar approaches, for examining how civil remedies can be used for crime-prevention purposes
- Highlights some of the untapped crime-preventive potential of civil remedies

For those interested in SCP, this analysis aims to make the following clearer:

- How SCP operates to prevent crime opportunities
- How it might operate differently if civil remedies were used more often
- How it might be expanded to look at the mechanics of the crime-prevention process in more detail

This entry defines, in detail, SCP, civil remedies, and TPP, and discusses classification systems that have been used to highlight important features of both SCP and civil remedies. It uses four different "approaches" based on SCP and other related opportunity theories to examine how civil remedies could be used as part of SCP. Approach 1 focuses on the 25 techniques of situational crime prevention, finding that almost all of the techniques appear to be compatible with the use of TPP. Approach 2 looks at 25 individual SCP measures, how they operate to limit crime, and how they might be able to be adapted to incorporate civil remedies. Approach 3 looks at the different types of routine-activity controls available in relation to a crime event, how these controls could operate to prevent crime in terms of the SCP techniques, and how civil remedies might be able to be used to assist this process. Finally, Approach 4 uses a single crime script to illustrate the application of a civil remedy to a crime-prevention initiative, for use in planning, implementation, and evaluation. In effect, it unpacks and repackages the multiple layers of control involved when civil remedies are used in crime prevention. Although the potential uses of civil remedies are very broad, the need for careful assessment of potential effectiveness and efficiency, and the possibility of overregulation and other unintended consequences, is stressed. Caution is also recommended with the possible expansion of the use of civil remedies into criminal lifestyles as part of an opportunity-theory approach.

Fundamentals

An Opportunity-Theory Approach: Situational Crime Prevention and Related Approaches

"Situational crime prevention" refers to the attempt to alter settings or "situations," in which potential offenders might otherwise find themselves, so that it is less likely for crimes to occur at that time or place, in that manner, against that target, or with as severe an effect (see Clarke chapter in Wortley and Mazerolle 2008). It does not rely on understanding any long-term motivational characteristics of these potential offenders. Instead, it uses a variety of techniques to block crime opportunities or otherwise make crime less attractive for potential offenders. Its focus is on the more proximal causes of behavior and is intended to limit the harm before it occurs.

A number of different theoretical and practice perspectives have influenced the development of SCP theory and practice since it was developed initially by Ron Clarke and others (see discussion in Clarke 2010). Only two of these perspectives are discussed at length here – the routine activity approach (RAA) and the rational choice perspective (RCP). Although other perspectives, e.g., designing products against crime (see Ekblom 2005), and crime pattern theory (see, e.g., Brantingham and Brantingham chapter in Wortley and Mazerolle 2008), are useful for understanding the full scope of the use of civil remedies in crime prevention, RAA and RCP are particularly useful for highlighting the features of SCP that provide the sharpest focus for understanding civil remedies and third-party policing in terms of SCP and the opportunity-theory approach.

Civil Remedies

A civil remedy has been defined as "an action taken by an authoritative body – a legislature, a court, or an administrative agency – to enforce compliance with prescribed conduct or to impose a cost for failure to comply" (Mann 1992, p. 1908). As Cheh (1991) has pointed out, traditionally, with the exception of behaviors subject to regulatory safeguards, those who harm another person have been potentially subject to criminal law sanctions brought by the state or to civil law remedies sought by the harmed party. Only in the last 40 years or so has the civil law been used by the government, in her example, by the US federal government to bring divestiture and treble damage actions against businesses run by both white-collar and organized-crime offenders under the Racketeer Influenced and Corrupt Organizations Act (RICO). At its best, the threat of this civil action is meant to disrupt criminal activities in a manner that may not be achieved through the ordinary workings of the criminal law where punishable behavior is set out in a statute and then punished if carried out. The civil aspect of this legislation not only focuses on other aspects of the crime or crimes, it allows the government to use a branch of the court system that has a lower standard of proof than is required in a criminal proceeding, where guilt beyond a reasonable doubt is used. Although Cheh (1991) pointed out some legal limitations to their use, civil remedies can be used both when the focus of the remedy is on the primary actor, the offender, and when they are used against another party who is not directly involved in the criminal activity. Because they are available and legal, however, does not necessarily mean that they are suitable, in terms of their crime-preventive effects, or desirable, in terms of their unintended consequences. These issues are discussed below in the section entitled "Potential Controversies."

Third-Party Policing

"Third-party policing" (TPP) is the term used by Buerger and Mazerolle (1998, p. 301) to refer to "police efforts to persuade or coerce nonoffending persons to take actions which are outside the scope of their routine activities, and which are designed to indirectly minimize disorder caused by other persons or to reduce the possibility that crime may occur." They noted that, in practice, this persuasion or coercion most often operates through civil law regulations. The "third party" is the one who is forced to do the "policing." The term refers to actions undertaken through a variety of partnerships between police and other agencies and addresses both crime-preventive aims and crime control (Mazerolle and Ransley 2005). One could argue also that efforts by other crime-fighting elements within the government, such as a prosecutor carrying out community prosecution (Boland 2001), are also included under the TPP umbrella.

Civil Remedies and Crime Prevention

Civil remedies include such a broad group of actions that they do not easily fit under a single label developed for discussing crime prevention. For example, in the well-known typology of crime prevention developed by Brantingham and Faust (1977), which uses a public-health-model approach to the prevention of crime, a civil remedy such as a zoning code could be used as a form of primary prevention to prevent clubs from opening in an area, thereby changing what might otherwise be a potentially favorable environment for selling drugs. Another civil remedy, such as a curfew, might be used as a secondary crime-prevention measure to prevent teens (a group at risk for delinquency) from driving after midnight except in special circumstances. While a third civil remedy, such as an offender-notification provision, might be targeted at convicted felons as a form of tertiary prevention (of those who have displayed the unwanted behavior or condition).

Civil remedies can also be examined according to the means used to achieve a crime-prevention result. In terms of social prevention, the civil remedy is used to change social relations as a way of preventing crime in the future. For example, civil injunctions have been used to prevent gang members from associating with other gang members (Maxson et al. 2005), with the intention of disrupting the group's social relations. In terms of situational crime prevention (SCP), civil remedies can be used as part of an effort to reduce the opportunities for a potential offender to do crime in a particular setting or situation. An example of this type of application is when a city uses housing-code enforcement to get a landlord to alter the physical environment around a public housing unit so that many of the conditions favorable to drug sales are eliminated (Green 1996). Although civil remedies can operate within a variety of crime-prevention frameworks, in this entry, the focus is on using the integrated concepts associated with SCP, and its related theoretical perspectives, to examine how civil remedies are or can be used to effect opportunity reduction.

One major advantage of the situational approach, when compared to traditional rehabilitative or early-prevention approaches, is that it does not require preventers to be able to identify who the offender or potential offender is. This has implications for the use of civil remedies as part of a situational scheme, namely that preventive techniques or interventions that do require identifying offenders are not, strictly speaking, examples of SCP. Thus, they are subject to the same types of constraints related to detection that traditional post-offense policing faces. But they face another type of constraint or criticism that does not usually relate to SCP: The state is trying to impose a negative sanction on behavior without providing criminal law safeguards, as noted previously (see, e.g., Cheh 1991; Mann 1992).

Just as not all civil remedies fall under an SCP umbrella, not all SCP fit under a TTP umbrella. While a situational measure *can* be carried out by non-policing parties who are subject to civil sanctions for nonperformance of some required action, SCP does not *require* that the measures be carried out either with non-policing third parties nor does it *require* a threat of civil sanctions for these parties to carry out situational measures. Figure 1 sets out a simple framework for understanding the usual legally derived mechanisms that operate to influence potential

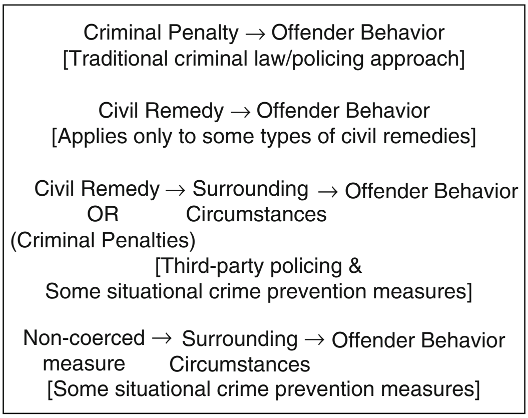
Criminal Penalty → Offender Behavior
[Traditional criminal law/policing approach]
Civil Remedy → Offender Behavior
[Applies only to some types of civil remedies]
Civil Remedy → Surrounding → Offender Behavior
OR
Circumstances
(Criminal Penalties)
[Third-party policing &
Some situational crime prevention measures]
Non-coerced → Surrounding → Offender Behavior
measure
Circumstances
[Some situational crime prevention measures]

Civil Remedies, Fig. 1 Simple comparative framework of the causal influences of different types of incentives on potential offenders' criminal behavior

criminal behaviors. The "surrounding circumstances" that civil remedies (and criminal penalties) may influence are the situational aspects of behaviors, including those controlled by third parties.

Legal Framework for Understanding Civil Remedies

In their book on third-party policing (TPP), Mazerolle and Ransley (2005, pp. 74–75) looked at the broad range of legal actions that could be used for some type of crime-prevention enforcement through a third (non-law enforcement) party. In their table on legal frameworks (pp. 74–75) for TPP, they presented a typology of the four types of authority that can be involved in legal enforcement actions:

1. Statutory provisions (such as a statewide or nationwide law)
2. Subordinate or lower-level legislation (such as a citywide ordinance)
3. Contracts (agreements between individual parties)
4. Torts (based on the legal duties between parties, which can arise based on the relationships between parties or on statutorily imposed duties of care)

Statutory provisions and subordinate legislation can include both civil and criminal ramifications for those whose behavior falls within the provisions set out. Contracts and torts are both subject only to civil law enforcement. Law enforcement agencies, on the other hand, do not ordinarily use contracts and torts as formal enforcement measures, although these agencies may work in partnerships with other parties who can use contract or tort enforcement to try to achieve a crime-preventive purpose. It is important to understand the legal framework underlying the civil remedy used when planning a preventive campaign that seeks to use such a remedy because this will help the preventer gauge the mechanism of the control available, as well as its obvious legal limitations.

Types of Civil Remedies

When planning for a crime-prevention campaign, it is also very important to understand all of the different types of civil remedies that might be available for use. Within the legal framework they set out, Mazerolle and Ransley (2005, pp. 74–75) also provided examples of types of interventions that use the civil law as an enforcement mechanism. They list seven types of statutory remedies:

1. Orders to control behavior of offenders, which includes both precursor behaviors and behaviors associated with more serious types of criminal actions
2. Movement and association limits, which includes limits imposed on types of people and types of situations with which a party can interact
3. Conduct licensing, which relates primarily to retail or service establishments
4. Formalized surveillance, which focuses on those who have been convicted of a crime
5. Property controls, which relate to real property owned or managed by private or public parties
6. Mandatory reporting by professional service providers
7. Civil forfeiture, which focuses on the proceeds of crime

As noted previously, these types of civil remedies fall under the general umbrella of SCP to the extent that they are situationally, and not offender, focused.

Mazerolle and Ransley (2005) also included three categories of subordinate legislation:

(1) orders under regulatory codes primarily related to health and safety issues; (2) product and service standards such as those that might make crimes associated with the products easier to carry out; and (3) controlled zones that limit or allow certain types of behavior within them. Product standards legislation falls under the general category of designing crime out of products, an area of opportunity theory that has seen a great deal of theoretical expansion recently (see Ekblom 2005). Crime pattern theory (CPT), which centers on the theory and practice of environmental criminology (see Brantingham and Brantingham chapter in Wortley and Mazerolle 2008), emphasizes the study of movement patterns of potential offenders and crime targets, along with other types of crime patterns. This aspect of opportunity theory is particularly useful for studying the effects of zoning – a civil regulatory remedy.

Approach 1: Focusing on the Situational Crime-Prevention Techniques Themselves

To understand how SCP works in practice and how this practice relates to civil remedies and TPP, one place to start is with a general introduction to the types of opportunity-reducing techniques that are considered to be situational and how they may operate to reduce crime. Here, the focus is on the most-recent SCP classification scheme – 25 techniques organized according to how they operate to affect offender decision making (see Cornish and Clarke 2003). Cornish and Clarke set out a table that organized the 25 specific techniques into five general categories of techniques with five specific techniques each. Smith and Clarke (2012) have recently produced a parallel table that sets out an explanation describing each of the 25 techniques and the five organizing categories. Here, the five general categories from Cornish and Clarke (2003) are provided along with summaries of their descriptions from Smith and Clarke (2012, pp. 306–307). The descriptions from Smith and Clarke included references to the concepts of offender, crime target (or victim), and target guardian derived from the routine activity approach (RAA), which is discussed in detail in this entry under "Approach 3," and in the following section.

- **Increase the effort** – which focuses on blocking the *offender's* actions or movement, thereby stopping the crime from happening – includes the following techniques:
  1. **Target harden** – affect how an offender might get to or use a target
  2. **Control access to facilities** – limit how an offender might gain access to a place
  3. **Screen exits** – monitor how an offender might exit from a setting where a crime can occur
  4. **Deflect offenders** – change potential offender's movement patterns
  5. **Control tools/weapons** – limit access to or use of things that make it easier to do the crime
- **Increase the risks** – which focuses on the *guardian's* potential role in detecting the crime and capturing the offender – includes the following techniques:
  6. **Extend guardianship** – encourage unofficial guardians
  7. **Assist natural surveillance** – make it easier to see criminal behavior
  8. **Reduce anonymity** – increase likelihood of identifying offenders
  9. **Utilize place managers** – use new or existing employees
  10. **Strengthen formal surveillance** – provide official guardians or increase their effectiveness
- **Reduce the rewards** – which focuses on decreasing the value of the *target* for an offender or otherwise making it harder for the offender to achieve his or her purpose – includes the following techniques:
  11. **Conceal targets** – limit offenders' ability to see targets
  12. **Remove targets** – take potential targets away or reduce their value
  13. **Identify property** – mark potential targets
  14. **Disrupt markets** – make the transfer of proceeds difficult
  15. **Deny benefits** – make it difficult to use targets for the purposes intended

- **Reduce provocations** – which focuses on the *situational stimuli* that may, in some situations, lead to crime – includes the following techniques:
  16. **Reduce frustrations and stress** – make settings and procedures calm
  17. **Avoid disputes** – limit situations that encourage disputes
  18. **Reduce emotional arousal** – limit insulting behaviors in a setting
  19. **Neutralize peer pressure** – lessen stimuli that encourage the desire to gain approval through crime
  20. **Discourage imitation** – limit details of modus operandi that can be copied
- **Remove excuses** – which focuses on strengthening the *situational controls* that may limit the offender's ability to excuse or ignore his or her criminal behavior in that situation – includes the following techniques:
  21. **Set rules** – detail which behaviors are unacceptable
  22. **Post instructions** – set out how to behave in the setting
  23. **Alert conscience** – put up reminders about behavioral requirements
  24. **Assist compliance** – make it easier to do acceptable behaviors
  25. **Control drugs and alcohol** – make it easier to limit intake of behavioral disinhibitors

In terms of TPP, only strengthen formal surveillance (technique 10) on its face, focuses on more formal types of policing. Yet, even with this technique, two of the three measures set out in the original table by Cornish and Clarke (2003, p. 90) – "burglar alarms" and "security guards" – can be used by third parties.

Approach 2: Looking at Particular Situational Measures and Civil Remedies

This approach also focuses on the 25 SCP techniques set out in Cornish and Clarke (2003), but it explores in more detail whether civil remedies could, at least theoretically, be used to help achieve the type of modification to the situation intended with the SCP measure. Table 1 looks at each of the 25 techniques, focusing on one measure set out as an example in the original table (Cornish and Clarke 2003). Each row of Table 1, performs the following: (1) links these measures to a crime, (2) sets out how the measure operates (drawing on the discussion in Smith and Clarke 2012), (3) includes the focus of the preventive measure (discussed in Smith 1998), (4) identifies the controller, using concepts derived from the routine activity approach (RAA) that are discussed in the next section, (5) provides the probable super controller (drawing on the work in RAA and also discussed in detail below), (6) describes the potential source of authority of a civil remedy (Mazerolle and Ransley 2005, discussed above), and (7) includes the type of intervention, also using the terminology of Mazerolle and Ransley (2005) set out earlier in this entry.

In Approach 1, which looked just at the technique itself and how it tended to work to affect a potential offender's behavior, only one technique (technique 10 – strengthen formal surveillance) appeared to be unlikely to allow TPP incentives to be used with it. Looking specifically at the examples of potential measures that were set out in the original table, however, demonstrated that this conclusion was overbroad because individual measures that use this technique could be carried out by third parties and not just police personnel. The analysis set out in Table 1 confirms that it is important to look at specific measures in order to understand the types of incentives that might be used to carry them out. Here, another type of measure, roadside speed display boards (a measure that can be used to alert conscience (technique 23)) may be used primarily by police agencies and may not be amenable to the use of a civil remedy incentive, even though the technique itself is not so limited.

Table 1 has several other uses as well. It allows a prevention planner to see how different types of measures could be used in terms of their theoretical and operational purposes, unpacking the variety of controls that may be operating when a situational technique is used. It also illustrates that many of the measures in the original Cornish and Clarke (2003) table are linked to place and consequently to place managers. This

Civil Remedies, Table 1 Using civil remedies to implement 25 situational crime-prevention measures

Technique (Cornish & Clarke, 2003)	Crime to be prevented	Measure (Cornish and Clarke 2003)	How the measure works (Smith and Clarke 2012)	Focus of control (Smith 1998)	Controller of this measure (RAA)	Usual super controller (Sampson et al. 2010)	Potential sources of authority (Mazerolle and Ransley 2005)	Potential type of civil remedy (Mazerolle and Ransley 2005)
Increasing the effort (for the offender)								
1. Target harden	Car theft	Steering-column locks	Harder for thief to start the car	Target	Guardian	Regulatory	Statute	Product standard
2. Control access to facilities	Bombing	Baggage screening	Harder to get offender and needed instrument to place	Prop/ place	Prop controller/ manager	Regulatory	Statute/subordinate legislation	Controlled zones
3. Screen exits	Fare evasion	Ticket required for passenger exit	Controls leaving place after offense	Place	Manager	Financial/ organizational	Subordinate legislation	Service standards
4. Deflect offenders	Prostitution	Street closures	Alters parties' movement patterns	Place	Manager	Regulatory	Subordinate legislation	Controlled zones
5. Control tools/ weapons	Illegal arms dealing	"Smart" guns	Alters access to instrumentalities	Prop	Prop controller	Regulatory	Statute	Product standards
Increasing the risk (that a guardian will act)								
6. Extend guardianship	Street robbery	Go out in group at night	Encourages unofficial guardians	Target	Guardian	Groups	Contract	Incentives – e.g., group discounts
7. Assist natural surveillance	Street robbery	Improve street lighting	Increases the potential for guardians to see	Place	Manager	Organizational	Subordinate legislation/contract/ tort	Regulatory code order/enforce conditions/action for negligence
8. Reduce anonymity	Sexual assault	Taxi driver IDs	Increases the potential for identification	Place	Manager	Regulatory	Subordinate legislation	Conduct licensing
9. Utilize place managers	Vandalism	CCTV for double-deck buses	Uses employees as guardians	Place	Manager	Organizational	Statute/subordinate legislation	Product standards
10. Strengthen formal surveillance	Traffic violations	Red light cameras	Allows more effective official guardians	Place	Manager	Organizational	Subordinate legislation	**Regular policing – no TPP**
Reducing the rewards (of the target)								
11. Conceal targets	Car theft	Off-street parking	Limits ability to see targets	Target/ place	Guardian/ manager	Regulatory	Subordinate legislation	Regulatory code order

12. Remove targets	Theft from cars	Removable car radio	Removes target from car	Target	Guardian	Markets	Statute	Product standards
13. Identify property	Burglary	Property marking	Makes items traceable to owner	Target	Guardian	Financial/ organizational	Statute	Product standards
14. Disrupt markets	Theft	Monitor pawnshops	Makes property transfer difficult	Target	Guardian	Regulatory	Subordinate legislation	Conduct licensing
15. Deny benefits	Shoplifting	Ink merchandise tags	Takes away value of property	Target	Guardian	Financial/ organizational	Contract	Incentives
Reduce provocations (related to situational stimuli)								
16. Reduce frustrations and stress	Harassment	Efficient queues and polite service	Encourages efficient and calm setting	Place	Manager	Organizational	Tort	Actions for negligence
17. Avoid disputes	Assaults	Separate enclosures for rival soccer fans	Limits situations that promote conflict	Place	Manager	Organizational	Tort	Actions for negligence
18. Reduce emotional arousal	Sexual assaults	Controls on violent pornography	Limits insulting behaviors in a setting	Place	Manager	Regulatory	Subordinate legislation/contract	Controlled zones/condition enforcement
19. Neutralize peer pressure	Drunk driving	"Idiots drink and drive"	Lessens desire to gain acceptance through crime	Place	Manager	Organizational	Statute	Conduct licensing
20. Discourage imitation	Vandalism	Rapid repair of vandalism	Limits access to details of how to do crime	Place	Manager	Organizational	Subordinate legislation	Regulatory code order
Remove excuses (by strengthening situational controls)								
21. Set rules	Drug use by visitors	Rental agreements	Lists unacceptable behaviors and consequences	Place	Manager	Organizational	Contract	Condition enforcement
22. Post instructions	Trespassing	"No parking"	Provides information about setting requirements	Place	Manager	Organizational	Subordinate legislation/tort	Regulatory code orders/actions for trespass
23. Alert conscience	Speeding	Roadside speed display boards	Provides situational reminders of rules	Place	Manager	Organizational	Subordinate legislation	**Not TPP – unless on nonpublic road**
24. Assist compliance	Library book theft	Easy library checkout	Makes it easier to do acceptable behaviors	Place	Manager	Organizational	Contract – e.g., service provider and institution	Condition enforcement
25. Control drugs and alcohol	Drunk driving	Breathalyzers in pubs	Makes it easier to limit the taking of mind-altering substances	Place	Manager	Organizational	Subordinate legislation	Conduct licensing

suggests the importance of place in crime prevention (since common measures were in the table) and also suggests the overall importance of place in crime commission.

Approach 3: Focusing on Controllers and Third-Party Policing (TPP)

Examining the *manner* in which a situational change is affected (which is what SCP has traditionally focused on) has limitations for understanding how SCP and opportunity theory can be linked to the use of civil remedies. This is because it does not necessarily dictate *who* can deliver or implement the needed change. Even the descriptions set out in Smith and Clarke (2012) do not cover all of the possible controls that might be operating with each technique. This distinction can be illustrated using SCP technique 18, reduce emotional arousal. For example, enforcing good behavior on the soccer field (the second measure listed under technique 18 in Cornish and Clarke's (2003) original table, not shown here) can be carried out by someone who controls the venue, someone who is protecting the target of the potential offense, or by someone who can influence an offender, or by all three types of people present in the situation. This focus on *who* "controls" the elements that converge to produce a crime event is central to the routine activity approach and is also central to understanding how civil remedies are, or could potentially be, used to persuade or coerce a third party to act to carry out an SCP measure.

The Routine Activity Approach (RAA)

The routine activity approach has developed from the theory set out by Cohen and Felson in 1979 that looked at, among other things, the three minimum elements that had to converge for a crime to occur at a particular time and place: the presence of a likely (or motivated) *offender*, the presence of a *target* (or victim) for the crime, and the *absence of a guardian* capable to protect the target and interrupt or prevent the crime. Since 1979, RAA has been expanded to include the following: (a) an additional element needed for a crime, i.e., place; (b) additional parties who could operate on these necessary elements to prevent a crime event from occurring, i.e., a handler to control an offender, a manager to control a place, and a props controller to deal with the instrumentalities of a crime; and (c) the sources of the types of incentives that can motivate guardians, handlers, managers, and props controllers. Most of these developments in RAA were discussed by Brunet (2002). The explanation in this section on RAA builds on Brunet's work by also providing a detailed explanation of how SCP measures may be able to be carried out by those different types of controllers and, in turn, how civil remedies may sometimes be used to induce other actors to change the situation so that the actors closest to crime events can prevent them. This approach focuses on connecting TPP to a measure when viewing the measure from the perspective of the RAA role.

The Target Guardian

The actions of target guardians, or those who protect the target or victim from victimization, can increase the risk of an offender being detected, captured, and convicted. The mere threat of these consequences occurring may also operate to block or limit crime events (see, e.g., Poyner 1992, for his study of the use of cameras on buses to prevent vandalism by schoolchildren). The ultimate goal is to get the guardian to act, but the focus of the intervention itself, in this case the civil remedy, may be on another controller (such as the place manager, discussed below).

This risk-increaser aspect of the guardian's role is well explored in Cornish and Clarke's 25-technique table and in Table 1 (techniques 6 through 10, above). What may be less initially apparent, however, is that target guardians can also operate to increase effort (through technique 1) by making it harder to get to or use a target, such as a car, by strengthening locking devices and to reduce rewards (through techniques 11, 12, and 13) by concealing, removing, or marking a potential target to make it less visible, not present, or not suitable for the potential offender's purposes. Target guardians can also help to reduce provocations in situations through helping those present avoid disputes (technique 17) and reduce emotional arousal (technique 18, as noted previously). The inducements available with TPP may be easily envisioned for some of these measures. For example, legislation may be passed that requires car owners to put steering-column locks on their vehicles (see Webb 1997). Regulations that focus on steering-column locks for new vehicles require the participation of different types of target guardians than would be true for legislation requiring the device be placed retroactively on used vehicles; however, in both cases the guardian would not need to be physically present to influence the likelihood of the event occurring.

The Handler

In 1986, Felson expanded the players affecting the commission of crime events by discussing the role that offender handlers could play in controlling a potential offender. Although Felson took an expansive view of the role of the handler as a social control agent, more recently Tillyer and Eck (2011) have argued that it may be useful to look at the preventive and explanatory role of handlers, particularly in terms of the situational aspects of these roles. And, although the context is situational, the influence itself may rely to some extent on the existing social relationship between the offender and the handler. Tillyer and Eck (2011) set out four primary factors they saw as possibly effecting handler effectiveness: (a) social closeness; (b) willingness to intervene; (c) opportunity to intervene; and (d) knowledge of the situation and environment in relation to the offense. Thus, if one is expecting offenders to be influenced by those likely to be present when crime events are about to happen, then, given the factors identified as possibly important by Tillyer and Eck, it is friends, girl friends, and potential co-offenders – and possibly teachers, parents, or bystanders – who would be available to be utilized as handlers in this context. Although this was not discussed by Tillyer and Eck (2011), each handler is a potential focus of a third-party policing initiative that might operate as a legal inducement to handling potential offenders, if an appropriate means of regulating the behavior of these potential handlers could be identified. The use of civil or criminal parental responsibility laws, or curfews that punish parents for their child's noncompliance, could be used as inducements for potential handlers (see Brunet 2002).

There is another way that SCP theory may provide some guidance for thinking about the possible roles of handlers in TPP. In keeping with Tillyer and Eck's (2011) contention that the situational roles of handlers have not been often explored, most of the examples of SCP measures set out in Cornish and Clarke's (2003) 25-technique table would not normally be carried out by handlers. Nevertheless, there are a number of techniques that appear, on their face, to be open to handlers. For example, handlers could be used to do the following: deflect offender movements away from potential targets (technique 4); control the tools or weapons offenders might use (technique 5); and reduce the anonymity of potential offenders by convincing them to wear name tags or uniforms in particular types of settings (technique 8). In addition, four of the techniques that focus on providing situational controls (removing excuses) appear to fit fairly directly with the role of a handler who has some relationship with a potential offender, in particular: setting rules (technique 21); alerting conscience (technique 23); assisting compliance (technique 24); and controlling drugs and alcohol (technique 25). Similarly, four of the techniques in the "reducing provocations" category – avoiding disputes (technique 17); reducing emotional arousal (technique 18); neutralizing peer pressure (technique 19); and discouraging imitation (technique 20) – could be carried out, in some situations, by handlers. Interestingly, these last two categories of techniques – reducing provocations and removing excuse – were seen by Cornish and Clarke (2003) as special cases. Reducing provocation, in their view, involves life-threatening cases or cases in which there is repeated exposure to situationally provoking stimuli. Removing excuses would tend to be used for minor criminal acts, either ambiguously criminal or antisocial, and for novice offenders or those not commonly doing serious crimes. These types of acts might be amenable to interventions from handlers present in the situation, perhaps precisely because they are special cases either of particularly unusual situations that might not be normally covered by other types of guardians or because handlers might be better able to convince fairly novice or minor offenders not to do the offense. If handlers are present, they may have "knowledge about the situation and environment that allow and/or provoke the offense," "social closeness to the potential offender," "willingness to intervene due to personal investment, either economic or emotional," and "an opportunity to intervene" (Tillyer and Eck 2011, p.184). When viewed in terms of TPP, two questions arise: (1) How would such a civil inducement be framed to focus precisely on the needed control? (2) Is a civil remedy the most effective inducement to use for each of these situations? The general issues raised by these questions are discussed briefly at the end of this entry.

The Place Manager

Another expansion of the routine activity approach came when Felson (1986) discussed the importance of *place* in the convergence of offender and target, without a capable guardian present. Building on this growing number of crime event elements and guardians, Eck (1995) identified the importance of *place managers* in controlling the conditions in which crime opportunities might be present.

The power of the place manager role can be seen by looking back at Table 1 and noting that techniques 1 through 5, and 15 through 25 can easily be performed by someone who controls the place in which the prevention measure can be carried out. For example, if the target of a residential burglary is an apartment house, then the owner could be the place manager who installs sturdy window and door locks (target hardening the house – technique 1) to make it more difficult for an offender to enter. Place managers can also reduce frustrations (technique 16) by providing more seating in a bus terminal or set rules (technique 21) and require hotel guests to register. In addition, techniques 7 (assisting natural surveillance), 8 (reducing anonymity), and 9 (utilizing place managers) can each involve place managers as controllers, depending on the particular measures used. Examples of measures that fall within these classes include cutting foliage around parking lot fences (to prevent theft of or from cars), requiring those attending schools to wear uniforms (to allow school administrators to identify when students are leaving the school building to be truant), and setting up CCTV to monitor student behavior on buses. Similarly, place managers can disrupt markets (technique 14) by requiring employees to check serial numbers on goods in pawnshops and deny benefits (technique 15) by requiring graffiti cleanup on their property.

In terms of TPP, many of the regulations commonly used in communities center on places – where people live, work, spend their leisure-time activities, or travel through on their way to these places. For example, housing and building codes provide guidelines for construction and upkeep. Safety regulations set up restrictions on the types of protective gear that must be worn at certain workplaces. Licensing requirements determine what types of beverages are sold in an establishment and fire codes set capacity limits on the number of patrons who can be present. Vendor licensing limits who can sell goods on the sidewalks and where the sellers can operate. The enforcement of each of these types of restrictions can, in some cases, induce place managers to act in ways that make crime less likely in a particular place.

Props Controllers

Smith (1998), in her analysis of the role that civil remedies can play in SCP, identified a fourth type of controller, the props controller. The importance of props in the crime commission process had been identified previously in the SCP classifications tables of Clarke (1992) and Clarke and Homel (1997) in the technique of "controlling facilitators," and more recently in the 25-techniques table of Cornish and Clarke (2003) in the technique of "controlling tools/weapons." Cornish (1994) used the term "props" – drawing on the dramaturgical aspect of the

Civil Remedies, Table 2 Planning for an anti-vandalism campaign

Scene/ function	Script action	Situational control	Mechanism	Type of intervenor	Focus of control	Medium of control	Type of inducement	Inducement enforcer
Preparation	Buy spray can	Sales regulation	Increase the effort	Businesses	Prop	Controller	Criminal law	Government agency
	Find good setting	City paint-out program	Increase the effort	Multiagency partners	Scene/ place	Manager	Civil remedy	Government agency
Entry	Enter setting	Access control	Increase the effort	Individual owner	Scene/ place	Manager	Non-coercive inducement	Individual owner
		Entry-exit screening	Increase the effort	Individual owner	Scene/ place	Manager	Non-coercive inducement	Individual owner
Precondition	Loiter	Surveillance	Increase the risk	Individual owner	Scene/ place	Manager	Non-coercive inducement	Individual owner
Instrumental precondition	Select target	Remove target	Reduce the rewards	Individual owner	Target	Guardian	Non-coercive inducement	Individual owner
Instrumental initiation	Approach target	Surveillance	Increase the risk	Individual owner	Scene/ place	Manager	Non-coercive inducement	Individual owner
Instrumental actualization	Reach target	Protective screens	Increase the effort	Individual owner	Target	Guardian	Non-coercive inducement	Individual owner
		Legal target provided	Remove excuses	Individual owner	Scene/ place	Manager	Non-coercive inducement	Individual owner
Doing	Spray graffiti	Graffiti-resistant paint	Increase the effort	Individual owner	Target	Guardian	Non-coercive inducement	Individual owner
Post-condition	Get away quickly	Moisture-activated alarm	Increase the risk	Individual owner	Target	Guardian	Non-coercive inducement	Individual owner
Exit	Leave setting	Entry/exit screening	Increase the effort	Individual owner	Scene/ place	Manager	Non-coercive inducement	Individual owner
Doing (later)	"Getting up"	Rapid cleaning	Reduce the rewards	Individual owner	Target	Guardian	Non-coercive inducement	Individual owner

Source: Adapted from Smith (1998) and Cornish (1994)

crime script – to describe these instrumentalities of crime. Despite the importance of these crime facilitators, the role of the "props controller" has not gained wide acceptance in the RAA literature (but see Mazerolle and Ransley, 2005). This is probably because props are not a necessary element of all crime events and the term "controller" has developed a more generic meaning. Nevertheless, "props controller" is used here in keeping with its more crime-specific role (see Table 2, as part of the regulation of spray paint sales to disrupt a graffiti tagging crime script).

Which Type of Controller?

Sometimes it is difficult to determine exactly how a particular civil remedy operates to affect the prevention mechanism. One example of such a complex case is where local regulations require taxi drivers to post their identification in their vehicles. If the crime that is being prevented is sexual assault of female passengers by drivers, then the question becomes who or what is being controlled. Is it the offender, the place, or the victim? In effect, it can be all three, although usually one of these elements is the one that is primarily affected and the other two elements may change, or be changed, in response. In the case of displaying taxi driver credentials, the potential offender, if he is a licensed driver, is controlled because he loses his anonymity and the risks associated with a sexual assault increase. In effect, the offender has been "handled" by the display of his identity information. The potential target, the passenger, can act as her own guardian as she has the information needed to help her determine whether the driver is bogus and she needs to get out, or better yet, not enter. But both of these reactions to the control cannot occur unless the place is first controlled by the place manager, usually the owner of the taxicab, who is being induced to post driver credentials by the regulation. The owner of the cab may be its driver, but he or she would be acting in the role of the place manager, not the potential offender, by posting the identification. Thus, an issue that first appeared as one of "which type of controller" is operating, can be seen as a situation in which multiple levels of control must operate – the place controller and the handler or the place controller and the guardian – for the crime to be prevented.

Super Controllers

Recently Sampson, Eck, and Dunham (2010) have identified a different type of level of factors that they see as affecting all three categories of controllers – handlers, guardians, and managers. They term these factors "super controllers." This classification system is similar to "inducement type" identified by Smith (1998, p. 76) as "non-coercive encouragement, civil remedies and criminal penalties," but it is more extensive (with 11 types) and covers a larger variety of possible influences on controllers. These incentives include the following: (a) formal – organizational, contractual, financial, regulatory, and courts; (b) diffuse – political, markets, and media; and (c) personal – groups and family. The three main categories are types of controls used. The super controller is given the name of the incentive source, with the incentive playing the role of a super controller of controllers. This is a different type of role than that seen previously in RAA where the roles were less abstract and more easily assignable to a real person either in planning or in practice.

This conception of incentives expands RAA and in a very neat manner ties together RAA and RCP. Sampson et al. (2010) do this by suggesting that super controllers manipulate the same rational choice factors – effort, risk, reward, provocation, and excuses – when seeking to influence controllers' likelihood of controlling potential crime elements that were identified as important for potential offenders in the SCP 25-techniques table of Cornish and Clarke (2003). This linking together of the elements from rational choice theory which influence potential offenders, with the decision-making elements of ordinary actors who may act as controllers, provides support for Cornish's (1993) contention that the rational choice perspective should be seen as a meta-theory, or a theory of action, for criminological theory. Sampson and her colleagues said that these five rational choice factors could be used to help understand why controllers – handlers, guardians, and managers – may not always act to control their associated crime elements – offenders, targets/victims, and places – and to help preventers figure out how to overcome these influences.

Thus, there are two ways that this new view of incentives can be useful to TPP: (1) it focuses on the processes that influence controllers (potential TPP actors) – i.e., risk, reward, effort, provocations, and excuses; and (2) it can highlight what is already going on to control crime and this can allow preventers to see if more can be done through employing the types of TPP discussed by Mazerolle and Ransley (2005). This view of super controllers can be applied to TPP even though: (a) TPP overlaps with only two categories of super controllers – regulatory incentives and courts incentives – and (b) regulatory super controllers include not only governmental regulators but also private regulators.

Approach 4: Being Crime Specific – Crime Scripts and Civil Remedies

The rational choice perspective of Cornish and Clarke (see chapter in Wortley and Mazerolle, 2008) was developed to help explain why SCP works. Using concepts such as "purposive behavior" and "bounded rationality," it adopted a crime-specific approach to examining three different types of involvement decisions related to crime commission – involvement, habituation, and desistance – and two different types of models of the crime event – the event model and the crime script. Applying lessons from research, it stressed the importance of risk, reward, and effort in the decision making of potential offenders. This last aspect of RCP is reflected in the 25-techniques table of Cornish and Clarke (2003), which was discussed in relation to Approaches 1 and 2, above. The crime script, initially developed by Cornish (1994), takes a different approach and focuses on the requirements of the crime itself, the processes that need to occur for the crime event to be completed. When Cornish listed these required actions, he also showed that one could link other factors related to these needed actions, such as reasons for an action to fail or SCP measures that could block an action. In 1998, Smith expanded the number of factors that could be linked to an SCP measure, such as the type of intervenor, the focus of the SCP control, the medium of control (or controller, as discussed in Approach 3 above), the type of inducement (essentially the type of super controller, also discussed above), and the inducement enforcer. Table 2 has been constructed based on the discussion in Smith (1998) as a tool for examining the part that a civil remedy commonly plays in a prevention initiative – in this case, disrupting a tag-writing script.

The distinctly crime-specific approach of the expanded crime script is particularly good for planners to use when one single technique is unlikely to be able to block the crime in the setting where the prevention initiative is needed. Because Table 2 attempts to "unpack" each measure that could be used, in addition to its use in planning, it may also assist at the implementation and evaluation stages when it is important to be clear how each measure may be affecting each action in the crime script. Thus, it is particularly well suited for the application of civil remedies that may require the intervention of a number of different stakeholders or members of a partnership in addition to the usual police intervenor.

Potential Controversies

By noting the many ways that civil remedies might be able to be used to block crime opportunities does not necessarily mean that all of these uses are either problem free or desirable. They may not be the most effective means of inducing someone to act nor the most efficient in terms of money and effort, and there are potentially many unintended consequences possible with both the use of civil remedies and SCP. While there are a number of sources in which these issues are discussed, Mazerolle and Ransley (2005) examine many of the arguments related to the use of civil remedies, particularly those related to overregulation, and von Hirsch, Garland, and Wakefield (2000) present arguments questioning the use of SCP measures that may overreach in terms of privacy interests. Arguments in defense of SCP are clearly set out in the chapter by Clarke in Wortley and Mazerolle (2008). Given the wide range of both civil remedies and SCP measures, it would be wise for planners to assess the types of civil remedies and SCP measures to be used in a particular location carefully prior to implementation. Planners may choose to favor noncoercive measures over the use of civil remedies whenever possible or choose the less-intrusive form of opportunity reduction where there is little difference expected in the crime-reduction outcome. Unintended consequences, such as an increase in fear of crime following an initiative, should not be ignored by planners just because they may be difficult to assess. The potential impact of a measure on area users, as well as potential offenders, must be considered.

Unresolved Issues and Future Controversies

At several points in this analysis, it was not clear exactly how an intervenor would attempt to frame a civil remedy to induce a controller to act. For example, how exactly do you convince a friend or a co-offender to "handle" an offender with a civil remedy? The situation, with its signals of possible danger, might be enough inducement for a co-offender to act, but what about a friend? Not only might it be difficult to envision what might work in such a situation, other than a desire to protect a friend from him or herself, it is not at all clear that the most desirable tool to use in this situation is one that involves coercion by the threat of legal action.

Another unresolved issue or future controversy centers on the place that lifestyle should hold in an opportunity-theory approach. All of the involvement models of RCP list "lifestyle" as an important consideration in the three main stages of offending – initiation, continuation (or habituation), and desistance (see Cornish and Clarke chapter in Wortley and Mazerolle 2008). Cornish (1994) also discussed lifestyle as a future consideration in terms of developing crime scripts. This factor has not been explored widely in SCP, but may be an important factor to consider when thinking about the broad scope of some civil remedies. Caution, however, is warranted since the use of civil remedies to disrupt "criminal lifestyles" would tend to produce more regulation and this regulation might not always be targeted to those parts of the lifestyle that are most closely linked to, or part of, a crime script. Limiting regulation to activities linked closely to crime commission means that it will be less likely to impinge on liberties, such as those involving protected aspects of movement or association.

Related Entries

▶ Civil Remedies
▶ Integrating Rational Choice and Other Theories
▶ Rational Choice Theory
▶ Third Party Policing and School Truancy

Recommended Reading and References

Boland B (2001) Community prosecution in Washington: The U.S. Attorney's fifth district pilot project. U.S. Department of Justice, Office of Justice Programs, National Institute of Justice, Washington, DC

Brantingham PJ, Faust FL (1977) A conceptual model of crime prevention. Crime & Delinquency 22:284–296

Brunet J (2002) Discouragement of crime through civil remedies: an application of a reformulated routine activities theory. Western Criminol Rev 4:68–79

Buerger M, Mazerolle LG (1998) Third-party policing: a theoretical analysis of an emerging trend. Justice Quart 15:301–327

Cheh M (1991) Constitutional limits on using civil remedies to achieve criminal law objectives: understanding and transcending the criminal-civil law distinction. Hastings Law J 42:1325–1413

Clarke RV (ed) (1992) Introduction. Situational crime prevention: successful case studies. Harrow and Heston, Albany, NY

Clarke RV (2010) Crime science. In: McLaughlin E, Newburn T (eds) The Sage handbook of criminal theory. Sage, London

Clarke RV, Eck JE (2005) Crime analysis for problem solvers in 60 small steps. Community Oriented Policing Service, U.S. Department of Justice, Washington, DC

Clarke RV, Homel R (1997) A revised classification of situational crime prevention techniques. In: Lab SP (ed) Crime prevention at a crossroads. Academy of Criminal Justice Sciences and Anderson, Highland Heights, KY/Cincinnati, OH

Cohen LE, Felson M (1979) Social change and crime rates: a routine activity approach. Am Sociol Rev 44:588–608

Cornish DB (1993) Theories of action in criminology: learning theory and rational choice approaches. In: Clarke RV, Felson M (eds) Routine activity and rational choice. Transaction Press, New Brunswick, NJ

Cornish DB (1994) The procedural analysis of offending and its relevance for situational prevention. Crime Prevention Studies 3:151–196

Cornish DB, Clarke RV (2003) Opportunities, precipitators and criminal decisions: a reply to Wortley's critique of situational crime prevention. Crime Prevention Studies 16:41–96

Eck JE (1995) Examining routine activity theory: a review of two books. Justice Quart 12:783–797

Ekblom P (2005) Designing products against crime. In: Tilley N (ed) Handbook of crime prevention and community safety. Willan, Cullompton, UK

Felson M (1986) Linking criminal choices, routine activities, informal control, and criminal outcomes. In: Cornish DB, Clarke RV (eds) The reasoning criminal: rational choice perspectives on offending. Springer, New York

Felson M (1987) Routine activity and crime prevention in the developing metropolis. Criminology 25:911–931

Green L (1996) Policing places with drug problems. Sage, London
Mann K (1992) Punitive civil sanctions: the middleground between criminal and civil law. Yale Law J 101:1795–1873
Maxson C, Hennigan K, Sloane D, and Kolnick K (2005) Can civil gang injunctions change communities? A community assessment of the impact of civil gang injunctions. Report to the U.S. National Institute of Justice
Mazerolle L, Ransley J (2005) Third party policing. Cambridge University Press, Cambridge
Poyner B (1992) Video cameras and bus vandalism. In: Clarke RV (ed) Situational crime prevention: successful case studies. Harrow and Heston, Albany, NY
Sampson R, Eck JE, Dunham J (2010) Super controllers and crime prevention: a routine activity explanation of crime prevention success and failure. Security J 23:37–51
Smith MJ (1998) Regulating opportunities: multiple roles for civil remedies in situational crime prevention. Crime Prevention Studies 9:67–88
Smith MJ, Clarke RV (2012) Situational crime prevention: classifying techniques using "good enough" theory. In: Welsh BC, Farrington DP (eds) The Oxford handbook of crime prevention. Oxford University Press, New York
Tillyer MS, Eck JE (2011) Getting a handle on crime: a further extension of routine activities theory. Security J 24:179–193
von Hirsch A, Garland D, Wakefield A (eds) (2000) Ethical and social perspectives on situational crime prevention. Hart, Oxford
Webb B (1997) Steering column locks and motor vehicle theft: evaluations from three countries. In: Clarke RV (ed) Situational crime prevention: successful case studies, 2nd edn. Harrow and Heston, Guilderland, NY
Wortley R, Mazerolle L (eds) (2008) Environmental criminology and crime analysis. Willan, Cullompton, UK

CJS

▶ Drug Abuse and Alcohol Dependence Among Inmates

Clinical Trials

▶ Multisite Trials in Criminal Justice Settings

Closed-Circuit Television

▶ CCTV and Crime Prevention

Coalition

▶ Communities That Care

Cocaine

▶ Drug Abuse and Alcohol Dependence Among Inmates

Coercion

▶ Police and the Excessive Use of Force

Coevolution

▶ Innovation and Crime Prevention

Cognitive Change

▶ Bayesian Updating and Crime

Cognitive Forensics: Human Cognition, Contextual Information, and Bias

Itiel E. Dror[1] and Reinoud D. Stoel[2]
[1]University College London and Cognitive Consultants International, London, UK
[2]Netherlands Forensic Institute, The Hague, The Netherlands

Synonyms

Bias in Forensic Science; Confirmation bias; Contextual influences; Forensic Science and bias

Overview

Human judgment stands at the center of criminal justice. Forensic science is no exception; it is the human examiner who is the main instrument of analysis in most forensic disciplines. It is the forensic expert who compares visual patterns and determines if they are "sufficiently similar" to provide evidence that they originate from the same source (e.g., whether two fingerprints were made by the same finger, whether two bullets were fired from the same gun, whether two signatures were made by the same person). Such determinations are governed by a variety of cognitive processes. Without objective scientific criteria and quantification instruments, these judgments are subjective.

The cognitive nature of subjectivity is that it can be influenced and biased by extraneous contextual information. Forensic scientists work within a variety of such influences: from knowing the nature and details of the crime to being indirectly pressurized by detectives, from seeing the "target" to working within and as part of the police, from computer generated meta-data to appearing in courts within an adversarial criminal justice system, and so on and so forth – the contextual influences are many and they come in many forms, some of which are subtle. After over 100 years of using forensic evidence, only recently has the forensic community begun to acknowledge the role of cognition and bias in forensic work, establishing a new field we term as Cognitive Forensics.

Cognitive forensics does not only refer to issues of interpretation and bias, but includes a whole array of cognitive issues relating to forensic work. Therefore, Cognitive Forensics deals with how cognition relates to forensic science and how cognitive knowledge can guide and enhance forensic work. In this entry, we provide some cognitive background, illustrate their relevance to forensic domains, show different types of contextual influences, and suggest ways of minimizing their effects so as to increase objectivity in forensic science and enhance its value.

Fundamentals of Human Cognition

Information Processing

People, in everyday life and in their professional work, process information. We hear, see, touch, and receive input information from a variety of sources. The brain and cognitive system process this information and often produce an output in a form of a decision or an action. The information coming into the cognitive system is processed as "bottom-up" information. One of the things that make humans intelligent is that they do not process such information in a vacuum. Along with the "bottom-up" components, the human cognitive system uses a variety of "top-down" processes. These include past experience, knowledge, expectation, and a variety of other factors that take part and greatly influence how incoming information is processed.

Top-down cognitive information processing is widespread and takes part in most cognitive operations. This is critical because the brain has limited information processing capacity, and the top-down processing helps to deal with the huge amount of information coming into the brain. The incoming bottom-up input is too much information for the brain to fully process, and a variety of cognitive mechanisms, such as selective attention, chunking, and automaticity, enable the brain to effectively handle the bottom-up information.

These top-down mechanisms stand at the core of human intelligence and expertise. For example, cognitive attention mechanisms select the more relevant and important information based on our past experience, context, expectations, and a whole host of top-down information. While certain information is selected, other information is ignored and disregarded.

Expertise

As we gain more experience and knowledge, our top-down cognitive mechanisms get more and more powerful, until we eventually become experts. Expertise entails top-down cognitive mechanisms that improve performance and result in superior abilities.

However, with more and more reliance on top-down cognitive mechanisms and more effective information processing, potential weaknesses and vulnerabilities arise. Expertise and top-down processes entail computation trade-offs, and they – along with overall superior performance – can also result in restricted flexibility and limited control, may cause the experts to miss and ignore important information, introduce tunnel vision and bias, and can cause other effects that degrade performance and can result in error. Such phenomena are not specific to forensic experts, they are inherent cognitive side-effects of expertise, per se, and are a professional challenge in the medical domain, as well as in the military, policing, and other highly skilled expert domains (Dror 2011).

It is important to emphasize that the nature of the cognitive mechanisms that cause the vulnerabilities work at a level without awareness, meaning that the experts do not intentionally and consciously take these actions. They take place automatically and without awareness. This is critical, because not understanding the underlying cognitive mechanisms involved in these phenomena often results in ineffective ways to countermeasure their effects. For example, considering cognitive biases as an ethical issue is not only an ineffective way of combating them but is also unfair to the experts because it suggests that they are doing something wrong intentionally and that they can stop doing it by mere willpower – a misconception and misunderstanding of the very nature and mechanisms of cognitive bias (see, e.g., Page et al. 2012).

Forensic Evidence

Although forensic evidence is very powerful and important in criminal justice, we need to understand its limitations and to improve it based on cognitive understanding of its vulnerabilities. Objectifying forensic investigations through data based analyses is an important development; however, it is not sufficient. Since the forensic examiner is the main instrument of analysis in many, if not all, forensic impression and pattern evidence disciplines, cognitive factors play a major role in forensic casework, stressing the need for Cognitive Forensics: A field that overarches the different forensic disciplines and addresses a whole range of cognitive issues that relate to forensic work.

This is because the evidence from the crime scene (be it a fingerprint, a bullet, a handwritten note, etc.) does not ever match the suspect perfectly – there are always variations and thus they are never totally identical. It is up to the forensic examiner to determine the evidential strength with respect to the question whether they come from the same source or not – based on their relative similarity. However, there is almost always no objective instrumentation or quantitative data available to determine the evidential strength of the comparison (e.g., "sufficient similarity" or discrepancies that can (or cannot) be explained). It is to some degree subjective and in the "eye of the beholder." Therefore, the cognitive mechanisms of the human examiner play a critical role in this determination – the new area of Cognitive Forensics.

Cognitive forensics not only plays an important role in the final determination of forensic identification, but even in the initial analysis of the evidence. For example, when expert fingerprint examiners determine the minutia features in a fingerprint, they do not only differ among themselves, but the same expert – examining the same prints – often vary in the features they find. Therefore, the lack of objective criteria in some forensic disciplines introduces not only inter-examiner inconsistencies but also intra-examiner inconsistencies (Dror et al. 2011; Ulery et al. 2012).

As we discussed earlier, the cognitive top-down mechanisms play a large role in such expert determinations, and because they are susceptible to influences, expert forensic examiners may be biased, as we detail below. However, it is important to emphasize from the onset that such influences affect different examiners in different ways and that they are most powerful in difficult and hard to call cases (e.g., in DNA mixture interpretation; see Dror and Hampikian 2011). Furthermore, the existence of influences and bias does not mean that it determines the final decision outcome. There are many factors that are involved and they interact with each other in a variety of ways.

Forensic traces, by their very nature, are often collected at crime scenes, and therefore are often far from ideal. These traces may then become evidence examined by forensic examiners who in most cases are part of the police or work within an adversarial legal system (see Giannelli 1997; National Academy of Sciences 2009). The examiners cognitive processes, subjectivity, and susceptibility to contextual influences has not been well understood by the forensic community for a long time, and therefore appropriate measures have not been taken. For example, often the examiners are exposed to a whole range of contextual information which is sometimes irrelevant and not needed for their work, but may affect and bias their judgments.

It is important to realize that much of this contextual information, like confessions, eyewitness testimonies, and outcomes of other forensic investigations, may well be very important for the case (see ▶ Forensic Science and Criminal Investigation), that is, for the judge or jury to reach their ultimate decision (or even for the detective, to determine lines of investigation, etc.), but not relevant to – and potentially psychologically contaminating – the forensic examiner. What the forensic expert can provide is a piece of the puzzle: The evidential value of their forensic evidence, based on estimates of the probabilities of the evidence given at least two hypotheses. This evidential value determination can then be used by others (e.g., a judge, jury, detective) to make legal or investigatory decisions. One probabilistic framework to do this is the Likelihood Ratio approach for the interpretation of forensic evidence (see Evett 1998; Stoel and Sjerps 2012), and the forensic science community is slowly progressing toward such an approach.

Cognitive Forensics

Four things have contributed to the emergence of Cognitive Forensics:

Theoretical Work

Theoretical work based on cognitive science understanding of the brain and the cognitive system, and research in other expert domains, has suggested that forensic science work may be affected by extraneous information and may be biased. Such work covers many forensic domains and some of it goes back a long time. For example:

- In handwriting, back over a hundred years – Hagan (1894) suggested that “Where the expert has no knowledge of the moral evidence or aspects of the case in which signatures are a matter of contest, there is nothing to mislead him, or to influence the forming of an opinion.”
- In fingerprinting, also a hundred years ago – Faulds (1912) suggested that “In finger-print cases... mistakes may be made... over-anxiety to prove his case that may distort his view.”
- In forensic odontology, “It appears that the current practice of bitemark analysis is rich in sources of potentially biasing influences... fundamental recognition that some form of bias is likely to exist” (see Page et al. 2012).
- In fire investigation, “Various circumstances conspire to make fire scene cause and origin investigation particularly susceptible to the affects of cognitive biases” (see Bieber 2012).
- Etc...

A landmark paper theoretically addressing these issues was a 2002 paper by Risinger, Saks, Thompson, and Rosenthal, which discussed “observer effects.” It made the theoretical case that based on cognitive and social psychology many forensic examination results are doubtful because they are vulnerable to distortion by the context and the state of the forensic examiner.

Empirical Research

There has been sporadic and anecdotal empirical research examining whether forensic examiners are indeed affected by contextual information. This research consisted of an experimental set up that involves showing different examiners evidence within different contexts as part of a research or training (e.g., document examination (Miller 1984), fingerprinting (Langenburg et al. 2009), and hair examination (Miller 1987)). Such between-subject experimental design compares different examiners, and therefore results can be attributed to individual differences among the different examiners.

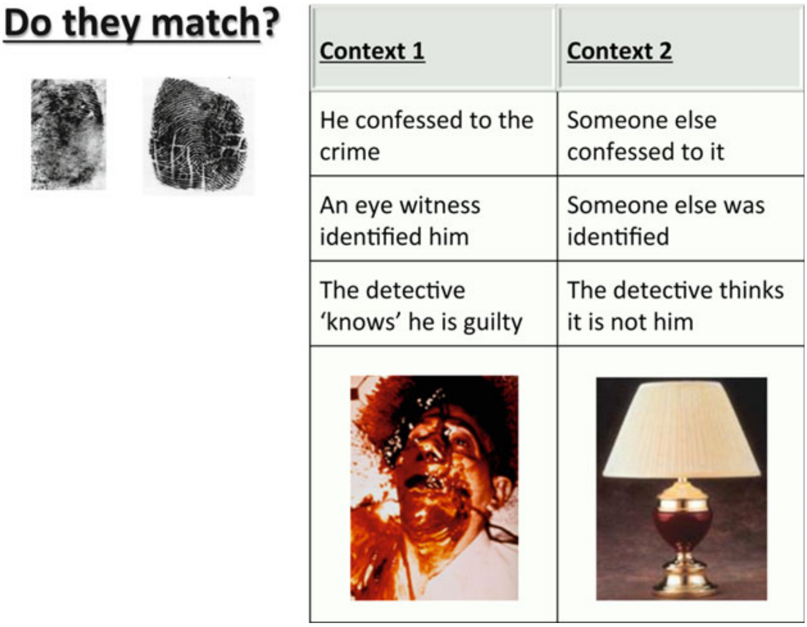

Context 1	Context 2
He confessed to the crime	Someone else confessed to it
An eye witness identified him	Someone else was identified
The detective 'knows' he is guilty	The detective thinks it is not him

Cognitive Forensics: Human Cognition, Contextual Information, and Bias, Fig. 1 The same examiners can reach different conclusions when the same evidence is presented within different extraneous contexts

A strong demonstration of biasing effects in forensic science was established in 2006 when data was collected from examiners as routine casework without their knowledge. This is very critical, because when examiners are not doing casework and taking part in research, they perform differently. As Dror and Charlton (2006) state "If you want to know how people drive, then their performance during a driving test is not very insightful and revealing, neither is their driving when they know they are near speed cameras or radars. One must try to observe and examine performance as well as collect data in the normal routine setting with minimal (or no) knowledge of the people involved" (p. 604).

This is especially vital when studying contextual influences – for them to affect performance, the participants must really believe them. This occurs in real casework, but not in contrived experiments. Furthermore, and perhaps critically important, in Dror et al. (2006) and Dror and Charlton (2006) studies, the same experts examined the same exact evidence on two different occasions (see Fig. 1). Depending on the difficulty of the case, the level of contextual information, and the type of conclusion, some examiners reached different decision outcomes. With all being equal (same examiners, same evidence) except the contextual information, one can confidently attribute the different conclusions to the effect of context.

Although these studies (as well as others) have found bias in forensic science, other studies have not found such affects. The existence of the latter studies with null findings, along with potential criticism of the former studies (e.g., small sample sizes) may suggest that there is no definite proof that forensic investigations are affected by contextual information. However, one must understand and remember that:

1. The existence of bias in forensic science is not only based on research in forensic science, but it rests on a large amount of theoretical and empirical research in a variety of expert domains, as well as on understanding the brain and cognitive mechanisms involved in expertise and making such decisions.
2. The existence of bias does not mean it necessarily affects the final forensic determination. As we state earlier, there are many factors that affect the forensic decision and interact with bias (such as difficulty of the case, type and strength of the bias, the examiner). Therefore, although bias affects the decision process, it does not necessarily affect the decision outcome.

Errors in Casework

Along with the theoretical and empirical research, a number of high-profile cases of errors in forensic identification brought attention to contextual influences and bias in forensic science work. Perhaps the most influential case was of Brandon Mayfield. He was incorrectly identified as the Madrid bomber by a number of FBI fingerprint examiners. A number of inquiries into this case all concluded that bias was a contributing factor.

The Office of the Inspector General (2006) inquiry concluded that "a significant cause of the misidentification was that the LPU examiners' interpretation of some features in LFP 17 was adjusted or influenced by reasoning "backward" from features that were visible in the known prints of Mayfield. This bias is sometimes referred to as "circular reasoning," and is an important pitfall to be avoided" (p. 7), and "working on a high-profile case influenced Green's initial judgment and created a mind-set in which his examination became biased by an expectation that the prints were a match" (p. 128).

Inquiries and Expert Working Groups

There have been a number of official inquiries into specific errors of forensic identification that further established the area of Cognitive Forensics. For example, the recent inquiry into the McKie erroneous identification (Campbell 2011) Recommendation 8 (Para 35.139) states that: "The SPSA should consider what limited information is required from the police or other sources for fingerprint examiners to carry out their work, only such information should be provided to examiners, and the information provided should be recorded" (p. 741), as well as "The SPSA should review its procedures to reduce the risk of contextual bias" and "The SPSA should ensure that examiners are trained to be conscious of the risk of contextual bias" (Recommendations 6 and 7 (Para 35.137 and Para 35.138), p. 741).

There have also been general inquiries into the state of forensic science. Most notable is the US National Academy of Science Inquiry (see ► Critical Report on Forensic Science). It concluded that: "a body of research is required to establish the limits and measures of performance and to address the impact of sources of variability and potential bias. Such research is sorely needed, but it seems to be lacking in most of the forensic disciplines that rely on subjective assessments of matching characteristics. These disciplines need to develop rigorous protocols to guide these subjective interpretations and pursue equally rigorous research and evaluation programs. The development of such research programs can benefit significantly from other areas, notably from the large body of research on the evaluation of observer performance in diagnostic medicine and from the findings of cognitive psychology on the potential for bias and error in human observers" (National Academy of Sciences 2009, p. 8).

Furthermore, a number of professional working groups have been established to examine these issues. The US National Institute of Standards and Technology (NIST) and the US National Institute of Justice (NIJ) have jointly formed a working group consisting of a large group of leading scientists and practitioners to examine these issues. After 2 years of work, this working group has acknowledged these issues and the need to deal with them.

The Forensic Regulator in United Kingdom recently concluded that "cognitive bias (also referred to as contextual bias or observer effects) is an issue that is relevant to forensic science" and that "organisations who undertake fingerprint examination should demonstrate within their accredited quality management system that they understand the potential for cognitive bias and build into their technical procedures safeguards to minimise the risk of bias and peer pressure" (Forensic Regulator 2011, p. 12).

Wider Influences on Criminal Justice: The Biasing Snowball Effect

The growing acknowledgment that bias may play a role in forensic work has helped establish the area of Cognitive Forensics and its importance. However, Cognitive Forensics does not stand alone in isolation. It affects the rest of the criminal justice system, which then affects it back, by what we term the "Biasing Snowball Effect." Many lines of evidence play a role in criminal justice, not only a variety of different forensic domains but also other lines of evidence, such as eyewitnesses and confessions. These should all be independent, but in reality, they can often affect and bias each other, resulting in a violation of evidentiary independence.

A firearm examiner, for example, knowing that the suspect's fingerprint was found on the gun, may be biased in their investigation of the bullets. Across lines of evidence, knowing the findings of one line of evidence can contaminate and bias another line of evidence. For instance, take the Willingham case in Texas, where erroneous forensic arson conclusions influenced eyewitness testimonies (Grann 2009), or the Jackson case in Las Vegas where an innocent person confessed to a crime he did not commit because of erroneous DNA identification (e.g., Mower and Mcmurdo 2011).

Bias may effectively give the right conclusions, but for the wrong reasons. So, although it may be "helpful" in some ways, it can also cause and reinforce miscarriages of justice (see examples above). Thus, by taking into account contextual information, forensic experts may well become more likely to interpret their evidence correctly, in the sense that they reach a conclusion that correspond to what actually happened. However, in doing this, the ability of the trier of fact to determine the truth is undermined. The crux of this criminalist paradox is that "By helping themselves be 'right' such analysts make it more likely that the justice system will go wrong. By trying to give the 'right' answer, they prevent themselves from providing the best evidence" (Thompson 2010, p. 130), therefore undermining themselves as objective scientists as well as their role in the judicial system.

Bias Countermeasures

Minimizing the effects of bias and other contextual influences is important and needed in forensic science. Nowadays, with the growing acceptance that these issues are real and relevant, it seems appropriate to take steps to deal with and countermeasure them. With increased implementation of such steps, they should be the accepted norm, and forensic evidence from laboratories that will not take such actions should be treated with doubt and reservation.

Best Practices and Standard Operating Procedures

The everyday working procedures of forensic laboratories must take into account bias and cognitive influences. As a first step, forensic examiners should not be exposed to information that is irrelevant to their work but that can bias them (for example, whether or not the suspect confessed to the crime or not, whether or not there are eyewitnesses or other evidence against the suspect, what the investigating detective thinks, and many other case details). It is relatively easy to deal with this kind of bias.

More complex situations arise when information that is potentially biasing may also be relevant to the examiner's work. In such cases, one must consider the relative importance of this information vs. the relative biasing influences it may have. If such a cost-benefit analysis does not provide clear results, one can examine whether or not this information makes a difference by giving this additional information only after initial conclusions have been reached. If the conclusion is maintained with this extra information, then that strengthens the results; if the new information suggests revising the conclusion, then the examiner must explain and detail the reasons. These types of procedures force to detail and account for the cognitive factors in the forensic decision making processes (Dror 2012).

There are more complex types of biases, some of which are very difficult to combat, for example, base rate information. Base rate information is organization- and discipline-specific information by means of which estimates may be made about the outcome prior to any investigation. For example, in airport security, the X-ray examiners have a base rate information that bombs will be detected very rarely. In the criminal justice system, for example, base rate information can cause a very high expectation for verification.

Best practices may involve artificially inserting fake cases into the normal workflow (with opposite conclusions) to the control and counter the effect of base rate information (Dror 2009). Such solutions are easy to implement in some domains, for example, using Threat Image Projection (TIP) software that occasionally adds threat items, such as knives and guns, to routine airport X-ray of ordinary bags (Schwaninger 2006). In forensic science it may be more tricky to implement because of the difficulty to create realistic cases, with the experts not being aware that the case under consideration is a fake case.

It is important that bias minimizing procedures, like context management, are both cognitively informed and pragmatic. They must be cognitively informed because forensic examiners who do not specialize in human cognition are limited in their ability to develop procedures to deal with bias. One must understand these phenomena in order to determine the best way to deal with them. On the other hand, cognitive scientists do not understand the realities of forensic laboratories and can come up with unrealistic procedures that cannot be implemented. Therefore, a cooperative effort between forensic practitioners and cognitive scientists is needed in order to move forward effectively.

Training

Best practices and appropriate procedures are critical, but without training their effectiveness is limited. Forensic examiners must understand the reasons and cognitive science behind the procedures and why they are needed. Training about cognitive factors in making forensic comparisons is essential. Indeed, forensic laboratories in different countries have begun to provide such training to their forensic examiners (e.g., New York and Los Angeles in the United States, Victoria Police in Australia, Greater Manchester Police and the London Metropolitan Police in the United Kingdom, and the Netherlands in Europe).

However, most forensic laboratories do not provide cognitive training to their examiners. As a result, most examiners lack cognitive understanding and its role in their work. An example of that is that a common, but incorrect, believe among forensic examiners is that cognitive bias does not effect them (i.e., they can be exposed to contextual information without being affected – see Dror et al. 2013 for debating this issue), or that bias is an ethical matter and that it can be dealt by mere will power (e.g., Thornton 2010).

Furthermore, proper cognitive training can help examiners understand when bias is most influential. This is important as it enables not to treat all cases in the same way, requiring the same bias countermeasures to be taken, thereby helping to selectively apply rules that are fit to purpose. A triage approach can be a practical and effective way of combating potential bias (for details, see Dror 2009).

Technology

Technology can be an important ally in combating bias. Computers carry out their algorithms "blind" to context. However, one must remember that they are programmed by humans and therefore may have biases within their algorithms. Furthermore, the technology may bias the human examiners working with it.

For example, Automated Fingerprint Identification Systems (AFIS) provide a list of potential candidates for the human experts to examine. However, recent empirical research has revealed that forensic experts are biased by this meta-data: They spent less time as they go down the list and are biased that if there is a correct match it will be at the top of the list. Therefore, they are more likely to make an erroneous identification on fingerprints that are on the top of the list and miss identifications that are lower down on the list (Dror et al. 2012).

Such technology-induced bias can be effectively reduced by providing shorter lists, by randomizing the order of candidates, or by planting correct matches further down the list as test cases to cognitively engage the examiners and as a quality control measure. This is just an example of how cognitive research and understanding is important for forensic science, not only to identify problems but also to provide solutions.

Selection

Being a forensic scientist is a highly sought-after profession, and many people apply to get into this domain. Nevertheless, there are not enough scientifically based and validated tests that quantify the specific abilities needed for doing the job. Most laboratories do not use any testing, and those who do, either use general tests that do not measure the specific abilities that are needed for the job, or they use inappropriate forensic evidence as test items. For example, many latent print laboratories use a test called Form Blindness. However, this test is built to test defects, not for quantifying special abilities and talent to perform fingerprint examination. This is not only a conceptual and theoretical difference, but it has practical implications. For example, "defect" testing, such as visual acuity and the Form Blindness test, should only test for minimal threshold requirement (i.e., if you do not pass, you should not be a fingerprint examiner). Such "defect" testing does not provide scores that enable to judge and rank the relative talent of the candidates. Furthermore, the Form Blindness test was developed over 80 years ago. This is decades before cognitive psychology had even emerged and very long before brain scans and other cognitive neuroscience tools and methodologies had been developed. In the past 80 years, our knowledge and understanding of the human brain and the cognitive system, and in particular visual cognition, has increased substantially. And finally, it lacks proper validation even in the area it was designed for implementation and use, that is, for handwriting examination.

It is important that forensic disciplines to have tests that are scientifically developed and validated to select the right people for the profession. Such tests should measure the specific abilities and skills needed to do the job, that is, the raw talent.

Research

Finally, more research is needed to understand how cognition relates to forensic work and how it can benefit forensic science (for example, when considering what factors determine the effects of bias (the examiners, difficulty of the case, types of bias, etc.) and how they interact). Through better understanding of these cognitive phenomena, we can determine how forensic science can best countermeasure their effects.

For example, the use of ACE (Analysis, Comparison, and Evaluation) in fingerprints (similar approaches exist in other forensic disciplines) is often not carried out linearly, one step at a time, sequentially, and independent of each other. In the initial, "Analysis" stage of ACE-V, a latent is assessed for its suitability for comparison. Suitable prints are then compared to potential matching prints. Cognitive understanding suggests that suitability judgments may differ in the presence of a comparison print due to the cognitive influence of contextual information (e.g., Folk et al. 1992; Awh et al. 2012).

We suggest that the ACE be conducted (and documented) linearly and sequentially, each phase independent of each other (see Dror 2009, for details). Although most laboratories do not subscribe to linear ACE, such changes have been implemented by the US Federal Bureau of Investigation (FBI) and the Netherlands Forensic Institute (NFI). For example, the revised Standard Operating Procedures (SOPs) of the FBI "include some steps to avoid bias: examiners must complete and document analysis of the latent fingerprint before looking at any known fingerprint" and "instructs examiners conducting analysis of a latent fingerprint to analyze it for evidence of distortion, determine whether it is 'of value,' and document the data used during analysis" (OIG 2006, p. 27).

Because the initial latent analysis in isolation may lack the benefit of direction guided by the comparison print, Dror (2009) suggests that examiners may be allowed to return and revisit the analysis stage, but they must document and justify it. Indeed, the Office of the Inspector General (2011) clearly takes this cognitively informed approach on board, citing this approach in its report: "a solution to bias may be requiring initial analysis of the latent fingerprint in isolation from the known fingerprints, but also permitting, with clear and detailed documentation, some 'reanalysis' of the latent print after comparison" (p. 28). A recent Expert Group set up by the National Institute of Standards and Technology (2012) has reached similar conclusions and has recommended that: "Modifications to the results of any stage of latent print analysis (e.g., feature selection, utility assessment, discrepancy interpretation) after seeing a known exemplar should be viewed with caution. Such modifications should be specifically documented as having occurred after comparison had begun" (Recommendation 3.2; see National Institute of Standards and Technology 2012). Furthermore, Dror recommends that examiners be restricted to the extent that such reanalysis be allowed (e.g., that "clear" features during analysis not be changed, but "ambiguous" ones can benefit from hindsight cognitive attention); for details, see Dror (2009).

This is an example of how research and cognitive understanding can directly suggest practical ways to enhance forensic work. Cognitive forensics, the domain of using such cognitive insights to enhance forensic science, covers a whole range of issues because human cognition plays many roles in forensic work.

Summary and Conclusions

Forensic evidence is an important component of criminal justice. However, it still often lacks objective criteria and quantification instruments and remains therefore subjective. With subjectivity comes the potential of being influenced and biased by extraneous information. Erroneous identifications, theoretical and empirical research, and official inquiries – all in the past few years – have made these issues very apparent. There is growing understanding and acceptance of these issues by the forensic community, which we term Cognitive Forensics.

Cognitive forensics cuts across forensic science and addresses the examiners – which are the instrument of analysis. It focuses on the underlying cognitive processing and how this can be optimized to produce the highest possible forensic results. With the growing field of Cognitive Forensics, more and more improvements will emerge which will help support criminal justice through strengthening forensic science.

Acknowledgement This work has been supported in part by a grant provided to I. Dror by the US National Institute of Standards and Technology (NIST) through the Department of Defense (DoD/TSWG) grant #N41756-10-C-3382.

Related Entries

- ▶ Critical Report on Forensic Science
- ▶ Expert Witnesses: Role, Ethics, and Accountability
- ▶ Forensic Science and the Paradigm of Quality
- ▶ Forensic Science and Criminal Investigation
- ▶ Forensic Science and Miscarriages of Justice
- ▶ Forensic Science Culture
- ▶ Legal Rules, Forensic Science and Wrongful Convictions

Recommended Reading and References

Awh E, Belopolsky AV, Theeuwes J (2012) Top-down versus bottom-up attentional control: a failed theoretical dichotomy. Trends Cogn Sci 16(8):437–443

Bieber P (2012) Measuring the impact of cognitive bias in fire investigation. International symposium on fire investigation, science and technology, 3–15

Campbell A (2011) The fingerprint inquiry report. APS Group, Scotland

Dror IE (2009) How can Francis Bacon help forensic science? The four idols of human biases. Jurimetrics J 50:93–110

Dror IE (2011) The paradox of human expertise: why experts can get it wrong. In: Kapur N (ed) The paradoxical brain. Cambridge University Press, Cambridge, pp 177–188

Dror IE (2012) Combating bias: the next step in fighting cognitive and psychological contamination. J Forensic Sci. 57(1):276–277 doi:10.1111/j.1556-4029.2011.01940.x

Dror IE, Charlton D (2006) Why experts make errors. J Forensic Identif 56(4):600–616

Dror IE, Cole S (2010) The vision in 'blind' justice: expert perception, judgment and visual cognition in forensic pattern recognition. Psychon Bull Rev 17(2):161–167. doi:10.3758/PBR.17.2.161

Dror IE, Hampikian G (2011) Subjectivity and bias in forensic DNA mixture interpretation. Sci Justice 51(4):204–208. doi:10.1016/j.scijus.2011.08.004

Dror IE, Rosenthal R (2008) Meta-analytically quantifying the reliability and biasability of forensic experts. J Forensic Sci 53(4):900–903. doi:10.1111/j.1556-4029.2008.00762.x

Dror IE, Charlton D, Peron A (2006) Contextual information renders experts vulnerable to making erroneous identifications. Forensic Sci Int 156(1):74–78. doi:10.1016/j.forsciint.2005.10.017

Dror IE, Champod C, Langenburg G, Charlton D, Hunt H, Rosenthal R (2011) Cognitive issues in fingerprint analysis: inter-and intra-expert consistency and the effect of a 'target' comparison. Forensic Sci Int 208:10–17. doi:10.1016/j.forsciint.2010.10.013

Dror IE, Kassin SM, Kukucka J (2013) New application of psychology to law: Improving forensic evidence and expert witness contributions. J Appl Res Mem Cog 2(1):78–81. doi.org/10.1016/j.jarmac.2013.02.003

Dror IE, Wertheim K, Fraser-Mackenzie P, Walajtys J (2012) The impact of human-technology cooperation and distributed cognition in forensic science: biasing effects of AFIS contextual information on human experts. J Forensic Sci 57(2):343–352. doi:10.1111/j.1556-4029.2011.02013.x

Evet IW (1998) Toward a uniform framework for reporting opinions in forensic science case work. Sci Justice 38:198–202

Faulds H (1912) Dactylography, or the study of fingerprints. Halifax, Milner and Co., London

Folk CL, Remington RW, Johnston JC (1992) Involuntary covert orienting is contingent on attentional control settings. J Exp Psychol Hum Percept Perform 18:1030–1044

Forensic Regulator (2011) Developing a quality standard for fingerprint examination. Fingerprint Quality Standards Specialist Group, United Kingdom, 20 Dec 2011

Giannelli PC (1997) The abuse of scientific evidence in criminal cases: the need for independent crime laboratories. Virginia J Soc Policy Law 4:439–478

Grann D (2009) Trial by fire: did Texas execute an innocent man? The New Yorker

Hagan W (1894) A treatise on disputed handwriting and the determination of genuine from forged signatures. Banks & Brothers, New York

Langenburg G, Champod C, Wertheim P (2009) Testing for potential contextual bias effects during the verification stage of the ace-v methodology when conducting fingerprint comparisons. J Forensic Sci 54(3):571–582

Miller L (1984) Bias among forensic document examiners: a need for procedural change. J Police Sci Adm 12:407–410

Miller L (1987) Procedural bias in forensic examination of hair. Law Hum Behav 11(2):157–163

Mower L, Mcmurdo D (2011) Las Vegas police reveal DNA error put wrong man in prison. Las Vegas Rev-J

National Academy of Sciences (2009) Strengthening forensic science in the United States: a path forward. NAS, Washington, DC

National Institute of Standards and Technology (2012) Expert working group on human factors in latent print analysis. Latent print examination and human factors: Improving the practice through a systems approach. U.S. Department of Commerce, NIST, Washington, DC

Office of the Inspector General (2006) A review of the FBI's handling of the Brandon Mayfield case. U.S. Department of Justice Office of the Inspector General, Washington, DC

Page M, Taylor J, Blenkin M (in press) Context effects and observer bias—implications for forensic odontology. J Forensic Sci. doi:10.1111/j.1556-4029.2011.01903.x

Risinger M, Saks M, Thompson W, Rosenthal R (2002) The Daubert/Kumho implications of observer effects in forensic science: hidden problems of expectation and suggestion. Calif Law Rev 90:1–56

Schwaninger A (2006) Threat image projection: enhancing performance? Aviation Security International, December, 36–41

Stoel RD, Sjerps M (2012) Interpretation of forensic evidence. In: Roeser S, Hillerbrand R, Sandin P, Peterson M (eds) Handbook of risk theory: epistemology, decision theory, ethics, and social implications of risk. Springer, New York, pp 135–158

Thompson W (2010) What role should investigative facts play in the evaluation of scientific evidence? Aust J Forensic Sci 43(2–3):123–134. doi:10.1080/00450618.2010.541499

Thornton JI (2010) Letter to the editor—a rejection of "working blind" as a cure for contextual bias. J Forensic Sci 55(6):1663

Ulery BT, Hicklin RA, Buscaglia J, Roberts MA (2012) Repeatability and reproducibility of decisions by latent fingerprint examiners. PLoS One 7(3):e32800

Cognitive Transformation

► Examining the Effectiveness of Correctional Interventions

Cognitive/Information Processing Theories of Aggression and Crime

L. Rowell Huesmann
Institute for Social Research, University of Michigan, Ann Arbor, MI, USA

Overview

Over the past several decades, it has become clear that an understanding of how human cognitive processes operate is necessary for an understanding of human social behavior in general and aggressive and criminal behavior in particular. Extensive empirical research on cognitive/information processing coupled with theoretical elaborations from cognitive science constructs has led to the emergence of a unified model of cognitive/information processing in aggressive behavior. The model identifies three processes in social problem solving during which emotional arousal, activated schemas, and situational cues interact to make aggressive and criminal behavior less or more likely : (1) cue attention and interpretation, (2) script retrieval, and (3) script evaluation and selection. In addition, two processes have been identified that change stored cognitive schemas, scripts, and beliefs to increase or decrease the long-term likelihood of aggressive and criminal behavior: (1) observational learning through imitation, inference, and emotional desensitization, and (2) conditioning that increases or decreases the accessibility of encoded cognitions. While these processes may first require cognitive control in the developing child, they eventually seem to operate as relatively automatic cognitive processes.

People attend to environmental cues differentially and interpret the cues differently as a function of predisposing neurophysiological factors, their emotional arousal, the kinds of cognitive schemas they have acquired, and which schemas are activated. More aggressive individuals tend to focus on fewer cues and cues that are more frequently symptomatic of hostility, tend to interpret ambiguous cues more readily as symptomatic of hostility, and tend to believe that the world is more hostile. This is particularly true when the individual is angry, either because of situational factors or a predisposition toward more general hostility. More aggressive individuals also have a greater proportion of aggressive scripts encoded in memory with more accessible links to everyday cues. They have been found to rehearse more often their aggressive scripts through aggressive fantasizing and to recall more often aggressive scripts from ambiguous cues. It has been shown that, while young children do not have well-defined or stable normative beliefs about the appropriateness of aggression, older children do have well-formed beliefs, and those beliefs influence how they evaluate retrieved scripts. Finally, aggressive individuals often do not expect their own aggressive behaviors to have bad consequences for them.

From the social/cognitive perspective, once a youth begins to perceive the world as hostile or antisocial, begins to acquire scripts and schemas emphasizing aggression and antisocial behaviors, and begins to believe that aggression and law-breaking is acceptable, the youth enters a vicious cycle that is difficult to stop. Cognitions, behavior, observations of others, and the responses of others, all combine to promote aggression and criminality. If not interrupted, the cycle can be expected to continue into adulthood, maintaining aggressive and antisocial behavior throughout the life span.

Introduction

One of the essential ways in which humans are different from all other species is in their well-developed ability to symbolically represent, process, and communicate information. The psychological processes that humans invoke to perform these tasks are called *cognitive processes*, and the internal representations of information utilized in these processes are denoted as *cognitions*. Some reflexive behaviors may usually involve only peripheral processing, but more central cognitive processes often override these reflexes, e.g., one can will oneself to keep one’s hand on a hot coal. From this perspective, all human social behavior, including aggressive and criminal behavior, is mediated by the cognitions and cognitive processing of the participants. This does not mean that cognitive processes “cause” social behavior. Rather as mediating processes they connect biological, environmental, and situational inputs to behavioral outputs. However, different patterns of cognitive processing are more conducive to one kind of social behavior than to another.

An information-processing model of social behavior is a description of the cognitive data structures a person utilizes and the sequence of cognitive operations the person uses in order to process whatever information is encountered by the person in order to decide on a social behavior. It is most analogous to a computer program that describes the output a computer produces from given input. Cognitive/information-processing theories of aggression and criminal behavior explain the occurrence of aggression and crime and individual differences in the risk for occurrence in terms of individual differences in these cognitive data structures and the associated information-processing operations. The cognitions and information-processing operations are not viewed as ultimate causes of crime or aggression but rather as what connects innate predisposing causal factors and past and current environmental causal influences to current behavior, whether antisocial or prosocial.

History of Cognitive/Information-Processing Models of Social Behavior

The use of cognitive/information-processing models to describe human behavior first emerged in the late 1950s in studies of human problem solving (Newell and Simon 1972; Simon 1969) probably because the expertise in computer programming that such modeling requires tended to be concentrated in that area. However, it was quickly recognized that the formalization and detailed specification of processes required in information-processing models make them more valuable than either natural language or mathematical statistics for modeling any kind of human behavior. In the early 1960s, two books – Feigenbaum's (1963) *Computers and Thought*, and Miller, Galanter, and Pribram's (1960), *Plans and the Structure of Behavior* – stimulated thinking about information processing in many areas of psychology, and by the mid-1960s Abelson (1968), Gullahorn & Gullahorn, Loehlin, Simon (1969), and others had begun to describe psychological models of a variety of social behaviors in terms of underlying cognitive structures and information processes. By the 1970s and early 1980s, more specific information-processing theories of different kinds of social and abnormal behavior were appearing (Carroll and Payne 1976; Huesmann 1982). The term "social cognition" became popular sometime in the late 1970s (Wyer and Carlston 1979), and with the publication of more often Fiske and Taylor's (1984) book on social cognition, cognitive/information-processing theorizing became part of the mainstream of thinking about social behavior.

The early 1980s also brought the first cognitive/information-processing analyses of aggression. Huesmann (1982, 1998; Huesmann and Eron 1984), building both on Bandura's ideas and Newell and Simon's (1972) thinking, offered a general information-processing model for aggression focusing particularly on observational learning, and Dodge (1993; Crick and Dodge 1994) offered a general information-processing model for aggression focusing particularly on perceptions and attributions. Shortly afterward, Slaby and Guerra (1988), building on Shure and Spivac's analysis of adolescents' adjustment problems as failures in social problem solving, expanded these ideas to analyze delinquent's criminal behaviors in terms of underlying cognitive/information-processing abnormalities. Since then, a variety of elaborations of the models have emerged (e.g., Anderson and Bushman 2002) and a reasonable consensus has developed about the ways in which human cognitions and cognitive processes mediate aggressive and criminal behavior. But before turning to a more detailed discussion of those models, we must review a number of general principles of information processing.

Fundamentals of Cognitive/Information-Processing Models

Cognitive/information-processing models formally lay out the sequence of cognitive processes involved in the occurrence of a behavior much as a computer program lays out the sequences of operations in a computer that produces a particular output. A basic assumption is that the human mind can be viewed as analogous to a computer. Behavior is the output of software programs operating within the biological hardware constraints of the brain. A closely related assumption is that behavior is best modeled with hierarchical levels of explanation (Simon 1969). For example, an aggressive interaction might first be modeled in terms of relevant interacting behavior sequences (e.g., she shouted at him; he told her to shut up; she shouted again; he hit her). The behavior sequence might then be modeled in terms of a higher level program for behavior (e.g., follow rule to retaliate when insulted by female). The operation of such programs might then be modeled by more fundamental cognitive processes, e.g., retrieval of rules by spreading activation; these more fundamental processes might be modeled in terms of neurophysiological reactions which in turn might be modeled in terms of biochemical processes. Different theories of social behavior may use different levels of explanation within, but generally most theories adopt a level analogous to programming in a high level computer language. The formal properties and limitations these programming languages and symbol systems place upon the modeled processes are well understood, and the natural language interpretations of the programs are also usually easily followed by other researchers.

Cognitive/information-processing theories of social behavior assume that a person's social behavior is completely determined by the configuration of hardware, software, knowledge structures, and environmental inputs that the organism has experienced. However, because all these elements are never completely known, behavior can never be perfectly predicted. Specific models sometimes include stochastic elements to model individual variations, but, even without them, complex variations in behaviors and cognitions are obtained because of the wide diversity of experiences encountered by different individuals. Cognitive/information-processing models are models of individuals' social behaviors, not of group means. Different individuals have different hardware, software, and data structures. Different situations generate different inputs. Most often models are developed that assume similar programs and structures for different individuals with well-defined loci for individual differences, e.g., perhaps the procedure for deciding whether or not to attribute hostility to another is the same for most individuals, but the database of past experiences on which they base their decision varies.

Every cognitive/information-processing model has an *executive program* (i.e., operating system) that specifies the overall flow of cognitive processes that are hypothesized, *subroutines* that represent the detailed operation of specific cognitive processes that are available, and multiple data structures in *memory* that represent the modeled individual's cognitions. Some cognitive/information-processing models may include mechanisms for self-modification of the cognitions through a variety of learning processes.

Connectionist Alternatives

One limitation of cognitive/information-processing models is that the relations between the basic information processes and the underlying neurophysiological processes of humans remain obscure. A second related problem is that cognitive/information-processing models most often hypothesize multiple programs that operate sequentially instead of in parallel (probably because until recently the machine languages available to simulate cognitive processes operated sequentially). However, there is significant evidence that many human cognitive processes occur in parallel.

One consequence of these problems has been the emergence of what have come to be called "connectionist" models of cognitive/information processing (Feldman and Ballard 1982). Such models attempt to explain social behaviors using simulated neural networks in which parallel processing is a natural property. The ability of such simulated neural networks to accurately model anything more than the most rudimentary social behaviors remains to be demonstrated. An alternative approach to incorporating parallel processing into cognitive/information-processing models has been to simulate parallel processing with rapid sequential processing. This approach is derived from the empirical finding that it is very difficult to tell if particular information processes in the human brain are operating in parallel or in rapid sequential order (Simon 1969). Information-processing theorists point out that it is important not to confuse "sequential" processing with "linear" models. Nonlinear relations can be modeled quite adequately with sequential information-processing models.

Basic Processes and Data Structures in Cognitive/Information-Processing Theorizing

Cognitive/information-processing models of social behavior have drawn on empirical knowledge about human cognition and human social behavior to define a set of basic processes and data structures that seem to explain different kinds of human social behavior. As mentioned earlier, one can think of any cognitive/information-processing system as consisting of a *memory* in which *data and programs* can be stored and an *executive program* which distributes resources and has overall control over the system. In controlling social behavior, this cognitive/information-processing system accepts inputs of social stimuli that define a particular social situation. The executive program calls upon appropriate subroutines to process the inputs, search memory for relevant information, and generate potential social behaviors for that situation. If a social behavior passes a final test of appropriateness, then it is performed (Anderson 1983).

Memory Structures

Conceptually, human memory consists of a network of nodes and links that represent encoded propositions. The meaning of each node is defined by its associated links which represent labeled attributes whose values are other nodes. Real-world meaning is ultimately given to the symbolic encodings by the connections of the network nodes to representations of external stimuli – visual, auditory, etc., but more immediate meaning is given by the connections to other symbolic nodes. *Elaborative rehearsal* of information generates more links to the rest of the network and more firmly encodes information in the network. Elaborative rehearsal goes beyond simple repetition of information and involves consideration of how the new information fits with other knowledge already encoded.

Memories are *retrieved* by being *activated*. One can imagine a *spread* of activation emanating out from one node to connected nodes. A cue, either an external stimulus or an internally activated schema or mood, activates the first node. As each successive node becomes activated, it is said to be retrieved, but the strength of activation diminishes as greater distance separates linked nodes. Clearly multiple links enhance the likelihood of recall although multiple nodes with similar links may produce inaccurate recall due to *interference*. However, retrieval of information generally is not a blind spreading activation process but rather a more directed *heuristic search*. Particular branches of the network are followed that appear particularly promising and the activated nodes are tested to see if they meet a criterion for what is being sought. Spreading activation occurs automatically at a subconscious level. Activated nodes can emerge into consciousness but can also exert influence while remaining subconscious. When the activation of a node is increased by an external stimulus or a related thought, we say that the node has been *primed*. Within this system, there is a conceptually distinct *short-term or working memory* where information being actively processed resides but from which information decays rapidly. If information is activated long enough in short-term memory, it is likely to be integrated into the unlimited-capacity *long-term memory*. Thus, attention to social information (i.e., maintaining it in short-term memory) can have a direct effect on the likelihood that it will be encoded in long-term memory.

Within this framework, *schema* is a term used to refer to any cognitive knowledge structure encoded in memory that represents substantial knowledge about a concept, its attributes, and its relations to other concepts. Each individual's memory contains "self-schemas," which are organized knowledge about themselves, "world schemas," which are organized knowledge about the world, "event schemas" which are organized knowledge about relations between events and behaviors, "belief-schemas," which are organized sets of beliefs, and so on.

Information-processing theorists have shown that a very few basic information processes operating on such schemas and external inputs can account for very complex behaviors. Among the most basic processes are the *generate-and-test* process (sometimes denoted TOTE – test, operate, test, execute) which involves retrieving information packets repeatedly from memory until one finds one that passes a logical test. Another is the *test and branch* operation which involves changing the sequence of processing operations based on a logical test. Still, another is the *linking or assignment* process which involves connecting two nodes in memory – for example, connecting a node representing an object to an emotion such as anger. Individual differences in the ability to control and sequence these basic operations are common and are often referred to as individual differences in *executive functioning*. A variety of studies including f-MRI studies have suggested that executive functioning is centered in the prefrontal cortex and may be decremented by damage to the prefrontal cortex.

Retrieving Social Information

The ease with which any kind of information or response can be accessed in the human central nervous system depends on how elaborately it has been encoded. Even simple S-R pairings are more likely to be elicited if there are multiple paths from stimulus to response. As described above, activation of social information in memory can be viewed as spreading activation processes with the spread of activation often being guided by already encoded schemas (heuristic search). For example, if you see someone you know as a librarian walking down the street, your mental search for explanations for what they are doing will be influenced by your schema for librarians.

Such memory search is activated by immediate *social cues* which provide the input to begin the search, e.g., the librarian in the above example. However, the speed and success of retrieving social information is also influenced by recent stimuli that may have activated relevant schemas in memory, making them more accessible. As mentioned above, this process is known as *priming*. For example, you may be more likely to recognize the librarian on the street if you have just previously walked by the library and had the library schema activated. *Emotions and moods*, such as anger or depression, also serve as cues that activate related schemas or scripts.

The activation of a particular schema or node in memory has automatic effects on social inference and behavior, but another process called *filtering* may limit the effect of any activated information. Filtering is the testing process through which the executive routine decides if the activated information is the information sought or whether the search should continue. This is an example of the basic generate-and-test information-processing operation described above. Filters may be simple tests of accuracy for the information or they may be tests of appropriateness of the scripts or behaviors activated. The cognitive schemas that guide specific filtering for social behavior have been denoted as self-regulating internal standards by Bandura (1986) or normative beliefs by Huesmann and Guerra (1997).

Automatic and Controlled Cognitive Processing

Studies of cognitive processing have revealed that at least two different modes of processing exist with quite different speeds and demands on conscious resources (Schneider and Shiffrin 1977). *Automatic processes* occur very rapidly, without using many cognitive resources, and without any conscious executive decisions being made about the process. Reading is one example. Many social-perception processes are automatic including "schema-triggered affect" (Fiske and Taylor 1984), memory priming effects, and spontaneous trait inferences (Bargh 1982). *Controlled processes* occur more slowly, require more cognitive resources, and demand conscious executive control. Conscious planning of how to deal with a social situation would be one example. The boundaries are sometimes fuzzy, and many controlled processes may become automatic either as part of a child's development or with repeated practice or rehearsal. For example, in a novel social situation, a person may engage in a controlled search for scripts that could be used in that situation. But highly familiar situations quite probably automatically trigger scripts that fit the situation.

Stress, Mood, Arousal, and Information Processing

The level of arousal that the human information processor is experiencing seems to change information processing in a number of ways. As arousal levels become higher than normal, attention seems to be directed more narrowly at a few cues that seem to be the most salient. Very high levels of arousal seem to decrease working memory capacity, narrow memory search, make activation of weakly associated schemas less likely, and make the activation of the best connected schemas and scripts most likely. For simple tasks, this restriction can produce enhanced performance which may provide an evolutionary explanation for the phenomenon; however, for complex tasks requiring extensive heuristic search for solutions, performance will worsen. Of course, stress is one of the most common causes of high arousal. As a result, in stressful social situations, one can expect that a person will focus on what seem to be only the most salient social cues, activate only the schemas and scripts that are most closely connected to those cues, and generate a narrow range of behaviors. Thus, when stress is coupled with emotions like fear or anger and coupled with situational cues linked to aggressive scripts, only violent schemas are likely to be activated. For example, a young male whose life is being threatened by another young male with a gun may find it very difficult to attend to any other cues in the environment besides the gun, or to retrieve any other scripts than simple scripts for fighting or fleeing.

The Cognitive/Information-Processing Theory of Aggressive and Criminal Behavior

The accumulated knowledge about human cognitive/information processing and social cognition as outlined above provides important insights into the processes involved in many social behaviors including crime. But it is important to realize that social cognition is NOT a cause of aggressive or criminal behavior. Social cognition is a *mediating* process that connects external situations, internal schemas, and social behavior in predictable ways. More generally, the existing research suggests that habitual aggressive and antisocial behavior is a product of *innate predisposing cognitive and emotional factors*, interacting with *exposure to an aggression-promoting social environment* that encourages the development of social cognitions promoting aggression, interacting *with situationally specific precipitating factors*. From a social cognitive/information-processing perspective, the variety of predisposing factors discussed above may make the emergence of certain specific cognitive routines, scripts, and schemas more likely, but these cognitions develop through interactions of the child with the environment and are designed to respond to different environmental situations. In early childhood, these cognitions promote aggressive behavior below the delinquent and criminal level and later promote criminal and violent behavior. Consequently, child aggression predicts adult criminality (Huesmann et al. 2002).

Slightly different cognitive/information-processing explanations for the occurrence of antisocial behavior have been formulated by Huesmann and colleagues (Huesmann 1982, 1998; Huesmann and Eron 1984; Huesmann and Guerra 1997), by Dodge and his colleagues (1993; Crick and Dodge 1994), and by Anderson and colleagues(Anderson and Bushman 2002). However, while they have had different emphases, all draw on Bandura's (1986) social/cognitive theorizing emphasizing self-regulating internal cognitions and draw on Berkowitz's (1990) neo-associationist ideas. Furthermore, all agree that there is a common set or core social cognitions and information-processing operations that explain how certain predisposing factors, socializing environments, and precipitating stimuli lead to aggressive and antisocial behavior. Here, one overall summary model is presented that combines the common cognitive/information-processing elements that have been theorized to explain aggressive and criminal behavior.

This model incorporates the premise that social behavior is controlled to a great extent by cognitive *scripts* (Abelson 1981) that are stored in a person's memory and are used as guides for behavior and social problem solving. As described above, a script incorporates both procedural and declarative knowledge and suggests what events are to happen in the environment, how the person should behave in response to these events, and what the likely outcome of those behaviors would be. It is presumed that while scripts are first being established, they influence the child's behavior through "controlled" mental processes (Schneider and Shiffrin 1977), but these processes become "automatic" as the child matures. Scripts that persist in a child's repertoire become increasingly more resistant to modification and change as they are rehearsed, enacted, and generate consequences. More aggressive and criminal individuals are presumed by this theory to have encoded a much higher proportion of social scripts utilizing aggressive and criminal actions.

World schemas stored in a person's memory are the second kind of cognitive schema that plays an important role. World schemas are the database that individuals use to make attributions about others' behaviors. Individuals with schemas representing the world as a more hostile place are more likely to attribute hostility to the behaviors of others and to behave more aggressively. *Normative beliefs* are the third kind of cognitive schema hypothesized to play a central role in regulating aggressive behavior.

Normative beliefs are cognitions about the appropriateness of social behaviors. They are related to perceived social norms but are different in that they concern what is "right for you." Normative beliefs are part of one's self-schema and serve as part of the database used to filter potential scripts that are retrieved from memory. A social script that suggests an aggressive behavior may not be employed if it violates the individual's normative beliefs.

Cognitive/Information Processing Proximal to a Social Behavior

These internalized cognitive schemas are utilized in a specified sequence of information-processing steps at the moment a social problem is encountered to determine how the individual responds. First, individuals attend to some of the environmental cues and evaluate the cues to which they attend on the basis of their world schemas and their own internalized emotional states. Then they retrieve social scripts that might seem to fit the situation and with a heuristically guided generate-and-test process. Each retrieved script is then evaluated not only on the basis of its projected objective outcome but also on the basis of its consistency with the individual's normative beliefs. Thus, at the moment when a social problem is encountered, there are three main processing loci at which individual differences in internalized cognitions and environmental stimuli can influence behavior: (1) cue attention and evaluation, (2) script retrieval, and (3) script evaluation and selection.

Cue Attention and Evaluation. The objective situation for any social problem is defined by the social problem and the environmental cues. However, which cues are given most attention and the interpretation of those cues may vary from person to person and may depend on a person's neurophysiological predispositions, current emotional state, and previous learning history as reflected in their activated world schema. Because emotional states may persist for some time, a person may enter a social interaction in an emotional state that is unrelated to the current situational cues. For example, a person exposed repeatedly to frustrating situations, who attributes the goal-blocking to the actions of others, may enter a social interaction in an aroused state with hostile feelings toward everyone. Independent of one's current emotional state, environmental stimuli cue the retrieval from memory of emotions and cognitions that have been associated with that cue in the past. For example, the “sight of an enemy” or the “smell of a battlefield” may provoke both instantaneous physiological arousal and the recall of thoughts about the “enemy” that give meaning to the aroused state as anger. That emotional state may influence both which additional cues the person attends to and how the person interprets the cues to which he or she does attend. A highly aroused, angry person may focus on just a few highly salient cues and ignore others that convey equally important information about the social situation. Then the angry person’s evaluation of these cues may be biased toward perceiving hostility when none is present. A person who finds hostile cues the most salient or who interprets ambiguous cues as hostile will be more likely to experience anger and activate schemas and scripts related to aggression.

Script activation and retrieval. It is presumed that more aggressive and antisocial individuals have encoded in memory more extensive, well-connected networks of social scripts emphasizing aggressive and antisocial problem solving. Therefore, such a script is more likely to be retrieved during any search. However, the search for a script is also strongly affected by one’s interpretation of the social cues, one’s activated schemas, and one’s mood state and arousal. For example, anger, even in the absence of supporting cues, will make the retrieval of scripts previously associated with anger more likely; the presence of a weapon, even in the absence of anger, will make the retrieval of scripts associated with weapons more likely; and the perception that another person has hostile intentions will activate scripts related to hostility.

Evaluation of scripts and script selection. The third locus for the expression of individual differences and situational variation occurs after a script is activated. Before acting out the script, one evaluates the script in light of internalized activated schemas and normative beliefs to determine if the suggested behaviors are socially appropriate and possible to do and if the expected outcome is likely to be desirable. People with different normative beliefs may thus evaluate the appropriateness of a script quite differently, and the same person may evaluate the same script differently at different times.

On the average, however, the habitually aggressive are expected to hold normative beliefs condoning more aggression and thus will employ more aggressive scripts. For example, if a man suddenly discovers that his wife has been unfaithful, he may experience rage and access a script for physical retribution. However, whether or not the man executes that script will depend on his normative beliefs about the appropriateness of “hitting a female.” More general normative beliefs may also be relevant. The man who believes in “an eye for an eye” and perceives himself as “an avenger” is more likely to retrieve a script emphasizing aggressive retaliation. Even within the same person, different normative beliefs may be activated in different situations and different mood states. The person who has just been to church may have activated quite different normative beliefs than the person who has just watched a fight in a hockey game on TV.

Scripts predict likely outcomes, but people also differ in their capacities to think about the future, in their concern with the future, and in their evaluation of the desirability of an outcome. The more a person focuses on immediate consequences and the less the person is concerned with the future, the more palatable a self-centered solution to a social problem may seem. For example, if the man above who hit his spouse was concerned about what his in-laws will think of him if they hear about it in a few days, he might not have hit her. Of course, people may also misperceive the likely consequences of aggressive acts simply because their scripts are inaccurate in predicting consequences for the current situation. Additionally, people differ in their evaluations of the desirability of an outcome. For some people, the respect of their peers may be the most important evaluative dimension for an outcome from a social conflict while for others, simply ending the conflict may be most important. For other people, the thought of violence may be so anxiety provoking that aggressive scripts are filtered out. However, as described later, those who have been exposed to violence repeatedly and have become emotionally desensitized to violence experience little negative affect when evaluating aggressive scripts and, thus, are more likely to use them.

Emotional arousal, schema activation, script generation, and evaluation. Cognitive/information-processing models of social behavior are sometimes thought of as “cold” models with little room for affect. However, the unified cognitive/information-processing model for aggression described here hypothesizes a key role for emotion at each stage of cognitive/information processing. Both an individual’s absolute arousal level and the valence of the arousal affect aggressiveness at each stage. Existing negative affect interpreted as anger will bias cue interpretation toward hostility, will prime the retrieval of aggressive scripts, and will cue normative beliefs more supporting of aggression. As described earlier, in high states of arousal, individuals search less widely and deeply for scripts and retrieve the best connected scripts; so aggressive behavior becomes even more likely in highly arousing situations for persons with predominately aggressive scripts. For the individual with a large well-learned repertoire of simple, direct aggressive scripts for solving social problems and a smaller less-well-learned repertoire of complex, indirect scripts for prosocial solutions, the arousal associated with anger will make the selection of an aggressive script even more likely than priming from the anger would by itself. Furthermore, one can expect similar effects for script evaluation. A person in a highly aroused state of rage because of a provocation will evaluate scripts less carefully and focus on retaliation beliefs in evaluating scripts because his/her rage has primed such normative beliefs.

Cognitive/Information Processing for Acquiring and Maintaining Aggressive Scripts and Schemas

According to social cognitive/information-processing theory, a habitually aggressive or criminal person is presumed to be someone who regularly retrieves and employs aggressive or criminal scripts for social behavior. The theory allows a number of factors that might promote the retrieval and utilization of aggressive and criminal scripts. It may be, for example, that the cues present in the environment trigger the recall only of aggressive scripts. However, the regular retrieval and use of aggressive scripts would suggest above all that a large number of aggressive scripts have been stored in memory. Similarly, the regular execution of such scripts would suggest that normative beliefs and other schemas supporting aggression have been acquired and encoded. How does this happen?

While a variety of preexisting neurophysiological factors can predispose individuals toward particular modes of cognitive processing or toward particular schemas and scripts, cognitive/information-processing theory asserts that the child’s early learning experiences play a critical role in the acquisition of scripts and schemas for social behavior just as early learning plays the key role in the acquisition of procedural and declarative knowledge relevant to intellectual life. Evolutionary forces and random variation create communalities and individual differences in the biological mechanisms underlying cognition and behavior. These individual differences in biology interact with individual differences in environment to mold the schema encoded in memory into those that promote nonaggressive prosocial behavior or those that promote aggressive antisocial behavior. Both enactive learning processes (i.e., conditioning, elaborative rehearsal) and observational learning processes (i.e., imitation, desensitization) are involved in the encoding of these social cognitive schema.

As Huesmann and Kirwil (2007) have described, social scripts, world schemas, and normative beliefs are all initially acquired through imitation and inference (observational learning). Imitation is an innate process in humans. It is established that human infants begin mimicking behavior when they are only a few months old but advance in early childhood to imitating entire scripts and inferring world schemas and normative beliefs from what they see others doing. Imitation is not a low level cognitive process but rather a complex high level process in which the scripts, schemas, and beliefs detected to be held by others are encoded into one’s own memory.

During the observational learning process, the person's current emotional state and current memory contents influence which existing schemas are activated. The activated schemas in turn influence how well the observed schemas can be encoded and integrated into memory. If the activated schemas are discrepant with the observed schemas, encoding is difficult; if they are consistent, it is easier. When highly aroused and angry, for example, persons may view a physically aggressive or antisocial script as more appropriate than they would otherwise. A young boy who can only recall seeing aggressive behaviors is more likely to encode a newly observed aggressive behavior than is a boy whose mind is filled with memories of prosocial solutions. A child with normative beliefs accepting of aggression is much more likely to encode new aggressive scripts for behavior. Consequently, violent and criminal behaviors become contagious. Violent and criminal behaviors tend to beget other violent and criminal behaviors in those who observe them.

Once it is encoded, the accessibility of a script in memory is influenced by the extent to which its use produces desired consequences, i.e., by instrumental learning. One might think that, because aggressive or criminal behavior very often produces negative consequences for the aggressor, the retrieval of aggressive scripts would extinguish. However, instrumental learning depends as much on how the individual interprets society's response to the behavior as it does on the response. As described above, very often, because of the schemas activated, the aggressor does not attribute the negative reaction of society to the specific script that the aggressor employed, and no learning takes place. The boy who is harshly punished by a teacher for pushing another child out of line will not unlearn the behavior if he interprets the cause of his punishment as dislike by the teacher.

As with all learned information, for a script or schema or normative belief to stay highly accessible, it needs to be rehearsed from time to time. The rehearsal may take several different forms, from simple recall of the original scene, to fantasizing about it, to play acting. The more elaborative, ruminative type of rehearsal characteristic of children's fantasizing is likely to generate greater connectedness for the script, thereby increasing its accessibility in memory. Also, through such elaborative rehearsal, the child may abstract higher-order scripts representing more general strategies for behavior than the ones initially stored. Of course, rehearsal also provides another opportunity for reevaluation of any script. It may be that some scripts initially accepted as appropriate (under specific emotional and memory states) may be judged as inappropriate during rehearsal.

Empirical Data on Cognitive/ Information processing and Aggression

Empirical Data on Cue Attention and Evaluation

Extensive research shows that more aggressive individuals tend to perceive hostility in others where there is no hostility, i.e., display a hostile attributional bias (Dodge 1993). Longitudinal analyses have suggested that a propensity toward hostile attributional bias mediates the relation between early childhood aggression and later antisocial behavior. Furthermore, recent experiments with random assignment of subjects have shown that improving the accuracy of intent attributions decreases the likelihood of aggressive behavior in children.

Empirical Data on Script and Schema Acquisition and Activation

While it has been methodologically more difficult to assess the kinds of scripts that individuals have encoded, one can assess the kinds of scripts they are most likely to retrieve and make inferences from those data. The available evidence suggests that, in fact, the more accessible scripts for more aggressive children are more aggressive. The scripts retrieved by more aggressive children to solve hypothetical problems tend to incorporate more physical aggression and manipulation actions. Priming by negative intent cues is more likely to activate an aggressive script in aggressive children. Aggressive children are less likely to generate more subtle prosocial scripts to solve social problems and there is some evidence that, as hypothesized, a narrower search process for a script is associated with more aggressive behavior.

There is also extensive evidence that the observation of aggressive or criminal scripts in real life or in the mass visual media leads to the encoding of such scripts. Children growing up observing violence around them in their neighborhood behave more violently (Guerra et al. 2003) and those observing such violence frequently in the mass media behave more criminally even 15 or 22 years later (Huesmann et al. 2003; Huesmann and Eron 1992). Numerous laboratory and field experiments and everyday observations of "copycat" crimes have demonstrated the encoding of specific scripts from observations of violence and crime. It is also clear that new aggressive scripts are abstracted out of the elements of specific scripts being observed. Thus, the aggressive scripts that children display after being exposed to violent scenes are not exactly the same as the scripts observed. As cognitive/information-processing theory predicted, rehearsal of antisocial scripts through fantasizing about them has also been shown to increase their accessibility and likelihood of use (Huesmann and Eron 1984, 1992).

There is also significant empirical support for the premise that even quite different specific aggressive scripts and schemas are linked together in one's memory network by a common "hostility" node and thus can be primed by other aggressive ideas or cues, even if they have no substantive connection. For example, Berkowitz and LePage (1967) showed that even the brief sight of a "gun" increases the likelihood that a provoked person will behave more aggressively in ways that have nothing to do with guns.

Empirical Data on Evaluation of Scripts

Once a script has been activated, it still may not be employed if it is evaluated negatively. The behaviors involved may be evaluated as inappropriate when filtered through an individual's normative beliefs about aggressive and antisocial or criminal behavior. Its expected outcome may be evaluated as undesirable when environmental responses are considered; or it may be evaluated as undoable when filtered through conceptions of self-efficacy. A variety of studies have shown such filtering effects. Huesmann and Guerra (1997) showed that normative beliefs that aggression is unacceptable reduce the likelihood of a youth behaving aggressively as does a priming of self-schemas that the person is nonaggressive. On the other hand, Baumeister and colleagues have demonstrated that a self-schema that includes an extremely positive evaluation of oneself (a narcissistic evaluation) promotes the selection of aggressive scripts when a person threatens that self-evaluation. Not only are self-schemas relevant to script evaluation, schemas about others are relevant too. For example, schemas about others which promote deindividuation allow the utilization of aggressive scripts which might otherwise be unacceptable.

The evaluation process for a generated script also includes an assessment of outcome "desirability." In fact, there is accumulating evidence that more aggressive children tend to believe that aggressive behavior will have more desirable outcomes, and prosocial behavior will have less favorable outcomes (Crick and Dodge 1994). Desirability also appears to be evaluated in terms of how the anticipated behavior and outcome "feel" to the person. If thinking about behaving criminally or violently makes one feel bad, one is less likely to behave that way. Repeated exposure to others behaving violently or criminally emotionally desensitizes people so they do not feel so bad when they think about it (Huesmann and Kirwil 2007). Thus, it becomes easier for them to behave criminally or violently.

The above studies have shown that aggressive youth differ from other youth in the beliefs and schemas they use to evaluate the appropriateness and effectiveness of potential scripts. However, there is also evidence that some youth are more aggressive because they do not devote much cognitive effort to this filtering step. For example, empirical data suggest that certain children, e.g., ADHD children, are neurophysiologically predisposed toward minimizing this evaluation step. The result is what many observers would call impulsive behavior (Kendall and Braswell 1985). If the first scripts activated are aggressive or criminal scripts, the result is aggressive or ciminal behavior even if the child "knows" upon reflection that the behavior was inappropriate.

Empirical Data on Gender Differences in Aggression and Crime

Males are much more likely to behave physically aggressively and much more likely to commit crimes. Some of these differences can be explained by differences in encoded cognitive schemas and differences in social cognitive development (Bennett et al. 2006). While many people hold normative beliefs that aggression by males is appropriate in some situations, that is not so true for females. Additionally, females acquire social cognitive skills earlier in life than males and consequently have better elaborated prosocial scripts for social problem solving. Such scripts may also work better for females because of females' superior (on average) verbal ability. Finally, differential socialization by society which disapproves of aggression in females leads to lower accessibility of aggressive scripts for females.

Related Entries

- ▶ Career Criminals and Criminological Theory
- ▶ Cognitive Forensics: Human Cognition, Contextual Information, and Bias
- ▶ Economic Theory of Criminal Behavior
- ▶ Empathy and Offending
- ▶ Evolutionary Theories of Criminal Behaviors
- ▶ Fear of Crime and the Psychology of Risk
- ▶ Genes, Crime, and Antisocial Behaviors
- ▶ Hormones, the Brain, and Criminality
- ▶ Juvenile Violence
- ▶ Longitudinal Studies in Criminology
- ▶ Neurology and Neurochemistry of Crime
- ▶ Pathways to Delinquency
- ▶ Psychopathy and Antisocial Personality Disorder
- ▶ Rational Choice and Prospect theory
- ▶ Social Control and Self-Control Through the Life Course
- ▶ Social Learning Theory
- ▶ Synthesizing Biological and Social Theorizing
- ▶ Theories of Punishment

Recommended Reading and References

Abelson RP (1968) Computer simulation of behavior. In: Lindzey G, Aronson E (eds) Handbook of social psychology, vol II. Addison-Wesley, Reading, MA

Abelson RP (1981) Psychological status of the script concept. American Psychologist, 36:715–729

Anderson JR (1983) The architecture of cognition. Harvard University Press, Cambridge, MA

Anderson CA, Bushman BJ (2002) Human aggression. Annu Rev Psychol 53:27–51

Bandura A (1986) Social foundations of thought and action: a social-cognitive theory. Prentice-Hall, Englewood Cliffs

Bargh JA (1982) Attention and automaticity in the processing of self-relevant information. J Personal Soc Psychol 43(3):425–436

Bennett S, Farrington DP, Huesmann LR (2006) Explaining gender differences in crime and violence: the importance of social cognitive skills. Aggress Violent Behav 10(3):263–288

Berkowitz L (1990) On the formation and regulation of anger and aggression: a cognitive-neoassociationistic analysis. Am Psychol 45(4):494–503

Berkowitz L, LePage A (1967) Weapons as aggression-eliciting stimuli. J Pers Soc Psychol 7:202–207

Carroll JS, Payne JW (1976) Cognitive and social behavior. Erlbaum, Hillsdale

Crick NR, Dodge KA (1994) A review and reformulation of social information processing mechanisms in children's adjustment. Psychol Bull 115:74–101

Dodge KA (1993) Social cognitive mechanisms in the development of conduct disorder and depression. Annu Rev Psychol 44:559–584

Feigenbaum EA, Feldman J (1963) Computers and thought. McGraw-Hill, New York

Feldman JA, Ballard DH (1982) Connectionist models and their properties. Cognit Sci 6:205–254

Fiske ST, Taylor SE (1984) Social cognition. Addison-Wesley, Reading

Guerra NG, Huesmann LR, Spindler AJ (2003) Community violence exposure, social cognition, and aggression among urban elementary school children. Child Dev 74(5):1507–1522

Huesmann LR (1982) Information processing models of behavior. In: Hirschberg N, Humphreys L (eds) Multivariate applications in the social sciences. Lawrence Erlbaum, Hillsdale, pp 261–288

Huesmann LR (1998) The role of social information processing and cognitive schema in the acquisition and maintenance of habitual aggressive behavior. In: Geen RG, Donnerstein E (eds) Human aggression: theories, research, and implications for policy. Academic Press, New York, pp 73–109

Huesmann LR, Eron LD (1984) Cognitive processes and the persistence of aggressive behavior. Aggress Behav 10:243–251

Huesmann LR, Eron LD (1992) Childhood aggression and adult criminality. In: McCord J (ed) Advances in criminological theory: crime facts, fictions and theory. Transaction, New Brunswick, pp 137–156

Huesmann LR, Guerra NG (1997) Childrens' normative beliefs about aggression and aggressive behavior. J Pers Soc Psychol 72(2):408–419

Huesmann LR, Kirwil L (2007) Why observing violence increases the risk of violent behavior in the observer. In: Flannery DJ, Vazsonyi AT, Waldman ID (eds) The Cambridge handbook of violent behavior and aggression. Cambridge University Press, Cambridge, UK, pp 545–570

Huesmann LR, Eron LD, Dubow EF (2002) Childhood predictors of adult criminality: are all risk factors reflected in childhood aggressiveness? Crim Behav Ment Health 12(3):185–208

Huesmann LR, Moise J, Podolski CP, Eron LD (2003) Longitudinal relations between childhood exposure to media violence and adult aggression and violence: 1977–1992. Dev Psychol 39(2):201–221

Kendall PC, Braswell L (1985) Cognitive-behavioral therapy for impulsive children. Guilford, New York

Miller GA, Galanter E, Pribram K (1960) Plans and the structure of behavior. Holt, New York

Newell A, Simon HA (1972) Human problem solving. Prentice-Hall, Englewood Cliffs

Schneider W, Shiffrin RM (1977) Controlled and automatic human information processing: I & II. Psychol Rev 84:1-66, 127–190

Simon HA (1969) The sciences of the artificial. MIT Press, Cambridge, MA

Slaby RG, Guerra NG (1988) Cognitive mediators of aggression in adolescent offenders: I. Assessment. Dev Psychol 24:580–588

Wyer RS Jr, Carlston DE (1979) Social cognition, inference, and attribution. Erlbaum, Hillsdale

Collective Action

▶ Social Disorganization and Terrorism

Communal Context

▶ Transitional Justice

Communities

▶ Rational Choice, Deterrence, and Crime: Sociological Contributions

Communities and Crime

▶ Neighborhood Effects and Social Networks

Communities and the Police

Jack R. Greene
School of Criminology and Criminal Justice, Northeastern University, Boston, MA, USA

Overview

> *One generation plants the trees; another gets the shade* – Chinese Proverb

This entry considers how research has come to understand the idea of "community" and its implication for civic governance, most particularly in respect to the public police. At the same time, this entry considers what has been learned about the role that each (community and the police) plays in the loosely connected network of social control, and the implications of communities' characteristics and processes for policing more broadly, as well as for their capacity to validate the law and the police.

For several years, reformers, community advocates, and policy makers have emphasized that communities have broader roles to play in policing (Greene 2000). This prompts a consideration of whether communities can indeed coproduce public safety, particularly in residential neighborhoods, and what constrains or facilitates communities in adopting such roles? Lastly, in a late-modern world the question arises, "Can we expect communities to be fully engaged in such activities, or have our notions of social cohesion and integration been inextricably altered?" Such questions explore the ways in which communities contribute to policing, broadly cast, and the limitations or restraints on such contributions.

This entry considers the idea of "community" as a central concept of social theory, more specifically in crime theory, the ways that the idea of community has evolved, and the implications of such evolution for social control and policing. It is an attempt to address what are often juxtaposed ideas: the mechanisms of informal (community) and formal (police) social control and how they impact one another in neighborhood settings. It also addresses the idea that to be effective and accepted, legal sanctions and legal actors need community validation and support, particularly in democratic societies, posing an additional criterion of institutional legitimacy in the discussion of the interactions between community and police. This entry first considers shifting conceptions of community and how they are related to crime and crime control theories. Second, the focus shifts to how the police have come to the idea of community attachment, and some of the complexities of such interactions. Third, this entry considers how police and communities form alliances for crime prevention and what evidence supports the effectiveness of such an approach. And, lastly a consideration of whether there is a broader role for the community in such matters is undertaken.

The Community Perspective: Key Issues

> *The impersonal hand of government can never replace the helping hand of a neighbor* – Hubert H. Humphrey

The idea of the community or *gemeinschaft* was introduced in the late 1800s by Ferdinand Tonnies (1887), a German sociologist. Tonnies' *Gemeinschaft* refers to associations among individuals that orient their identity and behavior to the common mores, customs, and beliefs helping to unify the collective, what he referred to as the "unity of will." Being part of this collective, bound by shared beliefs and shared places, undergirds ideas about what it means to be part of a community. While the family is often seen as the model for *Gemeinschaft*, other associations not based on birth status include communities defined by proximity, as well as those more dispersed, yet connected by beliefs or association (*Gesellschaft*). It was Tonnies' central contention that modern societies gradually moved from *Gemeinschaft* to *Gesellschaft*; from groups based on primary associations (such as kin and close neighbors) to secondary associations (such as workgroups).

In such arrangements, social action, as defined by Weber (1991), takes into consideration the behavior of others, orienting social and group relations. Communities shape and reinforce social relations, providing identity and meaning to those in the community. The idea of community, then, is a social construction, creating shared social realities and interpretations among members, and orienting individuals to others, typically through face-to-face interaction. Such conceptions of community have historical referents, the applications of which are not always clear in the modern world. Nonetheless, the notion of the *Gemeinschaft* community in the conceptions of Tonnies and Weber persists, symbolically and nostalgically.

In discussions of what "community" means, a prime consideration is the type and strength of solidarity in social relations, and the type and degree of social integration in a society. Solidarity, it is said, binds people to social relations (their neighbors and external others) outside the immediate family, and the level and strength of solidarity produces a range in group cohesion and collective identity. Durkheim (1949) was concerned with how to reconcile individual autonomy and collective action in society, that is, how the social order takes shape and is maintained. Durkheim (1949) distinguished between mechanical and organic solidarity as ways in which the social order and social cohesion are affected by the division of labor. In simple societies, kinship and shared values produce social cohesion ("conscience collective") and solidarity, generally through "likeness," that is, the similarity of people in social groups. In more complex societies, an organic solidarity is produced by linking of people through similar work, educational backgrounds, religious affiliation, or lifestyle. Social solidarity, then, arises through kinship ties as well as broader affiliations associated with work or through the interdependence of social arrangements, producing a range of communities, the older ones structured by close ties, and more modern communities being loosely coupled.

The idea of community has been explored from many perspectives, most of which see communities as having dense social and emotional connections, a sense of shared norms and beliefs, social reciprocity, and varying degrees of social responsiveness (Day 2005) (i.e., the *Gemeinschaft*' type). Of course, most of these characteristics of communities are also shaped by close geographic proximity – people living together.

While there is considerable variation in discussions about the most central elements of community, and how these elements have been transformed in the modern and late-modern worlds, the layperson conception of community persists; a collection of persons who share membership (sometimes tightly, sometime loosely) in a social grouping, often but not exclusively constrained by geography, and reinforced by social integration, accommodation, and responsiveness. Whether such definitions can persist in a cyber-networked world or in the face of declining community affiliation remains an open question. Nonetheless, over the past 35 years or so, governments have been systematically reorienting themselves, symbolically if not substantially to better integrate themselves within communities. The rise of the community-government era, with community-health, community-education, community safety, and community policing, among others, is a testament to the power of the idea of community in social, political, and economic life.

Criminology has embraced the idea of community influences in crime for many years, beginning with the work of Shaw and McKay (1942) and continuing to the present. Ideas about crime, most particularly its community roots, suggest that social cohesion in neighborhood or community settings shapes the collective efficacy of that neighborhood, or the level of trust among neighbors, as well as their willingness to intervene in matters of local social control (Sampson et al. 1997). Collective efficacy expresses the idea of neighborhood social cohesion, but perhaps in a less complex and intense way than is the case in previous ideas about community social cohesion. Here the focus is on situated neighborhood collective action rather than broader images of community social action and responsiveness emerging from deep emotional and social bonds. In the modern case, collective efficacy is situational and focused, not broadly cast and diffuse. Nonetheless, the effects of concentrated disadvantage and residential instability with violence are largely mediated by collective efficacy (Sampson et al. 1997).

Collective efficacy also speaks to the ability of communities to exercise whatever social capital is available, that is, to leverage social relations through cooperation and confidence building to achieve larger collective ends. Whereas collective efficacy can be seen as the strength in horizontal bonds between neighbors, social capital can been seen as the capacity of local social networks to vertically integrate within larger political, economic, and social networks, thereby achieving or maximizing local benefit. Knowing something about the level of collective efficacy and social capital in communities, then, tells us something about how that community manages its social affairs and with what effects. Of course, socially dysfunctional and disorganized communities lack both conditions, often succumbing to deviance and criminality, but this occurs in degrees and over time can change.

Community dynamics, then, shape local affiliation and orientations toward collective action, as well as strengthen social networks beyond the local community. From the perspective of modern policing, increasing community collective efficacy and social capital strengthens local social control, the first line of defense in matters of deviance, social disorder, and crime. So as a strategic matter, supporting the development of community collective efficacy and social capital should occupy police attention.

However, as discussed below, policing is not particularly well equipped to directly engage in "community building" other than, perhaps, to attempt to control unruly and disorganized places in the hope of creating an envelope in which community cohesion can grow and strengthen. While it has been argued that community policing has the potential for creating trust in the police, thereby raising community collective efficacy, the conundrum for modern policing is how to use formal processes to support the informal development of "community," or at least some representation of local social control that can strengthen neighbor-to-neighbor bonds, and build social capital. These are difficult arrangements to support. If socially disorganized communities are characterized by poverty, population heterogeneity, transience, and disorderly and criminal conduct, then addressing first-order conditions, such as controlling the streets by the police, does not necessarily lead to improving the stability of these areas. Oftentimes aggressive police action in communities produces mixed responses, with some supporting police interventions and others not. Moreover, it is not clear that in pursuing partnerships with "the community," police and community values can actually align (Thacher 2001).

Moreover, in undertaking such an approach, how communities react to police interventions is not well understood, particularly in light of variations in community composition, structure, and level of social disorganization. How the range of communities are to "partner" with outside agencies is also problematic, given the general findings that (1) identifying a clear voice for community representation is problematic, (2) and community participation in crime-oriented activities and meetings wanes over time.

Understanding what internal community capacities to coproduce local public safety are present and how these capacities (if they exist at all) can be enhanced and sustained are also topics with sparse intellectual and program attention. Lastly, two questions bearing on these matters remain unanswered, (1) what is the tipping point, when communities slide into a place where their capacity to participate in crime prevention erodes substantially, and conversely, (2) what is the tipping point when community social integration is sufficient to create collective efficacy and social capital, thereby enabling communities to engage effectively in such matters?

Such questions assume that police are adept in helping communities build capacity. As Goodman his and colleagues in attempting to measure community capacity in health care interventions (1998, 260) suggested, such capacities include "participation and leadership, resources, social and inter-organizational networks, sense of community, understanding of community history, community power, community values, and critical reflection." Such a wide range of community developmental capabilities are likely beyond the capacity of the police to build, but in partnerships with other community serving agencies, such capacities might be improved. This remains a major obstacle in creating effective police and public crime prevention partnerships, most especially in the communities who need such interventions the most.

Communities as Geographic or Ecological Areas

> *A good neighbor is a fellow who smiles at you over the back fence, but doesn't climb over it* – Arthur Baer

The idea of community has evolved considerably over the years. In some ways, it is a rather elastic concept, a broad umbrella idea expressing patterns of social communication, interaction, and value acquisition in any number of settings. Nonetheless, at its root, it has always implied propinquity, that is, people living together in the same social space, creating patterns of social interaction, from which meaning is derived. In the study of deviance and crime, ecological criminology has centered on the level of community organization, collective efficacy, and social capital as important determinants of effective community action in stemming delinquency and crime (Bursik and Grasmick 1998). The ecological and social-psychological traditions of the Chicago School identify community dysfunction as a primary inhibitor to social control, suggesting that community building could be an important attribute of police-neighborhood policy. Simply put, geography matters, as crime varies geographically, often most prevalent in places ill equipped to marshal social control.

Like criminals, places have criminogenic characteristics. That is to say, some places like some people account for a high incidence of crime and social disorder. Armed with this evidence, the police have identified crime ”hot spots” and attempted to systematically address them, often through aggressive police tactics. The evidence on “hot spots” suggests that targeted police interventions in constrained areas evidencing high rates of crime and social disorganization result in crime reduction, with mixed evidence as to whether some crimes are displaced to adjoining areas, or rather whether such adjoining areas enjoy a “diffusion of benefit,” that is, a similar reduction in crime to the target areas (Weisburd and Braga 2006; Ratcliffe et al. 2011).

While the research literature on the suppression effects of “hot spots” policing has garnered considerable police support and academic attention, more often than not such approaches fail to improve community social networking capacities. Rather, crime is aggressively pursued sometimes displacing its effects to other neighborhoods and sometime not. As it turns out, getting and sustaining community participation in neighborhood policing or community crime prevention efforts is quite difficult (see Skogan and Hartnett 1997, 110–160). This is especially the case in communities where social capital and collective efficacy are in short supply. Moreover, in these same communities, the police and the public are not always in agreement as to what the police should do, nor do they share the same value sets (Thacher 2001). While geography has become a major tool for the analysis of crime, most police responses rely on the police, not the community, for addressing crime and disorder problems, and often through the application of aggressive police tactics. Where the community is not fully “on board” with such actions, or when they are sprung on the community problems emerge, and the legitimacy of the police is called into question. From an analytic and practical perspective geography matters, but what to do with that geography (often referred to as a “community”), is not particularly clear. Moreover, it also appears that in communities where police legitimacy is low, the police are more likely to use force (Kane 2005).

Communities of Interest: Breaking Geographic Boundaries

Give a person a fish and you feed them for a day; teach that person to use the Internet and they won't bother you for weeks – Author Unknown

Consistent with Durkheim's idea of organic solidarity, the modern world is less linked by what were presumed to be deep social and emotional ties at the community level, and more through affiliations associated with work or personal interest. Some have pointed to a significant decline in community affiliation in civic affairs (Putnam 2000) in the latter half of the twentieth century, some of which is related to technological advances in social networking, that is, linking with others, often absent face-to-face interaction.

In a world marked by social networks, the idea of communities has shifted from place to ideas, life styles, near-instant electronic communication, and social fragmentation. Whereas once to be in a community, people actually had to meet, talk, and share their interests, now from the comfort of their own homes, people participate in social communities that span the globe, and perhaps not with the negative side effects of social isolation.

Communities of interests can be organized functionally, for example, business districts or other commercial areas; while being proximate to one another, they are not systematically engaged. Communities can be more dispersed, involving what is now called membership in affinity groups organized to work together but not in a hierarchical way, but rather in a more flexible and decentralized way. Such groups have the characteristics of coalitions, individuals linked to group identities and norms selectively and with indeterminacy, that is, participating in the coalition to the extent that group membership achieves the individual's goals or fosters identity. Such groups are transient, they ebb and flow episodically and they are formed and abandoned as identities shift or one's sense of group affiliation diminishes.

In a Durkheimian sense, such groups do not exhibit mechanical solidarity; rather, they pivot from one form of organic solidarity to another, expressing some social cohesion, but often with limited time horizons. Nonetheless, social networks are social structures (individuals, groups, organizations, nation states) connected by interdependencies, such as friendship, kinship, common interest, financial exchange, dislike, sexual relationships, or relationships of beliefs, knowledge, or prestige. Studies have identified social networks at many levels of social organization, and have demonstrated that social networks are important in determining the ways problems are identified and get solved, as well as the ways network members achieve their goals and understand their social conditions, that is, how they make sense of circumstances and then act upon them. Such social networking also provides for cultural transmission better linking social network and social culture analysis.

Policing as a Community Activity

> *The police are the public and the public are the police; the police being only members of the public who are paid to give full time attention to duties which are incumbent on every citizen in the interests of community welfare and existence* – Robert Peel

Policing has always been a community activity whether or not formally acknowledged. As part of the social control continuum, the police are intimately connected to community notions of acceptable and unacceptable behavior, order and disorder, as well as interpretations of lawfulness and deviance. In turn, the police, the formal apparatus of the state, are directed through representative processes to evenhandedly enforce the law with public support for such enforcement. While the public rarely has a formal referendum on police activity (although news accounts and public protests do occur), people routinely assess the legitimacy of the law and law enforcement – both substantive and procedural justice (Tyler 2006). They seek fairness in the law and in its application.

In democratic societies, the social contract authorizes government to provide an envelope of security, safety, and access for the populace, and expects people to conform to social and legal expectations. As a consequence, governments regulate many forms of social, political, and economic behavior, including those associated with deviance and crime. At the same time, members of the society generally comport with social expectations emphasizing the observance of duly constituted law and social regulation. From a criminological perspective, self-control theory suggests that early socialization and social learning involving the acquisition of values, norms and beliefs, increases self-control, and thereby reduces deviance and crime. Of course, social learning has an important community context as well.

Shaw and McKay (1942) found that delinquency increased in neighborhoods characterized by poor economic conditions, ethnic heterogeneity, and residential instability, the structural correlates of crime. Such neighborhoods were said to be socially disorganized and therefore less capable of controlling delinquent behavior. Refinements of these ideas suggest that individual, family, neighborhood, and community identities are at the forefront of social control, and where communities are socially disorganized, lacking collective efficacy and social capital such identities are the most vulnerable (Sampson and Groves 1989) and at risk.

But even in socially disorganized places, a fragile community precedes the police in matters of social regulation and social control, whether done well or not. Simply put, the community is there first, at the forefront of observing, confronting, and sometimes shaping social behavior, pro-social, deviant, and criminal. While communities are legally discouraged from vigilantism or other aggressive self-protective methods, it is fully anticipated that individuals and communities are self-policing.

That is to say, in democratic systems, the state gains compliance, and some degree of social conformity, from the vast majority of its citizens, through consent, not through police force.

From a social-psychological point of view, compliance is conditioned by beliefs about the legitimacy of the state, its laws and its mechanisms of enforcement, as well as social understandings of rightful behavior and personal comportment. This is seen as social and legal conformity or the voluntary aligning of an individual's behavior with widely held social expectations or norms. Of course, norms may compete with one another across social, economic, and political lines, but generally speaking, conformity involves discerning how others behave and what they expect, and then adjusting one's behaviors accordingly.

In democratic states, such conformity is arrived through social consensus; in totalitarian states, such compliance is conditioned by fear. In either case, governments require self-policing, that is, a large proportion of the populace conforming to the law and social convention, whatever its genesis. While the police enforce the law typically against law breakers, or those suspected of such behavior, it is individuals in community settings that engage in self-policing as well as social regulation that stems from community interactions and local social ethos. This is not just a matter of philosophy, but also a matter of practicality. There is no way that the police would have enough capacity to enforce the rules, absent informal social control. Police in the USA number about 2.5 per 1,000 inhabitants, and in Europe, the range is from 1 per 500 to 1 per 200. While in either case the police are outnumbered and their capacity to control large populations is circumscribed, it does not really matter how many police there are – it is the very essence of a social grouping that the primary way of ensuring conformity to most rules is through personal and informal social control, resulting in self-policing.

The distinction between "police" and "policing" that is made here is important, as the former involves activities of government, while the latter refers to a wide range of individual, civic, private, and public arrangements, each of which contributes to preventing, deterring, or otherwise shaping crime and social deviance. In simple terms, the police are charged with maintaining law and order, while policing is too big to be left to the police alone. On its' broadest scale, "policing" is another term for "maintaining conformity" or ensuring that most of the people follow most of the rules most of the time. So, as a matter of practice many, not few, are involved in policing (Loader 2000). While much of the rhetoric and program focus regarding police crime prevention has been in residential community settings susceptible to crime, other forms of policing permeate residential, commercial, military, recreational, and industrial communities, more broadly, in what Brodeur (2010, 17–42) has called the "web of policing."

Ironically, it is interesting to note that while policing can be seen as having deep connections to community norms, sentiments, and ethos, historically, the police have had a passive-aggressive relationship with their civic constituents. That is to say in several times and places, formalized policing has evidenced an institutional "personality," revealing a pattern of negative attitudes toward the community, accompanied by resistant behavior in the face of community calls for change (Greene 2000).

The police have been the object of reform for more than a century; some reforms have improved policing and police organizations while others have been resisted with prevailing functions and structures unaltered (Mastrofski and Willis 2011). And, of course, in some instances, the police flatly reject ideas associated with community coproduction of public safety and community/police partnerships (Lyons 2002; Herbert 2006). Such rejection has been most evident in communities of color, or those marginalized by ethnicity or economics, in part as a rejection of community capacities or willingness to work with the police but also as a reflection of resistance to police institutional change. In either case, such complexity speaks to the likelihood of the police and community working together on matters of local crime control and social order as being more episodic than permanent.

Policing in Search of Community

> *It is vain to talk of the interest of the community, without understanding what is the interest of the individual* – Jeremy Bentham

While it has been important for policing to consider its "community roots," those roots are complicated by individual variation in values and concerns within any geographic place. Managing value conflict in communities and in larger political settings is a major undertaking in the process of governance. Communities rarely speak with one voice. Consequently, assessing community dynamics is a necessary condition to understanding the underlying value consonance and dissonance in a communal setting. As the police search for "community," they must be wary not to be drawn from one community voice, over others, lest their credibly in the community be compromised.

Moreover, having a police who see themselves as "citizens in uniform" rather than as an occupying force is also a critical component in gaining community acceptance. Organized public policing from its onset recognized that the "police are the public and the public the police" a quote attributed to Sir Robert Peel, founder of the Metropolitan Police of London, the forerunner of modern-day policing. Like the military, the police are "citizens in uniform," a democratic governance idea meaning that the police and the military draw their resources – human, financial, and normative – from the body politic. Correspondingly, the police and the military owe allegiance and fidelity to the society in which they are deeply embedded. Neither the police nor the military "occupy" their civic sponsorship, for to do so would break the social contract empowering democratic governance, while also turning government on the civic body.

In a rather ironic way, the modern police have escaped their community foundations. Instead, the formal and social organization of the police has been distanced from those to be policed despite the anticipated relationship between democratic government and the body politic. There are many reasons for the police to take some distance from the public, some related to a need for institutional and legal independence, and some related to police fear of oversight and public criticism. In either case, policing, a community-focused activity, has gravitated from informal means of social control, to public police as a visible and formal mechanism to achieve the same ends, while at the same time drifting from a close affinity between public and police to one that is often strained, especially in marginal communities. Simply put, public policing has its greatest challenges of connecting with marginal communities when the community and the police do not trust one another. Such social distance poses considerable barriers for the police to lead community renewal efforts.

To some such distancing may seem appropriate to the extent that the impartial enforcement of the law and other formal social regulation under democratic principles requires that the enforcers of the law maintain independence, lest their actions seem biased, or politically motivated. When policing becomes an adjunct to local politics, it divides civic membership between those to be supported and those to be opposed. In the earliest formations of policing, particularly in big cities in the USA, the police were clear adjunct to political machines. This has also been the case in nearly every history of Western policing to one degree or another (see, Critchley 1978). Calls for "taking the politics" out of policing stress the occupational and political distancing of the police and the public, at least symbolically, if not substantively most particularly for the purpose of reducing overt politics in police decision making and practice.

Police grasp of the professional mantel also helped to distance professional from client, while at the same time affirming the status of the professional. As the public police have sought public recognition as a profession, so too have those to be policed become distanced from those doing the policing. This is rather ironic, as historically the police came from the very classes that formal policing seeks to oversee and control.

Finding the appropriate distance (or closeness) of the police and the public has been complicated and at times elusive. Should the police become too close to the public, fear of the corrupting influences of community takes center stage; if the police are seen as too distant from the public, fear of an independent and repressive police rise. So, a too close or too distant police invariably raises concerns of politics and accountability; to who are the police accountable and by what methods?

Despite efforts to distance the police from the public, the police are inherently captured by the social contexts in which they find themselves including those of community life. Policing is localized, it occurs somewhere – neighborhood, highway, business district, or back ally. The social contours of these places shape police actions within them, as the police have and continue to use perceptual shorthand in understanding places as well as the people that inhabit them. Moreover, several have commented on the emergent styles of police that are associated with types of communities.

Such distancing from the public can have adverse consequences as well. Since the times of Sir Robert Peel, it has been recognized that an acceptable public police must engender the respect and trust of the public, lest they be seen as agent provocateurs or spies. Given that the police are outnumbered, meaning there are more of *us* then *them*, public order and community safety require that people police themselves, as well as through a complex system of regulation provided from other agencies, public and civic. That is to say, behavior is partly shaped by the police and partly by the labyrinth of social and institutional regulation part of modern society.

Today, we are regulated from many quarters, some focused on deviance and criminality, but most on broader issues of social regulation (education, health, recreation, and the like) and public management (zoning, taxation, product control, and the like). The role of communities in the web of social regulation under such circumstances is not always clear. Once the bastion of social affiliation and hence social control, communities have been transformed in the post-modern world such that community influence in regulating social behavior is often fragmented and diffuse.

Still policing needs community support to be both effective and legitimate. From the perspective of police effectiveness, it has long been recognized that police are largely mobilized by the community. That is to say, policing in response to community needs is significantly affected by the community's capacity and willingness to seek police assistance. While patrol responses to troubled neighborhoods seek to provide support for "hot spots," knowing where the "hot spots" are, and more importantly what they mean, requires community input. "Hot spots" are based on crime reports, and community members or businesses are the necessary ingredient for reporting crime and other worrisome behavior. Absent community amplification of crime and social problems leading to crime through the identification of crime situations and reporting them, the police would be responsible for detecting crime and/or the circumstances underlying crime independently. This is not the model of public policing most have come to understand. So, if a community lacks the capacity or will to contact the police about criminal victimization, unruly neighborhood circumstances, or business concerns, the police have little in the way of focusing their efforts on particular places, problems, or circumstances. In this way, the police are highly dependent on the community to learn about crime.

Of course what motivates community involvement with the police is a belief that the police can do something to improve circumstances and a level of trust that the police will respond appropriately. Community acceptance of police legitimacy sets the underlying conditions of not only police mobilization, but also acceptance of the police as a source of support, cooperation, or collaboration. Unfortunately, in many socially disorganized communities which often also experience high crime, civic trust of the police is low (for a review, see, Weitzer 2000). What we know about police activity in socially disorganized neighborhoods suggests that the police behave differently there as compared to more stable and/or more affluent communities. Those behaviors are to *under- or over*-police these neighborhoods, leading to much community dissatisfaction with the police and complaints about police misuse of authority and force (Weitzer 2000; Kane 2005).

This discussion calls into question our understandings of what it means to be a community, how communities are shaped and in turn shape individuals, and how this social construct – community – actually works. To call for greater police involvement with the community, community mobilization, community coproduction of crime prevention, or community policing assumes a working and generalizable definition of communities, their social and social-psychological dynamics, especially related to social control, and their ability to influence government policies and actions. If police are to work with communities, then issues of representation, communication of interests, capacity, and sustainability arise. In all cases, communities need to have some way to clarify and then represent their interests and communicate them to others. At the same time, these communities need to have some capacity to bring to bear on the problem at hand, in concert with the police and other service providers, and this capacity needs to be sustainable over time.

The Police and the Community: The Complexity of Alliance

> *The only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good, either physical or moral, is not sufficient warrant* – John Stuart Mill

Government alliances with communities are invariably complex. Generally speaking they are asymmetric in terms of power and communication, and likely fraught with concerns about the corruption of government policy or the co-optation of local community sentiment toward the government. Depending on whether one takes a conflict or consensual view of the relationships of people to government, communities as an intermediate social aggregate can come to view government as a benign or oppressive force in daily life. Such views are often stratified by race and social class as well as by prior experience with government agents, most particularly with the police. So building community alliances is a difficult proposition for modern-day policing. It is art and science: art in connecting community interests with government agendas, and science in fitting police actions to community needs and local social tolerances.

Building community-police alliances requires “winning the hearts and minds” of both the police and the community, at least in areas where social control and neighborhood safety are at stake. Of course the rhetoric of community/police partnerships often outstrips the reality, for any number of philosophic, theoretical, and pragmatic reasons (see for example, Mastrofski 2006, 44–73).

At the philosophic level, a fundamental dilemma for democratic policing is public acceptance of the imposition of law by the police (Manning 2010). Moreover, as policing has come to be distanced from communities, most especially in disadvantaged communities, the police have also often come to be viewed as an occupying force in unsettled neighborhoods, at least by portions of that community. Of course, much of what the police do occurs in these often neglected communities, placing the police in distressed and often conflict-ridden areas, where support for the police is at best tentative. And, while the police have often genuinely attempted to assuage community concerns with a range of police interventions, it might be said that the police are always one bad incident away from community criticism. Such a precarious relationship adds tensions to police and community interactions, each seeking to protect individual and institutional dignity, as well as social and institutional legitimacy. These circumstances continue to shape police and community expectations about the others’ motivations, as well as conditioning what is expected of each in securing neighborhoods safety.

The idea of the precariousness of community support for the police is illustrated in a long history of community reactions to police interventions. In assessing the wave of riots that swept across urban America in the 1960s and 1970s, for example, the Warren Commission noted that a number of these riots were precipitated by questionable or overly aggressive police actions – actions which ultimately sparked the social unrest witnessed. In a similar fashion in 1991 following the video-taped beating of Rodney King and the riots it sparked, the Christopher Commission documented conflict between the LAPD and racial communities in Los Angeles and widespread distrust of the police. The King incident was perhaps not new to minority communities in Los Angeles, but its widespread viewing not only confirmed others' experiences, but became a rallying point for community riots in that city.

More recently, police interventions in Oakland, California, in 2011 and other US cities have significantly strained police and community relationships once again. This pattern is not restricted to the USA, similar concerns are apparent in the UK as well (Morrell et al. 2011) where youth reported that some were motivated by wanting "to get back at the police" for the earlier shooting of a young man in Tottanham. Similar findings emerge from assessments of the 2007 Paris riots and the role of the French National Police in dealing with these riots. According to Mouhanna (2009) the federal nature of the French Police Service and most particularly the riot police, the complexity of policing in Paris and its immediate surrounds, and the marginalization of minority communities, particularly Islamic communities in France, all contributed to the French riots. As important, many of these riots were seen as being initially triggered by police actions.

In the USA, starting with the political and administrative reforms of the police during the first half of the twentieth century, and in police adoption of the "professional mantel," the nexus between local informal social control and its formal counterpart in the public police and law was strained and perhaps even severed. A managerial gloss over policing led to police fixation with response times, case clearance and arrest rates, all efforts of the police, while at the same time ignoring the supposed outcomes of these efforts, namely, safe and secure communities. In large cities, the police responded to community crime in a less than systematic way, and often have to return to the same communities for reasonably speaking the same reoccurring problems. Police were treating the symptoms of community crime and disorder problems, not the underlying sources of those problems (Goldstein 1990). Perhaps more important, the police rarely sought to reinforce community social control norms, portraying such efforts as being "social work," not real police work. Such attitudes persist today.

Surely communities continue to constrain social behaviors when and where they can, but massive population shifts and the effects of increased urbanization strained community capacity for social control, especially in large cities. Such strain was accompanied by police laying greater claim to being singularly responsible for controlling crime. Crime was on the rise, social turmoil visibly present, and public fear of victimization daily present. This period witnessed a dramatic shift in crime control expectations and governmental apparatus (Garland 2001).

Beginning in the 1990s, community policing, or variations of it, became a national mantra of the American police (see Greene 2000), ultimately being exported to many parts of the world. The language, symbolism, and programs of community policing have sprung up in urban, suburban, and even rural police departments and ideas associated with community policing became a new orthodoxy for cops.

Simultaneously ambitious and ambiguous, community policing promised to radically change the relationship between the police and the public, by address underlying community problems, and improving the living conditions of neighborhoods. The police were admonished to deal with crime, of course, but also with "community quality of life," a circumstance thought injured by physical and social incivility, or in the parlance of the times, "broken windows." The organizing themes of community policing called for law enforcement to be more focused, proactive, and community sensitive, sometimes often competing objectives.

Community policing also suggested significant changes in the social and formal organization of policing. On the level of social organization, community policing was thought to break down the barriers separating the police from the public, thereby impacting formal organization through the use of police generalists and the decentralization of police services, while at the same time inculcating police officers with a broader set of community service ideals seeking to change the social organization of policing as well. Whether community policing simply joined the pantheon of presentational strategies of the police (Manning, 1977) or was an added circumlocution (Klockars, 1988), its political appeal was far reaching. In the US during the 1990s the federal government committed to putting 100,000 community policing officers on the streets and spent nearly $9 billion doing so.

The promises of community policing are many. They include: strengthening the capacity of communities to resist and prevent crime and social disorder; creating a more harmonious relationship between the police and the public including some "power sharing" with respect to police policymaking and tactical priorities; restructuring police service delivery by linking it with other municipal services; reforming of the model of police organization; creating larger and more complex roles for individual police officers; reducing police overreaction and personalizing police services, and increasing civic respect for the police and law through a model emphasizing responsiveness, transparency, and the upholding of democratic values. This new style of policing is said to produce more committed, empowered, and analytic police officers, flatten police hierarchies, and open the process of locally administered justice to those who are often the object of justice decision making. This shift also sought to affirm crime prevention as the ascendant goal of policing, not crime suppression, and in important ways reaffirm the traditional roles of the police (Innes 2007).

While the anticipated outcomes in community policing are indeed broad, there is considerable breadth in the definition of what constitutes community policing as well. The range is rather remarkable, given the centrality of community, often lost in this discussion. Community policing has variously been defined as neighborhood patrol or "policing the patch," reassurance policing, community involvement in crime prevention activities, as well as zero-tolerance policing strategies seeking to deal with small but reoccurring crime and social disorder, such as panhandling, street-level prostitution, public intoxication, gambling, and less serious property and personal crime. Police crackdowns, sweeps, and aggressive police interventions have also been associated with community policing, to the extent that they are supposed to provide the community with welcomed relief from the tyranny of street thugs and gang members.

Under the rubric of broken windows policing, a wide array of police interventions are linked to improvements in communities, although on closer examination, the community is not participant to these strategies; rather, the community is the target of such strategies. With such concept elasticity, it is perhaps inevitable that many activities seek harbor in community policing symbolizations.

However, strategies which do not engage the community or do so only marginally fail to account for variations in community social integration, and rather focus on the community as criminogenic and therefore warranting aggressive police intervention, generally fail the test of improving community collective efficacy or social capital. In using such police tactics, once the visible and immediate problems (crime) are gone or displaced, the community is rarely seen as in need of continued support. This is all the more problematic as recent research suggests that a small number of communities have a consistent trajectory of crime and social disorder (Weisburd et al. 2004), suggesting that these communities may actually cycle through periods of crisis and remission, only to realize crises again. In some unfortunate ways, broadly cast community policing strategies absent community attachment return us to a previous era of "fire-brigade policing," that is, showing up after the fire has been reported, as opposed to preventing the fire from its onset. In some important ways, the idea of "community policing" has been highjacked by historical approaches to policing, saturation patrol, and aggressive police street tactics, masked as community interventions.

Community partnerships with the police are also complicated by the police themselves. There is considerable evidence that the police while publically vocal in support of such connections, privately, within the police occupational culture eschew such community linkages (Herbert 2006). While there are likely many reasons for such contradiction in the public and private faces of police, it is clear that community policing still struggles with "winning the hearts and minds" of line-level police officers, those who would presumably be the front-line boundary spanners in community-police partnerships (Lurigio and Skogan 1994). More recent studies of police attitudes toward community policing suggest that the police do not embrace these ideas, and that police work group cultures either downplay or demean the practice of community policing (Herbert 2006; Lyons 2002).

In other parts of the world, even when community policing practices are observable, the police, when interviewed, cannot articulate the philosophy or the practical implications of community policing. While it is clear that policing in much of the western world has embraced the symbolism of community policing, problem solving, and civic engagement, it is less clear that such ideas have permeated the police organizational and work group cultures, where crime control through aggressive police action continues to dominate police practice.

While it is true that the police struggle to find their footing in the philosophy and implied practices of community policing, so too does the community struggle with its new role in the coproduction of public safety. In some respects, the public is of two minds: The police should continue to "fight crime" when and wherever it appears, but when such actions affect me and my community, the police should become more conciliatory and inclusive. Perhaps, such is the nature of most public policy; collective goods are difficult to individualize, except when you are in need of them or are the object of their interventions. Partnerships require the willingness and capacity of the police and the community to engage in a dialogue on community safety, each bringing a contribution to addressing community crime and disorder problems. While a nostalgic view of community and historical policing may afford some expectation of such capacity, their presence and alignment in today's world is, of course, more complex.

Policing the Community: The Evidence

A wise man proportions his belief to the evidence –
David Hume

Community policing is the product of great pessimism and great optimism – pessimism about the police having become estranged from the community and concerns about community/police reconciliation, and optimism that recontextualizing the police in community trappings would ultimately improve neighborhood safety as well as the relationships between the police and the public. Beginning in the early 1990s and continuing in some measure today, policing in the USA and in other parts of the world embraced the idea and programs of "community policing," a derivation of policing emphasizing, among other things community mobilization against crime and disorder, police connectedness in neighborhood settings where the police had been historically absent, improved police/community interactions and willingness to share information, targeted programs to strengthen community social control and crime resistance, and greater transparency and accountability of the police to the public – big ambitions indeed.

The evidence of the success of such shifts in policing or in police/community partnerships is rather thin. Partly this is the case because community policing interventions have been diffuse in approach and sporadic in application. Partly this is due to concerns with the scientific rigor with which community policing has been assessed. And, part of this struggle to define and measure the implications of community policing can be associated with conceptual or theoretical gaps in how the questions of community policing have been framed.

While there is a rather voluminous literature associated with community policing, the evidence suggests that while such programs may affect fear of crime, they do not impact crime as directly. Moreover, research suggests that improvements to community cohesion stemming from community policing are fleeting (see Taylor 2006). The national evaluation of community policing conducted in the USA (see, Roth and Ryan 2000, 2) found:

> Building partnerships with communities by COPS grantees was commonplace in many of the agencies visited but, all too often, partnerships were in name only or simply standard, temporary working arrangements. Most visited agencies did some form of problem solving, but form and visibility varied widely from agency to agency. In observed sites, crime prevention efforts abounded primarily manifested as traditional programs now subsumed under the community policing label.

The struggles of community policing are conceptual and operational. They involve assumptions about the capacity and willingness of the police and the community to work with each other. They also invoke ideas about community life and police agency receptivity to change which are tenuous at best. Finally, they require considerable investment to accomplish, as well as a reorientation of what we mean by partnership and neighborhood safety.

Is There a Broader Role for Communities in Policing?

> *A community is like a ship; everyone ought to be prepared to take the helm – Henrik Ibsen*

In the search for a more defined and anchored role for communities in the prevention of crime and in responding to crime and social disorder, many questions arise. These questions largely ask about the inclination and capacity of communities to engage with the public police in establishing greater levels of social control, most particularly in areas where such controls are weak or do not exist. Moreover, these questions seek to better understand the working parameters of community coproduction to assure that the community does not become absent from such discussion (community crime control) or entirely responsible for neighborhood public safety (for example, see Taylor 2006).

So, do communities have broader roles to play in policing? Can these communities coproduce public safety, particularly in residential neighborhoods? What constrains or facilitates communities in adopting such roles? And, in a late-modern world, can communities be expected to be fully engaged in such activities, or have notions of social cohesion and integration been inextricably altered? Such questions ask about the ways in which communities can contribute to policing, broadly cast, and the limitations or restraints on such participation.

To begin to consider the role of the community in the maintenance of social control first requires some idea of how communities might broaden social investment and outcomes. That is to say, what animates community social investment leading to social cohesion and collective efficacy? Here we are drawn back into the nostalgia of "community," a place of deep social and emotional bonds, indelibly shaping social relations and meanings. Unfortunately, such ideas of community no longer capture modern social life nor are they seen as exerting tremendous influence on individual identities. What is shown is that levels of social capital and collective efficacy help to distinguish levels of community crime (Sampson and Groves 1989), suggesting a variable community role in such matters, depending on the presence or absence of crime and community social cohesion.

Continuing studies of communities and crime reinforce these earlier findings, yet questions concerning mechanisms for adequately increasing collective efficacy and social capital, and especially those involving the police remain unanswered. Nonetheless, there is evidence that the police need to be concerned with their instrumental *and* expressive contributions to community life if changes in community social cohesion and collective efficacy are to improve.

Community policing has been associated with improving community social conditions to the extent that in addressing community crime and social order problems, community fear of crime appears to subside. In short, community policing has been shown to reduce fear of crime, although not crime itself. These findings suggest that the police's focus on crime in neighborhood settings have had the expressive by-product of reducing fear of crime.

In addition to the general effects of community policing on fear, "hot spots" policing models and assessments of them clearly represent an instrumental way of thinking (see Weisburd and Braga 2006, 225–244), that is, these findings apply where the police are evaluated by what they directly produce, most particularly their ability to make arrests and reduce reported crime. The focus on crime analysis, Compstat, and similar technologies see crime reduction through targeted and often aggressive police actions as the products of public policing. Such findings do not suggest that community social capital or collective efficacy necessarily improves as crime and fear subside; rather, these findings anticipate that such police actions will influence the underlying conditions thought to detract from community cohesion, especially community fear of victimization or retaliation. For example, Markowicz and colleagues (2001) suggest that social cohesion, disorder, and fear are linked to the extent that fear of crime may reduce an individual's community involvements, thereby decreasing community collective efficacy.

Such findings, coupled with a long series of analyses suggesting a fear reduction outcome for community policing, support the idea that such efforts by the police might help communities find their own collective voice once some modicum of order is established (in this case by the police). But how these efforts translate to collectively stronger communities is not known. In this regard, the instrumental achievements of the police are necessary but perhaps not sufficient. Evidence of this causal sequence remains thin.

An alternative expressive model of police suggests that public confidence in the police is tied more to public concerns about community cohesion and the role of the police in protecting community values and norms than to the instrumental efforts and outcomes of the police. Here research by Jackson and Sunshine (2007) show that public evaluations of community cohesion, moral trust, and social consensus significantly condition community support for and trust of the police. Such reasoning suggests that the community is looking for police help and support for the community and its sense of control over community life. Bellair (2000) finds that informal community surveillance by neighbors – neighboring – together with residential stability positively influences neighborhood surveillance activities. But Wells and colleagues (2006) found that "... residents of geographic areas characterized by lower levels of collective efficacy are no more or less likely to intervene in the face of local problems than are residents of other areas (540)."

Expressive values are central to ideas of collective efficacy as it is the attachments and involvements of community members as well as the meaning such interaction provides, which define collective efficacy. Unfortunately, evidence to date does not support the notion that police support for community by reducing low levels of social disorder necessarily increases community social cohesion or collective efficacy as broader social forces shape insecurity (Jackson et al. 2009). Nonetheless, if communities are to have more involvement in matters of public safety, it is clear that expressive values need to be included in this discussion.

This entry has examined the distinctions between police and policing, the role that communities might play in policing, and the limitations on such participation, as well as the changing role of the public police in the broad fabric of social control and public safety. Whereas 50 or so years ago, the public police had considerable monopoly on these matters, today these are shared undertakings. People are self-policing; communities and other socially based institutions shape public behavior and attitudes including those associated with crime and deviance, the justice system in all its fragments exercises control, as do other civic and economic institutions. As stated preciously, the police are charged with enforcing the law, but policing is too big to be left to the police alone. Garland (2001) refers to this as the "responsibilization strategy" suggesting that governments are relinquishing power and sovereignty by enlisting participation from citizens and nongovernmental actors to control crime. Aas (2007, 134–36) suggests the regulatory state is now composed of new technologies and rationalities shifting social control from governments to a broader base of civic involvement.

Correspondingly, conceptions of communities have changed over the years as well. Whereas the idea of community conjures ideas of social integration and collective social action formed by dense and complex social and emotional ties, today the idea of community is being reshaped in a global and technologically wired world. No doubt, some idyllic communities continue to exist, but in the main, they have become ideas in a sea of practical constraints on communal social action. Despite claims to community resurgence in social and political life, the evidence does not fully support the claim. In a late-modern world, communities and community identities have been fractured.

Globalization, information capitalism, and a culture of consumption have altered the social condition and hence our sense of what it means to be part of a community. Communities of interest, practice, communications, and belief are broadly cast and not restricted to confined geographic areas. While policing has been place bound since its inception, the policing of such a far-flung enterprise perhaps calls for new forms of policing and new forms of community involvement in social control. Such forms of social control are likely to produce new strains between the police and those policed. Of course under such circumstances, it will be difficult for the police to "occupy" the community as has been the general strategy for many years. Instead sharing information and at times following community leadership will challenge the institution of the police, which has been characterized as police centric in its ways of doing business.

In a globalized world with internationalized crime and much broader community connections policing in the twenty-first century and beyond will likely be confronted with the need to expand its role from crime fighting to peacekeeping, while at the same time being engaged in what are new and emerging forms of governance (see Aas 2007, 186–189). The crime implications of globalization include among others cyber-crime, terrorism, and the global trafficking of people, weapons, and drugs. At the same time, globalization also shifts people across the world and seeks greater connection among groups who were previously marginalized. In the Western world and elsewhere, this has involved issues of immigration, and often difficult responses to migrant peoples. As these "new communities" emerge supported by a wide array of human rights groups, policing will need to once again consider the underlying values that have historically undergird state-sponsored police. As Aas suggests (2007, 189) "issues of national justice can no longer be separated from issues of global justice." In a globalized world, crime and community identity are far reaching and notions of communities and the police yet evolving.

Related Entries

- ▶ Biased Policing
- ▶ Broken Windows Thesis
- ▶ Causes of Police Legitimacy
- ▶ Community Policing
- ▶ Neighborhood Effects and Social Networks
- ▶ Order Maintenance Policing
- ▶ Problem-Oriented Policing
- ▶ Reassurance Policing and Signal Crimes

Recommended Reading and References

Aas KF (2007) Globalization & crime. Sage, Thousand Oaks, CA

Bellair PE (2000) Informal surveillance and street crime: a complex relationship. Criminology 38:137–167

Brodeur J-P (2010) The policing web. Oxford University Press, New York

Bursik Jr RJ, Grasmick HC (1998) Neighborhoods and crime: the dimensions of effective community control. Lexington Books, New York

Critchley TA (1978) A history of police in England and Wales. Constable, London

Day G (2005) Community and everyday life. Routledge, New York

Durkheim E (1949) The division of labor in society. Translated by George Simpson. The Free Press, New York

Garland D (2001) The culture of control: crime and social order in a contemporary society. University of Chicago Press, Chicago

Goldstein H (1990) Problem-oriented policing. McGraw-Hill, New York

Goodman RM, Speers MA, Mcleroy K, Fawcett S, Kegler M, Parker E, Smith SR, Sterling TD, Wallerstein N (1998) Identifying and defining the dimensions of community capacity to provide a basis for measurement. Health Educ Behav 25:258–278

Greene JR (2000) Community policing in America: changing the nature, structure and functions of the police. In Criminal justice 2000, Volume 3, Policies, processes, and decisions of the criminal justice system. National Institute of Justice, Washington, DC, pp 299–370

Herbert S (2006) Citizens, cops and power: recognizing the limits of community. University of Chicago Press, Chicago

Innes M (2007) The reassurance function. Polic: J Policy Pract 1(2)

Jackson J, Sunshine J (2007) Public confidence in policing: a neo-Durkheimian perspective. Br J Criminol 47(2):214–233

Jackson J, Bradford B, Hohl K, Farrall S (2009) Does the fear of crime erode public confidence in policing? Polic: J Policy Pract 3(1):100–111

Kane R (2005) Compromised police legitimacy as a predictor of violent crime in structurally disadvantaged neighborhoods. Criminology 43(2): 469–498

Klockars CB (1988) The rhetoric of community policing. In Greene JR, Mastrofski SD (eds) Community policing: rhetoric or reality. Praeger, New York.

Loader I (2000) Plural policing and democratic governance. Soc Leg Stud 9(3):323–345

Lurigio AJ, Skogan WG (1994) Winning the hearts and minds of police officers: an assessment of staff perceptions of community policing in Chicago. Crime Delinq 40(July):315–330

Lyons W (2002) The politics of community policing: rearranging the power to punish. University of Michigan Press, Ann Arbor, MI

Manning PK (2010) Democratic policing in a changing world. Paradigm Publishers, Boulder, CO

Markowitz EF, Bellair P, Liska AE, Liu J (2001) Extending social disorganization theory: modeling the relationship between cohesion, disorder and fear. Criminology 39:293–319

Mastrofski SD (2006) Community policing: A skeptical view. In Weisburd D, Braga AA (eds) Police innovation: contrasting perspectives. Cambridge, New York, pp 44–73

Mastrofski S, Willis J (2011) Police organization continuity and change: into the twenty-first century. Polic: Int J Police Strat Manag 34(3):55–144

Morrell G, Scott S, McNeish D, Webster S (2011) The August riots in England: understanding the involvement of young people. National Centre for Social Research, London

Mouhanna C (2009) The French police and urban riots: is the national police force part of the solution or part of the problem? In: Waddington D, Jobard F, King M (eds) Rioting in the UK and France: a comparative analysis. Willan, Cullompton, pp 173–182

Putnam R (2000) Bowling alone: the collapse and renewal American community. Simon and Schuster, New York

Ratcliffe JH, Taniguchi T, Groff E, Wood J (2011) The Philadelphia foot patrol experiment: a randomized controlled trial of police patrol effectiveness in violent crime hotspots. Criminology 49(3):795–831

Roth Jeffrey A, Ryan JF (2000) The COPS program after 4 years—national evaluation. Research in brief. National Institute of Justice, Washington, DC

Sampson RJ, Byron Groves W (1989) Community structure and crime: testing social-disorganization theory'. Am J Sociol 94(4):774–802

Sampson R, Raudenbusch S, Earls F (1997) Neighborhoods and violent crime: a multilevel study of collective efficacy. Science 277(15):918–924

Shaw C, McKay H (1942) Juvenile delinquency in urban areas. University of Chicago Press, Chicago

Skogan WG, Hartnett SM (1997) Community policing Chicago style. Oxford University Press, New York

Taylor RB (2006) Incivilities reduction policing, zero-tolerance and the retreat from coproduction: weak foundations and strong pressures. In Weisburd D, Braga AA (eds) Police innovation: contrasting perspectives. Cambridge, New York, pp 98–114

Thacher D (2001) Conflicting values in community policing. Law Soc Rev 35(4):765–798

Tönnies F (1887) Gemeinschaft und Gesellschaft: Abhandlung des Communismus und des Socialismus als empirischer Culturformen. Reisland, Leipzig. In: Harris J (ed) Community and civil society. Cambridge University Press, New York, 2001.

Tyler TR (2006) Why people obey the law. Princeton University Press, Princeton, NJ

Weber M (1991) The nature of social action. In: Runciman WG (ed) Weber: selections in translation. Cambridge University Press, New York

Weisburd DL, Braga AA (2006) Hot spots policing as a model for police innovation. In Weisburd DL, Braga AA (eds) Police innovation: contrasting perspectives. Cambridge University Press, New York

Weisburd D, Bushway S, Lum C, Yang S-M (2004) Trajectories of crime at places: a longitudinal study of street segments. Criminology 42(2):283–321

Weitzer R (2000) Radicalized policing: residents' perceptions in three neighborhoods. Law Soc Rev 34:129–55

Wells W, Schafer JA, Varano SP, Bynum TS (2006) Neighborhood residents' production of order: the effects of collective efficacy and responses to neighborhood problems. Crime Delinq 52(4): 523–550

Communities That Care

J. David Hawkins, Richard F. Catalano and
Margaret R. Kuklinski
Social Development Research Group, University
of Washington, Seattle, WA, USA

Synonyms

Coalition; CTC; CTC prevention coalition; CTC prevention system

Overview

Communities That Care (CTC) is a community-based operating system for decreasing risk, enhancing protection, and reducing problem behavior in youth community wide. The approach has been developed and modified through community input and formative evaluation over 20 years. CTC provides communities with a structure for engaging stakeholders to improve youth development; a process for establishing a shared community vision; tools for assessing levels of risk and protection experienced by community youth; and a framework for prioritizing risk and protective factors, implementing evidence-based programs to address them, and monitoring and assessing attainment of specific, measurable community goals. CTC has been implemented in over 500 communities across the United States and countries including Australia, Canada, Croatia, Germany, the Netherlands, and the United Kingdom. All materials needed to implement CTC can be downloaded from http://www.communitiesthatcare.net. CTC has been found effective in reducing youth delinquency, violence, and tobacco and alcohol use in the Community Youth Development Study (CYDS), a randomized controlled trial conducted in 24 communities in seven states across the United States. This entry describes the theory underlying CTC, key features of successful community mobilization efforts, the five phases of CTC implementation, system- and youth-level findings from the CYDS, implementation challenges, and steps to broader dissemination.

Foundation

Over the past 35 years, prevention science has emerged as a discipline built on the integration of life course development research, community epidemiology, and preventive intervention trials. Researchers have identified longitudinal predictors, like family conflict or early academic failure, that increase the likelihood that young people will engage in problem behaviors associated with significant morbidity, and in some cases, mortality (Catalano et al. 2011; Hawkins et al. 1992b; Patton et al. 2009). Other predictors, like strong bonds to positive adults or the development of specific competencies, are protective, associated with reducing problem behaviors and increasing favorable outcomes like academic success, even in the presence of risk (Catalano et al. 2011). Many risk and protective factors are predictive of multiple adolescent problem behaviors, suggesting that improvements can affect a broad set of outcomes simultaneously.

Researchers have designed interventions to reduce risk and enhance protection and in controlled trials have found a number of preventive interventions to be effective in reducing prevalent and costly problems, including adolescent substance use and delinquency. There is a growing evidence base of what works to prevent youth problem behaviors and promote healthy development, and several lists of tested and effective programs have been made available to the public (see, for example, the list of programs identified at www.colorado.edu/cspv/blueprints, www.cochrane.org, and www.campbellcollaboration.org).

In spite of these advances, tested and effective approaches to preventing youth problem behaviors have not been widely adopted (Ringwalt et al. 2009). Reasons include a lack of knowledge about scientifically proven prevention programs, difficulty marshaling resources for science-based prevention and health promotion efforts, and failure to implement proven programs with fidelity. The Communities That Care (CTC) prevention system, which mobilizes community stakeholders to collaborate on the implementation of a science-based community prevention system, was developed to address these issues (Hawkins et al. 1992a) and has been refined over more than 20 years through community input and formative evaluation. This entry describes the CTC approach to prevention, steps involved in implementation, major findings from a randomized controlled trial, challenges in implementing CTC, and next steps in the dissemination of CTC.

Community Mobilization Efforts to Prevent Youth Problem Behavior

Community mobilization has been promoted as a strategy for reducing risk, enhancing protection, and improving youth outcomes (David-Ferdon and Hammond 2008; Fagan et al. 2011b; Hawkins et al. 2010). Typically, at the center of these efforts are locally based coalitions composed of diverse stakeholders committed to positive youth development. Community mobilization is promising because of coalitions' intimate connections to and knowledge of the local environment, capacity for sharing information and resources among service providers and others, and potential to efficiently provide prevention services in an organized manner and without duplication. Coalition responsibilities often focus on planning and coordinating prevention activities, which may involve implementing specific prevention programs to address local priorities, altering the environment by changing local policies and norms, or both.

In spite of their promise, community mobilization strategies have often had little success in meeting prevention goals, and many have been ineffective in reducing risk and problem behaviors (David-Ferdon and Hammond 2008; Fagan et al. 2011b; Hallfors et al. 2002; Merzel and D'Affliti 2003). Lack of success has been attributed to insufficient guidance in implementing prevention strategies, a mismatch between services and goals or inadequate service "dosage," and failure to select and use tested and effective prevention programs and strategies (Fagan et al. 2011b). Hallfors et al. (2002) suggested that coalitions could increase their chances of success by adhering to several criteria: (a) establishing focused and manageable goals, including well-defined, measurable outcomes; (b) using evidence-based programs, receiving adequate training to carry them out effectively, and monitoring implementation with attention to both dose and quality; and (c) evaluating impact by using meaningful indicators that correspond to overall goals as well as the programs chosen. CTC's design incorporates these recommended features.

The Communities That Care Prevention System

CTC mobilizes a broad base of community stakeholders to improve youth outcomes through the development and implementation of a science-based community prevention system. CTC "communities" are geographically defined places large enough for educational and human services to be delivered at that level. They can be incorporated towns or suburbs, or neighborhood or school catchment areas within larger cities. The community population for a CTC initiative should range from 5,000 to 50,000.

CTC is guided by the social development model, a theory of youth development which hypothesizes that to develop healthy, positive behaviors, young people must be immersed in family, school, community, and peer environments that communicate healthy beliefs and clear standards for behavior. When strong bonds exist between youth and prosocial individuals and groups, youth are more likely to adopt positive beliefs and healthy behaviors. Bonds are fostered when youth are provided with *opportunities* to be actively involved in meaningful, developmentally appropriate activities, *skills* to be successful in those activities, and *recognition* for their efforts, achievements, and contributions to the group (Catalano and Hawkins 1996). In addition to

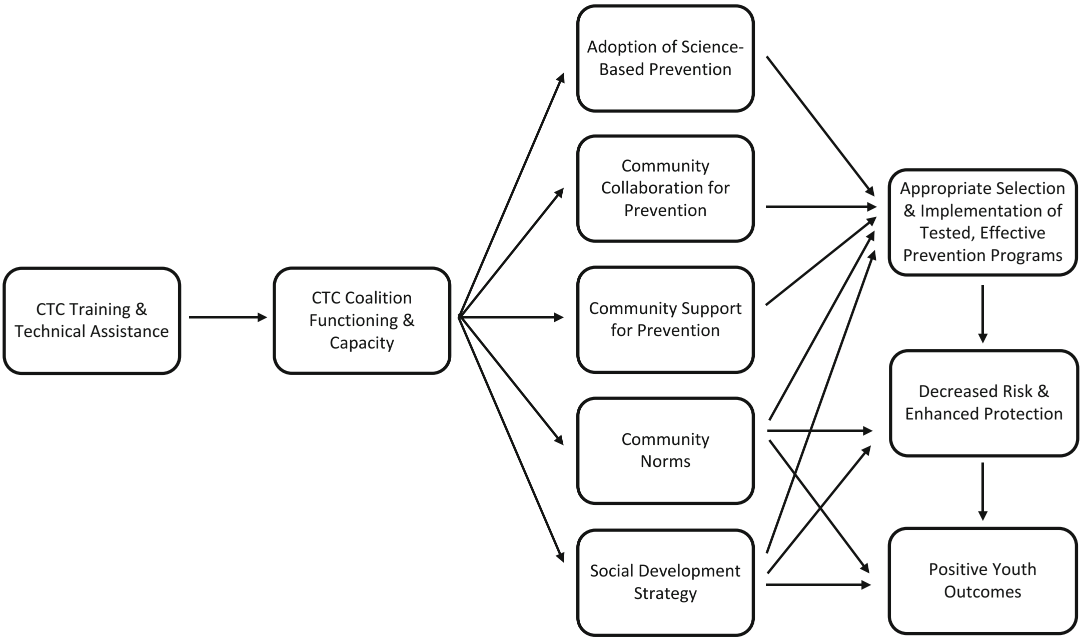
CTC Training & Technical Assistance
CTC Coalition Functioning & Capacity
Adoption of Science-Based Prevention
Community Collaboration for Prevention
Community Support for Prevention
Community Norms
Social Development Strategy
Appropriate Selection & Implementation of Tested, Effective Prevention Programs
Decreased Risk & Enhanced Protection
Positive Youth Outcomes

Communities That Care, Fig. 1 CTC model of community change

supporting the implementation of evidence-based prevention practices that reduce risk and enhance protective factors, CTC encourages community stakeholders to adopt the social development model in their daily interactions with young people and ensure that they are provided developmentally appropriate opportunities, skills, and recognition, as well as healthy standards for behavior. The social development model also underlies CTC community mobilization and training efforts. CTC creates opportunities for all interested stakeholders to participate in developing a shared vision for positive youth development based in prevention science. Diverse community representatives develop skills to work together effectively and are recognized for their efforts and contributions to youth development. Social bonds among coalition members enhance their commitment to science-based prevention work.

CTC's theory of change is depicted in Fig. 1. Training and technical assistance by certified CTC trainers serves as a catalyst for the development of the community prevention board, which commits to using a risk- and protection-focused approach to preventing youth problem behaviors. The coalition utilizes survey data to identify elevated risk factors and depressed protective factors within the community and then prioritizes which to address through evidence-based prevention programs targeting their priorities. The coalition facilitates broader prevention service system transformation by encouraging collaboration and the adoption and faithful implementation of tested and effective prevention programs aimed at prevention goals. Community prevention actions are expected to lead to reductions in risk and enhancements in protective factors, which in turn are expected to lead to more positive youth outcomes. Community-level impacts on risk and protective factors are expected 2–5 years following faithful implementation of tested and effective programs, and community-wide changes in youth outcomes are expected to occur within 4–10 years.

CTC Implementation: Providing Communities with the Tools, Skills, and Knowledge to Achieve Local Prevention Priorities

A major challenge for prevention scientists committed to applying research in the "real world" is increasing the use of tested and effective prevention practices while recognizing that communities differ from one another and need to decide locally what effective policies and programs to use. Local control is built into CTC from the beginning. CTC provides education and tools to build community capacity for using prevention science to decide which risk and protective factors and outcomes to prioritize and which tested and effective programs and policies to implement to address community-specific concerns. It takes communities approximately 1 year to complete CTC's five training phases and develop the skills and knowledge to choose and faithfully implement evidence-based prevention programs addressing local priorities (Hawkins et al. 2008b; Quinby et al. 2008). Each phase has specific milestones and benchmarks and includes training and technical assistance provided by a certified CTC trainer.

Phase 1**: **Get Started. In this phase, community leaders concerned with preventing youth problem behaviors assess community readiness to adopt the CTC system, as well as any barriers to implementation. Readiness requires collaboration among multiple stakeholders and a shared belief that reducing risk and improving protection can prevent problem behaviors. Community stakeholders need to be willing to establish shared goals and collaborate to achieve them. If these attitudes are not present, community readiness needs to be improved before implementing CTC. Other major activities include identifying one or two key leaders to champion CTC, hiring a coordinator to manage CTC activities, and obtaining school district support for conducting a youth survey that will provide local epidemiological data on risk, protection, and youth behaviors.

Phase 2**: **Organize**, **Introduce**, **and Involve. The major task in Phase 2 is to identify and train two key groups from the community in the principles of prevention science and the CTC prevention system. The first group consists of key community leaders and influential stakeholders (the mayor, police chief, school superintendent, and business, faith, community, health, social service, and media leaders) who are introduced to CTC during a half-day *Key Leader Orientation* that includes information about prevention science and the relationship between youth risk and protective factors and problem behaviors. Key leaders are responsible for securing resources for preventive interventions identified through the CTC process and identifying candidates for the CTC Community Board, a demographically diverse and broad-based coalition charged with carrying out CTC planning and prevention activities. The second group, the Community Prevention Board, participates in a 2-day *Community Board Orientation*. The board develops a vision statement to guide its prevention work and establishes workgroups to perform core implementation and maintenance activities: board maintenance, risk and protective factor assessment, resource assessment and analysis, public relations, youth involvement, and funding.

Phase 3**: **Develop a Community Profile. In Phase 3, the CTC board develops a community profile of risk, protection, and problem behavior among community youth; targets two to five risk and protective factors for preventive action; and identifies existing prevention resources and gaps. The first two activities are accomplished in a 2-day *Community Assessment Training*. The major source of data for the community profile is the CTC Youth Survey (Arthur et al. 2007), a school-based self-report questionnaire administered to students in Grades 6, 8, 10, and 12 that measures risk and protective factors experienced by young people, as well as their involvement in problem behaviors like substance use, delinquency, and violence. These data are summarized in a community profile, like the one shown in Figs. 2, 3, and 4. Each profile is unique to the community. Among 10th-grade students in the alternative high school depicted in Fig. 2, survey data indicate a number of elevated risk factors compared to school district averages. Many are drug-related risk factors, and not surprisingly, as shown in Fig. 4, both lifetime and current drug use by 10th-grade students in this school were considerably more prevalent than among 10th graders nationally, as measured by the Monitoring the Future survey (Johnston et al. 2000). Figure 3 shows that, while a large majority of

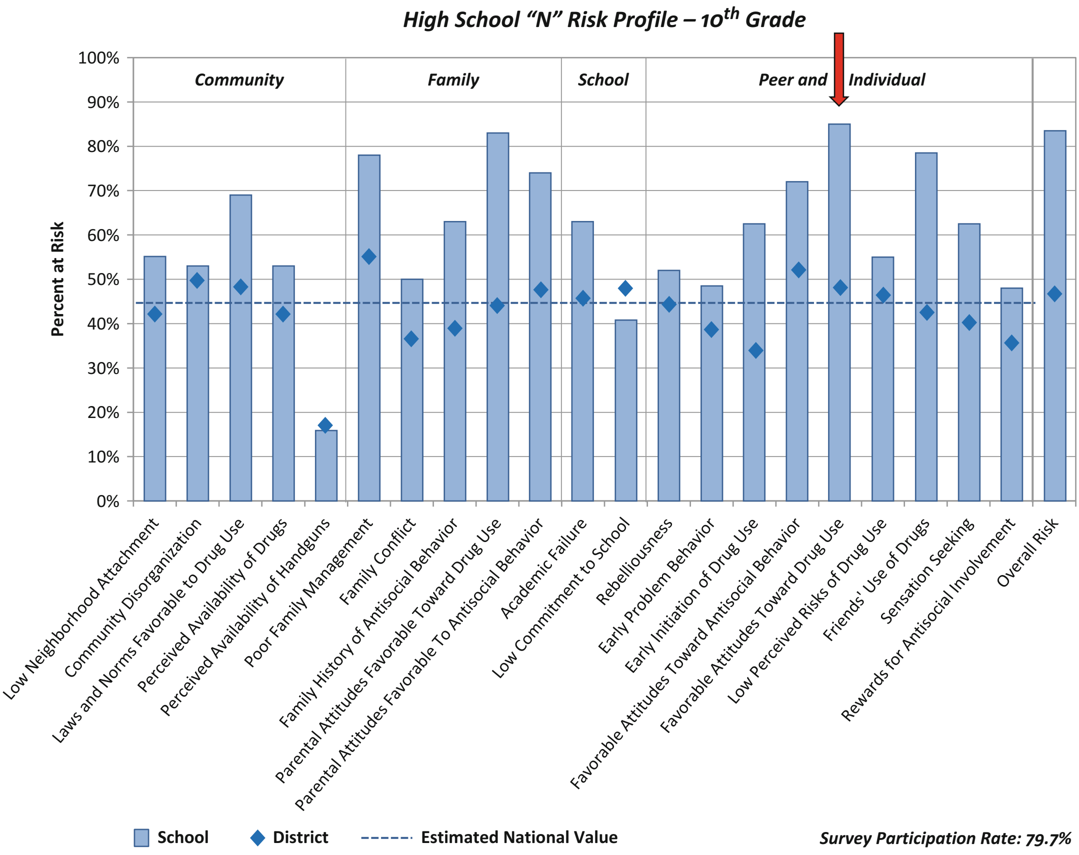

Communities That Care, Fig. 2 Sample community risk factor profile

10th-grade students in this alternative school reported that the school provided many opportunities for active involvement and much recognition for student effort and achievement, many fewer students had strong social skills, including skills to avoid or refuse trouble or problem behaviors. Based on these data, this community's CTC board chose to reduce the risk factor "favorable attitudes toward drug use" and strengthen the protective factor "social skills."

Following a *Community Resource Assessment Training*, board members survey service providers who deliver preventive intervention targeting their prioritized risk and protective factors to assess whether there are existing research-based prevention programs already available in the community. This ensures that any new prevention programs chosen by the board fill existing gaps rather than duplicating effective prevention programs already in place in the community.

***Phase 4**: **Create a Community Action Plan**.* In Phase 4, board members use information gathered in Phase 3 to develop a Community Action Plan. A 2-day *Community Plan Training* provides tools and knowledge to select tested and effective programs, policies, and actions targeting priorities established in Phase 3 to fill gaps in prevention services. The board chooses programs from the *CTC Prevention Strategies Guide*, a compendium of prevention programs (including parenting skills, school curricula, mentoring, after-school, and community-based programs) found effective in changing risk, protection, and problem behaviors in at least one high-quality

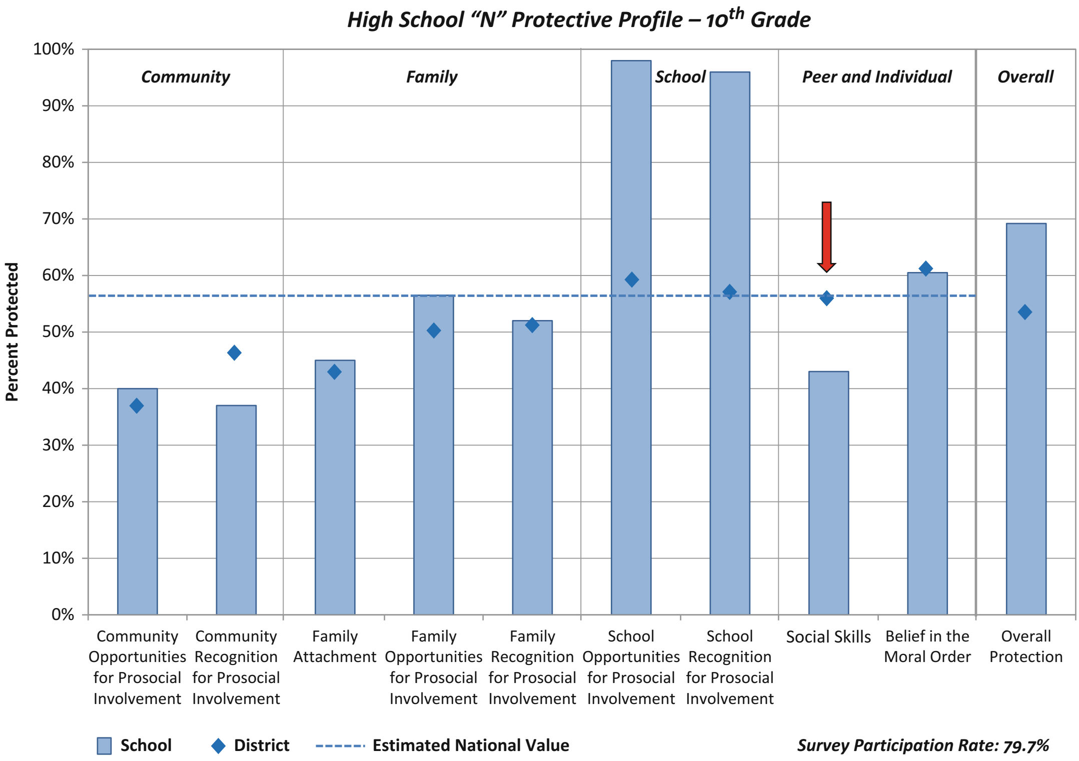

Communities That Care, Fig. 3 Sample community protective factor profile

controlled trial. The Community Action Plan specifies how the coalition will implement the chosen prevention programs, monitor implementation quality, and assess improvement in risk, protection, and problem behaviors.

Phase 5**: **Implement and Evaluate the Community Action Plan. The last phase consists of implementing the Community Action Plan. The *Community Plan Implementation Training* emphasizes the importance of adhering faithfully to the content, dosage, and high-quality delivery specified in program protocols. Communities are directed to prevention program developers for training and technical assistance in the programs selected. Board members and program implementers learn to track implementation progress, assess changes in participants, and make adjustments to achieve program objectives. Monitoring is accomplished through the use of program-specific implementation checklists, observations, and participant pre- and posttests. During this phase the board also engages local media to educate the community about risk and protective factor priorities and generate public support for the preventive interventions selected.

When they complete Phase 5, communities have the knowledge, tools, and skills to faithfully implement tested and effective prevention policies and programs to address locally prioritized risk, protection, and youth problem behaviors. The CTC process is ongoing. Every 2 years the CTC Youth Survey is readministered, and resource assessment data are updated. The CTC board reviews these data to evaluate progress and revise action plans as needed. As previously noted, community-level changes in youth risk and protection are expected to occur 2–5 years after tested and effective prevention programs are implemented. Community-level effects on youth behaviors are expected 4–10 years following implementation.

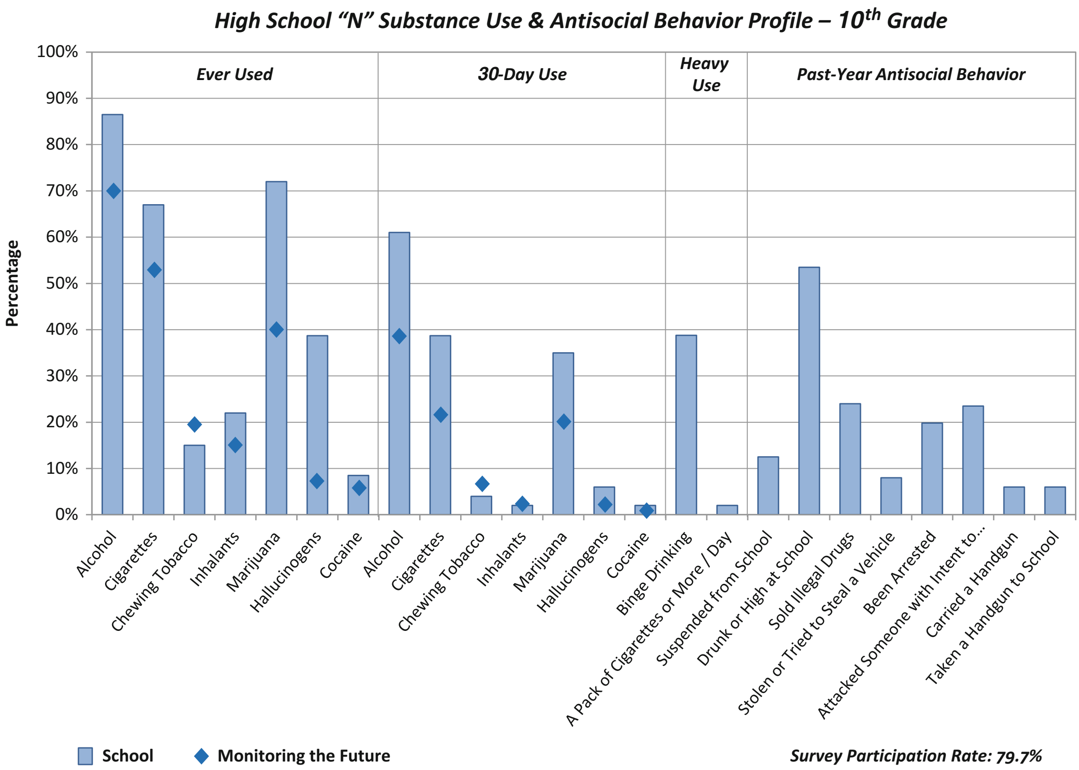

Communities That Care, Fig. 4 Sample community problem behavior profile

The Community Youth Development Study: A Test of Communities That Care

CTC has been evaluated in the Community Youth Development Study (CYDS), a multisite community-randomized trial initiated in 2003 involving 24 communities in Colorado, Illinois, Kansas, Maine, Oregon, Utah, and Washington. Participating communities are small- to moderate-sized incorporated towns with their own government, education, and law enforcement structures (population range: 1,500–50,000, $M = 15{,}633$, $SD = 10{,}147$). Communities were matched in pairs within state on population, racial and ethnic diversity, economic indicators, and crime rates. After receiving commitments from each community's mayor or city manager, the police chief or lead law enforcement officer, and the superintendent of schools to participate in the study, one community from each matched pair was assigned randomly to the intervention (CTC) or control condition. At the study's start, none of the communities had advanced in the use of science-based prevention to the point of selecting and using tested, effective preventive interventions to address prioritized community risks (Brown et al. 2009; Hawkins et al. 2008b). Youth in CTC and control communities did not differ at baseline on key outcome variables including substance use and delinquency (Hawkins et al. 2008b).

To date, intervention effects on prevention systems and youth risk and problem behaviors have been evaluated during the intervention period and up to 20 months after intervention support from the study ended. Prevention system effects were evaluated based on reports of key community leaders, prevention coalition chairs, coordinators, and board members from CTC and control communities. Intervention effects on youth risk and problem behaviors were evaluated based on reports of a longitudinal panel of 4,407 children from intervention and control communities who completed the CTC Youth Survey annually from Grade 5 through Grade 10. Major intervention effects are summarized in Table 1 and are described in the following sections.

CTC Leads to Sustained Prevention System Effects

The CYDS evaluated community efforts to faithfully implement (a) the core principles of the CTC prevention system and (b) tested and effective prevention programs with respect to content and delivery specifications. Broader effects on community prevention systems were also evaluated. CTC communities were supported by training in the first year and technical assistance throughout the first 5 years of the trial and achieved high implementation fidelity at the system and program levels.

High-Fidelity Implementation of the CTC Prevention System. At baseline, CTC and control communities did not differ in their use of a science-based approach to prevention (Hawkins et al. 2008b). The CTC Milestones and Benchmarks Survey was used to track progress in implementation of core components of the CTC prevention system. In each intervention year, CTC communities implemented an average of 90 % of the key features of CTC, including developing a community board, prioritizing risk and protective factors, selecting tested and effective preventive interventions from the *CTC Prevention Strategies Guide*, implementing selected prevention programs with fidelity, and periodically assessing risk and protective factors and child and adolescent well-being through surveys of students (Fagan et al. 2009; Quinby et al. 2008). Control communities did not make this progress over time in completing CTC milestones and benchmarks, implementing scientifically proven prevention programs, and monitoring program impacts (Arthur et al. 2010). CTC communities were able to sustain these gains 20 months after study-provided training and technical assistance ended (Gloppen et al. 2012).

Faithful Implementation of Tested and Effective Prevention Programs. Over the course of the trial, the 12 CTC communities demonstrated faithful implementation of 17 different school-based, after-school, and parenting interventions (shown in Fig. 5) selected from a menu of 39 possible tested and effective programs for fifth-through ninth-grade students contained in the *CTC Prevention Strategies Guide* (Fagan et al. 2011a, 2008, 2009). On average, CTC communities implemented 2.75 tested and effective prevention programs per year (range: 1–5). High rates of fidelity were achieved consistently over time with respect to adherence to program objectives and core components (average = 91 %–94 % per year) and dosage (number, length, and frequency of intervention sessions; average = 93 %–95 % per year) and were sustained 2 years after CYDS support ended. Although control communities did not implement as many tested and effective prevention programs as CTC communities, they, too, reported implementing tested, effective prevention programs during the intervention and 2 years into the sustainability phase of the study. However, compared to CTC communities, control communities did not reach as many children and parents, nor were they as likely to monitor prevention program implementation quality during the intervention and sustainability periods (Fagan et al. 2011a).

Prevention Service System Transformation. CTC's theory of change illustrated in Fig. 1 suggests that several characteristics of the prevention service system in communities are important to achieving desired improvements in youth risk, protection, and problem behavior. By the third year of the intervention, important signs of system transformation were observed in CTC but not control communities. Key leaders in CTC communities reported a significantly higher stage of adoption of science-based prevention in their communities, meaning that they had a greater understanding of major prevention science concepts like risk and protection, and reported that their communities empirically assessed community levels of risk and protection, used tested and effective prevention practices, and monitored prevention program implementation and outcomes more than did key

Communities That Care, Table 1 Major findings from the Community Youth Development Study

	Prevention system outcomes[a]			
	Randomized controlled trial phase	Related references	Sustainability phase[b]	Related references
	CTC communities:		CTC communities:	
Prevention system implementation	Implemented >90 % of CTC prevention system milestones and benchmarks, significantly more than control communities	Arthur et al. (2010), Fagan et al. (2009), Quinby et al. (2008)	Eleven surviving coalitions implemented significantly more milestones and benchmarks than controls 20 months after support ended	Gloppen et al. (2012)
Adoption of science-based approach to prevention	Key leaders showed significantly greater understanding of prevention science concepts, assessment of community levels of risk and protection, and implementation and monitoring of tested and effective prevention programs	Brown et al. (2011)	Sustained significantly greater adoption 18 months after support ended	Rhew et al. (2013)
Implementation of evidence-based prevention programs	Achieved high-fidelity implementation (>90 % adherence/year to objectives, core components, and dosage) of 17 different evidence-based school-based, after-school, and parenting programs	Fagan et al. (2011a), Fagan et al. (2008), Fagan et al. (2009)	Sustained high-fidelity implementation 2 years after the randomized controlled trial ended	Fagan et al. (2011a)
Implementation monitoring	Were more likely to monitor prevention program implementation quality	Fagan et al. (2008)	Continued to be more likely to monitor prevention program implementation quality 2 years after support ended	Fagan et al. (2011a)
Prevention program reach	Implemented more evidence-based prevention programs and reached more children and parents	Fagan et al. (2011a)	Sustained greater implementation and reach 2 years after support ended	Fagan et al. (2011a)
	Youth risk and behavior outcomes[a]			
	Eighth grade – randomized controlled trial phase	Related references	Tenth grade – sustainability phase	Related references
	Youth exposed to CTC:		Youth exposed to CTC:	
Risk factors	Reported significantly lower growth in risk factors targeted for preventive intervention	Hawkins et al. (2008a, 2009)	Reported significantly lower levels of targeted risk	Hawkins et al. (2012)
Delinquency initiation	Were 24 % less likely to initiate delinquent behavior	Hawkins et al. (2008a, 2009)	Were 18 % less likely to initiate delinquent behavior	Hawkins et al. (2012)
Alcohol use initiation	Were 32 % less likely to initiate alcohol use	Hawkins et al. (2009)	Were 29 % less likely to initiate alcohol use	Hawkins et al. (2012)
Tobacco use initiation	Were 33 % less likely to initiate cigarette smoking	Hawkins et al. (2009)	Were 28 % less likely to initiate cigarette smoking	Hawkins et al. (2012)
Prevalence of delinquency, violence, and substance use	Were significantly less likely to use alcohol in the past month, binge drink in the past 2 weeks, commit fewer delinquent behaviors in the past year	Hawkins et al. (2009)	Were significantly less likely to commit delinquent or violent behavior in the past year	Hawkins et al. (2012)

(*continued*)

Communities That Care, Table 1 (continued)

	Prevention system outcomes[a]			
	Randomized controlled trial phase	Related references	Sustainability phase[b]	Related references
Universal effects	Generally benefited equally from CTC, regardless of differences in gender, baseline risk, early substance use, or early engagement in delinquency	Oesterle et al. (2010)		
Benefit-cost analysis	CTC was estimated to return $4,259 per youth in net present benefit or $5.30 per $1.00 invested	Kuklinski et al. (2012)		

[a] Communities were matched in pairs within state on population, racial and ethnic diversity, economic indicators, and crime rates. They were randomly assigned by coin toss to the CTC or control condition. At baseline, CTC and control communities did not differ in their knowledge and use of science-based prevention. CTC and control community youth did not differ with respect to levels of risk, protection, and key outcome variables including substance use and delinquency. Youth risk and behavior outcomes were from a panel of 4,407 youth initially constituted in Grade 5 and surveyed annually through Grade 10

[b] Prevention system sustainability phase outcomes are reported at latest measurement

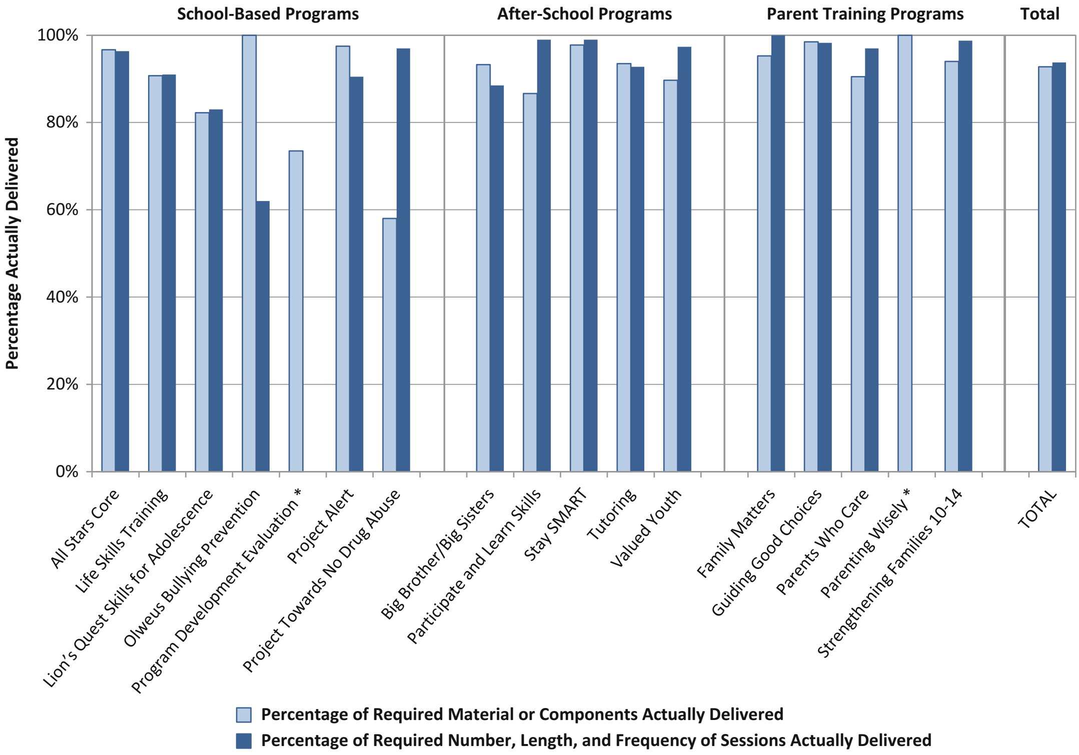

** No session fidelity forms submitted.*

Communities That Care, Fig. 5 CTC communities implemented prevention programs with fidelity (average results over four intervention years) leaders from control communities (Brown et al. 2011). Adoption of a science-based approach to prevention is hypothesized to be a key to achieving better youth outcomes through CTC. Key leaders in CTC communities also were willing to allocate a greater percentage of funding for prevention programs than key leaders in control communities, providing some evidence of increased community support for prevention in CTC communities. Although levels of collaboration did not differ in CTC and control communities at the third year of intervention, high levels of collaboration were observed in both. Effects on adoption of a science-based approach to prevention and desired prevention funding were sustained 18 months after the intervention phase of the Community Youth Development Study ended, and, at that point, key community leaders in CTC communities reported stronger community norms against adolescent drug use than did key leaders in control communities (Rhew et al. 2013).

CTC Leads to Sustained Prevention of Youth Risk, Delinquency, and Substance Use

CTC communities each prioritized 2–5 risk factors to target with tested and effective prevention programs, as shown in Table 2. Effects on youth risk and problem behaviors were first observed in Grade 7, after less than 2 full years of intervention and earlier than CTC's theory of change suggested. Effects have been sustained through Grade 10, 1 year after the trial's intervention phase ended.

Effects on Risk Factor Exposure. Longitudinal panel youth in CTC and control communities reported similar levels of targeted risk in Grade 5, when the intervention began (Hawkins et al. 2008a), but targeted risk exposure grew more slowly for youth in CTC communities between Grade 5 and Grade 10 (Hawkins et al. 2012). Significantly lower levels of targeted risk were first reported by CTC panel youth 1.67 years into the intervention, in Grade 7, (Hawkins et al. 2008a) and have been sustained in CTC panel youth reports through Grade 10.

Effects on the Initiation of Delinquency and Substance Use. Panel youth from CTC and control communities also reported similar levels of delinquency, alcohol use, and cigarette smoking at Grade 5 baseline. However, between Grades 5 and 10, CTC had significant effects on the initiation of these behaviors. Significant reductions in the initiation of delinquency were first observed in the spring of Grade 7 (Hawkins et al. 2008a). Panel youth from CTC communities were 25 % less likely than panel youth from control communities to initiate delinquent behavior, and they remained 24 % less likely to do so in Grade 8 (Hawkins et al. 2009). Significantly lower delinquency initiation rates were sustained through Grade 10 (Hawkins et al. 2012), when panel youth from CTC communities were 18 % less likely to initiate delinquency than panel youth from control communities.

Preventive effects on alcohol use and cigarette use were first observed in the spring of Grade 8, 2.67 years after intervention programs were implemented. Grade 8 youth from CTC communities were 32 % less likely to initiate alcohol use and 33 % less likely to initiate cigarette smoking than Grade 8 youth from control communities (Hawkins et al. 2009). Preventive effects were sustained through Grade 10 (Hawkins et al. 2012), when CTC panel youth were 29 % less likely to initiate alcohol use and 28 % less likely to initiate cigarette smoking than panel youth from control communities.

As shown in Fig. 6, differences in the initiation of delinquency, alcohol use, and cigarette smoking from Grade 5 through Grade 10 led to cumulatively lower rates of initiation over time: 62 % of 10th-grade youth in the panel from CTC communities had engaged in delinquent behavior compared with 70 % of 10th-grade youth in the panel from control communities; 67 % versus 75 % had initiated alcohol use; and 44 % versus 52 % had smoked cigarettes.

Reductions in the Prevalence of Delinquency, Violence, and Substance Use. CTC also significantly reduced the prevalence of youth problem behaviors in Grades 8 and 10. In Grade 8, the prevalence of alcohol use in the past month, binge drinking (five or more drinks in

Communities That Care, Table 2 Risk factors targeted for prevention by communities in the CYDS

	CTC community											
Risk factor	1	2	3	4	5	6	7	8	9	10	11	12
Community												
Laws and norms favorable to drug use									✓			
School												
Low commitment to school		✓	✓	✓		✓	✓	✓	✓		✓	✓
Academic failure				✓	✓			✓		✓	✓	
Family												
Family conflict	✓	✓					✓					
Poor family management			✓	✓						✓		✓
Parental attitudes favorable to problem behavior						✓						
Peer												
Antisocial friends	✓	✓			✓				✓	✓	✓	✓
Peer rewards for antisocial behavior	✓							✓				
Individual												
Attitudes favorable to antisocial behavior	✓					✓				✓		
Rebelliousness	✓							✓	✓			
Low perceived risk of drug use								✓				✓

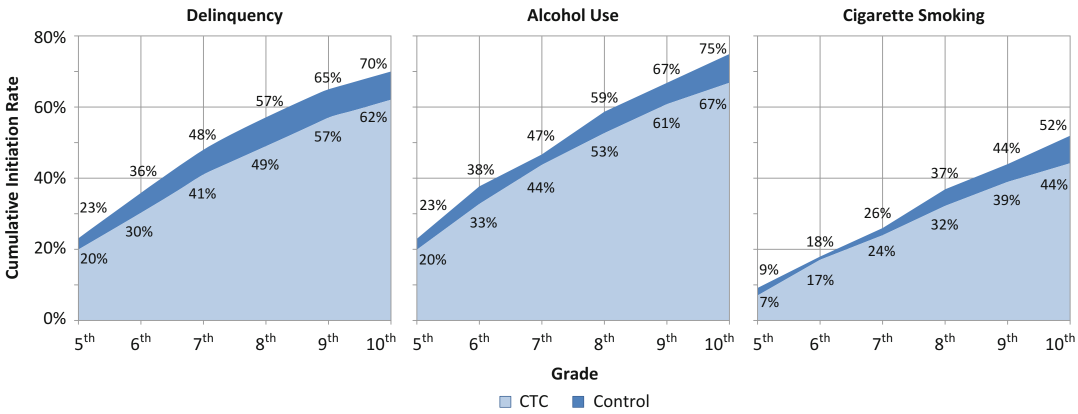

Communities That Care, Fig. 6 CTC leads to significant reductions in the cumulative initiation of delinquency and substance use

a row) in the past 2 weeks, and the variety of delinquent behaviors committed in the past year were all significantly lower in CTC panel youth compared to control community panel youth (Hawkins et al. 2009). A subset of the delinquency items was used to create a measure of violent behavior. The CYDS found significant effects of CTC in reducing the prevalence of delinquent behavior and violence in the past year in the spring of Grade 10 (Hawkins et al. 2012). Tenth-grade students in CTC communities had 17 % lower odds of reporting any delinquent behavior in the past year (t (9) = −2.33; *p* = .04; AOR = .83) and 25 % lower odds of reporting any violent behavior in the past year (t (9) = −2.51; *p* = .03; AOR = .75) compared to students in control communities.

CTC's Effects Are Universal. CTC was designed as a universal intervention, and the universality of its effects on the prevalence of substance use and delinquent behavior was tested based on outcomes observed in panel youth in the spring of eighth grade (Oesterle et al. 2010). Four groups of students were established using baseline Grade 5 data: boys and girls, high baseline risk versus all others, early substance users versus nonusers, and early engagers in delinquency vs. non-engagers. Results generally supported the universal effectiveness of the CTC intervention with respect to the prevalence of delinquent behavior and substance use within groups. There were two exceptions. The effect of CTC on reducing the prevalence of substance use in eighth grade was stronger for boys than girls, and the effect on the prevalence of delinquency was stronger for students who were non-delinquent at baseline.

Communities That Care, Table 3 CTC is a cost-beneficial intervention

	Stakeholders			
	Participants	Taxpayers	Others	Total
Per-youth benefits[a]				
Smoking	$671	$140		$812
Delinquency		$2,033	$2,405	$4,438
Total benefits	$671	$2,173	$2,405	$5,250
Per-youth costs[a]				$991
Per-youth net present benefit[a]				$4,259
Benefit-cost ratio				$5.30

[a]All benefits and costs are expressed in discounted 2004 dollars

CTC is a Cost-Beneficial Intervention

A cost-benefit analysis was undertaken to determine whether CTC is a sound investment of public dollars based on significant preventive effects on cigarette smoking and delinquency initiation found in Grade 8 (Kuklinski et al. 2012). Findings are summarized in Table 3. CTC's long-term financial benefits from reduced initiation were compared to a conservative CTC implementation cost of $991 per youth over 5 years. Most communities spent significantly less, roughly $513 per youth. Approximately 37 % of the total went toward a CTC coordinator and the coalition; 35 % to prevention programming; 23 % to CTC training, technical assistance, and monitoring; and 4 % to miscellaneous expenses. Implementation costs were more than offset by CTC's estimated financial benefits of $5,250 per youth. The benefit-cost ratio indicates $5.30 is returned to society for every $1.00 invested, evidence that CTC is a cost-beneficial investment.

Challenges to Implementing CTC

Like other community mobilization strategies, CTC can be challenging for communities to implement and sustain. In our experience, communities need to overcome five types of challenges to successfully own and operate CTC.

Be Ready to Implement a Comprehensive Approach to Prevention. Communities must be ready to initiate comprehensive evidence-based prevention in order to fully implement CTC. This requires readiness along several key dimensions. Differing views among stakeholders about the most effective approach to addressing youth problems must be considered. If particular groups advocate only exclusionary approaches such as arrest and incarceration, or exclusive reliance on treatment approaches, they may need further education about the potential effectiveness of prevention prior to engaging in the CTC process. If communities have an unsuccessful history of collaboration, characterized by conflict and competition for scarce resources, building trust through collaborative efforts to achieve smaller specific goals, like closing a crack house or getting a crosswalk installed on a busy street, may be necessary prior to initiating CTC.

Recruit and Engage Key Leaders. If CTC is to succeed, key community leaders who control resources and shape opinions, like the mayor, school superintendent, and police chief, must decide to bring the CTC system into their community. Their recruitment and engagement is essential to CTC's success. Without ongoing "buy-in" from key leaders and influential champions who have access to resources, CTC will not be viable because local resources will need to be committed and often redirected to prevention.

Evidence-Based Prevention Often Competes with Usual Practice. The board needs to advocate for a risk- and protection-focused approach to preventing problem behavior and for the use of tested and effective preventive interventions to address local priorities. At times, this can mean giving up programs that are rooted in community tradition or are local favorites if they lack research evidence of effectiveness.

Faithfully Implement and Evaluate. Achieving prevention goals requires high-quality implementation of the CTC system and prevention programs and monitoring outcomes over time. Monitoring community-wide implementation and making needed adjustments can be a challenge when programs are implemented in multiple classrooms, agencies, and community settings. Evaluation of changes in risk, protection, and outcomes at the community level is also a challenge. At the local level, it may be several years before community-wide reductions are achieved, requiring long-term commitment to the CTC approach.

Ensure Adequate Resources. CTC coalitions and community coordinators represent critical human resources needed to run CTC. Maintaining board engagement and recruiting and orienting new members to CTC and the coalition's progress to date are essential to long-term sustainability. Securing ongoing funding for coalitions and coordinators, prevention programs, and training and technical assistance can be a challenge for coalitions. In the CYDS and in a second quasi-experimental study of CTC in Pennsylvania (Feinberg et al. 2008), coalitions were aided in initial implementation efforts by seed money for coalition organizing and training and implementing chosen effective prevention approaches. Many coalitions in both studies have demonstrated success in generating ongoing resources for CTC. Those most likely to obtain funding were well functioning and planned for financial sustainability as part of their regular operations (Feinberg et al. 2008).

Steps to Broader Dissemination of CTC

All manuals and materials needed to implement CTC have been placed in the public domain by the US Substance Abuse and Mental Health Services Administration of the Department of Health and Human Services and are freely available at http://www.communitiesthatcare.net. CTC is currently being implemented in communities across the United States and in countries including Australia, Canada, Croatia, Germany, the Netherlands, and the United Kingdom. The CTC tests in both the Community Youth Development Study and the Pennsylvania quasi-experimental study were conducted with certified providers of training and technical assistance, as well as monitoring and oversight of implementation, including reviewing and providing feedback on annual plans. State and federal governments, local philanthropic organizations, the United Way, and other local funders could assist in broader dissemination of CTC by supporting training, technical assistance, and monitoring of processes and outcomes. Efficiencies can be achieved by offering CTC training through state governments, who can then train state prevention workers and contractors to deliver CTC training and technical assistance to local communities. A national CTC training and certification center could support the training of state and community personnel in CTC and share lessons learned from the implementation of CTC.

Conclusion

Policymakers and researchers increasingly recognize the need for clear and compelling evidence about how best to prevent prevalent and costly problem behaviors in youth. Delinquency and substance abuse are problems that begin in late childhood and adolescence, often with lifelong consequences. The results of the Community Youth Development Study show that *Communities That Care*, a coalition-based prevention system employing a risk- and protection-focused approach, is a cost-beneficial, effective way to prevent delinquency, violence, and tobacco and alcohol use in youth. Intervention communities were able to faithfully implement the CTC prevention system and evidence-based programs designed to address community-specific prevention priorities, and they sustained this success after the intervention phase of the Community Youth Development Study was over. These efforts were related to significant, meaningful, and sustained reductions community wide in the percentage of youth in intervention communities initiating delinquent behavior, alcohol use, and cigarette smoking, and in the prevalence of current involvement in smoking, delinquency, and violence when compared to youth from control communities.

Acknowledgments The Community Youth Development Study was supported by research grant R01 DA015183-03 from the National Institute on Drug Abuse, with co-funding from the National Institute of Child Health and Human Development, the National Cancer Institute, the National Institute of Mental Health, the Center for Substance Abuse Prevention, and the National Institute on Alcohol Abuse and Alcoholism. The content of this entry is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

Related Entries

- ▶ Drug Abuse and Alcohol Dependence Among Inmates
- ▶ Poverty, Inequality, and Area Differences in Crime
- ▶ Randomized Block Designs
- ▶ Randomized Experiments in Criminology and Criminal Justice
- ▶ School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis
- ▶ School Bullying as a Risk Factor for Later Criminal Offending
- ▶ Tobacco Abuse in Correctional Settings

Recommended Reading and References

Arthur MW, Briney JS, Hawkins JD, Abbott RD, Brooke-Weiss BL, Catalano RF (2007) Measuring risk and protection in communities using the Communities That Care Youth Survey. Eval Program Plann 30:197–211

Arthur MW, Hawkins JD, Brown EC, Briney JS, Oesterle S, Abbott RD (2010) Implementation of the Communities That Care prevention system by coalitions in the Community Youth Development Study. J Community Psychol 38:245–258

Brown EC, Graham JW, Hawkins JD, Arthur MW, Baldwin MM, Oesterle S, Briney JS, Catalano RF, Abbott RD (2009) Design and analysis of the Community Youth Development study longitudinal cohort sample. Eval Rev 33:311–334

Brown EC, Hawkins JD, Arthur MW, Briney JS, Fagan AA (2011) Prevention service system transformation using Communities That Care. J Community Psychol 39:183–201

Catalano RF, Hawkins JD (1996) The social development model: a theory of antisocial behavior. In: Hawkins JD (ed) Delinquency and crime: current theories. Cambridge University Press, New York, pp 149–197

Catalano RF, Haggerty KP, Hawkins JD, Elgin J (2011) Prevention of substance use and substance use disorders: the role of risk and protective factors. In: Kaminer Y, Winters KC (eds) Clinical manual of adolescent substance abuse treatment. American Psychiatric Publishing, Washington, DC, pp 25–63

David-Ferdon C, Hammond WR (2008) Community mobilization to prevent youth violence and to create safer communities. Am J Prev Med 34:S1–S2

Fagan AA, Hanson K, Hawkins JD, Arthur MW (2008) Implementing effective community-based prevention programs in the Community Youth Development Study. Youth Violence Juvenile Justice 6:256–278

Fagan AA, Hanson K, Hawkins JD, Arthur MW (2009) Translational research in action: implementation of the Communities That Care prevention system in 12 communities. J Community Psychol 37:809–829

Fagan AA, Arthur MW, Hanson K, Briney JS, Hawkins JD (2011a) Effects of Communities That Care on the adoption and implementation fidelity of evidence-based prevention programs in communities: results from a randomized controlled trial. Prev Sci 12:223–234

Fagan AA, Hawkins JD, Catalano RF (2011b) Engaging communities to prevent underage drinking. Alcohol Res Health 34:167–174

Feinberg ME, Bontempo DE, Greenberg MT (2008) Predictors and level of sustainability of community prevention coalitions. Am J Prev Med 34:495–501

Gloppen KM, Arthur, MW, Hawkins, JD, Shapiro, VB (2012) Sustainability of the Communities That Care prevention system by coalitions participating in the Community Youth Development Study. J Adolesc Health 51:259–264

Hallfors D, Cho H, Livert D, Kadushin C (2002) Fighting back against substance abuse: are community coalitions winning? Am J Prev Med 23:237–245

Hawkins JD, Catalano RF Jr et al (1992a) Communities That Care: action for drug abuse prevention, 1st edn. Jossey-Bass, San Francisco

Hawkins JD, Catalano RF, Miller JY (1992b) Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: implications for substance-abuse prevention. Psychol Bull 112:64–105

Hawkins JD, Brown EC, Oesterle S, Arthur MW, Abbott RD, Catalano RF (2008a) Early effects of Communities That Care on targeted risks and initiation of delinquent behavior and substance use. J Adolesc Health 43:15–22

Hawkins JD, Catalano RF, Arthur MW, Egan E, Brown EC, Abbott RD, Murray DM (2008b) Testing Communities That Care: the rationale, design and behavioral baseline equivalence of the Community Youth Development Study. Prev Sci 9:178–190

Hawkins JD, Oesterle S, Brown EC, Arthur MW, Abbott RD, Fagan AA, Catalano RF (2009) Results of a type 2 translational research trial to prevent adolescent drug use and delinquency: a test of Communities That Care. Arch Pediatr Adolesc Med 163:789–798

Hawkins JD, Shapiro VB, Fagan AA (2010) Disseminating effective community prevention practices: opportunities for social work education. Res Soc Work Pract 20:518–527

Hawkins JD, Oesterle S, Brown EC, Monahan KC, Abbott RD, Arthur MW, Catalano RF (2012) Sustained decreases in risk exposure and youth problem behaviors after installation of the Communities That Care prevention system in a randomized trial. Arch Pediatr Adolesc Med 166:141–148

Johnston LD, O'Malley, PM, Bachman, JG (2000) The Monitoring the Future national survey results on adolescent drug use: overview of key findings, 1999 (NIH Publication No. 00-4690). National Institute on Drug Abuse, Bethesda

Kuklinski MR, Briney JS, Hawkins JD, Catalano RF (2012) Cost-benefit analysis of Communities That Care outcomes at eighth grade. Prev Sci 13:150–161

Merzel C, D'Affliti J (2003) Reconsidering community-based health promotion: promise, performance, and potential. Public Health Matters 93:557–574

Oesterle S, Hawkins JD, Fagan AA, Abbott RD, Catalano RF (2010) Testing the universality of the effects of the Communities That Care prevention system for preventing adolescent drug use and delinquency. Prev Sci 11:411–424

Patton GC, Coffey C, Sawyer SM, Viner RM, Haller DM, Bose K, Vos T, Ferguson J, Mathers CD (2009) Global patterns of mortality in young people: a systematic analysis of population health data. Lancet 374:881–892

Quinby RK, Fagan AA, Hanson K, Brooke-Weiss B, Arthur MW, Hawkins JD (2008) Installing the Communities That Care prevention system: implementation progress and fidelity in a randomized controlled trial. J Community Psychol 36:313–332

Rhew IC, Brown EC, Hawkins JD, Briney JS (2013) Sustained effects of Communities That Care on prevention service system transformation. Am J Publ Health 103:529–535

Ringwalt C, Vincus AA, Hanley S, Ennett ST, Bowling JM, Rohrbach LA (2009) The prevalence of evidence-based drug use prevention curricula in U.S. middle schools in 2005. Prev Sci 10:33–40

Community Control: Criminality, Community Social Organization, Cultural Organization, Deviant Opportunities, Directs Controls

▸ Control Theory

Community Courts

Rachel Swaner
Center for Court Innovation, New York, NY, USA

Overview

A community court, often called a neighborhood or community justice center, is a neighborhood-focused court that applies a problem-solving approach to local crime and safety concerns. Emphasizing improved outcomes for offenders (e.g., lower recidivism) and for communities (e.g., safer neighborhoods), community courts strive to prevent crime by addressing its underlying causes (Berman and Feinblatt 2005; Rottman and Casey 1999). Most community courts handle low-level criminal cases arising in a set of target neighborhoods. Some community courts are multi-jurisdictional, handling housing disputes, juvenile delinquency matters, or other case types in the family court system as well. Community courts are typically located in a separate courthouse situated within the targeted neighborhoods, although some community courts do operate out of a centralized "downtown" court.

In 1993, the first community court was started in Manhattan, New York City. It was opened to address quality-of-life crimes such as prostitution, illegal vending, vandalism, and minor drug possession in the central business district of Midtown, which included Times Square, then the home of much low-level crime. Since then, over 60 community court projects have opened worldwide. In the United States alone, there about 40, while there are at least 17 in South Africa, 13 in England and Wales, and one each in Australia and Canada (Henry and Kralstein 2011). These courts seek to bring a neighborhood-focused, problem-solving approach to local crime, social disorder, and safety concerns, focusing on alternatives to jail and fines, instead giving sanctions such as community service, social service, and treatment for drug addiction.

This entry outlines the principles and theories of community courts and summarizes the current research regarding the effectiveness of community courts on expanding sentencing options, improving compliance rates for court mandates, lowering recidivism, and reducing costs. As will be discussed in further detail, the evaluation literature shows that community courts do indeed dispense more types of sanctions (e.g., community and social services) than regular courts; though jail is still used, it is mostly as a resentencing option. Mandate completion rates have shown to be higher than in more traditional courts, though more research may be needed in this area. Findings on differences in recidivism rates have been mixed; while some studies have shown lower recidivism rates for community court participants, other studies have failed to show an effect, which may in part be due to small sample size or no true comparison group. Finally, studies have shown mixed results on cost savings as well.

Background Description

One of the main components of community courts is the use of alternative sanctions, the goals of which are twofold: first, to pay back the community that may have been harmed by the offense through visible community service work projects in the neighborhood, such as cleaning local parks and painting over graffiti. Second, social services such as drug treatment and job training are assigned to link defendants with programs that address the problems that may have played a role in leading them to commit crimes. This combination of accountability and help seeks to reduce chronic offending. In order to understand local problems and priorities, community court staff also commonly venture out into local neighborhoods, attending community meetings and serving as a convener and problem-solver between criminal justice agencies (e.g., police, prosecutors, or probation) and the citizens they serve.

However, there is no one concrete model for community courts, which makes it difficult to develop one standard definition. But as Berman (2010) outlined, community courts are usually guided by the following six key principles that differentiate them from more traditional courts:

1. *Enhanced Information* – Community courts seek to make as much information as possible available about the individual defendant and the community context of the crime at the defendant's first court appearance. This allows for improved decision-making by judges, attorneys, and other justice officials, enabling them to match the defendant to appropriate sanctions and services.
2. *Community Engagement* – Community courts seek to assign offenders to community service projects in places where neighbors can see what they're doing, welcome observers and visitors, make information about courtroom activities easily accessible to the public, and involve the community in addressing crime-related problems.
3. *Collaboration* – Government agencies, social service agencies, and other community organizations work together under a single roof, where the physical proximity facilitates closer and more coordinated working relationships.
4. *Individualized Justice* – Community courts link defendants to individually tailored, community-based sanctions such as community service, job or educational training, and drug treatment.
5. *Accountability* – Community courts insist on regular and strict compliance monitoring, with clear consequences for noncompliance. Additionally, community courts often escort offenders immediately to the place where they will receive their community or social service assignment, greatly increasing the likelihood that they will comply with their

sanction. These mechanisms work to improve the accountability of the offenders.

6. *Outcomes* – Data collection and analysis are important tools at community courts. Data help to answer questions related to case processing (how many and how quickly), status and compliance, as well as to perceptions of the court, procedural fairness, and local quality of life.

Expanding on the sixth point above, community courts usually follow one or more of the following theories of change for attempting to achieve these outcomes, as outlined by Lee et al. (2013).

Procedural Justice and the Social Construction of Legitimacy

As Tyler (2001) suggested, public trust and confidence in police and courts is not related to performance or outcomes but to how fairly people feel they were treated. The idea of fairness vis-à-vis justice was clearly defined by John Rawls (1971) in his *A Theory of Justice*. He said that perfect procedural justice had two characteristics: (1) an independent criterion for what constitutes a fair or just outcome of the procedure and (2) a procedure that guarantees that the fair outcome will be achieved. Tyler (1990) put forth that citizens generally hold favorable views towards institutions that are perceived as unbiased, while holding negative views of those that are believed to be partisan or discriminatory. Procedural fairness is present when people perceive that they are experiencing the following in their interaction with judges (Tyler 2004: 443–47):

- ***Respect***: People are treated with dignity and their rights are respected.
- ***Neutrality***: Honest and impartial decision-makers base their decisions on facts.
- ***Participation***: Each party has an opportunity to express his or her viewpoint to the decision-maker, and conversely, decision-makers clearly articulate the basis of their decisions in ways that litigants can understand (without jargon).
- ***Trustworthiness***: Decision-makers appear benevolent, caring, motivated to treat parties fairly, and sincerely concerned about people.

Research has endeavored to assess the meaning of procedural justice for those who come in contact with the criminal justice system. As Wissler (1995), Lind and Tyler (1998), and Paternoster et al. (1997) have demonstrated, people are more willing to accept decisions when they think criminal justice officials or legal institutions are acting fairly. Similarly, Tyler (1990) has argued that citizens generally hold favorable views towards institutions that are perceived as unbiased, while holding negative views of those that are believed to be partisan or discriminatory. Elsewhere, Tyler (2001) has suggested that public trust and confidence in police and courts is related to how fair people feel they were treated; performance and outcomes are secondary factors.

Research has also attempted to understand the relationship between procedural justice and the public's law-related behavior, as there is growing concern that perceived injustice itself causes criminal behavior (LaFree 1998; Mann 1993; Russell 1998; Tyler 1990). Tyler and Huo (2002) have proffered that when citizens perceive justice system agencies to be fair, they are more likely to comply with the law, legal authorities, and court mandates, increasing institutional confidence. Compliance, Tyler (2004: 307) further explains, is linked to legitimacy – "the property that a rule or an authority has when others feel obligated to defer voluntarily" – and that to the extent that people regard the courts as legitimate, they are more willing to accept the directives of the courts.

To put it another way, legitimacy develops out of the use of fair procedures and the provision of respectful treatment. Procedural justice, then, is the key antecedent to legitimacy, voluntary compliance, and lower recidivism: the greater the legitimacy, the less the likelihood of defiance, hostility, and resistance to laws, legal authorities, and legal institutions.

Community Involvement

In addition to building legitimacy through procedural fairness, community courts seek to strengthen ties to the community so that residents and merchants see themselves as allies of the court and are therefore more likely to reinforce social norms that the court espouses. When the whole community comes together to reinforce norms that do not accept crime and disorder, neighborhoods will become safer as crime decreases. When community courts actively involve the community in decision-making and give back to the community in some tangible way, there will be improved perceptions of the court. Community courts engage the community in four primary ways:

- ***Visible community service***: By sanctioning offenders to community service work in places where neighbors can identify them as offenders and see what they are doing, community members will better understand the court's role in restoring the community that had been harmed. These community service projects usually involve work such as graffiti removal or trash cleanup, serving as a tangible reminder of the offender's membership in and responsibility to the group (see Herrschaft 2012).
- ***Giving the community voice***: By regularly asking the community what their needs and concerns are – through neighborhood surveys (see Swaner 2010), advisory boards, and community meetings – community courts are better able to offer programs and services that are relevant and address the desires of local residents and merchants.
- ***Collaboration***: Allowing community-based organizations to have space to run programming in the community court building, supporting community events organized by local organizations, and partnering with local groups to offer comprehensive programming and co-sponsoring events helps develop a strong neighborhood identity as well as community-based informal social controls that research links to crime reduction.
- ***Voluntary programming***: Community courts offer voluntary programs for all persons living within the court's jurisdiction – regardless of whether or not they have a court case. These programs can include after-school programs for local youth, job training, educational classes, an open computer lab, health and mental health services, and mediation sessions. Offering these programs helps the court to be seen as an asset to all members of the community, not just a place for those who commit crimes.

Deterrence

In 1982, Kelling and Wilson's now famous "broken windows" theory was introduced to the criminology world, and it is the theoretical underpinning of many community courts. The basic premise behind this theory is that when the physical and visible conditions of a neighborhood – graffiti, illegal dumping, public intoxication – are not controlled and monitored, it produces more social disorder, and, eventually, these neighborhood problems escalate to more severe crimes. Community courts focus on trying to fix the problems when they are small so that further petty crime and antisocial behavior will be deterred and, hence, will not lead to more severe crimes. In this way, community courts tend to focus on cleaning up minor "quality-of-life" crimes under the assumption that this will lead to reductions in other types of crime as well.

Focusing on "broken windows" also helps to reinforce that a particular community does not accept crime. Looking at the surrounding environment – whether it is clean or dirty – gives cues to individuals as to the social norms of the community. If it is a place that has broken windows and other visible signs of disorder, this is a signal to potential lawbreakers that any crime they commit is likely to be overlooked. If, on the other hand, it is a place that does not have broken windows, this is a signal that the community does not accept crime, and there would be certain punishment for committing a criminal act, serving as a deterrence to violating social norms. This means that in order to design an effective deterrent, community courts must consider the *certainty* of punishment. Research indicates that the certainty of punishment is the most important element in the potential criminal's decision whether or not to commit a crime (Nagin and Pogarsky 2001). Criminal behavior is not deterred if lawbreakers do not think they will be caught or given meaningful punishment if they are.

Accordingly, community courts replace sentences that lack ongoing obligations of any kind with meaningful sanctions for even the most minor of offenses. They have an intense focus on misdemeanor offenses that create visible signs of disorder and also focus on sanctioning offenders to community service that repairs those conditions of disorder.

Intervention

The intervention theory of change assumes that sanctioning offenders to social service interventions will help them address the underlying problems that may have caused them to commit crimes (or have other legal problems) in the first place. For example, drug offenders may have an underlying addiction that causes them to offend. Assigning them to a drug treatment modality such as outpatient/detox or residential treatment may help them overcome the addiction that led them to commit a crime. Other offenses such as public intoxication, driving under the influence, and drug possession also frequently result from addiction to alcohol or drugs. Providing people with appropriate interventions such as short- or long-term treatment, anger management workshops, job training programs, or GED programs may help them break cycles of recurrent criminal behavior caused by social dislocations such as drug addiction, lack of employment opportunities, and educational disadvantages. The same principles animate community court practices when handling non-criminal cases. For instance, landlord-tenant disputes that escalate to a case in housing court may reflect the tenant's need for employment assistance or for help accessing public benefits. Thus, community courts that handle housing matters commonly provide tenants with resources and referrals that may help them meet their payment obligations.

State of Art

Due in part to the recent proliferation of community courts, there is scant research literature on their impact. Comprehensive impact evaluations have been carried out in only Midtown, Manhattan (Hakuta et al. 2008; Sviridoff et al. 2000a, b, 2001); Hennepin County, MN (Eckberg 2001; Weidner and Davis 2000); Philadelphia, PA (Cheesman et al. 2009, 2010); Yarra, Australia (Ross et al. 2009); and Red Hook, Brooklyn (Lee et al. 2013). Methodologies applied in these sites included process evaluations; quasi-experimental studies testing impacts on case processing, outcomes, offender compliance, and/or recidivism; cost-benefit analyses; community surveys; and ethnographic observations and interviews with local offenders. The results of these impact evaluations and other smaller-scale evaluations are summarized by key themes below.

Expanded Sentencing Options

Confirming the central role of diversified sentencing options, a global survey of community courts (Karafin 2008) found that 92 % of community courts routinely mandate defendants to community service, and 84 % mandate defendants to social services, including treatment readiness classes (64 %), individual counseling (64 %), job skills (64 %), life skills (56 %), anger management (52 %), and substance abuse treatment (48 %). Two separate evaluations of the Midtown Community Court – one focusing on its early years and the other on recent impacts – both found that the court made significantly greater use of "alternative" sentences than the centralized Manhattan court. These studies also found that Midtown made less use of jail and less use of "walks," defined as sentences such as fines or time served that lack any ongoing obligation (Hakuta et al. 2008; Sviridoff et al. 2001).

Yet, despite sentencing a lower percentage of its defendants to jail, both studies found that when Midtown did use jail, the resulting sentences were longer on average. In addition, one evaluation found that Midtown was more likely to impose meaningful jail time as a "secondary sanction" due to noncompliance with what was initially an alternative sentence (Sviridoff et al. 2001). These dynamics meant that Midtown did not ultimately produce a significant net reduction in jail days served by its defendants.

Similar results were found in an evaluation of the Red Hook Community Justice Center (RHCJC) (Lee et al. 2013). Red Hook grants fewer "walks" than the downtown court.

Convicted RHCJC defendants are much less likely than downtown defendants to receive a conditional discharge with no real conditions other than staying out of trouble (15 % of convictions vs. 26 %) or a sentence of time served (3 % of convictions vs. 32 %). Red Hook defendants are also less likely than downtown defendants to have their cases dismissed (17 % of cases vs. 21 %) or to receive an adjournment in contemplation of dismissal (which involves a promise of dismissal within 6 months) (ACD) without community service or social service conditions attached (31 % of ACDs vs. 76 %). Of those defendants who do receive a sanction, however, RHCJC sends a larger share to community service and social service programs and a smaller proportion to jail – largely using jail as a resentence option with those defendants who are initially noncompliant.

Though only 1 % of Red Hook defendants receive jail as an initial sentence, when used as a resentence, the average number of days is longer than downtown (75.1 days, compared with 3.88 days). The net effect of this difference in the use of "secondary jail" sanctions is that fewer RHCJC cases ultimately receive a jail sentence (7 % vs. 17 %), but RHCJC cases ultimately average more time sentenced to jail (4.75 vs. 3.06 days among *all* defendants or 82.3 vs. 39.9 days among just those defendants who receive a jail sentence of any length).

Compliance with Court Conditions

As compared with the local centralized courts, the Midtown and Hennepin County community courts produced significant increases in offender compliance with community service mandates, from 50 % to 75 % in Midtown (Sviridoff et al. 2000a, b, Sviridoff et al. 2001) and 29–54 % in Hennepin (Weidner and Davis 2000). At the Red Hook Community Justice Center, between 2000 and 2009, 80 % of defendants mandated to community service and 69 % of defendants assigned to social service sanctions other than long-term drug treatment successfully completed their mandates.

At the Neighbourhood Justice Centre (NJC) in Yarra, successful completion of community-based orders (CBOs) over the course of 1 year (September 2008–August 2009) at the NJC was 75 %, compared to a 65 % successful completion rate for all CBOs statewide. For community service by offenders, the completion rate for assigned hours is marginally higher than the state average (88 % vs. 85 %), and each NJC offender performs an average of 105 h of unpaid work compared with 68 h for Victorian offenders in general (Ross et al. 2009).

An evaluation of the East of the River Community Court (ERCC) in Washington, D.C., found that between 2007 and 2009, 60 % of defendants who stayed on the ERCC calendar successfully completed the ERCC program (Westat 2012).

Effects on Recidivism

An evaluation of the Midtown Community Court failed to detect an effect on reoffending by individual offenders but did detect a drop in prostitution and illegal vending crime in the Midtown neighborhood, perhaps due to a displacement effect (Sviridoff et al. 2001). A study of recidivism of misdemeanants convicted in the Philadelphia Community Court (Allen and Schulman 2009) was conducted on offenders who completed all of the conditions of their sentence in 2007, reporting an overall 9 % recidivism rate, with lower rates reported for offenders that were required to complete a program. Cases removed from the Philadelphia Community Court before completing all of the terms of their sentence reported a higher 17 % recidivism rate. No recidivism data on a counterfactual comparison group, however, was conducted in that study.

Two evaluations, respectively of the Seattle (WA) and Liverpool (England) community courts, showed mixed results. In Seattle, there was not a significant difference in the probability of rearrest, but there was a smaller average number of rearrests among those processed in the community court than among those processed in a centralized court during a pre-implementation period (Nugent-Borakove 2009). Similarly, in Liverpool, there was not an impact on the reconviction rate, but the community court produced a small reduction in the total number of reoffenses (falling just short of statistical significance; see Jolliffe and Farrington 2009).

A recidivism study for the community court in Yarra showed that offenders sentenced by the Neighbourhood Justice Centre who received services at the center showed a lower rate of reoffending than a comparison sample of offenders sentenced at other courts (34 % vs. 41 %). However, this estimate was based on a relatively small sample of cases (100) and the difference between the NJC and comparison group was not statistically significant. In addition it was not possible to take into account any variations in the risk of reoffending associated with the NJC and comparison offenders (Ross et al. 2009).

At the East of the River Community Court, a survival analysis showed that recidivism was significantly lower among the ERCC diversion program and treatment court defendants than among the Metropolitan Police Department 5th District defendants. However, these findings only included data from the D.C. CourtView online records system. Analyses were conducted on the recidivism rates of only those defendants that participated in an ERCC diversion program compared to a matched group of defendants arrested in the MPD 5th District. Reoffending was 60 % lower among ERCC defendants than among MPD 5th District defendants 360 days after the case filing date. During the second follow-up period, recidivism was 42 % lower among ERCC defendants than among MPD 5th District defendants 360 days after the trigger case disposition date (Westat 2012).

A quasi-experimental evaluation of the Red Hook Community Justice Center showed that over a 1 year period, RHCJC defendants appeared less likely to be rearrested than the downtown sample (28 % vs. 31 %) and were also arrested fewer times on average (0.57 vs. 0.66). Over a 2 year period, differences between the RHCJC and the downtown court were somewhat greater. RHCJC defendants were significantly less likely than downtown defendants to be rearrested (36 % vs. 40 %) and averaged significantly fewer rearrests (0.95 vs. 1.16). An additional survival analysis demonstrated that the lower rate of recidivism for RHCJC defendants persisted over more than a 4 year maximum tracking period (Lee et al. 2013).

Cost Savings

Evaluations in Midtown, Hennepin, Yarra, and Red Hook included cost studies, but their findings differed. In Midtown, the evaluation identified approximately $1.3 million in annual savings based on a reduction in pre-arraignment detention time, reduced jail sentences on shoplifting cases (jail time was not reduced on other cases), and reduced prostitution arrests in the Midtown neighborhood (Sviridoff et al. 2001). The evaluation of the Neighbourhood Justice Center in Yarra also found that the court saved money; for every Australian dollar invested, the expected return was AUS$1.09–1.23. For the activities included in the NJC evaluation to return a positive net benefit, a change in reoffending behavior needs to be maintained for just over 4 years (Ross et al. 2009). The evaluation of the Red Hook Community Justice Center included a cost-efficiency analysis that found that the community court was cost-effective relative to the downtown court, but its cost-efficiency is reduced when it serves a small number of defendants (Lee et al. 2013). The Hennepin study, on the other hand, found that the community court was more expensive than regular case processing (Weidner and Davis 2000).

Possible Controversies

While many of the evaluations have shown positive results, it is important to understand some of the critiques that have been made of community courts.

First, community courts involve criminal justice agents and social service professionals working together in a coordinated way to best serve the offender and the community. But as Berman (2010) discussed, the guiding philosophies of criminal justice agencies and social service providers often differ in significant ways. Whereas law enforcement officials tend to promote harsher sanctions for defendants when they fail to comply, social workers, who have a better understanding of mental health and addiction, believe in giving second chances when an offender relapses.

This tension plays out in other ways as well. While community court judges have greater sanctions at their disposal, this also gives them control over assigning social service interventions to offenders. While judges are seen as being experts in law and the administration of justice, they may not have the social work training necessary to know which mental health interventions are most appropriate – and for how long – to meet offenders' needs. Community courts, then, may inappropriately extend the reach of the court into people's lives, transforming the judge into a powerful figure that transcends that of a traditional judge.

Moreover, while some social service sanctions seek to address the underlying problems that caused someone to commit a crime, many times larger socioeconomic factors – not individual pathology – are what contribute to crime. While community courts may offer job training or resume-building workshops, when there is a bad economy and a lack of available jobs, such social service sanctions will not be effective prevention measures.

Another critique is that one of the underlying program theories of community courts – "broken windows" – itself is flawed. This theory links disorder and crime prevention by using coercive social control to manage social disorder. However, just because disorder and crime are correlated, it does not mean that one causes the other. Moreover, research has shown that the link between the two is very weak. Broken window policing policy, then, tends to unfairly criminalize the poor, minorities, and homeless, and, as Malkin (2003) points out, "quality of life policing is not necessarily what residents envision as a response to their quality of life concerns" (1585).

Finally, money may be a source of controversy. Community courts may be competing with community-based social service agencies and other local organizations for limited public and private resources. Given their potentially higher level of political and cultural capital, community courts may be taking funding away from community groups who may be better suited to efficiently and effectively run the programming (see Malkin 2003). Community courts also require an initial influx of money to start-up – money to open a new building or renovate an existing one, hire new staff, and obtain necessary technology. Additionally, there are significant costs to run the social service programs that are essential to community courts. Critics might suggest that most of the key elements of community courts could be done more efficiently and with lower case management, staffing, and program costs within a centralized courthouse than a new, separate court.

Related Entries

▶ Drug Courts
▶ Mental Health Courts
▶ Problem-Solving Courts
▶ Reentry Courts
▶ Therapeutic Jurisprudence

Recommended Reading and References

Allen S, Schulman S (2009) Recidivism of Philadelphia community court misdemeanor cases closed in 2007. Personal communication with community court coordinator, Bill Babcock

Berman G (2010) Principles of community justice: a guide for community court planners. Center for Court Innovation, New York

Berman G, Feinblatt J (2005) Good courts. The New Press, New York

Cheesman F, Durkin M, Maggard S, Rottman D, Sohoni T, Rubio D (2009) Process evaluation of the Philadelphia community court. National Center for State Courts, Williamsburg

Cheesman F, Rottman D, Gibson S, Durkin M, Maggard S, Sohoni T, Rubio D, Curtis R (2010) Philadelphia community court evaluation final report: outcome/impact analysis and update on process evaluation. National Center for State Courts, Williamsburg

Eckberg D (2001) Hennepin county community justice project: summary report of short-term evaluation. Hennepin County District Court Research Department, Hennepin County

Hakuta J, Soroushian V, Kralstein D (2008) Testing the impact of the midtown community court: updating outcomes a decade later. Center for Court Innovation, New York

Henry K, Kralstein D (2011) Community courts: the research literature. Center for Court Innovation, New York

Herrschaft BA (2012) New York City community cleanup: the impact of a program for low-level offenders. Center for Court Innovation, New York

Jolliffe D, Farrington DP (2009) The effects on offending of the community justice initiatives in Liverpool and Salford. London. Available on request from the Ministry of Justice

Karafin DL (2008) Community courts across the globe: a survey of goals, performance measures and operations. Center for Court Innovation, New York. Report submitted to the Open Society Foundation for South Africa

Kelling GL, Wilson JQ (1982) Broken windows: the police and neighborhood safety. The Atlantic March:29–38

Lee C, Cheesman F, Rottman D, Swaner R, Lambson SH, Rempel M, Curtis R (2013) A comprehensive evaluation of the red hook community justice center. National Center for State Courts, Williamsburg

Lind E, Tyler T (1998) The social psychology of procedural justice. New York: Plenum

Llewellyn-Thomas S, Prior G (2007) North Liverpool community justice centre: survey of local residents. Ministry of Justice, London. Ministry of Justice Research Series 13/7

LaFree G (1998) Losing legitimacy: street crime and the decline of social institutions in America. Westview, Boulder

Malkin V (2003) Community courts and the process of accountability: consensus and conflict at the red hook community justice center. A Crim Law Rev 40:1573–1593

Mann CR (1993) Unequal justice: a question of color. Indiana University Press, Bloomington

Nagin DS, Pogarsky G (2001) Integrating celerity, impulsivity, and extralegal sanction threats into a model of general deterrence: theory and evidence. Criminology 39(4):865–892

Nugent-Borakove E (2009) Seattle municipal community court: outcome evaluation final report. Justice Management Institute, Denver

Rawls J (1971) A theory of justice. Harvard Press, Cambridge

Ross S, Halsey M, Bamford D, Cameron N, and King A (2009) Evaluation of the Neighbourhood Justice Centre, City of Yarra. Final report. University of Melbourne, Melbourne

Rottman D, Casey P (1999) Therapeutic jurisprudence and the emergence of problem-solving courts. Nat Inst Justice J 240:12–19. Reprinted in Palacios W et al. (eds) (2002) Crime and justice in America: present realities and future prospects. Prentice-Hall

Russell K (1998) The color of crime: racial hoaxes, white fear, black protectionism, police harassment, and other macro-aggressions. New York University Press, New York

Sviridoff M, Rottman D, Ostrom B, Curtis R (2000) Dispensing justice locally: the implementation and effects of the midtown community court. Center for Court Innovation, New York

Sviridoff M, Rottman D, Weidner R, Cheesman F, Curtis R, Hansen R, Ostrom B (2001) Dispensing justice locally: the impacts, cost and benefits of the midtown community court. Center for Court Innovation, New York

Swaner R (2010) Community perceptions of red hook, Brooklyn: views of quality of life, safety, and services. Center for Court Innovation, New York

Tyler T (1990) Why people obey the law. Yale University Press, New Haven

Tyler T (2001) Public trust and confidence in legal authorities: what do majority and minority group members want from the law and legal institutions? Behav Sci Law 19:215–235

Tyler T (2004) Procedural justice. In: Sarat A (ed) The Blackwell companion to law and society. Blackwell, Malden, pp 435–452

Tyler T, Huo YJ (2002) Trust in the law. Russell Sage, New York

Weidner R, Davis C (2000) Benefits and costs of the Hennepin county community court – a preliminary analysis. Institute on Criminal Justice, University of Minnesota Law School, Minneapolis

Westat (2012) East of the river community court evaluation, final report. Westat, Rockville

Wissler R (1995) Mediation and adjudication in small claims court. Law Soc Rev 29:323–358

Community Organization

▶ Neighborhood Effects and Social Networks

Community Policing

Jan Terpstra[1], Bas van Stokkom[2,3] and Kees van der Vijver[4]

[1]Institute of Criminology and Criminal Justice, Radboud University Nijmegen, Nijmegen, The Netherlands

[2]Radboud University Nijmegen, Nijmegen, The Netherlands

[3]VU University Amsterdam, Amsterdam, The Netherlands

[4]University of Twente, Enschede, The Netherlands

Synonyms

Community-oriented policing; Neighborhood policing; Personalized policing; Proximity policing

Overview

This entry deals with the characteristics of community policing, discusses its varieties and its problems and possibilities. Subsequently, the following aspects are examined: historical backgrounds, related police models (such as problem-oriented policing, disorder policing and reassurance policing), the Chicago Alternative Police Strategy, implementation, governance and accountability, and the effectiveness of community policing.

Introduction

Community (or community-oriented) policing has been a very popular term during the past few decades in many parts of the world. It has even been described as the new organizing paradigm in public policing, a revolution in police work, or even a new orthodoxy (Bayley and Shearing 1996). However, despite its popularity, community policing has a multitude of meanings (Rosenbaum 1998; Ponsaers 2001). This has given some authors occasion to describe the concept as vague or as nebulous, one that seems to justify more or less any police program or activity. Brogden and Nijhar (2005) describe community policing as a buzzword that comes in all shapes and sizes.

Three factors are relevant to understanding why it proves so difficult to have an unequivocal definition of community policing, one that is generally recognized and accepted. First, the vagueness of the concept is probably one of the factors that contributes to its popularity. As Rosenbaum (1998) said, community policing seems to be something that everyone can identify with. Who would be against "community"? Secondly, in practice, the activities known as community policing do not rest on one single model of policing, but on a multitude of them (Ponsaers 2001). Finally, it is often assumed that community policing does not represent a more or less coherent or identifiable set of practices and activities; rather, it is a more general organizational strategy for the police. In line with this, it is seen more as a sort of police philosophy, which cannot be reduced to certain specific activities. In Skogan's (2006) terminology, community policing is more a process than a product.

Nevertheless, there have been many attempts to define community policing. One of the best known is the one by Trojanowicz and Bucqueroux (1990): "Community policing is both a philosophy and organizational strategy to allow community residents and police to work together in new ways to solve problems of crime, physical and social disorder and neighborhood decay."

In addition to this definition, there have been many other endeavors to describe what community policing is and what should be seen as its main ingredients. Many of these broader definitions, however, lack a certain coherence, for example, because they not only concern police strategies, but also the organization of the police force. For instance, both elements can be found in the well-known definition by Skogan and Hartnett (1997), who make a distinction between five central elements of community policing: organizational decentralization, a two-way communication between the police and the public, a broadly focused, problem-oriented policing strategy, responsiveness to citizens' demands, and a commitment to helping neighborhoods solve crime problems on their own.

Another factor that may contribute to a lack of coherence is that definitions of community policing may also contain elements of other police models. An example of this is the definition of community policing given by Bayley (1994), who speaks of CAMPS: consultation, adaptation, mobilization, and problem solving. However, the latter element can better be seen as belonging to the problem-oriented police model (see also below).

If elements of other police models and organizational features are left aside, there remain five elements of community policing that are often referred to in the international literature. Despite all kinds of varieties in the organization and practice of community policing, there seems to be a general consensus about the relevance of these elements: bridging the distance between the police and citizens; a focus of the police on a broad range of problems in the neighborhood (not only crime, but also social disorder and feelings of insecurity); a preventive approach in addition to a reactive one; cooperation with other agencies in the prevention and control of crime, disorder, and other neighborhood problems; and the promotion of citizen involvement. The question of whether or not these five elements are achieved in practice is a matter for empirical research.

Backgrounds

Community policing started in the United States in the 1970s as a reaction to the failures of the dominant police styles of those days. Kelling and Moore (1988) divide the history of American policing into three different eras. First there is what they call the Political Era, which continued until the early 1900s. In those years, American police forces gained their authority from local political leaders and were adjuncts to local political machines. The police provided a wide range of services to citizens. The primary tactic of the police was the foot patrol. Although the police were organized at a local level, their organizations were highly centralized and quasi-military. This political strategy had significant weaknesses, such as the rise of police corruption, discrimination against strangers, and relative inefficiency.

The second era is called the Reform Era, which started in the early decades of the twentieth century with the work of A. Volmer and later on O.W. Wilson, who wanted to reform the police organization and strategy. These reformers rejected politics as the basis of police legitimacy. Instead, the law and police professionalism were presented as the new basis for the legitimacy of the police. As a result, police departments became relatively autonomous organizations. The reform policy narrowed the main function of the police to law enforcement and control of crime. The police officers' work became standardized and uniform. The relation between the police and citizens came primarily to be defined in neutral and remote professional terms. Citizens were seen as passive recipients of professional crime control services. Preventive patrol by car and rapid response became the main strategies of this reformed police force.

In the 1960s and 1970s, it became clear that this professional police model (as it was called, although it was not at all very professional, as Stone and Travis (2011) have argued), had serious shortcomings. Several factors contributed to this: the rising level of crime and fear of crime, complaints – especially from members of ethnic minorities – about the way they were treated by the police, and the new critical awareness as a result of civil rights and antiwar movements. Moreover, scientific research revealed that many of the central strategies of this professional police force (such as preventive patrol by car or standardized criminal investigation) proved to be highly ineffective.

The third era, starting in the 1970s, may be seen as a reaction to the failures of the standard police model. In this Community Policing Era (Kelling and Moore call it a Problem-Solving Era), there is a renewed emphasis on close relations with the community and the citizens. The definition of the police function broadens into a community strategy. According to Kelling and Moore (1988), organizational decentralization is inherent to community policing. The involvement of police officers in the diagnosis of and response to community problems makes it necessary to locate the power of operational and tactical decision making at lower levels within the organization. Community policing requires intimate relations with citizens and instead of a claimed monopolistic responsibility for crime control, there should be strategic alliances with neighborhood and community groups. Community policing is based on a wide variety of strategies, such as foot patrol (rediscovered as an important instrument for regaining citizens' trust and promoting feelings of security), information gathering, problem solving, victim services, and others.

Although this analysis by Kelling and Moore deals with developments in the United States, there are certainly similarities with the background to the rise of community policing elsewhere. The failures of the standard police model, a loss of police legitimacy and citizens' trust, and higher levels of crime and feelings of insecurity contributed to the adoption of this new police model, in addition to the wish of police organizations elsewhere to imitate and follow the American police models, which enjoyed a high reputation.

The historical backgrounds to community policing are even more complex than this standard view of the history of American policing reveals. The shift from a standard police force to community policing is also related to changes in the prevailing views on democracy and the relation between the state and society. Moreover, in some cases, especially in the United States, the model of community policing may be related to more or less romantic notions of community (Herbert 2006). This seems to be of less relevance in some European countries, for example, where community or neighborhood is more often viewed as relatively neutral scales for the organization of the police and police-citizen relations.

Related Police Models

For the past few decades, community policing has become a dominant policing perspective in many countries. Other policing models were introduced, which are often seen as related or even integral to this community policing perspective. These models may also be viewed as responses to the failure of the (so-called professional) standard police model, but they diverge in their strategies and priorities. Nevertheless, such models are conceptually close to community policing and often they are assumed to be innovations in community policing. Three such models are briefly discussed here: problem-oriented policing (Goldstein 1990), disorder policing (or broken windows policing) (Wilson and Kelling 1982), and reassurance policing (Innes 2004).

Problem-Oriented Policing

The concept of problem-oriented policing was introduced by Goldstein (1990). In his view, much police work suffers from what he calls a means-over-ends-syndrome: police work is often defined more in terms of the "means" of police work (such as rule enforcement, surveillance, or rapid response) than by its goals. According to Goldstein, the work the police do requires them to deal with a wide range of behavioral and social problems that arise in the community.

Often police work mainly consists of ad hoc responses to a seemingly endless stream of isolated and recurring incidents. A problem-oriented approach is meant to make an end to this unproductive process. The police should conduct a detailed, systematic analysis of the causes and factors underlying a series of comparable events and incidents. Interventions by the police should be based on such detailed analyses and will differ for each problem (Goldstein 1990). This should contribute to a more preventive form of policing. Problem-solving strategies should not only be based on the traditional instruments and resources available to the police, but also on those of other agencies. In this respect, problem-oriented policing is a form of partnership or multiagency policing.

Several reviews of studies that evaluate police work have concluded that community policing may only be effective if it is applied in combination with a problem-oriented approach (Eck 2004; Skogan and Frydl 2004). In spite of this, there are also important differences between problem-oriented and community policing. First, the relevance of a problem-oriented approach is not confined to community policing (Tilley 2003). Moreover, one of the main goals of community policing is to contribute to an improvement in or restoration of the relations between the police and citizens. Insofar as citizens are involved in a problem-oriented approach, there is only a functional relation between the police and the citizens, who primarily are seen as the "eyes and ears" of the police (Tilley 2003).

Disorder Policing

According to the model of disorder policing, the police should focus on the restoration of physical and social order in neighborhoods. In this way, the police might reduce feelings of insecurity among citizens, enhance their trust in the police, and prevent and counter crime. This theory achieved recognition, thanks to a classical article by Wilson and Kelling published in 1982, entitled "Broken Windows."

The broken windows theory assumes that forms of physical and social disorder (such as incivility, physical deterioration, or disorderly social conduct), especially if they accumulate and persist for a relatively long time, will result in feelings of insecurity and the breakdown of informal social control in the community. Wilson and Kelling (1982) explain that a broken window acts as a signal to potential offenders that people living in this area are indifferent to what is happening in their neighborhood. As a consequence, one broken window (which is not speedily repaired) will result in more broken windows and more serious forms of crime in the neighborhood. Instead of reducing the police to playing the role of the criminal justice system's "front end," keeping the peace and maintaining social order should be restored as the force's main tasks (Kelling and Coles 1996). The main goal for the police should be to promote informal social control.

Many scholars view broken windows policing and community policing as opposites. Broken windows policing is identified as an aggressive approach to minor offenses and disturbances. In that case, community-oriented policing is associated with a more "gentle, friendly, concerned, or low-key method."

According to a second view, disorder policing is a specific instantiation of community policing. According to this line of thought, the broken windows theory is considered a source of inspiration for the well-known form of community policing that was developed in Chicago in the 1990s.

Reassurance Policing

In the last decade, a new model of policing has come to the fore, called reassurance policing. A central assumption in this model is that the presence of visible, uniformed authorities will reassure citizens, reduce feelings of insecurity, and restore trust in the police. This model was presented in the UK in reaction to the fact that since the mid-1990s, the objective level of crime had improved, while many citizens still perceived public safety as a serious, even worsening problem (Tuffin et al. 2006).

According to this model, the police should focus on problems that have the greatest impact on citizens' feelings of security ("signal crimes"). According to Innes (2004), visible and troublesome incidents are occupying the "collective memory" of residents; they function as warning signs for future threats. The police should develop "control signals," actions that communicate a sense of regained peace and order. These signals convince citizens that police officers are concerned about local security.

Although this model places great emphasis on the importance of visibility, accessibility, and proximity of the police, these conditions, although perhaps necessary, are not seen as sufficient to promote feelings of security and trust in the police. Evaluations of reassurance policing in the UK show that, in addition to police visibility, other important factors are problem-solving strategies, citizen involvement, and the promotion of citizen initiatives (Tuffin et al. 2006). Extending the notion of reassurance with problem-solving strategies, cooperation with other agencies, and the focus on "signal crimes" makes this theory quite complex.

The Chicago Experiment: CAPS

Disorder policing takes the "community needs" of residents as its starting point. In Wilson and

Kelling's terms, the neighborhood indicates what the "appropriate level of public order" should be. The police should make use of the "preventive capital" of citizens. In *Disorder and Decline* (1990), Wesley Skogan reflected on the problematic nature of determining order. In heterogeneous neighborhoods, neither the police nor citizen-organizations can claim that their vision of order is authoritative and liable to impose compliance. In these conditions, Skogan says, it is a challenge to involve citizens and other relevant parties explicitly in determining and selecting disorder problems that should be tackled. Order is negotiated rather than imposed. In this way, policing becomes the object of an open, deliberative process.

Actually this view formed the starting point of an ambitious, intensive police innovation program, the Chicago Alternative Police Strategy (CAPS). This program "reinvented" some of the core features of earlier community policing approaches (prevention, cooperation, problem solving, etc.), but added some "new" aspects: encouraging residents and other stakeholders to find solutions for local security problems and giving priority to problems of disorder.

Since 1993, CAPS has been implemented in all of Chicago's 279 police beats. All beats have monthly meetings attended by residents, police officers, and other professionals, the aim being to identify and prioritize local problems together, and to develop plans for dealing with them.

The police explicitly accepted the broken windows philosophy, and consequently also the expansion of the police mandate. Nevertheless, CAPS was launched by the city administration. The most difficult task was to recruit the police force's support for the program (Skogan 2006). The innovative element of CAPS lies not so much in its methods of citizen participation, but rather in the rigorous, long-term organization and political will to establish and maintain partnerships in every neighborhood.

Evaluation studies (Skogan and Hartnett 1997; Skogan 2006) show that the monthly beat meetings are attended by an average of 20 citizens. CAPS attracts residents who had not participated earlier in other programs. Increased participation of women, blacks, and poor citizens was noted. A remarkable finding is that participation in black neighborhoods is just as high as in white ones, even in high crime districts where the police were unpopular.

Between 1994 and 2003, confidence in the police increased considerably. Citizens' views of police effectiveness, responsiveness and demeanor within the "three Chicago's" (white, black, and Latino) improved substantially (Skogan 2006). After studying Chicago beat meetings, Fung (2004) pointed out several problems. Some beats failed to address priority problems. Police resources were often allocated inequitably. The "natural" course of participation led to "conflictual paralysis" in one beat and domination by wealthy and well-educated residents in another. Often, energetic facilitators succeeded in breaking through this "laissez faire, first come, first served" style, managed to put problems of under-represented subgroups on the agenda, and called in intervention teams to take action against drug houses (for example). Thus, beat meetings need powerful facilitators. Fung argues that even in neighborhoods that lack resources or have serious internal conflicts, citizens gain more from the new deliberative arrangements than from the bureaucratic organizational forms that existed prior to the introduction of CAPS.

In a case study of a neighborhood watch group in Beltway, a suburban district in Chicago, Carr (2005) shows that citizens received support from the police and other local public services only after CAPS had been introduced. As a result, small groups of active citizens were able to mobilize resources and initiate partnerships. Carr challenges the assumption that neighborhoods need strong social cohesion in order to get crime and disorder under control. Without CAPS, Carr stresses, Beltway activism would have been impossible. The main factor that drives citizen participation is trust in public professionals. Even in poor ethnic neighborhoods, consultation and partnerships may be successful. What is important is to overcome "legal cynicism" and the tendency of residents to tolerate antisocial behavior (Carr 2005).

Implementation

A recurrent complaint about community policing concerns its implementation. In Mastrofski's (2006) view, police work in the United States still does not meet the high ambitions of community policing, despite more than two decades of reform and the formal commitment of many police organizations to this model. In practice, there are many obstacles to full and effective implementation. According to Mastrofski, community policing in the United States did not transform the police in the way that was originally envisaged. Klockars (1988) was still more pessimistic, even in the early days of community policing. In his view, the rhetoric of community policing wraps the police in romantic, powerful, and unquestionably good aspirations, which the police, however, cannot realize in practice but only in what he calls "*ersatz terms*."

Although the past few decades have seen many studies on community policing, there is a shortage of detailed research on the practical implementation of this police model. Often its implementation is treated as a black box. Many studies focus on formal or organizational aspects of this form of police work, while seemingly disregarding the reality of community policing at street level.

One of the main findings in the limited number of studies that have focused on the street level of community police work is that community police officers often continue to spend a great deal of time doing traditional police tasks. This may be the result of several factors, such as the prevailing police culture (Greene 2000) or limited expertise (Eck 2004). Police leaders may be rather cynical about community policing or may fear that the increased discretionary room allowed to community officers may have adverse implications for their own position. However, the most serious problems in the implementation of community policing may arise in improving relations between the police and citizens, especially in those urban areas that most need community involvement.

Studies in other countries reported comparable results. For example, in Denmark, Holmberg (2002) found that in practice, many of the community policing principles are often embraced only very partially or in unintended ways. The main impediments in this country included large working areas, the burden of a lot of paperwork, and the priority of other tasks.

In the Netherlands, the five main ambitions of community policing (proximity, focus on a wide range of problems in the neighborhood, preventive strategies, cooperation with other agencies in the control of crime and disorder, and promotion of citizen involvement) were only carried through to a limited extent (Terpstra 2010). In contrast to the United States, these problems were not so much caused by a (supposed) police culture nor by romantic notions of community (Herbert 2006). More important was the fact that the concept of community policing still has a vague professional image, even among Dutch community police officers themselves. Moreover, community police work is often rather open and indefinite in nature. Although community policing usually has high ambitions, the officers involved often have only limited resources, skills, and expertise to achieve them. In addition, the support they receive from their management, superiors, and colleagues is often quite meager, and sometimes contradictory.

Governance and Accountability

A common element of community policing initiatives is attempting to bridge the distance between the police and citizens. It is often assumed that community policing may encourage local governance and accountability of the police (Skogan and Hartnett 1997; Somerville 2009). A main argument underlying CAPS was that it would enable citizens to provide local knowledge and resources to the police and that they would be better able to monitor police officers and hold them accountable. Following this line of argument, community policing may even be presented as a form of democratization.

However, community policing is not always seen as beneficial to police accountability. Organizational decentralization, a growing intimacy between police officers and citizens, and the increased use of discretion by police officers may also be viewed as potential threats to police accountability (Brogden and Nijhar 2005; Herbert 2009).

In fact, community police officers are confronted with very divergent forms of accountability, reflecting hierarchical, egalitarian, and more individualist cultures. This complexity results from the fact that a multiplicity of actors, each with their own views and interests, have a stake in the implementation of community policing. The (local) government, (internal) superiors, colleagues, external partners and (representatives of) the public try to "govern" the community officers and hold them accountable. As a result, these officers are not only held accountable within the traditional hierarchical relations, but also horizontally and vertically downward by partners or citizens (Somerville 2009).

In countries where the police organization is not strictly local (which is generally the case outside the United States), it is useful to distinguish between nonlocal and local forms of governance and accountability in community policing. Nonlocally initiated forms are mainly bureaucratic and managerial, representing hierarchical and individualistic cultures, respectively. In addition, community police officers may not only be accountable to the local government and politicians, but also to other police professionals, partners, and (groups of) citizens. These latter local forms of governance and accountability are usually more egalitarian (although there may also be some hierarchical elements). The traditional culture of command-and-control is hard to align with the complexity of community policing, the need for a flexible, tailor-made approach, and the need for officer discretion.

In the 1990s, new managerial forms of governance and accountability were introduced in many police organizations, such as systems of performance management with quantitative targets. These forms do not fit in with the localized nature of community policing work. They often create unintended consequences that are difficult to combine with community policing, such as a standardization of the work or a fixation on those activities that can more easily be quantified and measured.

Professional types of governance and accountability are often presented as a solution to the problems associated with hierarchical and managerial forms. Here the professional discretionary power of the community police officer is highly valued. Community police officers are not only asked to follow rules and procedures, but also to use their autonomy in a professional way and to show their moral involvement with local problem solving. However, these professional forms of governance and accountability are scarcely ever realized in practice, despite being what many community police officers would prefer in their relation with their superiors.

For the past few decades, many countries have witnessed a growing interest in citizen participation in relation to policing and public safety. However, citizen participation scarcely contributes to local police governance/accountability. Control over the police by partners or citizens may be perceived as conflicting with the views, routines, and priorities of the police (Terpstra 2011). In the United States, Herbert (2006) interpreted this as the result both of a police culture that emphasizes police self-protection combined with lofty and unrealistic expectations about "community" and local democracy. Comparable problems were reported in the Netherlands, where more modest and pragmatic meanings of community and the relation between community policing and local democracy prevail. Here the minimization of citizen influence on the police was used by community police officers to reconcile conflicting demands and expectations about their work.

Effectiveness

The issue of effectiveness of community policing raises several important questions. First, what should be meant by the effects of community policing? Generally it is assumed that an evaluation of the impact of community policing should not only focus on the question of whether it results in a reduction of crime and social disorder, but also whether it contributes to feelings of security and to greater levels of citizens' trust in the police. Secondly, how are the (potential) effects of community policing to be measured? Studies evaluating the impact of complex programs like community policing require a longitudinal and quasi-experimental research design. An adequate interpretation of the results of such an impact evaluation demands in addition that a detailed process or implementation evaluation should be used. If there is a shortage of reliable information about the implementation of the intervention or program, the effects or impact will be hard to interpret.

The internationally best known evaluation of community policing is that of the Chicago Alternative Policing Strategy (CAPS) done by Skogan and his colleagues. This study shows that opinions about the police in Chicago steadily improved in the years that CAPS was implemented. An increasing number of citizens feel that the police in their area are dealing with problems that concern residents and that the police are prepared to cooperate with residents in solving these problems. The increasing visibility of the police (partly a result of more frequent foot patrols) improved citizens' evaluation of the police. Moreover, many residents feel less insecure, partially because they see the police more often. Fear of crime declined substantially among the groups experiencing the greatest levels of fear. In Chicago, many forms of crime (both violent crime and property crime) declined sharply. However, in general, the fall was comparable to trends in other American cities. The evaluation of CAPS also showed that residents in certain areas reported sharp improvements in neighborhood conditions. African-American residents in particular reported improvements in relation to, for instance, abandoned buildings or graffiti, and different forms of social disorder. In summary, community policing according to the CAPS model had a positive impact on citizens' opinions about the police, feelings of security, and neighborhood problems and disorder (Skogan and Hartnett 1997; Skogan 2006).

Since the late 1990s, especially in the United States, there have been several meta-evaluations of the effectiveness of community policing, often comparing it with other police models and strategies. Sherman et al. (1997) showed that community policing which does not explicitly focus on specific risk factors, generally, has no impact or only a minor one on the level of crime. According to this review, the effectiveness of community policing may be improved if it is supplemented with a problem-solving approach or if residents are involved in the assessment of police priorities in their area.

Skogan and Frydl (2004) concluded that in general, the effectiveness of community policing is hard to determine because this concept is used for a wide variety of police practices and strategies. They found no evidence that methods that are often associated with community policing, like neighborhood watch, general foot patrol, storefront offices, or community meetings, contribute to a reduction of crime. However, community policing may have a positive impact on neighborhood problems. Improved relations between the police and residents may also diminish fear of crime and may cause residents to perceive less social and physical disorder in their neighborhood. According to Skogan and Frydl (2004) community policing may have a greater impact on the reduction of crime if it is supplemented with elements of two other police models: hot spot policing and problem-oriented policing.

Although available research gives important indications of the potential impact of community policing, there are still major questions left. Community policing is an internationally heterogeneous concept and practice. It is often unclear if conclusions from mainly American studies may be generalized to other jurisdictions. There are also indications that the effectiveness of community policing may differ between different types of neighborhoods and for different categories of citizens.

Future Direction

The conclusions of studies on the implementation and effects of community policing have raised serious questions about the value and practical feasibility of the community policing model. They have tempered naïve and optimistic ambitions that were sometimes connected to the model (Herbert 2006). In some countries, such as Norway, community policing never gained a dominant position. The reform of policing in that country for the past two decades was more strongly motivated by a problem-oriented approach (Larsson 2010).

In addition, in some European countries like the Netherlands, both an increasing emphasis on managerial discourse and a shift to a more repressive and punitive political climate meant that community policing lost much of its dominance (although it remains an important part of police work) (Terpstra 2011). Moreover, since the introduction of this model, police organizations were confronted with other fundamental changes that did not fit neatly within the framework of community policing (such as cybercrime). As a result, community policing lost some of its attraction as an overarching framework for the police (Stone and Travis 2011).

Despite these developments, one might expect that community policing is here to stay, although perhaps in a more pragmatic, modest way. The fundamental notion was and remains that for the promotion of a legitimate, fair, and effective police, it is essential to close the distance between citizens and the police and to create different forms of cooperation between them at the neighborhood level. It may be expected that as a result, community policing will continue to be an important element in the future, although adapted to very different and changing circumstances.

Related Entries

- ▶ Broken Windows Thesis
- ▶ Communities and the Police
- ▶ Democratic Policing
- ▶ Impacts of Community-Oriented Policing
- ▶ Law of Community Policing and Public Order Policing
- ▶ National and Local Policing
- ▶ Order Maintenance Policing
- ▶ Problem-Oriented Policing
- ▶ Reassurance Policing and Signal Crimes
- ▶ Theories on Policing and Communities

References

Bayley DH (1994) Police for the future. Oxford University Press, New York/Oxford

Bayley DH, Shearing CD (1996) The future of policing. Law Soc Rev 30(3):585–606

Brogden M, Nijhar P (2005) Community policing. National and international models and approaches. Willan, Cullompton

Carr PJ (2005) Clean streets: crime, disorder and social control in a Chicago neighborhood. NYU Press, New York

Eck JE (2004) Why don't problems get solved? In: Skogan WG (ed) Community policing. Can it work? Wadsworth, Belmont, pp 185–206

Fung A (2004) Empowered participation: reinventing urban democracy. Princeton University Press, Princeton

Goldstein H (1990) Problem-oriented policing. McGraw-Hill, New York

Greene JR (2000) Community policing in America: changing the nature, structure, and the function of the police. In Horney J (ed) Policies, processes, and decisions of the criminal justice system. Criminal justice, vol 3. U.S. Department of Justice, Washington, DC, pp 299–370

Herbert S (2006) Citizens, cops, and power. The University of Chicago Press, Chicago

Herbert S (2009) Community policing and accountability. In: Grabosky P (ed) Community policing and peacekeeping. CRC Press, Boca Raton

Holmberg L (2002) Personalized policing. Results from a series of experiments with proximity policing in Denmark. Policing: Int J Police Strat Manag 25(1):32–47

Innes M (2004) Signal crimes and signal disorders: notes on deviance as communicative action. Br J Sociol 55(3):335–355

Kelling GL, Coles CM (1996) Fixing broken windows, restoring order and reducing crime in our communities. The Free Press, New York

Kelling GL, Moore MH (1988) The evolving strategy of policing. National Institute of Justice, Washington, DC

Klockars CB (1988) The rhetoric of community policing. In: Greene JR, Mastrofski SD (eds) Community policing: rhetoric or reality. Praeger, New York, pp 239–258

Larsson P (2010) Ideology as a cover up: community policing in Norway. J Pol Stud 3(16):233–241

Mastrofski SD (2006) Community policing; a sceptical view. In Weisburd D, Braga AA (eds) Police innovation. Contrasting perspectives. Cambridge University Press, Cambridge, pp 44–73

Ponsaers P (2001) Reading about "community (oriented) policing" and police models. Policing: Int J Police Strat Manag 24(4):470–496

Rosenbaum DP (1998) The changing role of the police: assessing the current transition to community policing. In Brodeur JP (ed) How to recognize good policing. Problems and issues. Sage, Thousand Oaks, pp 3–29

Sherman LW et al (1997) Preventing crime: what works, what doesn't, what's promising: a report to the United States Congress. National Institute of Justice, Washington, DC

Skogan WG (1990) Disorder and decline: crime and the spiral of decay in American cities. The Free Press, New York

Skogan WG (2006) Police and community in Chicago. A tale of three cities. Oxford University Press, Oxford/New York

Skogan WG, Frydl K (2004) Fairness and effectiveness in policing. The evidence. The National Academies Press, Washington, DC

Skogan WG, Hartnett SM (1997) Community policing, Chicago style. Oxford University Press, New York/Oxford

Somerville P (2009) Understanding community policing. Policing: Int J Police Strat Manag 32(2):261–277

Stone C, Travis J (2011) Toward a new professionalism in policing. National Institute of Justice/Harvard Kennedy School, Boston/Washington, DC

Terpstra J (2010) Community policing in practice: ambitions and realization. Pol: J Pol Pract 4(1):64–72

Terpstra J (2011) Governance and accountability in community policing. Crime, Law Social Change 55(1):87–104

Tilley N (2003) Community policing, problem-oriented policing and intelligence-led policing. In: Newburn T (ed) Handbook of policing. Willan, Cullompton, pp 311–339

Trojanowicz R, Bucqueroux B (1990) Community policing: a contemporary perspective. Anderson, Cincinnati

Tuffin R, Morris J, Poole A (2006) An evaluation of the impact of the national Reassurance Policing Programme (No. 296). Home Office Research, Development and Statistics Directorate, London

Wilson JQ, Kelling GL (1982) Broken windows: the police and neighborhood safety. Atlantic Monthly 243:29–38

Community Service in Europe

Frieder Dünkel[1] and Tapio Lappi-Seppälä[2]
[1]Faculty of Law, University of Greifswald, Greifswald, Germany
[2]National Research Institute of Legal Policy, Helsinki, Finland

Overview

Community service orders have been introduced in England and Wales 1972 and since then have become one of the most important community sanctions in many European countries and outside Europe. Community service has gained importance, in particular, for offenders who are not able to pay a fine, but also to replace short-term imprisonment. It has an essential role as the principal community sanction in Spain, Hungary, England and Wales, Scotland, Latvia, and all Nordic countries. The scope of application varies considerably when looking at the range of hours that can be imposed. In Europe, one regularly finds a scope between 40 and 240 h, in the USA, between 100 and 500 h. Empirical studies have shown that the implementation of community service has been successful. In this regard, it is important that most offenders assigned to community service programs are completing their working hours. The available evidence suggests that community service is (at least) a promising alternative in terms of reducing recidivism. Several results point to significant differences between offenders who are sentenced to imprisonment and offenders who participate in community service. Recent meta-analyses demonstrate a stronger effect size (.07) than earlier ones, which may be explained by the better structured programs considering "what-works" criteria. The authors advocate to extend community service and to ensure that it comes to replace imprisonment (instead of replacing other noncustodial sanctions).

Historical Development

Historically forced labor has a long tradition in human history. In penal history, it was one of the core elements of prison life. Forced prison labor is still widely accepted and exempted from the ban on forced labor as expressed in international human rights standards. First efforts to ban forced labor date back to 1930 when the International Labour Organisation (ILO) initiated the Convention No. 29 and in 1957 a second one (Convention No. 105) after ILO had been linked to the UN in 1946. Both Conventions aimed at fighting against slavery and any other form of involuntary (forced) labor (see de Jonge 1999, pp. 321 ff.). The ban on forced labor can also be found in art. 8 of the International Covenant on Civil and Political Rights (ICCPR) and in art. 4 of the European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR). Art. 8(3) (c) of the ICCPR, however, allows for forced or compulsory labor "normally required of a person who is under detention in consequence of a lawful order of a court, or of a person during conditional release from such detention." The ECHR in the same wordings allows for forced labor only during detention and on conditional release. Strange enough to permit forced labor during the post-release stage, both Conventions do not clearly differentiate between remand and sentenced prisoners. Theoretically, one could think of forced labor in remand custody, but as a consequence of the presumption of innocence (see art. 6 ECHR), it is a common opinion and jurisprudence to exclude remand prisoners from compulsory work. The ILO Convention No. 29 had allowed forced labor only for sentenced prisoners, not for remand prisoners or those conditionally released (see de Jonge 1999, p. 327). The German Constitution as many other national laws allows forced labor only for sentenced prisoners (see art. 12(3) Federal Constitution, *Grundgesetz*).

Therefore, from an international human rights view, it might be difficult to justify community service as an independent sanction as far as it is not strictly bound to replace a prison sentence. In that case, it can be justified with the argument *a maiore ad minus* that forced labor in the community is a less intrusive sanction than imprisonment plus forced labor. If, however, community work does not replace a prison sentence, but other alternatives such as fines or conditional discharge without intervention, it can be questioned as being a violation of the ban on forced labor.

The German Constitutional Court has found a way out of this dilemma by going back to the roots of the ban of forced labor in the German Constitution after World War II: The constitutional ban was implemented as a fundamental human right with regard to forced labor in the preceding Nazi-era, where thousands of people were killed by hard labor in concentration camps and elsewhere. The legislator wanted to outlaw such inhuman hard labor, which in no way is comparable with some few hours of labor, in particular if its purpose is the rehabilitation (resocialization, education) of the offender (see *Bundesverfassungsgerichtsentscheidungen*, BVerfGE 74, pp. 102 ff., 122 ff.).

Germany did not introduce community service as an independent sanction in the Criminal Law for adults of at least 21 years of age, but only as a substitute sanction to prevent imprisonment for fine-defaulters (1983). This is different in the field of juvenile justice where the German Juvenile Justice Act (JGG) provides for community service as an independent sanction as well as a condition of diversion or of a suspended sentence (see Dünkel 2011a). The practice extensively uses community service: 44 % of all sentenced juvenile and young adult offenders (aged 14–21) in 2010 received a community service order (see Dünkel 2011a; Heinz 2012).

First community service programs were introduced in the USA with female traffic offenders in Alameda County, California, in 1966, with local initiatives following in several counties throughout the USA (Wright 1991, p. 40). The aim of different initiatives was to create community service programs as a viable alternative to incarceration (van Ness 1986, p. 194).

Community service as an alternative to prison in Europe first has been introduced in 1972 in England and Wales. Other European countries followed in the 1980s: Italy 1981, Denmark and

Portugal 1982, France 1983, the Netherlands 1985 (see Albrecht and Schädler 1986).

With the exception of Denmark, the Scandinavian countries introduced community service only in the 1990s. Denmark started an experiment with community service – as mentioned – already in 1982. However, the sanction was not introduced in a wider scale in the Nordic sanctions systems until the 1990s. Community service appears in different forms. Finland and Norway treat community service as an independent sanction (although Norway renamed "community service" in 2001 as "community punishment"). In Denmark and Sweden, community service is attached either to conditional imprisonment or a probation order. In Finland, long (over 1 year) conditional prison sentences may be combined with a short (20–60 h) community service order. In Denmark, community service can be combined also with fines and unconditional imprisonment. In addition, community service may be attached with separate conditions concerning residence, school attendance, or work. Also Norway allows specific conditions regarding the offender's dwelling, work, and treatment. The maximum number of community service hours varies from 200 (Finland) to 420 (Norway). The range of community service in Denmark was from 30 to 240 h, but the maximum was raised to 300 h in 2012.

Countries differ also in the sense, how strictly community service has been defined as an alternative to imprisonment, and not to other sanctions. Finland has followed the strictest policy in this respect, by adopting a specific *two-step procedure*. First, the court is supposed to make its sentencing decision by applying the normal principles and criteria of sentencing without considering the possibility of community service. Second, *if* the result of this deliberation is unconditional imprisonment (and certain requirements are fulfilled), the court may transform the sentence into community service. In principle, community service may therefore be used only in cases in which the accused would otherwise have received a sentence to unconditional imprisonment. Other countries are less strict in this point. In connection of the 2001 reform in Norway, the scope of community punishment was extended to be used, not only instead of imprisonment, but as a replacement penalty for juvenile offenders who had previously been sentenced to supervised conditional imprisonment. In Iceland, the decision over community service is taken by the prison administration, not the courts. This avoids effectively the risks of net-widening (but could raise questions from constitutional point of view in some other countries).

Community Service in Europe, Table 1 Community service in Scandinavian countries 2008–2009

The Use of Community Service Sentences in Scandinavia in 2008–2009

Figures per 100,000 pop and % of court dispositions

	Fin %	Swe %	Den %	Nor %
Unconditional prison	130 (11 %)	158 (19 %)	160 (24 %)	219 (45 %)
Community service	64 (5 %)	64 (8 %)	59 (7 %)	54 (13 %)

In the turn of the 2000s, several Scandinavian countries completed law reforms in order to increase the use of community service. Sweden created a combination of community service and suspended sentence, thus increasing the number of annual cases from 2,000 to around 4,000. Denmark changed its policy in 2000 by allowing community service to be used also for drunken driving (which was previously forbidden). Within a period of two years, this increased the number of sentences from 1,000 to 4,000. Norway, in turn, tried to increase the credibility of community service by changing the title to community punishment by including also other elements in the sentence and by expanding the scope of application also to drunken driving. This resulted in an increase from around 500 cases to the present little over 2,500 cases (Table 1).

Relative to population community service is used roughly on a same scale in all countries, with Finland and Sweden on the lead (64) and Denmark and Norway following (59 and 54). A longer trend from 1997 onwards shows that three out of four countries expanded the use of community service in the shift of the millennium, while Finland had used this alternative more extensively already from the mid-1990s (Fig. 1).

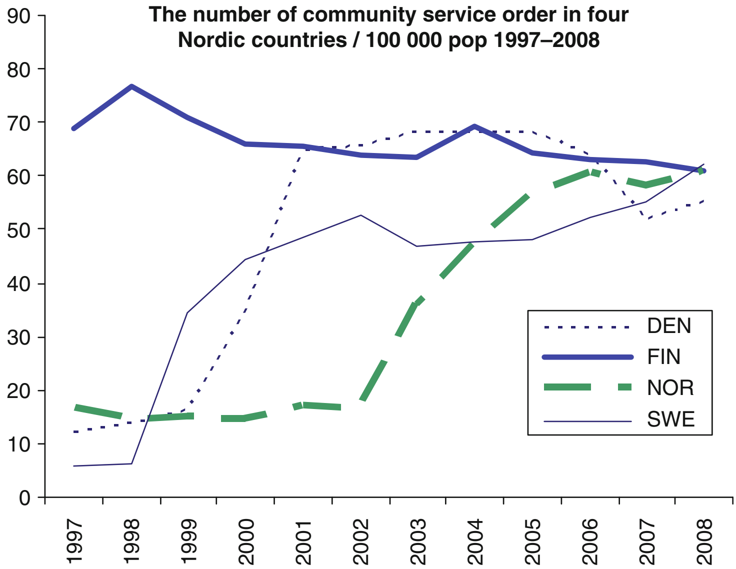

Community Service in Europe, Fig. 1 The number of court-imposed community service orders/ per 100,000 population

Non-European countries have introduced community service as well, in particular in the Anglo-Saxon world (Australia, New Zealand, South Africa, and USA), but also in South America, recently, community service is developed as an alternative sanction. In these countries (as in Europe already earlier), juvenile criminal law is a forerunner for reforms of the sanctions system (all juvenile justice jurisdictions dispose of community service orders, see Tiffer-Sotomayor 2000).

Legal Conditions and Basic Philosophies of Community Service

Community service involves the performance of unpaid work, during leisure-time and within a given period, for the good of the community. The offender "shall pay back to the community via unpaid work" (Goldson 2008, p. 78; Dünkel et al. 2011, p. 1673). Thus, the original philosophy of community service fits very well with the special preventive aim of punishment (resocialization or with regard to juvenile justice: education) as well as to the idea of restorative justice. In the context of restorative justice, another orientation is outlined by Wright (1991, p. 44): "The emphasis of community service is not on punishment nor on rehabilitation; rather, it is on accountability." It focuses "not on offenders' needs but their strengths; not on their lack of insight but their capacity for responsibility; not on their vulnerability to social and psychological factors but their capacity to choose. These differentiate a rehabilitative response from a restorative/community service response to crime. And punitive elements of community service orders may attend its imposition, within a restorative system, only as by-products of the offender's commitment of time and effort" (Wright 1991, p. 44).

However, in the last 20 years, some countries have redefined community service and also renamed it as "punishment in the community" or "community punishment" and thus emphasizing the repressive goal of retribution. This was particularly the case in England and Wales (1998), but also in Norway (2001). Therefore, the nature and content of community service may be rather different from jurisdiction to jurisdiction.

The status and contents of community service may vary. It may:

- be imposed as an independent sanction or as an adjunct to another sanction
- replace only prison sentences or also other penalties

Although community service is widely accepted in the field of juvenile justice, some countries explicitly exclude the very young age groups, e.g., under 16 (England and Wales) or 15 (South Africa).

The *range of hours* (and implicitly the range of prison sentences that can be replaced) varies considerably:

In Europe, the range mostly is from 40 to 240 h, often with a maximum of 120 h for juvenile offenders. Even in countries with normally great similarities in their penal policy such as the Scandinavian countries, large variations can be seen: In Finland, the maximum is 200; in Denmark, 300; in Norway, 420 h.

In South Africa, the range is from 50 to 240 h (to be completed in 1 year).

In New Zealand, the range with 40–400 h is considerably wider. Up to 200 h should be computed in one year, more than 200 h in two years. If offenders do not meet the requirements of the sentence, up to three months of imprisonment or a fine of up to 1,000 $ (NZ) can be imposed (see http://www.corrections.govt.nz/community-assistance/corrections-in-the-community/community-work.html).

In the USA, the range is regularly between 100 and 500 h (to be computed in one year).

International Human Rights Standards Concerning the Imposition and Execution of Community Service

Human rights aspects are of major importance in the decision-making process as the principle of proportionality (*Verhältnismäßigkeitsgrundsatz*), derived from the principle of *Rechtsstaat* (rule of law, art. 20(3) of the Constitution), is to be considered for the imposition of community service, i.e., the number of working hours imposed must be related to the seriousness of the offense. In most countries, therefore, the maximum number of hours is fixed by law. In Germany, where community service as an independent sanction is available only in juvenile justice, no maximum is fixed by law. Therefore, the principle of proportionality is of special importance to avoid unjustified high number of several 100 h.

The Recommendation of the Council of Europe concerning Community Sanctions or Measures (CSM) of 1992 (Rec. (92)16) stipulates that "the nature, content and methods of implementation . . . shall not jeopardise the privacy or the dignity of the offenders or their families, nor lead to their harassment. . . . Safeguards shall be adopted to protect the offender from insult and improper curiosity or publicity" (Rule 23).

The European Rules for Juvenile Offenders Subject to Sanctions or Measures (ERJOSSM) of 2008 (Rec. (2008)11) – as earlier Recommendations of the Council of Europe and the UN – assess that "sanctions or measures shall not humiliate or degrade the juveniles subject to them" (Rule 7). And: "Sanctions or measures shall not be implemented in a manner that aggravates their afflictive character or poses an undue risk of physical or mental harm" (Rule 8). In the commentary to these rules, reference is made to some forms of community work that can stigmatize offenders such as wearing special uniforms which identify them as offenders and which would not be consistent with the rules (Council of Europe 2009, p. 37). The practice of wearing orange uniforms as it is the case in England and Wales as well as in the Netherlands (and in the USA or South Africa) therefore is a violation of these human rights standards. From a European point of view, totally violating basic human rights standards has been the practice in some US states (and still is in Arizona) to make offenders work in so-called chain gangs, i.e., stigmatizing them by working in a kind of medieval outfit tethered together.

Two more human rights issues have to be mentioned in this context: Rule 37 of the ERJOSSM states that the costs of implementation shall not be borne by the juveniles or their families. And community work shall not be undertaken for the sole purpose of making a profit (Rule 45 ERJOSSM). Therefore, the demand of the South Australian Corrections Service to make the communities pay 200 $ per a day of community service with good reasons has been rejected by the local authorities (see http://www.abc.net.au/news/2012-04-05/community-service-charge-council/3934500).

Statistical Developments: Europe and Scandinavia

The use of community service as compared to other community sanctions and measures and imprisonment can be examined most effectively on the basis of enforcement statistics. The Council of Europe's penal statistics Space I (survey year 2009) and Space II (survey year 2010) provide information both on stock rates and flow rates (/100,000 population) in Europe.

Figure 2 displays the number of persons serving community service or other community sanctions under the supervision of probation services at a given day (31.12.2010). Figure 3 displays the same information on the base of flow statistics.

Community service has a noteworthy role in over 20 European nations with data in the Council of Europe's Space II survey. It has an essential role as the principal community sanction in Spain, Hungary, England and Wales, Scotland, Latvia, and all Nordic countries.

Many of the countries with high rates of community sanctions are also countries with relative high incarceration rates (see Dünkel/Lappi-Seppälä/Morgenstern/van Zyl Smit 2010, 1058 ff). Figure 4 compares the use of community service and imprisonment on the base of enforcement statistics.

Stock- and flow-statistics give the same result. There is no correlation between the number of people in community service and in prison at any given day (stock). Neither is there a correlation between the number of people starting community service or entering in prisons during the year (flow). The latter finding is, however, heavily influenced by one outlier (Spain). Should this observation be excluded, one would find a weakly positive correlation. In other words, countries that use extensively community service use also imprisonment more. This finding is confirmed by a comparison between all community sanctions and measures and imprisonment (below) (Fig. 5).

It seems that, in general, the use of imprisonment is not compensated by community sanctions. Countries with high incarceration rates are often (but not always) countries with high community sanctions rates. And countries with low incarceration rates seem to manage also with lower rates of community sanctions. However, one needs to take into account that this comparison includes only community sanctions with supervision under the probation service. It excludes formal types of sanctions such as suspended and conditional sentences without supervision, as well as fines. Should these measures be taken into account, low incarceration countries like the Nordic countries and Germany would rank high in the scale of noncustodial sanctions, but low in incarceration (see Lappi-Seppälä in this volume in the chapter on fines).

Evaluation Research: Recidivism and Reintegration into Society

Completion of Community Service Orders, Implementation of Programs, etc.

Community service in most countries is organized by the probation service and/or private nonprofit organizations of offender rehabilitation. One major problem in implementing community service is to organize working places and to allocate offenders to adequate working facilities. In particular, in large rural countries or local municipalities, it is sometimes difficult to find working possibilities close to the offender's residence. As community work has to be unpaid, the kind of work should be of a nature not to compete with regular paid work in the first labor market. Therefore, trade unions are keen to restrict community service to extra work which cannot be provided by normal enterprises.

Empirical studies have shown that the implementation of community service has been successful in many countries (see above) and today is the most important alternative sanction apart from fines and suspended sentences (probation). In this regard, it is important that most offenders assigned to community service programs are completing their working hours. Only a small minority fails. Factors for noncompliance or failure to complete the sentence are a too high number of hours imposed

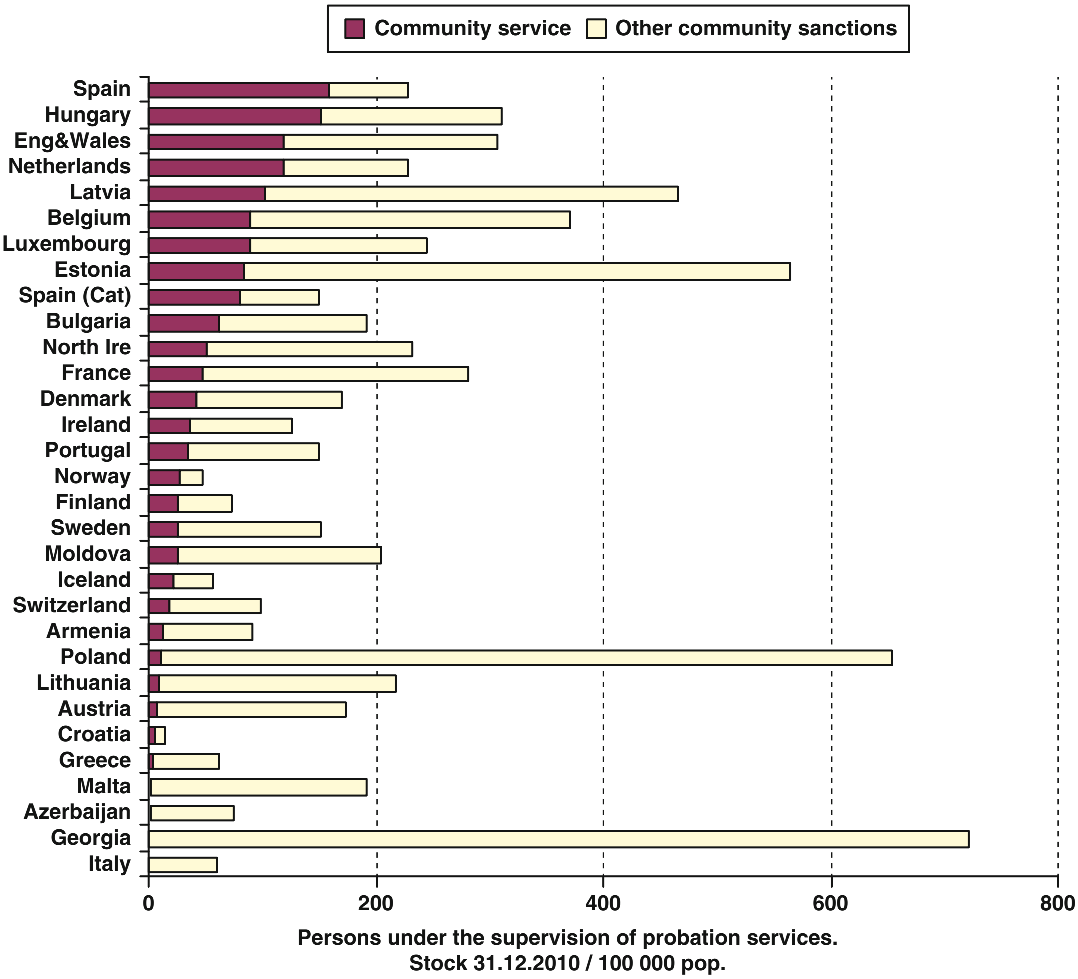

Community Service in Europe, Fig. 2 Persons under the supervision of probation services. Community service and other community sanctions and measures. Stock 31.12.2010 (Space II 2010 and Hildebrandt 2012 for the Nordic countries)

and/or personal problems of the offender to get to work (problems with alcohol and/or drugs, personality problems, unstableness, lack of practical experience with work, refusal to work etc.). The experience of the probation services shows that offenders in these programs often show multiple problems of reintegration which demonstrates the necessity of further rehabilitative measures compared to the sole allocation to a working facility.

One remarkable program in Germany should be mentioned. In Germany, community service is provided only for fine-defaulters as a substitute to imprisonment for nonpayment of a fine (Norway has in 2012 introduced a similar type of sanction ("Bötetjeneste") and also Finland currently is considering one). The German programs go back to the year 1983, but until the mid-1990s, most federal states ran those programs on a rather limited scale. The federal state of Mecklenburg-Western Pomerania had a specific problem as 22 % of prisoners in adult prisons were fine-defaulters. Therefore, in 1998, a state-wide project was founded that systematically created working facilities and in particular a system of contracting offenders who did not pay and react to the imposition of fine-default imprisonment. The restructuring of probation and private offender rehabilitation services resulted in

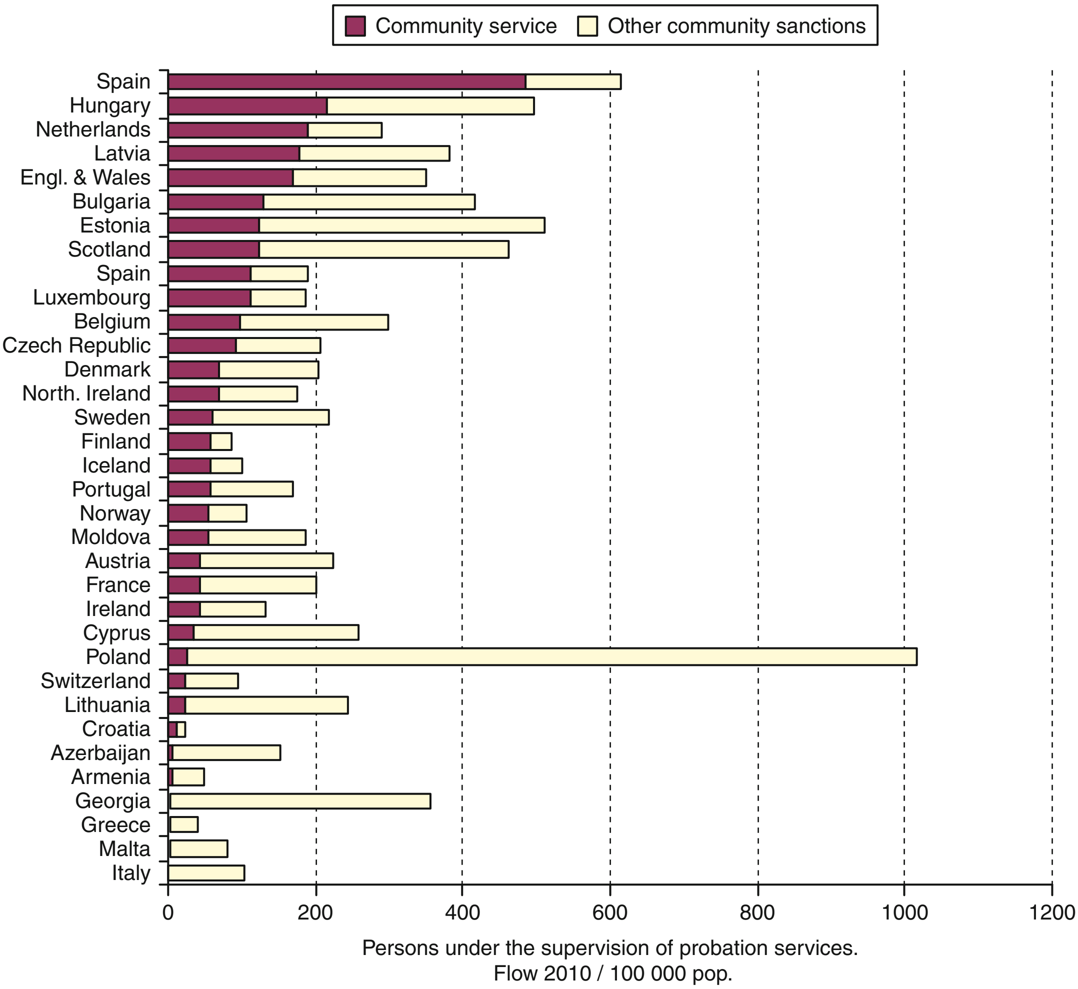

Community Service in Europe, Fig. 3 Persons under the supervision of probation services. Community service and other community sanctions and measures. Flow 31.12.2010 (Space II 2010 and Hildebrandt 2012 for the Nordic countries)

a structure of systematic allocation to 1,600 working facilities all over the large federal state of Mecklenburg-Western Pomerania. One special element of the project was to deliver social work and care for specifically difficult offenders who otherwise would not have completed community work. About 70 facilities have established such intensive care programs for fine-defaulters. Another pillar of the project was that fine-defaulters, even after they had been sent to prison, were allocated to community service programs and therefore often were released after a few days. The results during the 3-year pilot project starting in 1998 were remarkable: Daily numbers of fine-defaulters decreased from around 120–100 to 50–70, i.e., were reduced by half (see Dünkel and Scheel 2006). The results were so convincing that the Ministry of Justice decided to permanently appoint six social workers who are occupied with electing and allocating offenders to working facilities. The project has achieved a sustained success (the last numbers are from 2010), keeping fine-defaulters in prison almost at a low level achieved in 2001. The placement in community service facilities was in about 75 % of the cases successful, i.e., the offender completed at least partly community service and thus the state saved money by

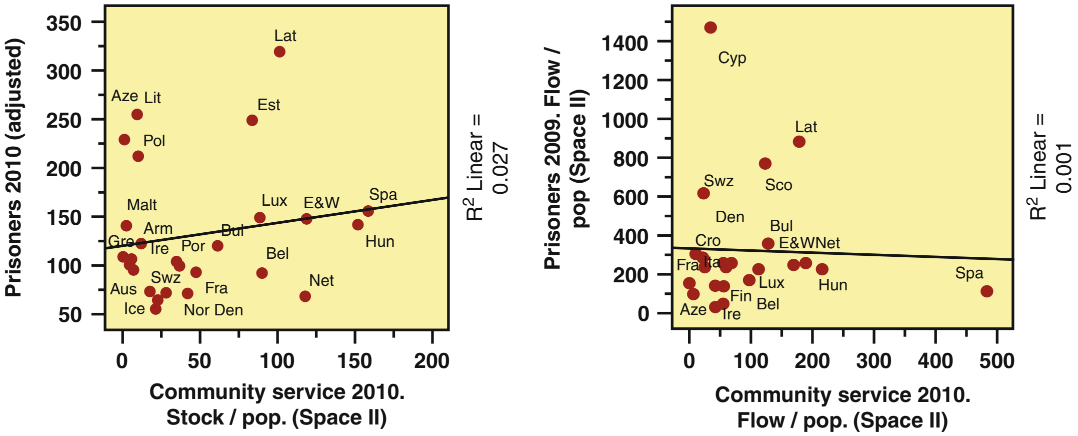

Community Service in Europe, Fig. 4 Prisoners and persons under community service Stock 31.12.2010 and flow 2009 (Source: Council of Europe, Space I and II)

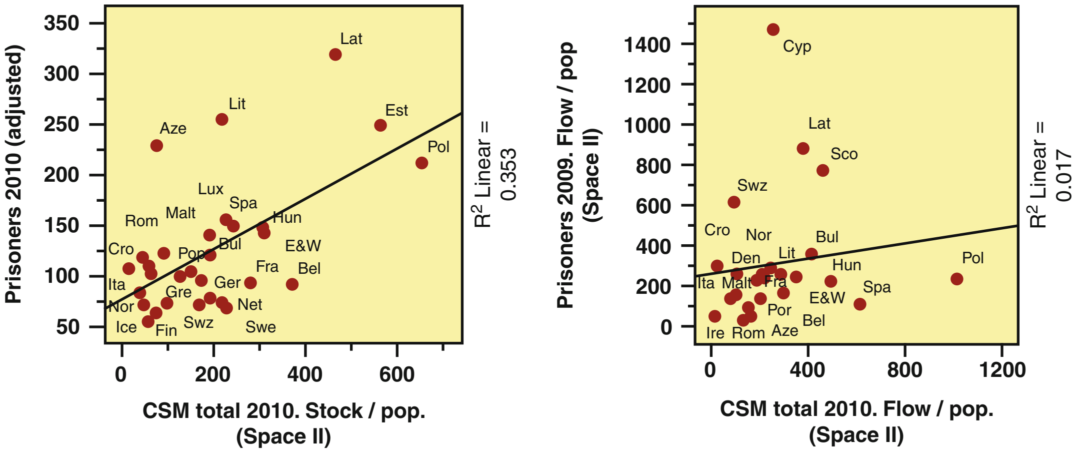

Community Service in Europe, Fig. 5 Prisoners and persons under supervision (all sanctions). Stock 31.12.2010 and flow 2009 (Source: Council of Europe, Space I and II)

avoiding imprisonment. In total, the state of Mecklenburg-Western Pomerania in 2010 saved about 1.6 million €, whereas the cost for the project (mainly the six social workers mentioned above) is about 450,000 € per year. Thus, the cost-benefit is more than one million € per year (see Dünkel 2011b).

Replacing Imprisonment or Net-Widening? The Example of Scandinavian Countries

All Nordic countries adopted or increased the application of community service in the 1990s and early 2000s. The following diagrams illustrate how these changes in the sanction structures were reflected in the application of other community sanctions and imprisonment in Denmark, Finland, Norway, and Sweden (see in more detail Lappi-Seppälä 2008) (Figs. 6, 7).

In Finland, community service came to replace prison sentences in 1993 without any notable net-widening. In Denmark, the extension of community service coincided with the reform that abolished the short-term (1–2 weeks) prison sentences. As it appears, the community service came to replace part of these prison sentences. But it also seems that part of previous

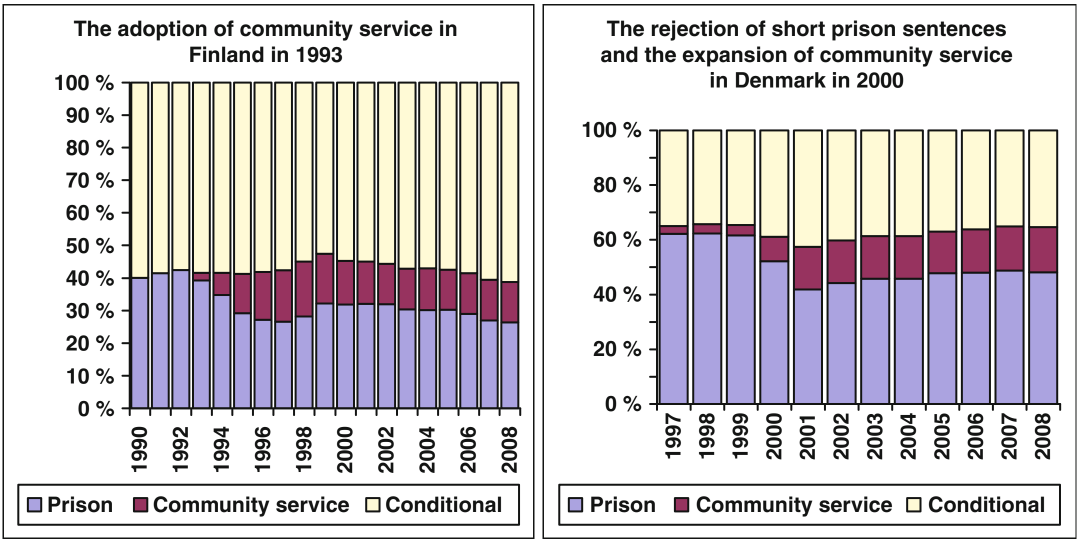

Community Service in Europe, Fig. 6 The relative share of community sanctions and imprisonment (/100,000) – Finland and Sweden (Source: Lappi-Seppälä 2012)

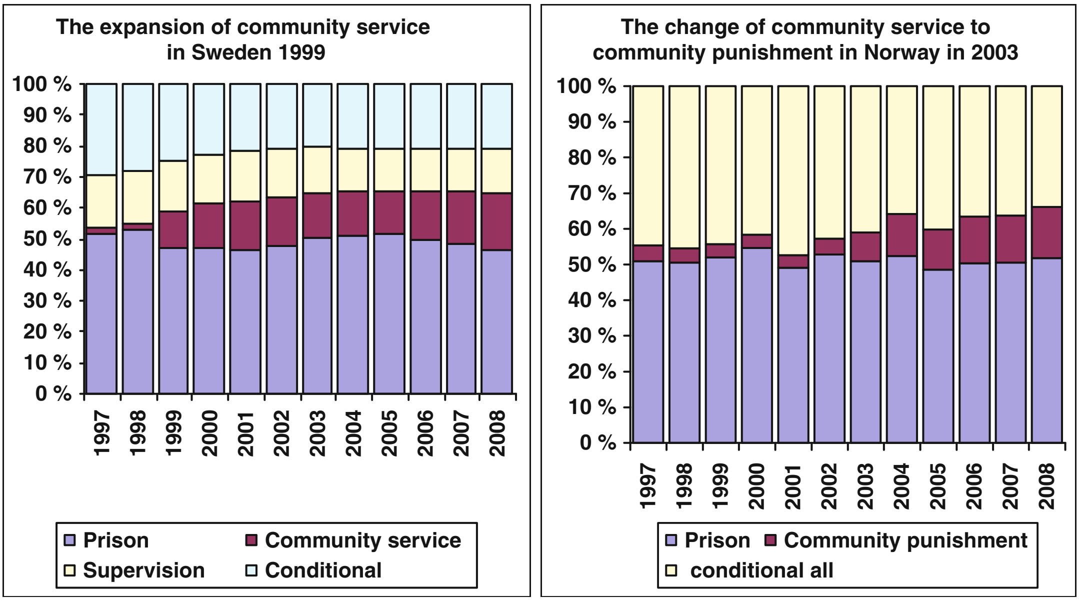

Community Service in Europe, Fig. 7 The relative share of community sanctions and imprisonment (/100,000) – Sweden and Norway (Source: Lappi-Seppälä 2012)

short-term prison sentences was replaced also by conditional sentences (Figs. 6, 7).

In Sweden, the expansion of community service in the year 2000 has replaced both conditional sentences and prison sentences. In 2002 Norway changed the name of "community service" to "community punishment" in order to enhance the credibility of the sanction. This increased the use of the sanction. However, this increase does not seem to be reflected in the share of prison sentences, but rather in the decline in the use of conditional imprisonment.

Community Service in Europe, Fig. 8 Reconviction rates after prison and community service in Finland, 1992

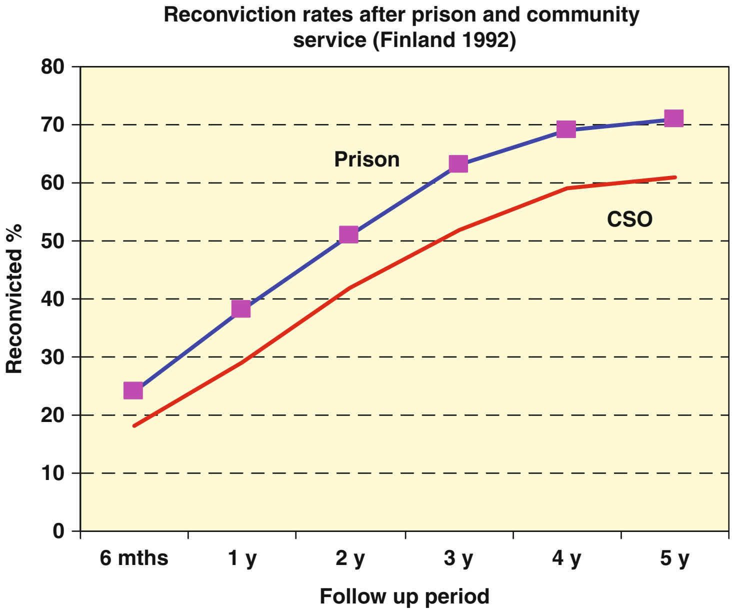

It looks like four out of five Nordic countries (Iceland included) have been successful in replacing prison sentences by community service. This success is associated with legislator's aims and intentions. In Finland, it was clearly stated that community service should be used only in cases where the accused would otherwise have received an unconditional sentence of imprisonment. This was ensured with specific legislative arrangements. Results were also seen in the statistics. Along with the increase in the number of community service orders from zero to 3,500, the number of unconditional sentences of imprisonment decreased by a little over 3,000. In Finland, annually some 3,500 community service orders are imposed by the courts. This represents around 35–40 % of the sentences of imprisonment which could have been converted (sentences of imprisonment of at most 8 months). This corresponds to some 400–500 prisoners (10–15 % of the prison population) of the daily prison population.

Assessing Reoffending Effects and Other Benefits

A Finnish study used quasi-experimental design and compared two matched groups of offenders: one sentenced to community service in that part of the country where community service was in use on experimental basis, and the other group of offenders with similar background and convicted for similar offenses (mainly drunken driving, which has been the major offense in Finland for community service). The follow-up period was 5 years. Only new sentences leading to conditional or unconditional imprisonment or community service were counted as recidivism. The study revealed a constant pattern showing that community service group had fewer reconvictions throughout the follow-up period. The differences in reconviction rates varied depending on where the counting was started. If started from the court's decision the difference after 5 years was 60 % for community service and 66 % for prison group. If started from the completion of the sentence, the figures were 62 % and 72 %. And if counting of the follow-up period in the community service group starts from the court's decision and in the prison group from the release to parole (which would be sensible), the difference in reconvictions would be 60 % (community service) and 72 % (prison, see Muiluvuori 2001) (Fig. 8).

In Denmark, Susanne Clausen (2007) has examined the preventive effect of community service compared to imprisonment. The study is limited to two offending categories: violent crime and traffic violations, including drunk-driving. In the case of traffic offenders, a change in the Danish penal code in 2000 made it possible to convert short prison sentences (up to 60 days) to CS. Capitalizing on this reform, Clausen compares recidivism levels between individuals sentenced to CS *after* the reform took place and those sentenced to unconditional imprisonment *before* the reform. As far as violent offenders, the quasi-experimental properties of the research design emanate from a change in the judicial policy that took place in the late 1990s. Within a relative short time period, the Danish courts increased the use of CS for violent offenders by a significant margin. In Clausen's research, the treatment group consists of violent offenders sentenced to CS in 2000–2001 while the control group includes individuals sentenced to prison in 1997–1998. To further enhance the comparability between the offender groups, Clausen's research controls for the impact of a large number of common risk factors, including age, gender, ethnicity, drug or alcohol misuse, mental problems, housing situation, education, employment, family type, and income. The subjects were tracked for 2 years for new crimes, starting from the date of conviction or release from prison (see Clausen 2007).

Findings from multivariate models suggest that, in general, characteristics like age, income, and prior criminality are more important as predictors of recidivism than the nature of the criminal sanction. On the other hand, the results provide some evidence of reduced recidivism associated with CS. For example, young traffic offenders (under age 25) were 36 % less likely to reoffend than those sentenced to prison. Among violent offenders, the unemployed seemed to benefit the most from community service. To explain these findings, Clausen suggests that CS protects young traffic, but also violent offenders from the stigmatizing from the stigmatizing effects of a prison sentence. On the other hand, by providing mandatory employment, CS may restore a sense of responsibility and routine for the previously unemployed offenders, thus enhancing their ability to reintegrate in the mainstream society. These benefits notwithstanding, Clausen's study failed to confirm that CS helps to lower recidivism across most offender populations. The modest effects associated with CS may reflect the fact that it was compared to either conditional or very short prison sentences which, in most cases, are unlikely to create much damage. At the same time, there is little reason to suspect that CS will *increase* recidivism. Thus, to the extent it is less costly and considered more humane than prison sentence, it seems like an attractive alternative. In a methodologically advanced study, Killias et al. 2010a divided offenders randomly in community service and control group (prison). Recidivism was studied using four indicators: (1) whether offenders were convicted and (2) the number of convictions; whether (3) offenders were arrested and the (4) number of arrests. In addition, the authors compared how much the offenders had advanced, how many arrests they had before and after the sentence. By all measures, the community service group survived better. However, the small size of the sample kept the statistical significance rates low. In 2010, Killias and colleagues published evidence from two controlled, randomized experiments that assess the causal effect of community service and electronic monitoring on different post-sentencing outcomes (Killias et al. 2010a, b). While the first study (Killias et al. 2010a) finds no difference between community service and traditional custodial sentences, the finding of the second study (Killias et al. 2010b) suggests marginally significant differences between two types of noncustodial sanctions, electronic monitoring and community service, in favor of the former. However, the evidence is still relatively limited, particularly as most of the existing studies rely on rather small samples, which may explain the frequent result that the effects found are nonsignificant.

Problems related to sample sizes have been effectively avoided in a largest study based on longitudinal official record data on adult offenders in the Netherlands (Wermink et al. 2010). The study compares recidivism after community service to that after imprisonment with a sample of 4,246 offenders. To account for possible bias due to the selection of offenders into these types of sanctions, a large set of confounding variables was checked using a combined method of "matching by variable" and "propensity score matching." The findings demonstrate that offenders recidivate significantly less after having performed community service compared to after having been imprisoned. In relative terms, community service leads to a reduction in recidivism of 46.8 % compared to recidivism after imprisonment. This finding holds both on the short term and long term, and for male and female offenders (Wermink et al. 2010) (Table 2).

The available evidence suggests that community service is (at least) a promising alternative, in terms of reducing recidivism (using the Maryland University methodology scoring, see, for example, MacKenzie 2006). Several results point to significant differences between offenders who are sentenced to imprisonment and offenders who participate in community service. The recent meta-analysis of Bonta et al. (2006) demonstrates a stronger effect size (.07) than earlier meta-analyses, which the authors explain by the better structured programs since 1995 considering "what-works" criteria (Bonta et al. 2006, p. 115).

In connection with community service, also other "non-reconviction benefits" need to be taken into account. These other beneficial features include positive contact to work life (and the resulting enhancement of offender's economical situation), better self-control over substance abuse, and better preservation of family ties. "Importantly, results also indicate that community service participants fare better with regard to long-term social benefit dependency and income, especially when we disregard the possible effects of time trends" (Andersen 2012, p. 26). In addition, violent offenders and offenders convicted for misdemeanors showed lower recidivism rates than the imprisonment group. A problem, still deserving attention is how to deal with offenders, whose substance abuse prevents the use of community service. One answer is provided in the form of Swedish contract treatment.

Community Service in Europe, Table 2 Recidivism rates after community service compared to imprisonment

Study	Results
Muiluvuori (2001) Finland	*Recidivism: New prison sentence in 5 years from the court decision*
Quasi-experimental (during the experimental reform)	Community service (60 %)
	Prison (66 %), nonsignificant
Matched groups by age, sex, and prior convictions	*Recidivism: New prison sentence in 5 years from the completion of the sentence*
	Community service (62 %)
	Prison (72 %), significant
Clausen (2007) Denmark	*Recidivism: New recorded offense 2 years from the court decision*
Quasi-experimental (before and after the reform)	Traffic. Reoffending risk for young traffic offenders (drunk-driving) 36 % lower in CS group
Multivariate analyses (central socio-demographic factors)	Violence. Largest benefit of CS for unemployed
Killias et al. (2000) Switzerland	*Recidivism: Convicted in 2 years*
Randomized experiment	Community service (21.4 %)
	Prison (25.6 %), nonsignificant
	Recidivism: Number on convictions in 2 years
	Community service (0.39)
	Prison (.64), nonsignificant
	Recidivism: Arrested in 2 years
	Community service (33.3 %)
	Prison (38.5 %), nonsignificant
	Recidivism: Number of arrests in 2 years
	Community service (0.76, but before 1.66)
	Prison (2.18, before 1.69), nonsignificant
	But see marked change in arrest rates before and after (above)
Wermink et al. (2010). Netherlands	Community service leads to a reduction in recidivism of 46.8 % compared to recidivism after imprisonment
Matched comparisons variable by variable	

(*continued*)

Community Service in Europe, Table 2 (continued)

Study	Results
Andersson and Alexandersson (1994) Sweden	No significant difference in recidivism rates in regression analyses
Bonta et al. (2002) Canada	Meta-analysis (37 studies)
	Effect size – 0.03 (= 3 % less recidivism)
Bonta et al. (2006) Canada	Meta-analysis (57 studies)
	Effect size – 0.07 (= 7 % less recidivism)

Outlook: The Potential of Community Service to Reduce the Prison Population and Increase the Reintegration of Offenders

European sanction policies have been characterized by two different trends during the 1990s and 2000s: an increasing use of prison and the adoption of new community sanctions. The first one reflects the growing punitive and populist trends in national crime policies; the latter seeks to counteract this development by offering more constructive, rational, and humane substitutes to incarceration. The overall effect of the expansion of community sanctions may appear disappointing: The number of alternatives has grown along with the increasing number of prisoners. Neither does it seem that countries that use extensively community alternatives have – in general – lower incarceration rates. Still, it would be premature to conclude that the efforts to enhance the use of community sanctions have been futile. Some jurisdictions have gained true success, giving thus also information of the models that do work. Failures, on the other hand, may offer important lessons on what not to do. In short, community sanctions do provide one functional tool against the overuse of imprisonment, once a number of lessons from different jurisdictions have been drawn upon. The key questions are: (1) how to ensure that these sanctions are applied in the first place, (2) how to ensure that they come to replace imprisonment (instead of replacing other noncustodial sanctions), and (3) how to uphold and maintain the general credibility of these sanctions. Current problems and achievements have been summarized as follows (Lappi-Seppälä 2003, see also van Kalmthout 2000):

1. Extra barriers should be constructed in order to ensure that the new alternatives are really used instead of imprisonment. In most countries, community service seems to substitute prison sentences only in roughly 50–60 % of cases. This rate can be improved by demanding directly – as is the case in Finland – that only prison sentences may be commuted to community service (leading to a "replacement rate" of over 90 % in Finland). Another way would be to define new alternatives as modes of enforcement of prison sentences, as has been done in Iceland.
2. *Effective use of new alternatives and coherent sentencing practices require clear (statutory) implementation criteria.* The courts should be given clear guidance as to when and for whom new sanctions are to be used. The role and position of new alternatives in the existing penal system (how they relate to other sanctions) should also be clarified.
3. *The overall success of any community sanction requires resources and proper infrastructure.* Community-based sanctions can only be applied within a community-orientated infrastructure geared to the specific requirements of these sanctions. Their implementation is dependent on the existence of an organization like the probation service. Often cooperation with private, semipublic, and public organizations or institutions is also required. The state and the local communities should provide the necessary resources and financial support.
4. *Supervision, support, and swift reactions are needed in order to keep the failure rates down and to maintain the general credibility of new sanctions.* The less control and supervision, the higher is also the dropout rate. There should also be a clear and consistent practice when the conditions of the sentence are violated. Varying and sloppy practices create

mistrust and resistance on the part of public prosecutors, the judiciary, and the public.

5. *Arrangements should be made in order to enhance the motivation of the offender for cooperation and mutual trust.* New alternatives usually require the offender's consent and cooperation. Treating the offender, not as a passive object of compulsory measures, but as an autonomous person, capable of reasoned choices, is a value by itself, and as such, it should be encouraged whenever possible. In addition, experience indicates that explicit and well-informed consent is a highly motivating factor for the offender. Through his/her consent, the offender has also become committed to the required performance in a manner that gives hope for good success rates.
6. *Issues of equality and justice must not be neglected.* Community sanctions may often lead to discrimination, since they are easily used for socially privileged groups of offenders. Measures must be taken to avoid social discrimination. Clear and precise implementation rules and procedures are one important means to this end. Another way is to tailor the system of community sanctions to meet the demands of different offender groups with their different problems. Sweden, for example, has a specific sanction – "contract treatment" – for those who suffer from drug or alcoholic addiction as a substitute for short-term prison sentences. Finland adopted in 2011 electronically monitored supervision order, supported by program work for offenders who are excluded from community service due to their addiction problems.
7. *The idea has to be sold over and over again.* Initial success does not guarantee that things will go well also in the future, just by themselves. Prosecutors and judges may lose their confidence, the enforcement agencies may lose their motivation, and the general public may withdraw its support. Maintaining the general credibility of the community sanctions and demonstrating their appropriateness is an ongoing process which does not end with the adoption of the requisite legislation and the arrangement of an initial training phase. The key groups, responsible for the implementation of the sanctions, must be given constant training and general information of the general benefits of community sanctions and the drawbacks of the wide use of custodial sanctions. Taking care of community relations is also important: The community should be informed of the benefits and crime control potential of community sanctions. Also the value of volunteer work needs a clear recognition. Finally, the practices must be subject to impartial scientific evaluation in order to obtain necessary information for further development.
8. *Do not turn community service into a mere punishment in the community.* The original objective of community service was to provide a socially constructive and less damaging alternative to imprisonment. Community service has rehabilitative and reintegrative potential, which must not be "sold" during the efforts to widen its applications and finding political support. Policy planners should resist the temptation of presenting community service to the public only as a "credible" and "demanding," if not outright humiliating, alternative – as has taken place in some jurisdictions. Community service should be oriented toward the restorative justice and the rehabilitation ideal and implemented that way (which would be in line with international human rights standards) and not be left to penal populism and repressive orientations of "punishment in the community."

Recommended Reading and References

Albrecht H-J, Schädler W (1986) Community service – a new option in punishing offenders in Europe. Max-Planck-Institut für ausländisches und internationales Strafrecht, Freiburg i. Br

Andersson T, Alexandersson, L (1994) Samhällstjänst som alternativ till fängelse. Utvärdering av 1990-1992 års försöksverksamhet. BRÅ-PM 1994:3.

Andersen SH (2012) Serving time or serving the community? Exploiting a policy reform to assess the causal effects of community service on income, social benefit dependency and recidivism. University Press of Southern Denmark, Odense

Bonta J, Wallace-Capretta S, Rooney J, McAnoy K (2002) An outcome evaluation of a restorative justice alternative to incarceration. Contemorary Justice rev 5(4), pp 319–338

Bonta J, Jesseman R, Rugge T, Cormier R (2006) Restorative justice and recidivism: promises made, promises kept? In: Sullivan D (ed) The handbook of restorative justice: a global perspective. Routledge, New York, pp 108–120

Clausen S (2007) Samfundstjeneste – Virker det? Jurist- og Økonomforbundets Forlag, København

Council of Europe (2009) European Rules for juvenile offenders subject to sanctions or measures. Strasbourg, Council of Europe Publishing.

Dünkel F, Lappi-Seppälä T, Morgenstern C, van Zyl Smit D (2010) (eds.) Kriminalität, Kriminalpolitik, strafrechtliche Sanktionspraxis und Gefangenenraten im europäischen Vergleich. Mönchengladbach, Forum Verlag Godesberg.

Dünkel F (2011a) Germany. In: Dünkel F, Grzywa J, Horsfield P, Pruin I (eds) Juvenile justice systems in Europe. Current situation and reform developments, vol 2, 2nd edn. Forum Verlag Godesberg, Mönchengladbach, pp 547–622

Dünkel F (2011b) Ersatzfreiheitsstrafen und ihre Vermeidung. Aktuelle statistische Entwicklung, gute Praxismodelle und rechtspolitische Überlegungen. Forum Strafvollzug – Zeitschrift für Strafvollzug und Straffälligenhilfe 60, pp 143–153

Dünkel F, Pruin I, Grzywa J (2011) Sanctions systems and trends in the development of sentencing practices. In: Dünkel F, Grzywa J, Horsfield P, Pruin I (eds) Juvenile justice systems in europe. Current situation and reform developments, vol 4, 2nd edn. Forum Verlag Godesberg, Mönchengladbach, pp 1,649–1,716

Dünkel F, Scheel J (2006) Vermeidung von Ersatzfreiheitsstrafen durch gemeinnützige Arbeit: das Projekt "Ausweg" in Mecklenburg-Vorpommern. Mönchengladbach, Forum Verlag Godesberg

Goldson B (2008) Dictionary on youth justice. Willan, Cullompton

Heinz W (2012) Das strafrechtliche Sanktionensystem und die Sanktionierungspraxis in Deutschland 1882–2010 (Stand: Berichtsjahr 2010) Version: 1/2012', Internet-Publication http://www.uni-konstanz.de/rtf/kis/Sanktionierungspraxis-in-Deutschland-Stand-2010.pdf

Hildebrandt S (ed) (2012) Nordisk statistic for kriminalforsorgen I Danmark, Finland, Island, Norge og Sverige 2006–2010. Direktoratet for Kriminalforsorgen, Copenhagen

de Jonge G (1999) Still 'slaves of the state': prison labour and international law. In: van Zyl Smit D, Dünkel F (eds) Prison labour: salvation or slavery? Dartmouth, Aldershot, pp 313–334

Killias M, Gilliéron G, Villard F, Poglia C (2010a) How damaging is imprisonment in the long-term? A controlled experiment comparing long-term effects of community service and short custodial sentences on re-offending and social integration. J Exp Criminol 6, pp 115–130

Killias M, Gilliéron G, Kissling I, Villettaz P (2010b) Community service versus electronic monitoring – what works better? Br J Criminol 50, pp. 1155–1170

Klement C (2010) Samfundstjeneste. En effektevaluering. Justitsministeriets Forskningskontor. http://www.justitsministeriet.dk/sites/default/files/media/Arbejdsomraader/Forskning/Forskningsrapporter/2011/Samfundstjeneste_-_En_effektevaluering.pdf. Accessed Aug 2011

Lappi-Seppälä T (2003) Fines. Community sanctions as a means to restrict the use of imprisonment? In: European Conference for Probation (Ed) Probation in Europe. Bulletin of the conférence permanente Européenne de la Probation, no 27. http://www.cepprobation.org/uploaded_files/juni%202003%20PDF%20format%2027%20-%20E.pdf Accessed June 2003

Lappi-Seppälä T (2008) Crime prevention and community sanctions in Scandinavia. Tokyo: UNAFEI (Annual report for 2007 and resource material series no 61). http://www.unafei.or.jp/english/pdf/RS_No74/No74_00All.pdf

Lappi-Seppälä T (2012) Penal policies in the Nordic Countries 1960–2010. J Scand Stud Criminol Crime Prev 13:85–111.

MacKenzie DL (2006) What works in corrections: reducing the criminal activities of offenders and deliquents. Cambridge University Press, Cambridge

Muiluvuori M (2001) Revidivism among people sentenced to community service in Finland. J Scand Stud Criminol Crime Prev 2(1), pp 72–82

Space I (2010) Council of Europe. Annual penal statistics by Aebi M F, Delgrande N. University of Lausanne, Switzerland. http://www3.unil.ch/wpmu/space/files/2011/02/SPACE-1_2010_English1.pdf

Space II (2010) Council of Europe. Annual penal statistics. Persons serving non-custodial sanctions and measures in 2010. Survey 2010, by Aebi MF, Marguet Y, Delgrande N. University of Lausanne, Switzerland. http://www3.unil.ch/wpmu/space/files/2011/02/Council-of-Europe_SPACE-II-2010-E1.pdf

Tiffer-Sotomayor C (2000) Jugendstrafrecht in Lateinamerika unter besonderer Berücksichtigung des Jugendstrafrechts in Costa Rica. Forum Verlag Godesberg, Mönchengladbach

van Kalmthout AM (2000) Community sanctions and measures in Europe: a promising challenge or a disappointing Utopia? In: Council of Europe (ed) Crime and criminal justice in Europe. Council of Europe Publishing, Strasbourg, pp 121–133

van Ness D (1986) Crime and its victims. Inter Varsity Press, Downers Grove

van Zyl Smit D, Dünkel F (eds) (1999) Prison labour: salvation or slavery? Dartmouth, Aldershot

Villettaz P, Killias M, Zoder I (2006) The effects of custodial vs. non-custodial sentences on re-offending: a systematic review of the state of knowledge. Campbell Syst Rev 2006, p 13

Wermink H, Blokland A, Nieuwbeerta P, Nagin D, Tollenaar N (2010) Comparing the effects of community service and short-term imprisonment on recidivism: a matched samples approach. J Exp Criminol 6, pp 325–349

Wright M (1991) Justice for victims and offenders: a restorative response to crime. Open University Press, Bristol

Community Supervision

▶ Behavioral Management in Probation

Community-Oriented Policing

▶ Community Policing

Comparative

▶ Comparative Incarceration

Comparative Case Analysis

▶ Linkage Analysis for Crime

Comparative Courts and Sentencing

Rasmus H. Wandall
Faculty of Law, University of Bergen, Bergen, Norway

Overview

Comparing sentencing arrangements is a vital part when developing and studying criminal sentencing. Few reform programs move ahead and little research is done without consulting experiences within other jurisdictions. To some degree, the United States represents an exception to this rule. This is because comparative research between federal states has become more common than earlier. Comparisons with other countries, be they European, South American, Asian, or African, remain less frequent. This is despite the efforts of a number of leading scholars who refer to experiences and sentencing models in other countries.

This entry gives an account of the subject matter of comparing criminal sentencing, the purposes of comparing, and the typical approaches of comparing criminal sentencing. In so doing, the entry discusses the issues at stake when comparing and the limitations that flow from the different approaches to comparing criminal sentencing. The entry's last section devotes special attention to the relevance of globalization to the act of comparing across borders.

The Subject Matter of Comparing Sentencing

For the present purposes, criminal sentencing covers actions through which it is determined what penal sanctions an offender must be subjected to following a criminal offence. Although this description, like any other, is ethnocentric, it does cover the process of police, prosecutorial, and court decision-making, the sanctioning that follows, just as it includes the organizational structures that sustains and develops the decision-making processes.

Typically, three different subject matters come into play when comparing criminal sentencing. The most frequent is the outcome of sentencing: the sanction type (prison, alternative sanctions, fines, etc.) and the amount of sanction (length of prison term, size of fine, hours of community service). For example, comparing the rate of imprisonment per capita or the number of prison sentences in the United States and in England and Wales. Another subject matter of comparison targets the legal structures that govern the sentencing decision-making, for example, comparing the many different constructions of numerical sentencing guidelines that have been built in different jurisdictions in the United States since 1980 or comparing the Swedish statutory sentencing law with any of these numerical guideline systems. A third subject matter of comparison takes seriously the social practices of the everyday handling of cases and the organizational framework of sentencing, in court as well as in prosecution. Inspired by the traditions of socio-legal scholarship, these studies rely neither on the law nor aggregate sentencing outcomes but compare the organization, operations, and views of those who handle criminal cases, for example, comparing judges' views on sentencing in Finland and in England and Wales or comparing court systems and their political roles in industrialized nations. These different subject matters should not be mistaken for one another. They bring different consequences to the comparative analysis and most importantly provoke us to ask how things would look with a different subject matter of comparison.

Purposes of Comparing Criminal Sentencing

As in comparative criminology in general (Nelken 2011), we can discern three typical purposes of comparing criminal sentencing decision-making: (1) the policy purpose, (2) the social scientific positivistic purpose, and (3) the contextual or interpretive purpose. The three purposes often overlap, yet are distinct.

First, there is a long tradition for using comparative methodology in policy research. In most sentencing reforms policy-makers look to other jurisdictions' experiences as a tool to find the best solution to their own challenges of sentencing reform. This practice of transferring sentencing experiences from one jurisdiction to another raises a number of comparative concerns. These are addressed later in the entry. Second, the social scientific positivistic purpose is about comparing sentencing decision-making and its organization to test theories about relevant causes and effects of criminal sentencing. For example, testing the explanatory value of income inequality to the level of prison rates in different countries or explaining the change from discretionary to numerical guideline informed sentencing in many jurisdictions in the United States. The third and interpretivist purpose of comparing is to unravel the contextual embeddedness of criminal sentencing and studying how sentencing decision-making, its regulation, and its procedures and sanctions are differently constructed. One example would be to show that sentencing discretion and its boundaries are intrinsically linked to local cultures.

Approaches to Comparing Criminal Sentencing

The approach to comparing is intertwined with both the purpose and the subject matter of comparing. Indeed, as the headings below confirm, there is a close connection between the different purposes of comparing sentencing and the different approaches to it. Nevertheless, it is both important and helpful to make the distinction when presenting three widespread and different approaches:

1. The legal categorical approach
2. The predictive and explanatory social scientific approach
3. The interpretivist approach

Legal Categorical Approach

A legal categorical approach is deeply connected to the legal structure as subject matter and has legal scholarship and dogmatics as its basis. The approach is used to categorize sentencing laws and systems, typically based on functional characteristics and their legal historical heritage. This approach helps to elucidate important structural properties of sentencing laws, but it also functions to cover the many relevant social, cultural, and political differences between jurisdictions. The best-known categorization in wider comparative studies is that by comparative legal scholars Kötz and Zweigert (1998) in their introduction to comparative law. They systematically organize the different legal systems in legal families: the Anglo-American, the romanistic, the germanistic legal family, the Nordic legal family, then add the law in the Far East, and finally religious legal systems. In comparative sentencing research, the categorization is most often seen in the simple distinctions between Anglo-American common law and continental models of civil law. The Nordic countries are sometimes placed in the continental and civil law category and sometimes recognized as also carrying elements recognizable in the Anglo tradition of law.

Though it is disputed whether distinctions such as civil law and common law and adversarial and inquisitorial law can serve as units of analysis, it is indisputable that these distinctions are actually used and that they at least historically carry some legitimacy. Civil law and inquisitorial models arose from Roman law and were subsequently shaped by European continental legal developments. Common law and the adversarial model have their origin in English law which subsequently spread to the United States of America and other places where the English sought influence. Common law is characterized by the central role of the trial, the judicial decisions, and the recorded precedents, all of which are important vehicles for the development of law. Contrary in civil law models, through codification, every offence is ideally preconceived in written legal statutes and the judge has the role of reading, understanding, and applying the statute content correctly – and most importantly has the final word. In a sense, the key axis in this model is that between the legal subject and the written rule, whereas in the common law model, it is rather that between the legal subject and the trial.

How does this help to compare criminal sentencing? First of all, a distinction is often made between the civil law tradition of structuring sentencing through legal statutes and the common law tradition of governing sentencing through trial experiences. This is a frequent way of comparing continental European and Anglo-American sentencing law. Indeed, France and Germany do have complete criminal codes defining every offence and stipulating the sentencing options for each. Many Anglo-Saxon jurisdictions still have no general criminal code, and still place the courts and the trial in a comparatively more significant role in sentencing. The basic structures of sentencing law are different. One could point to the adversarial nature of plea bargaining in England and Wales and in the United States and contrast it to the continental styles of legality in sentencing procedure. Both in Germany and in France, it is the judge who has the final decision, even when procedures involve consensual elements. Problems with these distinctions between common and civil law arise when they are used to describe entire countries and sentencing systems. Despite the historical heritages, sentencing laws and practices have evolved remarkably over time, and today it is problematic at best to describe entire systems as common or civil law and as inquisitorial or adversarial. As Pizzi (2008) illustrates, it is difficult to categorize the entire sentencing systems of the United States as common law systems as such. More than half of its jurisdictions have, since 1980, implemented legal sentencing rules that place the court under substantial legal control. The techniques employed include minimum sentencing laws and administrative sentencing guideline systems combined with various administrative governance institutions and enforced monitoring mechanisms. In England and Wales legislative structuring of sentencing has increased since the middle of the twentieth century. Today the sentencing approach towards many criminal offences is governed by binding sentencing guidelines, promulgated by a national sentencing council, and enforced by the upper courts. It is more descriptive then in these jurisdictions to emphasize the joint structuring of sentencing by the upper courts, the more recent administrative sentencing guidelines, and the legislature.

In the same way it makes little sense to categorize the Italian system as a civil law system as the country has implemented plea bargaining in a wide area of criminal cases. Most other continental countries have implemented some version or other of consensual procedures, eating away at the traditional principle of legality and central role of the judge assuring the law to rule sentencing (Thaman 2010). This does not mean that these basic distinctions between civil and common law and between inquisitorial and adversarial systems serve no analytical purpose. First of all, the distinctions are a welcome tool to reduce the otherwise dominant use of jurisdictional borders as a key distinction (Nelken 2007). Second, if we handle these concepts as representing legal traditions in the way that Glenn (2010) does in his comparative legal theory, we could use the distinctions as carriers of norms and legal forms, travelling from one country to another, constantly evolving with the local and different pressures of law, politics, and culture in each jurisdiction, and only expressing themselves in much more concrete ways as rules, practices, and arrangements that reflect more than a singular legal tradition (Field 2009). As such these distinctions do serve the purpose of sorting out general lines of traditional sentencing and procedure to further investigate their concrete implementations.

There are attempts at breaking free from the dominant distinctions of civil law and common law. For example, Frase (2008) explores the use of the established comparative criminal procedure model developed by Damaska (1991). Justice systems are categorized on two diagonal dimensions. In one dimension, systems are tending to be hierarchical or to be coordinate systems of justice. In the other dimension, systems are tending to be policy implementing or conflict solving. This does allow more flexibility with respect to legal categorical comparisons. However, it does not change the problem that legal categorizations of entire systems using such simple categories or dimensions lack significant descriptive and analytical power.

More fundamentally, comparative legal theory has evolved to challenge the legal categorical approach. Recognizing the increasing traffic between borders of legal ideas, constructions, and institutions, advanced comparative legal theory today seeks to accommodate for the mixed legal systems that have resulted. It is accordingly put forward that most legal systems are mixes of different legal constructions and solutions. They therefore represent different traditions and countries from which ideas and models for legal constructions and institutions have been borrowed, imposed, codeveloped, or other. The above described widening use of plea bargaining or similar consensual procedures are good examples, as are sentencing commissions and councils, sentencing guidelines systems, sentencing information systems, problem-solving courts, alternative dispute resolution mechanisms, and prosecutorial diversion programs. The headlines that are used to make sense of these developments include "legal transplants" (Watson 1993), "legal irritants" (Teubner 1998), "law as transposition" (Örücü 2002), and "legal borrowing" (Nolan 2009). Partly the concern here is to accommodate for the significant role that borrowing plays in comparative studies. Partly the concern is if such transport of legal institutions from one jurisdiction to another makes sense at all within the framework of comparative legal theory. The fundamental question is if a legal rule is bound to its social and cultural meanings, how can it move elsewhere without lousing this meaning?

Yet, despite these significant developments, the distinctions between civil and common law and between adversarial and inquisitorial procedures continue to play a substantive role in scholarly writings on comparative criminal sentencing law.

Predictive and Explanatory Social Scientific Approach

As positivistic social science in general, this approach compares variation in sentencing arrangements between jurisdictions, such as countries, states, or regions, in order to identify the causes and consequences of sentencing arrangements. Different from the legal classificatory approach, all social phenomena can be used in the study of sentencing whether legal or not. The approach is used to compare various aspects of sentencing. For example, comparing the degrees of voluntariness in sentencing guidelines to see the effects on levels of punishment, comparing fine levels or the frequency and length of prison sentences, comparing different paths that reduce prison sentences, comparing the use of alternatives to imprisonment and their sentencing arrangements, comparing the effects of sentencing reforms on court practices. comparing to explain increases in prosecutorial authority in criminal sentencing, comparing the consequences of political culture on criminal sentencing and court practice, identifying the causes and effects of different sentencing policies and the structural reforms of sentencing, (Tonry and

Hatlestad 1997; Tonry 2007, 2009; Tonry and Frase 2001), and comparing in order to test arguments about correlations between penal levels and sentencing ideologies. Though these many approaches differ from one another, they can all be broken down to accounts of variation between jurisdictions in the correlation between dependent and independent variables.

Four issues are particularly sensitive when using this approach. The first concerns the sample and population. Predictive and explanatory models make use of sample countries to generalize, but the extensive variation between countries makes it difficult to generalize both regionally and globally about sentencing. For example, Cavadino and Dignan (2007) argue that penal levels, measured by overall imprisonment rates, vary with the type of political economy. The four proposed typologies are the neoliberal (USA, UK, Australia, New Zealand, South Africa), the conservative corporatist (Germany, France, Italy, the Netherlands), the social democratic (Sweden, Finland), and the oriental corporatism (Japan). Their aim is to find similarities between groups of countries on significant variables that simultaneously distinguish those countries from others belonging to different typologies. Using sample countries from North America to east Asia, the authors can make the claim of sampling a global scale of modern industrialized countries. However, many countries pointing in different directions are not included. For example, what would happen if China is in the comparison? China has a very high rate of imprisonment, but does not adhere to a neoliberal societal model (Nelken 2011). Within the African scale of modern industrialized countries, it is only South Africa that is included in the comparison. However, Kenya is equally a comparatively large market economy in Africa, yet with much lower prison rates than South Africa.

Similarly, Lappi-Seppälä (2011) compares correlations between prison rates and a number of social and political structures in 30 different countries. He argues that among others, moderate penal policies are related to consensual and corporatist political cultures with high levels of social trust and political legitimacy. The sample consists of 30 countries from continental Europe (hereunder Russia), North America, and it includes Australia and New Zealand. Yet, it does not include any Asian (apart from Russia) or African countries. For the present purposes we are reminded that explanatory models are limited to the countries chosen for comparison and that it is problematic to generalize both regionally and globally about correlates to sentencing and sanctioning.

The second issue concerns the dependent variable. Whether it is prison rates, the use of alternative sanctions, fines, the legal framework of sentencing, or sentencing uniformity, it is central to a positivistic approach that the dependent variable is comparable across jurisdictions. When comparing prison rates, what counts as prison? Is weekend prison and open prison with daily outside working or educational activities outside included? Is electronic monitoring and home imprisonment included? Does it make sense to compare sentences served under the conditions in local prison facilities in the state of Virginia with those of Germany? Furthermore, increasingly important is the measuring of alternative sanctions. Its features, characteristics, and use have expanded rapidly over the last decades. But while many of the headlines are the same across borders, the content of alternative sanctions rarely is. Some are keenly aware of this challenge (Farrington et al. 2004). Taken wider, the issue here is that looking at criminal sentences across nations assumes sufficient non-variation both in sanctioning mechanisms and in competing social sanctioning mechanisms. For example, in the context of Japan in which disciplinary culture varies markedly from that of England and Wales and that of the United States, there is a range of other measures that are used for social control. South Africa, even though it does represent a punitive system, also represents local governance technologies that are applied to situations in which other countries, like England and Wales, France, or Norway, would use formal sanctions of imprisonment and community service. A social scientific predictive or explanatory approach must consider these factors.

The third issue that must be considered relates to the independent variables. When comparing two or more jurisdictions, everyone of the independent variables needs to be comparable across jurisdictions and relevant variations between them must be reflected. This includes both criminal offence variables, procedural variables, and variables related to the criminal offenders, and it includes the social, political, and cultural structures of the sentencing jurisdiction. Both legally and empirically, violent offences may look different, the offenders may belong to different groups of society and the form of trial may vary from place to place. In some countries less serious offences are not categorized as crimes but as administrative wrongs and are handled accordingly. Some countries rely on police and prosecution to finish many cases before they reach court. In some jurisdictions, like England and Wales, timely confessions are awarded measurable reductions in criminal sentencing, while in other countries confessions only affect the form of evidence and trial. Likewise, in some jurisdictions, there is little space for the inclusion of a personal investigation report, while in others it takes center stage. A further and important concern with independent variables is the covariance between them. For example, Lappi-Seppälä (2011) relates low levels of inequality, democracy, high social trust, and welfare society with moderate penal levels. However true this may be, it is associated with considerable methodological concern that there is a high covariance between these particular variables.

A related concern is to what extent we can meaningfully distinguish sentencing from any of the relevant independent variables. One important example is the ties between sentencing and procedural arrangements. Many criminal justice systems allow for a range of alternatives to trial: case diversions, conditioned charge withdrawals, and plea agreements, to mention but a few examples. Many such procedural dispositions are closely related to the expected sentence. This makes sentencing a procedural variable also and thus makes it difficult to distinguish between sentencing and procedure. Another example is the connection between criminal sentencing, strategies of other justice institutions, and overall penal policies. The typical assumption is that penal policy – to some degree – determines the direction of sentencing practices and that sentencing is distinct from activities of other justice institutions. However, in most Western countries the individual branches of penal administration have increased their interdependencies in a number of different ways such as shared strategies, policies, goals, measures, daily interacting operations, and overall common infrastructures. The consequence is an increasingly larger difficulty of distinguishing the effects of one agency from the other. This naturally makes measuring such variables across borders difficult.

The fourth issue with the positivistic approach is the interpretivist concern that comparing variable elements of sentencing regimes is problematic because every element is subject to the individual meaning structures of the compared systems, societies, and cultures. For example that judicial discretion is deeply imbedded in the cultural constructs of the society in which it is practiced, as are the meaning of "crime" and of "community" sanctions and of "youth" problems. See further below about the interpretivist approach.

There are no easy answers to any of these concerns and questions. Yet, this should not discourage the use of this approach. Rather, important limitations of what can be inferred from such comparisons and the extraordinary care we must take when designing and interpreting comparisons should be borne in mind.

Interpretivist Approach

The interpretivist approach takes off where the explanatory and predictive approach stops. It lays bare the cultural and contextual embeddedness of penal arrangements. For example, on the basis of comparing the United States with France and Germany, arguing that the punitive level in the United States is conditioned by the particular American resistance to state power and drive towards egalitarian social status (Whitman 2003). Another example could be contrasting the judicial discretion in Islamic law courts in

Morocco with that of courts in the United States' jurisdictions, showing how local (cultural) concepts and ideas inform the exercise of judicial discretion. Comparing judges' accounts and views on sentencing in different countries can show that any understanding has to be sensitive to the judges' different perceptions of harm, crime, sanction, and sentencing objectives.

Characteristically, interpretivist approaches use the sentencing and penal arrangements as indicators of wider cultural and political arrangements. However, taking culture seriously is not reserved to the interpretivist approach. Also predictive and explanatory approaches make use of culture. It is oftentimes difficult to spot the difference between the approaches, and in many studies the two are mixed.

The interpretivist approach firstly requires the observation and understanding of local cultural concepts and meanings and of how they manifest themselves in the different structural layers of criminal sentencing. This is a methodologically difficult task demanding linguistic, cultural, and historical sensitivity that only few have and only the fewest have the months and years to invest. Is one barred from taking this approach if one does not speak the language? How long must one stay in a given place? These are relative questions that limit the reliability of interpretive comparative studies, but that should not bar them. Some scholars use the collaboration between two or more colleagues, representing the legal cultures compared, as the central methodology.

Another important issue of the interpretivist approach is that the boundaries of what counts as legal and penal culture around sentencing can at most be loosely defined. We may cautiously characterize it as conveying the sets of ideas, expectations, and values that in practice and in policy-making is held with regard to penal law, the criminal process, and the sentencing.

A third important issue is that far from everything comes down to local cultures. Comparative studies today take seriously that techniques, ideas, concepts, and meanings travel and are sometimes shared by many jurisdictions, however, different concrete shape they acquire locally. In relation to the specific transport of legal concepts and ideas, see also above about the legal categorical approach. For example, many scholars have described a Western development towards curtailing judicial discretion, increasing the level of standardization and formalization of criminal sentencing. Aas (2005) calls this a development towards "sentencing at a distance." A significant body of research has looked at how such transnational trends and ideas merge with local legal and political cultures. Some have described how the policy implementation of three-strikes policies takes on local particularities in different countries. Others have described how transnational movements of new court types result in remarkably different forms of courts when institutionalized in different cultures of criminal justice. Yet another example is the different legal constructions of a principle of equality in sentencing that has emerged in different countries with the otherwise transnational development of standardization and formalization of criminal sentencing. The state of Minnesota provides one of the most prominent examples of numerical sentencing guidelines with a strong institutional tie to prison capacity constraints. Equality takes on the character of equality between formal categories, distancing sentencing from an individualized conception of equality. In Denmark (or Germany, Norway, or Sweden), the narrative guidelines and close ties between legal and appellate guidance of sentencing, institutionalized in the professionalization of judging and sentencing, maintain equality in sentencing in its individualized conception. It is a particular strength of the interpretivist approach that it is well adapted to grasp the reflexivity of transnational and local ideas, concepts, and meanings.

Globalization and Comparative Criminal Sentencing

Disregarding purpose or approach, it is today difficult to carry out comparative studies of any area of criminal justice without regard to the transnational or globalized field that the compared fields of practice are in. Globalization may be described as the way in which societies and communities over the last centuries are increasingly interconnected and in which processes of human and societal development are intertwined. Focusing on the processes that stem from this development, Savelsberg (2011) illustrates with three examples of globalization: well-documented conditions of late modernity, global-level scripts offering solutions to challenges across nations, and a new type of global law has emerged in vast areas of application, including penalty bringing with it both new doctrines and institutions (e.g., the international permanent and ad hoc criminal courts).

Globalization first of all affects a traditional approach to comparative criminal sentencing by questioning the nation state and the jurisdictional borders as key units of comparative analysis. Significant processes involve transnational developments, cross border activities, and entire fields of norms and institutions that have moved to a level of international problem-solving. Does that mean that we cannot compare nation states and their practice of criminal sentencing? No. For sure, an argument can be made that it is significantly problematic not to include a transnational dimension in a comparative study of criminal sentencing. However, globalization does not necessarily move the field of study away from the local level of governance and practice. Rather, globalization has become one of the many forces and institutions that matter on a local level, suggesting that we should rather accept and include in our studies the subtle processes between the local and the global. In that way, it modifies the object of study and it forces us to expand on the methods of our comparison. Fourcade and Savelsberg (2006) have developed well-suited themes to such a process perspective in sentencing: first, increasing the focus on agency which locally constructs global domains through local actions; second, recognizing the mutually constitutive relationship between global and local practice; third, emphasizing that the process of globalization itself is tied to transnational power flows with the United States as the power holder; and fourth, bearing in mind that every transnational and global convergence trend is locally manifested and accordingly is interpreted by local national institutions through the lenses of local historical and local political structures. Judicial courts and sentencing cannot be understood as one or the other but rather as dialectic between forces on the local and the global level.

Conclusion

Comparative approaches are important to policy and research about criminal sentencing. Globalization processes challenge the way that we conceptualize national borders and compare countries. But at the same time, globalization underlines the deep necessity of comparing, however, reflexively. This entry has touched upon important aspects of comparing criminal sentencing, laying bare significant possibilities and issues at stake with the different purposes of comparing, and with the different approaches to comparing sentencing. Accordingly, a legal categorical approach serves valuable purposes of organizing legal traditions but also suffers from the blindness to political, social, and cultural differences. An explanatory or predictive social scientific approach serves to advance our understanding of general questions of sentencing arrangements and provides valid descriptions of causes and effects of national sentencing and sanctioning practices. Nevertheless, it comes with serious concerns about its scientific assumptions, ranging from the units of analysis, its samples, as well as its construction of variables. Finally, an interpretivist approach suffers from the inevitable inability to generalize, yet provides an unparalleled means of understanding the depth and technologies of the relationships between courts, sentencing, and the local cultural context. As such, the interpretivist approach serves to challenge and extend the understanding derived from positivistic approaches. And as with most other things, it is the combination and not the choice of approach that provides the best result.

Related Entries

- ▶ Comparative Legal Cultures
- ▶ Comparative Political Economy of Punishment
- ▶ Comparing Police Systems Across the World
- ▶ Drug Courts
- ▶ Effective Community Supervision Programming
- ▶ Historical and Comparative Perspectives on Incarceration
- ▶ International Sentencing
- ▶ Managerial Court Culture
- ▶ Mitigation and Aggravation at Sentencing
- ▶ Multisite Trials in Criminal Justice Settings
- ▶ Penal Philosophy and Sentencing Theory
- ▶ Plea Bargaining
- ▶ Problem-Solving Courts
- ▶ Prosecutorial Discretion
- ▶ Sentencing as a Cultural Practice
- ▶ Sentencing Guidelines in the United States
- ▶ Sentencing Research
- ▶ Structured Sentencing Outside the United States
- ▶ Theories of Punishment

Recommended Reading and References

Aas KF (2005) Sentencing in the age of information. From Faust to Macintosh. Glasshouse, London

Aas KF (2007) Globalization and crime. Sage, London

Cavadino M, Dignan J (2007) Penal systems. A comparative approach. Sage, London

Damaska MR (1991) The faces of justice and state authority. A comparative approach to the legal process. Yale University Press, New Haven

Farrington D, Langan P, Tonry M (eds) (2004) Cross-national studies in crime and justice. U.S. department of justice. Office of justice programs. Bureau of justice statistics

Field S (2009) Fair trials and procedural traditions in Europe. Oxf J Leg Stud 29(2):365–387

Fionda J (1995) Public prosecutors and discretion: a comparative study. Clarendon series. Oxford University Press, Oxford

Fourcade M, Savelsberg J (2006) Introduction: global processes, national institutions, local bricolage: shaping law in an era of globalization. J Law Soc Inq 31(3):513–519

Frase R (2008) Sentencing and comparative law theory. In: Jackson J, Langer M, Tillers P (eds) Crime, procedure and evidence in a comparative and international context. Essays in honour of Professor Mirjam Damaska. Hart, Oxford, pp 65–79

Glenn P (2010) Legal traditions of the world. Sustainable diversity in law, 4th edn. Oxford University Press, Oxford

Hardie-Bick J, Sheptycki J, Wardak A (2005) Introduction: transnational and comparative criminology in a global perspective. In: Sheptycki J, Wardak A (eds) Transnational and comparative criminology. Glasshouse, London, pp 1–17

Jacob H, Blankenburg E, Kritzer H, Provine DM (1996) Courts, law, and politics in comparative perspective. Yale University Press, New Haven

Koichi H, Harris R, Hough M, Ville R, Zvekic U (1995) Probation round the world. A comparative study. Routledge, London

Kötz H, Zweigert K (1998) An introduction to comparative law, 3rd edn. Oxford University Press, Oxford

Lappi-Seppälä T (2011) Explaining imprisonment in Europe. Eur J Criminol 8:303–328

Nelken D (2007) Comparing criminal justice. In: Maguire M, Morgan R, Reiner R (eds) Oxford handbook of criminology, 3rd edn. Oxford University Press, Oxford, pp 139–157

Nelken D (2009) Contrasting criminal justice. Ashgate, Aldershot

Nelken D (2010) Comparing criminal justice. Making sense of difference. Sage, London

Nelken D (ed) (2011) Comparative criminal justice and globalization. Farnham, Surrey, England

Nolan JL (2009) Legal accents. Legal borrowing. The international problem-solving court movement. Princeton University Press, Princeton

Örücü E (2002) Law as transposition. Int Comp Law Q 51(2):205–223

Pakes (2010) Comparative criminal justice, 2nd edn. Willan, Cullompton

Pizzi WT (2008) Sentencing in the US: an inquisitorial soul in an adversarial body? In: Jackson J, Langer M, Tillers P (eds) Crime, procedure and evidence in a comparative and international context. Essays in honour of Professor Mirjam Damaska. Hart, Oxford, pp 65–79

Savelsberg J (2011) Globalization and states of punishment. In: Nelken D (ed) Comparative criminal justice and globalization. Farnham, Surrey, England, pp 69–86

Tata C, Field S (eds) (2010) Connecting legal and social justice in the neo-liberal world? The construction, interpretation, and use of pre-sentence reports. Punishment and society. Int J Penology 12(3)

Teubner G (1998) Legal irritants: good faith in British law or how unifying law ends up in new divergences. Mod Law Rev 61(1):11–32

Thaman SC (ed) (2010) World plea bargaining. Consensual procedures and the avoidance of the full criminal trial. Carolina Academic, Durham

Tonry M (ed) (2007) Crime and justice. A review of research. Volume 36. Crime, punishment, and politics in a comparative perspective. Chicago University Press, Chicago

Tonry M (2009) Explanations of American punishment policies. A national history. Punishment and society. Int J Penology 11:377–394

Tonry M, Frase R (eds) (2001) Sentencing and sanctions in western countries. Oxford University Press, New York

Tonry M, Hatlestad K (eds) (1997) Sentencing reform in overcrowded times: a comparative and cross-national perspectives. Oxford University Press, New York

Watson A (1993) Legal transplants, 2nd edn. University of Georgia Press, Athens, Georgia

Whitman JQ (2003) Harsh justice. Criminal punishment and widening divide between America and Europe. Oxford University Press, New York

Comparative Criminal Justice

► Comparative Political Economy of Punishment

Comparative Incarceration

Douglas B. Weiss[1] and Doris Layton MacKenzie[2]
[1]University of Maryland, College Park, MD, USA
[2]Justice Center for Research, The Pennsylvania State University, University Park, PA, USA

Synonyms

Comparative; Cross-national; Incarceration; Incarceration rates

Overview

One of the most distinctive features of the criminal justice system in the United States is the large number of individuals incarcerated in its prisons and jails. Since the early 1970s, the United States has experienced approximately 40 years of nearly uninterrupted growth in prison and jail populations. There were approximately 1.6 million people incarcerated in state and federal prisons, and approximately 750,000 individuals were held in county jails in 2010 (Minton 2011). Only in recent years has the growth in incarceration slowed.

Incarceration rates in the United States are several times greater in magnitude than those of other advanced Western democracies. A variety of explanations have been proposed for why incarceration is so much greater in the United States than in similarly developed nations. Some suggest that the United States has much greater levels of crime than in other Western democracies, while others argue that a "tough on crime" movement that has increased penalties is responsible for this growth. Research tends to indicate that the high rates of incarceration are mostly due to differences in sentencing policies and practices and not to exceptionally high levels of crime.

Further evidence that policy changes can have a significant impact on incarceration rates can be found by examining the experience of other countries. In the early 1950s, Finland had an incarceration rate several times greater than that of its Nordic neighbors. A profound shift in the thinking about the use of incarceration led to a series of changes to sentencing practices in Finland. As a result of these changes, Finland has gradually reduced its incarceration rate over the past 50 years and now has one of the lowest incarceration rates in all of Europe. Germany is another nation that has controlled prison population growth through policy changes. The experience of Germany and Finland in controlling growth in incarceration suggests that changes to current sentencing practices in the United States can be effective at reducing prison populations.

Incarceration Rates in the United States and other Western Democracies

The United States incarcerates more of its population than any other country in the world. Despite being home to just 5 % of the global population, the United States incarcerates about 25 % of the world's prisoners (Walmsley 2009). At year-end 2010, there were more than

1.6 million prisoners in US federal and state prisons for a rate of 497 prison inmates per 100,000 US residents (Guerino et al. 2011). As of midyear 2010, there were nearly 750,000 inmates held in local jails for a rate of 248 jail inmates per 100,000 US residents (Minton 2011). These rates are slightly down from their peaks in 2007 when prison and jail incarceration rates were 506 and 259, respectively. When prison and jail inmates are combined, the incarceration rate in the United States at year-end 2010 was approximately 730 inmates per 100,000 US residents (International Centre for Prison Studies nd).

The current level of incarceration in the United States is exceptional by both historical and comparative standards. Current levels of incarceration in the United States are the highest they have ever been in history. Throughout most of the twentieth century, incarceration rates in the United States remained relatively stable at a level of approximately 100 inmates per 100,000 US residents – a rate comparable to those of other Western democracies at the time. The stability of incarceration levels in the United States led Blumstein and Cohen (1973) to theorize that societies maintain stable levels of punishment. The theory did not gain much traction, however, as incarceration rates began nearly four decades of continuous growth almost simultaneous to the publication of the stability of punishment hypothesis.

Incarceration rates in the United States are also exceptional in comparison with other nations. The United States has the highest prison population rate in the world according to the International Centre for Prison Studies. In 2010, the nations that came closest to the level of incarceration in the United States are Rwanda, Georgia, and Russia. Drawing comparisons between these nations and the United States, however, would be misguided as these nations differ along a number of different dimensions such as levels of economic and social development. When one compares the United States to nations that are more similar in terms of economic and social development, however, the difference in incarceration rates is even greater. Incarceration rates of advanced Western democracies such as England, Germany, the Netherlands, Sweden, and Australia are considerably smaller.

While the United States has an exceptionally high incarceration rate relative to other advanced Western democracies, it is not the only nation that has experienced increases in incarceration rates in recent history (International Centre for Prison Studies nd). Incarceration rates in the nations of England and Wales, Scotland, and Australia increased by approximately 50 % between 1992 and 2010, while Netherland's incarceration rate has nearly doubled during that time. Only a few nations have managed to reduce or maintain relatively stable incarceration rates as Finland and Germany have done. Between 1992 and 2010, Finland's prison population rate has dropped by more than 10 %. Germany did experience an increase in incarceration since 1992; however, there has been a recent downward trend since 2004.

Explaining Cross-National Differences in Incarceration Rates

It is important to first acknowledge the difficulty inherent in making comparisons between nations in terms of incarceration use. Differences observed between nations in incarceration may be an artifact of different recording and reporting practices rather than a real substantive difference. One important difference between nations is the determination of who is included in the calculation of the incarcerated population. Some nations may include jail inmates in their calculation of their incarceration rates while other nations may exclude jail inmates. The determination of who should be counted as a prisoner can have a substantial impact depending on the size of the jail population.

With this caveat in mind, there is a relatively small body of literature that has examined and sought to explain cross-national differences in incarceration rates. This research has investigated the relationship between incarceration rates and a variety of social, political, and economic factors. Findings regarding which factors affect incarceration rates have been somewhat inconsistent due to methodological differences between studies. For instance, some studies draw comparisons between a large number of countries which differ from each other along a number of dimensions, while others focus on a smaller number of countries that are more comparable in terms of economic and social development. It is important to keep this in mind as making comparisons between vastly different nations may lead one to draw erroneous conclusions.

The factor most often thought to affect incarceration rates are crime rates. Homicide rates in the United States are about as twice as high as those of other advanced Western democracies, such as Finland, which has one of the highest homicide rates in Western Europe (United Nations Office on Drugs and Crime 2011). However, homicide accounts for a very small proportion of total crime and should not have a very large impact on incarceration rates. Other than homicide, crime and victimization rates in the United States are not significantly different from those of other advanced Western democracies (Farrington et al. 2004; Van Dijk et al. 2007).

Few studies have found support for the notion that differences in crime rates can account for disparities in incarceration rates (Lynch 1988; Neapolitan 2001). Neapolitan (2001) found a significant relationship between homicide rates and imprisonment rates in his study of 148 nations. Lynch (1988) also found differences in incarceration rates between the United States, England, Canada, and West Germany could be explained in large part by differences in the levels and types of crime in these nations.

Most research has found, however, that crime rates are not directly associated with incarceration rates (Young and Brown 1993; Lappi-Seppala 2007). Incarceration rates and crime rates in the United States have moved independently since the early 1970s. Incarceration rates have been increasing in the United States since the early 1970s regardless of changes in crime rates. Despite similarities in crime trends since 1960, Canada has maintained a relatively stable incarceration rate while the United States experienced its incarceration boom (Doob and Webster 2006).

It is more probable that crime rates indirectly affect incarceration rates by affecting attitudes toward punishment among politicians and the general public. Rising crime rates may contribute to increased punitive attitudes and increasingly severe criminal penalties that would result in increased use of incarceration. This may explain the divergent trends in incarceration between the United States and Canada. Doob and Webster (2006) attribute Canada's stable incarceration trend since the 1960s to the lack of changes in sentencing practices despite increasing crime rates. In contrast, the indeterminate sentencing scheme used in the United States came under fire from a number of different fronts in the 1970s. The Martinson (1972) report, which claimed nothing works in rehabilitation, and Judge Marvin Frankel's (1973) criticism of indeterminate sentencing contributed to profound changes in sentencing practices in the United States including the adoption of determinate sentencing, mandatory minimums, habitual offender laws, and the reduction or elimination of parole (MacKenzie 2006).

There is good evidence that suggests these changes to sentencing and correctional policies were largely responsible for incarceration growth in the United States. Blumstein and Beck (1999) attribute 88 % of the increase in imprisonment in the United States between 1980 and 1996 to changes in sentencing practices. During this time, there was an increase in both the number of offenses for which one may be imprisoned and in the length of prison time served.

The type of legal system used in a country is another factor that can affect incarceration rates. Ruddell (2005) and DeMichele (2010) found the common-law legal system used in the United States and other Anglo nations to be associated with significantly higher incarceration rates than nations that operate under a system of Romano-Germanic or Nordic law. Incarceration rates tend to be low under Nordic law regimes (e.g., Finland) and moderate under Romano-Germanic law legal regimes (e.g., Germany). DeMichele (2010) attributes the difference in incarceration rates between legal regimes to several factors including the balance of power among the courtroom workgroup, differences in courtroom practices (e.g., trial structure), and the involvement of lay participants in the judicial process.

Certain aspects of the common-law legal regime, such as the jury trial, increase the involvement of lay participants in making decisions regarding punishment. Heinous crimes that receive heavy media coverage may result in public demands for increasing the severity of punishment. Jacobs and Kleban (2003) found higher incarceration rates occur in nations where the general public has the greatest influence over corrections decisions. The lowest incarceration rates were found in nations where decisions about corrections were made by unelected experts, e.g., judges.

Another political factor that may affect incarceration rates is the level of conservative political strength. Studies done in the United States tend to find a relationship between conservative political strength and incarceration (Sutton 2000; Jacobs and Helms 1996). However, comparative cross-national research has not found this relationship to be significant (Jacobs and Kleban 2003). Jacobs and Kleban (2003) speculate that the failure to find this relationship in other advanced democracies may reflect the fact that decisions about corrections are insulated from popular opinion.

Economic factors provide another potential explanation for differences in national incarceration rates. However, research has failed to find a clear association between economic factors and incarceration. Ruddell (2005) found no relationship between incarceration rates and income inequality, unemployment, or inflation in a sample composed of the richest 100 nations as determined by GDP. Jacobs and Kleban (2003) also fail to find a relationship between incarceration and unemployment. However, they find that the most economically developed nations with the highest per capita GDP are more likely to have higher proportions of their population in prison. Incarceration rates in these nations may simply be higher because they have more resources to build and stock prisons.

Incarceration rates may also be affected by structural factors of the economic system. Sutton (2004) found a relationship between incarceration rates and unemployment, inflation, and income inequality in a study of 15 affluent capitalist democracies. However, the relationship between incarceration rates and these factors disappeared when measures of institutional structure, such as union and left party strength, are considered. Other factors that may affect incarceration rates include out-of-wedlock birthrates (Jacobs and Kleban 2003) and the degree of population heterogeneity in a nation (Jacobs and Kleban 2003; Ruddell 2005).

In sum, there are a variety of social, cultural, and economic factors that account for the disparity in incarceration rates between the United States and other advanced Western democracies. There is mixed evidence that suggests incarceration rates are related to crime rates or economic factors such as unemployment or inequality. Instead, there is a good body of research that suggests incarceration rates are affected by changes in criminal justice policies and the type of legal system used in a nation. Further evidence that changes to criminal justice policies can have a significant impact on incarceration rates will be presented in the next section.

Controlling Prison Population Growth

It is clear that changes to criminal justice policies and practices can have a significant impact on incarceration rates. The operation of the criminal justice system in the United States underwent profound changes in the early 1970s that contributed to incarceration growth. These changes include the switch to a determinate sentencing system, the adoption of mandatory minimums and habitual offender laws, and truth-in-sentencing legislation. Combined, these policies have increased the number of offenses for which one could be imprisoned and increased the amount of time offenders serve in prison.

If these policy changes were responsible for the growth in incarceration, then it is important to ask whether policy changes could reverse the past 40 years of incarceration growth in the United States. The answer to this question may be found by examining the experience of other nations that have dealt with extraordinarily high incarceration rates. In particular, Finland's experience with incarceration could serve as a guide for how incarcerated populations in the United States might be reduced.

In the early 1950s, the incarceration rate in Finland was around 200 prisoners per 100,000 inhabitants. While this rate is a fraction of the current imprisonment rate in the United States, at the time it was nearly twice the rate in the United States and was three to four times greater than other Nordic countries. Lappi-Seppala (2001) identified three possible factors that explained Finland's exceptionally high incarceration rate at that time. These factors include (1) a unique "cultural climate in which severity was not measured on the same scale used in the other Nordic countries" (106), (2) a rigid penal system characterized by high minimum penalties, and (3) severe sentencing practices for relatively common crimes such as drinking and driving.

These three factors also seem characteristic of criminal justice practices in the United States today. Criminal sentences in the United States tend to be more severe than those of other advanced Western democracies, and the greatest disparities in punishment exist among less serious property crimes. In a comparison of punitiveness among eight advanced Western democracies, Blumstein, Tonry, and Van Ness (2005) find that offenders convicted in the United States between 1980 and 1999 served considerably longer periods of time in prison than did offenders who were convicted in other countries. Further, they find the disparity in time served between the United States and other nations is greatest for less serious crimes such as motor vehicle theft. Lynch (1993) had previously found that average sentence lengths in the United States in the 1980s were considerably longer than those in Australia, Canada, England and Wales, and West Germany. However, the amount of time served was relatively similar for violent crimes in these five nations. Prisoners serving sentences for property crimes in the United States spent more time in custody than those in Australia, Canada, and England and Wales.

Changes in these three factors were responsible for the gradual decarceration that took place in Finland since the mid-1950s (Lappi-Seppala 2001). At the time, there was a profound shift in the ideological thinking of criminal justice policy which had till then emphasized (coercive) rehabilitation of offenders. Unlike the movement against rehabilitation in the United States where the punishment philosophy was replaced by the goals of incapacitation and deterrence, the backlash against rehabilitation in Finland was replaced by a philosophy of restitution that emphasized (1) the minimization of costs of crime and crime control to society and (2) creating a fair distribution of costs among the offender, society, and the victim. Rather than serving as the primary instrument of criminal punishment, incarceration was seen to be one possible option among others. This change in philosophy was accompanied by the recognition that the most effective crime prevention may not occur through fear of punishment (deterrence) but rather through increasing the legitimacy of the law and subsequently enhancing the internalization of values and morality.

This change in philosophy and the inability to justify the exceptionally high rates of incarceration in Finland relative to other Nordic countries prompted a series of legislative and policy reforms that would ultimately reduce incarceration levels in Finland. These reforms included reduced penalties for several common crimes, the expanded use of noncustodial alternatives, and the development of sentencing alternatives specifically designed to reduce the number of offenders sentenced to imprisonment. A complete discussion of these reforms can be found in Lappi-Seppala (2001) and Tornudd (1993).

Reduction of penalties for theft and drunk-driving offenses were two reforms that had a significant impact in reducing the number of offenders in Finnish prisons (Lappi-Seppala 2001). These reforms were designed to both (1) reduce the proportion of convicted offenders who would receive a custodial sentence and (2) reduce the sentence lengths of those who were incarcerated. Through a legislative reform enacted in 1972 that redefined serious forms of theft and introduced a new penalty range, the percentage of convicted offenders who received a custodial sentence for theft dropped from 38 % in 1971 to 11 % in 1991. In addition, the median sentence for offenders convicted of larceny dropped from 7.4 months in 1971 to 2.6 months in 1991. Reforms pertaining to drunk-driving offenses focused on replacing custodial sentences with noncustodial alternatives such as community service. Similar to the success of the reforms directed at theft offenses, drunk-driving-related reforms reduced the percentage of convicted offenders sentenced to unconditional imprisonment from 70 % in 1971 to 12 % in 1981.

The expanded use of noncustodial alternatives was another major reform that contributed to reduced incarceration rates in Finland. Specifically, Finland has relied heavily on the use of community service as an alternative for short terms of imprisonment Lappi-Seppala 2004. This alternative was effective at reducing incarceration because it was specifically designed to ensure that the only offenders who would receive it would be those destined for incarceration. While sentencing alternatives have been used in the United States, they have not been very effective at reducing the use of incarceration as a punishment. Instead of diverting convicted offenders from custodial sentences, the use of sentencing alternatives in the United States has contributed to a "net-widening" effect since those offenders receiving sentencing alternatives would not be incarcerated otherwise.

Finland adopted a two-step procedure during the sentencing phase to ensure that only those who were going to initially receive a custodial sentence were eligible for the community service alternative (Lappi-Seppala 2004). In the first step, the court determines a sentence for the offender using the normal criteria for sentencing. If the offender is sentenced to unconditional imprisonment, then the court can decide whether to commute the sentence into community service provided a number of criteria have been met. These criteria include (1) the consent of the convicted person, (2) the determination that the convicted offender is able to carry out the requirements, and (3) that there is nothing in the prior criminal history of the convicted person that would preclude the use of community service (e.g., violent history). After several years of being implemented on an experimental basis, community service orders started being used nationwide in 1994. An examination of trends in Finnish court sentencing practices suggests that community service orders, used in this way, have been successful in reducing the flow of offenders into prison. Since 1994, approximately 30 % of convicted offenders sentenced to an unconditional prison sentence have received a community service order instead.

Even as Finland took these steps to reduce its incarceration rate, crime trends in Finland were remarkably similar to those of its Nordic neighbors (Falck et al. 2003). Crime rates in Finland, Norway, Sweden, and Denmark increased from the 1960s through the early 1990s when they reached a more stable trend. While the increase in crime in Finland was unremarkable when compared to its neighbors during this time, the decline in incarceration rates during this same period was exceptional. In 1950, the incarceration rate in Finland was 187 inmates per 100,000 population. By the late 1990s, the incarceration rate in Finland dropped below 60 – a level comparable to those of its neighbors. Incarceration rates in Finland have since increased; however, all of the Nordic nations have experienced similar increases (National Research Institute of Legal Policy 2011).

Germany is another nation that has taken measures to control the growth of incarceration at a time when crime rates were increasing. Although Germany has not experienced as high of an incarceration rate as has Finland, it has faced problems related to the heavy use of incarceration such as severe overcrowding (Graham 1990). Heinz (2006) traces the history of sentencing practices in Germany as far back as 1882, where nearly 80 % of convicted offenders were sentenced to unconditional prison sentences. Since then, Germany has gradually rolled back its use of custodial sentences in favor of noncustodial alternatives such as fines. Since the mid-1980s, less than 10 % of convicted offenders in

German courts have received unconditional prison sentences, while approximately 70 % of offenders were ordered to pay fines.

While the use of imprisonment has declined throughout the twentieth century, modern sentencing practices in Germany can be traced back to legislation enacted in 1969 (Weigend 2001). The criminal law reforms that took effect in 1969 were explicitly designed to restrict the use of imprisonment to only the most serious cases. This aim was to be achieved through several measures including (1) the abolishment of prison sentences of less than 1 month; (2) the decriminalization of many petty offenses, such as public order and traffic offenses, which were subsequently turned into administrative infractions punishable by fines only; and (3) the discouragement of sentences of less than 6 months. Noncustodial alternatives, primarily fines and probation, were used in lieu of short terms of imprisonment.

The criminal law reforms of 1969 also began a trend in the development of sanctions other than those that have been traditionally used, i.e., imprisonment, probation, and fines (Weigend 2001). Some of these alternative sanctions have not been used very often, such as warning with suspension, while others are being increasingly used by the German courts, such as community service and restitution programs. Not all of these additional sanctions are directed at reducing the number of convicted offenders sentenced to imprisonment. For example, the sanction "warning with suspension of punishment" allows courts to suspend punishments consisting of dayfines and sentence the offender to probation instead. On the other hand, legislation enacted in 1994 allows the courts to mitigate the penalty or suspend sentences of less than 1 year of imprisonment if the convicted offender either participates in a victim-offender reconciliation program or makes some effort to provide financial restitution to the victim.

Finland and Germany's efforts to control and reduce prison populations may serve as an example for the United States to reduce its incarcerated population. While these two nations differ from the United States along a variety of dimensions, they both demonstrate that policy changes can be effective at controlling and even reducing prison populations. Certain features of the United States, such as its common-law legal system, may make it more difficult to achieve the level of success that Finland has had in its history. However, it is unlikely that incarceration rates in the United States will drop to levels of other advanced Western democracies absent any changes to sentencing policies and practices.

Conclusion

The incarceration rate in the United States far surpasses that of any other advanced Western democracy. This disparity in incarceration can be attributed to differences along a variety of factors including sentencing policies and the common-law legal system used in the United States. While some factors may be difficult to control because they are foundational, e.g., common-law legal system, there is evidence that changing sentencing and corrections policies can have a considerable impact on incarceration rates.

Recommended Reading and References

Blumstein A, Beck AJ (1999) Population growth in U.S. prisons, 1980–1996. Crime & Justice 26:17–61.

Blumstein A, Cohen J (1973) A theory of the stability of punishment. J Crim Law Criminol 64(2):198–207

Blumstein A, Tonry M, Van Ness A (2005) Cross-national measures of punitiveness. In: Tonry M, Farrington DP (eds) Crime and punishment in Western countries. University of Chicago Press, Chicago, pp 1980–1999

DeMichele M (2010) Three worlds of western punishment: a regime theory of cross-national incarceration rate variation, 1960–2002. Doctoral dissertation

Doob AN, Webster CM (2006) Countering punitiveness: understanding stability in Canada's imprisonment rate. Law Soc Rev 40(2):325–368

Falck S, von Hofer H, Storgaard A (2003) Nordic criminal statistics, 1950–2000. Department of Criminology, Stockholm University, Stockholm

Farrington DP, Langan PA, Tonry M (2004) Cross-national studies in crime and justice. Bureau of Justice Statistics, Washington, DC

Graham J (1990) Decarceration in the Federal Republic of Germany: how practitioners are succeeding where policy-makers have failed. Br J Criminol 30(2):150–170

Guerino P, Harrison PM, Sabol WJ (2011) Prisoners in 2010. Bureau of Justice Statistics, Washington, DC

Heinz W (2006) Penal sanctions and sanctioning practice in the Federal Republic of Germany 1882–2004. The Konstanz repositories on crime and sanctioning. Retrieved 30 Oct 2009, from http://www.uni-konstanz.de/rtf/kis/sanks04_eng.htm

International Centre for Prison Studies (nd) World prison brief. Retrieved 10 Feb 2012, from http://www.prisonstudies.org/info/worldbrief/

Jacobs D, Helms R (1996) Towards a political model of incarceration: a time-series examination of multiple explanations for prison admission rates. Am J Sociol 102:323–357

Jacobs D, Kleban R (2003) Political institutions, minorities, and punishment: a pooled cross-national analysis of imprisonment rates. Soc Forces 82(2):725–755

Lappi-Seppala T (2001) Sentencing and punishment in Finland: the decline of the repressive ideal. In: Tonry M, Frase RS (eds) Sentencing and sanctions in western countries. Oxford University Press, Oxford

Lappi-Seppala T (2004) Penal policy and prison rates: Long-term experiences from Finland. The National Research Institute of Legal Policy, Helsinki.

Lappi-Seppala T (2007) Trust, welfare, and political economy: cross-comparative perspectives in penal severity. In: Traskman PO (ed) Rationality and emotion in European penal policy: Nordic perspectives. DJOF, Copenhagen

Lynch JP (1988) A comparison of prison use in England, Canada, West Germany, and the United States: a limited test of the punitive hypothesis. J Crim Law Criminol 79(1):180–217

Lynch JP (1993) A cross-national comparison of the length of custodial sentences for serious crimes. Justice Q 10(4):639–660

MacKenzie DL (2006) What works in corrections: reducing the criminal activities of offenders and delinquents. Cambridge University Press, Cambridge

MacKenzie DL, Weiss DB (2009) Other countries have successfully reduced incarceration rates without increasing crime: we can do it! Vict Offenders 4(4):1–7

Minton TD (2011) Jail inmates at midyear 2010 – statistical tables. Bureau of Justice Statistics, Washington, DC

National Research Institute of Legal Policy (2011) Crime and criminal justice in Finland 2010. National Research Institute of Legal Policy, Helsinki

Neapolitan JL (2001) An examination of cross-national variation in punitiveness. Int J Offender Ther Comp Criminol 45(6):691–710

Ruddell R (2005) Social disruption, state priorities, and minority threat: a cross-national study of imprisonment. Punishm Soc 7(1):7–28

Sutton JR (2000) Imprisonment and social classification in five Anglo-American democracies. Am J Sociol 106:350–386

Sutton JR (2004) The political economy of imprisonment in affluent western democracies, 1960–1990. Am Sociol Rev 69(2):170–189

Tornudd P (1993) Fifteen years of decreasing prisoner rates. National Research Institute of Legal Policy, Helsinki

United Nations Office on Drugs and Crime (2011) 2011 Global study on homicide: trends, contexts, data. United Nations Office on Drugs and Crime, Vienna

Van Dijk JJM, van Kesteren JN, Smit P (2007) Criminal victimization in international perspective: key findings from the 2004–2005 ICVS and EU ICS. Boom Legal Publishers, The Hague

Walmsley R (2009) World prison population list. International Centre for Prison Studies, London

Weigend T (2001) Sentencing and punishment in Germany. In: Tonry M, Frase RS (eds) Sentencing and sanctions in western countries. Oxford University Press, Oxford

Weiss DB, MacKenzie DL (2010) A global perspective on incarceration: how an international focus can help the United States reconsider its incarceration rates. Vict Offenders 5(3):268–282

Young W, Brown M (1993) Cross-national comparisons of imprisonment. Crime Justice 17(1):1–49

Comparative Legal Cultures

David Nelken
Cardiff University Law School, Cardiff, UK

Overview

This entry offers an overview of the main problems in defining and using the concept of legal culture, with special attention to how this may be relevant for scholars of criminal justice.

Scholars of criminal justice may be interested in comparing legal cultures to make sense of similarities and differences in crime levels or punitiveness (Nelken 2000, 2010). Alternatively, they may want to understand how trends in criminal justice spread by means of legal transplants or efforts at harmonization or collaboration as well as how they are reinterpreted or resisted. This entry will first seek to clarify the meaning of the term legal culture and discuss how it may be used in doing research. It will then consider more specifically three of the key issues that arise in carrying out such investigations concerned with generating substantive findings about patterns of law or criminal justice in different jurisdictions. These concern units and boundaries, elements and aggregates, and coherence and change.

What Is Legal Culture?

Studies of law in relation to culture can cover a large range of topics including the role of culture in law (and the challenges of legal pluralism and multiculturalism) to the part played by law as culture, seen as a way of making meaning. Those interested in the relationship between law and culture might wish to study law as a cultural artifact, examine the way it becomes present in everyday life and experience, or through the media, or consider the role of law in accommodating cultural defenses or protecting cultural treasures. This entry, however, is mainly interested in seeing how the concept can be of use in research into cultural variation in how law is experienced in social life and, especially, differences in the role law plays and expected to play. There are obvious implications here for students of criminal justice seeking to tease out the relationship between formal and informal social controls.

Research into comparative legal cultures can involve investigations into the extent to which law is party or state directed (bottom-up or top-down), the role and importance of the judiciary, or the nature of legal education and legal training. It may concern ideas of what “law” means (and what law is “for”). It is possible to discern legal culture in different approaches to regulation, administration, and dispute resolution. There may be important contrasts in the degree to which given controversies are subject to law, the role of other expertises, and the part played by “alternatives” to law, including not only arbitration and mediation but also the many “infrastructural” ways of discouraging or resolving disputes. With reference to criminal justice, the use of legal, non-state, or “informal” forms of social control deserves attention.

But legal culture remains a highly debated term. Even Lawrence Friedman, who pioneered its use in social science, describes it as “an abstraction and a slippery one” and that if he were to start over he might not use it again (Friedman 2006). There are several kinds of disagreements about what legal culture means or should mean (Nelken 1995, 2006, 2007, 2012). There is, for example, the question whether or not legal culture is a useful term as compared to alternative concepts such as legal tradition or legal ideology. Then there is discussion about what the term actually refers to – for example, to attitudes and/or behavior, or whether legal culture refers more appropriately at the popular or at the institutional level.

Confusion can easily arise where what appears to be a debate about the correct interpretations of a given legal culture in fact trespasses into these other areas. Blankenburg, for example, takes Friedman to be arguing that it is “folk” culture – popular pressure for legal change – that shapes differences in legal development and response. He offers, by contrast, a comparison of the very different use of courts in Germany and the Netherlands, countries that he argues, have similar folk cultures and very different institutional outputs. For him, infrastructural arrangements represent the key to differences in legal culture, and in his view, “there is no legal culture outside of institutions” (Blankenburg 1997). Yet, what Blankenburg wants to call legal culture others would call institutions or structures, as seen in the related debate over whether the low level of use of law in Japan is to be attributed to deliberate cultural avoidance of litigation or is rather a result of structural arrangements that block access to the courts (Nelken 1997).

It is also not always easy to tell where definitional questions end and empirical enquiries begin. How are we to distinguish legal culture from political, economic, or religious culture? Or take Friedman’s important distinction between “internal legal culture” which refers to the role in the law of legal professionals and “external legal culture” which refers to those individuals or groups who bring pressure to bear on the law to produce social change. Do lawyers belong to “internal legal culture” (as servants of the courts) or to “external legal culture” as agents of social groups and of individual litigants.

What about the relationship between external legal culture and culture in general? What of the way judges incorporate lay definitions of appropriate behavior into their activities?

It can also be important to distinguish between legal culture and “the culture of legality” (stressing the “legal” before the word culture). The latter term – which corresponds very roughly to what in English is called “the rule of law” – is particularly common in those jurisdictions or parts of jurisdictions, for example, in the former Soviet Union, Latin America, or the south of Italy, where state rules are often avoided or evaded. These are places where – from a state perspective – there is a culture of illegality, and the point of talking of “legal culture” in such cases is to point to the normative goal of getting “legality” into the culture of everyday social and political life. This move seeks to reorient the behavior of such populations toward (state) law and/or encourage state law to respect certain limits of action. Whereas legal culture is a descriptive/explanatory term, the “culture of legality” is a normative and evaluative one. Keeping descriptive and normative meanings apart is necessary if we are to examine the sorts of legal culture that are more or less conducive to creating “the culture of legality”.

But it can also be argued that talk about culture or legal culture is rarely merely descriptive and is more often an interested intervention in debates. For the observer, cultural ideas are contested and connected to relations of power. Cultural repertoires include polarities, such as values and practices, ideas and habits, and innovations along with commonsensical ways of doing things. Culture is the product of historical influences rather than evolutionary change. It is marked by hybridity and creolization rather than uniformity or consistency (Merry 2012). At the same time, however, for social actors themselves, “law” and “culture” are words whose interpretation and definition have illocutionary effects (“this is the law” “that behavior is inconsistent with our culture”). Thus, the term “legal culture” may itself be used by judges, politicians, or others, in the course of making claims about what is or is not consonant with a given body of law, practices, or ideals. This use, as much prescriptive as descriptive or prescriptive through being descriptive, can “make” the facts it purports to describe or explain. Some scholars think the term must capture what these legal actors are trying to do (Webber 2004).

As this suggests, culture is a term that easily lends itself to misuse, providing an easy alibi for resistance to change. For Patrick Glenn, the idea of culture is suspect (as a replacement for talk of race) because of both its origins and its consequences. Usage also includes “essentialist,” overdetermined, overbounded, and xenophobic applications. With cultural analyses, scholars are tempted to orientalize behavior as foreign and irrational and ignore or downplay the importance of economic and related political drivers of behavior (Glenn 2003, 2004). The German word *kultur* indeed emerged as a defensive term used in romantic opposition to the French universalizing idea of civilization (for which today’s discourses of democracy and human rights could be considered equivalents). Yet, there is nothing to stop the scholar from describing (and commenting on) such uses of the concept. And Friedman’s use of the word culture is in fact much closer to the nonessentializing French idea. He stresses convergence and the role of modernity as we move to a global legal culture based round individualism, equality, and human rights (Friedman 2006).

Given that the goal of comparative work is to take us beyond ethnocentrism, we also need to ask whether there can be a universal concept of legal culture. How far is any given conception of legal culture inevitably itself a “folk concept” rather than an “observer’s” concept? For example, the Western idea of law as linked to the nation-state fits uneasily into places with overlapping plural legal systems. Instead of accepting Friedman’s own view of internal and external legal culture as neutral analytical categories, David Engel instead offers a cultural reading of them (Engel 2012). “Two aspects of Friedman’s imagery,” he tells us, “are particularly noteworthy. First, law is spatialized. It has an inside and an outside, and legal culture forms a kind of membrane between the two spaces.

Second, the internal space is dead, desiccated, and inert. It is a place of bones without flesh and words without life. By contrast, the external space is alive, vital, and active. Outside the law, one finds life forces that act like water in the desert and bring to life the otherwise barren world of law." Referring by contrast to the approach taken by the noted anthropologist Clifford Geertz (Geertz 1973), Engel points to Friedman's circumscribed view of law as a dusty, entombed skeleton versus Geertz's broad and unbounded view of classical Indian law as sun and cattle! "Friedman's colorful imagery of law and legal culture," he argues, "presupposes a pluralism of legal and normative spaces, whereas Geertz's description of Indian law presupposes the opposite – a unity of law, government, culture, and belief" (Engel 2012).

For Engel, this concern with law having an "inside" and an "outside" is itself a reflection of the "project" of modernity, aimed inter alia at securing liberal legalism and secularism. Friedman's conception of legal culture is linked to the aspiration to keep politics and economics subservient to law as well as the value of civil society playing its proper role as the source of the legal. Maintaining the metaphorical separation of inside and outside, and thereby affirming the "relative autonomy" of law, is central to the project of modernity. "The theory of modern law "presupposes that no one group in the society has a privileged access to religion and moral truth," and the proper role of law is to establish a secular and neutral "process for conflict resolution" rather than to endorse one set of cultural practices or religious beliefs over another." On the other hand, Friedman's distinction may not carry all the "liberal legalist" implications that Engel attributes to it. It has been Friedman's abiding concern to deny law's autonomy so to show that it is what lies outside the law that actually animates and shapes it.

Working with the Idea of Legal Culture

For this concept to be of value heuristically, it is essential to do more than treat legal culture as a descriptive synonym for the legal system or some part of it. If our goal is explanation, it can be helpful to start from what appear to be puzzling features of the role and the rule of law within a given society (Nelken 2004). Why do the UK and Denmark complain most about the imposition of EU law but then turn out to be the countries which have the best records of obedience? Why does Italy, whose public opinion is most in favor of Europe, have such a high rate of noncompliance? Why does Holland, otherwise so similar, have such a low litigation rate compared to neighboring Germany? Why in the United States and the UK does it often take a sex scandal to generate official interest in doing something about corruption, whereas in Latin countries why does it take a major corruption scandal to excite interest in marital unfaithfulness? Such contrasts can lead us to discover new insights – into the importance of "enforcement" as an aspect of law, the reasons people use courts, or the way shame and guilt cultures condition the role of law.

How we then put the term legal culture to use in such enquiries will depend on the disciplinary framework developed. To decide the appropriate strategies and methods to adopt, eyes must focus on disputes between (as well as within) continually evolving disciplines such as sociology, political science, or anthropology. For example, in line with competing approaches to social theory, legal culture will manifest itself through institutional behavior, as a factor shaping and shaped by divergences in individual legal consciousness, as a pattern of ideas that lie behind behavior, or as another name for politico-legal discourse itself. Studies of legal culture can also have different aims. Trying to make comparisons of legal systems more sociologically meaningful requires avoiding the risks of reifying culture as determinant and constraining of individual choices. A concern, by contrast, with illustrating the ongoing process of meaning – making in and through law overlaps with an interest in studying legal consciousness (Silbey and Ewick 1998, Silbey 2001).

For many authors, legal culture is measurable by asking people questions about their attitudes to the law. But we may also need to make sense of patterns of behavior of which the participants themselves are unaware. Even well-informed people living in India mistakenly think that the reason courts are slow is because the country has such a (relatively) high rate of litigation (Galanter and Krishnan 2003). Americans are convinced that their tort system regularly produces excessive and undeserved awards. But it turns out that, in large part, this impression is manufactured by the media (Holmes and McCaan 2004). Those societies where legal professionals show least concern for what Anglo-American writers describe as the "gap" or gulf between the "law in books" and "law in action" may not be those where the gap is least problematic but those where the gap is in fact overwhelming.

A problem that in regularly leveled against explanations that rely on the idea of culture is that of tautology. Why do Japanese do what they do – because they are Japanese! It adds little to say that Italians suffer extensive court delays because of their legal culture if legal delays are part of what we are referring to when we speak of Italian legal culture. The danger of tautology needs to be overcome if we hope to develop policy-relevant social science explanations showing how variables produce outcomes. If the concept of legal culture accounts for why a transplant takes or not, can the success of transplants provide clues to legal culture? (Kurchiyan 2009). Interpretative approaches, however, may be less embarrassed by using culture as "cause" and consequence. If social actors take "culture" to represent or require certain behavior or values, this may provide a reason for their actions. In the 1980's, the appearance of league tables of relative levels of incarceration induced Finland (interested in being seen as similar to other Scandinavian countries) to move toward the average European imprisonment rate by reducing its prison population; Holland for its part felt enabled to do the opposite. Likewise, how far East European legal cultures come to resemble Western legal cultures, assuming this is desirable, in part depends on how far politicians, policy makers, and legal actors there believe they can escape the patterns inherited from the past.

Another way of avoiding tautology is to see legal culture as something that needs to be explained rather than as itself the explanation. Take again the question why Italy suffers such long court delays. Treated as part of legal culture, the issue is to discover the legal, economic, political, religious, and factors that help reproduce this feature of Italian legal life. In the first place, there are relevant laws, especially those that have to do with civil and criminal procedure. There is also the management and organization (or lack of organization) of the courts and legal profession and claims about the supply of law not keeping pace with the demand, economic interests, political priorities, etc. (Nelken 2004). The problem of course is that it can then prove surprisingly difficult to decide what factors are the crucial ones, especially as the relevant facts can often be elusive and are open to very different interpretations. Large companies can make use of judges as (paid) arbitrators outside the normal trial system. But small businessmen (the backbone of economic life in Italy) would seem to gain little from the current situation. If so, why do they not put more pressure on the politicians to do something? Lawyers and banks profit from this situation. But do they have veto power on reform? Hence, rather than the idea of culture providing pat solutions to our question, it leads (as perhaps is appropriate) to an infinite regress of causal puzzles regarding what shapes it.

Controversies: Three Issues in Comparing Legal Cultures

Research on comparative legal cultures necessarily mobilizes some assumptions about what it is that holds a culture together. But it may also help to revise these. We shall consider here questions connected to units and boundaries, elements and aggregates, and coherence and change.

Units and Boundaries

What is the unit we want to compare? In particular, is it (still) safe to identify the boundaries of legal culture with those of national jurisdictions? The search to understand and explain legal culture at the level of the nation-state certainly continues to be an important ambition of comparative law and comparative sociology of law, as seen in titles such as the "Japanese approach to law," "Dutch legal culture," and "French criminal justice". But the student of legal culture will often want to focus at levels below and above that of the nation-state. On the one hand, the focus could be on the culture of the local courthouse, the working norms of different social and interest groups and professional associations, and the networks of individuals involved in pursuing, avoiding, or mediating disputes. On the other, we could study international institutions and regulators and the so-called third cultures of international trade, communication networks, and other transnational processes.

It can also be rash to assume any necessary "fit" between law and its environing national society or culture. Legal systems always have been modified by a variety of processes of borrowing, imitation, and imposition. Nation-states have never been the exclusive or even predominant source of norms. Their increasing insertion in larger bilateral or multilateral structures and networks means that there is now a wide gap between the (global) sites where issues arise and the places where they are managed (the nation-state). Different kinds of units emerge as objects and as agents of control. Instead of governments, the talk now is increasingly of "governance," of how power is exercised at a series of other levels and by other institutions, in collaboration or otherwise with state bodies. The "denationalization" of rule making means that transnational public and semipublic networks substitute, to an increasing extent, for national governments. Rule formulation and settlement increasingly take place within new agencies of transnational governance, such as NAFTA, the OECD, and the WTO, but also in many lesser-known public-private forums.

Naturally, this process varies by different areas of legal and social regulation. A contrast is often made between, on the one hand, those areas of law that are relatively internationalized, such as international business contracts, antitrust and competition policy, Internet and new technology, labor law, social law, and environment law, and, on the other hand, family law and property law. But experts in these latter fields frequently report evidence of international trends and cross-cultural influences even there.

As the world increasingly is tied together by trade and communication, people have the sense of living in an interdependent global system marked by borrowing and lending across porous cultural boundaries that are saturated with inequality, power, and domination. All this means that the purported uniformity, coherence, or stability of given national cultures will often be no more than an ideological projection or rhetorical device used by some of those within or outside a given society or other context (Coombe 1998). So we need to avoid reifying national or other stereotypes and recognize that these may often be no more (or less) than "imagined communities" or "invented traditions". We may also need to be cautious about using terms that suggest boundaries at a time when many argue that it would be more appropriate to speak of "flows" (Appudurai 1995).

But claims about the decline of the nation-state should not be taken too far. Many new states have been created recently, and "failing states" are more often the cause of concern rather than seen as harbingers of a new order. Differences between legal cultures may mobilize or reflect wider social and cultural patterns that roughly coincide with national political boundaries. Such boundaries often coincide with language and cultural differences and represent the source of common statistics. The imposition of a common legal code and the common training of legal officials form part of attempts to achieve and consolidate national identity. And "borders" continue to play important instrumental and symbolic roles, not least in responding to immigration. There is even an empirical basis for psychological differences in national traits in the way people relate to each other (Hofstede 1980). Such different historically conditioned sensibilities may persist over quite long periods (though careful research is needed to avoid confusing short-term and long-term trends).

Even if we accept that the state is – in many respects – losing its centrality, it does not follow that what happens in the field of criminal justice necessarily follows this general logic. Criminal law continues to be a powerful icon of sovereignty, and the nation-state persists as a key site where the insecurities and uncertainties brought about by globalization are expected to be "resolved." It is even claimed that the state may "act out" in responding to some crime problems precisely because it has lost power elsewhere (Garland 1996). More functionally, some argue that states are obliged to enforce a new harsher type of order required because of the dismantling of welfare protections mandated by neoliberalism (Wacquant 2009a, b). Each country may also have its own reasons for increasing punitiveness. If the United States has seen "governing through crime" in a range of domestic settings (Simon 2007), in many European societies state power has been used mainly to criminalize noncitizen flows. In places such as South Africa, the state has to underline its ability to provide public safety in order to convince the "global economy" that it is a "safe place" in which to do business.

One of the most important tasks of the student of legal culture is to try to capture how far globalization represents the attempted imposition of one particular legal culture on other societies. Importing countries are offered both by the Anglo-American model whose prestige is spread by trade and the media and by national versions of the more intellectually impressive continental legal systems embodied in ready packaged codes. The Anglo-American model is characterized by its emphasis on the link between law and the economy (rather than law and the state) and its reliance on legal procedures that prioritize orality, party initiative, and negotiation inside law. More than any particular feature of legal procedure, however, what does seem to be spreading is the common law ideology of "pragmatic legal instrumentalism," the very idea that law is something which does or should "work," together with the claim that this is something which can or should be assessed in ways which are separable from wider political considerations.

Elements and Aggregates

A problem linked to the task of identifying the unit(s) of legal culture is that of distinguishing the elements of (legal) culture as opposed to treating it as an aggregate. Cotterrell famously criticized Friedman for treating legal culture both as an element – discussing why people turn to law – and as an aggregate-as when speaking of American legal culture or modern or global legal culture (Cotterrell 1997). Glenn too tells us that what he finds particularly problematical is the employment of culture as a "holistic signifier" and as a "variable" (Glenn 2004). On the other hand, the distinction between aggregates and elements is not a hard and fast one. All wholes can be incorporated into yet larger ones, just as all elements can be broken into yet smaller ones. For example, the group "attitudes" toward the use of law that are at the center of Friedman's use of legal culture can also be broken down into smaller elements. Indeed, Friedman thinks it is plausible to speak of each individual's legal culture. And these individual attitudes or opinions are in turn themselves composed of measurable responses to a range of particular questions. Whether it is appropriate to go down or up in levels of abstraction will depend on the purpose of an enquiry, for example, whether are, comparing whole societies, or are seeking to explain interconnections within them.

Recently, Merry has argued that we need to break up the idea of legal culture into what she calls four "social dimensions." The first is the practices and ideologies within the legal system, everyday way of getting things done, shared assumptions about good and bad clients, and other internal rules and practices, some of which are based on legal doctrine and others on categorizations shared by the wider society, such as ideas of race and gender. (As she says, this corresponds to Friedman term "internal legal culture.") Then there is the public's attitude toward the law. Is the legal system a source of corruption and ethnic preference, for example, or viewed as an institution that offers the rule of law for all people equally, regardless of their background? This, she suggests, is somewhat similar to what Friedman calls external legal culture. Thirdly, there is the question of legal mobilization, which refers to how readily people define their problems in legal terms, when they turn to the law for help. A fourth dimension is legal consciousness, the extent to which an individual sees him or herself as embedded in the law and entitled to its protections. Experience with the law, both good and bad, can change legal consciousness. It may encourage further use or may drive the litigant to avoid the law next time. Merry argues that these last two aspects offer the best way to understand the cultural dimensions of law and its relationship to a social context as well as providing a more satisfactory analysis of the processes of translation across legal fields and the hybridism of these fields (Merry 2012).

Nevertheless, others insist on the need to grasp the aggregate as a whole. Stewart Field, for example, argues that "actors do not live in a world of differentiated elements of institutions, formations, structures of feeling and traditions, however useful the distinctions may be as a heuristic device. The ultimate and impossible challenge that legal anthropologists and comparative lawyers must set themselves is to try to get something of the subjective feel of the normative pressures operating on the legal 'other' while making those pressures explicit in a way that native legal actors would not and perhaps could not do." The goal for him is to show how "the interpretation of a wide range of operative legal concepts is shaped by distinct social connotations, because these practices only 'make sense' within particular sets of legal and broader social contexts and relationships. No doubt, these meanings are fragile, contested, and subject to change. But the argument is that there are distinctive cultural 'logics' at work, distinct ways of seeing into which the researcher must struggle to enter" (Field 2012).

Coherence and Change

The last way of thinking about what gives culture its hold on us is to ask what makes for the (alleged) coherence of legal cultures and the way this relates to the possibilities of change. Do aspects of law in society come in "packages"? How do elements turn into a whole? What gives a unit its "unity"? For many writers, coherence (others might call it path dependency) is the key to "explaining" continuities in patterns of ideas and practices over time. It predicts how it is likely to respond to attempts at legal transfers. But for those critical of the concept of legal culture, the presumption of coherence is the problem not the solution.

Comparative lawyers and philosophers of law tend to define the term with reference to the activities of the various legal professionals and jurists who bear the responsibility of (re)producing purported coherence in legal materials. But they say little about what type or degree of coherence is required in actual practice, and their way of using the term is somewhat narrow for socio-legal enquiry. A variety of possible phenomena may be involved – opinions and attitudes, behaviors, texts, and institutional and organizational norms, ideas, and ideals. What are at stake may be psychological pressures to consistency, pressures for group conformity, institutional and organizational controls or routines, or even loyalty to legal or religious texts.

For some scholars, there can be interdependence between the elements that make up a given legal culture (as between the number of lawyers and the number of trials). Alternatively, coherence may be imposed by commentators, for example, when they categorize types of legal culture or legal "communities" in terms of ideal types. But others think that any connection exists (only) insofar as participants talk about it "as if" it is real. Merry, as we have seen, sees legal culture as no more than a framework with different overlapping elements with no necessary overall coherence – little more than a series of topics to investigate concerning institutions, attitudes, mobilization, and consciousness. For her, "Legal practices tend to be hybrid and creolized, formed of borrowings, transplants and translations of other legal practices in other places and times. Hence, it is unlikely that it will be coherent. Indeed, as with culture more generally, lack of coherence is what gives actors room for manoeuvre and innovation" (Merry 2012).

Legrand takes "the notion of "culture" to mean the framework of intangibles within which a community operates, which has normative force for this community (even though not completely and coherently instantiated), and which determines the identity of a community as community" (1997). But if culture is, to a large extent, a matter of struggle and disagreement, the purported uniformity, coherence, or stability of given national or other cultures will often be no more than a rhetorical claim manipulated by members of the culture concerned or projected by outside observers. In particular, as Roger Cotterrell tells us, it is judges and lawyers who attempt to sell the ideology of law as a "gapless system," whatever they may know from their everyday practice in their offices or in the courts (1997).

Much depends on scale and perspective. Under scrutiny, even defined areas of law, such as those governing family relations, are far from coherent. But, seen cross-nationally, even apparently unconnected branches of law may in fact manifest remarkable levels of cultural similarity within a given society. As Whitman has claimed recently, in replying to criticisms of his "culturalist" approach to penal law, "the pattern that we see in comparative punishment is also the pattern we see in many other areas of the law. Indeed, I would claim it as a virtue of my book that it shows that punishment law cannot be understood in isolation from the rest of the legal culture. For example, American workplace harassment law differs from German and French workplace harassment law in very much the same way. The same is true of comparative privacy law . . .just as it is true of the law of hate speech and everyday civility" (Whitman 2005).

But even if legal culture possesses coherence, this does not stop it from changing. It can be salutary to recall the rapid transformation in attitudes toward "law and order" in the short period that elapsed from Weimar to Hitlerian Germany. What is clear is that any serious work on legal culture must make sense of both continuity and change. The strain toward coherence – whatever causes it – explains relative lack of change, the difficulty of change, and even the direction of change. But it does not stop all change. To go back to the previous examples of country differences, in legal culture, England and Wales have recently witnessed some important corruption scandals to do with parliamentary expenses that were not connected with sex. And Berlusconi, Italy's last prime minister, is now being pursued for sex scandals apart from any corruption implications. Nonetheless, it is still fair to say that, in the UK, scandals concerning business-politics links are still much less salient than in Italy and Berlusconi's fall from power depended more on the financial crash than his private life. In an ever more interconnected world, even resistance to outside influence is an active process rather than a result of cultural inertia.

Related Entries

- ▶ Cross-National Performance in Policing
- ▶ Cultural Criminology
- ▶ Effective Community Supervision Programming
- ▶ International Policy Transfer
- ▶ International Responses to Victims in Criminal Justice
- ▶ Sentencing as a Cultural Practice
- ▶ Transnational Exchange of Forensic Evidence

Recommended Reading and References

Appudurai A (1995) Modernity at large: cultural dimensions of globalization. University of Minnesota Press, Minneapolis

Blankenburg E (1997) Civil litigation rates as indicators for legal culture. In: Nelken D (ed) Comparing legal cultures. Dartmouth, Aldershot, pp 41–68

Coombe RJ (1998) Contingent articulations: a critical cultural studies of law. In: Sarat A, Kearns P (eds) Law in the domains of culture. University of Michigan Press, Ann Arbor, pp 21–65

Cotterrell R (1997) The concept of legal culture. In: Nelken D (ed) Comparing legal cultures. Dartmouth, Aldershot, pp 13–32

Engel DM (2012) The uses of legal culture in contemporary socio-legal studies: a response to Sally Engle Merry. In: Nelken D (ed) Using legal culture. Wildy, London

Field S (2012) Comparative youth justice and legal cultures. In: Nelken D (ed) Using legal culture. Wildy, London
Friedman L (1994) Is there a modern legal culture? Ratio Juris 7:117
Friedman L (2006) The place of legal culture and its position in the sociology of law. In: Freeman M (ed) Law and sociology. Oxford University Press, Oxford
Galanter M, Krishnan JK (2003) Debased informalism: *Lok adalats* and legal rights in modern India. In: Erik GJ, Thomas CH (eds) Beyond common knowledge: empirical approaches to the rule of law. Stanford University Press, Stanford, p 96
Garland D (1996) The limits of the sovereign state: strategies of crime control in contemporary society. Br J Criminol 36:445–471
Geertz C (1973) The interpretation of cultures. Basic Books, New York
Glenn HP (2003) The nationalist heritage. In: Legrand P, Munday R (eds) Comparative legal studies: traditions and transitions. Cambridge University Press, Cambridge, p 4
Glenn HP (2004) Legal cultures and legal traditions. In: van Hoeck M (ed) Epistemology and methodology of comparative law. Oxford University Press, Oxford, p 32
Hofstede G (1980) Culture's consequences: international differences in work related values. Sage, Beverly Hills
Holmes W, McCaan M (2004) Distorting the law. Chicago University Press, Chicago
Kurchiyan M (2009) Russian legal culture: an analysis of adaptive response to an institutional transplant. Law Soc Inquiry 34:337
Legrand P (1997) Review of David Nelken's comparing legal cultures. Camb Law J 56:646
Merry SE (2012) What is legal culture? An anthropological perspective. In: Nelken D (ed) Using legal culture. Wildy, London
Nelken D (1995) Understanding/invoking legal culture. In: Nelken D (ed) Special issue on legal culture, diversity and globalization. Social Legal Stud 4(4):435–52
Nelken D (1997) Puzzling out legal culture. In: Nelken D (ed) Comparing legal cultures. Dartmouth, Aldershot
Nelken D (ed) (2000) Contrasting criminal justice. Ashgate, Aldershot
Nelken D (2004) Using the concept of legal culture. Austl J Leg Phil 29:1–28
Nelken D (2006) Rethinking legal culture. In: Freeman M (ed) Law and sociology. Oxford University Press, Oxford, pp 200–224
Nelken D (2007) Defining and using the concept of legal culture. In: Orucu E, Nelken D (eds) Comparative law: a handbook. Hart, Oxford, pp 109–132
Nelken D (2010) Comparative criminal justice: making sense of difference. Sage, London
Nelken D (2012) Using legal culture: purposes and problems. In: Nelken D (ed) Using legal culture. Wildy, London
Silbey S (2001) Legal culture and legal consciousness. In: Encyclopedia of the social sciences. Elsevier/Pergamon Press, New York, pp 8623–8629
Silbey S (2005) After legal consciousness. Annu Rev Law Soc Sci 1:323
Silbey S, Ewick P (1998) The common place of the law: stories from everyday life. University of Chicago Press, Chicago
Wacquant L (2009a) Prisons of poverty. University of Minnesota Press, Minneapolis
Wacquant L (2009b) Punishing the poor: the neoliberal government of social insecurity. Duke University Press, Durham
Webber J (2004) Culture, legal culture, and legal reasoning: a comment on Nelken. Austl J Legal Philos 29:25–36
Whitman JQ (2005) Response to Garland. Punish Soc 7:389

Comparative Police Systems

► Democratic Policing

Comparative Political Economy of Punishment

Nicola Lacey
All Souls College Oxford, University of Oxford, Oxford, UK

Synonyms

Comparative criminal justice

Overview

Recent decades have seen a revival of interest in the broad social, political, and economic forces which are shaping criminal justice policy in Western democracies. The global economic changes which began in the 1970s – recession, the contraction or collapse of manufacturing industries, the growth of unemployment, and the creation of a large sector of people either unemployed or employed in insecure forms of work – it is argued, have eroded the consensus which sustained postwar penal welfarism. As rising crime across Western countries gradually produced a situation in which the experience of criminal victimization and of managing the risk and fear of crime became normal features of everyday life, crime became increasingly politicized, leading to exclusionary criminal justice policies (Pratt 2007).

In making this argument, scholars like Garland (2001), De Giorgi (2006), Reiner (2007), Wacquant (2009), and Young (1999) have refined the Marxist account deriving from Rusche and Kirchheimer (1969) with an analysis of cultural and political factors, thus avoiding the pitfalls of economic reductionism. Yet in each case, an emphasis on structural forces tends to direct attention away from variations in the institutional framework through which those forces are translated into concrete social policies in different countries. While each of these accounts, therefore, suggests hypotheses which are ripe for comparative investigation, a comparative focus does not feature prominently in their analyses (though see Garland 2007).

Not all "late modern" democracies, however, have opted for a neoliberal politics (Lacey 2012). And many have managed to sustain relatively moderate, inclusionary criminal justice systems, while the British and American systems have, albeit to different degrees, been moving toward ever-greater penal severity. Even as between Britain and the USA, the differences in terms of the overall scale and quality of punishment are striking. Countries like Denmark, Germany, or Sweden fulfill the prophecy of "penal populism" yet less accurately (Lacey 2008; Cavadino and Dignan 2006).

But how do these resemblances across types of political economy hold together over time, and why do they produce systematically different patterns of punishment? Comparative institutional approaches seek to answer these questions. They argue that progress can be made toward a genuinely explanatory account by drawing on the political-economic analysis of comparative institutional advantage and of the capacities for strategic coordination inherent in differently ordered systems. Structural forces are seen as mediated not only by cultural filters but also by economic, political, and social institutions: it is the institutional mediation of cultural and structural forces, and its impact on the interests and identities of relevant social actors, which is argued to produce the persistent variety in criminal justice which, notwithstanding globalization (Lacey 2011), we see across systems at similar stages of capitalist development (Sutton 2004; Tonry 2007).

Institutional Approaches

Recent comparative institutional analysis (Lacey 2008) builds on the distinction between "liberal" and "coordinated" market economies developed by political scientists Peter Hall and David Soskice (2001). A "coordinated market economy" (CME) functions in terms primarily of long-term relationships and stable structures of investment, not least in education and training oriented to company- or sector-specific skills, and incorporates a wide range of social groups and institutions into a highly coordinated governmental structure. Such a system is argued to be more likely to generate incentives for the relevant decision-makers to opt for relatively inclusionary criminal justice for it is a system which is premised on incorporation and hence on the need to reintegrate offenders into society and economy. Typically, moreover, the interlocking institutions of coordination of the CMEs conduce to an environment of extensive informal social controls; this in turn supports the cultural attitudes which underpin and help to stabilize a moderated approach to formal punishment.

A "liberal market economy" (LME) – of which the extreme case, significantly for any argument about criminal justice, is the USA – is typically more individualistic in structure, is less interventionist in regulatory stance, and depends far less strongly on the sorts of coordinating institutions which are needed to sustain long-term economic and social relations. In these economies, flexibility and innovation, rather than stability and investment, form the backbone of comparative institutional advantage. It follows

		Imprisonment per 100,000			Homicide per 100,000
	2002/3	2006	2009/10	2011	2008
Liberal Market Economies					
USA	701	737	743[å]	-	5.2
New Zealand	155	186	203	199	1.3
England & Wales	141	148	154	155	1.2
Australia	115	125	133	-	1.2
Scotland[a]	128	140	153	155	2.2*
Co-ordinated Market Economies					
Netherlands	100	128	94	-	1*
Germany	98	94	88	87	0.8
Sweden	73	82	78	-	0.9
Denmark	58	77	71	74	1.4*
Finland	70	75	61	59	2.5
Norway	58	60	72	73	0.6*

[å] Prison rate for 2009

* Homicide rates for 2007

[a] Prison rates for Scotland are estimated on the basis of average daily prison population for 2002/3, 2006/7, 2009/10

Adapted from: Hall and Soskice (2001), Barclay and Tavares (2003), Cavadino and Dignan (2006), Hall and Gingerich (2004), International Centre for Prison Studies (accessed 2011). European Sourcebook of Crime and Criminal Justice (2010); UNODC (accessed 2010); General Register for Scotland (accessed 2011); Prison Statistics Scotland (accessed 2011)

Comparative Political Economy of Punishment, Fig. 1 Punitive measures in cross-national comparison

that, particularly under conditions of surplus unskilled labor – conditions which LMEs are also more likely to produce – the costs of a harsh, exclusionary criminal justice system are less than they would be in a CME (Fig. 1).

Political Economy, Imprisonment, and Homicide

This institutional approach sees the distinction between different varieties of advanced capitalist political economy as a powerful tool in building an understanding of the inclusionary and exclusionary dynamics of criminal justice systems. For, particularly as developed in the most recent political economy literature, which explores the relationship between the regime of economic production and labor market institutions and a cluster of social and political institutions, the distinction has an analytic reach into a wide range of political and economic variables. The various factors which characterize coordinated and liberal systems are closely intertwined: the number of variables further implies that there may be paradigm and less central cases of each type of system. The features of the LME/CME distinction of greatest relevance to criminal justice are as follows:

1. *The structure of the economy: production regimes, labor markets, education and training, and disparities of wealth*

   The departure point for an analysis of two main varieties of capitalism – an individualistic liberal model whose comparative advantage lies in flexibility and a coordinated model in which comparative advantage depends on long-term investment in workers and their skills – lies in the changing nature of economic relations. In Britain and the USA, for example, the dynamics of an LME have accelerated markedly over the last 30 years, as many of the attitudes and values which sustained, respectively, the postwar welfare state settlement and the Great Society programs have come to be eroded by a more aggressively market-oriented culture (Reiner 2007). This culture is itself premised in part on the imperative of high performance amid increasing global economic competition, with the collapse of Fordist production regimes and the availability of cheap manufactured goods from countries like South Korea, China, and India. The inevitable upshot is structural economic insecurity for low-skilled workers. In a short-term economic culture, the bottom third of the workforce risks becoming a socially as well as economically excluded group.

In the CMEs, by contrast, a longer-term economic culture appears to have survived increased international competition and the collapse of Fordism. Within the political economy of comparative advantage, this is seen as a function of several interlocking factors: the nature of the economic activities in which these countries have concentrated their efforts, the close incorporation of employers as well as unions in the management of the economy, and the implications of each of these factors for the structure of education and training. Unlike the increasingly flexibilized LMEs, many CMEs excel in producing high-quality goods which depend on industry-specific, nontransferable skills. In this context, employers have strong reason to invest in education, training, and apprenticeship systems. They also have strong reason to use their considerable bargaining power with government to press for generous welfare provision for workers who are temporarily unemployed but whose skills remain necessary to the economy. With the higher levels of investment in education and training typical of these economies, which also demonstrate lower disparities of wealth and higher literacy rates (Iversen and Soskice 2009), the costs of pursuing socially exclusionary policies in areas such as criminal justice are relatively high. This implies that many of the general theories of increasing penal severity are based on an account primarily applicable to LMEs as opposed to CMEs, whose high-skill production regimes were less strongly affected by the collapse of Fordism.

In the LMEs, increasing relative deprivation consequent on flexibilization of labor markets and growing disparities of both income and skills pose a huge challenge for inclusionary criminal justice policies (Reiner 2007, Chap. 3), particularly in a world in which mass communications and increased levels of education imply the cultural inclusion of the relatively deprived within the individualistic values of a consumer society from which they are economically excluded. What is more, this relative deprivation has increased along a number of dimensions, accentuating both differences between the richest and the poorest and the difficulty of moving from a position in the bottom one third or so into the relatively advantaged majority (Young 1999). In this context, prisons have become a mechanism for disciplining those excluded from the legitimate economy. During this period there has been a large increase in the absolute and relative size of the harsher end of the American and, to a lesser extent, British criminal justice systems. In England and Wales, the imprisonment rate virtually doubled – from 79 to 152 per 100,000 – between 1970 and 2011 (http://www.prisonstudies.org/info/worldbrief/wpb_country.php?country=169 accessed August 25th 2011), while in the USA, it quintupled – from 153 to 743 per 100,000 – between 1974 and 2009. This is reflected not only in the scale of imprisonment and policies such as mandatory minimum sentences but also in a weakening of political sensibilities in favor of human rights and decent conditions for prisoners.

2. *Political systems: electoral arrangements and the bureaucracy*

The analysis of two systematically different variants of contemporary capitalist economy deploys not only economic variables but also a variety of other institutional variables. Among the most important of these for criminal justice is electoral politics. Given the increasing salience in many countries of criminal justice to politics and the force of electoral discipline on democratic governments, the relationship between popular attitudes to crime and punishment and the political system constitutes an important variable in any attempt to understand the differences between contemporary penal systems in relatively similar societies. This is the case even if popular attitudes conducing to fear of crime and to penal severity are on occasion stimulated by government rhetoric and policy (Beckett 1997; Roberts and Hough 2002). Whatever the causation here – it seems likely that it moves in both directions – once certain popular attitudes and expectations are created, they in turn create significant electoral constraints. How directly they are registered in the electoral system (Barker 2009; Miller 2008), and hence exert discipline on governing parties, is therefore likely to be an important factor in explaining the institutional capacity of different systems to sustain moderate criminal justice policies.

In this context, it is significant that there is, empirically, an association between CMEs and proportionally representative (PR) electoral systems, whereas LMEs tend to have first-past-the-post, winner-takes-all systems (the key exceptions being New Zealand and Scotland: Lacey 2012). The policy-making autonomy of CMEs is constrained by the need to negotiate with groups incorporated in the governmental process. CMEs with PR systems are more oriented to effective participation in and contribution to policy making – at least for groups integrated within subsisting sociopolitical structures – than are LMEs whose electorate gets a one-shot say in policy making at election time. But this consensus-building dynamic may make the CMEs less heteronymous in the light of swings of popular opinion. While decisive winners of first-past-the-post elections in LMEs may feel relatively unconstrained by popular opinion early on in their terms, their unmediated accountability at the ballot box makes them highly sensitive to public opinion as elections loom.

What is more, as party affiliations among the electorate weaken, governments' increasing dependence on the approval of a large number of "floating" or reluctant voters, who regard crime as a threat to their well-being, may feed into the political salience of criminal justice. Under the conditions attendant on the collapse of Fordism since the 1970s and in the light of the salience of increasing relative deprivation to the scale and the perceived seriousness of crime problems, there may thus be a stronger association between the politicization of criminal justice and the impact of penal populism in majoritarian, two-party LMEs such as the USA, with decisive implications for the harshness of punishment. In this context, the further empirical fact that PR-based CMEs are more likely to elect left of center governments and to display lower disparities between the best and worst off is significant (see Iversen and Soskice 2006). Electoral structure, in other words, has implications for both partisanship and the substance of political, social, and economic outcomes.

There is further reason to think that this difference in the electoral structure of LMEs and CMEs may have some important implications for upward pressure on punishment. As many commentators have observed, one of the developments which has fed the trend toward penal populism in several LMEs is the emergence of well-organized single-issue pressure groups, notably those representing victims of crime (Pratt 2007, Chap. 3; Gottschalk 2007, Chaps. 5–7). On the face of it, one would expect that single-issue political groups would find it harder to get their voices heard in a majoritarian, two-party system than in a PR system which incorporates a number of smaller parties. This expectation would not apply, however, where a particular single issue appeals widely to floating or reluctant voters. Such has been the case of crime in the USA over the last 30 years. As Simon (2007, pp. 26ff, 159ff) has shown, both the popularity of harsh criminal policy among median voters and the relative simplicity of enacting such policy – the lack of need, for example, to develop complex new bureaucracies to administer or implement increased criminalization – have proved a potent temptation to US politicians and other elected officials.

Furthermore, while concerns about crime reflected in victims' movements may well find a footing in the PR environment, particularly among smaller parties who may hold the balance of power, their adoption and implementation will ultimately have to be negotiated out in the complex bargaining process typical of PR systems and will hence tend to be more insulated from the dynamics of emotive campaigns than is typical in the majoritarian systems. What is seen in the latter is a cycle of mutual reinforcement, grounded in a set of incentives conducing to politicians' focus on single issues such as criminal justice which are, superficially, easy to demonstrate that they have acted upon, such action in turn leading to heightened public identification of the salience of crime problems and to heightened expectations of governmental capacity to resolve them through tough criminal policy. This dynamic is especially strong where both major parties adopt a tough stance on law and order (Newburn 2007). The scale and impact of this sort of cycle is vividly illustrated by the fact that, in addition to its unplanned expansion of the prison system, the British Labor government has been estimated to have enacted no fewer than 3,000 additional criminal offences in the 8 years between its election and the end of 2005 (Morris 2006).

It is also worth noting a further difference between the political systems to be found in LMEs and CMEs, itself correlated with the PR/majoritarian distinction. In most CMEs, deference to the expertise of the professional bureaucracy – that is, the civil service, often including not only policy advisers, penal system officials, and prosecutors but also judges – tends to be higher than in LMEs. This is in part because the coalition politics typical of PR systems imply a less polarized political environment in which governments feel less need to retain total control of policy making. By contrast, particularly in recent years, the tendency in majoritarian systems has been for governments to prefer to work with their own, politically appointed advisers, and to ignore the advice of technically neutral civil servants wherever this interferes with political expediency. Particularly in the UK, the increasing domination of parties by their leaders has fed this dynamic. This not infrequently leads to a situation in which pressure groups such as those representing victims of crime are constructed as the relevant "experts" for the purposes of consultation in the development of policy, weakening both an important constraint on ad hoc policy making and coordination with criminal justice professionals (Pratt 2007, Chap. 3; Simon 2007, Chap. 3). In LMEs, this feeds into a dynamic in which politicians' decisions become ever less insulated from the flow of perceived public opinion – a factor which has been a crucial driver of penal harshness in several countries. In CMEs, by contrast, the strength of a professional bureaucracy, along with deference to expertise, have been identified by a number of scholars as conditions key to the maintenance of moderate criminal justice policies (Cavadino and Dignan 2006, pp. 35–36; Savelsberg 1994, 1999; Tonry 2007, pp. 31–32).

3. *The welfare state*

Another key difference between contemporary capitalist democracies lies in the institutions associated with the welfare state (Esping-Andersen 1990; 1996). Here again, political economies at relatively similar levels of development, characterized by broadly liberal-democratic political structures, have taken markedly different paths. In terms of Esping-Andersen's famous typology, while countries with liberal welfare regimes like the USA and Britain have adopted neoliberal policies committed to "rolling back the state" and curtailing public expenditure (Hillyard and Tombs 2005), the Nordic, social-democratic countries have maintained their welfare states more or less intact (Pratt 2008a; Pratt 2008b), with Northern European countries such as Germany adopting a pattern closer, in terms of generosity of provision and scope of coverage, to their Nordic than to their British neighbors.

Among variables in political-economic structure, the welfare state has received most sustained attention from comparative criminal justice scholars. It is obviously plausible that the impact of relatively generous welfare provision on the reduction of poverty has a knock-on effect on crime. Less obviously, there is evidence that it is also associated with levels of punishment. Downes and Hansen (2006) have shown, in a study covering 18 countries, that those spending a higher proportion of their GDP on welfare have lower imprisonment rates – a relationship which has grown stronger over the last 15 years. Similarly, Beckett and Western (2001, pp. 44, 48, 55; Western 2006) have demonstrated systematic differences among states within the USA, in which those with relatively low social welfare spending are also those with higher prison populations.

But the precise causal mechanisms here are unclear. Sometimes the argument takes a cultural form: the inclusionary instincts represented in generous welfare state policies are likely to be reflected in criminal justice policy. Beckett and Western (2001), for example, argue that welfare regimes vary according to their commitment to including or excluding marginal groups: the more inclusive systems exhibit both higher welfare spending and lower imprisonment rates. Albeit making some observations about the relationship between the welfare state and the structure of the economic and political systems, Pratt (2008a), in his recent analysis of "Scandinavian exceptionalism," attributes the generosity of the Scandinavian welfare systems primarily to a "culture of equality" with long historical roots. But what explains this varying commitment to egalitarianism and inclusion and underpins the political support necessary to sustain it?

Clearly, long-standing institutional arrangements are typically articulated with, and stabilized by, distinctive cultural attitudes such as the strong Scandinavian commitment to social solidarity and equality (Bondeson 2005, p. 189). But there is also evidence that the distinctive structures of welfare states are articulated with the political and economic dynamics already discussed. In other words, it may be not just possible, but good economic sense for some countries to maintain generous welfare provision even in the face of increasing competition from countries who are not investing public resources in this way. Certain economic arrangements, in short, themselves foster a culture of solidarity or support for the welfare state – a culture which is in its turn important in sustaining the political support needed to sustain generous welfare institutions.

A range of explanations has focused on precisely such an articulation of the welfare state to the structure of the economy. Within a liberal market system in a flexible economy, governments have chosen to maximize incentives to rejoin the labor market – a strategy that has had sufficient plausibility with a critical mass of the electorate because of the high degree of transferable skills within the workforce. Within the labor markets of countries with less flexibility, where long-term investment in less transferable skills (as in Germany), or an extensive public sector providing employment for women and services for dual-career families (as in the Nordic countries), is still key to comparative advantage, it makes sense to give relatively generous support to workers who experience periods of unemployment, rather than encouraging them to retrain or to find work in new sectors of the economy. Generosity of welfare provision and relatively secure employment relations are, under certain conditions, just as good a basis for economic success and stability as flexibilization and welfare cuts. And though there has been some recent reduction of welfare benefits in several of the European and, to a lesser extent, Nordic countries, with generosity of provision in the countries like Germany depending heavily on status as an insider to the high-skill economy, the remaining differences between European welfare regimes remain significant and seem unlikely to be eroded in the near future (Hall 2007, p. 39; Iversen 2007, p. 278).

4. *Constitutional structure: decision-makers, veto points, and constraints on criminalization*

Beyond political and economic institutions, the constitutional structure which provides parameters for the institutional environment already considered, and for the legal system through which much criminal policy is implemented, may also be important in explaining national differences in criminal justice policy. Three such variables might be thought to entail systematic differences in countries' capacities to develop and sustain moderate penal practices. These are the distribution of decision-making power among different actors; the structure of legal institutions, and in particular of the tenure and selection of the judiciary and prosecutors; and the impact of constitutional framework on the definition of crime.

The discussion of political systems above indicated that one important variable in shaping the reception of popular concern about crime into criminal justice policy is the degree to which political decision-making is insulated from the flow of electoral opinion. But a further factor seems likely to be important here. This is the distribution of veto points across the system, with consequent potential for checks and balances to be invoked, and for reactive policies to be blocked or delayed, thus leading to a more reflective and negotiated style of policy making.

On the face of it, such veto points might be taken to be correlated with federal as opposed to unitary systems and with bi- or multicameral as opposed to unicameral legislatures. But this is not inevitable. For example, in the USA, most criminal justice policy is developed at state level or even at city or county level, so that the federal structure cannot exert the sort of inhibiting power which it has in Germany. And in the bicameral legislature of the UK, the constitutionally privileged place of the House of Commons restricts the degree to which the upper House can act as a block on policy formation. This is not to say that, for example, US federalism has been unimportant to the way in which criminal policy has developed (Miller 2008; Stuntz 2001). The need to address multiple constituencies can, as in the case of campaigns to abolish capital punishment (Garland 2010), place barriers in the path of criminal justice reform, while the highly decentralized form of US government allows local politicians to emphasize popular issues the costs of which their own constituents will not have to bear (Soskice 2009; Lacey 2011; Stuntz 2011). And the creation of federal structures can itself feed upward trends in punishment: this appears to have been the case in the USA through both direct and indirect (via state-imitation) effects of the federal Uniform Determinate Sentencing Act and, in Europe, in recent EU initiatives arguing for cross-union statutory minima for certain serious crimes (Pratt 2007, p. 169).

There is scope here for a careful empirical analysis of the ways in which systems with more and less diffused structures of criminal justice decision-making have responded to external pressures toward penal severity. A priori, the distribution of veto points and the need to coordinate decision-making points in federal systems where key aspects of criminal justice policy have to be centrally or constitutionally determined would appear to be important variables. The size or scale of the system is also likely to be relevant. In relatively small jurisdictions like the Netherlands or Scotland, both the capacities for central negotiation and the intensity and influence of elite policy networks may also be important factors. They are differences which may favor either the stabilization of moderate policies via coordination between elite networks, as in the 1970s in the Netherlands, or rapid shifts of policy through precisely the same mechanism, as in that country in the late 1980s (Cavadino and Dignan 2006, p. 123). Leaving the question of scale aside, however, the key feature of the long-established PR systems, whose parties tend to represent sectoral interests, is their resulting orientation to negotiation across groups. This is likely to conduce to a wider distribution of veto or delaying points, and hence underpins an association between CMEs and checks on the politicization of criminal justice.

Second, it is plausible that the selection, training, and tenure of judges and other key criminal justice officials will have distinctive implications for the environment in which penal policy is developed and implemented. To take only the most obvious example, a system like that of the USA, in which many judges are elected, is one in which a key barrier between popular demands for punishment is, if not removed, seriously weakened. This becomes important under conditions in which criminal justice is highly politicized. We might draw an analogy here between the election of many US judges and the fact that the vast bulk of criminal cases in Britain are heard by lay magistrates. This arrangement would be unthinkable in the highly professionalized systems of Northern Europe or Scandinavia. While English magistrates are not answerable to popular opinion in the style of elected officials, they are less buffered by a professional expertise and culture. Their place in the administration of criminal justice suggests that relatively low importance attached to expertise may find its roots deep in the history of British sociopolitical arrangements rather than merely being a product of the recent dynamics of criminal policy.

In many Northern European and Nordic CMEs, the judiciary continues to be regarded as a key partner in the development as well as the implementation of criminal justice policy. In LMEs like the USA and the UK, by contrast, the rise of penal populism has seen an increasingly hostile and unstable relationship between government and judiciary. The judiciary conceives their independence as inconsistent with any overt incorporation in governmental negotiations, and the government is accordingly inclined to regard the judiciary as an irksome and even irresponsible thorn in the flesh of its criminal policy. Though constitutional or human rights structures such as the European Convention or the US Constitution may provide judges with some tools to resist certain government excesses of punishment or criminalization, these have tended to be relatively weak in the face of a determined executive with a clear legislative majority (Lacey 2009). In this context it is interesting that judiciaries in the Anglo-Saxon, common law, liberal market countries with strong traditions of judicial independence appear to have suffered a decline in public status and authority in recent years (see Simon (2007, Chap. 4); while their European, civilian cousins, traditionally of lower status and more intimately linked to the state bureaucracy, appear so far to have escaped a similar fate.

Third, systems exhibit markedly variable constitutional constraints on the content of criminal law and punishment. These constraints have deep historical roots. Whereas in most European countries, the process of constitutional modernization entailed a clear differentiation between criminal justice and administrative "police power," in common law systems such as Britain and the USA, no such differentiation has ever been fully institutionalized (Dubber 2005; Whitman 2003). Legal and constitutional constraints on the substance (as opposed to the procedure) of both criminalization and punishment are, accordingly, weaker in these systems. Again, there is an association between this variable and the LME/CME distinction.

5. *Institutional capacity to integrate "outsiders"*

Another key feature of contemporary societies is the increasing mobility of the social world from the late 1960s on. This mobility has a number of dimensions. In a wealth-valuing culture and flexible economy, with relatively high levels of education, there may be more mobility between social classes; in a globalizing economy characterized by transnational political structures like the EU, marked by relatively cheap international travel and mass communications, there is more geographical mobility. These developments have added new layers of complexity to one of the central challenges for any democratic system of criminal justice: that of "reintegrating" offenders into society and economy.

This is a complex issue. But it is an important corrective to what might the temptation to think that the highly coordinated systems of Europe and Scandinavia today are necessarily better placed to sustain democratically acceptable levels of penal moderation than their liberal market Anglo-Saxon counterparts. For the structure of this problem is significantly different in the two sorts of system. While the laisser-faire and individualistic culture typical of LMEs may well make it relatively easy to integrate geographical or "cultural" "outsiders" like recent immigrants wherever they find access to the labor market, the more intensively group- and skills-based system of the CMEs may well pose significant challenges in terms of integrating newcomers into the representative and decision-making structures which have helped to sustain a relatively moderate criminal justice policy with relatively high institutional capacity for reintegration. CMEs are, in short, good places to be incorporated insiders, but hard systems to enter from the outside (Lacey 2008, Chap. 3).

Future Directions

The comparative political-economic analysis of crime and punishment, using the tools of contemporary political science, is at a relatively early stage of its development. But it has opened up an extensive research agenda which promises to refine the understanding of how the dynamics of crime and punishment are shaped by forces external to the criminal process and of how those forces themselves relate to patterns of institutional structure and incentives which vary systematically across national borders. At a time at which much criminal justice scholarship is preoccupied with globalization, institutional approaches grounded in comparative political economy are a telling reminder that vast differences between criminal justice systems of the developed world continue to obtain, as well as a crucial resource for understanding how they are produced and, accordingly, for assessing whether they are likely to persist.

Related Entries

- ▶ Comparative Courts and Sentencing
- ▶ Comparative Legal Cultures
- ▶ Comparing Police Systems Across the World
- ▶ Historical and Comparative Perspectives on Incarceration
- ▶ History of Corrections
- ▶ Informal Social Control
- ▶ Public Opinion about Sentencing
- ▶ Sentencing Commissions
- ▶ Sex Offenders and Criminal Policy
- ▶ Social Control Theory of Sexual Homicide Offending

Recommended Reading and References

Barker V (2009) The politics of punishment: how the democratic process shapes the way America punishes offenders. Oxford University Press, New York

Beckett K (1997) Making crime pay. Oxford University Press, New York

Beckett K, Western B (2001) Governing social marginality. In: Garland D (ed) Mass imprisonment: social causes and consequences. Sage, London, pp 35–50

Bondeson U (2005) Levels of punitiveness in Scandinavia: descriptions and explanations. In: Pratt J, Brown D, Brown M, Hallsworth S, Morrison W (eds) The new punitiveness. Willan Publications, Cullompton

Cavadino M, Dignan J (2006) Penal systems: a comparative approach. Sage, London

De Giorgi A (2006) Rethinking the political economy of punishment. Ashgate, Aldershot

Downes D, Hansen K (2006) Welfare and punishment in comparative perspective. In: Armstrong S, McAra L (eds) Perspectives on punishment. Oxford University Press, Oxford

Dubber MD (2005) Police power. Columbia University Press, New York

Esping-Andersen G (1990) The three worlds of welfare capitalism. Polity Press, Cambridge

Esping-Andersen G (1996) Welfare states in transition. Sage, London

European Sourcebook of Crime and Criminal Justice Statistics – 2010, fourth edition (4th ed. Den Haag: Boom Juridische uitgevers. Onderzoek en beleid series, no. 285, Ministry of Justice, Research and Documentation Centre [WODC]). www.europeansourcebook.org/ob285_full.pdf

Garland D (2001) The culture of control. Oxford University Press, Oxford

Garland D (2007) High crime societies and cultures of control. In: Ostermeier L, Paul B (eds) Special Issue, Kriminologisches Journal

Garland D (2010) Peculiar institution: America's death penalty in an age of abolition. Oxford University Press, New York

General Register Office for Scotland. Revised Mid-year Population Estimates 1982–2000. www.gro-scotland.gov.uk/statistics/theme/population/estimates/mid-year/revised1982-2000.html

Gottschalk M (2007) The prison and the gallows. Cambridge University Press, Cambridge

Hall PA (2007) The evolution of varieties of capitalism in Europe. In: Hancké B, Rhodes M, Thatcher M (eds) Beyond varieties of capitalism. Oxford University Press, Oxford

Hall PA, Soskice D (2001) An introduction to the varieties of capitalism. In: Hall PA, Soskice D (eds) Varieties of capitalism. Oxford University Press, Oxford, pp 1–68

Hillyard P, Tombs S (2005) Towards a political economy of harm: states, corporations and the production of inequality. In: Hillyard P, Pantazis C, Tombs S, Gordon D (eds) Beyond criminology: taking harm seriously. Pluto Press, London, pp 30–54

International Centre for Prison Studies. World prison brief. http://www.prisonstudies.org/info/worldbrief/

Iversen T (2007) Economic shocks and varieties of government responses. In: Hancké B, Rhodes M, Thatcher M (eds) Beyond varieties of capitalism. Oxford University Press, Oxford

Iversen T, Soskice D (2006) Electoral Institutions and the politics of coalitions: why some democracies redistribute more than others. Am Polit Sci Rev 100:165–181

Iversen T, Soskice D (2009) Distribution and redistribution: the shadow of the nineteenth century. World Polit 61(3):438–486

Lacey N (2008) The prisoners' dilemma: political economy and punishment in contemporary democracies. Cambridge University Press, Cambridge

Lacey N (2009) Historicising criminalisation: conceptual and empirical issues. Mod Law Rev 72(6):936–960

Lacey N (2011) Why globalisation doesn't spell convergence: models of institutional variation and the comparative political economy of punishment. In: Crawford A (ed) International and comparative criminal justice and urban governance. Cambridge University Press, Cambridge

Lacey N (2012) Political systems and criminal justice: the prisoners' dilemma after the coalition. Current Legal Problems 2012. doi:10.1093/clp/cus002

Lacey N (forthcoming 2012) Punishment, (Neo-) liberalism and social democracy. In: Simon J, Sparks R (eds) The sage handbook of punishment and society. Sage Publications

Miller LL (2008) The perils of federalism: race, poverty, and the politics of crime control. Oxford University Press, New York

Morris N (2006) Blair's "Frenzied Law-Making". The Independent 16 August

Newburn T (2007) "Tough on Crime": penal policy in England and Wales. In: Tonry M (ed) Crime, punishment and politics in comparative perspective: crime and justice: a review of research, vol 36. University of Chicago Press, Chicago, pp 425–470

Pratt J (2007) Penal populism. Routledge, London

Pratt J (2008a) Scandinavian exceptionalism in an era of penal excess: part I the nature and roots of Scandinavian exceptionalism. Brit J Criminol 48: 119–137

Pratt J (2008b) Scandinavian exceptionalism in an era of penal excess: part II does Scandinavian exceptionalism have a future? Brit J Criminol 48:275–292

Reiner R (2007) Law and order: an honest citizen's guide to crime and control. Polity Press, Oxford

Roberts J, Hough M (eds) (2002) Changing attitudes to punishment: public opinion, crime and justice. Willan Publishing, Cullompton

Rusche G, Kirchheimer O (1969) Punishment and social structure, (first published, in German, 1939). Russell Sage, New York

Savelsberg JJ (1994) Knowledge, domination, and criminal punishment. Am J Sociol 99:911–943

Savelsberg JJ (1999) Knowledge, domination and criminal punishment revisited. Punishm Soc 1:45–70

Simon J (2007) Governing through crime: how the war on crime transformed American democracy and created a culture of fear. Oxford University Press, New York

Soskice D (2009) American exceptionalism and comparative political economy. In: Brown C, Eichengreen B, Reich M (eds) Labor in the era of globalization. Cambridge University Press, New York, pp 51–93

Stuntz WJ (2001) The pathological politics of criminal law. Mich Law Rev 100:505–600

Stuntz WJ (2011) The collapse of American criminal justice. Harvard University Press, Cambridge

Sutton J (2004) The political economy of imprisonment in affluent western democracies, 1960–1990. Am Sociol Rev 69:170–189

The Scottish Executive. Statistical Bulletin Criminal Justice Series (August 2007) CJr/2007/7. www.scotland.gov.uk/Resource/Doc/196743/0052707.pdf

The Scottish Government. Statistical Bulletin Criminal Justice Series (20 December 2010). www.scotland.gov.uk/Publications/2010/12/15154652/24

Tonry M (2007) Determinants of penal policies. In: Tonry M (ed) Crime and justice: a review of

United Nations Office on Drugs and Crime. UNODC Homicide Statistics. www.unodc.org/unodc/en/data-and-analysis/homicide.html

Wacquant L (2009) Punishing the poor: the neoliberal government of social insecurity. Duke University Press, Durham

Western B (2006) Punishment and inequality in America. Russell Sage, New York

Whitman JQ (2003) Harsh justice. Oxford University Press, Oxford

Young J (1999) The exclusive society. Sage, London

Comparative Punishment

► Institutional Theories of Punishment

Comparing Police Systems Across the World

Rob I. Mawby
University of South Wales, Newport,
South Wales, UK
Department of Natural Sciences, University of
Gloucestershire, Cheltenham, UK

Overview

Until comparatively recently, most of the writing – and indeed research – on the police had focused on the situation in the United States and England and Wales. Indeed, it was only with David Bayley's (1985) seminal text that any systematic cross-national approach was adopted. This lack of an international comparative dimension was surprising, given the centrality placed on comparative analysis by sociologists and political scientists. It was also a notable omission given that senior police administrators across the world, such as Raymond Fosdick in New York and Sir Charles Jeffries, a British colonial official, had long been aware of differences in the nature of police systems between societies and used their awareness to adopt new ways of policing (Mawby 1990).

Practitioners are still involved in international comparisons. However, more of the recent contributions have been made by academics. International comparative research has become an essential ingredient of the established academic agenda of police studies. One method of so doing, of which the following entries are excellent examples, is to compare specific policing issues cross-nationally. Another is to compare different police systems. However, ironically, the very fact that police systems and policing methods are discussed so regularly though the international – academic and practitioner – police families lends weight to the argument that in postmodern societies, convergence has led to a reduction in the contrasts between different police systems. Having previously compared Anglo-American police systems with four other models, the continental, communist, colonial, and Far Eastern systems (Mawby 1990, 1999a, 2011a), the issue is revisited here. Specifically, this entry considers how far the continental, communist, and colonial systems, all examples of control-dominated systems, are still relevant to a contemporary critique of policing across the world.

Models of Alternative Police Systems

As other entries in this volume testify, there is a distinction between policing – as a process – and the police as an institution that might be responsible for many other services that are only tenuously related to maintaining order or preventing crime. Focusing on the police as a state-based organization, the nature of "the police" varies markedly between countries and over time (Bayley 1985; Findlay and Zvekic 1993; Mawby 1990, 1999a, 2011a). Distinctive models can be distinguished in terms of police legitimacy, structure, and function. Legitimacy implies that the police are granted special authority by those in power, whether this is an elite within the society, an occupying force, or the community as a whole. Structure implies that the police is organized, with some degree of specialization and with a code of practice within which, for example, the extent to which use of force is legitimate is specified. However, the extent of organization or specialization, and the types of force considered appropriate, will vary. Finally, function implies that the role of the police is concentrated on the maintenance of law and order and the prevention and detection of offenses, but there might be considerable differences in the balance between these and in the extent to which other duties are assigned to the police.

On these three criteria, the public police model that emerged on mainland Britain and the USA in the nineteenth century differed from the centralized, autocratic arm of state authority that preceded it on continental European. Equally, a control-dominated system can be identified with traditional policing in the colonies established by Britain and its European neighbors and communist Europe.

However, there are, arguably, as many differences between countries within a model as there are between alternative ideal types, a point made forcefully by Anderson and Killingray (1991, 1992) in the context of a colonial police system. Equally, there are often variations within a country. For example, in Canada marked variations exist between the centralized RCMP and local urban and provincial police.

The Canadian example is a useful one, because it challenges the assumption that the control-oriented models cited above are categorically different from Anglo-American police systems. While the latter might be identified with a democratic, community-oriented police system, the task of assigning any specific police system to a community-oriented model is contentious. Although this may be the type of democratic policing that many aspire to, and Western democracies have been keen to influence such police developments elsewhere (Bayley 2006; Marenin 1998; Pino and Wiatrowski 2006), it is difficult to nominate any one country as even approaching achieving it (Brogden 1999; Mawby 1990). This is evident if we disaggregate the core components of a community-oriented system.

Such a police system is one where the main function of the police is to provide a public service that addresses the wider needs of the community. Maintaining order is important, but the emphasis is more on crime as symptomatic of community problems than as an affront to authority. Such a model assumes that the police are accorded considerable legitimacy by local communities. The police are consequently generally organized and managed locally, and barriers between police and public are minimized. Community policing and problem-oriented policing typify this approach.

Anglo-American police clearly fall short of these ideals. In England and Wales, for example, often eulogized as the home of community policing, the modern police system emerged at least in part as a means of maintaining order in the midst of working class protest and was brutally deployed following the first World War to break the national miners' strike, a strategy revisited when Margaret Thatcher's government used the police to break the miners' strike in the early 1980s (Fine and Millar 1985). And in many cases, police were recruited from rural areas to work in the cities, undermining the claim that they were local citizens in uniform. In the USA, where police systems have been traditionally locally based, personnel have been recruited locally, and officers have engaged in a wide range of "noncrime" responsibilities; the image of the police as a militaristic body charged with fighting the "war" against crime is equally pervasive. Increased militarization of the police to "fight" the "war" against terrorism has reconfirmed this, as Belur notes in this volume. Elsewhere, Bayley's (1991) early presentation of the Japanese police as community based and welfare oriented has been questioned (Aldous 1997; Leishman 1999; Miyazawa 1992). It may, therefore, be that the key strength of specifying a community-oriented model is as an ideal type, in the Weberian sense, so as to better evaluate police systems and changes within them.

In stark contrast to a community-oriented system, a control-dominated system is one where the main function of the police is to maintain order and where the population generally fails to recognize the legitimacy of the state and its agents, the police. In such societies, the police may carry out a range of administrative tasks on behalf of the state but rarely provide a public service that addresses the welfare needs of the community. The police are, consequently, generally organized and managed centrally and have many paramilitary qualities. In some cases, the distinction between police and military is negligible.

Nevertheless, it can be argued that the democratization of postcommunist and postcolonial societies and the cross-national interchange of ideas within the international police family have resulted in police structures and methods from one country being imported to others, leading to convergence. It is this argument that is more fully discussed in this entry. The focus is on three police models that have traditionally been considered control oriented: the European continental system, communist police, and the colonial system. In each case, the original model is described before moving on to consider how specific police systems have changed and how far democratic, community-oriented policing is more evident.

The Continental Model: The Greek Police Example

Discussions of an alleged continental European policing system have a long history, and Fosdick's (1969) account of continental police at the beginning of the twentieth century is the first of many attempts to identify key characteristics of the police systems of continental Europe (Mawby 1990). In terms of function, the role of the police in continental societies has traditionally tended to be wide-ranging, with a particular emphasis upon political control, termed "high policing" in the French context, in addition to crime control. Continental systems have also been associated with a range of administrative responsibilities, with relatively less emphasis on welfare or service functions. However, there is a marked difference between the French, Italian, and Spanish police on the one hand, as classic examples of this *ideal type*, and their counterparts in countries such as the Netherlands or Scandinavia.

The need for strong policing might imply that continental police systems would also be characterized as centralized and paramilitary, but this was not always the case. For example, Iceland and Switzerland have, respectively, district and canton-based systems, and the Netherlands reorganized in 1993 into 25 regional forces (Jones 1995). More characteristic of the traditional continental model is a structure where one centralized, militaristic force is counterbalanced by either a second or by local city forces. The French are traditionally identified with this model, where the national *Gendarmerie*-covered rural areas and urban areas were policed initially by local forces and latterly by the *Police Nationale*; the maintenance of at least two police forces allowing governments to ensure that no one institution achieved too much power. While the police in most continental countries carry firearms, it is also the case that in many countries, there is at least one centralized force that evidences significantly greater militaristic qualities. In France, for example, the *Police Nationale* traditionally came under the Ministry of the Interior, whereas the *Gendarmerie* has been a military force under the Ministry of Defence, with a two-tier entry system, barrack accommodation, and impressive armaments.

In the past, continental police systems were also distinguished in terms of their lack of public accountability, being directly responsible to the head of state. While this is less easily reconciled with the liberal democracies of postwar Europe, it is still the case that public accountability is more restricted in countries where the police are more centralized and militaristic. Moreover, the move toward democracy does not inevitably bring with it a more democratic police system. The Greek situation well illustrates this (Mawby 2011b).

An assessment of the Greek police as it developed after gaining independence from the Ottoman Empire in 1830 suggests that the system was steeped in the continental European tradition. This was partly the result of external influence (especially from Britain, France, and Russia), partly the law and order problems prioritized internally.

The *Chorafylake*, formed in 1833, was modeled on the French *Gendarmerie* and charged with the task of supporting the army in protecting the fledgling state from insurrection. It remained the major policing agency for the next 150 years, adapting and being reformed to meet new challenges perceived to threaten public order. For example, it was also used to counter the emerging "threat" of communism and to assert Greek sovereignty in newly "acquired" territories (Rigakos and Papanicolaou 2003). However, the prioritization of high policing meant that the *Chorafylake* seemed unwilling or unable to deal with conventional crime problems, most notably in the expanding cities. As a result, municipal police forces were created in the late nineteenth century. They were disbanded in 1906 but then in 1920, a national force was reconstituted to cover the main cities and the island of Corfu. Known as the *Astynomia Poleon*, it was conceived as a civilian force, modeled on the London Metropolitan Police, with its own training system and a pay structure designed to present it as a more professional alternative to the *Chorafylake*. To a certain extent, the establishment of separate forces, a paramilitary *Gendarmerie* to police rural Greece and a civilian urban equivalent, paralleled developments in many other European countries. Both forces were armed and explicitly conservative and anticommunist, the latter tendency being reinforced in the immediate period after World War 2 as the police were positioned as a force against communism, both internally and externally, and specialist units, the Hellenic National Intelligence Service (KYP) and the LOK Special Forces, were created with alleged supported from the US. LOK was actively involved in the 1967 coup that led to the military dictatorship of the colonels (1967–1974) and a further shift toward paramilitary policing with the creation of the Greek Military Police. The return to a democratic system in the 1980s and the creation of a civilian police system thus emerged from the foundations of a centralized and militaristic police, mandated to uphold right-wing governments, with little or no broader accountability.

The structure of the new police was encapsulated in the 1984 Act with the *Chorafylake* and *Astynomia Poleon* merged into the new *Hellenic Police*. Unlike in postcommunist societies (see below) though, there was no dramatic change to police personnel. However, as Rigakos and Papanicolaou (2003: 286) note, this new police system was intended to become a "genuine social service." But the ensuing structure contained a hybrid mix of its forbearers, with innovations constrained by political and social expediency. Thus, while it was answerable to a new Ministry of Public Order, the *Hellenic Police* continued the centralized, militaristic tradition of the *Chorafylake*, being justified with reference to the universal mantra of a "war on crime" and the only slightly less common assumption that the military model was a safeguard against corruption.

In one sense, any justification for such a model appears weak. Thus, Greece had, and continues to have, a relatively low crime rate. The International Crime Victim Survey (ICVS), which included Greece for the first time in 2003/2004, ranked its rate for most crimes as well below average, albeit levels of fear were higher than in most other countries (van Dijk et al. 2008). It was, rather, the political threat, both communist and Turkish inspired (and more recently riots fuelled by the financial crisis), and concerns over cross-border contagion (illegal immigration and drug trafficking) that have been used to justify the need for strong policing.

Reflecting such concerns, a number of developments in the past 20 years have been aimed at toughening security responses, including:

- The reformulation of the Special Suppressive Anti-Terrorist Unit (EKAM), originally introduced in 1978, when the *Hellenic Police* was created
- The creation of the Border Guard Police Service (BGPS) in 1998 to tackle illegal immigration
- The formation of the Special Guard Service (SGS) in 1999 to protect sensitive sites from terrorist attacks
- The introduction of Criminality Prevention and Repression Squads (CPRS), with a particular focus on Romas and illegal immigrants
- The creation of the Department of Police Special Controls (DPSC) in 1995 with Special Controls Squads (SCS), mainly active in areas of social deprivation

There have, of course, been countertrends. Nevertheless, the Greece police has maintained its tradition as a powerful, centralized, and militaristic organization. It is against this backcloth that we need to contextualize the response to incidents such as the fatal shooting of Alexis Grigoropoulos in December 2008 that provoked riots both at the time and a year later and the confrontations on the streets of Greece's major cities following the financial crisis in the Spring of 2010 and subsequently. More broadly, it is illustrated in the feelings of the Greek population, expressed in the ICVS, where the Greek police rate poorly (van Dijk et al. 2008).

Moving Out of Communism: The USSR and the New Europe

The claim that there is, or was, a distinctive communist police system has also been the subject of debate. For example, it is plausible to argue that it was in many ways conceived out of the police models of continental Europe. There are also clear differences between the police systems of Russia, the "colonial power" of Central/Eastern Europe, and the People's Republic of China (PRC) (Mawby 1990). As elsewhere, then, the extent to which countries' police systems can be categorized into an ideal type is influenced by a host of variables such as social structure, prevailing culture, the influence of other countries and, in the case of communist countries, the nature of the previous regime and the distinctive form taken by the revolution.

The emergence of communism took starkly different paths in the USSR and PRC (Kowalewski 1981). While in China, decades of civil war culminated in the creation in 1949 of a rural-based popular government that was dependent upon peasant support and saw the Mass Line as a mechanism for forging conformity; in the USSR the revolution of 1917 was all but completed by 1920, leading to minority Bolshevik control of a country where grassroots support for the new regime was minimal. These conditions therefore paved the way for a highly centralized economy and state, with control exerted through a party elite and where the former Tsarist Secret Police provided a model for the new secret police, the *cheka*, which operated above the centralized militia.

Clearly a major role of the police was to deal with crime. However, the extent to which crime control was *the main role* of the police is questionable, with the police playing a key political role in maintaining the regime. The *cheka*'s mandate was to control the border and to prevent internal counterrevolutionary activity, a mandate it accepted with such ruthlessness that it was disbanded in 1922, although its replacements – the GPU, MGB, and, later, the KGB – were also counterintelligence organizations with considerably more power than the militia. Meanwhile the militia played a major part in political control through the regulation of passports and identity cards, through which it had a direct influence on restricting population movement.

The political role of the police was not the only way in which the functions of the police were wide-ranging. However, whereas in China, these included welfare and social service roles; in the USSR the emphasis was on administrative responsibilities. The centralized nature of policing was also evident. Any degree of local autonomy that existed in theory appears to have been purely fictional. Furthermore, the USSR's militia was essentially militaristic, being routinely armed with a rank structure equivalent to the military and recruits drawn from either the party or the armed forces. This acknowledgement of party membership was crucial to an understanding of the legitimacy of the police in communist societies. The interrelationship of party and police was perhaps the most significant feature of the system. In both the USSR and PRC, the police included a disproportionate number of party members, but in addition the party independently exerted considerable influence on police practices. In the USSR, with an early emphasis on a written constitution, the militia was in theory accountable to the law, although the security forces have traditionally operated above the law. However, the influence of the party on the entire criminal justice system meant that the concept of separation of powers was entirely absent. Similarly, attempts to involve the public in the policing process seemed to result in either additional party control through the creation of police aides (at best) or a complex informer system (at worst).

Clearly the emphasis changed as the balance of power shifted and new priorities gained ascendancy. For example, Lenin's preference for apportioning some power to factory committees illustrated an example of locally based crime control that was dismantled in Stalin's ruthless pursuit of a totalitarian central regime, but reemerged under Khrushchev, with the introduction of comrades' courts, the campaign against Parasites and the People's Guard.

The police systems of Warsaw pact countries were closely modeled on that of the USSR and thus unsurprisingly incorporated political and administrative responsibilities, a strong centralized secret police and a centralized, militaristic uniformed police. Being accountable to the party, there was little popular mandate. In postcommunist societies they were commonly acknowledged to have been "repressive" systems (Mawby 1999b).

It is also notable that during the mid-1980s, as these regimes faced more overt public opposition, the repressive political role of the police became *more* pronounced. In Czechoslovakia, the police were associated with brutality in the period leading up to 1989. In the GDR the police played a key role in protecting the Honiker government in the early 1980s, while in Poland the police attempted to repress *Solidarity*.

Just as political reform in the former USSR and Eastern Europe has been varied, so changes to the police have differed in scale. On the one hand, it appears that change in Russia and many of its former states has been minimal (Pustintsev 2000; Beck and Chistyakova 2002; Galeotti 2003). On the other hand, police reform in some, but by no means all, of its Central/Eastern European "colonies" was accorded priority in the embryonic periods of the new regimes. However, the move from a communist to a democratic police system has not been plain sailing in these societies in transition (Zvekic 1996, 1998).

Given the importance of the police in sustaining the "old" system, it is scarcely surprising that the new democratic regimes of Central and Eastern Europe should have prioritized changing them, and in the early days of the new regimes, changes to police systems were considered crucial (Mawby 1999b). However, while the most radical changes occurred in East Germany, where its police were absorbed into West Germany's, elsewhere changes were less radical. Changes to police *personnel* appear to have been widespread in the early phase of development. But the function, structure, and legitimacy of the police have been affected rather less. Some changes *have* taken place, but they have perhaps not been as radical as was envisaged at the time of the collapse of communism in Central and Eastern Europe. This reflects both the difficulty of radically changing established institutions and the changing priorities in postcommunist societies, where governments have identified an escalating crime problem.

Social, economic, and cultural changes added to the pressure on criminal justice systems in general and the police in particular. Economic inefficiency, collapse of living standards, and social dislocation bred a criminogenic environment, and rising levels of crime were well illustrated in official statistics. Victim survey data, while being relatively recent, also suggest that from comparatively low rates under communism crime rates in societies in transition rapidly caught up with those in the West. Certainly, public anxiety increased (Zvekic 1996, 1998; Mawby 1999b), albeit the most recent data suggest that both crime rates and public anxiety have been reduced (van Dijk et al. 2008). The ICVS also suggests that in some former Warsaw pact countries public attitudes toward the police have improved, although they still remain relatively negative.

While these factors tended to put a break on police reform, the influence of Western Europe, especially as societies in transition sought EU membership, and the USA is also important. Political alliances with the West, through Interpol, EU membership, and joint training operations have drawn postcommunist societies closer to the West, and especially closer to Western Europe and the USA (Marenin 1998). In the latter case, however, US concern to emphasize the "war" against crime, and especially terrorism and the international drug problem, has taken priority over any emphasis upon democratizing and decentralizing the police.

Colonial Policing: UK and US Colonialism

A third police system that has been consistently recognized in the literature is the colonial model.

In many respects it corresponds to the continental model – not surprising given that much of the administrative and legal structure of European states was based on earlier Roman institutions, where the Romans were themselves colonists. This also serves to remind us that the British were not the only colonists (Cole 1999). However, there has been relatively little research on the police of colonial Spain, Portugal, and France or indeed of the colonial model adopted by the USA in its "unincorporated territories," such as Puerto Rico, which came under US control in 1898 following the Spanish-American War.

Focusing on the British Empire, it is arguable that the Westminster government created a different type of police system for its colonies, one that was more appropriate for the control of a subjugated population. The model it used was the one first established for Ireland, where the police could not rely on public consent, which was then introduced, with modifications, throughout British-controlled Africa, Asia, and the Caribbean.

Colonial police may be characterized as, in terms of structure, relatively centralized and militaristic (e.g., armed and living as units in barracks), in functional terms, giving more priority to public order tasks, but also having a number of administrative responsibilities, and deriving their legitimacy from their colonial masters rather than the indigenous population. There has, however, been considerable debate over the distinctiveness of a colonial model. On the one hand, Brogden (1987) has argued that the differences between British and British colonial policing have been exaggerated. On the other hand, Anderson and Killingray (1991, 1992) claimed that the differences between colonial systems exceeded the similarities.

What is clear is that the British government saw the establishment of a strong police apparatus as central to the establishment of control and legitimacy across its empire. To enforce control, the police were formed according to a militaristic model, although not necessarily armed, nor was the police system necessarily centralized: in India the provinces or states had their own police forces although the Indian Police Service was responsible for the recruitment, training, and deployment of senior ranks; in Nigeria, at least two forces operated in different parts of the country. However, clearly the police acted on behalf of the British government and had little local mandate, frequently operating with draconian powers. A further illustration of the lack of local influence was the common practice of recruiting staff from either the British military or from elsewhere in the empire, ensuring that the police did not establish close relationships with the indigenous population. The police were part of the administrative structure, and their roles reflected this. They were involved in ensuring that local government ticked over quietly. While crime control might have been important, especially where it involved British nationals as victims, maintaining order and eliminating dissent were pivotal (Arnold 1986). The police were engaged in putting down political protest, including labor disputes, while in Hong Kong, guarding against the threat of communism was central to their mandate (Anderson and Killingray 1991; Travers and Vagg 1993). While crime and disorder issues within the indigenous community received little priority, this meant – ironically – that in some countries, such as Hong Kong, communitarian forms of self-policing were tolerated.

The fact that the British government experienced similar problems throughout its empire provided an important push toward conformity. This was strengthened, though, with centralized administration and control through the Colonial Police Service, based in London, central training for officers, the formation of the Inspector General of Colonial Police in 1948, and the practice of transferring senior officers between different countries.

However, just as there were differences in emphasis between different parts of the empire, so the model shifted at different points in time. In particular, in the conflict building up to independence, which in many cases culminated in armed insurrection, the response of the British government was to accentuate the key features of the model. So, for example, police numbers were increased, central control was strengthened, police arms were improved, and links between police and military were enhanced (Anderson and

Killingray 1992). That this should happen at a time when the legitimacy of the British government and its police was being challenged is scarcely surprising. What is equally important to stress, though, is that it makes problematic the transition from a colonial to a democratic, community-based police system. Moreover, given the inevitability of conflict and disorder after independence, the social control functions of the colonial police came to be valued by the new regimes. In consequence, the strategy, for example, in India (Arnold 1986), seems to have been to replace the (British) officer class but preserve core features of the system. Thus, even with significant political change, police systems often continued as before, an issue illustrated in Northern Ireland (Independent Commission on Policing in Northern Ireland 1999).

This is also evident in the Caribbean. In Jamaica, for example, which gained independence from Britain in 1962, the Jamaica Defence Force (JDF) continued the colonial police tradition. In the context of political unrest and corruption, a significant illegal drug trade, high homicide rates, and the establishment of the rigidly defined geographical zones known as garrisons, safe havens controlled by organized crime, paramilitary policing has been evidenced through joint police-military operations and special units such as the Mobile Reserve and the Crime Management Unit (CMU) (Harriot 2000, 2003). In Puerto Rico, despite significant political changes, high crime rates, especially homicide, and the existence of public housing projects that are virtually no-go areas, have been used to justify a continued military-style police system, symbolized through the *Mano Dura Contra el Crimen* (Hard Hand against Crime) (Dinzey-Flores 2011) and the brutal shooting by a police officer of community leader, Miguel A. Cáceres Cruz in 2007. External intervention, particularly by the USA, has been significant in many parts of the Caribbean, with funding and expertise provided in an attempt to move toward more effective, and possibly more democratic, forms of policing (Bayley 2006; Bowling 2010; Pino and Wiatrowski 2006). However, while Wilson et al. (2011) report some evidence of improvement in Trinidad and Tobago, perhaps the key message here is the difficulty of introducing significant changes, particularly in the context of internal and cross-border crime and disorder issues.

Resistance to Change: The Continuation of Control-Dominated Police Systems

The evidence discussed here suggests that control-dominated police systems are still evident across the world and that it is still relevant to talk about the continental, communist and colonial models, albeit the key features of these have been modified to varying degrees. The question then arises as to why there has, apparently, been resistance to change.

Two broad generators of change can be identified: internal pressure and external influence. Internal changes may be predicated by regime change. This is particularly well illustrated in the case of former Eastern Bloc countries and postcolonial societies but also applies to the Greek example. In each case, the transformation of the traditionally control-dominated police into a community-oriented police was commonly advocated by aspiring leaders. However, a perceived threat to law and order led to a dismantling of police reform agendas, for example, in Eastern Europe (Mawby 1999b), with subsequent reforms less radical than had been anticipated (Beck et al. 2006), while in the case of postcolonial societies, new governments also sometimes retained old police systems in order to establish and assert their authority (Anderson and Killingray 1992). The Greek and Caribbean examples also illustrate the extent to which governments' "identification" of major crime problems and political threats has been used to justify the continuation of a militaristic police.

The broader processes leading up to regime change are illustrated in Fig. 1. During the death throes of colonial and communist societies, regimes under threat acted to preserve the status quo by enhancing the militaristic features of the police, for example, by police-military cooperative ventures, the creation of elite, specialist public order police squads, and increases to police

	Continental	*Communist*	*Colonial*	*Community policing*
Function	Control	Control	Control	Service
Structure	Militaristic Centre/Local	Militaristic Centralised	Militaristic Varied	Civilian Local
Legitimacy	Central government	Party	Occupying power	Local Community

Comparing Police Systems Across the World, Fig. 1 Police systems in transition: the problems of changing the police

hardware. The result was that such systems contrasted even more with a community-oriented system. The possibility of a steady transition from control to community oriented thus faced even more obstacles.

In this situation, the importance of external influences becomes crucial. External influence is particularly important in postmodern societies where similar influences are prevalent across national boundaries and where examples of innovative developments in one society are readily available as examples of best practice elsewhere. Formal pacts add a further impetus to change. For example, the emergence and expansion of the European Union has involved greater cross-border cooperation and, consequently, increased pressure toward the harmonization of policy. Allied to this, Europol was ratified in 1999 as the EU organization for cross-border coordination between national law enforcement agencies, providing collation, analysis, and dissemination of information, and a European Police College (CEPOL) was established in 2001. A further source of external pressure involves Western democracies, with their alleged democratic police systems, offering a supportive role in the transformation. As noted above, the USA has been influential in the Caribbean, while it also founded a police college in Budapest, providing training for key staff from societies in transition. The UK has fulfilled a similar role, establishing the Knowhow Foundation to provide support in its former colonies and later extending the scheme to Central and Eastern Europe.

However, it is easy to identify tensions here without acceding to Brogden’s (1987) dismissals of the differences between Anglo-American and control-dominated systems. Western democracies have a long history of sacrificing democratic principles when other interests are perceived as paramount, and this equally applies to police reform (Bayley 2006; Pino and Wiatrowski 2006). One aspect of this relates to organized cross-border crime, including drug trafficking, people smuggling, and money laundering, where the “war” against crime may be advocated in advance of democratic policing ideals. Examples of this are evident in the content of US training priorities at its Budapest college and in US policy in the Caribbean. Of even more significance is the *political* threat. As Aldous (1997) demonstrated for postwar Japan, US policy oscillated between prioritizing the democratization of the police through decentralization and encouraging a strong police as barrier against the communist threat, the latter point being reiterated in South Korea (Lee 1990). Similar fears concerning the spread of communism within Europe can be identified vis a vis US support for hard-line policing in Greece. In the light of such examples, the influence of US policy on the democratization of Islamic states in the aftermath of the Arab Spring should be greeted with caution. While internal pressures toward democratic policing are by no means inevitable, external influences are often ambiguous. The assumption that police systems across the world are inevitably converging toward a democratic, community-oriented ideal is thus highly questionable. Variations between the police systems of different countries are still evident, with the continental, communist, and colonial models still relevant for any analysis of variations between the police systems of different countries.

Related Entries

- ▶ British Police
- ▶ Causes of Police Legitimacy
- ▶ Communities and the Police
- ▶ Community Policing
- ▶ Cross-National Performance in Policing
- ▶ Democratic Policing
- ▶ Dutch Colonial Police
- ▶ French Colonial Police
- ▶ Gendarmerie Policing
- ▶ German Police Until 1918
- ▶ National and Local Policing
- ▶ Order Maintenance Policing
- ▶ Police Use of Firearms
- ▶ Policing of Peacekeeping
- ▶ Police in the Police State
- ▶ Policing in Developing Democracies
- ▶ Privatization of Policing in an International Context
- ▶ Role and Function of the Police
- ▶ Strategies of Policing Terrorism
- ▶ Women in Policing

Recommended Reading and References

Aldous C (1997) The police in occupation Japan. Routledge, London

Anderson DM, Killingray D (eds) (1991) Policing the empire. Manchester University Press, Manchester

Anderson DM, Killingray D (eds) (1992) Policing and decolonisation. Manchester University Press, Manchester

Arnold D (1986) Police power and colonial rule: Madras 1859–1947. Oxford University Press, Oxford

Bayley DH (1985) Patterns of policing. Princetown University Press, New Brunswick

Bayley DH (1991) Forces of order: policing modern Japan. University of California Press, Berkeley

Bayley DH (2006) Changing the guard: developing democratic police abroad. Oxford University Press, New York

Beck A, Chistyakova Y (2002) Crime and policing in post-Soviet societies: bridging the police/public divide. Policing Soc 12:123–137

Beck A, Chistyakova Y, Robertson A (2006) Police reform in post-soviet societies. Abingdon: Routledge

Bowling B (2010) Policing the Caribbean: transnational security cooperation in practice. Oxford University Press, Oxford

Brogden M (1987) The emergence of the police: the colonial dimension. Brit J Criminol 27:4–14

Brogden M (1999) Community policing as cherry pie. In: Mawby RI (ed) Policing across the world: issues for the twenty-first century. UCL Press, London, pp 167–186

Cole B (1999) Post-colonial systems. In: Mawby RI (ed) Policing across the world: issues for the twenty-first century. UCL Press, London, pp 88–108

van Dijk J, van Kesteren J, Smit P (2008) Criminal victimisation in international perspective: key findings from the 2004–2005 ICVS and EU ICS. The Hague: Boom Legal Publishers (accessed at http://rechten.uvt.nl/icvs/pdffiles/ICVS2004_05.pdf)

Dinzey-Flores ZZ (2011) Criminalizing communities of poor, dark women in the Caribbean: the fight against crime through Puerto Rico's public housing. Crime Prev Community Saf 13:53–73

Findlay M, Zvekic U (eds) (1993) Alternative policing styles: cross-cultural perspectives. Kluwer, Boston

Fine B, Millar R (eds) (1985) Policing the miners' strike. Lawrence & Wishart, London

Fosdick RB (1969) European police systems, Revisedth edn. Patterson Smith, Montclair

Galeotti M (2003) Russian police reform: centralisation, paramilitarisation and modernisation. Crime Justice Int 19(70):17–19

Harriot A (2000) Police and crime control in Jamaica: problems of reforming ex-colonial constabularies. The University of the West Indies Press, Kingston

Harriot A (ed) (2003) Understanding crime in Jamaica: new challenges for public policy. The University of the West Indies Press, Kingston

Independent Commission on Policing in Northern Ireland (1999) A new beginning– policing in Northern Ireland (Patten Report). Independent Commission on Policing in Northern Ireland, Belfast (see also www.belfast.org.uk/ report.htm)

Jones T (1995) Policing and democracy in the Netherlands. Policy Studies Institute, London

Kowalewski D (1981) China and the Soviet Union: a comparative model for analysis. Stud Comp Communism 14:279–306

Lee SY (1990) Morning calm, rising sun: national character and policing in South Korea and in Japan. Police Stud 13:91–110

Leishman F (1999) Policing in Japan: east Asian archetype? In: Mawby RI (ed) Policing across the world: issues for the twenty-first century. UCL Press, London, pp 109–125

Marenin O (1998) United States police assistance to emerging democracies. Policing Soc 8:153–167

Mawby RI (1990) Comparative policing issues: the British and American experience in international perspective. Routledge, London

Mawby RI (ed) (1999a) Policing across the world: issues for the twenty-first century. UCL Press, London

Mawby RI (1999b) The changing face of policing in central and eastern Europe. Int J Police Sci Manag 2:199–216

Mawby RI (2011a) World policing models. In: Natarajan M (ed) International crime and justice. Cambridge University Press, New York, pp 409–415
Mawby RI (2011b) The Greek police in context. In: Cheliotis LK, Xenakis S (eds) Crime and punishment in contemporary Greece: international comparative perspectives. Peter Lang, Oxford, pp 491–503
Miyazawa S (1992) Policing in Japan: a study on making crime. State University of New York Press, New York
Pino N, Wiatrowski MD (eds) (2006) Democratic policing in transitional and developing countries. Ashgate, Burlington
Pustintsev B (2000) Police reform in Russia: obstacles and opportunities. Policing Soc 10:79–90
Rigakos GS, Papanicolaou G (2003) The political economy of Greek policing: between Neo-Liberalism and the Sovereign state. Policing Soc 13:271–304
Travers H, Vagg J (1993) Crime and justice in Hong Kong. Oxford University Press, New York
Wilson DB, Parks RB, Mastrofski SD (2011) The impact of police reform on communities of Trinidad and Tobago. J Exp Criminol 7:375–405
Zvekic U (1996) Policing and attitudes toward police in countries in transition. In: Pagon M (ed) Policing in Central and Eastern Europe: comparing firsthand knowledge with experience from the west. College of Police and Security Studies, Ljubljana, pp 45–59
Zvekic U (1998) Criminal victimisation in countries in transition. UNICRI, Rome

Compliance and Corporate Crime Control

Wim Huisman
School of Criminology, VU University
Amsterdam, Amsterdam, The Netherlands

Synonyms

Corporate crime control; Regulatory compliance; White-collar crime

Overview

Legal definitions of economic crime have long been in dispute. Formal prosecution of a crime needs an (individual) accused and a legally defined deviant act. Common parlance and media, however, speak of "economic crime" when they refer to corruption, financial manipulation, or fraudulent enrichment. As a reflection of typical status of the offenders, their crimes are dubbed "white-collar crime." Moreover, the media often blames corporations and public agencies rather than individuals. Braithwaite (1984: 6) describes corporate crime as "the conduct of a corporation, or of employees acting on behalf of a corporation, which is proscribed and punishable by law." Corporate crime, therefore, encompasses a wide array of illegal activities that are criminally, civilly, and administratively proscribed and which may be committed by individual managers as well as by the firm they work for. Thus, modern regulations have introduced fines and conditional exclusion (e.g., from competition) for well-defined organizational misconduct.

Additionally, enterprises may try to control misconduct by internal self-regulation and may penalize deviant employees and managers, using internal sanctions. Their regulatory definition of "misconduct" need not be restricted to criminal deviance but may include all sorts of agency (particularly company) activities which can lead to social damage and to public shaming (such as for health risks, environmental pollution, etc.). Consequently, corporate misconduct can be considered to something of a sliding scale of wrongdoing: crime, punishable by state sanction to organizational misconduct inspected by internal compliance management of corporations and other agencies. It is obvious that this scale ranges from "hard" criminal law with strict procedural rules (such as formal burden of proof and the right to defense) to "soft" regulation with managerial techniques of control. These concepts of "hard" and "soft," however, do not indicate the severity of imposed sanctions but rather the legal definition of the rules and the procedures applied to control them. As is well known, the threat of hard sanctions, particularly if the misconduct is hard to prove, regularly leads to plea-bargaining or an out of court for a settlement (Cullen et al. 2006).

It follows from these definitions that there is quite some overlap of criminological research on corporate crime with business studies on compliance with regulation in corporations and other organizations. Corporate crime theories combine insights from criminology and organizational sciences. The leading paradigm blames organizations' environments and structural characteristics for inducing managers and employees to commit crimes in a business context while otherwise behaving as law-abiding citizens. Prevention measures as well as repressive intervention focus on corporate strategy, structure, and culture to identify and change crime-fostering organizational conditions. Recently, however, there is an academic trend toward analyzing personality traits and criminal careers of corporate actors. Research on the frequency and causes of regulatory (non) compliance and corporate crime has to consider criminal justice interventions as well as corporate compliance management.

Research on Corporate Compliance

The dependent variable in studying compliance with corporate regulation may be seen on a gliding scale from the "hard" side of criminology which labels rule violations as "corporate crime" or "white-collar crime" to the "soft" side of political science and management which look at obedience with increasing mass of internal and external rules. As a result, the study of regulatory (non)compliance is rich and diverse, applying various methodological and theoretical approaches.

Parker and Nielsen (2011) distinguish two methodological approaches. The first is straightforward testing of theories which identify organizational characteristics and external factors that are associated with (non)compliance. This approach is concerned with measuring the prevalence and levels of (non)compliance as dependent variable. The explaining variables are typically individual motivation, organizational characteristics, and environmental factors such as industry culture and the regulation itself. The outcomes of such rigorous empirical studies are often used to design normative models for better regulation, which are expected to produce higher levels of compliance. Such "evidence-based" regulatory policies rely on explanatory theory building about causes of compliance and noncompliance.

The second approach aims at the social construction of the very notions of compliance and noncompliance. It tries to uncover multiple interpretations and meanings of compliance together with the power relations that lead to more or less social acceptance or denial of legitimacy. They use discourse analyses and empirical observation of interactions which define, maintain, or change notions of compliance. For instance, in her study on the enforcement of occupational safety and health regulation in the United Kingdom, Hutter concluded that defining, achieving, and maintaining compliance are an interactive process between regulator and regulated (Hutter 1997).

Both approaches complement each other. Internally, corporations use the term "compliance" for activities by which they try to ensure internal regulations, designing procedures and employ compliance officers. Externally, "compliance" stands for a style of regulation and enforcement, as opposed to a "deterrent" enforcement style.

Prevalence and Causes of (Non) Compliance and Corporate Crime

The Prevalence of Compliance and Noncompliance

In criminology, studies of new forms crime usually start with assessments of the scale in terms of the prevalence of offenses and the level of damage. Edwin Sutherland (1949/1983) is seen as the academic founding father of crimes committed in the context of otherwise legitimate business. He coined the term "white-collar crime" when he conducted a first study on regulatory noncompliance by America's 70 largest corporations. He found sanctions for violations of government regulation in all of them, 97 % were recidivist and 60 % he even labeled as "habitual" criminals with more than four violations. Clinard and Yeager (1980) replicated this study by measuring noncompliance by America's 500 largest corporations. They found widespread violation within several branches of industry with a concentration on a small number of habituals: 13 % of the offenders were responsible for about half of all violations.

It is extremely difficult to estimate the prevalence of corporate crime or corporate noncompliance. Since organizational crime is generally less visible and harder to define, the "dark numbers" are even higher than with ordinary crimes. Official statistics direct much more attention to conventional crime than to white-collar crime. Victims of manipulation, fraud, and embezzlement often do not detect their losses, and even if they become aware, may be reluctant to file reports due to fear of negative publicity. The police, customs, and other agencies have to handle variable or ambiguous legal definitions so that corporate crime is not uniformly defined, reported or recorded. Specific forms of business regulation are dispersed over several agencies which not always coordinate their enforcement (e.g., environmental regulation, occupational safety regulation, trade regulation, consumer protection regulation, etc.) Each of the numerous agencies may have its own forms of record keeping reflecting the expertise and priorities of the regulated field, but not reflecting the actual prevalence of noncompliance. Nevertheless, registrations of these enforcement agencies show that corporate rule breaking is widespread in many areas of regulation and branches of industry.

Motivation for Compliance and Noncompliance

Criminological theories on the causes of criminal behavior look at two types of explanatory variables: *motivation* – why people *want* to commit crimes, and *opportunity* – why people *are able* to commit crimes. The same variables can be found at the positive end of the compliance spectrum. Business regulation requires corporations to take certain action; corporations also have to be willing and able to comply. Thus, corporate deviance is often due to omission: Corporations violate rules as a result of failure to act in circumstances where that is necessary, e.g., financial reporting, installing prescribed equipment, or avoiding environmental pollution.

There is some debate as to whether to look for motives at the level of the organization or individual managers and employees. Managers' social programming tends to the same action in same situations (Coleman 1987), and thus, their motives are identified with the goals of the organization as a whole. Some scholars even see the motivation of individuals as "irrelevant" in explaining corporate crime (Braithwaite 1984; Coleman 1987).

At the firm level, three types of motives for compliance and noncompliance are distinguished: economic, social, and normative motives. Economic motives refer to the extent to which the firm is committed to maximizing its own economic or material utility, such as increasing turnover and profit. Social motives refer to the extent to which a firm is committed to earning the approval and respect of significant people with whom an actor interacts including other businesses, trading partners, employees, customers, local communities, and the wider public. Normative motives refer to the extent to which the firm is committed to obeying the regulation for its own sake because of a sense of moral agreement with the specific regulation or a generalized sense of moral duty to comply. Empirical research has made it clear that there are a wide range of relevant motives and they interact with each other in reciprocal ways: Personal motives of managers are extrapolated to the motivations of the organization as a whole, and on the other hand, these shape the motivations of individual managers (Parker and Nielsen 2011).

The financial crisis and large accounting fraud cases have renewed the attention for personality traits that influence individuals' motives for corporate crime. Large financial accounting frauds such as ENRON have been related to the personality traits of corporate executives. Public, political as well as academic debate on the causes of the financial crisis of 2008 have indicated the roots of the current credit crunch to be the result of excessive risk taking, moralistically referring to "greed" (Huisman 2011). Research has even linked masculinity to risk-taking financial markets (Coates and Herbert 2010). To prevent excessive risk taking and to change “masculine” corporate cultures, new laws prescribe certain quota of the members of corporate executive boards to be female. In recent studies, several personality traits are related to white-collar crime, such as the need for control, bullying, charisma, fear of falling or failing, company ambition, lack of integrity, narcism, and a lack of social conscience (Blickle et al. 2006; Bucy et al. 2008; Piquero et al. 2010; Ragatz and Fremouw 2010).

Opportunities and Organizational Capabilities

Opportunities for crime can be identified as means to commit crime as well as the lack of mechanism preventing it. While the general perception of organizational deviance sees it in the light of intention and economic motives, empirical studies have identified less serious causes such as widespread *routine*, and other types of organizational shortcomings (Van de Bunt and Huisman 2007; Vaughan 2011). Motivation to comply is of secondary importance if a firm does not possess the capacity to comply (Parker and Nielsen 2011). In this view, corporate rule breaking is the outcome of *bounded rationality*. Managers do not possess all the relevant information on the cost and benefits of options for action and organizational procedures lead to suboptimal outcomes. Landmark cases of organizational deviance – such as the Ford Pinto and the NASA Challenger – have falsified the initial explanation of these cases as being the result of rational and amoral calculation on the basis of cost-benefit assessments. Instead, the accidents were framed as being the result of organizational failure.

Most studies focus on the meso-level and try to identify criminogenic organizational strategies, structures, and cultures. These studies show that corporations set their ambitions high while tolerating poor practice regarding the means by which these goals should be attained (Wang and Holtfreter 2012). This dichotomy of ends and means varies with the amount of autonomy of departments. However, lack of internal control shows up in diverse organizational structures: In highly bureaucratic structures, managerial responsibility for shop-floor level misconduct tends to be diffused over several hierarchical layers, while in decentralized – *loosely coupled* – organizational structures the, more autonomous, subsidiaries might escape the monitoring of the parent company (Keane 1995). Organizational cultures may transmit rule-transgression by “normalizing” lawbreaking by neutralization and rationalization (Shover and Hochstetler 2002).

Multilevel Framework

Several authors brought the explanations together in an integrated framework of three categories of explanatory variables: motivation, opportunity, and (lack of) control (Coleman 1987; Shover and Bryant 1993). Integrative theories on the causes of corporate crime stress the interaction of variables on the institutional macro-level, the organizational meso-level, and the individual micro-level. Kramer and Michalowski (2006) place the three variables on three levels in a matrix with nine cells. To give some examples, on the *macro*-level, *motivation* is operationalized by the culture of competition characteristic for business, and that on the *meso*-level poses economic pressure on corporations to set ambitious targets which on the *micro*-level can provide individual managers with neutralizations to break rules to attain these goals. On the *macro*-level, the structure of the market determines the distribution of legitimate and illegitimate means to accomplish goals, while at the *meso*-level, the *opportunities* to attain goals are determined by the organizational structure. On *micro*-level, this can result in individual employees experiencing illegal means as the most attractive option for achieving goals. At the *macro*-level, the element of *control* points to the lack of administrative, political, and social control on the conduct of. If at the *meso*-level, the internal control structure is ineffective, it will result in an organization where on the *micro*-level, managers and employees rationalize illegal means for attaining organizational goals.

Corporate Crime Control

Academic models have suggested various strategies for monitoring compliance and controlling corporate crime. Some of these are used in practice. Generally, they distinguish two ideal-types of corporate crime control: a deterrence strategy and a compliance strategy. Alternative models develop integrative strategies. These models are based on assumption about the causes of compliance and noncompliance, as discussed in the previous section.

Deterrence Strategies

The deterrence style is referred to as "command and control": Regulations are drafted by government agencies and imposed on business, and compliance is strictly monitored. Regulatory enforcement is concentrated on detecting and sanctioning violations. It is based on the assumption that compliance and violation are led by economic motives. Companies are seen as rational actors calculating costs and benefits of compliance and noncompliance. A deterrence strategy assumes that offenders will refrain from future offending, if detection and punishment occur with sufficient frequency and severity. However, the strategy is counterproductive when it takes away the voluntary motivation to comply. Simpson (2002), who spent much of her academic career testing the corporate deterrence theory, concluded that we know very little about whether, when, and why deterrence might work as a strategy for compliance enforcement. Compliance decisions are not determined by objective cost and benefits of compliance and noncompliance, but rather by the subjective perceptions of corporate managers. However, there is hardly any empirical evidence that deterrent enforcement styles succeed in influencing this perception. Instead, industry and organizational characteristics have stronger effects on compliance (Simpson 2002).

The application of civil or administrative law for regulatory enforcement can have punitive or even incapacitating business effects, for instance, issuing administrative fines or revoking licenses. Nevertheless, the deterrence style is mainly associated with criminal law. In practice, however, criminal prosecution of corporate offenders is scarce. Most cases of corporate crime are not prosecuted or at best, they lead to a financial settlement. The reasons for lack of priority compared to street crime are multilayered: Specialized expertise is necessary due to the technical complexity of cases of white-collar crime; there are coordination problems due to the many agencies involved in the enforcement of economic laws; white-collar offenders and large corporations might be able to hire better lawyers capable of finding weaknesses in the prosecution's case; and finally because of legal difficulties, such as establishing who is liable in corporate entities (proving *mens rea)* and as well as the legal ambiguity of many regulations aimed at businesses. Nevertheless, heightened public "blaming" has increasingly led to prosecution and sanction in spectacular cases such as accounting fraud at ENRON, the Ponzi scheme of Madoff, and the large-scale bribery at Siemens.

The deterrence strategy also assumes that potential violators will be deterred by the punishment of offenders. In a landmark study, Thornton et al. (2005) found limited support for this assumption. In a survey of 233 American firms, it was examined whether companies learn about cases of severe penalties, in the same industry, and whether this knowledge changes the firms' compliance-related behavior. They found that less than half of the companies could identify the "signal cases" of severe legal penalties. They concluded that "explicit general deterrence" does not enhance the perceived threat of legal punishment, but it may serve as a reminder to check on reliability of internal compliance routines and as a reassurance that compliance is not foolish.

Due to the difficulties of directly influencing compliance-related behavior by legal penalties, contemporary regulators have set their hopes on reputation damage as a result of the "naming and shaming" effect of publicly sanctioning companies. Substantial reputation damage might result from publicly shaming corporate misconduct, even though there is not much knowledge about the impact in terms of financial and social reputation damage for business (Van Erp 2013).

Compliance Strategies

While official rhetoric is tough on tackling corporate crime and noncompliance, empirical studies show that most agencies regulating business apply enforcement styles that rely on cooperation rather than confrontation; they use advice and negotiation rather than coercion and formal legal action. "Compliance assistance" is the preferred official enforcement strategy of several regulatory agencies (Van de Bunt and Huisman 2007).

Where violation results from organizational shortcomings, negotiation and consultancy as enforcement style can be used to teach a corporation how to comply with regulations. Cooperative regulation is used to stimulate companies to maintain internal compliance management systems. Compliance management systems are viewed as important tools for companies to organize and monitor regulatory compliance.

This might possibly lead to the introduction of "meta-regulation" at a distance: Instead of a "rule-based" approach with detailed prescriptions for business operations, the government creates "principle-based" regulation, containing broad duty-of-care provisions and the obligation for businesses to maintain compliance management systems. This is the case when regulation no longer dictates with what kind of equipment a certain level of emission should be reduced but prescribes a compliance system that ensures a sustainable business operation. Enforcement is then based on reports provided by the companies themselves, and it is aimed at ascertaining whether compliance mechanisms are working effectively. However, such an approach may prove counterproductive in the case of intentional violation (Gunningham 2011).

From the rationality paradigm, lenient enforcement and negotiation were initially rejected. Academic studies on the causes of noncompliance have given legitimacy to addressing causes of noncompliance: organizational incompetence and corporate cultures failing to reinforce the value of compliance.

Whether compliance management systems have had a positive impact on noncompliance has to date received little attention in empirical research. Compliance programs create awareness and give instructions on how to comply. Kaptein (2011) found a positive effect of ethical codes of conduct on compliance. It was observed, however, that the effectiveness of corporate compliance officers could be compromised when they report to the business line management, rather than the corporate board. Compliance programs could even be considered to be more cosmetic than real, largely serving a PR function. In some cases, they may actually increase or facilitate corporate offenses, by regulatory capture or keeping up appearances (Friedrichs 2010).

Integrative Models

The criticism of different styles of regulation and enforcement is as old as the propagation of "new," "smarter," and more "responsive" models. Academics who suggest these models try to find an optimal mix of compliance and deterrence. A currently adopted model is that of "responsive regulation" (Ayres and Braithwaite 1992). The central notion of "responsive regulation" is that enforcement styles should be tuned to the motives and competences of individual companies laid down in an "enforcement pyramid." Regulators should start at the bottom of the pyramid, assuming that a company is willing to comply voluntarily. When this is shown not to work, regulators should escalate up to deterrence-oriented strategies. The regulatory pyramid has become the paradigm for a "sequential" combination of compliance and deterrence styles of enforcement. It has been adopted by regulatory and enforcement agencies worldwide.

In academic debate, the model has had a mixed reception. Empirical studies on its use show that it is very difficult to apply the entire pyramid in practice. It is hard to assess the actual motives of companies, it is difficult to shift between enforcement styles, and it ignores other environmental influences on compliance than state-regulation (Gunningham 2011).

As a result of the above insights, new and more sophisticated models have evolved such as "smart regulation": adopting the principles of responsive regulation, but expanding regulation to third parties as surrogate regulators. As all modern regulatory theories, smart regulation advocates government to facilitate and mobilize other network members rather than direct intervention. Smart regulation has to organize public enforcement so that it can reinforce third parties to help with regulating (Gunningham and Grabosky 1998).

Situational Crime Prevention

Responsive models seem to be primarily oriented at influencing the motivation behind compliance and noncompliance. The bottom of the enforcement pyramid aims at normative motives for compliance (the general attitude that one should comply with regulation), the next step aims at social motives (the realization that others also comply), and the top is aimed at economic motives (increasing the costs of violation).

As motivation has proven difficult to measure, criminology has shifted the focus on studying and changing opportunities. An opportunity perspective focuses on the specific characteristics of a situation, and studies the process through which an offense is committed. In other words, it studies the "how," rather than the "why." Situational crime prevention addresses the environments which shape crime opportunities, and the modifications which may diminish criminal opportunities. It identifies five characteristics of criminal opportunity: effort required to carry out the offense; perceived risk of detection; reward from committing the offense situational conditions that may encourage criminal action; and excuses and neutralizations of the offense. Preventive measures are best found in the immediate environment of the crime. Crime prevention, then, becomes a matter of five principles: increasing the effort and risk of detection, reducing circumstantial provocation, and reward and removing excuses.

Situational crime prevention theory departs from the assumption that potential offenders decide whether or not to engage in crime, based on the attractiveness of criminal opportunities, in terms of costs and benefits of crime. White-collar crimes may be committed without clear criminal intent, since regulation is often ambiguous. Nevertheless, offenders are assumed to make more or less rational choices to commit crime, or at least, are to some degree sensitive to the changes in their environment which influence costs and benefits.

Benson and Madensen (2007) propose situational crime prevention theory, which has been widely used to analyze traditional forms of street crime, to white-collar crime. They argue that the analysis of the opportunity structure of white-collar crime through situational crime prevention theory can provide useful starting points for prevention. In fact, they argue that an opportunity perspective is *more* fruitful than an approach based on offender characteristics or motive: "we suggest that focusing on how is likely to be more productive than focusing on why" (Benson et al. 2009, p. 176). They suggest that crime prevention by means of the alteration of opportunity structures represents "a more fundamental way of thinking about the problem of white-collar crime control" than those of competing schools of thought, in the sense that it "implicitly underlies" other approaches (Benson and Madensen 2007, p. 623). Furthermore, this analysis leads to newer and more effective methods of prevention and control, because potential points of intervention are illuminated. Consequently, they urge white-collar crime scholars to apply the situational prevention approach, and have made the first step with a brief analysis of opportunity structures in health care fraud. Others have tried to apply situational crime prevention to environmental crimes (Dorn et al. 2007). Considering the strong claims by Benson et al., more empirical research is needed to test the usefulness of situational crime prevention theory for prevention of particular types of white-collar crime.

Future Developments and Directions for Research

Most of the relevant variables in explaining corporate compliance and noncompliance have been identified in different strings of research, often in critical response to one another. The main and current research question is how all these relevant variables interact in producing compliance and noncompliance. New and more advanced research designs are needed to unravel the existing process of interaction. It has to integrate measurement by objective indicators which try to establish the causal relations with independent variables focusing on hypotheses about what produces compliance on the one hand, with interpretative approaches which appreciate that compliance and noncompliance are defined in processes of social construction on the other hand. The ambiguities in the definition of compliance could be one of the variables in explaining noncompliance.

Research on the regulation of business and regulatory enforcement has long moved beyond the dichotomy of compliance and deterrence. Scholars are still searching for the optimal mix of enforcement styles. The two-dimensional problem has been expanded to a three-dimensional challenge, surpassing another dichotomy – the dyadic relationship between regulator and the regulated – and by introducing third party regulators. Harnessing actors and resources outside the public sector also fits in situational crime prevention.

Yet, although these more sophisticated models are based on empirical insights as to compliance and noncompliance, the models themselves still have to be tested in empirical research.

Related Entries

- ▶ Corporate Crime Decision-Making
- ▶ Corporate Crimes and the Business Cycle
- ▶ Corporate Liability
- ▶ Crimes of Globalization
- ▶ Deterrence of Tax Evasion
- ▶ Environmental Regulation and Law Enforcement
- ▶ Managerial Court Culture
- ▶ Naming and Shaming of Corporate Offenders
- ▶ State-Corporate Crime
- ▶ Victims of Corporate Crime
- ▶ Women and White-Collar Crime

Recommended Reading and References

Ayres I, Braithwaite J (1992) Responsive regulation: transcending the deregulation debate. Oxford University Press, Oxford

Benson ML, Madensen TD (2007) Situational crime prevention and white-collar crime. In: Geis G, Pontell H (eds) International handbook of white collar crime. Springer, New York, pp 309–326

Benson ML, Madensen TD, Eck JE (2009) White-collar crime from an opportunity perspective. In: Simpson SS, Weisburd D (eds) The criminology of white-collar crime. Springer, New York, pp 175–193

Blickle G, Schlegel A, Fassbender P, Klein U (2006) Some personality correlates of business white-collar crime. Appl Psychol Int Rev 55(2):220–233

Braithwaite J (1984) Corporate crime in the pharmaceutical industry. Routledge and Kegan Paul, London

Bucy PH, Formby EP, Raspanti MS, Rooney KE (2008) "Why do they do it?" The motives, mores, and character of white-collar criminal. St John's Law Rev 82:401–571

Clinard MB, Yeager PC (1980) Corporate crime. Free Press, New York

Coates J, Herbert J (2010) Endogenous steroids and financial risk taking on a London trading floor. Proc Natl Acad Sci 105(16):6167–6172

Coleman JW (1987) Toward an integrated theory of white-collar crime. Am J Sociol 93(2):406–439

Cullen F, Cavender G, Maakestad W, Benson M (2006) Corporate crime under attack: the fight to criminalize business violence, 2nd edn. LexisNexis Matthew Bender/Anderson Publishing Company, Newark

Dorn N, Van Daele S, Vander Beken T (2007) Reducing vulnerabilities to crime of the European waste management industry: the research base and the prospects for policy. Eur J Crime Crim Law Crim Just 15(1):23–36

Friedrichs DO (2010) Trusted criminals: white collar crime in contemporary society. Wadsworth Cengage Learning, Belmont

Gunningham N (2011) Strategizing compliance and enforcement: responsive regulation and beyond. In: Parker C, Nielsen V (eds) Explaining compliance: business responses to regulation. Edward Elgar, Cheltenham, pp 199–221

Gunningham N, Grabosky P (1998) Smart regulation: designing environmental policy. Oxford University Press, Oxford

Huisman W (2011) Corporate crime and crisis: causation scenarios. In: Sociology of crime, law and deviance, Economic crisis and crime. Emerald/The JAI Press, Bingley, UK, pp 107–125

Hutter BM (1997) Compliance: regulation and environment. Clarendon, Oxford
Kaptein SP (2011) Towards effective codes: testing the relationship with unethical behavior. J Bus Eth 99(2):233–251
Keane C (1995) Loosely coupled systems and unlawful behaviour: organization theory and corporate crime. In: Pearce F, Snider L (eds) Corporate crime; contemporary debates. University of Toronto Press, Toronto, pp 168–181
Kramer RC, Michalowski RJ (2006) The original formulation. In: Michalowski RJ, Kramer RC (eds) State-corporate crime. Wrongdoing at the intersection of business & government. Rutgers University Press, New Brunswick, pp 18–26
Parker C, Nielsen V (2011) Explaining compliance: business responses to regulation. Edward Elgar, Cheltenham
Piquero NL, Schoepfer A, Langton L (2010) Completely out of control or the desire to be in complete control? An examination of how low self-control and the desire for-control relate to corporate offending. Crime Delinq 56(4):627–647
Ragatz L, Fremouw W (2010) A critical examination of research on the psychological profiles of white-collar criminals. J Forensic Psychol Pract 10:373–402
Shover N, Bryant KM (1993) Theoretical explanations of corporate crime. In: Blankenship MB (ed) Understanding corporate criminality. Garland Publishing, New York, pp 141–176
Shover N, Hochstetler A (2002) Cultural explanation and organizational crime. Crime Law Soc Change 37:1–18
Simpson SS (2002) Corporate crime, law and social control. Cambridge University Press, Cambridge
Sutherland EH (1949) White collar crime. Dryden Press, New York
Sutherland EH (1983) White collar crime. The uncut version. Yale University Press, New Haven
Thornton D, Gunningham N, Kagan R (2005) General deterrence and corporate environmental behavior. Law & Policy 27:262–288
Van de Bunt HG, Huisman W (2007) Organizational crime in the Netherlands. Crime Just Rev Res 35:217–260
Van Erp J (2013) Naming and shaming. In: Encyclopedia of criminology and criminal justice. Springer, New York
Vaughan D (2002) Criminology and the sociology of organizations. Analogy, comparative social organization, and general theory. Crime Law Soc Change 37:117–136
Vaughan D (2011) Theorizing the dark side. In: Linstead S, Marechal G, Griffin R (eds) Organization studies, special issue on the dark side of organizations. SAGE publications, Thousand Oaks
Wang X, Holtfreter K (2012) The Effects of corporation- and industry-level strain and opportunity on corporate crime. J Res Crime Delinq 49(2):151–185

Compstat

James J. Willis
Department of Criminology, Law and Society,
George Mason University, Fairfax, VA, USA

Overview

Since its implementation in 1994 in the United States under former Commissioner William Bratton and Deputy Commissioner, Jack Maple, of the New York City Police Department, Compstat has become widely recognized as a major innovation in US policing and abroad. Compstat is a strategic management system whose reform elements are designed to overcome some of the traditional constraints and limitations of bureaucratic administration and make police organizations more responsive to changes in their crime environment. According to its doctrine, Compstat decentralizes decision-making to middle managers operating out of districts, holds these managers strictly accountable for their performance, and increases a police organization's capacity to identify, understand, and respond to crime problems as they emerge. The evidence that Compstat reduces crime is not clear and consistent, but its originators and proponents credit Compstat with impressive crime drops and improvements in neighborhood quality of life in New York City. Based on the publicity it has received, including a 1996 Innovations in American Policing Award sponsored by the Ford Foundation and Harvard University's John F. Kennedy School of Government, other police leaders, delegates, and politicians have flocked to New York City to observe Compstat in action. Impressed by what they have seen, they have implemented Compstat in their own agencies. Consequently Compstat has become a global commodity and many countries, including England, Australia, and Canada, have since adapted methods of assessing performance from the NYPD Compstat model. Research on Compstat's relationship to another popular reform, community policing, indicates that they may work independently when implemented in the same police organization: each having little effect on the other. This suggests opportunities for their integration, but what form such a model would take is uncertain. What is clearer is that Compstat may be evolving in a new and potentially different direction. According to its supporters, "predictive policing" is an extension of Compstat's focus on using analyses of timely crime data to drive police strategies. In this case, crime and noncrime data are made readily available and combined with forecasting, modeling, and sophisticated statistics to help make predictions about where crime is likely to occur in the future. What specific role Compstat plays in predictive policing has not been studied systematically, but this new development suggests interesting avenues for future research.

Compstat's Doctrinal Elements and Its Diffusion

When Commissioner Bratton assumed leadership of the NYPD, his assessment of the organization was that it suffered from bureaucratic dysfunction. Functional specialization impeded coordination, an inflexible hierarchy hindered the rapid flow of information, timely crime data were unavailable, and the department had lost sight of its overarching mission, namely, to reduce serious crime (Bratton and Knobler 1998: 209; Silverman 1999). To streamline operations, Bratton borrowed state-of-the-art management doctrines espoused by organizational development experts in the private sector. These principles included a commitment to establishing priorities, finding creative approaches to new challenges, and using information scientifically to drive decision-making (Willis et al. 2007). According to this "reengineering" approach (Hammer and Champy 1993), successful organizations respond to uncertainties in their environments by reprioritizing their goals and revamping core structures for their successful accomplishment. Compstat, whose name comes from a computer file name, was the centerpiece of Bratton's reform efforts (Silverman 2006: 221).

At the core of NYPD's Compstat model are four crime reduction elements that are designed to make police organizations rational and more responsive to management direction: (1) accurate, timely information made available at all levels in the organization; (2) selection of the most effective tactics for specific problems; (3) rapid focused deployment of people and resources to implement those tactics; and (4) relentless follow-up and assessment to learn what happened and make subsequent tactical assessments if necessary (Bratton and Knobler 1998). Fundamental to this Compstat approach is the delegation of decision-making authority to middle managers (or precinct/district commanders) with territorial responsibility. These commanders are then held directly accountable for reducing crime in their precincts and given the necessary resources (specialist units, detectives, etc.) for accomplishing this goal.

Alongside these elements, the NYPD also developed eight crime control and quality-of-life strategies that were disseminated throughout the department. Their primary purpose was to provide commanders with guidance on how to craft specific approaches to reduce crime and disorder problems in their precincts, including youth violence, drug and gun crimes, and public nuisances such as "squeegee window washers" (Safir, n.d.: 6). However, case studies on Compstat in other departments suggest these features of the NYPD model have not been as widely implemented as its frequent "crime control strategy meetings" that are the centerpiece of the Compstat process (Willis et al. 2007). During these meetings precinct commanders appear before the department's top echelon to report on crime problems in their districts and what they are doing about them. This occurs in a data-saturated environment where crime analysts collect, analyze, and map crime statistics to spot crime trends and help precinct commanders identify patterns among crime incidents. Top administrators then use this information to quiz district commanders on the crime in their beats and to hold them responsible for solving them. Failure to provide satisfactory responses to their inquiries may lead to stern criticism or possibly removal from command. Based on what those who developed Compstat have written, as well as what those who have studied Compstat have observed, researchers have identified six core elements that have emerged as central to the development of Compstat (Weisburd et al. 2003):

- *Mission clarification*. Top management is responsible for clarifying and promoting the core features of the department's mission. Mission clarification includes a demonstration of management's commitment, such as stating those goals in specific terms for which the organization and its leaders can be held accountable – for example, reducing crime by 10 % in a year.
- *Internal accountability*. Operational commanders are held accountable for knowing their commands, being well acquainted with the problems in the command, and accomplishing measurable results in reducing those problems – or at least demonstrating a diligent effort to learn from that experience. Those who fail to do so can suffer adverse career consequences such as removal from command.
- *Geographic organization of operational command*. Operational command is focused on the policing of territories, so central decision-making authority about police operations is delegated to commanders with territorial responsibility (e.g., districts). Functionally differentiated units and specialists (e.g., patrol, community police officers, detectives, vice) are either placed under the command of the district commander or arrangements are made to facilitate their responsiveness to the commander's needs.
- *Organizational flexibility*. The organization develops the capacity and the habit of changing established routines to mobilize resources when and where they are needed for strategic application.
- *Data-driven problem identification and assessment*. Crime data are made available to identify and analyze problems and to track and assess the effectiveness of the department's response.
- *Innovative problem-solving tactics*. Police responses are selected because they offer the best prospects of success, not because they are "what we have always done." In this context, police are expected to look beyond their own experiences by drawing upon knowledge gained in other departments and from innovations in theory and research about crime prevention.

On occasion, this list has been expanded to include a seventh Compstat element or "external accountability." This refers to Compstat's capacity for making police decision-making more transparent to the public. Under Compstat, departments are externally accountable to the degree that they provide stakeholders with accurate and timely information about how well they are accomplishing their official crime control mission.

Evidence suggests that Compstat has diffused rapidly. In a national survey of large (>100 sworn) police departments in the USA administered by the Police Foundation in 2000, a third of agencies reported they had implemented a Compstat-like program with a quarter claiming they were intending to do so. This study confirmed that Compstat had "literally burst onto the American policing scene" and was following a diffusion process as rapid as "innovations in other social and technological areas" (Weisburd et al. 2004: 15). Nor does the spread of Compstat appear to be slowing, according to another survey conducted in 2006 by scholars at George Mason University and funded by the Department of Justice's Office of Community-Oriented Policing Services, 60 % of large police departments in the USA had implemented Compstat or a Compstat-like program (Willis et al. 2010a).

Compstat in Practice

It is virtually a truism in organizational theory that organizational change, including police reform, rarely works in intended ways. Accounts of the New York City Police Department suggest that Compstat's elements function like a well-oiled machine (Henry 2002; McDonald et al. 2002), but others characterize these evaluations as "advocacy" studies that lack rigorous scientific analysis by disinterested observers (Jang et al. 2010). A small but rich body of systematic research in other police departments that have sought to replicate the NYPD model has revealed some unintended effects and internal paradoxes in how it operates (Willis et al. 2007; Dabney 2010).

Using data from surveys sent to a stratified sample of 615 police agencies nationwide, Weisburd et al. (2003) concluded that Compstat reinforces the paramilitary model of police organizations. Most notably, Compstat appeared to strengthen the command hierarchy by making middle managers more responsive to top leadership direction through fear of punishment for poor performance. This finding was significant because it showed that Compstat appeared to preserve, rather than transform, the "bureaucratic" or "paramilitary" model of police organization criticized by reformers as too rigid and punitive. Follow-up fieldwork at three police agencies of different size and geographic location confirmed the survey finding that Compstat's accountability mechanism was generally the element most strongly implemented, along with a clear crime control mission and the display of crime statistics and maps at regular Compstat meetings (Willis et al. 2007). To some this could be interpreted as clear evidence of significant police reform. Compared to past police practices, Compstat reinforced the police crime-fighting mission, held middle managers more accountable for performance, and advanced the timely use of crime data in decision-making. These changes notwithstanding, the fact that departments had placed less emphasis on other key Compstat elements (geographic organization of operational command, organizational flexibility, and innovative problem-solving strategies) tempered any claims of radical change under Compstat.

One possible explanation for Compstat's uneven implementation in these three agencies was that those elements of Compstat that were most likely to be adopted were also those that were most likely to help the agency to appear progressive in the eyes of powerful stakeholders in the agency's environment (e.g., local politicians, community leaders, citizens' groups). Like many public organizations, police agencies that appeal to powerful values and beliefs in their environment about what police structures and practices *should* look like stand to gain legitimacy and increase their chances of survival and of securing valuable resources. Increasing the agency's attention to crime control, the spectacle of weekly performance evaluations, and glitzy displays of crime statistics help give the powerful impression that these three departments were doing *something* to reduce crime even though other structures designed to strengthen this focus remained fundamentally unaltered. Those elements which were least likely to be implemented were also those that represented the greatest departure from taken-for-granted beliefs about how police organizations should operate. So, for example, favoring experimentation with different problem-solving approaches over traditional law enforcement responses to crime could expose the organization to risk (as innovations rarely succeed the first time) and challenge widely held expectations that the primary role of the police is to engage in preventive and reactive patrol and make arrests.

An in-depth case study of one police department applying Max Weber's conceptualization of the modern bureaucratic organization to Compstat also revealed a paradox in how Compstat operated: the power of Compstat's accountability mechanism undermined several of its other elements including collaboration (by fostering competition) and innovation (by shrinking tolerance for failure). This case study also revealed how Compstat's strategic approach to adapting quickly to emerging crime problems clashed with classic bureaucratic structures for maintaining efficient routines (such as clear lines of authority and stable job assignments) (Willis et al. 2004).

Compstat's Effect on Crime

Compstat's implementation in the NYPD corresponded to a greater drop in New York City's crime rate than the national average. According to Silverman, "During 1994, the overall New York City crime decline was 12 %, compared to 2 % nationally, and 17 % for 1995" (Silverman 1996: 1). Its supporters claimed that this was clear evidence of Compstat's success as a mechanism for reducing crime and improving neighborhood quality of life (Bratton and Knobler 1998), but the association between Compstat and declining crime levels does not necessarily imply causation. Subsequent research seeking to assess Compstat's effect on crime suggests a more complicated picture. This is owed in no small measure to the fact that Compstat often combines agency reorganization with the implementation of a wide range of crime control strategies (e.g., order maintenance policing, zero tolerance policing, hot spot policing) making it difficult to identify the specific Compstat mechanism responsible for any crime reduction.

Applying the same descriptive analyses used by its supporters, Eck and Maguire challenge the claim that Compstat played a key role in reducing serious crime in New York City (2000). They examine data from the United States' Uniform Crime Reports between 1986 and 1998 and show that the decline in New York's homicide rate began long before Compstat was implemented. They also note that other large cities experienced significant declines in their homicide rates during the early to mid-1990s but had not implemented a Compstat program. For example, San Diego, San Antonio, and Los Angeles all experienced sharp drops in their homicide and violent crime rates during this period (Harcourt 2001). Eck and Maguire conclude, "On balance, the data do not support a strong argument for Compstat causing, contributing to, or accelerating the decline in homicides in New York City or elsewhere" (2000: 233). Similarly, Willis et al. using a simple pre-/posttest noted that crime was already in decline at the agencies where they conducted their on-site fieldwork and the rate of decline was the same or less steep following Compstat's implementation (2007).

More rigorous attempts to assess Compstat's effectiveness as a crime control strategy do not provide clear and consistent results. Using piecewise linear growth models, Rosenfeld et al. (2005) examined the impact of three law enforcement initiatives, including Compstat, on homicide trends in three cities while making systematic comparisons to crime trends in other cities and controlling for other factors influencing crime trends (e.g., measures of social and economic disadvantage and police density). Based on their analysis, they concluded that there was insufficient evidence in support of the NYPD Compstat model's impact on homicide trends. Others are also skeptical of Compstat's capacity to reduce crime (Dixon 1998; Chilvers and Weatherburn 2004).

Recent studies in Queensland, Australia, and Fort Worth, Texas, are more sanguine. They use time-series analyses to assess the impact of Compstat on crime. In Queensland, Compstat's management accountability features were combined with a problem-oriented policing approach, but this was not the case in Fort Worth. Using crime data spanning a 10-year period (1995–2004 inclusive), the Queensland study concluded that Compstat generally reduced reported crime and was cost effective, but it was most effective at reducing property-related offenses. Similarly in Fort Worth, Compstat significantly decreased property and total index crime rates but had little effect on violent crime rates. What still remains unclear are the independent effects of the different elements of the multidimensional Compstat process (e.g., accountability, crime analysis, order maintenance strategies) on crime.

In summary, when combining these findings on Compstat with a growing body of evidence on police effectiveness, it appears that when Compstat's accountability and performance measurement structures support the implementation of police practices targeted and tailored toward specific offenses, offenders, or places, Compstat can lead to reductions in crime and disorder (Braga and Bond 2008: 599).

Compstat's Relationship to Community Policing

Compstat has emerged during the same period as community policing, another popular engine of police reform in the USA and around the globe.

Community policing can be characterized as a philosophy and an organizational strategy designed to reduce crime and disorder through community partnerships, problem solving, and the delegation of greater decision-making authority to patrol officers and their sergeants at the beat level. It varies more than Compstat from place to place in response to local problems and community resources. Given the visibility of these reforms, it is natural to consider the relationship between the two when they are implemented in the same police organization: do they operate together, that is, one reinforcing the other, or are there points of conflict, where pursuing one makes it harder to pursue the other successfully? Alternatively, do they work separately, that is, each having little consequence for the other? Some have claimed that Compstat complements and supports community policing and even improves it (McDonald et al. 2002). According to this perspective, both reforms share the same concern with crime and quality-of-life offenses and seek to decentralize operations geographically. Where there are discrepancies, such as community policing's weakly developed accountability mechanism, the other reform is simply able to compensate.

A second group of researchers is less convinced that these reforms are compatible. Wesley Skogan, a well-known researcher on community policing, has commented on the tensions between Compstat's data-driven accountability elements and the different currents that push community policing in the opposite direction (2006). Community policing's focus on local neighborhood problems, developing police-community partnerships, and creative problem solving at the grassroots level are not key features of Compstat as it has evolved. It is also possible that these reforms operate independently from one another in order to minimize conflicts to existing organizational structures and routines.

The body of empirical research on this issue is very small. Willis et al. identify seven core elements that the full implementation of Compstat and community policing demands and assess where their respective reform doctrines stand on these elements (see Table 1). Briefly, Compstat focuses the police mission narrowly on serious crime, while community policing broadens it to include a wider variety of objectives including community problems, reductions in social and physical disorder, and citizens' fear of crime. Furthermore, in contrast to Compstat, community policing requires that community members play a key role in defining what the police should be trying to accomplish.

When it comes to making people feel responsible for their performance, Compstat is first and foremost a method to hold middle managers accountable for knowing what is going on in their areas and for devising timely, effective solutions to the most pressing problems. By contrast, there is less concern in the community policing literature with holding police feet to the fire of accountability.

Both community policing and Compstat promote decentralizing decision-making authority, but Compstat concentrates on the delegation of authority to middle managers, while community policing has been far more interested in decentralizing decision-making to those responsible for doing the lion's share of the police organization's work: the rank and file.

Organization flexibility refers to a department's capacity to determine where a problem is and to change or disrupt department routines to do this. Community policing requires flexibility to meet the varying demands of different constituencies, while Compstat requires flexibility to put the key resources in the hands of the right people and to alter procedural routines to do what must be done to be effective in controlling crime.

Both reforms emphasize the use of timely data to drive decision-making, but there are some important differences. Compstat relies on a department's existing data systems to focus attention on serious crime, while community policing solicits input from community residents to identify a broader range of minor crimes and social disorder deserving of police attention. Moreover, community policing places higher value on sergeants and patrol officers participating in crime analysis, unlike Compstat which empowers district commanders to identify and analyze problems.

Compstat, Table 1 A comparison of the doctrines of Compstat and community policing (Adapted from Willis et al. 2010a)

Reform element	Community policing	Compstat
Mission clarification	Broadening of police mission to include wide variety of objectives Key role of community in defining police priorities	Specific and highly visible crime reduction goal
Internal accountability	Not well developed	Middle managers held strictly accountable for crime performance by leadership
Decentralization of decision-making	To lowest level in organization, especially patrol officers, in order to customize solutions to local problems	To middle managers
Organizational flexibility	Capacity to accommodate innovation and differing needs within communities resulting in strong neighborhood-level focus	Capacity to reallocate resources for effective accomplishment of crime control objectives (putting resources in the hands of key decision makers)
Data-driven problem identification and assessment	Systematic identification of crime and disorder problems by local residents and follow-up assessments of success	Systematic identification of serious crime problems and follow-up assessments of success
	High value placed on rank-and-file officers participating in crime analysis	High value placed on middle managers doing crime analysis
Innovative problem solving	Innovation at lowest levels of organization expected and valued Significant role of community in problem-solving process	Innovation by middle managers expected and valued
External accountability	Police consult with community on objectives and progress toward them	Police publicize traditional crime statistics as measures of agency performance

Under Compstat and community policing, crime data are supposed to provide a basis for searching and implementing creative solutions to crime and disorder problems. Where they differ is that Compstat assigns most of this responsibility to middle managers using traditional crime and calls-for-service data, while community policing emphasizes the role of rank-and-file officers and the public in using customized sources of data and information to tailor specific solutions to local neighborhood problems.

Finally, Compstat and community policing's attempts to make police operations more transparent work in different ways. Under Compstat, departments are externally accountable to the degree that they provide stakeholders with accurate and timely information about how well they are accomplishing their crime control mission. According to community policing doctrine, external accountability goes far beyond merely providing citizens with standard crime measures. Under this model, police are accountable to the degree that they create a collaborative environment with local residents that is directly responsive to their concerns and to the degree that they foster an open dialog on what the police are doing and how well they do it.

Based primarily on observations conducted from 2006 to 2007 at seven police agencies in the USA, Willis et al. describe how Compstat and community policing operated in relation to each of these core elements and assessed their level of integration (not at all integrated, low, moderate, or high). They concluded that Compstat and community policing operated largely independently from each other. Their simultaneous operation helped departments respond to a broader set of goals and wider variety of tasks than had they implemented just one reform. So, for example, community policing provided opportunities to meet with community groups and identify their concerns, while Compstat helped focus the police organization's energies on fighting crime. However, those problems identified as most pressing by communities were rarely subject to the kind of intense and systematic scrutiny at regular Compstat meetings as serious crime. These findings provided the basis for recommendations on how Compstat and community policing might be integrated in order to be mutually reinforcing (e.g., constructing measures of community policing that can be prioritized, measured, and reported on at regular Compstat meetings) (Willis et al. 2010b). However, to date, these recommendations have not been subjected to empirical testing and thus remain speculative.

Predictive Policing and Future Research

At the time of this writing (2012), Compstat's emphasis on using timely crime data to address current crime problems may be evolving into a policing approach that attempts to forecast trends and then prevent crime from occurring in the *future* (Bratton and Malinowski 2008: 264). Currently in the stages of its early development, this model known as "predictive policing" uses "advanced analytics" in an attempt to anticipate or predict crime and act as a guide to "risk-based deployment" (Beck and McCue 2009: 19–20). Its elements include integrating crime data with other information sources inside and outside of the police agency (e.g., economic data on housing foreclosures); analyzing these data to identify and anticipate crime patterns related to people, places, or events; mobilizing resources; implementing strategies in response to these predictions; and linking accountability to performance targets rather than just past outcomes (Bratton et al. 2009). Similar to Compstat, this model draws on market practices commonly used in the private sector and specifically attempts to understand and predict consumer behavior. Companies like Walmart use business analytics to forecast demand and then make adjustments to their supply lines so that any increase in demand can be met successfully (Beck and McCue 2009). So, for example, in the event of a large weather event like a hurricane, Walmart stores stock up in advance on water, batteries, and other items to ensure that they have adequate supplies available. According to Charlie Beck, Chief of the Los Angeles Police Department where predictive policing originated, "the predictive vision moves law enforcement from focusing on what happened to focusing on what will happen and how to effectively deploy resources in front of crime, thereby changing outcomes" (Pearsall 2010).

The emergence of predictive policing opens up a rich vein of research opportunities including questions about the nature of its implementation in different police agencies and its effectiveness in reducing crime. Early anecdotal evidence regarding the latter is encouraging. The Santa Cruz Police Department in California reported an 11 % drop in burglaries for the 5-month period following the implementation of its predictive policing model compared to the same period the previous year. This model runs historical data on the location and time of burglaries through an algorithm. This then generates a list of locations and times with the highest probability of burglaries and auto thefts that is distributed to officers at roll call (Baxter 2011).

Any evaluation of predictive policing should also include an assessment of the consequences of its adoption on the legitimacy of the police in the eyes of citizens and other community stakeholders. Of particular concern here should be the degree to which predictive policing heralds a more pronounced move toward the profiling of certain types of people than has been observed under its predecessor. Given predictive policing's potential to link police databases with a variety of other information systems, such as the census, schools, or social services, the likelihood increases that specific groups will be targeted for additional police attention. Not only will this likely exacerbate existing racial disparities in the American criminal justice system, it also advances a troubling conception of justice and constitutional rights based on the risk of future offending rather than past practice (Harcourt 2007: 3). Some have suggested that predictive policing might unintentionally undermine the Fourth Amendment protections for some individuals (Ferguson 2011). Under current law in the USA, police officers can stop someone based on “reasonable suspicion” that a crime is being, has been, or is about to be committed. According to the Supreme Court, an important contextual factor for determining reasonable suspicion is whether or not the stop occurs in a high-crime area. Thus predictive policing opens up the possibility that individuals living, visiting, or working in such areas will “have a lesser expectation of privacy than those in other non-high crime neighborhoods” (Ferguson 2011: 1).

It might be that the promise of crime analysts applying advanced statistical methods to vast amounts of data, thereby improving the accuracy of results, convinces police chiefs that this is a viable crime prevention strategy. An alternative possibility is that ethical concerns about predictive policing’s analytic methods being used “to target individuals inappropriately for future crimes, or bad acts that they may commit but have not” could impede its progress (Beck and McCue 2009:23). Either way, the emergence of predictive policing presents an intriguing case for an assessment of Compstat as it appears to be developing, one that may lead to important empirical and theoretical insights about the changing nature of crime control policy and practice in the contemporary USA and elsewhere.

Related Entries

▶ Information Technology and Police Work

Recommended Reading and References

Beck C, McCue C (2009) Predictive policing: what can we learn from Walmart and Amazon about fighting crime in a recession? Police Chief 76:18–24

Braga A, Bond B (2008) Policing crime and disorder hot spots: a randomized controlled trial. Criminology 46:577–607

Bratton WJ, Knobler P (1998) Turnaround: how America’s top cop reversed the crime epidemic. Random House, New York

Bratton WJ, Malinowski SW (2008) Police performance management in practice: taking Compstat to the next level. Policing: A Journal of Policy and Practice 2:259–265

Bratton WJ, Morgan J, Malinowski S (2009) Fighting Crime in the Information Age: The Promise of Predictive Policing. Discussion draft, November 18. Available online: http://publicintelligence.net/lapd-research-paper-fighting-crime-in-the-information-age-the-promise-of-predictive-policing/

Baxter S (2011) Santa Cruz Police Have Success With Predictive Policing. Santa Cruz Sentinel, July 18. Available online: http://www.santacruzsentinel.com/localnews/ci_18502786.

Chilvers M, Weatherburn D (2004) The New South Wales Compstat process: its impact on crime. Aust N Z J Criminol 37:22–28

Dabney D (2010) Observations regarding key operational realities in a Compstat model of policing. Justice Quarterly 27:28–51

Dixon D (1998) Broken windows, zero tolerance, and the New York miracle. Current Issues in Criminal Justice 10:96–106

Eck J, Maguire ER (2000) Have changes in policing reduced violent crime? An assessment of the evidence. In: Blumstein A, Wallman T (eds) The crime drop in America. Cambridge University Press, Cambridge, pp 207–265

Ferguson AG (2011) Predictive policing and the fourth amendment. Am Crim L Rev. http://www.american-criminallawreview.com/Drupal/blogs/blog-entry/%E2%80%9Cpredictive-policing%E2%80%9D-and-fourth-amendment-11-28-2011. Accessed 3 Jan 2012

Hammer M, Champy J (1993) Reengineering the corporation: a manifesto for business revolution. Harper Business, New York

Harcourt BE (2001) Illusion of order: the false promise of broken windows policing. Harvard University Press, Cambridge, MA

Harcourt BE (2007) Against prediction: profiling, policing, and punishing in an actuarial age. University of Chicago Press, Chicago

Henry VE (2002) The Compstat paradigm: management accountability in policing, business, and the public sector. Looseleaf Publications, New York

Jang H, Hoover LT, Joo H-J (2010) An evaluation of Compstat’s effect on crime: the forth worth experience. Police Quarterly 13:387–412

Mazerolle L, Rombouts S (2007) The impact of COMPSTAT on reported crime in Queensland. Policing: An International Journal of Police Strategies and Management 30:237–256

McDonald P, Greenberg SF, Bratton WJ (2002) Managing police operations: implementing the New York crime control model, Belmont, CA

Pearsall B (2010) Predictive policing: the future of law enforcement? NIJ Journal 266:16–19

Rosenfeld R, Fornango R, Baumer E (2005) Did Ceasefire, COMPSTAT and Exile Reduce Homicide? Criminol Publ Pol 4:419–450

Safir H (n.d.) The Compstat process. Office of Management and Planning, NYPD

Silverman EB (1996) Mapping change: how the New York City police department re-engineered itself to drive down crime. Law Enforcement News 15:1

Silverman EB (1999) NYPD battles crime: innovative strategies in policing. Northeastern University Press, Boston

Silverman EB (2006) Compstat's innovation. In: Weisburd D, Braga A (eds) Police innovation: contrasting perspectives. Cambridge University Press, Cambridge, pp 267–283

Skogan WG (2006) Police and community in Chicago: a tale of three cities. University of Chicago Press, Chicago

Weisburd D, Mastrofski SD, McNally AM, Greenspan R, Willis JJ (2003) Reforming to preserve: COMPSTAT and strategic problem solving in American policing. Criminol Publ Pol 2:421–456

Weisburd D, Mastrofski SD, Greenspan R, Willis JJ (2004) The growth of Compstat in American policing. Police Foundation, Washington, DC

Willis JJ, Mastrofski SD, Weisburd D (2004) COMPSTAT and bureaucracy: a case study of challenges and opportunities for change. Justice Quarterly 21:463–496

Willis JJ, Mastrofski SD, Weisburd D (2007) Making sense of COMPSTAT: a theory-based analysis of organizational change in three police departments. Law and Society Review 41:147–188

Willis JJ, Kochel TR, Mastrofski SD (2010a) The co-implementation of Compstat and community policing: a national assessment. Department of Justice, Office of Community-Oriented Policing Services, Washington, DC

Willis JJ, Mastrofski SD, Kochel TR (2010b) Recommendations for integrating Compstat and Community Policing. Policing: A Journal of Policy and Practice 4:182–193

Computational Criminology

► Agent-Based Assessments of Criminological Theory

Computer Forensics

D. Kall Loper
Department of Computer Science and
Engineering, Lyle School of Engineering,
Southern Methodist University, Dallas, TX, USA

Synonyms

Low-level data forensics

Overview

The 1s and 0s that make up digital data are incomprehensible to most people. Even though we have heard about them many times, we often have no idea how the 1s and 0s turn into useful data. Forensic analysts must make sense this data and present it to persuade others. This entry explains how the binary data (1s and 0s) are exactly equal to a slightly easier way of showing the data: hexadecimal. Hexadecimal is commonly seen as the most fundamental representation of the data. Once expressed, the organization of the data becomes easier to understand. Common structures and methods of discovering and explaining other structures are explained and then shown in two examples: carving a lost or deleted file and finding hidden data in a common JPEG photograph. In the first example, a step-by-step set of commands is used to find and recover a deleted picture. The commands demonstrate the same steps used by automated forensic packages. The second example examines data embedded in a JPEG photograph byte by byte. Using publicly available resources, the structures of embedded data are explained. With this information, an analyst's ability to present data exceeds the ability of commonly used forensic packages. Along with the examples is a discussion of how these operations are handled by automated forensic tools. The whole is brought together in a discussion of how these techniques can improve testimony and the investigative effectiveness of a forensic analyst.

Introduction

Digital information in its raw form is expressed as ones and zeros, or more correctly as on's and off's. Each single piece of information: one or zero is called a bit. This understanding is part of popular culture, but is essentially useless. Almost all meaningful information is gathered into bytes or "words." With Intel processors, the standard word is 2 bytes or 16 bits. Even though the fundamentals of chip architecture and low-level file structure may also seem useless to the day-to-day business of an investigator or forensic analyst, the structure of data is critical to understanding digital evidence. While the majority of useful evidence comes from active user files, the majority of disk space is unallocated. This means that either direct interpretation of unallocated space forensic recovery or file carving is necessary to access that space. All three of these activities benefit from some knowledge of reading a hexadecimal presentation of the data. The term "hexadecimal presentation" is specifically chosen because it is merely another way to look at the data on the disk and in no way changes the content of the file.

Fundamentals of Data: An Explanation of Hexadecimal Presentation

Data "presentation" refers to the way data is displayed to a user. A native presentation is the way the file was intended to be displayed to a user. Hexadecimal presentation shows the data in a raw form but summarized into couplets of hexadecimal digits. Hexadecimal digits include the familiar 0–9 of decimal but also include A–F. Where the more familiar decimal repeats its digits every 10 counts, hexadecimal repeats every 16. Since these counting systems use the same characters, it is common to specify which system is being used. Hexadecimal numbers are often followed by "xh" to denote hexadecimal (e.g., 00xh); similarly, decimal notation can be noted with "xd" if it is intermixed with hexadecimal.

There is no mathematical difference between counting in hexadecimal (base 16), decimal (base 10), or even binary (base 2). Each notation can be converted to the other with absolutely no loss of information. For example, 255 (decimal) is exactly equal to FF (hexadecimal) which is exactly equal to 1111 1111 (binary). Internet addresses also demonstrate this fact. 192.168.0.1, a valid IP address, is expressed in decimal-dot notation. It can also be expressed in binary as 1100 0000. 1010 1000.0000 0000.0000 0001. However, it can most efficiently be expressed as C0.A8.0.1 in hexadecimal. Hexadecimal notation is especially convenient for information based on 8-bit bytes. As can be seen in the IP address, two hexadecimal digits – a hexadecimal couplet – can completely express 8 bits. Rather than presenting eight digits in a row, the system uses two digits, making the information easier to interpret.

The underlying structure of a file is often based on a byte structure. Either an 8-bit word (byte) or some multiple of them is used to hold data. In Windows programming, a 16-bit word is called a WORD, a 32-bit word is called a DWORD, and a 64-bit word is called a QWORD. When reading outside resources about file structure, these terms are often used freely with the assumption that the reader understands them. Resources that do not deal specifically with the Windows Application Programming Interface (WinAPI) may use "word" in a different way.

Even when it only requires one or two digits to account for all possible values, a full byte or even a WORD or DWORD may be used to maintain byte alignment. This means that new fields within the data tend to start on the 0th, 4th, 8th, or 12th byte of a row of 16 bytes. Even when only single bytes are used, a full WORD or DWORD may be filled with padding to maintain byte alignment. Although any value can be used for padding, it is most common to find nulls (00xh). When working with tools that allow a user to change the number of columns displayed in hexadecimal, it can be useful to maintain the standard 16 byte width to allow easy recognition of byte aligned file structures.

Basic File Structures

File types are structured according to the needs of their creators; thus, there are no universal rules about file structures. However, most file types use a header and body structure. Headers contain information about the file. A file signature value or "magic number" is often used to identify the file type. There is no guarantee that a signature value is unique to a given file type. The Web site wotsit.org maintains lists of file types by extension and gives signatures values when available. Other commonly found fields in the header include file version number, count of records contained in the file, logical size of the file, checksums of file contents, and many others. Header information usually contains only information necessary to interpret the file structure and identify the file type.

The body of the file contains the data used by the application interpreting the file. All of the possible data types cannot be listed here, but there are a few general types. Text-based file types tend to be presented in Unicode or ASCII equivalent. Most hex tools provide an ASCII interpretation to the right of the hex itself. Unicode corresponds to ASCII with Latin characters (the letters used in English). The only difference noticeable is that Unicode has columns of null (00xh) between the letters. Unicode needs the extra bits to encode accented Latin characters and non-Latin characters. However, in English they are not used so the bits are empty, thus null.

The body of a file may also contain blended text and non-text regularly interspersed. This is the hallmark of a file with discrete records. Each record can either follow a precisely established format or contain its own header information. The contents of the record may need to be interpreted or may be readable as text. Repeating record structures make it much more likely that a file fragment will provide useful information. For example, partial Web browser history files can be recovered and read from the hex without recovering an intact file. In many cases, the human eye can spot repeating patterns or legible fragments of files that cannot be identified by automated file carving software. A few regular expression searches (RegEx) and a little knowledge of file structure can produce accurate and precise results.

Encapsulated File Structures: Encryption, Compression, and Encoding

Sometimes it is not immediately apparent that data is part of a known file type. This may be because the data is encapsulated or layered in other forms of processing. It is common practice in computer science to break a complex task into layers of activity. Each layer is responsible for a discrete task and either passes up or down to the next layer without interacting with the contents. In this way, multiple layers can be imposed on the data found on the disk. Processes that encrypt, compress, or merely convert data can make it impossible to directly analyze the data. The data on disk in no way resembles the original file data when viewed directly, but the original file data can be easily recalled by reversing the encapsulation or layering process. For useful analysis, the data must be converted to a readable format. Automated forensic packages can process this data back through the layers that were imposed on it or it can be done by isolating the file and processing it with a stand-alone tool.

Accessing the Data

There are many tools available that allow an analyst to access data with a hexadecimal presentation. Rather than try to exhaustively list them, an example of each type can serve to illustrate them. XXD is a utility found on Linux systems that will present a file in hexadecimal. It can also be used in Windows with the Cygwin utility. XXD is called from the command shell (e.g., Bash). Unless the file to be examined is small, it is best to send the output to “less” or a GREP search as follows: XXD/evidence/image.dd | grep ffd8ffe1. This line will search a forensic image for a JPEG signature characteristic of a JPEG picture with EXIF data embedded in it. WinHex is a full featured hex editor that is commercially available. Care should be taken when using hex editors to not write to the file being examined. The interface is familiar to Windows users. Familiar menus and dialog boxes allow the user to open a file in hex presentation. Additional dialogs allow a user to open a partition or physical disk the same way. FTK Imager is a commercial-grade imaging and data access tool provided at no cost by AccessData. Imager can open physical devices (disks, CDs, flash drives, etc.), partitions, or files. A file system hierarchy view similar to Windows Explorer allows a user to traverse a file structure and display data in hexadecimal or in its native presentation. FTK Imager is intentionally incapable of performing searches in the data.

Automated Forensic Packages and Hexadecimal-Level Analysis

Automated forensic packages allow analysts to rapidly conduct searches across multiple data sets and access archive files (layered data) without intermediate steps. These tools are not rivals or replacements for lower-level tools; quite the contrary, they include low-level tools with other functions that add greater convenience for the analyst. Automated tools can contextualize the data within the file system or within search results. Automated packages can also prevent human error found in complex, repetitive tasks; however, they can mask unique aspects of the data through consistent presentation. For example, some forensic software takes a file that was never meant to be viewed by a user and presents it as a table of values or an html presentation (shows it as a Web page). While there is no "native" view of the file, hexadecimal presentation is the most accurate. If the software cannot interpret a field of data, it may simply be skipped in the "user-friendly" presentation. If the user relies on the tool, then there may never be a chance to see the skipped data until an opposing expert brings it forward. Most forensic experienced users are aware of the warnings to "know" their tools and the tool's limitations, but without a firm basis in interpreting the hexadecimal data, they are left thinking that comparing the output of two forensic tools works as validation; it does not.

Automated packages can also introduce a bias for intact data that can be cleanly presented by the package. There is no denying that automated packages are necessary to process the massive amounts of data found on even a moderately sized corporate system (very roughly 40 GB). It is self-evident that automated packages are needed for home systems that can run into the multiple terabytes. Automated packages tend toward a clean presentation of native views. With huge amounts of data available, there may be justification in searching visible user files first. As this practice is taught to new examiners and becomes ingrained in experienced examiners or, worst of all, becomes part of policy in a police department or forensics practice, it leads to a bias toward intact data of known types. It is easy to automate searches in EnCase™ with EnScripts™. If policy is in place to exclude unknown file types or search only intact data, large amounts of relevant data can be completely missed. It is easy to avoid this with the full text indexing and searches found in FTK: the Forensic Tool Kit™ by AccessData. However, users soon learn to ignore or even explicitly filter out results from unallocated space or files fragments in favor of those that can be cleanly presented. Users of either package can overcome these limitations, but there is a tendency to become complacent with the tool. By getting in the habit of examining "corrupt" files, unknown file types, and fragmentary data, a forensic examiner builds a mastery of forensic examination that goes beyond document review or tool dependence.

Interpreting File Content

The two lines in Fig. 1 are output from a hex editor. The first line identifies the column (in hexadecimal, there are 16xd columns). The second is the first 16 bytes of a JPEG photograph with EXIF data embedded. The JPEG signature is FF D8 FF E1xh (bytes 0–3). The JPEG type identifier is 45 78 69 66xh (bytes 6–9). This is followed by two null bytes: 00 00xh (bytes A–B). The next two bytes identify the order of following information 4D 4Dxh (bytes C–D). The ASCII interpretation of the bytes (4D 4Dxh) as MM indicates that values should be read left to right as read by Motorola chips (called big endian byte order). The other possible value is "II," indicating that values should be read right to left as read by Intel chips (called little endian byte order). Much more information can be found in the following lines like the make and model of the camera used

```
0   1   2   3   4   5   6   7   8   9   A   B   C   D   E   F          ASCII
FF  D8  FF  E1  2F  F7  45  78  69  66  00  00  4D  4D  00  2A     ÿØÿá/÷Exif  MM *
```

Computer Forensics, Fig. 1 The first sixteen bytes of a JPG photograph with the hexadecimal and ASCII presentations of the data

```
0   1   2   3   4   5   6   7   8   9   A   B   C   D   E   F          ASCII
37  B0  3D  65  B9  FF  D9                                         7°=e¹ÿÙ·······
```

Computer Forensics, Fig. 2 The last seven bytes of a JPG photograph with hexadecimal and ASCII presentations of the data

to create this photograph. The GPS coordinates are also recorded in this particular photograph (not shown). Without an EXIF aware viewer, an analyst would not know this data was present without viewing it in hexadecimal; the user-friendly presentation shows only the photograph.

There are thousands of file types in common use, it would be impossible to catalog them all in this entry, but technical details can be found on the Web in a few common resources. Microsoft file formats can be found in Microsoft TechNet or in the Microsoft Developer Network (MSDN). To identify an unknown file signature or extension, wotsit.org maintains voluminous catalogs of file types. Details of open-source file types can often be found through their SourceForge projects. Some file types are maintained by work groups comprised of representatives from many companies, academic institutions, governmental bodies, etc. Web searching these resources or simply looking for them in the results of a Web search for the first 4–8 bytes, in hexadecimal, from an unknown file will help find authoritative and reliable sources for file formats.

Recovering and Interpreting Data

Files that have been deleted or corrupted can be recovered by using the techniques described in this entry. File carving is the process of recovering a file or fragments of a file from unallocated space (where the file system has marked the space as unused). The easiest way to do this is to search for a header signature (FF D8 FF E1xh for a JPEG picture) and copy all the data between it and the file footer signature (FF D9xh for a JPEG) as seen below in the last line of a JPEG file (Fig. 2).

Recovering a File from Start to Finish

In this example, a JPEG photograph was copied to 128 MB flash drive formatted with the FAT32 file system. An image of the physical drive was created with the Linux/Unix utility “dd” using a SMART boot disk on a host system. SMART is a commercial forensic package created by ASR data, but any forensically sound Linux boot disk will work. The raw, bit-stream copy of the flash memory is analyzed using hexadecimal tools mentioned previously. The purpose of this example is to comment on various file structures along the way to recovering the file. A few pointers on the commands will help in understanding what is happening. All commands are issued in a terminal window in the Ubuntu operating system. Multiple commands can be strung together with a “pipe” character (|). Commands can also have options that change the way they act. Each of the following commands defines an operation in the recovery of a file. These same operations are used by automated packages. The automated packages do not actually implement the Linux/Ubuntu commands, but they have the same functions.

The first step is to find the file to be recovered. As seen previously, the signature of a JPEG file is

FFD8FFxh. Choosing to include the fourth couplet, either E0xh or E1xh, limits the number of files found. Removing it has other interesting effects as will be seen later. “xxd” converts a file from a binary stream to a hexadecimal stream. By refining it with command options, it can be used to extract specific information. “grep” is a Unix environment search tool that has been ported out to many other environments. By sending the output of an entire file in a hexadecimal stream to grep, an analyst can search that stream very efficiently. The image file is named “image_dd.001.”

```
UBUNTU Command: xxd -g 0 image_dd.001 | grep ffd8ff
```

Output:

```
0400400:
ffd8ffe000104a464946000101010048......
JFIF.....H
```

By default xxd lists 16-byte lines with a hexadecimal line number. In this case, the line number is 0400400xh. It is not obvious, but that line number is a good sign for carving a complete file. Each sector is 512xd bytes. 512xd equals 200xh. Thus, any line number evenly divisible by 200xh is the first line of a sector. Most files start on sector boundaries, not in the middle of a sector. The first four bytes are ffd8ffe0xh. This is one of the possible JPEG signatures. These bytes are followed by the remaining 12 bytes in hexadecimal followed by the ASCII conversion of the data. One other consideration is that xxd gives output in a conveniently grouped byte pattern by placing a space between each hexadecimal couplet. “xxd –g 0” removes the grouping and makes the output searchable by grep.

It is necessary to convert this line number to decimal. In Windows, open calculator and choose scientific view. Enter 400400 in hexadecimal and convert to decimal. In a Linux command line, also use a calculator: bc. Echo directs output to stdout (the terminal or a pipe). “bc” is the binary calculator. The following command tells echo to repeat the information “ibase = 16; 400400” to bc. bc interprets the input base (ibase) as hexadecimal (16) and the number as 400400 with no other operations. The default base for bc is 10, so the output base stays 10. The output is the 400400xh converted to decimal. When using bc, remember the default input in hexadecimal is uppercase, but the default output of xxd is lowercase. This will be shown in the next example.

```
UBUNTU Command: echo "ibase = 16; 400400" | bc
```

Output:

```
4195328
```

Once the header location is known, the footer signature location must be found to locate the end of the file. Line 400400xh was the first to contain a JPEG signature, so all previous lines can be skipped. Once again, xxd and grep can locate this value.

```
UBUNTU Command: xxd -s 4195328 -g image_dd.001 | grep ffd9
```

Output:

```
0406 fc0: 20 86 73 13 7a ee 7f ff d9 00 00 00 00 00 00 00.s.z...........
```

The line number is 0406fc0xh. The first nine bytes are part of the JPEG file. Bytes number eight and nine are the footer signature value (FF D9xh), marking the end of the file. The trailing nulls (00xh) are file slack space. It cannot be seen from the output above, but there are actually 55 bytes of file slack space before the end of the sector and in this case, the beginning of the next file. Those bytes could contain fragments of previous file(s) stored in that sector. In this case, they contain nulls because the media was securely deleted before use in this example.

Again, it is necessary to convert the line number to decimal, but also to add the 9 bytes found in that line. The direct output of the line number needs to be converted to uppercase or produced from xxd using the “-u” switch to make the hexadecimal digits uppercase. The default output of xxd is lowercase and bc requires uppercase. Since decimal and hexadecimal use the same characters for 0–9, it may be hard to tell that the additional nine bytes are added in hexadecimal.

```
UBUNTU Command: echo "ibase = 16; 406FC0 + 9” | bc
```

Output:

```
4222921
```

A final calculation will yield the file length. This can also be accomplished with a simple calculator.

```
 0   1   2   3   4   5   6   7   8   9   A   B   C   D   E   F          ASCII
FF  D8  FF  E0  00  10  4A  46  49  46  00  01  01  01  00  48     ÿØÿá··JFIF·····H
```

Computer Forensics, Fig. 3 A JPG photograph header showing the JPG signature in hexadecimal and a file type identifier intended to be interpreted in ASCII

UBUNTU Command: echo "4222921 - 4195328" | bc

Output:

27593

With all the file location information available, the file itself can be carved out of the image and exported to a file system. In this case, it is simply exported to the working directory containing the image, but the output can be directed to any mounted file system.

UBUNTU Command: dd if = image_dd.001 of = carved.jpg skip = 4195328 bs = 1 count = 27593

Output:

27593 + 0 records in

27593 + 0 records out

27593 bytes (28 kB) copied, 0.515814 s, 53.5 kB/s

There is also a new file called “carved.jpg.” It can be viewed with the xview command or with the tools of the desktop environment. To be clear, there is no difference between the carved file and the original. This can be seen by viewing the JPEG file and seeing the picture. It can also be mathematically confirmed by running an MD5 checksum of the original file and the carved file. In this case, the MD5s match, confirming that not one bit is different.

Interpreting Information in a JPEG File

Aside from the picture itself, there is often other data embedded in a JPEG file. Two of the many types of JPEG file are discussed here. The first is the standard “JFIF” file. This is the type of file seen in the previous example. To comply with the JFIF standard, the first two bytes must be the image marker FF D8xh. The second two bytes are the application marker FF E0xh. These bytes are the header signature value. A five-byte identifier follows these: 4A 46 49 46 00xh; the identifier has the ASCII values JFIF followed by a null terminator. This can be seen in (Fig. 3).

The second type of JPEG file contains additional information embedded within the file. EXIF data is often included with Tagged Image File Format (TIFF) files and JPEG/TIFF files. With photographs, it is most often used to capture camera settings, manufacturer, model, and other details relevant to photographers. As cellular phone cameras have become more common, the data has begun to include other elements like GPS coordinates. EXIF data also includes metadata describing the JPEG file. It is possible to get creation date for the image from the camera rather than from the file system containing the image. Much of this information is typically available through media management software, but automated forensic packages have been slow to update the types of EXIF data offered. This means that without exporting each JPEG file and viewing them in media management software (not forensically sound software), a forensic examiner may miss critical data like the GPS coordinates of where the picture was taken or embedded data possibly used on social media sites. As an alternative, the analyst can learn how to decipher the contents of that data directly.

In the following example, the first 48 bytes of a JPEG picture are used to illustrate how EXIF information is stored in TIFF and JPEG files (Fig. 4) (Table 1).

The designers of the JPEG file, the Joint Photographic Experts Group, created the JPEG file to be versatile. A number of options were left open for future development or customization. Because of these fields could not be anticipated and listed in the original file specification, some fields were reserved specifically to describe subsequent field’s structures and locations. Image File Directories (IFDs) may either contain brief

```
 0   1   2   3   4   5   6   7   8   9   A   B   C   D   E   F         ASCII
 FF  D8  FF  E1  2F  F7  45  78  69  66  00  00  4D  4D  00  2A   ÿØÿá/÷EXIF··MM·*
 00  00  00  08  00  09  01  0F  00  02  00  00  00  08  00  00   ················
 00  9E  01  10  00  02  00  00  00  08  00  00  00  A6  01  12   ·············|··
```

Computer Forensics, Fig. 4 These are the first 48 bytes of a JPG photograph. The structure of these bytes is explained in Table 1, non-repeating elements of the header, and Table 2, repeated structures in Image File Directories (IFD's)

Computer Forensics, Table 1 JPEG file header and partial application data

Field	Bytes	Sample	Notes
SOI marker	2	FF D8	Start Of Image marker, also part of the file signature
APPO marker	2	FF E1	Application use marker, indicates the type of data in the file
APPO length	2	2FF7	Length of application field: 12,279xd (varies with each file)
Identifier	5	45 78 69 66 00	"EXIF" and null terminator (00xh)
Byte order	2	4D 4D	"MM" indicates "big endian" byte order
Version	2	00 2A	Version is always 42xd
Offset to IFD	4	00 00 00 08	Image File Directories (IFD) contain EXIF data

Computer Forensics, Table 2 Image File Directories (IFD)

Field tag		Field type		Length of field	Data or pointer	
(2 bytes)		(2 bytes)		(4 bytes)	(4 bytes)	
01 0Fxh	271xd	00 02xh	ASCII	00 00 00 08xh	00 00 00 9Exh	158xd
01 10xh	272xd	00 02xh	ASCII	00 00 00 08xh	00 00 00 A6xh	166xd
01 12xh	274xd	00 03xh[9]	WORD	00 00 00 01xh	00 01 00 00xh	1xd
01 1Axh	282xd	00 05	Ratio	00 00 00 01xh	00 00 00 AExh	174xd
Other IFDs omitted						
87 69xh	34665xd	00 04xh	DWORD	00 00 00 01xh	00 00 00 0Cxh	204xd

data (IFD resident data) or indicate where more extensive data stores may be found by following a pointer to another location in the file. The IFD tag is shown in bold in the example above. Each IDF is 12 bytes. The table below explains the first two IFDs from the example above. The first field of the IFD identifies it as tag number 271 (10 F in hexadecimal equals 271 in decimal). This value can be found in the TIFF standard on page 35 and in a table on page 117. The TIFF standard document indicates that tag 271 is the make of the camera or scanner. The second field in the first IFD is the field type. According the TIFF standard on page 15, type 2 is 7-bit ASCII text. That means that the camera or scanner make will be reported in ASCII text. The third field of the first IFD is the length. The make of the camera will be reported in 8 ASCII characters. The fourth and final field of the first IFD is either the data indicated (make of the camera) or a pointer to the make of the camera. The last field has four bytes; if the data length is greater than four, the value of this field is a pointer to where the data can be found. Counting 158 bytes from the APPO section (not the beginning of the file), there are 8 ASCII bytes that identify the camera's make: "HTC-8900." This cannot be seen the sample; the sample data ends before the third IFD's field type is shown (Table 2).

The second IFD indicates the model of the camera or scanner. With the camera phone that created this photograph, the make and model are identical. The one difference that can be seen in the IFD is the pointer. The previous pointer was 158 bytes and the data occupied 8 bytes; the new pointer is 166 bytes (158 + 8 = 166). The third IFD is the orientation of the picture. The field type indicates that this is a WORD (2 bytes). The length of field indicates that only one unit is used to store the data. Since the field type is WORD and only one unit is used, the data for this IFD occupies one WORD or 2 bytes of data. The data or pointer field will contain data if it is less than four bytes. This IFD has resident data. Only the first two bytes are used, so the remaining two bytes are padded with a null (00xh).

The final IFD in the first group will indicate that there are no more IFDs with a pointer of 00 00 00 00xh or will give a pointer to the next IFD table. The tag type 34665 is reserved to indicate that the IFD does not contain displayable content but either a final entry or a pointer to the next set of entries. Tags higher than 32,768xd are considered private tags. They are typically used by organizations, hardware makers, or software makers for special data to be stored. There are no firm standards on where in the file an IFD can be stored. The presence of unique IFDs at unpredictable locations means that an analyst who is completely reliant on a given tool will likely miss data.

Both Finding and Interpreting Data in a JPEG

Another lesson presented by exploring the EXIF data is that pointers within a file indicate the location of data anywhere within the file. EXIF data can be placed anywhere within the Application Use area, as defined by the APPO marker and length. The JPEG file standards document, used to decipher the bytes found in the JPEG header, indicates that a JPEG is a wrapper for content data. Part of that data is one or more pictures, but part of it can be text or some other form of data. The file examined for EXIF data actually contains two images. A search for the image marker: FF D8xh shows that there can be three markers in a JPEG file. There is one marker at the beginning of the file, another at the beginning of a thumbnail, and a third at the beginning of the full-sized image. A fragment with a complete thumbnail may appear to be only a few lines of the greater image because the thumbnail does not display. The thumbnail may provide confirmation that the image was present on the system or it may show enough detail to prove useful. Without some knowledge of the underlying file structure, it is easy to ignore the image as being corrupt or incomplete.

Forensic Examination and Testimony

It is rare that a forensic examiner making conclusions supported by the data is challenged. The most common scenario in law enforcement forensics is the simple production of documents or images that speak for themselves. The challenge often lies in finding the data and backing the results with properly documented procedure, but not always. The most common scenario for civil forensic examiners is simple preservation of data in a forensically sound manner and production of documents for review by either an attorney or a content expert. Again, the challenge lies in producing a result in a consistent and searchable format with well-documented procedure, but not always. For this type of case, an automated forensic package with consistent results from a process that can be replicated is the tool of choice. They allow large volumes of data to be processed reliably.

For a subset of investigations in both civil and criminal forensics, the analyst must go deeper to understand what has happened to a particular file or in the system itself system. In the example above, the EXIF data contained in the picture includes GPS coordinates of where the picture was taken. The EXIF data in that picture also contains date and time stamps. Such times might differ from the file times that were created when the picture was downloaded from the camera. The EXIF data contains the make and model of the camera which can help match the camera to one seized at a crime scene. All of this information can help make a case, if the investigator is aware of it. In civil forensics, an “expert” may testify that all of a given type of file was produced to the opposing side during discovery. Without the ability to interpret and understand file types that are not known to the automated package used, the “expert” may miss an entire file type with crucial data. Almost worse, the “expert” may not warn the attorney of the presence of such information. An entire legal strategy may fail based on a key fact that was absent to the uninformed side.

Using the data from the JFIF and EXIF examples, it has been shown that files in unallocated space can be identified and recovered in a process known of file carving. By first locating the header signature of a file, then locating the footer signature, and copying all the data in between, an analyst can carve a file from unallocated space. This allows the analyst to demonstrate an understanding of the process used by automated tools that carve files. If segments of the file are stored in more than one location, a state called “fragmented,” an analyst can still derive useful information from EXIF data found in the initial fragment. Garfinkle (2007) published data that estimated 92 % of files found on an NTFS volume were intact or fragmented once. This means that most files can be found and carved intact. The remaining, fragmented, files can still yield useful results. If fragmentation is minimal, then file structure may indicate how to carve and combine the pieces.

Conclusion

Understanding the structures of digital files is more than an academic exercise. Analysts who limit themselves to the output of the tool do not bring any special skill to their analysis. Automated tools play an important role in rapidly processing large amounts of information, but they should not be allowed to set boundaries on the evidence. Restricting the data considered to easily displayed files places boundaries on the data. The inability to interpret file fragments places boundaries on the data. Even self-evident results can be successfully challenged by calling into question the process that produced them or the knowledge of the person who chose the process. Sometimes, an opposing expert can even pray on the ignorance of an analyst. The analyst either questions his or her own results or argues beyond his or her knowledge. The best defense to such challenges is a deep and thorough understanding of how the tools work and the structure of the data found as it sits on the disk.

Related Entries

- ▶ Automated and Manual Forensic Examinations

Recommended Reading and References

Adobe Systems (1992) TIFF: Revision 6.0, http://exif.org/TIFF6.pdf

Aware Systems (2008) Tiff tag reference, GPS tags, http://www.awaresystems.be/imaging/tiff/tifftags/privateifd/gps.html

Becker R, Payne S (2011) International data forensic solution center: computer forensics investigation training. International Data Forensic Solution Center, Phillipsburg

Garfinkle S (2007) Carving contiguous and fragmented files with fast object validation. Digital Invest 4(S 1):2–12

Maischein M (2011) The TIFF file format: I-TIFF, http://www.fileformat.info/format/jpeg/corion.htm.

Murray J, Van Ryper W (1996) Encyclopedia of graphics file formats, 2nd edn. O’Reilly Media, Sebastopol, CA

National Digital Information Infrastructure & Preservation Program (2011) Still images: tags for tiff and related specifications, http://www.digitalpreservation.gov/formats/content/tiff_tags.shtml

Nugent T, Moolenaar B (1996) XXD manpage. http://www.unix.com/man-page/OpenSolaris/1/xxd/

Pal A, Memon N (2006) Automated reassembly of file fragmented images using greedy algorithms. IEEE Trans Image Process 15(2):385–393

Pal A, Sencar H, Memon N (2008) Detecting file fragmentation point using sequential hypothesis testing. Digital Invest 5(S 1):2–13

Concept of the Police

- ▶ Democratic Policing

Conceptualizing of Police

Fabien Jobard
Centre de Recherches Sociologiques sur le Droit
et les Institutions Pénales (CESDIP), CNRS,
Guyancourt, France

Overview

Police as an institution cannot be said to have inspired in-depth conceptualization efforts from criminology and criminal justice so far. Scholars in this area show a clear preference for empirically exploring what individual police officers do or think, and the policing concept they most readily converge towards tends to hinge on the use of force, a notion developed by ethnomethodologist, Egon Bittner. This concept holds that what constitutes policing as such is "the distribution of nonnegotiably coercive force." In another tradition, stemming from European political philosophy, police are heir to a dual dimension, i.e., its relationship to both knowledge and dogma.

However, numerous affinities between these two schools of thought – the Anglo-American sociological school, with its interactionist inclinations, and the European school of political philosophy – suggest that the concept of police should in fact be understood through a dualist epistemological approach. Both in legal and knowledge terms – two fundamental dimensions of policing – police as an institution is highly idiosyncratic in that it concomitantly harbors both (a) the rule of law and violence and (b) knowledge and ignorance.

Fundamentals of the Police Concept

The concept of police currently used in social science has been derived mainly from the legacy of American ethnomethodologist Egon Bittner, who considered the use of physical force as the "core of the police role."

Egon Bittner and Nonnegotiable Coercive Force

From his field observations of policing patrols, Egon Bittner was able to characterize the police by their distinctive function: "The role of the police is best understood as a mechanism for the distribution of non-negotiable coercive force employed in accordance with the dictates of an intuitive grasp of situational exigencies" (Bittner 1974/1990: 131). Situational exigencies are not defined in and of themselves but are left to the appreciation of the police the field: "something-that-ought-not-to-be-happening-and-about-which-someone-had-better-do-something-now" (Bittner 1974/1990: 249).

Virtually all policing scholars today rely on this definition as their starting point. With it, the concept of police follows in the footsteps of early twentieth century German social scientist Max Weber, who had defined the state as an organization that retained a "monopoly on the legitimate use of physical force."

This concept is currently known as the Police Use of Force Paradigm (PUFP) and has several implications. First, whatever force is employed ought to be minimal, i.e., the exact amount of force required by the situational exigencies that are being dealt with, which should not be exceeded. Such a force is neither destructive nor annihilative, as opposed to military force.

Besides all this, force is the natural, mechanical outcome of common-sense behavior in a critical situation. Bittner here still appeals to policemen's intuition: to deal with troubles, policemen, ordinary individuals in a pacified society, draw upon their intuitive perception of situations in order to anticipate what the community is expecting from them. Force, with PUFP, is thus located on the demand, instead of the supply, side: "citizen demand is a factor of extraordinary importance for the distribution of police service" (Bittner 1974/1990: 252).

It follows that PUFP, as a concept, pertains to police self-restraint, which gives sociological resonance to European legal theories on self-restraining state power. In this respect too, it is linked to the concept of legitimate violence introduced by Max Weber. The basic assumption, however, is that the police are the institution that intervenes in situations where the ultimate capacity of the police to dominate is not in doubt, as opposed to circumstances of armed rebellion and warfare, where military force is necessary. To Bittner, policing rests on social consent, which is why he considers that the police should not be resorted to in civil war situations, for instance, or even in the context of crowd control – a function that to him is devoted to military corps such as the gendarmeries in Europe or the National Guard in the USA.

Controversies on the PUFP Model

Canadian scholar Jean-Paul Brodeur has produced a powerful critique of the PUFP (Brodeur 2010: 103–128). He criticized Bittner for not discriminating between competence and performance. Indeed, empirical research has consistently shown that force is actually quite rarely resorted to by even the uniformed police patrolling the skid-row areas of large metropolises. Brodeur argued that a concept of police cannot be grounded in what the actual police rarely or never do: the concept should target what policing is, not what it may, or might or might not be (assuming a situation where the police, for want of public consent, would be unable to resort to force. . .). To Brodeur, the key and actual feature of any police was "legal lawlessness" (Brodeur 2010: 6) "Policing agents are part of several connected organizations authorized to use in more or less controlled ways diverse means, generally prohibited by statute or regulation to the rest of the population, in order to enforce various types of rules and customs that promote a defined order in society, considered in its whole or in some of its parts" (Brodeur 2010: 130). This concept is by no means incompatible with the PUFP; in fact, it proceeds from it: force, considered by Bittner to be the "core of the police role," here becomes one of the many means of expression of the "legal lawlessness" granted to the police.

Other authors, such as English historian Robert Reiner, have voiced suspicion towards more normative aspects of the PUFP. Reiner considers such concepts as policing by consent and demand-side force (Reiner 1992: 139–146) as delusive. Reiner's critique targets an entire trend of thought in Anglo-American social science which, like Michael Banton for instance, concludes that: "The policeman on patrol is primarily a 'peace officer' rather than a 'law officer'. Relatively little of his time if spent enforcing the law in the sense of arresting offenders; far more is spent 'keeping the peace' by supervising the beat and responding to requests for assistance" (Banton 1964: 127).

Reiner, on the contrary, holds that if policing were done out of consent, there would be no need for resorting to force. Contemporary policing, however, is increasingly focusing on force and coercion. While most cases of domestic violence used to elicit mere verbal warnings, they are legally repressed today, and the police shall use coercion. Hence, to Reiner, the core of the police concept is not the potential capacity to use force, but "order maintenance," i.e., the capacity to prevail over individuals as well as crowds. In both the USA (Kraska and Kappeler 1997) and Europe (Porta and Fillieule 2004), policing organizations appear to be increasingly militarized, especially for crowd control operations, as opposed to having their role restricted to social work and peacekeeping. Neither peace nor consent are given; actually, they might be might be further removed today than ever before.

New Perspectives: Consent or Conflict?

In a series of interviews with Brodeur, Egon Bittner has elaborated upon his concept of police (Brodeur 2007). His approach was an etymological one, strongly associating the concept of "police" to the Greek etymology "polis," meaning "city." In this perspective, what the police are dealing with is "a 'polis' that already redefines the minimal state, where it literally consists of the organization of conditions of urban life." In this setting, what then the state provides is "creating conditions for the orderly coexistence of strangers" (Bittner in Brodeur 2007: 111).

With such a definition, Bittner knows he is getting exposed to a conceptualization of the police that is not consensual but agonistic, i.e., a conception of policing that no longer rests on trust as a necessary preamble to any society but on conflict understood as a key element of said society. In this perspective, policing could be said to be defined no longer by its means but by its ends, and its primary end or finality is to maintain order, i.e., ensure the coexistence of social groups in the city, in other words to perpetuate the (social or economic) domination of a given group over minorities. This perspective is very much in line with minority threat theories, which are quite popular in the USA and have used econometric analysis to show a strong correlation between police strength and the presence of black minorities (Kent and Jacobs 2005). The notes published by African-American writer James Baldwin, in which the police patrolling in urban ghettos are compared to "an occupying soldier in a bitterly hostile country," are one in many testimonies of this (Baldwin, More Notes of Native Son, 1962, quoted by Sklansky 2005: 1772). In Europe, historical research usually focuses on the riot-control origins of policing organizations, whose primary goal was to control the urban proletariat on behalf of the "peace-loving propertied classes" mentioned by Silver (1967).

In this perspective, the concept of police is correlated to supply, i.e., the state as the organization that perpetuates the domination of a given group over the others. In Bittner's words, "the thorny question is whether police force can be justified on the supply side" – in Brodeur 2007: 118. It is in that spirit that Peter K. Manning, while still claiming to get his inspiration from a Bittner-inspired concept of police, offered the following conceptualization: "The police as an organization in Anglo-American societies, constituted of many diverse agencies, are authoritatively coordinated, legitimate organizations (...). As such, they require compliance to command from lower personnel and citizens and the ability to proceed by exception." (Manning 2010: 68). Clearly, the order that the police are supposed to maintain is not just any order, or even the status quo, but a substantially defined one. In a nutshell, the police embody one order (the political order of the state) and defend another (the social order as is). Twice anointed, the police are "sacred": to Manning, "the police engage in a form of magic" (2010: 59), and do so by embodying a transcendental entity – public authority, the state, the collective – and deploying an imagery of self-sacrificial symbolism (whose most visible expression is the ritualized death of a policeman – see Manning in this volume). To Loader and Mulcahy (2007), this political disposition translates into local culture-abiding behavior: social order is not created by the police, but simply consolidated by them.

To perform these political functions, the police may resort to exception. This notion originates in the work of German jurist Carl Schmitt, a dissident disciple of Weber and a controversial as well as influential figure, given his full allegiance to the Nazi regime. The notion of exception is useful to get a proper grasp of the specific nature of the police – both inside and outside the law – and to explore the concept of police in greater depth (Jobard 2012). Because it is unequal, social order allows the deployment of unequal policing. In relegated social spaces, where demand for policing is strongest, not only is policing supply scarcest, it is also the least democratic of all: ruthless, biased, undisclosed, and unfair. Most importantly, given the lack of resources and credibility of residents in these areas, police action is seldom likely to be challenged in court. In that sense – and barring exorbitant displays of police violence or irrefutable evidence such as video footage (Skolnick and Fyfe 1993) – police in these areas may act without risk of being disciplined by public justice, i.e., in an exceptional legal space. As summed up by Philip Stenning, "while public police are theoretically accountable for excessive uses of force through the processes of the criminal and civil law, it is well known that in practice it is very difficult to obtain either criminal convictions or civil judgments against public police officers for such wrongdoings" (Stenning 2000: 337).

According to Schmitt, however, any entity that may behave this way in an exceptional situation is nothing short of the "sovereign," i.e., the owner of the political, who designs the political borders of a society, defines who the enemy is in this society and who the friend is (Schmitt 1932). Hence, exceptional actions – i.e., decisions made with no risk of punishment in exceptional circumstances or spaces – are the actions that, according to Schmitt, assert the political in our societies. In this perspective, assuming and recognizing that the police has a capacity to act in the context of exception, then the police can only be said to be the institution that brings to modern, rational, bureaucratic orders an irrational, passionate, untamed dimension, which is that of the political.

The Many Expressions of the Concept of Police

The way of the concept of police in social science has been a meandering one. Some contend that policing has more to do with social work than force nowadays. To others, force is just one element in the bigger picture, certainly not the whole picture, and policing action is characterized by its unpunished infringements on that element. What characterizes the police, then, is not coercive force but legal lawlessness. To others still, the police have to do with a disposition to order maintenance. In this perspective, police as an institution is rooted in adversity against the dangerous classes and is in charge of protecting the dominant groups. In another acceptation, more grounded in political theory, the concept of police would not be distinct from that of political order: the police institution embodies and defends the political order and can even be said, more specifically, to embody the political and its own particular passion in modern, rational, disenchanted societies. The police, from this point the view, is an institution that, despite being laden with modernity, rationality, and technology, carries along a political essence that escapes the Weberian modernization and disenchantment of the world (Weber's "Entzauberung der Welt").

These various contemporary understandings are rooted in older, more European conceptions, those associated with policing sciences that started to develop in mainland Europe as early as the Renaissance.

The Police: History of a Polysemous Concept

We have been concerned so far with how social sciences may conceptualize policing as it occurs in contemporary societies. Academic usage would rather have had us describe the history and genesis of a concept before examining its most current developments.

Contemporary conceptualization of the police, however, is peculiar in that it has been developing as if detached from historical knowledge. There are two reasons for this. First, the field has been dominated by North American social scientists, whose policing institutions, as we will see, have cultivated some distance with their mainland Europe counterparts during their development. As a matter of fact, however, the latter are what the police concept has been drawing upon ever since the seventeenth to eighteenth century. The second reason is the specific fields of said North American social scientists: being sociologists, ethnomethodologists, ethnologists, and sometimes jurists, their concern has never been to anchor their conceptualizations in the history of political ideas. What was being favored was the conceptualization of whatever was immediately observable. Yet the concept of police does have a history, and this history helps to understand certain aspects of policing today, especially its relationship to the political.

Police Concept and Police Science

Among the seminal texts on the history of the concept of police is Michel Foucault's 1979 University of Stanford lecture, in which the French philosopher intended to describe "the rationality of state power (which) was formulated especially in two sets of doctrine: the *reason of state* and the *theory of police*" (Foucault 1981). For theoreticians of seventeenth- and eighteenth-century European police, "the police includes everything (. . .), sees to the living": "'Polizei' is a positive task: it has to foster both citizens' lives *and* the state's strength (. . .) Life is the object of the police: the indispensable, the useful, and the superfluous. That people survive, live, and even do better than just that, is what the police has to ensure" (Foucault 1981).

Theoreticized a century later by Hegel, "policing" is this government competency that is used to improve the living conditions of the people, perpetuate the population, and give it strength.

Everything falls within the realm of the police: the trade of grain, ideas, and beauty; the movements of men, capital, and cattle; the increase of well-being; the decrease of child mortality, illness, or intemperance; security in the widest possible sense; and the "abolition of disorder" (Knemeyer 1980). In such a context, policing knows no border. The police institution, as defined in "police science" textbooks, is totalitarian in that it aspires to take charge of the entire life of individuals as well as their interpersonal relationships. The policing project, the project of police science craftsmen, is to consider the population as an asset and its growth as a goal. Therefore, in European absolutist states, "police" encompasses repression and prevention, the mores, the culture, and the well-being; "police" is anything related to what is referred to as "social policy" or "social control" (Pankoke 1986).

This is where the connection with the *reason of state* lies. The police are comprised of all the government institutions that contribute to state power towards the well-being of the population. However, the latter may quickly be relegated to the backdrop of the former. The object of policing is knowledge on things and people, the police is "the eye" that public authorities keep on society (L'Heuillet 2007). This is precisely where policing meets the *reason of state*: since the police contribute to the efficiency and profitability of governmental action, they are a reason of the state, supported by new knowledge such as statistics or accounting; since it conceives of no obstacle to the growth of the happiness of constituents, it stops at no means and may resort to force at any time.

From this huge but scattered pool of knowledge that fed "police science" handbooks, one area was particularly successful in its development, i.e., the branch dealing with the forecasting, management, or repression of disorders and irregularities. The mad, the sick, the vagrant, and the young – unmarried and unattached – were increasingly targeted by scientists in these fields (Pasquino 1991). This branch of knowledge bridged the gap with "social policy" as a whole and, all along the nineteenth century, conquered its autonomy within policing knowledge as it is known today: knowledge on crime and its various forms; bolstered surveillance of specific social environments; focus on juvenile crime or prostitution and later on foreigners; development of specific techniques for the tracking of individuals, from passports to physical descriptions through the spoken portrait, followed at the turn of the century, under the influence of Italian Lombroso and Frenchman Bertillon, by the development of various anthropometric techniques aiming to spot and arrest criminals and even "born criminals" (Piazza 2011).

The Shapes of Police Science in the Classical Age

The notion of police was born with the absolutist state. Starting from there, however, various versions of police science started to branch out, even leaving the Anglo-Saxon world aside.

The most significant version, in terms of both the sheer size of its knowledge legacy and its influence on subsequent political philosophy (including Hegel's, already mentioned), is that of German *Polizeiwissenschaften*, sometimes also referred to as "Kameralwissenschaften" or cameralism (named from the prince's "chamber," his cabinet, which points at the secrecy of consultations with or by his advisors). Around the late seventeenth century, in German states, "cameralists" started to shape the body of knowledge that was to lead to modern-day governmentality – as described by Michel Foucault – as well as police science and its focus on maximizing the productivity of both the social body and the state (Wakefield 2009). German cameral science has bequeathed us a huge corpus: more than 3,000 German language books on policing are known to have been published from 1600 to 1800 (Pasquino 1991; Stolleis 1996).

The French kingdom has also produced an impressive number of reports, which however were established by public agents who had more to do with the policing function in its modern sense. Indeed, as early as 1667, the French monarchy had established a police authority in Paris, followed by a nationwide body of police commissioners and inspectors: the reports and other written documents that constitute French policing science are in fact their correspondence. Nicolas Delamare, often mentioned by Michel Foucault, was one of them.

"Policing science" may be an overstatement here, since – as opposed to what was going on in Prussia – this knowledge was neither formalized nor academic. It was practical knowledge on specific urban areas, focusing less on the general economy of the nation than on public order in a given city. Already a motley network of informants was playing a substantial part in the manufacturing of this knowledge on the city and the street.

Strangely enough, policing science never reached the shores of Albion. While a few thinkers have used the term, thus testifying to the existence of a form of police science, they were quick to downplay its scope. Patrick Colquhoun, in his *Treatise on the Police of the Metropolis* (1795), for example, stated that "Police in this country may be considered as a new Science; the properties of which consist... in the Prevention and Detection of Crimes; and in those other Functions which relate to Internal Regulations for the well ordering and comfort of Civil Society." Similarly, here is how Edwin Chadwick introduced the police in the report he produced for Sir Robert Peel in 1929 on "preventive police": "the first great object of a Police, that to which every practical adoption should conduce, is to prevent the commission of crime. The second is, when crime has been committed, to detect and to bring to Conviction the perpetrators of it" (both quoted in Neocleous 1998: 440). In England, this rejection of policing as a state science is but the corollary of the rejection of continental mercantilism and more broadly the notion that the development of society should be mediated by the inquisitive knowledge and unlimited power of the state. In the spirit of the economic "laissez-faire" doctrine, the Anglo-Saxon notion of policing was quickly reduced to institutionalized force.

Still, the linkage between police as an instituted force and police science – i.e., between ancient and modern conceptions of the police – is more than incidental. In our view, it is actually central.

Police Science and Modern Police Agencies

What is the actual kinship between this police of the whole, as conceptualized by classical age police science, and contemporary policemen – patrolling our streets and collecting samples from bloodstains in our TV shows (as they also do in real life, only less frequently)? Both a conceptual link and an institutional historical process exist.

First, let us explore the institutional process. The entire body of knowledge collected by police science, as well as knowledge tools such as accounting or statistics, all contributed to the birth of bureaucracy, this multifaceted government body overseeing whatever pertains to the knowledge required to ensure the prosperity of the state and its constituents. Alongside bureaucracy, the police pursued its development, but did so as an institution that had specialized in one function that of preventing: preventing disorder, preventing crime, and helping justice elucidate crimes. It follows that the police is not a branch of the bureaucracy: while bureaucracy belongs to the sphere of dogma, policing is circumstantial; and while bureaucracy is a vector of Weberian modernization, policing retains an *ancien* regime flavor, grounded in secrecy, dissimulation, and the general techniques of absolute power.

France is a case in point. The French Revolution had no choice but to retain the police force established in 1667 by the absolute monarchy, with its urban brand of policing based on the surveillance of the territory and its dwellers, hinging on secrecy and malice, shielded from the institutions of control that the bureaucracy was starting to implement on its side. The point is that the police, in the minds of the 1789 revolutionaries (the decree on municipal police was promulgated on 14 December 1789), is a preamble to liberty: only enforced public tranquility makes it possible for a constitution (understood as a solemn break from absolutism) to be respected. Tranquility conditions liberty, and proper policing is a practical requirement for implementing a future constitution. Policing is thus circumstantial, paving the way for the rule of law and therefore preceding it. This is a decisive point, as it establishes the link between

Policey and modern police, between classical age concepts and contemporary conceptualizations.

The Police and the Legal Order

The reader may remember Bittner quoted as saying that the police realm of competence begins when something crops up "that-ought-not-to-be-happening-and-about-which-someone-had-better-do-something-now." A few centuries earlier, French philosopher Montesquieu contended that "the business of the Police consists in affairs which arise every instant, and are commonly of a trifling nature: there is then but little need of formalities" (The Spirit of Laws, 1748: §25). Certainly, the classical age understanding of policing and the function of the police as it is conceived of in contemporary sociology are related. Continuity here is best expressed in urgency and the imperiousness of action (in §25, Book 26, Montesquieu gets more specific: "The actions of the Police are quick; they are exercised over things which return every day").

By so doing, with regard to the legal order, policing gets characterized by a singularity: police action cannot be entirely prescribed by the rule of law. Obviously, in contemporary democratic regimes, policing increasingly tends to be framed by legal standards; however, its fundamental characteristic is that it always eludes the law – otherwise the police would be powerless. Prudence is the name of the game (and prudence is one of the most constant traits theoreticians have been associating the police with, ever since the seventeenth century). With prudence, certainly, but also, as dictated by the circumstances, without restraint.

As French MP Limodin put it while discussing some bill of law on policing in 1795: "policing implies, if I may say so, bordering on the arbitrary" (quoted in Napoli 2003: 220).

Policing power thus appears to have been born in a legal interstice, namely, the "measure" (Napoli 2003, 2009). The measure is a legal text that does not really belong in the pyramid of standards; it is the form chosen by the government to regulate the minutest details in the lives of citizens. The measure aligns, regulates, and administrates first and foremost what can be exchanged; in particular, it governs public roads and squares, as well as marketplaces. It will thwart whatever prevents exchanges and trade (crime as well as fraud or swindling – e.g., bad scale calibration or inaccurate bills) and, then, once the predators have been neutralized, will attempt to foster trade, increasing produced and consumed wealth.

This precedence of policing over the rule of law and – to a stronger degree even – the subordination of the very existence of liberty to a policed order is what has prompted German philosopher Walter Benjamin to formulate an "ignominious" concept of police – which, although it has not been very popular with social scientists, has reached fame in philosophy: "(. . .) two forms of violence are present in an institution of the modern state, the police. True, this is violence for legal ends (it includes the right of disposition), but with the simultaneous authority to decide these ends itself within wide limits (it includes the right of decree). The ignominy of such an authority (. . .) lies in the fact that in this authority the separation of lawmaking and law-preserving violence is suspended (. . .). It is lawmaking, because its characteristic function is not the promulgation of laws but the assertion of legal claims for any decree, and law-preserving, because it is at the disposal of these ends. The assertion that the ends of police violence are always identical of even connected to those of general law is entirely untrue. Rather, the 'law' of the police really marks the point at which the state, whether from impotence or because of the immanent connections within any legal system, can no longer guarantee through the legal system the empirical ends that it desires at any price to attain (. . .). Its power is formless, like its nowhere-tangible, all-pervasive, ghostly presence in the life of civilized states" (Benjamin 1921/1996: 242–243).

Such a bifid and paradoxical disposition, which consists in making use of violence within the rule of law but also in making possible the very existence of the rule of law by resorting to violence, has always proved terribly challenging to reformers in their attempts to establish democratic forms of policing. Though the police, by necessity, resist the imposition of external rules, they still have to yield to the rule of law; otherwise, policing becomes pure violence, governed only by the arbitrary, and tips the political regime scale to the side of the police state. In this tension between emancipation and domestication, the relationship between policing and knowledge becomes a key issue.

Police and Knowledge: Knowledge and Culture

The police are a dual institution, harboring both the rule of law and violence. Its relationship with knowledge is of a similar nature. Knowledge and the rule of law – as well as their opposites, ignorance, and violence – are constitutive elements of the concept of police.

First, it should be noted that the police institution is also heir to classical age policy in that one of its tasks is to have knowledge of the society so as to anticipate whatever may threaten it. As early as the seventeenth and eighteenth centuries, surveillance activities have led policing apparatuses to produce considerable amounts of knowledge. Surveillance does not pertain exclusively to "high policing," i.e., the political police. It trickles down to all levels of policing action, such as that of the patrolman engaged in a community policing program, who does his best to get a grasp of the life and people in the area. Knowledge on others is thus a prominent policing skill, so much so that some contemporary sociologists actually consider the police to be "knowledge workers" (Ericson and Haggerty 1997), because of their specific relationship to prevention, and thus to the forecasting, detection, and knowledge of the various risks that threaten society.

As can be seen, the rule of law/violence pair is not the only one at stake in the conception of policing: knowledge also plays a significant part. If the notion of knowledge is to be expanded to a wider dimension – e.g., the cognitive dimension of policing – then the concept of police should include the notion of policing culture and give it prominence. This includes not only surveillance-produced knowledge but also the cognitive environment surrounding policing institutions.

The creation of a judicial police body was debated in Parliament during the French Revolution. The discussions partly hinged on whether this force should be autonomous or, on the contrary, subordinate to judicial power. One of the MPs, although a conservative, once declared that police action should not be restrained by the law and saw no other conceivable boundaries to policing power than the civic mindedness of the police themselves: "Be wise in picking up your police officers, and then let them enjoy the latitude without which their functions are void." (Bon-Albert Briois de Baumetz, quoted in Napoli 2009). Powerlessness in codifying policing with external regulations is answered by the necessary "wisdom" of the agent. This resignation of policymakers when confronted with the police question exactly echoes the following remark by Anglo-Saxons police sociologists, reporting the very words of the police they had been observing: "you can't police by book" or, as stated by a veteran police officer from the city observed in John Van Maanen's seminal study on learning the trade: "There's only two things you gotta know around here. First, forget everything you've learned in the academy 'cause the street's where you'll learn to be a cop; and second, being first don't mean shit around here" (Van Maanen 1978: 306).

The French MP from 1791 considered the policeman's civic mindedness as the only conceivable guarantee that civil rights would be respected by the police institution; the Union City veteran police observed by Van Maanen in the 1970s established that a good policeman must remain impervious to external ideas and in particular those taught at the academy. This is the birth of a central notion, namely, police culture. More than institutional arrangements or legal prescriptions, what defines the police and their institution, in a system where "police discretion" is both necessary and feared, is the culture (Wilson 1968). As a consequence, one of the most insightful concepts for explaining the use of deadly force by US police has been that of "administrative climate," the local culture that actually determines which brand of policing is at work (Sherman 1980). Contrary to what happens with other, more tightly regulated institutions, the occupational environment is not a tempering element of policing: it is a *key* element of policing.

Thus, after having attempted to make urban police forces more democratic by playing either on their technical equipment, the relationship between their top management and elected officials or their organizational systems, US reformers have thought it more relevant to consider, following in the footsteps of the French revolutionary MP, that the "police officer is the agent of change" and to design police reform "in order to embrace police discretion and to find ways to make its exercise more informed, methodical, and collectively self-reflective, rather than to control discretion from above or from outside (the idea behind judicial oversight and civilian review boards)" (Sklansky 2005: 1776–1777). The fact that these attempts have all failed – beyond demonstrating the power of police unions, whose explicit purposes include protecting the immunity of policing agents – clearly shows that the question of knowledge aporetically converges towards the question of the rule of law. In the words of Napoli: "Ignoring legal texts is (...) the ultimate resource of the policing measure. At the end of the day, police action is based on an instance of non-knowledge" (Napoli 2009: 166).

To put this into perspective and relate it to the wider picture, policing could be said to entertain a two-pronged relationship to knowledge. Knowledge is among its key modes of action, one that, ever since the conceptualizations of the seventeenth and eighteenth centuries, has turned the police into a formidable surveillance institution. However, policing has also been known to rely on its occupational or even vocational culture to counter all external regulatory attempts. The policemen's intuition, flair, or common sense are, in their own and their leadership's eyes, their core knowledge asset.

Conclusion : A Dual Concept of Policing

Policing certainly ranks as one of the most challenging social objects to conceptualize. This resistance, of course, owes a lot the secrecy opposed by the police institution to the external gaze. It owes even more, however, to its dual nature. Dualism, a philosophical movement whose early proponents included Descartes and Leibniz, has established that certain phenomena consist in the reunion or coexistence of two irreducible and irreconcilable components. In this perspective, policing is the reunion of two pairs of opposites: in it coexist the rule of law and violence and knowledge and ignorance. What most consistently shapes the concept of police is precisely this irreducible duality.

Recommended Reading and References

Banton M (1964) The policeman in the community. Basic Books, New York

Benjamin W (1996) Selected writings. Harvard University Press, Cambridge, MA

Bittner E (1990) Aspects of police work. Northeastern University Press, Boston

Brodeur J-P (2007) An encounter with Egon Bittner. Crime Law Soc Change 48:105–132

Brodeur J-P (2010) The policing web. Oxford University Press, Oxford, UK

Ericson R, Haggerty K (1997) Policing the risk society. Clarendon, London

Foucault M (1981) 'Omnes et singulatim': towards a criticism of political reason. In: McMurrin S (ed) The tanner lectures on human values, vol 2. University of Utah Press, Salt Lake City

Jobard F (2012) Propositions on the theory of policing. Penal Fields (online publication) Champ pénal/Penal field, Vol. 9, 2012. http://champpenal.revues.org/8286

Kent SL, Jacobs D (2005) Minority threat and police strength from 1980 to 2000. Criminology 43(3): 731–760

Knemeyer F-L (1980) Polizei. Econ Soc 9:2

Kraska P, Kappeler V (1997) Militarizing American police: the rise and normalization of paramilitary units. Soc Probl 44:1–18

L'Heuillet H (2007) Baja política, alta policía: un enfoque histórico y filosófico de la policía. Prometeo, Buenos Aires

Loader I, Mulcahy A (2007) Civilizing security. Cambridge University Press, Cambridge

Manning P (2010) Democratic policing in a changing world. Paradigm, Boulder

Napoli P (2003) Naissance de la police moderne: pouvoir, normes, société. La Découverte, Paris

Napoli P (2009) Misura di polizia. Un approccio storicoconcettuale in erà moderna. Quad Stor 44:523–547

Neocleous M (1998) Policing and pin-making: Adam Smith, police and the state of prosperity. Polic Soc 8(4):425–449

Pankoke E (1986) Von ‘guter policey’ zu ‘sozialer politik’. ‘Wohlfahrt’, ‘glückseligkeit’ und ‘freiheit’ als wertbildung aktiver sozialstaatlichkeit. In: Sachßse C, Tennstedt F (eds) Soziale sicherheit und soziale disziplinierung. Beiträge zu einer historischen theorie der sozialpolitik. Suhrkamp, Frankfurt a.M

Pasquino P (1991) Theatrum politicum. The genealogy of capital. Police and the state of prosperity. In: Burchell G, Gordon C, Miller P (eds) The Foucault effect: studies in governmentality. Harvester Wheatsheaf, Hemel Hempstead

Piazza P (2011) Bertillonnage: savoirs, technologies, pratiques et diffusion internationale de l’identification judiciaire. Criminocorpus, revue hypermédia (online, last consultation 20 July 2012). http://criminocorpus.revues.org/347

Porta D, Fillieule O (2004) Policing social protest. In: Snow D, Soule S, Kriesi H-P (eds) The Blackwell companion to social movements. Blackwell, Oxford, pp 217–241

Reiner R (1992) The politics of the police, 2nd edn. Toronto University Press, Toronto

Schmitt C (1932/2007) The concept of the political. Chicago University Press, Chicago

Sherman L (1980) Perspectives on police and violence. Ann Am Acad Polit Soc Sci 452:1–12

Silver A (1967) The demand for order in civil society: a review of some themes in the history of urban crime, police, and riot. In: Bordua D (ed) The police: six sociological essays. Wiley, New York

Sklansky D (2005) Police and democracy. Mich Law Rev 103:1699–1830

Skolnick J, Fyfe J (1993) Above the law: police and the excessive use of force. The Free Press, New York

Stenning P (2000) Powers and accountability of private police. Eur J Crim Policy Res 8:325–352

Stolleis M (ed) (1996) Policey im Europa der frühen neuzeit. Vittorio Klosterman, Frankfurt-am-Main

Van Maanen J (1978) Observation son the making of policemen. In: Manning P, Van Maanen J (eds) Policing. A view from the streets. Newbery Award Records, New York

Wakefield A (2009) The disordered police state: German cameralism as science and practice. Chicago University Press, Chicago

Wilson J (1968) Varieties of police behavior: the management of law and order in eight communities. Harvard University Press, Cambridge

Concurrent Validity

▶ Self-Reported Offending: Reliability and Validity

Conditional Conviction

▶ Probation and Community Sanctions

Conditional Release

▶ Early Release from Prison

Confessions

▶ Interview and Interrogation Methods Effects on Confession Accuracy

Confirmation Bias

▶ Cognitive Forensics: Human Cognition, Contextual Information, and Bias

Constabularization

▶ Police and the Military Nexus

Construction Site Theft and Burglary

▶ Theft at Construction Sites

Content Analysis

▶ Quantitative Studies on Media and Crime

Contextual Influences

► Cognitive Forensics: Human Cognition, Contextual Information, and Bias

Contingency Management

► Behavioral Management in Probation

Continuity and Change

► Individual Characteristics and General Strain Theory

Control of Police Misconduct

Sanja Kutnjak Ivković
School of Criminal Justice, Michigan State University, East Lansing, MI, USA

Synonyms

Police accountability; Policing the police

Overview

Control of police misconduct is a complex undertaking that includes a series of heterogeneous functions such as establishing supervision and accountability, detecting and investigating misconduct, and cultivating police culture intolerant of misconduct. These functions of control and accountability could be performed by a network of diverse institutions such as the police agency itself, courts, mayors, media, and independent commissions.The traditional approach toward the study of control mechanism has been to classify the institutions performing the control functions into internal control mechanisms and external control mechanisms. The key *internal control mechanisms* (in which control functions are performed by the police agency itself) include the establishment of official rules, the roles performed by the police chief and the police administration, supervisors, and the internal system of control. The key *external control mechanisms* (in which control functions are performed by the organizations, institutions, and individuals outside of the police agency) include the legislature, prosecutors, and criminal courts, civil courts, the Supreme Court, and independent commissions. Lastly, the key *mixed mechanisms* (in which the functions are performed by an outside institution or agency with police officers constituting at least a part of the membership) include citizen reviews and the Commission on Accreditation for Law Enforcement Agencies.

Introduction

Control of police misconduct includes a number of heterogeneous functions, from setting official agency policies and enforcing them, establishing supervision and accountability, detecting and investigating misconduct, and cultivating police culture intolerant of misconduct, to providing resources for control, controlling police agency's own efforts to control misconduct, and limiting opportunities for misconduct. Some of these functions are reactive (e.g., disciplining police officers), while others are preventive (e.g., setting official policies, cultivating police culture intolerant of misconduct).

The functions are performed by a network of diverse institutions such as the police agency itself, courts, mayors, media, and independent commissions. The traditional approach to the study of control mechanism has been to classify the institutions performing the control functions into internal control mechanisms (housed within the police agency) and external control mechanisms (housed outside of the police agency). In addition, there are a few mechanisms of control that have elements of both internal and external mechanisms (those housed outside of the police agency, but potentially having police officers as members).

Internal Mechanisms of Control and Accountability

One of the most important roles in controlling misconduct and achieving accountability should be performed by the police agency itself. Yet, police agencies vary dramatically with respect to the intensity and thoroughness with which they control their employees' conduct, as various commission reports have indicated (see, e.g., Christopher Commission 1991; Knapp Commission 1972; Mollen Commission 1994; Pennsylvania Crime Commission 1974).

Police agencies have many diverse and interconnected ways in which they can strive to control police misconduct and enhance accountability. The control system "starts with the recruitment and selection process and continues with training and supervision, incorporating various aspects of rule establishment, communication, and enforcement that stimulate, allow, or prevent police officers from turning their propensity toward corruption into actual corrupt behavior" (Kutnjak Ivković 2005, p. 68).

1. *Administrative Rules and Policies.* Administrative rules and policies, usually compiled in the police agency's standard operating procedure manual (SOP), are typically made by the police chief. The policies relevant for the control of police misconduct traditionally channel the use of discretion, describe appropriate conduct of police officers, prohibit inappropriate conduct, instruct officers to complete written reports after critical incidents, and require supervisory oversight (see, e.g., National Research Council 2004; Walker and Katz 2008).

   Over the last two decades, there seems to have been an increase not only in the number of police agencies that have official policies, but also in the extent of the official policies. A study by Barker and Wells in 1982 reveals that about one-quarter of the police agencies at the time had no written policies at all, and having a set of written policies seems to have been related to the agency's size. A more recent 2007 Bureau of Justice Statistics survey indicates that, regardless of the size of the population they serve, more than 95 % of the local police agencies had written policies regulating the code of conduct and appearance, and the use of lethal and nonlethal force by their employees.

   In itself, promulgation of rules does not suffice to control misconduct. Indeed, both the ambiguity of the rules and the presence of unofficial rules in conflict with the official ones could lead to misconduct. Various independent commissions (e.g., Knapp Commission 1972; Mollen Commission 1994; Pennsylvania Crime Commission 1974) recorded examples of police administrators creating unofficial rules which trumped the official ones (Kutnjak Ivković 2005). Empirical research suggests that police officers in the police agencies characterized as corrupt were more likely to say that the rules were not clear than the police officers employed in the police agencies relatively free of corruption were.

   Empirical studies (e.g., Fyfe 1979) report the reduction in the use-of-deadly-force incidents in the aftermath of the introduction of the more restrictive use-of-deadly-force rules. The administrative regulation of the high-speed pursuits seems to be directly related to the frequency with which officers engage in high-speed pursuits. For example, the introduction of a more restrictive policy of high-speed pursuits in the Metro-Dade Police Department substantially reduced its number of pursuits. By contrast, the introduction of a more relaxed policy of high-speed pursuits in the Omaha Police Department resulted in a dramatic increase of the number of pursuits the next year. The effect of administrative rulemaking on misconduct has been explored in several other areas (e.g., use of force, domestic violence); the evidence therein is more limited and shows less success (National Research Council 2004, p. 285).
2. *Police Chief and Administration.* The police chief and other highly ranked police administrators are the key to the police agency's successful misconduct control. The nature of their

contributions can span across all aspects of traditional managerial functions such as planning, organizing, coordinating, and controlling.

Although what police chiefs and highly ranked administrators may do could be limited by the mayor and other politicians, the public, media, and police unions as well as by the legal statutes, laws, court cases, and civil service rules, "[the police chiefs] may exert a substantial influence on the recruitment standards, training in ethics, leadership and management style, supervisory accountability and standards, internal control mechanisms, discipline, and rewards" (Kutnjak Ivković 2005, p. 70). Independent commissions (e.g., Christopher Commission 1991; Knapp Commission 1972; Mollen Commission 1994; Pennsylvania Crime Commission 1974) provide examples of police administrators' failure to perform these functions and adhere to the expected standards by turning a blind eye to police misconduct.

If in contradiction with the official rules, the police chief's actual behavior sends a clear message to the police officers (see, e.g., Knapp Commission 1972, pp. 170–171). The majority of the police officers participating in a 2,000 nationwide survey agreed that "a chief's strong position against the abuse of authority can make a big difference in deterring officers from abusing their authority" (Weisburd et al. 2000, p. 6).

3. *Supervisors*. While police chief and the highly ranked police administrators deal with the general issues potentially affecting the whole police agency, first-line supervisors are required to oversee their subordinates. Specifically, they are expected to monitor police officers under their command, review their reports, advise police officers when their performance is less than satisfactory, and file a report when they are aware that police officers had violated the rules (Walker and Katz 2008, p 482). As the overwhelming majority of the respondents (90 %) participating in a nationwide survey of police officers agreed, supervisors' role in preventing misconduct is crucial (Weisburd et al. 2000, p. 6).

Independent commissions that investigated police misconduct in the 1960s/1970s documented cases of supervisors actively taking part in police misconduct (e.g., Knapp Commission 1972; Pennsylvania Crime Commission 1974). More recent independent commissions (e.g., Christopher Commission 1991; Mollen Commission 1994) tended to be more likely to find supervisors turning "a blind eye" or failing to take a stance on police misconduct. The Mollen Commission documented that Michael Dowd (a former NYPD police officer who stole money and drugs, participated in larger drug rings, became a drug dealer, and was eventually caught, tried, and sentenced to 14 years in prison), received glowing performance evaluations from his supervisors who even recommended him as "a role model" for other police officers (Mollen Commission 1994, p. 118). The Mollen Commission further found that supervisors in the NYPD regularly failed to review the search and seizure forms, arrest forms (p. 29), and overtime payment forms (p. 39) and question line officers under their supervision who were falsifying these forms. Similarly, the Christopher Commission (1991) concluded that the LAPD supervisors omitted to monitor the subordinates' racist discussion on the MDT. In the aftermath of the Rampart Division scandal, the LAPD's Board of Inquiry documented failures in the supervisors and wrote that "the practice of officers printing or signing a sergeant's name to booking approvals and arrest reports was a particularly glaring illustration of poor CRASH supervision" (Los Angeles Police Department 2000, p. 61).

Independent commissions (e.g., Christopher Commission 1991; Los Angeles Police Department 2000; Knapp Commission 1972; Mollen Commission 1994) reported that supervisors in these police agencies not only were overlooking line officer misconduct, but also were not held accountable by the police chief and the highly ranked police administrators for their failure to supervise. In fact, the Mollen Commission (1994) concluded that the supervisors who were proactive and actually reported misconduct of their subordinates suffered informal punishment for their proper conduct.

Since 1994, CompStat (Computer Statistics) has been used not only as a mechanism to reduce crime rates, but also as a mechanism to increase supervisor accountability. Developed by Bill Bratton and his top administrators in the NYPD, CompStat centers around regular meetings in which the top administrators require of middle managers to discuss crime problems in their areas and hold them accountable for its success or failure in dealing with the problems. CompStat became an instant hit and by 1999, one-quarter of the sample of police agencies with 100 or more employees have already implemented CompStat, and an additional one-third planned to do so in the future (Weisburd et al. 2003). However, the concept of accountability, as envisioned by CompStat, relates to the changes in the crime levels and does not extend to the control of rule-violating behavior of the subordinates. Moreover, whereas CompStat could operate as an accountability mechanism for the middle managers who participate in the CompStat meetings (e.g., Silverman 1999), it does not operate as an accountability mechanism for first-line supervisors who do not participate in the CompStat meetings (National Research Council 2004, p. 188).

4. *Internal System of Control*. Each police agency has a separate system set up within the agency to investigate allegations of police misconduct. Depending on the police agency size, overall public service demands, available resources, and number of complaints, the internal system of control can vary from a single police investigator assigned on the case-by-case basis to an elaborate network of police investigators housed in the main office and additional field offices. According to the 2,000 LEMAS, the majority of local police agencies (79 %) and state police agencies (84 %) had separate, permanent internal affairs offices. These internal affairs offices are traditionally directly accountable to the police chief.

The primary purpose of the internal affairs unit is fact-finding, while some other unit or individual (e.g., immediate supervisors, a board, citizen review) makes the decision about the case. The work of the internal affairs units could be classified as either reactive (e.g., investigation initiated by a complaint) or proactive (e.g., integrity tests). In reality, proactive investigations are more of an exception than the rule (Kutnjak Ivković 2005).

The official procedure typically starts with the filing of a complaint by a citizen or a police officer, or with the submission of a report by a supervisor. During the investigation, the police investigators collect and examine physical evidence, interview witnesses, analyze records, and interview the accused police officer. As a consequence of the *Garrity* ruling (*Garrity v. New Jersey*, 385 U.S. 483 (1967)), police agencies separate investigations into a criminal investigation and an administrative investigation. In the administrative investigation, the accused police officer is not allowed to claim the Fifth Amendment privileges and is ordered to answer the questions truthfully. At the end of the investigation, the police investigator completes the report and turns over the evidence to the decision-making body. The disposition of the complaint is made either by the police officer's chain of command review or by an administrative board through a disciplinary hearing. The nature of the hearing (e.g., adversarial) and the rules regulating it vary greatly across agencies. When the complaint is found sustained (i.e., there is sufficient evidence to prove that the police officer engaged in the rule-violating behavior), the decision-making body will mete out discipline. The severity of the discipline likely depends on the seriousness of violation, aggravating and mitigating circumstances, and the police officers' prior history.

Complaint rates vary greatly across the cities; Pate and Hamilton compared the complaint rates per 100 sworn officers across the six largest police agencies (1991, p. 144) and reported variation in the complaint rates from 5.5 in Philadelphia to 36.9 in Houston. A 2002 survey of large local and state police agencies shows that the rates for the use of force complaints varied from 1.3 for the state agencies to 9.5 for municipal agencies (Hickman 2006, p. 2). Typically, police agencies sustain between 0 % and 25 % of all complaints (Dugan and Breda 1991; Hickman 2006, p. 4; Pate and Fridell 1993, p. 42; Perez 1994), with 8–10 % being typical or average (e.g., Pate and Fridell 1993; Hickman 2006).

It seems that complaint rates are influenced by so many factors other than the actual rate of misconduct that these rates likely can tell more about the agency's openness to complaints, the ease with which citizens could file complaints, and the level of legitimacy of the police, than about the actual level of misconduct. Wide differentials of complaint rates across the cities prompted scholars to recommend extreme caution in cross-agency comparisons (e.g., Hickman 2006; Klockars et al. 2006; Pate and Hamilton 1991; Perez 1994; Walker 2001).

Having the system of internal control put in place does not guarantee that it will operate properly. Research studies and reports by independent commissions (e.g., Christopher Commission 1991; Knapp Commission 1972; Mollen Commission 1994; Pennsylvania Crime Commission 1974) documented many failures, from not establishing written guidelines and providing resources and manpower to the internal affairs units, to failing to investigate complaints, ignoring information, and openly hiding complaints (see, e.g., Christopher Commission 1991; Knapp Commission 1972; Mollen Commission 1994; Pennsylvania Crime Commission).

Early warning systems have recently been added to the internal system of control in a number of police agencies. Early warning systems are proactive in nature. Police agencies collect information about individual police officers, such as use of force reports, accident reports, complaints, financial records, and other information (see, e.g., Walker et al. 2000; Walker and Katz 2008). Once a police officer has a certain number of these items (e.g., sustained complaints), red flags are raised. The police agency engages in an intervention (typically informal counseling by the supervisor or retraining) and post-intervention (monitoring the police officer for a certain period of time; Walker et al. 2000). The idea behind the early warning systems rests on the premise that if potential problems are spotted and addressed early, they would not become serious problems later. The limited empirical research (e.g., Walker et al. 2000) implies that the early warning systems might be effective in terms of reducing the number of use of force reports and citizen complaints by the officers subject to early warning systems.

External Mechanisms of Control and Accountability

A number of institutions an organizations housed outside of the police agency take part in the control of police misconduct.

1. *Legislature*. The role of the legislature, both at the federal and state level, is to enact the laws that criminalize police misconduct and/or establish civil liability for police misconduct. First, the rules of various federal and state criminal statutes (e.g., criminalization of theft, assault, murder), enacted by the legislature, could be applied to every citizen, including police officers, unless the police officers can successfully claim that they performed the activity as part of their legitimate police work.

   Second, federal and state codes also establish certain crimes which only public officials can commit (e.g., bribery of public officials and witnesses, Title 18 of the U.S. Code, Section 201, 1999; extortion by public officials, Title 18 of the U.S. Code, Section 872, 1999; deprivation of civil rights, Title 18 of the U.S. Code, Section 242, 1999). The legislation includes such acts as the Hobbs Act (Title 18 of the U.S. Code, Section 1951, 1999) and the

Racketeer Influenced and Corrupt Organizations Act (RICO; Title 18 of the U.S. Code, Sections 1961–1968, 1999). The advantage of RICO relative to other penal codes is the possibility to prosecute for a "pattern of racketeering activity" (Title 18 of the U.S. Code, Section 1961, 1999) and, therefore, to present the picture of systematic exploitation of the official position, instead of being forced to prosecute each and every act separately.

Third, federal and state statutes also establish the grounds for civil liability for deprivation of civil rights (e.g., deprivation of civil rights, Title 42 of the U.S. Code, Section 1983; conspiracy to interfere with civil rights, Title 42 of the U.S. Code, Section 1985).

Fourth, the Violent Crime Control and Law Enforcement Act of 1994 (Title 42 of the U.S. Code, 1994) authorizes the Attorney General to collect the data about the use of excessive force by the police officers and publish the summary. Consequently, the Bureau of Justice Statistics and National Institute of Justice started sponsoring the national police use of force database, collected by the International Association of Chiefs of Police and the police public contact survey, as well as expanded the LEMAS survey to collect the data on formal citizen complaints about the use of force (Hickman 2006).

One of the problems with the legal codes is that the rules are vague and open for ambiguous interpretations. The Knapp Commission reported that in the 1970s, the rules governing the operation of bars were sound in principle, but were "so vague and ill-defined that they lend themselves to abuses in practice" (Knapp Commission 1972, pp. 147–148). Furthermore, the rules could be difficult to enforce. In the words of Patrick Murphy, a former Commissioner of the NYPD (Pennsylvania Crime Commission 1974, p. 420), "[b]y charging our police with the responsibility to enforce the unenforceable, we subject them to disrespect and corruptive influences, and we provide the organized criminal syndicate with illicit industries upon which they thrive." However, once the legal codes are enacted, the crucial question is whether they are enforced (see Criminal Courts; Civil Courts).

2. *Prosecutors and Criminal Courts.* The criminal justice system enforces the laws enacted by the legislature and provides the boundaries of police officers' acceptable/unacceptable behavior by holding them criminally responsible for police misconduct. Like other citizens, police officers could be tried for "traditional crimes" like robbery, assault, or theft. In addition, unlike other citizens, police officers are public servants and could also be prosecuted, tried, and convicted for crimes that require that the person who committed the crime is a public employee (e.g., extortion, 18 U.S.C. 1951; criminal liability for deprivation of civil rights, 18 U.S.C. 242).

There is no systematic database, either at the federal or the state level, that contains the data about all prosecutions and convictions for police misconduct. At the federal level, there were fewer than 50 convictions each year for law enforcement corruption in the period from 1992 to 1998 (Kutnjak Ivković 2005, p. 59), At the state level, the convictions are almost equally sparse and seemingly suggest that corruption is not a problem. However, the findings by the independent commissions contradict these reports. The Knapp Commission (1972) reported that corruption in the NYPD was widespread, ranging all the way to the rank of a captain. Yet, the prosecutors initiated the prosecution in only about 30 cases per year. Furthermore, severe sentences are unlikely; only one out of five police officers prosecuted, tried, and convicted received a prison sentence of more than a year (Knapp Commission 1972, p. 252). The Pennsylvania Crime Commission (1974), which investigated the extent and nature of corruption in the Philadelphia Police Department and found widespread corruption in the agency, reported that there were on average only seven arrests per year for corruption (Pennsylvania Crime Commission 1974, p. 446). The data also suggest that criminal prosecutions for the use of *excessive force*, either federal or state, are rare (e.g., Cheh 1995, p. 241; Human Rights Watch 1998) and that the conviction rates are low. Out of about 8,000 police misconduct complaints that the Department of Justice receives annually (Cheh 1995, p. 241), about 3,000 are investigated and only 50 presented to the grand jury.

3. *Civil Courts*. Civil courts enforce the laws enacted by the legislature and provide the boundaries of acceptable/unacceptable behavior by holding police officers, police departments, and city administrators civilly liable for police misconduct. Compared to the number of contacts between the police and citizens, civil lawsuits are rare. Yet, the amounts paid to the successful plaintiffs vary across the country, from an average of 1.6 million annually in Cincinnati to an average of 35.8 million annually in Los Angeles (Kappeler 2006, p. 10).

   The legal basis for the civil lawsuit against the police could be found in both the federal codes and the state codes. Section 1983 lawsuit (Title 42 U.S. Code 1983) is considered a crucial tool in the fight against police misconduct. In *Monroe v. Pape* (365 U.S. 167 [1961]), the Supreme Court determined that police officers could be held liable for deprivation of the 4th Amendments rights under the civil rights statute. In *Monell v. Department of Social Services* (436 U.S. 658 [1978]), the Supreme Court established that municipalities could be held liable for police misconduct if it was pursuant to the agency's policy or custom. Thus, these two decisions allowed citizens to sue both individual police officers and police departments, and according to the estimates, the number of the Section 1983 lawsuits increased dramatically since the 1960s (Cheh 1995, p. 250).

   As a consequence of the Supreme Court decision in Lyons (*City of Los Angeles v. Lyons*, 461 U.S. 95 [1983]), citizens can obtain only compensatory and punitive damages (Cheh, 1995, p. 255). The effect of civil lawsuits on individual police officers is limited; the compensatory damages (and sometimes punitive damages) resulting from the lawsuit are typically covered by the city government. The effect of civil lawsuits on the police agencies is limited as well. Empirical research covering several jurisdictions shows minimal effect on police agencies. The Los Angeles County Sheriff's Department seems to be an exception from this rule; in 1993, the Office of the Special Counsel – a form of citizen review – was established to investigate problems, recommend reforms, and reduce the costs of litigation. The reports imply positive changes.

   The situation has changed with the enactment of the 1994 Violent Crime Control Act (42 U.S.C. 14141, 1994), which has authorized the Department of Justice to act as a plaintiff and sue a police department when there is "... a pattern or practice of conduct by law enforcement officers ... that deprives persons of rights, privileges, or immunities secured or protected by the Constitution." The most recent data available by the Department of Justice (January 31, 2003) indicate that the investigation of 14 police agencies is ongoing, that, out of the 10 lawsuits, 4 ended with consent decrees and 6 with out-of-court settlements (Department of Justice 2010). The settlements or consent decrees require of the police agencies to engage in systematic and widespread reforms of the police agency, such as revising the use of force reporting system, establishing the early warning system, revising the complaint procedures, and improving training. In addition, each police agency is assigned a court-appointed monitor to ensure the implementation of these changes. Existing research did not examine the overall effect of these "pattern or practice" lawsuits on police agencies. The Vera Institute, serving as a monitor for the Pittsburgh Police Department, generally reports that the police are on track with the required changes (Vera Institute of Justice 2002). On the other hand, the reports concerning police agencies in Los Angeles and Washington D.C. indicate that they failed to meet some of the deadlines (Walker and Katz 2008, p. 503).

4. *The Supreme Court*. The chances that the Supreme Court will hear any particular case are miniscule (the Supreme Court grants *certiorari* in less than 5 % of the cases filed). Compared to the lower-ranked trial courts, the Supreme Court, as the highest court of the country, has a wider role to play. The effect of its decisions is far reaching, well beyond the specific case under consideration. Thus, by establishing a nationwide precedent for future court cases, the Supreme Court performs not only a reactive role, but also a potentially even more important, preventive role.

   The relevant Supreme Court cases regulate proper conduct (e.g., the Fifth Amendment warnings given to the arrestees, as was decided in *Miranda v. Arizona* (372 U.S. 436 [1966])), as well as police misconduct (e.g., exclusion of illegally obtained evidence, as was decided in *Mapp v. Ohio* (367 U.S. 643 [1961]). The cases may focus on substantive law (e.g., sentencing for the violation of King's constitutional rights under color of law, as was decided in *Koon v. United States* (518 U.S. 81 [1996])) or procedural law (e.g., prohibition of a denial of the Fifth Amendment rights to a police officer during a criminal case, as was determined in *Garrity v. New Jersey*, 385 U.S. 483 (1967)).

   Empirical research typically explores the influence of a specific landmark decision on police officer behavior. One such decision is *Miranda v. Arizona* (372 U.S. 436 [1966]), in which the Supreme Court held that a confession obtained during custodial police interrogation constitutes a violation of the 5th Amendment right against self-incrimination, unless the police provided specific warnings to the persons that they have the right to remain silent, that anything they say could be used against them, and that they have the right to counsel. The Miranda warnings have been routinely publicized in TV shows and motion pictures and have become a part of the American popular culture (see, e.g., *Dickerson v. United States*, 530 U.S. 428 [2000]). The early research, already in progress when the Miranda decision was made, suggested that police officers rarely gave Miranda warnings. Later studies (Leo 1998) showed that the warning was issued routinely, but the style in which the warning was issued tended to be superficial. The most recent study (Leo 1998), relying on direct observation, confirmed that the police issued the Miranda warnings in about 96 % of the cases and thus showed that the Miranda decision has had a long-term effect on police behavior.

5. *Independent Commissions*. Independent commissions are established in the aftermath of a critical incident which resulted in a scandal (e.g., the Christopher Commission was established in the aftermath of the worldwide airing of the Rodney King beating tape; the Mollen Commission was established in the aftermath of the arrest of six police officers). The political pressure created pushes the mayor or the city manager to create the commission, composed of prominent community members and potentially policing scholars, with the purposes of investigating the nature and extent of police misconduct and proposing recommendations for the changes in the police agency. Prominent examples of independent commissions include the President's Crime Commission (1967), Knapp Commission (1972), and Pennsylvania Crime Commission (1974).

   The work of independent commissions has the potential to reach beyond the individual case; they are in a position to establish the standards that could affect the way policing is done across the country (see, e.g., National Research Council 2004; Walker and Katz 2008). Yet, independent commissions are temporary and have no power to enforce their own recommendations (e.g., Kutnjak Ivković 2005; National Research Council 2004; Walker and Katz 2008). For example, although one of the recommendations by the Christopher Commission (1991) was the establishment of an early warning system in the LAPD, more than a decade later, the

LAPD did not have the early warning system put in place (Walker 2005, p. 179). Furthermore, the work of independent commissions could be severely affected by the lack of political independence (e.g., Pennsylvania Crime Commission 1974), insufficient legal authority (e.g., Knapp Commission 1972, p. 44), and inadequate resources (e.g., Pennsylvania Crime Commission 1974, p. 762).

Mixed Mechanisms of Control and Accountability

The mixed mechanisms of control and accountability share some characteristics of external mechanisms (e.g., housed outside of the police agency) and some characteristics of internal mechanisms (e.g., have police officers as members).

1. *Citizen Reviews.* Citizen reviews are established with the purpose of providing an independent review of citizen complaints. Yet, empirical research suggests that about 23 % (15 out of 65) of citizen reviews have police officers as members (Walker and Kreisel 2001). This idea is not new; although it was first developed in the 1960s, the majority of the existing citizen reviews were established in the last two decades (e.g., Walker 2005). Walker (2001, p. 6) claims that by 2,000, "[o]ver 100 different agencies exist, covering law enforcement agencies that serve nearly one-third of the American population, and they are found in about 80 % of the big cities of this country."

   Walker has classified citizen reviews based on functions they perform (Walker 2001). Only about 34 % of citizen reviews conduct the initial fact-finding completely independently from the police (Class I citizen review; Walker 2001). An additional 46 % provide input in the police investigation (Class II citizen input; Walker 2001; Walker and Kreisel 2001), and 17 % perform the function of the appellate review once the police investigation is completed (Class III citizen monitors). Lastly, Class IV citizen reviews ("citizen auditors"), which constitute only 3 % of the citizen reviews, "do not investigate individual complaints, but are authorized to review, monitor, or audit the police department's complaint process" (Walker 2001, p. 62).

   The focus of Class I to Class III citizen reviews is on individual cases. Even when they go beyond the single case and engage in the policy review as well, their work is bound by the issues raised in the cases. On the other hand, Class IV citizen reviews, which are authorized to review and monitor the police agency's own complaint system, have a greater potential to provide meaningful feedback to the police agency. The San Jose Police Auditor and the Special Counsel to the Los Angeles County Sheriff's Department are evaluated in the literature as more successful auditors, while the Seattle Police Auditor and the Albuquerque Independent Counsel are evaluated as the less successful (Walker 2005, pp. 165–166).

   Empirical research exploring the effectiveness of citizen reviews is limited. Whereas Kerstetter and Rasinski (1994) report that the level of public confidence in the complaint process should increase with the establishment of a citizen review, Sviridoff and McElroy (1989) find that both citizens and the police officers were unhappy with the New York City Civilian Review Board and thought that the board was biased against them. Another avenue of research tried to compare the citizen review outcomes in individual cases with the outcomes of individual cases decided by the police agency itself. A serious challenge with this type of empirical research is finding matching cases. Hudson (1972) compared the work of a citizen review ("Police Advisory Board," which has subsequently been abolished) with the police investigation in Philadelphia and concluded that the citizen review sustained a *lower* percentage of complaints than the internal affairs unit did. However, the types of the cases handled by two institutions were not similar.
2. *CALEA*. The Commission on Accreditation for Law Enforcement Agencies (CALEA),

established in 1979, is an independent agency housed outside of the police. It is composed of 11 police officers and 10 community members (e.g., judges, professors, politicians). Its primary task is to serve as a source of self-regulation for the police.

CALEA publishes standards and model official rules. Some of the rules and standards are required for police agencies seeking CALEA accreditation, while others are only recommended. For example, CALEA now requires the early warning systems for accreditation (CALEA 2006). As of September 2009, CALEA has 463 standards (CALEA 2010). The estimates are that about 500 police agencies have been accredited by 2008 (Walker and Katz 2008, p. 492).

Participation in the CALEA accreditation program is voluntary. The drawback of this approach is that the police agencies in greatest need to change the official rules (e.g., characterized by widespread corruption) are the least likely to do so following CALEA standards. Furthermore, the CALEA standards provide just the minimum standards and do not even try to assess the optimal or ideal standards (Walker and Katz 2008, p. 494).

Empirical research on the influence on CALEA accreditation on misconduct and accountability in the police agencies is very limited. Walker and Katz (Walker and Katz 2008, p. 493) provide examples of how accreditation reduced insurance costs, enhanced the use of force reporting, and improved procedures for juveniles in several agencies.

Related Entries

- ▶ History of Police Unions
- ▶ Law of Police Searches
- ▶ Law of Police Seizures and the Exercise of Discretion
- ▶ Law of Police Use of Force
- ▶ Police and the Excessive Use of Force
- ▶ Police Corruption
- ▶ Police Culture
- ▶ Police Discretion and Its Control
- ▶ Police Lying and Deception

Recommended Reading and References

Barker T, Wells RO (1982) Police administrator's attitudes toward definition and control of police deviance. FBI Law Enforc Bull 51:8–16

CALEA Commission on Accreditation for Law Enforcement Agencies (2006). Standards for Law Enforcement Agencies. Fairfax, VA: CALEA

CALEA Commission on Accreditation for Law Enforcement Agencies (2010). Standards for Law Enforcement Agencies. Fairfax, VA: CALEA

Cheh MM (1995) Are law suits an answer to police brutality? In: Geller WA, Hans T (eds) And justice for all. Police Executive Research Forum, Washington, DC, pp 233–259

[Christopher Commission]. Independent commission on the Los Angeles Police Department (1991). *Report of the independent commission on the Los Angeles Police Department*, Los Angeles, Author

Department of Justice (2010). FAQ. http://www.justice.gov/crt/split/faq.php#primsource

Dugan JR, Breda DR (1991) Complaints about police officers: a comparison among types and agencies. J Crim Justice 19(2):165–172

Fyfe JJ (1979) Administrative interventions on police shooting discretion: an empirical examination. J Crim Justice 7:303–323

Hickman, M. J. (2006). Citizen complaints about police use of force, Washington, D.C, Bureau of Justice Statistics. Available at http://bjs.ojp.usdoj.gov/content/pub/pdf/ccpuf.pdf

Hudson J (1972) Organizational aspects of internal and external review of the police. J Crim Law, Criminol Police Sci 63:427–432

Human Rights Watch (1998) Shielded from justice: police brutality and accountability in the United States. Human Rights Watch, New York

Kappeler VE (2006) Police civil liability, 2nd edn. Waveland Press, Long Grove

Kerstetter WA, Rasinski KA (1994) Opening a window into police internal affairs: impact of procedural justice reform on third-party attitudes. Soc Justice Res 7(2):107–127

Klockars CB, Kutnjak Ivković S, Haberfeld MR (2006) Enhancing police integrity. Springer, Dordrecht

[Knapp Commission] Commission to investigate allegations of police corruption and the city's anti-corruption procedures (1972). Report on Police Corruption. New York, G. Braziller

Kutnjak Ivković S (2005) Fallen blue knights: controlling police corruption. Oxford University Press, New York

Leo RA (1998) The impact of miranda revisited. In: Leo RA, Thomas GC (eds) The miranda debate: law, justice, and policing. Northeastern University Press, Boston

Los Angeles Police Department (2000). Board of inquiry into the rampart area corruption incident: executive summary. http://www.lapdonline.org
[Mollen Commission] New York city commission to investigate allegations of police corruption and the anti-corruption procedures of the police department (1994). Commission Report. Author, New York
National Research Council (2004). Fairness and effectiveness in policing: the evidence. committee to review research on police policy and practices. Wesley Skogan and Kathleen Frydl, editors. Committee on law and justice, division of behavioral and social sciences and education. The National Academies Press, Washington, DC
Pate AM, Hamilton HE (1991) The big six: policing America's largest cities. Police Foundation, Washington, DC
Pate A, Fridell L (1993) Police use of force: official reports, citizen complaints, and legal consequences. Police Foundation, Washington, DC
Pennsylvania Crime Commission (1974) Report on police corruption and the quality of law enforcement in Philadelphia. Author, Saint Davids: PA
Perez DW (1994) Common sense about police review. Temple University Press, Philadelphia
Silverman EB (1999) NYPD battles crime: innovative strategies in policing. Northeastern University Press, Boston
Sviridoff M, McElro JE (1989) Processing complaints against police in New York city. Vera Institute of Justice, New York
Vera Institute of Justice (2002). "Pittsburgh's experience with police monitoring." http://www.vera.org/project/project1_1asp?section_id=2&proejct_id=13
Walker S (2001) Police accountability: the role of citizen oversight. Wadsworth Publishing Company, Belmont
Walker S (2005) The new world of police accountability. Sage Publications, Newbury Park
Walker S, Alpert G, Kenney D (2000) Responding to the problem officer: a national evaluation of early warning systems. National Institute of Justice, Washington, DC
Walker S, Katz CM (2008) The police in America: an introduction, 6th edn. McGraw-Hill, Boston
Walker Samuel, Betsy Wright Kreisel (2001). Varieties of citizen review. In: Roger G Dunham, Alpert G (ed) Critical issues in policing,4th edn. Waveland Press, Prospect Heights, IL, pp 338–355
Weisburd David, Rosann Greenspan, Edwin E. Hamilton, Hubert Williams, Kellie A. Bryant (2000) Police attitudes toward abuse of authority: findings from a national survey. Washington, DC: United States Department of Justice, National Institute of Justice, Research in Brief. http://www.ncjrs.gov/pdffiles1/nij/181312.pdf
Weisburd D, Mastrofski SD, McNally AM, Greenspan R, Willis JJ (2003) Reforming to preserve: compstat and strategic problem-solving in American policing. Criminol Public Policy 2:421–456

Control Theory

Marc Le Blanc
School of Criminology and School of Psychoeducation, Université de Montréal and Boscoville 2000, Montreal, QC, Canada

Synonyms

Community control: criminality, community social organization, cultural organization, deviant opportunities, directs controls; Criminal phenomenon: crime, criminal, criminality; Event control: criminal event, routines activities, presentness, occasions; Personal control: criminal behavior, social control, self-control, models, constraints

Overview

The central notion of our theory is control. Control as the exercise of restraining and directing influences over the criminal phenomenon. The criminal phenomenon involves three embedded levels: the crime, the criminal, and the criminality. In appropriate contexts, control triggers conformity in harmony with social and moral expectations. The restraining and directing influences for each levels of the criminal phenomenon are the result of four mechanisms: bonding, unfolding, modeling, and constraining. Our theoretical statement is an expansion of existing psychological and sociological theories and a mixture of various control constructs. As a consequence, our control theory is integrative. It is also isomorphic because the four mechanisms are expressed in particular ways for each levels of the criminal phenomenon. Finally, our theory is developmental because it uses the laws of thermodynamics to address the question of the explanation continuity and change in the criminal phenomenon.

Introduction

Over the last 20 years, criminology has not witnessed any major theoretical innovations. Numerous theoretical perspectives were available: social disorganization, strain, control, cultural deviance, differential association, social learning, criminal personality, labeling, deterrence, and so on (Shoemaker 2005). These perspectives were elaborations of ideas of early twentieth-century theorists: Quetelet, Lombroso, Durkheim, Freud, Marx, Tarde, and others. Most of the theoretical perspectives explained why individuals became criminals. Durkheim, Freud, Reiss, Nye, and Hirschi have presented control theories. They proposed specific constructs and they accepted the same basic assumptions concerning human nature. Over the last four decades, control theory became and remains the most prominent empirically based criminological theory.

Control theories share four assumptions about human nature and social order. The first assumption is that humans are unsocialized at birth, and during the life course their socialization is never perfect. These conditions favor the emergence and maintenance of a criminal propensity in individuals and communities. The second assumption is that a social order always implies some consensus on values and informal and formal mechanisms of interactions. However, social order is always transient. The third assumption is that communities and individuals are self-serving and have mutual influences on each other. The fourth assumption is that the situation, the individual, and the community modulate each other as much as there are forces internal to each of them.

Our theory is multilayered. It targets three levels of definition and explanation of the criminal phenomenon: the crime, the criminal, and the criminality. Our theory is integrative. It combines constructs from various disciplines and theoretical perspectives. We constructed our theory assuming that different words and terms are different for various theorists but that their theoretical meanings and operational definitions are synonymous. Our theory is isomorphic. The similarity of the explanatory structure from one level of definition of the criminal phenomenon to the other corresponds to the view of the fractal nature of the universe. We propose six generic constructs: two categories of exogenous factors, environment and background, and four control mechanisms, bonding, unfolding, modeling, and constraining. Finally, our control theory is developmental. Most criminological theoretical perspectives do not address the question of the explanation of continuity and change in the criminal phenomenon. We define a contextual perspective that allows for developmental principles like orthogenesis, sensitivity to the original state, and epigenetic probability.

In this entry, we define levels of explanation of the criminal phenomenon: the crime, the criminal, and the criminality. Starting from these layers, we elaborate an integrative and developmental control theory. We define the components of the theory and their structural organization. This paper proposes static and a dynamic formulation of our theory. An extended version of our integrative and developmental multilayered control theory of the criminal phenomenon is available (Le Blanc 1997a, b, 2006, 2009). The Social Sciences and Humanities Research Council of Canada supported the theoretical and empirical research on our integrative and developmental multilayered control theory of the criminal phenomenon over four decades.

The Levels of Definition and Explanation of the Criminal Phenomenon

The French criminologist Jean Pinatel, in a masterly effort to define the bases of criminology, proposed, in 1963, that criminologists should distinguish among three levels of the criminal phenomenon, namely, the criminality, the criminal, and the crime. Each level of the criminal phenomenon has its own perspectives, its own rationales, and its own methods. American criminologists rediscovered this fundamental principle (Hirschi 1979; Short 1985). However, the emphasis has been mainly on two levels, the criminality and the criminal.

The micro-level interpretation of the criminal phenomenon, the *crime*, refers to a small part of the criminal career of a person or of criminality. The criminal event has a beginning, a development, and an end, and the task of criminology is to ascertain the control mechanisms that sustain its appearance in a particular situation.

The meso-level of the criminal phenomenon, the *criminal*, includes the study of his personal biological, psychological, and social characteristics in relation to offending. The dependent variable is any descriptive index of the criminal activity, participation, frequency, onset, and duration or any developmental measures, activation, aggravation, desistance, and trajectories (Le Blanc and Fréchette 1989). Independent control variables are numerous and vary during the life course.

The macro-level of the criminal phenomenon, the *criminality*, is the sum of all offenses committed and criminals at a given time in a particular place, a school, a community, and a society. The rate of criminality is influenced by demographic, economic, political, and other macro control factors. The independent variables are indicators of the state of a society, a community, and a school.

Accepting the need to specify the level of the criminal phenomenon when elaborating theoretical propositions implies that they are distinct and their explanatory factors matched to them. Psychologists distinguish three levels: the milieu, the person, and the situation. Sociologists refer to society, community, and institutions. In consequence, we posit that each level of explanation is only pertinent to a particular level of definition: a crime in a situation, a criminal with his offending, and a rate of criminality of a society.

To respect the logic of the ecological fallacy and the fallacy of reductionism, explanatory variables of one level do not explain the dependent variable of another level. For example, societal variables may affect the person; however, only person variables can influence directly individual offending, and person variables cannot affect the rate of criminality. Each level of explanation is also autonomous. However, explanatory variables are constantly interacting, and it is also implicit from a dynamic point of view that levels of definition reciprocally influence each other. At a specific point in time, the characteristics of a person and of his community, associated with previous behavior, will partially determine subsequent behavior that, in turn, will modify the characteristics of the person and of his community.

The Integrative Multilayered Control Theory of the Criminal Phenomenon

Lenski (1988, p. 168) defines a multilayered theory as “one in which a broadly inclusive general theory establishes a covering principle from which a series of more limited special theories can be derived.” We propose to apply this definition to the levels of definition and explanation of the criminal phenomenon. A broadly inclusive general theory is what Wagner and Berger (1985) call an orienting strategy; it discusses guidelines for understanding the criminal phenomenon, what notions to include and how to relate them to each other. In criminology, there are many such strategies according to Shoemaker (2005). These limited theories present a plausible body of theoretical statements offered to explain a particular layer of the criminal phenomenon. In this entry, we develop a general control theory and limited theories for the explanation of the crime, the criminal, and the criminality.

The Structural Statement of an Integrative Multilayered Control Theory

The Structure of the Generic Control Theory

Much of the recent literature forgets formulations of control theory that preceded Hirschi’s statement (1969) that limits the notion of control to Durkheim’s definition of the bond to society. All control theorists would agree with Empey’s statement that the core of control theories is (1978, p. 207) “…their emphasis upon the idea that delinquent and conformist behavior is a function of the ability of the child to control his antisocial impulses. They start from the assumption that children require training if they are to behave socially. Delinquent behavior will result either if a child lacks the ability for effective training or because he has been trained badly." This statement fits particularly well Durkheim's definition of control in his 1934 book "L'éducation morale" and the last version of control theory that adds the notion of low self-control to the notion of bonds (Gottfredson and Hirschi 1990). As formulated, control theory applies to the meso-level of definition of the criminal phenomenon, the criminal. Some authors developed a control theory that applies to communities, cities, or society: Sutherland, Trasher, Shaw, McKay, and Kornhauser.

Our general control theory of the criminal phenomenon states that, in a favorable context, control mechanisms are operand and will change in harmony with social and moral expectations, as a consequence conformity results and is maintained. Conversely, in an unfavorable context, control mechanisms are insufficient and inappropriate, and the criminal phenomenon emerges and persists.

We use the term "control" according to its third literal definition in Webster's dictionary (p. 245, 3b): "a mechanism used to regulate and guide the operation of a system." The term "control" then refers to the definition of the central notion that Gibbs proposes (1989, p. 23): "…control is overt behavior by human in belief that (1) the behavior increases the probability of some subsequent condition and (2) the increase or decrease is desirable." For Gibbs, the commission of an act or its omission is overt behavior. Overt behaviors manifest themselves in the form of an inanimate thing, human and nonhuman organisms, self-controls and external controls, proximal, sequential, or social. Gibbs argues that this notion of control is central for all the behavioral and social sciences. Gibbs' definition of control is compatible with its literal definition: the exercise of restraining and directing influences. This definition has two advantages. First, it keeps us away from the strictly individual level formulation of the definition of control in the Durkheimian tradition, that is, the bond to society and internal and external constraints used during socialization. Second, its level of abstraction facilitates the formulation of our control theory for the three levels of definition of the criminal phenomenon. Individuals, communities, and events can produce behaviors, acts, circumstances, or conditions that are purposive and desirable. Bonds and constraints are only one type of such overt behaviors.

Gibb's notion of control is also central to psychology. Lytton (1990) uses the umbrella of control system theory to review the literature on child development. Horowitz (1987) proposes a structural/behavioral control model of development. The notion of control is also dominant in criminology. Attachment and supervision for bonding theorists are forms of such regulating mechanisms. We could also argue that arrests, the perceived certainty of a sanction or opportunities, are forms of restraining influences for subsequent offending. They are constructs proposed by labeling, deterrence, and strain theorists. Favorable and unfavorable definitions for differential association theorists and reinforcements for learning theorists are also controls because they are directing influences on criminal behavior.

In our generic control theory, there are four mechanisms of control: bonding, unfolding, modeling, and constraining. Two types of context modulate the action of these mechanisms, the environment and the background. Each category of control mechanism and each type of contexts represent numerous factors that have a potential impact on one particular level of the criminal phenomenon. The definitions of the control mechanisms are the following. Their name is written in capital letters in Fig. 1.

Bonding refers to the various ways by which individuals are held together in a community, in a group, or interpersonally. Unfolding is the natural growth and development toward a desirable state of greater quality, the growth of a community or the development of the person according to social expectations. Modeling is the existence of patterns that can shape conformity, opportunities, models, or occasions that are available in a community or to individuals. Constraining is the regulation of conformity through various direct and indirect restrains; they are limits defined by a community or imposed by the social network of the person. These mechanisms are simultaneously and causally interacting

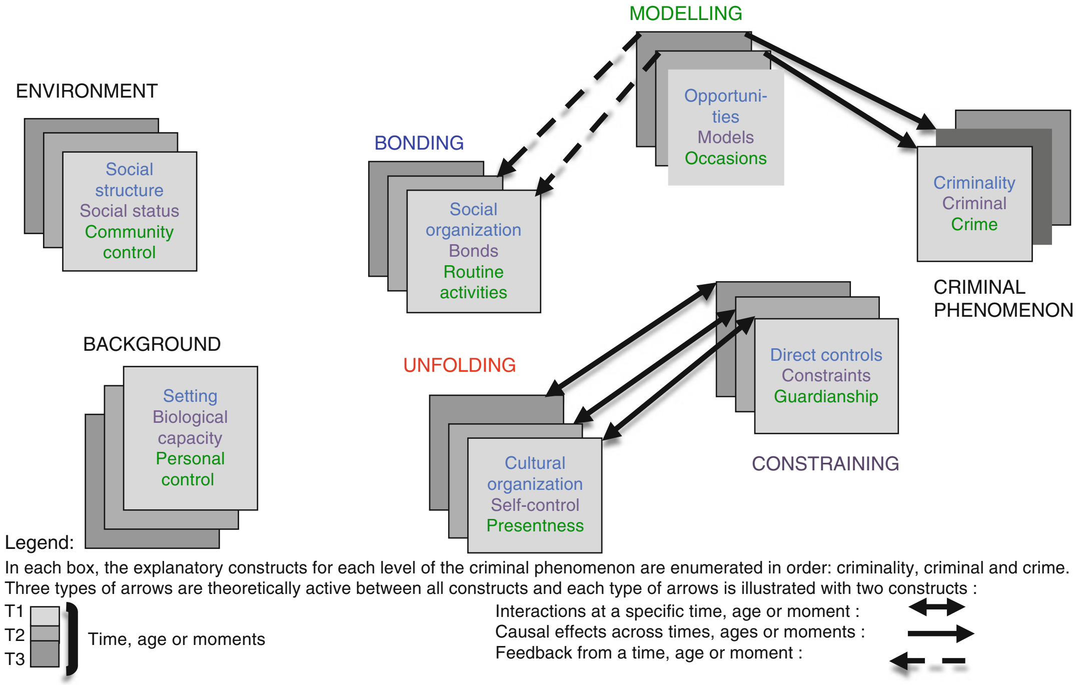
MODELLING
Opportunities
Models
Occasions
ENVIRONMENT
Social structure
Social status
Community control
BONDING
Social organization
Bonds
Routine activities
Criminality
Criminal
Crime
CRIMINAL PHENOMENON
BACKGROUND
Setting
Biological capacity
Personal control
UNFOLDING
Cultural organization
Self-control
Presentness
Direct controls
Constraints
Guardianship
CONSTRAINING
T1
T2
T3
Time, age or moments

Legend:
In each box, the explanatory constructs for each level of the criminal phenomenon are enumerated in order: criminality, criminal and crime.
Three types of arrows are theoretically active between all constructs and each type of arrows is illustrated with two constructs :
Interactions at a specific time, age or moment : ↔
Causal effects across times, ages or moments : →
Feedback from a time, age or moment : ⇠

Control Theory, Fig. 1 The structure of the integrative multilayered control theory of the criminal phenomenon

to produce conformity (all the arrows in Fig. 1). They also have their own life or ontogeny (the superimposed boxes in Fig. 1). This theory is systemic in the sense that it defines a structure, a sequence between its components, as well as reciprocal and directional relationships and feedbacks. It is also a dynamic theory because over time there is continuity and change within the control mechanisms as well as because of their mutual influences.

In Fig. 1, the structure of the control theory is illustrated by the proximity of the mechanisms from the criminal phenomenon. This structure depends on the principle of prerequisites, on the distinction between continuity and change, and on existing empirical knowledge. The theory states that the exogenous factors do not have a direct impact on the criminal phenomenon. They are the environment (the complex of climatic, edaphic, and biotic factors) and the background (circumstances that modulate the control mechanisms). Two of the mechanisms of control, bonding and unfolding, are prerequisites; their impact is indirect on the criminal phenomenon. They are the foundations of the general control mechanism. Without bonds, models cannot be significant and constraints cannot be operand. In consequence, an unbounded community or individual cannot be sensitive to direct controls or influenced by the models. In addition, since the unfolding mechanism refers to a desirable state, this precedes the influence of available models and constraints. The bonding and the unfolding mechanisms modulate the criminal phenomenon through the mechanisms of modeling and constraining. These mechanisms are proximal causes of the criminal phenomenon because they are more specific to the space-time dimension. Their nature changes with time. The bonding and the unfolding mechanisms are in a situation of reciprocal causation at a specific moment. The modeling and the constraining mechanisms are in the same situation. In sum, the bonding and unfolding mechanisms are the foundations and the continuity component of control, while the modeling and constraining mechanisms are catalysts of conformity. We postulate that the four mechanisms of control are in a synergetic relation. They interact to produce an overall level of control of the criminal phenomenon. This synergy, as illustrated in Fig. 1, emerges also from the causal effects, the reciprocal relations, and the feedback effects. The superposed boxes containing the constructs represents the changes on the time, age, and moment dimensions of the evolution of controls.

This structure of the theory applies to the three levels of the definition of the criminal phenomenon. In a multilayered theory, the components are present at all the levels of explanation of the criminal phenomenon. There is isomorphism in the content of the mechanisms of control at the layers of the crime, the criminal, and the criminality.

The Content of the Control Theory of the Criminal Behavior

Reiss' (1951) statement of control theory proposed the distinction between social and personal controls. Hirschi's formulation of the bond (1969) did not include psychological constructs. This deficiency is overcome with the notion of self-control (Gottfredson and Hirschi 1990). Since the 1970s, the vast majority of etiological publications in criminology develops and tests these notions.

At the level of the criminal, conformity to conventional standards of behavior occurs and persists, on one hand, if self-control is robust and the bond to society is firm and, on the other hand, if constraints are appropriate and models are prosocial. This self- and social regulation of conformity is conditional to the biological capacities of the person and his position in the social structure. Alternatively, criminal behavior emerges and continues when low self-control persists, the social bond is tenuous, the constraints are insufficient, and the deviant model is abundant. These causes of offending will be more efficient when the individual has some biological deficiencies and when he comes from a lower social class. The definitions of these six constructs are presented and operationalized in Le Blanc (1997a, b, 2006) and in numerous studies in criminology.

The structure of the individual level control theory can be deduced from many empirical results. In causal modeling term, bonds (attachment and commitment) and self-control are in a situation of reciprocal causation. They are developmental prerequisites of conformity. They are also continuous primary needs during the life course. This is not the case for the modeling (involvement and peers and adults models) and constraining (rules, supervision, punishment) mechanisms. During the life course, the models change – attitudinal and behavioral demands of society for children, adolescents, and adults – and constraints are transformed – from external to internal controls – with age and with the evolution of society. In consequence, modeling and constraining are consequences of the bonding and self-control mechanisms; they are proximal causal factors of conformity. There are also retroactive effects illustrated in Fig. 1. A retroactive effect runs in the opposite direction of the sequence defined by the structure of the theory, for example, from actual individual offending on subsequent state of bonding, self-control, modeling, and constraining.

The Content of the Control Theory of the Criminality

We now move on to the level of criminality. We elaborate our community control theory with the content and the logic used for the personal control theory. Kornhauser (1978) argues that the social disorganization perspective is basically a control theory, even if it is primarily about communities. She indicates that this theoretical perspective has an overwhelming emphasis on the community context. She demonstrates that social disorganization theory includes the major components of a control theory: the strength of social bonds in the community as the foundation of control, the importance of direct controls, the weakness of culture, and the defective socialization to cultural values. Her model is based on three constructs: exogenous variables (economic status, mobility, heterogeneity), cultural disorganization, and social disorganization. She specifies the relationships between the exogenous variables and the structural and cultural community organization constructs without stating clearly the connection between these two forms of community organization. Our version of community control elaborates on this base.

The dependent variable is the rate of criminality in a particular community at a specific moment and its evolution over time. This rate refers either to an overall rate or to rates for specific types of criminal acts. The independent variables, as in the personal control theory, comprise six constructs: social structure, setting, social organization, cultural organization, opportunities, and direct controls. These constructs represent, respectively, the environment and the background contexts, the bonding, the modeling, the unfolding, and the constraining mechanisms (Fig. 1).

Our community control theory assumes that a high rate of conformity to conventional standards of behavior persists when the social organization is sound and the cultural organization robust, when direct controls are efficient, and when there are sufficient legitimate opportunities. This regulation of conformity is conditional on the quality of the setting and on the position of the community in the social structure. Alternatively, a high rate of criminality exists when social disorganization and cultural disorganization are persistent, when direct controls are inappropriate, and when deviant subcultures and opportunities are numerous. These causes of a high rate of criminality in a community will be more efficient when the setting is physically degraded and when the social status of the community is low. The rate of conformity will vary over time and between communities according to changes in the position of the community in the social structure and the quality of the setting and to variations in the levels of social and cultural organization, direct controls, and opportunities.

The construct of social structure refers to the population and urbanization dimensions, the socioeconomic composition of the community, the residential stability of the residents, and the racial and ethnic heterogeneity of the neighborhood. The setting is the isomorphic construct to biological capacity at the level of the individual. It is defined traditionally by the density and crowding dimensions and by the physical deterioration of the inner city. It could also involve such characteristics as the level of pollution, traffic, noise, and so on. Social disorganization refers to the weakness of informal networks; it is the loss of community control over members and the erosion of informal networks, the quality of primary relational networks (intimate informal primary groups: family, friends, neighbors), and the strength of secondary relational networks (broader local interpersonal networks and the interlocking of local institutions). The distinction between structure and culture has long been established, particularly by Parsons. Various notions are present in the criminological literature that refers to cultural organization. There is Sutherland's constructs of economic and political individualism, Sellin's notion of culture conflict, Merton's anomie construct, and Angel's social integration notion. Cultural disorganization refers to the attenuation of societal cultural values as controls. The desirable state for any society is the presence of robust and influential cultural values. In a deteriorated setting and a low socioeconomic status community, various deviant models and numerous illegitimate opportunities are available to residents. The notion of opportunities includes the subcultures, their representative gangs, and the corresponding illegal markets (stolen goods businesses, drugs, prostitution, and so on). However, it also involves black markets, such as undeclared work, and deficiencies in resources for conventional activities, such as work, playgrounds, sport organizations, and art classes. The notion of opportunities also refers to the large availability of suitable targets because of the deficiencies in social and cultural organization and formal and informal direct controls. The existence of illegitimate opportunities and the scarcity of legitimate opportunities encourage the use of repressive direct controls, and they sustain a high rate of criminality. Kornhauser (1978, p. 74) defines direct controls as "...purposive efforts to ensure conformity or limit deviance...." Hunter (1985) proposes three levels of such controls. The private level refers to relationships among peers and adults. The parochial level points to the broader set of local interpersonal networks of neighbors and interlocking of local institutions, such as voluntary organizations, stores, schools, and churches. The public level involves the ability of the community to secure public goods and services (health services, social services, policing, and so on).

We argue that an integrative community control theory contains six constructs: social structure, setting, social organization, cultural organization, opportunities, and direct controls. In addition, based on past theoretical statements and empirical studies, we proposed many interdependencies between these constructs (Fig. 1). There is a consensus among the specialists of community control that the social structure of the community and its setting are exogenous factors to community processes. These sets of variables directly influence the social and cultural organization of the community, the nature of available opportunities, and the nature of direct controls. In consequence, the social structure and the setting do not have a direct impact on the rate of criminality. Communities, like humans, have two primary needs, self-conservation and integration to its larger society. First, the primary need of self-conservation has its origin in the setting and the social structure, and they are the main sources of community organization. As a result, the setting and social structural contexts encourage the bonding (social organization) and the unfolding (cultural organization). These mechanisms imply that, from a particular setting and a specific position in the social structure, the community develops its social and cultural organization. Secondly, the primary need of integration of the community into its larger society leads also to the nature of the social and cultural organization. There is also a consensus that the available opportunities and the nature of direct controls are consequences of the social and cultural organization of the community.

The two primary needs of communities, self-conservation and integration, are persistent during history. This is not the case for the modeling (opportunities) and constraining (direct controls) mechanisms. Over time, the models change and the constraints alter with the evolution of society. For example, gangs of today are different from gangs of previous decades. During history, constraints have been transformed from repressive to humanitarian. As a consequence, Fig. 1 states that the modeling and the constraining mechanisms are influenced by the social organization and cultural organization mechanisms. Figure 1 also indicates that this theory is dynamic and interactional. Reciprocal, directional, and retroactive arrows represent interactions.

The Content of the Control Theory of the Criminal Event

An event control theory has common grounds with the personal and the community control theories. Hirschi (1986) argues that there is no fundamental opposition between rational choice and control perspectives. They share the same image of man as a self-seeking individual. Bursik and Grasmick (1993) state that the routine activities and the social disorganization perspectives complement each other. They argue that the community dynamics relate naturally to the offender/target/capable guardian convergence so important for routine activities theorists. In sum, a criminal event is a function of the community in which it takes place and of the individual who commits it. In these circumstances, it is natural to move to the micro-level explanation of the criminal phenomenon with a control perspective.

The offense control theory assumes that conformity to conventional standards of behavior in a specific situation occurs when routine activities are conventional, when the individual presentness is low, when there is no occasion for the perpetration of a criminal act, and when guardianship is reliable. This regulation of conformity is conditional on the quality of community and personal controls. Alternatively, a crime is likely when the person’s presentness is high, when his routine activities are unconventional, when there are numerous occasions to commit crimes, and when possible targets are unprotected. These causes of the perpetration of an offense will be more potent when the person has a high propensity for crime (low personal control) and lives in a disorganized community (low community control) (Fig. 1).

The dependent variable is a particular criminal event. This event can be various, characterized according to the nature of the crime: the mechanics of the perpetration of the act (planning, use of instruments, accomplices, and so on) and the psychological reaction before, during, and after the event (Le Blanc and Fréchette 1989). The explanatory constructs of community control, personal control, presentness, routine activities, occasions, and guardianship are isomorphic with the constructs of the other two levels of explanation of the criminal phenomenon. Personal and community controls represent the exogenous variables. Routines' activities delineate the bonding mechanism; the presentness component manifests the unfolding mechanism; the occasion's construct defines the modeling mechanism; and the guardianship dimension represents the constraining mechanism.

The construct of routine activities refers to habitually enacted public activities. This construct focuses particularly on the individual lifestyle, the daily activity patterns that disperse the person away from his family, the household situation, and the involvement in conventional activities. According to routine activity theory, these activities will bring the person in contact with numerous targets for the commission of a crime that is improperly guarded. It incorporates public and private and institutional and noninstitutional activities. The construct of presentness is defined as "the idea that people differ in the extent to which they are vulnerable to the temptation of the moment" (Gottfredson and Hirschi 1990, p. 87). It characterize people with low self-control as "...impulsive, insensible, physical, risk-taking, shortsighted, and nonverbal" (p. 90). The Chicago school tells us that in crime-prone communities, there are numerous occasions for the commission of criminal acts. The individual makes a rational choice about the suitability of the target; he evaluates the degree of effort involve, the amount and immediacy of the reward, the likelihood and severity of punishment, and the moral costs. We expect that the more a target seems suitable, the higher the probability of the commission of an offense, particularly if the person's self-control is low and his routine activities are public and noninstitutional. In Cohen and Felson's (1979) routine activity theory, three elements are necessary for the commission of a criminal act: a likely criminal, a suitable target, and the absence of a capable guardian. The absence of guardianship is a proximal cause of a crime, and that situation is likely to be perceived as such when the person has high presentness, when his routine activities are dominantly unconventional, and when there are suitable targets.

Our integrative offense control theory relies greatly on Felson's (1986) discussion of the relationships between criminal choices, routine activities, indirect control, ecology, and social control theories. Figure 1 represents the interdependencies between the constructs from a dynamic point of view (the superimposed boxes) and in an interactional perspective (the directional, bidirectional, and feedbacks arrows). Le Blanc (1997a, 2006) discusses the interdependencies between the constructs of the community control and the personal and event control theories and the personal control and the event control theories.

The Developmental Statement of the Integrative Multilayered Control Theory of the Criminal Phenomenon

The structural statement of our multilayered control theory has defined its constructs and structure. Most criminological theories do not discuss the dynamics of continuity and change of the criminal phenomenon, and the criminological data are rarely longitudinal. Only recently, theories include a developmental perspective. We believe that the time is ripe to further these endeavors through our multilayered control theory. To do so, we used the frameworks provided by theoretical developmental psychology and chaos theory (Le Blanc 1997a, 2006).

Crime is embedded in criminality and individual offending, and criminals are constituent parts of the crime rate. Event control is part of personal and community control, and personal control is a component of community control. Figure 2 represents embeddedness by three parallel spirals. Development is contextual and probabilistic. This means that the influence of the changing context on the trajectory of development is partly uncertain and that development must be defined in terms of "... organism-context

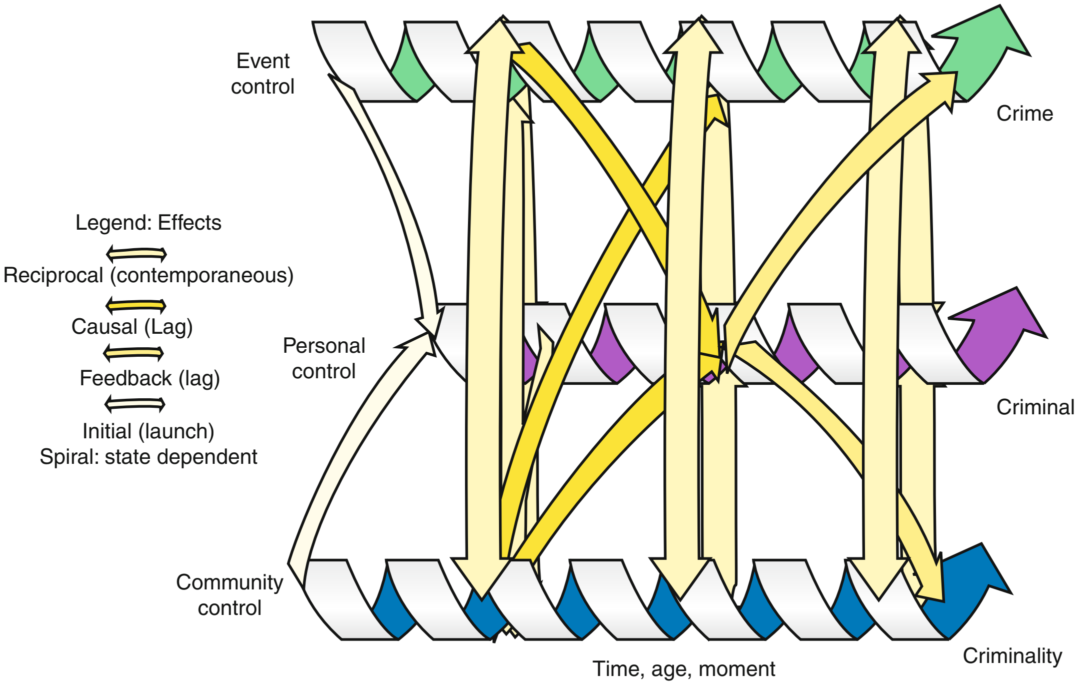

Control Theory, Fig. 2 The developmental multilayered control theory of the criminal phenomenon

reciprocal or dynamic-interactional relations" (Lerner 1986, p. 69). However, the organization and the internal coherence of the organism limit the probabilities of different trajectories. These principles are also those of the chaos perspective (Gleick 1987; Briggs and Peat 1989). This perspective talks about structured randomness, complex systems, nonlinear dynamics, and inner rhythms. In sociology, Fararo (1989) develops that line of thinking under the notion of dynamic social systems and formalizes his theory in terms of the mathematics of nonlinear dynamics.

According to the contextual perspective on development, we can state that the level of community, personal, and event control is specific to a particular time and space. According to chaos theory, we can state that changes at one layer of control will affect changes at the other layers of control. Spirals in Fig. 2 represent the coevolution of the changes at the various layers of control. Arrows from one spiral to the others represent that changes in event control will modify the level of personal and community control and vice versa. Studies in criminology document contextual effects.

Whatever the level of explanation of the criminal phenomenon, its developmental trajectory reflects the orthogenic principle stated by Werner (1957, p. 126): "…whenever development occurs it proceeds from a state of relative globality and lack of differentiation to a state of increasing differentiation, articulation, and hierarchic integration." Psychologists will accept that proposition and sociologists will recognize the pertinence of that statement for societal change. The evolution toward more complexity is governed by two principles: sensitivity to the initial condition and probabilistic epigenesis. Criminologists demonstrated that past criminal activity explains subsequent offending (Le Blanc and Loeber 1998). Individual offending is not the only variable affected by this principle; a significant proportion of the explained variance of explanatory variables – attachment to parents, commitment to education, association with delinquent peers, and others – is accounted for by the past level of the same variables (Le Blanc 2006). Superimposed boxes in Fig. 1 and turns in spirals in Fig. 2 represent this developmental mechanism. As a consequence, each subsystem is partly self-organizing and self-perpetuating. This is the case for the bonding, unfolding, modeling, and constraining systems and their subsystems. Concerning the principle of probabilistic epigenesis, results about individual offending clearly show that there are some normative stages but that the outcome of individual development is only probable, never certain.

Finally, development is interactional as recognized by criminologists (Thornberry 1987) and contextual developmentalists (Lerner 1986). The development of event, personal, and community control implies interactions that are represented in Fig. 2. Interactions take various forms: reciprocal interdependencies among constructs at a specific time; causal relationships between constructs over time, such that constructs will become alternatively independent and dependent variables; state dependencies for each construct; and, retroactive, the impact of the criminal phenomenon at time 2 on the four control mechanisms at time 3.

After reviewing the principles that govern the development of the control of the criminal phenomenon, it is necessary to specify the nature of its course. In our developmental criminology paradigmatic paper (Le Blanc and Loeber 1998), we propose a developmental analysis of individual offending. Le Blanc (2006) argues that we can apply the methods of study for the analysis of within-individual change to the course of event and community control. The course of controls can take the forms of quantitative and qualitative changes, a distinction fundamental for developmentalists and specialists of social change.

Quantitative changes are usually termed “trends” at the community level and “growth curves” at the individual level. First, quantitative changes are the degree of change on any construct of our multilayered control theory. Second, they assess the direction of change, progression or regression. Finally, they also refer to the rate of change, the relationship between the degree of change and time. We can measure the degree, direction, and velocity of change for community control, for example, concerning the social structure: variations in the racial heterogeneity, in its social organization; changes in participation in voluntary organization, in its direct control; and changes in the number of police patrols. We could assess the quantitative measures of change in attachment, self-control, and number of delinquent friends for the personal control. We could describe quantitative changes in guardianship, availability of targets.

Qualitative changes refer to something that is different from what went on before, something that is more complex according to the orthogenetic principle. These changes in nature are habitually subdivided in a developmental sequence that comprises a certain number of stages. When sociologists talk about industrial, postindustrial, developing, postmodern societies, and so on, they refer implicitly to such stages. When social ecologists talk about a community moving from a middle-class status to a working-class status and from a homogeneous underclass status to a gentry’s status, they define implicitly stages of development. Psychologists refer explicitly to stages when they talk about sensorimotor or preoperational intelligence; oral, anal, or phallic functioning; and conventional or postconventional reasoning. In each of these examples, social and behavioral scientists call attention to a universal developmental sequence divided into a limited number of stages. At the level of analysis of events, there is probably a sequence in planning and organization, for example, between a forged check to a sophisticated credit card fraud, from an unplanned breaking and entering to a professional one, and from events of bullying during childhood to gang fights during adolescence and wife battering later on. We assume that a normative developmental sequence exists for the communities, individuals, and events constructs of our multilayered control theory.

Figure 2 represents the dynamics of control of the criminal phenomenon. This figure integrates the mechanisms of the course of development and the developmental processes to the structure of the multilayered control theory of the criminal phenomenon. The dynamics of control are of two categories. First, Fig. 2 represents continuity and change over time; in that case, the figure must be read from left to right. Second, Fig. 2 also shows the interactions between the layers of control; in that case, the figure must be read vertically.

The horizontal reading of Fig. 2 refers to minutes and hours in the event control spiral; to days, weeks, months, and years in the personal control spiral; and to years and decades in the community control spiral. These spirals are metaphors that represent the action of the mechanisms we proposed to specify the course of development of the criminal phenomenon. The time dimension that is associated with the spirals shows orthogenesis. The beginning of the spirals represents the initial condition, while the rest of the spirals introduce the sensitivity to the initial condition. The independent spirals are there to indicate that event, personal, and community control are self-organizing phenomenon. Coevolution is indicated by the placement of the spirals on three dimensions. Finally, along each spiral, there are probabilistic quantitative and qualitative changes.

The vertical reading of Fig. 2 implies that we think in terms of interdependencies between the layers of control of the criminal phenomenon. In our discursive statement of our theory, synergy implied embeddedness and reciprocal relations between the layers of control. In Fig. 2, embeddedness is represented by the fact that the crime spiral is placed in-between the criminal and the criminality spirals. This position indicates that an offense is a constituent part of individual offending and of the rate of criminality and that individual offending is part of criminality. In Fig. 2, interactions are shown by the large arrows; each arrow specifies these interdependencies. Synergy also exists at all points along the time dimension, and as a consequence, the multilayered control of the criminal phenomenon looks like a torus attractor such as proposed by chaos theorists to represent continuity and change (Briggs and Peat 1989; Le Blanc 2006).

Figure 2 also represents the interactional perspective. The discursive statement of our multilayered control theory identifies various types of relations: reciprocal, causal, state dependent, and retroactive. Large arrows indicate reciprocal relations, while thin arrows show causal relations and retroactions. A discursive statement of the relations showed in Fig. 2 would be the following. Insufficient community control and tenuous personal control will diminish the level of event control of a person and a crime is more likely to be committed. Conversely, the commission of a crime will alter the actual level of personal control for that person and modify the actual level of control in his community. Notwithstanding these relationships between layers of control, there are also some independent within layers changes relative to the initial condition. In addition, a change in a person level of event control will alter his subsequent level of personal control and the subsequent level of control in his community. A change in community control will also affect the subsequent level of personal and event control for the person. Finally, a change in the level of a person's personal control will affect the subsequent level of event control for that person and the following level of control in his community. Notwithstanding these changes, the level of the criminal phenomenon also modifies the subsequent level of control.

Conclusion

Our control theory covers three levels of definition and explanation of the criminal phenomenon: the crime, the criminal, and the criminality. We constructed an integrative theory with different words and terms from different theorists, but their theoretical meanings and operational definitions are synonymous. Our theory is fractal because of the similarity of structure from one level of definition of the criminal phenomenon to the other. Six generic constructs and two categories of exogenous factors are defined and organized in a similar structure. The control mechanisms are bonding, unfolding, modeling, and constraining. We define a contextual perspective that allows for developmental principles of orthogeny, sensitivity to the original state, and epigenetic probability. The course of development has quantitative and qualitative dimensions and is nonlinear.

As presented, this formulation of our multilayered control theory has several deficiencies. The first concern is its generic and discursive status. The statement of our control theory is not yet specific enough, and additional work is needed to facilitate an operational formulation of the theory. A theory is also never complete before a formal test identifies its logical inconsistencies. A formalization of the individual level control theory exists (Le Blanc and Caplan 1993), but the same operations should take place for community and event control and our generic control theory as a whole. Only then will it be possible to discuss its testability and start empirical tests. However, numerous limited tests of the structure our control theory do exist in the criminological literature.

Related Entries

- ▶ Age-Crime Curve
- ▶ Anomie and Crime
- ▶ Career Criminals and Criminological Theory
- ▶ Causes
- ▶ Criminal Careers
- ▶ Desistance from Crime
- ▶ Early Chicago School Theory
- ▶ General Strain Theory
- ▶ Genetic Basis to Self-Control
- ▶ Group-Based Trajectory Models
- ▶ Growth Curve Models with Categorical Outcomes
- ▶ Informal Social Control
- ▶ Integrating Rational Choice and Other Theories
- ▶ Interactional Theory of Delinquency
- ▶ Labeling and Deviance
- ▶ Lifestyle Theory
- ▶ Longitudinal Crime Trends at Places
- ▶ Measurement of Self-Control
- ▶ Modus Operandi of Sex Offenders
- ▶ Moffitt’s Developmental Taxonomy of Antisocial behavior
- ▶ Neighborhood Effects and Social Networks
- ▶ Pathways to Delinquency
- ▶ Poverty, Inequality, and Area Differences in Crime
- ▶ Psychopathy
- ▶ Rational Choice Theory
- ▶ Situational Action Theory
- ▶ Social Capital and Collective Efficacy
- ▶ Social Control
- ▶ Social Control and Self-control Through the Life Course
- ▶ Social Control Theory of Sexual Homicide Offending
- ▶ Synthesizing Biological and Social Theorizing

Recommended Reading and References

Briggs J, Peat FD (1989) Turbulent mirror. Harper and Row, New York

Bursik RJ, Grasmick H (1993) Neighborhoods and crime: the dimensions of effective community control. Lexington Books, New York

Cohen LE, Felson M (1979) Property crime rates in the United States: a macrodynamic analysis 1947–1977, with ex ante forecasts for the mid – 1980. Am J Sociol 86:90–118

Durkheim E (1934) De l’éducation morale. Alcan, Paris. English edition: Durkheim E (1961) Moral education (trans: Wilson EK, Schnuer H). The Free Press, New York

Empey LT (1978) American delinquency. Dorsey Press, Homewood

Fararo TJ (1989) The meaning of general theoretical sociology, tradition and formalization. Cambridge University Press, New York

Felson M (1986) Linking criminal choice, routine activities, informal control and criminal outcomes. In: Cornish DB, Clarke RV (eds) The reasoning criminal: rational choice perspectives on offending. Springer, New York

Gibbs JP (1989) Control sociology’s central notion. University of Illinois Press, Chicago

Gleick J (1987) Chaos, making of a new science. Vicking Penguin, New York

Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Hirschi T (1969) Causes of delinquency. University of California Press, Berkely

Hirschi T (1979) Separate and unequal is better. J Res Crime Delinq 16:34–38

Horowitz FD (1987) Exploring developmental theories: toward a structural/behavioral model of development. Lawrence Erlbaum, Hillsdale

Hunter AJ (1985) Private, parochial and public school orders: the problem of crime and incivility in urban community. In: Suttles GH, Zald MN (eds) The challenge of social control: citizenship and institution building in modern society. Ablex Publishing, Norwood

Kornhauser RR (1978) Social sources of delinquency: an appraisal of analytic models. University of Chicago Press, Chicago

Le Blanc M (1997a) A generic control theory of the criminal phenomenon, the structural and the dynamical statements of an integrative multilayered control theory. In: Thornberry TP (ed) Developmental theories of crime and delinquency, vol 7, Advances in theoretical criminology. Transaction Publishers, New Brunswick, pp 215–286

Le Blanc M (1997b) Socialization or propensity: a test of an integrative control theory with adjudicated boys. Stud Crime Crime Prev 6:200–224

Le Blanc M (2006) Self-control and social control of deviant behavior in context: development and interactions along the life course. In: Wikstrom P-O, Sampson R (2005) The social contexts of pathways in crime: development, context, and mechanisms. Cambridge University Press, Cambridge

Le Blanc M (2009) The development of deviant behavior, its self-regulation. Mon Kriminol Strafr 91:117–136

Le Blanc M, Caplan M (1993) Theoretical formalization, a necessity: the example of Hirschi's social control theory. Adv Criminol Th 4:329–431

Le Blanc M, Fréchette M (1989) Male criminal activity, from childhood through youth: multilevel and developmental perspectives. Springer, New York

Le Blanc M, Loeber R (1998) Developmental criminology upgraded. In: Tonry M (ed) Crime and justice handbook, vol 23. University of Chicago Press, Chicago, pp 115–198

Lenski G (1988) Rethinking macrosociological theory. Am Sociol Rev 53:163–171

Lerner RM (2002) Concepts and theories of human development. Lawrence Erlbaum, Mahwah

Lytton H (1990) Child and parent effects in boys' conduct disorder: a reinterpretation. Dev Psychol 26:683–697

Messner SF, Krohn MD, Liska AE (eds) (1989) Theoretical integration in the study of deviance and crime: problems and prospects. State University of New York Press, Albany

Pinatel J (1963) Traité de criminologie. Dalloz, Paris

Reiss AJ (1951) Delinquency as the failure of personal and social controls. Am Sociol Rev 16:196–207

Shoemaker DJ (2005) Theories of delinquency: an examination of explanations of delinquent behavior. Oxford University Press, New York

Short JF (1985) The level of explanation problem in criminology. In: Meier RF (ed) Theoretical methods in criminology. Sage, Beverley Hills

Thornberry TP (1987) Toward an interactional theory of delinquency. Criminology 25:863–892

Wagner DG, Berger J (1985) Do sociological theories grow? Am J Sociol 90:697–728

Werner H (1957) The concept of development from a comparative and organismic point of view. In: Harris DB (ed) The concept of development. University of Minnesota Press, Minneapolis

Controlled Experiments

▶ Randomized Experiments in Criminology and Criminal Justice

Co-offending

Peter J. Carrington
Department of Sociology and Legal Studies,
University of Waterloo, Waterloo, ON, Canada

Overview

This entry is a general introduction to the subject of co-offending. In-depth treatments of specific topics in co-offending are provided in the articles by Hochstetler (decision-making), van Mastrigt (attributes of co-offenders), and Weerman (theorizing co-offending).

Co-offending is the commission of a crime by more than one person. It is often called "group crime," although that term is somewhat misleading, as the vast majority of co-offenses are committed by only two offenders (see Size, below). Co-offending is common and has been observed and studied by criminologists for at least a century, since Breckinridge and Abbot (1912) remarked on its high frequency among delinquent boys appearing in the juvenile court of Cook County, Illinois. (Shaw and McKay 1931) also commented on the prevalence of co-offending among boys appearing in the Cook County juvenile court in 1928. The study of co-offending became established in the 1960s, grew slowly through the rest of the twentieth century, and has begun to flourish in the early twenty-first century, partly because of a growing awareness of the applicability of the theories and methods of social network analysis (Carrington 2011).

Co-offending is studied as a phenomenon of interest in itself and also because it can provide insight into other criminological phenomena, such as problems in the measurement of crime, social contagion of criminal attributes and behavior, and criminal networks, gangs, and organized crime.

Prevalence

Writers on co-offending are unanimous that it is very common, especially among young offenders. However, empirical estimates of its prevalence vary widely – from 10 % of criminal incidents to over 80 % of offenders – probably because of variations in the definition of prevalence and in the types of data, as well as the variety of places, times, and samples, that have been used to measure it (Table 1). van Mastrigt and Farrington (2009) point out that in estimating prevalence, offenses (that is, criminal incidents) must be distinguished from offense participations, or the participation of one person in one offense. Since co-offenses involve multiple offense participations, prevalence estimates based on counting co-offenses will be lower than estimates based on counting offense participations. In addition, some authors report a third type of prevalence: the percentage of subjects in the sample who ever co-offended during the period of observation. Each definition of prevalence provides a different perspective on the same phenomenon. Estimates of prevalence also differ because they are derived from self-reported versus official (recorded) data (or, occasionally, victim reports), because the sample is limited to a certain age range, such as (commonly) young people, or because different types of crime are included (Table 1).

Early estimates of the prevalence of co-offending tended to be relatively high. Breckinridge and Abbot (1912) commented that lone offending was rare among the delinquent boys that they studied. Shaw and McKay (1931) reported that 82 % of the offense participations of a sample of boys who came before the Cook County juvenile court in 1928 were in co-offenses, and 81 % of the boys had at least one co-offense in their delinquent careers. Erickson (1971) reviewed 11 studies of juvenile delinquency in the United States prior to 1935 and found that the reported prevalence of co-offending ranged between 70 % and 80 % of offenses; his own research on self-reported delinquency found that 65 % of offenses were co-offenses. Of the offenses that were self-reported by a sample of 13–16-year-olds in the National Survey of Youth in 1967, 70 % were co-offenses (Warr 1996).

Estimates based on more recent data are lower. Sarnecki (2001) reports that 56 % of the young offenders registered by the Stockholm police between 1991 and 1995 were involved in at least one co-offense during that period. Fifty-one percent of the recorded offenses committed between the ages of 10 and 32 by the boys in the Cambridge Study were co-offenses (Reiss and Farrington 1991), as were 37 % of recorded offense participations involving young persons in the British police data for 2002–2005 analyzed by van Mastrigt and Farrington (2009) and 25 % of offenses involving young persons in Carrington (2002a) Canadian police data for 1992–1999.

The prevalence of co-offending is much lower in the very few studies that analyze population samples of offenders of all ages. van Mastrigt and Farrington (2009) report that only 10 % of offenses by offenders of all ages were co-offenses; Carrington (2002a) reports 12 %; and Becker and McCorkel (2011) report 18 %.

Co-offending and the Measurement of Crime

The phenomenon of co-offending has important implications for the validity of simple crime

Co-offending, Table 1 Prevalence of co-offending in various studies

Study	Country	Data source	N[a]	Age of offenders	Type of crime	Percent co-offenses[b]	Percent co-offenders[c]	Percent ever co-offended[d]
(a) Young offenders								
Shaw and McKay (1931)	USA	Court records	5,480 offenders 3,517 incidents	Juveniles	All		81.8	81.0
Erickson (1971)	USA	Self-reports	150 respondents	15–17	18 items	65.0		
Warr (1996)[e]	USA	Self-reports	847 respondents 3,065 incidents	13–16	12 items	70.2		
Sarnecki (2001)	Sweden	Police records	22,091 suspects 36,941 participations 29,209 incidents	5–20	All			56.0
Carrington (2002a)	Canada	Police records	648,339 offenders 480,457 incidents	3–17	All	24.5	44.1	
Pettersson (2005)	Sweden	Police records	1,253 suspects	<21	Violent			89.4
van Mastrigt and Farrington (2009)	UK	Police records	26,997 participations	10–17	All		36.5	
Carrington (2009)	Canada	Police records	55,336 offenders 109,265 participations 93,470 incidents	5–17	All	45.0	53.0	64.0

(b) Various ages								
Hindelang (1976)	USA	Victim reports	182,290 offenses	Unknown	Victimizations	48.0		
Reiss (1980)	USA	Victim reports	22,107 offenses	Unknown	Victimizations	35.7		
Reiss and Farrington (1991)	UK	Court records	411 respondents 683 offenses	10–32	All	51.0		
Carrington (2002a)	Canada	Police records	2.9 million offenders 3.4 million incidents	3–89	All	11.8	24.5	
van Mastrigt and Farrington (2009)	UK	Police records	61,646 offenders 120,274 participations 105,348 offenses	10–74	All	10.4	21.6	30.2
Becker and McCorkel (2011)	USA	Police records	20.2 million offenders 16.2 million offenses	All ages	All	17.6	34.0	

Later studies using the same data as reported studies, studies in which the prevalence is set by design, and studies not reporting the overall prevalence of co-offending, are omitted

[a] In most studies where the number of offense participations is not explicitly given, the term "offenders" refers to offense participations

[b] Percent of offenses that were co-offenses

[c] Percent of offense participations that were in co-offenses

[d] Percent of offenders that ever co-offended during the study period

[e] Calculated by the author from data provided in Table 1

statistics, whose calculation often implicitly assumes that one incident means one offender, so that the count of criminal incidents equals the count of offenders. Ignoring co-offending can also bias more advanced statistics, such as the effect parameters in regression models of sentencing outcomes (Waring 2002).

The Age-Crime Curve

The well-known age-crime curve is a plot of age-specific crime rates by the year of age of the offenders. Invariably, it relies on counts of offenders, not of incidents – or, more accurately, on counts of offense participations. However, since co-offending is more common among young people, the age-crime curve exaggerates the rate of youth *crimes*, while accurately depicting the rate of young *offenders* (Reiss and Farrington 1991; van Mastrigt and Farrington 2009). An age-crime curve that is adjusted for differential age-related rates of co-offending also exhibits the characteristic peak during adolescence, but it is considerable less sharp than the unadjusted age-crime curve (van Mastrigt and Farrington 2009, Fig. 2, p. 565): that is, the elevated rate of co-offending in adolescence accounts partly, but by no means entirely for the characteristic shape of the age-crime curve during adolescence.

Probability of Conviction

Co-offending also affects the calculation of the probability of conviction, which is used by researchers and policy-makers to assess flow through the criminal justice system and to assess the effectiveness of crime prevention and crime control policies and programs (van Mastrigt and Farrington 2009). Often the probability of conviction is calculated by dividing the number of persons convicted by the number of reported criminal incidents – again, implicitly assuming that one incident means one offender (a further inaccuracy is introduced when incident counts are weighted by the number of victims). Co-offending results in inflated conviction probabilities: for example, if all incidents involved two co-offenders, and all offenders were convicted, the probability of conviction would be calculated as 2.0, instead of the correct figure of 1.0.

The Dark Figures of Co-offending

Co-offending also results in inaccurate offender-based crime statistics because of the various ways in which co-offenders are omitted from reported counts of offenders. As with solo offenders, co-offenders may be omitted because the incident was not reported to, or recorded by, police; or because recorded incidents were not cleared, so that the number and attributes of the offender(s) are not known. In addition, in co-offenses that are recorded and cleared, only a subset of the co-offenders may be identified and counted. This results in biases in statistics on group crime, age and crime, possibly gender and crime, and any other aspects of crime that are correlated with co-offending. However, some writers suggest that co-offenses have a higher probability than solo offenses of being reported to police; if true, this would result in further biases in the statistics discussed above (van Mastrigt 2008).

Characteristics of Co-offending Groups

Size

The majority of co-offenses involve only two co-offenders (Carrington 2002a; Reiss 1988; Reiss and Farrington 1991; Shaw and McKay 1931; van Mastrigt 2008). Table 2 summarizes the distributions of co-offending group sizes from recent large-scale studies in three countries. In police-reported data from England and Canada, for all types of offenses and all ages of offenders, three-quarters of co-offenses involved only two co-offenders, and less than 10 % involved more than three co-offenders. The proportion of multiple-offender groups is much higher in the American data on violent victimizations: more than 30 % of victimizations involved more than three offenders. The discrepancy may be due to the use of a rather loose definition of “involvement” of co-offenders in a crime by victims reporting to a victimization survey.

Co-offending, Table 2 Distributions of co-offending group sizes in three countries

	Canada, youth, 1992–1999[a]	Canada, 1992–1999[b]	England, 2002–2005[c]	USA, violent, 2008[d]
Lone offender	75.5	88.2	89.6	81.3
Co-offense	24.5	11.8	10.4	18.7
Total	100.0	100.0	100.0	100.0
Number of co-offenders				
2	71.2	74.9	76.2	39.6
3	20.0	17.3	16.5	29.4
4+	8.8	7.8	7.3	31.0
Total	100.0	100.0	100.0	100.0

[a] Recalculated from data on co-offenders under 18 years of age in Carrington (2002a), Table 1, p. 283, based on 2,891,695 criminal incidents recorded by police parts of Canada in 1992–1999

[b] Recalculated from data on co-offenders of all ages in Carrington (2002a), Table 1, p. 283, based on 2,891,695 criminal incidents recorded by police parts of Canada in 1992–1999

[c] Recalculated from data in van Mastrigt (2008), Table 4.1, p. 53, based on 105,348 notifiable criminal events recorded by police in South Yorkshire, UK, in 2002–2005

[d] Recalculated from data in Bureau of Justice Statistics (2011), Table 37, based on 4,581,260 incidents of violent criminal victimization reported by victims in the USA in 2008

Composition

Co-offending groups have a strong homophilous tendency: that is, co-offenders tend to be similar in age, gender, ethnicity, place of residence, and criminal experience (Pettersson 2005; Reiss 1988; Sarnecki 2001; van Mastrigt and Farrington 2011; Warr 2002; Weerman 2003; see also van Mastrigt, this volume). This observed similarity lends itself to interpretations in terms of co-offender selection, or a preference for similar people as accomplices (Reiss and Farrington 1991; Warr 2002; see also Weerman, this volume). On the other hand, the results of stochastic modeling suggest that a not-inconsiderable part of the observed homophily can be accounted for simply by the composition of the pool of potential accomplices. This is particularly the case for the frequent observation that female offenders exhibit lower co-offending homophily than males (Carrington 2002b).

Other processes than homophilous preference or random selection have been invoked to explain the composition of co-offending groups. One explanation, based on rational choice theory, emphasizes a conscious, rational decision-making process in the choice of co-offenders (Tremblay 1993). Other writers, influenced by routine activities theory, propose that offenders encounter potential accomplices in "convergence settings" such as neighborhoods, schools, or other "hangouts." Thus, offenders become co-offenders through a process of constrained choice.

Stability

Co-offending relationships tend to be short-lived: the majority of co-offending pairs commit only one (recorded) offense together (McGloin et al. 2008; Reiss and Farrington 1991; Sarnecki 2001; van Mastrigt 2008; Warr 1996). However, some professional criminals such as burglars maintain relatively long-term criminal collaborations.

Correlates of Co-offending

The main characteristics of crimes and offenders that are associated with the frequency of co-offending are the type of crime and the age of the offender. There is also some evidence of weak relationships between co-offending and the sex or gender, ethnicity or race, and criminal experience of the offender.

Type of Crime

The strongest predictor of co-offending is the type of crime that is being committed. Some

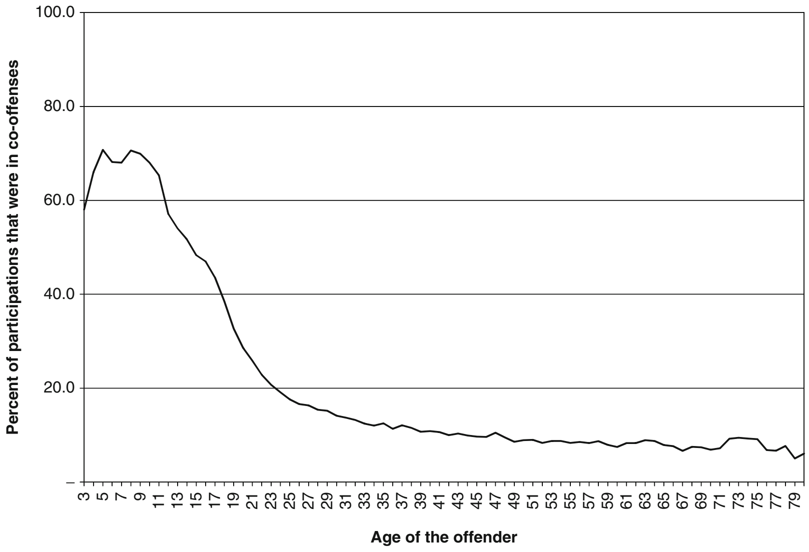

Co-offending, Fig. 1 Co-offending and the age of the offender (Source: Data on 3.4 million offense participations reported by police in parts of Canada. Statistics Canada, Uniform Crime Reporting Survey, 1992–1999)

types of crime, such as burglary, robbery, arson, and theft of or from a motor vehicle, are characterized by high rates of co-offending (Carrington 2002a; Reiss and Farrington 1991); and these high rates are not accounted for by confounding factors such as the offender's age and criminal experience (Carrington 2009; van Mastrigt and Farrington 2009). Other types of crime have characteristically low rates of co-offending, for example, sexual assault and other sex crimes, drug possession, drink-driving, and administrative offenses such as breach of conditions of bail or probation. Generally, offenses committed by co-offenders tend to be more serious than those committed alone (Erickson 1971) – a phenomenon called social amplification. Carrington (2002a) found that group crimes were more likely than solo crimes to involve a weapon, particularly a firearm, and to involve major property damage or loss, or major injury or death. Vandiver (2010) found that co-offending with a boy increased the level of seriousness of sexual offenses committed by girls.

Age of the Offender

The relationships between co-offending and characteristics of the offender are reviewed in detail by van Mastrigt (this volume), so they are only briefly summarized here. The age of the offender is by far the strongest offender-related correlate of co-offending. Co-offending is much more prevalent among young than adult offenders (compare columns 1 and 2 of Table 2): the age/co-offending curve rises to a peak in adolescence, and falls monotonically with age thereafter (Fig. 1). However, the age/co-offending curve varies by type of crime and the offender's criminal career (Carrington 2002a, 2009; van Mastrigt and Farrington 2009).

The Offender's Criminal Experience

Co-offending is most common at the onset of the criminal career and decreases with criminal experience (Reiss and Farrington 1991), except among child offenders under 12 years of age (Carrington 2009). This pattern is not accounted for by the decrease with age in co-offending, although age and experience are correlated. However, more active ("high-rate") offenders are more likely to co-offend than low-rate offenders, and this tendency masks the decline in co-offending in their careers (Carrington 2009).

Gender and Ethnicity of the Offender

The probability of co-offending has also been found to be weakly related to the offender's gender or sex, with female offenders being slightly more likely to co-offend. There is also a small body of evidence suggesting a relationship between co-offending and the offender's ethnicity or race (see van Mastrigt, this volume).

Social Contagion

Higher rates of co-offending observed at the onset of the criminal career and the sustained co-offending exhibited by some persistent offenders (Reiss 1988; Carrington 2009) suggest that co-offending may be a mechanism through which criminal attitudes and behavior are transmitted to criminal neophytes by more experienced and influential accomplices. This process has variously been conceptualized as tutelage, recruitment (Reiss 1988; van Mastrigt and Farrington 2011), instigation (Warr 1996), social learning, peer influence (Warr 2002), and mentorship (Morselli 2009). A variant account of social contagion is the idea that co-offending with more experienced criminals provides neophytes with criminal opportunities – access to criminal capital in the form of expertise and networks – just as collaboration with more experienced and established practitioners does in other occupations (Becker and McCorkel 2011).

The ability of more experienced and committed criminals to influence impressionable neophytes is often attributed to a difference in age, but has also been attributed to gender roles in the case of female offenders who are hypothesized to have been influenced or coerced into co-offending with their male romantic partners, particularly in cases of child sexual abuse.

Co-offending Networks

Co-offending data have been used to study the structure and dynamics of gangs, criminal or delinquent groups, organized crime groups, and criminal markets. Linking co-offenses together by the offenders that they have in common results in a co-offending network, which is one type of criminal network. In the hypothetical example shown in Fig. 2, there are five co-offenses involving seven offenders, some of whom such as Jill are repeat offenders (Fig. 2a). Linking together incidents connected by common offenders such as Jill results in a bipartite, or two-mode network (Fig. 2b), so-called because the nodes, or vertices, are classified into two mutually exclusive classes (offenders and incidents), and the lines, or edges, exist only the classes. This type of network is also known as an affiliation network, because it represents individuals' affiliations – in this case, to criminal events. A two-mode or affiliation network can be analyzed as it is, but criminal network researchers usually work with the simpler one-mode person-to-person network shown in Fig. 2c, which can be derived from the two-mode network by a simple operation of matrix algebra. Real-life co-offending networks may have hundreds or even thousands of nodes (e.g., Sarnecki 2001).

The theories and techniques of social network analysis are used to analyze these networks in order to identify static and dynamic structures and test hypotheses about them (Carrington 2011). Whether such networks are, or contain, gangs or organized crime groups is a matter of conceptual definition and empirical investigation. Research in this area has identified various structural features of co-offending networks and interpreted them in terms of issues in the study of gangs and organized crime.

In this extension of co-offending research, the phenomenon of co-offending is of interest not in itself, but as evidence of connections among offenders (Fig. 2c), which may be used instead of, or in addition to, other evidence derived from wiretaps, surveillance, interviews, etc., to study the social organization of crime and criminals. Co-offending data have played only a very limited role in the study of gangs and organized crime. Co-offending per se constitutes only one kind of connection among criminals, and only a small part of the activities of organized criminals and enterprises; consequently, co-offending

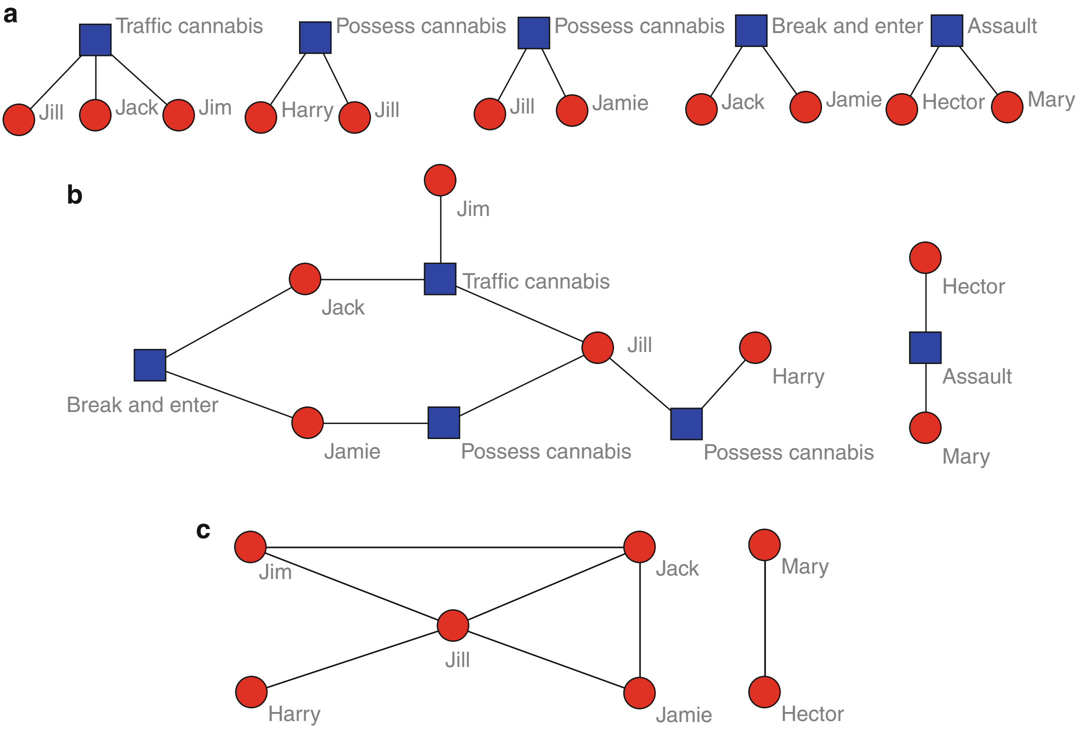
a
Traffic cannabis
Jill
Jack
Jim
Possess cannabis
Harry
Jill
Possess cannabis
Jill
Jamie
Break and enter
Jack
Jamie
Assault
Hector
Mary
b
Jim
Traffic cannabis
Jack
Jill
Break and enter
Jamie
Possess cannabis
Harry
Possess cannabis
Hector
Assault
Mary
c
Jim
Jack
Jill
Harry
Jamie
Mary
Hector

Co-offending, Fig. 2 Hypothetical co-offending network. (**a**) Five co-offenses involving seven offenders. (**b**) Derived two-mode co-offending network. (**c**) Derived one-mode co-offending network

networks provide only a skeletal, incomplete, and potentially biased view of underlying criminal networks. Most network-oriented research on organized crime and criminal markets has relied on data other than co-offending, such as self-reports, police wiretaps and other intelligence, and reports of police investigations (see, e.g., the reviews in Carrington 2011; Morselli 2009). Some researchers have compared the value of co-offending data to that of other sources of data on organized crime (Malm et al. 2010).

Gangs as Cohesive Subgroups

One approach to the question of the existence and structure of so-called gangs in co-offending networks is to identify cohesive subgroups within the network. A cohesive subgroup is a set of nodes that are connected together with relatively dense ties. In Fig. 2c, Mary and Hector form one cohesive subgroup, and the other offenders form a second. Within the second subgroup, there are two smaller subgroups whose members are all directly interconnected (termed “cliques”): Jill, Jamie and Jack, and Jill, Jim and Jack. These four offenders could be seen as the core, with Harry being somewhat peripheral. Research on real-life co-offending networks has generally found that they consist of many small groups of less than a dozen members that are loosely linked together by relatively sparse connections into much larger networks (McGloin 2005; Reiss 1988; Sarnecki 2001; Warr 1996).

Criminal Performance

A central question in the study of organized crime and criminal markets is that of criminal performance, or achievement: what factors affect the likelihood and degree of success of individuals and groups in their criminal activities? The study of criminal networks has adapted ideas from the wider literature on networks and success in legitimate business, notably Ronald Burt's theory of structural holes and brokerage, and Nan Lin's network theory of social capital.

Nonredundancy, Structural Holes, and Brokerage

The sparse connections between cohesive subgroups in co-offending networks create nonredundancy in some individuals' own ego-networks: that is, co-offending is rare among the focal individual's co-offenders. Opportunities for criminal brokerage arise or are created when an individual forms a unique bridge between otherwise unconnected groups that possess reciprocally valued resources. In the subgroup of five co-offenders in Fig. 2c, one of Jill's co-offenders, namely, Harry, does not co-offend with anyone else. This makes Jill's connection with Harry nonredundant, or necessary, for the cohesiveness of the subgroup, and makes Jill a structurally important member of the group. To put it differently, before Jill co-offended with Harry, there was a structural hole, or disconnect, between Harry and the others, which his co-offense with Jill bridged. By bridging this structural hole, Jill became a broker, or liaison, between Harry and the other three members of the subgroup, and may benefit from this structural advantage. There remains a structural hole between Mary and Hector and the other five offenders, and if one of the five co-offended with Mary or Hector, he or she would become a broker between the two formerly disconnected subgroups, and structural advantage would accrue to him or her.

Morselli (2009) and his colleagues and Malm and her colleagues (Malm and Bichler 2011; Malm et al. 2010) have explored the antecedents and consequences of criminal brokerage based on individuals' positions in criminal groups and networks, and in criminal transaction networks, although co-offending is only a small part of the data on organized crime networks that they analyze.

Embeddedness and Criminal Capital

The concept of criminal capital is adapted from that of social capital, which refers to the resources that are available as a result of one's connections to others. By extension, criminal capital refers to the knowledge, skills and other resources that are derived from embeddedness in criminal networks. The value of criminal capital both for individuals and for criminal groups has been demonstrated with data on co-offending networks (McCarthy and Hagan 1995; Waring 2002).

Centrality has been used as an indicator of an individual's embeddedness in a criminal group. Individuals who occupy central positions in co-offending groups or networks tend to be more experienced and prolific offenders (Sarnecki 2001), to have the most pro-criminal attitudes, and to have the highest risk of violent victimization. Boys tend to be more central than girls.

Interdiction

At least since the publication of Harper and Harris's (1975) paper on link analysis, researchers and practitioners have appreciated that knowledge of connections among criminals, including co-offending networks, can be useful in identifying criminal groups and their membership, and in understanding their organizational structure and dynamics in order to disrupt or destroy them. McGloin's (2005) research has shown the potential effectiveness of identifying, targeting, and removing or neutralizing "key players," who are individuals who occupy key positions in the criminal network, as central members or as "cut points," whose removal would cause disconnection of the network. For example, in Fig. 2c, Jill is both a central figure and a cut point, although the high degree of connectedness among Jill, Jack, Jim and Jamie means that removal of Jill might have only limited success in disrupting the network, by isolating Harry. On the other hand, research by Morselli (2009) and his colleagues calls this strategy into question by research showing the flexibility and adaptability of criminal networks; the implication is that these networks can reconfigure themselves to recover from key player attacks. Malm and Bichler (2011) show that different strategies are implied by the different structures found at different parts of illicit supply chains.

Related Entries

▶ Co-offending and Offender Attributes
▶ Theories of Co-offending

Recommended Reading and References

Becker S, McCorkel JA (2011) The gender of criminal opportunity: the impact of male co-offenders on women's crime. Fem Criminol 6:79–110

Breckinridge SP, Abbot E (1912) The delinquent child and the home. Charities Publication Committee, New York

Bureau of Justice Statistics (2011) Criminal victimization in the United States, 2008. Statistical Tables. National Crime Victimization Survey. http://bjs.ojp.usdoj.gov/index.cfm?ty=dcdetail&iid=245. Accessed 5 Nov 2001

Carrington PJ (2002a) Group crime in Canada. Can J Criminol 44:277–315

Carrington PJ (2002b) Sex homogeneity in co-offending groups. In: Hagberg J (ed) Contributions to social network analysis, information theory and other topics in statistics. Stockholm University Press, Stockholm, pp 101–116

Carrington PJ (2009) Co-offending and the development of the delinquent career. Criminology 47:1295–1329

Carrington PJ (2011) Crime and social network analysis. In: Scott J, Carrington PJ (eds) SAGE Handbook of social network analysis. Sage, London, pp 236–255

Erickson ML (1971) The group context of delinquent behavior. Social Problems 19:114–129

Harper WR, Harris DH (1975) The application of link analysis to police intelligence. Human Factors 17:157–164

Hindelang MJ (1976) Criminal victimization in eight American cities. Ballinger, Cambridge

Malm A, Bichler G (2011) Networks of collaborating criminals: assessing the structural vulnerability of drug markets. J Res Crime Delinq 48:271–297

Malm A, Bichler G, Van De Walle S (2010) Comparing the ties that bind criminal networks: is blood thicker than water? Secur J 23:52–74

McCarthy B, Hagan J (1995) Getting into street crime: the structure and process of criminal embeddedness. Soc Sci Res 24:63–95

McGloin JM (2005) Policy and intervention considerations of a network analysis of street gangs. Criminol Public Policy 4:607–636

McGloin JM, Sullivan CJ, Piquero A, Bacon S (2008) Investigating the stability of co-offending and co-offenders among a sample of youthful offenders. Criminology 46:155–188

Morselli C (2009) Inside criminal networks. Springer, New York

Pettersson T (2005) Gendering delinquent networks. A gendered analysis of violent crimes and the structure of boys' and girls' co-offending networks. Young Nord J Youth Res 13:247–267

Reiss AJ (1980) Understanding changes in crime rates. In: Feinberg SE, Reiss AJ (eds) Indicators of crime and criminal justice: quantitative studies. Bureau of Justice Statistics, Washington, DC

Reiss AJ (1988) Co-offending and criminal careers. Crime Justice 10:117–170

Reiss AJ, Farrington DP (1991) Advancing knowledge about co-offending – results from a prospective longitudinal survey of London males. J Crim Law Criminol 82:360–395

Sarnecki J (2001) Delinquent networks: youth co-offending in Stockholm. Cambridge University Press, Cambridge, UK

Shaw CR, McKay HD (1931) Social factors in juvenile delinquency, vol II, Report on the causes of crime. National Commission on Law Observance and Enforcement, Washington, DC

Tremblay P (1993) Searching for suitable co-offenders. In: Clarke RV, Felson M (eds) Routine activity and rational choice, vol 5, Advances in criminological theory. Transaction, New Brunswick, pp 17–36

van Mastrigt SB (2008) Co-offending. Relationships with age, gender, and crime type. PhD dissertation, Institute of Criminology, University of Cambridge

van Mastrigt SB, Farrington DP (2009) Co-offending, age, gender and crime type: implications for criminal justice policy. Br J Criminol 49:552–573

van Mastrigt SB, Farrington DP (2011) Prevalence and characteristics of co-offending recruiters. Justice Q 28:325–359

Vandiver DM (2010) Assessing gender differences and co-offending patterns of a predominantly "male-oriented" crime: a comparison of a cross-national sample of juvenile boys and girls arrested for a sexual offense. Violence Victims 25:243–264

Waring EJ (2002) Co-offending as a network form of social organization. In: Waring EJ, Weisburd D (eds) Crime and social organization, vol 10, Advances in criminological theory. Transaction, New Brunswick, pp 31–47

Warr M (1996) Organization and instigation in delinquent groups. Criminology 34:11–37

Warr M (2002) Companions in crime: the social aspects of criminal conduct. Cambridge University Press, Cambridge, UK

Weerman FM (2003) Co-offending as social exchange: explaining characteristics of co-offending. Br J Criminol 43:398–416

Co-offending and Offender Attributes

Sarah van Mastrigt
Department of Psychology and Behavioral Sciences, University of Aarhus, Aarhus, Denmark

Synonyms

Accomplice crime; Group crime; Joint/group offending

Overview

Co-offending, the act of committing crime alongside one or more accomplices, is a well-documented and widespread phenomenon (Carrington & van Mastrigt, 2013). However, it is far from uniform. Research has uncovered sizeable variations in the rates and forms of co-offending observed both between and within individuals. In this entry, empirical evidence linking these variations to the offender attributes of sex, ethnicity, age, and criminal career is reviewed and discussed in relation to its implications for theory and policy. It is concluded that these offender characteristics generate, and sometimes reflect, meaningful differences in co-offending and that treating co-offending as a unitary phenomenon risks masking some of its most interesting, and policy relevant, features. Open questions and directions for future research are identified, together with a need for greater attention to person-by-situation interactions.

Background

Research dating back nearly a century has noted that many crimes are committed in groups, particularly during adolescence (Shaw and McKay 1931). This finding is so robust across time and geography that co-offending is now widely regarded as criminological “fact.” Even so, relatively little research has been dedicated to understanding co-offending patterns and their variations between and within individuals. This is likely due, in part, to the fact that many of the early investigations of co-offending were cross-sectional in nature and based on small samples of predominantly white, young males. Although these studies were quick to identify variations in rates of co-offending observed for different crime types (which tend to be high for property crimes like robbery, burglary, and theft of motor vehicles and very low for sex crimes), their restricted samples often precluded detailed analyses of the relationships between co-offending and key *offender* characteristics like sex, age, ethnicity, and criminal history.

It is only relatively recently that growing empirical interest in the topic of group crime, and the availability of larger and more inclusive datasets, has led to considerable advances in our understanding of attribute-based variations in co-offending. To date, most research has focused on identifying variations in the *extent* of co-offending; however, a growing body of literature suggests that there are also notable individual-level variations in the *nature* of co-offending with respect to the composition and character of one’s co-offending involvement. Understanding these variations is important not only for theory development (see Weerman, this volume) but also because accurate estimates of crime rates and key criminal justice statistics, like the probability of conviction and the effects of incapacitation, must take account of varied rates of co-offending for different social groups. Because an offense committed by two offenders is counted twice in many criminal justice statistics (e.g., arrest and conviction data), the share of crime attributed to offender groups that have a greater tendency to co-offend will be artificially inflated, and estimates of interventions targeting these offenders will be subject to greater distortions (Andresen and Felson 2009; van Mastrigt and Farrington 2009; Zimring 1981). Thus, as McCord and Conway (2005: 11) have noted, “crime. . .cannot be accurately portrayed unless co-offending is accurately recorded” and disaggregated for different offender groups.

The importance of information on individual-level variations in co-offending is clear. Unfortunately, knowledge in this area is still underdeveloped (in particular that exploring within individual co-offending across the criminal career). Nevertheless, a number of consistent attribute-based variations in the prevalence and form of co-offending have been identified in the extant literature. These are outlined below.

Empirical Evidence

Sex

Compared to the vast literature dedicated to exploring other gender differences in crime, relatively little work has considered how patterns of group crime vary for males and females.

With respect to sex differences in the prevalence of co-offending, the bulk of the existing evidence suggests that females are slightly more likely than males to offend with accomplices, although a handful of studies have reported the opposite. Citing West German police statistics from 1982, for example, Reiss (1988: 134) reported that across age and crime type, males displayed higher aggregate rates of co-offending compared to females (32 % vs. 24 %), a finding that was later replicated (to a lesser extent) in a number of Swedish investigations of juvenile co-offending (Pettersson 2005; Sarnecki 1990, 2001). More commonly, however, females are found to have higher co-offending rates. In one of the earliest studies to explicitly compare male and female co-offending patterns, it was concluded that in the US context, “females in both urban and rural settings tend to have group violation rates that are higher than their male counterparts” (Erickson and Jensen 1977: 268). Similar conclusions have since followed from research conducted in other countries, including Holland (Hakkert 1998: 466), Scotland (Emler et al. 1987: 104), and England (Budd et al. 2005), although the magnitude of difference is typically small (often less than 10 %).

Most studies exploring gender differences in co-offending, including those cited above, have been based on juvenile samples and have been limited to bivariate analysis of the sex/co-offending relationship. Because the age distributions and types of offenses committed by male and female offenders is known to differ, however, it is important to consider whether the demonstrated sex/co-offending relationship could be spurious and, in fact, reflect variations in the ages or types of crimes committed by male and female co-offenders. The findings from two recent multivariate investigations of co-offending suggest that this is not likely to be the case. Based on a large-scale analysis of Canadian Uniform Crime Report (UCR) 2 incident-based data, Carrington (2009: 1306) reported that with the exception of sex offending, females’ rates of juvenile co-offending were higher across nearly all major categories of offense and that this pattern was maintained across age and criminal experience. His overall conclusion that “the [juvenile] offense participations of females were more likely to involve accomplices than those of males, but the difference is not large (57 % vs. 51 %)” echoes that reported by van Mastrigt and Farrington (2009: 565) across a wider age range (10–74 years); in their study, females had approximately 20 % higher logged odds of co-offending, even when controlling for age and crime type. These multivariate findings provide the strongest evidence yet that females are (slightly) more likely than males to engage in group offending and that the sex/co-offending relationship cannot be attributed solely to the types of crimes that males and females commit, or to the age of particular study populations. What has yet to be determined is what *can* explain this sex difference. Whether females are more likely to co-offend because they are more socially oriented in general, more susceptible to peer pressure, or something else entirely remains an empirical question.

In addition to variations in the extent of male and female co-offending, a number of sex-linked variations in the composition of males’ and females’ co-offending groups have also been identified. One of the most consistent findings to emerge in this regard is that across age, males tend to offend in larger groups. Although numerous studies have shown that offending in groups of two is the overwhelming norm for both sexes, pair offending is “somewhat…more common among females and offending in groups of three or more is slightly more common among males” (Carrington 2002a: 289; Hakkert 1998; Sarnecki 2001).

Another interesting difference concerns the gender composition of males’ and females’ offending groups. In general, while both males and females appear to choose co-offenders of the same sex in the majority of cases (especially during youth), “females are [nonetheless] more often found in the company of males than vice versa” (Warr 2002: 79, also see Pettersson 2005; Reiss and Farrington 1991; Sarnecki 2001). In their analysis of 400 Philadelphia juveniles’ court records, for example, Conway and McCord (2002) reported that whereas 95.5 % of male co-offenders had a male accomplice in their first recorded co-offense, only 80.6 % of females had female co-offenders. In older samples, even more striking differences are evident. In Carrington’s (2002b) Canadian analysis, 87 % of co-offending males compared to only 50 % of co-offending females were involved in same-sex incidents. Often, these findings are interpreted, either explicitly or implicitly, to indicate that males have a much stronger preference for same-sex accomplices, perhaps because they are viewed as more competent or able crime partners (Steffensmeier and Terry 1986). However, in a population highly skewed in favor of males, as the offending population is, a preponderance of all-male groups and a scarcity of all-female groups would be expected even if offenders chose their accomplices at random (Carrington 2002b). Recent analyses have shown that when the observed sex compositions of offending groups are compared with those expected based on the gender distribution of the offending population, males do, in fact, exhibit a slightly greater (statistical) bias for same-sex accomplices overall, but this difference is relatively small, and for offending groups with two offenders (the vast majority), this bias is greater for females (van Mastrigt and Carrington 2013). These findings highlight the importance of considering chance expectations when comparing the compositions of males’ and females’ co-offending groups.

Such considerations will also be relevant for the interpretation of other recent findings reported in the co-offending literature, for example, the tendency for boys involved in violent offending to be tied to co-offenders from the same neighborhood to a greater extent than females (Pettersson 2005), the suggestion that when males are involved in mixed-sex offending it is usually with family members or romantic partners (Reiss and Farrington 1991; Warr 2002), and the finding that mixed-sex offending groups appear to be slightly more likely to re-offend than same-sex groups of either gender (van Mastrigt 2008:203). Going forward, it will be important to determine whether these patterns reflect meaningful sex differences in the *preference* for particular types of co-offenders, or differential *opportunity* or social network structures that render some types of accomplices more “available” to one sex or the other (Andresen and Felson 2009; Weerman 2003).

Additional attention is also warranted with respect to the offending roles adopted by males and females engaged in group crime. More than 25 years ago, Reiss (1988) stressed the importance of identifying potential co-offending recruiters or instigators, that is, individuals who introduce others into crime and/or play a primary role in initiating offenses. Although there is often a suggestion that males are more likely than females to adopt this role and that male instigation plays a particularly important part in female offending (Sarnecki 2001), very little empirical work has examined this question. Although an earlier analysis of National Youth Survey data showed that females were less likely than males to report being an instigator of crime (Warr 1996), two more recent multivariate analyses of recruitment and instigation, respectively, failed to identify a gender effect when controlling for other person and situational-level variables (McGloin and Nguyen 2012; van Mastrigt and Farrington 2011).

Taken together, the findings outlined above suggest that there are a number of consistent gender differences in the prevalence and character of co-offending and that sex-linked variations deserve serious attention from both researchers and theorists. At the same time, many of the identified differences are small, and a number of the group crime patterns observed across sex are more similar than different (e.g., with respect to the overall tendency for offending groups to be small, homogeneous, and unstable). Given that many features of crime are very strongly differentiated by gender (including the basic tendency to offend), it is perhaps surprising that the sex variations noted here are not of a greater magnitude. This in itself raises interesting questions for future work. Clearly, there is much more to be learned about how gender shapes the dynamics of co-offending.

Ethnicity

One of the attributes that has received the most attention in the general criminological literature and the least in the co-offending field is ethnicity. Very little systematic research has explored ethnic or racial variations in the extent and form of co-offending. Citing US victimization statistics from 1982, Reiss (1988:134) was the first to identify the potential importance of this attribute, noting that for violent offenses, black offenders were "somewhat less likely than whites to be solo offenders" (60 % vs. 72 %). More recently, Felson (2003: 154) reported a similar finding based on his analysis of the same data from the year 2000, noting that "although nonwhites constitute 20 % of the total U.S. population, violent victims of lone offenders report their attackers to be nonwhite 34 % of the time. By comparison, victims of *group violence* describe their attackers as nonwhite or mixed in 56 % of reports." Unfortunately, comparative prevalence analyses for non-US populations, nonviolent crimes, and other ethnic and racial groups are lacking. Thus, while there is preliminary evidence that race may play a role in whether American offenders commit their violent crimes alone or in groups, the extent to which this is the case more generally is unclear. Additional data and multivariate analyses which can disentangle the effects of ethnicity and related variables (e.g., SES, crime type) will be needed before definitive statements can be made regarding ethnic and racial differences in the tendency to co-offend; however, it appears likely that some variation exists.

In addition to impacting whether an offense involves co-offenders, a handful of studies suggest that ethnicity also plays a role in the *selection* of those co-offenders. In the American context, individuals from different racial backgrounds are rarely suspected of committing offenses together. Based on the same US victimization data noted above, Reiss (1988: 135) reported that less than 6 % of violent group victimizations were perpetrated by mixed-race groups, leading him to conclude that offending with an accomplice of a different race is even less likely than offending with a co-offender of the opposite sex. In a more recent study of violent crime committed by juveniles in Philadelphia, co-offending was similarly characterized by marked racial homogeneity such that 96 % of the accomplices linked to African-American suspects and 83 % of those linked to Whites and Latinos were of the same ethnic background (Conway and McCord 2002). Outside of North America, the results are more mixed. In the UK-based Cambridge Study in Delinquent Development, high levels of homogeneity were reported for whites, but not blacks, a finding that the authors attributed in part to "the relative distribution of whites and blacks in the London population" (Reiss and Farrington 1991: 391).

In the most comprehensive analyses of ethnicity and group crime published to date, the co-offending networks of all youths (under the age of 21) suspected of committing a violent offense in Stockholm, Sweden, in 1995 were constructed (Pettersson 2003; Sarnecki 2001). Using country of birth and citizenship status to code offenders into one of eight ethnic groups, the authors found that the probability of choosing a person from one's own ethnic group as a co-offender was about three times as high as would have been expected if the choices had been made at random. Further, some differences between ethnic groups and crime types were observed. Individuals with an African or a South/Central American background were particularly likely to choose co-offenders of the same ethnicity, and assaults and threatening behavior were more likely than muggings to involve ethnically homogeneous groups.

These findings suggest that complex relationships between ethnicity, crime type, and co-offending are likely to exist. However, the varied ethnic profiles of previous samples together with a nearly exclusive focus on violent crime make detailed conclusions about the ethnicity/co-offending relationship, and hypotheses regarding its roots, impossible to draw at this stage. What does seem clear is that overreliance on simple racial dichotomies like "black" and "white" may fail to detect important co-offending variations like those reported in the Swedish data. As research in this area develops, it will be important to consider the most appropriate ethnic categorizations for different geographical areas, and how to best compare them. Another important issue for future investigations of the ethnicity/co-offending relationship and, in particular, the ethnic composition of offending groups will be to control for the effects of potential confounds like geographic racial segregation. As Sarnecki (2001: 136) has warned, given that many co-offending relationships are formed within the confines of one's local neighborhood and activity spaces (see also, Reiss and Farrington 1991), it could be the case that the "the ethnic dimension of co-offending is actually governed in the first instance not by the ethnic background of the actors. . .but rather by the type of residential segregation to be found in a society." Investigations which assess the effect of ethnicity above and beyond that of propinquity (and other variables) will be needed to clarify their respective roles in the selection of co-offenders.

Age

One of the most consistent findings in the co-offending literature is that youths are significantly more likely than adults to commit crimes in groups. Numerous studies conducted across historical time periods and geographical areas have shown that in the aggregate, the age/co-offending curve rises sharply in early adolescence, peaks in the mid to late teens, and decreases steadily thereafter, mirroring the shape of the general age/crime curve (Andresen and Felson 2012; Carrington 2002a; Piquero et al. 2007; Stolzenberg and D'Alessio 2008; van Mastrigt and Farrington 2009). Although this aggregate decrease in co-offending with age is observed across all major studies, there are considerable variations in the specific prevalence rates reported by year of age in different investigations. On the one hand, many studies suggest that "solo offending. . .does not become the modal form of offending until the late teens or early twenties" (Reiss 1988: 151, see also Andresen and Felson 2012). In the Cambridge Study in Delinquent Development, for example, 75 % of offense participations committed by boys at ages 10–13 involved co-offenders, solo offending became the norm between ages 21 and 24 and accounted for 72 % of offense participations at ages 37–40 (Piquero et al. 2007:104). In contrast, some studies suggest that co-offending accounts for less than 50 % of offending at all ages. In an influential Canadian study, lone offending exceeded co-offending from the age of 11 onwards (Carrington 2002a: 289), and two recent large-scale analyses of police data from the USA (Stolzenberg and D'Alessio 2008) and the UK (van Mastrigt and Farrington 2009) revealed that, in the aggregate, the majority of even the youngest offenders committed their offenses alone (see also, Carrington & van Mastrigt 2013). The reasons for these varied findings are somewhat unclear, but as Andresen and Felson (2012) have stated, "whether or not co-offending is reported to be the *norm* at any given age, all existing research agrees that on the whole it is nonetheless *more common* amongst younger compared to older offenders." Further, this holds even when controlling for potential confounds like gender and crime type (Carrington 2009; van Mastrigt and Farrington 2009). The decrease in co-offending with age thus appears to occur independently of variations in other offender and offense attributes.

As Reiss and Farrington (1991) discuss, aggregate declines in co-offending with age could either reflect changes in the population of offenders (e.g., if co-offenders tend to desist at early ages while solo offenders persist) or behavioral changes within offenders (e.g., if offenders become more likely to commit crimes alone with age). In the first empirical test of this question, Reiss and Farrington analyzed prospective longitudinal data from the Cambridge Study. They found a significant negative correlation between age and average number of co-offenders for even the most persistent offenders in their sample, a finding that led them to conclude that the decrease in co-offending with age was a developmental process, which was "not caused by the persistence of solo offenders and/or the dropping out of co-offenders, but instead reflects changes within individual criminal careers" (1991:382; also see Piquero et al. 2007).

More recently, however, evidence for both selective desistance and shifts in individual careers has emerged (McGloin et al. 2008: 174). Trajectory analyses of juvenile co-offending careers in Philadelphia revealed that approximately 50 % of juveniles who demonstrated a pattern of decreasing co-offending in late adolescence desisted, while the other half appeared to shift their criminal careers towards solo offending. Still others exhibited a pattern of increased co-offending behavior over time. The varied findings that have emerged from this and other (e.g., Carrington 2009) studies highlight the risk that "aggregate indicators of co-offending over time can be analytically insensitive and misleading since they mask unique and varying developmental trajectories" at the individual level (McGloin et al. 2008). The need for more disaggregated longitudinal analyses of the age/co-offending relationship, and further research on the impact of early co-offending trajectories for the subsequent criminal career, is paramount (see next section).

In cases where within-individual changes from joint to lone offending do occur, the obvious question is: what underlies the shift? One possibility is that the age-related change in co-offending reflects variations in the normative influence and salience of peers across the life course. Some argue that youths are simply more gregarious and peer oriented than adults; thus, offenders tend to offend alone more often as they grow older because they tend to do *everything* less often with peers as they age (Stolzenberg and D'Alessio 2008; Warr 2002). As Carrington (2009: 1299) states, "children and teenagers like to do things with their peers and in groups, and committing crimes is just another of their group activities. With maturity comes greater autonomy and, therefore, more solo offending."

An alternative explanation is that the shift from joint to lone offending is related to the accumulation of criminal experience (Reiss and Farrington 1991). According to this view, as offenders develop criminal expertise with age, the need for and/or attractiveness of accomplices decreases, particularly in cases where crime proceeds would need to be shared. While this may, indeed, be the case, the current evidence suggests that criminal experience alone does not explain the age/co-offending relationship. In the most recent analysis of the Cambridge Study, for example, Piquero and colleagues (2007) concluded that "the rate of co-offending remains relatively constant for up to eight offenses before declining...[thus]...age may be more important than experience in explaining the decline in co-offending." Recent multivariate evidence from a Canadian juvenile criminal career study also showed that age and experience had independent effects in the prediction of co-offending, although these variables were also related in complex ways (Carrington 2009).

To recap, the current evidence suggests that the offender attribute of age is one of the strongest aggregate-level predictors of co-offending (typically, second only to crime type). The shape of the aggregate age/co-offending curve is robust, observed across gender and criminal experience, and for most major crime types. When the age/co-offending relationship is examined within individuals, however, it is considerably more varied. Many individuals' co-offending trajectories mirror the aggregate curve while others diverge, for reasons that have yet to be identified. Thus, exclusive reliance on aggregate analysis of the age/co-offending relationship may neglect important offender groupings (e.g., individuals with increasing rates of co-offending with age). These "atypical" offenders warrant further attention.

In addition to consistent evidence that co-offending prevalence changes with age, a growing body of research also shows that co-offending dynamics change with age such that, overall, the offending groups formed at older ages tend to be smaller and more heterogeneous. With respect to group size, for example, whereas the majority of co-offenses involve two offenders at all ages, incidents with three or more offenders are fairly common in adolescence, but become very unlikely after age 20 (Reiss and Farrington 1991). In addition to changes in the *number* of accomplices selected by offenders of different ages, there is also strong evidence that the *types* of accomplices selected vary with age. Research from the USA (Warr 2002), Sweden (Sarnecki 2001), and England (van Mastrigt 2008) all reveal a "slight tendency for the age difference between actors and their co-offenders. . .to increase as actors became older." This finding was explored in more detail in a recent analysis of age homophily (van Mastrigt and Carrington 2013) in which youths were found to exhibit significantly greater clustering as compared to adults, even when controlling for chance expectations based on the age distribution of the offending population.

Patterns of gender homogeneity also appear to change with age. Among Swedish juveniles, for example, most co-offending dyads up to age 20 were unisexual (and in particular, all male); however, the likelihood of committing crimes with accomplices of the opposite sex was significantly increased for the oldest offenders (Sarnecki 2001: 54). Similar evidence has also emerged in England, where adults (≤18) in one study were found to have more than twice the odds of offending in mixed-sex groups compared to youths (van Mastrigt and Farrington 2009).

Research further suggests that the locations from which co-offenders are selected varies with age, such that geographic proximity plays a more fundamental role in the selection of co-offenders in adolescence, and becomes less important later in life. Whereas 100 % of co-offenders aged 10–13 in the Cambridge Study lived in the same postal district in London, only 53 % of those aged 21–32 lived in such close proximity. Thus, "although it is clear. . .that propinquity in the selection of co-offenders declines with age, at all ages a sizable proportion of all co-offenders in any offense reside in fairly close proximity to each other" (Reiss and Farrington 1991:389). This finding raises interesting questions about the shared (and divergent) "offender convergence settings" and accomplice pools that facilitate co-offender selection at different ages (see Andresen and Felson 2012).

The aggregate tendency for individuals to select co-offenders less similar to themselves as they age likely reflects the well-known sociological fact that friendships and relationships tend to diversify with age. In childhood, individuals tend to form large, predominately same-sex friendship groups with other kids from the same neighborhoods or schools. As children become teenagers and then adults, they gain increased mobility, enter the workforce, and start to form more lasting friendships and relationships with members of the opposite sex. All of these milestones act to expand the boundaries of individuals' social networks which, in turn, increases their exposure to less similar others, including potential accomplices. More research is needed to explore whether the age-related changes that have been observed in offending groups simply reflect structural constraints related to the changing constellations of individuals' general social networks (from which accomplices are often drawn) or whether these patterns reflect different, crime-specific processes related to aging and co-offending. Teasing these issues apart will require longitudinal comparative analyses of co-offenders' conventional and illicit network ties across the life course. Unfortunately, such data is extremely rare.

In sum, there is clear evidence that in the aggregate, joint participation in crime peaks in adolescence, drops sharply in early adulthood, and settles to a low plateau in later life. However, disaggregated analyses show that the age/co-offending relationship varies somewhat more when examined from the perspective of individual criminal careers. With respect to changes in the composition of offending groups across age, there is evidence for both similarities and differences between youths' and adults' co-offending groups. While the vast majority of all co-offending groups are small, unstable, and relatively homogeneous at all ages, adults nonetheless tend to offend in slightly smaller, more heterogeneous groups than youths.

Criminal Careers

The notion that co-offending involvement generates and/or reflects differential criminal career trajectories is long standing but rarely investigated empirically (not least due to a shortage of the requisite data). In their infamous description of Sidney's criminal career, Shaw and McKay (1931) noted clear changes in both the level and form of Sidney's criminal behavior that were closely tied to his changing accomplice associations. Despite frequent references to this finding across the criminological literature, co-offending remains "one of the most ill studied of all criminal career dimensions" (Piquero et al. 2007: 120). In this section, various dimensions of the criminal career are framed as offender "attributes" and considered as either causes or effects of co-offending. Although research in this area is still in its infancy, a number of key criminal career features including offending style, persistence, and versatility appear to be important.

Co-offending careers may take one of three forms: (1) an exclusively solo career, where the individual always engages in crime alone; (2) an exclusively co-offending career, where the individual never offends alone; and (3) a mixed co-offending career, where the individual commits a combination of joint and lone offenses. For offenders with numerous offenses, "the typical criminal career is a mix of offenses committed alone and with others" (Reiss 1988: 117; Reiss and Farrington 1991; Sarnecki 2001). At the same time, it appears that active offenders also exhibit "a style of offending" (i.e., predominantly group or loner) over time (Warr 1996: 22; Reiss and Farrington 1991; Weerman 2003). In a recent examination of juvenile co-offending careers in Philadelphia, for example, considerable consistency was observed in many individuals' co-offending behavior, such that "some classes [of offenders] contain relatively high-rate co-offenders. . .whereas others comprise relatively low-rate co-offenders," although a number of individuals also exhibited a high degree of versatility in their co-offending behavior with some increasing and others decreasing their co-offending activity over time (McGloin et al. 2008: 174).

The observed between- and within-individual variability in co-offending trajectories begs two questions: what might explain these differences? And what do they matter for the continued criminal career? With respect to the first question, one possibility is that individual differences render particular "types" of offenders more or less likely to engage in co-offending across the criminal career. According to Moffitt's (1993) model of adolescence limited (AL) and life course persistent (LCP) offenders, for example, LCP offenders might be expected to offend alone to a greater extent than AL offenders because they are more internally motivated to offend and less likely to require the social support or practical assistance of accomplices. According to this population heterogeneity view, patterns of co-offending are assumed to *reflect*, rather than shape, varied offender trajectories. Thus, in the AL vs. LCP example above, offenders with *low* co-offending involvement might be expected to exhibit more life course persistent criminal careers. Alternatively, involvement in co-offending might *lead* offenders down the road to persistent offending because it increases criminal opportunities, skills, and/or awareness (Andresen and Felson 2009; Carrington 2009; McGloin and Nguyen 2012). In this case, *high* levels of co-offending involvement might be expected to predict more persistent and chronic offending careers.

A number of studies have investigated the co-offending/persistence relationship, with mixed results. In support of the hypothesis that co-offending should be negatively correlated with offending frequency and persistence, Knight and West (1975) found that in their long-term follow up of London boys, offenders with a history of lone offending were, in fact, less likely than joint offenders to be "temporary" delinquents. Similarly, based on her analysis of UK police data, van Mastrigt (2008: 172) found that over a 3-year period, there was a negative relationship between the number of co-offenders per offense and offending frequency. In contrast, co-offending *increased* the likelihood of persistent criminality in the Philadelphia study, particularly when combined with an early age of onset (Conway and McCord 2002). Age was also found to play an important role in conditioning the relationship between co-offending and offending frequency in Carrington's (2009: 1317) Canadian study. His analysis revealed two distinct delinquent careers (high activity and low activity), which were characterized by different relationships between co-offending, criminal experience, and age; whereas co-offending increased slightly with overall criminal activity for high-activity offenders, it decreased with experience for low-activity offenders. Further, for nearly three-quarters of offense participations – those involving low-activity offenders committed during the teenage years – co-offending decreased with the offender's age, criminal experience, and overall career activity. In contrast, for offense participations of high-activity offenders as teenagers, co-offending decreased with age but varied little with experience and career activity. Finally, for offense participations committed during childhood, co-offending decreased with increasing career activity but varied only slightly, if at all, with the offender's age and criminal experience. These complex differences highlight the need to consider potential interactions between multiple criminal career attributes and co-offending.

In addition to studies identifying positive and negative relationships between co-offending and career activity, a handful of studies have demonstrated null effects. For example, Reiss and Farrington (1991: 376) and Piquero and colleagues (2007: 110) found no relationship between average number of co-offenders per offense and total number of convictions in the Cambridge Study. The conclusion that "the proportion of an offender's offenses that are committed with others does not differentiate chronic from non-chronic offenders" was further supported in a recent analysis of the Racine cohort data (McGloin and Stickle 2011: 14). Interestingly, although chronic and non-chronic offenders had comparable levels of co-offending with peers in the Racine data, less active offenders were significantly more likely to report offending "because of peers." Thus, whether or not high- and low-activity offenders have similar co-offending participation *rates*, it seems likely that the *role* of co-offending differs in the criminal careers of chronic vs. non-chronic offenders. Whereas co-offender influence may directly motivate offending for less serious offenders, accomplices likely play a less causal, but more practical, role in the criminality of chronic offenders, for example, by making some crimes easier to commit.

Irrespective of the relationship between co-offending and the *persistence* of the criminal career, it seems clear that involvement in co-offending with particular types of accomplices has the potential to shape the *character* of one's subsequent criminal career in important ways. In the Philadelphia co-offending study, for example, offending with a violent co-offender in the first offense was a significant predictor of future violence, even when the joint offense itself was not violent (Conway and McCord 2002). In another study, offenders who could identify a criminal mentor (approximately ¼ of whom were also accomplices) tended to report enhanced criminal earnings and reduced periods of incarceration, a finding which led the authors to conclude that "tutelage guides the criminal maturation process" (Morselli et al. 2006: 19). Further evidence for the impact of co-offending comes from a recent study in which juveniles who were tied into large, nonredundant co-offending networks exhibited more versatile criminal careers than solo offenders (McGloin and Piquero 2010). This finding is consistent with other aggregate-level evidence that, compared to solo offending, the distribution of co-offending is more varied across crime types (Andresen and Felson 2012). Together, these findings suggest that offending with others has the potential to enhance exposure to diversified criminal skills and deviant contacts and that the specific content of these exposures matters for the development of the criminal career.

Various offender and criminal career attributes may also impact the role one adopts within co-offending groups. In a recent investigation of self-reported instigation among a sample of incarcerated male offenders in the USA, both early age of onset and crime-specific expertise predicted instigation within crime type, a finding that highlights the importance of considering how individual and situational predictors may work together to shape co-offending behavior in particular offenses and across the criminal career (McGloin and Nguyen 2012).

In sum, the current evidence suggests that co-offending can act as both an important feature, and determinant, of the criminal career. Framing key dimensions of the criminal career as offender attributes and linking them to co-offending behavior over time may hold promise for development of both the co-offending and criminal career literature. Unfortunately, focused studies of the criminal career/co-offending relationship are rare at present. Better longitudinal data on co-offending, and more information about variations in the selection and impacts of co-offenders across the life course, will be particularly important for future development in this area.

Implications and Future Directions

Taken together, the current evidence provides strong support for Waring's (2002: 321) skepticism that "co-offending can or should be treated as a unitary phenomenon." To a greater or lesser degree, sex, ethnicity, age, and several key criminal career features have all been shown to vary with co-offending prevalence and form. This is significant for a number of reasons.

From a theory-development perspective, the fact that at least some of these person-based variables maintain simultaneous and independent effects on the prediction of co-offending, even when controlling for well-established situational-level variables (such as crime type), highlights the importance of integrated models of co-offending. Offender attributes, although important, do not exist in a vacuum and are only one piece of the co-offending puzzle. Explanations which are sensitive to the effects and interactions of individual offender attributes, routine activities, social networks, and situational contexts are likely to shed the best light on co-offending patterns and processes (see Weerman; Hochstetler, this volume). Further, given that observed variations in the extent of co-offending are sometimes greater (and sometimes smaller) than differences in the form of co-offending, theoretical frameworks that allow for conceptual distinctions between the factors that determine whether a crime is committed with others and the factors that determine the *selection* of those others may be particularly powerful.

The co-offending variations outlined here are also important from a policy perspective. First, because co-offending prevalence appears to be higher for some social groups (i.e., females, ethnic minorities, young people, and individuals exhibiting particular criminal career trajectories), these groups will look "worse" in criminal statistics that do not take account of co-offending (for recent discussion of this issue see Andresen and Felson 2009; McCord and Conway 2005; van Mastrigt and Farrington 2009). As outlined earlier, these offender groups will be more affected by the multiple counting procedures inherent to many offender-based crime measures, and their share of crime will be distorted accordingly. Secondly, the fact that accomplice selection processes appear to vary somewhat across offender groups (and also, perhaps, within criminal careers) means that one-size-fits-all efforts to disrupt co-offending networks may not provide the greatest returns. To give but one example, efforts to hamper co-offending via situational crime prevention will likely be effective only to the extent that these efforts are sensitive to identifying how "offender convergence settings" vary across age and gender (Felson 2003). Similarly, targeted interventions aimed at individuals with particular types of co-offending careers may be more effective than those aimed at accomplice networks in general. Individuals who consistently exhibit recruitment or instigation behavior may be of particular interest (although even interventions targeted at these "extreme" co-offenders must recognize the importance of situational variables). In sum, to treat co-offending as invariant across and within individuals may not only obscure important features of the phenomenon from an academic perspective, it may also hinder attempts to address it in practice.

At the same time, attribute-based variations in co-offending should not be overstated. Many of the between-group differences identified in the current literature are small and should not deter efforts to identify general risk factors for co-offending. There is tremendous room for growth in the co-offending literature and need for additional research at all levels of analysis (individuals, offenses, environments). With respect to offender attributes specifically, future directions are endless. Only a handful of the individual-level variables most readily available in official records have been discussed here; however, studies based on other data sources are already beginning to identify other potential correlates of co-offending, including economic adversity and cooperative orientation (McCarthy et al. 1998), being unmarried and undereducated (Alarid et al. 2009), and exhibiting lower levels of psychosocial maturity and higher levels of anxiety and psychopathy (Goldweber et al. 2011). As the literature in this area continues to develop, an even greater number of attribute-level and situational-level variables are likely to emerge as significant correlates of group crime. The challenge for future co-offending researchers, theorists, and policy makers will be to evaluate the role of these variables and the person-by-situation interactions that tie them together. To navigate this complexity will be no small feat.

Take Home Message

There is notable individual-level variation in the extent to which offenders commit crimes alone and with others, and the form that their co-offending takes. Offender attributes such as sex, ethnicity, age, and criminal history have all been shown to impact co-offending, but questions remain regarding the underlying reasons for these effects, and potential interactions with situational and structural factors. There is tremendous room for development in this field; as a first step, it is recommended that information on a wider range of offender, offense, and situational correlates are collected in future co-offending research.

Related Entries

▶ Co-offending
▶ Co-offending and Offender Decision-Making
▶ Theories of Co-offending

C

Recommended Reading and References

Alarid LF, Hochstetler AL, Burton VS (2009) Group and solo robberies: do accomplices shape criminal form? J Criminal Just 37:1–9

Andresen MA, Felson M (2009) The impact of co-offending. Brit J Criminol 50:66–81

Andresen MA, Felson M (2012) Co-offending and the diversification of crime types. Int J Offender Ther Comp Criminol 56(5):811–829

Budd T, Sharp C, Mayhew P (2005) Offending in England and Wales: first results from the 2003 crime and justice survey. Home Office research study, vol 275. Home Office, London

Carrington PJ (2002a) Group crime in Canada. Can J Criminol 44:277–315

Carrington PJ (2002b) Sex homogeneity in co-offending groups. In: Hagberg J (ed) Contributions to social network analysis, information theory and other topics in statistics. A festschrift in honour of Ove Frank. Stockholm University Press, Stockholm, pp 101–116

Carrington PJ (2009) Co-offending and the development of the delinquent career. Criminology 47:1295–1329

Carrington PJ, van Mastrigt SB (2013) Co-offending in Canada, England and the United States: a cross national comparison. Global Crime. doi:10.1080/17440572.2013.787926 (Online First)

Conway KP, McCord J (2002) A longitudinal examination of the relation between co-offending with violent accomplices and violent crime. Aggress Behav 28:97–108

Emler N, Reicher S, Ross A (1987) The social context of delinquent conduct. J Child Psychol Psychiat 28(1):99–109

Erickson ML, Jensen GF (1977) Delinquency is still group behavior! Towards revitalizing the group premise in the sociology of deviance. J Crim Law Criminol 68:262–273

Felson M (2003) The process of co-offending. In: Smith MJ, Cornish DB (eds) Theory for practice in situational crime prevention, vol 16, Crime Prevention Studies. Criminal Justice Press, Monsey

Goldweber A, Dmitrieva J, Cauffman E, Piquero AR, Steinberg L (2011) The development of criminal style in adolescence and young adulthood: separating the lemmings from the loners. J Youth Adolesc 40 (3):332–346

Hakkert A (1998) Group delinquency in the Netherlands: some findings from an exploratory study. Int Rev Law Comput Technol 12:453–474

Knight BJ, West DJ (1975) Temporary and continuing delinquency. Br J Criminol 15:43–50

McCarthy B, Hagan J, Cohen LE (1998) Uncertainty, cooperation and crime: understanding the decision to co-offend. Soc Forces 77(1):155–184

McCord J, Conway KP (2002) Patterns of juvenile delinquency and co-offending. In: Waring R, Weisburd D (eds) Crime and social organization. Transaction Publishers, New Brunswick, pp 15–30

McCord J, Conway KP (2005) Co-offending and patterns of juvenile crime. Research in Brief. NCJ 210360. National Institute of Justice, Washington, DC

McGloin JM, Nguyen H (2012) It was my idea: considering the instigation of co-offending. Criminology 50:463–494

McGloin JM, Piquero AR (2009) I wasn't alone: collective behaviour and violent delinquency. Aust N Z J Criminol 42(3):336–353

McGloin JM, Piquero AR (2010) On the relationship between co-offending network redundancy and offending versatility. J Res Crime Delinq 47:63–90

McGloin JM, Stickle WP (2011) Influence or convenience? Rethinking the role of peers for chronic offenders. J Res Crime Delinq 48(3):419–447

McGloin JM, Sullivan CJ, Piquero AR, Bacon S (2008) Investigating the stability of co-offending and co-offenders among a sample of youthful offenders. Criminology 46:155–188

Moffitt TE (1993) Life-course persistent and adolescence-limited antisocial behavior: a developmental taxonomy. Psychol Rev 100:674–701

Morselli C, Tremblay P, McCarthy B (2006) Mentors and criminal achievement. Criminology 44(1):17–43

Pettersson T (2003) Ethnicity and violent crime: the ethnic structure of networks of youths suspected of violent offenses in Stockholm. J Scand Stud Criminol Crime Prev 4:143–161

Pettersson T (2005) Gendering delinquent networks. A gendered analysis of violent crimes and the structure of boys and girls co-offending networks. Young 13:247–267

Piquero AR, Farrington DP, Blumstein A (2007) Key issues in criminal career research: new analyses of the Cambridge Study in Delinquent Development. Cambridge University Press, Cambridge

Reiss A (1988) Co-offending and criminal career. In: Morris N, Tonry M (eds) Crime and justice, vol 10. University of Chicago Press, Chicago

Reiss AJ, Farrington DP (1991) Advancing knowledge about co-offending: results from a prospective longitudinal survey of London males. J Crim Law Criminol 82:360–395

Sarnecki J (1990) Delinquent networks in Sweden. J Quant Criminol 6(1):31–50

Sarnecki J (2001) Delinquent networks. Cambridge University Press, Cambridge

Shaw CR, McKay HD (1931) Report on the causes of crime. US Government Printing Office, Washington, DC

Steffensmeier D, Terry RM (1986) Institutional sexism in the underworld: a view from the inside. Sociol Inq 56(3):304–323

Stolzenberg L, D'Alessio SJ (2008) Co-offending and the age-crime curve. J Res Crime Delinq 45(1):65–86

van Mastrigt SB (2008) Co-offending: relationships with age, gender and crime type. Ph.D. Dissertation, University of Cambridge

van Mastrigt SB, Carrington PJ (2013) Sex and age homophily in co-offending networks: opportunity or preference? In: C Morselli (Ed.) Crime and networks. Routledge, New York, NY

van Mastrigt SB, Farrington DP (2009) Co-offending, age, gender and crime type: implications for criminal justice policy. Brit J Criminol 49:552–573

van Mastrigt SB, Farrington DP (2011) Prevalence and characteristics of co-offending recruiters. Justice Quart 28:325–359

Waring EJ (2002) Conceptualizing co-offending: a network form of organization. In: Waring E, Weisburd D (eds) Crime and social organization: advances in criminological theory, vol 10. Transactions Publishers, New Brunswick

Warr M (1996) Organization and instigation in delinquent groups. Criminology 34:11–37

Warr M (2002) Companions in crime: the social aspects of criminal conduct. Cambridge University Press, Cambridge, MA

Weerman F (2003) Co-offending as social exchange: explaining characteristics of co-offending. Brit J Criminol 43:398–416

Zimring FE (1981) Kids, groups, and crime: some implications of a well-known secret. J Crim Law Criminol 73(3):867–885

Co-offending and Offender Decision-Making

Andrew Hochstetler
Department of Sociology, Iowa State University, Ames, IA, USA

Overview

This entry reviews literature on group crime and co-offending. This includes occurrence, possible causal significance, and group formation.

It restricts the subject matter to discrete, short-lived street crimes committed by two or more principals in close collaboration while in each others' immediate proximity. It argues that shared memories between participants and the meaning attributed to them (group identities) influence many criminal decisions in groups. The entry reviews what is known of leadership and on function and form of crime groups with an eye toward demonstrating that groups alter individual calculus in criminal decisions in many ways.

Fundamentals

This entry is about group crimes, or co-offending. For conceptual convenience it restricts the subject matter to discrete, short-lived street crimes committed by two or more principals in close collaboration while in each others' immediate proximity. This definition is useful because it eliminates a great many acts that might otherwise be termed group crime. A person who trades in stolen goods with thieves arguably commits group crime and co-offends in each transaction, but does not meet the close collaboration, level of organization, and distance criteria. An offender who retaliates alone against an insult upon his gang has motive that might lead some to conclude that the avenger commits group crime, but this gangster is excluded. A drug dealer distributing to a loose network of smaller-scale dealers is excluded. White-collar offenders often collaborate, but are not street offenders. An accessory who disposes of a killer's firearm might be called a co-offender by some; likewise, a conspirator who lays plans but does not accompany the principals to a crime scene might be included in more expansive discussions than used here. Set these complications aside in favor of a narrow focus.

Background

The study of co-offending has a long and theoretically central role in the development of criminology, a field that traditionally is closely linked to sociology. Because many offenders do not commit crime alone, friendship groups often commit crime together, criminal activities are learned by socialization, and interaction plays a significant role in many criminal decisions, the topic of group crime begs theoretical inquiry. Obviously, the study of crime groups bears upon many key theories about how criminal motivation forms. Empirical interest in the topic is long-standing not only for theoretical reasons but also because of the recurring observation that explanations that focus exclusively on the formation of individual intent may be imperiled by the prevalence of offending groups. Use of aggregate data that does not account for co-offending risks misinterpretation, and therefore attention to the prevalence co-offending is important. For example, the involvement rates of youth in crime or the youthfulness of the offenders responsible for crimes that occur in a given territory may be exaggerated by failing to account for the fact that youth are more likely to co-offend than others.

Another reason that examining and problematizing group crime is the suspicion that many offenses occur as a result of processes of collective behavior where individuals may sacrifice capacity to make personally rational decisions to an emergent collective will. Most crime groups are small groups, so literature on crowds is not especially relevant, however. Nevertheless, there is a rich literature in sociology showing that group decision-making is shaped by processes of social interaction in small groups. Put differently, the decisions of groups are much more complex than a simple aggregate of the preferences of the individuals that compose them. In example, group polarization is a tendency of group decision-making. This is where the decisions that a group makes are more extreme (sometimes more risky) than are the decisions of the individual members. In addition, even short deliberations in a group can shift individual opinions, and those who are adamant and outspoken or thought by group members to be knowledgeable of how to handle the problem at hand exercise the most influence. Individuals shift their opinions most dramatically when interaction or their preconceptions lead them to believe that they hold a minority opinion in the group. There are dozens of well-established findings about how decision-making in groups alters the decisions that the individual members would make if alone, but here it will suffice to point out that decisions made by groups and individuals differ significantly in process and outcome. When an offender rationalizes that a personal decision to engage in crime would not have been made if not for the presence of co-offenders, there is a scientific leg to stand on.

Description

Group crime is common. As early as 1931, criminologists were aware that more than 80 % of offenders in Chicago's Juvenile Courts offended with others (Shaw and McKay 1931). This and later findings from victimization surveys revealed that the young are the most likely to offend in group, that solo offending is comparatively uncommon at young ages, and that lone offending is not the modal form of crime until the late teens (Reiss 1986). Offenders under the age of 13 are more likely to commit crime in groups than are those in later teenage years; about 40 % of juvenile offenders commit the majority of their crimes with others (Gonzales et al. 2005). Of course, young people compose a disproportionate share of the population of active offenders, and this age distribution of criminal offending is partially responsible for the fact that much crime is group crime. Despite its concentration among the young, the group nature of offending is seen in many offenses both trivial and deadly, and group offending is not rare in any age group.

An examination of personal robberies in England and Wales revealed that 62 % had multiple offenders and in 30 % of robberies there were more than two offenders (Smith 2003). A descriptive 1999 study of US carjacking revealed that 55 % were committed by two or more offenders and 45 % by a lone offender (Klaus 1999). Somewhat more surprisingly, one US state's analysis of domestic violence incidents revealed that 18 % involve two or more offenders acting in concert (Idaho State Police 2003).

As can be seen, the mean level of co-offending varies considerably by offense. Generally, levels of co-offending are high for acquisitive crimes including burglary, robbery and theft (especially for motor vehicle theft), and for arson. (The latter is an acquisitive crime in some cases.) Group offending levels are low for sexual assault and similar sex crimes and, perhaps counterintuitively, they also are low in drug offenses (Carrington 2009). Findings are not altogether consistent when it comes to which crimes are most often group crimes, and results seem conditional on data sources; for example, some studies indicate that violent offenses have a higher number of offenders than nonviolent crime (Mcgloin and Piquero 2009).

Group crime undoubtedly is common, but the preceding discussion illustrates that lone offenders are responsible for a great many crimes, so the group nature of crime perhaps should not be overstated. In 2004, of all completed violent crimes reported in the National Criminal Victimization Survey where victims could estimate the number of offenders, 79 % were committed by a lone offender, 9 % by two offenders, 5 % by three, and 4 % by four or more. It is probably safe to say that the lion's share of crimes (as measured by the ratio of group crimes to total crimes) is committed by lone offenders, but it is abundantly clear that the extent to which individuals offend in groups is ignored at peril to the analyst of crime in part because a group offense must have at least two offenders, but in 9 % of crimes there are three or more (Bureau of Justice Statistics 2005). Interaction with co-offenders certainly influences many criminal decision-makers.

Crime group composition largely reflects what is known of the distribution of criminal offending. For example, crime groups disproportionately are composed of the young and of males. In 2009, of the violent crimes involving multiple offenders, 32 % all were under the age of 20, 19 % all were 20–29, 10 % were all 30 or older, and 28 % involved offenders of mixed age (Bureau of Justice Statistics 2009). Female offenders are more likely to commit crime in groups than their male offender counterparts so that a higher percentage of female crimes are group crimes (Mastrigt and Farrington 2009).

Nevertheless, in 2009, 62 % of group crimes had all male offenders and only 11 % contained all female offenders, reflecting the disproportionate distribution of criminal participation by gender (Bureau of Justice Statistics 2009). It is self-evident that females are much more likely to offend in mixed sex groups than males. Perhaps, this is because by most measures females are less criminal than males, and being with males, therefore, enhances chances of offending. Males score high on many of the troublesome psychological attributes related to crime and externalizing problem behavior. Moreover, they may be particularly susceptible to the temptation to show off by sensation-seeking and risk-taking behaviors, characteristics easily linked theoretically to crimes committed both alone and with others.

Conceptual Issues

The discussion of co-offending is part of the larger theoretical exchange in criminology as it speaks to the most important questions about why individuals offend. Co-offending may be a mechanism whereby criminal attitudes are passed along, where offenders learn the rewards and costs of crime, and for transferring criminal knowledge, as well as a context where social backgrounds impinge on decisions. Perhaps the thorniest question concerning real-world crime groups, as opposed to decision-making groups in laboratories or experimental vignettes, is whether or not they are greatly significant in shaping individuals' decisions or are incidental/instrumental gatherings of criminally inclined individuals. To speak to this issue, it is important to understand what is known about how crime groups are formed and proceed.

Significance of Group Effects

A primary reason for skepticism about the importance of groups is belief that crime groups are mere instruments for carrying out individual motivations. A related issue is that what are seemingly group effects may be an artifact of the types of individuals selecting into crime groups. This discussion is similar to the flocking or feathering debate that occurs in the study of peer effects. Scholars continue to debate whether the influence of peers shape criminal choice or if it just seems to be the case because those inclined to crime select friends with similar predisposition, although evidence for the former is very strong (Haynie 2002). Likewise, considerable ink has been spilt contesting whether co-offending is a significant means of criminal socialization, primarily important only in its effects on individuals' situational calculations of criminal payoff, or an incidental outcome of individual criminal predisposition. This debate continues although the most obvious reconciliation is an integrated acknowledgement that crime develops from individual proclivities, group selection and socialization, and from situational inducements and altered decision-making that can occur in groups.

Because solo offending is common and because youths are gregarious and do many activities in groups, not just crime, it is easily argued that groups should not be viewed as particularly significant in the development of crime or as a form of criminal tutelage. To the extent that groups do have effects, some believe that their influence is straightforward; group effects exist because the copresent can change situational inducements to crime at the margins and not because crime groups have lasting, particularly influential, or decisive effects on individual decision-making (Stolzenberg and D'Allesio 2008).

On one hand, there is something to be said for caution against exaggerating the place of groups in criminal decisions. It is apparently true that many persons are unlikely to find themselves in circumstances where friends recruit them to burglary or shift their willingness to take the risk of an armed robbery. Stable personality characteristics and attributes that predate group interaction take causal precedence. Few become repeat felons without subtle, and often, lifelong inculcation of attitudes that make such behaviors seem, at least circumstantially, acceptable. Thus, individual attitudes probably serve as a more central means for transferring criminal disposition socially than situational interaction in crime groups. Therefore, becoming a burglar has little to do with the group that accompanies to any given offense – rather birds of a feather flock together.

It is true also that offenders change their accomplices often as their criminal careers unfold. This implies that the influence of any particular group is at best transitory and adds credence to the view that the influence of crime groups in most cases is best understood as a moderately significant situational inducement.

On the other hand, there is good reason to think that co-offending plays a significant role in socialization to crime. For example, the majority of offenders in their first co-offense commit delinquency with an older accomplice, suggesting that crime groups are an important mechanism of socializing younger youth into the delinquent ways of their older peers (McCord and Conway 2002). In further evidence, once a youth co-offends with a violent offender, the chances of that youth subsequently committing a violent crime increase (McCord and Conway 2002). In fact, the distribution of co-offending across crime types is more varied than the distribution of solo offenses (Andresen and Felson 2011).

Moreover, the importance of criminal capital, usually construed as the abilities and networks to accomplish crime, increasingly is seen as being a significant exogenous factor shaping the development of criminal careers and decisions. It would be difficult to imagine that co-offending does not contribute to the development of criminal capital. This suggests that co-offenders have lasting influence on criminal careers even if no single event is transformative. Peer attitudes and behaviors conducive to crime via socialization probably go hand-in-hand with the formation of successive crime groups and operate reciprocally. Indeed, satisfactorily isolating effects of peer socialization, incident-level group effects, and individual characteristics in surveys to discover why or how criminal friends and associates are important would be an extremely difficult problem to handle empirically.

Without undermining the flocking perspective, or diminishing the fact that even the young often offend alone, it is safe to say that persons belonging to peer groups that share and instill crime-conducive attitudes likely will find themselves in a crime group eventually and that their behaviors will be more likely to conform to those of willing and copresent friends as the group decides whether to seize a particular criminal opportunity. Groups may be important in forming predisposition to offend, teaching attitudes and techniques for offending, and for determining whether or not an individual decides to commit a crime on a particular day. The relative importance of competing theoretical characterizations of groups in explaining crime and how crime group membership should be cast in causal modeling remains to be sorted out. It would be precipitously premature, therefore, to discount the longstanding contention in criminology that direct exposure to others' criminal behavior serves as a significant form of tutelage and criminal motivation.

Group Assembly

While investigators have devoted considerable empirical attention to general selection of criminal peers, they have devoted far less to the process of assembly, selection, and recruitment of co-offenders. This is regrettable as the decision to enter a group that eventually commits a crime may be a more significant decision point in shaping criminal decisions than consideration of foreground variables like target characteristics. Background characteristics that predict criminality may work their way into the foreground of discrete criminal choices by way of the decision to join a crime group.

Certain environments are fertile for the formation of crime groups, partially explaining why some areas produce more crime than others. For example, crowded impoverished neighborhoods also are those where young persons and adults congregate on public street corners and engage in unstructured activities, like "hanging out" most often. Of course, open air drug markets where unemployed gangsters, dealers, users, and chronic addicts congregate are the perfect social and spatial setting for robbery and burglary groups to form. The degree to which unstructured activity pervades in an area amplifies the opportunity for crime and the formation of crime groups and thereby affects the local rate of group crime (Osgood and Anderson 2004). Crime in the form of co-offending often springs from "offender convergence spaces" (Felson 2003). These settings allow offenders to find and recruit each other to crime and thereby offer an ongoing local structure for criminal cooperation even though most crime groups may be short-lived and informally organized. Clearly some settings are conducive to crime group formation, but thus far the correlates of these locales are not thoroughly understood; for example, some preliminary evidence suggests that on the whole urbanity (urban environments) decrease co-offending behavior. One possible reason is that where people are interpersonally estranged and close to few, they have a more difficult time recruiting co-offenders; this may offset the socio-spatial conveniences of the city that make group crime likely (D'Allesio and Stolzenberg 2010).

The activities that persons engage in routinely and lifestyles that persons carry out shape the formation of their social networks and the persons they contain. Persistent offenders often devote their days and nights to drug and drink and avoid structures of disciplined work lives. They inhabit social spheres where they are sure to encounter others who do the same, and therefore, it comes as no surprise that many criminal schemes are hatched in the context of shared drugs and drink. In these immediate environments, say inner-city heroin shooting galleries or rural swimming holes where midday partying occupies the time, offenders are likely to encounter friends and potential co-offenders. A great many crimes begin with groups of unemployed young men driving aimlessly about town or loitering on street corners while others are at offices, factories, and construction sites. The foreground circumstances and preceding activities where crime groups form probably have as much to do with how most crimes develop than explicit and intentional recruitment to carry about previously formed criminal intentions.

It should come as no surprise that very little thought goes into planning the typical crime. This spontaneity of crime means only that crime is seldom preceded by careful attention to detail or lengthy deliberation and discussion. One reason is that some of the members in crime groups might have offended together in the past, making criminal opportunity in the group implicit and the intent to commit crime partially formed by the mere assembly of compatriots. In addition, persistent offenders often are at the ready to commit crime as they carry on with their daily lives should the opportunity arise, sometimes even having a tentative mental list of particular targets that they are considering. Should they encounter willing accomplices or a need, they already have loosely laid plans. These realities qualitatively alter the meaning of any statement that crime seldom is planned or that crime groups typically form without criminal intentions.

Crime groups often move toward crime incrementally by way of decisions where the possibility becomes increasingly clearer to members, but crime's occurrence often is far from a foregone conclusion up to the last instant. Groups that begin the day partying together may change composition as they move toward crime, for example. The reluctant and cautious may drop away as trouble becomes increasingly likely. Males may leave behind more cautious females; revelers who have responsibilities to attend to may depart, leaving the more fully committed to continue their misbehavior. The stage is set for crime by framing a shared understanding of the potential found in a group's members and in how activities transform its composition (Hochstetler 2001). Young men with growing arrest records who are 2 days into a methamphetamine binge and unready to sleep can be fairly certain that their current associates see criminal possibility in the group; those still present are likely to mark the rewards of emergent criminal opportunities as attractive. These incremental developments are noted here to provide a qualitative sense of how crime groups form and to point out that the preceding activities of a group contribute to the decisions it makes.

Like individuals, groups have identities. Crime can be a significant shared secret that overshadows other meanings and expectations ascribed to particular company. It is not easily forgotten once a group commits a crime. The next time the group gets together after each offense, criminal possibilities need less consideration. The coming together of the group may be construed as the first indications of criminal possibility or opportunity. This is more than conjecture. Empirically, it appears that crime groups can be typified by offenses that their members are willing to commit, and this leads to specialization beyond what would be expected if crimes were selected randomly by groups (Warr 1996). In other words, groups perceive criminal opportunities based on what they already have done. Law enforcement detectives also can tell that they often know when they have an active criminal group in a community not only because crime reoccurs but also because of the way it is done (modus operandi), and because apparent division of labor replicates recent events. Interviews with offenders reveal that sometimes friendship groups have been gathering for some time without offense but become active crime groups that offend frequently after successfully committing a first crime together (Hochstetler 2001). One might say that criminal possibility and modus operandi become primary parts of the group's repertoire or identity.

The background of individual members may play a large part in crime group formation (Thornberry et al. 1993). Individuals select compatriots based on proximity, family and neighborhood ties, education, and other cultural similarities. They often prefer to associate with people who have similar life experiences. It is not uncommon for siblings or other relatives to offend together or for young people to first encounter those who will be co-offenders in the future while playing in their neighborhoods. The familiarity of many offenders with their co-offenders is reflected in the fact that co-offenders within a group typically match each others' race. Among the juveniles in one study, the ethnic identity of co-offenders and accomplices matched for 96 % of black offenders, 83 % of white offenders, and 83 % of Hispanic or other offenders (Mccord and Conway 2002). Crime groups are predominantly unisexual and tend to be age homogenous. This homogeneity reflects that crime groups form from those close at hand.

Although many offenders know their crime partners well, this need not be the case. Longtime and frequent offenders often have little trouble finding and selecting willing and able accomplices. They are aided in doing so by the identifiable signs of cultural and criminal capital. For example, body markings (scars, tattoos), dress, language patterns, and interests allow offenders to solve the special dilemma they face in identifying and recruiting willing crime partners and associates. Fashions change but cultural markings may nevertheless indicate not only that one is potentially willing to commit crime but also those who portend to abide by codes of the underworld (Gambetta 2009).

For the same reasons, some of us avoid certain persons the potential criminal collaborator may be attracted to signals of criminality. The offender may be drawn to persons whose social and legal histories imply criminal willingness and repelled by indicators of intolerance for crime. In many instances, criminal reputation precedes and this obviously aids in recruitment of accomplices. Just as criminal opportunity is associated with the assembly of criminally experienced groups, it is also found in the presence of persons known or thought to be offenders individually. Possible criminal interpretations of emergent situations and opportunities are contingent in large part on meanings ascribed to the presence of certain people, the history of the group, the nature of one's contacts, and the form of situations and companions in activities immediately preceding criminal situations.

Those heavily engaged in crime are likely to know many potential accomplices and therefore have little difficulty recruiting. Unsurprisingly, persons who commit crime frequently have large networks of potential accomplices (Warr 1996). Offenders who belong to open networks, as opposed to networks where there is a great deal of redundancy in members' relations, exhibit greater offending diversity in the group crimes they commit (McGloin and Piquero 2010). This suggests that open networks free offenders to engage in diverse pursuits by expanding collaborative potential. Offenders that have many crime partners also are more likely to commit violent crime (McGloin and Piquero 2009). Because the larger network structure surrounding group formation shapes opportunities, it is important for the analyst of crime groups to remember that they are not only an immediate context but also part of the larger context of offenders' networks.

Peers' crime-condoning attitudes and criminal behavior have powerful effects on friends' attitudes and behaviors, evidencing the importance of peers for criminal socialization. Membership in some criminal groups also predicts offending patterns. For example, gang-involved youth are more criminally active than other youth by far (Espensen and Huizinga 1993). In addition, entering a gang elevates individual offending by comparison to prior levels, and offending decreases upon leaving a gang (Thornberry et al. 1993). It is arguable that street gang membership and participation in informally organized crime groups are analytically distinct behaviors, but what is found of gangs is at least illustrative of general processes that probably are present in other crime groups.

Traditionally, law and criminal justice procedure recognize the hazards of criminal association in both criminal organizations and in ordinary interactions. Socialization between criminal offenders has long been discouraged in a variety of rules that govern who known offenders associate with and allowable activities in these interactions. For example, states grant parole with the stipulation that the parolee not associate with anyone that they know to have a felony record; one reason for this clearly is that when a person who is predisposed to crime encounters or socializes regularly with others who are the same, the chances of parole failure increase. Risk assessment instruments used to classify offenders as more or less dangerous often include the degree to which a person socializes with offenders as compared to more upstanding associates as considerations.

Instigators, Leaders, and Followers

Recall that one of the reasons for focusing on the group nature of many crimes is that making decisions in groups affects what individuals decide to do. In other words, the collaborative nature of many offenses suggests that explanations based solely on the psychological characteristics of offenders fall short inevitably. As the small group decision-making tradition of inquiry has shown, some members have more influence over the decisions that a group will make than do others even when members occupy equivalent positions. In crime groups, participants may be consistent instigators or relatively reluctant across crime events. Some may be leaders while others are inclined to follow.

Instigators are those who spark the motivation for crime by presenting criminal ideas and opportunities, by assembling tools for crime, by persuading and allaying fears or encouraging, and by taking actions that constrain options for the copresent so that crime becomes a more attractive option. Leaders are those who exercise disproportionate influence over group actions. As Warr (1996, p. 17) explains of delinquency, "conventional theories of delinquency seem to portray delinquent groups as mere aggregates of like-minded or similarly motivated individuals. Yet it is difficult to believe that all group members are equally motivated or inclined to break the law on any given occasion" (Warr 1996, p. 17). What makes a leader or instigator can vary from crime to crime, but it is clear that such roles exist. Indeed, for group sex offending by juveniles, 1/3 of the crimes reportedly had a clear leader who orchestrated the crime as it happened (Bijleveid et al. 2007). Most youthful offenders who commit crime in groups assert that one person first suggested the crime, reflecting the rather obvious reality that groups seldom act with completely collective and mutual will (Warr 1996).

The extent to which there are consistent group instigators and leaders across multiple crimes is a different matter. As crime groups are informally organized, there may not be great consistency in who occupies the leadership or instigator role from crime to crime as a single group continually offends; there may be even less consistency in the role taken as an individual moves from crime group to crime group. The degree that stability exists in roles is a question that has generated an ongoing exchange in contemporary criminology.

Much of this discussion centers on the assertion that high-rate offenders are particularly influential in criminal group formation, recruitment, and decision-making and, therefore, "offer a potentially important target for intervention efforts" (Reiss 1988, p. 117). Those offenders who are engaged in frequent offending and who are persistent offenders may train and influence others to commit crime. High-rate offenders may serve a key role in inspiring discrete criminal events and in passing criminal learning across various groups and time; high-rate offenders can serve as role models for peers and influence those who are marginally inclined to offend over their tipping point. For example, life-course persistent offenders are thought by some to not only have distinct patterns of early involvement in crime and background predictors that are different than those of other groups but also to recruit and exercise more decisive influence over criminal decision-making. Low-rate offenders who tend to have fleeting careers in crime and who hail from less disadvantaged circumstances may adopt a more passive role. Low-rate offenders sometimes are portrayed not as the influential but as the influenced in characterizations of offender types and their behavioral manifestations.

Evidence reveals differences between high- and low-rate offenders in co-offending patterns (Warr 1996). Co-offending is most common in youth and decreases with age, perhaps suggesting that as offenders age their need for accomplices diminishes; moreover, this effect is pronounced among low rate offenders as they progress in a career. The relationship between co-offending and offending with high frequency generally is found to be weak or nonexistent (Reiss and Farrington 1991; Piquero et al. 2007), but the relationship between co-offending and age appears consistently and is strong.

When offenders are divided in groups based on offending frequency, high-activity offenders' co-offending is slightly less common to start and remains at consistent levels as they gain criminal experience (Carrington 2009). Low-activity offenders commit a higher proportion of crimes with co-offenders at the beginning of their careers but commit fewer crimes with others as they gain experience. This suggests that "offending plays a different role in the careers of high and low-activity offenders" (Carrington 2009, p. 1321). One interpretation is that low-rate offenders gain the confidence in their initial criminal forays to go it alone in the future. Another is that they learn that crime pays more when alone as they move from expressive to instrumental motivations. Higher-rate offenders by contrast may have learned early on that certain crimes are better done with others and other crimes work better alone. They may have developed instrumental reasons for crime early in careers, find occasional utility in co-offending, and find partners as needed. Put differently, when it comes to deciding to commit crime in groups or alone, high-rate offenders do not seem to specialize in group or solo crime, but low-rate offenders begin with others and proceed toward lone offending.

Experienced offenders may participate in smaller groups on average. The mean number of co-offenders declines with experience (Piquero et al. 2007, p. 110). When investigators calculate the odds that an offender would be a solo or group offender based on an array of independent variables including criminal experience, they find that the mean number of offenders and offenses are negatively correlated (Van Mastrigt and Farrington 2009). If co-offending helps in the commission of some crimes but is not viewed as necessary in most, it makes sense that group size declines with criminal experience. Typically, few accomplices are needed for crime so that in crimes with instrumental motives as opposed to expressive motives, the payoff is likely to be greater and risks lower with fewer partners.

Offenders who commit crimes with a large number of co-offenders as they move from crime to crime are deemed "recruiters" in one study (van Mastrigt and Farrington 2011). These tend to participate in groups that are heterogeneous and stable, to commit property crimes, and to be older than other offenders. The profile of the recruiter, therefore, resembles somewhat that of the persistent offender or persistent property offender. This group already is known to be particularly troublesome and difficult to integrate into conventional life, but evidence that they recruit those who are different means that they are even more costly than most suppose. If this group of offenders could be turned to more conventional pursuits, the results on aggregate rates of crime potentially could be exponential across generations.

Evidence for identifiable roles in crime groups and that some individuals tend to occupy them across groups should not be overstated. In group crime incidents roles are unequal, a fact that is apparent and recognized in centuries of laws dealing with differential legal culpability in crime. But, roles in crime groups may not be associated consistently and strongly with types of persons. It is doubtful if consistent and reliable psychological constructs will be found that clearly distinguish leaders from followers in group crimes in any prospectively predictive sense. One reason is that variation in measureable and stable psychological traits is constrained by the fact that all who commit group crime are offenders, and thus the comparison group of lone offenders is similar in most respects.

In addition, identification of psychological differences would be surprising given that group crimes emerge from dynamic and informal decision-making and action processes. The person who ends up being a leader and a follower in a group may be highly contingent on situational development. Indeed, one can easily imagine that the persons considered leaders or followers may shift from second to second as individuals' enthusiasm for the offense is likely to fluctuate as the situation unfolds. Before finding stable characteristics of leaders and followers, or instigators and followers, we will at least need to come to agreement on firm definitions and measurements for these terms. Nevertheless, common sense should tell us that there are persons with individual characteristics that contribute to the development of criminal situations and that these characteristics can make them more likely to play instigator and leadership roles.

Functions and Form

There may be very rational components to the formation and use of some crime groups assembled by individual participants. This clearly is true of the very rare professional thieving and drug distribution crews who may put together the best team that they can for the task they need to accomplish. Although this sort of organization makes for great caper movies, it is exceedingly rare. More common is for crime groups to have only some design elements that evidence intent and functionality. The opportunity for some crimes clearly is contingent on the number of co-offenders present, for example. Offenders likely consider whether they have the ability to overcome their adversaries in muggings and street fights where the number of allies matters. Some burglars will not go into a home or to the second floor without an outdoor lookout or getaway driver. It takes two to tangle with an ATM machine, a strong chain and a pickup truck in tow-chain ATM heists.

Persistent offenders may assemble a crew with some consideration of the best persons available to them to accomplish certain tasks or some consideration of their partners' trustworthiness. Although many crime groups form without careful consideration and some form in an instant, it may be wise for criminologists to keep in mind that there are some criminally connected and crafty individuals in street life; these persons may go so far as to deceptively distance themselves from the more unpleasant and risky tasks when putting together a crime. In addition to the functional advantages for such manipulators, committing crime in a group may have material and psychological advantages such as increased confidence gained from planning and committing crime together that one was fearful to do alone. Among the robbers in one study, co-offenders reported a greater sense of control during their crimes and co-offending led to increased planning (Alarid et al. 2009).

Overrepresentation of the young and of males in crime groups may shape the typical processes of crime group decision-making. In some tasks, male risk-taking is enhanced by the presence of other males, while in others the presence of a female increases the tendency to show off. Females already known to males, such as girlfriends, may be a restraining influence, but in other circumstances the desire to impress other males and females is apparent (see Frankenhuis et al. 2010 for a review). There can be little doubt that the desire to impress others by engaging in what is perceived to be bold and fearless behavior is often associated with being masculine, and this has a great deal to do with how group effects can influence crime, particularly for the young.

It would defy logic and the findings from decades of small group research to think that group dynamics are not shaped by cross-age interactions in one way or another. Street offenders also exist in very patriarchal worlds, and it is likely, therefore, that the influence of male offenders in mixed gender crime groups may be disproportionate. Role assignment also may reflect this fact; in robberies with mixed gender groups, participating males surely are more likely to wield weapons while females more often are assigned to serve as getaway drivers or lookouts. Age composition also may shape crime group interaction. Younger persons often defer to older ones, and older persons may take on characteristics of the young in their presence, such as enhanced risk-taking. Age and other indicators of experience or skill may lead other offenders to defer equal participation in decision-making. One thing is clear: there is much to be learned about the process of co-offending and how it alters individual criminal calculus.

Related Entries

- ▶ Co-offending and Offender Attributes
- ▶ Co-offending and Offender Decision-Making
- ▶ Gangs and Social Networks
- ▶ Theories of Co-offending

Recommended Reading and References

Alarid LF, Hochstetler A, Burton V (2009) Group and solo robberies: do accomplices shape criminal form? J Crim Justice 37:1–9

Andresen MA, Felson M (2011) The impact of co-offending. Br J Criminol 50:66–81

Bijleveid CCJH, Weerman FM, Looije D, Hendriks J (2007) Group sex offending by juveniles: coercive sex as group activity. Eur J Criminol 4:5–31

Bureau of Justice Statistics (2005) National criminal victimization survey crime incident report, 2004. U.S. Government Printing Office, Washington, DC

Bureau of Justice Statistics (2010) National criminal victimization survey crime. Incident report, (2009). U.S. Government Printing Office, Washington, DC

Carrington PJ (2009) Co-offending and the development of the delinquent career. Criminology 47: 1295–1329

D'Allesio SJ, Stolzenberg L (2010) Do cities influence co-offending? J Crim Justice 38:711–719

Espensen FA, Huizinga D (1993) Gangs, drugs, and delinquency in a survey of urban youth. Criminology 31:565–589

Felson M (2003) The process of co-offending. In: Smith MJ, Cornish DB (eds) Theory for practice in situational crime prevention, crime prevention studies, vol 16. Criminal Justice Press, Monsey, pp 149–167

Frankenhuis WE, Dotch R, Karremans JC, Wigaboldus DHJ (2010) Male physical risk-taking in a virtual environment. J Evol Psychol 8:75–86

Gambetta D (2009) Codes of the underworld: how criminals communicate. Princeton University Press, Princeton

Hochstetler A (2001) Opportunities and decisions: interactional dynamics in robbery and burglary groups. Criminology 39:737–764

Idaho State Police, Statistical Analysis Center (2003) Intimate partner violence: a NIBRS analysis. Idaho State Police, Meridean

Klaus P (1999) Carjacking in the United States, 1992–1996. Bureau of Justice Statistics Special Report, NCJ 171145. Bureau of Justice Statistics, Washington, DC

McCord J, Conway KP (2002) A longitudinal examination of the relation between co-offending with violent accomplices and violent crime. Aggress Behav 28:97–108

McGloin JM, Piquero AR (2009) I wasn't alone: collective behaviour and violent deliniquency. Aust N Z J Criminol 42:336–353

Osgood DW, Anderson AL (2004) Unstructured socializing and rates of delinquency. Criminology 42:519–550

Piquero AR, Farrington DP, Blumstein A (2007) Key issues in criminal career research: new analyses from the Cambridge study in delinquent development. Cambridge University Press, Cambridge

Reiss AJ (1986) Co-offender influences on criminal careers. In: Blumstein A, Cohen J, Roth JA, Visher CA (eds) Criminal careers and "career criminals", vol 2. National Academy Press, Washington, DC

Reiss AJ (1988) Co-offending and criminal careers. In: Tonry MH, Morris N (eds) Crime and justice, vol 10. University of Chicago Press, Chicago

Reiss AJ, Farrington DP (1991) Advancing knowledge about co-offending: results from a prospective longitudinal survey of London males. J Crim Law Criminol 82:360–395
Shaw CR, Mckay HD (1931) Social factors in juvenile delinquency. Report on the causes of crime, vol 2. National Commission on Law Observance and Enforcement, Washington, DC.
Smith J (2003) The nature of personal robbery. Home Office Study 254. Home Office, London
Stolzenberg L, D'Allesio SJ (2008) Co-offending and the age-crime curve. J Res Crime Delinq 45:65–86
Thornberry TP, Krohn MD, Lizotte AJ, Chard-Wiershem D (1993) The role of juvenile gangs in facilitating delinquent behavior. J Res Crime Delinq 30:55–87
Van Mastrigt SB, Farrington DP (2009) Co-offending, age, gender and crime type: implications for criminal justice policy. Br J Criminol 49:552–573
Van Mastrigt SB, Farrington DP (2011) Prevalence and characteristics of co offending recruiters. Justice Quart 28:325–359
Warr M (1996) Organization and instigation in delinquent groups. Criminology 34:11–37

Corporate Crime

▶ Corporate Liability
▶ Crimes of the Powerful
▶ White Collar Crime
▶ Women and White-Collar Crime

Corporate Crime Control

▶ Compliance and Corporate Crime Control

Corporate Crime Decision-Making

Nicole Leeper Piquero[1] and Eric J. Connolly[2]
[1]The University of Texas at Dallas, Richardson, TX, USA
[2]Florida State University, Tallahassee, FL, USA

Synonyms

White-collar crime

Overview

Corporate crime is often viewed as a specific form of white-collar crime and consists of violations of either criminal, civil, or administrative law committed by an organizational actor (e.g., the corporation) or by a representative of the organization (e.g., an employee acting on behalf of the company) to further the interests and goals of the organization. In other words, the organization and not the individual benefits from the illegal act(s). The rational choice theoretical framework has frequently been used to explain and test the corporate crime decision-making process. Generally, this framework assumes that crime and deviance is a choice or an outcome reached as the result of a costs and benefits calculation and that the decision-making process is influenced by the situation. In other words, decision-makers (e.g., managers, CEOs) choose to break the law when the perceived benefits attached to the act outweigh the perceived costs attached to the act. In the corporate crime context, the individual actor has multiple loyalties, such as to the company, to his or her family, and to religious and moral beliefs, and as such the decision to commit corporate crime will be affected by both individual- and firm-level characteristics and influences. Research examining corporate crime decision-making as a rational choice calculation tends to indicate that individual moral evaluations inhibit criminal behavior, whereas a criminogenic corporate culture and benefits of the act to both the firm and the individual actor promote or encourage violations. More research exploring direct and indirect effects of informal and formal sanctions on individual corporate offending is needed to help further the understanding of the process of decision-making in an organizational context.

Fundamentals

The study of corporate crime can be traced back to the influential work of Edwin Sutherland. He introduced the term "white-collar crime" during his presidential address in 1939 to the American

Sociological Society to draw attention to the crimes not commonly regarded as criminal at the time; those would be the crimes committed by individuals in the "upper classes." Sutherland (1983, p. 7) would go on to define the concept as "a crime committed by a person of respectability and high social status in the course of his occupation." Even though Sutherland's offender-based definition of white-collar crime focused on the characteristics of the individual actor, his seminal publication (1983), *White Collar Crime*, tabulated data from court and administrative commission decisions against the 70 largest manufacturing, mining, and mercantile corporations in the United States to illustrate his point that those from the upper classes also engage in criminal behavior. His examination included offenses that today would be categorized as corporate crimes including antitrust violations, false advertising, infringement of patent, trademarks, and copyrights, as well as unfair labor practices (as defined by the National Labor Relations Board). Ultimately, he found that over the life careers of each of the 70 corporations, each company had at least one adverse decision against it with a mean of 7.8 decisions per corporation (Sutherland 1945). Sutherland used this data to further his argument that sociologists and criminologists needed to turn their attention to the crimes committed by people in high-level corporate positions. Since then, scholars studying white-collar and corporate crime have honored Sutherland's vision and advanced the understanding of white-collar and corporate crime. Within this line of investigation, one area that has received specific attention has been corporate crime decision-making.

Scholars who study white-collar and corporate crime often contend that these offenders are distinctly different, at least qualitatively, from ordinary street offenders. Two data collection efforts from the 1980s examined federally convicted white-collar offenders that allowed for the first time an analysis of white-collar offender characteristics (Forst and Rhodes n.d.; Weisburd et al. 1991). These studies revealed that white-collar offenders were generally from middle-class backgrounds, had higher education levels than street offenders, and were repeat offenders. White-collar offenders are also typically believed to be rational calculators, and as such they are believed to be more amendable to both risks and rewards, that is, they are believed to be somewhat rational in nature. In contrast, street offenders typically tend to come from disadvantaged backgrounds, they are undereducated, they seek to gain immediate gratification, and they tend to be more impulsive rather than rational in nature. Two empirical studies comparing street offenders to white-collar offenders confirmed that both types of offenders come from different social and demographic backgrounds (Benson and Kerley 2000; Weisburd et al. 1991). Thus, the questions are how do individual differences affect the decision-making process of offenders and do these processes vary across offenses occurring within an organization or on the street.

Contextual factors surrounding the decision-maker are believed to effect the individual's evaluation of likely benefits and consequences for committing crime. Past research has shown that a much juvenile street crime occurs in groups where drugs and alcohol are typically present, and peers build on the immediate excitement and thrill associated with deviant behavior while downplaying the possible negative outcomes attached to their actions (Cromwell et al. 1991). In contrast, corporate crime occurs within an organizational setting where managers tend to focus on the long-term consequences of actions (Clinard and Meier 1979). According to Braithwaite and Geis (1982, p. 302), this is the case because corporations are future oriented, concerned about their reputations, and make decisions to violate the law based on the perceived future benefits for the organization. Therefore, where young street criminals tend to misjudge potential costs and benefits linked to crime, corporate criminals are believed to be more rational in their calculation of likely outcomes because they have an eye toward advancing both long-term individual- and organizational-level interests.

Key Issues

The decision or choice to engage in crime (or not) is at the heart of the rational choice theory of crime. This theory is an extension of the deterrence doctrine which tends to focus only on formal sanction threats. Rational choice theory is based on the expected utility principle and includes not only formal sanction threats as costs of crime but also allows for the inclusion of informal sanction threats including shame and morality, and it takes benefits of crime into consideration during the calculation. Rational choice theory asserts that crime is more likely to occur when the anticipated benefits for the offender outweigh the anticipated costs for the offender. At its core, the theory assumes that an internal “choice” is made based on the decision-maker’s rational calculation of costs and benefits. However, rational choice theorists argue that individual rationality is “limited” and “less than perfect” due to several factors (Simon 1957). These factors include the following: (1) the decision-maker’s capacity to inaccurately perceive natural conditions, (2) the decision-maker’s capacity to imperfectly process information, and (3) the lack of necessary information provided to the decision-maker regarding the costs and benefits of the act. The net of these bounding factors is that the theory purports that individuals are self-motivated and are somewhat rational in their decision-making.

In accordance with this perspective, costs and benefits have both subjective and objective dimensions. For example, a subjective benefit of crime may be the thrill of getting away with it, whereas a subjective cost may be the recognized shame from family or friends. An objective benefit of crime may be monetary gain, whereas an objective cost may be a formal sanction. The evaluation of costs and benefits within subjective and objective dimensions becomes increasingly diverse when dealing with corporate crime decisions. For instance, in deciding whether or not to commit a corporate crime, one has to deliberate between expected costs and benefits for oneself as well as for the corporation in which they are employed. Examples of micro level costs of corporate crime may be formal punishment (civil, criminal, or regulatory) for the individual such as jail time or financial penalties; social criticism from friends, family, and colleagues; and/or a loss of self-respect. In terms of macro level costs, corporations may be subjected to regulatory, civil, and/or criminal prosecution; decrease in revenue; loss of ground to foreign competitors; and diminished firm prestige. Individual benefits of corporate crime may be career advancement or promotion, increase in personal income, and increase in self-image. Benefits to the corporation may include such things as an increase in firm increased profits and a gain of corporate prestige.

Braithwaite (1984, p. 6) defines corporate crime as the “conduct of a corporation, or of employees acting on behalf of a corporation, which is proscribed and punishable by law.” Therefore, according to this definition, corporate crime is committed to further corporate interests rather than individual interests. For example, managers that violate the law may do so as a rational pursuit for achieving organizational-level goals, maintaining the organization’s reputation, and/or stabilizing organization profits (Reed and Yeager 1996). It is presumed that organizational-level costs and benefits are also taken into consideration when the individual actor evaluates the costs and benefits of crime. In some cases, however, individual-level costs and benefits may align with those of the organization, thus forming a symbiotic relationship. Therefore, what is ultimately beneficial or costly to the corporation is beneficial or costly to the individual and vice versa. Corporate crime scholars suggest this to be the case for much of corporate offending.

The rational choice theory of crime also holds that the decision to commit crime is influenced by the context in which the decision is made. This facet of the theory is especially relevant to understanding corporate crime because the social and environmental context become salient factors when looking at criminal decision-making in an organizational setting. More specifically, within a corporate culture, competition, fear, and commitment to compliance may influence managerial decision-making and affect the likelihood of engaging in unethical behavior. For example, a corporate culture that emphasizes maintaining a competitive edge against foreign companies may set unreasonable goals for the firm. As a result, this pressure to attain hard to reach goals may generate a climate of fear among managers to create "innovative" solutions to achieve company objectives (Reed and Yeager 1996). Although these "innovative" solutions may or may not be legal, they may be regarded as acceptable depending on the organization's commitment to compliance. On the other hand, if an organization has recently been sanctioned or the company does not tolerate misconduct from its employees, criminal behavior may be less common (Paternoster and Simpson 1996). Corporate crime intentions may also be more likely when supervisors express a "fear of falling." That is, they are motivated by the fear of losing what they have worked so hard to attain (Wheeler 1992; Piquero 2012). This could occur when managers recognize the organization's financial health is in jeopardy, or if they perceive the organization may be losing ground to foreign competitors or otherwise declining in the marketplace. Quite simply, fear of falling may be a motivation to engage in corporate acts when any adverse situation may affect the organization's future well-being. Corporate managers choose to act on their own, may receive orders from top executives or supervisors, or may demand subordinates to engage in misconduct to maintain or advance firm-level interests.

In sum, corporate crime decision-making has often been viewed from a rational choice perspective. That is, managers make decisions about whether to engage in a crime based on an assessment of anticipated benefits and anticipated costs attached to a crime. Grounded in this framework are subjective and objective costs and benefits for both the individual and the organization in which they are employed in. Social and environmental factors within an organization become salient factors when looking at corporate crime decisions. The next section discusses empirical research that has linked the rational choice model of crime to corporate offending.

Future Directions

While some research on corporate crime decision-making exists, it pales in comparison to the vast body of literature that examines decision-making for common street offenders. Limited data and lack of funding for empirical research on corporate crime have made it challenging for scholars to understand of this specific decision-making process. Scholars who have examined it tend to rely on extant criminological theories and apply them, adapting them if necessary, to the corporate crime context. As previously noted, a great deal of this research has focused on the rational choice decision-making model of crime because of the characteristics associated with corporate crime. For example, Braithwaite and Geis (1982) have argued that corporate offenders are particularly sensitive to balancing costs and benefits of crime because they have a greater stake in conventionality and consequently have more to lose. While Kadish (1977) proposed that corporate crime is rational, instrumental, and directed at achieving economic gain, thus, corporate offenders are more likely to be deterred by costs linked to violating the law such as formal sanction certainty or severity.

Extant research that has applied a rational choice model to corporate crime has suggested that the specific environmental circumstances in a corporate setting affect decision-maker's evaluation of costs and benefits. A prime example of this is Paternoster and Simpson's (1993) *subjective utility theory of corporate offending*. As with other rational choice explanations, this model focuses on the individual actor as the decision-maker who is affected by both personal and organizational factors. These factors include perceived or subjective certainty and severity of formal and informal sanctions for the individual and the organization, organizational and environmental factors that are present during the decision-making process, internalized morals against the act, and perceived sense of fairness or legitimacy of the act. Research that has tested this model of corporate offending has found support for rational choice theory by suggesting that both individual and organizational factors affect criminal intentions (Paternoster and Simpson 1996; Simpson 2002; Simpson et al. 1998).

Individual moral evaluations have been shown to be an important source of inhibition not only for street offending but also for corporate crime (Elis and Simpson 1995; Simpson and Piquero 2002; Smith et al. 2007; Paternoster and Simpson 1996; Piquero 2012). Research evidence suggests that moral reservations condition how the critical decision-maker evaluates the costs and benefits of offending by prescribing some potential outcomes as morally unacceptable. For example, managers who perceive the illegal act in question conflicts with their moral code will not choose to offend because they believe the act is off limits. This means that, independent of the costs and benefits factored in, moral considerations condition the impact of instrumental concerns (Paternoster and Simpson 1996).

Additional research looking at unethical corporate decision-making has consistently found that fear of losing ground to foreign competitors increased the likelihood of corporate offending (Elis and Simpson 1995; Paternoster and Simpson 1996). A reason for this finding could be the inherent competeveness in the world of business. As corporations aim to be the top competitor in their industry, sometimes in order to achieve this goal laws must be broken, ignored, or altogether disregarded. This could be viewed as a macro level version of the fear of falling hypothesis. That is, from an organizational standpoint, the constant pressure to assure stakeholders of the corporation's stable financial and marketplace position may cause supervisors to view corporate crime as a necessary means to maintain or advance the organization's bottom line and public image.

Research evidence also indicates that managers are significantly more likely to offend when they are directly ordered to engage in misconduct by a supervisor (Elis and Simpson 1995; Smith et al. 2007). A possible explanation for this finding could be the manager's "appeal to a higher loyalty." As mentioned before, managerial decision-making is different from ordinary street crime decision-making because managers consider organizational interests when choosing whether or not to commit crime. As such, offending is more likely when managers are pressured to act in the best interest of the organization. This may very well be the case if the firm is not doing well financially and supervisors perceive that violating the law will gain substantial profits for the firm. Therefore, managers would be more likely to choose to commit corporate crime because their primary loyalty is to the company and the company's goals.

Moral climate within an organization has also been examined as to its effects on corporate crime decision-making. In general, these studies tend to find that the organizational culture affects criminal motivations. For instance, corporate crime is more likely to occur when it is regarded as a common occurrence in an organization (Elis and Simpson 1995; Paternoster and Simpson 1996). Managers are more likely to choose to engage in misconduct when they are surrounded by a corporate culture that encourages misconduct or embodies supervisors who turn a blind eye to how objectives are achieved. Therefore, if the company's ethos regards unethical behavior as an acceptable means to achieve goals, managers will tend to follow suit and approach goals with an "any means necessary" frame of mind. Vaughan (1996) has shown that over time, consistent illegal activity becomes commonplace within the corporate culture and consequently is viewed as acceptable. On the other hand, corporate offending is observed to be less likely when managers have access to a working compliance program or corporate misconduct is highly criticized by the organization in which they are employed in (Simpson and Piquero 2002).

While rational choice theory has commonly been applied to corporate crime decision-making, it is not the only model to have been tested. Other theoretical models that have begun to be examined include social learning, techniques of neutralization, low self-control, and desire for control. For example, Piquero et al. (2005b) examined both differential association theory and techniques of neutralization theory in order to understand corporate crime. According to social learning theory, individual behavior should be guided by peer reinforcement, while techniques of neutralization would suggest that neutralizations (or justifications) will be used during the decision-making process. Using a sample of Masters of Business students, respondents were presented with a vignette describing the sale and promotion of a hypothetical pharmaceutical drug that was about to be recalled off the market. They were asked to report the extent to which they would further or inhibit the distribution of the drug. They found, as expected, that the agreement of coworkers and the board of directors was positively associated with furthering the sale of the drug despite the indication that the drug was about to be recalled. Thus, work peers and bosses encouraged and supported the continuance of corporate profit. The results also showed that negative perceptions attached to corporate crime from close friends and business professors (peers outside the workplace environment) actually increase the likelihood of corporate offending. This is a finding not anticipated. Piquero and colleagues (2005b, p. 181) explain this finding as a result of "the extensive amounts of hours spent at the work and immersed in the corporate climate, as well as the fact that survival instincts are likely to kick in when the actions they take at work decides what type of life they will be providing for themselves and their families."

In recent years, the individual characteristics have been examined as explanations of criminal behavior. Gottfredson and Hirschi's (1990) theory of low self-control has received a great deal of attention. The core of the theory contends that individuals who evince low levels of self-control are more likely to engage in criminal and analogous behaviors. Research examining the influence of low self-control on offending propensities has found consistent support for this assumption among street offenders (Pratt and Cullen 2000). However, this line of research has not been as successful at explaining white-collar and corporate crime. Benson and Moore (1992) examined a sample of white-collar and street offenders and found differences across the two groups in terms of rates of offending with the street offenders offending at higher rates. Simpson and Piquero (2002) found that organizational characteristics were better predictors of corporate crime offending intentions than indicators of low self-control within a sample of business students.

Other scholars have adopted concepts from disciplines outside of criminology to help explain corporate criminality. Particularly, the notion of "desire for control," a concept found in psychological literature, has been recently integrated into the rational choice framework to help explain corporate crime. The desire for control is the wish to be in control of everyday life events. According to this model, individuals who exhibit high levels of desire for control are expected to be assertive in nature, occupy positions of power, and prefer to manipulate events to ensure desired outcomes (Burger and Cooper 1979). On the other hand, individuals who exhibit low desire for control are expected to be nonassertive, occupy subordinate-level positions, and may prefer others to make their daily decisions for them. Arguably then, individuals with high levels of desire for control in a business environment would be likely to be found in high-level positions such as a manager or an executive.

Research examining desire for control has found that individuals who evince high levels of desire for control were more likely to report intentions to commit corporate crime (Piquero et al. 2005a, 2010). This finding suggests that organizational actors who desire more control than others may be more likely to resort to criminal options rather than solve certain issues through legal means. More importantly, consistent with the expectations of desire for control, research evidence shows that high levels of desire for control affect several individual rational choice considerations such as the threat of informal and formal sanctions and individual morality and shame (Piquero et al. 2005a). Although studies applying desire for control to corporate crime decision-making are limited, this line of research offers much promise for further understanding the role of individual differences in corporate offending.

Related Entries

- ▶ Corporate Crimes and the Business Cycle
- ▶ Crimes of the Powerful
- ▶ White-Collar Crime

Recommended Reading and References

Benson ML, Kerley KR (2000) Life-course theory and white-collar crime. In: Pontell HN, Schichor D (eds) Contemporary issues in crime and criminal justice: essays in honor of Gilbert Geis. Prentice Hall, Upper Saddle River, pp 121–136

Benson ML, Moore E (1992) Are white collar and common offenders the same? An empirical and theoretical critique of a recently proposed general theory of crime. J Res Crime Delinq 29:251–272

Braithwaite J (1984) Corporate crime in the pharmaceutical industry. Routledge & Kegan Paul, London

Braithwaite J, Geis G (1982) On theory and action for corporate crime control. Crime Delinq 28: 292–314

Burger JM, Cooper HM (1979) The desirability of control. Motiv Emotion 3:381–393

Clinard MB, Meier RF (1979) Sociology of deviant behavior. Holt, Rinehart & Winston, New York

Cromwell P, Olsen JN, Avary DW (1991) Breaking and entering: an ethnographic analysis of burglary. Temple University Press, Philadelphia

Elis LA, Simpson SS (1995) Informal sanction threats and corporate crime: additive versus multiplicative models. J Res Crime Delinq 32:399–424

Forst B, Rhodes W (n.d.) Sentencing in eight United States district courts, 1973–1978. Codebook. (Interuniversity Consortium for Political and Social Research Study No. 8622). University of Michigan, Ann Arbor

Gottfredson MR, Hirschi T (1990) A general theory of crime. Standford University Press, Stanford

Kadish SH (1977) Some observations on the use of criminal sanctions in the enforcement of economic regulations. In: Geis G, Meier RF (eds) White-collar crime: offenses in business, politics and the professions. Free Press, New York

Paternoster R, Simpson SS (1993) A rational choice theory of corporate crime. In: Clarke RV, Felson M (eds) Advances in criminology theory: routine activity and rational choice, vol 5. Transaction, New Brunswick, pp 37–58

Paternoster R, Simpson SS (1996) Sanction threats and appeals to morality: testing a rational choice model of corporate crime. Law Soc Rev 30:549–583

Piquero NL (2012) The only thing we have to fear is fear itself: investigating the relationship between fear of falling and white collar crime. Crime Delinq 58:362–379

Piquero NL, Exum M, Simpson S (2005a) Integrating the desire for control and rational choice in a corporate crime context. Justice Quart 22:252–280

Piquero NL, Tibbetts SG, Blankenship MB (2005b) Examining the role of differential association and techniques of neutralization in explaining corporate crime. Deviant Behav 26(2):159–188

Piquero NL, Schoepfer A, Langton L (2010) Completely out of control or the desire to be in control? How low self-control and the desire for control relate to corporate offending. Crime Delinq 56(4):627–647

Pratt TC, Cullen FT (2000) The empirical status of Gottfredson and Hirschi’s general theory of crime: a meta-analysis. Criminology 38:931–964

Reed GE, Yeager PC (1996) Organizational offending and neo-classical criminology: challenging the reach of a general theory of crime. Criminology 34:357–382

Simon H (1957) Models of man. Wiley, New York

Simpson SS (2002) Corporate crime, law, and social-control. Cambridge University Press, New York

Simpson SS, Piquero NL (2002) Low self control, organizational theory, and corporate crime. Law Soc Rev 36:509–548

Simpson SS, Paternoster R, Piquero NL (1998) Exploring the micro–macro link in corporate crime research. In: Bamberger P, Sonnenstuhl WJ (eds) Research in the sociology of organizations: deviance in and of organizations. JAI Press, Greenwich

Smith NC, Simpson SS, Huang C (2007) Why managers fail to do the right thing: an empirical study of unethical and illegal conduct. Bus Ethics Quart 17(4):633–667

Sutherland EH (1945) Is white collar crime crime? Am Sociol Rev 10:132–139

Sutherland E (1983) White-collar crime: the uncut version. Yale University Press, New Haven

Vaughan D (1996) The challenger launch decision. University of Chicago Press, Chicago

Weisburd D, Wheeler S, Waring E, Bode N (1991) Crimes of the middle classes. Yale University Press, New Haven

Wheeler S (1992) The problem of white collar crime motivation. In: Schlegel K, Weisburd D (eds) White collar crime reconsidered. Northeastern University Press, Boston, pp 108–123

Corporate Crimes and the Business Cycle

Michael Levi
Centre for Crime, Law and Justice at Cardiff
School of Social Sciences, Cardiff University,
Cardiff, Wales, UK

Overview

This entry examines the (in fact only partial) Global Financial Crisis and its possible universal and differential direct impacts on frauds and on corporate crimes of various kinds and on control responses which, as routine activities perspectives would suggest, should have an impact upon financial and corporate crime levels. The entry examines the limited evidence about reporting processes and argues that some frauds whose commission long preceded the crisis will be brought into victim and/or public consciousness as a result of the credit squeeze; some "organized criminals" may be drawn into greater confidence in making fraud participation offers to insiders or blackmailing them because of the latter's inability to repay debts and because they believe that people are more corruptible at times of economic stress; some fraud opportunities linked to workplaces will be reduced because if people motivated to defraud have lost their jobs, they can no longer commit internal frauds, but in other cases, temptations are greater because of the desire not to lose lifestyle and social status. The net effect of changes in guardianship, motivations, and opportunities is difficult to determine, and most fraud data – other than plastic card fraud – are too dependent on changing probabilities of recognition, reporting, and recording to enable confident inferences about trends to be drawn. It seems plausible that more "slippery slope" insolvency frauds occur in times of recession, as some company directors and professionals seek to preserve income and wealth from the economic consequences of the downturn. On the other hand, a bonus-driven culture makes "strain" a perpetual condition irrespective of the business cycle. There is no evidence that the GFC has had or is likely to have a major impact on increasing the cost of fraud or levels of fraud overall in the areas about which we have the best data and knowledge: North America, the UK, and Australia.

Background

This entry reviews the relationship between economic cycles and a variety of white-collar and corporate crimes. This is not an area that has received very much criminological attention. An established criminological tradition examines the connection between economic crises and crime generally (Deflem 2011). In the early days, it was influenced by Marxist analysis of the endemic crises of capitalism, commencing with the work of Bonger (1916) and including some early empirical work by Radzinowicz (1939). Box (1987), anticipating later work on inequality and crime, argued that the significance of unemployment will vary depending on its duration, social assessments of blame, previous experience of steady employment, perception of future prospects, comparison with other groups, etc. Hence, there is likely to be a causal relationship between relative deprivation and crime, particularly in subcultures where unemployment is perceived as unjust and hopeless by comparison with the fortunes of other groups. Hale (1998) argued that increased opportunity or availability of targets had a long-run relationship with property crimes, while unemployment had a short-run impact on property crimes via its impact on motivating offenders. Arvanites and Defina (2006) showed that the strong economy of the 1990s reduced all four index property crimes and robbery by reducing criminal motivation. Business cycle growth produced no significant opportunity effect for any of the crimes studied. Cleck and Chiricos (2002) found that opportunity levels were generally unrelated to property crime rates and do not appear to mediate the unemployment-crime relationship. Of these authors, only Box looked at corporate crimes, arguing that recession motivated more corporate crimes for survival.

In a review of the evidence that was conducted in the wake of the previous Australian recession, Weatherburn (1992: 8) concluded that:

> [T]here would appear to be no reason to expect the criminogenic effects of sustained economic deprivation to be any quicker to dissipate than they were to arrive.... Uncontrolled and rapid economic development, chronically high levels of unemployment in certain sectors of the economy, and the existence of a marginalised under-class of individuals and families living in poverty are all conditions which can just as easily be found in booming as in contracting economies.... Nothing could be more inimical to law and order than an economy which generates rapidly rising living standards for some and an abundance of criminal opportunities for the remainder.

Turning to white-collar crimes – broadly construed – its relationship with economic cycles has been examined less directly, via unemployment or strain theory, using data from before the most recent severe economic downturn, which most closely resembles the 1930s (when both criminal justice action and the nature of the economy were markedly different). Langton and Piquero (2007) thoughtfully examined the relationship of general strain theory and white-collar crimes of different levels of complexity, finding that the relationship varied for different sorts of offenses. Schoepfer and Piquero (2006: 223–234) found that

> more unemployment was associated with less and not more embezzlement [measured by recorded crimes from the UCR]. Though not necessarily as expected by IAT [institutional anomie theory], this negative effect, however, would be inline for most forms of white-collar crime, including embezzlement, since employment provides the opportunity for offending. Hence, this effect makes sense within the context of this particular crime type (e.g. white-collar crime). . .. higher rates of polity weakened the effect of unemployment on embezzlement.

However, the growth and globalization of financial instruments and industrial production of this century may have altered the relationship, and further work needs to be done on understanding economic cycles and corporate crimes, bearing in mind the immense problems of data validity.

The Global Financial Crisis and Financial/Corporate Crimes

There are at least two ways of examining the relationship between corporate/financial crimes and economic cycles, most recently the Global Financial Crisis whose effects, though not whose fundamental causes, were first strongly felt in 2007–2008. One is to look at the extent to which financial crimes (e.g., in the subprime mortgage issuance market and/or in the bundling of these and other derivatives into financial instruments like Collateralized Debt Obligations – CDOs) influenced the financial crisis (Braithwaite 2010; Nguyen and Pontell 2010). Another is to review the impact that the crisis has had and might plausibly have on financial and corporate crimes. It is the latter, possibly less exciting, issue that will be explored in this entry. One aspect of this is the impact on public sentiments and government managerial reaction to corporate crimes generally and to subsets of financial crimes in particular (see, e.g., Levi 2009). There are many forms of corporate crime, which have different impacts and offender backgrounds, and about whose prevalence and incidence we have variable information. These are outlined below:

Types of Economic Crime

1. Harm government/taxpayer interests
2. Harm *all* corporate as well as social interests
   - That is, systemic risk frauds that undermine public confidence in the system as a whole, domestic and motor insurance frauds, maritime insurance frauds, payment card and other credit frauds, pyramid selling of money schemes, and high-yield investment frauds
3. Harm social and some corporate interests but benefit other "mainly legitimate" ones
   - Some cartels, transnational corruption (by companies with business interests in the country paying the bribe)
4. Harm corporate interests but benefit mostly illegitimate ones
   - Several forms of intellectual property theft – sometimes called "piracy" – especially those using higher quality digital media

There are additionally many corporate crimes such as environmental crimes (Gibbs et al. 2010) or “safety crimes” (Tombs and Whyte 2007) that involve occupational deaths and diseases but that are not easily connected empirically to the Global Financial Crisis, though they are plausibly connected to changes in the cycle of industrialization and rates of profitability in the Global North and in the globalized supply chain from the Global South to North (and, increasingly, within the Global South and the BRICS): as, for example, in the shift from industrial production to financial and other services. This also has implications for individual national statistics, since as industrial production has shifted to Asia, the industrial safety records of Asian countries – which often are not readily available or verifiable – can be expected to worsen and that of Western economies to improve, even if the inspection of factories and extractive industries were equivalent, which they are not. Thus it may look as if there has been no effect of the crisis on health and safety in North American and Western Europe, and indeed, it is not obvious what health and safety regulators in the Global North can do about industrial deaths in China, except to impose production chain corporate criminal liabilities.

A key aim of this entry is to examine what the evidence is that would enable us to judge what has happened and what is plausibly likely to happen to financial and corporate crimes in the context of what is often referred to as the Global Financial Crisis (GFC) – though as it transpired, it has not dramatically affected Australia, China, Singapore, or much of South America – whether as a result of that crisis or as a result of other factors that coincide with it. The Australian banking system was less hard-hit than elsewhere, to the extent that it briefly overtook the USA to take second spot after the UK as a financial services heavyweight, with the only net positive score among the leading countries in the Financial Development Index; by 2011, initial public offering (of securities) activity and insurance enabled Hong Kong SAR to overtake the United States and the United Kingdom to top the index (World Economic Forum 2012).

Research has reviewed general corporate crime data in the USA, including the effects of general and specific industrial prosperity. Using the original dataset of Clinard and Yeager from the mid-1970s, Wang and Holtfreter (2012) suggest that the effect of corporation-level strain is more pronounced for corporations marked by lower levels of diversification and corporations that operate in industries experiencing higher levels of strain. They found no evidence of corporation-level strain on violations; that corporations in rapidly growing, but not decaying, industries had higher violation rates; that the effect of corporation-level strain was less pronounced in corporations characterized by a higher level of diversification; that financially strained corporations had even higher violation rates in industries experiencing a decline in their financial performance; and that corporations experiencing growth in financial performance had higher violation rates in rapidly growing industries. Finally, they found no support that financially strained corporations had by far higher violation rates in industries characterized by more prevalent corporate illegality. Their work contains a solid review of general evidence on strain and corporate crime, though it is worth noting that the financial services sector is not singled out and had far less significance in the 1970s compared with this century.

The primary focus will be in areas where at least some reasonably valid data are available which, unfortunately, are mostly related to volume frauds in the private and public sectors rather than to management frauds and crimes *by* corporations, and care should be taken about generalization elsewhere. Thus, to the extent that measures against the financial crisis in general and fraud/corruption in particular have an impact on development aid and on opportunities for world trade, they might directly or indirectly stimulate or suppress fraud/grand corruption in developing countries: but data are not available to examine seriously that proposition. Other kinds of corporate crime data – for example, health and safety and/or environmental crimes – might be interesting to correlate with economic trends, but the research base is underdeveloped. It seems plausible that marginalized firms would be the most likely ones to commit not just frauds but also to cut corners on worker safety, consumer safety, and waste dumping. However, in a bonus-driven culture with predatory venture capitalist firms and short-term securities holdings, squeezed margins may be a sufficient condition, but they are by no means a necessary condition for corporate crime. The conventional wisdom attributes the rise of fraud to the impact of globalization, but "globalization" is insufficiently linear or specific to be useful as an explanatory variable. Black (2010) is critical of white-collar criminology's shift since Sutherland from elite corporate misconduct and what he terms "control frauds" to more blue-collar embezzlement and frauds since Cressey's time, and though it is arguable that criminology should cover both these areas, it is indeed important to remember what *forms* of business crimes one is examining, however they are labeled.

Estimating Financial Crime Costs

Despite all the victimization surveys that have been carried out in Australia, Europe, and the USA, and to a far lesser extent elsewhere, the collection of information on impacts (beyond reporting and "confidence in the police") has been quite crude and simplistic. Moreover, harm (and threat) is not just about how much things cost us. "Organized fraud," for example, intermingles actual impacts with the social construction of malevolence and "dangerous groups." But what if we do not know whether or not particular fraudsters are connected with "organized crime groups" or are financing terrorism? Should we take the agnostic position that they are not connected and therefore are less harmful than they would have been? And why should having such connections be worse than the frauds of schemers like Bernie Madoff who stole roughly $15 billion from a variety of charities and relatively wealthy individuals and firms over a 15-year period, despite numerous attempts to alert the authorities?

Zemiology – the study of social harms – has been treated as a useful tool with which to batter the obsessions of conventional criminology with household and street crimes: the aim of some of its advocates is to transform our priorities so that we might do more about road and industrial accidents than about terrorism and perhaps more about corporate food poisoning than about what is normally thought of as criminal poisoning. However, Sparrow (2008) emphasizes that it is absurd to neglect the relevance of perceptions of future dangerous intentions (as in "Islamic violence" post 9/11) when assessing harm, for this is part of their cultural impact and also ought to trigger our strategic interventions. It is arguable that we should not ignore planned (and foreseeable unplanned) harm doing in constructing risk assessments and crime control methodologies, however problematic politically it may be to incorporate such crime risk elements into general social planning (Dorn and Levi 2006; Edwards and Levi 2008). But irrespective of the current operational use of harm data, the latter (and crime seriousness studies) remain sociologically interesting and socially important.

Many estimates of fraud and money laundering are based on very limited evidence, derive from institutional profile raising, and take on a life of their own as "facts by repetition," seldom critiqued either by media always hungry for sensational headlines (Levi 2006) or by pressure groups. On the other hand, if we rely on cases brought to conviction, the data may be absurdly low in both harm caused and case complexity, the most subtle cases of fraud and money laundering being hardest to prove beyond a reasonable doubt in the minds of juries or judges. We must be careful to avoid "throwing the baby out with the bathwater" by using only very low figures from convictions and treating them as "the figures" rather than just as validated minimum figures (see Levi and Burrows 2008).

The elapsed time from fraud to discovery and then from discovery to criminal justice action (if any) means that many of the larger frauds coming to light in 2008–2013 will have been committed some years before. One classic notable example is the Madoff's long-running Ponzi scheme (van de Bunt 2010) though this was atypically long for this type of investment fraud.

Although such schemes all come to an end eventually when incoming funds are insufficient to meet current payments, the link with the economy is that drops in confidence and increases in need for funds derail the criminal business model. In some other cases where fraud has been alleged, only careful case-by-case examination can enable us to understand when the fraud began in relation to the business cycle: but one might expect the frauds to result from a combination of ambitious expansion plans with unexpected reductions in business opportunities and/or cash flow. This time attrition for frauds differs from "volume crimes" such as credit and debit card fraud, which usually come to light promptly (though seldom as promptly as burglaries and vehicle thefts).

Normally, we would look to statistical data on recorded/surveyed crime and/or cost of crime trends to enable us to judge whether a problem is getting better or worse. However, as will be seen, even in the best countries, fraud data are quite poor, and this is an area that requires some redress if empirical criminology is not to continue to neglect white-collar crimes (see, more generally, Simpson and Weisburd 2009). Despite activities in train to improve government fraud statistics in the UK (Levi and Burrows 2008; NFA 2012) and in Australia, but not yet in the USA, these recent improvements cannot be applied retrospectively to past data. Payment card fraud data have the advantage that there is little time lag from crime to discovery, but in higher value management frauds and cartels – which arguably are far more serious – there typically is a much longer time before detection and even longer for criminal justice outcomes, if any: so any given year's data will be a mix of crimes occurring at very variable years. This is a particular problem for evaluating the impact of economic trends, and there has been no opportunity here for the individual case analysis that would be a desirable alternative to aggregate data. Some safety crimes have immediate impacts, but environmental and industrial diseases may take years to emerge.

Australian fraud data are only intermittently available (Levi and Smith 2011; ABS 2012) and therefore cannot be "trended" without unacceptable speculative leaps. In the USA, very little effort has been made to generate better fraud data outside the populist consumer area of "identity theft" (Langton 2011), whose direct costs to the 5 % of the adult US population victimized over a 2-year period were estimated at $17 billion in 2008 (Langton and Planty 2010). The UK National Fraud Authority is seeking to make costs of fraud data annual, and the current estimate of the costs of fraud is £73 (US$118) billion (Levi et al. 2007; Levi and Burrows 2008; NFA 2012). Currently, however, reliable time series cost data are available in the UK only for payment card and check frauds, some aspects of identity frauds, and for external frauds against government departments and local authorities. No similar efforts have been conducted in other countries around the world, and fraud cost and incidence surveys carried out by Ernst & Young, KPMG, Kroll, and PwC in many parts of the world are both intermittent and have been applied only to large corporate victims, who are their primary clients and target markets for awareness raising (Levi et al. 2007; Levi and Burrows 2008). Indirect costs have not been examined seriously in any of these studies. Finally, there are the usual problems of whether it is possible to place within the category of fraud acts and actors that have not yet been subject to any criminal justice verdict: the "no-fault" settlement in 2010 of the action for civil fraud against Goldman Sachs by the Securities and Exchange Commission being a case in point (Dorn and Levi 2011). Such elite cases pose questions about the limitations of the relationship between crime and social harm and, if included as "fraud," would have a dramatic effect on the composition of the dependent variable. If we take the effort to collect data as an indicator of what information the state considers important, as Jeremy Bentham intended criminal statistics to be in the nineteenth century, one inference is that fraud data are plainly considered unimportant by most governments most of the time, both to inform their resource allocation decisions and for public information about levels of "crime." (Some skeptics may believe that this is a conspiracy to keep white-collar crimes out of the public eye, but there is no evidence for this proposition).

Identity Theft and Fraud

Identity frauds and their explanation are reviewed in detail elsewhere (Vieraitis et al. 2012). Here, we focus simply on trends. The Personal Fraud Survey for 2010–2011 (ABS 2012) found that Australians lost $1.4 billion due to personal fraud (which includes credit card fraud, identity theft, and scams). The survey results estimated that three in five victims of personal fraud (713,600 persons) lost money, an average of $2,000 per victim. Around 6.7 % of the population aged 15 years and over were a victim of at least one incident of personal fraud in the 12 months prior to interview. This is an increase from 5 % in 2007. The survey results show:

- 3.7 % (662,300) of Australians were victims of credit card fraud, an increase from 2.4 % in 2007.
- 0.3 % (44,700) of Australians were victims of identity theft, a decrease from 0.8 % in 2007.
- 2.9 % (514,500) of Australians were victims of scams, an increase from 2.0 % in 2007.

These trends are demonstrated elsewhere (see http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/4500.0Chapter25July+2011+to+June+2012).

However, it seems unlikely that these trends are the result of the financial crisis, not least because this has not occurred in Australia, due to the more limited involvement of banks in the derivatives market and to the boom in extractive industries sales to Asia.

In the USA, trends in identity fraud may be measured by victimization survey data, which show no obvious, consistent link with the business cycle. Below are the trends (Langton 2011: 2) (Fig. 1):

In the UK, data collected by the not-for-profit fraud prevention service CIFAS indicated a rise in identity takeover frauds, as “new credit” became harder to get, generating displacement to impersonating existing account holders. Such frauds emerge quite quickly. In 2009–2010, applications containing lies or supplying false supporting documentation fell 23 %; the use of false identity details and cases where an innocent victim has had their identity details used fraudulently rose 10 % (CIFAS 2010). Since then, the frauds recorded on a collective industry database rose by a quarter in 2009 and remained relatively unchanged thereafter: they were 77,500 in 2007 and 113,250 in 2011 (CIFAS 2012). No equivalent data are available for any other countries, because no equivalent organizations exist there to integrate credit application data.

Payment Card and Allied Credit Frauds

Since 2006, UK payment card fraud data have displayed a broad downward trend in fraud on lost or stolen cards, due to the introduction of Chip and PIN in Europe, a more recent drop in counterfeit frauds on skimmed and cloned cards, which previously had risen substantially, mainly being used overseas to sidestep Chip and PIN controls. Total fraud losses on UK cards fell by 7 % between 2010 and 2011 to £341 million. This is the lowest annual total since 2000 and follows on from a fall of 17 % in 2009–2010. While card usage and transaction volumes continue to grow, card fraud losses against total turnover – at 0.06 % – continue to decrease and fell by 33 % in 2009–2011 (Financial Fraud Action UK 2012). More generally, payment card fraud in Europe increased after 2006, peaking in 2008. The levels in 2011 are €121 million more than in 2006. A notable exception is the UK, which accounted for 45 % of the total in 2006 and now accounts for 29 %, a reduction of €177 million. It appears that criminals have shifted their focus to new opportunities within Europe, as antifraud measures in the UK became stricter (FICO 2012).

In financial year 2006, card and check fraud in Australia totaled A$142.6 m.; by 2011, it had almost doubled to A$251.5 m. As a percentage of the total value of transactions over the same period, it doubled to 0.0132 % in 2011 (http://www.apca.com.au/payment-statistics/fraud-statistics/archive-releases/2011-financial-year). None of these Australian or UK changes have any obvious relationship to the GFC but are more readily explained by prevention actions and by the introduction of Chip and PIN in Australia (Levi and Smith 2011).

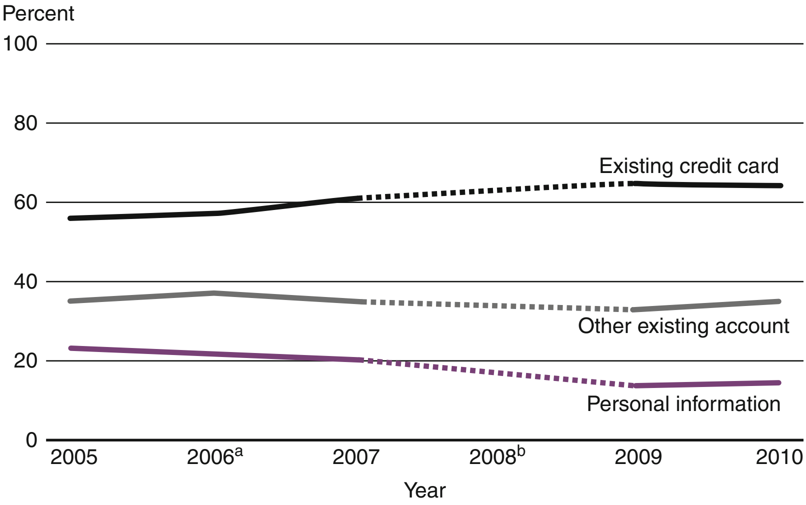

Corporate Crimes and the Business Cycle, Fig. 1 Trends in household identity theft victimizations, USA

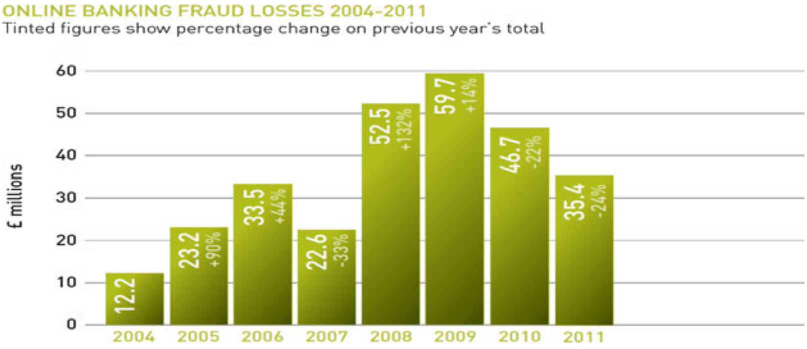

Corporate Crimes and the Business Cycle, Fig. 2 Online banking fraud losses 2004–2011

Canadian data for debit card fraud show that debit card frauds rose since the recession, before falling significantly in 2011 to where they were in 2005 – C$70 million – though with a lower average cost and much higher debit card use (http://www.interac.ca/media/stats.php, accessed 21 September 2012).

There are trend data in the report of the US Federal Trade Commission (FTC 2012: 3), showing a continuing rise in complaints since 2007 (and before), though identity thefts – which do not always result in frauds – have varied somewhat. For the 12th year in a row, identity theft complaints topped the list of consumer complaints in the USA. Of complaints filed in 2011, 279,156 (15 %) were identity theft complaints, compared with 86,250 (though a higher 27 % of complaints) in 2001.

The fact that the upward trend long preceded the financial crisis shows the danger of examining only data after the date of the crash. They suggest a rising secular trend, though not self-evidently any particular reason. Rather they are consistent with a dynamic model of routine activities theory within which both criminal actors and commercial crime preventers seek to maximize gains within their capacity and minimize losses, respectively. This may be seen in banks' online UK fraud losses (Financial Fraud Action UK 2012), which again show no particular correlation with the business cycle (Fig. 2).

Mortgage Frauds

The FBI (2012) notes that "during 2011, mortgage origination loans were at their lowest levels since 2001, partially due to tighter underwriting standards, while foreclosures and delinquencies have skyrocketed over the past few years.

Corporate Crimes and the Business Cycle, Fig. 3 CoreLogic – estimated fraudulent loans by origination year, 2006–2010

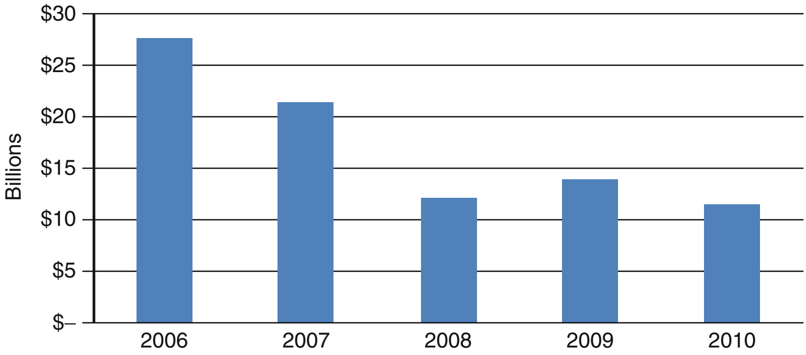

So, distressed homeowner fraud has replaced loan origination fraud as the number one mortgage fraud threat in many FBI offices. Other schemes include illegal property flipping, equity skimming, loan modification schemes, and builder bailout/condo conversion. During FY 2011, we had 2,691 pending mortgage fraud cases." The mortgage fraud report (FBI 2011) provides the following chart, neatly illustrating the decline of mortgage originations alongside the recession that they helped to generate (Fig. 3):

There is a lagged effect of the fraud identification process, and the frauds began typically before the financial crisis, though the general report usefully outlines the precipitating factors behind the mortgage fraud crisis and suggests that the rate of deception accelerated as brokers sought to retain their commission bonuses in a falling market. (Though the latter is plausible, the data are not available to demonstrate it). Many of the situational opportunities to defraud are fairly widespread over time, so variation is likely to be accounted for in terms of the spread of techniques among the wider population, expectations of the corruptibility of others, and changes in motivation, including the fear of downward economic mobility.

Control Frauds

"Control frauds" are seemingly legitimate entities controlled by persons who use them as a tool for fraud. Black (2010) argues that fraudulent lenders produce guaranteed, exceptional short-term "profits" through a four-part strategy: extreme growth (Ponzi-like), lending to uncreditworthy borrowers, extreme leverage, and minimal loss reserves. These false profits defeat regulatory controls and allow the CEO to convert the assets of the firm to his personal benefit through seemingly normal compensation mechanisms. The short-term profits also cause the CEO's stock options holdings to appreciate. Each element of the strategy dramatically increases the eventual loss. The record "profits" allow the fraud to grow rapidly for years by making bad loans. The "profits" allow the managers to loot the firm through exceptional compensation, increasing losses.

Another example of much longer elapsed time from fraudulent to business-detected awareness is "rogue trading" in the form of illicit (and – perhaps more importantly – unsuccessfully!) betting on market trends, for example, by Nick Leeson at Barings in 1995 (http://www.nickleeson.com) and by Jérôme Kerviel at Société Générale in 2007–2008. (In 2010, Kerviel was imprisoned for his deceptions, despite his attempts to blame it on the bank management). In both cases, the primary accused alleged that more senior managers turned a blind eye to the risk taking, because they were incompetent (Barings) or because they stood to gain if the risks paid off (both Barings and Société Générale). The popular explanation for this (and other excrescences of the financial sector) is that it is caused by greed. However, unless it is a tautology, greed is not a helpful discriminator between those who both see and take fraud opportunities and those who do not. Some might argue that offenders and non-offenders are equally greedy, but that the latter are more risk averse by temperament/social training/testosterone levels (Coates 2012). Nor is it obvious that "the recession" – local or global – had any impact at all on these sorts of alleged frauds, except indirectly, inasmuch as credit and controls tend to be tighter in recessionary times. In essence, they were cases in which ill-supervised traders found ways of breaching their dealing limits without detection and, having lost significant sums in early trading, carried on trading in the hope that they would be able to recoup those losses. Whether others behaved similarly but were more fortunate in their later dealing (or gambling) and remained forever undetected is unknown.

The Global Financial Crisis (GFC) and Its Economic and Social Impact

Australia was less affected by the GFC than were many other Western OECD countries such as the UK and the USA. Before looking at the implications for fraud, it might be helpful to see what the general financial crime-relevant effects might be expected to be. Institute for Fiscal Studies research (Muriel and Sibieta 2009) shows that low-skilled "elementary" occupations have suffered most since 2008 in the UK, followed by skilled trades and sales. In contrast, managers and senior officials have seen their unemployment increase by only 1 %, and "white-collar" professional unemployment has increased by just 0.7 %, about one sixth of the increase in the other groups. This does not mean that the "white-collar workers" are not fearful: but the impact of these fears on their propensity to defraud or launder money remains unresolved. The rise in unemployment among those with housing commitments and "negative equity" in their mortgaged homes constitutes a serious economic trap and could lead to increased personal bankruptcies and/or crime to support living standards. However, social security fraud and payment card fraud are the two types most readily open to such persons as criminal opportunities without significant skills barriers.

A demonstration that increased unemployment – whether generally or in a particular sector from which fraudsters came – always tended to be followed by increased crime rates would provide more compelling evidence of a causal relationship than would mere cross-sectional data for a particular point in time. However, it would have to overcome unreported and unrecorded fraud data problems, some but not all of which are being mitigated in the UK by the Action Fraud Reporting Centre, which is based partly on the USA and Canadian NWC3/RECOL reporting centers, though these latter are more web than telephone based.

The GFC and Models of Fraud Explanation

There has never been a simple link between recession, poverty, and crime. Barbara Wootton (1959) observed that if there were, then widows and elderly people would have the highest offending rates. Becker's valuable contribution to the economic explanation of crime has been eroded over time by the understanding from behavioral economists that his individualistic focus on crime and rational choice is mistaken, since our beliefs about what other people are doing and will find socially acceptable significantly mediate our behavior and even our desires. The reaction of people to often dramatic changes in their financial circumstances and their social prestige (and in their *expectations* of unbroken progress toward "the good [consumer] life") is a good test of the adequacy of criminological theories: in the particular case of white-collar crimes (see Simpson and Weisburd 2009), the anxieties among well-paid (and often highly wealth-oriented) professionals about the personal and corporate impact of the GFC gives plausible grounds for concern about what extra crime risks will be generated. These might take the form of greater willingness to trade off the risk of money-laundering prosecution against doing extra business. There is also the possibility that lower paid white-collar workers – for example, in call centers and banks – may be tempted to engage in fraud and money laundering on behalf of "organized crime" groups. These phenomena are not new: what is being proposed is that there may be a significant escalation in take-up of such opportunities.

Clarity is important here. To the extent that crimes are *occupational*, one must have an occupation in order to commit them. To illustrate this, we might examine the extent to which fraudulent chief executives such as Lord Conrad Black, Alan Bond, the chiefs of Enron, and "dot.com" bubble chiefs were able to allocate to the company expenditures that in fact were largely or wholly personal. If others have confidence in them, such entrepreneurs can develop new businesses that may generate new manipulative possibilities, but this would usually take longer at times of recession. At a lower level, staff in call centers (whether physically located in Western jurisdictions or in English-speaking South Asian countries) cannot so easily copy and extract personal data of account holders if they are no longer employed in the call centers. If still employed, they may be more tempted to defraud if they consider that they may shortly become unemployed and/or that the company will show no loyalty toward them. (Though their *ability to offend* may be reduced by physical opportunity controls such as the absence of USB and CD drives on computers and rapid integrity checks). Under such circumstances, *voluntary* compliance via procedural legitimacy (Tyler 2009) becomes much harder to achieve.

Knowledge of criminal techniques should be differentiated from *opportunity and motivation to offend*, though offending may be inhibited if people do not believe that they have enough knowledge to commit any particular sort of fraud that they contemplate. *It is the expectation or fear that one may become unemployed*, *not post-employment status*, *that presents the extra risk here*. There may be some crimes – computer hacking, for example – where employment status may be unimportant, though the importance of social engineering in getting passwords or other elements of crime may suggest that having a job in which one *knows* these current gateways is useful. (Past gateways may still be current if controls are not changed immediately when people are "let go"). Impacts of the GFC on professionals elsewhere may impact any country because, given international finance and trade – both for business and individuals – they may be the victims of frauds originating elsewhere, and expatriate professionals who lose their jobs may return home and exercise their skills illicitly. The greater time available to the unemployed in any country can give greater room for experimentation, and criminal attempts seldom are punished or even logged by targets or financial intermediaries.

The "crime triangle" of motivation, opportunity, and capable guardians can be applied to fraud, and it is here in human and technological changes to these parameters that one may search for an explanatory framework. A priori, it would appear that different skill sets, commercial/personal background, and formal qualifications will be needed for different fraud offenses, and the barriers to entry depend on the starting point of any given individual or network *in relation to the practical opportunity and criminal justice obstacles that confront them*. The longer and more intensive an investigation, the more likely it is that surveillance will generate an accurate model of interactions between players in the network. In most countries of the world, a distinction is made between "laissez faire" opportunities to set up and work in commerce and some restrictions applied to people who want to open or work in the financial services sector, largely on the grounds that the latter can directly steal funds from the public. It is important to understand such restrictions in a global context rather than the traditional nation-state perspective of regulation and criminal justice.

What some offenders are able to do is simply to deploy the range of global corporate mechanisms available in a free enterprise society where there are insufficient "capable guardians" to stop them misusing the disguises offered by the corporate form or the authority and power of a corporate role. Barry (2001) highlights the role played by Bond's ingenuity and the (skillfully manipulated) delays of law in enabling him to exploit the rule whereby gifts "not for value" could be set aside only if they preceded bankruptcy by 2 years. The lawyers who enabled that delay were paid by generous international benefactor friends and by a Swiss who was alleged to have acted as a nominee for Bond, using a variety of foreign trust companies. Enron offered a dazzling array of special purpose vehicles to hide transactions: this was long before the GFC, and it is moot whether the impact on fraud of the crisis was affected by the Sarbanes-Oxley and other supervision measures introduced in the aftermath (see Tillman 2009, for some insights).

Levels of Indebtedness and Regulation

Corporate indebtedness is a major trigger for the revelation of fraud, as it becomes impossible to hide large-scale fraud (or at least to hide great loss, since it may be possible to misrepresent fraud as legitimate commercial failure among all the other corporate collapses, especially if creditors can be strung out long enough to risk "throwing good money after bad"). At the individual level, levels of indebtedness and illiquidity are major components that make some aspects of the current GFC different, and the reduction in lending capacity has affected the depth of the economic crisis. One would not want to overstress this – the great scammers of the 1980s were highly leveraged on their corporate assets and gave personal guarantees that turned out to be worthless (Barry 2001). However, not only are wealthy entrepreneurs of the current era (e.g., Russian "oligarchs") highly mortgaged on their assets and therefore more pressurized to lie in order to stay afloat long enough to avoid a forced sale, but also a broad range of people in many walks of life are hugely indebted compared with the 1980s. The long period of prosperity is one reason for this, alongside the rising housing market that tempted many to borrow for current optional expenditure beyond their capacity to repay. Business, consumers, and regulators fell victim to the convenient myth that the "boom and bust" cycle had been abolished.

The rise in visible mortgage frauds (Carswell and Bachtel 2009; Nguyen and Pontell 2010) and in consumer/investment scams has energized the regulatory process, assisted by forensic linking software developments which make it easier proactively to search out connections between banking and insurance fraud networks. Since Ponzi investment pyramids rely on a high rate of incoming investments to sustain payouts, a fall in the rate of *increase* of investments or a *reduction* in the rate of reinvestment of imaginary profits causes them to collapse earlier.

Conclusions

The impact of the Global Financial Crisis upon fraud depends upon what one includes within the latter category. The damning verdict of the Valukas (2010) report into the factors leading up to the collapse of Lehman Brothers and the civil charges of fraud against Goldman Sachs by the Securities and Exchange Commission (treated as criminal by the media) make the fluctuating boundaries of fraud even more apparent (Dorn and Levi 2011). Frauds and white-collar crimes need to be broken down in conformity with the purposes for which one wants to use the typology, especially if a link to the business cycle is to be established.

In their powerful analysis of "animal spirits," Akerlof and Shiller (2010) identify confidence, fairness, corruption and bad faith, money illusion, and stories as the key drivers of economic change, both positive and negative. As they acknowledge, their work is of a high level of generality, and to turn them into operationalizable models capable of correlating with financial crimes of various kinds is too major a task for this study (and plausibly for my lifetime). What seems plain is that generalizations about the correlation between economic crises and fraud (or other types of financial crime) are far too glib to be sustainable in the round. More limited evidence exists of some subsets of fraud, but even here, the recession serves to mask the complexity of trends, and once the lagged awareness, reporting, and investigation of larger frauds are taken into account, the relationship is shown to be less strong than it initially appeared. A working generalization is that the most serious financial frauds involving credit are "crimes of prosperity" at which time controls are relatively weak and borrowing is relatively easy. The fear of falling also acts as a stimulant to some of those who have the opportunity to defraud (Piquero 2012). In times of financial decline, when controls both on fraud and on the availability of fresh credit tighten, these long-term frauds are exposed, and this creates the appearance of their being "crimes of adversity." Smaller scale volume frauds may be stimulated by factors relatively independent of the business cycle, but again the business cycle may provoke changes in the control environment which, alongside controller perceptions of the rate of change in risk, may lead to reductions in opportunity. It may be that biosocial elements influence animal spirits in the hour between dog and wolf (Coates 2012). But the relative lack of watchdogs and the absence of barking in the night or of prosecutions in the days since the financial crisis began should reasonably baffle those who see a great deal of routine prosecution of relatively trivial crimes elsewhere in the social system and who wonder why electronic surveillance and rapid response repression are only justified when elites are not under suspicion.

Related Entries

- ▶ Anomie and Crime
- ▶ Compliance and Corporate Crime Control
- ▶ Corporate Crime Decision-Making
- ▶ Crimes of Globalization
- ▶ Deterrence of Tax Evasion
- ▶ Fraud Victimization
- ▶ Identity Theft
- ▶ Routine Activities Approach

Recommended Reading and References

ABI (2009) General insurance claims fraud. Association of British Insurers, London, http://www.abi.org.uk/Media/Releases/2009/07/40569.pdf

ABS (2012) Personal fraud. 2010–2011. Australian Bureau of Statistics, Canberra, http://www.abs.gov.au/ausstats/abs@.nsf/mf/4528.0/

Akerlof G, Shiller R (2010) Animal spirits: how human psychology drives the economy, and why it matters for global capitalism. Princeton University Press, Princeton

Arvanites TM, Defina RH (2006) Business cycles and street crime. Criminology 44(1):139–164

Barry P (2001) Going for broke. Bantam, Melbourne

Barbara Wootton (1959) Social Science and Social Pathology (London: Allen and Unwin)

Black W (2010) Epidemics of 'Control Fraud' lead to recurrent, intensifying bubbles and crises. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1590447

Bonger W (1916) Criminality and economic conditions. Little, Brown, New York

Box S (1987) Recession, crime and punishment. Palgrave Macmillan, London

Braithwaite J (2010) Diagnostics of white-collar crime prevention. Criminol Public Policy 9(3):621–626

Carswell AT, Bachtel DC (2009) Mortgage fraud: a risk factor analysis of affected communities. Crime Law Soc Change 52:347–364

CIFAS (2010) Figures reveal the bleak and complex reality of fraud in the UK today. http://www.cifas.org.uk/default.asp?edit_id=1048-57. Accessed 27 Nov 2010

CIFAS (2012) Is identity fraud serious? http://www.cifas.org.uk/is_identity_fraud_serious

Cleck G, Chiricos T (2002) Unemployment and property crime: a target-specific assessment of opportunity and motivation as mediating factors. Criminology 40(3):649–680

Coates J (2012) The hour between dog and wolf: risk-taking, gut feelings and the biology of boom and bust. Fourth Estate, London

Deflem M (2011) Introduction: criminological perspectives of the crisis. In: Deflem M (ed) Economic crisis and crime. Emerald Group, Bingley

Dorn N, Levi M (2006) From Delphi to Brussels, the policy context: prophecy, crime proofing, impact assessment. Eur J Crim Policy Res 12:213–220

Dorn N, Levi M (2011) Theorising the security and exchange commission's 2010 settlement with Goldman Sachs: legislative weakness, reintegrative shaming or "temporary business interruption" for financial market power elites? In: Antonopoulos G, Groenhuijsen M, Harvey J, Kooijmans T, Maljevic A, von Lampe K (eds) Usual and unusual organising criminals in Europe and beyond. Profitable crimes: from underworld to upper world. Liber amicorum Petrus C. van Duyne. Maklu Uitgevers nv, Antwerp

Edwards A, Levi M (2008) Researching the organisation of serious crimes. Criminol Crim Justice 8(4):363–388

FBI (2011) 2010 Mortgage fraud report: year in review. http://www.fbi.gov/stats-services/publications/mortgage-fraud-2010/mortgage-fraud-report-2010

FBI (2012) Financial crimes report to the public: fiscal years 2010–2011. http://www.fbi.gov/stats-services/publications/financial-crimes-report-2010-2011

FICO (2012) Evolution of card fraud in Europe. www.fico.com/fraudeurope

Financial Fraud Action UK (2012) Fraud – the facts 2012. Financial Fraud Action UK, London

FTC (2012) Consumer sentinel network data book January–December 2011. Federal Trade Commission, Washington, DC

Gibbs C, McGarrell E, Axelrod M (2010) Transnational white-collar crime and risk: lessons from the global trade in electronic waste. Criminol Public Policy 9(3):543–560

Hale C (1998) Crime and the business cycle in post-war Britain revisited. Br J Criminol 38(4):681–698

Langton L (2011) Identity theft reported by households: 2005–2010. Government Printing House, Washington, DC, http://bjs.ojp.usdoj.gov/content/pub/pdf/itrh0510.pdf

Langton L, Piquero NL (2007) Can general strain theory explain white-collar crime? A preliminary investigation of the relationship between strain and select white-collar offenses. J Crim Justice 35(1):1–15

Langton L, Planty M (2010) Victims of identity theft: 2008 (NCJ 231680). http://bjs.ojp.usdoj.gov/index.cfm?ty=pbdetail&iid=2222

Levi M (2006) The media construction of financial white-collar crimes. Br J Criminol 46:1037–1057

Levi M (2008) The phantom capitalists: the organisation and control of long-firm fraud, 2nd edn. Ashgate, Andover

Levi M (2009) Suite revenge? The shaping of folk devils and moral panics about white-collar crimes. Br J Criminol 49(1):48–67

Levi M, Burrows J (2008) Measuring the impact of fraud, a conceptual and empirical journey. Br J Criminol 48(3):293–318

Levi M, Smith R (2011) Fraud vulnerabilities and the global financial crisis. Trends & issues in crime and criminal justice no. 422. Australian Institute of Criminology, Canberra

Levi M, Burrows J, Fleming M, Hopkins M (with the assistance of Matthews K) (2007) The nature. Extent and economic impact of fraud in the UK. Association of Chief Police Officers, London. http://www.cardiff.ac.uk/socsi/resources/ACPO%20final%20nature%20extent%20and%20economic%20impact%20of%20fraud.pdf

Muriel A, Sibieta L (2009) Living standards during previous recessions. IFS Briefing Notes No. 85. http://www.ifs.org.uk/publications/4525

NFA (2012) Annual fraud indicator. National Fraud Authority, London

Nguyen TH, Pontell HN (2010) Mortgage origination fraud and the global economic crisis: a criminological analysis. Criminol Public Policy 9(3):591–612

Piquero NL (2012) The only thing we have to fear is fear itself: investigating the relationship between fear of falling and white-collar crime. Crime Delinq 58(3):362–379

Radzinowicz L (1939) A note on methods of establishing the connexion between economic conditions and crime. Sociol Rev 31(3):260–280

Schoepfer A, Piquero NL (2006) Exploring white-collar crime and the American dream: a partial test of institutional anomie theory. J Crim Justice 34(3):227–235

Simpson S, Weisburd D (eds) (2009) The criminology of white-collar crime. Springer, New York

Sparrow M (2008) The character of harms. Harvard University Press, Cambridge

Tillman R (2009) Reputations and corporate malfeasance, collusive networks in financial statement fraud. Crime Law Soc Change 51:365–382

Tombs S, Whyte D (2007) Safety crimes. Willan, Cullompton

Tyler T (2009) Self-regulatory approaches to white-collar crime: the importance of legitimacy and procedural justice. In: Simpson, Weisburd (eds) The criminology of white-collar crime. Springer, New York

Valukas A (2010) Lehman Brothers Holdings Inc. Chapter 11 proceedings examiner's report. http://lehmanreport.jenner.com/

Van de Bunt H (2010) Walls of secrecy and silence: the Madoff case and cartels in the construction industry. Criminol Public Policy 9(3):435–453

Vieraitis L, Copes H, Birch I (2012) Identity theft (this volume)

Wang X, Holtfreter K (2012) The effects of corporation- and industry-level strain and opportunity on corporate crime. J Res Crime Delinq 49(2):151–185

Weatherburn D (1992) Economic adversity and crime. Trends and issues in criminal justice no. 40. Australian Institute of Criminology, Canberra

World Economic Forum (2012) The financial development report 2012. World Economic Forum, Davos

Corporate Criminal Liability

► Corporate Liability

Corporate Liability

Kathleen F. Brickey
Washington University School of Law,
Washington University, St. Louis, MO, USA

Synonyms

Corporate crime; Corporate criminal liability; Organizational liability

Overview

The principle that corporations can be held criminally responsible for crimes committed on their behalf has been well established in Anglo-American law for more than a century. After exploring the historical and theoretical underpinnings of this phenomenon together with the deterrent and retributive rationales for imposing criminal liability on corporations, this entry examines Justice Department policies regarding the prosecution and punishment of corporate crime and how those policies are implemented in practice.

Fundamentals of Corporate Criminal Liability

What is corporate crime? Although this question seems a logical place to begin, perhaps it assumes too much. It might be more apt to ask in the first instance whether it is possible for corporations to commit crimes. Corporations are, after all, creatures of the state. They are legal fictions that are created through corporate charters issued and approved by government officials. A corporation's charter defines its powers and the conditions under which it can operate. It goes without saying, of course, that corporate charters do not (and cannot) confer the power to commit crimes. Thus, criminal wrongs committed on behalf of the corporation are legally beyond the organization's authority to act. And because a corporation is a mere legal abstraction, it has no soul to damn and no body to kick (Coffee 1981). It should come as no surprise, then, that if pure logic were to prevail, the likely consensus would be that corporations don't commit crimes. People do.

While this observation may seem intuitively correct, American courts and legislatures have seen fit to treat corporations as "persons" under the law. And once that step has been taken, it logically follows that recognition of a corporation's status as a legal person necessarily leads to corollary principles relating to corporate rights and liabilities. Thus, for example, a corporation can own property, enter into contracts that impose binding legal obligations, sue and be sued in its own name, and even be prosecuted and held responsible for crimes committed on its behalf.

Respondeat Superior Theory

Before turning to the larger practical and philosophical questions, it is essential to first consider the underlying legal principles, which fall under the rubric of the respondeat superior doctrine and provide the theoretical foundation for holding corporations criminally liable.

Notwithstanding conventional wisdom to the contrary, it is now well established under the respondeat superior doctrine that corporations can be held criminally responsible for wrongs committed in their names. The basic idea is that even though corporations are legal fictions, they are treated as "persons" under the law and are thus deemed capable of committing wrongful acts and forming criminal intent. Thus, for example, just as an accused murderer must be shown to have struck the fatal blow with intent to cause the victim's death, a corporation charged with price fixing must be shown to have acted in concert with another company to engage in anticompetitive conduct for the purpose of restraining trade. In the latter context, the corporation's criminal wrongdoing is established through the doctrine of respondeat superior – a legal sleight of hand that allows attributing the wrongful acts and intent of the agents who plan and implement the price-fixing scheme to the corporation itself.

Historical Underpinnings

The respondeat superior theory of organizational liability has its origins in the early English and American law of nuisance. Under English law, corporations were not accountable for criminal acts. As Chief Justice Holt wrote in 1701, "A corporation is not indictable but the particular members of it are" (*Anonymous*, 88 Eng. Rep. 1518 (K.B. 1701)). His oft-quoted statement reflected the belief that criminal liability was properly attributable to corporate owners, operators, and agents who committed criminal wrongs in their individual capacities rather than to the corporation itself.

But as corporations became more numerous and increasingly engaged in activities that had widespread effects on the public welfare, the idea that corporations could rightly be held criminally responsible for the consequences of their business operations came to be seen in a somewhat different light. Corporate business activities polluted rivers with rubbish and dead animal carcasses, operated malodorous slaughterhouses, caused the deterioration of bridges and roads, obstructed public highways with illegal buildings and railroad cars, and endangered public health and welfare in countless other ways. Because these wrongs harmed the public at large, the only procedure the law recognized to vindicate the public's interest in redressing this species of harm was criminal prosecution. For while the civil law allowed individuals to sue for particular injuries they had suffered, they had no legal right to sue on behalf of others to remedy the larger harm to the community.

The other obvious alternative – holding the individuals who directed or performed the offending acts – was deemed impractical for a number of reasons. Principal among them were that (1) the corporation, not the individual agent who acted on its behalf, was the primary (and usually sole) beneficiary of the wrongdoing; (2) the individual agents who permitted debris to fall into the river and obstruct navigation would be hard to identify and in all likelihood would lack both the financial and practical means to clean up the pollution; and (3) if the remedy involved a major economic outlay, it only made sense to make the corporation a party to the prosecution so it would have a fair opportunity to defend its actions – as, for example, to argue that a rogue employee who authorized or committed the wrongful acts was trying to sabotage the company. Of greater practical significance, however, was the absence of any legal procedure to enforce a judgment against a corporation – in this context, a judgment ordering the corporation to abate the nuisance – unless the corporation had been made a party to the proceeding. And so it was that the camel's nose entered the tent.

Legal Doctrine

Organizational liability for crimes and injuries caused by agents has a century-long history both in England and America. But it was in 1909, in the seminal case of *New York Central and Hudson River Railroad Company v. United States*, 212 U.S. 481 (1909), that the United States Supreme Court formally recognized the respondeat superior doctrine in corporate prosecutions. In *New York Central*, a railroad company was convicted of violating a federal statute that prohibited giving rebates to shippers who paid them to transport their goods. The statute expressly authorized criminal prosecution of corporations and their agents for giving the rebates, which in this case had been paid to sugar refiners to retain the refiners' business. The Supreme Court upheld the company's conviction by borrowing from the civil law doctrine of respondeat superior, which held employers financially responsible for injuries caused by employees who were acting in the course of their employment. Lawyers for the railroad argued that the statute should be declared unconstitutional because to hold the corporation criminally responsible for the illegal acts of the offending employee was, in effect, to punish innocent stockholders for crimes that neither they nor the board of directors had – or indeed could have – legally authorized. Rejecting that reasoning, the Court wrote:

> Since a corporation acts by its officers and agents their purposes, motives, and intent are just as much those of the corporation as the things done. If, for example, the invisible, intangible essence of air, which we term a corporation, can level mountains, fill up valleys, lay down iron tracks, and run railroad cars on them it can intend to do it, and can act therein as well viciously as virtuously. *Id.* at 492-93 (quoting Bishop's New Criminal Law § 417)

The Court went on to note that corporations can only act through their officers and agents and that it was clearly fair to hold a corporation accountable for wrongs committed by those to whom it had entrusted the authority to act in the matter at hand – here, setting rates for transporting sugar. But the Court also found that as a practical matter, it was necessary to hold the railroad liable for giving illegal rebates because of the increasingly powerful role that corporations had begun to play. As the Court put it: "[T]he great majority of business transactions in modern times are conducted through these bodies, and... interstate commerce is almost entirely in their hands.... It would be a distinct step backward to hold that Congress cannot control those who are conducting... interstate commerce by holding them responsible for the intent and purposes of the agents to whom they have delegated the power to act in the [matter]" (*Id*. at 495–96).

Thus, the principle that the acts and intent of corporate agents who have authority to act in a particular matter can be attributed to the corporation in a criminal prosecution gained the imprimatur of the United States Supreme Court and quickly became the predominant theory of corporate criminal liability in American law.

Limits on Respondeat Superior Liability

Although *New York Central* was essentially about whether Congress had the power to impose criminal liability on corporations, the Supreme Court's opinion firmly established that respondeat superior was an appropriate standard for recognizing this species of criminal responsibility. But that is not, of course, the end of the story. The Court's opinion did not attempt to define the outer boundaries of corporate liability, and it stands to reason that a rule that holds corporations responsible for the acts of their agents should not be unlimited. It would not make sense, for example, to hold a pizza parlor liable for assault if its delivery man shot a customer who refused to give him a tip. Thus, while respondeat superior quickly became the established doctrine for holding corporations criminally responsible, it remained for the lower courts to develop a coherent set of rules to contain this potentially expansive doctrine within reasonable bounds and balance the need to make corporations accountable for the wrongful acts of their agents against the need to avoid overly broad liability rules.

The Scope of Authority Requirement The primary limitations on corporate criminal liability are that the offending employee or agent must have acted within the scope of employment and with intent to benefit the corporate employer. Thus, for example, if the pizza delivery driver shot a customer who refused to give him a tip, the restaurant's corporate owner would not be liable for assault. Although the shooting occurred while the driver was on the job, committing an assault was not within the scope of his authority, that is, assaulting customers was not part of his job description. He was authorized to deliver pizzas, collect the tab, and accept a tip, but retaliating against a customer who refused to tip was clearly beyond the scope of his authority. Similarly, if Wall Street brokers conspired to engage in insider trading, they would be acting outside the scope of their authorized functions even though they might have legitimately become privy to inside information during the course of their employment.

But as might be expected, most cases are not so clear-cut and raise the knotty question of what constitutes an employee's scope of authority. As a practical matter, courts ordinarily give the term "scope of authority" a broad interpretation. Rather than simply looking at the employee's official job title or express authority, courts routinely examine the agent's apparent authority to engage in the offending conduct to determine whether the corporation should be held criminally liable. This rule reflects the not surprising observation that corporations rarely grant their employees specific authority to engage in criminal conduct, which in turn leads to the principle that a corporation may be held criminally responsible for conduct that exceeds what the employee has been expressly directed or authorized to do.

To return to the pizza delivery illustration, suppose the corporate owner of a pizza parlor chain adopts a marketing policy that guarantees free pizza if the delivery takes more than 30 min. But suppose that, in addition, the corporation's employee evaluation policy penalizes drivers who deliver a higher than average number of free pizzas. If our hypothetical driver speeds to avoid delivering too many free pizzas, his reckless driving is carried out at least in part to benefit the company by enabling it to sell more pizzas than it gives away. Thus, his reckless driving might well be attributed to the company. But that raises the corollary question of what happens if the corporate agent – here, the delivery man – has been specifically forbidden to engage in the wrongful conduct. Thus, for example, suppose that in addition to the free delivery marketing strategy and the employee evaluation policy described above, the corporation also has a strict rule that forbids drivers from speeding while they are delivering pizzas.

Or suppose that a hotel chain adopts a policy that expressly forbids its purchasing agents from giving preferential treatment to suppliers who pay a trade association's assessment. The purchasing agent not only disobeys specific instructions not to boycott recalcitrant suppliers but admits that he did so out of "anger and personal pique toward the individual representing the supplier" (*United States v. Hilton Hotels Corp*., 467 F.2d 1000, 1004 (9th Cir. 1972)). Under these circumstances, should the corporate owner be held responsible for the illegal boycott?

By and large, courts have concluded that it should. Even if a corporation has an "extensive program" to ensure employee compliance with the law, courts have rarely recognized the company's due diligence as a valid defense to corporate criminal liability. An employee's scope of authority includes "acts done on behalf of a corporation and directly related to the performance of the type of duties the employee has general authority to perform" (*United States v. American Radiator & Standard Sanitary Corp*., 433 F.2d 174, 204–05 (3d Cir. 1980)). Thus, "a corporate compliance program, however extensive, does not immunize the corporation from liability when its employees, acting within the scope of their authority, fail to comply with the law" (*United States v. Ionia Management S.A*., 555 F.3d 303, 310 (2d Cir. 2009) (internal quotations omitted)).

What policy supports such a seemingly harsh liability rule? First, notwithstanding the hotel chain's efforts to require its employees to comply with its policy, the underlying purpose is to provide the corporate employer a stronger incentive to implement its compliance policies and monitor its agents to ensure that they follow the applicable requirements of the law. Second, if the purchasing agent's strong-arm tactics were to succeed, the corporation would stand to benefit from increased business generated by members of the trade association. The hotel chain should not be permitted to escape responsibility for the purchasing agent's violation of corporate policy and the letter of the antitrust laws simply "by issuing general instructions without undertaking to enforce [them] by means commensurate with the obvious risks" (*Id*. at 1007).

Similarly, in the pizza delivery example, the corporation sent the drivers conflicting signals that likely reflected policies devised by different departments that could well have had incompatible interests and goals.

The Intent to Benefit Rule A second important limitation on the respondeat superior doctrine is the "intent to benefit" rule. For a corporation to be bound by the criminal acts of its agents, the acts must be performed with intent to benefit the corporate entity. But just what does that mean in this context? Like the scope of authority requirement, the intent to benefit rule is generally interpreted expansively by the courts. It does not require proof of any minimal level of actual or anticipated benefit to the corporation. Indeed, the corporation need not receive any benefit to be held vicariously liable for the agent's crime. Instead, the touchstone is the agent's intent. If the crime is intended to benefit the corporation, that is enough. Thus, even though the agent's conduct may actually harm the corporation, the wrongful acts are attributable to the company as long as the agent was acting within the scope of authority (broadly construed) and with intent to benefit the corporation.

To illustrate, consider a bank vice-president who embezzles millions of dollars from various accounts for personal use, the Wall Street brokers who trade on inside information in violation of their fiduciary duty to their corporate employer, and the pizza delivery driver who assaults the nontipping customer. All of the agents in these hypothetical situations are acting purely out of self-interest. The bank vice-president and the Wall Street insiders are motivated solely by a desire to reap a personal financial gain. Similarly, when the pizza delivery driver assaults the customer, his sole motive is to retaliate because he feels cheated by the customer's refusal to tip.

But what about corporate executives who "cook the books" to improve their company's reported financial condition? The executives are motivated primarily by a desire to artificially inflate the corporation's stock prices by making the company look more profitable than it actually is. Even if they derive a collateral benefit from the accounting fraud (e.g., receipt of bonuses or other forms of performance-based pay that are tied to the company's reported profitability), the corporation and its shareholders will be the principal beneficiaries because the essence of the fraud is to overstate the company's revenue to keep its stock price artificially high.

Relationship Between Individual and Organizational Liability

Since conventional wisdom holds that a corporation is liable for wrongs committed by agents who are acting on its behalf – and it can act and think only through its agents – it should come as no surprise that lawyers would argue that because the agents' acts were the acts of the corporation itself, the agents should not be held individually responsible for wrongs they committed on the company's behalf. Thus, the argument ran, established principles of agency law require courts to look "to the principal" – here, the corporation – "rather than the mere agent" (*State v. Morris & Essex R.R.*, 23 N.J.L. 360, 369 (1852)). Although this, too, may seem intuitively correct, intuition fails once again.

Modern law recognizes that the criminal liability of a corporation is distinct from the liability of the offending agent. Thus, because a corporation and its employees have distinct existences and the employees are the "human face" of the organization, the argument that the agents could not be held individually responsible because their acts were deemed in law to be the acts of the corporation was to no avail. A corporate agent who acts in a representative capacity is not entitled to use the corporate entity as a shield against personal responsibility for his own misconduct. Stated differently, holding the corporation legally responsible for criminal wrongdoing does not serve as a proxy for holding the culpable agent who committed the crime personally responsible as well, notwithstanding that the agent was acting on behalf of the corporation itself.

But the corollary principle is that the respondeat superior rule does not condition corporate liability on the prosecution and conviction of the responsible agent. On the contrary, the agent need not be charged with the crime or may have even been acquitted of committing it. Yet notwithstanding this seeming incongruity, an unsuccessful prosecution of the agent will not automatically vitiate the corporation's conviction. Even though the verdicts are inconsistent, the acquittal of the agent could be the product of jury mistake or lenity. Indeed, one court was incredulous that the jury had acquitted any of the accused agents and suggested that the acquittals could well be attributable to "considerations not rational at all" (*United States v. Austin-Bagley Corp.*, 31 F.2d 229, 233 (2d Cir. 1929)). Thus, rather than trying to second guess which of two unexplained and inexplicably conflicting verdicts reflects the "correct" result, the default rule is that as long as there is sufficient evidence to prove that someone acting on the corporation's behalf committed the crime, the corporation's conviction may legitimately stand.

And, as will be discussed in a later section, it may be difficult or impossible to identify the particular agent or agents who committed the violation because of the diffusion or compartmentalization of responsibilities that is typical of modern corporate hierarchies.

Rationales for Corporate Criminal Liability

The classical goals of criminal justice are premised on society's need to express moral disapproval of antisocial behavior and impose a punishment that serves, variously, the goals of retribution, rehabilitation, deterrence, and incapacitation. In the context of corporate prosecutions, perhaps the strongest rationales are deterrence and retribution.

Deterrence

As is true in the context of individual liability, there are two types of deterrence – specific and general – that are appropriate goals of corporate prosecution and punishment.

General Deterrence The theory of general deterrence is that punishing one actor will discourage other similarly situated parties (here, other corporations) from committing the same or similar crimes in the future. In the corporate context, a criminal indictment or conviction can have a negative impact on stock prices and the company's business reputation. Thus, for example, Arthur Andersen, a big-four accounting firm, collapsed under its own weight in the wake of its conviction for obstructing justice in the Enron accounting fraud scandal. Similarly – more than three decades earlier – E.F. Hutton, which was among the two or three largest brokerage firms in the United States, also vanished from the scene after pleading guilty to 2,000 counts of mail and wire fraud in conjunction with charges that the firm had obtained interest-free use of millions of dollars by engaging in what was at bottom an elaborate check-kiting scheme.

If these prosecutions were to serve as object lessons, one of the dominant themes would be a strong admonition to corporate managers that the consequences of crossing the line between corner cutting (e.g., engaging in "sharp practices") can be swift and severe. Although the managers at Andersen and Hutton may have believed their firms could survive the scandals, they clearly could not. Thus, it could be said that the goal of general deterrence – here, to exert a positive influence on high-level managerial decision making – was arguably well served by these prosecutions because the expectation is that other companies will learn from the bad examples of firms that have been caught up in the web of a major criminal case.

The electrical equipment scandals of the 1950s are a case in point. During the late 1950s, a number of electrical equipment suppliers – including industry leaders General Electric (GE) and Westinghouse – engaged in an elaborate price-fixing conspiracy to illegally coordinate bids on major government contracts and artificially maintain profits and market share. This was an example of an industry-wide practice that had widespread effects on competition that could not easily be remedied without making visible public examples of the offending corporations and their executives. In consequence, 29 electrical equipment companies – including GE and Westinghouse – were convicted along with 44 of their executives for the roles they played in the flagrant price-fixing scheme (*Corporations*: *The Great Conspiracy*, TIME, Feb. 17, 1961, http://www.time.com/time/magazine/article/0.,9171,826890,00.html).

Although by today's standards the penalties imposed were far from severe, they sent a strong message to the corporate community that major companies and their executives were not immune from prosecutorial scrutiny for conduct that crossed the line from what many corporate chieftains had considered to be aggressive (if possibly unethical) tactics, to what quickly came to be recognized as not only illegal but also criminal conduct.

Specific Deterrence The goal of specific deterrence is to ensure that the convicted party will not re-offend. In the corporate context, imposition of a criminal penalty on the company is intended to provide management a strong incentive to institute policies and oversight mechanisms that will prevent the offending activity from occurring again. This goal can be advanced in various ways, including specific sanctions or remedies that the court imposes as part of a corporation's criminal sentence or by collateral consequences that arise by operation of law or through discretionary actions undertaken by government agencies. And, of course, a convicted corporation may experience a change of fortune because of adverse market reactions to the scandal that led to the prosecution.

The most immediate and direct consequences that follow a criminal conviction are court-imposed sanctions. These can include a virtually limitless array of punitive and remedial measures such as a potentially crippling fine; appointment (at corporate expense) of a court-approved monitor who has expertise in the problematic part of the company's operations (e.g., accounting controls or environmental compliance) and will have oversight responsibilities to ensure that effective compliance mechanisms are put into place; requiring disciplinary action against responsible corporate agents; requiring corrective advertising or publication of a public apology for the offending conduct; and other measures specifically designed to address conditions that led to the criminal violation.

Sanctions may also come in the form of collateral consequences that are imposed by external forces. In addition to loss of reputation that may flow from the fact of conviction, other consequences may include mandatory or discretionary sanctions such as a Defense Department decision to debar a military contractor from bidding on a lucrative government defense contract. These considerations contributed to big-four accounting firm Arthur Andersen's refusal to plead guilty to a felony charge of obstruction of justice. Although a guilty plea might have brought an end to the relentless adverse publicity attendant to the high-profile criminal investigation of the firm's destruction of Enron records, Andersen refused to plead guilty because the SEC would not guarantee that it would waive an agency rule that would disqualify the firm from auditing public companies if it were convicted of a felony. That, in turn, would only exacerbate concerns about the viability of the firm and thus contribute to further erosion of Andersen's client base.

Similarly, in the E.F. Hutton case, the collateral consequences of Hutton's conviction were catastrophic. As part of its guilty plea to engaging in a massive mail and wire fraud scheme, Hutton agreed to pay a $2 million fine, $750,000 for the costs of the government's investigation and $8 million in restitution to the affected banks. In addition, Hutton incurred nearly $100 million in legal fees as it tried to extricate itself from its regulatory nightmare.

Although Hutton might have survived these financial consequences of the criminal investigation, it also faced countless civil law suits and intense regulatory scrutiny – including investigations (and ultimately censure) by the New York Stock Exchange, the US Department of Labor, and several congressional committees – as well as looming threats by a number of state agencies to forbid Hutton from doing business in their respective states and to subject the firm to additional regulatory checks.

But Hutton's loss of reputation after the plea agreement also costs it dearly, especially in one of its strongest units – its municipal finance division. Illustrative examples include the New York Metropolitan Transportation Authority's (MTA) decision to remove Hutton from a list of firms that were eligible to participate in MTA's future underwritings; New York City's removal of Hutton as an underwriter for a $592 million bond financing – the City's then-largest financing ever; and the Massachusetts Housing Financing Agency's threat to remove Hutton from a select list of investment banks that were allowed to serve as managing underwriter on the Agency's bonds.

That said conditions were ripe for a perfect storm.

Retribution

Retribution is another classic justification for recognizing corporate criminal liability because it is seen as an appropriate way to punish a corporation that financially benefits from its criminal activities. Retribution, of course, has ancient roots in the Biblical concept of "an eye for an eye, a tooth for a tooth."

Punishment is considered an appropriate rationale for imposing criminal liability on a corporation because the illegal conduct commonly enriches the corporation, and punishment is deemed mete out "just deserts." Some commentators have even suggested that in the context of corporate criminal liability, retribution should be the primary motivation for punishment because modern corporations develop distinctive personae through increasingly sophisticated marketing techniques (Friedman 2000). For example, a corporation may affect public perception of the organization by advertising its products and packaging as "green" or environmentally sustainable. Consumers who consider themselves environmentally conscious may then reward these businesses by purchasing their products over other, cheaper – or perhaps objectively superior – alternatives. Thus, the argument runs, it only seems appropriate that if society can give relative moral approval to a corporation's "good deeds," it should likewise have a mechanism for expressing moral disapproval for wrongs committed in the company's name – especially if the corporation's environmental claims turn out to be bogus.

Justice Department Corporate Prosecution Guidelines

Because corporations are persons only in a juristic sense, factors other than the traditional rationales for criminal prosecution and punishment often come into play. Many of these nontraditional factors are incorporated in Justice Department corporate prosecution guidelines that are designed to aid federal prosecutors in deciding whether to bring criminal charges against a corporation. Although the guidelines confirm that traditional considerations such as the potential deterrent value of a criminal prosecution should remain part of the equation, they also recognize that because corporations are "persons" in a uniquely different sense, additional considerations should be taken into account as well.

Those considerations include the following: (1) the nature and seriousness of the offense and whether it posed a risk of harm to the public; (2) whether the offense was an isolated instance of wrongdoing within the corporation and whether the organization's managers condoned or participated in it; (3) whether the corporation had previously engaged in similar misconduct and – if so – whether it had been prosecuted, sued, or otherwise subject to regulatory scrutiny because of the prior incident; (4) whether the corporation had voluntarily disclosed the wrongdoing and cooperated in the investigation of those who may have been responsible for it; (5) whether the corporation had an effective compliance program in place when the wrongdoing occurred; (6) what steps the corporation had taken to correct the conditions that led to the violation, including disciplining or firing responsible agents; (7) whether a prosecution would be likely to harm shareholders, employees, and others who were not personally responsible for the violation; (8) whether prosecuting the responsible individuals would be an adequate response to the wrongdoing; and (9) whether other remedies short of criminal prosecution would likely be sufficient (United States Department of Justice, Principles of Federal Prosecution of Business Organizations § 9–28.300 (1997)).

Two points should be kept in mind, however. First, like all Justice Department guidelines, the principles of corporate prosecution are nonbinding and are not legally enforceable against the Department if prosecutors fail to follow them. And second, the guidelines do not purport to be statements about when a corporation can be prosecuted under established respondeat superior principles. Instead, they are intended to provide prosecutors general guidance on the desirability of filing criminal charges against a corporation under the particularized facts and circumstances of the case under consideration.

Practical Considerations

Beyond the classical justifications for imposing criminal punishment, there are unique practical reasons why courts and legislatures have fashioned rules for holding corporations criminally liable for crimes committed on their behalf. First and foremost is that in large, complex organizations, individual responsibility is – of necessity – diffused. It is a fact of modern corporate life that corporations must rely on an extensive system of delegation to carry out the organization's ordinary, day-to-day business activities. Subordinate employees often have considerable autonomy in the performance of their responsibilities and may actually have (and exercise) greater power to act in matters relating to the corporation's everyday operations than high-level managers. Yet despite the considerable influence that lower-level operatives may wield over the conduct of the corporation's day-to-day business, they often remain shielded from personal liability by a complex web of corporate hierarchies that make it difficult if not impossible to identify which of them violated the law.

Although the corporate fraud scandals of the early 2000s stand out as a notable exception to the rule (scores of corporate executives and mid-level managers went to prison in the flurry of post-Enron prosecutions), high-level managers ordinarily will not be directly involved in the wrongdoing. Thus, for example, although the board of directors or high-level managers of a shipping company may be indifferent to environmental compliance issues, it is unlikely that there will be evidence that they expressly authorized subordinate employees to discharge untreated waste into the ocean. At the same time, even though there may be a corporate policy against illegal discharges, any one of many subordinate crew members or engineers might decide on their own to cut corners, defy corporate policy, and engage in "midnight dumping" on the sly. Although it may be evident that the dumping has occurred, the task of identifying the crew member who opened the valve that permitted the release of the waste can be daunting. And it may be equally difficult to establish that managers in the upper echelons of the company's hierarchy (or, for that matter, the ship's captain) either knew of or tacitly approved the illegal shortcut by turning a blind eye to parts of the ship's operations where ocean dumping and other environmental violations are most likely to occur.

Corporate Prosecution and Punishment

No discussion of corporate criminal liability would be complete without at least brief consideration of how the underlying theories supporting recognition of corporate liability translate into practice.

Prosecution

Critics of corporate criminal prosecutions lament that current doctrine is overly broad and unfair. Yet despite the expansive liability principles they lament, the respondeat superior doctrine has not led to unchecked prosecutorial zeal for charging corporations with criminal misconduct. On the contrary, corporate prosecutions are relatively rare. Although there are few comprehensive sources that systematically track the incidence of corporate prosecutions and convictions, some specialized studies help to fill the void. One such study, which was conducted by the Department of Justice Bureau of Justice Statistics (DOJ Study), focused on federal enforcement of environmental laws between 1994 and 1997. The study found that only 17 % of those charged with environmental or wildlife offenses were organizations.

The results of the DOJ study also included the following findings relating to investigations of environmental protection violations and wildlife offenses. In 1997, federal prosecutors concluded 871 investigations of environmental or wildlife violations and declined to prosecute in more than 60 % of those cases, but the declination rate in corporate investigations was much higher than it was for individuals who were investigated for comparable crimes. In contrast with decisions to decline prosecution in 55 % of cases in which individuals were investigated for environmental crimes and 38 % of cases in which they were investigated for wildlife offenses, prosecutors declined to charge corporations investigated for violations in 70 % of environmental protection investigations and in 67 % of investigations for wildlife offenses (U.S. Dept. of Justice Office of Justice Programs, *Bureau of Justice Statistics Special Report*: *Federal Enforcement of Environmental Laws*, *1997*, at 4 (1997)).

A similar disparity between individual and corporate prosecutions was reported in an empirical study of hazardous waste prosecutions under the federal Resource and Conservation Recovery Act that was conducted by this author (Brickey Study). Relying primarily on data compiled by the Environmental Protection Agency's National Enforcement Investigations Center for the decade ending in 1992, the Brickey Study found that of roughly 140 hazardous waste prosecutions, about 30 % of the defendants were corporations (Brickey 2001).

In addition to traditional criminal prosecutions, Deferred Prosecution Agreements (DPAs) – which have proliferated in recent years – provide an alternative procedure for addressing corporate wrongdoing. Deferred Prosecution Agreements are conditional contracts between prosecutors and a corporation that is facing potential criminal prosecution.

These agreements grant corporations provisional pretrial diversion from prosecution in exchange for cooperation with the Justice Department's investigation. In many instances the agreement is conditioned on the corporation's payment of a substantial fine, the appointment of a corporate monitor to oversee problematic aspects of the corporation's operations, or structural or personnel changes within the organization. DPAs do not ordinarily result in an immediate agreement to refrain from criminally prosecuting the corporation. Instead, they effectively place the corporation under government supervision for a specified period of time and defer the decision whether to actively pursue criminal charges pending a determination whether the corporation will successfully fulfill the terms and conditions of the agreement.

There has been a dramatic increase in the use of DPAs as a form of corporate oversight in the wake of the recent corporate accounting fraud scandals. Between 1974 and 2003, there were only 18 documented DPAs and similar pre-prosecution agreements (Nanda 2010). In contrast, between 2003 and 2008, the Justice Department entered into at least 103 such agreements with corporate defendants.

Punishment

Although deferred prosecution agreements almost invariably involve exacting some concession from the corporation such as payment of a fine or restitution, these exactions are a product of ad hoc decisions by prosecutors about the most effective and efficient means to resolve a case through a negotiated agreement rather than through formal punishment imposed by a court upon conviction.

Since corporations cannot be imprisoned, corporate punishment usually takes the form of a criminal fine. As would be expected, the size of the fine depends on the severity of the crime. Although the maximum (and occasionally minimum) fine is customarily set by statute (e.g., the maximum corporate fine for price fixing is $100 million, but no minimum fine is specified), courts have considerable discretion in determining the size of the fine to be imposed within the statutorily authorized range.

While it is not uncommon for individuals who are convicted of crimes to be sentenced to pay a fine, imposition of a criminal fine is more prevalent in corporate prosecutions. In 2010, for example, 77 % of corporate defendants were fined, while only about 10 % of individuals convicted in federal court were sentenced to pay a fine (United States Sentencing Commission Annual Reports 2010, Ch. 5 at 38–39). Not surprisingly, corporate fines tend to be more substantial than fines imposed on individuals in the run-of-the-mill criminal case. In 2010, the average criminal fine imposed against corporations convicted of violating federal law was about $16 million. The highest fine, which was levied against a pharmaceutical company for federal food and drug law violations, was nearly $1.2 billion. These figures reflect a recent upward trajectory in federal criminal fines. Yet while fines remain a mainstay of corporate criminal punishment, court-ordered restitution to the victims of a company's wrongdoing also plays an important role in corporate sentencing. In 2010, restitution was ordered in nearly 25 % of sentences imposed on corporations, and the average amount that corporations were required to pay was nearly $18 million (*Id*. at 38).

While monetary sanctions in corporate prosecutions are common, corporations convicted of crimes are also routinely subject to nonmonetary sanctions like probation or mandatory ethics and compliance training programs. In 2010, slightly more than 70 % of corporations sentenced under the Federal Sentencing Guidelines received some form of probation (*Id*. at 39). This number generally tracks with previous years, during which a majority of corporate sentences included a term of probation. See generally 2003–2010 United States Sentencing Commission Annual Reports. The terms of corporate probation vary both in scope and length, but all have the universal goal of ensuring future compliance and, if need be, changing a lax "corporate culture" that allowed the criminal activity to occur.

But there are less conventional punishments in the Sentencing Guidelines' panoply of corporate sanctions. Perhaps most notable among them is the so-called corporate death penalty, a sanction that is strictly reserved for what the Guidelines term "criminal purpose organizations," that is, organizations that operate primarily for a criminal purpose or primarily through criminal means. Thus, for example, a hazardous waste company that has no legitimate way to dispose of hazardous waste and systematically dumps drums of dangerous chemicals in rural woods and streams would likely be deemed a criminal purpose organization. And if the company is found to be a criminal purpose organization, the Guidelines require the court to impose a fine that is high enough to divest the company of all of its assets, thus causing it to cease to exist.

Conclusion

And so the discussion has come full circle. It began with what some would consider a highly dubious premise that corporations possess sufficient attributes of personhood that – like their human counterparts – they are capable of committing crimes. The narrative ends on what is perhaps an even more surprising note – that the theory of corporate personhood has become so deeply entrenched that, in the eyes of the law, this "invisible, intangible essence of air" can even be sentenced to death.

Related Entries

- ▶ Corporate Crime Control
- ▶ Corporate Crime Decision-Making
- ▶ State-Corporate Crime
- ▶ Theories of Punishment
- ▶ Victims of Corporate Crime

Recommended Reading and References

Brickey KF (1982) Corporate criminal liability: a brief history and an observation. Wash Univ Law Q 60:393–423

Brickey KF (1987) Death in the workplace: corporate liability for criminal homicide. Notre Dame J Law Ethics Pub Policy 2:753–790

Brickey KF (1992–1994) Corporate criminal liability, 2nd edn. Clark Boardman Callaghan, Deerfield, 3 vols; annual supplements 1995–present

Clinard MB, Yeager PC (2006) Corporate crime. Transaction Publishers, New Brunswick

Coffee JC Jr (1981) No soul to damn: no body to kick: an unscandalized inquiry into the problem of corporate punishment. Mich Law Rev 79:386–459

Cullen FT, Cavender G, Maakestad WJ, Benson ML (2006) Corporate crime under attack: he fight to criminalize business violence, 2nd edn. Lexis-Nexis, Albany

Friedman L (2000) In defense of corporate criminal liability. Harv J Law Pub Policy 23:833–858

Nanda VP (2010) Corporate criminal liability in the United States: is a new approach warranted? Am J Comp Law 58:605–630

O'Reilly JT, Hanlon JP, Hall RF, Jackson SL, Lewis ER (2009) Punishing corporate crime: legal penalties for criminal and regulatory violations. Oxford University Press, New York

Orland L (1995) Corporate and white collar crime: an anthology. Anderson Publishing, Cincinnati

United States Department of Justice (1997) Principles of federal prosecution of business organizations

United States Department of Justice, Bureau of Statistics (1995) Sourcebook of criminal justice statistics 1995

(2003–2010) United States sentencing commission annual reports

Wells C (2001) Corporations and criminal responsibility, 2nd edn. Oxford, New York

Correction

▶ Punishment as Rehabilitation

Correctional Administrators

► Prison Warden Stress and Job Performance

Correctional Education of Programs

Daniel R. Lee
Department of Criminology, Indiana University
of Pennsylvania, Indiana, PA, USA

Overview

Correctional education programs include strategies and programs to improve the knowledge, skills, and abilities of incarcerated individuals. These programs are administered at various levels of instruction and include general and specific topics. The programs can best be categorized as academic or vocational. Levels of academic instruction include basic education, secondary education, and post-secondary education. Vocational programs vary in level of instruction, too, and they range from basic employment skills to advanced, professional certification. Additionally, many correctional education programs have separate or integrated components related to enhancing life skills, cognitive abilities, and other pro-social behaviors. Recent reports indicate that correctional populations are less educated than the general public and that nearly 40 % of prisoners have earned less than a high school diploma (Harlow 2003). Released prisoners who have earned at least a high school education are known to have lower rates of rearrest, reconviction, and reincarceration (Beck and Shipley 1989; Harlow 2003). These relationships have placed correctional education programs in a controversial position within the discussions of which services should be provided to inmates that might ease reentry, improve social reintegration, and reduce recidivism and how the costs of prison administration can or should be controlled by eliminating or reducing the frequency of custodial services and programs.

Brief History

Correctional education programs, especially those within the United States or North America, have an extensive history that dates to the origin of incarceration as a source of punishment. In 1787, educational instruction was offered by clergyman William Rogers in the Walnut Street Jail (Gehring and Eggleston 2006, Hanneken and Dannerbeck 2007). Later, Zebulon Brockway used educational programming as a rehabilitative tool during his administration as superintendent at the Elmira Reformatory in New York (Brockway 1912, Gehring n.d.). The modern era of correctional education likely began with the research of Austin MacCormick and the publication of *The Education of Adult Prisoners* (1931–1976). MacCormick supported Brockway's principles and went on to found the Correctional Education Association first as a part of the American Correctional Association and later as a separate organization (Correctional Education Association n.d.). The use of correctional education programs as a part of rehabilitation initiatives was dampened in the latter part of the twentieth century as the United States experienced philosophical shifts regarding crime and punishment and as a result of Martinson's (1974) report on the effectiveness of rehabilitation. Today, it is reported that approximately 80 % of state and federal prisons offer some type of academic program (Harlow 2003).

At times, legislative changes have contributed to the history of correctional education programming within the United States. Notably, the Violent Crime Control Act of 1994 included provisions that eliminated Pell Grants for inmates enrolled in post-secondary correctional education (PSCE) programs. (Federal Pell Grants are based upon documented financial need and provided to low-income undergraduate students to promote higher education; these grants began as Basic Education Opportunity Grants under the Higher Education Act of 1965 (Department of Education, n.d.).) Ubah and Robinson (2003) report that, at the time of the passage of the VCCA, Pell Grants provided the primary source of funding for PSCE, and their legislative elimination carried political and social implications for correctional education programming. Initial assessments of the impact of this legislation showed a near immediate decline in funding for and administration of PSCE programs (Tewksbury and Taylor 1996). Later, Tewksbury et al. (2000) reassessed the relationship and attempted to measure the impact that the elimination of Pell Grants had on PSCE programs. In their research, Tewksbury et al. (2000) surveyed directors of adult education programs regarding the extent of PSCE programs and their perceptions about funding for PSCE programs. They found that after the elimination of Pell Grant funding, the number of certificate, associate degree, baccalaureate degree, and graduate degree programs had decreased. Additionally, they reported decreases in the percentage of inmates participating in PSCE programs and a greater decrease in the percentage of eligible inmates who had actually enrolled in PSCE programs. Directors' perceptions of programs were negative. Over 40 % of the directors indicated that the elimination had changed their programs, and fewer than 20 % indicated that the elimination had not changed their programs. Similarly, Ubah and Robinson (2003) reported that changes to funding in PSCE have occurred, but many states have been able to continue administering programs by finding alternate sources of funding. For example, some programs began to rely on state-based sources of funding for their PSCE programs. Related to this, Erisman and Contardo (2005) reported that as of 2003, the enrollment of inmates in post-secondary education programs had returned to the levels that existed prior to the elimination of Pell Grant funding.

Spangenberg (2004) reported that another piece of federal legislation, the Workforce Investment Act, had placed caps on the funding for correctional education programs. She stated that with the WIA, the previous minimum of 10 % of adult basic education funds for state programs was changed to become a 10 % maximum. Her report presented information that, in several states, legislation provided mandates that inmates with certain levels of literacy receive education, but despite the mandates, some states became challenged with providing funds for these programs because local economic conditions warranted more restrained spending. Despite this, her report stated that many states were able to continue correctional education programming, despite the legislative changes in federal funding, by identifying alternate streams of funds.

In 2008, the Second Chance Act (P.L. 110–199) was authorized and included provisions to fund programs related to the successful reentry of inmates returning to communities (Justice Center, n.d.). These funds are designed to develop and provide resources to former inmates that include employment assistance, housing, substance abuse counseling, and other services. A portion of the Second Chance Act funds have been used to establish a study by the RAND Corporation to assess the impact of correctional education programs on the successful reintegration of former inmates (Turner and Davis 2011).

Extent of Correctional Education Programs and Participants

In 2007, the American Correctional Association conducted a survey of correctional education programs (Corrections Compendium, 2008). The results of this survey indicated that approximately 27 % of inmates at that time were participating in correctional education programs and many more were on wait-lists to join future programs. Many programs included incentives for inmate participants, including good time credits for program completion, monetary rewards or pay-per-day for participation, employment opportunities within the institution, and/or transfer to facilities closer to family members. Across state systems at the time, available correctional education programs included adult basic education (ABE), General Educational Development (GED), vocational/technical, advanced academic or Adult Secondary Education (ASE) (e.g., 2-year degree, 4-year degree, and postgraduate), and other specialized instruction (e.g., special education, English as a second language (ESL), prerelease, and job readiness), with ABE, GED, and vocational/technical programs being the most prevalent. Despite established needs among correctional populations, correctional education programs are not available across all institutions (Harlow 2003).

The population of correctional education participants includes many who qualify for special education (Harlow 2003). Leone et al. (2008) report that accurate measurement of the proportion of inmates in need of special education services is difficult to obtain, but those with disabilities are entitled to accommodations under the Americans with Disabilities Act and Section 504 of the Vocational Rehabilitation Act. Similar provisions are included in the Individuals with Disabilities Act of 1997. Leone et al. (2006) report that special consideration must be provided to juvenile offenders, and often, the issues related to their education are regulated by state or federal mandates tied to program funding.

Assessment and Evaluation of Correctional Education Programs

Tracy and Steurer (1995) developed a model evaluation instrument with the intent to assess existing correctional education programs. The purpose of their instrument, funded by the National Institute of Justice, was to create an evaluation tool that would identify the strengths of previous assessments and evaluations of correctional education programs and would be applicable across all state system and federal educational programs. The interest in developing this tool was to overcome some of the disadvantages of past evaluations. They had identified that independent evaluations of correctional education programs had been done, but individual agencies or facilities had conducted these evaluations with singular purposes related to immediate agency needs. That is, a specific agency, facility, or program would be evaluated with no interest or applicability in comparing specific results to other programs, agencies, or extant research findings. Based on a review of the existing literature, their model evaluation instrument focused on the best of existing evaluation designs (e.g., standardizing uniform measures of recidivism and other offender behaviors, identifying appropriate control and/or comparison groups, incorporating statistical control variables, and administering intake and exit measurements to identify causal patterns) while minimizing other methodological concerns (e.g., selection bias related to inmates' personality or interest in succeeding in correctional education programs).

Evaluations of correctional education programs include the goal of identifying the success of the program, but there is no common measure or sentiment about how success should be defined. Possible measures of success include improved skills among inmates (e.g., grade level increases, literacy, and earned certification or degree), reduced administrative costs (for the facility or correctional system administration), reduced recidivism, or enhanced social position such as employment, higher wages, home ownership, or enrollment in higher education upon release. The varied measures of success create challenges for researchers and policy analysts who want to summarize or generalize findings across studies.

Identifying the relationship between correctional educational programs and recidivism is a typical objective for assessments of program success, but the measurement of recidivism is not standardized. The measurement of recidivism across evaluations includes post-release self-reported measures of offending, arrests, convictions, and reincarceration. This variation across evaluations in the measurement of recidivism makes it difficult to draw comparisons and summary conclusions. Within this measure, another evaluation concern is associated with studies that measure offending and distinguish inconsistently the differences between new crimes and technical violations of parole conditions that lead to revocation. Additionally, the time of exposure or time following release from prison or jail must be considered, as the follow-up period of evaluation varies across studies; longer periods of exposure could be associated with increased rates of recidivism that might not be related to educational programming.

Another methodological concern is related to the structure of correctional education programming. Many correctional education participants are enrolled in multiple courses at the same time, and as this happens, it is difficult to identify which courses might contribute to the relationship between correctional education curricula and recidivism. Also, it is possible that multiple courses or modes of instruction could have interactive as well as additive relationships. As mentioned above, enrollment in correctional education courses might include incentives for pay or early release in addition to the intrinsic value of education. It could be suggested that correctional educational programs prepare inmates for successful reintegration by providing credentials and skills that transfer to successful employment, but another possible relationship is that education produces cognitive changes and creates social capital that promotes pro-social behaviors and relationships. As incentives and mandates for participation exist, it can be challenging for evaluators to recognize the true sentiment of the correctional education participants. This becomes a concern for evaluators if inmates' sentiment is what causes the inverse relationship with offending; that is, the effect of correctional education programs could be spurious or indirectly operating through variations in offender sentiment. Additionally, evaluations of correctional education programs must recognize and control for the fact that some programs are directly connected to reintegration and reentry services that might enhance the effect of educational programming.

Although many individual program assessments exist, an adequate summary of their findings is difficult due to the conceptual and methodological challenges presented above. Comparative assessments and meta-analyses can be used to establish the general benefits of correctional education across settings and programs, but these are rare. Steurer et al. (2001) analyzed data from three states to determine the effect of correctional education programming on recidivism and employment across three states (i.e., Maryland, Minnesota, and Ohio). The study is unique in that participants from multiple states were assessed in the same evaluation. In this study, offenders were surveyed prior to their release, and state agencies contributed data on recidivism, including arrest, conviction, and reincarceration, and employment for 3 years post-release. The results indicated that correctional education participants had lower rates of recidivism and higher earned wages than the nonparticipants.

Meta-analyses allow for cross-study comparisons that can summarize and highlight the general findings associated with correctional education program success. Cecil et al. (2000) conducted a meta-analysis of adult basic education and life skills programs. They reported that many of the individual program evaluations that were included in the meta-analysis showed positive effects on recidivism, but the effect varied by the population characteristics of the participating inmates and the groups included as comparisons. Their review identified the hazard within this body of literature that the results must be considered in the context of the quality and rigor of the methodology for evaluation.

Bouffard et al. (2000) evaluated the literature related to vocational education and employment programs among adult offenders. Using the Maryland Scientific Measures Scale (see Sherman et al. 1997), they rated the methodology of each program evaluation included in the assessment to weight the individual conclusions according to their scientific rigor. Based on their review, they concluded that vocational education programs and correctional industry programs that incorporate vocational education are related to reductions in recidivism, but they concluded that community-based employment programs have not been evaluated with enough scientific rigor to offer summary conclusions.

In a more general meta-analysis that included evaluations of many types of correctional educational programs, Wilson et al. (2000) reviewed 33 experimental or quasi-experimental assessments of correctional education programs (including programs of various academic and vocational instruction). They concluded that the effects of the correctional education programs are variable, but in general, the programs are related to a lower likelihood of recidivism among program participants when compared to nonparticipants. Similarly, their analysis identified that program participants were more likely to be employed post-release. Wilson et al. (2000) presented these findings with a cautionary interpretation that the results, despite being based on many quasi-experimental or experimental evaluations, provide equivocal evidence that correctional education programs are successful. With many of the studies included as a part of the meta-analysis, the differences between program participants and nonparticipants that were statistically controlled often were limited to demographic characteristics such as age or race; a contrast that was excluded is the offenders' personal motivation for program participation and desistance from crime. If participants have a personal sentiment related to pro-social behavior, the results could be biased in favor of program and participant success.

Other researchers have tried to identify other summary conclusions related to the effect and importance of correctional education programs. Drake et al. (2009) presented a meta-analysis that focused on the ability of various criminal justice policies and programs, including correctional education, to reduce crime and administrative costs. They included in their review several types of correctional education programs, including basic education, post-secondary education, vocational education, and life skills training. They determined that correctional education showed general evidence that the programs have criminal justice and economic benefits. This conclusion was based upon reductions in recidivism associated with correctional education programs that translate into reduced costs for criminal justice program administration and victims of crime.

In 2006, MPR, Inc., a private educational policy firm, received funding from the US Department of Education to develop a common set of data definitions that measure the characteristics of correctional education programs and participants (Lee et al. 2012). The project was designed to facilitate comparative assessments of correctional education programs across states by requesting that state departments of corrections identify and collect data regarding a core set of variables with standardized definitions. In 2011, a revised data guidebook was established by MPR in a partnership with researchers at Indiana University of Pennsylvania. The data guidebook defined key variables across inmates, facilities, and state departments of correction, and when possible, the guidebook included variables that were consistent with existing, mandated educational program data reporting requirements. This revised data guidebook was pilot tested with seven states to determine the feasibility of having states report correctional education data in this standardized format. Lee et al. (2012) found that many states are willing and able to adjust their data collection and reporting practices to promote the cross-state documentation and evaluation of correctional education programs. In the future, standardized data collection and reporting practices will facilitate comparative analyses and program evaluations.

Related Entries

- ▶ Early Release from Prison
- ▶ History of Corrections
- ▶ Punishment as Rehabilitation
- ▶ Reentry
- ▶ Rehabilitation
- ▶ Theories of Punishment

Recommended Reading and References

Beck A, Shipley B (1989) Recidivism of prisoners released in 1983. US Department of Justice, Washington, DC, NCJ#116261

Bouffard J, MacKenzie D, Hickman L (2000) Effectiveness of vocational education and employment programs for adult offenders a methodology-based analysis of the literature. J Offender Rehabil 31:1–41

Brazzell D, Crayton A, Mukamal D, Solomon A, Lindahl N (2009) From the Classroom to the Community: exploring the Role of Education during Incarceration and Reentry. Urban Institute, Washington, DC

Brockway Z (1912) Fifty years of prison service: an autobiography. Charities Publication Committee, New York

Cecil D, Drapkin D, MacKenzie D, Hickman L (2000) The effectiveness of adult basic education and life-skills programs in reducing recidivism: a review and assessment of the research. J Correct Educ 51:207–226

Coley R, Barton P (2006) Locked up and locked out: an educational perspective on the U.S. prison population. Educational Testing Service, Princeton

Correctional Education Association, n.d. http://www.ceanational.com/history.htm. Retrieved 20 June 2012

Corrections Compendium (2008) Inmate education programs. Correct Compend 33:9–25

Center on Crime, Communities & Culture (1997) Education as crime prevention: providing education to prisoners. https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=172486

Department of Education (n.d) Federal Pell Grant Program. http://www2.ed.gov/programs/fpg/index.html. Retrieved 20 June 2012

Drake E, Aos S, Miller M (2009) Evidence-based public policy options to reduce crime and criminal justice costs: implications in Washington state. Victims and Offenders 4:170–196

Erisman W, Contardo J (2005) Learning to reduce recidivism: a 50 state analysis of postsecondary correctional education policy. Institute for Higher Education Policy, Washington, DC

Gehring (n.d) What works and why. http://coe.csusb.edu/programs/correctionalEd/documents/WhatWorks.pdf. Retrieved 20 June 2012

Gehring T, Eggleston C (2006) Correctional education chronology. California State University San Bernardino, San Bernardino

Hanneken D, Dannerbeck A (2007) Addressing offenders' educational opportunities and challenges. Correct Compend 32(1–4):37

Harlow CW (2003) Education and correctional populations. U.S. Department of Justice, Washington, DC, NCJ # 195670

Justice Center (n.d) The second chance a. http://reentrypolicy.org/documents/0000/1277/2.14.12_Second_Chance_Act_Fact_Sheet_.pdf. Retrieved 20 June 2012

Lee D, Giever D, Tolbert M, Rasmussen L. Bridging the correctional education information gap: lessons learned from piloting a voluntary correctional education data collection system. MPR Associates, Berkley

Leone P, Wilson M, Krezimen M (2008) Understanding and responding to the education needs of special populations in adult corrections. Paper presented at the John Jay Reentry Roundtable on Education. Accessed 31 Mar–1 April 2008

MacCormick A (1976) The education of adult prisoners: a survey and a program. AMS Press, New York

Martinson R (1974) What works? Questions and answers about prison reform. Public Interest 35:22–54

Sherman LW, Gottfredson D, MacKenzie DL, Eck J, Reuter P, Bushway S (1997) Preventing crime: what works, what doesn't, what's promising. National Institute of Justice, Washington, DC

Spangenberg G (2004) Current issues in correctional education. Council for the Advancement of Adult Literacy, New York

Steurer S, Smith L, Tracy A (2001) Education reduces crime: three state recidivism study. Correctional Education Association, Lanham

Tewksbury R, Erikson D, Taylor J (2000) Opportunities lost: the consequences of eliminating pell grant eligibility for correctional education students. J Offender Rehabil 31:43–56

Tewsbury R, Taylor J (1996) The consequences of eliminating pell grant eligibility for students in post-secondary education programs. Fed Probat 60:60–63

Tracy A, Steurer S (1995) Correctional education programming: the development of a model evaluation Iinstrument. J Correct Educ 46:156–166

Turner S, Davis L (2011) Improving the educational experiences of incarcerated adults and juveniles. Paper presented at the 63rd Annual Meeting of the American Society of Criminology. Washington, DC

Ubah C, Robinson R (2003) A grounded look at the debate over prison-based education: optimistic theory versus pessimistic worldview. Prison J 83:115–129

Wilson D, Gallagher C, MacKenzie D (2000) A meta-analysis of corrections-based education, vocation, and work programs for adult offenders. J Res Crime Delinq 37:347–368

Correctional Interventions

▶ Examining the Effectiveness of Correctional Interventions

Correctional Treatment

▶ Examining the Effectiveness of Correctional Interventions

Correctional Workers in the Organizational Change Process

Danielle S. Rudes and Jill Viglione
George Mason University, Fairfax, VA, USA

Overview

The amassed literature on organizational change casts little doubt that change is a premier topic in the scholarship on organizations. Despite a litany of studies focusing on the organizational change process and outcomes, cumulative knowledge about change is in its relative infancy. Perhaps this stems, at least in part, from the considerable debate in prior and ongoing research about how best to define and study organizational change. On one side, scholars present change as a rupture to existing and stable organizational environments. In this view, organizations operate by adhering to a sequential order where a distinct set of events or routines form the basis for organizational operations. When revised or new events upend current practices through planned or spontaneous actions, the organization and its workers endure change. In this input-process-output or unfreezing-moving-refreezing model, "synoptic" accounts present organizational change as an accomplished or failed event, or set of events, rather than an ongoing process.

An alternative view of organizational change presents organizations as a series of continual processes that together create an organizational entity with its own identity. In this view, culture sensemaking and resistance provide three of the many ways organizational actors continually grow and develop to help organizations fortify or redesign their identity. In this model, change is a dynamic and fluid process that is a normative part of organizational life. Organizations and organizational actors continually adjust to changes as they occur and/or possess enough agency (capacity of individuals to act independently and to make their own free choices to bring about change themselves. Scholars in this framework consider the dynamic interactions present within workers' micro-processes (tacit/everyday knowledge) carried out in practice and routines (repetitive behavioral patterns that are habitual and often stagnant for forging understanding about how change processes operate). Although not every change is successful, processual change is an organizational certainty.

Within the second theoretical and conceptual framing outlined above, empirical research on organizational change within criminal justice, and particularly correctional systems, regularly considers how organizations and their workers manage change. This entry focuses exclusively on three prominent themes present in organizational change research at the individual level of analysis: (1) culture/climate, (2) individual and group resistance, and (3) sensemaking/perception. The following paragraphs outline what is known about how workers, particularly correctional workers, perceive, understand, and navigate organizational change. In conclusion, what remains unknown is discussed while providing some suggestions for future directions.

The Organizational Change Process

Generating intentional organizational change through policy or practice reform or implementation is a complex and often lengthy process that requires extensive planning and training for those involved. If even one department within an organization is not ready for change, implementation efforts will likely fail. Key to success is staff willingness and ability to adopt and implement an innovation or change into their existing routine. Major challenges associated with generating organizational change include facilitating change within the existing organizational culture/climate, helping staff understand and make sense of proposed changes, and overcoming staff resistance. To date, prior research on organizational change within correctional agencies and institutions has paid scant attention to change processes among correctional workers (e.g., correctional officers, parole/probation officers). That is, there is considerably more literature about workers during the change process within non-correctional industries. However, relatively little is known about how correctional workers perceive, understand, adapt, adopt, implement, and, at times, resist changes and reforms within their workplaces.

Workers and Change in Correctional Settings

Why Study Change in Correctional Settings?

Historically, US correctional goals have oscillated between rehabilitative and punitive ideals.

In the period before the late 1960s, much of American corrections focused on offender rehabilitation. However, due to a growing concern that rehabilitation was ineffective at reducing recidivism and producing better post-custody citizens, corrections shifted toward a more punishment-oriented approach with an eye on retributive and deterrence goals. Throughout these shifts, the significance and function of prisons – as a part of the larger justice system – have also changed over time with restructured and reinterpreted goals and practices in response to the temporal problems and issues that interest society at any time. At present, financial strain has again shifted correctional goals toward reducing prison populations and accompanying costs with an outcome goal of recidivism reduction. In line with the growing weight of accountability, many correctional agencies are coupling their rehabilitation goals with evidence-based practices (EBPs) that use cognitive-behavioral approaches *and* risk management strategies. The resulting redesigned correctional models incorporate both rehabilitation and punishment creating a complex new goal structure that requires correctional organizations to undergo change.

Successful EBP implementation of new or revised practices and procedures requires a change in daily operations and a shift in business mentality. Prior work on change suggests successful implementation rests in the presence of three key components. These include (1) motivated and committed leadership, (2) organizational development involving the reworking of mission statement and values, resulting in an organizational culture transformation, and (3) collaboration between criminal justice agencies and supporting service providers, public safety agencies, and other stakeholders. Within these components, one of the biggest challenges facing organizations when implementing EBPs is altering the existing organization in ways that successfully support the new practice. In corrections, this may mean changing the organization's core ideology and goals from a focus on authority and control to more therapeutic, rehabilitative environments. For example, Simon (1993) discusses how parole has changed over time from a "disciplinary parole" model that focuses on punitive and controlling methods, "clinical parole" which focuses on individualized treatment, to "managerial parole" which focuses on achieving internal efficiency. Individual workers rely on workplace culture and climate to make sense of the shifting organizational emphases. That is, as parole changes, successful implementation means that workers also shift their understandings of their duties and the parolees they supervise. If this does not occur, change is unlikely.

Existing Research on Change in Correctional Settings

Research examining organizational change in correctional settings, focusing on staff-level variables throughout the change process is limited. Table 1 displays an overview of the studies reviewed that examine the experience of correctional staff throughout the change process. Although a systematic literature search was not completed, preliminary findings presented in this entry offer evidence of scarce attention to correctional workers in the change process relative to what is available regarding other criminal justice workers, namely, police. In the sections that follow, research is presented on correctional workers during the change process. This work is imparted in accordance with the three themes noted previously: culture/climate, sensemaking/perception, and resistance.

The Change Process in Corrections: Culture and Climate

Perhaps one of the most elusive yet monolithic concepts in organizational studies is culture or climate. Defined as the marriage of an organization's formal and informal structures combined with organizational actors and both intra- and interorganizational contingencies, an organizations' culture or climate combines ideals, practices, routines, goals, norms, and influences. This unique mix creates a normative environment where organizational actors both develop and adhere to formal and informal cultural guidelines. As an important facilitator of or barrier to organizational change, organizational culture

Correctional Workers in the Organizational Change Process, Table 1 Overview of literature on correctional workers in the organizational change process

Author/year	Description	Main topic area(s)	Correctional workers	Content/process	Change outcome	Suggestions how to improve change outcome
Battalino et al. (1996)	Case study on a change approach, meant to create attitudinal changes (away from a fear-based environment) in a correctional system	Org.al culture Staff perceptions	Prison staff and medical staff from prison colony and a female institution	Content	Preliminarily improved performance, networking Change goal well known, but misunderstood Cynicism Lack of adoption in some sites	Strategic assessment and tightly link change to org.al goals Structured learning models Expect conflicting inputs and prepare to address them
Ekland-Olsen and Martin (1988)	Detailed case study of administrative responses to litigated reform efforts directed at the TX department of corrections	Staff resistance Org.al culture/climate Perceptions	Prison administrators and prison guards	Process	Administrators controlled information, used political connections, made judicial counterattacks to resist change Prison staff perceived reform as illegitimate Prison staff violated law	Decrease degree of isolation from broader community Involve prison personnel in change Expect reforms
Farrell et al. (2011)	Examines the relationships between org.al factors and use of evidence-based practices in juvenile correctional agency	Staff cynicism Perceptions Org.al climate	Juvenile probation officers and supervisors	Content and process	Lower staff cynicism for change, more favorable perceptions of supervisory leadership and greater integration with community-based service providers were related to greater use of service-oriented practices Climate was not significantly related	Clarify mission of juvenile justice Address factors that make staff question the values and leadership of org. Resources, training, team meetings, on-site work teams important
Ferguson (2002)	Implementation of a risk and needs assessment tool as part of daily practice	Loss of discretion Staff resistance Staff perceptions	Adult probation department	Process	Improved workload management Systematized assessment More time on high-risk cases	Strong change commitment necessary Need committed resources Listen to staff Expect challenges
Lawrence and Johnson (1990)	Examine the shift from indeterminate sentencing to imposition of sentencing guidelines	Staff attitudes Role conflict Perceptions	Adult probation officers	Process	Role conflict found between PO role in PSI, plea negotiations, working under "just deserts" philosophy Perceive sentencing guidelines as ineffective POs held negative views of guidelines	

Lerch et al. (2011)	Examines the impact of a continuous on-site training model to advance the implementation of evidence-based practices in correctional settings	Org.al culture/ climate Resistance	Correctional officers Case managers	Content and process	Eventual improved climate by second year follow-up No improvement in org.al commitment Noncustodial staff shifted to use of more positive communication strategies Minimal change in communication strategies among custodial staff	Pair formal policy change with continuous, persistent, and consistent training Invest time and energy to change Expect a gradual, difficult change process Focus attention on improving org. commitment
Palumbo et al. (1984)	Examination of the Oregon Community Corrections Act to determine what factors are key in achieving successful achievement of the goals of the Act	Perceptions Org culture Resistance	Upper-level administrators (i.e., judges) Street-level administrators (i.e., POs)	Process	Wide variation in way legislation is implemented and degree to which goals are achieved General policy goals are met, but specific statutory goals are not More successful counties have POs with higher levels of program commitment, ability to meet local needs, and a higher degree of external support	Decrease differences in commitment and support between upper-level and street-level implementers Improve perceived job security of street-level workers Attempt to decrease lack of support for community corrections
Pullen (1996)	Evaluation of the reasoning and rehabilitation cognitive skills development program involving observations and questionnaires	Staff attitudes Program Implementation	Seven different sites, juvenile probation officers	Content and process	Program not adequately implemented Content was delivered, but knowledge and skills not imparted to offenders Lack of offender improvement	Need org.al infrastructure to support program Need support by administration Revise training approach, reallocate resources, monitor program, and quality control Make program more suitable for target audience Select proper POs to deliver services
Rudes et al. (2011)	Assesses custodial and noncustodial staff in prison reentry facility during change processes	Culture Resistance	Adult prison staff	Process	Staff unready for change Staff unaware or uninterested in work and change processes Staff follow guidelines inconsistently	More attention to pre-change culture and context Change takes a long time; longitudinal (mixed method) design

(*continued*)

Correctional Workers in the Organizational Change Process, Table 1 (continued)

Author/ year	Description	Main topic area(s)	Correctional workers	Content/ process	Change outcome	Suggestions how to improve change outcome
Steiner et al. (2011)	Assesses parole officers' perceptions of transition to a sanction grid	Staff resistance Staff cynicism Loss of discretion Staff perceptions	Adult parole officers	Process	POs dissatisfied with loss of discretion PO perceptions of implementation and intentions most influential in shaping views of reform	Include intent of reform in training Increase involvement of line staff in change
Turpin-Petrosino (1999)	Examined whether parole decision making complied with the 1979 New Jersey Parole Act, which attempted to limit parole discretion through mandates	Discretion Resistance	Adult parole release hearing boards	Process	Discretion and decision making has not been modified Parole boards continued to make decisions based on crime factors (plea bargaining, crime aggravation, type of crime) Board was inclined to deny parole in majority of cases even without appropriate evidence	Make legal standards more explicit to decrease policy uncertainty Structure rather than dismantle discretion Provide clear policies and procedures outlining how to use discretionary decision making without violating objective standards

represents both the "way things *get* done" and the way things become entrenched.

Research within correctional settings suggests that culture may assist with change implementation under certain circumstances. For example, evidence-based practice reforms occur more often in correctional organizations that present open learning environments with a focus on performance. Likewise, correctional workers who give primacy to treatment quality and are committed to their organization are more likely to use evidenced- based, effective treatment practices. Additionally, correctional staff who embrace cultural shifts toward rehabilitation are generally more likely to achieve successful cognitive-behavioral change.

In an examination of juvenile probation officers, Farrell et al. (2011) find that officers working in organizations with lower levels of staff cynicism for change and more favorable perceptions of supervisory leadership are more likely to use the service-oriented practice innovation their organization was implementing. Similarly, greater organizational integration with community-based service providers also relates to the greater use of service-oriented practices among juvenile probation officers (POs). Farrell et al. (2011) found no significant relationship between office climate or the perception that the organization is open to new ideas and innovations, and the use of the innovation.

Likewise, other features of the organization may also matter. Hepburn and Albonetti (1980) found that role conflict among correctional staff in a maximum-security prison did not determine whether staff fulfilled treatment or custodial roles. However, role conflict was predictive according to organizational goals. Culture also links to other factors important in organization change such as readiness for change and resources. In one study of correctional culture, Rudes et al. (2011) find that culture impedes correctional change when staff in an adult reentry facility undergoing change report low readiness for change. Likewise, prior work suggests that the amount of resources organizations provide (e.g., wraparound services – additional complimentary services in conjunction with current treatment), often aligns with organizational cultural norms. That is, organizational workers spend resources (e.g., money, time) on activities and outcomes that best align with existing organizational cultural norms.

In recent years, a focus on what criminology scholars refer to as "The New Penology" provides another way of considering organizational culture and the metaphoric pendulum that scholars' discuss swinging back and forth between the oppositional justice goals of rehabilitation and incapacitation. The new penology presents a slow change to broad organizational culture that moves community corrections agencies, like parole, from a focus on rehabilitation and/or law enforcement toward a new focus on people processing, waste management, and actuarial management systems. In this framework, correctional workers' objectives are "to sort and classify, to separate the less from the more dangerous, and to deploy control strategies rationally" (Feeley and Simon 1992: 452). Combined with statistically formulated classification schemes and the use of intermediate sanctions, the new corrections' approach focuses on "managing and recycling selected risk populations" (Feeley and Simon 1992: 465) in a way that minimizes risk while setting aside community reintegration goals.

Finally, studies measuring organizational *climate* via change processes have done so through surveys that assess the degree to which employees perceive their organization as supportive of new ideas and open to change. An organization's climate can signal to employees what is expected of them and potential consequences for different behaviors. Scott and Bruce (1994) found that employees who perceived their organization as supportive of innovations were more likely to support and implement new ideas and strategies themselves. Similarly, Friedman et al. (2007) found that organizational climates more conducive to learning relate to greater use of evidence-based practices.

All told, scholars consider organizational culture or climate as a paradox that explains both everything and nothing within organizational environments. The rich, contextual nature of organizational normative ordering does provide some leverage for explaining how and why organizational change processes succeed or fail. However, culture and climate are excessively difficult to quantify, study, and understand for culture is only as valuable as the meaning it holds to the individuals that access it.

The Change Process in Corrections: Perceptions/Sensemaking

Organizational cultures are not intrinsic. Instead, staff play a large part in how cultures both develop and endure over time. As such, staff are an important part of change, reform, and implementation efforts within organizations. Part of the role staff plays in the change process includes their understanding or perception of the current culture alongside how and why that culture is changing. Staff sensemaking often has enormous bearing on the success of implementation efforts.

Grounded in the works of past symbolic interactionist scholars and more recently, Weick (1995) and Gioia and Chittipeddi (1991), sensemaking involves a continual, interpretive process individuals use to manage disruptions to existing ideologies, routines, and activities. While not confined to organizations, sensemaking is often the principal means workers use for negotiating workplace change, innovation, and reform. Sensemaking, then, is a framework for workers to process new information against existing information to construct and reconstruct meaning. Conceptually, sensemaking intimately entangles with organizational change, particularly in individual workers' response. As such, sensemaking is a "springboard" "where meanings materialize and inform action" (Weick et al. 2005: 409). In studies of organizations, researchers often consider sensemaking among top and middle managers. At the management level, supervisors deduce change and then distribute their interpretations to subordinates. However, street- or line-level workers make sense of policy and practice changes, too. In many cases, low-level workers act as policy interpreters. They do this by drawing upon their own experiences, beliefs, and workplace ideologies. In turn, their sensemaking influences their decisions and actions.

Worker sensemaking and perception rest in their understanding of current and changing goals, values, and practices and are rooted in their prior experiences and personal characteristics. Previous research on correctional staff suggests individuals with higher education, and advanced degrees are more likely to implement innovations. Other studies find that educated staff are more likely to view evidence-based therapies positively while also favoring the use of complex medically assisted treatments (e.g., buprenorphine and methadone) in combination with counseling.

Akin to research on employee sensemaking and perceptions of current and changing goals, values, and practices of the organization, prior studies have examined how individual-level factors (such as race, gender, tenure) affect the staff perceptions and attitudes. Specifically, qualitative research on gender differences among correctional officers (COs) finds that females deal with inmate confrontations using interpersonal communication skills that possibly exert a "calming effect" on inmates. Common perception of male COs, on the other hand, is that they are more capable of successfully carrying out the responsibilities of a CO, even when supervising female inmates. More generally, female COs are often believed to be less supportive of punishment and more supportive of a human services orientation as a means to reduce future crime. It is important to note, however, that there is a smaller, competing body of research that argues female COs are actually more strict, authoritative, and confrontational in their interactions with inmates.

Previous research has also found a link between race and perceptions, with several researchers reporting that white COs display more punitive attitudes, while nonwhite COs display attitudes aligned more with treatment and rehabilitation. Older correctional workers are often found more supportive of treatment and rehabilitation as opposed to younger COs. Along the same times, COs who have had a longer tenure on the job also tend to be less punitively oriented and more supportive of rehabilitation and treatment ideals. Finally, COs with more education typically hold more positive perceptions about the inmates they supervise.

While not directly linked to the perceptions of change, it is crucial to understand possible factors that influence the perceptions and attitudes of correctional workers. Positive attitudes and behaviors in general link to positive outcomes. These include (1) positive views of inmates and rehabilitation, (2) more pleasant relations between CO and inmates, and lastly (3) more positive conditions within the facility.

Although limited in number, several studies have considered correctional worker perceptions of a change approach within their organization. In one study of corrections' workers, Steiner et al. (2011) surveyed parole officers to understand their perceptions of an agency transition to a sanction grid that imposed graduated sanctioning guidelines on parole officers. Survey results demonstrated a variety of different PO perceptions. One major PO concern was that the sanction grid placed restrictions on their discretionary decision-making abilities. These officers were less supportive of the innovation, overall. POs generally felt left out of the process used for choosing the sanction grid for their organization. They did not feel like anyone heard, or wanted to hear, their ideas or perceptions. In contrast, POs who felt respected as professionals and understood the intent and purpose of the sanction grid held more favorable perceptions of the innovation.

Within treatment settings – corrections' organizational cousin – there is a large body of research examining organizational change. This research is important when discussing correctional workers as one of the goals of corrections includes a rehabilitative/treatment component. Examining the challenges associated with the change process in the treatment field provides insight into challenges correctional workers face that are further complicated by the coexisting opposing goals of corrections (incapacitation vs. rehabilitation). Staff training is again an important theme in this literature, as the more relevant and engaging the training, the more likely staff are to have positive perceptions and attempt to actually implement the innovation into their work. Greater treatment provider support for change is possible when staff perceive an actual need for the improvement as well as the opportunity for professional growth. Similarly, when staff members perceive leadership or the organizational culture positively, they are more likely to associate positive attitudes toward change. Treatment counselors are more open to change when they have greater confidence in their own skills and abilities, as well as when the benefits of the innovation are reinforced. Likewise, more educated treatment counselors are also more likely to accept change. Rowan-Szal et al. (2007) also report that allowing treatment staff to participate in the change process and help in the planning stage results in more positive perceptions and a greater willingness to implement the innovation into practice. A significant barrier to change within treatment settings includes a lack of time, specifically when serving a large number of clients. A perceived divergence between usual care and the proposed change can also hinder implementation efforts within treatment settings.

Aside from personal and professional characteristics affecting perception and sensemaking, sometimes correctional workers oppose change more strongly than just misunderstanding. In these cases, staff resist change.

The Change Process in Corrections: Resistance

Just as culture/climate and sensemaking/perception are assistive with organizational change, they also enhance, support, and facilitate worker resistance. Staff regularly resist change when they experience difficulty reconciling prior organizational goals and ideologies with new ones. As the natural opposite to organizational control, resistance represents contested power relations within an organizational structure whereby workers exhibit behavior that does not align with management or organizational change goals. At a formal level, resistance emerges in organizational life via strikes, collective action, grievance filing, and work stalling or stopping. At the more prevalent, yet less visible, informal level, workplace resistance is what some scholars dub “routine resistance.” This type of resistance is generally unplanned and is sometimes covert. Although informal resistance is more difficult to study because it is virtually imperceptible, it is here that mundane, everyday resistance takes shape. It is also here that workers make steady progress toward retarding or halting organizational change.

In correctional settings, the complexity of the change process is due in part to the well-documented and unfailing tension between two competing organization goals – treatment (rehabilitation) and security (incapacitation). Scholars note this robust goal rift in prisons, parole and probation. On a correctional ideology continuum, staff often connect to one of the two goals more closely. When change or reform proposes movement from one side of the continuum toward the other, correctional staff feel strain and try to find ways to reconcile or eliminate it.

How and why staff resist change and reform closely intertwines with how their organizational culture/climate and sensemaking frameworks align with the change. In this regard, prior work such as Turpin-Petrisio’s (1999) research on parole decision making–suggests that correctional workers, namely, probation/parole officers, often continue employing pre-change behavioral structures and related decisions despite reform efforts. In these studies, POs perceive reform as ineffective when it does not first consider their needs, prior decision-making patterns, or when they are less committed to their organization. Related work by Latessa (2004) notes the importance of how managers and supervisors communicate the change to individual workers in ways that help workers understand how the change will positively affect their work.

More specifically in correctional settings, Ferguson (2002) examined the experience of a probation department as they implemented a risk/needs assessment tool as part of their daily routine. While the department eventually experienced success with integrating the assessment tool into practice, significant challenges arose during the change process. One major hurdle was staff perceptions of the innovation. Staff were concerned that the use of this tool meant that they were losing their discretionary decision-making ability and no longer had any input when it came to writing presentence recommendations or case plans. Although this was a major challenge, trainers acknowledged this issue and focused on the idea that the structured assessment tool, combined with their professional judgment, is critical to an effective assessment. There was also some staff resistance to change, as workers expressed concern that as soon as they adjusted to the current change, more changes would follow. Staff also worried about managing their workload once the new change became routine. Trainers addressed this concern by emphasizing the risk/needs tool as a new way of looking at offender risk and needs as opposed to creating additional work. Akin to Latessa’s (2004) suggestion for managers, trainers also explained how the tool assists POs in doing their jobs more effectively.

In another corrections study, Battalino et al. (1996) emphasize the difficulty that comes with generating change within a public organization, the California Department of Corrections (CDC) prison system. In their evaluation, the authors examine the implementation process of the Treatment of People Initiative (TOP), the goal of which is to transform the organizational culture of this correctional system from a more fear-based work environment to one that promotes constructive and supportive interactions and values the contributions of each employee (Battalino et al. 1996). The TOP initiative met staff resistance as employees felt it was impossible to operate within a correctional setting without using an authoritarian style. Some staff members believed that they had been using many of the TOP values already and felt that the current push for the initiative was an insinuation that they had done something wrong or that their organization had failed. Staff also perceived the change initiative to be “soft” and in opposition to the context of the existing organizational culture. This study on change within corrections highlights the complexity of generating change within a large, bureaucratic organization that traditionally operates with a competing goal structure (Battalino et al. 1996). It also indicates several triggers for staff resistance. Ultimately, the TOP initiative was unsuccessful in the CDC as the change was only attitudinal without linking to on-the-ground behavioral change.

Finally, Pullen (1996) evaluated the implementation of the reasoning and rehabilitation (R&R) cognitive skills development program in a Juvenile Intensive Supervision Probation (JISP) department. The RR program, a mandatory component of JISP, is an educational program provided by POs to teach cognitive skills such as problem-solving and critical reasoning to juvenile offenders in groups of six to eight. POs received training on delivering the R&R program by ensuring they understand the cognitive skills they would be teaching and helping POs to prepare lesson plans, overheads, and other class materials for the session. Through survey and observational data, Pullen (1996) found unsuccessful implementation of the R&R program. POs received the information given to them in the training, but did not develop the actual skills needed to run the program. POs often prepared right before the group session or had to do so at home, reporting that there was no decrease in other job responsibilities in order to make time for their new responsibilities. Pullen (1996) reported that POs delivered the content of the material but often read directly from the manual suggesting that they did not have a complete understanding of the material. POs also lacked enthusiasm for the program, focusing on the fact that it was mandatory rather than making it relevant and interesting to the juveniles (Pullen 1996). Pullen (1996) concluded that the shortfalls in the implementation of the R&R program can undermine the effectiveness of the program and cognitive skill development overall.

Resistance of many forms, from foot dragging to monkey wrenching, exists within organizational environments, and corrections does not escape this reality. Although workers may not admit or even allude to direct forms of resistance, resistance often occurs indirectly through lack of buy-in and commitment, feelings of mistrust, and questions about how and why reform is necessary and occurring.

Summary of What Is Known About Change Among Correctional Workers

A common theme throughout these studies is that staff try to understand change while working within the framework of their own perceptions and their organizational culture/climate. They become resistant to change when they cannot align prior goals and values with new ones or when they perceive a loss of their discretionary decision-making capabilities or their ability to use professional judgment. For some staff, implementing change is burdensome and difficult to do while balancing already busy schedules, work, or caseloads. At these times, change seems not only unrealistic but also impossible.

Change progress, on the other hand, is possible when staff members understand the intent and purpose of the innovation. The research by Battalino et al. (1996) highlights a challenge unique to correctional settings in that correctional workers are resistant to altering the organizational culture to be less authoritarian because they felt it did not align with the goals and context of the organization as a whole. This research reveals that staff perceptions regarding how a change initiative will affect them personally, how it is implemented in their daily routines, and how it fits into their current understanding of the organization they are embedded in are influential in shaping their views concerning the efficacy of the reform and openness to implementing and adopting it.

Improving Change Processes and Outcomes Among Correctional Workers: Better Ways of Studying Change

Given what is known about organizational change, Van de Ven and Poole (2005) suggest a typology of complimentary approaches for studying change that considers organizations as both things (nouns) and processes (verbs). In building their typology, the authors suggest that prior work on organizational change operates according to Mohr's (1982) distinction between variance and process. In this thinking, organizational scholars generally study organizations in one of two ways: First, they consider how organizations vary (using variance theories) (Mohr 1982) looking at variables and causation. Second, they focus on organizational processes (using process theories), assessing historical and sequential narratives. Van de Ven and Poole (2005) formulate a four-part typology that suggests a way of expanding current knowledge by incorporating both process and variance theories into studies of organizations, that is, considering organizations as both a noun and a verb. In this framing, scholars would simultaneously consider differences between what Barnett and Carroll (1995) call *content* and *process* within organizational change. This would infinitely broaden the scope of current understanding of organizational change by considering the select dimensions within organizations that are changing, while also taking account of how and why the change develops and unfolds.

What Is Missing? Next Steps

While prior studies of organizational change highlight organizational culture/climate, staff perceptions/sensemaking, and resistance as key contingencies for change, prior research does not adequately address the specific complexities present within correctional agencies. Culture and perception are both critical factors affecting implementation, particularly when staff consider the compatibility of a change with existing practice. They may look for the value added or may look for ways to make both old and new policies or practices work simultaneously. While some prior work establishes that staff education level and prior experience affect willingness of staff to accept reform efforts, little is known about the processes staff go through to perceive, understand, and, at times, resist workplace change. Prior studies also do not address how staff think or perceive reform in environments where staff are not well educated – which is common in correctional organizations such as prisons.

Specifically, frontline staff are a pivotal part of the change process as they are often principally responsible for policy and practice implementation. At present, the limited knowledge of change processes among correctional workers does not include much scholarship on staff perceptions as part of the overall organizational culture and as a pathway toward or from resistance. This entry presents information on the perceptions and attitudes of correctional workers generally. One hypothesis might include a question about how perceptions and attitudes regarding treatment and rehabilitation align with openness and staff perceptions of change. This, though, is just one area warranting additional attention.

That attention requires forethought. The history of scholarship on organizational change among correctional workers and within the broader organizational literature builds upon studies that answer the "what" question (e.g., what change occurred and what were the causes and effects of the change). However, more studies that use a longitudinal design to answer the "how" question are needed. These studies will explore process issues that consider the temporal and contextual richness of change processes and environments. Of course, these studies, mostly qualitative in nature, come with specific challenges including labor intensity, stretching researcher capacity and resources, and (when qualitative) concerns about generalizability, reliability, bias, and validity. However, with appropriate and transparent sampling, coding, and analysis strategies, the yields from studies of this nature far outweigh the challenges.

Another possibility for advancing research on correctional workers in the change process involves a different lens for examining workers role in change. Martha Feldman (2000, 2003) is nearly alone in her work on mindfulness with organizational environments, specifically within change processes. Whereas much of the prior work on change suggests that agency-less workers engage in change processes by mindlessly accepting it, apathetically ignoring it, or overtly or covertly resisting it, Feldman (2000, 2003) suggests that workers use *mindful* actions that are a source of, rather than a barrier to, change. Future work on change processes may consider this conceptualization as a framework for studying change. Doing so may require a grounded theory approach and an analysis strategy that allows for exploring workers' perceptions and behaviors using nontraditional lenses.

Another possible route for advancing change work in corrections comes from Iverson (1996). He suggests that change acceptance increases with organizational commitment, harmonious climate, education, job motivation, job satisfaction, job security, and positive affectivity. He also finds decreased support for organizational change accompanied by union membership, role conflict, tenure, and environmental opposition. Although prior work considers many of these factors, there is little research about the role union membership plays in correctional change processes except Page's (2011) book on the California Correctional and Peace Officers Union (CCPOA), which does not explore change processes directly.

Finally, scholars of organizational change processes must consider increasing focus on organizational communication via knowledge creation process through "continuous dialogue between tacit and explicit knowledge" (Nonaka 1994: 14) and through middle managers' "role as key interpreters and sellers of strategic change. At both the street- and management-levels, communication is a conceptual partner to perception/ sensemaking and culture/climate. Horizontal and vertical information flow within organizations and among key and non-key organizational actors presents a critical mechanism for understanding how and why change occurs or fails to do so. Studies that consider networks, knowledge transfer, and organizational learning would improve knowledge of how communication works within change processes.

Conclusion

One obvious unstated point in this entry on correctional workers in the organizational change process is the importance of viewing change through both a criminological (social science) and organizational lens. As Pettigrew and colleagues (2001) note, "Dynamism has been difficult to study, and social science has developed quite comfortably as an exercise in comparative statics" (697). What is needed in studies of organizational change processes are research projects that do the harder work of formulating designs and collecting/analyzing data that speak to the historical and processual changes that occur within organizational contextual environments. Examining culture/climate, perception/ sensemaking, and resistance is not an easy undertaking. However, the payoff for improving knowledge about how change works among correctional workers is large. This knowledge can vastly progress practice or policy implementation through improved training programs – affecting both content and design. Organizations who know more about how their workers understand and cope with change can also do more to work with workers on related issues that could be resistance points if unknown. Finally, considering not just the independent and dependent variables but thinking through the interactional dynamics of organizational change provides a crucial link for understanding how and why workers make the choices they make during organizational change processes and how to make those choices work in the organization's best interest.

Related Entries

▶ History of Corrections
▶ History of Probation and Parole in the United States
▶ Innovation and Crime Prevention
▶ Managing Innovation of Policing
▶ Organizational Change and Police Legitimacy
▶ Prison Warden Stress and Job Performance
▶ Probation and Parole Practices
▶ Probation Officer Decision-Making

Recommended Reading and References

Barnett WP, Carroll GR (1995) Modeling internal organizational change. Annu Rev Sociol 21:217–236

Battalino J, Beutler L, Shani AB (1996) Large-system change initiative: transformation I progress at the California department of corrections. Public Prod Manag Rev 20(1):24–44

Ekland-Olsen S, Martin SJ (1988) Organizational compliance with court-ordered reform. Law Soc Rev 22(23):59–384

Farrell JL, Young DW, Taxman FS (2011) Effects of organizational factors on use of juvenile supervision practices. Crim Justice Behav 38(6):565–583

Feeley MM, Simon J (1992) The new penology: notes on the emerging strategy of corrections and its implications. Criminology 30(4):449–474

Feldman MS (2000) A performative perspective on stability and change in organizational routines. Ind Corp Chang 12:727–752

Feldman MS (2003) Organizational routines as a source of continuous change. Organ Sci 11:611–629

Ferguson JL (2002) Putting the "what works" research into practice. Crim Justice Behav 29(4):472–492

Friedmann PD, Taxman FS, Henderson CE (2007) Evidence-based treatment practices for drug-involved adults in the criminal justice system. J Subst Abus Treat 32(3):267–277

Gioia DA, Chittipeddi K (1991) Sensemaking and sensegiving in strategic change initiation. Strateg Manag J 12(6):443–448

Hepburn JR, Albonetti C (1980) Role conflict in correctional institutions. Criminology 17(4):445–460

Iverson RD (1996) Employee acceptance of organizational change: the role of organizational commitment. Int J Hum Resour Manag 7(1):122–149

Latessa EJ (2004) The challenge of change: correctional programs and evidence-based practices. Crim Public Policy 3(4):547–560

Lawrence R, Johnson SL (1990) Effects of the Minnesota sentencing guidelines on probation agents. J Crime Justice 13(2):77–104

Lerch J, Viglione J, Eley E, James-Andrews S, Taxman FS (2011) Organizational readiness in corrections. Fed Probat J 75(1):5–10

Mohr LB (1982) Explaining organizational behavior. Jossey-Bass, San Francisco

Nonaka I (1994) A dynamic theory of organizational knowledge creation. Organ Sci 5(1):14–37

Page J (2011) The toughest beat: politics, punishment, and the prison officer's union. Oxford University Press, Oxford

Palumbo DJ, Maynard-Moody S, Wright P (1984) Measuring degrees of successful implementation: achieving policy versus statutory goals. Eval Res 8(1):45–74

Pettigrew AM, Woddman RW, Cameron KS (2001) Studying organizational change and development: challenges for future research. Acad Manag J 44(4):697–713

Pullen S (1996) Evaluation of the reasoning and rehabilitation cognitive skills development program as implemented in juvenile ISP in Colorado. Division of Criminal Justice, Boulder

Rowan-Szal GA, Greener JM, Joe GW, Simpson DD (2007) Assessing program needs and planning change. J Subst Abus Treat 33(2):121–129

Rudes DS, Lerch J, Taxman FS (2011) Implementing a reentry framework at a correctional facility: challenges to the culture. J Offender Rehabil 50(8):467–491

Scott SG, Bruce RA (1994) Determinants of innovative behavior: a path model of individual innovation in the workplace. Acad Manag J 37(3)

Simon J (1993) Poor discipline: parole and the social control of the underclass, 1980–1990. The University of Chicago Press, Chicago

Steiner B, Travis LF, Makarios MD (2011) Understanding parole officers' responses to sanctioning reform. Crime Delinq 57(2):222–246

Turpin-Petrosino C (1999) Are limiting enactments effective? An experimental test of decision making in a presumptive parole state. J Crim Justice 27(4):321–332

Van de Ven AH, Poole MS (2005) Alternative approaches for studying organizational change. Organ Stud 26(9):1377–1404

Weick K (1995) Sensemaking in organizations. Sage, Thousand Oaks

Weick KE, Sutcliffe MK, Obstfeld D (2005) Organizing and the process of sensemaking. Front Organ Sci 16(4):409–421

Corrections

▶ Desistance and Supervision

Corruption

Michele Riccardi[1] and Federica Sarno[2]
[1]Transcrime, Joint Research Centre on Transnational Crime, Università Cattolica del Sacro Cuore di Milano, Milan, Italy
[2]Transcrime, Joint Research Centre on Transnational Crime, Universitá Cattolica del Sacro Cuore di Milano and Universitá degli Studi di Trento, Milan, Italy

Synonyms

Bribery; Graft

Overview

Ancient philosophers and poets described corruption as part of the human behavior and therefore as almost endemic in every community. However, it was only in the 1990s that corruption began to attract the interest of scholars, governments, international organizations, and NGOs. Since then, a far-reaching debate has developed on the issue, giving rise to a large number of academic publications and political programs.

The most important issues of this debate are examined in this entry. In particular, section “Defining and Classifying Corruption” provides a review of the numerous definitions and classifications of corruption adopted in the literature. Section “Measuring Corruption” discusses the main difficulties in measuring corruption, while section “Understanding Corruption” reviews the cultural, institutional, and economic models developed by researchers to interpret the causes and the effects of corruption. Section “Experiencing Corruption” provides a summary of corruption experiences across regions and sectors, and section “Policing Corruption” finally suggests policy and regulatory strategies with which to prevent and combat corrupt practices worldwide.

Defining and Classifying Corruption

Before analyzing the various aspects of corruption, it is necessary to define the phenomenon. Quite frequent in the literature is the definition of corruption as *the abuse of power for private gain*. This definition implies the existence of three elements – namely, *power*, *abuse*, and *private gain* – that are worth discussing in detail.

Firstly, a precondition for corruption is the existence of a person who is somehow entrusted with discretionary power, i.e., the authority to allocate scarce benefits or to provide restricted goods or services (Rose-Ackerman 1978). Secondly, an abuse, i.e., a violation of the role assigned to this entrusted person, must be committed. Finally, the abuse of the role must be intended to obtain some kind of private benefit, i.e., an advantage for the entrusted person or a third party (Rose-Ackerman 1978). To be noted in this regard is that the nature of the *private gain* may vary (UNODC 2011): it may be tangible (e.g., a house) or intangible (e.g., a career advancement) and it may take the form of monetary payment (e.g., a bribe in cash) or in-kind compensation (e.g., food and drink, sex, preferential treatment, a job).

However, the above definition is still very broad. In an attempt to specify it better, researchers have often resorted to certain operational classifications which have proved fruitful and are now summarized.

Active Versus Passive Corruption

The first classification regards the distinction between active and passive corruption. Active corruption refers to the *supply* side of the corrupt transaction, i.e., the party who offers, promises, or gives an undue advantage to the person entrusted with discretionary power so that the latter abuses his/her role (e.g., a company promising a bribe to a public official in order to be selected in a public bidding process). By contrast, passive corruption concerns the *demand* side, i.e., the party who requires, receives, or accepts the undue advantage (in this case, the public official).

This classification is not intended to identify who instigates the act (the corruptor) (European Commission 2011: 6). With reference to bribes, for example, the “initiative for bribing can originate with the bribe giver or the bribe taker,” the latter case being often regarded as extortion (Transparency International 2006: 16). To be noted, moreover, is that in many criminal justice systems, both behaviors are punished.

Public Versus Private Corruption

Depending on whether the entrusted power is abused by a public official (e.g., a politician or a civil servant) or a private party (e.g., a company executive), it is possible also to distinguish respectively between public corruption and private corruption.

Private corruption involves only individuals operating in the private sector (“private-to-private” corruption), while public corruption is generally defined as the *abuse of public office for private gain* (Tanzi 1998: 564). In this sense, depending on whether the public official is a civil servant or a politician, public corruption can be further distinguished between *administrative* (or *bureaucratic*) and *political* corruption.

Administrative corruption generally affects the legal implementation of laws, regulations, and/or policies, while political corruption mainly concerns their formulation. Political corruption frequently involves the electoral cycle, implying, for example, practices of buying/selling votes and/or affecting election funding (Transparency International 2006). On the other hand, in administrative corruption cases, civil servants may act "according to the rules" (e.g., they receive a bribe to accelerate an authorization process) or "against the rules" (e.g., they are corrupted to break the law) (Pope 2000).

Grand Versus Petty Corruption

The third operational classification refers to the extent of the corrupt practice: *grand* corruption generally occurs at the highest level of government, affecting the decision-making process; and, as such, it usually involves large sums of money. By contrast, *petty* corruption usually occurs at the administrative level. This is, for example, the case of a low-ranking public official who accepts or extorts a bribe of a small amount from an individual (European Commission 2011).

To be noted is that *petty* corruption may be indicative of broader and more dangerous *grand* corruption schemes occurring at higher levels of governance (Pope 2000: XIX). Moreover, it should be stressed that the amount of money paid is not necessarily related to the extent of damage caused by corruption. As pointed out by many authors, small bribes may have big consequences (Rose-Ackerman 2003).

Some other authors (see, e.g., Tanzi 1998: 565) do not distinguish between administrative and petty corruption and between political and grand corruption. Moreover, further operational classifications exist, referring, for example, to the spread of corrupt acts in a certain community (*structural* vs. *situational* corruption), to its frequency (*isolated* vs. *systemic*), or to its nature (*cash* vs. *in-kind* corruption, see above). Consequently, it is difficult to reach universal consensus on how to define corruption. This consideration also applies to the *legal* definitions of corruption.

Legal Definitions

Contrary to the categories analyzed so far, legal definitions of corruption do not provide broad classifications, but rather focus on identifying those corruption-related criminal acts (such as bribery, embezzlement, fraud, unfair competition) which can be recognized as criminal offenses and hence prosecuted.

It should be accordingly emphasized that finding a *legal* definition of corruption with a maximum of consistency across countries is very difficult, given that different labels and headings are adopted in different jurisdictions (Jehle and Harrendorf 2010: 160). This is all the more the case of *private* corruption: in fact, while public corruption exists as a criminal act in almost all jurisdictions, private corruption often appears under different labels (e.g., unfair competition).

Many conventions have been adopted at regional and international levels in an attempt to provide a legal framework within which to prevent and prosecute corruption. However, the scope of their provisions varies greatly. Some of them (i.e., the UN Convention against Corruption and the Council of Europe's Criminal Law Convention on Corruption) cover a broad range of offenses extending from active and passive briberies of foreign and national officials to embezzlement, illicit enrichment, obstruction of justice, and trading in influence. Others (such as the OECD Convention on Combating Bribery of Foreign Public Officials in the International Business Transactions) are instead highly specific in their scope, in that they criminalize only certain kinds of corrupt practices (i.e., the *promising*, *offering*, or *giving* of a bribe to *foreign* public officials) (OECD 2008).

Moreover, not all the provisions are mandatory. Some of them are optional, which means that signatory states have only to take account of their potential implementation (OECD 2008: 14). As a consequence, some jurisdictions do not establish certain corrupt practices as specific criminal offenses and therefore do not even prosecute them.

Measuring Corruption

Having defined corruption and in order to gain better understanding of its causes and consequences, it is necessary to quantify its extent. But measuring corruption is not an easy task. Firstly, what is measured depends exactly on how corruption is defined; and as seen above, various classifications are possible. Moreover, there are a number of methodological challenges which apply not only to *direct* measurements, such as administrative statistics, but also to more *indirect* measurements of corruption such as experts' assessments or victimization surveys. These problems are now reviewed.

Administrative Statistics

Firstly, corruption, like other non-conventional crimes, is a concealed form of behavior difficult to detect. Administrative statistics, such as police or prosecution data on cases of corruption, enable the most direct measurement of the phenomenon, but they represent only a fraction of the actual magnitude of the crime because many offenses are not discovered or not reported to the relevant authorities (the so-called dark number).

The dark figure of corruption is usually higher than for other offenses. It is so for two main reasons. Firstly, in comparison with more conventional crimes (e.g., burglary, theft, robbery), victims of corruption are less likely to report the offense, either because they perceive these behaviors as normal practices and are reluctant to modify them or because they themselves are somehow involved in the offense (e.g., they have paid a bribe requested by a public official) (UNODC 2009). Secondly, in comparison with other crimes, the boundaries between illicit and licit conducts may not be easily distinguished, so that it is difficult for a victim to understand whether or not a behavior should be reported (UNODC 2009; Recanatini 2011).

Cross-country measurements of corruption are even more problematic since, as noted above, there is no common definition of corruption across different jurisdictions (Jehle and Harrendorf 2010: 160). For example, only half of the 42 European countries covered by the European Sourcebook of Crime and Criminal Justice Statistics could provide police and prosecution data on public corruption (Aebi et al. 2010: 83), and the way in which this offense is defined differs markedly from country to country (Aebi et al. 2010: 34). Besides legal definition problems, also cultural factors (e.g., the differing likelihood that citizens will report offenses to the police) and statistical ones (e.g., the statistical counting rules used to collect crime data) often hamper cross-country measurements of the phenomenon.

Indirect and Alternative Measures

Owing to all these problems, official data on reported cases of corruption do not provide a reliable picture of the actual magnitude of the phenomenon. Hence, alternative and indirect measures of corruption should be considered, such as composite indicators, sample surveys, and experts' assessments.

Composite Indicators

This approach combines different variables and information in a single index so as to provide a synthetic measure of corruption in a certain country or region. One of the most popular composite indicators is the CPI – Corruption Perception Index – developed by Transparency International, which combines different sources such as polls, experts' assessments, and media reports. It has been calculated on a yearly basis since 1995 and in 2012 covered 176 countries (http://cpi.transparency.org/cpi2012/). Other indicators are the Corruption Control Indicator (CCI), one of the six Worldwide Governance Indicators developed by the World Bank, or the Global Integrity Index (GII), produced by Global Integrity.

Composite indicators provide a synthetic measure of the corruption level and also make it possible to produce comparisons across countries and over time. However, they also have weaknesses. Firstly, they usually measure how corruption is *perceived* by a range of subjects (experts, public officials, researchers, journalists, the public) rather than the actual level of corruption itself. Secondly, they combine information from different sources referring to different topics (e.g., corruption, governance, press freedom, company transparency), with the consequence that it is sometimes difficult to understand what exactly these indicators measure. Finally, the construction of these indicators entails a number of methodological choices (e.g., selecting sources, selecting how to weight the average in computing the results) which increase the level of subjectivity of these measurements (UNODC 2009: 5).

Sample Surveys

Indirect measures of corruption are also provided by sample surveys. These may be carried out on a number of subjects, interviewed both as *witnesses* of corruption events and/or as *victims* (in so-called victimization surveys). Respondents can usually be distinguished into three categories: households, businesses, and public officials. While households are usually interviewed as victims of petty corruption practices, business surveys are addressed to business executives or employees, who are asked to report their experiences of corruption during their corporate activities. Civil servants or public officials are asked for information on various issues, including corruption in recruitment and promotion practices, job mobility, training frequency, work incentives, salary, and wages.

The main advantage of sample surveys is that, with respect to official statistics, they usually present a lower "dark number," since they are anonymous. The respondents are therefore less reluctant to provide information on their experiences of corruption than they are when dealing with the police. Surveys can therefore provide a more reliable picture of the actual extent of corruption in a country or sector.

Another strength is that they allow the collection of both quantitative and qualitative information. The former makes it possible, among other things, to compute *prevalence* rates in a certain region or business area (e.g., the share of a population of individuals and/or companies that paid a bribe in the previous year); the latter sheds light, for example, on the types of corruption (e.g., in cash, in kind; see section on "Defining and Classifying Corruption"), on the amounts of money involved (e.g., average percentage of the GDP per capita paid in bribes), and on the nature and characteristics of the victims or the offenders (e.g., gender, age, social status, job position).

However, sample surveys also have weaknesses. Firstly, given the reluctance of victims of corruption to report it (see section on "Administrative Statistics"), also victimization surveys have their own "dark figure." Consequently, the sample size should usually be larger than in other surveys measuring other types of offense. Secondly, certain types of corruption (e.g., embezzlement, unfair competition, abuse of function) do not generate individual victims, so that it is not possible to apply victimization surveys in these cases. Thirdly, the results may vary considerably depending on the sampling method adopted: for example, overweighting small and medium enterprises (SMEs) in business victimization surveys may result in higher corruption levels given the higher vulnerability to corruption of SMEs compared with larger companies (UNIDO and UNODC 2007: 7).

Despite these limitations, the importance of sample surveys in measuring corruption has been widely recognized, and in recent years various investigations have been specifically focused on this crime, such as UNODC's surveys on households and businesses in the Western Balkans, Afghanistan, and in several African countries (UNODC 2009: 7), the Crime and Corruption Business Surveys (CCBS), as well as the HEUNI study on the Finnish-Russian border.

Also to be mentioned are business victimization surveys which measure corruption together with other types of crime, such as the International Commercial Crime Survey (ICCS), the International Crime against Businesses Survey (ICBS), the EU Business Victimization Survey, the Swiss BCS, and the Mexican BCS.

Experts' Assessments

Another approach to measuring the *perception* of corruption is based on assessments by experts, who are asked to provide a quantification of the level of corruption (or of its trends) in selected regions or countries. Given its mainly qualitative nature, this methodology entails a high degree of subjectivity and a weak statistical significance. Nevertheless, it is an inexpensive and easy-to-perform instrument with which policy makers can obtain an overview of the phenomenon and collect further insights and suggestions.

Expert assessments are often performed by commercial risk rating agencies, consulting groups, media companies (e.g., Economist Intelligence Units) and international organizations, such as the World Bank (World Development Report and the Enterprise Survey), the WEF (Executive Opinion Survey), and the EBRD (Business Environment and Enterprises Performance Survey).

Understanding Corruption

Now that corruption has been defined and the main difficulties in its measurement have been described, this section focuses on the causes and effects of corruption through an overview of the main theoretical models and empirical studies developed by researchers in the field.

Models and Theories

Theoretical models used to explain corruption are usually drawn from economic theory or political sciences, and they may be considered as derivations of Becker's "crime and punishment" theory.

Generally speaking, these models envisage the coexistence of several elements: first, the presence of someone with discretionary power (see section on "Defining and Classifying Corruption"); second, the existence of economic rents associated with this discretionary power; and third, the presence of a system of sanctions in case of violations. More recently, further elements have been introduced, such as an information asymmetry and/or a lack of enforcement, which make it difficult to monitor persons with discretionary power.

Although a large number of theoretical models have been developed (for a review see Bardhan 1997), two main categories can be distinguished and are now described:

Principal-Agent Models

In principal-agent models, the principal delegates to an agent some discretionary power in allocating resources and/or administrating the regulation that the agent could use for his/her own interest rather than for the principal's sake. Classic examples are the electorate (principal) that elects a legislator (agent) or a city mayor (principal) who delegates discretionary power to a public official (agent).

In the case of corruption, a triple relationship can be identified among a principal, an agent, and a client (Klitgaard 1988). The principal gives the agent the power to pay a premium to a client who satisfies the principal's interest (e.g., the electorate delegates a public official to select the most efficient company in a public procurement procedure), but the agent may choose, in return for a bribe, an inefficient company in contrast with the principal's interest.

The agent may benefit from the high monitoring costs that the principal must bear or from some information asymmetry (e.g., the principal does not know the third party as well as the agent does). In a world of complete information, in fact, the agent could be fully monitored by the principal, and hence, corruption events could be avoided (Rose-Ackerman 1978).

Rent Allocation Models

Public intervention in markets may give rise to rents of different kinds: authorizations to operate a business, licenses to import goods, etc. Entrepreneurs, as rational agents, compete for these rents by devoting time and part of the firm's resources to the *rent-seeking* activity (Krueger 1974).

While competition for rents may be absolutely legal, it often takes illegal forms as corruption, bribery, and/or market abuse (e.g., the payment of a bribe to obtain a commercial license). A large number of studies have therefore identified and quantified the costs to the economy of rent-seeking activity (e.g., Krueger 1974).

Empirical Studies

Numerous studies have empirically tested the above-mentioned models in order to determine the causes and consequences of corrupt practices and the factors that influence them.

Advances in the measurement of the level of corruption (see section on "Measuring Corruption") have undoubtedly improved empirical analysis of the correlation between corruption and a range of explanatory variables. Most of these studies are cross-country investigations adopting composite indicators, such as Transparency International's CPI, as dependent variables, and other socioeconomic variables as regressors (see Lambsdorff 2006, for a review).

In this regard, it is important to stress that the use of composite indicators as dependent variables is subject to the methodological problems already described in section on "Measuring Corruption", i.e., the high degree of subjectivity incorporated in such indexes and their hybrid nature comprising not only corruption but also other patterns (e.g., corporate transparency, governance, press freedom). Moreover, even if a significant correlation between the level of corruption and other variables has been identified, it is often difficult to establish the direction of the causality, i.e., to understand whether these variables are the *causes* or *effects* of corruption. For example, there is a significant correlation between corruption indicators and GDP per capita, but it is not possible to distinguish clearly whether it is corruption that generates poverty or the latter that facilitates corrupt practices (Lambsdorff 2006).

The determinants and factors correlated with corruption can be divided into three main categories: moral-cultural factors, institutional factors, and economic factors.

Moral-Cultural Determinants

Researchers have firstly sought to understand whether the presence of certain values or norms in some cultures facilitate the acceptance of corrupt practices. For example, the custom of gift giving, either as a form of acknowledgment or as a lubricant, may be accepted in one country but unethical elsewhere (UNODC 2011: 12). Reed (1996: 396) identifies the culture of gift giving and the importance of personal networks as two drivers facilitating the spread of corruption in Japan. Networking also appears important in explaining the high corruption level in Italy, together with other cultural determinants such as the lack of trust in politics, the emergence of *business* politicians and organized crime (Della Porta 1996).

Also the level of trust within a society has been demonstrated to be closely correlated with corruption, in the sense that a lack of trust among bureaucrats, politicians, and private citizens may incentivize corrupt practices (La Porta et al. 1997). For example, as noted by many authors, where the public has a low opinion of politicians, corruption may be more tolerated because it is considered inevitable (Mény 1996: 319). In this sense corrupt practices appear to be more frequent in hierarchical societies (Husted 1999) or in ones dominated by religions which La Porta et al. (1997, p. 337) consider particularly hierarchical: the Catholic, Eastern Orthodox, and Muslim religions. Conversely, higher levels of civic engagement and of media freedom appear to be less correlated with the spread of corruption (e.g., Treisman 2000; Brunetti and Weder 2003).

Interestingly, a negative correlation has been found between corruption and the degree of involvement of women in the society (in the labor force, private and public institutions, etc.), which remains robust at a cross-country level even after controlling for employment status (Swamy et al. 2001).

Institutional Determinants

Institutional determinants of corruption are usually related to the structure or the characteristics of the political system in a given country or region.

According to the principal-agent theory, democracy should guarantee closer control over civil servants by the electorate, thus discouraging corrupt practices. A significant negative correlation between the level of democracy and corruption has been shown by Treisman (2000) in a cross-country analysis. Brunetti and Weder (2003) have confirmed this finding also in a time series dimension.

Key constraints on corruption are the power exerted by the parliament and by political parties. Kunicova and Rose-Ackerman (2005) evidence that parliamentarism, with respect to presidentialism, may provide disincentives to corruption. The same can be shown with respect to the strength of political parties, since weak political groups can facilitate the commission of corrupt practices.

Close attention has been paid to the role of voting systems, i.e., the ways in which an electorate selects and monitors its political delegates. Generally speaking, there is evidence that electoral systems entailing lower competition among politicians, e.g., systems characterized by smaller voting districts or the use of party lists (Persson et al. 2003) or by proportional representation (Kunicova and Rose-Ackerman 2005), are usually correlated with higher corruption levels.

Economic Determinants

However, most empirical studies focus on the economic determinants of corruption (for a comprehensive review, see Lambsdorff 2006).

As anticipated above, resource allocation models hypothesize that public intervention in the economy results in rent-seeking activities, which in their turn may incentivize corrupt practices. However, empirical findings provide little evidence of this relationship (see Lambsdorff 2006, for a review), and there are many exceptions to the model (e.g., Scandinavian countries, which are characterized by very low levels of corruption but large public sectors).

The focus has therefore shifted to the structure of the market: several authors have shown that the level of competition in an economy is negatively correlated with corruption. Corrupt practices are also less frequent in markets characterized by a high degree of openness to international investment and trade (Treisman 2000).

Corrupt practices also appear to be correlated with the level of wages in the public sector, in the sense that low salaries may induce civil servants to supplement them with bribes or illicit fringe benefits. Evidence in this regard has been provided by Klitgaard (1988).

Effects of Corruption

Besides the causes of corruption, attention has also focused on its effects. A small number of authors suggests that corrupt practices could provide benefits to economies, especially in "greasing the mechanisms" of a society and in avoiding the inefficiencies of bureaucracy.

However, most studies have concentrated on the negative consequences of corruption. Firstly, it harms the job market: in the presence of corrupt practices, human resources are allocated inefficiently and without meritocratic criteria. Corruption has also been demonstrated to be detrimental to investments in education (Gupta et al. 2002). Secondly, the ineffectiveness of public expenditure has been recognized as well (see Rose-Ackerman, for a review): corruption distorts the system of public procurements, so that it fails to minimize the costs and maximize the quality of the services provided to the community (e.g., corrupt public officials do not select the *best* company in public bids). Accordingly, corruption may also lead to an uncontrolled growth of public expenditure (Tanzi and Davoodi 1997).

Also distortions are negative results of corruption. For example, distortions in the allocation of public and private resources: corrupt policy makers may divert public resources to unproductive, profit-seeking (DUP) activities (e.g., large-scale infrastructural projects instead of investments in research, innovation, education) (Tanzi and Davoodi 1997). Secondly, distortions occur in the distribution of wealth: corrupt practices incentivize the redistribution of wealth to non-deserving subjects. Gupta et al. (2002) provide evidence of a strong correlation between corruption and increased inequality within the population. High corruption rates may also spur the growth of illegal activities (e.g., black marketeering) and discourage Foreign Direct Investments (FDI), because corruption in the recipient country has proved detrimental to investments by foreign countries and multinationals (Wei 2000).

Experiencing Corruption

While section "Understanding Corruption" reviewed theories and empirical studies so as to understand the causes and the effects of corruption, this section provides an overview of corruption experiences across countries and markets.

Corruption Across Countries

In the 1960s and the 1970s, corruption was perceived as a "disease" mainly affecting developing countries. Most early empirical and theoretical studies focused on emerging regions (Krueger 1974; Bardhan 1997, for a review), demonstrating that corruption was a major obstacle to their development and, as described in section on "Understanding Corruption," a cause of inequality and the ineffective allocation of financial resources and wealth distribution (Tanzi and Davoodi 1997; Gupta et al. 2002).

Attention has been also paid to communist and post-communist countries. Researchers have often hypothesized that communism provided structural incentives for corruption (Sandholtz and Taagepera 2005) which the transition to market economies has not been able to dismantle. On the other hand, the privatization process in these countries during the 1990s offered numerous opportunities for corrupt practices (Rousso and Steves 2006).

Developing and post-communist countries are invariably ranked in the lowest positions by corruption composite indicators such as Transparency International CPI. In these countries, corruption is often perceived as the most severe problem after poverty and unemployment (UNODC 2011: 7). Consequently, for the past 50 years, local governments, international organizations, and NGOS have pursued two objectives: the development of mass programs to eradicate corruption (see section on "Policing Corruption") and the conduct of large-scale surveys to collect more detailed measurements and information (especially concerning "petty" corruption).

As regards the latter, surveys have been carried out by UNODC in several African countries (for a review, see UNODC 2009: 7) and by the World Bank in African and Central and South American countries (see Recanatini 2011, for a review). As regards communist and post-communist countries, studies by UNODC have investigated corruption practices in Afghanistan, Albania, Bosnia and Herzegovina, Croatia, Kosovo, Montenegro, Serbia, and FYR Macedonia (UNODC 2011). Close attention has also been paid to Russia and, more recently, China (see, e.g., Sandholtz and Taagepera 2005).

However, Western countries are not immune to corruption. In the 1980s and the 1990s, major scandals involving cases of "grand" corruption (such as the Italian *Tangentopoli* (Bribesville) scandal of 1992–1995) induced researchers to study corrupt practices in Western Europe, highlighting the relationship between corruption and transition processes – for example, in Spain after the dictatorship (Heywood 1997) or in Germany after reunification (Seibel 1997) – or between corruption and organized crime, as in Italy (Della Porta 1996).

Corruption has always been central to the political history of Japan, given that 9 of the 15 Prime Ministers between 1955 and 1993 were involved in corruption scandals (Nelken and Levi 1996: 2). Researchers have made great efforts to explain the spread of corruption in Japan, often citing cultural determinants to do so.

The analysis of corruption in the United States has often considered episodes of "grand" corruption during the prohibition era in the 1920s and political scandals in the 1970s and 1980s. More recently, corruption cases have been studied in the context of such 1990s and 2000s financial scandals as Enron, MCI-Worldcom, or the subprime crisis.

Corruption Across Sectors and Markets

Corruption has many facets. In the same country, it may be a major problem in one sector (e.g., public procurements) and a minor problem in another (e.g., public education). It would therefore be erroneous to consider it as uniformly distributed in a community. Composite indicators provide synthetic measures of corruption, but they do not help identify which sectors are severely affected and which are not.

In fact, the perception of corruption may vary radically within the same community. For example, a survey in Guatemala showed that the police were believed to be "honest" by only 10 % of the households in the sample, while universities were considered "honest" by 60 % of those households and the church by nearly 80 % (Recanatini 2011: 45). In the Western Balkans, the local government is regarded by the population as at most risk of corruption, while the military sector is not (UNODC 2011). Again, a survey conducted in several countries of Africa and Latin America showed that "petty" corruption was the main concern in some countries, such as Guinea and Sierra Leone, while nepotism and purchase of public positions were in others, such as Paraguay (Recanatini 2011).

On adopting a market perspective, it can be hypothesized that, as anticipated in section on "Understanding Corruption," a higher risk of corruption exists in activities entailing a greater degree of intervention by the government or the regulator. Corrupt practices are widespread wherever licenses or authorizations are necessary to operate a business. This is particularly the case in the import–export sector or when foreign direct investments (FDI) are involved. Bribes of millions of USD have been paid by multinationals to local governments in order to obtain authorizations to invest (Tanzi 1998) and corruption scandals involving well-known companies, such as Siemens, have led to massive anti-corruption programs.

Public procurement is probably the area most exposed to corruption, both because of its high economic value (on average, around 10–20 % of GDP) and because of the crucial role played by public officials and local governments in that they often have a great deal of discretionary power (Ware et al. 2011). New regulations have been introduced in order to prevent corrupt practices, but there still persist gray zones where companies participating in public bidding processes can influence the competition or obtain secret information on rival tenders (Soreide 2005).

New businesses or markets are also susceptible to corruption. Recent studies have focused on green economy activities, which have recorded rapid growth in the past 10 years but also frequent cases of corrupt practices (e.g., in reducing emissions programs, forestry business, and wind power sector) (Caneppele et al. 2013).

Policing Corruption

To conclude, there follow considerations on how corruption could be combatted by policies and regulations. As shown above, corruption is a complex and multi-faceted phenomenon. Anti-corruption policies therefore need to be tailored to the specific weaknesses of a region or a business area (UNODC 2011: 13).

In accordance with resource allocation models (see section on "Understanding Corruption"), many anti-corruption policies seek to reduce public intervention in the economy and to increase the degree of openness in the market as means to abate the incentives for corruption (Rose-Ackerman 2003). Deregulation programs have proved effective in reducing corruption in some countries: for instance, Singapore, where corruption in customs procedures decreased after duty-free imports were extended (Bardhan 1997: 1335). Likewise, the introduction of competitive pressure among US police officials reduced corruption in the drug market (Bardhan 1997: 1337). In some cases, however, deregulation policies and competitive strategies may not be feasible or desirable solutions: for example, in the supply of essential goods or services in poor countries (Bardhan 1997: 1335).

As described in section on "Understanding Corruption," low salaries in the public sector may encourage corruption. Reforms in the wage structure for civil servants or a premium system for virtuous officials are recognized as important strategies for fighting corruption effectively (Rose-Ackerman 2003; Bardhan 1997).

From an administrative point of view, the creation of independent bodies and anti-corruption agencies specifically aimed at preventing and/or prosecuting corruption is a pillar of many anti-corruption programs. Among others, Hong Kong's Independent Commission Against Corruption has been widely recognized as able to eradicate corrupt practices at the highest level of government. However, when replicated in other countries, the Hong Kong model has not produced the same positive results. In particular, it seems that anti-corruption agencies are less effective where corruption is more systemic (Rousso and Steves 2006: 247–248).

Enhancing public awareness of corruption and promoting a "culture of integrity" (i.e., through anti-corruption campaigns), in both the public and the private sectors, are other crucial measures. However, to obtain concrete results (i.e., the disclosure of corrupt practices by the employees of a company), it is necessary to put effective accountability mechanisms (e.g., whistleblower legislation and protection systems) in place (Rose-Ackerman 2003; Bardhan 1997).

Further tools for use in fighting corruption, such as the application of anti-money laundering regulation or the confiscation of assets resulting from corrupt practices, should also be considered.

Related Entries

- ▶ Police Corruption
- ▶ Transnational Corporate Bribery

Recommended Reading and References

Aebi M et al (2010) European sourcebook of crime and criminal justice statistics, vol 241, 4th edn, Onderzoek en beleid series. Boom Juridische uitgevers, Ministry of Justice, Research and Documentation Centre WODC, Den Haag

Bardhan P (1997) Corruption and development: a review of issues. J Econ Lit 35(3):1320–1346

Brunetti A, Weder B (2003) A free press is bad news for corruption. J Public Econ 87(7–8):1801–1824

Caneppele S, Riccardi M, Standridge P (2013) Green energy and black economy: mafia investments in the wind power sector in Italy. Crime Law Soc Change 59(3). doi:10.1007/s10611-013-9418-1. http://link.springer.com/article/10.1007/s10611-013-9418-1

Della Porta D (1996) Actors in corruption: business politicians in Italy. Int Soc Sci J 48(149):349–365

European Commission (2011) Supporting anti-corruption reform in partner countries concepts, tools and areas for action. Concept Paper No. 2, Tools and Methods Series. European Union, Luxemburg

Gupta S, Davoodi H, Alonso-Terme R (2002) Does corruption affect inequality and poverty? EconGov 3:23–45

Heywood P (1997) From dictatorship to democracy: changing forms of corruption in Spain. In: Della Porta D, Mény Y (eds) Democracy and corruption in Europe, Social change in Western Europe series. Pinter Pub, London

Husted BW (1999) Wealth, culture and corruption. J Int Bus Stud 30(2):339–359

Jehle JM, Harrendorf S (eds) (2010) Defining and registering criminal offences and measures – standards for a European comparison. Göttingen Stud Crim Law Justice 10

Klitgaard R (1988) Controlling Corruption. University of California Press, Berkeley

Krueger AO (1974) The political economy of the rent-seeking society. Am Econ Rev 64(3):291–303

Kunicova J, Rose-Ackerman S (2005) Electoral rules and constitutional structures as constraints on corruption. Br J Polit Sci 35(4):573–606

La Porta R, Lopez-de-Silanes F, Shleifer A, Vishny RW (1997) Trust in large organizations. Am Econ Rev Pap Proc 137(2):333–338

Lambsdorff JG (2006) Causes and consequences of corruption: what do we know from a cross-section of countries? In: Rose-Ackerman S (ed) International handbook on the economics of corruption. Edward Elgar, Cheltenham/Northampton

Mény Y (1996) 'Fin de siècle' corruption: change, crisis and shifting values. Int Soc Sci J 48(149):309–321

Nelken D, Levi M (1996) Corruption of politics and the politics of corruption: an overview. J Law Soc 23(1):1–17

OECD (2008) Corruption. A glossary of international standards in criminal law. Organisation for Economic Co-operation and Development, Paris

Persson T, Tabellini G, Trebbi F (2003) Electoral rules and corruption. J Eur Econ Assoc 1(4):958–989

Pope J (ed) (2000) TI sourcebook 2000. Confronting corruption: the elements of a national integrity system. Transparency International, Berlin

Recanatini F (2011) Assessing corruption at a country level. In: Graycar A, Smith RG (eds) Handbook of global research and practice in corruption. Edward Elgar, Cheltenham

Reed SR (1996) Political Corruption in Japan. Int Soc Sci J 48(149):395–405

Rose-Ackerman S (1978) Corruption: a study in political economy. Academic, New York

Rose-Ackerman S (2003) Corruption. In: Rowley CK, Schneider F (eds) Encyclopedia of public choice, vol 1. Springer, Dordrecht

Rousso A, Steves F (2006) The effectiveness of anti-corruption programs, preliminary evidence from the post-communist transition countries. In: Rose-Ackerman S (ed) International handbook on the economics of corruption. Edward Elgar, Cheltenham

Sandholtz W, Taagepera R (2005) Corruption, culture and communism. Int Rev Sociol 15(1):109–131

Seibel W (1997) Corruption in the federal republic of Germany before and in the wake of reunification. In: Della Porta D, Mény Y (eds) Democracy and corruption in Europe, Social change in Western Europe series. Pinter Pub, London

Soreide T (2005) Grey zones and corruption in public procurement: issues for consideration. In: OECD (ed) Fighting corruption and promoting integrity in public procurement. OECD, Paris

Swamy A, Knack S, Lee Y, Azfar O (2001) Gender and Corruption. J Dev Econ 64(1):25–55

Tanzi V (1998) Corruption around the world. causes, consequences, scope, and cures. IMF Papers, vol 45, No. 4. International Monetary Fund

Tanzi V, Davoodi H (1997) Corruption, public investment, and growth. IMF working paper no. 139

Transparency International (2006) Handbook: curbing corruption in public procurement. Transparency International, Berlin

Treisman D (2000) The causes of corruption: a cross-national study. J Public Econ 76(3):399–457

UNIDO, UNODC (2007) Corruption prevention to foster small and medium-sized enterprise development.Providing anti-corruption assistance to small businesses in the developing world, vol 1. United Nations Industrial Development Organization and the United Nations Office on Drugs and Crime, Vienna

UNODC (2009) Quantitative approaches to assess and describe corruption and the role of UNODC in supporting countries in performing such assessments. Background paper prepared by the secretariat, Nov 2009. United Nations Office on Drugs and Crime, Doha

UNODC (2011) Corruption in the western Balkans: bribery as experienced by the population. United Nations Office on Drugs and Crime, Vienna

Ware GT, Moss S, Campos JE, Noone G (2011) Corruption in procurement. In: Graycar A, Smith RG (eds) Handbook of global research and practice in corruption. Edward Elgar, Cheltenham

Wei SJ (2000) How taxing is corruption on international investors. Rev Econ Stat 82(1):1–11

Cost-Benefit Analysis

Matthew Manning
School of Criminology and Criminal Justice,
Griffith University, Mt. Gravatt, QLD, Australia

Overview

The use of economic analysis (EA) in crime and justice research has increased in recent years from the need to quantify the effectiveness and efficiency of interventions and policies. The New Welfare Economics model pursues an economic decision rule known as the *Kaldor-Hicks criterion*. This rule forms the underlying rationale for cost-benefit analysis (CBA). In short, CBA is used to evaluate options (e.g., interventions, programs, or policies) by comparing the total benefits with the total costs of the respective options. Results of EA assist in informing policy on the economic feasibility of the options in question. This entry briefly summarizes the underlying conceptual foundations of CBA and why CBAs are important to policy decision-making from a societal perspective. To illustrate the use of CBA, this entry provides two examples of CBAs conducted of third-party policing interventions highlighting potential limitations of the applied method.

Fundamentals

Importance of Economic Analysis

Economic analysis (EA) is increasingly used as a regulatory and policy development tool to assist decision-making in the area of crime and justice. Since the 1980s, there has been a requirement that major regulatory initiatives (in the USA) incorporate cost-benefit studies to supplement the noneconomic evidence, for example, Executive Order 12291 promulgated by President Regan and, more recently, Executive Order 12866 issued by President Clinton. Factors including political ideology and limited fiscal and personnel resources inevitably affect policy decisions. However, the inclusion of EA provides salient information with respect to the economic impact of decisions on individuals and more broadly society (Boardman et al. 2006).

EAs provide decision-makers with information about the most economically efficient policies and programs by assessing comparative costs and benefits (Manning et al. 2011). Such information provides an economic rationale for the allocation of scarce public resources. Further, EAs promote economic efficiency and good fiscal management by assessing available options to ensure the greatest return on investment. EAs also allow policy makers to judge the economic implications of existing policies and/or programs.

Key Economic Concepts

Studies that observably lack the important economic concepts that form the foundation of economic analysis are problematic. These studies tend to be exercises in accounting rather than EAs. The important economic concepts include Pareto efficiency, potential Pareto efficiency, willingness to pay, and opportunity costs.

Pareto Efficiency

Boardman et al. (2006) explain that "…an allocation of goods is Pareto efficient if no alternative allocation can make at least one person better off without making anyone else worse off" (p. 26). That is, we achieve an efficient allocation of resources when one person is made better off, but not at the cost of another. Figure 1 illustrates this concept by splitting a fixed amount of money between two people. The vertical axis measures the amount ($) received by person 1, and the horizontal axis the amount received by person 2. The sloping line represents the potential Pareto frontier. The frontier represents all feasible combinations of allocations between the two people. If the two agree, each person receives the full amount as indicted by the intersection on the x and y axes ($1,000 each). However, if they do not agree, then they receive only $250, which represents the status quo (A). The shaded triangle represents Pareto improvements over the status quo. Therefore, the shaded triangle illustrates all the alternative allocations that make at least one person better off (than the status quo) without making the other worse off. However, Pareto improvements are not Pareto efficient, as there is the possibility of further Pareto improvements. Only when the potential Pareto frontier is reached is it not possible to make further reallocations (Manning 2008).

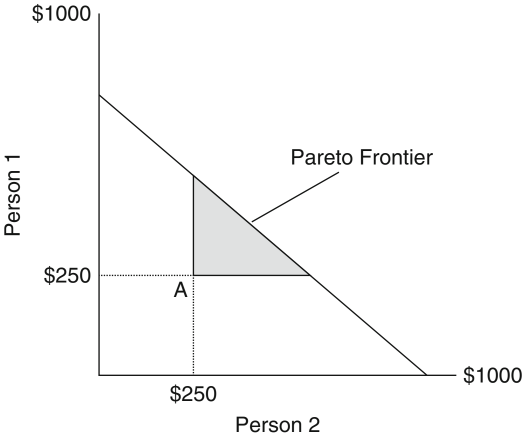

Cost-Benefit Analysis, Fig. 1 Pareto efficiency. Adapted from: Boardman et al. (2006)

Kaldor-Hicks Criterion

Adopting the strict Pareto efficiency rule for CBA requires that policies are only adopted if they yield positive benefits and that full compensation or a set of transfers can be made that makes at least one person better off without making anyone else worse off. This requires analysts to value all impacts in terms of, for example, willingness to pay and value inputs with respect to opportunity costs. This ideal holds that there are only winners and no losers. As appealing as this is, difficulties arise given (1) the burden on analysts to measure aggregate costs and benefits inferred from observing prices and the quantification of market and nonmarket goods, (2) the unfeasibility of valuing all the costs and benefits to each person affected, (3) the difficulty in identifying the administrative costs of policy and the pragmatic and ethical limitations of identifying and allocating potential transfers or compensation models, and (4) the incentive for individuals to cheat by overstating the costs and understating the benefits they would receive from policies (Boardman et al. 2001; Mishan 1972). If this criterion were adopted, the result would be forgoing most policies with potential net benefits to society at large and too much effort placed on identifying transfers or compensation for losers.

Consequently, the framework of the old economic doctrine (the neoclassical Cambridge School) gave way to New Welfare Economics model arguing that CBA utilize an alternative decision rule with possibly less conceptual appeal but more practicability. Known as the

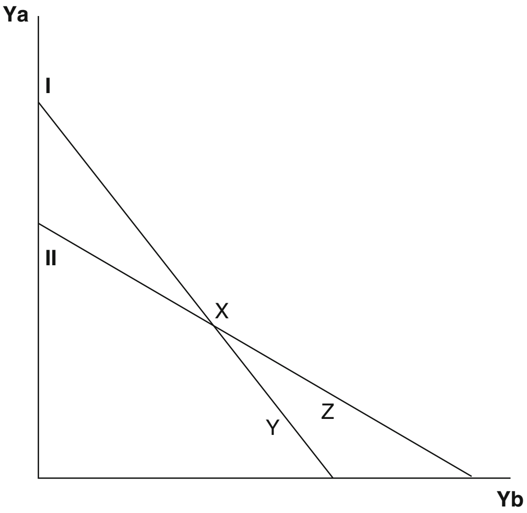

Cost-Benefit Analysis, Fig. 2 Kaldor-Hicks improvements. Adapted from: Brent (2008)

Kaldor-Hicks criterion, or the net benefits criterion, a policy is *potentially* Pareto efficient and should be adopted if there are positive net benefits. Moreover, as long as there are net benefits, then it is possible to compensate losers so that the policy could be Pareto improving. In short, the key difference between the Pareto efficiency and the *Kaldor-Hicks* criterion is with respect to compensation, so that transfers can be made to those who are worse off under the policy. Given this, every Pareto improvement is a *Kaldor-Hicks* improvement, but the majority of *Kaldor-Hicks* improvements are not Pareto improvements. Figure 2 provides an example of the Pareto improvement process. Budget line I is the income before the policy/intervention and budget line II is income after the policy/intervention. The figure demonstrates that the allocation for x and y are not comparable as there is more Ya at x and more Yb at y. However, if income is redistributed along II from x to z, there is more Ya and Yb at z than at y. Therefore, point z is a potential Pareto improvement (i.e., both could be made better off). In summary, under the Pareto improvement criterion, there must be positive net benefits to both participants and that no one is made worse off under the new policy.

Given the unlikelihood for a policy to make no one worse off, under the *Kaldor-Hicks* criterion, compensation can be made to gain net benefits. For example, person 1 receives a utility of ten (arbitrary score) and person 2 a utility of 100. If the new or proposed policy changes the utility received from person 1 and 2 to 20 and 99, respectively, the change would not be Pareto efficient as person 2 is now worse off. However, a *Kaldor-Hicks* improvement is achieved as person 1 could theoretically compensate person 2 (between one and ten utils) to accept the situation. A full review of the *Kaldor-Hicks* criterion including a thorough critique of the compensation test and the distribution of income is available in Brent (2008).

The *Kaldor-Hicks* criterion forms the underlying rationale for CBA. For example, CBA is used to evaluate a project or policy by comparing the total costs of the program with the total benefits. Thus, if benefits exceed costs, then the program would normally go ahead. This is effectively an application of the *Kaldor-Hicks* criterion because it is equivalent to requiring that the benefits received by winners should be enough to compensate the losers. The *Kaldor-Hicks* criterion is used for decision-making because (1) policies with net social benefits maximize aggregate wealth; (2) a redistribution of a normal good indirectly helps those who are worse off in society because richer societies have a greater capacity for redistribution; moreover, *ceteris paribus*, if redistribution is a normal good and because people want more as their wealth increases, then their capacity and willingness to help may increase; (3) government activity will have different sets of winners and losers, but the costs and benefits of different policies may average out over the full set of government activity for each person; and (4) an equal distribution of wealth or income may be possible through direct transfers after a number of “efficiency-enhancing” policies are adopted—transfers can be done wholesale as opposed to retail (Boardman et al. 2001; Zerbe et al. 2006).

Willingness to Pay

Calculating the net benefits associated with a policy requires estimating inputs and outputs.

Willingness to pay enables the modeling of how a person values the outputs associated with a policy. Through a series of questions, information about the payments required or potentially received by a person is elicited, so that they become indifferent between what is already occurring (status quo) and the policy in question (with payments). In short, willingness to pay identifies how much (i.e., dollar amount) a person is willing to give up to acquire a good or a service. For example, consider a policy that is relevant to three people. Person 1 says that they would be indifferent between the policy and status quo at $1,000. Person 2 says that they would be indifferent between the policy and the status quo at $2,000. However, person 3 does not like what the policy offers so they would have to receive a payment of $2,500 (−$2,500) to feel just as well off under the status quo. The sum of the willingness to pay values (persons 1, 2, and 3) is a measure of the net benefits of the impact of the proposed policy. In this example, benefits equal $3,000 and costs equal $2,500. Therefore, net benefits of the policy are positive (+$500). This scenario, however, is not Pareto efficient, as person 3 is made worse off with respect to the status quo. But this could be altered if person 1 was to give $750 and person 2 was to give $1,750 to person 3. This is now a Pareto efficient solution as persons 1 and 2 are better off than the status quo (+$250) and person 3 is no worse off. The point is that if aggregate net benefits of a policy are positive for all concerned, then solutions exist to make the policy a Pareto improvement over the status quo (Boardman et al. 2001).

Opportunity Costs

Objectives are achieved through the use of inputs. Economists use costs in CBA to value inputs, assess their efficient use, or to assess the required inputs to produce or implement a policy or program (Boardman et al. 2006). Costs are seen as *opportunity costs*. That is, if the input is not used in one context, it could be used in another (Manning et al. 2006). In other words, the opportunity cost of using an input is its value in its best alternative use. The best measure of the opportunity cost of a resource is its market value or price (Drummond et al. 1987).

Returning to the example of three persons whose willingness to pay for a given policy was $500, if the opportunity cost of inputs to produce the policy were $750, then some people would have to forgo $750 for goods valued at this price. In this scenario, the policy does not produce adequate benefits to the three people to allow them to compensate those who must forgo the $750 ($500 in net benefits to the three persons minus $750 in opportunity costs = −$250). In this scenario, the opportunity cost is too high and does not produce enough benefits to compensate all. However, if the opportunity cost was $200, then net benefits in society would be positive ($300) and make it possible to compensate all those who bear a cost so that no one is made worse off, and some people are made better off under the policy (Boardman et al. 2001).

Common Tools Used for Economic Analysis

Four common methods used to assess the economic efficiency/effectiveness of program alternatives are cost-savings, cost-effectiveness, cost-utility, and cost-benefit analysis. A number of less rigorous economic methods, which provide useful information to policy makers, include cost analysis, cost-of-illness analysis, and cost-consequence analysis. The various economic tools address somewhat different decision-making questions, require different data (depending on the question and stakeholder), and represent different degrees of analytical sophistication (Manning et al. 2006). Table 1 provides a brief summary of cost-savings, cost-effectiveness, cost-utility, and cost-benefit analysis with respect to the measurement of costs and outcomes and strengths and weaknesses of the various approaches. A full discussion of the four main methods is provided in Manning (2008).

Discounting Costs and Benefits

After assessing costs and benefits, both must be discounted to allow for potential time differentials. For example, since we perceive costs incurred in the present to be less of a burden than costs

Cost-Benefit Analysis, Table 1 Common economic analysis techniques

Type of analysis	Measure of cost	Measure of outcomes	Strengths	Weaknesses
Cost-savings	Monetary value of resources	Monetary savings resulting from impact of intervention	• Good for assessing the savings generated to stakeholders	• Difficult to place monetary values on salient life benefits
Cost-effectiveness	Monetary value of resources	Units of effectiveness (e.g., reduced recidivism)	• Easy to incorporate standard evaluations of effectiveness	• Hard to interpret when there are multiple measures of effectiveness
			• Good for alternatives with small number of objectives	• Only useful for comparing two or more alternatives
Cost-utility	Monetary value of resources	Units of utility (e.g., improvements in social-emotional development, quality of life measures)	• Incorporates individual preferences for units of effectiveness	• Difficult to arrive at consistent and accurate measures of individual preferences
			• Incorporates multiple measures of effectiveness into single measure of utility	• Cannot judge overall worth of a single alternative
Cost-benefit	Monetary value of resources	Monetary value of benefits (full evaluation of all direct, indirect benefits—including a discussion of intangible outcomes where they cannot be monetized)	• Can judge absolute worth of a project	• Difficult to place monetary values on salient life benefits
			• Can compare CB results across a variety of projects	

Adapted from: Levin and McEwan (2001)

incurred in the future (Manning 2004), future costs must be discounted to properly compare them with present costs. One calculates alternative investment patterns by calculating their present value. The formula for estimating the present value of future costs is presented in Eq. 1:

$$P = \sum_{n=1}^{n} Fn(1+r)^{-n} = \frac{F1}{(1+r)} + \frac{F2}{(1+r)^2} + \frac{F3}{(1+r)^3} \qquad (1)$$

where $P =$ present value, $F_n =$ future costs at year n, and r = the discount rate. The choice of discount rate has always been controversial but three conceptual approaches are available: (1) returns to consumer savings options (cash rate) (Levin and McEwan 2001), (2) average returns to investment made by the private sector (Boardman et al. 1996), and (3) weighted average of methods 1 and 2 (Levin and McEwan 2001). A full discussion of these methods is provided in Manning (2008).

Net Present Value

When assessing the economic rationale of a program or policy, Manning et al. (2006) propose that evaluators measure the costs and potential benefits using net present value (NPV). NPV equals the difference between the present value of benefits relative to the present value of costs. In short, the NPV criterion is simple: *adopt the program if the NPV is positive*. Specifically, one would adopt a project if the NPV = present value (benefits) – present value (costs) > 0 (Boardman et al. 2006).

NPV has a caveat; NPV does not necessarily identify the most efficient allocation of resources. Rather, NPV identifies the *more* efficient allocation of resources. For example, an intervention for which program alternatives vary with respect to potential output Q is considered in Fig. 3. The benefits and costs associated with alternatives are represented by the functions (BQ) and (BC), respectively. Moving from Q_0 to Q^* increases efficiency, but at a decreasing rate. That is, $NPV(Q^*) > NPV(Q_2) > NPV(Q_1) > NPV(Q_0)$. However, as we move beyond the

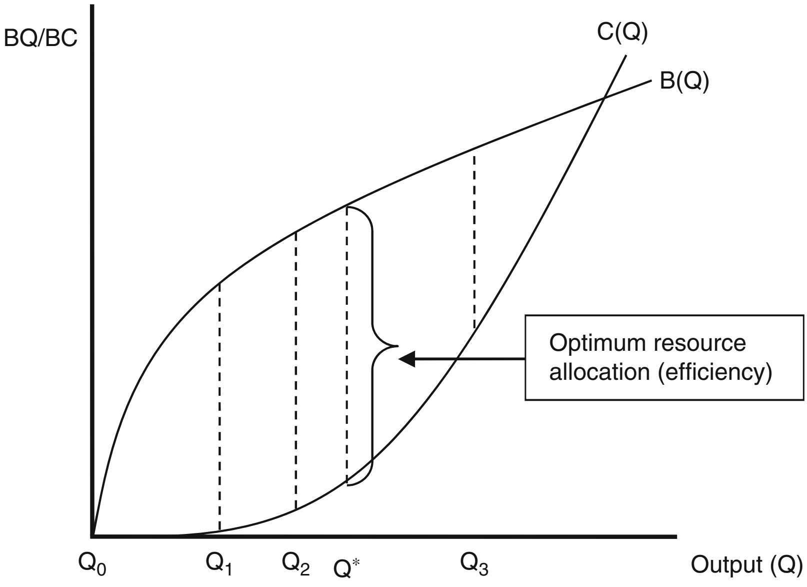

Cost-Benefit Analysis, Fig. 3 Efficient resource allocation. Adapted from: Boardman et al. (2006)

optimal scale (Q^*), the net benefits decrease. Net benefits remain positive until the point where the curves C (Q) and B (Q) intersect. The problem is that Q* may not have been selected as it was not included in the set of evaluated options. Therefore, NPV identifies the more efficient alternative among the group of alternatives evaluated but does not identify the most efficient alternative. The optimum alternative (Q*) may have been omitted because of bounded rationality problems (i.e., cognitive limitations of both knowledge and cognitive capacity) (Manning et al. 2011), the optimum level may not have been known, or budgetary or political constraints may have limited the range of alternatives considered for evaluation. This considered, however, NPV is still an appropriate method as it provides accurate answers with respect to the alternatives that are evaluated (Boardman et al. 2006).

Incremental Cost-Comparison Analysis

Any economic analysis should include an incremental cost-comparison analysis of the available options. The term incremental cost analysis refers here to a change in the scale of activity; in other words, the difference in cost or effect between two or more options being compared.

Figure 4 provides a diagrammatic representation of the incremental cost curve, where $IC_{a'}$, Q_1 is the cost of option 1 at quantity (activity level) Q_1, and $IC_{b'}$, Q_1 is the equivalent estimate for option 2. The incremental difference (cost) between option 1 and option 2 at quantity (activity level) Q_1 is IC_{a-b}, Q_1.

Figure 4 highlights that the incremental cost is equal to the change in total cost divided by the change in quantity. This may appear to be a marginal cost; however, the incremental cost is distinguished from the marginal cost by holding Q constant at $Q = 1$ (Q_1). Therefore, the alternatives (activities) are compared at $Q = Q_1$.

The incremental cost is represented mathematically in Eq. 2:

$$IC = \Delta TC/\Delta Q$$

where

$$TC = \sum_{i=1}^{n} (Pi \times Qi) \tag{2}$$

Note, IC = incremental cost, Q = quantity/activity, Δ = change in activity, TC = total cost, P_i = price of inputs (derived from Table 3') and Q_i = quantity of inputs.

Cost-Benefit Analysis, Fig. 4 Incremental cost analysis. Adapted from: Boardman et al. (2006)

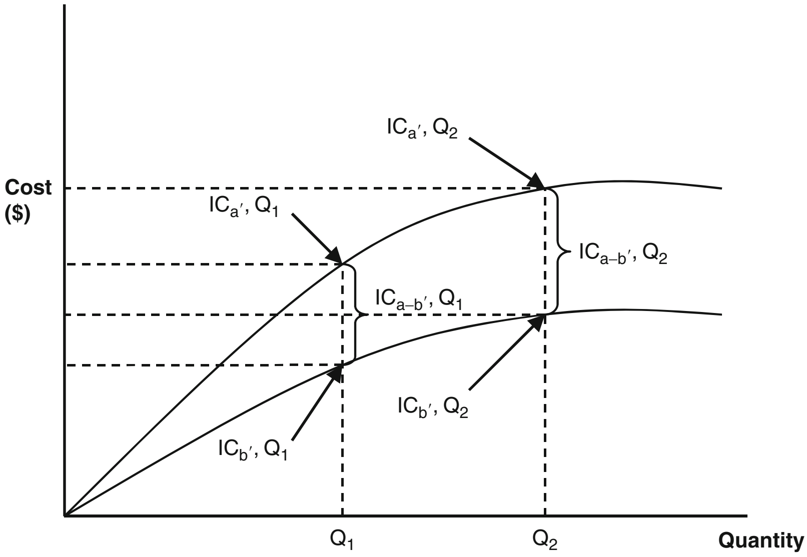

The Optimal Choice of an Action

The optimal choice of an action for a decision-maker is to maximize the benefits they receive less the opportunity cost. Comparing the addition to benefits with the addition to opportunity costs identifies the optimal choice. These are commonly referred to as the marginal benefit and the marginal cost of an action. The rule for optimal decision-making is that the program should be funded if the additional benefits from the program are equal to or greater than the additional costs. A full discussion of making the optimal decision and marginal CBA is available in Boardman et al. (2006).

Examples of Cost-Benefit Analyses

Below, two economic analyses of two third party policing (TPP) interventions are discussed to illustrate the use of cost-benefit analysis and its potential limitations. For those not familiar with TPP, it is defined as

> . . .police efforts to persuade or coerce other regulators or non-offending persons, such as health and building inspectors, housing agencies, property owners, parents, and business owners, to take some responsibility for preventing crime or reducing crime problems (Buerger and Mazerolle 1998, p. 301).

Månsdotter et al. (2007) conducted an economic evaluation of a multicomponent TPP alcohol prevention program implemented in licensed premises in Stockholm, Sweden. The study aimed to quantify the societal savings directly generated from the intervention. Savings were estimated as potential monetary consequences of victims of violence to the judicial system (e.g., police, public prosecutor, courts, prisons), production changes (e.g., sickness absence, other effects on working hours), health care (e.g., transport from crime scene, emergency treatment, inpatient care, outpatient care, pharmaceuticals), and other damage (e.g., technical support, personal assistance, and personal belongings). In addition, victims of violence were surveyed to ascertain the health effects of the intervention. The victim's health state was measured before, 2 weeks after the event, and present using the EQ-5D questionnaire.

The EQ-5D questionnaire is a multi-attribute health status classification system that includes five health attributes: mobility, self-care, usual activity, pain/discomfort, and anxiety/depression. Each attribute consists of three levels: no problem, some problems, and major problems, defining 245 possible health states including the possible states of unconsciousness and dead (Drummond et al. 1987). The results were converted to a weighted state health index by using values of health profiles from a UK general

Cost-Benefit Analysis, Table 2 Total costs of DWI incidents by severity and cost

	Fatal injury	Non-fatal injury	PDO vehicle	Costs for all cases
Medical	$6,693	$4,203	-	$4,127M
Emergency services	$930	$194	$24	$218M
Productivity	$665,453	$7,919	$35	$19,391M
Employer costs	$6,679	$530	$31	$23M
Administrative	$48,337	$1,259	$127	$2,474M
Legal	$70,925	$1,703	-	$2,869M
Travel delay	$387	$187	$107	$543M
Property damage	$8,059	$3,231	$1,157	$7,082M
Monetary costs	$807,473	$19,244	$1,481	$37,490M
Quality of life	$1,977,529	$48,367	-	$80,832M
Comprehensive costs	$2,785,002	$67,611	$1,481	$118,322M

Source: McKnight and Streff (1994)

population study (Dolan et al. 1996). Health gains were reported in quality of adjusted life years (QALYs). QALYs is a generic outcome measure that is calculated by adjusting the length of time affected through the health outcome by the corresponding utility value of the resulting level of health status (Drummond et al. 1987). The costs of implementation included administration (e.g., salaries for project staff, offices, travel, and conferences), studies of alcohol serving practices (i.e., remuneration to actors), community mobilization (e.g., advisory group participation), responsible beverage service training (i.e., 2-day course on responsible beverage service for 572 participants), and stricter alcohol law enforcement (i.e., joint controls by the licensing board and police and increase of notification letters to licensees).

Månsdotter and colleagues estimated the cost of the program to be 769,000 euros. The average cost of violent crime was estimated to be 19,049 euros, which equates to an approximate saving of 31.314 million euros. Savings were attributed to a reduction in judicial system costs (78%), production losses (15%), health-care issues (5%), and damages issues (2%). The cost saving ratio of the program was 1:39. Overall, health gains were estimated to be 236 QALYs (discounted QALYs), corresponding to 3–4 lives in normal health. As illustrated in the study, the intervention excluding savings was cost-effective in terms of health outcomes (assuming EUR 54,000 per quality of life – 795828/236 is approximately EUR 3,000 per QALY).

McKnight and Streff (1994) assessed the effects of a TPP intervention that enforced laws, using trained servers, to prohibit the service of alcohol to already intoxicated patrons. Specifically, the study addressed the following questions: (1) What is the effect of enforcement upon the service of alcohol to intoxicated patrons? (2) What impact does any change in service to the intoxicated have upon the relative numbers of DWIs coming from bars and restaurants? (3) How do the costs of enforcement relate to the estimated savings yielded by projected reductions in alcohol-related incidents?

The benefits of the intervention were calculated using Eq. 3:

$$\mathrm{B} = \mathrm{a} \times \mathrm{b} \times x \times \mathrm{ni\ ci} \tag{3}$$

where B represents the benefits from reduced drinking and driving, a equals the proportional reduction in DWIs (from taverns) (0.361), b represents the proportion of current DWIs (from bars and restaurants) (0.35), x is a factor that controls for the percentage of alcohol-related incidents that could have occurred in the absence of alcohol consumption (0.853), and $\sum ni\ ci$ represents the number of DWI incidents in a given year multiplied by the average costs per incident ($118.3 billion) (total comprehensive costs of all DWI incidents). Table 2 provides a summary of total DWI incidents by severity and cost.

Results show a cost-savings ratio of 260:1; that is, each dollar invested in the intervention

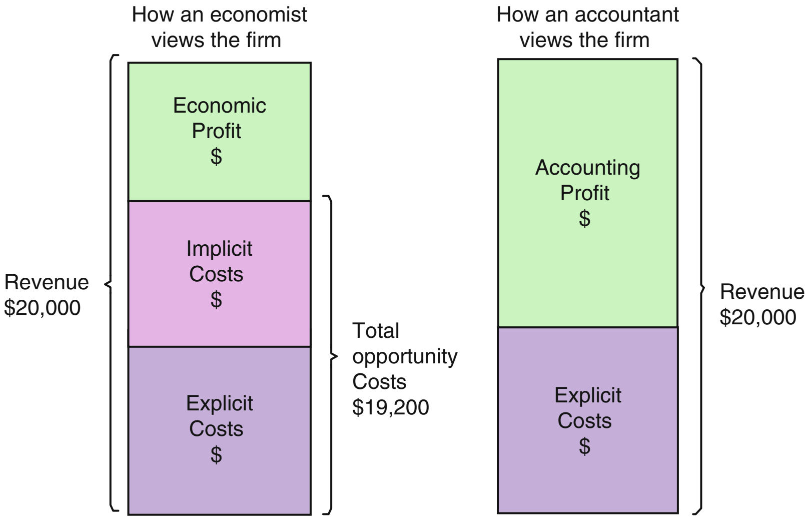

Cost-Benefit Analysis, Fig. 5 Economic vs. accounting costs. Adapted from: Frank et al. (2007)

generated a savings of approximately $260. Data for calculating this figure was derived from a study by Levy and Miller (1992), where savings ($260) equals $12.8billion (total cost to everyone involved in alcohol-related incidents) divided by $48,400 (total cost of effort needed to bring about a decrease in alcohol service violations and DWIs from bars. A number of assumptions were made with respect to estimated savings: (1) only reduced traffic risk was used—this excluded assault, household injury, and other risks associated with alcohol consumption; (2) there was a low percentage of patrons who were refused drinks drinking elsewhere; (3) "traffic safety benefits from refusals of service are confined to reductions of those matching BACs in excess of 0.10%"; (4) Washtenaw County is representative of the United States; (5) the level of enforcement is consistent with the effort as that employed in the Washtenaw County study; and (6) program effectiveness and the benefits of the intervention are the same over time.

Issues to Consider Ex Ante CBA

To highlight the potential limitations of CBAs and emphasize important issues *ex ante* CBA, the CBA examples presented above are used as an illustration. This discussion is disaggregated by the cost-benefit equation, where the benefit-cost ratio (BCR) is calculated by dividing the benefits of a program/intervention (B) by the costs of the program/intervention (C). The reader should note, however, that due to space limitations the discussion is restricted to a sample of limitations (specific to the CBA methodology) that may distort results. Boardman et al. (2006) provide a full discourse of limitations.

The Cost Side of the Equation

Månsdotter et al. (2007) noted that their cost estimates were based on ". . .account statement, average cost estimates and various kinds of estimations by individuals" (p. 622). Costs should not be limited to accounting costs, which are based on an account statement. Rather, costs should be economic costs. Economic costs include the following: implicit costs (a cost that is represented by lost opportunity in the use of one's own resources, excluding cash—these are sometimes referred to as intangible costs as they are not easily accounted for, and no actual payment is made) and explicit costs (a direct payment made to another party in the course of operation, e.g., wage, rent, and materials), where accounting costs only include explicit costs (see Fig. 5).

Economic costs are important for a number of reasons: (1) the information contained in accounting budgets does not include all the cost information on all the resources used (e.g., volunteer time, donated equipment and services, and any other unpaid inputs); (2) inputs that have already been paid for, or are not included in the budget, will not be discernable; (3) standard accounting budgets may distort the true cost of an input. That is, accounting budgets typically charge the cost of an input to the year in which the cost occurred. If the life of the asset is, for example, 20 years, then this cost should be distributed over the life of the asset; (4) costs are often embedded in a budget or expenditure statement that is larger than the project in question. For example, costs of an intervention (e.g., police time) may be embedded into the total operational costs of the enterprise, and as a consequence, these must be categorized according to economic classifications (e.g., variable costs); and (5) accounting documents often represent plans for how resources will be allocated as opposed to classifying expenditures after they have been used. That is, accounting statements often include budgeted accounts with actuals being left for inclusion until the end of the accounting period (Levin and McEwan 2001).

Total opportunity costs (TC) (implicit + explicit costs) should be classified into fixed (FC) and variable costs (VC). FCs represent costs that are independent of output (e.g., buildings and machinery). VCs are variable with output. VC generally varies at a constant rate relative to capital and labor. That is, VCs change as the quantity of output changes (e.g., if output = 0, then VC = 0; if output = 1, then output = 50; if output = 2, then output = 100). With this information, we are then able to calculate total costs (i.e., TC = FC + VC).

The Benefits Side of the Equation

The studies by Månsdotter et al. (2007) and McKnight and Streff (1994) raise a number of general problems with respect to assessing the benefits resulting from interventions/policies: (1) The first limitation is low response rate (e.g., surveys of victims of violence in the Månsdotter study). This may be problematic as it potentially influences results in the opposite direction; Månsdotter and colleagues state, "…one may imagine that the most harmed victims are most inclined to respond, but also that the severely harmed hesitated most about answering the questionnaire" (p. 622). (2) The savings and health gains attributed to an intervention are based on a decline in a given variable (e.g., violence). As such, there may be an underestimation in results, for example, spillover effects from not serving alcohol to underage patrons and staff turnover of RBS trained servers in the experimental area. On the other hand, there may be an overestimation where problem drinkers may be displaced to another area (Månsdotter et al. 2007). (3) Savings may be underestimated when an effectiveness study is limited to given times. For example, violence between 10pm and 6am and DWIs between 11pm and 5am Friday to Sunday. (4) Benefits of an intervention need to be comprehensive, that is, beyond the stakeholder perspective. For example, the Månsdotter study excluded benefits such as avoided damage and short-time closures. (5) Studies should include the long-term monetary benefits that are gained from an intervention. For example, the study by Månsdotter asked respondents to restrict their reports to 12 months after the violence—this represents only short-term benefits. (6) Studies should broaden the savings scope to include variables such as administration costs of sick leave and also medication provided by pharmacies (Månsdotter et al. 2007). (7) QALYs studies should expand their evaluation on a given independent variable. For example, examining whether unlawful threats involve more enduring violence could strengthen a study by potentially uncovering "more costs and health-damaging effects, than non-serious assaults" (Månsdotter et al. 2007, p. 622). (8) When assessing health gains (using the EQ-5D questionnaire) respondents in the Månsdotter study were asked to remember and report earlier and current health states. This may be problematic, particularly when respondents report better health at "present" than before the violence. This potential flaw certainly leads to an underestimation of health gains. (9) Studies should ensure that benefits of the intervention are assessed across all groups. For example, Månsdotter et al. (2007) did not include benefits to individuals who were prevented from purchasing alcohol. This may have uncovered both health gains and improvements in non-health-related quality of life (e.g., changes in life events had the individual been arrested for an alcohol-related crime). (10) Studies should be aware of the two potential shortcomings of research that may compromise the internal and the external validity of results, for example, (1) what else, other than participation in the program/intervention may have caused these results and (2) how general or representative are these findings to other groups?

Final Remarks

CBA is a technique that when used properly is a powerful policy decision-making tool. The technique is a little more subtle and sophisticated than most noneconomists recognize, and as such one should be careful in its application. Further, when using the results of CBA to inform public policy, one should understand that economic efficiency/effectiveness is not the only component of the decision-making protocol. Other goals such as equity and fairness should carry weight. Further, the maximization of social benefits that are identified in the economic study must be weighed against other competing interests such as moral and political objectives and how the goals of the system are conceived.

Related Entries

- ▶ Measuring Police Unit Performance
- ▶ Police-Led Interventions to Enhance Police Legitimacy
- ▶ Pulling Levers Policing
- ▶ Third Party Policing and School Truancy

Recommended Reading and References

Boardman AE, Greenberg DH, Vining AR, Weimer DL (1996) Cost-benefit analysis: concepts and practice, 2nd edn. Prentice Hall, Upper Saddle River

Boardman AE, Greenberg DH, Vining AR, Weimer DL (2001) Cost-benefit analysis: concepts and practice, 2nd edn. Prentice Hall, Upper Saddle River

Boardman AE, Greenberg DH, Vining AR, Weimer DL (2006) Cost-benefit analysis: Concepts and practice, 3rd edn. Pearson-Prentice Hall, Upper Saddle River

Brent R (2008) Applied cost benefit analysis. Edward Elgar Publishing, Northhampton

Buerger M, Mazerolle L (1998) Third party policing: a theoretical analysis of an emerging trend. Justice Q 15:301–328

Dolan P, Gudex C, Kind P, Williams A (1996) The time-trade-off method: results from a general population study. Heal Econ 5:141–154

Drummond MF, Stoddart GL, Torrance GW (1987) Methods for the economic evaluation of health care programmes. Oxford University Press, Oxford

Frank RH, Jennings S, Bernanke BS (2007) Principles of microeconomics. McGraw-Hill Irwin, Sydney

Levin HM, McEwan PJ (2001) Cost-effectiveness analysis, 2nd edn. Sage Publications, London

Levy D, Miller T (1992) A cost-benefit analysis of increased enforcement to reduce serving intoxicated patrons. University of Baltimore and the Urban Institute, Baltimore

Manning M (2004) Measuring the costs of community-based developmental prevention programs in Australia. Griffith University, Brisbane, Unpublished Masters (Hons)

Manning M (2008) Economic evaluation of the effects of early childhood intervention on adolescent outcomes. Griffith University, Brisbane, Unpublished PhD

Manning M, Homel R, Smith C (2006) Economic evaluation of a community-based early intervention program implemented in a disadvantaged urban area of Queensland. Econ Anal Pol 36(1 & 2):99–120

Manning M, Homel R, Smith C (2011) An economic method for formulating better policies for positive child development. Aust Rev Pub Aff 10:61–77

Månsdotter A, Rydberg M, Wallin E, Lindholm A, Andréasson S (2007) A cost-effectiveness analysis of alcohol prevention targeting licensed premises. Eur J Public Health 17(6):618–623

McKnight A, Streff F (1994) The effect of enforcement upon service of alcohol to intoxicated patrons of bars and restaurants. Accid Anal Prev 26:79–88

Mishan E (1972) The futility of pareto efficient distributions. Am Econ Rev 62(5):971–976

Zerbe R, Bauman Y, Finkle A (2006) An aggregate measure for cost-benefit analysis. Ecol Econ 58(3):449–461

Counterterrorism

Laura Dugan
Department of Criminology, University of Maryland, College Park, MD, USA

Overview

This entry divides terrorist countermeasures into deterrence-based, conciliatory-based, and situational-based strategies. Deterrence-based strategies attempt to reduce terrorism through punishment and the threat of punishment. While politically popular, empirical evidence suggests that punishment or repression directed toward terrorists or their constituency can lead to increases in terrorist violence. Conciliatory-based counterterrorism strategies offer rewards for legitimate alternatives to terrorist behavior. Examples include actions directed toward terrorists, such as leniency that might be offered to them for publicly renouncing their organization or conciliation that might be directed toward the terrorists' constituency, such as providing medical assistance at a time of need. The two studies that evaluate the effects of conciliation show that these actions are, indeed, related to less terrorism (i.e., in Israel and Turkey). Situational-based counterterrorism refers to efforts to reduce opportunities for terrorist attacks (e.g., installing metal detectors in airports). While evidence suggests that target hardening has effectively reduced some forms of terrorism (aerial hijacking), terrorists are often innovators who can develop alternative methods of attack. Finally, scholars are still learning what counterterrorism strategies have worked in the past to reduce terrorism as they begin to develop datasets that document the prior actions by governments to control terrorist violence.

Introduction

Concerns about terrorism have grown substantially since the late 1960s when the Popular Front for the Liberation of Palestine (PFLP) hijacked an Israeli El Al flight that was in route from Rome to Tel Aviv, demanding the release of Palestinian prisoners and creating a drama that permeated major media capturing the world's attention. It signaled to the world that terrorist violence is no longer contained in known conflict zones (Hoffman 2006). Prior to that event, most terrorist campaigns were localized, affecting only a single country or region and targeting only those individuals who were relevant to the terrorists' grievances (e.g., Basque Fatherland and Freedom [ETA] in the Basque region of Spain, Corsican National Liberation Front [FLNC] in the French island of Corsica, the National Liberation Front [FLN] in Algeria, the Irish Republican Army [IRA] in the United Kingdom). Once governments realized that terrorist violence was becoming a transnational problem, efforts began to develop more robust strategies for protecting citizens from being targeted by terror. This was especially apparent after the 1972 massacre of members of the Israeli Olympic team by the Palestinian terrorist organization Black Friday when Western Europe began to formulate antiterrorism policies and create counterterrorism units.

In the more than 40 years since the PFLP hijacking, reports on terrorist attacks have regularly pervaded media outlets, leading to global concerns that most persons are potentially vulnerable targets of terrorism. Despite this increasing concern, it is surprising how little is known about effective strategies to reduce terrorist violence. Schmid and Jongman (1988) reviewed more than 6,000 published works that examine terrorist violence and found that most of it is impressionistic making broad generalizations based only on episodic evidence. This conclusion is disheartening given that officials had to regularly develop counterterrorism strategies based on anecdotal evidence rather than empirically sound evaluations. As of 2004, this problem had only marginally improved. Lum and colleagues (2006) conducted a thorough and systematic review of counterterrorism evaluations as part of the Campbell Collaborative program. From the more than 20,000 written documents on terrorism between 1971 and 2004, only seven studies met their criteria of being moderately rigorous evaluation studies. This is especially surprising because a similar review of criminal justice evaluations found more than 500 rigorous and scientifically sound program impact evaluations (Sherman et al. 1997), raising a stark contrast between what we know about what works in criminal justice and what little we know about controlling terrorism.

To be fair, one of the primary reasons terrorism research had relied so heavily on case studies rather than empirical analysis is because comprehensive data that documents terrorist activity simply did not exist. This changed shortly after the beginning of the twenty-first century when RAND began to chronicle both international and domestic attacks and the National Consortium for the Study of Terrorism and Responses to Terrorism (START) compiled the Global Terrorism Database (GTD). As data became more widely available and extensive, scholars began to evaluate the effects of known government efforts to stop terrorist violence. The following sections describe different strategies that governments have used to control terrorism and the studies that have evaluated their effectiveness.

Types of Countermeasure Approaches

In this section, three major strategies are outlined to addressing terrorist violence. The first – *deterrence-based counterterrorism* – is by far the most appealing to officials because it is consistent with strategies to enforce criminal law and it demonstrates the strength of the government, increasing its popularity across constituencies. The second strategy – *conciliatory-based counterterrorism* – is typically less popular among officials and their constituents, but might be more effective in reducing violence. The third approach – *situational-based counterterrorism* – is similar to situational crime prevention, in that it hardens likely targets making it more difficult for terrorists to gain access to them. All three approaches are explained in greater detail below.

Deterrence-Based Counterterrorism

In his famous speech after the Iranian hostage crisis, President Ronald Reagan warned, "Let terrorists beware that when rules of international behavior are violated, our policy will be one of swift and effective retribution" (Stanik 2003: 33). This sets the tone for the Reagan administration's and future US administrations' response to international terrorism, as evidenced by the 1986 bombing of Libya, the 1993 bombing of Iraq's military intelligence headquarters, and the 1998 missile attacks in Afghanistan and Sudan. While rhetorically bold, Reagan's proclamation was consistent with hundreds of years of practice under Western criminal law traditions, which calibrates levels of punishment to best deter lawbreaking. Given this history, the appeal of repressive responses to terrorism is unsurprising, especially after devastating attacks like that of 9/11. In fact, President Bush's approval ratings soared as he assertively offered phrases like, "wanted dead or alive," "smoke them out," and "bring 'em on" (CNN 2001; Knowlton 2001; Associated Press 2003). With the popular support of the US public, the United States invaded Afghanistan demonstrating its resilience and strength, while decimating al-Qaeda training camps and forcing the Taliban into hiding.

Other efforts to deter terrorist behavior include apprehending terrorists and extending their prison sentences, passing antiterrorism laws, targeted assassinations, imposing curfews and containment, retaliating with violent military repression, invading territory, and indiscriminate repression.

While deterrence is appealing to many, others have argued that repression can lead to backlash. For example, since the US invasion of Afghanistan, opinions about its effectiveness are mixed. It is true that the organization al-Qaeda, as it was in September 2001, no longer exists; however, its successors remain highly networked and now have offshoots planning attacks across the globe. This new global jihad reaches supporters through its teachings on various Internet sites and its regularly published online English magazine, Inspire. Despite the rise in the global jihad movement, to date, the "new" al-Qaeda has been unable to replicate any attacks as sensational as the 9/11 attacks in the United States or elsewhere in the world, suggesting that the US repression response may have perhaps been an effective deterrent strategy. However, it can be argued that the damages caused by the attacks in Afghanistan and Iraq are far worse than any benefits because they have undermined the legitimacy of the United States, as the collateral damage gives the impression that the United States places little value on Afghani and Iraqi lives. Further, the military strikes seem to have increased al-Qaeda's popularity throughout the region, attracting large numbers of recruits into the global jihad (Malvesti 2002). In fact, some go as far as to claim that Osama bin Laden's intent behind the 9/11 attacks was to elicit a US response that would kill Muslims leading to retaliations (Malvesti 2002). This sort of "jujitsu" strategy is purposely designed to elicit a harsher response than the precipitating attack, increasing grievances and strengthening any broad-based loyalty to the terrorist organization (McCauley 2006). In other words, repressive policies adopted by the United States and other governments might actually cause more terror than they avert.

Thus, despite the appeal of deterrence-based counterterrorism strategies, the above argument suggests that when directed at terror, punishment will sometimes fail to deter and could even lead to more violence. This sort of backlash effect has been found in responses to other criminal behaviors and has been well explained by criminological theory. For example, labeling theorists argue that punished offenders will begin to identify more thoroughly with their role as lawbreakers solidifying their criminal (or terrorist) identities (Farrington 1977). Others have claimed that punishment can compromise the punisher's perceived legitimacy and consequently elicit acts of defiance (Tyler 2006), which is consistent with McCauley's (2006) argument that terrorists sometimes deliberately use a "jujitsu" strategy in order to sabotage a government's legitimacy.

Evidence of this sort of backlash effect from repressive actions has also been found in scholarly research. Several studies of the Republican terrorist campaign in Northern Ireland find that efforts by the British government to suppress terrorist violence increased rioting (Peroff and Hewitt 1980) and terrorism (LaFree et al. 2009). In fact, one Irish Republican Army member has been quoted as saying, "The British security forces are the best recruitin' officer we have" (Geraghty 2000: 36). Elsewhere, analysis has shown that during the Iranian revolution, repression led to a long-run increase in the frequency of protests (Rasler 1996). Piazza and Walsh (2009) found that countries that violate human rights are more likely to suffer terrorism than countries that adhere to human rights. Bartholomew (2011) found that laws that increased punishment for abortion-related terrorism failed to act as a deterrent, and in certain instances, they were associated with increases in property destruction and violence. Given the above evidence, one can argue that terrorism emerges as a response to perceived injustices, such as government repression, that inspire groups to mobilize in retaliation. In fact, even in cases where repression seems to work, it rarely comes without some costs. For example, Eppright (1997) found evidence that Israel's 1996 incursion into Lebanon led to a significant drop in Hezbollah's rocket attacks in Israel. Yet, Israel's actions in Lebanon also increased dramatically the local and international support for Hezbollah, undoubtedly strengthening the terrorist organization's propensity for future attacks.

Despite these mixed findings, one point seems clear from the research conducted thus far. Most terrorist organizations are unlikely to be deterred by traditional sanctions, especially since they are often wholly willing to exchange their lives or their freedom to strike a blow against their enemies. In fact, of the eight reasons that Cronin (2006) gives for a terrorist group's decline, only two are directly related to deterrence: the capture or killing of the leader and military force or repression. Empirically, Dugan and Yang (2011) show that the risk of continued terrorist attacks by Sendero Luminoso dropped substantially after the Peruvian government captured its leader, Guzman. While this result is promising, Cronin (2006) warns that the effectiveness of capturing one organization's leader depends upon its structure, the charisma of the leader, and the presence of a viable successor. This warning can be better heeded when we consider the effects of Israel's 1992 deportation of top Hamas leaders. Afterward, more radical mid-level leaders took over which eventually produced more deadly attacks against the Israelis (Hoffman and Cragin 2002). Capturing or killing a leader can also backfire if the group raises its captured or killed leader to the status of martyr, motivating further attacks (Cronin 2006). To date, we have yet to determine the effects of the killing of Osama bin Laden on the strength of the global jihad. Unsurprising, rhetoric suggests that attacks will increase, yet without effective leadership, the threats could be innocuous.

Despite the empirical findings described above, it is difficult to draw strong conclusions about deterrence without accounting for alternative approaches to countering terror. For example, LaFree et al.'s (2009) study of Northern Ireland selected only six actions by the British government from 1968 through 1992 to test their impact on Republican terrorism. The six actions were chosen because they were repeatedly cited as important in news accounts and scholarly research. Yet, by only selecting six actions over a 25-year period, the study ignored all less publicized – and perhaps less repressive – attempts by the British government that could have also influenced the rise and fall of Republican terrorism. Without data on the more subtle actions by governments, scholars may be getting biased results. Moreover, without including conciliatory as well as repressive government actions, we cannot assess the impact of alternatives to repression on reducing terrorist violence.

Conciliatory-Based Counterterrorism

Conciliatory-based counterterrorism is a logical alternative to deterrence-based strategies because it is similarly sourced in rational choice theory (see Dugan and Chenoweth 2012). Both strategies appeal to the rational actor, but instead of punishing lawbreaking behavior, conciliatory-based strategies reward legitimate alternatives to terrorism. Examples of earlier conciliatory efforts are found in Europe in the 1980s when Spain pardoned imprisoned ETA members after they publicly renounced the organization and its use of violence. The "reinsertion" policy allowed ETA members to live normal lives, free from the terrorist organization. Similarly, the Italian government offered leniency to members of the Red Brigades when they provided information that lead to the apprehension of other members (Cronin 2006). In both examples, governments strategically provided rewards for good behavior rather than punishing bad behavior. More recently, deradicalization programs have been aiding convicted terrorists by engaging them in religious dialogue to dismantle the ideological beliefs that justify terrorism. Concurrently, the programs work closely with detainees' families to prepare them to lead nonviolent lives by providing financial support to educate their children, train their wives, and help to reintegrate detainees into the community (Kruglanski et al. 2010).

Despite the promise offered by the theory behind conciliatory-based counterterrorism, little is known about its effectiveness because little is known about the types of actions that governments have engaged in to reduce terrorism. In an effort to remedy this empirical gap, Chenoweth and Dugan (2011) have been developing a comprehensive database that documents government's terrorist-related actions ranging from fully conciliatory to fully repressive in select countries. Along with my colleague Erica Chenoweth, I have collected the Government Actions in Terror Environments (GATE) data for the years 1988 through 2004 in Israel, Turkey, Algeria, Lebanon, and Egypt, and we plan to expand the collection to cover more countries over more years. Because this is the only effort to systematically capture a range of counterterrorism activity, the collection method will be briefly described below.

In order to attempt to capture all government actions that are relevant to any terrorist conflict for each country, we relied heavily on open sources. Only through media reports it is possible to learn about the more subtle, yet relevant government activity in a terror environment. The data were collected using Textual Analysis by Augmented Replacement Instructions (TABARI), which searches news articles and identifies observations that match the criteria of an extensive set of dictionaries designed to capture international and domestic activity (Schrodt 2006). TABARI is an automated text-coding program that codes the first sentence of news articles based on verb and noun pattern recognition. TABARI was used to code hundreds of thousands of Reuters articles downloaded from Factiva using country-level search terms (243,448 stories for "Israel*," 67,107 for "Leban*," 52,575 for

"Algeria*," 109,694 for "Egypt*," and 152,998 for "Turk*") as the sole search criterion for the period June 1, 1987, to December 31, 2004 (Chenoweth and Dugan 2011).

After TABARI identified relevant news articles, the output (i.e., actor, verb, and target) was filtered to select only those actions that government actors implemented toward substate targets. All verbs within these criteria were kept as relevant in order to ensure that all unexpected actions would be captured and that a wide range of actions were included even if they may not immediately seem like counterterrorism (e.g., Israeli attempts to allow developers to build more efficient water wells in the Palestinian territories). After autocoding the verbs to fall on a conciliatory-repression scale, every lead sentence was examined for proper coding (Chenoweth and Dugan 2011). During this cleaning process, each government action was attributed to politicians, the military, the judiciary, or the police. The resulting file contains the article's lead sentence, the actor, action, target, and new codes measuring level of conciliation and repression, whether the action was material and whether the target was discriminate (or specific) or indiscriminate (or general) for 6,063 Israeli government actions, 1,856 Turkish actions, 680 Algerian actions, 307 Lebanese actions, and 624 Egyptian actions for the years 1988 through 2004. This dataset gives an action-by-action view of state attempts to resolve conflicts with various non-state actors.

Chenoweth and Dugan (2011) present figures that compare conciliatory and repressive trends of government actions with that of terrorist attacks from the GTD for each country. In general, the three trends seem to track one another. For example, the GATE data show a rise in repressive actions during the First Intifada in Israel that was later combined with conciliatory actions. Further, during the Second Intifada, the Israeli government acted with a record high frequency of repressive actions. In Turkey, it appears as if repressive actions increased corresponding to terrorist attacks during the height of the Kurdish insurgency, which accounts for most of the terrorist attacks and most of the Turkish government's actions. It also appears that conciliatory actions only began to rise in the mid- to late 1990s, at the termination of the PKK-led Kurdish insurgency. In Algeria, the vast majority of repressive actions seem to accompany the Algerian Civil War through the 1990s, with repressive policies declining near the end of that decade as (primarily Islamist) terrorist attacks increased. In Egypt, the primary terrorist activity occurred in the mid-1990s, when a variety of Islamist groups launched an internal war against Hosni Mubarak's regime. These acts were accompanied by massive increases in repression, while conciliatory acts remained virtually nonexistent in Egypt. Finally, the patterns in Lebanon appear to be more mixed, which is likely due to the relatively low number of terrorist-relevant government actions compared to the other countries.

Thus far, the GATE data have only been used for two evaluations. In their study of Israel, Dugan and Chenoweth (2012) found, using a monthly time series analysis, that conciliatory actions are associated with decreases in Palestinian terrorism and repressive actions are related to increases in such violence. These findings suggest that conciliatory-based counterterrorism might be promising and affirms that repressive strategies should be used with caution. Preliminary findings on Turkey suggest a similar pattern, although the number of conciliatory actions is so few that the findings related to those actions fail the tests of statistical significance.

Situational-Based Counterterrorism

When we turn our attention away from the motivation of the offender and just assume that terrorists will always want to attack, we are free to consider traditional crime control strategies that are based on limiting offending opportunities. Clarke and Newman (2006) present this strategy of controlling terrorism by suggesting that we adopt a situational approach to preventing terrorism – similar to that used to control more common crimes. By systematically analyzing the opportunities that terrorists exploit, they suggest that we can direct our efforts toward developing reasonable means to block those opportunities. Of course, their strategy is adaptive because terrorist organizations will always exploit new avenues of attack. For example, Enders and Sandler (1993) and others (Dugan et al. 2005) show that after metal detectors were installed in US airports, the number of hijackings decreased. However, Enders and Sandler (1993) also demonstrated that metal detectors also led to increases in other kinds of hostage-taking tactics, suggesting that terrorists adapt to changes in situational factors. Jackson (2005) argues that terrorist organizations are innovative and that with enough resources will evolve beyond current barriers to their success. Thus, while situational-based counterterrorism is an important strategy, it should be accompanied by a better understanding of terrorist tactics that have been used in the past to overcome protective barriers (Clarke and Newman 2006). Regardless of the sophistication of the perpetrators, it is good policy to make it more difficult for terrorists to successfully attack vulnerable targets.

Conclusions

For this entry, I explored three strategies to control terrorism: deterrence-based, conciliatory-based, and situational-based counterterrorism. While it seems that the most popular approach is to rely on deterrence strategies, it is unclear from the empirical research that this is the most effective strategy. In fact, most evidence suggests that deterrence perpetuates more violence rather than facilitating its ending. Unfortunately, evaluations of the other two approaches are scant, making any strong conclusions premature. Regardless, it is safe to say that the problem of terrorism is complex and that a broad range of tactical strategies should be considered before taking action.

The literature on terrorist countermeasures is currently in its infancy. However, several developments show the potential to better understand the effects of a broad range of approaches to reduce terrorist violence. First, the availability of incident-based terrorism data, such as that of the GTD, allows scholars to finally have an objective empirical dependent variable to use in evaluation studies. Second, more scholars are developing datasets to document a broad range of efforts by governments and others to reduce terrorist violence. As these datasets are completed and made available, I expect an increase in the number of evidence-based studies, providing policy-makers with the much-needed substantiation to inform their decision-making.

Related Entries

- ▶ Deterrence
- ▶ Ecoterrorism
- ▶ Homegrown Terrorism in the United States
- ▶ Incarcerating Terrorists
- ▶ Innovation and Crime Prevention
- ▶ Preventing of Terror at Shopping Malls
- ▶ Rational Choice Theory
- ▶ Rational Choice, Deterrence, and Crime: Sociological Contributions
- ▶ Social Disorganization and Terrorism
- ▶ Strategies of Policing Terrorism
- ▶ Terrorist Organizations

Recommended Reading and References

Associated Press (2003) Bush: 'Bring on' attackers of U.S. troops. 2 July 2003

Bartholomew B (2011) The utility of deterrence-based sanctions in the prevalence of violence against abortion providers: testing a blended model of deterrence and backlash. Doctoral Dissertation, University of Maryland

Chenoweth E, Dugan L (2011) Exploring counterterrorism in the Middle East and North Africa: a new data set. Paper presented at the American Political Science Association annual meeting, Sept 2011

Clarke RV, Newman GR (2006) Outsmarting the terrorists. Praeger Security International, Westport

CNN (2001) Bush: Bin Laden wanted dead or alive, 17 Sept 2001

Cronin AK (2006) How al-Qaida ends. Int Secur 31:7–48

Dugan L, Chenoweth E (2012) Moving beyond deterrence: the effectiveness of raising the expected utility of abstaining from terrorism in Israel. Am Sociol Rev 77(4) (forthcoming)

Dugan L, Yang S-M (2011) Introducing group-based trajectory analysis and series hazard modeling: two innovative methods to systematically examine terrorism over time. In: Lum C, Kennedy L (eds) Evidence-based counterterrorism policy. Springer, New York

Dugan L, LaFree G, Piquero A (2005) Testing a rational choice model of airline hijackings. Criminology 43:1031–1066

Enders W, Sandler T (1993) The effectiveness of antiterrorism policies: a vector-autoregression-intervention analysis. Am Polit Sci Rev 87(4):829–844

Eppright CT (1997) Counter terrorism and conventional military force: the relationship between political effect and utility. Stud Confl Terrorism 20:333–344

Farrington DP (1977) The effects of public labeling. Br J Criminol 17:112–125

Geraghty T (2000) The Irish war. Johns Hopkins University Press, Baltimore

Hoffman B (2006) Inside terrorism. Columbia University Press, New York

Hoffman B, Cragin K (2002) Four lessons from five countries. RAND Rev 26:42–43, Summer 2002

Jackson BA (2005) Aptitude for destruction, volume 1: organizational learning in terrorist groups and its implications for combating terrorism. RAND Corporation, Santa Monica

Knowlton B (2001) Terror in American/'We're going to smoke them out': president airs his anger. New York Times, 19 Sept 2001

Kruglanski AW, Gelfand M, Gunaratna R (2010) Detainee deradicalization: a challenge for psychological science. Observer 23:1

LaFree G, Dugan L, Korte R (2009) The impact of British counterterrorist strategies on political violence in northern Ireland: comparing deterrence and backlash models. Criminology 47:501–530

Lum C, Kennedy LW, Sherley AJ (2006) Are counter-terrorism strategies effective? The results of the Campbell systematic review on counter-terrorism evaluation research. J Exp Criminol 2:489–516

Malvesti ML (2002) Bombing Bin Laden: assessing the effectiveness of air strikes as a counter-terrorism strategy. Fletcher Forum World Aff 26:17–29

McCauley C (2006) Jujitsu politics: terrorism and response to terrorism. In: Kimmel PR, Stout CE (eds) Collateral damage: the psychological consequences of America's war on terrorism. Praeger, Westport, pp 45–65

Peroff K, Hewitt C (1980) Rioting in northern Ireland. J Confl Resolut 24:593–612

Piazza JA, Walsh JI (2009) Transnational terrorism and human rights. Int Stud Q 53:125–148

Rasler K (1996) Concessions, repression, and political protest in the Iranian revolution. Am Sociol Rev 61(1):132–152

Schmid AP, Jongman AJ (1988) Political terrorism: a new guide to actors, authors, concepts, databases, theories and literature. North-Holland, Amsterdam

Schrodt PA (2006) Twenty years of the Kansas event data system project. Unpublished manuscript, University of Kansas

Sherman LW, Gottfredson D, MacKenzie DL, Eck J, Reuter P, Bushway S (1997) Preventing crime: what works, what doesn't, what's promising: a report to the United States congress. National Institute of Justice, Washington, DC

Stanik JT (2003) El Dorado Canyon: Reagan's undeclared war on Qaddafi. U.S. Naval Institute Press, Annapolis

Tyler TR (2006) Psychological perspectives on legitimacy and legitimation. Annu Rev Psychol 57:375–400

Counterterrorism and the Media

Revital Sela-Shayovitz
David Yellin Academic College and The Hebrew university, Jerusalem, Israel

Overview

The role of the media is particularly highlighted during critical periods in which there are extreme levels of terrorism threats, since the media are the primary source for gaining a perspective of the situation. The importance of the media was keenly demonstrated during the terrorist attacks on September 11, 2001, when the American public heavily relied on news reports and followed them more closely than before. Indeed, CNN and FOX News were the major news networks to learn about the events (Castro 2006). Furthermore, for the police, cooperation with the media is essential in order to successfully deal with a terrorism crisis. Media reports of anti-terror policing may help police navigate events, reduce feelings of panic, and raise the sense of confidence in society. At the same time, during high threat levels of terrorism, the media stress the coverage of public trust in police and contribute to the legitimacy by underscoring the effectiveness of police counterterrorist efforts. However, as the terrorism threat subsides, counterterrorism coverage becomes more complicated and may harm police legitimacy.

Faced with these realities, what can police organizations do in order to improve their cooperation with the media in their counterterrorist efforts? Fortunately, there is much that *can* be done. Police organizations must take into account the media when engaging in anti-terror policing and may develop various strategies to effectively respond to a terrorist strike with the media's help. In addition, police organizations can take advantage of the media in order to secure public assistance for carrying out antiterrorism policing.

Fundamentals of Police and the Media

The media are highly prone to focus their attention on the police in news programs as well as in the realm of fiction (e.g., in crime drama television series). Indeed, police work provides cases that are newsworthy and involve drama and violent action. Consequently, public perceptions of crime are distorted and characterized by an overrepresentation of severe and violent crime, and an underrepresentation of other types of offenses.

However, the media's tendency to concentrate on the police also derives from the perception that the police represents rule-governed order and even symbolizes the state's moral authority. Furthermore, the police force is the most visible organization of the criminal justice system, and therefore the onuses of dealing with crime and social control tend to fall disproportionately within their mandate (Ericson 1991; Chermak and Weiss 2005).

While the news media provide a wide forum in which police work is covered, they also present a conflicting portrayal of police. In some reports, the police image is highly professional and even heroic, while in others policing is portrayed as ineffective, marked by internal conflicts and corruption. In addition, the police image is also affected by the fictional depiction of television police serials (e.g., *CSI* and *NYPD Blue*), which profile the police in an unrealistic manner (Dowler and Zawilsk 2007). Overall, it seems that the media tend to represent police work as unsuccessful and incompetent rather than professional and effective (Fox and Van Sickel 2001; Gallagher et al. 2001).

Another important aspect is related to the role of the media in reshaping the legitimacy of the police in the public sphere. Police legitimacy plays an essential role in public compliance and cooperation with the police. Accordingly, if people perceive the police as legitimate and place trust in them, they are more willing to obey the law and collaborate with the police (Sunshine and Tyler 2003). In this context, the media assume a key function in "policing the police" by exposing cases of police wrongdoing and mismanagement. As a result, they stimulate public discourse regarding procedural propriety, which urges police accountability, and this in turn contributes to legitimacy. Therefore, the coverage of these cases serves as a mechanism of regulation of police work and the compliance of its members. Moreover, the very existence of journalists as vigilant watchdogs over police forces on behalf of the public interest also contributes to the legitimization of the police (Ericson 1991; Reiner 1985).

However, the media can sometimes become very critical and may in fact erode police legitimacy for a long period of time. Even a single high-profile event of police misconduct may dramatically undermine the public trust and seriously damage police prestige. Findings indicate that following well-publicized incidents of police misbehavior cases, there was an immediate and dramatic increase in public disapproval of the police, which persisted for a long period of time (Sunshine and Tyler 2003). Furthermore, the media's exposure of internal conflicts and management problems in the police organization may damage the latter's legitimacy. As a result, police spokespeople consistently struggle to control the police's image in the public sphere.

There is a continuous debate on the relationship between the police and media organizations. Police-media interaction seems to serve the needs of both types of organizations. The police are dependent upon the media in order to gain public cooperation, while the media need the police in order to obtain information on crime and terrorism, which are interesting and newsworthy items (Chermak and Weiss 2005; Lovell 2001). However, the dynamic interaction between these organizations may be depicted as coexisting relationships that ebb and flow in terms of dominance and control. This is because the police and the media organizations have different agendas: the police engage in controlling information in order to decipher crime, while the media are devoted to providing the "sensational" details of crime and the exposure of police misbehavior. Nevertheless, police organizations have increasingly developed a growing acknowledgement of the need to improve their interaction with the media. For example, police organizations have established structures and allocated appropriate resources to this end, such as public information officers (PIOs) who have trained in media communication in order to advance the organization's goals. In some countries, the police have adopted media education-training programs for police officers in order to improve their performance skills in the media. Furthermore, some police organizations have collaborated with the media in solving crime, e.g., by partaking in investigation television programs which promote police-media cooperation. Thus, although the interaction between the police and the media is sometimes tense, in general, both organizations report high levels of satisfaction with respect to the quality of their relationship (Chermak and Weiss 2005).

Public Opinion About Policing and the Media

There is scholarly consensus in respect to the influence of the media on the public views of the police. However, there is considerable debate about the direction, course, and degree of the effect. In this regard, three main hypotheses have been offered. The *hypodermic needle* hypothesis assumes that coverage has a powerful and long-term effect, since people mainly rely on the media for their sources of information. The *limited effects* approach suggests that people evaluate information from the media about the police on the basis of the context of what they already know from other sources, such as direct contact with the police. Therefore, reports about the police must contend with their preexisting knowledge. The *subtle/minimal effects* hypothesis maintains that the effect of the media is neither dominant nor minimal, but in fact is achieved differently. To elaborate, the media guide the perceptions of people regarding the most important issues involving crime and policing by highlighting particular topics and certain crimes. Hence, by framing police work in a certain way, the media reshape public perceptions and expectations about policing.

Studies have found that there is evidence for each of the three approaches. The overwhelming majority of the public consider the media as their main source of information about police and crime (Fox and Van Sickel 2001; Gallagher et al. 2001). At the same time, it was found that the media's influence is moderated by race and preconceived attitudes toward the police. People who have negative views about the police or belong to minority groups are more likely to consume the coverage of police misconduct cases, which in turn may reinforce negative perceptions and lead to generalize these cases as representative of police work. Thus, the impact of media coverage depends on people's perceptions and their prior experience with the police (Weitzer and Tuch 2004).

Media Framing Counterterrorism: International Perspectives

Terrorist attacks offer a powerful narrative for the media while often providing significant ratings. News reports provide terrorist attacks with meaning by reconstructing them within a social and political context that can be understood in a simple way (Nacos 2002). However, not all terrorist events receive coverage, and two main characteristics were found to determine the likelihood of broadcasting a report: the death toll and the extent of destruction. Accordingly, terrorist events occasioning a high death toll or involving significant destruction will produce considerable amounts of news coverage (Chermak and Gruenewald 2006). Furthermore, the coverage of terrorist attacks is liable to be distinguished by several aspects: the report usually appears in the headlines of the news broadcasts; the construction of newspaper headlines often becomes emotional and features dramatization; and news articles are longer and accompanied by further visual aspects, such as headline colors and photographs (Sela-Shayovitz 2007).

While the framing of terrorism has been extensively studied, the coverage of police counterterrorism efforts has largely been neglected. The discussion on this topic can be divided into two main categories: the news framing of anti-terror policing under high levels of terrorist threat and the coverage of anti-terror policing as a routine activity.

The news coverage of an extreme terrorist threat is often characterized by assigning a great priority to anti-terror policing. Journalists frequently tend to spotlight police performance in anti-terror events on the front pages of the newspapers rather than focus on the traditional roles of police work (Sela-Shayovitz forthcoming). Indeed, during times of a security threat, people are more concerned about the success of police performance in counterterrorist operations rather than take an interest in other aspects of police work (Tyler and Huo 2002). Thus, it appears that news editors are attentive to public concerns and, as a result, stress the police's ability to combat the danger.

A further central aspect is the role of the media in the *rally 'round the flag* effect. The *rally effect* relates to the impact of the security threat on the increase in public support for leaders and governmental institutions (Mueller 1973; Sigelman and Conover 1981). For example, following the September 11th terrorist attacks, the American public was united behind President George W. Bush, and his approval ratings soared to unprecedented levels. Furthermore, President Bush's policies benefitted from high levels of support from all across the political spectrum at this time (Gaines 2002). Similarly, in research conducted in Israel, during a period of intense terrorism in the Second Intifada, there was an increase in the Jewish public's positive evaluations of the Israeli police (Jonathan 2010). Several studies have demonstrated the key role of the media in the "rally effect." For instance, the CNN's massive coverage of the Persian Gulf crisis in 1990–1991 substantially relied on official sources and reshaped public approval of US policy by providing justifications for a military response (Livingston 1994). Additionally, immediately following the 9/11 terrorist attacks, the ten largest American newspapers reflected Bush's policy of "war on terror," and editors refrained from reporting criticism about the practical consequences of military intervention (Ryan 2004; Ruigrok and Van Atteveldt 2007).

In much the same manner, at these times, the media contribute to the "rally effect" and foster the legitimacy of the police in the public sphere. Research shows that during the period of an extreme terrorist threat in Israel, there was a significant increase in the favorable coverage of different aspects of police legitimacy, such as trust in police, police performance, interpersonal treatment, and the fairness of procedural processes. Journalists highly emphasized trust in police by frequently quoting statements of government ministers and citizens that reflected confidence in antiterrorism policing. In addition, the reporters refrained from engaging in criticism about police work. Yet, as the threat of terrorism declined, the media no longer remained positive and highly stressed their criticism of police (Sela-Shayovitz forthcoming).

Another interesting issue is the media's tendency to become patriotic when terror strikes. Immediately after the 9/11 terrorist events, the American media framing of the situation was repeatedly characterized by a network of associations among ideas, images, and symbols, such as *war* and *homeland security*, which stimulated popular support of the government (Eatman 2003). Furthermore, the American media tended to be more patriotic in comparison to the media in other countries that suffered from terrorism, such as Britain (Ruigrok and Atteveldt 2007).

The coverage of anti-terror policing is also distinguished by the use of terms and symbols of war. During periods of severe terrorism threats, journalists often use phrases, symbols, and associations that foster patriotism and an affinity with the police forces in counterterrorism, such as *police's struggle with terror* or *police combat with suicide bombers*. In addition, the newspapers glorify the achievements of anti-terror policing, employing terms associated with heroism. For example, "police officers prevent terrorist attack with their own bodies" or "blues to those who wear the blue uniform" (Sela-Shayovitz forthcoming). Thus, the media tend to embrace the police forces by underscoring their efforts in combating terrorism. This manner of coverage policing may encourage public solidarity and cooperation with the police forces. However, it is also possible that it results from the tendency of reporters to mainly rely on police sources, which consequently draw attention to police successes. In this context, research indicates that during these times, press editorials tend to adopt the official point of view provided by the government (Dor 2004).

Key Issues of Covering Anti-terror Policing

Following the decrease in terrorism threat, the covering of anti-terror policing becomes much more complex, since it involves ideological and political aspects. Indeed, there is an ongoing debate regarding the implications of anti-terror activities on human rights (Tyler et al. 2010). Thus, the media may serve as a forum where different perspectives about counterterrorism policies are debated and contested in the public discourse.

However, anti-terror strategies are less concerned with the infringements of human rights, whereas journalists are often devoted to the watchdog role of defending people's rights (Chermak and Weiss 2005; Ericson 1991). Therefore, the report may weaken police legitimacy by exposing cases of human rights violations in counterterrorism. Following the 9/11 terrorist attacks, antiterrorism policing often became a justification for undermining civil liberties, and the exposure of these cases in the media led to public disapproval with counterterrorism (McCulloch and Pickering 2005). In a similar vein, heavily publicized cases of the abuse of human rights in Britain have considerably damaged police legitimacy (Loader and Mulcahy 2003; Reiner 2000).

It is important to note that the coverage of the violations of human rights in counterterrorism specifically causes harm to police relations with minority groups, since they are more likely to consume police misconduct cases in the media, which in turn reinforces their negative opinion of the police (Weitzer and Tuch 2004). Indeed, minority groups, such as Muslim communities, are often perceived as high-risk groups and become a focus of anti-terror policing due to their ethnic background. Consequently, their relations with the police are delicate and fragile (Tyler et al. 2010).

On the other hand, the media can provide assistance for successfully dealing with terror both from an organizational aspect and in particular cases. From an organizational perspective, the media can help sustain public opinion regarding the necessity to persist in counterterrorism efforts, since this type of policing is an extension of police duties (Bayley and Weisburd 2009). Consequently, as the risk of terrorism subsides, it might be relegated to a lower priority because the public becomes once again more concerned with traditional policing (e.g., crime control and maintaining the social order). Therefore, the police can take advantage of the media's power to promote public awareness in carrying out this mission. Furthermore, police organizations may use the media for lobbying to obtain additional resources for anti-terror activities.

The media can help the police in specific cases in recruiting public cooperation and assistance. For instance, police organizations can use the media for issuing emergency warnings and other messages to the public, announcing clarifications about anti-terror operations (e.g., closing public areas for anti-terror drills), and promoting civil assistance in witnessing incidents, sharing suspicions, and providing information about terrorism.

Conclusions and Future Research

During times in which there is a high threat of terrorist attack, police-media cooperation is crucial in order to deal effectively with the situation and reduce panic. Additionally, media coverage of counterterrorism has a key role in police legitimacy and people's readiness to obey police instructions. However, although the debate about crime and the media is long running, little scholarly attention has been paid to the coverage of police counterterrorist efforts. Recent findings indicate that under an intense terrorist threat, the media become favorable toward counterterrorism policing, and legitimacy significantly increases in the press (Sela-Shayovitz forthcoming). These findings highlight the importance of additional inquiries into the role of the media in profiling anti-terror policing.

From a practical perspective, research also raises the question of whether police organizations can improve their collaboration with the media in regard to counterterrorism. In this context, several aspects are particularly noteworthy.

Police organizations must take the media into account in the management of anti-terror policing, since police-media collaboration is crucial for gaining control over terrorist events. Moreover, the police should take advantage of a suspension in terrorism to train police officers about media performance in order to respond more effectively at the time of crisis. Second, it would be more beneficial to develop a framework of cooperation between the police and the media in order to formulate strategies of the management of emergency events and in handling sensitive information. In fact, several police organizations have developed models of cooperation with the media, such as the Scotland Yard and Israel (Castro 2006). Third, police forces may enhance their capacity for working with the media in counterterrorism efforts by engaging a specific spokesperson police officer who is a professional in this field. This is based on the fact that general police spokespersons cannot address the specific needs of this type of policing. Indeed, counterterrorism policing is a very complicated task, and police spokespersons who are proficient in their work are vital for cultivating public assistance and cooperation (Chermak and Weiss 2005; Lovell 2001).

Another important aspect is related to the necessity of fostering dialog with Muslim communities in respect to antiterrorism policing. In this context, police organizations can promote their relations with the local media and particularly those media outlets that are consumed by minority groups in order to enhance dialog and collaboration.

However, due to the media's scrutiny of the violation of human rights, the police must be well equipped with strategies to manage the threat and minimize damage to the public trust. Although anti-terror policing is less accountable than other types of traditional policing, the police can use a variety of strategies to deal with these cases. For example, police can issue public announcements about investigations the case in order to draw their own lessons and pacify critical voices. Furthermore, the police can generate a process of regulation concerning the violation of human rights, which in turn reinforces legitimacy. Finally, police organizations may use the recourse of dismissing policepersons who have been accused of the abuse of human rights. Thus, by engaging in several techniques under the watchful eye of the media, the police can preempt the likelihood of overall criticism and restore public confidence in the police.

To conclude, future research should more closely examine the interaction between the police and the media in times of crises in order to develop strategies for effective cooperation during terrorist events. In addition, further research is required for a better understanding of the role of the media in shaping public perceptions about anti-terror policing. Most research in this field usually focused on one media outlet, whereas people from different social groups consume a diversity of media forms (national and local television, newspapers, radio, and the Internet). Thus, expanding the analysis to the impact of the various media genres on public opinion regarding counterterrorism may provide a deeper insight into this topic. It is also necessary to further examine the effect of the news coverage of police anti-terror activities and police legitimacy in order to enhance existing knowledge in this field.

Recommended Reading and References

Bayley D, Weisburd D (2009) Cops and spooks: the role of the police in counterterrorism. In: Weisburd D, Feucht TE, Hakimi I, Felson ML, Perry S (eds) To protect and to serve: policing in the years of terrorism, and beyond. Springer, New York

Castro CD (2006) The impact of police and media relations on a crisis. Naval Postgraduates School Monterey, California

Chermak S, Gruenewald J (2006) Media coverage of police misconduct. Int J Police Strateg Manag 29:261–281

Chermak S, Weiss A (2005) Maintaining legitimacy using external communications strategies: an analysis of police-media relations. J Crim Justice 33:501–512

Chermak S, McGarrell E, Gruenewald J (2006) Media coverage of police misconduct and attitudes toward police. Polic Int J Police Strateg 29:261–281

Dor D (2004) Intifada hits the headlines: how the Israeli press misreported the outbreak of the second Palestinian uprising. Indiana University Press, Bloomington

Dowler K, Zawilsk (2007) Public perceptions of police misconduct and discrimination: examining the impact of media consumption. J Crim Justice 35: 193–203

Eatman RM (2003) Cascading activation: contesting the White House's frame after 9/11. Political Communication 20:415–432

Ericson RV (1991) Mass media, crime, law, and justice: an institutional approach. Br J Criminol 31:219–249

Fox RL, Van Sickel RW (2001) Tabloid justice: criminal justice in an age of media frenzy. Lynne Rienner Publishers, USA

Gaines BJ (2002) Where's the rally? Approval and trust of the president, cabinet, congress, and government since September 11. PS Polit Sci Polit 35:531–536

Gallagher C, Maguire E, Mastrofski S, Reisig M (2001) The public image of the police. International Association of Chiefs of Police, Alexandria

Jonathan T (2010) Police involvement in counterterrorism and public attitudes toward the police in Israel: 1998–2007. Br J Criminol 50:748–771

Livingston S (1994) The terrorism spectacle. Westview Press, Boulder, CO

Loader I, Mulcahy A (2003) Policing and the condition of England: memory, politics and culture. Oxford University Press, Oxford

Lovell JS (2001) Police performances: media power and impression management in contemporary policing. Unpublished doctoral dissertation. The State University of Rutgers, Newark, NJ

McCulloch J, Pickering S (2005) Suppressing the financing of terrorism: proliferating state crime, eroding censure and extending neo-colonialism. Br J Criminol 45:470–486

Mueller JE (1973) War, presidents and public opinion. Wiley, New York

Nacos BL (2002) Mass-mediated terrorism: the central role of the media in terrorism and counter-terrorism. Rowman and Littlefield, Lanham

Philo G, Berry M (2004) Bad news from Israel. Pluto Press, London

Reiner R (1985) The politics of the police. Harvester, Brighton

Reiner R (2000) Crime and control in Britain. Sociology 34:71–94

Ruigrok N, Van Atteveldt V (2007) Global angling with a local angle: how US, British and Dutch newspapers frame global and local terrorist attacks. Booksurge Publishers, Charleston

Ryan M (2004) Framing the war against terrorism: US newspaper editorials and military action in Iraq. Gazette 66:363–382

Sela-Shayovitz R (2007) Female suicide bombers: Israeli newspaper reporting and the public construction. Crim Justice Stud 20:197–215

Sela-Shayovitz R (forthcoming) Police legitimacy under the spotlight: media coverage of police performance in the face of a high terrorism threat. J Exp Criminol

Sigelman L, Conover PJ (1981) The dynamics of presidential support during international conflict situations. Polit Behav 3:303–318

Sunshine J, Tyler TR (2003) The role of procedural justice and legitimacy in shaping public support for policing. Law Soc Rev 37:513–548

Tyler TR, Huo YJ (2002) Trust in the law. Russell Sage, New York

Tyler T, Schulhofer S, Huq AZ (2010) Legitimacy and deterrence effects in counterterrorism policing: a study of Muslim Americans. Law Soc Rev 44:365–401

Weitzer R, Tuch SA (2004) Race and perceptions of police misconduct. Soc Probl 51:305–325

Counterterrorism Strategies

► Strategies of Policing Terrorism

Courts

► Identification Technologies in Policing and Proof

Covert Policing

► Law of Undercover Policing

Creativity

▶ Innovation and Crime Prevention

Credibility Assessment

▶ Evaluating Truthfulness: Interviewing and Credibility Assessment

Crime

▶ Situational Action Theory

Crime and Justice

▶ Victimization, Gender, and the Criminal Justice System

Crime and Justice on Television

Nickie D. Phillips
Department of Sociology and Criminal Justice,
St. Francis College, Brooklyn, NY, USA

Overview

The past two decades have seen the emergence of a variety of crime-based reality television shows that purport to give watchers an insider's view of the criminal justice system. The convergence of private and public in a mediated context, particularly in the form of infotainment, is a consequence of late modernity. It is true that we are less able to separate our lives from the mediated reality in which we exist. As Mark Deuze (2011) points out, we are increasingly living life *in* media as opposed to *with* media. For example, it is becoming ever more common for police to be active participants in the commodification of crime as entertainment. For crime-based reality television shows featuring police officers, law enforcement agencies maintain editorial control over the final product and, in the process, strive to present themselves in the best light. Unseemly police behavior has been edited from the shows before air date, leaving a rather sanitized version of police work.

The merging of crime and entertainment on TV raises interesting questions. Among them is how "real" the depiction is of criminal justice and law enforcement in reality shows, with the acknowledgement that in a mass-mediated, late-modern society reality is a fluid concept. As the role of policing and media merge, questions arise as to what responsibility the police officers who appear in them, as well as the makers of the programming, have to produce a well-rounded and honest portrayal of the system.

Generally, crime-based reality programming reinforces a "crime-control" ideology, which places its emphasis on reducing the crime in society through increased law enforcement, prosecutorial powers, and harsher sentences. The shows tend to portray law enforcement, and other components of the criminal justice system, as working as intended – successfully implementing justice for the good of local communities. However, at times, these shows depict due process as an obstacle and require the outside support of citizen-participants. Ultimately, however, the shows feature a narrative in support of the status quo, one that reinforces punitive attitudes while providing little, if any, critical analysis of the system itself or discourse conducive to questioning existing criminal justice policies.

Background

Described on its website as "the rawest and the realest of all so-called reality shows," *COPS* was a successful forerunner among crime-based reality programs. The show premiered during primetime on the Fox television network in

1989 and became one of their most successful and longest running programs. The *COPS* aesthetic approach, cinéma vérité style camera work accompanied by a "Bad Boys" soundtrack, became an iconic representation of gritty, life-on-the-line police work. *COPS* is notable for being the first crime-based reality television show in the United States "to use actual video footage as opposed to reenactments" (Doyle 1988, 96). The show is filmed in various cities around the country and features shaky, handheld camera footage of police officers responding to calls in progress and engaging in order-maintenance policing. The viewer is on a virtual ride-along as the officers enter local communities to assist in such scenarios as spousal abuse, low-level drug busts, and alcohol-inspired disorderly conduct.

Crime-based reality television shows have proliferated since the late 1980s. *America's Most Wanted*, a show that enlisted the help of its audience to capture fugitives, debuted in 1988 and aired for 23 years. Since then, entire networks and cable channels have devoted their air time to crime – and justice-related programming. These shows vary in format, from the "unscripted" to the news magazine-style formatting of *Dateline*. They also vary in focus, with some dedicated to highlighting only federal crimes, female offenders, high-profile or celebrity cases, or corrections. Launched in 1991, the cable channel Court TV was designed to provide live coverage of trials, largely criminal, with the intention of helping inform individuals about the operations of the courtroom. The extensive, round-the-clock coverage of high-profile court cases, with attention to irrelevant and often salacious personal details of the various participants involved, has led to a media-frenzied environment characterized as "tabloid justice" (Fox et al. 2007). More recently, Court TV has been supplanted by TruTV, a cable channel self-described as "television's destination for real-life stories told from an exciting and dramatic first-person perspective and features high-stakes, action-packed originals." The network features a slew of crime-based reality shows, including *Police POV*, hyped by the network as a "groundbreaking" and "the most intense police series ever." Discovery network's *Investigation Discovery Channel* (known as ID) offers 24-h programming devoted to crime issues, featuring syndicated crime docudramas and original programming such as *Deranged*, *Life of a Crime*, and *Most Evil*. Further, about one-third of the shows now airing on the A&E network are crime-based reality shows such as *Cold Case Files*, *The First 48*, and *Beyond Scared Straight*.

Understanding the ever-increasing popularity of these crime-based shows requires a broader look at the genre of reality television. Given the diversity in format and content, it is not surprising that scholars maintain that there is no singular, clear definition of the genre of reality television. In response, Nabi et al. offered the following characteristics of reality television shows (see also Baruh 2009; Nabi et al. 2003):

> (a) people portraying themselves (i.e., not actors or public figures performing roles), (b) filmed at least in part in their living or working environment rather than on a set, (c) without a script, (d) with events placed in a narrative context, (e) for the primary purpose of viewer entertainment. This definition excludes programs captured by other genres. (p. 304)

Although their definition excludes "news programming, talk shows, and documentaries, as well as programs featuring reenactments (e.g., *America's Most Wanted*) and simple video clips not placed in a narrative context (e.g., *America's Funniest Home Videos*)" (p. 304), other scholars hold a broader definition of reality programming. In their seminal work on crime-based reality television, Mark Fishman and Gray Cavender (1998) place reenactment shows that encourage audience participation squarely in the realm of reality television. Specifically, they distinguish two types of crime-based reality shows: reenactments and narratives encouraging audience participation (e.g., *America's Most Wanted*) and ride-alongs with officers (e.g., *COPS*). Whether scripted or unscripted, these shows share one defining feature: the blurring of fact and fiction in an entertainment context.

The "Infotainment" Factor

Regardless of their specific content or format, the proliferation of crime-based reality shows is a testament to the prominence of "infotainment," a blending of "informative" programming (often news-like, with the purported intent of informing viewers of important social issues) with entertainment. Ib Bondebjerg points out that reality television shows elevate "private life stories" into "public discourse" while blurring the boundaries between fact and fiction (Bondebjerg 1996, p. 28). Bondebjerg does not necessarily find this hybridization troubling as long as the conditions of communication are clear. For example, the title cards at the beginning of the show could clearly delineate that the stories are based on true accounts but that the scenes consist of reenactments. For Bondebjerg, our everyday experience is being merged with more traditional forms of public service information into a hybrid form that at once informs and entertains the viewer. Shows fall on a continuum with regard to the amount of useful information given to inform the viewer versus dramatization designed to sensationalize. Scholars have generally been critical of this blurring of boundaries, pointing out that although television and other new media may afford us access to much more information – particularly with regard to the operations of the criminal justice system – we are not necessarily more knowledgeable about how the system actually operates. Given that a functioning democracy depends on an informed citizenry, some scholars have pointed out that entertainment disguised as information is not just misleading but rather poses a potential danger to democracy (Gamson et al. 1992; Postman 2005).

Crime-based reality television has been found to incorporate stylistic elements that blur the distinction between fact and fiction, with the intention of making the show seems more realistic. For example, some shows such as *America's Most Wanted* feature logos and set designs that are similar to official law enforcement regalia and often feature hosts that seem to have credibility either through their celebrity or personal story of victimization (Cavender 1998). Some shows rely on official police statistics, intimating that a close relationship exists between law enforcement and the hosts. (Cavender 1998) also notes that shows such as *America's Most Wanted* and *Unsolved Mysteries* usher in a sense of urgency by encouraging viewers to "call in now," which also gives the appearance of the shows being broadcast live even though they were prerecorded. In shows such as *COPS*, the action appears to be occurring in real time, even though the raw footage is heavily edited; sounds are blended to give the illusion that time is continuous (Doyle 1988).

While reality television merges fact and fiction, viewers are keenly aware that the shows do not portray an unadulterated "reality." Researchers have found that the public tends to view the reality claims made by such programs with a skeptical eye (Hall 2006; Nabi et al. 2003). Viewers recognize that although unscripted, the producers often manipulate circumstances with the intention of increasing dramatic effect. Police officers and other agents of the criminal justice system, like all participants in reality television shows, alter their behavior in the presence of cameras. In fact, in her study of viewer perceptions of reality programs, Alice Hall (2006) notes that realism is a fluid concept and that viewers acknowledge that the presence of the camera affects the behavior of participants, that circumstances are frequently staged, and that the final product has endured selective editing. Perhaps most significant, Hall found that while the viewers remarked that most situations were probably staged, they nevertheless contended the reactions of the participants were "real" and, as such, a reflection of their true nature and personality. Overall, though, scholars suggest that the more realistic the program is believed to be, the greater the effect on the viewer's attitudes and beliefs and that viewers were significantly more likely to characterize crime-based reality shows as more realistic than fictional programming (Oliver and Armstrong 1998).

Crime-Control Ideology

Scholars have suggested that the emergence and popularity of these shows are reflective of the political climate that arose during the 1970s and continued into the following decades. This tough-on-crime era is characterized by an escalating war on drugs, increased mandatory prison sentences, a shift away from rehabilitation, and the implementation of quality-of-life policing focused on reducing social and physical disorder. Crime-based reality television shows reinforce the status quo through support of crime-control policies and ultimately work to sustain crime myths. Further, like most media coverage of crime, the shows tend to overrepresent violent crime and are focused heavily on lower-class street crime to the exclusion of corporate crime, white-collar crime, and crimes committed by the state. For example, Oliver (1994) conducted a content analysis of 76 crime-based reality programs, including *COPS*, *America's Most Wanted*, *Top Cops*, *FBI*, *The Untold Story*, and *American Detective*, and found that, consistent with prior research, the shows overrepresented violent crimes, along with an unusually high rate of successful resolution.

Shows such as *Dallas/Kansas City/Detroit* SWAT and *DEA* feature paramilitary police units that are engaged in high-risk assignments such as enforcing the war on drugs, responding to barricade situations, and apprehending dangerous suspects. The *SWAT* franchise goes to great lengths to humanize the officers (most of whom are male), featuring on-screen mentions of their home life, including wives/girlfriends and children, and a website containing a feature where viewers can "meet the SWAT team." The suspects, however, are portrayed as interchangeable and anonymous one-dimensional characters. There is little to no consideration given to the (continuing) failure of the war on drugs. Crime-based reality shows have long been criticized for presenting crime in a decontextualized manner, with a reliance on use of force as the go-to option for resolving conflicts (Doyle 1988; Grant 1992). By presenting crime solely from the perspective of law enforcement officers, there is little room for a critical assessment of the system and its operation. Crime-based reality television shows provide a neatly packaged narrative wherein good versus evil is on stark display, and law enforcement is on the side of the angels. That police officers serve as the thin blue line protecting citizens and ultimately restoring social order is a compelling narrative for viewers.

In her study examining the ways in which crime-based reality television shows contribute to our nation's mythology surrounding crime control and the administration of justice, Jessica Fishman (1999) notes that the genre is not ideologically monolithic. While the worldview presented is one of dangerousness and where retributive justice prevails, the shows differ in the role of the state and civilian participants in the quest to eradicate crime. For example, in her study of *COPS* and *America's Most Wanted*, Fishman finds that *COPS* portrays state agents as heroic social protectors. The restoration of social order is reserved for police officers with special knowledge, skills, and resources to tackle the crime problem. From this perspective, civilians are ill-equipped and inadequately prepared to deal with crime. Conversely, *America's Most Wanted* provides viewers with a worldview in which law enforcement has initially failed in its mission to protect society and, therefore, citizens are empowered to actively participate in crime control. For *America's Most Wanted*, the most important catalyst in fighting crime is the alert citizen (Fishman 1999).

Some researchers have found that viewers who exhibit personality traits consistent with authoritarianism report greater enjoyment of reality programs, including crime-based shows (Nabi et al. 2003, p. 326). For example, Oliver and Armstrong (1995) surveyed viewers of crime-related programs and tested the concept of "disposition theory," the idea that "viewer enjoyment is strongest when liked characters are portrayed as winning and disliked characters portrayed as losing" (1995, p. 561). The researchers found that when compared to fictional programs, the crime-based reality programs were "most enjoyed by viewers who evidenced higher levels of authoritarianism, reported greater punitiveness about crime, and reported higher levels of racial prejudice" (1995, p. 565).

Researchers have also found that attitudes toward police may be impacted based on consumption of crime-based reality television shows, although the effect differs by race. For example, white viewers of crime-based reality shows reported higher levels of confidence in police compared to African-American viewers who demonstrated no increase in confidence (Escholz et al. 2002). These findings demonstrate that viewers are not simply passive recipients of manufactured ideological information. Rather, they actively interpret images of crime and justice, and as such, meanings surrounding criminal behavior and paths to justice are constantly negotiated within a cultural context.

Some have postulated that reality television, particularly crime-related programming, may be attractive to viewers because of its voyeuristic appeal. Researchers investigating the idea, however, have found mixed results. In seeking to assess the validity of the claim, the researchers first made the distinction between the clinical definition of voyeurism, which contains pathological, psychosexual elements of viewing unknowing victims, and the more common usage of voyeurism to mean watching other persons as they live their lives (Nabi 2003). While researchers have found some evidence indicating a voyeuristic gratification in the more common usage related to watching reality-based television shows, this is not necessarily the most important motivator for the viewers. In fact, Nabi, Stitt, Halford, and Finnerty suggest that for crime-based reality television, viewers may be less likely to tune in for the purposes of voyeurism and are more likely interested in learning about "criminal behavior and the criminal justice system" (2006, p. 442).

For crime shows such as *COPS*, the structure of the programming lends itself to what Baruh calls a "fly on the wall" approach in which camera positioning places the viewer in the position of a detached onlooker. This technique contributes to the idea that the viewer is bearing witness to a gritty reality that is often dangerous and always enveloped in larger notions of justice.

Race and Gender Stereotypes

Crime-based reality television shows have frequently been criticized for perpetuating racial and gender stereotypes. In *COPS*, for example, researchers found that the type of offense presented varied by the race of the offender: Whites were found to be more likely portrayed in offenses related to "cars and alcohol," while African-Americans were more likely to be shown committing burglary and grand theft (Monk-Turner et al. 2007). Researchers also found, after analyzing 8 h of episodes, that aside from the appearance of one female officer, all the police officers portrayed on *COPS* were white males. There was a conspicuous absence, on any episodes analyzed by the researchers, of any female minorities appearing as either police officers or offenders. Researchers have also found that while minorities remain underrepresented as cops, they are not necessarily overrepresented as suspects. However, when aggression was used by officers, minorities were "significantly more likely than white criminal suspects to suffer from unarmed physical aggression from police officers" (Oliver 1994, p. 9).

In their analysis of *COPS* and *World's Wildest Police Videos*, Prosise and Johnson (2004) suggest that the shows serve as a means to legitimize and justify questionable police behavior such as racial profiling. For example, the authors point out that police officers often rely on hunches, or "officer intuition," to justify police stops. These hunches are portrayed as an effective means of making a good arrest and implicating a guilty suspect. Hunches that prove to result in bad stops or arrests are likely to end up on the editing room floor.

Female police officers are more likely to be featured in their own spin-off shows, such as *Female Forces* and *Police Women of Broward County*, rather than fully incorporated into the more traditional, male-centric shows. Rabe-Hemp (2011) examined the ways in which gender was negotiated on the show *Female Forces*, airing on *Biography* network. The show, described on the network website as "Brains, beauty, and a badge...," follows female police officers in Naperville, Illinois, "as they fight a full gamut of big city crime in the suburbs of Chicago." In contrast to other, male-dominated crime reality television shows, *Female Forces* was more likely to show vignettes featuring the officers engaged in order-maintenance style policing as opposed to responding to calls regarding violent crimes. While this portrayal more closely aligns with the "reality" of crime (i.e., according to official reports, there are more calls for service for noncriminal and/or order-maintenance incidents than for violent crimes), it illuminates the extent to which the male-dominated shows tend toward a macho, aggressive style of policing focused heavily on crime fighting.

In her analysis, Rabe-Hemp explores how stereotypes of female police officers are both reinforced and challenged. By virtue of portraying women in an authoritative position, the patriarchal structure that privileges males as the sole figures of authority is challenged. At times though, this resistance was negotiated through emphasized femininity. For example, Rabe-Hemp describes how the female officers insisted that it was possible to remain feminine and still be a competent police officer, all the while engaging in personal grooming such as hair brushing and applying makeup (Rabe-Hemp 2011).

Current Issues and Controversies

Future Directions

Compared to the courts and corrections, generally, most media coverage of the criminal justice system is focused on law enforcement. The relative lack of attention to corrections is likely due to the lack of willingness on the part of correctional officials to open up the facilities to the media (Surette 2011). Surette (2011) noted that generally, prisons are presented in the media rather negatively, focusing on prison riots and misconduct by correctional officers, in portrayals that are unlikely to garner public support for the institutions. Perhaps not surprisingly, most research on crime-based reality television programs also centers on the front end of the criminal justice system, including the early, highly successful shows such as law enforcement-focused *COPS* and *America's Most Wanted*. Other reality television shows that feature corrections, such as *Lock Up* or *Inside American Jails*, are often neglected by researchers. Future research may examine the ways in which prison-focused reality shows feature inmates as spectacle and doomed to recidivate. Researchers may explore whether the shows provide a critical discourse on the failures of the system or engage in promoting a view of criminal behavior that is reduced to a simple issue of individual responsibility, while neglecting the insurmountable obstacles that most offenders face upon reentry.

The Role of the Media and the Role of Law Enforcement

Other future issues for criminologists to explore may center less on the extent to which shows are misrepresenting "reality" and more on acknowledging that criminal justice is increasingly a lived experience on-camera. It is in this context that the distinction between the role of the media and the role of law enforcement has become increasingly muddled. With the proliferation of crime-based reality shows, members of police departments are taking on a more proactive role *as media*, while members of the media are serving as pseudo-law enforcement agents. For example, police officers who have participated in *COPS* have reported to researchers that they felt that the television crew was *on their side*, even ready to jump into a fight if necessary (Hallett and Powell 1995) [emphasis added].

The consequences of the fusion between media and law enforcement are no more apparent than in the top-rated show *To Catch a Predator*. As scholars have noted, in that show, the delineation between civilian vigilante group, law enforcement, and the media has become dangerously blurred. The show featured Dateline NBC correspondent Chris Hansen as he confronted alleged sexual predators who have made contact with underage participants in stings set up by the program in cooperation with local law enforcement agencies. The alleged predators were lured to the decoy house and subject to Hanson's interrogations. Law enforcement agencies partnered with a civilian group known as Perverted Justice, whose members were dedicated to stamping out pedophilia by posing online as underage victims and exposing the adults who engage in explicit communication with the decoys. The program's popularity stems in part from its reliance on shame and humiliation as public spectacle that has become a staple of reality television (Kohm 2009).

The show found itself the recipient of unwanted media attention as a result of the suicide of Bill Conradt, an assistant district attorney, caught exchanging explicit internet chats with a presumed underage decoy in one of the show's sting operations. Conradt shot himself as the SWAT team and camera crews closed in on his home. Investigations into the botched sting targeting Conradt revealed the extent to which media directs and influences law enforcement activity. Though Conradt had, for whatever reason, declined to actually show up at the decoy house, Hansen pressed further, requesting the police obtain arrest and search warrants to be executed at Contradt's home the next morning. From the hastily written arrest and search warrants that were riddled with errors to the involvement of a SWAT team – likely influenced by the show's usage of dramatic takedowns in situations in which less a show of force would suffice – it becomes apparent that concern for spectacle was among the primary motivations for the actions of law enforcement. In fact, when determining whether to press charges on the other 24 alleged perpetrators captured in the sting, the prosecutor was forced to drop the cases citing issues of illegality in police procedure and venue problems.

A lawsuit filed by the show's own producer alleged unethical relationships between the show, Perverted Justice, and law enforcement. The suit claimed that the Perverted Justice group did not always supply the producers with full transcripts of the chats and that the show covered up questionable police behavior. The producer alleged she was fired for speaking out about the unethical practices of the show. Although her lawsuit was eventually dismissed, a second wrongful death lawsuit filed by Conradt's sister was settled for an undisclosed amount. The suit alleged that "the network interfered with police duties and then failed to protect her brother's safety." When ruling that the wrongful death case could move forward, Judge Denny Chin pointed out "that NBC had 'placed itself squarely in the middle of a police operation, pushing the police to engage in tactics that were unnecessary and unwise, solely to generate more dramatic footage for a television show.'" Future research could continue to document the ways in which crime and criminal justice is "dangerously confounded with its own representation" (Ferrell et al. 2008).

Shame and Humiliation

One of the hallmarks of crime-based reality television shows is the extent to which suspects are shamed and humiliated on-screen. Future research may continue to explore the ways in which moral sentiments are mass-mediated and the public spectacle of punishment has been revived in the context of reality television. Crime-based reality shows have managed to merge crime and entertainment into "humiliation TV," often airing segments with no other end than to ridicule the offender. Shows that compile caught-on-tape moments such as *Disorder in the Court* and *World's Dumbest...*, highlighting the misadventures of criminals and other ne'er do wells, are prime examples. For some, such as those implicated in *To Catch a Predator*, the consequences of shame and humiliation may prove deadly.

In her analysis of *COPS*, Mariana Valverde (2006) points out that the show portrays police officers as engaged against the enemy "others" who represent stereotyped "criminal classes." The social context from which crime emerges (i.e., larger structural constraints of poverty, race, class, and gender) is neglected in favor of a focus on the apprehension of lower-class criminals, frequently intoxicated, and virtually always surrounded by social and familial disorder. Humiliation and shame are part and parcel of the show as the camera focuses an inordinate amount of time on the incoherent speech and movements of those who are intoxicated for the viewers' pleasure. Many subsequent shows such as *Dallas/Kansas City/Detroit SWAT*, *DEA*, *and Police Women of Broward County* have replicated *COPS* cinema verité style along with other conventions and are subject to similar criticisms.

Increasingly, shame and humiliation have extended to reality shows that focus on ethical dilemmas that at times may overlap with criminal behavior. Shows such as *Intervention*, *Celebrity Rehab*, and *Sober House* focus on the shame and humiliation of addiction yet decouple drug use and addiction from criminal behavior, putting on full display our cultural ambivalence surrounding drug use.

Cultural criminologists have suggested that part of the pleasure of viewing reality television is that it provides a means of privately transgressing. In a vicarious manner, viewers are relating to and experiencing deviance in a socially sanctioned way – they are offered a means of reaffirming social boundaries and for condemning deviance through the humiliation and shame of the other. In fact, Hall (2006, p. 204) found that the viewers' enjoyment of reality television shows is, in part, "a strong element of schadenfreude." Cultural criminologist Mike Presdee suggests that watching "real" people being humiliated calls up a "different order of emotions, pleasures and desires" than that of fictional humiliation (Presdee 2001, p. 80). However, given that there is such fluidity between viewers' perceptions of what constitutes "reality" television from fictional television, it is likely that those emotions, pleasures, and desires are not so distinct. Reality shows then are part of a larger social context in which crime, shame, humiliation, and violence are sold as commodities to a willing public eager to be entertained.

Related Entries

▶ Cultural Criminology

Recommended Reading and References

Baruh L (2009) Publicized intimacies on reality television: an analysis of voyeuristic content and its contribution to the appeal of reality programming. J Broadcast Electron 53(2):190–210. doi:10.1080/08838150902907678, doi: papers://227A529C-089D-4314-B924-24AB50571CE2/Paper/p512

Bondebjerg I (1996) Public discourse/private fascination: hybridization in 'true-life-story' genres. Media Cult Soc 18(1):27–45. doi:10.1177/016344396018001003

Cavender G (1998) The shadow of shadows: television reality crime programming. In: Fishman M, Cavender G (eds) Entertaining crime: television reality programs. Aldine De Gruyter, New York, pp 79–94

Deuze M (2011) Media life. Media Cult Soc 33(1):137–148. doi:10.1177/0163443710386518

Donovan P (1998) Armed with the power of television: reality crime programming and the reconstruction of law and order in the United States. In: Fishman M, Cavender G (eds) Entertaining crime: television reality programs. Aldine De Gruyter, New York, pp 117–137

Doyle A (1988) Cops: television policing as policing reality. In: Fishmann M, Cavender G (eds) Entertaining crime: television reality programs. Aldine De Gruyter, New York

Eschholz S, Blackwell B, Gertz M, Chiricos T (2002) Race and attitudes toward the police: assessing the effects of watching. J Crim Just 30:327–341, (12355527252820796286related:fqMK9ZKmd6sJ), doi: papers://227A529C-089D-4314-B924-24AB50571CE2/Paper/p446

Ferrell J, Hayward K, Young J (2008) Cultural criminology: an invitation. Sage, Thousand Oaks

Fishman J (1999) The populace and the police: models of social control in reality-based crime television. Crit Stud Mass Commun 16:268–288

Fishman M, Cavender G (1998) Entertaining crime: television reality programs. Aldine Transaction, New York

Fox R, Van Sickel R, Steiger T (2007) Tabloid justice: criminal justice in an age of media frenzy. Lynne Rienner, Boulder

Gamson WA, Croteau D, Hoynes W, Sasson T (1992) Media images and the social construction of reality. Annu Rev Sociol 18:373–393

Grant J (1992) Prime time crime: television portrayals of law enforcement. J Am Cult 15(1):57–68

Hall A (2006) Viewers' perceptions of reality programs. Commun Q 54:191–211

Hallett M, Powell D (1995) Backstage with "Cops": the dramaturgical reification of police subculture in American crime "info-tainment". Am J Police 14(1):101–129

Jewkes Y (2010) Media & Crime. Sage, Thousand Oaks

Kohm S (2009) Naming, shaming and criminal justice: mass-mediated humiliation as entertainment and punishment. Crime Media Cult 5:188, (related: woyeB3m3gTsJ), doi: papers://227A529C-089D-4314-B924-24AB50571CE2/Paper/p80

Kooistra P, Mahoney J, Westervelt S (1998) The world of crime according to "Cops". In: Fishman J, Cavender G (eds) Entertaining crime: television reality programs. Aldine De Gruyter, Hawthorne, pp 141–158

Kort-Butler L, Hartshorn K (2011) Watching the detectives: crime programming, fear of crime, and attitudes about the criminal justice system. Sociol Q 52:36–55

Monk-Turner E, Martinez H, Holbrook J, Harvey N (2007) Are reality TV crime shows continuing to perpetuate crime myths. I J Criminol 6:132

Nabi RL, Biely EN, Morgan SJ, Stitt CR (2003) Reality-based television programming and the psychology of its appeal. Media Psychol 5(4):303–330

Nabi RL, Stitt CR, Halford J, Finnerty KL (2006) Emotional and cognitive predictors of the enjoyment of reality-based and fictional television programming: an elaboration of the uses and gratifications perspective. Media Psychol 8(4):421–447

Oliver M (1994) Portrayals of crime, race, and aggression in "reality-based" police shows: a content analysis. J Broadcast Electron 38(2):179–192

Oliver M, Armstrong G (1995) Predictors of viewing and enjoyment of reality-based and fictional crime shows. Journal Mass Commun Q 72(3):559–569, doi: papers://227A529C-089D-4314-B924-24AB50571CE2/Paper/p414

Oliver M, Armstrong G (1998) The color of crime: perceptions of Caucasians‘ and African-Americans' involvement in crime. In: Fishman M, Cavender G (eds) Entertaining crime: television reality programs. Aldine De Gruyter, New York, pp 19–36

Postman N (2005) Amusing ourselves to death: public discourse in the age of show business. Penguin (Non-Classics), USA

Presdee M (2001) Cultural criminology and the carnival of crime. Routledge, USA

Prosise T, Johnson A (2004) Law enforcement and crime on cops and world's wildest police videos: anecdotal form and the justification of racial profiling. West J Commun 68(1):72–91

Rabe-Hemp CE (2011) Female forces: beauty, brains, and a badge. Fem Criminol 6(2):132–155. doi:10.1177/1557085111398471

Surette R (2011) Media, crime and criminal justice: images, realities and policies, 4th edn. Wadsworth, Belmont

Valverde M (2006) Law and order: images, meanings, myths. Rutgers University Press, New Brunswick

Crime and the Racial Composition of Communities

Charis E. Kubrin
Criminology, Law and Society, University of California, Irvine, CA, USA

Overview

One of the most significant social facts about crime is that it is not randomly distributed across neighborhoods within a city. Rather, crime tends to cluster in certain areas or neighborhoods. This is why most residents are able to identify the "bad section" of town. For over a century, scholars have attempted to understand the spatial patterning of crime. A common approach involves investigating the characteristics of neighborhoods with high crime rates. One frequently examined neighborhood characteristic is racial composition. Scholars ask, how is the racial and ethnic composition of a neighborhood related to its crime rate? This entry reviews the literature that attempts to address this question, the theoretical approaches that hypothesize a relationship between the two, and some challenges and obstacles associated with studying neighborhood racial composition and crime. In the concluding section, the entry offers some important future directions researchers can take to better understand crime and the racial composition of communities.

Key Findings in the Literature

A consistent finding in the neighborhoods and crime literature is that racial composition matters. Although racial composition is a broad construct that can reflect a variety of things, with few exceptions, studies published to date focus primarily on one of two measures to capture community racial composition: racial heterogeneity and percent black.

As far back as the early 1900s, researchers investigated the effects of racial and ethnic heterogeneity within communities (Shaw and McKay (1969 [1942]; Sellin 1938). In their classic work *Juvenile Delinquency and Urban Areas*, Shaw and McKay (1969 [1942]) argued that three structural factors – ethnic heterogeneity among them – led to the disruption of local community social organization, which in turn accounted for variations in rates of crime and delinquency. Nearly 100 years later, studies that analyze racial heterogeneity and crime continue to document a strong positive relationship between criminal violence and an area's level of racial diversity, although it is fair to say the findings reveal a more complex relationship than previously assumed.

For example, using victimization data from 57 neighborhoods to examine the relationship between neighborhood characteristics, including racial heterogeneity, and rates of violent crime and burglary, Smith and Jarjoura (1988) show that poverty and ethnic heterogeneity, among other factors, predict neighborhood victimization rates, although their influence is more conditional than direct and varies by type of crime. Similar findings are reported by Warner and Pierce (1993) using data on calls to the police in 60 Boston neighborhoods. Warner and Pierce find that although racial heterogeneity has a positive effect on burglary rates, its effects are conditional on the level of poverty. Further evidence of interaction effects between racial heterogeneity and poverty are documented in both Warner and Rountree (1997) and Kubrin (2000) who examine neighborhoods in Seattle, Washington. Warner and Rountree (1997) show that heterogeneity has a significant direct effect on assault when poverty is at its mean, but its effect is greatest in nonpoor neighborhoods and is less strong in high-poverty neighborhoods. Likewise, Kubrin (2000) finds that in neighborhoods characterized by low poverty levels, racial heterogeneity increases violent crime, whereas in high-poverty neighborhoods, heterogeneity actually decreases crime.

Far more studies have operationalized racial composition in a more limited way – as reflective of the percentage black in a community. As with racial heterogeneity, findings from this literature indicate that this measure of racial composition is associated with crime rates across communities. In their recent review of the neighborhood literature on race and crime, Peterson and Krivo (2010) conclude, “Rates of crime are higher in local areas where African Americans are more heavily concentrated and, conversely, lower where African Americans are constitute a smaller portion of the residents, whether in Atlanta, Baton Rouge, Chicago, Cleveland, Columbus, New York, Philadelphia, or Seattle” (p. 14; see this discussion for a list of relevant studies in this area). Peterson and Krivo note that often the differentials in crime across communities of different racial composition are quite stark. For example, McNulty’s (2001) study of block groups in Atlanta reveals that murder rates were a startling six times higher in predominantly black neighborhoods than in predominantly white neighborhoods, while rape and robbery rates in African American areas were over three times those in white areas and aggravated assaults were about twice as common in the former than the latter.

As this (admittedly brief and incomplete) overview of studies on crime and the racial composition of communities reveals, race, place, and crime are inextricably linked in the United States. What accounts for this linkage and what mechanisms underlie these relationships? In the remainder of the entry, I discuss three theoretical approaches – social disorganization, labor-market stratification, and cultural/subcultural explanations – that attempt to account for the racial composition-crime connection across neighborhoods.

Social Disorganization Theory

Social disorganization theory has emerged as the critical framework for understanding the relationship between community characteristics and crime in urban areas. According to the theory, certain neighborhood factors can lead to social disorganization, defined as the inability of a community to realize the common values of its residents and maintain effective social control (Kornhauser 1978, p. 120). Communities with ineffective social control have higher crime rates.

Social control serves two purposes: to ensure residents’ conformity to norms and to regulate deviant behavior. According to the theory, two types of social control – formal and informal – operate to achieve these ends. Formal control is exercised by institutions such as the police to maintain stability in the neighborhood. Informal social control is exercised by families, peers, and neighbors who “look out” for the community by discouraging deviant behavior. Informal social control, at the heart of the theory, is said to provide the first line of security for community members.

The dynamics of informal social control are discussed by Greenberg and her colleagues (1985) who identify three ways it may impact crime: (1) via informal surveillance, or the casual but active observation of neighborhood streets that is engaged in by residents during daily activities; (2) via movement governing rules, or the avoidance of areas in or near the neighborhood (or in the city as a whole) that are viewed as unsafe; and (3) via direct intervention, or through the questioning of strangers and neighborhood residents about suspicious activities. Direct intervention may also include chastising adults and admonishing children for behavior defined as unacceptable.

From its inception, social disorganization theory has considered a community's racial composition as among the most salient predictors of neighborhood crime rates. In particular, arguments center on the community's level of racial diversity among residents. The theory proposes that racially heterogeneous communities have higher crime rates in large part due to diminished informal social control. It is argued that in communities with diverse racial groups living side by side, interaction between members will be low, or at least lower than in neighborhoods with individuals from the same racial background. It is also assumed that racial heterogeneity can undermine ties between neighbors, limiting their ability to agree on a common set of values or to solve commonly experienced problems. Reasons for both arguments point to cultural differences between racial groups, language incompatibility, and the fact that individuals prefer members of their own race to members of different races. As noted earlier, studies that explicitly examine racial heterogeneity's impact consistently find it is significantly and positively associated with neighborhood crime rates (Kubrin 2000; Smith and Jarjoura 1988; Warner and Pierce 1993; Warner and Rountree 1997).

Labor-Market Stratification Theories

A second theoretical perspective that attempts to account for the racial composition-crime connection across neighborhoods focuses on labor-market characteristics in neighborhoods with significant numbers of racial minorities. Researchers ask. What are the key labor-market characteristics associated with high rates of violence that help to explain high crime rates in minority neighborhoods?

Theorists interested in answering this question draw heavily on dual labor-market theory, whose underlying assumption is that not all jobs are created equal. The theory differentiates between primary sector jobs, or those desirable jobs with high wages, good work conditions, and job stability, among other advantageous features, and secondary sector jobs, or those less desirable jobs with low wages, poor work conditions, job instability, and other unattractive features. The stability of primary sector jobs, it is argued, fosters strong social relationships with peers in the workforce. In contrast, in secondary sector jobs, as a result of job instability, workers fail to develop strong ties to their coworkers and the workplace.

The implications of this for crime and violence are critical, according to scholars such as Crutchfield (1989) and Crutchfield and Pitchford (1997), who theorize a link between labor-market characteristics and violence. They identify two mechanisms by which labor-market characteristics may be associated with crime. The first occurs at the individual level among workers and reflects core arguments of social control theory. Secondary sector jobs, they argue, have low wages, low skill levels, and bleak promotion prospects. As such, they are less likely to produce the kinds of job commitments (commitments to conventional lines of behavior) and interpersonal ties to coworkers (attachments to conventional others) enjoyed by primary sector occupations. Moreover, secondary sector workers have more idle time and are often unconcerned with losing their jobs or losing respect among coworkers. In contrast, primary sector workers receive higher wages, have more complex jobs, and have promotion and career prospects. As such, they are likely to develop strong job commitments and ties to coworkers and are less likely to engage in behavior that will jeopardize their jobs or employment relations (Crutchfield et al. 2006, p. 203). Given the nature and characteristics of their employment, and in line with social control theory, secondary sector workers are more likely to engage in crime.

The second mechanism linking labor-market characteristics and crime operates at the neighborhood level and occurs when large numbers of marginally employed individuals live in proximity to one another. Residential segregation and other forces such as discrimination produce high concentrations of jobless and secondary sector male workers, who move in and out of the labor force, work less than full time when employed, and develop few commitments to jobs. Freed from these commitments and constraints, these individuals have time on their hands with little to do. It is argued that such workers frequently pass the time by hanging out on street corners, in pool halls, bars, and nightclubs – places that can be staging grounds for violent encounters. In some neighborhoods, then, concentrations of marginally employed secondary sector workers result in a "situation of company" in which high concentrations of young men experiencing unstable employment are conducive to violence (Crutchfield et al. 2006: 204).

At this point, race has not factored into the equation. Where does race enter in and how does this all relate to crime and the racial composition of communities? Race becomes salient because, due to discrimination, inequality, lower education levels, and less social capital, blacks are disproportionately employed in the secondary sector labor market. Crutchfield and his colleagues (2006) argue that, collectively, these factors and processes produce neighborhoods with high concentrations of African Americans, secondary sector workers, and unemployed workers and are at risk for high rates of violence. They describe this as a four-step process:

> First, macro processes of dual labor markets and residential segregation give rise to neighborhoods with concentrations of marginally unemployed Black adults and youth. Second, high rates of job instability and the accompanying low levels of economic resources in such neighborhoods give rise to a "situation of company," in which high concentrations of jobless and marginally employed Black males have time on their hands and weak institutional commitments. Third, labor instability and low levels of economic resources result in increased violent crime through the presence of these "situations of company." Finally, collectively these processes produce the association of neighborhood racial composition with violent crime. (p. 204)

Empirical research largely supports the proposition that labor instability and violent crime rates are related (Crutchfield 1989; Crutchfield et al. 1999). In particular, researchers find that relationships between poverty and violent crime and income inequality and violent crime depend heavily on the distribution of workers in the primary and secondary occupational sectors and upon levels of unemployment (Crutchfield 1989, p. 505). What has not been empirically demonstrated, however, is whether (and to what extent) the effect of racial composition on crime across neighborhoods can be accounted for by labor-market characteristics – until very recently. In their analysis of Seattle neighborhoods, Crutchfield et al. (2006) examine the relationships among neighborhood racial and ethnic composition, labor-market indicators, and violent crime rates, examining the extent to which labor-market concentration accounts for racial differences in neighborhood rates of violence. Among the many findings, they show that the effects of racial composition are, to some extent, explained by labor-market characteristics in line with the theoretical arguments described above.

An important caveat must be noted here. Although Crutchfield et al. (2006) find that labor instability helps to account for the racial composition-crime relationship across Seattle neighborhoods, it does not *completely* explain why rates of violence are higher in African American (and Latino) communities than elsewhere. They argue, "Even after taking into account labor instability and social disorder, African American and Latino neighborhoods still have higher rates of violence than other neighborhoods" (p. 217). This raises an important question: What else may account for these racial differences in levels of violence? Crutchfield and his colleagues (2006) encourage researchers to consider the role of culture and cultural processes, which brings us to a final theoretical approach related to crime and the racial composition of communities.

Cultural and Subcultural Theories

Both the social disorganization and labor-market perspectives constitute structural theories that address the link between racial composition and crime rates across neighborhoods by focusing on the crime-producing characteristics of communities such as poverty and joblessness. Cultural and subcultural theories take a different approach. Scholars advocating a cultural approach tend to go in many directions but perhaps the most popular are arguments related to a "black subculture of violence." Cultural theories of this sort range from the more liberal, oppression/power/anomie-based theories to those that criticize black culture, attacking permissive values which have undermined personal responsibility and character. The more liberal theories, which I discuss below, identify a subculture of violence where, due to a variety of structural conditions, residents resort to violence to defend status, honor, or reputation.

Perhaps the most recognized culturally based perspective related to crime and the racial composition of communities comes from Elijah Anderson (1999). In *Code of the Street*, Anderson's landmark ethnography of inner-city Philadelphia communities, he argues that as a result of worsening conditions in inner-city communities over the last several decades, black youth in disadvantaged communities have created a local social order – or street code – comprised of their own set of norms and rituals of authenticity. This street code articulates norms and characterizes social relations among residents, particularly with respect to violence. The street code helps us understand why rates of violence are so high in many impoverished African American communities.

According to Anderson (1999), among those who ascribe to the street code, respect typically forms the core of the person's self-esteem, especially when alternative avenues of self-expression are limited. One way respect is acquired is by developing a reputation for being violent, by creating a self-image based on "juice" (p. 72). On the streets, the image one projects is paramount and must communicate the message that one is capable of violence when the situation requires it. At the top of the street code hierarchy is the "crazy," "wild," or "killer" social identity. In Anderson's study, he found that youth often created altercations with the sole purpose of building "juice" and to be known as having a reputation for being quick tempered. According to this perspective then, violence is considered the single most critical resource for achieving status among those who participate in street culture.

Creating a credible violent reputation not only commands respect but also serves to deter future assaults. However, when challenges do arise or do transgressions occur, violence is viewed as an acceptable, appropriate, and even obligatory response. If a person is assaulted, for instance, it is essential in the eyes of his peers and others to seek revenge. As a result, youth who ascribe to the code often feel constrained to pay back, or seek revenge, after a transgression. Their identity, self-respect, and honor are tied up with the way they perform on the streets during and after such encounters.

Violent social control, or what Black (1983) long ago referred to as self-help, is directly related to the availability and effectiveness of authoritative parties of dispute resolution such as the police. Self-help crimes are more likely where law is less accessible, for example, in poor minority communities where residents are often at odds with the police. When called, the police may not respond, which is one reason, Anderson argues, many residents feel they must be prepared to defend themselves and their loved ones against others. Kubrin and Weitzer (2003) document this in their study of extremely disadvantaged communities in St. Louis. They found that problems confronting residents were often resolved informally – without calling the police – and that neighborhood cultural codes supported this type of problem solving, even when the “solution” was a retaliatory killing.

It is often quite difficult to empirically test the core cultural arguments laid out by theorists such as Anderson. However, work by Stewart and Simons (2006) attempts to do just that. Using two waves of data from 720 African American adolescents in 529 neighborhoods, they investigated whether neighborhood context, family type, and discrimination influenced adoption of the street code among youth. More relevant to the discussion above, they also assessed whether adoption of the street code mediated the effects of neighborhood context, family characteristics, and racial discrimination on violent delinquency. Consistent with Anderson’s arguments, they found that neighborhood disadvantage, living in a “street” family, and discrimination significantly predicted adopting street code norms. Moreover, the street code mediated about one-fifth of the effect of racial discrimination and about 4 % of the effect of family characteristics on violent delinquency. In short, the results of their study were generally consistent with Anderson’s code of the street thesis.

In understanding the linkages among race, place, and violence, and in explaining why rates of violence are higher in African American communities, perspectives such as Anderson’s emphasize the importance of local culture in the production of violence. It is important to stress, however, that neighborhood structural conditions generate this subculture; in this sense, cultural differences are themselves adaptations to broader structural inequality, which itself is an outgrowth of social, political, and economic forces such as globalization and deindustrialization, redlining and residential segregation, punitive criminal justice policy, and discrimination.

Challenges and Future Directions

Despite significant advances in research on crime and the racial composition of communities over the past several decades, much of which is linked to the three theoretical perspectives just discussed, obstacles and challenges remain. As is obvious from the review of the literature in this area, apart from those studies which measure racial heterogeneity and a handful of others that incorporate additional racial and ethnic minorities into their studies, the vast majority of research refers to the effects of percent black on crime rates, ignoring other racial categories. To some extent, this is understandable given that most large cities in the USA have had only small numbers of other minority groups such as Asians or Latinos. In such situations, then, percent black – by default – has reflected the racial heterogeneity in the area. However, in recent decades with increased mobility and migration, the population composition in communities throughout the USA is changing. Beyond black and white differences, many larger (and some smaller) cities are becoming racially diverse. Further, this change is magnified in some regions of the country such as the Southwest and West Coast, where percentages of Hispanics and Asians have increased dramatically. For these reasons, scholars increasingly underscore the need to move beyond black-white comparisons in neighborhood crime studies (see, e.g., Peterson et al. 2006; Peterson and Krivo 2010). An obvious next step, then, would be to broaden the racial focus as a means to deepen our understanding of just how race, ethnicity, and crime are related across communities.

Yet this may be difficult without first addressing a more fundamental (though often neglected) issue in neighborhood research on racial composition and crime: how “race” and “ethnicity” are defined. Although these terms are frequently used interchangeably, in fact, race and ethnicity is not the same thing. Definitions of race include populations that are believed to be distinct in some ways from other populations based on real or imagined *physical* differences such as skin color or facial characteristics. An individual is typically externally classified (i.e., someone else makes the classification) into a racial group rather than the individual choosing where they belong as part of their identity.

For many social scientists, though, race is more than just biology – it is a social construct or invention that changes as political, economic, and historical contexts change. That is, race is considered a socially significant category, one that is vital to understanding the organization and consequences of social relationships. In this sense, race is, among other things, a sorting mechanism for marriage, dating, and friendship and an organizing device for mobilization to maintain or challenge systems of racial stratification.

Although related to race, ethnicity refers not to physical characteristics but social traits that are shared by a human population. These social traits include things like nationality, tribe, religious faith, shared language, shared culture, and shared traditions. Unlike race, ethnicity is not usually externally assigned by other individuals. The term ethnicity focuses more upon a group's connection to a perceived shared past and culture.

Definitions of race and ethnicity are critical because they identify classifications of offenders (and victims), and thus impact conclusions about the linkages among race, ethnicity, and crime across communities. Moreover, difficulties created by the absence of a uniform definition of race are compounded by the data collection practices of government agencies and programs. As LaFree and Russell (1993) explain by way of example, "... both the US Census Bureau and the National Crime Survey allow respondents to provide data on their racial origin. Neither organization applies uniform criteria (e.g., mother's race) or requires external verification. Likewise, the Uniform Crime Reports (UCR) do not provide law enforcement officials with standard guidelines for coding racial origin" (pp. 281–282). Also in UCR data, ethnicity is ignored all together as Hispanics are classified as white, nonwhite, or other race. Of course compounding all of this is the fact that multiracial Americans are one of the fastest growing demographic groups in the country. How are these individuals to be defined and classified?

Another important challenge for researchers who study crime and the racial composition of communities deals with teasing out race versus class effects in neighborhood crime studies. This is due, in large part, to the significant overlap between race and socioeconomic status in the United States. Because of racial inequality, neighborhoods with high concentrations of African Americans are also the most disadvantaged, experiencing greater social isolation from mainstream society, joblessness, family disruption, and less exposure to conventional role models and fewer working- and middle-class families to serve as social buffers in the community. In fact, it is argued that racial differences in concentrated disadvantage are so strong that the "worst" urban contexts in which whites reside are considerably better than the average context of black communities (Sampson 1987, p. 354).

According to many scholars, the structural conditions in black communities are central to understanding the large observed racial differences in urban crime and likely explain why racial composition – especially as reflected in percent black – and crime are associated in neighborhood studies. The argument follows that if similar conditions prevailed in white neighborhoods, they too would exhibit high crime rates. As plausible as it may be, this thesis has gone largely untested because whites rarely live in extremely disadvantaged communities, rendering it difficult to obtain appropriate samples of black and white neighborhoods for comparison.

One exception to this is a study by Krivo and Peterson (1996), who examine racial composition and neighborhood crime in Columbus, Ohio (one of only a handful of cities that has a sufficient number of poor black *and* poor white neighborhoods for comparison). They compare crime levels in extremely poor and disadvantaged black and white neighborhoods with crime in their counterparts. They find that racial differences in structural disadvantage largely account for black-white differences in crime rates across communities. In other words, the fact that black communities are more likely than white communities to experience poverty, social isolation, joblessness, and a lack of conventional role models explains higher rates of crime in those communities. This finding implies that neighborhood racial composition may not matter as much as previously believed.

At the same time, however, they also find that structural disadvantage does not *completely* explain the racial gap in violence. Crime rates for racially distinct neighborhoods generally approach one another when structural conditions are controlled but they do not become equal – similar to what Crutchfield and his colleagues (2006) found with respect to racial composition, labor-market conditions, and rates of violence in Seattle. That notable gaps in violence remain unaccounted for by racialized community conditions suggests that a focus on differentiation in structural conditions *within* white, African American, and other-race neighborhoods may be limited. In this context, in subsequent work, Peterson and Krivo (2010) question the following: How much do the distinct spatial contexts in which white, African American, and other-race neighborhoods are located contribute to varying crime rates within these communities?

Using crime and census data for 9,593 neighborhoods in 91 large cities for the year 2000, Peterson and Krivo (2010) examine how inequality in *nearby* neighborhoods contributes to patterned racial and ethnic differentials in crime. They argue that a common feature of many African American neighborhoods, whatever their internal character, is proximity to communities with characteristics typically associated with higher crime rates, such as disadvantage and residential instability. In contrast, white areas are often surrounded by neighborhoods where crime-promoting conditions are absent and factors that discourage crime, such as external community investments, are prevalent. In line with these arguments, a key finding of their study is that white neighborhoods benefit from the dual privileges of low internal disadvantage as well as embeddedness within a context of other white and advantaged areas, whereas African American (and other minority) neighborhoods suffer a double jeopardy: they are at risk of greater violence stemming from their own internal – often highly disadvantaged – character *and* they bear the brunt of isolation from violence-reducing structures and processes because they are surrounded by disadvantaged areas.

Peterson and Krivo's (2010) study begins to address the challenges associated with teasing out race versus class effects in neighborhood crime studies and offers an additional explanation for why racial composition and crime remain associated across neighborhoods, even after controlling for disadvantage, labor-market characteristics, and other related factors. Yet researchers still have a long way to go before we are able to fully understand the dynamics that underlie the race, place, and crime nexus in US communities.

Related Entries

- ▶ Disadvantage, Disorganization and Crime
- ▶ Neighborhood Effects and Social Networks
- ▶ Race and Ethnicity in Social Disorganization Theory
- ▶ Race, Ethnicity, and Youth Gangs
- ▶ Social Network Analysis and the Measurement of Neighborhoods
- ▶ Spatial Models and Network Analysis

Recommended Reading and References

Anderson E (1999) Code of the street. Norton, New York

Black D (1983) Crime as social control. Am Sociol Rev 48:34–45

Crutchfield RD (1989) Labor stratification and violent crime. Soc Forces 68:489–512

Crutchfield RD, Pitchford SR (1997) Work and crime: the effects of labor stratification. Soc Forces 76: 93–118

Crutchfield RD, Glusker A, Bridges GS (1999) A tale of three cities: labor markets and homicide. Sociol Focus 32:65–83

Crutchfield RD, Matsueda RL, Drakulich K (2006) Race, labor markets, and neighborhood violence. In: Peterson RD, Krivo LJ, Hagan J (eds) The many colors of crime: inequalities of race, ethnicity, and crime in America. New York University Press, New York, pp 199–220

Gans H (1968) The balanced community: homogeneity or heterogeneity in residential areas? People and plans: essays on urban problems and solutions. Basic Books, New York, pp 166–181

Greenberg S, Rohe WM, Williams JR (1985) Informal citizen action and crime prevention at the neighborhood level: executive summary. US Department of Justice/National Institute of Justice, Washington, DC

Kornhauser R (1978) Social sources of delinquency. University of Chicago Press, Chicago

Krivo LJ, Peterson RD (1996) Extremely disadvantaged neighborhoods and urban crime. Soc Forces 75:619–648

Kubrin CE (2000) Racial heterogeneity and crime: measuring static and dynamic effects. In: Chekki DA (ed) Research in community sociology, vol 10. JAI Press, Stamford, pp 189–218

Kubrin CE, Weitzer R (2003) Retaliatory homicide: concentrated disadvantage and neighborhood culture. Soc Probl 50:157–180

LaFree G, Russell KK (1993) The argument for studying race and crime. J Crim Justice Educ 4:273–288

McNulty TL (2001) Assessing the race-violence relationship at the macro level: the assumption of racial invariance and the problem of restricted distributions. Criminology 39:467–490

Peterson RD, Krivo LJ (2010) Divergent social worlds: neighborhood crime and the racial-spatial divide. Russell Sage, New York

Peterson RD, Krivo LJ, Hagan J (eds) (2006) The many colors of crime: inequalities of race, ethnicity, and crime in America. New York University Press, New York

Sampson R (1987) Urban black violence: the effect of male joblessness and family disruption. Am J Sociol 93:348–382

Sellin T (1938) Culture conflict and crime. Social Science Research Council, New York

Shaw CR, McKay H (1969 [1942]) Juvenile delinquency and urban areas. University of Chicago Press, Chicago

Smith DA, Jarjoura RG (1988) Social structure and criminal victimization. J Res Crime Delinq 25:27–52

Stewart EA, Simons RL (2006) Structure and culture in African American adolescent violence: a partial test of the 'code of the street' thesis. Justice Q 23:1–33

Warner BD, Pierce GL (1993) Reexamining social disorganization theory using calls to the police as a measure of crime. Criminology 31:493–517

Warner BD, Rountree PW (1997) Local social ties in a community and crime model: questioning the systemic nature of informal social control. Soc Probl 44:520–536

Crime Concentration

▶ Prediction and Crime Clusters

Crime Investigations by the International Criminal Court

Michel De Smedt and Anja Schröder
Office of the Prosecutor, International Criminal Court, The Hague, The Netherlands

Synonyms

ICC; International Criminal Court

Overview

The International Criminal Court (ICC) was established by the Rome Statute and aims to deal with the most serious crimes of concern to the international community as a whole. There are five elements rendering the investigation of war crimes, crimes against humanity, and genocide at the ICC distinct from common criminal investigations: first, the gravity of international crimes creates challenges to cover all crimes at stake and to meet the expectations of all victims. Second, the size and type of the organizations responsible for crimes impacts in different ways on the focus of the investigation. Third, protecting persons at risk due to their interaction with the Court is a particularly challenging requirement. Fourth, the Court highly depends on cooperation from national jurisdictions, given the lack of enforcement mechanisms inherent to the Rome Statute. Fifth, investigations are often performed in the context of ongoing conflicts where the "peace versus justice" debate impacts on the willingness to support investigations. Criminology could support international investigations and prosecutions in various ways: it could assist in measuring scale and impact of crimes. Furthermore, it could develop models to analyze and prove the role of the most responsible perpetrators. In addition, it could map links between organized crime and conflict economies. Finally, criminology could develop tools to assess the effectiveness of the justice system created by the Rome Statute and the ICC as part of that system.

Fundamentals

International criminal justice has come a long way from the International Military Tribunals of Nuremberg and Tokyo via the ad hoc International Criminal Tribunals for Rwanda and the former Yugoslavia, and the hybrid international tribunals for Sierra Leone and Cambodia to the final emergence of the International Criminal Court (ICC), the first permanent international criminal court in the world (Schabas and Bernaz 2011). The ICC, which is based on the Rome Statute, an international treaty adopted in 1998 in Rome, became operational in 2002. The ICC's capacity of being the first permanent international criminal court in the world brings with it one major difference compared with previous tribunals: the former focused primarily and are still focusing a posteriori on events well defined and limited in terms of time and place. By contrast, the ICC has been assigned a dual role: through having jurisdiction over the most serious crimes of concern to the international community, the Court shall not only end impunity by bringing the perpetrators of past crimes to justice. In addition, the Court is intended to deal with ongoing crimes and contribute in this dual way to the prevention of crimes. These crimes may comprise a broad variety of offenses, often occurring in the context of prolonged armed conflicts or protracted violence in different parts of the world.

As a consequence, the ICC's Office of the Prosecutor (OTP), with a workforce of almost 300 staff from over 80 countries, deals with numerous "situations" in most diverse countries at the same time. As of April 2012, seven situations are under investigation before the Court (Uganda, Democratic Republic of the Congo [DRC], Central African Republic [CAR], Darfur, Kenya, Libya, Côte d'Ivoire). Fifteen cases have been brought before the Court, of which six are currently at trial stage. Eight situations are currently under preliminary examination in four different continents (Colombia, Honduras, Guinea, Georgia, Afghanistan, Nigeria, Korea, Mali), meaning that the OTP is analyzing information to determine whether an investigation should be opened or not. In addition, as of 3 October 2011, the OTP had received 9,303 communications in which individuals or organizations have submitted information on alleged crimes potentially falling within the jurisdiction of the Court.

While the ICC's scope may appear relatively broad at first glance, there are important limitations as to when and where the Court can intervene. First, the Court has jurisdiction only over crimes having occurred since 1 July 2002. Next, the Preamble of the Rome Statute proscribes that the ICC only deals with "the most serious crimes of concern to the international community as a whole." These are genocide, war crimes, and crimes against humanity. In addition, the Court will be able to exercise jurisdiction over the crime of aggression once the provision adopted by the Assembly of States Parties enters into force in the future. Furthermore, the ICC can only investigate and prosecute individual persons, as opposed to organizations as a whole. Finally, other than where the Security Council refers a situation, the Court's treaty-based personal and territorial jurisdiction does not extend its reach to all 193 countries in the world. While a number of major States such as the USA, Russia, and China have not yet joined the Rome Statute, the ICC has witnessed a steep increase from 60 States Parties in 2002 to 121 States Parties as of April 2012.

There are three ways for the OTP to become active. First, a State Party can refer a situation to the Court, as happened in the Democratic Republic of the Congo (DRC), Uganda, and the Central African Republic (CAR) (Rome Statute, Art. 13(a), 14). Second, the Prosecutor can open an investigation proprio motu subject to the authorization of the Pre-Trial Chamber (Rome Statute, Art. 15). For both a State Party referral and proprio motu authorization, the Court can only exercise jurisdiction if the alleged crimes occurred on the territory of a State Party or were committed by a national of a non-State Party or on the territory of a non-State Party which has lodged a declaration accepting the Court's jurisdiction as in the case of Côte d'Ivoire (Rome Statute, Art. 12). Third, the UN Security Council can refer any situation to the Court pursuant to its Chapter VII powers under the UN Charter, including in relation to a non-State Party, as happened in 2005 with Darfur, Sudan, and in 2011 with Libya (ICC Statute, Art. 13(b)). Irrespective of how a situation is brought before the Court, in all three instances, the Prosecutor must be satisfied that the information available provides a reasonable basis to proceed with investigations (Rome Statute, Art. 15(3), 53(1)). In any other situation not covered by any of the above mentioned, the Court cannot become active.

The Rome Statute also sets as an admissibility requirement that any case brought forward for prosecution must be of sufficient gravity (Rome Statute, Art. 17(1)(d)). In addition, the ICC can only proceed with a specific case if the State concerned is either inactive or otherwise unwilling or unable to carry out genuine proceedings (Rome Statute, Art. 17(1)(a)–(c)). This touches on one of the most fundamental principles of the Rome Statute, namely, complementarity. The Court is designed to complement national criminal justice systems, not replace them. Therefore, the ICC functions as a court of last resort, which should only become active in the exceptional situations foreseen in the Statute.

All these aspects and additional ones to be discussed in-depth in the following render investigations and prosecutions by a permanent international tribunal distinct from those conducted by either national or hybrid tribunals. This entry will identify five main peculiarities: the gravity of crimes, the size and type of alleged responsible groups, the protection challenges, the international cooperation, and the "peace versus justice" debate. It will also assess to what extent criminology could become more involved in addressing these challenges, despite the fact that it has been rather reluctant to study international crimes so far (Haveman and Smeulers 2008). Finally, one remark is in order: the fact that the jurisprudence is still developing creates a specific challenge as such. However, this aspect will not be dealt with here, as it is not idiosyncratic to international investigations, but rather related to the emergence of a new legal framework.

Four fundamental principles define the OTP's strategy: positive complementarity, focused investigations and prosecutions, addressing the interests of victims, and maximizing the impact of the Office's work (OTP 2010a). The OTP's action is further guided by three key notions: independence, impartiality, and objectivity (OTP 2010b). That is, the OTP shall act independently of "instructions from any external source" according to Art. 42(1) Rome Statute. Furthermore, from the notion of impartiality implicit in Art. 21(3) of the Rome Statute, it follows among others that the Office shall "apply consistent methods and criteria irrespective of the States or parties involved or the person(s) or group(s) concerned" (OTP 2010b). Finally, objectivity in accordance with Art. 54(1) Rome Statute means that the OTP must investigate incriminating and exonerating circumstances equally. As ICC Prosecutor Fatou Bensouda has stated, objectivity aims at establishing the truth, as opposed to securing "convictions at any cost" (Bensouda 2010).

Key Issues

Gravity of International Crimes

According to the Rome Statute, the cases selected for investigation and prosecution have to be of sufficient gravity. The OTP considers four indicators for assessing gravity in determining whether an investigation shall be opened. These are scale, nature, manner, and impact of the crimes committed. In the case of ICC crimes, the scale tends to be massive. For instance, in the DRC, the OTP announced in September 2003 that it had identified the crimes in the Ituri region as warranting the focus of an investigation, with reports indicated the particular gravity of the crimes, resulting in inter alia an estimated 5,000 civilian deaths since 2002. The massive scale of the precursor to this violence in the DRC is further evident in the reported total number of both direct and indirect conflict deaths in the whole DRC between 1998 and 2002, which has been estimated at 3.3 million people (Geneva Declaration on Armed Violence and Development 2008). In the case of the postelection violence in Kenya in 2007–2008, which happened over a period of only a few weeks, 1,200 persons were allegedly killed, in addition to at least 1,000 reported rapes and 450,000 persons displaced. Apart from the scale of the crimes, the manner of commission is considered for the assessment of gravity. In Ituri, the Office considered information on summary executions, burning of people alive, physical mutilation, and the specific targeting of vulnerable groups, in particular women and children. Examples of extreme brutality can also be found in other situations. In the Bemba case, victims were "raped at gunpoint, in public or in front of or near their family members" (Pre-Trial Chamber II 2009). In Darfur, babies were reportedly burnt in the presence of their families. Finally, impact refers to the consequences of a given crime. This includes social, economic, and environmental damage of crimes; the intention of spreading terror among the civilian population; or crimes committed with the aim or consequence of increasing the vulnerability of civilians, as alleged by the OTP in the Banda and Jerbo case concerning an attack against peacekeepers (OTP 2010b).

Nonetheless, gravity can sometimes be difficult to measure. Quantitative data, such as figures of people killed or raped, are often rare and, if available, may be difficult to ascertain. This is due to a variety of factors. First, some areas may not be covered by any monitoring mechanism. This may be the case for certain locations not accessible in times of ongoing conflict. Potential bias may also result from a possible interest of the parties to a given conflict to underreport crimes by one side and overreport those by another. Moreover, there may be a reluctance to report crimes altogether, which is particularly true for incidents of sexual violence. Here, an assumed systematic underreporting can be traced back to a number of reasons, including social stigma attributed to rape victims in certain communities. This is the case, for instance, in Libya, where victims of sexual violence have reportedly been at risk of reprisals and stigmatization. As a consequence, it can often be difficult to assess the full scale of sexual violence allegedly committed during the armed conflict of 2011. While the credibility of information, particularly numeric information, may often require scrutiny, the reliability of sources may often be equally hard to assess. In this context, methodologies derived from criminology could help to further refine indicators for assessing gravity, including models for measuring scale and impact on societies affected by international crimes. See for an example thereof the work of John Hagan (Hagan et al. 2005) who won the Stockholm Prize in Criminology in 2009 for his work.

In well-functioning national jurisdictions, it is hard to imagine that serious crimes are not investigated or that they go unpunished. In situations of massive violence that the OTP faces, however, there will often be large amounts of information concerning multiple perpetrators at varying levels of responsibility committing large-scale and widespread crimes against numerous victims. It is not the responsibility or role of the Prosecutor to investigate and prosecute each and every criminal episode within a situation. This is not only practically unfeasible: it would also run counter to the notion of a system of courts combating impunity through complementary action at the international and national level, as foreseen in the preamble of the Statute, and the primary responsibility of those States for the investigation and prosecution of such crimes (Rome Statute, Preamble; OTP 2003). The OTP, thus, must exercise discretion in determining which cases should be selected and prioritized for investigation and prosecution. As a consequence, the OTP must strategically focus its activities on a limited number of cases at the highest levels of responsibility and on the most serious crimes. This selection process comes with positive and negative implications. On the one hand, maintaining focus allows expeditious investigations and prosecutions while limiting the number of persons put at possible risk due to their interaction with the Court. It also prevents excessive spending of resources by investigating too broadly. On the other hand, a limited pool of witnesses may give rise to problems for a case should a witness cease cooperating at a later stage. Therefore, it is necessary to balance the use of documentary and other evidence with the value of witness testimony and the reliability and stability of such witnesses. Focused investigations may further mean that the cases will not capture the full extent of all crimes suffered by all victims nor enable prosecution for all possible perpetrators. While it remains to be seen how the ICC judges will respond to the OTP's focused approach, it is safe to assume that the OTP cannot meet the expectations of all victims. However, it does certainly seek to ensure in the cases it brings that the incidents selected for prosecution are reflective of the main types of victimization.

In the case of massive crimes, the number of direct victims and indirectly affected communities demanding justice from the Court is high with victims approaching the Court in a single situation often in thousands. Expectations must thus be managed from the beginning not only by clarifying the limitations of the mandate (such as temporal jurisdiction) but also by explaining the prosecutorial policy of the Office. In this context, it is important to note that under the Rome Statute, victims have for the first time in the history of international criminal justice the possibility of taking an active role as witnesses, participants in the proceedings, and beneficiaries of reparations. However, due to the Office's policy of focused investigations and prosecutions, there may be victims who suffered harm as a result of crimes other than those actually included in the charges selected for prosecution. The interests of these victims, who therefore do not qualify for participation in the pretrial or trial proceedings in a particular case, need to be addressed by the Office in other ways. In addition to outreach sessions in affected communities in situation countries such as Uganda or DRC, victims may present their views early on through submitting information on crimes or participation in town hall meetings, which have at times contributed to the definition of incidents and charges brought forward by the Prosecution. Second, through considering the broader impact of crimes, the Office seeks to address the interests of a wider community of victims. Third, at the reparations stage, the OTP favors a wider approach to allow participation of victims and representations from or on behalf of victims and other interested persons who suffered harm as a result of crimes other than those included in the charges selected for prosecution (OTP 2010a). In this context, it should be noted that the ICC, as the first international criminal tribunal awarding reparations, will have to decide on how to provide former child soldier victims with compensation in the case of the former Commander-in-Chief of the *Forces patriotiques pour la libération du Congo*, Thomas Lubanga Dyilo. In addition, stimulating supplementary solutions to close the impunity gap will be necessary, either through national proceedings or alternative mechanisms such as truth and reconciliation commissions. A comprehensive strategy is needed to determine how situation countries can (be helped to) deal with all crimes committed.

Perpetrators of International Crimes

Massive crimes require extensive planning and usually involve a broad variety of direct and indirect perpetrators from a State apparatus or an organization. For instance, three State armies and 20 armed groups were involved in Congo Wars over the last two decades, the remnants of which can be witnessed in the ongoing conflict in eastern DRC. In the Darfur conflict, one State army, eight government-allied forces, and 14 armed opposition groups have been active. The strength of such groups can range from 50 to 50,000 fighters. Besides quantitative criteria, groups vary in their nature, particularly their level of organization, which may range from network-like, temporary structures to militias up to a fully fledged state apparatus.

The share in the commission of gravest crimes needs to be carefully investigated and evaluated for each group against the gravity threshold to identify the most responsible perpetrators. In the situation of Kenya, for instance, the Office has investigated simultaneously persons affiliated with both the government and opposition side. In the situation of Libya, the ICC's first Prosecutor, Luis Moreno-Ocampo, pointed to allegations of crimes not only by pro-Gaddafi forces but also by NATO forces and forces affiliated with the National Transitional Council. In the situation of Côte d'Ivoire, the Office has emphasized that the investigation of alleged crimes committed by different parties to the conflict will be conducted irrespective of political affiliation. However, both focused investigations and resource limitations may also require prioritization. For instance, regarding the complex conflict in Ituri, in which many militias have allegedly committed crimes, the OTP focused its investigation on those militias allegedly responsible for the most serious crimes, which resulted in cases concerning the *Union des Patriotes Congolais* ("UPC") and its military wing, the *Forces Patriotiques pour la Libération du Congo* ("FPLC"), and the *Force de Résistance Patriotique en Ituri* (FRPI) and the *Front de Nationalistes et Intégrationnistes* (FNI). In Uganda, upon analyzing the gravity of crimes allegedly committed by different armed forces and groups, the OTP found that alleged crimes by the LRA were of highest gravity. The OTP therefore started with an investigation of the LR-A. Case selection is thus based on a number of legal and policy principles, including overriding principles based on objectivity, impartiality, and independence, as well as considerations of gravity focused on the most serious criminal episodes.

A similar process is applied once a specific organization or group has been identified for investigation. The OTP has stated in its 2003 policy paper that "as a general rule, *the Office of the Prosecutor should focus its investigative and prosecutorial efforts and resources on those who bear the greatest responsibility*, *such as the leaders of the State or organisation allegedly responsible for those crimes*" (OTP 2003). Accordingly, individuals investigated and prosecuted before the ICC have included senior civilian leaders and military commanders, two Heads of State, and one former Head of State. In some cases the focus of an investigation by the Office may go wider than high-ranking officers, if investigation of certain type of crimes or those officers lower down the chain of command is necessary for the whole case. Nonetheless, as a general rule, the Office encourages national prosecutions, where possible, for the lower-ranking perpetrators or otherwise seeks to work with the international community to ensure that the offenders are brought to justice by some other means. The goal is to develop comprehensive anti-impunity strategies combining international and national efforts to combat impunity. If former perpetrators are not dealt with through comprehensive strategies, which may involve a variety of transitional justice tools, there is a risk that they may rejoin criminal subcultures or migrate to third countries to form newly emerging criminal networks elsewhere. Here, criminology could research further offender integration in relation to perpetrators of international crimes at different levels of responsibility.

To link suspects to a given crime, the guiding model for investigations by the OTP puts emphasis on a comprehensive crime analysis enabling one to piece together crime patterns and chains of command and to collect the type of evidence needed to establish nature and degree of responsibility (OTP 2003). While a fundamental requirement in this context is to prove charges beyond reasonable doubt, the criminal responsibility of leaders is not always easy to establish. For instance, in many conflict-ridden environments, sexual violence is a frequently reported phenomenon. However, the context of leadership responsibility for individual incidents is not always clear: rape may be committed as an opportunistic crime by private citizens or members of an armed group or state apparatus against the backdrop of a general breakdown of the rule of law or disciplinary failures, or rape may be committed or condoned as an organized campaign by a state apparatus or organization to further specific policy objectives.

A further challenge arises from proving criminal responsibility at the highest level. The more sophisticated a group, the more its members will cover their traces in order to make issues of attribution difficult. In other situations, a criminal network may benefit from a conflict in undetected or underreported manners. Ignoring such a link may result in organized crime reinforcing itself in a conflict area, rendering it also more difficult to fight it at the national level, particularly in countries far removed from the conflict.

In this area, criminology has a twofold task: first, it can assist in bringing together various disciplines including organizational thinking, organized crime models, and military organizational models to better understand how the most responsible perpetrators fulfill their role in a given organization and to define different ways of proving criminal responsibility. Second, criminology could build on the work of Edwin Sutherland and other criminologists to apply the study of crimes by corporations and State crime to the linkage between organized crime, international crimes, and armed conflict.

Protection of Persons at Risk

The OTP often operates in areas of ongoing conflict or post-conflict situations. This comes with a variety of challenges including poor infrastructures coupled with often remote and inaccessible locations, volatile security situations, and a precarious information security environment. In addition to operational challenges, cultural differences need to be taken into account. For instance, the Office has to communicate effectively with witnesses in local languages, some of which have no corresponding words for legal terminology, while finding qualified professional translators requires exceptional efforts. However, while many of these challenges do not differ from those encountered by other agencies working in the same environments, the Court faces one distinct challenge. This is the legal duty to protect persons at risk on account of their interaction with the court, predominantly victims and witnesses. As the OTP has increased its activities and judicial proceedings have evolved from opening investigations towards actual trials and conviction, the number of persons threatened due to their interaction with the Court has also increased, as well as the capabilities of suspects to track, monitor, and influence witnesses. The risk of threats and actual victimization of witnesses implies that the OTP has to approach witnesses without exposing them, which includes identifying safe sites for interviews, securing discrete transportation for investigators and witnesses, or checking for possible relationships between drivers and hotel owners with suspects.

While witness protection is a challenge in national jurisdictions as well, it becomes even more complex when one cannot rely on a well-established law enforcement structure. Take the case of Darfur, in which two arrest warrants have been issued against the Head of State, President Bashir. Due to security concerns, the OTP had to conduct investigations into this situation without being able to secure access to witnesses and crime scenes inside Darfur. Instead, it collected information from witnesses in third countries. This scenario arises when one operates in ongoing conflicts and post-conflict situations without state support or when one is operating globally in multiple places with limited resources. A particular difficulty is related to the fact that the OTP is usually investigating the powerful, i.e., members of armed groups controlling a given territory or members of a state apparatus. Here, the Office is dealing with alleged perpetrators who see themselves as (and are de facto) in control of the law. In addition to challenges posed by powerful suspects, individuals at risk due to their interaction with the Court such as witnesses sometimes may not fully appreciate the dangers posed by the security situation or have a different sense of risk due to their ingrained behavior in living in a conflict zone, which may ultimately lead them to undermine protective measures. In addition, there is a risk that certain individuals may see protection through the Court as a way to improve their own circumstances. Inevitably, security concerns may influence the speed of an investigation. While theoretically speaking, quickly securing an arrest warrant could have a deterrent impact in restraining the further commission of crimes, unfortunately the execution of an arrest warrant is not always automatic.

Cooperation with National Jurisdictions and Other Institutions

Cooperation is the fourth aspect rendering investigations and prosecutions by the OTP distinct. Cooperation is needed for a number of reasons. Next to the general support to the Court's decisions, operational cooperation from national authorities and international organizations is a core requirement for conducting investigative activities such as gathering information and collecting evidence, dealing with logistic challenges in the field, responding to security issues, or building capacity several thousand kilometers away from the ICC Headquarters in The Hague. Overall, OTP operations benefit from cooperation with a broad range of partners, including referring States, third States involved in conflicts related to the situation under investigation, States Parties, international organizations, war crimes units of national jurisdictions, and other experts. For instance, in the situation of Libya, as of 3 October 2011, the OTP had conducted 50 missions to 15 different countries. In other situation countries such as the DRC, to facilitate missions, cooperation mechanisms had to be established with other partners including the peacekeeping mission and other relevant organizations.

In addition to support needed on the ground, the OTP has benefited in the past from its network with national law enforcement agencies and other specialized organizations such as national forensic institutions. At the same time, the OTP has also provided assistance where possible and subject to witness security considerations to national jurisdictions, such as in the case of the *Forces Dé mocratiques pour la Libération du Rwanda* (FDLR), an armed group operating in the DRC, where the OTP successfully contributed to the investigation by the German authorities of two senior leaders living in Germany, while the OTP itself investigated a third senior leader residing at the time in France. Also, as part of its policy of positive complementarity, the OTP has provided assistance to the Ugandan authorities in preparation of their first domestic war crimes case before the newly established International Crimes Division of their High Court, against an alleged mid-level LRA commander.

While these are positive examples of cooperation, the interaction with national jurisdictions is hampered at times by a number of factors: first, there are the ordinary challenges of law enforcement cooperation. These constraints are equally known to other types of international investigations such as organized crime investigations. A recent unpublished study conducted under the guidance of Interpol together with the ICC identified major challenges to international investigations including different legal frameworks, different capabilities and resources, incompatibility of procedures, lack of trust, and cultural obstacles. The study further identified factors that can improve such cooperation including the ability to assess the impact of transnational crime, strategic investment into international cooperation including training programs, and other ways of promoting interpersonal contact as well as joint planning (Police Academy of the Netherlands 2011).

The level of support by a given country or organization can have considerable impact on the ability of the OTP to investigate and prosecute. While the duration of an investigation depends on a number of factors such as focus and complexity of the investigation, the assistance from national jurisdictions or absence thereof can be decisive. Nonetheless, the OTP has also demonstrated that it can pursue its mandate in the face of intransigence or active opposition from the State concerned: for example, in both Darfur and Libya, where it showed that it was able to pursue investigations against the highest level of responsibility without the assistance of the territorial State.

The most critical challenge for prosecuting international crimes, however, is the execution of an arrest warrant. As often pointed out, the ICC lacks an own police force. In fact, the Court depends entirely on the assistance of States to ultimately fulfill its arrest mandate. The level of support in the area of arrests depends on both the ability of a given country to control its own territory and its willingness to surrender a suspect to the Court. There have been positive examples, such as three nationals of the DRC who were surrendered to the Court by the DRC's national authorities, or Belgium, which implemented an arrest warrant against the President and Commander-in-Chief of the *Mouvement de libération du Congo*, Jean-Pierre Bemba Gombo. Another positive example is the surrender of former President Laurent Gbagbo by the national authorities of Côte d'Ivoire following a warrant of arrest issued under seal in November 2011. Yet warrants remain pending against Sudanese President Bashir, regarding which the Pre-Trial Chamber has issued several decisions regarding noncompliance following the failure by some States Parties to cooperate with the Court. With regard to Libya, the OTP has received support from Interpol through three “red notices” for the arrest of ICC suspects Muammar Gaddafi (until his death), Saif Al-Islam Gaddafi, and Abdullah Al-Senussi. The OTP has also been able to secure the voluntary appearance of nine suspects before the Court, namely three rebels from Darfur and six public figures in Kenya. Despite concerted efforts, however, a total of 11 suspects sought by the ICC were still at large as of April 2012. The execution of arrest warrants, nonetheless, is a perennial challenge for all international courts and tribunals. The arrests of Radovan Karadzic and Ratko Mladić and their surrender to the International Criminal Tribunal for the Former Yugoslavia, for example, occurred 13 and 16 years after their first indictment in July 1995, respectively. Nonetheless, they were ultimately brought to face justice.

While the OTP’s timing of an arrest warrant application will in the first instance depend on the status of the evidence collected, other factors will also come into play. These include the window of opportunity to actually arrest a suspect.

Despite such challenges, international criminal justice has witnessed the gradual emergence of a paradigm shift away from the rule of power towards the rule of law. Needless to say, such a shift creates tensions. As first ICC Prosecutor Luis Moreno-Ocampo has pointed out, “Reality has demonstrated that the Office’s independent decisions have triggered conflicts of interests for States. Leaders who are using crimes to retain power have criticized the Court and managed to mobilize some international support to this end. States Parties have struggled to prioritize their commitment to international justice over more immediate economic or political interests” (Moreno-Ocampo 2011).

Against the backdrop of a general lack of knowledge of what the Court can and cannot do, coupled with high expectations about justice being brought about, insufficient levels of cooperation may ultimately lead to the perception of double standards.

Peace Versus Justice?

Finally, the fact that ICC investigations and prosecutions often take place against the backdrop of ongoing conflict routinely triggers an old debate revolving around a difficult question: what is more important – peace or justice? This debate is inevitable considering that the OTP may be investigating and prosecuting in an arena where other actors are focused on political settlements to end conflicts, on securing access for the provision of humanitarian assistance, or on enabling implementation of cease-fire agreements or the deployment of peacekeeping personnel. It is often presented as a dilemma difficult to resolve, but the dilemma itself may also often be illusory and intentionally instrumentalized by those in power to avoid justice. While supporters of the paradigm “peace first, justice second” argue that peace is more important to achieve security, proponents of the paradigm “no peace without justice” emphasize that if justice is not ensured at all levels, tensions may often continue to fester and foster to new violence based on previous patterns of impunity for violent crimes. The long-term logic of favoring short-term solutions needs also to be considered: what deterrent message is conveyed to the perpetrators of genocide, crimes against humanity, or war crimes by signaling that they can ensure impunity for their crimes by holding peace a hostage to threats to commit more crimes? It is worthwhile mentioning here that the UN Security Council referrals of two situations to the ICC, namely, Darfur and Libya, have both stressed the need to account for past crimes. As described in the OTP policy paper on the Interests of Justice of September 2007, “The issue is no longer about whether we agree or disagree with the pursuit of justice in moral or practical terms: it is the law. Any political or security initiative must be compatible with the new legal framework insofar as it involves parties bound by the Rome Statute” (OTP 2007). Moreover, in an April 2009 report on mediation, UN Secretary-General Ban Ki-Moon advised all mediators that they must take into account and respect the ICC jurisdiction and action in accordance with the Rome Statute: “Where serious crimes have been committed, pursuing international justice during mediation can generate considerable tension and affect the outcome, since indicted parties may cease cooperation and actively obstruct the process. Ignoring the administration of justice, however, leads to a culture of impunity that will undermine sustainable peace. Now that the International Criminal Court has been established, mediators should make the international legal position clear to the parties. They should understand that, if the jurisdiction of the ICC is established in a particular situation, then, as an independent judicial body, the Court will proceed to deal with it in accordance with the relevant provisions of the Rome Statute and the process of justice will take its course" (UN Security Council 2009).

Future Directions

In light of the numerous challenges, it seems that the ICC is still at the beginning of a long-term process to bring an end to impunity. While the Court has recently completed its first trial in the case against the former Commander-in-Chief of the *Forces patriotiques pour la libération du Congo*, Thomas Lubanga Dyilo, and nears completion in two others, many of the challenges the OTP faces will continue to exist. This includes high expectations towards and knowledge gaps regarding the Court, protection challenges, and cooperation issues, as well as the recurring peace and justice debate. Furthermore, at a time of an already high caseload, the Court must effectively respond to demands by some States Parties for budget restraints in an overall stretched economic climate. Both the first and current prosecutor have noted that the OTP has surpassed the limit of resources in comparison to the number of cases.

Despite numerous challenges outside of its control, the ICC and with it the Rome System, which celebrated their 10th anniversary in 2012, will and must be ultimately judged based on the ability to achieve the goal of the preamble of the Rome Statute, namely, the determination to end impunity and contribute to the prevention of the most serious crimes in the world. In this context, criminology could play a leading role in assessing the effectiveness of the two complementary protagonists of the Rome System, i.e., the Court on the one hand and States Parties on the other. Here, academic research may assist in developing performance indicators to find answers to a number of open questions, for instance, as follows: is the Court having a deterrent effect in modifying behavior? Has the existence of the Court started to lead to an increase of national proceedings, as a result of complementarity? Is the Court being perceived by the broader public as bringing and safeguarding justice?

Successful investigations and prosecutions of genocide, crimes against humanity, and war crimes require the integration of a broad variety of academic disciplines and fields of work usually not found. Those most relevant in the current context are international relations, investigative work, security studies, prosecutorial approaches, and organizational studies. Criminology is the field which, due to its idiosyncratic, interdisciplinary orientation, would be well suited to assist in improving our understanding of international crimes and their perpetrators, our ability to prove responsibility of the most responsible ones, including criminal networks whose operations are sustained and fuelled by conflicts around the world.

Finally, if the ultimate goal of the ICC is, according to the principle of positive complementarity, not to intervene due to genuine national proceedings, criminology will be needed to conduct research on how to prepare national jurisdictions not only to deal with gravest crimes but also to prevent them from happening in the first place. As noted in the OTP's policy paper of 2003, "the absence of trials by the ICC, as a consequence of the effective functioning of national systems, would be a major success" (OTP 2003).

Related Entries

- ▶ Crimes of the Powerful
- ▶ Development of International Criminal Law and Tribunals
- ▶ Genocide
- ▶ International Sentencing
- ▶ Network Analysis in Criminology

▶ Organized Crime, Types of
▶ Resistance to State Crime
▶ Scientific Evidence Before International Criminal Tribunals
▶ Social Network Analysis of Organized Criminal Groups
▶ State Crime
▶ Transnational Exchange of Forensic Evidence
▶ Victims of State Crime
▶ Victims and the International Criminal Court

Recommended Reading and References

Bensouda F (2010) Challenges related to investigation and prosecution at the international criminal court. In: Bellelli R (ed) International criminal justice. Law and practice from the Rome statute to its review. Ashgate, Farnham/Burlington, pp 131–142

Geneva Declaration on Armed Violence and Development (2008) Global burden of armed violence. http://www.genevadeclaration.org/fileadmin/docs/Global-Burden-of-Armed-Violence-full-report.pdf. Accessed 4 Jan 2011

Hagan J, Rymond-Richmond W, Parker P (2005) Criminology of genocide: death and rape in Darfur. Criminology 43(3):525–561

Haveman R, Smeulers A (2008) Criminology in a state of denial – towards a criminology of international crimes: supranational criminology. In: Smeulers A, Haveman R (eds) Supranational criminology: towards a criminology of international crimes. Intersentia, Antwerp/Oxford/Portland, pp 3–26

Moreno-Ocampo L (2011) Address to the assembly of states parties, tenth session of the assembly of states parties. New York. http://www.icc-cpi.int/NR/rdonlyres/7F1DD418-2E1A-4216-BFA5-D8F618AFAE95/284107/ASPLMOFinalrevised_2_.pdf. Accessed 4 Jan 2012

OTP (2003) Paper on some policy issues before the office of the prosecutor. http://www.icc-cpi.int/NR/rdonlyres/1FA7C4C6-DE5F-42B7-8B25-60AA962ED8B6/143594/030905_Policy_Paper.pdf. Accessed 4 Jan 2012

OTP (2007) Policy paper on the interests of justice. http://www.icc-cpi.int/NR/rdonlyres/772C95C9-F54D-4321-BF09-73422BB23528/143640/ICCOTPInterestsOfJustice.pdf. Accessed 3 Jan 2012

OTP (2010a) Policy paper on victims' participation. http://www.icc-cpi.int/NR/rdonlyres/9FF1EAA1-41C4-4A30-A202-174B18DA923C/281751/PolicyPaperonVictimsParticipationApril2010.pdf. Accessed 4 Jan 2012

OTP (2010b) Draft policy paper on preliminary examination. http://www.icc-cpi.int/NR/rdonlyres/9FF1EAA1-41C4-4A30-A202-174B18DA923C/282515/OTP_Draftpolicypaperonpreliminaryexaminations04101.pdf. Accessed 4 Jan 2011

Police Academy of the Netherlands (2011) Summary report: pearls in policing, The Hague. http://spl.politieacademie.nl/Portals/0/docs/Internationaal/PEARLS%20SUMMARY%20REPORT%202011.pdf. Accessed 13 Apr 2012

Pre-Trial Chamber II (2009) Decision pursuant to Article 61(7)(a) and (b) of the Rome statute on the charges of the prosecutor against Jean-Pierre Bemba Gombo (ICC-01/05-01/08). Public document. http://www.icc-cpi.int/iccdocs/doc/doc699541.pdf. Accessed 4 Jan 2012

Rome Statute of the International Criminal Court (1998) UN doc. A/CONF.183/9, 2187 UNTS 90, entered into force 1 July 2002

Schabas WA, Bernaz N (eds) (2011) Routledge handbook of international criminal law. Routledge, London/New York

UN Security Council (2009) Report of the secretary-general on enhancing mediation and its support activities (S/2009/189). http://daccess-dds-ny.un.org/doc/UNDOC/GEN/N09/278/78/PDF/N0927878.pdf?OpenElement. Accessed 10 May 2012

Crime Linkage Analysis

▶ Linkage Analysis for Crime

Crime Location Choice

Wim Bernasco[1,2] and Stijn Ruiter[1]
[1]Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam,
The Netherlands
[2]Department of Spatial Economics,
VU University Amsterdam, Amsterdam,
The Netherlands

Synonyms

Discrete spatial choice

Overview

Most behavior of interest to social scientists is choice behavior: actions people commit while they could also have done something else. In geographical and environmental criminology, a new framework has emerged for analyzing individual crime location choice. It is based on the principle of random utility maximization as developed in economics. It integrates the study of spatial crime distributions with journey-to-crime research, and it is used to explain the offender's choice of where to commit an offense. It allows the analyst to simultaneously assess the role of location attributes and the role of these attributes in relation to offender characteristics, such as their age, ethnicity, gender, criminal experience, and where they live. Initial applications of the model have shown that the decision of where to commit an offense can successfully be described as a function of characteristics of the decision-maker (i.e., the offender), the potential crime target locations, and their interactions, including the distance that separates them. Other applications have established that physical and social barriers inhibit the journey to crime, that railways facilitate the journey to crime, and that offenders are more likely to offend near former anchor points (past homes). New developments in the area of crime location choice include a focus on small spatial units of analysis and the assessment of spatial spillover effects.

Introduction

Many research questions in the social and behavioral sciences, including criminology, deal with understanding and predicting individual choices. Political scientists aim to understand why people vote and what makes them choose a particular political party (Palfrey and Poole 1987). Sociologists want to understand what makes people decide in favor of a particular education, occupation, or marriage partner (Jepsen and Jepsen 2002). In marketing research, understanding and predicting consumer choice is core business (McFadden 1980). Transportation researchers aim to find out what it is that makes commuters choose to travel by bicycle, car, bus, or train (Train 1980). Behavioral ecologists try to find out what influences a nonhuman animal's choice of where to forage, rest, or reproduce (Krebs and Davies 1993).

Making choices requires agency, that is, the capacity of actors to make decisions. That capacity does not necessarily imply consciousness of the choice process on the part of the decision-maker. Behavioral ecologists, for example, study nonhuman animals' choices but do not make any assumptions about the mental processes that give rise to these behaviors.

Choice is also a central concern in criminology, victimology, and criminal justice research. The following questions provide some examples. How do judges or juries decide on whether a suspect is guilty or not, and on the type and the severity of a sanction? What makes people decide in favor of perpetrating crime? What makes them select a particular victim?

This entry is about crime location choice, which addresses the questions where offenders go to commit crime and why they go there instead of somewhere else. The aim of the chapter is to provide a concise but complete discussion of the various issues involved in using the discrete choice framework to understand the spatial decision making of criminals. The remainder of this text first discusses how the issues of crime location choice have been addressed in the literature. It subsequently outlines random utility maximization theory and the discrete choice framework. The section that follows discusses what has been learned substantively from recent applications of the discrete choice framework to crime location choice. The final section discusses pending issues that have yet to be resolved and most of which require additional empirical research.

Different Approaches to Study Crime Location Choice

Crime location choice addresses the questions where offenders commit crime, and why there instead of somewhere else. Before the discrete choice framework was introduced in the geography of crime, there were three separate approaches to the study of crime location choice (Bernasco and Nieuwbeerta 2005). Each of these approaches is characterized by a specific unit of analysis and a specific dependent variable.

The *offender-based* approach uses either the offender or the offense as the unit of analysis. The dependent variable is the length of the journey to crime, which has always been operationalized as the distance between the home of the perpetrator and the place where the offense was perpetrated. The most basic question answered is the question how far from home offenders perpetrate their offenses. A somewhat more complicated question is whether and how steep the offending intensity decreases with the distance from home. More generally, this question addresses the form of the distance decay function, because it has also been claimed that this distance function is upward sloping near the offender's home (within a "buffer zone") and only starts to slope downwardly at greater distance. Still more complicated questions compare the distance between different types of offenders and different types of offenses. Empirical research has shown that most offenders commit their offenses rather close to their homes, that there is distance decay, and that the average distance traveled varies across offense types and across offender types. The theoretical basis of the offender-based approach is small. It starts from the assumption that travel is costly and that for that reason offenders prefer to travel as little as possible to the locations of their offenses. Because this implies that most offenders would perpetrate crime on their doorsteps, and this appears not to be the case, an additional hypothesis has been introduced which states that offenders do not commit crimes close to their own homes for fear of recognition by victims or bystanders.

The offender-based approach is limited for establishing where offenders perpetrate crime and for understanding that choice. It is limited because at any given distance, there are a number of alternative locations (all located on the circle around the offender's home) that can be selected, and the offender-based model does not provide any further help in understanding the actual location that was chosen. It is also limited because it assumes that distance (and thus travel cost) is the main criterion that determines where offenders perpetrate crime (in addition, offenders would avoid perpetrating very close to their homes for fear of being recognized by victims or bystanders). Thus, the offender-based approach ignores that location choices may be driven by geographic variations in the availability of criminal opportunities (i.e., the presence of suitable victims and targets and the absence of formal and informal guardians) or by characteristics of the offender (e.g., awareness space).

The *target-based approach* uses potential target locations as units of analysis. The dependent variable is the crime rate or the victimization rate at the target location. This approach actually includes the bulk of studies in geographic criminology in which the crime locations are aggregated to larger areas and related to the characteristics of these aggregated areas, in particular characteristics that signal the presence of suitable victims and targets (e.g., entertainment areas or shopping centers) and the absence of formal and informal guardians (e.g., police presence and likelihood of social control exercised by residents). Thus, while the target-based approach does analyze crime location as a choice outcome, it typically does not consider offender mobility and thus ignores that some areas are vulnerable because they are more accessible to offenders than other areas that are identical in every other respect. For example, a shopping strip may be an ideal place for a late-night street robbery for motivated offenders who live nearby, but if the nearest motivated offender lives 5 miles away, it may not be a suitable location for any offender.

The third approach is the *mobility-based approach*. The mobility-approach uses pairs of geographical locations as the units of analysis. The dependent variable is the number of crime trips from one of both locations to the other. This approach applies spatial interaction models, also known as gravity models (Haynes and Fotheringham 1984), to model volumes of offender travel within and between the neighborhoods, traffic zones, or other geographic entities in the study area (Elffers et al. 2008; Reynald et al. 2008; Smith 1976). These studies estimate regression models of trips from a particular zone of origin to a particular zone of destination. The variables include the number of observed trips, which is the dependent variable, and crime attractors (services and people that attract criminals from elsewhere to the destination), crime producers (services and people that produce criminal motivation in the origin), and impedance variables. Impedance variables measure the "friction" or obstacles that must be overcome to move from the origin to the destination, including the distance, the travel time, or the cultural or ethnic similarity between the origin and the destination. In the mobility-based approach, distance and other impedance indicators are accounted for as choice criteria: They are independent variables that determine, among other things, how likely an offender from a particular origin location is to perpetrate a crime in a particular destination location. Spatial interaction models analyze aggregate flows of crime, but are not able to appropriately model variation across categories of offenders or among individual offenders. For example, to answer the question whether distance differently affects the crime location choice of juveniles and adults, one would have to estimate separate models for juveniles and adults. This is not a tractable option when the analysis is to include more than just a few categories.

The *discrete choice framework* constitutes the fourth approach to the issue of crime location choice. The dependent variable in this approach is the choice outcome, that is, which alternative of a countable set of alternative locations does the offender select, and the unit of analysis is the individual decision-maker.

There have thus far been published eight empirical studies that use the discrete choice framework to analyze crime location choice, including Bernasco and Nieuwbeerta (2005); Bernasco (2006); Clare et al. (2009); Bernasco and Block (2009); Bernasco (2010a); Bernasco (2010b); Bernasco and Kooistra (2010); and Bernasco et al. (2012).

Random Utility Maximization and Discrete Choice

In many disciplines, the *discrete choice framework* (Ben-Akiva and Lerman 1985) has emerged as a powerful and elegant approach to theorize about choice, to statistically model it, and to predict its outcomes. The discrete choice framework is a combination of random utility maximization (RUM) theory and a statistical model of the family of discrete choice models (e.g., conditional logit, multinomial logit, nested logit, and mixed logit models). The models can be used without RUM theory for other purposes than choice modeling, but the major strength of the discrete choice model is that the statistical model is directly derived from RUM theory. Because the theory is a formal (mathematical) theory, its predictions are quite precise and can be tested quantitatively. The intimate link between the theory and the statistical model makes it relatively straightforward to test new elements and conditioning clauses when they are added to the theory. While these features may sound like obvious requirements, most theories in the social sciences are informal and are only loosely linked to the statistical models that are used to test them.

The discrete choice framework is a set of assumptions and methods to model a decision-maker's choice among a set of alternatives that are mutually exclusive and collectively exhaustive. This means that the decision-maker must select exactly one alternative, and that the alternatives do not overlap, so that by choosing an alternative all other alternatives become unavailable.

Most of the assumptions in the discrete choice framework are based on RUM theory (McFadden 1973). RUM theory is the microeconomic theory of behavior in which a random component is added to the utility function. The random component represents incomplete information on the part of the analyst, not on the part of the decision-maker. By making certain assumptions about the distribution of this random component, the theory is directly linked to a statistical model that allows probabilistic statements to be formulated and tested.

The discrete choice framework was developed in the 1970s by McFadden and others working in the field of travel demand, and the first applications of discrete choice were in the study of travel mode choice (i.e., the choice between train, bus, car, or airplane). Later, the model was also applied to the choice of a travel routes and travel destinations (Ben-Akiva and Lerman 1985).

The discrete choice framework defines four elements of a choice situation (Ben-Akiva and Bierlaire 1999):

1. Decision-makers. The decision-maker is the person or agent that makes a choice.
2. Alternatives. The decision-maker must choose one alternative from the choice set, that is, the set of available alternatives that are mutually exclusive and collectively exhaustive.
3. Attributes. Alternatives have attributes that affect the utility that the decision-maker derives from them when they are chosen. The decision-maker evaluates the utility of all alternatives. The decision-makers themselves can also have attributes that may affect the utility they derive from the alternatives.
4. Decision rule. According to RUM theory, the decision-maker chooses the alternative that maximizes (expected) utility (net gain, profits, satisfaction) when chosen.

In further discussing discrete choice modeling, the notation of Train (2009) is followed. For stylistic reasons and clarity of argument, the decision-maker is referred to as a "he" and the researcher is referred to as a "she."

A decision-maker, labeled n, must make a choice among the J alternatives in the choice set. Decision-maker n obtains a level of utility (profits, satisfaction) U_{ni} from alternative i if that alternative is chosen. The decision rule of utility maximization theory asserts that the decision-maker decides in favor of the alternative i if and only if he expects to derive more utility from alternative i than from any other available alternative. Thus, if he decides in favor of alternative i, he must expect to derive less utility from each of the other alternatives.

$$U_{ni} > U_{nj} \forall j \neq i. \tag{1}$$

The utilities are assumed to be known by the decision-maker, but not by the researcher. The researcher only observes the actual chosen alternative i, the set of J alternatives, some attributes a_{ni} of the alternatives, and some attributes d_n of the decision-maker; and she can specify a function V, often called *representative utility* or *systematic utility*, that links these observed attributes to the decision-maker's utility:

$$V_{ni} = V(a_{ni}, d_n) \forall i \tag{2}$$

The researcher incompletely observes utility, so that generally $U_{nj} \neq V_{nj}$. The utility can be written as the sum of representative utility V_{nj} and a term ε_{nj} that captures the factors that determine utility but are not observed by the researcher and that is treated as random.

$$U_{ni} = V_{ni} + \varepsilon_{ni} \tag{3}$$

The probability P_{ni} that decision-maker n chooses alternative i is the probability that the utility associated with choosing i is greater than the utility associated with any other alternative in the choice set:

$$P_{ni} = \Pr(U_{ni} > U_{nj} \forall j \neq i) \tag{4}$$

Substituting Eq. 3 in Eq. 4 yields

$$P_{ni} = \Pr(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \forall j \neq i)$$

$$P_{ni} = \Pr(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \forall j \neq i) \tag{5}$$

This is the most general formulation of the discrete choice model. Statistical models that implement this include not only the workhorses of this family, the conditional and the multinomial logit models, as special instances, but also various others, such as nested logit, mixed logit, and multinomial probit. Details of these models are described in Train (2009). All applications of the discrete choice model to crime location choice have thus far used the conditional logit model, which is also referred to as "the (multinomial) logit model with variables that vary over alternatives." This model assumes that the unobserved utility term ε_{ni} follows an extreme value type 1 distribution, from which the following formula for the probability that decision-maker n chooses alternative i can be derived.

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_{j=1}^{J} e^{V_{nj}}} \tag{6}$$

This formula relates the utility derived the alternative that is chosen (the numerator) to the total utility of all alternatives in the choice set (the denominator).

For computational convenience, representative utility V_{ni} is usually assumed to be linear in the K parameters:

$$V_{ni} = \beta' \mathbf{X_{ni}} = \sum_{k=1}^{K} \beta_k X_{kni} \tag{7}$$

Application of Discrete Choice Models to Crime Location Choice

The application of discrete choice models to crime location choice implies that the four elements of choice situations are specified. A collection of potential crime locations constitutes the choice set. In most applications of the model thus far, this collection of locations is formed by the set of "neighborhoods" in a single city or region. Thus, the problem addressed can be phrased as "given that an offender will commit an offense in the study area, in which of the neighborhoods in the area will s/he commit it?" Substantive theory is needed to specify what the characteristics of potential targets are that make them attractive to offenders.

Thus far, the discrete spatial choice model has been applied mostly to burglary and to robbery. Studies on burglary include Bernasco and Nieuwbeerta (2005); Bernasco (2006) and Bernasco (Bernasco 2010a) in The Hague, the Netherlands; and Clare, Fernandez, and Morgan (Clare et al. 2009) in Perth, Australia. Bernasco (2010a) and Bernasco and Kooistra (2010) studied commercial robbery in the Netherlands, and Bernasco and Block (2009) and Bernasco et al. (2013) studied street robbery in Chicago, USA. Bernasco (2010b) studied burglary robbery, theft from vehicle, and assault.

Target Characteristics

The discrete choice framework was introduced in criminology by Bernasco and Nieuwbeerta, who used data on 548 burglaries committed by solitary (non-co-offending) offenders in any of the 89 neighborhoods of the city of The Hague, the Netherlands, by offenders who were themselves residents of the city. They argued that burglars would be attracted to targets located near their homes (indicated by distance and by distance to the city center) that were affluent (indicated by property value) and accessible (indicated by percentage single-family dwellings) and where they were least likely to be disturbed by residents or bystanders (indicated by low residential mobility and low ethnic heterogeneity). They found that offenders were attracted to neighborhoods located near their own homes and near the city center, to ethnically mixed neighborhoods, and to neighborhoods with a relatively high percentage of single-family dwellings.

Bernasco (2006) extended this analysis by adding data on co-offender groups, by covering a longer period, and by using factor analysis to group neighborhood characteristics into the sets of variables indicating affluence, physical accessibility, and social disorganization. Only the physical accessibility of neighborhoods was found to significantly increase their odds of being targeted.

The spatial anchor point of offenders structures their choices because it defines the distances that they must overcome to reach target locations. Because offenders live in different locations, the distances between their homes and potential targets vary between potential targets (for the same offender) as well as between offenders (for the same potential target).

An important difference between journey-to-crime research and criminal location choice research is that in the former, variation in distance to crime is the explanandum; distance is the dependent variable. In criminal location choice, distance is part of the explanans; the distance to a potential target is one of the arguments that offenders consider when they decide on where to commit an offense.

Barriers

The findings on burglary in The Hague mentioned above already indicate that distance plays an important role in criminal location choices. Distance is actually a poor measure of the impedance encountered when traveling between two places. Arguably, better measures include monetary cost, travel time, and lack of comfort. Studying 1761 burglaries in 292 suburbs in Perth, Australia, Clare et al. (2009) demonstrated that physical barriers (such as rivers or highways that form obstacles for cross-travel) and connectors (such as public transport lines) affect offenders' journeys to crime. They also found that the influence of impermeable barriers increases with proximity of these obstacles to the offender's point of origin.

More abstract barriers to mobility are social in nature. A social barrier deters offenders to commit crime in a potential target area because the potential target area is socially (culturally, ethnically, economically) different from the offender or from the area where the offender lives. A study on Chicago robbers (Bernasco and Block 2009) demonstrated that, controlling for distance from home, offenders displayed a strong tendency to commit robberies in places where the majority of the population was of their own ethnic category. There are two plausible reasons for this finding. First, because birds of a feather flock together, ethnic minorities tend to live together in areas or congregate elsewhere. For this reason, most individuals spend a considerable amount of time in the proximity of their peers, and it is only logical that they commit crime around the places they normally visit. The second reason is that offenders like to disguise their motives and that they may be much less conspicuous in areas where people live who are like they are.

Awareness Space

Crime pattern theory (Brantingham and Brantingham 2008) suggests that for an individual to commit a crime at a place, the place must not only offer criminal opportunities but must also be part of the individual's awareness space, which is shaped by current and recent routine activities. Recent research utilized the discrete spatial choice model to investigate whether offenders are more likely to target areas near their former residences (Bernasco 2010b; Bernasco and Kooistra 2010). It was demonstrated that offenders were not only likely to offend near their current residence, it was also demonstrated that they were more likely to offend near former homes, especially if they had lived there for a long time and had moved away only recently. This type of idiosyncratic spatial knowledge need not be restricted to offender's home locations but may include the homes of their friends and family members, as well as other major anchor points such as schools, workplaces, or entertainment and business areas they regularly visit during presumably legal activities.

Offenders must also learn about the criminal opportunities of places where they commit crimes. Indeed, it has been suggested that the phenomena of repeat victimization and near repeat victimization (whereby in the wake of a crime, victimization risk is elevated in the vicinity of the crime site) may be explained by the postulated tendency of offenders to return to their prior targets (Bernasco 2008; Bowers and Johnson 2004). This tendency might be lower for locations where offenders were arrested.

Small Spatial Units of Analysis

From geography, it is well known that physical and social processes are not constant across spatial scales. The finding that the outcomes of both descriptive and explanatory analyses strongly depend on the size and shape of the spatial units of analysis has been coined the modifiable areal unit problem (or, more neutrally, phenomenon) (Openshaw 1984). As most theories of crime seem to focus on small spatial units, criminologists are urged to collect and analyze data on small spatial units (Weisburd et al. 2009).

In spatial discrete choice, the MAUP has not been given much attention, and studies of location choice, whether in human migration, recreation, or firm location, have typically based the scale decision on convenience and data availability. In crime location research, Bernasco (2010a) and Bernasco et al. (2013) discuss the complexities involved in using small geographical units. The first challenge is mainly computational. Small spatial units of analysis typically imply a large or very large choice set, which may generate computational challenges because the matrix that is used in the maximum likelihood estimation of the parameters has N × J records, where N is the number of decision-makers and J is the mean number of alternatives available per decision-maker. Bernasco et al. (2013) demonstrate that in their analyses of street robbery at census block level, they would have to join a set of 12,000 offenders with a set of 25,000 census blocks, yielding a matrix of 300 million records to be included in the maximum likelihood estimation procedures. Fortunately, the coefficients of a conditional logit model can be estimated consistently on a random subset of the alternatives (McFadden 1978). Therefore, sampling from alternatives has become an agreed standard in applications with large numbers of alternatives, including location choice. Guevara (2010) developed methods to sample from alternatives for more complex discrete choice models, such as nested logit.

The second challenge is theoretical. Small spatial units imply that spatial spillover effects may be more salient, especially if spillover effects display strong distance decay. For example, a retail store may not only attract street robbers to the exact location where it is located but also to blocks across the street and around the corner. A robber may follow a customer from a store in one block to a more isolated place in an adjacent block before he attacks, or wait in ambush around the corner. In analyses at higher levels of aggregation, such as census block groups, census tracts, or neighborhoods, such short-distance effects are much weaker because the large majority of the short-distance spillover takes place within the boundaries of the larger geographical units.

Related Entries

- ▶ Agent-Based Models to Predict Crime at Places
- ▶ Co-offending
- ▶ Crime Mapping
- ▶ Crime Science
- ▶ Econometrics of Crime
- ▶ History of Geographic Criminology Part I: Nineteenth Century
- ▶ Rational Choice Theory

Recommended Reading and References

Ben-Akiva ME, Bierlaire M (1999) Discrete choice methods and their applications to short term travel decisions. In: Hall RW (ed) Handbook of transportation science. Kluwer, Norwell, MA, pp. 5–34

Ben-Akiva ME, Lerman SR (1985) Discrete choice analysis: theory and applications to travel demand. MIT Press, Cambridge, MA

Bernasco W (2006) Co-offending and the choice of target areas in burglary. J Invest Psychol Offend Profil 3:139–155

Bernasco W (2008) Them again? same offender involvement in repeat and near repeat burglaries. Eur J Criminol 5(4):411–431

Bernasco W (2010a) Modeling micro-level crime location choice: application of the discrete choice framework to crime at places. J Quant Criminol 26(1):113–138

Bernasco W (2010b) A sentimental journey to crime; effects of residential history on crime location choice. Criminology 48:389–416

Bernasco W, Block R (2009) Where offenders choose to attack: a discrete choice model of robberies in Chicago. Criminology 47(1):93–130

Bernasco W, Kooistra T (2010) Effects of residential history on commercial robbers' crime location choices. Eur J Criminol 7(4):251–265

Bernasco W, Nieuwbeerta P (2005) How do residential burglars select target areas? a new approach to the analysis of criminal location choice. British J Criminol 45:296–315

Bernasco W, Block R, Ruiter S (2013) Go where the money is: modeling street robbers' location choices. J Econ Geogr 13(1):119–143. doi:10.1093/jeg/lbs005

Bowers KJ, Johnson SD (2004) Who commits near repeats? a test of the boost explanation. Western Criminol Rev 5(3):12–24
Brantingham PJ, Brantingham PL (2008) Crime pattern theory. In: Wortley R, Mazerolle L, Rombouts S (eds) Environmental criminology and crime analysis. Willan, Cullompton, pp 78–93
Clare J, Fernandez J, Morgan F (2009) Formal evaluation of the impact of barriers and connectors on residential Burglars' macro-level offending location choices. Aust N Z J Criminol 42:139–158
Elffers H, Reynald D, Averdijk M, Bernasco W, Block R (2008) Modelling crime flow between neighbourhoods in terms of distance and of intervening opportunities. Crime Prevent Commun Safety 10:85–96
Guevara CA (2010) Endogeneity and sampling of alternatives in spatial choice models. PhD Thesis. MIT, Department of Civil and Environmental Engineering, Cambridge, MA
Haynes KA, Fotheringham AS (1984) Gravity and spatial interaction models. Sage, Beverly Hills
Jepsen L, Jepsen C (2002) An empirical analysis of the matching patterns of same-sex and opposite-sex couples. Demography 39(3):435–453
Krebs JR, Davies NB (1993) An introduction to behavioural ecology, 3rd edn. Blackwell, Oxford
McFadden D (1973) Conditional logit analysis of qualitative choice behavior. In: Zarembka P (ed) Frontiers in econometrics. Academic Press, New York, NY, pp. 105–142
McFadden D (1978) Modeling the choice of residential location. In: Karlkvist A, Lundkvist L, Snikars F, Weibull J (eds) Spatial interaction theory and planning models. North-Holland, Amsterdam, pp 75–96
McFadden D (1980) Econometric models for probabilistic choice among products. J Business 53(3): S13–S29
Openshaw S (1984) The modifiable areal unit problem. Geo Books, Norwich
Palfrey TR, Poole KT (1987) The relationship between information, ideology, and voting behavior. Am J Polit Sci 31(3):511–530
Reynald DM, Averdijk M, Elffers H, Bernasco W (2008) Do social barriers affect urban crime trips? the effects of ethnic and economic neighbourhood compositions on the flow of crime in the Hague, the Netherlands. Built Environ 34:21–31
Smith TS (1976) Inverse distance variations for the flow of crime in urban areas. Soc Forces 54(4): 802–815
Train KE (1980) A structured logit model of auto ownership and mode choice. Rev Econ Stud 47(2): 357–370
Train KE (2009) Discrete choice methods with simulation. Cambridge University Press, New York, NY
Weisburd D, Bernasco W, Bruinsma GJN (eds) (2009) Putting crime in its place: units of analysis in geographic criminology. Springer, New York

Crime Mapping

Spencer Chainey
Jill Dando Institute of Security and Crime Science, University College London,
London, UK

Synonyms

Geographical analysis; Geography of crime

Overview

Crime mapping is the direct application that comes from considering two particular characteristics about crime – it has an inherent geographical quality and crime does not happen randomly. That is, all crime occurs at some location, and it is the features of this location and the nature of the crime that has been committed, including the decision-making of an offender and their interaction with a victim (or other target such as a building), that make the study of its geography useful for law enforcement and public safety. This can include responding to calls for service, identifying crime hot spots, and supporting the investigation of a crime series.

This section begins by discussing the fundamental theoretical principles that underpin crime mapping, offers a brief history of its development, and then illustrates examples of its application. The final section identifies future directions for its development.

Theoretical Fundamentals

The inherent geographical quality of crime can be observed in many theoretical forms. For example, the four dimensions of crime state that for a crime to occur it must have a legal dimension (a law must be broken), a victim dimension (someone or something has to be targeted), an offender dimension (someone has to do the crime), and a spatial dimension (it has to happen somewhere) (Brantingham and Brantingham 1981). This therefore highlights the value that an assessment of the location of the crime can play, alongside the study of the offender and their behavior, and the vulnerability of the person or property that was targeted.

The nonrandom geographical nature to crime can also be explained theoretically by drawing from Routine Activity Theory, Rational Choice Theory, Crime Pattern Theory, and principles of least effort. Routine Activity Theory (Cohen and Felson 1979) states that for a crime to occur, three components must be present. There must be the presence of a likely offender, the presence of a suitable target, and the absence of a capable guardian. These three components must meet in time and space, to formulate the necessary "chemistry" for crime (Felson 1998). Routine Activity Theory therefore states that the risk of crime changes over time with the movement of people throughout their daily routine activities, meaning that the geography relating to these activities helps to determine features relating to victim vulnerability and offender motivation. Most researchers take the existence of likely offenders as a constant in our society and attempt to explain and prevent crime by examining the two remaining components. Victim and target suitability can change over space and time, and the possibilities for introducing suitable guardianship into an area as a way to prevent crime, or to make existing guardians more suitable, offer a means for considering ways for reducing crime opportunities. The consideration of these three components of crime has recast Routine Activity Theory into the Crime Problem Analysis Triangle (Eck 1995) – stating that all crime problems have an offender, victim, and a place component (incorporating both the geographical location and temporal features of the crime), with the notion of guardianship in this model acting as a handler to help reduce the vulnerability of the victim or target.

Rational Choice Theory (Clarke and Felson 1993) provides a framework to consider offender decision-making (and its relationship with space) when a crime opportunity is presented. That is, the commission of crime by a person is a (fairly) rational decision, with the aim being to achieve some sort of desire or goal. This goal may be to derive personal financial gain (e.g., as in a theft of cash), personal pleasure (e.g., stealing a car for joyriding), or through the imposition of power (e.g., through the extension of bullying by stealing possessions off a person or through violence). This criminal decision-making is considered to involve two parts. There is a long-term, multistage decision to become generally involved in criminal activity (criminal involvement decision) and a shorter-term, more immediate decision (the criminal event decision) to grasp an opportunity that is presented (Cornish and Clarke 1986). In a geographical sense, factors relating to social and economic disadvantage and social disorganization (Burgess 1925; Shaw and McKay 1942) may have a role to play in the long-term decision-making process, whereas the immediate decision-making of an offender is primarily based on the environmental cues from the prospective target that can change from place to place.

Crime Pattern Theory provides a useful convergence of Routine Activity Theory and Rational Choice Theory by helping to explain the geographical behavior of an offender, their interaction with potential targets, and offense distribution. That is, it helps examine the "relationship of the offence to the offender's habitual use of space" (Bottoms and Wiles 2002, p. 638). Crime Pattern Theory suggests that offenders are influenced by the daily activities and routines of their lives, so that even if they are searching for a criminal opportunity, they will tend to steer toward areas that are known to them (Brantingham and Brantingham 1984). In their day-to-day activities, they will be watching for targets that have no guardians or place managers (Eck 1995). Crime Pattern Theory considers the notion of awareness space (the areas that are familiar, through carrying out routine activities in these areas), opportunity space (those places which offer a target-rich environment, i.e., where there is a concentration of suitable targets), and the places where these two overlap. This is illustrated in Fig. 1, showing a hypothetical model of an offender's awareness space (consisting mainly of the places they

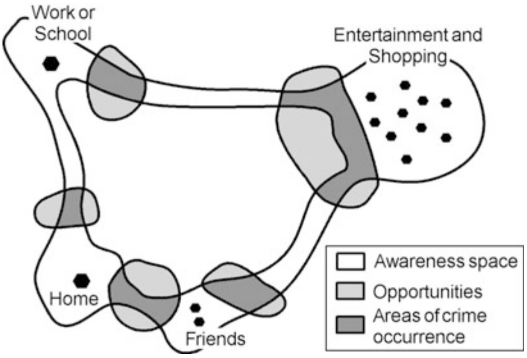

Crime Mapping, Fig. 1 A hypothetical model of the creation of criminal occurrence space where offender awareness space and opportunities coincide (Source: Chainey and Ratcliffe 2005)

routinely frequent and the routes between these places) and the overlap with opportunities for crime. For this offender, they are most likely to commit crime in those locations where their awareness space and the opportunity space overlap. The opportunity space extends to areas outside of their awareness, but for this offender, crime that they commit is unlikely in these areas because they are not familiar with the opportunities, nor comfortable with the higher risk taking that may be involved (i.e., lack of familiarity of escape routes and a higher perceived risk of being caught).

The final fundamental theoretical consideration behind crime mapping is the least effort principle. This states that most journeys we take in our daily activities are short and that the frequency of these activities decays with distance. That is, while we may make long distances to carry out activities, the majority of things we do involve us seeking out trips that minimize this distance. For instance, we may need a carton of milk. While there are many places that sell milk, most people will purchase milk by going to a local store, rather than expending additional effort and travelling to a store some additional distance away. This least effort principle is also evident with offending behavior. Research into offenders' journeys to crime shows that the majority of trips are short (within a few miles) (see Rossmo 2000 for a comprehensive review of journey to crime literature). This concept is also captured in Fig. 1 through the illustration of boundaries to the awareness space. If there was no concept of the least effort principle, the awareness space would extend to infinity in all directions.

These theoretical principles provide the backdrop to the application of crime mapping. However, while these theories help to conceptualize the spatial behavior and interaction between offenders and their targets, there will always be exceptions, e.g., an offender may not commit crime within their own awareness space due to the influence of peers who introduce the offender to new areas. For the majority of offenders, broad behavioral tendencies do exist and can be explained as aggregate criminal spatial behavior. Crime mapping therefore draws from geographical and environmental criminology concepts to help explain why crime occurs, and makes use of geospatial tools (such as Geographical Information Systems – GIS) to capture, analyze, and visually interpret these geographical qualities. Examples of its application are provided in a following section. Before this a short history of GIS and crime mapping is provided to help capture how its use has evolved.

GIS and a Brief History of Crime Mapping

A GIS is "a computer system for capturing, managing, integrating, manipulating, analysing and displaying data which is spatially referenced to the Earth" (McDonnell and Kemp 1995, p. 42). A GIS provides the computing environment within which crime data can be layered with base maps and other geographic data that relate to the landscape and conditions of the area under study. This other geographic information may include population data from a census, the locations of bars and nightclubs, or data describing the local land use patterns. These data can be represented as individual layers that can be manipulated, analyzed, or displayed as separate entities, or could be combined with other layers to be displayed together, integrated to provide a new perspective of the area that they represent, or analyzed against each other to explore relationships.

GIS began to emerge as a discipline in the 1960s, with its origins being in the development of land use applications in Canada, growing then through applications that included the automation of the cartographic drafting process used by mapping agencies such as the British Ordnance Survey and acting as a platform to aid military intelligence gathering, display and analysis from satellite, and other remote sensing imagery (e.g., aerial photography and radar). The use of GIS in policing did not begin to emerge until the 1980s, in large part due to computer hardware prices being too prohibitively expensive before this time, the lack of suitable GIS software that could be used to support policing requirements, and the lack of electronic recording of police records. Even then, it took until the late 1990s for the use of GIS in policing to gather pace, mainly due to being held back by organizational and management problems, issues with sharing information, technical problems, and geocoding issues (see Chainey and Ratcliffe 2005 for a more detailed assessment). Many of these problems are still apparent today, but through developments in the practical use of GIS, several innovators have identified how these can be overcome, and have shared this knowledge across a network of users who engage via forums, conferences, training courses, and publications that since 1998 have begun to capture evidence of how crime mapping can be effectively used to support policing and crime prevention.

Crime Mapping Applications

Crime mapping contributes to policing, law enforcement, and crime reduction by helping to generate a real understanding of criminal activity and the direction in tackling it (Chainey and Ratcliffe 2005). The further reading section provides details of many examples that have captured its operational use – both for tactical and strategic purposes. In this section, examples of its use are provided relating to several of the main operational functions in law enforcement and public safety:

- Performance and accountability – CompStat
- Response – directing calls for service
- Prevention – the analysis of crime hot spots
- Detection – geographic profiling
- Evaluation – through the use of analysis that explores the impact, displacement, and diffusion of benefit from a geographically targeted policing or crime prevention initiative.

Performance and Accountability: CompStat

CompStat (short for Computer Statistics) is an operational management process that combines computer technology, operational strategy, and managerial accountability to determine the provision of crime reduction policing (Walsh 2001). From its early origins with the New York City Police Department, the CompStat idea has been adopted in many police agencies across the world. Figure 2 captures the CompStat process in operation at Thames Valley Police Force in England.

The basic aim of CompStat is to provide detailed intelligence to operational commanders on a regular basis so that they can determine an appropriate crime reduction and law enforcement strategy and then subsequently be held accountable for the success of the strategy. The computer system to analyze and map crime patterns is at the core of the CompStat process (Walsh 2001, p. 352). Crime maps form a useful basis around which those in attendance can visualize operational decisions, resource allocation, and managerial accountability. This can include the use of live and interactive crime mapping as a way to help focus and direct discussions around particular problems, deployed tactics, and performance in relation to the current local crime reduction priorities.

Response: The Use of GIS for Calls for Service

Pinpointing the location of an incident and ensuring that it is identified with minimal ambiguity is vital for a quick and direct response. Many police agencies use gazetteers that are integrated into

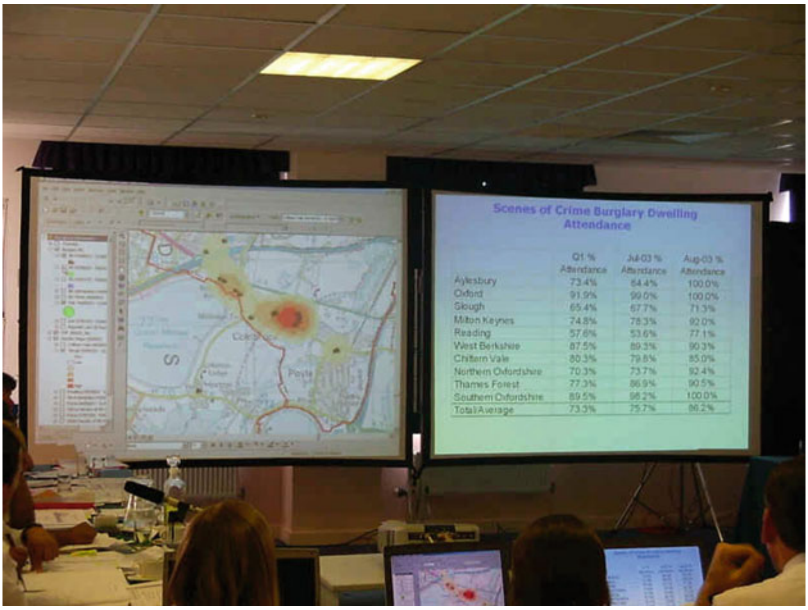

Scenes of Crime Burglary Dwelling Attendance

	Q1 % Attendance	Jul-03 % Attendance	Aug-03 % Attendance
Aylesbury	73.4%	64.4%	100.0%
Oxford	91.9%	99.0%	100.0%
Slough	65.4%	67.7%	71.3%
Milton Keynes	74.8%	78.3%	92.0%
Reading	57.6%	53.6%	77.1%
West Berkshire	87.5%	89.3%	90.3%
Chiltern Vale	80.3%	79.8%	85.0%
Northern Oxfordshire	70.3%	73.7%	92.4%
Thames Forest	77.3%	86.9%	90.5%
Southern Oxfordshire	89.5%	98.2%	100.0%
Total Average	73.3%	75.7%	86.2%

Crime Mapping, Fig. 2 Thames Valley Police (England) CompStat meeting

their mapping systems to search for and view the location of an incident when a call comes in to their command and control center. Increasingly, police forces are also utilizing Global Positioning Systems (GPS) to provide a real-time link to the command center to show the location and details of response teams that are on patrol (e.g., see Home Office 2005). The use of mobile computing technology also helps police agencies organize responses to incidents. In-car computer technology has become the norm in law enforcement agencies in the United States, permitting live access to central databases, allowing for the display of recent events, and permitting the recording of details about incidents while on out on patrol.

Crime Prevention: Hot Spot Analysis

Hot spot mapping is a popular analytical technique used by law enforcement, police, and crime reduction agencies to visually identify where crime is highest, aiding decision-making that determines where to target and deploy resources. Its application has been used to support the operational briefing of police patrols (Home Office 2005; LaVigne and Wartell 2000; Harries 1999), informs the creation of intelligence products and problem solving (Home Office 2005; Chainey and Ratcliffe 2005), has been used as a tool to capture the measurement and analysis of crime patterns for crime auditing purposes in the UK's crime reduction partnerships (Chainey and Ratcliffe 2005; Chainey 2001), and supports performance analysis (e.g., see Chainey and Ratcliffe 2005; Home Office 2005; Walsh 2001). In essence, hot spot mapping is a technique that is used to help determine where crime may happen next, using data from the past to inform future actions. In this sense it acts as a basic technique for predicting where crime may occur, using the premise that retrospective patterns of crime are a useful indicator for future patterns.

There are many different mapping techniques that can be used for identifying and exploring patterns of crime. Attempting to do this via point mapping has become outdated since the proliferation of GIS software and the increasing sophistication of mapping techniques. Three of the most common techniques include the following:

- Thematic mapping of geographic administrative areas (also referred to as choropleth mapping): this approach uses geographic units that are defined for administrative or political use (e.g., police beats, census blocks, wards, or districts). Crimes that are geographically referenced as points can be aggregated to these geographic unit areas and then shaded in accordance with the number of crimes that fall within them. These types of maps are quick to produce and allow for additional information relating to the geographic unit (such as population) to be linked to explore relationships or calculate a crime rate – increasing their versatility for analysis. However, the varying size and shape of most geographical boundaries can lead the map reader to misinterpret the true distribution of crime due to issues associated with the distribution of crime across the geographical division of boundary areas (Chainey and Ratcliffe 2005) and due to issues associated with the Modifiable Areal Unit Problem (MAUP). This is where changes in the boundaries themselves can directly affect the patterns shown on the map. Thematic mapping of boundary areas continues to see widespread application, from being used for comparing the different volumes of unique and repeat burglaries across a study area's census zones (Ratcliffe and McCullagh 2001) to being used for comparing vehicle theft in relation to land use in Overland Park, Kansas (Harries 1999, p. 47). An example of a thematic map, generated for census areas, is shown in Fig. 3a.

- Grid thematic mapping: in order to address the problems associated with different sizes and shapes of geographical regions, uniform grids can be drawn in a GIS and used as the unit of aggregation for point geo-referenced data (see Fig. 3b). This approach does have some limitations (e.g., it suffers from the same MAUP problems outlined above), but its use has been illustrated for identifying vulnerable residences where target hardening was then implemented (Bowers et al. 2001) and for mapping the volume of emergency calls and violent offenses per square mile in North Carolina (LeBeau 2001).
- Kernel density estimation: this technique has become increasingly popular due to its now wide availability in standard GIS software products, its accuracy for hot spot identification (Chainey et al. 2008), and the aesthetic look of the resulting map in comparison to other techniques (Eck et al. 2005). Point data (offenses) are aggregated within a user-specified search radius and a continuous surface that represents the density of crime events is calculated. A smooth surface map is produced, showing the variation of the crime density across the study area, with no need to conform to geometric shapes such grid units. Examples of the use of KDE are now widespread, with popular crime mapping texts showing its use for a variety of applications (see Chainey and Ratcliffe 2005; Eck et al. 2005; Harries 1999). An example of a KDE crime hot spot map is shown in Fig. 3c.

The use of hot spot analysis continues to advance, particularly in terms of exploring how geographic information that represents the underlying causal factors of crime can be modeled alongside recorded incidents of crime. Risk Terrain Modeling is one of these techniques and involves combining separate layers of risk variables (e.g., transport termini, schools, bars, and nightclubs) to produce a composite map showing the presence, absence, or intensity of all risk factors across a study area (Kennedy et al. 2011).

Detection: Geographic Profiling

Geographic profiling is an investigative technique that uses the locations of a linked series of crimes to determine the most probable area of offender residence (Rossmo 2000). Geographic profiling is a method of strategic information management that prioritizes the offender search efforts to maximize the efficiency of serial crime investigations. Until recently, this technique's application has very much focused on serious violent or sexual crimes such as rape and murder. However, it is also now more widely applied to nonviolent, high-volume crime (e.g., criminal damage and burglary) where a series is evident.

Geographic profiling is based on research from the fields of criminology, geography, forensic and environmental psychology, mathematical modeling, statistical analysis, and criminal investigation.

Crime Mapping, Fig. 3 Common hot spot mapping techniques. (**a**) Thematic mapping of administrative units, (**b**) grid thematic mapping, and (**c**) kernel density estimation

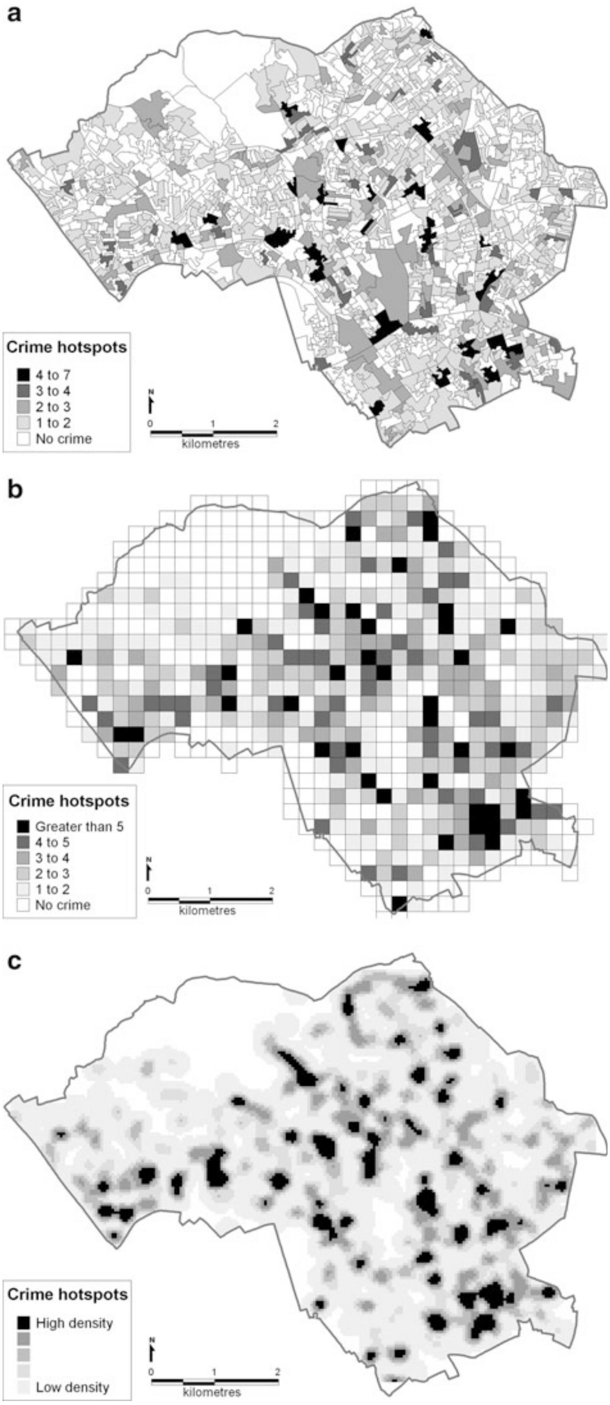
a
Crime hotspots
4 to 7
3 to 4
2 to 3
1 to 2
No crime
N
0
1
2
kilometres
b
Crime hotspots
Greater than 5
4 to 5
3 to 4
2 to 3
1 to 2
No crime
N
0
1
2
kilometres
c
Crime hotspots
High density
Low density
N
0
1
2
kilometres

In a criminal context, each site associated with a single offender (e.g., victim encounter site and attack site) represents a location that is most likely to be present within the offender's awareness space. Using a series of sites linked to the same offender, a probability map can be constructed that predicts the location of an offender's anchor point – usually the offender's residence.

A geographic profile is constructed using a mathematical algorithm that models the offender's likely journey to crime distances and the relationship between crime location sets and offender residence. The process assigns scores to the various points on a map that represent the offender's hunting area. These scores are based on the crime site locations and certainty in the linkages between crimes. The map output of a geographic profile identifies the most probable area of offender residence. The geographic profile can then be used to help recommend investigative strategies that will support the detection of the offender. This could include suspect prioritization based on the cross-referencing of the geographic profile with suspect lists, the targeting of DNA swabbing, or patrol saturation and operational stakeouts. Daniell (2008) provides an example that illustrates the use of geographic profiling that led to catching a serial sex offender in Bath, England.

Evaluation

Many crime prevention initiatives and police operations are area-based: resources are targeted to a particular area in an attempt to deal with its crime problems. Crime mapping offers a useful tool to help monitor if the initiative or operation has been a success by analyzing the before and after picture of the crime levels in the area. With the use of techniques such as the Weighted Displacement Quotient (Bowers and Johnson 2003), an analysis can also be conducted to discover if there has been any displacement or diffusion of benefit as a result of an area-based targeted initiative. For example, in an analysis of alley-gating schemes in Merseyside, England (involving the installation of security gates across alleyways or footpaths that run behind properties), Bowers et al. (2004) used concentric buffer rings around the alley-gating areas to explore if and how far any burglary was displaced (see Fig. 4). The results showed there to be a diffusion of benefit to properties in the surrounding areas. The evaluation also enabled a rudimentary cost-benefit analysis to be calculated, showing that in the areas where the gates had been installed, the cost-benefit ratio was around 1.86 (i.e., a saving of £1.86 for every pound spent).

Future Directions

Advances in computing technology come with it opportunities to apply more sophisticated analysis of geographical patterns and apply new technology that supports improvements in policing, law enforcement, and crime prevention. For example, greater use of Internet-based crime mapping tools can help support the publication of neighborhood-level crime statistics to the public to provide positive messages of reassurance and alert residents to spates of crime in their neighborhood that may place them at a heightened risk of being victimized and the precautions they can take to minimize this risk (Chainey and Tompson 2012).

To date, geographical analysis of crime patterns in policing has focused on the generation of descriptive material that helps to identify geographical patterns of crime (such as crime hot spots), but has yet to fully embrace more sophisticated forms of geographical information analysis. New developments that are expected in crime mapping include the following:

- Analysis of the *significance* of geographical patterns – spatial significance testing identifies geographical patterns that are considered to be particularly unusual by applying the principles of significance testing (e.g., see Chainey 2011).
- A harmonization of the geographical and temporal exploration of crime, so that each is explored together and as a continuum rather than in isolation.
- The development of local spatial regression techniques to help examine relationships between crime patterns and other variables – to explain why crime occurs at certain places.
- The development of techniques that predict where and when crime will occur in the future.
- Spatial modeling processes that test "what if" scenarios such as examining the impact that a crime prevention initiative may have on crime levels (e.g., see Birks et al. 2012).

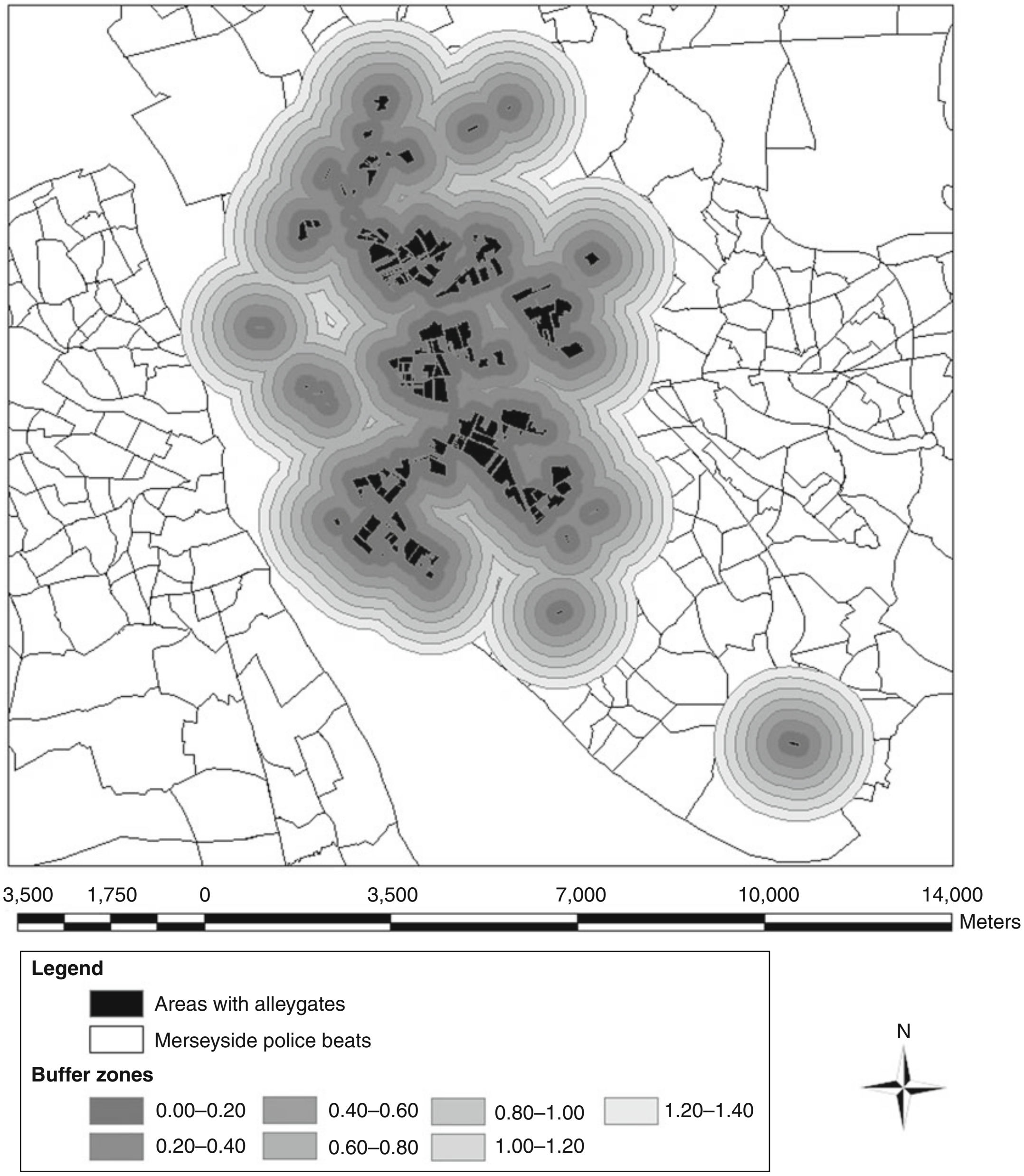
3,500
1,750
0
3,500
7,000
10,000
14,000
Meters
Legend
Areas with alleygates
Merseyside police beats
Buffer zones
0.00–0.20
0.20–0.40
0.40–0.60
0.60–0.80
0.80–1.00
1.00–1.20
1.20–1.40
N

Crime Mapping, Fig. 4 Alley-gated areas in Merseyside and concentric buffers used to detect displacement or diffusion of benefit (Source: Bowers et al. (2004))

Related Entries

- ▶ Agent-Based Models to Predict Crime at Places
- ▶ Applied Geographical Profiling
- ▶ Biological Geographical Profiling
- ▶ Compstat
- ▶ Crime Location Choice
- ▶ Crime Science
- ▶ Criminal Investigative Analysis
- ▶ Early Chicago School Theory
- ▶ Effectiveness of Situational Crime Prevention
- ▶ History of Geographic Criminology Part I: Nineteenth Century

- ▶ Hot Spots and Place-Based Policing
- ▶ How to Make Crime Mapping More Inferential
- ▶ Information Technology and Police Work
- ▶ Longitudinal Crime Trends at Places
- ▶ Longitudinal Studies in Criminology
- ▶ Near Repeats and Crime Forecasting
- ▶ Outdoor Serious Violence: The Role of Place
- ▶ Place-Based Randomized Trials
- ▶ Prediction and Crime Clusters
- ▶ Predictive Policing
- ▶ Problem-Oriented Policing
- ▶ Rational Choice Theory
- ▶ Routine Activities Approach
- ▶ Spatial Models and Network Analysis
- ▶ Spatial Perspectives on Illegal Drug Markets

Recommended Reading and References

Birks D, Townsley M, Stewart A (2012) Generative explanations of crime: using simulation to test criminological theory. Criminology 50(1):221–225

Bottoms AE, Wiles P (2002) Environmental criminology. In: Maguire M, Morgan R, Reiner R (eds) The oxford handbook of criminology. Oxford University Press, London, pp 620–656

Bowers K, Johnson SD (2003) Measuring the geographical displacement and diffusion of benefit effects of crime prevention activity. J Quant Criminol 19(3): 275–301

Bowers K, Newton M, Nutter R (2001) A GIS-linked database for monitoring repeat domestic burglary. In: Hirschfield A, Bowers K (eds) Mapping and analysing crime data – lessons from research and practice. Taylor and Francis, London

Bowers KJ, Johnson SD, Hirschfield AFG (2004) Closing off opportunities for crime: an evaluation of alley-gating. Eur J Criminal Policy Res 10:285–308

Brantingham PJ, Brantingham PL (1981) Environmental criminology. Waveland Press, Prospect Heights

Brantingham PJ, Brantingham PL (1984) Patterns in crime. Macmillan, New York

Burgess EW (1925) The growth of the city: an introduction to a research project. In: Park RE, Burgess EW, McKenzie RD (eds) The city. University of Chicago Press, Chicago, pp 47–62

Chainey SP (2001) Combating crime through partnership; examples of crime and disorder mapping solutions in London, UK. In: Hirschfield A, Bowers K (eds) Mapping and analysing crime data – lessons from research and practice. Taylor and Francis, London

Chainey SP (2011) Advanced hotspot analysis – spatial significance mapping using the Gi* statistic. In: Presentation at the 2011 international crime and intelligence analysis conference. Available at: http://www.ucl.ac.uk/jdi/events. Accessed 17 Apr 2012

Chainey SP, Ratcliffe JH (2005) GIS and crime mapping. Wiley, London

Chainey SP, Tompson L (2012) Engagement, empowerment and transparency: publishing crime statistics using online crime mapping. Policing J Policy Pract

Chainey SP, Tompson L, Uhlig S (2008) The utility of hotspot mapping for predicting spatial patterns of crime. Security J 21:4–28

Clarke RV, Felson M (eds) (1993) Routine activity and rational choice. Advances in criminological theory, vol 5. Transaction Books, New Brunswick, NJ

Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44:588–608

Cornish D, Clarke R (1986) The reasoning criminal: rational choice perspectives on offending. Springer, New York

Daniell C (2008) Geographic profiling in an operational setting: the challenges and practical considerations, with reference to a series of sexual assaults in Bath, England. In: Chainey SP, Tompson L (eds) Crime mapping case studies: practice and research. Wiley, Hoboken

Eck J (1995) A general model of the geography of illicit retail marketplaces. In: Weisburd D, Eck JE (eds) Crime and place. Criminal Justice Press, New York, pp 67–93

Eck JE, Chainey SP, Cameron JG, Leitner M, Wilson RE (2005) Mapping crime: understanding hot spots. National Institute of Justice, USA. Available online at www.ojp.usdoj.gov/nij

Felson M (1998) Crime and everyday life: impact and implications for society. Pine Forge Press, Thousand Oaks

Harries K (1999) Mapping crime: principle and practice. United States National Institute of Justice, Washington, DC. Available online at http://www.ojp.usdoj.gov/nij/maps/pubs.html

Home Office (2005) Crime mapping: improving performance. Home Office Police Standards Unit, London. http://www.ucl.ac.uk/scs/people/academic-research-staff/spencer-chainey/Slides/Home_Office_CrimeMapping. Accessed 17 Apr 2012

Kennedy LW, Caplan JM, Piza E (2011) Risk clusters, hotspots, and spatial intelligence: risk terrain modeling as an algorithm for police resource allocation strategies. J Quant Criminol 27(3):339–362

LaVigne N, Wartell J (eds) (2000) Crime mapping case studies: successes in the field, vol 2. Police Executive Research Forum, Washington, DC

LeBeau JL (2001) Mapping out hazardous space for police work. In: Bowers K, Hirschfield A (eds) Mapping and analysing crime data – lessons from research and practice. Taylor and Francis, London

McDonnell R, Kemp K (1995) International GIS Dictionary. GeoInformation International, Cambridge, UK

Ratcliffe J, McCullagh M (2001) Crime, repeat victimisation and GIS. In: Bowers K, Hirschfield A (eds) Mapping and analysing crime data – lessons from research and practice. Taylor and Francis, London, pp 61–92
Rossmo DK (2000) Geographic profiling. CRC Press, Boca Raton
Shaw CR, McKay HD (1942) Juvenile delinquency and urban areas. University of Chicago Press, Chicago
Walsh WF (2001) Compstat: an analysis of an emerging police managerial paradigm. Policing An Int J Police Strat Manage 24(3):347–362

Crime on Public Transport

Andrew Newton
Applied Criminology Centre, University of
Huddersfield, Huddersfield, UK

Synonyms

Mass transit; Public transit; Public transportation; Rapid transit

Overview

This entry examines how crime concentrates and is potentially specialized within and around the public transport system. It begins by defining public transport for the purposes of this discussion and outlines why this is an area of importance. It also discusses some of the reasons why there is a paucity of evidence into crime on public transport and the unique challenges this system poses for identifying and analyzing concentrations of crime. It describes how public transport networks provide a number of unique settings (places and times) across which crime and disorder can occur. These include in and around stations and stops and on board moving vehicles. The public transport network brings increased accessibility to places, and this creates distinctive patterns of offending. This entry reviews the international evidence for the manifestations of crime in these situations and considers the theoretical reasons and explanations for such events and the extent to which these may be specialized. It then examines the prevention opportunities offered by the crime concentrations and specializations identified, and some emerging trends and future avenues for research.

Introduction

This entry considers crime and disorder on public transport, the ways in which it concentrates and the extent to which it is specialized. It begins by defining public transport for the purposes of this discussion, and emphasizes the importance of reducing crime and disorder and associated fears of personal security on the public transport network. It will then review the empirical evidence for the concentrations and specializations of crime and disorder evident on public transport systems, and assess the theoretical perspectives that help explain these patterns. It will describe how the public transport system provides unique settings (places and times) that subsequently create distinctive crime patterns that encompass a variety of changing environments. These include stations, stops, and interchanges and "en route" and "line of route" offenses, and therefore, a description is given of how the public transport network serves settings that incorporate both static and dynamic situations of crime and disorder (Newton 2005). This entry will explore how concentrations of crime manifest across each of these settings and highlight some of the unique challenges faced when attempting to explain and prevent crime on the public transport network. It will conclude by discussing some emerging trends and avenues for further research in this field.

For the purposes of this entry, the term *public transport* is used to capture what North American readers may think of as "public transit," "mass transit," "rapid transit," or "public transportation" systems. There is no clear consensus of a definition of public transport, and therefore the following are adopted as forming key components of the public transport system:

- Public transport is used to describe a system used by the public, often a means of transporting passengers in mass numbers, generally a for-hire system that occurs across a fixed route or line.

- It consists of a range of transport modes, including railway (railroads, light rail, metro/subway/underground railway, high-speed rail, and intercity rail), buses, trolleybuses, and trams; ferries; coaches; airlines; water taxis, gondolas; and pedicabs.
- Bicycle hire schemes are becoming more popular in urban areas and could be included as part of the public transport system although they are not discussed here.
- In some regions "collective transport" is also considered as a form of public transport, for example, a minibus or fixed group taxi (e.g., South America and Russia).
- "Paratransit" is an expression sometimes used in areas of low demand and for people who need a door-to-door service.
- There are distinctions between multimodal and intermodal systems: a multimodal system accommodates different modes of transport, whereas an intermodal system ensures strategic connections between different modes.
- There is a debate as to whether or not taxis are part of the public transport system although they are not discussed within the context of this entry.

Why Examine and Reduce Crime and Disorder on Public Transport?

There are a number of reasons why understanding and preventing crime on public transport should be of interest. Firstly, public transport has an important role to play in reducing social exclusion, by providing access to facilities such as work, health, leisure, and employment. In Great Britain, for example, the 2011 National Travel Survey estimated that about one quarter of households do not have access to a car. There are also obvious environmental benefits in promoting public transport as a means of sustainable travel, and hence, the use of public transport may be expected to grow over time rather than decrease. A number of surveys have identified that fear of crime and personal security is a major inhibiting factor to the use of public transport, second only in many surveys to reliability and accessibility. Indeed, a report in the UK by the then Department for Environment, Transport and the Region (DETR) suggested that reducing fear of crime could increase patronage by 3 % at peak and 10 % at off peak times (Newton 2004). One of the clear messages from this is the significance of making passengers actually feel safe, and Smith and Clarke (2000) suggest this should be a mandate shared across all transport systems.

It is evident in the literature concerned with the fear of crime, however, that reducing actual levels of crime does not necessarily equate to reductions in people's perceptions of crime risk. The factors that contribute to worries over personal security on public transport are not disparate from those associated with fear of crime outside of the transport network, and the dynamics of vulnerability on public transport include variation by gender, ethnicity, age, familiarity with, and levels of public transport usage (Yavuz and Welch 2010). However, there are some features of the public transport network that are perhaps distinctive. As will become more apparent in this entry, the dynamic nature of the public transport system creates unique environments, through which specific modes of transport traverse, transporting potential targets and victims, on a system that passes through areas with different levels of crime risk, and therefore continuously receives different inputs and outputs over time. This creates a unique, potentially specialized, and certainly concentrated arena within which crime and disorder may occur.

Context and Scope

While the definitions of public transport alluded to earlier suggest that it is a for-hire system of mass transportation used by the public, in its broadest sense the public transport journey should actually be viewed in its entirety, what the then DETR called the "whole journey approach" (Newton 2004). This not only consists of travel on actual vehicles or different modes of transport, but when waiting at a stop, station, or interchange, and also walking to and from stops and stations at the start and end of journeys, or even potentially transferring between different modes of transport. As suggested by the DETR, negative experiences across any part of the journey might actually influence fears of personal security and change travel patterns, perhaps to the extent that passengers may stop using public transport altogether. While the notion of the whole journey approach is crucial for examining crime on the public transport journey, the majority of the limited empirical evidence is focused on the "at stop or *station*," "en route," and "line of route" offenses. Therefore, the content of this entry will naturally focus on these situations in detail – there being a paucity of research into the *walking* parts of the journey. In addition, as much of the literature is focused on bus and rail travel, the scope is also limited to these two modes of transport. However, many of the concentrations and potential specialism's identified and the theoretical concepts used to explain these can be readily applied to other parts of the public transport network and other modes of public transport vehicles. Finally, this entry does not examine security at a wider level than crime and personal security, and therefore, terrorism is excluded from this discussion.

The Nature and Extent of Crime on Public Transport

There is limited information available as to the extent of crime and disorder on public transport, and furthermore, discrepancies exist as to the actual levels of crime and disorder on public transport. The most likely reasons for this include the following: the multiple agencies responsible for maintaining and operating the various systems, the lack of standardized reporting and recording of crime and disorder on public transport, difficulties in analyzing the available data, the different policing and security arrangements in place across the network, and the likely levels of underreporting on the system (Newton 2004).

One of the definitive references in this area (Smith and Cornish 2006) identifies six categories of crime that are typically evident on the public transport network, these being antisocial behavior; crimes against passengers including theft, robbery, and assault; crimes against employees including assault and robbery; vandalism and graffiti; line of route crimes which are offenses along routes that cause delay or affect safety; and a final category of "other offences." In addition, it is useful to distinguish "en route" offenses from those at stations and stops (Newton 2005). This entry will review the empirical evidence of the concentrated nature of these six types of offenses, both in time and place. It demonstrates the spatial and topological patterns of crime evident on the public transport network, and it is contended that these are a result of the movements on the system in terms of offender and targets, and the increased accessibility it provides to the places the network serves. A question this raises is whether this leads to unique opportunities for offenders to specialize on and near to the public transport network.

There are sizeable differences in published figures regarding the extent of crime and disorder on public transport. A study in the UK by the Department for Transport (DfT 2010) suggested that during 2008–2009, there were 12 crimes per million passengers on the bus network and 13 crimes per million on the underground and overground services. This also indicated that the rate of offenses varies by type of crime and analysis of the 2006/2007 figures estimated that the number of offenses (per million passenger journeys) were as follows: on the rail network, 11.9 for violence against the person offenses against passengers, 4.5 against staff, and 39.1 for theft from person and robbery offenses. On the bus network, the rates were 9.0 per million passenger journeys for violence against the person offenses against passengers, 4.1 against staff, and 15.01 for theft from person and robbery offenses. By contrast, a study in Los Angeles suggested an average of 1.55 crime incidents per 100 rides (Loukaitou-Sideris et al. 2002). A 1990s passenger study in England and Wales by Easton and Smith indicated 5 % of passengers had been threatened with violence and 4 % had been the victim of theft, and in Victoria in Australia, a 1990s study by Carr and Wright identified 690 crimes against persons on public transport over

12 months in the context of 300 million passenger journeys per year. While discrepancies will exist in crime rates across countries and different cities, it is suggested those found with the figures presented here are due to differences in recording and reporting practices rather than international or even national variations.

There are additional difficulties with estimating and identifying the levels and extent of crime and disorder on public transport. Levels of underreporting are unknown, and Levine, Wachs, and colleagues examined this in California in the 1980s and found figures may underestimate actual levels of crime by 25–30 times. There are clear difficulties in identifying true levels of crime and disorder on public transport, and, for reasons that are explored next, this is perhaps more problematic than the well established levels of underreporting found in police-recorded crime figures. However, it is believed that levels of crime on public transport have, for the most part, followed the internationally declining levels of crime rates in general (Transit Cooperative Research Program 2009).

The Limited Evidence Base

There are a number of reasons why there have been limited studies into the nature of crime and disorder in public transport. These in part relate to the fragmented nature of regulation across the various systems, the number of agencies and organizations who are responsible for providing and regulating public transport services, difficulties in capturing reliable and accurate information on crime and disorder incidents, the different bodies involved in policing and the security of public transport systems, the levels of underreporting of crime and disorder on these networks, and the analytical challenges faced when attempting to investigate patterns of crime and disorder on public transport (Newton 2004). While the following examples relate to the organization of public transport in the UK, similar fragmented systems exist internationally.

An examination of public transport on the rail network in the UK reveals the wide range of organizations responsible for delivering this system nationally, these include the following: National Rail, the Train Operating Companies who operate under Strategic Rail Franchise Agreements, Network Rail who are responsible for national infrastructure, and the Rail Safety and Security Board (set up to improve the health and safety performance of the railways). There are also some privately owned systems such as light rail (e.g., the London Docklands and the Manchester Metrolink). However, on the rail network there is a consistent set of information collated on crime and disorder as the British Transport Police (BTP) are responsible for policing the rail networks, and they provide monthly and annual figures on crime on the network (both on trains and at stations).

On the bus network, this system is more fragmented in the UK. Outside of London (which is regulated by Transport for London), the deregulation of buses which came into force in 1986 as part of the Transport Act 1985 has resulted in a number of operators who can provide commercial or subsidized services. In the large metropolitan areas, these are coordinated by the seven Passenger Transport Executives (PTEs), and outside of these areas Local Authorities regulate buses. In one of the large metropolitan areas, there were over 50 bus operators providing services, and outside of London there is no requirement to report crime incidents on buses, and this often relies on the goodwill and time of both operators and police forces. Indeed in 1998 only 19 police forces (less than 40 %) provided such information, and only 16 bus companies out of over 100 made returns on staff assaults (Newton 2004). While this information is dated, there is no evidence to suggest that the situation has improved considerably, and a key factor is that in the UK (and elsewhere) there is no statutory dedicated category for "crime on public transport" to be recorded as an incident in its own right by the police. Some operators retain the services of private security contractors to maintain safety and security on their systems which adds a further layer of incident reporting (which may or may not be passed onto the police).

A third and final complication that contributes to the paucity of evidence on crime and disorder on public transport is the unique analytical challenge that it poses, particularly when attempting to identify spatial concentrations of crime. In addition to the lack of consistent information available, the range of settings and environments in the transport network within which crime can occur presents a major analytical challenge. Examples include the following: criminal damage at a stop or station at an unknown time, graffiti or the slashing of a seat on a vehicle at both an unknown time and location as it may only be reported when a bus or train returns to the depot at the end of the day, and pocket picking that occurs as part of a passengers journey, somewhere between departure and arrival. In addition it is very difficult to pinpoint the location of a crime that occurs on a moving vehicle, although there are studies that attempt to analyze dynamic crime events (Newton 2005), for example, the development of "hot routes" or "hot lines" (Newton 2008; Tompson et al. 2009).

Theoretical Perspectives: Why and How Is Crime Concentrated on Public Transport?

The environments of the public transport networks are in some ways distinct from other areas of public space due to the locations they serve, the dynamic nature of the system, and the different inputs and outputs to it. However, there is support within the literature for adopting a number of well-established theoretical criminological perspectives to examine crime on public transport, and this entry now adopts some of these perspectives as a framework for examining the manifestations and concentrations of crime and disorder evident on public transport. A useful starting point is to consider the problem analysis triangle (Clarke and Eck 2005) to conceptulalize patterns of crime and disorder on public transport, and three questions arise here:

- Who are the potential targets and victims, and who can act as capable guardians?
- Who are the likely offenders, and who can act as handlers to restrain the otherwise motivated offenders?
- What places and times (settings) within the public transport network create juxtapositions of offenders and victims that increase or reduce crime opportunities, and who can act as place managers for these?

Victims, Targets, and Guardians

There are a number of potential victims or targets that may experience crime and disorder on the public transport system. The most obvious of these are passengers and staff, and it is important to consider this in the context of all peripatetic staff who may work on the system. The risk to each of these varies across different sections of the transport network. For example, relative to the risk of crime for passengers during busy periods, the risk of crime is likely to be quite different for lone passengers at night and lone staff (at day or at night) who may or may not be carrying cash. There is also risk to the infrastructure of the transport system which includes criminal damage and graffiti of vehicles, stops, and stations, or damage to fixed lines (e.g., obstacles placed on routes which cause damage or obstruct routes) or even theft of copper lines (discussed later under emerging trends).

Offenders and Handlers

There is a paucity of empirical evidence about the nature of offending on public transport, in particular the characteristics of the offenders, and whether they perhaps specialize in crime committed within the public transport network. Who are the potential offenders and why they may be present at particular places and times, and do they use the public transport network as part of their journey to crime? This is an under-researched area that warrants future research as discussed at the end of this entry. A 1990s study by Belanger, one of the few conducted on this topic, examined the extent to which public transport may facilitate the movement of criminals between areas. This study looked at the travel behavior of a sample of serious offenders across the New York City subway system and found that the distance travelled from home or anchor point did vary and that 56 % of offenders travelled within their own borough to commit crime, 29 % travelled to

Manhattan from another borough, and 12 % travelled from one borough to another excluding Manhattan. This raises the question of whether the distance travelled to commit crime differs for those trips that public transport is used and those for which it is not.

Places, Times, and Place Managers?

When examining the places on the public transport network where crime might occur, it is useful to categorize these settings by the three components of the transport journey: the walking, waiting, and en route settings. There is empirical evidence that crime is concentrated at a small proportion of transport stations, stops, and hubs (Smith and Clarke 2000; Loukaitou-Sideris et al. 2002; Newton 2004; Smith and Cornish 2006; Newton and Bowers 2007) and "en route" at particular sections of public transport journeys (Newton 2008; Tompson et al. 2009). These concentrations vary by both crime type and time of day, as will be discussed in more detail below. The *waiting* environment may also include transfer between transport modes within a transport interchange or hub, and this is perhaps distinctive from transfer between transport modes that occurs outside of transport interchanges as this may be classified as part of the *walking* environment. These external transfers and the *walking* environment more generally, for example, from point of departure to stop, or stop to place of arrival, are unlikely to be recorded as part of a public transport journey, and limited information is available on crime and disorder in these components of the public transport journey.

The above discussion has considered who are the likely targets and potential offenders on the public transport and the places, times, and situations where these are more likely to coincide. The opportunities this offers for prevention will be discussed in more detail later but include consideration of the following: who can act as capable guardians for these different targets, who may be an appropriate handler for offenders, and the importance of place management and environmental in reducing crime and disorder on the public transport system.

Crime Pattern Theory and Routine Activities

Two theoretical perspectives that can readily be applied to the public transport network are crime pattern and routine activity theories. Crime pattern theory defines people's activity spaces in terms of three main components: nodes, paths, and edges (Clarke and Eck 2005). The term "node" is used to describe where people (both offenders and victims) travel to and from, and the idea of personal activity nodes closely resembles the ideas of a person's routine activities. These nodes are linked by paths, which would include roads or other transport routes. Edges define the boundaries around nodes, within which people reside, work, or are routinely active in. According to crime pattern theory, crimes are likely to concentrate where offender routine activity spaces intersect with those of suitable targets of crime.

Nodes and paths are particularly relevant in the context of the public transport network, with nodes represented by rail stations and bus stops (and so on), while transport routes correspond to the paths that connect the transportation nodes. Edges may even delineate the boundaries of the public transport network, and there may be regulated and controlled entrances and exits to the system (Newton 2004). Edges can be considered the boundaries of the public transport network, but these become less geographically distinct when considering the walking component of the whole journey approach discussed earlier. Another concept that has been applied to transport nodes is the notion of risky facilities (Clarke and Eck 2005), further discussed below.

The transport network can be thought of as having the potential to shape the activity spaces (and hence the spatial and temporal concentrations) of offenders and potential targets and thus to increase the potential of crime occurring on the network. There is clear evidence, reviewed below, of crime concentrations on the network at two of the critical sections of the network, in and around stations and stops (nodes) and crime en route (paths). Moreover, the way in which the transport network influences the potential for crime is not limited to where it occurs but also what types of crimes are most likely to occur and potentially which offender specialism's may occur in and around the system.

Crime Attractors, Crime Generators, and the Public Transport Network

A final key theoretical perspective is notion of "crime attractors," "crime generators," and "crime neutral" areas (Clarke and Eck 2005). A crime generator is an area that attracts large numbers of people for reasons other than to commit a crime. At particular times and places, this concentration of victims and offenders produces an unplanned opportunity for offending. Crime attractors are places that offenders visit due to knowledge of the area's criminal opportunities, such as bars and prostitution areas, in order to commit a crime. A key question of relevance here is do places on the public transport system act as crime attractors or crime generators. Research has suggested that transport hubs can act as both attractors and generators and that this varies by crime type and by time of day (Smith and Cornish 2006).

Research by Levine and colleagues in California in the '1980s identified that for some crime types, for example rape, homicide, and robbery, the risk was greater when guardianship was likely to be low or where it was impeded by the design of the location (Newton 2004). Examples of this include when there were low levels of pedestrian traffic, low levels of surveillance, and many concealed areas. A study of graffiti and vandalism by Wilson and Healy in the 1980s on the rail network in New South Wales, Australia, found offenders were principally juveniles and that most damage occurs in unsupervised areas during off peak hours. Burrell (2007) highlights that violent crime on public transport tends to happen in the late evening/nighttime when there is less supervision, whereas pocket picking and purse snatching are more frequent during the rush hour. The above examples all demonstrate how transport hubs may act as attractors and or generators of crime, at different times of the day, and for different types of crime.

In order to examine these theoretical perspectives in more detail and also highlight the empirical evidence that exists as to how crime is concentrated and potentially specialized on the transport network, two of the key settings of the public transport network will now be examined in more detail: the first of these is crime in and around stops and stations, and the second is "en route" crime on a moving (and sometime stationary) vehicle.

Crime at Stops and Stations

There are a number of studies that have examined how the design of transport stations can influence levels of crime. Examples include the Paris Metro (Clarke 1996), the Metro in Washington DC (La Vigne 1997), and the metro in Gothenburg, Sweden (Ceccato et al. 2011). Similar studies have examined over ground rail stations in the country of Wales(UK) (Cozens et al. 2004). A clear finding from these studies, particularly in relation to the Washington DC example, was how unusually low crime rates (compared to other systems and local crime rates) could be explained by reference to aspects of environmental design.

In addition to the internal design of stations, there have been studies that examine the link between rail stations and their external environments or nearby surroundings. Studies in Chicago (Clarke 1996; Block and Block 2000) into street robbery in the vicinity of rapid transit stations found that street robbery is concentrated near (but not immediately outside) rapid transit stations and that robbery varied temporally and concentrated late at night (between 11:00 and 12:00 p.m., with a peak time of 2:00 a.m.). Block and Block contend that although the behavior of potential victims and offenders and the time of day and other individual features of each situation will affect risk, it is the place and the surrounding or nearby space that is crucial in bringing all these factors together in what they term the environs of rapid transit stations. A recent doctoral thesis by Herrmann (2011) who examined robbery hot spots in New York provides support for this notion, and this research demonstrated that although all identified hot spots were near subway stations, those that occurred during the daytime (3:00 p.m. peak time) were near high schools but that those occurring in the nighttime (1:00 a.m.) were not, instead being around late night businesses.

In addition to studying crime near railway stations, there have been a number of studies that have examined crime and disorder in and around bus stops and bus stations (Loukaitou-Sideris et al. 2002; Newton 2004; Newton and Bowers 2007; Vu 2009). Loukaitou-Sideris examined the link between criminal activities at bus stops and associated nearby environmental characteristics in a North American city and found that the ten bus stops with the most crime accounted for 18 % of all crimes, and although passenger levels at these stops were high, other nearby stops with high patronage exhibited little or no crime. They identified that there were an abundance of "negative" environmental factors and a general lack of "defensible space" near these high-crime bus stops, including liquor stores, bars, check cashing establishments, seedy motels, pawnshops, vacant lots/buildings, and adult book stores and movie theaters.

This is supported in the UK by Newton and Bowers (2007) who found criminal damage to bus shelters was concentrated at a small number of bus shelters. Examining the criminogenic and socioeconomic characteristics of places, they also identified that bus shelter damage was related in known and predictable ways to characteristics of local neighborhoods. Indeed shelter damage risk was higher in areas with high levels of antisocial behavior, a lack of capable guardianship, and the presence of youths. For example, they found a positive relationship between bus shelter damage and proximity to parks, children's play areas, and schools (particularly those whose unauthorized truancy levels were above the national average), and a negative relationship with the presence of pubs and bars, nightclubs, and off-licenses (late night liquor stores).

An interesting feature of the two above studies if that some of the so-called negative environmental features found to increase crime in the US study, namely, the presence of off-licenses (liquor stores), actually reduced the risk of crime in the UK study. This finding is explained best when considering that the US research examined crimes such as assault and violence, whereas the UK study was concerned with criminal damage to bus shelters. The late night land use associated with these features increased violence and assault in one study, but actually served as a deterrent for criminal damage in another study, by increasing the number of persons in the area, whose presence therefore may have actually served as a form of guardianship against criminal damage to bus shelters. It also serves as a clear example of how spatial crime concentrations vary by type of offense and time of day.

"En Route" Crime

Few studies have examined crime that occurs while a public transport vehicle is moving, predominantly due to the analytical challenges this poses. A 1970s Home Office study by Mayhew examined damage to buses in Manchester (UK) and found that damage was greatest on buses without a conductor, more prevalent on the upper decks of the bus, and especially at the rear of the bus. On buses with a rear staircase, graffiti and vandalism was more prominent upstairs and at the front of the bus. This study concluded that a lack of supervision was an important factor in the occurrence of vandalism and graffiti on buses, supporting the idea that a lack of guardianship or place management on the transport network acts a contributory factor to criminal damage. This study considered the internal environment of a vehicle, but not the external environments it may pass through. There are, however, some studies which have examined how the external environments a moving vehicle passes through may influence crime risk on that vehicle.

Crime and disorder is concentrated on particular routes and sections of routes on the public transport network. A 1980s study by Pearlstein and Wachs examined crime on the bus network in California and concluded that only 88 out of 233 routes (less than 40 %) experienced serious incidents of crime and that crime mostly occurred on routes that traversed areas with high crime rates. They also found crime was disproportionately high during late evenings when violent crimes were more prominent and that theft and robberies were most likely during the rush hour (Newton 2005).

Additional studies, such as those by Levine and colleagues in California in the 1980s, support the notion that incidents of bus crime are highest for bus routes that pass through high-crime areas. Moreover, Newton (2008) identified that not only is bus-related crime positively correlated with levels of crime in the areas a bus traverses, and that the risk of crime on those routes that traverse high-crime areas is greater than on other routes, but furthermore that the risk propensity is heightened on routes that go through high-crime areas and have a higher numbers of stops within these areas (multiple entry and exit points). It should be noted that this risk is not just confined to the passengers and staff who travel on public transport but also to the actual infrastructure of the public transport network.

An analysis of criminal damage to vehicles using data of objects that were projected or thrown at buses (including for example rocks, stones, bricks, and eggs) also found that risk was highest to buses passing through areas that experienced high crime rates (Newton 2004). In this situation the theoretical edges or boundaries to the public transport network become more blurred, and the interaction between the internal (inside the vehicle) and external environments (that a vehicle passes through) is different to the inputs and outputs of passengers and staff to the system (which occurs at transport stops and stations or nodes). The latter can be regulated by controlling the entrance and exit points to the system (by offenders of victims), while the former cannot, and this provides an additional layer of complexity to the study of crime on the public transport system.

Crime Concentrations on Public Transport and Prevention Implications

Perhaps a key significant feature of the concentrations and potential specializations of crime and disorder that are evident is the prevention opportunities these offer. Crime and disorder has been shown to be concentrated within certain components of the public transport network, and these include along particular route sections of journeys, nearby and within certain stations and stops, and within particular parts of stations and interchanges, and these all vary by offense type and time of the day. There is also a clear interaction between the movement of vehicles and the environments through which they traverse, and regulating access to and from the system (the entry and exit points) can serve as a useful mechanism for reducing crime and disorder on the network.

The design, environment, and management of stations and stops can influence crime rates, and there is an established evidence base of strategies and measures that can be adopted to improve the design of public transport infrastructure in order to reduce crime levels (Clarke 1996; LaVigne 1997; Smith 2008; Ceccato et al. 2011) including better lighting and illumination and removing dark corners and hiding places. Furthermore, effective place management of stations (Clarke 1996), the introduction of capable guardians (Newton et al. 2004), and the use of effective handlers (Van Andel 1989) can all significantly reduce risk on the public transport network. Many aspects of the public transport network have regulated entrance and exit points that connect the internal environment of a transport journey with its external environment. These offer clear prevention opportunities but these must be contextualized within an intelligence-led or problem-solving approach (Clarke and Eck 2005), and as already demonstrated in this entry, these differ significantly for offense type and time of the day.

The design of transport systems and the introduction of new technologies can also influence crime rates. For example, a study on the London Underground in the 1980s by Clarke and colleagues revealed that the introduction of subway slugs (foil wrapped around 10 pence coins to create fake 50 pence coins) became widespread as new automated vending machines were introduced. Remedial measures were highly effective in reducing this, but greater anticipation and investment at the design phase may have prevented this. Interviews by Ekblom with offenders on the London Underground (Clarke and Eck 2005) revealed that offenders would stand by signs warning passengers of "pickpocketing in operation," as after reading these signs passengers would frequently pat their own pockets, which was of considerable assistance to the offender in identifying which pockets a wallet was likely to be in.

Emerging Issues and Future Debate

There are a number of recent developments that should be mentioned, as they offer insights into potential avenues for further research into crime patterns on the public transport network. One of the issues that was raised but perhaps not answered within this entry is whether public transport systems act as a generator or attractor of crime, or perhaps both. A recent paper (Bernasco and Block 2011) investigated the distribution of street robberies in Chicago and how this was influenced by crime generators, crime attractors, and offender anchor points. They suggested that stations provide an increased level of accessibility to areas and found that street blocks with a station had four times as many robberies as similar street blocks without them. However, they suggested that the impact of stations alone was not the significant factor, and the combined impact of a number of additional features near to stations was also contributory. Indeed, research in both Vancouver and New Jersey (Sedelmaier 2003) found no evidence that, on their own, the introduction of a new station to an area increases crime risk.

Block and Block (2000) describe the environs of transport stations and importance of the nearby environment, and Robinson and Giordano (2011) discuss the importance of what they term spatial interplay, the interaction of land uses in relation to crime incidents around transport stations. As described in the Herrmann (2011) study, it is not just the location of subway stations that influenced when and where street robbery occurred but also what was near to the station. A clear avenue for future research is to examine this spatial interplay in more detail, of the interaction between crime risk; and the environs of public transport stops and stations and associated nearby land use types.

Newton et al. (forthcoming) examine this interaction by considering pocket picking and theft on the London underground, and the relationship between what happens above the ground and below the ground, which could be termed the transmissibility of crime on public transport. This research shows that there are distinctive patterns of theft, both outside underground stations and on the network itself. The interesting issue here is that these are policed separately (the underground by British Transport Police and aboveground by London Metropolitan Police), yet offenders do not distinguish between these two settings, in the same way they do not distinguish between the geographical boundaries of different police forces. This demonstrates the need for information sharing between these policing units and also potentially the benefit of joint operations.

It is clear that the environs of public transport and the potential transmission of crime around this system is an area for further research, particularly when examining how offenders may potentially use the public transport as part of their journey to and or from crime. This is a key question in examining the spatial and temporal patterns and concentrations of crime evident on public transport and perhaps, moreover, in determining the extent to which offenders may specialize on public transport networks. There is little information at present as to whether offenders specialize on the public transport network, and if this is different to or perhaps an extension of their usual offending patterns, and to contextualize this within the journey to crime research.

The recent work of Reynald (2011) into the relevance of guardianship raises new issues that are particularly relevant to public transport. One of the findings of her research is that the three key aspects of capable guardianship are the willingness to supervise, the ability to detect potential offenders, and the willingness to intervene. Familiarity of context was found to be a very important component of this, but the public transport network provides a dynamic environment which is constantly changing. This therefore makes transport nodes highly dynamic contexts for which perceptions of familiarity may not be time stable. When on board a moving transport vehicle, new people may enter or leave at each stop, and a vehicle may move through both familiar and perhaps unfamiliar environments. This is certainly an area that warrants future research activity.

More recent technological innovations, including the use of mobile apps such as barcode scanners and the potential for introducing new payment forms (such as scanning mobile phones to pay for entrance/exit to transport systems), may also alert offenders to more attractive handsets and smartphones, akin to the Ekblom interviews with offenders who stood near signs warning of pocket picking. Gentry (2012) examined theft of electronic products on the Massachusetts rapid transit subway and found that theft of electronic devices was concentrated at a small number of stations. As new technologies are introduced to the system, these will offer new opportunities for offending that should be researched.

A final issue that has become increasingly important on the public transport network is line of routes offenses, not en route offenses that occur on a moving vehicle, but those that disrupt lines and delay journeys, for example, damage and obstructions to lines and routes (Smith and Cornish 2006). The increasing value of metals has resulted in a growing increase in the number of metal thefts, and the rail infrastructure in particular has seen a dramatic rise in the number of copper cable thefts over recent years (Sidebottom et al. 2011). This is an example of an emerging and specific type of crime that may be expected to be concentrated spatially, particular to only a few types of environments, the transport system being a key one, and a crime that is hence likely influenced by criminal opportunities associated with the configuration of the transport network.

Related Entries

- ▶ Crime Location Choice
- ▶ Crime Prevention Through Environmental Design
- ▶ Crime Science
- ▶ Criminology of Place
- ▶ Hot Spots and Place-Based Policing
- ▶ Intelligence-Led Policing
- ▶ Routine Activities Approach

Recommended Reading and References

Bernasco W, Block R (2011) Robberies in Chicago: a block-level analysis of the influence of crime generators, crime attractors, and offender anchor points. J Res Crime Del 48(1):33–57

Block R, Block C (2000) The Bronx and Chicago: street robbery in the environs of rapid transit stations. In: Goldsmith V, McGuire P, Mollenkopf J, Ross T (eds) Analyzing crime patterns; frontiers of practice. Sage, Thousand Oaks, pp 137–152

Burrell A (2007) Violence on and around public transport. UCL Jill Dando Institute, University College, London

Ceccato V, Uittenbogaard AC, Bamzar R (2011) Safety in Stockholm's underground stations: the importance of environmental attributes and context. Secur J Advance Access 26 December 2011: 1743–4645 http://www.palgrave-journals.com/sj/journal/v26/n1/abs/sj201132a.html Security Journal (2013) 26, 33–59. doi:10.1057/sj.2011.32; published online 26 December 2011

Clarke R (1996) Preventing mass transit crime. In: Crime prevention studies, vol 6. Willow Tree Press, Monsey

Clarke R, Eck J (2005) Crime analysis for problem solvers in 60 small steps. Office of Community Oriented Policing Services, United States Department of Justice, Washington, DC

Cozens P, Neal R, Hillier D, Whitaker J (2004) Tackling crime and fear of crime while waiting at Britain's railway stations. J Pub Transport 7(3):23–41

Department for Transport (DfT) (2010) Estimated costs to society of crime on public transport in England in 2006/07: final report on findings. DfT, London

Gentry K (2012) Smartphones and safe subways: preventing cell phone theft within the MBTA subway system. Paper presented at the Academy of Criminal Justice Sciences (ACJS) Conference, New York City, 13–17 Mar 2012

Herrmann C (2011) Risky businesses: a micro-level spatiotemporal analysis of crime, place, & business establishment type. Unpublished Ph.D. thesis, The City University of New York, New York

LaVigne N (1997) Visibility and vigilance: metro's situational approach to preventing subway crime. National Institute of Justice, Washington, DC

Loukaitou-Sideris A, Liggett R, Hiseki I (2002) The geography of transit crime: documentation and evaluation of crime incidence on and around the green line stations in Los Angeles. J Plann Educ Res 22(2):135–151

Newton A (2004) Crime and disorder on buses: towards an evidence base for effective crime prevention. Unpublished Ph.D. thesis, University of Liverpool, Liverpool

Newton A (2005) Crime on public transport: static and non-static (moving) crime events. West Criminol Rev 5(3):23–40

Newton A (2008) A study of bus route crime risk in urban areas: the changing environs of a bus journey. Built Environ 34(1):88–103

Newton A, Bowers K (2007) The geography of bus shelter damage: the influence of crime, neighbourhood characteristics and land use. Int J Criminol http://www.internetjournalofcriminology.com/; http://www.internetjournalofcriminology.com/Newton%20and%20Bowers%20-%20The%20Geography%20of%20Bus%20Shelter%20Damage.pdf

Newton AD, Johnson SD, Bowers KJ (2004) Crime on bus routes: an evaluation of a safer-travel initiative. Policing Int J Police Strat Manag 27(3):302–319

Newton A, Partridge H, Gill A (forthcoming) Above and below: examining the transmissibility of theft risk on the London underground

Reynald D (2011) Guarding against crime: measuring guardianship within routine activity theory. Ashgate, Surrey

Robinson J, Goridano L (2011) Spatial interplay: interaction of land uses in relation to crime incidents around transit stations. In: Andresen M, Kinney B (eds) Patterns, prevention, and geometry of crime. Routledge, London, pp 175–199

Sedelmaier C (2003) Railroaded: the effects of a new public transport system upon local crime patterns. Unpublished Ph.D. thesis, Rutgers University, New Jersey

Sidebottom A, Belur J, Bowers K, Tompson L, Johnson SD (2011) Theft in price-volatile markets: on the relationship between copper price and copper theft. J Res Crime Del 48(3):396–418

Smith M (2008) Addressing the security needs of women passengers on public transport. Secur J 21:117–133

Smith M, Clarke R (2000) Crime and public transport. Crime Justice A Rev of the Res (27) pp 169–233.

Smith M, Cornish B (2006) Secure and tranquil travel: preventing crime and disorder on public transport. UCL Jill Dando Institute of Crime Science, London

Tompson L, Partridge H, Shepherd N (2009) Hot routes: developing a new technique for the spatial analysis of crime. Crime Mapp J Res Pract 1(1):77–96

Transit Cooperative Research Program (TRCP) (2009) - Chapter 2: Passenger perception of crime and terrorism. TCRP synthesis 80: transit security update. Transportation Research Board, Washington, DC, pp 10–15

Van Andel H (1989) Crime prevention that works: the care of public transport in the Netherlands. Brit J Criminol 29(1):47–56

Vu Y (2009) Bus stops and crime: do bus stops increase crime opportunities in local neighbourhoods? Unpublished Ph.D. thesis, Rutgers University, New Jersey

Yavuz N, Welch EW (2010) Addressing fear of crime in public space: gender differences in reaction to safety measures in train transit. Urban Stud 47(12):2491–2515

Crime Prevention

▶ Situational Crime Prevention and the Wild West

Crime Prevention Through Environmental Design

Rachel Armitage
Applied Criminology Centre, University of Huddersfield, Huddersfield, UK

Overview

This entry examines an approach to crime reduction which differs from many others in that it focuses not on the offender or their reasoning for committing an offense but upon the environment in which an offense takes place. This approach also differs in its consideration of who should be held responsible for the reduction of crime, with a focus not solely upon the traditional criminal justice system agencies but also upon planners, architects, developers, and managers of public space. The approach is based on the presumption that offenders will maximize crime opportunities, and therefore, those opportunities must be avoided (in the first place) or removed (following the emergence of a crime problem). In the 2001 publication "Cracking Crime Through Design," Pease introduces the concept of design as a means of reducing crime, but more importantly the premise that it is the moral responsibility of many different actors and agencies to improve the lives of those who may fall victim to crime, those who live in fear of crime, and (less obviously) those who will, through the presentation of unproblematic opportunities, be tempted into offending. In the case of crime prevention through environmental design (CPTED), it is the planners, designers, developers, and architects who risk acting (as Pease paraphrases the poet John Donne) as the gateway to another man's sin.

CPTED is an approach to crime reduction (that may be described as a measure, program, or intervention) which aims to reduce crime through the design and manipulation of the built (and sometimes natural) environment. It focuses predominantly upon "designing out" opportunities for crime before they occur – for example, at the preplanning or planning stage – although some interventions take place post-development, in response to a crime problem which has emerged.

A commonly used formal definition is that used by Tim Crowe who defines CPTED as: "The proper design and effective use of the built environment, that can lead to a reduction in the fear of incidence of crime and an improvement in quality of life. . .The goal of CPTED is to reduce opportunities for crime that may be inherent in the design of structures or in the design of neighbourhoods" (Crowe 2000, p. 46). Ekblom (2011) proposes a redefinition and presents the following alternative, which introduces several points not included within Crowe's definition, including the balance between security and contextually appropriate design and the possibility of intervening at different stages between preplanning and post-construction. Ekblom states that CPTED is: "Reducing the possibility, probability and harm from criminal and related events, and enhancing the quality of life through community safety; through the processes of planning and design of the environment; on a range of scales and types of place, from individual buildings and interiors to wider landscapes, neighbourhoods and cities; to produce designs that are 'fit for purpose', contextually appropriate in all other respects and not 'vulnerability led'; whilst achieving a balance between the efficiency of avoiding crime problems *before* construction and the *adaptability* of tackling them through subsequent management and maintenance" (Ekblom 2011, p. 4).

Early proponents of CPTED largely approached the subject from a planning, urban design and architecture perspective (Jacobs 1961; Newman 1973) with Jeffery coining the phrase "crime prevention through environmental design" in his 1971 book of the same title. While this period saw the development of what are still considered to be the key principles of CPTED, the main focus of these studies was upon the influence of the environment on human behavior in general (Coleman 1986) as opposed to the reduction of specific crime types.

The move towards a practical application of the principles, focused largely upon the prevention of acquisitive crimes (such as burglary and car crime) and tailored towards crime reduction practitioners (through training and guidance), emerged in the 1980s with (among others) work conducted by Brantingham and Brantingham (1981), Poyner (1983), and Poyner and Webb (1991).

More recently, research within the field of CPTED has focused upon the *effectiveness* of both the individual and collectively applied principles of CPTED measures in reducing crime and the fear of crime (by authors such as Armitage, Hillier, Pascoe, Cozens, and Kitchen), the *process* of applying CPTED principles within the police and planning environments (by authors such as Kitchen and Monchuk), the development of CPTED-based *risk assessment tools* to predict (and prevent) risk (by authors such as Winchester and Jackson, Van der Voordt and Van Wegen, and Armitage), and a wider approach to the potential benefits of such interventions including the impact upon environmental and social *sustainability* (by authors such as Cozens, Kitchen, Dewberry, Pease, and Armitage and Monchuk).

Given a widening of the focus to include the process of application and consideration of benefits beyond crime reduction, such as social and environmental sustainability, a more appropriate definition of CPTED might be as follows: the design, manipulation, and management of the built environment to reduce crime and the fear of crime and to enhance sustainability through the process and application of measures at the micro (individual building/structure) and macro (neighborhood) level. This definition is represented in Fig. 1 (below).

Principles of CPTED

The principles of CPTED have been presented by several authors, including, but not exclusively, Poyner (1983), Crowe (2000), and Cozens et al. (2005), and adapted across different countries to form the attributes of safe places/environments within planning policy and guidance (e.g., *Safer Places* within England and Wales). Poyner (1983) outlined the principles as surveillance, movement control, activity support, and motivational reinforcement. Cozens et al. (2005) extended this to include the seven principles of defensible space, access control, territoriality, surveillance, target hardening, image, and activity support. The following summary presents a basic introduction to these terms.

Defensible Space

Defensible space is the creation of buildings/enclosures/spaces which allow/facilitate/help the residents of that space to keep potential offenders out. The term was coined by Oscar Newman (1973) who suggested that the physical design of a neighborhood can either increase or inhibit people's sense of control over the spaces in which they reside. Newman categorized space into public (e.g., the road in front of a property), semipublic (e.g., the front garden), semiprivate (e.g., the back garden), and private (inside the property). He argued that if space is defensible, it will be clear to the owner/user of that space, and to non-legitimate users, who should and who should not be in this space. CPTED interventions ensure that space is clearly demarcated, that it is clear who has control/ownership/rights over that space, and that potential offenders have no excuse to be in that space (see discussion of "permissions" in Wortley and Tilley, this volume). CPTED interventions would rarely achieve this through the installation of physical barriers; rather, interventions would include the more subtle measures such as a change in road color and texture or a narrowing of the entrance to the development to mark the area as private.

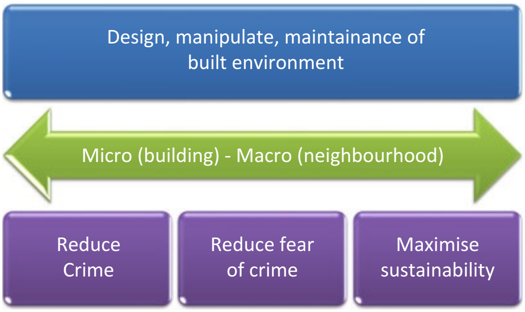
Design, manipulate, maintainance of built environment
Micro (building) - Macro (neighbourhood)
Reduce Crime
Reduce fear of crime
Maximise sustainability

Crime Prevention Through Environmental Design, Fig. 1 Crime prevention through environmental design

Territoriality

Territoriality involves the human emotion/response to the space which they define as their own. Physical responses to territoriality might include a resident marking an area as their own through the installation of a house sign or gate. Emotional responses to territoriality would include a resident's feelings of intrusion or infringement should a person enter what they consider to be their space. Thus, territoriality refers to the human motivation to control the space which they believe is theirs, be that through the legal ownership of that space or through their adoption and management of that space. While Cozens et al. (2005) separate defensible space and territoriality, a more concise summary of CPTED principles might categorize defensible space alongside territoriality, given that the physical creation of defensible space aims to create territorial control over that space.

Access Control

Access control refers to the design of buildings and space to actively keep people out. While this principle has traditionally been referred to as "access" control, perhaps due to its routes in more traditional situational crime prevention measures to restrict entry into buildings and rooms within buildings, within CPTED the aim is much wider. What has been referred to as access control encompasses the following aims: (1) to limit the likelihood that offenders will become aware of that area as a potential target (see discussion of awareness space within Wortley and Tilley's summary of Crime Pattern Theory), (2) to make it more difficult for offenders to navigate into, out of, and within an area should they select it as a target, (3) to increase the *physical* difficulty of entering a building/space should offenders become aware of the area as a target, (4) to increase the difficulty *psychologically* for offenders to enter and move around an area without feeling conspicuous, and (5) to remove any excuse for potential offenders to be within a private or semiprivate space and maximize the legitimate users' confidence in challenging non-legitimate users of space. Given the wider aims of this principle, "access control" would appear too limited a definition. A more appropriate term might be the "limitation of access, egress, and through movement."

Surveillance

Surveillance refers to the way that an area is designed to maximize the ability of formal (security guards, police, employees) or informal (residents, passersby, shoppers) users of the space to observe suspicious behavior. These formal and informal users are referred to in Routine Activity Theory as capable guardians. Within situational crime prevention more generally, surveillance may include the installation of CCTV or the use of formal security guards. Within CPTED, surveillance rarely relates to formal measures but refers more to the informal surveillance created through measures such as ensuring that dwelling entrances face the street, that rooms facing the street are active (such as the kitchen or living room), and that sightlines are not obstructed by shrubbery or high walls. Linked with territoriality, the principle of surveillance requires users of that space to realize that an individual is behaving in a suspicious manner (be that through their behavior or simply their presence within a private/semiprivate area) and to have the confidence to challenge them or intervene. Therefore, the term "surveillance" includes the operational tasks of active (formal) and passive (informal) surveillance, the surveillability (Ekblom 2010) of that space, and the creation of the perception among offenders that they are being observed.

Target Hardening

Target hardening is often referred to as physical security and includes the initial design, or retrofit upgrade, of doors, windows, fences, and other physical structures to increase the difficulty for offenders in entering a building or space.

Image

Cozens et al. (2005) use the term "image," while others have used "management and maintenance" to cover the principle of creating buildings/spaces which are physically free from litter, graffiti, vandalism, and damage but are also areas without stigma or a poor social reputation. It is difficult to allocate a specific label to these concepts as image refers to a state and management and maintenance to the activities which create that state.

Activity Support

Activity support relates to the creation of an environment which increases the likelihood that legitimate users will make use of space and subsequently act as additional surveillance. Although activity support is included by many as a distinct principle of CPTED, the ultimate aim is to enhance surveillance and so the two principles can be combined.

Given this discussion of definitional issues and a desire to condense the list presented by Cozens et al. (2005), the principles of CPTED might be summarized as:

- Physical security – securing buildings and spaces to a level which is appropriate to risk and where possible products which are tested to the relevant security standards should be utilized
- Surveillance – designing building and space to allow both formal and informal surveillance from users of that space and to create a feeling of unease among non-legitimate users of the space
- Movement control – limiting access, egress, and through movement

- Management and maintenance – ensuring that buildings and the surrounding spaces are designed to create a positive image and to ease future maintenance of the space and ensuring that programmed management and maintenance systems are in place
- Defensible space – ensuring that spaces have a clearly defined ownership, purpose, and role to enhance feelings of territoriality among residents and legitimate users

Empirical Support for CPTED

Ekblom (2009) presents an excellent discussion of where CPTED needs to be improved both conceptually and practically, and these are addressed in the conclusion to this entry. One weakness identified by Ekblom is the need to improve the evidence upon which CPTED is based. While it is acknowledged that there are still areas which require clarification, the evidence base has been strengthened over the last decade, with several rigorous examinations of the impact of individual design features on crime levels. Using the five principles defined above, the following section outlines the evidence base to support the efficacy of CPTED in reducing crime.

Physical Security

Research on security measures as a means of preventing residential burglary is mixed, with several studies suggesting that the actual home break-in requires little in the way of technical sophistication and that physical security is a low priority for burglars when searching for targets (e.g., Reppetto 1974). However, several studies suggest that with all other factors being equal, burglars would prefer to offend against properties with lower levels of physical security (Cromwell et al. 1991). Budd's (2001) analysis of the British Crime Survey found that security devices are very effective in reducing the risk of burglary victimization. Budd's study found that, in England and Wales in 1997, 15 % of households without security measures were burgled, compared to 4 % of households with basic measures in place and 3 % with higher levels of security.

Surveillance

Research suggests that surveillance and visibility play a major part in offenders' decision-making processes when selecting properties to offend against. Reppetto (1974) interviewed 97 convicted burglars and found that the most common reason for avoiding a target was that there were too many people around. Offenders stated that the possibility of neighbors watching them deterred them from selecting a property and that they would select targets where they felt less conspicuous and where there was less visual access from neighboring properties. When assessing the design characteristics of victimized properties, several studies have identified a lack of surveillance or poor levels of visibility as key features of crime-prone homes (e.g., Armitage 2006; Brown and Altman 1983; Winchester and Jackson 1982).

Movement Control

The efficacy of the principle of limiting movement is less clear-cut. Although the majority of research suggests that limiting through movement will reduce crime, some studies suggest that encouraging pedestrian and vehicular movement will provide informal surveillance of the area – something that Jacobs (1961) refers to as "eyes on the street." The mechanisms through which limiting movement might reduce crime are as follows: firstly, that an area with high levels of through movement provides ease of entry/escape for offenders, and therefore, reducing through movement would decrease the risk of offending. This supposition is supported by interviews with offenders and analysis of victimization levels within residential areas (e.g., Poyner and Webb 1991; Taylor and Gottredson 1987). The second mechanism suggests that an area with high levels of through movement is more likely to be within an offender's awareness space. Offenders are likely to have passed through the area while conducting their day-to-day activities and to have become familiar with the suitability of targets. Limiting through movement would thus decrease the likelihood of offenders being aware of suitable targets. This supposition is supported by interviews with offenders and analysis of patterns of crime within residential areas (e.g., Poyner and Webb 1991; Wiles and Costello 2000). The third rationale for limiting movement is based upon the premise that an area with high levels of through movement will allow offenders to feel less conspicuous and to blend in with the activities of legitimate users of the space. This is supported by several research studies (e.g., Poyner and Webb 1991; Taylor and Gottredson 1987).

Although there have been many studies into the impact of through movement on crime levels within residential areas, two particularly rigorous studies, conducted in England, have found that burglary is higher where properties are located within an area with a higher number of connections to other areas and lower on what are referred to as "true" culs-de-sac – those with no connecting roads of pathways (Armitage et al. 2010; Johnson and Bowers 2010).

In a review of the evidence relating to the impact of through movement on crime, Taylor (2002) concludes that "Neighbourhood permeability is . . . one of the community level design features most reliably linked to crime rates, and the connections operate consistently in the same direction across studies: more permeability, more crime" (Taylor 2002, p. 419). However, this assertion is not universally supported as there is some research – particularly research conducted in the last decade and using space syntax techniques – that has concluded that increased levels of through movement have a beneficial impact upon crime. Several studies have concluded that crime is concentrated in more isolated and less accessible streets (e.g., Hillier and Sahbaz 2009). One explanation for the disparity in research findings is that, although space syntax allows a greater number of properties to be analyzed (Hillier and Sahbaz looked at 101,849 properties), it also means that, for some aspects of layout (including particularly footpaths and through movement), presumptions are made about movement and patterns. For example, the remote assessment of a neighborhood may show a cul-de-sac development with no connecting footpaths, and that would be labeled a true cul-de-sac. However, on physically assessing a development (i.e., completing an assessment on-site – as was the case with research completed by Armitage et al. 2010), it may become apparent that, although there are no official footpaths, residents and users of the space have created informal footpaths because they connect the development to nearby shops.

Management and Maintenance

Several studies have suggested that if low-level disorder such as vandalism and litter are not addressed, they can act as a catalyst for more serious crimes. Skogan (1990) refers to this as the contagion theory, suggesting that the "presence of vandalism stimulates more vandalism" (p. 39). Wilson and Kelling (1982) refer to this contagious effect as the "broken windows theory" (p. 16). Broken windows suggests that an area with existing deterioration such as graffiti and vandalism conveys the impression that (a) nobody cares so apprehension is less likely and (b) the area is already untidy so one more act will go unnoticed. This conclusion is supported by Taylor and Gottredson (1987) who found that physical incivilities indirectly influence offenders' perception of risk in that they portray residents' level of care or concern for the area in which they live and thus act as an indicator for the likelihood that the residents will intervene if they detect an offense taking place. Armitage (2006) assessed 1058 residential properties in England and found that those which showed signs of poor management and maintenance had experienced higher levels of prior burglary.

Defensible Space

As was discussed above, defensible space is a term used to describe the design features of an area that increase territorial behavior among residents and users of that space. Specific defensible space measures include maximizing the perception that a space is private or semiprivate through subtle design features such as a narrowing of the road entrance or a change in road color or texture. These environmental features are sometimes referred to as symbolic barriers as they do not physically keep people out. Brown and Altman (1983) and Armitage (2006) found that, compared with non-burgled houses, properties which had been burgled had fewer symbolic barriers, as well as actual barriers such as fences and locked gates, protecting private territory from public access. Brown and Bentley interviewed offenders, asking them to judge (from pictures) which properties would be more vulnerable to burglary. The results revealed that properties showing signs of territorial behavior (such as the installation by of a gateway at the front of the property or a sign on the gate/door marking the area as private) were perceived by offenders to be less vulnerable to burglary.

Practical Application

While CPTED is founded on an agreed set of theories and assumptions, the way that CPTED is applied varies across and even within different countries. It is beyond the scope of this entry to cover the different international approaches to implementing CPTED. Therefore, the focus in this section is upon the three countries England/Wales, Australia, and the Netherlands. England/Wales has been chosen because it has developed a holistic approach to incorporating CPTED into the planning system – through the publication of specific planning policy and guidance, the promotion of an award scheme, and the provision of specific police resources to ensure that CPTED is considered within local development. Australia, and specifically the state of New South Wales (NSW), is considered for two reasons. The first is that in NSW there is a legislative requirement for new developments to be assessed for crime risk; the second is that NSW has adopted a very different approach to implementing this requirement, one which involves very little police resources or even involvement, with crime risk assessments largely conducted by private crime reduction consultants. The Netherlands is considered for its comprehensive model of delivering CPTED within the planning system. The model includes legislative requirements, incentives, and process of delivery.

Across England and Wales there are 43 police forces, and within each of these there is at least one individual whose role involves reviewing the planning applications which are submitted to the local planning authority (within the local council) and offering CPTED advice to mitigate any potential crime risks associated with the proposed development. This role is referred to as Architectural Liaison Officer (ALO) or Crime Prevention Design Advisor (CPDA). The distinction is generally geographical, with northern police forces using the term “ALO” and southern forces using the term “CPDA.” Even within England and Wales (which share a government and associated laws and policies), the role of ALO/CPDA varies between police forces, with some ALO/CPDAs dedicated entirely to this role, while others have numerous additional roles. The role can also vary in terms of process, with some local planning authorities *requiring* preplanning consultation (e.g., the local authorities within Greater Manchester), while other forces have a more reactive response, with the consideration for crime prevention being entirely dependent upon the ALO/CPDA seeking out current planning applications and contacting the planning office to offer CPTED advice. Nationally, within England and Wales, the consideration for crime prevention within planning decisions is not a requirement. The planning system in England and Wales is guided by national policy – at the time of writing the National Planning Policy Framework. This policy states that local planning policies and decisions should aim to create developments which are (among other considerations) safe and where crime, disorder, and the fear of crime do not undermine quality of life. This should be considered in local and neighborhood planning decisions but is not a legislative requirement. England and Wales also has planning guidance which directs local planning authorities, and those working within the built environment profession, as to *how* to develop safe neighborhoods. This guidance is entitled *Safer Places – The Planning System and Crime Prevention.* While this is a typical model of delivery of CPTED within England and Wales, there is one police force which has taken a very different approach – this is Greater Manchester Police (GMP) in North-West England. The GMP approach to delivering the ALO/CPDA role involves a team of consultants who, while being civilian, are based within GMP Police Headquarters and have access to police recorded crime data, Neighbourhood Policing Teams (NPTs), and other relevant police intelligence. As was suggested above, the GMP approach also differs from other police forces in England and Wales in the emphasis placed upon preplanning consultation between the planning authority, developers, and the GMP team. Within the ten Greater Manchester local authorities, where a client wishes to apply for planning permission, the application which is made to the local planning authority must adhere to national planning policy but also to the requirements made by each local authority (through what is referred to as a validation checklist). Crucially, for Greater Manchester Police, each local authority *requires* that the submission of a major planning application be accompanied by a Crime Impact Statement (CIS) – a document which includes an analysis of crime statistics, reports from site visits, and local policing knowledge of the area to highlight potential crime risk and to make CPTED recommendations to mitigate those risks. The final unique element is that the GMP team is able to charge for their service. This has helped to protect the provision of this service within the current economic climate, and while other ALO/CPDA services are seeing cuts, this team has managed to sustain its services and even expand.

In addition to policy, guidance, and allocation of police resources, England and Wales also implement an award scheme to encourage developers to design out crime at the planning or preplanning stage. The Secured by Design (SBD) scheme is managed by the Association of Chief Police Officers Crime Prevention Initiatives (ACPO CPI) and run on a day-to-day basis by local police ALOs or CPDAs whose role is to ensure that developments are designed and built to certain specifications. SBD is based upon the key principles of CPTED, and the standards and guidance follow those principles of physical security, surveillance, access/egress, territoriality, and management and maintenance. There have been five published evaluations of the effectiveness of the SBD scheme (see Armitage and Monchuk 2011 for a review), each concluding that SBD confers a crime reduction advantage.

In countries such as Australia, delivery varies dramatically from state to state. The state of New South Wales is selected as an example for this entry because of the model of delivery which includes a legislative requirement for a crime risk assessment to be conducted for developments considered by the local council to pose a crime risk. While this legislation shows a clear commitment to the importance of CPTED, the process of embedding this within the planning and policing system differs greatly to England and Wales. In New South Wales there is no equivalent of the ALO/CPDA role, and the closest position to this is the Crime Prevention Officer. In a similar vein to the Crime Reduction/Prevention Officer role in England and Wales, the post includes a variety of roles and responsibilities. Within New South Wales, this post also has the additional burden of covering a large geographical area. This means that in practice, the Crime Prevention Officer cannot systematically assess all planning applications from a crime prevention perspective. Therefore, the role of conducting the required crime risk assessment and recommending alterations based upon crime risk is conducted either by private crime prevention consultants, planning companies, or the developers themselves. Clancey et al. (2011) conducted a review of 33 crime risk assessments submitted between January 2007 and October 2010 and found that these were conducted by 24 companies – 11 of which were planning firms, eight were social planning firms, seven were development companies, five were private crime prevention consultants, and two were engineering firms. The review also identified that while the guidance specifies that crime risk assessments should measure risk using police recorded crime data, hot-spot analyses, and socioeconomic data, only 16 of the 33 contained any reference to crime data, and for those which did, the analysis was broad and shallow with little indication of specific crime risk in terms of crime type, location, or modus operandi. The New South Wales model of delivery has greater legislative strength than that of England and Wales, with a requirement in legislation for crime risk assesssments to be conducted where a local council considers there to be a crime threat from a potential development. The process of delivering this also differs to the predominantly police-based model of England and Wales, with a greater responsibility placed upon external private consultants. Strengths of this method include legislative power to require consideration of crime as well and less pressure on police resources. Potential weaknesses relate to the threat to independence where a Crime Risk Assessment is conducted by a company who may have a vested interest in seeing the development go ahead with little or no alterations to the planning application.

The Netherlands has one of the most comprehensive approaches to embedding CPTED within the planning process, and this applies to regulation, award schemes, and the process of delivery. In terms of regulation all new-built homes in the Netherlands have to comply with specific security regulations for windows and doors, and from the 1st of January 1999, planning permission could only be obtained if the application met the legal requirements for built-in security. The Netherlands also has an award scheme (similar to the UK's SBD scheme) entitled Police Label Secure Housing. Unlike the SBD scheme, this award (which was originally owned and managed by the police) is managed by the Dutch government who adopted the police label into their planning policy guidelines, and (since 2004) every new estate or dwelling must be built in accordance with the police label or an equivalent label. Although the award was modeled on SBD, there are several distinctions which mark the two schemes apart. The first is that the label is split into three different certificates – Secured Dwelling, Secured Building, and Secured Neighborhood. These can be issued separately, but together they form the Police Label Secure Housing award. The label is also less prescriptive than SBD with more flexibility for developers aiming to achieve a secure development. The list of requirements is set out under five categories (urban planning and design, public areas, layout, building, dwelling), and these include performance requirements (what) and specifications which indicate the way in which those requirements will be met (how). As a means of encouraging creativity and avoiding the risk of developers "designing down" to specific requirements, where a developer offers a solution which differs from that set out in the "how," but can still demonstrate the same preventative effect, then this will be considered. The scheme also differs in that it is valid for ten years only, and after this period, a reassessment is required. In terms of the delivery of the scheme, the system is very similar to that within England and Wales. Until 2009, each police region had a number of Building Plan Advisors (Bouwplanadviseur) whose role was very similar to the ALO/CPDA role. As a response to budget cuts, the role has been civilianized and is run by the municipalities either through the employment of external consultants or civilian Building Plan Advisors located in-house.

A Critique of CPTED

While many of the wider criticisms of situational crime prevention in general apply to CPTED, there remain some fundamental, unresolved weaknesses in the theory and application of this method of crime reduction. This entry concludes with a brief presentation of the main criticisms of CPTED and a discussion of how these might be addressed.

Lack of flexibility: One area where CPTED has remained open to criticism is in the lack of flexibility in its principles and guidance as well as the application on the ground. This lack of flexibility may, in part, relate to the agencies traditionally involved in delivering CPTED (police, security consultants, ex-police) whose professional background and training focuses upon the requirement to follow and not challenge instructions. It may also relate to a lack of confidence, training, or experience among those carrying out the role of delivering CPTED that, in practice, leads to the rigid application of standards as opposed to adapting the design to fit a particular context.

Failure to clarify confusion on specific CPTED principles: This criticism relates to the confusion which has surrounded one particular CPTED principle – that being the impact of through movement on levels of crime. The debate centers upon the benefits of facilitating movement within an area weighed against the risks of potentially criminogenic design. For those who advocate increased connectivity, the rationale does not necessarily relate to crime reduction. The primary purpose of designing connected developments is to ensure that people can get from A to B without the need of a vehicle, thus reducing carbon emissions and the visibility of the car, and to avoid the need for residents to take unnecessarily lengthy routes. While the *cul-de-sac* layout is favored in the majority of the criminological literature on the subject, urban designers would argue that there are many negative features of this layout. It increases travel distance and therefore reliance upon the motor vehicle, it is an inefficient use of land, and it increases the difficulty of ensuring that public transport can travel close to these residential properties. While research will always present differing findings and subsequently viewpoints, this debate has led to polarized and often overstated statements regarding the crime risk of culs-de-sac versus through roads. Not only has such simplification proved unhelpful for those tasked with reducing crime through the design and manipulation of the environment, it has also led to unnecessary confusion regarding a subject for which the academic evidence appears to be relatively unambiguous.

Non-standardized delivery: Perhaps linked to the lack of professionalization of the role of implementing CPTED at ground level, there is a concern regarding the considerable variation in the process of delivery, both nationally and internationally. There is not only a difference in *who* is in the CPTED delivery role (England and Wales being predominantly warranted police, the Netherlands civilians based within municipalities, and New South Wales security consultants) but also a difference in *how* CPTED is being applied. Some countries have taken the approach of introducing legislation or building regulations to require the specific security standards within residential dwellings (the Netherlands and Scotland); others have introduced legislation to require the consideration of crime risk (New South Wales) or incentivized the consideration for crime prevention with awards such as SBD (UK) and Police Label Secured Housing (the Netherlands). While it is understandable that the delivery will differ between countries and even between states, there still remains a lack of consistency within countries such as England, which share the same legislation and regulations.

Lack of clarity in scope: Perhaps less of a criticism than a note of caution that the boundaries and scope of CPTED must be made clear. While the principles cover design, building, and future management and maintenance of an area, the extent to which CPTED interventions can realistically influence so many factors is debatable. As Ekblom (2009) clearly highlights, “There is a tendency to use the label CPTED indiscriminately to cover everything that aims to prevent crime in the built environment…this is not conducive to focused thinking” (Ekblom 2009, p. 9). An example which highlights this is the principle of management and maintenance. The SBD scheme has historically stated that developments must have a programmed system in place to manage and maintain the area. Yet, unless this is social housing, how is this program established and, more importantly, maintained? It appears to be a principle which fits well in theory, but cannot be consistently applied in practice. While warning of the risks of extending its net too wide, there would be some merit in CPTED integrating the wider social approach addressed by what has been labeled as second-generation CPTED (see Saville and Cleveland 2003).

Failure to align with other agendas: Although this criticism has begun to be addressed through work conducted by (among others) Cozens, Pease, Armitage, and Monchuk, the CPTED community has been very slow to adapt their focus to fit with contemporary issues such as social, economic, and environmental sustainability. While the two agendas of minimizing crime and maximizing sustainability may appear distinct, achievement of one (reducing crime) ultimately contributes to the other (maximizing sustainability). Crime is carbon costly, including the carbon costs of police mileage in response to a crime, the replacement of stolen and damaged property, the health and other costs to victims in high crime, the costs of moving home in response to crime or fear of crime, and the maintenance and refurbishment of void properties. Crime has a huge carbon footprint, and given the prominence which the carbon reduction agenda has been given in government policy, media attention, funding for research, and the priorities of the general public, it would appear that there has been a missed opportunity to enhance the priority afforded to the importance of the consideration for crime reduction within the built environment.

Failure to innovate and adapt to change: The final criticism relates to the failure of CPTED to respond to the social and economic environment in terms of its model of delivery, the principles upon which it is based, and its focus. The first example of the failure to innovate relates to the current economic crisis and how this has impacted upon funding for public services such as the police. Of the 43 police forces within England and Wales, only one has adapted its model of delivery not only to survive these cutbacks but to thrive and grow through innovation. There has also been slow progress among CPTED research, policy, and practice to adapt to the changing nature of crime, with the focus remaining on acquisitive crimes, while more common disorder issues have been sidelined.

Conclusions and Future Research

While CPTED may lack some consistency in process and application, research suggests that the principles upon which it is based can work, both alone and combined with other interventions, to reduce crime and the fear of crime and to maximize social, environmental, and economic sustainability. In moving forward, CPTED must evolve, but in the words of Ekblom (2009), it must lose its historical baggage first. While there is always room for further research, the CPTED community can begin to confidently challenge some of the debates which have dominated this field and which extensive, independent, and methodologically rigorous research has clarified. Attention should now be focused upon building upon examples of good practice both in terms of by *whom* and *how* CPTED is delivered on the ground. Implementation should be adapted to context and designed to suit the social and economic challenges of different communities. Future thinking should focus upon new models of delivery which can be implemented with limited public funding and within political environments which favor restricted legislation, regulation, and governmental interference. CPTED must also adapt to changing concerns regarding crime. The traditional focus upon acquisitive crime must widen to address public concern regarding low-level crimes and antisocial behavior and also governmental priorities such as terrorism and violent extremism. But these challenges can be seen as opportunities. Where there are problems, there is scope to develop solutions, and CPTED is a practical, cost-effective crime reduction measure which, research has shown, can adapt to many different problems and contexts.

Related Entries

- ▶ Agent-Based Models to Predict Crime at Places
- ▶ Applied Geographical Profiling
- ▶ British Police
- ▶ Broken Windows Thesis
- ▶ Co-offending and Offender Decision-Making
- ▶ Crime Mapping
- ▶ Crime Science
- ▶ Designing Products Against Crime
- ▶ Effectiveness of Situational Crime Prevention
- ▶ History of Geographic Criminology Part I: Nineteenth Century
- ▶ Innovation and Crime Prevention
- ▶ Public Housing and Crime Patterns
- ▶ Rational Choice Theory
- ▶ Residential Burglary
- ▶ Routine Activities Approach
- ▶ Situational Crime Prevention
- ▶ Theories for Situational and Environmental Crime Prevention

Recommended Reading and References

Armitage R (2006) Predicting and preventing: developing a risk assessment mechanism for residential housing. Crim Prev Community Saf Int J 8(3):137–149

Armitage R, Monchuk L (2011) Sustaining the crime reduction impact of secured by design: 1999 to 2009. Secur J 24(4):320–343

Armitage R, Monchuk L, Rogerson M (2010) It looks good, but what is it like to live there? Assessing the impact of award winning design on crime. Spec Vol Eur J Crim Policy Res 17(1):29–54

Brantingham PL, Brantingham PJ (1981) Environmental criminology. Sage, Beverley Hills

Brown BB, Altman I (1983) Territoriality, defensible space and residential burglary: an environmental analysis. J Environ Psychol 3:203–220

Brown B, Bentley D (1993) Residential burglars judge risk: the role of territoriality. J Environ Psychol 13:51–61

Budd T (2001) Burglary: practice messages from the British crime survey. Home Office, London

Clancey G, Lee M, Fisher D (2011) Do crime risk assessment reports measure crime risks? Curr Issues Crim Justice 23(2):235–254

Coleman A (1986) Utopia on trail: vision and reality in planned housing. Hilary Shipman, London

Cozens P, Saville G, Hillier D (2005) Crime prevention through environmental design (CPTED): a review and modern bibliography. Prop Manag 23:328–356

Cromwell PF, Olson JN, Avary DW (1991) Breaking and entering: an ethnographic analysis of burglary. Sage, Newbury Park

Crowe T (2000) Crime prevention through environmental design: applications of architectural design and space management concepts, 2nd edn. Butterworth-Heinemann, Oxford

Ekblom P (2009) Redesigning the language and concepts of crime prevention through environmental design. http://reconstructcpted.wordpress.com/publications-and-other-papers/. Accessed 21 Aug 2012

Ekblom P (2011) Deconstructing CPTED. . .and Reconstructing it for practice, knowledge, management and research. Eur J Crim Policy Res 17:7–28

Hillier B, Sahbaz O (2009) Crime and urban design: an evidence-based approach. In: Cooper R, Evans G, Boyko C (eds) Designing sustainable cities. Wiley-Blackwell, Chichester

Jacobs J (1961) The death and life of great American cities. Random House, New York

Johnson S, Bowers KJ (2010) Permeability and burglary risk: are Cul-de-Sacs safer? Quant J Criminol 26(1):89–111

Newman O (1973) Defensible space: people and design in the violent city. Architectural Press, London

Pease K (2001) Cracking crime through design. Design Council, London

Poyner B (1983) Design against crime: beyond defensible space. Butterworth, London

Poyner B, Webb B (1991) Crime free housing. Butterworth, Oxford

Reppetto TA (1974) Residential crime. Ballinger, Cambridge, MA

Saville G, Cleveland G (2003) An Introduction to 2nd generation CPTED: Part one. CPTED Perspect 6(1):7–9

Skogan WG (1990) Disorder and decline: crime and the spiral of decay in American neighbourhoods. University of California Press, California

Taylor R (2002) Crime prevention through environmental design (CPTED): yes, no, maybe, unknowable and all of the above. In: Bechtel RB, Churchman A (eds) Handbook of environmental psychology. Wiley, New York

Taylor R, Gottredson SD (1987) Environmental design, crime and prevention: an examination of community dynamics. Crime Justice Annu Rev Res 8:387–416

Wiles P, Costello A (2000) The ‘road to nowhere’: the evidence for travelling criminals, home office research study 207. Home Office, London

Wilson JQ, Kelling GL (1982) Broken windows: the police and neighborhood safety. Atlantic Mon 127:29–38

Crime Preventive Effects of Incapacitation

Hilde Wermink[1], Robert Apel[2],
Arjan A. J. Blokland[1,3] and Paul Nieuwbeerta[1,4]
[1]Department of Criminology and Criminal Law, Leiden University, Leiden, The Netherlands
[2]School of Criminal Justice, Rutgers University, Newark, NJ, USA
[3]The Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, The Netherlands
[4]Utrecht University, Utrecht, The Netherlands

Overview

Among the manifold goals of penal confinement, incapacitation aims to impose a period of “time out” from the criminal career, by removing the opportunity for an individual to commit crime in

This chapter is largely based on two articles that are published elsewhere (Blokland and Nieuwbeerta 2007; Wermink et al. 2012). Some parts are incorporated integral.

the community for the duration of his or her sentence. After a period of empirical stagnation, questions concerning the incapacitation effect of imprisonment have recently become salient. This recent empirical attention has been spurred, in part, by the unparalleled growth in the use of incarceration in Western society over the last three decades. This growth is most stark in the United States. Yet in Europe as well, two-thirds of the 35 countries surveyed experienced growth in their incarceration rate during the first half of the 2000s (Aebi et al. 2006).

A growing literature has become attentive to the effects of incapacitation. This entry gives an overview of the current literature and pays attention to the (a) effects of general incapacitation and (b) effects of specific incapacitation of frequent offenders.

Effects of General Incapacitation Effects of Prison

Approaches to Estimation of General Incapacitation Effects of Prison

A key goal for a study of incapacitation is to estimate the number of crimes that an incarcerated offender *would have committed*, where he/she is free in the community rather than confined in prison. This is taken as an estimate of the incapacitation effect, defined as the number of crimes averted by physically isolating an offender from society at large (Blumstein et al. 1978). This task would seem straightforward at first glance, but the problem on closer inspection is that this quantity is a counterfactual which can never, even in principle, be directly observed.

Estimation of the incapacitation effect of imprisonment is fundamentally an actuarial exercise. The empirical challenge is one of overcoming the *counterfactual problem*, that is, providing a credible estimate of behavior that is not observed because an individual is locked up in jail or prison rather than being free in the community. The key quantity for incapacitation is known as lambda, λ, representing the annual offending frequency conditional on active offending, which can be taken as an estimate of the number of crimes avoided through incarceration. Zimring and Hawkins (1995, p. 81) observe that there are two approaches to estimating lambda: *[T]o determine the level of crime that would have occurred if a particular group had not been confined, one must either study the criminal activity of the same group at a different time in their lives to estimate what that group would have done if not confined, or one must study the behavior of persons other than those confined to approximate the crimes avoided by imprisonment in the past.*

It can be referred to the first as a *within-person counterfactual approach*, and to the second as a *between-person counterfactual approach*. Point of departure for both approaches mentioned is that the crimes the offender would have committed while free are actually prevented by incarcerating the offender. As such, these methods do not account for the mitigating effects of replacement – other offenders filling in the position left vacant by incarcerating the original offender – or the fact that co-offenders might have committed the particular crime regardless of the absence of the incarcerated offender. Replacement however seems especially relevant for drug offenses and is deemed largely irrelevant for the large majority of crimes. Group offending in turn might be more relevant for juveniles than for adults since juvenile offenders more often co-offend.

Within-Person Approach

The most influential incapacitation studies are universally of the first kind in Zimring and Hawkins' (1995) typology. These studies provide within-person counterfactual offending rates from self-report surveys of arrestees or prison inmates. With this approach, targeted individuals are questioned about their criminal activity during the months leading up to their arrest or confinement. These estimates are taken as the number of crimes they committed on an annual basis when they were free (prior to incarceration) and by implication, the number of crimes they would have committed per year during the time that they were incarcerated. Therefore, the counterfactual offending rate for incarcerated individuals is their own offending rate in the months preceding their confinement.

The best-known source of a within-person counterfactual is the Rand Corporation's second inmate survey, a study of over 2,000 male inmates in California, Michigan, and Texas (Chaiken and Chaiken 1982). This study, for example, revealed that annual offense frequencies for ten different crimes were highly skewed, with active offenders at the median committing 15 offenses and offenders at the 90th percentile committing 605 offenses (these figures exclude drug offending). They estimated the mean offending rate to be between 187 and 278 crimes per year, depending on how ambiguous survey responses were treated. Based on these figures, incarceration was shown to have the capacity to substantially incapacitate criminal behavior. In fact, on the basis of this research, policy-oriented criminologists began to advocate selective incapacitation of high-rate criminal offenders as an explicit penal policy.

These incapacitation studies had limitations, however. First and foremost is the question of the reliability of offender self-reports of their criminal behavior (Spelman 1994; Zimring and Hawkins 1995). Specifically, it is unclear whether incarcerated offenders, particularly high-rate offenders, can accurately recall their prior criminal activity. Even if they can, they may not be motivated to report it honestly to interviewers. Spelman (1994), in fact, found evidence of both overreporting and underreporting in the second Rand inmate survey, but by individuals at different locations in the distribution of offending rates. Second, the presence and nature of crime spurts can introduce serious distortions in estimates of lambda (Blumstein et al. 1986) and therefore in estimates of incapacitation effects. Research has shown that offenders experience relatively short periods of high-rate offending immediately prior to incarceration, meaning that incapacitation effects will be severely overstated, especially if reporting windows are comparatively narrow. Moreover, a portion of the pre-incarceration crime spurt appears to be artifactual rather than behavioral, that is, a function of the way that individuals are filtered through the criminal justice system and selected for custodial sentences rather than reflective of genuine growth in offending frequency.

Finally, these early incapacitation studies were guilty of overly optimistic assessments of the incapacitative benefits of incarceration. The studies showed self-report offending rates to vary widely from one jurisdiction to another. For example, Chaiken and Chaiken (1982) reported the median total nondrug crime frequency among prisoners (not including jail inmates) in California, Michigan, and Texas to be 42, 17, and 9, respectively. The median offender reported committing 4.2 nondrug offenses per year in Nebraska (Horney and Marshall 1991), 4.4 nondrug offenses in New Orleans (Miranne and Geerken 1991), and 12 nondrug offenses in Wisconsin (DiIulio 1990). Relatedly, the incapacitation estimates from this research tend to assume constant offending during the reference period, whereas Horney and Marshall (1991) demonstrated that most offenders were actively involved in crime only intermittently and that failure to account for this intermittency would inflate estimates of the incapacitative benefits of imprisonment when offense rates are annualized.

Between-Person Approach

Until recently, within-person counterfactual studies exhausted all empirical research in the incapacitation tradition. Three studies in the recent years have adopted the second approach in Zimring and Hawkins' (1995, p. 81) typology, that of estimating the offending rate of persons other than those confined to approximate the crimes avoided by imprisonment.

Sweeten and Apel (2007) relied on propensity score matching to select incarcerated and non-incarcerated individuals who closely resembled each other on a wide variety of background variables, including criminal history. Using data from a contemporary, nationally representative sample of American youth, they limited their attention only to individuals who were not incarcerated prior to the reference conviction. They estimated lambda among the non-incarcerated sample to be about nine offenses per year among 16–17 year olds and about six offenses per year among 18–19 year olds.

Owens (2009) took advantage of a change in the sentencing guidelines in Maryland that lowered the age at which an offender's juvenile record could be used to add criminal history points for the purpose of criminal sentencing. This policy change effectively reduced the recommended sentence faced by certain groups of young adults with a juvenile criminal history. In her study, individuals who were sentenced after the policy change accumulated 2.8 arrests per year (with an implied Index offending rate of 1.5 offenses per year) during the time that they would still have been confined under the old sentencing regime.

Wermink et al. (2012) relied also on the between-person approach, but combined it with the within-person approach using a sample of all offenders convicted of a crime in the Netherlands in 1997. Using the between-person counterfactual approach, offenders who were convicted but not imprisoned in 1997 were matched to those individuals convicted to imprisonment in 1997. To estimate the incapacitation effect, they investigate the criminal behavior of the matched controls after their matched counterpart was sent to prison. Using the within- and between-person counterfactual approach, their best estimate is that between 0.17 and 0.21 convictions are prevented per year of first-time incarceration. Distinguishing for different social groups, their results show larger first-time incapacitation effects for male offenders compared to female offenders and for juveniles compared to adults. Their incapacitation estimates are considerably smaller than those of earlier studies.

Remarks on Effects of General Incapacitation

In conjunction with these prior findings, the current results inform both penal policy and the public discourse surrounding the use of imprisonment as the societal reaction to crime. The current results suggest that in as far as imprisonment is used with the primary goal of reducing crime, a general increase of imprisonment as the sanction of choice is not likely to yield high crime control benefits. While the incapacitative effects of imprisonment are likely to be larger when specific offender groups are selectively targeted, studies examining the effects of such selective policies mostly find rather disappointing effects. In addition, prior studies indicate that even a relatively short prison sentence may have long-lasting detrimental effects in terms of increased offending. Given the incapacitation effect of first imprisonment is not that large to begin with, it is therefore likely to be offset – if not overruled – by the long-term collateral effects of incarceration. This is not to say that imprisonment does not have any function in the criminal justice system. It does however clearly point to its limitations and downsides, underscoring that imprisonment is to be regarded not as a panacea to the crime problem, but merely as an ultimum remedium.

Selective Incapacitation of Frequent Offenders

Criminal justice policies also aim for effects of selective incapacitation, that is, imprisonment policies specifically targeting some predefined group of offenders. Selective incapacitation policies are increasingly being implemented, both in the USA and in Europe. The rationale behind these policies is straightforward: reducing crime at lower cost. The more skewed the distribution of offense frequency in the offender population, the more offenses can be prevented by selectively incapacitating those offenders that are at the high end of this distribution. If frequent offenders are also more serious and violent offenders, as both theory and empirical research suggests, the benefits of removing these offenders from society may be even greater (Piper 1985). Selective incapacitation however is not uncontroversial. Ethical objections have been raised to the sanction disparity inherent to these selective policies (e.g., Moore 1986). Others have stressed the "stochastic" selectiveness that is already built-in to the judicial system, predicting low yields of an additional focus on frequent offenders (Blumstein et al. 1993). Finally, researchers have pointed to the rising prison costs, wastage of prison capacity, the costly care of the aging prison population, possible problems in monitoring growing groups of inmates serving long-term sentences, and the growing number of appeals lodged by those offenders meeting the selective criteria that may result from selective policies.

Research into the effects of incapacitation has either been at a macro-level, combining aggregate data on prison populations and crime rates, or at the micro-level using models combining information on individual offense rates, probability of conviction, and sentence length. Aggregate research is faced with the problem of separating the effect of imprisonment on crime from the effect crime has on imprisonment. Additionally, this type of research can only address the combined deterrent and incapacitative effects of imprisonment. Micro-level research does focus solely on imprisonment's incapacitative effect, but finds that outcomes vary widely depending on the estimates of the various criminal career characteristics used. Also, thus far, these latter studies pertain to adult offenders only.

Macro-level Studies on Effects of Selective Incapacitation

Studies into the effects of selective incapacitation on crime have used one of two approaches: macro- and micro-level studies. Macro-level or, as Spelman (2000a) refers to them, top-down studies have relied on aggregate data on society-level crime rates and prison populations. These studies typically estimate the relation between crime rates and prison populations, while controlling for other variables that may affect crimes rates like the population's age distribution or its unemployment rate using multivariate regression techniques.

Researchers using aggregate data have specifically focused on the implementation of selective policies, most often the "three strikes" laws (Clark et al. 1997; Vitiello 1997; Zimring et al. 2001). As with the studies on imprisonment in general cited above, many of the studies into selective imprisonment did not control for simultaneity. This may lead effects of these selective policies to be underestimated, since outcome measures may also be affected by preexisting increasing crime trends that gave rise to implementing the selective policy in the first place. Many studies also have a short follow-up period. However, as Kovandzic (2001) notes, the incapacitative effects of selective policies most probably do not occur for many years after the policy is implemented, since most offenders selectively imprisoned under the new policy would have been imprisoned for at least some period even under prevailing nonselective policy. Using long-term data from Florida, Kovandzic (2001) finds little evidence for a general deterrent effect of Florida's habitual offender laws on crime and concludes that its limited effects on crime are mostly due to incapacitation.

It must be noted, however, that while policy shifts on imprisonment in the USA have been both quick and dramatic, these data are far from experimental (Spelman 2000a). Decisions to expand imprisonment or to implement a selective policy did not occur randomly, and therefore results of these changing conditions are confounded with existing between-state differences in crime trends and prison rates and all other factors that may have influenced either one of these variables. Results, even of the most complex models, are therefore still potentially subject to bias.

Micro-level Studies on Effects of Selective Incapacitation

Instead of relying on aggregate data, micro-level, or bottom-up, studies try to model the workings of the criminal justice system in more detail. Making estimates of the offense rate per year, the probability of arrest and conviction, and the average sentence length, these studies calculate the proportion of the criminal career that a typical offender will spend imprisoned. This proportion then represents the benefits of imprisonment in terms of decreased crime rates. Unlike most aggregate studies, these micro-level studies focus solely on the incapacitative effect of imprisonment.

Recent studies by DiIulio and Piehl (1991), Spelman (1994), and Piehl and DiIulio (1995) have yielded elasticity estimates ranging from −0.16 to −0.26. However, recognizing the still limited precision of the estimates used, Spelman (2005) concludes that it is more reasonable to expect a 0.1–0.3 % decrease in crime for a 1 % prison expansion. Micro-level studies necessarily make assumptions about several criminal career characteristics. Small – and plausible – changes have been demonstrated to considerably affect the studies' outcomes (Spelman 1994). While over the years these studies have gained validity, as a result of more detailed data on criminal career becoming available, they still need to rely on simplifications regarding the distribution and interrelationships of the various criminal career dimensions (Spelman 1994, p. 110).

In an effort to circumvent uncertainties and simplifications tied to the use of criminal career estimates, Bernard and Ritti (1991) designed a rolling cohort methodology that allowed them to estimate the incapacitative effect of selective policies using actually observed data. Their analyses were based on count data from the Philadelphia birth cohort. In brief, they simulated the implementation of some hypothetical selective policy and then counted the actual number and type of offenses recorded after the youth had been incapacitated under the hypothetical policy. These offenses would make up the incapacitative effect of the selective policy since they would not have occurred under the selective policy, but actually did occur under existing, nonselective policy. Based on only the "chronic" offenders in the Philadelphia cohort (those offenders who accumulated at least five police reports), a selective policy which entailed imprisoning a youth after his second arrest until his eighteenth birthday resulted in a 25 % decrease in all reported crimes and a 35 % decrease in the serious crimes reported for this cohort. A similar study using Swedish birth cohort data showed that imprisoning every recidivist in this cohort for a period of 2 years would reduce registered crime for this cohort by 28 % (Andersson 1993). However, while both studies found that selective incapacitation would reduce crime, both studies also showed implementation of a selective policy would markedly inflate prison populations by as much as 22 times the rate under nonselective policy.

Finally, Blokland and Nieuwbeerta (2007) use data from the Netherlands Criminal Career and Life-course Study to estimate the incapacitative effects of alternative selective prison policies. Using the rolling cohort method, implementations of various penal scenarios differing in selection rate, sentence disparity, and selective accuracy are simulated. Results show that it is hard for selective policies to yield a positive societal result: costs of imprisonment typically exceed benefits gained from crimes prevented. Selectively incapacitating presumed frequent offenders on the basis of the number of offenses they have committed in the past – or in the preceding 5 years – leads to a substantial decline in crime. This decline is the greatest in the first years after the introduction of the selective policy as a rising share of offenders are selectively imprisoned. After some time, depending on the duration of the standard prison term, some of the offenders having completed their prison term rejoin the population, thus causing crime to rise again. The ultimate result of our simulations – assuming only incapacitative effects – is a 25 % decrease in crime under the strictest regime. However, these benefits are offset by the fact that the introduction of a selective policy causes the prison population to rise considerably – up to 45 times the population under prevailing nonselective policy for a very strict selective regime. As a result, if a selective policy based on the total number of prior convictions is to be efficient, the average conviction prevented by the selective incapacitation of frequent offenders would have to generate four to eight times as much in benefits than it would cost detaining an offender for 1 year.

The study by Blokland and Nieuwbeerta (2007), however, also shows that on the basis of the total crime costs for 2004, the benefits from preventing an average conviction are only approximately twice as high as the costs of 1 year of prison. Under the assumption of equal probability of conviction across all offenders, this would mean that even the selective incapacitation scenario with the most favorable CBR (five prior convictions in the entire criminal history, 5 years of prison) is only efficient if the costs per prisoner can be reduced from 190 to 58 Euros per day, which cannot be considered a realistic option. Allowing disparity in the probability of conviction, however, may – at current prison costs – result in a breakeven situation for some selective policies if frequent offenders are less than half as likely as non-frequent offenders to be convicted given an actual crime. Finally, risk estimation merely on the basis of criminal past generates high wastage of prison capacity. A lower selection rate (employing a stricter selection criterion) reduces this problem, but the resulting policy has hardly any added crime reduction value over prevailing nonselective policy.

Remarks on Selective Incapacitation

Selectively incapacitating offenders on the basis of an estimation of their future behavior rooted in their past criminal behavior points to the growing influence of risk-oriented thinking in criminal legislation and judicial policy. Frequent offenders belong to the category of offenders with a high risk of recidivism, and society thinks that they therefore ought to be removed from its midst. Offenders with a sufficiently extensive criminal history are thus penalized not for what they have done but for what they are expected to do in the future. This implies that a prison term is converted from a punitive into a preventive measure, one that our results suggest is used unnecessarily in many cases. Moreover, under a selective policy, there is no longer any relationship between the offense committed and the duration of the sentence. Selective incapacitation conflicts with a number of fundamental principles of criminal law systems. On the basis of our research results, which were mainly negative about the selective detention of frequent offenders, one may well wonder whether such a fundamental change to criminal law systems is actually opportune and even whether this policy has any empirical legitimacy. It is possible and even likely that criminological research would point to other policy alternatives, primarily those that do find support in criminological research, as a more obvious choice – for example, investing more in the socioeconomic functioning of prisoners by improving and expanding the probation and aftercare service.

Related Entries

- ▶ Estimating the Effects of Incapacitation
- ▶ Incapacitation
- ▶ Penal Philosophy and Sentencing Theory
- ▶ Theories of Punishment

Recommended Reading and References

Aebi MF, Aromaa K, de Cavarlay BA, Barclay G, Gruszczyñska B, von Hofer H, Hysi V, Jehle J-M, Killias M, Smit P, Tavares C (2006) European sourcebook of crime and criminal justice statistics, 3rd edn. Boom Juridische, Den Haag

Andersson J (1993) A longitudinal simulation study of incapacitation effects. Project Metropolitan No. 35. Department of Sociology, University of Stockholm, Stockholm

Bernard TJ, Ritti RR (1991) The Philadelphia birth cohort and selective incapacitation. J Res Crime Delinq 28(1):33–54

Blokland AAJ, Nieuwbeerta P (2007) Selectively incapacitating frequent offenders: costs and benefits of various penal scenarios. J Quant Criminol 23:327–353

Blumstein A, Cohen J, Nagin D (eds) (1978) Deterrence and incapacitation: estimating the effects of criminal sanctions on crime rates. National Academy of Sciences, Washington, DC

Blumstein A, Cohen J, Roth JA, Visher CA (eds) (1986) Criminal careers and career criminals, vol 1. National Academy Press, Washington, DC

Blumstein A, Canela-Cacho JA, Cohen J (1993) Filtered sampling from populations with heterogeneous event frequencies. Manage Sci 39:886–899

Chaiken JM, Chaiken MR (1982) Varieties of Criminal Behavior. Report No. R-2814-NIJ. Rand, Santa Monica

Clark J, Austin J, Henry DA (1997) “Three strikes and you’re out”: a review of state legislation. National Institute of Justice, Washington, DC

DiIulio JJ Jr (1990) Crime and punishment in Wisconsin. Wisconsin Policy Research Institute, Milwaukee

DiIulio JJ, Piehl AM (1991) Does prison pay? The stormy national debate over the cost-effectiveness of imprisonment. Brookings Rev 9:28–35

Horney J, Marshall IH (1991) Measuring lambda through self-reports. Criminology 29:471–495

Kovandzic TV (2001) The impact of Florida’s habitual offender law on crime. Criminology 39(1):179–203

Miranne AC, Geerken MR (1991) The New Orleans inmate survey: a test of Greenwood's predictive scale. Criminology 29:497–518

Moore MH (1986) Purblind justice: normative issues in the use of predictive or discriminating tests in the criminal justice system. In: Blumstein A, Cohen J, Roth JA, Visher CA (eds) Criminal careers and 'career criminals', vol 2. National Academy Press, Washington, DC, pp 314–335

Owens EG (2009) More time, less crime? Estimating the incapacitative effect of sentence enhancements. J Law Econ 52:551–579

Piehl AM, DiIulio JJ (1995) Does prison pay? Revisited: returning to the crime scene. Brookings Rev 13:20–25

Piper ES (1985) Violent recidivism and chronicity in the 1958 Philadelphia cohort. J Quant Criminol 1(4):319–344

Spelman W (1994) Criminal incapacitation. Plenum Press, New York

Spelman W (2000a) What recent studies do (and don't) tell us about imprisonment and crime. Crime Justice 27:419–494

Spelman W (2000b) The limited importance of prison expansion. In: Blumstein A, Wallman J (eds) The crime drop in America. Cambridge University Press, Cambridge

Spelman W (2005) Jobs or jails? The crime drop in Texas. J Policy Anal Manage 24(1):133–165

Sweeten G, Apel R (2007) Incapacitation: revisiting an old question with a new method and new data. J Quant Criminol 23:303–326

Vitiello M (1997) Three strikes: can we return to rationality? J Crim Law Criminol 87:395–481

Wermink H, Apel R, Nieuwbeerta P, Blokland AAJ (2012) The incapacitation effect of first-time imprisonment: a matched samples comparison. J Quant Criminol. doi:10.1007/s10940-012-9189-3

Zimring FE, Hawkins G (1995) Incapacitation: penal confinement and the restraint of crime. Oxford University Press, New York

Zimring FE, Hawkins G, Kamin S (2001) Punishment and democracy: three strikes and you're out in California. Oxford University Press, New York

Crime Profiling

▶ Criminal Investigative Analysis

Crime Repetition by Youth

▶ Specialization in Juvenile Offending

Crime Scene Analysis

▶ Criminal Investigative Analysis

Crime Scene Interpretation

▶ Criminal Investigative Analysis

Crime Scene Profiling

▶ Behavioral Investigative Advice

Crime Science

Gloria Laycock
Jill Dando Institute of Security and Crime Science, University College London,
London, UK

Overview

With detection, arrest, and prosecution as the universal response to crime, despite their manifest failure to reduce the problem, the time is right for a complete rethink about crime and crime reduction. This entry considers the need for a new discipline – crime science – which it is argued is necessary if crime control is to be improved. The entry suggests that greater emphasis on experimentation and a more scientific approach to crime problems under the banner of "crime science" will make a significant difference. Examples are provided of the advantages of scientists working together to address crime problems but the entry also argues for greater emphasis on ethical and aesthetic issues if we are to avoid potential pitfalls of this new approach.

Introduction

This entry discusses aspects of crime science, which is a new approach to the control of crime, although its roots go back a further 30 or more years. It was formally established in 2001 at the suggestion of Nick Ross – a journalist and broadcaster who with others established the Jill Dando Institute of Crime Science in memory of his murdered colleague. A discussion of crime science in an encyclopedia on criminology may seem out of place but criminology and crime science are close in background, interests, aims, and objectives which makes them close bedfellows, and followers of each discipline should see each other as potential collaborators and close associates rather than as rivals for students, grants, or kudos (Smith and Tilley 2005). Indeed many criminologists could, if they so chose, also describe themselves as crime scientists rather as a rocket scientist could also call himself a physicist or just a scientist.

The next section describes crime science – it looks at some of the history of the subject and the reasons for its development, what it is, what it is not, and how it relates to criminology and other mainstream disciplines. The following section describes the "state of the art," briefly outlining some of the key findings of crime science most of which derive from the roots of the subject in environmental criminology (which is not directed at the study of crimes against the environment; it is concerned with the immediate "environmental" determinants of criminal behavior). After that, we look at some controversies – as with any science, there are a number of issues relating to the subject matter itself but in the case of crime science there are also questions around whether there really is a need for the discipline and how it can achieve one of its aims which is to adopt a multidisciplinary approach to the problem of crime control. The final section considers some "open questions" associated with crime science needing further thought or greater academic attention.

Background Description

It is suggested that there are four major and related elements linking science and crime (Laycock 2005). Firstly, science can help us by improving our understanding of crime and its causes. This is the domain of the social sciences in general and criminology, psychology, and sociology in particular. Secondly, it can help us by making crimes more difficult to commit through the technologies that are developed on the back of it. Examples come from CCTV systems, car deadlocks and immobilizers, luggage scanners at airports, and so on. Thirdly, it can help us in catching offenders more quickly and bringing them to justice. The obvious example here is forensic science including forensic psychology, fingerprinting, and genetics and DNA technologies. Finally, it can help us in thinking like scientists think, that is, in testing hypotheses and developing a body of knowledge enabling the articulation of theory. It is this last point which is common to all those who see themselves as crime scientists and in this there are commonalities with experimental criminologists (Sherman 2009), which are expanded upon further below.

Crime science grew from environmental criminology and crime analysis, both of which have a historical focus upon crime prevention, although crime science takes a somewhat broader view of prevention and includes the detection and disruption of crime. It is thus an outcome-focused approach rather like medical science, with its emphasis on the control of disease through prevention, diagnosis, and treatment. Also like medical science, a range of professions is associated with the achievement of the aims. This is particularly so in preventive medicine where parents teach their children to wash their hands, local authorities provide drains and sewage systems, and the government provides inoculation programs, all of which contribute to the maintenance of good health in society. Similarly with crime prevention, individuals lock their homes and vehicles, look after their children and families, and support their communities in

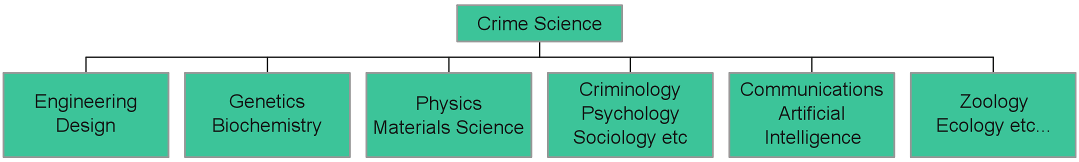
Crime Science
Engineering Design
Genetics Biochemistry
Physics Materials Science
Criminology Psychology Sociology etc
Communications Artificial Intelligence
Zoology Ecology etc...

Crime Science, Fig. 1 Relationship between crime science and other disciplines

preventing crime and disorder, while the central and local government provide a police service and criminal justice system and might also encourage commercial organizations and industry to play their part in crime prevention through the better design of goods, services, and management systems.

Medical science also makes use of a range of scientific approaches in addressing its problems, so we have medical physics and engineering where products are designed to deal with specific medical problems such as heart monitors and limb replacements. Chemistry also plays a major part in medical science in the development of drug treatments, as increasingly do biology and genetics. Again, crime science takes the same approach in encouraging other scientists to think about the application of their disciplines to the control of crime. Figure 1 below shows diagrammatically what we have in mind.

There is no significance to subjects being in the same box and it is not an exhaustive list, but note that the social sciences are included, particularly criminology and psychology which have an obvious special relevance to the control of crime.

Crime science is primarily about the testing of hypotheses associated with the understanding, prevention, disruption, and detection of crime. In the social context, this is a much more challenging task than in a laboratory, where control of extraneous variables is perhaps more easily achieved. Indeed some might argue that experimentation in society is neither possible nor desirable. Crime scientists would disagree and, as is demonstrated in the section below, not all experimentation has to be conducted in the real world. In order to carry out experiments or to speculate about cause and effect, we start with the collection of data. These data then need to be manipulated logically and rationally drawing on the currently best available evidence. In this way hypotheses can be developed and tested, theories articulated, and understanding increased.

Fundamentally, this is the scientific method (Townsley et al. 2003) which has been promoted to police practitioners and their partners under the acronym SARA (Eck and Spelman 1987) standing for scanning, analysis, response, and assessment. At each stage in this process, we are developing and discarding hypotheses. Take, for example, a police commander who may feel that in his police area there is a problem of theft of vehicles from outside residences overnight. This is his starting hypothesis, which it is possible to test using police data. His investigation (part of the scanning process) may show that there is indeed such a problem but that it is in only in the northern part of the police area and there is a much larger issue of theft of motor bikes and scooters in the south. It is then possible to formulate various hypotheses about why the thefts may be happening there (the analysis), to test them, and to implement a response or responses on the basis of the subsequent analysis (the response element of SARA). We then have the hypothesis that our action will reduce the problem by firing a particular mechanism (Pawson and Tilley 1997), which can then be refuted or supported (the final assessment stage of SARA). So we see that at each stage of the SARA process, hypotheses are being developed, tested, and discarded or, if supported, then acted upon.

The formulation of hypotheses is not as simple or straightforward as it might sound. It involves asking questions about what might be causing what, in what context, and then testing the resulting hypotheses. Experience of teaching crime science to practitioners over the past decade has shown it to be one of the most challenging aspects of the process. Articulating hypotheses involves asking clever questions, which is arguably what the best scientists do. Newton's simple question – why did the apple fall – which led to the discovery of gravity is a stunning question. Indentifying hypotheses is the point at which science involves judgment, which is not in itself a science. Judgment can be thought of as common sense, which as we all know is not too common; indeed Huxley is quoted as saying that "Science is simply common sense at its best, that is, rigidly accurate in observation, and merciless to fallacy in logic" (Attributed to Huxley: see http://quotes.prolix.nu/Science/).

It is through the articulation and testing of hypotheses that theories can be developed. The theoretical basis of crime science is that developed by the environmental criminologists and draws upon four "theories" or perspectives on crime. These are described in detail elsewhere in this volume [insert page number and reference]. In summary, they are the rational choice perspective, routine activity approach, crime pattern theory, and social/environmental and behavioral psychology. They are all concerned with the reduction of crime and the first three are sometimes called "opportunity theories" or the criminologies of everyday life (Garland 1996). Their focus is on the effects of the immediate environment on behavior, and although crime science would include the study of more remote causes of crime involving longitudinal studies, for example, it is largely concerned at this stage in its development with achieving more immediate reductions.

Another feature of crime science, perhaps obvious from its focus on the reduction of crime, is its practical nature. Although concerned with theory development, it is also concerned with practical outcomes. There are implications from this for the ways in which research is carried out and particularly for the ways in which it is presented. In this, crime scientists can take lessons from the work produced by the British Home Office Police Research Group which for over a decade from 1992 published short research reports which were intended to be read by intelligent practitioners, particularly the police, who may not have had specific research training and who would not have ready access to academic journals or even the interest in reading them if they did. Similar short summaries are now also available from a number of other agencies including the Australian Institute of Criminology and the US COPS Office website (www.popcenter.org).

One of the most widely publicized themes from the body of work carried out under the auspice of the Home Office Crime Prevention Unit and later the Police Research Group was related to repeat victimization (Laycock 2001), which specifically tested a variety of hypotheses over a 15–20-year period, all intended to develop understanding of the statistically observable fact that crime concentrates on victims, places, and products and that by concentrating on their subsequent protection crime can be reduced. In keeping with the desire to influence police practitioners, much of this work did not immediately feature in the academic literature, and this in itself causes problems for academics who are often rewarded in the universities not for the practical influence of their work but for its publication in top journals. A specific example of this is the booklet written by Clarke and Eck and first published as "Become a Problem-Solving Crime Analyst" in 2003 and later, with minor changes, published as a US version. It has since been translated into almost 20 languages and has been hugely influential in supporting the development of police problem solving and crime analysis: But it does not feature in the academic literature and most academics would not, therefore, be aware of it.

State of the Art

This section presents some of the more recent research carried out by crime scientists and which are at what we would currently see as the frontier of the research agenda. Examples are chosen from the four areas suggested as those on which crime science draws – social sciences which help our understanding of crime, physical sciences including computer science which can assist in the prevention of crime, forensic sciences which support detection, and the use of scientific method to test hypotheses, which all the sciences do but which are discussed here in the context of crime control.

Perhaps one of the most significant observations from criminology over the past few decades has been the extent to which crime concentrates in space (Sherman et al. 1989). This began as not much more than an interesting observation. Sherman's early work, for example, showed the extent to which there were "hot spots" of crime in major cities of the United States and which he suggested might direct police patrol. More specifically, as crime scientist Ken Pease (Pease 1998) demonstrated, crime also concentrates on individuals, and as noted above this phenomenon of repeat victimization has become the focus of a number of hypotheses about ways in which to prevent crime (Laycock 2001). The conclusion being: Protect victims and crime goes down. This appears to be the case in relation to a number of different offenses and can be observed in a range of different jurisdictions (Farrell and Pease 2008).

Recent work has investigated the way in which these spatial crime clusters are built up. For example, Bowers et al. (2004) have demonstrated that crime clusters in space *and* time. Domestic burglary hot spots are built up in ways similar to those seen in the foraging of animals (Johnson et al. 2009). The burglar will "hit" an area and then move on before he or she can be arrested. The researchers draw on the statistical methods used by epidemiologists in analyzing their data. They argue that rather like an outbreak of disease, the crimes occur in spates, which are often small enough for the police to patrol.

In a series of related studies, Johnson and Bowers have shown that repeat victimization, which is partially explained by their work described above, has also led to near neighbors of burglary victims being at heightened risk for a short period following the initial offense. Working with international colleagues, they have demonstrated that the same effect can be found in other jurisdictions (Johnson et al. 2007). All of this work has clear implications for crime prevention and detection.

We can see from this substantial body of work the extent to which initial statistical observations from criminologists provided fertile ground for the development of hypotheses by crime scientists, which could then be developed to both inform practice and further the development of theoretical understanding of the phenomenon of crime itself. It is noted here that Johnson, Bowers, and the majority of their international collaborators have relatively little formal training as criminologists.

There are a number of innovative approaches to crime control that might be discussed on the basis of technologies developed from various scientific discoveries. One of the more recent and certainly more controversial is the introduction of full-body scanners at airports and other transport hubs, which are intended to reduce the terrorist threat or to address smuggling and other offenses which involve carrying contraband on the body. The majority of these scanners are based upon either backscatter X-ray technology, which detects radiation reflected from the body, or passive or active millimeter wave technology which interprets natural radiation from bodies (in the case of passive millimeter scanners) or which reflects extremely high-frequency radio waves off the body (for active millimeter scanners) in order to detect illicit material that may be hidden underneath clothing. The potential threat to health is an obvious issue, which is still the subject of discussion, although the extent of the threat depends upon the system that is used. A problem common to all such systems is, of course, the potential invasion of privacy. The fact that these scanners can operate relatively quickly and do not require the subject to remove clothing is likely to increase their use, without any particular "just cause" as might otherwise be expected. This is an example of the importance of alerting the scientists and technologists to the social consequences of some of their inventions. Although it could be argued that innovators are not responsible for the use to which their innovations are put, there is increasing interest in encouraging scientists to think through some of these ethical issues as is discussed more fully below.

Computer science offers a particularly interesting contribution to crime control first through the realization that prevention must be paramount. Given the ubiquity of the use of computers and the (apparent) anonymity offered by the World Wide Web, effort spent in detecting crimes, which may have been committed by offenders in different countries, is easily seen as futile under most circumstances. So it is very clear that the design of the system itself must be such as to reduce the opportunities for crime and other abuses. We thus see arrangements for online banking, shopping, e-mailing, etc. continuously subject to improvements in the security arrangements associated with them. One observation in relation to the development of computer security systems is the need to take greater account of so-called wetware, that is, humans. Designers of computer password systems make a point of ensuring that the systems will refuse to recognize easily remembered passwords, and as a consequence people tend, for example, to write down passwords on bits of sticky paper or in other insecure places, which of course defeats the purpose. Having the designers of Internet security systems work together with psychologists in the development of these systems should improve the systems themselves and reduce the risk to security. This is a clear example of the advantage of individuals from more than one discipline working together to address a common problem.

Forensic science is now a well-established approach to the detection of crime, and with the developments associated with fingerprint technology and DNA, we see some powerful techniques becoming available to the police and other security experts. Forensic science has a long history of scientists from different disciplines working together, all of whom might call themselves crime scientists but who might also benefit from some of the additional perspectives offered by the social sciences. So, for example, in deciding how to apportion scarce forensic resources when investigating domestic burglary, priority might be given to repeat crimes where we know that there is around an 80 % chance that the same burglar has returned. Such a burglar would almost certainly have a record on the UK National DNA Database and we also know that although a DNA trace is found in only about 5–6 % of burglary offenses, when it is found, there is an over 40 % chance of finding a link to a known offender (Burrows et al. 2005): This percentage is likely to rise in the case of repeat instances of burglary. So a little knowledge of criminology on the part of the forensic scientist would arguably improve their day-to-day performance.

Turning to what is new in forensics, we might look at forensic geosciences, which offer the promise of a whole new branch of science in support of crime detection (Morgan and Bull 2007). Drawing on insights from geology, they explore the distribution of soils, sediments, and dust from crime scenes and address the extent to which they transfer from one place or context to another. It is a fundamental of geography that every contact leaves a trace, and this is what much forensic science is based upon. It has been recently demonstrated, for example, that sand and soil particles have a "fingerprint" or characteristic profile depending upon where they are found. It is thus possible to challenge an assertion by a suspect that he or she was not present at a particular site by comparing the profile of granules found at a crime scene with those on the footwear of the suspect.

All of the sciences test hypotheses but as noted earlier this can be problematic in the real world. It may be unethical, impractical, or politically unacceptable to carry out controlled experiments in the social context. Recent developments in computer science have pointed to a possible way out of this dilemma. Groff and Birks (2008) suggest that new crime prevention approaches could be tested in an artificial computer-generated world before being tried in reality. In a more recent article, Birks et al. (2012) provide a good introduction to computational modelling or more specifically agent-based modelling (ABM). They tested whether the hypothesized mechanisms of environmental criminology were sufficient to explain the observed characteristics of residential burglary – particularly its spatial concentration (hot spots), its concentration on victims (repeat victimization), and the characteristic journey to crime curve. The results showed that the propositions of the routine activity approach, rational choice perspective, and crime pattern theory provide a viable generative explanation of those independent crime characteristics.

This work is an exciting first step in what looks like becoming a rich research vein for crime scientists. The article goes on to provide a number of further projects that might be conducted and through which we might explore the theories and mechanisms associated with crime reduction. One of the advantages of the approach is that it forces the experimenter to be explicit about the mechanisms that are presumed to lead to the reduction in crime and to be highly specific in the articulation of the hypotheses that are being tested (Sidebottom and Tilley 2011).

Controversies

To call this section "controversies" might be overstating the case, but there are a number of issues with various degrees of contention, which merit some discussion. First, and perhaps most controversially, why not stay within the criminological discipline? Why do we need crime science as a separate enterprise? There are several of reasons. Perhaps the best has been well said by Ken Pease (Pease 2008) in a discussion of crime science in which he reviews the psychological concept of "framing" and the way in which it can be used to describe the partitioning of academic disciplines. Such partitioning, while helpful to universities and employers in facilitating the organization of teaching and in informing assumptions about subsequent knowledge, is also contributory to constraining thinking into "silos" which do not help the development of science in general or problem solving in particular. As Pease puts it:

> The cognitive structures into which thinking is organized are known to psychologists as frames. Frames limit cognitive commerce. Thus the primary disadvantage of organizing the study of the natural and social worlds into disciplines is that the frames come to act as artificial barriers to thinking. It is difficult to overstate the importance of framing in the organization of knowledge. Problems are "framed" in ways that make them susceptible to the skills of the framer. (Pease 2008, p. 4)

He goes on to argue that criminology is rooted in sociology (with links to psychology and law), and in line with the notion of framing, criminology's "solutions" to crime problems and indeed the conceptualizations of crime itself are similarly (over-)influenced by those frames of reference. Crime science is thus an explicit attempt to overcome the negative effects of framing and to make explicit the advantages of working across disciplines.

In his further justification for the emergence of crime science, Pease is critical of criminology. To quote Pease again:

> The decision to champion such a discipline shift is rooted in the perception. . ..that the status, progress, and influence of the understanding of crime and criminality achieved by scholarship labeled criminology has. . ..been disappointing both in itself and in its application to the urgent social problems to which it is addressed. (Pease 2008, p. 14)

Pease is also critical of some of the subgroups of criminology and singles out for particular comment the latest iteration of curious thinking from the so-called cultural criminologists. They, he argues, see crime as culturally defined and take scholarship to mean critical reading of the written word rather than empirical, data-driven investigation. If crime science is at one extreme on a continuum, then critical criminology appears to be at the other.

So Pease argues for the establishment of crime science and a break from criminology on three main grounds: First, criminology suffers like most disciplines from the effects of framing; secondly, it has failed to deliver what is required in contributing to the control of crime. On this later point, he is joined by a number of criminologists including past presidents of the American Society of Criminology. They have chosen to try to change the discipline from within while Pease prefers to leave it. Thirdly, some (but Pease concedes not all) criminologists are in his view (and mine as it happens) totally wrong in their approach to crime control. They are members of the criminology "club" and any club that accepts them as members is one to which Pease would rather not belong, as he puts it (paraphrasing Groucho Marx).

The relationship with criminology described in the definition of crime science outlined in this entry is different from that of Pease, being better described as cherry-picking collaboration. As argued above, and as applies to other social sciences, criminology has much to offer the crime scientist in its insights about crime, ethics, and experimentation as exhibited, for example, by the experimental criminologists. But crime scientists argue that there is a need to do more than work with criminologists if the concern is the control of crime (its prevention, disruption, and detection) ethically and with due deference to good design (i.e., we do not like the Fort Knox option, effective though that might be). We would wish to embrace all the sciences and in particular to approach crime control using scientific methods, that is, to experiment.

Why call it crime science? What is wrong, for example, with police science as suggested by Weisburd and Neyroud (2011)? The arguments for the label crime science have been rehearsed elsewhere, see, for example, Laycock (2012); but in a nutshell, while we share many of the methodological and philosophical ideas and approaches of the police scientists, we would include the operation of the courts, prisons, and probation services as of interest to crime science. We would also wish to address issues of concern to communities such as minor disorder or other problems, which would hopefully fall short of policing. So unless we also want to create prisons scientists, probation scientists, and criminal justice scientists (the list could be longer), we have opted for crime scientists. Furthermore, we draw an analogy with medical science, which is notably not called doctor science and which concerns itself with the prevention, detection, and treatment of diseases. In doing so, it does not rely solely on doctors but, as noted above, expects support from the central and local government, from parents, from schools, teachers, and others in advocating healthy lifestyles. There are direct and obvious analogies with crime control.

There are other issues associated with crime science stemming from its roots in environmental criminology, a substantial component of which is situational crime prevention (SCP). A classic criticism of SCP is that it is no more than target hardening and pays no attention to social issues but concentrates on turning social space into a fortress. This criticism is wrong on at least two counts. First, it is not the case that SCP is no more than target hardening. There are currently five major approaches to SCP, all of which have been demonstrated in different contexts to reduce crime – these are the mechanisms (Pawson and Tilley 1997) through which crime is reduced. They are:

1. Increase the perceived *effort* required of the potential offender to commit the crime (this includes but is not restricted to target hardening).
2. Increase the perceived *risk* to the offender (this includes, but again is not restricted to, increasing surveillance such as introducing CCTV).
3. Reduce the expected *reward* (e.g., putting dye capsules on clothing to deter shop theft).
4. Remove *provocation* (e.g., controlling taxi queues around night clubs to deter queue jumping).
5. Remove *excuses* (e.g., setting clear rules for children about the unacceptability of bullying in schools).

Under each of these headings can be found a wide variety of tactics (Tilley and Laycock 2001) but their adoption should depend upon a proper understanding of the crime or disorder problem being addressed, an appreciation of the context within which the tactic is to be applied, and some judgment about its appropriateness such as the extent to which it might cause an oppressive environment, be unethical or an overreaction to an otherwise relatively minor problem. At this point, the decision whether or not to implement becomes one of judgment.

It is also perhaps worth noting here that the default response from the police to the prevention of crime is typically increased patrol, CCTV, or walls, barriers, and other target-hardening approaches. This is perhaps why SCP has become seen as so restricted. Experience of teaching problem solving and the use of the SARA process over a number of years suggests that it can be quite difficult to persuade the police to broaden their response options beyond increasing the perceived effort and risk to the offender.

There have been a number of points so far when the notion of judgment has been mentioned. This has arisen in relation to deciding on which problem to address, which hypotheses to test, and which responses to adopt. Good judgment is an integral part of scientific development. Science is not the barren, formulaic process it is sometimes supposed to be. It should be a creative and vibrant activity that is exciting for the scientist and those who work with them. Some of the latter include the politicians whose job is to decide whether and if so when to adopt the innovations of science and technology and to implement them in society. This again is a judgment: It often requires balancing the rights of individuals with the need to keep society safe or the aesthetic appearance of a neighborhood with the need to reduce crime. In making these judgments, politicians or those who work on their behalf need a reasonable estimate and understanding of risk. It is often this that they are balancing. So, for example, if the risk of a terrorist attack is low, is the cost of implementing a potentially intrusive surveillance system defensible? For the politician, that cost is more than financial: It would include the political cost, that is, they can lose their job. They also would want to take account of precedent, public reaction and what the media might say. This is the policy context within which science and technologies are developed, and although it may be the responsibility of the scientist to ensure that the decision makers are aware of the scientific and technological options, it does not follow that decision makers will simply follow the most scientifically rational route. If scientists aspire to influence judgments (as crime scientists do given their wish to influence policy and practice), then it is perhaps appropriate if they understand this wider decision-making context and how it can be used to further scientific endeavor; science does not sell itself.

The final issue to be discussed here in the current debate on the development of crime science is the level within the universities at which it is taught. The first university to teach crime science was University College London, which is the home of the UCL Jill Dando Institute of Security and Crime Science (JDI). The teaching there has so far been largely although not exclusively restricted to postgraduate training. This has meant taking postgraduate students trained in an acknowledged scientific discipline and exposing them to the challenges of crime control. An advantage of this is that the students either come ready versed in scientific method or are positive about the approach. It compares with the early development of forensic science, for example. Social sciences differ in the extent to which they welcome experimentation, but even within those disciplines that do not have a hard science core, there appears to be a ready market for this more practical and scientifically based style. Other universities are now offering similar programs as crime science slowly develops with some providing courses at undergraduate level.

There is a strong argument for keeping with the postgraduate option and not encouraging the development of undergraduate courses. To do so seriously risks establishing just another academic silo with all the framing difficulties as set out by Pease. We would basically be back to the beginning as far as the encouragement of multidisciplinary working was concerned. Nobody should underestimate the difficulty of establishing multidisciplinarity, which is one of the current buzzwords in research funding circles. While we might be able to point to outstanding examples of cross-disciplinary working, which have led to major scientific breakthroughs such as the work on DNA by Watson and Crick, those collaborations were close to serendipitous. Actually creating an academic context within which similar joint working might be fostered is a real challenge, particularly in today’s world when there is less opportunity in universities for the casual conversations in common rooms seen as a feature of the rosy academic past.

UCL has been extremely fortunate in this area. The Engineering and Physical Sciences Research Council in the UK, which is one of the major funders of science, provided funding to UCL's JDI in 2009 to establish a Doctoral Training Centre in Security Science. This involved a £7 million grant to support 10 fully funded PhD students per year for 4 years. UCL undertook to double those numbers. The funding enabled the recruitment of scientists from a range of disciplines to be offered studentships and to work in the same physical space, with all the day-to-day interaction opportunities that this implies, and to share common courses about crime and the ways to control it. This is an ongoing program and its success or otherwise will be closely watched. So far, it has brought computer scientists, mathematicians, political scientists, linguists, and engineers together to carry out research.

Open Questions

Following from the previous section, one of the open questions has to be the extent to which crime science can avoid the "silo pitfalls" of the past and genuinely produce novel solutions to today's problems which demonstrate sensitivity to ethical concerns and good design. On this, the jury has to be still out. A second related question is whether or not crime science can add value to crime control through investment in a multidisciplinary approach – indeed at this stage, it is an open question whether different disciplines will be working together on common problems or whether their interactions will be restricted to coffee and biscuits. On this, the recent launch of a crime science open access journal might in the longer term prove helpful in providing an appropriate academic outlet for joint papers which might otherwise struggle for acceptance in the well-established mainstream disciplinary journals. Fundamentally crime science is attempting to generate a common language and understanding of the potential methods of crime control across disciplines that might not normally exist.

The aspirations of crime scientists to affect policy and practice are laudable, if potentially challenging, not least because there is a fundamental difference in the contingencies under which practitioners, policy makers, and researchers operate. For example, practitioners are not typically rewarded for reading research papers or for their understanding of science, while policy makers increasingly have an eye on the media representation of their policies rather than the extent to which they may reflect risk or rationality as a scientist may see it. Whether or not crime scientists can better influence policy and practice than, say, criminologists have managed to do is therefore an open question. The way to deal with this, however, is to change those contingencies not to complain about the lack of influence of research over policy and practice. The questions are as follows: How do we make practitioners (in the present case largely the police) more interested in what they perceive as irrelevant research? How do we motivate researchers better to communicate with the practitioners? How do we encourage policy makers to pay more attention to scientific research results? Until these questions are addressed, the lack of connection of the police with science, as described by Weisburd and Neyroud (2011), for example, will remain.

There is of course a huge range of further open questions when we turn from what might be described as administrative issues to the more academic ones. First, much more work needs to be done on ethics and legitimacy specifically in relation to the implementation of various technological solutions to crime problems. A tendency to exaggerate the risk or threat of crime (particularly terrorism) leaves societies vulnerable to the unfettered implementation of "hard" solutions to crime problems, which may be effective but which in the longer term may also be undesirable.

The most important research questions are ultimately a matter of opinion and inclination, but for the writer, as a crime scientist with an interest in crime prevention as a means of controlling crime rather than detection, arrest, sentencing, and ultimately imprisonment, the key question has to be how to design a social system which minimizes criminal opportunities while protecting rights, freedoms, and aesthetics. This might remain an open question for some time to come.

Recommended Reading and References

Birks D, Townsley M, Stewart A (2012) Generative explanations of crime: using simulation to test criminological theory. Criminology 50(1):2012–2221

Bowers KJ, Johnson SD, Pease K (2004) Prospective hotspotting: the future of crime mapping? Brit J Criminol 44(5):641–658

Burrows J, Hopkins M, Hubbard R, Robinson A, Speed M, Tilley N (2005) Understanding the attrition process in volume crime investigations. Home Office Research Study 295, Home Office, London

Clarke RV, Eck J (2003) Become a problem solving crime analyst, UCL Jill Dando Institute. Available from Routledge publishers or www.popcenter.org. Accessed 23 May 2012

Eck J, Spelman W (1987) Problem-solving: problem-oriented policing in newport news. Police Executive Research Forum, Washington, DC

Farrell G, Pease K (2008) Repeat victimization. In: Wortley R, Mazerolle L (eds) Environmental criminology and crime analysis. Cullompton, Willan. ISBN 978-1-84392-280-3

Garland D (1996) The limits of the sovereign state: strategies of crime control in contemporary society. Brit J Criminol 36(4):445–471

Groff E, Birks DJ (2008) Simulating crime prevention strategies: a look at the possibilities. Policing J Policy Pract 2(2):175–184

Johnson SD, Bernasco W, Bowers KJ, Elffers H, Ratcliffe J, Rengert G, Townsley MT (2007) Near repeats: a cross national assessment of residential burglary. J Quant Criminol 23(3):201–219

Johnson SD, Summers L, Pease K (2009) Offender as forager? A direct test of the boost account of victimization. J Quant Criminol 25:181–200

Laycock G (2001) Hypothesis based research: the repeat victimization story. Crim Just Int J Policy Pract 1(1):59–82

Laycock G (2005) Defining crime science. In: Smith MJ, Tilley N (eds) Crime science: new approaches to preventing and detecting crime, Crime science series. Willan Publishing, Devon, UK

Laycock G (2012) Happy birthday? Policing J Policy Pract 6(2):101–107

Morgan RM, Bull PA (2007) The philosophy, nature and practice of forensic sediment analysis. Prog Phys Geog 31(1):43–58

Pawson R, Tilley N (1997) Realistic evaluation. Sage, London

Pease K (1998) Repeat victimisation: taking stock. Home office crime prevention and detection Series Paper 90. Available from http://webarchive.nationalarchives.gov.uk/20110218135832/http://rds.homeoffice.gov.uk/rds/prgpdfs/fcdps90.pdf. Accessed 24 May 2012

Pease K (2008) Crime science. In: Shoham SG, Knepper P, Kett M (eds) International handbook of criminology. CRC Press, Boca Raton

Sherman L (2009) Evidence and liberty: the promise of experimental criminology. Criminol Crim Just 9(1):5–28, ISSN 1748–8958

Sherman LW, Gartin PR, Buerger ME (1989) Hot spots of predatory crime: routine activities and the criminology of place. Criminology 27:27–55

Sidebottom A, Tilley N (2011) Further improving reporting in crime and justice: an addendum to Perry, Weisburd and Hewitt (2010). J Exp Criminol. doi:10.1007/s11292-011-9128-6

Smith M, Tilley N (eds) (2005) Crime science. Willan, Cullompton, Devon

Tilley N, Laycock G (2001) Working out what to do: evidence-based crime reduction. Crime reduction series paper 11. Home Office, London. ISSN: 1468–5205, ISBN 1-84082-792-0

Townsley M, Johnson SD, Pease K (2003) Problem orientation, problem solving and organizational change. Crime Prev Stud 15:183–212

Weisburd D, Neyroud P (2011) Police science: toward a new paradigm. National Institute of Justice, Washington, DC. Available from https://www.ncjrs.gov/pdffiles1/nij/228922.pdf. Accessed 24 May 2012

Crime Simulation

► Simulation as a Tool for Police Planning

Crime Spates

► Prediction and Crime Clusters

Crime Specialization, Progression, and Sequencing

Christopher J. Sullivan and Shaun M. Gann
School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA

Overview

This entry describes what is known about three key dimensions of the active criminal career:

frequency, specialization, and progression/ sequencing. Each of these topics is considered in terms of its theoretical implications and the empirical research that has been conducted on it. A great deal has been learned about these topics since they emerged early in the discussion of criminal careers. There are some questions about the ability to measure and assess them and the implications of doing so, however, and those are reviewed here as well.

Background on Dimensions of Active Criminal Careers

The Study of Criminal Careers

Understanding the criminal career is important as it can offer insight into the nature of offending generally and help in explaining the behavioral paths of offenders over time. This can, in turn, inform policy and practice that relates to career criminals as well as those who offend for only a limited period of time. The idea of a criminal career was raised in early life history research with delinquents. Later on, a report by Blumstein and colleagues (1986) outlined a number of relevant characteristics of criminal careers that might be described or explained by criminologists, and the idea of a criminal "career" became a conceptual framework against which empirical findings could be judged and theories could be developed (Piquero et al. 2003). This began a trend toward quantitative analysis of the dimensions of long-term careers in offending. Among those that were initially defined in that report or have received attention since are (a) the age of onset of careers, (b) the length and continuity in careers, (c) the process of desistance from criminal behavior, (d) the frequency of offending, (e) specialization or versatility in the types of offenses committed, and (f) progression of careers in terms of possible escalation or stepping down in the seriousness of offending. Each of these can serve as a touchstone for theory or policy/practice with respect to crime and criminality.

According to Smith et al. (1991:8), "the criminal career paradigm is not a theory of crime but rather a framework that permits more detailed study of criminal behavior and its causes by identifying dimensions of individual criminal careers." One advantage of separating the dimensions of the criminal career is that criminological theories can be applied to a specific dimension (or dimensions), allowing theorists to be precise in their propositions. For example, one theory may explain why offenders become involved in criminal activity, while another may explain the variation in frequency among active offenders.

This entry will focus on the last three dimensions of crime described above (frequency, specialization, and escalation), which are generally designated as dimensions of active criminal careers. Blumstein et al. (1986) asserted that high-frequency offenders were worthy of study from a career perspective because they "contribute disproportionately to the total measured number of crimes" (18). They also suggested that, within those careers, it is important to investigate questions of how crimes mixed and were arrayed as a possible means of better understanding offenders. This entry first provides some definitions for these concepts before moving into a synthesis of what is known about them. That discussion also considers potential controversies or disagreements that have come up over time. Finally, some "open questions" in each of these areas are discussed in order to provide a sense of where research and practice might go in the future.

Dimensions of the Active Criminal Career

The dimensions of the criminal career discussed here are concerned with the nature of offending from the standpoint of relative seriousness (including frequency) or type of crime across an extended period of time (Piquero et al. 2003). In the current discussion, "frequency" of offending is concerned with individual levels of criminal activity during a given timeframe. Knowledge of the frequency of offending has important implications for the potential costs associated with an individual's pattern of behavior. It has been at the heart of discussion of criminal careers because it is useful theoretically, and practically, to identify and explain those who offend at high rates as they tend to commit a large proportion of crime from an aggregate perspective as well (Piquero et al. 2007). For example, in their

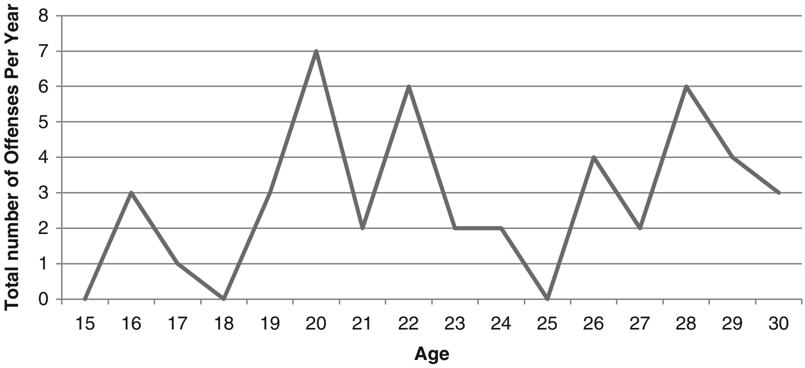

Crime Specialization, Progression, and Sequencing, Fig. 1 Illustration of individual frequency of offending (California Youth Authority Data)

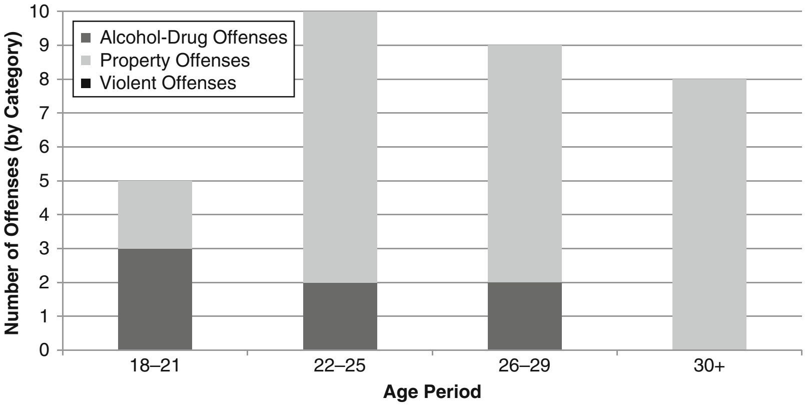

Crime Specialization, Progression, and Sequencing, Fig. 2 Illustration of specialization in offending (California Youth Authority Data)

study of almost 10,000 adolescent males in Philadelphia, Wolfgang and colleagues (1972) concluded that 627 of the boys committed a majority of the total crimes. Figure 1 shows a career pattern for an individual case drawn from data on California Youth Authority parolees (see Sullivan et al. 2009) comprised of years with relatively high and relatively low offending frequency values. This pattern would be that of an offender with a fairly high frequency who offends across a significant part of the life course. Given this individual's overall frequency of offending across the career, he might qualify as one of the minority of individuals who commit the majority of offenses identified in the Wolfgang study.

Offending specialization is primarily concerned with the nature of the crimes committed by individual offenders and the degree to which they form patterns across a career (or a portion of a career). To Osgood and Schreck (2007), specialization reflects systematic differences in the type of offenses committed by individuals. Systematic differences are those that are sizeable enough to distinguish one offender from another statistically. In most cases, specialization reflects some discernable pattern in form or type of offending across a defined period of time. Figure 2 offers a general sense of the contrast between specialization and versatility in offending. Over four time periods, this individual engages in no violent offenses and, although he starts with a mix of property and substance use offenses, is predominantly a property offender.

The issue of "progression" in criminal activity also reflects the presence of offending patterns but generally considers the nature of the acts in terms of seriousness as opposed to just type. The ideas of "escalation" and "de-escalation" in the seriousness of offending fall into this general

Crime Specialization, Progression, and Sequencing,
Fig. 3 Illustration of escalation in offending (California Youth Authority Data)

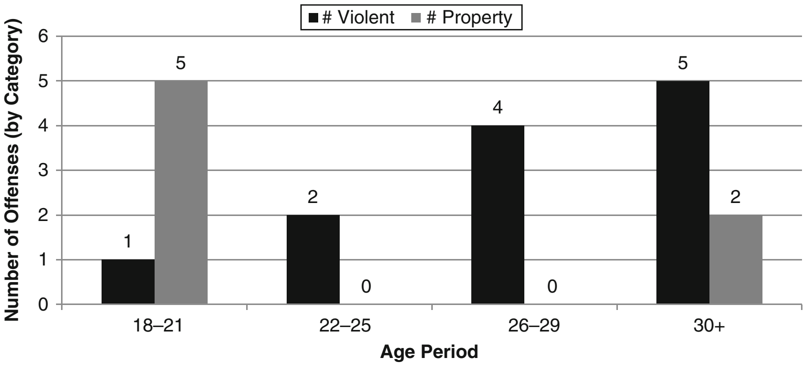

framework (Blumstein et al. 1986; Piquero et al. 2003). Although escalation can be framed in terms of the increased frequency of offending, it is more generally be defined as a tendency to commit more serious offenses as the criminal career unfolds (Piquero et al. 2003). The idea of escalation/de-escalation inherently asks the question of whether an individual's criminal behavior is getting "better" or "worse." If one assumes that violent offenses are more serious than property offenses, the notion of escalation can be observed in Fig. 3 by contrasting the relative prevalence of the two offense types over time. Initially, this individual is involved in far more property crimes, but these wane over the observed time period, while violent crimes increase over time.

State of the Art and Controversies

Frequency in Offending Careers

Measuring and making sense of the frequency of offending from a criminal careers perspective would appear to be a straightforward endeavor. Still, there are different ways of framing questions regarding the frequency of criminal behavior. One of the earliest controversies in research on criminal careers hinged around the distinction between total observed crime and individual frequency of offending (Blumstein et al. 1988; Gottfredson and Hirschi 1986). A society's aggregated age-crime curve is necessarily composed of a number of different individual events. Sometimes these represent actions of individuals who are involved for the first and only time, but, more often, they represent repeated acts by the same individuals (see discussion of Wolfgang study above). The objective of criminal career research rests partly in distinguishing the latter group from the former.

The number of offenses per year for inactive offender has been given a label of *Lambda* (λ) by some scholars (Piquero et al. 2007), and a great deal of the thinking and research in the study of criminal careers has focused on identifying and explaining the behavior of those high rate offenders. Because offending frequency research is concerned with active offenders, researchers employing the criminal career paradigm do not rely on aggregate crime rates. Per capita crime rates are calculated by dividing the total number of crimes committed in a specified area by the total population for the area. As such, they do not distinguish between those who commit crime (active offenders) and those who do not. Instead, frequency research relies on identifying and understanding Lambda.

Gottfredson and Hirschi (1990) argued that a single latent construct, criminal propensity, was the main influence on all aspects of the criminal career, including frequency. According to this view, the frequency of offending increases as criminal propensity becomes more elevated. Contrary to Gottfredson and Hirschi's argument, however, early criminal career proponents posited that different factors influenced each of the dimensions, specifically that different factors influenced participation and, among active offenders, frequency (Smith et al. 1991). As such, they called for research to determine whether different variables were correlated with the separate dimensions of the criminal career and, if so, which of those affected each dimension. Smith and colleagues (1991) attempted to determine whether the factors associated with participating in crime and delinquency were different from those associated with the offending frequency of active offenders. The results were inconclusive, indicating that most of the correlates of frequency were also correlates of participation. There were, however, a few variables associated with one dimension and not the other. For example, age was associated with participation but not with frequency. Overall, the authors concluded that their data did not support the argument that a general propensity toward crime explained each of the dimensions of the criminal career.

Lambda values vary significantly among individual offenders and are skewed to the right (Blumstein et al. 1986). In other words, the majority of offenders commit relatively few crimes each year, while a small portion of offenders commit crime on a more frequent basis. For example, using a representative sample of prison and jail inmates in California, Michigan, and Texas, (Chaiken and Chaiken 1982) found that among offenders who committed at least one burglary prior to their incarceration, the median Lambda value was relatively small at 5.5 burglaries per year, whereas offenders in the 90th percentile committed over 230 burglaries per year.

Research has found a few factors that are consistently correlated with increased frequency rates, including early age of onset, frequent drug use (Rhodes 1989), unemployment, and previous high levels of criminal involvement (Blumstein et al. 1986). However, research on the relationship between frequency and certain demographic variables – age, sex, and race – is mixed. In their report, Blumstein et al. (1986) concluded that these variables were highly correlated with an individual's participation in criminal activities but only marginally associated with individual frequency. For example, although an offender's sex is typically a strong predictor of participation in criminal activity (Liu et al. 2011), research has shown that the offending frequencies of active females are not significantly different from those of active males (Blumstein et al. 1986; Rhodes 1989). Like Blumstein and colleagues, Rhodes (1989) reported that active male and female offenders had almost identical offending frequencies. Regarding race, Blumstein et al. (1986) concluded that among active offenders, the frequency rates of black and white offenders were nearly identical. However, contrary to the conclusions from Blumstein and colleagues, Rhodes (1989) found that nonwhite offenders had significantly higher Lambda values than whites.

The relationship between age and the frequency of offending has been heavily debated; in part, these discussions hinge on disentangling individual offending from aggregate patterns. Gottfredson and Hirschi (1986); Blumstein et al. (1988a, b) took differing perspectives in that the overall pattern observed in the age-crime curve might arise out of different processes, each with its own etiology and distinct implications for theory and practice. Gottfredson and Hirschi (1986) attributed the decline in the aggregate age-crime curve that occurs after its peak in late adolescence to decreased individual offending frequencies or individuals "maturing out of crime." Proponents of the criminal career paradigm, however, argue that the causes of the decline are more extensive than simply maturing out of crime (Smith et al. 1991). They posit that the decline in aggregate crime rates may be influenced by both the number of offenders participating in crime and individual frequency rates. During early adulthood, many individuals terminate their criminal careers, while those that remain active commit crime at the same or elevated frequency. After reexamining the Glueck data, Blumstein et al. (1988a, b) found that annual arrest frequencies remained stable – ranging from 0.86 to 1.14 – for a sample of boys followed from ages 14 to 29.

The idea of career criminals, or offenders who commit relatively frequent crimes over an extended time period, is important to policymakers. The belief is that if these career criminals can be identified, policies aimed at preventing, controlling, or curtailing their careers – either through incarceration or other programs designed to prevent offending or recidivism – may produce a sizeable decrease in crime rates. One such policy that has been suggested (and incorporated in many jurisdictions) is selective incapacitation, which aims to identify and incarcerate serious career criminals in order to prevent the potentially significant number of crimes they would commit if free in the community (Gottfredson and Hirschi 1986). An example of selective incapacitation is habitual offender laws, or “three strikes” laws, in which repeat offenders are given a significantly longer sentence after their third felony conviction. In the original criminal careers report and other earlier works, Blumstein and colleagues (1986, 1988a, b) argued that additional research was needed to identify the legal variables correlated with high values of Lambda in order to rigorously consider and guide selective incapacitation policies.

Another possible approach is to use some of the identified correlates of high frequency of offending to identify targets for prevention efforts. For example, early age of onset is a noted marker of frequent offending, meaning that targeting resources toward adolescents who have early contact with the justice system may be one avenue by which later crime might be prevented (Ayers et al. 1999; Liu et al. 2011). Additionally, research typically finds that violent offenders have higher Lambda values than their nonviolent counterparts (Piquero 2000). As such, Piquero argued that prevention programs aimed specifically at reducing violent crimes might not be necessary. Instead, general prevention programs aimed at reducing offending frequency would concomitantly reduce violent crime as well since offenders with high frequencies tend to commit violent crime at a disproportionate level.

Although frequency research has become prominent since the original criminal careers report, there is still debate on which subset of offenders should be used in studying offender frequency. Studies typically draw their samples from one of two populations: current inmates or arrestees (Blumstein et al. 1986; Miles and Ludwig 2007). Inmate studies typically use retrospective interviews to determine the number and type(s) of crimes that inmates committed during a specified time period. Conversely, studies that sample arrestees (whether incarcerated or not) use official arrest reports to determine the number and type(s) of crimes for which an offender was arrested. Because frequency in the criminal career paradigm measures the number of crimes committed per year as opposed to the number of arrests per year, some researchers have argued that the most accurate way to estimate Lambda is by using retrospective inmate interviews where the inmates can describe crimes for which they were never arrested.

Both methods used to measure frequency rates have flaws. Self-reports suffer from problems with offenders accurately recalling past criminal acts and under- or overrepresentation of the number and type of crimes committed. However, the most serious potential problem with self-reports is selection bias. There may be a significant difference between offenders willing to share their criminal history and those unwilling to reveal their past criminal activities. In addition, self-report data on offense history is almost exclusively collected from incarcerated offenders. This poses problems when trying to generalize results to the entire offender population, most of whom never serve time in prison (Rhodes 1989).

Similarly, frequency values obtained from official arrest data may be biased due to the fact that arrests occur only for a small percentage of the number of crimes committed and/or the reality that part of an offender’s career may be spent in prison or jail. As such, the possibility of underestimating Lambda is a potential problem in using arrest data. One way to ensure more precise estimation of frequency rates is to combine the two methods by using arrest data to serve as a check on the accuracy of inmates’ self-reported criminal activities. Miles and Ludwig (2007) suggest that measuring Lambda is fraught with problems and, consequently, is of limited use in considering crime prevention policy.

As discussed above, offending frequency research has led to discussion of controversial selective incapacitation policies. These policies may be difficult to implement in practice;

however. Some have argued that by the time these high-rate offenders are identified and trigger habitual offender laws, they may have done considerable damage but be on the downside of their criminal careers (Gottfredson and Hirschi 1990; Miles and Ludwig 2007). Limitations in empirical research and the very real possibility of false positives (i.e., predicting an offender will be a high-rate offender in the future and s/he does not) ensure that the controversy surrounding selective incapacitation laws will remain.

Specialization/Versatility in Offending Careers

Although typological and taxonomic identification of offender groups has long been a subject of criminological research and justice practice, this aspect of offending careers has also grown as a topic of interest since the 1986 "Criminal Careers" report. The degree to which offenders specialize or generalize in turn has implications for understanding the etiology of criminal offending in terms of explaining and responding to crime. The main interest from a criminal career perspective is whether, and to what degree, criminals concentrate their offending in particular offense types. In other words, do certain offenders engage predominantly, or exclusively, in drug offenses, sex offenses, violent offenses, or property offenses? If the answer to this question is yes, offenders might be placed in distinct groups based on their focus on certain crime types. If not, it may be futile, or even counterproductive, to develop strategies targeting particular subgroups of offenders based on the types of crimes in which they are presumed to solely or predominantly engage (e.g., sex offender treatment, drug offender treatment). In other words, absent specialization, criminologists should focus primarily on general explanations of crime, such as Gottfredson and Hirschi's (1990) self-control theory, and it would be best if the system treated all offenders similarly no matter their instant offense – contingent on their frequency of offending. Conversely, the presence of specialization might suggest the use of more fine-grained explanations of offending that recognize the potential for qualitatively distinct groups of offenders. Policy-makers and practitioners might then look to make use of such grouping information in developing sentencing and treatment strategies.

Broadly speaking, two types of theories make implicit or explicit assumptions about the possibility of specialization. In contemporary criminology, general theories that make few distinctions between offenders in terms of the nature of their offending and the explanations derived from these theories typically suggest that offenders do not discriminate in the types of crimes that they commit. The General Theory of Crime suggests that offending is only one of a number of actions that might be undertaken in pursuit of short-term benefits at the expense of potential long-term consequences (Gottfredson and Hirschi 1990). Offenders are believed to have low levels of self-control which leads them to engage in a variety of criminal offenses without much regard for the form that they take. Consequently, there should be no recurring patterns of particular offense types. Offenders are unlikely to show specialization as this short-term, pleasure-seeking orientation will manifest itself differently depending on the situation (Britt 1994).

Conversely, taxonomic perspectives of criminal careers are open to the possibility that some distinctions among offenders exist – including those related to the types of crime committed. The criminal career paradigm defined by Blumstein and colleagues (1986) focuses on individual-level variation in offending across the life course. It also suggests that characteristics like specialization are worthy of theoretical explanation. This view holds that there are certain groups of offenders who warrant distinct explanations and may possibly require specialized treatment or intervention. For example, there may be a small group of serious, generalized offenders who engage in a continuous and varied track of antisocial behavior over time. At the same time, it is possible that certain individuals will engage in less serious forms of delinquency within the boundaries of adolescence before desisting around the time that they reach adulthood. This is the premise of Moffitt's (1993) taxonomy of antisocial behavior, which is an example of a theory that outlines subgroups of offenders that are distinct in terms of the nature and etiology of their offending. Moffitt's adolescent-limited offenders may specialize in relatively less serious antisocial behavior (e.g., status offenses, substance use), while life-course persistent offenders engage in a variety of offense types across a longer portion of their lives. The notion of offender types, however, has been criticized by those who espouse general views of crime and criminals. Even the use of the term "career," which implies some coherent pattern of behavior over time, has been the subject of questions in this regard (Gibbons 1988).

Probably the biggest controversy in this area is whether those interested in understanding criminal behavior should bother to study specialization at all. Gibbons (1988) suggests that the typological schemes offered by both criminologists and criminal justice practitioners tend to fall short when tested against empirical data and, in general, the literature on this characteristic of the criminal career tends to show only modest levels of offender specialization (Piquero et al. 2003). Nevertheless, the fact that some important theoretical propositions in criminology have differing views on specialization, coupled with the reality that offenders are often grouped by type in policy initiatives and treatment programs, has led to continued empirical research.

Some early research in this area found that information about the nature of an offender's previous criminal behavior was not a good predictor of later offense types (Wolfgang et al. 1972). Farrington and colleagues (1988) assessed the potential for specialization in a sample of juvenile court records and found an average coefficient value of 0.10, which falls well short of the 1.0 that would be expected if there was complete specialization in measured offending. They also compared the observed number of specialized offenders to what might be found by chance alone; roughly 20 % of their cases were characterized as specialists in that analysis. Blumstein et al. (1988a, b) found slightly higher levels of specialization in adult offenders, but these individuals represented just a minority of the population. Identifying limitations in the previous approaches, Britt (1996) found that offenders showed a much greater tendency to repeat the same offense than switch to another on their subsequent criminal act. Mazerolle et al. (2000) analyzed diversity index values, which were first used to account for similarity/dissimilarity among species, to assess specialization and found that their sample exhibited fairly diverse offending profiles. Later work with data from the Cambridge Study in Delinquent Development likewise identified little evidence of violent specialization (Piquero et al. 2007).

Despite these findings, some recent research suggests that answering questions about the degree of specialization in offending may require a slightly more nuanced view. In his ethnographic research, Shover (1996) found that, although it may be obscured by the variety that is apparent across an entire criminal career, offenders seem to make short-term changes in their offending indicative of specialized behavior. Francis et al. (2004) also identified some offender "types" in their study of career-length crime patterns. Osgood and Schreck (2007) report consistent evidence of specialization in violent offending using general population data on self-reported delinquency. Like Shover (1996), these studies suggest that offenders may alter their preferences over time, which adds up to a conclusion of versatility when looked at across a lengthy career but also suggests that it might be useful to consider these short-term, specialized patterns in more depth.

While these findings suggest some degree of specialization in offending, it is important that they are viewed in light of the larger body of research that has found that offenders tend to commit a variety of different criminal acts over their careers. Britt (1994) points out that even those studies that identify specialized offenders suggest that they comprise only a small portion of the overall population. Over time, research on specialization has drawn on a variety of methods, and, in some cases, it appears that the observed differences in results may be partly attributable to alternative approaches to measurement or analysis (Sullivan et al. 2009). Some studies that have identified support for the notion of specialization (e.g., Francis et al. 2004; Osgood and Schreck 2007) have used measurement and analytic techniques that may provide a better test for the presence and nature of specialization than those used previously. Also, more recent research on specialization has begun to better incorporate theory which helps in moving past the blunt question of whether there is any specialization while also illustrating the potential utility of understanding this aspect of criminal careers. For example, a recent study considered the question of specialization within the context of family violence using clear theoretical and preventive foundations (Piquero et al. 2006). Synthesizing the results of empirical studies, it is clear that most offenders will commit a variety of different types of crime over their careers, but there may be parts of the career where they tend to engage in one type or another and those patterns should be considered in theoretical development around criminal careers (Armstrong and Britt 2004).

Progression and Sequencing in Criminal Careers

Research on progression and sequencing in criminal careers is also concerned with identifying and explaining patterns across criminal careers (Piquero et al. 2003). While the question of how careers progress or sequences unfold can relate to specialization, it is primarily concerned with the possibility that there are trends in seriousness across the career. Although implicit in a number of approaches to dealing with offending behavior (e.g., diversion programs that seek to intervene with less serious offenders to prevent more serious offenses later on), few perspectives have formalized the notion of escalation in offending theoretically. Loeber and Hay (1997) offered a pathways model that did consider the possibility of escalation to violent behavior after a period of time engaging in less serious forms of delinquency. Their expectation is consistent with the idea of heterotypic continuity where less serious forms of aggression are believed to later give way to more serious violent behavior at a later stage of development. This theory also laid out several precursors to later escalation in aggression (e.g., temperament, cognitive factors). There are also some notions regarding escalation in certain subtypes of offenders. For example, sex offenders and men who engage in domestic violence are perceived to be at risk of escalation in the nature or frequency of their behavior over time.

Of the three criminal career dimensions considered in this entry, progression has received the least amount of attention (Armstrong and Britt 2004; Liu et al. 2011). Piquero et al. (2003) reviewed the evidence on the potential for escalation in offending and found the results to be decidedly mixed. Any observed patterns appear to be age-graded in that the seriousness of offending often trends upward for juveniles. Farrington (2003) asserted that most juveniles begin their criminal career by committing relatively minor offenses but tend to commit progressively more serious offenses until the age of 20. Conversely, the typical progression of offense seriousness for adult offenders appears to be stable or declining over time (Farrington 2003; Piquero et al. 2003).

Britt (1996) offered a new modeling strategy for identifying and testing specialization and escalation, conducting a series of analyses that were similar to those previously used to understand social mobility. Using data from a sample of Michigan offenders and a crime-specific breakdown, he identified statistical trends showing general support for a pattern of greater seriousness in offense type from one to the next. A somewhat similar recent study by Massoglia (2006) sought to understand offending patterns in the transition from adolescence to adulthood. Although this study only covered a portion of possible criminal careers, it provides insight into patterns of escalation and de-escalation. Massoglia identified groups that represented clear instances of escalation. Specifically, three subgroups of individuals in that study (26 %) had transition patterns that fit with notions of escalation (i.e., moved to a more serious offending class). This included cases that moved from normative deviance (e.g., truancy, disorderly conduct) to a predatory offending or illicit drug use class in early adulthood. Piquero and colleagues (2006) studied the question of escalation within the framework of spousal assault using data from victim interviews. This analysis focused on the severity of injury experienced by the victim in a second assault. They identified a high probability of escalation in these successive incidents of domestic violence in one site, suggesting that this pattern is prevalent. In other sites, there was a mix of escalation and de-escalation in the injuries experienced by victims. In general, this contradicted the notion that escalation is endemic to that particular form of criminal behavior.

Policy and practice around escalation are primarily focused on the fact that, if there is a progression from less to more serious offending, the earlier behavior can be used as a marker for intervention and the more serious and harmful behavior on the part of that individual may be prevented. For example, using conviction data from the England and Wales Offenders Index, Liu et al. (2011) examined the seriousness of offenses committed by over 4,000 offenders to determine whether escalation and/or de-escalation of offense seriousness is present over the criminal career. They concluded that offenders with few convictions (1–5) showed de-escalation over time, whereas offenders with more than five convictions tended to escalate in offense seriousness. As such, the authors argued that intervention policies should be aimed at persistent offenders because they are more likely to eventually escalate into committing serious, violent crimes, a view that is shared by other researchers (Ayers et al. 1999).

Conclusions and Future Research

In general, criminologists now know considerably more about the criminal career than was known 25 years ago when research and discussion in this area really began to take off. At that point, the idea of a criminal career was one that made sense intuitively, and Blumstein and colleagues (1986) summarized a good deal of research to that point that supported the usefulness of the perspective. Still, at that time, there were questions in terms of its fundamental connection to observed data and its potential implications for a general understanding of crime and related policy and practice. For example, there are those that asserted that longitudinal research of the type described above was too expensive and time-consuming and would prove to be of limited use to theory and practice (Gottfredson and Hirschi 1986).

There are, however, some important conclusions that can be reached upon reflection a couple of decades later. It is clear that most crime is committed by a small pool of high-frequency offenders and a number of correlates of those high-frequency careers have been identified (e.g., early age of onset, career length, and frequent drug use). In terms of specialization, it is safe to say that most serious offenders will engage in a variety of different crime types over the course of a lengthy criminal career, but it would be an overstatement to say that there is no group of specialized offenders – particularly when disaggregating patterns of behavior that can extend years or decades. Similarly, the research on progression in offending that includes a pattern of escalation is fairly mixed. Some evidence has been found to support the idea of escalation, but it seems to be contingent on the context of the study.

So it is the case that criminologists currently know more about the three dimensions of the criminal career considered here. At the same time, there appears to be some limitations in knowledge as well as some question around the utility of pursuing these research questions given the methodological challenges inherent in their study. For example, although prevention may be more desirable than longer-term incapacitation as a method of dealing with high-frequency offenders, the question of early identification continues to loom large. Resolving the question of false positives and false negatives is a difficult one as predicting behavior is a challenging endeavor – even with high-quality information on risk factors. Consequently, the identification and response to high Lambda offenders continues to be a difficult empirical and practical question. Additionally, while advances in measurement and analytic techniques have helped in considering the question of offender specialization and progression more thoroughly, efforts to consider patterns in criminal offending in a clearer theoretical framework are also emerging with increasing regularity. It will be important in the future to determine whether the identification of a small group of specialists and those with clear patterns of offending escalation have important implications for theory, policy, and practice or are mainly just products of intuitively appealing notions about the nature of criminal behavior. This will be more helpful in terms of understanding criminal careers than a simple "yes/no" answer to the question of whether specialization or escalation in offending exist. In general, the next iteration of research on these topics should continue to explicitly embed an understanding of the dimensions of the active criminal career in the pursuit of knowledge about the processes by which criminal behavior is sustained over time while also attending to any implications for practice that might emerge.

Related Entries

- ▶ Age-Crime Curve
- ▶ Criminal Careers
- ▶ Career Criminals and Criminological Theory
- ▶ Criminal Careers and Public Policy Responses
- ▶ Onset of Offending

Recommended Reading and References

Armstrong TA, Brit CL (2004) The effect of offender characteristics on offense specialization and escalation. Justice Q 21(4):843–876

Ayers CD, Williams JH, Hawkins JD, Peterson PL, Catalano RF, Abbott RD (1999) Assessing correlates of onset escalation de-escalation and desistence of delinquent behavior. J Quant Criminol 15(3):277–306

Blumstein A, Cohen J, Roth JA, Visher CA (eds) (1986) Criminal careers and "career criminals". National Academy Press, Washington, DC

Blumstein A, Cohen J, Das S, Moitra SD (1988a) Specialization and seriousness during adult criminal careers. J Quant Criminol 4:303–345

Blumstein A, Cohen J, Farrington DP (1988b) Criminal career research: its value for criminology. Criminology 26(1):1–35

Britt CL (1994) Versatility. In: Hirschi T, Gottfredson M (eds) The generality of deviance. Transaction, New Brunswick

Britt CL (1996) The measurement of specialization and escalation in the criminal career: an alternative modeling strategy. J Quant Criminol 12:193–207

Chaiken JM, Chaiken M (1982) Varieties of criminal behavior. Rand Corporation, Santa Monica

Farrington DP (2003) Developmental and life-course criminology: key theoretical and empirical issues—the 2002 Sutherland award address. Criminology 41(2):221–255

Farrington DP, Snyder H, Finnegan T (1988) Specialization in juvenile court careers. Criminology 26:461–485

Francis B, Soothill K, Fligelstone R (2004) Identifying patterns and pathways of offending behaviour: a new approach to typologies of crime. Eur J Criminol 1:47–87

Gibbons DC (1988) Some critical comments on criminal types and criminal careers. Crim Just Behav 15(1):8–23

Gottfredson M, Hirschi T (1986) The true value of lambda would appear to be zero: an essay on career criminals, criminal careers selective incapacitation cohort studies and related topics. Criminology 24(2):213–234

Gottfredson M, Hirschi T (1990) The general theory of crime. Stanford University Press, Palo Alto

Liu J, Francis B, Soothill K (2011) A longitudinal study of escalation in crime seriousness. J Quant Criminol 27(2):175–196

Loeber R, Hay D (1997) Key issues in the development of aggression and violence from childhood to early adulthood. Annu Rev Psychol 48:371–410

Massoglia M (2006) Desistance or displacement? The changing patterns of criminal offending from adolescence to adulthood. J Quant Criminol 22:215–239

Mazerolle P, Brame R, Paternoster R, Piquero A, Dean C (2000) Onset age persistence and offending versatility: comparisons across gender. Criminology 38:1143–1172

Miles TJ, Ludwig J (2007) The silence of the lambdas: deterring incapacitation research. J Quant Criminol 23:287–301

Moffitt TE (1993) Adolescent-limited and life-course-persistent antisocial behavior: a developmental taxonomy. Psychol Rev 100:674–701

Osgood DW, Schreck CJ (2007) A new method for studying the extent stability and predictors of individual specialization in violence. Criminology 45:273–312

Piquero AR (2000) Frequency specialization and violence in offending careers. J Res Crime Delinq 37(4):392–418

Piquero AR, Farrington DP, Blumstein A (2003) The criminal career paradigm. In: Tonry M (ed) Crime and justice: a review of research, vol 30. University of Chicago Press, Chicago

Piquero AR, Brame R, Fagan J, Moffitt TE (2006) Assessing the offending activity of criminal domestic violence suspects: offense specialization escalation and de-escalation evidence from the spouse assault replication program. Public Health Rep 121:409–418

Piquero AR, Farrington DP, Blumstein A (2007) Key issues in criminal career research: new analysis of the Cambridge study in delinquent development. Cambridge University Press, Cambridge

Rhodes W (1989) The criminal career: estimates of the duration and frequency of crime commission. J Quant Criminol 5(1):3–32

Shover N (1996) Great pretenders: pursuits and careers of persistent thieves. Westview, Boulder

Smith DA, Visher CA, Jarjoura GR, O'Leary V (1991) Dimensions of delinquency: exploring the correlates of participation frequency and persistence of delinquent behavior. J Res Crime Delinq 28(1):6–32

Sullivan CJ, McGloin JM, Ray JV, Caudy MS (2009) Detecting specialization in offending: comparing analytic approaches. J Quant Criminol 25:419–441

Wolfgang M, Figlio R, Sellin T (1972) Delinquency in a birth cohort. University of Chicago Press, Chicago

Crime Sprees

▶ Prediction and Crime Clusters

Crime Trends

▶ Youth Homicide in the United States

Crimes Against Animal Life

Ragnhild Sollund
Department of Criminology and Sociology of Law, Faculty of Law, University of Oslo,
Oslo, Norway

Synonyms

Poaching; Wildlife trade; Wildlife trafficking

Overview

This entry addresses the trade and trafficking in animals and discusses this multivariate crime through the lens of animal abuse and speciesism, as these phenomena are discussed within green, eco-global criminology. The "wildlife" trade is one of the fastest increasing illegal transnational trades and a threat to a great number of species, as animals are used, for example, in Asian traditional medicine, as "pets" or as part of collections, for meat, for trophies, or for their fur. The trafficking causes tremendous suffering to the individuals which are abducted from their habitats to be trafficked, and great numbers die before they reach their destination. Offenders are both local villagers and large organized groups which either concentrate on this crime alone or combine it with other organized crime, for example, the drug trade. The "wildlife" trade is controlled by means of the CITES convention, which sees "wildlife" as natural resources, the trade in which must be regulated in order to preserve species from extinction. Through this convention, animals are not accorded rights, only value as "specimen" representing a species. This entry concludes by encouraging criminological research which includes a non-speciesist perspective, taking into consideration the harms of the animal trade, whether legal or illegal.

Key Issues

As the field of "green" (Lynch 1990; South 1998) or "eco-global" (White 2008, 2011) criminology has developed and expanded since the 1990s, so has the number of articles and publications related to so-called "wildlife" crime, both criminological, from other academic fields and from NGOs (e.g., Zimmerman 2003; Warchol et al. 2003; Lemieux and Clarke 2009; Westerhuis in press; Pires and Clarke 2011a, b; Traffic 2008; IWAF 2008; Gonzales 2003; Wyatt 2009, 2011; Sollund 2011; Wellsmith 2011). A reason is the growth of the trade, placing it variably as the second or third largest illegal trade worldwide, competing with the illegal drugs and arms trades and with human trafficking. As with other harm encompassed by green criminology, this phenomenon is far from one dimensional. It is local, regional, and transnational, organized or part of individual and culturally rooted exploitation, reflecting its broad spectrum of actors, their varied motivations, the causes of the trade and also its impact, for example, the massive suffering of the victims and species loss. This entry will discuss some typical findings in the literature describing the field and see them through the lens of animal abuse and speciesism.

First, some linguistic clarifications are required. The term "wildlife" is anthropocentric and antagonistic and implies that those animals who are not living under human control and domestication in fundamental ways are different from and dangerous to humans and that (often) they must be controlled for human protection. It also implies the "othering" of these animal(s) and species as "nature" in contrast to human "culture."

Generally, the word "poaching" is used to describe the act of illegally killing or abducting animals from their habitats. This implies that other animals are regarded as the property of humans, regardless of the negative impact this has on the well-being of each individual victim, such as pain and death. To be closer to the true character of killing and removing animals in and from their habitats, the terms theriocide (Beirne 1999, 2009) and abduction (Sollund 2011) are preferred. This emphasizes that nonhuman animals have equal interests to humans in living a full life free from pain and abuse in their natural, local domain (Singer 1975; Regan 1983). The word "animal" is a rejection of the fact that humans are also animals. Authors have solved this difficulty in different ways, for example, by applying the terms "nonhuman animals" and "animals other than humans." These concepts can still be criticized for giving humans priority over other species, as humans are the yardstick other species are measured against, and thereby categorizing other animals as the same, despite the obvious interspecies differences. In the absence of better alternatives and for consistency of style, the term "animal," when referring to animals who are not human, will still be applied.

Animal trafficking is a splendid case to illustrate why a broader perspective than that represented by traditional, mainstream criminology is necessary in addressing the transnational crime of animal trafficking. Green or eco-global criminology represents a perspective encompassing a variety of counter-hegemonic approaches that are concerned not only with the health of the environment but also with the animals living in environments, acknowledging their inherent value and that animals and the environment are mutually interdependent (Beirne 2011: 353).

The introduction of CITES (the Convention on International Trade in Endangered Species of Wild Fauna and Flora) in 1973 (in force from 1975) established a divide in the perception of the animal trade and trafficking. It introduced a move from the legally free exploitation of "wild" nonhuman species to a requirement for regulation, recognizing the need to preserve species from extinction due to overexploitation from hunting and "poaching" (Hutton and Dickson 2000; Reeve 2002). Previously, the wildlife trade phenomenon, much as other "green crimes," for example, those related to pollution, waste, and land degradation (Beirne and South 2007; White 2008, 2011), was neglected by criminologists, probably because the "wildlife" trade like much other harm was not a crime and the trade and trafficking in animals has a long history in which animals have been regarded as resources that can rightfully be exploited – "harvested" like a crop – by humans and used for a wide variety of purposes, for example, refined as cultural apparel, in traditional medicine, for adornment, clothing, and food.

What determines whether trade and trafficking of a nonhuman individual is legal or criminalized is the degree of protection attributed to its species and the growing scarcity of individuals belonging to a species, defined by which of the three CITES appendices the species is listed in, if any (see below). The CITES convention seeks not to protect individuals from suffering, only as representatives of a species.

Trafficking in "wildlife" is partly a consequence of speciesism and entails massive animal abuse, death, and species extinction. The three phenomena – "wildlife" trade, animal abuse, and speciesism – are thus interrelated and inseparable, as animals cannot be traded without abuse and pain, and pain and abuse would not be inflicted upon them were it not for the ideological basis of speciesism permitting and legitimating it. Some space will therefore be dedicated to elaborate upon the phenomena of animal abuse and speciesism and how these are understood in a green criminology framework.

Background Description: Animal Abuse, Speciesism, and Animal Trafficking

Beirne (1999) should be credited with seriously introducing the topic of animal abuse to criminology in an article published in 1999. He argued that criminology had been speciesist and anthropocentric and that animal abuse should be included as a legitimate field of criminological research because animal abuse entails pain and suffering, a violation of rights, and is one of several forms of oppression, like sexism and racism (Beirne 1999: 140). Beirne's call for a nonspeciesist criminology is by no means fully responded to in mainstream criminology, but with "green criminology" (Beirne and South 2007; South and Beirne 2006) or "eco-global criminology" (White 2008, 2011) and the introduction of terms such as "speciesism," "ecological justice," and "species justice," it appears to be a move toward a criminology that researches animal abuse, not because it may be a symptom of interhuman violence but because animals increasingly are regarded as victims when abused (Beirne 2008).

This includes more than philosophical and ideological arguments for why animals should have rights (not to be abused) and proceeds to include empirical studies where animal abuse is studied in its own right.

Beirne builds on Robert Agnew's (1998) definition of animal abuse as

> any act that contributes to the pain, suffering, or death of an animal or that otherwise threatens its welfare. Animal abuse may be physical, psychological, or emotional; may involve active maltreatment or passive neglect or omission; and may be direct or indirect, intentional or unintentional. (1999:121)

With this broad understanding, there is hardly any area of society where animals can be found that would be exempt from research, yet to narrow it down, one could start by limiting the studies to *human*-inflicted pain. Undoubtedly many animals suffer as prey, and, for example, a mouse being victim of a cat's playful torture will likely suffer both physically and psychologically. The suffering caused by "cruelty play" by one animal against another is beyond the scope of this entry, though interesting and meriting attention. Nonhuman animals and humans share an interest in killing for food. However, in the case of the 30 % of nonhuman animals that are carnivores, this is because of a *need*, while humans may hunt and kill for a number of other reasons, for example, for "sport" and status. It is highly debated whether humans need meat and rejected by many; nor do humans need to fight to position themselves as leaders of a pack or flock. Therefore, animal-inflicted suffering and death and human-inflicted suffering and death are differently motivated and not comparable. This discussion is therefore pursued no further here.

Most of the harm inflicted on animals is *legal*, yet as equally harmful and fatal as the very few *criminalized* acts directed against animals. Areas of animal abuse that should be prioritized in research are those where animals are abused by humans through a deliberate act that the human should understand would entail discomfort, suffering, or pain for the animal, or through omission, regardless of the motivations that could be sought, excused, or neutralized by appeal to higher loyalties, for example, in the case of vivisection. One could/should, for example, research all kinds of *systemic* animal abuse, as in the meat industry, research laboratories, the fur industry, as well as unsystemic, domestic animal abuse. The aim would be to start to look beyond human-centered understandings and interpretations of pain, in which only pain attributed to animals that does not counter human interests is defined as abuse by animal welfare legislation.

Thus green eco-global criminology entails an expansion of the traditional foci of criminology, from breaches of the law (whether from a control, offender, or victim perspective) to include *harm* as a central category (Beirne and South 2007; White 2008, 2011). In this, the analysis of abuse must be broadened to include not only those animals who are ascribed a social status as companions to humans (and consequently whose status is thus "elevated" to imply that they may also be ascribed victim status) but all sentient beings, regardless of their actual or potential, direct, or indirect value for humans. In such a perspective, the topic mainly addressed in this entry – the legal and illegal abduction, trade, trafficking and theriocide of animals (theriocide is a term introduced by Piers Beirne, as a parallel to homicide, referring to the acts in which nonhuman animals are killed by humans) (Beirne 2009: 17), not only that which is criminalized, finds its place.

It is estimated that nine and a half billion animals are killed annually in the USA alone, of which nine billion are killed for food, while twenty to one hundred million animals are killed in experiments (Agnew 1998: 180). Animals destined to become "meat" or "research tools" live under conditions that are equivalent to torture, confined in narrow cages, deprived of offspring, seldom allowed to ever move freely or feel the sun, and exposed to painful and fatal experiments. These are conditions which would never have been accepted if applied to humans and under such circumstances would be criminalized. Agnew developed a theory that could grasp not only the illegal abuse but also the vastly dispersed and culturally accepted legal animal abuse. In his model (1998: 182) are several factors that are important for a person's involvement in abusive acts toward animals: (1) his/her *social position* – gender, "race" [sic] education, occupation, and place of residence, urban/rural; (2) *individual traits*, most importantly empathy and self-control, socialization models, nature of animals; and (3) *ignorance re abusive consequences of behavior*, beliefs justifying abuse and perceived benefits/costs of abuse. The last, ignorance, may be the most important factor, as it accounts for much of the systemic abuse taking place in industrial complexes, on which, for example, the consumption of meat and production of medicine and cosmetics rely. However, this ignorance may also be the consequence of a *desired* ignorance and denial, as confronting oneself with the harmful effects of one's consumption may be harder than to act upon the knowledge of the harm this consumption entails, and end it (Sollund 2008).

Neutralization techniques are important in minimizing and justifying abusive acts – as appeals to higher loyalties, for example, in the case of vivisection (Agnew 1998; Regan 1983), and the condemnation of the condemners – for example, attacking animal rights activists as "extremists" and "terrorists." These techniques of neutralization, however, would not make sense, if it were not for the fact that they are underpinned by a governing ideology and doxic practice (Bourdieu 1995) – speciesism.

Speciesism

The term speciesism was coined by Richard Ryder in 1970 and has since been expanded and defined in various ways. Peter Singer made the term *speciesism* famous with his book *Animal Liberation* (1975), which also ignited the animal liberation movement. Speciesism can be understood as a prejudice or biased attitude favoring the interests of the members of one's own species against those of members of other species. As with sexism and racism, speciesism rests on domination and subordination (Nibert 2002), rooted in patriarchal ideology, based on the fact that other species are not human and paralleling the way that women are seen as different and inferior to the white male (Donovan and Adams 1996; Noske 1997).

The most important issue regarding a prejudice is not necessarily the prejudice itself but whether the prejudice is turned into practice. As an example, racists may not transform racist prejudice into racist acts if they find themselves in a social environment where racist acts are socially unacceptable. In that case, their prejudices, though racist, may not have much significance because informal and/or formal social control will usually prohibit them from coming out in the open. In the case of speciesism, ideology legitimates and upholds prejudice, thus more easily converting it into abusive acts and exploitative practice Nibert (2002). Speciesism (like racism, sexism, and classism) is a set of widely and socially shared beliefs that result from and support oppressive social arrangements. These include the objectification of animals (e.g., as "meat" or "fur"), thereby obscuring animals as beings with individual selves, desires, and needs as well as characteristics based on a sex. Their exploitation is facilitated by the denial of any individuality.

When considering the treatment of animals regarded as suitable for consumption as "meat," it is interesting to question how this is compatible with the love and care directed toward domestic pet animals. However, as already noted, there is a difference in the perceptions of harm inflicted on humans and animals related to their status: *[. . .] some harms are defined as criminal, others as abusive, but not criminal, and still others as neither criminal nor abusive* (Beirne 2009: 200). This state of affairs is generally unquestioned and hence facilitates speciesism in a way that means it operates even without conscious prejudice. Critics of this status quo would argue that *not* to take a stand is also to *take* a stand, by refusing to reflect upon the consequences of one's actions. Such refusal may be caused by emotional or other reactions to consideration of the harms done to animals and for that reason be connected to various forms of denial, but it is also facilitated by the meat industry which distributes advertisements of "happy cows in pasture" far remote from the reality of intensive and factory farming. All of this contributes to the *cultural denial* in which whole societies partake in collective denial of atrocities (Cohen 2001: 11).

Speciesism can thus be defined either as ideology or practice – or as the two in tandem– supported by discourse that polices and maintains the human-nonhuman boundary. There are, for example, numerous derogatory terms referring to animals, such as "stupid as a hen," "stupid cow," "sly as a wolf," "fat as a pig," or "old dirty pig" when referring to an old male human with an appetite for young girls. In fact, such terms are so integrated into our language(s) that it takes a very conscious mind to avoid them. Terms are often sexist, thus combining prejudice against animals and female humans, like "bitch." Derogatory terms serve not only to denigrate the individual they are directed to but also to prepare the ground for use, abuse, and objectification of this individual and others of the same kind, and thus its legitimacy (Cohen 2001; Sollund 2008). When animals, reptiles, and birds are referred to as "specimens" in the CITES convention, this is another way of objectifying them and depriving them of individuality and victim status, through which their only value is as part of a specie.

State of the Art: The Trafficking and Trade in Animals

The growing number of criminological articles about the abduction of animals from their habitats usually refers to these acts as *poaching* (Lemieux and Clarke 2009; Pires and Clarke 2011a, b; Warchol et al. 2003). This term includes both abduction of "wild" living animals from their habitats and also killing them on the spot – theriocide. They are subsequently trafficked and traded, and the purpose of the live abduction or killing determines whether their entire bodies are trafficked or only part of them. These actions are encompassed under the terms *wildlife trafficking* (Wyatt 2011; Warchol et al. 2003) or/and *wildlife trade* (in print; Zimmerman 2003; Westerhuis' in print; Wyatt 2009), or the narrower *wildlife crime* (Wellsmith 2011; Zimmerman 2003), emphasizing that the trade in animals represents a *crime*. This indicates that the abduction, trade, and trafficking in animals are *criminal* acts; however, this is not always the case, as a significant part of the animal trade is indeed taking place *in accordance* with regulations. Actually, the *legal* "wildlife trade," excluding "fisheries," is variously estimated to be worth from 5 to 50 billion US dollars annually (Reeve 2002: 10) and to 160 billion US dollars per year (Duffy in White 2011: 55). Whether the trafficking and trade in an animal is legal or illegal will depend on how threatened his/her species is.

The CITES convention (also referred to as the Washington Convention) is an agreement that was initiated in 1973, now signed by 175 states (parties to the convention) which through the convention are obliged to pass legislation meeting the requirements of the convention, and consequently also to punish those breaching this legislation. The aim of CITES is to *regulate* trade, not to *prevent* trade, so that the trade in animals can be *sustainable* and species conserved. As quoted from the CITES webpage, *Its aim is to ensure that international trade in specimens (*sic*) of wild animals and plants does not threaten their survival*. CITES can be criticized for legitimating trade and trafficking in animals and for prolonging a situation encouraging animal abuse and species decline through its anthropocentric approach to other species (Sollund 2011; Hutton and Dickson 2000). In the CITES convention, species are accorded protection depending on the threat of the trade, which is defined in the appendices. Those listed on Appendix I may be traded only in exceptional cases (as scientific research with the prospect of the significant harm this implies), because they are close to extinction. Appendix I species can only be traded if accompanied by an export and import certificate, or a reexport certificate issued for individuals who have previously been imported (Wijnsiekers 2001, here in Wyatt 2011). Appendix II lists those species which are not yet necessarily threatened with extinction but that may become so unless trade is closely controlled. It also includes so-called look-alike species, that is, species that look like those listed for conservation. International trade in Appendix II species may be authorized by the granting of an export permit or reexport certificate. No import permit is necessary for these species under CITES. On Appendix III are those species which there at the request of one of the parties to the convention, which may imply that the species is not necessarily threatened worldwide but in this state, and this state needs the cooperation of other parties to prevent the species from becoming extinct. Through the conferences of the parties, species may be suggested for inclusion on the appendices and also moved between them, for example, when a species is moved from Appendix I to Appendix II. There are also recommendations on how large the quota for export of individuals can be and directions regarding the purpose of the abduction, theriocide, trade, or trafficking. As an example, at the conference of 2007, it was suggested that the *Panthera pardus* (leopard) in Uganda be transferred from Appendix I to Appendix II with the following annotation:

1. For the exclusive purpose of sport hunting for trophies and skins for personal use, to be exported as personal effects
2. With an annual export quota of 50 leopards for the whole country

As this example shows, CITES does not regard the capture and killing of animals as a crime. However, from the perspectives of ecological justice and species justice (White 2008), though not breaching laws, these acts justify the use of the term *crime* (ibid), thereby emphasizing the serious nature of such actions. For the leopards involved, it makes no difference that killing them is not a crime. Whether legal or illegal, both entail tremendous harm for the victims and for the environment because removing species from their habitats ruins ecosystems (e.g., Herbig 2010).

As mentioned in the introduction, there exists a body of literature on the animal trade with contributions from the natural sciences, such as biology and ornithology. These form a necessary basis for criminological discussion together with the work from Traffic and UWAF (Pires and Clarke 2011a, b; Sollund 2011; White 2011) although in the following the focus is on the main features and findings of a selection of studies from criminology.

From Local Abductors to Transnational Organized Crime

A principal finding in the literature reporting research on animal trafficking is its multidimensional character. The markets are multiple; animals are trafficked for a large number of reasons, as rare objects for collectors, as trophies, for medicine (e.g., gall blathers and rhino horn), for experiments in western laboratories, for the “pet” market, for collectors, for their fur, for meat, for ornament, and for falconry (Warchol et al. 2003; Wyatt 2009, 2011; Pires and Clarke 2011a, b; Herbig 2010; Sollund 2011; Reeve 2002; Hutton and Dickson 2000; Reeve 2002).

Illegal animal trafficking is categorized as follows: first, local farmers who abduct and sell individuals to supplement their income; second, larger, mafia-like groups who buy from local peasantry and sell the animals at great profit; and third, major international smuggling rings which are often involved in other illegal trades (Zimmerman 2003: 1668). These rings tend to use violence, have considerable resources, are aware of smuggling routes, and consequently pose the largest threat to the regulation of the illegal "wildlife" trade. For example, there are three ways in which the drug trade may be linked with the "wildlife" trade: by parallel trafficking along shared smuggling routes, by the use of "legal" "wildlife" shipments to conceal drugs, and by the use of "wildlife" products to barter for drugs, and the exchange of drugs for "wildlife" to launder the proceeds of drug trafficking (Reeve 2002: 12–13; South and Wyatt 2011). Animals are used as receptacles for illegal drugs, whether they are alive or entailing their death, and planeloads of birds are exchanged for drugs; in Brazil the police estimate that perhaps 40 % of illegal drug shipments are combined with "wildlife" traffic (Reeve 2002: 13). With established distribution chains and buyers, such organized trade has far greater impact on the numbers of individuals traded than caused by unorganized individual abductors. Such groups are vigorous, for example, in Africa and Asia, in the illegal ivory trade (Lemieux and Clarke 2009), and in Colombia where the same routes can be used to smuggle primates as drugs (Zimmerman 2003: 1672). Estimates show that between 200,000 and 600,000 primates are illegally exported from Colombia every year, destined for laboratories and research centers (Zimmerman 2003).

In addition to the harm done to the individuals that are victims of trafficking, this trade can also lead to infection of domestic species with contagious diseases (Herbig 2010: 126–128). Whether on a small or large scale, the motivation is economic though the gain will vary with the position of the trafficker in the distribution chain. For example, poor indigenous villagers acting as abductors can earn $5–25 for each salmon-crested cockatoo, but by the time the birds reach the infamous bird markets in Indonesia, they will each cost up to $150 dollars (Metz 2007; Low 2003; Pires and Clarke 2011). A scarlet macaw or hyacinth parrot can retail for $10–15,000 in the USA or Europe (Speart 1993 in Warchol et al. 2003: 6); rare parrots can be sold for even 100,000 $ per pair (Reeve 2002: 11). The value of illegal ivory has grown significantly from US$200 per kilogram in 2004 to up to US$1,700 per kilogram in 2009 (Grossman 2009). The hunting of "bushmeat" now represents a considerable threat to biodiversity, especially in Central Africa, where the existing impact of hunting for sustenance by indigenous groups is being added to by commercial interests which abduct and export animals from the forests to markets and restaurants in Europe. Elsewhere, Australian native species are on the tables of restaurants in Europe and the USA (Westerhuis in print).

There is a point to making a distinction between legal and illegal animal trafficking for whether an act is a breach of law or not may be central to discussion of how it can be combated, for example, by means of law enforcement strategies and by studying the market and exploring how CITES actually works (Warchol et al. 2003; Pires and Clarke 2010; 2011; Zimmerman 2003; Lemieux and Clarke 2009; Wellsmith 2010, 2011). However, in the case of animal trafficking, the legal/illegal distinction is insufficient, and it is necessary to build an approach from a more holistic, multidisciplinary basis in order to expand understanding of the field, and this entails embracing the concept of harm rather than (just) crime (Beirne and South 2007). It is also important that the underpinning rationale of cultural practices and ideology – speciesism – is also explored.

Although they may be investigating animal trafficking, the perspective adopted by some researchers seems anthropocentric, and their terminology implies that the animal trafficking is a crime more directed to *humanity* than to the animal victims themselves (see White 2011: 68–69). For example, by uncritically and dryly referring to abduction and theriocide as "poaching" (Lemieux and Clarke 2009; Pires and Clarke 2010, 2011), the suffering of the victim is not central, but instead the focus is on someone appropriating a "common good – a common property." The term "wildlife" creates a social distance between local animals and humans and domesticated animals. Investigating animal trafficking appears important, not because of the pain and suffering it involves but *because* it is a "crime," and thus fit for criminological analyses (Pires and Clarke 2011). The reason why this is an insufficient approach will prove evident in the following when some of the consequences of animal trafficking are sketched out.

The Suffering Caused by Animal Trafficking

Trade and trafficking in combination with habitat loss and habitat degradation constitute the largest threat to a great number of species (Du Plessis 2000). It is maintained that the earth now faces mass extinction of species on a scale potentially akin to the disappearance of the dinosaurs, which will result in the loss of 50 % of the earth's species by the end of the century if unhindered. UNEP (United Nations Environment Programme) states that 24 % of all mammals (1,130 species) and 12 % of birds (1,183 species) are globally threatened (Reeve 2002: 7, 8). Of the 90 parrot species at risk of extinction, 81 % are threatened by habitat loss, while 43 % by the trade in live birds for the "pet" trade (Du Plessis 2000: 21). The growth in the trafficking industry witnessed during the past two decades cannot be disconnected from the expansion of the World Wide Web and its significant role as a marketplace (IWAF 2008).

These numbers, though great, say nothing about the suffering involved for the victims. For example, one report describes how antitank weapons are used to blow the heads off elephants in the hunt for "bushmeat" and how mammals, birds, and reptiles which are seized by customs are found dead, malnourished, or starved, with broken limbs and bones, with lesions and frozen to death (Warchol et al. 2003: 8). Regarding parrots abducted for the "pet market," a report about the parrot trade in Mexico revealed that 65,000–78,500 Mexican parrots are captured each year and [that] the overall mortality rate for trapped parrots exceeds 75 % before reaching a purchaser. This means that 50,000–60,000 parrots per year die in the trade (Weston and Memon 2009). Up to 75–90 % of illegally trafficked parrots die during transportation (Gonzales 2003). A study of the trade of parrots listed on the CITES appendices estimated that 1.2 million birds were exported between 1991 and 1996, and this is likely to be a gross underestimation of the real figures (Wright et al. 2001). Nearly a third of the 145 parrot species in the Neotropics (Mexico and Central and South America) are threatened, making them among the most endangered groups of birds worldwide (Reeve 2002). In the case of reptiles, studies show a mortality rate of 90 % during the first year in captivity (Herbig 2010: 120).

Future Directions: How to Approach the Crime of Animal Trafficking?

While entailing loss in biodiversity, destruction of ecosystems, and generally a far poorer world, the treatment of the animals traded for profit – whether on a small or big scale – entails unimaginable suffering and abuse. Or one can *try* to imagine what it is like to be abducted from the nest and the flock to have one's eyes sewn closed (see Wyatt 2009, 2011), the arms (wings) strapped to the body, the mouth (beak) taped, and to be stuffed in a bottle just big enough to fit the body in this strapped condition, and then stuffed into another box or a suitcase with hundreds of other victims for days without food or water. If one survives the journey, the rest of life will be spent in a cage or on a leash. This important dimension of animal trafficking is rarely discussed. However, there are exceptions, as in Westerhuis' examination of Australian legislation and discussion of the moral question of the "wildlife" trade. Westerhuis concludes [. . .] the anthropocentric nature of environmental ethics in contemporary law and regulation allows the killing and trading of native wildlife, and does not recognize the sentience of wild animals. When tested, Preston's hypothesis that a new ecological ethic would replace the traditional anthropomorphic ethic cannot be supported (Westerhuis in print). By carefully examining the criminological literature addressing animal trafficking, sympathy for the victims at times shines through – animal abuse and suffering is mentioned (Warchol et al. 2003; Wyatt 2011: 116), and may also be a motivation for the research (Wellsmith 2011). Generally, however, empathy does not seem to be an important reason for studies in this field. This may be related to speciesism. Those researchers who are not themselves speciesist and are therefore interested in the field but do not address the abuse may fear of collegial attitudes and repercussions, including getting funded or published. Instead, to enhance their case and broaden its perceived relevance, they may center discussion around the harm done to *humans* or to the *environment* through the loss of biodiversity. Or it may be agreed in principle that this is another environmental crime, like pollution and the trade in toxic waste, which is therefore interesting for a criminologist. Or trafficking in animals may be studied because it is an expanding global transnational *crime*, which should and could be studied by criminologists because studying crime, to simplify, is what criminologists do, and this is an under explored field. However, animal abuse is increasingly becoming a legitimate field of study in criminology and as it does one can ask whether it is possible to research the animal trade without discussing the abuse it entails? To draw a parallel, does it make sense, for example, to research genocide, as in the case of Rwanda, without taking into account death, suffering, and power relations?

No matter what the motivation for research, further study is required in order to determine what measures must be taken to combat animal-related crime. It is clear though that different actors and differing motivations call for refined and differentiated preventative measures, reactions, and sanctions. While organized criminal groups may be deterred by long prison sentences and huge fines, in the case of local peasants, other measures such as teaching them the value of their victims and encouraging eco-tourism which could provide employment as guides are more likely to succeed in combating the animal trade. It has been suggested that in the case of the illegal trade in raptors in Russia that breeding programs should be encouraged to reduce the suffering of the birds who are victims of illegal trafficking and to preserve species from extinction (Wyatt 2011). However, such legal interventions and breeding programs can bring their own risks as exemplified by the CITES convention which permits rather than prohibits harmful trade in animals. When such trades depend on certificates and export permits, these can lead to criminal enterprise producing forgeries (Zimmerman 2003; Weston and Memon 2009; Wyatt 2009), and it is up to customs to distinguish between "legal" and "illegal" species which is not always a dependable (or honest) process. By introducing a total ban on trade in "wild" animals, customs and law enforcement agents would not have to distinguish between forged and authentic certificates and permits and between look-alike and protected species. A total ban would, however, encourage collectors and traffickers as the price of rare species would increase, as occurs when species are moved from Appendix II to Appendix I (Reeve 2002).

Nonetheless, by legitimating animal trafficking (e.g., as through CITES), speciesist ideology accepts and promotes animals as "natural resources" and as a "crop" to be harvested by humans (Sollund 2011). If research into animal trafficking is undertaken without paying attention to the victims of the trade and its ideological basis – speciesism – the trade will prevail. To increase awareness of the inherent value of other animals is to do them justice and acknowledge their place in the world and their intrinsic right to not suffer abuse, whether this be one-on-one harm, institutionalized harm, or harm arising from human action (White 2011: 23).

Related Entries

▶ Green Criminology
▶ Organized Crime and the Environment

Recommended Reading and References

Agnew R (1998) The causes of animal abuse: a social psychological analysis. Theor Criminol 2(2):177–210

Beirne P (1999) For a nonspeciesist criminology: animal abuse as an object of study. Criminol Interdiscipl J Official Publication Am Soc Criminol 37(1):117–149

Beirne P, South N (ed) (2007) Issues in green criminology. Devon: Willan

Beirne P (2009) Confronting animal abuse. Rowman and Littleman, Lanham

Beirne P (2011) Animal abuse and criminology: introduction to a special issue. Crime Law Social Change 55(5):349–357

Bourdieu P (1995) Outline of a theory of practice. Cambridge, Cambridge University Press

Cazaux G (1999) Beauty and the beast: animal abuse from a non-speciesist criminological perspective crime. Law Soc Change 31(2):105–125

CITES convention http://www.cites.org/eng/disc/text.shtml Accessed 14 June 2011

Cohen S (2001) States of denial. Polity Press, Cambridge

Donovan J, Adams C (1996) Beyond animal rights. A feminist caring ethic for the treatment of animals. New York: Continuum

Du Plessis MA (2000) CITES and the causes of extinction. In: Hutton J, Dickson B (eds) Endangered species threatened convention. Earthscan, London

Gonzales JA (2003) Harvesting, local trade, and conservation of parrots in the Northeastern Peruvian Amazon. Biol Conserv 114(2003):437–446

Grossman A (2009) Tusks and trinkets: an overview of illicit ivory trafficking in Africa. Afr Secur Rev 18(4):50–69

Herbig J (2010) The illegal reptile trade as a form of conservation crime: a South African criminological investigation. In: White R (ed) Global environmental harm: criminological perspectives. Willan Publishing, Devon

Hutton J, Dickson B (eds) (2000) Endangered species threatened convention. Earthscan, London

IWAF (International fund for animal welfare) (2008) Killing with keystrokes. An investigation of the Illegal wildlife Trade on the world wide web. http://www.ifaw.org/Publications/Program_Publications/Wildlife_Trade/Campaign_Scientific_Publications/asset_upload_file64_12456.pdf. Accessed 14 June 2011

Lemieux AM, Clarke R (2009) The international ban on ivory sales and its effects on elephant poaching in Africa. Br J Criminol 49:451–471

Low R (2003) The worldwide trade in wild-caught parrots. In: Proceedings to the International Aviculturists Society. http://www.funnyfarmexotics.com/IAS/2003Proceedings/Trade_Parrots_Low.pdf. Accessed 8 June 2010

Lynch M (1990) The greening of criminology: a perspective on the 1990s. Crit Criminol 2:1–5

Memon (2009) The illegal parrot trade in Latin America and its consequences to parrot nutrition, health and conservation. Bird Popul 9:76-83

Metz S (2007) Rehabilitation of Indonesian parrots from the Illegal Wild Bird Trade: early experiences on Seram island, Indonesia. In: Proceedings of Association of Avian veterinarians, Australian comittee. http://www.indonesian-parrot-project.org/pdf_files/AAV_07_6-29-07.PDF. Accessed 29 June 2010

Nibert D (2002) Animal rights/human rights. Rowman and Littlefield, Lanham

Noske B (1997) Beyond boundaries. Humans and animals. Montreal: Black Rose Books

Pires S, Clarke R (2011a) Sequential foresting, itinerant fences and parrot poaching in Bolivia. Br J Criminol 51:314–335

Pires S, Clarke R (2011b) Are parrots CRAVED? An analysis of parrot poaching in Mexico. J Res Crime Delinq, online 15 March p 1–22

Reeve R (2002) Policing international trade in endangered species: The CITES treaty and compliance. Earthscan, London

Regan T (1983) The case for animal rights. University California Press, Berkeley

Singer P (1975) Animal liberation. Pimlico, London

Sollund R (2008) Causes for speciesism: difference, distance and denial. In: Sollund R (ed) Global harms. Ecological crime and speciesism. Nova science publishers, New York

Sollund R (2011) Expressions of speciesism: the effects of keeping companion animals on animal abuse, animal trafficking and species decline. Crime, Law and Social Change, p 1–15, Online ISSN 0925-4994

South N (1998) A green field for criminology: a proposal for a perspective. Theor Criminol 1998(2):211–233

South N, Beirne P (2006) Green criminology. Aldershot: Ashgate

South N, Wyatt T (2011) Comparison of the Illicit wildlife and Drug trades: an exploratory study. Deviant Behav 32:538–561

Traffic (2008) What's driving the wildlife trade? A review of expert opinion on economic and social drivers of the wildlife trade and trade control efforts in Cambodia, Indonesia, Lao PDR and Vietnam

Traffic (2011) http://www.traffic.org/home/2011/6/10/experts-urge-better-regulation-of-bushmeat-trade.html Accessed 14 June 2011

Warchol GL, Zupan LL, Clarke W (2003) Transnational criminality: an analysis of the illegal wildlife market in Southern Africa. Int Crim Justice Rev 13:1–26

Wellsmith M (2011) Wildlife crime: the problems of enforcement. Eur J Crime Policy 17:125–148

Westerhuis' D (in press) Trade in wildlife: a moral question reviewed

Weston MK, Memon MA (2009) The illegal parrot trade in Latin America and its consequences to parrot nutrition, health and conservation. Bird Popul 9:76–83
White R (2008) Crimes against nature: environmental criminology and ecological justice. Willan Publishing, Devon
White R (2011) Transnational, environmental crime: toward an eco-global criminology. Routledge, London
Wright T et al (2001) Nest poaching in neotropical parrots. Conservation Biology 15(3):710–720
Wyatt T (2009) Exploring the organization of Russia Far East's Illegal wildlife trade: two case studies of the illegal fur and illegal falcon trades. Global Crime 10(1):144–154, Vol. 10 (1–2) 144–154
Wyatt T (2011) The Illegal raptor trade in the Russian federation. Contemp Justice Rev 14(2):103–123
Zimmerman ME (2003) The black market for wildlife: combating transnational organized crime in the Illegal wildlife trade. Vanderbilt J Transnational Law IVOL 36:1657

Crimes Against Humanity

► Human Rights Violations in Criminal Court

Crimes of Empire

► Crimes of the Powerful

Crimes of Globalization

David O. Friedrichs[1] and Dawn L. Rothe[2]
[1]Department of Sociology/Criminal Justice, University of Scranton, Scranton, PA, USA
[2]Department of Sociology and Criminal Justice, Old Dominion University, Norfolk, VA, USA

Overview

The topic addressed in this entry is not one of the conventional topics within surveys of the terrain of criminology and criminal justice literature. But a criminology that aspires to remain relevant in a rapidly changing world during the course of the twenty-first century must transcend the traditional boundaries of criminological inquiry. Some forms of crime, broadly defined, have been visible to scholarly students of crime from the outset, with homicide as the classic example. But other forms of crime, broadly defined, have been far less visible – and in some cases virtually invisible – to criminological inquiry. This entry addresses one form of such crime: crimes of globalization and more specifically the crimes of international financial institutions. These "crimes of globalization" are situated within the context of the broad range of crimes that are linked to and driven by the expanding conditions of globalization. Some attention is given to two key concepts and phenomena: "crime" and "globalization". The fact that the crimes of international financial institutions cannot be readily classified under either the heading of "corporate crime" or "crimes of states" is a core reason for the relative invisibility of this type of crime. Nonetheless, as "corporate crime" and "crimes of states" are both classified as organizational crimes, international financial institutions do fit within the realm of research on "white-collar crime" or "the crimes of the powerful." Crimes of globalization are intimately interrelated with crimes of states and political white-collar crime (especially corruption), as well as the interests of powerful private sector entities such as corporations and investment banks.

How is such crime best explained and understood? The basic parameters of an integrated theory of crimes of globalization are identified here. The overall lack of accountability and the profound challenges involved in controlling crimes of globalization are also addressed. Some concluding observations address the potential value of further criminological attention to crimes of globalization.

Crimes of Globalization: The Fundamentals

Defining Crimes of Globalization

Conventional crime and its control is principally a local phenomenon and somewhat less so a state and national (or federal) phenomenon.

Accordingly, most criminologists are not especially receptive to adopting a global framework for defining crime or for the study of crime and its control. In the conventional view, crime as defined by law on the state level – or locally or nationally – is the proper domain of criminology, along with its control, and it is through the study of such crime and its control that criminology is most likely to make its contribution to the advancement of human knowledge and to produce knowledge specifically useful for identifying optimal public policies in response to crime.

Crimes of globalization are those demonstrably harmful policies and practices of institutions and entities that are specifically a product of the forces of globalization and that by their very nature operate within a global context. Although these crimes can involve violations of criminal laws on the state or international level, they may also incorporate harms not specifically addressed by statutory law. The vastly disproportionate influence of elite interests over the formal criminal law is accordingly taken into account in a definition that transcends the boundaries of such law. It is not typically the specific intent of those who engage in crimes of globalization to cause harm. Rather, the devastating harm to vulnerable people in developing countries is a consequence of the skewed priorities of institutions and entities which favor the interests of the powerful and the privileged.

The concept of crimes of globalization is not synonymous with two formulations that have received significant recent attention: the "globalization of crime" and "globalization and crime" (e.g., Aas 2007; Friman 2009; Larsen and Smandych 2008). The first of these terms refers broadly to long-standing forms of crime now carried out in an increasingly global context, and the second term refers broadly to the influence of globalization on crime, conventionally defined. Simon Mackenzie (2006) has introduced the term "systematic crime" in his discussion of the broad forms of global harm emanating from the practices of international financial institutions, and their complicity in denying the link between supporting interests of advanced economies and harm in developing countries.

The concept of crimes of globalization, as originally formulated, was limited to the demonstrably harmful activities of international financial institutions, with a special focus on one of these institutions, the World Bank. However, these crimes intersect with a range of other forms of crime engaged in by powerful entities, including crimes of states, political white-collar crime, and state-corporate crime (Friedrichs 2010). Multiple complex interconnections exist between these different types of globalized harm.

Some refinement of the definition of crimes of globalization seems warranted. In the interest of greater clarity, the notion of crimes of international financial institutions specifically is best classified as a subtype of the broader category of crimes of globalization. The two principal international financial institutions are the International Monetary Fund (IMF), which seeks to maximize financial stability, and the World Bank, primarily focused on promoting development (Woods 2006). The World Trade Organization (WTO) is often aligned with these international financial institutions and has many parallel attributes and issues but strictly speaking is an international regulatory entity, with its primary formal mission being to foster trade. In a rapidly changing global economy, the roles of the international financial institutions have been increasingly questioned. These institutions have many ties with each other, and the lines of demarcation between their activities can become quite blurred.

The World Bank and the International Monetary Fund: International Financial Institutions

The World Bank is here addressed in some detail, to exemplify the fundamental nature of an international financial institution (Goldman 2006; Marshall 2008; Weaver 2008). The *World Bank*, formally the International Bank for Reconstruction and Development (IBRD), was established at the Bretton Woods Conference in 1944 to help stabilize and rebuild economies ravaged by World War II. Eventually it shifted its focus to an emphasis on aiding developing nations. The

Bank makes low-interest loans to governments of its member nations and to private development projects backed by those governments with the stated aim to benefit the citizens of those countries. The World Bank claims to contribute to the reduction of poverty and to the improvement of living standards in developing countries. Today the Bank is a large, international operation with more than 10,000 employees, 180 member states, and annual loans of $30 billion.

The World Bank was established, along with the International Monetary Fund, at the behest of dominant Western states, with little input from developing countries. They are disproportionately influenced by or manipulated by elite economic institutions and entities – e.g., transnational mining companies – and have been characterized as agents of global capital. In developing countries, they deal primarily with the political and economic elites of those countries with little direct attention to the perspectives and needs of indigenous peoples. They have loaned money to ruthless military dictatorships engaged in murder and torture and denied loans to democratic governments subsequently overthrown by the military. They have favored strong dictatorships over struggling democracies because they believe that the former are more able to introduce and see through the unpopular reforms their loans require. Borrowers typically are political elites of developing countries and their cronies, although repaying the debt becomes the responsibility of these countries' citizens, most of who do not benefit from the loans. The Bank has lost $100 billion due to fraud and corruption, in one estimate, over a period of several decades. The privileged in developing countries have been the principal beneficiaries of World Bank and International Monetary Fund loans, not poor people in those countries.

The World Bank and the International Monetary Fund have been the target of much criticism, especially in the recent era. They have been characterized as paternalistic, secretive, and counterproductive in terms of their claimed goals of improving people's lives. They have been charged with complicity in policies with genocidal consequences, with exacerbating ethnic conflict, with increasing the gap between rich and poor, with fostering immense ecological and environmental damage, with neglecting agriculture crucial to survival in developing countries, and with the callous displacement of vast numbers of indigenous people in these countries from their original homes and communities.

Critics claim that many less developed countries that received World Bank and International Monetary Fund loans are worse off today in terms of poverty and that the severe austerity measures imposed on borrowing countries, deemed necessary to maximize the chances of loans being repaid, impact most heavily on the poorest and most vulnerable citizens. The "structural adjustment agreements" in developing countries have been shown to also impact negatively on human rights in those countries. The building of dams has been the single most favored World Bank project, but even its own experts concede that millions of people have been displaced as a result of these dams. In many of these dam projects, resettlement plans either have been nonexistent – in violation of the Bank's own guidelines – or have been inadequately implemented. In one notorious case in the 1970s, anti-dam protesters in Guatemala were massacred by the military.

At a World Bank meeting in Berlin in 1988, protesters called for the establishment of a Permanent People's Tribunal to try the World Bank (and the International Monetary Fund) for "crimes against humanity." An American anthropologist has characterized the forced resettlement of people in dam-related projects as the worst crime against them, short of killing them. An American biologist characterized the World Bank's report on the environmental impact of a dam project in a developing country as "fraudulent" and "criminal."

Globalization in Relation to Crimes of Globalization

Globalization clearly has many different dimensions. Those most pertinent within the realm of white-collar crime generally, and crimes of globalization specifically, include the following:

(1) the growing global dominance and reach of neoliberalism and a free market, capitalist system that disproportionately benefits wealthy and powerful organizations and individuals; (2) the increasing vulnerability of indigenous people with a traditional way of life to the forces of globalized capitalism; (3) the growing influence and impact of international financial institutions and the related, relative decline of power of local or state-based institutions; and (4) the nondemocratic operation of international financial institutions, taking the form of globalization from above instead of globalization from below.

Crimes of Globalization, Transnational Crimes, and International Crimes

The relationship of crimes of globalization to the familiar but sometimes unclearly invoked terms transnational crimes and international crimes requires some attention here. *Transnational crimes* are essentially forms of crime that are increasingly carried out across borders and via international or global networks. The forces of globalization are transforming and amplifying structures of opportunity for a wide range of different forms of criminal activity (Aas 2007). Some of these forms of crime (e.g., human, arms, and drug trafficking) are hardly new, but the transnational dimension of them has expanded. Some of these forms of crime are quite new (e.g., cybercrime), but their transnational dimension greatly enhances the challenges of controlling them.

Organized crime has long had transnational dimensions. The processes of globalization have been transforming some dimensions of such crime, with these transnational dimensions increasingly central to the operation of emerging forms of organized crime. But Hazel Croall (2005) is surely correct in arguing that there has been a disproportionate amount of criminological attention to the globalization of organized crime, when the global activities of multinational corporations and financial institutions cause far greater harm. Certainly the threat of transnational or global terrorism is substantial, but the argument can be made that it also has received disproportionate attention relative to other forms of transnational or global harm. Especially since 9/11, there has been a huge amount of attention to transnational or global terrorism.

International crimes are best conceived of as violations of international law, which in their generic form (e.g., genocide, war crimes, crimes against humanity, and massive violations of human rights) have a long history. Such crimes have often been committed within national boundaries but are increasingly carried out globally. International crimes are most typically thought of as crimes of states but may also be committed by insurgencies, militias, and other parties. Corporations – and increasingly multinational corporations – are also complicit in international crimes. For example, some of the corporations operating in Nazi Germany and its occupied territories played a role in the Holocaust, serving as classic cases of such crimes. In the more recent era, such corporations as Blackwater (XE), Sandline, and Halliburton have been accused of violations of international law (Rothe 2009). The conditions of globalization produce expanding opportunities for such crime.

The lines of demarcation between crimes of globalization, transnational crimes, and international crimes are sometimes fluid and complex. But the key actors typically involved, and the bodies of law violated, tend to be different.

Any coherent discussion of "crimes of globalization" must also address this phenomenon historically and cross-culturally. Some activities carried out in ancient times can be described as early forms of globalized white-collar crime. Monumental historical crimes were committed in the name of imperialism and colonial expansion globally, over a period of many centuries. Obviously the global slave trade in conjunction with this, continuing into the nineteenth century, was a crime of epic proportions. Furthermore, imperialistic enterprises continued through the twentieth century into the twenty-first, and slavery and a thriving slave market have hardly become extinct in the contemporary world, despite the collapse of traditional colonial regimes and forms of slave trade. More broadly, "crimes of empire" are a feature of our world today.

Crimes of Globalization, State-Corporate Crime, and Crimes of States

The crimes of international financial institutions (IFIs) have a generic relationship to state-corporate crimes insofar as they are cooperative ventures involving public sector and private sector entities and in some respects are hybrid public/private sector entities. The literature on state-corporate crime (e.g., Michalowski and Kramer 2006) has focused on crimes arising out of cooperative ventures involving states and corporations. In one sense, crimes of globalization could be characterized as a neglected, cognate form of such crime: i.e., state-international financial institution crime.

While the above categories may appear to be separate phenomenon, we suggest the connections between them are not so easily separated. The intersection of business and government has led to increased cases of a "globalized criminality." In the recent era, Western states as well as corporations have promoted neoliberalism or a supposed "free market" model for the global political economy. Within such an environment, the crimes of globalization of international financial institutions are intertwined with crimes of states. The policies and practices of the international financial institutions are largely driven by the global agenda of powerful developed states such as the United States.

In many of the developing countries, corrupt political oligarchs facilitate the promotion of this agenda, despite it being largely at odds with the interest of their citizens. Cases of such corrupt practices have been especially pronounced and well documented in the case of sub-Saharan countries of Africa, such as the Democratic Republic of Congo, Uganda, Rwanda, and Senegal (Rothe 2010b). But altogether these corrupt practices are a global phenomenon.

Origins of the Concept of Crimes of Globalization

The concept of "crimes of globalization" was first put forth in a paper for the American Society of Criminology Annual Meeting in 2000 (subsequently published as an article in *Social Justice* in 2002) with the specific title "The World Bank and Crimes of Globalization: A Case Study" by David O. Friedrichs and Jessica Friedrichs. The specific inspiration for this concept was the experience of the junior author, Jessica Friedrichs, living among poor fishing people in northeast Thailand whose way of life was being destroyed by a dam at least partially financed by the World Bank. Jessica Friedrichs returned to the United States in April 2000 from her junior year experience in Thailand, 2 months earlier than anticipated, to make a presentation on the Pak Mun Dam situation at an antiglobalization (or global justice) demonstration in Washington, D.C., that month. The senior author of the article, David Friedrichs, with a long-standing interest in crimes of the powerful, realized that while at least some of the policies and practices of international financial institutions such as the World Bank had demonstrably harmful consequences, this form of harm had been quite wholly neglected by criminologists.

The Pak Mun Dam Case

In the Pak Mun dam case, the World Bank helped finance the building of the dam in eastern Thailand in the early 1990s. The process of planning, constructing, and operating this dam was undertaken without obtaining input from the fishermen and villagers who lived along the river. The construction of the dam had a detrimental effect on the environment, flooding the adjacent forests. This effect violated the World Bank's own policies on cultural property destruction. Many edible plants upon which locals were dependent for their sustenance and for income were lost. Villagers who used the river waters for drinking, bathing, and laundry developed skin rashes. Most importantly, a severe decline in the fish population occurred. As a consequence, the way of life of indigenous fishermen dependent upon abundant fish for food and income was annihilated. The resettlement of the fishermen and compensation for their losses were wholly inadequate. Traditional communities began to disintegrate. Many of those affected by these developments organized protest villages and engaged in other actions calling for the Thai government and the World Bank to take responsibility for the devastation they caused by building the dam, which cost far more than expected and generated far less electricity than had been anticipated.

Crimes of Globalization: Subsequent Studies

In the wake of the original article focusing upon the World Bank and the case of the Pak Mun dam, a number of criminologists have applied the concept of crimes of globalization to other circumstances. For example, (Rothe et al. 2006) conducted research that explored the interrelations between the International Monetary Fund and the World Bank and legacies of colonialism along with foreign policies that set the stage for large-scale atrocities and crimes of states. Exploring the circumstances leading to the sinking of the ferry *Le Joola*, the authors demonstrated that the state of Senegal itself had core liability for this maritime tragedy, with its dramatic loss of lives. The government readily admitted its errors and several ministers either stepped down or were removed from their positions. However, despite unequivocal governmental responsibility, Rothe, Muzzatti, and Mullins advance the case that the sinking could not be characterized simply as a case of state crime. Rather, a thorough investigation and analysis of the reasons and forces behind the *Le Joola* sinking suggested that international financial institutions bore some clear culpability for the disaster. In response to structural adjustment programs (SAPs) imposed by the International Monetary Fund, the Senegalese government was forced to cut spending in many areas. These spending cuts extended to ferry programs central to transportation in Senegal, especially in relation to its geographic location. This had a direct impact on the upkeep and return of the *Le Joola* to open waters. The ferry capsized with only one of its two engines functioning, resulting in the deaths of 1863 passengers. This was the second largest maritime disaster in history. Most crucially, the authors of this study demonstrated why scholars need to examine the criminogenic effects of policies and practices of international financial institutions in developing countries such as Senegal. These policies and practices privilege capitalistic profit over human lives and a better quality of life for people in developing countries. Accordingly, this is crime against vulnerable human beings.

An article by (Rothe et al. 2009) took a parallel approach, exploring the role of international financial institution policies in the conditions leading to the Rwandan genocide in 1994. While the World Bank and the International Monetary Fund did not seek to instigate economic collapse or to promote genocide, their policies and their systematic inattention in Rwanda set the stage for political and economic disaster as well as the genocide itself. The authors suggested that these international financial institutions knowingly violated their own standards as well as international human rights principles. Through the imposition of harsh conditions tied to their financial aid, they facilitated criminal activities on a massive scale.

In an article published in 2008, Ezeonu and Koku also adopted the crimes of globalization concept. They demonstrated the key contributing role played by the neoliberal policies of international financial institutions in sub-Saharan Africa, in expanding the vulnerability of people in this region to HIV infection. They called for more systematic criminological attention to the victimization of people in developing countries as a consequence of the promotion of neoliberal policies and practices in an increasingly globalized world (see also Ezeonu 2008).

In a similar vein, Rothe (2010a, b) has provided an analysis of the complicity of international financial institutions in heightened levels of corruption and the suppression or violation of human rights in developing countries. Analyzing such complicity seems especially important given that these institutions claim to be engaged in combating corruption in developing countries, including those linked to transnational and multinational corporations. The anticorruption initiatives include threatening to withhold much needed economic aid and loans in the absence of action taken against corrupt activities in these countries. Rothe has illustrated the specific role of the international financial institutions in the illegal expropriation of the rich natural resources of the Democratic Republic of Congo by the neighboring countries of Uganda and Rwanda. Beyond theft on a grand scale, Rwandan and Ugandan state forces and militias also engaged in especially atrocious human rights violations conducted against civilian populations, including forced labor, systematic rape, and widespread killing. Through their funding of African states engaged in crimes against both their own citizens and those of neighboring countries, the international financial institutions bear some responsibility for these crimes.

Parallel circumstances have arisen in other parts of the world. Stanley (2009) has analyzed the role of the international financial institutions in Indonesia. They directed some $30 billion to the Suharto regime, despite its known record of massive corruption, false accounting, and a militaristic appropriation of aid funds. As the World Bank's focus was on supporting Indonesia, the state was able to use funds supposedly intended to reduce poverty in its brutal campaign against civilians in the state of Timor-Leste. This campaign had as its purpose terrorizing people to deter them from voting for independence from Indonesia. One could identify many other cases in Asia and other parts of the world where the international financial institutions have been complicit in supporting corrupt, authoritarian regimes and facilitating their massive violations of human rights.

The concept of crimes of globalization has also been adopted in relation to forms of crime that occur in the context of globalization but do not specifically involve the international financial institutions. Wright and Muzzatti (2007) have addressed the global restructuring of agriculture and food systems – agri-food globalization – with some specific attention to the victimization of huge numbers of animals: e.g., 58,000 sheep stranded at sea for almost 3 months in 2003, in violation of animal welfare law. Altogether, policies and practices relating to the global restructuring of agriculture and food systems were driving up food prices, pushing tens of millions of people toward hunger and starvation, and developed country farm subsidies were driving large numbers of farmers in developing countries into desperate circumstances – to the advantage of corporate and high finance interests in the wealthy countries of the world.

The notion of crimes of globalization has also been implicitly adopted by some authors who are not criminologists and who are addressing a broad public audience. A book by John Perkins (2005), entitled *Confessions of an Economic Hit Man*, became a best seller following its publication in 2005. He defined economic hit men (EHM) as "highly paid professionals who cheat countries around the globe out of trillions of dollars" (Perkins 2005, p. ix). In his case, as an employee of an international consulting firm, Perkins claims to have participated in a range of activities involved in funneling funds from international financial institutions and international aid organizations into the hands of major transnational corporations and a small number of wealthy and influential families in developing countries. Economic hit men are engaged in persuading developing country leaders to become part of a vast global network that ultimately serves the interest of US-based corporations and US businesses generally, at the enormous expense of the people of the developing countries. In a subsequent book, *The Secret History of the American Empire*, Perkins (2007) further explores some of these themes. Although these two books have been criticized as self-dramatizing, they may have succeeded in raising the consciousness of new audiences about crimes of globalization.

In *A Game as Old as Empire: The Secret World of Economic Hit Men and the Web of Global Corruption* (Hiatt 2007), a book inspired by the Perkins' best seller, various authors address aspects of "the corporatocracy" ("the powerful people who run the world's biggest corporations, the most powerful governments, and history's first truly global empire," p. 20). These authors address such matters as the hundreds of billions of dollars that developing countries spend annually for servicing their debt, the world of offshore banking, the expropriation of Africa's oil wealth, the role of export credit agencies in boosting overseas sales for multinational corporations, and the mirage of debt relief.

An Integrated Theory of Crimes of Globalization

On the global level, any initiatives at explanation should begin with a sense of humility: i.e., the globalized world we live in is so endlessly complex, with countless different variables interacting on multiple different levels, that full-fledged explanation (and prediction) is tremendously difficult. Nonetheless, the simplistic and one-dimensional explanations of crime advanced by some criminologists – e.g., low self-control theory – are quite useless in relation to crimes of globalization. To the extent that we can hope to understand crimes of globalization, an integrated theoretical approach is necessary.

There have been attempts to provide integrated theories to explain traditional street crime. Integrated theories have also been applied to organizational offending, including corporate crime, state-corporate crime and crimes of states, or violations of international criminal law (Rothe 2009). These integrated theories provide a foundation for the integrated theory of crimes of globalization outlined below.

An integrated model of offending takes into account motivations, opportunities, controls, and constraints at four levels of analysis (interactional, organizational, structural, and international). In doing so, the integrated theory combines insights from criminological theories and other disciplines to explain the multiple levels at play within each specific case. Social learning (e.g., differential association), anomie, strain, rational choice, routine activities, techniques of neutralization, and control theories have been brought together within an integrated approach to address some specific components of organizational offending.

An integrated theory of crimes of globalization also draws from sociological organizational analysis (e.g., network and system analysis) and from political economy models to explain policies and practices pursued at the state and international levels. This integrated approach recognizes that organizations are not monolithic entities and do not operate within a vacuum. Instead, it recognizes agency at the interactional level while simultaneously noting the impact of organizational culture and the broader structure on the individual agency level of decision-making. Key components of an integrated theory of crimes of globalization include the motivating forces of pursuit of both legitimacy and profit as organizational goals, the internal structure and reward systems of organizations, and external pressures which are imposed on organizations.

The specific political context within which international financial institutions operate is one in which the interests of the powerful countries that provide most of the funding for their programs are aligned with those of the political and business elites of the developing countries which are being aided. On a structural (or state-based) level, then, international financial institution programs are ultimately skewed to advance or protect the interests of elites while all-too-often inflicting harm on ordinary citizens and indigenous populations in developing countries.

On the organizational level, a criminogenic environment is fostered within the international financial institutions. The officers and staff of these institutions are rewarded for getting out aid and loans and implementing programs in developing countries, not for ensuring that these programs do not cause long-term harm for the people of the countries being aided. The “success” of the programs is measured more in terms of their size and scope rather than their ultimate effect. International financial institutions reward their personnel for technical proficiency rather than for concerning themselves with the perspectives and needs of the ordinary people of developing countries. Despite professed noble objectives and the absence of a specific intent to do harm, the mode of operation of the international financial institutions is intrinsically criminogenic and undemocratic. Their key deliberations are carried out secretly and with an absence of accountability.

Controlling and Responding to Crimes of Globalization

Throughout the course of the twentieth century and into the early stages of the twenty-first century, it became widely recognized that local, state, and national (or federal) institutions of social control were no longer sufficient for the challenges arising in an increasingly globalized world. The expanding adoption of transnational, international, and global institutions to address a broad spectrum of harmful activities – from cross-border trafficking to crimes of war – has been one of the defining attributes of the contemporary era. But the international financial institutions operate with a singular absence of effective accountability and oversight. There are a range of international declarations and accords that prohibit some of the harmful activities engaged in by the international financial institutions ranging from fostering corruption to complicity in violations of human rights. As specific examples, we have the *United Nations Convention Against Corruption* Article 2 (c) which defines officials of public international organizations in a way that could be applied to international financial institutions. Other such international accords define transnational criminal activities and human rights violations in a form that can also be interpreted as applying to the international financial institutions. However, no international institutions or tribunals specifically have jurisdiction over, take complaints on, or adjudicate the broad range of harmful activities engaged in by the international financial institutions. The powerful countries that dominate these institutions – notably, the United States and Western European countries – are highly unlikely to call them to account for their harmful activities, since their policies and practices are aligned with and advance the economic interests of these countries.

The Global Justice Movement and Crimes of Globalization

To date, global justice (or antiglobalization) activists and their protest rallies have been the principal entities holding international financial institutions accountable for their harmful policies and practices in developing countries. A World Bank Bonds Boycott is one manifestation of this, since the World Bank raises most of its funds by issuing bonds, and these bonds are purchased by ordinary citizens through pension funds, labor unions, churches, municipalities, universities, and other entities.

In a world where citizens of developing countries increasingly have access through the Internet to specific information about the exploitative activities of Western institutions in their own countries, demands for addressing crimes of globalization are likely to increase. Populist challenges to autocratic, corrupt regimes and the immense economic inequality that they sustain will surely be one of the defining themes of the twenty-first century. In developing countries, growing numbers of people are characterizing themselves as victims of monumental crimes carried out by these regimes, and Western governments and institutions are seen as complicit in fundamental ways in these crimes. During 2011-2012 this situation played out in Egypt, Tunisia, Libya, Syria, Bahrain, Yemen, and other Middle Eastern countries. The gross maldistribution in terms of consumption of natural resources that exists between the developed and developing world is also highly likely to be subjected to increasing challenge. Going forward, it seems highly likely that Western entities, including international financial institutions, will increasingly be "indicted" for their perceived crimes against the people of developing countries.

Concluding Observations

The "crimes of globalization" concept has not generated significant controversy to date because it continues to lie outside the focus of almost all students of crime and criminal justice. But criminologists are in collective denial if they fail to recognize that the crimes of the powerful – including those of international financial institutions – will attract growing attention from ordinary people all over the globe. In an increasingly globalized world, international financial institutions are immensely important players and are likely in the future to assume even greater importance and influence globally. A core premise for this entry is that international financial institutions, whatever their stated missions, have in their policies and practices been complicit in very large-scale forms of social harm and that this harm should be recognized as a significant form of crime, and accordingly it is both a legitimate and a useful project for criminologists to apply a criminological framework to the understanding of these crimes of globalization and the challenges arising in relation to the prevention and control of such crime.

Related Entries

- ► Crimes of the Powerful
- ► Human Trafficking
- ► International Crime Victimization Survey

Recommended Reading and References

Aas KF (2007) Globalization and crime. Sage, Los Angeles

Croall C (2005) Transnational white-collar crime. In: Sheptycki J, Wardak A (eds) Transnational and comparative criminology. Glasshouse Press, London, pp 247–268

Ezeonu I (2008) Crimes of globalization: health care, HIV and the poverty of neo-liberalism in Sub-Saharan Africa. Int J Soc Inquiry 1:113–134

Ezeonu I, Koku E (2008) Crimes of globalization: the feminization of HIV pandemic in Sub-Saharan Africa. The Global South 2

Friedrichs DO (2010) Trusted criminals: white collar crime in contemporary society, 4th edn. Cengage, Belmont

Friedrichs DO, Friedrichs J (2002) The World Bank and crimes of globalization: a case study. Soc Justice 29:1–12

Friman RH (2009) Crime and the global political economy. Lynne Rienner, Boulder

Goldman M (2006) Imperial nature: the World Bank and struggles for justice in the age of globalization. Yale University Press, New Haven

Hiatt S (ed) (2007) A game as old as empire: the secret world of economic hit men and the web of global corruption. Berrett-Koehler, San Francisco

Larsen N, Smandych R (eds) (2008) Global criminology and criminal justice. Broadview Press, Toronto

Mackenzie S (2006) Systematic crimes of the powerful: criminal aspects of the global economy. Soc Justice 33:162–182

Marshall K (2008) The World Bank: from reconstruction to development to equity. Routledge, New York

Michalowski RJ, Kramer RC (eds) (2006) State-corporate criminality: wrongdoing at the intersection of business and government. Rutgers University Press, New Brunswick

Perkins J (2005) Confessions of an economic hit man. Ebury Press, London

Perkins J (2007) The secret history of the American empire: the truth about economic hit men, jackals, and how to change the world. Penguin, London

Rothe DL (2009) State criminality: the crime of all crimes. Lexington Books, Lanham

Rothe DL (2010a) Facilitating corruption and human rights violations: the role of international financial institutions. Crime Law Soc Change 53:397–412

Rothe DL (2010b) The dragon rising: international financial institutions and the emerging role of China. In: Chambliss WJ, Michalowski R, Kramer RC (eds) State crime in a globalized age. Willan Press, London, pp 152–169

Rothe DL, Muzzatti S, Mullins CW (2006) Crime on the high seas: crimes of globalization and the sinking of the Senegalese ferry Le Joola. Crit Criminol 14:159–180

Rothe DL, Mullins CW, Sandstrom K (2009) The Rwandan genocide: international finance policies and human rights. Soc Justice 35:66–86

Stanley E (2009) Torture, truth and justice, the case of Timor-Leste. Routledge, London

Weaver C (2008) Hypocrisy trap: the World Bank and the poverty of reform. Princeton University Press, Princeton

Woods N (2006) The globalizers: the IMF, the World Bank and their borrowers. Cornell University Press, Ithaca

Wright W, Muzzatti S (2007) Not in my port: the “death ship” of sheep and crimes of agri-food globalization. Agr Hum Val 24:133–145

Crimes of the Powerful

David Kauzlarich[1] and Dawn L. Rothe[2]

[1]Department of Sociology and Criminal Justice Studies, Southern Illinois University Edwardsville, Edwardsville, IL, USA

[2]Department of Sociology and Criminal Justice, Old Dominion University, Norfolk, VA, USA

Synonyms

Corporate crime; Crimes of empire; Crimes of globalization; Organizational crimes; State crime; State-corporate crime; White-collar crime

Overview

Crimes of the powerful have occurred throughout humanity’s existence, from the ancient armies

Crimes of the Powerful, Fig. 1 Crimes of the powerful by type

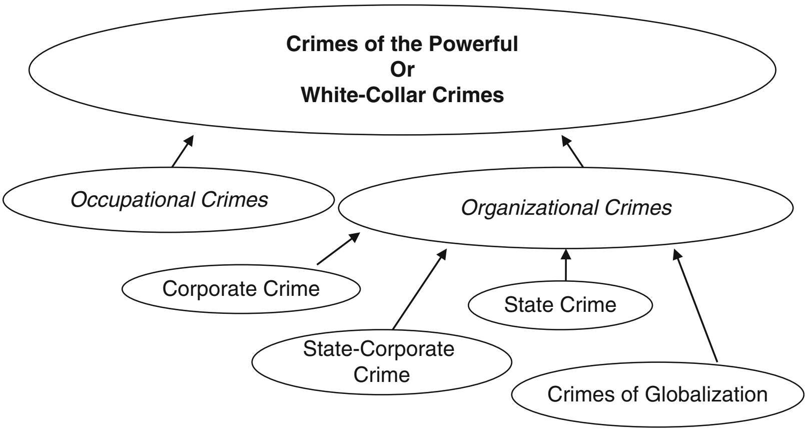

and Roman leadership to the aristocratic criminal of the eighteenth Century to the modern corporate CEO and head of state. While these behaviors did not come under the scrutiny of criminologists until the last century, there is a growing body of critical theoretical frames that have been used to explain the etiological factors associated with crimes of the powerful.

Fundamentals

Theoretical explanations of the crimes of the powerful are not straightforward or universally applicable. As such, it helps to first understand the typical forms by which these types of criminality are generally categorized as well as the levels of analysis with which theories can be applied. Crimes of the powerful, also referred to as white-collar crimes, can be broken down into two main categories: occupational and organizational crimes. Occupational crimes are those committed by individuals within the context of their occupation for their own self-interests whereas organizational crimes are those committed by individuals within the context of their occupation for the interests of or benefit of the organization. The latter does not imply that the individual may not benefit as well, however, a major component of organizational crime is the inclusion of organizational benefit (Fig. 1).

Organizational crimes can be broken down into four major subtypes: (1) state crime, (2) corporate crime, (3) state-corporate crime, and (4) crimes of globalization.

State crimes are illegal or socially injurious act of omission or commission by an individual or group of individuals in an institution of legitimate governance which is executed for the consummation of the operative goals of that institution of governance (Kauzlarich and Kramer 1998). Genocide, human rights violations, war crimes, illegal war, and crimes against humanity are actions that fall under the category of state crime. State crimes are historically and contemporarily ubiquitous and result in more injury and death than traditional street crimes such as robbery, theft, and assault. Consider that genocide during the twentieth century in Germany, Rwanda, Darfur, Albania, Turkey, Ukraine, Cambodia, Bosnia-Herzegovina and other regions claimed the lives of tens of millions and rendered many more homeless, imprisoned, and psychologically and physically damaged (Rothe and Kauzlarich 2010).

Corporate crimes are committed on behalf of business interests, sometimes by individuals, sometimes by groups; they surface among the self employed and among executives of companies large and small. Corporate crime may include theft, fraud, and violence. Famous cases of financial fraud include those committed by Enron and Adelphia in which the manipulation of stocks and various forms of deceptive accounting resulted in the theft of billions of dollars from investors. Violations of worker

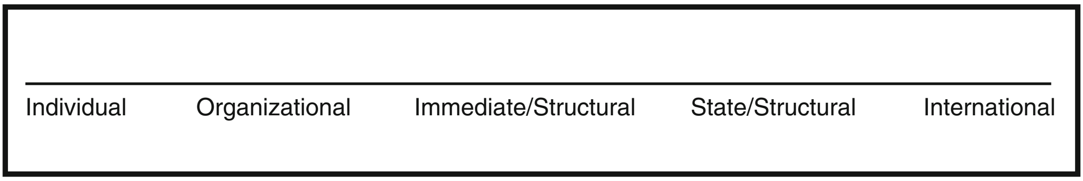

Crimes of the Powerful, Fig. 2 Continuum of theoretical levels of analysis

safety and environmental regulations are also major forms of corporate crime.

State-corporate crimes are illegal or socially injurious actions that result from mutually reinforcing interaction between institutions of political governance and institutions of economic production and distribution (Kramer and Michalowski 1991: 5); also see Michalowski and Kramer (Michalowski and Kramer 2006) and Matthews and Kauzlarich (2000).

The concept of state-corporate crime has been used to examine the space shuttle Challenger explosion (Kramer 1992; Michalowski and Kramer (2006), the environmental devastation caused by US nuclear weapons production (Kauzlarich and Kramer 1998; Michalowski and Kramer (2006), and the deadly fire at the Imperial Food Products chicken processing plant in Hamlet, North Carolina (Aulette and Michalowski 1993; Michalowski and Kramer (2006). Other examples of state-corporate crime include the I.G. Farben Company's involvement with Nazi atrocities (Michalowski and Kramer (2006), the Wedtech case involving defense contractor fraud (Friedrichs 1996), and the violent and deadly crash of ValuJet flight 592 in May of 1996 (Matthews and Kauzlarich 2000).

Crimes of globalization are activities that lead to a range of physical and economic injury committed by international agencies in the interest of facilitating global capitalism (Friedrichs and Friedrichs 2002). Most criminological literature has examined the harms resulting from the World Bank's funding of capitalist expansion into less developed countries, which results in the marginalization of indigenous peoples, higher rates of income inequality, environmental disaster, health crises, and political corruption (Ezeonu and Koku, Ezeonu and Koku 2008; Rothe 2010; Rothe et al. 2006; Rothe et al. 2009). Although most of the harms are not directly criminalized by political authorities as in the case of most traditional forms of corporate crime, expanded definitions of crime rooted in critical criminology are used to frame the activities as subject to criminological examination.

Occupational crimes are committed against an employer are also understood by most critical criminologists as the outcome of the need for and fetish with money in capitalist societies. While not profit-seeking in the same way as organizations, both traditional street crime and occupational white-collar crime stem from economic factors, whether real or perceived.

This distinction between the various forms of crimes of the powerful reviewed above is important for any discussion or application of theory. For example, a theory explaining the motivation of someone stealing 500.00 dollars from their employer for their own financial gain is unlikely to involve the same situational and motivational factors of a head of state committing genocide. Nor are the opportunity structures or controls going to be the same. Likewise, theories apply to different levels of analysis. Each of these can be thought of as operating along a continuum from the individual or interactional level of analysis to the supra-macro level or international level (Fig. 2).

The distinction between these various levels of analysis is relevant to the type of crimes being analyzed as well as the theoretical model that can be applied. For example, a theoretical approach that explains individual level behavior may have limited explanatory power in the case of corporate crime, in which organizational culture or pressures facilitating or "causing" motivating forces. As such, theories that explain occupational crimes for example differ from those that attempt to explain organizational crimes such as corporate, state-corporate, state crime or crimes of globalization.

Critical Theories

Given the range of theories that have been applied to the crimes of the powerful over the last century, it is impossible to review each approach. The emphasis here is on critical and integrated theories of crime, although their relationships to non-critical explanations will be briefly noted. What distinguishes critical theories of crime from other explanations is their opposition to – not just interest in – unequal fundamental political, economic, and social structures and relationships. The major forms of critical theories include Marxist, left realism, feminist, postmodernism, cultural, and peacemaking perspectives. Respectively, their critiques are centered on capitalism, stratification and inequality, patriarchy, modernity, positivist criminology, and warmaking. Critical theories of crime have roots in general sociological conflict theory, and those that focus on crimes of the powerful are almost exclusively inspired by Marxist sociological theory (now often referred to as political economy theory).

While Karl Marx said little about crime, some criminologists, especially critical criminologists, recognize a substantial debt to this nineteenth century scholar. Marx believed that a society's mode of economic production directly influenced the manner in which relations of production are organized and thus determines in large part the organization of social relations, the structure of individual and group interaction. Under a capitalist mode of production, there are those who own the means of production and those who do not. The former group is known as the bourgeoisie and the latter as the proletariat. The bourgeoisie, or ruling class, controls the formulation and implementation of moral and legal norms, and even ideas. Both classes are bound in relationship to one another, but this relationship is asymmetrical and exploitive. This relationship affects law and crime in fundamental ways. Laws are created by the elite to protect their interests at the expense of the proletariat. Marxists might point out that even presumably simple and well supported laws may not work in the interests of the have-nots, though they may be perceived to be a representation of the collective will of a society.

Over a century ago an intellectual follower of Marx, Willem Bonger, applied some of Marx's arguments to crime in capitalistic societies. In Criminality and Economic Conditions (published in English in 1916), Bonger observed that capitalistic societies appear to have considerably more crime than do other societies. Furthermore, while capitalism developed, crime rates increased steadily. Under capitalism, Bonger argued, the characteristic trait of humans is self interest (egoism). Given the emphasis on profit maximization and competition, and the fact that social relations are class structured and geared to economic exchange, capitalistic societies spawn intraclass and interclass conflicts as individuals seek to survive and prosper. Interclass conflict is one sided, however, since those who own and control the means of production are in a position to coerce and exploit their less fortunate neighbors. Criminal law, as one instrument of coercion, is used by the ruling class to protect its position and interests. Since social relations are geared to competition, profit seeking, and the exercise of power, altruism is subordinated to egoistic tendencies.

Applied to crimes of the powerful, Bonger's argument, which is still a basic assumption of modern Marxists, is that those in positions of corporate power are almost always focused on making, sustaining, or increasing profit. Financial crimes by corporations are therefore seen as a logical extension of capitalist logic as are violations of worker and environmental regulations and laws. Quite simply, from a basic Marxist approach, corporate crime is caused by the overriding allegiance to profit no matter who might be hurt along the way. State and state-corporate crime can also be viewed from a Marxist perspective. Considering state-corporate crime, states are seen as primarily facilitators of capital accumulation for corporations. This is especially true in the case of state-facilitated state-corporate crime in which "governmental regulatory institutions fail to restrain deviant business activities, because of direct collusion between business and government, or because they adhere to shared goals whose attainment would be hampered by aggressive regulation" (Kramer and Michalowski 1991: 6). As Barnett (1981: 7) and Chambliss and Zatz (1993) have noted, a major goal of the US state has been to promote capital accumulation, and the state's regulatory function "must not be so severe as to diminish substantially the contribution of large corporations to growth in output and employment." For example, while state regulatory agencies have been created to help protect workers (Occupational Safety and Health Administration), the environment (Environmental Protection Agency), and consumers (Consumer Product Safety Commission), these agencies will not undermine an industry's fundamental contributions to the functional requirements of the economy (Matthews and Kauzlarich 2000). In sum, while laws and regulations governing corporate behavior are plentiful, Marxist scholars will point out that the laws lack serious enforcement and associated penalties which therefore do little in the way of deterring corporations from crime. This is not a coincidence since capitalist states are more interested in protecting capital than they are the environment, workers, and consumers. It is important to note that most Marxist scholars now see the state as having some relative autonomy in that it is not always directly working with and for capitalists, but rather the long term survival and expansion of business and capital in general. As Chambliss and Zatz (993: 10) note "Should the state represent only the interests of capitalists, the conflicts will increase in intensity, with workers pitted against the state.... Were the state to side with the workers. . .the system would likewise collapse and a new social order would have to be constructed. Faced with this dilemma, officials of the state attempt to resolve the conflict by passing laws, some which represent the interests of capitalists and some the interests of workers."

Similar logic is used by Marxist or political economy critical criminologists to explain the basic causal properties of crimes of globalization. In the Marxist view, states will favor business in policy making, regulatory agency mandates, and international activities. The latter is especially important in recent decades as the process of globalization spreads capital and culture throughout the far reaches of the globe. International agencies such as the World Bank, the International Monetary Fund, and the World Trade Organization are clearly pro-capitalist, free market neoliberal agencies that see private ownership and corporate control, rather than socialist and democratic collectivist initiatives, as the main ways to link and grow economies. As such, many crimes of globalization are directly related to the economic interests of those in power rather than the welfare of indigenous populations.

The most recent, cutting-edge critical theories of crimes of the powerful have been developed under the umbrella of "crimes of empire." Michalowski (2009: 308) defines this as "the construction of development strategies, citizenship, sovereignty and culture in subordinate spaces through agreements with local elites if possible, and by force if necessary, in order to facilitate processes of capital accumulation and distributions of power that are disproportionately beneficial to a dominant cosmopolitan center." Examples of crimes of empire include many illegal military invasions, most forms of imperialism and colonialism, training and support for totalitarian regimes by capitalist states, and the suppression of human rights across the globe. Major recent analyses of crimes of empire have occurred in the context of the United States invasion of Iraq in 2003 (Kramer and Michalowski 2005). Many powerful states, like powerful corporations, are in a constant search for growth, which may include expanded territory, expedited avenues for capitalist accumulation, and increased social and cultural power and resources. The history of state crimes is replete with these interests, whether we are considering historical or more contemporary empire building by European, North American, Middle East, or African countries.

A constant thread in imperialism and "war for empire" is the use of military force not only against combatants but also civilians, which is a grave crime against humanity, and one that is often normalized in the context of patriotism or self defense in communicative discourse (Kramer 2010). Very early on non-critical scholars of traditional street crime theorized about the manner in which deviance or crime can be neutralized (Sykes and Matza 1957). Modern critical scholars of state crime in particular have found that Sykes and Matza's (1957) theory is very useful in understanding crimes of the powerful. Originally, Sykes and Matza (1957) proposed the concept of techniques of neutralization, which involves mental and social rationalizations both before and after crimes are committed. The key forms of neutralization and justification include (1) denial of responsibility, (2) denial of injury, (3) denial of victim, (4) condemnation of the condemner, and (5) appeal to higher authority. These techniques can best be understood in terms of the simple process of rationalizing one's own behavior, whether in response to cognitive dissonance, as a precondition to acting, aiding a cost-benefit analysis, or as a post-action to minimize or attempt to legitimize or neutralize the guilt of a person's behaviors or after the fact accountability. While techniques of neutralization are often used to explain occupational crimes, this approach also aids in our understanding of specific individuals within organizations.

Scholars of state crime and crime of the powerful in general have found evidence of the operationality of techniques of neutralization (Cohen 2001; Kauzlarich et al. 2000; Kramer 2010; Simon 1999; Vaughn 1996). Indeed, one of the most important things which separate state victimizers from their victims is their power to exert their will. Most often, victimizers do not acknowledge the degree to which their policies have caused harm while assessing the effectiveness of their policies to bring about desired change or to maintain their position of dominance. Unjust and injurious domestic and international policies can also be downplayed by neutralizing reasonable categorical imperatives (e.g. do no harm) by employing bankrupt utilitarianism, which may or may not be guided by ethnocentric paternalism. Often times harms are neutralized by denying responsibility, dehumanizing the powerless for purposes of exploitation, and appealing to higher loyalties (i.e. the capitalist political economy and "national security") (Kauzlarich et al. 2000; Simon 1999).

With respect to the criminal historical treatment of Native Americans within the United States, "colonists quickly justified their violence by demonizing their enemies" Takaki (1993: 43). However, the transference of one's own negative tendencies to another group is not something new. While Native Americans were seen as unruly, "God-less" savages, Takaki (1993) notes that the atrocities committed by the civilized whites against the Native Americans, were, in fact savage. It is in this light, then, that Native Americans became an enemy worthy of indiscriminate killing. In much the same manner, the indiscriminate killing of the "God-less" communists of Central America during the US imperialist project there were also justified in a similar manner.

In sum, the most longstanding and powerful critical analysis of the crimes of the powerful in critical criminology hinges on the assumption that corporations pursue goals of profit and expanded market shares and along the way cause significant legal and illegal harm and injury. States do so as well, either for empire building or the facilitation of corporate capitalism. Some critical scholars have taken issue with this and suggest that not all crimes of the powerful can be delineated down to economic motivations as this ignores religious, political, ideological, and cultural factors. Further, classic Marxist theory ignores other levels of analysis as well as obvious factors such as social controls and opportunities for crime. This has led to the development of critical integrated theories of crime. While these models are designed to explain state crime, they are also applicable to many other forms of organizational crime.

Critical Integrated Theoretical Perspectives

The earliest attempts to generate a theoretical model of state crime date back to Ronald Kramer and Raymond Michalowski's (1990) work on state-corporate crime (see Michalowski and

Kramer 2006 for a complete history). The theory was later expanded on and clarified by several other criminologists and remains the only comprehensive integrated theory for addressing organizational crimes, particularly state crimes, to date. For example, in 1998, David Kauzlarich and Ronald Kramer provided a detailed schematic of their proposed integrated theoretical frame which incorporated anomie and political economy theories at the state/structural level, organizational theory at the meso level and strain, rational choice, differential association, and routine activities at the interactional level of analysis. Kauzlarich and Kramer discuss how motivation is affected by one's socialization within that environment, the social meaning given to his or her behavior, an individual's goals, and issues of personality such as personal morality and obedience to authority.

Borrowing from Sykes and Matza (1957), reviewed above, they include techniques of neutralization as a variable of control. At the organizational level, Kauzlarich and Kramer draw heavily from organizational theorists to include instrumental rationality, role specialization, and task segregation. At the structural or institutional level of analysis, the major social institutions and social structure are included, particularly the political and economic institutions and their interrelationship as well as anomic conditions. Kauzlarich and Kramer (1998: 146) suggest the primary assumption of that perspective is that the very structure of corporate capitalism provides the impetus toward organizational crime, thus becoming crimes of capital (Michalowski 1985). They further propose that the political economy perspective stresses the shaping and/or constraining influences of the broader historical structure of a society as a factor of organizational behavior. This includes factors such as the culture of competition, economic pressure, and performance emphasis.

While this integrated model has proved useful in examining numerous cases of organizational crime (see Michalowski and Kramer 2006), the theory has been critiqued for its heavy emphasis on the criminogenic nature of capitalist social organization, which limits the theory to those crimes associated primarily with the corporate culture or within the United States. Further, Dawn L. Rothe and Christopher Mullins have suggested that the theory fails to recognize several other key factors associated with crimes of the powerful, such as weakened and transitional states, the involvement of militias, ideological and religious motivating forces, and international relations. To strengthen and expand the explanatory power of Kauzlarich and Kramer's (1998) theory, Rothe (2006, 2009) and Rothe and Mullins (2008, 2009) proposed an expanded version of the integrated model. This included expanding the levels of analysis to include the international to incorporate the increasingly international nature of organizational criminality (as recognized by global political economy theory) and by recognizing both formal controls and informal controls or constraints on organization as well as individuals.

Another difference between Kauzlarich and Kramer's model and that of Rothe and Rothe and Mullins is the inclusion of an often ignored component of some forms of crimes of the powerful – social disorganization. While much attention has been paid to organizational context and decision-making processes by scholars of organizational crime, there is a similarly rich criminological tradition which examines how social forces work within communities that are disorganized to produce criminal actions and actors. Coming out of the Chicago School of thought, social disorganization theory (Bursik and Grasmick 1993; Shaw and McKay 1942) suggests that when communities possess a diminished capacity to create and enact informal mechanisms of social control, crime rates increase. The expanded integrated model incorporates this to provide an explanatory tool for a host of crimes of the powerful including those of heads of state operating in specific conditions as well as other organizations including corporate endeavors. Others have followed this avenue by examining social disorganization in relation to private military contractors' (PMC) criminality in fields of operation (Rothe and Ross 2010). It has been noted that PMCs often operate within a disorganized environment. This not only includes the community level, but the state and organizational ones as well. After all, war-torn areas are by definition disorganized.

In sum, Rothe and Mullins use many of the theoretical concepts advanced by Kramer, Michalowski, and Kauzlarich, but expand the theory to incorporate international variables, factors that are not associated with capitalistic endeavors of corporations and states, and components of social disorganization. In addition, they propose other factors and theoretical concepts that allow for the adaptation of catalysts that may be unique to specific cases (e.g., paramilitary groups, insurgencies, militias, postcolonial conditions, and weakened or illegitimate governments).

Summary

Crimes of the powerful are complex. While there are numerous non-critical theories of crime (rational choice, routine activities theory, anomie theory, to name a few) that have been used to explain occupational and organizational crimes, the focus here has been on critical and integrated approaches. Critical theories focus on power, and in the study of white-collar crime, power is mostly considered to be synonymous with some form of capital accumulation or the practices designed to protect or expand economic resources.

Marxist or political economy theorists have been criticized for being too simple in causal logic. Surely the desire for profit or money is a strong motivation for behavior, yet not all corporations and states commit the same amount or type of crime. Moreover, it has been noted that ideological and religious motivations can be as strong as economical drives. Further, contexts for crime are important for they help explain under what particular situations crime is more or less likely to occur. This is where integrated theories have helped enrich the study of the causes of occupational and organization crime. Crime is multi-dimensional, even if it stems from the same source, and using theories that attend to all levels of analysis help us understand how organizations and people within them come to learn and practice harmful or injurious ways in the courses of their personal and professional lives.

Related Entries

▶ Corporate Crime Decision-Making
▶ Crimes of Globalization
▶ Marxist Criminology
▶ Nuclear Weapons and State Crime
▶ State Crime
▶ State-Corporate Crime

Recommended Reading and References

Almond G (1990) A discipline divided: schools and sects in political science. Sage, Thousand Oaks

Aulette JR, Michalowski RJ (1993) Fire in hamlet: a case study of state-corporate crime. In: Tunnell K (ed) Political crime in contemporary America. Garland, New York, pp 171–206

Barak G (ed) (1998) Integrating criminologies. Ashgate/ Dartmouth, Aldershot

Barnett H (1981) Corporate crime, corporate capitalism. Crime Delinquency 27(1):4–23

Bonger W (1916) Criminality and economic conditions. Little, Brown, Boston

Bonger W (1969) Criminality and economic conditions, abridged version. Indiana University Press, Bloomington

Bursik RJ Jr, Grasmick HG (1993) Economic deprivation and neighborhood crime rates 1960–1980. Law Society Rev 27(2):276–280

Chambliss W, Zatz M (1993) Making law: the state, the law, and structural contradictions. Indiana University Press, Bloomington

Cohen S (2001) States of denial: knowing about atrocities and suffering. Polity Press, Cambridge

Ezeonu I, Koku E (2008) Crimes of globalization: the feminization of HIV pandemic in Sub-Saharan Africa. The Global South, 2, 2, Africa in a Global Age, 112–129

Friedrichs D (1996) Trusted criminals: white collar crime in contemporary society. Wadsworth, New York

Friedrichs DO, Friedrichs J (2002) The World Bank and crimes of globalization: a case study. Soc Justice 29(1–2):1–12

Green P, Ward T (2004) State crime: governments, violence and corruption. Pluto Press, London

Kauzlarich D, Kramer R (1998) Crimes of the American nuclear state: at home and abroad. Northeastern University Press, Boston

Kauzlarich D, Matthews R, Miller W (2000) Toward a victimology of state crime. Crit Criminol 10:173–194

Kramer R (1982) Corporate crime: an organizational perspective. In: Wickman P, Daily T (eds) White collar and economic crime. Lexington Books, Lexington, pp 75–94

Kramer RC (1992) The space shuttle challenger explosion: a case study of state-corporate crime. In: Schlegel K, Weisburd D (eds) White collar crime reconsidered. Northeastern University Press, Boston, pp 214–243

Kramer R (2010) Resisting the bombing of civilians: challenges from a public criminology of state crime. Soc Justice 36(3):78–97

Kramer R , Michalowski R (1990) Toward an integrated theory of state-corporate crime. Presented at the American Society of Criminology, Baltimore

Kramer RC, Michalowski RJ (1991) State-corporate crime: case studies in organizational deviance. Unpublished manuscript

Kramer R, Michalowski R (2005) War, aggression, and state crime: a criminological analysis of the invasion and occupation of Iraq. British J Criminol 45:446–469

Matthews R, Kauzlarich D (2000) The crash of Valujet flight 592: a case study in state-corporate crime. Sociol Focus 33(3):281–298

Michalowski RJ (1985) Order, law and crime. Random House, New York

Michalowski RJ (2008) Power, crime and criminology in the new imperial age. Crime Law Soc Change 51:303–325

Michalowski R (2009) Power, crime and criminology in the new imperial age. Crime Law Social Change 51(3–4):303–325(23)

Michalowski RJ, Kramer R (1990) Toward an integrated theory of state-corporate crime. Paper presented at the American Society of Criminology, Baltimore

Michalowski RJ, Kramer R (eds) (2006) State-corporate crime: wrongdoing at the intersection of business and government. Rutgers University Press, Piscataway

Pearce F (1976) Crimes of the powerful. Pluto Press, London

Rothe DL (2006) The Masquerade of Abu Ghraib: state crime, torture, and international law. PhD dissertation, Department of Sociology, Western Michigan University, Kalamazoo

Rothe DL (2009) The crime of all crimes: an introduction to state criminality. Lexington/Roman and Littlefield, Lanham

Rothe DL (2010) Facilitating corruption and human rights violations: the role of international financial institutions. Crime, Law Soc Change 53(5):457–476

Rothe DL, Kauzlarich D (2010) State-level crime: theory and policy. In: Barlow HD, Scott D (eds) Crime and public policy: putting theory to work, 2nd edn. Temple University Press, Philadelphia, pp 166–187

Rothe DL, Mullins CW (2008) Genocide, war crimes and crimes against humanity in Central Africa: a criminological exploration. In: Haveman R, Smeulers A (eds) Supranational criminology: towards a criminology of international crimes. Intersentia, Antwerp, pp 135–158

Rothe DL, Mullins CW (2009) Toward a criminology of international criminal law: an integrated theory of international criminal law violations. Int J Comp Appl Crim Justice 33(1):97–118

Rothe DL, Ross JI (2010) Private military contractors, crime, and the terrain of unaccountability. Justice Q 27(4):593–617

Rothe DL, Mullins CW, Muzzatti S (2006) Crime on the high seas: crimes of globalization and the sinking of the Senegalese Ferry Le Joola. Crit Criminol Int J 14(2):159–180

Rothe DL, Mullins CW, Sandstrom K (2009) The Rwandan Genocide: international finance policies and human rights. Soc Justice 35(3):66–86

Shaw C, McKay HD (1942) Juvenile delinquency and urban areas. University of Chicago Press, Chicago

Simon D (1999) Elite deviance, 6th edn. Allyn and Bacon, Boston

Sutherland E (1949) White collar crime. Holt, Rinehart and Winston, New York

Sykes G, Matza D (1957) Techniques of neutralization: a theory of delinquency. Am Sociol Rev 22:664–670

Takaki R (1993) A different mirror: a history of multicultural America. Little Brown, Boston

Vaughan D (1982) Toward understanding unlawful organizational behavior. Mich Law Rev 80:1377–1402

Vaughan D (1992) The macro–micro connection in white-collar crime theory. In: Schlegel K, Weisburd D (eds) White collar crime reconsidered. Northeastern University Press, Boston, pp 124–145

Vaughn D (1996) The challenger launch decision: risky technology, culture, and deviance at NASA. University of Chicago Press, Chicago

Crimes Without Victims

► Moral Crimes

Criminal Activity of Sex Offenders

► Criminal Career of Sex Offenders

Criminal Career

► Desistance from Crime

Criminal Career of Sex Offenders

Patrick Lussier
Laval University, Quebec City, QC, Canada

Synonyms

Criminal activity of sex offenders; Development of sex offending

Overview

The interest for the criminal career is not new and several commentaries and observations about sex offenders' criminal activity have been made for quite some time. Most of these commentaries and observations were focused on the same underlying questions, that is, sex offenders' dangerousness. The issue of dangerousness has been addressed by examining sex offenders' likelihood of sexual recidivism using different methodologies. These descriptive studies of sex offenders' criminal records were not supported by an organizing conceptual framework which led to the emergence of controversies among researchers about sex offenders' nature and extent of their criminal behavior. These controversies certainly did not help to challenge common myths and false beliefs about sex offenders' criminal behavior which have, in some instances, serve as the foundation to develop new criminal justice policies to tackle the problem of sexual violence and abuse. The current study aims to introduce the criminal career approach and, in doing so, aims to provide a common organizing framework for policy makers as well as researchers from various disciplines such as psychology, psychiatry, sociology, social work, criminal justice, and criminology. Although there is a long history of criminal career research with the publication of *Criminal Careers and Career Criminals* in 1986 by Dr. Al Blumstein and colleagues, such a framework was introduced to the field of criminal justice and criminology. The criminal career approach is concerned with the description and explanation of the longitudinal sequence of offending. While the criminal career approach has been around for quite some time in criminological circles, it would take some time, however, before this framework would be introduced more explicitly to the field of sexual violence and abuse (Blokland and Lussier 2012; Lussier et al. 2005). Building on the criminal career approach proposed by Blumstein and colleagues, the current review examines the current state of knowledge regarding the criminal activity of sex offenders. While the criminal career approach should not be seen as a cure-all approach, it provides a conceptual framework to organize existing study findings, to guide future empirical research, as well as to help to think more clearly about sex offenders' criminal behavior. Therefore, this chapter aims, first, to introduce researchers from the field of sexual violence and abuse to the criminal career approach; second, to organize the empirical knowledge on the criminal activity of sex offenders using a criminal career approach; and, third, to review the state of empirical knowledge on various dimensions of the criminal career of sex offenders.

The Criminal Career Approach

Over the past five or six decades, no other offender type has been under more scrutiny from researchers than sex offenders have been. Researchers from the field of sexual violence and abuse have measured a wide range of factors to describe individuals having committed sex crimes, such as family background, victimization experiences, exposure to deviant models, attachment bonds, parental practices, childhood behaviors, psychiatric symptoms, personality traits and disorders, intelligence and cognitive skills, sexual arousal and interests, mood and temperament, cognitive distortions, sexual development and sexual behaviors, coping skills and coping strategies, and pornography use, to name just a few. One striking observation that can be made is that while sex offenders have been described in so many ways along so many dimensions and

Criminal Career of Sex Offenders, Table 1 Descriptive summary of criminal career parameters

Parameters	Definition
Prevalence	The proportion of a given population committing a crime during a specific time period
Age of onset	The age at first offense
Frequency (volume)	The number of crimes committed
Lambda	The number of crimes committed taking into account the time at risk (i.e., excluding periods during which the offender did not have the opportunity to offend (i.e., hospitalization, incarceration, death))
Continuity	The passage from juvenile offending to adult offending
Career length	The length of time between onset and termination of offending
Versatility (diversity)	The number of different crime types committed
Seriousness	The degree of gravity of the criminal activity
Specialization	The tendency to limit offending to one particular form of crime
Desistance	Termination of offending

factors, comparatively speaking, the very behavior that clinical researchers aimed to explain, sexual offending, has been largely neglected. This is not an overlooking but illustrates the fact that most theoretical views of sex offending is based on the assumptions that there is a stable propensity to commit sex crime and theoretical models should only be concerned by the description and the explanation of this propensity. These models, therefore, do not recognize the importance of distinguishing such aspects as prevalence, age of onset, persistence, frequency, seriousness, and desistence. This trait-like approach is not well suited to describe and explain why offenders start or stop offending and whether the same factors explain both. Some theoretical models do make distinction for various offending stages but have been limited to a few aspects of sex offending, such as onset and persistence (e.g., Marshall and Barbaree 1990). The criminal career approach provides a framework to think about how sexual offending starts, develops over time, and stops and whether such distinctions are theoretically, clinically, as well as policy relevant (Table 1).

The Onset of Sex Offending

The age of onset of sex offending refers to the age at which sex offending is initiated. The age of onset is particularly interesting because it marks the origins of the behavior and allows examining why and under what circumstances the criminal behavior was initiated. The age of onset has been discussed in several empirical studies on sex offenders, but its operationalization has not always been clear and straightforward. In earlier investigations, clinical researchers have been concerned with the age of onset of sexual problems of adult offenders. Using the term sexual problems is problematic because it encompasses behaviors such as the onset of deviant sexual arousal, the onset of deviant sexual fantasizing, the onset of deviant sexual behaviors, as well as the onset of sexual offending. Early theoretical models put much great emphasis on the role of deviant sexual fantasies as a precursor to sexual offending. Clinical research, however, has shown that only a small proportion of adult offenders report deviant sexual fantasies and/or a paraphilia, and an even fewer proportion of them report having experienced such deviant fantasies prior their offending (Marshall et al. 1991). This reinforces the importance of distinguishing deviant fantasies, deviant sexual behaviors, and sex offending.

Early studies looking at the onset of sex offending has described adult sex offenders as grown-up juvenile sex offenders. For example, in the Prentky and Knight study (1993), 49 % of their sample of adult rapists reported an onset prior age 18, while the rest of the sample reported an onset in adulthood. These results mirror those reported in the Groth et al. study (1982), which showed an average age of onset of 19 years old for a sample of sexual aggressors against women, while in the Abel et al. (1993) study, it was 22 years old. The self-reported onset age for child molesters appears to be different than the one reported for rapists, but the findings are not stable across studies. In the Prentky et al. (1993) study, whereas 49 % of adult rapists were JSOs, that number increased to 62 % for child molesters. Therefore, given these results, one could expect that the average onset age for child molesters would be younger than the one reported for rapists. This is not the case and this could be attributable to sampling differences. More precisely, these earlier studies showed some discrepancies across child molester types. To illustrate, in the Marshall et al. (1991) study, the self-reported age of onset was 24 years old for extrafamilial offenders against boys, 25 years old for extrafamilial offenders against girls, and 33 years old for incestuous fathers. Similar numbers were reported by Smallbone and Wortley (2004) suggesting that the onset of extrafamilial child molestation starts sooner than the onset of intrafamilial child molestation. These differences may be explained by opportunity structure of the offense as one needs to have a biological child to offend against him or her. These studies used retrospective data to estimate the age at which adult sex offenders started their offending behavior. From these self-report studies (see, e.g., Abel et al. 1993), however, it is not always clear whether the onset refers specifically to sex *offending* or to some other behavior such as the onset of deviant sexual interests, the onset of deviant sexual fantasizing, and the onset of deviant sexual arousal.

Not surprisingly, the age of onset based on self-report data is younger than those based on official data (e.g., Gebhard et al. 1965; Lussier et al. 2005; Smallbone and Wortley 2004). When looking at the official age of onset, results clearly indicate that it significantly varies across sex offender types. Reports suggest that sexual aggressors of women are typically charged for a first offense in their late twenties, while for sexual aggressors of children, it is typically in their late 30s. There is a gap between the age of onset reported in self-report studies with adult sex offenders and those found in studies based on police data. That gap, however, is relatively unknown given that self-report and official data on the age of onset are not typically analyzed in the same study limiting the conclusions that can be drawn. Further, the utility of other sources of information, such as the police report and the victim statement, has not been examined in prior research. Lussier and Mathesius (2012) provided evidence that the official age of onset, as measured with criminal justice data, provides a distorted view on the actual onset of offending, at least for some sex offenders. Their claim was based on the observation that official data of offending does not take into consideration the offender's ability to avoid and/or delay detection. In their study, official data, police data, and victim's account were analyzed to compare and contrast the official and actual age of onset. On average, it was found that there is a gap of about 7 years between actual and official age of onset in sex offending. The findings showed that the gap between actual and official onset was much more important for child molesters, more specifically incestuous and pseudo-incestuous fathers. These findings may suggest that victims may take significantly longer to report the crime to the authorities (if they do) when the crime is committed by a parental figure. For the most part, while the actual age of onset does not vary across sex offender types, it does for the official age of onset, suggesting differential investment in detection avoidance across offenders. Further, findings show that close to 20 % of sex offenders have already desisted or are in the process of desisting by the time they are first charged for their sex crime.

Frequency of Sex Offending

Frequency of sex offending refers to the number of sex crimes committed. This refers to the offender's volume of crimes. It is interesting to note that researchers have not spent much time describing this aspect of sex offender's offending. Like many other crime types, measuring frequency of sex offending is not as straightforward as it may seem. This may be due to the nature of the offending behavior and various offending strategies adopted by sex offenders (Lussier et al. 2011). As such, frequency of sex offending may refer to the number of victims an individual has offended against. Frequency can also refer to the number of sex crime event, or the number of times an individual has sexually offended against a person. Some offenders may adopt an offending strategy in which different victims are offended against on a very limited number of occasions (once or twice). This victim-oriented strategy may characterize offenders who offend against strangers. Other offenders may decide to limit the number of victims they offend against choosing instead the maximization of a single offending opportunity by reoffending on multiple occasions against the same person. This strategy may be referred to as the event-oriented strategy and characterizes incest and pseudo-incest offenders as well as offenders offending against their partner. It is difficult to estimate the frequency of offending using available data found in past research because too often criminal justice indicators are used, such as the number of arrests (or convictions) for a sex crime. Such indicators may not reflect well the actual behaviors that lead to the arrest. For example, a sex offender may have been convicted once for sex crime which involves the abuse of a child over a 5-year period in which the victim was molested on more than 500 occasions. Relatedly, an individual may also have only one conviction for a sex crime that involves the rape of three women, each raped on a single occasion. Frequency can also be influenced by the time at risk or the time an offender was at risk of perpetrating a crime (e.g., not incarcerated, not hospitalized, not dead). As such criminal career researchers have used the term lambda or annual frequency of offending that takes into account the time at risk.

Empirical studies have shown much heterogeneity in the frequency of sex offending across sex offenders. In the Lussier et al. (2011) study using a sample of convicted adult male sex offenders, the average number of victim was 1.8, but the number of victims varied between one and 13. Similarly, Groth et al. (1982) reported that the number of self-reported sex crimes reported by their sample of sex offenders varied between one and 30. Even greater variance was found in the Weinrott and Saylor (1991) study where the range of self-reported victims varied between one and 200. Heterogeneity in the volume of sex offending can also be found across sex offender type. Such discrepancies between aggressors of women and children using official data were not found in the Weinrott and Saylor (1991) study. They found that on average, sexual aggressors of women had 1.8 victims, while it was 2.0 for child molesters. Similarly, Groth et al. (1982) using self-reported data found that rapists and child molesters had a similar average number of sex crimes (about five). This may have something to do with sample composition. Pham et al. (1999) also reported that their group of intrafamilial offenders had a mean number of 1.6 victims, much lower than what was reported for extrafamilial child molesters. Indeed, intrafamilial child molesters tend to have a significantly lower number of victims (Abel et al. 1987). Intrafamilial offenders' lower average number of victims may be a function of structuring opportunity factors. Indeed, the average number of victims corresponds to the number of children typically found in contemporary western-country families. This may suggest that these offenders are less likely to offend outside the family setting. It may well be also that incestuous fathers are seeking event-oriented opportunity where they can maximize the number of offending opportunities against the same victim. It could be, therefore, that the sample used in the Weinrott and Saylor (1991) study included a high proportion of intrafamilial child molesters.

Weinrott and Saylor (1991) study findings are also insightful about the sex offenders' volume of sex offending because their study also used self-reported data. In other words, they asked these men to report their number of victims. Hence, while sexual aggressors of women had, on average, about two official victims, these men reported, on average, having offended against close to 12 victims. In other words, the mean number of self-reported victims was more than five times the number of official victims. It should be noted, however, that the median number of victims was 6, or three times the average number of official victims. The median is much lower than the mean, most probably because of the presence of small group offenders with a disproportionately high number of victims. The discrepancies between official and self-report data were also observed for child molesters. If their sample of child molesters had, on average, offended against two victims, these men reported having offended, on average, against seven victims. Similar numbers were reported by (Marshall et al. 1991) as well as by Groth et al. (1982). Nonetheless, the number of self-reported victims reported in the Weinrott and Saylor study is a far cry from those reported by Abel et al. (1987). The Abel et al. study reported that extrafamilial offenders against girls had about 20 victims on average, while for extrafamilial offenders against males, it was 150. Note that the median numbers of victims for the two groups were 1 and 4, suggesting that 50 % of their sample of extrafamilial offenders against girls self-reported only one victim, while 50 % of the sample of extrafamilial offenders against boys reported no more than four victims. Perhaps, for some reasons, a small group of offenders included in the Abel et al. (1987) exaggerated the volume of their offending. Hence, while Abel et al.'s (1987) mean number of victims looks impressive, the inspection of the median suggests that most child molesters, to the exception of those offending against boys outside the family setting, typically offend against a single victim. This is not meant to say that there are no prolific sex offenders.

Lussier et al. (2011) investigated the presence of prolific offenders in a sample of adult males convicted for a sex crime. Using several sources of information (self-report, police investigation, victim statement), the frequency of offending (and lambda) was estimated. The study was insightful about the presence of prolific sex offenders for several reasons. The authors examined frequency in terms of the number of sex crime events offenders had been involved in. The study findings highlighted that about 11 % of their sample had been involved in over 300 sex crime events as opposed to about 40 % who had been involved in only one crime event. There were no differences in the number of different victims offenders of both groups had offended against, suggesting that maximizing the number of victims is independent of the offender's decision to maximize the number of crime events. In other words, some offenders take advantage of low-risk short-term opportunities with different victims, while others opt to benefit from a single offending opportunity by repeatedly offending against the same victim over and over. The findings revealed that the most prolific sex offenders were older and had a more conventional background characterized by a stable relationship with an adult partner, a job at the time of the offense(s), no drug issues, and no prior record for a sex crime. The convention image of the prolific sex offender contrasts with the typical image of the chronic offender stemming from empirical research with prison populations of general offenders as someone who is young, single, unemployed, has significant alcohol and/or drug issues, and a lengthy criminal record. The findings of (Lussier et al. 2011) also contrast with the media portrayal of the "sexual predator."

Continuity in Sex Offending

Continuity refers to the persistence of sex offending from adolescence to adulthood. Continuity, therefore, informs about the proportion of juvenile sex offenders who become adult sex offenders as well as the proportion of adult sex offenders who were juvenile sex offenders. The continuity hypothesis stipulates that today's juvenile sex offenders are tomorrow's adult sex offenders, and several policies are based on this assumption. The discontinuity however suggests that most juvenile sex offenders do not go on to become adult sex offenders and that, for the most part, juvenile sex offenders' offending behavior is limited to the period of adolescence. The lesson learned from criminal career research is that retrospective data with samples of adult offenders tend to overestimate continuity of offending because such studies fail to include youths who have desisted from offending in adolescence. It is well accepted by criminologists that there is much discontinuity in antisocial behavior, but antisocial personality disorder in adulthood virtually requires antisocial behavior in youth. Whether this conclusion applies also to sex offending is unclear. It has in fact been argued that JSOs do not continue their sexual offending in adulthood. While this is certainly true for most JSOs, empirical studies indicate that a small fraction (between 5 % and 10 %) may indeed continue their sexual offending in young adulthood (e.g., Lussier et al. 2012). The proportions of JSOs continuing in adulthood gradually increase with a longer follow-up period in adulthood (between 10 % and 15 %) (e.g., Hagan et al. 2001). Bremer (1992) reported a 6 % reconviction rate in a sample of serious JSO, but the recidivism rate rose to 11 % when based on self-reports. Therefore, while the use of official data underestimates recidivism rates, it is unlikely to be able to explain the fact that the vast majority are not rearrested or caught again for a sex crime. Taken together, there appears to be sizeable discontinuity in sexual offending from adolescence to adulthood superimposed on a little continuity. Hence, while aggregate data indicate that the overwhelming majority of juveniles do not persist their sexual offending in adulthood, it does not inform practitioners and policy makers about those few juvenile offenders who do persist.

Versatility in Sex Offending

Versatility refers to offenders' tendency to commit a wide array of offenses. Criminal versatility is concerned with the crime-switching patterns of offenders as offending persists over time. Two types of crime-switching patterns can be distinguished. First, crime-switching can be analyzed in the context of sex offenders' whole criminal activity. Hence, from that perspective, researchers are concerned with the number of different types of non-sex crimes sex offenders are involved in. An underlying theme would be to examine the crime mix, or the nature of the different types of criminal activities characterizing sex offenders' criminal repertoire. The scientific literature on the criminal versatility of sex offenders has been described in details elsewhere (e.g., Lussier et al. 2005). Secondly, crime-switching can also be examined strictly in the context of sexual offending. This would refer to sex offenders' tendency to limit themselves to one form of sex crimes. Within the context of sex offending, crime-switching patterns can occur along several dimensions such as victim's age, gender, relationship to the offender, nature of sexual acts committed by the offender, and level and type of coercion used. Others have also used the term sexual polymorphism to describe a sexual criminal activity characterized by much versatility (see Guay et al. 2001; Lussier et al. 2008). Few studies have examined the level of sexual polymorphism and crime-switching patterns in the sexual criminal activity of persistent sex offenders. Based on the current state of knowledge, there are three broad conclusions that can be drawn in regard to the offending pattern of persistent sexual offenders.

First, Soothill and colleagues came to the conclusion that while sex offenders are generalists in their criminal offending, they tend to specialize in their sexual offending, confining themselves to one victim type (Soothill et al. 2000). Similarly, (Radzinowicz 1957) also found specialization in victim choice in that only 7 % of his large sample of sex offenders had convictions for crimes against both male and female victims, a finding consistent with those of Gebhard et al. (1965). More recently, Cann et al. (2007) found that only about 25 % of their sample of incarcerated sex offenders was versatile when considering victim's age and gender as well as the offender-victim relationship. On the other hand, crime-switching patterns may vary as a function of the dimension of the sexual polymorphism considered. For instance, while they also found much stability as to the victim's gender, Guay et al. (2001) reported considerable versatility for those targeting adolescents. While offenders targeting children and those targeting adults remained in the same category, those offending against adolescents were likely to switch either to adults or to children. (Guay et al. 2001) hypothesized that adolescents may be a sex surrogate choice when the preferred partner was not available.

Empirical studies conducted in clinical settings have shown a divergent picture of the sex offenders' crime-switching pattern. Weinrott and Saylor (1991) as well as Heil et al. (2003) have argued that official data hide an enormous amount of sex crimes. Using official data only, Weinrott and Saylor (1991) found that only 15 % of their sample of offender was versatile considering only three categories: adult female, extrafamilial children, and intrafamilial children. Using a self-reported computerized questionnaire, however, that number rose to 53 %. Similarly, (Heil et al. 2003) reported that incarcerated offenders in treatment are not versatile as to victim's age (7 %) and gender (8.5 %) when assessed with official data but are when interviewed using a polygraph (70 % and 36 %, respectively). Less dramatic numbers were reported for parolees which might be explained by sampling differences (i.e., incarcerated offenders were more serious offenders) and the fact that admitting a crime was a prerequisite to enter treatment. Abel's well-known study conducted under strict conditions of confidentiality showed that 42 % of their sample targeted victims in more than one age group, 20 % targeted victims of both gender, and 26 % committed both hands-on and hands-off crimes (Abel and Rouleau 1990). Similar results have been reported elsewhere in a sample of sex offenders assessed in a forensic psychiatric institution (Bradford et al. 1992).

The overlapping nature of different forms of sexually deviant acts found in the clinical studies is counterintuitive to current typological models of sex offenders based on the characteristics of the offense. The victim's gender, the victim's age, the offender-victim relationship, the level of sexual intrusiveness, and the level of force used during the commission of the crime are some examples of criteria that have been used over the years to classify sex offenders (e.g., Gebhard et al. 1965; Knight and Prentky 1990). Smallbone and Wortley (2004) came to the conclusion that diversity in paraphilic activities may be a function of general deviance. Indeed, looking at different activity paraphilia (e.g., voyeurism, frotteurism, sexual sadism) in a sample of child molesters, they found that a scale measuring the versatility of sexual deviance correlated significantly and positively with nonsexual offending. In other words, as the frequency of offending increases, so does the versatility in paraphilic interests and behaviors. Similarly, (Lussier et al. 2005) found in a sample of adult sex offenders that versatility in sex offending was strongly related to versatility in nonsexual nonviolent offending as well as versatility in nonsexual violent crime. Furthermore, using structural equation modeling, they found that such pattern of general versatility was related to an early onset and persistence of antisocial behavior. In other words, sex offenders characterized by a life-course-persistent antisocial tendency were more likely to show much versatility in their sexual offending. In that regard, (Guay et al. 2001) hypothesized that crime-switching in sexual offending might be partly explained by low self-control.

Lussier et al. (2008) analyzed crime-switching patterns of a sample of convicted adult sex offenders using transition matrices and diversity indexes. They observed that crime-switching in sex offending is multidimensional in that diversity index tends to be relatively independent from one another. In other words, if an offender offends against victims from different age groups, it does not imply that this person will offend against both males and females. Therefore, contrary to Abel and Rouleau (1990) conclusions, little evidence was found suggesting that sex offenders are sexual deviates offending against different types of victims in different contexts. Furthermore, the study findings highlighted that crime-switching patterns vary across dimensions of sex crimes. On the one end of the continuum, victim's gender and level of physical force are relatively stable across crime transitions. The notion of preference is relevant and of importance in the understanding of persistence in sexual offending. The study findings were in line with those of Soothill et al. (2000) suggesting that some specialization can be found for certain aspects of sex offending and that some aspects of offending are far from being random upon certain situational contingencies. On the opposite end of the continuum, the victim's age and sexual intrusiveness involve more crime-switching. In other words, sex offenders are not prone to switching from males to female victims (and vice versa) or to change their level of violence across offenses. It is the nature of sexual acts (hands-on, hands-off, oral sex, penetration) and victim's age that tend to fluctuate the most across the longitudinal sequence of sex crimes committed by sex offenders. The concept of sex surrogate might also play a part in stimulating crime-switching (Guay et al. 2001). This appears to be especially true for those having offended against adolescent victims, who might represent the second best option in the absence of the preferred victim type (i.e., children or adults). This situation appears to be true for both child molesters and rapists. Of importance, and in keeping with the sex surrogate hypothesis, is that very few child molesters also offended against adults and vice versa. Finally, the study highlighted that crime versatility in sex offending tends to increase as a function of persistence of sex offending, especially for victim's age, offender-victim relationship, and sexual intrusiveness. Hence, the more sex offenders offend against different victims, the more their sexual criminal repertoire will diversify along those dimensions. This might partly explain discrepancies reported in earlier studies as clinical samples including more serious and persistent offenders should report more evidence of crime-switching.

Specialization in Sex Offending

Crime specialization is another important aspect of the criminal career of sex offenders. Various definitions of crime specialization have been proposed over the years. Criminal career researchers have generally defined specialization as the probability of repeating the same type of crime (Blumstein et al. 1986). Crime specialization is important from a crime control perspective for obvious reasons. If offenders do specialize in a particular crime type, then it gives support to the implementation of crime prevention strategies targeting known offenders involved in such crime type. The sex offender registry is an example of assuming that sex offenders are sex crime specialists. Recording personal information in a database helping to track down convicted sex offenders is considered useful from a law enforcement standpoint. It can be used as a tool to prioritize suspect by assisting in the criminal investigation of new cases of sexual assault and abuse. According to the specialization hypothesis, if the criminal activity of a sexual offender persists, it would be primarily in sexual crime. Crime specialization is understood as the tendency for some crime to involve a higher or lower level of repetition over time. In that context, researchers may be interested in comparing whether specialization in sex crime is similar or different than specialization in burglary, drug-related offenses, driving under the influence, auto-theft, etc. The concept of crime specialist, on the other hand, refers to individuals with a higher probability of repeating the same crime over time. In that context, researchers may be interested in determining the proportion of sex crime specialists among sex offenders and their characteristics. Specialization can be studied in two contexts: (a) whether sex offenders tend to specialize in the type of sex crime they commit (e.g., Soothill et al. 2000; Lussier et al. 2008) and (b) whether sex offenders tend to specialize is sex crimes when considering their whole criminal activity. The latter has been the subject of closer empirical scrutiny by researchers.

Lussier (2005) reviewed the scientific literature on crime specialization in sex offending and concluded that empirical studies have provided little empirical evidence supporting the specialization hypothesis (see also Simon 1997). More precisely, sex offenders do not limit themselves to sex crimes, quite the contrary. This is not to say that all sex offenders are criminally versatile or that sex offenders' criminal record always includes other crime types. Sex offenders who have a persistent criminal activity tend to be involved in other crime types that are not sexual in nature. This conclusion, however, requires closer scrutiny. Recidivism studies do show that sex offenders have a greater likelihood of being rearrested for a sex crime than non-sex offenders are. For example, in the (Langan et al. 2003) study, using a 3-year follow-up of close to 270,000 prisoners, 5 % of ASOs were rearrested for a sex crime compared to 1 % for non-sex offenders. These results do not take into account the fact that the recidivism rates significantly vary across sex offenders. For example, Quinsey et al. (1995) showed that for comparable follow-up periods, the sexual recidivism rate of incest offenders was 8 % as opposed to 18 % for child molesters offending against girls and 35 % for child molesters offending against boys.

Recidivism studies are impacted measures of crime specialization because it does not consider the whole criminal activity of an offender but only two successive crimes. Others have used transition matrices and have reported on the probabilities of being rearrested for a sex crime while taking into account the offender's entire criminal career. Results have shown that crime specialization is much lower for rape (e.g., Blumstein et al. 1988) than when using a broader definition of sex crimes that includes child molestation (e.g., Stander et al. 1989). This suggests that child molesters are more likely to specialize in sex crimes than rapists. Such conclusion seems to be reinforced by the analysis of their criminal record and the importance of sex crimes in their entire criminal activity. Indeed, studies have shown that sex crimes represent about 4–14% of the entire criminal activity of rapists, while it is about 40 % for child molesters (Gebhard et al. 1965; Lussier et al. 2005). These numbers should be seen as tentative given the small number of studies having examined specialization in sex crime using a ratio of sex crimes to all crimes committed by sex offenders.

Desistance from Sex Offending

The study of the criminal career of sex offenders is still in its infancy. Not surprisingly, therefore, several dimensions of the criminal career of sex offenders have not been subject of much empirical research. Of importance, patterns of escalation and de-escalation in sex offending have been largely overlooked. Another key criminal career dimension that has been overlooked until recently is desistance (e.g., Kruttschnitt et al. 2000; Laws and Ward 2011). Desistance refers to the termination of the criminal career and can be understood here as the termination of sex offending. The concept of desistance is important to describe the age at which sex offenders stop their sex offending; it allows us to determine the length of sexual offending, whether desistance from sex offending is accompanied by desistance in other crime types. Criminal career researchers understand desistance as a discrete event (the period after which offending has stopped). Developmentalists, however, understand desistance as a dynamic process by which offending slows down and becomes more specialized until complete termination. Studies having discussed the issue of desistance in the context of sex offending have generally relied on sexual recidivism indicators to determine desistance, that is, desistance is implied for the absence of a new charge or conviction during the follow-up period. This approach is somewhat misleading because with a longer follow-up period, offender considered to be desistors may become sexual recidivists. Recidivism is also problematic in the context where desistance is seen as a process as the presence of a new conviction for a sex crime may not inform about whether or not offending is less frequent and less serious over time.

In the field of sexual violence and abuse, desistance has generally been discussed in terms of the presence (or absence) of an age effect on the risk of sexual recidivism (Lussier and Healey 2009). In other words, it is possible that with age, aging, and the passage of time, sex offender's propensity to commit a sex crime changes. Researchers generally agree on the recidivism rates of the younger adult offenders and older offenders, but there is controversy about the age effect occurring for other offenders. Three main points have been at the core of the debate about the link between aging and reoffending in adult offenders: (a) identification of the age at which the risk of reoffending peaks, (b) how to best represent the trend in risk of reoffending between the youngest and the oldest group, and (c) the possibility of differential age-crime curves of reoffending. One hypothesis states that, when excluding the youngest and oldest group of offenders, age at release and the risk of sexual recidivism might be best represented by a plateau. Thornton (2006) argued that the inverse correlation revealed in previous studies may have been the result of the differential reoffending rates of the youngest and oldest age groups, rather than a steadily declining risk of reoffending. In this regard, one study presented sample statistics suggesting a plateau between the early 20s and the 60s+ age groups (Langan et al. 2003) No statistical analyses were reported between the groups, thus limiting possible conclusions for that hypothesis. Another hypothesis suggested there might be a curvilinear relationship between age at release and sexual recidivism, at least for a subgroup of offenders. (Hanson 2002) found evidence of a linear relationship for rapists and incest offenders, and a curvilinear relationship was found for extrafamilial child molesters (see also Prentky and Lee 2007). Whereas the former two groups showed higher recidivism rates in young adulthood (i.e., 18–24), the latter third group appeared to be at increased risk when released in the subsequent age bracket (i.e., 25–35). This led researchers to conclude that, although rapists are at highest risk in their 20s, the corresponding period for child molesters appears to be in their 30s. These results, however, have been criticized on methodological grounds, such as the use of small samples of offenders, the presence of a small base rate of sexual reoffending, the use of uneven width of age categories to describe the data, the failure to control for the time at risk after release, and the number of previous convictions for a sexual crime (Barbaree et al. 2003; (Thornton 2006). The controversy over the age effect led researchers to question whether risk assessors should consider the offender's age at the time of prison release and, if so, how the adjustment should be done (Barbaree et al. 2007; Harris and Rice 2007).

Subsequently, two prominent schools of thought emerged, and two main hypotheses have been used to describe and explain the roles of propensity, age, and reoffending in sexual offenders: (a) the static-maturational hypothesis and (b) the static-propensity hypothesis. The static-maturational hypothesis suggests that sex offenders' risk of reoffending is subject to a maturation effect, as this risk typically follows the age-crime curve (Barbaree et al. 2007; Lussier and Healey 2009). Importantly, the maturation hypothesis is based on the assumption of a stable propensity to reoffend, but the offending rate can change over life course. In other words, the rank ordering of individuals (between-individual differences) on a continuum of risk to reoffend remains stable, but the offending rate decreases (within-individual changes) in a similar fashion across individuals. It was determined that the offender's age at release contributes significantly to the prediction of reoffending, over and above scores of various risk factors said to capture sex offenders' propensity to reoffend. Multivariate analyses showed that when controlling for prior criminal history, the rate of sexual reoffending decreases by about 2 % for every 1-year increase of the offender's age at release (Thornton 2006). Adjusting for sociodemographic and criminal history factors, (Meloy 2005) replicated this finding for probation failure and for nonsexual reoffending, but not for sexual reoffending. This could be explained by the low base rate of sexual reoffending for this sample (i.e., 4.5 %). Other studies indicated that age at release contributes significantly to the prediction of reoffending, even after adjusting for actuarial scores (e.g., Barbaree et al. 2003). Similar to (Thornton 2006; Hanson 2006) reported that after adjusting for the scores on Static-99, the risk of sexual reoffending decreased by 2 % for every 1-year increase in age after release. No interaction effects were found between scores of the Static-99 and age at release. Though these preliminary results provide evidence in favor of the maturational hypothesis, many questions remain unanswered. The key question is whether sex offenders identified as high risk are also subject to an age effect. Because previous studies did not test the maturational hypothesis separately for high-risk offenders and considering that high-risk sex offenders constitute only a small minority of all convicted sex offenders, researchers might have been limited in finding a differential age effect.

The static-propensity hypothesis suggests that, by using historical and relatively unchangeable factors, adult sex offenders can be distinguished based on their likelihood of reoffending. The main assumption is that criminal propensity is stable over life course, and therefore, risk assessment tools should only be used for measuring the full spectrum of this propensity. An important point of contention for the static-propensity hypothesis is whether younger offenders at high risk to reoffend show the same or similar recidivism rates as older offenders with the same risk to reoffend. According to the static-propensity hypothesis, older offenders with high scores on risk assessment tools represent the same risk of reoffending as younger offenders with similar scores (e.g., Harris and Rice 2007). For static-propensity theorists, the only age factor that risk assessors should include are those reflecting a high propensity to reoffend, such as the age of onset of the criminal activity. For example, Harris and Rice (2007) argued that the effect of aging on recidivism is small. In fact, they argued that age of onset is a better risk marker for reoffending than age at release. In other words, those who start their criminal career earlier in adulthood show an increased risk of reoffending. Their findings showed that the offender's age at release did not provide significant incremental predictive validity over actuarial risk assessment scores (i.e., VRAG) and age of onset. This could be partly explained by the fact that age of onset and age at release were strongly related, that is, early-onset offenders are more likely to be released younger than late-onset offenders. The high covariance between these two age factors might have limited researchers in finding a statistical age at release effect in multivariate analyses. Furthermore, (Barbaree et al. 2007) found that, after correcting for age at release, the predictive accuracy of actuarial tools significantly decreased, suggesting that an age effect was embedded in the risk assessment score. Actuarial tools have developed by identifying risk factors that are empirically linked to sexual reoffending. If the risk of reoffending peaks when offenders are in their 20s, it stands to reason that characteristics of this age group are most likely to be captured and included in actuarial tools. Consequently, scores of risk assessment tools might be more accurate with younger offenders but overestimate the risk of older offenders.

The findings of Lussier and Healey (2009) were generally consistent with the age-crime curve. Most sex offenders do not reoffend sexually after being released from prison and the Lussier and Healey (2009) study provided additional evidence of this. Congruent with the findings of Kruttschnitt et al. (2000), their finding was further evidence against the argument that sex offenders respond to nothing but long-term imprisonment and intensive community supervision. All offenders eventually desist, albeit at a different rate. The factors or the mechanisms explaining desistance remain tentative (Laws and Ward 2011). To illustrate the importance of age on desistance, Lussier and Healey (2009) compared the predictive accuracy of the offender's age at release to that of the scores of an actuarial tool (i.e., Static-99) that includes one item reflecting the offender's age at release. The findings showed that, by itself, age at release was as good a predictor of reoffending as the score of the Static-99, an actuarial tool designed to determine the risk of reoffending in sexual offenders. These results suggest that the offender's age at release should be an important component considered by risk assessors when considering cases for long-term incapacitation and intensive community supervision. It is plausible that even if the age-crime association is quite general, it is not necessarily invariant and some offenders might deviate from that pattern. Hence, it is possible that the age effect might not operate the same way for individuals characterized by different offending trajectories. Future studies should examine whether the age-crime curve is present for sex offenders characterized by different offending trajectories and whether the age effect has the same impact on sexual recidivism across these groups. The results of the Lussier and Healey (2009) study do not provide empirical evidence for a strategy of selective incapacitation aimed at sex offenders but rather highlight the limited understanding of the role of aging and the process of desistance in sex offenders.

Conclusion

The study of sex offenders' criminal activity using a criminal career approach is still in its infancy. Therefore, the conclusions drawn here should be interpreted accordingly. Longitudinal studies have shown that most juvenile sex offenders do not become adult sex offenders. In fact, studies suggest that about 10 % of convicted juvenile sex offenders become adult sex offenders. Similarly, retrospective longitudinal studies with adult sex offenders suggest that most adult sex offenders were not previously juvenile sex offenders. Said differently, there is not much continuity in sex offending from adolescence to adulthood. Such continuity appears to be more important for sample of adult sex offenders sampled in maximum-security psychiatric hospital suggesting that continuity in sex offending is associated with a mental health disorder. Empirical studies suggest, therefore, that the onset of sex offending for the majority of adult offenders occurs in adulthood. The investigation of the onset of offending shows a gap of about 7 years between the actual onset of sex offending and the age at first conviction for a sex crime. This gap varies across sex offender type. While there are not much age differences in terms of the actual onset of sex offending, there are differences in terms of age at first conviction for a sex crime. Child molesters are more likely to be first convicted for a sex crime later than sexual aggressors of women. Such differences may explain why child molesters are often found to be older than sexual aggressors of women and are most probably due to the fact that child molesters go undetected for longer time periods than sexual aggressors of women. The study of offending frequency reveals that most sex offenders will commit one crime against a single victim. Of those who persist, studies reveal two main offending strategies: one based on maximizing the number of victims (victim-oriented), while the other is based on maximizing the number of events (event-oriented). Those opting for a victim-oriented strategy tend not to reoffend against the same victim, while those following an event-oriented strategy will target few victims that will be abused repeatedly, sometimes over long time periods that can span over a decade. The study of desistance suggests that all sex offenders eventually desist from sex offending but at a different rate. Empirical research also suggests that age and aging does impact sex offender's risk of reoffending over time in a traditional fashion. Older sex offenders have a significantly lower probability of reoffending than younger offenders, yet the criminal justice system is often imposing the most stringent conditions and sentences to older offenders which may impact their ability for successful community reintegration. Such conclusions, however, are based on official data on recidivism and subject to known limitations about official statistics on crime. Using other sources of information in future research to measure crime represents an important challenge for researchers in the field of sexual violence and abuse. Greater understanding of the criminal career of sex offenders will inform policy makers about these offenders' pattern of offending over time but also how to best tackle the problem of sexual violence and abuse.

Related Entries

▶ Sex Offender Treatment
▶ Sexual Recidivism
▶ Specialization and Sexual Offending

Recommended Reading and References

Abel GG, Becker JV, Mittleman MS, Rouleau JL, Murphy W (1987) Self-reported sex crimes of non-incarcerated paraphiliacs. J Interpers Violence 2:3–25

Abel GG, Rouleau JL (1990) The nature and extent of sexual assault. In: Marshall WL, Laws DL, Barbaree HE (eds) Handbook of sexual assault: issues, theories, and treatment of the offender. Plenum Press, New York, pp 9–21

Abel G, Osborn CA, Twigg DA (1993) Sexual assault through the life span: adult offenders with juvenile histories. In: Barbaree HE, Marshall WL, Hudson SM (eds) The juvenile sex offender. Guilford, New York, pp 104–117

Barbaree HE, Blanchard R, Langton CM (2003) The development of sexual aggression through the life span. Ann N Y Acad Sci 989:59–71

Barbaree HE, Langton CM, Blanchard R (2007) Predicting recidivism in sex offenders using the VRAG and SORAG: the contribution of age-at-release. Int J Forensic Ment Health 6:29–46

Blokland A, Lussier P (2012) Sex offenders: a criminal career approach. Wiley-Blackwell, Oxford

Blumstein A, Cohen J, Roth JA, Visher CA (1986) Criminal careers and "career criminals". National Academy Press, Washington, DC

Blumstein A, Cohen J, Das S, Moitra SD (1988) Specialization and seriousness during adult criminal careers. J Quant Criminol 4:303–345

Bradford JM, Boulet J, Pawlak A (1992) The paraphilias: a multiplicity of deviant behaviours. Can J Psychiat 37:104–108

Bremer JF (1992) Serious juvenile sex offenders: treatment and long-term follow-up. Psychiatr Ann 22:326–332

Cann J, Friendship C, Gozna L (2007) Assessing crossover in a sample of sexual offenders with multiple victims. Leg Criminol Psychol 12:149–163

Gebhard PH, Gagnon JH, Pomeroy WB, Christenson CV (1965) Sex offenders: an analysis of types. Harper & Row, New York

Groth AN (1982) The incest offender. In: Sgrol SM (ed) Handbook of clinical intervention in child sexual abuse. D.C. Heath and Co, Lexington, pp 215–239

Guay JP, Proulx J, Cusson M, Ouimet M (2001) Victim-choice polymorphia among serious sex offenders. Arch Sex Behav 30:521–533

Hagan MP, Gust-Brey KL, Cho ME, Dow E (2001) Eight-year comparative analyses of adolescent rapists, adolescent child molesters, other adolescent delinquents, and the general population. Int J Offender Ther Comp Criminol 45:314–324

Hanson RK (2002) Recidivism and age: follow-up data from 4673 sexual offenders. J Interpers Violence 17:1046–1062

Hanson RK (2006) Does static-99 predict recidivism among older sexual offenders? Sex Abuse-J Res Treat 18:343–355

Harris GT, Rice ME (2007) Adjusting actuarial violence risk assessments based on aging or the passage of time. Crim Justice Behav 34:297–313

Heil P, Ahlmeyer S, Simons D (2003) Crossover sexual offenses. Sex Abuse: J Res Treat 15:221–236

Knight RA, Prentky RA (1990) Classifying sexual offenders: the development and corroboration of taxonomic models. In: Marshall WL, Laws DR, Barbaree HE (eds) Handbook of sexual assault: issues theories and treatment of the offender. Plenum Press, New York, pp 23–54

Kruttschnitt C, Uggen C, Shelton K (2000) Predictors of desistance among sex offenders: the interaction of formal and informal social controls. Justice Q 17:61–87

Langan PA, Schmitt EL, Durose MR (2003) Recidivism of sex offenders released from prison in 1994. US Department of Justice, Washington, DC

Laws DR, Ward T (2011) Desistance from sex offending: alternatives to throwing away the keys. Guildford, New York

Lussier P, Healey J (2009) Rediscovering Quetelet, again: the "aging" offender and the prediction of reoffending in a sample of adult sex offenders. Justice Q 26:827–856

Lussier P, Mathesius J (2012) Criminal achievement career initiation and cost avoidance: the onset of successful sex offending. J Crime Justice, in press

Lussier P (2005) The criminal activity of sexual offenders in adulthood: Revisiting the specialization debate. Sexual Abuse: A Journal of Research and Treatment 17:269–292

Lussier P, LeBlanc M, Proulx J (2005) The generality of criminal behavior: a confirmatory factor analysis of the criminal activity of sexual offenders in adulthood. J Crim Justice 33:177–189

Lussier P, Leclerc B, Healey J, Proulx J (2008) Generality of deviance and predation: crime-switching and specialization patterns in persistent sexual offenders. In: Delisi M, Conis P (eds) Violent offenders: theory, public policy and practice. Jones and Bartlett Publishers, Boston, pp 97–140

Lussier P, Bouchard M, Beauregard E (2011) Patterns of criminal achievement in sexual offending: unravelling the "successful" sex offender. J Crim Justice 39:433–444

Lussier P, van den Berg C, Bijleveld C, Hendriks J (2012) A developmental taxonomy of juvenile sex offenders for theory research and prevention: the adolescent-limited and the high-rate slow desister. Crim Justice Behav, in press

Lussier P, Mathesius J (2012) Criminal achievement, career initiation, and cost avoidance: The onset of successful sex offending. J Crim Justice 35: 376–394

Lussier P, van den Berg C, Bijleveld C, Hendriks J (2012) A developmental taxonomy of juvenile sex offenders for theory, research and prevention: The adolescent-limited and the high-rate slow desister. Crim Justice Behav 39:1559–1581

Marshall WL, Barbaree HE (1990) An integrated theory of the etiology of sexual offending. In: Marshall WL, Laws RD, Barbaree EH (eds) Handbook of sexual assault: issues, theories, and treatment of the offender, applied clinical psychology. Plenum Press, New York, pp 257–275

Marshall WL, Barbaree HE, Eccles A (1991) Early onset and deviant sexuality in child molesters. J Interpers Violence 6:323–335

Meloy ML (2005) Sex offender next door: an analysis of recidivism, risk factors, and deterrence of sex offenders on probation. Crim Justice Policy Rev 16:211–236

Pham TH, DeBruyne I, Kinappe A (1999) Évaluation statique des délits violents chez les délinquants sexuels incarcérés en Belgique francophone. Criminologie 32:117–125

Prentky RA, Knight RA (1993) Age of onset of sexual assault: criminal and life history correlates. In: Hall GCN, Hirschman R, Graham J, Zaragoza MS (eds) Sexual aggression: issues in etiology assessment, and treatment. Taylor & Francis, Washington, DC, pp 43–62
Prentky RA, Lee AFS (2007) Effect of age-at-release on long term sexual re-offense rates in civilly committed sexual offenders. Sex Abuse: J Res Treat 19:43–59
Quinsey VL, Lalumière ML, Rice ME, Harris GT (1995) Predicting sexual offenses. In: Campbell JC (ed) Assessing dangerousness: violence by sexual offenders, batterers, and child abusers. Sage, Thousand Oaks, pp 114–137
Radzinowicz L (1957) Sexual offences: a report of the Cambridge department of criminal justice. Macmillan, London
Smallbone SW, Wortley RK (2004) Onset, persistence, and versatility of offending among adult males convicted of sexual offenses against children. Sex Abuse: J Res Treat 16:285–298
Simon L (1997) The myth of sex offender specialization: An empirical analysis. New England Journal on Criminal and Civil Commitment 23:387–403
Soothill K, Francis B, Sanderson B, Ackerley E (2000) Sex offenders: specialists, generalists or both? Br J Criminol 40:56–67
Stander J, Farrington DP, Hill G, Altman PME (1989) Markov chain analysis and specialization in criminal careers. Br J Criminol 29:317–335
Thornton D (2006) Age and sexual recidivism: a variable connection. Sex Abuse: J Res Treat 18:123–135
Weinrott MR, Saylor M (1991) Self-report of crimes committed by sex offenders. J Interpers Violence 6:286–300

Criminal Careers

▶ Onset of Offending

Criminal Careers and Public Policy Responses

Paul Mazerolle
Arts, Education, and Law, Griffith University,
Griffith, QLD, Australia

Overview

The past three decades of research and theorizing in Criminology has provided an opportunity to learn and further understand the dimensions of criminal careers as well as offending behavior across the life course. First articulated by Alfred Blumstein and colleagues in 1986, the criminal career approach to understanding patterns of offending behavior has provided a wealth of information that has fostered greater understanding of criminal behavior, greater knowledge to foster theoretical development, and realignment as well as important implications for the development of age-graded policy responses to crime and criminal behavior.

While the strongest emphasis of the criminal career approach has been on generating a wealth of information about criminal careers, including factors related to their initiation, factors related to components of active offending activity, as well as factors related to career desistance or termination, the approach has important implications for informing policies and programs aimed at controlling or preventing crime and criminal behavior.

This entry provides an opportunity to describe the criminal career approach as well as chart how the criminal career paradigm provides a useful vehicle for informing and facilitating the development of effective policies and practices aimed toward controlling and preventing crime and criminal behavior across the life course.

Introduction

The criminal career paradigm has emerged over the past three decades and is one of the most dominant perspectives for facilitating empirical research about criminal offending behavior. The modern understanding of the criminal career approach dates to work by Alfred Blumstein and colleagues (1986); however, the precursors to this form of thinking had its origins in prior work examining offending birth cohorts (cf. Wolfgang et al. 1972) as well as research examining controversial policy responses to crime (Greenwood and Abrahamse 1982).

The rise of the criminal career approach and the proliferation of research on aspects of criminal careers have coincided with the rise of the life course approach in Criminology. Life course Criminology, which incorporates Developmental Criminology (LeBlanc and Loeber 1998), represents an approach for examining offending behavior over time, through age-graded sequences as well as contexts that are both meaningful and subject to change (Piquero and Mazerolle 2001). According to this approach the life course incorporates age-graded roles, responsibilities, and institutional contexts and sequences that develop over time (Elder 1995). The life course approach provides a rich framework for understanding behavioral change and continuity over time and across the age-graded life span. It includes concepts such as trajectories over the life course as well as transitions and turning points (Sampson and Laub 1993).

A related way of conceptualizing the life course approach involves embracing the concept of offending pathways which vary over time and are influenced by a range of social contexts. Overall, the life course perspective presents unique opportunities for developing a comprehensive understanding of criminal behavior. The life course approach is entirely consistent with the criminal career paradigm, and both approaches have implications for informing policy directions aimed at preventing and controlling criminal behavior.

In the following sections, a presentation of the criminal career approach is provided. This is followed by an examination of policy responses to crime and criminal behavior which relate to key dimensions of criminal careers across the life course.

The Nature of Criminal Careers

The criminal career approach examines dimensions of offending behavior and aims to categorize such meaningful dimensions into concepts that foster empirical assessment and understanding. The criminal career approach differentiates individual patterns into three broad phases, including the onset or initiation of offending, the active offending phase, as well as the desistance or termination phase. Criminal careers differ with regard to when they begin (early or late onset), how they proceed (highly specialized vs. versatile, high or low frequency, etc.), and whether and when they finish or terminate (desist or persist).

Over the past 20 years, a wealth of research has uncovered various aspects of criminal careers. For example, research supports the finding that offenders with an early age at onset to offending are at an enhanced risk for a serious and persistent criminal career (Moffitt 1993; Patterson and Yoerger 1997). Offenders with a later-onset age tend to have less prolonged offending careers and stronger prospects for completing school, securing employment, and leading a life with diminished offending risks in adulthood. Importantly, the precursors for early and late onset offending vary and consequently have implications for prevention efforts.

The bulk of the criminal career can be described by an individual's offending frequency, the nature of their offending with respect to offense type (e.g., violence, property crime), as well as whether offending is escalating in seriousness over time. Indeed, research consistently finds that serious offenders are those that begin their careers early, exhibit a high frequency of offending across diverse or versatile types of offending including violence, over a prolonged or persistent period of time (Moffitt 1993). Importantly, it is crucial to ascertain the nature, dimensions, and precursors for high-frequency offending, as well as factors that relate to versatile or specialist offending patterns because the opportunities for prevention are informed by the diverse nature of offending causation.

In terms of the later stages of criminal careers, research illustrates factors related to desistance or career termination (Laub and Sampson 2003; Sampson and Laub 1993) as well as factors related to the process of desistance (Bushway et al. 2001). Increasingly, research reveals the value of employment skills and experiences, as well as the protective influences and consequent informal social control related to developing meaningful intimate relationships including marriage (Horney et al. 1995). Clearly, understanding when, how, and why active offenders refrain from further offending is significant for informing a range of programs and practices to minimize offending persistence.

The value of the criminal career approach rests in the meaningfulness of the concepts for capturing criminal career dimensions (e.g., onset, frequency), the ability to identity correlates for related career dimensions, the ability to inform extant theories about criminal behavior, as well as in utilizing the knowledge gained to inform policy and practice directions to foster greater crime prevention and control. It is to this issue we turn in the next section.

Criminal Careers and Crime and Justice Policy and Practice

The past 20–30 years has borne witness to an increasing emphasis on understanding criminal careers as well as an emphasis on understanding crime across the life course. These developments have co-occurred during a period whereby an increasing emphasis on evidence-based policy and practice has occurred across the western world.

Criminal justice policy making has been observed to be oftentimes a haphazard exercise fraught with ideological influences as well as misinformed conceptions about crime, its causes, as well as what constitutes effective crime control and prevention programs (Cullen and Gendreau 2001; Currie 1993). Over the past 15 years, there has been an increasing push toward evidence-informed policy responses, and many criminologists have been instrumental in supporting such directions. These movements have also been reinforced by a public increasingly frustrated with ineffective and/or wasteful crime policy responses. It is within this context that the role of research knowledge about criminal careers across the life course is well placed to inform responses to criminal behavior. Indeed, if the precursors of criminal career dimensions vary, then the policy and practice responses need to be differentiated to address various aspects of criminal career dimensions. For example, the policy responses for reducing onset to criminality may well need to differ from the policy responses that reduce violent offending, frequent offending, or chronic offending. A large body of research has indeed addressed these issues. We turn now to considerations of policy responses to crime across the life course.

Policy and Practice Focus Areas Across Criminal Career Dimensions and the Life Course: Preventing or Delaying Offending Onset

Given that criminal careers are often established early in the life course, programs aimed at reducing or delaying offending onset represent important areas for investment. The past decade has seen a push toward early intervention programs as such efforts seek to provide effective supportive conditions and protective influences around young people at risk for embarking on an offending pathway (i.e., onset). Programs and policies designed to prevent onset to an offending pathway are generally targeted toward family or school settings because they represent important institutional contexts whereby effective interventions can occur.

There are several examples of programs aimed at preventing offending onset. For example, a Nurse-Family Partnership program designed by David Olds has been aimed at supporting pregnant mothers through enhanced visits before and after the birth to assist them with supportive information and assistance in dealing with young children. Three randomized controlled trials have been conducted related to this program, and improvements in health and economic outcomes have been observed which have subsequent benefits for reducing crime. A study on the crime reduction benefits of this program demonstrates discernible reductions in arrests and convictions as well as reduced drug- and alcohol-related problems in comparisons to control subjects (Olds et al. 1998).

Another area of focus for delaying or preventing onset involves parent training programs. Parent training programs are aimed at providing effective parenting models and skills which subsequently translate into improved socialization of the child. These programs are especially aimed at fostering improved social and emotional competence in young people, readiness for school, as well as addressing conduct and behavioral challenges. One of the developers of an effective program, Dr. Carolyn Webster-Stratton has been at the forefront of research demonstrating the effectiveness of such approaches (Webster-Stratton and Taylor 2001). Numerous studies have demonstrated that early family/parent training programs are effective in reducing crime in adolescence and adulthood.

A range of additional programs have shown good success in supporting children in their development and preparation for school, with the Perry Preschool Project a notable standout (Schweinhart 2007). Additionally, a number of school-based delinquency prevention programs have shown good promise in preventing delinquency initiation and programs which target cognitive processes and critical thinking and behavioral skills appear to be worthy of further investment (Gottfredson 2001).

Reducing High-Frequency and Chronic Offending

Offenders who engage in serious, violent, frequent, and persistent criminal behavior generate a high degree of policy and program attention. Such chronic, high-rate offenders generate a high volume of the crime which is managed through the criminal justice system and which creates a high degree of harm for the community. In response, several policies and programs have been implemented with varying degrees of success. The 1980s, for example, included a push toward selective incapacitation approaches for chronic offenders (Greenwood and Abrahamse 1982). Selective incapacitation aimed to further prolong prison sentences for chronic, high-frequency offenders to ensure their offending risks could be removed from community exposure. Such approaches, while popular in some circles, failed to garner support due to concern over the effectiveness and fairness of such approaches.

A further example of a “get tough” policy approach relates to the three strikes and you’re in policy across many jurisdictions. Such policies, while popular among a subset of the community, tended not to persist because of cost concerns, concerns about their minimal impact of crime across the community, as well as concerns over sentencing equity and consistency.

A further area of challenge in designing programs and responses for serious offenders relates to conceptions of specialized offending patterns which require program attention. For example, a range of programs exist for dealing with sexual offenders or domestic violence offenders but much of the evidence suggests that such programs are of mixed effectiveness. A major explanation for such equivocal results is that most high-frequency, serious offenders exhibit versatile offending patterns and therefore require responses that reflect their generalized criminal tendencies. Thus, more generalized policy and program responses have preventative implications for domestic violence as well as sexual offending.

While there are encouraging results emerging from a range of prevention programs (e.g., Communities That Care, youth mentoring), the move toward multisystemic therapy (MST) programs among several program providers is especially encouraging. MST was developed by Dr. Scott Henggeler. The program provides intensive family- and community-focused treatment which aims to confront the institutional systems that influence chronic and violent juvenile offenders. MST views individuals embedded within various systems involving family, community, peers, etc. Thus, the intensive program response often includes teachers, friends, as well as family members. In general, the MST program is aimed at the high-frequency, high-rate, and serious offender group. Over a period of several years and studies, MST reveals consistently positive results across various domains including school, family, and crime. For example, evaluations of MST have revealed improvements in keeping youth in school, improvements in family relationships, reductions in drug and alcohol use, as well as large reductions in rearrests of up to 70 %.

Consequently, MST is one of the most impressive programs available for addressing serious, violent, and chronic offenders.

Finally, important opportunities exist for correctional rehabilitation programs to target areas of intervention for persistent, high-risk offenders. While correctional rehabilitation has been out of favor at times, the weight of research evidence demonstrates the important prosocial outcomes which emanate from effective programs (Andrews et al. 1990). A series of studies consistently reveal the positive results that can be observed through investing in programs that target key areas of intervention need (antisocial attitudes, critical cognitive processing skills, etc.), especially for high-risk offenders, and in programs that are responsive or tailored to the learning styles of offenders (Andrews et al. 1990).

Facilitating Desistance and Career Termination

Fostering desistance from crime represents an ongoing public policy challenge. Research on criminal careers illustrates how offending changes across the life course and reveals multiple areas for prevention, in particular in the emerging adulthood phase of the life course. Key areas of focus for fostering desistance relate to the need for employment skills, educational attainment, substance abuse prevention, as well as opportunities for meaningful relationship development.

Criminologists John Laub and Rob Sampson have been at the forefront in undertaking research on how informal social control in adulthood can foster desistance. Experiences in acquiring employment and meaningful intimate relationships (e.g., marriage) have been shown to support offending desistance (Sampson and Laub 1992) in adulthood. Thus, investing in the employment skills of established or active offenders represents an important opportunity for fostering desistance. A similar observation holds for education. Increased levels of education are consistently related to desistance and investments in educational opportunities for active offenders, much like job skills, represent important opportunities to strengthen social control and social ties to prosocial institutions. With increased educational attainment, expanded employment opportunities arise, and such investments in conformity can lead to reductions in re-offending.

Given the ongoing relationship between drugs and crime, it appears that desistance is not easily achieved among offenders with drug abuse problems. Thus, efforts at fostering drug abuse treatment among active offenders are an important investment toward reducing persistent offending over time. Moreover, there is clear and established evidence which shows that among serious offenders with drug abuse problems, participating in treatment is associated with reductions in drug use, higher employment rates, as well as reductions in crime.

Challenges and Implications

The explosion of research on criminal careers has shaped understanding about criminality across the life course. Beyond its direct contribution to shaping understanding as well as theories about criminal behavior, this knowledge is highly significant for informing practical responses to crime and ensuring effective crime policies are enacted, implemented, and supported.

There are a range of implications regarding the importance of placing criminal career research findings into a policy and practice context. For example, while the evidence about effective policy responses to crime and criminal behavior evolves over time, the embrace of politicians and policy leaders to such knowledge is highly variable. At times, evolving knowledge is met with evolving acceptance. At other times, evidence-informed policy and practice appears to take a backward step. A case in point relates to the embrace of correctional boot camps for juvenile offenders. The research from the 1990s to 2000s consistently revealed problems associated with boot camps and that they were not effective in addressing juvenile offending (MacKenzie et al. 2001). Yet, a recently elected state government in Australia has campaigned on a *get tough on crime* election platform and is implementing a "new style" boot camp to address serious juvenile crime. It is unsurprising that the leader in charge of this initiative has a background in the military and obviously knows what works, despite evidence to the contrary. In summary, gains have been hard fought and achieved in relation to the embrace of evidence to inform policy and practice around the western world, but such gains can be quickly lost. Thus, it is important to be ever mindful and vigilant about the fragility of such gains as new generations of politicians and political masters require exposure to evidence about the nature of criminal careers as well as evidence about what works to prevent and control criminal behavior across the life course.

While there are strong connections between criminal career dimensions and areas for public policy focus and investment (e.g., delaying onset), it is important to understand the role of social context and related factors that can influence criminal career dimensions as well as the required policy responses to prevent and control crime. This observation simply calls attention to the point that context matters and rarely is it the case that patterns of offending are invariant across gender, racial groups, and social class differentiations among other considerations. Indeed prior research has recognized the significant variations in criminal career dimensions as well as various risk factors across gender, race, and social class (Piquero et al. 2007).

Consider, for example, the role of gender. Although a number of common factors are found to increase risk for both males and females, some differences do appear in the overall sensitivity to certain risk and protective factors as well as in the developmental course and nature of antisocial behavior between the genders. For example, community-level factors such as socioeconomic disadvantage have been found to affect girls as young as 5 years old. There is also some evidence that females may be more responsive than males to the protective influence of factors such as prosocial role models, reduced community violence, and enhanced extracurricular opportunities. Females appear to be more sensitive to the adverse effects of poor familial attachment and conflict but also more sensitive than males to the protective effect of strong attachment to family.

Female juvenile offenders tend to present with multiple and chronic family-level risk factors and a family environment which is typically more adverse than that experienced by male juvenile offenders. Females associating with delinquent male peers, particularly females experiencing an early-onset menarche, are at significantly greater risk of adolescent-onset conduct problems (Moffitt et al. 2001). Moreover, puberty presents a time of enhanced risk for females, particularly those with preexisting externalizing behavior problems or otherwise salient risk factors such as academic difficulties, a history of abuse, or parental criminality/psychopathology (Chesney-Lind and Sheldon 2004).

At an individual level, the risk associated with neurocognitive difficulties seems to be more salient for males (Moffitt et al. 2001). Also, males typically exhibit higher trait levels of impulsivity and negative emotionality than females. Importantly, it is this difference which appears to primarily underlie the marked difference between males and females in the prevalence of adolescent antisocial behavior and conduct disorder (Moffitt et al. 2001).

There is also some evidence that developmental trajectories to antisocial behavior and delinquency differ between the genders. Moffitt (1993) distinguished between two etiological trajectories to adolescent antisocial behavior problems: one which begins and typically ends in adolescence (adolescence-limited) and one with an early, childhood-onset which continues on into adulthood (life course persistent). However, more recently other researchers (e.g., Silverthorn et al. 2001) have argued that while Moffitt's (1993) adolescent-limited and life course persistent taxonomy captures the timing of both male and female antisocial behavior/offending patterns, it may be failing to identify a subgroup of adolescent-limited females who exhibit a more serious pattern of offending/antisocial behavior problems as well as a more problematic childhood risk profile.

The above description illustrates that diversity exists in the criminogenic risk factors for males and females, and such diversity has implications for the flow on toward criminal career development and related dimensions. Such diverse contexts and influences on criminal career initiation and related dimensions obviously transcend gender and involve further contexts including race, age, and social class, among other areas. Prevention and crime control efforts developed to address certain aspects of criminal career dimensions (e.g., onset, frequency) require flexibility and adaptability to accommodate for contextual diversity. It is clearly the case that when policy simplicity tries to address behavioral complexity, very poor crime reduction results are observed. What communities and governments need to strive for are ensuring that policies and programs aimed at controlling and preventing crime are informed by the best knowledge around behavioral complexity of criminal career dimensions and diversity to support appropriate translation into comprehensive policy and programmatic development and delivery that moves importantly beyond the one size fits all approach so often observed around the world.

Conclusion

The criminal career paradigm provides a useful approach for understanding unique dimensions of offending behavior. The approach enables researchers to examine correlates of unique career dimensions (e.g., onset, persistence, specialization) and uncover diversity in the nature of criminal careers. At the same time, the criminal career approach has several implications for fostering knowledge about policies and practices aimed toward crime control and prevention and in this way can contribute directly to effective crime reduction policies.

This entry has demonstrated a range of policy and program approaches that emerge from considering the criminal career approach. In the main, responses that are informed by key aspects of career dimensions are best placed to impact upon crime and criminal behavior. A range of policy and program options exist for addressing the initial stages of criminal careers. Examples include nurse home visitation programs, parent training, as well as early school transition programs. By contrast, more established criminal careers require specific approaches in response to career dimensions. For example, the frequent offender, the persistent, or the chronic offender requires more concentrated criminal justice related interventions. Examples of efforts that appear to be ineffective are evident (e.g., selective incapacitation, boot camps, three strikes), however, there are encouraging programs worthy of investment including multisystemic therapy (MST) as well as a range of cognitive behavioral treatment programs aimed at the prison-based population of offenders who will one day return to their communities.

Responses aimed at fostering desistance and criminal career termination require attention to the criminogenic factors that prolong criminal offending and drug abuse treatment programs show good promise in this area. Moreover, greater attention to fostering informal social bonds to employment experiences as well as to intimate relationships represent good opportunities for turning points away from crime. Thus opportunities for acquiring job skills and education represent useful investments aimed toward fostering the process of desistance over time.

Knowledge about criminal career dimensions and their understanding continues to evolve. While current understanding about effective, evidence-based policy responses to diverse crime problems is strengthening over time, a constant challenge is in ensuring that evidence about what works to control and prevent crime is embraced by the relevant authorities. Importantly, the criminal career approach is ideally placed for fostering knowledge about varying aspects of offending across the life course as well as illustrating a range of useful policy investments for preventing and controlling criminal activity.

Related Entries

- ▶ Age-Crime Curve
- ▶ Career Criminals and Criminological Theory
- ▶ Communities That Care
- ▶ Crime Specialization, Progression, and Sequencing
- ▶ Criminal Career of Sex Offenders
- ▶ Criminal Careers and Public Policy Responses
- ▶ Desistance and Supervision
- ▶ Desistance from Crime
- ▶ Desistance from Gangs
- ▶ Evidence-Based Policy in Crime and Justice
- ▶ Identification Issues in Life Course Criminology
- ▶ Measuring Crime Specializations and Concentrations
- ▶ Offense Specialization: Key Theories and Methods
- ▶ Officer Safety, Health, and Wellness
- ▶ Onset of Offending
- ▶ Pathways to Delinquency
- ▶ Risk Factors for Adolescent Sexual Offending
- ▶ School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis
- ▶ Sex Offenders and Criminal Policy
- ▶ Specialization and Sexual Offending
- ▶ Specialization in Juvenile Offending

Recommended Reading and References

Andrews DA, Zinger I, Hoge R, Bonta J, Gendreau P, Cullen F (1990) Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. Criminology 28:369–404

Blumstein A, Cohen J, Roth J, Visher C (1986) Criminal careers and 'career criminals', vol 1. National Academy Press, Washington, DC

Bushway S, Piquero A, Broidy L, Cauffman B, Mazerolle PI (2001) A developmental framework for empirical research on desistance. Criminology 39:491–515

Chesney-Lind M, Sheldon R (2004) Girls, delinquency and juvenile justice, 3rd edn. Wadsworth, Belmont

Cullen FT, Gendreau P (2001) From nothing works to what works: changing professional ideology in the 21st century. Prison J 81:313–338

Currie E (1993) Reckoning: drugs, the cities, and the American future. Hill and Wang, New York

Elder G Jr (1995) The life course paradigm: social change, and individual development. In: Moen P, Elder G Jr, Luscher K (eds) Examining lives in context: perspectives on the ecology of human development. American Psychological Association, Washington, DC, pp 101–135

Gottfredson DC (2001) Schools and delinquency. Cambridge University Press, New York

Greenwood PW, Abrahamse A (1982) Selective incapacitation. Rand, Santa Monica

Horney J, Osgood DW, Marshall IH (1995) Criminal careers in the short-term: intra-individual variability in crime and its relation to local life circumstances. Am Sociol Rev 60:655–673

Laub JH, Sampson R (2003) Shared beginnings, divergent lives: delinquent boys to age 70. Harvard University Press, Cambridge, MA

LeBlanc M, Loeber R (1998) Developmental criminology updated. In: Tonry M (ed) Crime and justice: an annual review of research, vol 23, pp. University of Chicago Press, Chicago, pp 115–198

MacKenzie DL, Wilson DB, Kider S (2001) Effects of correctional boot camps on offending. The Ann Am Acad Polit Soc Sci Novemb 578:126–143

Moffitt TE (1993) Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. Psychol Rev 100:674–701

Moffitt TE, Caspi A, Rutter M, Silva PA (2001) Sex differences in antisocial behaviour: conduct disorder, delinquency, and violence in the Dunedin Longitudinal Study. Cambridge University Press, Cambridge

Olds D, Henderson C Jr, Cole R, Eckenrode J, Kitzman H, Luckey D, Pettitt L, Sidora K, Morris P, Powers J (1998) Long-term effects of nurse home visitation on children's criminal and antisocial behavior: 15 year follow-up of a randomized controlled trial. JAMA 280:1238–1244

Patterson GR, Yoerger KL (1997) A developmental model for later-onset delinquency. In: Deinstbeir R, Osgood DW (eds) Motivation and delinquency. University of Nebraska Press, Lincoln, pp 119–177

Piquero AR, Mazerolle P (2001) Life-course criminology: contemporary and classic readings. Wadsworth, Belmont

Piquero AR, Farrington DP, Blumstein A (2007) Key issues in criminal career research: new analyses of the Cambridge Study in delinquent development. Cambridge University Press, Cambridge, UK

Sampson R, Laub JH (1993) Crime in the making: pathways and turning points through life. Harvard University Press, Cambridge, MA

Schweinhart LJ (2007) Crime prevention by the High/Scope Perry Preschool Program. Vict Offenders 2:141–160

Silverthorn P, Frick PJ, Reynolds R (2001) Timing of onset and correlates of severe conduct problems in adjudicated girls and boys. J Psychopathol Behav Assess 23:171–181

Webster-Stratton C, Taylor T (2001) Nipping early risk factors in the bud: preventing substance abuse, delinquency, and violence in adolescence through interventions targeted at young children (0–8 years). Prev Sci 2:165–192

Wolfgang M, Figlio R, Sellin T (1972) Delinquency in a birth cohort. University of Chicago Press, Chicago

Criminal Careers of Places

Sue-Ming Yang[1], David Weisburd[2,3] and Elizabeth R. Groff[4]
[1]Department of Criminology, National Chung-Cheng University, Min-Hsiung, Chia-Yi, Taiwan
[2]Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA
[3]Faculty of Law, The Hebrew University, Mt. Scopus, Jerusalem, Israel
[4]Temple University, Philadelphia, PA, USA

Synonyms

Criminology of places; Hot spots

Overview

Though both individuals and places have long been the foci of criminological research, the majority of criminological research focuses on individual criminal involvement (Reiss 1986). As a result, criminologists often assume that places play a relatively minor role in explaining crime compared to an individual's criminal propensity. In fact, there has been a long history of studying of crime at places dating back to the nineteenth century.

In the recent decades, there has been a gradually increasing interest in academia to study crime at places. Particularly, research has focused on the distribution of crime across geographic places as well as the explanatory factors of crime at places (Weisburd et al. 2012; Sampson et al. 1997). To provide an overview for the recent development, this entry reviews literature related to crime concentration, longitudinal crime rates at places, and empirical findings as well as challenges faced when studying the criminal career of places.

Among studies of crime distributions at places, one of the most well-known and widely agreed-upon findings is the concentration of crime at a small number of places. The earliest work known to demonstrate this empirically was from French scholar Michel-André Guerry and the Venetian cartographer Adriano Balbi. They published three maps on the distribution of crime in France in the years 1825–1827. In 1833 the influential *Essai sur la statistique morale de la France* was published, in which Guerry examined whether poverty and density of population might explain higher crime rates (Guerry 1833). His map shows clear patterns of differential concentration patterns of crime. The rich northern *departements* were confronted with higher property crime rates than the poor *departements* in the south of France. He concluded that the level of poverty was not the direct cause of crime. Similarly, his data suggested that population density was not a cause of crime. In 1836 his friend Alexandre Parent-Duchâtelet published an empirical study containing maps on the distribution of prostitution from 1400 until 1830 in Paris (Parent-Duchâtelet 1836). Because of the official control of brothels by the Paris authorities, systematic data were available on prostitutes, especially from the years 1817 to 1827. Even information regarding the *departements* where they came from was collected. Not surprisingly, the center of the city had the highest number of prostitutes. He used neighborhoods, as defined by administrative boundaries, as the unit of analysis.

Other scholars also studied these issues in the 1830s. Most notably, the work of Adolphe Quetelet (1831[1984]) examined crime rates across provinces and countries in France and found that places with higher property crime rates tended to be wealthier. He pointed out on his maps that crimes, just like wealth, are not distributed equally across places. Guerry and Quetelet could be considered as pioneers of crime and place research, and their works set the stage for later theorizing linking crime and place.

Almost 100 years after the work of Guerry and Quetelet, in a classic study of juvenile delinquency, Shaw and McKay examined the residential locations of juvenile offenders and then pinned the addresses on a big map of the city of Chicago (Shaw and McKay 1942). The geographic distributions of those addresses showed clear clustering patterns in their residential locations. Through their close observations and explorations of the city, Shaw and McKay concluded that not all areas were plagued with the same amount of crime. The so-called transition zone outside of the center of the city had the highest amount of crime in the whole city. Specifically, places with higher juvenile crime rates tended to be closer to the center of the city, disorganized, occupied by the poor, and also possessed other social illnesses. Moreover, Shaw and McKay (1942) argued that as long as the structural factors of a place, such as poverty, racial heterogeneity, urban decay, and population turnover rates, remain stable, crime rates will as a result also remain stable over time.

Despite the preponderant evidence found on the stability of crime places, others have found that crime rates do change under certain circumstances. For example, the argument that crime remains stable in Chicago was challenged by studies done in later years (Bursik and Webb 1982). Specifically, Bursik and Webb reanalyzed crime rates in Chicago neighborhoods with an extended dataset compared to the original one used by Shaw and McKay and found that the stability of crime in Chicago was purely a historical artifact. Crime rates in Chicago started to fluctuate after WWII. The lack of empirical credibility, the focus on individuals as the center of study, and other theoretical flaws contributed to the decline of the traditional social disorganization theory formulated by Shaw and McKay. However, in the 1980s a group of young scholars led by Albert Reiss started to take a new interest in the study of crime places, rather than individuals. Editing an early volume in the *Crime and Justice* series, Reiss and Tonry (1986) sought to bring *Communities and Crime* to the forefront of criminological interests. Reiss sought to raise a new set of questions about crime that had been ignored in earlier decades: "Recent work on communities and crime has turned to an observation that Shaw and McKay neglected: not only do communities change their structures over time but so often do their crime rates. . .a recognition that communities as well as individuals have crime careers" (Reiss 1986, p. 19). This volume and other work developed in this period drew upon the identification of neighborhoods and communities to expand insights about the development of crime (Bursik and Webb 1982; Clarke 1983).

Sampson (1993) also pointed out the importance of linking time and place in the study of crime. Criminology had been dominated by methodological individualism; as such, the understanding of crime had been decontextualized. Putting crime back to its place has then become salient for a more holistic understanding of crime. Another major problem was the lack of a temporal perspective, as well as the availability of longitudinal data, in crime and place research. Without taking into account the actual sequence of actions, it is not possible to understand the causal mechanisms behind the phenomenon (Sampson 1993, p. 430). To address this issue, Sampson argues that communities have "careers" in crime, and longitudinal study is needed to disentangle changes in crime rates in communities (see also, Reiss 1986). The initiative to refocus criminological inquiries at places has led to two different but tightly linked tracks of study examining crime at places. The first line of research focuses on the hot spots phenomenon that examines stability of crime across geographic units to identify where the concentration of crime locates. The second line of research moves one step further and focuses on tracing the trajectories of crime trends of places over time.

Crime Concentration at Places

Crime concentration has been found in many studies across different places at different geographic levels. Early research done by Guerry, Quetelet, and Shaw and McKay clearly demonstrated that crime was concentrated in a few places. In more recent years, the research findings on crime concentration and the empirical findings about hot spots of crime enhanced development of theoretical ideas about crime distribution at places. The introduction of new theoretical ideas, such as situational crime prevention

(Clarke 1983), routine activities theory (Cohen and Felson 1979), environmental criminology, and crime pattern theory (Brantingham and Brantingham 1984), invited people to explore the linkage between crime and places. At the same time, the popularity of personal computer and the affordable mapping software allow for timelier and more in-depth crime analysis to visualize the concentration patterns of crime for both researchers and practitioners. This theoretical development and technological advancement jointly promoted the research of crime at place.

Subsequent research continued to identify high-crime neighborhoods within cities (Bursik and Webb 1982; Sampson and Groves 1989). The concentration of crime has been found at various geographic levels, not just the neighborhood level. The law of concentration also applies to specific types of "places." Spelman (1995) examined calls for service at public places such as schools, housing projects, subway stations, parks, and playgrounds, over a 4-year time period in Boston. Analysis results showed that there were strong place-to-place variations on crime and disorder rates. Specifically, the worst 10 % of public places were responsible for about 30 % of crime. In addition to the concentration of crime, Spelman also found that long-run risks at places remained quite stable and explained a substantial amount of variation in citizen's calls for service. Concentration patterns have also been observed in "high-risk" places such as bars/taverns. For example, Sherman et al. (1992) examined tavern crimes in Milwaukee and found that as few as 15 % of taverns in the city accounted for more than 50 % of the crimes that occurred in all taverns.

The extent of concentration is even more salient at small geographic areas like addresses and places, than in larger aggregations of geography such as neighborhoods or census tracts. In one of the pioneering studies in this area, Lawrence Sherman and colleagues found that the majority of citizens' calls for service were generated by about 3 % of addresses in Minneapolis, Minnesota.

Fifteen years later, using street segments as the unit of analysis, Weisburd et al. (2004) demonstrated similar concentration rates in Seattle, WA. That is, about 4–5 % of street segments accounted for about 50 % of all crime each year. This percentage remained quite stable during their 14-year observation period. Using 16 years of data and adding refinement to the definition of street segments, Weisburd et al. (2012) analyzed extended data from Seattle and reconfirmed the concentration of crime at the street-segment level. In Seattle, over 9 % of the segments did not evidence a single crime incident over this 16-year period. Consistent with the Pareto Principle (a.k.a. the 80–20 rule), they found that 80 % of crime incidents were found on between 19 % and 23 % of segments across their study period, while 100 % of incidents fell on between 60 % and 66 % of segments in a given year. Another study by Weisburd and Amram (forthcoming) found that 5 % of the street segments in Tel Aviv were responsible for 50 % of the crime incidents, a statistic remarkably similar to the results from Seattle. These studies and others (Brantingham and Brantingham 1984; Clarke 1983) have established crime places as an important focus of criminological inquiry and practical crime prevention.

The concentration of crime at "hot spots" is also observed for specific crime types. Braga et al. (2010) examined gun crimes in Boston and concluded that street segments and intersections account for less than 5 % of the total streets in the city but had over 75 % of gun violence incidents within them. Additionally, Weisburd and Mazerolle (2000) found that approximately 20 % of all disorder crimes and 14 % of crimes against persons were concentrated in just 56 drug-crime hot spots in Jersey City, New Jersey, an area that comprised only 4.4 % of street segments and intersections in the city. Similarly, Eck et al. (2000) found that the most active 10 % of places (in terms of crime) in the Bronx and Baltimore accounted for approximately 32 % of a combination of robberies, assaults, burglaries, grand larcenies, and auto thefts. Finally, examining juvenile arrests over a 14-year period, Weisburd et al. (2009) also found an extreme concentration rate – 3–5 % of street segments were responsible for all juvenile arrests during any given year. In short, the findings from studies provide strong support for the idea of a "law of concentrations" for crime at places.

To date, the available empirical evidence clearly shows that crime does not randomly distribute across places. Rather, research has found that a small number of places contribute a disproportionate share of crime. This finding echoes the results from the famous Philadelphia Cohort Study, in which Wolfgang et al. (1972) found that less than 6 % of individuals were responsible for more than 50 % of the arrests. The comparison result led Sherman to argue that future crime is "six times more predictable by the address of the occurrence than by the identity of the offender" (1995, pp. 36–37). He then asked "why aren't we doing more about it? Why aren't we thinking more about wheredunit, rather than just whodunit?" As such, many scholars proposed the need to study of crime places using criminal career perspective.

Stabilities and Changes of Crime at Places

Following the tradition of Chicago school, scholars like Robert Bursik and his colleagues started to focus on how crime develops in communities. Specifically, some scholars began to apply the concepts drawn from the criminal career paradigm to examine whether places, just like individuals, also have "criminal careers" (Reiss 1986, p. 19; Sherman 1995). For example, Schuerman and Kobrin (1986) applied a dynamic model to explaining neighborhood crime characteristics over time. Recognizing the importance of understanding developmental crime trends, Sherman (1995) pointed to the relevance of the criminal career paradigm to the study of longitudinal crime rates at micro places. Other important concepts have also been borrowed from the criminal career paradigm to study the onset, continuity, specialization, and desistance of crime at places (Spelman 1995).

Among the research of crime at places, stability of crime trends is the most common finding across different geography. For example, Griffiths and Chavez's (2004) study of Chicago neighborhoods also found support for the argument that places might follow very different trajectories with regard to homicide trends between 1980 and 1995. Specifically, they found that different neighborhoods exhibited different homicide trends – some had stable trajectories, some had increasing trends, while others had experienced high number of homicides. But the majority of the neighborhoods demonstrated stable homicide trends, while a smaller percentage of neighborhoods showed high increasing homicide rates during the same time.

The stability of crime at places is also found at smaller geographic units. Weisburd and colleagues applied the criminal career concepts to examine stability and changes in crime rates of street segments in Seattle, WA (Weisburd et al. 2004, 2012). Specifically, Weisburd et al. (2004) found chronic hot spots and chronic cold spots of crime over 14 years of analysis.

With additional 2 years of data, (Weisburd et al. (2012)) reconfirmed the previous findings using group-based trajectory analysis (see Nagin and Land 1993) to classify street segments into different trajectory groups based on their levels and directions of crime trends. They found 22 distinct groups of street segments following different developmental trends. Out of the trajectory groups, almost half of the street segments (49.6 %) were classified into the "crime-free" trajectory group as they had crime counts at or near zero for every year of the study. The second biggest trajectory group was trajectories that featured with "low stable" crime trends. Totally, 32 % of street segments with relatively low crime rates that remained stable over the 16-year study period were classified into this category. The next category was the "moderate stable" trajectory that had a moderate crime rate that was flat over time and contained 1.2 % of the street segments in Seattle. Importantly, the stability of crime is not limited to low-crime places. In fact, they also found that about 1 % of street segments, the "chronic high" trajectory, evidenced a stable high crime rate over time. Despite the nationwide crime drop observed in the early 1990s, those places continued to experience high crime problems over that period of time. In sum, Weisburd et al. (2012) found that over 80 % of streets in Seattle had extremely stable crime patterns over time. Though the hot spots phenomenon has been supported by numerous studies, this was the first time research was able to empirically demonstrate the long-term stability of hot spots (and cold spots) at such a micro level of geography.

The extent of stability of crime trends is astonishing over such a long period of time. Nonetheless, they also found places with fluctuating crime rates. Consistent with the crime drop phenomenon across the US cities through the 1990s and early 2000s, they found places that experienced declines in crime. The "low-rate decreasing" category accounted for 9.3 % of segments that had relatively low crime rates which evidenced declines in crime. The "high-rate decreasing" group contained 2.4 % of segments that had high rates which decreased over time. While most segments remained stable or experienced crime drops over this period, there were some segments which experienced crime spikes over the study period. In the city of Seattle, about 3.8 % of street segments featured "low increasing" crime rates, while another 0.9 % of segments experienced "high increasing" crime rates over the 16-year period.

The findings from Weisburd et al. (2012) produced two major conclusions about criminal career of places. First, crime is not only concentrated at small places, these "hot spots" remain stable across time. In fact, the majority of places feature extremely stable crime patterns longitudinally. This means that crime prevention practitioners can focus their resources on relatively few crime hot spots and deal with a large proportion of the crime problem. Importantly, places are not "moving targets." Place-based crime prevention provides a target that "stays in the same place." Thus, crime prevention policies targeting high-risk places could be beneficial. Secondly, despite the stability and continuity of crime rates at most places, there is still a great deal of variability *across* street segments over time in Seattle. Thus, it is important to understand factors that determine the changes in crime at those places.

What Factors Predict Crime Change?

While scholars have provided a strong empirical basis for the assumption that crime is strongly clustered at crime hot spots and that there are important developmental trends of crime at place, existing research provides little insight into the factors that underlie these patterns. For example, we could identify only three prior published studies that specifically examined developmental patterns of crime at micro places over time. One study conducted by Spelman (1995) looked at specific places such as high schools, public housing projects, subway stations, and parks in Boston, using 3 years of official crime information. Taylor (1999) examined crime and fear of crime at 90 street blocks in Baltimore, Maryland, using a panel design with data collected in 1981 and 1994. These studies are limited only to a small number of locations and to a few specific points in time. To provide a comprehensive examination of risk and protective factors that could affect the developmental patterns of crime at places, Weisburd et al. (2012) collected 25 data sources which represented key concepts from two major place-based theoretical perspectives – social disorganization theory and opportunity theory. They examined relationships between the developmental crime patterns derived from group-based trajectory analysis and social disorganization and opportunity variables. The explanatory power between variables was compared and contrasted to identify important risk and protective factors in explaining specific development processes of crime at place. Multinomial logistic regression was used to compare the two most distinct crime places, the crime-free places versus the chronic-crime segments, and see what factors best distinguish between the two places. The following sections report on findings from a comprehensive study done by Weisburd et al. (2012).

Opportunity Measures and Crime Hot Spots

The importance of opportunity theories for understanding crime at place has a long history in criminology (Brantingham and Brantingham 1984; Clarke 1983; Cohen and Felson 1979). A focus on crime naturally leads scholars to specific places or situations and the opportunities that situations and places provide for crime. Based on opportunity theory, the key elements of the crime triangle are motivated offenders, potential targets, and capable guardians (Cohen and Felson 1979).

To test the validity of opportunity theory in predicting crime trends at places, Weisburd and colleagues applied the theoretical framework to the Seattle data and found a general support for the theory. Firstly, an increase in motivated offenders, as represented by an increase in high-risk juveniles on a street segment, has a strong and significant impact on the likelihood of the street segment being classified into the high-rate chronic hot spots pattern. Indeed, for every additional high-risk juvenile found on a street segment, the likelihood of being in this group as opposed to the crime-free group more than doubles. A positive change over time in the number of high-risk juveniles on a street segment also increases the likelihood of being a chronic-crime pattern street segment. Of course, the fact that high-risk juveniles on a street increases crime risk does not mean that these juveniles are the culprits. It may be their friends who commit crimes on the street in the course of their routine activities (possibly visiting the high-risk juveniles who live on a particular street segment).

Secondly, with all else being equal, it would be expected that as the number of suitable targets increases, the number of crimes would also increase. Weisburd and colleagues used four measures to capture the number and attractiveness of targets on a street segment: employment, public facilities, residential population, and retail business sales. The results show that employment is the single most important variable in explaining the likelihood of a street falling in the high chronic-crime group as opposed to the crime-free group. For every additional employee on a street segment, the odds of falling in this pattern increase by 8 %. The change in number of employees over time also has a strong and significant impact in the same direction, though of a somewhat smaller magnitude. The presence of public facilities within a quarter mile of a street segment also significantly increases the probability of being in the chronic-crime-trajectory pattern. Having a public facility such as a community center, park, library, middle or high school, or a hospital within a quarter mile increases the likelihood of being in the chronic-crime group (as contrasted with the crime-free group) by almost 25 %.

Residential population also has a very strong impact on the likelihood of being a crime hot spot in the data. As would be expected by opportunity theory, the larger the residential population, the more likely a street segment is to be in the chronic-crime group. In contrast, the measure of retail sales was not statistically significant in predicting street segments to be in the chronic high group. It may be that the amount of retail sales does not reflect the number of patrons or visitors to the street or that after accounting for employment, public facilities, and residential population, most of the variability has been captured on street segments. Nonetheless, the presence of suitable targets as indicated by employees, facilities, and residential population is a key factor in explaining crime hot spots.

The findings for guardianship, however, are less clear. Guardianship was measured by the existence of a police or fire station within a quarter of a mile of a street segment. Another common measure used in the situational crime prevention research is street lighting, which was measured by wattage. However, both measures of guardianship reached statistical significance, but not in the expected direction. The presence of a police or fire station increases the likelihood of a street segment being in the chronic-crime pattern. Additionally, more wattage of lighting on a street is associated with a higher likelihood of being a crime hot spot. Weisburd et al. argue that these inconsistent findings could be a result of the fact that the presence of police/fire station and high wattage lighting are both more likely to be found in areas with higher population density. Moreover, it is also possible based on situational crime prevention theory that having a police station, or increased street lighting, in location with more crime problems were responses to higher crime rates. Thus, the findings could be spurious and confounded with population density or prior crime problems.

In addition to the key elements of opportunity theory, it is also found that accessibility and urban form, like types of street, are significant variables in predicting that a street segment will be a crime hot spot. For example, the number of bus stops and arterial roads both increase the likelihood of being in the chronic-crime group as contrasted with the crime-free group.

Social Disorganization and Crime Hot Spots

As mentioned earlier in the entry, social disorganization theory is another major theoretical approach for the study of places. A series of variables reflecting structural components of social disorganization have been tested in the past, including socioeconomic status, population heterogeneity, mobility, and female-headed households. A neighborhood that is poor, more disadvantaged, heterogeneous in terms of racial or ethnic composition, and more urbanized tends to be more vulnerable to crime and other social problems. Using the Seattle data, Weisburd et al. (2012) tested social disorganization theory at the street-segment level to see whether the core theoretical concepts help predict chronic-crime places. They found that socioeconomic status represented by the value of residential property on a street segment and the amount of housing assistance is both strongly and significantly related to a street segment falling in the chronic-crime group.

The other two structural dimensions reflecting social disorganization were mixed land use and racial heterogeneity. However, both of these measures were not statistically significant in differentiating between the chronic-crime and crime-free groups. In terms of the extent of urbanization, it was captured by the distance of the street away from the city center. It had an overall impact on predicting crime, in that areas closest to the center of the city had the highest crime rates. These were the areas where new immigrants and poorer residents were concentrated, where social control was weak, and juvenile delinquency and crime problems would accordingly be concentrated. In addition to those structural characteristics, Weisburd et al. also included description of physical conditions in the model. Physical disorder such as litter, trash, graffiti, and abandoned cars are the most direct indicators of social disorganization (Shaw and McKay 1942[1969]), and its relationship to developmental trajectories of crime at street segments is very strong. A street segment is much more likely to be in the chronic-trajectory pattern as opposed to the crime-free pattern if it has higher reports of physical disorder incidents. Additionally, increases in physical disorder are also positively associated with higher likelihoods of being in the chronic-crime trajectory.

Recent conceptualizations of social disorganization theory draw distinctions between the structural characteristics of areas and the mediating factors that connect the structural factors and outcome variables like crime (see Bursik and Grasmick 1993; Sampson et al. 1997; Sampson and Groves 1989). The strength of social ties among residents, also called social capital, determines the extent to which social control functions in areas. Different measures identified in prior studies have been used to conceptualize the intermediating mechanism including participation in local organizations (Sampson and Groves 1989), willingness (or perception of responsibility) to intervene in public affairs (Sampson et al. 1997), local friendship networks (Sampson and Groves 1989), mutual trust (Sampson et al. 1997), and unsupervised teens wandering on the street (Sampson and Groves 1989; Sampson et al. 1997). These factors are believed to condition the effects of structural disadvantage on local crime problems.

The empirical validity of these measures has been tested in many studies. However, Weisburd and colleagues (2012) were first to test the idea at the micro geographic level over a long period of time. Based on the data, they found that truant juveniles on a street segment significantly increased the likelihood of it being in a crime hot spot, while high level of collective efficacy (as measured by percentage of active voters) decreased the chance of a place being a crime hot spot. It seems, accordingly, that the more involved the residents are in public affairs, the less likely the streets are to have chronic-crime problems. Thus, social control could be built at street level to reduce the crime problem.

Conclusion

Overall, research has found that crimes are clustered in a small number of places, regardless of the geographic unit of analysis examined. The "hot spots" phenomenon suggests that we could identify and deal with a large proportion of crime problems by focusing on just a very small number of places. In addition to the concentration of crime, recent studies have found that crime hot spots remain very stable over time. However, despite the preponderant evidence for crime stability at places, other scholars have also identified a smaller proportion of places with changing crime rates (Griffiths and Chavez 2004; Weisburd et al. 2004, 2012). These studies find that changes in social characteristics and contextual factors can also lead to change in crime rates at places. Specifically, social disorganization theory and opportunity theory are both relevant in predicting crime at places. For example, the number of residents, unsupervised teens wandering the streets, the more (or less) committed citizens are to public affairs, and having more potential crime targets in an area all determine the occurrence of crime at any given place.

Another recent development of place-based criminology is the emphasis on studying micro geographic areas like addresses or street segments, rather than larger geographic units such as communities, neighborhoods, or census tracts. Weisburd et al. (2012), Braga and colleagues (2010), Taylor (1999), and other scholars have demonstrated a stronger crime concentration effect at micro places and have thus argued that focusing on micro places could be promising and more efficient for crime prevention efforts.

Crime at place is very predictable, and therefore, it is possible not only to understand why crime is concentrated at place but also to develop effective crime prevention strategies to ameliorate crime problems at places. Criminologists and crime prevention practitioners can identify key characteristics of places that are correlated with crime. At a policy level, it is important to focus on initiatives like "hot spots policing" that address specific streets within relatively small areas. If police become better at recognizing the "good streets" in the bad areas, they can take a more holistic approach to addressing crime problems.

Related Entries

- ▶ Criminology of Place
- ▶ History of Geographic Criminology Part I: Nineteenth Century
- ▶ History of Geographic Criminology Part II: Twentieth Century
- ▶ Hot Spots and Place-Based Policing
- ▶ Law of Crime Concentrations at Places
- ▶ Longitudinal Crime Trends at Places

Recommended Reading and References

Braga AA, Papachristos AV, Hureau DM (2010) The concentration and stability of gun violence at micro places in Boston, 1980–2008. J Quant Criminol 26(1):33–53

Brantingham PL, Brantingham PJ (1984) Patterns in crime. Macmillan, New York

Bursik RJ Jr, Grasmick HD (1993) Economic deprivation and neighborhood crime rates, 1960–1980. Law Soc Rev 27(2):263–284

Bursik RJ Jr, Webb J (1982) Community change and patterns of delinquency. Am J Sociol 88:24–42

Clarke RV (1983) Situational crime prevention: its theoretical basis and practical scope. Crime Justice 4:225–256

Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44:588–608

Eck J, Gersh J, Taylor C (2000) Finding crime hot spots through repeat address mapping. In: Goldsmith V, McGuire P, Mollenkopf J, Ross T (eds) Analyzing crime patterns: frontiers of practice. Sage, Thousand Oaks

Griffiths E, Chavez JM (2004) Communities, street guns and homicide trajectories in Chicago, 1980–1995: merging methods for examining homicide trends across space and time. Criminology 42(4):941–978
Guerry A-M (1833) Essai sur la statistique morale de la France. Crochard, Paris
Nagin DS, Land KC (1993) Age, criminal careers, and population heterogeneity: specification and estimation of a nonparametric, mixed Poisson model. Criminology 31(3):327–362
Parent-Duchatelet AJB (1836) De la prostitution dans la ville de Paris. Dumont, Bruxelles
Quetelet A (1831) Recherches sur le penchant au crime aux diff´erents ˆages. Hayez, Brussels
Quetelet A (1833) La precision des r ´ esultats croit comme la racine ´carree du nombre des observations. Annales d'Hygiene Publique 'et de Medecine L ´ egale ´ 9:308–336
Reiss Jr. AJ (1986) Why are communities important in understanding crime? In: Tonry M, Morris N (eds), Communities and crime. Crime and justice: a review of research, vol 8. University of Chicago Press, Chicago, pp 1–33
Reiss AJ Jr, Tonry M (1986) Communities and crime. University of Chicago Press, Chicago
Sampson RJ (1993) Linking time and place: dynamic contextualism and the future of criminological inquiry. J Res Crime Delinq 30:426–444
Sampson RJ, Groves WB (1989) Community structure and crime: testing social disorganization. Am J Sociol 94(4):774–802
Sampson RJ, Raudenbush SW, Earls F (1997) Neighborhoods and violent crime: a multilevel study of collective efficacy. Science 277:918–924
Schuerman L, Kobrin S (1986) Community careers in crime. In: Reiss AJ Jr, Michael T (eds) Communities and crime, vol 8, Special refereed issue of crime and justice. University of Chicago Press, Chicago, pp 271–311
Shaw CR, McKay HD (1942) Juvenile delinquency and urban areas. University of Chicago Press, Chicago
Sherman LW (1995) Hot spots of crime and criminal careers of places. In: Eck J, Weisburd D (eds), Crime and place. Crime prevention studies, vol 4. Willow Tree Press, Monsey, pp 35–52
Sherman LW, Weisburd D (1995) General deterrent effects of police patrol in crime 'hot spots': a randomized study. Justice Q 12(4):625–648
Sherman LW, Gartin PR, Buerger ME (1989) Hot spots of predatory crime: routine activities and the criminology of place. Criminology 27(1):27–55
Sherman LW, Schmidt JD, Rogan DP, Gartin PR, Cohn EG, Collins DJ, Bacich AR (1992) The variable effects of arrest on criminal careers: the Milwaukee domestic violence experiment. J Crim Law Criminol 83(1):137–169
Spelman W (1995) Criminal careers of public places. In: Eck JE, Weisburd D (eds) Crime and place: crime prevention studies, vol 4. Willow Tree Press, Monsey
Taylor R (1999) Crime, grime, fear, and decline: a longitudinal look. Research in brief. National Institute of Justice, Washington, DC
Weisburd D, Amram S (forthcoming) The law of concentrations of crime at place: the case of Tel Aviv-Jaffa. Police Pract Res
Weisburd D, Mazerolle LG (2000) Crime and disorder in drug hot spots: implications for theory and Practice in policing. Police Q 3(3):331–349
Weisburd D, Bushway S, Lum C, Yang S (2004) Trajectories of crime at place: a longitudinal study of street segments in the city of Seattle. Criminology 42(2):283–321
Weisburd D, Morris N, Groff E (2009) Hot spots of juvenile crime: a longitudinal study of arrest incidents at street segments in Seattle, Washington. J Quant Criminol 25:443–467
Weisburd D, Groff ER, Yang S (2012) The criminology of place: developmental patterns and risk and preventive factors. Oxford University Press, New York
Wolfgang ME, Figlio R, Sellin T (1972) Delinquency in a birth cohort. University of Chicago Press, Chicago
Yang S (2010) Assessing the spatial-temporal relationship between disorder and violence. J Quant Criminol 26(1):139–163

Criminal Cases Review Commission

▶ Innocence Commissions

Criminal Defense Profession

Alissa Pollitz Worden and Andrew L. B. Davies
School of Criminal Justice, University at Albany,
State University of New York, Albany, NY, USA

Overview

A defendant's right to legal counsel in a criminal prosecution is one of the most familiar of the Constitution's due process protections. In 1963, the Supreme Court declared in *Gideon v. Wainwright* that

> There are few defendants charged with crime, few indeed, who fail to hire the best lawyers they can get to prepare and present their defenses. That government hires lawyers to prosecute and

> defendants who have the money hire lawyers to defend are the strongest indications of the widespread belief that lawyers in criminal courts are necessities, not luxuries.

This entry reviews the criminal defense profession. The first section reviews the historical origins of counsel in the English legal system, tracing its evolution into colonial courts. The history of the American legal profession is marked by the rapid changes of a nation that adjusted, over the course of a few generations, to political independence, frontier expansion, abolition of slavery, immigration, and the Industrial Revolution. The tiny legal profession of the 1700s grew rapidly during the 1800s and by 1900 had evolved into a stratified guild system that remains in place today. The second section describes the contemporary structures and characteristics of the criminal defense bar and examines how lawyers define their work and how they think about their clients. The third section considers means of measuring the quality of legal representation, comparing legal, occupational, performance-based, and client perspectives on effective (and ineffective) counsel. A substantial majority of criminal defendants are represented by publicly paid counsel, through a patchwork of state and local policies and programs; the fourth section describes the variety of these systems, maps the boundaries between public and privately retained representation, and reviews the empirical literature that compares the performance of lawyers working in these diverse settings. Two concluding sections explore the political context of criminal defense work and current and emergent challenges in defining the scope and appropriate roles for criminal lawyers in contemporary courts.

The Criminal Defense Profession in the United States

The Evolution of Criminal Defense

Most of United States legal structure and doctrines were borrowed and adapted from the common law and practices of England. Medieval adjudication processes were governed by clergy (and attended largely to violations of religious laws), and the inquisitorial nature of those proceedings reflected faith in divine intervention to guide the hand of justice. But following the consolidation of English villages under William the Conqueror, law enforcement and prosecution increasingly targeted those who violated the King’s laws, not those of the church. A more secular and compartmentalized judicial system took root, staffed by appointed justices and eventually moderated by lay juries. While the origins of today’s adversarial and jury-based processes are not completely documented, it appears that the shift to formal adversarial procedures by the fifteenth century established the distinctive roles of prosecution (representing in either a private or public capacity the interests of the complainant) and defense lawyer (speaking on behalf of the accused).

However, lawyers’ access to the courts was more limited than their access to clients (Donahue 1964). English lawyers were formally recognized as practitioners at the Inns of Courts by the 1400s. But even when retained for a fee, their presence in the courtroom was limited and sometimes prohibited, in part due to fears that their participation would obscure, not reveal, the truth. In subsequent centuries, English custom wavered on the issue: common law tended to treat lawyers as important if not essential, but legislators and judges often did not. In 1789, the authors of the US Constitution rejected British ambivalence about the right to counsel when they drafted the 6th Amendment, establishing an unambiguous right to retain an attorney:

> In all criminal prosecutions, the accused shall enjoy the right to a speedy and public trial, by an impartial jury of the State and district wherein the crime shall have been committed, which district shall have been previously ascertained by law, and to be informed of the nature and cause of the accusation; to be confronted with the witnesses against him; to have compulsory process for obtaining witnesses in his favor, and to have the Assistance of Counsel for his defence.

Originally, the guarantee of assistance of counsel meant that counsel could, and should, be present at all stages in criminal proceedings in federal prosecutions. Later, this right was extended to state prosecutions, though many of the colonies had such a guarantee in place since the early 1700s. In principle, this history suggests the triumph of earlier English common law reasoning that counsel was an essential element of due process. Until the twentieth century, however, this principle coexisted uneasily with economic facts about American lawyers and defendants: the former were private businessmen, engaged in a for-profit profession, and the latter were often too poor to buy their services. There is scant historical evidence about the fates of defendants who went to court without paid counsel during the 1800s and early 1900s; one can only speculate that the legal services they received were perfunctory at best. Contemporary ideals about the right to counsel (regardless of ability to pay) did not take hold in the United States until the twentieth century (see, e.g., Smith 1919). A cascade of landmark cases established, at least in principle, the obligation of states to provide counsel in capital cases (*Powell v. Alabama*, 1932), felonies (*Gideon v. Wainwright*, 1963), many juvenile proceedings (*In re Gault*, 1967), and most misdemeanors (*Argersinger v. Hamlin*, 1972) (for a very useful review, see Drinan 2009).

When the Constitution was ratified, the American legal profession was small, and its practitioners were known more for scholarship than for their legal practice. This changed quickly, as populations grew, moved westward, and established settlements that were swiftly incorporated into new states. The need for legal advice and advocacy exploded during this era of rapid social change: lawyers were needed to settle property disputes, manage business deals, and of course defend miscreants and troublemakers. New villages and new statehouses also bred new opportunities for aspiring politicians, and quickly lawyering became not only a profitable profession for adventurous young men but also a path to public office, insofar as it provided courtroom opportunities for showcasing oratorical skills (Friedman 2007).

The barriers to entering the profession were low in the nineteenth century. In most states, after studying law with a local judge, a young man had only to earn the approval and certification of that judge to join the bar. This apprenticeship system began to fade in the mid-1800s, when attorneys began to recognize the value of more restrictive standards for admission to the bar. Within a generation, enterprising educators established law schools, advertising for students and institutionalizing the notion that formal education was requisite for practice. This development took a momentous turn in 1876 when Harvard reconstructed its law school curriculum, under the leadership of Christopher Columbus Langdell, to focus less on the practical issues of practice and common law understandings in favor of the "science" of precedent, jurisprudence, and legal logic. This educational model – still known as the Langdell method – quickly caught on in the curricula of new law schools. With a few adjustments, this model guides the standard 3-year degree program of today. This method did rather little to prepare lawyers for the reality of courthouse negotiations and client management, and even today most law graduates endure grueling postgraduation study programs to actually pass state bar exams. But Langdell left an important legacy: lawyers were thereafter honored with the academic status of *Juris Doctor*.

During the three decades following Langdell's revolution, state bars began to establish educational requirements for admission to the bar. This move paralleled those of other professions (such as medicine) that sought to increase legitimacy and prestige by establishing higher standards for professional certification. By the 1920s, most states required passing a standardized exam to practice law (Friedman 2007).

By the early 1900s, the legal profession experienced another significant change that had implications for criminal defense practice: the field began to specialize and stratify in alignment with the social and economic status of clients. The solo general practitioner was gradually supplanted by firms that marketed their services to corporations, local businesses, and upper-middle-class families; their work involved handling contracts, taxes, probate, and other work organized around financial acquisition and planning. At the lower end of the status continuum stood clients whose legal problems tended to be immediate, unexpected, and problematic: family crises, injuries and accidents, and, of course, arrests and criminal charges (Heinz and Laumann 1994; Heinz et al. 2005). This trend has continued into the twenty-first century: research suggests that, with some important exceptions, lawyers who take on criminal clients occupy low rungs on the professional ladder, are less likely than others to practice in large firms, and are most likely to combine criminal work with other "personal plight" areas of law, such as personal injury.

Characteristics of Contemporary Criminal Defense Lawyers

Americans are ambivalent about lawyers generally and criminal defense lawyers in particular. Shakespeare (Henry VI, Part 2) and, reportedly, Jesus had few kind words for lawyers (Book of Luke 11:46). One of the most prominent scholars in the field of legal studies recently published an academic treatise on the history of lawyer jokes (Galanter 2005). But Americans also idealize lawyers. Abraham Lincoln's homespun oratorical skills were crafted in village courthouses and bought him enough political capital to become a presidential contender. In the twentieth century, fictional defense lawyers such as Atticus Finch and Perry Mason were heroic figures who resolutely confronted skeptical jurors and riled communities to represent clients who, though innocent, faced long odds. Today, law school has become increasingly expensive and entry-level positions more scarce, and yet law school applications and enrollment have inclined steadily over the past 50 years.

The most recent census data on employment reports that the nation has over one million practicing lawyers (excluding those who may hold law degrees but do not represent clients; Bureau of Labor Statistics 2011), a number that has doubled in three decades. But a contemporary census of the legal profession would provide only a partial picture of the criminal defense bar. This is in part because specializations are not uniformly recorded, in part because the mix of cases that lawyers accept varies greatly, and in part because there are few sharp lines between public defenders – full-time salaried attorneys representing the poor on behalf of the state – and attorneys who represent criminal defendants under less-structured programs, in combinations of publicly paid and privately retained arrangements.

Who are these lawyers, where do they work, and how do they view their clients? Private criminal defense work remains concentrated in small local firms. Over the past 50 years, the expansion of indigents' right to counsel has meant that a large share of criminal defense cases that once went unrepresented or nominally represented by private practitioners is now in the hands of publicly paid attorneys. National data on practice settings indicate that lawyers who self-identify as legal aid or public defender attorneys constitute only 1 % of practicing attorneys, a small subgroup of the 6 % who begin their careers practicing any sort of public interest law (National Association for Legal Professionals 2006). Between 1980 and 2000, the percentage of all lawyers who worked in small private firms (of fewer than ten practitioners), where one might expect to find most private criminal defense work, has remained around 50 % (American Bar Association 2010).

The profession remains predominantly white: today only 5 % of practicing lawyers are African American, and 4 % are Asian; independent of racial identity, 3 % identify as Latino. In 2010, women comprised 32 % of the licensed bar (Bureau of Labor Statistics 2011), a percentage that has risen steadily since the 1970s as law school cohorts have become nearly 50 % female. Some have argued that a largely white, middle-class profession lacks the perspective to effectively represent defendants who are disproportionately low-income people of color, and a small body of empirical research suggests that lawyers in general, but particularly white lawyers, have lower expectations about the outcomes of cases of hypothetical black clients (see, e.g., Eisenberg and Johnson 2003).

Research also suggests that women and men differ slightly in their perspectives on the criminal defense work: men more fully embrace a strong due process orientation, and women are more likely to perceive moral dilemmas working with criminal clients (Siemsen 2004). But generally, interviews with criminal lawyers reveal higher levels of commitment to due process and opposition to punitive sentencing policies (Worden 1998; Ehrhard 2008). Contradicting stereotypes of co-opted public defenders, empirical studies report zealous advocacy orientations in this group (McIntyre 1987; Emmelman 1996).

Defining Effective Counsel and Performance

The Supreme Court of the United States has held that the Sixth Amendment requires not only that counsel be present but also that it be "effective." In the words of Justice O'Connor in *Strickland v. Washington* (1984), it is not enough that "a person who happens to be a lawyer is present at trial alongside the accused." This question has been addressed not only by the courts but also by professional associations who have developed standards and guidelines on what it means to provide counsel effectively. Empirical work, meanwhile, has examined some of the assumptions which underlie these standards – in particular that effective representation will result in more favorable outcomes for the accused.

Constitutional/Appellate Court Definitions

Writing for the majority in *Strickland v. Washington*, Justice O'Connor specified a two-pronged test to determine whether the assistance of counsel the defendant received had been ineffective.

> First, the defendant must show that counsel's performance was deficient. This requires showing that counsel made errors so serious that counsel was not functioning as the "counsel" guaranteed the defendant by the Sixth Amendment. Second, the defendant must show that the deficient performance prejudiced the defense. This requires showing that counsel's errors were so serious as to deprive the defendant of a fair trial, a trial whose result is reliable.

The holding in *Strickland* was intended to set a high standard for claims of ineffectiveness. It stifled defendant claims by stipulating that courts should generally defer to the judgment of trial attorneys' strategic reasons for presenting (or not) certain defenses. In Washington's case, the trial attorney had chosen not to enter mitigating evidence because of the danger that other prejudicial facts would be brought to light, and the Court held that this constituted a "strategic" exception. Moreover, the facts omitted were, in the opinion of the Court, insufficiently likely to have changed the verdict in the case. Accordingly, the appellant was held not to have met either prong of the test.

The court returned to the question of when a decision might reasonably be considered "strategic" in *Wiggins v. Smith* (2003). The defense attorney was held to have erred by ceasing his search for mitigation evidence on the basis of a "strategic" focus on proving the client's innocence. The court cited the capital defender standards promulgated by the American Bar Association, assessing the attorney's conduct against the standard of "prevailing professional norms," and ultimately concluded that his failure to investigate represented "inattention" rather than "strategy." Strategic decisions to cease to pursue mitigation avenues must not precede, but should only follow full investigation.

Professional Definitions

Professional groups such as the American Bar Association and many state bars have compiled lists of standards and performance guidelines that define effective representation. The National Study Commission on Defense Services, commissioned by the US Department of Justice, composed its *Guidelines for Legal Defense Systems in the United States* in 1978. These have since been followed by a range of other sets of standards and guidelines, some of which were compiled in 2000 as the *Compendium of Standards for Indigent Defense Systems*. Both the ABA and the NLADA now maintain online libraries of professional standards devoted not only to defender practice and system design but also to specific practice areas (such as juvenile or capital representation) and specific practice elements (such as the need to advise clients of collateral consequences of conviction such as employment restrictions and placement on offender registries, and the need to sustain a healthy attorney-client relationship).

Professional standards stipulate a variety of practice characteristics ranging from attorney qualifications, client interactions, case preparation methods, and contact with other court professionals. In addition, standards have been developed concerning the organization of defender systems themselves. These have focused less upon the work defenders do and more upon the systemic burdens under which they practice: standards of supervision and training, resources available for the investigation and preparation of cases, and, most prominently of all, caseloads.

Performance-Based Definitions

Empirical studies of the quality of indigent defense tend to focus on the observable characteristics of cases which are understood to represent quality in defender services, as defined by the legal and professional standards aforementioned. Examples include the extent of motion practice, early intervention by counsel, the time spent by attorneys preparing cases, investigative work, client-attorney interactions (their number, duration, and quality), and the provision of ancillary services, such as assistance with civil legal matters or access to social work services (e.g., Flemming 1986).

Researchers have also gone beyond the question of whether attorneys adhere to the procedures stipulated by professional standards and have attempted to find out if these better practices produce better case outcomes for clients. These outcomes are related directly to defendants' fates: convictions, sentences, bail decisions, time to adjudication, days spent incarcerated, diversion into alternative programs, and success following reentry.

Research findings on this question are inconsistent, likely reflecting the problems of sampling and measurement in court data. Nationally representative samples of defendants are hard to come by, and studies of indigent defense case outcomes are typically conducted on samples compromised by selection bias (such as incarcerated populations) or limited to the operations of specific parts of the courts system (such as federal courts) or types of cases (such as homicides). With these caveats in mind, researchers have repeatedly investigated two questions which tap common assumptions about publicly funded defenders: first, whether defendants retaining attorneys privately do better in court than those represented by publicly funded lawyers and second, whether institutional providers employing salaried attorneys (such as public defender offices) obtain better outcomes than those that make ad hoc assignments of private attorneys paid in some other way.

In contradiction to the stereotype that publicly funded defenders provide inferior services to privately retained ones, most studies have found little or no support for this thesis after differences between clients and their cases are controlled (Hartley et al. 2010). Publicly funded attorneys may obtain improved outcomes for clients in certain contexts, particularly when considering intermediate case outcomes such as pretrial release or charge bargaining in isolation from sentencing itself (Hartley et al. 2010). Hanson et al. (1992) found a lower likelihood of incarceration for defendants retaining private attorneys, but no difference in other outcomes. Nevertheless, the presumption that private attorneys, who may spend significantly more time on cases than publicly funded ones, obtain better outcomes for their clients has little basis in empirical evidence.

The absence of findings of differences in the quality and effectiveness of defense attorneys has prompted speculation regarding other possible differences in the working environment or methods of publicly funded counsel in comparison with privately retained attorneys. It may be the case that publicly funded attorneys, because of their frequent appearance in court, develop working relationships with other court actors which allow them to work to their client's advantage through informal means where traditional adversarial ones fail them. Flemming et al. (1992) observed the ways in which defenders necessarily formed and maintained working relationships with judges and even prosecutors. If this is the case, "repeat players" in the private bar would demonstrate similar collegial behavior and positive results for their clients.

Within public defense systems, comparisons of public defender offices with systems that rely on the private bar have also found few differences in case outcomes (Houlden and Balkin 1985). Some exceptions exist for specific demographic groups (Hartley et al. 2010), and several recent, well-designed analyses have established significant evidence of the detrimental effects for defendants of representation by assigned counsel in comparison to institutionalized providers in the context of the federal courts (Iyengar 2005) and among felony defendants in urban counties (Cohen 2011). The sources of such differences remain a point of controversy, though aside from the infrastructural advantages that public defender offices may provide, authors such as Iyengar (2005) also point to differences in the attorneys that each system attracts, and differences in the ways the operation of market forces change the composition of each. Though incipient, this research suggests that different systems may attract different defenders, who in turn obtain different outcomes for their clients.

Clients' Perceptions

Beyond examining case outcomes, other scholars, beginning with Casper (1971), began asking defendants themselves what they thought of their lawyers. These studies revealed that client perceptions are molded not only by the favorability of sentencing decisions but also by the interactions clients had with their attorneys. Predictably, defendants' skepticism – their common assumption that appointed lawyers are at best incompetent and at worst indifferent state lackeys – presents particular obstacles in building the trusting relationship presumed to be required for effective defense (Flemming 1986). Notwithstanding the well-intentioned nature of the advice defenders frequently offer their clients, the failure to couch that advice in a reassuring client-attorney relationship dulls the appreciation that defendants might otherwise be expected to feel for their attorneys.

Reassuring defendants that their attorneys are motivated, interested, and honorable is more than a matter of wanting to make clients happy, however. Researchers have also adduced evidence that the procedural fairness of court practices has other significant benefits. Defendants who are satisfied that their verdicts were arrived at via fair processes are more likely to accept their sentences, feel more confidence in criminal justice generally, and may be more likely to comply with its mandates (Casper et al. 1988). Attention to client satisfaction, therefore, is not merely a matter of pleasing consumers. It has been sustained by concern with assuring that defendants themselves – many of whom will eventually reenter society – regard the criminal justice system positively and are able to reintegrate successfully into society.

Attorney-client interactions and their consequences for defendants' evaluations have come under considerable scrutiny, and research suggests that attorney and client expectations about what constitutes effective service are at odds. Quite apart from gaining clients' confidence, an obstacle which state-appointed attorneys find difficult (Flemming 1986; Casper 1971), attorneys vary in the style and degree of attention paid to the client and their willingness to involve them in decision-making. More controversially, some scholarship has also sought to demonstrate the ways in which attorney decisions may be guided by factors other than client interests, such as the fee structure by which they are reimbursed (Tata 2007). Accordingly, the extent to which attorneys are able to foster positive client experiences and the possibility that structural factors such as the source of their reimbursement act as potential impediments to effective representation remain issues of concern.

Criminal Defense as Public(ly Funded) Function

The immediate result of the 1963 *Gideon* decision was to create a massive demand for publicly provided representation. At times this demand has overwhelmed some states and communities. The retroactive nature of the decision meant that many inmates were retried or, when this proved impractical, simply released (Eichman 1966). Quickly, however, a nationwide patchwork of defender providers was created.

Though data on the immediate implementation of the Gideon mandate are scarce, the earliest national survey revealed that by 1972, 16 states had developed statewide programs with the remainder delegating the responsibility for defense provision to counties and localities (Benner and Lynch 1973). State financial support for these systems also varied, and in 1975, states funded just 41.5 % of defense costs, leaving the remainder for localities to supplement (Sourcebook of Criminal Justice Statistics 1977). In the decades since the 1970s, the general trend has been toward the centralization of both organizational and funding responsibility for defender services within states. Recent data show that these trajectories are far from complete, however, with a total of 22 states using statewide systems in 2007, funding an average of 66 % of system costs in 2008 (Langton and Farole 2010).

National estimates on uptake of publicly funded defender services are scarce. Eligibility criteria themselves were left to the discretion of providers, with *Gideon* and subsequent decisions stipulating only that the defendant be "too poor to hire a lawyer." As a result, estimates of indigence among the population of defendants are bedeviled by a lack of standardization, as well as a more fundamental lack of systematically collected data. Moreover, states vary in their definitions of the right to counsel: in some, the right has exceeded that mandated by the Supreme Court through state statutory revisions and case law. Most estimates suggest, however, that the proportion of defendants using free legal services averages in the range from 65 % to over 80 %, depending on the type of case.

The freedom afforded states in determining the manner of delivery for defender services has resulted in a variety of structural approaches to service provision both among and within states. Historically, public defense was typically provided through informal ad hoc assignment of counsel, frequently – though not always – with little attention to attorney qualifications or compensation. These traditional "assigned counsel" systems made no meaningful distinction between private and publicly funded lawyers, since most defenders mingled publicly funded assignments with cases on private retainer. Indeed, the term "public defender" did not always connote payment from the public purse, but also applied to lawyers working pro bono to "defend" the "public." The institutionalization of states' obligation to compensate counsel led to the necessity of systematized billing processes for attorneys as well as new bureaucratic structures with new responsibilities: assigning attorneys to cases, assuring they provided representation, and paying them for their services. This formalization brought the opportunity for greater oversight and systematization in the provision of defender services through assigned lawyers, though the extent to which these systems have expanded beyond simple billing functions to oversee quality in service provision is highly variable.

Among these evolving assigned counsel programs, a now-familiar approach gained popularity: the Public defender offices are permanent offices staffed by salaried attorneys who provide legal representation in all cases assigned to them. Public defender offices were not uncommon in large cities by the early part of the twentieth century. Public defender offices seem to promise economies of scale, which probably explains why they are concentrated in urbanized areas (Worden and Worden 1989). Staff in such offices may be full or part-time and may or may not be permitted to retain private practices in addition to their work as state or local employees.

Concern with efficiency has led to yet a third approach: contracts with attorneys or law firms to provide representation. The terms of such contracts may vary, but the approach is distinctive from the archetypal assigned counsel or public defender approaches because attorneys contract with local governments (typically counties) to handle all or a percentage of the jurisdiction's caseload, during a set period of time. How they manage that caseload is up to them; how local governments choose a bidder is up to those governments. Contracting offers mixed benefits. A bidding process may reduce costs for localities, though the consequence of that process may be that quality is compromised, particularly if the contract offers a flat fee for services regardless of case complexities. In *State* vs. *Smith* (1984), a low-bid contract for the provision of defense services in Mojave County, Arizona, was struck down on the basis that selecting a provider on a pure cost basis was insufficient to guarantee effective assistance of counsel. Moreover, as Worden (1993) showed, many counties are too small to generate an effective bidding process for contracts, so contracting can actually increase rather than decrease costs. Nevertheless, contracting has become increasingly popular, resulting in considerable interest in the processes of bidding and contract composition which avert the harm to quality that might result.

A single provider of defender services is generally insufficient to provide all legal representation in a jurisdiction, however, because cases with multiple defendants present conflicts of interest. Except in the case of traditional assigned counsel systems (where each defendant can be assigned to a different lawyer from the pool), institutional defender systems need backup systems. Rules for the determination of conflicts differ, but in general a provider cannot provide representation to a defendant if it has previously represented either a codefendant or a witness in the case. Backup systems generally come either in the form of assigned counsel plans, additional institutionalized "conflict defender" offices, or firms or attorneys contracted for the purpose.

The Political Context of Criminal Defense

Crime and justice policy is, by definition, the product of political processes, and the organized bar, legal advocacy groups, and individual attorneys have been at the center of many reform debates. The American Bar Association (and many state and local associations as well) has advocated for changes in legal processes (including, but not limited to, higher standards for criminal defense of the poor). In many states, public and nonprofit organizations provide assistance to defense attorneys, notably in capital cases and appeals. Organizations such as American Civil Liberties Union, The Innocence Project, the National Legal Aid & Defender Association, and the Death Penalty Information Center draw heavily on the contributions and experience of lawyers as well as legal scholars. Many such groups provide or support provision of legal representation in individual cases and also actively lobby for policy changes in substantive and procedural law.

Most landmark changes in criminal procedural law have resulted not from lobbying and legislation, however, but from litigation. The due process revolution was waged by criminal defense lawyers representing convicted clients in high-level appeals, and these lawyers were motivated by the opportunity to change policy, typically in the direction of greater due process protections or equality of treatment for defendants. Scholars and professionals refer to this sort of work as "cause lawyering" or "impact litigation." Clearly some of these initiatives have made lasting changes in operational definitions of justice and fairness, and indeed, some experts argue that litigation is a more promising strategy for effecting legal reforms than efforts in legislatures. However, others caution that while legal rulings can dramatically redefine defendants' rights, courts have very limited enforcement powers, and litigants (and their lawyers) have little incentive to ensure that sweeping rulings are, in fact, transformed into routine practices that change how police, prosecutors, judges, and juries do their work (Scheingold 1974; but see also Scheingold 2004).

More generally, the last four decades witnessed a sharp increase in adjudication and sentencing in the aftermath of a real increase in crime rates during the 1970s and 1980s and a much more prolonged increase in punishment into the current century. Scholarship on the causes of the latter (and on variability across states and communities as well) suggests not only increasing (and sometimes overheated) awareness of crime as a policy problem but also social, economic, and cultural differences in how political entities responded to crime and justice issues (Davies and Worden 2009). The work of criminal defense lawyers is shaped by the legal contexts in which they practice. But while researchers have investigated the consequences of changing sentencing codes and caseloads for law enforcement and corrections, they know little about how prosecutors and defense lawyers have adapted to these shifts.

Current Issues and Controversies

Today's lawyers face challenges and controversies that may redefine the scope and the standards for their work with criminal defendants. Some of these are structural: at what point is representation required? Others entail revisiting the standards for effectiveness (or ineffectiveness) of counsel. New (or expanding) clienteles, defined by charges and by personal characteristics, call for new expertise and skills. In a very different vein, some argue that recoding some minor crimes into violations or civil offenses might dramatically reduce the need for lawyers, privately paid or publicly provided. Finally, defense lawyers confront new ethical and professional questions as alternative courts recast their traditional identity as adversarial champion. The examples here illustrate some of these challenges.

When Does the Right to Counsel Begin and End?

The question of when the 6th amendment right to counsel applies to cases is answered differently, in law and in practice, across jurisdictions. Counsel at trial is unquestioned, but there are legal and practical disputes about whether a defendant needs, and is entitled to, counsel in juvenile court, at arraignment and bail hearings, and in appeals (Worden et al. 2010–2011). Attorney General Eric Holder recently called for close scrutiny of court practices and customs that oblige defendants to make court appearances, possibly plead, and have bail set without the benefit of counsel. Policies on post-conviction review of cases that hinge on DNA evidence, and cases that involve sex offenders facing civil commitment proceedings, vary across states; as such petitions become more numerous, expectations for legal assistance may increase.

Revisiting "Effective Counsel" in the Context of Plea Negotiations

As noted above, "effective" counsel has historically been defined, simply, as the absence of egregiously "ineffective" counsel, and the bar for a ruling of ineffective counsel has been placed rather high. A pair of recent Supreme Court rulings signals a shift in this thinking, however. In *Laffler v. Cooper* and *Missouri v. Frye*, both decided by 5–4 votes, the court established that standards of effective representation should include judgments about the quality of lawyers' advice about rejecting plea offers in cases that ended in trial convictions.

New Clients with New Problems

Today's defense lawyers face a more complex tableau of indictments, sanctions, and evidentiary questions than did their predecessors. While the last generation's draconian drug laws are slowly being retracted, and street crime rates have fallen dramatically, today's clients present a new mix of social, cultural, and legal problems. Law enforcement's focus on illegal immigration adds a layer of complexity and anxiety to some defendants' cases – an aspect of representation given special urgency by the Court's recent *Padilla v. Kentucky* (2010) decision requiring that immigrant defendants be counseled on the immigration consequences of a guilty plea. The last decade's construction of new sex offender laws carrying collateral sentencing consequences – registries, residency restrictions, and civil commitment – raise the stakes for lawyers negotiating for their clients' futures. Lawyers (and prosecutors) face new questions about technology, evidence, and admissibility as it becomes easier to inadvertently leave an electronic trail of communications, travel, or financial transactions (and also, of course, easier for authorities to follow those trails; see, e.g., *United States v. Jones* (2012)).

Reformed Legal Codes and Fewer Lawyers?

The cost of providing lawyers in minor cases is causing some states to revisit the question of whether they are necessary at all. In reforms referred to collectively as "reclassification" or "decriminalization," minor offenses such as sleeping in a public park, low-level harassment, or, most conspicuously, marijuana use have been reduced to the level of civil infractions without the possibility of jail time. (Absent a threat of incarceration, the Supreme Court does not compel states to provide legal counsel.) Reclassification is often accompanied by elimination of other legal consequences of conviction – for example, by providing assurances that the case will not create or add to one's criminal record. Some believe that reclassification saves public tax dollars even as it reverses the criminal stigmatization of minor offenders. At its grandest, the move toward reclassification is seen as an attempt to begin to undo some of the overbuilt, "industrialized" aspects of the criminal justice system that sprang up in conjunction with the rise in crime and incarceration since the 1960s. This initiative also taps into the sentiment that lawyers, and indeed courts, may be unnecessary for the resolution of many minor infractions and disputes.

New Models of Adjudication and Legal Representation

Specialized courts (sometimes called therapeutic courts) are designed as rehabilitative alternatives to the adversarial adjudicative process. Since the late 1980s, drug courts, domestic violence courts, and mental health courts have proliferated. While they vary in many ways, they have in common a commitment to addressing defendants' social, psychological, medical, and economic problems as a route away from recidivism. A central tenet of most of these courts is the necessity of a collaborative effort on the part of defendants, judges, social services, prosecutors, and defense lawyers. These experiments offer the appealing promise of destigmatization, reform, and resource recovery (since rehabilitated offenders will commit fewer future crimes and require fewer criminal justice resources to process and punish). However, they also create new professional and ethical questions for defense lawyers. A spirited defense of a client's innocence or an attack on a prosecutor's evidence is hardly the fitting prelude to this sort of diversion and treatment; indeed, a guilty plea and, typically, some evidence of remorse are prerequisites for these alternatives. These specialized courts have typically identified particular psychological and medical problems (addiction, mental illness) or specific offenses (domestic violence, prostitution). The newest experiment in this genre is the veterans' court, which defines the experience of combat as a special condition meriting specialized treatment.

Conclusions

From its beginnings as a modest and tradesmanlike craft, defense lawyering has become much more than only a profession. It has come to stand for both the worst and the best of the law: the morally questionable imperative to provide criminals with expert assistance in avoiding conviction and punishment, on the one hand; and, on the other, some of the Supreme Court's finest prose on the need to equalize access to justice for rich and poor, state and citizen, alike.

This ambivalence of idealism mixed with the appearance of disreputability continues to characterize the defense bar to this day. In order to pursue a case to the highest court and secure a significant policy victory, lawyers must be prepared to subject a single client to years of litigation – instead of simply encouraging them to plead guilty and dispose of their matter efficiently. Advocates argue that defense reform will not be complete until every criminal defendant is afforded the full criminal trial to which they are nominally entitled. But of course, lawyers stand to benefit a great deal from expansion of the right to counsel and guarantees of more generous compensation. What results, therefore, is a continuation of old arguments over how much defense counsel is really necessary to justice: whether it serves clients or causes and whether it protects the guilty or the innocent. These arguments conjure up incongruous images: on the one hand, courageous Atticus Finch in a hostile courtroom, and on the other, florid fast-talking lawyers promising fast relief for drunk driving arrests on late-night advertisements.

One reason that these old arguments have been so hard to resolve is that defense lawyering, unlike many other elements of the criminal justice system, has not been subjected to significant empirical scrutiny over its central claims. Police organizations, for example, have sought sophisticated approaches to proving their effectiveness; offender treatment programs are regularly subjected to performance evaluations. But the defense bar has engaged in far less self-examination, and outsiders get few opportunities to evaluate lawyers' work.

Furthermore, the realities of criminal practice are more complex than most people, including lawyers themselves, may acknowledge. Lawyers' place in adversarial adjudication is straightforward enough: zealous defense of clients' interests, which usually means strategic pursuit of acquittal or minimal punishment. But lawyers usually operate in the more complex context of plea negotiations and in a court system that has become increasingly punitive, even as (especially recently) alternative adjudication systems, diversion programs, and treatment options have proliferated. Serving clients' "best interests" has become more complicated.

Defense lawyers have a unique interest in protecting their clients from the cycle of cumulative disadvantage – the hardships that start with the stigma of arrest, accrue with conviction and sanctions, and are cemented when reintegration and rehabilitation fail. Simple structural reforms – such as organizational changes which assure counsel reaches a defendant early or reclassification measures which assure defendants are not detained needlessly – can make a difference in whether a defendant is able to go home, keep a job, maintain custody of children, and even reintegrate into society in the long term. Broader awareness of these changes (and evaluations of their effectiveness) will not resolve historical, cultural, and economic dissension over defense lawyers' roles. It might, however, emancipate lawyers from simplified ideological prescriptions about their work and lead to a more nuanced and realistic public understanding of their work.

Related Entries

- ▶ Comparative Legal Cultures
- ▶ History of the Death Penalty
- ▶ Innocence Commissions
- ▶ Juries, Lay Judges, and Trials
- ▶ Measuring Wrongful Convictions
- ▶ Mental Health Courts
- ▶ Plea Bargaining
- ▶ Problem-Solving Courts
- ▶ Public Defenders
- ▶ Scientific Evidence in Criminal Adjudication
- ▶ Scientific Evidence in Criminal Prosecutions
- ▶ Sentencing Research
- ▶ Sex Offenders and Criminal Policy
- ▶ Therapeutic Jurisprudence

Recommended Reading and References

Benner L, Lynch B (1973) The other face of justice. National Legal Aid and Defenders Association, Washington, DC

Bureau of Labor Statistics (2011) Occupational employment statistics. http://www.bls.gov/cps/cpsaat11.pdf

Casper J (1971) Did you have a lawyer when you went to court? No, I had a public defender. Yale Rev Law Soc Action 1:4–9

Casper J, Tyler T, Fisher B (1988) Procedural justice in felony cases. Law Soc Rev 22:483–507

Cohen T (2011) Who's better at defending criminals? Does type of defense attorney matter in terms of producing favorable case outcomes. US Bureau of Justice Statistics, Washington, DC, Working Paper Series. http://freealexander.org/wp-content/uploads/2011/12/SSRN-id1876474.pdf

Davies A, Worden A (2009) State politics and the right to counsel: a comparative analysis. Law Soc Rev 43:187–220

Donahue C Jr (1964) An historical argument for right to counsel during police interrogation. Yale Law J 73:1000–1057

Drinan C (2009) The third generation of indigent defense litigation. N Y Univ Rev Law Soc Change 33:33–427

Ehrhard S (2008) Plea bargaining and the death penalty: an exploratory study. Just Sys J 29:313

Eichman C (1966) Impact of the Gideon decision upon crime and sentencing in Florida. Florida Division of Corrections Research and Statistics Section, Tallahassee

Eisenberg T, Johnson S (2003) Implicit racial attitudes of death penalty lawyers. DePaul Law Rev 53:1539–1556

Emmelman D (1996) Trial by plea bargain. Law Soc Rev 30:335
Flemming R (1988) Client games: defense attorney perspectives on their relations with clients. Am Bar Assoc Res J 1988:253–277
Flemming R, Nardulli P, Eisenstein J (1992) The craft of justice: politics and work in criminal court communities. University of Pennsylvania Press, Philadelphia
Friedman L (2007) History of American law, 3rd edn. Simon & Schuster, New York
Galanter M (2005) Lowering the bar: lawyer jokes and legal culture. University of Wisconsin Press, Madison
Hanson R, Ostrom B, Hewitt W, Lomvardias C (1992) Indigent defenders get the job done and done well. National Center for State Courts, Williamsburg
Hartley R, Miller H, Spohn C (2010) Did you get what you pay for? Type of counsel and its effect on criminal court outcomes. J Crim Justice 38:1063–1070
Heinz J, Laumann E (1994) Chicago lawyers: the social structure of the bar, revised. Russell Sage Foundation and Chicago: American Bar Foundation, New York
Heinz J, Nelson R, Sandefur R, Laumann E (2005) Urban lawyers: the new social structure of the bar. University of Chicago Press, Chicago
Houlden P, Balkin S (1985) Cost and quality of indigent defense: ad hoc vs coordinated assignment of the private bar within a mixed system. Justice Sys J 10:159–172
Iyengar R (2005) An analysis of the performance of Federal Indigent Defense Counsel. Princeton University, working paper
Langton L, Farole D (2010) State public defender programs, 2007 NCJ 228229. Bureau of Justice Statistics, Washington, DC
McIntyre L (1987) The public defender: the practice of law in the shadows of repute. University of Chicago Press, Chicago
National Association for Legal Professionals (2006) Employment patterns – 1982–2004. NALP Bulletin, June 2006
Siemsen C (2004) Emotional trials: moral dilemmas of women criminal defense attorneys. Northeastern University Press, Boston
Smith RH (1919) Justice and the poor. Carnegie Foundation, New York, Bulletin # 13
Sourcebook of Criminal Justice Statistics (1977) Washington DC: Bureau of Justice Statistics
Spangenberg Group website: http://www.spangenberggroup.com/
Tata C (2007) In the interests of clients or commerce? Legal aid, supply, demand, and 'ethical indeterminacy' in criminal defense work. J Law Soc 34:489–519
Worden A (1993) Counsel for the poor: an evaluation of contracting for indigent criminal defense. Justice Q 10:613–637
Worden A (1998) Representing the accused: professional values and professional choices of small-town lawyers. Crim Just Rev 23:1–28
Worden A, Worden R (1989) Local politics and the provision of indigent defense counsel. Law Policy 11:410–424
Worden A, Davies A, Brown E (2010–2011) A patchwork of policies: Justice, due process, and public defense across American states. Albany Law Rev 74:1423–1463

Criminal Interviewing

▶ Interview and Interrogation Methods Effects on Confession Accuracy

Criminal Investigation

▶ Investigative Psychology

Criminal Investigative Analysis

Roger L. Depue
FBI Retired

Synonyms

Crime profiling; Crime scene analysis; Crime scene interpretation; Criminal personality profiling; Profiling; Psychological profiling

Overview

This entry will discuss Criminal Investigative Analysis (CIA) from an historical perspective concentrating primarily on how it began and developed in the Federal Bureau of Investigation (FBI) Behavioral Science Unit (BSU), and continues into the present day (2013) in the FBI National Center for the Analysis of Violent Crime (NCAVC), the International Criminal Investigative Analysis Fellowship (ICIAF), and the Academy Group, Inc. (AGI).

The idea of drawing personality characteristics of the criminal from an analysis of the crime scene and the behavior exhibited there had been around since before the days of Sir Arthur Conan Doyle and his fictional supersleuth Sherlock Holmes (1887). More recently (1968), Dr. James A. Brussel, a New York psychiatrist, prepared a profile of the Mad Bomber for the New York City Police Department. It turned out to be amazingly accurate in the description of perpetrator George Metesky. Other early attempts to describe a criminal by the analysis of the behavior exhibited during the commission of the crime were carried out by psychologists and psychiatrists, especially those connected with the judicial process. These professionals were usually engaged by the judicial system to ascertain the mental capacity of a defendant in order to determine if he was competent to stand trial. They were trained in diagnostic and treatment skills and their language was often replete with technical terms, affording little insight to investigators about how and why the crime was committed.

The Early Days of Profiling

The FBI's contribution to behavioral interpretation of crime scenes began in the training division at the old post office building in the heart of Washington, D.C., in 1971. At that time, Special Agent (SA) criminologist Howard Teten, a former police officer from California, was working on the behavioral assessment of violent crime scenes in preparation for teaching classes to law enforcement officers of the FBI National Academy Program. His friend and colleague SA psychologist Patrick Mullany soon joined him in this work. The two instructors team-taught an applied criminology course to veteran law enforcement officers.

In 1972, the FBI opened the new FBI Academy complex located in the center or the United States Marine Corps Base at Quantico, Virginia. A new and more comprehensive training and educational program was instituted with courses affording college credit due to an affiliation with the University of Virginia. There were five academic units (departments) at the academy: Legal Instruction, Education and Communication Arts, Forensic Science, Management Science, and the Behavioral Science Unit (BSU).

The BSU was a multidisciplinary training unit consisting of agents with educational backgrounds in psychology, sociology, criminology, and political science. Courses were taught in subjects such as Community Relations (SSA Jack Pfaff), Urban Police Problems (SSA Larry Monroe), Crisis Management (SSA Dick Harper), Hostage Negotiation (SSAs Conrad Hassel and Tom Strentz), and the refined course of Applied Criminology that would eventually evolve into CIA. During the next few years, other members of the BSU followed the lead set by Supervisory Special Agents (SSA) Teten and Mullany and began contributing to the further development of crime scene analysis and behavioral interpretation techniques.

A Crime Epidemic in the United States

Meanwhile in the mid-1970s, violent crime had begun rising steadily in the United States, until reaching epidemic proportions in the 1980s. For example, in 1962, there were 8,404 homicides; twenty years later, there were 21,012 homicides. Homicides cleared by arrest were 93 % in 1965 and 74 % in 1982 (FBI Uniform Crime Reports). The incidence of homicide was growing by leaps and bounds, and the clearance rate (as measured by arrest) was decreasing. A relatively new phenomenon emerged called "serial murder," a term coined by BSU SSA's John Douglas and Robert Ressler. Serial murder was defined as three or more homicides committed over a period of time with an emotional "cooling off" period between murders (as compared to mass murder where multiple victims are killed at one time). The alarming statistics for other crimes of violence were following the homicide trend.

In an effort to deal with this epidemic of violence and decreasing clearance rates, the Attorney General of the United States, William French Smith (1981–85), ordered the directors of each federal law enforcement agency to submit to him ideas and plans for addressing this catastrophic state of affairs. Since the FBI BSU had been conducting training in violent criminal behavior since its inception, Executive Assistant Director John Otto turned to the BSU for input to plan a response to the directive.

As part of the BSU training programs, case consultation naturally evolved as law enforcement students asked questions about their unsolved cases. Officers would describe a crime, e.g., a brutal murder, and ask for the instructor's opinion about what kind of person would commit such a crime. The instructor would speculate about the characteristics and traits of the unknown perpetrator. Instructors from different academic disciplines began to join together to contribute ideas about an unsolved violent crime from their respective disciplines and "street" experience as former field investigators. Their techniques were refined, and they soon gained a reputation for rendering valuable assistance in the analysis and solution of unsolved cases. An informal profiling program was begun.

FBI Research on Violent Criminals and Their Crimes

During classes at the FBI National Academy and during training seminars given around the country throughout the 1970s and 1980s, police students asked questions about practical aspects of cases that could not be answered by simply turning to the existing literature of criminology. Questions such as: Does the killer return to the scene of the crime and if so, why would he do that? Is the killer in the crowd when the body is found or when the crime scene is being processed? Does a killer ever insert himself into the investigation? BSU instructors recognized that answers to these and other similar questions would give investigators the ability to make predictive estimates of criminal behavior based on a fuller, deeper analysis of the crime scene and the probable mindset of the criminal perpetrator. These and other questions caused two members of the BSU, SSA's Robert Ressler and John Douglas, to begin an informal effort to obtain some answers by going into prisons and conducting interviews of several incarcerated murderers. Convinced of the value of information obtained, they enlisted the aid of behavioral scientists (especially Dr. Ann Burgess) from major universities to design research projects and interview protocols to be used for subsequent interviews of targeted perpetrators imprisoned for having committed multiple homicides.

First, since empirical research of serial violent offenders and their crimes had never been conducted from the criminal investigator's point of view, the initial research required the thorough examination of multiple, often horrific, crime scenes of a perpetrator in great detail prior to the prison interview. The crime scenes were carefully compared to one another, noting similarities and differences. Questions about how and why they were committed were listed.

Once the detailed examination of the offender's crimes was completed, other documentations (i.e., pre-sentence investigation reports, mental health records, arrest histories, etc.) were studied. Agents realized that conducting interviews without this comprehensive preliminary work would be the equivalent of trying to gain insight into the mind of a great artist (Vincent Van Gogh) by interviewing him without ever looking at his actual work. The quality of a work of art made the artist a famous person of interest to the connoisseur and likewise, the details of the crime scenes made the killer an infamous person of interest to the investigator. An additional important reason for a close examination of the crimes concerned the accuracy of the information obtained from the offender during the prison interviews. Skilled investigators know that criminals are notorious liars, so a secondary benefit from the careful preliminary examinations was that the offender had more difficulty lying to interviewers about his crimes. Aristotle noted that those who wish to succeed must ask the right preliminary questions (Metaphysics II).

Secondly, researchers wanted to understand the developmental history of the perpetrator to try to determine how he became a violent offender. They wanted to gain insight into the behaviors and thought processes of the violent criminals. The protocols were developed to attain these ends. Interview protocols sought information about the subject in categories such as developmental history, health, family, education, occupation, mental health, arrest and conviction, and the forms of mental aberration. Questions concerned things such as at what age did violent fantasies begin, and at what age did the violent behavior start?

A subunit was formed in the BSU for conducting additional research and was initially led by SSA Ressler and later by SSA Richard Ault, Jr. Over a period of years, many BSU instructors went into the prisons with the comprehensive research protocols to conduct face-to-face interviews of violent criminals such as assassins, serial killers, rapists, child abductors, sexual sadists (and their wives), and, eventually, other criminals such as traitors and spies. The BSU agents sought to learn the "tradecraft," fantasies, and habits of the perpetrators. The first interviews concerned assassins and serial killers and were initially conducted by SSAs Robert Ressler and John Douglas. SSAs Robert "Roy" Hazelwood and Kenneth Lanning conducted the initial research interviews of rapists and child abductors. Soon the research program expanded to include espionage and was supervised by SSA Richard Ault, Jr.

The three approaches that were responsible for increasing the reservoir of new knowledge in the BSU were also the three primary tasks of the BSU itself: training, case consultation, and empirical research. These functions came to be referred to as "the three-legged stool" because each approach was interdependent with the others and simultaneously reinforced by the others. Training acquired investigative ideas and questions from course attendees that could be added to research protocols. Research interviews provided insight into the mind of the criminal regarding his thought processes and behaviors before, during, and after his crime. Casework allowed a venue for the application of new knowledge to ongoing cases. These functions, in turn, created "cutting-edge" materials for training programs. The three-legged stool became the foundation for sophisticated crime analysis and interpretation of criminal behavior that was to become known as "profiling." The profiling theory stated that behavior could be identified in the crime scene: extracted, examined, interpreted, and finally described in a profile of the unknown perpetrator.

The Profiling Process

During an investigation of an unsolved crime, the profiling process consisted of a detailed examination of all available case materials and was intended to answer four questions: (1) what happened? (a listing of all significant behaviors of offender and victim); (2) how it happened? (a reconstruction of the crime with supporting evidence for each step); (3) why it happened? (the reasons for each behavior and the formation of motive); and (4) who is responsible? (the type of person who would have committed this type of crime, in this manner, for these reasons). Characteristics of a perpetrator listed in the profile might include such descriptors as approximate age, gender, race, marital status, occupation, hobbies, vehicle information, appearance and grooming, arrest history, residential information, life style, psychological features and disorders, as well as intelligence level, emotional adjustment and ability to interact socially, etc. A fifth question was added to the above four questions: Where would such a person as described in the profile be found in the environment? This question allowed for the extrapolation of behavioral information drawn from the crime to be applied to lifestyle. Answering these questions provided insights and practical suggestions to investigators, leading to a more thorough and comprehensive investigation.

Students were arriving at the FBI Academy with unsolved violent crime cases for analysis, and additional cases were being received from throughout the country and from several nations where students who had attended BSU classes resided. At this point, it became necessary to implement a procedure to systematically handle the influx of cases. Bureau case management policies were put in place. It was not long before the caseload continued to grow and the need for additional personnel became evident. But Bureau administrators wanted to evaluate the program before proceeding.

In order to ascertain the value of the profiling program, a 1981 study was undertaken by the FBI

Institutional Research and Development Unit headed by SSA criminologist Howard Teten (the former BSU pioneer) who had been promoted to that specialized unit. Two hundred nine violent crime cases had been submitted by one hundred ninety-two law enforcement agencies to the BSU for analysis. They were followed upon by a survey to see if in the opinion of the case investigators, the crime analysis and subsequent profile furnished by the BSU had been helpful. These cases were primarily unsolved "old dog" cold cases where all traditional and logical investigative practices had been tried but failed to bring closure. It would have been remarkable if even 1 in a 100 of these cases were solved. As it was, the survey showed that 88 suspects had been identified largely because the profiling process had expanded and refocused the investigation. Even more remarkable was the fact that in 15 of the cases, profiling was reported to have identified a suspect outright and thereby led to the immediate solution of the cases. Needless to say, the funds necessary for expansion became available.

The FBI Investigative Profiler Program

Soon additional personnel were assigned to the unit. SSA John Douglas supervised a subunit of four "investigative profilers" brought into the BSU from the field and trained in the profiling process to work as permanent investigative profilers. The first four investigative profilers were SAs Jim Horn, Blaine "Mac" McIlwaine, Bill Hagmaier, and Ron Walker. In order to facilitate examination of cases from around the country, 55 field agents were brought back to the academy and trained in the process of case evaluation to facilitate submission of cases to the BSU. They were called "profile coordinators" and assisted in the examination and selection of cases in their divisions of the country to determine if they would likely benefit from the profiling process.

By the mid-1980s the term "profiling" took on a negative connotation (i.e., racial profiling) due to the practice of some law enforcement officers using racial discrimination to unfairly target minority persons suspected of being involved in criminal activity. In order to more accurately describe the investigative profiling process, it was renamed Criminal Investigative Analysis (CIA). CIA became the umbrella term that incorporated all of the tasks included in profiling (i.e., equivocal deaths analysis, cold case analysis, staged crimes, indirect personality assessment, planning interrogations, expert testimony in court, etc.).

A Crime Classification Manual was developed by the FBI BSU practitioners and associates in order to standardize terminology so that more accurate communication could take place within the criminal justice community. Examples of several efforts to distinguish terms from one another as well as to clarify understanding of terms commonly used came about. For example the term "modus operandi (MO)" was distinguished from "signature." MO was described as a learned set of behaviors the offender develops and maintains because they work operationally to facilitate the efficient commission of the crime. MOs are dynamic and malleable. "Signature" was described as the offender's "calling card" in that it was an individualized set of indicators that usually point to an offender's personality (i.e., behaviors that are not necessary to the efficient commission of the crime, but satisfy some idiosyncratic need of the offender such as taking a souvenir from the victim). They generally do not change. "Staging" was described as purposely altering of the crime scene prior to the arrival of police in order to redirect the investigation away from the most logical suspect, or to protect the reputation of the victim and the family (i.e., removal of pornographic materials from the death scene in an autoerotic death to make it look like suicide).

FBI Police Fellowship Program

Worldwide, law enforcement investigators expressed a growing interest in learning CIA skills and the Police Fellowship Program was born. The CIA skills were too complex to be taught in a short period of time. BSU SSA Roy Hazelwood, a veteran profiler, designed a program of instruction and case analysis for high-caliber, experienced investigators who would be invited to come to the FBI Academy for 1 year to work as Fellows in the BSU. The philosophy of the program was to train officers coming from large law enforcement agencies to become CIA specialists so that they could return to their departments and establish similar CIA functions there. The program began on a trial basis with the selection of one officer (Detective Sam Bowerman of the Baltimore County, Maryland, Police Department), and was soon followed by a group of four more investigators from other large police departments (Ed Richards, Texas Department of Public Safety, Ray Pierce, New York City Police Department, Dennis Cremins, Los Angeles Police Department, and Eric Witzig, Washington D.C. Metropolitan Police). Thereafter, they were followed by groups of five to six carefully selected men and women investigators from large police agencies from the United States, Canada, Australia, and Europe. Sergeant Ron McKay of the Royal Canadian Mounted Police of Canada became the first international member of the Fellowship Program.

The attendees of the Fellowship were given 3 months of classroom instruction with classes at the academy, the University of Virginia, and the Armed Forces Institute of Pathology. For the next 3 months of training, the attendees were assigned crime scenes from solved crimes of violence and asked to apply the CIA skills they had acquired. In the last 6 months, the attendees were assigned active unsolved cases to work with veteran profilers as regular CIA members of the BSU. Gradually, the yearlong program was shortened to 6 months. After most of the pioneering BSU CIA agents had retired, the program was discontinued despite its success and popularity.

Beginning of the International Criminal Investigative Analysis Fellowship

In 1985, the members of the first full Fellowship Program decided to form a group of investigators who had completed the Fellowship Program, a National Criminal Investigative Analysis Fellowship. In time, with the addition of attendees from other countries, the organization of diverse alumni became formalized and the International Criminal Investigative Analysis Fellowship (ICIAF) was formed. Their goals were simply stated: to maintain the ethical standards of the FBI program and to assure that the quality of the CIA process would be continued.

When the FBI Fellowship Program ended, ICIAF was well equipped to conduct their own CIA Fellowship training program, allowing for the acquisition of CIA knowledge and skills. Leadership in the organization, such as presidents Sam Bowerman and Ron McKay, developed a process for the careful selection of exceptionally qualified criminal investigators and affording them high-quality training. Conferences were held to bring together FBI and ICIAF specialists and other reputable forensic behavioral science speakers. Additionally, at regular regional and national meetings, ICIAF members invited law enforcement agencies to bring unsolved violent crime cases to the meetings where they were examined by ICIAF investigators for behavioral interpretation and profiling. ICIAF eventually expanded its purview to include a second division of CIA in order to deal with the developing field of geographic profiling.

The National Center for the Analysis of Violent Crime

By the mid-1980s, the BSU instructors had identified other projects to be joined to their CIA work. A Community Analysis Worksheet (a community profile) was developed to assist police executives to conduct a detailed examination of their communities for planning efficient delivery of police services (SSA Roger Depue). Since profiling had been successful for identifying and dealing with individual criminals, a group analysis protocol (a group profile) was developed to assess the characteristics and dynamics of criminal groups for purposes of successful penetration and exploitation (Depue). In 1984, following a meeting of law enforcement practitioners at Sam Houston State University, a Violent Criminal Apprehension Program (VICAP) created to track unsolved homicides and to link them to solved and unsolved cases from different jurisdictions (the idea of former LAPD homicide detective Pierce Brooks was added to the BSU). Among the first members of the VICAP team joining Pierce Brooks were investigators from police departments around the country, including Terry Green of Oakland, California; Ken Hanflan of Salem, Oregon; Jim Howlett of Charlotte, North Carolina; and Winston Norman of the Washington Metropolitan Police Department (who was an FBI BSU instructor following his retirement from the police department). The computerized Arson Information Management (AIM) system developed by Dr. David Icove was brought to the BSU, and the first attempt to use artificial intelligence to augment profiling was made with use of a VAX 11-785 state-of-the-art computer. BSU CIA investigators observed that officers working extremely violent cases often developed personal problems themselves. Because of that fact, an innovative program eventually entitled Stress Management in Law Enforcement (SMILE) was developed by BSU SSA instructors (John Mindermann, James Reese, Robert Schaefer, and James Horn) to address these concerns. An effort to project crime trends into the future was developed by SSA Bill Tafoya. SSAs Terry Ethridge, Tom O'Maley, and Joe Harpold developed innovative crime prevention strategies and added them to the BSU repertoire.

Assistant Director Jim McKenzie and his deputy, Dr. Jim O'Connor, cleared the way to bring all of the programs under one roof, as the BSU was gradually becoming a national center for the analysis of violent crime. The BSU was divided into two separate entities, one specializing in training and research (Depue) and the other in investigative support (SSA Alan "Smokey" Burgess). The original BSU had grown from 12 to 55 members. The national recognition of the innovative and important work of the BSU came in 1984 when President Ronald Reagan, speaking at the National Sheriff's Association Annual Conference in Hartford, Connecticut, announced the institution of the FBI National Center for the Analysis of Violent Crime (NCAVC) to be located at the FBI Academy at Quantico, Virginia. Its mission was to address unusual, vicious, bizarre, or repetitive violent crimes. Roger Depue, long-time chief and member of the BSU, was made its first administrator.

Additional field agents from around the country were brought back to Quantico to work as CIA specialists in the NCAVC, and others were given a 2 week training in-service in CIA fundamentals so that they could function as CIA coordinators (previously called profile coordinators) between Quantico, FBI field offices, and law enforcement agencies within their areas. CIA work was separated into specialized units involving different crimes of violence. The United States was divided into geographic regions, with NCAVC CIA agents providing oversight and direction to the FBI CIA coordinators and law enforcement investigators in the regions. During the 1990s, the FBI NCAVC grew in size, became a field office support resource, and moved offsite from Quantico to neighboring Aquia, Virginia. Eventually, the NCAVC was joined to other operational functions from the FBI Academy (e.g., the Hostage Rescue Team, Special Operations, Conflict Management, and Hostage Negotiation) and the entire operation became a separate FBI entity placed under the mantle of the new investigative support group called the Critical Incident Response Group (CIRG).

Expansion into the Private Sector

In 1989, retired BSU chief Dr. Roger Depue consulted with Dr. Bertram Brown (former Director to the United States National Institute of Mental Health and White House psychiatrist) and Conrad Hassel, Esquire, (retired chief of FBI Special Operations and Research) about the feasibility of starting a center for CIA to assist private sector organizations in dealing with problems caused by aberrant and violent behavior. Shortly thereafter, The Academy Group, Inc. (AGI) was begun as the first forensic behavioral sciences services company to bring CIA expertise to the private sector. AGI practitioners began to apply profiling techniques to the problems of violence in workplaces and schools, authorial attribution of anonymous communications, expert behavioral analysis and testimony in a wide variety of civil cases, and the analysis of criminal cases when requested by security and law enforcement agencies. Many veteran agents retiring from the NCAVC joined their colleagues at AGI such as CIA practitioners Peter Smerick (current Chairman of the Board) and R. Stephen Mardigian (current President), and Michael Napier, an interview and interrogation specialist. Fellowship alumni also joined AGI to continue doing CIA work such as Ken Baker, United States Secret Service, and Larry McCann, Virginia State Police. AGI located in Manassas, Virginia, allowed for continued contact with the FBI BSU and NCAVC.

Work of the Critical Incident Response Group NCAVC

Meanwhile back at the FBI during the early 2000s, the CIA units of the CIRG NCAVC were named Behavioral Analysis Units (BAU) and four of these units were designated each with separate responsibilities. BAU 1 contained SSA's specializing in threat assessment, especially in the analysis of written documents. Linguist SSA Jim Fitzgerald (who joined AGI following his retirement) led the field in this important work. This unit dealt with the threatening communications that often preceded or accompanied dangerous acts. Writing is a form of behavior that can be analyzed to reveal personality characteristics of the writer. Communications that are anonymous can be examined as to writing style and content and often afford important clues as to the identity of the unknown authorship.

BAU 2 specialized in violent crimes against adult victims. Most violent crimes fall into this category. Both the victims and perpetrators are adults. Also in this category there was a good deal of concentration on crimes against the other most vulnerable victims in society, the elderly (as developed by SSA Mark Safarik, Dr. John Jarvis and Dr. Kathleen Nussbaum).

BAU 3 was exclusively dedicated to crimes against children. Both adult and child perpetrators prey on children. This unit works closely with the National Center for Missing and Exploited Children. It is noted that BAU's 2 and 3 continue the work against serial murderers originally begun in the BSU of the 1980s.

BAU 4 deals with counterterrorism. Since 2001 and the terrorist attack of the World Trade Center in New York City, much emphasis and many resources have been directed toward this most serious problem. These NCAVC BAUs continue their CIA work until the present time (2013).

Meanwhile at the FBI Academy BSU, SSA Andy Bringuel instituted the Terrorist Research Analysis Project (TRAP) expanding the "group profiling" concept to a more comprehensive Group Analysis Protocol that improved understanding of foreign and domestic terrorist groups.

ICIAF Understudy Program

Under the leadership of presidents such as Ron McKay, Kate Lines, Steven Conlon, and Keith Howard (2013), the ICIAF continued to identify, select, and train new investigators from around the world in the art and science of CIA. An Understudy Program has been created whereby applicants are vetted and trained by qualified ICIAF CIA practitioners. The process includes nomination and interview of a candidate, practicum of case assessments, reading requirements, video and audio reviews and analyses, academic and field practice exercises, examinations, review of analytical notes and reports, and final individual candidate screening board interviews.

In 2001, the geographic profilers originally trained by the Vancouver Police Department (VPD), Canada, became part of the ICIAF (information provided by ICIAF Geographic Profiling Division members). They formed a separate Geographic Profiling Division with its own vice president, and eventually adopted the Geographic Understudy Program developed by VPD in 1997. The program was designed to provide comprehensive training to members of those agencies wishing to establish their own geographic profiling capability. Suitable candidates embark upon a year of study under the tutelage of a mentor who must be a fully qualified geographic profiler. The training program is divided into four blocks: (1) probability and the *Rigel* software; (2) serial crime and crime linkage; (3) violent sex offenders and criminal profiling; and (4) quantitative spatial techniques and geographic profiling.

The first three blocks are done through distance education, under the supervision of the understudy's mentor. The fourth block (four months) is a residency at the mentor's agency, involving both reviews of previous files and casework in active investigations. The understudy must pass a qualifying examination at the end of the training period in order to successfully conclude the program. Continuing education is important for the ongoing development of professional skills and expertise, and the ICIAF community, listserv, and conferences play a key role in this process. Geographic profilers from Canada, the United States, the United Kingdom, and The Netherlands are now members of the Fellowship, and they have worked on serial crime cases from around the world.

As a result, ICIAF membership continues to grow in size and sophistication (2013).

Response to Negative Critiques of Profiling

Profiling may mean different things to different people. As reported above, the FBI had an inhouse research capability, the Institutional Research and Development Unit (IRDU), to look at programs to determine if they were cost effective and whether or not they should be continued. The definition of profiling as set forth by the IRDU was: "As used in this study psychological profiling is defined as the process of identifying the gross psychological characteristics of an individual based on an analysis of the crime(s) he or she committed and providing a general description of the person utilizing those traits."

In 1981 the FBI IRDU did "a cost benefit study to determine the extent to which the service has been of value to the users." The service (profiling) was provided by special agent behavioral science investigators who were instructors at the FBI Academy (the profilers). The service was provided to the users (primarily police investigators) at their request because the crimes (primarily murder) apparently could not be solved by conventional investigative methods. From a practical standpoint FBI management was interested in whether or not the police investigators thought the service was valuable to them and should be continued to be provided by the FBI.

It should be pointed out at this time that the profiling process would have been discontinued if it was not demonstrated to have been of value to the users.

Questionnaires were distributed to 192 law enforcement agencies regarding the 209 cases that had been analyzed and profiles of unknown offenders had been prepared for them. Results revealed that in 88 cases suspect(s) had been identified and the law enforcement investigators (users) reported that the profiles assisted in focusing the investigation properly and had helped in locating possible suspects in the majority of cases. In 15 cases they reported that the profile actually identified a suspect. Needless to say the profiling program was continued.

In the FBI profiling was developed by the examination of numerous cases, analysis of the crime scenes and by identifying the behavior therein. Agents became adept at identifying the behaviors and constructing profiles of the unknown offenders. Research projects were conducted whereby agents interviewed offenders following a thorough review of crimes perpetrated by these offenders in order to gain insight into the thought processes operating in the offenders before the crime, during the crime, and after the crimes were committed.

This empirical research contributed to the profiling process being refined and expanded. The FBI then brought outstanding investigators from federal, state and local law enforcement agencies into the FBI Academy for an intensive Fellowship Program where the process and techniques discovered by the agents were shared with them.

Since that time many people have claimed to be "profilers" without working with law enforcement, and without doing the meticulous work of carefully examining numerous cases and crime scenes, and without getting feedback on the value of their efforts. Furthermore they have not analyzed large numbers of cases and crime scenes of known perpetrators and then gone into the prisons to interview these offenders in order to gain an indepth understanding of their criminal behavior.

Now "research efforts" aimed at determining the value of "profiling" are being conducted by behavioral scientists and academics.

The results of this research often cast doubts on the “profiling” process studied.

From the practitioner perspective, the one true method of determining the value of profiling is to ask the users of the profiling process whether or not it was helpful to them in solving crimes. If it is not, that profiling process should be discontinued.

(The information quoted in the IRDU study is from an internal document of the FBI entitled “EVALUATION OF THE PSYCHOLOGICAL PROFILING PROGRAM” and reported by the Institutional Research and Development Unit of the FBI Academy, Quantico, Virginia in December of 1981).

Conclusion

It would be an understatement to say that CIA has caught the imagination of the media and entertainment industry. People are fascinated by the subject and crave information about how the process is done and examples of its effectiveness. The first fiction author to spend time at the FBI academy and the BSU researching cases and trying to gain insight into the profiling process and the mind of the serial killer was Thomas Harris (1982). His best-selling book *Red Dragon* created national interest in the BSU. His next best-selling book *Silence of the Lambs* and the subsequent movies about serial killer Hannibal Lector and FBI Special Agent Clarice Starling brought international attention to the BSU and the NCAVC. Media interests abounded. Additional books, magazines, television programs, movies, and games proliferated and have increasingly become available for public consumption. As with the box office success of the film *Top Gun* when there was a groundswell of young men and women eager to become the next generation of fighter pilots, young students of behavioral sciences and criminal justice aspire to become the next wave of CIA practitioners.

Today worldwide, more and more perpetrators of violent crime are being identified, apprehended, prosecuted, and incarcerated earlier in their criminal careers. In the United States, the violent crime rate has dropped significantly and much of it must be attributed to CIA skills. It is largely due to a community of astute investigators with impressive CIA knowledge and skills who are conducting more proficient crime analysis, behavioral interpretation, and construction of accurate profiles of unknown dangerous offenders. These CIA skills have increased the sophistication and professionalism of investigators of violent crime in their struggle to assure that a higher level of security and justice is being achieved for the citizens of their respective communities.

Related Entries

- ▶ Criminal Investigative Analysis
- ▶ Criminal Profiling
- ▶ FBI Influence on State and Local Police
- ▶ History of Criminal Investigation
- ▶ Investigative Psychology
- ▶ Modus Operandi of Sex Offenders
- ▶ Profiling Arson
- ▶ Psychological Autopsy of Equivocal Deaths
- ▶ Risk Assessment, Classification, and Prediction

Recommended Reading and References

Depue RL, Schindehette S (2005) Between good and evil. Warner Books, New York

Douglas J, Olshaker M (1995) Mindhunter. Simon & Schuster/Pocket Books, New York

Douglas J, Olshaker M (1997) Journey into darkness. Simon & Schuster/Pocket Books, New York

Douglas J, Olshaker M (1999) The anatomy of motive. Simon & Schuster, New York

Douglas J, Ressler RK, Burgess AW, Burgess AG (1992) Crime classification manual. Lexington Books, New York

Hazelwood R, Burgess A (1987) Practical aspects of rape investigation: a multidisciplinary approach. Elsevier, New York

Hazelwood R, Michaud SG (2001) Dark dreams. St. Martin’s Press, New York

Hazelwood R, Dietz PE, Burgess A (1983) Autoerotic fatalities. Heath Inc, Boston

McCrary GO, Ramsland K (2003) The unknown darkness. New HarperCollins, New York

Michaud SG, Hazelwood R (1998) The evil that men do. St. Martin’s Press, New York

Monahan J (1995) The clinical prediction of violent behavior. Jason Aronson Inc, Northvale

Ressler RK, Shachtman T (1992) Whoever fights monsters. St. Martin's Press, New York

Ressler RK, Shachtman T (1997) I have lived in the monster. St. Martin's Press, New York

Ressler RK, Douglas J, Burgess A (1988) Sexual homicide: patterns and motives. Lexington Books, New York

Criminal Investigative Profiling

► Behavioral Investigative Advice

Criminal Justice System

► Drug Abuse and Alcohol Dependence Among Inmates

Criminal Law

► Identification Technologies in Policing and Proof

Criminal Personality Profiling

► Behavioral Investigative Advice
► Criminal Investigative Analysis

Criminal Phenomenon: Crime, Criminal, Criminality

► Control Theory

Criminal Profiling

► Behavioral Investigative Advice

Criminal Punishment

► Institutional Theories of Punishment

Criminal Thinking Model

► Lifestyle Theory

Criminalistics

► Forensic Science and Miscarriages of Justice
► History of Forensic Science in Policing

Criminalization and Occupational Health and Safety

Fiona Haines[1] and Paul Almond[2]
[1]School of Social and Political Sciences, University of Melbourne, VIC, Australia
[2]School of Law, University of Reading, UK

Synonyms

Health and safety law; Workplace safety crimes

Overview

Work can be a dangerous place and provides the context for a multitude of deaths, injuries, and illnesses. These include those arising from major disasters, such as the chemical plant explosion at Bhopal, India, in 1984 or at Piper Alpha in the North Sea in 1988 and more recently in the Gulf of Mexico in 2010. There are the fatalities reported in local newspapers: deaths from a fall at a building site, a tractor rollover, the death of a truck driver on the road, and death from asphyxiation due to a trench collapse. Fatal and chronic illnesses are also important outcomes of work: heart attacks, cancers, lung disease, and the multitude of chronic ailments that arise from exposure to chemicals and other hazards both in the course of work and as a consequence of events like Bhopal and Chernobyl. Then, there are the legacies from particular occupations that lead to chronic disability: back pain from nursing, broken knees from carpet-laying, and repetitive strain injury from typing. Indeed, some ailments take their name specifically from work-related tasks: potter's lung, policeman's heel, and housemaid's knee.

The relationship between the harms above and criminalization as a means to reduce this toll of deaths, injuries, and illnesses is complex. Understanding criminalization in the context of occupational health and safety (OHS) requires analysis of the economic, social, and political context within which organizational and individual practices arise that lead to death, injury, and illness. Further, laws – including criminal prohibitions – are themselves a result of the interrelationship between economic, political, and social factors. Indeed, these elements often are evident in the contests around what should and should not be proscribed and prescribed by law. This entry begins with a brief overview of the nature of the problem of safety globally. Then, literature on the history of OHS law and enforcement is reviewed in order to tease apart the multiple layers of the connection between criminalization and safety at work. This history illustrates the enduring importance of key elements. These elements include the need to pay close attention to the demands of a capitalist economy and its impact on health and safety at work, the concern governments have to retain legitimacy through enacting (but not necessarily enforcing) law, the challenges of dealing with the letter of the law, and placing the variety of regulatory techniques that have been developed over time in an appropriately local and global context.

Occupational Health and Safety: A Global Problem

Rates of OHS-related physical harm remain problematically high within modern industrial societies. In 2007 (the last year for which full comparative figures are available), there were 5,785 deaths at work across the 27 EU member states; this translates into an adjusted rate of approximately 3.5 deaths per 100,000 workers (Eurostat 2010: 347). In the USA, the Occupational Safety and Health Administration's (OSHA) Census of Fatal Occupational Injuries placed the number of fatal injuries to workers for 2008 at 5,214, or 3.7 deaths per 100,000 workers. In Canada, there were 1,036 work-related deaths in 2008, in Australia 444 in 2009 (286 "working" fatalities, 117 out-of-work fatalities, and 44 public or bystander fatalities), and in Japan 1,075. Death, illness, and injury due to poor working conditions extend well beyond the industrialized world. While the ILO recorded 18,067 fatal occupational "accidents" in 2003, the true rate is thought to be as high as 357,948, with a further 1.95 million deaths from work-related ill health (cited in Hämäläinen et al. 2009: 129). The vast majority of these deaths occur in the industrializing world, with China (97,248 accidents and 332,454 ill health), India (46,928 accidents and 355,863 ill health), and Indonesia (15,873 accidents and 61,572 ill health) featuring prominently. Africa is thought to account for over 470,000 work-related deaths each year.

These statistics, though, should be treated with caution. The data are incomplete, as many injuries and deaths are not reported or recorded, even in industrialized nations with established administrative systems; the statistics available for many industrializing nations, which lack this infrastructure, are so sketchy as to be almost unusable. What counts as a death or injury at work also differs. Some jurisdictions (Australia being an example) have an inclusive definition, including workers, nonworkers, and deaths on the road (as a working driver and as a commuter), while most others have a much narrower definition (the UK excludes deaths on the road, and the USA excludes deaths to nonworkers). Incidents of occupational ill health (and resultant deaths) also tend to fall outside the statistics, primarily because they are harder to quantify accurately (see Tombs and Whyte 2007). For this reason, any comparison between countries and jurisdictions needs to be treated carefully. Nonetheless, it is clear that the problem is substantial and requires criminological attention.

The Criminalization Phenomenon: The Legacy of History

Criminologists can learn much from the history of OHS regulation – and from interpretations of that history. These interpretations include those who see primarily a progression towards greater civilization and higher standards. These authors point to the way the law has extended formal protections to an ever-widening portion of the working population, thereby reflecting broader democratizing trends within society (Thomas 1948). Others are more skeptical and see rather a complex deference to the demands of capital, contingent on local conditions (Carson 1979, 1980). Through a series of papers, Kit Carson argued that laws developed in the context of the industrial revolution in a manner that allowed industrialization and laissez-faire capitalism to thrive. Law reform served various purposes. The laws mediated competition within the emerging industrial class, as well as disciplining factory workers to obey the clock rather than the seasons (Carson 1979). But reform needed to take place in a way that ensured that the political legitimacy of the incumbent government was preserved. This was no easy task since critical voices in the Parliament argued that the toll on health that resulted from factory labor threatened the military capacity of the nation.

Critically, both these interpretations of history, the civilizing and the critical, pointed to an important finding – namely that law reform did not necessarily mean adequate enforcement (Almond 2013; Carson 1979; Thomas 1948; Tombs and Whyte 2007). Law reform might be a powerful symbol of change, but implementation was another matter. Diverse interpretations of the industrial revolution in England, in short whether there was an identity of interest between workers and employers, were echoed in subsequent reforms that took place in the 1970s, when the debate centered on whether there was, or was not, a consensus between workers and employers regarding the need for a safe workplace. This period has been identified as the "third wave" of safety regulation, which followed periods of regulation by contract (first wave) and by prescriptive rules (second wave) (Tucker 1995). A key figure of this period, Lord Robens, argued strongly that laws and regulations needed to be framed to enhance this common interest (Robens 1972). In particular, strict prescriptive rules determining what was, or was not, safe should give way to a focus on safe *outcomes* (through performance standards) and the processes by which they could be assured.

The debates around the value of prescription or performance standards can be viewed in several ways. The first way is to evaluate the various arguments to identify which approach is more effective in reducing death, injury, and illness. From a perspective of evaluation prescription can play an important role, for example, in restricting access to confined spaces to prevent asphyxiation. But prescriptive standards can miss critical issues such as risky work *processes* (such as factory production lines). Further, prescriptive standards can quickly become obsolete as technology changes over time. Viewed in light of these limitations to prescription, the philosophical shift to performance standards begun in the 1970s was timely. However, it was also important in light of its economic and political context. The development of the Robens model needs to be understood in the context of 1970s Britain and in particular the centrality and subsequent decline of the coal industry. Robens, with his history in the union movement and his chairmanship of the National Coal Board was seen as "the voice of industrial common sense." His status and biography, straddling both union and management, meant that his ideas could gain traction. Key among these ideas was the need for a flexible regulatory approach that had at its core a focus on the need for safety, the essence of a performance approach (Tucker 1995). Indeed, "Robens-style" legislation, albeit with some important differences in emphasis, was adopted broadly across anglophile countries.

A key difference in emphasis between countries was the need for employee representation and in the role of the unions. The state of Victoria, Australia, emphasized the need for clear worker representation in OHS at the same time as it moved to institutionalize performance standards. The level of control provided to workers within Robens-style legislation was seen as critical to its effectiveness. But an emphasis on worker representation also needs to take account of what kind of representation it is. An important influence on the development of OHS after Robens in the UK was the decline of union power during the Thatcher period in the 1980s. This is because performance standards emphasize the need to internalize safe systems of work and a safety management culture within corporate processes but, in doing so, they also "normalize" OHS risks by assuming that "the risks associated with increased efficiency and profitability can be responsibly managed and contained" (Silbey 2009: 343). This leaves unexamined the fundamental role played by the economic power of employers in determining exposure to risk in the first place. In this way, the neoliberal approach of contemporary governments is embedded via the movement of OHS away from the contested political sphere of prescription standards and industrial relations and towards the relatively uncontested arena of internal business procedures. The impact on the decline of union power on the development of performance standards, and indeed, levels of prescription, is complex. It is possible, for example, to understand prescriptive standards as the end point of negotiation about what it is and is not acceptable for an employer to ask an employee to do. Unions play a prominent role in this process. It is interesting to note that there remains a higher level of prescriptiveness within European health and safety regulations (where union representation in general is higher), a difference which causes friction within the EU (Walters 1996).

In short, any claim of "consensus" around safety between employers and employees needs to be viewed as contingent and fragile. Without adequate worker participation, or where levels of industrial conflict are high, "consensus" quickly breaks down (Gunningham and Sinclair 2009). A vivid example of this is in the role and effectiveness of health and safety representatives. Health and safety representatives, together with health and safety committees, can be an effective way for employees to voice their safety concerns to management and have them resolved quickly and effectively (James and Walters 2002). Yet, an increasing neoliberal emphasis on individual responsibility in the OHS regulatory framework can change the safety representative's task, reducing their role to one of policing fellow workers' "unsafe behavior," such as nonuse of protective safety equipment, and fining them for "breaches" of locally generated safety procedures (Gray 2009). This is an inversion of the role as it was conceived within places such as Victoria, Australia, as the "consensus" shifts from tripartite (employer, employee, and government) responsibility for safety to an overriding emphasis on worker responsibility to ensure a safe workplace. Techniques and regulatory strategies (in this case the role of health and safety reps) must be understood and interrogated in context.

Debates around the need for industrial manslaughter legislation to ensure a safe workplace are similarly complex. These debates are often couched in the need to emphasize adequate *enforcement* of OHS law (rather than just a focus on content) and the need for significant penalties to apply (Tombs and Whyte 2007). Many argue that an industrial manslaughter offense, a version of a homicide offense, is required to ensure safety at work. This would allow the communicative power of the "mainstream" criminal law to be applied to an individual or company following a workplace fatality. An offense of this sort was introduced into UK law via the Corporate Manslaughter and Corporate Homicide Act 2007 (Almond 2013) and was successfully used for the first time to convict a corporation of manslaughter in early 2011. Yet, for others, the demand for a "new" criminal offense concedes too much to history. That is, occupational health and safety legislation, properly conceived, always was and should remain criminal (Johnstone 2007). These authors argue that OHS law has been improperly *conventionalized* as quasi-criminal and that reform and enforcement should focus on using it and reemphasizing its essentially criminal nature. However, over time, the persistent demands for separate industrial manslaughter offenses suggest that the quasi-criminal nature of OHS law remains prominent in the public imagination – at least in some jurisdictions. Recent surveys undertaken in the United Kingdom have shown that OHS breaches are not necessarily regarded as crimes per se, and commonly applied OHS laws are not regarded as having the status or symbolic effect associated with criminal law (Almond and Colover 2010). In general terms, safety laws are regarded rather ambiguously. Nonetheless, this ambiguity is not something that should be accepted unproblematically; it is a product of the struggles seen throughout the history of OHS law reform (Wells 2001).

The problematic status of OHS law as criminal extends also to the consideration of just who - or what – should be made accountable. The intricacies of this debate are well outlined in several works including those by Fisse and Braithwaite (1993) and Celia Wells (2001). At issue here is the degree to which individuals and/or organizations should be held accountable. Wells (2001) explores the way criminal law itself has developed with a bias towards the identification of single individuals who need to be made criminally liable for their actions. Her account of history shows how this skews criminal liability onto the weak, while civil liability for physical harm (if any) is placed on the strong. In the OHS context, she points out the difficulties of individual liability within the workplace context. Often in worksites, it is difficult to identify a single individual who is responsible for a death in the workplace. Further, it is not easy simply to add up responsibilities of several individuals and apportion blame accordingly (the so-called aggregation model of responsibility). Rather, there is a system of work or an organizational culture that gives rise to corners being cut and pressures placed on individuals that can lead to disaster and death (Gray 2009; Silbey 2009). On the other hand, those who view the potential of criminal liability in instrumental terms, as a means of achieving a change in behavior via deterrence, advocate the pursuit of individual rather than corporate liability strategies because individual directors and those whose decisions lead to safety violations are more directly deterred by the threat of imprisonment, fines, and reputational sanctions (Gobert 2005; Tombs and Whyte 2007). Teasing apart who should be held responsible, whether there should be corporate criminal liability or whether it should be the most senior manager or executive who should be accountable, forms a considerable part of the relevant literature here. What is clear is that industrial manslaughter offenses, which invariably derive corporate liability from the individual, have proved most successful against small business where lines of communication are clear and noncomplex (Wells 2001). A recognition of this fact was a key factor shaping the recent law reform process in the UK, where a specific “corporate manslaughter” offense, with liability based on a “management failure” model of liability, was introduced in 2007 to replace the existing common-law offense of manslaughter, thereby attempting to remove the requirement that a responsible individual be found as a basis for corporate liability.

The history of reform, both that of targeted OHS law and that of developing a separate industrial manslaughter provision, reemphasizes the importance of taking account of the competing interests that find their way into the law. Certainly, there is no end in sight for these reforms. There is a helpful body of literature that explores whose interests end up written into law as a result of these struggles, as well as the necessary economic conditions for reform to be effective. Dan Curran’s (1993) history of mine safety legislation in the United States shows how reform that could improve safety required not only a strong union presence, but a union with a strong safety focus as well as a high demand for coal and a tight labor supply. And James Gobert’s (2005) work in the English context shows how potentially effective industrial manslaughter provisions were reduced or weakened under pressure from employers. In any case, whether “regulatory” or “criminal,” the laws that emerge in a field like health and safety tend towards compromise reflect the competing interests and levels of influence of the groups that are party to debates in this field.

The Complex Division Between Criminal Behavior, Regulatory Breaches, and Desirable Behavior

Drawing bright lines in law and regulation between what is criminal, what is harmful and what is desirable is difficult. Part of the difficulty here is the way that the benefits of a particular activity or industry are often intertwined with harm. The embedded nature of harm within benefit and the unequal distribution of those harms and benefits means that it is difficult (both politically and practically) to develop precise law or finely targeted regulation that can separate the two. Further, those with interests in the benefits often have unique leverage over policy-making within government. For example, a mine in a particular locale may provide necessary and welcome employment and inward investment to a town, and its continued presence supported by many residents as a result. It may well also provide significant profits for the mine's owners. Yet, it is miners that are at most risk from death, illness, and injury. So, workplace safety breaches are categorized as side effects of an otherwise purposeful form of wealth-creating industrial activity (which is encouraged), and risks are managed through regulation rather than mining prohibited outright (Hawkins 2002; Kagan and Scholz 1984). Context-rich analyses of criminalization in the context of OHS illustrate well the complexity of the distinction between what is "criminal" behavior and prohibited and what is "merely" a regulatory breach of the law.

While debates around the need for criminalization are important, they can distract attention away from the dynamic within which regulatory strategies (including criminalization) emerge. Again, it can be helpful to go back to Kit Carson's work (1980), which shows how industrialists' resistance to criminalization generated new forms of law in an attempt to bring them to account. Notions of *mens rea* and responsibility were developed in order to frame criminality in terms of individual pathology rather than sociological factors such as poverty and inequality. But safety inspectors found that proving fault in relation to workplace safety breaches was exceedingly difficult and argued for the need for strict liability offenses that did not require intent or *mens rea* in order to prove a breach. But once in place, the industrialists' argued that a truly *criminal* offense must require proof of intent, and so these safety "crimes" were different from "ordinary" deviance in legal form; this meant that they could then be classified as normatively different due to the lack of individual pathology involved (Carson 1979). This change of status was fundamental in legitimating the new laws in the eyes of the industrialists. Understanding enforcement in the OHS context means understanding why enforcement is responsive to the regulated – including strategies to deal with the level of political influence the industry is able to exert. Regulators find it difficult to rigorously enforce safety standards in the face of political ambivalence (Haines 2011). Hence, in the face of resistance, it is not surprising that enforcement officers seek legal provisions that are more easily proved, which gave rise to debates around ambiguity and criminality.

This can be characterized as a problem of *assimilation* and *differentiation*; the problems of assimilating industrial offenders into existing concepts of fault proved insurmountable due to the fundamental individualism of the criminal law. From one perspective, strict liability offenses weakened the law, but from another, they at least allowed inspectors to prove an illegality against a corporate defendant. This "double shift" towards and then away from criminalization was born out of necessity and the strictures of a criminal system developed to deal with working-class "unruliness." Hence, the "quasi-criminal" or regulatory character of the law developed from this point. The impact of this history of ambiguity on current government policy continues to be felt. For example, similar arguments attached to the "reasonable practicability" defense contained in the UK Health and Safety at Work Act 1974 (HWSA). The absolute duty for employers to provide a safe workplace was mitigated by requiring that the costs of breaching the law outweigh the cost of prevention. This had the effect of legitimating the HSWA in the eyes of business without introducing any kind of "moralizing" *mens rea* element into play; at the same time, this requirement acts as a de facto fault element, in that it allows regulators to differentiate between "respectable" businesses and the morally culpable "bad apples" (Hawkins 2002). But such a requirement arguably also shields companies from rigorous enforcement by introducing a utilitarian calculus into the question of safety protection, allowing what should be an absolute moral obligation to be sidestepped where the costs of doing so would be detrimental to business competitiveness.

There is another important result of this "responsiveness"; the diversity of law and enforcement techniques implemented in various jurisdictions. In the past 40 years or more, a proliferation of different techniques has developed in the OHS context. These techniques span different forms of penalties, such as "on the spot" fines and enforceable undertakings measures, and those that attempt to govern organizational behavior such as safety management systems or individual action in minute detail such as behavior-based safety training. Also important to understand are self-regulatory systems that include a wide array of accreditation schemes (often put in place by business defensively to preempt legislative reform) through to what Ayres and Braithwaite (1992) term enforced self-regulation: the safety case model in the regulation of major hazard facilities such as chemical plants and oil refineries. This latter case, when properly implemented, needs to be understood as far more than self-regulation. Under this model, each plant must come up with fine-grained and clearly articulated rules around safety at their facility. These are checked and, if satisfactory, approved by the regulator who then enforces this unique set of rules (the "safety case") as it applies to each individual plant. Each of these initiatives above can, and should, command attention in terms of how far they provide an incentive for improved OHS standards on the ground.

Legal, organizational, and behaviorally focused techniques also need to be understood in context. This context includes contests over: (i) the definition of what laws should, or should not, pertain to the workplace; (ii) what is, and is not, acceptable in terms of both risk and behavior; and ultimately (iii) if death, injury, or illness occur, who should be held responsible and for what. There are a broad range of actors with potential influence here. There are the three major players: the government, the employers, and the employees. Further, inspection and enforcement officers need separate consideration to that of the government as their practices, formal and informal policies, and decision-making processes are significant in shaping regulatory outcomes. In particular, the work of scholars associated with the UK's Oxford Centre for Socio-Legal Studies has shown how regulators utilize discretion, negotiation, and bargain and bluff, in order to maximize their influence while facing severe resource constraints (Hawkins 2002; Hutter 1997). In such cases, they need to find creative ways to implement the law and to generate higher standards. This can prove difficult, if not impossible.

Placing Regulation and Law in Context: The Direct Impact of Economic Conditions

Important as law, regulation, and enforcement are to OHS standards, it is important to recognize that economic conditions also have a significant impact on safety. We noted above that small businesses were at most risk from prosecution under industrial manslaughter provisions in the UK. But small businesses in some sectors can have lower safety standards because of their weak position in the market place, both as discrete entities and as part of a contracting chain (Haines 1997). They have neither the financial nor human resources to invest in safety to the same degree as their large business counterparts. As a result, the safety afforded to employees of small businesses can suffer. This finding on small business also needs to be extended to include the contracting relationships prevalent in some industries (e.g., construction) and increasing in others due to increasing levels of contract labor and the decision by some companies to contract out non-core parts of their business operations, such as cleaning or maintenance (James et al. 2007). This can create problems for safety. Firstly, these businesses may be small and so less able to ensure the safety of their employees. Secondly, the degree to which they can focus on safety may be dependent upon the contract price they are able to obtain; when stiff competition drives this price down, levels of safety can decline. Because of this, certain jurisdictions have enacted what is known as "chain of responsibility" legislation, making principal contractors responsible for the safety of subcontractors' employees. This is a good example of law that is sensitive to economic conditions. But not all jurisdictions have these laws, or implement them effectively, and where they are absent, real problems in safety levels can arise (James et al. 2007).

Finally, there has recently been an increased attention paid to workplace stress and its impact on ill health. For some, categorizing workplace stress as a health and safety issue is a step too far, but stress has been associated with significantly decreased life expectancy (La Montagne et al. 2007). In particular, work which is time pressured, but where individuals have little control over the pattern of their work, is particularly hazardous. Also, being at the bottom of the hierarchy in terms of power over the conditions of work within a workplace has been shown to have detrimental health effects, a key finding of the prominent Whitehall studies (see, e.g., Marmot and Smith 1991). The implication of this research is that a narrow intervention of the kind often associated with health and safety initiatives is unlikely to be effective. Rather, what is needed is a wholesale reorganization of work itself. Achieving such change seems at a considerable distance from the debates around the proper boundaries of the criminal law. But they share a similar feature in that reorganizing work in a manner consistent with the promotion of health, as with determining what is and is not criminal behavior, is politically fraught and likely to engender strong opposition from those who benefit from current ways of working.

Globalization, Criminalization, and Workplace Safety

Much current OHS research focuses on deaths, illnesses, and injuries that take place in the industrialized world, but (as stated at the outset of this entry) these are global problems. There are many high-profile examples of industrial disasters in industrializing countries, like Bhopal, mentioned at the beginning of this entry. Another example is provided by the Kader toy factory fire that occurred in Bangkok, Thailand, in 1993 (Haines 2005), which resulted in 183 deaths. Many lethal factory fires share similar features, including poor building design, oppressive working conditions, and substandard fire safety procedures. The role the criminal law can play in the aftermath of these disasters is also noteworthy and problematic, particularly in terms of who those laws seek to target. In the case of the Kader fire, it was a worker that was jailed for 10 years for causing the fire because he failed to extinguish the butt of his cigarette. His contribution was minor and pales into insignificance in light of inadequate building and fire safety standards at the Kader factory. Criminalization, in an international context, can act to scapegoat individuals at the expense of dealing with systemic problems. The international diffusion of corporate criminal liability has been driven in no small part by the prompting of international organizations like the OECD. This has the potential to "drive up" the value and effectiveness of regulatory processes in developing nations, but their implementation in very different jurisdictions can lead to problematic outcomes.

The globalization of trade also has an appreciable impact on OHS. There are ever increasing numbers of global supply chains where the most risky jobs are located in rapidly industrializing contexts (James et al. 2007). China provides a salutary lesson in this regard; in the scramble for traction in the new world economy, the maximization of economic growth has taken precedence over the enforcement of effective safety regulations. Chinese industry seeks to undercut production costs (and hence safety standards) elsewhere in the world and, in doing so, *sets the floor* for workplace conditions in competing developing nations. Governments of the Global South thus prioritize the need to attracting inward investment to spur economic growth over demands for improvement in safety levels. The challenge of OHS from a global perspective contains similar complexities to those reflected in the discussion earlier in this entry, in that harm is embedded within the benefits of international trade, including employment and income for individuals, extended families, and local communities. But this should not excuse the behavior of multinational corporations which scour the globe in order to find where production is cheapest and safety and environmental standards low. The strength and weaknesses of various industrializing states under global competition needs careful analysis, however. It should not be assumed that states such as Thailand, China, or India have no room to develop an effective OHS regime; indeed, local responses that are sensitive to local conditions can be more effective than those simply transplanted from the Global North (Haines 2005). Hence, a failure, or success, in improving safety needs to be evaluated in light of local political contests in addition to the impact of global economic conditions.

Conclusion

This entry has argued that the connections between processes of criminalization and occupational health and safety need to be understood within the historical context of the recurrent struggles that have always accompanied law reform in this area. The ambiguous legal standing of OHS law is a result of the struggles in defining what is seen as acceptable within the workplace environment, as well as perceptions of the law's capacity to effectively secure acceptable behavior on the part of the regulated. But at the same time, regulation here is about the value of human life and the capacity for one person to place another at risk. It is about the pursuit of a fair and just system of economic citizenship, and there are important debates about how best to advance these values and the role that the notionally "moral" criminal law should play in this process. Is it more important to try and reinforce the normative value of health and safety regulation via the communicative power of the criminal law or to avoid doing so because of the disruptive, individualizing, and distracting impact that such action can have? The contest around criminalization should not simply be one around the need for industrial manslaughter but to understand the history of OHS law in different contexts and the nature of criminal law within the workplace itself.

The struggles and contests also reflect broader economic, political, and social change. Just as prescription can be seen as a result of various contests, so too the Robens model's emphasis on the consensus of interests needs to be understood within the context of the welfare state in Britain in the 1970s. It is not surprising that this expansive view has been challenged by the rise of neoliberalism and the regulatory state. But this is not simply a return to the past. Hence, the importance and effectiveness of health and safety representatives and committees depends critically upon local context. In one area, it can represent a very real opportunity for representation of those whose lives are most at risk. Yet, in another, it can signal the return of an emphasis on individual responsibility and victim blaming. These struggles and contests are reflected in attempts to introduce new criminal offenses and penalties to sanction egregious failing by employers for their negligence. The push to criminalize is met with resistance, which in turn is met on occasions with capitulation but also with creativity.

It is in this political milieu that multiple techniques have been adopted to try and raise standards. These need careful scrutiny. In particular, the same initiative can work quite differently depending on economic and political context – including whether strong unions that priorities safety are, or are not, present. The constantly changing nature of production and consumption, too, places significant challenges in the path of improving standards. There is now much greater complexity to relations between businesses and a real need for OHS to be included in relations between businesses, not just within the confines of one firm. Finally, economic globalization and the presence of supply chains for particular goods and services spanning the globe make law reform and adequate enforcement of standards difficult. It is a context, however, in which a criminological analysis that is cognizant of the economic and political dimensions of the problem has much to offer.

Related Entries

- ▶ Compliance and Corporate Crime Control
- ▶ Corporate Crimes and the Business Cycle
- ▶ Corporate Liability
- ▶ Environmental Regulation and Law Enforcement
- ▶ Victims of Corporate Crime

Recommended Reading and References

Almond P (2013) Corporate manslaughter and regulatory reform. Palgrave Macmillan, London

Almond P, Colover S (2010) Mediating punitiveness: understanding public attitudes towards work-related fatality cases. Eur J Criminol 7(5):1–16

Ayres I, Braithwaite J (1992) Responsive regulation: transcending the deregulation debate. Oxford University Press, New York

Carson WG (1979) The conventionalization of early factory crime. Int J Sociol Law 7(1):37–60

Carson WG (1980) The institutionalization of ambiguity: early British factory acts. In: Stotland E, Geis G (eds) White collar crime: theory and research. Sage, Beverly Hills, pp 142–173

Curran DJ (1993) Dead laws for dead men: the politics of federal coal mine health and safety legislation. University of Pittsburgh Press, Pittsburgh

EUROSTAT (2010) The social situation in the European Union 2009. Publications Office of the European Union, Luxembourg

Fisse B, Braithwaite J (1993) Corporations, crime and accountability. Cambridge University Press, Cambridge

Gobert J (2005) The politics of corporate manslaughter – the British experience. Flinders J Law Reform 8(1):1–38

Gray GC (2009) The responsibilization strategy of health and safety: neo-liberalism and the reconfiguration of individual responsibility for risk. Br J Criminol 49(3):326–343

Gunningham N, Sinclair D (2009) Organizational trust and the limits of management based regulation. Law Soc Rev 43(4):865–900

Haines F (1997) Corporate regulation: beyond 'punish or persuade'. Clarendon, Oxford

Haines F (2005) Globalization and regulatory character: regulatory reform after the Kader toy factory fire. Ashgate, Dartmouth

Haines F (2011) The paradox of regulation: what regulation can achieve and what it cannot. Edward Elgar, Cheltenham

Hämäläinen P, Saarela KL, Takala J (2009) Global trends according to estimated number of occupational accidents and fatal work-related diseases at region and country level. J Safety Res 40:125–139

Hawkins K (2002) Law as last resort: prosecution decision-making in a regulatory agency. Oxford University Press, Oxford

Hutter B (1997) Compliance: regulation and environment. Clarendon, Oxford

James P, Walters D (2002) Worker representation in health and safety: options for regulatory reform. Ind Relat J 33(2):141–156

James P, Johnstone R, Quinlan M, Walters D (2007) Regulating supply chains to improve health and safety. Ind Law J 36(2):163–187

Johnstone R (2007) Are occupational health and safety crimes a hostage to history? An Australian perspective. In: Brannigan A, Pavlich G (eds) Governance and regulation in social life (Essays in honour of W.G. (Kit) Carson). Cavendish-Routledge, London, pp 33–54

Kagan RA, Scholz JT (1984) The 'criminology of the corporation' and regulatory enforcement strategies. In: Hawkins K, Thomas JM (eds) Enforcing regulation. Kluwer-Nijhoff, Boston, pp 67–96

La Montagne A, Keegal T, Louie A, Ostry A, Landsbergis P (2007) A systematic review of the job stress intervention evaluation literature 1990–2005. Int J Occup Environ Health 13(3):268–280

Marmot MG, Smith GD (1991) Health inequalities amongst British civil servants: the Whitehall II study. Lancet 337(8754):1387–1401

Robens L (1972) Safety and health at work: report of the committee ('The Robens Report'). HMSO, London

Silbey S (2009) Taming Prometheus: talk about safety culture. Annu Rev Sociol 35:341–369

Thomas MW (1948) The early factory legislation: a study in legislative and administrative evolution. Thames Bank, Leigh-on-Sea

Tombs S, Whyte D (2007) Safety crimes. Willan, Cullompton

Tucker E (1995) And defeat goes on: an assessment of third-wave health and safety regulation. In: Pearce F, Snider L (eds) Corporate crime: contemporary debates. University of Toronto Press, Toronto

Walters D (1996) Health and safety strategies in Europe. J Loss Prevent Process Industr 9(5):297–308

Wells C (2001) Corporations and criminal responsibility, 2nd edn. Oxford University Press, Oxford

Criminological Theory

▶ Juvenile Violence

Criminology of Place

David Weisburd[1,2], Elizabeth R. Groff[3],
Sue-Ming Yang[4] and Cody W. Telep[5]
[1]Department of Criminology, Law and Society,
George Mason University, Fairfax, VA, USA
[2]Faculty of Law, The Hebrew University,
Mt. Scopus, Jerusalem, Israel
[3]Temple University, Philadelphia, PA, USA
[4]Department of Criminology, National Chung-Cheng University, Min-Hsiung, Chia-Yi, Taiwan
[5]School of Criminology and Criminal Justice,
Arizona State University, Phoenix, AZ, USA

Overview

The primary focus of research in criminology has traditionally been on individuals and why they become involved in crime (Eck and Eck 2012; Sherman 1995). Weisburd and Piquero (2008), for example, found that studies that sought to predict crime were more likely to focus on individuals than any other units of analysis. Examining articles in the journal *Criminology* that tried to model crime, they found that 94 of 169 articles (55.6 %) published between 1968 and 2005 used the individual as the unit of analysis. Other units commonly found in the study included cities or counties (14 %) and neighborhoods (9 %). More recently, Eck and Eck (2012) reviewed all of the research papers in *Criminology and Public Policy* since its inception and found that none addressed crime at place.

In recent decades, however, a growing number of scholars have begun to take a very different approach to the crime problem. This approach begins not with the people that commit crime but rather with the places where crime occurs. This is a radical reconceptualization of the crime problem, but one that that is warranted by the reality of crime in the city. Why is it necessary to reconsider the "person-focused" crime and justice model of the last century? The main reason is simply that the yield of this approach has been questioned for more than three decades. Weisburd and Piquero's (2008) review found that despite the focus primarily on individuals, prediction levels are fairly low and have remained low over time. Lawrence Sherman and colleagues (1989) categorized this new approach as the "criminology of place." The focus here is on very small geographic areas within cities, often as small as addresses or street segments, and understanding their contribution to the crime problem. It focuses attention on "hot spots of crime," or crime concentrations in such micro geographic areas and the factors that explain these concentrations over time, as well as the best ways to address these high crime locations. Drawing upon work by Weisburd, Groff, and Yang (2012), this entry will review the emergence of the criminology of place in recent decades and theoretical perspectives for understanding why places are such an important part of the crime equation. This entry will also review the main findings from a longitudinal study in Seattle, Washington, by Weisburd et al. (2012) examining the factors that contribute to why streets are high crime (or low crime) over time.

The Emergence of the Criminology of Place

Criminologists have traditionally sought to understand why certain people as opposed to others become criminals (e.g., see Gottfredson and Hirschi 1990) or to explain why certain offenders become involved in criminal activity at different stages of the life course or cease involvement at other stages (e.g., see Laub and Sampson 2003). Criminologists have also been interested in places, but they have historically focused on larger geographic areas like neighborhoods or even larger areas like cities or states (see Weisburd et al. 2009). For example, criminologists have often tried to explain why certain types of crime or different levels of criminality are found in some communities as contrasted with others (e.g., see Bursik and Grasmick 1993).

While the individual and "macro" units of place, such as the community, have long been a focus of crime research and theory, only recently have criminologists begun to explore crime at very small "micro" units of geography. The roots of such approaches can be found in the efforts of scholars to identify the relationship between specific aspects of urban design (Jeffery 1971) or urban architecture (Newman 1972) and crime, but broadened to take into account a much larger set of characteristics of physical space and criminal opportunity (e.g., Brantingham and Brantingham 1981). These studies drew important distinctions between the site in question and the larger geographical area (such as the neighborhood, community, police beat, or city) that surrounds it.

But the key to the origins of the criminology of place is a group of emerging theoretical perspectives that developed as a reaction to the limitations identified in offender-based criminology in the 1970s (e.g., see Martinson 1974). In a groundbreaking article on routine activities and crime, for example, Cohen and Felson (1979) suggested that a fuller understanding of crime must include a recognition that the availability of suitable crime targets and the presence or absence of capable guardians influence crime events. Cohen and Felson were to turn traditional conceptions of the crime problem on their head by suggesting that crime could be prevented without changing the supply or motivation of offenders in society.

In the routine activities model, victims, offenders, and guardians were all essential parts of the crime equation. While traditional conceptions had focused on offenders, using data on crime rates in the United States in the post-World War II period, Cohen and Felson (1979) illustrated that changes in other parts of the crime equation could impact the level of crime in society. For example, they found that changes in the value of property or the ease of which it could be stolen affected burglary rates. Decreasing size and weight of electronic devices (e.g., televisions) and the increasing value of goods found in homes were found to be related to increases in victimization. They also found that changes in the presence of "capable guardians" influenced crime. In this case, the entrance of women into the work force and the resulting lack of "guardianship" at homes were associated with significant increases in house burglaries. The routine activities perspective pushed criminologists to extend their vision beyond the traditional concern with the causes of criminality and their focus on offenders. Indeed, Cohen and Felson (1979) took a pool of motivated offenders as a given and focused their attention more on the suitability of targets and the presence or absence of capable guardianship.

Having introduced other elements of the crime equation, Cohen and Felson's (1979) work naturally led to a focus on crime opportunities. Victims, offenders, and guardians are likely to intersect in physical space. The spatial component of crime thus became a key component of this perspective. As Cohen and Felson (1979, 589) note, "we take criminal inclination as given and examine the manner in which the spatiotemporal organization of social activities helps people to translate their criminal inclinations into action. Criminal violations are treated here as routine activities which share many attributes of, and are interdependent with, other routine activities."

Researchers at the British Home Office, in a series of studies examining "situational crime prevention," also challenged the traditional focus on offenders and communities (Clarke 1983). Situational crime prevention moved the focus of the crime equation away from the people who commit crime and instead considered the context of criminal events to be critical. At the core of situational prevention is the concept of rational choice (Clarke and Cornish 2001). In contrast to offender-based approaches to crime prevention, which usually focus on the dispositions of criminals, situational crime prevention begins with the opportunity structure of the crime situation (Clarke 1995). Such opportunities affect whether potential offenders will choose to commit crime. Importantly, such choices are seen to be the result of a decision employing "bounded rationality" (Clarke and Cornish 2001). That is, potential offenders use the limited knowledge about victims or guardians in specific situations that is easily available to consider relative costs and benefits.

Situational crime prevention is concerned with the "opportunity structures" of specific contexts and places. By opportunity structure, advocates of this perspective are not referring to the broad societal structure of opportunities that underlie individual motivations for crime but to the immediate situational components of the context of crime. In this context, crime prevention may involve efforts as simple and straightforward as target hardening or access control and often follows a straightforward and commonsense notion of how to deal with crime problems (Clarke 1995). Importantly, place at a "micro" level is key to situational crime prevention theory since it focuses on the immediate opportunities for crime, which are generally structured within very small geographic areas.

Around the same time as routine activities theory and situational crime prevention developed, Paul and Patricia Brantingham published their seminal book *Patterns in Crime*, which emphasized the role of place characteristics and human activity in shaping the type and frequency of human interaction (Brantingham and Brantingham 1984). Environmental criminology, also known as crime pattern theory, explores the distribution and interaction of targets, offenders, and opportunities across time and space (Brantingham and Brantingham 1981). Both rational choice and situational opportunities play a key role in crime pattern theory. The concept of place, in turn, is essential to crime pattern theory. Not only are places logically required (an offender must be in a place when an offense is committed) but also their characteristics are seen to influence the likelihood of a crime. Crime pattern theory links places with desirable targets and the context within which they are found by focusing on how places come to the attention of potential offenders.

The Importance of Examining Crime at a Micro Geographic Level

One implication of these emerging perspectives was that places at a "micro" geographic level should be an important focus of criminological inquiry. While concern with the relationship between crime and place is not new and indeed goes back to the founding generations of modern criminology, the "micro" approach to places suggested by recent theories had just begun to be examined by criminologists (see Weisburd et al. 2009). Places in this "micro" context are specific locations within the larger social environments of communities and neighborhoods (Eck and Weisburd 1995). They are sometimes defined as buildings or addresses (e.g., see Sherman et al. 1989); sometimes as block faces, "hundred blocks," or street segments (e.g., see Weisburd et al. 2004); and sometimes as clusters of addresses, block faces, or street segments (see Braga et al. in press).

The criminology of place has from the outset fit its units of analysis to its theoretical interests. Routine activities theory (Cohen and Felson 1979), situational crime prevention (Clarke 1983), and environmental criminology (Brantingham and Brantingham 1981) naturally push the cone of geographic resolution to very low levels. Importantly, however, while pushing the geographic study of crime to much lower levels than traditional criminology, these theories did not specify a precise geographic unit for studying crime.

At what level should crime at place be studied? There is an important trend over time toward the study of crime at place at smaller units of geography. But does that trend reflect a fact about the level of geography that is important for understanding crime, or is it simply a result of the specific data available and theoretical interests of scholars? Or is the best approach one that is eclectic in its understanding of crime at place? While the relevance of studying varying geographic units in coming to a more complete understanding of crime at place should not be discounted, it is important to recognize at the outset that studying crime at the "wrong" geographic unit may lead to a very misleading portrait of how place and crime interact. In this context, examining crime patterns at larger geographic levels, even commonly used "smaller" units such as census tracts or census block groups, may mask significant lower order geographic variability.

Importantly, beginning with a micro-level approach allows the researcher to examine the influences of larger geographic units, while starting at higher levels of geography may preclude examination of local variability. Collecting data at the lowest geographic level, or smallest units of analysis, allows one to examine how those trends relate to larger units of analysis at higher levels of geography. But data collection at larger units would not allow conversion to more micro units of analysis. For this reason, Weisburd et al. (2012) used a micro geographic unit – the street segment – as the basic unit of analysis in their study of Seattle. The street segment, however, is not the smallest unit of geography that can be used to study crime. Some scholars have argued for the importance of examining crime patterns across individual addresses or facilities (e.g., see Sherman et al. 1989). Policing data may include significant coding errors regarding a crime's occurrence at the address level, and so one can be much more confident in the accuracy of crime incident data at the street-segment level, since getting the street of a crime's occurrence wrong is much less likely than mistaking the specific address location for a crime. There is also strong theoretical justification for the identification of street segments as key behavior settings for crime (see below).

Social Disorganization and Opportunity Theories: Recognizing the Importance of Theoretical Integration

Criminology of place scholars have looked primarily to what can be termed "opportunity theories" (Wilcox et al. 2003) as an explanation for why crime trends vary at places and as a basis for constructing practical crime prevention approaches (e.g., see Sherman et al. 1989; Weisburd et al. 2004). Routine activities theory, situational crime prevention, and crime pattern theory all place great emphasis on the specific opportunities offered by specific places and situations. In contrast, study of crime at higher geographic levels has placed emphasis on the social characteristics of places, for example, the socioeconomic levels of people who live in certain areas (e.g., Bursik and Grasmick 1993) or the degree to which there is strong population heterogeneity (Shaw and McKay 1942). Such perspectives may be grouped more generally as social disorganization theories (see Sampson and Groves 1989). These themes are repeated again and again in traditional macro studies of place and crime.

Scholars who study the criminology of place have virtually ignored social disorganization theories in empirical analysis and theoretical discussion. In one sense, this is understandable, since the impetus for study of micro crime places came from opportunity theories. Such theories justified examination of small geographic units because of their emphasis on the specific situations and contexts that make crime possible.

In this context, the neglect of social disorganization in the study of the criminology of place can be traced to what some scholars have called "theoretical competition" (Bernard and Snipes 1996). All theories cannot be right in this perspective, and accordingly it is the job of criminologists to advance a single theoretical paradigm that can explain the phenomenon at hand. But a different perspective is interested in "theoretical integration." As Bernard and Snipes (1996, 302) write, "(w)e argue that integration is the appropriate approach because the theories primarily make different but not contradictory predictions." In the context of theoretical integration, one would draw from multiple theories that can increase overall understanding of the crime problem at micro levels of place.

Theoretical integration of opportunity and social disorganization theories in a model for understanding crime at street segments would not make sense if social disorganization was a concept that is irrelevant to the criminology of place. This seems to be the position of many scholars in this area. Sherman et al. (1989, 30), for example, argued in introducing the idea of a criminology of place, that "(t)raditional collectivity theories [termed here as social disorganization theories] may be appropriate for explaining community-level variation, but they seem inappropriate for small, publicly visible places with highly transient populations." Often this position is taken indirectly, simply by ignoring social disorganization theories (e.g., Eck and Weisburd 1995). Sometimes scholars have specifically recognized the potential relevance of social disorganization to the understanding of variations in crime at micro geographic levels (e.g., Smith et al. 2000). However, such interest is the exception rather than the rule.

There is reason to believe that social disorganization theories are relevant at both micro- and macro-level geographies. Street segments do not simply represent physical entities. They are also social settings, or following Wicker (1987, 614) "behavior settings," which can be seen as "small-scale social systems." In this context, street segments can be seen as examples of small-scale communities (see Taylor 1997). People who frequent a street segment get to know one another and become familiar with each other's routines. Residents develop certain roles they play in the life of the street segment (e.g., the busybody, the organizer). Norms about acceptable behavior develop and are generally shared. Blocks have standing patterns of behavior, for example, people whose routines are regular like the mail carrier or the shop owner. In this context, we can see street segments as "micro communities" as well as "micro places." They have many of the traits of communities that have been seen as crucial to social disorganization theory, in that these physical units function also as social units with specific routines.

If the street segment can be seen as a type of "micro community," then social disorganization theory has direct relevance to the criminology of place (see Rice and Smith 2002; Smith et al. 2000; Taylor 1997). For example, street segments, like communities, are often dynamic with people moving in and out as well as shops opening and closing. Such transitions have often been seen to represent heightened social disorganization in studies of communities. Similarly, poverty and social disadvantage have been identified as strongly related to crime at higher levels of geography. Is variability in wealth or social class at the street-segment level also related to trends in crime? Do characteristics reflecting social disorganization vary at a street-segment level of analysis, as they do across communities and neighborhoods? There are critical empirical questions addressed in the Weisburd et al. (2012) study discussed below.

Predicting Crime at Place: Findings from a Longitudinal Study

Weisburd et al. (2012) recently focused on the criminology of place in an effort to advance both theory and practical crime prevention in this area. They examined crime incidents at the street-segment level over a 16-year period in Seattle, Washington, assessing the concentration of crime over time and the geographic variability of crime across the city. They then combined data on a number of physical and social characteristics of the street segments, using variables from both opportunity and social disorganization theories to better understand what factors help explain the distribution of crime at the street level. Their study overall reached five major conclusions that are all discussed in more detail below:

1. Crime is tightly concentrated at crime hot spots, suggesting that one can identify and deal with a large proportion of crime problems by focusing on a very small number of places.
2. These crime hot spots evidence very strong stability over time and thus present a particularly promising focus for crime prevention efforts.
3. Crime at places evidences strong variability at micro levels of geography, suggesting that an exclusive focus on higher geographic units, like communities or neighborhoods, will lead to a loss of important information about crime.
4. It is not only crime that varies across very small units of geography, but it is also the social and contextual characteristics of places.

The criminology of place in this context identifies and emphasizes the importance of micro units of geography as social systems relevant to the crime problem.

5. Crime at place is very predictable, and therefore it is possible to not only understand why crime is concentrated at place but also to develop effective crime prevention strategies to ameliorate the crime problem at place.

The Tight Coupling of Crime at Place

Criminologists have traditionally assumed that crime is "loosely coupled" to place. The idea of loose and tight "coupling" has been used in many disciplines to identify the extent to which parts of systems are linked or dependent to one another. What is meant here by "loose coupling" of crime at places is that criminologists have traditionally not seen the bonds that tie crime to place as very strong, even though it has been clear from the outset that crime occurs in specific settings. Traditionally, criminologists have viewed crime opportunities provided by places as too numerous to make crime prevention strategies targeting specific places of utility for theory or policy.

Recent studies point to the potential theoretical and practical benefits of focusing research on crime places and suggest a "tight coupling" of crime with the places where crime occurs. A number of studies, for example, suggest that there is a very significant clustering of crime at places, irrespective of the specific unit of analysis that is defined. The extent of the concentration of crime at place is dramatic. In one of the pioneering studies in this area, Sherman et al. (1989) found that only about 3 % of the addresses in Minneapolis, Minnesota, produced 50 % of all calls to the police. Fifteen years later, in a study in Seattle, Washington, Weisburd and colleagues (2004) reported that between 4 % and 5 % of street segments in the city accounted for 50 % of crime incidents for each year over 14 years. Weisburd et al. (2012) found that about 50 % of crime is found at just 5–6 % of street segments each year in Seattle over 16 years. More than 20 % of crime incidents were found at just 1 % of street segments.

Research also demonstrates that crime is not only concentrated at a small number of places but also that these concentrations remain stable over time, also reinforcing the idea of tight coupling. Weisburd et al. (2012) found similar levels of concentration for each of the 16 years they observed crime at street segments in Seattle. It is in some sense startling that over 16 years, about the same number of street segments produce about the same proportion of crime, especially since crime declined by more than 20 % in the city during this period. The 1 % of street segments that produce about 23 % of total crime in Seattle began the observation period as the most serious hot spots in Seattle and remained hot spots throughout the time period observed. The relative intensity of crime at these places and the stability over time of this developmental pattern suggest that there are specific characteristics of these places that generate or attract crime. Reflecting a similar underlying structure of tight coupling between crime and place was the more than 80 % of the street segments in the city which had very little or no crime throughout the 16-year period. Here, there would seem to be factors that help places resist or discourage crime.

Street-by-Street Heterogeneity in Crime

While a number of theoretical approaches in recent decades stress the importance of using small units of geography, it is important to demonstrate empirically that a focus on micro places is relevant to understanding the distribution of crime. That is, it could be the case that crime hot spots at a micro geographic level are just proxies for larger community-level hot spots. One could imagine a scenario, for example, where all the hot spots in a city are concentrated in a single neighborhood, which would suggest that neighborhood-level processes might be more relevant to consider.

Weisburd et al. (2012) used a series of geographic analyses to demonstrate that the processes explaining crime patterns at street segments do not come primarily from higher geographic influences such as communities. There were indications of the influence of higher order trends in the data, for example, in the fact that higher crime street segments were not distributed at random and were more likely to be closer to each other than would be predicted simply by chance. But these indications of macro geographic influences were much outweighed by evidence of the importance of looking at crime at the micro level (i.e., street segments). There is strong street-to-street variability in crime patterns in the data, and such variability emphasizes the importance of studying crime at place at a micro unit of analysis.

Predicting Hot Spots of Crime

Weisburd et al. (2012) were the first to examine the distribution of opportunity and social disorganization factors at the street-segment level over time. The key question here was whether or not variables in these two perspectives were concentrated at street segments. That is, are there hot spots of crime opportunities and hot spots of social disorganization at the street-segment level? And if so, are these hot spots found throughout the city, as crime hot spots are?

They examined four main dimensions of opportunity theory: (1) motivated offenders, (2) suitable targets, (3) guardianship, and (4) accessibility/urban form. Opportunity measures were generally highly concentrated and evidenced strong variability across places. For example, 50 % of high risk juveniles (a proxy for "motivated offenders") were consistently found on between 3 % and 4 % of the total number of Seattle street segments. In turn, half of all the employees (a proxy for "suitable targets") in the city were located on less than 1 % of Seattle street segments. These concentrations of opportunity factors also remained fairly stable over time, suggesting one reason for the stability of crime patterns longitudinally.

Weisburd et al. (2012) also found tremendous concentration and variation in most of the social disorganization measures they examined. While prior theory suggested that opportunity factors would be highly concentrated, the results for social disorganization were more surprising. Looking both at structural and mediating variables of social disorganization, they found that there are hot spots of social disorganization at the street-segment level. For example, social disorder is one of the most direct indicators of the inability of a social system to control behavior. Fifty percent of reports of physical disorder in Seattle were found on between 1.5 % and 3 % of street segments. Fully 50 % of truant students (a mediating variable used as a proxy for "unsupervised teens") were consistently found to live on between 2 % and 3.5 % of the total street segments during the study period. And these hot spots were not simply part of contiguous hot spots at larger geographic levels. They were not found only in specific neighborhoods; rather they were distributed across the city landscape. This suggests that a perspective that has generally been seen as relevant only higher levels of geography shows concentration and variability at the street-segment level.

How do these findings relate to crime concentrations in the city? Weisburd and colleagues (2012) used a multinomial logistic regression to examine the association between opportunity and social disorganization factors and street-segment crime trends. The overall model examined the influence of these factors on eight patterns of trajectories of crime. The focus here is on the comparison between the chronic street segments (that 1 % of street segments responsible for over 20 % of crime) and the approximately 80 % of street segments that were largely crime free over 16 years. How do opportunity and social disorganization variables help to differentiate whether a street segment is likely to be crime free or a chronic crime hot spot? The findings from the multinomial regression analysis overall suggest a high degree of statistical fit between theories of opportunity and social disorganization and crime at street segments. Using a pseudo R^2 measure suggested by Nagelkerke, for example, Weisburd et al. (2012) found that the model explained about 68 % of the variance in pattern membership. This suggests that crime at place is very predictable, and even in a fairly early stage of theorizing and empirical research, predicting crime at place can be done with more precision than most models predicting crime in individuals (see Weisburd and Piquero 2008). This also suggests the relevance of an "integrated theoretical approach" (Bernard and Snipes 1996) to the criminology of place. Both opportunity measures and social disorganization measures provide important information for understanding variation in developmental patterns of crime at place.

The multinomial regression results suggested a number of opportunity and social disorganization variables significantly predict whether a street segment is likely to be a crime hot spot compared to being crime free. The two most important predictors of crime hot spots were drawn from the opportunity perspective. Both related to the suitability of places as crime targets. The larger the residential population of a street segment, the more likely it was to be a crime hot spot. Similarly, when there were more employees on a street segment, it was much more likely to become a hot spot of crime. These findings reinforce a more general conclusion of the opportunity perspective. The factors that increase the risks of crime relate directly to the situational opportunities that places present. When more people live on a block, there is more potential for crime because there are more potential victims (and perhaps higher numbers of motivated offenders). When more employees work on the block, they increase the amount of crime on that street, because they are likely to increase the volume of targets on the street. Another indicator of crime opportunities, arterial roads, also increased the risks of crime on street segments markedly. Arterial roads were much more likely to become crime hot spots. Arterial roads are more likely to bring together motivated offenders and suitable targets, because they are easily accessible.

Three variables that reflect the social disorganization perspective also had very large impacts on the likelihood of a street segment being a crime hot spot. The higher the level of physical disorder on a street segment, the greater the likelihood of it being a high crime street segment. This is very much consistent with the idea that an inability of the small social systems to control disorderly behavior will increase the risk of crime. High socioeconomic status, in contrast, acted as a protective factor for crime, with crime hot spots much less likely to be found at places with wealthier residents (as reflected in higher property values) who assumedly are able to bring into play both formal and informal social controls more effectively. Perhaps most interesting was the very strong impact of collective efficacy (see Sampson et al. 1997), as reflected by voter participation. Collective efficacy seems to act as a strong protective factor for crime at place. Accordingly, the direct situational opportunities that increase crime risk are one part of the crime equation at places, but social factors that act to insulate places from crime risks are also important.

Implications

If crime is strongly concentrated at place, and such concentrations are stable over long periods, then it is possible to assume that a focus on crime hot spots can have important crime prevention benefits. The concentration of crime suggests that police or other crime prevention agents can focus on a very small number of targets and have a large impact on crime problems overall. The fact that crime hot spots are stable across long periods reinforces the potential utility of such approaches, because it implies that if crime prevention agents did not intervene, the problem would persist. These are the implications of the tight coupling of crime and place that have been observed in prior studies.

But the potential for the criminology of place in crime prevention is not just theoretical. These basic research findings have been reinforced by evaluations of practical crime prevention efforts. A series of randomized field trials demonstrate that police efforts focused on hot spots can result in meaningful reductions in crime and disorder (see Braga and Weisburd 2010). Braga et al. (in press) found in their Campbell Collaboration systematic review of the hot spots literature that 20 of 25 tests from 19 experimental or quasi-experimental evaluations of hot spots policing reported noteworthy crime or disorder reductions. This strong body of rigorous evaluations led the National Research Council Committee to Review Research on Police Policy and Practices (2004, 35) to conclude "a strong body of evidence suggests that taking a focused geographic approach to crime problems can increase the effectiveness of policing." Importantly, Braga et al. (in press) also reported little evidence of crime displacement as a result of hot spots programs. Crime did not simply move around the corner as a result of police intervention.

But if important causal mechanisms underlying developmental patterns of crime at place can be found in factors such as economic deprivation or collective efficacy, then a much broader set of social interventions may also be required to change the trajectories of crime at crime hot spots. The focus on the specific places where crime problems are found provides an opportunity to "lower the scale" of social interventions and accordingly to make such interventions relevant to crime prevention practitioners. It is one thing to attempt change in the social conditions of an entire neighborhood or city. It is another to try to ameliorate problems on specific blocks. Perhaps the criminology of place provides a scale of intervention that can rekindle interest in the importance of social and structural interventions in doing something about crime. The work of Weisburd et al. (2012) suggests the importance of focusing crime prevention, whether it is at the level of local police agents or in terms of the development of social programs for hot spots of crime. Whatever the approach that is taken, it is time to recognize the need to focus crime prevention resources at micro places such as street segments.

It is important to note that despite the wealth of data Weisburd and colleagues (2012) were able to gather on street-segment characteristics, they were limited because of retrospective data collection. These data availability issues suggest the importance of a prospective longitudinal study of crime at place that would capture at specific times both the characteristics of places and people (see Weisburd, Lawton, and Ready 2012). The National Institute on Drug Abuse within the National Institute of Health recently provided funding for such a longitudinal study to take place. Weisburd and colleagues were awarded over $3 million to undertake a 5-year study of how living in drug or violent crime hot spots influences personal health, mental health, HIV and sexually transmitted disease rates, safe sex practices, drug use, crime, and other antisocial behaviors. The study will also develop knowledge on why places become drug or violent crime hot spots and how characteristics of street segments and their residents impact upon developmental trends of health, drug use, and crime. Results from this study should provide exciting new insights into the criminology of place.

Related Entries

- ► Geography of Crime and Disorder
- ► History of Geographic Criminology Part I: Nineteenth Century
- ► History of Geographic Criminology Part II: Twentieth Century
- ► Hot Spots and Place-Based Policing
- ► Law of Crime Concentrations at Places
- ► Longitudinal Crime Trends at Places

Recommended Reading and References

Bernard TJ, Snipes JB (1996) Theoretical integration in criminology. In: Tonry M (ed) Crime and justice: a review of research, vol 20. University of Chicago Press, Chicago, pp 301–348

Braga AA, Weisburd DL (2010) Policing problem places: crime hot spots and effective prevention. Oxford University Press, New York

Braga AA, Papachristos AV, Hureau DM (in press) The effects of hot spots policing on crime: an updated systematic review and meta-analysis. Justice Q

Brantingham PJ, Brantingham PL (eds) (1981) Environmental criminology. Sage, Beverly Hills

Brantingham PJ, Brantingham PL (1984) Patterns in crime. Macmillan, New York

Bursik RJ Jr, Grasmick HG (1993) Neighborhoods and crime: the dimensions of effective community control. Lexington Books, New York

Clarke RV (1983) Situational crime prevention: its theoretical basis and practical scope. In: Tonry M, Morris N (eds) Crime and justice: a review of research, vol 14. University of Chicago Press, Chicago, pp 225–256

Clarke RV (1995) Situational crime prevention. In: Tonry M, Farrington DP (eds) Building a safer society: strategic approaches to crime prevention. Crime and justice: a review of research, vol 19. University of Chicago Press, Chicago, pp 91–150

Clarke RV, Cornish DB (2001) Rational choice. In: Paternoster R, Bachman R (eds) Explaining criminals and crime. Roxbury Publishing, Los Angeles

Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44:588–608

Eck JE, Eck EB (2012) Crime place and pollution: expanding crime reduction options through a regulatory approach. Criminol Public Policy 11(2):281–316

Eck JE, Weisburd D (1995) Crime places in crime theory. In: Eck JE, Weisburd D (eds) Crime and place. Crime prevention studies, vol 4. Willow Tree Press, Monsey, pp 1–33

Gottfredson M, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Jeffery CR (1971) Crime prevention through environmental design. Sage, Beverly Hills

Laub JH, Sampson RJ (2003) Shared beginnings, divergent lives. Delinquent boys to age 70. Harvard University Press, Cambridge

Martinson R (1974) What works? Questions and answers about prison reform. Publ Interest 35:22–54

National Research Council (2004) Fairness and effectiveness in policing: the evidence. Committee to review research on police policy and practices. In: Skogan W, Frydl K (eds) Committee on Law and Justice, Division of Behavioral and Social Sciences and Education. National Academies Press, Washington, DC

Newman O (1972) Defensible space: crime prevention through environmental design. Macmillan, New York

Rice KJ, Smith WR (2002) Sociological models of automotive theft: integrating routine activity and social disorganization approaches. J Res Crime Delinq 39:304–336

Sampson RJ, Groves WB (1989) Community structure and crime: testing social-disorganization theory. Am J Sociol 94:774–802

Sampson RJ, Raudenbush SW, Earls F (1997) Neighborhoods and violent crime: a multilevel study of collective efficacy. Science 277:918–924

Shaw CR, McKay HD (1942) Juvenile delinquency and urban areas. A study of rates of delinquency in relation to differential characteristics of local communities in American cities. University of Chicago Press, Chicago

Sherman LW (1995) Hot spots of crime and criminal careers of places. In: Eck J, Weisburd D (eds) Crime and place. Crime prevention studies, vol 4. Willow Tree Press, Monsey, pp 35–52

Sherman LW, Gartin P, Buerger ME (1989) Hot spots of predatory crime: routine activities and the criminology of place. Criminology 27:27–55

Smith WR, Frazee SG, Davison EL (2000) Furthering the integration of routine activity and social disorganization theories: small units of analysis and the study of street robbery as a diffusion process. Criminology 38:489–523

Taylor RB (1997) Social order and disorder of street blocks and neighborhoods: ecology, microecology, and the systemic model of social disorganization. J Res Crime Delinq 34:113–155

Weisburd D, Piquero AR (2008) How well do criminologists explain crime? Statistical modeling in published studies. In: Tonry M (ed) Crime and justice: a review of research, vol 37. University of Chicago Press, Chicago, pp 453–502

Weisburd D, Bushway S, Lum C, Yang S-M (2004) - Trajectories of crime at places: a longitudinal study of street segments in the city of Seattle. Criminology 42:283–321

Weisburd D, Bruinsma GJN, Bernasco W (2009) Units of analysis in geographic criminology: historical development, critical issues, and open questions. In: Weisburd D, Bernasco W, Bruinsma GJN (eds) Putting crime in its place: units of analysis in geographic criminology. Springer, New York, pp 3–31

Weisburd D, Groff ER, Yang S-M (2012) The criminology of place: street segments and our understanding of the crime problem. Oxford University Press, New York

Weisburd D, Lawton B, Ready J (2012) Staking out the next generation of studies of the criminology of place: collecting prospective longitudinal data at crime hot spots. In: Loeber R, Welsh BC (eds) The future of criminology. Oxford University Press, New York, pp 236–243

Wicker AW (1987) Behavior settings reconsidered: temporal stages, resources, internal dynamics, context. In: Stokels D, Altman I (eds) Handbook of environmental psychology. Wiley-Interscience, New York, pp 613–653

Wilcox P, Land KC, Hunt SC (2003) Criminal circumstance: a dynamic multicontextual criminal opportunity theory. Walter de Gruyster, New York

Criminology of Places

► Criminal Careers of Places

Critical Criminology

► International Responses to Victims in Criminal Justice

► Marxist Criminology

► So What Criminology?

Critical Incidents

Tom Mijares[1] and Ron McCarthy[2]
[1]Texas State University-San Marcos,
San Marcos, TX, USA
[2]Director Emeritus, National Tactical Officers
Association, Los Angeles Police Department
(Retired), Doylestown, PA, USA

Overview

This entry begins with a brief look at the traditional approach to critical incident management by individual law enforcement officers, placing particular emphasis on the recognition by the judiciary that special circumstances require special responses in the form of highly selective processes for selecting, instructing, and equipping the personnel tasked with these duties in an increasingly more violent and complex society. This entry then describes how social and technological changes led to the development of the current approach to incident control and problem resolution. The authors also summarize the controversies coming from this approach which features small, well-trained, and well-disciplined groups working as a cohesive and coordinated unit without depleting the law enforcement agency's ability to respond to more routine calls for police service. The entry describes projections for the future of critical incident management based on the direction, frequency, and magnitude of current trends. The authors conclude with recommendations for future operations.

Historical Patterns

Traditional methods of American law enforcement have centered on the beat cop making his rounds on foot and later in a motorized vehicle to take the appropriate action through his own initiative and limited resources. Backup assistance and radio calls to the station for advice were technically impossible until the relatively recent times of the middle twentieth century. Specialized response capabilities were initially limited to horse-mounted police officers and were improved later by the advent of the automobile and the "flying squads" of major city detective bureaus.

Labor unrest during the 1930s demonstrated the need for specialized police units possessing the means for immediate deployment, containment, and arrest capabilities to deal with unusual circumstances such as barricaded suspects, demonstrations, strikes, and other forms of unrest in a rapidly changing society.

In 1931 the National Commission on Law Observation and Enforcement, commonly referred to as the Wickersham Commission, had been charged with the responsibility of reporting on the enforcement of the Eighteenth Amendment (Prohibition). The Commission's 14-volume series of reports probed deeply into the American criminal justice system to offer the causes and effects of the situation and to suggest possible ameliorative efforts (Bopp and Schultz 1972).

As police administrators recognized the need to maintain group integrity and cohesion for these critical incidents, specialized units were developed. Despite the assumption of elitism in the creation and maintenance of these units, the following shortcomings of specialized operations had already been identified by the Wickersham Commission:

1. The selection process for most police appointments and assignments was often based on nepotism and favoritism.
2. Training was irregular, inconsistent, and incomplete. As a result, a greater amount of force was used than what may have been necessary.
3. No continuous and concerted effort was made in large departments to integrate these units with other sections of the police departments.
4. Little research was conducted to find improvements to existing techniques and technology.
5. Specific organizational orders were not developed to define the deployment procedures to be followed by specialized units, and the lines of authority, responsibility, accountability, and communications were obscure.

An Increase in Violence in a More Complex Society

From the middle and late 1960s and early 1970s to the present, social and technical factors worked to change the complexion of law enforcement in general and tactical policing in particular. First, the war in Vietnam and other conflicts resulted in a quantum increase in the number and variety of weapons available for law enforcement applications. Unfortunately, criminal elements throughout the world had (and continue to have) access to the same technology without the legal, social, and economic constraints endured by law enforcement agencies (Dobson and Payne 1982).

Second, America's space exploration program and its by-products have also had an impact on the technology of law enforcement and criminality. Limited space for orbiting payloads resulted in smaller, more efficient radio communications systems. Landline call boxes and one-way radios in police cars were expanded to full two-way portable communications. Where law enforcement had traditionally been dependent on the unplanned and unrelated efforts of patrol officers acting individually, the improvement in communications allowed all involved law enforcement officers the opportunity to coordinate their efforts more effectively in surveillance, pursuits, high-risk warrant service, tactical operations, and other activities that optimally require several individuals to act as a unit.

Third, the frequency and magnitude of special threats increased beyond the imagination of the beat cop of a few decades ago. Regardless of the size of a law enforcement agency's jurisdiction, it is extremely naive for a police administrator to ignore the possibility of extreme violence. History has already shown that these events are not limited by demographic and geographic factors.

Finally, the effects of technical sophistication were accelerated by the development of criminal groups from all points on the political-philosophical spectrum and a willingness to collaborate against a perceived common enemy, i.e., the "establishment." What is particularly disturbing for law enforcement has been the realization that irrespective of their philosophies and ultimate goals, some of these groups share the commonalities of impressive armories of military weapons, the sophisticated training necessary to employ them, a proclivity for violence, and a total commitment to their causes. The multifaceted composition of the American populace suggests that even more dissident groups will develop.

Events such as the Watts Riot in 1965, the Texas Tower incident in 1966, and the Howard Johnson shootout in New Orleans in 1973 held several separate but interwoven ramifications for the conduct of future law enforcement operations. First, they demonstrated that law enforcement technology was not yet sufficient to deal with many current problems. No longer was military surplus equipment adequate or appropriate for domestic applications. The Texas Tower incident in particular established an awareness of the need for a more sophisticated approach to communications.

Second, these incidents also led to an awareness of the need for firm, effective crowd and traffic control. Considering how numerous people could have wandered into the field of fire in these commercial and residential areas during these critical incidents, a major factor in the successful resolution of the incident may be attributed to the fact that the responding officers were able to isolate the perpetrators and that they made considerable efforts to block traffic from the area.

Third, these events illustrated how the complexion of many types of crimes was changing. The availability of large-capacity firearms and the sometimes direct involvement of a great number of people, both as perpetrators and as victims, were examples of how criminal activity was evolving and how law enforcement was forced into modifying its approaches accordingly.

Fourth, alterations in criminal activity could also be seen by changes in overall strategies and specific tactics employed by the perpetrators. For the first time in the modern history of American law enforcement, responding police forces were confronted with perpetrators who were actually willing to die for their cause.

Fifth, where news coverage of these events had previously been limited to next-day reporting and analysis, technical advances in the news and entertainment media provided immediate and on-scene telecast capability. Critics of law enforcement were given a means to record and scrutinize the substance of what had occurred and how the responding police forces responded to the incident. Improved news media coverage fostered another phenomenon: religiously, philosophically, and politically differing groups such as the Symbionese Liberation Army, the Branch Davidians, and The Order often achieved public martyrdom among some segments of society for their actions.

Sixth, responding to critical incidents on an ad hoc basis led to a breakdown in the normal chain of command. Without this necessary element of organizational management, most of the officers who had been assigned respond to critical incidents simply were forced to standby until appropriate orders were received. However, during the Watts Riot a very small number of officers from the Metro Division of the Los Angeles Police Department immediately recognized the confused, overly cautious, and politically dictated approach displayed by the command staff. These officers had been assigned to the Division in part because of their long-established ability to exercise legitimate and purposeful initiative and because of their skill in employing aggressive patrol techniques. The officers recognized the threats and hazards of allowing the rioters to continue unchecked. Because these officers had often worked together as a cohesive unit, they were able to use an established principle of crowd control by identifying and removing the leaders of the riot, thus reducing the crowd to a collection of individuals without direction.

Finally, these incidents pointed to the need for law enforcement specialists who are trained to plan for and deal with unusual events as a coordinated unit without depleting the law enforcement agency's personnel or ability to respond to routine calls for police service. Special circumstances require specialized preparation in terms of personnel selection, technology, training, and tactics. Under the guidelines of *Downs v. United States* (1975) and *City of Winter Haven v. Allen* (1991), a law enforcement agency could no longer assume that every member is prepared or equipped to perform any task on an ad hoc basis in response to this sort of increasingly complex and heinous criminal activity. This need has proved to be particularly true in large metropolitan areas where multiple calls for this type of police service have become commonplace. Specialized preparation revolves around an organization whereby all functions associated with these perplexing problems can be performed in an orderly fashion without disruption to the other elements of the overall law enforcement agency. The term organization implies more than a recognized position in an organizational chart. The term carries with it the implication that any specialized component possesses the appropriate authority and official means to perform any tasks for which it is responsible and that it receives the necessary administrative support to perform these tasks. Events such as the Munich massacre in 1972 demonstrated the same law enforcement shortcomings on an international basis (Mijares and Mijares 1994).

Present Trends in the Response to Critical Incidents

Over the past 40 years, police administrators have progressively relied upon the concept of rigorously selected personnel who have been trained and equipped with special weapons and tactics (SWAT) to respond to these incidents while allowing the uninterrupted ability to respond to more routine calls for service. Since its formative years the SWAT concept has increased dramatically from a small number of large urban police departments to an almost universal acknowledgement of the guidelines defined in *Downs v. United States* and *City of Winter Haven v. Allen* recognized that special circumstances require special responses. When law enforcement agencies were financially unable to develop a full-time, organizationally dedicated tactical response capability, part-time SWAT units and mutual assistance programs were developed (Perkins and Mijares 1994).

Where officers were originally wearing military surplus fatigue uniforms and baseball caps, today's operators don padded, fire- and ballistic-resistant, and slash-impervious protective gear and fiberglass helmets. Where the original tactical personnel were armed with revolvers and pump shotguns, present SWAT personnel may select from a wide array of weapons and other equipment. Where yesterday's officers were trained only on localized barricaded suspect and crowd control scenarios, the currently available training classes are many and varied through the state, regional, and national tactical training associations.

Law enforcement managers and SWAT commanders have recognized that the future, as it relates to law enforcement's obligation to be prepared to protect the citizens it serves, can be predicted. Economics and joblessness are commonly used predictors of crime in our society. Unpredictable calamities that are certain to take place, such as natural disasters, leave uncertainty as to the exact location, date, and time, but it is relatively certain that earthquakes will occur in California, tornados will occur in the Plains States, and floods will occur along the rivers. Accompanying these catastrophes and the damage they do is the predictable lawlessness on the part of some elements in society who see destruction of security, property, and chaos as opportunity to loot and assault. It is the mission of SWAT teams to provide a means for law enforcement to respond efficiently and effectively to all types of critical incidents in such a manner that routine calls for police service need not be ignored (Mijares and McCarthy 2008).

The changing culture in all countries introduces crimes at a level and intensity not seen before. Active shooters have become a real and very dangerous problem throughout the world. The tragedies at Columbine High School and at the Texas Tower are two early examples. Societies that believe they are culturally immune from this phenomenon are not. More recent example took place in Oslo, Norway, Virginia Tech University, and Aurora, Colorado. These latest attacks demonstrated how a single suspect can easily kill dozens of people. Thus, it is highly probable that mass murderers will continue to attack, especially where it is perceived that "It can't happen here," is a tenet of incompetent law enforcement managers and unrealistic politicians.

Terrorism driven by both religious fanaticism and political extremism and the forecast of mass murder and civil disturbances created by extreme ideology are happening and will most likely continue. International terrorist groups such as Al Queda and its affiliates have planned and accomplished large-scale attacks in major cities throughout the world, e.g., New York in 2001, Madrid in 2004, Mumbai in 2008, and Moscow in 2009. Terrorist communications have been intercepted that tell us they intend to hit smaller targets in locations that are less prepared. The intelligence gathered on these groups suggests that large cities must continue to build and refine their tactical response capabilities and smaller jurisdictions must catch up.

The Las Vegas Metro Police, Los Angeles Sheriff's Department, Los Angeles Police Department, and Orange County California Sheriff's Department have created a response known as Multiple Assault Counter-Terrorism Action Capabilities (MACTAC). These very large and capable agencies have joined together and combined their resources and training to be able to deploy their SWAT teams and other assets immediately in the jurisdiction that is under attack. This tactic allows the city or county team that is initially on scene to expect that as deployment is initiated, three additional teams from large agencies are already on their way with tactics and techniques have been practiced and will be immediately utilized. Some smaller agencies have begun to emulate this model and regionalize their training, logistics, tactics, and communications, so they have the same type of response capability. The MACTAC concept is being introduced nationally.

A similar approach is being used in the area of gathering and analyzing intelligence in the law enforcement context. Usually described as "fusion centers," this concept is the result of a joint project among elements of the Department of Homeland Security and the Department of

Justice and assigned personnel from state, county, regional, and local police departments. This approach involves the sharing of information on matters of mutual concern. Unlike the infamous "red squads" of a previous generation of law enforcement, fusion centers do not merely gather and store information but instead they analyze and disseminate it when verified and validated.

Civil disturbance is a serious threat at all times in our free and open society with the myriad special interests, cultures, and human greed that exist in a heterogeneous society. By comparison, Japan had very little uncivil behavior and virtually no looting during the latest natural disasters. When there is an "excuse" for a riot in an open society, there is no shortage of willing participants. The victory of a sports team, either college or professional, can elicit mass civil disturbance. A hurricane can be the "opportunity" for the unprincipled elements in our society to assault, steal, burn, and murder.

Law enforcement SWAT units must be ready and must train in joint exercises to protect those who cannot protect themselves and have a right to live in peace. Because of their high standards and advanced training, their organizational flexibility, and their disciplined approach to problem resolution, today's SWAT teams have become much more than their original predecessors. They have become the operational first-response arm to thwart terrorism on the domestic front. Their ability to continue in this role is dependent on the perception of a threat by law enforcement leaders and elected officials. It is also dependent on a universal realization that the privacy rights and freedom of movement desired by all citizens in a free and open society also extend to those who would take them from us. Finally, it will depend on the resolve and ability of law enforcement personnel to balance the obligation to protect and serve with the constraints and privileges of the United States Constitution.

The SWAT concept of small units has not been without its criticism. Much of this reproach has focused around claims that there has been an overuse of military metaphors and intimidation, intraorganizational elitism, and allegations that the concept contradicts the community policing model, and that the overall approach is unnecessarily expensive and even counter-productive on the basis of cost-benefit analysis (Kraska and Kappeler 1997). However, these criticisms have been rejected on both methodological and substantive grounds that have led to inaccuracies in conclusions (Klein 2005).

Policing Critical Incidents in the Future

Past events and current trends can provide a reasonable projection of likely scenarios and conditions under which tactical operations will be conducted in the future. Like all other aspects of modern society, police operations in general and tactical operations in particular will be greatly affected by advances in technology. Equally important, changes in society, fluctuations in the economy, variations in political demands, and continuous modifications of legislation and judicial interpretations will also make an impact on the conduct of law enforcement policies and procedures. Unfortunately these advances are likely to benefit the criminal perpetrator as well as the law enforcement responders to criminal actions. Consequently, there is no shortage of problems likely to be encountered in the future.

Increased Encounters with Juveniles

A review of the *Uniform Crime Reports* over the past several years has shown a decrease in the crime rate among the adult population. However, a similar review would produce evidence that the tendency among juvenile offenders is the opposite. Accordingly, likelihood is high that there will be an increase in encounters with heavily armed juveniles by SWAT units both in terms of frequency and intensity.

A particularly disturbing aspect of this projection is the observed increase in acts of violence involving multiple victims perpetrated by juveniles against other juveniles. These acts have been manifested through schoolyard and theater shootings in various parts of the country. They have also been seen among the youthful participants in the fiercely competitive subculture of illegal drugs. Although these events appear to have been committed independently, a disturbing set of common characteristics has developed which may have important consequences for law enforcement in general and tactical operations in particular. First, the economic and social backgrounds of the perpetrators vary considerably. In addition, the geography of these incidents suggests that similar incidents can take place at any location. Where most young criminals surrendered to the superiority of responding police units in previous incidents, today's youthful offenders often refuse to lay down their weapons and show no reluctance to stand fast in an escalation of the situation into full-scale armed combat with the police. Whether this tendency is the product of misguided juvenile feelings of invincibility and immortality or simply a reflection of rising overall societal violence, the police face a difficult dilemma. If the officers are forced into using fatal force against juvenile offenders, they will be characterized as overzealous and trigger-happy. If they approach a critical incident with the deliberation shown by first responders to the Columbine High School incident of 1997 when they attempted to subdue and arrest the perpetrators, they will be criticized for being too slow to respond and uncaring about the perpetrator's victims. Irrespective of the action taken, the responding law enforcement personnel will be criticized more severely and from more directions than ever before.

Suicide by Cop

Suicide is hardly a new phenomenon. Various politicians, mental health professionals, and attorneys filing law suits against the police occasionally suggest that the police officers who confront suicidal suspects should use something besides deadly force to neutralize the situation. However, officers cannot be expected to accomplish immediate psychological evaluations and identify the mental and emotional problems leading to these suicidal incidents nor can they be expected to discover a solution to the perpetrator's problems in a few minutes when psychologists and psychiatrists who have given several hours of clinical sessions to their patients spread out over several months cannot do so. Some professionals such as psychologists, attorneys, and academicians have suggested that the police should have the skills to deal with the suicidal individual. At least they suggest that the police should be able to call a mental health professional to the scene to be utilized in some way to prevent a negative outcome. However, the availability of such a resource simply does not exist for most police departments. What mental health professional is available on Saturday at 2:20 a.m.?

The law enforcement element of the criminal justice system, including well-trained tactical officers, may not be able to address the psychopathological needs of suicidal individuals in general or victim-precipitated shootings by police in particular. However, through training, education of the public, and continuous research, the impact of this phenomenon on law enforcement can be assuaged sufficiently to allow its etiology to identify appropriate remedial and preventative measures.

Terrorism and Police Involvement with Weapons of Mass Destruction

In the absence of a formal declaration of war, most terrorist incidents that could be perpetrated within the borders of the United States are criminal acts. Although they may differ from the usual criminal acts in their gravity, the enormity of the criminal acts committed by these heinous offenders is so extraordinary that many citizens expect a military response or from a federal law enforcement agency. However, the response, particularly at the initial stage of the incident, will inevitably be performed by local law enforcement officers. Because of the magnitude of these events, a tactical unit will be directed to resolve the situation. Whether it is a mass homicide with firearms or a suicide bombing, the accelerated violence characteristic of a terrorist incident will require a response beyond the capability of a single officer or even group of individual officers. It will require the attention of a cohesive and coordinated unit. In addition, the extreme violence of terrorism is likely to be imitated and displayed by members of the illegal drug trade as well as by delusional individuals who have acquired the tools of the terrorists. Even without a political, social, or religious agenda, these individuals cannot be taken lightly by law enforcement.

Thus, it will become increasingly important for police officers to stay abreast of world events as to develop and hone their tactical skills. Where police officers may have relied solely on their respective intelligence units for information on criminals, tactical officers will need to stay more active in their pursuit of information.

Terrorist events could be catastrophic. Only with a coordinated and cooperative response can the situation be managed. If such an incident were actually to take place, direct involvement by SWAT personnel alone may be unlikely. Representatives from fire departments, public health agencies, hazardous material units, and explosive ordnance disposal squads would be called on to neutralize the threat and to relieve any effects. However, Hillman (1999) suggested that SWAT personnel will have specific roles to fulfill in order to allow these other units to complete their missions safely and expeditiously:

1. They will accompany threat neutralization and medical treatment personnel into contaminated areas, particularly when there is a need to confront suspects and to control injured victims.
2. They will provide security for these specialized personnel.
3. They will provide much of the labor-intensive tasks of identifying, seizing, marking, and preserving physical evidence in such a manner that the custody and chain of evidence can be legally maintained for later criminal prosecution.
4. They will assist with the evacuation of intended but unaffected victims. These individuals, though not necessarily symptomatic, may require quarantine procedures and may be in a highly agitated and irrational state. The demeanor, expertise, and methods employed by police personnel will be a major factor in maintaining relative calm and avoiding an intensification of the problem.

Changing Environments

During the formative years of the SWAT concept, most of the violent encounters were against barricaded suspects in homes or businesses. These encounters were usually the result of escalated family disputes or aborted robberies. While these trends continue today, there is a need for a tactical response in locations that historically were not just unthinkable, but were often regarded even by most criminals as sacrosanct. The shooting of 90 young people at a youth camp in Norway is a case in point. Whatever the motivation, first responders must become familiar with operating in environments and under conditions that are very different from the traditional residence and business place. This familiarization will involve not only differences in physical environments but also diversity in the motivations of perpetrators as well as the reactions by the victims. In most cases, the use of violence by criminals was primarily intended to be an aid for escape with an occasional outburst by an extremist crusader with an agenda. Based on statements made by captured perpetrators, each copycat use of violence appears to be an attempt to be even more grandiose than its predecessors and to leave a mark that will not be soon forgotten.

Crime as a Diversion

It is reasonable to assume that criminals and terrorists know and understand at least some of the basic requirements of a police tactical response. With this information it would also be reasonable to assume that terrorists or criminals would precede a major incident with a smaller but extraordinary event that would divert police assets to a geographical location within a jurisdiction that would make response to the primary event limited and very difficult. It is also prudent to train and familiarize other officers with at least a rudimentary ability to respond to further incidents when the primary tactical unit is already engaged. Officers assigned to the function of providing a holding action need not be permanently assigned SWAT personnel but could be taken from a list of officers who may have been selected for training and later permanent assignment to the unit.

Recommended Changes in Approaches to Critical Incident Resolution

Maintaining Standards for the Profession

A certain measure of creativity and modification are often expected and occasionally necessary for the resolution of the myriad of tactical problems expected to be encountered by a modern SWAT team. However, the actual execution of plans must stay within the parameters of the United States Constitution, existing legislation, current case law, department policy, and standards of the profession. Standards must be met in a variety of ways through organizational readiness (*City of Winter Haven v. Allen* 1991), personnel procedures (*Moon v. Winfield* 1973), equipment renovation and replacement, and tactical applications (*Downs v. United States* 1975), in order to be consistent with standards of the profession. It is incumbent upon all law enforcement personnel to stay abreast of these standards and in any changes in the state of the art.

Such information is readily available through several different media. Attending the conferences of the various national, regional, and state tactical officers associations provides current information on the development of new techniques as well as an opportunity to view demonstrations of innovations in the technology. Equally important, these conferences provide association members an opportunity to exchange case histories and opinions regarding approaches to be followed under similar circumstances. Printed articles in professional magazines such as *Law Enforcement Technology*, the *Tactical Edge*, and the *FBI Law Enforcement Bulletin* furnish information on the newest developments in equipment and its most effective use. Academic journals such as *Police Liability Review* and *Journal of Contemporary Criminal Justice* provide commentary on relevant and directly applicable court decisions and legislation. In short a large volume of substantive information is available for the active learner in this area.

Modifications in Approaches to Training

The anticipated problems plus unforeseen others that will certainly occur will require tactical units to incorporate a variety of innovative approaches to situational preparation. The types of problems anticipated to be encountered in the future by SWAT teams suggest that the resolution will frequently require more than one organization. The trends toward incidents of greater scale and which often cross jurisdictional lines will require the cooperative effort of several law enforcement agencies and support organizations. Some police departments automatically mobilize assistance from the local fire department and emergency medical services. However, the likelihood of incidents involving weapons of mass destruction will involve the responses from state and federal personnel.

Whether the support comes from within the organization or from external sources, the possession of individual skills and knowledge is insufficient without coordination of effort. Without the coordination coming from supervised joint training and participation in mock disaster drills, the effectiveness of any attempts to resolve these situations is very limited.

Conducting Continuous Research

Scientific research is an organized search for the truth involving problem identification, parameter definition, data collection and analysis, and the realization of a research decision. In the field of criminal justice in general and tactical operations in particular, applied research could address issues such as techniques, technology, and legal issues. While there may not be a legal requirement specifically mandating any form of research, the ability to display and document a continuous effort to find solutions for improving SWAT responses is certainly a helpful approach in promoting the image of any tactical unit and its management.

As an example, a tactical unit may wish to improve its ability to respond to barricaded suspect situations by examining the efficacy of various pieces of intelligence-gathering equipment or sensory-enhancing technology such as a thermal imaging surveillance system. The research decision would very likely involve a recommendation to conduct follow-up studies on the most effective techniques for using the equipment during a tactical scenario. It would likely also include a recommendation to conduct further research on the issues associated with the legal use of the equipment.

Conclusion

Extreme violence requiring a coordinated police response is a phenomenon with no foreseeable end. Most of the critical incidents requiring a tactical response are still committed by individual perpetrators as a result of either aborted criminal activity or personal psychological disturbances. Consequently tactical preparations in the form of training, equipment, and tactical responses continue to focus on the individual criminal.

However, a disturbing trend has been a growth in the number and strength of various well-organized domestic terrorist groups. Often using freedom of religion as a justification for their crimes and a self-serving brand of patriotism as a rallying cry, these groups seem to be particularly adept at attracting popular support by concentrating on their philosophy and ignoring their criminal activity.

Law enforcement management will continue to be responsible for protecting the peaceful majority of society from those who would use the available opportunity and technology to further their criminal motivations. Whether the criminal is an individual or a group, motivated by economic gain or political and religious philosophy, rational and calculating, or psychologically maladjusted is irrelevant. Law enforcement will continue to deal with these problems as they arise. The managers of police agencies will be required to anticipate these problems, develop proactive solutions, stay within a limited budget, and still provide routine services with minimal disruption. Of vital importance is the requirement that all of these tasks be done while staying within the guidelines of existing case law and legislation and protecting the agency and its employees from avoidable litigation.

The success of future police tactical operations during critical incidents will require active and supportive management. Success or failure will be highly dependent on how well management can assess agency and community needs through scientific research, provide relevant training for police personnel, utilize current and developing technology from many different fields, coordinate both internal and external activities toward the achievement of agency goals, and account for the entire organization with both objectivity and sensitivity.

Related Entries

- ▶ History of Technology in Policing
- ▶ Hot Spots and Place-Based Policing
- ▶ Juvenile Violence
- ▶ Police and the Military Nexus
- ▶ Police Use of Firearms
- ▶ Policing in Developing Democracies
- ▶ Policing of Peacekeeping
- ▶ Policing Terrorism and Legitimacy
- ▶ Risk Management in Policing
- ▶ Strategies of Policing Terrorism
- ▶ Terrorist Organizations

Recommended Reading and References

Bopp WJ, Schultz DO (1972) A short history of American law enforcement. Charles C. Thomas, Publisher, Ltd, Springfield

City of Winter Haven v. Allen, 541 So.2d 128 (Fla. App. Dist. 1991) and 689 So.2d 968 (Fla. App. Dist.1991)

Dobson C, Payne R (1982) The terrorists: their weapons, leaders, and tactics. Facts on File, New York

Downs v. United States, 522 F.2d 990 (6th Cir.1975)

Hillman M R (1999) Biological/chemical terrorism and SWAT response. The Tactical Edge 17(3)

Klein G (2005) The militarization of the police? Or ten ways Dr. Kraska got it wrong. Police Forum (14, 5)
Kraska P, Kappeler V (1997) Militarizing the American police: the rise and normalization of paramilitary units. Social Problems (44,1)
Mijares T, McCarthy R (2008) The management of police specialized tactical units, 2nd edn. Charles C. Thomas, Publisher, Ltd, Springfield
Mijares T, Mijares K (1994) The Munich massacre: its effects on law enforcement. Command, Summer
Moon v. Winfield, 306 F. Supp. 843 (1973) and 383 F. Supp. 31 (1974)
Perkins D, Mijares T (1994) Police liability issues associated with inter-agency mutual assistance pacts. Police Liability Review 8

Critical Report on Forensic Science

Robert E. Gaensslen and A. Karl Larsen
Forensic Science Group, Department of Biopharmaceutical Sciences, College of Pharmacy, University of Illinois at Chicago, Chicago, IL, USA

Overview

In the USA, the National Research Council (NRC) Committee on Identifying the Needs of the Forensic Sciences Community issued a lengthy report based on several years of hearings and study. The report contained 13 major recommendations covering a wide spectrum of issues and disciplines within the forensic sciences. The NRC is an arm of the National Academies, which are often regarded as the most prestigious and authoritative bodies on matters of science, medicine, and engineering. The NRC has previously issued reports on forensic DNA technology (NRC 1992, 1996), but had never undertaken a comprehensive evaluation of forensic sciences before.

The NRC report was directed by the US Congress, prompted by a number of factors and developments. Two important ones were the advent of DNA profiling of biological evidence and a major US Supreme Court decision in 1993 (Daubert et ux. etc. 1993) setting down the criteria for the admissibility of scientific evidence in the courts. *Daubert* was the first case on admissibility of scientific evidence ever considered by the Supreme Court. The decision emphasized that admissibility was to depend on a sound scientific basis for the proffered evidence. DNA typing was always considered to have a sound scientific basis, and so became a sort of "paradigm" for the rest of the forensic sciences to follow. Almost everyone, critics and defenders alike, agreed that the forensic sciences were under-resourced. Thus, was born the idea that a comprehensive study was needed and that NRC was the body to conduct it. Authorized in 2005, the study was completed and the report issued in 2009. This entry will mainly discuss the NRC report on forensic science, highlighting the main recommendations, some of the background leading up to it, and some of the reaction to it.

Background and Introduction

For almost 200 years, forensic laboratory and medicolegal services have been provided by government agencies. In the nineteenth century, most of these services were provided by medicolegal institutes in continental Europe and in a few of the larger cities in the US Forensic laboratories, as we know them today, developed during the twentieth century. In the years before the Second World War, several forensic laboratories were established in the USA. They were housed within the Federal Bureau of Investigation (FBI) and in some larger municipal police departments, such as the Los Angeles Police Department. With time, state and county forensic-science labs were established. The Omnibus Crime Control and Safe Streets Act of 1968 (P.L. 90-351) provided a considerable amount of funding for forensic laboratory expansion, mainly through block grants to the various states. More recently, considerable federal funding for the support and expansion of DNA technology in forensic labs has been available.

This entry will mainly discuss the NRC report on forensic science, some of the background leading up to it, and some of the reaction to it. Another report, on the forensic science services (FSS) in the UK, will be addressed in another contribution. Unlike the US report, the emphasis on the British report (from a Committee of the House of Commons) is on the implications of the government's plan to shut down the forensic science service (FSS), essentially ending government-supported forensic-science services in the UK (see ▶ Organizing and Supplying Forensic Science Services). Recently, Ross commented on the NRC report from the standpoint of the forensic sciences in Australia and New Zealand (Ross 2011). He was looking at the forensic sciences in those countries through the lens of the relevant issues considered by the NRC committee that produced the report in the USA. Of special interest in this treatment is the fact that Australia has a National Institute of the kind recommended by the NRC report for the USA.

The National Research Council (NRC) of the National Academies is a highly prestigious body that assembles expert panels from the scientific, medical, and technological world to look at difficult or controversial questions and make recommendations. This is the first time the NRC has comprehensively examined forensic sciences and services, and its findings and recommendations have been widely cited and quoted. The US Congress, various executive branch agencies, and the courts are likely to rely on them.

As noted, a number of factors coalesced to prompt Congress to commission the NRC report. In the *Daubert* decision, the US Supreme Court emphasized the need for the "scientific" results proffered as evidence in a courtroom to be based on solid scientific principles gathered through hypothesis testing. The court emphasized the concept of empirical testing of a potential falsifiable hypothesis, a notion set forth by Popper (1989). In subsequent decisions (General Electric v Joiner 1997; Kumho Tire v Carmichael 1999), the court extended these principles to "technological" evidence elicited by way of expert testimony. In effect, the "Daubert trilogy," as the three cases are sometimes called, modified and sharpened the focus of Rule 702 of the Federal Rules of Evidence. The Federal Rules of Evidence are established by the federal judiciary and govern the federal courts. There are several of these rules that pertain to forensic science and expert witnesses, but 702 is probably the most critical one. Rule 702, Testimony by Expert Witnesses, states: A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if: (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue; (b) the testimony is based on sufficient facts or data; (c) the testimony is the product of reliable principles and methods; and (d) the expert has reliably applied the principles and methods to the facts of the case.

Until *Daubert*, an older 1923 District of Columbia appellate court case called *Frye v. US* was widely used as the standard for scientific evidence admissibility (Frye 1923). Under *Frye*, the determinant was "general acceptance in the scientific community" to which the subject area belonged. There were legal debates through the years over what constituted "general acceptance" and over the definition of the relevant "scientific community." *Daubert* set down six criteria that a judge could use in deciding on admissibility: (1) Is the proposition testable? (2) Has it been tested? (3) Are there accepted standards? (4) Has there been peer review and/or publication? (5) Are the procedures generally accepted? (6) Does the procedure have a known error rate, and if so, what is it? The decision made the judge the "gatekeeper." He/she has considerable latitude in using the stated criteria to decide on admissibility. The Court has made clear these criteria do not constitute a "checklist." Strictly speaking, the US Supreme Court decision only applied to the federal courts, but many states (around 30 as of 2010) have adopted it (or its basic principles) as well. Those that have not adopted it continue to use the Frye precedent, or a modified version of it. In the wake of *Daubert*, there was considerable commentary that DNA typing and profiling constituted a sort of paradigm of proper scientific method, and other forensic disciplines should be held to similar scientific standards. Criticism of the pattern evidence specialties, handwriting comparison, fingerprint comparison, toolmark and firearms identification, and evidence such as tire or footwear impressions and bite marks, has been especially pointed in this regard. There was scant discussion of chemistry-based specialties, such as controlled substance identification, toxicology, and forensic chemistry. Concerns were also raised about measurement of error rates, transparency concerning the accuracy and precision of tests, and validation of procedures.

In the 1980s, the FBI, in collaboration with the forensic labs doing DNA testing at the time, had assembled a peer standard-setting body called the Technical Working Group on DNA Analysis Methods (TWGDAM). The goals of TWGDAM included standardizing and validating methods and choosing consensus protocols that would enable smooth operation of the DNA database (called CODIS, for *Co*mbine *D*NA *I*ndexing *S*ystem). Most people would agree that this TWG worked well and realized its objectives. For a time, the TWGDAM was supplanted by a committee known as the DNA Advisory Board (DAB), established by the DNA Identification Act of 1994 (P.L. 103-322). This Board was disbanded by law about 5 years later, and the Technical Working Group on DNA came back into existence, now as the Scientific Working Group on DNA Analysis Methods (SWGDAM), to establish and monitor consensus standards in DNA. Building on the TWGDAM model, the National Institute of Justice assisted in forming a number of other TWGs, and the FBI formed additional SWGs in this period. The purpose was to establish consensus standards in various forensic-science disciplines. There are now quite a number of these committees.

The exoneration of over 250 convicted felons in 34 states prompted by investigations carried out by the Innocence Project through DNA analysis (Innocence Project visited 9/12/11) shed some harsh light on practices in the criminal justice system that had led to the original convictions. In some cases, these included poor or unacceptable forensic lab practices (Garrett 2008; Collins and Jarvis 2008). Note that the extent to which unacceptable forensic practices contributed to the wrongful convictions in these cases has been challenged (Collins and Jarvis 2008). Reports of incompetent scientists and even fraud by people working in forensic laboratories and on forensic cases, and of almost complete meltdowns in a few laboratories, also came to the fore (www.corpus-delicti.com/forensic_fraud.html, 2011).

In this climate, there was also concern about the limited funding available for forensic-science services, the limited funds available for forensic-science-related research, and how forensic-science academic programs and education fit into the overall picture. A census of the public forensic science laboratories in 2002 (Peterson and Hickman 2005) indicated that the labs were encountering serious backlogs due to increased workloads and decreased resources. Looking at the calendar year 2002, forensic labs had about 289,900 cases in backlog as of January 1, received nearly 2,707,000 new cases during the year, completed over 2,495,000 of them, and ended the year with a backlog of over 501,000 cases. A follow-up study using almost identical methodology for 2009 (Durose et al. 2012) showed that labs started the year with over 1,131,000 cases backlogged and ended it with almost 1,154,000 despite having worked 3,632,000 cases.

All these forces and developments persuaded Congress that forensic science needed some review and attention from the federal government, and thus, in late 2005, the US Congress authorized the National Academy of Sciences to conduct the inquiry that led to the NRC report. The report was published in 2009.

The NRC Report

In charging the National Academy to conduct its inquiry, Congress stated that

> While a great deal of analysis exists of the requirements in the discipline of DNA, there exists little to no analysis of the remaining needs of the

> community outside of the area of DNA. Therefore... the Committee directs the Attorney General to provide [funds] to the National Academy of Sciences to create an independent Forensic Science Committee. This Committee shall include members of the forensics community representing operational crime laboratories, medical examiners, and coroners; legal experts; and other scientists as determined appropriate.

This Forensic Science Committee was to:

1. Assess the present and future resource needs of the forensic science community, to include State and local crime labs, medical examiners, and coroners;
2. Make recommendations for maximizing the use of forensic technologies and techniques to solve crimes, investigate deaths, and protect the public;
3. Identify potential scientific advances that may assist law enforcement in using forensic technologies and techniques to protect the public;
4. Make recommendations for programs that will increase the number of qualified forensic scientists and medical examiners available to work in public crime laboratories;
5. Disseminate best practices and guidelines concerning the collection and analysis of forensic evidence to help ensure quality and consistency in the use of forensic technologies and techniques to solve crimes, investigate deaths, and protect the public;
6. Examine the role of the forensic community in the homeland security mission;
7. [Examine] interoperability of Automated Fingerprint Information Systems (AFIS) and
8. Examine additional issues pertaining to forensic science as determined by the Committee

There were 17 members of the Committee. Two were forensic pathologists, four were academic or practitioner representatives of forensic science, three were attorneys (one co-chairperson was a federal judge), and the remainder were science and engineering professors. They held a number of meetings and received both written and verbal comments from a wide variety of interested parties and representatives of interested organizations. The report, issued in February of 2009, has 11 chapters. The first four concern defining forensic science, making a case for the integration of different services, the legal environment around admissibility of forensic-science evidence, and a discussion of scientific method and data interpretation. Next, some of the forensic-science subdisciplines are described. For purposes of this inquiry and report, forensic science was construed to include toxicology, firearms and toolmarks, questioned documents, trace evidence, controlled substances, biological and serology screening (including DNA analysis), fire debris and arson analysis, impression evidence, blood pattern analysis, crime scene investigation, medicolegal death investigation, and digital evidence, following NIJ 2006.

The report is summarized and discussed below under headings corresponding to the chapter subjects. Some of these chapters contained recommendations, while others did not. The major recommendations are stated in their entirety and are integrated into the chapter headings in which they occur in the original report.

The Forensic-Science Community and the Need for Integrated Governance

One of the problems identified is the breadth of scientific, technical, and practice specialties that fall under “forensic science” as broadly construed. From molecular biology, chemistry, and physics, to medicine, nursing, pharmacology, and toxicology, to fingerprints and firearms, to crime scene investigation, the umbrella covers a variety of practitioners from Ph.D.-level scientists to police personnel trained in crime scene processing and investigation (see ▶ Forensic Science Culture). Any proposal for uniformity covering such a wide spectrum of areas and interests presents a challenge, to say the least.

Public-sector lab and medicolegal services are also delivered at different levels of government, from federal to municipal, with correspondingly different resources. Some state laboratories primarily serve local law enforcement agencies, because that is where the majority of criminal cases are investigated. There are private, for-profit laboratories in the USA. It is fair to say that more private testing labs have grown up for DNA analysis because the need was there. A number of public laboratories have contracted with private labs to conduct the DNA profiling for data banking; to work old, cold, and backlogged cases; and in some cases even to work current cases as a way of catching up. In addition, there are many "identification" units or "fingerprint" units in law enforcement agencies that are not part of larger forensic labs.

Lack of adequate funding is a theme running throughout the forensic lab and medicolegal services community. This has led to chronic backlogs in the laboratories as noted above. The National Institute of Justice (NIJ), which funds virtually all forensic-science research, has managed the infusion of millions of dollars to the public-sector labs and medical examiner offices over the past several years to improve their DNA programs, to help investigate and hopefully solve cold cases, and for general service improvements. The Coverdell program is the main vehicle for laboratory improvement funding, part of which is competitive. The program has helped materially, but has not solved the underlying resources problem. Much more research is needed in forensic science than is able to be funded with NIJ's current resources. NIJ's budget is minimal if compared with that of the National Institutes of Health (NIH) or the National Science Foundation (NSF), and it is generally able to award 20 or 30 grants a year for forensic-science research, not including Coverdell or President's DNA Initiative funds. In fiscal year 2011 (the year from September 1, 2010, to August 31, 2011), for example, NIJ awarded about 385 grants and cooperative agreements totaling over $200 million. Only a small part of these funds was given for forensic-science research as such. NIJ also manages the activities of several funded centers through its Forensic Resource Network. The research funds administered by the Office of Science and Technology are competitive, and expert panels review the applications. NIJ recently announced it will develop review panels whose members serve multi-year terms, similar to NIH's study sections, instead of constituting new panels for each review cycle. It is to be noted that the FBI maintains an active research and development section of its Laboratory Division that involves in-house scientists and collaborations with university and other government agency scientists. The National Institute of Standards and Technology (NIST) is also involved in some forensic-science activities that fall under its overall mission, such as devising standard reference materials. To try and address both coordination and oversight of this myriad of activities, the Committee recommended that a National Institute of Forensic Science be created.

Recommendation 1. *To promote the development of forensic science into a mature field of multidisciplinary research and practice, founded on the systematic collection and analysis of relevant data, Congress should establish and appropriate funds for an independent federal entity, the National Institute of Forensic Science (NIFS). NIFS should have a full-time administrator and an advisory board with expertise in research and education, the forensic science disciplines, physical and life sciences, forensic pathology, engineering, information technology, measurements and standards, testing and evaluation, law, national security, and public policy. NIFS should focus on:*

(a) Establishing and enforcing best practices for forensic science professionals and laboratories;

(b) Establishing standards for the mandatory accreditation of forensic science laboratories and the mandatory certification of forensic scientists and medical examiners/forensic pathologists—and identifying the entity/entities that will develop and implement accreditation and certification;

(c) Promoting scholarly, competitive peer-reviewed research and technical development in the forensic science disciplines and forensic medicine;

(d) Developing a strategy to improve forensic science research and educational programs, including forensic pathology;

(e) Establishing a strategy, based on accurate data on the forensic science community, for the efficient allocation of available funds to give strong support to forensic methodologies and practices in addition to DNA analysis;
(f) Funding state and local forensic science agencies, independent research projects, and educational programs as recommended in this report, with conditions that aim to advance the credibility and reliability of the forensic science disciplines;
(g) Overseeing education standards and the accreditation of forensic science programs in colleges and universities;
(h) Developing programs to improve understanding of the forensic science disciplines and their limitations within legal systems; and
(i) Assessing the development and introduction of new technologies in forensic investigations, including a comparison of new technologies with former ones.

This recommendation sits at the core of the report, and many of the other recommendations are conditioned upon its implementation. This NIFS would be a new breed of federal agency, having research and educational support as well as oversight and enforcement powers and responsibilities. The Committee elaborated what it believed to be minimal criteria for a federal coordinating agency. They noted they had considered NSF and NIST, but did not think either met the criteria. NIJ was rejected because it is part of the US Dept. of Justice whose primary mission is law enforcement and prosecution. The Committee recognized that budgetary and other constraints could delay or even prevent the implementation of this core recommendation. NIJ argued in its response to the report that it was structured to handle the tasks described and that many of its initiatives followed the recommendations (NIJ 2011).

There have been two legislative initiatives in the US Congress that directly responded to this recommendation (THOMAS – Library of Congress). One, a bill in the US Senate (S132 introduced 1/25/11 by Sen. Patrick Leahy of Vermont), the Criminal Justice and Forensic Science Reform Act, contains provisions to accomplish many of the recommendations in the NRC report, though not using the same language and terminology. Senate Bill S3378 (and an identical bill in the House, HR 6106) set out to accomplish many of the same goals but differ from S132 in the administrative structure. Sen. Leahy's bill would create an Office of Forensic Science within the US Attorney General's Office and an appointed Board to oversee all the proposed activities. Sen. Rockefeller's bill would create a National Forensic Science Coordinating Office at NSF. There would be substantial involvement from NIST in the creation of standards. Neither of these bills has progressed out of committee, and there does not appear to be much chance that either of them will be enacted.

The Admission of Forensic-Science Evidence in Litigation

The report has a lengthy, detailed, and informative review of the transition from *Frye* to *Daubert* criteria for admissibility and the evolution of Federal Rule 702 (which many state courts follow, along with the federal courts). This subject was sufficiently discussed above for purposes of this summary. A number of decisions on the admissibility of different categories of evidence are discussed. An important point recognized by the Committee is that courts and admissibility rulings are not the appropriate mechanism for establishing solid scientific underpinnings for the forensic sciences; that must be done by carefully controlled research.

The admissibility discussion leads naturally to the next chapter of the report, concerning scientific principles and data interpretation.

The Principles of Science and Interpreting Scientific Data

This entry discusses scientific method and the principles of method development and validation that characterize the traditional sciences. In the view of the Committee, these must be followed by forensic sciences. There is a review of the scientific method and of the subsequent validation of scientific findings and results. Emphasis is placed on the validation of methods and of the methods used by forensic sciences to classify/identify substances, to associate/dissociate evidentiary items with potential sources, and sometimes to attempt to reconstruct aspects of past events based on the physical evidence record. Method validation should be, in the view of the committee, published in peer-reviewed literature.

It is suggested that the principles of ISO 17025 for testing laboratories be used as the basis for forensic method validation. These principles include calibration of reference standards and materials, comparison of results on one method with another, interlaboratory comparisons, systematic assessment of factors that influence results, and assessment of uncertainty. Many forensic labs are now using ISO 17025 principles as the basis for accreditation. The QA (quality assurance) guidelines developed for DNA analysis by TWGDAM/SWGDAM (discussed above) are enumerated as a model. There are now many TWGs/SWGs for different categories of evidence that have established consensus principles for the analysis and interpretation of the particular type of evidence (e.g., SWGDRUG, drug and controlled substance identification and quantitation analysis; SWGFAST, fingerprints and other friction ridge patterns; SWGGUN, toolmark and firearms identification; SWGMAT, materials such as fibers).

It is noted that forensic labs should be reporting measurement error with results, and keeping track of error rates to the extent possible. Techniques and comparisons need to be subjected to blind testing to ferret out the rate of false positives and false negatives.

Bias and sources of bias, particularly as applicable to comparisons where "source/not source" conclusions are reached, are discussed. Context bias and observer bias have been discussed more recently in the context of forensic-science examinations (see ▶ Cognitive Forensics: Human Cognition, Contextual Information, and Bias, Saks et al. 2003; Crime Lab 2009b; Thornton 2010). The larger problem outlined by those critical of forensic sciences contains the "law-enforcement context bias" as well. It has to do with the effects of unconscious influences of bias on an observer seeing, recording, and interpreting data.

Descriptions of Some Forensic-Science Disciplines

This entry discusses the forensic analysis of several categories of evidence, including biological, controlled substances, friction ridge pattern analysis—usually fingerprints, other patterns such as footwear, toolmarks and firearms, hairs, fibers, questioned documents, paint, explosives residues, fire debris, odontology, and bloodstain patterns—and digital media analysis. Each one is described with respect to its sample data and collection, the methods of analysis employed, the scientific interpretation and reporting of results, and the committee's conclusions.

There is broad variety across the disciplines regarding techniques, methods, reliability, error rate assessments, reporting language and guidelines, underlying research, and the educational background of practitioners.

Improving Methods, Practice, and Performance in Forensic Science

This entry focuses on improvements in practice across disciplines, with a view toward bringing more scientific approaches and methods to bear, particularly but not exclusively, in the pattern evidence comparison areas (see ▶ Forensic Science and the Paradigm of Quality).

Recommendation 2. *The National Institute of Forensic Science (NIFS), after reviewing established standards such as ISO 17025, and in consultation with its advisory board, should establish standard terminology to be used in reporting on and testifying about the results of forensic science investigations. Similarly, it should establish model laboratory reports for different forensic science* *disciplines and specify the minimum information that should be included. As part of the accreditation and certification processes, laboratories and forensic scientists should be required to utilize model laboratory reports when summarizing the results of their analyses.*

Although it may at first sound trivial, terminology is a very critical issue in forensic sciences, especially the words and phrases forensic practitioners use as "terms of art." Criminalists, for example, use the word "identification" to mean placing something into its proper class or category, such as "the stain represented as item # 6 was identified as human blood" or "item # 9 was identified as cat hair." Fingerprint and firearms identification specialists, on the other hand, use "identification" to mean "individualization," such as "the developed latent print represented as item # 8 was identified as the left middle fingerprint of John Doe" or "the bullet represented as item # 4 was fired by the 9 mm semiautomatic pistol, item #2." We also talk about "identifying the human remains as that of Jane Doe." In cases where there are comparisons made between a known specimen and an evidentiary one, the terminology issue can become even more confusing. Suppose a 13-locus DNA profile of the male fraction from a vaginal swab is identical in every respect to the profile of the suspect. One could write or testify that "the profiles match," or that "the suspect could not be excluded as the source of the semen," or that "the suspect and evidence profiles are alike in every respect that was tested and compared." In a DNA comparison like this, these statements would be accompanied by an estimate of the probability of chance duplication. One could construct similar examples from paint comparisons, hair or fiber comparisons, etc. The issue here is whether these statements mean the same thing to a trier of fact. They probably do not. It was thus recommended that a standardized terminology be adopted and used in laboratory reports and in testimony, so that given the same results, the wording is always the same. It is understood that there might have to be a certain amount of within-discipline terminology standardization, that is, trace analysts might not be able to use exactly the same terminology and vocabulary as a fingerprint or firearms examiner.

The Committee noted that even if some or all of its core recommendations were ignored or delayed because of funding or political considerations, the terminology issue was one that should receive immediate attention. Another one was conformity of lab reports with external standards such as ISO 17025 (2005), where methods, materials, procedures, results, and conclusions are recorded, along with sources and measurements of uncertainty and levels of confidence where appropriate.

Recommendation 3. *Research is needed to address issues of accuracy, reliability, and validity in the forensic science disciplines. The National Institute of Forensic Science (NIFS) should competitively fund peer-reviewed research in the following areas*:

(a) Studies establishing the scientific bases demonstrating the validity of forensic methods.

(b) The development and establishment of quantifiable measures of the reliability and accuracy of forensic analyses. Studies of the reliability and accuracy of forensic techniques should reflect actual practice on realistic case scenarios, averaged across a representative sample of forensic scientists and laboratories. Studies also should establish the limits of reliability and accuracy that analytic methods can be expected to achieve as the conditions of forensic evidence vary. The research by which measures of reliability and accuracy are determined should be peer reviewed and published in respected scientific journals.

(c) The development of quantifiable measures of uncertainty in the conclusions of forensic analyses.

(d) Automated techniques capable of enhancing forensic technologies.

The wide range of different disciplines comprising "forensic science" comes with an equally wide range of bases for accuracy, reliability, requirements for validation of a technique, etc.

The recommendation calls for research to quantitate reliability measures. It also calls for examiners to provide quantitative statements of measurement uncertainty where possible. This recommendation also implies that the research should try to get at error rates for different techniques and procedures. It is designed primarily to abstract best practices from the research data.

The committee were especially concerned about disciplines like fingerprint comparison, where there have been claims of "100 % accuracy" or "zero error rate" (see ► Probability and Inference in Forensic Science). The Brandon Mayfield case discussed just below was a stark reminder of the possibility of misidentifications, even by highly experienced examiners. There has been lengthy discussion by lawyers, law professors, and fingerprint examiners around these questions (Koehler 2008; Haber and Haber 2008; Mnookin 2008; Dror et al. 2006; Dror and Charlton 2006).

Recommendation 4. *To improve the scientific bases of forensic science examinations and to maximize independence from or autonomy within the law enforcement community, Congress should authorize and appropriate incentive funds to the National Institute of Forensic Science (NIFS) for allocation to state and local jurisdictions for the purpose of removing all public forensic laboratories and facilities from the administrative control of law enforcement agencies or prosecutors' offices.*

Bias on the part of forensic examiners as an intrinsic result of being administratively located within law enforcement agencies or prosecutors' offices has been discussed for years. The problem may be made even more acute if the examiners are themselves agents of law enforcement (see ► Cognitive Forensics: Human Cognition, Contextual Information, and Bias). The poster-child case for this viewpoint is Brandon Mayfield (U.S. DOJ Inspector General 2006). The details are complicated, but briefly, FBI examiners mistakenly identified latent fingerprints from the Madrid, Spain, train bombing incident of March 11, 2004, as those of Brandon Mayfield, a Muslim US citizen living in Oregon. Critics point to this case as a clear example of context bias on the part of the FBI analysts. But critics also argue that this sort of bias can be more subtle, that analysts working in a law enforcement environment are bound to be influenced toward that viewpoint, and that objectivity is not possible. Many forensic scientists and lab administrators do not agree, and there are published discussions on both sides of the question.

Recommendation 5. *The National Institute of Forensic Science (NIFS) should encourage research programs on human observer bias and sources of human error in forensic examinations. Such programs might include studies to determine the effects of contextual bias in forensic practice (e.g., studies to determine whether and to what extent the results of forensic analyses are influenced by knowledge regarding the background of the suspect and the investigator's theory of the case). In addition, research on sources of human error should be closely linked with research conducted to quantify and characterize the amount of error. Based on the results of these studies, and in consultation with its advisory board, NIFS should develop standard operating procedures (that will lay the foundation for model protocols) to minimize, to the greatest extent reasonably possible, potential bias and sources of human error in forensic practice. These standard operating procedures should apply to all forensic analyses that may be used in litigation.*

Strengthening Oversight of Forensic-Science Practice

The four recommendations growing out of the "oversight" have to do with quality programs, accreditation, certification, proficiency testing, code of ethics, and further research into the quantitation of measurement variability and error.

Recommendation 6. *Congress should authorize and appropriate funds to NIFS to work with the National Institute of Standards and Technology (NIST), in conjunction with government laboratories, universities, and private laboratories, and in consultation with*

Scientific Working Groups, to develop tools for advancing measurement, validation, reliability, information sharing, and proficiency testing in forensic science and to establish protocols for forensic examinations, methods, and practices. Standards should reflect best practices and serve as accreditation tools for laboratories and as guides for the education, training, and certification of professionals. Upon completion of its work, NIST and its partners should report findings and recommendations to NIFS for further dissemination and implementation.

Recommendation 7. *Laboratory accreditation and individual certification of forensic science professionals should be mandatory, and all forensic science professionals should have access to a certification process. In determining appropriate standards for accreditation and certification, the National Institute of Forensic Science (NIFS) should take into account established and recognized international standards, such as those published by the International Organization for Standardization (ISO). No person (public or private) should be allowed to practice in a forensic science discipline or testify as a forensic science professional without certification. Certification requirements should include, at a minimum, written examinations, supervised practice, proficiency testing, continuing education, recertification procedures, adherence to a code of ethics, and effective disciplinary procedures. All laboratories and facilities (public or private) should be accredited, and all forensic science professionals should be certified, when eligible, within a time period established by NIFS.*

Recommendation 8. *Forensic laboratories should establish routine quality assurance and quality control procedures to ensure the accuracy of forensic analyses and the work of forensic practitioners. Quality control procedures should be designed to identify mistakes, fraud, and bias; confirm the continued validity and reliability of standard operating procedures and protocols; ensure that best practices are being followed; and correct procedures and protocols that are found to need improvement.*

Recommendation 9. *The National Institute of Forensic Science (NIFS), in consultation with its advisory board, should establish a national code of ethics for all forensic science disciplines and encourage individual societies to incorporate this national code as part of their professional code of ethics. Additionally, NIFS should explore mechanisms of enforcement for those forensic scientists who commit serious ethical violations. Such a code could be enforced through a certification process for forensic scientists.*

Education and Training in Forensic Science

The committee noted that major strides had been made recently in bringing some standardization to forensic-science academic programs through the Forensic Science Education Program Accreditation Commission (FEPAC). The FEPAC accreditation program and guidelines grew out of a NIJ-supported Technical Working Group for Education and Training in Forensic Science (NIJ 2004). There are still many programs offered, however, that are not FEPAC accredited (FEPAC accreditation is voluntary), and they vary in their uniformity and science content. For graduate programs in particular, more funding should be available if there is to be a proper research base developed.

Another issue is forensic-science education for attorneys and judges. More interactions and information exchange are encouraged. Law schools are in a position to enhance this interaction. In disciplines that require training as well as education, a firm foundation in scientific principles is recommended for both the educational and the training components (see ▶ Forensic Science Culture).

Recommendation 10. *To attract students in the physical and life sciences to pursue graduate studies in multidisciplinary fields critical to forensic science practice, Congress should authorize and appropriate funds to the*

National Institute of Forensic Science (NIFS) to work with appropriate organizations and educational institutions to improve and develop graduate education programs designed to cut across organizational, programmatic, and disciplinary boundaries. To make these programs appealing to potential students, they must include attractive scholarship and fellowship offerings. Emphasis should be placed on developing and improving research methods and methodologies applicable to forensic science practice and on funding research programs to attract research universities and students in fields relevant to forensic science. NIFS should also support law school administrators and judicial education organizations in establishing continuing legal education programs for law students, practitioners, and judges.

Medical Examiner and Coroner Systems: Current and Future Needs

Various problems exist in the USA with respect to death investigation, and this entry addressed them. There are two systems: medical examiner and coroner. Medical examiners are forensic pathologists, but there are few medical examiner jurisdictions compared with coroner jurisdictions. Coroners generally have little or no formal medical or forensic training. This situation creates a patchwork quilt, and there is considerable lack of uniformity across jurisdictions. The situation also inhibits the collection of accurate death statistics for deaths that come under forensic investigation.

There is now a Scientific Working Group for Medicolegal Death Investigation (SWGMDI), and it has already issued several reports. A recent draft report shows that there are not nearly enough forensic pathologists in the USA to meet current needs.

Recommendation 11. *To improve medico-legal death investigation*:

(a) Congress should authorize and appropriate incentive funds to the National Institute of Forensic Science (NIFS) for allocation to states and jurisdictions to establish medical examiner systems, with the goal of replacing and eventually eliminating existing coroner systems. Funds are needed to build regional medical examiner offices, secure necessary equipment, improve administration, and ensure the education, training, and staffing of medical examiner offices. Funding could also be used to help current medical examiner systems modernize their facilities to meet current Centers for Disease Control and Prevention-recommended autopsy safety requirements.

(b) Congress should appropriate resources to the National Institutes of Health (NIH) and NIFS, jointly, to support research, education, and training in forensic pathology. NIH, with NIFS participation, or NIFS in collaboration with content experts, should establish a study section to establish goals, to review and evaluate proposals in these areas, and to allocate funding for collaborative research to be conducted by medical examiner offices and medical universities. In addition, funding, in the form of medical student loan forgiveness and/or fellowship support, should be made available to pathology residents who choose forensic pathology as their specialty.

(c) NIFS, in collaboration with NIH, the National Association of Medical Examiners, the American Board of Medico-legal Death Investigators, and other appropriate professional organizations, should establish a Scientific Working Group (SWG) for forensic pathology and medico-legal death investigation. The SWG should develop and promote standards for best practices, administration, staffing, education, training, and continuing education for competent death scene investigation and postmortem examinations. Best practices should include the utilization of new technologies such as laboratory testing for the molecular basis of diseases and the implementation of specialized imaging techniques.

(d) All medical examiner offices should be accredited pursuant to NIFS endorsed standards within a timeframe to be established by NIFS.

(e) All federal funding should be restricted to accredited offices that meet NIFS-endorsed standards or that demonstrate significant and measurable progress in achieving accreditation within prescribed deadlines.

(f) All medico-legal autopsies should be performed or supervised by a board certified forensic pathologist. This requirement should take effect within a timeframe to be established by NIFS, following consultation with governing state institutions.

Automated Fingerprint Identification Systems

The primary issue under this heading is automated fingerprint identification system (AFIS) interoperability. There are several commercial AFIS system vendors, and the software and search algorithms are proprietary. Different jurisdictions have different systems, making it difficult to have a fully nationally integrated database and search system.

Recommendation 12. *Congress should authorize and appropriate funds for the National Institute of Forensic Science (NIFS) to launch a new broad-based effort to achieve nationwide fingerprint data interoperability. To that end, NIFS should convene a task force comprising relevant experts from the National Institute of Standards and Technology and the major law enforcement agencies (including representatives from the local, state, federal, and, perhaps, international levels) and industry, as appropriate, to develop*:

(a) Standards for representing and communicating image and minutiae data among Automated Fingerprint Identification Systems. Common data standards would facilitate the sharing of fingerprint data among law enforcement agencies at the local, state, federal, and even international levels, which could result in more solved crimes, fewer wrongful identifications, and greater efficiency with respect to fingerprint searches; and

(b) Baseline standards—to be used with computer algorithms—to map, record, and recognize features in fingerprint images, and a research agenda for the continued improvement, refinement, and characterization of the accuracy of these algorithms (including quantification of error rates).

Homeland Security and the Forensic-Science Disciplines

The Committee briefly explored potential roles of forensic sciences in homeland security, as that was part of its charge (see ▶ Forensic Science and Criminal Investigation and ▶ Forensic Science Culture). In doing so, it noted that these suggestions were preliminary and could comprise a separate study.

Forensic-science involvement in homeland security situations differs from its involvement in criminal cases. There needs to be federal agency—local laboratory coordination, earlier involvement in the investigations to help direct them, and specialized technical expertise having to do with chemical, biological, or nuclear threats. Digital forensic sciences are expected to have a major role in antiterrorism efforts.

Recommendation 13. *Congress should provide funding to the National Institute of Forensic Science (NIFS) to prepare, in conjunction with the Centers for Disease Control and Prevention and the Federal Bureau of Investigation, forensic scientists and crime scene investigators for their potential roles in managing and analyzing evidence from events that affect homeland security, so that maximum evidentiary value is preserved from these unusual circumstances and the safety of these personnel is guarded. This preparation also should include planning and preparedness (to include exercises) for the interoperability of local forensic personnel with federal counterterrorism organizations.*

Discussion

The Committee's report and recommendations have gotten a mixed reception in the forensic science and legal worlds. Some organizations have issued official or public responses or reactions to the report (ASCLD 2009; SWGFAST 2009; AAFS 2009; SWGDE 2009; White House OSTP 2010; Suresh 2012). So have various commentators and groups, some from forensic science and others from the legal profession (Crime Lab Report 2009a; Budowle et al. 2009; Kaye 2010; Chatman 2009; Koehler et al. 2011).

Some of the organizations generally supported the principles on which the NRC recommendations were based and stated that they, in effect, stood ready to be part of the forward progress toward better forensic science. Most supported the call for more resources. Some objected to certain recommendations and criticisms. SWGFAST, representing one of the disciplines that was most criticized in the report, disagreed with many of the conclusions about the insufficiency of a scientific basis for friction ridge comparisons. Some disagreed with the recommendation of separating forensic laboratories from law enforcement administrations, suggesting that the apparent intrinsic bias it represented was not in fact necessarily present and that the relationship did not automatically cause bias.

Not everyone agreed with the recommendation to form the NIFS. Some thought the necessary resources could be assembled from and within existing agencies that already have involvement in forensic-science research and support. NIJ, which was explicitly excluded by the NRC report as an appropriate setting for NIFS, responded with a lengthy justification for why it was actually the most appropriate place. NSF has stated, through the testimony of its director on the Rockefeller bill, that it stands ready to act as the primary federal agency for forensic science if the bill should be enacted.

The harshest critics of forensic science from academic settings and law schools take the report as a vindication of their long-held beliefs about and criticisms of the sorry state of forensic science.

Some have seen the report as a blueprint for progress and positive change, even if all the recommendations cannot be implemented (especially see Kaye 2010; Koehler et al. 2011).

Given the current state of the US economy and the concern with federal spending, it does not seem at all likely that the NIFS will come to exist. The failure of two congressional bills aimed at the NRC Committee's primary goals to emerge from Committee is an indication that big, new, comprehensive legislation focused on improving forensic science is not likely. The majority of forensic-science laboratories are operated by local or state government. It should also be understood that there are significant limitations on what the federal government can do to regulate their activities. Another unfortunate consequence of the economic recession is that state and local governments find themselves under extreme pressure to streamline services and reduce expenses. This situation cannot be seen as favorable for any increase in most forensic-science lab resources and could even mean that resources will decline.

Notwithstanding the different points of view and parochial interests of various parties and organizations, the report must be seen as a milestone in the history of forensic sciences in the USA. Never before has there been such a large, comprehensive look at the forensic sciences by anyone, and this one was done by the prestigious NRC. There is validation of and support for the TWG/SWG model, and those organizations will continue to make progress. There may be increased federal funding for forensic-science research and for operational improvements. Some of the recommendations will be implemented incrementally in and by the profession, even without a centrally focused federal forensic-science entity. The report will likely stand as the primary reference document for any discussion of forensic science laboratory improvement or funding for years to come.

Related Entries

- ▶ DNA Profiling
- ▶ DNA Technology and Police Investigations
- ▶ Drug Enforcement
- ▶ Expert Witnesses: Role, Ethics, and Accountability
- ▶ Fingerprint Identification
- ▶ Forensic Anthropology
- ▶ Forensic Environmental Evidence
- ▶ Forensic Facial Analysis
- ▶ Forensic Linguistics
- ▶ Forensic Palynology
- ▶ Forensic Science
- ▶ Forensic Science and Criminal Investigation
- ▶ Forensic Science and Miscarriages of Justice
- ▶ Forensic Science and the Paradigm of Quality
- ▶ Forensic Science Culture
- ▶ Forensic Science Effectiveness
- ▶ Legal Rules, Forensic Science and Wrongful Convictions

Recommended Reading and References

AAFS [American Academy of Forensic Sciences] (2009) Position statement in response to the NAS "forensic needs" report, 4 Sept 2009

ASCLD [American Society of Crime Laboratory Directors] (2009) Comments on the release of the NAS report on forensic science, 19 Feb. 2009

Budowle B, Bottrell MC, Bunch SG et al (2009) A perspective on errors, bias, and interpretation in the forensic sciences and direction for continuing advancement. J Forensic Sci 54(4):798–809

Chatman Q (2009) How scientific is forensic science? Champion, Aug 2009

Collins J, Jarvis J (2008) The wrongful conviction of forensic science. Crime Lab Report, 16 July 2008

Crime Lab Report (2009a) A response to the testimony of Judge Harry T. Edwards before the Senate Judiciary Committee April 1, 2009 by Crime Lab Report. [Crime Lab Report is an independently edited and produced report on crime laboratory and forensic science issues; its views often reflect those of the forensic laboratory community]

Crime Lab Report (2009b) Post-conviction activists 'contaminate' evidence in Texas [discussing context-bias on the part of fire investigators in a high-profile death-penalty case in Texas called "Willingham" in which the defendant was ultimately executed]

Daubert et ux. etc. et al. v. Merrell Dow Pharmaceuticals, Inc., No. 92-102 (1993), 509 U.S. 579, 113 S.Ct. 2786 (often referred to as "Daubert" or "Daubert vs Merrell Dow")

Dror IE, Charlton D (2006) Why experts make errors. J Forensic Ident 56(4):600

Dror IE, Charlton D, Péron AE (2006) Contextual information renders experts vulnerable to making erroneous identifications. Forensic Sci Int 156:74

Durose MR, Walsh KA, Burch AM (2012) Census of publicly funded crime laboratories 2009, U.S. Department of Justice, Bureau of Justice Statistics, 2012 August, NCJ 238252

Frye v. United States (1923) Frye v. United States, 54 App. D.C. 46, 293 F. 1013, No 3968, Court of Appeals of District of Columbia (often referred to simply as 'Frye")

Garrett BL (2008) Judging innocence. 108 Columbia L.R. 55

General Electric Co et al. vs Robert K. Joiner et ux., No. 96-188 (1997) 522 U.S. 136, 118 S. Ct. 512 (often referred to as "Joiner")

Haber L, Haber RN (2008) Scientific validation of fingerprint evidence under Daubert. Law Probab Risk 7(2):87

http://www.corpus-delicti.com/forensic_fraud.html, visited 9/12/2011

Innocence Project: http://www.innocenceproject.org/ visited 9/12/11

ISO 17025 (2005) ISO [International Organization for Standardization]/ IEC [International Electrotechnical Commission], General requirements for the competence of testing and calibration laboratories, 2nd edn. ISO, Geneva

Kaye D (2010) The good, the bad, the ugly: the NAS report on strengthening forensic science in America. Sci Justice 50:8–11

Koehler JJ (2008) Fingerprint error rates and proficiency tests: what they are and why they matter. 59 Hastings Law J 59:1077

Koehler J, Mnookin JL, Cole SA, et al (2011) The Need for a research culture in the forensic sciences, UCLA LR 58:725 (and available from the Northwestern University School of Law Faculty Working Papers)

Kumho Tire Co. Ltd., et al. v. Patrick Carmichael, etc. et al., No. 97-1709 (1999) 526 U.S. 137; 119 S. Ct. 1167 (often referred to as "Kumho Tire"). The Daubert case along with the Joiner and Kumho Tire cases are sometimes referred to as the "Daubert trilogy"

Mnookin JL (2008) The validity of latent fingerprint identification: confessions of a fingerprinting moderate. Law Probab Risk 7(2):127

NIJ [National Institute of Justice] (2004) Education and training in forensic science: a guide for forensic science laboratories, educational institutions, and students, office of justice programs, NCJ 203099

NIJ [National Institute of Justice] (2006) Status and needs of forensic science service providers: a report to congress, NCJ 213420; this report followed a 180-Day Study Report on the Status and Needs of U.S Crime Laboratories by the American Society of Crime Lab Directors [ASCLD] in May 2004

NIJ [National Institute of Justice] (2011) The national institute of justice response to the report of the national research council: strengthening the national institute of justice, NCJ 234630

NRC [National Research Council] (1992) Committee on DNA technology in forensic science, DNA technology in forensic science. National Academies Press, Washington, DC

NRC [National Research Council] (1996) Committee on DNA in forensic science, the evaluation of forensic DNA evidence. National Academies Press, Washington, DC

NRC [National Research Council], National Academy of Sciences, Committee on Applied and Theoretical Statistics, Committee on Identifying the Needs of the Forensic Sciences Community (2009) Strengthening forensic science in the United States: a path forward. National Academies Press, Washington, DC, p 254

Peterson JL, Hickman MJ (2005) Census of publicly funded forensic crime laboratories, Bureau of Justice Statistics Bulletin, NCJ 207205

'Popper K (1989) Conjectures and refutations: the growth of scientific knowledge, 5th edn. Routledge, London, Karl Popper and Conjectures and Refutations in Wikipedia, retrieved 9/9/2011 from http://en.wikipedia.org/wiki/Karl_Popper and http://en.wikipedia.org/wiki/Conjectures_and_Refutations

Ross A (2011) Forensic science in Australia: can we learn from international reports? Austr J Forensic Sci 43(2–3):135–145

Saks MJ, Risinger DM, Rosenthal R, Thompson WC (2003) Context effects in forensic science: a review and application of the science of science to crime laboratory practice in the United States. Sci Justice 43(2):77–90

Suresh S (2012) Testimony before the U.S. senate commerce, science, and transportation committee on the science and standards of forensics, 28 March 2012

SWGDE [Scientific Working Group on Digital Evidence] Position on the National Research Council Report to Congress, 17 Sept. 2009

SWGFAST [Scientific Working Group on Friction Ridge Analysis, Study, and Technology] (2009) Position Statement on the NRC Report, posted 3 Aug. 2009

THOMAS—Library of Congress—http://thomas.loc.gov/home/thomas.php—visited 9/5/12, U.S. Congressional legislation can be queried and tracked from this site

Thornton JI (2010) A rejection of 'Working Blind' as a cure for contextual bias. J Forensic Sci 55(6):1163

U.S. DOJ [Department of Justice] Inspector General (2006) The Brandon Mayfield case, see at http://www.usdoj.gov/oig/special/s0601/PDF_list.htm; also see FBI Responds to the Office of Inspector General's Report on the fingerprint misidentification of Brandon Mayfield, National Press Release, 6 January 2006. More information and viewpoints: Crime Lab Report, Crime Labs under Police—Unresolved Issues, 15 October 2008; and S. Weinberg, Keystone Cops at the Police Lab, Miller-McCune.com, 18 June 2009

White House OSTP [Office of Science and Technology Programs] Response to NRC Report, October 2010

Cross-National

► Comparative Incarceration

Cross-National Crime Rates

► Institutional Anomie Theory

Cross-National Performance in Policing

Mark Brunger[1] and Graham Ellison[2]
[1]Department of Law and Criminal Justice, Canterbury Christ Church University, Canterbury, Kent, UK
[2]Institute of Criminology and Criminal Justice, School of Law, Queen's University Belfast, Belfast, UK

Overview

This entry considers some of the conceptual and methodological issues researchers might face when conducting cross-national performance comparisons of policing systems. In doing so, it asks whether any similarities can be identified across jurisdictions that would allow national police forces to be compared. Many jurisdictions now have some mechanism in place by which to assess police performance at the national level. Clearly, debates around performance, effectiveness, and accountability are important when modern police organizations command multimillion dollar budgets, but they are also important in terms of public legitimacy: how the police perform within the context of their mandate and the degree to which they inspire public trust and confidence. This entry suggests that considered in cross-national terms, there is considerable variation in the mechanisms used to assess and measure police performance. However, the trend towards internationalization and diversification in public policing has made the desire for cross-national analysis of policing systems more pressing. In the European Union (EU), for example, there have been attempts to both harmonize the Home Affairs policies of its 27 member states and standardize policing policies (particularly around drugs and antiterrorism). However, as we outline in the discussion, these efforts remain limited and are hindered by the differences among member states regarding how police organizations are structured, individual national political priorities, different legal systems, languages and political cultures, rivalries between national police organizations, and not to mention difficulties in compiling agreed performance criteria and targets that have a cross-national relevance. Quantitative cross-national crime data do exist, however, from which comparisons can be made, but such undertakings are also hampered by a multitude of problems, with reliability and validity issues particularly in abundance. These difficulties are also compounded by a relative lack of clarity about what exactly needs to be compared. The entry concludes by suggesting that while some correlations and similarities can be deduced, the existence of local cultural, political, and legal differences ensures that the act of comparison remains a difficult proposition.

Assessing the Key Data Sources

Academic Research

When considering cross-national analysis, the main question that arises is to whether any similarities and differences can be made measurable to any relevant extent. It is the case that some significant attempts have been made at cross-cultural comparisons and are, therefore, worthy of consideration. Bayley (1990), for one, has provided a comparative analysis of policing developments across different jurisdictions. Surprisingly, given their prominent role in maintaining social order across societies, he notes that the police have often been ignored in comparative analysis (Bayley 1990, p. 4). The main problem he identifies is that despite the fact that all policing systems have similar basic concepts, definitional elements, and organizational similarities, the actual practice of policing varies widely across jurisdictions, therefore, making any effort at cross-national comparison a difficult one (see Bayley 1990, Chapter 1). Nevertheless, there have been some attempts at measuring police performance internationally. For example, Dennis and Erdos (2005) in *Cultures and Crimes: Policing in Four Nations* compare policing methods in the UK, France, Germany, and the USA. They point out that what these nations have in common is that they have witnessed steep rises in crime and disorder from the late 1960s onwards. In order to make comparisons, they concentrate on measuring trends in easily comparable offense categories, such as robbery and burglary, for example. The authors come to the general conclusion that there is a clear link with between maintaining increases in police numbers and reducing crime. To this effect, they argue that continued rising crime rates in the UK can, in some part, be attributed to the lack of visibility of police "on the beat," which has made crime, in their terms, a low-risk activity. Furthermore, they argue that in France, Germany, and the USA, there has been a more "effective" attempt to deal with crime through a larger and sustained resourcing of the criminal justice system. However, like other analyses that make use of crime statistics to measure police (and policy makers) activity, it raises all sorts of problematic questions around the reliance on "officially recorded" crime data. The differential patterns of development in community and public/private policing practices and the diversity in legal systems have produced divergent and inconsistent ways of recording offenses and what offenses are actually deemed recordable in the first place.

Other studies have mapped the police use of force in a variety of international contexts (e.g., see Stenning 2003). Researchers assessed both the extent to which force, including deadly force, was used and the normative justifications for it. Again there are considerable variations in the use of force by police in the jurisdictions chosen for the study with established democracies much more likely to have a tighter regulatory framework in place, and within the EU, a statutory requirement to abide by supranational legal directives such as the European Convention on Human Rights. The researchers noted that given the variations in levels of political and economic development and the pressures facing the police in the jurisdictions studied, it was difficult to draw any definitive cross-national comparison about the circumstances under which force would be deployed and its character.

The *International Centre for the Prevention of Crime* (Lelandais 2007) has recently engaged in an attempt to compare police performance in eight jurisdictions: the USA (New York City & Chicago), the UK, Canada (province of Quebec), Chile, Belgium, France, Australia, and New Zealand. What the research demonstrates is that the instruments used to assess police performance vary wildly within and between jurisdictions: Some national governments use the same managerialist performance indicators that are applied across the public sector as a whole (e.g., the UK); some assess performance via existing financial or public administration legislation (e.g., Quebec, France), while others adapt varieties of Computer Statistics packages (or COMPSTAT) (e.g., the USA – New York, New Zealand, Australia). Chile and Belgium adopt aspirational guidelines and action plans, while within the USA there are significant differences with the city of Chicago, for example, preferring the *Alternative Policing Strategy (APS)* to COMPSTAT. Even the nomenclature of what is being assessed varies from "police excellence" in Belgium, "high performance" in New Zealand, "best value" in the UK, and "quality of service" in Chile (Lelandais 2007). In some jurisdictions "performance" is taken to mean effectiveness in dealing with crime; in others it relates to public satisfaction and confidence in the police (which may have nothing to do with crime) or to fiscal and budgetary probity and managerial dexterity. Even if it was accepted, based on this analysis, that it just *might* be possible to construct cross-national performance indicators, there remain significant problems insofar as it equates *policing* with the police. This position, which is discussed below, is becoming untenable given the increasing blurring of the boundary between highly variant modes of public and private policing provision.

There are therefore huge difficulties in conducting cross-national research into policing organizations which hinge on the differences that exist between police organizations nationally and *also* differences in the way that the police are organized within the *same* jurisdiction. While nations in the west and north may subscribe to a notion of democratic policing, this takes on a rather different form in those countries that have a British common-law heritage (the USA, the UK – excluding Scotland, Australia, Canada, Republic of Ireland, New Zealand) compared to a Napoleonic or Roman civil law tradition (Scotland, all of continental Europe and Latin America, the province of Quebec, and so on). Most notably, in the UK, for example, public policing is based on the doctrine of "public consent," while the "office of constable" which is rooted in common law ascribes all police officers irrespective of rank, a set of original, delegated powers which exist independently of their statutory authority or position on the bureaucratic hierarchy. Furthermore, even within the regions of the United Kingdom (Northern Ireland, Wales, Scotland, and England), there are considerable differences in how the police are organized. Scotland, for example, has a Roman law heritage, which impacts on the substantive nature of police powers and judicial procedure compared to

England, Wales, and Northern Ireland, which share a common-law legal tradition. In these jurisdictions, until relatively recently (1985 in England and Wales and 2000 in Northern Ireland), the police retained prosecutorial powers that were largely absent in the system of Procurator Fiscal in Scotland and even more so in the Napoleonic/Roman law tradition of independent institution between accuser and accused. However, other common-law jurisdictions (e.g., the USA and Canada), recognizing the anomaly in the British system, typically have had district attorneys (or their equivalent) to direct investigations for a considerably longer period of time. In addition, it is difficult to compare the style and organization of policing within the UK historically. Until the recent reforms associated with the peace process, Northern Ireland had a highly militarized form of public policing undertaken by the Royal Ulster Constabulary (RUC) that owed more to the colonial form adopted across the British Empire in the nineteenth century and the continental *gendarmerie* than to the archetypal "British Bobby."

The United States has a traditionally decentralized model of police organization with an amalgam of town, city, municipal, state, and federal forces. This system of tiering is also found in some European nations (Spain, Czech Republic, France, Italy, Germany, etc.) but not in others, for example, Norway, Sweden, the Republic of Ireland, and New Zealand who all maintain a single national police, albeit one that is demarcated into functional specialisms. Unlike many European police forces, the UK does not have a national police that is directly answerable to a central government department. Rather law enforcement is organized territorially in each legal system (England & Wales, Scotland, and Northern Ireland) with the police under the *guidance* of the relevant devolved administrations (Scotland and Northern Ireland) or the Home Office (England and Wales) but are operationally independent (in theory) from political direction.

In those European countries governed by a Napoleonic/Roman civil law tradition, the lines governing the responsibilities of the police and military authorities are also rather blurred with many countries (Austria, France, Portugal, Italy, Belgium, Switzerland, etc.) maintaining a *gendarmerie* that is administratively under the control of the civil authorities in some cases, but under military control in others. For example, *gendarmes* (individual officers) can be dispatched overseas as part of humanitarian and peacekeeping missions alongside the regular army, but in the majority of circumstances, they are used for regular policing duties. In some cases, the roles and responsibilities of the *gendarmerie* in relation to the civil police are unclear, and in Italy, for instance, you can report a crime either to the *Carabinieri* (gendarmerie), the *Polizia di Stato* (national civil police), or the *Vigili urbani* (municipal or local police). It is not unusual to find both the *Polizia di Stato* and the *Vigili urbani* turning up to the scene of a road traffic accident, and determining what organization will deal with the situation is often just a case of negotiation or more likely "who got there first." Furthermore, Italy also maintains a corps of the regular army that is dedicated exclusively to "police" duties – the *Guardia di Finanza* – that deals with financial crime as well as illegal immigration and border controls. Somewhat unusually, *Guardia di Finanza* officers have the right to demand and check customer receipts for purchases made in shops and restaurants. To complicate matters further, some European countries, such as Belgium, have disposed of their *gendarmerie* and moved towards decentralized and regionally autonomous police services, while others such as the Netherlands and Scotland (as a semiautonomous region of the UK) are seeking to nationalize and centralize their police operations.

There are also considerable variations in the degree of functional specialization found among national police organizations. In the UK, all police officers can find themselves performing crowd and riot control duties, but in France, this is the almost exclusive responsibility of a reserve unit of the *Police Nationale* – the *Compagnies Républicaines de Sécurité* (CRS). Again, however, the situation is muddied by the existence of a specialized unit of the *gendarmerie* (the *gendarmerie mobile*) that performs similar duties.

Even within Europe then, assessing cross-national police performance begs the following question: performance of what and of whom? There are simply too many organizations performing policing type functions and where the lines of authority and responsibility are unclear to make any kind of clear comparison. Fundamental differences are also evident when we start to go beyond the core European Union nations. Postcommunist policing systems within those countries that comprised the former Warsaw Pact have moved (or are moving) to less authoritarian and repressive dispensations, but even here there is considerable variations, with those states that have negotiated access to the EU being further along the developmental trajectory than those that have not. Compare the situation in the Czech Republic with that of Belarus, for example. Similarly, in Asia there is considerable differentiation in the organization and structure of the police, not to mention adherence to democratic norms and human rights standards. In South Korea and Japan policing is generally "rights based" and in accordance with democratic norms, but in other states, India, Sri Lanka, Pakistan, and Bangladesh, the situation is described by the *Commonwealth Human Rights Initiative* (2007) as "grim," with key concerns in and around human rights, accountability, and governance abound. Policing in Hong Kong represents a rather curious hybrid case, having structural similarities with both the British "colonial" and "metropolitan" models given the legacy of its history, but which also appears to be influenced by policing developments in mainland China (Lau 2004). Importantly, however, we know next to nothing about the structure and organization of policing in the People's Republic of China, which has received almost no attention in the comparative research literature (though see Tanner 2005). In the Middle East, policing structures reflect particular religious, cultural, and political sensibilities that make comparative analysis difficult, if not futile. In Islamic theocracies such as Iran and Saudi Arabia (and Afghanistan historically, though increasingly so again), the *Mutawwa'in* (religious police, though the religious police are referred to as *Basij-e Mostaz'afin* in Iran.) rigorously enforce Sharia law which applies to everyone irrespective of whether they are Muslim or not. Since there is no formal separation of church and state in Iran and Saudi Arabia, the *Mutawwa'in* have powers, duties, and responsibilities that rival, if not exceed, those of the regular police, not to mention a degree of influence that has simply no equivalent in Western democracies.

The argument being made here is that due to the huge variation in the organization of policing *between* and *within* nations, any kind of cross-national comparison is difficult if not nigh on impossible. This is further complicated by the variability of their core mandate (citizen safety, public order, or religious/cultural concerns) and huge disparities in adherence to democratic norms and international human rights standards. Academic research is further hampered by the rivalries between national police organizations or between the police and the gendarmerie that makes such comparative assessments difficult. One of us (Ellison) has had personal involvement with a reform endeavor involving the Turkish police which was seriously (though not fatally) hampered by the unwillingness of the *Jandarma* (gendarmerie) to participate in the study because of a legacy of hostility and mistrust of the Turkish National Police (TNP). This might not have mattered but for the fact that the *Jandarma* (gendarmerie) are the de facto police in much of eastern Turkey and perform all the tasks and functions that the TNP perform in the rest of the country. Therefore, the research was only able to glimpse a partial picture of the organization of public policing in Turkey and as such was limited in terms of the overall conclusions and recommendations that could be made.

Questions also arise in relation to the context of modern "policing" and what it should be taken to mean in the twenty-first century. Against a background of advancing economic liberalism, the European Union (EU), North American, and Australasian nations have shifted to forms of partnership policing, which has overseen a flattening out of policing across civil society to harness more private enterprise in its provision and delivery. Who or what "the police" can be taken to mean under these circumstances is becoming increasingly muddled. In the UK, for example, many police forces employ Police Community Support Officers (PCSOs) who are uniformed personnel, but not sworn officers, and who have a range of limited powers in relation to antisocial behavior, traffic, and so on. PCSOs have been described by the previous Commissioner of the London Metropolitan Police, Sir Ian Blair, as part of what he termed "the extended police family," but have been criticized by the police employees association, the Police Federation, for attempting to police on the cheap. The public meanwhile are unsure of what the role of PCSOs is supposed to be, since their legal powers are rather limited and their status somewhat vague. Indeed, in England and Wales the distinction between public and private is likely to be further distorted by the current government's reform agenda via the *Police Reform and Social Responsibility Bill*, which will likely result in a fundamental restructuring of the British police. Proposals are currently underway to outsource a number of key services to the private sector such as forensics, emergency call logging, crime information systems, as well as the widespread establishment of strategic partnerships with private firms being encouraged to engage in crime reduction initiatives. Already, in many UK towns and cities, parking and minor traffic enforcement duties are performed almost exclusively by private companies who are contracted by local authorities.

However, such public-private hybrid activities are not just confined to crime prevention and reduction. There is considerable evidence that national security policing (or high policing) is increasingly contracted to the corporate sector, and in the context of global insecurity, we are seeing something of a state-corporate symbiosis emerging whereby public and private actors engage in "a mutually beneficial application of game theory to achieve proximate objectives" (O'Reilly and Ellison 2006, p. 14). Increased privatization, marketization, commodification, and hybridization have all problematized the very nature of policing and its delivery across a myriad of domains. One cannot, for example, analyze the counterterrorism "performance" of the US police separately from the huge corporate infrastructure (security consultancies, security contractors, risk advisors, and so forth) that has grown up around homeland security. To put the point more simply, it is becoming much harder to track and measure who does what, when, and how, making efforts at constructing cross-national baselines that are empirically verifiable, reliable, and generalizable difficult, if not again, nigh on impossible. Furthermore, in those jurisdictions that are characterized by high levels of violent crime (South Africa and Brazil are cases in point), private security arguably eclipses public policing – at least for those that can afford it. In both jurisdictions thousands of heavily armed private security firms with alluring names such as "Blue Angels" provide security services in response to a near total lack of trust and confidence in the public police. Furthermore, high levels of crime in the Brazilian *favelas* have spurred grassroots forms of civil policing, with hundreds of authoritarian paramilitary style, *Autodefesas Comunitárias* (citizen defense committees), emerging to effectively take the law into their own hands and to deal with transgressors through a variety of extralegal punishments, such as shootings and beatings. This blurring of the boundaries between public and private policing/security further complicates any attempt at comparative analysis, unless we know in detail what these relationships are, and are able to control for the activities of public and private actors in a way that makes analytical sense.

Even focusing on specific models of policing does not lead to any productive conclusions. A general strategy of policing that we find being implemented across the globe is community orientated policing, which, therefore, might provide us with a base from which to make comparative analyses. Irrespective of whether it "works" or not, community policing allows police forces across the world to use a common language and as a philosophy offers a meeting point for "stakeholders" that becomes a forum for discussion and international collaboration via bi- and multilateral exchanges, reform efforts, conferences, seminars, and the movement of policing personnel within what Otwin Marenin has termed the "transnational policing policy community"(Marenin 2007). So at least in theory, it may provide a unified vision of global policing by establishing a series of international benchmarks. However, much of this remains aspirational, and as Brogden and Nijhar (2005) assert, owing to the different historical and political development of policing systems, the relationships between state institutions such as the police and the community are greatly differentiated between jurisdictions. This results in community policing becoming quite differently understood across and even within nations (2005, pp. 104–107). Brogden and Nijhar (2005, p. 110) expose such stark differences in the way community policing is deployed, particularly in regard to whether policing functions are centralized or decentralized that it is almost impossible to generalize from jurisdiction to jurisdiction. In some jurisdictions, such as Belgium, Northern Ireland, and Chicago, for example, attempts have been made to place community orientated policing strategies at the core of all police activities. But, in the most part, such strategies remain on the periphery of policing priorities, often playing second fiddle to national security priorities and the maintenance of social order. Certainly, the extent to which policing systems are autonomous from the state authority has a significant impact on the relationship between citizens and criminal justice system, skewing some police systems more towards maintaining social order and the political status quo than others.

Quantitative Databases

If we assume (at a stretch) that the core mandate of many police organizations is crime reduction, then sources of international crime data might be useful when attempting to make cross-national comparisons between individual forces. These include data collated by INTERPOL, the United Nations Crime Surveys, and the European Sourcebook (see Stamatel 2006). In the INTERPOL data, police representatives of the member nations complete quantitative returns on police records in 11 offense categories: murder, sex offenses (including rape), serious assault, all kinds of theft, robbery, breaking and entering, motor offenses, fraud, counterfeiting, and drug offenses. However, as Stamatel reports (2006, p. 16), although the definitions for each offense are rounded as closely as possible within clear definitional frameworks, the data process is a voluntary one which provides little in terms of quality control measures over offense categorization. Even INTERPOL itself notes that "the figures must be interpreted with caution" (Interpol 1999).

With regard to the United Nations Crime Survey, the UN has been collecting crime data since 1970 – with "a view to improving the analysis and dissemination of that information globally" (United Nations Office on Drugs and Crime 2007). The data is based upon multilingual questionnaires given to UN crime and police coordinators in member countries (Stamatel 2006, p. 17). They record aggregate national level crime figures within four areas: police, prosecution, courts, and corrections. Again, data is based upon offenses reported to the police and is collected within 10 offense categories: homicide, rape, assault, thefts, fraud, embezzlement, drug-related crime, bribery, corruption, and kidnapping. However, not all member states report back evenly nor are the data easily fixed within the same categories. For the data that compare prosecution and courts systems, it is not at all clear how the effects of different prosecution systems are controlled for, such as whether they are adversarial or inquisitorial, nor do they control for the effects of plea bargaining where some cases do not go to trial, which is used much more extensively in the USA and the UK than in other jurisdictions. Such data must be approached with caution, and it is difficult to make any cross-national assessments of the police or indeed of criminal justice agencies more generally.

Thirdly, the European Sourcebook of Crime and Criminal Justice Statistics started collecting data in 1990. It collates data within the categories of homicide, assaults, rape, armed robbery, motor vehicle theft, burglaries, drug-related offenses, and people trafficking. Unlike the other two sources, the European Sourcebook also collates information about the number of offenders by crime type and the percentage of offenders who are female, minors, and nonnationals (Stamatel 2006, p. 18). The sourcebook also claims to offer a higher level of quality control in the data collection effort (Stamatel 2006, p. 19). This includes standard classification schemes that are implemented over the course of the data collation process and the fact that data collection are easier to undertake within the European Union since most states have a closer affinity with regard to their crime classification. However, the sourcebook suggests that even within the European Union, there are multiple dimensions to crime classification and data collection that vary considerably between Western, Central, and Eastern European nations (Krajewski 2011). To take just one example, there is considerable cross-national variation in the statutory definition, reporting, recording, and prosecution of rape within European jurisdictions (Lovett and Kelly 2009).

International Victimization Survey

By its very nature, policing is a subjective exercise, located in huge amounts of discretion and partiality in what crimes are prosecuted and recorded. Thus, police statistics provide a skewed picture of offending and are limited to illustrating, at best, broader trends in crime (Maguire 2002; Skogan 1975). Moreover, it has been extensively documented that there are validity and reliability issues with official crime statistics as a means for assessing levels of crime and police performance in any given jurisdiction (see Maguire 2002; Skogan 1975).

So, in addition to police records that are the mainstay of the three sources indentified above, the International Crime Victimization Survey (ICVS) collects data on victimization rates (see Van Dijk et al. 1990). The ICVS provides researchers with some detailed points of comparison with officially recorded data and, therefore, is worth considering further. Most industrialized democracies have a long tradition of publishing victimization surveys (e.g., see British Crime Survey or the National Crime Victimization Survey). The ICVS surveys cross-national household experiences of crime, perceptions of policing, as well as general feelings of safety and security. The ICVS was initiated in 1987 by a group of European criminologists with expertise in national crime surveys (Van Dijk et al. 1990), and initially it was operationalized in 14 industrialized countries. The survey seeks to advance international comparative criminological research beyond the constraints of officially recorded crime data to produce estimates of victimization that can be used for international comparison. Surveys have taken place in 1992, 1996, 2000, 2005, and 2008, with 140 surveys now having been deployed in over 78 different countries – sampling over 320,000 citizens. The present database covers 325,454 individual respondents. Crime categories are broken down into vehicle-related crimes (theft of a car, theft from a car, theft of a motorcycle or moped, theft of a bicycle), burglary, attempted burglary, theft of personal property, and contact crimes (robbery, sexual offenses, and assault and threat).

Arguably, the ICVS is the most comprehensive instrument yet developed to undertake an international comparative study of volume crimes, victimization, and perceptions of the police and criminal justice agencies. Significantly, the data is derived from surveys among the general public and, therefore, not influenced by political or institutional prerogatives. For the IVCS, validity and reliability is better maintained through the standardization of questionnaires, and results are presented within statistical confidence margins (see Lynch 2006). But full standardization, in all aspects of design, has still to be proven, especially for the surveys deployed in developing countries. Even if the ICVS provides data that are fit for international comparison, we have to be wary of country-specific circumstances that will not always allow for fully standardized methodology to be applied. Moreover, although the ICVS provides a measure of common crimes, the comparatively small samples sizes precludes estimations of less prevalent crimes such as rapes or aggravated assaults. The ICVS also ignores victimization by complex crimes such as largesse forms of corruption or organized crime, which victimizes collective populations rather than individuals.

Victim surveys and recorded crime statistic invariably produce two quite different pictures of the volume and distribution of crime. Yet, it is not clear whether this is because victim-based statistics are more accurate than recoded crime statistics or vice versa (Skogan 1975). Although some might argue that levels of recorded crimes cannot be reliably used for comparing levels of common crime across countries, measurement procedures in both formats illustrate characteristic errors. Nevertheless, victim surveys seem a slightly more productive source from which to undertake cross-national comparative analysis, although this should not be exaggerated and we still need to be cognizant of a whole range of conceptual and methodological issues.

Dilemmas and Problems

Difficulties with Cross-National Analysis

The entry has so far described three main sources of data that might be used to help in the production of cross-national comparisons of policing systems. Besides some of the specific issues already observed within each method, a number of general problems with undertaking such analyses can be identified. For instance, there are no unilinear patterns of development that all nations follow, and likewise, in many developing nations the structures of public policing and criminal justice generally diverge considerably from those found in advanced liberal democracies. Critics have observed that "western" ethnocentric concepts and categories are often superimposed on less industrially developed countries and their translation varies enormously (Shichor 1990, p. 65; Nelken 2009). Cultural and geopolitical differences, therefore, make direct comparisons difficult (Nelken 2009). Still, it might be argued that cross-national analyses could be made across nations that have evolved into similar advanced democratic civic systems – such as the EU or North America – where measurable, complex, bureaucratic, and recording infrastructures such as performance indicators are deployed (see Collier 2001). For example, an inordinate number of indicators have been used to measure success and failure in crime reduction. Some indicators measure public satisfaction rates of police organizations, others crime clear up rates, while others track victimization rates. Certainly, patterns can be found on a number of key measures. Nevertheless, although liberal democracies are similar and more transparent, difficulties still arise, as performance indicators often track spurious clearance rates and measure arrests and detections rather than what has been achieved through say crime prevention measures. Performance indicators, however, usually coalesce around what have been termed "volume crimes" – due to their often uniformity in definition and large volumes in occurrence – and, therefore, are perhaps the easiest to measure. Again, differences arise in the way nation states not only record crimes but also in the administrative infrastructure available for recording to take place (see Krajewski 2011).

On another level, offense categories themselves often prove problematic, as they do not easily subscribe to exactly the same definitional boundaries. Dealing with antisocial behavior is a key issue for the UK police, frequently ranking as the number one priority concern among the public, but it proves problematic when trying to fit it into a singular definitional or legal framework. Reducing analysis to one variable that is common and uniform in most societies, such as "homicide," for example, might provide a good indicator, not just of rates between countries but also police performance in investigative terms. Indeed, various studies have used homicide to make cross-national/cross-cultural comparisons (e.g., see Avison and Loring 1986; Krahn et al. 1986). Still, what might look superficially similar is often framed within local, cultural, political, and legal contexts that limit our ability to make generalizations (Nelken 2009).

Even within the UK, for instance, Scotland defines homicide as including the offenses of "murder" and "culpable homicide," whereas in England and Wales, it is defined as including "murder," "manslaughter," and "infanticide" (Richards 1999). Elsewhere in the European Union, causing death by dangerous driving qualifies as manslaughter in some jurisdictions, but not all. In the UK suicide was only decriminalized in 1961 and prior to that time was regarded in law as a form of self-homicide. Currently, many jurisdictions regard death as the result of assisted suicide second-degree murder or manslaughter. However, in Australia's Northern Territory, the use of the so-called Deliverance Machine to perform an assisted suicide was legal between 1995 and 1997 with the death not recorded as homicide. Similarly, in Switzerland assisted suicide remains legal with critics claiming that current legislation contributes to "suicide tourism" from other jurisdictions. Even for those offenses deemed to be homicides, initially around 15 % are reclassified as "other" following police or court action (Richards 1999). In a major study of homicide statistics in 36 European nations (Anamort European Project 2008), the researchers found that the misclassification of homicides was a significant problem in 14 of the countries surveyed. The combined effects of such misclassification were felt to "lead to underestimation of the magnitude of the deaths due to homicides in all of these countries" (Anamort European Project 2008, p. 3). The main problem appeared to relate to the lack of accurate information provided to coders on the cause of death which had not been forwarded from the police or coroner's office. This appeared to be less of a problem among the established democracies of northern and Western Europe owing mainly to higher standards in forensic investigation/pathology but also, as noted above, to more efficient transmission of medicolegal data to national statistical offices (Anamort European Project 2008, p. 3).

Furthermore, as Richards (1999, p. 30) points out, in a number of jurisdictions, homicide statistics are compiled from when a death was registered as a homicide rather than when the death occurred. In the UK, all cases of suspected homicide are referred to the coroner's court. However, in complex cases it is not unusual for the determination of cause of death to be made up to several years and in some cases decades later. The Police Service of Northern Ireland's (PSNI) Historical Enquiries Team (HET) which was established to investigate over 1,800 cold-case deaths arising from the conflict is currently reinvestigating deaths that occurred 40 years ago. So far, a number of these cases have been referred to the coroner's court for a determination on the actual cause of death. Our point then is that even attempting to compare police investigative performance between jurisdictions for something as ostensibly straightforward as homicide that seems to be measurable across a number of jurisdictions is fraught with difficulties around definition, classification, and ultimately interpretation.

Conclusion

To reiterate the starting points, clearly, as policing has become a multimillion dollar and a more globally transnational exercise, the pursuit for efficiency has produced an expected desire to measure police performance on a cross-national basis. This entry asked whether any similarities could actually be identified across different nations in order to undertake such measurements. Three data sources have been considered: (a) academic comparative research, (b) official crime data, and (c) international victimization surveys. In these contexts, the entry has directed the reader to some of the difficulties that encompass carrying out any adequate and robust cross-national comparisons of policing systems. Comparison both requires understanding and interpreting what those in other jurisdictions are trying to achieve. Whether it is comparing a particular police organization's performance data or using crime statistics as a barometer of police performance, the cultural and administrative differences that characterize the organization of policing at the cross-national level have to be considered. Surveys such as the European Sourcebook can provide rich comparative data, which given the continued harmonization among states within the EU can lead us to perceive that international comparisons are becoming easier. However, within the EU there are inconsistent levels of technology and robust administration systems to accurately record crime which tends to give rather anomalous results in comparative terms. For example, some of the new accession states (Latvia, Lithuania, Estonia, Poland) appear on some indices to have even crime rates lower than that of Sweden, which traditionally within Western Europe has had low crime rates coupled with high levels of social welfare provision (Krajewski 2011). In addition, as noted above, some EU nations are shifting to decentralize their police (Belgium), some are tending towards enhanced centralization (the Netherlands, Scotland), and some are maintaining traditional organizational structures (Germany, France, Italy, Spain), while others are further privatizing aspects of the police role and function (England and Wales). These organizational differences are coupled with variant cultural understandings of "crime" and "justice," variations in recording practices, the lack of comprehensive national datasets, and different legal cultures and political systems that lead us towards the conclusion that, to say the least, an objective benchmark comparative analysis is a tough proposition. Yet, despite the problems that have been identified, aggregate crime statistics may provide the most useful indicator of cross-national police performance, although these need to be qualified heavily and interpreted with some caution. Overall, however, the international vagaries of policing mean that a definitive cross-national comparative analysis remains a rather elusive goal.

Related Entries

- ▶ BCS
- ▶ British Police
- ▶ Comparing Police Systems Across the World
- ▶ Crime Mapping
- ▶ Development of the UCR and the NCVS
- ▶ History of the Dutch Crime Victimization Survey(s)
- ▶ Methodological Issues in Evaluating Police Performance
- ▶ National Victimization Surveys
- ▶ NCVS

Recommended Reading and References

Anamort European Project (2008) Homicide-related deaths in an enlarged European Union, Anamort European Project, June. http://ec.europa.eu/health/ph_projects/2004/action1/docs/action1_2004_frep_22_a10_en.pdf

Avison WR, Loring PL (1986) Population diversity and cross national homicide: the effects of inequality and heterogeneity. Criminology 24:733–749

Bayley DH (1990) Patterns of policing: a comparative international analysis. Rutgers University Press, New Brunswick

Brogden M, Nijhar P (2005) Community policing: international models and perspectives. Willan Publishing

Collier P (2001) Police performance measurement and human rights. Public Money Manag 21(3):2001

Commonwealth Human Rights Initiative (2007) Feudal forces: democratic nations: police accountability in commonwealth South Asia. Available: http://www.humanrightsinitiative.org/publications/police/feudal_forces_democratic_nations_police_acctability_in_cw_south_asia.pdf

Dennis N, Erdos G (2005) Cultures and crimes: policing in four nations. Institute for the Study of Civil Society, London, Civitas

European Sourcebook of Crime and Criminal Justice Statistics (2007) http://www.europeansourcebook.org/

Interpol (1999) International crime statistics. ICPO-Interpol General Secretariat, Lyon

Krahn H, Hartnagel TF, Gartrell JW (1986) Income inequality and homicide rates: cross national data and criminological theories. Criminology 24:269–295

Krajewski K (2011) Crime and punishment in central and Eastern Europe during the last 20 years – is the region different from rest of Europe? Plenary paper given at the European Society of Criminology annual conference, Vilnius, Sep 2011

Lau RWK (2004) Community policing in Hong Kong: transplanting a questionable model. Crim Justice 4(1):61–80

Lelandais I (2007) Measuring police performance: international experiences. International Centre for the Prevention of Crime, October

Lovett J, Kelly L (2009) Different systems, similar outcomes? Tracking attrition in reported rape cases across Europe. Child & Woman Abuse Studies Unit, London Metropolitan University. Available: http://www.sccjr.ac.uk/projects/Different-Systems-Similar-Outcomes-Tracking-Attrition-in-Reported-Rape-Cases/63

Lynch JP (2006) Problems and promise of victimization surveys for cross-national research. Crime Justice 34:2

Lynch JP, Addington LA (2007) Understanding crime statistics: revisiting the divergence of the NCVS and UCR, Cambridge studies in criminology. Cambridge University Press, New York

Maguire M (2002) Crime statistics: the data explosion and its implications. In: Maguire M (ed) The Oxford handbook of criminology, 3rd edn. Oxford University Press, Oxford

Marenin O (2007) Implementing police reforms: the role of the transnational policy community. In: Goldsmith A, Sheptycki JW (eds) Crafting transnational policing: police capacity-building and global policing reform. Hart Publishing, Oxford, pp 177–203

Mayhew P (1987) Residential Burglary: a comparison of the US, Canada and England & Wales. National Institute of Justice/Government Printing Office, Washington, DC

Nelken D (2009) Comparative criminal justice beyond ethnocentrism and relativism. Eur J Criminol 6(4):291–311

Newman WL, Berger RJ (1988) Competing perspectives on cross-national crime: an evaluation of theory and evidence. Sociol Q 27:27–41

O'Reilly C, Ellison G (2006) Eye spy private high: reconceptualising high policing theory. Br J Criminol 46(4):641–660

Richards P (1999) Homicide statistics, House of commons research paper 99/56, House of Commons Library. Available: http://www.parliament.uk/documents/commons/lib/research/rp99/rp99-056.pdf

Shichor D (1990) Crime patterns and socioeconomic development: a cross-national analysis. Crim Justice Rev 15:64

Skogan WG (1975) Measurement problems in official and survey crime rates. J Crim Justice 3(1):17–31

Stamatel JP (2006) An overview of publicly available quantitative cross-national crime data. IASSIST Q

Stenning PC (ed) (2003) Police use of force and human rights. Simon Fraser University, Criminology Research Centre, Occasional Paper 2003–02. Available: http://www.sfu.ca/crc/fulltext/pufrep.pdf

Tanner MS (2005) Campaign-style policing in China and its critics. In: Bakken B (ed) Crime, punishment and policing in China. Rowman & Littlefield, Lanham, pp 171–188

United Nations Office on Drugs and Crime (2007) United Nations Surveys on crime trends and the operations of criminal justice systems. Retrieved 3 July 2007 from http://www.unodc.org/unodc/en/crime_cicp_surveys.html

Van Dijk JJM, Mayhew P, Killias M (1990) Experiences of crime across the world. Key findings from the 1989 international crime survey. Kluwer Law and Taxation Publishers, Deventer

Cryptology and Security

Jonathan C. Hoskin
Clemson Computing and Information Technology, Clemson University, Clemson, SC, USA

Overview

Commerce in the modern world is based on virtually instantaneous electronic transmission of financial data including movement of money. The Internet has allowed us to move away from physical transactions via checks and currency. However, it is dependent on an unprecedented level of trust with an equally unprecedented potential of catastrophic financial loss. Illegal movement of money in the electronic form requires no physical access nor leaves physical evidence and can be accomplished from the other side of the world. Thus, the need for unquestionable security has arguably equaled or eclipsed that of military needs from just a few decades ago. With so much to gain and so little to lose, the criminal element has been transitioning to take advantage of every opportunity afforded by less than adequate security measures.

Security must be well planned, implemented, monitored, enforced, and reviewed. Use of encryption and decryption is an absolute necessity and thus must be more dependable than any time in the history of the planet. The cryptographic algorithms used ideally must not simply be rock-solid, but beyond question, thus our dilemma. At best we know that cryptography is a work in progress and we unfortunately cannot prove it is perfect, – we can only prove it is imperfect. We use computer systems and all the rest of the infrastructure knowing that it is flawed by virtue of being made by humans.

This short introduction notes some of the early methods which often translate into present-day concepts and vulnerabilities. Use of letter substitutions and transpositions defines early cryptology but was nevertheless a dependable method when used correctly. It has always been the misapplication of techniques that rendered cryptographic systems vulnerable. Topics include stream and block ciphers, passwords, and symmetric versus Asymmetric cryptography.

Fundamentals and Future Directions

One of the more dependable themes in modern entertainment is the thriller in which the supposedly impossible was done. Security of a bank, museum, or fortified building is breached, and something is taken defying incredible odds. Invariably many types of security systems are circumvented, and by luck, guile, and intelligence, the deed is done. Its comparable to an airplane crash in that not one, two, or even three things went wrong but a combination and other factors that caused the accident, any of which if recognized would have changed the outcome. Computer security can be viewed in much the same way. Human intelligence is a marvelous thing, but considering every possibility requires that people are perfect and humans can at best overcompensate for their inadequacy. Recognizing that advances in technology may swiftly undermine security, overcompensation should not be considered a luxury.

This short survey of cryptology begins with the above statement because it reflects reality. Every week it seems that another incredible security weakness has been exploited, usually resulting in or expecting to cause significant financial loss. Computer magazine, the flagship publication of the IEEE (the Institute of Electrical and Electronic Engineers) Computer Society noted in its October 2011 issue that “Hackers steal $13 million in 1 day via online security breach (Garber).” History has shown that criminals constantly get better; they find and attack the weak parts of physical and computer/networked systems where they exist and circumvent systems when they do not. Contributing on the positive side are white-hat computer hackers that test and evaluate systems, often discovering weaknesses and security holes that plague software written with too little attention paid to security.

The following document addresses the topic believing it to be fundamentally sound and capable of providing the security needed. However, like everything else in life, it is invariably the implementation and correct use of the system that complete the process. A simple password renders the best encryption system useless, and readers should understand any limitations will be exploited. At best cryptography and related security is a work in progress.

In the modern world, encryption and decryption via use of credit cards, security cards, passwords, and other tokens are part of everyday life and becoming ever more important. No one would ever think to leave a $100 note bill in clear view in their automobile or on their desk at work, but somehow they are not as concerned about a password which may represent significantly more financial risk. This disconnect with reality is the target of social engineering, attempting to obtain such information via sociological means. Although social engineering is not the specific topic of this document, it should be considered at each point of a security audit.

There are many approaches to address the topic of cryptology. A mathematical explanation is appropriate for those with the required mathematical background; however, this document will minimize formulaic expressions. Mathematics will be limited to numbers of possibilities and understanding the significance of large numbers (for a mathematical view, see Menezes et al. (1996)).

Several years ago a story was circulated around the Internet that began: *“Aoccdrnig to a rscheearch at Cmabrigde Uinervtisy, it deosn’t mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae.”* While the attribution was wrong, the scrambled message makes a very important point. Breaking an encrypted message does not necessarily require getting every "t" crossed and "i" dotted. As a consequence, linguists and puzzle solvers were critically important in this field, and it is only the recent development of the computer-based cryptology that has reduced their need.

Keeping messages private has always been an obvious requirement for military or diplomatic endeavors. The general study of such matters is called cryptology. It can be divided into two disciplines: obscuring or inhibiting messages from being understood by some manner of encryption, the study of which is called cryptography, and the breaking of those coded messages called cryptanalysis. In general use, cryptology is often referred to as cryptography which may lead to some confusion. This may be further complicated by steganography, the concept of hiding a message in some fashion and which itself may be encrypted.

Often, the modern study of cryptology gives less recognition to early methods and origins, progressing quickly to complex computational methods which when used correctly can surpass the capability of being broken by even the most sophisticated computing system. While a general understanding of computer-based models is appropriate, its utility is beyond many criminals although those with a technical understanding or ability to employ those technically savvy should recognize that in fact it can be unbreakable. This document will therefore note various aspects of the historical development because the many possibilities remain useful in everyday life (Callery 2008). The classic document of cryptology leading into the computer age is *The Code-Breakers* (Kahn 1967) although some may find *The Code Book* a smaller and more digestible volume (Singh 2000). A shorter review can be found in the journal Cryptologia (Al-Kadi 1992.)

Perhaps the earliest and most obvious method of obscuring information was the written form when few people could read. Related to that was use of a different language. The royal courts in Europe invariably spoke a different language than the common mostly illiterate person of a given country, thus protecting conversations from workers and limiting interaction to all but the educated. Changing languages throughout the conversation or in written messages was almost guaranteed to protect information when the unintentional listener/reader is only marginally literate. We continue to employ this technique today to keep information about birthday parties or similar concerns from our children. This may also be accomplished by using more complex wording without necessarily needing to change languages, but children (and criminals) learn quickly.

During World War Two, the United States used the Native American dialect of the Navaho to communicate securely between marines. This technique could only work if the opposing side had no expertise in the language chosen which had been carefully predetermined. Navaho was selected from among a wide variety of American languages because the population of speakers was high enough to meet military needs in terms of number of recruits. Traditional Navaho words also had to be reinterpreted to describe modern military hardware and other issues which were not part of the language. In a similar way, the "language of the street" incorporates words that replace others such as drug names or various criminal acts.

Another simple method of passing such messages is to physically hide it such that the other side cannot find it. In early Roman times, a slave could have his hair cut, have a message tattooed on his head, and after hair regrowth, the message was hidden. Recutting the hair revealed the message after traveling to an appropriate location. Clearly not a mechanism for timely disseminating of information, it can have the advantage of hiding the message from the slave. Unfortunately for him, he need not be alive to convey the message.

Using a locked strong box was a viable option for sending messages and other more tangible items, but rather than a single lock, a system was devised to use two locks. The owner placed items needing protection in the box and secured it with their lock. It was sent to another person, and instead of opening the box, something he or she could not do without the appropriate key, they added their own lock via a second hasp and sent it back unopened to the sender. Upon the return of the box, the original lock was removed and resent to the previous sender. It could then be opened because only the original receiver's lock remained on the box. However, this technique could be exploited by what is known as a "man in the middle attack." This required the box be diverted to another person before reaching its original target, a second lock was added by the interloper, and it was returned to the original sender. If the owner had no verification that the box reached its original destination, its contents could be obtained by that intervening person, and the sender might never know what happened. This was a concern hundreds of years ago and remains an issue today with regard to digital information.

There are numerous other ways to physically hide information. A message can be scratched into a material such as wood then covered with a layer of wax to hide it. A heat source will reveal the message after the wax drips away. In early roman times, the use of pigmented beeswax was an artistic technique now known as encaustic; thus, a message could masquerade as an artistic work. A slight twist to the above process would include the use of invisible ink. Typically, this has involved the use of "ink" which is clear when initially used. The ink may be as simple such as lemon juice or a specific chemical compound or mixture. Revealing the message involves treatment of heat or exposure to a chemical spray. Today, we can add notes to computer documents, and they remain hidden in plain sight because most people do not know about or consider their existence.

Although the above methods work, eventually a need developed for a more capable system to convey complicated or lengthy information in a routine fashion. Usually, individuals encrypt and decrypt information themselves which implies another security concern – more than one party will know the mechanism. The ultimate goal of a cryptography system is to protect information even if adversaries understand how it works.

A fundamental concept was the substitution cipher where letters are substituted to obscure the message. The Atbash cipher used in the bible was perhaps not strictly for security purposes. Its name comes from its use, substituting the last for the first letter, the second from the last for the second letter, continuing as defined by the Hebrew alphabet. Another simpler cipher is ROT13 which rotates back on M the thirteenth letter. If you cut the alphabet in half between M and N and then slide the second half under the first half were A aligns with N, B with O, etc., you have a cipher that is easier to use. The message "hello" becomes URYYB and there is no confusion to complicate decryption. Letters can be transposed rather than substituted within words somewhat as noted above referencing Cambridge, but the anagrams generated are perhaps even less secure.

Another transposition example will show the concept with a very short message. For the message ABCDEFGHI, the first nine letters of the alphabet, simply write the first three on the first line, the second three on the second line, and the remaining three on the third. Reading down the first column and then the second and finally the third, we get ADGBEHCFI for the encrypted message. The mechanism or cipher needs to allow for uneven rows and is not limited to only three lines. Obviously, there are many other possible ways of manipulating the message, reading down one column and then up the following column or going backwards. The one limitation is that all letters in the original message are present and there are a limited number of words and messages that can be made from them. It is easy to understand that compared to transposition, substituting letters can make the encrypted message much more difficult to decrypt. However, adding simple transposition to other building blocks of encryption can dramatically increase complexity to the process.

An early device allowed for the simple sliding of one alphabet against another. This became known as the Caesar cipher because Julius Caesar used it with his generals. It consisted of two disks, one smaller rotating on the top of the other and both having the alphabet imprinted there on.

Cryptology and Security, Table 1 Plain and encrypted text created using simple methods

Encrypting data is one way to keep it secure	Plain text
GPETARVKPI FCVC KU QPG YCA VQ MGGR KV UGEWTG	Encrypted text created by shifting two letters right
└▲□●┬○◙▄╥─_├╓◙╓_▄▪_━▲└_╧╓┬_◙▄ _◊└└○_▄◙_▪└□░●└	Encrypted text using atypical characters (with underlines indicating spaces for clarity)

This simple substitution is known as a cipher which is a series of steps followed to encrypt or decrypt a message. The original information is known as plaintext (often shown in lowercase letters) with the enciphered version (shown in small uppercase letters) known as the ciphertext. We can envision they can be lined up so that in English, the "a" on the top disk is aligned with the "a" on the bottom disk. Turning the smaller disk two letters to the left (Caesar used three letters), we can then use this cipher disk to encrypt a message. This movement of letters by two is the auxiliary information usually called the key or sometimes the cryptovariable (shown in uppercase letters). This is easier than Atbash or similar techniques because the letters remain in alphabetic order.

Using this substitution cipher above, the word "ate" becomes "CVG" and the line of text "encrypting data is one way to keep it secure" becomes "GPETARVKPI FCVC KU QPG YCA VQ MGGR KV UGEWTG," which initially seems imposing (Table 1). This would likely be unreadable to someone with no cryptography skills but is nevertheless very weak. Removing the spaces will make the message look even more difficult but is no match for the cryptanalysis methods below. Difficulty of code breaking may also appear to be increased by using alien or nontraditional characters such as those taken from other languages, cultures, or additional fonts. This is deceptive only for those not familiar with encryption.

Each language has their own peculiarities, there are a limited number of letters which are repeated, and some letters occur in one order but never reversed. In English a double "h" is unlikely compared to a double "e," and a "u" will follow a "q" but never in the reverse order. More important in deciphering messages are the frequency of the letters which cannot be hidden in the above examples. The standard frequency distribution of letters is different for a given language. Generally, in English the letter frequency order is E, T, A, O, I, N, S, H, R, D, L, C, U, M, W, F, G, Y, P, B, V, K, J, X, Q, and Z. A message of sufficient length or multiple simple substitution messages combined allow you to break the message. Simply replace the most frequent letter with E and the second most frequent with T and continue with all letters, and even if not perfect, the result is probably decipherable.

Another way of representing letters is to use not one but two or three characters. The benefit of this technique is a single letter can be represented by multiple combinations. When each letter is represented by two alphabetic characters, the number of possibilities is 26^2 or 676, thus providing 26 combinations for each letter if distributed evenly. Using three characters, the number is 26^3 or 17,576 with even greater possibilities. Thus, the letter "a" might be "AEL," "MLT," and "ZYE," for example.

When two or three characters are used to represent one letter, not only can each letter have multiple representations but they might also key for words. Thus, "AML" might be code for "bank," and "LTL" could translate to "dollars." To decrypt a mixture of substituted letters and code words requires its' nomenclator. Certainly one of the most historically important encrypted messages was known as the Zimmermann telegram and used a numeral-only nomenclator with numbers up to 99,999. The purpose of that telegram was to entice Mexico into World War One on the German side and keep the USA out of Europe long enough to subdue England. Having used the same nomenclator previously, it had already been partially broken, and its decryption was a significant point in the war.

The usefulness of this possibly more secure technique using a nomenclator is sometimes diminished by the necessity of agreeing on all likely code words beforehand and requiring a code book. If security is the purpose, the nomenclator must be stored and used securely to not fall into the wrong hands. But as noted above, increasing use results in decreasing security.

To expand the possibilities of the Caesar cipher and increase its complexity and thus security simply requires not keeping the disk or substitution the same. If, for example, the first letter is moved by two and the second by four, “ee” now becomes “GI,” and if repeated, the result is not so obvious. That initially makes the cryptanalysis more difficult, but using frequency analysis it would eventually be recognized that there are two different groups of letters – the simplest polyalphabetic cipher. The natural end to this technique is to use all the possibilities and was ultimately done by Blaise de Vigenère (see Table 2). Continuing on the same road, what if a word was used for the cipher key? If the word “CIPHER” was the cipher, that would mean lining up the “a” with the “c” and using the wheel to determine the first letter and then lining up the “a” with the “i” for the second letter. After the sixth letter a lined up with “r,” the same code is repeated until the message is complete. As the code word cipher becomes longer, cryptanalysis becomes ever more difficult but not impossible.

Up to this point in the development of cryptology, the cryptanalysts were winning but that was about to change. Consider what happens if the code keyword is 100 random letters from a to z, now the encrypted message becomes extremely difficult to break. True randomization makes the encrypted message random characters and thus unbreakable when used correctly. One important requirement becomes access to truly random numbers. Before computers became available, the process of generating random numbers might be similar to the numbered balls used for lottery drawings. However, because encryption has become ubiquitous and the need for random numbers dramatically increased, we are limited to computer-generated or pseudorandom numbers. Such numbers may not be perfectly random but are sufficiently random to not reduce the security of the encryption. For cryptographic purposes, the process of random number generation must be verified to ensure quality, or unbiased random numbers are generated.

Cryptology and Security, Table 2 Vigenère square

Key	abcdefghijklmnopqrstuvwxyz
1	BCDEFGHIJKLMNOPQRSTUVWXYZA
2	CDEFGHIJKLMNOPQRSTUVWXYZAB
3	DEFGHIJKLMNOPQRSTUVWXYZABC
4	EFGHIJKLMNOPQRSTUVWXYZABCD
5	FGHIJKLMNOPQRSTUVWXYZABCDE
6	GHIJKLMNOPQRSTUVWXYZABCDEF
7	HIJKLMNOPQRSTUVWXYZABCDEFG
8	IJKLMNOPQRSTUVWXYZABCDEFGH
9	JKLMNOPQRSTUVWXYZABCDEFGHI
10	KLMNOPQRSTUVWXYZABCDEFGHIJ
11	LMNOPQRSTUVWXYZABCDEFGHIJK
12	MNOPQRSTUVWXYZABCDEFGHIJKL
13	NOPQRSTUVWXYZABCDEFGHIJKLM
14	OPQRSTUVWXYZABCDEFGHIJKLMN
15	PQRSTUVWXYZABCDEFGHIJKLMNO
16	QRSTUVWXYZABCDEFGHIJKLMNOP
17	RSTUVWXYZABCDEFGHIJKLMNOPQ
18	STUVWXYZABCDEFGHIJKLMNOPQR
19	TUVWXYZABCDEFGHIJKLMNOPQRS
20	UVWXYZABCDEFGHIJKLMNOPQRST
21	VWXYZABCDEFGHIJKLMNOPQRSTU
22	WXYZABCDEFGHIJKLMNOPQRSTUV
23	XYZABCDEFGHIJKLMNOPQRSTUVW
24	YZABCDEFGHIJKLMNOPQRSTUVWX
25	ZABCDEFGHIJKLMNOPQRSTUVWXY
26	ABCDEFGHIJKLMNOPQRSTUVWXYZ

During World War Two, making a cipher unbreakable involved using a one-time pad, a piece of paper with key codes. Critically important is that it is only used once. Codes repeatedly used create patterns cryptanalysts use to break the code. Judicious use of a sufficiently long and truly random one-time pad may fulfill the cryptographic needs, but it creates other problems – it is only as secure as the one-time pad, and transmission of such information creates its own problems. When the adversary finds 30 days or more of one-time pads, because they may already be disseminated to ships at sea, for example, you may have lost the ability to keep messages secret. Likewise, given the length of the one-time pad should be at least as long as the message itself, this unbreakable method is unfortunately of limited utility in the current world.

By the early 1900s, there had been sufficient interest in cryptography to entice inventors to attempt making devices to accomplish this task. The best documented is the Enigma machine used by the Germans in World War Two which was an electromechanical device to encrypt/decrypt messages. It might be simply described as an electrical–mechanical version of multiple Caesar cipher disks called rotors, chained together such that when a typewriter key is pushed, the result is an illuminated letter. It had the possibility of being nearly unbreakable, but the allies had one so they knew how it worked and it was misused. The Germans essentially repeated a portion of the same message each day because the daily weather report began with the same words. Thus, if settings on the Enigma could be manipulated to decrypt the ciphertext back to those words, they could decrypt the entire message. Although this was not an easy task given the state of computing devices, it allowed messages to be broken with hard work. Not every transmitted message would be decrypted, and those that were might not have been decrypted in time to be useful.

The above examples are based on traditional characters or ones that replace those characters. With the advent of computers, all characters were replaced by a binary code, strings of ones and zeros. Seemingly elementary this enabled a vast and wide-ranging new world of unprecedented complexity. For the uninitiated, a simple example is no further than a typical garage door remote. To set or change the code, you open it and move the 8, 12, or more switches in both the sender and receiver. (Leaving the original settings unchanged is the same as keeping the default password on a computer – it is only marginally better than not having a password or code.) Eight similar switches grouped together are known as a computer byte.

Cryptology and Security, Table 3 Number of possibilities for selected numbers of bits

Bits	Bytes	Math	Number of possibilities
1	1/8	2^{1}	2
2	1/4	2^{2}	4
3	3/8	2^{3}	8
4	1/2	2^{4}	16
5	5/8	2^{5}	32
6	3/4	2^{6}	64
7	7/8	2^{7}	128
8	1	2^{8}	256
12	3/2	2^{12}	4,096
16	2	2^{16}	65,536
24	3	2^{24}	16,777,216
32	4	2^{32}	4,294,967,296
64	8	2^{64}	1.84467×10^{19}
128	16	2^{128}	3.40282×10^{38}
256	32	2^{256}	1.15792×10^{77}
512	64	2^{512}	1.34078×10^{154}
1,024	128	$2^{1,024}$	1.79769×10^{308}
2,048	256	$2^{2,048}$	3.23170×10^{616}
4,096	512	$2^{4,096}$	$1.04439 \times 10^{1,233}$
8,192	1.024	$2^{8,192}$	$1.09075 \times 10^{2,466}$
16,384	2.048	$2^{16,384}$	$1.18973 \times 10^{4,932}$

A byte consists of 2 × 2 × 2 × 2 × 2 × 2 × 2 × 2 possibilities also represented as 2^{8} or mathematically as 256 possibilities. This respectively represents a doubling of possibilities with each additional character or 2, 4, 8, 16, 32, 64, 128, and 256. This is a key concept because the code possibilities increase geometrically, while the ability to test a code or pass-phrase is linear (Table 3). This has a dramatic effect as passwords get longer, and the difficulty of breaking passwords becomes increasingly difficult depending on the current state of the art of cryptanalysis. If we spend twice as long inputting pass-phrases, we have only attempted twice as many codes. Ultimately, the trade-off is defined by the cost of breaking the code (the computational expense) vs. the worth of the protected information.

By choosing the option of 12 bits on a garage door opener, we effectively have a byte and a half or 2 × 2 × 2 × 2 × 2 × 2 × 2 × 2 × 2 × 2 × 2 × 2 possibilities also represented as 2^{12} or 4,096 possibilities. This is not an overwhelming number, and a device that produces all possible signals simulating all garage door-opening possibilities is easy to envision. Another byte (2^{20} or 1,048,574 possibilities) of additional complexity would make it increasingly unlikely that a code generator could produce all the possible code signals in a short span of time. The garage door opener is a security device meant to keep honest people honest and cannot be considered sufficient to protect someone that has been targeted by those with criminal intent.

Although computer technology may have been pressed into the cryptology arena very early because of world wars, passwords were not the initial concern. Nevertheless, we use them every day and they contain information which may be built upon to understand others' aspects of cryptology. Portrayal of the mathematical complexity of password encryption is difficult and most obviously incorrect in popular media. Breaking a password is often presented as identifying each character, one at a time in sequence, which might accurately represent a safe cracker listening to the tumblers fall in place using stethoscope on an old-style safe but has nothing to do with passwords. This is particularly disconcerting because, as any computer or ATM (automated teller machine) user inherently understands, you either get the password right or you do not. Unfortunately, the underlying message conveyed to the public is that passwords cannot be safe and are easily broken, thus unintentionally undermining the need for using good passwords. Believing you know all but the first or last character of a password is the same as not knowing the password. At that point, it seems only a brute force attempt at breaking the password or portions believed to be unknown will allow it to be determined. A brute force attack means trying every possibility until the correct one is discovered. The goal of the cryptographer is to limit access to only the correct password or a brute force attempt because the latter is defined by the state of computing technology and the implementation of the system. But this only works dependably if the user has a good password, hereafter known as a pass-phrase, and keeps it protected.

Use of pass-phrase instead of password is an important change meant to entice and empower the user into choosing more secure passwords. The user need not have to remember a pass-phrase; it can be recorded on a piece of paper and subsequently stored in a wallet or purse. Thus, it is considered as important as money, credit cards, and other information which the user intentionally keeps out of reach by others. Regardless of complexity, the user ultimately remembers the pass-phrase even if it takes several days. However, it should therefore be obvious that requiring frequent pass-phrase changes discourages use of strong pass-phrases and conveys the entirely wrong message.

With regard to pass-phrase parameters, two important factors are complexity, known as entropy, and length. Entropy is dependent on the possibilities of the pass-phrase. If a user is limited to only numbers or 0 to 9, the difficulty of breaking it is much reduced. Normally, systems should allow pass-phrases consisting of the alphabet containing 26 upper- and 26 lowercase letters along with the ten digits and a variety of special chapters. The standard for keyboard input is the ASCII (American Standard Code for Information Interchange) table and consists of 128 possibilities. Key combinations with the Ctrl, Alt, and other keys can dramatically increase the number of possibilities. But because every input device does not have every character or modifying keys, even the standard ASCII list of 128 may be impractical.

The much more important factor is pass-phrase length. If the pass-phrase is limited to 26 letters (case being irrelevant), ten numbers, and four special characters (e.g., @, #, $, and %), that would be 40 characters. Therefore, using only one character, the number of possibilities is 40, and with two characters (40^2), the number of possibilities increases to 1600. Moreover, using six characters (40^6) dramatically increases the possibilities to over four billion. (Note that using repeating characters such as 111111 for a pass-phrase is more likely to be observed, and thus, systems will generally limit the number of repeated characters to two.) Similarly, if upper- and lowercase letters can be used

(52) and numbers (10) and eight special characters, one additional character increases the possibilities by a factor or 70 and even more substantially with additional characters.

Assume you have 70 possible characters and you are required to have an eight-character pass-phrase. That would be 70^8 or 576,480,100,000,000 possibilities. You could correctly guess the pass-phrase the first time or the last time, but on average it would require trying half of the possibilities or take 288,240,050,000,000 attempts – a brute force attack. If you try one every 10 s, it would take 48,040,008,333,333 min or 91,400,302 years to guess the pass-phrase on average for someone using a good pass-phrase. That seems like reasonable security risk but assumes the system will let you continuously attempt to input the pass-phrase without slowing down or stopping, an obvious implementation error. But more importantly, it is a good idea because we do not know how technology might change in the future.

Knowledge of a person can sometimes be used to guess a pass-phrase: the name of an owner's dog or other known or easily obtainable information is a poor choice, and their use perpetually hounds the lives of celebrities. Generic bad pass-phrase lists can be found on the Internet and invariably include the word password. The next and most likely mechanism to correctly guess the pass-phrase is by what is known as a dictionary attack. They involve trying all words and probably some simple derivatives also including foreign language words. That files exist purposely for dictionary attacks should be considered from a positive point of view; they emphasize the need for a good pass-phrase. Although many types of specific crypt-analysis attacks exist, they are relegated to longer texts, and the only remaining possibility for a pass-phrase is a brute force attack. Guidelines for choosing pass-phrases exist, but only systems which require complexity (at least one letter, number, and special character), define minimum length, and limit character repetition can ensure sufficiently secure pass-phrases are used. Often they allow pass-phrases to be reused after some number of changes or allow changing to very similar pass-phrases, another potential risk.

Secure systems do not allow continually trying pass-phrases which is a necessity but also problematic in itself. If a system only allows three attempts before blocking an account, a given account can be rendered useless if anyone can try three one-character pass-phrases to block it. The price for this type system must include the cost of an administrator required to frequently reset pass-phrases and both time and effort diverted from the user's efforts. Another method to deal with such problems requires the user to wait 5 min before he or she can again attempt to login, and only after a greater number of attempts is the account blocked requiring an administrator reset. Giving the user a method for self-resetting of pass-phrases is another option but is itself a security consideration.

The mechanism used by the system to process pass-phrases is often not known by users, they just assume it is recorded somewhere, and when the correct pass-phrase is input, the user gains the appropriate access. It should be obvious that if the user's pass-phrase is compared to a recorded version, then the pass-phrase must be available in some way which is necessarily a security issue. Thus, for security reasons, such a process is no longer used, and in fact generally pass-phrases should not be known by anyone except the user.

Instead of keeping the pass-phrase on the computer to compare it with input, a hash of the pass-phrase is saved. As the word hash implies, the pass-phrase is essentially chopped like the meat and potatoes combination with the same name. A computer hashing algorithm (used interchangeably with cipher) converts the input to a string of characters that cannot be reconstituted. It is a secure one-way function.

The hashing process is similar to the CRC or cyclic redundancy check that quickly checks to determine if a file is believed to be the one associated with the file name. However, CRC is not intended to be a secure check as with the pass-phrase hashes noted above. Nevertheless, it is extremely difficult to create an illegitimate file

Cryptology and Security, Table 4 Examples of text hashed with CRC-16 and SHA-1 algorithms

	Hashing Algorithm result	
Text (underline indicates differences)	**CRC-16**	**SHA-1**
Encrypting data is one way to keep it secure	AA1C	C799E6694ACA06AB82A390D5746645E1A2FE74F3
Encrypting data is one way to keep it secure	5225	ACC68E63A2986E8979E05FCEDB87E174A48D8AA7
Encrypting data is one way to keep it secure	6838	6284A2EC6C476A99C42C0AC838E443299D76A9CF
Encrypting data is one way to keep it Secure	CA1B	7D4D20D0272C2F3C359272EEA92076FE71FE23F8
Encrypting data is one way to keep it secure	E3AB	F3DF22162D6DD7FE589E0B996B349B9FC6DE5C9F

with the same CRC value also having the same name as what it is attempting to impersonate. This is what makes the CRC value such a useful tool to monitor file changes.

For the hashing algorithm (examples include MD-5, Whirlpool, and CRC-16) to work as a test of pass-phrase accuracy, the same pass-phrase must always result in the same hash value. Fortunately, system administrators need not know a pass-phrase for a specific user; they merely need access to change a pass-phrase. Changing the pass-phrase to something different and subsequently requiring the user to change it again within a few uses helps to ensure the system limits what system administrators can do when all such activity is recorded or logged. However, hackers routinely erase or modify logs in an attempt to hide their tracks.

Several examples of text that has been hashed give the reader an appreciation for the technique shown in Table 4. The examples are derivatives of the text “Encrypting data is one way to keep it secure.” Each line is hashed using two algorithms. First is the shorter CRC-16 (cyclic redundancy check) of 16-bit or 2-byte length and designed for quickly checking to see if a file has changed. Converting the larger text to only four characters of the resulting hash shows the power of a simple hashing algorithm. (When the results do not include letters above F, it indicates that characters are hexadecimal or base 16 which includes numbers 0–9 and letters A, B, C, D, E, and F.) For this simple example above, I used the free program Easy Hash v. 1.6 by Tomasz Kapusta to generate the hashes, but similar programs can be found on the Internet. The more complex SHA-1 is a widely used hash algorithm and has a 20-byte or 160-bit length equivalent to 40 characters. It was designed for the US National Security Agency in 1995. Although each line in the example only differs from the first line by an extra space or the case of one letter, the result of the hash is dramatically different. SHA-1 received extensive testing before being selected by the NSA, but, like all widely used algorithms, eventually attacks will reveal weaknesses, and this will soon be replaced by a more secure standard.

Unfortunately, users can create situations that require a correct pass-phrase to be used without the ability to simply change it. An example of this issue is when a drive, folder, or file is securely encrypted and does not come under the control of an administrator and the person in question has lost/forgotten the pass-phrase or is no longer available. If the pass-phrase is cryptographically sound, then the protected entity essentially no longer exists because breaking that pass-phrase maybe virtually impossible. As encryption of data becomes more popular with individuals, important family information is put into jeopardy if that person has not given the pass-phrase or key to their lawyer or other family member.

A more secure way of protecting critically important information is to use a technique whereby multiple users are required to find the correct pass-phrase. This can be envisaged using an X, Y grid. Assume the correct pass-phrase is located at zero on the Y axis and is equivalent to some number X. For simplicity we will use the pass-phrase at 10, 0 (X, Y) on the grid. A straight line can be drawn through that point and through points 11, 1 and 8, −2. Knowing any two points anywhere along the line will define the line and thus define the pass-phrase as 10 because that is where the line passes through the Y axis. Ten different people could be given different X–Y coordinates, and any two can determine the pass-phrase, but one person alone could not define the Y axis location. In a similar fashion, use of other mathematical functions can require three or more sets of coordinates to define where the line crosses the Y axis. Thus, collusion or compromising one point does not immediately compromise the pass-phrase, but its complexity does not warrant use for anything except critical data. However, whoever controls the system is therefore another security consideration.

Given the limitation of some number of characters for a pass-phrase, it is possible to generate all the possibilities for a given hashing algorithm, saved as a "rainbow table." These are available on the Internet but are large, in the range of a terabyte or more. Thus, if one can obtain the hashed pass-phrase, locating it in the table will produce the originating pass-phrase. This works because a comparison of results of hashing programs should generate the same hash although in use, and unbeknownst to users, the result can purposely be different. Just as adding salt to food hash will change it, adding a salt (an additional factor) to a hashing algorithm changes its result as well. For example, simply appending an additional single character to the end of each pass-phrase will change the hashed result dramatically, as implied above and thus rendering a hash table useless. Because the same character is added each time to each pass-phrase, the resulting hash will be the same for each specific pass-phrase but will be different from the non-salted pass-phrase.

Probably the most simplistic and yet elegant operation in computer-based cryptology is the Exclusive Or Logical Operator also known as XOR (pronounced X-Or). It is symmetrical, meaning it works the same for both encryption and decryption. Remembering that computers deal with 0 s and 1 s, it is obvious that they work well with logic table based on true and false (Table 5). Given the result, this is a deceptively simple operation. The message (line 1), a string of 0 s and 1 s, is XORed with the key (line 2) also a string of 0 and 1 s. The key and the encrypted message are lined up, and for each bit, if both are 0, or both are 1, the resulting encrypted bit (line 4) is 0. Conversely, if they are different, the unencrypted bit is a 1.

Cryptology and Security, Table 5 Encryption and decryption using the exclusive or logical operator

1	Message	0	1	0	0	0	0	0
2	Key	1	1	0	1	0	1	1
3	XOR							
4	Encrypted	1	0	0	1	0	1	1
5	Key	1	1	0	1	0	1	1
6	XOR							
7	Unencrypted	0	1	0	0	0	0	0

Un-encrypting the encrypted message works the same way. The result is the original message (line 7). This is a very fast operation with security dependent on the key. A short simple key of eight bits (256 possibilities) can be broken easily, while a non-repetitive random key is essentially the one-time pad and cannot be broken when used correctly. Therefore, we continually increase the key length for secure functions with 256 bits (1.15×10^{77} possibilities) on the horizon and even longer keys available as that eventuality is expected.

Most of the above consists of cryptographic basics and how they may be used. Many operations are chained together to increase the difficulty of cryptanalysis. Much of what is used consists of the two major types of modern ciphers which include stream and block types. The stream cipher may be considered similar to the XOR system above and the one-time pad. It encrypts and decrypts the message bit by bit but is very fast. Unlike the one-time pad which requires a key as long as the actual message, the stream cipher uses a key stream generated with a much smaller key. An example of a stream cipher is RC4.

In contrast to the stream cipher, the block cipher, as the name implies, processes the message blockwise or equally sized chunks. Over time the block size has been increasing to stay ahead of more capable cryptanalysis techniques and computer advances. The minimum length has moved to 128 bits but will continue to increase. Actual use may be similar to the polyalphabetic substitution noted above, but as expected it is not that simple.

Block ciphers might be compared to a chain knot in that each subsequent knot is dependent on what goes before it and is therefore unlike the stream cipher in that a single missed bit might not obfuscate the whole message. Security is enhanced by repeating the simple function known as a round. Multiple rounds and other manipulations the security. Recognizing the need for such a cipher for nonmilitary or official uses led the US government to release a standard, the Digital Encryption Standard in 1976. With only a 65-bit key length, it was considered obsolete no later than 1998 when it was broken, but a variant – Triple DES – was used in its place. A competition resulted in the Advanced Encryption Standard or AES being released in 2001 after several years of testing. Other well-known block ciphers include Blowfish, Serpent, and Twofish.

Both stream and block ciphers are symmetrical, meaning they are used for both encryption and decryption. Cryptography had only been symmetrical until the 1970s when asymmetrical ciphers were first created. Conceptually, this would be more secure because previously the key was the same in both directions, and therefore, more than one person would need to know the key.

One inherent security problem that existed was the inability to protect a symmetric key. This was finally solved using a key pair or what is known as public key cryptography (Diffie and Hellman 1976). Initially, its strength relied on the difficulty of factoring prime numbers of very large numbers and more recently elliptic curves. A key pair is generated with one used as a private key and the other a public key. The significance of the keys is simply that at least one is kept secret and which is irrelevant. The most important consideration is that a cryptographically secure key needs to be much larger than others mentioned above, 1,024 bits or 128 bytes (representing 2^{1024} or 1.8×10^{308} possibilities) which makes it a noticeably slower, on the order of a thousand times slower. But the additional time and key length is used to encrypt a key which is subsequently used in the much faster stream cipher above to encrypt or decrypt the message.

Consider a simple scenario with email; someone sends you a message which is encrypted with your public encryption key which might be posted on the Internet for anyone to use. Note the similarities between this example and the locked box scenario above. You in turn decrypt it with your personal key. Because a different key is needed to decrypt the message, anyone could send a message to you, but only you, by virtue of the fact that you have the private key, can decrypt it. Although this sounds of little use by itself, consider that the original message is encrypted using the sender's personal key and also encrypted using your public key. The result is a message that could only be sent by the originator because it is encrypted with their private key (you decrypt it with their public key) and subsequently only you can decrypt it with your private key because it was also encrypted with your public key. While governments and corporations may question its use, consider that the vast majority of email is unwanted spam and that public/private key encryption could dramatically reduce unwanted email. It could significantly impact both network bandwidth concerns and infrastructure costs but is unlikely to become a standard.

The above represents a very limited introduction to cryptology without considering possible future developments. A new technology like quantum computing having the ability to much more quickly break codes makes constant vigilance a necessity for cryptographic security professionals. Ramifications of these or other new technologies make cryptographic security always a work in progress. But no matter how much we focus on the technology, that cannot be the end or as Bruce Schneier (2004) put it "I've realized that the fundamental problems in security are no longer about the technology; they're about how to use the technology."

Related Entries

- ▶ Automated and Manual Forensic Examinations
- ▶ Computer Forensics

Recommended Reading and References

Al-Kadi IA (1992) The origins of cryptology. The Arab contributions. Cryptologia 16(2):97–126

Callery S (2008) Codes and Ciphers: all the tips and tricks you need to make them and break them. Harper Collins Publishers, NY, NY.

Diffie W, Hellman ME (1976) New directions in cryptography. IEEE Trans Inf Theory 22(6):644–654

Garber L (2011,October) News briefs. Hackers steal $13 million in one day via online security breach. Computer p 15. IEEE Computer Society

Kahn D (1996) The code-breakers: the comprehensive history of secret communications from ancient times to the internet. Revised edn. Scribner, NY, NY.

Menezes A, von Oorschot P, Vanstone S (1996) Electronic format 1997 (www.cacr.math.uwaterloo.cc/hac). Handbook of applied cryptography. CRC Press

Schneier B (1996) Applied cryptography: protocols, algorithms, and source code in C, 2nd edn. Wiley, New York

Schneier B (2004) Secrets & lies: digital security in a networked world. Wiley, New York, p. 398

Singh S (2000) The code book: the science of secrecy from ancient Egypt to quantum cryptography. Anchor Books, Random House

CTC

▶ Communities That Care

CTC Prevention Coalition

▶ Communities That Care

CTC Prevention System

▶ Communities That Care

Cultural Criminology

Jeff Ferrell
Texas Christian University, Fort Worth, TX, USA
University of Kent, Canterbury, UK

Synonyms

Labeling and Deviance; Labeling: History and Concept; Modern Marxist and Radical Theory; New Media and Crime Images; Postmodern Criminology; Reality Based Television Shows

Overview

Cultural criminology is designed to place issues of meaning, symbolism, and interpretation at the forefront of the criminological enterprise. Cultural criminologists argue that both crime and crime control operate as cultural endeavors, with their personal and social consequences constructed out of contested processes of collective representation and interpretation. Given this, criminological inquiry must encompass more than crime and criminal justice as conventionally understood; it must also focus on transgressive subcultural practices, media representations of crime and victimization, public displays of anger and elation, and public performances of justice and injustice. In order to fully theorize such phenomena, cultural criminologists contend, criminology must in turn resurrect older traditions of cultural and interactionist analysis while also drawing on newer perspectives from visual theory, cultural geography, and related fields. In this way cultural criminology offers a critique of conventional criminology and its circumscribed range of theories and methods and at the same time attempts to reconstruct criminology as a more expansive, interdisciplinary, and intercultural undertaking. Cultural criminology likewise attempts to reorient criminological analysis to the distinctive social and cultural dynamics of the present late modern period and so to enhance criminology's ability to penetrate the contested global politics of crime and crime control as they are now developing.

Fundamentals of Cultural Criminology

While cultural criminology has emerged as a distinct theoretical and methodological orientation within criminology only during the past two decades, the intellectual roots of cultural criminology can be found amidst earlier perspectives within sociology and criminology. For cultural criminologists, the goal has been to reclaim these perspectives as essential components of criminological inquiry, to reimagine and reinvigorate them in the context of contemporary crime and crime control trajectories, and to enliven them with new perspectives from other disciplines.

Intellectual History

At its core, cultural criminology constitutes a synthesis of two theoretical orientations: the first North American and the second British. The North American orientation dates to the mid-twentieth century and the emergence of symbolic interactionism and labeling theory in the analysis of deviance and crime. For interactionist and labeling theorists, the appropriate subject matter of criminology was not crime as an individual act, but rather the web of social interactions and interpretations by which any act might or might not be constructed as crime. At issue was the meaning of crime and the exercises of power and discrimination by which that meaning was determined and redetermined within an ongoing social process. As regards crime and crime control, the "symbolic" dimensions of symbolic interaction suggested that a transgressive act could be made to symbolize heroism or horror, self-defense, or assault; likewise, the practice of crime control could be presented as a matter of good policing, patriotic participation, or political repression. For these theorists, the reality of crime was inherently a social and cultural reality – and a political reality as well. By understanding the symbolic construction of crime and crime control, criminologists could begin to penetrate the proselytizing claims of moral entrepreneurs, the discriminatory dynamics of the criminal justice system, and they lived experiences of those labeled criminal. These lived experiences could in turn be documented through "naturalistic" case studies and careful ethnographies that revealed patterns of shared meaning and symbolic communication within marginalized subcultures. Here, then, was one foundation of cultural criminology: an attentiveness to symbolism and meaning in the construction of crime and a critical analysis of crime, culture, and power.

Influenced by the work of the North American interactionists, sociologists and criminologists associated with the National Deviancy Conference, the Birmingham School of Cultural Studies and the "new criminology" in Great Britain began in the 1970s to develop what would become cultural criminology's second intellectual foundation. Sharpening the critical potential of interactionist and labeling approaches, they carefully investigated the cultural avenues through which power was by turns enforced and resisted. Situating crime and crime control in historical and political context, they likewise analyzed the shifting ideological utility of crime, exposing the ways in which crime concerns and anticrime campaigns were made to fit within larger political agendas and historical trajectories. These researchers also engaged in the sort of case study and ethnographic research developed among interactionist criminologists, finding amidst the leisure worlds and illicit subcultures of their subjects moments of stylized defiance and collective resistance to legal authority. Taken as a whole, this work not only continued the interactionists' reconceptualization of crime as a social and cultural construction; it also began to reconceptualize the very nature of power and social control in contemporary society. Power in many ways now denoted the power to control the meaning of public events and social control the ability to control the terms by which the activities of marginal groups came to be understood and interpreted.

In the 1990s these two orientations were synthesized for the first time into a distinct "cultural criminology" that explicitly drew on both the North American and British traditions in an attempt to theorize contemporary intersections of crime and culture. To supplement and expand these earlier approaches, cultural criminologists now included a variety of additional disciplinary and interdisciplinary perspectives: anarchist criminology and other critical criminologies, postmodern theory, cultural geography, and others. Utilizing this synthetic approach, they focused especially on the components of style, language, and symbolic meaning that animate illicit social worlds and on the ways in which legal and political authorities in turn utilize mediated representation to criminalize such worlds (Ferrell and Sanders 1995). Cultural criminologists also sought to revitalize the tradition of ethnographic research into illicit subcultures, both by theorizing the epistemic and emotional dynamics of this approach in distinction to quantitative research methods and by engaging in in-depth ethnographic explorations of graffiti writers, neo-Nazi skinheads, street-level sex workers, and other groups.

Current Theoretical Orientations

Contemporary cultural criminology has spawned a variety of theoretical models that build from earlier work on the situated meanings and symbolic politics of crime and crime control. A number of these models explore both the immediate, sensual experience of crime and the dynamics that link such illicit experience to larger patterns of cultural meaning and historical change. The concept of *edgework*, for example, is utilized by cultural criminologists to analyze incidents of extreme, risky behavior that are generally seen as deviant or illegal (Lyng 2005); cultural criminologists have employed this model in investigating graffiti writing, street racing, BASE jumping, anorexia, sadomasochistic sexual practices, and other contemporary phenomena. In contrast to conventional understandings of such practices as simply destructive or individually out of control, cultural criminologists argue that these practices often involve a conscious exploration of the "edge" between chaos and order, and danger and self-determination. Essential to this exploration is the way in which such practices blend high levels of risk with the practiced skills necessary to negotiate such risk; for participants, great risk requires great skill, and the greater the skill, the greater the risk that can be taken. This dynamic of risk and skill in turn appeals to more and more people precisely because both risk and skill are increasingly expunged from contemporary life; low-skill, service sector jobs predominate in the emerging consumer economies of late capitalism, and whether at work or play, "risk management" strategies now regulate most every aspect of social life. If in this way the concept of edgework links illicit experiences with larger social and economic changes, it also helps to explain a particular crime control problem. Cultural criminologists have found repeatedly that edgework produces for participants an alluring "adrenalin rush" out of its mixture of extreme risk and practiced skill – but because of this, crime control strategies meant to stop such activities often serve only to heighten the risks and sharpen the skills involved and so to enhance the appeal of edgework activities for those engaged in them.

This edgework model resonates with the broader notion of the *seductions of crime* (Katz 1988). From the view of cultural criminology, much traditional criminological theory has been constituted back to front; that is, such work has theorized the background factors that lead up to the criminal event but has had little to say about the immediate dynamics of the criminal event itself. Cultural criminologists argue that this immediate "foreground" of crime, animated as it is with interactional exchanges and powerful emotions, is indeed worthy of analytic attention. Within it, participants can come to be "seduced" by the momentary meanings and emotions that emerge; within the criminal event, issues of stigma, honor, and respect can become a powerful, if fleeting, impetus for violent, exploitative, or courageous behavior. As with the model of edgework, the notion of crime's seductions suggests that crime cannot be reduced to simple, individualistic explanations; instead, moments of criminality often harbor a host of complex, negotiated meanings that both reflect and reconstruct larger cultural forces. Moreover, these meanings may drive the very dynamics by which a criminal event is resolved; more than physical violence or monetary gain, they are often what matter most to those involved.

A similar intertwining of meaning, emotion, and culture emerges with the cultural criminological theory of the *carnival and crime;* here the frame of reference is also historical and comparative (Presdee 2000). Cultural criminologists note that, historically, carnival has served in many societies as a well-defined time of vulgarity, ridicule, and ritualized excess. Because it was predefined and regulated, though, carnival functioned both to celebrate and encourage misbehavior and to contain it within social and cultural boundaries; in this sense, carnival maintained the sort of delicate balance between dangerous human desires and defined social roles essential to the vitality of the societies in which it flourished. In contemporary Western societies, though, the delicate balance of carnival has largely been upended, in some cases through the criminalization of carnivalesque activities and in other cases through the legal and commercial cooptation of carnival into commodified experience and abstract spectacle. As a result, some aspects of carnival's historical practice can now be bought and consumed, in the form of explicit pornography or degrading "reality television" programming. Other aspects are today enacted as crime – drug taking, sexual predation, and "joyriding" in stolen automobiles – but now all the more dangerously because freed from their containment within community ritual.

As with the models of edgework and the seductions of crime, the model of carnival in this way embodies another key concept in cultural criminology: the idea of *transgression*. From the view of cultural criminology, there is often more at stake in crime and crime control than simple law-breaking and the legal response to it. Instead, crime seems often to embody a breaching of social and cultural boundaries, a dynamic of crossing over or breaking through – and if those boundaries to be breached encode dominant economic arrangements or long-standing cultural norms, then there is indeed much more at issue than simple law-breaking. Yet, as the model of carnival shows, transgression also involves a moving dialectic between boundaries made and boundaries breached; in this sense, transgression constitutes an ongoing social and cultural process, a series of negotiated actions and reactions, more than a single act. In addition, as the edgework model suggests, those engaged in transgressive activities sometimes cross the boundaries of safety or respectability with an emerging intentionality; they begin to understand the world and their place in it differently precisely because of the boundaries they breach and so acquire from transgression a critical lens for seeing anew those arrangements that engender it.

Two final theoretical approaches incorporate the historical perspective of the carnival of crime model but with an eye in particular toward the distinctive dynamics of the present, late modern period. The theory of *exclusion/inclusion* posits that contemporary society suffers from a particularly problematic contradiction (Young 1999). On the one hand, the social order is defined by increasing economic and political inequality and with this the increasing economic and legal exclusion of larger and larger segments of the population from full membership in mainstream society. Ongoing financial and housing crises, high levels of unemployment and homelessness, the prevalence of low-wage and temporary work among those who remain employed, and mass incarceration and systematic legal disenfranchisement in the United States and elsewhere all serve to exclude many among the poor, ethnic minorities and the formerly middle class from full participation in conventional social worlds. At the same time, though, these and other groups are, by design, culturally included. The power of the mass media, and especially mass advertising, is such that these groups are systematically taught to desire the same consumer goods as others. Moreover, like others in late capitalist society, they are encouraged to see these goods as essential markers of lifestyle success and to define their own status and identity by goods and services consumed. The result of this contradiction is a culturally induced epidemic of resentment, insecurity, and humiliated desire and with this a parallel explosion in economic crimes of acquisition and expressive crimes of retaliation and frustration. With this model of exclusion and inclusion, then, cultural criminologists attempt to build from Merton's foundational formulation of socially induced strain and adaptations to it while at the same time reimagining it within the contours of late capitalist culture and economy.

The model of *media loops and spirals* likewise examines the distinctive cultural dynamics of late capitalist/late modern society, with a focus especially on the emerging interplay of media, crime, and criminal justice (Ferrell et al. 2008). According to this model, the links between crime, criminal justice, and the media in contemporary society are neither linear nor straightforward; for criminologists, therefore, it is no longer a matter of simply inquiring into how accurately the media reports on a particular type of crime, for example, or to what degree individuals imitate crimes that they see portrayed in the media. Instead, the saturation of everyday life with media-producing technology, and the nonstop consumption of mediated images and information, is such that distinctions between an event and its mediated representation often collapse. Criminologists seeking to understand crime and media, then, are increasingly confronted by a self-perpetuating interplay of criminal acts, mediated representations, and mediated responses – by situations, that is, where crime and its image circle back on one another – and so must consider the nature of media loops and spirals if they are to understand the contemporary construction of crime and criminal justice. As recent research in cultural criminology has shown, graffiti practitioners blend illegal street graffiti with sophisticated, legal media productions; gang members stage violent assaults so as to record and sell copies of them on the Internet; "reality television" shows set up violent confrontations and at times entrap their participants in assaults and arrest; and police officers utilize Facebook postings to track suspects, alter their street enforcement strategies due to their own police car cameras, and employ televised policing shows as points of reference in their own street policing. Moreover, these looping entanglements of crime, criminal justice, and media regularly reproduce themselves over time and so become ongoing spirals of image and activity. An initial cell phone image of a criminal act, for example, may over time become an Internet posting, a piece of legal evidence, a segment in a video compilation to be viewed or purchased, and an image within a crime segment on a televised news program; similarly, a police car video may later be subpoenaed as part of legal proceedings, utilized for police training, or compiled into a "world's wildest police chases" video. In such a world, cultural criminologists argue, criminological analysis must appreciate the power of mediated representation and must be sensitive to the ways in which such representation feeds back into the practice of crime and criminal justice.

Methods of Cultural Criminology

The research methods utilized by cultural criminologists embody the theoretical orientations that define cultural criminology itself. Given their analytic focus on shared meaning and emotion, cultural criminologists require methods that can take them inside the illicit situations where meaning and emotion are negotiated and that can attune them to the subtleties of symbolic interaction among people and groups. To understand the allure of edgework or the seductions of criminal moments, cultural criminologists argue, they must find ways to be present when such seduction is underway and to share as best possible in this seductive emotional process. Likewise, the focus on symbolism and mediated representation means that cultural criminologists must find ways to decode the complex symbolic universes of marginal groups and to penetrate the processes by which mediated representations are made and circulated. In all of this, cultural criminologists argue for a general "methodology of attentiveness" – that is, a methodological orientation that enables close attention to the nuances of human culture and human transgression. Broadly speaking, cultural criminologists in this sense prefer methods that embody something of the phenomenological imperative – the imperative to encounter "the thing itself," on its own terms and in situ – rather than methods that spawn artificial circumstances, broad generalizations, and abstract conclusions.

Because of this, cultural criminology's methodological orientation has become also a critique of conventional methods within mainstream criminology; in arguing for and employing their own attentive methods, cultural criminologists have also emerged as perhaps the most vocal critics of contemporary methodological practices in criminology (Ferrell et al. 2008). From the view of cultural criminology, the most widely used methods in criminology – survey research, statistical analysis of survey results, statistical analysis of governmental data, and the massive data sets that result – primarily serve to distance criminology from its subject of study. By their own design and internal logic, such approaches preclude on-the-ground engagement with situations of crime and crime control and with the meanings and emotions that emerge within these situations. Likewise, such methods systematically miss the subtleties of symbolism and style that circulate within illicit worlds, as they reduce these subtleties to a set of simplistic choices among survey answers or governmental record-keeping categories. As such, these methods in turn reduce the human identities and human interactions that construct the reality of crime and justice to statistical abstractions, all while removing researchers and their own identities from creative engagement with the research process. Most damningly, cultural criminologists suspect that such methods are perhaps so widely and popularly employed precisely because they do keep criminological inquiry at a distance, forfeiting critical engagement with the subject matter in the interest of political leverage and agency funding and thereby reproducing stereotypes and assumptions more than investigating them.

Building from the pioneering field research of the earlier North American and British theorists whose work set the course for cultural criminology, cultural criminologists turn to ethnography – long-term, in-depth field research with subjects of study – in place of more conventional methods of survey research and statistical analysis. Through deep immersion in the lives and interactions of criminals, crime victims, and crime control officials, cultural criminologists seek to become part of the process by which such people and groups construct the various meanings of crime. Likewise, cultural criminologists seek through ethnographic research to attune themselves to the shared languages, symbolic codes, and public performances that emerge within and between groups as they work to make sense of crime and crime control. This sort of ethnographic research embodies another, more controversial dimension as well. As already glimpsed in the models of edgework and the seductions of crime, cultural criminologists argue that the emotional dynamics of crime are often essential to the meaning and experience of it – and because of this, they argue, ethnographic researchers should also seek a degree of emotional intimacy with subjects and situations of study. This notion of subjective or appreciative understanding is embodied in Max Weber's classic sociological concept of *verstehen* and specifically in what cultural criminologists call *criminological verstehen* (Ferrell and Hamm 1998). With this concept in mind, cultural criminologists have in recent ethnographies analyzed, for example, their own emotional participation in the "adrenalin rush" of graffiti writing, the macho pleasures of high-powered weapons use among off-duty police officers, the stark alienation of right-wing terrorist groups, and the dismissive stigmatization of phone sex workers and street prostitutes. Significantly, the controversial nature of this research stems not only from its deep immersion in the lives and emotions of illicit groups but from its reversal of the "objective" and "value-free" underpinnings of more conventional forms of social scientific research. Here it is not the "objectivity" of surveys and statistical analysis that produces research insights, but the opposite: the emotional subjectivity achieved through ethnographic research.

Cultural criminologists have in addition sought to reimagine and reinvent ethnographic research in light of current theoretical models and in the context of contemporary social and cultural developments. If the situated seductions of crime can emerge within a criminal event and evaporate with its termination, for example, then perhaps ethnography must be organized not only around long-term researcher participation but researcher attentiveness to immediate situational dynamics. Cultural criminologists refer to this approach as *instant ethnography* – ethnographic research that delves deeply into moments and situations, deconstructing them in search of their larger implications and understanding them as instances in which crime and its meaning are performed. Likewise, given the fluid, dislocated nature of many social interactions and identities in the late modern world, cultural criminologists argue that ethnography's traditional focus on single, unitary social groups must now be supplemented by forms of *liquid ethnography.* This sort of ethnography pays special attention to issues of migration and movement, to the ways in which seemingly solid social identities are made and remade, and to the interplay of mediated images in the ongoing construction of individual and group perceptions. Finally, as suggested by the notion of *verstehen* and by cultural criminologists' analyses of their own emotional involvement with crime issues, *autoethnography* has also emerged as a viable research method within cultural criminology. With autoethnography, the ethnographer's own self and identity become a focus of study – but not as form of isolated self-examination. Instead, autoethnography demands that researchers account for the ways in which their participation in illicit groups or dangerous situations has altered their own sense of self or status or otherwise forced them to confront previously unexamined privileges and assumptions. In this way, autoethnography is designed to create a fuller ethnographic account of crime and crime control, one in which the ethnographer's own experiences become yet another window into larger social and cultural patterns.

Paralleling the distinction between survey research/statistical analysis and ethnographic inquiry is the distinction between conventional approaches to research on media and crime and those approaches taken by cultural criminologists. Traditionally, criminologists have utilized content analysis – the measuring of static, discrete content categories within media texts – in studying, for example, the amount of media coverage given to a particular type of crime or trends in the amount of violence in popular films. In the context of contemporary media saturation and in light of the model of media loops and spirals already seen, though, cultural criminologists find that this approach is often unable to capture the ongoing, fluid interplay of contemporary crime and media. Calculations of media content and its amount miss the broader cultural and aesthetic frames of reference within which such content takes on meaning. In addition, utilizing content analysis in an attempt to measure the degree of divergence between the "real" nature of crime and its "biased" media representation ignores the spiraling process by which the "real" and the representational increasingly come to constitute one another.

To supplement conventional content analysis, cultural criminologists utilize two types of alternative methods. *Ethnographic content analysis* conceptualizes mediated representation as an unfolding process of decision making, action, and reaction among a variety of groups and so undertakes the study of media and crime from the view of researcher immersion in this ongoing process. The approach likewise advocates an ongoing interplay between researchers and the media texts they study; the method is meant to provoke deep involvement with media texts, such that researchers over time develop thoroughgoing accounts of texts and their emergent meanings. In this way ethnographic content analysis allows researchers to understand media texts not as isolated embodiments of content but as emergent cultural process incorporating a variety of political and cultural dynamics. While this method enables researchers to identify and analyze textual patterns, as with content analysis, it also accounts for the liquid, looping process that emerges between crime, media, and justice. A second alternative approach links crime and media analysis even more closely to the overall practice of ethnography. Here, in the spirit of liquid ethnography, researchers engage in fieldwork with criminals, criminal justice workers, and crime victims as these groups engage with the mass media, investigating the ways in which such groups attempt to manage images and accounts of their lives and experiences or, increasingly, undertake to produce their own mediated accounts of crime and justice. Similarly, cultural criminologists at times conduct ethnographic research inside media organizations, so as to be present as decisions are made and accounts developed regarding the representation of crime and justice issues.

Substantive Themes and Research Trajectories

As cultural criminology has matured over the past two decades, its theories and methods have been applied to a wider and wider range of phenomena. Certainly research is still being conducted on subcultural interactions and aesthetics, popular culture representations of crime and violence, and the seductive dynamics of edgework and other forms of situated transgression. In addition, though, cultural criminologists are both refining their focus on particular forms of crime and crime control and widening their analysis of contemporary social and cultural changes as regards crime and social control.

One emerging focus within cultural criminology is the relationship between crime, crime control, and everyday life (Ferrell et al. 2008). Cultural criminologists note that the nature of law and the criminalization process is such that the magnitude and meaning of a criminal activity remains under construction – and that because of this, seemingly inconsequential or invisible crimes often merit as much investigation as those defined at any one time as consequential. In addition, cultural criminologists point out that contemporary legal and social control is increasingly encoded in the small situations of everyday life. Pervasive surveillance/CCTV cameras in urban areas, CPTED (crime prevention through environmental design) programs in parks and train stations, corporate risk-prevention strategies, computerized data and web monitoring, cell phone tracking devices – these and other phenomena build often unnoticed ideologies of legal and cultural control into the practices of daily life. Likewise, popular contemporary policing approaches like the "broken windows" model focus police attention on everyday situations and low-order criminality, though generally with little reflection on the assumptions underlying such approaches. In light of all this, cultural criminologists argue that if we are to understand the meanings of crime and the ideologies of crime control, we must investigate their presence in everyday life; that is, we must investigate the dynamics of routine traffic stops, the layout of city centers, the daily policing of the homeless, and the prevalence of everyday crimes like shoplifting or loitering. Such research provides added benefits as well; it serves as a corrective to the media focus on sensational and unusual crime, and it encourages scholars and students alike to engage in accessible criminological inquiry.

This mention of city centers and urban CCTV cameras highlights a second substantive focus within cultural criminology: the city and its spatial arrangements (Hayward 2004). Cultural criminologists note that, with the dramatic and increasing urbanization of the world over the past few decades, many conventional criminological analyses of crime and crime control may now need to be reconceptualized as analyses of distinctly urban patterns of crime and control. Drawing especially on work in cultural geography, cultural criminologists have, for example, investigated the ways in which new urban economies organized around high-end consumption reorganize the city's spatial arrangements while also fostering new models for policing these consumerist spaces. Cultural criminologists have also explored these spaces as sites of legal and cultural conflicts that pit recent arrivals against long-time residents, corporate developers against public space activists, and automotive traffic patterns against advocates of alternative transportation – with these conflicts in turn entangled with ongoing mediated representations, arrests, and court proceedings. Recent ethnographic work in cultural criminology has likewise explored the cultural practices of illicit urban subcultures like those of skateboarders, graffiti writers, and homeless trash scroungers, with special attention to the process by which such practices remake urban space, imbuing it with illicit meaning and remapping it along the lines of subcultural status and values.

Two additional trajectories of research attempt to engage cultural criminology directly with the distinctive political and economic arrangements, and political and economic crises, that define the contemporary late modern/late capitalist period. The first of these is a developing cultural criminology of the state and of global political conflict. Recent work in cultural criminology has explored, for example, the increasingly mediated nature of military recruitment, military training, and military warfare itself; it has also shown that the boundary separating such mediated militarism from the worlds of popular culture and entertainment is becoming blurred at best. Similarly, cultural criminologists have documented the activities of militant or terrorist groups in self-producing sophisticated media and have shown how these and other images of war and violence loop and spiral through the global mediasphere. Linking these sorts of insights to the more traditional concerns of criminology, other recent research has explored the culture of punishment in the United States and other countries – a culture which flows in and out of the prison and into popular media, public perceptions of crime and justice, and even the tourism and entertainment industries (Brown 2009). Similarly, research into gangs and gang culture has documented the role of law enforcement agencies in defining perceptions of gang threat and in moving gang members and their culture back and forth across national boundaries; it has in turn documented gang members' attempts at self-empowerment and resistance to these state practices.

A second trajectory has cast cultural criminology into the broader cultural currents of late modernity. Building in part from the theory of exclusion/inclusion, cultural criminologists have found a complex of values, practices, and identities that coalesce around drift, dislocation, and the particular forms of transgression that accompany them. In this shifting world, status and identity are increasingly cut loose from traditional anchors of steady career or home-based stability; indeed, the notion of "home" itself loses its conventional comfort as more and more people migrate from city to city or country to country in search of something other than part-time work and political insecurity. Especially in light of the ongoing global economic crisis, aspirations of upward achievement are often replaced by a desperate sense of surviving with what is at hand or with what might come to be at hand – yet advertisers continue to aggressively market identities that, while made desirable, are neither affordable nor achievable. In such a world, cultural criminologists argue, insecurity and uncertainty become a way of life, and in response, some people seek to find identities in acts of individual risk-taking and transgression; others engage in expressive crimes fueled by resentment, fear, and failure. In such a world crime control also changes, becoming increasingly a matter of risk management and the containment of transitory populations. Here is an emerging global milieu permeated not only by inequality and injustice, exclusion and inclusion, but by a seemingly intractable culture of ambiguity and uncertainty; to make sense of it, cultural criminologists argue, criminology may itself have to become comfortable with ambiguous findings, liquid analysis, and alternative forms of inquiry (Young 2007; Ferrell et al. 2008).

In this context cultural criminology has continued to build from its original cross-cultural roots in North American and British criminology, with significant bodies of theoretical and substantive work now coming out of the Netherlands, Australia, Brazil, and other countries. In keeping with the early intellectual interplay between North American and British scholars, this contemporary work is not only geographically dispersed in its origins and substantive concerns but intentionally conversational, with cultural criminologists from around the world sharing ideas and working toward new intellectual syntheses (Carvalho et al. 2011) that can account for emerging social circumstances. Likewise, cultural criminologists continue to challenge orthodox approaches within criminology and to develop new means of engaging in criminological analysis. Drawing on film and visual theory and revisiting long-standing traditions of documentary photography, much contemporary work in cultural criminology is oriented toward both analyzing existing images of crime and producing an alternative body of informed visual representation as regards crime and justice (Hayward and Presdee 2010). Other cultural criminologists continue to experiment with the narrative forms of criminology, producing in place of conventional criminological discourse a range of manifestos, vignettes, poems, and short stories in both print and digital media. In all of this, the intent is to bring cultural criminology solidly into the present and the future and to position it for ongoing, critical engagement with the cultural dynamics of crime and crime control.

Recommended Reading and References

Brown M (2009) The culture of punishment. New York University Press, New York

Carvalho S, Neto MP, Mayora M, Linck JAG (2011) Criminologia cultural e rock. Sao Paulo, Lumen Juris

Ferrell J (1999) Cultural criminology. Ann Rev Sociol 25:395–418

Ferrell J, Hamm M (eds) (1998) Ethnography at the edge. Northeastern University Press, Boston

Ferrell J, Hayward K (eds) (2011) Cultural criminology: theories of crime. Ashgate, Farnham

Ferrell J, Hayward K, Morrison W, Presdee M (eds) (2004) Cultural criminology unleashed. Glasshouse, London

Ferrell J, Hayward K, Young J (2008) Cultural criminology: an invitation. Sage, London

Ferrell J, Sanders CR (eds) (1995) Cultural criminology. Northeastern University Press, Boston

Hayward K (2004) City limits: crime, consumer culture, and the urban experience. Glasshouse, London

Hayward K, Presdee M (eds) (2010) Framing crime: cultural criminology and the image. Routledge, London

Katz J (1988) Seductions of crime. Basic Books, New York

Lyng S (ed) (2005) Edgework. Routledge, New York

Presdee M (2000) Cultural criminology and the carnival of crime. Routledge, London

Young J (1999) The exclusive society. Sage, London

Young J (2007) The vertigo of late modernity. Sage, London

Customary Justice

► Post-Conflict Traditional Justice

D

Decision Making Theory

▶ Integrating Rational Choice and Other Theories

Defiance and Motivational Postures

Valerie Braithwaite
Regulatory Institutions Network (RegNet),
The Australian National University,
Acton, ACT, Australia

Overview

Motivational postures are the signals that people send to authorities, including criminal justice authorities, to indicate their liking for that authority and their willingness to defer to the authority's rules and processes. These signals change in response to the actions of authority. Five motivational postures describe the way in which individuals and groups position themselves in relation to authority. The posture of commitment represents belief in the authority, its goals and purpose. The posture of capitulation involves acquiescing to authority because it brings the least trouble. Resistance is a posture of protest and anger about how an authority operates. Disengagement is a posture of withdrawal, of severing the relationship with authority to the point where the authority is irrelevant. Game playing challenges authority by circumventing rules and laws while appearing to do what is expected. These postures are openly shared and coexist. They combine to form a complex signaling system that authorities can read and respond to in an emotionally intelligent way to form a more effective criminal justice system.

Fundamentals

What Are Motivational Postures?

When authorities take action to create order or enforce law, individuals display a range of responses. The response of prime concern in criminology is whether people obey the law. Yet other responses also play a role in the success of an authority in crime control and prevention. These responses include whether individuals trust authority, believe authority is legitimate, or cooperate with authority. Motivational postures are related to this class of responses.

Motivational postures are socially shared thoughts and feelings that become organized into well-crafted signals to authority about different kinds of approval of and deference to authority (Braithwaite et al. 1994; Braithwaite 1995, 2003, 2009). The distinctiveness of the motivational postures concept revolves around their multifaceted nature, providing individuals with a suite of responses. They can be used within a single encounter with authority, across several encounters, or even across a range of authorities. Because motivational postures are outward displays of

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

approval or deference, they can be used to communicate about the quality of relationships and negotiate new relationships with authority.

Social Distancing

At the heart of motivational posturing theory is the notion that people choose how much social distance they place between themselves and an authority, just as they choose how much social distance they place between themselves and another person (Bogardus 1928). Sometimes people may be prepared to approach an authority, to listen and be open to its message. Other times they may keep their distance, being wary of the authority's purpose and turning a deaf ear. As people change their social distance, they are adopting positions that best protect them from an authority's power. In Harris's terms, individuals distance themselves from authority in such a way as to protect or enhance their ethical identity (Harris 2011). At the same time, authorities want individuals to position themselves closely, particularly when an authority wants the public to be responsive to its message (e.g., during a natural disaster or security crisis). Motivational postures are the signals that are sent by individuals and groups; they can be read by authorities and can be used to establish better working relationships.

Some motivational postures signal alignment with authority; others are oppositional. Within most of us, postures coexist. Early socialization teaches us that authority has the power to both help and hinder as we find our way in the world. Authority maps out paths of safety and success. Authority also hinders through rules and their enforcement. Authority in this sense always poses a potential threat to our freedom to act as we want. In response to that threat, we feel ambivalence. We see value in positioning ourselves closely to signal we are in accord with the authority. Then, as we observe the authority acting in ways with which we disagree, we increase our distance as we become less certain that our thoughts and actions are at one with the authority.

Five Motivational Postures

Five motivational postures have been identified empirically with some consistency across the domains of different authorities. *Commitment* and *capitulation* are postures that represent willingness to go along with authority. *Commitment* conveys a belief that the authority's purpose is sound and that, in principle, the authority and its goals should be valued and supported by everyone. Commitment is a posture that enables individuals and groups to go beyond compliance, to do more than an authority expects or asks in the interests of furthering the accomplishment of shared goals.

Capitulation is the posture of doing what is asked, without necessarily understanding or caring about purpose and goals. Similar to McBarnet's (2003) usage, capitulation reflects acquiescence to the powers that be. Capitulation in the posturing context also incorporates acceptance of the idea that authority has superior knowledge and knows what is best.

Commitment and capitulation are postures that signal alignment with authority and openness to cooperation, be it through shared beliefs or acquiescence. These are the most common postures in a healthy democracy. At the same time, however, people may respond with anger or annoyance when they are directly affected by the decisions and actions of authorities. *Resistance* is an expression of hostility toward an authority. When displeased with how an authority is performing, the posture of resistance communicates grievance, usually that the authority treats people in a manner that is unreasonable and unfair. Resistance is a posture that is less likely to take issue with the broader purpose of the authority and more likely to focus on the way the authority uses its power. Acceptance of the authority's purpose gives rise to hope that the authority will mend its ways in the face of criticism.

Once hope is abandoned, the more socially distant postures emerge. Individuals conclude that change will not occur and that their interests – or those of others – would be better served without the authority. Through the posture of *disengagement*, individuals withdraw from having any relationship with the authority. They take no notice of what the authority says or does. The posture of disengagement communicates rejection of the authority's goals and processes. Disengagement may be a posture of anomie (Durkheim 1897[1952]) where people have lost meaningful connection with the norms and values of the authority – and the authority with them. As a result, individuals live their lives apart from the authority and remain impervious to its powers.

While disengagement has a degree of fatalism about it (e.g., the authority will do what it will do and I am not going to have any part in it), *game playing* has a combative agenda. The objective is to outsmart the authority and assert independence over the authority while technically playing within the rules. It is the posture that gives rise to creative compliance (McBarnet 2003). The posture of game playing, while paying attention to the letter of law, shows little respect for the spirit of the law. In adopting the posture of game playing, individuals cleverly sidestep deference to the authority.

Using Motivational Postures and Signaling Defiance

We may move from one posture to another in response to what an authority says or does, or we may simultaneously embrace two or more postures. We may feel committed to the goals of an authority (e.g., preventing terrorism) but at the same time disapprove of how the authority enforces the law (e.g., abusing human rights). We may be prepared to capitulate to an authority in which we have little interest (e.g., paying tax) but fall into a pattern of game playing because it is the norm among our peers (e.g., signing up to tax avoidance schemes).

As postures wax and wane in response to circumstance, it is nevertheless the case that postures overall will vary in their salience. Authorities govern effectively in democracies because the accommodating postures of commitment and capitulation tend to be more salient and dominate the postures of resistance, disengagement, and game playing, at least for most people most of the time. On occasions, however, the oppositional postures of resistance, disengagement, and game playing may assume dominance over the accommodating postures of commitment and capitulation. Oppositional posturing represents defiance (Braithwaite 2009). Defiance sends a message to authority that communicates a state of overriding emotional and rational rejection of authority. Authority sees defiance commonly as one unified whole that needs to be squashed. Motivational posturing theory allows us to approach defiance with a more analytic and discriminating lens. Defiance can be resistant or dismissive. Importantly, the message of each type of defiance is different, as is the most appropriate response.

When the posture of resistance dominates postures of commitment and capitulation, resistant defiance is displayed (Braithwaite 2009: 113–115, 262–269). Resistant defiance is driven by grievance about how the system is working. Responding to grievance and improving the integrity of the system, particularly through procedural reforms such as dealing with people more fairly, respectfully, and openly, reduce levels of resistant defiance.

When the postures of disengagement and game playing take hold, a different kind of defiance emerges – dismissive defiance (Braithwaite 2009: 113–115, 262–269). Dismissive defiance is driven by the belief that the authority is blocking opportunity illegitimately. Dismissive defiance is not readily turned around because the authority is seen as unable to redeem itself. People who are dismissively defiant question the power held by authorities. Most notably, they question why an authority should exist. Dismissive defiance is not reduced by displays of procedural justice, as occurs with resistant defiance. Such displays are either likely to be ignored or considered insincere by the dismissively defiant.

Why Do Motivational Postures Matter?

There are three ways in which motivational postures add value to the array of social science concepts on which criminologists draw. First, motivational postures provide a more nuanced appreciation of an authority's influence than singular concepts such as trust or legitimacy or cooperation. Seldom do people have only one posture. Individuals learn to adapt to institutional life through experiences in schools, religious groups, families, work, and leisure. In the process, individuals acquire the full suite of motivational postures.

The multidimensionality of postures is a reminder to authorities to consider which postures they wish to engage as they go about their business of crime prevention and control. What is more, a person's suite of postures will not necessarily be in perfect alignment. Lack of alignment creates cognitive dissonance and room for deliberation. Engaging with the full range of motivational postures provides opportunity for resolution of differences through persuasion and dialogue. Responsive regulation (Ayres and Braithwaite 1992) and restorative justice (Braithwaite 2002) are approaches to crime control that recognize and respond to multiple motivational postures in those being policed or regulated. Through being able to read the posturing of individuals and groups and understand how postures can be shifted, authorities improve the likelihood that they will elicit cooperation and improve their effectiveness (Braithwaite et al. 2007a, b). They may avoid the escalation of unnecessary conflict, for example, through responding constructively to resistance while strengthening bonds forged through commitment and capitulation. Or an authority may increase its legitimacy and compliance with laws through refining its mission and improving its enforcement regime, thereby tackling game playing while building commitment to new shared goals.

Some critics may say that government is neither flexible nor insightful enough to be responsive to motivational posturing in the way described here. In an era of regulatory capitalism, however, agents of government are doing more of the regulating and enforcement and can invest in understanding how to read posturing, build commitment, deal responsively with resistance, and find ways of reining in disengagement and game playing. Restorative justice is an approach that operates on such principles.

Postures provide useful social cues for improving communication between authorities (or their agents) and individuals at the micro level. They are also useful at the macro level. Motivational postures add value through providing insight into how well an authority is engaging with the public. Motivational postures can be measured and aggregated to give an indication of the social distance between an authority and the communities it serves. An analysis of an authority's motivational posturing profile within and between different communities provides insight into how an authority is falling short in its bid to win public confidence. Authorities that have good relations with the communities they serve and are achieving valued social goals should attract high levels of commitment and capitulation, some level of resistance (an appropriate countermeasure to an authority's power in a democracy), and low levels of disengagement and game playing. Problems in engaging with communities can be identified through departures from this profile. High resistance signals that authorities need to listen more and be responsive to concerns. Unusually high levels of disengagement or game playing signal a more fundamental problem in which authorities have become disconnected from the norms and values of a substantial segment of the community. Reversing the situation involves critically assessing the authority's mission, the moral purpose that underpins it, the laws and rules that supposedly reflect purpose, and the authority's capacity to enforce laws and rules in a respectful way.

The third insight provided by motivational posturing theory is recognition of the two types of defiance: resistant defiance which does not challenge the purpose of the authority but expresses grievance over the way in which the authority carries out its duties and dismissive defiance which challenges the existence of the authority and undermines its effectiveness. If authorities are to show emotional intelligence in how they deal with lawbreakers, it is important that they don't misread the defiance they are dealing with and its consequences, particularly mistaking the more benign form of resistant defiance for dismissive defiance. Resistant defiance is argumentative and annoying for authorities, but it can be turned around if the authority is prepared to listen to grievances and respond in a way that makes the system work more fairly and reasonably. Authorities that work at maintaining their public integrity will manage resistance routinely, allowing them to dedicate more resources to the challenges created by disengagement and game playing. It is the dismissive defiance associated with disengagement and game playing that is most difficult to address constructively because dismissive defiance places lawbreakers psychologically beyond the reach of influence of authority. Dismissive defiance in both taxation and occupational health and safety is associated most strongly with individuals failing to comply with the law (Braithwaite 2009: 270–272, 2011; Braithwaite et al. 2007b).

Background

Motivational postures were discovered in the context of regulation (Braithwaite et al. 1994; Braithwaite 1995, 2003, 2009). Regulation to secure compliance with criminal law (e.g., policing) is a subset of regulation as it has been studied in the motivational posturing context.

As governments introduced laws to improve the standard of care in nursing homes, those working in the nursing home industry were preoccupied with making sense of how they would be affected and treated under the new laws. Change was inevitable, but with it came uncertainty and threat that they would not meet the standards and not receive the approval of the new authority. They adopted positions for dealing with the new authority – motivational postures.

Embracing the new laws involved accommodation to the new standards and processes and in principle commitment to quality care. Becoming reconciled to change and new inspection regimes that hopefully would be benign for those who tried to do the right thing manifested as capitulation. This posture reflected a sense of having no escape from the authority. Feelings of grievance and resentment over enforced change found expression as resistance to authority and the way inspectors used their power. Despair and dismissiveness toward the authority and its goals for better quality of care brought disengagement, most notably among those who felt their business would be untenable in the future. Subsequent research revealed that these postures were present in other regulatory contexts as well. Research in the field of taxation a decade later revealed game playing as a fifth posture at work. It has been replicated in other contexts.

The coherence and regularity of the five postures was identified through factor analyzing the responses of individuals to a motivational posturing questionnaire. These clusters of beliefs, attitudes, preferences, interests, and feelings have been assessed through some 30 statements rated on a five-point strongly disagree-strongly agree Likert scale. Postures have been studied in this way across a range of contexts including taxation (Braithwaite 2003, 2009; Hartner et al. 2008), environmental preservation (Bartel and Barclay 2011), occupational health and safety (Braithwaite 2011), policing (Murphy and Cherney 2012), and child protection (Ivec et al. 2011) as well as nursing home regulation (Braithwaite et al. 1994; Braithwaite et al. 2007a). Observational and qualitative analyses of posturing have been undertaken in research on small business taxation compliance (Harris and McCrae 2005), agricultural reform (Cartwright 2011), war making and peace building in Indonesia (Braithwaite et al. 2010), child protection (Harris 2012), and the resettlement of South Sudanese refugees (Losoncz 2011).

Research has shown that motivational postures are in part shaped by a person's values, norms, and expectations and in part by the actions taken by authorities (Braithwaite 2009). The term "postures" was chosen to capture the socially shared and acceptable nature of these signals – they were not deep dark secrets. The term "motivational" captures their purpose of protecting the individual from the potential threat that authority poses to his or her freedom. Postures are part of the psychological armory that individuals put on to allay any fears they might have in their next encounter with authority.

Applications of Motivational Posturing Theory

A study by Bartel and Barclay (2011) used motivational postures to demonstrate how interventions with farmers who were not complying with environmental laws could be designed in a more considered, targeted, respectful, and responsive way.

Bartel and Barclay identified qualitative differences in the responses of Australian farmers to more stringent regulation and criminalization in environmental protection. One cluster of farmers aligned themselves with authority through postures of commitment and capitulation. They not only complied with the basic legal requirements but also went "beyond compliance" in furthering the objectives of the regulations and the authority. They felt positively toward government intervention, were involved in newer industries, and better positioned economically to cope with change in the management of their smaller properties.

Other farmers clustered into the defiant groups of resistance, disengagement, and game playing. Consistent with motivational posturing theory, the game playing cluster of farmers was the least likely to preserve land for conservation purposes, opposed change, opposed government, and the law. They were older, well-established farmers of their district, and less educated. The resistant farmers and disengaged farmers were more ambivalent in how they approached regulation. Bartel and Barclay (2011) observed defiance as being linked to jurisdiction and industry. The resistant and disengaged farmers tended to be crop growers, dealing with drought and pests and a threatened livelihood. For Bartel and Barclay, motivational postures provided a fine-grained analysis of types of opposition to environmental laws and reasons for that opposition, as well as spotting where entrenched forms of defiance may lie.

In the context of policing, the motivational posture of disengagement has been used to explain the limits of procedural justice in dealings with ethnic minorities (Murphy and Cherney 2012). Murphy and Cherney found that when ethnic minorities believed that laws were not legitimate, procedural justice was counterproductive in eliciting cooperation. Disengagement from authority was the factor that explained why procedural justice was proving counterproductive as a means of increasing cooperation. Disengagement by ethnic minorities from the police meant that police were unable to find a foothold to start building a cooperative relationship through procedural justice.

Tyler and colleagues have shown convincingly that procedural justice can build legitimacy of laws and elicit cooperation with authorities (Tyler 1997). However, when no relationship is in place, as reflected in a posture of disengagement, procedural justice provides an opportunity for game playing with authorities rather than cooperating.

Harris (2012) used motivational posturing theory to better understand the responses of parents to their first visit from a child protection officer. Child protection authorities placed importance on their officers using an assessment framework which was expected to deliver consistency in decision making and "court readiness" should there be need to remove the child from the family. Harris observed parents expressing greater defiance over assessment procedures than over the visit from child protection authorities. Harris' thesis was that where assessment was experienced as being intrusive, good will on the part of parents to cooperate with the investigation was lost, and this was evident through motivational posturing.

Within Harris's (2012) sample of parents receiving their first visit, some were predominantly positive about the assessment process. These parents were more likely to see benefits in the intervention by child protection authorities (commitment posture). Others were less comfortable with the assessment measures but accepted that the investigation would occur and were resigned to putting up with it (capitulation posture). A third group took offense at the assessment process and responded with criticism and anger over the investigation (resistance posture). A fourth group was equally critical, but instead of fighting, withdrew from the authority, expressing no hope that the investigation or anything flowing from it could help them or their child in any way (disengagement posture). Whether parents dealt with what they saw as unreasonable intrusiveness through actively resisting assessment requests or feigning cooperation, they successfully increased the workload of overly stretched child protection officers and undermined effectiveness in protecting children.

Bartel and Barclay (2011) and Harris (2012) identified groups who were experiencing threat from authority at one point in time. Efforts to capture the dynamic possibilities of motivational posturing have been made by Robinson and

McNeill (2008). They propose a model to examine formal, substantive, and long-term compliance with community penalties. Robinson and McNeill argue that when one form of compliance is privileged over others (e.g., formal compliance such as attending scheduled appointments), those serving community sentences may be more likely to engage in a form of posturing (e.g., capitulation to the system of surveillance) that does little to engender commitment to abiding by conditions of community sentencing (substantive compliance) or being law abiding in the longer term.

Robinson and McNeill (2008) propose that offenders with community sentences move between these different levels of compliance. Changes in levels are hypothesized as reflections of their interaction with various actors in the correctional system; different postures come to the fore in response to different experiences and treatment.

Some evidence in support of Robinson and McNeill (2008) proposition comes from qualitative research on how Australian tobacco farmers responded to government closure of their industry in a small rural community. The government turned its back on the tobacco growing industry as it struggled to compete internationally and as it fell into disrepute with the rise of the anti-smoking lobby. Government removed protective tariffs and imposed heavy taxes on the tobacco growers, seriously threatening their livelihood. Recognizing discontent in the region, buyers operating an illicit “chop-chop” market saw an opportunity to move into the community, offering to buy tobacco for a very attractive price while circumventing excise tax.

As government authorities cracked down on the chop-chop market, farmers were incensed that government could be so merciless in its treatment of them. Defiant postures emerged in response to the intrusive and tactical maneuvers of the authorities. The result was that growers distanced themselves from government officials and closed ranks in silence over what they knew about the chop-chop market. The chop-chop industry flourished. Cartwright (2011) documented the postures of resistance and disengagement of the community in her interviews and heard stories of game playing by growers who had thrown their lot in with the chop-chop buyers. She also observed widespread tempering of defiance with capitulation. Growers wanted to stay on the right side of the law and feared the coercive measures taken by chop-chop buyers who demanded a steady supply of tobacco from their suppliers.

In spite of police, customs, and tax officer presence in the community, authorities were unable to gain support. Their strategy was narrowly focused on threatening and catching growers and buyers of chop-chop. There was less evidence of authorities acknowledging the law-abiding postures of growers and helping ensure that those who were trying to stay within the law could keep their farms viable in the future. Interestingly, Cartwright (2011) could find no evidence of commitment to government authority or policy among any of her participants.

The dynamic of how postures can change over time as a result of how authorities are seen to be acting also has been illustrated in the context of tax compliance (Braithwaite 2009, Chapter 8). Perceptions of deterrence and procedural justice (integrity) have been shown to affect defiant posturing. The trajectory of change for resistant defiance is different from dismissive defiance.

An analysis of how resistant defiance gathered momentum over time identified two pathways that competed for dominance over a three-year period. One was a consistent barrier to defiance, a pathway of moral obligation that was theorized as representing a “moral” self – a self that was law abiding and that had nothing to fear from authority. Competing with the moral self was a pathway hypothesized as representing the democratic collective self. The democratic collective self saw unfairness in the system and expressed grievances, including disillusionment with the democracy and dissatisfaction over the tax paid given the goods and services that government provided. The democratic collective self wanted improvements to the system and was prepared to protest in spite of the restraints urged by the moral self. This was how resistant defiance evolved over a three-year period.

Within these dynamics, deterrence played a dual role. It fuelled grievance in the early period, thereby boosting postures of resistance; but over time deterrence strengthened the value of being law abiding, boosting commitment, and capitulation. Procedural justice similarly was shown to have a mixed fate. Those who had distanced themselves from government and felt oppressed by the system were unlikely to acknowledge procedural justice (integrity) in the system. But once acknowledged, postures changed in a more cooperative direction. Over time, resistant defiance was lowered by perceptions that the authority honored its commitment to procedural justice.

A similar but simpler story emerged for dismissive defiance. A weak pathway represented the moral self that upheld ideals of doing the right thing and meeting tax-paying obligations. The stronger pathway fuelling dismissive defiance represented a desire to achieve, particularly aspirations for social standing or wealth. This is referred to as the status-seeking self. At first, the status-seeking self was high on grievance, but within a short time span, this turned into interest to find ways of avoiding tax without breaking the law. Aggressive advisers became the ideal "alternative authority" for the status-seeking self. Only two avenues emerged for keeping dismissiveness in check. First was the weakened moral self. The second was deterrence. Again, deterrence initially fuelled grievance, and only in the longer term did it reduce dismissive defiance and bolster the moral self.

For tax authorities, this study emphasized the importance of system-wide coverage of procedural justice because it is a way of ensuring responsiveness to the posture of resistance and to resistant defiance. Just as important, however, is critical scrutiny of the law, enforcement strategies, and penalties that are necessary for dealing with the less common, but more substantial problems that arise when postures of disengagement and game playing assume dominance and manifest as dismissive defiance. Authorities need to be watchful of the postures developing in the community and adjust their enforcement strategies to preserve a "firm but fair" regime.

Key Issues and Controversies

Motivational posturing is purposely an amalgam of more basic psychological concepts such as attitudes, beliefs, norms, expectations, and needs. The advantage of an amalgam of more basic concepts is that it is more accessible to practitioners. Motivational postures provide a scientifically sound yet practically useful tool for criminologists and others involved in taking communities with them in developing or implementing policy. Yet there remain questions around theorizing change and potential controversy about the normative aspects of posturing.

Theorizing Change

Research has shown how postures can be shaped by individual characteristics such as values, worldviews, and circumstances (Bartel and Barclay 2011; Braithwaite 2009). They are also responsive to the actions of authorities – the rules they enforce and how they enforce them (Braithwaite 2009; Braithwaite et al. 1994). The actions of authorities are shaped, in turn, by jurisdictional laws, administrative instruments, and enforcement cultures (Bartel and Barclay 2011). All three – individual characteristics, actions of authorities, and the regulatory system – vary across domains, for example, policing, peacekeeping, nursing homes, taxation, occupational health and safety, and the environment. While postures have been shown to operate across all these domains, they do so differently. Understanding changes in how postures operate across domains, systems, and cultures requires further theoretical development.

Normative Value of Postures

If postures are responsive to characteristics of the person, the authority, and the jurisdiction, what then constitutes a desirable posture? From the perspective of an authority, commitment and capitulation are the signals authorities want to receive, while resistance, disengagement, and game playing are less welcome. An argument has already been made for why authorities should embrace resistance as essential feedback on performance. But the important normative questions to address are the following: Are postures of disengagement and game playing necessarily undesirable? Are there not occasions when dismissive defiance is necessary to overcome oppression and bad government?

Why Do Authorities Fear Postures That Promote Dismissive Defiance?

In the broader context of establishing law and order from local neighborhood policing through international peacekeeping, it is important to recognize that authorities create defiance through their very existence. Defiance is not just something that offenders experience and display because they are overly emotional or lack self-control. When authorities make their presence felt in the lives of individuals, individuals must manage the experience of intrusion on their freedom – they must manage the fact that they are expected to obey the authority even if they don't want to or believe what is being asked of them is morally wrong.

Low trust in governments, rising individualism, plural societies, and/or cultural heterogeneity (LaFree 1998) mean that it is easy to deal with the threat posed by authorities through demonizing them as not worthy of respect and rationalizing defiance as the only way forward. Whether the problem is child abuse, domestic violence, gang violence, street protests, tax evasion, or financial crimes on Wall Street, defiance, once widespread, undermines the morale of law enforcers and eventually the authority's public legitimacy. The global financial crisis revealed how the financial crimes of the powerful can be engineered with such defiance that neither government nor their regulatory authorities could or would effectively challenge practices and avert disaster. Authorities have reason to fear defiance. And there is good reason for thinking that dismissive defiance may not serve the public well.

The Other Side of the Argument: Dismissive Defiance Brings Change

But defiance, even dismissive defiance, in itself is not necessarily undesirable. Healthy democracies value and embrace defiance among their citizenry: Failure to do so undermines democracy (Durkheim 1961). Defiance may lead to injustice being challenged through formal processes including the courts. Or it may lead to a challenge on the streets, involving lawbreaking and criminal arrests (Lovell 2009). Suffragettes were arrested as they battled for equal rights for women. Rosa Parkes spearheaded the US Civil Rights Movement when she was arrested for refusing to obey a bus driver who directed her to give up her seat for a white passenger. Nelson Mandela and the anti-apartheid activists too were punished harshly by authorities for their defiance in fighting for freedom and justice. The Arab Spring has revealed the courage and defiance of ordinary people fighting for democracy. Lovell (2009) has provided a sympathetic analysis of "seasoned activists who are willing to transgress the law in the pursuit of social justice" (xi) and who carve out their niche at the poorly researched intersection of politics and criminology.

We can only conclude that the normative value of defiance depends on its manifestation and its purpose and the change that it brings to society. As Arendt (2000) points out, we may hope that change will be for the better, but it is not something of which any of us can ever be sure.

Future Directions

An important way of extending motivational posturing theory is to consider how these ideas operate at a group (community, national, or corporate) level as opposed to individual level and how posturing is used by authorities to control, rightly or wrongly, the communities that come under their influence. Such developments show promise in relation to three ongoing strands of work. Planned social change on a large scale requires the marshaling of collective hope (Braithwaite 2004). Marshaling the hopes of a few (the likes of Nelson Mandela or Osama Bin Laden) into collective hope involves building support around shared goals, engendering collective confidence that these shared goals can be achieved, and finding the pathways to progress the agenda such that individual efforts are coordinated to produce an outcome that is more than the sum of its parts (Braithwaite 2004). Such processes unfurl amidst numerous setbacks and hiccups, and arguably most fail. Motivational postures provide a framework for tracking the journey of collective hope in terms of its ascendancy, threats to ascendancy, or demise.

A second strand of work involves regulatory ritualism and understanding how organizations and nation-states can replace ritualism with more productive and authentic action (Braithwaite et al. 2007a). According to Merton (1968), ritualism means acceptance of institutionalized means for securing social goals while losing all focus on achieving the goals or outcomes themselves. In nursing homes, for example, new policies and in-service training programs may be introduced to respond to problems of noncompliance, but the new policy may never be implemented and the in-service training program may not address inspector's concerns. Regulation has spawned many rituals of comfort but not of compliance for good outcomes (Braithwaite et al. 2007a). A regulatory regime that settles for postures of capitulation to the neglect of postures of commitment will be at risk of regulatory ritualism. When postures of capitulation dominate at the expense of postures of resistance, the likelihood of self-learning and self-initiated change becomes impossible (Braithwaite et al. 2007a). The dialogue and conversations that can unsettle conventional practice and move people out of their comfort zone do not take place, and ritualism prevails.

In a similar way, regulatory ritualism threatens the effectiveness of many international "social justice" initiatives particularly in developing countries. Such countries are urged to become signatories to international agreements for human rights or equity in education or health care, sometimes with rewards from donors once agreements are signed. Capitulation without commitment means that meaningful change is unlikely, replaced by token gestures and empty rituals (Charlesworth 2011).

A third strand of work focuses on the contribution of motivational posturing to the goal of crime prevention through strengthening communities' efficacy and resilience. Within this agenda, sparks of individual defiance need to be harnessed and redirected toward goals that benefit communities. In such contexts, the posture of commitment within communities comes to the fore. Commitment gives rules and laws a sense of meaning and purpose beyond the situation people find themselves in. Commitment to ideas can drive out anomie. Anomie often prevails in communities that have declared "war" on authorities. Kolodziej (2011) has shown empirically that in Poland where tax authorities and taxes are held in low regard, commitment to taxation is related to being knowledgeable about the way the economy works. This understanding of the bigger picture enables individuals to adopt a more positive attitude to having a tax system and move beyond the complaints associated with a system that is not currently meeting public expectations.

Theoretical developments around hope, ritualism, and resilience require a synthesis of micro and macro social processes, the individual psychology of cooperative engagement and the mobilization of meaning and purpose on a larger scale. It remains to be seen how well motivational posturing theory can contribute to this synthesis. It also remains to be seen whether low crime societies manifest a politics of hope and commitment. Do they eschew the cynicism of ritualism, the lure of opportunity (Shover 2007) and gaming of law? That is a much bigger research agenda, one that we have yet to grasp.

Related Entries

- ► Procedural Justice, Legitimacy, and Policing
- ► Psychology of Procedural Justice and Cooperation
- ► Shame in Criminological Theory

Recommended Reading and References

Arendt H (2000) "What remains? The language remains": a conversation with Günter Gaus. In: Baehr P (ed) The portable Hannah Arendt. Penguin, New York

Ayres I, Braithwaite J (1992) Responsive regulation: transcending the deregulation debate. Oxford University Press, New York

Bartel R, Barclay E (2011) Motivational postures and compliance with environmental law in Australian agriculture. J Rural Stud XXX:1–18

Bogardus ES (1928) Immigration and race attitudes. D. C. Heath, Boston

Braithwaite V (1995) Games of engagement: postures within the regulatory community. Law Policy 17(3): 225–255

Braithwaite J (2002) Restorative justice and responsive regulation. Oxford University Press, New York

Braithwaite V (2003) Dancing with tax authorities: motivational postures and non-compliant actions. In: Braithwaite V (ed) Taxing democracy: understanding tax avoidance and evasion. Ashgate, Aldershot, pp 15–40

Braithwaite V (2004) The hope process and social inclusion. Ann Am Acad Pol Soc Sci 592:128–151

Braithwaite V (2009) Defiance in taxation and governance: resisting and dismissing authority in a democracy. Edward Elgar, Cheltenham/Northampton

Braithwaite V (2011) Motivations, attitudes, perceptions and skills: pathways to safe work. Safe Work Australia, Australian Government, Canberra

Braithwaite V, Braithwaite J, Gibson D, Makkai T (1994) Regulatory styles, motivational postures and nursing home compliance. Law Policy 16(4):363–394

Braithwaite J, Makkai T, Braithwaite V (2007a) Regulating aged care: ritualism and the new pyramid. Edward Elgar, Cheltenham/Northampton

Braithwaite V, Murphy K, Reinhart M (2007b) Taxation threat, motivational postures, and responsive regulation. Law Policy 29(1):137–158

Braithwaite J, Braithwaite V, Cookson M, Dunn L (2010) Anomie and violence: non-truth and reconciliation in Indonesian peacebuilding. ANU E-Press, Canberra

Cartwright S (2011) Policy, obligation and justice in Australia's tobacco growing industry. PhD dissertation, Australian National University, Canberra

Charlesworth H (2011) Swimming to Cambodia: justice and ritual in human rights after conflict. ALTA law research series (ALRS), 20 Sept. http://www.austlii.edu.au/au/journals/ALRS/2011/4.html

Durkheim E (1897[1952]) Suicide: a study in sociology. Routledge & Kegan Paul, London

Durkheim E (1961) Moral education. Free Press, New York

Harris N (2011) Shame, ethical identity and justice interventions: lessons from research on the psychology of influence. In: Karstedt S, Loader I, Strang H (eds) Emotions, crime and justice. Hart, Oxford

Harris N (2012) Assessment: when does it help and when does it hinder? Parent's experiences of the assessment process. Child Fam Soc Work 17(2):180–191

Harris N, McCrae J (2005) Perceptions of tax and participation in the cash economy: examining the role of motivational postures in small businesses. Australian National University, Canberra. Regulatory Institutions Network Working Paper 80

Hartner M, Rechberger S, Kirchler E, Schabmann A (2008) Procedural fairness and tax compliance. Econ Anal Policy 38(1):137–152

Ivec M, Braithwaite V, Reinhart M (2011) A national survey on perceptions of how child protection authorities work 2010: the perspective of third parties – preliminary findings. Australian National University, Canberra. Regulatory Institutions Network Occasional Paper 16

Kolodziej S (2011) The role of education in forming voluntary tax compliance. Gen Prof Educ 1:22–25

LaFree G (1998) Losing legitimacy: street crime and the decline of social institutions in America. Westview, Boulder

Losoncz I (2011) Blocked opportunity and threatened identity: understanding experiences of disrespect in South Sudanese Australians. Aust Rev Afr Stud 32:121–145

Lovell JS (2009) Crimes of dissent: civil disobedience, criminal justice, and the politics of conscience. New York University Press, New York

McBarnet D (2003) When compliance is not the solution but the problem: from changes in the law to changes in attitude. In: Braithwaite V (ed) Taxing democracy: understanding tax avoidance and evasion. Ashgate, Aldershot, pp 229–244

Merton RK (1968) Social theory and social structure. Free Press, New York

Murphy K, Cherney A (2012) Understanding cooperation with police in a diverse society. Br J Criminol 52:181–201

Robinson G, McNeill F (2008) Exploring the dynamics of compliance with community penalties. Theor Criminol 12(4):431–449

Shover N (2007) Generative worlds of white-collar crime. In: Pontell HN, Geis G (eds) International handbook of white-collar and corporate crime. Springer, New York, pp 81–97

Tyler TR (1997) The psychology of legitimacy: a relational perspective on voluntary deference to authorities. Person Soc Psychol Rev 1(4):323–345

Defiance Theory

Leana A. Bouffard[1] and
Lawrence W. Sherman[2,3]
[1]College of Criminal Justice, Sam Houston State University, Huntsville, TX, USA
[2]Institute of Criminology, Cambridge University, Cambridge, UK
[3]University of Maryland, Baltimore, MD, USA

Overview

In 1993, Lawrence W. Sherman proposed defiance theory, which laid out the factors leading to defiant or deterrent reactions to punishment. His original theory may be interpreted as a counter to deterrence/retribution approaches to achieving compliance with the law in that it proposed that while deterrent effects may result from sanctions, a combination of factors may also produce defiant effects for that same sanction. Criminological theory and criminal justice policy have long focused on the relationship between sanctions and criminal behavior. Deterrence and labeling, in particular, are two major theoretical traditions that emphasize sanctions as a key explanatory factor, providing contradictory predictions about the impact of those sanctions on behavior. Deterrence theorists predict that sanctions, especially those which are swift, certain, and proportionally severe, will deter or reduce further criminal behavior. Additionally, criminal justice policy is often predicated on the assumption that sanctions deter offenders. Labeling theory, on the other hand, predicts that sanctions will stigmatize the offender, producing increased offending (i.e., secondary deviance) in the future.

Defiance Theory (1993)

A large body of research examining the deterrence and labeling theoretical perspectives has produced inconclusive results. Research suggests that, in varying instances, sanctions may either deter or increase future offending. Recognizing this diversity in the effect of sanctions, Lawrence W. Sherman (1993) argued that the pattern of sanction effects observed in existing research exhibits two themes. First, the impact of sanctions appears to depend on perceptions of fairness, in that sanctions viewed as unfair are more likely to increase offending. Second, sanctions appear to increase crime among out-groups while deterring crime among in-groups. Suggesting that existing theory is incapable of accounting for these patterns, Sherman proposed *defiance theory* to explain the conditions under which sanctions will increase criminal activity versus deterring offending.

The starting point in Sherman's original defiance theory is the concept of defiance, which he defined as "the net increase in the prevalence, incidence, or seriousness of future offending against a sanctioning community caused by a proud, shameless reaction to the administration of a criminal sanction" (p.459). Defiance may take several forms. Similar to deterrence, defiance may be either specific (i.e., the reaction of an individual to his or her own punishment) or general (i.e., the reaction of a group to the punishment of a group member). Additionally, individuals may exhibit either direct defiance, reacting against the sanctioning agent, or indirect defiance, reacting against another individual who vicariously represents the sanctioning agent. Sherman provides examples of the different types of defiance. For example, an individual who assaults a police officer during an arrest is exhibiting specific, direct defiance. An individual who assaults his or her spouse following a domestic violence arrest is exhibiting specific, indirect defiance. General, direct defiance is illustrated by the South African ambush killings of police officers, who were viewed as the tools of oppression during the apartheid era, a perception that has lingered. The 1992 Los Angeles riots following the acquittal of the four police officers who beat Rodney King during a traffic stop provide an example of general, indirect defiance.

Conditions for Defiance

Sherman argues that defiance is likely to occur when the sanctioned individual views his or her punishment as unfair or illegitimate, is poorly bonded, and denies the shame of the punishment. In contrast, sanctions are expected to produce deterrence when the sanctioned individual is well bonded, views the sanction as legitimate, and accepts the shame he or she feels. In this case, the individual remains proud of his or her connection to the community, recognizes the harm that his or her actions have caused, and attempts to repair that bond. Sanctions are irrelevant to future offending when these factors are evenly balanced. For example, when a well-bonded offender denies the shame associated with the unfair sanction, the expected outcome will likely be irrelevance, not deterrence. In this instance, the perceived unfairness of the sanction and the failure to accept the shame that accompanies the sanction will nullify any deterrent effect produced by the strong social bond. With this discussion, Sherman is able to theoretically account for the mixed effects of sanctions apparent in the research literature.

Defiance theory focuses on explaining the defiant reaction to a sanction. Specifically, there are four necessary conditions for defiance to occur: (1) the sanction must be defined by the offender as unfair, (2) the offender must be poorly bonded to society, (3) the sanction must be viewed by the offender as stigmatizing, and (4) the offender must refuse to acknowledge the shame produced by the sanction. In proposing his theory and identifying these four conditions, Sherman has borrowed from John Braithwaite's theory of reintegrative shaming, Tom Tyler's concept of procedural justice, and Thomas Scheff and Suzanne Retzinger's discussion of the role of shame and rage in destructive conflicts.

Perceptions of Fairness

According to Sherman, one key theme in understanding whether a sanction produces defiance or deterrence is the perceived fairness or legitimacy of the sanction or sanctioning agent. Unfairness may be related to disrespect by the sanctioning agent or a perception that the punishment is arbitrary, discriminatory, or otherwise unjust. Whether the unfairness of a punishment is substantive or perceptual, unfair or unjust sanctions may not have their intended deterrent effect. According to Tyler's procedural justice perspective, sanctions that are perceived as unfair reduce the legitimacy of law enforcement or the criminal justice system, which reduces the likelihood of compliance. If an individual perceives a punishment as unjust, he or she may begin to question the law itself and feel justified in disregarding it. Scheff and Retzinger contend that societal disapproval, when expressed disrespectfully or to a person with weak social bonds, may evoke anger. Reintegrative shaming theory likewise argues that sanctions that stigmatize and label the offender may weaken existing social bonds and produce increased offending. Thus, perceptions of the fairness of a sanction and the experience of being stigmatized by a sanction may interact with an individual's social bonds to produce defiant effects.

Social Bonding

Procedural justice argues that when the legitimacy of formal sanctions breaks down, social sanctions are expected to take their place. Thus, social bonding also plays a large role in defiance theory. In particular, Sherman relies on some elements of reintegrative shaming theory to explain the connection between unfair or stigmatizing sanctions and social bonds. For Braithwaite, individuals who have strong social bonds (i.e., interdependency) may be more likely to experience reintegrative sanctions, which are rejecting of the act but avoid applying a label to the individual. Thus, reintegrative sanctions are likely to be viewed as fair and to produce deterrence. Disintegrative sanctions, however, are rejecting of both the act and actor, stigmatizing the sanctioned individual. These sanctions are more likely to be viewed as unfair and disrespectful. Sherman likewise recognizes the potential criminogenic effect of stigmatizing sanctions, especially among individuals with weak social bonds. He argues that individuals with strong social bonds will not react defiantly to a punishment perceived as unfair so as not to jeopardize those bonds. On the other hand, individuals with weak social bonds are more likely to deny the shame of being sanctioned and respond with indignation and anger. This angry, prideful reaction sets the stage for defiance and increased offending.

Experiencing Shame

An individual's reaction to the shame of a sanction is the final link in the explanation of defiance. Sherman highlights the role of shame, pointing both to reintegrative shaming theory and to Scheff and Retzinger's work on the master emotions of shame and pride. Both Braithwaite and Scheff and Retzinger argue that individual reactions to the shame of a sanction will vary depending on an individual's level of social bonding. Similarly, labeling theory suggests that secondary deviance (i.e., defiant effects for Sherman) occurs as a reaction to the experience of being sanctioned, in that individuals may perceive their punishment as an attack and may act defiantly as a defense to society's disapproval.

Scheff and Retzinger criticize early versions of labeling theory for failing to take emotions, especially shame, into account. For these authors, societal disapproval is described as a threat to the sanctioned individual's social bonds. If the individual accepts the shame that he or she feels and recognizes the harm he or she has caused, the individual may seek to avoid that behavior in the future (i.e., deterrence). Braithwaite's theory of reintegrative shaming presents a similar argument. On the other hand, if the person refuses to acknowledge or rejects that shame, he or she may respond with self-righteous anger. Scheff and Retzinger describe a shame/rage spiral, in which rage is a protective measure against shame, a way of rejecting the shame, and a defense against a perceived attack. Thus, this shame/rage spiral occurs when a person's bond is threatened, the shame is not acknowledged, and behavior is interpreted as an attack. This produces violence, hatred, and resentment which may lead to defiance. For Sherman, shame, or the refusal to acknowledge shame, is the primary causal mechanism in explaining defiant effects.

Empirical Evidence for Defiance Theory

Sherman concludes his theoretical formulation by noting that "until recently, the science of sanction effects has been short on facts and even shorter on theory. Now, it seems, the available theory has gotten ahead of the facts" (p. 468). Despite the promise of defiance theory in explaining variation in sanction effects, there have been no complete tests of the theory since its development. Most of the evidence that can be marshaled in support of the theory is derived from studies not originally designed to examine its propositions. Some research supports the notion that perceptions of unfairness, either to the law being imposed or to the sanction itself, are likely to lead to more criminal offending (i.e., defiant effects). Sherman highlights research suggesting that previously sanctioned individuals are less likely to be deterred. It may be that, because few people are formally sanctioned for offending, those who do receive a punishment perceive their treatment as comparatively unfair and respond defiantly by engaging in further delinquency.

Additional research examining police-citizen encounters indirectly tests some of the propositions articulated by defiance theory. These studies primarily focus on the offender's (or citizen's) perceptions of fair treatment by police officers in their encounters. Raymond Paternoster, Robert Brame, Ronet Bachman, and Lawrence Sherman (1997) examined the effect of arrest on the likelihood of engaging in subsequent domestic assaults and found that the offender's perceptions of fair treatment by police were important determinants of future offending. Other studies of police-citizen interactions support the premise that individuals who feel that they are unfairly treated by police are more likely to be resistant. In other words, the perceived legitimacy of a police officer's action is an important predictor of citizen compliance or resistance. When the police are perceived to be respectful to citizens, compliance is more likely. Confrontational and physical actions on the part of police, on the other hand, are more likely to produce resistance, possibly because the actions are interpreted as unfair and stigmatizing. This body of research supports Sherman's argument that defiant effects are more likely to occur when sanctions are perceived as unfair. While these results are suggestive, they do not address the key to defiance theory, which is an individual's perception of the sanction.

A more recent study examined the perceptual nature of defiance theory and the impact of those perceptions on future offending more closely and has provided the most complete test of the theory to date. Leana Bouffard and Nicole Leeper Piquero (2010) found that individuals who perceived a sanction as unfair and were poorly bonded had higher rates of future offending. While much existing research demonstrates that perceptions of unfairness and social bonding have a strong connection to offending, the role of shame in producing defiance remained unclear in this study. Other research, however, has supported the role of shame in offending.

In 2006, David Brownfield examined the propositions of defiance theory as they relate to gang membership and found that all elements of defiance theory were linked to gang membership. More specifically, perceptions of legitimacy and pride were important. Gang members reported being happy about the possibility of being arrested. A study of recidivist drunk driving in Australia likewise found evidence that shame experienced as a result of the sanction may produce a sense of self-righteous anger, which leads individuals to question the legitimacy of the sanctioning agent (Freeman et al. 2006). These studies have also noted the importance of delinquent peers, who might provide justification for defiance and a social reaction that enhances the likelihood of a prideful response. Unfortunately, the existing research generally provides only piecemeal support for defiance theory. Studies specifically designed to link the theory's propositions together are necessary.

Recent Developments in Defiance Theory

While empirical evidence provides some support for the propositions of defiance theory (Sherman 1993), the explanation is limited in its position as an explanation of lawbreaking, especially in response to sanctions. In 2010, Sherman expanded the original conceptualization of defiance to position it as the independent variable in a broad, general theory of criminology that encompasses "lawmaking, law-breaking, and responses to law-breaking" (Sherman 2010, p. 360). Sherman comments that Sutherland's original definition of criminology included all three of these elements. No theory has yet been proposed, however, to unify all three in one causal framework. Thus, this reconceptualization is an ambitious project designed to provide a more comprehensive theoretical perspective that applies to all three components of criminology.

This reconceptualization may be viewed as an elaboration of the 1993 version of defiance theory, broadening and extending the original through the process of consilience, in which predictions from one domain of criminology (i.e., lawbreaking) are applied to other domains (i.e., lawmaking and responses to lawbreaking). As Sherman contends, observations from all three domains point to the same causal force, the sense of moral obligation to resist the status quo, what Sherman refers to as defiance.

The major elaborations of the original theory include broadening the definition of defiance as the moral sense of "obligation or justification to defy the status quo" (p. 364). Sherman recognizes a continuous contest between defiant and deterrent actors that encompasses both lawful and unlawful efforts to create or change law, the use of power by agents of law, and confrontations between lawbreakers and law-enforcers (e.g., police-citizen interactions). Because the general defiance theory incorporates all elements of criminology and reciprocal causation, this is the major restatement that allows the more general theory to cover all domains within criminology.

Sherman's formal restatement proposes that "criminal laws are more likely to be made, broken, obeyed and enforced when people intuitively feel that their actions are morally founded" (p. 366). He defines defiant effects in a similar framework as deterrent effects, in that a general reduction in crime is assumed to be a deterrent effect without evidence that the deterrence process (i.e., the sanction imparts a fear of future sanctions) is operating. Likewise, a defiant effect would be defined as "active resistance to a countervailing force" (p. 369) or resistance to the status quo. These are a broad class of observable behaviors reflecting decisions to make law or policy, break laws, enforce laws, or a choice not to enforce laws. Defiance then is an "individual or collective sense of moral obligation, including indignation and empathy" (p. 369). Individuals selectively perceive harm caused to themselves or others, and those perceptions of harm may produce moral intuitions of indignation, empathy, or both. These facilitate a definition of the status quo as illegitimate and produce an obligation to defy that status quo.

Sherman provides a number of specific predictions or hypotheses based on his general defiance theory that encompass the three domains of criminology:

1. Conduct will be criminalized to the extent that opponents of the conduct promote a feeling of moral obligation to oppose that conduct as illegitimate.
2. Laws will be enforced to the extent that enforcement agencies collectively and individually

accept the legitimacy of the moral obligation to do so.

3. Laws will be broken to the extent that potential offenders feel a moral obligation to defy agents making and enforcing laws as illegitimate.
4. Morally infused campaigns of law-breaking can lead to changes in law, to the extent that the violations promote a feeling of moral obligation to allow them.
5. Law-makers and enforcers will also break laws, or re-make them, to the extent that they feel moral indignation, empathy, or a moral obligation to do so.
6. Changes in moral indignation or empathy can be fostered by events causing an epiphany revealing the harm caused by some conduct, law or punishment.

(Sherman 2010, p. 375)

Types of Defiance

As in the original defiance theory, Sherman (2010) distinguishes a number of potential defiant effects. He retains a sense of specific as opposed to general defiance that refers to the individual or collective response to an encounter with the status quo, either its content or the violation or enforcement of law (see p. 370). In the original version of defiance theory, Sherman defined direct and indirect defiance. Here, he retains the initial meaning but renames these forms as direct and displaced defiance, referring to behavior aimed either at those who have committed a moral wrong (i.e., direct) or at those not directly known to have committed the harm but belonging to and representing the same group (i.e., displaced).

A distinction between concealed and identifiable defiance refers to efforts by the actor to hide their actions and is related to how much they are willing to sacrifice with their defiance. For example, those individuals who flee the scene of the crime to avoid capture are engaging in concealed defiance. Those who openly identify themselves and seek attention for their actions engage in identifiable defiance. For identifiable defiance, Sherman gives the example of suicide bombers who create video- or audiotape recordings prior to committing the act. Defiance may also be categorized as active (i.e., requiring action) or passive (i.e., requiring the withholding of action). In terms of law-breaking or offending, commission of most crimes would be considered active; however, failure to wear seat belts or helmets would be examples of passive defiance. For example, despite helmet laws, some motorcyclists may choose not to wear helmets as a defiant withholding of action related to perceptions that the government has overstepped their authority in regulating behavior. In terms of law-enforcing, Sherman describes the use of police discretion (e.g., under-enforcement or choosing not to arrest in certain situations) as passive defiance. On the other hand, excessive use of force by police may be defined as active defiant law-enforcing. Finally, Sherman distinguishes resistant and dismissive defiance based on perceptions about the moral authority of the wrongdoing agent or agency. Resistant defiance recognizes the moral authority of the agent but advocates for change, while dismissive defiance rejects the agent altogether, disabling or eliminating the authority.

Defiance and Other Theoretical Perspectives

In his 2010 restatement, Sherman retains a focus on the legitimacy of sanctioning agents or other agents of authority. In contrast to the earlier version, which viewed the acknowledgement of shame as the key predictor of defiance, the 2010 version seems to place legitimacy as the main element in understanding defiant effects. Thus, Tyler's work on procedural justice and other research exploring issues of fairness retains a place of importance within the General Defiance Theory. Sherman also positions General Defiance Theory as a complement to subcultural and conflict theories in that it recognizes both competing systems of morality within society and the use of the power and material resources of authorities to shape moral obligation through such forces as moral panics. Additionally, the theory borrows from techniques of neutralization in recognizing the justifications offered for defiance or resistance to the moral force.

Future Directions

Researchers have highlighted the links between Sherman's (1993) original defiance theory and other explanatory mechanisms. For example, within psychology, the personality construct of grandiosity (i.e., exaggerated perceptions of self-worth) may inform the path to defiance, in that grandiose or self-centered individuals may be more likely to reject the sanctioning agent and the shame associated with being sanctioned, resulting in a defiant response. In criminology, research also suggests that individuals with low levels of self-control are more likely to perceive sanctions as unfair and to respond with anger (Piquero et al. 2004). Though not specifically addressed by Sherman or necessarily intended, one advantage of defiance theory is that in linking concepts, like emotion and social bonding, it offers an integrative perspective that accounts for either defiance or deterrence.

From the life-course and criminal career perspective, defiance may also be seen as an explanation of continuity in and desistance from offending. The theory provides an explanation for continuity, which is the defiant response of a poorly bonded offender who defines his or her sanction as unfair and stigmatizing and refuses to acknowledge the shame he or she feels. These individuals may continue or escalate their offending, becoming involved in secondary deviance (i.e., persistence). The original theory can also explain desistance by arguing that if an individual defines a sanction as unfair and stigmatizing but has strong social bonds, that person may accept the shame that he or she feels or be unwilling to jeopardize his or her bonds through a defiant reaction. According to Sherman, these individuals will be deterred from future offending (i.e., they will desist). In the 2010 restatement, Sherman extends this argument by suggesting that desistance may be thought of as an individual defying their own prior behavior. In other words, those who have previously engaged in lawbreaking may eventually recognize the harm that they have caused to themselves and others, producing a moral obligation to resist their own previous behavior (i.e., a turning point in the language of the life-course perspective). Thus, exploring defiance theory from a longitudinal, life-course framework is another promising avenue for research.

James Unnever and Shaun Gabiddon (2011) also highlight the role of defiance in providing an understanding of the overrepresentation of African-Americans, especially men, in the criminal justice system. In their Theory of African-American Offending, these authors highlight perceptions of the legitimacy of criminal justice agencies as a key explanatory factor. Due to a history of racial injustices at the hands of the criminal justice system, African-Americans view law enforcement and other criminal justice actors and agencies as having little legitimacy, perceiving that the law and the criminal justice system are racist, disrespectful, and unfair to African-Americans. Greater levels of experienced and/or vicarious racial injustice interfere with bonding to predominantly white social institutions like school and work. Thus, offending by African-Americans is viewed as a defiant response to racial injustice. Sherman (2010) likewise points to the history of race and law in the United States as an illustration of various aspects of his General Defiance Theory. For example, in terms of lawmaking, Sherman argues that his theory is capable of accounting for both the justification/rise and the demise of slavery within the USA by positioning these decisions within a consideration of the shifting public sentiments from the moral justification of slavery to a moral obligation to oppose the enslavements of human beings. Both of these perspectives suggest the relevance of the concept of defiance and of General Defiance Theory in exploring the relationship between race and crime, criminal law, and the criminal justice system.

Conclusion

It is difficult to truly assess the value of Sherman's original defiance theory with the paucity of studies that directly test its propositions. Rather, the existing research provides suggestive evidence supporting some elements of the theory, particularly perceptions of fairness and social bonding. Additionally, the 2010 reconceptualization into a General Defiance Theory of criminology adds layers of complexity that make testing its propositions more difficult. As a new theory, no researchers have yet attempted to assess the propositions set forth in Sherman's (2010) restatement. What remains is to explicitly design studies that link the elements of defiance together as proposed by the theory and to explore the connections between this and other theories at both the specific level of the original version (i.e., in predicting lawbreaking) and at the more general level of the restatement (i.e., in also predicting lawmaking and law-enforcing). At this point, the theory is relevant to understanding the importance of the interplay between perceptions of fairness and legitimacy, social bonding, and the experience of shame, but future research is necessary to fully explore these relationships and to assess how well the concepts of defiance, moral obligation, etc., extend to all domains within criminology.

Related Entries

- ▶ Compliance and Corporate Crime Control
- ▶ Defiance and Motivational Postures
- ▶ Shame in Criminological Theory

Recommended Reading and References

Bouffard LA, Piquero NL (2010) Defiance theory and life course explanations of persistent offending. Crime Delinq 56:227–252

Braithwaite J (1989) Crime, shame, and reintegration. Cambridge University Press, Cambridge

Brownfield D (2006) A defiance theory of sanctions and gang membership. J Gang Res 13:31–43

Freeman J, Liossis P, David N (2006) Deterrence, defiance and deviance: an investigation into a group of recidivist drunk drivers' self-reported offending behaviours. Aust N Z J Criminol 39:1–19

Paternoster R, Brame R, Bachman R, Sherman LW (1997) Do fair procedures matter? The effects of procedural justice on spouse assault. Law Soc Rev 31:163–204

Piquero AR, Gomez-Smith Z, Langton L (2004) Discerning unfairness where others may not: low self-control and unfair sanction perceptions. Criminology 42:699–734

Scheff TJ, Retzinger SM (1991) Emotions and violence: shame and rage in destructive conflicts. Lexington Books, Lexington

Sherman LW (1993) Defiance, deterrence, and irrelevance: a theory of the criminal sanction. J Res Crime Delinq 30:445–473

Sherman LW (2010) Defiance, compliance, and consilience: a general theory of criminology. In: McLaughlin E, Newburn T (eds) The Sage handbook of criminological theory. Sage, Thousand Oaks.

Tyler TR (1990) Why people obey the law. Yale University Press, New Haven

Unnever JD, Gabbidon SL (2011) A theory of African-American offending: race, racism, and crime. Routledge, New York

Defining Disorder

Bernard E. Harcourt[1] and Danielle Wallace[2]

[1]Law School, University of Chicago, Chicago, IL, USA

[2]School of Criminology and Criminal Justice, Arizona State University, Phoenix, AZ, USA

Synonyms

Incivilities

Overview

There's a lot of order in disorder and a lot of disorder in order. The problem, it seems, is in the definition – in defining disorder, or order, that is. Prominent scholars have remarked that disorder is a "slippery" concept (Skogan 1990), and a few quick observations confirm that. Red-light districts – areas that have traditionally been associated with the notion of disorder – are often more orderly than expected (Skogan 1990, pp. 61–63). Neighborhoods that are governed by organized crime or other tight social networks often turn out to be some of the safer neighborhoods in an urban setting (Suttles 1972). Conventional indices of neighborhood disorder, for instance, people loitering at the corner, often turn out to be forms of social control that ensure safety in the community (Patillo 1998). And perceptions of disorder, it seems, are often the product of the racial composition of the neighborhood, not the level of danger: individuals often perceive Black and Hispanic neighborhoods as far more disorderly than any objective measurement would establish (Sampson and Raudenbush 2004).

This makes it extremely difficult to define disorder – in fact, it may make it impossible. Disorder, it turns out, is in the eye of the beholder. It is a normative rather than purely descriptive category. It functions, most often, as a statement of preference. A good example is New York City under mayor Rudolph Giuliani and former police commissioner William Bratton: during the first 3 years of broken windows policing, from 1994 to 1996, the rate of robbery victimization fell about 60 %, and many observers claimed victory for order, but during the same period, allegations of police misconduct rose by about 60 % (Harcourt 2001, pp. 167–170). Whether to describe that period as orderly or disorderly tells us more about, say, our preferences and normative values, than it does about the scientific measure of disorder. But perhaps the most telling illustration of the difficulty of defining disorder is the multiplicity of definitions employed by scholars: even though there has been extensive work trying to define both disorder and order (see, e.g., Ross and Mirowsky 1999), as Kurbin (2008) notes "variability in how disorder is understood and conceptualized across studies is the rule rather than the exception" (p. 205). Indeed, researchers generally do not even agree on whether to call the phenomenon "disorder" or "incivilities" (Kubrin 2008). In this entry, we catalogue some of the main issues surrounding the definition of disorder.

Important Controversies

Creating Disorderly Subjects

Generally, disorder is thought of as containing two dimensions: physical and social disorder. Physical disorder pertains to the decaying urban environment – for example, abandoned buildings, rundown sidewalks, decaying storefronts, or broken windows; social disorder embodies any public behavior of individuals that can be seen as threatening (Sampson and Raudenbush 1999; Harcourt and Ludwig 2006). Cues in the neighborhood environment that are viewed as social disorder tend to result in the categorization of individuals as orderly or disorderly. Orderly or law-abiding people are generally understood to be those residents who, when confronted with disorder, retreat into their homes to avoid criminals and victimization. Disorderly people, on the other hand, are those who see disorder as an opportunity to commit crime (see Harcourt 2001 for a discussion on how broken windows theory creates these categories). Unfortunately, the line between the orderly and disorderly is often hard to define, and these categories become especially troubling when we consider that residents may tend to see disorder through the lens of race and ethnicity.

This point is illustrated well through example. In *Off the Books*, Sudhir Venkatesh (2000) shows how residents of a disadvantaged community in Chicago earn an income through illicit activities. Illicit income may come from activities such as drug selling, repairing cars in alleyways, or selling stolen goods from a van near a neighborhood park. Clearly, these activities are illegal and would draw police attention. But do the residents of the neighborhood see it that way? These individuals offer goods; some of those goods can be seen as necessities, like socks or beauty supplies, at prices that enable the residents to buy them where they otherwise could not. Are the residents seeing disorder? Most likely not. Are the police seeing disorder? Most definitely, it is these kinds of activities that broken windows policing explicitly targets, with the expectation that taking disorderly people off the streets will prevent serious crime from happening (Wilson and Kelling 1982). But in this case, the individuals selling stolen goods, while clearly engaged in illegal activities, are merely making money; additionally, they are providing low-cost goods to residents. Thus, rather than being disorderly, they are viewed as a positive economic force in the neighborhood. Again, the line between disorderly and orderly blurs.

One of the unique aspects of the broken windows theory is its temporal structure – disorder breeds serious crime. Apply this to individuals (i.e., social disorder), and the theory would suggest that disorderly individuals will either themselves will commit serious crime or attract other criminals. Policing disorderly individuals, in theory, effectively removes serious criminals from the street. In the process, disorder becomes a degree of crime: breaking a window, littering, and jumping a turnstile become grades along a spectrum that leads to homicide. And since disorder theory and its implementation defines who is orderly and who is disorderly, this, combined with its temporal aspects, has a deleterious, long-standing stigmatizing effect on individuals (Harcourt 2001, p. 150).

This is especially problematic in communities were disorder does not represent "criminal activity" the way the police or nonresidents would think of it, specifically in communities where illegal activities (i.e., drug selling, prostitution, or panhandling) are a means to an end for daily subsistence (St. Jean 2007; Venkatesh 2006) or when neighborhood social networks are interwoven with individuals who are criminally active (Patillo 1998; Pattilo-McCoy 1999). The policing of disorder removes breadwinners and the protective agents of the community – furthering the dire economic conditions in the community and making the illegal economic activities that spurn policing more steadfast in the neighborhood (Rose and Clear 1998). Here, policing disorder and not acknowledging what disorder means to community residents redefine residents in the eyes of law enforcement and even mainstream society. Residents in these communities, therefore, are looked upon as hard-core criminals in the making, in need of policing and punishment. Regardless of individual variation in perceptions, policing disorder redefines individuals in disorder communities as disorderly individuals, rather than residents. Creating disorderly subjects, coupled with variation in individuals' perceptions of disorder, becomes even more nefarious because policing disorder in these communities does not reduce disorder; rather, it reinforces the need for policing.

Seeing Disorder

One of the primary reasons why theories of disorder have gained so much traction is their simplicity; generally, the presence of disorder in neighborhood signals criminal opportunities and a lack of social control (Wilson and Kelling 1982). The common sense appeal of disorder theories is evident. We have all been in "bad" neighborhoods, looked at our surroundings, and felt certain of the potential for crime and scared for our safety. It is this simple association between the presence of disorder and subsequent knowledge of neighborhood conditions in the minds of individuals that keeps disorder theory at the forefront of criminology and urban sociology. As a result, disorder theories are able to use "disorder" as a proxy for informal social control, neighborhood withdrawal, or other neighborhood characteristics related to crime or disorganization.

Yet, an important ingredient in the association between disorder, neighborhood outcomes, and crime is often assumed: residents must perceive, interpret, and act on disorder similarly in order to affect the neighborhood-level outcomes disorder is associated with. This assumption is made not by individuals but by academics. Disorder is at the core of neighborhood social dynamics because it represents deviance from prescribed norms about neighborhood like. Burisk and Grasmick (1993) suggest that all that is needed for neighborhood social control is the shared notion of a life that is free of personal harm. Disorder, it is assumed via its definition, is the visual reminder within a neighborhood that safety is not present. But does disorder uniformly represent this to individuals?

The broken windows theory will be used as an example of this, though it is certainly not the only disorder theory that assumes this process. The broken windows theory suggests that disorder leads to crime in a neighborhood because of resident withdrawal (Wilson and Kelling 1982). That withdrawal is based in residents perceiving their community as unsafe or that no one cares about what happens in the neighborhood and, perhaps most importantly, no one cares about what happens to *them* in the neighborhood. Because of what disorder signals, residents are afraid of potential victimization and left uncertain of help in a crisis; therefore, they restrict their activities in a neighborhood (Wilson and Kelling 1982).

Once withdrawal happens, informal social control in a neighborhood is diminished; it is this process that the broken windows theory suggests lead to crime.

Rested in the above process are three components related to individuals, though which are rarely discussed and often assumed: (1) perceptions of disorder cues, (2) how individuals interpret disorder cues, and (3) how people act on their interpretation of disorder cues. Each of these steps is crucial to the link between disorder and crime, though, to date, they have been left relatively unaccounted for by disorder theory. Additionally, each of these steps builds upon each other; therefore, we begin with perceptions.

For residents to ultimately restrict their involvement within a neighborhood, they first need to "see" the disorder around them. In much disorder research, there is an assumption that people see only the disorder that is present; more specifically, people only perceive the objective levels of disorder in their environment (Ross and Mirowsky 1999). Most researchers report that there is little within-neighborhood variation of disorder perceptions; in other words, people in the same neighborhood see disorder similarly (Skogan 1990). Yet, given the results of some recent studies, this is clearly not the case (Hipp 2010; Sampson and Raudenbush 1999, 2004; Wallace, in press). Sampson and Raudenbush (1999) report a 0.6 (approximately) correlation between subjective and objective measures of disorder; in another study, their results show within-neighborhood variation in disorder perceptions (Sampson and Raudenbush 2004). Even when constraining the ecological unit to about the size of a city block, effectively controlling for exposure to difference sections of a neighborhood, Hipp (2010) also found variations in disorder perceptions. In essence, research has clearly demonstrated that disorder perceptions vary. Researchers, with few exceptions (Taylor 2001), have yet to fully theorize why variation in disorder perceptions is so problematic for disorder theory: when people see disorder differently, it suggests that disorder is interpreted differently.

Just as with perceptions, for disorder to negatively affect residents' behavior within their neighborhood, to be in accordance with the broken windows theory, they first must see disorder and then interpret it as threatening. A great deal of research suggests that the presence of disorder in a neighborhood makes residents feel uneasy and at risk for victimization (Kelling and Coles 1996; Skogan 1990; Wilson and Kelling 1982). Hunter (1978) suggests that individuals interpret disorder to mean that both citizen groups and public agencies within and outside of the neighborhood cannot maintain satisfactory conditions within the neighborhood and subsequently residents feel vulnerable. When disorder is employed in sociopsychological studies, most studies suggest that individuals interpret disorder as the potential for crime and victimization, which generates fear and stress thereby causing negative personal outcomes, such as reduced health and well-being or increased depression and mistrust (Ross 1993; Ross and Jang 2000; Ross and Mirowsky 2001). While the link between disorder and fear of crime is established, it is unclear if individuals' interpretations is what generates fear, especially given that fact that individuals often are unable to distinguish disorder and crime (Gau and Pratt 2008) and researchers often conflate the two in disorder measurements (Harcourt 2001). As research shows that perceptions of disorder are not universal, it is unreasonable to assume that *interpretations* of disorder are universal as well.

Finally, we reach the third step in the process where perceptions and interpretations of disorder combine to impact individuals' behavior. If individuals perceive disorder and interpret it in a threatening way, the broken windows theory, as well as other theories, suggest they begin to withdraw from neighborhood life and restrict their territoriality (Skogan 1990; Wilson and Kelling 1982). Of course, this is predicated on an interpretation of disorder that will elicit such behavior. Should disorder be perceived and interpreted by individuals in any other light, their assumed behavioral reaction does not emerge. Indeed, there are several urban ethnographies that show that people do not necessarily withdraw from disorder. For example, in *Black Picket Fences*, Mary Patillo shows that neighborhood residents interact and work with local gang members selling drugs to arrange places and times in the neighborhood were drugs can be sold that do not effect neighborhood life, such as basketball games or use of a playground. In *Off the Books*, residents often use services that in other neighborhoods would be considered disorderly, such as fixing cars in alleys or panhandling, because the services are economical and a means to help other residents get by (Venkatesh 2006). In sum, once variation in perceptions or interpretations of disorder does not cause individuals to withdraw from their neighborhood or be fearful, the link between disorder and neighborhood outcomes, such as crime or informal social control, is broken.

From the delineation of the above steps, it is easy to see how *any* kink in the three steps is a serious threat to validity for the various theories of disorder. There have been calls for research to address the above points (Taylor 2001), but only recently has scholarship begun to study them. To date, current research has established that there are variations in disorder perceptions, though has had little success in explaining why they occur. At best, we know that variations in perceptions of disorder are not primarily a result of neighborhood exposure or use (i.e., one's routine activities) but perhaps partially due to how individuals employ racial stereotypes that are attached to place (Sampson and Raudenbush 2004). Unfortunately, this explanation does not cover all the variation we see; for example, women, those highly educated, and married perceive more disorder, while older adults and non-Whites perceive less (Hipp 2010; Sampson and Raudenbush 2004; Wallace, in press). What we are left with is variation in perceptions without explanations. In essence, until we know how people see disorder, we are left uncertain of the effects disorder has on individuals and neighborhoods.

Race and Disorder

As noted in the previous section, research has shown that disorder perceptions are imbued with race (Sampson and Raudenbush 2004). The stereotype of Black Americans as violent and criminal is both well established and common (Eberhardt et al. 2004), and it should be no surprise that this stereotype has bled into the perception of place (Wacquant 1993) given segregation and slavery (Loury 2002). As a result, individuals' perceptions of disorder are filtered through "stigmatized groups and disreputable areas" (Sampson and Raudenbush 2004).

There is both direct and indirect evidence of this. First, when testing whether neighborhood racial characteristics influence disorder perceptions, Sampson and Raudenbush (2004) show that even when taking objective levels of neighborhood disorder into account, there is a positive association between disorder perceptions and percent of Black in a neighborhood. In essence, the more Black faces people see in their community, the more problematic they perceive disorder. As a result, they suggest that "residents supplement their knowledge (of disorder) with prior beliefs informed by the racial stigmatization of modern urban ghettos" (p. 336). As a consequence, Sampson and Raudenbush (2004) express concern about using the broken windows theory to inform disorder reduction strategies: if disorder perceptions and fear are racially motivated, reducing disorder will not reduce fear given that fear is being produced by another visual cue – race.

We also see evidence that race is involved in disorder perceptions via policing literature. The broken windows theory has motivated a type of policing strategy called order-maintenance policing that targets disorder cues, particularly social disorder cues, as a way to reduce serious crime in the neighborhood. The reasoning will not be repeated here as it is part of early sections, but there is an inherent problem with policing disorder, namely, the act of subject creation. When policing disorder, we run the risk of moving away from thinking that the act of loitering, for example, is a crime and toward thinking that the loiter is criminal. Harcourt (2001) expands on this as such: "Rather than judging the act of loitering, we may attribute to the person who is loitering certain propensities – certain tastes, attitudes, and values" (p. 163). Couple the issue of subject creation with the evidence of the racialization of disorder perceptions and making policing disorder in a race neutral way becomes difficult, if not impossible. The broken windows policing creates racial animus: "The prominence of race in the decision to stop citizens may not rise to the threshold of racial profiling, but it does seem to create a racial classification of 'suspicion'" (Fagan and Davies 2001). Bringing about order to a "disorderly" neighborhood becomes impossible when the neighborhood has a minority racial composition; in effect, removing disorder will not remove the association of disorder with race. So long as black and brown faces remain, the neighborhood will be seen as disorderly regardless of the amount of disorder present.

The Paradox of Disorder

One aspect of disorder that has been continuously demonstrated by urban ethnographers and systematically ignored by disorder theorists is when residents see disorder as a positive force, or rather, as *order*. As discussed above, in disadvantaged neighborhoods, disorder has become an adaptation while still remaining disorder in the eyes of authority. Making order from disorder is a process aligned with the durability of social isolation and segregation. Generations of residents of disadvantaged neighborhoods are exposed to deleterious neighborhood conditions (Wilson 1987, 1996). It is this exposure that enables disorder to be viewed in a different light – rather than being a neighborhood problem, it is a form of daily survival (Anderson 1999) or a means of making money (St. Jean 2007; Venkatesh 2006).

A ripe example of this process is how some components of disorder were normalized in the Robert Taylor Homes housing project in Chicago (Venkatesh 2000). In *American Project*, Sudhir Venkatesh details life in a housing project in Chicago. Venkatesh shows how the identity of the housing project shifted over the course of 40 years. At the beginning, residents of housing projects saw their living environment as a move away from the decay of the surrounding areas. Where they once were sharing kitchens, now, residents had their own, self-contained apartments. Residents worked toward betterment of the community and the project by forming formal and informal groups that dealt with issues of safety, quality of life, and amenities in the projects. Yet, as time passed, the city and the Chicago Housing Authority (CHA) slowed the resources available to the projects. The efficacy residents once had in dealing with institutional forces faded – any solution to their housing problems had to come from their hands. With less social power, the solutions residents offered would not pull them out of poverty or their environment but only offer to make that environment safer or more negotiable. For instance, a group of powerful women in the Robert Taylor Homes created a law enforcement body of men who were staying with other residents and were not listed on the lease. (As a condition of their lease, residents were not allowed to have any individuals stay with them that were not listed on the lease.) In order not to be reported to the CHA, these women would often call upon the men to deal with problems, like domestic violence or snitching to the police (Venkatesh 2000). Soon, even this efficacy faded. Hustling – or making money through illicit means – became prominent and soon dominated public space. And as soon as hustling gained a foothold, residents stopped seeing disorder, like loitering or panhandling, and started to see hustling. This process was a direct result of opportunities and resources being pulled from the housing projects, leaving residents to cope with their situation in any way they can.

In disadvantaged communities, the economic situation so depressed and isolated that illicit economies provide one of the few reliable sources of income (Anderson 1999; Venkatesh 2000, 2006; Wilson 1996). Often conducted in public space, illicit economies often take the outward shape of social disorder. When occurring in public, illicit economies are not simply a panhandler or a drug dealer; they are richer and include more than the simple notions of what the classic definition of disorder leads us to imagine. Venkatesh (2006) paints an in-depth picture of an illicit economy working in a neighborhood park on the south side of Chicago:

> Pimps brought their sex workers to an abandoned building near the park. Carliss, a car mechanic, moved his outdoor 'Oil and Tire Change' operation to the alley next to the park's basketball court. Two gun brokers come to a nearby abandoned building once a week to see handguns and pistols. A few men sold stolen car stereos, guns, and other electronic equipment from the back of two beat-up beige vans that were always stationed at the park entrance. Mo-Town, the local hot dog vendor, and Charlie, who sold stolen cigarettes and beauty products, set up their respective carts at the edge

> of the park. And now the drug sales, as Big Cat had promised, round the clock. All of this was secured by placement of Big Cat's rank and file around the area: all were armed, they physically searched and harassed passersby, and they drank and smoked marijuana until the early morning hours with loud music blaring from their stereos. They also charged a fee to each entrepreneur based in and around the park. (Venkatesh 2006, p. 71)

Here we see economic opportunities come in many forms, forms often associated with disorder. In disadvantaged neighborhoods, the illicit economy is pervasive and often condoned by residents, regardless of whether they are in or out of the licit economy (Venkatesh 2006). Condoning does not mean acceptance of the illicit economy for many residents, though they understand the necessity of it. When working on public and accessible public space in the neighborhood, residents understand attacking the illicit economy means undoing someone's income: "residents may weigh delinquent activity that has an economic dimension differently than, say, crimes of passion like domestic violence and assault. This does not mean that all underground activity is tolerated. But of the activity generates income, any ethical dilemmas it creates must also be judged in terms of how the activity supports a household and ever the wider community" (Venkatesh 2006, p. 74). People turn a blind eye to the drug trade in inner-city communities: "people residing in the drug-infested, depressed inner-city community may understand the economic need for the drug trade" (Anderson 1999, p. 132). People only become concerned with the drug trade when it results in violence – then their intolerance comes to the surface. Now, illicit economic endeavors that are, to an outsider, disorderly are not seen in that light for neighborhood residents. Even when these activities are not welcomed, they are seen more as a means to an end than as disorder. The outward publicness of illicit economy shifts the disorderly aspects of its workings to a more complex problem of economic survival of individuals and communities. Hence, disorder is not linked to fear of crime, neighborhood decay, or crime; rather, it is rather the opposite: livelihood.

The handful of concerns about defining disorder that we have been able to address in this limited space – and there are many more – should make scholars wary of using the concept of disorder in its current form. Disorder is long overdue for reconceptualization, and careful scholarly work should aim to do just that.

Recommended Reading and References

Anderson E (1999) Code of the street: decency, violence, and the moral life of the inner city. W. W. Norton, New York

Eberhardt JL, Goff PA, Purdie VJ, Davies PG (2004) Seeing black: race, crime, and visual processing. J Personal Soc Psychol 87(6):876–893

Fagan J, Davies G (2001) Street stops and broken windows: terry, race, and disorder in New York City. Fordham Urban Law J 28:457–504

Gau JM, Pratt TC (2008) Broken windows or window dressing citizens' (in) ability to tell the difference between disorder and crime. Criminol Public Policy 7:163–194

Harcourt BE (2001) Illusion of order: the false promise of broken windows policing. Harvard University Press, Cambridge, MA

Harcourt BE, Ludwig J (2006) Broken windows: new evidence from New York City and a five-city social experiment. Univ Chic Law Rev 73(1):271–320

Hipp JR (2010) Resident perceptions of crime and disorder: how much is "bias", and how much is social environment differences? Criminology 48:475–508

St. Jean PKB (2007) Pockets of crime: broken windows, collective efficacy, and the criminal point of view. University of Chicago Press, Chicago

Kelling GL, Coles CM (1996) Fixing broken windows: restoring order and reducing crime in our communities. Simon and Schuster, New York

Kubrin CE (2008) Making order of disorder: a call for conceptual clarity. Criminol Public Policy 7(2):203–214

Loury GC (2002) The anatomy of racial inequality. Harvard University Press, Cambridge, MA

Patillo ME (1998) Sweet mothers and gangbangers: managing crime in a black middle-class neighborhood. Soc Forces 76(3):747–774

Pattilo-McCoy M (1999) Black picket fences: privilege and peril among the black middle class. University of Chicago Press, Chicago

Ross CE (1993) Fear of victimization and health. J Quant Criminol 9(2):159–175

Ross CE, Jang SJ (2000) Neighborhood disorder, fear and mistrust: the buffering role of social ties with neighbors. Am J Commun Psychol 28(4):401–420

Ross CE, Mirowsky J (1999) Disorder and decay: the concept and measurement of perceived neighborhood disorder. Urban Aff Rev 34(3):412–432

Ross CE, Mirowsky J (2001) Neighborhood disadvantage, disorder, and health. J Heal Soc Behav 42(3):258–276

Rose DR, Clear TR (1998) Incarceration, social capital, and crime: implications for social disorganization theory. Criminology 36:441–480

Sampson RJ, Raudenbush SW (1999) Systematic social observation of public spaces: a new look at disorder in urban neighborhoods. Am J Sociol 105(3):603–651

Sampson RJ, Raudenbush SW (2004) Seeing disorder: neighborhood stigma and the social construction of "broken windows". Soc Psychol Q 67(4):319–342

Skogan WG (1990) Disorder and decline: crime and the spiral of decay in American neighborhoods. University of California Press, Berkeley

Suttles GD (1972) The social construction of communities. University of Chicago Press, Chicago

Taylor RB (2001) Breaking away from broken windows: Baltimore neighborhoods and the nationwide fight against crime, grime, fear and decline. Westview Press, Boulder

Venkatesh S (2000) American project: the rise and fall of a modern ghetto. Harvard University Press, Cambridge

Venkatesh S (2006) Off the books: the underground economy of the urban poor. Harvard University Press, Boston

Wacquant LJD (1993) Urban outcasts: stigma and division in the black American ghetto and the French urban periphery. Int J Urban Reg Res 17:366–383

Wallace D (in press) A test of routine activities and neighborhood attachment explanations for bias in disorder perceptions. Crime Delinq

Wilson WJ (1987) The truly disadvantaged: the Inner City, the underclass, and public policy. The University of Chicago Press, Chicago

Wilson WJ (1996) When work disappears. Knopf, New York

Wilson JQ, Kelling GL (1982) Broken windows: the police and neighborhood safety. Atl Mon 127:29–38

Delinquency

▶ Network Analysis in Criminology

Democratic Policing

Peter K. Manning, V. H. Elmer and
Eileen M. Brooks
College of Criminal Justice, Northeastern
University, Boston, MA, USA

Synonyms

Cause of police legitimacy; Concept of the police; Comparative police systems; Organizational change and police legitimacy; Origins of modern policing; Policing in developing democracies; Role and function of the police

Overview

The study of democratic policing might well be seen as the answer to the question "What are the police good for?" That is, what good are they expected to do in a democratic state? It surely cannot be dismissed by arguing that this required good is solely or even primarily that it must produce "crime control." Police are a fundamental force in a democratic society in which in the course of their practice they are obligated, as the Patten Report states, to "perceive their job as the protection of human rights." Furthermore, their obligation is collective and shared and cannot be judged in that sense by ad hoc criteria that reflect narrow interests and political positions. They reflect, at best, what is valued, or regulate social obligations we have to one another. The role of the police in democracy is variable over time and culture, not a constant. Police may enhance democracy by sustaining employment, security, and democratic procedures such as voting and demonstrations; they may decrease its viability by corrupt practices and collusion to destroy democratic competition; and, ideally, they may be models of decorum, propriety, and restraint. As a conservative force in general, of course, it is most likely that their role at best may be neutrality on behalf of the state. Why and how they contribute to these necessary functions has not been systematically addressed. Because such questions have not been raised in regard to stable democratic states, and the role of the police in nurturing democracy has rarely been explored, much preliminary ground work is needed. The primary impediment to serious scholarship is the addiction to the false idea that above all else police must "control" crime and be assessed by that toxic criterion. To discuss democratic policing, it is necessary to trace briefly the development of the field of police studies and its failures to address what is democratic about policing in a democracy. This will permit a definition of police, the identification of the features of Anglo-American policing, the application of criteria for democratic policing derived from the work of John Rawls, and the discussion of some issues requiring future research.

The Development of the Field of Police Studies

The systematic study of policing in North America began in the early part of the twentieth century as a result of the pioneering work of Raymond Fosdick, Bruce Smith, and later, Augustus Vollmer. They approached the topic as reformers, as students of law and public administration, and were pragmatic men of action. Later, O.W. Wilson, who approached the study of policing as a student of public administration, served as Commissioner of Police in Chicago. He later wrote the first book on police administration, thus inventing the term demarcating the field. It is generally accepted that the sociological study of policing was initiated by the fieldwork-based study of policing in Gary, Indiana, in the ‘50s by a University of Chicago graduate student, William Westley. Westley later published this study as *Violence and the Police: A Sociological Study of Law, Custom and Morality* (1971). Some 20 years after Westley' s initial field work, the results of which were known as a result of a series of published brilliant papers, books written by social scientists Michael Banton, James Q. Wilson, Jerome Skolnick, and Arthur Niederhoffer set out the questions still shaping the field. They were in many ways looking at policing from the bottom, from the perspective of the patrol officer or constable. A book of lasting importance, the *Democratic Policeman* by George Berkeley (1969), was overlooked perhaps as a result of the growing concern in the late ‘60s with riots, disorder, and crime. This book, based on comparative analysis of policing in several European countries, was the first to address the principles of democracy that should ideally be reflected in police administration.

At this time, systematic social science addressed at least some of the lasting concerns of political philosophers from the time of Plato: What are the requirements for a fair, just, even-handed police? Who watches the watchers? From the Greeks onward, this question was couched implicitly at least in the context of democracy as a compelling system of governance. The question raised by Plato and Aristotle was again vibrant and commanding. Here, we encounter an oddity: as lasting as this concern has been, no tight definition of "democracy," "police," or "democratic policing" has been accepted. Political scientists, economists, and conservative apologists bypassed this thorny issue, and moved directly to impassioned advocacy of the idea. Here, policy and advocacy collide with empirical analysis and open-minded exploration of data. It has been known, for example, that some democratic policing practices can be found in nondemocratic societies, and that nondemocratic practices can be found in democracies, especially in times defined as crises. The assumption has been that police are neutral on behalf of the state, committed only to "law enforcement," and loyal. In a stable, economically sound society, the following questions might be considered: Loyalty to what principles? Action on behalf of any government? Commitment to any law, regardless of its impact? Available in any crisis, including a revolution? Preserve and protect what? Whose rights, property, and vested interests? In many respects, policing grew within large European cities as a kind of order-maintaining force, a source of steady employment, and an unpretentious and often violent governmental resource.

In part, this blindness to the broader parameters of policing is due ironically to the success of the political efforts of Vollmer and his students between the two great world wars. They sought diligently to encourage and persuade the public to accept the police as a nascent profession based on science (social, biological, and chemical), carried out by well-educated officers and experts, and focused diligently and relentlessly on chasing and jailing criminals. "Crime" rose to the surface as the "problem" for which they were designed. Ironically, given his later infamy, J. Edgar Hoover was among the brilliant innovators who shaped most "professional policing" by lobbying for new federal laws, opposing corruption, developing a statistical capacity within the FBI, and tightening the administrative controls over agents. The standardized appearance, conservative suit, white shirt, dark tie, black shoes, and hat were no small part of his ingenious impression management. This leadership at the federal level was important as a symbolic brand and image for policing in general. As such, the organization argued successfully that they were deserving of respect, honor, and decent wages consistent with their skills and duties. This position was augmented some 30 years later by the US government's creation of the Law Enforcement Assistance Administration (LEAA), later to become the National Institute of Justice within the Justice Department, a source of research funding, technical information support to police departments, and expert knowledge. With the publication of the report of the President's Crime Commission (PCC), several ideas emerged that were to have considerable influence and staying power over the coming years. The commission argued that the police, courts, prisons and jails, and probation and parole were not disparate competing units with diverse interests, training, aims, and tactics, but rather were part of an integrated, coherent, meaningful, and orderly *criminal justice system*. The assumption they made that it was just and produced justice was unchallenged. The commission viewed this "system" as flawed, but nevertheless it was created and implemented by just and honest people; it only required a bit of reform. It took a legal realist position that the system produced legal outcomes on a case by case basis. Since legal processes were the source of justice, the system produced justice of a sort. This system, they argued, had failed for lack of resources and needed an infusion of money, material, especially scientifically based technology, and personnel. The reformers hoped that this system could be reshaped by the application of science and technology, and they placed special emphasis on the role of technology in fighting crime. An entire volume of the final report was devoted to the imagined role of science and technology. This was social engineering applied to the problem of crime defined by officially sanctioned, gathered, and presented data. If reformed, they felt, this system could manage and control crime. Again, "crime" became the featured player in the drama. The report argued for enormous investment in human capital and technologies, and the government of the day was prepared to invest. Finally, the Commission viewed education and training as needed to provide "criminal justice professionals." As a result of the report of the commission, legislators created federally funded colleges within several large state Universities that granted degrees in a new field called "criminal justice." In due course, Ph.D. programs were begun, and the field of criminal justice, highlighted by the field of police studies, emerged.

Broadening the Scope of the Field

The subsequent years saw a rise in officially recorded crime (ORC), protest, disorder, and rioting that in part were caused and exacerbated by violent police interventions in traffic, domestic disorder, and demonstrations. It was widely thought that policing was too soft, and the recommendations of the PCC and reforms of the Johnson and Nixon administration were failures. When later in the 1980s the concept of community policing arose as a panacea, its aims were local, parochial, and vague. They at best were tactical and adopted in spite of no evidence of their efficacy. It was not until the early 1990s that policing again emerged as a powerful force believed to serve, protect, and reduce the specter of crime. As a result of the claims of Commissioner Bratton of the New York City Police and a book written by his advisor and colleague George Kelling, a new view of policing emerged. The media validated the claim that "smart management" focused on arrests connected to "disorder" rationalized the process of crime control and resulted in reduced crime. Policing rode the headlines as a positive, rational scientific organization devoted to the general good. There was no question of its democratic qualities, procedures, or negative effects. This was to come later. In this media frenzy, the claims of the broken windows perspective and crime control were wedded and dominated research for almost 10 years. By the time that it was well established that the dip in crime could not be separated from other factors, including manipulating the standards for reporting, recording, and investigating crime, and the research revealed the empty claims of the broken windows perspective, the field had moved on to other concerns, other issues, and other sources of funding. There is no question that heavy concentration of officers in targeted areas can reduce crime in the short term of 6 months to a year, but this comes at costs in regard to staggering misdemeanor arrests, stops and jailing of minority youths, and questions of fairness and race profiling.

Concern with policing in the context of democracy and development began to escalate in the 1990s with mini-wars in the Balkans, the Middle East, and Somalia to which armed United Nations Forces responded. As David Bayley (2005, Bayley and Perito 2010) has argued, such low-intensity conflicts required policing of a new sort – policing that may serve as a model for other new nations; policing that is peace-keeping, somewhere between crime-based prevention and low-intensity conflicts; policing as an export commodity to be provided variously by cooperating nation states, private security companies such as Halliburton, Blackwater, and others; policing as a kind of nation-building instrumentality; policing as a basis for civic governance. Some claimed this was a moment of exporting a commodity, and policing was merely the next version of Coca-Cola and Mickey Mouse (Brogden and Nijhar 2005). These police-like operations were carried out experimentally in the sense that each new venture was undertaken by different nation states as suppliers of force; each economy and culture encountered was in uneven state of development; and there were no models for such operations. The closest analogy was not "peace-keeping," but suppression of rebellions against colonial powers in Malaya, Palestine, Kenya, and earlier in Ireland. Furthermore, there was a sense in which the policing mission as it unfolded was open-ended. New innovations had to be developed to cope with unexpected contingencies. In addition, sociopolitical changes captured the imagination of scholars and refocused interest in policing and new forms of policing. These included the growth of the European Union and its cooperative policing arrangements, emergent police force; globalization of crime, especially drug-related crime, and regional policing responses to this in Latin America and the Caribbean; growing police cooperation around globally significant meetings and occasions such as international economic summits, Olympics, World Cup soccer matches, and large-scale rioting; and not of least importance, inexpensive networked communications; travel and the growth of global transnational corporations. The insular and parochial view of the United States, of the United Nations, international law, and international cooperation did not reduce the movement of "police peace-keeping": the claim was that such activities were bringing democracy to the world via some new and malleable forms of policing.

Democratic Policing as a Problem Emerges

At about the same time scholars, including notably Professors Jerome Skolnick and David Bayley who were employed by the State Department, began to reflect, to consider, and reconsider the nature of "democratic policing." It was a new enterprise insofar as the actions of policing preceded the concern for rationalizing and defining the concept. Since the early '70s, however, scholars such as David Bayley, David Sklansky, Tim Newburn and Trevor Jones, Clifford Shearing, Hsi Liang, and George Berkeley were characterizing democratic policing. They were focused on programmatic and prescriptive features, rather than empirical features that could be assessed. There was a new and an abiding concern with justice and democracy in the context of policing. There was also a growing awareness that past scholarship on policing had been restricted to anthropological, sociological, or historical studies of policing, often crime or order based, that had *assumed* the democratic nature of policing. The field of what was to become police studies was an odd amalgam of Anglo-American ideas, somehow connected to the idea that "policing" was essentially an invention of Sir Robert Peel.

This idea ignored several powerful ideas of great importance. Peel also invented the most powerful, vicious, and violent gendarmerie in Anglo-American society, the Peace Presentation Force, an Irish Constabulary that morphed into the RIC and RUC. This innovation gave rise to the varieties of policing, including the Garda, various state policing systems in Australia as well as the Australian Federal Police, and the Royal Canadian Mounted Police (RCMP). As Sinclair (2006) has cogently argued, this was the model for all colonial policing throughout the British Empire. These forces had little concern with "prevention," and they wrought destructive effects of colonial, post-colonial societies, and emerging nations. Finally, these gendarmerie-like forces had enormous and insidious effects on the differential means of police accountability that arose in the Anglo-American world. These were ignored in conventional treatments of "policing." In summary, the assumption that "democratic policing" is unequivocally the case in North America obscured the varieties of policing that developed from the Peel model, the negative and powerful effects of this form of policing, and characterized nominally preventive policing mistakenly as the only consequence of his innovations. The textbook version of modern Anglo-American policing as a direct consequence of Peel's ideas is misleading and partial.

Also unexamined, except in crude asides like the claim that it is obvious that the primary role of the police is to combat crime, is the matter of studying policing's achievements beyond their struggles to manage crime. Criminologists have accepted the claims of the police and their professionalism and have focused largely on their crime management efforts. What else do they or can they do that might sustain or improve the quality of life in democratic societies? The prior question remains: why study police in a democratic context, or as a democratic institution? The police are one organization in a network of institutions that shape social life, such as religion, family, economy, education, and are but one mode of formal social control. They do not stand alone, nor can they control any aspect of social life unilaterally. They deal with failures; they are failure-processors who cope with the vagaries of other forms of social control. However, it is clear that they perform a variety of functions of significance. They increase the life chances and possibilities of the advantaged classes, the middle class and above, by overlooking, reducing, or cooperating to minimize their crimes; they increase the negatives that are lodged against the minorities, the poor, the disadvantaged of all sorts. This is the case, not because police are "racists," elitists, unsympathetic, or incompetent; it is rather as Bittner (1970) says that they are designated to cope with and manage society's difficulties, or its "dirty work" as Everett Hughes (1958) writes. Furthermore, police capacity to control crime is highly limited and reductions and rises are revealed in reports showing brief dips in official statistics. These are partial, misleading, and necessary. The mandate of the police in a democratic society cannot be reduced to crime control, although it remains central; their obligations extend well beyond this, and their role cannot be reduced to visible patrol functions.

To address the mandate, it is necessary to review events of the last 20 years. They suggest the following:

- Crime in North America is declining, and there is no consensus on the explanation for this.
- Policing is rising in status, in part because of the drop in crime and in part because its top command are articulate and well-educated. Their claims to professionalism have been validated.
- Terrorism, globalization, and international police cooperation are increasing awareness of the isolation of American policing from international trends of cooperation, exchange of information and personnel, and increased and more systematic training.
- High levels of immigration and migration within the United States make clear it is no longer an isolated island in the world.
- There are concentrated levels of violence within ghettoized sections of large American cities, and inequality and poverty are rising to the highest levels in 50 years.
- The drop in crime combined with increased inequality is inconsistent with conventional criminological theories.

These issues have brought to mind the need to reflect on the mandate of the police in the United States. The fads – community policing, broken windows policing, and policing as an export commodity – are exhausted, and the sponsored research field is quite dazed.

Police Defined and Policing as an Organization Outlined

The previous discussion of the putative functions of democratic policing assumes the need for a fairly precise definition. Unfortunately, most definitions of policing in Anglo-American societies are loose or misleading, as has been argued, insofar as they focus on crime control to the exclusion of other functions. Furthermore, this begs the questions of the negative consequences of reduced crime and/or the failure to carry out other functions. What indeed are the negative consequences of policing as it is represented currently? It is important that a definition contains their violence potential, the organized character of their practices, their local character, and their focus on order and ordering. Both public and private police exist and operate under the canopy of the law. Finally, police require compliance, within their organization and from citizens, and in a democracy this is essentially limited by the notion that civil rights are universal and procedural guarantees present. Ironically, as Egon Bittner has argued, police in Anglo-American societies are given an enormous range of freedom to deal with the unexpected matters arising and others matters they judge may escalate and cause even greater damage to the social fabric. Manning (2010: 68) has proposed this definition of policing:

> Police organizations in Anglo-American societies, constituted of many diverse agencies, are authoritatively coordinated and legitimate. They stand ready to apply force up to and including fatal force in politically defined territories. They seek to sustain politically defined order and ordering via tracking, surveillance, and arrest. As such, they require compliance to command from lower personnel and citizens, and the ability to proceed by exception.

From this general definition, it is necessary to confine subsequent discussions to the public police. Private police are anomalous insofar as they are not paid to be fair. It is not a small matter that public police are permitted to proceed by exception because their mandate is elastic, negotiated, and local in large part. The mandate is the source of legitimacy and it is a kind of negotiated contract between the several publics they serve, not a reified single "public," and their own definition of the nature of the work and its obligations. It therefore expands and contracts over time and changes in response to public demands and legal constraints. The mandate in a democracy is fraught with contradictions because of the various expectations of police – service, crime control, order management, control of traffic, and demonstrations – and the shifting scope of their practices. That is, the police in a democracy are subject to changes in public opinion, law, political decisions, and media, and their elasticity is essential. In part because of this the democratic police have been most successful when they have convinced the middle class "respectable public" that their primary role is to control or at least manage officially recorded crime. The crimes of concern are of course what might be called "decent nineteenth century street crime": homicide, burglary, robbery, and assault. And in the late twentieth century, this includes policing "drugs." In many respects, this is code word for the patrolling the life styles of the poor in large cities. In many respects, this is an impossible mandate because of the limits of official modes of crime control. In fact, public trust in the police and maintenance of the public trust are probably the most significant "product" of democratic police. They assess trustworthiness in others; they monitor sources of distrust, both in groups and individuals; they are a repository of trust by citizens in their symbolic expressive role. It is for this reason that corruption, crime, and self-serving actions of police are judged more harshly than the actions of other citizens. Finally, policing in Anglo-American societies is less violent and more restrained in spite of periodic eruptions and killings. This suggests at least a rethinking of the Bittnerian definition of policing (Brodeur 2010).

Eleven Features of Anglo-American Policing

Given this definition and outline of the essential features of democratic policing, there still remains what might be called the constraining functions of democratic policing in any society. By stating the problem in this way, the scope of this discussion can be widened to include problems associated with the transformation of police systems to democratic modes of policing. Here is a brief list of proposed features (Manning 2010: 65–66) that can be used to judge police practices:

1. Police are a public entity and are accountable directly to the people.
2. Individual citizens, their demands, or requests cannot alone guide the organization because its obligations are collective and general, not individual. Callers are not "customers" or clients because they cannot choose to refuse the "service" or opt out of paying for it. There is no free market in police services.
3. Police cannot eschew their symbolic role as representatives of governance and governing. This is central to the belief in fair practice within a democracy.
4. Police violence is generally cautious and limited. Media amplification serves to confuse viewers and produce periodic crises in large cities.
5. Political territory of policing is problematic and cannot be restricted in advance to the limits of a legally defined territory. This ambiguity has increased as a result of international police cooperation; concerns about terrorism and its global dimensions and nonnational origins; policing under the guise of foreign aid and rapid inexpensive travel.
6. Police manage order: they do not produce it and may destroy it.
7. Police are not a neutral political force; they act on behalf of the state in crises.
8. Compliance with command and loyalty in the face of danger are essential to consistent policing, but they are always problematic.
9. Police are highly respected for carrying out dangerous work and given wide tolerance for their actions.
10. Police tracking and surveillance of citizens are increasing and largely unknown and unmonitored. This is concentrated in lower class areas and on lower class groups, but not restricted to them.

In summary of these considerations surrounding a definition of democratic policing, one can conclude the following. The police organization as presently constituted in Anglo-American societies has evolved historically and these patterns are reflected in the structure, function, and image of the police. The development of the role on the ground is a function of the emerging concern for managing risky situations that transcend in meaning the current on-view encounters and entail some form of risk management (Ericson and Haggerty 1997). In the police world, this means gathering "intelligence," information gathered prior to an event, mining data bases, compiling records from other organizations such as insurance companies and schools, and developing even more forms of information gathering and analysis. This means that the organization has evolved a heavily loaded observational role for the patrol officer, granted a wide range of choice for the officer in the event, and has gathered the general support of tradition and courts for trust in the observational and intuitive skills of such officers. Variations in the pattern of sanctioning are a function of the sanctions available as well as the targets seen as sources of uncertainty. These are the shifting targets of police when they act proactively. As nation states develop, sources of conflicts move from local to national and transnational, and these require a kind of reasoned neutrality in practice that may violate local norms and expectations for conflict resolution. The police in a democracy are double coded as both the source of violence and a protection against it. This is complicated by transcendental rules and norms propagated by the nation state. Police are at the same time no longer obligated to kinsmen and their local norms and practices and obligated to other orders – feudal loyalty, state, or legal loyalty. If we think of policing as encoding uncertainty, that is, responding to what it implies for order, then the job is to render manageable uncertainty in terms of continuous procedures of some kind. This, in turn, engenders trust in the police. In addition, as historical studies suggest, continuity in the job as a full-time paid and responsible agent, associated with continuity of procedures, also implies continuity in the office or role and its functions. These functions are in modern times associated strongly with the Weber (1947) idea of a rational bureaucracy, even though functionally police are punishment centered, not rational-legal in their operation. The connection of public police with some form of accountability and direction in the interests of the "people" seems a tacit expectation of democratic police, but not of policing in general, nor policing in fact. This question of accountability raises another one – how do the police sustain their role as a neutral arbitrator of conflicts? In other ways, the police are transducers or means of converting one sort of fact into another; they process facts into information: facts placed within a context. In some sense, they must display authority whilst being rooted in the everyday lives of citizens. The police mediate and manage conflicts in terms inconsistent with the ways in which they are conceived by citizens.

These ideas are putative and descriptive; they do not say what is actually done, why, or for what reason. Let us now consider democratic policing more specifically.

Policing in Accord with the Difference Principle

It is necessary to step back from the historical development of police, the faddish nature of police studies, and the drive for police professionalism to ask what analytic principles might guide judgment of police practices. It is police practices, not public attitudes, "transparency," or "accountability" that should command attention. Monitoring and legal constraints do not alone change police practices. The most important statement of political philosophy in this century is the work of John Rawls as found in his *Theory of Justice* (1970 TJ) and a condensed version of the argument entitled *Justice as Fairness* (2001 JF). These works presented two fundamental principles of justice. The first claims that all positions in general, and in this case the police, should be open to all on the basis of competition and the second claims that whatever passes for policy should be based on the difference principle. The difference principle can be summarized as stating that any implicit or explicit policy affecting extant inequalities should be to the benefit of the least-advantaged (JF: 42–43). Since it is unlikely that the police can actually reduce inequality, and in fact have no obligation to do so, a political philosophy with policing in mind and based on the justice as fairness idea requires some modification. One might argue that the justice as fairness principle in regard to policing should be refocused in a manner consistent with the Hippocratic Oath: the police should strive to minimize harm. The working version of this abstraction in regard to policing means that any action, planned stated, or enacted, should not increase inequalities. How can this grand working principle be grasped as a set of objectives or guide lines? Expectations of policing are questions of function: if the below principles or rules of thumb are observed, one might expect of (domestic) democratic policing that it function:

- Constrained in dealing with citizens and fair in procedure. These dealings should entail a degree of civility in interactions and in police practices. This excludes under virtually all conditions, torture, mass detentions, "round ups" based on political beliefs, not behaviors, and lengthy suspensions of habeas corpus for citizens.
- Largely reactive to citizens' complaints – reticent rather than sporadic – and not given to frequent secret proactive interventions, crackdowns, sweeps, and militaristic "operations."
- Equal in its application of coercion to populations defined spatially and temporally. The level of coercion is based on minimalistic criteria, much as counter insurgency tactics, rather than a mechanistic "use of force continuum."
- Fair in hiring, internal evaluation, promotion and demotion, transfers, and disciplinary treatment of employees, officers, and civilians.
- Competitive in an environment which includes private police, vigilante groups, posses, ad hoc policing under the guise of "self-help" and revenge. It may include the National Guard

and the armed services (army, navy, coast guard, and the air force). This implies formal and informal modes of cooperation rather than unified and unrelenting actions.

- Accountable and responsible for their actions individually and organizationally.

Finally, such broadly based ideas are inconsistent with the echoes of free market ideology that the police should be efficient, and effective. If efficient means the best usage of resources within the organization, it is clear that police use modest budgeting tools, are locked into invariant strategies (answering 911, random patrol, and investigating crime) and tactics (ways of delivering such services), and in fact have an open-ended remit with regard to overtime pay and resources whenever a crisis occurs. Add to this the facts that sick leave, disability, and days off are contractual, budgets are determined by nonmarket-based criteria, and not in the hands of police supervisory ranks. It is virtually impossible to fire a police officer short of criminal conviction. They cannot be efficient in the nature of the mandate. Effectiveness assumes a criterion against which functions and costs can be judged. If this is defined, for example, as the percentage of known crime for which police make arrests or charges, and rewards were made for more arrests, the consequences would be unacceptable in a democracy as being unfair in regard to risks to the less-advantaged classes given current laws and regulations.

Conclusion

Democratic policing is a concept in need of specification. Police studies throughout the twentieth century remained focused on American or at least Anglo-American police organizations. The onset of peace-keeping in low-intensity conflicts created some reflection on the nature of policing and its mandate and raised questions about the conventional concern for crime management as a paramount function. In this entry, police organizations are defined and their features enumerated. These led us to examine the governing ideas of democratic policing. These are predicated on the Rawlsian theory of justice with emphasis upon the difference principle. Finally, a list of six expectations was outlined by which policing practices could be judged. Efficiency and effectiveness are not among them.

Recommended Reading and References

Bayley D (2005) Changing the guard. Oxford University Press, New York

Bayley D, Perito R (2010) The police in war. Lynne Reiner, Boulder/London

Berkeley G (1969) The democratic policeman. Beacon, Boston

Bittner E (1970) The functions of police in modern society. USGPO, Washington, DC

Brodeur J-P (2010) The policing web. Oxford University Press, New York

Brogden M, Nijhar P (2005) Community policing: national and international approaches. Willan, Devon

Ericson RV, Haggerty KD (1997) Policing the risk society. University of Toronto Press, Toronto

Hughes EC (1958) Men and their work. Free Press, Glencoe

Liang H (1992) The rise of the police and the European state system from Metternich to the second world war. Cambridge University Press, Cambridge

Manning PK (2010) Democratic policing in a changing world. Paradigm, Boulder

Sinclair G (2006) At the end of the line: colonial policing and the imperial endgame, 1945–1980. Manchester University Press, Manchester

Weber, M (1947) The theory of economic and social organization. The Free Press, Glencoe

Westley W (1971) Violence and the police: a sociological study of law, custom and morality. MIT Press, Cambridge, MA

Design

▶ Innovation and Crime Prevention

Design Out Crime

▶ Designing Products Against Crime

Designing Out Crime

▶ Designing Products Against Crime

Designing Products Against Crime

Paul Ekblom
Design Against Crime Research Centre,
Central Saint Martins College of Arts and
Design, University of the Arts London,
London, UK

Synonyms

Design Out Crime; Designing out crime; Hot products

Overview

Much practice and research exists that addresses how the design of the built environment increases, or decreases, the risk of crime, particularly through the architectural approach of ▶ Crime Prevention Through Environmental Design (CPTED) and place-based ▶ situational crime prevention (SCP). But this entry covers the domain of *products*, essentially two- or three-dimensional objects that have been designed and manufactured in some way and which may be portable (e.g., laptops), mobile (e.g., cars), movable (e.g., home cinema TV sets), incorporated in another product (e.g., a tamper-evident lid for a medicine container), or installed in a place (e.g., a cash machine). There are however some transferrable concepts with architectural approaches: for example, certain products such as handbags or vehicles can be considered "enclosures" with equivalent "access control" issues to buildings. However, such transfer is often difficult because CPTED has developed its own terminological traditions which are not always clear or well connected with other crime prevention or design approaches.

The first section of this entry begins by defining key terms relevant to design against crime, such as risk and risk factors. It then reviews how the latter feature in situational crime prevention notably via the phenomenon of hot products, the underlying causes of elevated risk, and the risk life cycle of products. The second part covers the response to elevated risk, notably via intervention through design, covering both content and process; anticipation of future risks; and evidence of effectiveness. The important role of businesses in creating or reducing crime opportunities in manufactured products, and the difficulties of influencing their "design decision-making" to give some weight to security, is covered only briefly; more is in Ekblom (2012a) and Hardie and Hobbs (2005). The creativity and ▶ innovativeness of criminals themselves is well addressed in Cropley et al. (2010). The role of government in incentivizing and otherwise leading on design is discussed in Clarke and Newman (2005) and was exemplified in the UK Home Office's Design and Technology Alliance (see www.designcouncil.org.uk/our-work/challenges/security/design-out-crime/ for useful case studies).

Fundamentals of Risk and Risk Factors

Crime, unlike mercy, doesn't fall like a gentle rain evenly covering the land – it gathers in pools. Risk of crime is concentrated in particular ▶ places, on particular ▶ victims, and on particular products, the focus of this entry. This concentration has two kinds of implication. On the one hand, it gives strong clues about the causation of criminal events, whether concerning the targets or tools of those crimes or the insecurity of their immediate situation; on the other, it guides the kind of ▶ situational crime prevention strategy that can be adopted. That strategy can be developed either in reaction to an established pattern of risk or in anticipation, but in either case the underlying rationale is the same. If you – as policy maker, police officer, designer, manufacturer, or consumer – can identify the targets and tools at elevated risk of featuring in crime, then you can respectively concentrate your preventive policies and practices, direct your costly operational resources, design and incorporate elevated security performance in particular products, and choose the make and model of product you buy according to security ratings, as happens, say, with the UK car theft index (Laycock 2004).

Definitions

The term "risk" is used loosely in SCP, normally covering probability alone, and implicitly the risk of *harm to offender*. But risk can be decomposed into possibility (the nature of the undesired events), probability, and harm (Ekblom 2012b), and prevention can address each. Eliminating the *possibility* includes, say, replacement of tempered beer glasses with toughened or plastic ones so they cannot be misused as weapons (see, e.g., www.designcouncil.org.uk/our-work/challenges/Security/Design-out-crime/Alcohol-related-crime/). Reducing the *probability* includes providing clips at tables in bars to combat theft of customers' bags (Ekblom et al. 2012, and see www.designcouncil.org.uk/our-work/challenges/Security/Design-out-crime/Case-studies1/Stop-Thief-Chair-and-Grippa-Clips/). Reducing the *harm* includes designing an easy-to-use backup system for numbers stored on mobile phones (the numbers on stolen phones would otherwise be lost). Crime prevention approaches have until recently underemphasized harm reduction, focusing instead on "cutting the numbers of incidents." Harm is both something for designers to avoid, reduce, or mitigate and a consideration in setting priorities within the design process. Harm can be further divided: With a stolen purse harm could befall the user (if assaulted during a snatch), the product (the bag handle ripped), contents taken or damaged (e.g., phone), and a whole new range of victims through "propagated" offenses (e.g., identity theft from bank cards or burglary using stolen keys).

"Risk factors" is a concept derived from health (e.g., risk factors for heart disease). It is one of the few concepts used right across the crime prevention field. Coverage ranges from risk (and protective) factors for offending (such as poor parenting) through to criminogenic properties and features of places (e.g., the Burgess scale (Armitage 2006)) covering the immediate environment of houses (such as whether they are sited on a corner plot) and through to products (e.g., their value-to-weight ratio – Cohen and Felson 1979).

Risk Factor Approaches in Situational Crime Prevention

Classes of items at elevated risk of crime have been dubbed "hot products" (Clarke 1999), covered below. Products may be hot by virtue of their intrinsic material value (such as jewelry or bronze statuary), their manufactured-in value (such as a mobile phone), or some combination. In either case, this "reward" value (using ▶ rational choice perspective terms – Cornish and Clarke 1986) is often accompanied by some wider elements of opportunity, enabling the product to be taken with relatively little effort or risk to the offender. Of course, risk and effort may partly reside in the nature of the environment in which the products are typically found, such as whether guardians of targets (Cohen and Felson 1979) or other kinds of crime preventer (Ekblom 2011) are present, capable, and motivated. But much of that opportunity may reside in the rewarding and/or vulnerable design of the product itself; and even if the design is not obviously "culpable" (e.g., an easy-to-steal car or a provocative poster), design solutions may be the most reliable and/or cost-effective remedy.

The risk factor approach to products and crime was pioneered, like much else, by Cohen and Felson (1979) and Clarke (1999). Cohen and Felson generated the first of many acronyms, in the shape of VIVA. Value, inertia (i.e., weight), visibility, and access are a set of risk factors empirically backed by, among other things, correlating the decreasing weight of particular products (like TV sets) in the Sears catalogue with increasing crime risk. Clarke extended the analysis in a report using target-of-crime data from, for example, the British Crime Survey, to generate the more widely used CRAVED acronym:

- **C**oncealable
- **R**emovable
- **A**vailable
- **V**aluable
- **E**njoyable
- **D**isposable

CRAVED focuses more clearly on how the properties of particular classes of product connect to criminal opportunity from the offender's perspective, both as rewarding ends in themselves (valuable, enjoyable) and as means to achieving those ends via reduced risk and effort (concealable, removable, available, disposable). Examples of hot products include mobile/cell phones and cash.

Other risk factor acronyms have since emerged to demonstrate the versatility of this approach including AT CUT PRICES (Gill and Clarke 2012). This characterizes fast-moving consumer goods like batteries:

- **A**ffordable
- **T**ransportable
- **C**oncealable
- **U**ntraceable
- **T**radeable
- **P**rofitable
- **R**eputable
- **I**mperishable
- **C**onsumable
- **E**valuable
- **S**hiftable

A complementary "protective factors" approach was developed by Whitehead et al. (2008) who summarized the crime-resisting properties to design into mobile/cell phones. IN SAFE HANDS describes phones with these characteristics:

- **I**dentifiable – by owner, for example, through marking.
- **N**eutral – anti-theft design features should not adversely affect user's experience or elevate risk of other crimes.
- **S**een – to be protected – deterrence.
- **A**ttached – mechanical/electronic links to its owner.
- **F**indable – lost/stolen product can be tracked and found.
- **E**xecutable – can be deactivated if lost/stolen.
- **H**idden – for example, about the person, and used covertly.
- **A**utomatic – security built-in/automated.
- **N**ecessary – to be the owner and to be able to use a product, for example, via mechanical keys, codes, and biometrics.
- **D**etectable – make it obvious that product is being/has been stolen, for example, via alarm.
- **S**ecure – protection itself should not be easily removable or hackable.

Most of these factors describe the causal mechanisms whereby risk is reduced. Neutral and automatic relate to avoiding interference with other design requirements for the user, and secure describes self-protection of the security function. The latter connects with Ekblom's (2012b) distinction between a security feature on a product being "in function" (i.e., delivering the protection as intended) and "as object" (being a target of an attack intended to disable the security, steal the product for its valuable material, or vandalize it).

The risk factors approach, or at least how it is realized in practice, can be criticized. For example, concealable has a very different role depending on *when* in the theft *script* (Cornish 1994) the concealment occurs (Ekblom and Sidebottom 2008): when the thief is seeking a target (concealed in owner's pocket) versus when the thief is making off with a target (concealed in the thief's pocket). This has significant practical implications, introducing a design conflict between making the product concealable for users but not for thieves. Resolutions may involve active discrimination, for example, with computers which "report themselves" to owners or recovery companies such as Immobilise when separated from their registered place of use or cash-in-transit boxes which churn out smoke when stolen.

Risk factor lists are also somewhat ad hoc. To make the approach more systematic and to augment the capacity for exploring empirical risk patterns and generating new lists, an abstraction can help. Ekblom (e.g., 2008) devised the Misdeeds and Security framework to characterize very generic crime risks (and corresponding prevention opportunities) originally for assessing the criminogenic/criminocclusive impact of innovations in science and technology. The misdeeds are broad ways whereby products can feature in crime: a camera phone, say, can be:

- **M**isappropriated or stolen
- **M**istreated or deliberately damaged

- **M**ishandled, for example, smuggled
- **M**isbegotten or counterfeited
- **M**isused as a tool for crime, for example, in anonymous drug deals
- **M**isbehaved with, for example, cyber-bullying

Even broader is to treat the product as a *target* of crime (the first 4) or *contributor* to crime (the last 2). The contributor concept connects with the "crime facilitators" – tools or weapons – of ► SCP and "resources for offending" in the Conjunction of Criminal Opportunity framework (Ekblom 2011). A product could, of course, alternatively act as a *resource for crime preventers* as discussed below.

Underlying Causes of Risk

Risk factors are *correlates* of heightened possibility, probability, and harm attending particular products. What causal mechanisms underlie them? This is important for connecting the nature and design of products with underlying *theory*. Theory confers the capacity to generate variety in candidate modifications and innovations which are a priori plausible rather than "shots in the dark."

Sometimes elevated risk comes predominantly from *exposure* factors – some products tend to be left unattended (e.g., cars) or worn or used in risky environments (bling jewelry in clubs, mobile/cell phones on late-night streets). More localized exposure factors include being close to markets for stolen goods, being in locations where drug addicts steal to fuel their habit, or being the subject of aggressive and incautious sales techniques that leave valuable items accessible on the retailer's shelves.

Otherwise, predominant causes center on products themselves. *Value* is obviously important here – even if a satnav is concealed in a compartment in a car, offenders may seek it out. Value of course can change, as with steep fluctuations in commodity prices (Sidebottom et al. 2011). Other motivating causes engendered by products include acting as *precipitators* in the crime situation. These include *prompts* (drawing attention to a stealable object, such as a flashy new bike) and *provocations* (e.g., a coin-operated drink dispenser that swallows money and doesn't deliver).

In yet more cases, some sort of inherent *vulnerability* attends the product. The term vulnerability has been used variously, but Ekblom and Sidebottom (2008), attempting to define a consistent suite of product security concepts, suggest it be confined to covering all criminogenic properties (those enhancing probability of crime) of products *except* the motivation they engender. These normally relate to being seen and taken by the offender. In the case of harm, the product can be inherently *susceptible* to the actions of the offender – easily damaged, tampered with, etc.

These causes often reinforce one another. For example, properties that make phones inherently attractive to offenders, such as small size and portability, also make for ease of theft. Perhaps the most generic cause of elevated risk of mass-produced products is what might be called their *promiscuity* – they can be bought by anybody, sold on by anybody, and used by anybody, and virtually identical copies may be found throughout the community.

Vulnerability and susceptibility are not absolutes but depend on the resources the offender can bring to bear in taking, damaging, or manipulating the target product. These include other, misused products (e.g., portable cutting tools) and strength and dexterity (e.g., breaking anchorages and picking locks). Reflecting this understanding, insurance specifications for secure products like vehicles nowadays are stated as *performance criteria* (e.g., "resist attack by currently available tools for a minimum of 5 min") rather than technical construction (e.g., "lock must be made of manganese steel").

The Risk Life Cycle of Products

Felson (1997) observed that, besides a life cycle of legitimate use, products have a criminal one too:

1. Product does not exist.
2. Product exists, but few consumers know how to use it.
3. Product spreads, gains interest, worth stealing.
4. Product everywhere, no longer worth stealing.

This should produce an "inverted U curve" over time, of accelerating then decelerating risk. However, the "criminal nirvana" of saturation rarely occurs in reality. Both fashion and marketing/manufacturing tricks to get people to buy the latest model (not to mention true obsolescence and product unreliability) continue to drive both legitimate consumption and theft long after everyone possesses their first mobile phone.

Fundamentals of the Response to Elevated Risk

There are several broad strategies for responding to elevated risk of crime associated with particular products. The products can be:

- Safe – kept in a guarded or locked environment, like bullion.
- Secured after sale – protected by a dedicated security device (e.g., a "crooklock" linking steering wheel and brake pedal of unattended cars).
- Secured in production – incorporating specialized security components like anti-counterfeiting stickers.
- Security adapted – where design features explicitly reduce vulnerability (like anti-picklocks) or lower value (like the folding Puma bike whose diagonal down-tube is replaced by a tensioned steel cable, which unlocks at one end to wrap around the bike stand and which, if cut, renders it unrideable and unsaleable – see www.designagainstcrime.com/projects/puma-bike/).
- Inherently secure – by virtue of weight or bulk, for example, home cinema TV sets are currently awkward to carry off (but future technologies may see roll-up versions).

Applying the Misdeeds and Security framework cited above (fuller treatment is in Ekblom 2008), products can be:

- Secured against misappropriation – for example, vehicles with built-in immobilizers
- Safeguarded against mistreatment – for example, street signs that avoid couching regulations in provocative, confrontational terms
- Scam-proofed against mishandling and misbegetting – for example, fold-over airline baggage labels concealing holidaymakers' addresses from burglars' touts; or anti-copying functions within DVDs
- Shielded against misuse and "sivilized" against misbehavior – for example, "once-only" syringes and waste bins that reveal their contents – including hidden bombs (Lulham et al. 2012, and see www.designcouncil.org.uk/our-work/challenges/Security/Design-out-crime/Case-studies1/An-anti-terrorist-rubbish-bin/) or metro station seating shaped to discourage rough sleeping

Design and the Design Process

Links with Situational Crime Prevention

Product design connects closely with, and applies, many of the 25 techniques of ▶ situational crime prevention (see www.popcenter.org/25techniques/). It can even extend this list, for example, with the concept of "target softening" – with, say, the lock whose bolt can swivel in its housing, causing hacksaw blades to slip.

More theoretically, product design engages with the risk, effort, and reward (as encountered or perceived by the offender) of the Rational Choice perspective (Cornish and Clarke 1986). The Karrysafe handbag (www.designagainstcrime.com/projects/karrysafe/) has a Velcro fastening which increases the risk of the owner hearing or feeling the thief's action. An anchor cable for securing laptop to table leg increases the effort and resources (a cutter is needed to release it). Ink tags clipped to expensive clothing spoil the reward when shoplifters try to remove them.

SCP also seeks to manipulate *crime precipitators*, which add emotional/motivational/perceptual influences to the opportunity to act out the emotion or realize the criminal goal thus awakened. A frustrating door entry system can provoke damage from "machine rage"; a stylish new mobile phone can prompt thoughts of theft.

Even if the design is not obviously "culpable" for a product's elevated crime risk, design solutions may be the most reliable and/or cost-effective remedy, for several reasons:

- Design potentially removes the burden of effort from guardian of target products: with central locking, for example, car owners no longer must remember to lock all the individual doors of their vehicle.

- Mass production potentially enables incorporation of security into a huge proportion of product classes and individual items, covering many individual crime situations.
- Such mass coverage can supply the "herd immunity" needed for impact – when the proportion of a particular kind of items protected is high enough for thieves to give up on the whole category.

Designers can confer criminocclusive or harm-reducing properties on their products by intervening in the materials (e.g., resistant to damage), structural features (e.g., concealable), or functionality of a product (incorporating security or securing functions). Even the products' packaging can be recruited for security (Segato 2012). Criticisms of the design approach center on "paranoid products" (Gamman and Thorpe 2007) where the security focus is excessive, fear arousing, ugly, or inconvenient. But these apply to *poor* designs and clunky "engineering" solutions. Done properly, design can resolve a range of contradictions or "troublesome tradeoffs" (Ekblom 2008, 2012a, b), for example, between security and aesthetics, cost, convenience, and social inclusion (e.g., locks usable by elderly/disabled people).

The important thing is that designers capture all these requirements in the design *process*, which should begin with a penetrating analysis of the diverse stakeholders' interests. This should provide the basis for an effort to maximize the meeting of the often-contradictory requirements – not by seeking compromises but by applying ingenuity and creativity in a process of iterative development and testing (see Ekblom 2012c). The end result should be simultaneously user friendly while abuser unfriendly (Ekblom 2005). Further accounts of the design against crime process are in Thorpe et al. (2009) and www.designagainstcrime.com/methodology-resources/design-methodology/#users-abusers; and on the UK Design Council website. See the "double-diamond" model of design at www.designcouncil.org.uk/designprocess and the guide to designing out crime at www.designcouncil.org.uk/our-work/challenges/Security/Design-out-crime/Design-out-crime-guide/.

To support a more efficient, and potentially more effective, design effort, several preconditions must be established:

- First, designers require a "think thief" (or "think terrorism") mindset (Ekblom 2005), which may not come naturally if they assume all those who will encounter their product are legitimate users and honest, well-behaved citizens.
- Second, they need the simultaneous guidance and constraint offered by theoretical principles, so they reliably come up with plausible ideas as a starting point (though we should preserve the "wonky thinking" and importation of "foreign" ideas that generate true novelty).
- Third, they need "design freedom" to innovate, which is best served by performance-based requirements (as described) and tested theoretical principles rather than detailed technical specifications or specific exemplars of products that have successfully resisted crime – though particular security features can be innovatively recombined and tweaked to adapt to new products or new contexts of use.
- Finally, developing a clear design rationale (Ekblom 2012c) is important both to sharpen thinking and communicate with other designers and clients. One such rationale is the Security Function framework (Ekblom 2012a). This systematically describes products in terms of:
  1. *Purpose* (what are they for?)
  2. *Security niche* (how do they fit with the "ecology of security" – are they, for example, security products, securing products like "Stop Thief" chairs for cafes with notches to securely hang bags behind one's knees (see www.designcouncil.org.uk/our-work/challenges/Security/Design-out-crime/Case-studies1/Stop-Thief-Chair-and-Grippa-Clips/), or inherently secure products?)
  3. *Mechanism* (how do they work in cause-effect/theoretical terms?)
  4. *Technicality* (how are they constructed and how do they operate?)

Retrospective descriptions of products using this framework include the Grippa clip for securing customers' bags to bar tables (Ekblom et al. 2012); Meyer and Ekblom (2011) use it to provide a prospective specification for an explosion-resistant railway carriage.

Some of the above preconditions may arguably militate against individualistic deployment by talented designers of sheer intuitive genius. But their establishment enables society to build a broader design against crime capacity, one that enables more of the "field of emerging products" to be covered on a more routine basis. We thus raise the overall ground level of design fitness rather than achieving a few spectacular, but isolated, peaks.

Anticipating Risk Through Design

Every new product design is a bet on the future, whether concerning market success or undesired side effects like crime. Continual arrival on the market of new, naively insecure products generates what Pease (2001) calls *crime harvests*, followed by hasty retrospective efforts to cope with the crime and clumsily patch the damage by remedial design. The classic example has been with mobile/cell phones. While older "phone cloning" leaks are now plugged with the switch from analogue to digital systems, arguably the early vulnerabilities enabled the establishment of a crime market, with a persistent corpus of criminal expertise, criminal service providers, and criminal networks.

Advances in technology also produce a steady stream of new resources for crime, like cordless drills or pocketable 12 V batteries (which can be misused to energize car door locks). Previously secure items become vulnerable overnight. In fact, adaptive, entrepreneurial offenders and a changing technological backdrop set the scene for arms races between offenders and preventers, especially designers. The classic case is the evolution of the safe. The pace of the race is boosted by dissemination through the Internet – there are, for example, many lock-picking forums. Such arms races are described in Ekblom (2005). The basic strategy for handling them is (1) to develop ways to anticipate offender moves and countermoves and (2) to build designers' capacity (and manufacturers' will) to out-innovate the offenders and more generally to design variety and upgradeability into products.

The risk factor approach naturally primes anticipation. An ambitious attempt to develop a theft-proofing approach for personal electronic products was the EU-funded Project MARC (Armitage 2012). The basic plan was to devise a system for (1) determining the anticipated risk of theft attending some new product exemplar and then (2) incorporating, at the design stage, a commensurate level of security (obviously, products judged to be potentially at elevated risk of theft from their own properties and/or context of use should be given correspondingly higher security specifications). An attempt was made to try this out with a sample of existing mobile personal electronic products, rated by diverse experts, but various difficulties arose. For one thing, it was judged that a security checklist approach would impose an artificial ceiling on the exercise of ingenuity and skill in crime preventive design. This could lead designers to design-down to the level of security required by the checklist and militate against innovation. The security checklist also understates the degree to which security is specific to product type, which rather removes the justification for standardization. Nevertheless, this first serious attempt at crime proofing of products is unlikely to be the end of the story. The original lead researcher, in a recent reprise (Armitage 2012), makes a strong case, with practical suggestions, for taking a modified approach forward.

Horizon-scanning and foresight approaches (Department of Trade and Industry 2000) acknowledge the need to entertain diverse possibilities when making products robust to the future: specific predictions will likely be wrong. One such case was the TV set-top box, designed to allow analogue TV sets to receive digital signals. This seemed a likely hot product, until the service providers changed their marketing strategies from selling the boxes at cost to heavily subsidizing the price and recouping revenue from the additional service. Given such uncertainties, designers could cope by incorporating into their products some flexibility and upgradeability of the security function.

Mobilizing Design of Products Against Crime

Designs are often intended to work with human crime preventers. The Grippa clip (Ekblom et al. 2012), for example, requires bar customers to fasten their bag to it. Others work against unintentional crime promoters. The M-shaped "caMden" bike stand (designed by Adam Thorpe (Thorpe et al. 2009) on the basis of extensive research into locking behavior and secure parking configurations) nudges cyclists to lock their bike securely (both wheels and the frame) rather than relying on a single lock in the middle of the crossbar, which leaves the wheels removable and the frame liable to misuse as a tool for its own theft – serving as a lever to snap the lock.

Mobilization of preventers is challenging. Ekblom (2012d) describes how various mobilization failures among bar customers, bar staff, management, and senior company executives of one company left the Grippa clips unused (though not so in other companies' venues).

Governments could play various roles in modifying criminogenic products (Clarke and Newman 2005). Policy justifications for governments to mobilize companies center on "polluter pays" principles where a company that generates crime opportunities that fall as "externalities" on other victims or taxpayers is required to modify their products or to compensate in some other way (e.g., Newman 2012; Roman and Farrell 2002). Mobilization can be motivated through incentives including tax breaks, regulations, and "naming and shaming" of criminogenic designs. A useful review was undertaken by the UK Home Office (2006). An example of incentivization of vehicle manufacturers to improve security through awakening market demand is the UK Home Office's Car Theft Index (latest version covering 2005–2006 at http://data.gov.uk/dataset/car-theft-index-2004-2006). This encourages buyers to take theft rates of particular makes and models into account when choosing which to purchase. (Learmount (2005) makes the wider case for demand-led influence in a review of the field.) Similar pressures from insurance companies have also been effective (Hardie and Hobbs 2005).

Does Design Against Crime Work?

Assessment and feedback from workshop tests, field trials, user and service engineer experience, and ultimately sales, profitability, and market leadership are inherent to the iterative process of directed improvement that is product design (Thorpe et al. 2009). In impact evaluation and cost-effectiveness terms normally applied to crime prevention, however, there is unfortunately little hard evidence that relates to product design as opposed to "target hardening" and other situational approaches in general. Such evidence as exists is often characterized by weak research designs; formally evaluated products were summarized in Clarke and Newman (2005, Table 4), and few such studies have emerged since.

One reason is that prototypes are expensive to produce and test in sufficient quantities to support an impact evaluation of sufficient statistical power (Bowers et al. 2009). Another is the timescale for developing a product then evaluating it within a typical research funder's time frame.

Circumstantial, correlational evidence points to the contribution of vehicle security technology towards the substantial and sustained reduction of theft of cars in the UK in recent years, following implementation of a European Directive on compulsory factory fitting of immobilizers from 1998 (Webb 2005). British Crime Survey figures (Home Office 2007, Table 2.01) show theft of vehicles reduced by 65 % from 1995 to 2006–2007 following the design of improved security into the vehicle. None of the case studies commissioned by the Design Council for the Home Office have been formally evaluated.

Other evidence is more anecdotal but almost entirely self-evident (Clarke and Newman 2005). An example is the fabric curtain between certain London Underground train carriages, retrospectively fitted to stop boys riding the couplings. A glance reveals nowhere left to stand. But self-evidence cannot be taken for granted and gives no information on comparative cost-effectiveness.

A recent study (Sidebottom et al. 2009) of attempts to reduce bike theft by installing advisory stickers on the bike stands has yielded reliable intermediate outcome evidence, important where behavioral change of people acting as crime preventers or promoters (Ekblom 2011) is sought. The stickers were designed after systematic observation of bike-locking behavior and analysis of perpetrator techniques. The simple advice – lock both wheels and frame to the stand – yielded significant and substantial reduction (from 62 % to 48 % of observations) in the proportion of bikes locked insecurely (available funding did not, however, cover evaluation of impact on theft).

Conclusions and Future Research

The practice of systematically designing products against crime and the research into that practice have barely left their infancy. Concepts and frameworks are beginning to emerge (and will need further integration with one another and with other crime-oriented disciplines like SCP or schools of practice like CPTED). They have only superficially tapped the wealth of knowledge and experience that is the field of design.

The emerging domain faces a significant challenge in motivating designers, producers, and consumers to press for, and to use, secure designs. And any knowledge that does accumulate must be considered a "wasting asset," vulnerable to becoming out of date with social and technical change and adaptive offenders. Finding ways to incentivize designers and their clients, and developing and building innovative capacity among designers, is the only way to keep ahead in the long run.

We still lack a sufficient range of rich and rigorous case studies to build on. In particular, the effort to find hard evidence of the cost-effectiveness of product design against crime must continue. Only then will design against crime fare better in obtaining sustained funding and attention from the government. The evidence may also help convince consumers to favor products so designed and manufacturers to routinely include security in their requirements capture.

Related Entries

- ▶ Crime Prevention Through Environmental Design
- ▶ Hot Spots and Place-Based Policing
- ▶ Innovation and Crime Prevention
- ▶ Motor Vehicle Theft
- ▶ Naming and Shaming of Corporate Offenders
- ▶ Problem-Oriented Policing
- ▶ Rational Choice Theory
- ▶ Repeat Victimization
- ▶ Routine Activities Approach
- ▶ Situational Crime Prevention
- ▶ Theories for Situational and Environmental Crime Prevention

Recommended Reading and References

Armitage R (2006) Predicting and preventing: developing a risk assessment mechanism for residential housing. Crime Prev Community Saf: Int J 8: 137–149

Armitage R (2012) Making a brave transition from research to reality. In: Ekblom P (ed) Design against crime: crime proofing everyday objects, vol 27, Crime prevention studies. Lynne Rienner, Boulder

Bowers K, Sidebottom A, Ekblom P (2009) CRITIC: a prospective planning tool for crime prevention evaluation designs. Crime Prev Community Saf 11:48–70

Clarke R (1999) Hot products: understanding, anticipating and reducing demand for stolen goods, vol 112, Police Research Series. Home Office, London

Clarke R, Newman G (2005) Modifying criminogenic products – what role for government? In: Clarke R, Newman G (eds) Designing out crime from products and systems, vol 18, Crime prevention studies. Willan, Cullompton

Cohen L, Felson M (1979) Social change and crime rate changes: a routine activities approach. Am Sociol Rev 44:588–608

Cornish D (1994) The Procedural analysis of offending and its relevance for situational prevention. In: Clarke R (ed) Crime Prevention Studies, vol 3. Criminal Justice Press, Monsey

Cornish D, Clarke R (1986) The reasoning criminal. Springer, New York

Cropley D, Kaufman J, Cropley A, Runco M (2010) Creativity: the dark side. Cambridge University Press, Cambridge, UK

Department of Trade and Industry (2000) Turning the corner. Report of Foresight Programme's Crime

Prevention Panel. Department of Trade and Industry, London

Ekblom P (2005) Designing products against crime. In: Tilley N (ed) Handbook of crime prevention and community safety. Willan, Cullompton

Ekblom P (2008) Designing products against crime. In: Wortley R, Mazerolle L (eds) Environmental criminology and crime analysis. Willan, Cullompton

Ekblom P (2011) The 5Is framework for crime prevention, security and community safety. Palgrave Macmillan, Basingstoke

Ekblom P (2012a) The private sector and designing products against crime'. In: Welsh B, Farrington D (eds) The oxford handbook on crime prevention. OUP, Oxford, pp 384–403

Ekblom P (2012b) The security function framework. In: Ekblom P (ed) Design against crime: crime proofing everyday objects, vol 27, Crime Prevention Studies. Lynne Rienner, Boulder

Ekblom P (2012c) Happy returns: Ideas brought back from situational crime prevention's exploration of design against crime. In: Farrell G, Tilley N (eds) The reasoning criminologist: essays in Honour of Ronald V. Clarke, Crime Science series. Willan, Cullompton, pp 163–198

Ekblom P (2012d) Citizen participation in crime prevention: capturing practice knowledge through the 5Is framework. In: Coester M, Marks E (eds) International perspectives of crime prevention 4. Contributions from the 4th and the 5th Annual International Forum 2010 and 2011 within the German Congress on Crime Prevention. Forum Verlag Godesberg GmbH, Mönchengladbach

Ekblom P, Sidebottom A (2008) What do you mean, 'is it secure?' Redesigning language to be fit for the task of assessing the security of domestic and personal electronic goods. Eur J Crim Policy Res 14(61–87):20

Ekblom P, Bowers K, Gamman L, Sidebottom A, Thomas C, Thorpe A, Willcocks M (2012) Reducing handbag theft in bars. In: Ekblom P (ed) Design against crime: crime proofing everyday objects, vol 27, Crime Prevention Studies. Lynne Rienner, Boulder

Felson M (1997) Technology, business and crime. In: Felson M, Clarke R (eds) Business and crime prevention. Criminal Justice Press, Monsey

Gamman L, Thorpe A (2007) Profit from paranoia: design against 'paranoid' products. European Academy of Design conference 07: dancing with disorder: design, discourse, disaster. Izmir, Turkey. www.bikeoff.org/2007/04/30/profit-from-paranoia-design-against-paranoid-products/

Gill M, Clarke R (2012) Slowing thefts of fast-moving goods. In: Ekblom P (ed) Design against crime: crime proofing everyday products, vol 27, Crime Prevention Studies. Lynne Rienner, Boulder

Hardie J, Hobbs B (2005) Partners against crime: the role of the corporate sector in tackling crime. In: Clarke R, Newman G (eds) Designing out crime from products and systems, vol 18, Crime Prevention Studies. Willan, Cullompton

Home Office (2006) Changing behaviour to prevent crime: an incentives-based approach. Online report 05/06. Home Office, London www.homeoffice.gov.uk/rds/pdfs06/rdsolr0506.pdf

Home Office (2007) Crime in England and wales 2006/2007. Statistical Bulletin 11/07. Home Office, London

Laycock G (2004) The UK car theft index: An example of government leverage. In: Maxfield M, Clarke R (eds) Understanding and Preventing Car Theft, Crime Prevention Studies 17. Monsey, NY, Criminal Justice Press

Learmount S (2005) Design against crime. In: Clarke R, Newman G (eds) Designing out crime from products and systems, vol 18, Crime prevention studies. Willan, Cullompton

Lulham R, Duarte O, Dorst K, Kaldor L (2012) Designing a counterterrorism trash bin. In: Ekblom P (ed) Design against crime: crime proofing everyday objects, vol 27, Crime Prevention Studies. Lynne Rienner, Boulder

Meyer S, Ekblom P (2011) Specifying the explosion-resistant railway carriage: a desktop test of the security function framework. J Transp Sec 5:69–85

Newman G (2012) A market approach to crime prevention. In: Ekblom P (ed) Design against crime: crime proofing everyday objects, vol 27, Crime prevention studies. Lynne Rienner, Boulder

Pease K (2001) Cracking crime through design. Design Council, London

Roman J, Farrell G (2002) Cost-benefit analysis for crime prevention: opportunity costs, routine savings and crime externalities, vol 14, Crime prevention studies. Willan, Cullompton

Segato L (2012) Packaging against counterfeiting. In: Ekblom P (ed) Design against crime: crime proofing everyday objects, vol 27, Crime prevention studies. Lynne Rienner, Boulder

Shepherd J (1994) Preventing injuries from bar glasses. Br Med J 308:932–933

Sidebottom A, Johnson S, Thorpe A (2009) Using targeted publicity to reduce opportunities for bicycle theft: a demonstration and replication. Eur J Criminol 6:267–286

Sidebottom A, Belur J, Bowers K, Tompson L, Johnson SD (2011) Theft in price-volatile markets: On the relationship between copper price and copper theft. Journal of Research in Crime and Delinquency 48:396–418

Thorpe A, Gamman L, Ekblom P, Johnson S, Sidebottom A (2009) Bike Off 2 – catalysing anti-theft bike, bike parking and information design for the 21st century: an open innovation research approach. In: Inns T (ed) Designing for the 21st century, vol 2, Interdisciplinary methods and findings. Farnham, Gower

Webb B (2005) Preventing vehicle crime. In: Tilley N (ed) Handbook of crime prevention and community safety. Cullompton, Willan

Whitehead S, Mailley J, Storer I, McCardle J, Torrens G, Farrell G (2008) IN SAFE HANDS: a review of mobile phone anti-theft designs. Eur J Crim Policy Res 14:39–60

Desistance

▶ Desistance and Supervision

Desistance and Supervision

Fergus McNeill[1], Stephen Farrall[2],
Claire Lightowler[3] and Shadd Maruna[4]
[1]Scottish Centre for Crime and Justice Research, University of Glasgow, Glasgow, UK
[2]University of Sheffield, Sheffield, UK
[3]Institute for Research and Innovation in Social Services, Glasgow, UK
[4]Queen's University Belfast, Belfast, UK

Synonyms

Corrections; Desistance; Parole; Probation; Rehabilitation; Supervision

Overview

This entry explores the origins and development of arguments about the use of theories of and evidence about desistance from crime as a basis for developing approaches to offender rehabilitation and supervision. We begin by outlining arguments about the relative merits of offense-focused and desistance-focused approaches before going on to review the case for a "desistance paradigm" for correctional practice. In the final section, we outline the process and (some of the) outcomes of a recent project which aimed to further these debates by actively engaging ex-offenders, their families and supporters, frontline practitioners, managers, and policymakers in coproducing a set of recommendations for criminal justice reform, informed by their experiences of desistance and supporting desistance.

Introduction

Desistance theories and research seek to understand and explain how and why people stop offending – and stay stopped. The notion that such studies might provide a framework for offender supervision, and even for criminal justice interventions more widely conceived, has a long history – dating back at least to the work of the Gluecks (1937 [1966]). If, as they argued, desistance is about maturing out of criminal conduct, what could be done through criminal justice interventions to "force the plant" or to accelerate the maturational process? The question is, of course, a very good one, not least in the context of contemporary preoccupations with the economic, human, and social costs of reoffending by ex-prisoners (see, e.g., Ministry of Justice 2010) and with the broader challenges of ex-prisoner reentry (see Petersilia, this volume).

And yet, between the 1930s and the end of the twentieth century, hardly any use of desistance research to inform sentencing and correctional policy and practice is discernible. Instead, the story of this era is the familiar one of the rise and fall and rise again of rehabilitative interventions, a historical cycle linked to but not fully explained by debates about their effectiveness or ineffectiveness. Though these two topics – the process of desistance from crime and the effectiveness of rehabilitative interventions – are obviously linked in several ways, the connections between them did not begin to be properly explored until the turn of the century.

Today, debates and discussions about desistance and how to support it through criminal justice interventions seem to be bubbling up all around the world of corrections, not just in jurisdictions with deep cultural and historical connections, like the UK and the USA, but also in places as diverse as Norway and Singapore.

This entry does not aim to explain this upsurge of interest in desistance nor does it engage with the important and interesting question of when desistance research is relevant (and irrelevant) to criminal justice (see McNeill and Weaver 2010). Suffice it to say that unless criminal justice is concerned on some level with rehabilitation and reducing reoffending, desistance theory and research is unlikely to have much purchase. But to the extent that sentencing and correctional systems, and more specifically to the structure and practice of supervision, are concerned with these outcomes, understanding how and why people stop offending (with or without help or hindrance from the justice system) has obvious appeal. Throughout the entry, we use the US convention of referring to "sentencing and corrections," meaning the end of the justice process where sanctions are decided and then delivered. Our particular focus is on supervisory sanctions, that is, those sanctions or elements of sanctions, like probation and parole, which involve the supervision of the sentenced person in the community.

Rather than seeking to review theories of and evidence about desistance itself, this entry has the more modest aim of charting the emergence and development of the arguments advanced over the last 12 about the implications of this body of work for offender supervision.

Offense-Focused or Desistance-Focused Supervision

The emergence and development (or perhaps the revival) of debates about how desistance research could and should inform the development of supervision owes a great deal to the work of Stephen Farrall (2002) and of Shadd Maruna (2001). Both books were based in research projects which drew on and developed earlier work by Ros Burnett; her *Dynamics of Recidivism* study (Burnett 1992) was critical in generating new interest in desistance research in the UK. Another important early foray into the study of "assisted desistance" (as opposed to spontaneous or unaided desistance) was undertaken by Sue Rex (1999), who argued explicitly that:

> The knowledge we are beginning to acquire about the type of probation services which are more likely to succeed could surely be enhanced by an understanding of the personal and social changes and developments associated with desistance from crime. (Rex 1999: 366)

Thus, even while Farrall's and Maruna's research projects were ongoing, publications (like Rex's) had begun to emerge which engaged directly with the question of how desistance theory and research might inform supervision. An interesting early example was an edition of *Offender Programs Report* (volume 4, issue 1) which, among several interesting short articles, included a paper from Maruna (2000) in which he argued for a marrying of the desistance and "What Works?" literatures, taking from the former its analyses of the "micro-mechanisms of change" at the individual level and from the latter an appreciation of the general principles of effective rehabilitative intervention.

That marriage however looked ill fated when Farrall's (2002) book was published. It challenged the somewhat narrow and managerialized interpretations of "What Works?" research which, at that time, dominated correctional policy and practice in the UK. As well as developing a searching methodological critique of the "What Works?" research, Farrall's study (based on a qualitative longitudinal study of 199 probationers and their supervising officers) presented findings which suggested that motivation and social context were more clearly associated with desistance than probation supervision and that the focus of supervision (on risk factors and "criminogenic needs") neglected the crucial roles of relationships and social capital in the desistance process. Farrall's (2002) related proposition was that supervision should focus not solely on "offence-related factors" (or "criminogenic needs") but also on "desistance-related needs." The nature of the difference between the two approaches is perhaps best captured by one of the probationers in his study, in response to a question about what would prevent him from reoffending:

> Something to do with self progression. Something to show people what they are capable of doing. I thought that was what [my Officer] should be about. It's finding people's abilities and nourishing and making them work for those things. Not very consistent with going back on what they have done wrong and trying to work out why – 'cause it's all going around on what's *happened* – what you've already been punished for – why not go forward into something... For instance, you might be good at writing – push that forward, progress that, rather than saying 'well look, why did you kick that bloke's head in? Do you think we should go back into anger management courses?' when all you want to do is be a writer. Does that make any sense to you at all? *Yeah, yeah. To sum it up, you're saying you should look forwards not back.*

Desistance and Supervision, Table 1 Ideal-type contrasts: offense-focused and desistance-focused practice

	Offense-focused practice	Desistance-focused practice
Orientation	Retrospective	Prospective
Problem locus	Individual attitudes and behaviors	Individual problems and behaviors in social context
Practice focus	Individual attitudes and behaviors	Personal strengths and social resources for overcoming obstacles to change
Medium for effective practice	Rehabilitative programs (to which offenders are assigned on the basis of risk/needs assessment instruments)	Individual processes and relationships
Worker's Roles	Risk/needs assessor, program provider, case manager	Risk/needs/strengths assessor, advocate, facilitator, case manager
Intended outputs	Enhanced motivation	Enhanced motivation
	Pro-social attitudinal change	Changes in narrative/self-concept
	Capacity/skills development	Development of inclusion opportunities
Intended outcomes	Reduced re-offending	Reduced re-offending
		Enhanced social inclusion

> Yeah. I know that you have to look back to a certain extent to make sure that you don't end up like that [again]. The whole order seems to be about going back and back and back. There doesn't seem to be much 'forward'. (Farrall 2002: 225)

McNeill (2003), McNeill and Batchelor 2004), drawing not just on the work of Farrall and Maruna but on a wider range of desistance studies (*as well as* on "What Works?" research), sought to further elaborate what "desistance-focused probation practice" might look like. He argued that such practice would require thoroughly individualized assessment, focused on the interrelationships between desistance factors (linked to age and maturation, social bonds, and shifts in narrative identity), which build towards clear plans to support change. It would also require engaging in active and participative relationships characterized by optimism, trust, and loyalty, as well as interventions targeted at those aspects of each individual's motivation, attitudes, thinking, and values which might help or hinder progress towards desistance. Crucially, in McNeill's (2003) assessment, it would require work not just to develop personal capabilities but also to access and support opportunities for change, for example, around accommodation and employment. Finally, such practice would require approaches to evaluation which were themselves engaging, since such approaches would be vital in learning more from those involved about what persuaded them to desist and about the support that they needed to see their decisions through.

Beyond these practical prescriptions, and inspired by Farrall (2002), McNeill and Batchelor (2004: 66) went on to further elaborate the shift in practice dispositions or perspectives that desistance research seemed to suggest:

Table 1 above contrasts two notional "ideal types" of practice. McNeill and Batchelor (2004) were clear that this was intended only as a heuristic device; arguably neither of these approaches could or should exist in a "pure" form. Rather, the challenge, they argued, was to combine elements of both approaches in a case-sensitive manner. Hence, an offense focus must, of course, be necessary and appropriate given that, within any justice context, it is offending which occasions and justifies state intervention. However, being only or overly offense-focused might in some senses tend to accentuate precisely those aspects of a person's history, behavior, and attitudes which intervention aims to diminish. It may also, they suggested, tend towards misidentifying the central problem as one of individual "malfunctioning":

> Being desistance-focussed, by contrast, implies a focus on the purpose and aspiration of the intervention rather than on the 'problem' that precipitates it. It also tends towards recognising the broader social contexts and conditions required to

> support change. Thus, where being offence-focussed encourages practice to be retrospective and individualised, being desistance-focussed allows practice to become prospective and contextualised. (McNeill and Batchelor 2004: 67)

Although Maruna's (2001) book engaged less directly with the implications of his study for supervision, his ideas (conceived and elaborated along with his colleague and coauthor Tom LeBel) about "strengths-based" approaches to reentry and corrections were already developing in similar directions, informed both by desistance research and by a wider range of influences (Maruna and LeBel 2003, 2009). In essence, Maruna and LeBel (2003) exposed the limitations and problems associated with both risk-based and support- (or need-) based narratives for reentry. The former, they argued, casts the offender ultimately as a threat to be managed, the latter as a deficient to be remedied by the application of professional expertise. By contrast, "[s]trengths-based or restorative approaches ask not what a person's deficits are, but rather what positive contribution the person can make" (Maruna and LeBel 2003: 97).

Drawing on his *Liverpool Desistance Study* (LDS), Maruna et al. (2004) engaged more directly with the implications of the LDS for supervision, but reached similar conclusions. Probation discourse, they suggested, should move away from risks and needs and towards strengths, seeking to support and encourage redemptive and generative processes, such as those involved in constructive service or voluntary activities. Such a discursive shift could signal the positive potential of probationers not just to them but equally importantly to their communities. Although they supported Farrall's (2002) call for a more explicitly prospective or future-oriented form of supervisory practice, they also recognized the need for people to make sense of their pasts and therefore suggested the need for rehabilitative practices to support a reconstruction of the person's personal narrative, one which recognized and repaired wrongdoing but which refused to define or delimit the person by their previous (mis)conduct.

A Desistance Paradigm

By the middle of the first decade of this century, debate about the implications of desistance theory and research had developed to the point where "A Desistance Paradigm for Offender Management" was proposed (McNeill 2006). The "desistance paradigm" was written in the context of a peculiarly British debate about how probation practice should be reframed in the light of both changing evidence and normative arguments. As such, it engaged with two preceding paradigm-defining papers, the first of which (at the height of the "Nothing Works" era) argued for a "Non-Treatment Paradigm for Probation Practice" (Bottoms and McWilliams 1979) and the second deploying emerging evidence about effective intervention approaches to propose a "Revised Paradigm" (Raynor and Vanstone 1994). McNeill (2006) used both desistance research and normative arguments to seek to displace not their earlier paradigms but what he perceived as the misappropriation and misinterpretation of evidence in a managerialized and reductionist "What Works" paradigm that dominated probation policy and practice at that time (Table 2). He summed up the four paradigms as follows:

McNeill (2006: 56–57) summed up his central argument as follows:

> Unlike the earlier paradigms, the desistance paradigm forefronts processes of change rather than modes of intervention. Practice under the desistance paradigm would certainly accommodate intervention to meet needs, reduce risks and (especially) to develop and exploit strengths, but whatever these forms might be they would be subordinated to a more broadly conceived role in working out, on an individual basis, how the desistance process might best be prompted and supported. This would require the worker to act as an advocate providing a conduit to social capital as well as a 'treatment' provider building human capital. Moreover, rather than being about the technical management of programmes and the disciplinary management of orders, as the current term [in England and Wales] 'offender manager' unhelpfully implies, the forms of engagement required by the paradigm would re-instate and place a high premium on collaboration and involvement in the process of co-designing interventions. Critically, such interventions would not

Desistance and Supervision, Table 2 Probation practice in four paradigms

The non-treatment paradigm	The revised paradigm	A what works paradigm	A desistance paradigm
Treatment becomes help	Help consistent with a commitment to the reduction of harm	Intervention required to reduce reoffending and protect the public	Help in navigating towards desistance to reduce harm and make good to offenders and victims
Diagnoses becomes shared assessment	Explicit dialogue and negotiation offering opportunities for consensual change	"Professional" assessment of risk and need governed by structured assessment instruments	Explicit dialogue and negotiation assessing risks, needs, strengths and resources and offering opportunities to make good
Client's dependent need as the basis for action becomes collaboratively defined task as the basis for action	Collaboratively defined task relevant to criminogenic need and potentially effective in meeting them	Compulsory engagement in structured programs and case management processes as required elements of legal orders imposed irrespective of consent	Collaboratively defined tasks which tackle risks, needs and obstacles to desistance by using and developing the offender's human and social capital

> be concerned solely with the prevention of further offending; they would be equally concerned with constructively addressing the harms caused by crime by encouraging offenders to make good through restorative processes and community service (in the broadest sense). But, as a morally and practically necessary corollary, they would be no less preoccupied with making good to offenders by enabling them to achieve inclusion and participation in society (and with it the progressive and positive reframing of their identities required to sustain desistance).

McNeill's argument had been developed partly as the result of undertaking a literature review which aimed to explore the key skills required of supervisors charged with reducing reoffending (McNeill et al. 2005). Importantly, both that review and the paradigm paper included and attempted to integrate findings from desistance studies, "What Works?" correctional research, and the wider literature on the characteristics of effective psychosocial interventions more generally. Despite different methodologies and disciplinary orientations, some similar findings were emerging in desistance and "What Works?" research, for example, about the importance of the worker/client relationship in supporting change and about the need for "brokerage" of access to wider services to address practical needs.

More recently, Maruna and LeBel (2010) have engaged directly with the promise of developing and employing a desistance paradigm for correctional practice. Like McNeill (2006), they begin with the recognition that evaluation evidence (about what works to produce particular outcomes) is not the only form of evidence that matters in developing evidence-based practice (EBP). To focus exclusively on evaluation evidence in guiding intervention choices is to miss the importance of evidence about the very processes that interventions exist to support. Echoing Lewis (1990), they suggest a shift in focus "from programmes to lives," eschewing a correctional or medical model of change. In a more recent paper, McNeill et al. (2012a) have argued that it is simply wrong to treat desistance as the outcome of an intervention; interventions can contribute to change, but they do not "produce" it in any simple sense; desistance can and does exist "before, beyond, and behind" interventions. By focusing on and better understanding the change process, Maruna and LeBel (2010) argue, we may be able to better adapt supervision to contribute to (but not to "produce") the process. It follows that a desistance paradigm must place the person changing and *their* change process (and not the program) center stage, a message that finds support among prisoners and probationers themselves, who are often resistant to being processed through more programmatic interventions (Harris 2005).

Although the views and voices of "offenders" may not carry much weight with political or public audiences of correctional work, given the well-documented problems of supervision violation and program dropout, correctional practitioners would do well to heed the views and voices of those they are seeking to influence. Desistance research, in many respects, is one way (and only one way) in which these views and voices can be heeded.

Maruna and LeBel (2010: 72) sum up their argument for a desistance paradigm as follows:

> the desistance paradigm argues that the search for 'what works' should not begin with existing expert models of crime reduction, but rather should begin with an understanding of the organic or normative processes that seem to impact offending patterns over the life course. That is, if turning 30 is the 'most effective crime-fighting tool' (Von Drehle 2010), then we should seek to learn as much as we can about that process and see if we can model these dynamics in our own interventions.

However, recognizing that the practical implications of such a perspective remain seriously underdeveloped (see also Porporino 2010); Maruna and LeBel (2010) go on to articulate one form of desistance-supporting intervention. Drawing on labeling theory, and on their earlier work (cited above), they make the case for an approach to corrections which stresses pro-social labeling; as well as avoiding negative labeling, this requires practices and systems that expect, invite, and facilitate positive contributions and activities from people subject to supervision and that then certify and celebrate redemption or rehabilitation – de-labeling and de-stigmatizing the reformed offender.

McNeill (2012) has similarly argued, further to his original arguments for a desistance paradigm, that the field of corrections needs its own "Copernican Correction" – one in which supervision and support services revolve around the individual change process, rather than requiring offenders' lives to revolve around programs and interventions. Moreover, he argues for a shift away from seeing the "offender" as the target of the intervention (the "thing" to be fixed) to seeing the broken relationships between individuals, communities, and the state as the breach in the social fabric (or breach of the social contract) that requires repair. Importantly, this casts correctional agencies less as agents of "correction" and more as mediators of social conflicts. The objective becomes not the correction of the deviant so much as the restoration of the citizen to a position where she/he can both honor the obligations and enjoy the rights of citizenship.

From Paradigms to Practices or Not

The arguments reviewed in the previous section are important, but they are arguably more concerned with *reframing* supervision than with *redesigning* it in practice. The clearest attempts at tackling the latter challenge are perhaps to be found in reports prepared by McNeill (2009) and McNeill and Weaver (2010) for the Scottish Government and the National Offender Management Service of England and Wales, respectively. Both of these reports represent attempts to respond to policymaker and practitioner requests for a more explicit articulation of the practical implications of desistance research for supervision practice; both share a similar reticence in responding to these requests. As McNeill and Weaver (2010: 6) explain:

> One of the 'problems' with desistance research is that it is not readily translated into straightforward prescriptions for practice... As Porporino (2010: 61) has recently suggested: 'Desistance theory and research, rich in descriptive analysis of the forces and influences that can underpin offender change, unfortunately lacks any sort of organised practice framework.'
>
> However, though this is a practical problem, it is not necessarily a weakness. Even if we wished that there was a 'desistance manual' that could be prescribed for practitioners, there is not... [D] esistance research itself makes clear that offenders are heterogeneous, their needs are complex and their pathways to desistance are individualised. Overly generalised approaches to interventions therefore are themselves inconsistent with desistance research. It follows that evidence based practice can only really emerge from practitioners' reflective engagement and continual dialogue with those individuals with whom they work, and with the research that should inform how they work.

With this in mind, rather than offering a desistance manual or a desistance program, McNeill's (2009) paper *Towards Effective Practice in Offender Supervision* is an attempt to summarize evidence about both desistance and "What Works?" in order to inform evidence-informed reflective practice. Instead of offering a predesigned and therefore homogenized intervention, he attempts to articulate the range of issues and questions with which a reflective practitioner would have to engage on a case-by-case basis in seeking to support desistance. The "offender supervision spine" that he delineates includes explicit suggestions about how to approach the preparatory, relationship-building stage of supervision, as well as assessment, planning, intervention, and evaluation. McNeill's (2009) suggestion is that a supervisor working their way along this spine (in partnership with the supervisee) must be continually developing and testing evidence-informed "theories of change," seeking to work out together "why and how we think that doing what we propose to do will bring about the results that we seek" and then to implement that plan and to evaluate its progress.

Beyond such emergent models of the supervision process, a number of broader practical implications of desistance research have been identified in the literature. In a very recent overview, for example, McNeill et al. (2012b) identify eight broad principles, the genesis of which is apparent above:

1. Desistance, for people who have been involved in persistent offending, is a difficult and complex process, and one that is likely to involve lapses and relapses. Criminal justice supervision must be realistic about these difficulties and find ways to manage setbacks and difficulties constructively. It may take considerable time for supervision and support to exercise a positive effect.
2. Since desistance is an inherently individualised and subjective process, approaches to supervision must accommodate and exploit issues of identity and diversity. One-size-fits-all interventions will not work.
3. The development and maintenance not just of motivation but also of hope become key tasks for supervisors.
4. Desistance can only be understood within the context of human relationships; not just relationships between supervisors and offenders (though these matter a great deal) but also between offenders and those who matter to them.
5. Although the focus is often on offenders' risks and needs, they also have strengths and resources that they can use to overcome obstacles to desistance – both personal strengths and resources, and strengths and resources in their social networks. Supporting and developing these capacities can be a useful dimension of supervision.
6. Since desistance is in part about discovering self-efficacy or agency, interventions are most likely to be effective where they encourage and respect self-determination; this means working *with* offenders not *on* them.
7. Interventions based only on developing the capacities and skills of people who have offended (human capital) will not be enough. Probation also needs to work on developing social capital, opportunities to apply these skills, or to practice newly forming identities (such as 'worker' or 'father').
8. The language of practice should strive to more clearly recognize positive potential and development, and should seek to avoid identifying people with the behaviors we want them to leave behind.

This summary was produced as part of an ongoing project which aimed to develop a much more comprehensive and innovative response to the challenge to "operationalize" desistance (explained more fully in McNeill et al. 2012a). Rejecting the prospect of an often futile, counterproductive, and disrespectfully one-sided conversation between research and practice, the collaborators in this project have sought to foster a dialogue about supporting desistance involving academics, policymakers, managers, practitioners, ex/offenders, and their families and supporters.

The project is entitled "Discovering Desistance" and is funded by the UK Economic and Social Research Council (award no. RES-189-25-0258). Partnered in the USA by

Prof. Faye Taxman's Center for Advancing Correctional Excellence (http://www.gmuace.org/), it aims to explore the experience and knowledge of these different stakeholders in relation to desistance from crime and how correctional supervision in the community can best support it. To this end, the project involves three key elements:

1. Developing, with key stakeholders, user-friendly methods of disseminating existing research about desistance from crime and about supporting desistance in offender supervision. This involved working with ex/offenders, practitioners, and an independent film production company to make a documentary film about desistance (The Road from Crime; http://www.iriss.org.uk/resources/the-road-from-crime) which was released in July 2012.
2. Facilitating dialogue through the "Discovering Desistance" blog site (see http://blogs.iriss.org.uk/discoveringdesistance/) where academics, ex/offenders, and practitioners have all made key contributions to developing the discussion.
3. Running a series of stakeholder workshops (across four jurisdictions) which aim at coproducing a set of clear recommendations about the further development of "practice for desistance" and beginning to delineate the features of an organized framework for offender supervision practice to support desistance. This is the focus of the ongoing final stage of the project.

Though it is too soon to report the findings of these workshops here, it is interesting to note that the ideas and proposals that they have generated extend far beyond supervision practice and into much broader aspects of reentry and reintegration. Some of the recurring "propositions" suggested in the UK-based workshops include the following:

1. **Make greater use of reformed offenders**: Participants called for more meaningful service-user (i.e., supervisee) involvement in the design, delivery, and improvement of policies and provision across the criminal justice system, involvement that could be part of clear career routes and would include developing ways of recognizing and rewarding skills. Participants argued that greater involvement of ex/offenders in mentoring schemes should be a key part of this involvement.
2. **Reduce the reliance on imprisonment**: Participants argued that there is a need to reduce the prison population (especially women, black men, those with mental health issues, and those on short sentences), with freed-up money reinvested in community justice.
3. **Reorientate the philosophy of probation**: They called for a rethinking of probation, making it more "holistic" or "humanized," more focused on the service user's wants, strengths, and aspirations, as well as aiming for more community involvement and a greater degree of flexibility and creativity.
4. **Reconnect probation to local communities**: Participants argued that, in the future, probation offices and officers need to become better connected with local communities.
5. **Mobilize wider support networks**: All of society needs to take on a responsibility for helping people stop offending (organizations, families, and individuals).
6. **Focus on the positive**, **not the negative or risks**: Participants suggested that criminal justice needed to focus more readily on the positives and what people have achieved *and can achieve* in the future.
7. **Suggestions for supervision**, **release**, **and reintegration**: Community supervision needs to work to challenge inequality and promote equality, equalizing life chances and contributing to social justice (pursuing both substantive equality and equality in the criminal justice process).
8. **Redraft the** [UK] **Rehabilitation of Offenders Act**: Participants suggested the need to redraft the legislation around criminal records, so as to really encourage and recognize rehabilitation, not stand in the way of it. Reformed offenders should have the opportunity to have their record "spent" much earlier than is currently the case.

9. **Educate the general public about the processes of desistance**: They suggested that there is a need to better educate the general public about the process of leaving crime behind and the lives of current and former service users in order to break down the "them" and "us" mentality; this would ensure that there is better public understanding that people are capable of change and that we all have a part to play in supporting change. Criminal justice agencies ought to demonstrate that positive change is possible and show that it is common.
10. **Give people hope**; **show them they have a future**: The criminal justice system needs to become more acquainted with hope and less transfixed with risk, pessimism, and failure.

Conclusions

Though, as this entry demonstrates, many academics have worked hard in recent years to conduct and to disseminate desistance research, their voice in the debate about criminal justice reform is and should be just one among many. Others have different kinds of expertise to bring to this discussion. The Discovering Desistance project has been more about harnessing different forms of expertise than privileging or prioritizing one perspective. Perhaps its most important contribution has been to demonstrate the potential of "coproduction" in supporting change – at both the personal and the systemic levels (Weaver 2011).

As such, the impact on criminal justice – and more narrowly on correctional supervision – of engaging with desistance research is beyond the control of scholars and researchers; those working in and living with the correctional system have already started to talk and think about how and why people build new lives; the "desistance genie" is well and truly out of the bottle. In some respects, the work of academics in trying to stimulate new ways of thinking about supporting change in the criminal justice system and in the practice of supervision is complete. In other respects, it has barely begun. But while academics and researchers have plenty more work to do developing a robust, evidence-informed understanding of these processes and of what supports them, arguments over language, social attitudes, policy developments, and practice processes should not and cannot wait for research to provide "answers." Rather, all of the stakeholders engaged with supervision – policymakers, practitioners, parolees and probationers, and families – need to press on with the urgent business of working out what to do to continue discovering and supporting desistance together.

Related Entries

- ▶ History of Corrections
- ▶ Penal Justice and Social Injustice
- ▶ Penal Philosophy and Sentencing Theory
- ▶ Punishment and the State
- ▶ Rehabilitation
- ▶ Sentencing Research
- ▶ Theories of Punishment

Recommended Reading and References

Bottoms A, McWilliams W (1979) A non-treatment paradigm for probation practice. B J of Soc Work 9(2):160–201

Burnett R (1992) The dynamics of recidivism. University of Oxford Centre for Criminological Research, Oxford

Farrall S (2002) Rethinking what works with offenders: probation, social context and desistance from crime. Willan Publishing, Cullompton

Glueck S, Glueck E (1937 [1966]) Later criminal careers. Kraus, New York

Lewis DA (1990) From programs to lives: a comment. Am J Community Psychol 18:923–926

Maruna S (2000) Desistance from crime and offender rehabilitation: A tale of Two research literatures. Offender Program Rep 4(1):1–13

Maruna S (2001) Making good. American Psychological Association, Washington, DC

Maruna S, LeBel T (2003) Welcome home? Examining the "re-entry court" concept from a strengths-based perspective. West Criminol Rev 4(2):91–107

Maruna S, LeBel T (2009) Strengths-based approaches to reentry: extra mileage toward reintegration and destigmatization. Jpn J Sociol Criminol 34:58–80

Maruna S, LeBel T (2010) The desistance paradigm in correctional practice: from programmes to lives.

In: McNeill F, Raynor P, Trotter C (eds) Offender supervision: new directions in theory, research and practice. Willan, Cullompton

Maruna S, Porter L, Carvalho I (2004) The liverpool desistance study and probation practice: opening the dialogue. Probat J 51(3):221–232

McNeill F (2003) Desistance-focused probation practice. In: Chui W-H, Nellis M (eds) Moving probation forward: evidence, arguments and practice. Pearson Education, Harlow

McNeill F (2006) A desistance paradigm for offender management. Criminol Crim Justice 6(1):39–62

McNeill F (2009) Towards effective practice in offender supervision. Scottish Centre for Crime and Justice Research, Glasgow. Available online at: http://www.sccjr.ac.uk/documents/McNeil_Towards.pdf

McNeill F (2012) Counterblast: a Copernican correction for community sentences. Howard J Crim Justice 51(1):94–99

McNeill F, Batchelor S (2004) Persistent offending by young people: developing practice, issues in community and criminal justice, Monograph No. 3. National Association of Probation Officers, London

McNeill F, Weaver B (2010) Changing lives? Desistance research and offender management. Scottish Centre for Crime and Justice Research, Glasgow. Available online at: http://www.sccjr.ac.uk/documents/Report%202010_03%20-%20Changing%20Lives.pdf

McNeill F, Batchelor S, Burnett R, Knox J (2005) 21st century social work reducing reoffending: key skills. Scottish Executive, Edinburgh

McNeill F, Farrall S, Lightowler C, Maruna S (2012a) Reexamining “evidence-based practice” in community corrections: beyond “a confined view” of what works. Justice Res Policy 14(1):35–60

McNeill F, Farrall S, Lightowler C, Maruna S (2012b) How and why people stop offending: discovering desistance. Insights: evidence summaries to support social services in Scotland, No. 15, April 2012, Institute for Research and Innovation in Social Services. Available online at: http://www.iriss.org.uk/sites/default/files/iriss-insight-15.pdf

Ministry of Justice (2010) Breaking the cycle. Effective punishment, rehabilitation and sentencing of offenders. Ministry of Justice, London

Porporino F (2010) Bringing sense and sensitivity to corrections: from programmes to ‘fix’ offenders to services to support desistance. In: Brayford J, Cowe F, Deering J (eds) What else works? Creative work with offenders. Willan, Cullompton

Raynor P, Vanstone M (1994) Probation practice, effectiveness and the non-treatment paradigm. B J Soc Work 24(4):387–404

Rex S (1999) Desistance from offending: experiences of probation. Howard J Crim Justice 36(4):366–383

Weaver B (2011) Co-producing community justice: the transformative potential of personalisation for penal sanctions. B J Social Work 41(6):1038–1057

Desistance from Crime

Shawn D. Bushway[1] and Raymond Paternoster[2]
[1]School of Criminal Justice, University at Albany, State University of New York, Albany, NY, USA
[2]Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA

Synonyms

Criminal career; Desistors; Termination

Overview

Early understandings of the cessation from crime considered desistance to be the event of moving from a state of committing crime to a state of not committing crime. Gradually, however, scholars have begun to understand desistance not as an event but as a process. Fagan (1989) was the first to recognize this, differentiating the process of desistance, defined as the reduction in the frequency and severity of offending, from the event of quitting crime. Le Blanc and Fréchette (1989) also referred to desistance as a set of processes that leads to the cessation of crime, using the term deceleration to refer to a reduction in the frequency of offending prior to cessation. Continuing the dialog, Laub and Sampson (2001) explicitly separated the process of desistance from the termination of offending, which they viewed as the outcome of desistance. There are currently in the literature several excellent reviews of possible theoretical explanations for desistance – most notably Laub and Sampson (2003). This essay describes desistance by explicitly mapping the different processes of desistance to different stochastic time series models.

Describing Desistance

In its most basic form, a time series of individual offending can be *described* by the following equation, which is known as an autoregressive time series:

$$Y_t = \alpha + \rho Y_{t-1} + e_t \qquad (1)$$

where e_t is a time series of uncorrelated shocks. A key assumption of time series analysis is that the process is *stationary*, which simply means that the parameters of the model are stable throughout the time period. This kind of stationary process cannot create a long-term desistance trajectory that declines to zero over time. The path described by Eq. 1 will move to an equilibrium level and then stay flat with short-term variation around the equilibrium line – the very formulation means that there can be no meaningful shift in the level of offending. State dependence (past offending causes current offending) and individual heterogeneity (differences between individuals in the stable propensity to commit crime) as captured in the lagged Y term in Eq. 1 ***cannot*** explain desistance. Therefore, *desistance is inherently a nonstationary process.*

There are four broad classes of nonstationary time series. The first is a *series with a trend*. This trend is based on age. The trend predetermines the path. With a trend, Eq. 1 becomes Eq. 2

$$Y_t = \alpha_t + \rho Y_{t-1} + e_t \qquad (2)$$

This elementary model does not try to explain the existence of the trend, except in the most basic or general terms. The best example in criminology for a desistance theory that appeals to a basic trend is Gottfredson and Hirschi's (1990) theory of self-control. In this theory, any change in an individual's time series trend in offending over time is simply attributed to the "inexorable aging of the organism" (1990: 141). Since age is the time marker in this time series, saying age explains desistance is simply the same thing as saying that there is an undefined trend that mimics the trend in age (Bushway et al. 2001). Glueck and Glueck's (1974) maturational theory of the decline in crime over time is but one small step removed from Gottfredson and Hirschi's assertion about age (1990). They are careful to explicitly distinguish age from maturation – which means that the maturational process need not occur at the same age for everyone. However, all this statement does is to extend Eq. 2 to say that there is an unknown distribution of time trends in the population – there is a definitive time trend but everyone does not have the same time trend. This claim leaves open the possibility that this maturational process is preprogrammed and deterministic. Laub and Sampson (2003: 33), for example, characterize these kinds of developmental theories as being preprogrammed – essentially fixed trends: "[d]evelopmental accounts . . . focus on regular or lawlike individual development over the life span."

The second type of time series, a *cointegrated time series*, captures the counterargument to Sampson and Laub's characterization of the developmental path. Here, Eq. 3 is developed by adding a time-varying covariate X_t. The coefficient on X_t is time constant. This variable trends in the same way as criminal propensity.

$$Y_t = \alpha + \rho Y_{t-1} + \delta X_t + e_t \qquad (3)$$

This basic model in which time-varying covariates can explain the long-term pattern of desistance fits with the class of theoretical models in which theorists simply extended existing theories of the onset of crime to account for its desistance. For example, Agnew (2005) argued that the bulk of offenders desist from crime simply because the strains that they experienced as adolescents that launched them into crime in the first place (school, relationship, and job strains) diminish over time, and the ability to adapt in a conventional way to existing strains increases as they entered adulthood. The movement into adulthood, then, comes with both fewer and/or less intense strains and/or an increased capability to adapt to strain in a nondeviant way. Similarly, Akers (1998: 164) argued that the most important predictor of all dimensions of offending, including desistance, is involvement with delinquent peers: ". . . the single best predictor of the onset, continuation, or desistance of delinquency is differential association with law-violating or norm-violating peers." Most existing theories of crime responded to the new conceptual terrain brought about by the criminal career perspective, then, by simply insisting that they could just as easily explain desistance as they could onset or other dimensions of offending.

Developmental or maturational theories of crime can also be thought of as describing a cointegrated time series rather than a deterministic trend to the extent to which the theorist describes a variable or process that explains the change in propensity over the life course. For example, Gove (1985) posits that there are biological and psychological factors over time that peak and decline in the same manner as offending propensity. These factors are plausibly cointegrated with offending propensity.

Although they are skeptical about whether this can be done, Gottfredson and Hirschi (1990) acknowledge the possibility that time-varying covariates can explain long-term change. On the empirical side, Osgood (2005) advocates inserting time-varying covariates with time constant parameters into growth curve models in an attempt to explain the age crime curve. Within the growth curve framework, Osgood (2005) suggests testing to see if the time-varying covariates can detrend the data. This basic approach has been applied by Nieuwbeerta and Blokland (2005) where they look to see how much marriage and employment can explain the age crime curve. It is also seen in Sweelen et al. (2013) in which they look to see how much a set of time-varying covariates can explain the divergence between those who desist from and those who persist in crime. In each case, the researchers are looking to see if the time-varying covariates can make a nonstationary time series stationary – with time constant parameters, the only way this is possible is if the covariates themselves trend or track in the same manner over time as offending propensity.

The third type of time series process that can explain or accommodate nonstationarity is a *time series with a structural break*. A structural break implies that there are two or more sets of parameters, meaning that the causal process is different across different time periods.

$$Y_t = \begin{cases} \alpha_a + \rho_a Y_{t-1} + e_{at} \text{ if } & t<T \\ \alpha_b + \rho_b Y_{t-1} + e_{bt} \text{ if } & t> = T \end{cases} \tag{4}$$

Of course, there can be more than one structural break. There are elements of structural breaks that are harmonious with several desistance theories. For example, the notion of age-graded causal factors is entirely consistent with the idea that the value of coefficients on some time-varying variable changes over time.

A more general way of thinking about structural breaks is that some relatively time-stable component of an individual, such as self-control, changes over time. This is only relevant if life events and social context interact with self-control to affect behavior. In Thornberry's interactional model, for example, the exact nature of state dependence depends in meaningful ways on the individual's relatively stable characteristics (Thornberry 1987). In Thornberry's interactional theory, those individuals who are heavily embedded in crime are less "dynamic," in that, they are less responsive to changes in their environment, and, therefore, are also less state dependent (Thornberry and Krohn 2005). Nagin and Paternoster built on this idea in their own version of an interactional theory when they posited that the impact of sanctions depended in meaningful ways on the person's level of self-control (Nagin and Paternoster 1994). Subsequent empirical work by Wright and colleagues (2004), as well as by Hay and Forrest (2008), have all found evidence for an interaction between life events and stable individual characteristics such as self-control. If this basic preference function shifts over time in purposeful ways, as suggested by Hay and Forrest (2008) and Giordano et al. (2007), then the same inputs and opportunities lead to different behaviors at different times – and state dependent processes can start to head people in a different direction. This situation, where a person experiences different causal processes depending on changes in their underlying personal preferences, extends interactional theories to accommodate a structural break and strengthens the ability of these types of theories to explain long-term changes in offending propensity.

Another type of desistance theory that accommodates a structural break are theories which anchors a change in crime to an offender's change in personal identity (Giordano et al. 2002, 2007; Paternoster and Bushway 2009). The importance of identity theories from this perspective is that they provide an explanation for how fundamental individual characteristics such as self-control can change from one time period to another manifesting itself as a structural break. Changes in identity can trigger fundamental shifts in how people value the future (time discounting) or value their social contacts. Simply saying that preferences change is easy – explaining the mechanism by which they change is both important and difficult (see Akerlof and Kranton 2010). Identity theorists like Giordano and her colleagues (Giordano et al. 2002, 2007) and Maruna and Farrall (Maruna 2001; Farrall 2005) offer social psychological theories of desistance which revolve around structural breaks in the process that generates crime. Basing their views on a symbolic interactionist foundation, Giordano et al. (2002) argue that desistance requires substantial cognitive transformations or "upfront" cognitive work such as the development of a general openness to change, receptivity to "hooks for change," and consistent support from social others. In a later revisiting of this view, Giordano et al. (2007) developed a desistance theory that relies much more heavily on external, social processes – the regulation of emotions and the emotional identity (an "anger identity") of ex-offenders as they struggle with getting out of crime. Maruna (2001) also adopted a theory of desistance that relies on notions of the actor's identity, though not one premised on a change in identity. For Maruna (2001), "making good" does not so much involve an intentional change in the desister's identity from bad to good as it does a reinterpretation of one's criminal past to make it consistent with their current pro-social identity.

The fourth major type of nonstationary time series is a *random walk*, a well-known form that has been found occurring in many contexts, including the stock market price of a company and the financial status of a gambler. Random walks have a unit root:

$$Y_t = \alpha + Y_{t-1} + e_t \qquad (5)$$

According to Eq. 5, behavior in a given period is simply where you were in the previous period, plus a constant and a shock. The series has an infinite memory, since any shock is permanently incorporated into the time series. Random walks do not, therefore, return to any mean. The same formula can generate flat, increasing, decreasing, or U-shaped curves, depending entirely on the time series of uncorrelated shocks e_t.

This description of a random walk is consistent with Laub and Sampson's (2003: 34) characterization of life course theories of desistance as the result of a series of random events or "macro-level shocks largely beyond the pale of individual choice (for example, war, depression, natural disasters, revolutions, plant closings, industrial restructuring)." Random walks are inherently unpredictable, and as described by Laub and Sampson (2003: 33–34), this lack of predictability is the key factor which distinguishes life course trajectories from predetermined developmental trajectories:

> Developmental accounts...focus on regular or lawlike individual development over the lifespan. Implicit in developmental approaches are the notions of stages, progressions, growth and evolution... with the imagery being one of the execution of a program written at an earlier point in time. ... In contrast, life-course approaches ... emphasize variability and *exogenous* influences on the course of development over time that cannot be predicted by focusing solely on enduring individual traits. (emphasis added)

Another way to discuss the time series properties of life course theories is to consider the key life course assertion that the impact of life events depends on when they occur in a person's life. This is the notion that "timing matters." To the extent to which this timing dependence is predictable, it is consistent with time series models with structural breaks because the implication of timing dependence is that there are simply different models for different time periods. If there are a small number of changes, and these changes are tied to observable changes in identity, then this age-gradedness should be both predictable and identifiable. But if there are many structural breaks, and these breaks are tied to malleable social contexts, the age-gradedness becomes much more unpredictable. Indeed, a random walk can be characterized as a time series with *N* structural breaks, where *N* converges to the total number of periods in the time series.

The main difference between life course theories (random walks) and identity theories (structural breaks) is the number of breaks. In a world with many breaks, predicting long-term change is difficult. The result is a change in focus to explaining change in any given period, which is driven by these relatively exogenous life events. This conclusion is consistent with empirical practice – if a time series is a true random walk, with no trend and no cointegrated time series, the only feasible strategy is to explain period-to-period, that is, short-term, change. With a time series characterized by random walks, it is simply not possible to explain any long-term pattern because that long-term pattern is driven by random shocks. Ironically, this interpretation of life course theory implies that it is neither possible nor even interesting to study a life course "trajectory" since only period-to-period change contains interesting information – there is no meaning to a *life* course.

All theories of desistance must fit into one of the four basic categories of nonstationary time series models – trends, cointegrated series, series with a structural break, and random walks. Given the distinct empirical character of each of these four basic types of time series, a serious examination of individual time series characteristics should be a fruitful avenue for future empirical research. Further explication of theories within the framework provided by the extensive literature on time series processes should also help to clarify and formalize theories of desistance. Readers interested in seeing empirical examples of this approach should peruse Paternoster and Bushway (2009; Bushway and Paternoster 2012), where some basic illustrations are provided using data from the Cambridge Study in Delinquency Development (CSDD) data (Farrington et al. 2006).

Key Issues/Controversies

Bushway and Paternoster (2012) argue that the evidence shows that desistance is not inevitable as suggested by Sampson and Laub (1993; Laub and Sampson 2003) nor is it a simple part of the biological aging process as suggested by Gottfredson and Hirschi (1990). Moreover, they believe the evidence suggests that individuals stop committing crime not because they cannot physically commit crime anymore, but because they *choose* not to. Some choose to exit before others who wait until much later in their lives to quit crime, and as a result, there is a long right hand tail in the age distribution of offending. This long right hand tail also casts doubt on a strictly structural version of desistance which attributes the initial thrust into conformity to an acquiring of "turning points" or pro-social roles like jobs and marriages. While there is a convincing body of research that documents the ability of marriage and work to decrease crime, this work frequently does not speak very directly or clearly to the *causal mechanism* by which this effect occurs (Sampson et al. 2006). If the explanation is entirely or immediately structural, desistance would be highly correlated with the arrival rates of first marriages and stable employment during the 20s and into the 30s as people move into adulthood. And, indeed, a large portion of desistance clearly does occur between the ages of 20 and 40. *But, it cannot be ignored that employment and marriage have been available states for 20 years by age 40*. A simple matching or sorting story in which people desist when matched to jobs and spouses should not require more than 20 years before it reveals itself.

Of course, it is possible that work (and potentially marriage) has a differential impact depending on age such that work is involved in the desistance process but only when offenders reach a certain age. But this explanation would imply that something about the individual or her set of circumstances has changed with age, their identity and preferences, for example, and this change in turn leads to different choices by the individual. The typical interpretation is that the effect of these variables is age-graded, but this term simply describes what has to be explained. Another interpretation is that these factors (such as jobs and marriages) have a different impact on different kinds of people, and different kinds of people select into marriage and employment at different ages. Research on employment and crime is now increasingly showing that the established "fact" that employment is bad for youth (but good for adults) is entirely an artifact of selection. Strong controls for selection show that employment has the same modest *negative* impact on crime for youth as it does for adults. Entering into pro-social roles may have a role to play in desistance, but perhaps the acquisition of such roles is only part of the picture and comes later in the desistance process when other obstacles have first been overcome.

The facts of desistance state loudly and clearly that desistance cannot be explained either by strictly biological or structural explanations. If not biology and if not the immediate acquisition of pro-social roles, what then? One possible explanation lies in a person's identity and the corresponding changes this brings in how they weigh the inputs of their decision making, their preferences, and how they make choices (Akerlof and Kranton 2010).

Theoretical Frontiers: Identity Theory

There is a long intellectual tradition in sociology and social psychology which emphasizes the importance of one's identity (Stryker 1968). In recent years, economists have also argued that the preferences people have and ultimately the decisions that they make are influenced by who they think they are or who they want to become (Akerlof and Kranton 2010). Identity is important for numerous reasons, the most important for our concerns is that it motivates and provides a direction for behavior (Stryker 1968). A person's actions are seen as expressions of their self-identity – people intentionally behave in ways that are consistent with who they think they are. In interaction with others, therefore, people project an identity of who they are, and a primary vehicle for communicating to others who "one is" is through one's behavior.

Identities or selves vary in terms of their temporal orientation. Some selves are oriented toward the present as the *working self* (Markus 1977, 1983). The working self is that component of the self that can be accessed at the moment and is based upon the individual's here-and-now experience. In addition to a sense of who and what one is at the moment, or a self that is fixed on the present, people also have a sense of self that is directed toward the future. This future oriented self is defined positively as the self they would like to become and negatively defined as the self they would not want to become or fear that they might become. Markus and Nurius (1987) have defined this future orientation of the self as a *possible self*. The possible selves "are conceptions of the self in future states" (Markus and Nurius 1987: 157) and consist of goals, aspirations, anxieties, and fears that the individual has as to what he could become. While the working self is aware of what skills the person has and does not have and what the person can and cannot do in the present, the possible self is directed toward the future and what it is possible to be and what the person would not like to be. A person may, for example, see herself currently (the working self) as a thief, drug user, poor father, unskilled worker, but may see herself in the future as working in a job (though perhaps for minimum wage), legitimately buying things for her family, owning a used car, and ceasing to use drugs and commit crime. A person may, however, also fear that she may turn out to be a burned-out addict, riddled with disease, homeless, childless, jobless, and destined to die alone.

An important consequence of a possible self is that it provides directed motivation for one's behavior (Markus and Nurius 1987). Possible selves, both positive and negative, therefore, not only contain satisfying images of what the person would like to be or desperately fears becoming, they can also provide a specific and realistic set of instructions or a "road map" directing what one can do to achieve the positive future self and avoid the negative possible self. This is referred to as the self-regulating component of the possible self. The self is self-regulating because, among other things, it compares the past and current working self with the possible self and provides specific directions, strategies, or plans for narrowing any discrepancy between the two, thereby connecting the present with the future. Motivation is generated and is more likely to be successful, then, when the person not only has a goal of self-improvement but specific and realistic means to reach that goal. While the positive possible self is frequently a longer-term goal, an initial movement out of a deviant identity is more likely to be based on a motivation to avoid a feared self than it is a desired to achieve a positive self.

Though stable, identities clearly can and do change. A working identity as a criminal offender can change to a more conventional identity when the person thinks of a conventional identity as a positive possible self and an identity of a burned-out ex-con with no friends or possessions as a negative possible self or feared self. Contemplation of a possible self that does not include criminal offending in turn occurs when the working identity of criminal is perceived to be unsatisfying or disappointing. As one begins to find less success and satisfaction with the criminal identity, it is likely to conjure up negative possible selves – long terms in prison with young hoodlums, a violent death during a crime, small payoffs from criminal enterprises. These negative possible selves and the activation of positive selves – a working person, a person with a good spouse, a giving father, a law abider – can provide both the motivation and direction for change. Before one is willing to give up his working identity as a law breaker, then, one must begin to perceive it as unsatisfying, thus weakening one's commitment to it. This weakening of one's commitment to a criminal identity does not come about quickly, nor does it come about in response to one or two failures, but only gradually and only as the result of linking together of many failures and the attribution of those linked failures to one's identity and life as a criminal.

The process of desisting from crime first requires an offender to recognize that their working identity of offender is no longer satisfactory and their attachment to this identity must be weakened. The weakening of a criminal identity comes about gradually and comes about as a result of a growing sense of dissatisfaction with crime and a criminal lifestyle. The dissatisfaction with crime is more likely to lead to a conventional possible self when failures or dissatisfactions with many aspects of one's life are linked together and attributed to the criminal identity itself. It is not just that one has experienced failures but that diverse kinds of failures in one's life become interconnected as part of a coherent whole which leads the person to feel a more general kind of life dissatisfaction, the kind of life dissatisfaction that can lead to intentional identity change.

It is such a new understanding of one's life that leads to the effort to intentionally change it, or as Shover (1996: 132) put it: "[t]his new perspective symbolizes a watershed in their lives…[t] hey decide that their earlier identity and behavior are of limited value for constructing the future." The importance of this is that one consequence of this linking together of diverse failures in life, or what Baumeister (1994) calls the crystallization of discontent, is that after this occurs, the dissatisfactions that one has experienced now has implications for the future. Events that seemed atypical and isolated that have been linked are now seen as interrelated and therefore both less easily dismissed and seen as likely to continue to occur in the future. The projection into the future of continued life dissatisfaction leads the person to begin to seek changes.

Kiecolt (1994: 56) has argued that intentional self-change is unlikely to be successful without what she calls "structural supports" for change. These supports "provide individuals with means and opportunities for effecting self change" and include self-help groups and professional changers such as psychiatrists and social workers. As a separate condition for successful self-change, Kiecolt includes the assistance of social supports such as friends, family members, and spouses and partners.

Obviously if successful self-change is going to occur, the benefits of a new identity must outweigh the costs of leaving the old one.

However economically marginal a life of crime is, criminal offenders, particularly those with official records of arrest, conviction, and incarceration, find legitimate employment opportunities, even in the secondary labor market, very restricted. Some opportunity to secure a conventional job must be available for criminal offenders to desist, no matter how strong the motivation to change their identities and selves. Generally, anyone exiting one role or identity needs access to alternative sources of employment – nuns leaving religious orders no less than prostitutes leaving "the trade" must find outside employment. Without these kinds of structural supports, identity change becomes difficult. Social supports, whether in the form of friends, spouses/partners, jobs, or professional help, are important in self-change because they provide the one in the throes of a crystallization of discontent with an alternative existence or identity.

In an identity theory of desistance, changes in friendship networks and the securing of alternative jobs and vocations are important because they help maintain or bolster a fledging changed identity. To be clear, securing jobs, attracting new partners, and involvement with new friends come about *after* a change in identity has occurred. The change in identity has already occurred in the mind of the person; he has weighed the costs and benefits of the exiting identity and alternatives and is behaving in ways that conform to the new possible self.

Empirical Frontiers of Desistance Research: Long-Term Hazard Models

Bushway et al. (2009) show convincingly that the main types of growth curve models largely discard as noise information about change from the individual trajectories. This finding should be particularly troubling for desistance scholars, who are fundamentally interested in studying change. But how can researchers study change if individual trajectories are too imprecise and long-term trajectory models essentially ignore the very change that desistance scholars are interested in studying? Another possibility for examining desistance processes would be to turn to a study of recidivism. Thirty years ago, recidivism and desistance were complementary measures. Those who failed after a certain period were recidivists, and those who did not were desistors.

This static approach to thinking about recidivism and desistance has been effectively rejected. Now, cutting edge recidivism studies focus on hazard rates of offending over time and cutting edge desistance studies focus on measuring trajectories of offending rates over time. But, it is a well-known fact in statistics and quantitative criminology that these two models (hazards and trajectory-type models) are actually measuring the same concept, with hazard rate models focusing on *short-term* change in the propensity to offend and trajectory models focusing on *long-term* change in the propensity to offend. For example, having noted that the hazard rate focuses on the hazard of involvement in a given criminal event, Hagan and Palloni (1988) observe that

> (T)he expected number of criminal events during the age interval being examined is a unique function of these hazards. This expected number of criminal events is what Blumstein et al. are estimating when they calculate lambda (offending rate). So, lambda is a summary of the combined hazards of criminal events of various orders over a period time. (Hagan and Palloni 1988: 97)

As a result, the use of trajectories of rates to study desistance has brought the study of desistance conceptually very close to the study of recidivism.

In their article, Hagan and Palloni (1988) present arguments for focusing on the causal nature of the events, rather than on the rate of offending. At the time they made their argument, however, empirical methods only allowed for the estimation of time-stable rates for individuals. The ability to capture time variation in offending rates while controlling for individual heterogeneity, combined with the new emphasis on the *process* of desistance, provides a persuasive counterargument for a focus on the more long-term perspective.

The potential productivity of using hazard models with long-term data was highlighted by Barnett et al. (1989), who applied their insight about desistance and trajectories of offending to an analysis of recidivism using a hazard model. Barnett et al. (1989) examined the risk of recidivism until the 30th birthday among a small group of 88 offenders who had at least two convictions before their 25th birthday. Each offender was given a probability of a new offense as well as a desistance parameter that indicated the probability of instantaneously desisting after each event. Thus after each criminal event, the offender had the choice of continuing to offend at the given rate (λ) or desisting. By dividing the offenders into two groups, "frequents" (annual $\mu = 1.14$ or a 1 in 320 daily chance of offending) and "occasionals" (annual $\mu = 0.4$ or a 1 in 913 daily chance of offending), they were able to quite reliably predict future patterns of recidivism. The only complication in their models was a small group of "frequent" offenders who had appeared to desist from crime according to their predictions, but actually resumed a criminal career later in life. It was this small group of offenders they deemed "intermittent" for which their basic models were not adequate. They therefore called for "more elaborate models to incorporate the concept of intermittency, whereby offenders go into remission for several years and then resume their criminal careers" (p. 384).

Kurlychek et al. (2012) have attempted to learn about desistance in the short term by using survival models which can be tied to different models of desistance. Research on survival starts with a group of active offenders and then follows them for a period of time to model the risk of recidivism as well as the time (t) to recidivism. A hazard ratio is then estimated for each time period (t) as follows:

$$H(tg) = \frac{\#\ \text{arrested time}\ t}{\#\ \text{survived through time}\ t-1}$$

Those who have not failed by the end of the follow-up period may be assumed to have desisted from crime. However, it is also possible that they would have recidivated if they had been followed for a longer period of time, meaning that the observation was merely right censored. While much current recidivism research utilize the semi-parametric Cox regression strategy which does not force a functional form on the data over time (e.g., the models are more interested in explaining the effect of covariates over time), Kurlychek et al. (2010) suggest that the use of parametric methods might be more informative if one is attempting to explain the *actual form or time pattern of offending*.

This approach was first introduced to criminology by Maltz (1984) and extended by Schmidt and Witte (1988). For example, Schmidt and Witte (1988) applied a variety of functional forms to two cohorts of releases from the North Carolina prison system and were unsatisfied with the fit of any of the basic models. They identified the problem to be the basic assumption that everybody in the sample will fail if only followed up for a sufficiently long period of time. To address this issue, the authors then turn to what is known as a "split-population" or mixture model which allows for the fact that everyone does not fail. That is, some people do desist.

Split-population models include an extra parameter, often referred to by biostatisticians as the "cure" factor, which estimates the portion of the risk set that will never experience a failure (in other words, they will be "cured"). The cure factor is evidence of instantaneous desistance, or a structural break, particularly for individuals who have substantial rates of offending before the current offense. In this instance, individuals somehow decide (perhaps because they have changed their identity) to quit crime immediately. When applying split-population models to their data, Schmidt and Witte found that all split-population models outperformed their non-split model counterparts. However, Schmidt and Witte (1989) only follow their subjects for 5–7 years, not long enough to fully conclude that there has been desistance.

Kurlychek et al. (2012) estimated similar models using data with 18 years of follow-up from Essex County, NJ. They find that the two-parameter split-population exponential model fits the data almost as well as the more complex three-parameter lognormal counterpart and, in fact, out-performs this model in the later years of the data. It is striking how well this simple model can explain the observed behavior. Like the split-population lognormal model, the split-population exponential model assumes that there are two groups of offenders – those who have desisted at the beginning of the follow-up period and those who remain active. They find support for instantaneous desistance with the split-population lognormal and exponential model actually reaching quite similar conclusions about the size of the permanent desisting population at the outset of the follow-up period (the lognormal model is in the 20–23 % range while the exponential is 25–27 % range). This estimate is smaller than the estimates from Brame et al. (2003) looking at desistance after an arrest. However, it is still substantial. While the focus of most recidivism studies is on the high recidivism rates, the flip side here is that a full quarter of the sample of felony offenders desists after this conviction. Clearly, then, not all individuals are equally risky after a conviction. Indeed, because the exponential model assumes that the active offenders experience a constant risk of recidivism throughout the follow-up period, there is no evidence of declining hazard rates among the active offenders.

The length of the follow-up period in the Essex County dataset has a lot to do with the performance of the split-population exponential model. If the Essex County study had only followed offenders for 3, 4, or 5 years – typical follow-up periods for recidivism studies – the conclusions about the split-population lognormal and exponential models would have been different. Over this shorter window of time, the split-population lognormal model clearly performs better, but viewed over the entire 18-year follow-up period, the simpler, two-parameter split-population exponential model emerges as a formidable competitor. As more datasets with long follow-up periods are studied, it will be interesting to see how well the split-population exponential model performs, especially after the first few years of follow-up.

A final insight revolves around the concept of intermittency or reactivation of criminal careers after a period of dormancy or "temporary desistance" (Barnett et al. 1989). The concept of intermittency has been gaining ground in criminology in recent years and leads to certain theoretical and policy implications (e.g., the idea that desistance is always provisional). The Kurlychek et al. (2012) analysis is certainly consistent with the idea that a low rate offender can go for many years before committing a new offense. But intermittency is a particularly dynamic model of offending in which the offender goes from an active rate of offending to a zero rate of offending back to a fully active criminal career (what Laub and Sampson (2003) refer to as a "zigzag" criminal career). Barnett et al. (1989) moved to an intermittency explanation after they found evidence of a "fat" tail – higher rates of offending more than 5 years after the last offenses than could be explained by the exponential model. While Kurlychek et al. (2012) found support for their simple split-population exponential model, there was no fat tail even though they observed a more serious population over a longer follow-up period. As a result, they concluded that there is no evidence for intermittency, at least as described by Barnett et al. (1989). Replication and extension of these findings with other long-term datasets represents an important avenue for future research. In addition, there is a need to rejuvenate theoretical work in desistance. Previous efforts, while useful starting points, are not able to explain either the criminal patterns of contemporary offenders or the findings that have accumulated from recent empirical studies with different analytical models.

Related Entries

▶ Age-Crime Curve
▶ Cambridge-Somerville Youth Experiment

- ▶ Career Criminals and Criminological Theory
- ▶ Certainty, Severity, and Their Deterrent Effects
- ▶ Crime Specialization, Progression, and Sequencing
- ▶ Criminal Careers
- ▶ Criminal Careers and Public Policy Responses
- ▶ Desistance and Supervision
- ▶ Desistance from Gangs
- ▶ Examining the Effectiveness of Correctional Interventions
- ▶ Focused Deterrence and “Pulling Levers”
- ▶ Group-Based Trajectory Models
- ▶ Group-Based Trajectory Models and Developmental Change
- ▶ Growth Curve Models with Categorical Outcomes
- ▶ Identification Issues in Life Course Criminology
- ▶ Interactional Theory of Delinquency
- ▶ Interrupted Time Series Models
- ▶ Offender Change in Treatment
- ▶ Offender Decision Making and Behavioral Economics
- ▶ Parole and Prisoner Reentry in the United States
- ▶ Pathways to Delinquency
- ▶ Psychopathy and Offending
- ▶ Risk Factors for Prison Recidivism
- ▶ Transitional Work Programs for Ex-prisoners

Recommended Reading and References

Agnew R (2004) Why do criminals offend? A general theory of crime and delinquency. Roxbury Publishing, Los Angeles

Akerlof GA, Kranton RE (2010) Identity economics. Princeton University Press, Princeton

Akers RL (1998) Social learning and social structure: a general theory of crime and deviance. Northeastern University Press, Boston

Baumeister RF (1994) Meanings of life. Guilford Press, New York

Baumeister RF (1996) The crystallization of discontent. In: Heatherton TF, Weinberger JL (eds) Can personality change? American Psychological Association, Washington, DC, pp 281–297

Blokland A, Nieuwbeerta P (2005) The effects of life circumstances on longitudinal trajectories of offending. Criminology 43:1203–1240

Blokland A, Nagin D, Nieuwbeerta P (2005) Life span offending trajectories of a Dutch conviction cohort. Criminology 43:919–954

Bushway S, Paternoster R (2012) Understanding desistance: theory testing with formal empirical models. In: MacDonald J (ed) Measuring crime and criminality, advances in criminological theory, vol 17. Transaction Publishers, New Brunswick, pp 299–333

Bushway S, Piquero A, Broidy L, Cauffman E, Mazerolle P (2001) An empirical framework for studying desistance as a process. Criminology 39:491–516

Bushway S, Brame R, Paternoster R (2004) Connecting desistance and recidivism: measuring changes in criminality over the lifespan. In: Maruna S, Immarigeon R (eds) After crime and punishment: pathways to offender reintegration. Willan Publishing, Devon

Bushway SD, Sweeten G, Nieuwbeerta P (2009) Measuring long term individual trajectories of offending using multiple methods. J Quant Criminol 25(3):259–286

Doherty EE (2006) Self-control, social bonds, and desistance: a test of life-course interdependence. Criminology 44:807–834

Fagan J (1989) Cessation of family violence: deterrence and dissuasion. In: Ohlin L, Tonry M (eds) Crime and justice: an annual review of research, vol 11. University of Chicago Press, Chicago, pp 377–425

Farrall S (2005) On the existential aspects of desistance from crime. Symbol Int 28:367–386

Farrington DP, Coid JW, Harnett L, Jolliffe D, Soteriou N, Turner R, West DJ (2006) Criminal careers up to age 50 and life success Up to age 48: new findings from the Cambridge study in delinquency development. Home Office, London

Giordano PC, Cernkovich SA, Rudolph JL (2002) Gender, crime, and desistance: toward a theory of cognitive transformation. Am J Sociol 107:990–1064

Giordano PC, Cernkovich SA, Schroeder RD (2007) Emotions and crime over the life course: a neo-median perspective on criminal continuity and change. Am J Sociol 112:1603–1661

Glueck S, Glueck E (1974) Of delinquency and crime. Charles C. Thomas, Springfield

Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Gove WR (1985) The effect of age and gender on deviant behavior: a biopsychosocial perspective. In: Rossi AS (ed) Gender and the life course. Aldine de Gruyter, New York, pp 115–144

Hay C, Forrest W (2008) Self-control theory and the concept of opportunity: making the case for a more systematic union. Criminology 46: 1039–1072

Hirschi T, Gottfredson MR (1983) Age and the explanation of crime. Am J Sociol 89:552–584
Kiecolt KJ (1994) Stress and the decision to change oneself: a theoretical model. Soc Psychol Quart 57:49–63
Kurlychek M, Bushway S, Brame R (2010) Testing theories of desistance and intermittency by studying long-term survival models. Working paper
Laub JH, Sampson RJ (2001) Understanding desistance from crime. In: Tonry M (ed) Crime and justice: a review of research, vol 28. University of Chicago Press, Chicago
Laub JH, Sampson RJ (2003) Shared beginnings, divergent lives: delinquent boys to age 70. Harvard University Press, Cambridge, MA
Le Blanc M, Fréchette M (1989) Male criminal activity from childhood through youth. Springer, New York
Loeber R, LeBlanc M (1990) Toward a developmental criminology. In: Tonry M, Morris N (eds) Crime and justice: a review of research, vol 12. University of Chicago Press, Chicago, pp 375–473
Maltz M (1984) Recidivism. Academic, Orlando
Markus H (1977) Self-schemata and processing information about the self. J Pers Soc Psychol 35:63–78
Markus H (1983) Self-knowledge: an expanded view. J Pers 51:543–565
Markus H, Nurius P (1987) Possible selves: the interface between motivation and the self-concept. In: Yardley K, Honess T (eds) Self and identity: psychological perspectives. Wiley, New York
Maruna S (2001) Making good: how ex-convicts reform and build their lives. American Psychological Association Books, Washington, DC
Nagin DS, Paternoster R (1994) Personal capital and social control: the deterrence implications of a theory of individual differences in offending. Criminology 32:581–606
Osgood W (2005) Making sense of crime and the life course. The Annals 602:196–211
Paternoster R, Bushway S (2009) Desistance and the “feared self”: toward an identity theory of criminal desistance. J Crim Law Criminol 99:1103–1155
Piquero AR (2008) Taking stock of developmental trajectories of criminal activity over the lifecourse. In: Lieberman A (ed) The long view of crime: a synthesis of longitudinal research. Springer, New York, pp 23–78
Piquero AR, Blumstein A, Brame R, Haapanen R, Mulvey EP, Nagin DS (2001) Assessing the impact of exposure time and incarceration on longitudinal trajectories of criminal offending. J Adolesc Res 16:54–74
Sampson RJ, Laub JH (1993) Crime in the making: pathways and turning points through life. Harvard University Press, Cambridge, MA
Sampson RJ, Laub JH, Wimer C (2006) Does marriage reduce crime? A counter-factual approach to within-individual causal effects. Criminology 44:465–508
Shover N (1996) Great pretenders: pursuits and careers of persistent thieves. Westview Press, Boulder
Steffensmeier D, Ulmer JT (2005) Confessions of a dying thief: understanding criminal careers and criminal enterprise. Transaction Aldine, New Brunswick
Stryker S (1968) Identity salience and role performance: the relevance of symbolic interaction theory for family research. J Marriage Family 4:558–564
Sweelen, Gary, Alex Pifuerot Caupensel Stemberg (2013) Age and the explanation of crime, revisited J Youth Adolescence 42(6):921–938
Thornberry TP (1987) Toward an interactional theory of delinquency. Criminology 25:863–891
Thornberry TP, Krohn MD (2005) Applying interactional theory in the explanation of continuity and change in antisocial behavior. In: Advances in criminological theory, vol 14. Transaction, New Brunswick, pp 183–210
Wright BRE, Caspi A, Moffitt TE, Paternoster R (2004) Does the perceived risk of punishment deter criminally-prone individuals? rational choice, self-control, and crime. J Res Crime Delinq 41(2):180–213

Desistance from Gangs

Katie Feavel[1] and David Pyrooz[2]
[1]School of Criminology and Criminal Justice, Arizona State University, Huntsville, TX, USA
[2]College of Criminal Justice, Sam Houston State University, Phoenix, AZ, USA

Overview

Gangs present serious challenges to social and criminal justice institutions. The proliferation of gangs, coupled with their high involvement in criminal activity, fuels concern among citizens and policymakers alike. Indeed, gangs account for approximately one of five homicides in large US cities, and their members experience homicide victimization at rates 100 times greater than the general public (Decker and Pyrooz 2010). Further, gangs are a driving force for much of the nation’s gun crime as well as other pressing social issues (Howell 2007). The most recent law enforcement figures estimate that there are about 28,100 gangs and 731,000 gang members in the USA, demonstrating the scope of the problem (Egley and Howell 2011). Gangs continue to have a prominent presence in many neighborhoods, schools, communities, and families. As such, it is necessary to further understand both the characteristics and processes surrounding gangs and gang-related activity.

Despite the commonly held perception that gang membership is a lifelong commitment, *most individuals that join gangs also leave gangs*. In fact, involvement in gangs is typically short lived, with the majority of gang members remaining in their gang for 2 years or less (Krohn and Thornberry 2008; Pyrooz et al 2012). Involvement in gangs follows specific patterns: youth join gangs, they persist in gangs and participate in gang activities over some period of time, and then, more often than not, they leave their gang. That said, most research on gang membership has concentrated on either (1) the "ramping up" period of gang membership (i.e., when individuals enter into the gang trajectory) or (2) the persistence period of gang membership, focusing on the (mostly delinquent) activities of gang members. However, gang desistance – the process of exiting the gang – is an equally important aspect of gang membership. Indeed, hastening periods of gang membership will result in reduced rates of criminal offending and serious victimization.

This entry examines the process of *gang desistance*. It begins by answering: what is gang desistance? In doing so, it illustrates and characterizes the concept of gang desistance, drawing from the larger debates and issues experienced in the life-course criminology literature. Next, it discusses the major parameters of gang membership: onset (joining the gang), duration (time spent within the gang), intermittency (the rejoining of gangs), and termination (de-identifying with the gang). It applies key concepts from life-course criminology – trajectories, turning points, and transitions – to the gang context, which helps bring meaning to the gang desistance process. This is followed by examining the primary dimensions of gang leaving, including the *motives* for leaving the gang, the *methods* by which gang members execute leaving, the *abruptness* or rate at which this process transpires, the role of continued social and emotional *ties* that are retained despite having left the gang, and the process of shifting identities. This entry is concluded with a brief sketch of life after the gang, followed by a discussion on the importance of gang desistance research, the implications and benefits of understanding the many facets of gang desistance, and suggestions for future study, based on current gaps in the gang desistance literature.

Fundamentals

Gang desistance refers to the "declining probability of gang membership – the reduction from peak to trivial levels of gang membership" (Pyrooz and Decker 2011, p. 419). The component parts of this definition are derived from the life-course criminological literature (Kazemian 2007; Massoglia 2006), which decomposes desistance into two parts: (1) a reduction in the severity or frequency of participation in criminal activity and (2) an eventual, permanent end, or "true desistance" (Bushway et al. 2001, p. 492). Bushway et al. (2001, p. 500) stated that true desistance occurs when an individual's rate of offending is "indistinguishable from zero" – in other words, when there is no empirical difference from non-offenders.

Gang desistance, however, differs from crime desistance in important ways. First, gang membership is a *state*, while offending is an *event* (Pyrooz et al. 2010). Granted the state of gang membership is comprised of a host of events that display group allegiance, similarly, the confluence criminal events could be conceived as criminal states. Yet, it is clear that gang membership implies at least some degree of connection to a group as opposed to events. To be sure, desistance from gang membership refers to leaving groups, while desistance from crime means disengaging from criminal behaviors. For instance, desisting from the gang does not necessarily mean desisting from crime, nor does the termination of offending indicate a disengagement from the gang. Groups have unique structures, a set of rules, defined roles, and activities; they also have "social cement," a sense of cohesion within that specific group. In this sense, gang desistance is realized as individuals lessen their participation in group activities (e.g., meetings, social events, criminal events) as well as reducing the weight they place on the prominence of the gang and their identification with the gang, which leads to the second key difference between crime and gang desistance.

Operationally, gang membership is determined by self-nomination among study participants. In survey research, this typically takes on the following form: "Are you currently in a gang?" This measure has been found to be a "robust measure capable of distinguishing gang from non-gang youth" (Esbensen et al. 2001, p. 124). Those who have left the gang would have reported being a current gang member at some point in time, but no longer self-report gang membership. This is different from crime desistance because it is not necessary for "ex-criminals" to identify as having been a criminal. In other words, gang leaving is cognitive oriented in that it is determined by individual de-identification, while crime leaving is action oriented in that it is determined by individual behavior. This means that an individual can participate regularly in gang-oriented activities – flashing gang signs, recruiting and initiating new members, and selling drugs – but not be recorded as a gang member.

Criminal justice agencies, alternatively, determine gang membership differently than the research literature described above. Personnel in policing and correctional agencies focus on whether an individual meets a certain list of criteria to be entered into official gang databases. Self-reports are one of many measures that can exist for an individual to be entered into a gang database (Barrows and Huff 2009). In addition to self-report, it is common for agencies to consider gang-related tattoos, association with known gang members, flashing gang signs (observed directly or in pictures), the possession of gang paraphernalia, and information from informants. Those that meet two of the criteria are typically considered associates and those that meet three of the criteria are typically considered gang members. The number of necessary indicators and what attributes are included in the criteria vary between states (Barrows and Huff 2009). Leaving the gang, or getting removed from the gang database, is problematic because more emphasis has been placed on including rather than removing individuals in/from gang databases. As such, the regulation of gang databases and removing ex-gang members from such databases is a problem that criminal justice agencies will have to confront (Jacobs 2009). Regardless of the method of determining gang membership, there is movement both into and out of the gang. This movement is detailed in the following sections.

By defining gang desistance as a reduction in the probability of gang membership, it provides a broader view of the life course of gang membership in adolescence and young adulthood. In this sense, gang membership can be thought in terms of a distribution, with age along the x-axis and the probability of being in a gang along the y-axis. This is detailed in Fig. 1 using (somewhat) hypothetical examples of the intercept and slopes – or the points at which onset and termination occur. Gang desistance takes place from the peak level of involvement, the zenith of the curve around age 15, and slowly lowers to trivial level of membership around the late teens.

Couched with this hypothetical curve are two key parameters of gang membership: onset and termination. *Onset* refers to the first self-reported instance of identification with the gang. *Termination* refers to the first self-reported instance of de-identification with the gang. In relation to Fig. 1, these events boost or reduce the probability of gang membership above or below the 50 % threshold that would consider an individual a gang member.

In relation to life-course theory and research (Sampson and Laub 1993), onset and termination of gang membership take on added significance because they act as life-course *transitions*. Transitions are important events dotted throughout the life course that bring meaning to lives; events such as graduating high school, moving away to college, or having a baby are examples of significant events. Joining and leaving a gang are transitions because they are likely to constitute an important event in the life course. Moreover, these events are often formalized with getting "jumped into" or "blessed out" of the gang. Life events known as turning points, though, are key to understanding larger changes in the life course. Laub et al (2006, p. 314) noted: "turning points may modify trajectories in ways that cannot be predicted from earlier events." These events

Desistance from Gangs, Fig. 1 Key gang membership parameters

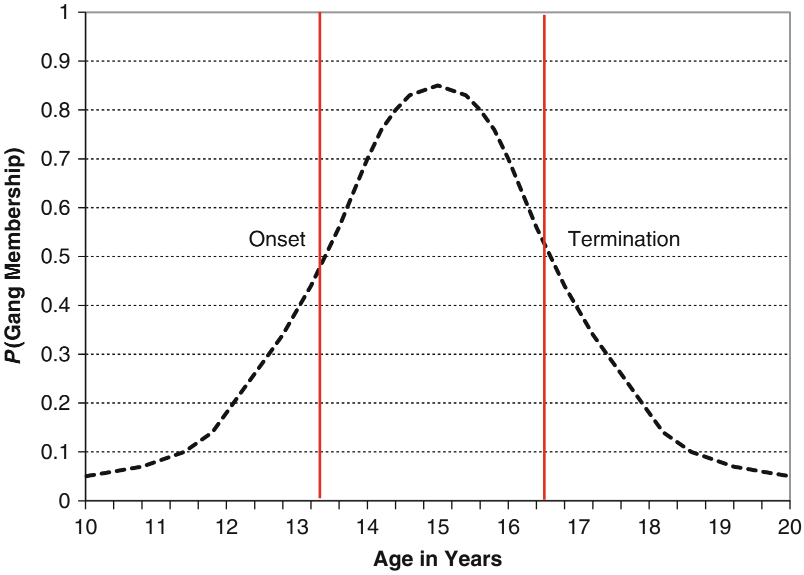

redirect life trajectories in significant ways. Thornberry et al. (2003) argued that gang membership acts as a turning point in the life course. Melde and Esbensen (2011) recently demonstrated this empirically using a sample of youth from multiple cities in the USA. They found that not only did gang joining have a significant impact on criminal offending, but it also influenced the routine activities of gang joining youth. Further, gang membership had a negative impact on the attitudes, emotions, and social bonds of these youth, which led Melde and Esbensen to conclude that gang joining does indeed act as a turning point. These effects, however, were not reversed once one left the gang, so it remains an open question if gang leaving can put lives back on track.

Duration refers to periods between onset and termination. In other words, duration answers the question: how long do gang members remain in gangs? In relation to life-course research, gang membership can be thought of as a trajectory because it is sustained over a period of time. Sampson and Laub (1993, p. 8) defined a trajectory as a "pathway or line of development over the life span." The gang trajectory is marked by formal periods of onset and termination, as noted in Fig. 1, but there are antecedent and ensuing connections to the gang – the latter is an integral dimension of the gang desistance process – which are denoted in areas below the curve yet above the x-axis.

To examine these time trends in the gang trajectory, it is necessary to have longitudinal data that systematically documents patterns of gang membership. Pyrooz et al (2012) reviewed a series of studies that examined the descriptive characteristics of continuity in gang membership, including the Gang Resistance Education and Training, Pittsburgh Youth Study, Rochester Youth Developmental Study, and Seattle Social Developmental Project (Thornberry et al. 2003). All of the aforementioned studies were carried out for at least 4 years and documented patterns of gang membership in Pittsburgh, Rochester, Seattle, and the 5-site G.R.E.A.T. study. In addition, Pyrooz et al. (2012) documented patterns of gang membership using longitudinal data collected in Philadelphia and Phoenix, based on the Pathways to Desistance study. As a whole, most youth did not remain in gangs for significant periods of time. The majority (48–69 %) remained involved with the gang for 1 year or less. Despite this, there was considerable variability around these figures, with 17–48 % reporting 2 years of membership, 6–27 % reporting 3 years of membership, and 3–5 % reporting 4 or more years of duration within the gang.

Previous research, however, only documented the descriptive characteristics of gang membership duration. Pyrooz et al.'s study modeled this relationship to determine what factors impact continuity and change in gang membership. They found males, minorities (blacks and Hispanics), Phoenix gang members, individuals with poor self-control, and those more deeply embedded within their gang persisted over longer time periods. In fact, that one standard deviation higher in gang embeddedness – which includes for involvement, importance, status, position, and activities in the gang – was associated with at least 1 more year of gang membership. These are important factors to consider as it is well established that gang membership is strongly related to violent offending and victimization.

Intermittency refers to the leaving and subsequent rejoining of a gang. Two studies have examined intermittency – Pyrooz et al. (2012) and Thornberry et al. (2003) – finding that of the individuals that reported gang membership for multiple waves, about 57–66 % were intermittent gang members. In other words, these individuals would have joined a gang at least two times and left the gang at least one time. The extent to which this is an artifact of measuring gang membership or a reality in the streets is unknown. However, it is safe to say that intermittency poses hurdles for understanding the key parameters of gang membership and gang desistance research more broadly. The fact that intermittency exists is a testament to the view of gang membership as a dynamic process.

Kazemian (2007) believes it is reasonable to hypothesize that *all* criminal careers experience some level of intermittency across the life course, making the issue even more conceptually applicable to gang desistance. Intermittent participation in gangs may lead to an illusion of desistance, since ex-gang members may rejoin the gang at some later point in time or have varying levels of participation throughout the life course without entirely leaving. Most available data sets lack the ability – for criminal involvement and gang involvement – to rule out later rejoining of the gang; ultimately, measuring desistance based on non-offending simply does not suffice. Studies that have been done on desistance followed up on individuals within a relatively limited time period, thus possibly generating cases of false desistance that may have led to inaccurate conclusions about the desistance process thus far. However, the length of follow-up required to study *true* desistance is problematic, as studies would need to extend longitudinally throughout the entire life course in order to control for intermittency.

Thus far, gang desistance has been defined and identified by the key parameters (onset, termination, duration, intermittency) of gang membership. This section focuses on gang desistance processes, that is, key factors that characterize moving an individual from current to former gang membership status. It points out that this is a dynamic and evolving process that can be characterized by motives for leaving, methods for leaving, abruptness of the departure, the residual social and emotional ties that persist despite having left the gang, and changes in identities (see Pyrooz and Decker 2011). This section refers to these concepts as the dimensions of gang leaving.

Motives for leaving refer to reasons that influenced a gang member to exit the gang, and this dimension can be thought of as the subjective component in this process. Furthermore, motives for desistance can be organized in terms of push and pull factors. Push motives are factors internal to the gang that make persistence in gang membership undesirable. Typical push motives include "getting tired of the gang lifestyle," "wanting to avoid trouble and violence," and "getting tired of always having to watch my back." The factors inspire one to seek out and select into other social arenas. Pull motives are factors external to the gang that steer or "yank" gang members away from the group. Typical pull motives include girlfriends, jobs, or children as the motivation for exiting the gang. Pyrooz and Decker found that about two out of every three gang members left their gang due to factors internal to the gang, or push factors. In other words, gang members didn't leave the gang because of social interventions or jobs, but because of factors internal to the gang itself.

Methods for leaving refer to modes employed in order to exit the gang. In other words, how one was able to exit the gang and whether that exit was met with resistance. It is commonly believed that gang departures are synonymous with "blood in, blood out" – shedding blood to enter and exit the gang. In fact, it is often believed that the only way to leave the gang is by getting murdered or murdering one's mother (Decker and Van Winkle 1996). While that might be an extreme example, it is not uncommon to hear that it is necessary to get "jumped out" of the gang. This process involves the "exiter" to have to endure punches, kicks, and other forms of interpersonal violence for a delimited time period. After these ceremonial actions, the individual is then "free to be on their way" having paid their debt to the gang. Another example of a departure method requires the exiter to commit a crime or a mission against a rival gang member. Many of these methods for leaving are similar to the methods for joining the gang (e.g., getting jumped in, going on a "mission").

Pyrooz and Decker (2011) described the above departures as "hostile" in that an individual was forced to engage in some type of behavior to formalize leaving. Based on interviews with over 80 former gang members in Phoenix, they found that most gang members walk away from the gang without any repercussions. In fact, only 20 % of former gang members reported being met with some type of resistance. Importantly, especially for practitioners, is that the methods for departure are often conditioned by the motive for leaving. Pyrooz and Decker found that of the individuals that left the gang due to pull motives (e.g., pregnant girlfriend, job), *none experienced a hostile method of leaving the gang*. As a whole, hostile methods are more of a myth than a reality, which is not uncommon when studying gangs and gang-related behavior (Howell 2007).

Abruptness of departure refers to the rate at which the gang desistance slope declines. In other words, how long does it take to exit the gang after reaching peak levels of involvement? Further, once an individual decides to leave the gang, how long does it take until that process comes full circle? There are two main categories for describing the abruptness of this process: "knifing off" (Laub and Sampson, 1993) and "drifting off" (Decker and Lauritsen 2002). Knifing off is characterized by suddenly leaving the gang, which can be fueled by a significant event, such as the death of a friend as a result of gang violence. Drifting off takes places gradually as a result of shifting ties and social connections to the gang over a longer period of time. In essence, the desistance slope declines a slower rate for "drifters" and at a faster rate for the "knifers." That said, Decker and Lauritsen (2002) found that it is more common for people to drift away from the gang rather than leaving quickly. Much of this is likely due to the natural progression from adolescence to adulthood where adolescents and young adults began to enter into new social arenas, continuing their education, taking jobs, and entering into relationships. These changes slowly disrupt social networks.

Continued ties persist despite having de-identified as being a gang member (Decker and Lauritsen 2002; Pyrooz et al. 2010). These ties to the gang can be both social and emotional; Decker and Lauritsen (1996) identified this as a "gray area" of leaving the gang. For example, some ex-gang members reported that they continued to hang out with the gang, participating in social activities such as drinking, smoking, and other social arrangement. Other ex-gang members reported that if their former gang was disrespected, they would respond to the disrespect. Similarly, if someone from their former gang network was assaulted, they would retaliate across rival gangs. Pyrooz et al. (2010) found that irrespective of how long an individual had left the gang, continued ties to the gang were associated with increased rates of violent victimization. In other words, the well-established pernicious effects of gangs persisted even if someone had renounced their allegiance.

Identity plays a rather obscure role in the desistance process, as it is not an easily observable dimension. However, gang members do use a number of outward symbols to identify themselves with their gang: hand signals, colors, tattoos, etc. The *name* of the gang also plays an important role in the gang member's identity construct (Bjerregaard 2002). Desisting from a gang often means shedding these symbols to which the individual identifies so closely. Not only do ex-gang members have to go through both cognitive and identity restructuring processes, but they must also navigate through changes in their daily routine activities. As desisters move away from the gang, they establish new social ties in the greater society, and the activities that take up their time tend to shift from gang-related to what would be considered more normative outlets, such as school or sports. Much of the continued tie to the gang pertains to the labels that are attached to gang membership. That is to say, gang membership is in many ways a "master status" that is difficult to shake (Decker and Lauritsen 2002). There are key players involved in assigning that label, including (1) the self, (2) the gang, (3) the neighborhood, (4) the family, and (5) agents of the justice system. Note that throughout the progression from the self to the justice system, there is greater distance in awareness leaving the gang. In other words, while this decision begins with player #1, the remaining players can seriously impact – both positively and negatively – the decision to leave. Even if one has denounced their allegiance to the gang, that decision is formalized by the gang, and one is no longer socially and emotionally tied to the gang; this does not mean that their identity shifts are consistent with the views of the neighborhood or criminal justice agencies.

What about life after the gang? Two seminal contributions to the gang literature – Thrasher (1927) and Decker and Van Winkle (1996) – focus mostly on *life in the gang*, but what about *life after the gang*? Stated another way, what are the life circumstances of gang members 5, 10, or even 20 years after joining a gang? After exiting the gang, do the disruptive activities that characterize gang membership cease and give way to conventional lifestyles putting lives back "on track"? Or, do the disadvantages that accumulate during periods of gang membership have a lasting impact, posing additional difficulties for ex-gang members? Malcolm Klein (1971) noted in his book *Street Gangs and Street Workers* that "[a]lthough the need is great, there has been no truly careful study of gang members as they move on into adult status." Fortunately, in the last 40 years since Klein's critique, several studies have been able to address these issues.

Moore (1978) and Hagedorn (1998) reported on the adult lives of current and former gang members in East Los Angeles and Milwaukee. Both studies followed up on earlier ethnographic research (Hagedorn 1998; Moore 1978) and found that gang members did not fare well in their adult life. Moore found that while some male and female gang members settled down, started families, and pursued conventional employment, this was not the norm. She reported that high rates of early parenthood, unemployment, literacy barriers, and failed relationships made the transition from adolescence to adulthood difficult for many individuals. Hagedorn found that his subjects had dismal high school graduation rates, high rates of unemployment, a reliance on illicit underground markets for income, a dependence on state welfare, and had children at young ages; nearly nine of ten female gang members were mothers in their early twenties. Both researchers relied heavily on Wilson's (1987) hypothesis regarding the impact of deindustrialization and concentrated disadvantage. They argued that their subjects, unlike earlier generations, were shut off from blue-collar job opportunities and thus relied on gang membership, illegitimate labor markets, and entry- and service-level employment for support.

Levitt and Venkatesh (2001) studied gangs operating in Chicago housing projects in 1991. They followed up on the members of their sample, which included gang and non-gang members, in 2000 to examine a host of outcomes and related changes that occurred over that 9-year period. Based on the 2000 interviews, Levitt and Venkatesh examined the effect of gang membership on nine life outcomes, including high school graduation; employment; current incarceration; ever incarceration; annual total, legal, and illegal income if not incarcerated; number of times shot; and housing project resident. The results of their analyses revealed that the negative outcomes tied to gang membership were associated with crime (incarceration, incidence of gunshot victimization, illegal income) rather than other social domains (high school graduate, current employment, public housing residence). This distinction between involvement in prosocial activities and involvement in crime is important for youth policy in general and gang policy in particular. The results suggest that gang membership has little impact on participation in prosocial activities, but long-term negative effects on involvement in crime and victimization.

Thornberry et al. (2003) and Krohn et al. (2011) followed a sample of about 1,000 at-risk youth attending middle schools in Rochester, NY, until their late twenties and early thirties. They argued that gang members would be less successful in accomplishing normative transitions from adolescence to adulthood, such as graduating high school and attending college, than youth who did not join a gang because of their involvement in "precocious" behaviors and risky activities. From their perspective, gang membership can be viewed as cutting off or limiting possibilities for youth, particularly in the key areas of education and employment. The outcomes these studies examined included high school dropout, teenage parenthood, early nest leaving, adult unemployment, welfare dependence, interpersonal problems in the household, cohabitation, adult offending, and adult arrest. Gang membership influenced all of these outcomes positively, increasing the likelihood of their occurrence. These findings, however, were stronger for male gang members and persistent gang members, compared to their female and transient counterparts.

Finally, Pyrooz (2012) explored the educational, economic, and employment trajectories of youth, focusing on the impact that adolescent gang membership had on these trajectories in early adulthood. He found that joining a gang had an immediate impact on educational attainment, particularly for graduating from high school and matriculating to college. While the differences in high school graduation lessened over time, gang joiners were less likely to attend college and earn a 4-year college degree. In summary, the net effect of gang membership on educational attainment was one-half year. That might sound minimal, but that was the difference between graduating from high school or not, as gang joiners completed 11.5 years of education compared to the 12 years completed by their similarly situated counterparts. Educational attainment had consequences for employment, where Pyrooz observed that gang members were jobless for longer periods throughout the year. While gang members, when working, earned as much and worked as many hours as non-gang individuals, over a 6-year period, they made $14,000 less because of their inconsistent patterns of joblessness.

In summary, this research paints a gloomy picture for the adult lives of adolescent gang members. This implies that the well-established consequences of gang membership are not entirely limited to the short term. That is, gang membership has lasting consequences that (1) extend across the life course and (2) cascade into other significant life domains, such as marriage, family, employment, and education. These consequences could impact not only the rates at which gang members desist from the gang but also complicate the motives and methods for desistance as well as the residual ties that persist after leaving.

Future Directions

This entry has explored a series of issues pertaining to leaving gangs. It has provided both conceptual and operational definitions of gang desistance, identifying key issues in the research literature. It explored the parameters of gang membership – onset, termination, duration, and intermittency – and, in doing so, placed these parameters in the broader context of the life course. The key portion of this entry detailed the multiple dimensions of leaving the gang. These dimensions include the motives for leaving, the methods for leaving, the abruptness of desistance patterns, the continued ties after leaving, and the process of identity reconstruction for ex-gang members. Based on what has been discussed above, there are several key conclusions that may help provide directions for future research.

First, leaving the gang is a dynamic and evolving process, one that takes time to realize fully.

This process involves factors that push and pull the gang member away from the nucleus of the gang – van Gemert and Fleisher (2005) refer to it as the "grip of the group" – helping to shred the social and emotional ties to the gang. However, notification of leaving the gang travels across key social players and paths at different rates: from the individual, to the gang, to the neighborhood, to the family, and finally, to the police. As this information transfers across each of these paths, there is the possibility for resistance and pushes and pulls *back into* the gang. For example, if a gang member on the desistance pathway was arrested in suspicion of committing a crime, he or she would be housed in a county jail facility – both jails and prisons are typically hotbeds for gang activity. This gang member would likely be segregated out of the general population into a pod consisting of non-rival gangs and likely fellow gang members, which could result in strengthened ties and increased embeddedness. Further, past antagonisms between former rival gangs may persist, and victimization may push one back into the gang lifestyle. To be sure, it may take several attempts before a gang member is able to break free from the group processes associated with gangs. Identifying intervention points where gang members' ties to their gang are weak, during violent episodes or after someone has "snitched," may be key to influencing decisions to leave, especially since internal factors appear integral to leaving the gang. Understanding these processes – especially with regard to complicating leaving the gang – should be a central task of gang researchers.

Second, leaving the gang might not be associated with benefits or virtues symmetrical to the impact of joining the gang. That is, while joining a gang can be viewed as a turning point in the life course due to wide-ranging negative effects, leaving the gang does not appear to necessarily return lives to the previous "unblemished" state. In this sense, the gang environment "takes on a history of its own" (Laub and Sampson 1993, p. 320) in that it changes lives in important ways. One might expect that while escaping the expectations of the group should decrease rates of criminal offending and victimization, the deficits accumulated during periods of gang membership may overwhelm any gain achieved from leaving. Nevertheless, given the robust overrepresentation of gang members in self-report and officially recorded rates of offending and victimization (Curry et al. 2002; Thornberry et al. 2003), it is probably no coincidence that the age-crime curve and the age-gang membership curve are tapering off at comparable or parallel rates. Despite the possibility that leaving the gang may not be associated with the drop in crime that is equivalent to gang joining, hastening periods of gang membership should remain a high priority given what is known about the serious consequences of gang membership in individual lives and gangs in communities.

Third, "desistance" is a process not unique to the gang context. The gang literature has drawn heavily from life-course criminology to aid in the development of gang desistance research. Yet, as Ebaugh (1988) noted, role exits contain a great deal of continuity across important life states, including retiring or switching careers to changing genders to leaving the religious convent. In other words, whether one is studying differences between exiting conventional and deviant networks or exiting within deviant network types, there are likely to be universal factors that characterize the exit. To be sure, leaving groups and roles is a global phenomenon. Ebaugh refers to four parts in the exit process, including (1) initial doubts, (2) seeking role alternatives, (3) turning points, and (4) establishing an "ex" identity. Understanding variability in these patterns within and across groups would bring a greater understanding to the difficulties in these processes and, ultimately, assist policies to help or hinder such exits. In the gang context, for example, are there organizational structural characteristics – such cohesion, hierarchies, or collective action (Decker and Pyrooz 2011) – that promote longer periods of gang membership? Or, are there characteristics of the gang that meet attempts to leave with serious resistance? Informal social controls in the group context will likely lend considerable insight into exit and desistance processes.

In the attempt to further develop an understanding of gang desistance, this entry has summarized and expanded on the current state of the gang desistance literature. In doing so, it has identified key concepts in the process of desistance and, in some cases, comparing these processes to research on crime desistance in the life course. Leaving the gang is a dynamic process that is not only characterized by motives and methods for leaving but also by broader factors that take place long before and after exiting the gang. This process has not received attention from the research community equal to the process of joining a gang, despite the equally important implications. Future research should further develop, both theoretically and empirically, the concepts and findings discussed in this entry.

Related Entries

► Desistance from Crime
► Gangs and Social Networks

Recommended Reading and References

Barrows J, Huff CR (2009) Gangs and public policy. Criminol Public Policy 8:675–703

Bjerregaard B (2002) Self-definitions of gang membership and involvement in delinquent activities. Youth Soc 34:31–54

Bushway SD, Piquero AR, Broidy LM, Cauffman E, Mazerolle P (2001) An empirical framework for studying desistance as a process. Criminology 39: 491–516

Curry GD, Decker SH, Egley A (2002) Gang involvement and delinquency in a middle school population. Justice Q 19:275–292

Decker SH, Lauritsen J (2002) Leaving the gang. In: Ronald Huff C (ed) Gangs in America, 3rd edn. Sage, Thousand Oaks

Decker SH, Pyrooz DC (2011) Gangs: another form of organized crime? In: Paoli's L (ed) Oxford handbook of organized crime. Oxford University Press, New York

Decker SH, Van Winkle B (1996) Life in the gang: family, friends, and violence. Cambridge University Press, Cambridge, UK

Ebaugh HRF (1988) Becoming an ex: the process of role exit. University of Chicago Press, Chicago

Egley A Jr, Howell JC (2011) Highlights of the 2009 National Youth Gang Survey. U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention, Washington, DC

Esbensen FA, Winfree LT, He N, Taylor TJ (2001) Youth gangs and definitional issues: when is a gang a gang, and why does it matter? Crime Delinq 47:105–130

Hagedorn JM (1998) People and folks: gangs, crime and the underclass in a rustbelt city, 2nd edn. Lakeview Press, Chicago

Howell JC (2007) Menacing or mimicking? Realities of youth gangs. Juv Fam Court J 58(2):39–50

Jacobs JB (2009) Gang databases: context and questions. Criminol Public Policy 8:705–709

Kazemian L (2007) Desistance from crime: theoretical, empirical, methodological, and policy considerations. J Contemp Crim Justice 23:5–27

Klein MW (1971) Street gangs and street workers. Prentice-Hall, Englewood Cliffs

Krohn MD, Thornberry TP (2008) Longitudinal perspectives on adolescent street gangs. In: Liberman AM (ed) The long view of crime: a synthesis of longitudinal research. National Institute Of Justice, Washington, DC

Krohn MD, Scmidt NM, Lizotte AJ, Baldwin JM (2011) The impact of multiple marginality on gang members and delinquent behavior for Hispanic, African American, and White Male Adolescents. J Contemp Crim Justice 27:18–42

Laub JH, Sampson RJ, Sweeten G (2006) Assessing Sampson and Laub's life-course theory of crime. In: Cullen FT, Wright JP, Blevins K (eds) Taking stock: the status of criminological theory, vol 15, Advances in criminological theory. Transaction, New Brunswick

Levitt SD, Venkatesh SA (2001) Growing up in the projects: the economic lives of men who came of age in Chicago public housing. Am Econ Rev 91:79–84

Massoglia M (2006) Desistance or displacement? The changing patterns of offending from adolescence to young adulthood. J Quant Criminol 22:215–239

Melde C, Esbensen FA (2011) Gang membership as a turning point in the life course. Criminology 49:513–552

Moore JW (1978) Homeboys – gangs, drugs, and prison in the barrios of Los Angeles (CA). Temple University Press, Pennsylvania

Pyrooz DC (2012) The non-criminal consequences of gang membership: impacts on education and employment in the life-course. Doctoral dissertation, School of Criminology and Criminal Justice, Arizona State University

Pyrooz DC, Decker SH (2011) Motives and methods for leaving the gang: understanding the process of gang desistance. J Crim Justice 39:417–425

Pyrooz DC, Decker SH, Webb VJ (2010) The ties that bind: desistance from gangs. Crime Delinq. Online First. doi:10.117/0011128710372191

Pyrooz DC, Sweeten G, Piquero AR (2012) Continuity and change in gang membership and gang embeddedness. J Res Crime Delinq. Online First. doi:10.1177/0022427811434830

Sampson RJ, Laub JH (1993) Crime in the making: pathways and turning points through life. Harvard University Press, Cambridge

Thornberry T, Krohn MD, Lizotte AJ, Smith CA, Tobin K (2003) Gangs and delinquency in developmental perspective. Cambridge University Press, Cambridge

Thrasher FM (1927) The gang: a study of 1,313 gangs in Chicago. University of Chicago Press, Chicago

van Gemert F, Fleisher M (2005) In the grip of the group. In: Decker SH, Weerman FM (eds) European street gangs and troublesome youth groups. AltaMira, Lanham, pp 11–30

Wilson WJ (1987) The truly disadvantaged: the inner city, the underclass, and public policy. University of Chicago Press, Chicago

Desistors

► Desistance from Crime

Detecting Deception with fMRI

George T. Monteleone
Center for Cognitive and Social Neuroscience,
The University of Chicago, Chicago, IL, USA

Synonyms

Detection of deception; Lie detection

Overview

The application of contemporary neuroimaging technologies to the detection of deception has garnered popular attention in recent years. Members of the scientific community have proposed that functional magnetic resonance imaging (fMRI) can be employed as signal detectors to predict behavior and cognitive states. This endeavor is often discussed with a tone of hopeful optimism, but it must be considered with adequate scientific rigor, proper understanding of the limitations of the tool being used, and good social responsibility. In the formative years of the field of criminology in the late nineteenth century, attempts to unify imaging techniques and physiological data for the purpose of human classification yielded questionable results and undesirable social influences. It is necessary not to repeat past mistakes in the excitement over a novel technique.

Imaging, Data, and Criminal Classification

The intersection of imaging, behavioral science, classification, and criminology has origins that date back to the late nineteenth century, when Sir Francis Galton described methods of composite photographic portraiture, a technical innovation at the forefront of imaging technology in its time (Galton 1879). Galton hoped to apply this tool to "elicit the principal criminal types by methods of optical superimposition of [. . .] portraits," in other words, to identify through composite imaging the physical appearance of a typical criminal. These experiments, though embraced with optimism by the scientific community, proved to be a failure. Galton ultimately observed that ". . .the features of the composites are much better looking that those of the components. The special villainous irregularities in the latter have disappeared, and the common humanity that underlies them has prevailed." While Galton's technical developments are impressive in their detail and rigor, his broader perspective on the applications of his techniques was flawed. Nearly a century later, scientist and historian Stephen Jay Gould would appropriately label Galton an "apostle of quantification," suggesting that his tireless measurements of human physical features in the attempt to achieve classification were simply too narrow in scope to serve as reliable models for the natural world (Gould 1981). Companions to Galton's experiments with composite portraiture were the notions of physiognomy and eugenics – ideas that suggested that human characteristics were biologically determined and socially controllable. These hypotheses not only failed to achieve any scientific credence in the twentieth century but also demonstrated the socially dangers political malleability of unsubstantial scientific ideas in later decades through their misappropriation.

The well-known nineteenth-century criminologist Cesare Lombroso was also no stranger to physiognomy and eugenics (Lombroso 1876–1897). A proponent of biological determinism and physiological classification, Lombroso's early scientific hypotheses extended from phrenology, a practice (now widely discredited) that attempted to predict and classify human characteristics through measurements of the shape of the skull (Gray 2004). This putative science is a distinct example of a sound fundamental theory (an existing relationship between the brain, personality, emotion, and behavior) coupled with a thoroughly unsound notion of how it could be measured and applied. Instead of being treated as a scientifically falsifiable hypothesis, phrenology was practiced under the assumption of its validity. Lombroso's foundational text also addressed numerous other external features that could be observed, quantified, and described in relation to the criminal type, such as tattoo markings on the body, linguistic and emotional behaviors, and nuances of handwriting. It is no surprise, given his interest in the indexical tabulation of observable features, that Lombroso dedicated a chapter of the 1884 edition of his text to "Photographs of Born Criminals," proposing that images could be employed to profile a criminal type.

Late nineteenth-century science was no doubt catalyzed by rapidly evolving complementary technological innovations such as advances in photographic techniques and printing that had, to that point, undergone several decades of development and refinement. The period was characterized by great eclecticism and imagination, though these positivist pursuits were not always tempered by healthy social skepticism; Galton and Lombroso were relatively unconfined to particular specializations, and their research practices branched into some peculiar explorations. Indeed, Lombroso's final project (published posthumously) was a positivist inquiry into hypnotism and spiritual phenomena, complete with photographic "evidence" that photographic technology enabled the visualization of ghosts, phantasms, apparitions, and ectoplasms (Violi 2004). Lombroso had been wary of such investigations earlier in his career, changing his position in the early 1890s after taking to attending séances. His reputation as a scientist ultimately overcame any skepticism about his association with the field of paranormal investigation, his authority instead lending credence to the field instead of being debased by it. Though not completely estranged from what was scientifically acceptable at the time, paranormal studies, like physiognomy and eugenics, have remained scientifically unsubstantiated over a century hence (Porter 2003).

Galton and Lombroso's particular attempts to unify physiognomy, photography, statistics, and criminology, though ambitious, were alloyed by bunk assumptions and demonstrably misguided, regardless of any lasting contributions they may have contributed in other fields. Galton, particularly, was a brilliant technical innovator, but failed to recognize that a useful tool is not tantamount to a useful model of human nature. Both eugenics and physiognomy, despite a general enthusiasm about their potential among the scientific communities of their times, have been demonstrated to be at best untenable, and at worse socially dangerous as evidenced by their appropriation in the false sciences of the Nazi party as propaganda to justify monstrous ideologies and acts of mass murder (Gray 2004). Such unsettling potential branches of the ideals of scientific positivism underscore the crucial importance of responsible social vision, good sense, and integrity in any applied science, despite any promises it may appear to offer for the immediate future.

As easy as it may seem to dismiss bad ideas from century-old science as antiquated and irrelevant in the present time, it is important not to fall into similarly narrow interpretations clouded by enthusiasm about current technological prospects. Functional magnetic resonance imaging (fMRI), the most contemporary imaging technology employed at the intersection of physics, physiology, psychology, and neuroscience, offers great promise to future developments in scientific understanding. However, improper application and misunderstanding of this tool undermine its potential; its technical complexity can grant it a false credibility that risks to be passed off to the public under the authoritative moniker of science.

Clarifying a Complex Tool

Commercial firms offering for-profit fMRI lie detection services are optimistic about its promise and technical merits. fMRI has been described in promotional material as a "direct measure of truth verification," an "unbiased method for the detection of deception and other information stored in the brain," a means to investigate "the science behind the truth" and "provide independent, scientific validation that someone is telling the truth." While the marketing decision to shift language from "lie detection" to "the science behind the truth" may be a clever one, it distracts from the important fact that a technology or procedure, no matter how sophisticated, cannot justify a model of nature that is fundamentally inaccurate. MRI is, simply, a tool, like a camera or a microscope, which can offer insights into the workings of the mind and brain when properly applied and reasonably interpreted in conjunction with other forms of inquiry. It is important to distinguish the use of fMRI to detect a signal, as in the case of deception detection, from the practice of scientific research, which attempts to test and refine abstract models of nature by repeatedly testing falsifiable hypotheses; using fMRI for signal detection is, rather, a kind of engineering, an attempt to develop a means to perform a desired function. An elephant in the room in fMRI diagnostics is a simple question: Is the putative signal, in fact, what it is assumed to be? Another elephant, perhaps the next room over, might follow with: Even if so, and even if it can be detected, is the proposed application an appropriate use of this resource? In any case, a clearer understanding of fMRI technology is in order in order to consider its potential, its limitations, and how it ought best (or ought not) be used.

In the interest of demystifying a complex and sophisticated technology, it can be considered in more familiar terms. An MRI brain image is similar in ways to a commonplace photograph. Indeed, some MRI technicians, when communicating to lay participants, refer to MRI image acquisition with the familiar language of "taking pictures." This metaphor is apt; MRI is, much like a camera, a means to record an index of a space within a given field of view (this technical term, field of view, is used in both photography and MRI imaging). In the case of photography, the field of view is determined by optics and perspective; an image of a three-dimensional object is projected onto a two-dimensional plane and recorded through some technical means. The photographic image is an index of a visual space, a record of the phenomenon of light, though an incomplete or slightly distorted version of it due to its optical projection. A two-dimensional photograph can never, despite any level of technical or optical sophistication, fully represent the three-dimensional space from which it originated (Arnheim 1954/1974). Furthermore, as simple as it may seem, it is important to distinguish the constitution of the image from its content: the image is often described as "the reality," but it is, in fact, an image and an index, not the reality itself. This consideration is important in relationship to the claim that an imaging technology self-validates by bringing the viewer "closer to the source." The translation of this problem to fMRI is elementary, but easy to overlook. Technology may afford us the ability to represent ("visualize") what was not possible in the past, but it is important not to confuse a model with what it represents. Scientific models, and even scientific images, will always contain some level abstraction, and in MRI this is absolutely the case. Images analyzed in fMRI studies are parts of a model: a complex, multilayered index of brain activity, and throughout the analysis process, this index becomes ever more abstracted. This abstraction is extremely useful for testing and interpreting results and formulating theories according to the most reasonable interpretations, but this image evidence, by the very structure of the scientific method, is not simplistic, hard, nor immutable. It is, at best, a means by which to formulate theories of what is likely to be happening in a system that is not directly observable.

Some basic misconceptions that surface in the popular discussion of fMRI can be done away with through a better basic understanding of the technology. For one, it does not index "blood flow," as many descriptions claim (including those published in certain commercial fMRI detection promotional material). Functional MRI signals are indices of shifts in local blood oxygenation levels, which in turn are interpreted as indices of neural activity due to the consumption of energy (via oxygen) during neural activity. The particular relationship between the immediate shift in local blood oxygenation levels and neural activity is not very well understood, but at the very least, the measure should be identified the same way it is by those who study it: the BOLD (blood oxygen level dependent) signal. A more specific model for the brain activity presumed to underlie the BOLD signal is at the cellular level, where neurons (brain cells) consisting of an axon tail extending from the cell body send electrical potentials (propagating voltages) along the axonal membrane. At the terminus of the axon, the signals influence neurotransmitter releases across intercellular space to receptors on the dendrites of other cells (receptors), leading to the buildup and propagation of subsequent electrical pulses by the recipient cell. This activity, compounded on the order of millions in cellular groups (ganglia), requires energy for fundamental physiological processes: some that are typical to cellular operation – such as metabolism – as well as some functions specialized to the neuron, such as the operation of "pumps" that move positive and negative charges across the cell membranes to propagate electrical signals.

The fundamental physics of MRI involve the scanner's sensitivity to differential alignment of atomic nuclei, which are manipulated by pulsing radiofrequency signals through an object in extremely strong magnetic field. Extending the photographic analogy, while a photograph is a record of light (photo: light; graph: drawing) and an MRI image is like a kind of "atomograph," a means of "drawing" an image of the nuclei of atoms in the magnetic field. Indices of blood oxygenation levels collected in fMRI are indices of shifts in the shape of hemoglobin, a macromolecule in the blood that carries oxygen to supply cells with energy. Hemoglobin takes different forms depending on whether or not it is carrying oxygen (it is either "oxygenated" or "deoxygenated"), and the MRI scanner, with its atomic recording properties, can be tuned to detect the differences in the molecule in these respective states. In order to create three-dimensional images, the scanner collects a sequence of two-dimensional slice images in sequence that are subsequently stacked into a three-dimensional volume, wherein the two-dimensional pixels constituting each slice become three-dimensional voxels according to the spacing parameters of the slices. The "functional" term (the "f" in fMRI) involves the incorporation of a time factor – the scanner collects a sequence of brain volume images, which are later reconstructed as a time series of three-dimensional volumes, just as a film is simply a sequence of still photographs. Functional MRI studies typically use two different types of image acquisition. A high-resolution anatomical image is acquired, which is of relatively precise detail, with cubes of around 1 mm per side, but requires several minutes to acquire, like a long exposure photograph. The functional images are acquired rapidly, with an entire brain volume (usually 30–40 slices) collected in 2–3 s. Because of the need for rapid acquisition, the images are much coarser, with the brain volume segmented into cubes around 3–4 mm per side. This trade-off of image quality for time analogous to photography and motion pictures; anyone who has operated digital still and video cameras has probably noticed the difference in image quality between the two formats, with reductions necessary in order to stream video through a limited bandwidth.

Though the BOLD signal is the starting point in fMRI analysis, these signals undergo a complex series of analytical processes and, ultimately, are translated to three-dimensional statistical maps distributed through a model of the brain. Data processing is typically described in two general stages: preprocessing and analysis. Preprocessing is a set of computations that are performed on the data time series to correct for temporal and spatial errors that might have occurred during data collection, improve signal-to-noise ratio in the BOLD signal, and to spatially normalize ("warp") data to a physiological template from which specific neuroanatomical regions can be estimated. Such steps may include (but not be limited to) slice-time correction to account for offsets in the serial acquisition of the slices comprising each three-dimensional image brain image in each temporal sample (think of it as a "frame" in a three-dimensional "brain movie"); co-registration of each three-dimensional map in the time series (an alignment of each "frame" of the "brain movie"); temporal filtering or linear de-trending of the time series (removal of "drift" artifacts that are generated by the scanner but not a product of any physiological activity); application of a Gaussian spatial smoothing to increase signal-to-noise ratio in regions of the brain presumed to be active at a scale larger than a single voxel; spatial normalization (a computational warping) to a documented anatomical coordinate system such as the Talairach-Tournoux or Montreal Neurological Institute template, resampling each voxel's native dimension to an isometric space of voxels of equal dimension; and finally any normalization of the signal to a standard measurement scale such as percent signal change. After preprocessing is completed, a statistical map of response to experimental conditions is obtained, typically by applying the general linear model (GLM) to estimate parameters given the particular design model of the stimulus timing presentation. The GLM is the application of a parametric model which accounts for the different factors in the study design by modeling a hypothesized response (referred to as the hemodynamic response or impulse response function) which is a set of statistical values determined by the fit between empirical data and the model. This series of fit values (beta values) is then mapped into the voxel space of the brain image, providing a statistical parametric map of the activity in each region. These beta values are then entered into subsequent calculations in order to test various contrasts between conditions. Subsequent statistical tests are carried out by averaging data using regions spatially defined according to a priori criteria, or tests are carried out independently within each voxel in the cortex, the results of which must be statistically controlled for the tens of thousands of repeated measures to obtain a corrected probability of error in the result (Nichols 2012).

As is the case in contemporary science, fMRI research involves the compounding of layers upon layers of theoretical models and can be misleading to characterize any of this as direct measures of reality. Indeed, in science we can hope to identify what is most likely and expect that we might always be wrong. A margin for error is always part of any quantitative model. It is not uncommon for a thorough research group to require several months to design and administer an fMRI study, refine the analyses, and develop a reasonable interpretation of the results prior submitting them to a peer review panel for publication, after which of course they would be open to discussion in the broader scientific community.

The administration of a complete fMRI study is by no means as simple as administering a traditional polygraph test, and even the fundamental task of data collection is wrought with caveats. Slight head movements can cause degradation of the signal or compromise the integrity of the spatial mapping of BOLD signal. It is not unusual for respondents with head motion greater than 8 mm (approximately one-third of an inch) in any direction to be discarded from a study due to loss of data integrity, despite the considerable cost to the researcher, a challenge heightened by the confined space of the magnet bore in which respondents must remain motionless during scanning. Thus, it is important to maximize the relative comfort of respondents in an otherwise uncomfortable circumstance; claustrophobia is a disqualifying factor in the selection of participants. Hearing protection should be worn, as acoustic noise can reach 120 dB(A), a sound pressure level adequate to cause hearing damage. Nevertheless, despite the discomfort of complete immobility in an enclosed space and extreme levels of auditory noise, participants can fall asleep during scanning, yielding useless (but expensive) data. The magnet, typically 3 Tesla in strength, emits a powerful magnetic field. As such, any ferromagnetic material in proximity to the magnet is extremely dangerous, both to the individual being scanned and the administrators of the scan. At risk of serious injury and/or death, it is impossible to scan any individual with metal implants, microscopic ferromagnetic remnants that might be present in the skin or mucous membranes (a concern for anyone who has spent appreciable time in metalworking or industrial settings), or even large tattoos which can contain trace amounts of iron oxides which can cause serious dermal burns (followers of Lombroso who would wish to use fMRI technology for criminal classification could only hope his suggested relationship between body markings and criminal behavior is specious, as his tattooed criminals would not qualify for interrogation.) Other risks are the generation of body heat caused by electromagnetic radiofrequency excitation and the risk of toxic cryogenic supercooling materials that could be released in the event of a technical malfunction of the magnet. Typically, fMRI magnets are maintained by institutions (such as hospitals) that keep a support staff of physicists, radiologists, and skilled technicians, supporting costs by renting scanning time to researchers and physicians with a considerable maintenance budget. A dedicated in-house fMRI lie detector would be wasteful and impractical, as the best model for such a resource-heavy device is shared use by a broad set of researchers and physicians with adequate budgets to support its administration costs.

Signal and Specificity

The complexity of collecting and analyzing functional MRI data raises a question: Why would such a technology be suited for the detection of deception? While brain imaging may appear to be a direct measure of cognitive activity, it would be better considered as an index of an extremely complex neurophysiological system from which cognition originates. Despite the expense and complexity of the technology and its power for making inferences about physiological activity from studies of groups of individuals, it has yet to be demonstrated as a reliable signal detector at the level of individuals, and its use in individual diagnosis still faces numerous challenges.

Any signal detection model has three possible outcomes: a *Hit*, in which the signal is present (the respondent is lying) and is properly identified (the detector registers a lie); a *Miss*, in which the signal (lie) is present but is not identified (the detector registers as “not-lie”); and a *False Alarm*, in which the signal is not present (the respondent is telling the truth) but is improperly identified (the detector registers a lie). In the extrapolation of such a model to the interrogation of suspects, the imperative is clear that such a signal detector would have to be perfect to avoid wrongful convictions (False Alarms).

Furthermore, a reliable signal detector needs to satisfy four criteria. First, it must be sensitive to the signal that it attempts to detect. Second, it must be specific enough to detect the signal that it purports to detect and should not detect a more abstract corollary of that signal. Third, it must be generalizable across individuals and contexts – it must not be a function of a given circumstance or sample of the population but must work in a variety of scenarios. Fourth, it must be robust enough to resist attempts to intervene on the signal – in the case of lie detection, this could be noise introduced into the signal or attempts by the participant to “deceive” the signal by enacting countermeasures, be they cognitive or physiological.[9]

It is worthwhile to consider some examples from the extant research literature on functional imaging and classification. Numerous studies have endeavored to identify the network of neural activity involved in truth telling and lying, and there is converging evidence that such a network can be identified as the likely set of neural components active during the cognitive state of lying. Research suggests that a network involving the ventrolateral prefrontal cortex, dorsolateral prefrontal cortex, dorsomedial prefrontal cortex, and anterior cingulated cortex is active during intentional deception (Phan et al. 2005). These regions, or a subset of them, have been discussed as possible target regions from which individual lies can be identified (Abe et al. 2007; Mohamed et al. 2006). At this point, it is necessary to clarify the practice of group classification from the individual. Of the voluminous body of functional

MRI research, the vast majority of studies involve group-level studies to identify brain regions active in a given task. The aforementioned results provide evidence that these regions, identified at the group-level, may be active during lying, but do not attempt classification at the individual level. Abe et al. (2007) did report lie classification at a 92 % success rate, but in fact used the traditional physiological polygraph to obtain this result (with a small sample of only 6 subjects), and simply reported aggregate results of the accompanying fMRI study without classification, underscoring the challenge fMRI faces to improve on such results and satisfy its promise of meeting the necessary standards of reliability, generalizability, and robustness.

Individual-level classification has been carried out by several researchers using sophisticated within-subjects paradigms that creatively extend the traditional group-level analysis methods employed in most functional MRI research. Kozel and colleagues reported an individual classification success rate of 90–93 %, and rates between 71 % and 85 % have been reported in other research (Langleben et al. 2005; Monteleone et al. 2009). These results suggest that fMRI is indeed well better than chance but remains imperfect. Furthermore, it is only with the utmost rigor that the network of regions employed in classification be distinguished from its role in other processes; the set of regions used to identify lies are just as likely active in a variety of other social and cognitive processes.

At the individual level, even if fMRI signal detectors are far from perfect, they offer the promise to satisfy the criterion of sensitivity. This, however, could also be said of the traditional polygraph operated by an expert, as classification rates above 90 % have also been reported using this method (Abe et al. 2007). The inability of the traditional polygraph to meet the requirements of generalizability and robustness is clear (Cacioppo 2003), and fMRI methods, despite their sophistication, still face these same challenges. According to the best of our limited understanding of the nature of cortical networks and function, fMRI could just as well prove a failure in terms of specificity. While the aggregate analysis of fMRI data is scientifically sound as a way to consider evidence that could converge on a better understanding of neural function through repeated studies and aggregates of results from research from other levels of modeling neural function (such as cellular-level recordings), the appropriation of specific task-related aggregate brain images to individual classification could risk to be no more justified than Galton's attempts to use aggregate photography to classify criminals. Over a century after Galton's failed imaging classification experiments, a surprisingly similar problem appears: the classifier must consider an aggregate image and make a reasonable conclusion about what it is and what it is not. One must consider that individual variability may not be accurately predictable in every case from an aggregate.

The difficulty in reliably identifying individuals from a putative network suggests that the network is not, in fact, specific to delivering false information, and alternative networks may be active during the presumed lie state. This is sensible, as the construct of "lying" is, in fact, a nonspecific term itself. Though adequate to describe a particular behavior, the colloquial idea of a lie must be reconsidered according to cognitive neuroscience models' neural and mental function. For example, lying could be considered a complex cognitive state composed of processes including (but not limited to) language, memory, and higher-order reasoning. Furthermore, the network of regions described by researchers as the "lie network" is by no means specific; these regions have been demonstrated to be active in a variety of cognitive tasks. For example, a useful counterpoint to any enthusiasm over the specificity of the prefrontal cortex loses momentum when we consider the other possibilities of the complex range of functions the prefrontal cortex plays in neural and mental function, from well-documented cases of its role in emotional processing (Damasio, 1994) to its function even in the "blank-slate" default network of brain function that is active when no cognitive task is at hand (Raichle and Snyder 2007).

The specificity problem is perhaps the most often overlooked amidst the enthusiasm about diagnostic fMRI. Given the complexity of brain activity, the limited understanding of it, and the index-of-an-index structure of the fMRI model, it would be foolhardy to presume that any signal collected in the brain is, a priori, specific to a particular behavior taking place when the signal was recorded. The theory that cortical function is not fixed, but instead is dynamic, modular, and adaptable, is not a new one. Over a half-century ago, neurologist Karl Lashley dedicated decades of research to the search for "the engram," a cortical trace of memory that could be defined in rodents. After years of carefully controlled experiments failed to detect a locus for a particular memory, even through the methodical excision of very part of the rodent cortex, he concluded that "there is no demonstrable localization of a memory trace" (Lashley 1950). Since Lashley's work, cognitive neuroscientists have embraced distributed processing models of brain activity, which propose that the brain may best be understood as indices of subcomponents of cognitive processes, instead of representing specific behaviors, thoughts, or cognitive states (Rumelhart et al. 1986). While regions participating in a network may be more reliable in their specificity of certain components, it could be reductive, limiting, or simply inaccurate to claim that a given region is tied to a particular module of a cognitive process.

Indeed, cognitive neuroscientists face the daunting challenge of synthesizing models of neural function from a variety of levels of inquiry: molecular, unicellular, multicellular, physiological, electrical, and computational, in the attempt to reconcile some sense of understanding of how a mass of trillions of electrically charged biological units give rise to the phenomena of cognition, emotion, and behavior. While estimates of the number of cells in the human brain are on the order of a hundred billion, the consideration of the interconnectivity of transmitted signals at the subcellular level is much larger, on the order of the hundred trillions. Given that these connections at the subcellular level are hypothesized to code neural information, the spatial precision fMRI signal, on the order of the size of a pea, is considerably coarse in comparison to the microscopic scale of individual neural connections. In light of this, functional MRI is but one imperfect tool in an extremely complex puzzle pursuing converging lines of evidence to best suggest the relationship between the physiology of the brain and the abstract functions of the mind. To reduce this endeavor to the simple detection of deception would be not only reductive but would propagate an oversimplification that would be counterproductive to scientific understanding.

Broader Cognitive and Social Caveats

Given the questionable scientific substance of lie detection, a reliable lie detector remains an idea of fiction. Presuming, however, that such a detector *could* even be devised lends little support for whether it would be prudent to implement it. Lombroso's ideas of biological determinism are reflected in other outmoded ideas from early twentieth-century scientific theories such as behaviorism, which still hold influence on some models of personality and social behavior. There exists a risk to overestimate the "nature" side of a dichotomous nature/nurture model for human behavior. It may be tempting to presume that physiology and the brain should be compartmentalized simply as a phenomenon of nature, but the brain is not a static entity. As a functional organ, a system of complex, time-dependent network signals, the brain appears to be anything but a fixed, deterministic system. Though brain *structures* may be essentially fixed, that does not mean brain *function* cannot be molded and modulated through experience and subject social influences. Furthermore, even consciousness and memory themselves are unreliable, as brilliantly demonstrated in the work of Elizabeth Loftus (Loftus 2003). Even if a sophisticated signal detector were to exist to trace the neural register of conscious memory, what a subject deems to be "the truth" cannot always be trusted.

Returning to the subject of fiction, literature has long warned of the dangerous malleability of "the truth" in a technocracy. This idea is brilliantly described in the dystrophic vision of George Orwell's classic, *Nineteen Eighty-Four*.

The Orwellian "Ministry of Truth" is, of course, an institution for the ironic promulgation of blatant falsehoods. Orwell, who was born at the end of Galton and Lombroso's generation and wrote his most visionary work during the turbulence of World War II (a decade replete with Nazi pseudoscience), was prescient in his consideration of the misappropriation of science and its potential for social injustice. But few authors could be considered more apropos to the particular discussion of technological classification than Philip K. Dick. Dick imagines the horrific social potential of misappropriated technological classification in several works: downloaded brain data are used to subject individuals to life-or-death trials (Dick 1966); electrophysiological prints are employed as classifiers in an oppressive eugenic society where lack of identity results in a lifetime of forced labor (Dick 1974); individuals are incarcerated for "precrimes" predicted by mysterious employees of a deterministic police state (Dick 1956); and a portable machine containing a battery of physiological measures is employed to test for humanity – failure results in classification as a rogue android, resulting in termination (Dick 1968). In any of the imaginary dystopias, an fMRI lie detector would seem, disturbingly, to be a perfect fit.

In his 1953 novella, *The Variable Man*, Dick presents an astute and more optimistic synecdoche of applied science in society: set in deterministic technocracy where wars are waged based on the quantitative predictions of a computer algorithm, the story describes a scenario eerily similar to the institutional application of a signal detection classification system. The titular "variable man," a nineteenth-century tinker who has been transported to the future through an unexpected accident, serves as an unpredictable parameter whose unique circumstance causes him to disrupt the computer's ability to make reliable classifications. Dick proposes that even the most sophisticated device may fail to predict with absolute accuracy the breadth of individual variability, as even a single outlier can undermine a system that is presumed – and required – to function perfectly. This simple yet viable suggestion, so easily overlooked by eager Galtonian and Lombrosian hypotheses of classification, is a crucial consideration in light of the broad variability observed in the human sciences – a variability that ought by all means be reflected in individual differences in brain activity. Dick's accompanying vision of the future, not unlike our own present, is wrought with hyper-specialization that begets a dangerous technological dependency: individuals in different areas of even the same general field are completely unable to understand the details of one another's professional expertise, and consequently, no individual is fully able to comprehend the workings of the very tools upon which they rely to make life-altering decisions. The social application of the thoroughly complex technology of fMRI could easily create a similar dilemma; indeed, even the relatively simple polygraph has been widely misapplied and misinterpreted during its controversial tenure as a putative classification tool (Cacioppo 2004). Dick's variable man possesses a unique savant-like intuition for technological devices amidst the otherwise narrow-minded, hyper-specialized future society in which he finds himself. As such, he is contracted by the government, which is engaged in a prolonged intergalactic conflict, to build a powerful weapon of war. But along with his technical genius, he possesses a simple but important perspective on not just how to innovate technically but how to best deploy his inventions. In lieu of the weapon which he is contracted to build, he creates a device with which he alters the past, circumventing the war altogether. It is such broad thinking that an institution of power, which has the resources and authority to steer the application of our technologies and resources, would do well to adopt.

The demands of functional MRI are extensive, and its reduction to a simple "lie detector" would be a narrow application of its capabilities, further complicated by the risk of misappropriation in light of achieving a particular goal. Furthermore, if fMRI signal detection could be refined to the point of reliable application, it would be more valuable applied as a means to diagnose and treat medical and psychological disorders. Errors in such a diagnostic circumstance, though hopefully minimized, would at worst yield a null result. An fMRI lie detector, by contrast, through its sheer complexity, expense, and technological impressiveness, is at risk to be easily misinterpreted, oversimplified, and abused as an invalid means of social control.

Related Entries

- ▶ Crime Science
- ▶ Eysenck's Personality Theory
- ▶ False Memories
- ▶ History of the Statistics of Crime and Criminal Justice
- ▶ Identification Technologies in Policing and Proof
- ▶ Marxist Criminology
- ▶ Synthesizing Biological and Social Theorizing

Recommended Reading and References

Abe N, Suzuki M, Mori E, Itoh M, Fujii T (2007) Deceiving others: distinct neural responses of the prefrontal cortex and amygdala in simple fabrication and deception with social interactions. J Cogn Neurosci 19(2):287–295

Arnheim R (1954/1974) Art and visual perception: a psychology of the creative eye. University of California Press, Berkeley/Los Angeles

Cacioppo JT (2003) Committee to review the scientific evidence on the polygraph. The polygraph and lie detection. National Academy Press, Washington, DC

Damasio A (1995) Descartes' error: emotion, reason, and the human brain. Harper Perennial, New York

Dick PK (1966) The unteleported man/Lies, Inc. Ace Books, New York

Dick PK (1968) Do androids dream of electric sheep? Doubleday, Garden City

Dick PK (1974) Flow my tears, the Policeman said. Doubleday, Garden City

Dick PK (1953) The variable man. Space Science Fiction 2(2)

Dick PK (1956) The minority report. Fantastic Universe

Galton F (1879) Composite portraits. J Anthropol Inst Great Brit Ireland 8:132–144

Gould SJ (1981) The mismeasure of man. W.W. Norton, New York

Gray RT (2004) About face: German physiognomic thought from Lavater to Auschwitz. Wayne State University Press, Detroit

Hakun JG, Ruparel K, Seelig D, Busch E, Loughead JW, Gur RC, Langleben DD (2009) Towards clinical trials of lie detection with fMRI. Soc Neurosci 4(6): 518–527

Kozel FA, Johnson KA, Mu Q, Grenesko EL, Laken SJ, George MS (2005) Detection deception using functional magnetic resonance imaging. Biol Psychol 58:605–613

Langleben DD, Loughead JW, Bilker WB, Ruparel K, Childress AR, Busch SI, Gur RC (2005) Telling truth from lie in individual subjects with fast event-related fMRI. Hum Brain Mapp 26(4):262–272

Lashley KS (1950) In search of the engram. Soc Exp Biol Symp 4:454–482

Loftus E (2003) Make-believe memories. Am Psychol 58:867–873

Lombroso C (1876–1897) Criminal Man. Duke University Press, Durham

Mohamed FB, Faro SH, Gordon NJ, Platek SM, Ahmad H, Williams JM (2006) Brain mapping of deception and truth telling about an ecologically valid situation: functional MR imaging and polygraph investigation-initial experience. Radiology 238(2):679–688

Monteleone GT, Phan KL, Nusbaum HC, Fistgerald D, Hawkley LC, Irick JS, Fienberg SE, Cacioppo JT (2009) Detection of deception using functional magnetic resonance imaging: well above chance, though well below perfection. Social Neuroscience

Nichols TE (2012) Multiple testing corrections, nonparametric methods, and random field theory. Neuroimage 62(2):811–815

Orwell G (1949) Nineteen eighty-four. Secker and Warburg, London

Phan KL, Magalhaes A, Ziemlewiz TJ, Fitzgerald D, Green C, Smith W (2005) Neurl correlates of telling lies: a functional magnetic resonance imaging study at 4 Tesla. Acad Radiol 12(2):164–172

Porter R (2003) "Marginalized practices", The Cambridge history of science: eighteenth-century science, The Cambridge history of science, vol 4. Cambridge University Press, pp. 495–497

Raichle ME, Snyder AZ (2007) A default mode of brain function: a brief history of an evolving idea. Neuroimage 27(4):1083–1090

Rumelhart DE, McClelland JL, the PDP Research Group (1986) Parallel distributed processing: explorations in the microstructure of cognition, vol 1 and 2. MIT Press, Cambridge, MA

Violi A (2004) Lombroso e i Fantasmi Della Scienza. In: Forum Avanzato di Ricerca. Universitaria Multimediale (rivista elettronica)

Detection of Deception

▶ Detecting Deception with fMRI

Deterrence

Robert Apel[1] and Daniel S. Nagin[2]
[1]School of Criminal Justice, Rutgers University, Newark, NJ, USA
[2]Carnegie Mellon University, Pittsburgh, PA, USA

Overview

The criminal justice system dispenses justice by apprehending, prosecuting, and punishing individuals who break the law. These activities may also prevent crime by three distinct mechanisms – incapacitation, specific deterrence, and general deterrence. Convicted offenders are often punished with imprisonment. Incapacitation refers to the crimes averted by their physical isolation during the period of their incarceration. Specific deterrence and general deterrence involve possible behavioral responses. Specific deterrence refers to the reduction in reoffending that is presumed to follow from the *experience* of actually being punished. However, there are many sound reasons for suspecting that the experience of punishment might instead increase reoffending. The *threat* of punishment might also discourage potential and actual criminals in the general public from committing crime. This effect is known as general deterrence and is the subject of this entry.

Key Concepts of Deterrence

Deterrence is a theory of choice in which would-be offenders balance the benefits and costs of crime. Benefits may be pecuniary in the case of property crime but may also involve intangibles such as defending one's honor, expressing outrage, demonstrating dominance, cementing a reputation, or seeking a thrill. The potential costs of crime are comparably varied. For example, crime can entail personal risk if the victim resists, and it may also invoke pangs of conscience or shame. The theory of deterrence is predicated on the idea that if state-imposed sanction costs are sufficiently severe, criminal activity will be discouraged, at least for some. Thus, one of the key concepts of deterrence is the severity of punishment. In this entry, the review of severity effects focuses on research findings concerning imprisonment.

Severity alone, however, cannot deter. There must also be some possibility that the sanction will be incurred if the crime is committed. For that to happen, the offender must be apprehended, usually by the police. He must next be charged and successfully prosecuted, and finally sentenced by the judiciary. None of these successive stages in processing through the criminal justice system are certain. Thus, another key concept in deterrence theory is the certainty of punishment. In this regard, the most important set of actors are the police – absent detection and apprehension, there is no possibility of conviction or punishment. For this reason, the present entry separately considers what is known about the deterrent effect of police.

One of the key conclusions that emerged from the 1960s- and 1970s-era deterrence literature was that the certainty of punishment was a more powerful deterrent than the severity of punishment. The analyses of this era generally used cross-sectional data on states and involved testing the effects on the statewide crime rate of the certainty and severity of punishment, along with other demographic and socioeconomic control variables. The certainty of punishment was measured by the ratio of prison admissions to the number of reported crimes, while the severity of punishment was measured by median time served of recent prison releases. The basis for the "certainty not severity" deterrence conclusion was that punishment certainty was consistently found to have a negative and significant association with the crime rate, whereas punishment severity generally had no significant association.

This conclusion at the time was probably based on faulty statistical inference. Two primary criticisms were leveled. The first was that the negative association between the certainty measure and crime rate was an artifact of the number of crimes appearing in the denominator of the certainty measure and the numerator of the crime rate. It can be mathematically demonstrated that errors in the measurement of number of crimes, of which there are many, will force a negative, deterrent-like association between the crime rate and certainty even if, in fact, the certainty of punishment had no deterrent effect on crime. The second involved the use of theoretically indefensible statistical methods for parsing out the cause-effect relationship between sanction levels and the crime rate. After all, sanctions may deter crime, but crime may also affect sanction levels. For example, perhaps overcrowded prisons might reduce the chances of newly caught offenders going to prison. However, subsequent findings from the so-called perceptual deterrence literature and economic studies of the effects of contact with the criminal justice system on access to legal labor markets provide a far firmer empirical and theoretical basis for the "certainty principle." Due to space constraints, this entry will not cover these research traditions.

The Deterrent Effect of Imprisonment

There have been two distinct waves of studies of the deterrent effect of imprisonment. As already noted, studies in the 1960s and 1970s examined the relationship of the crime rate to the certainty of punishment, measured by the ratio of prison admissions to reported crimes, and the severity of punishment as measured by median time served. These studies suffered from a number of serious statistical flaws that are detailed in Blumstein et al. (1978). In response to these deficiencies, a second generation of studies emerged in the 1990s. Unlike the first-generation studies, which primarily involved cross-sectional analyses of states, second-generation studies had a longitudinal component in which data were analyzed not only across states but also over time. Another important difference is that the second-generation studies did not attempt to estimate certainty and severity effects separately. Instead, they examined the relationship between the crime rate and rate of imprisonment as measured by prisoners per capita.

A review by Donohue (2007) identifies six such studies. All find statistically significant negative associations between imprisonment rates and crimes rates, implying a crime-prevention effect of imprisonment. However, the magnitude of the estimate varied widely – from nil for a study that allowed for the possibility of diminishing returns to an elasticity of −0.4. (By an elasticity of −0.4, it is meant that 10 % growth in the imprisonment rate reduced the crime rate by 4 %.) It is important to note that these studies are actually measuring a combination of deterrent and incapacitation effects. Thus, it is impossible to decipher the degree to which crime prevention is occurring because of a behavioral response by the population at large or because of the physical isolation of crime-prone people.

Donohue (2007) goes on to show that the small elasticity estimates imply that the current imprisonment rate is too large, while the high-end estimates imply the rate is too small. He lists a variety of technical shortcomings of these studies that make it impossible to distinguish among the widely varying effect size estimates. The most important is the degree to which the studies were successful in separating cause from effect. While imprisonment prevents crime through a combination of deterrence and incapacitation, crime also generates the prison population. This is an example of what is called the "simultaneity problem," whereby one wants to ascertain the effect of one variable (the imprisonment rate) on another variable (the crime rate) in a circumstance where it is known or suspected that reverse causation is also present, namely, that the crime rate simultaneously affects the imprisonment rate. Thus, statistical isolation of the crime-prevention effect requires properly accounting for the effect of crime on imprisonment. The Levitt (1996) study is arguably the most successful in this regard. It uses court-ordered prison releases as an instrument for untangling the cause-and-effect relationship. However, even the Levitt analysis suffers from many of the technical limitations detailed by Donohue.

More fundamentally, this literature suffers from more than just technical shortcomings that future research might strive to correct. It also suffers from important conceptual flaws that limit its usefulness in devising crime-control policy. Prison population is not a policy variable; rather, it is an outcome of sanction policies dictating who goes to prison and for how long, namely, the certainty and severity of punishment. In all incentive-based theories of criminal behavior, the deterrence response to sanction threats is posed in terms of the certainty and severity of punishment, not in terms of the imprisonment rate. Therefore, to predict how changes in certainty and severity might affect the crime rate requires knowledge of the relationship of the crime rate to certainty and severity as separate entities, which is not provided by the literature that analyzes the relationship of the crime rate to the imprisonment rate. The studies are also conducted at too global a level. There are good reasons for predicting differences in the crime-reduction effects of different types of sanctions (e.g., mandatory minimums for repeat offenders vs. prison diversion programs for first-time offenders). Obvious sources of heterogeneity in offender response include factors such as prior contact with the criminal justice system, demographic characteristics, and the mechanism by which sanction threats are communicated to their intended audience.

Three studies nicely illustrate heterogeneity in the deterrence response to the threat of imprisonment: The Weisburd et al. (2008) study on the use of imprisonment to enforce fine payment finds a substantial deterrent effect; the Helland and Tabarrok (2007) analysis of the deterrent effect of California's third-strike provision finds only a modest deterrent effect; and the Lee and McCrary (2009) examination of the heightened threat of imprisonment that attends coming under the jurisdiction of the adult courts at the age of majority finds no deterrent effect. These three important studies are considered in more detail below.

Weisburd et al. (2008) report on a randomized field trial of alternative strategies for incentivizing the payment of court-ordered fines. The most salient finding involves the "miracle of the cells," namely, that the imminent threat of incarceration is a powerful incentive for paying delinquent fines. The miracle of the cells provides a valuable vantage point for considering the oft-repeated conclusion from the deterrence literature that the certainty rather the severity of punishment is the more powerful deterrent. Consistent with the "certainty principle," the common feature of treatment conditions involving incarceration was a high certainty of imprisonment for failure to pay the fine. However, the fact that Weisburd and colleagues label the response the "miracle of the cells" and not the "miracle of certainty" is telling. Their choice of label is a reminder that certainty must result in a distasteful consequence, namely, incarceration in this experiment, in order for it to be a deterrent. The consequences need not be draconian, just sufficiently costly to deter proscribed behavior.

Helland and Tabarrok (2007) examine whether California's "Three Strikes and You're Out" law deters offending among individuals previously convicted of strike-eligible offenses. The future offending of individuals convicted of two previous strike offenses was compared with that of individuals who had been convicted of only one strike offense but who, in addition, had been tried for a second-strike offense but were ultimately convicted of a non-strike offense. The study demonstrates that these two groups of individuals were comparable on many characteristics such as age, race, and time in prison. Even so, it finds that arrest rates were about 20 % lower for the group with convictions for two strike offenses. The authors attribute this reduction to the greatly enhanced sentence that would have accompanied conviction for a third-strike offense.

For most crimes, the certainty and severity of punishment increases discontinuously upon reaching the age of majority, when jurisdiction for criminal wrongdoing shifts from the juvenile to the adult court. In an extraordinarily careful analysis of individual-level crime histories from Florida, Lee and McCrary (2009) attempt to identify a discontinuous decline in the hazard of offending at age 18, the age of majority in Florida. Their point estimate of the discontinuous change is negative as predicted, but minute in magnitude and not even remotely close to achieving statistical significance.

In combination, these three studies nicely illustrate that the deterrent effect of the threat of punishment is context-specific and that debates about whether deterrence works or not are ill posed. Instead, the discussion should be in terms of whether the specific sanction deters or not and if it does, whether the benefits of crime reduction are sufficient to justify the costs of imposing the sanction. To illustrate, while Helland and Tabarrok (2007) conclude that the third-strike effect in California is a deterrent, they also conclude, based on a cost-benefit analysis, that the crime-saving benefits are likely far smaller than the increased costs of incarceration. The Helland and Tabarrok study is an exemplar of the approach that should be taken in evaluating different sanctioning regimes.

The Deterrent Effect of Police

The police may prevent crime through many possible mechanisms. Apprehension of active offenders is a necessary first step for their conviction and punishment. If the sanction involves imprisonment, crime may be prevented by the incapacitation of the apprehended offender. The apprehension of active offenders may also deter would-be criminals by increasing their perception of the risk of apprehension and thereby the certainty of punishment. Many police tactics such as rapid response to calls for service at crime scenes or post-crime investigation are intended not only to capture the offender but to deter others by projecting a tangible threat of apprehension. Police may, however, deter without actually apprehending criminals because their very presence projects a threat of apprehension if a crime were to be committed. Indeed, some of the most compelling evidence of deterrence involve instances where there is complete or near-complete collapse of police presence. In September 1944, German soldiers occupying Denmark arrested the entire Danish police force. According to an account by Andeneas (1974), crime rates rose immediately but not uniformly. The frequency of street crimes like robbery, whose control depends heavily upon visible police presence, rose sharply. By contrast, crimes like fraud were less affected.

The Andenaes anecdote illustrates two important points. First, sanction threats (or the absence thereof) may not uniformly affect all types of crime and more generally all types of people. Second, it draws attention to the difference between absolute and marginal deterrence. Absolute deterrence refers to the difference in the crime rate between the status quo level of sanction threat and a complete (or near) absence of sanction threat. The Andenaes anecdote is a compelling demonstration that the absolute deterrent effect is large. However, from a policy perspective, the important question is whether, on the margin, crime deterrence can be affected by incrementally manipulating sanction threats.

Research on the marginal deterrent effect of police has evolved in two distinct literatures. One has focused on the deterrent effect of the aggregate police presence measured, for example, by the relationship between police per capita and crime rates. The other has focused on the crime-prevention effectiveness of different strategies for deploying police. These two literatures are reviewed separately.

Aggregate Police Presence and Crime

Studies of police hiring and crime rates have been plagued by a number of impediments to causal inference. Among these are cross-jurisdictional differences in the recording of crime, feedback effects from crime rates to police hiring, the confounding of deterrence with incapacitation, and aggregation of police manpower effects across heterogeneous units, among others. Yet the challenge that has received the most attention in empirical applications is the simultaneity problem referred to in the previous section – in the present case, the feedback from crime rates to police hiring.

The two studies of police manpower by Marvell and Moody (1996) and Levitt (1997) are notable for their different identification strategies. The Marvell and Moody (1996) study is based on an analysis of two panel data sets: one composed of 49 states for the years 1968–1993 and the other of 56 large cities for the years 1971–1992. To untangle the simultaneous causation problem, they regress the current crime rate on lags of the crime rate as well as lags of police manpower. If the lagged police measures are jointly significant, they are said to "Granger cause" crime. The strongest evidence for an impact of police hiring on total crime rates comes from the city-level analysis, with an estimated elasticity of −0.3, meaning that 10 % growth in police manpower produces a 3 % decline in the crime rate the following year.

However, regression analyses of this type do not generally provide a valid basis for making causal claims. But other forms of analysis can provide such a basis – one is instrumental variables regression. Levitt (1997) performs an instrumental variables (IV) analysis from a panel of 59 large cities for the years 1970–1992. Reasoning that political incumbents have incentives to devote resources to increasing the size of the police force in anticipation of upcoming elections, he uses election cycles to help untangle the cause-effect relationship between crime rates and police manpower. Levitt's model produces elasticities of about −1.0 for the violent crime rate and −0.3 for the property crime rate. Following Levitt's use of the electoral cycle as an instrument for the number of sworn police officers, other studies have employed alternative instrumental variables and reported comparable elasticities.

In recent years, a number of more targeted tests of the police-crime relationship have appeared. These studies investigate the impact on the crime rate of reductions in police presence and productivity as a result of massive budget cuts or lawsuits following racial profiling scandals. Each of these studies concludes that increases (decreases) in police presence and activity substantially decrease (increase) crime. By way of example, Shi (2009) studies the fallout from an incident in Cincinnati in which a white police officer shot and killed an unarmed African-American suspect. The incident was followed by 3 days of rioting, heavy media attention, the filing of a class action lawsuit, a federal civil rights investigation, and the indictment of the officer in question. These events created an unofficial incentive for officers from the Cincinnati Police Department to curtail their use of arrest for misdemeanor crimes, especially in communities with higher proportional representation of African-Americans out of concern for allegations of racial profiling. Shi demonstrates measurable declines in police productivity in the aftermath of the riot and also documents a substantial increase in criminal activity. The estimated elasticities of crime to policing based on her approach were −0.5 for violent crime and −0.3 for property crime.

The ongoing threat of terrorism has also provided a number of unique opportunities to study the impact of police resource allocation in cities around the world. The study by Klick and Tabarrok (2005) examines the effect on crime of the color-coded alert system devised by the US Department of Homeland Security (in the aftermath of the September 11, 2001, terrorist attack) to denote the terrorism threat level. Its purpose was to signal federal, state, and local law enforcement agencies to occasions when it might be prudent to divert resources to sensitive locations. Klick and Tabarrok (2005) use daily police reports of crime for the period March 2002 to July 2003, during which time the terrorism alert level rose from "elevated" (yellow) to "high" (orange) and back down to "elevated" on four occasions. During high alerts, anecdotal evidence suggested that police presence increased by 50 %. Their estimate of the elasticity of total crime to changes in police presence as the alert level rose and fell was −0.3.

To summarize, aggregate studies of police presence conducted since the mid-1990s consistently find that putting more police officers on the street – either by hiring new officers or by allocating existing officers in ways that put them on the street in larger numbers or for longer periods of time – has a substantial deterrent effect on serious crime. There is also consistency with respect to the size of the effect. Most estimates reveal that a 10 % increase in police presence yields a reduction in total crime in the neighborhood of 3 %, although studies that consider violent crime tend to find reductions ranging from 5 % to 10 %. Yet these police manpower studies speak only to the number and allocation of police officers and not to what police officers actually do on the street beyond making arrests. The next section proceeds from here by reviewing recent evaluations of deployment strategies used by police departments in order to control crime.

Police Deployment and Crime

Much research has examined the crime-prevention effectiveness of alternative strategies for deploying police resources. This research has largely been conducted by criminologists and sociologists. Among this group of researchers, the preferred research designs are quasi-experiments involving before-and-after studies of the effect of targeted interventions as well as true randomized experiments. The discussion that follows draws heavily upon two excellent reviews of this research by Weisburd and Eck (2004) and Braga (2008). As a preface to this summary, the theoretical link between police deployment and the certainty and severity of punishment is clarified. For the most part, deployment strategies affect the certainty of punishment through its impact on the probability of apprehension. There are, however, notable examples where severity may also be affected.

One way to increase apprehension risk is to mobilize police in a fashion that increases the probability that an offender is arrested after committing a crime. Strong evidence of a deterrent as opposed to an incapacitation effect resulting from the apprehension of criminals is limited. Studies of the effect of rapid response to calls for service (Spelman and Brown 1981) find no evidence of a crime-prevention effect, but this may be because most calls for service occur well after the crime event, with the result that the perpetrator has fled the scene. Thus, it is doubtful that rapid response materially affects apprehension risk. Similarly, because most arrests result from the presence of witnesses or physical evidence, improved investigations are not likely to yield material deterrent effects because, again, apprehension risk is not likely to be affected. A series of randomized experiments were conducted to test the deterrent effect of mandatory arrest for domestic violence. The initial experiment conducted in Minneapolis by Sherman and Berk (1984) found that mandatory arrest was effective in reducing domestic violence reoffending. However, findings from follow-up replication studies (as part of the so-called Spouse Assault Replication Program, or SARP) were inconsistent.

The second source of deterrence from police activities involves averting crime in the first place. In this circumstance, there is no apprehension because there was no offense. This is the primary source of deterrence from the presence of police. If an occupied police car is parked outside a liquor store, a would-be robber of the store will likely be deterred because apprehension is all but certain. Thus, measures of apprehension risk based only on enforcement actions and crimes that actually occur, such as arrest per reported crime, are seriously incomplete because such measures do not capture the apprehension risk that attends criminal opportunities that were not acted upon by potential offenders because the risk was deemed too high.

Two examples of police deployment strategies that have been shown to be effective in averting crime in the first place are “hot spots” policing and problem-oriented policing. Weisburd and Eck (2004) propose a two-dimensional taxonomy of policing strategies. One dimension is “level of focus” and the other is “diversity of focus.” Level of focus represents the degree to which police activities are targeted. Targeting can occur in variety of ways, but Weisburd and Eck give special attention to policing strategies that target police resources in small geographic areas (e.g., blocks or specific addresses) that have very high levels of criminal activity, so-called crime “hot spots.” Just like in the liquor store example, the rationale for concentrating police in crime hot spots is to create a prohibitively high risk of apprehension and thereby to deter crime at the hot spot in the first place.

Braga’s (2008) informative review of hot spots policing summarizes the findings from nine experimental or quasi-experimental evaluations. The studies were conducted in five large US cities and one suburb of Australia. Crime-incident reports and citizen calls for service were used to evaluate impacts in and around the geographic area of the crime hot spot. The targets of the police actions varied. Some hot spots were generally high-crime locations, whereas others were characterized by specific crime problems like drug trafficking. All but two of the studies found evidence of significant reductions in crime. Further, no evidence was found of material crime displacement to immediately surrounding locations. On the contrary, some studies found evidence of crime reductions, not increases, in the surrounding locations – a “diffusion of crime-control benefits” to non-targeted locales.

The second dimension of the Weisburd and Eck taxonomy is diversity of approaches. This dimension concerns the variety of approaches that police use to impact public safety. Low diversity is associated with reliance on time-honored law enforcement strategies for affecting the threat of apprehension, for example, by dramatically increasing police presence. High diversity involves expanding beyond conventional practice to prevent crime. One example of a high-diversity approach is problem-oriented policing. Problem-oriented policy comes in so many different forms that it is regrettably hard to define.

One of the most visible examples of problem-oriented policing is Boston’s Operation Cease Fire (Kennedy et al. 2001). The objective of the collaborative operation was to prevent inter-gang gun violence using two deterrence-based strategies. One was to target enforcement against weapons traffickers who were supplying weapons to Boston’s violent youth gangs. The second involved a more innovative use of deterrence. The youth gangs themselves were assembled (and reassembled) to send the message that the response to any instance of serious violence would be “pulling every lever” legally available to punish gang members collectively. This included a salient severity-related dimension – vigorous prosecution for unrelated, nonviolent crime such as drug dealing. Thus, the aim of Operation Cease Fire was to deter violent crime by increasing the certainty and severity of punishment but only in targeted circumstances, namely, if the gang members were perpetrators of a violent crime. Just as important, Operation Cease Fire illustrates the potential for combining elements of both certainty and severity enhancement to generate a targeted deterrent effect. Further evaluations of the efficacy of this strategy should be a high priority.

Conclusions

This entry has reviewed the evidence on the general deterrent effect of sanctions. Evidence of a substantial effect is overwhelming. Just as important is the evidence that the effect is not uniform across different sanctions, jurisdictions, and individuals. Both conclusions are important to devising crime-control policies that make effective use of sanctions to prevent crime. The first conclusion implies that a well-balanced portfolio of strategies and programs to prevent crime must necessarily include deterrence-based policies. However, the second conclusion implies that not all deterrence policies will be effective in reducing crime or, if effective, that the crime-reduction benefits may fall short of the social and economic costs of the sanction.

Future research on sanction effects will be most useful for policy evaluation if it moves closer to a medical model. Medical research is not organized around the theme of whether medical care cures diseases, the analog to the question of whether sanctions prevent crime. Instead, medical researchers address far more specific questions. Is a specific drug or procedure effective in treating a specific disease? Does the drug or procedure have adverse side effects for certain types of people? Furthermore, most such research is comparative – is the specific drug or procedure more effective than the status quo alternative? The analogous questions for deterrence research are whether and in what circumstances are sanction threats effective, and which threats are more effective and in what circumstances.

Related Entries

- ► Certainty, Severity, and Their Deterrent Effects
- ► Deterrence: Actual Versus Perceived Risk of Punishment

- ▶ Deterrent Effect of Imprisonment
- ▶ Econometrics of Crime
- ▶ Focused Deterrence and "Pulling Levers"
- ▶ Rational Choice, Deterrence, and Crime: Sociological Contributions
- ▶ Situational Action Theory

Recommended Reading and References

Andenaes J (1974) Punishment and deterrence. University of Michigan Press, Ann Arbor

Blumstein A, Cohen J, Nagin D (1978) Deterrence and incapacitation: estimating the effects of criminal sanctions on crime rates. National Academies Press, Washington, DC

Braga AA (2008) Police enforcement strategies to prevent crime in hot spot areas. Office of Community Oriented Policing, U.S. Department of Justice, Washington, DC

Donohue J (2007) Assessing the relative benefits of incarceration: the overall change over the previous decades and the benefits on the margin. Working paper. Yale Law School, New Haven

Helland E, Tabarrok A (2007) Does three strike deter? A nonparametric estimation. J Hum Resour 42:309–330

Kennedy DM, Braga AA, Piehl AM, Waring EJ (2001) Reducing gun violence: the Boston gun project's operation ceasefire. U.S. National Institute of Justice, Washington, DC

Klick J, Tabarrok A (2005) Using terror alert levels to estimate the effect of police on crime. J Law Econ 48:267–279

Lee DS, McCrary J (2009) The deterrent effect of prison: dynamic theory and evidence. Unpublished manuscript. Princeton University, Princeton

Levitt SD (1996) The effect of prison population size on crime rates: evidence from prison overcrowding litigation. Q J Econ 111:319–351

Levitt SD (1997) Using electoral cycles in police hiring to estimate the effect of police on crime. Am Econ Rev 87:270–290

Marvell T, Moody C (1996) Specification problems, police levels, and crime rates. Criminology 34: 609–646

Sherman LW, Berk RA (1984) The specific deterrent effects of arrest for domestic assault. Am Sociol Rev 49:261–272

Shi L (2009) The limits of oversight in policing: evidence from the 2001 Cincinnati Riot. J Public Econ 93: 99–113

Spelman W, Brown DK (1981) Calling the police: a replication of the citizen reporting component of the kansas city response time analysis. Police Executive Research Forum, Washington, DC

Weisburd D, Eck J (2004) What can police do to reduce crime, disorder, and fear? Ann Am Acad Pol Soc Sci 593:42–65

Weisburd D, Einat T, Kowalski M (2008) The miracle of the cells: an experimental study of interventions to increase payment of court-ordered financial obligations. Criminol Public Policy 7:9–36

Deterrence of Tax Evasion

Christian Traxler
Department of Economics, University of
Marburg, Marburg, Germany
Max Planck Institute for Research on Collective
Goods, Bonn, Germany

Overview

Building on the logic of Gary Becker's rational choice model of crime, Allingham and Sandmo (1972) developed the theory of income tax evasion. Their model studies the choice of a taxpayer who rationally trades off the benefits from evading taxes and the risky costs from detection and incurring fines. By increasing auditing frequencies or the magnitude of fines and penalties, tax authorities can deter evasion. This survey discusses three selected aspects of the literature that studies the general deterrence hypothesis in the context of tax evasion. First, the entry will summarize several empirically studies that convincingly solved the key identification problem. The majority of these studies provide evidence that strongly support the deterrence hypothesis. In a second step, the present survey will address the role of taxpayers' imperfect information about the actual enforcement policy. This point, which is closely related to the field of perceptual deterrence research, has recently gained attention in empirical work on deterrence in the context of evasion. The third and last aspect covered in this survey concerns the role of standard deterrence incentives as compared to the tax morale and alternative noneconomic factors that shape tax compliance. The entry concludes with a brief summary and remarks on future directions for research on tax enforcement.

Introduction

Four years after Becker's seminal work on crime and punishment, Michael Allingham and Agnar Sandmo published their analysis of income tax evasion (Allingham and Sandmo 1972). They considered a risk-averse taxpayer who decides how much of his income to report to tax authorities and how much to conceal. Taxes on concealed incomes are evaded. With a certain probability, however, the taxpayer is audited and all the concealed income is detected. In this case, the true taxes plus a fine – which is increasing in the concealed income – have to be paid. Rational taxpayers choose the optimal level of underreported income, trading off the benefits (saving on taxes) with the costs (higher fines) from evasion. Similarly as in Becker's (1968) work, there is scope for general deterrence: by increasing the probability of detection and the magnitude of fines (and other penalties), evasion becomes less attractive. A rational taxpayer will therefore conceal less income. Along these lines, tax authorities can enforce a certain level of compliance with taxes.

The work by Allingham and Sandmo (1972), one of the most cited articles in public economics, triggered a wave of theoretical studies that quickly extend the original analysis. (For detailed surveys of this literature, see Cowell (1990), Andreoni et al. (1998), and Slemrod and Yitzhaki (2002).) The empirical assessment of the model's implications, however, evolved at a slower rate. In particular, the test of the deterrence hypothesis in the context of tax evasion turned out to be quite complicated. First of all, evasion and tax enforcement – i.e., the level of auditing determining the detection risk as well as the size of the fines – are endogenously determined: evasion responds to enforcement, but enforcement policies also react to evasion behavior (exactly as with the police-crime circularity). By now, only a handful studies managed to break this endogeneity. Section 2 (Causal Evidence on Deterrence) discusses several of these contributions. Except for one study, these contributions provide evidence suggesting that deterrence works.

A second issue in empirical studies on deterrence in the context of tax evasion emerges from concerns about the limited knowledge of taxpayers about the actual enforcement policies. While the imperfect information of decision makers about the likelihood and magnitude of sanctions is of general interest for the research on deterrence (Bebchuk and Kaplow 1992), this point appears particularly relevant for the case of tax enforcement. In contrast to visible indicators for enforcement such as police on the streets, there are few cues that taxpayers can use to assess tax enforcement policies. In addition, tax enforcement agencies do not reveal details about their auditing practice and often – at least in many continental European countries –operate quietly in the background. Section 3 (Tax Enforcement and Perceptual Deterrence) discusses a very recent strand of literature on general deterrence in the context of evasion that explicitly addresses this point. This literature, which closely related to the field of perceptual deterrence research (see the survey in Nagin 1998), points to the importance of the channels through which information on detection and punishment is spread.

A further complication in the empirical analysis of deterrence emerges in the operationalization of the theoretical concept of "the detection risk." The problem is best illustrated with the group of (not self-employed) labor income receivers. In modern economies, the evasion rate in this subpopulation is almost zero (Slemrod 2007). At the same time, the effective auditing frequency in this group is close to zero, too. The observation of this evasion and enforcement pattern has motivated many researchers to question the validity of the model of rational taxpayers: given the low auditing rate, the standard model would predict more evasion. To explain this puzzle, the importance of noneconomic motives for compliance – often subsumed among the label "tax morale" – has been put forward (Torgler 2007). Section 4 (Tax Morale and Third-Party Information) will briefly summarize this literature. The section will also highlight that the modern literature classifies the zero-evasion, zero-auditing observation quite differently. In practice, an individual's auditing risk is not a fixed parameter (as assumed in the simple benchmark model), but rather a function of her evasion choice: the more income she conceals (given a certain level of true income), the higher is her auditing risk. Observing an effective auditing frequency that is close to zero corresponds to just one point of this auditing function (see, e.g., Kleven et al. 2011). Hence, full compliance can be a rational response to standard incentives from deterrence if small levels of underreporting would be detected with a high probability.

Causal Evidence on Deterrence

The identification of a deterrent effect from auditing and the severity of sanctions is complicated by several problems. First of all, it is obviously difficult to measure evasion (Slemrod and Weber 2012). Secondly, enforcement strategies are no exogenous: it is not only evasion, which should respond to enforcement; the choices of enforcement authorities are also determined by the (past, present, or expected future) level of evasion.

Early studies on the determinants of tax evasion were mainly concerned with the first issue (e.g., Clotfelter 1983; Feinstein 1991). Based on random auditing data from the US tax authority's Taxpayer Compliance Measurement Program (TCMP), these contributions derived individual level measures for evasion. In some studies, this dependent variable was regressed on regional measures of auditing frequencies or income type specific proxies (e.g., Dubin et al. 1990; Beron et al. 1992). In doing so, however, they did not cope with the second issue, the endogeneity of enforcement. Cross-sectional and intertemporal variation in regional auditing rates, for instance, might reflect variation in the unobserved determinants of the taxpayers' noncompliance that also drive evasion.

A more recent wave of contributions that studies the determinants of evasion behavior is based on randomized field experiments. These experiments provided truly exogenous variation in enforcement policies and therefore contributed to what Slemrod and Weber (2012) called the "credibility revolution in the empirical analysis of tax evasion." One of the first studies is based on the Minnesota Income Tax Compliance Experiment (Slemrod et al. 2001). In cooperation with a team of researchers, the Internal Revenue Service (IRS, the US tax authority) approached a randomly selected set of taxpayers from a predefined group of income earners with different types of letters. Next to a baseline letter that served as a control treatment, one treatment included a letter with the information that the tax return filed by the taxpayer was selected to be "closely examined." The comparison of the first-differences in reported incomes provided mixed results.

In line with the deterrence hypothesis, low- and middle-income taxpayers responded to the audit treatment with an increase in reported income. This effect was considerably stronger for those with more opportunities to evade income (e.g., self-employed income). For high-income taxpayers, however, the audit threat backfired: relative to the control group, it induced *lower* reported tax liabilities. One possible interpretation of this observation is that high-income taxpayers who know that they will be audited with certainty behaved strategically. By entering the bargaining-like interaction with the IRS with a lower reported income, these taxpayers might expect to end up with lower tax payments.

More clear evidence on deterrence is offered by a recent field experiment in Denmark. In cooperation with the Danish tax authorities, Kleven et al. (2011) gained access to a stratified and representative sample of more than 40,000 income tax filers. Half of the individuals were randomly assigned to similar audit threat treatments as in Slemrod et al. (2001), communicating either a 100 % or a 50 % chance of an audit. Relative to a control treatment that did not receive any letter, the audit threats triggered a significant increase in reported incomes, with a larger effect under the 100 % auditing risk (Kleven et al. 2011).

In contrast to the other contributions, the study by Fellner et al. (2013) does not rely on a random draw from a certain population but rather a selected sample of 50,000 potential evaders. They experimentally study the effect of various mailing types on the evasion of TV license fees. Relative to a neutral mailing, an audit threat treatment that also stressed the severity of sanctions had a significantly positive impact on compliance. The evidence in Fellner et al. (2013) therefore shows that deterrence also works for the most important group – those that self-selected into evasion.

All these contributions focused on independent individual or household level choices of evasion. The field experiment by Pomeranz (2011) documents an interesting spillover effect from deterrence in the context of joint, interdependent evasion decisions. Based on a large-scale field experiment with more than 400,000 small Chilean firms, she finds that audit threat letters not only reduces the evasion of Value Added Taxes (VAT) of the targeted firms (i.e., the firm that receives the letter), but also for untreated suppliers, who are also involved in the evasion decision. The logic behind this observation is straightforward: the potential detection of the VAT evasion of the treated firm increases the detection risk of the untreated firm. Hence, in the context of interdependent evasion decisions – as it is the case with VAT evasion – the deterrence effect from tax enforcement is multiplied.

Beyond this strand of research that experimentally varied the threat from an audit, there is still little evidence on the specific impact of actual audits. A rare piece of causal evidence on specific deterrence in the domain of tax enforcement is again found in the Danish study from above (for an earlier studies, see Erard 1992). In addition to and independently of the letter treatments, Kleven et al. (2011) randomly selected a group of taxpayers that was actually audited. Experiencing an audit turned out to have a substantial positive effect on the level of reported incomes in the years following the audit. Quantitatively, this effect was much stronger than the impact of the threat letters (Kleven et al. 2011). While the mechanisms driving this observation are not yet identified, the evidence suggests that actual audits have a much stronger and more long-lasting effect than simple audit threats (on the longevity of deterrence effects, see also Fellner et al. 2013).

Tax Enforcement and Perceptual Deterrence

As Becker's model of crime, the rational model of evasion assumes that decision makers know the precise detection probability set by the enforcement authority. Perceptual deterrence research provides some evidence that "there presumably is a positive relationship between actual and perceived levels of enforcement"; however, it is considered "implausible that individuals' probability estimates are generally accurate, particularly when the probability is extremely low" (Bebchuk and Kaplow 1992, p. 366). While this observation is of general interest for research on deterrence, it appears particularly relevant for the case of tax enforcement. In contrast to visible cues such as the number of policemen on the streets, there is hardly any publicly accessible information that allows potential tax evaders to accurately predict their true detection risk. This raises the question, how changes in the enforcement policy translate into changes in compliance behavior.

The question was addressed in the field experiment in Fellner et al. (2013). The researchers confronted subjects with the different letter treatments, in particular, letters with and without an audit threat (see above). Subjects were then asked to indicate their perceptions about the detection risk and potential fines, etc. The results from the perception survey neatly complemented the evidence on the behavioral responses observed in the field experiment: the audit threat considerably increased the perceived detection risk. The evidence suggests that the threats shaped perceptions and that individuals rationally adjust their compliance behavior to these perceptions.

While Fellner et al. (2013) provide evidence on the impact of a targeted policy on individual perceptions and compliance behavior, the impact of "hidden" policies – i.e., generally unobservable enforcement activities – in a society remains unclear. A theoretical analysis of this issue is provided by Sah (1991), who models agents that update their perceived detection risk based on information obtained from neighbors and acquaintances. A higher number of detections among this sample ceteris paribus result in an increased risk perception. In turn, the inclination to comply with the law increases. Implicitly, the model therefore highlights the crucial role of communication for the dispersion of information on enforcement activities.

A recent study that addressing behavioral responses to interpersonal communication in the context of tax enforcement is Alm et al. (2009), who study tax compliance in a laboratory experiment. They implemented a design in which there was no "official" information on auditing risks available; however, they allowed the experimental subjects to communicate. Alm et al. find that income reporting is sensitive to information on auditing frequencies that is spread via communication. More closely related to the model in Sah (1991) is the contribution by Rincke and Traxler (2011), who provide field evidence on a substantial deterrent effect that is mediated by word of mouth. They analyze the enforcement activities of field inspectors that approach potential evaders of TV license fees (compare Fellner et al. 2013) at their homes. These field inspectors unregularly visit different neighborhoods in different months. The presence of inspectors in an area is not announced; inspectors are not uniformed nor do they use any police-like cars. Hence, it is basically impossible to observe if inspectors are visiting an area.

Making use of monthly panel data on the activities of field inspectors in more than 1,000 small municipalities, Rincke and Traxler (2011) identify a substantial impact of these "hidden" enforcement activities. Following an instrumental variable approach, they find that three additional detected evaders induce one further evader (that was not approached by field inspectors) to switch to compliance. Making use of microlevel data, they provide further evidence that strongly suggests that the effect is driven by communication about detections among neighbors. In addition, they show that the enforcement spillover is concentrated in the close spatial proximity to the detected household.

The findings in Alm et al. (2009) and Rincke and Traxler (2011) provide two important insights. First and most importantly, the fact that enforcement activities are often "hidden" and that auditing rules are typically not announced by tax authorities does *not* imply that an increase in enforcement activities fails to deter evasion. Communication on experiences with enforcement authorities seems to be sufficient to convey the information about an increased detection risk. This implies – and that is the second point to take away – that patterns of communication and information dispersion become decisive for general deterrence.

Tax Morale and Third-Party Information

Third-Party Information

The rational model of evasion predicts that a sufficiently high detection probability should deter evasion. The comparison of evasion levels for different sources of incomes (Slemrod 2007), however, seems to provide a picture that first appears at odds with this prediction. In modern economies, the evasion of taxes on nonbusiness labor income is nearly zero. At the same time, the auditing rate of labor income receivers is close to zero, too. As noted earlier, this pattern has motivated many researchers to question the rational model of evasion: it has been argued that the standard model would predict "too much evasion." As an explanation for this "puzzle," noneconomic motives for compliance, often subsumed among the label tax morale, have been put forward (e.g., Torgler 2007). The modern taxation literature, however, provides a different explanation to the zero-evasion, zero-auditing observation: the presence of third-party income reporting, which makes many taxpayers unable to evade.

In its most simple version, the standard model of evasion captures the risk of detection in one single parameter, the auditing probability p. In practice, however, an audit does not necessarily result in the detection of all evasion. In addition, and more importantly, an individual's auditing risk is not given by a constant parameter, but rather a function of her underreporting and her true income. It appears plausible to assume that, for a given level of true income and a given income source, an individual's risk of detection increases with a higher level of underreporting (several extensions of the basic model follow this approach; see the surveys in Andreoni et al. 1998; Cowell 1990). If one thinks about the observation of zero-auditing and zero-evasion for labor incomes along the lines of this more realistic model, one arrives at a straightforward explanation.

In economies with a modern tax enforcement system, labor income is typically "third-party reported." This means that not only the labor income receiving taxpayer declares her incomes to the tax authority, but also the firm who pays these incomes (and often withholds taxes) reports the payments. Tax authorities can therefore easily compare the taxpayer's declared income with the income that is reported by her firm. For a taxpayer who exclusively receives labor income, this implies that she is basically unable to conceal income (see, e.g., Kleven et al. 2011). While there is a zero-auditing risk under full compliance (which is observed on average), any noncompliance will get detected basically with certainty. Given this nonlinear detection risk, full compliance is a rational response to standard economic incentives – and perfectly in line with a more realistic extension of the baseline model of rational evaders.

This point appears trivial. In fact, it was well understood among policy-oriented researchers for a long time (see, e.g., Long and Swingen 1990). For the theoretically motivated research, however, it took more time until it was acknowledged that the model by Allingham and Sandmo (1972) is a simplified model of a taxpayer that fully self-reports her, e.g., self-employed income, rather than a model that describes the evasion of third-party reported income. (For a theoretical analysis of the case where employer and employee jointly evade taxes on third-party reported incomes, see Kleven et al. 2009.) In line with the baseline model, fully self-reported incomes – in contrast to third-party reported incomes – are actually concealed to a significant amount (Slemrod 2007).

The availability of third-party information provides a convincing explanation for the high compliance and the low auditing risk that is observed for labor incomes. Hence, the argument that this pattern can *only* be explained by tax morale or social norms for tax compliance is not justified. This does not mean, however, that tax morale – or any other behavioral incentive beyond those captured in the standard model – is irrelevant for tax compliance. It is therefore interesting to consider possible noneconomic incentives and their relevance for tax enforcement. The remainder of this section therefore discusses the role of tax morale. Models of bounded rational taxpayers, probability weighting, and loss aversion are beyond the scope of this survey (see, e.g., Dhami and al-Nowaihi 2007, and references therein).

Tax Morale

The modern literature uses the term tax morale quite broadly, typically subsuming the idea of a social norm for tax compliance, perceptions about civic duties as "honest taxpayer," but also the notion of stigmatization (for a survey, see Torgler 2007). Theoretical approaches typically model tax morale in form of an additional, noneconomic cost of evasion (an overview of different modeling strategies is provided by Traxler 2010). In these frameworks, concealing income triggers economic consequences (in case of detection) as well as psychic (e.g., shame) or social cost (e.g., from social exclusion) that are motivated by tax morale. These costs might differ between different individuals – some agents might not be affected by tax morale, others might even perceive a moral obligation to evade taxes. In addition, it is typically assumed that the noneconomic costs of evasion depend on the compliance behavior of "relevant others" (Traxler 2010). In a society, where everybody complies with taxes, the noneconomic costs of evasion might be large. Hence, individuals experience strong incentives to comply with taxes. However, if evasion would be widespread in this society, the same individuals might perceive hardly any noneconomic costs of cheating and might evade taxes. A tax morale that depends on the behavior of others therefore has an important implication: it gives scope for a multiplicity of equilibria (or, more generally, social interaction in tax compliance; see Galbiati and Zanella 2012). For a given level of tax enforcement, a society might coordinate on a state where evasion is widespread or a state where there is a strong norm to comply with taxes (and, potentially, further equilibria between these extremes; see Traxler 2010).

These ideas on tax morale have several important implications for the effectiveness of deterrence. The first two are mainly of theoretical interest: changing the level of tax enforcement might eliminate the existence of one (out of several) possible equilibrium. Hence, minor changes in the enforcement policy could, in principle, produce large shifts in compliance behavior. Related to this point, one can think more generally about the interaction of the formal, legal and informal, private enforcement of compliance: if stricter tax enforcement is an expression of the societies' weight attached to compliance with this rule (an idea from expressive law), it might also crowd-in stronger psychic or social cost from tax evasion. Similar as with the private versus public provision of public goods, however, there might also be scope for crowding-out (Kube and Traxler 2011).

A third implication relates to the notion of the "relevant others." If the most relevant others (in terms of shaping perceptions about tax morale) are, for instance, politicians or high-profile members of the business world, the models highlight potential benefits from focusing enforcement policies on these groups: Enforcing a high level of compliance in a narrow group might contribute to a high tax morale among a broader population (Traxler 2010). Hence, there might be a positive spillover from deterrence. One should note, however, that stricter tax enforcement among these specific groups could also backfire. This is the case, since media reporting on detected evaders from moral reference groups could erode the tax morale in the general population.

A further interesting implication for deterrence is contained in models that consider tax morale and stigmatization (e.g., Kim 2003). These studies assume that stigma-related costs are only incurred in case of detection. Given that taxpayers want to avoid stigmatization and any further negative social consequence from being known a detected evader (e.g., ostracism, social exclusion), the fear of stigmatization will increase the overall level of (economic and non-economic) punishment and therefore renders an increase in the auditing frequency more effective. Relating the idea of stigmatization to the previous point, one might argue that the costs from stigmatization crucially depend on how many other taxpayers are detected evading. If stigmatization is more severe in a society where evasion is a rare phenomenon, this means that the deterrent effect from an increase in auditing is larger as compared to a society where evasion is widespread.

The discussion from above shows that tax morale has nontrivial implications for tax enforcement and deterrence. It is therefore worth asking, how important tax morale really is. Unfortunately, the empirical literature provides little causal evidence that allows assessing the role of tax morale. Beyond a huge number of correlational studies, there is only one very recent contribution that documents a causal impact of (survey measures of) tax morale on actual evasion behavior (Halla 2012). In addition, evidence from field experiments that tested approaches which rely on tax morale suggests that it is difficult to effectively apply this type of alternative enforcement strategies. The Minnesota Income Tax Compliance Experiment discussed above, tested – next to an audit threat letter – also different moral appeal letters. Relative to a control treatment, these moral letters turned out to be ineffective (Blumenthal et al. 2001). Similar evidence is provided in the field experiments of Fellner et al. (2013) and Pomeranz (2011). However, Fellner et al. (2013) found some evidence that an information letter, that emphasized the high level of compliance in the population, had a positive effect on compliance (relative to a neutral treatment which did not include this information).

Concluding Discussion and Future Research

This entry provided an overview of different strands of research on deterrence within the domain of tax evasion. Section 2 (Causal Evidence on Deterrence) summarized field studies that randomly varied the auditing risk that is communicated to potential evaders. While some of these studies point to the particularities that distinguish tax evasion from the deterrence of noneconomic crimes (e.g., the evidence on responses from high-income group in Slemrod et al. 2001), the overall picture provides clear support for the deterrence hypothesis: a higher auditing risk reduces evasion. These randomized studies made a major contribution to the empirical analysis of deterrence in the domain of tax evasion, as they established credible, causal evidence on a conceptually important issue (Slemrod and Weber 2012). Nevertheless, there is still a lot left for future research.

First of all, there is little causal evidence on the effectiveness of actual audits. Existing results suggest that there is a substantial effect of specific deterrence (Kleven et al. 2011). Due to the lack of data, however, the robustness and intertemporal persistence of this observation is still unclear. To analyze these points and to better understand the impact of audits, more studies with exogenous variation in auditing are needed.

Second, the empirical literature offers only little evidence on the impact of different sanction levels. Moreover, there is basically no field evidence on the optimal mix of the two policy tools for deterrence – the detection risk and the magnitude of the punishment. One theoretical solution – an enforcement system with infinitely high sanctions and an infinitesimally small auditing risk ("hang taxpayers with zero probability," as suggested in an early contribution), which is cheap in terms of auditing resources – will most likely not be tested empirically. However, one can easily imagine a randomized study that varies both dimensions of the enforcement system – the detection risk and, e.g., the size of fines. Such an experiment would make a huge contribution to the important policy debate on the composition of an optimal enforcement system as well as to the theoretical literature on the trade-off between certainty and severity of sanctions.

Section 3 (Tax Enforcement and Perceptual Deterrence) of this entry discussed a recent literature which links perceptual deterrence research with the empirical analysis of tax enforcement. The first contributions in this domain suggest that there is a clear link between policies and perceptions, as well as perceptions and behavior (Fellner et al. 2013). In addition, the crucial role of communication about own experiences with auditing and detection is highlighted (Rincke and Traxler 2011). In the context of imperfect perceptions about the objective auditing risk, word of mouth about detections crucially shapes the effectiveness of enforcement strategies.

Given the difficulty to obtain precise microdata, this young literature has mainly focused on communication within neighborhoods and spatial distances. It would be important for future research to bring this analysis to further layers of social interaction and to future measures of "proximity" (such as friends in a social network). This would allow assessing, for instance, the role of communication at the workplace or among business partners for spreading information about auditing policies and auditing technologies ("what can easily be detected").

Naturally, this research agenda seems to be linked with the growing empirical literature on networks. Given that communication between taxpayers is crucial for their compliance behavior, there emerges an interesting research question: how should an optimal auditing policy allocate resources between different taxpayers within a (neighborhood, social, workplace) network? Who is the key evader to target? Empirical studies that address these questions would provide important insights for the design of optimal enforcement strategies and would make major contributions to the applied work on networks.

Section 4 (Tax Morale and Third-Party Information) addressed the role of third-party information reporting and tax morale for tax compliance. It is now widely acknowledged that the benchmark model of evasion is tailored to the case of fully self-reported incomes and that the availability of third-party information explains the observed auditing and evasion pattern for different occupational groups and income sources – in particular, self-employed and business versus nonbusiness, labor incomes. A straightforward policy implication is that an extension of the application of third-party information reporting – or more generally: policies that give tax authorities broader access to matchable data – will render a tax enforcement system more efficient. In order to increase compliance, investing in clever information systems therefore appears to be an alternative to investments in stricter auditing.

Finally, several possible implications of tax morale for deterrence were discussed. Field experiments that tested moral appeals or other approaches to enforcement that are rooted in the notion of tax morale are generally found to be ineffective in turning evaders into compliant taxpayers. It would be misleading, however, to interpret this observation as an indicator for tax morale being irrelevant. Based on the available evidence, the question about the importance of tax morale for actual behavior remains unanswered. Again, this is an important issue to address in future research.

Related Entries

- ▶ Econometrics of Crime
- ▶ Certainty, Severity, and Their Deterrent Effects
- ▶ Deterrence: Actual Versus Perceived Risk of Punishment
- ▶ Rational Choice, Deterrence, and Crime: Sociological Contributions

Recommended Reading and References

Allingham M, Sandmo A (1972) Income tax evasion: a theoretical analysis. J Public Econ 1:323–338

Alm J, Jackson B, McKee M (2009) Getting the word out: enforcement information dissemination and compliance behavior. J Public Econ 93:392–402

Andreoni J, Erard B, Feinstein J (1998) Tax compliance. J Econ Lit 36:818–860

Bebchuk L, Kaplow L (1992) Optimal sanctions when individuals are imperfectly informed about the probability of apprehension. J Legal Stud 21:365–370

Becker G (1968) Crime and punishment: an economic analysis. J Polit Econ 78:169–217

Beron KJ, Tauchen HV, Witte AD (1992) The effect of audits and socioeconomic variables on compliance. In: Slemrod J (ed) Why people pay taxes. University of Michigan Press, Ann Arbor, pp 67–89

Blumenthal M, Christian C, Slemrod J (2001) Do normative appeals affect tax compliance? Evidence from a controlled experiment in Minnesota. Natl Tax J 54:125–138

Clotfelter C (1983) Tax evasion and tax rates: an analysis of individual returns. Rev Econ Stat 65(3):363–373

Cowell FA (1990) Cheating the government – the economics of evasion. MIT Press, Cambridge, MA

Dhami S, al-Nowaihi A (2007) Why do people pay taxes: expected utility versus prospect theory. J Econ Behav Organ 64:171–192

Dubin J, Graetz M, Wilde L (1990) The effect of audit rates on the federal individual income tax 1977–1986. Natl Tax J 43(4):395–409

Erard B (1992) The influence of tax audits on reporting behavior. In: Slemrod J (ed) Why people pay taxes. University of Michigan Press, Ann Arbor

Feinstein J (1991) An econometric analysis of income tax evasion and its detection. Rand J Econ 22(1):14–35

Fellner G, Sausgruber R, Traxler C (2013) Testing enforcement strategies in the field: threat, moral appeal and social information. J Eur Econ Assoc 11(3): 634–660

Galbiati R, Zanella G (2012) The tax evasion social multiplier: evidence from Italy. J Public Econ 96:484–495

Halla M (2012) Tax morale and compliance behavior: first evidence on a causal link, B.E. J Econ Anal Policy 12(1)

Kim Y (2003) Income distribution and equilibrium multiplicity in a stigma-based model of tax evasion. J Public Econ 87:1591–1616

Kleven H, Claus K, Emanuel S (2009) Why can modern governments tax so much? An agency model of firms as fiscal intermediaries. NBER Working Paper 15218

Kleven H, Knudsen M, Kreiner CT, Pedersen S, Saez E (2011) Unwilling or unable to cheat? evidence from a tax audit experiment in Denmark. Econometrica 79:651–692

Kube S, Traxler C (2011) The interaction of legal and social norm enforcement. J Public Econ Theory 13:639–660

Long S, Judyth S (1990) Third-party information reporting and compliance. In: Research conference report, how do we affect taxpayer behavior? Internal revenue service document 7302, Washington, DC, pp 3–91

Nagin D (1998) Criminal deterrence research at the outset of the 21st century. Crime Justice 23:1–42

Pomeranz D (2011) No taxation without information: deterrence and self-enforcement in the value added tax. Working Paper, Harvard Business School

Rincke J, Traxler C (2011) Enforcement spillovers. Review of economics and statistics 93:1224–1234

Sah R (1991) Social osmosis and patterns of crime. J Polit Econ 99(6):1272–1295

Slemrod J (2007) Cheating ourselves: the economics of tax evasion. J Econ Perspect 21:25–48

Slemrod J, Shlomo Y (2002) Tax avoidance, evasion, and administration. In: Auerbach AJ, Feldstein M (eds) Handbook of public economics, vol 3, Chapter 22, pp 1423–1470, Elsevier: Amsterdam

Slemrod J, Weber C (2012) Evidence of the invisible: toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. Int Tax Public Finance 19(1):25–53

Slemrod J, Blumenthal M, Christian C (2001) Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota. J Public Econ 79(3):455–483

Torgler B (2007) Tax compliance and tax morale: a theoretical and empirical analysis. Edward Elgar Publishing, Cheltenham

Traxler C (2010) Social norms and conditional cooperative taxpayers. Eur J Polit Econ 26:89–103

Deterrence: Actual Versus Perceived Risk of Punishment

Gary Kleck
College of Criminology and Criminal Justice,
Florida State University, Tallahassee, FL, USA

Overview

Extant evidence indicates that individual perceptions of the certainty, severity, and swiftness of punishment have essentially no correlation with actual levels of those measures of risk prevailing in the area in which the individuals reside. This suggests that public policies that are designed to reduce crime by increasing the deterrent effect of punishment are unlikely to succeed because they are not likely, in general, to increase prospective offenders' perception of the legal risks of committing crime. This does not mean that there are no deterrent effects of the threat of punishment, but only that variations in objective levels of punishment may not affect the magnitude of any deterrent effects that do exist.

Key Issues

There is now an enormous body of scholarly research on the question of whether higher levels of punishment for crime reduce rates of criminal behavior. While there are other mechanisms by which greater punishment levels could reduce crime, such as the incapacitative effects of incarceration or moral reformation, perhaps the most thoroughly studied mechanism is that of deterrence, in which persons otherwise inclined to commit crimes refrain from doing so because they fear the prospect of legal punishment. A person might be deterred from crime as a result of their own personal experiences with punishment, an effect commonly labeled "special deterrence" because it is specific to persons who have been punished. Or a person might refrain from crime as a result of their awareness of others being punished, and thus of the possibility that they might likewise be punished, an effect called "general deterrence," because it can affect anyone in the general population, not just those punished (Zimring and Hawkins 1973, pp. 72–73). For this reason, general deterrent effects are potentially far more potent sources of social control than incapacitative effects or special deterrent effects.

Research on the general deterrent effect of punishment on criminal behavior has largely fallen within two broad categories: (1) macro-level research using official crime statistics to assess the links between objective levels of punishment, such as the ratio of arrests to offenses or the average length of prison sentences, to crime rates, and (2) individual-level "perceptual" research using survey methods to assess the links between perceptions of punishment (e.g., its certainty or severity) and self-reported criminal behavior (Paternoster 1987; Nagin 1998). Many early researchers testing deterrence propositions using macro-level data assumed that any negative associations observed between punishment levels and crime rates were due to deterrent effects. That is, the researchers commonly assumed, but did not demonstrate, links between actual punishment levels and perceptions of punishment.

The second variety of research typically used individual-level survey research to assess the effect of perceptions of punishment on (self-reported) criminal behavior. It addressed the scientific question of whether the former affects the latter, but did not directly address the policy issue that commonly lurks behind the scientific debate: do higher levels of punishment imposed by the criminal justice system produce lower rates of criminal behavior? It is possible that, even if the individual-level research indicates that there is an effect of punishment perceptions on criminal behavior, higher levels of punishment still may not increase the deterrence of criminal behavior because punitive policy efforts fail to intensify perceptions of risk in the first place (Jacob 1979, p. 584; Nagin 1998). Complicating things still further, higher punishment levels may reduce crime rates, but not by increasing general deterrence. For example, crime may be reduced through incapacitation effects, i.e. the physically restraining effects of criminals being incarcerated, or a host of other mechanisms (see Andenaes 1952 for a classic listing). Incapacitation effects could well explain most of the past macro-level findings of an apparent crime-reducing effect of certainty of arrest, conviction, or imprisonment on crime rates.

There is a modest literature on the formation of sanction risk perceptions which has generally focused on the effects of (1) an individual's prior experience of sanctions, such as arrests or incarceration experiences, and of (2) the individual's prior criminal behavior on perceptions of risk. In general, researchers have concluded that the more crimes a person has committed, the more crimes they are likely to have committed without suffering punishment, which reduces perceived risks – a phenomenon described by some as an "experiential effect" (Paternoster et al. 1982). Conversely, others have claimed that people will adjust their estimates of the certainty of punishment upward when they personally experience sanctions such as arrests (Horney and Marshall 1992), though most research does not support this assertion (Pogarsky and Piquero 2003).

Unfortunately, there is little empirical evidence on exactly how individuals go about gathering information on punishment risks. Their perceptions of sanctions risks could be based on the experience of their criminal friends, as Cook (1980) speculated, or on the news media, or even on fictional sources like television and movies. Kleck et al. (2005) found that the frequency with which people watched television news programs was unrelated to their perceptions of risk of punishment, but it may be the specific content rather than frequency of exposure to news that shapes perceptions. Or perhaps there is more of a media impact among more criminally inclined people, for whom news about punishment is likely to be more salient. And if national media are the primary influences, this could explain why there is only a moderate amount of variation in perceptions of legal risk across different areas – everyone is influenced by the same national media messages, rather than varying local realities. Likewise, politicians might influence public perceptions of legal risks when they run on law-and-order themes. Incumbents might boast of the high punishment levels they have helped create, while challengers might reduce the public's perceived risk levels by criticizing incumbents for their supposedly lenient policies.

Unlike incapacitative effects, which occur regardless of how criminals perceive their punishment, deterrence can occur only to the extent that the risk of punishment is perceived by prospective offenders. The dimensions of punishment that influence its deterrent impact on criminal behavior have been summarized in the proposition that "the greater the certainty, severity, and swiftness (celerity) of punishment, the lower the crime rate will be" (e.g., Gibbs 1975, p. 5). However, a more precise restatement of this proposition, which stresses the essential role of perceptions, would be: "The greater the *perceived* certainty, severity, and swiftness of punishment, the lower the crime rate will be."

Of course, the abstract possibility of punishment for crime is perceived by virtually everyone. Some may believe that these risks are low, or that they are themselves unlikely to be caught and punished for any crimes they might commit, but almost everyone recognizes at least the theoretical possibility that they might suffer legal punishment if they violated the law. Nevertheless, this does not imply that increases in actual punishment levels (i.e., increases in actual certainty, severity, or swiftness of legal punishment) will increase deterrent effects and thereby reduce crime, since variations in actual punishment levels may not cause variations in the average perceived level of punishment among prospective offenders. Indeed, critics of deterrence as crime control have long pointed to this reality-perception gap as a key weakness in deterrence-based crime control policies (Zimring and Hawkins 1973, p. 45).

The answer to the simple yes/no question "Can punishment deter crime?" is almost certainly "Yes," since at least a few people almost certainly refrain from some crimes due to a fear of legal punishment. But this is an irrelevant question from a policy standpoint since no policymakers are asking the simple yes/no policy question, "Shall we punish crime?" The more relevant policy question is: "Shall we have, as a means to crime reduction, *more* punishment than we have now?" It is not obvious that more punishment produces a stronger deterrent effect. because it would be uncertain whether increased punishment levels would cause increases in the perceived risk of punishment.

This way of framing the issue is pertinent because much of the debate over crime control in recent decades has been confined to variations on the theme of increasing legal punishments, and some is even more narrowly confined to strategies for increasing the severity and certainty of punishment. Thus, legislators debate bills that would mandate minimum sentences for certain crimes, increased penalties for repeat offenders, enhanced penalties for crimes committed with guns or in connection with drug trafficking, and many other strategies for increasing the severity of punishment. Law enforcement officials lobby for increased budgets and enforcement authority so that they can arrest more criminals, thereby increasing the certainty of arrest for crime. Prosecutors make similar appeals for resources that would enable them to increase the certainty of conviction. And many advocates argue for building more prisons so that both the probability and severity (length) of prison sentences can be increased.

These policy proposals are justified at least partly on the grounds that they will reduce crime through increased deterrence, by "sending a message," "getting the word out" that crime will not tolerated, that criminals will be "taught a lesson" that crime will not be "coddled," and that punishment will surely follow crime. Thus, it is asserted that policy changes producing increases in punishment levels will reduce crime by means of deterrence, that is, by means of an increased perception of legal risk among prospective offenders.

Deterrence theorists have routinely hypothesized about variations in the strength of deterrent effects across individuals, suggesting that various attributes of humans moderate the impact of punishment threats on criminal behavior. For example, legal threats are thought to have stronger effects on persons with greater stakes in conformity and stronger social bonds to conventional others who would ostracize them as a result of legal sanctions being imposed (Zimring and Hawkins 1973, pp. 96–128; Paternoster 1987; Nagin 1998). One might extend this reasoning to the link between actual and perceived punishment levels – perhaps the degree to which the former affects the latter is likewise moderated by some of the same variables. But even though the effect of actual punishment levels on perceived levels might vary across individuals, all theories of general deterrence nevertheless assume that the average effect is a significant positive one.

No matter how inclined and able people may be to rationally process and weigh information, and to consider potential costs and benefits of various courses of action, they cannot actually decide and act rationally unless there is at least some accuracy to their perceptions of those costs and benefits and thus some correspondence between reality and their perceptions of reality. In some realms of human activity, it is perfectly reasonable to assume a fairly close correspondence between perceptions and the realities of costs and benefits. In the sphere of economic behavior, narrowly construed, the assumption is particularly plausible, mainly because there is such a large volume of relevant information available to actors, and a relatively high degree of accuracy to that information. Consumers generally know exactly the price of different brands of goods, while investors know not only the price of a share of stock in any given firm, but also a great deal about the assets, liabilities, and past profit performance of that company. Relevant information in the sphere of market behavior is comparatively voluminous, accurate, and easily obtained. Rational behavior, and predictable responses to changes in costs and benefits, are to be expected in such information-rich environments. Shaped by research experiences in this context, it is not surprising that some economists appear to consider it self-evident that there must be at minimum a significant positive, albeit imperfect, correlation between actual risks and perceived risks (Becker 1968; Ehrlich 1973).

Criminal behavior is quite different, especially with regard to one of the main risks associated with it, punishment. Information about legal risks is limited, often inaccurate, and hard to obtain, so the correlation of actual risks and perceptions of those risks is considerably weaker than in the realm of market behavior. If prospective offenders' perceptions of punishment risk bore no systematic relationship to punishment reality, variations in that reality would have no effect on deterrence of criminal behavior. People might well be deterred by the possibility of punishment, but they would be no more likely to be deterred in settings where risks were higher than in places where they were lower. Under such circumstances, investment in policies increasing punishment levels would be wasted, at least from a deterrence standpoint.

While a positive association between actual and perceived levels of punishment might seem patently obvious to some, there is good reason to question the linkage, and a fair amount of research on related topics that casts doubt on the assumption that the link is strong. Few people, whether criminals or noncriminals, are consumers of criminal justice statistics, and even criminals have only limited personal experience with crime and punishment. Further, depending on hearsay and gossip among their criminal associates may not be a reliable basis for forming even approximately accurate notions of levels or trends in CJS punishment activities.

Similarly, the news media provide neither criminals nor noncriminals with much reliable information on levels of either crime or punishment. At the macro-level, the amount of news coverage of legal *punishment* is unlikely to bear a very strong relation to the general level of actual punishment, since studies of the relationship between the volume of news coverage of *crime* and actual rates of crime find the relationship to be close to nonexistent (Garafalo 1981; Marsh 1989). If common punishment events such as court sentencings or admissions to prison receive even less publicity than the crimes that gave rise to them, it is unlikely that people could formulate even minimally accurate perceptions of punishment risks from news media coverage.

Indeed, various documented news media biases in coverage of crime and punishment could cause, in an irregular fashion, either overstated or understated perceptions of punishment risk. For example, some scholars have found that newspapers exaggerate the certainty of arrest by over reporting solved crimes (Roshier 1973, p. 37; Parker and Grasmick 1979, p. 371). Conversely, studies reviewed by Roberts (1992) indicate that news stories about suspects who "got off on a technicality" or who got a "slap on the wrist" sentence from a judge leads to the public perceiving the severity of sentences to be lower than it really is. Further, since estimates of the certainty of punishment necessarily reflect perceptions of the volume of criminal acts relative to punishments, the lack of correspondence between the volume of news media coverage of crime and actual crime rates may further weaken reality-perception correlations regarding the certainty of punishment.

Personal Experience of Punishment and Perceived Future Risk

On the other hand, already active criminals might draw on their own experiences and those of close associates to formulate their perceptions of punishment risk. To the extent that they accurately stored away these experiences when they occurred, and then accurately retrieved the information later, these experiences could improve both perceptions of past legal risks and forecasts of future risks. The research on personality traits common among known offenders, however, does not encourage a view of criminals as disciplined and careful processors of information, likely to systematically recall and assess such past experiences. Indeed, within a population of persons who already evince tendencies towards risk-taking, past experience of punishment could even lead to a variant of the gambler's fallacy: "My string of past bad luck in getting caught is due to end; my chances of avoiding arrest are bound to improve because I've exhausted my share of bad luck." And of course the information situation is far worse for noncriminals, the people that deterrence-based policies are supposed to keep law-abiding. Noncriminals have no personal experience of criminal behavior leading to either punishment or no punishment, and thus no individual information base at all to use in formulating perceptions of legal risk.

Does the Experience of Punishment Increase Perceived Risk of Punishment? – The Resetting Effect (The Gambler's Fallacy)

The special deterrence hypothesis asserts that the experience of being personally caught and punished for crime should increase the offender's fear of future punishment, which in turn reduces criminal behavior. After all, it seems obvious that being caught doing crime should reinforce the idea that you are likely to be caught again in future if you continue committing crimes. At minimum, punishment experience should increase the punished person's perceived certainty of punishment. Perhaps if people were thoroughly rational, this would be true. If they are not, this proposition is not at all obvious, and a good deal of empirical evidence indicates it is wrong.

A series of careful studies indicates that the experience of being caught causes many offenders to conclude that it is unlikely they would be unlucky enough to get caught the next time they try crime, almost as if they had "used up" their bad luck. Like many gamblers, they irrationally believe that a larger number of unlucky experiences somehow improve one's odds of good outcomes in the future. Consequently, the experience of being caught doing a crime leads them to *reduce* their estimates of the probability of being caught in future (Paternoster and Piquero 1995; Piquero and Paternoster 1998; Piquero and Pogarsky 2002; Pogarsky and Piquero 2003; Pogarsky et al. 2005; Matsueda et al. 2006). More recently, Kleck and Barnes (2013), using survey data from a nationally representative sample of US adults, found that persons who had previously been arrested reported significantly *lower* levels of perceived arrest risk than those who had never been arrested (though these authors could not control for criminal behavior). Other researchers generally find no effect of arrest on perceived certainty of punishment (Piliavin et al. 1986; Pogarsky et al. 2004), or obtain very mixed findings (Wood (2007) found no evidence of an effect of punishment experience on perceived certainty, but did find a positive association of punishment experience with perceived severity). And a few other studies found that punishment experiences do appear to increase perceived certainty, as expected (Paternoster et al. 1985; Horney and Marshall 1992; Heckert and Gondorf 2000; Lochner 2007).

In sum, the full body of evidence does not indicate that experiencing a punishment will increase one's perceived risk of future punishment. The effects of punishment on perceived risk are highly variable, and do not, across the samples of subjects studied to date, produce the results predicted by the special deterrence thesis. Instead, many caught offenders seem to "reset" their estimated probability of arrest back down to some lower level that prevailed prior to capture (Pogarsky and Piquero 2003). This may partly explain why criminogenic effects of legal punishment seem to outweigh the special deterrent and other crime-reducing effects. While personally experiencing punishment may make its unpleasantness more vivid and concrete as Andenaes suggested (this effect has never been systematically studied), its special deterrent effect is nevertheless weak because the experience of punishment does not, on average, increase the punished person's perceived certainty of future punishment.

The Effect of Actual Punishment Rates on Individual Perceived Risk of Punishment

While some self-report studies have treated perceptions of punishment as a dependent variable, only a handful have assessed the impact on those perceptions of actual punishment levels prevailing in the person's area. One early study examined, in a limited way, the association between actual and perceived punishment levels. Erickson and Gibbs (1978) surveyed a random sample of Phoenix residents, asking them to estimate the probability of arrest for ten different offenses. Comparing their collective estimates with police statistics on arrest probabilities, they found a 0.55 Pearson correlation (rho =.39) between objective and perceived certainty of arrest, across ten offenses (p. 259). This study addressed variation in perceptions only across offense types rather than across individuals, and in arrest certainty across offense types, rather than across areas or time periods.

The most extensive investigation of the link between actual punishment levels and individual perceptions of punishment risk was conducted by Kleck and his colleagues (2005), and examined the perceived risks of punishment of persons interviewed in a national telephone survey of a representative sample of urban US adults. Their perceptions of the risk of arrest, conviction, sentencing to prison, average prison length, and swiftness of punishment for four different crime types were compared with the actual risks prevailing in the survey respondents' areas, as measured with official police and court data. The findings indicated that associations between perceived and actual legal risks were not significantly different from zero, for any of the types of risk or offense types.

Thus, there appears to be no significant association between perceptions of punishment levels and the actual levels of punishment that CJS agencies achieve, implying that increased actual punishment levels do not routinely reduce crime through general deterrence mechanisms.

Are Average Collective Perceptions of Legal Punishment Risk More Accurate?

Scholars from a diverse array of disciplines, addressing a wide variety of specific subjects, have argued that there is a "collective wisdom" in large populations that is not evident when one studies only individuals. For example, students of public opinion have theorized that "averaging many individuals' survey responses... tends to cancel out the distorting effects of random errors in the measurement of individuals' opinions" (Page and Shapiro 1992, p. 15), resulting in aggregated public opinions regarding policies and political leadership that are stable, meaningful, and responsive to changes in available information.

Does a sort of "collective wisdom" concerning punishment risks prevail at the aggregate level for local populations that is not evident at the level of individual persons? Perhaps the *average* perception prevailing within a population may correspond closely to punishment realities, even if any one individual substantially over- or underestimates risks. In his review of early deterrence research, economist Philip Cook (1980) acknowledged that many individuals may not react to changes in punishment risks in rational or predictable ways, but argued that some do, and therefore *average* population-wide perceptions may well respond to shifts in actual risks generated by the criminal justice system in a predictable and "rational" way. With respect to the deterrence of robbery, he hypothesized that "an increase in the true effectiveness of the system results in a corresponding increase in the *mean* of robbers' perceptions of effectiveness, and an increase in the number of robbers who are deterred," mainly because "the likelihood that a prospective offender will observe one or more friends apprehended is increased when the overall effectiveness of the system increases" (p. 225). If Cook is correct, increases in actual punishment levels should produce increases in *average* levels of perceived punishment risk, and thereby increase deterrent effects, even if there is little correspondence evident at the individual level between perceptions of system effectiveness and actual effectiveness.

There are strong theoretical reasons to expect that this gap is a large one, and that perceptions of the risk of legal punishment are likely to have at best only a weak relationship with actual risks. First, people have limited capacities, or inclinations, to acquire, retain, and later make use of information, especially with respect to crime and punishment (Simon 1957; Kleck 2003). Defending the deterrence doctrine, Cook (1980) argued that deterrent effects could still exist in the face of limited information about the prospects of punishment, implying that incorporating these limits required only a modest revision of deterrence theory.

Second, people display certain consistent biases in acquiring, retaining, and using information, which tend to weaken the connection between actual contingencies such as legal risks, and peoples' perceptions of them. Tversky and Kahneman (1974; Kahneman and Tversky 1979) demonstrated in a long series of experiments that people faced with the need to decide under conditions of uncertainty make use of "a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations" (1974, p. 1124). These simplified decision rules may often produce reasonably successful outcomes, but they also result in consistent deviations from what utility maximization would predict, yielding decisions that, from the standpoint of orthodox rational choice theory, are "irrational." People deviated from simple rationality, in consistent ways, which the authors called "biases."

For example, people tend to ignore the prior probability of outcomes of a decision, and thus ignore the "base rate" or population-wide probability of it occurring. Regarding the deterrent effects of punishment, this implies that it may, for most people, be irrelevant what the population-wide probability of being caught when committing a crime is, perhaps because people give far more weight to their own experiences than to those of others, even if their own experiences may be very unrepresentative and unlikely to be repeated, and the experiences of a large population would provide a much more reliable basis for forecasting future risks for the individual. If the "base rate" information is irrelevant to decision-making, people are unlikely to acquire and retain such information in the first place.

The only direct test of the collective wisdom hypothesis found no support for it (Kleck and Barnes 2013). Regardless of whether one focuses on perceptions of individual persons, or the average of perceptions among large populations, there is generally no significant association between perceptions of punishment levels and actual levels of punishment. This in turn implies that increases in punishment levels do not routinely reduce crime through general deterrence mechanisms since the fundamental link between actual punishment levels and perceptions of punishment levels appears to be weak to nonexistent. Increases in punishment might reduce crime through incapacitative effects, through the effects of treatment programs linked with punishment, or through other mechanisms, but they are not likely to do so in any way that depends on producing changes in perceptions of risk.

These findings do not, on the other hand, imply that punishment does not exert any general deterrent effect. Rather, they support the view that any deterrent effect, however large or small it may be, does not covary with actual punishment levels to any substantial degree, since the perceptions of risk on which deterrent effects depend generally do not covary with punishment levels, within the range of levels prevailing in contemporary urban America. There may be some baseline level of deterrent effect generated by punishment-generating activities of the criminal justice system, but this level apparently does not consistently increase with increased punishment levels or diminish with decreased punishment levels. Thus, increased punishment levels are not likely to increase deterrent effects, while decreased punishment levels are not likely to decrease deterrent effects.

For those seeking ways to improve crime control, these findings suggest a need for either

(1) a shift in resources towards strategies whose success does not depend on general deterrence effects, or (2) different, nonroutine methods for generating effective deterrence messages. One approach in the latter category is to more narrowly target very specific deterrence messages at audiences who are at especially high risk of committing crimes in the near future. This was the main idea behind a program implemented in Boston and aimed at reducing youth gang violence. Rather than using apprehension, prosecution and punishment to send very broad "wholesale" deterrence messages aimed at the general population, the Ceasefire program delivered direct and explicit "retail deterrence" messages to a relatively small target audience of gang members and potential members. Unfortunately, evaluations of this sort of program are limited (Kennedy 1997).

It is also possible that unusually highly publicized punishment events may generate deterrent effects that the routine, largely unpublicized punitive activities of the criminal justice system do not. For example, it has been asserted that highly publicized executions exert an effect, albeit a possibly temporary one, on homicidal behavior (Phillips 1978; Stack 1987), and the same might be true of less extreme punishments, such as incarceration, if sufficiently publicized. There is, however, no persuasive evidence bearing on publicized punishment events other than executions or death sentencings. Further, there is a severe upper limit on how much publicity-dependent deterrent effects could be increased, since the very newsworthiness that is essential for gaining publicity would, in the absence of direct state control over news media, decline as soon as a given type of punishment event became more common. A few punishment events are highly publicized for a time, and may shift perceptions of legal risk upwards, but such effects are likely to be temporary, lasting only as long as the associated crime is the news media story *du jour*.

Criminals' awareness of legal risks may be largely confined to the most conspicuous features of their immediate environments at the time a crime is contemplated. They are aware of the presence of a police officer, patrol car, or bystander who might intervene or summon the police, but are not sensitive to the overall likelihood of arrest in their areas. These localized perceptions, however, may do little more than displace offenders to other places and times rather than deterring offenses altogether.

One can view prospective offenders' responses to punishment levels as characterized by a severely constricted rationality. While many people are capable of weighing perceived risks and rewards when deciding whether to do crime, they typically possess so little accurate information about key risks and rewards that this capacity for rational decision-making remains to a great extent inoperative.

Open Questions

The research linking perceptions of legal risks to actual risk levels has so far been cross-sectional in nature, and thus has not directly assessed the impact of changes over time in legal risk on perceived risks. Thus, while cross-area differences in objective risk appear not to influence differences in perceived risk across areas, further perceptual research, perhaps using a panel design, is needed to definitively establish whether changes in actual punishment levels produce changes in perceived risk.

It remains possible that prospective offenders are sensitive to changes in legal risks, but only for short periods of time. Thus, deterrent effects of legal threats might be elevated, but for so short a time that these effects are not evident in research using temporal units of analysis as long as years or months. For example, executions may deter homicides in the days just before or after the event, but analysts studying homicide counts for years would miss the effect because the number of homicides deterred would be tiny relative to annual murder counts. Thus, perceptual research examining crime levels in shorter time periods than years or months may prove useful.

Related Entries

- ▶ Certainty, Severity, and Their Deterrent Effects
- ▶ Deterrence of Tax Evasion
- ▶ Deterrent Effect of Imprisonment
- ▶ Focused Deterrence and "Pulling Levers"
- ▶ Offender Decision Making and Behavioral Economics
- ▶ Rational Choice and Prospect Theory
- ▶ Rational Choice Theory
- ▶ Rational Choice, Deterrence, and Crime: Sociological Contributions

Recommended Reading and References

Andenaes J (1952) General prevention: illusion or reality? J Crim Law Criminol Polit Sci 43:176–198

Becker G (1968) Crime and punishment: an economic approach. J Polit Econ 76:169–217

Cook P, Norval M (1980) Research in criminal deterrence: laying the groundwork for the second decade. In: Michael T (ed) Crime and Justice: an annual review of research, vol 2. University of Chicago Press, Chicago

Ehrlich I (1973) Participation in illegitimate activities: a theoretical and empirical investigation. J Polit Econ 81:521–565

Erickson M, Gibbs J (1978) Objective and perceptual properties of legal punishment and the deterrence doctrine. Soc Probl 25:253–264

Garofalo J (1981) Crime and the mass media. J Res Crime Delinq 18:319–350

Gibbs J (1975) Crime, punishment and deterrence. Elsevier, New York

Heckert A, Gondolf E (2000) The effect of perceptions of sanctions on batterer program outcomes. J Res Crime Delinq 37:369–391

Horney J, Marshall IH (1992) Risk perceptions among serious offenders: the role of crime and punishment. Criminology 30:575–592

Jacob H (1979) Rationality and criminality. Soc Sci Q 59:584–585

Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. Econometrica 47:263–291

Kennedy D (1997) Pulling levers: getting deterrence right. Natl Inst Justice J 236:2–8

Kleck G (2003) Constricted rationality and the limits of general deterrence. In: Blomberg T (ed) Punishment and social control: enlarged second edition. Aldine de Gruyter, New York

Kleck G, Barnes JC (2013) Deterrence and macro-level perceptions of punishment risks: is there a "collective wisdom?". Crime Delinq (in press)

Kleck G, Sever B, Li S, Gertz M (2005) The missing link in general deterrence research. Criminology 43: 623–660

Lochner L (2007) Individual perceptions of the criminal justice system. Am Econ Rev 97(1):444–460

Marsh HL (1989) Newspaper crime coverage in the U.S., 1983–1988. Crim Justice Abstr 21:506–514

Matsueda R, Kreager D, Huizinga D (2006) Deterring delinquents: a rational choice model of theft and violence. Am Soc Rev 71:95–122

Nagin D (1998) Criminal deterrence research at the outset of the twenty-first century. In: Michael T (ed) Crime and justice, vol 23. University of Chicago Press, Chicago

Page BI, Shapiro RY (1992) The rational public. University of Chicago Press, Chicago

Parker J, Grasmick H (1979) Linking actual and perceived certainty of punishment: an exploratory study of an untested proposition in deterrence theory. Criminology 17:366–379

Paternoster R (1987) The deterrent effect of the perceived certainty and severity of punishment: a review of the evidence and issues. Justice Q 4:173–215

Paternoster R, Piquero A (1995) Reconceptualizing deterrence: an empirical test of personal and vicarious experiences. J Res Crime Delinq 32:251–286

Paternoster R, Saltzman L, Chiricos T, Waldo G (1982) Perceived risk and deterrence: methodological artifacts in perceptual deterrence research. J Crim Law Criminol 73:1238–1259

Paternoster R, Saltzman L, Chiricos T, Waldo G (1985) Assessments of risk and behavioral experience: an exploratory study of change. Criminology 23:417–436

Phillips DP (1978) The deterrent effect of capital punishment: new evidence on an old controversy. Am J Soc 86:139–148

Piliavin I, Thorton C, Gartner R, Matsueda R (1986) Crime, deterrence, and rational choice. Am Soc Rev 51:101–119

Piquero A, Paternoster R (1998) An application of Stafford and Warr's reconceptualization of deterrence to drinking and driving. J Res Crime Delinq 35:3–39

Piquero A, Pogarsky G (2002) Beyond Stafford and Warr's reconceptualization of deterrence: personal and vicarious experiences, impulsivity, and offending behavior. J Res Crime Delinq 39:153–186

Pogarsky G, Piquero A (2003) Can punishment encourage offending? Investigating the resetting effect. J Res Crime Delinq 40:95–120

Pogarsky G, Piquero AR, Paternoster R (2004) Modeling change in perceptions about sanction threats: the neglected linkage in deterrence theory. J Quant Criminol 20:343–369

Pogarsky G, Kim KD, Paternoster R (2005) Perceptual change in the national youth survey. Justice Q 22:1–29

Roberts JV (1992) Public opinion, crime, and criminal justice. Crime Justice 16:99–180

Roshier B (1973) The selection of crime news by the press. In: Cohen S, Young J (eds) The manufacture of news. Constable, London

Simon H (1957) *Models of Man*. NY: Wiley

Stack S (1987) Publicized executions and homicide, 1950–1980. Am Soc Rev 52:532–540

Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. Science 185:1124–1131

Wood PB (2007) Exploring the positive punishment effect among incarcerated adult offenders. Am J Crim Justice 31:8–22

Zimring FE, Hawkins G (1973) Deterrence: the legal threat in crime control. University of Chicago Press, Chicago

Deterrent Effect of Imprisonment

Roberto Galbiati[1] and Francesco Drago[2]
[1]Department of Economics, CNRS-OSC and Sciences Po, Paris, France
[2]University Federico II Naples and CSEF, Napoli, Italy

Overview

During the last decades, societies have largely used incarceration as a central crime control tool. Between 1970 and 2008, the prison population per 1,000 inhabitants increased by a factor of more than 4.5 in the USA. Despite a dramatic difference in incarceration levels between the two sides of the Atlantic, also in European countries, prison population increased by a factor between two and three over the period 1970–2008 (Buonanno et al. 2011). This massive increase in incarceration had been coupled with a strong debate in social sciences over the magnitude of its impact on crime rates and over the reasons why crime rates might react to changes in prison population. Those favoring an increase in the severity of the criminal justice system have often argued that increasing prison sentences will lead prospective criminals to reduce their criminal activity. Criminals in fact would weight costs and benefits of crime, and an increase in the cost of criminal activities following longer expected prison sentences will induce a subset of them to refrain from engaging into crime. Those holding an opposite view maintained that an increase in incarceration might have a negative net effect. One of the main arguments against an increase in the severity of incarceration system maintains that while imprisonment may temporarily constraint inmates thus preventing them from criminal activity (*incapacitation*), the negative impact of longer or harsher incarceration for those who are incarcerated may enhance subsequent levels of criminal activity.

Which of the two views is more plausible crucially depends on the relative weight of the two effects: the behavioral response of prospective criminals with respect to an increase in expected sentences (general deterrence) and the behavioral consequences of longer or harsher prison sentences for those that experienced incarceration (specific deterrence).

Overall, estimating the relative weight of these two effects is fundamental to the understanding of whether the actual incarceration rates are at the optimal level. It is clear that prison sentences do not deter all the individuals from committing crime (otherwise prisons would be empty), but only a subset of all potential criminals. Therefore, an increase in the severity of prison sentences increases the incarceration rates. An incarceration rate is set at the optimal level if the social cost of incarcerating an additional individual (at that incarceration rate) is not below the social benefit deriving from the prevention of an eventual crime committed by that individual. While it is very difficult to compute the cost and the benefit from incarcerating an additional individual, it is very important to understand how potential criminals respond to prison sentence severity and the criminogenic effect of prisons. Both these effects enter the cost and the benefit from increasing the severity of prison sentences and incarceration rates.

Fundamentals of Deterrence Effects of Imprisonment

Definitions

In order to better understand what is the meaning of deterrence effects of imprisonment, it is useful to refer to the pioneer economic model of crime Gary Becker (1968). Becker's contribution is crucial because it provides a simple theoretical framework presenting the commission of a crime as a rational choice reflecting a cost-benefit comparison. In this simple framework, an individual chooses to commit a crime if the expected payoffs (weighted by a utility function) from the commission of a crime exceeds the payoffs from not committing a crime. The expected payoff from committing a crime depends on the prison sentence from that crime, the probability to be caught and convicted to a prison sentence, and the gain deriving from that crime. The expected payoff from not committing a crime depends on the payoff from legal activities (e.g., expected wage from a legal job, disutility from working). An individual compares the expected costs and benefits from committing a crime and behaves accordingly. If the potential criminal opts for the criminal activity, we infer that he faces a net expected gain from the activity despite the expected punishment. This simple reasoning highlights that the probability of being punished and the costs associated with being punished are crucial elements for a criminal's choice. Rephrasing this point, the rational choice theory predicts that changes in the certainty and the severity of the punishment will induce changes in criminals' behavior.

The probability of punishment will depend upon many factors (number of police forces, efficiency of the justice system, and apprehension technology among others). For what concerns the severity of the punishment, two elements are crucial. One is length of the expected sanction and in particular the expected length of imprisonment. The other fundamental aspect is given by the conditions of detention. Both longer and harsher expected prison sentences represent a higher cost for prospective criminals. According to this framework, by varying these three elements, a policy maker can affect criminals' choice and thus can affect the level of crime rates.

The *general deterrent effect of imprisonment* is the response of potential criminals to the expected cost associated with the decision to commit a crime. This is an ex ante perspective. The choice of committing a crime is affected by the expected level of punishment that in turn reflects the probability of apprehension and conviction, the expected length of imprisonment, and the expected harshness of the imprisonment.

Since a share of those that are apprehended and convicted will actually spend some time in prison, it is natural asking if the time spent in prison and the condition of detention will affect criminals' propensity to commit a crime once they are released. If we assume that incarceration leads criminals to update their beliefs about the consequences of punishment, we might expect that having experienced harsher punishment should reduce the inmates' propensity to recommit a crime once released. Under this perspective, imprisonment is an experience good. Criminals ex ante do not perfectly know the cost of spending time in. Once they are convicted to a prison sentence and they make the experience of incarceration, ex post, they update their knowledge of prison conditions. Experiencing imprisonment and staying in prison longer will affect criminals' knowledge of the prison environment and in turn their evaluation of expected costs of committing a crime once they are released. For this reason, those having experienced harsher prison conditions should be less prone to recommit a criminal act. This is known as the *specific deterrence hypothesis*. On the other hand, however, experiencing harsher prison conditions could also imply higher propensity of committing crime once released. In fact, harsher prison conditions may lead to a higher human capital deployment and worse labor outcomes for former inmates (Waldfogel 1994). Since these factors affect the relative benefits of engaging into legal or criminal activities, they will affect former inmates' criminal choice. In this case, the theoretical predictions are not clear-cut. Hence, understanding the *specific deterrence effects of imprisonment* is ultimately an empirical question.

Current Issues and Controversies

Understanding how and whether potential criminals respond to changes in expected prison sentences (general deterrence) and if prison experience affects former inmates' decisions to recommit a criminal act after release (specific deterrence) has been the object of a large and growing empirical literature. Most of the studies have focused on the American criminal justice system, but in recent years several studies have provided evidence on European ones. Many of these studies apply modern econometric techniques allowing to get rid of potential confounding factors that could affect the interpretation of the findings. These exercises thus provide important insights that could usefully inspire crime control policy interventions. Despite these relevant advances, international comparisons are still very difficult given the large diversity of the specific national contests.

Empirical studies aiming at understanding to what extent potential criminals respond to manipulations in expected prison sentences and/or to what extent the prison conditions experienced during incarceration by former inmates affect their post-release behavior need to face some major empirical challenge.

For what concerns *general deterrence*, longer prison sentences might deter potential criminals but might also reflect changes in the general attitude toward criminality or might anticipate expected increases in crime rates. For this reason, it is hard to disentangle if there is any causal effect of increasing expected prison sentences on criminal behavior. The same reasoning applies to the general deterrent effects of expected harshness of prison conditions. For example, overcrowded prisons might affect potential criminals' choices because more crowded prisons imply higher expected costs of detention but at the same time will mechanically reflect an increase in crime rates. Pinning down any *general deterrence* effect of harsher prison conditions is thus particularly challenging.

Understanding the *specific deterrent effect of prison* suffers from the same empirical challenges. Inmates suffering harsher prison conditions will probably react to them once they will be released, but being assigned to harsher prisons might reflect higher individuals' dangerousness. Thus, even in the presence of detailed individual level data, it is hard to understand if and to what extent prison conditions have any causal effect on former inmates' behavior without some clear empirical research design able to break the simultaneity between prison conditions and individuals' propensity to reoffend.

For these reasons, most of the recent research has focused on specific case studies allowing to quasi-experimentally disentangle the effect of the crucial policy manipulable elements on individuals' choices.

In particular researchers have investigated the general and specific deterrent effects of prison sentences' length and of prison conditions. Related to this last point, some interesting studies have focused on an important but largely neglected aspect of incarceration: the effect of peers' characteristics and behavior on inmates' post-release behavior. This aspect is crucial since there is a large and consistent evidence that peers' might affect people behavior in many domains, ranging from educational attainment to job search. If the same holds for crime, prisons might be a sort of criminal school where inmates learn or are influenced from each other. Thus, other inmates' characteristics or behavior might crucially affect ones' post-release behavior.

General Deterrence

Prison Sentences Length and Prison Conditions

Evidence on the effects of the severity of punishment on criminal activity is growing and consistently point out that, as predicted in the standard economic/rational choice approach to crime, potential criminals take into expected sentences' length when they decide whether to commit a criminal act. To understand this effect, most works in this field have studied the effect of incarceration rates on aggregate crime rates. In order to break the simultaneity between prison population and crime rates, researchers have resorted to case studies exploiting exogenous variation in prison population or expected prison sentences' length. An influential study based on US State level data, Levitt (1996), shows that releasing one prisoner is associated with an increase of 15 crimes per year. This estimate, however, includes deterrence and incapacitative effects. A major change in expected prison sentences in the USA have been introduced in some states through sentence enhancements' laws such as California's "three strikes and your out" law or Proposition 8. These laws provide researchers with some variation on expected prison sentences' length useful to disentangle deterrence and incapacitation. Among these studies, Kessler and Levitt (1999) exploiting sentence enhancements targeting the most frequent and dangerous criminals find that some crime rates fell by 4 % after sentence enhancement, which, for example, increased the sentence for any "serious" felony offender by 5 years. Exploiting the increase in expected sentences induced by the California "three strikes law," Helland and Tabarrok (2007) compare the future offending of individuals convicted of two previous strikeable offenses with that of individuals who had been convicted of only one strikeable offense but who, in addition, had been tried for a second strikeable offense but were ultimately convicted of a nonstrikeable offense. The study finds that arrest rates were about 20 % lower for the group with convictions for two strikeable offenses providing evidence in line with previous studies.

Since the cost of imprisonment does not depend only on the expected prison sentence length but also on expected prison conditions, an important issue is to understand if and how potential criminals respond to expected harsher prison conditions. This is an interesting aspect since, setting aside ethical considerations, prison conditions are more easily manipulable than sentences' length. Such a deterrent effect of harsher prison conditions seems to exist in the USA where recent research (Katz et al. 2003), using death rates among prisoners as a proxy for prison conditions, shows that more punitive facilities have a small but statistically significant general deterrent effect.

These and several other studies taking seriously the econometric challenges behind the identification of the deterrent effect of prison sentences' length show that, at least for USA, potential criminals take into account expected sentence length and prison severity when they have to decide whether to engage into a criminal act.

Evidence outside the USA is less systematic, but a recent large criminal justice policy intervention in Italy has provided researchers the opportunity to study for a large sample of former inmates what is their response to a manipulation of expected prison sentences.

The Collective Clemency Bill passed by the Italian Parliament in July 2006 provided an ideal case study to understand how people respond to exogenous variations in prison sentences. This law provided for an immediate 3-year reduction in detention for all inmates who had committed a crime before May 2, 2006. Upon the approval of the bill, almost 22,000 inmates – about 40 % of the prison population of Italy – were released from Italian prisons on August 1, 2006. The bill states that if a former inmate recommits a crime within 5 years following his release from prison, he will be required to serve the remaining sentence suspended by the pardon (varying between 1 and 36 months) in addition to the sentence given for the new crime. This is equivalent to a policy manipulating incentives to commit a crime since it commutes 1 month of time of the original sentence to be served in 1 month more of expected sentence for future crimes. More importantly, this institutional framework manipulates prison sentences at the individual level in a random fashion. In particular, conditional on inmates' original sentences, the variation in the remaining sentence at the date of the pardon (and hence in the expected sentence for any crime) depends only on the date of an inmate's entry into prison, which is plausibly non correlated with individuals' characteristics potentially affecting their propensity to recidivate. The variation in the remaining sentence at the date of the pardon can thus be used to identify the causal impact of a manipulation of expected prison sentences on individuals' propensity to recommit a crime. Research based on data from the Italian experiment (Drago et al. 2009) shows that a marginal increase in the remaining sentence reduces the probability of recidivism by 0.16 percentage points (1.3 %). This means that for former inmates, 1 month less time served in prison commuted into 1 month more in expected sentence significantly reduces their propensity to recommit a crime. Interestingly, this deterrent effect of prison sentences is quite homogeneous across inmates with different individual characteristics, but individuals convicted to relatively longer sentences do not seem to be deterred. This suggests that longer sentences ex ante deter more but experiencing longer time in prison might be deleterious in terms of former inmates' behavioral response to incentives.

Given this apparently consistent evidence suggesting that longer and harsher prison sentences deter crime, it could be tempting to draw immediate policy conclusions. However, it is extremely important to be cautious in making this further step. In fact it is still far to be clear for researchers what are the most cost-effective means to reduce criminal activity, and even just looking at deterrence, it is far from clear if criminals respond more to an increase in the severity rather than an increase in the certainty of punishment (Durlauf and Nagin, 2011). For example, in a pioneering study, DiTella and Shargrodsky (2009) show that electronic monitoring is more effective than incarceration in reducing ex post former prison inmates' criminal behavior. Moreover, existing evidence suggests that the effect of the aggregate impact of changing prison conditions on crime rates is relatively small (Katz et al. 2003). Finally, the evidence from the Italian case suggesting that those having spent long prison sentences do not respond to a manipulation in prison sentence after release suggests that the ex post effects of prison sentences might compromise the ex ante deterrent effect.

Since most criminals experiencing prison sentences experience multiple prison conditions during their life and constitute the core of the criminals' population, for the reasons explained above, understanding specific deterrence and the effect of incarceration on those that have experienced it is a crucial challenge for social scientists.

Specific Deterrence

Prison Conditions and Time Served

Opening the prison black box, we find very different punitive situations in terms of overcrowding, health services, social activities for inmates, and so on. All these elements might impact former inmates' propensity to reoffend. Those inmates having experienced harsher prison conditions might be more deterred by the threat of future sentences, but these same people might suffer from more human capital deployment and worse health conditions thus facing lower opportunity costs of criminal activity of engaging. While the issue of the deterrent effects of prison treatment appears particularly important for both researchers and policy makers, the empirical evidence is scarce. Only a few recent works analyze the effects of prison conditions on criminal behavior. The lack of evidence is mainly due both to the difficulty in obtaining access to reliable data on prison conditions and to the identification challenges faced by researchers interested in understanding how criminals respond to prison conditions. Nonetheless, the few existing studies focusing on the issue of prison conditions and former inmates behavior provide some crucial knowledge on the phenomenon. Focusing on women incarceration conditions, recent evidence exploiting the expansion of female penal system capacity in the United States (Bedard and Helland 2004) shows that, on average, increasing the distance of detention facilities from inmates' home tends to lower the female crime rate. Despite this results seem to conform to the deterrence hypothesis; from this case study, it is hard to understand if the deterrent effect is driven by the response of formers inmates or by the reaction of criminals who had never received a prison treatment (or both). Exploiting quasi-experimental designs and individual level data is crucial to separate general deterrence from specific deterrent effects of incarceration conditions. Studies using these two ingredients cast doubts on the capacity of harsher prison conditions to reduce former inmates' propensity to recidivate. By exploiting a discontinuity in the assignment of federal prisoners to security levels, American economists provide evidence that serving a sentence in a higher security prison implies a higher post-release propensity to commit a crime (Chen and Shapiro, 2007). In a similar vein, exploiting some randomness in the facility assignment rules and individual level data on inmates' post-release behavior, an Italian case study (Drago et al. 2011) shows that being assigned to a prison where mortality is higher increases the propensity to recidivate for former inmates.

For the analysis of the effect of the amount of time spent in prison, recent quasi-experimental evidence from the USA and France (Kuziemko, 2007 and Maurin and Ouss 2009) suggests that spending more time in prison all else equal should reduce the propensity to recidivate for former inmates. Taken together the results of recent empirical literature focusing on the specific deterrent effects of prison in terms of prison sentences' length and detention conditions quality are mixed. Thus, deriving any strong policy conclusion based on the current evidence is not only hard but also not recommended.

Peer Effects Between Prison Inmates

Experiencing incarceration might affect post-release inmates' behavior not only through the effects of time spent in physical prison conditions or in prison activities but also through the effect of interactions with other inmates. Social scientists have documented that the exposure to peers' actions and characteristics might have a strong impact on individuals' behavior in various domains ranging from school performance to labor supply decisions. Peers might influence one's decision by providing information and by affecting social norms, constraints to action, and various other channels. Individual choices to engage into criminal behavior are also plausibly affected by peers' influence since peers affect the criminal markets' conditions, provide information about criminal opportunities, and act together. Estimating the impact of peer effects on criminal behavior and understanding the mechanisms through which peers affect individual criminal participation are of primary importance in the design of effective policies to prevent crime (Manski, 1993; Glaeser et al. 1996). Understanding if peer effects between prison inmates exist and estimating their magnitude is crucial to understand how prison experience affects former inmates' behavior. In this sense potential peer effects are a crucial component of prison experience, and understanding them is particularly important to assess overall general *and specific deterrent effects of prison* and thus to understand how the prison system works overall.

Peers effects between former prison inmates might take place because once in prison criminals respond to the characteristics of those they spend time with. For example, juvenile offenders serving time in the same correctional facility in Florida seem to have some influence on each other's subsequent criminal behavior (Bayer et al. 2009). The data on juvenile inmates spending some time in 1 of 169 Florida correctional facilities provide a complete record of past crimes, facility assignments, and arrests and adjudications in the year following release for each individual. These kind of data have allowed researcher to solve the simultaneity problem that usually hamper the identification of peer effects. From the Florida case study, we learned that peers exert their influence by influencing individuals who already have some experience in a particular crime category (Bayer et al. 2009). This evidence supports the idea that prisons are a sort of criminals' school.

Another plausible mechanism underlying peer effects is the maintenance of prison peer groups after release and the presence of complementarities in post-release behavior (e.g., joint crime production). Sociological and qualitative research on prison gangs (Skarbek 2010; Leeson and Skarbeck 2010; Fleisher and Decker 2001) and on former Italian inmates' post-release networks (Baccaro and Mosconi 2004; Santoro and Tucci 2006) supports this interpretation. Moreover, exploiting data on the Italian prison experiment described above, researchers have shown that former inmates from the same nationality that spent time together in the same prison facility tend to influence each others' post-release behavior (Drago and Galbiati, 2012). In particular, criminals respond to the behavior of those they spent time in prison with by increasing their propensity to commit crime if others' criminal activity increases. This suggests that peer groups formed in prison remain the same after release, and thus peers will continue to influence each other even after release from prison.

Thus, peer effects are a crucial aspect of inmates' interaction that have to be carefully taken into account in order to understand the overall effect exerted by prison sentencing on individuals' criminal choices.

Directions for Future Research

Despite a growing research effort, clean evidence about general and specific deterrent effects of prison is still scarce. Some more systematic research and a larger number of specific country case studies are most welcome to fill the knowledge gap.

Given that identifying causal relation is particularly difficult in this domain, a more systematic access and use of individual level data is needed. Some particularly fruitful lines of research should focus on understanding the relative weight of severity and certainty of punishment in determining the overall general deterrent effect of prison sentences. Continuing in the effort to open the specific deterrence black box is crucial in order to understand the dynamics of future crime rates since a large share of those that have been incarcerated in the USA and Europe during the last decades will be released in the following years.

Finally these research building blocks could and hopefully will be used to build a more comprehensive approach to determine the immediate and dynamic future effects of manipulating the length and harshness of prison sentences.

Related Entries

- ▶ Certainty, Severity, and Their Deterrent Effects
- ▶ Deterrence of Tax Evasion
- ▶ Econometrics of Crime
- ▶ Rational Choice, Deterrence, and Crime: Sociological Contributions

Recommended Reading and References

Baccaro L, Mosconi G (2004) Il girone dei dannati: ovvero il fenomeno della recidiva. Rass Penit e Criminol 2:212–237

Bayer P, Hjalmarsson R, Pozen D (2009) Building criminal capital behind bars: peer effects in juvenile corrections. Q J Econ 124(1):105–147

Becker GS (1968) Crime and punishment: an economic approach. J Polit Econ 76:169–217

Becker G (1978) The economic approach to human behavior. University of Chicago Press, Chicago, Chapter 1

Bedard K, Helland E (2004) The location of women's prisons and the deterrence effect of harder time. Int Rev Law Econ 24:147–167

Buonanno P, Drago F, Galbiati R and Zanella G (2011) Crime in europe and US: dissecting the reversal of misfortunes. Econ Policy

Chen MK, Shapiro JM (2007) Do harsher prison conditions reduce recidivism? a discontinuity-based approach. Am Law Econ Rev 9(1):1–29

Di Tella, Rafael, Ernesto Schargrodsky. (2009). Criminal recidivism after prison and electronic monitoring. National Bureau of Economic Research Working Paper 15602

Drago F, Galbiati R (2012) Indirect effects of a policy altering criminal behavior: evidence from the italian prison experiment. Am Econ J Appl Econo 2

Drago F, Galbiati R, Vertova P (2009) The deterrent effects of prison: evidence from a natural experiment. J Polit Econ 117(2):257–280

Drago F, Galbiati R, Vertova P (2011) Prison conditions and recidivism. Am Law Econ Rev 13(1):103–130

Durlauf S and Nagin D (2011). The deterrent effect of imprisonment. In Cook P.J, Ludwig J and McCrary J eds Controlling crime: strategies and tradeoffs. NBER

Fleisher MS, Decker SH (2001) An overview of the challenge of prison gangs. Correct Manag Q 5(1):1–9

Glaeser EL, Sacerdote B, Scheinkman JA (1996) Crime and social interactions. Q J Econ 111(2):507–548

Helland E, Tabarrok A (2007) Does three strikes deter? a nonparametric estimation. J Hum Resour 22:309–330

Katz L, Levitt SD, Shustorovich E (2003) Prison conditions, capital punishment, and deterrence. Am Law Econ Rev 5:318–343

Kuziemko Ilyana (2007) Going off parole: how the elimination of discretionary prison release affects the social cost of crime Working Paper no. 13380 (September), NBER, Cambridge, MA

Leeson PT, Skarbek DB (2010) Criminal constitutions. Glob Crime 11(3):279–297

Levitt SD (1996) The effect of prison population size on crime rates: evidence from prison overcrowding litigation. Q J Econ 111(2):319–351

Levitt S, Miles T (2007) Empirical study of criminal punishment. In: Mitchell Polinsky A, Shavell S (eds) Handbook of law and economics, vol 1. North-Holland, Amsterdam

Manski CF (1993) Identification of endogenous social effects: the reflection problem. Rev Econ Stud 60(3): 531–542

Maurin Eric Aurelie Ouss. (2009) Sentence reductions and recidivism: lessons from the bastille day quasi experiment IZA Discussion Paper no. 3990 (February), Inst. Study Labor, Bonn

Santoro E, Tucci R (2006) L'incidenza dell'affidamento sulla recidiva: prime indicazioni e problemi per una ricerca sistematica. Rass Penit e Criminol 10(1):79–158

Skarbek D (2010) Putting the ‘con’ into constitutions: the economics of prison gangs. J Law, Econ Organ 26(2):183–211

Waldfogel J (1994) The effect of criminal conviction on income and the trust reposed on workmen. J Hum Res 29:62–81

Development of International Criminal Law and Tribunals

Kai Ambos
Department of Foreign and International Criminal Law, Institute for Criminal Law and Justice, Georg-August-Universität Göttingen, Göttingen, Germany

Overview

This entry offers a brief reconstruction of the development of International Criminal Law (ICL) and an equally brief description of the existing International Criminal Tribunals and Courts. The former part will be divided into two sections: starting with the Versailles Peace Treaty and historical precedents, the Nuremberg and Tokyo trials are looked at as the first stage of the development, focusing on international conflicts. The establishment of the UN ad hoc Tribunals in the 1990s accompanied by the development of ICL for non-international conflicts constitutes the second stage. The third stage, i.e., the response of ICL to transnational terrorism after 11 September 2001, will not be covered by this outline. The description of the existing tribunals starts with the UN ad hoc Tribunals for the former Yugoslavia and Rwanda, continues with a more detailed characterization of the International Criminal Court, and finishes with the so-called mixed or hybrid tribunals.

The Versailles Peace Treaty and Historical Precedents

Crimes against the basic principles of humanity are nothing new to the history of mankind. The crusades of the eleventh century may be considered as early forms of genocide. Other examples of international crimes are the Spanish and Portuguese Conquista of the Americas accompanied by the extermination of great numbers of the native population, the massacre of thousands of the French Huguenots during St. Bartholomew’s night 23 August 1572, and the massacre of Glen Coe in the year 1692. In all these cases, investigations never took place, and criminal sentences were never passed on the responsible persons except in the case of Peter von Hagenbach in the year 1474 (Cryer 2005, pp. 17 ff.).

At the beginning of the nineteenth century, the punishability of **piracy** was acknowledged under customary international law. Also, **slavery** was declared a crime of international concern due to numerous international treaties, which had been concluded since 1815 (Bassiouni 1999, pp. 305 ff.). Notwithstanding, the proposal of the president of the International Committee of the Red Cross (ICRC), Gustave Moynier, to set up an International Criminal Court (ICC) after the German-French War in 1870/1871 remained without any political resonance (König 2003, p. 60).

When the Allied and Associated Powers convened the 1919 Preliminary Peace Conference, the first international investigative commission was established, and the **Versailles Peace Treaty** was adopted. This treaty established a new policy of prosecuting war criminals of the vanquished aggressor state after the end of the hostilities. The legal basis of that policy was laid down in 1919 in the Paris Peace Treaties which created four groups of offences: crimes against the sanctity of the treaties, crimes against the international morals (which were not defined more precisely; Puttkamer 1949, p. 424), war crimes “in a narrow sense” (i.e., “violation of the laws and customs of war” according to Art. 228 of the Versailles treaty), and violations of the laws of humanity. The latter was not included in the Versailles treaty since the USA took the view that it could not be sufficiently precisely defined and thus was too vague as a basis for prosecutions (Bassiouni 1999, p. 65). At the same time, with the Versailles treaty, the individual criminal responsibility for crimes against international law was for the first time recognized on a treaty basis. It was further recognized that such responsibility had no limits of rank or position.

In 1920, the Allied Powers decided to hand over the prosecution to Germany, which passed new legislation to be able to prosecute German suspects before its own Supreme Court (the *Reichsgericht*), sitting at **Leipzig**. However, only 12 Germans were prosecuted for war crimes (Ahlbrecht 1999, pp. 42 ff.). Thus, the so-called Leipzig Trials have been considered a failure and are widely cited as proof of the German unwillingness to seriously prosecute their own war criminals.

In international law, the term "**crimes against humanity**" was for the first time used in the context of the **genocide of the Armenians**, which led to a joint declaration of France, Great Britain, and Russia on 24 May 1915, asserting that all members of the Ottoman Government and those of its agents found to be involved in those massacres would be held personally responsible for the crimes. Prosecuted on the basis of the Turkish penal code, several ministers of the wartime cabinet and leaders of the Ittihad party were found guilty by a court martial of "the organization and execution of crime of massacre" (Schabas 2009, p. 25). At the international level, the Peace **Treaty of Sèvres**, signed on 10 August 1920 between the Ottoman Empire and the Allies (France, Italy, Japan, UK), in many aspects similar to the Treaty of Versailles, contained, as a major innovation, offences which were later qualified as crimes against humanity (Article 230). The treaty, however, never took effect. It was replaced by the Treaty of Lausanne of 24 July 1923, which included a "Declaration of Amnesty" for all offences committed between 1 August 1914 and 20 November 1922.

The First Ad Hoc Tribunals: Nuremberg and Tokyo

During World War II, the prosecution of "war crimes" became a primary objective. In 1942, the Allied Powers signed a declaration in St. James Palace in London, which established the **UN War Crimes Commission** (UNWCC). We will return to it below. The "Declaration of St. James" also laid down the foundation for the International Military Tribunal (IMT). This was followed by the Moscow Declaration of 30 October 1943, which confirmed the Allied quest for prosecution. Finally, the "Declaration of London" of 8 August 1945 – concluded by the governments of Great Britain, USA, France, and the Soviet Union – gave birth to the IMT.

The first series of World War II trials, the **Nuremberg trials**, took place under the terms of a charter drafted in London between June and August 1945 by representatives of the USA, Great Britain, Soviet Union, and France. It was therefore called the "London" or "Nuremberg Charter". The Nuremberg Charter contained three categories of offences: crimes against peace, war crimes, and crimes against humanity. As to the defenses, Article 7 rejected official position and Article 8 superior orders as grounds for excluding responsibility. The Allies set up the IMT to prosecute the "Major War Criminals". Twenty-three defendants were initially charged and 19 convicted (Engelhart 2004, p. 734 ff.).

The **Tokyo Trials** were based on the charter for the Far East, or *Tokyo Charter*, which was proclaimed on 19 January 1946. The charter was, unlike the London Charter, not part of a treaty or an agreement among the Allies. Representatives of the Allied nations, which had been involved in the struggle in Asia (the USA, Great Britain, France, Soviet Union, Australia, Canada, China, the Netherlands, New Zealand, India, and the Philippines), created the IMT for the Far East (IMTFE). It was composed of judges, prosecutors, and other staff from the Allied nations. The IMTFE recognized the same offences as the IMT: crimes against peace (as defined in the London Charter); "conventional war crimes, namely, violations of the laws or customs of war"; and crimes against humanity. The definition of crimes against humanity differed from that of the IMT Charter in two ways: first, the IMTFE Charter expanded the list of crimes to include imprisonment, torture, and rape. Second, it eliminated the requirement that "crimes against humanity" had to be connected to war. As to possible defenses, the charter excluded – as did the IMT Charter – official position or superior orders. The prosecution selected 28 defendants; 25 were convicted (Osten 2003).

Post Nuremberg World War II Trials

The Nuremberg and Tokyo Trials were followed by a second series of prosecutions of Nazi leaders, pursuant to **Control Council Law** (CCL) No. 10. The most famous proceedings were the 12 trials before the US-American court in Nuremberg. Other important cases have been documented by the United Nations War Crimes Commission (UNWCC 1947–1949). It was formally established on 20 October 1943, and its task was basically twofold: on the one hand, to investigate war crimes, collect evidence, and identify those responsible and, on the other, to inform the Allied governments about the cases providing a sufficient basis for prosecution. In total, the UNWCC documented 89 war crimes trials on the basis of protocols of 2111 proceedings (Ambos 2002/2004, p. 140).

The proceedings instituted by the occupation powers ended a few years after the end of the war. However, the prosecutions of Nazi criminals have continued in and outside Germany until today. The most famous cases on the basis of universal jurisdiction were the trials against *Adolf Eichmann* (ILR 1968, pp. 5–14) and *Klaus Barbie* (Le Monde 5–6 July 1987, p. 1). The former was sentenced to death by the Jerusalem District Court on 15 December 1961, and the latter was first tried in absentia for war crimes and sentenced to death by the *Tribunal Permanent des Forces Armées de Lyon* in two judgements and later *in presence* sentenced to life imprisonment on 4 July 1987 for crimes against humanity. After having served 4 years of his sentence, Barbie died of leukemia in 1991. Other cases include that of *Paul Touvier* (ILR 1995, 338 ff., 357 ff.) in France, sentenced to life imprisonment before a *Cour d'Appel de Paris* in Versailles, France (20 April 1994), and that of *Imre Finta* (ILR 1995, 520 ff.) in Canada, finally acquitted by the Supreme Court (24 March 1994). Last but not least, *John Demjanjuk* was sentenced to 5 years imprisonment on 15 May 2011 but released pending an appeal; he died on 17 March 2012.

The principles resulting from the practical experience of the IMT were an important substructure for the upcoming development of ICL. The International Law Commission (ILC), founded in 1947 upon the recommendation of the Committee on the Progressive Development of International Law and its Codification (CPDIL), adopted seven principles on its second session in 1950 (ILC 1950, pp. 374 ff.). Those principles in conjunction with the Nuremberg Charter, the CCL 10, and the adjudication of the Nuremberg courts are called the "**Nuremberg Principles.**" They comprise rules on the general part (Principles I–IV, VI, and VII), on international crimes (VI), and on a procedural "fair-trial" norm (V).

The Development of International Criminal Law Prior to the Establishment of the UN Ad Hoc Tribunals

The Genocide Convention

Based on thoughts by Rafael Lemkin (1933, p. 117), Resolution 96 was adopted by the UN General Assembly on 11 December 1946. It declared genocide to be a crime of international concern and formed the basis for the drafting of a treaty by a group of experts. The Genocide Convention was adopted by the GA on 9 December 1948 and came into force on 12 January 1951. It is the most vital legal instrument on the crime of genocide (Schabas 2009, pp. 3 ff.).

The Hague and Geneva Law

First efforts to establish a "law of war" can be traced back to the middle of the nineteenth century focusing primarily on the humanization of war, first with regard to the admissible means and methods of warfare (so-called **Hague Law**) and then later increasingly with regard to the protection of the victims of armed conflict (so-called **Geneva Law**). While the Hague and Geneva laws regulate the situation of an armed conflict, i.e., the *ius in bello*, the law governing the resort to force is called *ius ad bellum* (MacCoubrey and White 1992, p. 217).

The **Hague Law** was developed in two Hague Peace Conferences of 1899 and 1907. Since then, it has been amended and updated by various conventions, including by the 1977 First Protocol Additional (PA) to the 1949 Geneva Conventions (GC), and contains some "Hague elements." In substance, the Hague Conventions provide for three important principles still valid until today (Bailey 1972, p. 63): first, the *Martens Clause* according to which, notwithstanding the absence of specific regulations, "populations and belligerents remain under the protection and empire of the principles of international law, as they result from the usages established between civilized nations, from the laws of humanity, and the requirements of the public conscience" (cf. Preamble of the 1907 Hague Convention (IV)); second, the right to injure "the enemy is not unlimited" (Article 22 of the annexed Regulations of the 1899 and 1907 Hague Conventions II and IV); and third, the prohibition "[t]o employ arms, projectiles, or material of a nature to cause superfluous injury" (Art. 23 (e) 1899 Convention) or "calculated to cause unnecessary suffering" (Art. 23 (e) 1907 Convention).

The **Geneva Law** emerged from the Geneva Conventions of 1864, 1906, 1929, and 1949. It deals with the protection of noncombatants (civilians) and former combatants who are no longer willing or able to fight. The (modern) Geneva Law consists of the *four Geneva Conventions* (GC I–IV) of 12 August 1949 and the *three Additional Protocols* (AP I, II, III) of 18 June 1977 (AP I and II) and 8 December 2005 (AP III), respectively. This body of law constitutes the modern International Humanitarian Law (IHL) (Sassoli and Bouvier 2006, Part I, Ch. 3, pp. 121 ff.). Penal provisions can be found in the conventions that apply in the case of an international (armed) conflict (GC I–IV and AP I), while AP II contains no penal provisions whatsoever. This distinction reflects the *traditional "two-box approach"* differentiating between an international and a non-international armed conflict. It has been overcome with the seminal interlocutory decision of the *Tadic* Appeals Chamber (IT-94–1-AR 72) of the International Criminal Tribunal for the former Yugoslavia (ICTY) holding that IHL, in particular common Article 3 GC I–IV, provides for penal prohibitions in the case of a non-international conflict under certain conditions (para. 88 ff.).

As to the **grave breaches regime**, Articles 49 GC I, 50 GC II, 129 GC III, and 146 GC IV oblige the state parties to penalize conduct amounting to a grave breach of the GC, while Articles 50 GC I, 51 GC II, 130 GC III, and 147 GC IV contain the acts covered. The state parties have either to prosecute these acts before national courts or extradite those responsible to another state party (*aut dedere aut iudicare*). It is controversial whether the "grave breaches" norms provide for direct *individual criminal responsibility* given that they are only addressed to the state parties, obliging them to enact the respective penal prohibitions and ensure criminal prosecution. More detailed provisions concerning individual criminal responsibility can be found in AP I. Articles 86 and 87 AP I provide for *command or superior responsibility*. The superior may be liable for a failure to prevent crimes committed by his subordinates. In subjective terms at least a kind of negligence is necessary.

With regard to *grounds excluding responsibility* in general, the Geneva Law rejects the exclusion of criminal responsibility pursuant to a superior order implicitly. The recourse to *military necessity* is only possible in exceptional cases, namely, if the actions taken were necessary and proportional. This will rarely be the case if international crimes are committed. The *reprisal* defense was declared entirely unacceptable with a view to the protection of certain groups and objects (Prosecutor v. Martic, IT-95–11-R 61, 8.3.1996, para. 8 ff., 15 ff; Kupreškic et al. judgement, No. IT-95–16-T, 14.1.2000).

The Geneva Law also recognizes, at least partially, the *principles of legality and culpability*. Article 67 GC IV recognizes on the one hand the prohibition of retroactivity (*nullum crimen sine lege praevia*); on the other hand, it links the penalty to the offence thereby taking up the sentencing element of the principle of culpability (the punishment must conform to the actual culpability of the convicted person; see also Article 68 GC IV).

The Draft Codes of the International Law Commission and Private Initiatives

In 1947 the ILC was assigned to prepare a "Draft Code of Offences/Crimes against the Peace and Security of Mankind." The first Draft Code was adopted in the sixth session in 1954 (*Draft Code 1954*), which is composed of only four articles containing provisions on the general part and some criminal offences. After a definition of aggression was agreed upon by the GA in 1974, the ILC was again instructed to draft a code in 1981. The second Draft Code was adopted in 1991 (*Draft Code 1991*) and included provision concerning the general part and 12 offences, some of which show the political nature of the Draft Code 1991 (Articles 15–26). The third *Draft Code of 1996* basically rests upon the Draft Code of 1991. Great changes were undertaken in the special part, particularly by a sharp reduction of the former 12 offences to only 5: aggression, genocide, crimes against humanity, war crimes, and crimes against the UN and associated personnel. Apart from the last crime, this is exactly the catalogue of offences which was later included in the ICC Statute as the so-called core crimes. Prior to the 1996 Draft Code, in *1994*, the ILC had submitted a *Draft Statute* for an International Criminal Code (Ambos 2002/2004, pp. 444 ff.).

The various *unofficial proposals* for the development of ICL law may be subdivided into substantive law "Draft Codes" and procedural law "Draft Statutes." The most influential unofficial proposals have been the drafts of the Association Internationale de Droit Pénal (AIDP/ISISC/MPI 1995/1996) and of the International Law Association (Ambos 2002/2004, pp. 475 ff.) (ILA 1988).

The UN Ad Hoc Tribunals

The International Criminal Tribunal for the Former Yugoslavia

In reaction to massive violations of IHL and human rights in the former Yugoslavia beginning in 1991, the UN established the "United Nations Commission of Experts Pursuant to Security Council Resolution 780," which led to the establishment of the ICTY by the Security Council on 25 May 1993 (UN Doc. S/RES/827). The court was established as a subsidiary body of the Security Council according to Article 29 of the UN Charter. While the work of the court was initially supposed to last until the reestablishment of peace and security in the former Yugoslavia, the Security Council has in the meantime set up a so-called (first) **completion strategy** that fixed a time limit of 31 December 2004 for the end of investigations, 31 December 2008 for the end of trials of the first instance, and 31 December 2010 for the end of trials on appeal. On 22 December 2010, the Security Council adopted Resolution 1966 (2010) establishing the International Residual Mechanism for Criminal Tribunals with two branches, one for the ICTY and one for the ICTR; they commenced to operate on 1 July 2013 and 1 July 2012, respectively (UN SC-Res. 10141). According to the latest completion strategy, the tribunal continues its downsizing process (ICTY President 2012). All trials are expected to be completed by mid-2012, except for the case of *Radovan Karadžić*, which is expected to be completed in late 2013. Most appellate work is scheduled to be finished by the end of 2014 (ICTY President 2012). As of 19 November 2012, the ICTY has indicted 161 persons and concluded proceedings against 128 persons.

The **ICTY Statute** (ICTYS) contains 34 articles dealing with questions of substantive law (Articles 2–7, 24 ICTYS), procedural law (Articles 1, 8–10, 18–23, 25–30 ICTYS), as well as the organization of the tribunal (Articles 11–17, 31–34 ICTYS). Besides the Statute itself, several other legal instruments such as the Rules of Procedure and Evidence (RPE) have been adopted by the tribunal. It is composed of different chambers – the three Trial Chambers and the Appeals Chamber – the Office of the Prosecutor (OTP), and the Registry (Article 11 ICTYS). The chambers have 16 permanent judges and up to 12 so-called ad litem judges (Article 12(1) ICTYS) or ad hoc judges who can be appointed by the secretary general upon request of the President of the Tribunal (cf. Article 13*ter*, and *quarter* ICTYS). The OTP – though formally part of the tribunal – shall act independently as a separate organ (Article 16 ICTYS). The Registry "serves" both the chambers and the Prosecutor (Article 11(c) ICTYS). One of its functions (Article 17 ICTYS) is also the setting up of an adequate defense for the accused, including the assigning and payment of council.

The tribunal's **jurisdiction** extends to all (natural) persons responsible for serious violations of IHL committed in the territory of the former Yugoslavia since 1991 (Articles 1, 6 ICTYS). According to Articles 2–5 of the ICTYS, the tribunal exercises jurisdiction *ratione materiae* over grave breaches of the four Geneva Conventions, violations of the laws, or customs of war, genocide, and crimes against humanity. The rule on **individual criminal responsibility** is provided for in Article 7 (1) ICTYS. It includes three groups of perpetrators: the politically responsible official person, the (military) superior, and the (committing) subordinate. Superior/command responsibility is laid down in Article 7 (3) ICTYS. Neither the official position of the accused nor the action pursuant to an order shall relieve a person from criminal responsibility (Article 7 (2), (4) ICTYS), but this fact may be considered in mitigation of punishment if the tribunal determines "that justice so requires" (Article 7 (4) ICTYS).

The International Criminal Tribunal for Rwanda

The International Criminal Tribunal for Rwanda (ICTR) was established by Security Council Resolution 955 of 8 November 1994. The **Statute** of the ICTR (ICTRS) resembles largely the one of the ICTY. Judges from both tribunals are assigned to the Appeals Chamber of both tribunals (Article 13 (3) ICTRS, Article 14 (4) ICTYS). The members of the Appeals Chamber of the ICTY shall also serve as the members of the Appeals Chamber of the ICTR (Article 13 (4) ICTYS).

The competence of the ICTR embraces the prosecution of serious IHL violations committed in the territory of Rwanda and by Rwandan citizens in the territory of neighboring states between 1 January 1994 and 31 December 1994 (Article 1 ICTRS). The ICTR exercises **jurisdiction**, similar to the ICTY, over genocide, crimes against humanity, and internal armed conflict crimes. Just like the ICTY, the ICTR introduced a first draft for a **completion strategy** in 2003, which has been continuously updated and developed since then. The latest report was submitted on 14 November 2012. The tribunal's president informed that "[a]s at 5 November 2012, the Tribunal has completed its work at the trial level with respect to 92 of the 93 accused. [. . .] The one remaining trial judgement will be delivered before the end of 2012, and appellate proceedings have been concluded in respect of 44 persons. The remaining appeals are projected to be completed by the end of 2014" (ICTR President 2012, Para. 3). As already mentioned above, the ICTR will then, like the ICTY, be transferred into the Residual Mechanism, which started operating on 1 July 2012.

The International Criminal Court

Negotiating History

In 1994 the UN General Assembly (GA) referred the ILC-Draft Statute to the "**Ad Hoc Committee** on the Establishment of an ICC." This committee presented a final report after two sessions in 1995. Then, the **Preparatory Committee** (PrepCom), established by the GA just after the 1995 report, took over. Its task was to prepare a draft for the Rome Conference to be held in 1998. The PrepCom held altogether six meetings from 25 March 1996 until 3 April 1998. Between the sessions, more or less formal meetings by states and delegations of states took place in order to smoothen out possible points of conflict before the actual sessions. In one of those "intersessionals," the important **Zutphen Report** was compiled (Sadat Wexler and Bassiouni 1998, pp. 7 ff., 129 ff.).

On 15 December 1997, the GA decided to arrange a **State Conference** for the establishment of the ICC in **Rome** (UN GA-Res 52/160). The conference was not only open to states but also for nongovernmental organizations (NGO). It commenced on 15 June 1998 and ended on 17 July 1998 with the adoption of the ICC Statute. One hundred fifty-nine governmental delegations and 250 delegations of NGOs which had merged into the "Coalition for an ICC" attended the conference. Until the cessation of the conference, it was not entirely clear whether the ultimate goal – namely, the adoption of an ICC Statute – could be reached due to the opposition of important states. However, after intense negotiations, the conference adopted the court's Statute by a vote of 120 in favor to 7 against (USA, China, Libya, Israel, Iraq, Qatar, and Yemen), with 21 abstentions. A non-recorded vote was requested by the United States.

The Statute **entered into force** on **1 July 2002** after the deposit of the 60th instrument of ratification (cf. Article 126). Until the Statute was closed for signature on 31 December 2000, 139 states had signed, and as of 21 July 2012, 121 states had ratified it. Following the Rome Conference, a **Preparatory Commission** (PrepCommis) was first established in order to compile further legal instruments and to prepare the first meeting of the Assembly of State Parties (ASP). In addition, a working group on the crime of aggression was set up in order to reach a consensus on the definition and the conditions of jurisdiction pursuant to Article 5 (2) ICC Statute (on the final agreement see below).

The Rome Statute, the Structure of the Court, and Other Legal Instruments

General

The ICC Statute consists of 13 parts and 128 articles. The ICC was established as a permanent institution in The Hague (Articles 1, 3). While it is not an organ of the UN, it is linked to the latter by a "relationship agreement" (Article 2). The court is made up of a Presidency, a Pre-Trial Chamber (PTC), a Trial Chamber (TC) and an Appeals Chamber (AC), an Office of the Prosecutor (OTP), and a Registry (Article 34). Although the defense is not an organ of the court, an Office of Public Counsel was set up at the Registry (Regulation 77 of the Regulations of the Court).

The Judges

The judges are elected from two lists (Article 36 (5)): list A shall consist of candidates with established competence in criminal law and procedures and the necessary relevant experience. List B shall consist of candidates with established competence in relevant areas of international law, such as IHL and human rights law, and extensive experience in a professional legal capacity which is of relevance to the judicial work of the court. Especially the List B requirements are increasingly criticized for allowing diplomats without criminal trial experience to become ICC judges (Bohlander 2009, pp. 532 ff.; Ambos 2012a, pp. 224 ff). In addition, the candidates shall be of "high moral character, impartiality and integrity" and "possess the qualifications of their national law for appointment to the highest judicial offices" (Article 36 (3)(a) ICC Statute). They must be fluent in English or French (Article 36 (3)(c)). The judges shall be selected by lot to serve 3, 6, or 9 years (Article 36 (9)(b)). Only the judges elected for a term of 3 years are eligible for reelection (Article 36 (9)(c)). Judicial impartiality shall be secured by not engaging in any other occupation of professional nature. The judges shall represent the main legal systems of the world. On 10 September 2004, the ASP adopted a resolution on the "Procedure for the Nomination and Election of Judges of the International Criminal Court" (ICC-ASP/3/Res.6), providing for rules for the nomination and election of the judges. However, the respective "Advisory Committee on Nominations" of the ASP (Article 36 (4)(c) ICC Statute) has not yet been established. Updated information on the judges and chambers can be found at http://www.icc-cpi.int/.

The Office of the Prosecutor

The OTP shall act independently and as a separate organ of the court. Its first head was the Argentinean Luis Moreno Ocampo – he is "The Prosecutor" (cf. Article 42). He was elected on 21 April 2003 and took office on 16 June 2003. The ASP also elected two deputy prosecutors, Mr. Serge Brammertz (Investigations) and Mrs. Fatou Bensouda (Prosecutions). After Brammertz' leave to Lebanon in January 2006 (before assuming the Office of Chief Prosecutor of the ICTY in January 2008), only Fatou Bensouda from Gambia stayed on as deputy prosecutor. On 1 February 2011, Phakiso Mochochoko from Lesotho was appointed as Head of the Jurisdiction, Complementarity, and Cooperation Division. The actual power now resides in an Executive Committee, composed of the prosecutor, the deputy prosecutor, and the heads of the different sections (JCCD, Investigation, and Prosecution) and supported by some external consultants. Moreno Ocampo's mandate expired in June 2012. Deputy Prosecutor Bensouda was elected unanimously as his successor on 12 December 2011 at the ASP's tenth session. She took office on 16 June 2012.

The OTP concluded several agreements with other organizations and persons, i.e., with the International Criminal Police Organization (Interpol) or so-called intermediaries (Ambos 2011a, p. 329 fn. 123). It cooperates with the UN by virtue of the UN-ICC agreement and the MONUC-Memorandum of the ICC.

Registry and Assembly of States Parties

The Registry is responsible for the administration and servicing of the court and is headed by the Registrar (Silvana Arbia, Italy, successor of Bruno Cathala, France). It consists of the Immediate Office of the Registrar, the Security and Safety Section, the Common Administrative Services Division, the Division of Court Services, Public Information, the Documentation Section, and the Division of Victims and Counsel (Lachowska 2009, p. 389).

The Assembly of States Parties (Article 112) is composed primarily of representatives of the states that have ratified and acceded to the Rome Statute. Other states, which have signed the Statute or the Final Act, may be observers in the Assembly (Article 112 (1)). The ASP is supposed to meet annually and can be seen as the decision-making organ of the court. It decides on various issues, such as the interpretation/application of the Statute, the adoption of legal texts and of the budget, and the election of the judges, the prosecutor, and the deputy prosecutor(s). Any dispute between two or more States Parties relating to the interpretation or application of the Statute shall be referred to the ASP. The Assembly may itself seek to settle the dispute or may make recommendations on further means of settlement of the dispute, including referral to the International Court of Justice in conformity with the Statute of that court (Article 119 (2)). The ASP shall also supervise the State Parties' compliance with their cooperation obligations under the Statute (Article 112 (2)(f) in relation to Article 87 (5) (b) and (7)). It is the only enforcement organ of the Rome system in this respect, unless the UN Security Council has referred a situation to the court (Article 87 (5)(b) *in fine*).

Legal Sources

Apart from the Statute itself, the two most important secondary legal sources of the ICC are the Elements of Crimes and the Rules of Procedure and Evidence, which were agreed on in the fifth session of the PrepCommis and were finally adopted in the first session of the ASP. In accordance with Article 9 ICC Statute, the **Elements of Crimes** shall assist the court in the interpretation and application of the core crimes (Articles 6, 7, 8, and 8*bis*). They are a subsidiary source, have to be consistent with the Statute (Article 9 (3)), and are subject to legal interpretation of the court. The **Rules of Procedure and Evidence** (RPE) complement the procedural regime of the Statute and have to be consistent with it (Article 51 (4)). In the event of conflict, the latter shall prevail (Article 51 (5)). They do not affect the procedural rules for any national court or legal system for the purpose of national proceedings.

On 26 May 2004, the judges adopted the **Regulations of the Court** (Article 52). The regulations were developed to fulfill the goal of speedy trials and to secure a fair trial for the accused. Regulations of the Registry and of the OTP were adopted on 3 March 2006 and 23 April 2009, respectively. Another legal source is the **Agreement on Privileges and Immunities** (Article 48) which grants certain immunities and privileges to the judges, the prosecutor and its staff, the Registrar and its staff, as well as to counsels, experts, witnesses, or any other person required to be present at the seat of the court. Other legal sources are inter alia: the Code of Professional Conduct for Counsel, the Code of Judicial Ethics, the agreement between the International

Criminal Court and the United Nations, and the agreement with the EU on cooperation and assistance.

The First Review Conference in Kampala

The first Review Conference took place in Kampala, Uganda, from 31 May to 11 June 2010, a year later than envisaged by Article 121. ICC States Parties, observer states, international organizations, NGOs, and other participants discussed proposed amendments to the Rome Statute and took stock of its impact to date. Debates focused on the impact on victims and affected communities, complementarity, cooperation, and peace and justice. Apart from these general discussions, the conference took three concrete decisions: most importantly, the crime of aggression was defined (Art. 8*bis* Statute), and the conditions for the exercise of jurisdiction were agreed upon (Arts. 15bis, 15ter; Ambos 2011b). Further, the deletion of Article 124 was rejected, and the war crimes of Article 8 (2)(b) (xvii) to (xix) were extended to a non-international armed conflict (Article 8 (2)(e) (xiii) to (xv) Ambos 2011a, p. 125).

Current Investigations

The OTP first initiated investigations mid-2004 with regard to two **situations**. One was the situation in **Uganda** closely connected with the activities of the so-called Lord's Resistance Army (LRA); the other one concerns crimes committed on the territory of the **Democratic Republic of Congo** (DRC) since 1 July 2002. In both situations, the respective governments made use of the possibility of a state referral in accordance with Article 13 (a), 14 ICC Statute. In June 2004 (DRC) and July 2004 (Uganda), the OTP determined that there is a reasonable basis to open a formal investigation into the situations (Article 53 (1)). On 21 December 2004, the **Central African Republic** (CAR) referred a situation to the ICC and requested an investigation by the OTP into the crimes committed on its territory since 1 July 2002. The OTP opened investigations on 22 May 2007. On 13 July 2012, Mali has self-referred a further situation to the ICC through the use of Article 14 ICC Statute. On 16 January 2013, the OTP formally opened an investigation into the alleged crimes, committed on the territory of Mali since January 2012.

While in all these situations, the court's jurisdiction was triggered by state self-referrals under Article 13 (a), 14 of the Statute, two other situations have been referred to the court by the Security Council pursuant to Article 13 (b), namely, **Darfur** (**Sudan**; Resolution 1593 of 31 March 2005) and **Libya** (Res. 1970 of 26 February 2011). The Darfur investigation was formally opened in June 2005. In the Libyan situation, the prosecutor announced very quickly, on 3 March 2011, that he will open a formal investigation. Last but not least, with regard to **Kenya**, the prosecutor acted for the first time proprio motu pursuant to Articles 13 (c) and 15. On 31 March 2010, Pre-Trial Chamber II authorized the prosecutor to open the investigation pursuant to Article 15 (4) with regard to crimes against humanity committed between 1 June 2005 and 26 November 2009.

Apart from these formal trigger mechanisms, Article 12 (3) offers non-States Parties the possibility to accept the jurisdiction of the court by a kind of **ad hoc declaration** "with respect to the crime in question." So far, this provision has been invoked in two cases. First, on 15 February 2005, **Ivory Coast** accepted the jurisdiction of the ICC with respect to alleged crimes committed from 19 September 2002, which was renewed by both the former President Laurent Gbagbo and current President Alassane Ouattara. On 3 October 2011, Pre-Trial Chamber III authorized the prosecutor to open an investigation into war crimes and crimes against humanity allegedly committed following the presidential election of 28 November 2010. On 22 February 2012, this authorization was expanded to include crimes allegedly committed between 19 September 2002 and 28 November 2010. Secondly, on 22 January 2009, the **Palestinian National Authority** lodged a declaration with regard to acts committed on the territory of Palestine since 1 July 2002, especially during the 2008/2009 Gaza war.

If an investigation is formally opened, several **cases** arise from the respective situation, persons are targeted, and, if they do not voluntarily surrender to the court, **arrest warrants** are issued. So far, arrest warrants have been issued regarding the situations in Uganda (five arrest warrants issued, one case terminated because of the death of the suspect), Dafur/Sudan (five), DRC (seven, of which five have been executed), Libya (three, one case terminated), CAR (one issued and executed), and Côte d'Ivoire (two issued and one executed). Thus, only seven arrest warrants have been executed so far. Trials started against *Jean-Pierre Bemba Gombo*, *Thomas Lubanga Dyilo*, *Germain Katanga*, and *Mathieu Ngudjolo Chui*. The first judgement was delivered on 14 March 2012 in the Lubanga case (Ambos 2012b). On 10 July 2012, Lubanga was sentenced to a total period of 14 years of imprisonment. On 18 December 2012, Mathieu Ngudjolo Chui was found not guilty of the charges brought against him. In other cases summons to appear have been considered sufficient, and the suspects appeared voluntarily before the court, i.e., in the case of Abu Garda, a member of the Sudanese Janjawid militia (ICC-02/05-02/09-2), Banda Abakaer Nourain (ICC-02/05-03/09-3) and Mohammed Jerbo Jamus (ICC-02/05-03/09-2) and with regard to the so-called Ocampo six (Prosecutor v. Kirimimuthaura et al. ICC-01/09-02/11-01, Samoeiruto et al. ICC-01/09-01/11-01).

The "Mixed" Tribunals

The Legal Bases

As a result of the increasing internationalization of the prosecution of serious human rights violations, many so-called mixed or hybrid tribunals have been established in several states. These tribunals have a mixed national-international legal basis and recruit national and international (foreign) prosecutors and judges. The tribunals are either part of a **transitional UN administration** (Kosovo, East Timor), or based on a **bilateral agreement with the UN** (Sierra Leone, Cambodia, Lebanon), or on legislative provisions adopted by an **occupying power** (Iraq). A purely **national tribunal** for international crimes was created in Bangladesh.

Kosovo and East Timor

Following the armed conflict between Serb authorities and the Kosovo Liberation Army, **Kosovo** was placed under the interim administration of the UN on 10 June 1999. The competence of the transitional UN administration (UNMIK) in Kosovo for "maintaining civil law and order" and the representation by a Special Representative of the Secretary-General (SRSG) derive from SC-Res. 1244 (6) and (11 (i)). On this basis numerous "regulations" and "administrative directions" have been enacted in order to define the applicable law (Bohlander 2003, pp. 24 ff.). As a result, the Provisional Institutions of Self-Government (PISG) were established, including a government, a president, a parliament, and a court system. As to criminal justice, there exist three instances for the adjudication of crimes of international concern. The attempt to set up a special tribunal ("Kosovo War Crimes and Ethnic Crimes Court") failed; instead, international prosecutors and judges were assigned to all district courts in a 2:1 proportion (two international, one local judge). The court system includes a constitutional court, a supreme court, five district courts, a commercial court, 25 municipal courts, 25 minor offence courts, and an appellate court for minor offences. Shortly before the declaration of independence by Kosovo on 17 February 2008, the European Union Rule of Law Mission in Kosovo (EULEX) was established in February 2008 to "monitor, mentor, and advise Kosovo institutions in all areas related to the rule of law and to investigate, prosecute, adjudicate, and enforce certain categories of serious crimes" (Article 3(a) and (d) of the Council Joint Action 2008/124/CFSP). Through EULEX, 31 international judges and 15 international prosecutors support local judges and prosecutors. There is one state public prosecutor's office, five district prosecutors' offices, and seven municipal prosecutors' offices. EULEX exercises its executive authority over a special prosecutor's office, which includes eight international prosecutors, and focuses on serious crimes including human trafficking, money laundering, war crimes, and terrorism.

In **East Timor** – in accordance with SC-Res. 1272 (1999) of 25 October 1999 – the "United Nations Mission in East Timor"

(UNTAET) had overall responsibility for the administration of East Timor and was empowered to exercise all legislative and executive authority, including the administration of justice. The organization of the courts in East Timor was reorganized, and at the same time, panels with exclusive jurisdiction over serious criminal offences (genocide, war crimes, crimes against humanity, murder, sexual offences, and torture) were established within the District Court in Dili. The panels had exclusive jurisdiction only for offences committed in the period between 1 January 1999 and 25 October 1999 and operated until 20 May 2005. Moreover, within the Office of the General Prosecutor, there was a "Special Prosecutor" called the Deputy General Prosecutor for Serious Crimes, who headed the Serious Crimes Unit (Othman 2003, pp. 87 ff.). After the independence on 20 May 2002, the laws launched by UNTAET remained in effect, and the judges have been appointed by the Supreme Council of the Judiciary of East Timor (von Braun 2008, pp. 137 ff.). In 2005, the mandate of the Serious Crimes Unit expired, and its investigative functions were resumed by the Serious Crimes Investigation Team, assisting the Office of the Prosecutor-General of East Timor. Until the expiration of its mandate, the Serious Crimes Unit had indicted 391 people. While 84 defendants were convicted, three were acquitted in trials before the special panels, and more than 300 indictees remained at large, almost all of them in Indonesia.

Sierra Leone

In SC-Res. 1315 (2000), the Security Council asked the UN-Secretary General to negotiate an agreement with the Government of Sierra Leone to create an independent court to prosecute persons responsible for the commission of serious violations of IHL and crimes committed under Sierra Leonean law during the country's civil war. On 16 January 2002, such an agreement, accompanied by a Statute of the Special Court for Sierra Leone (SCSLS), was signed (Kelsall 2009, pp. 254 ff.). Thus, the court is based upon a bilateral, international law agreement between an international organization and a state and not solely upon a UN SC Resolution. It has the power to prosecute persons who bear the greatest responsibility for serious violations of IHL and Sierra Leonean law committed in the territory of Sierra Leone since 30 November 1996 (Article 1 SCSLS). This covers precisely crimes against humanity, war crimes, certain other international crimes, and certain crimes under Sierra Leonean law (Articles 2–5 SCSLS). The Special Court and the national courts shall have concurrent jurisdiction. The Special Court shall have primacy over the national courts of Sierra Leone and may at any stage of the procedure request a national court to defer to its competence in accordance with the SCSLS and the Rules of Procedure and Evidence (Article 8 SCSLS). In 2003, the prosecutor issued 13 indictments, of which two were withdrawn due to the deaths of the accused in December 2003. Thus far, the trials of three former leaders of the Armed Forces Revolutionary Council (AFRC), of two members of the Civil Defense Forces (CDF), and of three former leaders of the Revolutionary United Front (RUF) have been completed, including appeals. On 31 October 2009, the Special Court's eight convicted people were transferred to Mpanga Prison, Rwanda for sentence enforcement. The judgement against former Liberian President *Charles Taylor* was rendered on 26 April 2012. In a unanimous judgement (the "dissenting opinion" of the alternate Judge El Hadji Malick Sow from Senegal, expressed after the verdict had been delivered, does not count as a vote), Taylor was found guilty on all accounts of aiding and abetting the RUF and AFRC rebel groups and/or Liberian fighters operating in Sierra Leone and of having planned attacks on civilians. On 30 May 2012, he was sentenced to a term of 50 years imprisonment. Upon the delivery of the final judgment in the Charles Taylor case, the Special Court's mandate will be complete and the SCSL closes. However, many of its legal obligations will not terminate with the conclusion of all cases. In August 2010, the United Nations and the Government of Sierra Leone agreed to establish a Residual Special Court for Sierra Leone, which will be responsible for fulfilling the Special Court's obligations. The RSCSL Agreement and the RSCSL Statute were ratified by the Parliament of Sierra Leone in December 2011.

The RSCSL shall, pursuant to Article 1.1 of its Statute: "[. . .] maintain, preserve and manage its archives, including the archives of the Special Court; provide for witness and victim protection and support; respond to requests for access to evidence by national prosecution authorities; supervise enforcement of sentences; review convictions and acquittals; conduct contempt of court proceedings; provide defence counsel and legal aid for the conduct of proceedings before the Residual Special Court; respond to requests from national authorities with respect to claims for compensation; and prevent double jeopardy." At the beginning of 2012, the Registrar established the Residual Special Court for Sierra Leone Transition Working Group in order to coordinate work relating to the transition to RSCSL and closure of the Special Court (SCSL Ninth Annual Report 2011-2012, p. 7). The RSCSL will have its interim seat in The Hague and an office in Sierra Leone. It will be headed by a president (chosen by RSCSL judges), a prosecutor and a registrar.

Cambodia

In Cambodia, after long-lasting negotiations, a bilateral agreement with the UN was signed on 06 June 2003 "concerning the prosecution in Cambodian law of crimes committed during the period of Democratic Kampuchea" (during the period from 17 April 1975 to 6 January 1979). Parallel to the negotiations with the UN, a "Law on the Establishment of the Extraordinary Chambers" was prepared in order to prosecute these crimes by national institutions. The law was first adopted on 10 August 2001 and later amended (ECCC-Law). The extraordinary chambers have the power to bring trials against suspects who committed genocide, crimes against humanity, grave breaches of the Geneva Conventions, as well as certain other enumerated international and national crimes (Articles 2–8 ECCC-Law). The chambers are established within the existing court structure; they operate as a court of first instance, the Supreme Court being an appellate court and final instance (Kashyap 2003, pp. 192 ff.).

The extraordinary chambers are unique in structure and composition. The Trial Chamber is composed of five professional judges, of whom three are Cambodian (with one as president) and two foreign (Article 9 (1) ECCC-Law). The Appeals Chamber is composed of seven judges, of whom four are Cambodian (with one as president) and three foreign (Article 9 (2) ECCC-Law). While the courts in East Timor and Sierra Leone take their decisions with a simple majority, in the ECCC national judges dominate and a supermajority (a simple majority plus one) is necessary (Article 14 ECCC-Law). The judges agreed on rules of procedure on 12 June 2007.

Personal jurisdiction is limited since only the *senior leaders of Democratic Kampuchea and those who were most responsible* shall be brought to trial. Those leaders are former heads of the Khmer Rouge (Heder and Tittemore 2004). In Case 001, *Kaing Guek Eav alias Duch* was sentenced – after an appeal by the co-prosecutors – to life imprisonment. In Case 002, the trial against *Nuon Chea*, *Ieng Sary*, *Khieu Samphan*, and *Ieng Thirith* began on 27 June 2011. On 4 March 2013, Ieng Sary died in hospital from natural causes, as the Co-Prosecutors announced on 2 April 2013. On 7 September 2009, the international co-prosecutor filed two Introductory Submissions, requesting the co-investigating judges to initiate investigation of five additional suspected persons. These two submissions have been divided into what is known as Case files 003 and 004.

Iraq

The Iraqi Special Tribunal was established by the US Coalition Provisional Authority; its Statute was issued on 10 December 2003 by the Iraqi Governing Council and approved on 18 October 2005 by the first freely elected Parliament. The name of the court was changed to the "Iraqi Higher Criminal Court," which is now financed exclusively by the Iraqi government. The tribunal is not part of the regular Iraqi judicial system but an autonomous organ with its own rules and an own administrative capacity. The tribunal has jurisdiction over any Iraqi national or resident of Iraq accused of the core crimes (Articles 11–13,) committed since 17 July 1968 (takeover of the Ba'ath party) and up until and including

1 May 2003 (official ending of acts of war) in the territory of the Republic of Iraq, or elsewhere, including crimes committed in connection with Iraq's wars against the Islamic Republic of Iran and the state of Kuwait (Article 1(b)). Moreover, the tribunal has the power to prosecute certain violations of Iraqi laws (Article 14). The tribunal consists of the Tribunal Investigative Judges, one or more Trial Chambers, and an Appeals Chamber (Article 3). The Statute resembles the ICC Statute in its substantive law and procedural law provisions.

Besides former president *Saddam Hussein* (who was sentenced to death on 5 November 2006 and executed on 30 December 2006), several high-ranking Iraqi officials were tried, among others in the following cases: the *Al-Dujail* case against eight accused, which started on 19 October 2005 (all eight convicted, four sentenced to death – one after appeal – 3–15 years imprisonment, and one defendant was acquitted), and the *Al-Anfal* case, which started on 24 June 2007, against members of the former Ba'th regime (three sentenced to death, two to life imprisonment, and one acquitted). The third trial started in August 2007, relating to the brutal crushing of a Shiite rebellion in 1991 (three of the defendants were acquitted, four were sentenced to life imprisonment, six were given a long prison sentence, and two were sentenced to death). The fourth trial dealt with the execution of 42 merchants who were accused of raising their prices during the period when UN sanctions had been imposed against Iraq (two defendants were sentenced to death, one to life, 3–15 years and 1–6 years imprisonment, one defendant was acquitted). Further proceedings were initiated in the following cases: against 14 accused (inter alia against *Ali Hassan al-Majid*, *Hashim Hassan al-Majid*, *Tarik Aziz Issa*, members of the former Ba'th regime, and of the militia) for the deportation and forced movement of families in 1984. The indictment was issued on 4 November 2008, and the judgement was rendered on 2 August 2009. In another case, an indictment was issued against 14 persons (inter alia against the former minister for the interior *Sadun Shaker*) for the killing and forced displacement of Falili-Kurds. The judgement was rendered on 29 November 2010. Furthermore, in the *Al Jeboor* case, four accused were found guilty of crimes against humanity.

Lebanon and Bangladesh

The **Special Tribunal for Lebanon** (STL) was created by SC-Res. 1757 (2007) of 30 May 2007. The provisions of the document annexed to it, and the Statute of the Special Tribunal thereto attached, entered into force on 10 June 2007. The STL is based in The Hague and is neither a subsidiary organ of the UN nor is it a part of the Lebanese court system. Rather, it supersedes the national courts within its jurisdiction (Article 4 (1)). It is a hybrid court in the sense that it is composed of both national and international judges. The STL is unique in that the applicable law is national in character, while the ICTY and ICTR are limited to prosecuting crimes in violation of international law and the (other) hybrid tribunals prosecute crimes under both domestic and international law. In addition, the STL is the first UN-assisted tribunal to combine substantial elements of both a common law and a civil law legal system (U.N. Doc. S/2006/893).

According to Article 1 of the Statute of the Special Tribunal for Lebanon (STL), "[T]he Special Tribunal shall have *jurisdiction* over persons responsible for the attack of 14 February 2005 resulting in the death of former Lebanese Prime Minister Rafic Hariri and in the death or injury of other persons." However, the temporal jurisdiction was extended to include other attacks bearing the same, or similar, characteristics of the Hariri assassination (U.N. Doc. S/2006/893). As to the substantive law, the Statute stipulates that the tribunal shall apply provisions of the Lebanese Criminal Code relating to the prosecution and punishment of acts of terrorism (a crime that so far has not been within the province of an international tribunal) and crimes and offences against life and personal integrity, among others. Following an indictment by former prosecutor Daniel Bellemare on 17 January 2011, on 30 June 2011, the tribunal issued four arrest warrants that have not been executed so far. The STL's first case is *The Prosecutor v. Ayyash,* *Badreddine, Oneissi and Sabra*. In this case, which awaits completion of the pre-trial phase, the Prosecutor accuses Ayyash, Badreddine, Oneissi and Sabra of criminal responsibility for the attack that killed former Lebanese Prime Minister Rafik Hariri and others on 14 February 2005 (STL President 2012-2013, p. 7). On 1 February 2012, the Trial Chamber had made an order for trial in absentia, which was confirmed by the Appeals Chamber on 1 November 2012 (STL President 2012-2013, p. 10). On 19 July 2012, the Pre-Trial Judge set 25 March 2013 as a tentative date for trial to start. However, many procedural problems, including incomplete disclosure and technical issues faced by the Defence in accessing certain disclosed material, caused the Pre-Trial Judge to postpone the commencement of trial (STL President 2012-2013, pp. 7, 9).

The jurisdiction of the **Bangladesh Tribunal** includes crimes against humanity, crimes against peace, genocide, war crimes, and "violation of any humanitarian rules applicable in armed conflicts laid down in the Geneva Conventions of 1949" and "any other crimes under international law" Article 3 (2) (a)-(f) International Crimes (Tribunals) Act 1973 (ICTA). Common law and customary international law are treated as primary sources of law, and the tribunal resembles the existing tribunals, albeit it conducts a purely domestic process (Linton 2001, pp. 221 ff.). On 20 November 2011, the first person was charged (Delwar Hossain Sayedee, a leader of Jamaat-e-Islami, an Islamist party opposed s Bangladesh's independence). Subsequently, the charges were "framed" (i.e., confirmed) by the tribunal. Further, the chief investigator Abdul Hannan Khan carries out investigations against 10 other suspects, including another six members of the Jamaat and two of the Bangladesh Nationalist Party. Apart from Sayedee, charges have been filed against Maulana Matiur Rahman Nizami, Ali Ahsan Mohammad Mujahid, Abdul Kader Mollah, Abdul Alim, Muhammad Kamaruzzaman, and Salahuddin Quader Chowdhury. Charges were framed against Kamaruzzaman on 4 June 2012 (the trial started on 2 July 2012), Nizami and Mollah on 28 May 2012, and Chowdhury on 4 April 2012.

Comparative Analysis

All these tribunals can be characterized as "mixed" not only because of their composition but also because of their organization, structure, and the applicable law. The tribunals apply **national and international law**. While there was first, in line with the ICTY and ICTR precedent, a certain preference for the common law system, in particular in terms of the applicable procedure, the ECCC introduced an inquisitorial-like French procedure, and the STL, as the first UN tribunal, combines elements of both legal systems.

It is common to all tribunals (with the exception of the STL) that they are **situated** in the state where the crimes of their subject matter jurisdiction took place. Thereby, a certain proximity to the local, crime-affected population is ensured. Either the tribunals are part of the local justice system (Kosovo, East Timor, Cambodia) or though special tribunals, somehow affiliated with the national system (Sierra Leone, Iraq, Lebanon). Contrary to the ICTY and ICTR, which have had no less than 15 years to terminate their proceedings, the mixed tribunals have a significantly **shorter time period** to conclude their work. For instance, the SCSL originally had only 3 years to fulfill its mandate. The problems become even more apparent if one looks at the **organizational problems** and the **tight resources** compared to the high operative cost they are bound to combat. Their budgets are remarkably lower than the ones of the ad hoc tribunals.

The crimes falling within the subject matter **jurisdiction** of the mixed tribunals are the **core crimes** genocide, crimes against humanity, and war crimes. The elements of the crime of these core crimes are related to the ICC Statute, but war crimes are in the majority of cases not as detailed codified as in Article 8 ICC Statute. In addition, all tribunals apply **specific violations of national law** depending on the situation: in East Timor, torture was added as an offence; in Kosovo incitement to national, racial, religious, or ethnic hatred, discord or intolerance ((1) of Regulation 4 (2000)), the illegal possession of weapons ((8) of Regulation 7 (2001)), and unauthorized border crossing ((3) of Regulation 10 (2001)) are criminal offences; in Cambodia, the destruction of cultural property during armed conflict can be prosecuted; in Sierra Leone, offences relating to the abuse of girls and setting fire to dwelling houses were included; and in Iraq, the wastage of national resources is a violation, and the STL even applies exclusively national law. Furthermore, in Sierra Leone, adolescents between 15 and 18 years old can be brought to trial (Article 7 SCSLS).

The existence of **extrajudicial mechanisms** for dispute resolution, for instance, through **truth and reconciliation commissions** (Bassiouni 2003, pp. 711 ff.), leads to concurrent jurisdictions or at least entails difficulties of delimitation. In Sierra Leone, overlapping jurisdiction was meant to be avoided by prosecution of only the persons most responsible by the Special Court itself, basically leaving child and adolescents to the competence of the truth commission. The legitimacy of the Iraqi Court has been highly criticized from the outset due to its establishment by the US occupying power and its legal source (Megally and Zyl 2003). While the acceptance of the court by the Iraqi people may indeed be questioned, this fact is not an anomaly but lies at the heart of an international criminal justice system dominated by ad hoc tribunals. In fact, in this respect, all these tribunals face a **dilemma**: on the one hand, the national judiciary is generally not able and very often unwilling to carry out proceedings for internationalized core crimes; on the other hand, the "internationalization" of the courts and procedures gives rise to a deficit in its legitimacy in relation to the local population.

Related Entries

- ▶ Genocide
- ▶ International Responses to Victims in Criminal Justice
- ▶ Victims and the International Criminal Court

Recommended Reading and References

Ahlbrecht H (1999) Geschichte der völkerrechtlichen strafgerichtsbarkeit im 20. Jahrhundert. Nomos, Baden-Baden

Ambos K (2002/2004) Der Allgemeine Teil des Völkerstrafrechts – Ansätze einer Dogmatisierung. Duncker & Humblot, Berlin

Ambos K (2011a) Internationales Strafrecht. C.H. Beck, München

Ambos K (2011b) The crime of aggression after Kampala. Ger Yearb Int Law 53: 463–509

Ambos K (2012a) International courts and their judges. Book review. Crim Law Forum 23:223–228

Ambos K (2012b) The first judgment of the International Criminal Court (Prosecutor v. Lubanga): a comprehensive analysis of the legal issues. Int Crim Law Rev 12:115–153

Ambos (2013) Treatise on International Criminal Law, Vol. I: Foundations and General Part, Ch. I. Oxford University Press, Oxford

Association International de Droit Pénal, Instituto, Superiore Internazionale di Scienze Criminali, Max-Planck-Institut für ausländisches und internationales Strafrecht (1995/1996) Siracusa-Draft and updated Siracusa-Draft

Bailey SD (1972) Prohibitions and restraints in war. Oxford University Press, London

Bassiouni MC (2003) Introduction to international criminal law. Transnational Publishers, Ardsley

Bassiouni MC (1999) Crimes against humanity in international criminal law, 2nd ed. Kluwer Law International, The Hague

Bohlander M (2009) Pride and prejudice or sense and sensibility? A pragmatic proposal for the recruitment of judges at the ICC and other international criminal courts. New Crim Law Rev 12:529–542

Bohlander M (2003) The legal framework of the prosecution and the courts. In: Ambos K, Othman M (eds) New approaches in international criminal justice – Kosovo, East Timor, Sierra Leone and Cambodia. ed. iuscrim, Freiburg i. Brsg, pp 21–60

Cryer R (2005) Prosecuting international crimes. Cambridge University Press, Cambridge

Engelhart M (2004) Der Weg zum Völkerstrafgesetzbuch – Eine kurze Geschichte des Völkerstrafrechts. Jura, pp 734–743

Heder S, Tittemore BD (2004) Seven candidates for prosecution – accountability for the crimes of the Khmer Rouge, 2nd edn. Documentation Center of Cambodia, Phnom Penh

ICTR President (2012) Letter dated 22 May 2012 from the President of the International Criminal Tribunal for Rwanda addressed to the President of the Security Council. S/2012/836. 14 November

ICTY President (2012) Letter dated 16 November 2012 from the President of the ICTY, addressed to the President of the Security Council. S/2012/847. 19 November

International Law Commission (1950) Yearbook of the International Law Commission, 1950, Vol. II, Report of the International Law Commission covering its second session, 5 June – 29 July 1950, pp. 364 ff., Available at http://untreaty.un.org/ilc/publications/yearbooks/1950.htm.

International Law Reports (1968), pp. 5 ff.; (1995), pp. 338 ff., 357 ff., 520 ff.

Kashyap S (2003) The framework of the prosecution in Cambodia. In: Ambos K, Othman M (eds) New approaches in international criminal justice – Kosovo, East Timor, Sierra Leone and Cambodia. Ed. Iuscrim, Freiburg i. Brsg, pp 189–205

Kelsall T (2009) Culture under cross-examination. Cambridge University Press, Cambridge

König K-M (2003) Die völkerrechtliche Legitimation der Strafgewalt internationaler Strafjustiz. Nomos, Baden-Baden

Lachowska A (2009) The support work of the Court's registry. In: Doria J, Gasser H-P, Bassiouni MC (eds) The legal regime of the international criminal court – essays in honour of Professor Igor Blishchenko. Nijhoff, Leiden, pp 387–400

Lemkin R (1933) Akte der Barbarei und des Vandalismus als delicta juris gentium. Int Anwaltsblatt 18:117–119

Linton S (2001) Cambodia, East Timor and Sierra Leone. Crim Law Forum 12:185–246

MacCoubrey H, White ND (1992) International law and armed conflict. Dartmouth, Aldershot

Megally H, Zyl Pv (2003) U.S. justice with an Iraqi face? Int Her Tribune 4

Osten P (2003) Der Tokioter Kriegsverbrecherprozeß und die japanische Rechtswissenschaft. Berliner Wissenschaftsverlag, Berlin

Othman M (2003) The framework of prosecutions and the court system in East Timor. In: Ambos K, Othman M (eds) New approaches in international criminal justice – Kosovo, East Timor, Sierra Leone and Cambodia. ed. Iuscrim, Freiburg im Breisgau, pp 85–112

Puttkamer EV (1949) Die Haftung der politischen und militärischen Führung des 1. Weltkrieges für die Kriegsurheberschaft und Kriegsverbrechen. Arch Völkerrechts 1:424–449

Sadat Wexler LN, Bassiouni MC (1998) Model draft statute for the International Criminal Court based on the Preparatory Committee's text to the diplomatic conference. Nouvelles Études Pénales, Rome

Sassoli M, Bouvier A (2006) How does law protect in war? – cases, documents and teaching materials on contemporary practice in international humanitarian law. Vol. 1: outline of international humanitarian law. Vol. 2: cases and documents, 2nd edn. International Committee of the Red Cross, Geneva

Schabas W (2009) Genocide in international law, 2nd edn. Cambridge University Press, Cambridge

STL President (2012-2013) Special Tribunal for Lebanon, Fourth Annual Report 2012-2013

United Nations War Crimes Commission (1947–1949). Law Reports of Trials of War Criminals, Selected and prepared by the United Nations War Crimes Commission 1947–1949, Available at http://www.loc.gov/rr/frd/Military_Law/law-reports-trials-war-criminals.html. vol 1–15

von Braun L (2008) Internationalisierte Strafgerichte – eine Analyse der Strafverfolgung schwerer Menschenrechtsverletzungen in Osttimor, Sierra Leone und Bosnien-Herzogowina. Berliner Wissenschaftsverlag, Berlin

Development of Sex Offending

► Criminal Career of Sex Offenders

Development of the UCR and the NCVS

Lynn A. Addington
Department of Justice, Law, and Society,
American University, Washington, DC, USA

Overview

The importance of the Uniform Crime Reporting Program (UCR) and the National Crime Victimization Survey (NCVS) cannot be emphasized enough as they are the only two ongoing national measures of crime in the United States that provide annual level and change estimates. While discussions of the UCR and NCVS often focus on the differences between the two, the two systems share many similarities in their development over time as well as their influence on how crime has been defined and studied. This entry examines the origins and development of these two data collection systems individually and describes the similarities between them. This entry concludes by examining how the UCR and NCVS have contributed to criminology by shaping views on what crimes "count," especially the focus on serious, violent street crimes. Alternative ways crime could be measured in the UCR and NCVS's existing frameworks are discussed.

Introduction

The Uniform Crime Reporting Program (UCR) and the National Crime Victimization Survey (NCVS) are the only two ongoing national measures of crime in the United States that provide annual level and change estimates. Data collected under the UCR use official records to measure crimes known to police. The NCVS data rely on survey responses to generate estimates of criminal victimization. This entry examines the origins and development of the UCR and NCVS individually and describes the similarities between the two data collection systems. This entry concludes by discussing how both systems have contributed to criminology by shaping what crime "counts," particularly the focus on violent, street crime. Finally, alternative ways are considered with regard to defining and measuring crime that could be used in the UCR and NCVS's existing frameworks.

Origins and Development of the UCR and NCS-NCVS

Uniform Crime Reporting Program

Before the UCR, very little information was available to study crime, and the existing data were primarily of prisoners collected by the Bureau of the Census. Information on individuals living in correctional facilities was included in the 1850 decennial Census (Rosen 1995). Before this national effort, a handful of states collected crime statistics using court and prison data. The first national collection of crime data using police records occurred in 1880 when the Census Bureau included homicides, arsons, and burglaries known to police as part of the 1880 decennial (Rosen 1995). Given the local level of data collection, a significant problem concerned the lack of uniformity in crime definitions across states. While this issue was acknowledged, it was not resolved with the early Census data collection efforts and these data were largely discounted (Rosen 1995).

By the 1920s, the need for national crime data had become apparent. Countries including France, Austria-Hungary, Sweden, Holland, Denmark, and Turkey were generating national crime statistics at this time (IACP 1929). Many criminologists and statisticians voiced their embarrassment that the United States lacked information on this important social indicator. In 1927, the National Crime Commission pronounced that "the United States had the worst criminal statistics of any civilized country" (as quoted by Rosen 1995, p. 220). Organizations such as the Social Science Research Council also supported the collection of crime data as a way "to provide a solid empirical base for 'scientific criminology.'" (Rosen 1995, p. 223). Police officials also demanded national statistics on the nature and extent of crime to "allay the public's undue fears" over media-created "crime waves" (Maltz 1977, p. 33).

While the need for national crime data was clear, the most appropriate way to generate these statistics in the United States was less so. Debates ensued over the appropriate data source to be used (see Rosen 1995, for a discussion). Initially interest focused on court records as police records were deemed to be unreliable (Maltz 1977). Ultimately the decision was made to utilize data closest to the criminal event, which supported the use of police records rather than court or correctional data (Rosen 1995; Maltz 1977). Along with the data source to be used, debates centered on who should collect and disseminate the data. The main two contenders were the Bureau of Investigation (the forerunner to the Federal Bureau of Investigation or FBI) and the Bureau of the Census (Maltz 1977). With the decision to collect police data, many favored the Census Bureau since it was a neutral organization (Maltz 1977). Law enforcement organization such as the International Association of Chiefs of Police (IACP) endorsed collection by the FBI (Maltz 1977). Eventually the FBI was selected due in part to lobbying by then FBI director J. Edgar Hoover (Maltz 1977).

While some scholars have questioned the emphasis on the IACP with regard to the original idea for the UCR (Rosen 1995), the IACP and its Committee on Uniform Crime Records undertook the immediate work for designing this data collection system. After the decision to use police data had been made, many technical issues remained including "the structure of the data base, the source of the data, crime measures to be used, geographical uniformity of crime definitions, rules for scoring and counting offenses, [and] the mode of data collection" (Akiyama and Rosenthal 1990, p. 50). The work to address these issues culminated in the first UCR manual in 1929. This 464-page volume devoted more than half of its text to addressing the issues of which crimes would be reported to the UCR and uniformity in crime definition and counting procedures (IACP 1929). The remainder of the volume described other aspects of the UCR program, many of which are practices that still remain today (Akiyama and Rosenthal 1990). The handbook also explained the data submission procedure. Law enforcement agencies provided handwritten tallies of aggregate counts of crime, which was consistent with technological capabilities of the time. To promote participation in the voluntary UCR system, the FBI provided the necessary forms for free as well as and return, postage-paid envelopes (IACP 1929).

Initially the UCR collected seven crimes known to police, known as Index offenses. These crimes were murder and nonnegligent manslaughter, forcible rape, robbery, aggravated assault, burglary, larceny-theft, and motor vehicle theft. Later in 1982, arson would be added as the eighth Index offense by congressional mandate (FBI 2004). The basic criteria used to select these crimes were the seriousness of the crime, the similarity of rates of occurrence throughout all geographic regions of the country, the frequency of occurrence, and the likelihood of coming to the attention of police (Poggio et al. 1985). Other crimes had been considered for collection such as carrying concealed weapons, arson, and kidnapping. The IACP explained that such crimes were excluded because "[e]ither they are concealed when committed so that the police frequently do not know they occur, or else their statutory definitions vary to such a degree it would be impossible to obtain reasonably comparable results. Unless the returns are limited to those offenses which can be relied upon to give a trustworthy picture of crime, their purpose will be defeated at the start." (IACP 1929, p. 180). The IACP did emphasize that simply because a crime was not included in the UCR, it did not prevent a local law enforcement agency from collecting data for its own records.

In 1930, the FBI began collecting UCR data based on the program outlined by the IACP. That year, 400 law enforcement agencies provided data (FBI 2004). Since that time, many changes occurred (see Barnett-Ryan 2007, for a complete accounting), but the essence of the UCR Program remained basically the same for decades including the submission of handwritten aggregate tallies (Poggio et al. 1985, p. 21). In the late 1970s, pressure mounted for the FBI to modernize the UCR in order to capitalize on both the innovations in the capability of law enforcement agencies to collect more detailed crime data as well as the improvements in the scholarly study and understanding of criminality (FBI, n.d.; Poggio et al. 1985, p. 21).

In response to these calls for an update, the Bureau of Justice Statistics (BJS) and the FBI commissioned a team of experts to reevaluate the UCR (Poggio et al. 1985). The recommended plan that came out of this review was not fully implemented. Its call for more detailed crime information at the incident level, though, did provide the foundation for what would become the National Incident Based Reporting System (NIBRS) form of data collection for the UCR. South Carolina participated in the FBI's pilot incident-level reporting program and in 1991 became the first state to submit its UCR data in NIBRS format (Barnett-Ryan 2007). By 1992, North Dakota, Iowa, and Idaho also were submitting their crime data in NIBRS format.

The NIBRS format covers a wider variety of offenses than captured by the summary report system's Index, or as they are referred to today Part I, offenses. As the FBI has moved away from a "Crime Index" toward violent and property categories, the term "Index offenses" is no longer used in favor of "Part I" offenses (FBI, n.d.). NIBRS collects incident-level details for 46 Group A offenses, which include the 8 former Index offenses. Examples of these additional crimes range from kidnapping and forcible sex offenses beyond rape (such as sodomy and sexual assault with an object) to vandalism, gambling offenses, and fraud offenses. In addition to expanding the number of crimes reported to the UCR, NIBRS captures incident-level characteristics not included in the summary reporting system. For Group A offenses, NIBRS can collect up to 53 distinct data elements to describe each criminal incident. These details include demographic information for victims, offenders, and arrestees; victim-offender relationship; information about weapons, injuries, stolen items; and clearance of the incident. NIBRS also includes information on up to 10 offenses that occur in a criminal incident rather than reporting on only the most serious offense as was previously the case under the summary system.

One of the biggest limitations with NIBRS has been the transition from summary-based to incident-based data collection. As with the summary reporting system, participation in the UCR remains voluntary under NIBRS and is not mandated by the FBI. Unlike the summary reporting system, law enforcement agencies must be certified before they are eligible to submit data in NIBRS format. In addition, states and agencies are under no deadline to convert to NIBRS. As a result of these factors, the conversion process has been gradual. As of 2012, 32 states are NIBRS certified and NIBRS agencies cover 27 % of the US population (JRSA, n.d.).

National Crime Survey-National Crime Victimization Survey

Just as the UCR played an essential role in shaping the national collection of US crime statistics, the NCVS and its predecessor the National Crime Survey (NCS) had a similar groundbreaking role. Before the NCS, no national crime statistics had been based on survey data. The idea for using victim surveys to measure crime originated out of a convergence of factors that included increased attention on crime as a social problem, decreased confidence in the reliability of UCR data, and overall advances in the field of survey methodology (see Cantor and Lynch 2000 for a discussion). In 1965, President Lyndon Johnson appointed two Presidential Commissions with charges that included "reduc[ing] the amount of crime that eluded the attention of police." (Cantor and Lynch 2000, p. 98). To accomplish these goals, the Commissions realized the need for more data to better understand and explain what appeared to be a growing crime problem (President's Commission 1967). The quality of US crime data at the time was strongly criticized by the Commission, which noted "the United States is today, in the era of the high speed computer, trying to keep track of crime and criminals with a system that was no less than adequate in the days of the horse and buggy" (President's Commission 1967, p. 123).

Victimization surveys offered a new source of data that addressed the Commissions' two pressing concerns. First, survey-based crime data could provide greater incident details than the existing aggregate-based UCR. Second, victimization data were separate from that collected by police. The unknown issue was whether victimization surveys would work, both in terms of respondents being willing to report victimization experiences and the ability to produce annual crime estimates. To test this method, the Presidential Commissions supported three sets of pilot victimization surveys: one household-based survey in Washington, DC; one group of city-based surveys; and one national sample. All of these pilots produced two important findings: (1) confirmation that victim surveys could provide reliable crime estimates and (2) identification that a significant amount of crime that occurred was never reported to the police (Cantor and Lynch 2000). These pilot surveys also helped to identify methodological issues, such as ways to facilitate recall, temporal placement, and collecting crimes against the household (Cantor and Lynch 2000).

Based on recommendations from the Commission and findings from methodological studies, the NCS was created and first implemented in 1972. Originally the NCS was a system of four victimization surveys – one national household sample (the Crime Panel), a city level household sample (the Central City surveys), and two Commercial Victimization Surveys (one at the national level and the other at the city level)

(Rennison and Rand 2007). Only the Crime Panel survived and is what is referred to here as the NCS. The original goals for the NCS concerned providing an indicator of the crime problem independent of police data and the UCR and an ongoing measure of victimization risk (Rennison and Rand 2007). The NCS also served other purposes that included shifting the focus of the criminal justice system to the victim and away from the offender and providing a measure to assess changes in reporting to police (Rennison and Rand 2007). As a result of this close tie to monitoring the UCR, the crimes covered by the NCS were closely linked with the UCR crimes. The NCS crimes included rape, robbery, aggravated and simple assault, personal larceny (pickpocketing and purse snatching), burglary, motor vehicle theft, and property theft.

The NCS was a nationally representative sample of households. Initially all household members aged 14 and older were personally interviewed to determine their victimization experiences over the past 6 months. Those household members aged 12 and 13 were interviewed by proxy. The Census field staff conducted the NCS interviews in person and over the telephone using a paper and pencil instrument. The NCS consisted of three parts: the control card, the screening questionnaire, and the incident report. The control card collected household information and respondent demographics. The screening questionnaire ascertained if the respondent experienced a criminal victimization in the past 6 months, and the incident report collected details of each victimization incident reported. Households remained in the sample for 3½ years and were interviewed at 6-month intervals. Since the sample was of households and not individuals, those who moved out of the household were not followed, but rather the new residents were included in the sample for the remainder of time the household was in sample.

As with the summary UCR system, the NCS also underwent a significant redesign effort (see Cantor and Lynch 2000, for an in-depth discussion). The redesign efforts began almost as soon as the survey went into the field (Cantor and Lynch 2000). The NCS was a significant event for both those interested in survey methodology as well as social statistics. One area of criticism came from academics who challenged the significant financial expenditure on a data system that provided limited analysis options and few variables that could be used for theory testing (Cantor and Lynch 2000). Critiques also came from the law enforcement community who questioned this new survey-based crime data collection method and the associated criticisms of police-based crime data (Cantor and Lynch 2000). At one point, the NCS had attracted enough criticism that its sponsor the US Department of Justice considered discontinuing the survey until it could be redesigned (Lynch 1990). While the survey was never suspended, it was the subject of a reevaluation by National Academy of Sciences and a 5-year program of research, instrument development, and redesign planning (Lynch 1990).

The redesign effort focused on developing a better screening instrument to promote respondent recall of victimization incidents and revising the incident report to collect additional details about the victimization event (Cantor and Lynch 2000). Other redesign work concerned research devoted to recall periods, respondent fatigue, and other methodological issues (Cantor and Lynch 2000). Research and development of these efforts began in the 1980s and the redesigned NCVS was introduced into the field in 1992. Unlike the UCR's NIBRS redesign effort which remains an ongoing project, the implementation of the redesigned NCVS instrument and methodology was completed in 18 months. The redesigned survey came with a new name: the National Crime Survey became the National Crime Victimization Survey.

The redesigned NCVS kept the same three part format (control card, screener, and incident report). Otherwise the survey had been completely overhauled, most noticeably in terms of the screener. The screener now included a series of short cues to prompt recall. In the NCS, each screener question corresponded to a particular crime, such as assault. In the NCVS, screener questions instead focused on aspects of the victimization such as where it occurred, what the respondent was doing at the time, and who committed the offense. The premise was to facilitate recall and to use various memory triggers to do so. In addition, new crimes were added as part of the screening process (vandalism and sexual assault) and specific prompts were added to better collect data on certain crimes that had been included in the NCS (rape and domestic violence). Rape and domestic violence were more clearly screened by including questions related to sexually based offenses and offenses committed by people known to the respondent. Additional crimes were also collected in supplemental questionnaires. These supplements increased after the redesign and concerned such topics as workplace violence, school crime, stalking, and identity theft. Other changes that occurred with the redesign concerned how the survey was administered with the introduction of computer-assisted telephone interviewing (or CATI) and how repeated (or series) crimes were counted. Rennison and Rand (2007) provide a comprehensive discussion of these and other changes.

The NCVS also has experienced additional changes since its redesign. Until recently, many of these changes centered on efforts to reduce costs as the NCVS has been flat funded for most of its existence since the redesign. The most significant of these changes has been the dramatic cut in sample size (Rennison and Rand 2007). Concurrent with the flat funding and sample cuts, external pressures demanded that the NCVS do more in terms of data collection and speed in releasing data. Additional questions were added concerning hate crimes, crimes against the disabled, computer crimes, and identity theft. The NCVS was also under pressure to produce its annual estimates more quickly (Rennison and Rand 2007).

The deep cuts to the NCVS program generated an unsustainable situation and a request by BJS for the National Academies' Committee on National Statistics (CNSTAT) for a panel to review the NCVS and "consider alternative options for conducting the [NCVS]." (National Research Council 2008, p. 23). The CNSTAT Panel's review found the NCVS was in need of more resources and restoration. Specifically "[a]s currently configured and funded, the NCVS is not achieving and cannot achieve BJS's legislatively mandated goal to 'collect and analyze data that will serve as a continuous and comparable national social indication of the prevalence, incidence, rates, extent, distribution, and attributes of crime . . .'" (National Research Council 2008, p. 78). The CNSTAT Panel offered many recommendations to restore the NCVS as well as an overall call to encourage appropriate funding of the survey. In response to this report, BJS began conducting a redesign effort to address the recommendations. These efforts have included restoring the sample size as well as efforts to generate subnational victimization estimates (BJS n.d.).

Similarities Between the UCR and NCS-NCVS in Their Origins and Development

While most discussions that examine both the UCR and NCS-NCVS focus on their differences and points where they diverge (see Biderman and Lynch 1991; Lynch and Addington 2007a), the two systems share many similarities with regard to their origins and development as well as the current demands and tensions under which each still operate. With regard to their origins, both the UCR and NCVS started from a need for better measures to inform the crime problem. For the UCR, law enforcement officials wanted to combat misperceptions of crime based on media-generated crime waves, and social scientists wanted empirical measures for studying crime. For the NCS, government officials and policymakers wanted to address a perceived increase in crime that could not be addressed with aggregate police data. Both also were developed as a way to obtain data that were closer to the crime problem than previously were available. For the UCR, previous crime statistics used data from courts or prisons. Decades later these police data would come under fire for being unreliable and subject to manipulation. Victimization data from the NCS provided a means to get closer to the criminal event and to avoid the police reporting filter.

In their development, both the UCR and NCS-NCVS systems underwent massive redesign efforts. While the specific changes differed, both were based on changes in technology and in improved knowledge of the crime problem. Changes in technology permitted the UCR to have the capability to collect more detailed information on crime incidents as well as to accommodate the millions of cases generated by an incident-based system. The summary reporting system originated in the 1920s when handwritten tallies were considered state of the art. By the 1980s and 1990s, police agencies had the ability to provide more sophisticated crime data as their own collection and analysis programs were increasingly becoming computerized (Poggio et al. 1985). The technology also permitted scholars, policymakers, and others the statistical and computing capabilities to manipulate and analyze these data. The NCVS similarly benefited from advances in technology that enabled CATI interviewing and improved data collection. More sophisticated statistical software packages – particularly programs that could adjust for the NCS-NCVS's complex survey sample – extended the utility of the data for researchers.

Improvements in the knowledge of the crime problem also prompted changes for both the UCR and NCS. For the UCR, incident details provided by the NCS highlighted the summary reporting system's limitations. With the available details from the NCS, researchers and policymakers began to demand the production of police data that provided incident-level information needed to explain the crime problem (Poggio et al. 1985). Solely providing aggregate level data was no longer sufficient. The NCS similarly was critiqued by scholars who demanded more details and explanatory variables to permit modeling of victimization risk and development of theories of victimization. Here the critiques came from the NCS itself. Once it was demonstrated that a victimization survey could produce reliable national estimates of crime, researchers and policymakers demanded more information. As with the UCR, annual estimates – even annual estimates coupled by additional incident details – were not enough.

Both the NCVS and UCR continue to face external demands and pressures, particularly to provide more information and to provide it more quickly. As discussed above, the NCVS has suffered the double burden of flat funding and increased demand for specific crime data and for producing these data more quickly. The UCR also has faced demands for faster release of crime data (e.g., Rosenfeld 2011) and for the collection of new crime types such as cargo theft (FBI 2011) and new crime definitions such as rape (FBI 2012). The UCR and NCVS must balance these demands for expansion and change with ongoing concerns. Both systems are charged with the task of generating annual level and change crime estimates. Data collection changes must be balanced with the need for series continuity. Both systems operate within budget and funding constraints that limit what each can do and how much data can be collected. Both also are essentially volunteer systems and must be concerned with overburdening their respondents – whether these data providers are individual law enforcement agencies and state data analysis centers (for the UCR) or individual survey respondents (for the NCVS).

Effect of UCR and NCVS on Defining What Crime "Counts"

Unlike other social indicators, no natural metric exists for crime and as a result, the classification of crime and which crimes are included is a very important decision (Lynch and Addington 2007b; much of the following discussion draws upon this chapter). Since counting every crime would be an impossible task, both the UCR and NCS-NCVS needed to identify those crimes that would (and would not) be included in their data collection systems. Literally, these decisions determined what crimes "counted." By doing so, certain crimes and aspects of those crimes are emphasized and others virtually disappear. In its summary system, the UCR elevated legal aspects of offenses and virtually excluded all other attributes of crime events. This decision was largely due to limits of police data at the time the UCR was created, but these choices continue to affect views of crime even though these same crimes might not be selected today. Given the NCS's initial focus as a check on the UCR, the crimes captured by the NCS largely mirrored those of the UCR. The redesign of both systems increased the number of crimes that were counted. NIBRS has expanded UCR crimes particularly in the areas of fraud, drug-related crimes, forcible sex crimes other than rape, and kidnapping as well as incident details that permit identification of crimes against children and those involving intimates. The NCVS expanded the crimes it collected through supplements as well as added questions to the screening questionnaire.

The decision of what crimes are counted in each system has had far-reaching effects on views of what crime is. The paradigm of crime problem is one dominated by street crime and more serious crime. This focus has largely ignored criminal activity that is less serious, but more prevalent such as bullying, vandalism, minor assaults, and threats. These crimes do not result in great loss or injury, but can be consequential to community, schools, and other social institutions as well as neighborhood cohesion and feelings of safety. In addition, such crimes may be leading indicators of more serious crime and social problems.

The UCR and NCVS, however, do not have to be limited to this paradigm of serious, street crime. As discussed by Lynch and Addington (2007b), an ignored potential of both NIBRS and NCVS is to go beyond recreating the UCR summary system's Index Crime Classification. By collecting incident-level details, the NCVS and NIBRS have the necessary information to experiment with alternative crime classifications. These alternative classifications can improve insights on crime and victimization for researchers and policymakers. Any crime classification needs to assess the risk that crime poses and does so in a way that can inform efforts to control this risk (Lynch and Addington 2007b). Rather than focusing on legal classifications, an alternative classification system could use categories that focus on relationship status, location, or activity at the time of the incident. Relationship status could compare strangers and intimates. Location could compare public and private places, and activities could focus on domains such as home, work, school, or leisure. These categories could be further defined using injury or property loss rather than legal definitions to separate violent and property crimes. Such a classification system emphasizes elements of the crime event rather than criminal behavior or legal definitions (Lynch and Addington 2007b). This scheme also directs attention away from a street crime paradigm of crime and permits more expansive consideration of the crime problem.

The NCVS and NIBRS already collect the necessary information to experiment with alternative crime classifications. Few researchers have capitalized on the opportunity to incorporate theses classifications in their work. This lack of work might be evidence of the pervasive effect of the original UCR crime classifications. Consideration of alternative classification models were mentioned in both redesign efforts. One basis for NCS redesign was to focus on victimization risk and ways for researchers to define crimes using different characteristics (Cantor and Lynch 2000). Designers of NIBRS identified several benefits of an incident-based system, one of which focused on the analytical flexibility that would allow "users to count and categorize crimes in ways they find meaningful" as well as "to explore a myriad of details about crime and law enforcement" (Poggio, et al. 1985, p. 4).

Conclusion

The importance of the UCR and NCS-NCVS cannot be emphasized enough, especially how they – separately and together – have shaped what is known about crime. These two measures of crime provide a rich source of information on trends over time, annual estimates of the crime problem, and details about the incident. While discussions of the UCR and NCS-NCVS often focus on the differences between the two and points where they diverge (Biderman and Lynch 1991; Lynch and Addington 2007a), the two systems share many similarities in their development over time. Both likely will continue to evolve over time as technology advances, as more is learned about crime, and as the demand for more detailed data continues.

The UCR and NCS-NCVS also have affected views of crime by defining what crimes count (and which do not). The UCR originally was viewed as speaking primarily to law enforcement agencies. It was not until its move to NIBRS were the needs of outside researchers and policymakers specifically acknowledged (Poggio et al. 1985). The NCS originated out of a need, in part, to monitor the UCR. Since the NCS's redesign, calls have been made for the NCVS to turn away from this monitoring function (Cantor and Lynch 2000). With a change in focus for both systems, attention could be devoted to ways to address the crime problem through utilizing alternative crime classification systems as well as exploring techniques for measuring crimes that affect citizens on a more routine basis than serious violent street crimes.

Recommended Reading and References

Akiyama Y, Rosenthal HM (1990) The future of the uniform crime reporting program: its scope and promise. In: MacKenzie DL, Baunach PJ, Roberg RR (eds) Measuring crime: large-scale, long-range efforts. SUNY Press, Albany

Barnett-Ryan C (2007) Introduction to the uniform crime reporting program. In: Lynch JP, Addington LA (eds) Understanding crime statistics: revisiting the divergence of the NCVS and the UCR. Cambridge University Press, Cambridge, pp 55–89

Biderman AD, Lynch JP (1991) Understanding crime incidence statistics: why the UCR diverges from the NCS. Springer, New York

BJS (n.d.) NCVS redesign. http://bjs.ojp.usdoj.gov/index.cfm?ty=tp&tid=91. Retrieved 4 Feb 2012

Cantor D, Lynch JP (2000) Self-report surveys as measures of crime and criminal victimization. In: Duffee D (ed) Measurement and analysis of crime and justice. National Institute of Justice, Washington, DC, pp 85–138

Federal Bureau of Investigation (2004) Uniform crime reporting handbook. FBI, Washington, DC. http://www.fbi.gov/about-us/cjis/ucr/additional-ucr-publications/ucr_handbook.pdf. Retrieved 4 Feb 2012

Federal Bureau of Investigation (2011) Cargo theft electronic data submission specifications. http://www.fbi.gov/about-us/cjis/ucr/Cargo_Theft_Elec_Submission_Specs_2011.pdf. Retrieved 4 Feb 2012

Federal Bureau of Investigation (2012) Revised rape definition announced for uniform crime reporting program. http://www.fbi.gov/news/news_blog/revised-rape-definition-approved-for-uniform-crime-reporting-program. Retrieved 4 Feb 2012

Federal Bureau of Investigation (n.d.) UCR general FAQS. http://www.fbi.gov/about-us/cjis/ucr/frequently-asked-questions/ucr_faqs. Retrieved 4 Feb 2012

International Association of Chiefs of Police (1929) Uniform crime reporting: a complete manual for police, 2nd edn. J.J. Little & Ives, New York

Justice Research and Statistics Association (n.d.) Status of NIBRS in the states. IBR Resource Center, http://www.jrsa.org/ibrrc/background-status/nibrs_states.shtml. Retrieved 12 Feb 2013

Lynch JP (1990) The current and future national crime survey. In: MacKenzie DL, Baunach PJ, Roberg RR (eds) Measuring crime: large-scale, long-range efforts. SUNY Press, Albany

Lynch JP, Addington LA (2007a) Understanding crime statistics: revisiting the divergence of the NCVS and the UCR. Cambridge University Press, Cambridge

Lynch JP, Addington LA (2007b) Conclusion. In: Lynch JP, Addington LA (eds) Understanding crime statistics: revisiting the divergence of the NCVS and the UCR. Cambridge University Press, Cambridge, pp 297–334

Maltz MD (1977) Crime statistics: a historical perspective. Crime Delinq 23:32–40

National Research Council (2008) Surveying victims. National Academies Press, Washington, DC

Poggio EC, Kennedy SD, Chaiken JM, Carlson KE (1985) Blueprint for the future of the uniform crime reporting program: final report of the UCR study. U.S. Department of Justice, Washington, DC

President's Commission on Law Enforcement and Administration of Justice (1967) Task force report: crime and its impact – an assessment. U.S. Government Printing Office, Washington, DC

Rennison CM, Rand M (2007) Introduction to the national crime victimization survey. In: Lynch JP, Addington LA (eds) Understanding crime statistics: revisiting the divergence of the NCVS and the UCR. Cambridge University Press, Cambridge

Rosen L (1995) The creation of the uniform crime report. Soc Sci Hist 19:215–238

Rosenfeld R (2011) The big picture: 2010 presidential address to the American society of criminology. Criminology 49:1–26

Developmental Criminology

► Integrated Cognitive Antisocial Potential Theory

Deviance Theory

► Transitional Justice

Differences-in-Differences in Approaches

James H. Derzon
Battelle Memorial Institute, Centers for Public Health Research and Evaluation, Arlington, VA, USA

Synonyms

Randomized control trials; RCTs

Overview

Although almost universally acknowledged as the most powerful research design for causal attribution, the randomized control trial (RCT) is not without limitations in applied research settings. Foremost among those limitations is the generalizability of study findings due to vicissitudes in tested sample, setting, intervention, and result. If generalizability cannot be assumed, the value of the finding as evidence for future action may be tempered.

In addition, several practical and theoretical limitations may be imposed by the RCT. Probabilistic equivalence on all measured and unmeasured characteristics is difficult to assume in many applied settings. The ethical requirement of clinical equipoise for many interventions mathematically limits discovery of effective interventions to 25–50 % of trials. RCT research tends to produce evidence on which recipients respond to the intervention being tested and not produce evidence on the intervention characteristics responsible for that response.

The preference for RCTs may bias what interventions are studied and what evidence is considered. Many interventions are not amenable to randomization, and evidence of their effectiveness may be discounted if only evidence from RCTs is valued. If only RCTs are valued, the increasingly available archival and electronic records data is similarly discounted as a source of evidence of what works as randomization must occur prior to intervention. Finally, as a design for causal proof, RCTs tend to focus solely on program effectiveness and not on other considerations of implementation and maintenance; as a prospective design, with measures often established prior to the research being conducted, it may be a design ill-suited to observing and documenting unintended consequences.

Two additional limitations result from a hyperallegiance to RCTs for evidence. When RCTs are applied in real-world settings, they often assess the links between complex interventions and the individual mediators and moderators of distal outcomes using complex statistical models and multivariate methods. Including such findings in formal meta-analyses is difficult and undermines the accumulation of evidence. Also reducing the accumulation of evidence is the limited number of researchers qualified to conduct formal RCTs.

To supplement the evidence available for evidence-based practice, an effectiveness-surveillance system using a synthetic difference-in-differences design is proposed. Such a system would provide associational evidence of effectiveness and provide a valuable complement to the evidence generated by RCTs.

The RCT: A Powerful Design for Cause

In an ideal world, or the idealized world of the laboratory study, it is difficult to argue against the logic of the RCT for drawing causal inferences. In its simplest form, a sufficient number of units (typically individuals) are selected to be representative of a known population and are randomly assigned to either a treatment or a control condition. Those in the treatment condition are then exposed to a stimulus (also known as the manipuland, independent variable, treatment, or intervention) that is withheld from those in the control condition. Under the logic of the design, the performance of the control group approximates what would have been the performance of the experimental group in the absence of the stimulus being tested. Since the experimental group cannot simultaneously both receive and not receive the stimulus, scientists rely on inference to attribute differences in performance to the presence and absence of the stimulus being tested. If, in a well-controlled trial, the outcome covaries with the presence and absence of the stimulus, then a causal relationship is generally inferred.

Since first proposed in its modern form by Sir Austin Bradford Hill (1952), the RCT as the premiere method of causal inquiry has grown in both importance and stature. RCTs reside below systematic review and meta-analysis at the peak of evidence-based medicine's evidence pyramids and have been granted priority by the Department of Education (Scientifically-based evaluation methods 2003). Well-implemented randomization creates two groups with "initial probabilistic equivalence," that is, "groups that are comparable to each other within the known limits of sampling error" (Cook and Campbell 1979, p. 341). This, in turn, allows the researcher to assume that differences in outcome are attributable entirely to the intervention and are not influenced (confounded) by preexisting subject characteristics. Moreover, if the sample prior to randomization is representative of a population, randomization allows the result of an RCT to be generalized to the population represented by the sample.

While the RCT is a powerful strategy for attributing cause in many research settings and has been effectively applied to answer many important research questions, it is merely one feature of a well-designed experiment. In and of itself, randomization does not ensure valid or reliable findings nor ensure inferences based on those findings are correct. Well-conducted randomization creates covariate-balanced groups equal in their probability of displaying the outcome(s) of interest and differing only in their receipt or nonreceipt of the treatment/intervention/stimulus being tested. What randomization does not control for are other vicissitudes of the research endeavor. These can include sample selection; research methods and procedures; the use of sensitive, reliable, and valid measures; application of appropriate statistics; appropriate interpretation of findings and their limitations; and, finally, complete and accurate reporting of the study and possible threats to the validity of the research findings, as discussed in detail elsewhere (e.g., Cook and Campbell 1979; Lipsey and Cordray 2000; Shadish et al. 2002). In other words, when applied to real-world research in real-world settings, unequivocal belief that in and of itself randomization is sufficient to produce credible evidence is not tenable.

Before accepting a knowledge claim resulting from an RCT (or any other evidence claim, for that matter), all aspects of the study design and execution must be examined and judged. As an exercise in applied logic, all parts of the research endeavor work together to form a conclusion. The result produced from a single well-conducted RCT has tremendous value and, arguably, no other single study design produces evidence of equal merit. However, the intrinsic value of the RCT design for generating knowledge should not be confused with the extrinsic value of the design for generating actionable knowledge (i.e., its capacity to document "worth"). Moreover, its elevation to the "the gold standard" for knowledge generation has had the unintended consequence of discounting other methods of inquiry (Gugiu and Gugiu 2010). The discussion that follows provides several potential, practical, and documented limitations of relying on RCTs for generating robust actionable knowledge in criminology and criminal justice, with some of these limitations also applying to quasi-experiments. This entry concludes by offering an alternative design for knowledge accretion that may produce evidence of sufficient merit and worth to meet the expanding need for evidence of effectiveness of what works, for whom, and under what conditions. In providing this perspective, we add to the already robust literature documenting the limitations of RCTs (e.g., Boruch 1975; Pawson 1994).

Key Issues/Controversies

In this section we discuss some of the key issues which may limit the extrinsic value of the RCT for creating actionable knowledge. After a brief introduction on the lack of standards for assessing external validity and generalizability, potential limitations of different aspects of the RCT for creating generalizable evidence are discussed. The section closes with brief discussions of a number of practical and theoretical limitations of relying on the RCT for knowledge generation.

Lack of Standards for Assessing Generalizability

Through its control of confounds, a sufficiently large and well-implemented RCT undoubtedly provides greater internal validity and better warrant for attributing results to the intervention tested than do other study designs. Moreover, there are clear standards for evaluating study quality, and these standards allow us to judge the internal validity of the knowledge claimed by the trial. However, the purpose of science is the identification of generalizable conclusions (Overton 1998), and "conclusions limited to a specific subset … are not scientifically informative" (Hunter and Schmidt 2000, p. 277). External validity estimates the extent to which the results obtained by a study generalize to an unstudied population; if subjects in an RCT are representative of the unstudied population, study results are likely to replicate well in these unstudied populations. Unfortunately, there are few established guidelines for estimating external validity (Rothwell 2005) and, as estimated in the social sciences, what guidelines do exist often seem reduced to simple sample demographics. In contrast, Rothwell postulates 39 issues that potentially affect external validity for clinicians to consider when evaluating clinical practice RCTs in medicine.

Potential Limitations of Generalizability

Below we discuss several potential limitations of generalizability in an RCT, particularly those that are designed to produce evidence of efficacy. These concerns, in particular, have been proposed as central to the failure of efficacy research to translate into effective behavioral and health promotion efforts (Glasgow et al. 2003). These limitations relate to generalizability of the sample, the setting, the intervention itself, and the result.

Generalizability of the Sample. How representative of unstudied populations are subjects who agree to participate in an RCT? For many practical reasons, study populations are likely selected on their proximity to research staff. If research-intensive settings (e.g., university towns) are fundamentally different from other settings, it may limit generalizability. Subjects in such settings, and those who participate in research generally, are likely more motivated to seek intervention than those not interested in participating and are willing to be randomized to a control condition in which they will receive no, limited, or delayed services. This willingness is ensured through the informed consent procedure, in which subjects voluntarily sign a plain-language form attesting to the fact that they understand the implications of participation. For many interventions this may not be problematic, but for efficacy research in which the sample is self-selected, homogeneous, highly motivated, and screened to control for comorbid conditions, it is reasonable to question the generalizability of the sample and if it is, in fact, representative of the population it claims to represent.

Generalizability of the Setting. In addition to much research being conducted in settings convenient to the researchers, it should be observed that all interventions occur within a political, economic, social, and cultural context. RCTs can be very effective at controlling these potential confounds within a study, but ignoring these external community-level characteristics when generalizing beyond the studied sample may ignore consequential causal factors. Simply put, even if demographic characteristics of the sample included in the RCT are similar to an unstudied population, generalizability will be compromised to the extent individual responsiveness to intervention is influenced by the political, economic, social, and cultural context of the community in which the intervention is tested relative to the community to which the findings are hoped to generalize. Since variance is often controlled by limiting an RCT to one particular setting, such trials provide no way of estimating if or how the unique circumstances of that setting influenced results. These potential influences are rarely discussed in the limitations sections of articles, and their consequence is likely unappreciated in federal initiatives that require implementation of science-based programs whose efficacy may have been established in a limited setting.

Generalizability of the Tested Intervention. Randomization controls for confounding by preexisting subject characteristics, but other factors associated with implementing and delivering the intervention can bias and limit the generalizability of a result. RCTs are often used to test whether an intervention is efficacious (is it effective under optimal conditions?); whether the intervention is effective under real-world conditions is an entirely different question. Often the intervention tested by an RCT is not the intervention that gets disseminated or adopted; developers often refine their interventions based on the results of their research, and even simple interventions are amalgams of activities that are typically adapted to local conditions (Durlak and DuPre 2008). If the intervention being implemented lacks fidelity to the intervention that was tested, the implemented intervention cannot be considered evidence-based.

Generalizability of the Result. The effect of a given cause is captured by the findings of a well-conducted RCT. Since findings are based on a comparison of outcomes in each arm of the experiment (e.g., a *t*-test that compares the means and standard deviations of each condition), each arm contributes equally to the observed finding. A fundamental implication of this assertion is that both conditions of an experiment must be fully specified (Reichardt 2011). Only if both arms of the experiment are fully understood and explicated can the results of an RCT be interpreted. This counterfactual definition of effect is known as the Rubin model of causality (Rubin 2005) and is integral to experimental design. Unfortunately, one seldom reads even a cursory description of what services were received by the counterfactual sample.

Although reliable statistics are lacking, it is likely that many RCTs in applied settings are corrupted by compensatory services within the comparison group. For example, a multisite randomized control study for high-risk youth (HRY) found that 21 of the 46 randomly assigned “control” sites (45.7 %) received more prevention services than the sites randomized to receive the HRY intervention. An initial analysis concluded that the HRY intervention had no effect, whereas statistically controlling for compensatory exposure disclosed a strong and significant positive effect (Derzon et al. 2005). Depending on which counterfactual is assumed, both results are arguably correct (the intervention was ineffective relative to other services available and more effective than not receiving prevention services). Note that if compensatory services had not been measured, an incorrect conclusion could have been that the intervention was not effective. If the counterfactual is not defined and not documented as carefully as the intervention, it is hard to imagine how the result of an RCT can be understood.

Practical and Theoretical Limitations to the RCT

Practical and theoretical limitations to the RCT are discussed below. These include assuming initial probabilistic equivalence, clinical equipoise, decisions about what evidence is considered, and limitations to knowledge accretion.

Assuming Initial Probabilistic Equivalence. Recall that the purpose of randomization is to create initial probabilistic equivalence. This is an idea that is closely tied to the theory of the intervention, its assumed causal relationship with the outcome, and the potential for other factors to produce the outcome or that influence subjects’ response to the intervention. As subject variability in response increases, confidence intervals tend to increase and commensurately larger samples are necessary for statistical power and avoiding a type II error (not finding an effect for interventions that are, in fact, effective). Thus, for interventions in which measures are robust and the causal relationship is mechanistic and deterministic and the effect is produced only by the cause being tested, the assumption of initial probabilistic equivalence may be strongly supported. However, if there are multiple sufficient causes to produce the outcome (overdetermination), or if the theory and measures used account for only part of the variability in outcome (underspecification), the assumption of initial probabilistic equivalence – especially in small-sample studies – may be considerably weakened. In such cases, randomization may not create the covariance balance necessary to assume initial probabilistic equivalence, and biased results are possible. Although RCTs provide a strong defense against the violation of this assumption, in criminology, and in evaluation research generally, there is a long history of null findings that are, in part, attributable to a lack of statistical power (Lipsey 1990).

Clinical Equipoise. From a researcher's perspective, uncertainty and clinical equipoise are moral imperatives for the conduct of ethical research (Weijer et al. 2000; cf. Veatch 2007). Equipoise is defined as genuine uncertainty as to which arm in an experiment is likely to prove more effective. Once a treatment is known to be effective, it is generally considered unethical to withhold that treatment. Mathematically, if there is true uncertainty of the effectiveness of the practice being tested relative to its alternative (some will be superior, some inferior, and some equal), adherence to equipoise limits progress in advancing evidence-based practice to discovering 25–50 % of successful treatments when they are tested in RCTs (Djulbegovic 2009). To the extent equipoise is achieved, the ethical implementation of RCTs may sorely constrain the advancement of knowledge.

What Evidence Gets Considered? The RCT focuses attention on the units that can be randomized and on interventions that are amenable to randomization. Thus, much of the RCT research focuses on recipient characteristics and who responds to the intervention being tested (e.g., publications that present multivariate evidence accounting for individual differences in displaying the outcome). Without minimizing the value of this information – which is useful for intervention targeting – such results provide little insight for improving service delivery or identifying the active ingredients of multicomponent interventions. Since service providers control these aspects of interventions, it seems that equal attention to these features is warranted. Multisite randomized trials can provide such evidence, but these are expensive and rare. Meta-analysis and systematic reviews of RCTs may also provide that evidence, but building the evidence base necessary is dependent on sufficient reporting of these details from the original trial. In sum, while the typical RCT may tell us for whom the intervention is effective, it may provide little information to advance our understanding of how and why interventions achieve their results.

A second implication of a commitment to RCTs as "gold standard" evidence may be an a priori filtering of the kinds of interventions that can have "gold standard" evidence of effectiveness. For example, environmental interventions are often not amenable to randomization. Thus, if only RCTs are accepted as credible evidence, our understanding of the impact of these interventions will be highly constrained. For example, Cozens, Saville, and Hiller in 2005 published a review of the effectiveness of crime prevention through environmental design (CPTED) strategies. Interestingly, only one title cited in the review's bibliography contains the word "randomized." Yet, while many studies of environmental interventions cited in this review do not meet the gold standard, the results that are included are encouraging. Equally important, as a practical matter, relative to individual-based interventions, the fidelity of environmental interventions does not degrade once they are implemented, nor do such interventions require booster sessions and additional resources to maintain impact. Had Cozens et al. dismissed evidence for CPTED practices not obtained through RCTs, their review would likely have been much shorter and a good deal of knowledge would have been lost to the field.

A third implication is that, by definition, RCTs are prospective studies. Randomized assignment must occur prior to the intervention, and this requirement obviates the use of retrospective data. Particularly in medicine – although they are increasingly available in other domains – electronic records provide a potential wealth of data for conducting comparative effectiveness studies. Commensurate with this growth are statistical matching methods that are increasingly robust in producing initial probabilistic equivalence permitting causal inference in retrospective and observational studies (Stuart 2010). Prospective data collection is expensive, and even small RCTs often require considerable costs to build effective sampling frames, to create buy-in with both controls and subjects, and to implement with fidelity. Providing incentives to a control or comparison group that receives limited, no, or delayed services may also be a nontrivial expense. The cost of collecting these data may, in part, explain the propensity of RCTs to focus almost exclusively on documenting intervention effectiveness, while often ignoring or omitting many other issues of concern to program adopters (such as cost, training, local acceptance of intervention theory and activities, and other issues of implementation or maintenance).

Finally, unintended consequences are often not discussed in RCT research. This is not to suggest that RCTs cannot collect and report this information, but merely an observation that RCTs typically focus on establishing the efficacy or effectiveness of a program through a circumscribed set of measures. These measures are defined by a logic model developed in advance of the trial and this focuses attention on the measures adopted prior to the start of a trial. If not associated with the measures defined in advance, unintended consequences may be ignored and are almost certainly unmeasured, and a biased understanding of the impact of the intervention may be promulgated. Morell (2005) offers some reasons for unintended consequences and guidance on how evaluation methodologies can be improved to account for these unforeseen and/or unforeseeable consequences.

Limitations to Knowledge Accretion. One of the major limitations of relying solely on RCTs for generating knowledge surrounds the accumulation of knowledge about effective practices. RCTs are best suited for establishing causal linkages between discrete phenomena (a manipulation and an outcome) that are linked by simple, short causal chains (Victora et al. 2004). In addition to the probabilistic constraints of equipoise, when RCTs are applied in real-world settings, they often assess the links between complex interventions and the individual mediators and moderators of distal outcomes. One method to account for that complexity is to use multivariate methods to report conditioned findings. That is, propensity adjustment may be used to account for selection bias, and multiple regression or other forms of statistical modeling are used to analyze the effects of the intervention. There are often good theory-building reasons for doing this, but because the partial or semipartial estimate of impact is determined by the variables in the model, findings from these studies cannot easily be used to examine the stability of evidence across multiple trials unless each researcher modeled the same variables. Suffice it to say, they do not, and the simple main effects that could be included in evidence synthesis are often not reported (Alford and Derzon 2012).

Also limiting knowledge accretion may be the number of researchers available to conduct RCTs and other experiments in criminology and criminal justice. Although the number of PhDs granted in the USA in criminology and criminal justice more than doubled between 1997–1998 and 2007–2008, still only 104 PhD degrees were granted in 2007–2008 (ASA, 2010). The June 2012 newsletter of the American Society of Criminology's Division of Experimental Criminology boasts 167 active members (AEC/DEC, 2012). While these numbers certainly do not represent the total number of researchers conducting or capable of conducting RCTs in criminology and criminal justice, it does seem fair to question whether the tremendous need to produce evidence, and to conduct original research and replications across all the varied settings and populations that exist, can be met by the supply of researchers available to conduct RCTs, or the funding available to support researchers conducting RCTs.

In the end, it is fair to ask how successful RCTs are in providing the evidence necessary to support evidence-based practice. Several federal and privately funded groups now conduct evidence reviews in which they evaluate the methods, procedures, and internal validity of RCTs to make best-practice recommendations. The Center for the Study and Prevention of Violence (CSPV, 2012) summarizes the results of 12 such efforts. Programs supported by RCTs that meet commonly accepted criteria for providing credible evidence are recommended as model, exemplary, effective, or perhaps promising programs. Although the list of effective programs compiled by CSPV is long (over 450 programs), when broken out by outcome or contrasted with the number of programs studied, the list takes on a different character. For example, of the 491 programs included on CSPV's list in 2010, only 46 programs assessed a school-based ATOD or violence program, had implementation materials, and had been tested in two or more minimal quality trials (Alford and Derzon 2012). The Blueprints for Violence Prevention project at CSPV has examined over 900 programs to identify 11 effective and 22 promising violence prevention programs (CSPV, 2012). The problem is not limited to drug or crime studies. The US Department of Health and Human Services' Office of Adolescent Health screened 1,000 studies to identify 28 programs that likely reduce the probability of teen pregnancy or birth (DHHS/OAH, 2011). As another example, the US Department of Education's Institute of Education Sciences created the What Works Clearinghouse in 2002 and has screened at least 461 programs to improve academic achievement. Of these, 5 programs showed positive effects with a medium-to-large evidence base (DOE/IES, 2011).

The low ratio of programs of documented effectiveness relative to the number of programs examined suggests that relying on gold standard RCTs is not an efficient or a particularly effective standard on which to build evidence for science-based practice. It is possible that much of the evidence identified or submitted in support of these programs was not eligible because it was not obtained using an RCT, or that the RCTs submitted did not meet study quality inclusion criteria. Regardless, these numbers represent an astonishing paucity of evidence relative to need. As a corollary, how much skill, time, and money will be required to conduct high-quality RCTs that meet these study quality standards as valid, reliable, robust, and generalizable information for practice?

An Alternative: Seeing Beyond the Trees

The commitment to science-based intervention does not require a commitment to the RCT as the only path to identifying effective strategies for social betterment (Concato et al. 2000). It may be required for causal proof, but this is a standard that should follow and not lead the development of science-based practices. Alternative approaches are tenable and have proven useful. Key among these are natural experiments (e.g., difference-in-differences designs, regression discontinuity), adjustment for selection (e.g., matching, case-control, regression, fixed effects [sibling/person as own control], propensity scoring, and doubly robust estimation), and instrumental-variables analysis. Because of its applicability to criminology, the difference-in-differences approach is discussed in greater detail below.

Difference-in-differences (DID) approaches are natural experiments that contrast change in outcome for the intervention group (or groups) over time to change in one or more nonrandomly identified comparison groups. It is a popular design in criminology that relies, as perhaps all comparative research does, on the plausibility of the attribution that differences observed can be attributed to the intervention (Victora et al. 2004).

Given the many reasons outlined throughout this entry and the pressing need to develop evidence of what works in real-world settings, it seems an appropriate moment in the development of science to consider how advances in scientific methods, accountability in social programming, and information technologies can be harnessed for science to improve the human condition. Increasingly, social programs are tracking and monitoring performance data for program improvement and accountability. The ubiquity of the personal computer, programs such as Microsoft Excel, and the development of the internet have made simple data management and the sharing of information over great distances possible. The past 20 years have seen a tremendous growth in the techniques and acceptance of meta-analytic methods for summarizing and analyzing findings generated by diverse instruments across uncoordinated settings. What is proposed is that these three phenomena be exploited to invert the phase model (e.g., Flay 1986; Greenwald and Cullen 1985) commonly adopted for advancing science to create a surveillance system that uses systematic data collection across multiple intervention settings to identify promising practices.

Using a synthetic DID model (S-DID) with distributed data collection and centralized analysis, local programs and evaluators would collect and submit summary performance data (i.e., pretest and posttest results) – and other relevant metadata (describing, e.g., the intervention, setting, and sample) – to a central repository that could then standardize those data and allow user-generated comparative effectiveness analysis. That is, findings from similar practices can be pooled to estimate average performance and the range of performance across multiple implementations, and these findings can be contrasted with performance scores from alternative practices. A user interface that graphically and quantitatively summarizes and displays the available evidence would give the local user tools to judge the breadth, depth, and consistency of evidence for each practice. Since findings are cumulative, the more settings that contribute, the more robust the estimates available. Homogeneous distributions of results are explained by sampling error, and metadata (user-provided information on setting, sample, and treatment characteristics) can be used by more sophisticated users to explore reasons for heterogeneity.

Persistent heterogeneity can be explored by adding categorical metadata variables to the system and entering results separately for each category. For example, if an intervention is hypothesized to be effective in urban but not rural settings, an urban/rural discriminator can be added to the metadata to distinguish subsequent submissions. As data in the two categories accrete, the homogeneity of evidence for each distribution can be tested and the results of the two distributions can be contrasted to confirm or refute the hypothesis. Hypothesized mediators that are continuous can likewise be entered as metadata and their influence on the outcome tested using correlational approaches.

Given the limitations of RCTs, the advantages of such a system are numerous. Interventions are tested in the settings in which they were adopted, and the diversity of adaptations, settings, and subjects provides an empirical basis for discerning what works where, when, and for whom. Because the diversity of interventions contributing to such a system is unlimited, opportunities for conducting ad hoc comparative effectiveness research – in near real time – are likewise unlimited. Since primary data are already being collected for performance-monitoring purposes, the cost of maintaining such a service, once engineered, would likely be modest. But the greatest benefit by far is that, by allowing local implementers and local evaluators to contribute evidence, the system democratizes evidence generation: all users who contribute create the evidence necessary for identifying evidence-based best practices. Interventions in real-world settings are tested, and the opportunity for innovation is distributed to all intervention practitioners.

What Is Required?

At the local level, data on outcomes are collected from service recipients. Summary statistics are generated for both pretest and posttest (this keeps data collection relatively unobtrusive and maintains subject confidentiality). Descriptive data that describe the sample, setting, activities, and context of the intervention are provided by program administrators, agents, or local evaluators.

Using a web interface, these locally derived data are then entered into a centralized data

Differences-in-Differences in Approaches, Fig. 1 An example of the possible output from S-DID

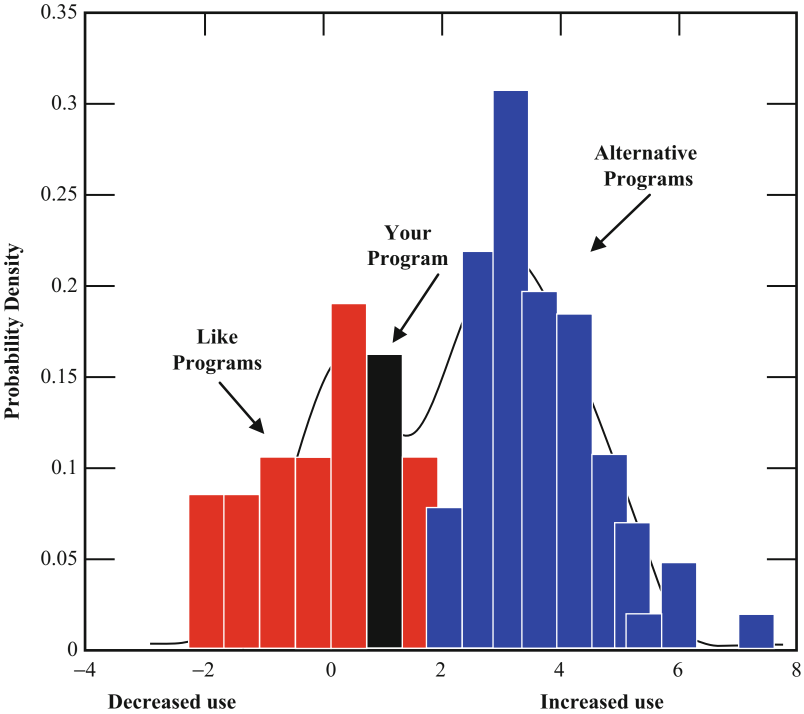

repository that permits the user to see how their change score compared with other change scores from similar interventions and then map that distribution against alternative interventions (see Fig. 1). Evidence for generalizability and testing of potential confounds could be accomplished using similarity scores (e.g., clustering based on proximity scores, with a commensurate loss of data from filtering) or meta-regression.

Limitations of S-DID

For all its potential, the S-DID approach is not without its own set of limitations. Foremost is the fact that it is a system dependent on distributed data collection. If the infrastructure necessary to support such a system were built, would local implementers and evaluators upload the data necessary for the system to work? Such a system is worthless unless populated with data from a variety of settings and samples. Moreover, would the data entered be reliable and valid? Although the summary statistics necessary to populate an S-DID database are fairly basic (e.g., means and standard deviations), whether local implementers and evaluators accurately collect data and convert their primary data into the necessary summary statistics is a nontrivial consideration. If the data entered are biased, flawed, or otherwise corrupt, they will be of little value to advancing knowledge.

Because evidence from the proposed system comes from multiple sites, it ameliorates, but does not obviate, the possibility of misattribution. Only the RCT, well-implemented, controls for third-factor influences and permits causal attribution. In this system, intervention selection is not random, and there may be unmeasured factors correlated with intervention selection that drive the result. Nonetheless, as interventions are tested across multiple settings and samples, the evidence generated in this system is self-correcting; if third factors determine the result, their impact will become apparent over time and through follow-on implementations. In exchange for causal attribution, the S-DID system allows plausible attribution and an early warning system to identify both likely effective and likely ineffective practices.

Differences-in-Differences in Approaches, Table 1 Comparison of RCTs with S-DID approaches to building evidence

Type of limitation	Randomized control trials	Synthetic difference-in-differences model
Participation	Only highly trained researchers produce effectiveness evidence	Practitioners and local programs produce effectiveness evidence
Politics	Proof of efficacy leads to approved solutions	Commitment to problem, not solutions
Power	Often underpowered and insensitive to impact	Pooling evidence increases the ability to detect effects that are present
Practicality	Distinguished by uniqueness and contribution to new knowledge	Only adopted interventions are tested
Price	Costs can be extraordinary, and much of the cost goes to not providing services	Costs likely minimal. Performance data are repurposed, remaining data are provided by program staff
Principles	Internal validity – did the intervention cause the result?	External validity – whether the intervention works, under what conditions, and for whom
Possibility	Burden of proof is "beyond a reasonable doubt" (e.g., $p < 0.05$), confounded by a publication bias toward successful trials	Burden of proof is the "preponderance of evidence" and the performance distribution of multiple implementations

Can It Work?

Two efforts currently under way are using the proposed evidence system to identify best practices. The CDC, through its Laboratory Science, Policy and Practice Program Office, is using both published and partner-submitted quality improvement data to identify best practices for laboratory medicine (https://www.futurelabmedicine.org/about/; Christenson et al. 2011), and Bill Hansen and the author are developing a web-based system for managing local substance use prevention trial data (Evaluation-based Effectiveness Testing; DHHS Grant # DA026219). The National Evaluation of Safe Schools/Healthy Students used an S-DID approach to estimate the correlates of effectiveness in a sample of 59 grantees (Derzon et al. 2012). Across a variety of applications, the approach is providing evidence of practices associated with effectiveness.

In sum, S-DID circumvents many of the limitations of traditional RCT research (see Table 1). It is not offered as a panacea for all the limitations of RCT research, but may be expected to supplement RCTs in our quest to understand what interventions are likely to be effective, under which conditions interventions tend to be effective, and for whom those interventions may be effective.

Related Entries

- ▶ Cambridge-Somerville Youth Experiment
- ▶ History of Randomized Controlled Experiments in Criminal Justice
- ▶ Place-Based Randomized Trials
- ▶ Propensity Score Matching
- ▶ Randomized Block Designs
- ▶ Randomized Experiments in Criminology and Criminal Justice
- ▶ Rational Choice Theory

Recommended Reading and References

Academy of Experimental Criminology, Division of Experimental Criminology (AEC/DEC) (2012) Newsletter. Academy of Experimental Criminology/The Division of Experimental Criminology Newsletter 7(1):1. Available at: http://gemini.gmu.edu/cebcp/DEC/AECDECJune12.pdf. Accessed 9/20/2012

Alford AA, Derzon J (2012) Meta-analysis and systematic review of the effectiveness of school-based programs to reduce multiple violent and antisocial behavioral outcomes. In: Jimerson S, Nickerson A, Mayer M, Furlong M (eds) Handbook of school violence and school safety, 2nd edn. Routledge, New York, pp 593–606

American Sociological Association (ASA) (2010) Report of the ASA task force on sociology and criminology programs. American Sociological Association, Washington, DC. Available at: www.asanet.org. Accessed 8/7/2012

Boruch RF (1975) On common contentions about randomized field experiments. In: Boruch RF, Riecken HW (eds) Experimental testing of public policy: the proceedings of the 1974 Social Science Research Council Conference on Social Experiments. Westview, Boulder, CO

Center for the Study and Prevention of Violence (CSPV) (2012) Center for the study and prevention of violence: blueprints for violence prevention. Available at: http://www.colorado.edu/cspv/blueprints/index.html. Accessed 9/7/2012

Christenson RH, Snyder SR, Shaw CS, Derzon JH, Black RS, Mass D, Epner P, Favoretto AM, Liebow EB (2011) Developing laboratory medicine best practices: systematic evidence review and evaluation methods for quality improvement. Clin Chem 57:816–825

Concato J, Shah N, Horwitz RI (2000) Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med 342: 1887–1892

Cook TD, Campbell DT (1979) Quasi-experimentation: design and analysis issues for field settings. Houghton Mifflin, Boston

Cozens PM, Saville G, Hillier D (2005) Crime prevention through environmental design (CPTED): a review and modern bibliography. Property Manag 23:328–356

Department of Education, Institute of Education Sciences (DOE/IES) (2011) What works clearinghouse. Available at: http://ies.ed.gov/ncee/wwc/. Accessed 12/14/2011

Department of Health and Human Services, Office of Adolescent Health (DHHS/OAH) (2011) Evidence-based [teen pregnancy prevention] programs. Available at: http://www.hhs.gov/ash/oah/oah-initiatives/tpp/tpp-database.html. Accessed 12/14/2011

Derzon JH, Yu P, Ellis B, Xiong S, Arroyo C, Mannix D, Wells ME, Hill G, Rollison J (2012) A national evaluation of safe schools/healthy students: outcomes and influences. Eval Prog Plann 35(2):293–302, 10.1016/j.evalprogplan.2011.11.005

Derzon JH, Springer F, Sale L, Brounstein P (2005) Estimating intervention effectiveness: synthetic projection of field evaluation results. J Prim Prev 26: 321–343

Djulbegovic B (2009) The paradox of equipoise: the principle that drives and limits therapeutic discoveries in clinical research. Cancer Control 16:342–347

Durlak J, DuPre E (2008) Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. Am J Commun Psychol 41(3):327–350, 10.1007/s10464-008-9165-0

Flay BR (1986) Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. Prev Med 15:451–474

Glasgow RE, Lichtenstein E, Marcus AC (2003) Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. Am J Public Health 93(8): 1261–1267

Greenwald P, Cullen JW (1985) The new emphasis in cancer control. J Natl Cancer Inst 74:543–551

Gugiu PC, Gugiu MR (2010) A critical appraisal of standard guidelines for grading levels of evidence. Eval Health Prof 33(3):233–255. doi:10.1177/0163278710373980

Hill AB (1952) The clinical trial. N Engl J Med 247(4):113–119. doi:10.1056/NEJM195207242470401

Hunter JE, Schmidt FL (2000) Fixed vs. random effects meta-analysis models: implications for cumulative research knowledge. Int J Select Assess 8:275–292

Lipsey MW (1990) Design sensitivity: statistical power for experimental research. Sage, Newbury Park

Lipsey MW, Cordray DS (2000) Evaluation methods for social intervention. Annu Rev Psychol 51: 345–375

Morell JA (2005) Why are there unintended consequences of program action, and what are the implications for doing evaluation? Am J Eval 26:444–463

Overton RC (1998) A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. Psychol Methods 3:354–379

Pawson R (1994) What works in evaluation research? Brit J Criminol 34(3):291–306

Reichardt CS (2011) Evaluating methods for estimating program effects. Am J Eval 32(2):246–272. doi:10.1177/1098214011398954

Rothwell PM (2005) External validity of randomised controlled trials: "To whom do the results of this trial apply?". Lancet 365:82–93

Rubin DB (2005) Causal inference using potential outcomes. J Am Stat Assoc 100(469):322–331. doi:10.1198/016214504000001880

Scientifically-based evaluation methods. Priority proposal by the US Department of Education. 68 Fed Reg 62445 (2003)

Shadish WR, Cook TD, Campbell DT (2002) Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin, New York

Stuart EA (2010) Matching methods for causal inference: a review and a look forward. Statist Sci 25(1):1–21. doi:10.1214/09-STS313

Victora CG, Habicht J-P, Bryce J (2004) Evidence-based public health: moving beyond randomized trials. Am J Public Health 94:400–405

Weijer C, Shapiro SH, Cranley Glass K (2000) For and against: clinical equipoise and not the uncertainty principle is the moral underpinning of the randomised controlled trial. Br Med J 23:756–758

Veatch RM (2007) The irrelevance of equipoise. J Med Philos Forum Bioethics Philos Med 32(2):167–183. doi:10.1080/03605310701255776

Differential Association Theory

Gerben Bruinsma
Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam,
The Netherlands
UV University Amsterdam, Amsterdam,
The Netherlands

Synonyms

Differential social organization; Early Chicago school theory; History of criminological theories: causes of crime; Social learning theory

Overview

The differential association theory (DAT) of Edwin H. Sutherland is one of the key theories in criminology. The theory and its empirical support, however, are not undisputed. There is much confusion about DAT in the criminological literature, caused partly by Sutherland who changed his theory several times. Early in his career, Sutherland embraced the multiple-factor approach and the interactionist theory of William Thomas (1863–1947), then leaned to social disorganization and culture conflict, and finally settled his own DAT in the late 1930s and 1940s. In this entry, changes in DAT are discussed and the complexity of DAT is illuminated by exploring the intellectual roots of the theory. Several propositions of DAT can be understood better when the theory is situated within broader sociological and psychological traditions. The entry continues by presenting the underlying assumptions of DAT and the arguments of its critics.

Introduction

The differential association theory (DAT) has a history that goes back to the 1920s when a scholar in sociology with a minor in economics was invited to write a textbook on criminology with less focus on European data and research (Bruinsma 1985; Gaylord and Galliher 1988; Goff and Geis 2011). Edwin Hardin Sutherland (1883–1950) published then his "Criminology" that would be influential in the field of criminology for decades (Sutherland 1924). In this first edition, hardly any elaborated perspective on crime causation was presented but in subsequent revised editions, the DAT has been posed, elaborated, and changed in just a few pages each. Besides that, Sutherland, who was by nature an intellectual doubter, tried to improve his DAT continuously. As a result, DAT has been published over the years in scattered works in sometimes slightly different versions, but also in radically changed statements. The concept of *differential association* was introduced in Sutherland's study on the professional thief (Sutherland 1937). Despite the confusion, DAT is still one of the most influential but complex theories in criminology (Geis 1976; Kubrin et al. 2009; Vold et al. 2002). Its complexity is also likely responsible for its abominable history of empirical testing and for the fact that criminologists simplified DAT in the last 60 years into "*the bad influence of delinquent peers on young people's crime involvement*" (Bruinsma 1992; Cressey 1952). This entry will go beyond that simplicity.

Development of the Differential Association Theory

DAT has a remarkable history because of the fact that Sutherland has published different versions of his explanation of crime in four successive editions of his textbook (*Principles of*) *Criminology* over a period of 20 years and in some other publications as well. DAT has been presented by Sutherland in different forms, using varieties in wording, in central concepts, and in causal mechanisms. The history of DAT reflects more or less the intellectual personality and life history of a great scholar, always doubting whether his theory can address the explanation of crime: "*It is a story of confusion, inconsistencies, delayed recognition of implicit meanings, and of much* *borrowing from and stimulation by colleagues and students*" (Sutherland 1956[1942], p. 13). Sutherland was sensitive to the theoretical influences of other criminologists and for critical remarks on his thoughts by his colleagues and friends. The editor of Lippincott Sociological Series invited Sutherland to publish his (*Principles of*) *Criminology*, a book that would be most influential in criminology for decades.

In the first edition of this textbook of 1924, Sutherland was, as he acknowledged later (1956), a follower of the multiple-factor approach. In his critical discussion of the state of art in criminology of that period, Sutherland was looking for information in the empirical research "*that enable us to state that such a person with such and such attitude in such and such situation will always become delinquent*" (Sutherland 1924, p. 82). According to Sutherland each individual is capable of committing crimes but "*it requires contacts and direction of tendencies to make either a criminal or a law-abiding person*" (Sutherland 1924, p. 118).

Furthermore, he adapted the perspective of Chicago sociologists about changing societies in which the people are constantly being influenced by different values and norms and that the family has becoming less influential in socializing young people: "*Social disorganization is a condition of progress as well as of delinquency*" (Sutherland 1924, p. 133). In summary, in the first edition some general assumptions of the DAT can be found:

(a) The search for a universal explanation of crime
(b) Attention to the interaction of the individual and his/her social environment
(c) Interest in cultural and macro social conflicts and their consequences for the individual
(d) The idea that crime – like all other behavior – is learned, and not the result of heritable defects

In the second edition, published ten years later with the changed title *Principles of Criminology*, Sutherland's own vision is more clearly present. Consistent with his thesis about crime as learned behavior, Sutherland added: "*Failure to follow a prescribed pattern of behavior is due to the inconsistency and lack of harmony in the influences which direct the individual*" (Sutherland 1934, p. 52). More strongly than ever before, he joined the so-called "Chicago school" of sociology and criminology, in which macro social (cultural) conflicts are assumed to be central in the explanation of human behavior.

In his classical study *The Professional Thief*, Sutherland (1937, pp. 206–207) for the first time put forward the concept of differential association: "*Differential association is characteristic of the professional thieves, as of all other groups. . .The differential element in the association of thieves is primarily functional rather than geographical. Their personal association is limited by barriers which are maintained principally by the thieves themselves.*" Interactions with other people are a necessary condition to enter the complex world of professional thieves on the border line of conventional society. Differential association is here defined in a narrow sense: *the underworld of criminals*.

In 1939, Sutherland opened his third, and largely revised, edition of *Principles of Criminology* with a chapter in which he tentatively presented his own theory in the form of seven propositions. With these propositions, he tried to connect three levels of explanations: (a) the *macro* level (that of culture conflicts), (b) the *meso* level (that of social disorganization), and (c) the *individual* level (that of having contacts with criminals, i.e., in 1939, differential association). He presented the following seven propositions as a general theory without naming it DAT (Sutherland 1939, pp. 4–9):

(1) The processes which result in systematic criminal behavior are fundamentally the same in form as the processes which result in systematic lawful behavior.
(2) Systematic criminal behavior is determined in a process of association with those who commit crimes, just as systematic lawful behavior is determined in a process of association with those who are law-abiding.
(3) Differential association is the specific causal process in the development of systematic criminal behavior.

(4) The chance that a person will participate in systematic criminal behavior is determined roughly by the frequency and consistency of his contacts with the patterns of criminal behavior.

(5) Individual differences among people in respect to personal characteristics or social institutions cause crime only as they affect differential association or frequency and consistency of contacts with criminal patterns.

(6) Cultural conflict is the underlying cause of differential association and therefore of systematic criminal behavior.

(7) Social disorganization is the basic cause of systematic criminal behavior.

He (1939, p. 9) summarized his theory on the same page as follows:

> Systematic criminal behaviour is immediately to differential association in a situation in which cultural conflicts exist, and ultimately to the social disorganization in that situation. A specific or incidental crime of particular person is due generally to the same process, but it is not possible to include all cases because of the adventitious character of delinquency when regarded as specific or incidental acts.

The crux of this version is that systematic criminal behavior is determined by the process of association with criminals just as systematic law-conforming behavior is developed in a process of association with law-conforming people. Cultural conflict is the underlying cause of differential association and social disorganization is the basic cause. In this version, Sutherland ascribed a limited sense to associations with criminal behavior patterns, namely, only associations with a criminal subculture. Belonging to such a subculture implies that the ratio of contacts turns in favor to one side.

The 1947 Version of Differential Association Theory

Edwin Hardin Sutherland published the final version of his theory in 1947. The years after 1947 Sutherland kept on searching for ways to improve his general theory of crime but his sudden and unexpected death by a stroke in 1950 prohibited that.

Sutherland introduced his nine propositions of DAT with the statement "*The following paragraphs state such a genetic theory of criminal behavior on the assumption that a criminal act occurs when a situation is appropriate for it, as defined by the person who is present*" (Sutherland 1947, p. 6). The added clarifications are from Sutherland (1947, pp. 6–7).

(1) *Criminal behavior is learned.*

According to Sutherland, this proposition implies that criminal behavior cannot be inherited. Besides that, he stated that a person who did not learn how to commit criminal behavior could not invent that behavior.

(2) *Criminal behavior is learned in interaction with other persons in a process of communication.*

In many respects, Sutherland added, this communication can be verbal as well as nonverbal.

(3) *The principal part of the learning of criminal behavior occurs within intimate personal groups.*

Negatively, this statement implies that impersonal communication like films or newspapers plays an unimportant part in the genesis of crime.

(4) *When criminal behavior is learned, the learning includes*:

(4a) *Techniques of committing the crime, which are sometimes very complicated and sometimes very simple.*

(4b) *The specific direction of motives, drives, rationalizations, and attitudes.*

(5) *The specific direction of motives and drives is learned from definitions of the legal codes as favorable or unfavorable.*

In some societies an individual is surrounded by persons who invariably define the legal codes as rules to be observed, while in others he is surrounded by persons whose definitions are favorable to the violation of the legal codes.

(6) *A person becomes delinquent because of an excess of definitions favorable to violation of the law over definitions unfavorable to violation of law.*

This is the principle of differential association. It has to do with criminal and anti-criminal associations. When somebody becomes criminal, it is because he has contacts with criminal behavior patterns and because of a relative isolation with anti-criminal behavior patterns. Sutherland added that everybody assimilates with the surrounding culture, unless other patterns are in conflict with this culture. This sixth statement also implies that neutral associations have none or little impact on the development of crime.

(7) *Differential associations may vary in frequency, duration, priority, and intensity.*

Associations can vary according to these modalities.

(8) *The process of learning criminal behavior by association with criminal and anti-criminal patterns involves all of the mechanisms that are involved in any other learning.*

This statement implies that learning criminal behavior is not limited to the process of imitation.

(9) *While criminal behavior is an expression of general needs and values, it is not explained by those general needs and values because noncriminal behavior is an expression of the same needs and values.*

With this statement, Sutherland opposed himself against criminologists who tried to explain the occurrence of criminal behavior by general needs and values, like searching for happiness, social status, or wealth.

Changes in the Differential Association Theory by Sutherland

Some remarkable changes compared to its predecessor can be observed in the final version of Sutherland's DAT (Bruinsma 1985; Chiricos 1967; Cressey 1952, 1964, 1969):

(a) All references to explanations of processes at higher levels of aggregation were removed from earlier statements of DAT.

(b) More than ever before, Sutherland emphasized the learning process of individuals. In statement one, he claims *that* criminal behavior is learned; in statement two, *by which* criminal behavior is learned; the third statement indicated the *context in which* the learning process take place; and in statements four and five, *what* is learned. In addition, statement eight finally supplies information on *the manner in which* criminal behavior is learned.

(c) Much confusion remained about the central concept of differential association (see § 7.2). The explanation of crime in the final version shifted from an excess of contacts with criminals to an excess of positive definitions. This is a remarkable shift in DAT. In 1939, Sutherland (Sutherland 1939, p. 6) wrote "*The ratio of criminal acts to lawful acts by a person a roughly the same as the ratio of the contacts with the criminal and with the lawful behaviour of others.*" Cressey (Sutherland and Cressey 1966) added that it is all about the presentation of positive and negative definitions and that the person who presents them not necessarily needs to be a criminal.

(d) Compared to the 1939 edition, Sutherland abandoned the idea that associations need to be criminal themselves ("crime is the cause of crime"). In the 1947 version, it is about the presentation of behavior patterns, despite the character of the person who presents these.

(e) In the final version of DAT, the modalities of associations (duration, priority, frequency, and intimacy) were introduced. Sutherland replaced the 1939 concept of consistency of associations by intimacy and added two new modalities in addition to the frequency of associations: priority and duration. "*In a precise description of the criminal behaviour of a person these modalities would be stated in*

> *quantitative form and a mathematical ratio be reached. A formula in this sense has not been developed and the development of such a formula would be extremely difficult*" (Sutherland 1947, p. 7).

Roots of Sutherland's Differential Association Theory

DAT cannot fully be understood without discussing its intellectual roots. The fundaments of DAT can be found in (symbolic) interactionism (Fisher and Strauss 1978). It is a sociological and social psychological perspective on society and individuals that originated in Chicago in the late nineteenth century and early twentieth century. Sutherland tried to translate the basic elements of this sociological perspective into the field of criminology. Some of these elements emerged from earlier works of sociologists and psychologists who developed (symbolic) interactionism, others from the publications of Sutherland's colleagues and students during his professional life. In the studies of Charles Cooley (1854–1929), George Mead (1863–1931), and William Thomas (1863–1947), the focus is on social processes and on the interaction of the individual with his environment. They all focus on primary intimate groups as the most important social environment of people (Warr 2001). By interacting and communicating with others, the self (that regulates one's social behavior) can develop into a generalized other: "*The organized community or social group which gives to the individual his unity of self may be called 'the generalized other'. The attitude of the generalized other is the attitude of the whole community*" (Mead 1974 [1934], p. 154). As in DAT, the face-to-face interactions and responses of others are emphasized in studying the social behavior of people, whether deviant, criminal, or conforming behavior. From others individuals learn how to define the social situations that they encounter in daily life. In general, symbolic interactionism emphasized the changeability of societies and the heterogeneity of values within them, exercising their impact on people's motives, attitudes, and behaviors. Mead emphasized the interactions with other people in which definitions (= normative constraints) are constantly reformulated and reworked in specific contexts (the process of learning). Being close to this principle, Sutherland views the development of a criminal self as a result of the individual's communications and interactions with the social environment.

According to Vold and his coauthors (2002, pp. 160–162), DAT adopted also another basic element from Mead's symbolic interactionism, that is, that the contents of what is learned (techniques, motives, drives, attitudes, and definitions) are cognitive elements. It was Mead's idea that human beings act toward objects based on the meanings that the objects have for them. People construct relatively permanent "definitions of their situation" out of their experiences they derive from particular situations, and form a relatively set of ways of looking at reality. Cognitive factors determine behavior or, as DAT stated that principle, an excess of positive definitions toward law violations. Mead argued that the self, as he called it, is socially constituted. The organized society is prior to the individual, and the self has a social structure in which the consciousness is socially organized from the social organization of society in general. Gongaware and Dotter (2005) demonstrated clearly that Sutherland owed much of the symbolic interactionist theory of George Mead and that Sutherland's principles can be perceived as a translation and elaboration of that theory in the field of crime causes.

Although Sutherland skipped the word "conflict" in his final version of DAT, he was convinced that culture conflict is the underlying cause of crime in society. His discussions with Thorsten Sellin (1896–1994) strengthened his view that culture conflicts are a consequence of the transition of society from homogeneous to heterogeneous and of the increasing social differentiation (Sellin 1938; Sutherland 1939). Migration processes accelerate changes in society and consequently culture conflicts will emerge. Sutherland nevertheless posits culture conflict as an important causal factor in his differential organization theory, explaining variations in crime rates of social and ethnic groups.

Assumptions of the Differential Association Theory

DAT is based on various assumptions. These assumptions are partly testable, partly not. The assumptions are that:

(i) Crimes are not inherited.
(ii) Crimes can be learned in social interactions.
(iii) Individuals are not just a tabula rasa in which the environment put definitions favorable to commit crimes and where crimes are the output of an unobservable process.
(iv) Individuals play not only a passive role in the learning process but also an active role in acquiring positive definitions, close to symbolic interactionism rejecting the idea of an oversocialized conception of individuals.
(v) Individuals do not invent criminal behavior.
(vi) DAT implies a recursive process in which committing crimes will continue as long as individuals have contacts with criminal behavioral patterns.
(vii) The mechanisms underlying the learning process are active continuously.
(viii) Personal traits or propensities like self-control or susceptibility play no role in the emergence of crime.
(ix) Processes at a higher societal level are the fundamental causes behind the causes at the individual level.

A number of these assumptions have been the subject of several debates after the death of Sutherland in 1950 (Cressey 1952, 1964, 1969). Most of these assumptions, however, can be translated in propositions and subsequently be tested empirically. An important assumption of Sutherland that personality characteristics play no role in the occurrence of crime needs to be clarified more fully in order to understand Sutherland's thoughts. Sutherland took the criticism very seriously that DAT cannot ignore personality characteristics, or psychological variables as he called them: "*I believe it is the most important and crucial question in criminological theory*" (Sutherland 1956[1942], p. 25). Originally he underlined that these variables should be incorporated in his DAT, later he rejected that idea (Sutherland 1956; 1956 [1942]).

Criticism on the Differential Association Theory

Criticism on the Nine Propositions

Sutherland theorized in his first proposition that crime is learned. This statement, however, offers no information on how criminal behavior is learned. Statement *one* contains no conditions that can lead to criminal behavior. Others have pointed at the ideological content of this proposition (Warr 2001, p. 184). However, from an explanatory point of view, the proposition is lacking to supply any information how the learning process is taking place and what is learned.

A similar critique is valid for the *second* statement of DAT. Every individual interacts with other people without generating crime. There is no specification of conditions by what kinds of interaction crime will occur. The statement informs us only that learning is not a solitary or isolated process, but a *social* process in which other people are involved. Who are involved is pointed at in the next statement: intimate groups. But Sutherland did not define in statement *three* in what intimate groups crime is learned. In practice, criminologists limit themselves almost always to "delinquent friends" or "peers" in general as empirical indicator for intimate groups. Just in statement *four*, a first condition of the occurrence of crime is mentioned. The variable "techniques to commit crimes" (whether complicated or simple) is assumed to be a necessary condition for committing a crime. Besides that, it is stated that the specific direction of motives, drives, rationalizations, and attitudes is relevant. However, which motives, drives, rationalizations, and attitudes remain unclear as well as how these have an impact on the occurrence of crime.

In statement *five*, some conditions about the origins of motives and drives are indicated. Statement five can be formulated as "If laws are defined positive, then motives and drives will emerge conform the law, and if laws are defined as negative then criminal motives and drives will emerge" (Opp 1974). Conditions for the occurrence of crime are stated in proposition *six*.

Sutherland claimed that when an excess of definitions favorable to violate laws over definitions unfavorable is present, crime will occur. In statement *seven* it is indicated that associations can be attributed different qualitative characteristics like frequency, duration, priority, and intensity, but contains no explanandum.

The nine statements of DAT vary considerably in scope and range. In modern epistemological perspectives, statements 4, 5, 6, and 8 can be classified as testable hypotheses. Statements 1, 2, 3, and 9 are general orientations or general theoretical outlooks in which clusters of undefined variables are related to each other and cannot be tested properly. The seventh statement is not only a general theoretical outlook but it also contains a metascientific sentence of methodological nature that quantifying is a heuristic useful activity in theory formation.

Criticism on the Concepts

The Excess of Positive Definitions to Violate the Law

Several complex and abstract concepts play a role in the nine propositions of DAT. Some are very problematic for the theory. The most worrisome is the concept of an excess of definitions favorable to violate laws that will cause crime. What is meant by the concept "an excess of positive definitions to violate laws over definitions unfavorable to violate laws"? To answer this question one has to clarify four issues:

(a) What are definitions favorable to law violations?
(b) What is an excess of definitions favorable to law violations?
(c) Is it the excess of definitions *of individuals whose criminal behavior has to be explained* or is it the excess of *presented* position definitions?
(d) How do individuals recognize and adopt these definitions from other people?

DAT is embedded in symbolic interactionism in which the learning of definitions is part of the development of the social self. People construct permanent definitions out of their experiences from particular situations and form a relatively enduring set of ways of looking at things. In communication with others these definitions are transferred (or, to do justice to the symbolic interactionist roots of DAT, the word of exchanging definitions positive to violating the law between people should be preferred, expressing the active role of individuals in building a set of attitudes) to the individual and consequently adapted to regulate his behavior in a given context. The definitions are cognitive elements like behavioral techniques, motives, drives, and attitudes. Sutherland only used the attitudes part of symbolic interactionism. These can be interpreted as a *kind of moral rules that guide a person's behavior*. If people would communicate to each other that a certain kind of behavior is not appropriate in a particular context, you may apply that moral rule as a consequence of that rule. If the social environment is proclaiming the attitude that a certain kind of behavior is appropriate in that context, you may apply that rule. The underlying concept of definitions resembles closely the concepts of *beliefs* of Hirschi's social control theory (2006[1969]) or of *morality* in situational action theory of Wikström (2010). It is common practice to include neutralization techniques being part of the concept of positive definitions, implying that part of the positive definitions to violate the law also consist of rationalizations that take away barriers to commit crimes. As such, they are a category of positive definitions.

The second complexity is the measurement of an "excess." Can an excess of positive definitions to violate laws be counted empirically? A number of critical comments (Bruinsma 1985; Glueck 1956; Matsueda 1982, 1988; Short 1960) encountered problems in the measurement of the excess of definitions.

At the theoretical level some additional critical questions can be raised. (i) Let us assume that Sutherland is right and that there is a kind of crossover in the formation of definitions; why is the crossover then put exactly on .5? Why not on .6 or .4? (ii) If we insist that a kind of crossover exists, then the chance that DAT will be refuted in empirical research will be 100 % because the number of conforming behaviors exceeds the number of criminal behaviors.

As a consequence, the definitions about these kinds of behaviors will also be unbalanced. (iii) Why does the presence of positive definitions, that are less than .5 of all definitions of a person, have no significance at all for the commission of crimes? (iv) We have to consider whether positive and negative definitions are about *all* violations of the law. Do people have a definition about all possible crimes? If yes, that would imply that someone must have an infinitive number of definitions. That would be very unrealistic. Lastly, (v) another objection can be put forward on this central concept of DAT. One can question whether an excess of positive definitions to violate the law is about general attitudes toward committing crimes or *specific* attitudes toward *specific* crimes. Is it possible that people have a positive definition to commit fraud and at the same time have negative definitions to other forms of crime? In DAT there is no sign whether these definitions are about specific crimes or crimes and rule breaking behavior in general.

However, the most important weakness of DAT is the indistinctness whether an excess of positive definitions is an attribute of the *person whose behavior has to be explained* or has to do with *his environment*. Almost all criminologists assume that the excess of positive definitions is of the person whose criminal conduct has to be explained, but in statement six of DAT, *no object* is mentioned to which the attribute of an excess of positive definitions favorable to violating the law is related. The debate has always been about the question whether not everybody who has contacts with criminal associations will become criminal. Cressey defended that important question not very clearly: "*Thus, he* (Sutherland) *does not say that persons become criminals because of associations with criminal behavior patterns; he says that they become criminals because an* overabundance *of such associations, in comparison with associations with anti-criminal behavior patterns*" and thus strengthened the confusion (Cressey 1964, pp. 25–26). This theoretical issue needs to be sorted out by solid empirical research.

The Modalities of Associations

Sutherland identified four modalities of associations in the seventh proposition of the 1947 version of DAT: It is stated that associations can vary in frequency, duration, priority, and intensity. The meanings of priority and duration are "*obvious and need no explanation*" (Sutherland 1947, p. 7). About priority he wrote "'*priority' is assumed to be important in the sense that lawful behaviour developed in early childhood may persist throughout life. This tendency, however, has not been adequately demonstrated, and priority seems to be important principally through its selective influence*" (Sutherland 1947, p. 7). Warr (Warr 2001, p. 185) suggested that Sutherland referred with this modality to the Freudian point of view that was accepted as a common fact at his time. It remains unclear whether the theory is about the frequency of the contacts a person has *now* or is it about the contacts he had f*or the first time in his life*. In the latter interpretation it is assumed that the younger an individual is when he first comes in contact with criminal behavior patterns, the more impact they have for the rest of his life. For instance, it is assumed that the parents have more impact on the development and acceptance of positive definitions to law violation than other persons later in life.

The modality *frequency of contacts* got no further specification by Sutherland. It is assumed that the more often you see people and talk to them, the more influential that association will be. Frequency can thus be either the number of times you meet other people or the amount of time people spend together in a defined period of time (day, week, month, year) or the number of times multiplied by the hours spent together. That last option is never used in criminological research. The third modality *duration* is left open by Sutherland with no further specification. If incorporated in the frequency of contacts, then duration will be redundant as a separate modality of contacts with criminal or conforming behavior patterns. Opp (Opp 1974) suggested in his modification of DAT to let out this modality for that reason.

The last modality, the *intensity* of criminal associations, also named *identification or attachment*, is the most complex of the four, because a (developmental) psychological notion is introduced in DAT. By intensity Sutherland meant the prestige of the source of association. With this modality he tried to overcome the critics of his time that police officers and prison guards have frequent contact with criminals but show no criminal behavior. The suggestion was that police officers seldom have intimate relationships with criminals and hardly ever attribute prestige to them. That makes them less vulnerable for the impact of them. Identification is generally regarded as the necessary condition for an individual to obtain preferences, attitudes, values, and norms from a group someone belongs to or wants to belong to. Acquiring of these happens in an interaction process called socialization. Crucial in this socialization process is the existence of real or symbolic models with which the individual identifies himself.

It can be concluded that in DAT it is left open what the function of this proposition is in the theory and what the effects of these modalities are on the excess of positive definitions favorable to criminal behavior. Probably it is implied that the more frequent, the longer, and the higher the priority and the more intense the contact with associations, the greater the chance that an excess of positive definitions over definitions negative to law violations will occur (the sentence is deliberately formulated without referring to any object, considering the discussion of the sixth proposition of DAT in the previous paragraph). There are no indications presented that one of the "modalities" is more important than one of the others, nor if, and how, these modalities are related to each other. Is priority more important than duration? And if so, how much more important? In the history of DAT, there has never been an attempt made to qualify the relative weights of the four modalities into mathematical formulas to assess the impact of each of them on the formation of positive definitions favorable to violating the law.

Differential Associations with Whom?

In DAT it is stated that the interactions and communications are in intimate social groups of significant others. That allows a great number of people with whom a person can have associations. In many research, DAT is perceived as the theory of the bad peers or delinquent friends (Bruinsma 1992). Sutherland never limited DAT to have contacts with just peers or delinquent friends. People interact with parents, siblings, grandparents, members of the extended family, neighbors and other neighborhood residents, schoolmates and teachers, colleagues, shop owners, sport mates and trainers, acquaintances, and strangers. All can potentially influence an individual by exchanging positive and negative definitions toward law violations through interactions and communications. It is surprising to observe that siblings have not attracted much more attention in empirical research. It can be suggested that siblings share many interactions and communications about definitions and behavioral techniques lasting for many years.

Conclusions

The differential association theory of Edwin Sutherland has a long history in criminology. It has been presented by Sutherland in different wordings and scope in scattered publications over the years, many times in a very brief statement. The differential association theory can be placed within the tradition of symbolic interactionism of early Chicago sociology. There is much criticism on the nine statements of the theory as well as on the concepts Sutherland has used. The theory is not undisputed in criminological theory. Hirschi and Gottfredson (1980, p. 9) stated that DAT neither predicts nor explains criminal behavior.

During the years, a number of modifications of DAT were published, of which some became very well known in the field of criminology. Gresham Sykes and David Matza modified proposition 4 of DAT into their Neutralization Theory (Matza and Sykes 1957; Sykes and Matza 1957), Dan Glaser into the Differential Identification Theory based on proposition seven and into the Differential Anticipation Theory (Glaser 1978), and Richard Cloward and Lloyd Ohlin into the Differential Opportunity Theory based on a combination of DAT and anomie theory of Robert Merton (Cloward and Ohlin 1960); De Fleur and Richard Quinney made a formalized model of DAT according to the principles of formal logic and set theory (Fleur and Quinney 1966); the German sociologist Karl-Dieter Opp reformulated the theory guided by the methodological rules of critical rationalism (Opp 1974, 1976) and one that is based on behavioral theory that is still present in today's criminology: the social learning theory (SLT) of Akers (Akers 1973, 1998). Akers have several times claimed that his social learning theory is superior to Sutherland's one: "*... social learning is not meant to be an alternative, competitive, or rival theory to Sutherland's position. It is, instead, a broader theory that integrates processes of differential association and definitions from Sutherland's theory, modified and clarified, with differential reinforcement and other principles of acquisition, continuation, and cessation from behavioral learning theory*" (Akers 1998, p. 47). This entry, however, demonstrated that this claim is not valid and that the roots of both are too distinct to compare. DAT is based on symbolic interactionism, social learning on behaviorism. They differ in scope, vision on human nature, and how people learn and decide to commit a crime in a certain setting.

Acknowledgement I thank professors Sally Simpson and Lieven Pauwels for their comments and suggestions for the improvement on a previous version.

Recommended Reading and References

Akers RL (1973) Deviant behavior. A social learning approach. Wadsworth Publishing, Belmont

Akers RL (1998) Social learning and social structure: a general theory of crime and deviance. Northeastern University Press, Boston

Bruinsma GJN (1985) Criminaliteit als sociaal leerproces: een toetsing van de differentiële associatietheorie in de versie van K.-D. Opp. Gouda Quint, Arnhem

Bruinsma GJN (1992) Differential association theory reconsidered: an extension and its empirical test. J Quant Criminol 8:29–50

Chiricos TG (1967) The concept of cause. A developmental analysis of the theory of differential association. Issues Criminol 3:91–99

Cloward RA, Ohlin LE (1960) Delinquency and opportunity: a theory of delinquent gangs. Free Press, New York

Cressey DR (1952) Application and verification of the differential association theory. J Crim Law Criminol Police Sci 43:43–52

Cressey DR (1964) Delinquency, crime and differential association. Martinus Nijhoff, Den Haag

Cressey DR (1969) Epidemiology and individual conduct. In: Cressey DR (ed) Delinquency, crime and social process. Harper & Row, New York, pp 557–577

Fisher BM, Strauss AL (1978) Interactionism. In: Bottomore T, Nisbet R (eds) A history of sociological analysis. Heinemann, London, pp 457–498

Fleur MJD, Quinney R (1966) A reformulation of sutherland's differential association theory and a strategy for empirical verification. J Res Crime Delinq 3:1–22

Gaylord MS, Galliher JF (1988) The criminology of Edwin Sutherland. Transaction Books, New Brunswick

Geis G (1976) Editorial: revisiting Sutherland's criminology (1924). Criminology 14(3):303–306

Glaser D (1978) Crime in our changing society. Rinehart & Winston, New York

Glueck S (1956) Theory and fact in criminology: a criticism of differential association. Brit J Criminol 7:92–98

Goff C, Geis G (2011) Edwin Hardin Sutherland: the development of differential association theory. In: Cullen FT, Johnson CL, Myer AJ, Adler F (eds) The origens of American criminology. Transaction, New Brunswick/London, pp 37–62

Gongaware TB, Dotter D (2005) Developing the criminal self: Mead's social psychology and Sutherland's differential association. Sociol Spectr 25:379–402

Hirschi T (2006[1969]) Causes of delinquency, 5th edn. Transaction, New Brunswick/London

Hirschi T, Gottfredson M (1980) The Sutherland tradition in criminology. In: Gottfredson M, Hirschi T (eds) Understanding crime. Current theory and research. Sage, Beverly Hills, pp 7–19

Kubrin CE, Stucky TD, Krohn MD (2009) Researching theories of crime and deviance. Oxford University Press, New York/Oxford

Matsueda RL (1982) Testing control theory and differential association: a causal modeling approach. Am Sociol Rev 47(4):489–504

Matsueda RL (1988) The current state of the differential association theory. Crime Delinq 34(3):277–306

Matza D, Sykes GM (1957) Techniques of neutralization: a theory of delinquency. Am J Sociol 22:664–670

Mead GH (1974[1934]) Mind, self & society from a standpoint of a social behaviourist, vol 1 (ed. and with an introduction by Ch. W. Morris, 19 edn. The University of Chicago Press, Chicago/London

Opp K-D (1974) Abweichendes Verhalten und Gesellschaftsstruktur. Luchterhand, Neuwied

Opp K-D (1976) Methodologie der sozialwissenschaften. Einführung in Probleme ihrer Theorienbildung. Rowohlt, Reinbeck bei Hamburg

Sellin T (1938) Culture conflict and crime. Social Science Research Council, New York

Short JF Jr (1960) Differential association as a hypotheses: problems of empirical testing. Soc Probl 8:14–24
Sutherland EH (1924) Criminology. Lippincott, Philadelphia
Sutherland EH (1934) Principles of criminology. Lippincott, Philadelphia
Sutherland EH (1937) The professional thief. The University of Chicago Press, Chicago
Sutherland EH (1939) Principles of criminology. Lippincott, Philadelphia
Sutherland EH (1947) Principles of criminology. Lippincott, Philadelphia
Sutherland EH (1956a) Critique of the theory. In: Cohen AK, Lindesmith AL, Schuessler K (eds) The Sutherland papers. Indiana University Press, Bloomington, pp 30–41
Sutherland EH (1956[1942]) The development of the theory. In: Cohen AK, Lindesmith LA, Schuessler K (eds) The Sutherland papers. Indiana University Press, Bloomington, pp 13–29
Sutherland EH, Cressey DR (1966) Principles of criminology, 7th edn. Lippincott, Chicago
Sykes GM, Matza D (1957) juvenile delinquency and subterranean values. Am Sociol Rev 26(6):712–719
Vold GB, Bernard TJ, Snipes JB (2002) Theoretical criminology, 5th edn. Oxford University Press, New York
Warr M (2001) The social origins of crime: Edwin Sutherland and the theory of differential association. In: Paternoster R, Bachman R (eds) Explaining criminals and crime: essays in contemporary criminological theory. Roxbury, Los Angeles, pp 182–191
Wikström P-OH (2010) Situational action theory. In: Fisher B, Lab S (eds) Encyclopedia of victimology and crime prevention. Sage, Thousands Oaks/London, pp 875–879

Differential Association-Reinforcement Theory

▶ Social Learning Theory

Differential Enforcement

▶ Biased Policing

Differential Group Organization

▶ Differential Social Organization

Differential Offending

▶ Juvenile Justice System Responses to Minority Youth

Differential Social Organization

Ruth Triplett
Department of Sociology and Criminal Justice,
Old Dominion University, Norfolk, VA, USA

Synonyms

Differential group organization

Overview

In this entry, the concept of differential social organization is traced from the publication of the fourth edition of *Principles of Criminology* by Edwin Sutherland in 1947. Sutherland viewed differential social organization as the explanation for group crime rates that corresponded to differential association, his individual-level theory of involvement in crime. Sutherland left the factors that comprise differential social organization largely unexplored, however, along with exactly how it should be linked to differential association. Along with some powerful criticisms of his macro theory as cultural deviance, this lack of attention by Sutherland may explain why differential social organization never received the attention that researchers and theorists alike have given to differential association. Recent work by Akers (1998), Matsueda (2006), and Sampson and Graif (2009), however, offer insights into the concept of differential social organization and address some of the questions that Sutherland left unanswered. These works, along with those of others such as Harding (2009); Browning (2009); Haynie et al. (2006); and De Coster et al. (2006), offer new avenues for exploration for those interested in explaining neighborhood variation in rates of crime, in particular, and encourage new questions about organization both for and against crime.

Understanding Differential Social Organization

Differential social organization, sometimes referred to as differential group organization, is a central concept in the fourth edition of Edwin Sutherland's *Principles of Criminology*. Though used to explain aggregate crime rates in a way that corresponded to his individual-level theory of crime, differential association, two key questions about differential social organization were left largely unexplored by Sutherland. First, Sutherland gives little guidance for understanding the key factors important in understanding organization both for and against crime. Second, he says little about how differential social organization and differential association should be linked. Perhaps as a result of this lack of attention, the concept has received little attention from other criminologists compared to differential association. Recent work though shows that it offers interesting possibilities for shedding new light on well-researched areas such as neighborhoods and crime.

In what follows, Sutherland's development of the concept of differential social organization is outlined along with its influence on the work of Cloward and Ohlin (1960). Their expansion of his ideas gives early insight into factors important in shaping organization for crime. In addition, criticisms of Sutherland's work that might well have hampered the willingness of criminologists to give it much attention are reviewed. Attention then turns to an explication of recent work that attempts to answer questions left unexplored by Sutherland. Understanding what Sutherland means by differential social organization, however, requires a short discussion of two other concepts, differential association and culture conflict. A review of the three concepts follows.

Background Description

In 1921, Edward Hayes, then chair of the Department of Sociology at the University of Illinois, asked Edwin Sutherland to write a textbook on criminology (Gaylord and Galliher 1988). Hayes was the editor of a book series for Lippincott, and it was this request that led Sutherland to devote himself to criminology. The first edition of *Criminology*, as his text was originally titled, was published in 1924. By 1939, he included in it a statement of his theory of differential association which he revised until the publication of the final version in 1947 when the fourth edition of *Principles of Criminology* was published (Gaylord and Galliher 1988). With the publication of this edition, three theories are offered to explain crime at three different but corresponding levels – differential association (individual), differential social organization (group), and culture conflict (societal).

Differential association is the most well known, developed, and tested aspect of the three ideas. In 1947, Sutherland laid out his theory of differential association in nine propositions. The theory explains individual involvement in crime as the result of a process of learning, through association. The learning includes both techniques and "…the specific direction of motives, drives, rationalizations and attitudes" (1947, p. 6) favorable to the commission of crimes. It predicts that an individual "…becomes delinquent because of an excess of definitions favorable to law violation over definitions unfavorable to violation of the law" (6).

After the statement of the ninth proposition, Sutherland writes that what was just stated at the individual level can also be stated at the group level. Differential social organization is founded on his belief that "…crime is rooted in social organization and is an expression of that organization. A group may be organized for criminal behavior or organized against criminal behavior. Most communities are organized both for criminal and anti-criminal behavior and in that sense the crime rate is an expression of the differential group organization" (1947, pp. 8–9). He writes of the link between differential social organization and differential association that – "The person's associations are determined in a general context of social organization. A child is ordinarily reared in a family; the place of residence of the family is determined largely by family income; and the delinquency rate is in many respects related to the rental value of the houses. Many other factors enter into this social organization, including many of the small personal group relationships" (1947, p. 8).

Sutherland did not develop his ideas about differential social organization to the extent that he did those surrounding differential association. In his discussion of differential social organization, he does not explain, for example, what the "...many other factors..." of social organization are, either for or against crime, that he considered important. However, there are two sources beyond *Principles of Criminology* that are important to examine for further insight into what Sutherland meant by differential social organization. The first is a chapter he wrote on theft during wartime published in 1943, 4 years before the fourth edition of *Principles of Criminology*. The second is what can be drawn from those individuals and works that influenced his ideas about social organization.

In a chapter entitled "Crime" in an edited volume entitled *America in Wartime*, Sutherland outlined a theory of theft during wartime that was based on the concept of differential social organization. Unhappy with the explanations offered by others, Sutherland (1943) offers differential social organization as an alternative which, he argues, can fit the known facts about theft during wartime. In doing so, he describes a number of aspects of organization both against theft and for theft that add to the meaning of differential social organization. Sutherland (1943) lists external opportunities relating to guardianship of homes and goods; socialization of children by parents, schools, and churches; supervised activities for children; attitudes towards property ownership by those who do not own much property; and clarity in the meaning of property ownership and ownership rights as factors important in the organization *against* theft. Factors important in organization *for* theft that he listed included various changes in contacts with criminal patterns. Sutherland argued that frequency of theft was increasing which increased the chance that any one individual would have contact with someone who stole and collusion of individuals with thieves, and that there had been an increase in the number of people stealing to sell to the black market. In addition, Sutherland included a cultural element in the social organization for crime. Sutherland proposed that during wartime, there were more people for whom theft was the only alternative to not surviving. Since culture allowed for this exception to the rule that people should not steal, he argued, there could be predicted an increase in theft. "The two aspects of organization together are called differential group organization. The balance between the opposed organizations determines whether the crimes committed, the reactions against crimes, and the convictions rates increase or decrease" (203). As Matsueda (2006) writes, it is clear from this entry that Sutherland saw differential social organization as consisting of a variety of different structural and cultural factors.

A final important part of Sutherland's ideas on differential social organization is found in the conclusion of this entry. Here he writes that "organization has two principal constituent elements, namely consensus in regard to objectives and implementation for the realization of objectives...Each of these is found in the organization for crime and the organization against crime" (203). This statement suggests that organization involves not simply structures and culture, which might be considered fairly stable, but the more dynamic process of interaction that must occur for a group to develop agreement on a common goal and take action towards achieving that goal (see Matsueda 2006 and the discussion of his work below). It is notable that these two components of social organization correspond with the definition of social disorganization used by contemporary theorists interested in neighborhood variation in crime rates that involves the ability to agree upon and work towards a common goal.

In addition to Sutherland's own writing, Matsueda (2006) argues that an examination of some of the figures whose work influenced him can give insight into the meaning of differential social organization. Most notable perhaps is the influence of Shaw and McKay (1942) who are the founders of the theory of social disorganization. In an essay on the development of his theory, Sutherland (1942[1973]) reports having originally "borrowed" the term social disorganization from Shaw and McKay. Later, he changes this for differential social organization under the encouragement of Albert Cohen because "…the organization of the delinquent group…is social disorganization only from an ethical or some other particularistic point of view" (1942[1973], p. 21). Gaylord and Galliher (1988) later argue that he changed the terminology in response to two factors. First critiques come out of social disorganization, some of which drew attention to the fact that it seemed to be evaluating the organization of certain areas, "theirs," as deficient to "ours." They report that Sutherland wanted to get away from the negative connotation associated with the concept. Second, they argue that he found the concept differential social organization better at conveying his idea that society can be organized both for and against crime. The change in terminology does not mean that he did not see value in Shaw and McKay's ideas or their research findings, both of which he continued to cite in his text. Matsueda (2006) argues, though, that he is reconceptualizing Shaw and McKay's social disorganization to be weak organization against crime.

A second important influence on Sutherland's ideas of differential group organization is found in Tannenbaum (1938). Most broadly, Sutherland saw value in Tannenbaum's work on identity and the role of social reactions against crime in the shaping of criminal identify (Gaylord and Galliher 1988). This is connected with the influence of symbolic interactionism on Sutherland and the role it plays in differential association. Specifically in terms of differential social organization, however, his influence can be seen in Sutherland's emphasis on reactions to crime as key part of differential social organization. On the one hand, the community response to crime is part of the organization against crime. On the other, the development of a criminal identity is part of organization for crime (Matsueda 2006).

Overall then a review of Sutherland's own writings and those who influenced his work suggest that differential social organization involves social structural and cultural elements. Also involved are the processes that lead to the development of a common goal and action towards the achievement of that goal. With safety from theft and violence as a common goal, control of the behavior of others, including reactions to misbehavior, is key element of the action towards achieving that goal.

Along with differential association and differential social organization, a third concept is also important in understanding Sutherland's work. This is culture conflict or, what Cressey (1960, 1968) later came to call, normative conflict. This refers to "…conflict between legal and other norms that arise through the societal growth process…" (Cressey 1968, p. 50). For Sutherland, culture conflict was the basic explanation for crime. Culture conflict then explains why there is crime in society; differential social organization explains why some groups have higher crime rates than others in that society; and differential association explains why a particular individual is involved in crime. How all three link up is left largely undeveloped by Sutherland.

Even after reviewing Sutherland's writings and influences, and placing the concept of differential social organization in relation to differential association and culture conflict, it remains clear that Sutherland never developed his ideas on differential social organization in the systematic way that he did his ideas of differential association. This lack of development left it open to interpretation by others, both supporters and critics, in ways that influenced the interest in and thus development of his ideas by other criminologists. One particularly important development of his ideas of social organization for crime is found in the work of Cloward and Ohlin (1960). One particularly important critique is found in Kornhauser's (1978) *Social Sources of Delinquency*. A discussion of each of these works follows.

Two early scholars whose work explores the ideas of differential social organization are

Cloward and Ohlin (1960). In *Delinquency and Opportunity*, they link Merton's theory of anomie with Sutherland's differential association and differential social organization to explain the nature, development, and stability of delinquent subcultures among males in lower class areas. Their theory centers on the idea that youths want to achieve success but have differential access to opportunities to do so. Cloward and Ohlin argue that Merton's theory of anomie recognized differences in conventional opportunities for achieving success but ignored differences in opportunities for crime. While Sutherland's differential association recognized differences in opportunity to learn about criminal opportunities, he ignored differences in legitimate opportunities. They felt that inclusion of both in one theory would enhance the ability of theory to explanation of the nature and existence of various types of subcultures.

Cloward and Ohlin (1960) posit that youths in the lower class, having little in the way of legitimate opportunities for success, struggle to find options for achieving higher status. Just what options they have were argued by Cloward and Ohlin (1960) to depend in large part of the social organization of the lower class neighborhood in which they found themselves. They noted two factors in particular. The first was integration of different age levels which helped with the transmission of delinquent skills and values. The second was integration of conventional and deviant values. They argued integrating illegitimate opportunities structures and values with those of the legitimate world enhanced the stability of criminal roles. Differences in the organization of a given area in terms of opportunity structures are what lead to differences in subcultures. They describe three ideal subculture types – delinquent, conflict, and retreatist. Delinquent subcultures involve theft and organized criminal activities as alternative ways of achieving status. These emerge in neighborhoods where there exist illegitimate opportunities for achieving status because of both an integration of age levels and conventional and deviant values. Conflict subcultures, center on the use of violence, and emerge where few illegitimate opportunities exist. Finally, it is the retreatist, revolving around drugs, which consists largely of those individuals who are not able to successfully compete in either the legitimate or illegitimate opportunity structure.

By positing the existence of different subcultures and relating them to access to opportunity for learning and integration into legitimate or illegitimate life, Cloward and Ohlin were starting on a path of exploration of what factors are key to understandings differential social organization. Though there work was highly influential both in criminology and in the arena of public policy, continued criticisms of social disorganization and cultural deviance theories alike were to turn criminologists' attention away from exploring the role of social organization in understanding crime. In particular, the influential work of Kornhauser (1978) did much to turn attention of criminology, in general, and those interested in social disorganization theory, in particular, away from differential social organization.

There are two parts to her critique that are important to the discussion here. The first is her critique of the work of Shaw and McKay (1942) and their "mixed" model of neighborhood crime. She points out that they present us with a model that combines two different theories – control theory and cultural deviance. The control part of the theory predicts that in high-crime-rate neighborhoods, neighborhood controls are weak, freeing youths to be delinquent. These youth then develop delinquent traditions in the neighborhood that are transmitted to others. The cultural deviance part, then, argues that youths in neighborhoods with delinquent subcultures commit delinquent acts because of their socialization into a delinquent way of life. She argues that this mixed model is not logically possible because control and cultural deviance theories are based on differing assumptions about human nature. On the one hand, control theories assume that youths are capable of delinquency when free from control and that no other motivation is necessary. On the other, cultural deviance theories assume youths are not capable of deviance. They are merely following the norms and values of a subculture into which they have been successfully and completely socialized. Kornhauser suggests that the cultural deviance part of the model is not necessary and, for reasons discussed below, not even logically tenable. She then presents us with a pure control model of neighborhood crime.

Kornhauser (1978) combines her argument for a social disorganization theory that is based on a pure control model with a powerful critique of cultural deviance theories. She uses Sutherland's differential association and his discussion of culture conflict as the prime example of a cultural deviance theory. Her argument starts with the assumptions she believes are behind cultural deviance theories. These are that "…man has no nature, socialization is perfectly successful, and cultural variability is unlimited" (p. 34). She then points out that the first is not actually consistent with what cultural deviance theorists argue. If socialization is perfect, she argues, it must be because people *by their nature* are perfectly adaptable. Though others have argued that she misinterprets Sutherland (see, e.g., Akers 1998), the critique resonated with others. Discussion and research on social disorganization largely followed the pure control model. Cultural deviance theories continued to fall into disfavor, including Sutherland's ideas about cultural conflict. Differential social organization received relatively little attention, and what it did receive did not differentiate between culture conflict and differential social organization. The lack of attention Sutherland gave to differential social organization relative to differential association is mirrored by the level of attention the field has given it. There are, however, several recent exceptions which show various ways of exploring this idea, and it is to these that the discussion now turns.

State of the Art

Recent work on differential social organization has offered new opportunities for development and research on the factors that may shape organization both for and against crime. Here, the focus is on three different works that contribute to the understanding of differential social organization. The first work, by Akers (1998), explores linking differential social organization with the individual-level differential association. Two others, Matsueda (2006) and Sampson and Graif (2009), critically examine various aspects of differential social organization.

Akers, who has spent his career developing and testing social learning theory as an elaboration of differential association, has now taken his work a step further by articulating the link between differential social organization and differential association. Though his initial work focused exclusively on rewriting the individual-level differential association, Akers (1998) has maintained a career long interest in integrating the individual-level explanation with the structural component. It is in the 1998 publication of *Social Learning and Social Structure* that Akers makes his most complete statement of that link as well as providing preliminary evidence regarding his proposed model. The model includes four different groups of social structural factors whose effects on crime he argues are mediated by social learning variables. The four social structural factors are differential social organization, differential location in the social structure, social disorganization and conflict, and differential location in primary, secondary, and reference groups.

Akers uses differential social organization to refer to "…ecological, community or geographical differences across systems…" (332) such as neighborhoods and the division between urban and rural. Importantly, Akers (1998) argues that differences in crime across these areas are related to structural and/or cultural differences, some of which are known and some of which are not. Differential location in the social structure refers to factors such as age, gender, race, and class that are indicative of a group's position in the social structure. Next, Akers draws from three major social structural theories to discuss the effects of social disorganization, anomie, and culture conflict. He argues that when social cohesion and integration are disrupted, disorganization and anomie result in an area or group, with a resulting increase in the crime rate. Further, social, political, and/or economic injustices may result in the development of different belief systems that lead to conflicts and, through conflicts, higher area or group rates of crime. Finally, Akers posits that individuals are differentially located in primary, secondary, and reference groups such as family, school, and peers, in such a way as to lead to differences in social control and socialization.

As is seen above, Akers leaves unspecified much of what comprises the essential aspects of differential social organization. A recent test of the theory by Lee et al. (2004) leaves questions about the important social structural and cultural factors that make up differential social organization largely unanswered. The one measure of differential social organization is size of the community. Akers explication of the link between differential social organization and differential association, though, and his empirical tests of the link are important steps in understanding how differential social organization contributes to our understanding of individual criminal involvement.

Another theorist whose work has advanced the understanding of differential social organization is Matsueda (2006), a longtime advocate of Sutherland's work. In the current work, his focus is on adding to the theory a dynamic component that is implied, but never explicitly recognized. He draws his thinking about the dynamic aspect of organization from Sutherland's (1943) ideas that social organization involves the development of consensus over goals to be achieved and the means by which to achieve that goal. He then draws from Mead's (1934) theory of symbolic interaction and ideas of social control to discuss just how it is that consensus over goals and means is built up and turned into action. He argues Mead's ideas are consistent with Sutherland's differential social organization for both involve the development of meaning regarding goals and means as well as the development of a self that is built up over time and in interaction with others as solutions to common problems are sought.

Central to this statement is how collective action as a solution to a common problem, such as crime control or safety, is built up. Collective action, which he views as a types of social organization, is defined as "those acts commonly defined as occurring outside of institutional contexts in informal groups or gatherings, tending to be more spontaneous and creative, and requiring the building of coalitions and consensus in the absence of a strong normative system" (19). Understanding how a problem results in collective action depends on several factors including the frame and individual thresholds. Framing refers to how the problem is defined. Individual thresholds deal with how individuals vary on the proportion of people in a collective action that it takes for them to decide to join in. One example he uses to illustrate his points deals with collective efficacy. To view this in terms of process and action, he argues, involves consideration of how residents frame the issue of crime in their neighborhood, how much value individuals in the neighborhood place on safety, and how capable individuals in the neighborhood are in winning others to their view, as well as the mix of individuals with various levels of individual thresholds. Collective efficacy does not depend then just on social networks in the neighborhood then on "efficacious individuals" embedded in the network and their work in framing a problem and engaging others.

While Matsueda radically expands our understanding of the role of processes in social organization, recent work by Sampson and Graif (2009) on types of social capital illustrates important ways of thinking about key components of differential social organization across neighborhoods. Echoing somewhat Matsueda's emphasis on "efficacious individuals," Sampson and Graif distinguish between types of social capital at the resident and leadership levels of neighborhoods. They argue that the social capital that may reside in the abilities of the residents is not the same as that which is involved with neighborhood leaders and the two types of social capital do not necessarily correlate. Analyses of social capital need to distinguish then between resource possibilities of social networks, leadership in the neighborhood, and organizational capacity. Along with these structural dimensions are cultural aspects of neighborhood life including the normative climate and expectations about safety. Using these different dimensions of social capital, their analysis uncovers four different types – cosmopolitan, conduct norms, urban village, and institutional alienation – that vary on the levels of social capital existing among residents and leaders.

Beyond these three highlighted works are others who have recently explored various aspects of social organization at the neighborhood level and its effects on crime. Browning's (2009) work on "negotiated coexistence" illustrates how the inclusion of offenders in neighborhood social networks which depend upon reciprocity and exchange can lower the ability to control the behavior of individuals in the network. Haynie et al. (2006) present a rare empirical assessment of neighborhood disadvantage and its effects on exposure to violent peers. They rightly note that while many predict neighborhood structural characteristics will affect whom youths will interact with and what they will be exposed to, empirical research testing these ideas is limited. Similarly, Harding (2009), drawing on work such as Cloward and Ohlin (1960) on integration of age groups, is interested in expanding knowledge on the transmission of culture in disadvantaged neighborhoods. He explores how neighborhood disadvantage increases time spent by adolescents with older youths. Finally, De Coster et al. (2006) are also interested in culture as found in the "street milieu." In their study, they examine that, for example, the effects of deviant peer relationships, easy access to firearms, witnessing violence, being the victim of violence, and expectations of death all affect involvement in violence by shaping the ability of families to control the children.

Controversies and Open Questions

How different types of social organization are related to crime rates across groups and areas is a question that has fascinated criminologists since the field's inception. An increasing emphasis on social process theories emerged in the 1970s, turning attention away from social structural theories as criticisms of these theories grew. Interest never died completely out though, and work on social organization continued, slowly but surely. Neighborhood research on social disorganization, collective efficacy, and social capital is an example of the continued vitality in this area. Though there is a great deal of interesting work being done, questions remain for exploration.

Researchers in interested in social disorganization theory and neighborhood crime have been exploring some aspects of neighborhood organization for and against crime since Shaw and McKay. Initially research focused on social structural factors such as poverty, mobility, and heterogeneity and connected them to social networks and organizational participation as factors important in understanding the ability of residents in a neighborhood to agree upon and work towards common goals. More recent work has radically expanded our ideas on key factors in understanding neighborhood organization – collective efficacy, bridging and bonding social capital, institutional capacity, neighborhood leadership – many of which are now being connected to neighborhood types. The focus here is on the ability to organize *against* crime found in resources and structures which aid in the development of a common goal and action towards achieving that goal. Many of the ideas need further explication and testing, but a solid ground has been laid for future work.

If Sutherland is right though, there is a need to explore the factors that shape organization *for* crime. Cloward and Ohlin (1960) saw integration of different age levels along with integration of the conventional and criminal worlds as important for opening opportunities for success in the criminal world. Harding's (2009) research on integration of age levels and its effects on exposure to neighborhood culture shows the importance of this. What other factors, structural or cultural, are key to understanding organization for crime? It seems that much of what has been done in this area is related to culture – alienation from the larger culture and the police, for example, as well as understanding of cultural adaptations to living in high-crime areas. Related to this is the need to examine how organization for and against crime are interrelated.

Perhaps the most interesting and difficult aspect of differential social organization is grappling with the role of culture. Most seem to have moved away from notions of a deviant subcultural where residents are predicted to adhere to beliefs supportive of violence or crime to more subtle notions of rationalization and adjustments to poverty and feelings of alienation from the larger society and the police. Key to continued development is distinguishing culture and its effects from those of social structure. Work thus far, though often identifying cultural and social structure factors, often fails to distinguish them in ways that aids in our understanding of their independent effects much less their relationship. Without a careful delineation, the ability to untangle the relationship between social structural factors and their effects from those of culture will continue to limit our understanding of both.

Tackling culture leads to another question. What about the relationship between social disorganization and social control theories on the one hand and social disorganization, differential social organization, and social learning on the other? Kornhauser's (1978) powerful criticism that Shaw and McKay originally presented an untenable mixed model because of the contradictory assumptions behind control and cultural deviance seemed to have the effect of banishing consideration of culture from most analyses of neighborhood crime rates under the social disorganization framework. In fact, social disorganization is most commonly argued to be the macro level equivalent of social control theories focusing, while cultural deviance and social learning are similarly linked. More and more it seems that social control and social learning are linked processes but are these different versions of each?

Related Entries

- ▶ Differential Association Theory
- ▶ Disadvantage, Disorganization and Crime
- ▶ Early Chicago School Theory
- ▶ Neighborhood Effects and Social Networks
- ▶ Social Capital and Collective Efficacy

Recommended Reading and References

Akers R (1998) Social learning and social structure: a general theory of crime and deviance. Northeastern University Press, Boston

Browning C (2009) Illuminating the downside of social capital: negotiated coexistence, property crime, and disorder in urban neighborhoods. Am Behav Sci 52:1556–1578.

Cloward R, Ohlin L (1960) Delinquency and opportunity. Free Press, New York

Cressey D (1960) Epidemiology and individual conduct: a case from criminology. Pacific Sociol Rev 3:47–58

Cressey D (1968) Culture conflict, differential association and normative conflict. In: Marvin W (ed) Crime and culture: essays in honor of Thorsten sellin. Wiley, New York, pp 43–54

De Coster S, Heimer K, Wittrock S (2006) Neighborhood disadvantage, social capital, street context, and youth violence. Sociol Q 47:723–753

Gaylord M, Galliher J (1988) The criminology of Edwin Sutherland. Transaction Publishers, New Brunswick

Harding D (2009) Violence, older peers and the socialization of adolescent boys in disadvantaged neighborhoods. Am Sociol Rev 74:445–464

Haynie D, Silver E, Teasdale B (2006) Neighborhood characteristics, peer networks, and adolescent violence. J Quant Criminol 22:147–169

Kornhauser R (1978) Social sources of delinquency: an appraisal of analytic models. University of Chicago Press, Chicago

Lee G, Akers R, Borg M (2004) Social learning and structural factors in adolescent substance use. West Criminol Rev 5:17–34

Matsueda R (2006) Differential social organization, collective action and crime. Crime Law Soc Ch 46:3–33

Mead GH (1934) Mind, self and society. University of Chicago Press, Chicago

Sampson R, Graif C (2009) Neighborhood social capital as differential social organization: resident and leadership dimensions. Am Behav Sci 52:1579–1605

Shaw C, McKay H (1942) Juvenile delinquency and urban areas. University of Chicago Press, Chicago

Sutherland E (1942[1973]) Development of the theory. In: Karl S (ed) Edwin Sutherland: on analyzing crime. University of Chicago Press, Chicago, pp 13–29

Sutherland E (1943) Crime. In: William O (ed) American society in wartime. University of Chicago Press, Chicago, pp 185–206

Sutherland E (1957) Principles of criminology, 4th edn. J.B. Lippincott Company, Chicago, revised

Sweetser F (1942) A new emphasis in neighborhood research. Am Sociol Rev 7:525–533

Tannenbaum F (1938) Crime and the community. Ginn and Co, Boston

Disadvantage, Disorganization and Crime

Karen F. Parker and Richard Stansfield
Department of Sociology and Criminal Justice,
University of Delaware, Newark, DE, USA

Overview

A fundamental community-level theory, social disorganization theory posits that crime and delinquency are more pronounced in areas characterized by persistent poverty, population heterogeneity, and residential mobility, which combine to disturb the capacity of neighborhoods to maintain informal social control. These ideas have been well investigated and empirically supported, leading social disorganization theory to become the most well-known theory of neighborhood crime in the field today. Yet as time passes, scholars have turned their attention to the ways that central cities have changed. This includes the influx of service-based economies to city centers, the spatial concentration of poverty, and how joblessness has overcome the economic prospects and hopes of so many urban residents. And while the removal of manufacturing industries affected jobless rates, preexisting racial residential segregation meant the ill effects of deindustrialization were predominantly felt among minorities and the poor. It is these arguments, directly associated with Disadvantage theory, that have led researchers to incorporate both disorganization and disadvantage into their community-level research on crime.

The interconnections between disadvantage and disorganization are the central focus of this entry. After providing a brief review of the fundamental ideas driving social disorganization and disadvantage theories, some of the advantages to incorporating both approaches jointly when attempting to capture community conditions, neighborhood processes, and crime rates will be discussed. At the same, it is important to acknowledge that neighborhoods are not influenced just by structures and process within (Kubrin and Weitzer 2003; Sampson et al. 2002). Community studies of disadvantage and disorganization must also account for changes in the larger society, such as political changes in crime policies, growing diversity in the US population, and severe downturns in the economy and/or times of economic recession. To illustrate, the 1990s saw unprecedented levels of immigration from Central and Southern America. The majority of new arrivals settled in urban neighborhoods within a few port cities. Yet despite facing high levels of community disadvantage in these neighborhoods, the arrival of new immigrant groups has not led to higher rates of crime, raising questions about additional factors overlooked by community-level research. These critical issues are offered as aspects for future research or open questions, as they have significant implications for disorganization and disadvantage.

The Fundamental of Social Disorganization and Urban Disadvantage Theories

Social Disorganization Theory

One of the most fundamental approaches to the study of violence emanates from the Chicago school research of Shaw and McKay. In essence, Shaw and McKay (1942) argued that neighborhood dynamics lead to social disorganization in communities, which account for the variations in crime and delinquency. Under this line of inquiry, Shaw and Mckay (1931, 1942) set out to explain variation in offending informed largely by the work of Park et al. (1925). Park was a proponent of the concentric zone modeling of cities, observing that industries tend to be located within inner-city zones and cities grow naturally from the inside out. Shaw and McKay believed that characteristics of the inner-city zone could explain delinquency and crime due to three elements – low economic status, ethnic heterogeneity, and residential mobility. First, communities with low economic status lack adequate money and resources to generate the needed formal and informal controls necessary to reduce crime and delinquency. Second, racial and ethnic heterogeneity, which is often accompanied by fear and mistrust, impedes communication and patterns of interaction in a community. Third, residential mobility was said to disrupt a community's network of social relations by acting as a barrier to the development of extensive friendship networks, kinship bonds, and local social ties.

Believing the conditions of disorganization to be structurally conducive to crime rates, Shaw and McKay (1931) examined robbery and mental illness and found that rates were higher in the inner-city zones, where residents lived adjacent to industry in conditions of population growth and environmental hazard. Rates progressively declined, moving away from the inner city into transitional zones however, and lowest of all on the outskirts of the city in the commuter zone. An equally important aspect of their work was that rates of delinquency remained high in the inner-city zones regardless of the ethnic group residing in those areas. That is, they suggested that structural conditions of disorganization were the key to understanding urban crime, not the characteristics or motivations of individuals. Over time, "social disorganization" became known as referring to the inability of a community structure to realize the common values of its residents and maintain effective social control (Bursik 1988). That is, social disorganization addressed the structural barriers that impede the development of the formal and informal ties that promote the ability of the community to solve common problems.

The social disorganization model however struggled to hold popularity within criminology throughout much of the 1970s and 1980s, not least because as Bursik and Grasmick (1993) argued, the authors failed to supply a refined concept of social disorganization and how its constituent parts were each related to crime and delinquency. In an effort to revitalize interest in the approach, Bursik (1988) noted 5 key reasons for the failure of social disorganization to be applicable to contemporary society and criminology. These included normative assumptions made about organization in neighborhoods and the dependency on official sources of data for testing social disorganization, noting that many neighborhoods have their own informal mechanisms of dealing with delinquency; the study was cross-sectional, yet cities are dynamic and not static; and simply at the time, criminology was more interested in theories of crime focused on explaining individual tendencies to offend. Perhaps the key argument, however, was that social disorganization had not reached its full potential as a community theory yet could do so if these flaws were addressed in future work.

A renewed interest in the theory has been witnessed and research flourishes today largely because of the ability of criminologists such as Robert Sampson to show the relevance of Shaw and McKay for understanding crime in contemporary society. This renewed interest has led to some important developments. First, researchers have extended the types of structural conditions in urban areas that impede social control, including population size, poverty, and family disruption (see Sampson and Groves 1989). As another example, Bursik and Grasmick (1993) expand Shaw and McKay's claims of the inability of communities to maintain social control, offering a "systemic model" linking community control to crime rates at three levels: private (friendships or households), parochial (organizational ties), and public (citywide efforts to regulate neighborhoods via police presence).

Third, the mediating linkages between structural conditions, social control, and crime rates have been further articulated. While Shaw and McKay identified the capacity of the community to control group-level dynamics as a key mechanism linking community characteristics with crime rates, Sampson and Groves (1989) clarified the types of social networks in a community at both the informal (e.g., friendship ties) and formal (e.g., participation in community organizations) level. They suggested that communities unable to (1) control teenage groups through collective social control, (2) form informal local friendship networks, and (3) offer local participation in formal and voluntary organizations will experience high rates of crime and delinquency (Sampson and Groves 1989). The authors found that structural characteristics did predict levels of disorganization, and in turn, the weakly organized communities had higher rates of crime. These findings were replicated 10 years later using updated responses from the British Crime Survey (Lowenkamp et al. 2003) highlighting the strength and stability of the findings.

More recently, Sampson et al. (1997) conceptualize community ties a bit differently to include the inability of community residents to collectively display trust and deal with problems (such as disorderly teens or drugs) via “collective efficacy.” Collective efficacy has been found to largely mediate the relationship between structural disadvantage and crime (Sampson et al. 1997). Not only does this new line of inquiry show why problems differentially exist between neighborhoods, but it also shows that the neighborhoods experiencing problems are less able to do anything about it. For example, the focus of trust and cohesion suggests that when thinking about intervening and exercising informal social control – neighbors are able to know the likelihood of a neighbor collaborating or backing them up based on trust.

The dynamic element of collective efficacy thus makes an enormous leap forward in explaining the disorganization, social control, and crime relationship. This is important given that the direct evidence on traditional disorganization factors such as residential instability and ethnic heterogeneity has been more mixed in recent research, particularly in the case of ethnic heterogeneity (Morenoff et al. 2001; Sampson et al. 2002). Yet there are limitations, particularly when identifying the key factors behind precisely how or why ties are activated and mobilized for the purpose of social control. Recent research also questions the causal pathway of disorganization and crime, in which crime may cause feedback effects on disorganization, disorder, and a community’s capacity to build social ties (Kubrin and Weitzer 2003). Thus, while these efforts expand upon and extend social disorganization theory, the field continues to call for further conceptual refinement, inclusion of other factors rarely examined by disorganization researchers, and rigorous methodology to unpack the dynamic process between disorganization and crime.

Disadvantage Theory

Numerous scholars have documented the growing disadvantages in urban areas and how these disadvantages are spatially concentrated in poor black areas particularly. Building on this foundation, criminology has attended to the structural aspects of concentrated disadvantage in producing urban crime. Wilson (1987) claims that the increase in economic marginalization of blacks in the inner city was due, in a large part, to a set of spatial and industrial changes in the political economy. In *The Truly Disadvantaged*, he points to deindustrialization, coupled with suburbanization of middle-class African Americans, as central to the rise in poverty and social isolation inner-city residents. Central to his arguments, Wilson illustrates the dramatic shift away from local manufacturing-based economies during the 1970s and 1980s, particularly in the Midwest and Northeast regions of the United States (Kasarda 1995). Advances made in technological industries contributed to this shift, as did the motivation to secure cheaper labor and the growing demand for service industries. That is, the process of deindustrialization marks a change in the industrial mix of urban areas or, more specifically, when the share of jobs shift from manufacturing to administrative and information services in many American cities (Kasarda 1995). The process of deindustrialization had a number of adverse outcomes on city residents.

First, the movement of manufacturing jobs into the suburbs or overseas and shift in industrial mix of local economies resulted in a spatial mismatch, which emphasizes the geographic changes in the number of low-skilled jobs in the inner city and limited mobility of African Americans due to residential segregation. Essentially a “mismatch” occurs as residential segregation prevents blacks from following employers to the suburbs. The low wages and earnings associated with manufacturing work served as a disincentive for inner-city residents to seek longer commutes for work outside their neighborhoods and city boundaries. Yet inside city boundaries, as central cities are transformed into service-based economies and centers for administration and information processing

(Kasarda 1995), these areas see a reduction in the number of potential employment opportunities available to the low-skilled worker. That is, the requirements of the new economy also created a "skills mismatch" between workers and the service jobs available, further disadvantaging blacks living in the inner city.

Second, while the removal of manufacturing jobs affected unemployment or jobless rates, preexisting segregation in labor markets through discrimination in hiring and denying equal opportunity meant that the ill effects of deindustrialization were predominantly felt among minorities and the poor. The shift in labor demand had real consequences for minority groups and women. Scholars have linked the high rates of joblessness among black males to the transformation of the US economy or industrial restructuring (Wilson 1987). And while the literature examining the influence of restructuring on women is small and inconclusive, some research suggests women benefited from industrial restructuring, as the shift increased the number of jobs in female-dominated service occupations, while other scholars argue that black women lost rather than gained with the expansion of service-based industries. For example, Smith and Tienda (1987) find that Latino and black women faced greater declines in jobs than Asian and white women, suggesting that the already high levels of economic marginalization among women was further concentrated among minorities. Adding to the crippling effects of industrial restructuring on women, the disabling of welfare or the elimination of federal welfare entitlements such as AFDC to black women with children whom were disproportionately represented, only further guaranteed that African American women would be vulnerable to shifts in the local economies. To capture the social transformation of urban areas, Wilson's term "concentration effects" reflects the continued deterioration of employment opportunities and job networks, decline in schools, and diminishing number of marriageable partners, thus contributing to higher levels of family dissolution, which only further aggravated the weak labor attachments and conventional role models for inner-city residents (Wilson 1987, p. 58).

Taken together, these scholars focus on how blacks are residentially concentrated in poor urban areas with limited access to job networks and opportunities to participate in formal local labor markets. Racial residential segregation and black economic disadvantage (joblessness, poverty concentration, and deindustrialization) are thus considered the macrolevel sources of racial discrimination in the urban environment, leaving blacks to face a more disadvantaged residential environment than the poor of other racial groups. A wealth of recent empirical work has documented that these precise structural conditions are uniquely related to urban violence (Krivo and Peterson 1996; Parker 2008).

In sum, there is strong evidence that factors of concentrated disadvantage and racial isolation remain direct predictors of many criminal outcomes. But while these measures provide a telling story in their own unique relationships to crime, they are not insulated from measures of disorganization. Dynamic neighborhood processes, particularly those related to collective efficacy, are not produced in a vacuum; rather, they appear to emerge mainly in environments with sufficient socioeconomic resources and residential stability (Sampson et al. 2002). As Sampson and Wilson (1995, p. 53) conclude, "the intersection of race, place, and poverty goes to the heart of our theoretical concerns with society and community organization."

Disadvantage and Disorganization: Why They Should Be Examined Together?

This section discusses the benefits and advantages to the understanding of community conditions and crime by assessing disadvantage and disorganization jointly. While neglecting other important issues, arguments here focus on three major advantages: (1) by placing the focus on local communities rather than individual circumstances or motives, we more fully understand the structural processes that lead to crime; (2) together, the role that race plays in the USA becomes more lucid, as race is largely infused with economic status and place throughout the

United States; and (3) the ability to move beyond cross-sectional examination of crime rates to incorporate how changes occur in communities and crime over time.

Community Focus

Sampson and Wilson (1995) largely accepted the tenets of Shaw and McKay's disorganization theory; however, they disagreed that the disorganization and growth out of the city was part of a "natural process" that cities went through. Instead they argued that racial discrimination, and the consignment of African Americans into inner-city neighborhoods based on economic and political decisions, was the key to understanding disorganization and crime in contemporary American society. Through making this connection, Sampson and Wilson have helped illustrate the connection and overlap between disadvantage and disorganization. A wealth of structural research has also suggested that variations in disorganization are intimately linked to racial inequality and racial segregation (Peterson et al. 2006). Accordingly, recent research has utilized disadvantage and disorganization together to provide a greater understanding of community processes. Grattet (2009), for example, uses Wilson's notion of concentrated disadvantage to understand social disorganization and defended neighborhoods, noting that neighborhoods without adequate resources lack the legitimate means (and informal control mechanisms) of defending its area from outsiders.

This is a positive trend in research, moving away from an exclusive focus on structural factors toward placing the community at the heart of the discussion of disadvantage and disorganization. As Sampson and Wilson (1995) describe, disadvantaged communities are not only more likely to be deprived of resources necessary to mobilize crime control, but the resulting isolation of communities affects the capacity to form cohesive networks based on mutual trust and common goals. Due to political and economic forces, the disappearance of jobs (Wilson 1987, 1996), and the increasing social isolation of the urban poor, cultural values in these communities have had the potential to shift away from middle-class norms of achieving success to subcultural norms where the pursuit of status and respect replaces conventional means (Anderson 1999). Indeed as Anderson describes, the structural conditions of the neighborhood are linked to public cynicism and trust in formal social control and serve to reshape residents' values. Studies from this perspective, including the work of Elijah Anderson (1999), highlights that structural conditions and cultural responses to these conditions in which people find themselves jointly shape neighborhood organization and neighborhood crime. As those with the misfortune of living in segregated, isolated, and disorganized communities are also then exposed to conditions and values that may permit their criminal involvement, a community focus is critical.

Race, Ethnicity, and Place

Racial groups continue to face different social and economic realities (see, e.g., Sampson and Wilson 1995; Wilson 1987). Discrimination and racial competition in opportunity structures are core issues that stratify racial groups in the urban context. That is, structural forms of racial discrimination, such as residential isolation, poverty concentration, and segregated labor markets, not only stratify groups spatially but can serve as a source of conflict between them. Disorganization and disadvantage arguments allow race to be examined as a structural yet local, as opposed to, individual or biological feature of communities. In this way, this literature advances the study of race, urban inequality, and violence by highlighting the core conditions that lead to racial differences in life circumstances and crime within neighborhoods and US cities.

Second, by combining theories of disorganization and disadvantage, scholars acknowledge the persistent disadvantaged position of African Americans as a structural feature of urban areas. For example, Massey and colleagues (see Massey and Denton 1993) provide evidence of whites' exclusion of black participation in the local labor market and whites' resistance to blacks residing in integrated communities. Research also reveals the inequalities across race and ethnic groups as they compete in the labor force and face differential treatment as a result of economic restructuring. A striking feature of contemporary United States is that black employment remains lower than white employment despite recent narrowing of educational and occupational inequalities. Another important finding is that blacks face greater barriers to residential mobility than other race and ethnic groups. According to Massey and Denton (1993), blacks remain spatially isolated and residentially segregated from whites at all levels of economic status, while Hispanics and Asians have experienced clear improvements as their economic status increases. Similarly, Wilson (1987) argues that blacks reside in areas of extreme poverty concentration, a reality not known to poor whites.

By examining disorganization and disadvantage together, one gains insights into the ways that race becomes embedded in the structural characteristics of areas, resulting in the racial disparities in crime rates so often found in empirical studies. Yet as Sampson and Wilson (1995) note, the ability of researchers to make meaningful comparisons about the conditions that produce racial violence is significantly hindered by the fact that so few whites reside in areas of concentrated disadvantage when compared to blacks. Thus, as research continues to explore the interconnections between disorganization and disadvantage, it is important to consider not only the level of disparities between racial groups in structural characteristics, opportunity structures like access to jobs, and rates of crime and delinquency but also what larger historical and society forces are allowing those disparities to persist.

Change

Fundamental to Chicago-style criminology is its attention to dynamic changes in urban neighborhoods and crime. In fact, the changing nature of, and level of spatial interdependency between, neighborhoods is embedded in the social disorganization tradition (Morenoff et al. 2001). Yet the element of change, once core to the theory, has been largely neglected in the existing research. In an effort to sway researchers, Byrne and Sampson (1986, p. 17) called for longitudinal analysis when they stated: "By far the vast majority of ecological studies have examined the relative effects of structural characteristics on crime in cross-sectional analysis at one point in time. This is somewhat ironic given that classic ecological theory is concerned with the processes of change in urban areas." Some 20 years later, Sampson (2002) reiterates this troubling trend toward cross-section research, further arguing that the lack of research has caused a series of fundamental questions concerning our understanding of cities and communities.

Cross-sectional designs largely dominate the macrolevel research on crime. In fact, one can find few examples of empirical research that examines changes in macrolevel characteristics and crime rates over time (see examples in Messner et al. 2005). By far the greatest impediment to longitudinal analysis is data, although some researchers have overcome this issue. Conducting national time-series analyses of white and black arrests rates, LaFree and colleagues find differences in the influence of economic indicators on white and black arrest rates over time. For example, based on their longitudinal analysis between 1957 and 1990, LaFree and Drass (1996) found that income inequality (as a measure of resource deprivation) has a positive impact on the changes in white and black arrest rates, but absolute economic well-being (their composite measure that combines median income and unemployment) did not significantly influence homicide rates for either racial group. While studies have begun to examine changes at the city level (or higher levels of aggregation), research focusing on community-level changes is still largely missing from the literature. Thus, examining changes in both communities and crime over time is central to disorganization and disadvantage arguments. The full set of conditions which may lead to disorganization can only be captured when examining long-term processes of urban disadvantage and development (Kubrin and Weitzer 2003).

Future Directions and Open Questions for Disadvantage and Disorganization

While much has been gained by combining arguments of disadvantage and disorganization in the study of urban crime, more work is warranted. So much of the empirical work on communities and crime has focused on how to delineate the connections between community characteristics and crime within the community itself. However, larger societal processes heavily influence communities as well. Three areas that should be accounted for in future work on disadvantage and disorganization include (1) trends in Hispanic immigration and growing diversity, (2) the move toward more conservation crime policy and the impact of mass incarceration on urban communities, and (3) the need to expand the ways we measure the local economy, particularly as the economic recession looms on.

Hispanic Immigration and Growing Diversity

As discussed, structural factors such as racial segregation act to concentrate and isolate disadvantaged minorities into communities with high rates of population turnover, systematic flight of the most qualified residents, few professional employment opportunities, and high rates of family disruption (Bursik 1988; Massey and Denton 1993; Sampson and Wilson 1995; Wilson 1996). As new immigrants have typically settled in these communities, and because immigration may be associated with conflicting norms and the inability of people to agree to common values and reach common goals (Bursik 1988; Sampson et al. 1997), recent immigration has been argued to predict higher rates of crime. New immigrants may also face additional economic barriers and blocked opportunities which make it difficult to abide by conventional societal norms. As an example, new arrivals may initially be faced with limited opportunities entering into the labor market due to language barriers, a lack of well-paying jobs within close proximity, and discrimination (Martinez 2002).

Despite the fact that new immigrants often reside in communities that are characterized by disadvantage and disorganization, research tends to show that the presence of new Latino immigrants results in either a negligible or negative effect on crime (see, e.g., Martinez 2002; Ousey and Kubrin 2009). Many explain the "Latino paradox" as the result of the importance placed on family and religious conservatism in Latino communities, which may counteract the criminogenic effect of the economic disadvantage that often accompanies the immigration experience. The emphasis placed on family and religion may serve to lower crime by increasing informal social control on family members through a greater sense of obligation, reference group identity, and community support. This may be particularly important for young Hispanics, a group that is disproportionately large as compared to the white population and at risk for criminal offending due to the link between age and crime.

Contemporary immigration is also vastly different from immigration of Shaw and McKay's day, not least because of the human capital and financial resources many immigrants bring. Even among the less-skilled and less-educated recent arrivals, research has documented high rates of employment among recent Hispanic immigrants, even if the employment happens to be at a low-wage level (Martinez 2002). Jobs that are considered "undesirable" and/or those offering wages that are low by American standards are often acceptable to recently arrived immigrant laborers as they are offer much better wages than similar jobs in their country of origin. In addition to basic economic sustenance, the higher rates of Hispanic labor market involvement are important for an understanding of offending patterns because employment provides stability and regulation of daily behavior by requiring that people follow consistent schedules and spend time engaging in conventional activities (Wilson 1996).

Information emerging from this body of immigration-crime literature thus provides interesting caveats to traditional theories of disadvantage and disorganization. Specifically, whether there is a positive or negative relationship of neighborhood conditions and crime for Hispanics may depend on a number of factors such as local labor market structures, the presence of racial discrimination facing different groups, and the presence of a dual frame of reference in a community, through which residents evaluate their socioeconomic conditions relative to past realities in their country of origin (Martinez 2002). The presence of vibrant co-ethnic communities awaiting recent arrivals, as in the case of the Cuban enclave in Miami, for example, offers another example of how some communities may be able to provide a protection against outside forces of discrimination while offering employment opportunities through informal social networks within. A challenge for criminologists lies in how to use social disorganization theory and theories of disadvantage together in order to explain how one ethnic group may have overcome conditions of disadvantage while many low-income blacks have yet to do so.

Crime Policy and Mass Incarceration

Another important open question for scholars of disorganization and disadvantage is the extent to which structural characteristics influence arrest and incarceration and the resulting implications back on the community itself. Surprisingly, criminologists have paid little attention to the determinants of arrest rates across time and space, despite the strong tradition of neighborhood studies examining the relationship between social and economic inequality and crime arrests (Mosher 2001). Yet it is hard to deny the impact of an era of mass imprisonment and the war on drugs on neighborhood organization and disadvantage. Mass incarceration may have one of two key implications for ecological analyses studying social disorganization and disadvantage. Firstly, in communities of high disadvantage, a significant proportion of the population may no longer be present in the neighborhood, as they were three decades ago. Recent research has also revealed that important distinct relationships exist between concentrated disadvantage and the use of social control against blacks, when compared to whites (Parker et al. 2005). These authors argue that the concentration of black disadvantage and racial isolation may influence black arrests because these factors serve to amplify perceived group differences, thus increasing antiblack dispositions, which in turn may increase the pressure on police to exert control. Given the racial and economic disparities in the experience of arrests and incarceration, some communities have experienced more concentrated levels of incarceration than others. And there have been well-documented neighborhood problems associated with the removal of significant numbers in a community, such as the removal of male role models for children and employable residents. If estimates of Clear et al. (2001) are any indication, among the worse-off communities as much as 30 % of the population may be gone.

In addition to the role crime policy and mass incarceration has played in removing family and community role models from neighborhoods, it is as important to consider that most of those incarcerated during the mass incarceration era are eventually released back into the community. This issue of reentry into communities, and its implication for the organization of neighborhoods, is also critical to consider. In interviews in Tallahassee, Florida, Clear and colleagues (2001) target understanding the implications of this process of removing and returning offenders, for social networks, social capital, and subsequently informal social control. The impacts go far beyond financial and fundamentally alter the shared sense of identity, interpersonal networks, self-esteem levels, and a community's reputation as a good place to live. The importance of both of these arguments is brought to light if we consider the more recent work on collective efficacy (Sampson et al. 1997). As a theory focusing on trust, community ties, and relationships with neighbors, both the removal and reentry of convicted community members are likely to have significant implications for family disadvantage, informal neighborhood networks, mutual trust, and shared goals in the neighborhood. Based on recent data from the National Institute of Justice that suggests that the era of mass incarceration may be stabilizing, and possibly even ending, considerations of mass incarceration and crime policy must be a critical element to future studies of social disorganization and disadvantage.

Economic Downturn and Times of Recession

Given the historical importance of residential patterns, economic conditions, and other community-level measures of disadvantage or disorganization, it is not surprising that the looming economic recession presents researchers with an important call to expand the ways we measure the local economy. To illustrate, the massive trend in housing foreclosures nationally could easily contribute to neighborhood conditions such as residential mobility and disadvantage. Research has begun to examine the possible effects of foreclosure on neighborhood levels of crime. As Katz et al. (2011) discuss, the concern is less about the act of foreclosure itself and more about the neighborhood conditions that a foreclosure is believed to create, compromising the social environment much in the same joblessness has been thought to do (e.g., Wilson 1996). In this vein, one may expect that neighborhoods with high rates of foreclosure will similarly begin to undergo a process in which residents withdraw from a community, both physically and psychologically, and in turn, informal social control mechanisms break down. A conclusion on the importance of such measures may be, at best, premature, especially as the preliminary evidence suggests that foreclosed homes may not have the long-term negative impact on crime that many scholars have feared (Wallace et al. 2012).

Nevertheless, within the current economic climate, new measures of disadvantage and disorganization can be helpful to expanding our knowledge of what structural factors are important to consider within particular time and spatial contexts. Traditional measures of disadvantage, social ties, and collective efficacy have proven effective, but may not fully account for variation in neighborhood crime in times of recession or severe economic downturn. Other factors are thus also important to consider, like housing foreclosure and lacking consumer sentiment. By studying the interconnections between disorganization and disadvantage, particularly as influenced by macroeconomic conditions, wider social patterns in migration, and in the context of decisions by political elites, the utility of disorganization and disadvantage remains more pertinent than ever.

Related Entries

- ▶ Crime and the Racial Composition of Communities
- ▶ Early Chicago School Theory
- ▶ History of Criminological Theories: Causes of Crime
- ▶ Neighborhood Effects and Social Networks
- ▶ Poverty, Inequality, and Area Differences in Crime
- ▶ Race and Ethnicity in Social Disorganization Theory
- ▶ School Structural Characteristics and Crime
- ▶ Social Disorganization and Terrorism
- ▶ Understanding Victimization Frequency

Recommended Reading and References

Anderson E (1999) Code of the street: decency, violence, and the moral life of the inner-city. WW Norton, New York

Burgess EW (1967) The growth of the city: an introduction to a research project. In: Park R, Burgess E, McKenzie RD (eds) The city. University of Chicago Press, Chicago

Bursik RJ Jr (1988) Social disorganization and theories of crime and delinquency: problems and prospects. Criminology 26:519–551

Bursik RJ, Grasmick HG (1993) Neighborhoods and crime: the dimensions of effective community control. Lexington, New York

Byrne JM, Sampson RJ (1986) The sociology ecology of crime. Springer, New York

Clear TR, Rose DR, Ryder JA (2001) Incarceration and the community: the problem of removing and returning offenders. Criminology 47(3):335–351

Grattet R (2009) The urban ecology of bias crime: a study of disorganized and defended neighborhoods. Soc Probl 56(1):132–150

Kasarda JD (1995) Industrial restructuring and the changing location of jobs. In: Harrell AV, Peterson GE (eds) Drugs, crime, and social isolation. The Urban Institute, Washington, DC

Krivo L, Peterson RD (1996) Extremely disadvantaged neighborhoods and urban crime. Soc Forces 75:619–648

Kubrin CE, Weitzer R (2003) New directions in social disorganization theory. J Res Crime Delinq 40(4):374–402

Lafree G, Drass K (1996) The effect of changes in intraracial income inequality and educational attainment on changes in arrest rates for African Americans and whites, 1957–1990. Am Sociol Rev 61(4):614–640

Lowenkamp CT, Cullen FT, Pratt TC (2003) Replicating Sampson and Groves' test of social disorganization theory: revisiting a criminological classic. J Res Crime Delinq 40:351–373

Martinez R Jr (2002) Latino homicide: immigration, violence and community. Routledge, New York

Massey D, Denton N (1993) American apartheid: segregation and the making of the underclass. Harvard University Press, Cambridge, MA

Messner SF, Deane GD, Anselin L, Nelson BP (2005) Locating the Vanguard in Rising and Falling Homicide Rates Across U.S. Cities. Criminology 43:661–696

Messner SF, Raffalovich L, McMillan R (2001) Economic deprivation and changes in homicide arrest rates for white and black youths, 1967–1998. Criminology 39(3):591–614

Morenoff JD, Sampson RJ, Raudenbush SW (2001) Neighborhood inequality, collective efficacy and the spatial dynamics of urban violence. Criminology 39:517–558

Mosher C (2001) Predicting drug arrest rates: conflict and social disorganization perspectives. Crime Delinq 47(1):84–104

Ousey GC, Kubrin CE (2009) Exploring the connection between immigration and violent crime rates in U.S. cities, 1980–2000. Soc Probl 56:447–473

Park R, Burgess EW, McKenzie RD (1925) The City. University of Chicago Press, Chicago

Parker KF (2008) Unequal crime decline: theorizing race, urban inequality and criminal violence. New York University Press, New York

Parker KF, Stults BJ, Rice SK (2005) Racial threat, concentrated disadvantage and social control: considering the macro-level sources of variation in arrests. Criminology 43(4):1111–1134

Peterson RD, Krivo LJ, Browning CR (2006) Segregation and Race/Ethnic Inequality in Crime: New Directions. In: Cullen FT, Wright JP, Blevins KR (eds) Taking Stock: The Status of Criminological Theory. Transaction Publishers, New Brunswick

Sampson RJ, Groves B (1989) Community structure and crime: testing social disorganization. Am J Sociol 94:774–802

Sampson RJ, Wilson WJ (1995) Toward a theory of race, crime, and urban inequality. In: Hagan J, Peterson RD (eds) Crime and inequality. Stanford University Press, Stanford

Sampson RJ, Raudenbush SW, Earls F (1997) Neighborhoods and violent crime: a multilevel study of collective efficacy. Science 277:918–924

Sampson RJ, Morenoff JD, Gannon-Rowley T (2002) Assessing "neighborhood effects": social processes and new directions in research. Annu Rev Sociol 28:443–478

Shaw CR, McKay HD (1931) Social factors in juvenile delinquency. Report on the causes of crime, national commission of law observance and enforcement. GPO, Washington, DC

Shaw CR, McKay HD (1942) Juvenile delinquency and urban areas. University of Chicago Press, Chicago

Smith SA, Tienda M (1987) The doubly disadvantaged: women of color in the US labor force. In: Stromberg A, Harkess S (eds) Working women, 2nd edn. Mayfield, Palo Alto

Wallace D, Hedberg EC, Katz CM (2012) The impact of foreclosures on neighborhood disorder before and during the housing crisis: Testing the spiral of decay. Social Science Quarterly 93(3):625–647

Wilson WJ (1987) The truly disadvantaged: the inner city, the underclass, and public policy. University of Chicago Press, Chicago

Wilson WJ (1996) When work disappears: the world of the new urban poor. Alfred A Knopf, New York

Disclosure

▶ Naming and Shaming of Corporate Offenders

Discrete Spatial Choice

▶ Crime Location Choice

Discriminant Validity of Disorder and Crime

Jacinta M. Gau[1] and Travis C. Pratt[2]
[1]Department of Criminal Justice, University of Central Florida, Orlando, FL, USA
[2]School of Criminology and Criminal Justice, Arizona State University, Phoenix, AZ, USA

Synonyms

Order maintenance; Quality of life; Zero tolerance

Overview

The broken windows thesis posits a causal relationship between disorder and crime whereby disorder occurs initially and, if left unchecked, ultimately produces serious crime. This causal sequence requires that disorder and crime be empirically separable phenomena; that is, they must possess discriminant validity. If discriminant validity is absent, then the broken windows thesis is fatally flawed because disorder cannot be said to cause crime if disorder and crime are two facets of the same underlying problem. This entry reviews three studies that have spoken to the issue of discriminant validity and have concluded that disorder and crime are not, in fact, unambiguously distinct factors. The future of the broken windows thesis is thus uncertain, and if order maintenance policing does reduce some types of crime, this might be accomplished through a net-widening effect rather than through the informal social control mechanisms that are keys to the thesis.

Introduction

For policy analysts, the question "Where does crime come from?" is subordinate to "What institution of control can be utilized to reduce crime?" In their 1982 *Atlantic Monthly* article that laid out the central tenets of what would become known as the "broken windows thesis," policy analysts James Q. Wilson and George Kelling selected the police as the institution that would get the limelight. Wilson and Kelling argued that the police were the keys to crime reduction and prevention at the community level and that the best way for them to accomplish this would be to proactively target disorderly people and conditions.

Today, the idea of proactive, order maintenance policing has become part of the common vernacular in discussions of police goals and activities. In 1982, however, Wilson and Kelling faced a serious obstacle that impeded their proposed line of reasoning: The professional model of policing and several constitutional rulings issued by the US Supreme Court had made it unacceptable and legally questionable for police officers to interfere in matters not pertaining directly to serious crime. Wilson and Kelling, then, could not merely assert that police should start concentrating their efforts on reducing annoying yet nondangerous, noncriminal behaviors like loitering and panhandling – police leaders and policymakers would have pointed to the various industry standards and court rulings and thrown their hands up at the suggestion. Wilson and Kelling therefore had to be creative.

And they were. They laid out a formula wherein (a) disorder was uncoupled from serious crime and (b) disorder led to serious crime in a causal fashion. The reasonableness of the claim that disorder is well within the legitimate purview of police enforcement activities was thus made obvious. It boiled down to a simple syllogism: Police are responsible for controlling crime, and disorder causes crime if nobody does anything about it, so police are responsible for controlling disorder.

The problem is that Wilson and Kelling offered no empirical evidence for the unlinking of disorder and crime; in fact, there really was no reason to think that Wilson and Kelling's formulation of the disorder-crime relationship was any more valid than, say, Garofalo and Laub (1979) contention that people's fear of "crime" is actually a very general feeling of anxiety and apprehension about neighborhood-level conditions of sociostructural malaise wherein crime and disorder are part of overarching judgments about quality of life. Wilson and Kelling, then, did not *prove* that disorder and crime are distinct from one another and that disorder causes crime – they *speculated* that such was the case.

Wilson and Kelling's oversight may not have been problematic had others stepped up to fill in the gaps, but this did not happen. Instead, the disorder-crime distinction and supposed causal sequencing was accepted as fact, and efforts were launched to put broken windows policing into action in the streets. Doing so violated the basic principle of scientific inquiry that requires theories and their constituent predictions (i.e., hypotheses) to be empirically tested repeatedly before being taken as valid. Scientific testing is hardly a minor detail. The broken windows thesis has assumed a position of preeminence in policing without any strong evidence that Wilson and Kelling were right. Worrall (2006a, p. 379) put it well when he argued that "just because theorists have pushed for the separation of crime and incivility indicators does not mean both concepts are distinct... the incivilities [broken windows] thesis has evolved with little attention to the validity of its measures."

This entry assesses the evidence pertaining to the *discriminant validity* (i.e., the empirical uniqueness) of disorder and crime. Discriminant validity is said to be present when two factors that are theorized to be separate and unique from one another actually do display these qualities in empirical tests. The question of discriminant validity is vital to the broken windows thesis (BWT) because this framework posits that disorder is an independent variable and crime is a dependent variable. This means that these two phenomena must be related but distinct and separable from one another. If disorder and crime possess discriminant validity, then this portion of the broken windows thesis holds up; if they do not, then the BWT argument is seriously weakened. Disorder can only cause crime if disorder is *not* crime, and crime, likewise, can only be caused by disorder if crime is *not* disorder.

Disorder and Crime Under the Broken Windows Thesis

There is no shortage of practitioners and researchers who claim that the BWT has been firmly empirically supported (e.g., Kelling and Coles 1996; Kelling and Sousa 2001; Skogan 1990). Most of the studies claiming to find support for the thesis, however, have not actually tested the thesis itself but, rather, have assessed the crime-reduction impact of police efforts to crack down on disorder. As will be discussed later, some evidence does suggest that order maintenance policing may be promising under certain conditions; however, it is not at all clear that the mechanism by which order maintenance policing may effectively reduce disorder and crime is actually that specified by the broken windows thesis. Researchers who have studied the thesis itself and the hypotheses derived from it have found little empirical support for it and have offered several reasons why the disorder-crime causal link posited by the BWT may be flawed (Harcourt 2001; Harcourt and Ludwig 2006; Sampson and Raudenbush 1999, 2004; Taylor 2001).

Perhaps the biggest challenge to most of the existing broken windows studies is researchers' tendency to rely on cross-sectional data. Cross-sectional data are those that are gathered at a single time point, such as when a survey asks respondents to report their perceptions about various neighborhood and community characteristics. Cross-sectional research designs make the establishment of causation difficult because there is no time lapse between the measurement of the independent variable and that of the dependent variable, and thus, it is very easy to mistake correlation (i.e., simple, noncausal coincidence) for causation. To the extent that areas plagued by serious crime also suffer high levels of social and physical disorder (Sampson and Raudenbush 2004), cross-sectional research may merely capture the convergence of these two phenomena in space and time and erroneously portray this contemporaneousness as evidence of causality. Alternatively, Taylor's (2001) longitudinal study showed that disorder measured at one time point did not consistently or strongly impact crime measured at a later date. This severely undermined the broken windows thesis and called its core tenet into serious question.

In the search for reasons why disorder may not, in fact, cause crime in the manner advocated by Wilson and Kelling (1982), the issue of discriminant validity stands prominent. As explained in the preceding paragraphs, disorder can only be a predictor of crime if disorder and crime are empirically distinct phenomena. Content overlap between these two factors would violate the principles of scientific inquiry because the independent and dependent variables would essentially be the same thing.

Some proponents of the broken windows thesis might argue that discriminant validity is not a prerequisite for the veracity of the claim that disorder causes crime. Two lines of argument might be advanced in this regard. First, it might be proposed that an absence of discriminant validity would merely signal that the broken windows process has occurred – disorder went unchecked, disorder caused crime, and now crime and disorder are both prevalent in the area. This line of reasoning, however, tautologically begs the question of whether disorder causes crime, as merely pointing to the co-occurrence of the two problems proves neither that disorder temporally and causally precedes crime nor that the two problem types are empirically distinguishable.

Second, it could be contended that disorder and crime are more like a continuum of problems than two separate issues, with crime being on the more serious side of the continuum and disorder on the less serious side. The idea of such a continuum is, in and of itself, quite plausible; however, it contradicts the logic of the BWT. The continuum model would render broken windows unfalsifiable by making it impossible to figure out exactly what the independent and dependent variables are. An accurate empirical test could never be conducted because there would always be doubt as to whether something that was measured as crime should have actually been considered disorder or vice versa. The crucial assumption made by Wilson and Kelling is that disorder causes crime. This premise requires that disorder and crime be two different problems; that is, it necessitates discriminant validity.

It is immediately apparent that this requirement is violated to at least some extent in the BWT, because many of the behaviors the thesis identifies as disorder (e.g., prostitution, tagging, vandalism, minor drug dealing) are crimes by statute or ordinance. There is, therefore, overlap between disorder and crime from a legal or definitional standpoint. This fact provides preliminary evidence against the scientific validity of the broken windows thesis.

The only way that the broken windows thesis can survive this dilemma is if disorder and crime are *perceived* as being unambiguously different in the public's view. Broken windows is grounded in social psychology and hinges on subjective perceptions and on the emotions that people feel and the inferences they draw when they see either the presence or absence of disorder in their neighborhoods of residence. In other words, disorder does not exist in some kind of objective "reality" or "truth" (Piquero 1999; Sampson and Raudenbush 2004), but, rather, disorder takes on whatever meaning is ascribed to it by local residents. If people *see* a difference between disorder and crime, then broken windows might be a viable theory of crime production despite the aforementioned statutory overlap.

One of the most significant demonstrations of the similarity between disorder and crime comes from Sampson and Raudenbush (1999), who relied on systematic social observations of disorder rather than on survey-based perceptual measures. Their results clearly and consistently showed a weak relationship between neighborhood-level disorder and crime once structural covariates were controlled. The only exception to this pattern was with respect to the crime of robbery, which was, interestingly, the only crime type linked to disorder in Skogan's study published in 1990 (Harcourt 2001). Moreover, it appeared that disorder and crime both sprung from the common origin of weak collective efficacy – one of the most consistently robust community-level predictors of such problems (Pratt and Cullen 2005). Far from the causal sequence predicted by the BWT, then, it appeared that disorder and crime were both "symptoms" of the same "disease": the breakdown of a community's ability to regulate its group members and enforce standards of civility and legal conformity.

Testing for discriminant validity between disorder and crime requires the use of a statistical modeling procedure that was created specifically for the purpose of testing proposed factor structures (e.g., Gau 2010). This analytical technique is called confirmatory factor analysis (CFA). Confirmatory factor analysis has allowed tests of discriminant validity within the BWT framework; without this modeling tool, tests of this sort would be much more difficult and the results would likely be harder to interpret. Because CFA is so central to the discriminant validity debate, the next section will briefly outline this analytical strategy and will be followed by summaries of the studies that have employed CFA to test for discriminant validity.

Confirmatory Factor Analysis

Confirmatory factor analysis is a statistical procedure that permits extensive examination of factor structure. A *factor* is something that is not observed directly but accounts for the similarity between a set of objects. For instance, a fire truck, a stop sign, and the pen a professor uses to grade student exams may all be objects representative of the factor "red in color." Certain variables or objects can be indicators of different factors depending on the full item set. If the fire truck were put into a group with a police car and an ambulance, then the factor would be "emergency vehicles." Confirmatory factor analysis permits researchers to propose hypothesized models by saying, "I think *this* variable (e.g., vandalism) is part of *this* factor (e.g., disorder) and I think that *that* variable (e.g., robbery) is part of *that* factor (e.g., crime)" and to then test these predictions.

The quality of a hypothesized model is assessed using fit indices and factor loadings. Fit indices are established criteria that have certain maximum or minimum values – depending on the index – that are indicative of good model fit. Factor loadings are similar to regression coefficients and measure the strength of the relationship between a variable and the factor it is predicted to represent. In the context of tests for discriminant validity in broken windows research, this type of validity would be present when all of the fit indices met their respective thresholds for good fit and the items thought to measure disorder all loaded highly on the disorder factor *and not* on the crime factor, while the crime items loaded well on crime *and not* on disorder. If the fit indices or factor loadings do not meet these requirements, then there is a problem with the model and discriminant validity might be absent.

Confirmatory factor analysis is a popular statistical procedure in psychological research and has gained recognition and use in criminal justice and criminology, as well. The software required to run these analyses is widely available, reasonably priced (as far as software goes), and sports user-friendly, point-and-click interfaces (Gau 2010). There is, in short, no good reason for CFA to not be a staple of broken windows research, which makes it surprising that only a handful of studies have used this method to assess disorder and crime factors for measurement validity (Armstrong and Katz 2010; Gau and Pratt 2008; Piquero 1999; Worrall 2006a). The next section details the studies that have used CFA to test for discriminant validity in perceptual measures of disorder and crime.

Empirical Inquiries into the Discriminant Validity of Disorder and Crime

Worrall (2006a) and Armstrong and Katz (2010) addressed the issue of discriminant validity by entering variables tapping survey respondents' perceptions of disorder, crime, and personal victimization experiences into a series of CFA models. Worrall employed data from 12 cities in 12 different states, and Armstrong and Katz used a sample of survey respondents in a single city in Arizona. The results of both studies painted a murky, inconclusive picture. Survey respondents did not make clear, consistent distinctions between the three phenomena, but neither did they unambiguously see them as one and the same. One-factor models positing disorder as being a single factor along with crime or victimization did not display good fit, yet neither did two-factor models allowing disorder to function as a factor apart from the other two. Overall, then, the authors of both studies concluded that discriminant validity had not been established and that this key premise of the broken windows thesis was not empirically supported.

Gau and Pratt (2008) added to the debate by examining survey data from respondents in neighborhoods in 21 municipalities in Washington State. The variables of interest came from 17 crime and disorder items that asked respondents to indicate whether each issue was *1 = no problem*, *2 = uncertain*, *3 = a problem*, or *4 = a serious problem* in their neighborhoods.

While some of the items were obvious indicators of crime (e.g., gun violence) and others were clearly measures of disorder (e.g., noise), many others (e.g., vandalism, youth gangs) were ambiguous and did not lend themselves clearly to either category. The authors therefore relied on prior literature (Giacopassi and Forde 2000; Reisig and Parks 2004; Sampson and Raudenbush 1999; Taylor 2001) to construct the disorder and crime factors. The disorder factor contained the following items: dogs running at large, drunk drivers on the road, people drinking to excess in public, groups of teens or others hanging out and harassing people, youth gangs, people using illegal drugs, vandalism, noise, traffic problems, and garbage/litter. The crime indicators were people's homes being broken into and things being stolen, people being robbed or having their purses/wallets taken, domestic/intimate partner violence, rape/sex crimes, child abuse, violent crime, and gun violence.

Gau and Pratt (2008) constructed two different models so that each one could be tested against the other to determine which one seemed superior to the other. This strategy is necessary for valid hypothesis testing; it is not a good idea to run only one model because then it is impossible to determine whether the tested model truly is the best one or whether some other model that was not tested might fit better. The first model that Gau and Pratt specified predicted that all of the disorder and crime variables were part of a single factor. This model represented the prediction that people see no difference between disorder and crime and that they instead tend to lump them together as part of the general condition of the neighborhood. Support for this model would undermine the broken windows thesis. The second model predicted that disorder and crime were two separate, distinct factors, each measured by the variables listed above. This model represented the original broken windows framework and predicted that disorder and crime possess discriminant validity. Support for this model would strengthen the broken windows thesis.

Both models displayed similar fit indices that offered lukewarm validation of each of them. Some indices lent the impression that the models were good, while others fell short of their ideal values. On the basis of these alone, then, it was not clear whether the one-factor model or the two-factor model was the better representation of crime and disorder.

There is, however, another way to evaluate model fit above and beyond fit indices: the between-factor correlation. A between-factor correlation of 0.85 or greater indicates that the factors cannot be justifiably separated (see Gomez et al. 2005) because discriminant validity is unacceptably low. It becomes impossible to tell whether items that are purported to be indicators of one factor are not, in fact, equally plausible indicators of the other factor. Gau and Pratt (2008) discovered that the correlation between the crime and disorder factors was 0.92. The two-factor model, therefore, had to be rejected because the correlation between the crime and disorder factors made factor separation indefensible. Perceptions of crime and perceptions of disorder appeared to constitute a single factor; that is, discriminant validity was not present.

It is possible that discriminant validity in individual-level perceptions of crime and disorder is only part of the story; that is, the separability of disorder and crime might be influenced by neighborhood – or community-level conditions. Wilson and Kelling (1982) hinted that there might be a tipping point whereupon a neighborhood becomes so overrun by disorder and crime that police cannot do much other than try to keep pace with calls for service. The tipping point concept may apply to resident perceptions, too. People living in areas where disorder has been present for some time may hardly notice these incivilities because they are simply part of the neighborhood landscape, whereas people residing in areas that have been disorder-free in the past may react very strongly when even the slightest sign of incivility springs up. In other words, the broken windows thesis might be more applicable to some neighborhoods or communities than to others, depending upon extant levels of disorder in those places.

Gau and Pratt (2010) investigated the possibility of spatially patterned discriminant validity by creating a difference score that captured the extent to which people saw disorder levels as being higher, lower, or similar to crime levels in their neighborhoods. They regressed the difference score on several variables, including respondents' perceptions of disorder and crime. The results indicated that the difference between crime and disorder increased as perceived neighborhood-level disorder increased. The authors then split the sample into low-disorder and high-disorder subsamples to test for interaction effects. Among the low-disorder group, increases in disorder caused a reduction in the difference score, while among the high-disorder group, increases in disorder were associated with a closing of the gap. Evidence for a perceptual tipping point was thus uncovered: People living in areas where disorder is mild or absent are more sensitive, and when they see visible signs of incivility, they are likely to interpret these signs as ominous harbingers portending the arrival of serious criminals. People accustomed to seeing disorder in their neighborhoods, on the other hand, were better able to make a distinction between disorder and crime and to more clearly see them as two separate phenomena.

Implications for the Broken Windows Thesis and Order Maintenance Policing

The absence of clear, consistent discriminant validity in the three CFA-based studies described above renders indefensible the claim that disorder is an independent variable and crime is a dependent variable. These two phenomena are no doubt interrelated; however, this relationship is likely correlational rather than causational (see also Sampson and Raudenbush 1999). The possibility that people's ability to discriminate between disorder and crime is dependent upon the amount of disorder present in their neighborhoods of residence (Gau and Pratt 2010) adds another layer of complexity to the matter. These findings have important implications for criminological theory, the measurement of concepts in broken windows research, and crime control policy.

Theoretical and Methodological Implications

The broken windows thesis insists that people see disorder as a visible indicator that the community is spiraling out of control. If, however, people do not differentiate between crime and disorder, then crime itself could serve as the visible indicator of the lack of informal social control in a community. If this is so, then broken windows theory is untenable because it is tautological to claim that crime causes itself.

The only way for the BWT to survive is for there to be a careful refinement of the measures of disorder and crime (Kubrin 2008). Wilson and Kelling (1982) outlined an interesting theoretical framework but they failed to take the quintessential step of defining their terms. It is perhaps not surprising, then, that there has been wide variation in the items used as indicators of disorder and of crime (see Armstrong and Katz 2010; Gau and Pratt 2008, 2010; Giacopassi and Forde 2000; Reisig and Parks 2004; Sampson and Raudenbush 1999; Taylor 2001; Worrall 2006a). It is difficult to test for a relationship between an independent variable and a dependent variable when one is not sure what the independent variable actually is or how it should be measured.

It also might simply be time to retire the BWT from active duty on the criminological stage. There is at this point no solid reason to believe – and plenty of reason *not* to believe – that disorder causes crime in the manner stated by Wilson and Kelling (1982). In addition to the apparent absence of discriminant validity between the BWT's core theoretical constructs, there are identifiable conceptual problems with the thesis. The broken windows idea was based on the notion that if a rock is thrown through a window and the damage done to the window is not repaired immediately, then pretty soon all the windows will be shattered. This seems logical on its surface, but there is a troubling question embedded in this scenario: Who threw the rock? If people are prowling about a neighborhood and maliciously throwing rocks at things, then it seems that a problem is already in place and that the newly smashed window is incidental to deeper, more pressing issues. Misplacing attention on the window diverts effort and resources away from identifying and addressing the genuine, underlying problems in a community that allow both disorder *and* crime to take root and flourish (Sampson and Raudenbush 1999).

Policy Implications

As stated at the outset of this entry, the broken windows thesis is a policy-based proposal (hence the limited attention paid to theoretical and measurement issues). Wilson and Kelling (1982, p. 36) made police the centerpiece of disorder and crime reduction by arguing that "Though citizens can do a great deal, the police are plainly key to order-maintenance." Broken windows theory has had a profound effect on the field of policing and has revolutionized some departments' missions and functions.

There is evidence that broken windows policing (often also called order maintenance policing, quality-of-life policing, proactive policing, and zero tolerance policing, though some broken windows advocates take issue with this last label due to its negative connotation) can reduce some types of crime, some of the time, in some jurisdictions. Sampson and Cohen (1988) discovered a relationship between order maintenance policing and reductions in robbery rates, though they were not entirely sure that this effect was attributable to deterrence or to broken windows with respect to the underlying theoretical mechanism. Braga et al. (1999) found that a disorder crackdown seemed to reduce crime, yet paradoxically did not reduce disorder or nuisance offending. Worrall's (2006b) analysis indicated that arrest rates for public order offenses had an immediate suppressive impact on assault, a delayed impact on burglary, and no effect on larceny. Kubrin et al. (2010) discovered a relationship between proactive policing and robbery rates. On the other hand, Novak et al. (1999) studied a disorder crackdown and found no evidence that the crackdown reduced any type of serious crime.

An important aspect of existing evaluations of disorder-based policing is the absence of attention paid to the mediating variables that were central to Wilson and Kelling's (1982) original specification of the BWT. True order maintenance policing that reduced crime in a way consistent with Wilson and Kelling's predictions would do so by enhancing social cohesion among area residents and stimulating social interaction and the frequenting of public spaces by law-abiding persons whose collective informal surveillance provided guardianship for all those present. Absent evidence that cohesion and informal social control emerged as a result of a concerted disorder-reduction effort by police, then order maintenance policing is empirically indistinguishable from crackdowns and hot-spots policing (e.g., see Sherman and Weisburd (1995), where a randomized crackdown-backoff tactic in crime hot spots reduced crime and disorder simply through sporadic, unpredictable increases in police presence).

Another alternative theoretical mechanism by which order maintenance policing might reduce crime in a manner inconsistent with the BWT was hinted at by Kelling and Coles (1996). They noted that nearly one quarter of the people arrested by police for disorderly behaviors were either carrying illegal weapons or had outstanding arrest warrants in connection to serious crimes. Order maintenance policing, then, might simply be a form of net widening that legitimizes police interference in people's freedom of movement for myriad reasons and thus permits them greater power to investigate people they feel as somehow suspicious or worthy of scrutiny. Net widening might accomplish the goals laid out by the BWT, but not via the same means, and therefore, it is not proper to call this "broken windows policing," strictly speaking.

The evidence to date that disorder and crime lack discriminant validity results also bears on the issue of citizen satisfaction with the police. If citizens do not see disorder as a problem apart from crime, then they are unlikely to understand the logic behind order maintenance policing. Victims of crimes like robbery and residential burglary may be unlikely to fully endorse police crackdowns on jaywalking and littering. Indeed, though citizens rate order maintenance as an important endeavor (Skogan and Hartnett 1997), they still see traditional law enforcement against serious crimes as the primary role of the police (Skogan 1990; Webb and Katz 1997).

A tunnel-vision focus on disorder could discredit police in the eyes of community residents.

Finally, the possibility that the discriminant validity of disorder and crime is dependent upon neighborhood context (specifically, the amount of disorder already present in the area) implies that order maintenance policing may be appropriate for better-off areas and inappropriate for more disadvantaged neighborhoods. Bucolic neighborhoods luxuriously free of dire problems might be an optimal place for police to adopt a quality-of-life approach in order to maintain the status quo and prevent disorder from seeping in. These are the places, too, where residents are most likely to hold positive attitudes toward police and to be willing to help police keep the social order in check.

By contrast, areas wherein disorder has already left its mark may, perhaps ironically, be the worst candidates for order maintenance policing simply because of the volume and severity of incivilities present in the area. Order maintenance policing places vast discretion in the hands of police and charges them with a mandate to intervene in the lives of people who do not pose an immediate or obvious threat to public safety. Exacerbating the danger inherent in the combination of discretion and low visibility is that order maintenance has been interpreted by many police departments as being best accomplished via the widespread use of *Terry*-type stop and frisks for suspicious behavior.

The convergence of these conditions – discretion, a disorder mandate, and reliance on *Terry* stops – has implications for police treatment of citizens in disadvantaged areas. Neighborhood-level disadvantage may encourage police misconduct (Kane 2002) and the use of more severe forms of coercive force (Terrill and Reisig 2003) by giving police the impression that the entire area has descended into lawlessness (Klinger 1997) and by offering the promise of impunity for bad deeds, given the social and political powerlessness of the people who reside there. In some disadvantaged, high-crime areas, the use of *Terry* stops has been racially discriminatory (Fagan and Davies 2000). Already-rocky relationships between the police and community can be further strained when officers stop people for what the people themselves see as "nothing" and, especially, when officers are rude and disrespectful during these encounters. This can worsen existing tensions and lead to even more crime and disorder as police systematically undercut their own legitimacy and authority (Gau and Brunson 2010).

Conclusion

When Wilson and Kelling wrote their *Atlantic Monthly* magazine article in 1982, they wanted the police to do something that the police could not do at that time: focus their efforts on the eradication of disorder. The professional model of policing and several US Supreme Court rulings had limited police duties to reactive response to calls for service and to the identification and arrest of people suspected of having committed serious crimes. In order to legitimize a broader conceptualization of the police role and to credibly argue that police can and should intervene in disorderly activities, Wilson and Kelling posited a restructuring of the disorder-crime relationship such that these two problems were linked sequentially and causally with disorder being a precursor to crime. The broken windows thesis thus casts disorder as an independent variable and crime as a dependent variable.

In order for the BWT's formulation of disorder and crime to be accurate, though, disorder and crime must be empirically separate phenomena; that is, they must possess discriminant validity. Preliminary evidence that discriminant validity is not present arises from the fact that many activities technically classified as disorder are, in fact, criminal by statute or ordinance. The broken windows thesis, however, is grounded in subjective, individual perceptions of disorder, so the thesis could survive the legal criticism if people see a difference irrespective of legal overlap.

Studies employing confirmatory factor analysis – a statistical modeling procedure that permits tests of factor structure and examinations of discriminant validity – have found either mixed support or no support for the central

BWT proposition that people discriminate between disorder and crime. Gau and Pratt (2008) discovered that the disorder and crime factors were too highly correlated to justify separation and that disorder and crime were, therefore, both part of a single factor. Worrall (2006a); Armstrong and Katz (2010) reported that one – and two-factor models alike fell short of the criteria necessary to establish good fit. Discriminant validity was either weak or was absent entirely across the different sets of models the authors ran. Finally, Gau and Pratt (2010) constructed a score measuring the difference between people's perceptions of disorder and those of crime and found that neighborhood disorder was a significant predictor of that difference score such that people living in low-disorder areas had trouble distinguishing between the two problem types, while those in high-disorder areas did seem to draw a line between them.

The apparent absence of discriminant validity between disorder and crime, and the evidence suggesting that perceptual discrimination may vary by neighborhood type, has several implications for the broken windows thesis and the policing strategy arising therefrom. The thesis itself needs to be refined and its measures more thoroughly defined. More radically but nonetheless on the table as an option, the thesis needs to be abandoned in favor of alternative explanations for the covariance between disorder and crime that are better conceptualized and, therefore, hold more promise for theoretical validity and useful policy. Order maintenance policing might, moreover, hold more promise in areas wherein disorder is not currently present and can be kept out of the neighborhood in a preventive fashion. Once disorder has taken hold, however, a new strategy might be required.

In sum, the broken windows thesis has focused national attention on disorder and crime. It is clear from a long line of research that both matter greatly to people's fearfulness and their satisfaction with their neighborhoods of residence; however, it is time to question the causal, disorder-to-crime sequence posited by Wilson and Kelling's (1982) broken windows thesis. It simply might not be true. Researchers and practitioners should consider alternative theoretical models that hold promise for reducing disorder and crime and improving people's quality of life.

Related Entries

- ▶ Broken Windows Thesis
- ▶ Defining Disorder
- ▶ Disorder: Observational and Perceptual Measures
- ▶ Social Disorder and Physical Disorder at Places

Recommended Reading and References

Armstrong T, Katz C (2010) Further evidence on the discriminant validity of perceptual incivilities measures. Justice Q 27:280–304

Braga AA, Weisburd DL, Waring EL, Mazerolle LG, Spelman W, Gajewski F (1999) Problem-oriented policing in violent crime places: a randomized controlled experiment. Criminology 37:541–580

Fagan J, Davies G (2000) Street cops and broken windows: *Terry*, race, and disorder in New York City. Fordham Urban Law J 28:457–504

Garofalo J, Laub J (1979) The fear of crime: broadening our perspective. Victimology 3:242–253

Gau JM (2010) Basic principles and practices of structural equation modeling in criminal justice and criminology research. J Crim Justice Educ 21:136–151

Gau JM, Brunson RK (2010) Procedural justice and order maintenance policing: a study of inner-city young men's perceptions of police legitimacy. Justice Q 27:255–279

Gau JM, Pratt TC (2008) Broken windows or window dressing? Citizens' (in)ability to tell the difference between disorder and crime. Criminol Public Policy 7:163–194

Gau JM, Pratt TC (2010) Revisiting broken windows theory: examining the sources of the discriminant validity of perceived disorder and crime. J Crim Justice 38:758–766

Giacopassi D, Forde DR (2000) Broken windows, crumpled fenders, and crime. J Crim Justice 28:397–405

Gomez R, Burns LG, Walsh JA, Hafetz N (2005) A multitrait-multisource confirmatory factor analytic approach to the construct validity of ADHD and ODD rating scales with Malaysian children. J Abnorm Child Psychol 33:241–254

Harcourt BE (2001) Illusion of order: the false promise of broken windows policing. Harvard University Press, Cambridge

Harcourt BE, Ludwig J (2006) Broken windows: new evidence from New York City and a five-city social experiment. Univ Chicago Law Rev 73:271–320
Kane RJ (2002) The social ecology of police misconduct. Criminology 40:867–896
Kelling GL, Coles CM (1996) Fixing broken windows. Simon & Schuster, New York
Kelling GL, Sousa WH Jr (2001) Do police matter? An analysis of the impact of New York City's police reforms. The Manhattan Institute, New York
Klinger DA (1997) Negotiating order in patrol work: an ecological theory of police response to deviance. Criminology 35:277–306
Kubrin CE (2008) Making order of disorder: a call for conceptual clarity. Criminol Public Policy 7: 203–214
Kubrin CE, Messner SF, Deane G, McGeever K, Stucky TD (2010) Proactive policing and robbery rates across U.S. cities. Criminology 48:57–97
Novak KJ, Hartman JL, Holsinger AM, Turner MG (1999) The effects of aggressive policing of disorder on serious crime. Polic An Int J Police Strateg Manag 22:171–189
Piquero A (1999) The validity of incivility measures in public housing. Justice Q 16:793–818
Pratt TC, Cullen FT (2005) Assessing macro-level predictors and theories of crime: a meta-analysis. In: Tonry M (ed) Crime and justice, vol 32. University of Chicago Press, Chicago, pp 373–450
Reisig MD, Parks RB (2004) Can community policing help the truly disadvantaged? Crime Delinq 50:139–167
Sampson RJ, Cohen J (1988) Deterrent effects of the police on crime: a replication and theoretical extension. Law Soc Rev 22:163–189
Sampson RJ, Raudenbush SW (1999) Systematic social observation of public spaces: a new look at disorder in urban neighborhoods. Am J Sociol 105:603–651
Sampson RJ, Raudenbush SW (2004) Seeing disorder: neighborhood stigma and the social construction of broken windows. Soc Psychol Q 67:319–342
Sherman LW, Weisburd D (1995) General deterrent effects of police patrol in crime "hot spots": a randomized, controlled trial. Justice Q 12:625–648
Skogan WG (1990) Disorder and decline: crime and the spiral of decay in American neighborhoods. University of California Press, Berkeley
Skogan WG, Hartnett SM (1997) Community policing, Chicago style. Oxford University Press, New York
Taylor RB (2001) Breaking away from broken windows. Westview Press, Boulder
Terrill W, Reisig MD (2003) Neighborhood context and police use of force. J Res Crime Delinq 40:291–321
Webb VJ, Katz CM (1997) Citizen ratings of the importance of community policing activities. Polic An Int J Police Strateg Manag 20:7–23
Wilson JQ, Kelling GL (1982) Broken windows: the police and neighborhood safety. The Atlantic Monthly, March 29-38
Worrall JL (2006a) The discriminant validity of perceptual incivility measures. Justice Q 23:360–383
Worrall JL (2006b) Does targeting minor offenses reduce serious crime? A provisional, affirmative answer based on an analysis of county-level data. Police Q 9:47–72
Xu Y, Fiedler ML, Flaming KH (2005) Discovering the impact of community policing: the broken windows thesis, collective efficacy, and citizens' judgment. J Res Crime Delinq 42:147–186

Disorder: Observational and Perceptual Measures

Donald L. Anthony and Douglas D. Perkins
Program in Community Research and Action,
Peabody College, Vanderbilt University,
Nashville, TN, USA

Overview

This entry provides a synopsis of disorder-related research methods, measures, and analytic frameworks. The entry is intended to give readers a better understanding of trends in disorder-related research and the many facets of disorder analysis; aspects covered in the literature range from the various ways in which individuals and communities conceptualize disorder to the roles of homeownership, organizational participation, and land use in predicting crime in neighborhoods. Methods employed in extensive studies in Chicago, Baltimore, and Salt Lake City are cited here for illustrative purposes. The entry concludes with a summary of major findings from those studies as well as suggestions for future disorder-related research. Literature on disorder frequently refers to social or physical *incivilities* (Taylor and Hale 1986), and this entry will use the terms *disorder* and *incivilities* interchangeably.

The Evolution of Disorder Research

Conceptualizations of neighborhood-level disorder and its association with crime have evolved over the last half century; likewise, methodologies employed in disorder-related research have evolved as well. Early research into disorder and crime was heavily influenced by Chicago-school sociologists such as Gerald D. Suttles, whose 1968 work *The Social Order of the Slums* examined the social and economic conditions of a poor Chicago neighborhood. Suttles employed a mixed-methods approach, analyzing quantitative demographic data and presenting ethnographic accounts of everyday life in the Chicago slums.

With Wilson and Kelling's advancement of incivilities theory in a 1982 issue of *The Atlantic*, interest in disorder-related research was renewed. As incivilities theory became a mainstay of the social sciences, disorder-related studies sought to test and refine incivilities theory largely by employing sharpened research methods and analytical tools. Methodological advances have allowed researchers and audiences to better understand connections between disorder and crime. Over the last half century, research on disorder and crime has employed diverse data collection methods and analytic frameworks. While gathering residents' perceptions and conducting passive surveys are still important in disorder-related research, recent studies tend to rely more on objective measures and active means of data gathering.

Analytic methods have been greatly improved by newer technologies. Geographic information systems (GIS) and hierarchical linear modeling (HLM) software are two examples of such technological advances; GIS gives researchers the ability to analyze spatial patterns and produce informative and interactive graphic representations of their findings, while HLM allows researchers to explore multiple levels of data and effects in an efficient manner. The following sections summarize trends in disorder-related research. These trends include clarification of the units of analysis, evolution of measurement tools, and advances in methods for analyzing data. Extensive studies undertaken by researchers in Chicago, Baltimore, and Salt Lake City serve as primary settings; other exemplary studies are cited when appropriate.

Units of Analysis

Neighborhoods have been conceptualized in many ways ranging from strictly geographic definitions to delineations that account for social and cultural contexts. Disorder-related studies tend toward the former, though questions about the appropriate scope for geographic boundaries persist. Early disorder-related studies typically grouped neighborhoods based on Census Bureau-defined boundaries. The use of such delineations has been problematic for researchers because census-defined neighborhoods – particularly census tracts – often include artificial or imposed boundaries. Overly broad census-defined neighborhoods may cross major roads, railroad tracks, and water bodies. Neighborhoods delineated in such a manner often extend far beyond finer, resident-recognized neighborhood boundaries such as street blocks.

Recognizing the need to define neighborhoods in ways that match residents' own definitions and are amenable to research, recent studies have focused on census-defined block groups (a collection of contiguous census blocks) and an even more precise unit of analysis, individual street block faces. When the basic unit of analysis is an individual street block face, a resident's neighbors are those on either side of one's street bounded by cross streets. (By "block" we generally mean street block as opposed to polygonal census block, which does not include neighbors across the street.) By examining lower geographic levels such as blocks and block groups, researchers have been able to better account for demographic characteristics that tend to be consistent within compact areas but vary – sometimes wildly – within larger areas such as census tracts. The studies highlighted in this entry utilized blocks as the basic units for their analyses. In the Baltimore study, for example, the authors focused their attention on 50 randomly selected street blocks (Perkins et al. 1992; Perkins and Taylor 1996). In the Salt Lake City study, researchers focused on 58 blocks in a transitional suburban setting (Brown et al. 2004b). Sampson and Raudenbush (1999, 2004) used the block group as the basic neighborhood delineator in a series of studies on disorder and crime in Chicago; however, like the studies in Baltimore and Salt Lake City, the Chicago-based study relied on individual street blocks for comparative purposes.

Using the street block as the primary unit of analysis gives researchers the ability to compare data at both the block level and the individual or household level. Such a multilevel approach allows researchers to determine the extent to which disorder as well as demographic and other potential mediating factors, such as collective efficacy, operate at the individual or block level. This is critically important because if virtually all the variance is at the individual level, it suggests that the impact of disorder and its mediators are mainly personal and subjective. If substantial block-level effects exist on top of any individual variation, it essentially shows that those effects are more than simply a matter of individual perception or psychology. They are "real" and "objective" and have an even greater impact insofar as being shared by one's neighbors. To obtain block-level data, researchers often aggregate individual or household-level data. Although less common, it is helpful to also identify and obtain measures that are not just aggregated but commensurate to the level being studied (Shinn 1990).

Data Gathering and Measurement Tools

Just as conceptualizations of neighborhood delineation have evolved, data gathering and measurement tools have changed as well. Disorder-related research has become increasingly focused on employing multiple methods of data gathering and measurement, often relying on advanced quantitative methods not available to previous generations of researchers. Technological and theoretical advances have offered researchers new ways of collecting and examining data. Video recordings, for example, allow researchers to verify their observations of street life and other neighborhood conditions. Thus, the *objectivity* of data-gathering methods has greatly improved.

Aggregation is not the only way to obtain more objective measures. Recent research has also compared residents' perceptions and independent measures of disorder and crime. While residents' perceptions continue to be important indicators, such subjective data do not form the sole basis on which disorder is measured. Disorder can be more objectively measured with trained independent observers rating settings using predefined scales or checklists. Crime can be measured by analyzing police reports, although such reports have often been charged with political bias and there is often an information gap since crimes are not always reported. Thus, even with the emphasis on objective measures, researchers tend to see a continuing need for resident surveys and interviews.

As disorder-related research evolved, researchers sought to increase objectivity by employing independent observers to conduct "windshield surveys" (Taylor et al. 1985). These surveys were typically conducted by independent raters who drove through neighborhoods and either made open comments or graded neighborhood characteristics on a number of predetermined criteria. While windshield surveys can be useful in a cursory sense, they may lead to inadvertent oversights on the part of the raters. Walking observations may be more useful as they allow the rater to take a slower pace and make closer observations, often while having impromptu interactions with residents and others in the neighborhood; such interactions can provide valuable insights not typically gleaned through passive windshield surveys.

Researchers have created various measures and scales for conducting independent visual surveys. In their study of physical decay in low, moderate, and high-income Baltimore neighborhoods, Taylor et al. (1985) conducted assessments of the physical conditions of 808 street blocks. Independent raters evaluated a wide-ranging set of neighborhood factors including building setbacks, traffic volume, types of land uses, graffiti, and building vacancies. The researchers created three neighborhood-level subscales using aggregated data collected in the block assessments. A street accessibility subscale aggregated measures such as street width, traffic volume, and street lighting. A land use subscale aggregated measures related to type of land use, that is, residential or nonresidential. A final subscale, density, included measures such as block length and building height. By aggregating data into three subscales, the authors were able to draw comparisons across blocks.

Other observational tools have been used to measure both physical and social disorder. One such tool is the Block Environmental Inventory (BEI). Researchers studying Baltimore neighborhoods in 1987 employed the BEI as a means of examining the social and physical conditions of 50 residential blocks, most of which were assessed by two trained raters to allow for inter-rater reliability tests (Perkins et al. 1992). The BEI included observations at both the house and block levels. At the house level, the BEI consisted of three main categories: incivilities, defensible space, and territorial markers. The incivilities subscale included items such as litter, vandalism, and dilapidation. The defensible space subscale included visibility, barriers, lighting, and security bars. The territoriality subscale consisted of signs warning of dogs, sitting places, plantings, and home decorations. Block-level items included people outside (by gender, activity, and general age group), open lot frontage (e.g., playgrounds, public gardens, institutional yards), abandoned cars, street trees, and vacant lots. The raters themselves were also part of the BEI; they were to observe whether people noticed them as they assessed the neighborhood (Perkins et al. 1992).

Another means of objective data gathering is through the use of Systematic Social Observation (SSO). Sampson and Raudenbush (1999, 2004) employed this method in a study of Chicago neighborhoods. The SSO method included the videotaping of social and physical disorder in 23,816 block faces across 80 neighborhoods. The authors asserted that SSO was an important methodological advance because it could be replicated. Of the 23,816 block faces that were videotaped by independent raters, 15,141 were coded and subsequently utilized in the study. Notably, because of the extensive coding required, SSO may be cost and/or time prohibitive for researchers on unfunded or underfunded projects.

Resident surveys continue to be popular in disorder-related research. Sampson and colleagues (1989, 1999, 2004) employed resident surveys in both Chicago and the United Kingdom to determine residents' perceptions of disorder and crime. Likewise, Perkins and colleagues (1990, 1993) in Brooklyn and Queens, New York; Perkins and Taylor (1992, 1996) in Baltimore; and Brown, Perkins, and Brown (2004a, b) in Salt Lake City employed resident surveys primarily as a means of comparing resident perceptions to independent observations. Surveys, particularly face-to-face interviews, often contextualize objective data, giving researchers unique insights into the subject neighborhoods.

Researchers have employed several analytical tools for determining actual occurrences of crime. Sampson and Raudenbush (2004) relied on publicly available data to help them objectively measure both crime and demographics. Particularly, they relied on Census Bureau data and police records of violent crimes. Likewise, Taylor et al. (1985) relied on police reports from the Baltimore Police Department to measure serious crime in their subject neighborhoods. In addition to reported crime and surveyed victimization, Perkins and Taylor (1996) also analyzed newspaper articles related to crimes in or near neighborhoods included in their study as a measure of media influence on localized fear.

Obtaining reliable measures of crime often proves more difficult than accurately measuring perceptions or media accounts. Police reports are perhaps the most common sources of crime data employed by researchers. However, because crime records may be under- or overcounted for budgetary or political ends and because reporting to police varies by type (e.g., higher-value property crimes and homicides more likely to be reported than lesser property and personal crimes), neighborhood, and city, resident surveys and interviews often provide very different information and insights.

Means of Analysis

Collecting data at both the individual and block levels allows researchers to employ complex statistical methodologies to find associations between disorder and crime. In some cases, however, simpler analytic techniques may yield useful data about the neighborhoods being studied. Using the Chicago, Baltimore, and Salt Lake City studies, this section provides a brief overview of the various statistical techniques employed in disorder-related research.

Simple correlations may reveal important relationships among variables. Taylor et al. (1985), for example, found that actual occurrences of crime measured by analyzing Baltimore police reports were positively correlated with physical decay, nonresidential land uses, and vacancy rates. Perkins et al. (1992) utilized correlations to examine relationships among actual disorder, perceived disorder, and perceived crime in Baltimore neighborhoods. After controlling for race, education level, tenure, and block size, four statistically significant correlations emerged at the block level: (1) the number of street lights was negatively correlated to perceived robberies, (2) the presence of yard decorations was negatively correlated to perceived drug dealing and fighting in the street, (3) yard decorations were negatively correlated to perceived burglaries, and (4) the presence of block watch signs was negatively correlated to perceived burglaries. Clearly, in both Baltimore studies, the use of simple correlations yielded helpful information: certain environmental conditions and features were associated with neighborhood crime reports and perceptions.

Multiple regression allows researchers to determine the degree to which a combination of independent variables are predictive of a dependent variable. Perkins et al. (1992) employed multiple regression to determine whether territorial functioning and defensible space features could supplement observed disorder in predicting perceived disorder and crime. The authors found that adding territorial markers and defensible space features to their model explained more of the variance in perceived disorder and crime.

Hierarchical linear modeling (HLM) and structural equation modeling (SEM) are examples of advanced statistical techniques employed by researchers. With HLM, the effects of predictors at multiple levels and/or over multiple points in time on outcomes can be determined. For example, Perkins and Taylor (1996) used HLM to compare the power of three different methods of measuring aggregated community disorder (surveyed resident perceptions, on-site observations by trained raters, and content analysis of crime- and disorder-related newspaper articles) to predict subsequent fear of crime, controlling for individual-level perceptions of social and physical disorder and independent ratings of physical disorder on respondents' properties. Sampson and Raudenbush (2004) used HLM to examine the relationship between independently observed social and physical disorder and perceived disorder both within and between neighborhoods.

SEM has been employed in disorder-related studies as a means of exploring complex models including latent variables. For example, Sampson and Groves (1989) used SEM in finding that socioeconomic status, ethnic heterogeneity, residential mobility, family disruption, and urbanization predicted crime and delinquency when mediated by local friendship networks, unsupervised teenage peer groups, and low organizational participation.

Other Common Disorder Research Methods

Other methods have proven useful in disorder-related research; among these are qualitative analyses and longitudinal studies. Qualitative data can give researchers insights that may be impossible to glean, or less richly developed, through surveys, systematic environmental assessments, crime reports, or other quantitative sources. In-depth open-ended or semi-structured interviews, ethnographic accounts, and oral histories are examples of such methods. Earlier disorder-related research conducted by Suttles and other Chicago-school sociologists relied heavily on ethnography; many researchers now utilize qualitative research as one component in a toolbox of data collection methods. Longitudinal studies allow the researcher to discern trends over a period of time. By following up snapshot data with data collected at a later time or at subsequent intervals, the researcher may gain a better understanding of phenomena that are not static in nature. Because neighborhoods are ever-changing, examining the neighborhood over a period of time can help the researcher contextualize and better understand phenomena and findings.

Several recent disorder-related studies have used many of the research and analytic methods discussed here. The following section highlights findings from those studies, focusing on the methodologies employed in both the data-gathering and data analysis stages. These studies are cited for illustrative purposes; references to additional studies of interest are provided following the entry's conclusion.

Major Findings Using Mixed Research Methods

This section provides a brief overview of recent major findings in the disorder-related literature. As a comprehensive review of the literature is beyond the scope of this entry, this section focuses on findings based on the above methods and means of analysis. Findings are categorized based on the following themes: objective versus subjective measures of disorder and crime, advanced statistical analyses, and challenges to accepted theories.

Objective and Subjective Measures

Early disorder-related research tended to focus on residents' perceptions of neighborhood disorder, occasionally comparing those perceptions to observations by trained raters. Recent research has in some ways attempted to standardize disorder-related research by balancing residents' perceptions with objective measures. Data in recent studies include results of independent neighborhood assessments, analyses of media accounts of crime, and analyses of police reports.

Taylor et al. (1985) examined the relationship between observed and perceived physical disorder in their study of Baltimore neighborhoods. Residents and independent raters evaluated four specific indicators of physical disorder: housing vacancies, empty lots, litter, and unkempt properties. After combining the indicators to form a single scale, the authors found a high correlation between resident perceptions and independent observations of physical disorder.

The Block Environmental Inventory (BEI) was used by Perkins et al. (1992) as a means of objectively measuring disorder, territorial markers, and defensible space features in Baltimore neighborhoods. Researchers obtained data on residents' perceptions by surveying 412 residents across 50 street blocks either by telephone or in person. Interviewers asked how big a problem (on a three-point scale) each of a list of 12 specific items related to physical disorder, social disorder, and neighborhood crime are on respondents' blocks. Perkins et al. used both correlation and regression analyses to determine whether the objective measures of disorder gleaned through the BEI were associated with subjective measures. The authors correlated objective and subjective measures of five specific forms of physical disorder: litter, vandalism, dilapidated exterior, vacant housing, and trashed lots. Correlations for all five forms of disorder were statistically significant. Litter was the strongest correlation, while dilapidated exterior was the weakest. After controlling for race, education, homeownership, and block size, all of the correlations were weakened but still statistically significant. The authors concluded that residents' perceptions and independent observations of physical disorder are strongly related but that the relationship is affected in part by block demographics.

In a 2004 study of Salt Lake City neighborhoods, Brown et al. considered associations among several predictors of crime: home attachment, observed physical disorder, perceived physical disorder, social ties and collective efficacy, and home ownership and demographics. Besides the demographic variables, all other predictors were scales or composites of multiple measures. Using multilevel modeling (HLM), the authors found that observed disorder and perceived disorder were not related at the individual level (i.e., people tend not to perceive their own properties as disordered). However, as expected, observed and perceived disorder were significantly related at the block level (Brown et al. 2004b).

Sampson and Raudenbush (1999, 2004) found little difference between observed and perceived disorder. For their study, social disorder included adults congregating, public intoxication, selling drugs, and prostitution. Physical disorder ranged from minor offenses (e.g., cigarettes on the street) to more prominent offenses (e.g., gang graffiti on buildings). The researchers employed SSO by having independent raters videotape block faces; the tapes were then coded for both social and physical disorder. The objective data were compared to survey data gathered from 3,864 residents. The authors found that residents' perceptions of both physical and social disorder were significantly associated with independent measures of disorder.

Specific findings regarding the relationship between objective and subjective measures of disorder have varied by study. However, all of the studies cited in this section have found at least some degree of correlation between independent observations and resident perceptions of disorder. Variation among the findings appears to be attributable at least partially to the particular research setting, research questions, analytic methods, and the variables or characteristics being studied.

Advanced Statistical Analyses

Quantitative modeling techniques such as multiple regression, hierarchical linear modeling (HLM), and structural equation modeling (SEM) have been utilized to analyze disorder and crime data. Each method brings its own set of strengths to data analysis. Multiple regression allows researchers to determine the predictive values of independent variables on a particular outcome. Hierarchical linear modeling allows researchers to compare data on multiple levels; change in individuals or neighborhoods can be traced over time, individuals can be compared with their neighbors, and blocks or neighborhoods can be compared to each other. Structural equation modeling allows researchers to determine the effects of latent variables. Both HLM and SEM are relatively new tools. The following subsections explore findings attributable to the use of these tools; specifically, the subsections address findings related to personal characteristics, place characteristics, home and income characteristics, and other important factors.

Personal Characteristics

Quantitative analyses reveal that certain personal characteristics are associated with individuals' perceptions of disorder and crime. Studies from Chicago, Baltimore, and Salt Lake City offer diverse and wide-ranging findings on the associations among crime, fear, disorder, and personal characteristics. With regard to sex, females are likely to perceive more disorder than males (Sampson and Raudenbush 2004). Likewise, females are likely to express more fear of crime than males at both the within-block and between-blocks levels (Perkins and Taylor 1996).

Age plays a role in perception and fear, though the findings have been somewhat inconsistent. At the block level, age is a significant predictor of fear of crime. In one study, however, age was not a significant predictor of fear at the individual level (Perkins and Taylor 1996). In another study, within a given block, older individuals perceived less disorder (Sampson and Raudenbush 2004), an unexpected finding that may be due to older residents being less aware of specific changing disorder cues.

Race and ethnicity have significant effects at multiple levels. African Americans tend to perceive less disorder than do other residents of their block; however, controlling for social context, blocks with more African Americans tend to have more disorder as perceived by residents of the block. Latinos are also positively associated with perceived disorder at the between-blocks level and are likely to perceive more disorder at the within-block level (Sampson and Raudenbush 2004). Race has been shown to be a significant predictor of fear of crime as well (Taylor et al. 1985; Perkins and Taylor 1996).

Marital status plays some role in perception of disorder. Separated and divorced individuals are likely to perceive more disorder at the within-block level than their married and single counterparts. Between blocks, marital status is not significantly associated with perceptions of disorder (Sampson and Raudenbush 2004). Clearly, sex, age, race and ethnicity, and marital status are all important personal characteristics that shape residents' perceptions of crime and/or disorder, although their effects tend to vary by location and model.

Place Characteristics

Taylor et al. (1985) found two important correlations between crime and place-based characteristics. First, the authors found a strong positive correlation between crime and vacant housing, indicating that the higher the percentage of vacant housing units in a neighborhood, the higher the amount of crime in that neighborhood. A second statistically significant positive correlation was found between crime and nonresidential land uses. Both vacancy and land use were significant predictors of fear. Perkins and Taylor (1996) found that the physical conditions of nonresidential lots were better predictors of fear of crime than the conditions of residential areas. Land use appears to be a consistent indicator of fear as shown by two separate Baltimore-based studies.

Home and Income Characteristics

Researchers have examined the associations between crime and home-based characteristics such as tenure and home attachment, which are both inversely related to observed disorder (Brown et al. 2004b). At the within-block level, homeownership is negatively associated with police reports of crime. At the between-blocks level, home attachment is negatively associated with police reports of crime (Brown et al. 2004b).

Socioeconomic status (SES) has been shown to be a significant predictor of perceived disorder, fear of crime, and neighborhood confidence. Fear and confidence are particularly affected by perception of disorder in moderate-income neighborhoods (Taylor et al. 1985). In a United Kingdom-based study, Sampson and Groves (1989) considered the association between SES and crime. The authors found that SES was a significant predictor of crime. However, a closer look revealed that the presence of unsupervised peer groups mediated the effects of SES on crime. In fact, unsupervised peer groups mediated 80 % of the effect of SES on mugging and street robbery, 34 % of the effect of SES on stranger violence, and 68 % of the effect of SES on total victimization. The presence of unsupervised peer groups had a worsening effect on both household and property victimization, while organizational participation had an inverse effect on both forms of victimization. These findings led Sampson and Groves to conclude that the presence of unsupervised peer groups and the absence of organizational participation were more important to remediate than low SES.

Other Outcomes

Brown, Perkins, and Brown (2004a) employed a longitudinal approach to examine the spillover effects of cleaning up a brownfield in Salt Lake City and replacing it with new housing. The researchers were interested in incumbent upgrading – the idea that residents will invest in their own properties if their neighbors do likewise. Independent disorder assessments and resident surveys were taken from 1993 to 1996 (T1) and again from 1998 to 1999 (T2). Police reports of crime were measured from 1995 to 1996 (during T1) and again from 1999 to 2000 (T3). The researchers examined physical disorder using objective raters; home attachment, homeownership, perceived disorder, and crime were measured using resident surveys.

Using correlation analysis, the researchers found that from T1 to T2, observed disorder in the areas surrounding the former brownfield decreased, especially in the blocks closest to the brownfield. HLM analysis indicated that crime reports decreased at T2 in the neighborhoods closest to the brownfield, while crime reports tended to increase in the neighborhoods farther away. The authors found that the presence of independently observed physical disorder on specific properties increased the likelihood of crime on those properties; further, physical disorder on a particular block increased the likelihood of crime on properties within that block. Conversely, the presence of new houses built on the former brownfield was associated with decreased crime and disorder on nearby blocks. The blocks in which disorder increased between T1 and T2 were more vulnerable to crime at T3. The authors argued that litter and unkempt lawns were two types of physical disorder that could lead to residents investing less in their homes.

In summary, recent research has not revealed any single, dominant predictor of crime. Rather, the research has created a patchwork of findings which – when taken as a whole – provide a clearer picture of how individuals perceive and experience disorder, fear, and crime in their local settings. While some personal characteristics appear to influence both degree of fear and perceptions of crime and disorder, territorial markers, defensible space features, homeownership, home and place attachment, organizational participation, and lack of activities for young people appear to work in some combination to explain the occurrence of crime within neighborhoods. Simple answers have not emerged in disorder-related research, regardless of the methods or analytic means employed.

Challenges to Accepted Theory

The studies included in this section have resulted in findings that challenge long-held assumptions and theories. Taylor et al. (1985) found that nonresidential land use was significantly correlated with crime and was a significant predictor of perceived physical decay. In a later study, Taylor et al. (1995) concluded that nonresidential land use in a neighborhood has a statistically significant effect on physical deterioration. Although parks well used by families may tend to deter crime, in general most nonresidential land uses act as magnets for neighborhood deterioration and subsequent loss of social control and ultimately (either directly through routine travel patterns of offenders or indirectly through increased disorder) for crime. This conclusion directly refutes Jane Jacobs' (1961) argument that businesses have positive impacts on the neighborhoods in which they are located. The authors did note that Jacobs wrote in a context in which neighborhoods and neighborhood businesses were typically more closely aligned ethnically. Further, in that context, longer hours kept by business owners may have discouraged loitering in the evening hours. Taylor et al. suggested that policy focus on preventing or slowing physical deterioration and loss of social control. All of the recommended policies begin with acknowledgement of the relationship between nonresidential land uses and disorder.

Sampson and Raudenbush (2004) directly challenged the broken windows and incivilities theories. In their study of Chicago neighborhoods, the authors employed SEM to examine the effects of structural characteristics – such as race, ethnicity, and class – on physical and social disorder. Observed disorder was a moderate correlate of predatory crime (e.g., robbery, aggravated assault, rape, and homicide) and varied with neighborhood demographic composition. When the demographic characteristics were taken into account, the connections between observed disorder and all but one predatory crime (robbery) were rendered statistically insignificant. The authors argued that concentrated poverty and lack of collective efficacy at the neighborhood level explained disorder and most types of predatory crimes. Rather than using disorder as a predictor of crime, Sampson and Raudenbush argued that disorder was – like crime – an outcome of broader structural factors. The authors concluded by arguing that the mere mitigation of disorder may be inadequate means of combating crime; addressing structural factors and collective efficacy would be a more appropriate response.

Troublesome Findings and Directions for Future Research

The relationship between disorder and crime is still a relatively new field of study for social scientists. Thus, many of the findings warrant closer examination. This section considers potential areas for future research. Three specific issues raised in the literature reviewed in this entry are presented for consideration: the spatial dimensions of disorder and crime, measurements of social disorder, and the future of *broken windows*.

While units of analyses have been discussed and debated among researchers, few studies have explicitly examined the extended spatial implications of neighborhood disorder. Brown et al. (2004a) did consider extended neighborhoods in their study of a former brownfield site in Salt Lake City. They found that blocks farther away from the cleaned up area had more crime and disorder. Their study was situated in a first-ring suburban community; it is uncertain whether studies set in an inner city or outer-ring suburb would result in the same or similar findings. Future research might focus on the spatial dimensions of disorder and crime in various settings, comparing perceptions, observations, and occurrences of crime across different types of neighborhoods in different types of communities.

Researchers have found it relatively simple to define physical disorder. After all, such incivilities are visible, sometimes highly so. Broken windows, graffiti, and even cigarette butts in gutters are observable; there would likely be agreement among researchers and residents that most of these are undesirable. Defining social disorder has proven more difficult. A clear way of characterizing social disorder has eluded researchers for the most part. Researchers who describe social interactions and relationships as disorders or incivilities risk pathologizing individuals, neighborhoods, and communities. Studies examined in this entry characterize noticeable offenses such as prostitution and drug dealing as social disorder. Yet, social disorder may extend beyond such egregious examples. Future research could delve further into social disorder, seeking to better understand and define it while making sure not to present marginalized and oppressed individuals, neighborhoods, and communities as abnormal or at fault. Interestingly, early disorder-related research was largely centered on the social aspects of disorder; the focus on physical disorder is more recent. In this way, clarifying and better defining social disorder represents a return to the roots of disorder-related research.

Perkins et al. (1990) join earlier researchers in pointing out that measuring social climate is a controversial issue in disorder-related studies. As this entry has shown, researchers often rely on measures of social climate to draw conclusions about groups and blocks. Yet such measures are typically collected at the individual level and aggregated to some higher level (e.g., street block or census tract). The validity and reliability of social climate in disorder-related research warrants scrutiny; future research may focus on issues of data aggregation (Shinn 1990).

Finally, the work of Sampson and colleagues has called key components of the broken windows theory into question. Their findings indicate that disorder and crime are not necessarily involved in a cause-and-effect relationship. Rather, disorder and crime are both outcomes of greater structural forces. In spite of these findings, broken windows theory continues to be a dominant framework in community policing. Future studies may wish to replicate Sampson's methods in other settings. By exploring disorder, demographic characteristics, and crime reports in multiple settings, researchers can corroborate or challenge Sampson and colleagues' findings. Additionally, researchers may choose to employ other data collection and analytic methods. For example, rather than replicating the SSO used by Sampson and Raudenbush (2004), researchers seeking to replicate the UK study may opt to use the BEI or some other data collection mechanism. Challenging a dominant paradigm and effecting change in policy and practice are not easy tasks; replicable studies in various settings may help shift the focus from broken windows to a more comprehensive discussion of social context and its manifestations in both disorder and crime.

Conclusion

This entry has provided an overview of recent disorder-related research, emphasizing the ways in which data have been collected and analyzed. Over the last few decades, researchers have supplemented resident perceptions of disorder with independent observations made by trained raters. Further, media accounts and police reports of crime have added points of comparison in some disorder-related research. Means of analysis have also evolved; the development of efficient analytic techniques such as hierarchical linear modeling and structural equation modeling has allowed researchers to explore layers of data and the effects of latent variables, respectively. Using these data-gathering methods and analytic techniques, researchers have uncovered personal and environmental characteristics that increase the likelihood of both perceiving and experiencing crime and disorder. Finally, implications recent studies have for future research include topics as diverse as geographic scope, clarification of social disorder, and the future of the broken windows theory.

Related Entries

- ▶ Broken Windows Thesis
- ▶ Crime Prevention through Environmental Design
- ▶ Defining Disorder
- ▶ Designing Products Against Crime
- ▶ Differential Social Organization
- ▶ Disadvantage, Disorganization and Crime
- ▶ Discriminant Validity of Disorder and Crime
- ▶ Drug Trafficking
- ▶ Informal Social Control
- ▶ Law of Community Policing and Public Order Policing
- ▶ Order Maintenance Policing
- ▶ Poverty, Inequality, and Area Differences in Crime
- ▶ Problem-oriented Policing
- ▶ Prostitution
- ▶ Public Housing and Crime Patterns
- ▶ Race and Ethnicity in Social Disorganization Theory
- ▶ Race, Ethnicity, and Youth Gangs
- ▶ Social Disorder and Physical Disorder at Places

Recommended Reading and References

Brown BB, Perkins DD, Brown G (2004a) Crime, new housing, and housing incivilities in a first ring suburb: multilevel relationships across time. Hous Policy Debate 15(2):301–345

Brown BB, Perkins DD, Brown G (2004b) Incivilities, place attachment and crime: block and individual effects. J Environ Psychol 24:359–371

Franzini L, Caughy MOB, Nettles SM, O'Campo P (2008) Perceptions of disorder: contributions of neighborhood characteristics to subjective perceptions of disorder. J Environ Psychol 28(1):83–93

Gau JM, Pratt TC (2010) Revisiting broken windows theory: examining the sources of the discriminant validity of perceived disorder and crime. J Crim Justice 38(4):758–766

Haney TJ (2007) 'Broken windows' and self-esteem: subjective understandings of neighborhood poverty and disorder. Soc Sci Res 36(3):968–994

Jacobs J (1961) The death and life of great American cities. Random House, New York

Kurtz EM, Koons BA, Taylor RB (1998) Land use, physical deterioration, resident – based control and calls for service on urban street blocks. Justice Q 15(1):121–149

Perkins DD, Florin P, Rich RC, Wandersman A, Chavis DM (1990) Participation and the social and physical environment of residential blocks: crime and community context. Am J Commun Psychol 18:83–115

Perkins DD, Meeks JW, Taylor RB (1992) The physical environment of street blocks and resident perceptions of crime and disorder: implications for theory and measurement. J Environ Psychol 12:21–34

Perkins DD, Taylor RB (1996) Ecological assessments of community disorder: their relationship to fear of crime and theoretical implications. Am J Commun Psychol 24:63–107

Perkins DD, Wandersman A, Rich RC, Taylor RB (1993) The physical environment of street crime: defensible space, territoriality and incivilities. J Environ Psychol 13:29–49

Robinson JB, Lawton BA, Taylor RB, Perkins DD (2003) Multilevel longitudinal impacts of incivilities: fear of crime, expected safety, and block satisfaction. J Quant Criminol 19(3):237–274

Ross CE, Mirowsky J (2001) Neighborhood disadvantage, disorder, and health. J Heal Soc Behav 42(3):258–276

Ross CE, Mirowsky J (2009) Neighborhood disorder, subjective alienation, and distress. J Heal Soc Behav 50(1):49–64

Sampson RJ, Groves WB (1989) Community structure and crime: testing social – disorganization theory. Am J Sociol 94(4):774–802

Sampson RJ, Raudenbush SW (1999) Systematic social observation of public spaces: a new look at disorder in urban neighborhoods. Am J Sociol 105(3):603–651

Sampson RJ, Raudenbush SW (2004) Seeing disorder: neighborhood stigma and the social construction of "broken windows". Soc Psychol Q 67(4):319–342

Schweitzer JH, Kim JW, Mackin JR (1999) The impact of the built environment on crime and fear of crime in urban neighborhoods. J Urban Technol 6(3):59–73

Shinn M (1990) Mixing and matching: levels of conceptualization, measurement, and statistical analysis in community research. In: Tolan P, Keys C, Chertok F, Jason LA (eds) Researching community psychology: issues of theory and methods. American Psychological Association, Washington, DC, pp 111–126

Suttles GD (1968) The social order of the slum: ethnicity and territory in the inner city. University of Chicago Press, Chicago

Taylor RB (1987) Toward an environmental psychology of disorder: delinquency, crime, and fear of crime. In: Stokols D, Altman I (eds) Handbook of environmental psychology, vol 2. Wiley-Interscience, New York, pp 951–986

Taylor RB (1996) Neighborhood responses to disorder and local attachments: the systemic model of attachment, social disorganization, and neighborhood use value. Sociol Forum 11(1):41–74
Taylor RB (1997) Social order and disorder of street blocks and neighborhoods: ecology, microecology, and the systemic model of social disorganization. J Res Crime Delinq 34(1):113–155
Taylor RB, Hale M (1986) Testing alternative models of fear of crime. J Crim Law Criminol 77(1):151–189
Taylor RB, Shumaker SA (1990) Local crime as a natural hazard: Implications for understanding the relationship between disorder and fear of crime. Am J Commun Psychol 18(5):619–641
Taylor RB, Shumaker SA, Gottfredson SD (1985) Neighborhood-level links between physical features and local sentiments: deterioration, fear of crime, and confidence. J Archit Plan Res 2(4):261–275
Taylor RB, Koons BA, Kurtz EM, Greene JR, Perkins DD (1995) Street blocks with more nonresidential land use have more physical deterioration: evidence from Baltimore and Philadelphia. Urban Aff Rev 31:120–136
Wilcox P, Quisenberry N, Cabrera DT, Jones S (2004) Busy places and broken windows? Toward defining the role of physical structure and process in community crime models. Sociol Q 45(2):185–207
Wilson JQ, Kelling GL (1982) Broken windows: the police and neighborhood safety. Atlantic 249(3):29–38

Disproportionate Minority Contact

▶ Juvenile Justice System Responses to Minority Youth

Dissocial Personality Disorder

▶ Psychopathy and Antisocial Personality Disorder

Distributional Analysis

▶ Quantile Regression Models to Analyze Experimental Data

DNA

▶ Identification Technologies in Policing and Proof

DNA Databasing

▶ DNA Technology and Police Investigations

DNA Fingerprinting

▶ DNA Technology and Police Investigations

DNA Identification

▶ DNA Profiling

DNA Profiling

Tracy Alexander
Cold Case Investigation, LGC Forensics,
Abingdon, UK

Synonyms

DNA identification; Forensic DNA matching; Genetic profiling

Overview

The use of DNA for identification purposes has often been the subject of controversy. This is particularly true in the UK at present with the imminent implementation of legislation to destroy a significant number of physical samples and to remove the related profiles from the national DNA database. This chapter covers the basic biology behind the use of DNA profiling and its use in the criminal justice system in the UK, particularly in terms of the relevant legislation and the provision of services by private laboratories. These issues are fundamental to the understanding of the problems that have always existed in terms of DNA interpretation, particularly in terms of mixed and partial profiles, plus emerging issues that have arisen with advances in research and technology.

Fundamentals of DNA Profiling in Forensic Science

DNA is a complex chemical compound comprised of relatively simple building blocks generally found within the nucleus of cells. The genome is the entirety of this cellular DNA, and encodes all the genetic information that governs an organism's structure and function, and is unique to the organism, except in genetically identical individuals. During sexual reproduction, the information determining the physical characteristics is inherited, half from each parent. Small sections of the molecule contain specific variations in the genetic code that can be statistically evaluated to assist in the process of individualization and therefore identification, most notably in human subjects. Analysis of the variable sections of DNA is frequently employed to determine paternity and to resolve immigration disputes, but public perception is focussed on its application as an aid to police investigative processes, to identify victims of crime and, by the transfer, exchange, and persistence of traces of DNA, the offenders responsible for crime.

The use of DNA identification methodologies has revolutionized crime scene investigation and the presence of forensic science in the courtroom. Historically, evidence of identity was limited to direct eye witness or, since the late 1800s, photographic recognition. Fingerprint identification has been in common use in the UK since the establishment of the Fingerprint Bureau in 1901, with the first conviction employing this technology being for a murder in 1905. Consequently, public confidence in the veracity of "dactyl" fingerprints and the legislation governing its utility has had considerable time to develop, whereas the revolution of DNA identification is still a relatively recent phenomenon and still presents legal and ethical issues that have yet to be fully resolved.

DNA

Deoxyribonucleic acid is a macromolecule containing the genetic instructions for the replication and function of all known living organisms (excepting a small number of viruses). Specific segments known as genes carry this genetic information and encode for proteins, for example, are known as genes. The presence of DNA in living material was originally recognized in 1869 by a Swiss physician, Friedrich Miescher. By 1937, the chemical composition of DNA had been identified and x-ray diffraction had demonstrated that DNA had a regular structure and the link between DNA and hereditary characteristics had been postulated. In 1953 Francis Crick and James Watson proposed the double-helix model of DNA structure. Two strands joined together by four distinct nitrogenous bases form a compact right-handed helix, thus allowing for a large amount of information to be encoded in a relatively small space. Crick and Watson's DNA breakthrough helix discovery was enabled through the work of two other scientists: Maurice Wilkins, who pioneered X-ray crystallography, and Rosalind Franklin who refined the technique for work with DNA. Together they identified the four bases which link the two strands of the helix; adenine, cytosine, thymine, and guanine (A, C, T, and G). These nitrogenous bases are attached to a ribose sugar on the phosphate-sugar backbone of each DNA strand.

Each molecule of DNA includes a pattern of these bases bonded together. Adenine forms hydrogen bonds with thymine while cytosine similarly engages with guanine. This base pairing holds the two strands of the double helix in close proximity to each other. It is the sequence of these base pairs within genes that is the genetic "blueprint" for the organism, encoding sequential instructions for amino acids, which are the building blocks of proteins. The information encoded in DNA is held in compact structures called chromosomes, of which humans have a normal complement of 46: 23 of maternal and 23 of paternal origin. Humans share almost all of their DNA sequence with other humans and much of the sequence information with other organisms. Only about 1 % is specific to humans when compared to a near relative such as a chimpanzee, and only about one tenth of that 1 % of DNA differs from one human to the next (with the exception of identical twins). However, this small fraction of the DNA still comprises some three million base pairs.

When analyzing person-specific DNA variations for forensic purposes, the sections of DNA examined are called short tandem repeats (STRs). As the name suggests, these are short sequences of DNA normally just four bases long repeated as adjacent blocks. It is predominantly the number of repeats of the core sequence that varies within individuals (e.g., GATA, GATA, GATA). STRs are, fortuitously, found in the noncoding regions of the genome (i.e., not within the genes) but the position (locus) of these on the chromosomes does not usually change from one person to the next. For crime scene investigation in the UK, a DNA profile is produced by measuring the physical length of the DNA at ten STR loci simultaneously, and the DNA analyzer displays the results as a series of peaks on a graph known as an electropherogram (EPG). The information carried therein is statistically very powerful, and the probability of a match being from someone other than the suspect and unrelated to them is given a numerical value. Assessments of match probabilities are made with reference to STR profiles separated into different racial groups in order that the most conservative figure is given.

DNA in the form of linear chromosomes exists within the nucleus of every nucleated cell. Several highly specialized cell types lose their nuclei and their DNA, for example, red blood cells are packed with hemoglobin while white blood cells retain their DNA, but in general DNA recovered from a white blood cell will be identical to DNA from any other cell type or tissue from the same individual. The nucleus is not the only source of DNA; cells contain tiny sausage-shaped organelles known as mitochondrion. These are the powerhouses of the cell specifically designed to make energy. Each mitochondrion has between one and ten copies of circular DNA that code for proteins required for the energy production and each cell may have hundreds of mitochondrion depending on its function (e.g., mitochondrion are found in large numbers in muscle cells). Hence in adverse conditions when the genomic DNA is too severely degraded to be useful, it is sometimes still possible to get some useful genetic information from the mitochondrion. The amount of genetic variation is far lower than genomic DNA but mtDNA can be very useful in identifying family members. The mitochondrion are inherited through the maternal line as at fertilization it is only the sperm's nucleus that is injected into the egg, not the mitochondrion.

While all nucleated cells contain DNA, not all nuclear DNA is contained within cells. Like many of the elements that go to make up an organism, the DNA is recycled. Dead cells are harvested and the DNA broken down. In this process, long stretches of DNA can be found in the plasma and recently also discovered in sweat. This discovery has far-reaching implications for the evidential value of "touch" DNA, particularly in terms of secondary and tertiary transfer between individuals and touched items (Quinones and Daniel 2012).

Development of Technology

The first paper reporting the use of DNA in a criminal context was published in 1985 by Alec Jeffreys, Professor of Genetics at Leicester University (Jeffreys 1985a).

Jeffreys was researching inherited variation in human DNA and he demonstrated how a DNA profile could be used to resolve issues of identity and kinship. Its initial use in legislative practice was to demonstrate that a child was the legal offspring of two individuals already granted asylum in the UK (Jeffreys 1985b).

DNA technology was the subject of research and development for the purposes of criminal investigation throughout the late 1980s and early 1990s primarily by the Forensic Science Service in the UK (FSS), previously known as the Home Office Forensic Service Laboratory (HOFSL). The FSS was the major provider and innovator in forensic science in the UK until its closure in March 2012. The multi-locus probe was introduced to routine casework in 1987 but required a large amount of sample material to produce a profile, although a full profile could give a likelihood ration of one in a million. This was followed by the single-locus probe in 1989 which allowed smaller samples to be tested and could produce statistics of one in 20–30 million.

However, both techniques were limited by the quantity of starting material required and regularly failed due to the poor quality and small amounts of DNA recovered from crime scenes. A major breakthrough came about with the development of the polymerase chain reaction (PCR) by American biochemist and Nobel Prize winner Kary Mullis. This enabled the biochemical copying of small amounts of DNA. This meant that sufficient target sequences could be amplified so as to be easily detectable by DNA analysis. During the PCR process the sample of DNA is treated with cycles of heating and cooling that denature the DNA and divide it into two separate strands, and a DNA primer is then used to anneal to these strands. Primers are short pieces of DNA containing sequences complementary to the target regions. The annealed primers together with a TAQ polymerase (an enzyme which enables a chain reaction to happen) allows two new copies of the sequence under investigation to be produced, one from each strand. This process of denaturing and rejoining was repeated again and again in a cyclic manner allowing amplifications of DNA even from samples with a relatively limited amount of starting material.

The flexibility of this method meant that only the targeted sections are copied and simultaneously those copies are marked by the addition of fluorescent tags. The fluorescent tags mean that the copied sections of DNA are visible when illuminated by laser light. In the UK, most DNA samples are subject to 28 cycles of PCR, doubling the amount of DNA available for analysis at each cycle.

The first use of PCR was in the detection of Human Leucocyte Antigen DQ alpha (HLA), introduced by the FSS in 1991, thereby allowing the analysis of much smaller stains. However, PCR has now become a fundamental part of all current profiling methodologies. The STR Quad system was introduced in 1994 which looked at four regions of DNA followed by, in 1995, Second Generation Multiplex (SGM) which gave a profile of six areas of an individual's DNA plus the sex indicator area, giving an average discrimination potential of 1:50 million. The SGM technique facilitated the introduction of a computerized database in 1995. SMGPlus®, used since 1999, is the main technique in use in the UK and looks at ten areas plus the sex indicator area and increases the discrimination potential to 1:1,000 million. However, many countries are moving to 16 or 17 loci multiplexes which the UK plans to do by 2014.

Use in the Criminal Justice System

In the UK, a Royal Commission was set up in 1993 to look at the opportunities that new DNA technologies might be able to offer the criminal justice system. The development of rapid, automated testing and the use of digital tools enabled interrogable databases to be quickly compiled. However, in order to deploy the technology, the police required the power to take DNA samples from those involved in criminal offences. The ability to take samples was restrictions in the Police and Criminal Evidence Act. PACE 1984 had been specific about the consent and authority required before a sample could be taken. Samples were categorized as "intimate" or "non-intimate" and regulations covered who could and could not take samples. The Criminal Justice and Public Order Act 1994 redefined intimate and non-intimate samples: mouth swabs were redefined as non-intimate and could be obtained without the person's consent. Though a suspect could refuse to open his mouth, it was then permissible to pluck head hairs with roots from which DNA could be obtained. Consent was still required to obtain blood and still required a qualified practitioner to take it. The Police Reform Act (2002) changed the regulations concerning the taking of samples: a police constable could now take non-intimate samples or could delegate this power to a "designated person" such as a civilian forensic officer. It also created the requirement for all new police officers to supply DNA samples to the Police Elimination Database.

In 1995, the evolution of DNA STR profiling technology and the subsequent change in legislation meant that the Home Office was in a position to create a database. The technology was simplified and automated, and the world's first criminal intelligence database. This was launched in April of that year: the UK National Criminal Intelligence DNA Database (NDNAD) (www.genome.wellcome.ac.uk/doc_wtd020879.html). Scotland and Northern Ireland have databases separate to that in England and Wales but all three are intersearchable.

First Use in a Criminal Investigation

In 1983 Lynda Mann was found raped and murdered in a deserted footpath in Leicestershire. Conventional grouping tests on semen samples from the body suggested that her killer was a person with blood type A and an enzyme type shared by approximately 10 % of males in the general population, but with no further evidence, the case remained unsolved. In 1986 the murder of another girl, also in Leicestershire, was linked by police through modus operandi. Police held a prime suspect, Richard Buckland, who confessed to the second murder but not the first. Jeffreys, in conjunction with the FSS, using extraction methods which enabled DNA from semen to be separated from DNA from vaginal cells, demonstrated that the murders were committed by the same person and that that person was not Buckland. Leicestershire Constabulary and the FSS began an investigation in which 5,000 local men were asked to volunteer blood or saliva samples, but after 6 months, no matches had been found. Later one of those men was heard bragging that he had been paid £200 to give a sample on behalf of Colin Pitchfork. Pitchfork was arrested in September 1987 and samples taken from him matched those of the double killer. Pitchfork admitted the murders and was convicted in 1988, becoming the first man to be convicted on DNA evidence, with Buckland being the first person to be proved innocent by DNA profiling. It was also the first time that the mass DNA screening of a population had been undertaken, a process that has been carried out on numerous occasions since. Even in cases where no suspect has been identified through this process, it has been beneficial in quickly eliminating a large number of individuals as being the donor of a profile believed to be crime related.

Persistence of DNA and Use in Historic Cases

Trace amounts of DNA can be recovered from bones as much as 5,500 years old, with opportunities within the forensic world for the identification of the victims and perpetrators of crime. In 1992, DNA testing gave compelling evidence linking remains recovered in Brazil in 1985 with Dr Joseph Mengele who died in 1979 by comparing a sample taken from the femur of the skeleton with samples taken from his widow and his son, which indicated full parental inclusion. In 1991, skeletal remains found in a shallow grave in Yekaterinburg were identified by Russian authorities as those of Tsar Nicholas II, the Tsarina Alexandra with three of their children. Remains discovered in a nearby smaller grave in 2007 were identified for the remaining two children using mtDNA, in part using samples from the Duke of Edinburgh who shares the same maternal link. Improvements in technology have meant that DNA is a frequently employed technique in resolving “cold cases,” such as in the recent conviction in the UK of David Burgess for the murder of Yolande Waddington in 1966.

Databases

In April 2007, responsibility for the NDNAD was transferred from the Home Office to the National Policing Improvement Agency (NPIA). The NPIA published the following statistics for the databases of England and Wales, Scotland, and Northern Ireland combined as at January 2012 (www.npia.police.uk/en/13338.htm).

Estimated total number of individuals retained on NDNAD	5,882,724
Total number of subject sample profiles retained on NDNAD	6,889,385
Total number of subject sample profiles retained on NDNAD from volunteers	43,915
Total number of crime scene sample profiles retained on NDNAD	409,695

In the USA, the Combined DNA Index System (CODIS) is the central database for

DNA profiles created by federal, state, and local crime laboratories and funded by the Federal Bureau of Investigation (FBI). Originally holding only the profiles of sex offenders, it has been extended to include a much wider range of offenses and encompasses:

- The Convicted Offender Index
- The Arrestee Index
- The Forensic Index (crime scene profiles)
- The Missing/Unidentified Persons Index
- The Missing Persons Reference Index

All 50 states have legislation governing the collection, storage, and retention of DNA and it is state law rather than federal law which governs which crimes qualify for CODIS. CODIS databases exist at local, state, and federal level and separate laboratories can retain or share information as they choose. However, CODIS is the largest database in the world.

Interpol also operates DNA Gateway, inaugurated in 2002, containing more than 117,000 profiles submitted by 61 member countries. Police from any of the 190 countries belonging to Interpol can access information (http://www.interpol.int/INTERPOL-expertise/Forensics/DNA).

Legislation

In the UK, legislation regarding the collection, storage, and use of data held on the NDNAD has developed over time, and no single piece of legislation covers every aspect legislative amendments have been made to old laws, and case law originating from judges' rulings has redefined the application of the legislation.

The Doheny and Adams ruling (1997) addressed the way in which DNA evidence should be presented in court. An expert could no longer give an opinion on whether a crime stain came from a suspect but had to explain its probability. In 2000, the Lashley judgement in the Appeal Court ruled that DNA evidence alone was insufficient to bring a conviction and supporting evidence was also required. However, this can be as limited as geographical proximity to the offence; living in or having visited the region where a crime scene stain is matched can be enough. Furthermore, in 2000, challenges to convictions in two cases, *R v. Wier* (murder) and *R v. "D"* (rape), sparked further reform. PACE (1984) required samples to be destroyed after acquittal or discontinuance; Wier and "D" were identified using unlawfully held DNA samples. The convictions were appealed, but the Lords found that it would have been against the cause of justice for the convictions to be set aside, and the Criminal Justice and Police Act 2001 amended PACE so that now all DNA data collected from persons arrested for an offence could be kept, whether found guilty or not guilty. The Criminal Justice Act (CJA) 2003 further extended this so that data could be logged from anyone arrested for an offence, irrespective of whether they were eventually charged. Within a short space of time, the database doubled in size, but the problem of holding records of innocent people was created. Samples and profiles could only be destroyed by application to the Chief Constable of the arresting force. In addition, the holding of samples for the prevention or detection of crime is exempt from the Human Tissue Act (2004), brought about in part as a response to the discovery of the retention of the organs of children without consent by the Alder Hey Children's Hospital.

Recent legislative changes will have a tangible impact on DNA identifications. The Nuffield Council on Bioethics published a report in September 2007 on "The forensic use of bioinformation: ethical issues" which recommended that proposals to extend police powers even further to include the taking of DNA for minor offences such as littering should not be implemented. In addition there has been a growing perception among civil liberties groups that the retention of samples on the NDNAD of individuals who were never convicted of an offence infringed the civil liberties of those whose DNA profiles were stored. The Appeals of "S" and Marper particularly apply: both were arrested in 2001 in separate incidents, but both cases were dropped. On application to the Chief Constable of South Yorkshire, both were refused the right to have their samples destroyed. Between 2002 and 2004, they were refused a judicial review of the decision, and their appeal was rejected first by the Court of Appeal and then by the House of Lords. However, in 2008, the European Court of Human Rights found in their favor, stating that the indefinite retention of profiles on a database interferes with a right to a private life and is particularly important for minors. Following consultation, this led to further amendments to PACE (1984) being promoted in the Crime and Security Act 2010 which passed into law but which has not been enacted to date. The coalition government elected in 2010 has further revised provisions for the rights of the individual in the Protection of Freedoms Act 2012, with far-reaching consequences for the NDNAD. The terms of the act require that approximately one million records of people on the database in England and Wales, and the copies held elsewhere, must be removed. The law does not require the removal of records of adults who have been convicted or have accepted a caution from the police, and people arrested for (but not convicted of) a serious offence can have their records retained for 3 years in the first instance, or a further two if there is the approval of a court (www.genewatch.org/sub-539488).

To date it has been possible to carry out a familial search on the UK NDNAD. On arrest, two buccal swabs are routinely taken, one to be kept as a backup in case the first sample fails to yield a profile. The "A" sample is processed; the remaining "B" sample is stored. When a crime scene sample has not given an immediate match on the database, it has been possible to look for previously loaded profiles that show similarities. Geographical factors and known information about the suspect, such as age, are taken into account to narrow the number of near matches. Once a manageable number of matches are obtained, it is possible to profile the "B" sample looking only at Y-STRs (DNA information obtained only from the Y chromosome and so only paternally inherited). This information allows the formulation of family trees indicating the presence of a male relative who fits the criteria for the offender but who has never been arrested for a recordable offence. The first successful prosecution relying on this procedure was in 2004 when Craig Harman was convicted of manslaughter for throwing a brick from a bridge which killed a lorry driver. Harman had left his blood on the brick (having injured his hand before taking it) but did not at that point have a police record. Forensic experts at the FSS found a profile with similar characteristics using the new techniques of familial searching through a relative whose profile was on the database, and as a result, the police traced Harman. Familial searching has always been limited to the most serious of cases and requires approval from the DNA ACPO lead. However, the application of familial searching in this way will no longer be available to police forces as following the Protection of Freedoms Act 2012, the UK has decided to destroy all "B" samples (over six million samples), even if taken from convicted offenders.

For further details, please refer to:

- Police and Criminal Evidence Act (PACE) 1984
- Criminal Justice and Public Order Act 1994
- Criminal Evidence Act 1997
- Criminal Justice and Police Act (CJPA) 2001
- Police Reform Act 2002
- Criminal Justice Act (CJA) 2003
- Serious Organised Crime and Police Act 2005
- Policing and Crime Act 2009
- Crime and Security Act 2010
- Protection of Freedoms Act 2012

Some rights of the suspect have been set aside when arrested or detained under certain sections of the Terrorism Act 2000, and specific rights apply to suspects under the age of majority, i.e., 18 (www.legislation.gov.uk).

It is interesting to note that the Association of Chief Police Officers (ACPO) recently announced, before the actual implementation of the Protection of Freedoms Act 2012, a new operation to capture the DNA of individuals whose profiles are not currently held on the database. Using powers under the Crime and Security Act 2010, which became law last year, the aim of Operation "Nutmeg" is to gather DNA profiles from criminals who were convicted before 1995 (when the database was launched). Initially the operation will target 11,993 criminals convicted of serious offences such as murder, manslaughter, and rape over the past 40 years. The success of the initiative is impossible to guess but there is obviously scope for further sampling, plus potential implications for the removal or retention of the samples currently targeted for destruction.

Provision of Services

The Forensic Science Service pioneered the use of the DNA database. Originally the Home Office Forensic Science Laboratory, it became an executive agency of the Home Office in 1991 but in 2005 changed its status from executive agency to a government-owned company. Demand for DNA services exceeded all expectations particularly after the launch of the NDNADB and large backlogs of samples gave private companies an opportunity to join the market by providing assistance. Police forces paid the FSS for their services but the increasing use of competitive tendering resulted in a loss of market share and the FSS ceased to be financially viable. It closed in March 2012 and services are now provided by a number of private companies such as LGC Forensics and Cellmark, who contract their services to the various constabularies in England and Wales. Private companies are able to invest in additional resources if it will have a beneficial effect on their profit margins and this, together with the competition in the forensic market, has led to some advantages such as reduced costs and turn round times, with a standard sample result being delivered in about 3 days. Forensic services in Scotland are provided by the Scottish Police Services Authority and in Northern Ireland by Forensic Science Northern Ireland. Each of these service providers will have separate provisions for the testing of different types of sample, and each must demonstrate the highest standards for preventing cross-contamination. Environmental testing is routinely carried out to monitor background levels and to inform processes to address any potential issues. It is also a requirement of the UK Accreditation Service (UKAS) and essential in order to maintain ISO 17025 accreditation, an assessment of internal standards, issued by the International Standards Organization and applied by UKAS in the UK.

Sampling

The NDNAD holds samples from three sources: personal samples from those arrested or charged, crime scene samples, and voluntary personal samples. Prior to 2004, these were non-evidential Criminal Justice (CJ) samples and required a confirmatory sample on rearrest. Since 2004, samples are taken under the provision of the Police and Criminal Evidence Act (PACE) and can be used for evidential purposes. Previously the taking of a confirmatory sample could entail a delay of up to 2 weeks during which time the suspect had the opportunity to abscond or to undertake multiple crimes, knowing their arrest was imminent.

There are three possible samples that can be taken from arrested persons for search against and inclusion on the database: blood, buccal scrapes (from the epithelial lining on the inside of the cheek against the buccal muscles), and pulled head hairs. Blood samples used to be taken from major crime suspects due to the likelihood of obtaining a usable profile, but improvements in technology have meant that this is no longer necessary and buccal scrapes are by far the most commonly taken. Pulled head hairs are an option if a suspect refuses to give a buccal sample voluntarily.

From a crime scene, there are many sources from which to obtain a DNA profile, with varying degrees of evidential value depending on the circumstances.

Blood is commonly found particularly at major crime scenes. The red blood cells contain the protein hemoglobin which bonds with oxygen to carry it around the body and there are about 4.5–5 million red blood cells per microliter (one thousandth of a milliliter) of blood, but they have no nuclear or mitochondrion DNA. It is therefore only the white blood cells, of which there are only 5–10,000 per microliter of blood, which can be analyzed for a DNA profile. In addition to the obvious opportunities such as blood left by suspects or victim's blood on weapons, the analysis of blood patterns (BPA) and subsequent profiling can be a vital tool in reconstructing a sequence of events.

Saliva does not contain DNA at point of production but epithelial cells from the inside of the cheek regularly slough off and are deposited in the saliva in sputum and on items coming into contact with the mouth, such as cigarette butts, drinking vessels, masks, gags, and on licked stamps and envelopes. Food is also an option but as saliva contains digestive enzymes, the material is often broken down, making odontology a potentially favorable evidential option.

Hairs with roots are a good source of DNA but dead hairs that have fallen out (telegen hairs) usually only contain mitochondrion DNA. The source of a hair may be identified by examining it in section. Determining the source may add value to an investigation, for example, finding head hairs on the boot of a suspect from a victim who has been kicked in the head may indicate the individual responsible for a fatal injury. Chest hairs on a particular knife in a multiple-stabbing incident with more than one suspect may indicate which weapon caused wounds to the chest and thereby which suspect is responsible for the chest injuries.

Sexual assaults are often resolved by finding an exchange of material between victim and suspect. Therefore a large number of samples are taken from victims in order to find traces from the suspect, either at a police station, hospital, or rape suite. These include internal and external vaginal and anal swabs plus samples from the mouth and any other pertinent areas, such as where the victim has been licked or bitten. Urine samples are also taken for toxicology but urine may be examined for DNA as some semen may be washed away with the urine during collection. On arrest, a suspect will also have numerous samples taken in order to find traces from the victim, including swabs from the glans and shaft of the penis. Exhibits are also collected from a crime scene to identify DNA from all parties involved. The ability to split DNA between semen and vaginal material is vital in the investigation of sexual assault using DNA evidence (the cellular and seminal fractions). In addition, a victim of rape is asked for permission to take a DNA sample from a child produced as a result in order to identify the offender, which is achieved by removing the mother’s profile from that of the child, leaving the profile of the child’s father. This can also be achieved after a sufficient period of development with the products of conception after a miscarriage or abortion.

Semen is only produced by postpubescent men with approximately one million spermatozoa per ejaculate. Vasectomized and naturally azoospermic males have a vastly reduced number of spermatozoa per ejaculate but it is often still possible to obtain a profile.

Footwear and clothing are often analyzed for “wearer” DNA as well as trace materials. This may come from cells sloughed off by close contact with the item or from sweat, as with “touch” DNA. Other items that have come into contact with biological material can also yield a DNA profile. One of the suspects in the Canary Wharf IRA bombings in November 1992 was identified from nasal debris found on a piece of green tissue paper recovered from a van containing a bomb that had failed to explode.

Trace amounts of DNA can be found in urine. They are hard to identify as they are contained only in discarded cellular material washed from the walls of the urinary tract and are diluted by other waste products. Similarly, very fresh fecal material may also yield trace amounts of DNA in mucal secretions on the outside of the stool.

DNA is also used for identification of the deceased. Venous blood is preferable but not always possible, particularly if the victim has bled profusely. Where a corpse is badly burned or heavily decomposed, samples may be sought in deep muscle tissue where nuclear DNA has not been exposed to degradation, or if there is none remaining, DNA may be obtained from bones or teeth.

In all circumstances, a protocol for the collection of samples must be strictly followed to ensure that there is no cross-contamination. Officers collecting samples will wear protective clothing, and once obtained, samples are sealed in sterile containers before processing. Crime scene samples and arrestee samples are processed separately, often at a completely different laboratory site.

Success Rates

It is assumed that subject samples should be almost 100 % successful in yielding a DNA profile.

Statistics for crime scene samples show approximate success rates for DNA profiles as follows:

Blood stains	93 %
Semen[a]	81 %
Saliva	76 %
Hair with roots	53 %
Other tissue	50 %
Other material[b]	26–30 %

[a] Not vasectomized or azoospermic individuals
[b] From unidentifiable source material

Processing and Analysis of Samples

Arrestee samples are usually buccal scrapes, one from the inside of each cheek. The two samples, “A” and “B,” are kept separately with the “B” sample being stored in the event that initial testing fails. Crime scene samples then go to a casework laboratory. Since the samples come in many forms, each sample will be treated differently to extract the DNA, and some will require presumptive testing, to locate the DNA for profiling and to establish whether a crime scene stain is from saliva, blood, or semen.

Stains and samples are then chemically treated to extract the DNA either manually or robotically and a process of quantification follows. A fresh sample will almost always provide an adequate amount of DNA. The smallest standard starting template permissible is a 1 ng (one millionth of a gram per milliliter) sample. This is subjected to the standard 28 cycles of PCR which provides sufficient amplified DNA to produce a profile which can be converted into a numerical code. If sufficient information from the profile is obtained, then this is sent to the NDNAD where it can be digitally compared to the millions of subject and crime stain profiles held there. Should there be no match, the profile from the crime scene sample will be held on the database indefinitely in case a match should appear in the future, either person to person, person to crime stain, or crime stain to crime stain. However, should the database indicate a match with a profile held on the database, then the information is passed on to the investigating force and a warrant for the arrest of that person is issued. Provided there is other supporting evidence, this means that the police can proceed with a prosecution.

High-Sensitivity Profiling

In some cases, the possibility of extracting DNA from crime scene samples will be reduced if the sample is old, degraded, or otherwise small in quantity. These can be subjected to more specialized techniques depending on the severity of the crime. All high-sensitivity work is performed in ultra-sterile conditions.

The FSS developed low copy number (LCN) to deal with samples containing insufficient good-quality DNA for standard profiling. The sample is subjected to 34 cycles of PCR to give more copies from which to draw a profile, but there are inherent difficulties. During 2007, a review was set up to examine the standards of science used in the analysis of LCN DNA, but before the review was complete, a challenge was made in the Northern Ireland case of *R v. Hoey*, a defendant charged with various offences in relation to the Omagh bombing in 1998. The case against him derived chiefly from DNA evidence using the LCN procedure, but the judge was not satisfied as to the integrity of the process and the prosecution case failed. This case was cited in the appeals of Reed, Reed, and Garmson in 2009 where it was found that in a very high proportion of profiles obtained using the LCN procedure, the profiles were not capable of robust and reliable interpretation.

DNASenCE (sensitive capillary electrophoresis) devised by LGC Forensics removes the impurities which interfere with the PCR process. The resulting profile is enhanced by a factor of 13. If necessary, the extracted material is loaded onto the capillary tube in greater concentration. This can enhance the resulting profile by a factor of 62. Other laboratories have devised similar enhancement procedures. The major advantage of post-PCR cleanup and enhancement methodologies is that the original sample remains for resampling using alternative techniques, whereas 34-cycle LCN uses up all the original material. The profile can then be searched against the NDNAD as with a standard SGM + sample (Gross et al. 2009).

In addition to DNA SenCE profiling, other methodologies have been developed, for example, Minifiler STR analysis which has been particularly useful in "cold-case" investigations. The eight mini-STRs examined are smaller versions of eight of the regions looked at in standard SGM + profiling. The smaller size means they are more robust and less prone to degradation but resulting profiles are still compatible with database searching. Y-STR profiling is also of benefit in sexual assault cases where there is semen present but no sperm or where the victim's profile is likely to mask the profile of the offender, for example, scrapings from a victim's fingernails.

Promega's Powerplex 16 and Applied Biosystems' Identifiler both examine the same ten SGM + sites plus the sex determination site, as well as an additional five "smaller" sites in common and one more which differs in each case. These processes are useful where a sample is degraded and where the new, smaller sites are likely still to be extant.

Next Generation Multiplexes

Next Generation Multiplexes (NGMs) examine an increased number of sites comprising those presently used for SGM + profiling plus additional sites. There is a directive from the European Network of Forensic Science Institutes (ENFSI) urging member states using DNA profiling for forensic science purposes to standardize their methodologies and profile storage across the European Union by applying a European standard set of alleles to improve cross-border compatibility. The aspiration is that this will encourage more cross-border searching. Several commercial companies have developed NGM kits that meet the ENFSI recommendation and in the UK the decision on which Next Generation Multiplex (NGM) kit will be adopted rests with the National Policing Improvements Agency. Once this has been decided, all subject and crime scene samples will be processed using an NGM and the NDNADB will be gradually "upgraded." The potential advantages of this will be a harmonization of approach across Europe as well as greater discrimination in individual matches. The major disadvantage is that the profiles will of necessity be more complex and therefore much more difficult to interpret, particularly in the case of mixtures.

Contrary to the impression given in "CSI"-type media, even a standard crime scene sample can still take 48 h to process, with further time taken for interpretation, depending on the complexity of the result. Under exceptional circumstances, this time can be reduced, but only with greatly increased demands on equipment time and personnel therefore incurring greatly increased costs. Obviously, the more complex the profiling method used, the longer it will take to interpret the findings.

Mixtures and Partial Profiles

Partial profiles and mixtures of profiles are particularly difficult to interpret. A partial profile occurs when there is insufficient good-quality DNA to produce a full profile. The match probability of such a profile can be significantly reduced to the point where it is not possible to reach any conclusion as to evidential or even intelligence value. In order to calculate the probability of a partial match, reference databases are used to estimate the proportion of the STR profile in the corresponding populations. The rarity of certain characteristics is also taken into account.

A notable case using LCN profiling involving a mixture of profiles was that of *R. v. Broughton* (2010), jailed for 2 years and 8 months in 1999 in connection with violent animal rights activism when police found a firebomb in his car. He was arrested again in 2007 and remanded in custody after incendiary devices were found in Oxford University colleges. A jury cleared him of possessing explosives but failed to reach a verdict on other charges and so there was a retrial in 2009.

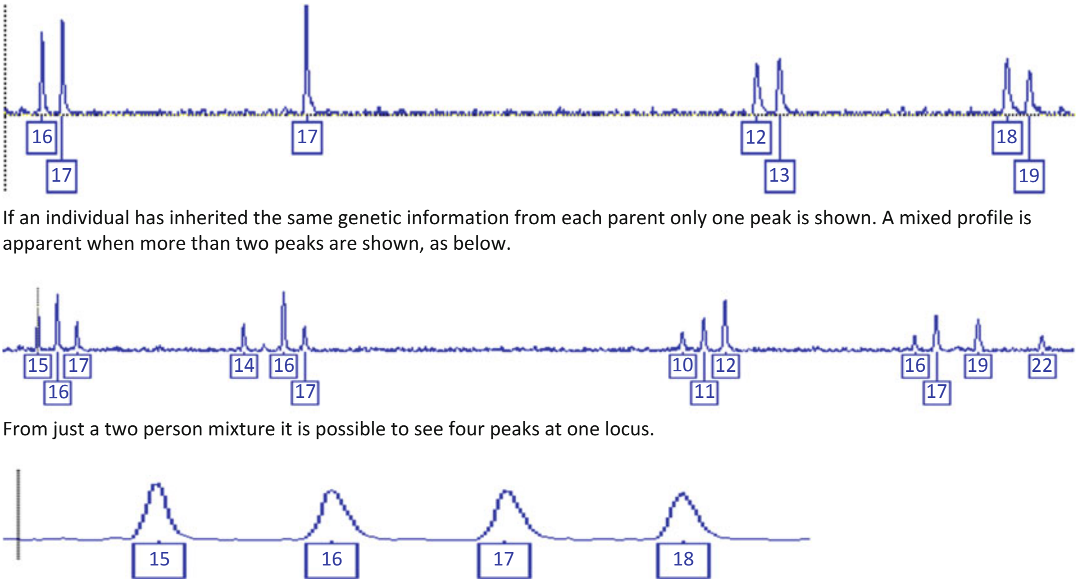

If an individual has inherited the same genetic information from each parent only one peak is shown. A mixed profile is apparent when more than two peaks are shown, as below.

From just a two person mixture it is possible to see four peaks at one locus.

From four peaks at one locus, there are these possible variations;

15,16 : 17,18
17,18 : 15,16
15,17 : 16,18
16,18 : 15,17
15,18 : 16, 17
16,17 : 15,18

DNA Profiling, Fig. 1 DNA electropherograms to demonstrate mixture interpretation

He was subsequently convicted and sentenced to 10 years in prison for conspiracy to commit arson. However, in 2010, this conviction was overturned on the basis that the DNA evidence had been unreliable. In the retrial, the uncertainty regarding the reliability of the DNA evidence hinged on differing interpretations of the results of processing which yielded the DNA profiles of more than one person on a matchstick which formed part of the incendiary device. The defense maintained that Broughton's DNA could not be present, and both prosecution and defense teams assembled impressive teams of DNA experts and statisticians using different software packages to calculate likelihood ratios. There was however fundamental disagreement on the question of how likely Broughton was to be included in the mixture or even how many profiles the mixture collected from the matchstick contained. Despite these differences of opinion, the jury returned a guilty verdict.

Mixed Profile Resolution

The components of a DNA profile are represented by a series of peaks that are measured and given a numerical value. A single-source profile will have two peaks at each locus, one from each parent (Figs. 1 and 2).

The more contributors to the mix, the more difficult interpretation becomes.

Without being able to subtract known profiles, such as from a victim or others known to have had the potential to contribute, the mixture has too many variables to interpret with certainty.

Interpretation

Each laboratory will set its own interpretation guidelines, giving a peak height at which a numerical value can be "read," but these are

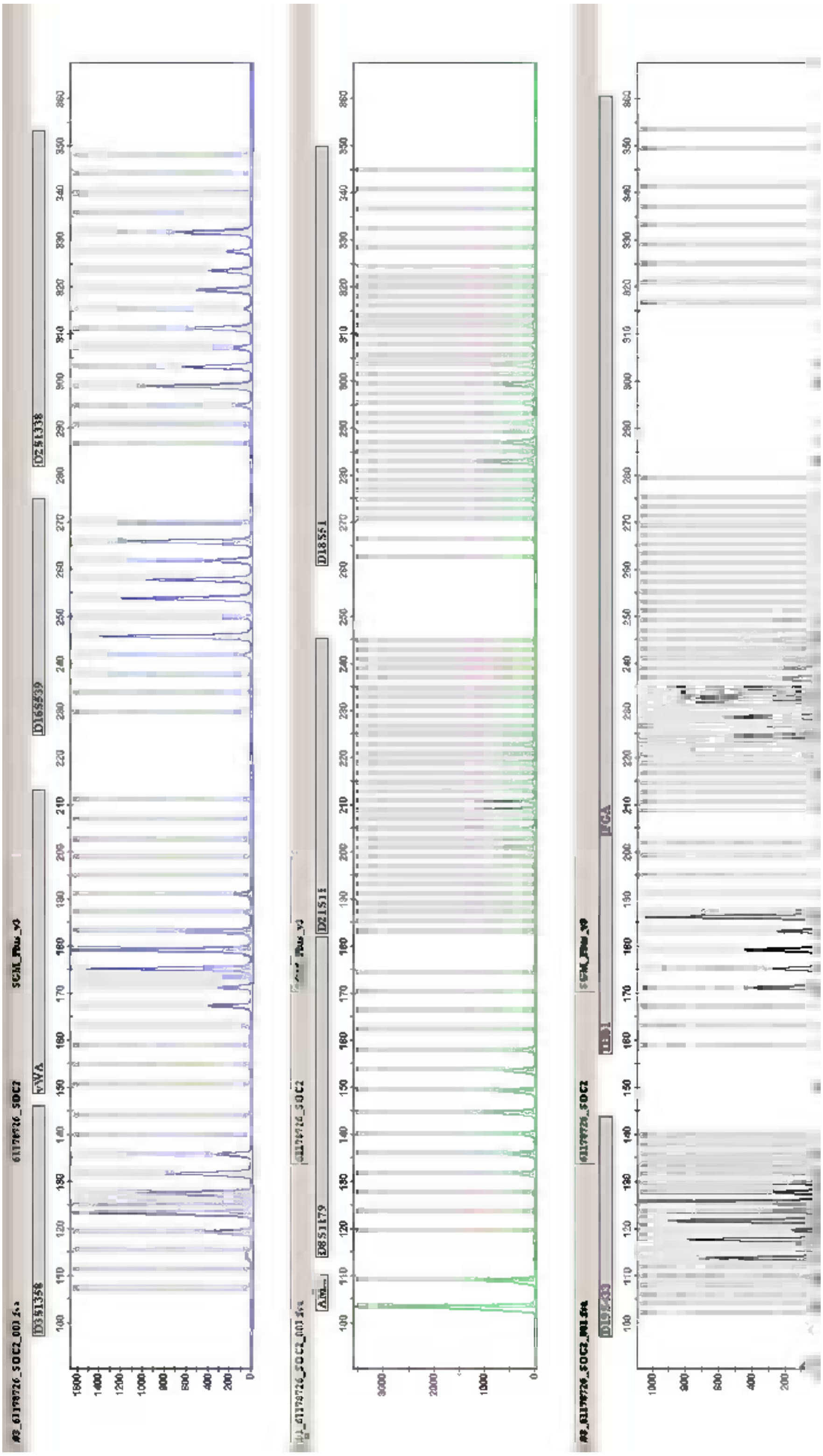

DNA Profiling, Fig. 2 Example of a complex mixture subject to variables including the nature of the offence and the source material. In addition, STR profiles can show "stutter," a term applied to peaks in a profile caused by the stochastic effects of PCR (Butler and Hill 2006). It is not certain why stutters occur, but they happen during the PCR process and show as a small peak usually one repeat smaller than the main band. Although a mostly clean, single-source profile may show stutter, this is easy to absorb in the overall interpretation, but the small bands often align with common alleles and so can be impossible to discriminate from true small peaks genuinely present from another profile as part of a mixture. Two peaks from the same source should be approximately the same size, but sometimes unidentifiable problems in profiling can lead to peak imbalance dramatically changing the appearance of a profile even though there is no background interference. In addition, stochastic effects can also occur. Referred to as "drop-in" and "dropout," drop-in is probably caused by contamination from an individual, the laboratory, or consumables. Dropout may be caused by a number of factors. There may be a problem with the primer leading to failure in amplification or the allele may be much larger than the usual size at a particular locus and is not seen with the others. DNA testing laboratories are continually attempting to adjust in order to cope with these difficulties but too little is known to be able to accommodate all eventualities successfully (Gill et al. 2000).

The Limitations of the System

Even with the incredible advances in technology, crime scene DNA profiling can be used to conclusively exclude a person from an inquiry, but even though it may provide compelling evidence of association, it is not proof of identity. Interpretation, particularly of complex mixtures, can be affected by many variables from interpretation guidelines differing between forensic providers to personal bias. In addition, it is estimated that the addition of partial profiles to the NDNADB may mean that approximately 0.1 % of matches are adventitious, that is, occurring by chance (Werrett 1997). The current sensitivity of profiling means that approximately 75 % of crime scene profiles generated are mixtures. Dr. Itiel Dror has recently expanded his studies on bias in fingerprint experts being dependent on external emotive factors to DNA-reporting officers (Dror 2012). In addition any process involving human beings is subject to human error, accidental or deliberate, and the pressure to perform efficiently in a commercial market place places additional pressures on forensic operators. Finally, despite extensive research in many different areas from collection to extraction to profiling, not enough is known about the propensity of individuals to leave their DNA and how that DNA is subsequently transferred between objects and individuals. While the Next Generation Multiplexes will increase the discriminatory power of profiling, the inherent increase in sensitivity will lead to more background DNA becoming part of a profile, so mixtures will be even more common, and the increase in the number of sites will make these mixtures even more difficult to interpret accurately and will take longer, with inevitable repercussions. Despite public perception, DNA profiling is never the complete answer to solving crime but continues to assist police forces as an aid to the criminal justice process.

Related Entries

▶ DNA Technology and Police Investigations

Recommended Reading and References

Butler JM. Fundamentals of forensic DNA typing

Butler JM, Hill CR (2006) Scientific issues with analysis of low amounts of DNA, 10th edn. American Medical Association

Dror IE (2012) Cognitive forensics and experimental research about bias in forensic casework. Sci Justice 52(2):128–130

Gill P et al (2000) An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. Forensic Sci Int 112:17–40

Gross T, Thomson J, Kutranov S (2009) A review of low template STR analysis in casework using the DNA SenCE post-PCR purification technique. Forensic Sci Int: Genet Suppl Ser 2(1):5–7

Jeffreys AJJ (1985a) Hypervariable "minisatellite" regions in human DNA. Nature 314(6006):67–73

Jeffreys AJA (1985b) Positive identification of an immigration test-case using human DNA fingerprints. Nature 317(6040):818–819

Quinones, Daniel (2012) Cell free DNA as a component of forensic evidence recovered from touched surfaces. Forensic Sci Int-Genet 6(1):26–30

Werrett DJ (1997) The national DNA database. Forensic Sci Int 88:33–42

DNA Technology and Police Investigations

Robin Williams and Matthias Wienroth
Centre for Forensic Science, Northumbria University, Newcastle Upon Tyne, UK

Synonyms

DNA databasing; DNA fingerprinting; DNA profiling; Forensic genetics

Overview

A forensic DNA profile is constructed by measuring highly polymorphic sequences of DNA in order to compare biological samples (especially blood, semen, skin cells, saliva, vaginal and nasal secretions, sweat, and other human tissue) found at a crime scene with samples taken from known individuals and those found at other crime scenes. First introduced in the 1980s, forensic DNA profiling is increasingly important to police investigations and criminal prosecutions in a large number of cases. However, despite its acknowledged successes in many criminal jurisdictions, debate continues about the general utility of forensic DNA technology to criminal investigations, the significance of due process and human rights challenges to the increasingly routine uses of DNA databasing, and the socio-ethical acceptability of some recent innovations in forensic DNA analysis.

History

The comparison of biological material recovered from scenes of crime with that taken from known individuals has a long pedigree in forensic science; the practice has been used widely to assist investigations and support the prosecution of offenders. Several different technologies, each seeking to capture varying degrees of distinctiveness exhibited by particular biological attributes, have been used to make these comparisons. Historically, the most widely accepted of these has been serological analysis in which blood samples are assigned to one of a small number of "ABO" blood types in order to include or exclude individuals as the possible sources of such material. However, in 1985, Alec Jeffreys and colleagues at the University of Leicester published two papers which demonstrated a new method for capturing individual differences at the genetic level (Jeffreys et al. 1985a, b). Described as "providing a level of individual specificity that was light-years beyond anything that had been seen before" (Newton 2004), the potential forensic science implications of what was first called "DNA fingerprinting" were quickly realized by Jeffreys and others.

The first, and highly prominent, deployment of this new technology in criminal investigations occurred only 2 years after Jeffreys' initial – and largely adventitious – laboratory discovery. Dawn Ashworth, a 15-year-old girl, went missing from her home in Northamptonshire on 31 July 1986. Her body was discovered 2 days later, and blood typing of semen recovered from her revealed identical features with semen obtained from the body of Lynda Mann who had been raped and murdered by an unidentified individual 3 years earlier (in both cases the semen donor was a Blood Group A secretor). The prime suspect for the murder of Dawn Ashworth was 17-year-old Richard Buckland, and following his arrest on 5 August 1986, he confessed to Ashworth's murder. However, Buckland was not Blood Group A and could not be linked to the semen recovered from Ashworth's body. In addition, he denied involvement in Mann's death. Faced with the contradiction between the biological evidence and Buckland's confession, investigators requested

Jeffreys to extract DNA from both recovered semen stains and to compare them to a blood sample taken from Buckland. Jeffreys' analysis concluded that Buckland's sample did not match the crime scene semen, but the semen taken from both crime scenes matched each other. Following Buckland's exoneration, the first ever mass DNA screening eventually resulted in the identification of Colin Pitchfork as the source of the semen, and Pitchfork was convicted of the two murders on 22 January 1988.

Despite this and other investigative successes in the UK and the USA, several technical limitations to this early form of forensic genetic analysis meant that it could be used only in a relatively small number of criminal investigations. These limitations included: the need to obtain relatively large quantities of DNA to undertake analysis; the time taken to complete the analytical process (several days or, in some cases, weeks); its unsuitability for use with degraded samples; the limited number of genetic markers that could be analyzed simultaneously; and the small number of samples that could be processed at one time. In addition, the criminal justice reception of forensic DNA fingerprinting was marked by an initial period of legal skepticism and methodological difficulty, especially concerning the estimation of "random match probability" and the communication of DNA profiling results by expert witnesses. The history of these "DNA Wars," especially the way in which they were conducted in several high-profile US criminal prosecutions, has been well documented by legal and human science scholars (see, for example, Kaye 2010; Lynch et al. 2008; Lynch and Jasanoff 1998).

Following these early controversies, and the stabilization of confidence in DNA profiling that resulted from their resolution, a series of subsequent changes in the methods for producing profiles also gradually overcame the technical shortcomings described above. The development of polymerase chain reaction (PCR), first reported by Kary Mullis in 1985/6, and its subsequent automation played a central role in this development by enabling the profiling of small and even degraded DNA samples through amplification. In the early 1990s, the shift to short tandem repeat (STR) multiplexes also meant that genetic information garnered through PCR became more easily comparable and digitizable, which made possible the automated computerized comparison of databased profiles.

DNA Evidence Recovery

In most jurisdictions, only trained crime scene examiners and forensic scientists recover biological material at scenes of crime. Some, however, also permit police officers to do so, at least in cases other than those of the most serious crimes. Regardless of these differences, investigators are increasingly conscious that aspects of the collection, packaging, recording, transporting, and laboratory handling of forensic genetic material can be subject to challenge in the course of judicial hearings, so all these practices are subject to strict protocols. What might otherwise be compelling DNA evidence can become "valueless if the authenticity of the samples used in the investigation cannot be confirmed" (Lincoln 1997). This means that very high levels of care are required to avoid prejudicing the weight otherwise given to the presence of DNA evidence recovered from crime scenes by lapses in continuity or by the use of inappropriate collection or preservations methods.

One clear feature of the trajectory of DNA profiling since the late 1980s has been the increasing success of forensic laboratories at being able to obtain analyzable quantities and qualities of DNA from an increasing variety of sources, including "trace" or "touch" DNA. In the early 1990s, Wiegand and associates were able to derive and profile DNA from debris obtained from fingernail scrapings and later from epithelial cells left on a victim's body following strangulation as well as cells left on strangulation tools. DNA analysis using an increased number of amplification cycles has been used successfully in UK forensic science casework since 1999 and a large number of studies have reported on the actual and potential uses of such "low copy number" (LCN) or low template (LT-DNA) analysis. While this technique has made possible the production of profiles from very small and also degraded samples of DNA, these profiles are subject to a series of technical effects which can make their interpretation very difficult. There has also been one significant UK judicial ruling that questioned the robustness of this particular practice (Weir, J. [2007] "The Queen v Sean Hoey," The Crown Court Sitting in Northern Ireland). For these reasons, and despite recent work conducted on behalf of the UK Forensic Regulator, the use of LT-DNA remains contested in many jurisdictions.

DNA Databasing

Throughout the late 1980s and into the early 1990s, all applications of forensic DNA profiling technology to criminal casework required the temporary storage of DNA profiles from nominated suspects, or other persons of interest, so that each individual's profile could be compared with those obtained from crime scene samples. The limited accounts of these practices suggest that such profiles (and usually the physical samples from which they were derived) were held for varying periods of time in local police or forensic laboratory collections. These collections were not usually regulated by formal mechanisms or by external bodies, but their existence caused little unease, perhaps because the practice was not widely known outside of restricted police and forensic science circles. However, it was not long before the use of DNA profiling expanded beyond reactive forensic casework in some jurisdictions. Forensic scientists in the UK, the USA, Austria, and New Zealand argued that searching DNA profiles recovered from crime scenes against profiles held in larger and centralized databases could, by the provision of "cold hits," facilitate the early identification of many more potential suspects (as well as the exclusion of some others as persons of interest). Once further technological advances made it possible to construct easily transportable digital representations of profiles and store them in continuously searchable computerized databases, the stage was set for a vastly expanded role for DNA profiling in many criminal investigations. In turn, this meant that the limitations of existing and varied local collections required them to be replaced by more extensive forensic DNA databases capable of contributing to the successful detection and prosecution of more offenders – at least in jurisdictions where there was the political will and administrative resolve to support such innovations (descriptions of the early days of DNA databasing in these and other jurisdictions can be found in Hindmarsh and Prainsack 2010).

The last 20 years have witnessed an increasing number of criminal jurisdictions in which such forensic DNA databases have been established – usually, but not always, at a national level. The first national forensic DNA database was created in England & Wales in 1995. Three years later, it was followed by the official launch of the US Federal Bureau of Investigation's Combined DNA Index System (CODIS), although all 50 US state databases were not fully connected through CODIS until 2004. Many other nation states established their own national forensic DNA databases during the last decade of the twentieth century, and each year sees the addition of more states authorizing or creating such collections. These hybrid scientific-legal innovations have usually required legislative changes and the provision of additional funds to create and populate such databases with profiles obtained from known subjects and those obtained from biological material recovered from crime scenes. Particular commercial actors – largely but not exclusively biotech companies – have also been prominent advocates and supporters of these state-driven criminal justice ambitions, and their interests have complemented the enthusiasm of many prominent policing stakeholders.

There is also widespread public support for the use of forensic DNA profiling in most contemporary democratic societies. Allowing for some simplification, the global trajectory of forensic DNA database expansion has followed a distinctive shape in which the types of people from whom DNA samples can be taken without consent, profiled, and retained have become greater. This has most often begun by sampling only those involved in the most serious crimes against the person, then moving on to include those involved (or suspected of being involved) in a range of property crimes. Police have also been authorized to take DNA samples at earlier points in investigative and judicial inquiries (for example, at the point of arrest rather than the point of charge, or even conviction). The period during which DNA profiles and samples can be retained has often been extended, and more DNA has been recovered from crime scenes. In addition to these developments, which together have resulted in the existence of much larger databases, the investigative applications of databased profiles have also expanded; there are increasing efforts to make possible the sharing of DNA profile information between criminal jurisdictions, for example, within the European Union via the Prüm Treaty, and beyond the EU through INTERPOL.

Current Databasing Practice

Taking DNA from suspects, and using these and crime scene DNA profiles as intelligence to support criminal investigations and as evidence to support prosecutions, are now central aspects of policing in a large number of countries. In those jurisdictions that have established a "national" DNA database, the vast majority of profiles entered are obtained from suspects during the investigation of crime. For those eager to promote and extend the powers of the police to sample criminal suspects, emphasis is placed on the immediacy of DNA profiling to exonerate individuals from, as well as implicate individuals in, further criminal investigations. The point is often made that enabling the police to obtain DNA samples from suspects introduces a reliable and objective method of evaluating their presence at a scene of crime. Yet, in order to obtain such a sample, the suspect must undergo a procedure to extract bodily material, and beliefs about the significance of this bodily intrusion color legislative decisions about forensic DNA sampling. Most states have enacted legislation that carefully specifies the situations in which the police may legally "interfere" with bodily integrity. Affording the police the authority to take DNA samples without consent involves making a number of legislative decisions. The first involves deciding at what stage in criminal procedure an individual should be subject to compulsory DNA sampling. A second, and related, issue is whether the police themselves should possess the authority to administer the collection, or if they should be required to obtain judicial approval. A third decision pertains to the types of offenses that should allow the compulsory sampling of suspects. And, following that, a fourth issue is whether such sampling should be relevant to the specific offense in question. A subsequent question arising is that of DNA retention: from whom, for how long, in what form (biological sample and/or digitized profile), and in what ways may the retained DNA material be used in the future.

Some states have reduced the significance of these issues and thus maximized the possible sampling of suspects. Equally, other states have minimized sampling because they have assigned more significance to them. The obvious example in the former category is the UK (England & Wales) which permits compulsory DNA sampling at the earliest point of investigation (upon arrest of a suspect) by the police for any "recordable" offense regardless of whether such a sample is relevant to the investigation. An example of a country in the latter category is France where nonconsensual sampling of suspects is completely prohibited.

In some states of the European Union (for example, The Netherlands, Luxembourg, and Malta), police are able to take compulsory DNA samples from individual suspects but such sampling requires judicial authority. Requiring the police to obtain this authority is a significant element in the distribution of powers across the criminal justice system. It prohibits the automatic sampling of criminal suspects by the police and transfers authority elsewhere, requiring the police to make a strong argument for compromising a person's bodily integrity. Nevertheless, it is increasingly common in many jurisdictions for the police to take DNA during the investigation of certain types of offenses. Most often, these are serious offenses which involve violence against persons. And some countries possess legislation which limits the collection of DNA from suspects in relation to specific – usually more serious – offenses. The situation in the USA is made more complex by differences between the states of the Union, but there seems a general trend there to establish "arrestee databases" in which DNA is taken, stored, and speculatively searched on arrest, even if subsequently removed (or at least sequestered) when individuals are not prosecuted, or if prosecutions do not result in findings of guilt. The many and constant changes in legislative frameworks governing forensic DNA databasing across the world are monitored by several police and civil society groups; from time to time, publications by INTERPOL and by a consortium of UK and US civil society groups provide serially updated accounts of these developments (see www.interpol.int and http://dnapolicyinitiative.org/).

Key Issues and Controversies: Claims-Making and the Measurement of Utility

Criminal justice researchers in the UK were among the first attempting to establish the value of DNA profiling and databasing to criminal investigations other than by citing conspicuous case successes. A series of studies, all funded by the Home Office, reported mixed success in this regard, although some UK government publications confidently asserted significant gains in the proportion of crimes detected when DNA evidence was available to investigators (McCartney 2006a, b). More recent studies have been carried out in the USA, mostly funded by the National Institute of Justice. Two of these have been especially ambitious. Roman and colleagues (Roman et al. 2008) carried out the first randomized control trial of the use of forensic DNA profiling across several US police districts in which the results of burglary investigations during which DNA evidence was collected but not analyzed were compared with those in which such evidence was made available to investigators. A second major study, of the role of forensic science in securing prosecutions, but also providing separate data specifically on DNA profiling, has recently been completed by Peterson and others (Peterson et al. 2010). This work followed the investigative process in a number of different kinds of crime (including homicide, rape, aggravated assault, robbery, and burglary) through the criminal justice system in order to assess the contribution of forensic science to the outcome of investigations and prosecutions. While a range of forensic evidence types was considered, particular attention was given to DNA analysis because of its ability to provide individualizing evidence capable of associating particular suspects to crime scenes. Despite these and other studies, there remains disagreement over the extent to which research has yet provided a robust account of the utility of DNA profiling and databasing to criminal investigations. The UK Human Genetics Commission, along with academic researchers in many jurisdictions, has commented on the need for better data to be provided by the custodians responsible for the operation of national DNA databases as well as by the police who are responsible for the use of DNA match information in support of individual investigations (Human Genetics Commission 2009).

Key Issues and Controversies: DNA Databases and Human Rights

All sampling and databasing of the genetic profiles of individual suspects by the police – especially those taken without consent – involves consideration of a number of legal, social, and ethical issues. As reflected in the comment of the Council of Europe Committee of Ministers from 10 February 1992, such sampling must "[t]ake full account of and not contravene such fundamental principles as the inherent dignity of the individual and the respect for the human body, the rights of the defence and the principle of proportionality in the carrying out of criminal justice." For some critics, forensic DNA sampling and databasing threaten the bodily integrity of citizens who are subject to the forced and nonconsensual sampling of their genetic material based on decisions of police and judicial actors prior to findings of guilt by relevant authorities. In addition, there is concern that DNA databasing also violates privacy rights by allowing the use of profiles and the storage of biological samples, storage which itself creates the potential for the future misuse of such samples held in state and privately owned laboratories.

Accordingly, these and other legal, social, and ethical issues have been explored in several academic and policy studies. In the UK, two major agencies – the Nuffield Council on Bioethics (2007) and the Human Genetics Commission (2001, 2002, 2009) – have both published substantial critical reports on the distinctively forceful legislative and operational developments in DNA databasing that occurred in England & Wales between the establishment of the National DNA Database in 1995 and the decision of the Council of Europe's European Court of Human Rights in the case of S & Marper v the UK Government in 2008.

In addition, the monitoring group Genewatch UK have also been actively interrogating official statements and statistics on the National DNA Database for a number of years as well as appearing before several House of Commons Select Committees that have inquired into aspects of forensic DNA profiling and databasing in England & Wales (see http:www.genewatch.org/). In the USA, the American Civil Liberties Union has frequently criticized the state and federal expansion of DNA collection and retention, and the American Society of Law, Medicine and Ethics sponsored a series of national workshops and conferences on US developments in DNA profiling, attended by many of the major academic, scientific, and operational experts in the field between 2003 and 2006.

Key Issues and Controversies: Recent Innovations in Forensic DNA Analysis

There are many investigations in which genetic material has been recovered from a crime scene but no matches with databased profiles have been made. In such circumstances, investigators may seek other ways to infer personal features of an unknown individual from the DNA that they have left at the scene. New forms of genetic knowledge, technological improvements in sample processing, and the premium on the investigative ingenuity necessary for the detection of "hard-to-solve" serious crime contribute the means and desirability for constant innovations in methods for interrogating the informational content of biological samples obtained from scenes of crime. At present, analysis can be undertaken to gain information about phenotypical attributes, "biogeographic ancestry," and "familial relationships." Some interrogations involve the direct examination of coding regions of the human genome – genes themselves – while others rely on new ways of using information from the noncoding areas already examined by conventional forensic profiling. The most significant of these approaches are briefly described in the following sections.

"Genetic Ancestry," "Population Groups," and Forensic Investigations. One particularly complex area of genetic information of interest to criminal investigators has been that of patterned human genetic diversity. Knowledge of the differential variability of genotypes according to population groups has informed the calculation of random match probabilities since the early days of DNA profiling. While population genetics are largely of interest to the specialized forensic community, the ability to infer the "biogeographic" origin of an individual who left otherwise unidentified biological material at a crime scene may be of significant interest to investigators. Reliable inferences of the "racial origins," "ethnic origin," "ethnic affiliation," or even "ethnic appearance" of such an individual can be used to focus subsequent inquiries, to determine an interview strategy, to compare with witness statements, or to design an intelligence-led mass DNA screen. However, there are significant conceptual and operational uncertainties surrounding such categorizations of individuals. Moreover, there is a danger, well

articulated, for example, by Duster and colleagues, that "race" will be reified in the attempt to define distinctive human population groups and subgroups. These critics also point to the ways in which questions of genetic "ancestral attribution" for these limited and pragmatic purposes can easily become confused with more ambitious theoretical assertions concerning the biology of "race" as well as "some old and dangerously regressive ideas about how to explain criminal conduct" (Duster 2003: 151).

Despite these problems, a number of forensic laboratories and agencies have added to their analysis of autosomal DNA STRs and SNPs, Y-chromosome STR and Y-chromosome SNP multiplexes for the analysis of loci whose polymorphic range is already databased by a variety of international consortia. The Y-chromosome is an especially suitable site for such investigations because of the low rate of recombination on this chromosome. This means that particular male-specific haplotypes are preserved across generations and vary systematically across different population groups. The research and reference databases used to inform developments of this kind are the Y-STR haplotype reference database, the US population database, and a European population database (see www.yhrd.org).

Autosomal Single Nucleotide Polymorphisms (SNPs). There have been several surveys of such polymorphisms, many of which are collected together in a global haplotype collection known as the "HapMap" (www.hapmap.org). SNPs have a much more limited polymorphic range than STRs, so that about four times as many SNPs are needed to produce profiles capable of discriminating individuality as those used by STR typing. Nevertheless, the establishment and expansion of SNP forensic databases alongside current STR collections is not out of the question, and the analytical scope of SNPs means that they can serve valuable forensic identification functions, especially in situations where samples are too degraded to make STR typing possible. The Y Chromosome Consortium has provided a "phylogenetic tree" which describes the history of 18 major lineages of diverse SNP haplotypes across human population, and other scholars recently discussed the development and use of Y-SNP multiplexes to support inferences of population origin.

It is difficult to determine the significance of these efforts to provide information about genetic ancestry. A clear preference for SNP markers over STRs seems to have emerged over the last few years, and large numbers of SNPs are now combined to form "ancestry informative markers," some of which have been used in forensic casework. However, until recently, there have been problems in standardizing such markers and in the standardization of haplotype nomenclature. Furthermore, the increase in populations of mixed origin is a feature of complex urban societies so that "indirect deductions about individuals are often unreliable" (Jobling 2001: 161). Even when the pattern of differential SNP distributions is used to "improve" the accuracy of such inferences, as in the case with "proportional ancestry" studies, significant uncertainties remain.

Inferring Specific Phenotypical Features. In addition to efforts at identifying the genetic ancestry of unmatched crime scene stains, forensic scientists and police investigators remain interested in whether interrogations of such stains may yield information about a wide repertoire of visible characteristics of their donors. For the purposes of police investigations, the ability directly to determine individuals' physical characteristics may be more appealing than inferring those characteristics from assumptions of biogeographic ancestry. The most frequently used method of direct interrogation – of the amelogenin locus – determines the biological sex of the DNA source and is already incorporated into the majority of multiplex systems. Aside from this test, however, the research literature reveals limited success in attributing phenotype from genotype in ways that are practically useful to investigators.

An initial review of forensic work in this field suggests that positive results remain scarce, and are focused on probabilistic inferences about pigmentation (see for example Kayser and de Knijff 2011). While analysis of the human melanocortin-1 receptor gene can be used to indicate "red hair" in the relevant subject, hair loss and hair coloring can make this test problematic when applied in investigative contexts. Both STR and SNP profiling are of interest to forensic scientists keen to develop predictive tests for a range of other observable physical characteristics including eye color, skin type, and height. It seems likely that SNP analysis may prove more successful than STR markers as the basis for such tests. This is not simply because most genomic mutations are single base changes but also because there is considerable research being carried out beyond the forensic community to identify SNP polymorphisms and their effects on a variety of human attributes.

Familial Searching. The term "familial searching," as used by forensic scientists and police officers, refers to a form of database searching reliant on knowledge about the probability of matches between the STR markers of two members of the same family (as opposed to the probability of matches between these markers when the individuals compared are unrelated). This practice makes use of understandings of inheritance that prefigured the discovery of the structure of DNA and which had been largely applied to understanding variation in human, animal, and plant phenotypical characteristics (for a summary account of these assumptions as applied to the forensic context, see Bieber et al. 2006). Because familial searching relies on identifying a pool of possible genetic relatives of a suspect, who are then subject to more direct investigation (typically by being interviewed by the police), forensic science policy makers in the UK and elsewhere have also acknowledged that a number of ethical issues need to be addressed when this strategy is being considered (see Greeley et al. 2006; Haimes 2006). Issues arise in both the searching of profiles on a database and in the subsequent investigative trajectories that follow the provision of a list of individuals derived from such a search. A genetic link between individuals might be previously unknown by one or both parties and police investigations may make such information known to them for the first time. Equally an investigation may reveal – to investigators, if not to informants – the absence of genetic links which participants assumed to have existed. There is also the question of whether this kind of use of an individual's databased DNA violates promises of privacy and confidentiality made when genetic material was originally donated voluntarily, for example, in the course of a mass DNA screen. Furthermore, assertions about criminality, geography, and familial relatedness that are central to the use of this forensic methodology are especially problematic – even if they do accord with the rhetorical endoxa of many detectives – and they reveal pervasive problems associated with the confusion between "genetic" and "social" relatedness ("families" are not only constituted through genetic lines but through clusters of non-genetically related individuals) as well as the implicit assumption that criminality is fostered because of such relatedness (either for genetic or social reasons).

Conclusion

This entry has outlined the history of the main DNA technologies that are currently used in police investigations along with their reception by criminal justice actors, especially in the UK and the USA. The uncertainties reflected in the "DNA Wars" of the 1980s and 1990s have diminished as scientific and legal agreements about the strength and limitations of many of these technologies have stabilized. Police enthusiasm for DNA technology has grown, and in many jurisdictions, legislators have increased police powers in order to maximize its potential uses to support criminal investigations and prosecutions. A variety of international bodies (especially the European Network of Forensic Science Institutes, The American Society of Crime Laboratory Directors, and the International Society for Forensic Genetics) support efforts to standardize forensic genetic practice and shape the education of relevant scientific and legal personnel. At the same time, new normative and empirical uncertainties arise as legislative, operational, and technical innovations are critically appraised by scientific, legal, and human science scholars.

Questions of scientific adequacy, investigative utility, and ethical acceptability, as well as the relationship between each of these questions, are sure to engage criminological interest for the foreseeable future.

Related Entries

- ▶ Cognitive Forensics: Human Cognition, Contextual Information, and Bias
- ▶ DNA Profiling
- ▶ Forensic Science and Criminal Inquiry
- ▶ Forensic Science Culture
- ▶ History of Forensic Science in Policing
- ▶ Identification and the Development of Forensic Science

Recommended Reading and References

Aronson JD (2007) Genetic witness: science, law and controversy in the making of DNA profiling. Rutgers University Press, New Brunswick

Bieber FR, Brenner C, Lazer D (2006) Finding criminals through DNA of their relatives. Science 312:1315–1316

Duster T (2003) Backdoor to eugenics, 2nd edn. Routledge, New York

Fraser J, Williams R (2009) Handbook of forensic science. Willan, Cullompton

Gerlach N (2004) The genetic imaginary: DNA in the Canadian criminal justice system. University of Toronto Press, Toronto

Greeley HT, Riordan DP, Garrison NA, Mountain JL (2006) Family ties: the use of DNA offender databases to catch offenders. J Law Med Ethics 34:248–262

Haimes E (2006) Social and ethical issues in the use of familial searching in forensic investigations: insights from family and kinship studies. J Law Med Ethics 34(2):263–276

Hindmarsh R, Prainsack B (2010) Genetic suspects: global governance of forensic DNA profiling and databasing. Cambridge University Press, Cambridge

Human Genetics Commission (2001) Whose hands on your genes? A discussion document on the storage protection and use of genetic information. HGC, London

Human Genetics Commission (2002) Inside information: balancing interests in the use of personal genetic data. Department of Health, London

Human Genetics Commission (2009) Nothing to hide, Nothing to fear? Balancing individual rights and the public interest in the governance and use of the national DNA database. HGC, London

Jeffreys AJ, Wilson V, Thein SL (1985a) Individual-specific ‘fingerprints’ of human DNA. Nature 316: 76–79

Jeffreys A, Wilson V, Thein SL (1985b) Hypervariable ‘minisatellite’ regions in human DNA. Nature 314: 67–73

Jobling MA (2001) Y-chromosomal SNP haplotype diversity in forensic analysis. Forensic Sci Int 118:158–162

Kaye DH (2010) The double helix and the law of evidence. Harvard University Press, Cambridge, MA

Kayser M, de Knijff P (2011) Improving human forensics through advances in genetics, genomics and molecular biology. Nat Rev Genet 12(3):179–192

Lincoln PJ (1997) Criticisms and concerns regarding DNA profiling. Forensic Sci Int 1997:23–31

Lynch M, Jasanoff S (1998) Contested identities: science, law and forensic practice. Soc Stud Sci 28(5–6):675–686

Lynch M, Cole S, McNally R, Jordan K (2008) Truth machine: the contentious history of DNA fingerprinting. Chicago University Press, Chicago

McCartney C (2006a) The DNA expansion programme and criminal investigation. Brit J Criminol 46:175–192

McCartney C (2006b) Forensic identification and criminal justice: forensic science, justice and risk. Willan, Cullompton

Newton G (2004) Discovering DNA fingerprinting 2004. http://genome.wellcome.ac.uk/doc_wtd020877.html Accesed 21 December 2012

Nuffield Council on Bioethics (2007) The forensic uses of bioinformation: ethical issues. The Nuffield Council, London

Peterson J, Sommers I, Baskin D, Johnson D (2010) The role and impact of forensic evidence in the criminal justice process. National Institute of Justice, Washington, DC

Roman JK, Reid S, Reid J, Chalfin A, Adams W, Knight C (2008) The DNA field experiment: cost-effectiveness analysis of the use of DNA in the investigation of high-volume crimes. Urban Institute, Washington, DC

Williams R, Johnson P (2008) Genetic policing: the use of DNA in criminal investigations. Willan, Cullompton

Domestic Burglary

▶ Residential Burglary

Domestic Terrorism

▶ Homegrown Terrorism in the United States

Domestic Violence

Russell P. Dobash[1,2,3] and
Rebecca Emerson Dobash[1,2,3]
[1]School of Law, University of Manchester,
Manchester, UK
[2]School of Social Justice and Inquiry, Arizona
State University, Glendale, AZ, USA
[3]Centre for Law, Crime and Justice, School of
Law, Strathclyde University, Glasgow, Scotland

Violence has long been a subject of interest for criminology, but it was not until the last quarter of the twentieth century that gender became a significant explanatory factor and domestic violence became a focus of research. Although an early, classic work by Wolfgang, Patterns of Criminal Homicide (1958), included a groundbreaking investigation of domestic murders, this was mostly ignored until the 1980s. In the USA, where most criminological research was conducted, the focus was on gang delinquency. This focus drew upon the pioneering work of Merton (1938) and produced a number of empirically informed accounts of gang delinquency and gang violence (Short and Strodtbeck 1965). The prevailing characterization of violent events was of a "face game" with disputes involving "honor" in which both victim and offender "agreed" that violence was an appropriate means of settling the conflict as verbal responses escalated to physical violence (Luckenbill 1977). The concept of "victim precipitation" defined the mutual and equally culpable participation of both parties as each "agreed" to the escalation from verbal encounter to physical violence.

When domestic violence (DV), now termed intimate partner violence (IPV), was "discovered" in the mid-1970s, the sparse accounts of physical and sexual violence against women within criminology relied heavily upon the concept of "victim precipitation" with its implicit conception of male-male encounters, provocation, and mutual agreement to violence, and this was applied to accounts of conflicts between men and women culminating in men's violence against women (e.g., Amir 1971). Similarly, within the psychiatric literature, women victims of domestic violence were deemed to be "provocative," "aggressive," and/or "masculine," and thus responsible for the violence men used against them (for a review see Dobash and Dobash 1979, 1992). Thus, as social scientists began to focus on domestic violence in the 1970s and 1980s, often in conjunction with activist community groups, many looked outside criminology to the emerging feminist literature that stressed gender, male domination, aggression, and violence as more relevant to this area of research. Such approaches have generally dominated the study of IPV, and the body of literature is now voluminous and includes criminology, socio-legal studies, sociology, anthropology, psychology, evolutionary psychology, health, and medicine. Here, the focus is upon the extent of the violence, the nature of violence and controlling behavior, the context in which it occurs, the characteristics of male perpetrators, and the response of the criminal justice system.

Since the "discovery" of this form of violence in the 1970s, there have been debates about definitions, terminology, research methods, and the resulting findings. Over time, several terms have been used: "wife beating," "wife abuse," "wife battering" which were followed by "woman beating, abuse, or battering," "domestic violence," and "intimate partner violence." There have been debates about *how much of this violence exists* in a given place at a particular time; *who commits the violence* (men, women, or both, i.e., symmetry or asymmetry in perpetration by men and/or women); a *narrow vs. broad definition of what counts as violence*, narrowly restricted to physical acts of violence or broadly expanded to include other acts that are emotional, financial, and the like; *what should be done about it and by whom*; as well as other debates not discussed here (see Dobash et al. 1992; Dobash and Dobash 2004).

Nature and Extent of Intimate Partner Violence

Here, the focus is on serious, physical "violence" against women rather than upon aggression, controlling behavior, emotional abuse, or financial deprivation. This is not to say that such nonviolent aggressive acts are not problematic or worthy of concern, intervention, or research. However, it is physical acts of violence that are most likely to be defined as illegal or criminal and to warrant intervention by the justice system including police, courts, probation, and prison. Around the world, several national and international studies have attempted to establish the nature and extent of intimate partner violence. When the focus is upon serious physical violence, the overwhelming evidence indicates that it is women, not men, who are significantly more likely to be the victims of violence from an intimate partner, to suffer serious consequences and injuries, and to require emergency attention and hospitalization (Krahe et al. 2005; Tjaden and Thoennes 1998; Watson and Parsons 2005).

National population-based surveys in several industrialized countries suggest that about one-quarter of adult women will at sometime in their life experience at least one act of violence from a male intimate partner (Backman and Saltzman 1995; Greenfield et al. 1998; Mirrlees-Black 1999; Tjaden and Thoennes 1998; Wilson et al. 1995; Walby and Allen 2004). Evidence collected by the World Health Organization reveals that intimate partner violence is a “common experience” for women throughout the world (Krug et al. 2002). A WHO meta-analysis of 50 population-based studies from 35 countries found lifetime estimates of between 10 and 69 %. In most countries this varied from 10 % to 50 % (Krug et al. 2002). A subsequent carefully conducted WHO study involving probability samples and standardized face-to-face interviews with 24,097 women at 15 sites in ten countries revealed a lifetime prevalence of 13–61 % for moderate or severe violence (acts capable of inflicting injury and hospitalization) from partners of ever cohabiting or married women with most countries reporting prevalence of 25–45 % (Garcia-Moreno et al. 2006).

Sexual violence against women in intimate relationships also occurs with the WHO research revealing that between 6 % and 59 % of women had experienced “forced sexual intercourse” by an intimate male partner, with most areas varying from 10 % to 50 % (Garcia-Moreno et al. 2006; Watson and Parsons 2005). Research conducted primarily in the USA indicates that 10–15 % of ever married or cohabiting women have been raped by an intimate partner, about one-quarter of all rapes involve intimate partners, and a considerable proportion of physically abused women also suffer sexual assault (Campbell 1999; Randall and Haskall 1995; Russell 1990; Ullman and Siegel 1993). Results of the comprehensive WHO study indicate that the violence is often severe and results in serious injuries, it occurs on a frequent basis, there is a strong relationship between frequency and severity, and it is intrinsically related to sexual violence (Garcia-Moreno et al. 2006). This definitive study corroborates and confirms the results of scores of other national and local surveys conducted in countries throughout the world.

Crime victimization surveys and evidence gathered from other sources have persistently shown that women are much more likely than men to be victimized by an intimate partner, to suffer injuries, and to require medical treatment. In the USA, National Crime Surveys have been conducted annually since 1972, and these and victimization surveys conducted in other countries have persistently shown that women constitute 70–90 % of all victims of assaults between intimate partners and that women are much more likely than men to report serious injuries (Archer 2000; Gaguin 1977–1978; Schwartz 1987; Johnson and Sacco 1995; Tjaden and Thoennes 1998; Worrall and Pease 1986). Evidence from other sources, such as police, court, and accident and emergency records, gathered in numerous countries over a number of years suggests that in an overwhelming proportion of cases, women are the victims of intimate partner violence (Abbott et al. 1995; Archer 2000; Dobash et al. 1992; Dobash and Dobash 2004; Nazroo 1995).

Extensive population-based surveys and intensive studies provide information about the specific nature of the physical acts of violence perpetrated by men in intimate relationships and repeatedly reveal a range from the more frequent use of pushing, shoving, slapping, punching, and kicking to the less frequent use of weapons and strangulation (Dobash and Dobash 1979; Dobash et al. 1998; Johnson 1996; Tjaden and Thoennes 1998).

Injuries

Research suggests that the most common injury from violent attacks involves bruising and lacerations of the face and body. Fractures, concussions, miscarriages, and internal injuries also occur although less frequently (Campbell 1998; Dobash and Dobash 1998; Coker et al. 2000; Kyriacou et al. 1999; Stark and Flitcraft 1992). Health and medical research indicates that permanent disfigurement, physical disability, and damage to hearing and vision sometimes occur, that abused women are six to eight times more likely to use health services than non-abused women, and that violence during pregnancy threatens the health of the woman and the fetus (Campbell 1998; Walton-Moss et al. 2005). Women are also likely to experience an ongoing sense of fear, helplessness, entrapment, and loss of self-respect (Dobash and Dobash 2004; Watson and Parsons 2005), as well as other long-term negative consequences to their health and well-being.

Constellation of Abuse

Intimate partner violence is frequently linked to other acts that do not involve physical violence but are controlling, intimidating, and coercive. Such acts do not break bones or cause bruising or bleeding, but may result in fear, intimidation, and damage to self-worth. The impact of coercive and controlling behaviors usually rests upon the fact that physical violence has previously been used and could be used again. This provides a firm foundation upon which to use intimidation in order to maintain authority and control. We have defined this combination of physical violence, injuries, and controlling behavior as the "constellation of abuse" (Dobash and Dobash 1984; Dobash et al. 2000). Our historical and intensive studies confirm the link between violence and other forms of intimation and coercion (Dobash and Dobash 1979; Dobash et al. 1998, 2000), and this link has been corroborated in several population surveys (Tjaden and Thoennes 1998; Walby and Allen 2004; Wilson et al. 1995). A survey by the World Health Organisation (WHO) revealed a direct relationship between violence against intimate partners and various forms of controlling and intimidating acts by men (e.g., restricting her mobility and access to friends and family). While levels of controlling behaviors vary from country to country (from 21 % to 90 %), systematic research from ten countries indicates a strong statistical relationship between "severe" restrictive controls and physical and/or sexual violence (Coker et al. 2000). Violence and other sustained forms of abuse not only result in physical injury and psychological distress but may also result in chronic mental health problems (Coker et al. 2000; Kyriacou et al. 1999).

This constellation of abuse may even extend beyond the "end" of a relationship as some men continue try to control and/or punish the woman for leaving or for beginning a new relationship. The woman may be stalked and/or subjected to further violence, and others (relatives, friends/neighbors, and new partners) may also be subjected to violence or intimidation in an ongoing effort to punish and control. A representative sample survey of 8,000 women in the USA revealed that 81 % of the respondents who reported having been "stalked" by a former partner indicated that they had previously been assaulted by the person who continued to harass them. Additionally, 31 % of these women also reported a previous incident of sexual assault (U.S. Dept. of Justice 1998a).

Type of Relationship: Marriage, Cohabitation, and Dating

The type of intimate relationship has been found to be important in several respects. Research on lethal and nonlethal violence suggests that co-cohabiting relationships are more at risk of lethal and nonlethal violence than marital relationships (Miethe and Regoeczi 2004; Shackleford and Mouzos 2005; Wilson et al. 1995). Additional research expanded to include dating, nonresidential relationships revealed that both cohabiting and dating relationships were at a greater risk of lethal violence than marital relationships (Dawson and Gartner 1998; Dobash et al. 2007; Johnson and Hotton 2003). The preponderance of evidence suggests that cohabitation and serious dating/engaged relationships have a greater risk of lethal and nonlethal IPV than state-sanctioned marriage. It may be that such relationships are more tenuous, involve less commitment, have fewer outside supports in the form of relatives or the state, and these circumstances leave the couple with fewer resources to deal with the inevitable conflicts of domestic life as well as reduce the likelihood of external intervention. Alternatively, any observed difference in the risk of IPV relating to the type of relationship may be associated with the distinct characteristics of those in these different types of relationship. Men and women who cohabit or are in a boy/girlfriend relationship are likely to be younger, poorer, and in other ways categorically different than those who are married (Brownridge and Halli 2002). It may be, however, that the observed differences are related both to the characteristics of the individuals involved and to the sociocultural factors associated with various types of relationships.

Conflict, Nonlethal, and Lethal Violence

Conflict usually precedes violent events, and conflicts (often chronic) in intimate relations characterized by violence include a number of recurring issues associated with daily life including: money, children, housekeeping, sex, fidelity, jealousy, possessiveness, authority, and the continuation of the relationship (Dobash and Dobash 1979; Dobash et al. 2000). Men's sense of entitlement, jealousy, and possessiveness are major issues and may be even more apparent in cases that end in murder (Block and Christakos 1995; Campbell et al. 2007; Dobash et al. 2007; Dobash and Dobash 2011; Polk and Ranson 1991; Serran and Firestone 2004; Wilson and Daly 1998). At the point when women attempt to leave, or terminate a relationship, issues of possessiveness and "ownership" become very apparent, and the combination of these factors appears to contribute to an elevated risk of lethal and nonlethal violence against women. Evidences from several studies of intimate partner murder suggest that at the time of the murder, one-third to one-half of women killed by a partner were either separated from their partner or had indicated their intention to leave the relationship (Browne et al. 1999; Dawson and Gartner 1998; Dobash et al. 2007 Johnson and Hotton 2003; Wilson and Daly 1993). The early stages of separation appear to be the most risky, although some men stalk and kill an ex-partner several years after separation.

Lethal violence in intimate relationships primarily involves men as perpetrators and women as victims, although abused women do sometimes kill their male abuser. In a number of countries, the ratio of male-to-female victims is around one to five, and in some societies there are no reports of women killing male intimate partners (Campbell et al. 2007; Daly and Wilson 1988; Dobash et al. 2007). The homicides of women intimate partners constitute 40–60 % of all murders of women, whereas no more than 10 % of men who are murdered are killed by an intimate partner (Campbell et al. 2007; Dahlberg and Krug 2002). In the USA such murders are one of the leading causes of premature deaths of women and the seventh leading cause of death for African-American women aged 15–45 (U.S. Office of Justice 1998b). When men murder an intimate partner, it usually occurs in the context of sustained, long-term violence and abuse. Research suggests that at least 60 % of all murders of women by intimate partners are associated with the chronic violent abuse by the male perpetrator (Campbell et al. 2007; Dobash et al. 2007; Moracco et al. 1998). A considerable proportion of these cases also involve a history of sexual violence which occurs in about 15 % of the murders of women (Campbell et al. 2007; Dobash et al. 2007; Mathews et al. 2004).

When women murder a male partner, it usually occurs in the context of the man's physical and sexual violence against her and frequently involves self-defense and/or retaliation (Browne 1987).

Collateral Murder Involving Intimate Partner Conflict/Violence

Men not only murder their women partners and ex-partners in the context of intimate partner violence and conflict, they also kill others, such as children or new partners, who might be described as collateral victims (Dobash and Dobash 2012; Langford et al. 1998). Children appear to be the most likely collateral victims, but research suggests that those who attempt to shelter or protect women, usually family and friends, as well as new male partners, are also at risk (Dobash et al. 2007; Dobash and Dobash 2012). The widespread availability of firearms, particularly in the USA, means that police officers attempting to intervene in "domestic disputes" are at considerable risk of injury or death. Men sometimes commit familicide, the murder of the entire family including their woman partner and children, often followed by suicide (Wilson and Daly 1998; Websdale 2010). Such actions are extremely rare among women. Most mass murders (more than two victims) are committed by men in the context of intimate partner conflict/violence, as are the majority of murders followed by suicide (Liem 2010; Websdale 2010).

Male Perpetrators of Intimate Partner Violence and Murder

Childhood and Adulthood

Clinical survey and longitudinal studies suggest that men who have committed violent acts against an intimate woman partner are significantly more likely than those who have not to have experienced adversity in childhood and various problems in their adult life (Browne et al. 1999; Moffit et al. 1998; Schumacher et al. 2001). Witnessing domestic violence, along with physical abuse and disadvantage in childhood, has been linked to the subsequent perpetration of lethal and nonlethal violence against a woman intimate partner (Ehrensaft et al. 2003; Gondolf 2002; Holtzworth-Munroe and Stuart 1994). Nonlethal and lethal violence against an intimate partner has also been associated with problems in adulthood including poor educational achievement, chronic unemployment, a history of arrest, and convictions for violence or other offenses, as well as problems in relationships with others, particularly intimates (Campbell et al. 2003; Dobash et al. 2000; Dutton and Hart 1992). Chronic substance abuse, particularly of alcohol, often features in the adult lives of both abusers and IPMurderers. It is not merely the consumption of alcohol but binge drinking and/or persistent alcohol abuse that are risk factors for violence, severe violence, and lethality (Fals-Stewart 2003; Finney 2004; Moracco et al. 1998; Walton-Moss et al. 2005). With respect to previous offending among those who commit IPMurder, studies have shown that a prior criminal record for any type of offense is a correlate of intimate partner murder (Dawson and Gartner 1998; Dobash and Dobash 2009; Moracco et al. 1998; Grann and Wedin 2002).

Personality

Focusing on personality, some studies have identified distinct characteristics of abusers, while others have found little difference in the characteristics of abusers and the wider population (Dutton and Kerry 1999; Gondolf 2002; Holtzworth-Munroe and Stuart 1994). Focusing only on men who have murdered an intimate partner, the results of several studies suggest that a considerable proportion of men who kill their partners are less likely to exhibit distinct personality problems than abusers and that a considerable proportion differ very little from the wider population (Dobash and Dobash 2009; Echeburua et al. 2003; Grann and Wedin 2002; Weizmann-Henelius et al. nd). While adverse experiences in childhood and problems in adulthood such as alcohol abuse may increase the risk of serious and more injurious forms of violence, clearly they cannot be considered necessary or sufficient conditions either for lethal or nonlethal violence.

While evidence suggests the importance of adversity in the backgrounds of the perpetrator's of IPV and IPMurder, demonstrating that a considerable proportion of these men are similar to those who, for example, murder other males, there is a body of evidence indicating that some men who assault and murder their partners may not be characterized by such backgrounds. Several studies indicate that a reasonable proportion of men who murder their intimate partner have grown up in a stable family, experienced few problems in childhood, were steadily employed, and appeared to be reliable partners and parents (Dobash and Dobash 2009; Dobash et al. 2004; Echeburua et al. 2003; Grann and Wedin 2002; Weizmann-Henelius et al. nd). This evidence is in sharp contrast to the men who murder other men where difficult backgrounds and criminal records are the norm (Dobash et al. 2004). In a UK study of men who murdered their partner, 40 % appeared to have relatively "conventional" backgrounds (Dobash and Dobash 2009). Additional evidence collected after the murder, however, suggests that around one in five of these men had assaulted their partner at least once at some point in the relationship and their orientations to relationships with women, their violence, and the victims of the murder parallel those of men who have experienced adversity in their backgrounds. These men reacted angrily and violently to their partner's attempts to terminate the relationship and were reluctant to express remorse or regret regarding the murder (Dobash and Dobash 2011, nd).

Criminal Justice and Intimate Partner Violence

The CJ system has long been the focus of concern for community and pressure groups attempting to assist women victims of IPV (Dobash and Dobash 1992). Activists and advocates learned that the CJ system was at best reluctant and at worst opposed to intervening in "domestic disputes." This is confirmed by systematic evidence from several countries which revealed that the norm was under enforcement of the laws of assault and that assaults between strangers were more likely than those between family members to result in arrest (Dobash and Dobash 1979). Violence against women in the home was not considered to be a "real crime" and was, instead, treated as a social and personal problem best dealt with by social services and others. Training guidelines, such as those of the International Association of Police Chiefs, suggested that this was a "personal matter" and that arrest should be a "last resort." If the legal system was to be involved, it was through civil remedies such as injunctions and protection orders. This legacy of legal ideas, principles, and practices constituted the family as a private domain where the man was "in charge" and usually immune from prosecution. Even near the end of the twentieth century, this legacy was boldly reflected in homicide "special immunity" statutes in some US states whereby a man who murdered his wife when she was engaged in a sexual act of infidelity was immune from prosecution (Daly and Wilson 1988).

Beginning in the 1960s, the orientations within criminal justice in the US were merged with "psychiatric ideals" with attempts to provide interventions that would "facilitate human helping" in "family fights" and "domestic squabbles" (Dobash and Dobash 1992). Short-term, immediate interventions, such as "mediation," were endorsed. Model programs were created around the ideals of "crisis intervention," and by the 1980s "crisis intervention" and mediation were standard policy and procedure in large police departments throughout the USA (Oppenlander 1982; Lerman 1984). Arrest was a "last resort" to be invoked only in cases of "wanton" or "brutal" assault, and even into the 1980s half of US police departments prohibited arrest in all but the most injurious types of domestic violence (Sherman 1992). Such "model" programs exemplified the approach to domestic violence that was neither seen as a serious criminal problem nor as the responsibility of the criminal justice system. However, subsequent lawsuits and class actions against the police for failure to protect women victims of violence (almost exclusively in the USA) and legislative and administrative mandates at national and local levels brought about changes in such policies.

Police officers in many countries are now given specialized training, often provided by the same pressure groups that lobbied for change. In some cases, protocols mandating enhanced police action are now in place with prescriptive instructions limiting police discretion and requiring actions to protect victims of violence and their children. In several jurisdictions, presumptive arrest of the perpetrator based on "probable cause" is now seen as an appropriate action, and arrest is mandated with such practices sometimes enacted by dedicated domestic violence police units (Iovanni and Miller 2001).

Criminal justice developments have occurred throughout the world. In the USA, the UK, and many other countries, pressure groups have contributed to changes in thinking and practice, particularly concerning the treatment of victims who had previously been deemed outside of the remit of police work. A comprehensive study of European Union countries found significant changes in criminal justice responses (European Commission 2010), and the authors concluded that the concept of "due diligence" to "protect, prosecute, and prevent" is widely accepted as is criminalization of IPV with mediation and conciliation no longer acceptable in most countries. In most, but not all, EU countries, policies and laws associated with domestic violence are usually gender specific, and it is generally acknowledged that women are the usual victims of intimate partner violence.

In some countries there have also been significant developments in the prosecution services, courts, and sentencing. Most importantly, intimate partner violence is now on the agenda and considered an important problem worthy of criminal justice consideration and intervention through prosecution and criminal courts. In some countries dedicated domestic violence courts have been created to facilitate the processing of such cases. New sentencing options have also been developed that mandate offender attendance to innovative feminist-oriented cognitive behavioral interventions that require offenders to deal with and acknowledge their violent acts and take responsibility for them (Dobash et al. 2000; Gondolf 2002). Such programs are an important part of the criminal justice response to intimate partner violence.

It is prudent to ask if such changes have made any difference. Has intimate partner violence been reduced? Are women and their children safer? Are offenders taking responsibility for their actions and changing their behavior? Systematic evaluations of innovations are rare, most have been conducted in the USA and most have focused on the effectiveness of arrest. In the 1980s, a number of studies in the USA attempted to compare the impact of arrest with doing nothing or responding in other ways such as mediation. Initially, arrest appeared to reduce repeat incidents of violence over a 12-month period, and subsequently a pro-arrest strategy was adopted in some jurisdictions (Sherman and Berk 1984). However, subsequent studies conducted in other cities reached conflicting conclusions ranging from studies finding that arrest had "no effect on repeat violence" to those showing a "reduced effect" and to those showing an "increase" in violence. Despite these contradictory findings, policies supporting arrest are now in place in the USA and elsewhere and are deemed to be a useful intervention, although practice has not always followed on from such policies (Ferraro 1989; Jaffe et al. 1993; Zorza and Woods 1994; Logan et al. 2006).

Summary and Conclusions

Violence against women is now recognized as a worldwide problem. While the prevalence of this violence may vary from country to country, it is clear that the problem exists in all those societies where it has been investigated. Violence against women is now recognized as an issue of human rights by the United Nations, UNESCO, the European Union, and other national and international bodies. Where once there was indifference to, or even outright support of, violence by men against women partners, this has now been drastically eroded although not eliminated.

Societal values and institutional responses have changed. It is clear that the efforts of women's groups throughout the world have made a difference in raising the issue and working toward more effective responses to women who have been the victims of intimate partner violence. In the USA, research indicates that there has been a significant reduction in intimate partner violence in those cities that have introduced meaningful interventions for IPV. In addition, national figures show a significant reduction in intimate partner murder.

Recovering from a shaky start, researchers within criminology have, along with those from other disciplines, played an important part in placing this problem on the academic and public agendas, in helping to transform ways of thinking about this problem, and in replacing uninformed speculation with solid evidence. Evidence reveals that violence against women by intimate partners is widespread and serious, but also that it is amenable to positive interventions from the justice system and others that work toward reductions in its frequency and severity and thus toward the well-being of women, their children, and, ultimately, the society at large.

Related Entries

- ► Surveys on Violence Against Women

Recommended Reading and References

Abbott J, Johnson R, Koziol-McLain J, Lowenstein SR (1995) Domestic violence against women: incidence and prevalence in an emergency department population. J Am Med Assoc 273:1763–1767

Amir M (1971) Patterns in forcible rape. Chicago University Press, Chicago

Archer J (2000) Sex differences in aggression between heterosexual partners: a meta-analytic review. Psychol Bull 126:651–680

Backman R, Saltzman LE (1995) Violence against women: estimates from the redesigned survey. Special report of the bureau of justice statistics. U.S. Dept. Justice, Washington D.C

Block CR, Christakos A (1995) Intimate partner homicide in Chicago over 29 years. Crime Delinq 41:496–526

Browne A (1987) When battered women kill. The Free Press, New York, Collier Macmillan

Browne A, Williams KR, Dutton DG (1999) Homicide between intimate partners: a 20-year review. In: Smith MD, Zahn MA (eds) Homicide: a sourcebook of social research. Sage, Thousand Oaks, pp 149–164

Brownridge DA, Halli SS (2002) Understanding male partner violence against cohabiting and married women: an empirical investigation with a synthesized model. J Fam Violence 17:341–361

Campbell JC (1998) Abuse during pregnancy: progress, policy, and potential. Am J Public Health 88:185–187

Campbell JC (1999) Forced sex and intimate partner violence. Violence Against Women 5:1017–1035

Campbell JC, Webster D, Koziol-McLain J, Block CR, Campbell D, Curry MA et al (2003) Risk factors for femicide in abusive relationships: results from a multisite case control study. Am J Public Health 93: 139–152

Campbell JC, Glass N, Sharps PW, Laughon K, Bloom T (2007) Intimate partner homicide: review and implications of research and policy. Trauma Violence Abuse 8:246–269

Cavanagh K, Dobash RE, Dobash RP (2007) The murder of children by fathers in the context of child abuse. Child Abuse Negl 31:731–746

Coker AL, Smith PH, McKeown RE, King MJ (2000) Frequency and correlates of intimate partner violence by type: physical, sexual, and psychological battering. Am J Public Health 90:553–559

Dahlberg LL, Krug EG (2002) Violence – a global health problem. In: Krug EG, Dahlberg LL, Mercy JA, Zwi AB, Lozano R (eds) World report on violence and health. WHO, World Health Organisation, Geneva, pp 1–21

Daly M, Wilson M (1988) Homicide. Aldine De Gruyter, New York

Dawson R, Gartner R (1998) Differences in the characteristics of intimate femicides: the role of relationship state and relationship status. Homicide Stud 2:378–399

Dobash RE, Dobash RP (1977/78) Wives: the 'appropriate' victims of marital violence. Victimology 2:426–442

Dobash RE, Dobash RP (1979) Violence against wives: a case against the patriarchy. Free Press & Open Books, New York

Dobash RE, Dobash RP (1984) The nature and antecedents of violent events. Br J Criminol 24:269–288

Dobash RE, Dobash RP (1992) Women, violence and social change. Routledge, London

Dobash RE, Dobash RP (eds) (1998) Rethinking violence against women. Sage, Thousand Oaks

Dobash RP, Dobash RE (2004) Women's violence to men in intimate relationships: working on a puzzle. Br J Criminol 44:324–349

Dobash RE, Dobash RP (2009) Out of the blue: men who murder an intimate partner. Fem Criminol 4:194–225

Dobash RE, Dobash RP (2011) What were they thinking? Men who murder an intimate partner. Violence Against Women 17:111–134

Dobash RE, Dobash RP (2012) Who died? The murder of collaterals related to intimate conflict. Violence Against Women 18:662–671, Jan

Dobash RP, Dobash RE, Wilson M, Daly M (1992) The myth of sexual symmetry in marital violence. Soc Probl 39(1):71–91

Dobash RP, Dobash RE, Cavanagh K, Lewis R (1998) Separate and intersecting realities: a comparison of men's and women's accounts of violence against women. Violence Against Women 4:383–414

Dobash RP, Dobash RE, Cavanagh K, Lewis R (2000) Changing violent men. Sage, Thousand Oaks

Dobash RP, Dobash RE, Cavanagh K, Lewis R (2004) Not an ordinary killer-just an ordinary guy: when men murder an intimate woman partner. Violence Against Women 10:577–605

Dobash RE, Dobash RP, Cavanagh K, Medina-Ariza JJ (2007) Lethal and non-lethal violence against an intimate partner: comparing male murderers with non-lethal abusers. Violence Against Women Int Interdisc J 13(4):1–27

Dobash RE, Dobash RP (nd) When men murder women. Oxford University Press, New York (forthcoming)

Dutton D, Hart S (1992) Risk markers for family violence in a federally incarcerated population. Int J Law Psychiatry 15:101–112

Dutton D, Kerry G (1999) Modus operandi and personality disorder in incarcerated spousal killers. Int J Law Psychiatry 22:287–299

Echeburua E, Fernandez-Montalvo J, Amor PJ (2003) Psychopathological profile of men convicted of gender violence. A study in the prisons of Spain. J Interpers Violence 18:798–812

Ehrensaft MK, Cohen P, Brown J, Smailes E, Chen H, Johnson JG (2003) Intergenerational transmission of partner violence: a 20 year prospective study. J Consult Clin Psychol 71:741–753

European Commission (2010) Standardise national legislation on violence against women, violence against children & sexual orientation violence: European Commission Directorate for Justice. Daphne, Brussels

Fals-Stewart W (2003) The occurrence of partner physical aggression on days of alcohol consumption: a longitudinal study. J Consult Clin Psychol 71:41–52

Ferraro K (1989) Policing women battering. Soc Probl 36:61–74

Finney A (2004) Alcohol and intimate partner violence: key findings from the research. Home office bulletin 216. Research, Development and Statistics Directorate, London

Gaquin DA (1977/78) Spouse abuse: data from the national crime survey. Victimology 2:632–643

Garcia-Moreno C, Jensen H, Ellsberg M, Heise L, Watt C (2006) Prevalence of intimate partner violence. Findings from the World Health Organisation multi-country study on women's health and domestic violence. Lancet 368:1260–1269

Gondolf EW (2002) Batterer intervention systems: issues, outcomes and recommendations. Sage, Thousand Oaks

Grann M, Wedin M (2002) Risk factors for recidivism among spousal assault and spousal homicide offenders. Psychol Crime Law 8:5–23

Greenfield L, Rand MR, Craven D, Klaus PA, Perkins CA, Ringel C, Warchol G, Maston C, Fox JA (1998) Violence by intimates: analysis of data on crimes by current or former spouses, boyfriends and girlfriends. US Dept. of Justice, Washington, DC

Holtzworth-Munroe A, Stuart GL (1994) Typologies of male batterers: three subtypes and the differences among them. Psychol Bull 11:476–497

Iovanni L, Miller SL (2001) Criminal justice system responses to domestic violence: law enforcement and the courts. In: Renzetti CM, Edelson JL, Bergen RK (eds) Sourcebook on violence against women. Sage, Thousand Oaks, pp 303–327

Jaffe PG, Hastings E, Reitzel D, Austin GW (1993) The impact of police laying charges. In: Hilton NZ (ed) Legal responses to wife assault: current trends and evaluation. Sage, Newbury Park

Johnson H (1996) Dangerous domains: violence against women in Canada. Nelson, Scarborough

Johnson H, Hotton T (2003) Losing control: homicide risk in estranged and intact intimate relationships. Homicide Stud 7:58–84

Johnson H, Sacco VF (1995) Researching violence against women: statistics Canada's national survey. Can J Criminol 37:281–304

Krahe B, Bieneck S, Moller I (2005) Understanding gender and intimate partner violence from an international perspective. Sex Roles 52:807–827

Krug EG, Dahlberg LL, Mercy JA, Zwi AB, Lozano R (2002) World report on violence and health. World Health Organisation, Geneva

Kyriacou D, Anglin D, Taliaferro E, Stone S, Tubb T, Linden J, Muelleman R, Barton E, Kraus JF (1999) Risk factors for injury to women from domestic violence. N Engl J Med 341:1892–1898

Langford L, Isaac N, Kabat S (1998) Homicides related to intimate partner violence in Massachusetts: examining case ascertainment and validity of the shr. Homicide Stud 2:353–377

Lerman LG (1984) Mediation of wife abuse cases: the adverse impact of informal dispute resolution on women. Harv Women's Law J 7:57–113

Liem MCA (2010) Homicide followed by suicide: a review. Aggress Violent Behav 15:247–261

Logan TK, Shannon L, Walker R (2006) Police attitudes toward domestic violence. J Interpers Violence 21:1365–1374

Luckenbill DF (1977) Criminal homicide as a situated transaction. Soc Probl 25:176–186

Merton RK (1938) Social structure and anomie. Am Sociol Rev 3:672–682

Miethe TD, Regoeczi WC (2004) Rethinking homicide. Exploring the structure and process underlying deadly situations. Cambridge University Press, New York

Mirrlees-Black C (1999) Domestic violence: findings from a new British crime survey self completion questionnaire. Home Office, London

Moffit TE, Caspi A, Silva PA (1998) Developmental antecedents of partner abuse: a prospective study. J Abnorm Psychol 107:375–389

Moracco KE, Runyan CW, Butts JD (1998) Femicide in North Carolina, 1991–1993: a statewide study of patterns and precursors. Homicide Stud 2:422–446

Nazroo J (1995) Uncovering gender differences in the use of marital violence: the effect of methodology. Sociology 29(3):475–494

Oppenlander N (1982) Coping or copping out? Criminology 20:449–465

Polk K, Ranson D (1991) The role of gender in intimate homicide. Aust N Z J Criminol 24:15–24

Randall M, Haskall L (1995) Sexual violence in women's lives. Violence Against Women 1:6–31

Russell DEH (1990) Rape in marriage. Indiana University Press, Indianapolis

Schumacher JA, Feldbau-Kohn S, Smith Slep AM, Heyman RE (2001) Risk factors for male-to-female partner physical abuse. Aggress Violent Behav 6:281–352

Schwartz MD (1987) Gender and injury in marital assault. Sociol Focus 20:61–75

Serran G, Firestone P (2004) Intimate partner homicide: a review of male proprietariness and the self-defense theories. Aggress Violent Behav 9:1–15

Shackleford TK, Mouzos J (2005) Partner killing by men in cohabiting and marital relationships: a comparative, cross-national analysis of data from Australian and the United States. J Interpers Violence 20:1310–1324

Sherman LW (1992) Policing domestic violence: experiments and dilemmas. Free Press, New York

Sherman LW, Berk RA (1984) The specific deterrent effects of arrest for domestic violence. Am Sociol Rev 49:261–272

Short JF, Strodtbeck FL (1965) Group process and gang delinquency. University of Chicago Press, Chicago

Stark E, Flitcraft A (1992) Spouse abuse. In: Last RJ, Wallace R (eds) Public health and preventive medicine, 13th edn. Appleton & Lange, Norwalk

Tjaden P, Thoennes N (1998) Prevalence, incidence, and consequences of violence against women: findings from the national violence against women survey. Research in brief (Nat. Instit. Just. & Centre for Disease Control/Prevention). US Dept. of Justice, Washington, DC

Ullman SE, Siegel JM (1993) Victim-offender relationship and sexual assault. Violence Vict 8:121–134

US Dept. Justice (1998) Stalking and domestic violence. The third annual report to congress under the violence against women act. US Dept. Justice, Washington, DC

US Dept. Justice-Office of Justice Programs (1998) Bureau of justice statistics factbook: violence by intimates-analysis of data on crimes by current or former spouses, boyfriends, and girlfriends. U.S. Department of Justice, Bureau of Justice Statistics, Washington, DC

Walby S, Allen J (2004) Domestic violence, sexual assault and stalking: findings form the British crime survey. Home office research study 276. Home Office, Research, Development and Statistics Directorate, London

Walton-Moss BJ, Manganello J, Frye V, Campbell JC (2005) Risk factors for intimate partner violence and associated injury among urban women. J Commun Health 30:377–389

Watson D, Parsons S (2005) Domestic abuse of women and men in Ireland: report of the national study of domestic abuse. Nat Crime Council and Economic and Social Research Institute, Dublin, pp 1–224

Websdale N (2010) Familicidal hearts: the emotional styles of 211 killers. Oxford University Press, New York

Weizmann-Henelius G, Gronroos M, Putkonen H, Eronen M, Lindberg N et al (nd) Gender specific risk factors for intimate partner homicide—a nationwide register-based study. J Interpers Violence (forthcoming)

Wilson M, Daly M (1993) Spousal homicide risk and estrangement. Violence Vict 8:3–16

Wilson M, Daly M (1998) Lethal and nonlethal violence against wives and the evolutionary psychology of male sexual proprietariness. In: Dobash RE, Dobash RP (eds) Rethinking violence against women. Sage, Thousand Oaks, pp 199–230

Wilson M, Johnson H, Daly M (1995) Lethal and nonlethal violence against wives. Can J Criminol 37:331–362

Wolfgang M (1958) Patterns in criminal homicide. University of Pennsylvania Press, Philadelphia

Worrall A, Pease K (1986) Personal crime against women: evidence from the 1982 British crime survey. Howard J 25:118–124

Zorza J, Woods L (1994) Mandatory arrest: problems and possibilities. National Center on Women and Family Law, New York

Double-Blind Administration

► Eyewitness Research

Dowry Deaths

► Lawful Killings

Drug Abuse and Alcohol Dependence Among Inmates

Javier A. Cepeda[1] and Sandra A. Springer[2]
[1]Center for Interdisciplinary Research on AIDS and Department of Epidemiology and Public Health, Yale School of Public Health,
Yale University, New Haven, CT, USA
[2]Center for Interdisciplinary Research on AIDS and Yale School of Medicine, Yale University, New Haven, CT, USA

Synonyms

AIDS; Alcohol dependence; Buprenorphine; CJS; Criminal justice system; Cocaine; Extended-release Naltrexone; Hepatitis C; HIV; Human Immunodeficiency Syndrome; Incarceration; Jail; MAT; Medication assisted therapy; Mental illness; Methadone; Methamphetamine; Naltrexone; Opioid dependence; Prison; Reentry; Release; STDs; TB; Tuberculosis; Vivitrol

Overview

The massive increase in the number of individuals within the United States (USA) criminal justice system (CJS) in the past 30 years can be largely attributed to the "War on Drugs" campaign with drug-related arrests increasing more than fivefold from 1970 to 2005. Many of the incarcerated individuals have a history of substance use disorders (SUDs), and have reported a use of drugs within the past month prior to arrest. According to the US Bureau of Justice Statistics (BJS), in 2002, the prevalence of offenders who used drugs more than once a week for at least 1 month prior to incarceration exceeded 60 % (Karberg and James 2005) and approximately 50 % of those incarcerated at the federal, state, and local levels met DSM-IV criteria for drug abuse or dependence.

The large burden of drug and alcohol dependence within the US CJS complicates treatment programs since drug and alcohol dependency is highly associated with comorbid psychiatric disorders and chronic infectious diseases. Effective pharmacotherapy for opioid and alcohol dependence exists in the community; however, access to treatment in the US CJS remains inadequate. Less than 10 % of local jail inmates received treatment (Karberg and James 2005).

The public health implications of SUDs treatment for offenders are numerous. Incarceration provides an opportunity to initiate or continue treatment for drug and alcohol dependency among inmates, whereupon both the individual and the community can be greatly benefited upon release. In addition to treatment, strong linkages should be established to ensure continuity of care. Failure to link services for offenders upon release to the community may increase recidivism, relapse to substance abuse, and increase transmission of sexually transmitted diseases (STDs) including human immunodeficiency virus (HIV), and viral hepatitis B and C to the uninfected.

This entry will concentrate mostly on the epidemiology, treatment, and public health implications of opioid and alcohol dependence among inmates in the US CJS. While abuse of others drugs, such as cocaine and methamphetamine, is highly prevalent among offenders, much of the clinical and epidemiological literature on incarcerated persons has been limited to research involving dependency on alcohol and opioids. Effective medically assisted therapy (MAT) for cocaine, crack, and methamphetamine dependence is still lacking, and to date, only behavioral interventions have been implemented to treat dependence for these substances. Conversely, there are FDA-approved pharmacotherapies and behavioral interventions for opioid- and alcohol-dependent persons within the community, and some of these have been successfully implemented in some incarcerated settings across the USA. This entry will discuss the strengths and limitations of various evidence-based practices of SUDs treatment administered in incarcerated settings.

Definition of Substance Dependency

The Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) classifies substance

Drug Abuse and Alcohol Dependence Among Inmates, Table 1 DSM-IV criteria for substance dependence and substance abuse. In using the DSM-IV criteria, one should specify whether substance dependence is with physiologic dependence (i.e., there is evidence of tolerance or withdrawal) or without physiologic dependence (i.e., no evidence of tolerance or withdrawal)

Dependence (3 or more in a 12-month period)	Abuse (1 or more in a 12-month period) Symptoms must never have met criteria for substance dependence for this class of substance
Tolerance (marked increase in amount; marked decrease in effect)	Recurrent use resulting in failure to fulfill major role obligation at work, home, or school
Characteristic withdrawal symptoms; substance taken to relieve withdrawal	Recurrent use in physically hazardous situations
Substance taken in larger amount and for longer period than intended	Recurrent substance related legal problems
Persistent desire or repeated unsuccessful attempt to quit	Continued use despite persistent or recurrent social or interpersonal problems caused or exacerbated by substance
Much time/activity to obtain, use, recover	
Important social, occupational, or recreational activities given up or reduced	
Use continues despite knowledge of adverse consequences (e.g., failure to fulfill role obligation, use when physically hazardous)	

dependency as displaying three out of the seven of the listed criteria in Table 1 in a span of 12 months. Substance abuse, as defined by DSM-IV, is demonstration of at least one of the criteria in Table 1 in a span of 12 months.

Epidemiology and Socio-demographic Profile of Drug and Alcohol Abuse and Dependence in the Criminal Justice System

Prevalence of drug abuse among offenders in the US CJS remains high at the county, state, and federal level. In 2005 alone, there were 1.65 million drug arrests. Among all US county jails, nearly 60 % of inmates reported a history of drug dependency at some point in their life and about half of all US inmates met DSM-IV criteria for drug dependency (Mumola and Karberg 2006).

The majority of those incarcerated are men; however, the incarceration rate for women has been increasing since the early 1990s. From 1990 to 1998, the number of women behind bars jumped 71 %. This escalation of incarcerated women was mostly attributed to a surge in the arrest of female drug users. In 2002, about 61 % of incarcerated women satisfied DSM-IV criteria for drug dependency, versus 54 % of incarcerated men in local jails (Karberg and James 2005).

Minorities, especially blacks and Latinos, are disproportionately represented in the criminal justice system. While only making up about a quarter of the US population, more than 60 % of the US criminal justice population is black or Latino. Conversely, white inmates have been reported as having a relatively higher prevalence of substance (alcohol and/or drug) abuse or dependence (78 %) compared to blacks (64 %) and Hispanics (59 %) (Karberg and James 2005).

Opioid Abuse and Dependence

Abuse and dependence to opioids continues to plague the USA. There are approximately 900,000 opioid-dependent persons within the USA. In 2004, there were approximately 1.2 million state prisoners, of whom 23.4 % had ever used heroin or opiates in their lifetimes, and 8.2 % of convicted inmates reported using heroin or other opiates 1 month prior to arrest (Mumola and Karberg 2006). Approximately 11 % of all male inmates, and 20 % of all female inmates reported using opioids daily in the 6 months prior to arrest. Since the 2000s, nonmedical abuse of prescription opioids has led to an upsurge in number of arrests. A nationwide survey from years 2002 to 2004 reported that 30 % of arrestees had used prescription opioids for nonmedical purposes.

Stimulant Abuse and Dependence

The prevalence of abuse of stimulants (cocaine/crack and methamphetamines) among inmates in state and federal prisons exceeds that of marijuana, opioids, hallucinogens, and other illicit depressants combined. In 2004, approximately one half of all federal and state prisoners with a history of drug use reported using cocaine and/or crack (Mumola and Karberg 2006). Additionally, methamphetamine (METH) abuse is a growing problem in the USA. From 1997 to 2004, use of METH was the only drug that increased across all measures collected in a nationwide survey (Mumola and Karberg 2006). There are stark differences in METH use by race/ethnicity. At the federal level, the prevalence of METH abuse among white, Hispanic, and blacks is 29 %, 5 %, and 1 %, respectively (Mumola and Karberg 2006). METH use also differs by gender. Among state inmates, 17 % of women used METH in the month prior to arrest, as opposed to 10 % of men (Mumola and Karberg 2006). Furthermore, a nationwide survey of a sample of local police officers revealed that METH was overwhelmingly their drug of concern compared to cocaine, marijuana, and heroin (48 % vs. 22 %, 22 %, and 3 % respectively). Unlike treatment for opioids and alcohol, there is no effective FDA-approved pharmacotherapeutic treatment yet for cocaine and METH dependency. This is particularly problematic, given the increasing abuse of methamphetamines and high prevalence of cocaine abuse among inmates in state and federal prisons.

Alcohol Abuse and Dependence

Alcohol use disorders (AUDs) encompass a wide variety of conditions including abuse, hazardous drinking, binge drinking, and dependence. Approximately 4 % of the US population meets criteria for alcohol dependence (Grant et al. 2004), while the prevalence of alcohol dependence in state jails, however, is roughly tenfold higher than in the general population (Mumola and Karberg 2006). In 2002, nearly 50 % of all inmates met criteria for alcohol abuse or dependence (Mumola and Karberg 2006). The prevalence of alcohol abuse or dependence among female inmates was 39 %, while half of all men met the criteria for dependence or abuse (Mumola and Karberg 2006). Whites were relatively more likely to abuse alcohol or be dependent (58.6 %), compared to blacks (42.7 %) and Hispanic inmates (41.8 %).

Management and Treatment of Opioid Dependence Within the Criminal Justice System

Despite the proven effectiveness of MAT to treat substance dependence in the community, long-term implementation is still rare across US incarcerated settings. The two most commonly administered pharmacotherapeutic options to manage and treat opioid dependence are opioid substitution therapy (OST) in the form of methadone and buprenorphine. The other options are antagonists at the mu-receptor in two formulations: once daily oral naltrexone (ReVia®), or the once monthly extended-release depot formulation of naltrexone (Vivitrol). See Table 2 for a summary of available MATs for the treatment of opioid dependence.

Methadone

Methadone is a full opioid mu-receptor agonist that produces similar effects to morphine and heroin. In terms of public health benefits, methadone has been associated with a reduction in risky injection practices (thereby reducing transmission of blood-borne viruses), improvement in overall health, and reduction in mortality. Methadone maintenance therapy (MMT) is also regarded as extremely cost-effective, with more than 50 % of the benefits affecting individuals who do not use drugs. In spite of the evidence demonstrating the advantages of MMT, it still remains difficult to obtain for the majority of heroin users. Access to methadone still remains a challenge for the majority of opioid-dependent persons in the USA for several reasons including long waiting lists and because some state Medicaid programs do not cover the cost of MMT.

Drug Abuse and Alcohol Dependence Among Inmates, Table 2 Medically assisted therapies for treatment of opioid or alcohol dependence

MAT	Mechanism	Form and frequency of administration	Advantages	Disadvantages
Methadone	Full opioid agonist	Liquid or tablet, daily	Inexpensive, no need to experience opioid withdrawal symptoms prior to administration, prevents relapse	Strict federal regulations, accessibility, diversion to illicit drug markets, tolerance, side effects
Buprenorphine	Partial opioid agonist, partial opioid antagonist	Sublingual tablet, film, daily or alternate day	Lower risk of diversion, administered in office setting (less stigma), prevents relapse	More expensive than methadone, patient must experience minimal withdrawal before administration, requires 8-h training to obtain ability to prescribe
Extended-release naltrexone	Full opioid antagonist	Once monthly intramuscular injection	Adherence advantage, no risk of dependence, no risk of overdose, FDA-approved for opioid and alcohol dependence treatment, reduces cravings, no special licensing or special training	Injection, expensive, 5–7 days of opioid abstinence needed prior to treatment for opioid dependence
Naltrexone	Full opioid antagonist	Tablet, daily or alternate day	No risk of dependence or overdose, reduces cravings, very effective for treatment of alcohol dependence	5–7 days of opioid abstinence needed, not found to be effective for preventing relapse for opioid dependence due to low adherence to oral form
Disulfiram	Inhibits acetaldehyde dehydrogenases involved in metabolizing alcohol	Tablet, daily	No risk of tolerance, possible use in treatment of cocaine dependence	Questionable efficacy, low compliance
Acamprosate	Blocks N-methyl D-aspartate receptors	Tablet, 2 tablets three times daily	Reduces relapse, can be used with opioid agonists for opioid dependence	Poor adherence, may only be effective with support groups

DSM-IV criteria for sub- stance dependence and substance abuse. In using the DSM-IV criteria, one should specify whether substance dependence is with physiologic dependence (i.e., there is evidence of tolerance or withdrawal) or without physio- logic dependence (i.e., no evidence of tolerance or withdrawal)

Authorities in the CJS still have strong reservations about offering methadone in prisons or jails. The predominating perception among criminal justice officials was that methadone itself “is replacing one addiction with another” (McMillan and Lapham 2005). Other reasons include concerns regarding cost, diversion, and overdose. Only a select few of incarcerated settings have set up the infrastructure necessary to implement a methadone maintenance treatment (MMT) program. Several research studies were conducted to investigate the effectiveness of the MMT program at Rikers Island (New York City) and found a decrease in risky injection practices 6 months post-release. Moreover, the implementation of a directly observed therapy approach of administration of methadone in the jail greatly alleviated correctional officials’ concern about possible diversion of methadone for illicit purposes (Tomasino et al. 2001). Furthermore, an RCT comparing (1) methadone *after* release, (2) methadone started *in* prison and transferred after release, and (3) counseling alone among opioid-dependent prisoners in Baltimore found that individuals who received only counseling were approximately seven times more likely to relapse to opioids and cocaine 12 months post-release to the community, when compared to individuals who received both counseling and had initiated methadone in prison (Kinlock et al. 2009). Overall, both of these studies lend evidence that initiation of MMT is not only feasible within correctional institutions in the USA, but also that MMT has broad implications in reducing risky and illicit drug-using behavior when prisoners return to the community.

Buprenorphine

In 2002, the FDA approved ***buprenorphine (BPN)**, a partial mu-receptor agonist and antagonist* for the treatment of opioid dependence. Providers need to take an 8-h course on BPN induction and maintenance practices to receive the necessary additional DEA license authorization to prescribe to patients. Since BPN can be administered in an office setting, there is less stigma associated with it compared to receiving daily doses of methadone at a federally sanctioned clinic. The dose response profile of BPN plateaus at a 32-mg dose (no additional effect for doses higher than 32 mg). Depending on the patient, BPN can be administered every 2–3 days, allowing for a more flexible maintenance regimen.

Due to its partial opioid agonist activity at the mu-receptor, BPN generally stimulates milder euphoric effects compared to methadone. BPN is most often administered sublingually, and the variant that includes naloxone (commercially known as Suboxone®) is thought to prevent misuse via injection of BPN. Recently, BPN has been introduced as a sublingual film that is more efficient in absorption than the tablet form. Furthermore, the chances of diversion are thought to be reduced due to a 10-digit bar code on the packaging that can be traced back to the patient should it be found in the illegal drug market and as a means to avoid accidental ingestion among minors.

Due to its relatively new arrival, research investigating the effectiveness of BPN in prison settings is still limited. Investigators conducted a RCT comparing BPN maintenance to MMT. The authors found that patients on BPN were more likely to attend community treatment centers compared to those randomized to receive methadone after release. Ninety-three (93) percent of the patients on BPN said that they intended to continue treatment after being released, while only 44 % of those on methadone expressed the same sentiment. While in jail, patients on BPN were significantly less likely to voluntarily drop out of receiving treatment than those on methadone. Despite these findings demonstrating more acceptability of BPN compared to methadone while the patients were incarcerated, there were no significant differences in terms of self-reported relapse to opioid use, rearrests, or reincarceration between the two groups. BPN was shown to reduce recidivism among inmates who initiated BPN while incarcerated compared to those who received no form of OST (Levasseur et al. 2002).

Implementation of a BPN maintenance program in a Baltimore prison highlighted some challenges distinct from MMT (Kinlock et al. 2010). The authors highlight diversion of BPN as the most significant challenge compared to methadone. Nurses had to wait 10 min per patient receiving a sublingual dose of BPN to ensure that the tablet had been completely dissolved. Methadone, on the other hand, is taken in liquid form, making diversion more difficult if the inmate is being directly observed. These findings are consistent with those from the Rikers Island study, with the authors reporting that approximately 10 % of the patients on BPN attempted to divert the drug, while only 1 % of the patients on methadone attempted diversion. Fears of diversion of BPN could be potentially mitigated if the sublingual film is used rather than the tablet, due to the former dissolving more rapidly than the latter. The FDA in 2012 ordered that tablet formulation of BPN would be replaced with the sublingual film in order to mitigate concerns for diversion and decrease deaths associated with inadvertent use of the sublingual tables by children.

Oral and Extended-Release Naltrexone

Buprenorphine and methadone are classified as types of opioid substitution therapy due to their affinity for the mu-receptor; however,

naltrexone, a full opioid antagonist, completely blocks the mu-receptor. Naltrexone can be administered orally daily, or once a month by intramuscular injection (extended-release naltrexone). The oral form, ReVia®, approved in 1984 for the treatment of opioid dependence, has NOT been shown to be effective in relapse prevention or in recidivism due to compliance issues mainly. In October of 2010, the FDA approved the use of extended-release naltrexone (XR-NTX) for the treatment of opioid dependence after results from a double-blind multicenter RCT performed in Russia. In this study, 126 opioid-dependent individuals were randomized to receive the XR-NTX compared to 124 subjects who received placebo. The XR-NTX group was 1.58 times more likely to be abstinent from opioids at end of the 24-week study period compared to the control group. Of note, 40 % of the subjects who received XR-NTX had HIV disease, and approximately 90 % had Hepatitis C infection without any serious liver function abnormalities. The extended-release version of naltrexone does provide a possibility of less adherence concerns, given its once monthly dosing as well as possible fewer side effects compared to the oral form of naltrexone as well as methadone and BP-N. Furthermore, because XR-NTX is an antagonist at the mu opioid receptor, there are no concerns for adverse events such as respiratory depression or overdose from the treatment as can be seen with methadone. Also due to its antagonist properties, there is not a concern for diversion as has been seen with buprenorphine. Therefore, XR-NTX may be a viable option for the CJS where such concerns predominate.

Before initiating ***naltrexone*** maintenance therapy, the patient must be free of opioids within the last 5–7 days. This is perhaps the most challenging aspect to naltrexone initiation, since many patients are fearful of experiencing any withdrawal symptoms and can often not resist the urge to continue using opioids in this period. While sudden abruption of not taking naltrexone does not produce withdrawal symptoms, the patient must be on oral naltrexone for at least 30 days to demonstrate efficacy in treatment outcomes with this form of naltrexone; however, it is unclear how long one should be maintained on XR-NTX at this time to demonstrate efficacy.

Thus far, only a few studies have examined the feasibility and effectiveness of oral or XR-NTX maintenance therapy among incarcerated populations. In one double-blind RCT, investigators compared oral naltrexone to another opioid antagonist called cyclazocine among 40 inmates with prior history of opioid addiction (Brahen et al. 1977). At the end of the study, inmates who were on naltrexone reported fewer side effects after induction with a placebo and very low toxicity. In fact, three patients on cyclazocine dropped out of the study due to complications arising from the treatment itself. None of the patients randomized to naltrexone dropped out of the study for these reasons (Brahen et al. 1977). Indeed, a recent meta-analysis reported that the evidence claiming oral naltrexone as a superior treatment to other therapies for preventing relapse is conflicting (Minozzi et al. 2011).

There has been only one publication, thus far, on the effectiveness of XR-NTX among incarcerated populations. In this Norwegian study, 46 heroin-addicted inmates were randomized to either XR-NTX or methadone (Lobmaier et al. 2010). At 6 months post-release, the investigators found similar reductions in frequency of heroin use and criminal behavior between the two groups, suggesting that XR-NTX may be as equally effective as methadone at treatment of opioid relapse and recidivism. RCTs evaluating the effect of XR-NTX among opioid-dependent incarcerated populations are ongoing at the present time.

Behavioral Interventions for Offenders with Drug Dependence

Oftentimes pharmacotherapy is integrated with psychosocial support, either in the form of SUDs 12-step counseling, cognitive behavioral therapy, case management, or enrollment in therapeutic communities. Prison-based SUDs treatment programs have been shown to reduce recidivism and drug use. Specifically, *cognitive behavioral therapy (CBT)* interventions for offenders focusing on reducing criminal "thinking" can lead to criminal behavior. The objectives of CBT are to prevent relapse, to recognize situations that can often lead to drug use or criminal activities, to assist patients on how to effectively deal with these situations, to enhance social support networks, and to advance feelings of self-efficacy.

Contingency management (CM) is another type of behavioral intervention employed among drug abusing offenders that aims to promote abstinence through positive reinforcement. The theory supporting CM as an effective strategy is that rewarding positive behavior will eventually replace punishable behavior. Findings from two meta-analyses lend evidence to this theory (Griffith et al. 2000; Lussier et al. 2006). Modeled after community-based therapeutic communities, *in-prison therapeutic communities (TCs)* are a type of intervention that move drug-using offenders into separate facilities from the rest of the incarcerated population. This kind of residential treatment allows for intensive interventional and psychosocial support. The two TC modalities focus on the 12-step self-help model and the relapse prevention model. An evaluation of the Amity prison TC revealed statistically significant better outcomes in terms of reincarceration among inmates enrolled in TC versus inmates who dropped out of the program versus inmates randomized to a control group. Unlike OST, TCs have gained wide acceptance in the prison system. Since 1994, many states have adopted the TC model and received federal funding to expand residential substance abuse treatment (RSAT) programs in prisons and jails.

Management and Treatment of Alcohol Dependence Within the Criminal Justice System

The prevalence of alcohol use disorders (AUDs) among incarcerated individuals is staggering, as previously discussed. Behavioral interventions have typically been viewed as the bastion for alcohol abuse treatment; however, these interventions have only been marginally effective in treating alcohol-related disorders (Springer et al. 2011). Specifically, behavioral treatment can include motivational enhancement therapy (MET), cognitive behavioral therapy, and self-help (e.g., 12-step) programs. MET, although less structured than CBT, has been shown to reduce cravings of alcohol with integrated pharmacotherapy using naltrexone (Monterosso et al. 2001). Currently, there are four FDA-approved pharmacotherapies available to treat alcohol dependence: naltrexone, extended-release naltrexone, acamprosate, and disulfiram. (These are summarized in Table 1).

Naltrexone

After being approved for opioid treatment by the FDA in 1984, oral naltrexone was approved 10 years later for the treatment of alcohol dependence after demonstrating a reduction in the pleasure produced from consumption of alcohol. Extended-release naltrexone (Vivitrol®) was FDA-approved for the treatment of alcohol dependence in 2006. Ethanol is believed to activate receptors in the opioid response system, which in turn activates numerous neurotransmitters, including dopamine. Since the efficacy of oral naltrexone is strongly contingent upon adherence, its usefulness as a treatment option for heavy drinkers has been questioned extensively. In many cases, compliance is the limiting factor in efficacy of oral naltrexone, while monthly injectable XR-NTX may overcome these challenges, due to fewer side effects and a better adherence profile (Mannelli et al. 2007). Although there has not yet been a head-to-head comparison of oral NTX and XR-NTX, evidence suggests that XR-NTX is highly efficacious for treatment of alcohol dependence, as well as compared to counseling or combined with counseling (O'Malley et al. 2007).

Disulfiram

Physiologically, disulfiram functions completely differently than naltrexone. Disulfiram blocks the oxidation of alcohol that results in increased levels of acetaldehyde. When a patient is administered disulfiram, yet continues to drink alcohol, the buildup of acetaldehyde can produce several negative effects such as nausea, vomiting, chest pain, and increased heart rate shortly after consumption. Similar to oral naltrexone, adherence to disulfiram is problematic unless the patients are highly motivated. Most noncontrolled trials among offenders demonstrated a beneficial effect of disulfiram, while one study did not (Bourne et al. 1966). One study among those on probation in Atlanta found 50 % of the subjects were abstinent from alcohol for over 3 months after daily observation of disulfiram administration (Bourne et al. 1966). Additionally, another study in New York Later studies, however, found that disulfiram was of no more benefit than placebo. One of these studies compared disulfiram to group therapy among alcoholic offenders in New Orleans and reported only a modest improvement in abstinence between the two groups.

Acamprosate

While both oral naltrexone and disulfiram have been available for decades, ***acamprosate*** has been available only since 2004. The biological mechanism of acamprosate's activity is still not fully understood, yet it has been shown in numerous randomized clinical trials to maintain abstinence among alcohol dependents (who have already been detoxified from alcohol) modestly better than those who were administered placebo. In summary, as evidenced by findings from a large multicenter clinical trial known as the COMBINE study, oral naltrexone coupled with psychosocial support is still considered the most effective pharmacotherapeutic treatment for alcohol dependence (Anton et al. 2006; Garbutt et al. 2005).

Contraindications and Severe Adverse Events of Naltrexone, Disulfiram, and Acamprosate

According to the manufacturer's instructions, naltrexone should only be administered at the recommended doses. Liver injury has been documented when doses are excessive. *Naltrexone* should NOT be administered to patients with acute hepatitis or liver injury and is contraindicated in persons with childs-pugh class C cirrhosis. Naltrexone is also contraindicated in patients undergoing acute opioid withdrawal.

Disulfiram is contraindicated in patients receiving metronidazole, paraldehyde, alcohol, or alcohol-containing preparations (such as cough syrups). Patients with severe heart disease, coronary occlusion, and psychoses should avoid disulfiram. Disulfiram should never be administered to a patient currently under the influence of alcohol or without his/her knowledge. Several cases of hepatitis and liver failure have been documented in patients administered disulfiram.

Acamprosate is contraindicated in patients who have previously demonstrated hypersensitivity to acamprosate and in patients with severe kidney impairment. Similar to naltrexone, there has not been a rigorous evaluation on the safety of acamprosate in pregnant women.

The safety of naltrexone, disulfiram, and acamprosate has not been rigorously evaluated in pregnant women and is generally considered contraindicated in pregnancy. If the decision is made to use these drugs during pregnancy, careful consideration of the benefits and risks of administering these medications during pregnancy is necessary.

Medical and Psychiatric Comorbidities Associated with Dependency to Alcohol and Drug Abuse

Incarceration provides not only an opportunity to curb drug and alcohol use disorders, but also to address other underlying comorbidities. Infectious diseases, such as human immunodeficiency virus (HIV), tuberculosis (TB), viral hepatitis B (HBV) & C (HCV), and sexually transmitted infections (STIs) such as syphilis, gonorrhea, and chlamydia, are all strongly associated with drug and alcohol abuse. The burden of these diseases among individuals who have passed through the criminal justice system is disproportionately high, compared to those who have not. In 2006, it was estimated that one out of seven individuals living with HIV in the USA had been through the CJS or 16 % of the incarcerated population (Maruschak 2008; Spaulding et al. 2009). For HCV and TB, more than a third of all those infected in the USA were either incarcerated or just released from a correctional facility.

Human Immunodeficiency Virus (HIV)

The rate of HIV and AIDS among those incarcerated is approximately three times and four times higher respectively than the US general population (Maruschak 2008, 2009; Spaulding et al. 2009). The high rates of HIV among incarcerated persons are in part due to the overlap of increased frequency of alcohol and drug use among persons who are incarcerated and the high-risk injection-drug-using behaviors and unprotected sexual behaviors that occur under the influence of these substances. Concurrent opioid and alcohol use is significantly associated with higher morbidity and mortality, decreased adherence to highly active antiretroviral therapy (HAART), and increased transmission of HIV among HIV-positive CJS populations. Among those with HIV disease who have comorbid alcohol or opioid dependence, adherence to HAART is low (Springer and Altice 2005). Thus, treatment of their chemical dependence should be prioritized among HIV-positive individuals in order to maintain adherence to HAART. In a small study among HIV-positive opioid-dependent incarcerated offenders in Connecticut, administration of both BPN and HAART resulted in low opioid positivity in urinalysis and high HIV viral load suppression at 3 months post-release (Springer et al. 2010; Springer et al. 2012). Currently, evaluation of administration of HIV treatment and MAT among incarcerated offenders before release is an active area of investigation.

Hepatitis C Virus (HCV)

With prevalence rates ranging from 16 % to 49 %, HCV is the most common blood-borne disease in the CJS. Injection-drug use is a strong risk factor for HCV acquisition. Most individuals with chronic HCV do not develop any symptoms; however, alcohol abuse can greatly accelerate end-stage liver disease (ESLD) and death (Springer et al. 2011). HCV treatment typically has been expensive and logistically difficult to implement in correctional settings due to its long duration. Newer and more effective treatments such as boceprevir and telaprevir that have shown more sustained virologic response when compared to standard treatment of pegylated interferon-ribavirin offer potential improved treatment opportunities for the incarcerated population as well. The CJS should consider evaluating the cost-effectiveness of HCV treatment, given the immense disease burden among inmates and potential risk to the community upon release.

HIV/HCV Coinfection

The prevalence of HIV/HCV coinfection is estimated to be between 50 % and 90 % among injection-drug users (IDUs). Estimates of prevalence of IDU among incarcerated adults range from 3 % to 28 %, and high-risk behaviors, such as sharing syringes, persist behind bars (Hammett 2006). Despite the potential for HIV and HCV acquisition in incarcerated settings, needle exchange programs are prohibited in American prisons and jails. Given the high burden of comorbid HIV and HCV disease in prisoners and the notion that ESLD is now the number one cause of death in HIV-positive persons, it behooves the CJS to consider treatment of HCV as well as treatment of drug and alcohol dependency to improve adherence to HCV treatment as a measure to decrease morbidity and mortality.

Tuberculosis (TB)

Transmission of TB is associated with unstable living conditions, poverty, and history of injection-drug use and is facilitated in crowded environments such as prisons and jails (Altice et al. 2010). The resurgence of TB in the last three decades can be largely attributed to the emergence of HIV (Altice et al. 2010). Susceptibility to TB is particularly high among those with high HIV viral loads due to suppression of the immune system from the virus (Friedland 2010). It has been estimated that in any given year, 40 % of all TB-positive individuals in the USA have circulated through the CJS (Hammett et al. 2002).

At intake, all inmates are screened for TB. Given the high prevalence of drug and alcohol users who are TB positive in correctional settings, treatment of TB is imperative. First-line treatments for latent TB infection, such as rifampin and isoniazid, are highly effective (95 % cure rate) with successful adherence (Friedland 2010). Incarceration provides an opportunity where patients can be treated by directly observed therapy (DOT) after patients have detoxed from alcohol or drug use. Chronic alcohol use in conjunction with TB treatment can cause severe hepatotoxicity. Moreover, once released into the community without proper linkage to health care services, resumption of alcohol or drug can disrupt compliance to treatment regimens and promote drug-resistant strains of TB (Friedland 2010). Therefore, it is imperative that alcohol treatment be instituted in concert with effective TB treatment to prevent adherence interruptions and liver toxicity.

Sexually Transmitted Diseases (STDs)

Previous research has shown that drugs and alcohol are associated with risky sexual behaviors that increase the likelihood of transmission of HIV and other STDs. HIV and other STDs disproportionately affect women within the CJS. Women are more likely to be incarcerated for crimes related to drugs or sex work, two behaviors that are strong correlates of STD transmission. Additionally, incarcerated adult women have high rates of chlamydia (6.3 %), syphilis (7.5 %), while juvenile women in detention facilities have high rates of gonorrhea (5.7 %). Screening and treatment of STDs in incarcerated settings should be administered in addition to treatment for opioid and alcohol dependency. Such interventions could prevent jails and prison from becoming incubators of STDs, given the strong association between drug/alcohol dependency, risky sexual behavior, and incarceration.

Mental Illness

In addition to the large burden of infectious disease comorbidities among drug offenders, substantial proportions also suffer from severe mental illness. Schizophrenia, major depression, and bipolar disorder are all highly associated with substance abuse. In 1998, it was reported that nearly a quarter of federal inmates, and over a third of jail inmates with psychiatric disorders, had a history of alcohol dependence (Ditton 1999). Upon release, individuals with comorbid mental illnesses often fare poorly in both mental illness treatment and substance abuse treatment programs due to the lack of integrated care in addressing both these ailments in one health center. Thus, untreated comorbid psychiatric disorders place released drug offenders at high risk for criminal behavior, parole violation, and consequent reincarceration.

Linkage to the Community Post-Release

The majority of CJS populations will eventually be released to the community. One of the greatest challenges facing recently released offenders is relapse to drug or alcohol use. Nearly 90 % of individuals with opioid dependence not on drug treatment will relapse within 12 months after being released (Kinlock et al. 2002). Moreover, overdose is the number one cause of death among all released incarcerated persons in Washington state (Binswanger et al. 2007). Additionally, relapse is correlated with criminal behavior and reincarceration.

During this critical transition period, a robust framework must be in place to coordinate access to drug treatment, mental health services, HIV care, or other medical services once an inmate returns to the community. Case management services are seen as an effective linkage between the criminal justice system and the community. Various case management models exist; however, several share core principles such as client advocacy, monitoring, referral, and planning. Several case management interventions have been able to successfully link substance users to treatment upon release (Lincoln et al. 2006), but participation in case management services offered in correctional settings still remains a significant barrier to its effectiveness. While in some instances inmates may be mandated to utilize these services as conditions of their parole,

Drug Abuse and Alcohol Dependence Among Inmates, Fig. 1 Schematic of reasons to initiate MAT to incarcerated persons with comorbid drug and alcohol dependence prior to release to the community

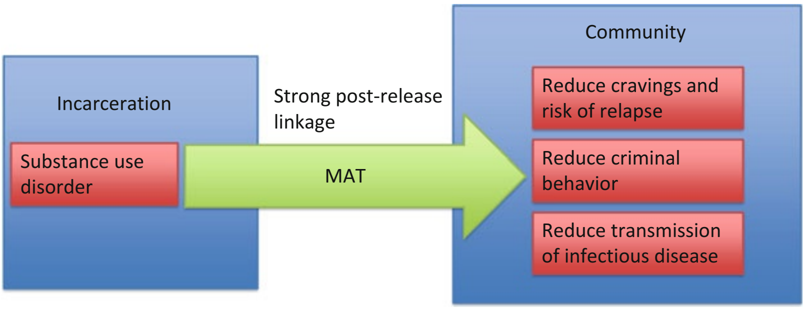
Incarceration
Substance use disorder
Strong post-release linkage
MAT
Community
Reduce cravings and risk of relapse
Reduce criminal behavior
Reduce transmission of infectious disease

those who volunteer to enroll in case management services often build stronger rapport with their case managers. Case managers are not viewed as representatives of the CJS; thus, clients are more willing to "open up" to them about illicit behavior. By establishing trust with their clients, case managers can advocate more efficiently on behalf of them and ensure greater retention with their services. Overall, it is not conclusive whether case management alone is sufficient to improve the health or curb drug use of released prisoners.

Future Directions

The "revolving door," in reference to repeat offenders with problematic drug and alcohol abuse, continues to plague the US CJS and imposes a heavy burden on taxpayer dollars. Correctional care in the form of drug and alcohol treatment among offenders should be prioritized in order to disrupt this cycle. The very high influx of drug and alcohol dependents cycling through the CJS should be seen as an opportunity to initiate treatment and rehabilitation. In conjunction with behavioral interventions, effective pharmacotherapy has been shown to sustain abstinence and reduce cravings for opioids and alcohol for recently released inmates. Despite this, it cannot be emphasized enough that several major hurdles must be overcome in order to efficiently deliver correctional substance abuse care for incarcerated persons. First, improving education of correctional staff would likely increase acceptance of opioid and alcohol pharmacotherapies among correctional staff. The percentage of CJS settings offering medication-assisted therapy in the USA is dismally low, despite evidence indicating positive outcomes for offenders. Furthermore with the approval of Vivitrol® for treatment of opioid dependence as well as alcohol dependence, treatment options are expanded for offenders and the opioid antagonist may be more appealing to the CJS, given its once monthly adherence benefit and no overdose or diversion concerns.

Second, participation in psychosocial support programs in jails or prisons should be increased. Psychosocial groups, especially 12-step programs such as AA, have been very successful in reducing relapse to alcohol and drug use as has therapeutic communities and CBI.

Third, before release, correctional health care providers and case managers should work in concert to ensure that adherence to medically assisted therapy and medications for other comorbidities is not jeopardized, particularly for HIV disease. Relapse to drug and alcohol use is a significant risk factor for noncompliance or treatment failure, yet rigorous follow-up and treatment of substance use disorders can decrease relapse and recidivism rates. See Fig. 1.

The evidence-based practices presented in this entry are mostly limited to opioid and alcohol dependence. Polysubstance use is common among offenders, yet currently there is no federally approved medically assisted therapy for other highly abused addictive substances such as cocaine or methamphetamine. CBT has shown promising results in treating cocaine dependence; however, an effective pharmacologic treatment for long-term abstinence does not exist as of yet. Disulfiram, a pharmacotherapy for alcohol dependency, is being investigated as a possible treatment for reducing cocaine use. Additionally, clinical trials are currently being conducted to evaluate the efficacy of an anti-cocaine vaccine. In a trial with 115 individuals however, less than 40 % of subjects reached the requisite antibody level needed to block cocaine's access to the brain. Treatment for methamphetamine dependency is of paramount importance now, as it is becoming a more emergent problem in many parts of the country. Potential immunotherapies for methamphetamine treatment have also been proposed; however, they still remain in the preclinical phase. Other pharmacotherapies show promise as well, such as XR-NTX and methylphenidate, but are still being evaluated.

Similar to management of most chronic diseases, treatment of substance use disorders among offenders requires a multidisciplinary approach inside and outside correctional settings. Moreover, a sizeable fraction of drug-using offenders suffer from numerous psychiatric and infectious comorbidities, which have broad societal and public health implications. Attention to inmates' medical needs must be addressed in prisons and jails as a measure of safety, not just for the individual, but also from a public health standpoint.

Recommended Reading and References

Altice FL, Kamarulzaman A, Soriano VV, Schechter M, Friedland GH (2010) Treatment of medical, psychiatric, and substance-use comorbidities in people infected with HIV who use drugs. Lancet 376(9738):367–387. doi:10.1016/s0140-6736(10)60829-x

Anton RF, O'Malley SS, Ciraulo DA, Cisler RA, Couper D, Donovan DM, Gastfriend DR, Hosking JD, Johnson BA, LoCastro JS, Longabaugh R, Mason BJ, Mattson ME, Miller WR, Pettinati HM, Randall CL, Swift R, Weiss RD, Williams LD, Zweben A (2006) Combined pharmacotherapies and behavioral interventions for alcohol dependence: the COMBINE study: a randomized controlled trial. JAMA 295(17):2003–2017. doi:10.1001/jama.295.17.2003

Binswanger IA, Stern MF, Deyo RA, Heagerty PJ, Cheadle A, Elmore JG, Koepsell TD (2007) Release from prison – a high risk of death for former inmates. N Engl J Med 356(2):157–165. doi:10.1056/NEJMsa064115

Bonczar TP, Glaze LE (2009) Publication probation and parole in the United States, 2008. US Department of Justice/Bureau of Justice Statistics, Washington, DC, NCJ 228230

Bourne PG, Alford JA, Bowcock JZ (1966) Treatment of skid-row alcoholics with disulfiram. Q J Stud Alcohol 27(1):42–48

Brahen LS, Capone T, Wiechert V, Desiderio D (1977) Naltrexone and cyclazocine. A controlled treatment study. Arch Gen Psychiatry 34(10):1181–1184

Coviello DM, Cornish JW, Lynch KG, Alterman AI, O'Brien CP (2010) A randomized trial of oral naltrexone for treating opioid-dependent offenders. Am J Addict 19(5):422–432. doi:10.1111/j.1521-0391.2010.00070.x

Ditton PM (1999) Mental health and treatment of inmates and probationers. US Department of Justice/Bureau of Justice Statistics, Washington, DC, NCJ 174463

Friedland G (2010) Infectious disease comorbidities adversely affecting substance users with HIV: hepatitis C and tuberculosis. J Acquir Immune Defic Syndr 55(Suppl 1):S37–S42. doi:10.1097/QAI.0b013e3181f9c0b6

Garbutt JC, Kranzler HR, O'Malley SS, Gastfriend DR, Pettinati HM, Silverman BL, Loewy JW, Ehrich EW (2005) Efficacy and tolerability of long-acting injectable naltrexone for alcohol dependence: a randomized controlled trial. JAMA 293(13):1617–1625. doi:10.1001/jama.293.13.1617

Grant BF, Dawson DA, Stinson FS, Chou SP, Dufour MC, Pickering RP (2004) The 12-month prevalence and trends in DSM-IV alcohol abuse and dependence: United States, 1991–1992 and 2001–2002. Drug Alcohol Depend 74(3):223–234. doi:10.1016/j.drugalcdep.2004.02.004

Griffith JD, Rowan-Szal GA, Roark RR, Simpson DD (2000) Contingency management in outpatient methadone treatment: a meta-analysis. Drug Alcohol Depend 58(1–2):55–66

Hammett TM (2006) HIV/AIDS and other infectious diseases among correctional inmates: transmission, burden, and an appropriate response. Am J Public Health 96(6):974–978. doi:10.2105/ajph.2005.066993

Hammett TM, Harmon MP, Rhodes W (2002) The burden of infectious disease among inmates of and releasees from US correctional facilities, 1997. Am J Public Health 92(11):1789–1794

Karberg JC, James DJ (2005) Substance dependence, abuse, and treatment of jail inmates, 2002. US Department of Justice/Bureau of Justice Statistics, Washington, DC, NCJ 209588

Kinlock TW, Battjes RJ, Schwartz RP (2002) A novel opioid maintenance program for prisoners: preliminary findings. J Subst Abuse Treat 22(3):141–147

Kinlock TW, Gordon MS, Schwartz RP, Fitzgerald TT, O'Grady KE (2009) A randomized clinical trial of methadone maintenance for prisoners: results at 12 months postrelease. J Subst Abuse Treat 37(3):277–285. doi:10.1016/j.jsat.2009.03.002

Kinlock TW, Gordon MS, Schwartz RP, Fitzgerald TT (2010) Developing and implementing a new prison-based buprenorphine treatment program. J Offender Rehabil 49(2):91–109. doi:10.1080/10509670903534951

Levasseur L, Marzo JN, Ross N, Blatier C (2002) Frequency of re-incarcerations in the same detention center: role of substitution therapy. A preliminary retrospective analysis. Ann Med Interne (Paris) 153(3 Suppl):1S14–1S19

Lincoln T, Kennedy S, Tuthill R, Roberts C, Conklin TJ, Hammett TM (2006) Facilitators and barriers to continuing healthcare after jail: a community-integrated program. J Ambul Care Manage 29(1):2–16

Lobmaier PP, Kunoe N, Gossop M, Katevoll T, Waal H (2010) Naltrexone implants compared to methadone: outcomes six months after prison release. Eur Addict Res 16(3):139–145. doi:10.1159/000313336

Lussier JP, Heil SH, Mongeon JA, Badger GJ, Higgins ST (2006) A meta-analysis of voucher-based reinforcement therapy for substance use disorders. Addiction 101(2):192–203. doi:10.1111/j.1360-0443.2006.01311.x

Magura S, Lee JD, Hershberger J, Joseph H, Marsch L, Shropshire C, Rosenblum A (2009) Buprenorphine and methadone maintenance in jail and post-release: a randomized clinical trial. Drug Alcohol Depend 99(1–3):222–230. doi:10.1016/j.drugalcdep.2008.08.006

Mannelli P, Peindl K, Masand PS, Patkar AA (2007) Long-acting injectable naltrexone for the treatment of alcohol dependence. Expert Rev Neurother 7(10):1265–1277. doi:10.1586/14737175.7.10.1265

Maruschak L (2008) HIV in Prisons, 2006 (NCJ 222179). US Department of Justice, Bureau of Justice Statistics, Washington, DC

Maruschak L (2009) HIV in prisons, 2007–08. US Department of Justice/Bureau of Justice Statistics, Washington, DC

McMillan GP, Lapham SC (2005) Staff perspectives on methadone maintenance therapy (MMT) in a large southwestern jail. Addict Res Theory 13(1):53–63. doi:10.1080/16066350512331328159

Minozzi S, Amato L, Vecchi S, Davoli M, Kirchmayer U, Verster A (2011) Oral naltrexone maintenance treatment for opioid dependence. Cochrane Database Syst Rev (4), CD001333. doi:10.1002/14651858.CD001333.pub4. http://www.thecochranelibrary.com/view/0/index.html

Monterosso JR, Flannery BA, Pettinati HM, Oslin DW, Rukstalis M, O'Brien CP, Volpicelli JR (2001) Predicting treatment response to naltrexone: the influence of craving and family history. Am J Addict 10(3):258–268

Mumola C (1999) Substance abuse and treatment of state and federal prisoners, 1997. US Department of Justice/Bureau of Justice Statistics, Washington, DC, NCJ 172871

Mumola C, Karberg JC (2006) Drug use and dependence, state and federal prisoners, 2004. US Department of Justice/Bureau of Justice Statistics, Washington, DC, NCJ 213530

O'Malley SS, Garbutt JC, Gastfriend DR, Dong Q, Kranzler HR (2007) Efficacy of extended-release naltrexone in alcohol-dependent patients who are abstinent before treatment. J Clin Psychopharmacol 27(5):507–512. doi:10.1097/jcp.0b013e31814ce50d

Rinaldo D (2008) Medicaid financing of medication-assisted treatment for opiate addiction. Accessed 28 Jan 2013

Spaulding AC, Seals RM, Page MJ, Brzozowski AK, Rhodes W, Hammett TM (2009) HIV/AIDS among inmates of and releasees from US correctional facilities, 2006: declining share of epidemic but persistent public health opportunity. PLoS One 4(11):e7558. doi:10.1371/journal.pone.0007558

Springer SA, Altice FL (2005) Managing HIV/AIDS in correctional settings. Curr HIV/AIDS Rep 2(4):165–170

Springer SA, Chen S, Altice FL (2010) Improved HIV and substance abuse treatment outcomes for released HIV-infected prisoners: the impact of buprenorphine treatment. J Urban Health 87(4):592–602. doi:10.1007/s11524-010-9438-4

Springer SA, Azar MM, Altice FL (2011) HIV, alcohol dependence, and the criminal justice system: a review and call for evidence-based treatment for released prisoners. Am J Drug Alcohol Abuse 37(1):12–21. doi:10.3109/00952990.2010.540280

Springer SA, Qiu J, Saber-Tehrani AS, Altice FL (2012) Retention on buprenorphine is associated with high levels of maximal viral suppression among HIV-infected opioid dependent released prisoners. PLoS One 7(5):e38335. doi:10.1371/journal.pone.0038335

Tomasino V, Swanson AJ, Nolan J, Shuman HI (2001) The Key Extended Entry Program (KEEP): a methadone treatment program for opiate-dependent inmates. Mt Sinai J Med 68(1):14–20

Drug Corners

► Spatial Perspectives on Illegal Drug Markets

Drug Courts

Michael Rempel
Drug Courts, Criminal Justice Intervention, Center for Court Innovation, New York, NY, USA

Overview

Drug courts are the oldest, most prolific, and most studied of the major alternative court models, which also include domestic violence, mental health, community, and reentry courts. Similar to these other models, drug courts organize their cases on a separate court calendar, presided over by a specially trained judge. What distinguishes drug courts is their focus on cases involving an underlying drug addiction. To treat the addiction, drug courts employ a combination of treatment and judicial oversight, generally for 1 year or longer. Judicial oversight generally involves regular drug testing, meetings with court-affiliated case managers, and status hearings before the judge. At these hearings, the judge and participant directly converse, while the attorneys often remain silent. The judge typically responds to progress with verbal praise or tangible incentives (e.g., certificates, journals, or gift cards) and to noncompliance with interim sanctions (e.g., more frequent status hearings, community service, or short jail stays). Drug courts are voluntary programs; those who do not wish to participate can have their cases handled in a conventional fashion instead.

The first drug court was an adult program that opened in Miami-Dade County in 1989. In 2009, two decades later, 2,459 drug courts had been established. These courts include 1,317 adult, 476 juvenile, 322 family dependency, 172 DWI, 89 tribal, and 83 other drug courts (Huddleston and Marlowe 2011). The original adult model enrolls an estimated 55,000 defendants per year in the USA (Bhati et al. 2008). Comparable estimates are unavailable for any other model. Internationally, drug courts have spread to countries as varied as Australia, Canada, Chile, Jamaica, Mexico, New Zealand, Norway, and the United Kingdom.

Numerous studies have shown that the original adult drug court model reduces re-offending, as compared with conventional prosecution (Gutierrez and Bourgon 2009; Mitchell et al. 2012; Shaffer 2011). Research suggests that adult drug courts reduce drug use as well (Government Accountability Office 2011; Rossman et al. 2011).

Over the past decade, scientific study surrounding drug courts has begun to evolve from evaluations of program impact (do they work) to studies of which target populations and program features enhance their success. Research has begun to coalesce, for instance, around the idea that the judge – through judicial status hearings and conversational interactions with participants – plays a particularly influential role (Marlowe et al. 2003; Rossman et al. 2011). This research bolsters the initial premise of the model that ongoing judicial oversight can substantially bolster the effects obtainable through community-based treatment alone.

This entry reviews why drug courts arose, describes their core policies and practices, surveys the research literature, and summarizes several current issues and controversies.

Origins

Several overlapping trends created conditions ripe for drug courts. Foremost among them was the strain on criminal justice systems nationwide, resulting from skyrocketing court and correctional caseloads. For instance, from the early 1980s to the late 1990s, criminal court caseloads increased by more than 50 % nationally (Ostrom and Kauder 1999). Over this same period, the increase in correctional caseloads was even greater. As compared with 500,000 jail and prison inmates in 1980, there were two million inmates at all points throughout the 2000s (Bureau of Justice Statistics 2012).

To explain these trends, key factors include the crack epidemic of the 1980s, combined with state and federal policies targeting drug-related crimes for aggressive enforcement. Beginning in 1998, data collected as part of the federal Arrestee Drug Abuse Monitoring (ADAM) Program confirmed the preexisting beliefs of many criminal justice officials that a high percentage of criminal arrestees were drug-involved. The 2003 survey found that 70 % of arrestees in approximately three dozen jurisdictions tested positive for at least one of nine illegal drugs, and 37 % of the arrestees were “heavy users,” defined by self-reported use of marijuana, cocaine, heroin, methamphetamine, or PCP in at least 13 of the past 30 days (Zhang 2003).

In this context, drug courts arose as a solution that could garner support from many sides of the ideological spectrum. For liberals, drug courts held the potential to expand access to treatment while counteracting a growing trend towards the mass incarceration of drug offenders. For conservatives, drug courts held the potential to increase public safety (through recidivism reductions) and to limit the costs of incarceration to state and local taxpayers. Not surprisingly, federal funding for drug courts has persisted across three presidential administrations.

The creators of the first drug courts, however, began with a relatively nonideological and practical objective: By routing drug cases to a single court calendar, the drug court model offered important efficiency advantages for overburdened court systems. Indeed, research demonstrates that efficiency was a critical objective motivating the earliest drug courts, but by the mid-1990s, efficiency was supplanted by the goals of rehabilitation and recidivism reduction (McCoy 2003).

Buttressing the case for rehabilitation as a legitimate court focus was the theory of *therapeutic jurisprudence*, which first gained currency in the early 1990s, just as drug courts began to spread (Wexler 1993). The theory argues that it is legitimate for the law to incorporate therapeutic considerations in addition to strictly legal ones that sentencing does not have to reflect "just desserts" alone (punishment proportional to offense) but can also embody a desire to address the underlying problems of the litigant. To facilitate the rehabilitation process, therapeutic jurisprudence embraces the idea that judges should adopt a nontraditional judicial role, including a willingness to engage in the classic drug court strategy of direct conversational interaction with defendants or other litigants.

Core Policy Components

Specific policies and practices vary from jurisdiction to jurisdiction, but in broad outline, the drug court model is relatively similar everywhere. To a large extent, this similarity reflects the influence of the National Association of Drug Court Professionals (NADCP), a trade association in existence since 1994 that produces drug court-related publications, provides technical assistance, and runs a popular annual training conference for drug court practitioners.

In 1997, NADCP convened a national group of experts and issued a document known as *Defining Drug Courts: The Key Components* (Bureau of Justice Assistance 1997). The document identifies ten policy components that all drug courts should adopt. Since the document was issued, federal and state agencies have routinely required drug courts to demonstrate that they follow all ten components in order to receive funding. The *Key Components* document has led drug courts to be defined by the following policies. (The list below does not verbatim reproduce the original "ten key components" but summarizes the resulting policies that now inform the field.)

- *Early Identification*: Potential participants are assumed to be especially receptive to the drug court intervention at the "crisis moment" comprised by the initial arrest or case filing. Therefore, drug courts screen and assess potential participants for drug dependence and other psychosocial problems as soon as possible after case initiation.
- *Community-Based Treatment*: Treatment is considered most effective when tailored to the individual. Therefore, drug courts link participants to community-based treatment, drawing on a continuum of possible outpatient and residential treatment modalities.
- *Legal Leverage*: Participants are considered most likely to remain engaged in treatment when faced with undesirable legal consequences for dropping out. Therefore, drug courts establish clear jail or prison alternatives for unsuccessful participation while using the positive incentive of a charge dismissal or reduction for graduates. (In family dependency drug courts, the legal incentives revolve around child reunification but are less cut and dry, since reunification must ultimately follow from the "best interests of the child.")
- *Judicial Status Hearings*: Participants are assumed to perform better when under close surveillance by the court and when routinely

reminded of their responsibilities. Therefore, drug courts require participants to report back to court regularly, often weekly or biweekly at the outset of participation, for judicial status hearings on their progress.

- *Direct Judicial Interaction*: As an authority figure, the judge is assumed to be in a unique position to motivate compliance. Therefore, the judge directly converses with participants during scheduled judicial status hearings, asks participants about their needs, acknowledges and praises progress, and admonishes participants for noncompliance.
- *Drug Testing*: Frequent and random drug testing is assumed to deter drug use and also to assist drug court staff in understanding whether participants are currently receiving a sufficient intensity of treatment. Therefore, drug courts administer random drug tests.
- *Interim Sanctions and Incentives*: Classic behavioral modification theory recommends the consistent use of sanctions and incentives. Therefore, the judge imposes sanctions for noncompliance (including short jail stays) and distributives incentives for progress (including regular symbolic incentives, such as praise and courtroom applause).
- *Multiple Chances*: The physiological effects of withdrawal, as well as psychological and other barriers to recovery, may make relapse-free drug court participation unlikely for many. Therefore, drug courts use interim sanctions to respond to initial relapses or other noncompliance while terminating participants only for repeated or severe misbehavior.
- *Case Management*: Court-affiliated case managers are used as a critical liaison between community-based programs and the court (i.e., judge and attorneys). Therefore, most drug courts employ staff members who devise the treatment plan, suggest modifications based on progress, and communicate with treatment providers. Case managers may also meet regularly with participants to motivate progress or address newly apparent needs.
- *Ancillary Services*: It is assumed that participants may have multitude of needs, including co-occurring mental health disorders, lack of education, vocational or employment deficits, and family dysfunction. Therefore, drug courts are prepared to link participants to additional services other than substance abuse treatment.
- *Collaboration*: Drug courts are predicated on the notion that the court, attorneys, and treatment providers should seek the same goals: the rehabilitation of the individual and the consequent reduction in that individual's threat to public safety. Therefore, once a participant is enrolled, all parties curtail the adversarial process and are supposed to work together to promote the recovery of the individual. To facilitate collaboration, most drug courts hold regular "staffing" meetings attended by an interdisciplinary "drug court team." In these meetings, the team discusses specific cases and makes decisions, usually by consensus, on how to respond to their compliance and service needs.

Research on Adult Drug Courts

More than 150 evaluations have been conducted to date, of which by far the greatest number has focused on the adult drug court model.

Do Adult Drug Courts Work?

Research has amply demonstrated the positive impact of adult drug courts on re-offending. Across more than 90 studies, including statewide evaluations in California, Indiana, Maryland, New York, Ohio, and Washington, adult drug courts have consistently produced recidivism reductions – although the precise magnitude of the impact has varied from site to site. Three *meta-analyses* – which synthesize the findings of other studies – variously estimated that adult drug courts produce an average reduction in the rearrest or reconviction rate of 8–13 percentage points (Gutierrez and Bourgon 2009; Mitchell et al. 2012; Shaffer 2011). (e.g., a 13-point reduction might involve a rearrest rate of 50 % in the comparison group declining to 37 % among those in drug court.) Most evaluations tracked defendants over 2 years or less, but several extended the follow-up period to 3 years or longer and still reported positive effects (Gottfredson et al. 2006; Rempel et al. 2003).

Only a handful of studies have directly examined effects on future drug use, but their results also were mostly positive (e.g., see Gottfredson et al. 2005). NIJ's *Multi-Site Adult Drug Court Evaluation*, a 5-year study of 23 drug courts and six comparison jurisdictions (Rossman et al. 2011), found that in the year prior to 18-month follow-up, drug court participants were significantly less likely to report any drug use (56 % vs. 76 %) or any "serious" use (41 % vs. 58 %). (Serious use omits marijuana and "light" alcohol use, with the latter defined as less than four drinks per day for women and less than five drinks per day for men.) The same study also detected positive effects when examining the results of oral swab tests that were conducted at the 18-month follow-up.

Among the plethora of single-site studies in the literature, three involve the random assignment of defendants to drug court and control conditions. The first of these randomized controlled trials took place in Maricopa County (AZ) and detected mixed results after 1 year (Deschenes et al. 1995) but more positive results after a 3-year timeframe (Turner et al. 1999). A randomized trial of a New South Wales (Australia) drug court reported a significant reduction in rearrests over 18 months (Shanahan et al. 2004). Finally, a randomized trial of the Baltimore drug court found that over 3 years, the court significantly reduced rearrests (Gottfredson et al. 2006) as well as drug and alcohol use (Gottfredson et al. 2005).

Do Adult Drug Courts Save Money?

Part of the original rationale for adult drug courts was to generate cost savings for increasingly overburdened criminal justice systems. Moreover, an array of cost-benefit studies almost universally confirms that drug courts save money, at least in the long term. An evaluation of nine adult drug courts in California found that the median drug court saved $5,139 per participant (Carey et al. 2005). This study also found that across multiple public agencies, including the court, prosecutor, public defender, law enforcement, probation, corrections, and treatment, the largest savings were accrued by corrections – through reductions in incarceration – and the only agency that incurred a net cost was treatment (e.g., through Medicaid/Medicare payments). Other cost studies have reported similar findings. For instance, across six sites in Washington State, five produced cost savings at an average of $3,892 per participant (Barnoski and Aos 2003). *NIJ's Multi-Site Drug Court Evaluation* detected average savings of $5,680–$6,208 per participant across its 23-site drug court sample. However, this last study came with an important asterisk: Drug courts entailed higher up-front costs (for treatment and other services) than comparison jurisdictions. Drug courts ultimately produced savings by reducing recidivism: that is, by reducing the costs that would otherwise have been produced by future crimes. These crime-related savings largely materialized, because drug courts achieved a significant reduction in the *most serious* future crimes that have otherwise produced substantial healthcare and property-related costs to victims. In sum, drug courts achieved a significant return on investment with high-risk offenders who might otherwise have committed serious crimes, but they produced far smaller savings, if any, with low-risk offenders.

Why Do Adult Drug Courts Work?

Much of the early research on adult drug courts focused on the bottom-line question of *whether* they work, but recent studies have yielded a rich array of findings concerning *why* they work – which theories of change explain their capacity to alter participant behavior for the better.

Treatment. A long-standing prior literature finds that when drug-addicted individuals are retained in treatment for significant periods – at least 90 days and ideally up to 1 year – those individuals tend to engage in less posttreatment drug use and criminal behavior. Research also shows that cognitive-behavioral therapy (CBT) is particularly effective in creating the pro-social thought, attitudinal, and decision-making changes that can, in turn, elicit pro-social behaviors (see Lipsey et al. 2007). Finally, research suggests that treatment is most effective when it does not adopt a “one-size-fits-all” approach but is tailored to the individual characteristics, learning style, and needs of each participant (Andrews and Bonta 2006). Despite the sizable preexisting literature on this subject, drug court research has yet to uncover a clear treatment effect. Instead, some research suggests that many drug courts refer participants to treatment providers that have failed to adopt evidence-based practices (Lutze and van Wormer 2007). A recent meta-analysis concurs that evidence-based practices are underutilized and finds that when they *are* used, drug courts produce larger recidivism reductions than otherwise (Gutierrez and Bourgon 2009). This analysis indicates that treatment *can* contribute to positive impacts, but because treatment quality varies substantially, it does not always produce its desired effects.

Deterrence. Based on research with other offender populations, it is possible to *deter* future misbehavior with legal sanctions that involve *certainty* (each infraction elicits a sanction), *celerity* (sanction imposed soon after the infraction), and *severity* (sanction is sufficiently undesirable to deter noncompliance but not so severe as to preclude upgrading to a more serious sanction after subsequent infractions) (e.g., Marlowe and Kirby 1999). In a drug court context, deterrence entails routine surveillance through judicial status hearings, drug tests, and case manager meetings; threat of interim sanctions; and threat of incarceration for final termination. Of these practices, several studies have found that drug testing and judicial status hearings are effective in reducing crime and drug use (e.g., Gottfredson et al. 2007; Marlowe et al. 2003; Rossman et al. 2011). Yet, several studies involving in-depth participant focus groups suggest that it is less the deterrent effect of surveillance and more the positive engagement effect of motivational interactions with a judge that leads judicial status hearings in particular to be effective (e.g., Goldkamp et al. 2001).

Providing a clearer test of deterrence theory, research also suggests that the threat of jail or prison for failing drug court altogether is a key factor motivating compliance. NIJ’s *Multi-Site Adult Drug Court Evaluation* found that participants who perceived themselves to face more severe consequences in the event of program failure engaged in less noncompliance than others while enrolled in the program and less crime and drug use at follow-up (Rossman et al. 2011). This research is consistent with previous studies, which generally link greater legal leverage to improved treatment outcomes (e.g., Rempel and DeStefano 2001; Young and Belenko 2002). Qualifying these findings, research also makes clear that it is not only the factual jail or prison sentence that participants face in the event of failure that influences their performance. Drug court participants who face similar legal consequences may have differing *perceptions* of those consequences due to what the participants were told, by whom, how often they were reminded of their responsibilities, and how well they actually understood those consequences. Research has shown that eliciting participant *perceptions* of legal pressure comprises the critical link to improved behavioral outcomes (Young and Belenko 2002).

As opposed to the threat of incarceration for final program failure, research has been less clear concerning the deterrent effect of *interim sanctions*. In determining when and how to use sanctions, drug courts often employ a great deal of individualized discretion – that is, using different sanctions with different participants and not necessarily imposing a sanction in response to each and every infraction (Rempel et al. 2003; Rossman et al. 2011). Yet, individualized discretion may vitiate the behavior modification principle of *certainty*, which holds that it is best to impose a sanction each time and to employ similar sanctions in response to similar infractions. Moreover, it is unclear at this time whether interim sanctions might yield clearer and more positive effects if drug courts adhered more consistently to best sanctioning practices.

Procedural Justice. Procedural justice concerns the fairness of court procedures and interpersonal treatment while a case is processed (Tyler and Huo 1990). Key dimensions include *voice* (litigants have their side heard), *respect* (litigants are treated with dignity and respect), *neutrality* (decision-makers are seen as trustworthy and unbiased), and *understanding* (court language and decisions are readily understood). Some research finds that when litigants believe the court process was fair, they become more likely to comply with court orders and to follow the law in the future (e.g., Tyler and Huo 2002). Procedural justice acts as the positive flip side to deterrence, eliciting compliance through fair procedures rather than through threat of consequences. The randomized controlled trial of the Baltimore drug court found that greater participant perceptions of procedural justice were associated with less crime and drug use at 3-year follow-up (Gottfredson et al. 2007). Similarly, in *NIJ's Multi-Site Adult Drug Court Evaluation*, drug court participants held significantly greater perceptions than the comparison group that the judge treated them fairly – and in turn, these perceptions strongly predicted reduced crime and drug use at follow-up. This same study found that specific drug courts in which the researchers independently rated the judge as having a more engaging demeanor – more respectful, fair, attentive, enthusiastic, consistent, predictable, caring, and knowledgeable – produced better outcomes than other drug courts. In light of this research, the federal Office of Justice Programs recently identified both *procedural justice* and *judicial interaction* as design features that all drug courts should fully embrace in order to be evidence-based (OJP 2012).

Research on Other Drug Court Models

This section reviews what has been learned to date about juvenile, family dependency, and DWI drug courts. In addition, reentry courts receive attention in a separate alternative courts entry (Lindquist et al., this volume).

As compared with the adult model, some have argued that juvenile drug courts may not be as successful, since juveniles do not tend to be addicted to drugs. Instead, they are more often casual drug users, especially of marijuana, who face a series of other social and psychological problems, including ties to deviant peer groups, low family functioning, poor educational performance, poor impulse control, and developmental disabilities. Still other juveniles may not have any severe problems of this nature but may instead be engaging in common teenage deviance that is likely to desist on its over time, in the absence of intervention. Indeed, results to date have been mixed. A recent meta-analysis of 34 juvenile drug court evaluations detected an average reduction in re-offending of eight percentage points; however, the average effect was smaller in the most methodologically rigorous studies, and when isolating drug-related re-offending in particular, there was not any significant impact (Mitchell et al. 2012).

Although these results are ostensibly disappointing, some researchers have found that when juvenile drug courts employ certain evidence-based practices, they can be more effective. Specifically, several studies have found that juvenile drug courts produce positive outcomes when they facilitate pro-social peer activities, limit contact with antisocial peers, involve family members in judicial status hearings, and employ evidence-based treatments such as multisystemic therapy, which entails comprehensive engagement with the youth, parents, teachers, and other systems in which the youth are involved (Salvatore et al. 2010; Schaeffer et al. 2010).

By comparison with juvenile drug courts, research findings have tended to be more positive for the remaining drug court models. Family dependency drug courts are somewhat unique in that they seek both to reduce substance abuse by the respondent parent and to achieve a positive permanency outcome for the child. Across several family dependency drug courts that have been evaluated, five of seven produced increased treatment completion rates for the respondent, and six of eight produced increased rates of parent–child reunification than equivalent comparison groups (e.g., see Fritsche et al. 2011; Green et al. 2009).

DWI courts have also yielded positive impacts. A recent meta-analysis isolated the impact of 28 DWI courts on re-offending and found that all except four reduced re-offending to at least some degree, with an average reduction of 12 percentage points (Mitchell et al. 2012). It is perhaps unsurprising that the results for DWI courts mirror those for the original adult model. These two models are highly similar, with both focusing on drug-involved adult criminal defendants – except DWI courts serve those whose court case involved DWI or DUI charges.

Issues and Controversies

Drug courts have spawned a large number of issues and controversies. Several prominent examples follow, of which some concern the legal and constitutional implications of the drug court model, whereas other issues concern specific findings in the evaluation literature.

The Therapeutic Judicial Role

Some have expressed concern that it may not be legally appropriate for judges to serve in any other capacity than as neutral arbiters of facts and legal questions. In a well-publicized commentary, the Honorable Morris B. Hoffman succinctly summarized this position as follows: "Judges have the right to exercise only those powers necessary to dispose of the cases before us. When we succumb to the very human temptation to do more – to fill the void that is so achingly apparent in so many of the dysfunctional people we see every day – we not only risk being wrong, but we risk being imperial (2000: 1478)." Hoffman further posits that when the judicial branch involves itself in social policymaking, or when individual judges apply therapeutic methods, it violates the constitutional separation of powers, which leaves it exclusively to the legislative and executive branches to develop and implement social policy (2000: 1479).

Others judges have sought to articulate how legal due process can be maintained even as judges extend their focus beyond legal process alone to the *outcomes* their decisions produce, such as recidivism reductions (see Hora et al. 1999). Moreover, a survey of more than 1,000 trial court judges nationwide found, on balance, support for judging methods that are common to drug courts and other alternative courts. Eighty percent of the responding judges believed it was very or somewhat important for judges to consider "the individual needs or underlying problems of the litigant," whereas only 25 % believed that that "problem-solving compromises the neutrality of the court" (Farole et al. 2008). Nonetheless, some would question whether majority opinion matters on these kinds of questions. The therapeutic judicial role that drug courts embrace – and especially the use of direct interaction between judge and participant coupled with a less adversarial process between the attorneys – remains highly controversial as legal practices. As a legal protection for drug court participants, while not opposing the therapeutic judicial role per se, the National Association of Criminal Defense Lawyers (NADCL) (2009) recommends that defense attorneys be present at all judicial status hearings, in the event that issues arise (e.g., related to sanctions or possible program termination) that require zealous defense advocacy.

Risk of Net Widening

A second concern, also rooted in legal due process principles, is that drug courts may inappropriately deepen and lengthen the criminal justice involvement of many defendants. Known as "net widening," this concern takes two essential forms. First, some have linked the establishment of drug courts to the increased criminalization of nonviolent drug behavior: that is, to rampant drug arrests and prosecutions (Drug Policy Alliance 2011). However, there is relatively little evidence to support this concern beyond a reported massive increase in drug arrests after the founding of the Denver (CO) Adult Drug Court (Hoffman 2000).

The second variant of net widening appears in the argument that, despite their widely promulgated status as an alternative to incarceration, the average drug court *increases* the criminal justice involvement of their participants – and increases rather than reduces time spent incarcerated (Drug Policy Alliance 2011). Here, critics can point to hard evidence supporting their position. A number of multisite studies found that adult drug court participation lasts about 15 months on average, representing a longer period of time than the jail or prison sentence that most participants would have otherwise faced under conventional prosecution (Rempel et al. 2003; Rossman et al. 2011).

These same studies also found that those who fail drug court receive a significantly longer jail or prison sentence than they would have received had they not enrolled in the first place. After considering the lengthier sentences imposed on those who fail with the complete avoidance of jail or prison sentences for those who graduate from drug court, the *average* effect on incarceration appears to be a wash. Neither *NIJ's Multi-Site Adult Drug Court Evaluation* nor a randomized controlled trial of the Baltimore drug court detected a net difference in incarceration sentences between drug court and comparison defendants (e.g., see Rossman et al. 2011). Further, in a study of six New York State drug courts, when combining those who graduated and failed the program, participants averaged significantly shorter incarceration sentences than the comparison group in three sites, a significantly longer sentence in one site, and no difference at all two sites (Rempel et al. 2003). Drug court proponents might counter that, even if drug courts do not produce an average reduction in incarceration on the initial criminal case, they may still reduce incarceration in the long term through recidivism reductions. This is exactly why the cost-benefit literature has found that adult drug courts save money in the long term. Nonetheless, those who are concerned with net widening tend to predicate their position on the basic legal fairness of the initial court outcomes, regardless of whether outcomes are positive in the long term.

In consideration of these issues, NACDL (2009) recently advocated a number of reforms. First, contrary to the practice in many drug courts, NACDL proposed that prosecutors should not be allowed to reject cases for drug court when formal charge and criminal history criteria indicate that the case is eligible to participate. NACDL noted that some prosecutors may tend to reject the highest-risk cases, even though such cases are the least vulnerable to net-widening concerns. More generally, NACDL recommended limiting drug court eligibility to those defendants who would otherwise face lengthy sentences under conventional prosecution while developing less intensive programming for defendants who face less legal exposure.

Adult Drug Court Volume

The net-widening critique typically translates into a policy recommendation to limit participation to those individuals who would otherwise face substantial legal exposure (Drug Policy Alliance 2011). Some social scientists, however, have suggested that drug courts should *increase* rather than reduce their numbers. One analysis estimated that in 2005, whereas 1,471,338 adult arrestees in the USA either abused or were dependent on drugs, only 55,365 (3.8 %) were enrolled in an adult drug court (Bhati et al. 2008). Even allowing that these estimates are now several years dated and were based on extrapolations from multiple data sources, this analysis makes clear that adult drug courts reach a small fraction of the eligible pool. This same study also projected that whereas adult drug courts in 2005 produced financial benefits to society worth about $624 million, if these courts had served all substance-abusing and substance-dependent defendants, the benefit might have reached $46 billion – resulting from an investment of $13.7 billion in treatment. Importantly, the implication that drug courts should serve more participants does not automatically contradict the aforementioned concerns about net widening. Some might reconcile these considerations in proposing that adult drug courts should reach a higher percentage of all *serious* cases – those that face substantial legal exposure – while carefully restricting eligibility where net-widening concerns might apply.

Uneven Treatment Quality

A long-standing research literature supports the general benefits of substance abuse treatment. Yet, drug court research has yet to provide clear confirmation of the positive effects of *treatment* – as opposed to judicial status hearings or other drug court practices. This does not mean that treatment is unimportant, for less than a handful of drug court studies to date have attempted a tease out its specific impact. Still, available research indicates that many drug court programs do not provide what national experts would term *evidence-based treatment*. Many treatment programs suffer from inadequate staff training, high turnover, lack of written curricula specifying what each treatment session should cover, and insufficient use of proven cognitive-behavioral therapy techniques (Lutze and van Wormer 2007). In addition, a serious drug addiction is a brain disease that can often be addressed in part through medication-assisted treatment (MAT). However, many drug courts do not have access to local treatment programs whose staff is trained to provide and monitor such treatment (Lutze and van Wormer 2007). Of final note, whereas drug court participants often possess multiple problems other than substance abuse, including pro-criminal thought patterns, ties to antisocial peers, antisocial personality patterns, and employment deficits, many drug courts neither conduct a rigorous assessment for these other needs nor are able to make appropriate treatments available.

Some might conclude that drug court staff should monitor and seek to improve the quality of treatment. Supporting such a conclusion, a recent review found that drug courts that do adhere to one or more evidence-based assessment and treatment practices produced significantly larger reductions in re-offending than other drug courts (Gutierrez and Bourgon 2009). However, in many jurisdictions, particularly small ones with a limited number of available providers, it may simply not be possible for a court to order or obtain best treatment practices. Accordingly, the question of how to improve the quality of treatment received by drug court participants is one that depends not merely on clear guidance from research but also on overcoming practical obstacles to the dissemination of evidence-based practices within local provider communities.

Appropriate Target Population

General research on offender interventions recommends varying program intensity based on risk of re-offense. Specifically, research indicates that high-risk defendants – those who are especially likely to re-offend in the absence of any intervention – require a particularly intensive form of treatment. By contrast, low-risk defendants may not be well served by an intensive and lengthy intervention such as drug court; since they are unlikely to re-offend in any case, low-risk defendants may be better served if they are left alone (Andrews and Bonta 2006).

Little research has put the risk principle to the test in drug courts. However, it is conceivable that the literature on juvenile drug courts is somewhat mixed, in part because these courts may order some teenagers to a year or more of intensive program participation, who might otherwise have desisted from crime on their own – that is, their risks and needs may have been too low to merit subjection to the intensive drug court model.

Regarding adult drug courts, one study of the Los Angeles program confirmed that it was more effective with high- than low-risk defendants (Fielding et al. 2002). Similarly, *NIJ's Multi-Site Adult Drug Court Evaluation* found some evidence that those who were particularly likely to commit serious crimes in the future and those who presented with a more severe addiction at baseline were especially likely to benefit from their participation. This same study uncovered few other differences in the magnitude of the drug court impact, based on participant demographics (age, race, or sex) or self-reported motivation at baseline (Rossman et al. 2011).

Interestingly, the social science-based concern that drug courts are best suited to high-risk defendants dovetails with the aforementioned defense bar position that drug courts should limit eligibility to high-risk/high-leverage defendants, who would otherwise face a substantial jail or prison sentence in the absence of drug court participation. In considering these issues, the National Association of Drug Court Professionals recently issued a two-part series of policy papers that essentially advised drug courts to abandon a "one-size-fits-all" program model in favor of multiple *tracks* (Marlowe 2012). In this system, programs would apply the full drug court model to defendants who pose a high risk of re-offending and have serious treatment needs related to a substance-dependence problem (track 1). Those who combine a high risk of re-offending in general but with less serious drug treatment needs in particular would receive intensive judicial oversight but less intensive community-based treatment (track 2). Conversely, those who combine a low risk of re-offending with serious treatment needs would receive less intensive oversight – that is, status hearings only when noncompliant – coupled with regular treatment (track 3). Finally, those who are both low risk and low need might receive general prevention services but would not be required to attend any form of intensive judicial oversight or treatment (track 4). Although the state of Missouri intends to implement a track system along these lines, it is unclear whether a new national movement will be launched in this direction. Moreover, research has not been conducted to determine whether a track-based system in fact yields improved drug court outcomes. In short, the development of tailored responses to different defendant populations is a potential cutting-edge policy direction that may or may not ultimately be the subject of broad experimentation nationwide.

Conclusion

As the first alternative court model, drug courts have existed since 1989, spawned the development of other alternative courts in the USA and internationally, and been the subject of a growing body of research and evaluation studies. At the same time, important unanswered questions persist. These include the effectiveness of models other than adult and DWI drug courts, the role of treatment in explaining drug court success and how to improve its quality, the appropriate target population for the full-length drug court model, and the solution to a number of competing legal and social science considerations that variously mitigate in favor of and against expanding the number of individuals who are enrolled in drug courts. Regardless of the effectiveness of any one model, the diffusion of drug courts nationwide has transformed the role that many courts are now prepared to embrace in devising solutions to vexing social problems.

Related Entries

- ▶ Community Courts
- ▶ Mental Health Courts
- ▶ Reentry Courts
- ▶ Therapeutic Jurisprudence

Recommended Reading and References

Drug Policy Alliance (2011) Drug courts are not the answer: toward a health-centered approach to drug use. Drug Policy Alliance, New York

Andrews DA, Bonta J (2006) The psychology of criminal conduct, 4th edn. LexisNexis, Newark

Barnoski R, Aos S (2003) Washington state’s drug courts for adult defendants: outcome evaluation and cost-benefit analysis. Washington State Institute for Public Policy, Olympia

Bhati A, Roman J, Chalfin A (2008) To treat or not to treat: evidence on the prospects of expanding treatment to drug-involved offenders. The Urban Institute, Washington, DC

Bureau of Justice Assistance (1997) Defining drug courts: the key components. Bureau of Justice Assistance, Washington, DC

Bureau of Justice Statistics (2012) Correctional populations: key facts at a glance. Washington, DC: Bureau of Justice Statistics. http://bjs.ojp.usdoj.gov/content/glance/tables/corr2tab.cfm

Carey SM, Crumpton D, Finigan MW, Waller MS (2005) California drug courts: a methodology for determining costs and benefits, phase II: testing the methodology: final report. NPC Research, Portland

Deschenes E, Turner S, Greenwood P (1995) Drug court or probation? An experimental evaluation of Maricopa County’s drug court. Justice Syst J 18:55–73

Farole DJ, Rempel M, Byrne F, Chang Y (2008) Problem-solving and the American bench: a national survey of trial court judges. Center for Court Innovation, New York

Fielding JE, Tye G, Ogawa P, Imam IJ, Long AM (2002) Los Angeles county drug court programs: initial results. J Subst Abuse Treat 23:217–224

Fritsche S, Bryan J, Kralstein D, Farley E (2011) The Bronx family treatment court 2005–2010: impact on family court outcomes and participant experiences and perceptions. Center for Court Innovation, New York

Goldkamp J, White MD, Robinson JB (2002) An honest chance perspective on drug courts: findings from drug court participant focus groups in Brooklyn, Las Vegas, Miami, Portland, San bernadino, and Seattle. Crime and Justice Research Institute, Philadelphia

Gottfredson DC, Kearley B, Najaka SS, Rocha CM (2005) The Baltimore city drug treatment court: 3-year outcome study. Eval Rev 29(1):42–64

Gottfredson DC, Kearley B, Najaka SS, Rocha CM (2006) Long-term effects of participation in the Baltimore city drug treatment court: results from an experimental study. J Exp Criminol 2(1):67–98

Gottfredson DC, Kearley B, Najaka SS, Rocha CM (2007) How drug treatment courts work: an analysis of mediators. J Res Crime Delinq 4(3):3–35

Green BL, Furrer CJ, Worcel SD, Burrus SWM, Finigan MW (2009) Building the evidence base for family drug treatment courts: results from recent outcome studies. Drug Court Review VI(2):53–82

Gutierrez L, Bourgon G (2009) Drug treatment courts: a quantitative review of study and treatment quality. Retrieved 30 Sept 2011 from the Public Safety Canada Web site: http://www.publicsafety.gc.ca/res/cor/rep/2009-04-dtc-eng.aspx

Hoffman MB (2000) The drug court scandal. North Carolina Law Rev 78:1439–1533

Hora PF, Schma WD, Rosenthal JTH (1999) Therapeutic jurisprudence and the drug treatment court movement: revolutionalizing the criminal justice system's response to drug abuse and crime in America. Notre Dame Law Review 74:439–537

Huddleston W, Marlowe DB (2011) Painting the current picture: a national report on drug courts and other problem-solving court programs in the United States. National Drug Court Institute, Washington, DC

Lipsey MW, Landenberger NA, Wilson SJ (2007) Effects of cognitive-behavioral programs for criminal offenders. The Campbell Collaboration, Oslow

Lutze FE, van Wormer JG (2007) The nexus between drug and alcohol treatment program integrity and drug court effectiveness: policy recommendations for pursuing success. Crim Just Pol Rev 18(3):226–245

Marlowe D (2012) Alternative tracks in adult drug courts: Matching your program to the needs of your clients: Part two of a two-part series. Alexandria: National Drug Court Institute

Marlowe DB, Kirby KC (1999) Effective use of sanctions in drug courts: lessons from behavior research. Natl Drug Court Inst Rev 2(1):1–31

Marlowe DB, Festinger DS, Lee PA, Schepise MM, Hazzard JER, Merrill JC, Mulvaney FD, McLellan AT (2003) Are judicial status hearings a key component of drug court? During-treatment data from a randomized trial. Crim Justice Behav 30:141–162

McCoy C (2003) The politics of problem-solving: an overview of the origins and development of therapeutic courts. Am Crim Law Rev 40:1513–1544

Mitchell O, Wilson DB, Eggers E, MacKenzie DL (2012) Assessing the effectiveness of drug courts on recidivism: a meta-analytic review of traditional and non-traditional drug courts. J Crim Just 40:60–71

National Association of Criminal Defense Lawyers (NADCL) (2009) America's problem-solving courts: the criminal costs of treatment and the case for reform. National Association of Defense Lawyers, Washington, DC

Office of Justice Programs (2012) Seven program design features: adult drug court principles, research, and practice. Washington, DC: Bureau of Justice Assistance, National Institute of Justice, and National Center for State Courts. http://research2practice.org/resources/

Ostrom B, Kauder N (1999) Examining the work of state courts, 1998: a national perspective from the court statistics project. National Center for State Courts, Alexandria

Rempel M, DeStefano CD (2001) Predictors of engagement in court-mandated treatment: findings at the Brooklyn treatment court, 1996–2000. J Offender Rehabil 33(4):87–124

Rempel M, Fox-Kralstein D, Cissner A, Cohen R, Labriola M, Farole D, Bader A, Magnani M (2003) The New York state adult drug court evaluation: policies, participants, and impacts. Center for Court Innovation, New York

Rossman SB, Roman JK, Zweig JM, Rempel M, Lindquist CH (eds) (2011) The multi-site adult drug court evaluation. The Urban Institute, Washington, DC

Salvatore C, Henderson MA, Hiller ML, White E, Samuelson B (2010) An observational study of team meetings and status hearings in a juvenile drug court. Drug Court Review 7(1):95–124

Schaeffer CM, Henggeler SW, Chapman JE, Halliday-Boykins CA, Cunningham PB, Randall J, Shapiro SB (2010) Mechanisms of effectiveness in juvenile drug court: altering risk processes associated with delinquency and substance abuse. Drug Court Review 7(1):57–94

Shaffer DK (2011) Looking inside the black box of drug courts: a meta-analytic review. Justice Quart 28(3):493–521

Tyler TR, Huo YJ (2002) Trust in the law. Russell Sage, New York

Wexler DB (1993) Therapeutic jurisprudence and the criminal courts. Wm Mary L Rev 35:279–299

Young D, Belenko S (2002) Program retention and perceived coercion in three models of mandatory drug treatment. J Drug Issues 22(2):297–328

Zhang A (2003) Drug and alcohol use and related matters among arrestees 2003. National Opinion Research Center, Chicago

Drug Courts' Effects on Criminal Offending

David B. Wilson[1], Ojmarrh Mitchell[2] and Doris Layton MacKenzie[3]

[1]Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA

[2]Department of Criminology, University of South Florida, Tampa, FL, USA

[3]Justice Center for Research, The Pennsylvania State University, University Park, PA, USA

Overview

Drug courts are problem-solving courts that divert generally low-level drug-involved offenders from conventional prosecution. These courts use their legal and moral authority to monitor offenders' abstinence from drugs and compliance with individualized treatment programs. This is done through regular status hearings before the court and routine drug testing. All drug court participants are expected to be engaged in some form of drug treatment. Successful complete of a drug treatment program is celebrated with a graduation ceremony and generally reward with dismissal of charges or expunging the initiating conviction.

This entry summarizes the findings from a recent systematic review and meta-analysis conducted by Mitchell et al. (2011) of the numerous studies that have examined the effectiveness of this criminal justice innovation. The systematic search identified 154 independent, eligible evaluations, 92 evaluations of adult drug courts, 34 of juvenile drug courts, and 28 of drunk-driving (DWI) drug courts. The findings most strongly support the effectiveness of adult drug courts, as rigorous evaluations find reductions in recidivism and these effects generally persist for at least 3 years. The evidence also cautiously suggests that DWI drug courts are effective in reducing recidivism, although few rigorous experimental evaluations have been conducted of DWI courts and the findings of these more rigorous evaluations are inconsistent. For juvenile drug courts, they found considerably smaller effects on recidivism than for adult drug courts but the evidence is suggestive of positive effects.

Background Description

Since they emerged in the late 1980s, drug courts have proliferated across the United States and more recently are being adopted elsewhere. Recent data indicate that there are over 2,400 drug courts in operation in the United States (Huddleson and Marlowe 2011). Other countries that have adopted drug courts include Canada, the United Kingdom, New Zealand, Australia, South Africa, Bermuda, and Jamaica (Berman and Feinblatt 2005).

Initially started entirely for nonviolent drug-involved adult offenders, the drug court model has grown to include specialized juvenile drug courts and DWI (driving while intoxicated) courts. The drug court model helped spawn a broader movement of therapeutic jurisprudence that includes a range of problem-solving courts, including mental health courts, prostitution courts, domestic violence, and others.

According to the National Association of Drug Court Professionals (1997, see also Hora 2002), there are several key components of a drug court: (1) early identification of eligible offenders; (2) a collaborative, non-adversarial, court processing; (3) regular judicial monitoring in the form of status hearings for drug court clients before the judge; (4) required drug treatment, (5) urine testing; (6) and the use of graduated sanctions for failure to comply with the court requirement and rewards for successful progress toward program completion. The specific treatment requirements are individualized to the needs of each client, and there is variability in how different drug courts implement these key components.

Eligible drug court clients are identified by prosecutors and diverted from traditional prosecution. The drug court client must, however, agree to the diversion to the drug court and the terms of drug court participation. The typical drug court client is a nonviolent drug individual charged with a drug or property offense, although some drug courts do accept individuals charged with or previously convicted of a violent offense. Most drug courts do not accept individuals with drug trafficking offenses. Thus, the ideal is that drug court clients are those individuals for whom drug use is presumed to play an important role in their criminal behavior.

As a non-adversarial judicial process, drug court clients must agree in some form to their guilt. One model is a pre-plea arrangement where no formal plea or other findings of guilt are formally entered but the client does stipulate that the facts of the case are correct, ensuring a successful prosecution if they are unsuccessful in the program. The case is dismissed if the participant completes the drug court. A post-plea/predisposition model requires a formal plea from the participant but delays sentencing. The case is dismissed for program graduates but sentence is imposed for those that fail. For more serious offenders, a post-plea/post-disposition model may be used. Under this approach, drug court participation is essentially a condition of probation and successful program completion is rewarded with a suspected sentence of incarceration, that is, he or she avoids a jail or prison sentence.

All drug court clients are required to participate in some form of drug treatment. A drug court may refer clients to a broad range of treatment services available throughout their community or may have a working relationship with a single treatment provider. The intensity of treatment varies and, ideally, is tailored to the needs of the client.

Drug court clients appear regularly before the judge for status hearings. These status hearings are crucial as it is here that the drug court judge and clients converse directly, and it is in these hearings where judges in collaboration with other court actors most clearly use the authority of the court to motivate treatment compliance. The court uses various rewards (e.g., praise, tokens of achievement, movement to the next phase of the program) and sanctions (e.g., increased treatment attendance or urine testing, short jail stays) to compel compliance to program requirements. Clients advance through three or more progressively less intense stages before completing the drug court, which typically takes about 12- to 18-months. Successful completion of the drug court is acknowledged with a formal graduation ceremony.

An early review of the evidence on drug courts conducted by the US General Accounting Office (GAO 1997) was unable to arrive at a firm conclusion, either for or against, regarding their effectiveness in reducing crime and drug use. In contrast, a 2001 review by Belenko drew a cautious but positive conclusion regarding the effectiveness of drug courts based on 37 evaluations. Both of these reviews, however, comment on the lack of evidence on outcomes measured post program. That is, most of the studies examined drug use and criminal behavior during the period of the drug court. Thus, it was not clear whether the reductions in drug use and criminal behavior would continue after graduation from a drug court.

Recent reviews have been more favorable. The GAO updated its review and concluded that drug courts were effective, at last during the program period, at reducing criminal behavior (U.S. General Accountability Office 2005). Furthermore, cost-benefit evaluations examined in this report also showed that drug courts yield a net benefit to society. The meta-analysis by Wilson et al. (2006) of 55 drug court effectiveness studies tentatively concluded that drug court participants have lower rates of recidivism than similar offenders who did not participate in drug courts. These findings held for evaluations that measured recidivism during and after program participation. Like the earlier reviews, these findings were tempered by the generally weak methodological rigor of the evaluations.

Evidence of Effectiveness of Drug Courts

This entry summarizes the findings of a recent meta-analysis and systematic review on the effectiveness of drug courts by Mitchell et al. (2011) published by the Campbell Collaboration (www.campbellcollaboration.org). This meta-analysis is an update of an earlier meta-analysis by the same authors (Wilson et al. 2006). This review examined the evidence of the traditional adult drug court and two newer adaptations of this model: juvenile drug courts and DWI drug courts.

Summary of Methods

The focus of the meta-analysis by Mitchell and colleagues (2011) was to synthesize all evaluations of the effectiveness of drug courts that used an experimental or quasi-experimental design and examined drug use or criminal behavior as an outcome. The evaluations must have compared a drug court (adult, juvenile, or DWI) to a comparison condition, such as probation. Studies that simply compared program graduates with program failures were excluded.

Mitchell and colleagues implemented a search strategy designed to find all published and unpublished evaluations that satisfied the

Drug Courts' Effects on Criminal Offending, Table 1 Mean odds ratio by type of recidivism measure

		95 % CI				
Outcome	Mean	Lower	Upper	*Q*	k^a	Tau^2
Adult drug courts						
General recidivism	1.66	1.50	1.84	442.19*	92	0.178
Drug recidivism	1.70	1.39	2.08	323.98*	42	0.368
Drug use	1.45	0.92	2.28	15.78*	4	0.165
Juvenile drug court						
General recidivism	1.37	1.15	1.63	66.31*	34	0.105
Drug recidivism	1.06	0.69	1.63	29.65*	14	0.357
Drug use	1.50	0.67	3.34	2.05	3	0.359
DWI drug court						
General recidivism	1.65	1.35	2.02	78.40*	28	0.159
Drug recidivism	1.59	1.22	2.09	16.84	14	0.054
Drug use	1.87	0.34	10.23	5.02*	2	1.227

*Distribution heterogeneous ($p \leq 0.05$)
[a] Number of odds ratios

above criteria. Their approach included keyword searching on numerous bibliographic databases and culling the reference list of prior reviews and eligible studies. In addition, websites of several organizations involved in promoting drug courts or organizations that conduct evaluations of drug courts were searched. The search was updated through August of 2011. This process yielded 181 studies that reported the results of 154 independent evaluations of drug court programs. See Mitchell and colleagues (2011) for a complete list of these references.

Results across studies were converted to odds ratios and these served as the basis for the meta-analysis. The odds ratio reflects the odds of success (or failure) in one group compared to another. In this case, the odds of success in the drug court versus the comparison. These were coded such that odds ratios greater than one reflected less criminal behavior or drug use among the drug court participants relative to the comparison condition and odds ratios less than one more criminal behavior. That is, values greater than one favor the drug court whereas values less than one favor the comparison.

Odds ratios were averaged across studies using the inverse variance weight method assuming a random effects model. This model gives greater weight to larger studies and assumes that the odds ratios vary both as a result of study level sampling error and true study level differences.

Summary of Results

Of the 154 evaluations included in the Mitchell and colleagues meta-analysis, 92 were evaluations of adult drug courts, 34 were of juvenile drug courts, and 28 evaluated DWI courts. Only 8 of these 154 evaluations were of drug courts outside of the United States (the other countries represented in the review included Australia, Canada, Guam, and New Zealand). The vast majority of these evaluations were relatively recent and have been conducted since 1999.

Mitchell and colleagues examined three different outcomes across the 154 studies: (1) general recidivism (typically measured as rearrest for any offense), (2) drug-related recidivism (typically measured as rearrest for a drug offense), and (3) drug use (usually measured as self-reported drug use or via urinalysis), for each type of court. Table 1 reports the mean odds ratio for each of these outcome types by each court type. All of these mean odds ratios favor the drug court over the treatment group, although the effect for the outcome of drug recidivism for juvenile drug courts is near a no different value. The pattern of evidence clearly favors the drug courts with most studies observing positive effects. For the general measure of recidivism, the percentage of studies observing positive effects was 88 % for adult drug courts, 70 % for juvenile drug courts, and 85 % for DWI courts.

These effects are easier to interpret if they are translating into simple percentages. For example, the mean odds ratio of 1.66 for general recidivism for the adult drug courts translates into a reduction in recidivism from 50 % (a common rate in the comparison conditions) to 38 % for the drug court participants. This is a small to moderate effect by most standards but arguably one of practical and policy significance. The comparable percent success rates for general recidivism for juvenile courts and DWI courts are 42 % and 38 %, respectively. All three of the mean odds ratios for the most general recidivism measure reported by each study are statistically significant.

An effect of drug courts on future criminal behavior is presumed to be mediated by reductions in drug use and drug dependence. That is, the core logic of drug courts is a presumed connection between drug use and more general criminal behavior. Thus, it is surprising that roughly half of the evaluations of drug courts did not report results specific to drug use. For both adult drug and DWI courts, the results for drug-specific measures of recidivism (rearrest for drug-related offenses) are quite similar to those for general measures of recidivism. The more direct measures of drug use, such as self-report or urinalysis, also produce roughly similar results for these two court types. The latter effects were not statistically significant; however, given the small number of studies contributing to the mean (and correspondingly large confidence interval), this result is not surprising.

In contrast, the pattern of mean odds ratios for juvenile drug courts across the three outcome types is counter-intuitive. The mean odds ratios are positive and statistically significant for general recidivism and more direct measures of drug use but essentially a value of no difference for drug-related measures of recidivism. There is not clear explanation for this pattern other than the smaller number of studies evaluating drug use outcomes.

Prior reviews of the drug court literature were critical of the high percentage of evaluations that only examined effects during the period of program enrollment. To address this concern, Mitchell and colleagues examined effects separately for different follow-up periods and clearly differentiated effects that overlapped either partially or completely with the period of program enrollment.

More specifically, they first compared results measured during the first, second, third, and fourth years post entry into the program using all available studies at each time period. Next they examined only those studies reporting results at 12- and 24-months post program entry and also those studies reporting results at 12-, 24-, and 36-months post program. This was only done for adult drug courts as there was insufficient data for these analyses for the juvenile and DWI courts. The analyses showed that the effects of adult drug courts on recidivism are not limited to the short term. Rather, the research suggests that drug court participants have reduced recidivism during and after drug court treatment, and these effects appear to last at least 3 years post-drug court entry.

Although the overall findings are encouraging and suggest that drug courts are generally effective, the findings were highly heterogeneous across studies. That is, there was great variability in the size of the effects across the different studies. This variability may be due to the numerous methodological differences across the studies or due to actual difference in the structure and design of the drug court.

Mitchell and colleagues examined the influence of numerous methodological features on study outcomes. Of greatest concern was whether selection bias might account for the overall positive findings. Because the vast majority of the studies included in this meta-analyses were quasi-experimental (i.e., did not randomly assign participants to the drug court and comparison conditions), it is plausible that naturally occurring differences between the groups might upwardly (or downwardly) bias the effectiveness estimates. For example, a selection mechanism resulting in the drug court program having more highly motivated individuals would result in a positive odds ratio even if the treatment was not effective. The evaluations included in this meta-analysis included studies with a low likelihood of selection bias, that is, experimental studies with random assignment to conditions, and studies with a high likelihood of selection bias, that is, studies with included drug court eligible offenders who elected not to participate.

Studies were grouped into four categories. The most rigorous category was randomized experiments. The second was high quality quasi-experimental designs. These rigorous quasi-experiment studies typically used subject-level matching on key variables or propensity score matching to ensure comparability of clients in each condition. The third category was standard quasi-experimental designs. These typically used either a historical comparison group that met drug court eligibility criteria constructed from archival data or a group of offenders who were eligible but not referred to the drug court program. The critical feature here is that the participants did not self-select into the drug court or comparison condition. The fourth and weakest category of studies was quasi-experimental designs that involved comparing drug court clients to drug offenders who were eligible for the drug court but declined participation (drug court refusers).

The evidence suggests that the overall mean results are at least somewhat positively biased. The weakest quality studies did, on average, produce the largest results across the three types of courts. Thus, it is important to focus more closely on the findings from the higher quality studies.

For adult drug courts, Mitchell and colleagues concluded that the more rigorous studies still provided evidence of general effectiveness. This finding was complicated by the inconsistent results across the three randomized experimental studies of adult drug courts. They concluded, however, that all three experimental evaluations of adult drug courts provide evidence of these courts' effectiveness in reducing recidivism. Two of the three evaluations find recidivism reductions in the first year after program entry, and the remaining evaluation finds a recidivism reduction at 3 years post-program entry. Furthermore, the vast majority of the rigorous quasi-experimental evaluations of drug courts find moderate reductions in general recidivism.

For juvenile drug courts, the picture is more complicated. The average effects for the 11 rigorous quasi-experimental studies and for the single experimental study were positive. Unfortunately, these effects were small overall (ranging from a mean odds ratio of 1.22 to 1.39 for general and drug-specific measures of recidivism) but not statistically significant. Thus, the most rigorous evidence suggests that juvenile drug courts have small effects but clearly additional research is needed to firm-up this conclusion.

For DWI drug courts, the rigorous quasi-experimental studies had a large and statistically significant mean odds ratio for both general recidivism and drug-specific recidivism (1.99 and 1.78, respectively). Unfortunately, the mean odds ratio for the four randomized studies was very small and statistically nonsignificant. However, Mitchell and colleagues noted that three of these four evaluations reported positive effects but that a single influential negative finding seriously attenuated the overall mean odds ratio. Thus, Mitchell and colleagues concluded the evidence of effectiveness for DWI courts was encouraging but that clearly additional high quality studies are needed.

As already discussed, drug courts vary substantially in how they implement the various key elements. Mitchell and colleagues explored numerous differences in the structure of the drug courts across studies and whether these differences were related to the observed effects. These analyses were guided by a framework proposed by Longshore et al. (2001). This framework has five dimensions and predicts that the most effective drug courts (1) will use the court's leverage (rewards and sanctions) to motivate offender change, (2) will serve populations with less severe problems, (3) will have high program intensity, (4) will apply rewards and sanctions predictably, and (5) will emphasize offender rehabilitation as opposed to other court goals like speedy case processing and punitive sanctioning.

Unfortunately, it was not possible to categorize the studies on all of these dimensions. Mitchell and colleagues were able to code variables that related to the first three dimensions but not the last two. Furthermore, the categorizations were often rather crude and were unable to cleanly differentiate the studies on each dimension.

The first dimension is the use of sanctions and rewards or what Longshore and colleagues refer to as leverage. To examine leverage, Mitchell and colleagues coded studies by the method of case disposition (pre-plea, post-plea, or mixed) and what happens to the charges/sentences upon graduation (dismissed/expunged or not dismissed/expunged). Mitchell and colleagues found mixed support for the leverage hypothesis. Consistent with the hypothesis, drug courts that dismiss/expunge charges upon graduation had higher mean odds ratios on both measures of recidivism. This difference is statistically significant for the drug recidivism measure, but is nonsignificant for the general recidivism measure. Counter to their hypothesis, however, drug courts that predominantly used post-plea case processing did not have greater reductions in recidivism than other courts. Post-plea courts are presumed to have greater leverage because offenders have been convicted and face immediate sentencing upon failure in drug court whereas the in the pre-plea model, the case must still be processed.

In practice, what may matter most are not these structural differences but the certainty and swiftness of sanctions and rewards. Mitchell and colleagues were not able to categorize courts on this dimension based on the information provided in the evaluation reports. Evidence supporting the importance of the certainty and swiftness of sanctions is emerging from studies of Project HOPE [see related encyclopedia entry], a variant on the drug court model. The key feature of Project HOPE is the certainty and swiftness of sanctions, typically a night in jail, for failure to remain drug free (i.e., a dirty urinalysis) or adhere to a simplified set of terms and conditions of probation.

The second dimension of Longshore and colleagues framework is the intensity of the program. Mitchell and colleagues categorized the studies on several variables designed to address this dimension. These included the length of the program, frequency of drug testing, frequency of status hearings, and number of phases through which the clients progressed.

The analyses did not generally support the hypothesis that more intense programs are associated with greater reductions in recidivism. The one exception was for analyses involving the number of status hearings. For the adult drug courts, courts with more than two status hearings in first treatment phase had larger effects of drug-related recidivism than other courts. The number of status hearings was also positively related to improved outcomes for juvenile drug courts. Other than the number of status hearings, these findings suggest that drug courts with greater program intensity are no more effective in reducing recidivism than other courts. Once again, however, these analyses do not fully capture the differences in the intensity of the drug treatment provided by the various drug courts, nor do they capture the differences in the quality of drug treatment services.

The third dimension of Longshore and colleagues' framework posits that drug court programs that serve less severe populations will be most effective. This hypothesis is in conflict with a core element of Andrews and colleagues' (1990) principle that more effective programs serve more severe populations. This "risk principle" has found empirical support (see Andrews and Bonta 1992). Mitchell and colleagues examined this issue by categorizing studies by whether individuals with violent criminal histories or with extensive prior convictions were eligible for the program.

Analyses of the general recidivism effect sizes support Longshore and colleagues' prediction. Specifically, samples that included only nonviolent offenders had statistically larger mean odds ratios on the general recidivism measure than samples that included violent offenders. Similarly, samples with minor criminal history had larger mean odds ratios than evaluations based on samples with more criminal history; however, this difference was not statistically significant. These findings were not replicated in the analysis of the drug-related recidivism outcome measure. Here there were no differences in the mean effects odds ratios on either of the measures of population severity. For the juvenile and DWI courts, there were no meaningful patterns across these categorizations. Thus, these findings are more supportive of Longshore and colleagues' hypothesis than Andrews and colleagues' risk principle although the finding is clearly weak and inconsistent across the different types of drug courts.

Key Issues and Controversies

The rapid expansion of drug courts has been based on a presumption that they are effective and on the local problems that they solve, such as jail overcrowding. The approach also resonated with the get-tough approach to crime that has dominated crime policy during that past quarter century. Drug courts also help address the problem of increasing prison and jail populations by diverting a portion of drug-involved offenders toward community sanctions and treatment. Thus, the drug court model satisfied both get-touch and rehabilitation ideologies, allowing for a broad base of support that facilitated adoption of the approach.

This general belief in the value of drug courts is not without detractors. For example, a recent report by the Justice Policy Institute (Walsh 2011) questioned our "addiction" to drug courts. The report does not question the basic effectiveness of drug courts but argues that there are unintended consequences to our use of them as a solution to what is essentially a public health problem: drug addiction. First, the report argues that drug courts have expanded the involvement of the criminal justice system into the lives of low-level offenders, or what is often called, net widening. Many of the individuals who enter a drug court program would have had their cases dismissed or been diverted to a community treatment program prior to the development of drug courts. Second, the report argues that criminal sanctions for individuals who fail in the drug court are often greater than what they would have received through conventional processing. Third, the report argues that a criminal conviction should not be a criterion for low-income and other disadvantages groups to gain access to needed drug treatment. And finally, the report argues that these negative consequences often affect minority groups more severely, both because they are less likely to be eligible given higher rates of offense histories that preclude eligibility and through higher failures rates for those admitted. The report argues for strengthening community-based treatment services for drug-involved individuals that do not require a criminal conviction and diverting many of the existing drug court clients to these community programs.

These criticisms of drug courts are specific to drug courts within the United States and reflect the broader ecosystem within which they operate. Addressing these criticisms would require larger changes within this ecosystem, primarily a major policy shift from a criminal justice approach to drug addiction to a public health approach. In the absence of such a shift, these courts are likely to continue to serve as an important means of facilitating access to drug treatment and improving treatment compliance through judicial sanctions.

Conclusions

The meta-analysis by Mitchell and colleagues draws generally positive conclusions regarding the effectiveness of drug courts for reducing criminal behavior and drug use. The vast majority of the programs that evaluate the effectiveness of these programs find that participants have lower rates of recidivism than nonparticipants. The analysis suggests that the typical effect is roughly a reduction from 50 % to approximately 38 % for adult drug courts.

For juvenile drug courts, the effects appear to be smaller, roughly a reduction from 50 % to 43.5 %. The findings for DWI courts are promising but not unambiguous, given the mixed findings and limited number of studies.

The findings are not without caveats. The number of randomized controlled trials (true experiments) is limited. These high quality studies do indicate, however, that drug courts can work. The high variability in effects across studies suggests differential effectiveness across variations in drug courts. That is, some courts may be highly effective whereas others may be ineffective. The meta-analysis by Mitchell and colleagues explored several hypotheses regarding differential effectiveness and found limited support for differential effects based on the leverage of the court. Additionally, there was weak support that drug courts will be more effective with less severe populations.

Despite the rather large number of studies examining the effectiveness of drug courts, this body of literature provides limited insight into the critical components of effectiveness. It is likely that the quality of the drug treatment services provided to drug court clients and the degree to which these treatments are evidence based are related to program's success. Furthermore, the findings from deterrence research in criminology and from behavioral psychology would imply the importance of certainty and swiftness in the application of judicial sanctions and rewards. The collection of studies reviewed by Mitchell and colleagues provides limited information on the role of these important features of the drug court model. Clearly, future research on the relationship between drug court elements and effectiveness is needed.

Related Entries

- ► History of Substance Abuse Treatment
- ► Problem-Solving Courts
- ► Reentry Courts
- ► Rehabilitation

Recommended Reading and References

Andrews DA, Bonta J (1992) The psychology of criminal conduct. Anderson, Cincinnati

Andrews DA, Zinger I, Hoge RD, Bonta J, Gendreau P, Cullen FT (1990) Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. Criminology 28(3):369–397

Berman G, Feinblatt J (2005) Good courts: the case for problem-solving courts. New Press, New York

Hora PF (2002) A dozen years of drug treatment courts: uncovering our theoretical foundation and the construction of a mainstream paradigm. Subst Use Misuse 37:1469–1488

Huddleston CW III, Marlowe DB, Casebolt R (2008) Painting the current picture: a national report card on drug courts and other problem-solving court programs in the United States. National Drug Court Institute, Washington, DC

Longshore D, Turner SW, Morral A, Harrell A, McBride D, Deschenes E et al (2001) Drug courts: a conceptual framework. J Drug Issues 31(1):7–26

Mitchell O, Wilson DB, Eggers A, MacKenzie DL (2011) Drug courts' effects on criminal offending for juveniles and adults. Campbell Syst Rev. http://www.campbellcollaboration.org

National Association of Drug Court Professional (1997) Defining drug courts: the key components. Bureau of Justice Assistance, Washington, DC

U.S. General Accountability Office (2005) Adult drug courts: evidence indicates recidivism reductions and mixed results for other outcomes. United States General Accounting Office, Washington, DC

U.S. General Accounting Office (1997) Drug courts: overview of growth, characteristics, and results. United States General Accounting Office, Washington, DC

Walsh Natasssia (2011) Addicted to courts: how a growing dependence on drug courts impacts people and communities. Justice Policy Institute, Washington, DC. http://www.justicepolicy.org

Wilson DB, Mitchell O, MacKenzie DL (2006) A systematic review of drug court effects on recidivism. J Exp Criminol 2(4):459–487

Drug Dealing Places

► Spatial Perspectives on Illegal Drug Markets

Drug Distribution

► Drug Trafficking

Drug Enforcement

Nicholas Corsaro
School of Criminal Justice, University of
Cincinnati, Cincinnati, OH, USA

Synonyms

Drug markets; Law enforcement; Policing

Overview

The US Office of National Drug Control Policy (ONDCP) has consistently allocated between US $12 billion to U2003S$17 billion per year to fund the Federal Drug Control Budget since the turn of the century (ONDCP 2011). In addition, the cumulative costs of drug law enforcement for federal, state, and local governments has been estimated to exceed US$35 billion annually (Boyum and Reuter 2005). These budgetary figures illustrate a persistent commitment for the continued "war on drugs" waged by different agencies within the law enforcement community to investigate, build cases, arrest and ultimately incarcerate individuals involved in illicit drug distribution.

Given this clear enterprise, it is important to understand the different practices and procedures utilized by drug law enforcement officials, including: (a) investigation techniques, (b) street-level enforcement, and (c) problem-oriented policing strategies designed to disrupt drug markets. Within this chapter, a variety of different drug enforcement approaches will be discussed in detail. In addition, potential drug law enforcement effectiveness and crime reduction impact will be examined within each sub-section, where appropriate and applicable.

Drug Investigations

In the past 20 years plus, there has been a shift away from reactive (i.e., unfocused and reactionary) policing toward more proactive police tactics. Proactive policing typically refers to self-initiated activities during uncommitted patrol time. A cornerstone role of drug law enforcement is to identify and eventually build cases against illicit drug offenders. A common practice of drug investigators is to engage in covert operations through undercover policing, and to rely on the use of confidential informants. In this section, these tactics and different policing approaches will be discussed in extensive detail.

Girodo (1985) notes that undercover investigations are typically either strategic or tactical in nature; strategic investigations are typically referred to as undercover probes that are used to target a group or geographic area in addition to standard surveillance activity. Tactical investigations target specific individuals, seek evidence regarding their intentions to commit a specific crime, and are geared toward eventual prosecution. In this case, the main purpose of conducting an extensive undercover drug investigation is to gather convincing evidence in order to meet the required standard of evidence (i.e., proof beyond reasonable doubt) needed for a conviction at trial (Sherman 1987, p. 88). In terms of a time-line, Girodo (1985) notes that undercover drug investigations can occasionally last a day or two, but most often occur over several weeks to a few months; and, in some cases can last years.

However, the use of undercover surveillance is not always clear-cut and certainly not without its controversies. Gary Marx (1988) illustrated that undercover policing has both intended and unintended consequences. Among other points of concern, Marx (1988) laments that individuals targeted by an investigation can be deceptively provoked to commit a crime. The use of coercive undercover techniques (i.e., coercion or temptation) by law enforcement has the potential to erode public confidence in police, particularly when they are used to target the politically affluent or those from marginalized communities. Thus, the short-term benefits of a single arrest will most likely not outweigh lingering doubts about procedural unfairness. Many have argued the lack of control over undercover policing should promote more stringent regulations.

Ultimately, legitimacy through the rigorous case-building and procedurally fair (i.e., uncorrupt) investigations that promote public safety and foster crime prevention is the ideal balance, which serves as the primary goal of drug investigations.

Marx (1988) also illustrated that effectiveness in undercover policing can be viewed as a paradox. Success or failure can be defined both as an arrest (because in this case a crime has been detected and justice pursued) as well as a non-arrest (i.e., the absence of an arrest may indicate a deterrent effect). Thus, evaluating the overall efficiency and utility of undercover investigations is an arduous task.

Another important aspect of investigative work by drug law enforcement is the use of confidential informants (CIs), which are usually individuals that are networked within the local drug trade. CIs typically assist police in their investigations, often in exchange for leniency. James Q. Wilson (1968) noted that the use of CIs often centers on the following primary functions: providing leads, casework facilitation, and occasionally testifying in court. Certainly, a great deal of research exists regarding the tactical components of covert drug investigations (see Miller 1987). Related specifically to illegal drug investigations, CIs are considered the ‘sine qua non’ of narcotics work.

In terms of outlining a framework of drug investigations, the primary function of CIs is to assist investigations by providing detailed information about the nature of drug market networks, which usually requires at least some intimate knowledge of the relationship between users and dealers. In a study that relied upon interviews with 24 undercover officers, Jacobs (1997) outlined a “contingent tie” perspective (i.e., perceptual bonds that may be either weak or strong depending on context and history between individuals) that can be used to explain the social-link between users and dealers, as well as police with CIs. In essence, the reliance on CIs as an investigative tool requires multiple steps and social interactions between drug investigators and potential CIs.

Jacobs (1997) also illustrated that police often first attempt to establish these contingent ties by purchasing drugs directly from potential lower-level distributers (e.g., pretending their ‘usual’ dealer was unable to provide their normal service and a single purchase would serve as a favor to the undercover officer). Then, investigators use threats of a potential arrest and likely conviction to encourage cooperation among these primary actors in order leverage them to serve as an intermediary informant with others involved in drug market networks. In addition, fictitious relationships between officers and CIs usually need to be established, which usually fall under one of three forms: jailhouse buddies (e.g., police and CIs pretend to know one-another from jail or prison), co-workers (e.g., acting as though they were introduced to one-another through work), or subcultural associates (e.g., conveying they have been long-term offending “colleagues”).

A failure to protect CIs (i.e., “burning informants”) as part of their investigation makes it virtually impossible to replenish this vital fact-finding resource. Jacobs (1997) noted that drug investigators will likely “cut out” (i.e., remove) CIs from their investigation by complaining about unfair (and fictitious) taxation directly to dealers. For example, undercover officers often ask dealers to engage in future transactions directly with them in order to “cut out” the mid-level person – who happens to also be the CI. This type of deception is used to capitalize on the CIs relationship with others involved in drug networks without putting them in the middle of the investigation, past the onset point. In addition, Manning (1977) illustrated that the reliance on CIs is usually temporary and requires a continued commitment given the high-turnover, and replacement found in illicit drug networks. Thus, law enforcement officials must be invested perhaps more in protecting rather than exploiting their relationship with CIs, particularly over the long-term.

Ultimately, investigations are a preliminary step in terms of initiating law enforcement driven approaches to disrupt illicit drug networks and open-air drug markets. While some have argued there should be additional constraints placed on the process of investigations, most researchers regard undercover investigations and the use of confidential informants a vital component in drug investigations. After a thorough inquiry is conducted, the next traditional step for police officials is to engage in street-level enforcement, which can take multiple approaches and has a variety of sensitive issues that are typically considered.

Street-Level Drug Enforcement

In a systematic review of over 130 street-level drug enforcement strategies, Mazerolle et al. (2007) synthesized their findings into a finite number of approaches that do not necessarily require interagency cooperation between police and other criminal justice organizations (i.e., those that tended to be "police-only"), some of which are discussed in detail here, including: drug seizures; crop eradication; crackdowns; raids; and search and seizures.

Drug seizures are often referred to as "supply-reduction" strategies that generally center on reducing trafficking, dealing, and the amount of drugs available in a specific geographic area. The seizure of drugs can take place during a routine investigation (described above) or can occur as part of a larger intervention or suppression effort. While evaluative research on seizures of general drugs is rather scant, a majority of evaluations have focused on the impact of heroin-based seizures. Wood et al. (2003) reported a non-significant association between police-led heroin seizures and price, availability, drug-related deaths (including overdoses), as well as overall rates of crime and arrest. Little empirical research exists demonstrating a tangible relationship between drug supply reduction strategies and crime associated with illegal drug markets. This is not surprising given that five-sixths of the crime and violence associated with illegal drug distribution is motivated primarily by drug money while only one-sixth is motivated by drugs themselves (see Caulkins and Reuter 1998). Thus, supply reduction strategies should also be evaluated, in part, on their periphery influence regarding financial increases within drug markets brought forth by street-level enforcement.

Another supply-reduction approach is crop eradication. Research on crop eradication has primarily focused on the removal, confiscation, and destruction of cannabis. Overall effectiveness is difficult to measure, primarily given the findings by Potter, Gaines, and Holbrook (1990) which illustrates how growers and distributers often adapt to the increased attention by police (e.g., moving cannabis plants from outdoors to indoors). Cost-effectiveness is also difficult to determine given the extensive amount of time devoted to most crop reduction strategies.

To this point, the drug law enforcement approaches discussed here have generally been restricted to investigations, undercover work, the use of CIs, and supply-reduction strategies. However, street-level drug enforcement can also involve both undercover and uniformed officers, and thus are usually very highly visible to the surrounding community. This is particularly true concerning the use of police crackdowns and raids. Sherman (1990, p. 1) defined a police crackdown as "a sudden increase in officer presence, sanctions, and the threat of apprehension either for specific offenses or for all offenses in specific places." In offense-specific crackdowns, police focus on a particular type of activity, such as public drunkenness, prostitution, and drug dealing. Location-based crackdowns are aimed at reducing crime in a targeted area. Again, Mazerolle et al.'s (2007) review of the evaluation literature shows that the vast majority of drug enforcement crackdowns generally target street-level drug markets (particularly focusing on "harder" drugs such as crack, cocaine, and heroin); however, some crackdowns also focused on drug distribution within indoor locations – including both residential and commercial sites (e.g., motels/hotels, clubs, etc.).

A synthesis of evaluation-based research suggests that crackdowns produce widely varying results in terms of altering drug and crime problems. Roughly 1/3 of the evaluations indicated a positive effect (i.e., a reduction in drug problems), 1/3 demonstrated no discernable effect, and 1/3 actually showed evidence of a negative effect (i.e., an unintended increase in drug problems). Of those studies that illustrated a reduction in drug problems, the following were consistent themes observed across the evaluations: (a) the reductions in crime were more likely observed with property and violent offenses rather than drug-specific incidents; (b) the impact appeared to be short-term; and, (c) potential effects were more likely observed in geographically contained areas near bridges, rivers, and borders where displacement was less likely to occur due to natural barriers (Mazerolle et al. 2007).

While crackdowns are described as a sudden increase in police presence, raids are equally intensive but are usually directed within specific indoor locations (e.g., a single residence within an apartment complex or a single bar). Like crackdowns, police raids appear to have a short-term impact on reducing the supply of drugs as well as crime and disorder problems. Cohen, Gorr, and Singh (2003) showed that police enforcement activities that target drug dealing in nuisance bars (i.e., drug hotspots) have the capacity to have a substantive impact on levels of drug dealing, at least during those periods of active raid-based enforcement. In the Cohen et al. (2003) study, there were distinct and active periods of police raids in nuisance bars conducted by the narcotics squad, followed by back-off periods. The authors' note that drug dealing in street markets that are associated with nuisance bars show considerable resiliency after raid and suppression strategies are withdrawn by law enforcement. However, over reliance on raids and crackdowns also has the potential to alienate even the most cooperative residents within high-crime areas, which can have long-term repercussions from a community crime perspective.

Problems with Intensive Street-Level Drug Enforcement

The continual use of police crackdowns, raids, and enhanced searches and seizures in order to control drug related problems in high-crime areas often comes at a tangible cost. Personal contact with police is perhaps the most important factor that shapes individuals' perceptions of law enforcement. Cao et al. (1996) also illustrated that community context plays a major role in terms of residents' appraisals of law enforcement. Citizens in disadvantaged and high-crime areas are often skeptical, concerned, and in most cases resistant to potentially intrusive police practices.

In terms of the intersection between race and community context, Weitzer (1999) found that Blacks and Whites in middle-class neighborhoods were less likely to perceive police to be abusive or to engage in enforcement-based misconduct than were Black residents in lower-class neighborhoods. In many cases, police seem to behave differently depending on the type of neighborhood they patrol. Terrill and Reisig (2003) found police were more likely to use force in disadvantaged neighborhoods on suspects. Similarly, Kane (2005) found police were more likely to engage in misconduct in distressed communities. Instances of over-policing (i.e., aggressive enforcement) and under-policing (i.e., lack of concern) can erode public confidence in formal mechanisms of social control.

However, research also consistently indicates a desire for integrative law enforcement responses to crime and drug problems in high-risk geographic contexts. In a study of the perceptions of Black residents in high-crime areas, Brooks (2000) found a paradox in that many African-Americans residing in poor neighborhoods and who were disproportionately affected by high-crime problems often associated with illicit street drug markets were frustrated with levels of crime, desired tough legal enforcement, but were also fearful of police. Carr et al. (2007) similarly found that youths from high-crime communities who were negatively disposed to police also were overwhelmingly in support of increased and tougher law enforcement approaches. Finally, in an interesting twist, Cooper et al. (2005) conducted a study that relied upon interviews with 40 drug-injecting residents and found that "study participants lauded the reductions in public drug activity that constant [police] monitoring produced" (Cooper et al. 2005, p. 681). Thus, there appears to be a constant tension in terms of maintaining a balance between levels of enforcement that are perceived appropriate by local residents contrasted with efforts that are viewed as excessive and unwarranted.

To illustrate, rapid response time by police are largely unrelated to on-scene arrests (some contend fewer than 29 for every 1,000 cases). However, expedient citizen reporting time significantly influenced the likelihood of on-scene arrests. Thus, outcome-driven processes that substantively improve police effectiveness are more important than simply focusing on traditional organizational police processes.

Integrative policing approaches that rely upon a diverse set of strategies as well as neighborhood residents to assist in the performance of their duties are also more likely to stimulate cooperation when citizens' view the police favorably and with legitimacy. Strategies that promote positive pro-social aspects of neighborhood integration appear to have a mediating effect on targeted neighborhood crime outcomes. In the next section, more extensive detail is provided for problem-oriented policing applications that have shown considerable promise in reducing drug related incidents.

Problem-Oriented Policing (POP) Strategies

Beginning in the 1980s, there was a groundbreaking shift in policing strategies that saw the advent of new approaches to law enforcement, including hotspots policing (i.e., strategically focusing on small geographic places to reduce high levels of crime), community policing (i.e., law enforcement approaches that promote community engagement) and problem-oriented policing (i.e., problem solving through the use of a variety of analytical approaches). Concerns with effectiveness and evaluation are consistent with the problem-oriented policing (POP) framework introduced by Herman Goldstein (1990).

Morris and Heal (1981) also advocated the use of "situational policing" as a potential approach to ensure the assimilation between various communities interests with police-led crime prevention activities. The POP model calls for police to use scanning, analysis, response, and assessment (SARA) in order to determine and define crime problems (i.e., root-causes) rather than reacting to crime incidents (i.e., symptoms). Goldstein (1990) provided an "inventory" of actions that he argued should be employed by law enforcement interested in adopting the problem-oriented model. Some key components of Goldstein's (1990) inventory included: concentrating on high-call locations; integration with additional government and private agencies; use of mediation (rather than simply using arrest-only); mobilization and collaborating with the local community; using civil and criminal law to control public nuisances; and, opportunity reduction (see also Clarke, 1997).

Eck (2006) more recently framed problem-oriented policing around three major principles: empirical, normative, and scientific. The empirical principle states that the public demands police handle a diverse range of crime problems, and that they are not invested in specifically how police handle such problems (i.e., arrest need not be the only "tool" available to police). The normative principle illustrates that police should address and ultimately reduce crime problems rather than simply respond to incidents. Finally, the scientific principle asserts that police should use scientific evidence to identify and address core crime problems. In essence, the tools officers utilize to address crime problems such as illegal drug distribution should be analytical, unique, and crafted specifically to each problem.

Mazerolle et al. (2006) conducted a rigorous meta-analytic review of street-level drug law enforcement strategies. The study assessed the relative effects of different approaches by comparing problem-oriented policing (outlined above) with community-wide policing, hotspots policing, standard (i.e., unfocused) law enforcement efforts. Effect size comparisons (i.e., changes in targeted outcomes) indicated that problem-oriented policing programs and geographically-based interventions that involved cooperative partnerships between police and third parties (e.g., community members, social service providers, additional government agencies, etc.) tended to be far more effective at controlling drug problems than traditional and law-enforcement only interventions. Mazerolle et al. (2006) concluded that policing initiatives that forge cross-agency and cross-community partnerships had the most pronounced and sustained effect on drug crime problems.

Recent development in POP drug enforcement strategies have focused on the use of "pulling levers" focused deterrence initiatives to disrupt the crime and violence associated with street-based drug markets. Pulling levers policing relies on actors across multiple agencies (i.e., police, prosecution, probation and parole, social service providers, and community leaders) to use data-driven approaches in order to target repeat and chronic offenders within high-crime locations (Goldstein 1990). Pulling levers strategies are designed to utilize specific deterrence in order to inform chronic and persistent offenders of the sanctions that are specifically available to criminal justice officials that can be used to obtain leverage and ultimately reduce recidivism (Kennedy 1997).

The initial pulling levers policing strategy that used this type of focused deterrence framework occurred in Boston where citywide levels of youth homicide significantly declined by roughly 66 % after implementation (Braga et al. 2001). A number of sites have replicated the approach, and a series of evaluations have provided further evidence of a significant violent crime impact. While the majority of these strategies and subsequent evaluations have focused predominantly on reducing youth, gang, and gun violence, officials have recently begun adapting the approach to disrupt levels of crime and violence associated with street level drug markets (Kennedy 2009).

Officials in High Point, North Carolina were the first to use the problem-oriented policing model while also incorporating a pulling levers intervention in order to disrupt neighborhood crime problems facilitated by local drug markets. More specifically, they combined the threat of enhanced sanctions to targeted individuals that were identified through a detailed drug investigation in order to emphasize the seriousness of their message through the use of a public neighborhood call-in session. In addition, the public communication strategy was designed to demonstrate interagency and community cooperation and encourage improved social service provisions for desisting offenders in an effort to permanently disrupt several open-air drug markets across the city (Kennedy 2009). Initial results from High Point indicated significant declines in drug and violent offending as well as perceptional changes among residents in target areas (Kennedy 2009). Replications of the High Point drug market program evaluated in Nashville, TN (Corsaro et al. 2010) and Rockford, IL (Corsaro et al. forthcoming) also demonstrated similar promise. Thus, the use of specific deterrence combined with detailed drug investigations appears to warrant additional attention as a potential mechanism to disrupt open-air drug markets. However, we also note that POP strategies are not always universally successful.

Problems with Problem-Oriented Policing

While a variety of promising POP strategies were presented here in terms of impacting drug related crime, it is also important to note that in practice POP has a number of limitations that should be carefully considered. Eck (2006) contends that problem-oriented policing was originally designed to be held within police headquarters, though in reality decentralization occurred where problem-identification and subsequent problem-solving tasks were shifted to police officers within operational units.

In addition, Cordner and Biebel (2005) conducted an organizational study of POP in practice within an organization that appeared to be devoted to the POP model. Study results indicated that patrol officers often engage in activity that could be defined as problem-solving; however, these same patrol officers tended to concentrate their efforts on symptoms of greater neighborhood problems (e.g., drug deals) and indeed very rarely formally used the SARA model in order to disrupt central community crime problems (e.g., personal crime, property crime, traffic violations). Cordner and Biebel (2005) found that while patrol officers engaging in problem-solving were thoughtful and analytical in their various approaches to specific neighborhood crime problems, they rarely engaged in a full range of POP activities since these require extensive data analysis, integrative and dynamic approaches, and evidence-based evaluations. In essence, the true POP model requires more time, effort, and energy than many officials are willing to invest.

Conclusion

Boyum and Reuter (2005) estimated that the imprisonment risk per drug transaction in the USA was about 1 in 4,500 for drug sellers that average roughly 1,000 transactions annually. In addition, they calculated the risk of imprisonment for distributing 0.2 g units of cocaine to be less than 1 in 15,000 per sale for distributers. When coupled with the combined state and federal annual budget averages that range between an estimated $30 and $35 billion per year, objective readers will likely question the efficiency, practicality, and sustainability of current criminal justice approaches to street level drug law enforcement.

In this chapter, a detailed review of drug enforcement strategies, history, processes, and evaluations of effectiveness were presented. Given that illicit drug networks and open-air drug markets appear to be unique across different locations, there does not seem to be a single catch-all drug law enforcement approach that solves the laundry list of drug crime problems. However, proper steps and continual commitment geared toward drug investigations, street level enforcement, community cooperation, and long-term commitment to solving drug crime problems has shown promising results.

In essence, effective POP strategies are built upon the total range of drug enforcement strategies discussed here. Investigations are a single step in the process; the use of CIs helps build stronger cases; suppression policing directed at specific individuals or within specified locations can have a short-term impact; inter-agency cooperation helps drive reductions in drug related crime problems; community notification sessions are designed to improve police-community relationships; and community informal social control helps sustain potential positive drug crime reduction gains. When the cumulative efforts of these drug enforcement strategies are present when policing illegal drug markets, the evidence suggests that crime reduction benefits are more likely to occur.

Related Entries

- ▶ Problem-Oriented Policing
- ▶ Spatial Perspectives on Illegal Drug Markets

Recommended Reading and References

Boyum D, Reuter P (2005) An analytic assessment of U.S. drug policy. AEI Press, Washington, DC

Braga AA, Kennedy DM, Waring EJ, Piehl AM (2001) Problem-oriented policing, deterrence, and youth violence: an evaluation of Boston's operation ceasefire. J Res Crime Delinq 38:195–226

Brooks RW (2000) Fear and fairness in the city: criminal enforcement and perceptions of fairness in minority communities. South Calif Law Rev 73:1219–1275

Cao L, Frank J, Cullen FT (1996) Race, community context and confidence in the police. Am J Police 15:3–22

Carr PJ, Napolitano L, Keating J (2007) We never call the cops and here is why: a qualitative examination of legal cynicism in three Philadelphia neighborhoods. Criminology 45:445–480

Caulkins JP, Reuter P (1998) What price data tell us about drug markets. J Drug Issues 28:593–612

Clarke RV (1997) Situational crime prevention: successful case studies. Harrow and Heston, Albany

Cohen J, Gorr WL, Singh P (2003) Estimating intervention effects in various settings: do police raids reduce illegal drug dealing at nuisance bars? Criminology 41:257–292

Cooper H, Moore L, Gruskin S, Krieger N (2005) The impact of a police drug crackdown on drug injectors' ability to practice harm reduction: a qualitative study. Soc Sci Med 61:673–684

Cordner GW, Biebel EP (2005) Problem-oriented policing in practice. Criminol Public Policy 4:155–180

Corsaro N, Brunson RK, McGarrell EF (2010) Evaluating a policing strategy intended to disrupt an illicit street-level drug market. Eval Rev 34:513–548

Corsaro N, Brunson RK, McGarrell EF (forthcoming) Problem-oriented policing and open-air drug markets: examining the rockford pulling levers deterrence strategy. Crime Delinq

Eck JE (2006) Science, values, and problem-oriented policing: why problem-oriented policing? In: Weisburd D, Braga A (eds) Police innovation: contrasting perspectives. Cambridge University Press, New York, pp 117–132

Girodo M (1985) Health and legal issues in undercover narcotics investigations: misrepresented evidence. Behav Sci Law 3:299–308

Goldstein H (1990) Problem-oriented policing. McGraw-Hill, New York

Jacobs BA (1997) Contingent ties: undercover drug officers' use of informants. Br J Sociol 48:35–53

Kane RJ (2005) Compromised police legitimacy as a predictor of violent crime in structurally disadvantaged communities. Criminology 43:469–498

Kennedy D (1997) Pulling levers: chronic offenders, high-crime settings, and a theory of prevention. Valparaiso Univ Law Rev 31:449–484

Kennedy D (2009) Deterrence and crime prevention: reconsidering the prospect of sanction. Routledge, New York

Manning PK (1977) Police work: essays on the social organization of policing. MIT Press, Cambridge, MA

Marx GT (1988) Undercover: police surveillance in America. University of California Press, Berkeley

Mazerolle L, Soole DW, Rombouts S (2006) Street-level drug law enforcement: a meta-analytical review. J Exp Criminol 2:409–435

Mazerolle L, Soole D, Rombouts S (2007) Drug law enforcement: a review of the evaluation literature. Police Q 10:115–153

Miller GI (1987) Observations on police undercover work. Criminology 25:27–46

Morris P, Heal K (1981) Crime control and the police, vol 67, Home office research study. Home Office, London

Office of National Drug Control Policy (2011) ONDCP available at http://whitehousedrugpolicy.gov/publications/policy/11budget/exec_summary.pdf. Accessed on 25 July 2011

Potter G, Gaines L, Holbrook B (1990) Blowing smoke: an evaluation of marijuana eradication in Kentucky. Am J Police 97:97–116

Sherman LW (1987) Reinventing probable cause: target selection in proactive investigations. J Soc Issues 43:87–94

Sherman LW (1990) Police crackdowns: initial and residual deterrence. Crime Justice Rev Res 12:1–48

Terrill W, Reisig MD (2003) Neighborhood context and police use of force. J Res Crime Delinq 40:291–321

Weitzer R (1999) Citizen perceptions of police misconduct: race and neighborhood context. Justice Quart 16:819–846

Wellford C, Pepper JV, Petrie C (2005) Firearms and violence: a critical review. The National Academies Press, Washington, DC

Wilson JQ (1968) Varieties of police behavior. Harvard University Press, Cambridge, MA

Wood E, Tyndall M, Spittal P, Li K, Anis A, Hogg R et al (2003) Impact of supply-side policies for control of illicit drugs in the face of the AIDS and overdose epidemics: investigation of a massive heroin seizure. Can Med Assoc J 168:165–169

Drug Marketing

► Drug Trafficking

Drug Markets

► Drug Enforcement

Drug Substitution Programs and Offending

Ambros Uchtenhagen[1] and
Nicole Egli Anthonioz[2]
[1]Research Institute for Public Health and Addiction, University of Zurich Medical School, Zurich, Switzerland
[2]School of Criminal Sciences, University of Lausanne, Lausanne, Switzerland

Synonyms

Drug-related crime; Heroin-assisted treatment; Opioid replacement treatment; Opioid substitution treatment

Overview

The prescription of opioids as a therapy for opiate dependence has a long history, and opioid substitution treatment is one of the best documented and researched therapeutic approaches in medicine. It is considered a cornerstone in the worldwide efforts to help opiate addicts and to relieve the burden for afflicted societies, mainly through a significant reduction of HIV infections in heroin injectors and the population at large and through a significant reduction of acquisitive crime by heroin users.

However, "fighting fire with fire" has always met strong opposition. Many only accept abstinence as an outcome of treatment. The research evidence of a substantial reduction of opiate-related crime through substitution treatment can facilitate widespread political acceptance by demonstrating the benefits of this approach for the societies as well as for addicts and their families.

This entry has two parts: the first describes the historical development of this treatment approach, and the second provides the international research evidence on the reduction in opiate-related crime associated with substitution treatment based on a systematic review of eligible studies comparing substitution treatment with diverse control groups. This entry examines the benefits of methadone, buprenorphine, and heroin substitution maintenance.

A Short History of Opioid Maintenance Treatment

Introduction

Maintaining opiate addicts on opiates has a long history. Opium was one of the most effective medications in ancient medicine, and its widespread use frequently led to chronic use and dependence. Medical and nonmedical rationing practice was introduced to satisfy the needs of those who were unable to discontinue its use. The discoveries of morphine and heroin and of intravenous injections were followed by even more widespread medical use and consequently concerns about chronic dependence. The idea of prescribing injectable opiates as a substitute for street heroin started in USA and was abolished on the basis of prohibitionist legislation, while it continued as regular practice in the UK. A new approach to maintaining opiate addicts on substitution therapy was initiated in the USA in 1963, with the prescription of oral methadone in the framework of a comprehensive treatment program. This approach slowly found increasing acceptance and is nowadays considered a cornerstone in the management of opiate dependence and for the prevention of HIV/AIDS in opiate injectors. The concept of heroin maintenance treatment was reactivated in order to reach out to treatment resistant heroin addicts. Based on the unanimously positive outcomes, heroin maintenance has become routine treatment for otherwise untreatable heroin addicts in Switzerland, the Netherlands, Denmark, and Germany.

The Origins: The Exceptional Power of Opium

Prescribing and using opiates as an effective medication for a range of ailments goes back to the origins of Chinese, Sumeric, and Egyptian medicine, and probably beyond (the first traces are known from Neolithic burial sites). In Homeric times, opium in wine was offered against all evils, and in ancient Egypt, sponges soaked in opium were used during surgery. Authoritative Roman and Arabic texts have praised the uses of opium throughout the Middle Ages, and the opium-made “theriak” was used as a panacea for many centuries, in spite of prohibitive warnings by the Church. Opium tincture (laudanum) was still considered to be the most essential drug at the dawn of modern medicine by Paracelsus. It is inevitable that its addictive potential became manifest, but we do not know when and where opiate addiction was first observed and when the use of opiates as a maintenance regime started. However, opiate dependence must soon have been recognized to be difficult to overcome (Seefelder 1996).

Medical and Nonmedical Maintenance by Rationing Systems

Maintenance was the logical answer to the failure in overcoming dependence. We do know that the Roman emperor Marc Aurel was maintained on opium by the eminent physician Galenus and that the Mugal emperor Jahangir received opium maintenance from his chief physician. And far into the twentieth century, opium dispensing to registered opium addicts was practiced in some Asian countries like Pakistan, providing obviously dependent people with their daily dose. This system was abolished after the adoption of the 1961 UN Single Convention.

A comparable system of an alcohol rationing system (Bratt system) was practiced in Sweden for many years. It became unpopular as a “paternalistic” interference of state and was abolished after the Second World War.

Morphine Maintenance

Opium is rich in alkaloids with different characteristics. They were identified successively and developed further synthetically. Morphine, distilled from opium in 1804, was a revolution in pain management and found widespread use especially after the invention of syringes for injection. Once heroin (diacetylmorphine) was developed in 1874, it was again used as an effective medication for many conditions, frequently prescribed as "patent medicines." In Europe, it was also considered to be a cure for morphine, cocaine, and opium dependence, until its own addictive potential became known. In the USA, 44 narcotic clinics were set up, most of them after the Harrison Act, which left the dependent persons without supply, and introduced tapering off morphine for detoxification purposes. By then the maintenance concept was reinvented, based on the frequency of relapses ("Tennessee system," 1914). While the southern clinics, treating mainly iatrogenic morphine addicts, were quite successful, the New York clinics failed in maintaining young disintegrated heroin addicts (Musto 1999; Mino 1990).

Prohibition Against Maintenance

All narcotic clinics, successful or not, were closed by law in 1923, while a few doctors continued to prescribe heroin to addicts until the Single Convention of 1961, when heroin became a controlled substance and its use restricted to scientific purpose. The concept of maintaining patients on an otherwise illegal or unwelcomed consumption was incompatible with the Puritan idea of prohibition.

The one country resisting the temptation of a strict prohibition in Europe was the United Kingdom where the Rolleston Committee in 1926 recommended heroin prescribing for chronic addicts (which at the time were mainly socially integrated patients who became dependent on morphine after treatment for pain management).

The Revival of the Maintenance Model: "Fighting Fire with Fire"

It was again in the USA where the maintenance treatment of opiate addicts was reintroduced, using methadone, a synthetic opioid, instead of morphine or diamorphine, on the basis of its advantages (oral application, longer half-life). According to the needs of the new target population – mainly socially disintegrated urban heroin injectors – a comprehensive health and social support program went along with methadone prescribing in contrast to just handing out prescriptions for unsupervised use (Dole and Nyswander 1965).

The positive effects of this new approach, especially on the health and criminal behavior of patients, confirmed by an increasing number of evaluation studies, led to a growing acceptance of the maintenance concept, in spite of "abstinence-only" arguments and opposition. But the decisive factor to speed up methadone (and buprenorphine) maintenance was the advent of the HIV epidemic. Drug injecting was recognized as a major factor for transmitting the viral infection (for hepatitis C as well). Throughout Europe, the idea of maintaining heroin addicts on opioids became increasingly acceptable, instead of restricting treatment to detoxification and abstinence-oriented approaches. The number of countries providing methadone maintenance increased from 7 in 1980 to 28 by 2005, and the number of countries providing buprenorphine maintenance rose within a few years to 21 (EMCDDA 2006). In 2008, there were ca. 670,000 patients in substitution treatment (EMCDDA 2010). The former main objective to reduce criminality and to curb the illegal heroin market was replaced gradually by a major public health concern (WHO/UNODC/UNAIDS 2004). Finally, in 2006 the World Health Organization succeeded in putting methadone and buprenorphine on the list of essential medicines, based on evidence about the safety and effectiveness of these substances.

Opioid maintenance treatment is the predominant treatment option for opioid users in Europe. It is generally provided in outpatient settings, though in some countries it is also available in inpatient settings and is increasingly provided in prisons. Opioid maintenance is available in all EU Member States. In most countries, specialized public outpatient services are the main providers of substitution treatment. However, office-based general practitioners, often in shared care arrangements with specialized centers, play an increasing role in the provision of this type of treatment (EMCDDA 2010). The only other psychotropic substance where maintenance is practiced is tobacco, using nicotine replacement patches.

Outside of Europe, we see a different picture. By 2009, opioid maintenance treatment was available in 70 countries, but only an estimated 8 % of drug injectors received it (Mathers et al. 2010). The reasons are manyfold: most of the research evidence comes from developed countries in the West, and injecting addicts are discriminated against as morally or legally deviant; therefore only abstinence is considered to be the legitimate goal of interventions, even at the cost of infringing on human rights and medical ethics. Compulsory nonmedical reeducation camps are still frequent in some Asian countries (WHO 2009).

The Quest for Prescribing the "Original Drug": New Research, New Practice

The increasing number of methadone patients inevitably has led to an increasing but less important number of "methadone-resistant" patients who continued to inject heroin in spite of adequate methadone dosages and care. As the systematic review of evaluations will show, methadone treatment by itself did not provide entirely satisfactory results. At the same time, the HIV epidemic made it a priority to increase coverage, i.e., to reach out to as many injectors as possible. Thus, prescribing heroin as the original and preferred substance of addicts was proposed and has been tested.

Even if the AIDS epidemic is at the origin of all initiatives to prescribe heroin in Switzerland, in the Netherlands, in Germany, in Canada, and in Spain, the idea of heroin maintenance was ready for a revival. In the UK, where the number of heroin addicts had increased and their characteristics changed, drug dependence clinics for maintenance treatment were opened in London in the 1960s, and notification of any addicts receiving a heroin prescription had to be provided to the Home Office. The much debated objectives were medical care and, at the same time, social control of addicts. For many reasons, heroin prescribing was more commonly replaced by methadone prescribing, until the AIDS epidemic led to a reconsideration and to new experimentation with injectable and smokable heroin, now in a perspective of public health interests (Strang and Gossop 1996).

Another debate started in Canada in the 1950s and again in the early 1970s, in the Netherlands and in the USA during the 1970s. The arguments were mainly focusing on heroin addiction as a chronic condition and the need to restrict its negative health and social consequences. Nonmedical options for providing heroin to addicts in the Netherlands (tolerated "home dealers" and "heroin bars") were started but then considered to be failures (Mino 1990).

Feasibility studies on heroin prescribing were initiated in Australia. Large-scale experimental studies were finally set up in Switzerland (1994–1996), the Netherlands (1995–1997), Germany (2002–2005), Spain (2003–2004), Canada, and the UK. The Swiss cohort study provided the first positive outcomes, acknowledged by an international expert committee set up by WHO (Ali et al. 1998). As they could not separate the effects of heroin prescribing from the effects of concomitant care, the other countries set up randomized controlled trials, thereby adding essential new findings to the already established positive outcomes (see below).

A common characteristic of the new experimental studies was the aim to cover heroin addicts unable to profit from other treatments including methadone maintenance, and the provision of pharmaceutical heroin in the framework of a comprehensive assessment and treatment program. In order to avoid overdose and diversion, the intake of injectables must be made under visual supervision of staff in the clinics. In Switzerland, the Netherlands, and Germany, heroin maintenance has reached the status of a routine treatment and is paid by health insurance. Belgium decided to start another trial, and Denmark has set up heroin maintenance without further research.

Benefits and Risks of Opioid Maintenance: A Summary Statement

A summary of findings confirms the feasibility of maintenance treatment in terms of patient satisfaction and widespread acceptance, the safety for patients and staff, the significant improvements in somatic and mental health, and a reduction in risk behavior and illicit drug use, in drug-related crime and in public nuisance offenses (Uchtenhagen 2003, WHO 2009). Maintenance therapies are increasingly integrated into treatment and care systems without negative effects for other approaches. Main risks are overdose mortality in the introductory phase of methadone maintenance and the diversion of prescribed opioids to the illegal market; they can be avoided by appropriate regimes. Economic studies have evidenced superior benefits compared to costs.

How the Treatment's Effect on Crime Was Evaluated Across the World

Methods

Under the umbrella of the Campbell Collaboration Crime and Justice Group, a systematic review was carried out (Egli et al. 2009) in order to investigate the effect on crime of substitution programs, be it substitution by methadone or heroin, as opposed to any other type of treatment (in particular, abstinence, detoxification, psychotherapy) or no treatment at all, as well as comparative effects between different substitution therapies. All primary references used in the systematic review can be found in the Campbell Collaboration publication (Egli et al. 2009); they are identified in the current text by the first authors' name and the year, without a space between.

A search strategy for the identification of studies was defined and followed. Relevant studies were identified through abstracts, bibliographies, and databases such as Medline, the National Criminal Justice Reference Service (NCJRS), Harm Reduction Journal, Journal of Substance Abuse Treatment, National Treatment Agency for Substance Misuse (NHS), and National Treatment Outcome Research Study (NTORS), as well as the bibliographies of relevant reviews. In order to be eligible, a study had to fulfill the following criteria:

– At least one of the study groups had to undergo a substitution program (using, e.g., methadone and/or opiates as substitution drugs).
– A measure of re-offending had to be given, since only such effects were included in the meta-analysis, as opposed to medical or social outcomes.
– Level 4 or higher on the Sherman scale had to be met.

The methods used, be it for review or for the analysis of the data, are as specified in Practical Meta-Analysis by Lipsey and Wilson (2001); odds ratios have been used for measuring the effects observed. An odds ratio greater than 1 stands for a reduction in the outcome measure (e.g., a positive effect with respect to the response variable). For the computation of mean effect sizes, the inverse variance method of meta-analysis (Lipsey and Wilson 2001) was used.

Illustrations (forest plots) have been created using R (www.r-project.org). Also, in these illustrations, the outcomes have been coded as follows: the criminal behavior (1: any offense; 2: Property crime; 3: theft) followed by the measure used (a: commission; b: arrest; c: conviction; d: incarceration; e: illegal income; f: cost of crime (estimated); g: ASI legal; h: charged) and finally the way it has been measured, in parentheses (1: self-report; 2: official sources; 3: known status; 9: unknown).

Results

Methadone Maintenance Treatment (MMT)

Seven RCTs (randomized controlled trials) and three high-quality quasi-experimental studies (i.e., level 4 on the Sherman Scale) were found for MMT. The control groups differed widely: three were wait-list control designs (Dole 1969; Schwartz 2007; Yancovitz 1991), one used a placebo control (Newman 1979), in one the control group received counselling (Kinlock 2008), in three others the control condition received detoxification (Daley 2000; Haglund 1978; Sees 2000), in one the control received treatment in the community (Bale 1980), and one received residential treatment (Teesson 2006). The results for these studies are shown in Fig. 1. Not surprisingly, the distribution is

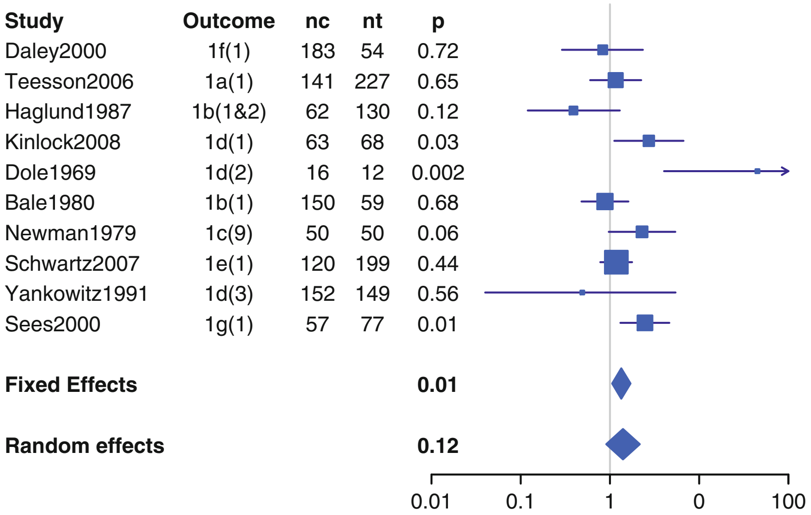

Study	Outcome	nc	nt	p
Daley2000	1f(1)	183	54	0.72
Teesson2006	1a(1)	141	227	0.65
Haglund1987	1b(1&2)	62	130	0.12
Kinlock2008	1d(1)	63	68	0.03
Dole1969	1d(2)	16	12	0.002
Bale1980	1b(1)	150	59	0.68
Newman1979	1c(9)	50	50	0.06
Schwartz2007	1e(1)	120	199	0.44
Yankowitz1991	1d(3)	152	149	0.56
Sees2000	1g(1)	57	77	0.01
Fixed Effects				**0.01**
Random effects				**0.12**

Drug Substitution Programs and Offending, Fig. 1 Descriptors and forest plot for studies fulfilling criteria 4 or 5 of the Sherman Scale where treatment group receives MMT and control groups receive no substitution treatment. The size of the boxes in the forest plot is proportional to the weight of the study in the summary measures. The confidence interval of Dole (1969) is shortened for representation (*arrow*). *nt* is the number of subjects in the treatment group, while *nc* is the number of subjects in the control group. Outcomes have been coded

heterogeneous ($p < 0.05$), indicating meaningful variation in effect across these studies. The random effects model does not indicate a significant effect of methadone maintenance with respect to these control groups; the mean effect measure is, however, in favor of MMT. Overall effects for the four different types of control groups have also been computed; none of these show a significantly greater effect of the maintenance treatment over the treatment in the control groups. All of these effects grouped by control treatment are, however, in favor of MMT.

Effects of Heroin Substitution Treatment

Six studies of heroin substitution programs were found, five of which were RCTs. In four of the RCTs, the control group underwent methadone maintenance (Dijkgraaf 2005; Hartnoll 1980; Löbmann 2008; March 2006), while in one (Perneger 1998), the control group underwent a conventional treatment. The results are shown in Fig. 2.

A very large effect was obtained for the study by Perneger (1998), along with a very large confidence interval. Both the small sample size and the variety of control treatments explain these two observations. Overall, for these RCTs, homogeneity is rejected ($p = 0.004$). If Perneger (1998) is not included in the analysis, due to the different treatment of the control group with respect to all other included studies, homogeneity is accepted ($p = 0.21$). The fixed effects mean effect size is then 1.55 [1.18; 2.02] ($p = 0.0015$). Here, a significant decrease in the criminality measures is, therefore, present for heroin over methadone maintenance treatment. However, in all of these trials except Dijkgraaf (2005), a selection effect is present that is more or less pronounced. In order to be admitted into these studies, subjects had to have followed (and failed) a program before (this selection effect is strongest for the study March (2006), where subjects were required to have had MMT at least twice before admission). Therefore, these results are overall more properly interpreted as showing a positive effect of HMT over MMT in subjects having already failed at a treatment, including MMT.

Effects of Buprenorphine Substitution Treatment

Three randomized controlled trials concerning the effect of buprenorphine maintenance on

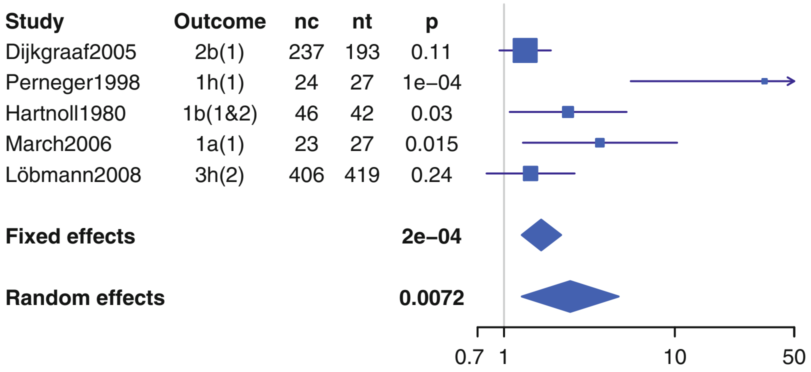

Study	Outcome	nc	nt	p
Dijkgraaf2005	2b(1)	237	193	0.11
Perneger1998	1h(1)	24	27	1e–04
Hartnoll1980	1b(1&2)	46	42	0.03
March2006	1a(1)	23	27	0.015
Löbmann2008	3h(2)	406	419	0.24
Fixed effects				**2e–04**
Random effects				**0.0072**

Drug Substitution Programs and Offending, Fig. 2 Descriptors and forest plot for studies where treatment groups receive heroin maintenance and the control group either MMT or another standard treatment. The size of the boxes is proportional to the weight of the study in the summary measures. The confidence interval for the study by Perneger (2008) has been cut for representation (*arrow*). The outcomes have been coded. *nt* and *nc* are defined as above

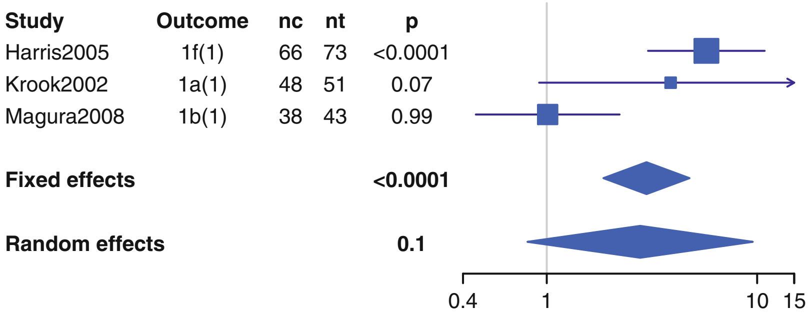

Study	Outcome	nc	nt	p
Harris2005	1f(1)	66	73	<0.0001
Krook2002	1a(1)	48	51	0.07
Magura2008	1b(1)	38	43	0.99
Fixed effects				**<0.0001**
Random effects				**0.1**

Drug Substitution Programs and Offending, Fig. 3 Descriptors and forest plot for studies where treatment group receives buprenorphine maintenance. The confidence interval for Krook (2002) has been cut on the right (*arrow*) for representation purposes. The outcomes have been coded and *nt* and *nc* are defined as above

criminal activities were found. In two of these studies, the control group was MMT (Harris 2005; Magura 2008), while in the third (Krook 2002), the control group received a placebo. The individual and overall effects are shown in Fig. 3. The homogeneity analysis rejects homogeneity ($p < 0.05$). The overall effect is positive, but not significant ($p = 0.10$). Only one of the three studies has a significant odds ratio, the study by Harris (2005), which shows the largest effect size, and is also the largest of the three studies.

When the study with a differing control group (Krook 2002) is excluded, homogeneity is still refuted (Q test $p = 0.0008$). Overall, there is no significant reduction in criminality when buprenorphine instead of methadone is used, although the findings suggest a slight advantage for buprenorphine relative to methadone (or placebo).

Discussion

Results show that MMT has a greater effect on criminal behavior than non-maintenance-based treatments, but not significantly so. Furthermore, the comparison between this maintenance treatment and heroin maintenance showed a significant advantage for heroin maintenance; this is true particularly for persistent users who have participated in a program before and failed. Buprenorphine maintenance also shows an advantage over alternative treatments, but one of these alternative treatments was placebo; over MMT, one study shows a significant advantage for buprenorphine, while a second study shows almost equivalent effects.

Two systematic reviews of substitution programs have been published by the Cochrane Collaboration: Ferri et al. (2006) and Mattick et al. (2009). While these reviews do not focus on crime as an outcome measure, a comparison of results with the findings from the Egli et al. (2009) review summarized above is relevant. In Mattick et al. (2009), three studies comparing methadone maintenance to no opioid replacement therapy with respect to their effect on criminal behavior are included. The results obtained are similar to the results obtained here in two respects: firstly, the effect of MMT seems to reduce criminal behavior more than the alternatives that do not include maintenance, and secondly, this effect is not significant.

In Ferri et al. (2006), four trials comparing methadone maintenance to heroin maintenance are included. One study showed a reduction in the risk of being charged when on heroin maintenance; this is in line with the results obtained here. Also, two studies considered criminal offending and social functioning in a multi-domain outcome measure, and again, heroin plus MMT yields better results than MMT alone. Again, this is in line with the results obtained here, in that heroin maintenance reduced criminality more than MMT.

Related Entries

- ▶ Drug Abuse and Alcohol Dependence Among Inmates
- ▶ Drug Courts
- ▶ Drug Courts' Effects on Criminal Offending
- ▶ History of Substance Abuse Treatment
- ▶ Randomized Experiments in Criminology and Criminal Justice

Recommended Reading and References

Ali R, Auriacombe M, Casas M, Cottler L, Fasrrell M, Kleiber D (1998) Report of the external panel on the evaluation of the Swiss scientific studies of medically prescribed narcotics to heroin addicts. World Health Organisation, Geneva

Dole VP, Nyswander M (1965) A medical treatment for diacetylmorphine (heroin) addiction. A clinical trial with methadone hydrochloride. J Am Med Assoc 193:80–84

Egli N, Pina M, Skovbo Christensen P, Aebi MF, Killias M (2009) Effects of drug substitution programs on offending among drug-addicts. Campbell Syst Rev 2009:3

EMCDDA (2006) European Monitoring Centre on Drugs and Drug Addiction EMCDDA: annual report 2006. Office for Official Publications of the European Union, Luxemburg

EMCDDA (2010) 2010 Annual Report.http://www.emcdda.europa.eu/online/annual-report/2010

Ferri M, Davoli M, Perucci CA (2006) Heroin maintenance treatment for chronic heroin-dependent individuals: a cochrane systematic review of effectiveness. J Subst Abuse Treat 30(1):63–72

Lipsey MW, Wilson DB (2001) Practical meta-analysis. Sage, Thousand Oaks

Mathers BM, Degenhardt L, Ali H, Wiessing L, Hickmann M, Mattick RP et al (2010) HIV prevention, treatment and care services for people who inject drugs: a systematic review of global regional and national coverage. Lancet 375: 1014–1028

Mattick R, Breen C, Kimber J, Davoli M (2009) Methadone maintenance therapy versus no opioid replacement therapy for opioid dependence (Cochrane Review). Cochrane Database Syst Rev 3

Mino A (1990) Analyse scientifique de la litérature sur la remise contrôlée d'héroïne ou de morphine (Scientific analysis of studies on the controlled prescription of heroin or morphine). Federal Office of Public Health, Bern

Musto DF (1999) The American disease. Origins of narcotic control, 3rd edn. Oxford University Press, New York

Seefelder M (1996) Opium. Eine Kulturgeschichte (Opium. A cultural history), 3rd edn. Ecomed, Landsberg

Strang J, Gossop M (1996) Heroin prescribing in the British system: historical review. Eur Addict Res 2:185–193

Uchtenhagen A (2003) Substitution management in opioid dependence. In: Fleischhacker WW, Brooks DJ (eds) Addiction. Mechanisms, phenomenology and treatment. J Neural Transmission suppl 66:33–60

Uchtenhagen A (2008) Heroin-assisted treatment in Europe: a safe and effective approach. In: Stevens A (ed) Crossing frontiers. International developments in the treatment of drug dependence. Pavilion, Brighton, pp 53–82

WHO (2009a) Assessment of compulsory treatment of people who use drugs in Cambodia, China, Malaysia and Viet Nam. An application of selected human rights principles. World Health Organisation, Geneva

WHO (2009b) Guidelines for the psychosocially assisted pharmacological treatment of opioid dependence. World Health Organisation, Geneva

WHO/UNODC/UNAIDS (2004) Position paper: substitution maintenance therapy in the management of opioid dependence and HIV/Aids prevention. World Health Organisation, United Nations Office on Crime and Drugs, Joint United Nations Programme on HIV/Aids

Drug Trade

► Drug Trafficking

Drug Trafficking

Henry H. Brownstein
NORC at The University of Chicago, Bethesda, MD, USA

Synonyms

Drug distribution; Drug marketing; Drug trade

Overview

In an organic sense, drugs are chemicals. When people ingest drugs, they pass through their body to their brain where they interfere with the neurotransmitters that transfer signals across synapses changing the messages that the brain sends to the body and thereby affecting the way the body works. In strictly economic terms, however, drugs are a commodity. That is, they are a product that is exchanged between people. Some drugs in some nations are socially and legally considered to have value for one reason or another and are deemed acceptable for use by some or all people under some or all circumstances. These are produced, distributed, and consumed under authority of law and sanctioned by less formal social norms for commercial purposes and personal use by approved individuals in approved circumstances. Others drugs are not acceptable for any purpose or any person under any circumstance. Today and for some time now, the production, distribution, and consumption of those drugs that are not authorized by law or sanctioned by social norms have been considered a serious problem for individuals, communities, and nations. There are very real public health concerns about drugs, all drugs but in particular illicit drugs or licit drugs used in illicit ways, and what they do to individuals and their minds and bodies. And there are very real public safety concerns about the impact of drugs on people and their communities and nations, especially with regard to the consequences of the commercial transactions involving those drugs that are not recognized by law. So it is not surprising that for the last century or so, there has been considerable attention among social scientists and public policymakers to questions and concerns about the trafficking and marketing of illicit drugs and their relationship to crime and violence.

Understanding Drug Trafficking and Illicit Drug Markets

As a commodity drugs need to be produced and distributed in order to reach consumers. Among those drugs that are illegal, there are differences in how and where they are produced, and therefore, how and at what scale they are distributed. For example, heroin may be sold to consumers in nations such as the United States as powder but it starts as an agricultural product cultivated in other parts of the world and requires a manufacturing process to convert the plant product into the form preferred by consumers and a distribution process that crosses national boundaries. Methamphetamine, on the other hand, can be produced by cooking the required but easily available chemicals at a fixed location with enough equipment to be called a laboratory, or simply by shaking the chemicals together in a 2-L bottle while sitting in the back of a moving car. Cocaine starts as a plant grown in places like South America, but by the time it reaches consumers in the United States, it is in the form of a powder manufactured in a laboratory setting or as little rocks known as crack that are made almost anywhere from a small amount of the powder for small quantity sales. So a broad commercial perspective is helpful for a general understanding of the trafficking and trade in illicit drugs.

Industrial economists distinguish *industry* from *markets*. Whereas industry refers to groupings of individual *businesses* that share common techniques and processes for the production of a certain commodity, markets refer to groupings of the *consumers* of the product of that industry (Andrews 1949). When the product is illicit drugs, whatever specific drug it is, the markets of consumers for those drugs are inextricably linked to the production and distribution techniques and processes and the organization and operation of the illicit drug industry. For example, people who purchase heroin from a local dealer in their own community are nonetheless connected to the individuals and organizations that are involved in the international, national, and regional agricultural, manufacturing, and trafficking operations that are necessary to produce heroin and bring it to market.

The illicit drug industry and its related markets have been studied from a number of perspectives. Looking at drugs as a commercial enterprise, economists have studied the illicit drug industry for one or more drugs and their various markets in terms of things like price and purity (e.g., Caulkins 2005; Caulkins and Reuter 1998; Rhodes et al. 2000). Public health researchers have focused on the impact of drug trafficking and markets on communities in terms of issues related to things like drug-related morbidity and mortality and drug treatment and prevention (e.g., Curtis et al. 1995; Sommers et al. 2006). Criminologists have studied the illicit drug industry and markets in terms of issues related to public safety, criminal activity, and violence (e.g., Goldstein et al. 1992; Weisburd and Mazerolle 2000). Overall the questions raised by this body of research ask how is the industry and how are the markets organized and how do they operate? How large are they in terms of production and in terms of economic value? Specifically in terms of trafficking, given that the product is illegal, how does it get from producers to consumers? And what is the impact of the trafficking and marketing of illicit drugs on people, communities, and nations?

Drug Trafficking

From the founding of the United States as a nation through to the late 1800s, it was common for home remedies to be freely available to citizens for dealing with personal problems of pain and illness. David Musto, a historian of drugs in America, called the late nineteenth century a period of “wide availability and unrestrained advertising” (1991, p. 42). During this period, the production and distribution of home remedies was not regulated by government, and many of these remedies contained palliative substances such as opium or stimulating substances such as cocaine (Inciardi 2007). Then in the early twentieth century, concerns for health and safety problems believed to be associated with their use resulted in the passage in 1906 by the United States Congress of the Pure Food and Drug Act, which imposed quality and packaging standards on all food and drug products (Inciardi 2007). Following that in 1914, the Harrison Narcotics Act was passed and in 1937 the Marijuana Tax Act, and together they gave the federal government some measure of regulatory control over the production and distribution of drugs (Musto 1999). Unfortunately while these government actions through taxation regulated the production and distribution of drugs that were earlier found in popular home remedies, they did not address the consumer demand for those drugs. So the illicit drug industry and markets filled the void to meet the demand.

As the illicit drug industry grew, the government of the United States increasingly became concerned about the impact of these drugs on public health and safety. In the middle of the twentieth century, there was not sufficient scientific evidence to know what that impact really was, but there were strong beliefs and opinions and those were used to support the direction of public policy toward illicit drug use and users and trafficking and markets. In 1971, President Richard Nixon formally declared war on drugs (Inciardi 2007). With his declaration, attention shifted from drug users to drug trafficking, and in 1973, Nixon created the Drug Enforcement Administration (DEA) as a component of the United States Department of Justice specifically charged with fighting the war on drug trafficking. Later President Ronald Reagan reaffirmed the war on drug trafficking, and then President George H.W. Bush with Congress passed the Anti-Drug Abuse Act of 1988 thereby forming the Office of National Drug Control Policy (ONDCP) as a policy agency to lead the national drug strategy. With ONDCP, the United States might finally have a Drug Czar, but the focus on drug trafficking as the problem and drugs as the enemy started with Nixon's declaration of war and his formation of the DEA to fight that war.

While the United States was intensifying its war effort against drugs and drug traffickers, not all nations followed suit. Many others favored a policy of harm reduction emphasizing efforts to manage the harm to drug consumers and their communities from drug use rather than trying to control the supply of drugs through trafficking (Inciardi and Harrison 2000). For example, in the Netherlands, drug policy historically has favored protecting the health of users and reducing their health risks by focusing on programs of prevention and treatment (van Laar et al. 2011). Given a tradition known as "gedoogbeleid" that favors discretion when dealing with drugs and drug users, since the late twentieth century, the Dutch have formally and systematically not enforced laws involving small quantities of cannabis and have even established guidelines allowing certain retail establishments to sell cannabis to consumers without fear of prosecution as long as they adhere to established guidelines, such as not advertising and not selling to minors (MacCoun and Reuter 1997). Under international pressure in recent years, the Dutch government has placed greater limits on how these establishments can operate and who they can serve, but the tradition and policy allow the trade to continue. This does not mean that the government of the Netherlands favors unregulated trafficking or trade in illicit drugs. A recent report issued by the Netherlands National Drug Monitor to the European Monitoring Centre for Drugs and Drug Addiction (van Laar et al. 2011) notes that along with addressing the need to prevent drug use, the needs of drug users for treatment and rehabilitation, the reduction of harms faced by drug users, and lessening any disturbances to public safety they might cause in their communities, combatting drug production and trafficking is one of its main objectives. But the report continues, the "primary aim of Dutch drug policy is focused on health protection and health risk reduction" (van Laar et al. 2011, p. 15). Similarly, in the United Kingdom, a harm reduction approach to drugs has been in place going back to a report issued in 1926, *The Report of the Departmental Committee on Morphine and Heroin* (Bennett 1988).

While sovereign nations appropriately and naturally each have their own policy toward drugs, drug users, and drug trafficking, there has been international cooperation going back to 1909 when 13 nations gathered in Shanghai for the International Opium Commission (Musto 1991). At that meeting, which focused on opium and opiates, no binding decisions were made and no treaties signed. But other meetings followed. In 1911 there was a meeting of 12 nations in Hague resulting in an agreement for each nation to enact legislation to control narcotics trade and to fund educational programs. Then starting in 1961, there was a series of three international Conventions held under the auspices of the United Nations. Through the first two, all previous multinational treaties that had been negotiated from 1912 to 1953 were consolidated, and controls over 100 different substances considered to be narcotic drugs were tightened, and through the third in 1988, a focus was set on international drug trafficking (United Nations 1988).

Despite national and international efforts to control or manage it, today the illicit drug industry is well established on a global scale with active local markets and large and profitable national and international corporate-type manufacturers and distributors. Systems and procedures are in place to move drugs that are produced in one part of the world to other parts of the world where consumers are waiting to purchase and use them. With the focus on drug trafficking, the question then is what is the nature and scope of international trafficking? What is the impact of the traffic in illicit drugs on both producer and consumer nations and on their citizens?

Since the industry and the markets operate outside of the law, there are no official records of how much is produced or how much is sold or purchased. There are no official records of costs related to production or distribution of any illicit drug product, and no official records of how many consumers there are or what consumers have paid or are willing to pay for their drugs. So statisticians, economists, and other social scientists use available data sources to try to calculate the scope and scale of the illicit drug industry as a whole, and for particular drug industries and markets, such as heroin or cocaine. They use these data to derive estimates of drug consumption and estimates of the supply of drugs. For consumption estimates, they sometimes use data from surveys of samples of people who tell an interviewer about their experience using drugs, or not. For example, in the United States, estimates are derived using data from the National Survey of Drug Use and Health (NSDUH), an annual nationwide household survey produced for the Federal government by the Substance Abuse and Mental Health Services Administration (SAMHSA) and administered to a random sample of 70,000 respondents age 12 and older asking them about their use of and experience with various illicit drugs. Estimates of consumer demand in the United States also come from the Treatment Episode Data Set (TEDS) similarly maintained by SAMHSA, though in this case, counting the drug treatment and using experiences of about 1.5 million people annually admitted to drug treatment.

In addition to survey data, estimates of drug consumption and supply are also derived from available official record data from Federal, state, and local government agencies that have operational involvement in some way or other with acknowledged users of various illicit drugs. These agencies include, for example, law enforcement agencies that arrest drug users and dealers or traffickers, and treatment service providers that work with drug users to try to help them stop using drugs. Law enforcement records, for example, have been used to derive estimates of the supply of drugs based on a variety of law enforcement activities including crimes known to the police and arrests, but also things like the amount of particular drugs seized by various local, regional, and national law enforcement agencies. In the United States, the agency record data for supply estimates comes from national aggregate counts of crime and agency counts of arrests from sources like the Uniform Crime Reports (UCR), an annual national count of crimes known to the police and arrests made by the police. More directly focused on the actual supply of drugs known to be in the country, another estimate comes from data on drugs seized and analyzed in laboratories by law enforcement for the DEA System to Retrieve Information from Drug Evidence (STRIDE) program (NDIC 2011), or the tactical intelligence collected by the El Paso Intelligence Center (EPIC), established by the DEA in 1974 for the collection and dissemination of information related to drug trafficking particularly along the United States border with Mexico.

In summary, what is known about the scope and scale of drug trafficking or even drug using is limited. No official records of activity in the illicit drug industry or drug markets are maintained. All estimates are indirect, derived from proxy measures. They are based on an accounting of the activity of individuals and agencies that in some way work with people who are involved in or with the drug industry. So there is some uncertainty about what is known or believed to be known about the illicit drug industry and drug trafficking.

Illicit Retail Drug Markets

Similarly there are no official records to support what is known or believed to be known about local drug markets and the retail trade in illicit drugs. However, there has been considerable research conducted over the last few decades on the organization and operation of illicit retail markets at the local level (National Institute of Justice 2003; National Research Council 2010, 2001). There are several ethnographic studies of markets in particular cities, and there are a number of studies that provide sociological, geographical, and economic analyses of market organization and operation. Important findings of these studies include the identification of differentiated roles among buyers and sellers of illicit drugs, the characteristics of social relationships and of structural forms in different local markets, and the variation among patterns of distribution and consumption across places like neighborhoods and by the different types of drugs being transacted.

An illicit retail drug market can be defined as the set of people, facilities, and procedures through which illicit drugs are transferred from sellers or dealers to buyers or users (National Research Council 2001, p. 160). As such they are economic enterprises and therefore operate in response to the forces of supply and demand. But unlike legitimate commercial enterprises, illicit drug markets participants are regularly faced with the an odd mix of rapid turnover and overlapping roles among sellers and buyers, a broad range of product price and quality in a limited geographic area, and an absence of any legitimate authority to settle disputes over things like market share or product quality. Consequently, the natural economic forces of supply and demand in such markets are tempered by the need of buyers and sellers to worry about things like the trustworthiness of the people they are dealing with and the fairness and security of those dealings. By definition in an illicit retail drug market, all transactions are criminal transactions. As a result every individual who participates as a buyer or seller in such a market risks a hostile encounter with law enforcement officers and criminal justice agencies. In addition they risk intimidation, coercion, or victimization by other buyers and sellers over market-related disputes or for just being there.

Drug Markets, Crime, and Violence

Estimates from official statistical data from surveys on drug using and the supply of illicit drugs known to law enforcement along with the evidence of a large body of social scientific research strongly suggest that there is a relationship between illicit drug trafficking and drug markets and crime and violence. For the most part, the research on the connection between drugs and crime and violence has focused on drug use, but there are also studies that directly link crime and violence with drug trafficking and drug markets. In an early study of homicide rates in the United States during the twentieth century, for example, criminologist Margaret Zahn found that the rate of homicide varied over time in relation to the establishment of markets for illicit products with rates of homicide being at their highest in the 1920s and 1930s, during the control of alcohol under Prohibition, and in the 1960s through 1970s, during periods of disruption in the heroin and cocaine markets (Zahn 1980). More recently, studies have found higher rates of crime and violence during periods of disruption in organization and operation of particular illicit retail drug markets. For example, in the late 1980s when crack cocaine markets were newly emerging in urban areas, researchers found that most of the crime and violence associated with those markets involved things like disputes between dealers over territory or disputes between sellers and buyers over the quality of the product sold (Brownstein 1996).

The focus of drug policy in the United States on drug trafficking has been linked over the years to a stated relationship between international drug trafficking and crime and violence. The National Drug Intelligence Center (NDIC) of the United States Department of Justice in its annual threat assessment has regularly reported on violence and crime related to transnational crime organizations in producer nations, notably in Latin America and Asia, and its impact in and on the United States (NDIC 2011). In recent years, concern has focused on the Transnational Criminal Organizations in Mexico with the NDIC reporting that seven such organizations in Mexico "control much of the production, transportation, and wholesale distribution of illicit drugs destined for and in the United States" and that among them "a dynamic struggle for control of the lucrative smuggling corridors leading into the United States [has resulted] in unprecedented levels of violence in Mexico" (NDIC 2011, p. 7).

Similarly crime and violence have been associated with domestic illicit retail drug markets. In the 1970s in Chicago, Patrick Hughes conducted a groundbreaking epidemiological study of heroin addicts and concluded, "... heroin dealers must have a reputation for violence, otherwise addicts and other deviants would simply take their drugs and their money" (1977, p. 31). Not only did he observe that heroin dealers would use violence in the conduct of their business but also that other participants in the market were at risk of being the perpetrators or victims of crime and violence themselves through confrontations with others including drug users, drunks, and gang members (Hughes 1977, p. 31). A body of research conducted since that time provides additional though inconclusive support for the notion that while drug markets more often than not are peaceful, there is an observable relationship between the organization and operation of illicit retail drug markets and criminal violence (see Reuter 2009).

In 1985, sociologist Paul Goldstein conceptualized the ways that drugs and violence might be related (Goldstein 1985). He suggested that violence might be a consequence of drug ingestion, a response by an addicted drug user needing drugs and having to use force to get them, or the product of the normal organization and operation of the illicit drug trade. The latter he called systemic and suggested it could include things like disputes between drug dealers over claims to territory or among dealers and buyers over the quality of the product being sold. Just about the time that Goldstein was writing and talking about his tripartite framework, the level of violence and crime began to rise dramatically in United States cities concurrent with the introduction of crack cocaine. Public policymakers, criminal justice and law enforcement practitioners, and the news media around the country became alarmed and responded in the belief that the increasing crime and violence could be linked to the use of crack cocaine in city neighborhoods (Brownstein 1996). But studies using the tripartite framework provided preliminary evidence that in fact the growing crime and violence had more to do with market dynamics than with user behavior (Goldstein et al. 1992).

Conclusions and Future Research

In summary, there are limits to what is known and can be known about the illicit drug industry, drug trafficking, and drug markets. To some extent those limits result from the lack of official records available to be used for scientific or policy analyses. This problem limitation is not likely to be overcome in that the industry exists outside of the law. In addition, the research conducted in this area to date largely has followed the personal agenda of individual researchers and research teams rather than being guided by an organized and systematic program specifically designed to guide researchers and their studies in the direction of questions that need to be answered to adequately and effectively inform policy and practice. The unfortunate consequence has been that much of the policy that is made and many of the practices and programs that are initiated are guided by unfounded assumptions or political principles rather than by scientific evidence.

Nonetheless, there are some things that researchers have found that are convincing. The drug industry, with its consumer markets and the trafficking operations that comprise it, is highly organized and does operate not only on the local level but also on a global scale. Depending on the drug (and among illicit drugs and drugs used in illicit ways, there are many distinctions in what they do and how they are obtained and used), in some way or ways, there is a relationship between local retail market outlets for the product, large- and small-scale producers or manufacturers of the product, and the organized enterprises and well-established procedures and practices put into place to move the product from producers to consumers. And as all this happens outside of the law and legitimate authority, things do not always go well and sometimes the outcome is harmful and a threat to personal and public health and safety.

While there may be one or more studies that have addressed a particular question, it does not necessarily mean that the question has been adequately answered. So questions important for understanding, explaining, and addressing the institution of drug trafficking and any social and personal problems related to it remain open. As noted earlier, among the questions probably worth asking but certainly not all of the questions that need to be asked are the following: How is the industry and how are the markets organized, and how do they operate? How does drug trafficking work? How do different drugs go from producers to consumers? How much of all and different drugs are being produced, and how great is the demand for them? What is the value of illicit drug production worldwide, and what is the value of all drugs being consumed? What is the impact of the illicit drug industry on people, communities, and nations? Are there any positive outcomes? What are the harms? Besides scientific evidence, what if any ethical or moral principles need to be considered when making public decisions about policies and practices related to drugs and drug trafficking? What if any practical limitations are there that may restrict what can or cannot be done about drugs and drug trafficking?

What needs to be done now is not just a continuation of the current practice of individual researchers raising individual questions based on personal interest or public attention and designing studies. What needs to be done is the work by teams of researchers, policymakers, and practitioners to design and develop a broader research agenda, probably international even more than national, to identify and study relevant questions about the drug industry and markets and drug trafficking to then allow policy to be made and practices implemented that will lead to outcomes that benefit and do not harm people, communities, and nations.

Related Entries

▶ Drug Abuse and Alcohol Dependence Among Inmates
▶ Drug Enforcement
▶ History of Organized Crime in Mexico
▶ History of Organized Crime in the United States
▶ History of Substance Abuse Treatment
▶ Italian Mafia
▶ Sex Offenders and Criminal Policy
▶ Trafficking in Human Beings

Recommended Reading and References

Andrews PWS (1949) Manufacturing business. Macmillan, London

Bennett T (1988) The British experience with heroin regulation. Law Contemp Probl 51:299–314

Brownstein HH (1996) The rise and fall of a violent crime wave – crack cocaine and the social construction of a crime problem. Harrow & Heston, Guilderland

Caulkins JP (2005) Price and purity analysis for illicit drug: data and conceptual issues. Heinz Research. Paper 25. http://repository.cmu.edu/heinzworks/25. Accessed 25 Oct 2011

Caulkins JP, Reuter P (1998) What price data tell us about drug markets. J Drug Issues 28:593–612

Curtis R, Friedman SR, Neaigus A, Jose B, Goldstein M, Ildefonso G (1995) Street-level drug markets: network structure and HIV risk. Soc Netw 17: 229–249

Goldstein PJ (1985) The drugs/violence nexus: a tripartite conceptual framework. J Drug Issues 14: 493–506

Goldstein PJ, Brownstein HH, Ryan PJ (1992) Drug-related homicide in New York: 1984 and 1988. Crime Delinq 38:459–476

Hughes PH (1977) Behind the wall of respect: community experiments in heroin addiction control. University of Chicago Press, Chicago

Inciardi JA (2007) War on drugs IV: the continuing saga of the mysteries and miseries of intoxication, addiction, crime and public policy, 4th edn. Allyn & Bacon, Boston

Inciardi JA, Harrison LD (eds) (2000) Harm reduction – national and international perspectives. Sage, Thousand Oaks

MacCoun R, Reuter P (1997) Interpreting Dutch cannabis policy: reasoning by analogy in the legalization debate. Science 278:47–52

Musto D (1991) Opium, cocaine, and marijuana in American history. Sci Am 265:40–47

Musto D (1999) The American disease, 3rd edn. Oxford University Press, New York
National Drug Intelligence Center (2011) National drug threat assessment 2011. No. 2011-Q0317-001. USDOJ, Washington, DC
National Institute of Justice (2003) Toward a drugs and crime research agenda for the 21st century. Special report. USDOJ, Washington, DC
National Research Council (2001) Informing America's policy on illegal drugs: what we don't know keeps hurting us. National Academy Press, Washington, DC, Commission on Behavioral and Social Sciences and Education
National Research Council (2010) Understanding the demand for illegal drugs. National Academy Press, Washington, DC, Committee on Law and Justice
Reuter P (2009) Systemic violence in drug markets. Crime Law Soc Chang 52:275–284
Reuter P, MacCoun R, Murphy P (1990) Money from crime: a study of the economics of drug dealing in Washington, DC. Drug Policy Research Center, RAND, Santa Monica
Rhodes W, Layne M, Johnston P, Hozik L (2000) What America's users spend on illegal drugs 1988–1998. Office of National Drug Control Policy, Washington, DC
Sommers I, Baskin D, Baskin-Sommers A (2006) Methamphetamine use among young adults: health and social consequences. Addict Behav 31: 1469–1476
United Nations (UN) (1988) United Nations conference for the adoption of a convention against illicit traffic in narcotic drugs and psychotropic substances. http://www.unodc.org/pdf/convention_(1988_en.pdf. Accessed 5 June 2012
van Laar M, Cruts G, van Gageldonk A, van Ooyen-Houben M, Croes E, Meijer R, Ketelaars T (2011) Report to the EMCDDA by the Reitox National Focal Point. Final version as approved by the Scientific Committee of the Netherlands National Drug Monitor (NDM) on the 22nd of December 2010. European Monitoring Centre for Drugs and Drug Addiction
Weisburd D, Mazerolle LG (2000) Crime and disorder in drug hot spots: implications for theory and practice in policing. Police Q 3:331–349
Zahn M (1980) Homicide in the twentieth century United States. In: Inciardi JA, Faupel CE (eds) History and crime. Sage, Beverly Hills

Drug-Related Crime

▶ Drug Substitution Programs and Offending

Dutch Colonial Police

Marieke Bloembergen and Ellen Klinkers
KITLV, Leiden, The Netherlands

Overview

This entry discusses the development of the Dutch colonial police forces in the late nineteenth and twentieth century in the context of colonial state formation. The differences in the organization of the police forces and the practice of policing in the Dutch colonies (the Dutch East Indies, Suriname, and the Dutch Antilles/Dutch Caribbean) were significant, but political policing and a difficult relationship between police and community were overall the result of a weak and inefficient colonial state. Decolonization did not imply a breach of Dutch involvement with the police in the (former) Dutch West Indies. In Indonesia, the independent republic that replaced the Dutch East Indies after a period of war, Japanese Occupation, military actions, and diplomacy (1942–1950), the breach was more definite. However, the history of the decolonization of the Indonesian police still needs to be investigated in order to understand the continuities and discontinuities in policing and police methods in colonial and postcolonial Indonesia.

International Perspectives and Key Issues

The end of the twentieth century marked a turning point in the research on colonial police and state formation. Failing humanitarian interventions and unsuccessful international peace missions in those years may have provoked this change. The image of a repressive police, effectively maintaining colonial power, dominated academic historical writings until well into the 1980s (Anderson and Killingray, 1991, 1992; Arnold, 1986), but since the late 1990s and early 2000s, historians are inclined to emphasize the fundamental weakness of the colonial state and thus the ineffectiveness of the police (Bickers 2004; Campion 2002; Chandavarkar 1998; Sinclair 2006). The research on Dutch colonial policing on which this present entry is mainly built (Bloembergen 2009, 2011a, 2012; Broek 2011; Klinkers 2011) links up with this approach and is in particular inspired by the insights of Rajnarayan Chandavarkar (1998) and David Campion (2002) that police violence was the result of a weak and inefficient state rather than a strong, decisive state. The failure of the colonial police was caused, as they argue, not only by inadequate management but also by a difficult relationship between police and the community.

The tensions within the twofold task of the police – to maintain order and to provide in the social need of security – were more manifest in colonial societies with their racial biased organization structures and fragmented power relations than elsewhere: the cooperation of the local population was required for providing safety and protection, but at the same time, the police had to maintain order on behalf of a foreign regime which could provoke hostility and popular opposition. The police was therefore faced with unsolvable dilemmas that incited repression, in such a way that the challenge became not how the colonial state maintained control through the police, but how they could achieve this in spite of the police.

It is a question whether the histories of the modern colonial police forces in the Dutch East Indies in Asia and the Dutch West Indies – consisting of Suriname and the Antilles (in the Caribbean world) – can be analyzed and understood within one integrated Dutch police system. On the one hand, the three organizations of police forces (for the Dutch East Indies, Suriname, and the Antilles) all emerged at the end of the nineteenth century in the context of modern colonial states-in-development; they shared as their main objectives the prevention and detection of crime, surveillance, and the maintenance of colonial power, while, in the context of a modern colonial state-in-development, colonial authorities and entrepreneurs waged similar debates about security and the organization and mission of the colonial police. The fear of a loss of control over the indigenous or colonized people in a changing colonial world and the colonial state's responsibility for the safety and protection of the inhabitants were key reasons for maintaining the colonial police. The state's concern for security legitimized colonial power, while security also seemed indispensable for economic development (Bloembergen 2009, pp. 71–106; 109–136).

On the other hand, however, the organization of the police forces in each of the Dutch colonies and their practice of policing differed importantly. Obviously, there were differences in scale and population. The colonial regime of the Dutch East Indies, after a period of military expansion until 1910, had to oversee a million population in an immense archipelago consisting of thousands of islands, stretching from the west of Sumatra to the east of New Guinea. However, the Dutch East Indies' modern colonial police forces, at the peak of their development in the early 1930s, consisted of only 54,000 men, at a population of 60 million souls (including less than 200,000 Europeans). Police control in the Dutch East Indies was therefore fragmented by definition and dependent on the effectiveness of local administration and security control. Meaningful, the semi-traditional village police – a service provided by male inhabitants of the villages – formalized by the colonial administration, was maintained throughout the colonial period.

Suriname is about 12 times smaller than the Dutch East Indies, and is about four times the size of the Netherlands, and thinly populated. The majority of the people live in the capital of Paramaribo and along the northern coast. Over three-fourths of Suriname is covered by tropical rainforest. The islands of the Antilles (Aruba, Curacao, St. Maarten, Bonaire, St. Eustatius, and Saba) are small, scattered widely over the Caribbean Sea, and densely populated. Until the 1950s, the colonial police in Suriname and the Antilles did not exceed a few hundred men each, with minimal specialization. These West Indian colonies were populated by Europeans and the descendents of African slaves.

In addition, Asian contract laborers and their descendents lived also in Suriname, while the interior was inhabited with small groups of Amerindians and Maroons. Thus, there were no indigenous systems of authority and control in Suriname and the Antilles.

The police in the Dutch East Indies was the largest and most professionalized of the three colonial police organizations. Nevertheless, the police in the Dutch East Indies was fragmented and revealed the weaknesses of the colonial state. Political consciousness and tensions in the colonies in the 1930s placed ideals of professionalization and modernization under pressure. Political policing became more prominent. The anxiety of the colonial state in the Dutch East Indies resulted in violent police actions, not diminishing but overshadowing policing out of care and protection. Policing in Suriname became in those years also subject to repression and gathering intelligence. In this period, the colonies shared politics of policing. Nevertheless, the outcome was different. While, after a violent decolonization war between 1945 and 1949, the independence of Indonesia was recognized by the Netherlands in 1949, a process of gradual decolonization of the Dutch West Indies was started.

As, in the end, the police forces in the Dutch East Indies and West Indies developed in distinct and mostly separate contexts, their histories will be dealt with in the following in more detail, separately.

Policing the Dutch East Indies

The lack of security was not an uncommon phenomenon in nineteenth century colonial Java. And policing had always been a tragic business. That is at least the idea that rises from colonialists' complaints about the dreadful performance of the Javanese *politieoppasser* (literal meaning: police caretaker) or the formidable sleep of the indigenous village wards – the two main tools of civil colonial policing during the nineteenth century. But with the army and a diverse flock of armed security guards and auxiliary forces at hand, the unsafe circumstances, especially in rural Java, never seemed a problem to the colonial authorities – at least until the last quarter of the nineteenth century. Only then, in the context of a "liberal" colonial state-in-progress and in the context of military, administrative, and economic expansion, worries were explicitly addressed – in public and governmental discussions – about rampant lack of security in Java. At the same time, a new ideal seems to emerge, the ideal of effective and good policing. The wracked result of this all was the police reform of 1897, the first of a series of three large police reforms – in 1897, in 1911–1914, and in 1918–1920 – by which the Dutch colonial state endeavored to get more control over the organization of security surveillance.

Fundamentals

The modern police of the Dutch East Indies, thus developed in the heyday of the "ethical policy" that held sway between 1900 and 1920, was the answer to a typical colonial problem: the struggle of a colonial state that wanted to be civilized but witnessed its legitimacy crumbling. Compelled to use force to enforce its authority, it nonetheless also sought to govern by consent. The modern police embodied these two contrary forces of the ethical policy, or the Dutch version of the *mission civilicatrice*, namely, simultaneous efforts to achieve development and control (Locher-Scholten 1981). Since the colonial police was the instrument used to pursue these diverse goals, it became a two-headed beast: in trying to safeguard the state's authority, it provoked resistance, while, in reaching out to fulfil society's need for security, it needed the cooperation of the local population (Campion 2002, pp. 1–2).

Three meaningful police reforms between 1897 and 1920 contributed to the installation of a *would be* modern police force, or the general police, that operated in the 1920s and 1930s. These reforms were all important steps in the colonial state's effort to gain control over security surveillance. After the reform of 1918–1920, the Netherlands Indies officially (although not effectively) had a centralized and uniform modern police force. It was in itself pluralistic: it consisted of (a) the *gewapende politie* (an armed militarized police force, in hands of European administration) and (b) the general police, with its many divisions – the administrative police (dating from the nineteenth century), a modern city police (created in 1912–1914, large towns), the field police (*veldpolitie*; created in 1918–1920 for the security surveillance of the rural areas), and, mainly meant for political control, the *gewestelijke recherche* (the regional investigation departments). The Attorney General was in charge of central control, the director of Interior Administration of the management of the police.

The modern Dutch East Indies' police force never had, nor could it have, a monopoly of security surveillance. Therefore, the colonial budget was too small, and the territory and the population number too large. Manned for more than 96 % by Indonesians and directed by a small minority of European staff, the size of the force peaked in the 1930s at 54,000, and this on a population of almost 60 million souls, which counted less than 200,000 Europeans. Only a very small minority of Indonesian policemen could reach the higher ranks of the police force. Through its organization and hierarchy, the police force therefore reflected an essential characteristic of the colonial state: namely, how this state worked both on principles of cooperation and inclusion, and on difference and exclusion (compare Cooper 2005). Due to the dualistic principle of colonial administration in the Dutch East Indies, the system of policing and surveillance remained moreover very much fragmented. The effectiveness of policing and surveillance depended to a large extent on the functioning of the local administration – a European administrative hierarchy or the *Binnenlands Bestuur* (Interior Administration), paralleled by an indigenous administrative hierarchy, or the *pangreh praja*, which was responsible for administrative police and village police. This village police, partly a colonial construction as well, was a forced service of guarding and patrolling, which male adult inhabitants of a village or of indigenous quarters (kampong) in the towns provided in turn. While ideally the modern colonial police tools were to encapsulate these local semi-traditional tools, the village police, whether effectively functioning or not, remained for economic reasons the core of colonial security surveillance at the very local level (Bloembergen 2012). Next to that, the modern colonial police force also depended on other local or private forms of security surveillance and on how the modern police tools interacted with these.

Trustworthy to her civilizing mission, the Indies' government, however, kept up the ideal of the police as a *good* police force, that is, "a force of men with high mental attitudes towards life, a good morale and a strong character" – this according to an official governmental enunciation in the Volksraad in 1931. The Police School in Soekaboemi, set up in 1914 (10 years before a comparable school existed in the Netherlands), was the symbol and promoter of this ideal. The curriculum of the police academy reflected the same ambivalent notion of fear and concern that shaped the modern colonial police, and thus the aims to empower and gain loyalty, and to civilize. Thus, police recruits, from the lowest to the highest rank, were trained to perform as a tool for (political) control and as a *civil* security tool – at least as a more civilized tool than the army. On the one hand, they were trained in semi-military discipline, the use of arms and physical exercises, and, after communists revolts erupted in West Java and on the west coast of Sumatra at the end of 1926 and early 1927, also on how to recognize a "communist conspiracy" – all reflecting the police needs of a police state; on the other hand, they were educated on the principles of a constitutional state, criminal law and justice, methods and rules of *civilized* modern policing, and the idea that postponement of violence should be a *leitmotif* (Bloembergen 2009:203–247; 299–332; Bloembergen 2011a, b). Also, they were meant to perform as a tool of "civilization": to bring order, safety, neatness, cleanliness – in short "light" in colonial society. This idea became part of the self-image of the policemen who were trained at the police academy.

The image of good, civilized police and the idea of civilization and neatness through policing seemed to have become more important even in the 1930s. With further development of Indonesian nationalism and fiercer political policing, it became harder for the colonial authorities to ignore opposition against colonial rule. After the violent repression of communist revolts in 1926–1927, mass arrests, and the internment of around 1,300 presumed communists without trial, the colonial government subsequently refined and extended the organization of political policing and enlarged the police force in general. The international economic crisis, which forced the government to cut down policing expenses, did not hamper the artificial image of *rust en orde* (law and order) or *zaman normal* (times of normalcy) by which the 1930s have been characterized. Meanwhile, the police, more visible because of the extension of political policing and being watched while watching, had become the standard for the quality of colonial government. For those groups who felt oppressed by the police, the police embodied "the dirty work of empire" (Orwell 2001 [1936]).

Consequences: Political Policing

Colonial anxiety was the main condition under which a centralized Dutch East Indies Intelligence Service was set up in 1916. In the context of modernizing colonial society, this anxiety was fed, in short, by peaks of panic about religious revolts and rural unrest in the countryside, a growing awareness of an indigenous movement towards association and progress (the *pergerakan*), the visible Japanese drive for expansion in Asia, and, therefore, double awareness of the insecure position of the Netherlands in Asia – a position which leaned on Dutch neutrality in international relations. After the installation of a Central Intelligence Service, set up during the Great War, and slightly reformed in 1919, the colonial government had – on paper – strong means to assess and control indigenous political activities: (1) a Central Intelligence Service (Politieke Inlichtingendienst, PID), in 1916, renamed Algemeene Recherche Dienst (General Intelligence Service, ARD), in the hands of the Attorney General; (2) a weekly survey of Javanese, Malay, and Malay Chinese newspapers (*Overzicht van de Inlandsche en Maleische Pers*, IPO); (3) the new Regulation for association and gathering, or article 8a, a tool in the hands of the police to act (=to attend/watch meetings and, officially when directed to do so, to intervene whenever "disturbances were expected") (Sunario 1926, p. 19); and (4) steady background intelligence, provided by the advisor for Native Affairs (Laffan 2002).

From ARD to the local administrators, from the modern police in all its diverse units to local spies, and back to the center, the will "to know" and means to arrive at knowledge were there. But did this result in perspicacity? Practice was rough.

In the night of 12 November 1926, the colonial police and administrators on Java were completely surprised by communist revolts that took place at several locations at the same time – in the main capital Batavia, in Bandung, and, most thoroughly, in Banten (West Java). Completely surprised they were again in January 1927, when another communist uprising occurred in the west coast of Sumatra. These uprisings, both organized by the Partai Kommunis Indonesia (Indonesian Communist Party, PKI), could occur despite the existence of a Central Intelligence Service, despite intense police surveillance, despite repression of strikes in the first half of the 1920s (organized by the Union of Tram and Railway workers), and despite forced exile of most of these organizations leaders. The latter measures could be implemented on the base of the so-called exorbitant rights of the governor-general, a regulation dating from 1854 which gave the governor-general the right to expel persons living in the Dutch East Indies when it was in the interest of colonial law and order. These violent interventions and repressive measures did have the effect that the preparations of the intended revolt were obstructed by loss of strong leadership, internal conflict, and miscommunications within the PKI divisions. Moreover, since the end of 1924, the colonial authorities notably feared organized unrest. Therefore, this was a double failure of the political police and the ARD: they were surprised by badly organized uprisings, which they had foreseen (Bloembergen 2009:247–259).

This clear failure of Governor-General Johan Paul van Limburg Stirum's expressed wish, in 1916, to be "informed [. . .] on what's going on in the Native's mind," was above all a failure of central control of the political police. The main reasons are as follows: due to the fragmentation of colonial authority, political policing remained to a large extent a local business. Miscommunication, due to language problems, unbridgeable distances, misunderstandings, and conflicting interests of police, administrators, and local spies did the rest. And this miscommunication was enhanced by the mechanisms of secrecy and misleading that accompanied spying and the PKI's preparations of a revolt. As one police expert would point out later on, despite ARD, central police control, and regional intelligence forces, there was no effective interlocal (nor supra-local) collaboration between police forces and administrations.

The way the colonial state subsequently repressed the communist uprisings was a typical solution of a colonial state that went on the defensive: what followed was mass repression, mass custodies, and mass internments to a special camp in Boven-Digoel (in short "Digoel") in New Guinea – and this without any form of trial. The PKI was no longer tolerated.

The communist revolts moreover stimulated a general enlargement of the police force: the set up of a refined system of intelligence gathering and reporting (the *Politiek-Politioneele Verslagen*) and the set up of stronger guidelines for recognizing "radicalism" and keeping firm while watching. The memory of the fierce repression of the communist uprisings and the spectre of "Digoel" moreover made clear to nationalist/anti-colonial opposition, where the government drew the line. In that sense, 1927 was a turning point: a more pronounced policy of political policing made political police and the police state-in-development more effective – or so it seemed (Bloembergen 2009, pp. 247–298, 2011a).

The system of political policing still fell prey to bureaucratic retarding and the inability to process more information, to miscommunication, and to misunderstandings as well. Moreover, because the political police were, in their task of political control, guided by the search for signs of communism, they further blinded the state. The system of political surveillance therefore stimulated the inability and disinclination of the majority of the colonial government to try to understand the aims, aspirations, reach and depth of the national movement, and other drives for progress among the indigenous population. The system was also the product of this inability and disinclination.

Moreover, precisely because of the scale, violence, and visibility of their repressive actions, the police also showed the fragile position and hampering legitimacy of the colonial state. While the police looked out, the public watched, criticized, and ridiculed the police. In the end, the colonial police therefore was not so much a vehicle for discipline, but a vehicle for the loss of prestige and corrosion of colonial authority. This is what made colonial policing a dirty task, and in the end, a tragedy.

Future Directions: Decolonization

This entry does not cover the history of the police during the Pacific War and the Japanese occupation of the Dutch East Indies (1942–1945) nor the issue of policing and decolonization in the Dutch East Indies/the Indonesian Republic, during the Indonesian revolution and Dutch military aggressions in the period 1945–1949. These histories still need to be investigated, although Ambar Wulan (2009) made a beginning with her study on the Indonesian Republican police and intelligence gathering during the Indonesian Revolution, 1945–1949. To be short, for these periods there are, to a certain extent, continuities with prewar colonial policing: the problem of the legitimacy of the colonial state, and thus of the police, only had become more acute. The practice of policing and the nature of the security problems, however, also changed, becoming even more complicated and more pluriform. The Dutch colonial police, for several reasons, could no longer play the role it had played before the Pacific War. They resumed their tasks in a society transformed and also unsettled by the Japanese occupation and war economy, a society involved in a decolonization war, and a society in a permanent state of civil wars. Moreover, the colonial police had to concur with the new police force, developed by the Indonesian Republic in the same period. These circumstances complicated the tasks of policing and the issue of police loyalties. A study of the colonial police in this period would have to address the relationships and interactions between this force and the colonial army and Dutch troops during the two military aggressions in 1947 and in 1948–1949 (notably referred to at the time as police actions) and those between this force, the Republican police force, and the various Dutch and Indonesian security forces developed in this period. The question concerning the functioning of the colonial police during this period is, therefore, not easy to answer and deserves new research.

Policing the Dutch West Indies: Suriname and the Netherlands Antilles/ Dutch Caribbean

The marechaussee in Suriname became operative on the first of July 1863, the same day that 33,000 slaves became free citizens in Suriname. This newly established police force symbolized the historical transformation of colonial society. The developments in the police force of Suriname in the late nineteenth century were closely intertwined with the changes in a society adapting to the abolition of slavery. Society was on the move. Social boundaries were no longer determined by slavery and freedom. Disciplining and law enforcement, until then largely based on military and plantation discipline, had to be reshaped. In addition, society became more complex because of the arrival of indentured laborers from Asia and the West Indies. Chinese, more than 34,000 British Indian laborers, and almost 33,000 Javanese immigrants from the Dutch East Indies moved to Suriname to work the plantations, while increasingly ex-slaves found a living outside the plantation economy. The unfolding developments of the Surinamese police force show the struggle of the colonial authorities to accept, mold, and structure the socioeconomic changes in colonial society in the aftermath of slavery. The search for a suitable colonial police force resulted in a series of reorganizations which came to a conclusion with the founding of the *korps gewapende politie* (armed police force) in 1895.

Maintaining colonial power was a major concern for the planters and local authorities. The police had to be efficient, leaning towards military discipline, but also decent and generous towards the people to legitimate state control and to realize its civilization mission. That is, to create a society, conform western European culture with people willing to work for the benefits of colonial prosperity (Klinkers 2011).

Fundamentals

In November 1862, more than 6 months before the abolition of slavery, the minister of Colonial Affairs gave permission to the governor of Suriname to introduce a new colonial police force, *het korps marechaussee* (marechaussee). On the Dutch island of Curaçao, a brigade of marechaussee had been operating since 1838. This brigade was not an institutional part of the ministry of war like the Royal Marechaussee in the Netherlands but stood under the command of the Attorney General. As a result, the colonial authorities in Suriname chose unanimously for the installation of a civilian police force with a strong military character. Military of European descent and encamped in Suriname were recruited to join the marechaussee. The policemen were supposed to live in barracks to keep a safe distance from the local community.

The Attorney General had a difficult time to maintain the marechaussee up strength. The solution for this shortage of personnel was sought in the foundation in 1868 of a second police force, the *korps inlandse politie* (inland police force). The inland police force would be manned with creoles, that is to say, descendents of former slaves in Suriname. The entrance of Creoles into the police force seemed inevitable, but their incorporation into the marechaussee was considered problematic. Racial prejudices certainly abounded. The obsession with the separation of local and European policemen in different police forces was remarkable, since Suriname had a long history of cooperation between black and white men in maintaining law and order. Black overseers were considered as mediators between slaves and whites at the plantations. Besides, patrols of black (both enslaved and free) and white men pursued runaway slaves in the interior of Suriname in the seventeenth and eighteenth century. The racial separation in the post-abolition police force can be understood as a redefinition of social boundaries in the aftermath of slavery (Compare Cooper and Stoler 1997, p. 7). The marechaussee was supposed to represent the colonial state and symbolized the unbroken supremacy of the white colonial population after emancipation.

However, there was no fundamental difference in the police practices of both forces. The removal of the institutional distinction between inland police and marechaussee took decades with endless discussions. Eventually Governor T.A.J. van Asch van Wijck decided to turn daily practice into policy and merged the two systems into one armed police force. The institutional changes did not fundamentally change the system. What remained was one police force with a military character, with a growing shift from European to local personnel through time. The colonial authorities failed to establish a strong police force during the twentieth century, because of the lack of financial means. Besides, the authorities seem to distrust the local policemen and clang to the army as a loyal ally of colonial power. In addition, the rise of (semiprivate) policemen fractured the system of law enforcement even more.

The initial idea to organize the police systems in the colonies of the Dutch West Indies in the same way faded to the background. The Antillean police force passed through several stages during colonial times, beginning with the *brigade koloniale marechaussee* (colonial marechaussee) (1839–1918), followed by *korps veldwacht* (rural police) (1873–1918, 1932–1949), *detachement marechaussee* (marechaussee) (1911–1918), *korps burgerpolitie* (civilian police) (1918–1949), and *korps militaire politietroepen* (military police troops) (1928–1949). The police was highly militarized and had, in contrast to the Surinamese police, a large influx of Dutch military and civil policemen throughout all its stages, except for the rural policemen. Personnel was distributed over six islands, but the majority was located on Curaçao. The expansion of the oil industry in the 1920s, requiring security and causing a large recruitment of people on the island, had been a reason for reorganization, expansion, and further militarization of the police system with the formation of a *korps militaire politietroepen* besides the already existing *korps burgerpolitie*. The bifurcated police system did not live up to the expectations of a strong and efficient police force. The weakness of the system of maintenance order became apparent when a small group of Venezuelans overpowered the police station in a fortress on Curaçao in 1929 (Broek 2011). The Venezuelans under command of R.S. Urbina not only succeeded in overpowering the fortress but also captured Governor L.A. Fruytier for almost 1 day. The incident was considered as an embarrassment for colonial power.

Consequences: Political Policing

Ideals of an effective and a respectable police force, of civilizing society and its people, were frequently at odds with maintaining colonial power and control. This happened in uncertain times, as in the years following the abolition of slavery, but became most apparent in the 1930s. The global economic crisis not only caused poverty and unemployment in Suriname and Curaçao but raised also political consciousness and opposition against the colonial administration. Surinamese laborers who had been employed in the oil industry on Curaçao in the 1920s lost their jobs and returned to their home country, disillusioned but inspired by the island's labor unions and left-wing press (Ramsoedh 1990, p. 33). The Surinamese colonial authorities feared that communism and nationalism would conquer influence, a fear that was stimulated by the communist revolts on Java and Sumatra in the Dutch East Indies in 1926 and 1927 (see above).

In Suriname, repression and political policing became more manifest than ever before. The press, labor unions, and other signs of political consciousness were not tolerated and repressed. An anti-revolution law was announced in 1933 to defy the feared threat of communism and to maintain colonial order. Most illustrious was the arrest of the Surinamese writer Anton de Kom who had been involved with communist and nationalist groups during his sojourn in the Netherlands. His decision to travel to Suriname to talk about the history of slavery caught the attention of the Dutch intelligence service that was surveying de Kom for some time already. They warned Governor A.A.L. Rutgers that he was suspected to poison the Surinamese people with anti-colonial thoughts. As a result, de Kom stood under police surveillance continuously, was forbidden to speak in public, and detained several times. A group of people gathered awaiting his release after he had been arrested again. The police ended this protest with force; 22 people were wounded and two were killed (Woortman and Boots 2009, pp. 63–135).

In this period of anxiety, the colonies shared politics of policing. The authorities in Suriname and the Antilles informed each other about possible security threats and measures. After the appointment of J.C. Kielstra as a governor of Suriname in 1933, the events in the Dutch East Indies gained importance. Kielstra had been a colonial administrator in the Dutch East Indies. Kielstra increased the repressive regime further, even though there were no indications that communism and nationalism would manifest themselves in Suriname as strongly as in the Dutch East Indies (Klinkers 2011, pp. 105–156).

Future Directions: Decolonization

Suriname and the Dutch Antilles became autonomous parts of the Dutch Kingdom in 1954 as stipulated in the 'Charter for the Kingdom of the Netherlands.' The Netherlands remained responsible for foreign affairs, defense, and good governance (Oostindie and Klinkers 2012, pp. 21–27). The police were internal Surinamese and Antillean affairs from then on. The Dutch gave up leading positions in the police organization in Suriname, which was already mainly staffed by local men reflecting the multicultural society. The Dutch input in the Antillean police forces was and would maintain relatively strong.

However, self-rule did not imply a breach of Dutch involvement in Surinamese police organization. Schooling and training, technical assistance, and cooperation between Dutch police and those in the West Indies intensified, while joint international crime investigation became more important. The organizations in Suriname and the Antilles started to expand and professionalize from the 1950s onwards. In Suriname, special branches, such as an intelligence service, children's police, and traffic police were founded. The armed police force changed its name into *korps politie Suriname* (Suriname police force) in 1973, emphasizing the civilian mission of the police, even though military aspects in its presentation and protocol would be uphold. The ineffectiveness of the fragmented Antillean police system had been acknowledged before. The three forces (civilian police, military police troops, and the village police) were transformed into the *korps politie Nederlandse Antillen* (Dutch Antilles police force) in 1949 (Broek 2011, pp. 147–149).

The fear that the independence of Indonesia would become a source of inspiration for the people in Suriname was a reason for gathering intelligence in the 1950s. But more than the struggle for independence, the Dutch government – encouraged by the United States – feared that communism would spread as an ink spot in the Caribbean area after the Cuban revolution of 1959. The Dutch government's secret intelligence service, *binnenlandse veiligheidsdienst* (BVD), assisted in the founding and professionalization of the Antillean police forces and Surinamese intelligence services (*veiligheidsdienst Nederlandse Antillen, centrale inlichtingendienst Suriname*), both subdivisions of the local police organizations. The cooperation was not entirely successful because of mutual distrust, reason for the Dutch BVD to work together with the marines on the Antilles and the Dutch troops in Suriname.

The Dutch government had to bring foreign policy into practice in a changing colonial reality. The work of army and police, intertwined during colonial times as both had been occupied with the internal security, had to be unraveled. The army or marines were still allowed to assist the police in times of crisis, but when and how this was supposed to happen was hard to determine. A conflict between Suriname and Guyana police force about territorial claims in the southwestern border area in 1968–1969 and a revolt on Curaçao in 1969 demonstrate how difficult it was to act in a mutually acceptable way and according to 1954 Charter.

The border dispute between Suriname and (British) Guyana harks back to the nineteenth century but came to a crisis after Guyana became independent in 1966. After the Guyana police force dispatched Surinamese workers out of the area, the Surinamese government demanded a military response to the Guyanese action. The Dutch government sympathized with the Surinamese territorial claims but refused military assistance which would intensify the conflict. The Surinamese government founded in response a special police unit, the *defensie politie* (defense police), to guard and protect the territory. Eventually, the conflict would never come to an armed clash, but simply lost its urgency. Without coming to a conclusion the territorial claims still lingers on (Klinkers 2011, pp. 157–244).

The labor conflict in Curaçao which ended in a revolt on May 30, 1969, challenged the Charter even more. Thousands of demonstrators marched through the streets of Willemstad, looting and burning. The police killed two people. Eventually, the local authorities called the Dutch Marines for assistance to end the uproar. This military action was successful in restoring order but was criticized at home and abroad for this neocolonial act (Broek 2011, pp.153–190; Oostindie 1999). The revolt changed the attitude towards the Charter. It made the Dutch government realize that their possibilities for intervention to guarantee good governance overseas were limited. Besides, the continuing dependency of Dutch financial aid and the unlimited stream of migrants were other consequences of the Charter. The government in The Hague headed for independence of the West Indies, which was accepted in Suriname in 1975. The Antilles, on the other hand, refused sovereignty and are connected with the Netherlands until today, even though the ties loosened and the islands are no longer united in one country since October 2010 (Oostindie 2011). Aruba took the lead and received a separate status in 1983 already; the *korps politie Aruba* was founded in 1985 as a result of new responsibilities and legal arrangements. As international crime, like drugs and human trafficking, increased, the cooperation with and investments in the (former) Dutch Caribbean police systems remained of major importance for both areas.

Related Entries

▶ French Colonial Police
▶ High Policing

Recommended Reading and References

Ambar Wulan G (2009) Polisi dan politik: Intelijen kepolisian pada masa revolusi tahun 1945–1949. Rajawali pers, Jakarta

Anderson DM, & Killingray D (eds) (1991) Policing the empire: Government, authority and control, 1830–1940. Manchester University Press, Manchester

Arnold D (1986) Police power and colonial rule: Madras 1859–1947. Oxford University Press, Delhi

Bickers R (2004) Empire made me: an Englishman adrift in Shanghai. Penguin, London

Bloembergen M (2007) The Dirty work of empire: modern policing and public order in Surabaya. Indonesia (Cornell University) 83:119–150

Bloembergen M (2009) De geschiedenis van de politie in Nederlands-Indië: Uit zorg en angst. Boom/KITLV, Amsterdam/Leiden

Bloembergen M (2011a) The perfect policeman. Colonial policing, modernity and conscience on Sumatra's West Coast in the early 1930s. Indonesia 2011:165–193

Bloembergen M (2011b) Being clean is being able: policing colonial *netheid* and gay vices in the Netherlands-Indies in the 1930s. In: van Kees D, Taylor J (eds) Cleanliness and culture: Indonesian histories. KITLV Press, Leiden, pp 117–146

Bloembergen M (2012) Vol, meurtre et action policières dans les villages Javanaises: Les dynamiques locales de la sécurité aux Indes néerlandaises dans les annees 1930. Genèses 82:8–36

Bowling B (2010) Policing the Caribbean. Oxford University Press, Oxford

Broek Aart G (2011) De geschiedenis van de politie op de Nederlandse Caribische eilanden, 1839–2010: Geboeid door macht en onmacht. Boom/KITLV, Amsterdam/Leiden

Campion DA (2002) Watchmen of the raj: The United Provinces police, 1870–1931 and the dilemmas of colonial policing in British India. Dissertation, University of Virginia

Chandavarkar R (1998) Police and public order in Bombay, 1880–1947. In: Rajnarayan Chandavarkar, imperial power and public order in Bombay. Cambridge University Press, Cambridge, pp 181–233

Cooper F (2005) Colonialism in question: theory, knowledge, history. University of California Press, Berkeley

Cooper F & Stoler A (eds) (1997) Tension of empire: colonial cultures in a bourgeois world. University of California Press, Berkeley

Cyrille F (1998) Politiemodellen beïnvloeden elkaar wederzijds: Het beleid aangaande het politiewezen in

Nederlandsch-Indië. Het Tijdschr voor de Polit 60:4–21
Gouvernementsblad Suriname
Klinkers E (2011) De geschiedenis van de politie in Suriname, 1863–1975: Van koloniale tot nationale ordehandhaving. Boom/KITLV, Amsterdam/Leiden
Laffan MF (2002) “A watchful eye”; The Meccan plot and changing Dutch perceptions in Indonesia. Archipel 63:80–108
Lamur HE (1973) The demographic evolution of Surinam 1920–1970: a socio-demographic analysis. Martinus Nijhoff, The Hague
Locher-Scholten E (1981) Ethiek in fragmenten: Vijf studies over koloniaal denken en doen van Nederlanders in de Indonesische archipel, 1877–1942. HES, Utrecht. [Ph.D.-thesis, Leiden University]
McCoy Alfred W (2009) Policing America’s empire: The United States, the Philippines, and the rise of the surveillance state. University of Wisconsin Press, Madison
Oostindie G (1999) Curaçao 30 mei 1969: verhalen over de revolte. Amsterdam University Press, Amsterdam
Oostindie G (2011) In: Spapens T, Groenhuijsen M, Kooijmans T (eds) Een koloniaal instrument: De politie in de Antillen en Suriname. Liber Amicorum Cyrille Fijnaut. Antwerpen/Cambridge, Intersentia, Universalis
Oostindie G, Klinkers I (2012) Gedeeld Koninkrijk: De ontmanteling van de Nederlandse Antillen en de vernieuwing van de trans-Atlantische relaties. Amsterdam University Press, Amsterdam
Orwell G (2001) Shooting an elephant. In: Davison P (ed) Orwell and politics: animal farm in the context of essays, reviews and letters selected from the complete works of George Orwell. Penguin, London, pp 17–24. [Originally published in 1936]
Ramsoedh H (1990) Suriname 1933–1944: Koloniale politiek en beleid onder gouverneur Kielstra. Delft, Eburon
Sinclair G (2006) At the end of the line: colonial policing and the imperial endgame 1945–80. Manchester University Press, Manchester
Sunario (1926) Het recht van vereeniging en vergadering der Indonesiërs. Perhimpoenan Hakim Indonesia, Leiden
Woortman R, Boots A (2009) Anton de Kom: Biografie. Uitgeverij Contact, Amsterdam/Antwerpen

Dutch Crime Networks

Toine Spapens
Tilburg University, Tilburg, The Netherlands

Introduction

This entry offers a brief history of Dutch crime networks. There have always been gangs in the Netherlands that were involved in serious crimes, such as violent robberies, extortion, large-scale theft, and the supply of illegal goods and services, and that is still true today. In that respect, Dutch networks do not differ from crime groups in most other countries. The reason to address these networks separately is the fact that, particularly since the 1970s, Dutch crime groups have managed to establish a far-flung international network for the import and reexport of different types of narcotic drugs. Furthermore, since the 1990s, the Netherlands has also developed into a major producer and exporter of synthetic drugs and cannabis. A key feature of the Dutch networks is that they organize these activities primarily in loose-knit cooperatives revolving around criminal entrepreneurs who transact business with others in shifting coalitions. To some extent, it is even more appropriate to view them as a single network (a *meso network*) from which temporary cooperatives of varying composition (*micro networks*) spring up (Spapens 2006, 2010, 2011, 2012).

Here, Dutch networks are defined as groups composed of lawbreakers who live in the Netherlands. This does not necessarily mean that they are Dutch nationals. What is important, however, is that they must have criminal connections to other network members. Indeed, without such “criminally exploitable ties,” they would be unable to take part in the coalitions. This definition therefore excludes itinerate gangs that also operate on Dutch territory and sometimes stay there for longer periods, but without visible connections to indigenous criminals.

Section “Bandits, Smugglers, and Local Providers of Illegal Goods and Services” starts with a brief description of the historical roots of Dutch networks. Section “The Emergence of ‘Dutch-Style’ Organized Crime” then addresses the emergence of Dutch-style organized crime from the 1970s onwards, resulting mainly from a shift towards drug trafficking. Section “Large-Scale Drug Production” focuses on the production of synthetic drugs and cannabis. Section “Schengen and Its Effects” goes on to discuss the most recent trends and developments in Dutch crime networks. Finally, Section “Conclusion” offers some concluding remarks.

Bandits, Smugglers, and Local Providers of Illegal Goods and Services

Criminal networks do not spring up out of the blue, and it is therefore important to take a brief look at the historical roots of organized crime in the Netherlands. Large gangs of bandits plagued the Low Countries as far back as the seventeenth and eighteenth centuries, as they did other parts of Northwest Europe (cf. Fijnaut 2013). In the Netherlands, such groups were particularly active in the present-day province of North Brabant, at the time a very rural and densely forested area. Although it became part of Holland after the end of the 80 Years' War, Brabant (as it was then known) was not a member of the States General and treated almost as a colony. During the Dutch Revolt, it had been a principal battleground, with both sides committing atrocities against the local population (Adriaensen 2007). As a result, the central government had little legitimacy with the people of Brabant. They supported the gangs, and the local authorities were susceptible to corruption. Moreover, the gangs usually set up camp on the borders of jurisdictions because this often led to disagreements between the respective magistrates about who should take action, resulting in inactivity. Finally, the brigands also made sure not to commit any crimes within the jurisdictions where they resided.

One of these bands, the White Feather gang based in Kaatsheuvel, totaled almost a hundred members, including women and children (Grootswagers 1983). Heavily armed and militarily organized – some of their ranks consisted of mercenaries and deserters who had stayed behind after the end of the 80 Years' War – the gangs committed burglaries and violent robberies in the countryside and extorted farmers by threatening to set fire to their farmhouses unless they paid them off (Egmond 1994). The gangs were highly mobile and operated in parts of the Netherlands outside Brabant, in what is now Belgium, and in the German border areas. Their demise finally came after 1798 when, under French influence, the State was unified and gendarmerie units were assembled to operate in the countryside. In only a few short years, the gangs were history. Some of their members' children, however, would carry on the legacy.

New opportunities arose when Belgium gained its independence in 1830, and it became lucrative to smuggle all sorts of goods back and forth across the new border. Sint Willebrord ('t Heike), a hamlet that had once been the lair of brigands, was located close by, and it rapidly turned into a notorious smugglers' den. Large-scale smuggling continued until 1960, when the Benelux Customs Union came into effect and border checks were abolished. After that, some smugglers reverted to large-scale theft, for instance, cars and armed robberies, or continued trafficking in illegal goods, notably amphetamines. Soon, however, they found other more rewarding illegal activities.

Other parts of the Netherlands proved less fertile ground for committing large-scale serious crimes. Government legitimacy was much stronger there, and because most people lived in towns, law enforcement was easier. Things changed when the Industrial Revolution – which only got going in the Netherlands late in the nineteenth century – led to the rapid growth of the cities in the western part of the country. Newcomers there faced poor living conditions, and poverty and the loss of existing social structures led to problems such as crime and alcoholism. Such problems were harder to control in the growing cities, especially because the police force had only just begun to refashion itself into a professional organization. Like the United Kingdom, which faced similar problems on a much larger scale, there was a revival of moral values in the Netherlands. From the second half of the nineteenth century onwards, the upper class did much to improve housing and hygiene as well as organize better welfare schemes for those who were unable to support themselves (De Swaan 1989). Furthermore, the government passed strict laws on vices such as gambling and prostitution. The government even considered a ban on stock market speculation, because it viewed it as a form of gambling (!), but parliament failed to pass the proposal. This helped lower the general crime figure on the one hand, but on the other, people still wanted to have "fun," and this inevitably created bigger market opportunities for those who could provide it. A *penoze* (urban underworld) emerged, and "local heroes" would occasionally crop up who were involved in all sorts of illegal activities simultaneously – illegal gambling, prostitution, trafficking in illegal firearms, fencing stolen goods, small-scale drug dealing, and sometimes extortion – and who also had a stake in bars and restaurants.

These activities were usually concentrated in specific neighborhoods, the Red Light District of Amsterdam being one example. Although the famous red-tinted windows did not appear until 1930, the district had always been a rough place where sailors and city folk alike came for prostitutes, gambling, and drinking. Moreover, the Red Light District was always an international meeting point. Opium use, for example, was already a problem in the Chinese community before the Second World War, and inevitably, friendly relations with other members of the districts' underworld allowed for the supply to Dutch customers as well, albeit on a small scale (Wubben 1986).

Still, these developments are not exceptional. Bands of brigands were also common in other parts of Northwestern Europe. Hamburg, for example, has its Reeperbahn, which is comparable to the Red Light District. And while in 1960s Amsterdam had "local heroes" like "Zwarte" (Black) Joop de Vries, London had the Kray Twins. Dutch networks as these are now known only emerged in the mid-1970s, and the key to this was the wholesale trade in hashish and marihuana in particular.

The Emergence of "Dutch-Style" Organized Crime

Youth culture developed rapidly in the Netherlands in the 1960s, as it did in other western countries. One result was the increasing scale of narcotic drugs use, particularly hashish and marihuana. These drugs had to be imported from abroad and the main source countries then were Pakistan, the Lebanon, and Morocco. At first, the drugs were smuggled in small quantities by young people, often users themselves, who had traveled to the east. As demand increased, traditional criminals also started to discover the market, and some of the "hippies" managed to expand their businesses as well. In 1974, for example, the coast guard intercepted a fishing trawler, the "Lammie," that carried two tons of hashish. Investigations revealed that some well-known members of the Red Light District underworld had organized the transport (Middelburg 2001).

Although the Netherlands was a party to the UN Single Convention on Narcotic Drugs, its government has always been relatively lenient towards drugs use and the possession of small quantities for personal use. It was agreed that drug addicts should be given help or treatment, and not serve as law enforcement targets (Leuw 1994, p. 34). With regard to cannabis, the government decided in 1976 to make a distinction between soft drugs (i.e., hashish and marihuana) and hard drugs (all other narcotic drugs). The Board of Procurators General subsequently issued a directive stating that the public prosecution service would no longer prosecute soft drugs possession as long as the amount did not exceed 30 g. Blanket legalization was impossible because that would violate the Single Convention. Logically, this also meant deciding how to act towards dealers selling soft drugs to customers. The 1970s, for example, saw the emergence of coffee shops selling hashish and marihuana. In 1979, another directive ordered prosecutors to refrain from active investigation of dealers and coffee shops unless they put public safety or health at risk, or openly tried to promote and expand their business (Spapens and Van de Bunt forthcoming).

Although historical research is lacking, it is clear that these changes also expanded the market for crime groups importing narcotic drugs. To begin with, coffee shops, particularly in Amsterdam and the border regions, started to attract large numbers of foreign youth who wanted to try out soft drugs without the risk of apprehension. As there was no system of licensing in place yet, the same crime groups that imported the drugs could also open coffee shops. Furthermore, criminals from the south of the country, who had accumulated capital from smuggling and armed robberies, teamed up with their counterparts in the western cities and started to invest in large shipments. The Dutch importers quickly succeeded in establishing business relations with wholesale producers in source countries.

In the mid-1980s, concern grew within the Dutch police force about such groups developing into organized crime syndicates of mafia-type proportions (Sietsma 1985). Moreover, when cocaine became fashionable in the 1980s, the importers began to serve that market too. In particular, Klaas Bruinsma's network seemed to be bringing increasing quantities of drugs into the country, and it tried to take over other businesses as well, such as the installation of gambling machines in bars and restaurants. Law enforcement response was slow, however, and it took until 1991 before the first special investigation squad was set up. That same year, Bruinsma was murdered, so the investigation instead focused on his "heirs," supposedly a trio that had taken over the "management of the organization," which was code-named Delta.

The competent authorities abruptly dismantled the investigation team in 1993 when an undercover operation spiraled out of control. It involved a criminal informer whom agents allowed to bring large quantities of drugs into the country in order to build trust with the alleged top leaders of the Delta group. That way, the police would be able to obtain evidence against them. After 2 years, however, the operation was not any closer to achieving its goal, whereas the informer had imported thousands of kilos of soft drugs, and possibly cocaine as well, with the consent of the police, who also allowed him to sell it on the illegal market. This resulted in a major scandal that led to a parliamentary enquiry.

Although one of the causes of the problem was the inadequate regulation of special investigative powers – the parliamentary enquiry revealed that other investigation teams had used similar methods – another was the fact that the police wrongly viewed these crime groups as well organized and hierarchically structured. Instead, they more closely resembled a network organization. "Entrepreneurs" within these networks continuously formed different coalitions in order to import and sell wholesale quantities of drugs. As far back as 1990, Dutch criminologists also came to the conclusion that "Dutch-style organized crime" was an amalgam of loosely knit networks instead of well-structured "firms," and this was confirmed in subsequent studies (Van Duyne et al. 1990; Van Duyne 1995; Fijnaut et al. 1996; Kleemans et al. 1998, 2002; Klerks 2000).

According to the literature, Dutch networks generally operate in line with the following basic model, which can be applied to drug trafficking and to other types of organized crime involving illegal goods.

A typical criminal cooperative consists of about 10–30 people. To begin with, most coalitions consist of a stable core made up of a criminal entrepreneur and a two or three well-trusted associates, usually family members or longtime friends.

Around this core, there is a second shell consisting of personnel carrying out the more crucial tasks, such as arranging cover-up loads. Unskilled workers, who are the most numerous, also stem from relatively close social circles, usually from the "extended family." Specialists, however, such as people who can set up complex money laundering schemes, often come from outside the direct criminal *milieu*. A good example of this is Willem Endstra, a real-estate dealer who also invested money on behalf of criminals. Specialists may get involved in criminal activities because of financial problems, or because they become friendly with a criminal entrepreneur and are attracted by the money or the excitement. Criminals may also deliberately maneuver people with specific skills into a situation of financial or other type of dependency and then more or less force them to cooperate (Spapens 2006).

The third shell of accomplices takes care of the high-risk tasks, such as the actual smuggling.

Although these persons know at least one of the other members of the group, they are kept in the dark about the extent of the activities.

The entrepreneur, aided by his core of long-term associates, strikes the business deals with other organizers at home and abroad, whereas the operational personnel varies with the requirements of the specific “projects.” These may differ depending on the type of crime – entrepreneurs are usually active in different fields simultaneously – but also on the agreements made in the business deals, for instance, who will be responsible for smuggling. If the other party, for example, has access to a proven smuggling route, it makes sense to let them take care of that part of the activities. Although the actual involvement of the workers varies, for practical reasons, they are often recruited from more or less the same pool of individuals.

The key persons of the crime groups are usually 30–40 years old, but some may continue their careers well past the age of retirement, interrupted of course by stints in prison. Being a leader of a criminal group involved in drug trafficking requires different types of human capital, such as a reputation in criminal circles, organizing skills, and contacts with suppliers and buyers abroad. In extended families, such capital is often transferred from one generation to another. Sons or sons-in-law – women still seem to play a modest role – can take advantage of the reputation and the network of their fathers, which enables them to engage in large illegal business deals quickly (Spapens 2006).

Immigration is another important factor. Nowadays, tens of different nationalities live in the Netherlands, and members of these communities may be able to bridge the gaps between supply and demand for specific illegal goods and services between their countries of origin and the Netherlands. For example, in the 1990s, Surinamese nationals were involved in the trafficking of cocaine to Europe; Chinese groups in the trafficking of migrants from China to the United Kingdom, but also in ecstasy production in the Netherlands; Turkish families in the trafficking of heroin from Turkey, but they also smuggled ecstasy back to the Turkish Riviera; Moroccan groups imported hashish; and former Yugoslavians smuggled illegal firearms to the Netherlands and narcotic drugs back to their countries of origin. Groups of different nationalities are most certainly not islands within the Dutch *meso* network, and all sorts of contacts and business relations exist. When it comes to the trading of illegal goods, criminal groups usually do not specialize in just one commodity.

In the Netherlands, levels of violence and corruption are relatively low. Most of the narcotic drugs produced in (see Section “Large-Scale Drug Production”) or imported into the Netherlands are exported abroad. There is no indication of turf wars over contacts with suppliers and buyers. Violence is usually the result of business deals going wrong, cheating, or the theft of a shipment of illegal goods. Corruption is also relatively rare. It usually involves relatively low-level customs officers who are able to ensure that containers in which illegal goods are concealed can be safely brought into the country. Criminals may also bribe police officers to provide information on ongoing criminal investigations.

Although it took some time for law enforcement agencies to gain a clear understanding of the composition and working methods of the Dutch networks in the early 1990s, and to adjust their investigative techniques accordingly, they gained experience over time and managed to bring several major cases before the courts, after which soft drug imports did appear to decrease. As it turned out, however, that was partly because the networks discovered drug production as a lucrative activity in the mid-1990s.

Large-Scale Drug Production

Drug production had already started back in the 1970s, with the manufacture of amphetamine, and expanded from the early 1990s onwards, when ecstasy production and cannabis cultivation took off. Production and sale, however, cannot be treated separately. After all, there is no point in producing large quantities of drugs (or importing them, for that matter) if you are unable to find wholesale customers. Export is the key to this, because the Dutch domestic market is relatively small.

In the mid-1970s, crime groups, particularly those in the south of the country, started to produce amphetamines. Amphetamines were only included in the Dutch Opium Act in 1975, much later than elsewhere in Europe. Consequently, Dutch providers started to smuggle "speed" to dealers abroad. With the sales network in place, they also succeeded in taking up illegal production when amphetamines were finally criminalized.

Large-scale use of MDMA, or ecstasy as it is popularly known, started in the second half of the 1980s. Here, the story is comparable to that of amphetamine. The United Nations had already added MDMA to the list of controlled psychotropic substances in 1985, but it was not included in the Opium Act until 3 years later. Other countries were quicker to respond to the UN requirement. Logically, this once again resulted in the Netherlands developing into a source country. After 1988, illegal production quickly took off and by the next year, the police had discovered the first clandestine laboratory (Weijenburg 1996). Consumer demand for ecstasy skyrocketed in the early 1990s. Foreigners preferred Dutch ecstasy because of its good quality, and this resulted in a thriving export trade. The Dutch-Belgian border area in the south of the Netherlands became a heartland of ecstasy production. Here, criminals had already built up experience with synthetic drug production, and soon, most networks in this part of the country switched to manufacturing ecstasy or started to combine it with existing illegal activities. Apart from the south, production also appeared to concentrate in the west of the country, particularly around Amsterdam (KLPD 2005).

In the first half of the 1990s, dealers in other Western European countries, particularly in the United Kingdom, were the most important buyers of Dutch-produced ecstasy. In the second half of the decade, the United States also emerged as a very attractive market. Members of the Jewish Diaspora played a crucial role in establishing connections between Dutch suppliers and American wholesale buyers. They included extended families whose members lived in the United States, the Netherlands, and Israel. Groups of Chinese living in the Netherlands got involved in delivering PMK, an essential controlled chemical for the manufacturing of MDMA, which they imported from China, and after the turn of the century, they also began to set up ecstasy laboratories and smuggling large quantities of MDMA to Canada (Spapens 2006).

Under increasing pressure from the United States – in 2000, President Clinton personally expressed his anger about Dutch ecstasy flooding into his country to Dutch Prime Minister Kok – the Dutch decided to crack down hard on the crime groups producing MDMA. This resulted in a notable reduction in production. To begin with, the police successfully investigated a number of key players, who were sent to prison for lengthy periods (at least by Dutch standards). Next, diplomacy resulted in the Chinese government taking more effective measures against the illegal production of PMK in their country. More prosaically, the Dutch lost the still lucrative US market to Chinese crime groups, who, having learned the production process in the Netherlands, set up laboratories in Canada. Only Australia then remained as a major overseas destination, although small shipments carried by couriers traveling by plane may still go anywhere in the world. It is also important to note that ecstasy is now far less popular than in the halcyon days of the 1990s. Finally, Dutch criminals also had an excellent and far less risky alternative that generated major profits: cannabis cultivation.

The Dutch climate is generally not very favorable to growing cannabis, and marihuana users used to prefer the "pot" imported from abroad. This changed in the early 1990s with the introduction of indoor cultivation methods. Americans who came to the Netherlands allegedly introduced the technique (Boekhout van Solinge 2008). The Dutch then improved the product by raising the percentage of THC, the working component of marihuana. Hobby growers started to set up nurseries and began to supply the coffee shops. At first, even the government welcomed this development because it reduced the need for imports, which were largely in the hands of the Dutch criminal networks. In hindsight, this was quite naïve, because the same groups of course quickly recognized the new market (Weijenburg 1996).

At first, Dutch networks started to set up relatively small-scale cannabis nurseries in the homes of members of their extended network, usually in economically weak neighborhoods (Bovenkerk and Hogewind 2003). Because the police put little effort into investigating cannabis cultivation and the penalties for those caught were lenient – often limited to a fine or community service – volunteers who wanted to earn extra money were lining up. However, the Dutch networks found managing a large number of nurseries for people who themselves had little growing experience more labor-intensive than they liked. They also attracted attention because the criminals sometimes resorted to violence, particularly if they suspected the growers of embezzling some or all of the harvest and selling it themselves to make a bigger profit.

Around the turn of the century, the Dutch networks found better solutions. On the one hand, they opened "grow shops" (Spapens et al. 2007). Here, growers could not only buy all the necessary equipment and get advice, but they could also purchase cuttings and sell the harvest. On the other hand, the professional criminals started to set up very large plantations – 5000 to 50,000 plants – themselves or, even better, gave someone with the necessary skills the equipment and had him run the plantation for them at his own risk. The production process was outsourced, and entrepreneurs could go back to focusing on trading the product. Today, they usually sell the best quality homegrown cannabis to coffee shops, whereas medium and low-grade products find their way to foreign dealers (Spapens et al. 2007).

Schengen and Its Effects

In 1995, the agreement between the Governments of the States of the Benelux Economic Union, the Federal Republic of Germany, and the French Republic on the gradual abolition of checks at their common borders, better known as the Schengen Agreement, came into effect. In 1998, the Treaty of Amsterdam further expanded the Schengen zone. The "open borders" had both visible and less-visible consequences for drug trafficking from the Netherlands to other parts of continental Europe and particularly Germany and France, because fixed customs controls at the Dutch-Belgian border had already been abolished in 1960. Things did not change with regard to the United Kingdom, but the country remained one of the important destinations for narcotic drugs from the Netherlands.

The most visible effect was a sharp increase in the number of Belgian, German, and French drug tourists who visited the border towns in particular. Foreign drugs users were naturally not a new phenomenon in the Netherlands. As far back as the 1970s, heroin addicts, especially from Germany, came to the country for good quality and low prices. To avoid the risk of being checked at the border, they generally did not take home major supplies for personal use. The same applied to foreigners who came to the Netherlands to buy hashish or marihuana in a coffee shop. Because of these limitations, most drug tourists came from the border areas. "Schengen" meant that the risk of taking quantities of drugs across the border virtually disappeared, and it thus became attractive for users living further inland to drive to the Netherlands regularly. Furthermore, general economic prosperity led to many French and German youth owning their own cars and having enough money to drive several hundreds of kilometers round trip to buy drugs in a Dutch coffee shop. A few years ago, thousands of drug tourists a day came to border towns such as Maastricht, Roosendaal, and Terneuzen (Spapens 2008; Fijnaut and De Ruyver 2008).

Another effect was the emergence of drug-dealing houses in the border towns. Of course, not every drug tourist was interested only in hashish or marihuana, and the amount they were permitted to buy in a coffee shop was limited to 5 g per person. Drug-dealing houses offered larger quantities and other types of narcotic drugs as well. In 2008, the number was estimated at 150 in the district of South Limburg alone (Fijnaut and De Ruyver 2008, p. 149).

Finally, suppliers operating in border areas periodically ship quantities of drugs, known as "kilo shipments," directly to dealers in Germany and France. A foreign dealer may have a weekly order of two kilos of marihuana, 1,000 XTC pills, 100 g of cocaine, and 50 g of heroin, depending on his customer base. The Dutch dealer will contact his suppliers of different types of illicit drugs to fill the order. Next, a courier will drive to the Netherlands to pick up the drugs. Because the risk of apprehension is virtually nonexistent, such deliveries have developed into an almost on-demand service (Spapens 2008).

Although it seems that the major entrepreneurs within the Dutch networks still control the import and export of wholesale quantities of narcotic drugs, there are indications that members from the "second shells" are now also organizing kilo shipments on their own. We do not have a clear picture of how this works, because tackling small-scale shipments – by Dutch standards – is not at all a priority with law enforcement officials, given the fact that their in-tray already overflows with cases involving hundreds or even thousands of kilos of drugs.

Generally, the authorities have also been slow to respond to these developments, mainly because "soft drugs" were long associated with "love and peace." Law enforcement agencies also found it difficult to crack down on cannabis cultivation because the sale in coffee shops was not prosecuted. Furthermore, painstaking criminal investigations resulted in relatively short prison sentences that hardly affected the broader network. Finally, there is less pressure from abroad than in the case of ecstasy production in the 1990s. Starting in 2004, however, some changes began to take place. To begin with, if a small-scale nursery was found in a private home, the dweller would be prosecuted, but also taxed for the extra income, and if he rented the house, he would also run the risk of eviction. Second, the police and the public prosecution service set up task forces to combat cannabis cultivation and the networks involved, aimed specifically at making it more difficult for people to grow cannabis. Third, "Joint Hit Teams," consisting of Dutch, Belgian, and French police officers, were established to patrol the main motorways used by drug tourists and kilo couriers and to take action against drug-dealing houses. Finally, in 2012, foreign customers were banned from coffee shops in the south of the Netherlands and plans are to extend this ban to the rest of the country as of January 1, 2013. However, there has also been fierce criticism of this measure, particularly because Dutch customers also need to register, which most of them refuse, and whether the government will go through with it is an open question.

Conclusion

According to the routine activities theory, crime requires motivated offenders, suitable targets or opportunities, and ineffective responses by the authorities (Cohen and Felson 1979). To begin with, we can conclude that in the Netherlands, there has been no lack of the first. Dutch networks nowadays consist of individuals from a wide range of backgrounds and nationalities. Second, Dutch networks have been able to exploit opportunities in the narcotic drug market, most notably those arising from the Dutch government's unusual approach towards soft drugs in the 1970s and the open borders of the Schengen Agreement in the 1990s. The extensive trade network that has since developed is the key to the specific characteristics of the Dutch networks. Finally, the authorities did indeed respond slowly to international requirements for criminalizing amphetamine and MDMA, and we can also conclude that it took the police and the public prosecution service a relatively long time to react to the development of the Dutch networks in the 1970s and 1980s, and to subsequent drug production in the 1990s. When law enforcement agencies did finally step up measures against ecstasy production, they proved beyond a doubt that Dutch networks are not invincible. In the past decade, however, these networks have broadened and diversified, making it more difficult to tackle them solely by means of law enforcement action. Measures that would dramatically reduce market opportunities, such as regulation of the sale and production of hashish and marihuana in other EU Member States, however, are politically unfeasible in the short and medium term.

Related Entries

- ▶ Careers in Organized Crime
- ▶ Drug Enforcement
- ▶ Drug Trafficking
- ▶ Gangs and Social Networks
- ▶ Network Analysis in Criminology
- ▶ Organized Crime, Types of
- ▶ Routine Activities Approach
- ▶ Social Network Analysis of Organized Criminal Groups
- ▶ Triads and Tongs

Recommended Reading and References

Adriaensen L (2007) Staatsvormend geweld. Overleven aan de Frontlinies in de Meierij van Den Bosch. Stichting Zuidelijk Historisch Contact, Tilburg

Boekhout van Solinge T (2008) De bestrijding van de cannabisteelt in Nederland in historisch perspectief.' In: Decorte T (ed) Cannabisteelt in de lage landen. Perspectieven op de cannabismarkt in België en Nederland. Acco, Leuven/Voorburg, pp 141–159

Bovenkerk F, Hogewind W (2003) Hennepteelt in Nederland: het probleem van de criminaliteit en haar bestrijding. Willem Pompe Instituut voor Strafrechtswetenschappen, Utrecht

Cohen L, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44:588–605

de Swaan A (1989) Zorg en de staat. Bert Bakker, Amsterdam

Egmond F (1994) Op het verkeerde pad. Georganiseerde misdaad in de Noordelijke Nederlanden 1650–1800. Uitgeverij Bert Bakker, Amsterdam

Fijnaut C (2013) The history of organized crime: a contribution to a more balanced approach. In: Paoli L (ed) The Oxford handbook of organized crime. Oxford University Press, Oxford,. (forthcoming)

Fijnaut C, de Ruyver B (2008) Voor een gezamenlijke beheersing van de drugsgerelateerde criminaliteit in de Euregio Maas-Rijn. Euregio Maas-Rijn, Maastricht

Fijnaut C, Bovenkerk F, Bruinsma G, van de Bunt H (1996) Eindrapport georganiseerde criminaliteit in Nederland. Sdu Uitgevers, The Hague

Grootswagers K (1983) De historie van het Loonse kerkdorp Kaatsheuvel. Uitgeverij Brabantia Nostra, Breda

Kleemans E, van den Berg E, van de Bunt H (1998) Georganiseerde criminaliteit in Nederland. Wetenschappelijk Onderzoek- en Documentatiecentrum, Ministerie van Justitie, The Hague

Kleemans E, Brienen M, van de Bunt H (2002) Georganiseerde criminaliteit in Nederland, Tweede rapportage op basis van de WODC-monitor. Wetenschappelijk Onderzoek- en Documentatiecentrum, Ministerie van Justitie, The Hague

Klerks P (2000) Groot in de hasj, Theorie en praktijk van de georganiseerde criminaliteit. Samson/Kluwer Rechtswetenschappen, Antwerpen

Korps Landelijke Politiediensten (2005) Criminaliteitsbeeldanalyse 2002–2004. Zoetermeer

Leuw E (1994) Initial construction and development of the official Dutch drug policy. In: Leuw E, Haen Marshall I (eds) Between prohibition and legalization. The Dutch experiment in drug policy. Kugler Publications, Amsterdam/New York, pp 23–40

Middelburg B (2001) De Godmother. LJ Veen, Amsterdam

Sietsma K (1985) Rechercheren naar de recherche. In: Fijnaut C, Kuijvenhoven A (eds) De recherche onder de loupe, inleidingen voor het recherchecongres op 14 en 15 oktober 1985 te Zutphen. J.B. van den Brink, Lochem

Spapens T (2006) Interactie tussen criminaliteit en opsporing. Intersentia, Antwerpen/Oxford

Spapens T (2008) Georganiseerde misdaad en strafrechtelijke samenwerking in de Nederlandse grensgebieden. Antwerpen/Oxford, Intersentia

Spapens T (2010) Macro networks, collectives and business processes: an integrated approach to organized crime. Eur J Crime Crim Law Crim Justice 18:185–215

Spapens T (2011) Interaction between criminal groups and law enforcement: the case of ecstasy in The Netherlands. Glob Crime 12:19–40

Spapens T (2012) Netwerken op niveau. Inaugural address Tilburg University, Tilburg

Spapens T, Bunt van de H (forthcoming) Soft drugs regulation and market dynamics in the Netherlands. European Journal on Criminal Policy and Research.

Spapens T, van de Bunt H, Rastovac L (2007) De wereld achter de wietteelt. Boom Juridische uitgevers, The Hague

van Duyne P (1995) Het spook en de dreiging van de georganiseerde misdaad. Sdu Uitgeverij, The Hague

van Duyne P, Kouwenberg R, Romeijn G (1990) Misdaadondernemingen: ondernemende misdadigers in Nederland. Gouda Quint, Arnhem

Weijenburg R (1996) Drugs en drugsbestrijding in Nederland. Vuga Uitgeverij BV, The Hague

Wubben H (1986) Chineezen en ander Aziatisch ongedierte, Lotgevallen van Chinese immigranten in Nederland, 1911–1940. De Walburg Pers, Zutphen

Dynamical Simulation in Criminology

Charlotte Gerritsen[1] and Michel C. A. Klein[2]
[1]Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, The Netherlands
[2]Department of Computer Science, VU University Amsterdam, Amsterdam, The Netherlands

Overview

Simulation techniques from the area of Artificial Intelligence can be beneficial for the field of Criminology because they can be used to gain more insights in criminological phenomena (that were not clear based on just the informal theory), without actually having to experiment with these phenomena in the real world. The main goal of this article is to explain the process of modelling and simulation step by step and to illustrate the benefits of this approach for the field of Criminology.

Introduction

Within the field of Criminology a number of standard research methods exist to study criminal behavior, e.g., victim surveys, offender surveys, social experiments, and the analysis of police data. Based on these methods multiple theories have been developed that provide insight into delinquent behavior. However, these theories are usually "informal," meaning that they are written in natural language or described graphically and, thus, in principle ambiguous. In contrast, researchers from the areas of Computer Science and Artificial Intelligence have recently started investigating whether theories from Criminology can be translated into a formal, unambiguous, machine-readable notation, so that they can be used for *simulation* (e.g., Gerritsen 2010; Liu and Eck 2008; Malleson and Brantingham 2009). The main assumption behind that work is that simulation techniques can be beneficial, because they can be used to gain more insights in criminological phenomena (that were not clear based on just the informal theory), without actually having to experiment with these phenomena in the real world. The main goal of this article is to explain the process of modelling and simulation step by step and to illustrate the benefits of this approach for the field of Criminology.

Modelling and Simulation

The main concept involved in the area of modelling and simulation is the concept of a *model*. A model is a representation of an object, system, or idea in some form other than that of the entity itself (Shannon 1975). Such a representation describes the most important concepts and their relations. However, it is almost never possible to describe all these aspects completely and unambiguously, e.g., because the exact relation between concepts in reality is not known or because not all factors that influence a concept are known. When building such a model it is important to realize that this has some important consequences:

1. A model is a simplification, not a complete representation of the reality.
2. A model is based on several choices and assumptions of the creator of the model.
3. Characteristics of the model are not necessarily characteristics of reality.

These consequences should be taken into account when working with models, as they imply that the conclusions that are drawn based on the model are based on the assumptions that are made.

In general, the main goal of a model is usually that it enables the user to study a certain process that exists in reality in a convenient manner. There may be several possible reasons why studying a model is more convenient than studying the process itself. One possible reason is that the process of interest does not yet exist. An example of such a situation is studying the impact of the introduction of surveillance cameras in a mall on people's behavior. Another reason can

Dynamical Simulation in Criminology, Fig. 1 General methodology for modelling

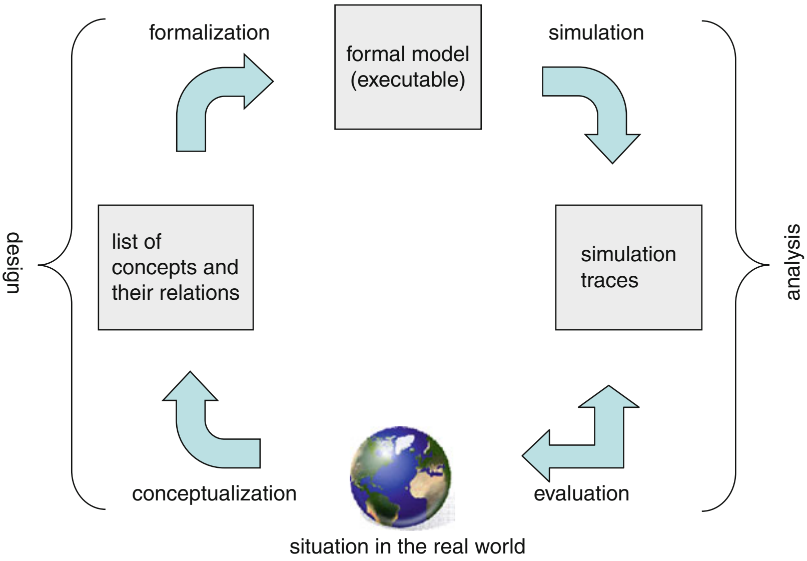
formalization
formal model (executable)
simulation
design
list of concepts and their relations
simulation traces
analysis
conceptualization
evaluation
situation in the real world

be that the process itself cannot be easily studied directly. For example, think about cognitive processes: it is not easy to measure things that go on in a human mind, while one can analyze relatively easily how a model of it behaves. In such a situation, a model may provide the researcher more insight in the process under investigation, even if it is not completely identical to the real world. One other common reason for studying models instead of the process in reality is prediction: one would like to see what happens in the future. Economical and weather models are examples of models that are often used for prediction, but also the prediction of the development of crime rates in a city is a possibility. Finally, experimenting using a model can be cheaper and more feasible than experimenting with real processes. For example, experimenting with different surveillance strategies is much more convenient in a simulated world than in the real world (Bosse and Gerritsen 2010).

In order to study a model, it is often useful if the model can imitate the dynamics of the process over time. This process is called *simulation*, and models that can be used to simulate behavior over time are called *dynamical models*. This type of model is the focus of this article. A simulation model imitates the dynamics of a process over time and thus helps to clarify the interaction between different aspects. The result of a simulation is a sequence of states of the model at subsequent time points, which is called a *trace*, *simulation trace*, or *simulation run*. The outcomes of a simulation are usually represented in a graphical form.

Modelling and Simulation Cycle

A general methodology for designing and analyzing models, the *modelling and simulation cycle*, is depicted in Fig. 1.

The modelling cycle consists of two main phases that comprise four steps. The first phase is the design phase in which the model is built. This phase consists of the conceptualization and the formalization of a process. The second phase is the phase in which the model is analyzed. This phase consists of the simulation of the process and the evaluation of the model. Below these four steps of the process are explained in more detail.

Conceptualization

Before starting the conceptualization, it is important to decide what exactly should be modelled. This means that it should be decided which process should be simulated and which questions need to be answered by using the model.

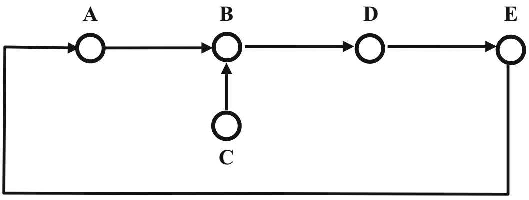

Dynamical Simulation in Criminology, Fig. 2 Graphical representation of the relationship between concepts

Then, there are two relevant aspects that need to be distinguished in the conceptualization step.

First, the relevant *concepts* need to be identified. What are the factors that play a role in a process? A concept can be an object but it can also be an event. Which concepts are important depends on the process that is modelled.

The second task is the specification of the *relationships* between the concepts that have been identified. At this moment it is specified *whether* different concepts influence each other, but not yet *how* they influence each other. This is specified in detail in the formalization step. The results of specifying the relationships between concepts can be a list of statements in the form of "A affects B," where it is possible that concepts are influenced by multiple concepts. It is also an option to represent these relationships in a graph. Each concept should be depicted as a node (circle) and the relationships as edges (arrows) between the nodes (see Fig. 2).

Formalization

In the formalization step, the concepts and relations between concepts specified in the conceptualization step are defined in more detail. The formalization depends on the representation mechanism that is used for the model. When considering existing simulation approaches, the following two classes can be identified (Bosse et al. 2007): *logic-oriented* approaches and *mathematical* approaches, usually based on difference or differential equations. Logic-oriented approaches are good for expressing qualitative relations but less suitable for working with quantitative relationships. Mathematical approaches are good for the quantitative relations, but expressing conceptual, qualitative relationships with them is hard to impossible. These representations will also be used in the case study presented in the next section. The exact form of the formalization (i.e., the formula's chosen) is a choice made by the modeller, based on the assumed form of relations between concepts. A number of standard ways of specifying relations are available to the modeller (e.g., logistic growth, weighted average).

Simulation Experiments

Depending on the representation of the model and the software tools used, it is possible to simulate the model and generate a simulation trace. To perform a simulation experiment a number of steps have to be taken. First the question or the pattern to be addressed by the experiment is formulated, for example, to investigate the amount of crime that is performed in a certain period. As a next step, choices have to be made with respect to the initial values of variables and the values of the parameters involved in the simulation model. Setting up a simulation experiment requires a structured plan on which values are chosen for these variables and parameters, thereby establishing a particular *scenario*. The settings of these values are determined by the process under investigation. If one is simulating the dynamics of crime over a city and the precise locations of the different houses are known, then this information can be used to set the parameters that represent the locations of the houses in the simulation. Alternatively, it is also possible to set parameters by using fictive information. Moreover, different scenarios can be used in order to compare the outcomes for different circumstances: changing the values in a number of different simulations. For a thorough evaluation, multiple scenarios need to be simulated.

Some models contain a stochastic element, i.e., parameters for which the values are determined by a probability. With this kind of models, running the same model several times results in possibly different traces, which all reflect the outcome of a different simulation for the same model.

Evaluation

Evaluation consists of verifying whether the model is a correct representation of the system that it represents. This can be done by formulating certain *properties* that hold in the actual situation and check whether they also hold in the model. These properties often express higher-level characteristics of the behavior of the process rather than the direct influences, for example, stable end situations (e.g., if all environmental factors are stable eventually also the amount of crimes will stabilize), or the effect of the occurrence of several events at the same time.

A distinction can be made between two types of properties:

1. Quantitative properties: statements about numerical characteristics of a model. Examples of quantitative properties are "the crime numbers never decline with more than 10 percent" or "the amount of burglaries increases with 5 percent each year."
2. Qualitative properties: statements about nonnumerical relations or characteristics of the model. Examples of qualitative properties are "eventually state property X will be true" (e.g., an offender will be arrested), "event A always occurs before event B," "parameter X (e.g., the crime rate) is always decreasing," "the simulation always reaches an equilibrium," etc.

Various computer programs exist that automatically check properties of a simulation model.

When the evaluation shows that the model does not resemble the real process well enough, there may be three possible causes, namely, (1) a modelling flaw, (2) wrongly chosen values for the parameters that describe the influence of the concepts on each other, or (3) an invalid theory that was used as a basis. In the first case, the modelling cycle has to be continued with a reconceptualization phase. In the second case, better values for parameters have to be chosen. This can be done heuristically, by trying different values that result in a more realistic simulation or by mathematical approaches that find optimal parameter values. This latter mechanism is called *parameter tuning*. To apply parameter tuning, realistic data about the simulated process are required. The tuning process then finds parameter values that, when used in the model, achieve results that are as close as possible to the empirical data. In the third case (i.e., there is a discrepancy between the theory used as a basis for the model and the phenomena observed in the real world), the model can be used as a tool to find out in more detail where the theory is incorrect (e.g., by comparing simulation results with empirical data). The result of this step is usually that the theory itself is altered or refined.

When the evaluation gives a satisfactory result, the validated model can be used as tool for studying the process that has been described. Different hypothetical scenarios can be simulated and compared to improve the understanding of the process or evaluate the effect of changes in the scenario.

Case Study: Simulation and Analysis of the Dynamics of Hot Spots

In this section the modelling and simulation cycle is applied to a concrete example: the spatiotemporal dynamics of crime, more particular the emergence of the so-called criminal *hot spots*. This is one of the main research interests within Criminology. Hot spots are places where many crimes occur (Sherman et al. 1989). After a while the criminal activities shift to another location, for example, because the police has changed its policy and increased the number of officers at the hot spot. Another reason may be that the passersby move away when a certain location gets a bad reputation. The reputation of specific locations in a city is an important factor in the spatiotemporal distribution and dynamics of crime (Herbert 1982). For example, it may be expected that the amount of assaults that take place at a certain location affects the reputation of this location. Similarly, the reputation of a location affects the attractiveness of that location for certain types of individuals. For instance, a location that is known for its high crime rates will attract police officers, whereas most citizens will be more likely to avoid it. As a result, the amount of criminal activity at such a location will

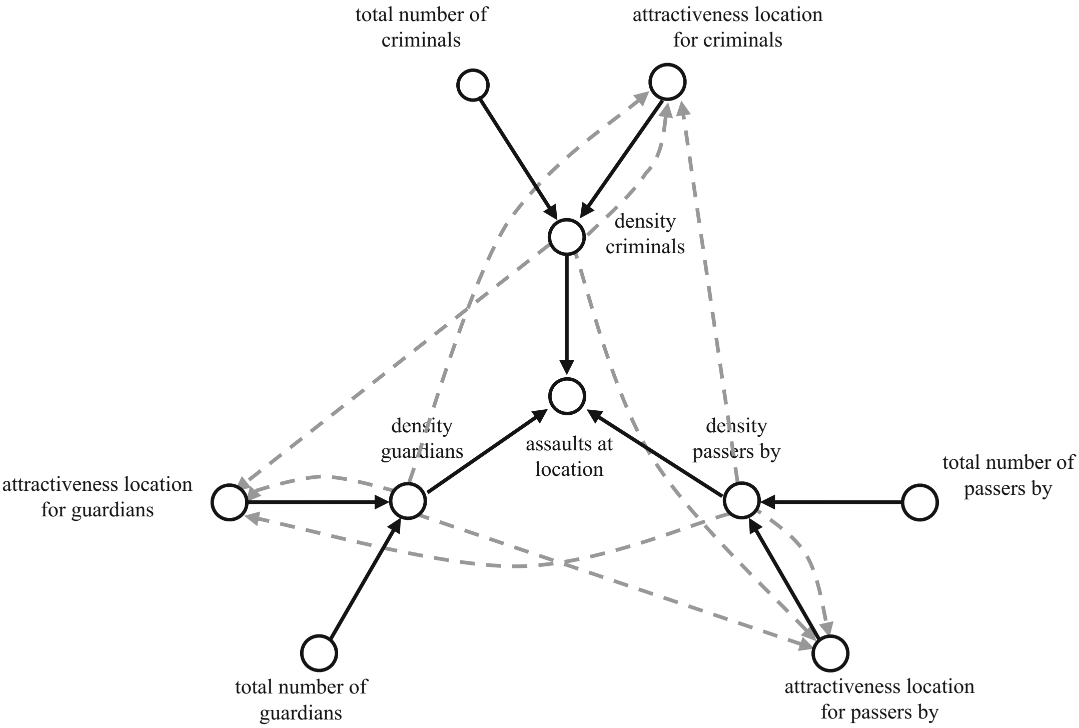
total number of criminals
attractiveness location for criminals
density criminals
assaults at location
density guardians
density passers by
attractiveness location for guardians
total number of passers by
total number of guardians
attractiveness location for passers by

Dynamical Simulation in Criminology, Fig. 3 Overview of simulation model for spatiotemporal dynamics of crime

decrease, which will affect its reputation again. Below, it is explained how the spatiotemporal dynamics of crime can be modelled by referring to the interaction of three groups, namely, *offenders*, *guardians*, and *passersby* (i.e., potential victims). This case study is inspired by the research presented in Bosse et al. (2011).

Conceptualization

In order to make a simulation model of the spatiotemporal dynamics of crime, several aspects are important. First, it is important to know the total number of the different groups involved, i.e., the *number of offenders*, *number of guardians*, and *number of passersby*. Next, it is assumed that the world (or city) that is addressed can be represented in terms of a number of different *locations*. It is important to know how many individuals (or *agents*) of each type are present at each location: the *density of offenders*, *density of guardians*, and *density of passersby*. Furthermore, to describe the movement of the different agents from one location to another, information about the *reputation* (or *attractiveness*) of the locations is used, so also this information is needed. This attractiveness is different for each type of agent. For example, passersby like locations where it is safe, e.g., locations where many guardians are present and no offenders. On the other hand, guardians are attracted by places where a lot of offenders are present, and offenders like locations where there are many passersby and no guardians. Finally, to be able to represent the idea of hot spots (locations where many crime takes place), the number of *assaults* per location is modelled. The idea is that more assaults take place at locations where there are many offenders and passersby and few guardians, cf. the *Routine Activity Theory* by (Cohen and Felson 1979).

The dynamic relationships are visualized in Fig. 3. To enhance clarity, the relationships that affect attractiveness of locations are depicted as gray dashed lines.

Dynamical Simulation in Criminology, Table 1 Mathematical formalization of concepts used for crime simulation model

Concept	Formalization
Total number of offenders	c
Total number of guardians	g
Total number of passersby	p
Density of offenders at location L at time t	c(L,t)
Density of guardians at location L at time t	g(L,t)
Density of passersby at location L at time t	p(L,t)
Attractiveness of a location L for a group x	β(L,x,t)
Attractiveness of a group y for a group x	B(x, y)
Assault rate at location L	assault_rate(L,t)

Dynamical Simulation in Criminology, Table 2 Logical formalization of concepts used for crime simulation model

Concept	Formalization
Density of a group at a location	density_of_at(G:GROUP, L:LOCATION, D:REAL)
Total number of a group	total_number_of(G:GROUP, N:REAL)
Attractiveness of a location for a group	has_attractiveness_for(L:LOCATION, G:GROUP, A:REAL)
Weight factor of attractiveness of group G2 for G1	are_attracted_by(G1:GROUP, G2:GROUP, B:REAL)
Assault rate at location L	assaults_at(L:LOCATION, A:REAL)

Formalization

In this section, as a next step, for each of the concepts introduced in the previous section, both a mathematical and a logical formalization (a logical atom) are introduced; see Tables 1 and 2. Note that these two forms of representation are presented purely as illustrations of the difference of logical and mathematical modelling, but both are formalizations of the same model.

Note that the first three concepts of Table 1 do not contain an argument t for time, since the values of these concepts remain constant over time.

Dynamical Simulation in Criminology, Table 3 Sorts used

Sort	Description	Elements
GROUP	Different groups in the society	{criminals, guardians, passers_by}
REAL	The set of real numbers	the set of real numbers
LOCATION	Different locations in the environment	{loc1, loc2, . . .}

Here, the variables A and B in the two attractiveness predicates are assumed to range over the interval [0, 1]. Note that some logical atoms make use of the so-called *sorts* (i.e., limited sets of elements that belong to a certain group, similar to types in programming languages). The specific sorts that are used in the presented model, and the elements that they contain, are shown in Table 3. Moreover, note that the attractiveness of certain groups for other groups is not shown in Fig. 3.

The dynamic relationships between these concepts are formalized below, both as mathematical formulae and as logical relationships. The first relationship determines how attractive each location is for the different groups (where it is assumed that $B11 + B12 + B13 = 1$):

SR1 (Simulation Rule 1)

at any point in time,	
if	the density of group G1 at location L is D1,
and	the density of group G2 at location L is D2,
and	the density of group G3 at location L is D3,
and	the total number of agents within group G1 is N1,
and	the total number of agents within group G2 is N2,
and	the total number of agents within group G3 is N3,
and	agents of group G1 are attracted by agents of group G1 with strength B11,
and	agents of group G1 are attracted by agents of group G2 with strength B12,
and	agents of group G1 are attracted by agents of group G3 with strength B13,
and	G1 and G2 and G3 are three distinct groups,
then	the current attractiveness of location L for group G1 is
	$B11^{*}D1/N1 + B12^{*}D2/N2 + B13^{*}D3/N3$

Mathematical formalization:

$$\beta(L,x,t) = B(x,c)\cdot c(L,t)/c + B(x,g)\cdot g(L,t)/g + B(x,p)\cdot p(L,t)/p$$

Logical formalization (note that the first line indicates the sorts to which the variables belong):

∀G1, G2, G3 : GROUP ∀L : LOCATION ∀
D1, D2, D3, N1, N2, N3, B11, B12, B13 : REAL
density_of_at(G1, L, D1) ∧ density_of_at(G2, L, D2)∧
density_of_at(G3, L, D3)∧
total_number_of(G1, N1)∧
total_number_of(G2, N2) total_number_of(G3, N3)∧
are_attracted_by(G1, G1, B11)∧
are_attracted_by(G1, G2, B12)∧
are_attracted_by(G1, G3, B13)∧
G1 ≠ G2 ∧ G1 ≠ G3 ∧ G2 ≠ G3
↠
has_attractiveness_for(L, G1, B11*D1/N1
+ B12*D2/N2 + B13*D3/N3)

This formula expresses that the attractiveness of a location is based on some kind of reputation of the location for the respective types of agents. This reputation is calculated as a linear combination of the densities of all agents. Several variants of a reputation concept can be used. The only constraint is that they are assumed to be normalized such that the total over the locations equals 1. For example, a natural parameter setting for offenders would be to have B(c,p) high since offenders need victims to assault, and to have B(c,g) low because offenders try to avoid guardians. For the guardians, B(g,c) is likely to be high since offenders attract guardians, whereas B(g,p) is rather high as well. Finally, for the passersby the B(p,c) can be taken low as passersby prefer not to meet offenders, and B(p,g) (and possibly also B(p,p)) high because guardians (and other passersby) protect the passersby.

Based on this attractiveness function, the movement of the different types of agents between the locations can be determined:

SR2 (Simulation Rule 2)

at any point in time,	
if	the density of group G1 at location L is D1,
and	the total number of agents within group G1 is N1,
and	the current attractiveness of location L for group G1 is A1
then	the density of group G1 at location L is
	D1 + η*(A1 − D1/N1)*N1*delta_t

Mathematical formalization:

$$c(L,t+\Delta t) = c(L,t) + \eta 1^{*}(\beta(L,c,t) - c(L,t)/c)^{*}c^{*}\Delta t$$
$$g(L,t+\Delta t) = g(L,t) + \eta 2^{*}(\beta(L,g,t) - g(L,t)/g)^{*}g^{*}\Delta t$$
$$p(L,t+\Delta t) = p(L,t) + \eta 3^{*}(\beta(L,p,t) - p(L,t)/p)^{*}p^{*}\Delta t$$

Logical formalization:

∀G1 : GROUP ∀ L : LOCATION ∀
A1, D1, N1 : REAL density_of_at(G1, L, D1)∧
total_number_of(G1, N1)∧
has_attractiveness_for(L, G1, A1)
↠
density_of_at(G1, L, D1 + η*(A1 − D1/N1)*N1*delta_t)

Here, the formula for offenders expresses that the density $c(L,t+\Delta t)$ of offenders at location L on $t + \Delta t$ is equal to the density of offenders at the location at time t plus a constant η (expressing the rate at which offenders move per time unit) times the movement of offenders from t to $t + \Delta t$ from and to location L multiplied by Δt. Here, the movement of offenders is calculated by determining the relative attractiveness *β(L, c, t)* of the location (compared to the other locations) for offenders. From this, the density of offenders at the location at time t divided by the total number c of offenders (which is constant) is subtracted, and the result is multiplied with c, resulting in the change of the number of offenders for this location. For the guardians and the passersby similar formulae are used.

In order to measure the assaults that take place per time unit, also different variants of formulae can be used, for example:

SR3 (Simulation Rule 3)

at any point in time,	
if	the density of offenders at location L is C,
and	the density of guardians at location L is G,
and	the density of passersby at location L is PB,
then	the amount of assaults that take place at L is the minimum of γ1*C−γ2*G and PB

Mathematical formalization:

$$assault\_rate(L,t) = min(\gamma_1 \cdot c(L,t) - \gamma_2 \cdot g(L,t), p(L,t))$$

Logical formalization:

$$\begin{aligned}&\forall \mathrm{L} : \mathrm{LOCATION}\ \forall \mathrm{C, G, PB} : \mathrm{REAL}\\&\mathrm{density\_of\_at(offenders, L, C)} \wedge\\&\mathrm{density\_of\_at(guardians, L, G)} \wedge\\&\mathrm{density\_of\_at(passers\_by, L, PB)}\\&\twoheadrightarrow\\&\mathrm{assaults\_at(L, min(gamma1^{*}C - gamma2^{*}G, PB))}\end{aligned}$$

Here, the assault rate at a location at time t is calculated as the minimum of the possible assaults that can take place and the number of passersby. Here the possible number of assaults is the capacity per time step of offenders (γ1) multiplied by the number of offenders at the location minus the capacity of guardians to avoid an assault (γ2) times the number of guardians. In theory this can become less than 0 (the guardians can have a higher capacity to stop assaults than the offenders have to commit them), therefore the maximum can be taken of 0 and the outcome described above.

Dynamical Simulation in Criminology, Table 4 Parameter settings

Simulation length	100
Locations	4
Number of offenders	800
Number of guardians	400
Number of passersby	4,000
Speed factor η (for all groups)	0.5
Δt	0.1

Dynamical Simulation in Criminology, Table 5 Initial distribution of agents

	L1	L2	L3	L4
Offenders	300	100	250	150
Guardians	50	150	125	75
Passersby	1,500	500	750	1,250

Simulation

This section presents an example simulation trace that was generated on the basis of the simulation model for the spatiotemporal dynamics of crime. The parameter settings used for the simulation are presented in Table 4. As shown there, the simulation comprises four locations (L1, L2, L3, and L4), 800 offenders, 400 guardians, and 4,000 passersby. The initial distribution of these agents over the four locations is shown in Table 5.

For the attractiveness function (SR1), the following formulae have been used:

$$\begin{aligned}&\beta(L,c,t) = p(L,t)/p \text{ for offenders}\\&\beta(L,g,t) = c(L,t)/c \text{ for guardians}\\&\beta(L,p,t) = g(L,t)/g \text{ for passers} - \text{by}\end{aligned}$$

Note that this corresponds to the following settings for the different B(x,y):

$$\begin{aligned}B(c,c) = 0, B(c,g) = 0, B(c,p) = 1\\B(g,c) = 1, B(g,g) = 0, B(g,p) = 0\\B(p,c) = 0, B(p,g) = 1, B(p,p) = 0\end{aligned}$$

Obviously, more complex attractiveness functions can be used as well.

A simulation trace for this situation is depicted in Fig. 4. Note that this simulation has been performed in the MATLAB simulation environment. The first three graphs depict the movement of, respectively, offenders, guardians, and passersby over the different locations. The last graph depicts the amount of assaults performed.

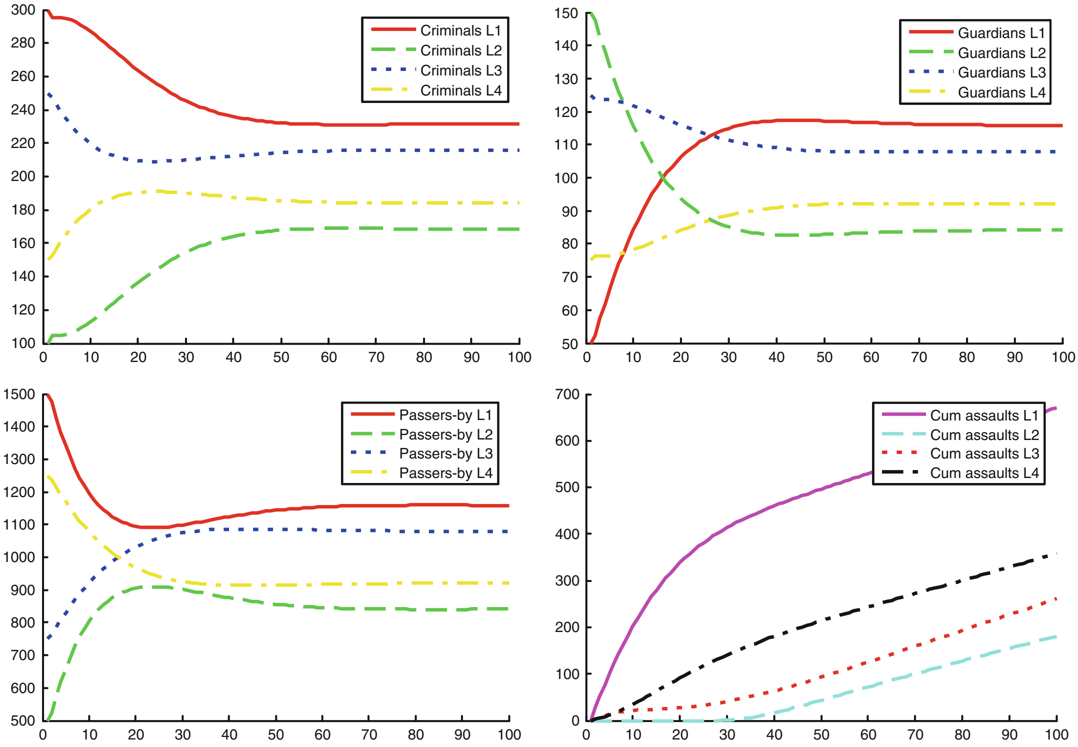

Dynamical Simulation in Criminology, Fig. 4 Example simulation trace for crime simulation model

As shown in Fig. 4, from the beginning of the simulation many passersby move away from location 1 (where there are many offenders and few guardians) and towards location 2 (where there are many guardians and few offenders). The guardians follow the opposite pattern: they move away from location 2 and towards location 1. As soon as the number of guardians at location 1 has increased, this location becomes more attractive for the passersby. The offenders first move away from location 1, towards location 2, but as soon as the passersby come back to location 1, a significant part of the offenders stays there. Eventually, all populations stabilize.

Although this is just one example simulation trace, it illustrates the power of simulation to study criminological processes. The way the simulation model has been set up enables the researcher to full in any desired combination of parameter settings, thereby reproducing (an artificial instance of) a particular real-world situation of interest. For example, one can manipulate the number of locations, the ratio between the different types of agents, or the attractiveness functions. By running simulations for these different settings, one can gain more insight in how the corresponding scenarios might develop in reality.

Evaluation

As shown, the model presented in the previous sections is capable of producing very interesting patterns. For example, eventually the number of offenders, guardians, and passersby on each location stabilizes. Nevertheless, as yet no guarantee has been given that these patterns are realistic. In other words, the model has not been validated (with empirical data). In order to fully validate a model, an obvious approach is to compare the output of the model with empirical data. For the presented case study, ideally this would mean that one would compare for each time point of the simulation whether the numbers of offenders, assaults, and so on would match the numbers if the process would take place in reality. However, for obvious reasons, such a detailed comparison is not always feasible. Moreover, it is even not always necessary, since the results of a simulation model that is not fully validated can still provide interesting insights to the researcher (as long as one is confident that the basic mechanisms of the model are plausible). For this reason, sometimes a more modest form of validation can be performed. For example, it is possible to specify higher-level requirements about the behavior of the model. These requirements might result from knowledge by domain experts. For example, if it is known from empirical data that hot spots never move if the characteristics of the locations stay the same, the following property might be formulated. This property expresses, in a semiformal notation, that after a certain time point t, the densities of all three groups stay around a certain equilibrium:

GP1 (Global Property 1): Stable Number of Groups at Locations

For all traces, there is a time point t such that	
for all time points t1 and t2 after t, for all groups g and locations l,	
if	the total amount of agents of group g is n,
and	at t1 the density of group g at location l is x,
and	at t2 the density of group g at location l is $x2$,
then	the difference between x and $x2$ is smaller than α % of n.

Using an automated tool for checking properties of simulation runs (Bosse et al. 2009b), property GP1 has been checked against the trace shown in Fig. 4. It was found, among others, that for an α of 1.0 (i.e., 1 %), stabilization of offenders occurs at time point 35, stabilization of guardians occurs at time point 28, and stabilization of passersby occurs at time point 38. Assuming that one would have similar information about a corresponding real-world process (e.g., one would know for a given scenario how long it takes until each group stabilizes), then a comparison of these numbers can be used to confirm that the model shows plausible behavior.

Also other types of properties can be specified. Some examples are the following (only in informal notation):

- "for each group, the attractiveness of each location is always between 0 and 1" (checking this property can be useful for the modeller to check whether the model is free of errors that make the variables go out of their ranges).
- "between time points t1 and t2, the density of group g at location l is monotonically increasing/decreasing."
- "for all traces $\gamma 1$ and $\gamma 2$, if there are more locations in $\gamma 1$ than in $\gamma 2$, then at the end of the simulation there will be less crime in $\gamma 1$."

Thus, checking certain properties allows the modeller to investigate whether the model behaves according to the expectations. Note that it is part of an iterative process: when the model does not behave according to the expectations, you return to the first step of the modelling and simulation cycle (conceptualization), in order to improve and refine its basic mechanisms.

Discussion

Within the area of Criminology, analysis of complex social and environmental processes, such as the spatiotemporal dynamics of crime, is an important challenge. For instance, criminologists are interested in the question where criminal hot spot may emerge and when they will emerge. Answering such questions is not easy since the large number of factors involved can make the process quite complicated. Moreover, it is often not feasible to perform experiments with these processes in the real world. The use of computer simulation can be a solution to this problem. Computer simulation may help to investigate complex (spatiotemporal) processes within criminology in a relatively fast and cheap manner.

In this article, the different methodological steps involved in modelling and simulation have been explained, and the methodology has been illustrated for a particular case study. Note that the main goal of the case study is to help researchers in their theory building, to shed more light on the process under investigation. As such, the presented model can be used as an analytical tool to see how certain aspects influence the spatiotemporal dynamics of crime. This tool can for example be used to investigate a fictive process, e.g., how do passersby, guardians, and offenders react to each other when their ratio is 4:1:2? And what happens when the number of guardians is increased?

As soon as the model can be shown to be sufficiently realistic (i.e., the global patterns produced by the model match reality), it may also become of interest for policy makers, e.g., to study certain "what-if" scenarios. For example, one may investigate how the crime level of a certain city will change if the policy makers invest in more surveillance in a certain area (Bosse et al. 2010b). Other interesting insights from simulation models in the criminological domain are the finding that anticipating strategies for police investments seem to be more effective than reactive strategies (Bosse et al. 2010b); the finding that school and parents have a relatively large influence on delinquent behavior among juveniles (Bosse et al. 2009a); and that for a reasonable amount of guardians, hot spot policing seems to be more effective on the crime rates than area hot spot policing (Bosse et al. 2010).

When one considers the level of detail of similar tools that have been proposed in the literature (such as Baal 2004; Brantingham et al. 2005; Groff 2005; Liu et al. 2005; Melo et al. 2005; Reis et al. 2006), it turns out that different perspectives are taken. For example, some authors have attempted to develop simulation models of crime displacement in existing cities, which can be directly related to real-world data (e.g., Liu et al. 2005), whereas others deliberately abstract from empirical information (e.g., Bosse et al. 2010). The point of view taken in the current article can be situated in between these two extremes. Initially, the simulation model was developed to study the spatiotemporal dynamics of crime per se. However, its basic concepts have been defined in such a way that they can be directly connected to empirical information, if this becomes available. As a result, an interesting challenge for future work will be to explore the possibilities to connect the basic concepts of the model to real-world data.

Related Entries

- ▶ Agent-Based Modeling for Understanding Patterns of Crime
- ▶ Agent-Based Models to Predict Crime at Places
- ▶ Simulation as a Tool for Police Planning

Recommended Reading and References

Baal PHM van (2004) Computer simulations of criminal deterrence: from public policy to local interaction to individual behaviour. Ph.D. Thesis, Erasmus University Rotterdam, Boom Juridische Uitgevers

Bosse T, Gerritsen C (2010) An agent-based framework to support crime prevention. In: Hoek W van der, Kaminka GA, Lespérance Y, Luck M, Sen S (eds) Proceedings of the ninth international conference on autonomous agents and multi-agent systems, AAMAS'10. ACM Press, pp 525–532

Bosse T, Jonker CM, van der Meij L, Treur J (2007) A language and environment for analysis of dynamics by simulation. Int J Artif Intell Tools 16(3):435–464

Bosse T, Gerritsen C, Klein MCA (2009a) Agent-based simulation of social learning in criminology. In: Sharp B (ed) Proceedings of the international conference on agents and artificial intelligence, ICAART'09. INSTICC Press, pp 5–13

Bosse T, Jonker CM, van der Meij L, Sharpanskykh A, Treur J (2009b) Specification and verification of dynamics in agent models. Int J Coop Inf Syst 18:167–193

Bosse T, Elffers H, Gerritsen C (2010) Simulating the dynamical interaction of offenders, targets and guardians. Crime Patterns Anal 3(1):51–66

Bosse T, Gerritsen C, Hoogendoorn M, Jaffry SW, Treur J (2011) Agent-based versus population-based simulation of displacement of crime: a comparative study. Web intelligence and agent systems: an international journal, IOS Press, to appear

Brantingham PL, Glässer U, Singh K, Vajihollahi M (2005) Mastermind: modeling and simulation of criminal activity in urban environments. Technical report SFU-CMPTTR-2005–01. Simon Fraser University

Cohen LE, Felson M (1979) Social change and crime rate trends; a routine activity approach. Am Sociol Rev 44:588–608

Gerritsen C (2010) Caught in the act: investigating crime by agent-based simulation. Ph.D. Thesis. VU University Amsterdam

Groff ER (2005) The geography of juvenile crime place trajectories. Ph.D. Thesis. University of Maryland

Herbert DT (1982) The geography of urban crime. Longman, Harlow

Liu L, Eck J (eds) (2008) Artificial crime analysis systems: using computer simulations and geographic information systems. Hershey, PA

Liu L, Wang X, Eck J, Liang J (2005) Simulating crime events and crime patterns in RA/CA model. In: Wang F (ed) Geographic information systems and crime analysis. Idea Group, Singapore, pp 197–213

Malleson N, Brantingham P (2009) Prototype burglary simulations for crime reduction and forecasting. J Int Symp Environ Criminol Crime Anal 2(1):47–66

Melo A, Belchior M, Furtado V (2005) Analyzing police patrol routes by simulating the physical reorganisation of agents. In: Sichman JS, Antunes L (eds) Multi-agent-based simulation VI, proceedings of the sixth international workshop on multi-agent-based simulation, MABS'05. Lecture notes in artificial intelligence, vol 3891. Springer Verlag, 2006, pp 99–114

Reis D, Melo A, Coelho ALV, Furtado V (2006) Towards optimal police patrol routes with genetic algorithms. In: Mehrotra S, et al. (eds) ISI 2006. LNCS 3975, pp 485–491

Shannon RF (1975) Systems simulation. Prentice-Hall

Sherman LW, Gartin PR, Buerger ME (1989) Hot spots of predatory crime: routine activities and the criminology of place. Criminology 27:27–55

Dyssocial Personality Disorder

► Psychopathy and Offending

Earlscourt Child and Family Centre (ECFC)

▶ Stop Now and Plan (SNAP®) Model

Earlscourt Girls Connection (EGC)

▶ Stop Now and Plan (SNAP®) Model

Early Chicago School

▶ History of Geographic Criminology Part I: Nineteenth Century

Early Chicago School Theory

Christopher R. Browning[1], Kathleen A. Cagney[2] and Katherine Morris[3]
[1]Department of Sociology, The Ohio State University, Columbus, OH, USA
[2]Departments of Sociology and Health Studies, The University of Chicago, Chicago, IL, USA
[3]Harvard University, Cambridge, MA, USA

Overview

The Early Chicago School of Sociology was a formative influence on subsequent criminological thought throughout the twentieth century and remains central to current debates. The concept of neighborhood *social disorganization* is perhaps the most enduring intellectual legacy of the Early Chicago School. In its most prevalent contemporary definition, social disorganization can be understood as the inability of a community to realize common values and maintain effective social controls. The origins of the concept can be traced to the foundational work of the French sociologist Emile Durkheim, particularly his seminal study *Suicide*. Drawing on Durkheimian themes, Thomas and Znaniecki were the first Chicago School researchers to articulate and apply the notion of social disorganization in their study of immigrants adapting to the rapidly changing early twentieth-century urban context. Subsequently, Park and Burgess pioneered the human ecology perspective contributing a spatial perspective on urban processes that lead some areas to face higher risk of social disorganization. Drawing on these influences and the work of Thrasher on urban gang formation, Shaw and McKay refined the social disorganization approach for the purposes of understanding neighborhood-level variations in crime rates. Specifically, they argued that neighborhood poverty, residential instability, and racial/ethnic heterogeneity compromised the ability of residents to realize shared values, including the control of crime in public space. Social disorganization theory has proved a remarkably resilient theoretical perspective within criminology and remains among the most influential explanations of variation in crime across urban space.

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

Introduction

The University of Chicago founded the first department of sociology in the United States in 1892, a time of increasing industrialization, domestic and international migration, and rapid urbanization. Under the leadership of its Chair, Albion W. Small, the department emphasized the necessity of merging social theory with empirical data collection and analysis in an effort to establish the *science* of sociology. The "Chicago School of Sociology" would later refer to the theories and methods practiced by faculty and students in the department during the first half of the twentieth century. The Chicago School is often conceptually split into two connected yet distinct cohorts – the "Early Chicago School,"1894–1940 during the Progressive Era, and the "Second Chicago School," 1945–1960 during the Postwar Period. The Second Chicago School, including graduate students Howard Becker and Erving Goffman, is often credited with developing symbolic interactionism from the work of Herbert Blumer and George Herbert Mead. For a more complete narrative of the Chicago School of Sociology, detailed biographies of its constituents, and discussion of the conflict over the existence of a second "Chicago School," see *Chicago Sociology: 1920–1932* by Robert E. Faris (1967); *The Chicago School of Sociology* by Martin Blumer (1984); and *A Second Chicago School?*: *The Development of a Postwar American Sociology* by Gary Alan Fine (1995).

Exemplified in the work of faculty members Robert E. Park and Ernest W. Burgess, the Early Chicago School (ECS) saw the surrounding urban environment as a natural social laboratory in which to investigate human behavior, including juvenile delinquency and adult crime. Park concludes his essay, "Human Behavior in the City Environment," with the following statement, "The city, in short, shows the good and evil in human nature in excess. It is this fact, perhaps, more than any other which justifies the view that makes the city a laboratory or clinic in which human nature and social processes may be most conveniently and profitably studied" (1915, p. 612). This perspective strongly influenced the qualitative and quantitative empirical studies conducted by faculty and students in the ECS during the first half of the century. Following the seminal work of Thomas and Znaniecki, Thrasher, Park and Burgess, and Shaw and McKay, as well as the ethnographies of Frederick Thrasher (1927), Harvey W. Zorbaugh (1929), and Walter C. Reckless (1933), among countless others, the ECS became synonymous with urban sociology and research on the effects of the neighborhood context for individual behaviors and outcomes. The ECS's research on urban social context would furthermore shape social scientists' approach to urban social problems and provide important theoretical groundwork for contemporary criminology.

This entry traces the origins and development of the key criminological theory to emerge from the ECS – social disorganization theory. Beginning with the foundational theoretical insights of Durkheim, the entry focuses on the central ECS figures that contributed significantly to the development of the social disorganization model, including Thomas and Znaniecki, Thrasher, and Park and Burgess. The work of Shaw and McKay and the realization of social disorganization theory in their research on neighborhoods and crime is then reviewed. Finally, the entry considers the implications of the ECS emphasis on the spatially situated nature of social phenomena for future directions in criminological research. Although Edwin Sutherland was a student and faculty member at the University of Chicago during the period addressed here, his seminal work on differential association theory began to take shape somewhat later than the period emphasized here, appearing fully formed in Sutherland's 1947 edition of *Principles of Criminology*. A more extensive discussion of Sutherland's important contribution to criminological theory can be found in the entry on ▶ Differential Association Theory.

The Roots of Social Disorganization Theory

The roots of social disorganization theory can be traced to Emile Durkheim, particularly his seminal work *Suicide* (1997). In this foundational contribution to the sociological discipline, Durkheim takes as his analytic focus variability in *rates* of suicide across communities, viewing these as "social facts" worthy of explanation in their own right. Suicide rates thus required explanation with reference to social phenomena, not the aggregation of individual psychological states. In Durkheim's view, periods of rapid social change – characteristic of modernity – attenuate the strength of the normative order structuring life expectations. For instance, rapid economic change – both declining *and* expanding economic fortunes – destabilizes expectations for life prospects, generating uncertainty and a state of what Durkheim calls "anomie," or normlessness. Durkheim's provocative and counterintuitive claim that rapid increases in wealth lead to increases in suicide rates stems from his conception of human nature. Fundamentally, Durkheim viewed human beings as naturally self-interested and capable of infinite aspirations in the absence of constraining social norms. Normatively destabilizing increases in wealth thus free the fertile human imagination to aspire in an unlimited way, inevitably leading to disappointment.

Although Chicago School research departed in significant ways from Durkheim, a number of important assumptions of the Chicago School approach can be linked to Durkheim's thinking. First, Durkheim's emphasis on the link between characteristics of social structures (e.g., economic change) and their implications for the strength of behaviorally consequential norms was a critical aspect of the Early Chicago School tradition. Relatedly, Durkheim relied on a self-interested and fundamentally Hobbesian view of human nature in describing the undesirable implications of normlessness for the functioning of societies. This image of the naturally self-interested human actor – inevitably seeking the most expedient method of satisfying basic appetites in the absence of social constraints – is later adopted as a key assumption of Early Chicago School approaches. Most influentially, these themes found expression in the work of Thomas and Znaniecki, Thrasher, and ultimately – drawing on the ecological and spatial insights of Park and Burgess – Shaw and McKay.

Early Chicago School Urban Research

Even before the work of Park and Burgess' research reached prominence, William I. Thomas and Florian Znaniecki's (1918–1928) five-volume ethnological study, *The Polish Peasant in Europe and America*, had a lasting influence on the methodology and theoretical framework of the department and the ECS. The Social Science Research Council named *The Polish Peasant* "the most significant piece of research in American sociology" in 1937 (Faris 1967, p. 17).

The Polish Peasant draws from Thomas' interests in social psychology and provides an in-depth account of the assimilation processes of immigrants arriving in urban ethnic communities from rural regions in Eastern Europe (Young 1962). Thomas and Znaniecki assert that the higher rates of delinquency in immigrant neighborhoods are not the result of personal deficiency or inherent criminal tendency. Rather, these outcomes are a product of the inevitable weakening of immigrant structures of social organization that occur when residents are transplanted from (largely rural) origin country milieus to urban environments. Thomas and Znaniecki's model of social disintegration within ethnic neighborhoods suggests that, in the face of rapid population growth and increasing racial, ethnic, and cultural heterogeneity, contradictory norms and behavioral expectations cause internal conflict among new immigrants and subsequently weaken the "control of the old social milieu" (1996 [1918–1920], p. 100). The decreasing effectiveness of traditional methods of social control – understood as positive and negative sanctions for conformity or deviation from rules and expectations for behavior – leads to increased rates of crime, domestic violence, and marital dissolution. Thus, in their view, social disorganization was a process characterized by attenuated normative structures precipitated by the interaction of heterogeneous cultural contexts and rapid change.

Thrasher's Theory of Gang Formation

Any discussion of the emergence of social disorganization would be incomplete without reference to the early work of Frederick Thrasher on gangs. An empirical investigation of 1,313 gangs, Thrasher's *The Gang* (1927) occupies a central place among influential ECS works informing the development of social disorganization theory. Thrasher's model was an early version of a social control theory, focused directly on the social and community conditions that led to gang formation.

In Thrasher's view, gangs are spontaneous, unstructured collectivities of adolescents emerging primarily in milieus characterized by ineffective family and institutional organization.

Although gangs can form in a variety of settings, in the urban context, Thrasher suggests that slums characterized by deterioration, high levels of mobility, and critically, weakened controls will tend to produce the highest concentrations of gangs. Gangs are not necessarily delinquent in this approach, as their primary function is social. Indeed, Thrasher's study offers a model for the emergence of gangs that, ironically, shifts attention away from the gang as a causal mechanism in the production of crime. Youth who participate in gangs are essentially *pre-selected* for such activity by weakened social controls – the principal explanation for their delinquent activity as well (which would likely occur in the absence of gang membership). Thus, the assumption of a natural tendency toward deviance emerges in Thrasher's work, articulated with reference to Thomas' "four wishes" – for fellowship, status, excitement, and security. In the absence of institutional structures (rooted in family and community life) in which to pursue these wishes, gangs will emerge to provide a form of inchoate organization. Although some gangs will develop more elaborate structure in the form of hierarchies, high levels of cohesion, and group norms, these are conventionally oriented and relatively rare. Most gangs will remain loosely organized and reliant on conflict with other groups for whatever minimal cohesion they maintain. Lacking structures for the development and enforcement of norms, gangs are likely to serve only as a social conduit of delinquent tendencies already in place due to weakened controls (Kornhauser 1978).

Human Ecology Theory

Thrasher's emphasis on weakened social controls in the social milieu of urban youth highlights his implicit interest in the role of neighborhoods in shaping gang formation. This incipient acknowledgement of the role of urban neighborhoods in structuring gang formation links Thrasher's research to the larger ECS theme of urban spatial differentiation. The foundational ECS research of Park and Burgess, however, took the spatial patterning of human behavior and urban social problems as a central theme of their important collection of essays – *The City: Suggestions for Investigation of Human Behavior in the Urban Environment* (1967 [1925]). They also take up the role of social disorganization as a process endemic to urban environments. For instance, they note that the social disorganization observed by Thomas and Znaniecki is not limited to immigrant communities but, rather, is inherent to the social processes of urban expansion, asserting that social disorganization is a normal component of neighborhood change (p. 54). Their theory of human ecology argues that residential mobility and change in land use patterns lead to social equilibrium imbalance, such that "the physical basis of social relations is altered, thereby producing social and political problems" (p. 64).

Human ecology theory is the study of the social and behavioral consequences of human interaction with the surrounding environment. Park and Burgess's human ecology model of the city is founded on the assertion that competition is the basic form of social interaction. Park drew on the Danish ecologist Warming's model of "plant communities" in which species struggle to survive in a complex and interdependent "natural economy." Such communities are characterized by a dynamic equilibrium marked by periods of both stability and change. In adapting this model to the urban environment, Park employed the metaphor of an organism to describe the "natural areas" – or neighborhoods – that emerge as a result of competition among groups with specific economic, occupational, race/ethnic, or cultural characteristics.

In *Human Communities*, Park (1952) elaborates the ecological model, emphasizing two key concepts. First, the city as a whole is depicted as a superorganism composed of natural areas linked by symbiotic relationships among its inhabitants. This insight is critical, as it highlights patterns of interdependency across urban areas, in addition to those characterizing the residents of a given neighborhood. For instance, small businesses may rely on one another in a local area, but areas may be characterized by commercial niches that create dependencies on other areas with other specializations. Second, Park argues that natural areas change as the city expands by a process of "invasion, dominance, and succession." For a discussion of the mechanisms of this process, see Park, Robert Ezra (1936a, b). Periods of relative stability may be disrupted when new residential populations or commerce moves into an area, potentially driving out previously dominant groups or businesses. Thus, Park's model, at its core, emphasizes the embeddedness and interdependence of urban residents and the areas they occupy as well as the ever-present potential for mobility and change.

These concepts were the foundation of Park and Burgess' (1967 [1925]) concentric zone model of the city, developed as an explanation of urban growth and expansion, neighborhood change, and segregation (pp. 51–54). The concentric zone model consists of five concentric circles: Zone I, Zone II, Zone III, Zone IV, and Zone V. The center circle in the model, Zone I, consists of the central business district and government offices. Zones II and III compose the "inner city," including the transition area or "area of deterioration"; this consists of manufacturing and new businesses and a surrounding residential area with the homes of workers. Zones IV and V are residential areas with apartment buildings and single-family homes in the former and suburban areas and satellite cities in the latter. Park and Burgess argue that the city expands by a process of invasion by the inner zones into the outer zones as areas become either more or less desirable for institutions and individuals. Park and Burgess found that neighborhoods in zones undergoing invasion or succession (i.e., with high residential mobility such as Zone II) also have higher rates of poverty, juvenile delinquency, adult crime, marital dissolution, and vice (p. 59). According to this model, the social equilibrium of the city is dependent on the rate of expansion in relation to changes in social organization. Park and Burgess, then, offer a sophisticated and explicitly spatial theory of the process by which urban areas develop over time.

Delinquency Areas and Social Disorganization

Park and Burgess's insights informed one of the most influential studies of juvenile delinquency ever undertaken. Clifford Shaw and Henry McKay's monumental work, *Juvenile Delinquency and Urban Areas*, (1942; 1969) combined the Durkheim-inspired and Thomas and Znaniecki-applied concept of social disorganization with Park and Burgess's theory of human ecology to generate an ecological, community-level explanation of urban variation in delinquency and crime. In this view, the differential exposure of certain social groups to settings conducive to engagement in crime accounts for variation in their rates of criminal activity.

Extending over multiple decades, this unprecedented effort to empirically assess the link between ecological factors and crime was rooted in Shaw's early insight that delinquents were not fundamentally different with respect to basic psychological or biological predispositions. For Shaw, delinquency emerged as a function of exposure to urban areas that could not effectively link youth to conventional groups and institutions. Lack of attachment to conventional sources of influence, in turn, led otherwise "normal" youth to participate in delinquent activities that would not have occurred had they resided in areas characterized by lower levels of "social disorganization."

In combination with his early neighborhood studies (*Delinquency Areas* (1929) and *Social Factors in Juvenile Delinquency* (1931)), Shaw's later collaborative work with McKay established the basis of twentieth-century ecological research on crime rates. In particular, a number of key findings characterized the Shaw and McKay corpus. First, areas characterized by high rates of delinquency were marked by comparatively low economic status (as measured by factors such as home ownership, median income, and receipt of public assistance). Although variation in economic conditions did not directly *cause* delinquency, Shaw and McKay demonstrated that the poorest areas of the city at any given moment (independent of the actual level of poverty) were characterized by higher rates of crime and other social problems.

A second key factor was residence in areas adjacent to heavy industry. These areas – Zone 2 in the Park and Burgess model – had low quality, dilapidated housing and population loss as commerce and industry occupied and compromised progressively more space within the area. Finally, a third factor related to the ethnic and racial composition of urban neighborhoods. Areas characterized by high concentrations of African American and foreign-born residents exhibited higher delinquency rates than other neighborhoods.

Among the key lasting impacts of Shaw and McKay's neighborhood research is their emphasis on the structural context of crime and its implications for understanding variation in behavior across race/ethnic groups. The structural factors Shaw and McKay identified as significant for understanding the distribution of crime across the city – poverty levels and associated population instability and composition – were thought to influence crime rates for a number of reasons. Centrally, poor neighborhoods characterized by rapid turnover in population were assumed to lack stable institutions for the socialization of youth (schools, churches), commitment of residents to improving the conditions of the neighborhood (as they are likely to attempt to move out as soon as finances allowed), and neighborly relations that would facilitate the social control of youth (e.g., informing parents if youth are witnessed misbehaving). Shaw and McKay saw Zone 2 as an area subject to chronic social disorganization as it was continually subject to invasion by new immigrant groups of differing cultural backgrounds. The resulting group heterogeneity further impeded neighborhood commitment and social ties, hampering the ability to maintain social control over youth.

A critical point that emerged from Shaw and McKay's research was the insight that higher rates of delinquency did not emerge from the predispositions – cultural, psychological, biological, or otherwise – of distinct ethnic or racial groups. As noted, Shaw and McKay engaged in research in the aftermath of a historic period of substantial immigration (around 23 million people) largely from southern, central, and eastern European countries. Shaw and McKay's empirical findings demonstrated that high delinquency areas remained so across the early decades of the twentieth century despite near total turnover in the composition of these neighborhoods (e.g., from predominantly northern European groups in the late nineteenth century to those from southern, central, and eastern Europe in the early twentieth century). As race/ethnic groups – previously occupying low economic status and high delinquency areas – moved out to more desirable zones of the city, their delinquency rates experienced corresponding declines. Since crime rates did not "follow" groups that exhibited high rates in delinquency areas, Shaw and McKay concluded that the source of higher crime could not be found in predispositions of any given group, but rather the delinquency fostering conditions of the areas in which they resided.

As the foregoing discussion suggests, the impact of Shaw and McKay's findings on the criminological discipline has been profound. Although the social disorganization perspective on crime – with which Shaw and McKay are inextricably associated – experienced something of a hiatus during the 1960s and 1970s, research on urban neighborhood context experienced a substantial resurgence after the publication of Kornhauser's *Social Sources of Delinquency*, one that continues unabated to this day. Although Shaw and McKay emphasized the process by which delinquent orientations were transmitted from older to younger neighborhood residents once social disorganization had taken root, Kornhauser dismissed this element of the model, drawing heavily on what she saw as a logically consistent "control" model at the heart of their approach. In its reduced form, the key neighborhood-level structural disadvantages identified – poverty, residential instability, and ethnic heterogeneity – lead to diminished informal social network ties among neighborhood residents, weakened institutions relevant for the socialization of youth, and attenuated conventional norms regarding appropriate behavior. In turn, attenuated institutional and network structures as well as conventional cultural orientations lead to increased delinquency and crime. Sampson's more recent theory of collective efficacy (Sampson et al. 1997) also draws heavily on Shaw and McKay's basic insights and Kornhauser's reinterpretation of the Shaw and McKay model.

Current Extensions of ECS Insights

The current relevance of Shaw and McKay and the Early Chicago School scholars that informed their work suggests the importance of revisiting their central insights. Below, some avenues for advancing the current application of social disorganization-inspired models of criminal activity are discussed, drawing on foundational themes of the ECS approach. In an influential analysis of the Chicago School's contribution to contemporary sociology, Abbott (1999) has argued that a consistent emphasis on the spatial and temporal embeddedness of social phenomena characterizes the work of the ECS. Contemporary sociology, Abbott suggests, could benefit from a return to this critical ECS theme. The insights of the ECS suggest three key directions for future neighborhood research on criminological outcomes.

First, research on neighborhood and crime might profitably move beyond the typical operationalization of "neighborhood." Park and Burgess emphasized that natural areas were, themselves, embedded in larger systems of cross-neighborhood exchange and interaction – processes that were crucial for understanding the functioning of urban "superorganisms." Although their insight is couched in the language of human ecology theory, the basic claim that neighborhoods do not operate autonomously to influence residents' outcomes is nevertheless apparent.

Neighborhoods are unlikely to contain the routine activity patterns of most urban residents, including youth. Contemporary neighborhood effects research in criminology tends to follow a research design where a neighborhood unit is specified – typically some census aggregation such as a tract or block group – and that unit is assigned to an individual to determine the characteristics of his or her residential context. Youth who reside in, for instance, a high poverty census tract – consistent with Shaw and McKay's original approach – are hypothesized to be at greater risk of engaging in delinquency and crime. This approach, however, assumes that the residential neighborhood is the sole spatial context relevant for understanding behavioral outcomes. Although some research has now moved toward more sophisticated models whereby contiguous neighborhoods are allowed to influence outcomes as well as "focal" or residential neighborhoods, even these models make assumptions about where youth spend time and what spatial exposures are likely to be most relevant. Drawing out the logic of the ECS emphasis on space and time, however, would argue for more precise information on the actual routine activity exposures of urban youth – information that is sorely lacking in contemporary research on urban neighborhood crime. Where (and how) do youth actually spend their time? To what extent do residential neighborhoods capture the spatial exposures of contemporary urban youth? Although Shaw and McKay pioneered the basic autonomous-neighborhood-influence model still in use, Park and Burgess's more nuanced discussion of space should inform thinking about contextual effects on crime.

Second, Park and Burgess's recognition of cross-neighborhood exchange and interaction points to the potential for identifying more complex ecological structures relevant for organizing urban activity patterns. For instance, the longstanding focus on neighborhood of *residence* neglects shared patterns of routine activity among youth that may or may not be organized by residential address. The locations of important day-to-day destinations such as school, work, grocery shopping, friends and relatives houses, and leisure destinations constitute the critical spatial exposures through which many contextual social processes are likely to operate. Just as some urban residents share a common residential location, they may also share a common set of routine activity locations, resulting in spatial patterning in the structure of shared activity locations. These *ecological communities* shape the likelihood of a variety of exposures for urban residents – to particular kinds of people, locations, activities, risks, and benefits. Thus, not only may urban residents spend a substantial amount of time outside their neighborhoods; they are likely to do so in systematic ways, resulting in patterned extra-residential exposures that may be relevant for understanding urban crime. Park and Burgess's ecological insights thus point us toward a more sophisticated model of contextual exposure than the currently prevailing approach.

Third, researchers are currently experiencing a critical turning point in their ability to track individual-level movements, through mechanisms such as GPS tracking, and obtain neighborhood- and municipal-level data on a grand scale. Social scientists, in collaboration with computer scientists and others, are just beginning to tap the resources of "big data" to examine the kinds of research questions motivated by ECS work. The increasing availability of these data may potentially lead to groundbreaking research on the role of context in criminal activity, research that moves far beyond the standard focus on the economic disadvantage, ethnic heterogeneity, and residential instability levels of administratively defined "neighborhood" units. In the City of Chicago, for instance, tremendous investments in data transparency are underway. This applies, crucially, to data on criminal activity, but also to data on 911 calls, 311 calls (i.e., non-emergency) public health data, transportation information, and social service availability. A critical feature of these newly available data is the fine-grained geographic resolution with which they will be reported, allowing for precise identification of urban conditions in space and time (down to the level of potholes and snowplow locations). Of course, not all of these data will be relevant to understanding patterns of criminal activity, but the increasingly rich information on all manner of phenomena will allow researchers to characterize urban spaces with unprecedented detail. Pinpointing geo-located service requests (e.g., police related) with response times, for instance, could provide new insights into the relationship of a neighborhood to its larger city and neighborhood residents to one another. The availability of data such as these may suggest new research questions and hypotheses about the role of context in crime and delinquency. These data also may allow social scientists to contribute research findings or provide more timely feedback to those who manage and maintain city services. This means that social science findings may be more readily incorporated into policies that shape cities.

Similarly, private sources of geo-referenced data provided by social media sources (e.g., Twitter, Foursquare) will provide yet another rich source of data characterizing patterns of mobility and space use. Although Park and Burgess's focus on natural areas within cities set the tone for the neighborhood studies of the next century, ECS researchers likely did not conceive of the possibilities that today's data resources allow. The capacity to characterize the internal and cross-neighborhood ecological dynamics of populations as they engage in their daily routines, using rich real-time "volunteered geographic information" provided by commercial social media sources, opens up tremendous possibilities for thinking about the spatial dynamics of crime. The timing and sequencing of social interaction, and the location and composition of group activity, may provide valuable information about how crime occurs. These data also may provide evidence of pro-social activities and organizations and a richness of resources not easily identified by conventional data sources such as social surveys or the census.

Conclusion

The ECS brought a concentrated period of development for urban sociological research and achievements of remarkable durability. The development of social disorganization theory was characterized by an adaptation of key Durkheimian theoretical innovations to the social processes of immigration and urban expansion characterizing the turn of the nineteenth-century American city. In turn, Park and Burgess's sophisticated conceptualization of the dynamics and consequences of urban growth provided an organizing spatial framework within which Shaw and McKay's pioneering application of the concept of social disorganization to crime could take shape. The now classic corpus of ECS research has motivated a vast research literature on neighborhood context and crime, successful and ongoing theoretical development of the perspective, and an increasingly ascendant role for contextual thinking within – and beyond – the discipline of criminology. For instance, research on health has drawn from these sociological perspectives to characterize the social and physical environment in which outcomes such as asthma, obesity, depression, and physical disability are manifested. Conditions once thought to follow largely from individual-level characteristics and behavior are now believed to be influenced by a host of factors at the neighborhood level.

Indeed, the application of social disorganization theory to outcomes other than crime was an innovation that Chicago School researchers also pioneered. Most notably, the work of Robert Faris and H. Warren Dunham (1939) introduced the notion that neighborhood social disorganization might be linked with social isolation and other mental health compromising conditions. Consistent with the expectations of social disorganization theory, these authors demonstrated that mental health outcomes and correlates such as schizophrenia and substance abuse were concentrated in areas characterized by factors such as ethnic conflict and residential mobility. A key legacy of ECS research on social disorganization can be seen in contemporary research focused on the contribution of neighborhood socioeconomic disadvantage as well as conditions such as social and physical disorder, social network density, and collective efficacy to health outcomes and healthcare service use (e.g., diabetes, rehospitalization) within disadvantaged neighborhoods.

Evidence of social disorganization theory's impact on health research also can be seen in the health-related interventions already in place. Current research on interventions aimed at making communities more "health enhancing" has focused on such actions as ameliorating physical hazards and creating bike- and walk-friendly environments – the foundation for these initiatives is rooted in ECS findings and contemporary elaborations which point to the importance of contextual effects for the emergence of a more general pro-social space (critical for ameliorating crime and promoting health).

The insights of the ECS continue to offer generative directions for the future of contextual research on crime. Combining the ECS legacy with emerging research on activity space and the promise of improved data resources described broadly as "big data" will likely yield fruitful advances. The ECS foundation will continue to inspire scholars who are focused on context and seek to contribute to a greater understanding of the mechanisms that link the social structural conditions of a community to crime and delinquency.

Related Entries

- ▶ Crime and the Racial Composition of Communities
- ▶ Criminology of Place
- ▶ Differential Association Theory
- ▶ Disadvantage, Disorganization and Crime
- ▶ Geography of Crime and Disorder
- ▶ History of Criminological Theories: Causes of Crime
- ▶ Informal Social Control
- ▶ Neighborhood Effects and Social Networks
- ▶ Poverty, Inequality, and Area Differences in Crime
- ▶ Race and Ethnicity in Social Disorganization Theory
- ▶ Social Capital and Collective Efficacy
- ▶ Social Network Analysis and the Measurement of Neighborhoods

Recommended Reading and References

Addams J (1895) Hull-house maps and papers: a presentation of nationalities and wages in a congested district of Chicago, together with comments and essays on problems growing out of the social conditions. Thomas Y. Crowell & Company, New York

Blumer M (1984) The Chicago school of sociology. The University of Chicago Press, Chicago

Brantingham PL, Brantingham PJ (1993) Environment, routine and situation: toward a pattern theory of crime. In: Clarke RV, Felson M (eds) Routine activity and rational choice: advances in criminological theory5, vol 5. Transaction Publishers, New York

Buchan E (1922) The delinquency of girls. Sociology. University of Chicago, Chicago

Bursik RJ (1988) Social disorganization and theories of crime and delinquency: problems and prospects. Criminology 26:519–552

Bursik RJ, Grasmick HG (1993) Neighborhoods and crime: the dimensions of effective community control. Lexington, Lanham

Durkheim É (1997 [1897]) In: Spaulding JA, Simpson G (ed) Suicide: a study in sociology. Free Press, New York

Faris RE (1967) In: Janowitz M (ed) Chicago sociology: 1920-1932. Chicago: The University of Chicago Press.

Faris Robert E, Warren Dunham H (1939) Mental disorders in urban areas: an ecological study of schizophrenia and other psychoses. University of Chicago Press, Chicago

Felson M, Cohen L (1980) Human ecology and crime: a routine activity approach. Hum Ecol 8:389–406

Fine GA (1995) A second Chicago school?: the development of a postwar American sociology. The University of Chicago Press, Chicago

Kornhauser RR (1978) Social sources of delinquency: an appraisal of analytic models. University of Chicago Press, Chicago

Kubrin CE, Weitzer R (2003) New directions in social disorganization theory. J Res Crime Delinq 40:374–402

Park RE (1915) Suggestions for the investigation of human behavior in the city environment. Am J Sociol 20:577–612

Park RE (1936a) Human ecology. Am J Sociol 42:1–15

Park RE (1936b) Succession, an ecological concept. Am Sociol Rev 1:171–179

Park RE (1939) Symbiosis and socialization: a frame of reference for the study of society. Am J Sociol 45:1–25

Park RE, Burgess EW (1967 [1925]) In: Janowitz M (ed) The city: suggestions for investigation of human behavior in the urban environment. University of Chicago Press, Chicago

Reckless WC (1933) Vice in Chicago. University of Chicago Press, Chicago

Sampson RJ (2012) Great American city: Chicago and the enduring neighborhood effect. University of Chicago Press, Chicago

Sampson RJ, Byron Groves W (1989) Community structure and crime: testing social-disorganization theory. Am J Sociol 94:774–802

Shaw CR, McKay HD (1972 [1942]) Juvenile delinquency and urban areas: a study of delinquents in relation to differential characteristics of local communities in American cities. University of Chicago Press, Chicago

Shaw CR, Zorbaugh F, McKay HD, Cottrell LS (1929) Delinquency areas. University of Chicago Press, Chicago

Snodgrass J (1976) Clifford R. Shaw and Henry D. McKay: Chicago criminologists. Br J Criminol 16:1–19

Thomas W, Florian Z (1996 [1918-1920]) The polish peasant in Europe and America: a classic work in immigration history. Zaretsky E (ed). University of Illinois, Urbana and Chicago

Thrasher F (1927) The gang: a study of 1,313 gangs in Chicago. University of Chicago Press, Chicago

Vold GB, Bernard TJ, Snipes JB (2002) Theoretical criminology, 5th edn. Oxford University Press, Oxford

Young K (1962) Contributions of William Isaac Thomas to sociology. Sociol Soc Res 47:3–24

Zorbaugh HW (1929) The gold coast and the slum: a sociological study of Chicago's near north side. University of Chicago Press, Chicago

Early Modern Police and Policing

Gerhard Sälter
Documentation and Research, Berlin Wall Memorial and UHK/BND, Philipps-Universität Marburg, Berlin, Germany

Overview

The concept of "police" in the early modern period was far removed from our present understanding. From the end of the Middle Ages as used in legislative texts, the term was used to mean the maintenance of good order, at first frequently in a local or regional context, extending later to a territory or kingdom. It entailed practices of authority notionally described as "police," associated with the concept of sovereignty and attempting to regulate the relations between individuals. In an era in which kings began to lay down laws defining social norms, this legislation extended to everything but the affairs of the court, the justice system, taxes, and the military. In practice, however, irrespective of the laws passed and contrary to their intention, social order long continued to be established and maintained at local or regional level.

Under the early modern model of social order, security tasks as well as preventive supervision and subsequent control were distributed across a large number of agencies which were responsible to varying, generally local authorities. These duties were frequently performed in addition to other occupations and sometimes voluntarily, while those who performed them were possessed of differing competences in differing locations. Moreover, social control long remained a task that, rather than being entrusted to specific institutions, was exercised on a day-to-day basis directly through the horizontal and vertical structures of society, normatively safeguarded and institutionally overseen by the courts. Policing by one element of society of another was embedded in their day-to-day social interaction. Up until and indeed even after the French Revolution – out of reach of the institutions of the state – it remained primarily a matter for the entirety of stakeholders in society irrespective of their social status and prestige.

Our present concept of the police as a supracompetent body, charged with maintaining good order and exercising a preventive task under uniform state command, was entirely absent in the early modern era. Its emergence was coupled with the development of a state monopoly on the use of force, which in turn is exercised and represented internally by the police. However, during this period in Western Europe, the precursors of today's police began to emerge in two forms. Towards the end of the Middle Ages, the first paramilitary police troops were formed, some aspects of which characterized the initial precursors of modern *gendarmerie* units. While from the late seventeenth century onwards, the nucleus was formed in Paris of the first modern police organization which was to set a pattern for the rest of Europe.

Public Security in the Premodern Period: Without Police

The State, the People, and the Making of Social Order

In learned discourse in premodern Europe, the term *police* or *Policey* described the exercise of authority in regulating day-to-day affairs. Included within this was the aspiration to achieve normative standardization and behavioral control through the medium of courts. Or to put it another way, the concept of police encompassed the formation of social order in areas not controlled by specialist administrative structures, such as the tax system and military organization. Under the Old Regime this concept meant the ongoing day-to-day practice in a local, particularly urban, context of establishing a system of order governed by laws. Thus, within the theory of monarchy, the concept embraced the dual meaning of order as on the one hand a social condition in which daily needs (food, economy, hygiene, security) could be satisfied and as on the other of the political authority capable of producing such order.

However, in view of the minimal effective reach of central state power and in deviation from monarchical theory, until well into the eighteenth century, it was not the norms promulgated by central instances of authority and the guardians of good order deployed by their local equivalents which materially regulated daily life, but the informal mechanisms by which social norms and social control were produced. Social norms reproduced themselves predominantly through day-to-day contacts and were complied with by everyone in the course of daily life. The vertical structures of social dominance played as much a part as horizontal daily contacts, for example, in urban neighborhoods or village communities, as well as relationships within corporations and church communities. All of these quotidian relationships in various ways inherently offered the opportunity for mutual supervision and the tangible imposition of sanctions (Sälter 2000).

The central instances of authority and the elites accepted these mechanisms of informal social control, provided that their own interests or specific moral concepts were unaffected. They regulated this form of social control primarily through the judicial system. The courts were set up by the ruling powers and staffed or overseen by members of social elites. This form of overseeing conformity with norms was reactive: confident as they were that the informal establishment and control of norms were in any case influenced and dominated by them, as far as formal social control was concerned, the elites restricted themselves to the role of arbitrator. They decided conflicts when these were brought to court by an injured party, and as Alfred Soman (1980) has concluded, they stabilized the established order and the mechanisms of social control in equal measure through the threat of punishment that could be imposed by the court. Only in certain fields did the courts actively seek to discover breaches of norms and impose sanctions. The reproduction of social order was thus based on autoregulation by local communities, both urban and rural, and the social dominance of the nobility, urban elites, dignitaries, ministers of religion, master craftsmen, burgesses, and heads of families, who in turn were overseen by instances of authority constituted as judicial bodies.

In the early modern era, there were various practices and discourses associated with the concept of police as an all-embracing description of public order. A variety of individual as well as collective participants joined in these discourses on the proper order of society which, to list but a few examples, were conducted at a legislative level, in petitions, through the medium of deviating conduct and disobedience, in conversation with neighbors, in publicly performed songs satirizing the agencies of authority, and in disputes before the courts. Such discourses and the associated practices were aimed at influencing the norms of social life to the advantage of those concerned and adapting the existing order to their needs. In this way they contributed to the reproduction of norms and the preservation of social order.

Public Security and Its Officials

Before the idea of sovereignty was spread through Europe with the advent of the Enlightenment and the French Revolution, a police force uniformly organized through an entire country would have been inconceivable. On the other hand the mostly locally oriented practices employed in reproducing good order on the principle of individual and collective self-help needed no such force. That which is occasionally described by scholars as a deficit in the premodern executive was, more properly, a form of defining the tasks and competences of the forces of order in a manner closely allied with society’s practices of maintaining social order.

In the states and territories of Europe in the early modern era, there were numerous coexistent forces of law and order, the demarcations between whose competences were generally blurred. Many of these offices were performed as ancillary occupations, leading to the accumulation identifiable in many places. Characteristically, too, these offices were generally responsible to local instances: a town council such as in Germany the *Stadtrat*, a rural official elected by local notables such as in England the *parish constable*, or a member of the nobility with their own jurisdiction, such as in France the *seigneur*. Accordingly, most of these upholders of good order possessed only local competence. Their activities were normatively regulated, and overseen in practice, by locally competent courts mostly composed of members of the elites.

With regard to their duties, there were three distinct groups of law enforcement officers:

- Firstly, court officials who formed the executive arm of the civil and criminal courts
- Secondly, units of the watch in cities and ndividual night watchmen in small towns and villages
- Thirdly, local officials who in towns and villages performed widely varying executive and supervisory roles in the field of *police*, such as supervising markets or preventing fires

Court Officials

The court officials – in England *court ushers* or *bailiffs*; in France *huissiers*, *archers*, or *sergents*; in Germany *Büttel*, *Gerichtsdiener*, *Schergen*, or *Gerichtsboten*; in Italy *uscieri del tribunale* or *sbirri*; and in Spain *aguaciles mayores* or *esbirros* – formed the executive arm of the court staff who administered civil and criminal justice. They served summonses, writs, and judgments; collected debts confirmed by the court; impounded goods and searched properties; applied court seals in, for example, inheritance matters; arrested suspects and convicts; pursued fugitives; guarded the homes of those arrested for debt in order to protect their property; were responsible for guarding prisoners; and served judges as an armed presence in court and on inspection tours. In some places they were responsible for publishing court judgments and official decrees, and they frequently performed other tasks associated with the exercise of municipal authority. On occasion they would patrol within the jurisdiction of the court, sometimes accompanied by a judge. In Paris the *sergents* were obliged to ensure that some of their number were present at certain locations in the city, so that citizens knew how to reach an officer when required.

There were also wide variations in the specific form taken by these offices. In France, for example, the *offices* could be purchased and the *officiers* were organized in corporations similar to the guilds, with the result that neither the crown nor the city council nor the court had much influence over how these posts were filled. And holding, as they did, a de facto heritable title to their office, the post holders were not easy to discipline. Thanks to their corporate organization, the court officials of Paris had the ability to negotiate collectively with their superiors and the crown regarding the extent of their competences and duties. Their position was further strengthened by the fact that their corporation had made substantial loans to the crown. A different situation applied in German cities and territories, where the courts could dismiss their officials if they failed to carry out their duties as instructed. Generally, the officials were responsible only to the court that employed them. The situation in England appears to have been similar.

Court officials generally received no salary with which they and their families could scrape a living. Their income, which varied from country to country and place to place, was composed of a basic wage (known in France as *gages*, being the interest on the purchase price for the office), sometimes a payment in kind (food, firewood, clothing), fees for specific duties, and a share in fines imposed. In England, they and the *watchmen* received fees for crimes reported if they could present the perpetrator before the court. It was not until the end of the eighteenth century that regular payment in money slowly took over.

Night Watches and Other Watchmen

In most towns and rural communities in Europe, the inhabitants were originally themselves responsible for public safety, to which end they appointed a night watch from among their number. However, this was frequently an unpopular service to perform, leading some citizens to send their apprentices, underage sons, or aged invalids to do duty in their place. In some cities it was common practice for citizens to engage paid replacements to act in their stead. Because of these structural weaknesses, but also because the central authorities became progressively less willing to tolerate citizens bearing arms, from the sixteenth to seventeenth centuries, in larger towns, these nightly duties were performed by semiprofessional, paid units of the watch. However, this led to conflict as these watchmen were frequently answerable not to the town elders but to the crown. For this reason for a time in Vienna two competing night watches existed side by side. In England from the thirteenth century onwards, every town was obliged to employ *watchmen*. In London that meant six in each of the city's 24 wards, as well as one unit that patrolled throughout the city.

The most important tasks of these night watchmen or watches were to warn their fellow citizens of an outbreak of fire and to prevent nocturnal intrusions and attacks. As the majority of European states became increasingly peaceful, it became less important to give timely warning of an approaching enemy. Those who manned the walls and watchtowers to fulfill this task either became redundant or were redeployed as fire watchers who also sounded a horn to mark the passing hours. In smaller towns they also played the role of jailers, with provisional lockups located in the towers. In some towns they ensured that the inns closed at the appointed time, and the houses were locked up at night. The watch did not remain awake throughout the night in every town – it is recorded that in some places, their duties ended at midnight when all was quiet.

There were also gatekeepers who controlled access to the towns and cities, checked passports, collected fees and customs duties, and ensured that no goods were illegally imported or exported. Only the larger cities could afford other guards in addition to the above who patrolled by day. These were not introduced in Paris until the eighteenth century, just as the first permanent guard post for the city watch was not established until 1720. In Vienna, by contrast, the *Stadtguardia* had a manned guard post in the seventeenth century. From 1705 several laws were enacted to improve the efficiency of London's watch, and towards the end of the eighteenth century, a guard that patrolled by day was instituted in the City of London. Many villages had watchmen, particularly in order to prevent thefts from the fields at harvest time.

All these watchmen were paid according to varying local custom partly in cash and partly in kind. In some places they received free accommodation. Most of them were dependent on earning a second income; or alternatively, a day laborer might undertake the duties of watchman as a secondary occupation. This sometimes limited their efficiency: complaints were voiced from time to time that after a full day's work, the poorly paid watchmen fell asleep while on their nightly duty. Another restriction on their efficiency, particularly in the cities, was rooted in their low social status. As members of the lower orders with correspondingly weak personal authority, they stood little chance of intervening in breaches of norms committed by members of the elites. For example, when the sons of the nobility or certain dignitaries roamed loud-voiced through the streets forcing innkeepers to open after hours, it was better for the men of the watch not to interfere. Even if they intervened in an obvious case of attempted rape and in doing so injured a member of a city elite, there was a risk that they rather than the malefactor might be hauled before the court.

From the seventeenth century onwards, the authorities became increasingly concerned about begging. Some communities employed inspectors of beggars to ensure that only authorized person, the so-called honorable poor, was in the streets to beg. They monitored the streets and squares to prevent unauthorized begging and arrest vagabonds. Beggars from elsewhere were brought before the court for sentencing and removed from the city. Given that begging was one of the normal temporary sources of income for the poorer members of society in times of seasonal or accidental unemployment, in many places these inspectors were the target of a corresponding hatred.

Other Officials

In towns and villages there were also various officials who performed duties under the heading of public order and public safety. Firstly, there were the heads of village, marketplace, and district communities, in some regions elected and in others appointed. In addition there were elected community officials such as the *parish constables* in England or *Rüger* in parts of Germany, whose tasks were focused on maintaining public order. Both formed a link between the superior administrative authority and its attempts to introduce norms and the local population, in that they were required to implement decrees and instructions on a day-to-day basis and see that these were complied with. They had often to restore harmony or mediate between the central legislation and local customs and interests. In many cases they could also fall back on armed personnel of the executive arm, though these were rarely available in large numbers.

Beyond the above there were also numerous officials charged with overseeing markets. They supervised the accuracy of the weights and measures used and monitored the quality of goods, particularly foodstuffs. While in many cities it was the guilds who performed these duties, in small towns and villages, the tasks fell to village dignitaries in addition to their other occupations.

The Military: Soldiers on Police Duty

It was characteristic of the premodern era that there was as yet no consistent demarcation of roles between the military and the police – something which developed only with the emergence of the modern constitutional state. Accordingly, the soldiers who with the creation of standing armies were in any case available were also deployed in a public security role. In England, which was late in forming a standing army, urban and rural militias known as *trained bands* composed of burgesses and landowners were used for this purpose. Such deployments were not restricted to revolts and unrest which were put down by loyal military units. In European states, particularly in the cities, soldiers were used on a daily basis to carry out the duties nowadays performed by the police: keeping order at markets and fairs, providing armed assistance in making arrests ordered by the courts, preventing and fighting fires, and acting as watchmen.

In Paris, for example, the 4,000 soldiers of the *gardes françaises* stationed in the city were used for such purposes. Especially in the suburbs, it would also seem that they were charged with general duties of ensuring good order. By the start of the eighteenth century, they in fact maintained permanent guard posts in some suburbs. They also supervised markets and fairs and were responsible for ensuring safety at the entrances to the city's three big theaters before and after performances and for keeping watch outside the buildings. In Lübeck a municipal military force was regularly called on by civil institutions to provide armed assistance, for example, in controlling traffic in the streets, providing security at major events, rounding up beggars, making arrests, making customs inspections, and preventing fires. Elsewhere regular troops were used for reasons of cost instead of a town watch. These examples seem to have been the rule, rather than the exception.

The Emergence of "Modern" Police Forces

Legislation, Centralization, and the Changing Face of Social Order

In the early modern period, some of the agencies of social control, namely, the central instances of authority in premodern Europe, slowly, and in conjunction with the formulation of absolutist theories of monarchy, developed an aspiration to impose norms on and oversee social life. They were partially supported in doing so by the social elites who however were also partially in competition with the crown and its administration. Efforts were made to standardize and remodel locally determined systems of order in a process which expressed itself from the sixteenth century onwards in an increasing number of normative interventions in the organically developed structure of the law.

This new legislation was initially aimed at documenting the state's claim to normative competence, but in the medium term it became associated with the expectation that the norms formulated in this legislation would be implemented, accepted, and adhered to. Still, these premodern states were far removed from either possessing or even aspiring to a monopoly on the use of force – this is a hallmark of the modern state. It took a certain amount of time for even the idea of a monopoly on the use of force to develop out of the concept of sovereignty formulated by Thomas Hobbes.

In the course of this development, public safety and security – formerly a matter for a large number of actors networked with one another in various ways – increasingly came within the purlieu of the central instances of power. The latter in turn tended to entrust the implementation of norms and behavioral controls to professional personnel under central management. In many cases this went hand in hand with efforts to increasingly subject local keepers of good order to central direction and central control. Thus, policing slowly ceased to be the responsibility of citizens and became a matter for professional policemen who were answerable not to the citizenry but to distant superiors in their country's capital. These processes entailed much friction and conflict and by the time of the French Revolution had yet to be completed in every European country.

Clive Emsley (1999), looking at the nineteenth and twentieth centuries, distinguishes between several types of police in Europe: civil police forces employed by the state, military police units employed by the state, and police employed by local authorities. The first precursors of such police forces and of such a police system came into existence in the premodern era. In particular the police system which had been established in stages in Paris and already set an example for other European states, with some changes survived the Revolution and following the Napoleonic wars, became a model for the whole of continental Europe.

Paramilitary Police Organizations

The Spanish and French monarchies were the first in Europe to set up paramilitary police organizations at the end of the Middle Ages. This at least in part reflected the desire to reinforce royal power and jurisdiction in competition with the rural nobility and urban elites. However, their first task was to bring the military to heel. Military units and demobilized soldiery roamed through both countries (in Spain in the aftermath of the Reconquista), living off the land by force of arms and reaching plague proportions in many areas. The new military police units were intended to intervene. In the long term, this intervention developed into an attempt to make the highways and thereby also interregional trade a little safer by instituting frequent patrols. The first task, that of controlling the military, led to these new units adopting a paramilitary form of organization and coming under the auspices of the ministries of war. Both these units had characteristics that are absent from modern police units. Each brigade was allocated judges who held special powers to administer summary justice against which there could be no appeal.

The Castilian crown founded the *Santa Hermandad* at the end of the fifteenth century. There are indications here that the self-organization of society was to some extent overridden as the newly created institution displaced the locally organized *hermandades* (brotherhoods) or *milicias populares* which had been maintained by the towns and cities to keep order in the countryside and in particular to protect pilgrims and merchants travelling the highways. They also served as protection against encroachments by the local nobility.

The *Santa Hermandad*, which had as many as 2,000 men under arms, was directed by the *Junta General* located in Toledo. Their units were spread across the rural regions under the control of the Spanish crown. Their judges, whose competence was restricted to areas outside of the towns and cities, had jurisdiction over such offenses as robbery, homicide, arson, and resistance to the power of the crown. Given that most of these were capital crimes and the judges frequently sentenced the delinquents on the spot, the justice they dispensed often took the form of severe corporal punishments or executions. These judges were feared accordingly. Later on their jurisdiction was transferred from the *Santa Hermandad* to the ordinary courts, and the brigades became purely police units which were replaced in Castile in 1835 by the *Guardia Civil*.

The *maréchaussée* founded in France in the sixteenth century combined the functions of a military police force with those of a rural police troop whose local commanders (*prévôts des maréchaux*) were also judges. The geographic competence of these *prévôts*, who were answerable to the Minister of War, extended only to the countryside and the highways, not to the towns and cities which administered their own justice. The crimes over which they had competence included offenses committed by members of the military as well as vagabondage, salt smuggling, housebreaking, highway robbery, and a variety of capital crimes and crimes of violence, not least among which were insurrection and resisting a public officer. In common with the *Santa Hermandad*, their judgments were reached after a summary hearing with no possibility of appeal. The *prévôt de l'îl* based in Paris was responsible within the city limits for desertion and some other military crimes; the brigades headed by him also watched over the periphery of the city.

By the seventeenth century the units of the *maréchaussée* were still widely spread out; however, some time later there began a process of distributing the companies and brigades evenly along the highways. From 1720 when the *maréchaussée* was reformed and its manpower increased, the expectation was that the *intendant* in every *généralité* should have a company at his disposal. In the eighteenth century the force was around 4,000 strong. The system of purchasing the positions of judges and officers was abolished in 1778. At the same time the command structure was simplified and modeled along military lines, as a result of which the previously used official titles adopted from the justice system also disappeared. At the latest from this time onwards, the brigades were also obliged to conduct daily patrols of their assigned areas.

The Habsburg states established similar police units at the end of the sixteenth century. In some territories the *Landprofosen* were also given the dual role of commander of a militarily organized police troop and judge with special powers. In most areas ruled by the Habsburgs, however, it proved impossible for this position to be maintained in the long term in face of the sometimes vehement protests of the local nobility who saw it as an encroachment on their own jurisdictional competence.

In the eighteenth century various German territories adopted these models and organized police troops along military lines, the commanders of which did not however have authority to act as judges. In Baden-Durlach, for example, in 1763, a corps of hussars was set up who, in addition to the bodies of citizens formed on an ad hoc basis and locally deployed police officers (*Hatschiere*) who had been in existence since around 1750, formed a police troop answerable to the central administrative authority. The smaller territories frequently balked at the cost of a special police troop. In some cases in Germany the *Reichskreise* set up their own troops of dragoons or hussars charged with police duties. By dint of carrying out patrols, these units were effective vis-à-vis both the local population and the regional elites and local authorities in representing and asserting the power of the state to impose order in the country.

The Origination of a Modern Police Force in Paris

A civil municipal police force originated for the first time as an independent institution in Paris at the end of the reign of Louis XIV and continued to develop in the eighteenth century. Its origination was not planned. On the contrary, it was the accidental result of an administrative reform aimed at asserting the power of the monarch. In 1667 the king created a new judicial office at the royal court in Paris under the name *lieutenance de police*, to which by the beginning of the eighteenth century, he then proceeded to transfer most of the functions involved in governing the city of Paris which had previously been exercised by a variety of instances closely associated with the city elites. The concentration of competences awarded to this office constitutes a stage in the centralization of political power that is typical of the Ancien Régime. This office became the nucleus of the first modern police organization in Europe.

Scholars have long interpreted this reform as a conscious reaction to the processes of urbanization, as well as a regulatory policy response to population growth, migration problems, and rising criminality. Leaving aside the fact there is no sustainable proof that this was the case, this was also not a reaction that would have accorded with the political mentality of the premodern period. It has been argued, as Clive Emsley (1991: 8) established with reference to England, "along lines similar to the police reformers of the nineteenth century, namely, that the system before 1800 had been in steady decline for perhaps as much as half a millennium." Corresponding accounts make a functionalistic connection between socioeconomic processes and political decisions which anachronistically transfers elements of the regulatory policy thinking of the nineteenth and twentieth centuries to the early modern period and the decision-makers of that time (Sälter 2002b).

By contrast the concentration of government competences in Paris represented the continuation of a tendency observable since the sixteenth century on the part of the crown to increasingly exclude local elites from holding power; a tendency which reached a provisional conclusion with the creation of the office of *lieutenant général de police*. The latter assumed control of the city of Paris as commissioner directly responsible to the crown. The object of this reform was not to materially change the way in which authority was exercised, nor did it include any new concept of municipal order. It was aimed exclusively at disempowering the urban elites who were in competition with the crown and who had opposed the king in the Fronde civil war and the agencies dominated by them.

Triggered by the change in institutional structure, in the decades that followed, the practices of authority as these related to social control slowly began to alter. The installation of the *lieutenant général* had prevented the urban-governing elites and prominent individuals allied with them from participating further in the process of shaping urban order. Prior to 1667 this system of order had been founded on the widest possible consensus embracing broad swathes of the population in a socially graduated model of cooperation. In return, those with a participating role in power and in the reproduction of urban order guaranteed that this system would prevail. When deprived of their power, they forfeited their previous role in overseeing social control on a day-to-day basis as well as their influence. They were no longer able to mediate between quotidian aspects of social control and those aspects that were institutionalized in the judiciary.

The *lieutenant général*, on the other hand, had neither the same access to nor communication with the life of the city as did the networks of urban elites with their numerous dependent social relationships. The centralization of state power denied the *lieutenant général* access to the discourses in which urban order had previously been shaped and reproduced. From the police perspective, this resulted in an increasingly palpable power vacuum, a restriction in the capacity of the political authorities to supervise day-to-day affairs, which the *lieutenant général* sought to fill with new forms of state control. Thus, it was the implantation of a new instance of authority in the Parisian center of power, an instance which was consciously removed from the networks of social relationships, that made it necessary to develop the offices and techniques which we nowadays refer to as police.

As a reaction to the deficiency in oversight perceived by the *lieutenant général*, from around 1700 as minor reforms were initiated on an ad hoc basis, the structure gradually developed of a police-like institution and the initial manifestations of police activities which deviated from previous practices that had been dominated by the judiciary. The nucleus of a modern police organization was established, and with this change something akin to a new concept of the reproduction of urban order slowly emerged.

The role of the courts in supervising compliance with norms had been reactive: the justice system mainly responded to a breach of norms if and when this could no longer be dealt with by the networks of social control that were firmly anchored in everyday life, leaving no option but to bring the matter before the court. The police now began themselves, albeit initially only in certain fields, to actively seek, detect, and punish infringements of the law. To this end the *lieutenant général* D'Argenson (1697–1718) from around 1700 established new posts which no longer corresponded with the functions of court officials or the city watch, even though it was from these groups that the first police officers were recruited. For the first time the duties with which the *inspecteurs de police* were tasked amounted to systematic preventive social control. Their tasks included monitoring the economy of the poor, gambling, strangers and foreigners, the illegal trade in goods on which the state had a monopoly, and tracking down clandestine protestants. Given that little information was forthcoming from the populace in respect of offenses of this kind, the police developed their own investigative arm which employed secret informers. In broad terms, one might describe this organization as a precursor of a criminal investigation force but also of a political police force.

The emancipation of the police associated with this development which distanced them from the judiciary also brought forth new methods of sanction. The police acquired a system of sanctions of their own which rendered them independent of the formalities of the criminal justice system. The *lieutenant général* having recourse to the power of the crown could impose penalties with no need for a court verdict and oversee the enforcement of these penalties for himself. As a result, penal jurisdiction was partially removed from the courts still dominated by the old elites. The police penalties also to some extent differed from court practices: in addition to expulsion from Paris and banishment to a specified place, the main punishments imposed were relatively long terms of imprisonment. Corporal punishments and public humiliation, on the other hand, were not imposed.

In the eighteenth century *Commissaires au Châtelet* were integrated into the new police system. This was an ancient office whose holders had been charged with traditional forms of policing since the Middle Ages. They had a dual function in the Parisian court system: They were a kind of examining magistrate with civil as well as criminal jurisdiction, to whom citizens and residents could bring suit, and in their role as members of the court, they also oversaw the fulfillment of traditional police duties such as street cleaning, food supplies, and Sunday observance. In the eighteenth century some of them were assigned portfolios within the police that were defined by subject area. In the long term in police hierarchy, they prevailed over the inspectors, and in the nineteenth century as *commissaires de police*, they became the senior functionaries in the new police system.

In summary it may be said that from around 1700 Paris produced Europe's first modern police. Their defining feature was that unlike the courts before them, they no longer functioned reactively with the object of supporting informal means of social control and excluding offenders deemed by the stakeholders concerned to be no longer tolerable (or as one might say today, no longer capable of reintegration or resocialization) from the community through formal sanction. The new police began to develop a form of social control which was formal, institutional, and preventative. They now decided who posed a risk to society; this decision was no longer left to the agencies of informal social control. This can be seen as the transfer of essential elements of social control from society to an arm of the state administration, which was to become a characteristic feature of the development of the police in the nineteenth century.

The Transition to the Nineteenth Century: The Continent and England

The development in Paris initially remained a special case in Europe. The attempt to transfer the Parisian system to the major towns and cities throughout France collapsed in 1699 in the face of resistance by local governing elites. Not until the Revolution was the Paris model, supplemented by the gendarmerie which had developed out of the *maréchaussée* and with modified official titles and certain adaptations, adopted across the country. Despite intense interest the model was not initially imitated in other European states. The Czar of Russia took an interest in the Paris model in 1718, as did the King of Prussia in 1742 and the Empress of Austria in 1751, and in some cases detailed studies were prepared. However in every country political resistance prevented anything more than fragmentary implementation. Not until after the French Revolution was the model spread across the continent at the hands of Napoleon's troops.

In England the Parisian system was always regarded as an expression of French tyranny, in stark contrast to the liberties won in England by the common man. For this reason (and because England was not occupied by Napoleon) there was substantial resistance to the adoption of the police system that was spreading across the continent and indeed to any change in the customary system of public security. Several police reform bills were rejected by Parliament. It was not until 1829 that the traditional system characterized by the *parish constables* and their *watchmen* was supplemented first in London and, from 1835 to 1839 throughout England, by uniformed police units organized on a standardized basis. However the old system continued to exist for a while in parallel both in London and in many towns and cities in England and Wales.

Conclusion and Suggestions for Future Research

The emergence of police organizations brought a fundamental change in the way European societies functioned. With the arrival of the police, policing became far more a function of the state than in the past. The procedures by which society had negotiated social order were replaced by state decision-making processes. The ground had already been laid since the sixteenth century by progressive legislation which sought to regulate ever more aspects of social life and expected ever greater obedience. However, so long as the implementation of these new laws went hand in hand with qualifying negotiation processes which respected local circumstances, the powers of these laws to reconfigure society were minimal. It was only with the arrival of a police force able and intended to compel compliance with state-imposed norms that the state in the long term displaced the informal processes by which norms were produced in society.

The practices of social control – for that is what policing means – changed as a result. Originally linked inseparably with the informal processes by which norms were produced, these practices were an integral part of day-to-day activities and communications (Sälter 2002a). With the creation of the police, the decision as to who was regarded as still capable of integration and who was deemed unacceptable passed from local stakeholders to state institutions. The criteria by which normal and deviant behavior were measured accordingly also came under greater state influence. The very existence of a police organization, which initially dominated and in the longer term largely displaced the direct negotiation of norms between social stakeholders of differing status, together with the legal system established in parallel, set these stakeholders apart from one another, since they were always aware of the presence of the state as a third protagonist between them who in the event of dispute would appear in the form a policeman (or indeed several).

Following this line of argument, one is left with some assumptions and requirements with regard to research into the police. One might consider whether the greater steps in centralizing and professionalizing the police ought not also to be explained in the context of power policy. What appears against a certain theoretical background to be a deliberate state reaction to socioeconomic processes may also be interpreted as the outcome of rivalry between certain social groups to achieve social dominance. It would surely be conducive in this regard to more strongly disengage police history from the normative burdens of earlier research with its affirmative references to the state and step back from the fixation with the crime rate as the prime motive for these developments. The criticisms leveled by police reformers at earlier institutions which they decried as lacking in efficiency should be understood as a strategy of legitimation. The frequently encountered uncritical acceptance of this justification by scholars is also likely to be linked with an underestimate of the competence and adaptability of traditional practices in establishing public security.

Most reforms of this nature in the early modern period were not primarily concerned with the common good but were aimed at reorganizing the processes of authority to give greater power to the political center, generally the crown. The implicitly or explicitly underlying assumption that the authorities in the early modern period were concerned solely with protecting the population against crime has already been refuted by Dirk Blasius (1988: 141) as displaying "an excess of simplification and naivety." Hay and Snyder (1989) on the other hand have emphasized the political character of the English police reform of 1829. In this case, as in Paris in the seventeenth century, one aim of the reform was to roll back the social power of urban and rural elites, which is why the latter, as Clive Emsley (1991) has shown, long maintained their opposition. It would consequently appear meaningful to embed police history more firmly into the context of research into the history of authority and how this has changed, and integrate specific social backgrounds.

This would clear the way for a reassessment of the relationship between police and society. If one assumes police organizations to be an institutionalized means of exercising authority, then scholars must disassociate themselves even more strongly than in the past from a purely institutional history perspective in which police practices lie hidden in a blind spot. Our interest should then be directed towards the process, contested on a daily basis by various social groups, of producing and reproducing order in society or, in the words of Alf Lüdtke (1992), towards the practices of policing. Concerning changes in police practices, which as an implementation of authority must adapt themselves to social changes, that is to say, socioeconomic processes, it may heuristically be assumed that these occurred not in the form of major reforms, but in a series of ad hoc revisions on a trial-and-error principle. Changes aimed at diversification and professionalization can only be studied through an analysis of day-to-day relations between police and populace, in which consideration should also be given to the power structures in society. Microanalyses of the actual activities of individual police organizations appear desirable.

Thus far in the study of police history, there has been no shortage of enquiry into the influence of social conditions on police organization. If the history of the police is interpreted as a part of political social history, two further effects must be considered. On the one hand it is to be anticipated that, as Philip Rawlings (1995) has emphasized, in most societies, policing takes place to a significant extent out of the hands of the police. This means that informal social control must be integrated into police history as the counterpart of state control of and via the police. It can be both allied with and in competition with the state's intended control over society. Taking the example of Paris in the early modern period, the transfer of power from corporative instances to the police reduced the opportunities open to the social elites to exercise control over the lower orders of society directly via their social conduct. On the other hand, forms of horizontal social control continued to exist within local communities and appeared to escape the influence of the police. This example shows that a change in informal social control, the manner in which it was influenced by institutional change in the police, and forms of cooperation between informal and official social control might prove interesting subjects for investigation with the capacity to provide valuable clues to the parameters within which the police operate.

Another source of influence is to be found in the reverse effect of institutional changes on the relationship between police and society and on society itself. In Paris the formation of a police organization brought about a change in the power relationships and constellations of society. Because this change resulted in a more effective assertion of the force of the state, new conditions emerged for private actions and new rules by which conflicts were played out. Among other things, individuals were more strongly dependent on state support in asserting their ideas of order and pursuing their interests and were less often able to achieve their aims by the use of force. They became increasingly obliged to involve the state, which itself was increasingly coming to act as controller of society, as arbitrator, or as mediator in private conflicts (Rousseaux 1996). As a result it became necessary for each individual that others, too, should assent to the rules for the nonviolent resolution of conflict. In turn these processes caused a shift in security thresholds and in perceptions of safety and security.

In the first place it was probably the burgesses who adapted themselves to the changing standards of behavior concomitant with the development of police. The centralization of the functions of authority and a more effective assertion of the state monopoly on force caused them to turn increasingly to the state in general and the functionality of the police in particular to assert their interests and preserve their personal safety. As Allan Silver (1967: 97) has argued: "With rising standards of public order has come an increasing intolerance of criminality, violence and riotous protest." Eric H. Monkkonen (1996: 208) similarly commented that ". . . the creation of the police reflected a growing intolerance for riots and disorder rather than a response to an increase of crime." Thus, it was the creation of the police that established a need for greater police efficiency. The elites subsequently had need of the police in order to monitor those elements of the population who were under suspicion of having adapted themselves to a lesser extent to the new standards. In their perception, the lower orders of society became a dangerous class which had to be controlled. This was reflected in England, as Hay and Snyder (1989) have identified, in an increasing number of complaints and reports to the police.

If one concurs with Silver and Monkkonen, it was the elites who insisted on the efficiency of the new police and observed them with a critical eye. For similar reasons, as David Garrioch (1994) has emphasized, the police in Paris in the eighteenth century were concerned to create a balance between the interests of the crown in asserting control and those of the city elites in maintaining security. As a result, by depriving the urban elites of power and requiring them to adapt to the exercise of force by the state, the police were compelled to act with a certain degree of efficiency. When reforms were enacted, local elites watched to see whether the state fulfilled its promise of greater security. They critically weighed the justification of the public good against the central grip on the use of force and demanded efficiency. This resulted in police practices being matched to their demands, in a process which in turn drove the adaptation of the police to meet the needs of society.

Some of the arguments expressed are based on research into the nineteenth and early twentieth centuries. This is partly because there has so far been too little research into the practices of social control in the early modern period. Fixated on institutions and fascinated by the sources they have produced, scholars have concentrated either on identifying the precursors of modern police organizations in the premodern world or on studying the practices of the courts. Despite numerous enticements, little attention has been paid to the practices of policing and social control. The fact that the norms of day-to-day life were for a longtime generated and overseen locally and informally as part of daily interaction is meanwhile acknowledged but has yet to be systematically investigated to a desirable extent. It would appear more necessary than ever to escape the temptation to project modern thought processes, motives, and concepts of order onto the protagonists of the early modern period.

What's more, with the possible exception of England, there has been a lack of research into the institutions and office holders who guaranteed public safety at local and regional level in early modern Europe. We know too little about how the security forces were charged with their duties and how they were paid, about their function and reputation, and their integration into society and the structures of the state. There also appears to me to be a lack of studies which address the techniques and practices of social protagonists jointly with those of state institutions in most European countries in the context by which they contributed to the maintenance of public security and the imposition of their ideas of order on their fellow men.

Rather than research which focuses on institutions and their premodern precursors and is pleased to discover deficiencies in the early modern period, there is an attraction in learning more about the actual practices in their social context and the associated discourses on good order. This would make no small contribution to our understanding of early modern societies.

Related Entries

- ▶ British Police
- ▶ Communities and the Police
- ▶ Community Policing
- ▶ Comparing Police Systems Across the World
- ▶ Conceptualizing of Police
- ▶ German Police Until 1918
- ▶ History of the Police Profession
- ▶ Policing of Peacekeeping
- ▶ Police Culture
- ▶ Role and Function of the Police

Recommended Reading and References

Bayley DH (1975) The police and political development in Europe. In: Tilly C (ed) The formation of the national state in western Europe. Princeton University Press, Princeton, pp 328–379

Beattie JM (2006) Early detection. The bow street runners in late eighteenth-century London. In: Emsley C, Shpayer-Makow H (eds) Police detectives in history. Ashgate, Burlington, pp 15–32

Blasius D (1988) Kriminologie und Geschichtswissenschaft. Bilanz und Perspektiven interdisziplinärer Forschung. Gesch Ges 14:136–149

Denys C (2002) Police et sécurité au XVIII^{e} siècle dans les villes de la frontière franco-belge. L'Harmattan, Paris

Emsley C (1991) The English police. A political and social history. Harvester Wheatsheaf, Hemel Hempstead

Emsley C (1999) A typology of nineteenth-century police. Crime Hist Soc 3(1):29–44

Garrioch D (1994) The people of Paris and their police in the eighteenth century: reflections on the introduction of a "modern" police force. Eur Hist Q 24:511–535

Härter K (2010) Security and "Gute Policey" in early modern Europe: concepts, laws, and instruments. In: Zwierlein C, Graf R, Ressel M (eds) The production of human security in premodern and contemporary history, Special issue of Historical Social Research, pp 41–65

Hay D, Snyder F (1989) Using the Criminal Law, 1750–1850. Policing, private prosecution and the state. In: Hay D, Snyder F (eds) Policing and prosecution in Britain, 1750–1850. Clarendon, Oxford, pp 3–52

Holenstein A (2003) Gute Policey und lokale Gesellschaft im Staat des Ancien Régime. Das Fallbeispiel Baden (−Durlach), 2 vols. Bibliotheca Academica, Tübingen

Holenstein A, Konersmann F, Pauser J, Sälter G (eds) (2002) Policey in lokalen Räumen. Ordnungskräfte und Sicherheitspersonal in Gemeinden und Territorien vom Spätmittelalter bis zum frühen 19. Jahrhundert. Vittorio Klostermann, Frankfurt

Iseli A (2003) "Bonne Police". Frühneuzeitliches Verständnis von der guten Ordnung eines Staates in Frankreich. Bibliotheca Academica, Tübingen

Lüdtke A (1992) Die "Sicherheit des Staates". Zu einer fatalen begriffsgeschichtlichen Kontinuität in Deutschland. In: Florath B, Mitter A, Wolle S (eds) Die Ohnmacht der Allmächtigen. Geheimdienste und Polizei in der modernen Gesellschaft. Ch. Links, Berlin, pp 18–35

Milliot V (2008) Mais que font les historiens de la police? In: Berlière J-M et al (eds) Métiers de police Être policier en Europe, XVIII^{e}-XX^{e} siècle. Presses Universitaires de Rennes, Rennes, pp 9–36

Monkkonen EH (1996) The urban police in the United States. In: Emsley C, Knafla LA (eds) Crime history and histories of crime. Studies in the historiography of crime and criminal justice in modern history. Greenwood Publishing Group, London, pp 201–228

Napoli P (2003) Naissance de la police moderne. Pouvoir, normes, société. La Découverte, Paris

Paley R (1989) "An imperfect, inadequate and wretched system"? Policing London before peel. Crim Justice Hist 10:95–130

Rawlings P (1995) The idea of policing: a history. Polic Soc 5:129–149

Rousseaux X (1996) Entre accommodement local et contrôle étatique: pratiques judiciaires et non-judiciaires dans le règlement des conflits en Europe médiévale et moderne. In: Garnot B (ed) L'infrajudiciaire du Moyen Age à l'époque contemporaine. Éditions Universitaires de Dijon, Dijon, pp 87–107

Sälter G (2000) Lokale Ordnung und soziale Kontrolle in der frühen Neuzeit. Zur außergerichtlichen Konfliktregulierung in einem kultur- und sozialhistorischen Kontext. Kriminol J 32:19–42

Sälter G (2002a) Gerechtigkeit und soziale Ordnung. Konflikte um individuelle Interessen und die rechte Ordnung der Dinge in Paris im frühen 18. Jahrhundert. In: Griesebner A, Scheutz M, Weigl H (eds) Justiz und Gerechtigkeit. Historische Beiträge (16.-19. Jahrhundert). StudienVerlag, Innsbruck, pp 61–74

Sälter G (2002b) Urbanisierung, Migration und Kriminalität als Begründungskontext für die Entstehung von Polizei. Zur Entstehung einer eigenständigen Polizei im Paris des Ancien Régime,

PoliceyWorkingPapers 5. www.univie.ac.at/policey-ak/pwp/pwp_05.pdf
Sälter G (2004) Polizei und soziale Ordnung in Paris. Zur Entstehung und Durchsetzung von Normen im städtischen Alltag des Ancien Régime (1697–1715). Vittorio Klostermann, Frankfurt
Silver A (1967) The demand for order in civil society: a review of some themes in the history of urban crime, police and riot. In: Bordua DJ (ed) The police: six sociological essays. Wiley, New York, pp 1–24
Soman A (1980) Deviance and criminal justice in western Europe, 1300–1800. An essay in structure. Crim Justice Hist 1:1–28
Williams A (1979) The police of Paris, 1718–1789. Louisiana State University Press, Baton Rouge

Early Quantitative Methods

► Quantitative Methods Used by Criminologists in the Nineteenth Century

Early Release from Prison

Frieder Dünkel
Faculty of Law, University of Greifswald,
Greifswald, Germany

Synonyms

Conditional release, Parole

Overview

Since the beginning of the history of prisons, early release and parole have been a major issue of concern. While at first being in the hands of the king or other sovereigns, in the nineteenth and early twentieth centuries early release and parole evolved into a judicial measure. The aim of early release has always been to improve the chances of rehabilitation. The idea of granting a favor was replaced by the idea of an evidence-based form of rehabilitation. The traditional mode of early release in the Anglo-Saxon world is parole after a (relatively) indeterminate prison sentence, while in the continental European world conditional or unconditional early release is provided after a specific part of the (determinate) sentence has been served (regularly two thirds or half). Both forms are optionally or often mandatorily linked to supervision by the probation service or similar services. Since the early nineteenth century, the voluntary (third) sector has traditionally played a role (e.g., charities, the church, and others). The purpose of this entry is to identify the extent to which a coherent and fair policy on release from prison is developing within the European Union and in the USA, and what can be seen as "good" or "best practice" in this field. Therefore, evaluation on recidivism and desistance from crime is also included.

Conceptual clarity has not been easy not least because the fine line between imprisonment and liberty is not always clear. Definitional challenges abound when it comes to comparing different forms of release. Only limited statistics are available.

Common themes include the need to clarify the purposes of conditional release, and the need to identify appropriate conditions and requirements. This entry urges for a greater use of early release to achieve the basic aims of punishment, especially the reintegration of offenders.

Introduction

This entry is devoted to the release of sentenced prisoners before they have served their sentence in full, but also to offenders released conditionally from psychiatric or other institutions that deal with offenders who are not criminally responsible. There are similar and overarching themes in all jurisdictions. The most basic of these is that early release has been a feature of penal practice for as long as sentences of imprisonment have been a primary form of punishment. What has characterized early release in Europe and elsewhere over the past one and a half centuries is the search for a fair and justifiable form of early release, which does not depend on the whims of a sovereign ruler but instead forms part of a coherent penal policy. Since the early nineteenth century, we can find regulations for early release by an act of grace of the sovereign (1813 in Bavaria, 1830s in Australia and the USA, 1847 in France, and 1853 in England, see Dünkel in Dünkel and Spieß 1983, p. 400 f.). "Juridification" began in the late nineteenth century with legal regulations for probation laws (famously the Probation of Offenders Act in England 1907), and for early release in Germany (1871) and Austria (1912). However, in continental Europe it took another half century to introduce the probation service (e.g., 1953 in Germany). Only in the second half of the twentieth century, a differentiated system of early release combined with supervision by the probation service and in cooperation with the prison service emerged when prison reform laws provided measures such as prison leaves and open regimes for the preparation of release. The Scandinavian prison reforms (e.g., Sweden in 1964) and reform laws in other countries (Germany in 1977) introduced the idea of transition and continuous care through prison leaves, open prisons (promoted by the UN Standard Minimum Rules for the Treatment of Prisoners in 1955), work release programs, and halfway houses. But the idea of rehabilitation somehow declined in the late 1970s in some countries and a more repressive understanding of what prisons should aim at prevailed in the Anglo-Saxon world (deterrence and incapacitation). Continental Europe somehow resisted the punitive turn and kept the rehabilitative philosophy of imprisonment. Although some scholars tend to interpret the emergence of risk assessment and managerialism as an expression of the punitive turn (Garland 2001; Tonry 2004), it must be recognized that the modern welfare philosophy at least in continental Europe has survived (see Snacken and Dumortier 2012) and is currently even expanding toward a coherent system of integrated services of the prison and probation/aftercare services. Evidently that has also to do with social control, in particular when more and more systems also rely on electronic devices and GPS-based control systems, but that is the truth only in those countries which do not provide the support of the probation services in these electronically supervised cases.

The traditional mechanism for releasing offenders early was – as mentioned above – the royal pardon or amnesty. In theory, an amnesty sets aside the conviction and sentence, while the pardon does not do so but merely allows immediate release of the prisoner or class of prisoners to whom it was granted. Notionally they are often both granted to a class of prisoners to celebrate some anniversary or other great national occasion, but in fact they are used as devices to reduce the prison population to a more manageable level. General pardons and amnesties have long fallen into disuse in countries such as the UK and Germany. In France and Belgium they were used regularly until fairly recently but there too the state, more specifically the former and the current president and king, respectively, were and are currently reluctant to use them, because they are seen as an unjustifiable use of executive power. Only in Italy, where there is a much higher level of skepticism about the use of state power to punish than in most other European countries (Nelken 2009), they are still an important element of penal policy.

While the constitutional objections against the use of pardons and amnesties may be well founded, the question remains whether other mechanisms for early release are performing the function of controlling the size of the prison population as effectively as they could or should. The evidence is mixed: In Finland, a well-designed system met the requirements of due process and was a key factor in reducing the prison population (see Lappi-Seppälä 2007; Dünkel et al. 2010a, b). The reform in Austria of 2008 has also contributed to a decrease of the prison population and at the same time has emphasized rehabilitation more strongly by excluding general deterrence as a consideration when deciding on early release after having served two thirds of the sentence. On the other hand, policy changes can also go in the opposite direction, as in Spain where the abolition of "good time" in 1996 has contributed to the increase of the prison population rate. In Germany, psychiatric hospitals experienced serious problems of overcrowding when in 1998 the legal criteria for early release were formulated more restrictively.

Early Release from Prison, Table 1 Recidivism of those released in Germany in 1994 and 2004 or sentenced to suspended sentences after 3 years

	1994 (%)	2004 (%)
Suspended prison sentence	39.7	37.1
Unconditional prison sentence	52.3	48.1
Suspended youth prison sentence	54.0	49.3
Unconditional youth prison sentence	74.8	66.4

The combination of information from various sources reveals that there is provision for early release in all European countries. It is, however, organized very differently in respect of who the decision makers are, the minimum period to be served, the extent of the remainder of the sentence during which the offender remains subject to some form of restriction, the question of whether release is (predominantly) automatic or subject to an individual prognosis/assessment, the possibility of combining the decision on early release with a requirement specifying what the offender may or may not do during a period of conditional release, and finally the question relating to the recall of conditionally released offenders and the imposition of sanctions for infringements of the conditions on which the release may have been granted. Dünkel et al. (2010b, p. 409 ff, Table 1) provide a synoptic overview of the legal requirements of early release which are summarized in this entry.

Legal Concepts and Basic Philosophies of Early Release

The aim of early release has always been rehabilitation and reintegration of the offender into society. This was well expressed by the German Federal Constitutional Court when it stated with regard to the principle of "resocialization" (rehabilitation): "From the point of view of the offender, this interest in resocialization grows out of his constitutional rights in terms of article 2 (1) in conjunction with article 1 of the Basic Law (that is, the right to develop one's personality freely in conjunction with the protection of human dignity). Viewed from the perspective of the community, the principle of the social state requires public care and assistance for those groups in the community who, because of personal weakness or fault, incapacity or social disadvantage, were retarded in their social development" (BVerfGE, *Bundesverfassungsgerichtsentscheidungen* 35, pp. 202, 235–236). Prisoners and (released) ex-prisoners belong to this group. So since the early 1970s, the idea of granting a favor by early release was replaced by the idea of rehabilitation as a constitutional right. Spain (Art. 25(2)) and Italy (Art. 27(3)) have explicitly introduced rehabilitation as constitutional rights. In recent days, rehabilitation by a system of graduated liberty with early release and aftercare is also seen as an evidence-based strategy for preventing reoffending (Andrews et al. 1990; Dünkel 2013; Lösel 2012; MacKenzie in this volume).

Other legal aims such as general prevention (deterrence) or "just deserts" should not be of relevance. Retribution (referring to the guilt of the offender), too, should be considered only in the sentencing stage and has no impact on the question of early release (see Dünkel 2013; Dünkel et al. 2010b, p. 404 f). The same applies to the role of the victim. The welcomed improvements of the position of the victim in criminal procedures should not go so far that the opinion of the victim determines the early release decision. Therefore, "victim impact statements" are not conforming to the rehabilitative idea of providing the reentry of prisoners into society depending on his development and resocialization.

Perhaps the simplest justification for early release is to reduce the prison population and cope with prison overcrowding. Ultimately, the importance of this aspect cannot be overlooked. Even in England and Wales, where there has been a significant increase in the prison population as a whole, an increasing number of steps were taken to release some categories of "less serious" offenders early in order to make room for the burgeoning overall prison population. The result of this *bifurcation* is a growing distortion between the actual implementation of sentences for different types of offenses depending on whether those that commit them are regarded as a category, as "dangerous" or are otherwise in disfavor with the authorities. The same pattern seems to be playing itself out in practice in Belgium where the new law governing the release of prisoners serving sentences of under 3 years has not been brought into effect, inter alia because a system administered by the prison authorities is able to keep a cap on a growing prison population by releasing these relatively short-term prisoners earlier than a more formal process is likely to do.

Looking at the modes of early release, two major systems have been developed. The traditional way of early release in the Anglo-Saxon world is parole after a (relatively) indeterminate prison sentence (see Allen et al. 1985; Petersilia 2003). This system came under pressure in the 1980s and was abolished in many US states. The somehow justified criticism against indeterminate sentences and an arbitrary system of release increased; however, the length of stay in prison as in the same time deterrence oriented policies and truth-in-sentencing-laws emerged which contributed to the American growth in the prison population ("mass incarceration", see for a summary Petersilia 2003, p. 221 ff).

In the continental European world, conditional or unconditional early release after having served a specific part of a determinate sentence (regularly two thirds or half) is provided. Both forms are optionally or often mandatorily combined with the supervision of the probation or similar services. The voluntary (third) sector has traditionally played a role since the early nineteenth century, for example, charities, the church, and others. There is one other important difference between the parole system and the continental European early release system. Parole boards are not necessarily composed of judges – the psychological and psychiatric or social pedagogic experts are also involved in the decision-making itself, whereas in the continental European law the decision is made by judges (however regularly after an expertise has been presented, in the more serious cases a forensic-psychiatric examination). This holds, however, only where the law endows the court with discretionary power as to release or not to release. Therefore, one has to differentiate automatic (or quasi-automatic) early release from systems relying to discretionary decisions, primarily based on prognostic examinations.

The Decision-Making Process of Early Release: Who Decides? The Role of Courts, Parole Boards, Expert Knowledge, Victims, and the Public

Different decision-making bodies are responsible for early release in various countries. It is not universally the case that, as required by the German Constitution (Art. 104(2) GG), an independent court must decide about early release. The German Constitution requires that any imposition as well as the termination of prison sentences as well as any other form of deprivation of liberty (e.g., remand prison and safety measures for psychiatric delinquents) by early release must be ordered by a judge. Therefore, a system of parole boards outside the judiciary would be unconstitutional. The question of transferring the competence of deciding on early release to the prison or other administrative authorities has never been an issue in Germany.

In Belgium, the prison administration makes the decision in all cases of imprisonment of up to 3 years; in other cases, the decision is made by Sentence Implementation Courts composed of three persons: a judge, an expert in social reintegration, and an expert in prison matters. Also in Denmark, England and Wales, Finland, Sweden, and Switzerland, some decisions are made by the prison administration or by a section of the Ministry of Justice. In England and Wales, as in many jurisdictions of the USA, the Parole Board, which decides cases of life and other indeterminate sentences, has an interdisciplinary composition like the Belgian Sentence Implementation Courts, and the Italian and Slovenian commissions (in the case of Slovenia, however, composed only of members of the judiciary). Elsewhere, the idea of a specialized judge or tribunal with experience of the implementation of sentences has been accepted (e.g., Croatia, France, Germany, Hungary, Italy, Poland, and Spain). Only in a few countries is the court of first instance, which is not a specialist in this regard, also responsible for this decision (the Czech Republic, Estonia, Greece, Lithuania, and the Netherlands). The Netherlands is a rather exceptional case – there, since the recent reform of 2008, the public prosecutor has been responsible (see Dünkel et al. 2010b, p. 409 ff, Table 1). Parole boards and other mixed bodies deciding on early release in theory guarantee a broader consideration of scientific prognostic evidence and examination. However, the influence of experts who are involved in preparing the judicial decision by a prognostic expertise (psychologists, psychiatrists, and also social workers) should not been underestimated if the decision is restricted to the judiciary. Research by Dünkel and Ganz (1985) in Germany demonstrated the major impact of negative prognoses from these experts, showing that in such cases a positive decision is almost never made in favor of early release for example. On the other hand, a positive recommendation by a prognostic expertise does not always result in early release being granted as the prosecutors are also consulted and they sometimes express serious reservations (because of the legal history of the offender, the crimes committed, etc.).

Therefore, one should not be blinded by legal formalism. In some European countries, conditional release is still sometimes decided directly by the prison authorities. This is true in Belgium for sentences of less than 3 years and in England and Wales when early release that goes beyond the statutory automatic period is considered. However, even where de jure this is not the case, the de facto influence of the prison authorities may be very significant. For example, temporary release may be used, as in Ireland, as a more permanent form of conditional release. Temporary release elsewhere is normally, but not necessarily, a preparation for conditional release. There is an important interplay between temporary and more permanent forms of early release (however categorized). Temporary release may in some jurisdictions be granted by the prison administration, and whether or not this can be achieved may have an important impact on subsequent decisions by parole boards or judges. The interaction of administrative, judicial, and quasi-judicial bodies is clearly a vital area for further research. The decisions that are "vital" to conditional release may in reality not be made by the formal body that is charged with making the final decisions. In England and Wales, where this has always been recognized by the courts, the practice is also of fundamental importance. For example, where the English Parole Board formally makes the all-important decision on the release of lifers, the reality is that if the prison authorities have not moved the lifer to an open prison and if the probation service have not completed a satisfactory release plan, the Parole Board is highly unlikely to recommend release (Padfield 2002). Thus the true gatekeepers are not only the formal Parole Boards, but also the administrative players who supervise the progress of the prisoner through the system. Yet the prisoner's rights (e.g., to legal advice and to full disclosure of documentation) may be significantly less for these so-called administrative decisions.

Significant in this regard is also whether the decision-making body is *automatically* involved or whether it only makes a decision when called upon to do so. While in Germany the chamber responsible for the implementation of punishment may be called upon to make a decision by the prisoner when he is about to complete two thirds of his sentence, in many other countries the prison administration or the prosecuting authority may also make applications. There are also examples of other persons or institutions having the right to make applications (see Tubex and Tournier 2003, p. 18 f).

Prisoners' procedural rights in respect of these applications are fairly similar. In most countries (Cyprus and Luxembourg are exceptions) the prisoner is heard, and mostly he takes part in the proceedings (a hearing or an oral procedure). The right of access to documentation (in some cases via a lawyer) is recognized in most countries, however, not in Bulgaria, Croatia, Cyprus, Latvia, Lithuania, Malta, Poland, Portugal, Slovakia, and Turkey. Practically without exception prisoners are told of the outcome (Tubex and Tournier 2003, p. 19 f), although the question, in what form, naturally remains.

A scientific risk assessment in respect of the risk of offenders committing further crimes in the future takes place in only a few countries (e.g., Denmark, in part in England and Wales, Germany, Italy, Norway, Poland, and Scotland) and certainly not in those countries where there is quasi-automatic release (e.g., Finland). Often such risk assessment is used only for particular classes of prisoners who are regarded as dangerous or those who are serving life sentences or safety measures.

Legal Preconditions of Early Release: Prognosis and Time Expired

The fundamental distinctions that are drawn in the requirements for prognoses underpin the wide penal policy implications of early release. In Europe, there are two fundamental types of early release, namely, the discretionary release system and the mandatory or automatic release system. Both are explicitly recognized by the 2003 Recommendation of the Committee of Ministers of the Council of Europe concerning Conditional Release. Discretionary release based on a positive individual prognosis is found particularly in Austria, Croatia, Germany, Estonia, Hungary, Poland, Russia, Slovenia, Spain, and the Czech Republic. More or less automatic release dispenses to a great extent with individual prognoses. These forms of release are gaining increasing importance in Europe. Thus, England and Wales, Finland, Greece, Sweden, and Turkey (where good conduct in prison is required) have introduced a generally mandatory form of early release. In England and Wales, one has, in addition to automatic release in the case of determinate sentences, the "classical" release decision-making process on the basis of prognostic evaluation by the Parole Board in the case of indeterminate sentences (life imprisonment and so-called extended sentences). The position in Scotland is similar, with early (i.e., at the two-thirds point) release of all prisoners serving 4 years or more, and of all life prisoners, being based wholly upon assessment of risk by the Parole Board.

Some countries do not dispense with prognostic evaluations entirely but in the face of practical problems with such evaluations reduce the requirements in cases of uncertainty to the absence of a negative prognosis. This means that early release occurs as the norm unless there is a clear negative prognosis that speaks against doing so. This form of regulation is found, for example, in Belgium, Denmark, and Switzerland (as well as in the Netherlands for cases dealt with under the old law until 2013). In Sweden, early release follows routinely after two thirds of the sentence, unless particular grounds militate against it. Since 1 January 2007, these grounds include the breach of a conditional sentence that results in it being revoked.

One can also include Belgium in the countries that – except for a particular group of sex offenders – have developed an automatic release system. In particular, prison sentences of up to 1 year are reduced to a period of 15 days to 3 months.

French criminal procedure has some noteworthy characteristics (see Art. 729 ff CPP). The reforms of 2000 expanded the grounds on which early release can be granted. Instead of a prisoner having to produce strong evidence of his resocialization, he now has to produce only evidence that he has attempted to reform, which appears more realistic. Conditional release is possible after half the sentence or in the case of recidivists after two thirds of the sentence. A significant expansion is also contained in the rule that a parent whose child was aged under 10 and living with him or her prior to sentence shall be given extra consideration for release. This is designed to strengthen family relationships and reduce the negative effects of imprisonment. In such cases, the minimum period that otherwise should be served falls away so that it is possible for someone to be released after a very short period. However, this form of early release appears to be used very rarely in practice. A similar family-oriented regulation is provided in Italy.

Features of the *temporal requirements* for conditional release are the absolute minimum time that has to be served on the one hand, and the relative part of the sentence that has to be served on the other. Most countries have very flexible requirements that sometimes vary in respect of particular groups of offenders. The US-parole system is probably one of the most flexible systems, as the proportion of the sentence to be served or the minimum time period can be very low compared to the sentence imposed by the court (see Petersilia 2003).

In Europe, the variation is considerable. Thus, for example, in Belgium at least a third has to be served, in the case of recidivists at least two thirds; in France, Poland, and the Czech Republic first offenders must serve half (also in Italy) and recidivists must serve two thirds (in Italy, three quarters). In Estonia and Russia, the minimum is half to two thirds depending on the seriousness of the offense. In Finland, prisoners are regularly released after half of the sentence, recidivists having served a prison sentence during the preceding 3 years after two thirds.

It is noteworthy that release after a third, which in Germany is only possible in juvenile criminal law, is possible routinely or at least by way of exception in Belgium, Denmark, Greece (one fifth or two fifths), Croatia, Lithuania, and Slovenia. Austria, Belgium, Croatia, England and Wales, Estonia, Finland (for those under the age of 21 years at the time of the offense), France, Greece, Hungary (in the case of sentences to more than 3 years), Italy, Lithuania, Poland, Russia, Scotland, Slovenia, the Czech Republic, and Turkeys, all allow release after half the sentence to a greater extent than in Germany, where release after half of the sentence remains exceptional (see s. 57(2) Criminal Law). The requirement of wider possibilities for release is apparent and predictable in countries with a priori harsher punishment and longer average sentences than Germany (e.g., Russia: see Dünkel 2013). While the 50 % requirement is rare, cases in which two thirds have to have been served are even rarer. It is the regular form of procedure in Denmark, Germany, Malta, the Netherlands (for more than 2 years of imprisonment), Romania, Scotland (this being the maximum term when release does not take place after half the sentence), and Switzerland (see Dünkel et al. 2010b, p. 409 ff, Table 1).

Longer shares that have to have been served before release, like three quarters or even four fifths, are only found exceptionally and even then only for particular groups of offenders, for example, in Armenia, Hungary (where a distinction is drawn between those serving sentences in prisons (three quarters) and penitentiaries (four fifths)), Romania, Slovenia, and Turkey. In Spain, as a rule, three quarters have to have been served, but exceptionally release is possible after two thirds or even half of the sentence.

In a few countries, there is no provision for absolute minimum periods at all. Otherwise they are mostly, as in Germany, from 2 to 6 months. Exceptions are Finland with only 14 days, Sweden with 1 month, and the Netherlands and Poland with – in certain cases – 1 year.

One special scenario is release from *life imprisonment*. Most countries that have life sentences provide a certain minimum term to be served: It is 10 years in Belgium (for recidivists 14 years); 12 years in Denmark and Finland; 15 years in Austria, Germany, and Switzerland (exceptionally 10 years); 18 years in France (for recidivists 22 years); 20 years in the Czech Republic, Greece (with a possible remission to 16 years), and Romania; 25 years in Poland, Russia, and Slovenia; 26 years in Italy; and 30 years in Estonia and in certain cases in Hungary (Dünkel 2013, § 38, notes 46 ff with further references). In England and Wales, and in Scotland, the judge who sentences someone to life imprisonment, whether the sentence is discretionary or mandatory, may and usually does set a minimum period (the “tariff”) and usually does set a period to be served. In Scotland, the judge must set a “punishment part,” which must be served in full before release on parole can be considered. In a small number of very serious cases, no minimum is set. For these cases, as for all cases of life imprisonment, in Lithuania, the

Netherlands, and Sweden early release is an act of grace and therefore no regulations containing legal criteria exist. In Ireland, the minimum sentence to be served in case of aggravated murder may be 40 years, in case of attempted aggravated murder 20 years. But the prisoner regularly receives a remission of one quarter (related to the minimum term) and further remission and parole can be awarded. Therefore, the time spent may be between 12 and 20 years.

Only Croatia, Norway, Portugal, and Spain have no provision for life sentences. However, the range of determinate sentences in Spain is up to 20, in exceptional cases up to 40 years. In Croatia, certain (recidivist) offenders and serious crimes also attract up to 40 years imprisonment, which in practice is likely to result in about the same periods actually being served compared to those served by prisoners sentenced to life imprisonment in countries that provide for this penalty.

There are other possibilities of shortening the time spent in prison which are connected to early release mechanisms, but should not be mixed: the so-called good time schemes, that is, the remission of a sentence for good behavior, particularly working efforts of the prisoner. In Greece, 1 day worked counts as 1.5 days of imprisonment, which in the case of working prisoners means a reduction of the prison sentence by more than half (if the prisoner performs well also in other areas, further reductions are possible). If one adds to this that prisoners are normally released from prison after having served two fifths of their sentences, the actual time served could in some cases be as low as one fifth of the original sentence.

In France as well, wide-reaching reductions for good behavior are possible: up to 7 days per month or 3 months per year (see Dünkel et al. 2010b, p. 409 ff, Table 1). Also in Ireland, regular remission can amount to 25 % of the sentence. In Spain, in contrast, reduction of a sentence for good behavior or labor was abolished in the 1990s, while in Italy the so-called *liberazione anticipata* that had been introduced in 1975 has remained in force. Here, the supervisory court can guarantee that a prisoner will be given a reduction of imprisonment of 45 days per 6 months, if the prisoner shows that he has participated in resocialization programs. What is problematic about good time regulations linked to prison work is that, in the light of the high rate of unemployment in prison (particularly in Central and Eastern European countries, see van Zyl Smit and Dünkel 2001), equality of opportunity cannot be guaranteed. The prison system should provide for full employment of prisoners so that the opportunity to reduce their sentences can be given to all prisoners. If the system fails to do so, prisoners who want to work but cannot do so should be compensated in other ways, which may be linked to early release.

Continuous Care and Aftercare of Released Inmates: The Role of the Probation and Aftercare Services

Most European countries have mechanisms that support offenders while they are serving part of their sentences in the community. In this regard, the requirements that are set for such persons in Germany are common in other countries, too, although without the doctrinal distinction between directives and obligations that German law draws in this regard. In some countries, such as Estonia, they are compulsory, while in others, such as Denmark, France, and Switzerland, they are used routinely. It is interesting that in some countries probationary supervision is compulsory, for example, in Belgium, Lithuania, Poland (for specific classes of recidivists), and Russia. It is not yet clear whether the supervision in Eastern European countries will be full-scale supervision by probation officers as we know it in the West or simply a form of reporting regularly to the authorities (i.e., the police). However, there are signs that a system of parole comparable to that in Western Europe is beginning to emerge in countries such as the Czech Republic, Estonia, Poland, and Slovenia (see van Kalmthout and Durnescu 2008).

The Netherlands until recently was an exception in this regard, as early release was not followed by probation supervision or the imposition of any directives or obligations. The reform of 2008 brought new opportunities to impose probationary supervision and directives to participate in training and treatment programs.

In some countries, early release is combined with electronic monitoring, for example, in the Netherlands, England and Wales, France, Spain, Sweden, and Switzerland. In as far as electronically monitored house arrest is not a community sanction imposed directly as an alternative to imprisonment, it may be used to reduce the term of imprisonment by a maximum of 2 months (Sweden), 4½ months (England and Wales), 6 months (Finland, the Netherlands, and Switzerland), or even 12 months (France). In Germany, this strategy has not been promoted as a policy option, as most prisoners that would qualify for it are sufficiently supported by regular parole (the probation service). Another reason is that it does not have the potential for reducing the prison population significantly.

The period of conditional release is usually the remainder of the sentence. However, often, as in Germany, there are also absolute minimum and maximum periods. In Finland, the maximum period is 3 years, and minimum terms have been abolished. In other countries (Austria, Hungary, Sweden, Switzerland, and the Czech Republic), the minimum period is shorter (1 year) than in Germany or in Poland (2 years). This is important in the case of short periods of conditional release in particular, as otherwise this could widen the net of social control disproportionately. It could also mean that prisoners simply refuse conditional release. Providing assistance and supervision in the first year after release should usually be sufficient, as most recidivism occurs then.

One of the major problems of reintegrating prisoners is the gap between the end of the prison sentence and the start of aftercare. Traditionally the prison services try to prepare prisoners somehow for release, but very often their efforts fail even as regards “simple” things like finding accommodation or employment. The additional psychosocial problems to cope with the stigma being released from prison, having lost social bonds and sometimes even family support contribute to high recidivism and revocation rates (see Petersilia 2003, pp. 105 ff, 139 ff). However, in the last 20 years increased efforts for integrating offenders into the labor market (Wirth 2009) or involving the probation service in release preparations and thus guaranteeing a continuity of care have been visible in some European countries (England and Wales, Germany, France). The English National Offender Management System (although sometimes criticized concerning its implementation and mixed outcomes, see Robinson 2005; Raynor 2012, see also Moore et al. 2006) has been a model for restructuring the probation service in order to warrant continuous care (for Germany, see Dünkel et al. 2008 with further references). The problem is that the probation services are only competent if the offender has been released early or conditionally. If he fully serves the sentence, there will be no compulsory supervision except in cases of dangerous offenders, where national law provides for intensive supervision as is the case in Germany and France.

This gap of taking care of ex-offenders in some countries is only partly filled by private nonprofit organizations of aftercare. The third sector has its roots in the early nineteenth-century charity organizations and is still active in many countries (e.g., Germany), although it is not a nationwide movement that has a quantitatively important impact.

Conditions of Supervision After Release: Obligations for Behavior, Paying Restitution, and Electronic Monitoring

With the exception of a few countries that provide mandatory unconditional release, release can be conditioned on the fulfillment of prohibitions and obligations. One (more or less obviously) is the obligation not to reoffend. Other conditions aim to prevent “criminogenic” lifestyles or situations that are associated with an increased risk of reoffending, for instance, the prohibition to drink alcohol, the prohibition to visit certain places (red-light districts or - in the case of pedophile offenders - kindergartens, etc.), or the restriction not to meet certain people.

Yet other conditions may require the released person to report to and to maintain contact with a supervisor, to reside at a particular place, to undertake psychiatric/psychological/medical treatment, to undergo regular testing for illegal drugs, or finally to pay a fine or to make different forms of reparation to the victim. However, conditions may be disproportionately or unnecessarily burdensome, and the more conditions that are imposed, the more likely it is that a prisoner will breach. Conditions may be imposed that are either incompatible or unenforceable – an offender overburdened with such conditions is less likely to respect the conditions as a whole (this is a conclusion of the English Inspector of Probation investigating high-profile supervision failures).

In some jurisdictions, supervision by a probation officer is obligatory (in particular in juvenile justice systems, see e.g., Germany), whereas in other jurisdictions it is at the discretion of the court. In these cases, referrals to the probation service are made if doing so will improve an otherwise negative or neutral prognosis. In many countries, certain offender groups (recidivist, sexual, and violent offenders) are regularly or mandatorily supervised by the probation service upon release. In the case of "dangerous" offenders, some countries such as France and Germany provide for offender supervision by the probation service even if the offender has fully served his sentence, sometimes combined with electronic monitoring.

Breach of Rules, Amendment of Supervision Rules, and Revocation

All systems that provide conditional early release also provide mechanisms for dealing with breaches or noncompliance. It is often the case that serious infractions (like severe reoffending) result in a revocation of the remainder of the sentence. Revocation for breaching conditions (e.g., not having regular contact with the probation officer, not paying reparation to the victim, not fulfilling community service) without however having reoffended is a problematic issue in this context. Many countries restrict revocation to cases of serious and repeated noncompliance and provide statutory rules that explicitly make revocation a measure of last resort (e.g., Scandinavian countries, Austria, Germany), instead of providing courts with powers to amend conditions or extend periods of supervision. In Finland and Sweden, revocation is only possible if the offender has committed a new crime leading to an unconditional prison sentence. It should also be noted that in the case of revocation in some cases only the remainder of the suspended rest of a sentence may to be served in prison, whereas in other countries the total suspended period of the sentence is to be served (e.g., Germany). This is contrary to the European Rules for Juvenile Offenders Subject to Sanctions or Measures of 2008 (ERJOSSM), which stipulate that where a revocation or modification of a community sanction is being considered, due account shall be taken of the extent to which the offender has already fulfilled the requirements of the initial sanction (Rule 48.4, see Council of Europe 2009).

International standards of the Council of Europe require that breaches in themselves should not constitute an offense (Council of Europe Rec. (92)16 on Community Sanctions and Measures, CSM, Rule 84) or at least "not automatically" constitute an offense (Rule 30.2 ERJOSSM). And the decision to revoke a community sanction shall not necessarily lead to imprisonment (Rule 86 of the CSM 1992). Where possible, modified or new community sanctions shall replace the previous ones (Rule 30.1 ERJOSSM). Furthermore, the offender and his family should not be charged with the costs of the supervision, for example, to pay for "renting" electronic devices (as it is the case in the USA).

The situation in the USA seems to be different from Europe as concerns revocation. Petersilia describes the alarming fact that in 1999 parole violators constituted 35 % of prison admissions (in absolute numbers about 200,000 offenders). She also expressed deep concern about the regional variations. In California, 66 % of offenders admitted to prisons were parole violators, whereas in Florida they counted only for 7 %

(Petersilia 2003, p. 239 f). This is another argument to get involved courts better and to restrict “the nearly unfettered discretion to revoke” of parole officers.

Statistics and Evaluation Research: Recidivism and Reintegration into Society

While at first glance it seems to be a rather easy task to count the numbers of those who are released early and those who fully serve their sentence, the database in Europe is rather limited. Many national statistics exhibit serious problems, which limit their validity and comparability. So, for example, German data sources indicate that around 30 % of prisoners are released early. However, the absolute numbers also contain fine defaulters who are not eligible for early release. The real figure therefore is more than twice as high (around 65 %, with large regional variations, see Dünkel 2013). Therefore, the numbers reported by the Council of Europe deserve particular reservation. They demonstrate, however, large variations in a comparative view, which in some cases are due to the legal conditions. The high proportion of early release in the Scandinavian countries (1999: 93 % in Denmark, 99 % in Finland, and 100 % in Sweden) is the result of quasi-automatic early release in these countries. On the other hand, in France the proportion of early release was only between 8 % and 14 % (see Tubex and Tournier 2003, 8 f). As in Scandinavia, Slovakia (83 %), Latvia (84 %), and Romania (86 %) have high early release rates, while the numbers given for Portugal (28 %), Scotland (29 %), Lithuania (51 %), and Spain (57 %) indicate a much more restrictive practice, which might partly explain the high prison population rates in these countries (see Dünkel 2013, § 38 notes 58 ff; Dünkel et al. 2010b; Dünkel 2013). In Austria, before the reform law of 2008 about 20 % of prisoners were released early. Since then, the proportion has increased to around 60 %. More recent data can be found in Padfield et al. (2010), which indicate further positive, but also negative developments. In Spain, from 1996 to 2008 early release rate declined to 22 % on average, and so was cut down to less than half compared to the 1990s. Also in Slovenia, a reduction from 46 % to only 32 % could be observed since the year 2000. In 2007, in Belgium, 94 % of prisoners were granted early release (a result of the quasi-automatic strategy), in Finland as in earlier years 99 %.

The situation in the USA is not clear as the number of releases on parole is not given in percentages of total releases. The sourcebook of criminal justice statistics gives longitudinal data of the population in jails and prisons and of persons under probation and parole (see http://www.albany.edu/sourcebook/pdf/t612010.pdf). The absolute numbers of the total correctional population since 1980 increased from 1.8 million to 7.1 million persons in 2010. The number of those in jail or prison increased from about 502,000 to 2,266,832 in 2010. In the same time span, the number of parolees increased from 220,000 to almost 841,000. In 2010, almost 709,000 sentenced prisoners were released from state and federal prisons, that is, about 44 % related to the daily prison population of 1.6 million prisoners in these prisons. (Another approximately nine million persons were released from jails; daily population there was about 749,000 in 2010). As stated by Petersilia (1999, 2003), the flow of entries and releases is remarkable and demonstrates the challenges for the probation service.

Recidivism rates are rather high: 47 % of prisoners released in 1983 and the same proportion of those released in 1994 were reconvicted within 3 years; however, only 25 % returned to prison (see http://www.albany.edu/sourcebook/toc_6.html). On the other hand, data from Petersilia (1999, 2003) indicate that about half of parolees fail to complete parole successfully and that their return to prisons at the end of the 1990s represented about one third of incoming prisoners (see above). It is also evident that the conditions imposed are more and more of a repressive nature, which makes failure almost inevitable. The probation services and aftercare services of the voluntary sector are often unable to deliver the necessary support, which contributes to the high failure rate (see the contributions in Hucklesby and Hagley-Dickinson 2007).

In Germany, statistical data about the recidivism of probationers are available with regard to two large statistical studies organized by the Federal Ministry of Justice. They cover all offenders sentenced or released in 1994 ($n = 947{,}000$) and in 2004 ($n = 1{,}045{,}000$; see Jehle et al. 2003, 2010).

The comparison of recidivism rates according to different sanctions does not allow the effectiveness of these sanctions to be assessed (selection problem).

However, within the same sentence category the 10-year follow-up gives an indication for the improved effectiveness of probation and parole (as well as of the treatment within prison, see Table 1).

Those sentenced or released, who have been put under supervision of the probation service, show remarkably low reoffending rates compared to those who (fully) served their prison sentence: The reincarceration rate of probationers was only 12 % compared to 21 % of the ex-prisoners group without supervision by the probation service. Looking at the groups of offenders released early with supervision by a probation officer, the reincarceration rate was only 14 % (own calculations according to Jehle et al. 2010, p. 186), although this group is deemed to be more at risk of reoffending than the group without supervision (see s. 57(3) German Criminal Law).

Again an estimation of "success" is difficult, but it can be shown that revocation rates of those under supervision have considerably decreased since 1970: The revocation rate dropped from 55 % in 1970 to 30 % in 2008, the revocation rate for probationers in the field of juvenile justice (14–21-year-old offenders) dropped from 43 % (1970) to 25 % in 2008 (see *Statistisches Bundesamt*, 2011, p. 17).

There is another "success-story" of the probation service in Germany: Originally, the probation service had been dedicated to deal with low-risk and first-time offenders. However, practice has changed and more and more high-risk offenders are being dealt by the probation service: The proportion of probationers who had already been under the supervision of the probation service increased from 59 % in 1963 to 85 % in 1990. Against the expectation that high-risk offenders would contribute to a decrease of the success rates, the opposite was the case! Success rates (no revocation) increased from 1963 to 1990 particularly for recidivist offenders (see Table 2).

Early Release from Prison, Table 2 Success rates (no revocation) of different groups of probationers in Germany

Offender group/year	1963 (%)	1990 (%)	Difference (%)
First-time offenders	70	84	+ 14
Probationers with prior convictions	45	69	+ 24
Of them: previously under supervision of the probation service	39	65	+ 26
All cases	55	71	+ 16

The interesting fact from Table 2 is that the groups with an a priori higher risk (those with prior convictions and with prior supervision by the probation service) profited the most by the liberalization of the scope of probation and early release: Their success rates improved by 24 and 26 % points compared to "only" 14 % for the group of first-time offenders. Excluding this problematic group from probation and parole would have resulted in an increase of the prison population (which has been successfully avoided in Germany since the late 1960s, see Dünkel and Morgenstern, in Dünkel et al. 2010a) and of the general recidivism rates.

The question whether early release reduces reoffending compared to fully serving the sentence remains a major methodological problem. There are a few empirical findings that underline the positive potential of early release (see for France Kensey 2004, 2007, p. 213 ff). Methodologically the problem is that those prisoners, who are released early, in contrast to those who serve their sentences in full, are a positively selected group and there is no immediate comparator. However, a point of comparison can be found in regional and temporal variations in approach. Thus in some Federal States or local districts prisoners are released, whereas in neighboring states or districts similar prisoners are not released, but, because of the application of more restrictive standards or traditions, they have to serve their terms in full. Also, over a period of time the criteria for release may change, so that one can – insofar as other conditions remain constant – make comparisons.

To this extent, studies in which there are control groups of some kind are more significant. The investigation by Dünkel in Berlin-Tegel examined for the first time the legal and social biographical characteristics (previous convictions, type of offense, age) of offenders in order to identify comparable groups, which in one case were released early and in the other served their full prison terms. The study showed that the former had a 13 % lower recidivism rate in the sense of a further sentence of imprisonment (see, in summary, Dünkel 2013, § 57 side notes 131 ff). Additional, albeit minor, differences were also apparent, depending on whether the released offenders were placed under the supervision of a probation officer or not. This research is confirmed by other German, sample-based studies. The German studies, notwithstanding their methodological shortcomings, are confirmed by research findings in other European countries that show the advantages of conditional release in terms of special prevention on the one hand, while not finding, on the other hand, any identifiable loss of security to the public as a result of extensive early release practices.

Kensey and Tournier in France showed lower rates of recidivism among those released conditionally compared to those who served their whole sentence in prison (−13 %, see Kensey and Tournier 2005). This applies in particular to first offenders imprisoned for the first time, of which 44 %, as compared to 77 % in the case of those serving their full sentences, were reconvicted.

Kensey found in another study that of those who were released conditionally, 23 % were sentenced to an unsuspended term of imprisonment again, while of those who served their full sentences, the share was 40 % (a difference of 17 %). Keeping the factors of age, type of offense, and previous convictions constant revealed a difference of 8 % in favor of the conditionally released group (see Kensey 2007, pp. 213 ff, 218). If further differentiation was made in terms of type of offense, the ratios in favor of early release were as follows: 11 % difference for theft, 25 % for robbery, 26 % for assault, and 22 % for rape. Overall the studies conducted by Kensey, like the study by Dünkel mentioned above, make it clear that the likelihood that those who are conditionally released will remain within the law is to some extent an effect of selection. Nevertheless, it remains the case that conditional release combined with subsequent assistance for and control of the offender by the probation service has a noteworthy net effect. The secondary Anglo-American meta-analyses confirm this trend (MacKenzie in Sherman et al. 1998 and in this volume; Petersilia 2003, p. 246 f; Harper and Chitty 2004). This is plausible insofar as offenders released early receive more support from the probation service, whereas the others sometimes lack any support and help by aftercare services. Raynor (2012), in summarizing the results of the end-to-end offender management support schemes, comes to the conclusion that these are more efficient than traditional, divided resettlement programs.

Finally it is worth noting that criminological research has made aware that early release may in practice be refused on grounds which are not necessarily justified by "objective" risk assessments. Thus, for example, foreigners may be refused early release not because of the risk they pose but because their early release is considered impracticable. Within the European Union, this may eventually be resolved to some extent by the new Framework Directive on mutual recognition in relation to "judgments and probation decisions with a view to the supervision of probation measures and alternative sanctions," which should allow conditional release that begun in one country to be continued in another. However, this Directive will only address a small share of the problems faced by foreigners in European prison systems. As the entry on Belgium makes clear, foreigners also often face additional procedural hurdles and exclusion from certain reintegration activities and language or cultural problems when preparing for release. The same issues arise in several other European countries, too. Also, ethnic minorities, even when they are not foreign citizens, may face similar issues. The detailed European provisions on nondiscrimination, which are specifically applied to all aspects of imprisonment by the European Prison Rules, should be followed strictly in respect of early release and the process lending up to it, to ensure that shortcomings are avoided.

Outlook: Extension of the Use of Early Release to Reduce Overcrowding, Increase Security, and Increase Reintegration of Offenders

The common problem of prison overcrowding that is found in many countries compels to consider penal policy initiatives to limit or reduce rates of imprisonment. Such initiatives can adopt either "front door" or "back door" strategies, in which either the number of sentenced prisoners is reduced (by increasing the use of alternatives such as fines, community service, or probation (or in terms of continental European systems, suspended sentences)) or the number of early releases is increased (see Dünkel 2013, § 38 side note 66 for further references). Regarding early release, in the light of both international practice and empirical criminological research, one could recommend that practice of releasing prisoners after they have served half the sentence should be expanded, and that "neutral" prognoses should be reduced by making release the rule, rather than an exception requiring justification. Nevertheless, in cases of very serious offenses against life or limb or sexual autonomy, one would, even where the recidivism rate was a priori relatively low, require an individual prognosis – notwithstanding its relative unreliability.

Allowing more or less automatic release after half or two thirds of the sentence seems to be more appropriate and has been followed in several countries. In this context, step-by-step strategies should be tried, as they systematically introduce relaxations of the prison regime (such as open prisons and temporary release) which make correct prognoses more likely and create a social space into which the offender is more likely to be received and integrated successfully.

Restrictions on purely general preventive grounds are not justified by international comparisons and standards. Criteria for conditional release (insofar as the system does not operate automatically in favor of the prisoner) should be determined solely on grounds of special prevention (rehabilitation).

In summary, it can be concluded that, in the light of the current state of empirical research on the effectiveness of sentences and of the prognoses of future behavior, it would be appropriate to extend the use of automatic, or almost automatic, early release instead of conditional release which relies on individual prognoses. Such quasi-automatic release has the additional advantage that the supportive measures taken by the probation service and the positive obligation placed on the offender can create a special preventive framework that is conducive to reintegration. In this regard, the pressure of overcrowding that can be found in so many countries can be seen as an opportunity for a rational and humane penal policy. A European "model" with release at the halfway point for those receiving determinate sentences, subject to supervision, would be a realistic option, although many obstacles of trends toward more punitiveness in some countries have to be faced.

Once it has been accepted that the "legitimate requirements of the sentence" entail reintegration, questions may be asked as to whether a term of imprisonment that jeopardizes that aim (by not providing for the possibility of release) is not in itself capable of constituting inhuman and degrading treatment. Therefore, from a (European) human rights perspective "life without parole" would be seen as an inhuman punishment violating human dignity.

This emphasis on reintegration as an element of European human rights law may stimulate further forms of release that are designed to better achieve this goal.

In their 2003 Recommendation concerning Conditional Release, the Committee of Ministers of the Council of Europe provided in some detail not only for better statistics on conditional release but also for wider research designed to "obtain more knowledge about the appropriateness of existing conditional release systems and their further development" (§ 40). What is urgently needed is a more enlightened policy-making on early release across Europe and other continents suffering from the same problems of prison overcrowding and a lack of a systematic and evidence-based resettlement policy.

Related Entries

- ► Electronic Monitoring
- ► Examining the Effectiveness of Correctional Interventions
- ► Probation and Community Sanctions

Recommended Reading and References

Allen H, Eskridge C, Latessa E, Vito G (1985) Probation and parole in America. McMillan, Inc, New York

Andrews DA et al (1990) Does correctional treatment work? A clinically relevant and psychologically informed metaanalysis. Criminology 28:369–404

Council of Europe (2009) European rules for juvenile offenders subject to sanctions or measures. Council of Europe Publishing, Strasbourg

Dünkel F (2013) Kommentierung §§ 38, 57 and 57a StG-B. In: Kindhäuser U, Neumann U, Paeffgen H-U (eds) Nomos Kommentar zum Strafgesetzbuch, vol 1, 4th edn. Nomos Verlag, Baden-Baden

Dünkel F, Ganz G (1985) Kriterien der richterlichen Entscheidung bei der Strafrestaussetzung nach § 57 StGB. Monatsschrift für Kriminologie und Strafrechtsreform 68:157–175

Dünkel F, Spieß G (eds) (1983) Alternativen zur Freiheitsstrafe. Max-Planck-Institut für ausländisches und internationales Strafrecht, Freiburg im Breisgau

Dünkel F, Drenkhahn K, Morgenstern C (eds) (2008) Humanisierung des Strafvollzugs. Konzepte und Praxismodelle. Mönchengladbach, Forum Verlag Godesberg

Dünkel F, Lappi-Seppälä T, Morgenstern C, van Zyl Smit D (eds) (2010a) Kriminalität, Kriminalpolitik, strafrechtliche Sanktionspraxis und Gefangenenraten im europäischen Vergleich, vol 1 and 2. Forum Verlag Godesberg, Mönchengladbach

Dünkel F, van Zyl Smit D, Padfield N (2010b) Concluding thoughts. In: Padfield N, van Zyl Smit D, Dünkel F (eds) Release from prison. European policy and practice. Willan Publishing, Cullompton, pp 395–444

Dünkel F, Grzywa J, Horsfield P, Pruin I (eds) (2011) Juvenile justice systems in Europe: current situation and reform developments, vol 1–4, 2nd edn. Forum Verlag Godesberg, Mönchengladbach

Garland D (2001) The culture of control. Crime and social order in contemporary society. University of Chicago Press, Chicago

Harper G, Chitty C (2004) The impact of corrections on re-offending: a review of 'what works'. Home Office HORS, London, 291

Hucklesby A, Hagley-Dickinson L (2007) Prisoner resettlement: policy and practice. Willan, Cullompton

Jehle J-M, Heinz W, Sutterer P (2003) Legalbewährung nach strafrechtlichen Sanktionen. Eine kommentierte Rückfallstatistik. Mönchengladbach, Forum Verlag Godesberg

Jehle J-M et al (2010) Legalbewährung nach strafrechtlichen Sanktionen. Mönchengladbach, Forum Verlag Godesberg

Kensey A (2004) Conditional release and the prevention of re-offending. In: Council of Europe (Ed) Crime and Crime Policy in Europe. Strasbourg, Council of Europe Publishing, pp 181–186

Kensey A (2007) Prison et récidive. Armand Collin, Paris

Kensey A, Tournier PV (2005) Prisonniers du passé? Cohorte des personnes condamnées, libérées en 1996–1997: examen de leur casier judiciaire 5 ans après la levée d'écrou (échantillon national aléatoire stratifié selon l'infraction). Paris, Ministère de la Justice, Direction de l'administration pénitentiaire, Coll. Travaux & Documents, no 68

Lappi-Seppälä T (2007) Penal policy in Scandinavia. In: Tonry M (ed) Crime, punishment, and politics in comparative perspective, vol 36, Crime and justice. University of Chicago Press, Chicago, pp 217–295

Lösel F (2012) Offender treatment and rehabilitation: what works? In: Maguire M, Morgan R, Reiner R (eds) The Oxford handbook of criminology, 5th edn. Oxford University Press, Oxford, pp 986–1029

Moore R et al (2006) Managing persistent and serious offenders in the community. Intensive community programmes in theory and practice. Cullompton, Willan

Nelken D (2009) Comparative criminal justice: beyond ethnocentrism and relativism. Eur J Crim 6:291–312

Padfield N (2002) Beyond the tariff: human rights and the release of life sentence prisoners. Willan, Cullompton

Padfield N, van Zyl Smit D, Dünkel F (eds) (2010) Release from prison. European policy and practice. Willan Publishing, Cullompton

Petersilia J (1999) Parole and prisoners reentry in the United States. In: Tonry M, Petersilia J (eds) Prisons, vol 26, Crime and justice, pp 479–529

Petersilia J (2003) When prisoners come home: parole and prisoner reentry. Oxford University Press, New York

Raynor P (2012) Community penalties, probation, and offender management. In: Maguire M, Morgan R, Reiner R (eds) The Oxford handbook of criminology, 5th edn. Oxford University Press, Oxford, pp 928–954

Robinson G (2005) What works in offender management? Howard Journal 44:307–318

Sherman LW et al (1998) Preventing crime, what works, what doesn't, what's promising? US Department of Justice, Office of Justice Programs, National Institute of Justice, Washington, DC, https://www.ncjrs.gov/works/

Snacken S, Dumortier E (eds) (2012) Resisting punitiveness in Europe? Welfare, human rights and democracy. Routledge, London

Statistisches Bundesamt (2011) (ed) Bewährungshilfe 2008. Wiesbaden, Statistisches Bundesamt

Tonry M (2004) Punishment and politics. Willan Publishing, Cullompton

Tubex H, Tournier PV (2003) Study of conditional release (parole) in the member states, analysis of replies to the general questionnaire. Council for Penological Co-operation, Strasbourg, PC-CP (2000) 24 Rev 4 (restricted)

van Kalmthout AM, Durnescu I (eds) (2008) Probation in Europe. Legal Publishers (WLP), Nijmegen

van Zyl Smit D, Dünkel F (eds) (2001) Imprisonment today and tomorrow: international perspectives on prisoners' rights and prison conditions, 2nd ed. Kluwer, Deventer/Boston

Wirth W (2009) Aus der Haft in Arbeit und Ausbildung. Bewährungshilfe 56:165–178

EBP

▶ Evidence-Based Policy in Crime and Justice

Eco-crime

▶ History of Green Criminology

Ecology of Crime

▶ Geography of Crime and Disorder

Econometrics of Crime

Giovanni Mastrobuoni[1] and Paolo Pinotti[2]
[1]Collegio Carlo Alberto, IZA, Netspar, Moncalieri, Italy
[2]Università Bocconi, Milan, Italy

Overview

The maxim "correlation is not equal causation" has by now been fully adopted by empirical researchers in the social sciences, including criminology. Correlations over time or across regions between crime rates and different measures of general deterrence are generally positive. These measures include the level of police forces, or prison population, or the severity of sanctions. This does not mean that higher numbers of police officers or an increased prison population lead to an increase in crime. It simply means that policy makers are constantly trying to reduce crime by altering deterrence based on the level of crime. Econometricians would say that the different measures of deterrence are endogenous (depend on crime) as opposed to exogenous or that there is a reverse causality problem. The theory of deterrence predicts that criminals would adjust their criminal behavior depending on the certainly and the severity of the punishment. Deterrence can be of two sorts. Specific deterrence refers to individual responses to actual punishments, while general deterrence refers to individual responses to expected punishments. The estimation of both types of deterrence is subject to endogeneity issues.

Inmates are usually not "randomly" assigned to different sanctions (sentence enhancements, prison conditions, etc.), but rather assigned to them based on their criminal history. And the criminal history is likely to influence future criminal behavior. If one was to regress criminal behavior on the type of punishment received without taking the endogeneity into account, a positive correlation between punishments and crime might stem from an omitted variable bias, where the omitted variable is the underlying individual criminal attitude.

This entry highlights different methods that have been proposed to overcome the endogeneity issue and, more generally, different econometric strategies that have been proposed to identify the causal effect of deterrence on crime.

Regression Models

The advances in econometrics that are going to be discussed can be found in many econometric textbooks. Useful econometric textbooks include Wooldridge (2010), Peracchi (2001), Angrist and Pischke (2008), and Cameron and Trivedi (2005). The purpose of this entry is to highlight how these methods can be used in the economics of crime literature, with a special emphasis on estimating deterrence effects. The focus is going to be on methods that use microlevel data, as opposed to time-series data. In recent years fine microlevel data has become available, and it fuels the most innovative research on deterrence. For simplicity an additional focus is on linear models, though many of the issues that are discussed apply to nonlinear models as well.

The Econometric Model

An econometric model of deterrence typically starts with the estimation of a linear model

$$y_{it} = \alpha + \beta_1 x_{1.it} + \beta_2 x_{2,it} + \cdots + \beta_k x_{k,it} + \varepsilon_{it},$$

where the dependent variable measures crime or recidivism and the independent variable measures the severity or the certainty of punishment.

This model has two possible interpretations:

1. Best linear predictor of Y
2. Conditional mean of Y given X

In order to get unbiased estimates of $\beta = (\beta_1 \ldots \beta_k)$, one needs the measures of deterrence to be uncorrelated with the error term (in matrix notation $E(x_k \varepsilon) = 0$ for $k = 1, \ldots, K$). According to the Gauss-Markov Theorem, if the rank condition is satisfied $rank\ (x) = k$ and the variance of ε_i is σ^2 (homoscedasticity), then ordinary least squares (OLS) coefficients are the best linear unbiased estimators (BLUE). This means that the OLS estimator is the one with the highest precision (the lowest standard errors) among linear estimators.

Violations of Ideal Conditions

The assumptions made by the Gauss-Markov Theorem can be violated by collinearity, omitted variable bias, simultaneity bias, measurement error, and heteroscedasticity.

Collinearity

When the explanatory variables are linearly dependent, for example, $y = bx + az + e$ and $x = wz$, then $y = (a + bw)z + e$, so any linear combination of a and b is a solution to the OLS. It is always a good idea to look at the correlation matrix of all regressors. Whenever there is multicollinearity, a simple solution might be to reduce the number of explanatory variables, unless omitting variables causes biased parameter estimates.

Omitted Variable Bias

Suppose the true model is $Y = X\beta + Z\gamma + \varepsilon$, but one estimates $Y = X\beta + u$, where $u = Z\gamma + \varepsilon$. In this case $\beta = E(X'X^{-1}X'Y) + E(X'X^{-1}X'Z)\gamma + 0$. The bias is given by $E(X'X^{-1}X'Z)\gamma$. Notice that the bias disappears if the correlation between X and Z is zero ($X'Z = 0$) or if the correlation between Y and Z is zero ($\gamma = 0$).

One example would be a recidivism (y) regression under alternative sanctions (x_i):

$$r_i = x_i\beta + \gamma z_i + \varepsilon_i$$

where the omitted variable is unobserved criminal attitude (z). More criminally prone people might be more likely to recidivate and at the same time be less likely to receive alternative sanctions. How does one solve this issue? Notice that this bias is sometimes called **selection bias** and other times it is called **endogeneity bias**.

Measurement Error

One way to deal with omitted variables is to use proxies (variables that measure, even imprecisely, the excluded variable), in the previous example past criminal history, psychological reports, etc. However, proxies introduce a measurement error bias. If by adding decent proxies the coefficient β does not vary by much, one can be more confident about the OLS results.

Heteroscedasticity

The OLS estimator assumes that all the errors, ε_{it}, are identically distributed with the same variance for each unit of analysis (IID assumption). Based on this, it estimates the variance of the β coefficient. Since the condition rarely holds, the estimated standard errors of the estimated coefficients are different than the real ones. Since significance is based on the size of the standard error relative to the coefficient, this leads to an error in inference. Huber (1967) and White (1980)'s observation led to the Huber-White Sandwich estimator for the standard errors.

A similar issue arises when several units of observation can be grouped and their outcomes are correlated within the group. Taking into account this requires a clustering of the standard errors at the group level; see Moulton (1986).

Fixed Effect Estimation

If the omitted variable is believed to be fixed across one dimension (time or space) and panel data (across time or space) are available, one can difference away the unobserved variable. Assume the model is

$$y_{it} = \alpha + \beta x_{it} + c_i + \gamma_t + \varepsilon_{it}$$

With t and i fixed effects, the identification comes from changes within i and t. Forgetting about the t effects and averaging the model over t, one gets

$$\bar{y}_i = \alpha + \beta \bar{x}_i + c_i + \bar{\varepsilon}_i.$$

With differencing one gets

$$y_{it} - \bar{y}_i = \beta\left(x_{it} - \bar{x}_i\right) + \varepsilon_{it} - \bar{\varepsilon}_i,$$

which can be estimated by OLS. The alternative is to add individual fixed effects in the regression.

Another popular estimator that eliminates individual fixed effects is the first difference (FD) estimator:

$$y_{it} - y_{i,t-1} = \beta\left(x_{it} - x_{i,t-1}\right) + \varepsilon_{it} - \varepsilon_{i,t-1}.$$

One can show that the two estimators are the same when the total number of periods (T) is equal to two. When $T \neq 2$, whether one or the other estimators are preferred depends on the speed at which one thinks that x affects y.

Up until now one implicitly assumes fixed effects (intercepts) that might vary across individuals and marginal effect of X on Y (slopes) that were the same across individuals. But there are a variety of random effect/coefficient models that exploit distributional assumptions to estimate more complicated panel models, like

$$y_{it} = \beta'_i x_{it} + \varepsilon_{it}.$$

The Counterfactual

Before introducing individual-specific coefficients in more detail and to understand the potential endogeneity issue, it is often useful to think in terms of counterfactuals, especially when a policy is discrete or binary (0 or 1). Suppose one is interested in estimating whether more police reduces crime. Ideally one would increase the number of police officers, measure crime rates, and compare them to what would have happened had there not been an increase. The issue is that one does not observe this counterfactual. If Y_i measures crime rates in region i and T_i measures the treatment of an increase in police forces, one only observes $Y_i = Y_{i0}(1 - T_i) + Y_{i1}T_i$ meaning the crime rates in treated regions when treated and the crime rates in control regions when not treated.

Taking the simple average between treated and non-treated units

$$\begin{aligned} E[Y_i|T=1] - E[Y_i|T=0] &= E[Y_{1i}|T=1] - E[Y_{0i}|T=0] \\ &= E[Y_{1i} - Y_{0i}|T=1] + E[Y_{0i}|T=1] - E[Y_{0i}|T=0], \end{aligned}$$

the first difference $E[Y_{1i} - Y_{0i}|T = 1]$ measures the average treatment effect on the treated

(ATET), and the second difference $E[Y_{0i}|T = 1] - E[Y_{0i}|T = 0]$ represents the selection, meaning the difference in outcomes between treated and non-treated regions in the absence of any treatment.

The identification of the treatment effect $Y_{i1} - Y_{i0}$, the change in crime for the same region with and without more police, is all about finding a reasonable counterfactual, a control group, that eliminates the selection. Whenever there is random assignment of treatment, the treatment and the control group are statistically equivalent, and the simple difference in outcomes between the two groups $E[Y|T = 1] - E[Y|T = 0] = E[Y_1 - Y_0]$, is going to equal the average treatment effect (ATE). Randomized controlled experiments are rare in the economics of crime literature. Treatment is generally not randomly assigned due to cost or legal/moral objections, but rather assigned based on a policy or a change in policy. These policies often lead to discontinuities in treatment and can be seen as "quasi-experiments."

But even if they were randomly assigned in such experiments, several problems can arise: (1) cost. These experiments tend to be very costly, which might reduce the size of the experimental group. (2) There might be noncompliance. Treated individuals/regions might decide not to get treated and the non-treated to get treated. (3) There might be nonrandom attrition among individuals when experimental units are observed over time. (4) The experiments might not last enough time to see a treatment effect. (5) The results of the experiment might not be generalizable (limited external validity). (6) There might be general equilibrium effects that would not be captured by an experiment.

Some of these limitations do not arise in "quasi-natural experiments."

Quasi-Natural Experiments

Natural experiments are usually policy reforms or specific rules that can be exploited to measure exogenous changes in x. These interventions often change the treatment status of marginal participants, that is, those to whom treatment will be extended. Whenever treatment effects are not constant across participants, the treatment effects estimated on the marginal participants are called local average treatment effect (LATE) and may be very different from the treatment effects if the interventions had to be extended to the entire population. For example, incapacitation effects of incarceration are usually estimated based on the release of marginal prisoners (see Levitt (1996) and Barbarino and Mastrobuoni (forthcoming)). It would probably be incorrect to extend such estimates to the entire prison population, as marginal inmates should be less dangerous than inmates who keep on staying in prison. See Imbens and Angrist (1994) and Angrist et al. (1993), and see Heckman and Vytlacil (2005) for further discussion.

Depending on the "quasi-experiment" under study, there are different empirical methodologies that have been proposed to solve the endogeneity issues. The most commonly used are instrumental variables, regression discontinues, difference in differences, and reweighting techniques (propensity score and matching).

The Simple Difference

Suppose an outcome, for example, crime, $y_{it} = \alpha + \beta P_{it} + \varepsilon_{it}$ depends on expected sanctions and that due to a policy reform, there is an across the board reduction in sanctions at time τ, such that $P_{it} = 1$ if $t \geq \tau$ and 0 otherwise. In this case, $\beta = E(Y|P = 1) - E(Y|P = 0)$ measures the change in recidivism. If there are only two periods available, before ($P = 0$) and after ($P = 1$) the reform, the main assumption is that the only changes that happen over time are due to the reform. If more time periods are available, one can compare the evolution before and after the reform. One can also go beyond just analyzing the differences in means and look at quantiles. The main issue of such simple differences is that the changes might be due to other factors or that the reform was due to changes or expected changes in crime. In such case the reform variation would be endogenous.

The Matching and the Propensity Score Estimators

Whenever simple differences are the only available option, treatment and control groups need to be as similar as possible; otherwise simple differences might just stem from these differences and not depend on the treatment status. Whenever one thinks that these differences are driven by observable traits as opposed to unobservable characteristics, one can simply control for such factors (in this case we have selection on observables x):

$$y_{it} = \alpha + \beta P_{it} + \gamma x_{it} + \varepsilon_{it}$$

An alternative is to compare individuals in the treatment group to individuals in the control group that have similar observable characteristics before treatment. Similarity can be defined based on the xs directly (e.g., nearest neighbor matching) or indirectly through likelihood of receiving treatment. The idea is that comparing individuals in the treatment and the control group that have the same estimated propensity to be treated should solve the selection issue. Once the likelihood of treatment (called propensity score) has been estimated, typically using logit or probit, there are several ways to match individuals in the two groups. Given the numerosity of matching procedures, it is important to show how robust the results are to the method used. In matching strategies, one pairs an individual in the treatment to one (one to one) or more (one to many) individuals in the control group, takes the difference in outcomes, and then averages across all treated individuals. For a discussion on efficiency, see Hirano et al. (2003). For an overview of matching estimators, see Smith and Todd (2001).

For such pairing to be even feasible, one needs a common support, that is, the range of propensities to be treated has to be the same for treated and control cases. This can be checked graphically by plotting the densities of the propensity scores by treatment status. If the overlap is not complete, one needs to limit the analysis to the smallest connected area of common support.

The propensity score can also be used to reweight observations in the control group such that their observable characteristics resemble on average the ones of the treatment group. Following Hirano et al. (2003), one can adjust for differences between the two groups by weighting observations by the (inverse) propensity score of assignment, that is, the probability of belonging to each group conditional on the observed covariates. Specifically, one can weight each unit by

$$T_i \frac{1}{P(X_i)} + (1 - T_i) \frac{1}{1 - P(X_i)},$$

where T_i is a dummy equal to 1 if the ith individual has been treated 0 otherwise and $P(X_i)$ is the propensity score, meaning the probability of belonging to the treatment group conditional on the vector of individual characteristics X_i. The next section elaborates more on this approach.

The Synthetic Control Approach

Due to the increasing availability of micro data on victims and offenders, the matching and difference-in-differences estimators mentioned above often exploit variation over large samples of cross-sectional units observed over a few periods of time. In some cases, however, individual-level data are not available. Moreover, important phenomena such as organized crime, drug trafficking, and corruption involve significant spillovers between individuals within the same local area. Finally, most crime-prevention policies vary only at the aggregate level.

In all these cases, the available data usually include a few units (e.g., states or regions) observed over longer time periods. Synthetic control estimators provide a useful tool in this context because they match treated and control units on both the longitudinal and cross-sectional dimensions. The idea is to compare (difference in differences) each treated unit with a weighted average of control units that replicates the initial conditions in the treated unit many years before treatment.

This method was originally devised by Abadie and Gardeazabal (2003) to estimate the economic costs of terrorism in the Basque Country. They do so by comparing the GDP per capita of a region, Y_{it} with a weighted average of the GDP per capita in the other Spanish regions not affected by the conflict:

$$\hat{\beta}_t = Y_{it} - \sum_{j \neq i} w_j Y_{jt}$$

where w_j is the weight attached to each jth region in the control group.

Following the notation of previous sections, $Y_{it} = T_{it}Y_{1it} + (1 - T_{it})Y_{0it}$, where Y_{1it} and Y_{0it} is the potential GDP per capita with and without terrorism, respectively, and T_{it} is a treatment indicator for the presence of terrorism. After the outbreak of terrorism in the Basque Country, $T_{it} = 1$ and $T_{jt} = 0 \ \forall j \neq i$, so that

$$\hat{\beta}_t = (Y_{1it} - Y_{0it}) + \left(Y_{0it} - \sum_{j \neq i} w_j Y_{0jt} \right), \quad (2)$$

where the first term on the right-hand side is the treatment effect of terrorism in region i and year t. The precision of $\hat{\beta}_t$ depends thus on the difference between the synthetic control and the (unobserved) development of the treated region in the absence of the treatment (the second term on the right-hand side), so the estimation problem amounts to choosing the vector of weights that minimizes the last difference on the right-hand side of equation (2).

A natural choice consists in minimizing the same difference over the previous period, in which neither region had been treated (i.e., $T_{it} = T_{jt} = 0$). As long as the weights reflect structural parameters that would not vary in the absence of treatment, the synthetic control provides a counterfactual scenario for the potential outcome Y_{0it}. Notice that an analogous identifying assumption, namely, that unobserved differences between treated and non-treated units are time invariant, is routinely imposed on difference-in-differences models. Indeed, Abadie et al. (2010) show that synthetic control methods generalize the latter by allowing the effect of unobserved confounders to vary over time according to a flexible factor representation of the potential outcome:

$$Y_{1it} - \beta_t = Y_{0it}^0 = \delta_t + \Theta_t X_i + \lambda_t \mu_i + \varepsilon_{it},$$

where δ_t is an unknown common factor with constant factor loadings across units, X_i is a $(r \times 1)$ vector of observed covariates (not affected by the intervention), Θ_t is a $(1 \times r)$ vector of unknown parameters, λ_t is a $(1 \times F)$ vector of unobserved common factors, μ_i is an $(F \times 1)$ vector of unknown factor loadings, and the error terms ε_{it} are unobserved transitory shocks at the region level with zero mean. The traditional difference in differences imposes that λ_t is constant for all t's.

Turning to the choice of the minimand, Abadie and Gardeazabal (2003) adopt a two-step procedure that minimizes the distance between the treated region and the synthetic control both in terms of pretreatment outcomes during the 1950s and a vector X of initial conditions over the same period (GDP per capita, human capital, population density, and sectoral shares of value added). Conditional on a diagonal matrix with nonnegative entries measuring the relative importance of each predictor, V, the optimal vector of weights $W^*(V)$ solves

$$\min \left(X_i - \sum_{j \neq i} w_j X_j \right)' V \left(X_i - \sum_{j \neq i} w_j X_j \right)$$

subject to $w_i \geq 0, \ \forall i$ and $\sum_i w_i = 1$; then, the optimal V^* is chosen to minimize the mean squared error of pretreatment outcomes:

$$\frac{1}{N} \sum_t \left(Y_{it} - \sum_{j \neq i} w_j^* Y_{jt} \right)^2,$$

where N_t is the number of years in the sample before the outbreak of terrorism.

After Abadie and Gardeazabal (2003), synthetic control methods have been used to study a wide range of social phenomena, including crime-related issues. As to the effects of (different types of) crime, Pinotti (2011) quantifies the costs of organized crime in southern Italy.

From a methodological point of view, Abadie et al. (2010) provide a throughout presentation of synthetic control estimators, while Donald and Lang (2007) and Conley and Taber (2011) propose alternative methods for dealing with small numbers of treated and control units in difference-in-differences models.

The Difference in Difference

If a new reform, or a new sanctioning rule, affects only some individuals (e.g., those in a given age group) but not others, a more convincing estimation strategy is to compare the changes in outcomes of the two groups. As in an experiment one can call the two groups the treatment and the control group. Estimating the following regression by OLS,

$$y_{it} = \alpha + \beta_P P_{it} + \beta_T T_{it} + \beta_{TP} P_{it} \times T_{it} \times \varepsilon_{it},$$

where P defines the pre/post period and T the treatment group β_{TP} is called the difference-in-differences estimator. This estimator is equal to

$$\begin{aligned}\beta_{TP} = & E(Y|T=1, P=1) - E(Y|T=0, P=1) \\ & - E(Y|T=1, P=0) - E(Y|T=0, P=0).\end{aligned}$$

The identification assumption is that the interaction between the reform and the treatment group is only due to the treatment or, in other words, that in the absence of treatment, the two groups would have shown similar changes in outcomes (parallel trend assumption). Generally this is more plausible when the two groups are similar in terms of observable characteristics and outcomes before treatment. One can usually improve these simple double differences performing robustness checks for the estimate, for example, using different control groups, looking at the evolution of outcomes over time, and exploiting sharp changes in treatment assignment.

One can also use difference-in-differences estimators coupled with matching estimators to reduce the initial differences between treatment and control groups (see, e.g., Mastrobuoni and Pinotti (2012)).

Instrumental Variables

Whenever one cannot credibly control for selection, researchers often start looking for an “exogenous variation” in x. Exogenous variations are usually driven by external factors (i.e., policy reforms, discontinuities in deterrence across individuals, across regions, or over time) that can plausibly thought to be unrelated to the dependent variable, other than through their influence on deterrence. Consider the model, where Z is an instrument for X: $y = \alpha + \beta_X + \varepsilon$.

For example, in the relationship between crime and police, a valid instrumental variable would be correlated with police but not with crime. Levitt (2002) used the number of firefighters as an instrument. If a one-unit change in the number of firefighters z is associated with 0.5 more police officers on the streets s and with a 100 decrease in total crimes, it follows that a one-unit increase in police officers is associated with $\frac{-100}{0.5} = -200$ fewer crimes. Another instrument for police is the exogenous shift in policing resources that has been used in Draca et al. (2011).

The instrumental variable coefficient is $\beta_{IV} = \frac{\partial y/\partial z}{\partial x/\partial z}$ and can be estimated by two-stage least squares, regressing y on z and x on z and taking the ratio between the corresponding marginal effects $\hat{\beta}_{IV} = \frac{(z'z)^{-1} z'y}{(z'z)^{-1} z'x} = \frac{z'y}{z'x}$. This simple formula can be generalized to the case of multiple regressors and multiple endogenous variables.

If one has more instruments than endogenous variables, one can test for overidentification. The intuition is that the estimated errors û can be obtained excluding one instrument and then regressing it on the excluded instrument. Under the assumption H_0 that û is consistent, one can test the exogeneity of the excluded instrument. In other words, such test rests on the assumption that for each endogenous variable, at least one instrument is valid.

It is important to always test for endogeneity of the explanatory variables. This can easily be done by regressing Y on X and on the corresponding predicted error term from the first stage, where one regresses X on Z. The t-statistic of the coefficient on the predicted error that is supposed to capture the endogeneity of X can be used to test endogeneity. The same test can be used for nonlinear second-stage estimation techniques such as probit or logit.

A final test that should be part of any IV estimation is related to the weak instruments issue. Instruments need to be correlated with the endogenous regressors, but whenever such correlation is too weak, it has been shown that the IV estimates converge to the OLS ones. The test statistics tend to be quite complicated, but in practice a simple rule of thumb is to have an F-statistic of the excluded instruments in the first stage (X on Z) that is larger than 10, though 15 seems to be a safer choice (see Stock and Yogo (2005)).

A special IV case is when z is a binary variable, $(0, 1\ :)\frac{\partial y/\partial z}{\partial x/\partial z} = \frac{\bar{y}_1 - \bar{y}_0}{\bar{x}_1 - \bar{x}_0}$. This estimator is called the Wald estimator.

Lochner and Moretti (2004) estimate the effect of education on crime by instrumenting education with compulsory schooling laws.

The Regression Discontinuity Design

Another difference estimator that is considerably more convincing than the simple one discussed above makes use of discontinuities in the assignment to treatment variable based on a, so-called, running or assignment variable. A classical example in the crime literature is the paper on deterrence and incapacitation effects by McCrary and Lee (2009). They use age of offenders as their running variable and age 18 as the threshold value where more punitive criminal sanctions take place (the treatment).

An important assumption is that individuals cannot manipulate the running variable and therefore assignment to treatment (they certainly cannot stop aging). In general one can examine the density of observations of the assignment variable around the threshold to test for manipulation McCrary (2008).

The other assumption is that at the discontinuity of the running variable, other variables, which might also be unobserved, do not change discontinuously. This means that unlike for the IV case, the running variable is allowed to influence Y but only in a smooth manner. There are two types of regression discontinuity (RD) design. In the first one, called the sharp design, as individuals move above the threshold, treatment status moves from 0 to 1 for everybody. In the second one, the conditional mean of treatment jumps at the threshold (fuzzy design).

The sharp design LATE estimate is equal to the average change in outcome at the discontinuity $E(y^+) - E(y^-)$, where "+" and "−" indicate observations just above and below the threshold. If in the fuzzy design the change in treatment is $E(x^+) - E(x^-)$, the local Wald estimator is $[E(y^+) - E(y^-)]/[E(x^+) - E(x^-)]$. Since the estimator is defined at the threshold, the main issue is about how to use the data efficiently without introducing too much bias, where the bias comes from moving away from the threshold. The default option is to run a regression of Y on local polynomials allowing for a jump at the threshold. The estimates are going to depend on the bandwidth choice (see McCrary (2008) for a discussion on this).

An RD design should start with a visual inspection of the discontinuity in treatment assignment. If such discontinuity is not clearly visible, a formal test that estimates $[E(x^+) - E(x^-)]$ should be reported. After the RD has been estimated, there are several tests that one can run to test the robustness of the results. One is to estimate the change in outcomes at random placebo cutoff points, and around 95 % should not be significant. One can also test whether other variables that should not be affected by the treatment show discontinuities at the threshold. One should also show how robust the results are to the choice of different bandwidths.

See Hahn et al. (2001) for an overview of the use of RD, and the special issue of the Journal of Econometrics (Imbens and Lemieux 2008a), and, especially, the practical guide by Imbens and Lemieux (2008b). Lee and Card (2008) explore the case when the regression discontinuity model is not correctly specified.

Related Entries

- ▶ Deterrent Effect of Imprisonment
- ▶ Network Analysis in Criminology
- ▶ Offender Change in Treatment
- ▶ Offender Decision Making and Behavioral Economics
- ▶ Randomized Experiments in Criminology and Criminal Justice
- ▶ Rational Choice, Deterrence, and Crime: Sociological Contributions
- ▶ Recent Perspectives on the Regression Discontinuity Design
- ▶ Sample Selection Models

Recommended Reading and References

Abadie A, Gardeazabal J (2003) The economic costs of conflict: a case study of the Basque country. Am Econ Rev 93(1):113–132

Abadie A, Diamond A, Hainmueller J (2010) Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. J Am Stat Assoc 105(490):493–505

Angrist JD, Pischke J-S (2008) Mostly harmless econometrics: an empiricist's companion. Princeton University Press, Princeton

Angrist JD, Imbens GW, Rubin DB (1993) Identification of causal effects using instrumental variables. NBER Technical Working Papers 0136, National Bureau of Economic Research, June 1993

Barbarino A, Mastrobuoni G (forthcoming) The incapacitation effect of incarceration: Evidence from several Italian collective pardons. American Economic Journal: Economic Policy

Cameron C, Trivedi PK (2005) Microeconometrics. Number 9780521848053 in Cambridge Books. Cambridge University Press, 3 2005

Conley TG, Taber CR (2011) Inference with difference in differences with a small number of policy changes. Rev Econ Stat 93(1):113–125

Donald SG, Lang K (2007) Inference with difference-in-differences and other panel data. Rev Econ Stat 89(2):221–233

Draca M, Machin S, Witt R (2011) Panic on the streets of London: police, crime, and the July 2005 terror attacks. Am Econ Rev 101(5):2157–2181

Hahn J, Todd P, Van der Klaauw W (2001) Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica 69(1):201–209

Heckman JJ, Vytlacil E (2005) Structural equations, treatment effects, and econometric policy evaluation. Econometrica 73(3):669–738, 05 2005

Hirano K, Imbens GW, Ridder G (2003) Efficient estimation of average treatment effects using the estimated propensity score. Econometrica 71(4):1161–1189, 07 2003

Huber PJ (1967) The behavior of maximum likelihood estimates under nonstandard conditions. In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, vol 1. University of California Press, Berkeley, pp 221–233

Imbens G, Lemieux T (eds) (2008a) The regression discontinuity design: theory and applications, vol 142. J Econom (Special issue). Elsevier, Feb 2008a

Imbens GW, Angrist JD (1994) Identification and estimation of local average treatment effects. Econometrica 62(2):467–475

Imbens GW, Lemieux T (2008b) Regression discontinuity designs: a guide to practice. J Econom 142(2):615–635

Lee DS, Card D (2008) Regression discontinuity inference with specification error. J Econom 142(2):655–674

Levitt SD (1996) The effect of prison population size on crime rates: evidence from prison overcrowding litigation. Q J Econ 111(2):319–351

Levitt SD (2002) Using electoral cycles in police hiring to estimate the effects of police on crime: Reply. Am Econ Rev 92(4):1244–1250

Lochner L, Moretti E (2004) The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. Am Econ Rev 94(1):155–189

Mastrobuoni G, Pinotti P (2012) Legal status and the criminal activity of immigrants. Working Papers 052, "Carlo F. Dondena" Centre for Research on Social Dynamics (DONDENA), Università Commerciale Luigi Bocconi

McCrary J (2008) Manipulation of the running variable in the regression discontinuity design: a density test. J Econom 142(2):698–714

McCrary J, Lee DS (2009) The deterrence effect of prison: dynamic theory and evidence. Berkeley olin program in law & economics. Working paper series, Berkeley Olin Program in Law & Economics, July 2009

Moulton BR (1986) Random group effects and the precision of regression estimates. J Econom 32(3):385–397

Peracchi F (2001) Econometrics. John Wiley and Sons, Ltd., New York, NY

Pinotti P (2011) The economic costs of organized crime: evidence from southern Italy. Bank of Italy Working Papers

Smith JA, Todd PE (2001) Reconciling conflicting evidence on the performance of propensity-score matching methods. Am Econ Rev 91(2):112–118

Stock JH, Yogo M (2005) Testing for weak instruments in linear IV regression. In: Identification and Inference for Econometric Models: Essays in Honor of Thomas

J. Rothenberg, eds Stock JH, Andrews DWK Cambridge University Press, New York, pp 109–120
White H (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48:817–830
Wooldridge JM (2010) Econometric analysis of cross section and panel data, volume 1 of MIT Press Books. The MIT Press, June 2010

Economic Dominance

▶ Institutional Anomie Theory

Economic Theory of Criminal Behavior

Nuno Garoupa
College of Law, University of Illinois at Urbana-Champaign, Champaign, IL, USA

Overview

The economics of crime started with the seminal article by Nobel Laureate Gary Becker in 1968. He suggested the well-known high-fine-low-probability result, in a framework where rational criminals compare the benefit of violating the law with the possible cost (in terms of probability and severity of punishment). Legal economists have developed a theory of deterrence in the last 40 years to explain optimal punishment in multiple contexts. The main results as well as new insights provided by behavioral law and economics are reviewed.

Basic Model

The economic theory of criminal behavior is an application of the neoclassical theory of demand. Formalized by Nobel Laureate Gary Becker in 1968, it states that potential criminals are economically rational and respond significantly to the deterring incentives by the criminal justice system. They compare the gain from committing a crime with the expected cost, including the risk of punishment, the possibility of social stigma, and eventual psychological costs. If Becker's theory is correct, increasing the resources that society devotes to the arrest, conviction, and punishment of criminals may be the best policy prescription for reducing the amount, and social costs, of crime.

Consider an example. Suppose an individual considers violating speed limits. S/he should consider, on one side, the benefit from such behavior (arriving earlier to his/her destiny or the simple pleasure of driving above the speed limit) and, on the other side, the possible punishment. However, a fine for violating speed limits is not paid with certainty, although there is a certain probability that such behavior is detected and effectively punished. Furthermore, there might be other elements to be considered such as reputation (which could go both ways, depending on the social group of reference) as well as possible consequences (for example, loss of one's driving license at some point in the future). If the benefit is more important than the costs, this individual is expected to violate speed limits. If the benefit is not sufficient to compensate for the costs, the individual is deterred and therefore no violation of speed limits takes place.

Becker's seminal paper posits that criminals follow rational utility maximization choosing in conditions of risk. The model resolves around a simple equation:

$$EU = p\ U(b - f) + (1 - p)\ U(b) \qquad (1)$$

where EU stands for expected utility (more rigorously, a Von Neumann-Morgenstern expected utility function), U for the utility function, p for the probability of capture and punishment, f is the severity of punishment, b is the income if undetected, and $b–f$ is the income if detected. All incomes include psychic components such as fear, excitement, and pain which are assumed to be convertible to monetary equivalents.

In Eq. (1) crime is a risky activity which is reflected by the probability and severity of punishment. Technically, crime is a lottery that offers an income b with probability $1 - p$ and an income *b*–*f* with probability *p*. If the punishment were zero, there would not be any risk. Hence, it can be said that severity of punishment increases risk. If the probability of punishment were zero or one, there would not be any risk (either because the individual is never punished or always punished). Consequently, when the probability moves away from zero and from one (presumably closer to ½), crime is riskier. For risk-averse individuals, crime imposes a risk disutility since they would be willing to accept a certain punishment in return for avoiding the uncertainty of the crime lottery.

Becker differentiated (Eq. 1) with respect to *p* and *f*, concluding that the negative partial differentials show that certainty and severity of punishment deter crime. Will an individual accept the gamble? The answer depends upon the individual's attitude to risk as embodied by the utility function *U*, the sizes of income *b*, probability *p*, and severity *f* as well as the opportunity cost (which can be normalized to zero for convenience). Even if an individual is risk averse, s/he will accept the gamble if the gains are sufficiently large (*b* is significant while *p* and *f* are relatively low).

In order to optimally deter crime, a fundamental result is proposed by Becker (1968): the fine *f* should be maximal since it is a costless transfer whereas detection (as expressed by the probability *p*) is costly. The fine refers to an amount of resources that is lost by the convicted criminal and gained by the government. However, the probability reflects the investment on an operational criminal justice system that can be quite expensive. Naturally the fine should be taken to the highest possible amount and the probability a mere complement. In the limit, the fine would equalize the entire wealth of a convicted criminal while the probability would be minimal.

Consider again the example of speeding. The policy implication derived by Becker (1968) is that if a driver is caught above the speed limit, s/he should lose his/her entire accumulated wealth (including the right to future earnings). However, the probability of such event taking place should be close to zero. The police should, for example, choose randomly 1 h per year to check speed limits and those caught violating them should lose their entire wealth. Although the intuition for this insight is powerful (probability and severity of punishment matter but the latter is much cheaper than the former), the policy implication seems hardly realistic. Not surprisingly, legal economists have taken an interest on this puzzle.

Polinsky and Shavell (2001, 2007) have developed an original economic model to assess Becker's claim in a variety of situations. In fact, the standard model of the economics of law enforcement (assessing the extent to which sanctions should be maximal and other legal implications) is now the so-called Polinsky-Shavell framework. This model has been developed throughout a long series of papers, not necessarily with the aim of an established unified theory of law enforcement, but more as an economic assessment of multiple problems encountered by scholars and policymakers in law enforcement.

The Polinsky-Shavell model is based on a couple of important assumptions. First, individuals are rational utility maximizing agents choosing in conditions of risk. They compare the benefit from not complying with the law with the corresponding costs. Second, optimal enforcement of the law should be derived from the maximization of social welfare as the objective of public policy. The social welfare function is given by the sum of the benefits minus the costs for everyone. It is usually assumed the gains from noncompliance should be considered when determining efficient law enforcement. The exclusion of the gains from noncompliance might undermine efficiency implications since it is outside of the standard Kaldor-Hicks analysis (a certain legal policy is said to be more efficient if the winners can potentially compensate the losers).

One important conclusion by Polinsky and Shavell is that there are many circumstances where nonmaximal sanctions are efficient. In other words, albeit seminal, Becker's results apply in a very limited set of assumptions. Under richer assumptions, increasing the probability and decreasing the sanction could actually be the efficient policy. It can be said that the main results offer interesting insights concerning the probability and severity of punishment in a multitude of situations.

Consider the most basic version of Polinsky-Shavell's model. Risk-neutral individuals choose whether to commit an act that benefits the actor by b and harms society by h. This act is punished with a sanction f and a probability p. A risk-neutral individual compares the gain b with the expected sanction pf. Everyone with a benefit higher than the expected sanction is not deterred; everyone with a benefit lower than the expected sanction is deterred. From a social perspective, all acts that generate more harm than benefit should be deterred; all acts that generate more benefit than harm should not be deterred. Therefore, if the expected sanction equals the harm, the private decision by each individual is socially efficient. However, this result presupposes that the probability and the severity of punishment are costless.

Assume that the distribution of parties by type is described by a density function $g(b)$ with a cumulative distribution $G(b)$; the distribution is known by the policymaker. The benefit b lies in the interval $[0, B]$. The probability of punishment p is costly, whereas the sanction is not (a fine is the simplest example). The cost function for the probability is $x(p)$ satisfying standard convexity properties. The maximum feasible sanction is F, which can be interpreted as the maximum wealth of individuals.

Risk-neutral individuals commit an offense if and only if b is greater than the expected fine, pf. Given individuals' decision of being honest or dishonest, social welfare is given by the sum of benefits minus costs as derived in the Polinsky-Shavell's framework and as shown by Eq. 2:

$$W = \int_{pf}^{B} (b - h)g(b)db - x(p) \tag{2}$$

Social welfare is maximized in the probability and in the fine. Immediately recognizing that the probability is marginally costly whereas the fine is not, it can be concluded that the optimal fine is maximal. The optimal probability satisfies the following first-order condition:

$$W_p = (h - pF)g(pF)F - x_p = 0 \tag{3}$$

There are two important results to be highlighted in Eq. 3:

1. The optimal probability is based on the harm caused by the offense and not on the illegal gain generated by noncompliance; that is because the objective is efficient deterrence and not full deterrence (further discussion by Polinsky and Shavell 1994);
2. The optimal expected sanction is less than the harm caused by the offense (there is underdeterrence); that is because enforcement is costly. Hence, the so-called multiplier principle ($p = h/F$) is inefficient. The following expression is easily derived from rearranging Eq. 3:

$$pF = h - x_p/Fg(pF) \tag{4}$$

When punishment is costly, the results are a bit more complicated. Suppose if an individual engages in the harmful activity, s/he faces some probability of being caught and fined and/or imprisoned. Imprisonment is costly. Therefore, the optimal fine is still maximal (since it is costless), but it could be efficient to supplement a fine with an imprisonment term (when the maximal fine is not very large due to lack of assets by offenders or when the marginal cost of imprisonment is reasonably small). The reason for this result is that the fine is socially costless; hence, it is advantageous to use it to its limits before using socially costly means of enforcement (Polinsky and Shavell 1984; Shavell 1985).

Extensions of the Basic Model

Since Becker (1968), the economic literature has taken the high-fine-low-probability result in the context of the assumptions and inquired as to its validity in a more general framework. Alternative setups have been considered including the possibility of legal errors; costs of imposing fines; general enforcement; marginal deterrence;

settlements (or plea-bargaining); self-reporting; repeat offenders; imperfect knowledge about the probability and magnitude of sanctions; corruption; incapacitation; costly observation of wealth; private enforcement; behavior of potential victims; social norms; and the fairness of sanctions (Garoupa 1997; Polinsky and Shavell 2001; 2007). It has been shown that the high-fine-low-probability result does not survive more realistic contexts.

For example, consider the model of marginal deterrence. In a seminal article, Stigler (1970) has noticed that individuals choose among a set of possible actions. Therefore, not only does general deterrence matter (compliance with the law), but also marginal deterrence (if an individual does not comply with the law, then s/he should commit the least serious offense). In other words, when faced with the possibility of committing several harmful acts, individuals should have clear incentives to choose the least harmful act. The important trade-off is that the marginal deterrence principle induces individuals to commit acts with lower harm (since they entail a lower expected sanction) but more individuals are induced to commit the act in the first place. Suppose, for example, that individuals may commit two acts if they are willing to. Some individuals may prefer to commit only one of them, but others certainly prefer to commit both. Repeat offenses could be framed in this context (although usually they are not discussed as a pure marginal deterrence problem).

A maximal fine should only be imposed on those who commit two offenses in order to satisfy the principle of marginal deterrence as explained by Polinsky and Rubinfeld (1991) in a nice application to recidivism. Polinsky and Shavell (1998) have a twist on the argument for enhancing penalties for recidivism based on the existence of records of prior convictions: When an individual contemplates committing an offense in the first period, s/he will realize that if s/he is punished, not only will s/he bear an immediate fine but a record will be issued (thus making her/him bear in the second period a sanction higher than it would be otherwise).

Another version of the model of marginal deterrence looks at situations where individuals have to choose between act 1 or act 2; they cannot commit both. Marginal deterrence issues arise as long as enforcement is general. If enforcement is specific to the act, then fines should be maximal and the probabilities chosen in the appropriate way. However, if enforcement is general, the optimal probability will be the same for both acts, and hence, the fines should reflect this interdependence (Shavell 1992).

A second example is punishment of attempts. An attempt is defined to be a potentially harmful act that does not happen to result in harm. If an individual commits an act, s/he will obtain a benefit and s/he will suffer a sanction with a probability that depends on whether or not s/he causes harm. Let us say that with probability q an act will result in harm. Hence, with probability $1 - q$, the act is only an attempt. Suppose that the fine for a crime is f and for an attempt is s. Also, the probability of apprehension for a crime is p and for an attempt is r. The expected sanction faced by an individual who commits an act is $qpf + (1 - q)rs$.

Punishing attempts increases deterrence by raising the probability of effectively imposing sanctions. Punishing attempts thus increases deterrence by expanding the set of circumstances in which sanctions are imposed. By the usual argument, there is an upper bound to sanctions. Once that bound has been set, the only way to deter it is by raising the probability of punishment. Therefore, punishing attempts is socially valuable. However, in many circumstance, courts cannot determine the probability of harm. Shavell (1990) shows that sanctions for attempts and for causing harm are uniquely determined in this scenario: (i) the sanction for attempts is never larger than the sanction for causing harm (hence, the sanction for attempts is typically nonmaximal); (ii) both are nondecreasing with the potential harm h.

A final example is the possibility of corruption. Corruption in law enforcement is a serious problem. It can include the payment of bribes or kickbacks to enforcers, threats to frame innocent individuals to get money from them (extortion), and actual framing (entrapment). In all these forms, corruption reduces deterrence and should be discouraged. One possibility is to make corruption a crime, thus creating a typical case of marginal deterrence. A second possibility is to reward enforcers for reporting violations; however, this option might create serious incentives to entrap and extort. Hence, the appropriate reward should balance the benefits of reducing bribery against the cost of expanding extortion and entrapment (Bowles and Garoupa 1997; Polinsky and Shavell 2001; Echazu and Garoupa 2010). Finally, a solution is to lower the sanction to eliminate the possibility of a bilateral agreement between an offender and a corrupt official; a corruption-proof solution requires the fine to equal the opportunity cost borne by the official and not the maximal upper bound (Becker and Stigler 1974).

Important extensions are also the application of the economic model to cooperative crimes. Two examples are relevant: corporate crime and organized crime. In the context of corporate crime, the debatable issue is the extent to which the fine should be paid by the managers or by the company. Corporate crime is not committed by companies, as such, but by different individuals within the corporation who are eventually criminally liable. In a perfect world, with complete contracting and without liquidity constraints, individual liability alone would induce efficient behavior. Consequently, corporate liability would not be necessary. Conversely, corporate liability is worthy of investigating when contracts are incomplete or when solvency matters. A socially optimal criminal sanctioning policy would favor large corporate fines for these individuals involved in the criminal activity (Garoupa 2000). This claim, based on Becker's analysis, assumes that corporate directors and shareholders who could be subject to large fines will provide the correct amount of employee monitoring, and eventually *ex post* sanctions on their employees to ensure that socially harmful crimes are not committed. However, large corporate fines could also induce corporate directors and shareholders to cover up for crimes committed by managers (in particular, when there is asymmetry of information, covering up might be the only available strategy to avoid liability). Therefore, large fines could be counterproductive and hinder rather than help deterrence.

A similar counterproductive effect can be detected in the context of organized crime. Severe punishment could shape criminal organizations in ways that make detection and conviction less likely. For example, it could induce strategies such as hostage taking or political corruption. If so, less severe punishment could be optimal when it assures an organizational structure that makes detection by law enforcement agencies easier or cheaper (Garoupa 2007).

Behavioral Criminal Law and Economics

On the modeling of criminal behavior, the economic approach has faced the standard criticism of using the expected utility framework, in particular the realism of economic rationality. The classical model assumes that individuals respond significantly to the incentives created by the criminal justice system. Nevertheless, there is some recent literature on behavioral criminal law and economics (Garoupa 2003; McAdams and Ulen 2010). This literature has shown that behavioral insights can improve the economic model to the extent that the rational model fails to gain empirical adherence in many contexts. However, the behavioral insights have not undermined the basic insights of the classical model.

Let us consider the optimism bias, a deviation from the rational choice model widely documented by behavioral economics. Such bias weakens deterrence by making individuals underestimate the probability of punishment. They could also overestimate the benefits derived from violating the law. Estimating that the costs are lower and the benefits are higher than they actually are, the overly optimistic individual commits more offenses than the rational criminal. As a result, optimal deterrence would require higher sanctions or higher probabilities (McAdams and Ulen 2010). Therefore, the optimism bias does not challenge the economic theory of deterrence, but the precise design of probability and severity of punishment.

Furthermore, Garoupa (2003) proposes a potentially offsetting effect. If the overly optimistic individual underestimates the probability of punishment, s/he might reduce avoidance activities (such as hiding illegal activities or other related precaution activities). A reduction in avoidance activities might increase the true probability of punishment, which partially cancels off the reduction of deterrence that excess optimism causes. As stated by McAdams and Ulen (2010), this example illustrates the importance of considering all the ways that some bias could affect the decision to violate the law.

Another example is cognitive dissonance and possible framing effects. Some individuals know crimes are fundamentally wrong and so they consider their behavior not to be a crime or to be morally justified. As explained by Garoupa (2003), similarly, framing could be important since most individuals do not like to feel responsible for death or violence, but they may accept a necessary death to achieve some particular goal or justified violence to reach a specific target. Manipulation of beliefs, framing effects, and cognitive dissonance impact economic rationality in significant ways. If criminal sanctions are to be understood as an incentive mechanism to induce compliance, efficient law enforcement needs to consider how individuals perceive their own criminal acts. Furthermore, social norms such as “ends cannot justify the means” or “there is no necessary death or violence” are consistent with the behavioral model.

It is important for an economic model of criminal behavior and enforcement to recognize that individuals have not only preferences over states of the world, but also over their beliefs of those states. Consequently, if the cognitive assessment of gains and losses is bounded by a reference point, it is the case that the individual objective function is no longer the expected utility functional. In such context, efficient enforcement is not only about shaping gains and losses to induce compliance, but also influencing possible reference points.

A third example mentioned in the behavioral literature is bounded self-interest, an important deviation from full rationality. Clearly individuals care about others in some, if not many, circumstances. Economists tend to use the notion of neoclassical altruism (for example, altruism is modeled in bequest decisions). Behavioralists propose a different approach by which individuals are less opportunistic if perceived to be treated fairly (positive reciprocity), but could instead be more spiteful if perceived to be treated unfairly (negative reciprocity). As observed by Garoupa (2003), formal economic models can tackle these different definitions and derive optimal law enforcement. A more challenging task is to understand how criminal law and procedure can be designed to enhance or reduce altruism (depending on the extent to which altruism increases or decreases crime rates). Certain crimes, such as hate crimes, are an illustration of negative reciprocity. It is likely to be efficient to penalize more hate motivated crimes in order to offset a possible existing negative reciprocity.

At the end of the day, the most significant challenge to the rational theory of crime as proposed by economists is empirical. The empirical literature on deterrence may be summarized as follows: increases in the probability of arrest, conviction, and punishment, and increases in the severity of punishment appear to have a deterrent effect on the population at large as well as on the small portion of the population that is most likely to commit crime (Levitt and Miles 2006, 2007). The exact magnitude of these deterrence effects is however controversial. Furthermore, the extent to which a deterrence effect can be obtained in a cost-effective manner still divides economists.

Related Entries

- ▶ Bayesian Updating and Crime
- ▶ Deterrence
- ▶ Deterrence: Actual Versus Perceived Risk of Punishment
- ▶ Rational Choice Theory
- ▶ Rational Choice, Deterrence, and Crime: Sociological Contributions

Recommended Reading and References

Becker GS (1968) Crime and punishment: an economic approach. J Polit Econ 76:169–217

Becker GS, Stigler GJ (1974) Law enforcement, malfeasance and compensation of enforcers. J Leg Stud 3:1–18

Bowles R, Garoupa N (1997) Casual police corruption and the economics of crime. Int Rev Law Econ 17:75–87

Echazu L, Garoupa N (2010) Corruption and the distortion of law enforcement effort. Am Law Econ Rev 10:162–180

Garoupa N (1997) The theory of optimal law enforcement. J Econ Surv 11:267–295

Garoupa N (2000) Corporate criminal law and organization incentives: a managerial perspective. Managerial and Decision Economics 21:243–252

Garoupa N (2003) Behavioral economic analysis of crime: a critical review. Eur J Law Econ 15:5–15

Garoupa N (2007) Optimal law enforcement and criminal organization. J Econ Behav Organ 63:461–474

Levitt S, Miles T (2006) Economic contributions to the understanding of crime. Annu Rev Law Soc Sci 2:147–164

Levitt S, Miles T (2007) The empirical study of criminal punishment. In: Polinsky AM, Shavell S (eds) The handbook of law and economics. Elsevier, Amsterdam, the Netherlands

McAdams R, Ulen TS (2010) Behavioral criminal law and economics. In: Garoupa N (ed) Criminal law and economics. Edward Elgar, Cheltenham, UK

Polinsky AM, Rubinfeld DL (1991) A model of optimal fines for repeat offenders. J Public Econ 46:291–306

Polinsky AM, Shavell S (1984) The optimal use of fines and imprisonment. J Public Econ 24:89–99

Polinsky AM, Shavell S (1994) Should liability be based on the harm to the victim or the gain to the injurer? J Law Econ Organ 10:427–437

Polinsky AM, Shavell S (1998) On offense history and the theory of deterrence. Int Rev Law Econ 18:305–324

Polinsky AM, Shavell S (2000) The economic theory of public enforcement of law. J Econ Literat 38:45–76

Polinsky AM, Shavell S (2001) Corruption and optimal law enforcement. J Public Econ 81:1–24

Polinsky AM, Shavell S (2007) The theory of public enforcement of law, In: Polinsky AM, Shavell S (eds) The handbook of law and economics. Elsevier, Amsterdam, the Netherlands

Shavell S (1985) Criminal law and the optimal use of nonmonetary sanctions as a deterrent. Columbia Law Rev 85:1232–1262

Shavell S (1990) Deterrence and the punishment of attempts. J Leg Stud 19:435–466

Shavell S (1992) A note on marginal deterrence. Int Rev Law Econ 12:345–355

Stigler GJ (1970) The optimum enforcement of laws. J Polit Econ 78:526–536

Ecoterrorism

Jennifer Varriale Carson
Department of Criminal Justice, University of Central Missouri, Warrensburg, MO, USA

Synonyms

Environmental extremism; Radical environmentalism

Overview

The radical environmental and animal rights movement in the United States has received a considerable amount of attention from federal law enforcement since September 11 and the subsequent "war on terror." In his 2002 testimony before the House Subcommittee on Forests and Forest Health, James F. Jarboe, the Domestic Terrorism Section Chief of the FBI, maintained that this movement was responsible for over 600 criminal acts. In 2006, FBI Director Robert S. Mueller stated that those motivated by a radical environmental and animal rights ideology continued to pose a threat. A survey conducted by Freilich, Chermak, and Simone (2009) found that 60 % of state police agencies in the United States reported that these groups represented a substantial hazard to the safety of their citizens, a threat that was only second to that of al-Qaeda and its affiliates.

Despite tending to be much more cautionary, scholars do acknowledge that these groups pose some threat. In 1996, Eagan felt that this activity had already peaked some 10 years earlier but cautioned that the Clinton administration's timber plan could lead to an increase in violence. Ackerman (2003), in a risk assessment of the Earth Liberation Front, argued that while this movement was not the most serious domestic threat in the United States, it had the capability to be much more destructive. In 2006, Liddick maintained that the animal rights movement in particular had become increasingly more violent.

Despite these concerns, environmental and animal rights extremists in the US have been a relatively nonviolent threat, responsible for only one suspected fatality and a handful of injuries since 1970. However, members of these groups have been responsible for a large amount of property damage. In fact, one series of arsons targeting a San Diego housing development resulted in over 50 million dollars in damages.

The radical environmental and animal rights movement, at one time classified as a top domestic terrorist threat primarily due to this immense ability to wreak economic havoc on their targets, remains an interesting case study of a homegrown movement that has engaged in terrorist tactics. This entry provides an overview of this movement with a discussion of the definitional issues involved, motivating philosophies, the group (or lack thereof) structure, preferred targets and tactics, and the federal government's response.

Definitional Issues

Despite the intense discussion surrounding radical environmental and animal rights groups from federal law enforcement and academics alike, few can agree on how to characterize this phenomenon. Most use the term "ecoterrorism," although the appropriateness of this phrasing has often been criticized. As defined by the FBI, ecoterrorism is "the use or threatened use of violence of a criminal nature against innocent victims or property by an environmentally orientated sub national group for environmental-political reasons, aimed at an audience beyond the target, and often of a symbolic nature" (US Congress, House Subcommittee on Forests and Forest Health). Several scholars contend that this language is both socially constructed and inappropriate due to the contention that environmental and animal rights extremists rarely target humans or use forceful means. In fact, some feel that using the word "terrorism" to describe the behavior of radical eco-groups has been part of a broader political movement, a movement linking environmentalism with terrorism. There is concern that such a political movement has had larger implications concerning the First Amendment and that labeling environmental and animal rights groups as terrorist organizations is a strategy to minimize their legitimacy. As evidence, many scholars point to the fact that ecoterrorism has only recently been added to the lexicon (per the Oxford dictionary in 1997) and labeled as such after the FBI expanded its own definition to include acts against inanimate objects in conjunction with the 2001 PATRIOT Act.

Thus, the arguments against applying the term terrorism, defined by LaFree and Ackerman (2009) as the "threatened or actual use of illegal force, directed against civilian targets, by non state actors, in order to obtain a political goal through fear, coercion or intimidation" (p. 1), to criminal incidents committed by these groups focus on two central themes. The first is directed toward the nature of the attacks, while the second concerns the intentions of those committing the attacks. Regarding the former, the majority of attacks perpetrated by those that adhere to a radical environmental or animal rights ideology constitute minor property damage rather than violent acts. As noted, only one suspected (and unconfirmed) fatality and a handful of injuries have resulted from radical eco-group incidents. While most attacks do involve property damage and certain incidents have cost victims millions, the nature of most acts is that of economic sabotage.

Second and perhaps more important is the issue of intention: members of radical eco-groups are selective about their targets, and those targets rarely include people. In actuality, only a small minority of attacks perpetrated by members of these groups is directed at specific persons. For the few cases where targets are not inanimate, most involve physical damage to executives' or researchers' offices or homes. This pattern is in direct contrast to most high-profile terrorist groups, who rarely have concern for civilian causalities and, in many cases, seek to harm the greatest number of innocent people to get their point across.

As a substitute for ecoterrorism, scholars have offered up alternatives like Vanderheiden's (2005) "ecotage," defined as the "economic sabotage of inanimate objects thought to be complicit in environmental destruction" (p. 432). The distinction here is viewed as a moral one, where once again targeting is focused on inanimate objects rather than people. Tree spiking is one example where such a moral distinction has been drawn. These devices, used to prevent the cutting down of trees, became responsible for injuring loggers, including a nearly fatal incident involving George Alexander, who was almost decapitated when his chainsaw struck a spike. As a result of this particular event, Judy Bari, a prominent leader of Earth First! at the time, acknowledged that changes needed to be made. She was later behind a larger nonviolent movement that publicly denounced tree spiking as a tactic through a partnership with the loggers. Nonetheless, and despite this development, tree spiking continues to be used as a tactic. In fact, environmental extremists have been known to use ceramic spikes in the place of metal ones in order to avoid detection.

Although the majority of scholars have argued against classifying attacks by radical environmental and animal rights groups as terrorism, few feel that this behavior can legitimately be construed as a form of civil disobedience. Others have pushed for labeling it as "political resistance." These terms, especially that of ecoterrorism and ecotage, are imperfect in that they do not represent the totality of behavior perpetrated by these groups. In other words, while there may be debate surrounding the applicability of the word terrorism to incidents involving these groups because of rare participation in events that fit the definition, this does not mean that these events *never* occur. On the other hand, several illegal activities do not meet the threshold of terrorist conduct; nonterrorist criminal incidents committed by radical eco-groups are nearly five times more common than terrorist attacks. Instead, as operationalized through Eagan's (1996) criteria (a position that will not compromise, status as a grassroots organization without any clear chain of command or any pay/benefits involved, and time and funds directed toward direct action rather than aimed at lobbying), this behavior is best referred to as the "illegal activities" or the "criminal conduct" of *members of radical environmental and animal rights groups, environmental and animal rights extremists, or members of radical eco-groups*. In addition to being less politically charged, the aforementioned terminology references the full spectrum of criminal behaviors, both terrorist and otherwise, committed by these groups.

Philosophical Underpinnings: Deep Ecology, Biocentrism, and Green Anarchy

Members of radical eco-groups are often motivated by the ideas behind "deep ecology" or the related concept of "biocentrism," a philosophy based on the belief that humans are only one element of the environment, an element that is not more important than any other but rather should peacefully coexist with all species. Consequently, biocentrism promotes the protection of all nature, as each species is vital to the survival of the ecosystem as a whole. Somewhat troubling about this philosophy is the implications it can have regarding the human population. What believers view as the ultimate goal (a pristine, preindustrial environment) is also an environment in which there are fewer people. However, there is little evidence that environmentalists interpret the canons of deep ecology in a way that would promote or justify violence.

The same cannot be said of the related albeit more extremist philosophy of green anarchy. Although sharing many of the same premises as deep ecology, green anarchy is blended with a much stronger anti-modernization and anticapitalist ideology. Developed in a magazine published in Eugene, Oregon, by the same name, this brand of thinking is tied to the promotion of harsher, sometimes even violent, strategies. The combination of standard environmental principles with a larger anarchist mentality expands the range of targets and encourages followers to work outside of the legitimate system. In fact, it is the legitimate legal, political, and economic system that with green anarchists take issue. Overall, there is some debate as to where the lines should be drawn between pure environmental versus animal-related versus anarchist philosophies, but most scholars maintain that there is considerable overlap.

Radical Environmental and Animal Rights Groups

As previously noted, Eagan (1996) argues that there are three main elements that all radical environmental and animal rights groups share: an uncompromising position, status as a grassroots organization without any clear chain of command or any pay/benefits involved, and time and funds directed toward direct action rather than aimed at lobbying. However, the environmental and animal rights extremist movement is, for the most part, just that a movement rather than a set of organized, hierarchically structured groups. There is often no one central leader, and the groups are predominantly composed of individuals or clusters of individuals who work separately without a formal chain of command. The cells share an ideology, but often attack independently and without knowledge of each other.

This lone wolf style has allowed for members of the radical fringe to maintain a certain amount of anonymity and, thus, avoid detection. This is perhaps best represented by the overall lack of successful convictions in comparison to the total amount of criminal activity perpetrated by environmental and animal rights extremists. However, the absence of a formal structure can also limit the amount of influence members have on each other. The one major exception to this informality is the "Family," an organized cell responsible for several arsons contributing to millions of dollars in damages and, consequently, the largest federal prosecution involving members of radical eco-groups in US history.

Two of the most prolific movements in the United States have been that of the Animal Liberation Front (ALF) and the Earth Liberation Front (ELF). ALF, originally formed in the mid-1970s in the United Kingdom as a splinter of the Hunt Saboteurs Association, has a primary objective of ending what it terms the "property status" of nonhumans. ALF encourages direct action in order to accomplish this mission, mainly through the rescue of animals and/or economic sabotage to the property of "animal exploiters." Their specific goals include liberating laboratory animals from suffering, causing economic harm to those that profit from animal exploitation, and bringing awareness to the plight of abused animals.

ELF is based on a similar premise to that of ALF, also promoting the destruction of the assets of those who, in their minds, threaten the environment. ELF, which splintered from Earth First! in the early 1990s in the United Kingdom, seeks to bring publicity through their various tactics to acts of environmental destruction. Perhaps most influential to the development of ELF was Edward Abbey's book, *The Monkey Wrench Gang*, which has become a guidebook of sorts to environmental extremists. ELF has focused its efforts primarily on combating the timber industry and, more recently, what they view as urban sprawl. ELF also claims to be strictly against any form of violence, but does promote illegal tactics if they are necessary to bring attention to, or to stop in some way, environmental harm. ELF has made the leaderless resistance style famous through its utilization of anonymous communiqués, sent to group spokesman Craig Rosebraugh and Leslie James Pickering and then later published on their website. Rosebraugh and Pickering claimed not be involved in the criminal activity, only to publicize it. Both have subsequently left to start another broader-based leftist organization that they named "Arissa."

The most recent addition to the environmental and animal rights extremist scene is the group Stop Huntingdon Animal Cruelty, or SHAC. SHAC was created in 1998 in the United Kingdom after a documentary aired on Huntingdon Life Sciences (HLS) that demonstrated its profound mistreatment of animals. Once the company moved its headquarters to New Jersey, SHAC became active in the United States through a series of incidents involving the harassment of HLS executives. SHAC has also been known to target businesses or organizations that support HLS, perhaps most famously was the campaign directed at Stephens Inc. After a series of incidents including the construction of a “StephensKills” website, the financial organization sold its shares of HLS. A similar situation then ensued with Marsh Inc., HLS’ primary insurer. The repeated harassment of employees, including a forced evacuation due to two smoke bombs, required the company to sever ties with HLS. The most important incident perpetrated by SHAC splinters occurred in 2003, when two pipe bombs were set off at Chiron, a biotechnology company and known associate of HLS. Although no one was injured or killed, the suspected mastermind Daniel Andreas “San Diego” was later added to the FBI’s most wanted terrorist list. The arrest of several SHAC leaders in 2006 (including the president Kevin Kjonaas) has limited its overall capabilities, even leading current president Pam Ferdin to maintain that the US branch would be disbanded. In fact, SHAC did not claim a single successful attack in either 2006 or 2007.

Targets and Tactics

As previously noted, the targeting of persons occurs in a minority of attacks perpetrated by members of radical environmental and animal rights groups, as do incidents involving violence. When people are targeted, everything from harassment (as with the aforementioned SHAC events) to the extensive damage of a researcher’s office is a strategy invoked by the radical fringe of the movement. In 1990, animal rights extremists were suspected in the shooting of a dean at the University of Tennessee due to the many threats he had received from these extremists prior to the incident. No one was ever charged with the homicide and the FBI has not directly connected the incident with a radical animal rights group. A handful of events committed by members of these groups have involved injuries, including the previously discussed nearly fatal incident involving George Alexander.

While violence directed at people has been a relatively infrequent occurrence, *threatening* violence has not been. Many scholars feel that the sentiment in communiqués and other propaganda from members of the movement has become increasingly radical and intolerant over time. Although seemingly hypocritical to a biocentrist ideology, the radical contingent of the environmental and animal rights movement has justified threats of violence by maintaining that certain corporate and government officials are beyond redemption. However, and despite this increasingly aggressive rhetoric, the behavior of members of radical eco-groups up to this point in time has remained mostly nonviolent.

A much more popular strategy of extremists is the targeting of symbolic entities in the form of physical structures; McDonald’s is one favorite of animal rights extremists. Other common targets include fur farms and research facilities, where a number of animal releases have taken place over the years. Radical environmentalists most often target the timber industry in one form or another, but have expanded in recent years to attacks on automobile dealerships and housing developments. One tactic specific to this movement is the use of the previously described tree spikes; long metal, or ceramic spikes driven into trees to prevent their removal. Upon contact with a chainsaw, as in the case of Alexander, tree spikes can be extremely dangerous.

A tactic common to both the radical environmental and animal rights movements is that of arson, the most destructive being the Vail ski resort attack in 1998 perpetrated by members of the Family and causing over 12 million dollars in damages. The burning down of urban sprawl targets, like the 2003 destruction of luxury homes under construction in Carmel Valley, California, is also a preferred strategy. Members of radical environmental groups have also turned their attention to setting car dealerships on fire, especially those that sell SUVs.

Bombings are a much less popular tactic despite the easy accessibility of how-to manuals on the Internet. Several incidents have been thwarted by police efforts or have taken the form of hoaxes rather than genuine attacks; the use of fake bombs may be more common than actual devices. Nonetheless, SHAC's use of pipe bombs against the Chiron Corporation and Shaklee Corporation parallels a number of other attacks using crude devices like the 2000 attack on luxury homes in Mount Sinai, NY, and the 2002 explosion of a poultry plant truck.

The use of CBRN weapons (chemical, biological, radiological, and nuclear) by environmental and animal rights extremists is also relatively rare in the United States. The most serious of these types of incidents involve the pouring of various acids (including muriatic and sulfuric) on researcher's equipment. One notable plot is the case of a small group of radical environmentalists in the 1970s that sought to eliminate the human race (save a select few that would not destroy the environment) through the release of infectious diseases, despite lacking the technical sophistication to be successful. Most of the significant cases of CBRN cases involving members of radical eco-groups have occurred outside of the United States, including one that utilized anthrax in a protest against the British government. However, scholars have cautioned that the use of CBRN weapons by environmental extremists could be a bigger issue in the future, despite philosophies that would seem to otherwise oppose the release of chemicals into the environment. If extremists felt it would further the cause in some way and save the earth from further destruction, they may employ these more serious tactics, particularly those groups that hold a more anarchist ideology.

Countermeasures and the Radical Environmental and Animal Rights Movement

Federal Legislation

The US governmental response to the criminal activities of members of radical environmental and animal rights groups has primarily come in the form of legislation. Two attempts at federally criminalizing this behavior, mainly due to their issues with the First Amendment, were unsuccessful: the Stop Terrorism Property Act of 2003 (STOP) and the Ecoterrorism Prevention Act of 2004. However, the tree-spiking clause added to the Anti-Drug Abuse Act of 1988 (ADA), along with the Animal Enterprise Protection Act of 1992 (AEPA) and its amended Animal Enterprise Terrorism Act of 2006 (AETA), were all signed into law.

All three of these federal sentencing acts have their own unique history and political context, yet all three share the same implication, namely; the use of sentencing enhancements for crimes, many of which were previously considered to be vandalism. Thus, this legislation has implications for a defendant's treatment and outcome in criminal justice proceedings. Most importantly, AETA in particular has consequences through treating environmental and animal rights offenders as terrorists in the federal system. Those charged as such, as shown in previous work, already receive longer sentences. In addition, the label of political criminal has been demonstrated as the most significant predictor of sentence length. After conviction, there are additional corrections-based consequences associated with those convicted including a significant loss in privileges. Therefore, ADA, AEPA, and AETA can all add to the offender's sentence through either prosecution under the specific legislation or through treating the offender as a political criminal in the courts and corrections systems.

The Anti-Drug Abuse Act of 1988 (ADA), as it name implies, was initially unrelated to radical environmental and animal rights groups. Motivated by the nearly fatal injury of Alexander, Idaho Senator James McClure proposed the inclusion of an addendum to the drug legislation that would federally criminalize tree spiking. Based on this single incident and despite the renunciation of the tactic by Bari and others, legislators argued that environmental extremists had injured a number of loggers and needed to be held accountable. ADA sets a penalty of up to

20 years in prison for the use of a "hazardous or injurious device" aimed at stopping the timber harvest in some way. The passing of this legislation was significant in regard to the rather large penalty tacked on to an act that was previously considered to be a misdemeanor. However, ADA's primary importance is tied to a larger political movement targeted at the radical environmental and animal rights movement; its hearings became the first public invocation of the term ecoterrorism. Therefore, while this act is focused on tree spiking, it was part of a broader effort to symbolically legislate the criminal conduct of members of radical environmental and animal rights groups at the federal level.

The Animal Enterprise Protection Act (AEPA) is another important legislative countermeasure that increased penalties on anyone that disrupts an "animal enterprise." These disruptions include the stealing or damaging of property, including the animals, that leads to a total loss of $25,000 or more (or that conspires to do so). Sentences are based on a four-tiered classification system (economic damage, major economic damage, serious bodily injury, and death), with a spectrum of punishments from fines and restitution to life imprisonment. Many activists were concerned about First Amendment violations after AEPA was passed, but supporters contend that legal protesting is not an act that is punishable under this part of the legislation.

The Animal Enterprise Terrorism Act of 2006 (AETA) served as an extension to AEPA, primarily in response to SHAC's strategy of targeting third parties. AETA once again enhanced penalties for disruptions of animal enterprises, but added language that penalized activities that lead to "reasonable fear" of an enterprise's owner. AETA also lowered the damage limit to $10,000 and broadened the offense categories of AEPA with a focus on interstate travel and the mail service. Most importantly, AETA specifically addressed the problem of third-party targeting by expanding the definition of animal enterprises. By passing AETA and increasing its scope, federal prosecutors hoped to have a much more useful tool for combating the problem of environmental and animal extremists.

Given this expansion, AETA has been the most controversial of the federal sentencing acts aimed at members of radical eco-groups. In a letter to Congress, the American Civil Liberties Union (ACLU) contended that AETA criminalizes rights protected under the First Amendment, such as peaceful demonstrations and boycotts. The ACLU argued that because the legislation was broad and somewhat vague given the terminology of "disruptive activities," it could be misused for the surveillance of law-abiding animal rights groups.

The cases of *U.S. v. Buddenberg* and *U.S. v. Demuth* demonstrate the controversy associated with AETA. In the former case, four individuals were charged under AETA based on the contention that they had trespassed, intimidated, and harassed University of California researchers and their families. Although some of their actions were criminal, many other actions listed in the indictment were protected by First Amendment rights including protesting, chalking a sidewalk, chanting, leafleting, and using the internet to find out information regarding biomedical researchers. The case was later dismissed without prejudice, giving the government power to charge the defendants in the future.

The latter case, that of *U.S.A. v. Demuth*, involves a sociology graduate student charged under AETA after refusing to testify about a 2004 laboratory break-in at the University of Iowa. Demuth was conducting research on environmental and animal rights activists at the time and would not divulge sources nor relay any information he received from his contacts. Using statements like "it's almost been a year since Iowa" from Demuth's social networking sites and his journal, the FBI maintained that he was involved in the break-in in some capacity. Several activist groups and his former academic advisor have protested the indictment.

Key Prosecutions

Given the FBI's focus on the radical environmental and animal rights movement, it is not surprising that over 30 individuals have been indicted since 2005. Perhaps the most well known of these cases is the charging of members of the Family in connection with 17 attacks from 1996 to 2001, including the Vail ski resort arson. Coined "Operation Backfire," several FBI field offices and the Bureau of Alcohol, Tobacco, and Firearms worked with local law enforcement to get defendants to plead to lighter sentences in exchange for information regarding the others involved. Most of the defendants were sentenced in 2007 and are serving anywhere from 37 to 188 months in a federal penitentiary.

Another notable conviction involves Eric McDavid, a green anarchist sentenced to 20 years in prison for plotting to sabotage the Nimbus Dam in California. The FBI used an informant to infiltrate the plot and arrested McDavid after he and his coconspirators bought bomb-making equipment. Although his defense claimed entrapment, McDavid's appeal has subsequently been denied.

A third and final important prosecution involved the aforementioned leadership of Stop Huntingdon Animal Cruelty, termed the SHAC-7. Notable as the first case charged under AEPA, six members of SHAC were sentenced to 3–6 years each and ordered to pay a million dollars in restitution for their involvement in a website that posted home addresses of HLS employees. Although the most direct link to criminal activity involved a phone call Kjonaas had made, members of the group were held responsible for the site that the prosecutor claimed incited violence. As previously mentioned, this high-profile prosecution was enough to disband the US chapter of SHAC despite the vocal opposition to the case from many free speech organizations.

The Future of the Radical Environmental and Animal Rights Movement

The 2010 hostage-taking situation involving James Lee, an activist described by some news outlets as an "environmental militant" and compared in others to the Unabomber, forever damaged perceptions of the environmental and animal rights movement. Lee, upset with the change in Discovery Channel programming to what he viewed as pro-birth and diversions to the global warming crisis, strapped a homemade bomb to himself and took three people hostage. In the end, only Lee was killed, but the real damage to the legitimate and law-abiding contingency of the environmental and animal rights movement was done. Despite any group taking responsibility for his actions, Lee became the new face of the movement, a movement that has long encouraged lone wolf type attacks and, consequently, had a difficult time defending itself from this attack.

All in all, cases like the one involving James Lee have been few and far between over the years. The environmental and animal rights movement, even its more radical and anarchist fringe, has thus far not proven to be a violent threat. What it has been, to many of its corporate targets, is an economic one and very destruction at that. Estimates range into the billions of dollars in damages, with some companies permanently destroyed from attacks.

The real question remains whether Lee will continue to be the exception rather than the rule. Although some scholars cite an increase in violent rhetoric, particularly among the green anarchists, counterterrorism experts maintain that the focus should be on past behavior rather than language, as the former is a much stronger predictor of what is to come. The rather steady decrease in total criminal incidents perpetrated by members of radical eco-groups since 2001 is additional evidence that this threat may be dissipating. Given these considerations, it may be best for US counterterrorism efforts to focus more on other homegrown threats that have proven to be more violent over time, such as the radical right-wing movement. It is possible that the environmental and animal rights extremists will be responsible for violent acts in the future, but it is likely that those acts will be very uncommon.

Recommended Reading and References

Abbey E (1975) The monkey wrench gang. Harper-Collins, New York

Ackerman G (2003) Beyond arson? A threat assessment of the earth liberation front. Terror Polit Violence 15(4):143–170

Amster R (2006) Perspectives on ecoterrorism: catalysts, conflations, and causalities. Contemp Justice Rev 9(3):287–301

Carson J, LaFree G, Dugan L (2012) Terrorist and non-terrorist criminal attacks by radical environmental and animal rights groups in the United States, 1970–2007. Terror Polit Violence 24(2):295–319

Eagan SP (1996) From spikes to bombs: the rise of eco-terrorism. Stud Confl Terror 19:1–18

Freilich J, Chermack S, Simone J (2009) Surveying state police agencies about terrorist threats, terrorism sources, and terrorism definitions. Terror Polit Violence 21(3):450–475

Jarboe J (2002) The threat of eco-terrorism. Retrieved 11 November 2011 from FBI website: http://www.fbi.gov/news/testimony/the-threat-of-eco-terrorism

LaFree G, Ackerman G (2009) The empirical study of terrorism: social and legal research. Ann Rev Law Soc Sci 5:347–374

Leader SH, Probst SH (2003) The earth liberation front and environmental terrorism. Terror Polit Violenc 15(4):37–58

Liddick DR (2006) Eco-terrorism: radical environmental and animal liberation movements. Praeger, Westport

Mueller RS (2006) Operation backfire press conference. Retrieved 11 Nov 2011 from FBI Web site: http://www.fbi.gov/pressrel/speeches/mueller012006.htm

Smith R (2008) Eco-terrorism? A critical analysis of the vilification of radical environmental activists as terrorists. Environ Law 38:537–576

Taylor B (2003) Threat assessments and radical environmentalism. Terror Polit Violence 15(4):173–192

Trujillo HR (2005) The radical environmentalist movement. In: Jackson BA, Baker JC (eds) Aptitude for destruction: case studies of the organizational learning activities of five major terrorist groups and a methodology for ascertaining what and why they learned. RAND Corporation, Santa Monica

U.S. Department of Justice and U.S. Department of Agriculture (1993) Report to congress on the extent of domestic and international terrorism in animal enterprises. Physiologist 36(6):207–208

Vanderheiden S (2005) Eco-terrorism or justified resistance? Radical environmentalism and the “War on Terror”. Polit Soc 33(3):425–447

Vanderheiden S (2008) Radical environmentalism in an age of antiterrorism. Environ Polit 17(2):299–318

Walker AB (2007) A field of failed dreams: problems passing effective ecoterrorism legislation. Vill Envtl Law J 18:99–118

Welchman J (2001) Is ecotage civil disobedience? Philos Geogr 4(1):97–107

Effective Community Supervision Programming

James McGuire
Institute of Psychology, Health, and Society, Division of Clinical Psychology, University of Liverpool, Liverpool, UK

Overview

This entry describes and discusses aspects of a comparatively recent innovation in criminal justice services, which has taken hold in a number of countries: the development and implementation of *structured programs* designed with the specific purpose of reducing the recidivism rates of participants. While such programs have been delivered both in prison and community settings, the present entry focuses primarily on the latter. Following an introduction of some basic concepts, the entry provides an account of the theoretical rationale that underpins this departure, and reviews sources of evidence on the principle of risk-need assessment which has informed the choice of methods and contents used in programs of this type. This is illustrated with reference to some well-established types of program, before turning to a discussion of some of the practical issues arising from engaging in work of this type.

Several key issues are then examined, including the following: the evaluation of program use and the outcome evidence so obtained, both from controlled experiments and following large-scale dissemination; the quality control of programs through accreditation processes; and the organizational context within which programs are delivered – what have been called *nonprogrammatic* influences. This is followed by a review of some of the implications of these results, focusing in particular on the capacity of community supervision to absorb larger numbers of offenders through transfer of resources from custodial sentences, in a process that has been labeled “justice reinvestment.” There is some brief discussion of strategies for enforcement of community supervision, and this entry concludes with discussion of some outstanding issues and controversies arising in this area.

The Emergence of Structured Programs

This entry will focus on one relatively recent development within the process of probation or community supervision: the advent and widespread dissemination of structured programs. These involve offenders in participatory, preplanned activities with the objective of enabling them to reduce their likelihood of committing further crimes. It is difficult to identify an exact date when this specialized type of practice arrived on the probation scene, but as an approximation it can be traced to the mid-1980s and the appearance of a number of publications which ran counter to the then prevalent belief that "nothing works": that there was no realistic prospect of successfully achieving offender rehabilitation. During that era, many legal systems were still imbued with an ethos informed by the erroneous impression created by research published in the 1970s, suggesting that persistent offending behavior could not be modified and that what had previously been called rehabilitation or reform of offenders was not practically feasible (Gaes 1998). Mounting evidence since then has contradicted those suppositions, and the return of what is loosely and perhaps misleadingly called "treatment" to the criminal justice arena further strengthened interest in and adoption of programs.

Probation or other community-based supervision of offenders was traditionally centered on humanitarian concerns, and the probation officer's principal role was to address the individual's personal and moral well-being. At a later stage, more concrete matters such as employment, accommodation, financial or family problems came to the fore, and supervision became more practical in its focus. Despite that switch of emphasis, behavior change was still thought to be accomplished largely through a relational process, and in that respect, working with individual offenders was considered to have similarities to counseling or psychological therapy. Thus, there was an accent on the supervisor's ability to develop a working relationship with offenders. It had always been realized that alcohol and other substance abuse disorders were associated with criminal offending, but as substance abuse became a problem in many societies and its close association with criminal offending came to be understood better, the incapacity of relational models to succeed in addressing and resolving such problems began to be exposed. Furthermore, as cost-conscious governments focused more on the accountability and cost-effectiveness of public services, community supervision agencies were carefully scrutinized regarding offender outcomes. This was paralleled by an emphasis on risk assessment, risk management, and public protection as core duties of community supervision staff.

The precise sequence whereby programs were adopted as a component of working with offenders is difficult to chart. Supervision became incrementally more structured, first through the advent of task-centered casework and an expansion in the use of group work, followed by the gradual emergence of greater explicitness and structure in the work that was done. The majority of the programs to be discussed here are delivered in small groups within which the climate takes less of a therapeutic inclination than of one that has more similarities to an educational process or a training session. However, relationships between participants and the role of the tutor or leader as a facilitator of collective learning still play a crucial, if underappreciated, part.

In this entry we will first consider definitions of some key terms. We will then consider the background and theoretical models that have influenced the development of programs, before looking more closely at the composition of some of the programs themselves. This will be followed by a review of how they are applied in practice and the kinds of findings that have been obtained from evaluations of their use. It is now widely accepted that programs alone are only part of effective services; therefore, we will also look briefly at what are called "nonprogrammatic" factors. Finally, there will be some discussion of outstanding and unresolved issues that still challenge this field.

Fundamentals

In what follows, several aspects of community-based programming will be discussed. A useful starting point is to consider definitions and the features that the specialized activities known as structured programs have in common. The word *program* as used in this context refers simply to a planned sequence of learning opportunities that can be reproduced on successive occasions (McGuire 2001). In criminal justice settings, its general objective is to reduce participants' subsequent criminal recidivism. Within that context, the typical program is a set of activities prepared in advance, with clearly stated objectives, and comprises a number of elements interconnected according to a planned design and with a demonstrable link to those objectives. So these are sometimes referred to as "structured" programs, to distinguish them, for example, from therapeutic communities which have preplanned elements but a lower level of detailed specification of content. For methods and exercises to be reliably replicated on successive occasions, they need to be described and recorded in an accessible, manual format that correctional practitioners will readily understand. Such "manualization" – essentially, preparing program materials in a portable form that would be readily usable on repeated occasions by those delivering them – has become a standard aspect of the use of programs. This has allowed critics of the movement towards programs to allege that supervision has been mechanized and that there has been a neglect of individual needs, where the use of structured methods instills a monolithic "one size fits all" approach.

If correct this would be a legitimate concern, but two responses can be made to it. First, program manuals vary considerably in the amount of detail they contain and in their level of prescriptiveness. Whereas some specify the minutiae of the actions of facilitators on an almost moment-by-moment basis (even, e.g., providing the text of statements made to participants at various points in the process), others are loosely or schematically formulated, affording sizeable latitude in how a method should be applied or a skill-training exercise should be run. This creates scope for adaptations to be made to take account of diversity among participants. It also creates opportunities for facilitators to exercise discretion as to the details of how sessions should be run, drawing on their own experience and their judgments of a group, though this must be kept within specified limits if the integrity of the program is to be preserved. Second, programs are not seen as solutions in themselves but as just one part of the overall supervisory process. In most contexts, participation in a probation program generally lasts for only a few weeks, usually at the commencement of a community sentence. Thus, a larger portion of time remains during which the individual's primary contact with the criminal justice system is through his or her probation supervisor. While supervising officers are not generally involved directly in program delivery, they should be aware of the nature and contents of programs their supervisees have attended. Thus, while a program should have a direct impact, it should also work as a catalyst for the longer-term supervisory experience and should instigate an orientation whereby learning and change can be sustained and an individual is motivated and equipped to continue working on problems he or she has identified.

Background Research and Theory

The growth of interest in programs and their dispersal was probably stimulated and sustained by research showing that individuals, even those with extensive criminal histories and apparently entrenched patterns of antisocial attitudes and behavior, are capable of change: efforts at "rehabilitation" can and do work. The pessimistic stance that dominated the correctional enterprise for two decades or so, commencing approximately in the mid-1970s, is now widely acknowledged to have been based on a number of errors and misinterpretations of the data that were extant at that time. Some scholars raised those points contemporaneously and challenged the escalating punitive ethos (e.g., Gendreau and Ross 1980). Since then, however, far more evidence has accumulated on this and related points.

The introduction of structured programs as a component of community supervision of offenders was grounded on the premise that such a development would have a discernible impact in reducing criminal recidivism. As a response to the controversies of an earlier era and the dismissal of rehabilitation from correctional discourse and practices, what is colloquially known as the "what works" research began to flourish from the mid-1980s onwards. This consists of a considerable volume of primary studies in which selected programs are evaluated with respect to their effects on re-offending rates of participant offenders, usually measured against an "untreated" or "business as usual" comparison group. The results of such studies have subsequently been integrated in surveys of research typically involving use of the statistical techniques of meta-analysis or systematic review.

There has been a steadily expanding number of such reviews of designated portions of this literature (McGuire 2008). At the time of writing, this comprises no fewer than 100 relevant meta-analyses of programmatic interventions (McGuire 2013). This excludes evaluations of other criminal justice approaches such as changes in legal procedures or in sentencing practices, situational crime-preventive measures, treatment of substance abuse conducted in healthcare settings, studies that focus purely on medical interventions, or interventions focused wholly or primarily on children below the age of criminal responsibility.

An overview of these findings reveals that on average, interventions yield a net reduction in offender recidivism that is modest, though statistically significant. This average figure is however somewhat higher when punitive interventions (which have either zero or negative effects) are excluded from it. More importantly, there are consistent patterns within the positive results that remain, such that there is justification for the claim that appropriately designed interventions can lead to reductions in subsequent recidivism of the order of 20–25 %. A proportion of the programs reviewed by Lipsey et al. (2007) which exhibited features of "best practice" produced an effect size (using the odds ratio statistic) equivalent to a reduction in recidivism among experimental as compared to control groups of 52 %.

There is a close association between background research, theoretical precepts, and everyday practice in the design of offending behavior programs (see, e.g., McGuire 2005). Their content is not derived from the now discarded notion that general differences can be found between samples of offenders and non-offenders. The concept of there being a distinctive "criminal personality" has been largely abandoned, and it is recognized that a majority of citizens break the law at some stage of their lives. At the same time tentative propensities towards involvement in crime may be unevenly distributed, and as a consequence of an interaction between temperament, socialization processes, and environmental variables, some individuals may be more likely to become repetitively or persistently involved in criminal conduct. Such a pattern has been found in many longitudinal and survey studies (Andrews and Bonta 2010), and models have been forwarded concerning the possible distinguishing features of those more prone to delinquency. Research and theory therefore focus more closely on identifying patterns within this variation.

This has led to the emergence of a model of criminal involvement focused on "risk factors," which are aspects of individual functioning that have significant and consistent associations with the likelihood of arrest and rearrest. Those most frequently discovered from both longitudinal studies and from meta-analysis of cross-sectional and treatment-outcome studies include the following: having a large proportion of associates who are involved in delinquency; low levels of self-control, with a tendency to be impulsive; antisocial attitudes and beliefs; and poor or underdeveloped skills in interpersonal problem-solving and social interaction. These are often conjoined with substance abuse and in a proportion of cases with other features of antisocial personality. When working with any given individual, it is considered essential to assess him or her with reference to the aforementioned risk factors, and a range of methods has been developed for undertaking this.

The nature of these associations and the causal pathways involved between them has been most clearly formulated in the *risk-needs-responsivity* (RNR) model (Andrews and Bonta 2010), which remains the principal theoretical model informing the design of programs. This model draws on concepts from social learning theory developed by Albert Bandura and others, with concepts of socialization and family influence as discovered in the work of the Oregon Social Learning Center and with other findings of developmental and life-course criminology. To an extent these ideas constitute a convergence of social learning theory and control theory as conceptualized in criminology but are further enriched by an understanding gained from research in personality and social psychology on the nature of person-situation interactions, differential associations, and the development of regular patterns of social cognition and self-appraisal.

Intervention methods that flow from this theoretical viewpoint are fundamentally grounded in the expectation that the patterns just described are largely (though not exclusively) products of development and learning. On that basis it is deduced that they can also be redressed by acquisition and application of new skills that will enable individuals to encounter high-risk situations and successfully negotiate a route through them without resorting to a criminal act. The central tenet of the attempt to achieve this is the realization that cognition (thoughts), emotion (feelings), and behavior (actions) are three discrete domains of activity and experience that are interdependent and in a constant state of interplay and mutual influence. Furthermore, there is also a causal (bidirectional) interplay between them and environmental events, a model called *reciprocal determinism*. Another way of describing this is to say that individual behavior and experience are a function of *person-situation interactions*.

The behavioral components of the programs derived from this draw on the fundamental principles of learning theory and the application of classical and instrumental conditioning to therapeutic change. The behaviorism of B. F. Skinner and other researchers is revised to incorporate social learning concepts such as modeling and internal representation thereby learning, e.g., about the effects of peer-group pressure, and acquiring skills of assertiveness to resist it. The cognitive components focus on the relationship between "automatic" and "controlled" processing of information, thought to lead to habitual patterns in the ways individuals perceive themselves and others, think about difficulties they are facing (or avoid doing so), and solve problems (or allow them to accumulate).

Program Designs and Contents

A majority of successful programs used today exploit some permutation of methods originating from cognitive-behavioral therapy (CBT). This is a theoretically driven, psychologically based approach to the amelioration of a broad spectrum of individual and family problems, covering the whole life span. The approach is widely used in education, mental health, and other services and to date remains the best researched and most firmly "evidence-based" of the different treatment models available in those fields. Variants of it have been adapted for use in criminal justice services. This is, however, by no means the only model that has been utilized in program designs, simply one that has become dominant as a result of a larger volume of published evaluation research.

The following are some examples of programs that are among the most widely used, some in adapted forms in prison as well as community settings. Fuller descriptions of many programs can be found in Hollin and Palmer (2006).

Reasoning and Rehabilitation: Usually simply called R&R, this is the longest established and most extensively employed program of this type. Developed in Canada, it has reportedly been used with over 50,000 offenders in 17 countries

(Antonowicz 2005). It consists of 38 sessions of 2 h each, covering a range of modules devoted (among others) to problem-solving, emotional control, negotiation training, values enhancement, creative thinking, and critical reasoning.

Enhanced Thinking Skills (ETS) is an abbreviated, 20-session version of R&R developed in the prison service of England and Wales but later used in probation, focusing mainly on problem-solving, self-management, and victim perspectives.

Think First: This is prepared in several variant forms for use in probation, prisons, and youth justice, respectively. The program is designed so that identified risk factors for criminal recidivism are focused on in successive phases (moving from problem-solving through self-management to social interaction skills, attitudes, and values) and progressively combined to enable participants to address different aspects of their offenses. Sessions incorporate analysis of offending behavior itself, that is, the sequence of events that occurred when an individual committed an offense, and what he/she was thinking and feeling at the time. This work is carried out in detail for each group member's most recent or most serious offense, then repeated across several incidents to examine any patterns that emerge.

Developed to work with groups with histories of a variety of criminal behaviors (individuals regarded as "versatile" offenders), the above three programs are sometimes called general offending behavior programs. Other programs however address specific types of offense, usually of a more serious (e.g., violent or sexual) nature, or with a more pronounced pattern in an individual's criminal history (e.g., substance misuse).

Several structured behavior change programs have been developed with a focus on recurrent violent offending. They include *Aggression Replacement Training*, an 18-session program initially developed for youth in the United States but more recently used with adults. It directs participants' attention to three aspects of events where they may have committed offenses, those of self-control ("what not to do"), social/interpersonal skills ("what to do"), and moral reasoning training ("why to do it"). For those with longer established patterns of serious violence, the *Cognitive Self-Change Program* may be used. Initially developed in prisons in Vermont, it is designed for individuals with a history of personal assaults who wish to address and alter their pattern of behavior; it involves 38 sessions lasting 2–2½ h. Where an individual's violent reactions are thought to derive from difficulties in anger control, many programs are available, including *Controlling Anger and Learning to Manage It* (*CALM*), a 24-session program focusing intensively on enabling individuals with histories of angry aggression to acquire skills in self-management of angry feelings.

There is a portfolio of programs now available for individuals who have committed sexual offenses. The *Sexual Offender Treatment Programme* (*SOTP*) entails a suite of five programs designed for the reduction of sexual offending, with forms available for use in both prisons and community. The original version was an intensive, multimodal intervention for use in prisons and comprised several versions for individuals varying by level of risk and level of functioning, together with supplementary sessions designed as "booster" or maintenance exercises. In England and Wales there are also several probation-based variants.

Substance Misuse Programs: Numerous programs have been developed to address substance abuse both in health and criminal justice settings. They differ according to their target problem (e.g., alcohol, illegal drugs), their planned outcome (abstinence, control), and their intensity (number and duration of sessions). Many include follow-up activities focused on relapse prevention for which there are additional methods and materials employed. Still others are designed for specific offenses arising from substance abuse, such as programs for driving while intoxicated, or for alcohol-related aggression.

Application and Dissemination

The majority of the studies reviewed in the meta-analyses briefly summarized above were conducted on a comparatively limited scale and often, though by no means always, under fairly well-controlled conditions. The question therefore arises as to whether similar findings can be secured when interventions of this type are propagated, organized on a larger scale, and assimilated into routine practice. Impressed by the findings, governments of several countries embarked on policy experiments in which specially designed programs or other services were introduced into community supervision (and in some instances also prison) sentences. Programs have been in use in many countries, in locations as far apart as the United States, Canada, England and Wales, Scotland, Ireland, Germany, Sweden, Finland, Norway, Estonia, Spain, Portugal, Turkey, Chile, Hong Kong, Australia, and New Zealand. In England and Wales, which has perhaps gone farthest in this direction, this formed part of a large-scale policy initiative, the *Crime Reduction Programme*, which was made operational from 2000 onwards. It entailed the gradual installation of structured interventions (designated "Pathfinder" programs) at many probation sites and an extensive menu of staff training to deliver them.

Given the plethora of program types, by the late 1990s the field of corrections rather quickly took on the appearance of a marketplace in structured interventions. As part of the drive towards transparency and accountability, some justice departments developed standards that programs were required to meet before they could be approved for use in justice services. The adoption of programs was accompanied by the establishment of "accreditation" processes to select the methods thought most likely to reduce recidivism. In England and Wales, an elaborate system was created for monitoring standards of delivery and evaluating outcomes. This led to the establishment of an independent group of expert consultants, currently called the *Correctional Services Advisory and Accreditation Panel* (CSAAP). The Panel devised a procedure for evaluating the suitability of programs and published a set of accreditation criteria for that purpose (Correctional Services Accreditation Panel 2010). Programs submitted to the Panel for accreditation are judged in terms of a set of 10 criteria, specifying that they should meet the following requirements:

1. *A clear model of change*: The program is based on a clear, explicit model of change with a sound basis in a broader theoretical framework.
2. *Selection of offenders*: There are clear and appropriate criteria for the selection and allocation of those who should take part.
3. *Targeting a range of dynamic risk factors*: The program's materials and methods must target a range of dynamic risk factors, that is, the program is multimodal in format and content.
4. *Effective methods*: The program entails use of demonstrably effective methods of intervention, derived from relevant research literature.
5. *Skills orientated*: Methods used in the program focus on the acquisition and development of relevant skills to enable individuals to avoid re-offending.
6. *Sequencing*, *intensity*, *and duration*: There are clear, explicit links between the sequencing, intensity, and duration of the program's separate components.
7. *Engagement and motivation*: The methods used take account of the need to engage and motivate participants and manuals specify how this will be done.
8. *Continuity of programs and services*: The program is integrated with other aspects of service provision, for example, sentence planning or community supervision.
9. *Process evaluation and maintaining integrity*: Provision is made for monitoring integrity of delivery, that is, to check that the sessions are being delivered in accordance with the planned model.
10. *Ongoing evaluation*: Arrangements are in place for data collection and analysis, used in ongoing and outcome evaluation.

In order to receive accreditation, a program submitted to the CSAAP must be accompanied by a considerable quantity of supporting documentation that clarifies precisely what it contains and what is involved in its delivery. Each program should have five manuals: (a) a *Theory Manual* which describes the intervention model and evidence supporting it; (b) the *Program Manual* which specifies the content, exercises, and materials used in the sessions; (c) an *Assessment and Evaluation Manual* showing how the applicants are assessed, their progress is monitored, and outcomes evaluated; (d) a *Management and Operational Manual* prescribing aspects of delivery and organization; and (e) a *Staff Training Manual*. The program materials are evaluated and allotted a score on each of the 10 criteria listed above (such that a score of 2 = fully met, 1 = partially met, 0 = not met). A program must gain a total score of 18–20 points to be awarded accredited status. A history of the Panel's work and an account of the challenges arising within it are given by Maguire et al. (2010). Similar arrangements involving expert panels or other validation procedures have been set up in Scotland, Canada, Sweden, and elsewhere.

Evaluation of Outcomes

Can the introduction of structured programs within community supervision make a meaningful difference to its effectiveness? It is one thing to obtain a good result from an experimental trial, but another to capture the same result when interventions are "rolled out" across multiple, widely dispersed sites (the difference between what have been called "demonstration" as compared to "practical" programs). It may be useful to examine this first on a "before and after" basis by comparing two sets of results. Figure 1 compares four types of sentencing "disposals" available to the courts in England and Wales during the 1990s. In each case, the predicted rate of reconviction within 2 years is compared with the actual rate over the same timescale. The predictions were calculated using the *Offender Group Reconviction Scale* (OGRS-2), a specially designed instrument developed on the basis of multivariate analysis of criminal histories of large samples of offenders. This combined a series of seven pieces of data into a prediction score, the percentage of offenders likely to be reconvicted within 24 months. The OGRS (now in its third revised version, OGRS-3) has been validated as among the most accurate predictors in recent meta-analytic reviews of the field of criminal reconviction (see, e.g., Yang et al. 2010).

The four types of sentences compared in the upper graph of the figure are those which were the principal disposals available to the courts in the 1990s: respectively, they were immediate imprisonment, community service order, probation supervision, and probation with additional requirements. The data were originally compiled in order to compare the differential effectiveness of sentences, primarily the relative impact of custodial versus community sentences on recidivism. The value of these data lies in the side-by-side comparison of *predicted* with *actual* rates within each sentence type. As can be seen in each case, the figures are very close; the rates at which individuals were reconvicted were very similar to the rates at which they were probably going to re-offend, and that occurred regardless of the type of sentence imposed upon them. The sentence of the court appeared to have no obvious bearing on the outcome; indeed it seems virtually unrelated to it. Community sentences were neither any better nor any worse than prison sentences in terms of recidivism outcome. The relevance of this for present purposes, however, lies in the finding that there was no evidence that probation supervision had had a beneficial effect in reducing reconviction rates. Such a finding is not unique, but fits with other findings obtained from elsewhere. In a study in the Australian state of New South Wales, supervision as traditionally conceived was not found to have any greater effect on reconviction rates than a court bond involving no direct supervision of offenders (Weatherburn and Triboli 2008). In a large-scale review of the use of intensive supervision programs (ISPs), with a study sample of 1,812 offenders across nine American states, in which a primary emphasis was placed on the surveillance component, similarly no evidence of a beneficial impact was found (Petersilia and Turner 1993). In a meta-analysis of studies of community supervision conducted for Public

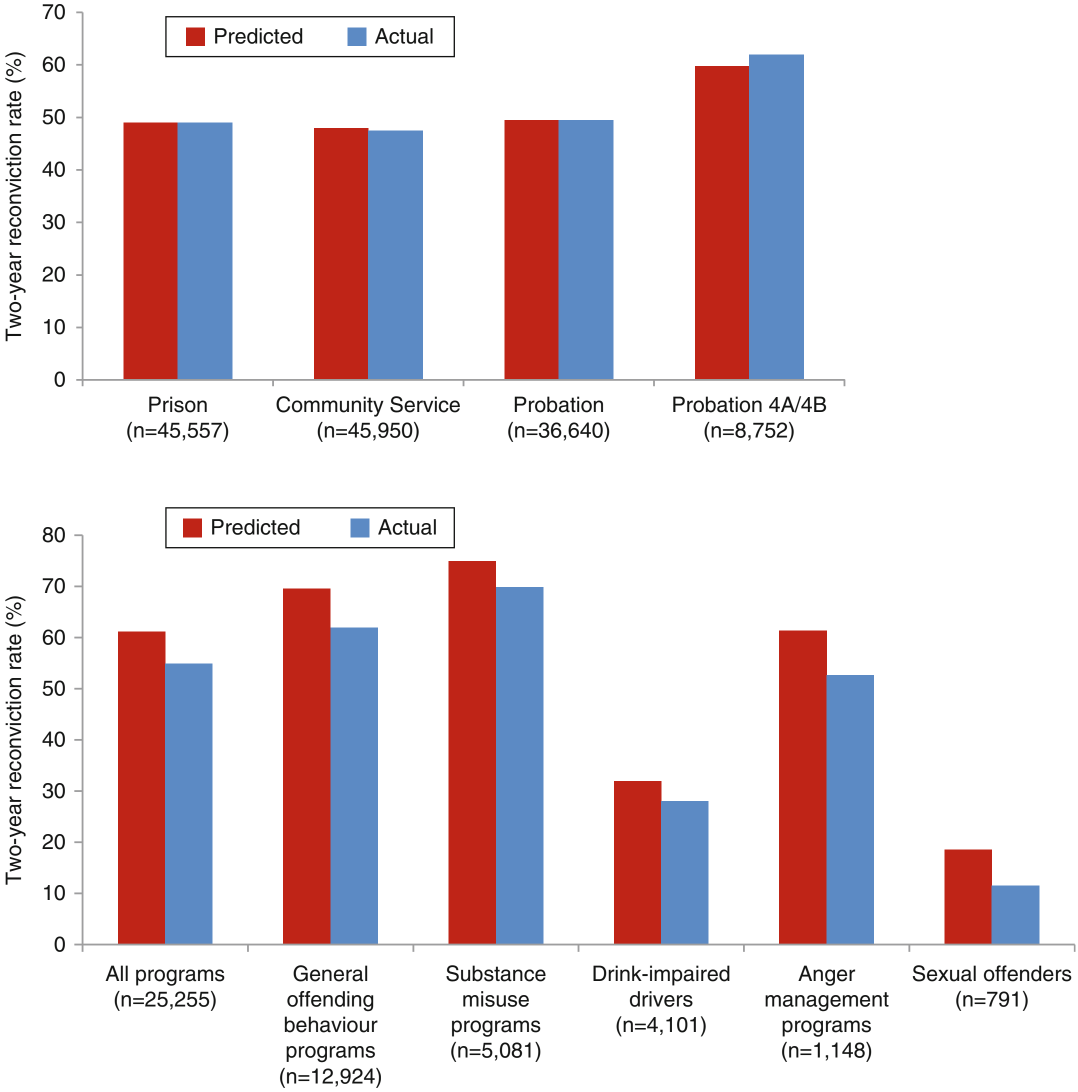

Effective Community Supervision Programming, Fig. 1 The *upper bar chart* shows predicted and actual 2-year reconviction rates for four types of sentence in England and Wales drawing on Home Office criminal statistics published in 1996 (Figure adapted from McGuire (2002)). The *lower bar chart* shows predicted and actual reconviction rates of participants in probation-based offending behavior programs (Based on data given by Hollis (2007))

Safety Canada, supervision alone was found to result in a reduction in recidivism of only 2 % points and effect size of zero for violent recidivism (Bonta et al. 2008).

The lower graph of Fig. 1 shows the findings of Hollis (2007) concerning the outcomes of probation sentences following the advent of structured programs. The sample size here is also quite large and the follow-up period is 2 years. Again, the predicted reconviction rates are based on use of the OGRS-2. Here the picture that emerges is different from that obtained as a result of sentences alone. For all offense types shown, for which specialized programs were available, there are significant reductions in actual as compared to predicted rates of reconviction. For program completers, the reductions were even larger: for general offending behavior programs, −26.4 %;

for substance misuse programs, −19.9 %; for drink-impaired driver programs, −28.3 %; for anger management programs, −25.5 %; and for sex offender programs, −61 % (Hollis 2007). While this is not a controlled experimental trial, it nevertheless offers a striking contrast to the effects of the sanction of supervision alone. In their evaluation of ISPs, Petersilia and Turner also found evidence that "…higher levels of program participation were associated with a 10–20 % reduction in recidivism" (1993, p. 315).

Side-by-side comparisons of offending behavior program participants with others not allocated to them have produced positive results in community settings. A major obstacle to evaluation however is the very high attrition rate among those allocated to attend, at least during the initial years following their inception. Fortunately in England and Wales there was evidence of significant improvement in this respect as programs became more firmly embedded in services, and completion rates doubled over a 5-year period (National Probation Service 2007). Evaluations of the probation "Pathfinder" programs showed that those who completed them were significantly less likely to re-offend than the comparison sample not allocated to attend (Hollin et al. 2008). While the evaluations could not be based on randomized experimental trials, questions arose as to whether this might merely be a selection effect. However, further analysis showed that the observed effects could not be explained solely in terms of prior features of the participants (McGuire et al. 2008), discerning a difference that was most likely an effect of treatment.

Another aspect of evaluation, increasingly emphasized as governments have proclaimed the need for gradually tightening constraints on public expenditure, is the relative cost of policies and practices when set alongside each other. Regarding criminal justice programs, however, the accumulated evidence suggests that the majority represent a net return on investment. The conventional statistic used for economic evaluation, the *benefit-cost ratio*, emerges as positive for the great majority of programs, and for some it is impressively high (Aos et al. 2001). Numerous studies have discovered superior benefit-cost ratios for community-based treatment of criminogenic risks and needs as compared with custodial sentences (McGuire 2012). Taking the United States as illustrative, for the fiscal year 2008, the annual *per capita* cost of imprisonment in a Bureau of Prisons facility was $25,894 and in a community correctional center $23,881. By contrast, the annual cost of supervision by a federal probation officer was $3,743 (U.S. Courts 2009). The differences between these figures indicates that, even allowing for increased costs of program delivery, there are adequate resources for larger numbers of offenders to be supervised in the community with program attendance as an ingredient of their sentences.

Nonprogrammatic Aspects

Results of these evaluations are nevertheless still weaker than those obtained in some "demonstration" studies. A possible explanation for this may reside not in the nature or effects of progams themselves, but in what have been called "nonprogrammatic" aspects of their implementation. This refers to a spectrum of issues including the use of well-validated assessment methods, appropriate allocation to programs and services, high-quality interactions between staff and participants, procedures for maintenance of program integrity, and provision of organizational support. There has been a consensus that these aspects of delivery and implementation have been neglected relative to the importance attached to programs themselves and that if insufficient attention is paid to them, especially when programs are diffused on a system-wide basis, integrity and consistency of delivery will be more difficult to maintain.

These areas have now been researched better, though given the intricacy and sometimes sensitivity of the issues to be explored, many questions remain unanswered. However, there is a steadily advancing grasp of the interconnections of program theory and design, on the one hand, and the factors that can optimize the chances of a program's success (or militate against them), on the other. There is firm evidence that contextual and other aspects of service delivery have a close association with outcomes. Andrews (2011) has reviewed the available evidence on the importance of these features, and there are persuasive grounds for accepting that if key aspects of implementation are marginalized or ignored, this can undermine the prospective positive effects of program participation.

There is evidence, for example, that the use of structured learning procedures; higher levels of staff interpersonal skill; effective use of modeling and problem-solving approaches; establishment of positive relationships; and use of effective reinforcement, disapproval, and authority where appropriate are all significantly associated with the effectiveness of programs. With reference to the relational dimensions, some primary studies in this area suggest that a balanced combination of usage of authority with openness and interpersonal responsiveness generates better effects than either an autocratic, rule-oriented style or a highly sensitive, but overly permissive approach. Such a balanced supervisory style was associated with lower rates of recidivism in evaluation of the New Jersey Intensive Supervision Program (Paparozzi and Gendreau 2005). However, it seems unlikely that such an integrated supervisory style is extensively practiced by probation staff, though proposals have been forwarded as to how training can be provided to encompass it. Of particular note is the model of *Strategic Training Initiative in Community Supervision* (STICS), developed in Canada and designed to translate the findings of the "what works" research into skills training for community supervisors (Bourgon et al. 2010).

Several recent reviews have underlined the importance of appropriate assessment and allocation to programs by risk level, the safeguarding of program quality, and adherence to the prescribed treatment model. There is evidence that when risk levels, model fidelity, and other factors are taken into account, programs based in the community achieve larger effects than those provided in institutional settings. Equally important, a "treatment-oriented" philosophy invariably proves superior to a punitive ethos: interventions based on "human service principles" and with features of best practice in this respect yield consistently better outcome results than deterrence-based practices (Lowenkamp et al. 2010).

Implications and Extrapolations

There is a body of knowledge of steadily growing magnitude concerning the possibility of altering offenders' lives, of enhancing community safety, and of reducing overall costs, through the provision of appropriate forms of intervention at the "tertiary" stage, that is, within criminal justice or correctional settings. Effective community supervision including structured programs focused on risk factors for offending has a sizeable role to play in bringing this about. One clear implication is for criminal justice or penal agencies to move away from a reliance on prison as the principal means of dealing with offenders. There are strong arguments instead for reserving it, as announced in so many promissory notes, to containing those who pose a serious threat to others or to themselves. Yet numerous datasets show that the prisons of many countries house large numbers of people who do not need to be there.

The continuing situation in which prison populations in some jurisdictions have swollen to unprecedented levels and resultant costs have spiraled, spent on a disposal route which evaluations continue to show is the least effective of the options available, is nothing short of remarkable. In the United Kingdom, carefully considered recommendations have been made by senior parliamentarians for a substantial reversal of the current trends and for a "reinvestment" in justice processes that have a more firmly anchored base in the community (House of Commons Justice Committee 2010). Evidence concerning the impact of programs should prove invaluable for any jurisdiction seeking to transfer resources and effort to reach a better balance of institutional and community-based provision. That holds, notwithstanding the sometimes weaker results of large-scale dissemination to date, which are almost certainly due to "failures of implementation" rather than to the contents or methods of programs in themselves (Andrews 2011). That is not of course to say that there is no room for improvement or no need to make further developments in the range of programs available. There are many questions still awaiting answers, and far more research is needed on these issues.

The concept of "justice reinvestment" can be interpreted broadly to provide a framework where the type of proposal being reviewed here can be incorporated as part of a wide portfolio of initiatives for transferring the burden of activity from custody to community. Allen (2011) suggests a series of seven possibilities, none of which would require any change in legislation: (a) increased use of diversion; (b) reduced use of remand, by providing alternative accommodation for homeless defendants; (c) making greater use of home curfew for others; (d) reducing the usage of short sentences which achieve very little in any case; (e) increasing the use of alternatives for those who breach community sentences; (f) reversing the steady decline in rates of parole; and (g) reducing rates of recall from parole. These can be set alongside and integrated with improvements in the delivery and content of probation and parole supervision. While as Kleiman (2011) suggests the latter has a dubious reputation in the eyes of many, based perhaps on the findings of Petersilia and Turner (1993), there are many more possibilities now in terms of what can be included within them. Unfortunately, in the past some had a track record of providing insufficient oversight, but Kleiman (2011) also describes far more efficient delivery and sanctioning strategies which can make a significant difference in these respects, illustrating this with reference to the very promising results of the HOPE (Hawaii's Opportunity Probation with Enforcement) program which showed greatly reduced rates of missed appointments, positive drug tests, new arrests, and probation revocation in a 1-year randomized controlled trial (Hawken and Kleiman 2009).

Another possible model is that of the Intensive Alternatives to Custody (IAC) scheme piloted in England and Wales. This entailed a high level of probation contact over an initial three month period, reducing steadily over a further nine months. Though not yet fully evaluated, preliminary findings from a process study (Wong, O'Keeffe, Ellingworth and Senior 2012) and a one-year outcome study (Khan and Hansbury 2012) are encouraging. Those eligible included relatively prolific (an average of 29 previous convictions), substance-abusing offenders whose lifestyles would otherwise have been chaotic. The rate of non-compliance was very close to that found in standard probation supervision where 25 % of orders are terminated for negative reasons (failure to comply with requirements or conviction for a further offence). The rate of reconviction at 12 months was somewhat lower (56.1 % as against 63.6 %) than that of a matched group given custodial sentences. Note that even if the latter rates were exactly the same, the cost saving from the IAC is considerable relative to imprisonment, and participants are able to maintain contact with their families.

Key Issues and Controversies

To conclude, it should be acknowledged that the generally positive account regarding the use of programs reflected in this entry is not universally shared. Programs have their critics and several reservations have been expressed concerning them. Some of the criticisms may be misplaced, for example, that programs are "not a panacea": it is difficult to find a source in which a claim was made that they were. The more cogent criticisms focus on the model or vehicle of change seen as central to how programs work and to aspects of their delivery in everyday practice.

For example, it has been held that the key mechanism in engendering individual change is the nature of the link between an offender and the probation worker, what may be called (transposing terminology from elsewhere) the "therapeutic alliance." But that had been the assumed vehicle of change in the numerous decades of probation work that preceded the advent of programs, and as described above there was a dearth of evidence concerning its impact. Furthermore, in the field of psychological therapy, while there is a consensus that a "working alliance" with the client may be a *necessary* condition of individual change, there is far less agreement concerning the extent of its importance or whether it amounts to a *sufficient* condition of change. Indeed based on systematic mining of the therapy process-outcome studies, even advocates for the importance of the alliance have conceded that it seems unlikely to account for more than 6–7 % of the variance in observed outcomes.

Possibly as a result of use of the word "treatment," offender programs have sometimes been typified as representing a species of medical model, in which the offending individual is conceptualized as deficient in some respect that has to be remedied, a pejorative attribution. It is argued that the individual is viewed as a passive recipient of things that professionals "do to" him or her. This is however a straightforward misunderstanding of the learning theory model that informs most programs. A deficiency of skill or of the exercise of habits of self-regulation is a product of accumulated learning experiences (or the absence of opportunities for them). Efforts to address this have more in common with the spirit of the classroom than that of the clinic, and like education they work best when the individual is an active participant in the process.

It has also been argued that the pathway to desistance from offending represents a personal journey and has to be understood as a process rather than something that can be realized by participating in a structured program. There are claims to the effect that recognition of this constitutes a "new paradigm." However, it is difficult to discern to what extent a proposal that involves a switch of emphasis can be held to constitute a new paradigm as such. There is nothing in the research literature on programs that conceptualizes individual change as an instantaneous event. Evidence on the changes that occur in individual offenders' lives as they attempt to avoid re-offending is perfectly compatible with the models that underpin community programs, which require honest self-appraisal, focused thought, and repeated practice to produce results.

Finally, at least in England and Wales, several criticisms have flowed from the fact that the dissemination of programs was conducted on a larger scale and faster pace than initially expected. Those challenges were compounded by the fact that the policy agenda described earlier was pursued alongside several other far-reaching organizational changes in parallel. The unfortunate result that programs are perceived as simply serving the purposes of centrally led micromanagement, and part of the broader trend towards heightened "managerialism," is in all likelihood a by-product of that process. There are however lessons that can be learned from this. Instigating too many changes simultaneously is a recipe for confusion and for damaging practitioner morale. The conjoined findings from programs themselves, and from the corpus of information now available on aspects of implementation, should ensure that efforts towards justice "reinvestment" can be more successful in future.

Further Reading

Edward J. Latessa and Paula Smith (2011) *Corrections in the Community* (5th edition, Cincinnati, OH: Anderson) is an excellent overview of the context and nature of work in this field and highlights some of the principal issues and controversies. For an overview of the background factors in theory and research that influence the design of programs and of outcome evidence regarding their effectiveness, see the book by Andrews and Bonta (2010), while for information on the rationale, contents, and outcomes of a range of specific types of program, see the volume edited by Hollin and Palmer (2006). The edited book by Joel A. Dvoskin, Jennifer L. Skeem, Ray W. Novaco, and Kevin S. Douglas (2012) *Applying Social Science to Reduce Violent Offending* (New York: Oxford University Press) challenges the current configuration of services in criminal justice and the overreliance on imprisonment, providing evidence-based and cogent arguments for a major change of policy. For a review of the "state of the science" in program implementation and evaluation, see the forthcoming book edited by Leam A. Craig, Louise Dixon, and Theresa A. Gannon, *What*

Works in Offender Rehabilitation: *An Evidenced-Based Approach to Assessment and Treatment* (Chichester: Wiley-Blackwell).

Related Entries

- ▶ Actualizing Risk-Need-Responsivity
- ▶ Effective Supervision Principles for Probation and Parole
- ▶ History of Probation and Parole in the United States
- ▶ Intensive Probation and Parole
- ▶ Justice Reinvestment
- ▶ Probation and Parole Practices

Recommended Reading and References

Allen R (2011) Justice reinvestment and the use of imprisonment: policy reflections from England and Wales. Criminol Public Policy 10:617–627

Andrews DA (2011) The impact of nonprogrammatic factors on criminal-justice interventions. Legal Criminol Psychol 16:1–23

Andrews DA, Bonta J (2010) The psychology of criminal conduct, 5th edn. Anderson Publishing, Cincinnati

Antonowicz DH (2005) The reasoning and rehabilitation program: outcome evaluations with offenders. In: McMurran M, McGuire J (eds) Social problem solving and offending: evidence, evaluation and evolution. Wiley, Chichester, pp 163–181

Aos S, Phipps P, Barnoski R, Lieb R (2001) The comparative costs and benefits of programs to reduce crime. Washington State Institute for Public Policy, Olympia

Bonta J, Rugge T, Scott T, Bourgon G, Yessine AK (2008) Exploring the black box of community supervision. J Offender Rehabil 47:248–270

Bourgon G, Bonta J, Rugge T, Scott T-L, Yessine AK (2010) The role of program design, implementation, and evaluation in evidence-based “real world” community supervision. Fed Probat 77:2–15

Correctional Services Accreditation Panel (2010) Report 2009–2010. Ministry of Justice, Correctional Services Accreditation Panel Secretariat, London

Craig LA, Dixon L, Gannon TA (eds) (2013) What works in offender rehabilitation: an evidenced-based approach to assessment and treatment. Wiley-Blackwell, Chichester

Dvoskin JA, Skeem JL, Novaco RW, Douglas KS (2012) Applying social science to reduce violent offending. Oxford University Press, New York

Gaes GG (1998) Correctional treatment. In: Tonry M (ed) The handbook of crime and punishment. Oxford University Press, Oxford, pp 712–738

Gendreau P, Ross RR (1980) Effective correctional treatment: bibliotherapy for cynics. In: Ross RR, Gendreau P (eds) Effective correctional treatment. Butterworths, Toronto, pp 3–36

Hawken A, Kleiman M (2009) Managing drug involved probationers with swift and certain sanctions: evaluating Hawaii’s HOPE. Office of Justice Programmes, National Institute of Justice, Washington, DC

Hollin CR, McGuire J, Hatcher RM, Bilby CAL, Hounsome J, Palmer EJ (2008) Cognitive skills offending behavior programs in the community: a reconviction analysis. Crim Just Behav 34:269–283

Hollin CR, Palmer EJ (eds) (2006) Offending behaviour programmes: development, application, and controversies. Wiley, Chichester

Hollis V (2007) Reconviction analysis of interim accredited programmes software (IAPS) data. Research Development Statistics, National Offender Management Service, London

House of Commons Justice Committee (2010) Cutting crime: the case for justice reinvestment. First report for session 2009–2010. The Stationery Office, London

Khan S, Hansbury S (2012) Initial analysis of the impact of the Intensive Alternatives to Custody pilots on re-offending rates. Research Summary 5/12. London: Ministry of Justice

Kleiman MAR (2011) Justice reinvestment in community supervision. Criminol Public Policy 10:651–659

Latessa EJ, Smith P (2011) Corrections in the community, 5th edn. Anderson Publishing, Cincinnati

Lipsey MW, Landenberger NA, Wilson SJ (2007) Effects of cognitive-behavioral programs for criminal offenders. Campbell Syst Rev. doi:10.4073/csr.2007.6

Lowenkamp CT, Flores AW, Holsinger AM, Makarios MD, Latessa EJ (2010) Intensive supervision programs: does program philosophy and the principles of effective intervention matter? J Crim Justice 38:368–375

Maguire M, Grubin D, Lösel F, Raynor P (2010) ‘What works’ and the correctional services accreditation panel: taking stock from an insider perspective. Criminol Crim Justice 10:37–58

McGuire J (2001) Defining correctional programs. In: Motiuk LL, Serin RC (eds) Compendium 2000 on effective correctional programming. Correctional Service Canada, Ottawa, pp 1–8

McGuire J (2002) Criminal sanctions versus psychologically-based interventions with offenders: a comparative empirical analysis. Psychol Crime Law 8:183–208

McGuire J (2005) The think first programme. In: McMurran M, McGuire J (eds) Social problem solving and offending: evidence, evaluation and evolution. Wiley, Chichester, pp 183–206

McGuire J (2008) A review of effective interventions for reducing aggression and violence. Philos T Royal Soc B 363:2577–2597

McGuire J (2012) Addressing system inertia to effect change. In: Dvoskin JA, Skeem JL, Novaco RW, Douglas KS (eds) Applying social science to reduce violent offending. Oxford University Press, New York, pp 265–290

McGuire J (2013) "What Works" to reduce reoffending: 18 years on. In: Craig LA, Dixon L, Gannon TA (eds) What works in offender rehabilitation: an evidence based approach to assessment and treatment. Wiley-Blackwell, Chichester, pp 20–49

McGuire J, Bilby CAL, Hatcher RM, Hollin CR, Hounsome JC, Palmer EJ (2008) Evaluation of structured cognitive-behavioural programs in reducing criminal recidivism. J Exp Criminol 4:21–40

National Probation Service (2007) Annual report for accredited programmes 2006–2007. National Offender Management Service Interventions and Substance Abuse Unit, London

Paparozzi MA, Gendreau P (2005) An intensive supervision program that worked: service delivery, professional orientation, and organizational supportiveness. Prison J 85:445–466

Petersilia J, Turner S (1993) Intensive probation and parole. Crime Justice Rev Res 17:281–335

US Courts (2009) Costs of imprisonment far exceed supervision costs. www.uscourts.gov/newsroom.2009/costsOfImprisonment.cfm

Weatherburn D, Triboli I (2008) Community supervision and rehabilitation: two studies of offenders on supervised bonds, Contemporary issues in crime and justice, number 112. New South Wales Bureau of Crime Statistics and Research, Sydney

Wong K, O'Keeffe C, Ellingworth D, Senior P (2012) Intensive Alternatives to Custody: Process evaluation of pilots in five areas. Ministry of Justice Research Series 12/12. London: Ministry of Justice

Yang M, Wong SCP, Coid J (2010) The efficacy of violence prediction: a meta-analytic comparison of nine risk assessment tools. Psychol Bull 136:740–767

Effective Supervision Principles for Probation and Parole

Jesse Jannetta[1] and William D. Burrell[2]
[1]Urban Institute, Washington, DC, USA
[2]Lawrenceville, NJ, USA

Overview

The content of parole and probation supervision, particularly in the United States, underwent an emphatic shift toward a focus on monitoring and surveillance starting in the mid-1970s. More recently, the field had focused increasingly on implementing supervision practices based on emerging evidence-based practices and the risk-need-responsivity model for changing offender behavior. However, changing everyday supervision practice to accommodate this emerging knowledge remains a challenge. Frameworks to translate this focus into concrete principles to guide the delivery of supervision have been created, but efforts to devise and test implementation models based on those principles are still in the early stages. Substantial questions remain about how to most effectively implement state of the art supervision practices at the agency level, what dosage and intensity of supervision is appropriate and effective for what types of supervisees, and what the effects will be of integrating new technologies into the practice of supervision.

This entry reviews the organizational context within which probation and parole supervision is delivered, examines the recent history of the field, presents a summary analysis of effective supervision strategies, and explores several unresolved issues related to supervision. It addresses only the supervision functions of probation and parole, although there are many other important services provided by probation and parole agencies.

Scope and Role of Probation and Parole

Probation and parole, referred to collectively as "community supervision," are often confused in terms of responsibility and function. In most US states, probation supervises individuals who have received community supervision in addition to or in lieu of an incarceration sentence, while parole supervises offenders who have been released from prison, either through discretionary release by a parole board or mandatory release onto supervision. Some states and the federal government have abolished the traditional discretionary parole release function and have renamed the supervision function.

Both probation and parole represent a form of "conditional liberty" in which an offender is in the community under the supervision of a probation and parole officer (PPO). The offender is subject to “conditions” of supervision imposed by the court or releasing authority. The conditions are obligations (e.g., report to a PPO and pay victim restitution) or restrictions (e.g., do not possess weapons or leave the jurisdiction without permission) on the offender. The responsibilities of probation and parole agencies generally fall into three categories: intervention, surveillance, and enforcement. Intervention involves providing probationers and parolees with access to treatment and other services intended to bring about behavior change. Surveillance is the monitoring of the environment around and behavior of supervisees. Enforcement entails the PPO monitoring compliance with the conditions of supervision and sanctioning noncompliance. Failure to comply can result in revocation of the conditional liberty and a sentence to jail or prison.

The United States has the highest incarceration rate in the world. However, incarcerated offenders represent only 30 % of the adults under correctional supervision in the United States according to the Bureau of Justice Statistics; the remaining 70 % are in the community under probation or parole supervision. This distribution has remained stable for more than three decades. One in 47 adults is on probation or parole. Probation is the most commonly imposed disposition for juvenile offenders, as well. Accurate national data on juvenile probation has only recently become available. The Census of Juvenile on Probation reports approximately 490,000 juveniles were on formal or informal probation in the United States in 2005.

The decentralized and fragmented nature of governmental structures in the United States makes it difficult to concisely describe the organizational context of probation and parole. Agencies at the federal, state, county, municipal, and tribal levels provide probation and parole supervision. There are private organizations that provide supervision as well, some nonprofit and others for profit. At each level of government, agencies are often split between the executive and judicial branches. Within a state, adult probation and parole may be combined in the department of corrections, while juvenile probation may be located at the county level within the courts. The possible combinations of branches and levels of government provide numerous possible configurations, most of which exist somewhere in the United States.

Despite carrying the majority of the correctional workload, probation and parole are not well understood by most elected officials, policy-makers, or the general public. This contributes to the chronically low funding levels for probation and parole and a general lack of political and public support. Probation and parole have also struggled with their mission over the past three decades, unable to articulate a compelling purpose and vision of public value created for the community.

Changing Currents in Parole and Probation Supervision

The history of the development of the mission of probation and parole supervision is critical to understanding the recent developments in terms of the principles of effective supervision and the challenges of implementing those principles. From their origins in the mid to late nineteenth century, both probation and parole were firmly based on the belief that with the proper guidance and assistance, offenders could be reformed while living in the community.

John Augustus, a Boston bootmaker, is credited with creating the modern concept of probation with his volunteer work in the Boston police court assisting both adults and juveniles that he deemed worthy of a second chance. In 1878, Massachusetts became the first state to authorize probation in statute. Other states followed and by the turn of the twentieth century, probation was an established component of the justice system (Latessa and Smith 2010). Zebulon Brockway, the superintendent of the Elmira, NY, reformatory, is credited with establishing the modern system of parole in 1877. As with probation, parole was designed to help reform inmates released from incarceration and was also firmly grounded on the rehabilitative principle. Parole also grew in popularity as the nineteenth century drew to a close and was soon also a common feature of the states' correctional systems. The development of the nation's first juvenile court in Cook County, IL, in 1899 signaled that juveniles were deserving of different treatment than adults, treatment that was overtly rehabilitative in nature. As a result, juvenile probation quickly developed in concert with the juvenile courts (Latessa and Smith 2010).

For the first century of probation and parole, rehabilitation was a core element of the mission and a prominent feature of practice. As the centennial of these institutions approached, things began to change. The firm commitment to offender treatment and rehabilitation was rolled back as the USA embraced a punitive, incarceration-driven response to crime. The end of the rehabilitative ideal came quickly, in large part due to the work of Robert Martinson, who published an article in the journal *The Public Interest* (1974), summarizing research on correctional treatment that he and two colleagues had conducted for the State of New York. While the results of the research have been debated extensively, the message that endured was that, in terms of treatment for offenders, "nothing works." This triggered a wholesale retreat from treatment by institutional and community corrections agencies.

As a result, the philosophy and mission of probation and parole changed profoundly, from a balanced approach featuring assistance in combination with monitoring and enforcement to an approach that was predominately about watching offenders and returning them to custody when they violated the conditions of their supervision. Yet the idea that parole and probation should have a rehabilitative component never disappeared entirely. A similar dynamic occurred in the United Kingdom, where probation in England and Wales became very surveillance and monitoring oriented, while probation in Scotland remained within the social work tradition (Robinson and McNeill 2010). The National Research Council (2008, 36) summarized the resulting mission uncertainty for parole in this way (which could apply just as well for probation): "Variations in agency mission and culture—as well as in the fundamental role of parole itself—point to the conflict for parole agents: they don't know whether they are expected to be law enforcement agents or social caseworkers. They have responsibility for enforcing the law and the conditions of parole, as well as assisting in released offenders' reintegration in society."

Over this period of mission uncertainty, the number of individuals that PPOs were responsible for supervising grew tremendously. From 1980 to 2009, a period that included the height of the "get tough" era and the war on drugs, the adult probation and parole population grew by 275 % according to the Bureau of Justice Statistics. This tremendous increase in community supervision populations, and the caseloads of individual PPOs, further hastened a move away from activities intended to assist offenders, which are often very time-consuming. A greater role was assumed by tools that could quickly and accurately determine adherence to conditions of supervision, particularly drug testing and electronic monitoring.

At the same time, questions emerged regarding the effectiveness of community supervision as a strategy. Intensive supervision programs were found to be ineffective, meaning that "more" supervision did not produce better outcomes, although results were more promising for approaches that combined monitoring and appropriate treatment (Petersilia and Turner 1993). Analyses also showed that parolees subject to supervision had rearrest rates no better than similar people released from prison without supervision (Solomon et al. 2005). Findings like these raised serious questions regarding the value community supervision delivered, at least as commonly practiced in the United States at that time.

While the USA was moving away from rehabilitation to more punitive approaches, a small group of Canadian psychologists and correctional treatment staff was continuing to conduct and collect research on correctional treatment. From this research, largely produced subsequent to Martinson's review, the Canadians found a growing number of programs that reduced recidivism, often by a substantial amount. In 1990, Don Andrews and his colleagues published an article based on this work that shifted the paradigm. Using meta-analysis, they introduced the principles of effective correctional treatment (Andrews et al. 1990). The first principles to be identified were the risk principle, the need principle and the responsivity principle. The risk principle stated that services should be directed toward higher-risk offenders. The need principle stated that treatment should target criminogenic needs, those issues that lead to criminal behavior. The responsivity principle states that cognitive-behavioral treatment methods should be employed with offenders, and interventions should be tailored to the offender's learning style, motivation, abilities, and strengths. Together, the risk, need, and responsivity principles are referred to as the RNR model, and adherence to the RNR model has consistently predicted intervention effectiveness in reducing re-offending (Andrews and Bonta 2007).

This new evidence supported the beliefs of many in the probation and parole field that Martinson's conclusions were wrong, that offenders can change and that correctional staff can help make that change happen. The research, labelled "what works" in response to Martinson's "nothing works", reinvigorated the dialogue about the purpose and processes of probation and parole supervision. In the 1980s, when "risk control" was in vogue, probation and parole sought to punish and control offenders, using tools such as intensive supervision programs, day reporting centers and house arrest with electronic monitoring. With the rising use of actuarial risk assessments to predict likelihood of re-offense in the 1990s, "risk management" became popular. As the evidence continued to grow that treatment of offenders worked, the concept of "risk reduction" through offender behavior change came to the fore as people began to absorb the findings of the *what works* research.

While corrections was using the term *what works* to describe the use of empirically proven strategies and techniques, other fields such as medicine and education adopted the term *evidence-based practices*, or EBP, to describe their efforts at ensuring that clinical services and classroom practices reflected the latest research. With the publication of a series of papers on implementing evidence-based practices in community corrections by the National Institute of Corrections, the shift to EBP reached community corrections (Bogue et al. 2004). While a commitment to EBP in community supervision, at least at the rhetorical level, has become widespread (Jannetta et al. 2009), the contrasting supervision paradigm that allows supervising officers to carry large caseloads through reliance on drug testing and surveillance technologies that make violations easy to detect and substantiate persists in many agencies.

Content of Effective Supervision

As the evidence basis for the effectiveness of correctional interventions grew, efforts began to enumerate general principles to guide the conduct of parole and probation supervision based upon the RNR framework and related research evidence on effective correctional interventions. These efforts were intended to present a package of mutually reinforcing strategies that would allow probation and parole to realize their potential to reduce re-offending. Principles defined by three prominent efforts are presented in Table 1. The first of was undertaken by the Reinventing Probation Council, a group of 13 veteran probation practitioners who met over 2 years to focus on "the systematic deficiencies of probation, and what strategies needed to be developed to revitalize probation" (Reinventing Probation Council 2000, 1). A similar effort to detail the ideal content of parole supervision was convened by the JEHT Foundation and the National Institute of Corrections in 2007 (Solomon et al. 2008). The Crime and Justice Institute (CJI) and the National Institute of Corrections put forth eight principles of effective intervention and detailed how they could be applied in community corrections (Bogue et al. 2004).

All three of these frameworks place reducing re-offending through changing offender

Effective Supervision Principles for Probation and Parole, Table 1 Parole and probation supervision strategies

Reinventing probation council strategies	CJI/NIC principles of effective intervention	13 Parole supervision strategies to enhance reentry outcomes
1. Place public safety first 2. Supervise probationers in the neighborhood, not the office 3. Rationally allocate resources 4. Provide for strong enforcement of probation conditions and a quick response to violations 5. Develop partners in the community 6. Establish performance-based initiatives 7. Cultivate strong leadership	1. Assess actuarial risk/needs 2. Enhance intrinsic motivation 3. Target interventions a. *Risk principle*: prioritize supervision and treatment resources for higher-risk offenders b. *Need principle*: target interventions to criminogenic (correlated to crime) needs c. *Responsivity principle*: be responsive to temperament, learning style, motivation, culture, and gender when assigning programs d. *Dosage*: structure 40–70 % of high-risk offenders' time for 3–9 months e. *Treatment principle*: Integrate treatment into the full sentence/ sanction requirements 4. Skill train with directed practice (e.g., use cognitive-behavioral treatment methods) 5. Increase positive reinforcement 6. Engage ongoing support in natural communities 7. Measure relevant processes/ practices 8. Provide measurement feedback	1. Define success as recidivism reduction and measure performance 2. Tailor conditions of supervision 3. Focus resources on moderate and high-risk parolees 4. Frontload supervision resources 5. Implement earned discharge 6. Implement place-based supervision 7. Engage partners to expand intervention capacities 8. Assess criminogenic risk and need factors 9. Develop and implement supervision case plans that balance surveillance and treatment 10. Involve parolees to enhance their engagement in assessment, case planning, and supervision 11. Engage informal social controls to facilitate community reintegration 12. Incorporate incentives and rewards into the supervision process 13. Employ graduated, problem-solving responses of violations of supervision conditions in a swift and certain manner

behavior as the primary focus of community supervision efforts, in contrast to the narrower focus on surveillance and enforcement that had come to typify community supervision. By doing so, they attempt to resolve the mission confusion that has bedeviled parole and probation agencies at least since the decline of the rehabilitative ideal. Articulation of protection of public safety as a primary goal is common among community supervision agencies, but few have put in place robust accountability mechanisms around recidivism reduction. Performance measurement in community corrections is largely episodic, tending to focus on routine activities rather than outcomes.

The backbone of current approaches to facilitating risk reduction through community supervision is the use of actuarial risk/need assessment instruments to establish relative likelihood of re-offending and identify dynamic criminogenic needs to target for intervention. The use of assessment instruments including a broad range of dynamic items (i.e., those that can be changed) related to established criminogenic needs is now common, although not universal (Jannetta et al. 2009). Major criminogenic need factors include history of antisocial behavior, antisocial personality pattern, antisocial associates, family and/or marital factors, school and/or work factors, leisure and/or recreation factors, and substance abuse.

Results of these assessments guide community supervision at both the case and strategic levels. At the case level, assessment results should inform the development of a supervision plan, or case plan, detailing the goals and objectives that the supervisee must meet. Such plans are "the backbone of the supervision process, the map for how agents and offenders will identify and solve offenders' problems" (Taxman et al. 2004, 31). These plans are a key method of ensuring that surveillance and treatment are combined as supervision techniques. This is crucial because a combination of surveillance and treatment is more effective at reducing recidivism than surveillance alone (Aos et al. 2006). It is also necessary to ensure that interventions facilitated by parole and probation target established criminogenic needs, because programs directed at non-criminogenic needs are ineffective. It is also important to embed assessment information in key tools of supervision, as the fact that assessments are conducted and the information is available to supervising officers is no guarantee that such information will actually be used in practice.

At the strategic level, assessment information is the foundation of the rational allocation of resources. In terms of persons, this means concentrating supervision and treatment resources on the people most likely to re-offend. Substantial allocation of resources is necessary to intervene with higher-risk offenders, because it is necessary to deliver high dosages of treatment, and to be prepared to address multiple criminogenic needs to improve their outcomes (Lowenkamp et al. 2006). It also steers intervention resources away from lower-risk offenders, whose outcomes can be worsened by intensive programming (Lowenkamp and Latessa 2004). The rational allocation of resources also refers to concentrating on the periods of time where the risk of failure is greatest, generally at the outset of the supervision period (National Research Council 2008). Additionally, allocation of resources includes initiating place-based supervision to focus on the locations in which risk factors for recidivism are more prevalent or simply where higher concentrations of supervisees are found.

Probation and parole supervision will have greater success in changing offender behavior if supervisees receive positive reinforcement. While many of the interventions needed to address criminogenic needs would likely require delivery by someone other than the supervising officer, enhancement of motivation and positive reinforcement can be delivered by the PPO. This includes providing incentives as well as utilizing techniques to enhance their motivation. Motivational interviewing, a technique to explore and resolve ambivalence about change and strengthen the motivation to change, has been widely adopted by correctional agencies, including community supervision agencies. Incentives should be more common than sanctions, with four positive reinforcements for every negative consequence (Taxman et al. 2004). Examples of incentives can include small things like bus tokens, certificates of achievement, or even verbal praise. More substantial incentives can include reduction of supervision reporting requirements and even discharge from supervision entirely. Involving parolees and probationers in developing their supervision and treatment plan is also a method for increasing their motivation to carry out the plan.

Working on enhancing motivation and supervisee buy-in to the goals of their plan requires attention to the responsivity component of the RNR model. Building a strong working relationship based on trust in the PPO to be firm but fair is an important element of effective supervision. It has long been established that elements of interaction between supervisees and supervising officers that are not programmatic in nature, such as officer decorum and interaction styles, have an important effect on offender outcomes (Palmer 1995). Behavioral management techniques, indicated by high levels of caring/fairness and trust, are associated with success, and toughness-authoritarianism approaches are associated with failure (Skeem et al. 2007).

Greater collaboration with external partners is a prerequiste for effective supervision. External partners include agencies that can assist with the monitoring and intervention aspects of supervision work, such as law enforcement and community service providers. For people coming onto supervision after release from incarceration, greater coordination with institutional corrections allows for a more seamless transition approach that fosters positive change. Informal pro-social supports such as family members, employers, and neighbors are also important informal partners to engage. Relationships with these informal social supports are typically more effective than formal controls such as parole and probation supervision in promoting positive individual change (National Research Council 2008).

Finally, supervision violations must be responded to in a way that is swift, certain, and seeks to solve problems. Swiftness and certainty in sanctions are more important the severity of sanctions in deterring unwanted behavior, yet sporadic use of the most severe sanction, return to incarceration, has been the norm in supervision. The concept of graduated sanctions is now a firmly established approach to dealing with violations to ensure that every violation meets with some meaningful response, but that returns to custody are reserved for the most serious violations. Sanctioning guidelines or matrices the incorporate a graduated approach to responding to violations are now common in the supervision field (Jannetta et al. 2009) and have been shown to generate positive outcomes (Martin et al. 2009). There is also evidence that swift and certain delivery of short jail stays as a sanction can deter violations, provided supervisees are carefully targeted and warned ahead of time (Hawken and Kleiman 2009).

Models to Incorporate Effective Supervision in Daily Practice

The work over the last decade or so to identify and implement EBPs for supervision has been challenging. Several jurisdictions have successfully implemented the EBP model and have had their success documented by outside evaluators (Taxman 2008; Fabelo 2009). Many other agencies have implemented some effective practices (Jannetta et al. 2009). Despite the seemingly wide agreement that EBPs are the preferred way to go, there remains a persistent reluctance on the part of many in the field to fully embrace the model and commit to its full implementation. This reluctance stems from two major issues – lack of consensus on the mission of supervision and implementation challenges.

The residue of the "nothing works" doctrine can be found in many probation and parole agencies. An entire generation of staff has grown up professionally in organizations that were unwavering in their commitment to a law enforcement/surveillance model of supervision. The staff never experienced the balanced approach with rehabilitation of the pre-Martinson era. In many jurisdictions, the post-Martinson generation includes the agency leadership. Some leaders question the research, others have concerns about the viability of the model and still others wonder whether such an approach can be sold to local elected officials who have conservative views on crime and justice. For these and other reasons, they are reluctant to undertake EBP. The issue is whether the leadership of probation and parole agencies are willing to make the necessary commitment to transform their organizations. Absent a clear organizational commitment to a mission that is based on risk reduction through behavior change, there is little impetus to adopt EBP.

Successful organizational change, such as adoption of EBP for probation and parole supervision, requires an effective implementation effort. Good programs will not produce good results if they are not implemented properly. This includes ensuring fidelity to the all of the elements of the original program design. The issue is whether probation and parole agencies have (or can develop) the staff capacity and leadership capability to accomplish large-scale, long-term organizational change such EBP. Most executives and managers in probation and parole agencies have long experience and expertise with the substance of their agency's core work; most do not have the experience with large-scale organizational development and change.

Fortunately, there are resources available to help. The National Institute of Corrections efforts in implementing EBP recognized this from the outset and subsequent efforts by NIC (Guevara et al. 2011) and the Council of State Governments Justice Center (Fabelo et al. 2011) have as well. Outside of the field of probation and parole, the National Implementation Research Network (NIRN) has been examining effective implementation strategies and techniques and offers assistance and resources.

Organizational commitment to the principles of effective supervision and EBP is not always sufficient to realign practice at the individual officer level. Even with training, officers may not base the supervision plan on the risk/need assessment and spend little time in meetings with supervisees addressing criminogenic needs (Bonta et al. 2008). Several operational models have been developed in an attempt to address this problem and ensure that supervision based on the principles of effective intervention is carried out consistently and with fidelity. There is some evidence that implementing such models can provide desired benefits even in the absence of fidelity, possibly by changing officer understanding of probationers and approach to violations that they incur (Harris et al. 2004).

Recent examples of operational approaches to supervision based on the principles of effective intervention indicate that they may be able to deliver both fidelity to the principles of effective supervision and improved outcomes. Implementation of Maryland's Proactive Community Supervision (PCS) had an impact on the way PPOs did their case planning, and offenders supervised through PCS had fewer warrants filed for violations and were on supervision longer before failure (when it occurred) than did a matched comparison group of non-PCS-supervised parolees and probationers (Taxman 2008). Probation officers trained in Strategic Training Initiative in Community Supervision (STICS) spent significantly more time in sessions with probationers focusing on criminogenic needs and procriminal attitudes than did control officers. Probationers supervised by STICS probation officers had lower recidivism and reconviction rates after 2 years (Bonta et al. 2010).

Current Issues and Future Directions

Dosage: How Much Is Enough?

Since probation and parole became organized governmental functions, the question of optimal caseload size has been discussed. How many offenders can/should a PPO supervise? This is a straightforward, but very complex question that cannot be answered easily. Recent research has shown that smaller caseloads with EBP can produce improved outcomes (Jalbert et al. 2011). This is a critical finding because experience and research have shown that EBP-based supervision models take more time. More substantive interaction is required of each contact between PPO and offender and that takes more time than the traditional check-in only type of contact.

The EBP research suggests benchmarks for hours of correctional treatment of offenders by risk level, with higher-risk offenders receiving increased amounts of treatment or "dosage" (Bogue et al. 2004). Similar research is required to determine if there is a reliable metric that will suggest how much PPO time is required to be effective. Such inquiries should also explore whether there are optimal levels in terms of frequency of contact, length of contact and length of supervision. Such metrics would be invaluable in determining the optimal caseload size for different types of offenders and levels of supervision. These inquiries need to account for the content of supervision as well. Research on intensive supervision programs proved that devoting more PPO time to working with offenders, but on surveillance and monitoring activities, is not effective. Unpacking the dosage question will require attention to both the type of supervisee and the content of the supervision, changing the question to the more nuanced, "How much of what kind of supervision-based intervention do we need to deliver to what kind of person?"

Responding to Violations of Conditions of Supervision

One of the core elements of the original designs of both probation and parole was that there would be consequences for an offender's failure to comply with the conditions of supervision.

The challenge posed to probation and parole staff (and the judges and parole officials who must decide whether to revoke supervision) is that not all instances of noncompliance are equivalent. Violations of conditions vary in their seriousness, frequency, and legal status. Committing a new crime is a violation, as is failing to attend mandated treatment or missing a scheduled appointment with a PPO. The latter are known as “technical violations” as they do not represent criminal activity.

Responding to violations has been one of the weakest areas of probation and parole practice. Historically, PPOs have had but one tool in their kit, filing a violation of probation or parole and returning the offender to the court or paroling authority for a revocation hearing and potentially, a sentence of incarceration. Some PPOs were reluctant to file a violation for less serious infractions, often sending the message that such minor violation behavior would be tolerated. In the get tough era of the 1980s and 1990s, many agencies adopted “zero tolerance” policies for violations. These resulted in increased numbers of offenders being sent to prison for technical violations of probation or parole.

Finding effective responses to violations of probation and parole is important for several reasons. First, Bureau of Justice Statistics data shows that more than a third of the new admissions to state prisons in the USA consist of parole (primarily) and probation violators, many of whom have committed technical violations. Prison is a very expensive and harsh response that should be reserved for the most serious offenders. Research and experience has shown that many of these violators can be safely managed in the community at a much reduced cost. The second reason that alternatives to revocation are important has to do with the conventional wisdom among many in criminal justice, including probation and parole, that technical violations are a reliable precursor to criminal activity. Countless violations of probation or parole are filed as preventative measures, and thousands of offenders are sent to prison and jail. There is no reliable evidence to support this conventional wisdom, and recent research from Washington State found no reduction in new criminal activity from confining technical violators (Drake & Aos, 2012). Substantial research suggests that swift, certain and measured sanctions short of revocation appear to be effective with technical violations.

There is an operational (or political) challenge and a research challenge here. The operational challenge is to provide a robust set of options for responding to violations beyond doing nothing or returning to custody. Otherwise, there will be too much use of both. The research challenge is to illuminate the relationship, if any, between technical violations of various kinds and the criminal offenses that PPOs have traditionally revoked supervision in an attempt to prevent.

Integrating Desistance Theory

In its thorough review of the state of knowledge regarding parole, desistance from crime, and community integration, the National Research Council (2008) noted the importance of explaining gaps between research findings on what influences desistance and evaluation findings regarding program effects. Desistance theory has been minimally integrated into supervision interventions at this point in time, with supervision interventions mostly based on the theories of individual behavior change that underlie the RNR model. Longitudinal research on desistance highlights specific conditions that lead to less offending: particularly good and stable marriages and strong ties to work. These findings struck the National Research Council (2008, 2–3) as “somewhat at odds with findings from program evaluation that individual-level change, including shifts in cognitive thinking, education, and drug treatment, are likely to be more effective than programs that increase opportunities for work, reunite families, and provide housing.” Although the probation and parole field has long way to go to fully conform to the RNR model, exploring a next generation of supervision practice that more fully integrates the findings of desistance theory could further enhance the effectiveness of the field.

Technology-Based Supervision

How can technological approaches such as GPS help promote desistance? The evidence on the effects of electronic monitoring on supervisee success are mixed (Padgett et al. 2006; Finn and Muirhead-Steves 2002). GPS and other monitoring technologies, including minimal-level reporting mechanisms for low-risk offenders such as reporting kiosks and telephone reporting, are proliferating rapidly in the field. Location monitoring technologies such as GPS have the potential to increase the focus on monitoring and surveillance, as they make it possible to verify compliance with conditions such as curfews as easily as drug testing made it to verify compliance with conditions related to drug and alcohol use. How new supervision technologies could potentially support supervision to change offender behavior is unclear. Further investigation of what types of electronic monitoring and technology-based supervision approaches are effective (or not) for what types of supervisees is a crucial task for supervision researchers.

Conclusion

The emergence of the evidence base underlying the RNR model ended a period of great pessimism regarding the possibility that corrections agencies, including parole and probation agencies, could change offender behavior and thereby make communities safer. Although implementation of supervision practices conforming to the RNR model remains uneven at best, the field appears to have clarity regarding where it needs to go. While probation and parole continue to work on the organizational change necessary to do so, the research community can support them by better exploring the nuance of responsive supervision. What kinds of supervision work best for what kinds of offender, and at what dosage? What kinds of supervision violations need to be responded to in what way? Given the central role community supervision plays in the correctional system, aligning its practices with what is already known to be effective while continuing to build knowledge about effectiveness has the potential to make a huge contribution to public safety.

Related Entries

- ▶ Behavioral Management in Probation
- ▶ Desistance and Supervision
- ▶ Intensive Probation and Parole
- ▶ Parole and Prisoner Reentry in The United States
- ▶ History of Probation and Parole in the United States
- ▶ Probation and Parole Practices
- ▶ Probation Officer Decision-Making
- ▶ Reentry

Recommended Reading and References

Andrews DA, Bonta J (2007) Risk-need-responsivity model for offender assessment and rehabilitation. Public Safety Canada, Ottawa

Andrews DA, Zinger A, Hoge RD, Bonta J, Gendreau P, Cullen FT (1990) Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. Criminology 28:369–404

Aos S, Miller M, Drake E (2006) Evidence-based adult corrections programs: what works and what does Not. Washington State Institute for Public Policy, Olympia

Bogue B, Campbell N, Carey M, Clawson E, Faust D, Florio K, Joplin L, Keiser G, Wasson B, Woodward W (2004) Implementing evidence-based practice in community corrections: the principles of effective intervention. National Institute of Corrections, Washington, DC

Bonta J, Rugge T, Scott T, Bourgon G, Yessine AK (2008) Exploring the black box of community supervision. J Offender Rehabil 47:248–270

Bonta J, Bourgon G, Rugge T, Scott T, Yessine AK, Gutierrez L, Li J (2010) The strategic training initiative in community supervision: risk-need-responsivity in the real world. Public Safety Canada, Ottawa

Drake EK, Aos S (2012) Confinement for Technical Violations of Community Supervision: Is There an Effect on Felony Recidivism? Olympia, Washington State Institute for Public Policy.

Fabelo T (2009) Probation evidence-based practices: a strategy for replication. Presentation to the Texas department of criminal justice judicial advisory council. Council of State Governments Justice Center, New York, justicereinvestment.org/files/**Texas**JudicialCouncilPresentation.pdf

Fabelo T, Nagy G, Prins S (2011) A ten-step guide to transforming probation departments to reduce recidivism. Council of State Governments, New York

Finn MA, Muirhead-Steves S (2002) The effectiveness of electronic monitoring with violent male parolees. Justice Q 19:293–312

Guevara M, Loeffler-Cobia J, Rhyne C, Sachwald J (2011) Putting the pieces together: practical strategies for implementing evidence-based practices. National Institute of Corrections, Washington, DC

Harris PM, Gingerich R, Whittaker TA (2004) The "effectiveness" of differential supervision. Crime Delinq 50:235–271

Hawken A, Kleiman M (2009) Managing drug involved probationers with Swift and certain sanctions: evaluating Hawaii's HOPE. Report submitted to the National Institute of Justice

Jalbert S, Rhodes W, Kane M, Clawson E, Bogue B, Flygare C, King R, Guevara M (2011) A multi-site evaluation of reduced caseload size in an evidence-based practice setting. Abt Associates, Cambridge

Jannetta J, Elderbroom B, Solomon A, Cahill M, Parthasarathy B, Burrell WD (2009) An evolving field: findings from the 2008 parole practices survey. Urban Institute, Washington, DC

Latessa EJ, Smith P (2010) Community corrections, 4th edn. Matthew Bender & Co, New Providence

Lowenkamp CT, Latessa EJ (2004) Understanding the risk principle: how and why correctional interventions can harm low-risk offenders. Topics in community corrections. National Institute of Corrections, Washington, DC

Lowenkamp CT, Latessa EJ, Holsinger AM (2006) The risk principle in action: what have we learned from 13,676 offenders and 97 correctional programs? Crime Delinq 52:77–93

Martin B, Van Dine S, Fialkoff D (2009) Ohio's Progressive sanctions grid: promising findings on the benefits of structured responses. Perspectives 33:32–39

Martinson R (1974) What works? Questions and answers about prison reform. Publ Interest 35:22–45

National Research Council (2008) Parole, desistance from crime, and community integration. Committee on community supervision and desistance from crime. Committee on law and justice, division of behavioral and social sciences and education. The National Academies Press, Washington, DC

Padgett KG, Bales WD, Blomberg TG (2006) Under surveillance: an empirical test of the effectiveness and consequences of electronic monitoring. Criminol Public Policy 5:61–92

Palmer T (1995) Programmatic and nonprogrammatic aspects of successful intervention: new directions for research. Crime Delinq 41:100–131

Petersilia J, Turner S (1993) Evaluating intensive supervision probation/parole: results of a nationwide experiment. National Institute of Justice, Washington, DC

Reinventing Probation Council (2000) Transforming probation through leadership: the "broken windows" model. Manhattan Institute, New York

Robinson G, McNeill F (2010) Probation in the United Kingdom. In: Herzog-Evans M (ed) Transnational criminology manual. Wolf Legal Publishers, Oisterwijk

Skeem JL, Louden JE, Polaschek D, Camp J (2007) Assessing relationship quality in mandated community treatment: blending care with control. Psychol Assess 19:397–410

Solomon AL, Kachnowski V, Bhati A (2005) Does parole work? Analyzing the impact of postprison supervision on rearrest outcomes. Urban Institute, Washington, DC

Solomon AL, Osborne JWL, Winterfield L, Elderbroom B, Burke P, Stroker RP, Rhine EE, Burrell WD (2008) Putting public safety first: 13 parole supervision strategies to enhance reentry outcomes. Urban Institute, Washington, DC

Taxman FS (2008) No illusions: offender and organizational change in Maryland's proactive community supervision efforts. Criminol Public Policy 7:275–302

Taxman FS, Shepardson ES, Byrne JM (2004) Tools of the trade: a guide to incorporating science into practice. National Institute of Corrections, Washington, DC

Effectiveness of Situational Crime Prevention

Kate J. Bowers[1] and Rob T. Guerette[2]
[1]Department of Security and Crime Science, University College London, London, UK
[2]School of Criminal Justice, Florida International University, Miami, FL, USA

Overview

This entry summarizes the evidence on the effectiveness of situational crime prevention measures. It uses existing reviews of the evaluation literature to draw broad conclusions about the types of situational measures that show promise and identifies some of the mediating factors in determining levels of effectiveness. It also defines and summarizes evidence on other outcomes that evaluations sometimes considered such as cost-effectiveness, sustainability, anticipatory benefit, and displacement/diffusion of benefits. The principles behind process evaluation, which examines how a scheme was implemented, are also outlined. The second part of this section discusses two different approaches to synthesizing evidence on what works – the approach taken by those who advocate detailed consideration of the validity of the evaluation research design and synthesis through meta-analysis (e.g., the Campbell collaboration) and those advocating a realist approach, which uses all relevant evidence and considers how distinct contexts and mechanisms impact on outcome (e.g., Pawson 2006). The entry concludes with an account of what still remains to be done in evaluating SCP, including gaps in our current knowledge and future methodological developments.

Introduction

Situational crime prevention (SCP) entails efforts which seek to alter the situational environment in an effort to disrupt the frequency and harmful effects of reoccurring crime behavior. Unlike other approaches which seek to reduce crime by focusing solely on offenders, the situational prevention approach recognizes the importance which the environment plays in facilitating and, in some cases, causing crime events. Environments where prevention measures can be implemented include not only geographical places but also include adaptations to manufactured products and electronic systems which readily present opportunities for offending. Among geographical environments, 11 general types have been identified (Clarke and Eck 2003), though the most common of these include residential, public ways, recreational, retail, and transport.

Several challenges exist when evaluating situational crime prevention efforts. Some of these are common in applied research settings generally, while others are unique to the evaluation of situational prevention efforts. In part, these challenges stem from the environmental orientation of situational crime prevention. Conventional research methodology to a large extent originated from research which focused on the individual level. The application of these same research methods to the evaluation of place, product, and system-based interventions is not always compatible. At best, conventional methods require adaptation when used in environmental settings. At worst, they are simply not applicable.

One challenge facing situational crime prevention evaluation is that SCP is a process rather than an intervention(s) in and of itself. This process relies on an action research methodology which calls for (i) an identification of problems, (ii) in-depth analysis of the factors that contribute to and facilitate the problem behavior, (iii) identification of suitable responses and a determination of those which are most feasible, (iv) implementation of those responses, and (v) an assessment of their impact. Inherent in this process is that the implementation of situational measures will be specifically tailored to the unique circumstances of the problem at hand. This means that problem profiles will likely differ across time, geographical settings, products, and systems. What is devised to work in one location or in response to one problem will likely not work at other places or in response to other problem types. Accordingly, random assignment of interventions, as required in classical experiments, runs counter to the deliberative process of situational crime prevention since it calls for blindly applying interventions at random.

A second challenge is that many situational prevention schemes rely on intervention "packages" consisting of multiple prevention measures. For instance, an initiative to reduce residential burglary may include an assortment of tactics such as a media campaign, target hardening measures including the use of security alarms and the fortification of dwelling entry points, and improved lighting, among others. This makes it difficult to identify the role of any single situational intervention as being responsible for any observed reductions in crime. At best, what can be revealed by any evaluation is that the assortment of situational alterations resulted in some change. To compound matters, a third challenge is that most evaluations of situational crime prevention initiatives are conducted retrospectively after the intervention has already taken place, thereby making random assignment impossible.

A fourth challenge is that despite situational crime prevention having a formal theoretical foundation and action framework for understanding crime problems and formulating situational interventions (Clarke 1992), many evaluations of situational prevention efforts include those which were not informed by the framework. While the underlying mechanisms for how and why those situational interventions should work can be understood by the SCP approach despite their ex parte implementation, the remaining caveat is that the situational interventions may not be accurately suited to the specific crime problem that they were intended to address. The intervention may be incomplete or simply irrelevant to the nature of the problem. This presents the possibility of producing false negatives (erroneous conclusions of ineffectiveness) when trying to determine the impact of the situational interventions that were implemented.

A fifth challenge is that there are a variety of developing concepts such as displacement (the movement of crime to other places, times, tactics, targets, or offenses), diffusion of benefits (a reduction in crime surrounding intervention areas), anticipatory benefits (the premature reduction in crime prior to the implementation of an intervention), and questions of cost-effectiveness (an assessment of the gains achieved by an intervention in relation to the expended cost) which also have to be accommodated in any comprehensive evaluation of situational crime prevention schemes. Each of these requires different research designs if they are to be adequately assessed. To study spatial displacement and diffusion effects, for instance, requires the use of a buffer zone surrounding treatment areas in addition to conventional comparison areas. For an assessment of anticipatory benefits, a time series design must be utilized in order to reveal any premature reductions in crime prior to the intervention having been put in place. Lastly, assessments of cost-effectiveness require determinations of the expenditures involved in the implementation of the prevention scheme as well as some of accounting of crimes prevented and their associated cost savings.

These challenges explain, to some degree, the distribution of research methods that have been commonly used in the evaluation of situational crime prevention to date. A review of SCP evaluations (see Guerette 2009) used over the last nearly four decades found that the majority of studies used some variety of a quasi-experimental design with many using a combination of a time series and before-after design with comparisons. Only 3 % (n = 6) of the evaluations relied upon randomized experimentation.

Despite these challenges, considerable progress has been made in understanding the effects of situationally focused crime reduction schemes. The following synthesizes the current state of situational crime prevention research:

- On the aggregate, the majority of SCP efforts are reportedly effective according to study authors. Roughly three out of four evaluations indicate beneficial reductions in crime (Guerette 2009).
- The effectiveness of SCP holds true even after adjusting for displacement and diffusion effects (Guerette and Bowers 2009; Hesseling 1994; Eck 1993).
- Diffusion of benefits is just as likely to occur as displacement (Guerette and Bowers 2009; Guerette 2009).
- SCP effects are often compounded by anticipatory benefits. Smith et al. (2002) found that in some cases crime reduction can be observed before the official start date of a situational prevention scheme.
- Situational prevention measures appear to have more enduring reduction effects compared to focused policing operations (Braga and Bond 2008).
- The large majority of SCP efforts evaluated comprised techniques which either increase the effort necessary to commit crime and/or increase the risk involved (Guerette 2009).
- Most evaluations of SCP rely on quasi-experimental designs. Randomized experiments remain rare (Guerette 2009).

- SCP measures have most commonly been implemented and evaluated in four place types: residential, retail, public ways, and transport (Eck and Guerette 2012; Guerette 2009).
- On the aggregate cost assessments of SCP measures are less common (less than half of evaluations compute or allow for such determinations (Guerette 2009)), though some focused reviews have found them to be cost efficient (e.g., Welsh and Farrington 2008a).
- Little is known about the sustained impact of SCP measures beyond 1 or 2 years. Most evaluations (nearly two-thirds) of SCP have post-evaluation follow-up periods of 24 months or less (Guerette 2009).

The following sections present in more detail the prevailing findings regarding the evaluation of situational crime prevention initiatives. The sections draw heavily from existing systematic reviews of situational crime prevention schemes and report on the nature of SCP evaluation, some emerging findings of effectiveness, and discuss what is known about the unique concepts involved in SCP evaluations, namely, those of displacement and diffusion, anticipatory benefits, sustainability, and cost-effectiveness.

The Current Evidence on What Works

A small body of systematic reviews of situational crime prevention evaluations has consistently found them to be effective. These reviews have examined SCP evaluations in a variety of ways. Some have assessed the collective conclusions reported by study authors, while others have sought to examine findings disaggregated by place and intervention types. Still others have examined more narrowly specific intervention types, such as CCTV and street lighting. Reviews of evaluations are always hampered by the underlying methodological rigor of the evaluation studies which are reviewed. Generally, those with stronger research designs are viewed as more trustworthy, and many reviews place thresholds on the levels of design rigor that will be included in the review. Despite the variability in the types of research designs and the focus of the review, on the aggregate, evaluations of SCP suggest that situational alterations at problem places can effectively reduce crime.

In a descriptive review of conclusions drawn by study authors of 206 SCP evaluations (Guerette 2009), it was found that three out of four (75 %, n = 154) concluded that the intervention was effective overall. Twelve percent (n = 24) concluded that the situational intervention was not effective, while 6 % (n = 12) reported mixed findings, and 8 % (n = 16) of the study outcomes were inconclusive. Relying on this same sample of evaluations, another review examined study findings across five common place types (Eck and Guerette 2012) resulting in a subsample of 149 evaluations. In that review variation in effectiveness across place types was found; however, for all place types, effectiveness was reported by at least 60 % of the evaluations (see Table 1). Though all recreational interventions appeared successful, the small number makes any meaningful conclusion unreliable, since more evaluations will likely find negative effects. Considered differently, situational efforts in public outdoor settings and in residential places had the most failures, but even in those settings, ineffective interventions were comparatively small compared to those found to be effective and those with mixed results.

One interesting finding in this review was that situational interventions at public places compared to other settings appeared less successful. There are many possible reasons for this, including (i) public places are more open and less contained; (ii) they may host a variety of crime behaviors some of which are not expected to be prevented by the situational measure; (iv) their implementation may be flawed; and (v) the techniques used were “off the shelf” and not informed by careful analysis of the targeted crime problem.

An examination of the success rate of commonly used interventions also found them to be mostly effective, although with some variation across intervention type. Table 2 presents interventions that were used in decreasing order of frequency (totals column) and shows the

Effectiveness of Situational Crime Prevention, Table 1 Effectiveness of place-based intervention evaluations by common place types

Place type	Percent of authors' conclusions (n)			
	Effective	Not effective	Mixed findings	Inconclusive
Residential (39)	77 (30)	10 (4)	10 (4)	3 (1)
Public ways (52)	62 (32)	12 (6)	19 (10)	8 (4)
Retail (25)	88 (22)	4 (1)	4 (1)	4 (1)
Transport (26)	88 (23)	0 (0)	8 (2)	4 (1)
Recreational (7)	100 (7)	–	–	–
Total (149)	77 (114)	7 (11)	11 (17)	5 (7)

Effectiveness of Situational Crime Prevention, Table 2 Effectiveness of the most used interventions

	Authors' conclusions % (n)				
Intervention	Totals n (%)	Effective	Not effective	Mixed	Inconclusive
CCTV	25 (37)	59 (22)	14 (5)	24 (9)	3 (1)
Lighting	14 (20)	55 (11)	15 (3)	15 (3)	15 (3)
CPTED	11 (16)	94 (15)			6 (1)
Mixed/other	10 (16)	93 (14)	7 (1)		
Access control	9 (14)	92 (13)			8 (1)
Place management	6 (9)	89 (8)		11 (1)	
Street redesign	4 (6)	67 (4)		17 (1)	17 (1)
Total	79 (118)	74 (87)	7 (8)	12 (14)	6 (7)

percent of each intervention type demonstrating different outcomes (effective, not effective, mixed, inconclusive). Seventy-nine percent of the 149 studies reviewed entailed the use of just seven situational intervention techniques. Interestingly, though most were found effective, the most frequently used interventions showed the lowest rate of effectiveness. Similar to before, this may be the consequence of the tendency to use "off-the-shelf" techniques rather than those tailored to the specific problem at hand.

It is important to note that these findings are drawn from the study authors' conclusions and must be taken with consideration of the types of evaluation methods that were used. Stronger evaluation designs allow for more confident determinations that it was in fact the situational intervention that was responsible for any observed crime reduction, while weaker evaluation designs leave open the possibility that something else caused the change. Because of this only gross determination of situational crime prevention effectiveness can be concluded.

In separate meta-analytic reviews of the two most common SCP techniques, closed circuit television (CCTV), and street lighting, similar positive findings were found (Welsh and Farrington 2008a, b). Using meta-analytical techniques, a review of 41 evaluations of CCTV found that overall the technique produced moderate significant reductions in crime though this too varied by place type. CCTV was found to be most effective when used in parking lots to reduce vehicle-related crime. The findings also revealed that they tended to be more effectively applied in the United Kingdom than in the United States. Using similar methods, a meta-analysis of 13 evaluations of the effects of improved street lighting found that they too achieved significant reductions in crime. This reduction was not restricted to nighttime crimes (which would be predicted to be most influenced by improved street lighting). Thus, a form of temporal diffusion of benefits was observed.

State of the Art: Measuring Outcomes Other Than Crime Reduction

As mentioned in the introduction, there are outcomes other than whether or not crime has been reduced in the area of intervention that are of interest to crime prevention practitioners. These are designed to answer the following questions: Is the intervention good value for money? How long does the crime reductive effect of an intervention last? Does the intervention shift the crime problem elsewhere? Particularly comprehensive evaluations attempt to answer these questions, but evidence here is limited because the data and the methods required in undertaking such research are complex. This section will summarize some of the issues and evidence where it exists under each of these headings, respectively.

Is the Intervention Good Value for Money?

In order to answer this question, a cost-benefit analysis is necessary (Roman and Farrell 2002). This involves estimating the cost of the intervention in monetary terms. This is no simple undertaking as there are many costs associated with setting up and running a scheme. These costs include personnel, transport, equipment, premises, training, and advertising. Ideally the costs should include everything from employment of full-time personnel and substantial hardware (such as CCTV cameras or alley-gates) to the cost of fuel and office supplies. A further complication is how to cost in-kind contributions or voluntary personnel. For example, a scheme might be (in fact is nearly always) introduced in an environment where personnel are already working on crime prevention tasks, so they might purely absorb the further activity without the need for new staff; however, there is still an opportunity cost associated with the time that they devote to the new project. There is also the matter of approximating local costs for personnel and staff; there is likely to be considerable variation in costs by region and to be accurate analysis should use local rather than national cost estimates. In order to reliably collect cost information, evaluators need to gather information from the very beginning – it is virtually impossible to get a reliable estimate of cost retrospectively.

Once the costs of the scheme itself have been collected, these costs need to be balanced against the financial benefit of the crimes that have been averted as a consequence of the preventative activity. This involves obtaining two estimates: one quantifying the number of crimes of each type that have been prevented and the other quantifying the costs that would result if those crimes had occurred. Once more, acquiring this information is no mean feat. The outcome effect size (e.g., the percentage reduction in crime) needs converting to a count of crimes prevented (see, e.g., Johnson et al. 2004). Following this, these counts need to be multiplied by the associated financial cost to society that would have been incurred as a result of the crime and summed to get an estimate of overall financial benefit. Data on the costs of crime are available for some countries. For example, the UK Home Office have estimates available for 2005 (Dubourg et al. 2005) for a range of crime types. The costs are national ones and include costs in anticipation of crime (e.g., security and insurance expenditure), as a consequence of crime (e.g., property stolen, lost output, and emotional/physical impact), and in response to crime (e.g., the police costs and those of the criminal justice system). For reference, this source estimates the cost of a common assault at £1,440, a dwelling burglary at £3,268, and a homicide at just over £1.4 million. There are obviously questions about these data in terms of how the estimates are made (particularly the more intangible costs such as emotional impact), how the costs change over time, and how useful national costs are at the local level.

There have been surprisingly few (published) studies that undertake cost-benefit analysis of situational crime prevention measures. Of note are the street lighting evaluation by Painter and Farrington (2001), the Kirkholt project evaluation (Forrester et al. 1990), an evaluation of an

Effectiveness of Situational Crime Prevention, Table 3 Cost-benefit analysis of situational measures

Study	Description	Benefit-cost ratio
Painter and Farrington (2001)	Installation of improved street lighting in Dudley and Stoke-on-Trent in the UK	A saving of $6.19 for every dollar spent in Dudley and a saving of $5.43 for every dollar spent in Stoke
Forrester et al. (1990)	Multi-prevention approach to reducing residential burglary on the Kirkholt housing estate. Included strategies focusing on victims and repeat victimization and target hardening measures	A saving of £5.04 for every pound spent
Clarke and McGrath (1990)	Robbery reduction intervention in betting shops. Target hardening measures such as fitting safes with time locks were introduced	A saving of $1.71 was estimated for every dollar spent
Skinns (1998)	Installation of CCTV cameras in Doncaster, UK	A saving of £3.50 was estimated for every pound spent
Gill and Spriggs (2005)	Covered seven programs in a national evaluation of CCTV cameras	Four of the schemes were found to be cost-effective; the other three were not (in other words the costs outweighed the benefits)
Bowers et al. (2004a)	Installation of alley-gates to the rear of terrace properties in Liverpool in the UK	After a year of installation, the gates became cost beneficial with an estimated saving of £1.86 for every pound spent

anti-robbery target hardening project (Clarke and McGrath 1990), two CCTV evaluations (Skinns 1998 and Gill and Spriggs 2005), and an evaluation of alley-gating (Bowers et al. 2004a). What is evident from these studies is that in the majority of cases, such measures were not only effective at reducing crime but also cost-effective in that the financial benefit of the intervention outweighed the cost. Of course, this particular set of findings is not likely to be a truly representative sample, but the results can be seen as giving some cause for optimism. See Table 3 for more details of the cost-benefit analysis estimates for these examples.

How Long Does the Crime Reductive Effect Last?

Answering this question involves dealing with a number of related concepts, some of which were mentioned in the introduction. These include anticipatory benefit (the potential reduction in crime before the official start of the scheme), sustainability (the length of time over which a reductive effect is observed), and residual deterrence (a reduction in crime which temporarily continues when the intervention has come to an end). As is the case for cost-benefit analysis, there are few SCP evaluations that systematically test for these possibilities.

Smith et al. (2002) outline a number of potential causes of anticipatory reductions in levels of crime and search for empirical examples. From a conceptual point of view, they highlight a number of different mechanisms: training that staff have received before the official start of the scheme being immediately utilized, pre-scheme fact finding activity (such as surveys), changing pre-implementation surveillance levels, or publicity causing changes in offenders perception of risks associated with offending in certain areas. Empirical evidence found a possible anticipatory effect in approximately half of series of situational prevention schemes (22 of 52) where enough information was available for such detailed pre-scheme assessment. Looking for anticipatory benefit needs to be done with caution to ensure that it is not a dummy effect being detected. In order to rule out the possibility of an “anticipatory effect” in fact being regression to the mean, a seasonal fluctuation, or a change in recording practice, suitable controls need to be used in the analysis.

Sherman (1990) reported that reductions in crime as a consequence of a police crackdown were realized during the schemes intervention period but that these also endured for some time after the implementation had ceased. This is attributed to a temporary change in offenders' perception of risk involved in committing crime that extends beyond the lifetime of the operation. A related point is the issue of sustainability. Unlike those that follow individual life courses of offending behavior, situational prevention evaluations rarely extend beyond 1 or 2 years following the original implementation of an initiative. This means that there is little information concerning the length of the period for which such measures remain effective. Situational measures which use long-lasting hardware or change the physical layout of the environment are both examples where there is scope for crime reductive effects to last over long periods of time. Using an evaluation method that stops monitoring outcome after 1 or 2 years runs the risk of significantly underestimating the crime reductive effect of these measures. It also limits the availability of information concerning the point at which hardware should be updated or replaced. A piece of equipment might have a 10-year warranty – but is it cost-effective to replace it at this point?

The Bowers et al. (2004a) alley-gating evaluation illustrates this point. The evaluation assessed outcomes at two different stages. The analysis separated alley-gates that had been in place for less than a year from those that had been in place a year or more. This demonstrated that in fact the gates only became cost-effective after they had been in for some time; the benefit-cost ratio was £0.96 for those in for less than a year and £1.86 for those in a year or more.

Does the Intervention Shift the Crime Problem Elsewhere?

One of the most common criticisms leveled at situational interventions is that crime will simply relocate to other times and places since they do not address what are seen as the root causes of crime. There are a number of different forms of potential displacement (e.g., Hesseling 1994; Barr and Pease 1990), of which spatial displacement (movement of crime to nearby areas) is the form which tends to be most commonly analyzed in evaluation studies. Other commonly listed possibilities include temporal displacement (crime moves to another time), target displacement (crime is directed away from one target to another), crime type displacement (one kind of crime is substituted for another), tactical displacement (one method of committing crime is replaced by another), and offender replacement (one offender is arrested but replaced by another). There is of course another possibility, which is that the positive effect of the scheme is larger than originally anticipated; this is known as "diffusion of crime control benefits" (Clarke and Weisburd 1994). For example, diffusion takes place when reductions in crime occur in untreated areas or targets that are close to those that have received treatment or in types of crime similar to those targeted by the scheme.

A number of reviews of displacement and diffusion have been conducted. Three of these were undertaken in the 1990s (Barr and Pease 1990; Eck 1993; and Hesseling 1994). These early displacement reviews were fairly consistent in finding that displacement was often not observed, and in cases where it was, it tended to be smaller in scale than the reduction shown in the intervention areas. Of the 33 studies reviewed by Eck (1993), only three (9 %) appeared to suggest a substantial amount of displacement (where the negative outcome associated with displacement outweighed the positive gain as a result of the intervention). Similarly, reviewing 55 studies, Hesseling (1994) found that in 40 % of cases, no displacement at all was reported, and six studies appeared to demonstrate a diffusion of benefits. Finally, Barr and Pease (1990) took a more conceptual approach in their review and considered the equity of crime distribution, arguing that even in the minority event of total displacement, a redistribution of crime can still achieve a desirable social gain.

A more recent review of displacement and diffusion effects among situational crime prevention measures sought to update these earlier findings (Guerette and Bowers 2009). This review reported evidence from 102 evaluations of situational focused crime prevention projects and involved analysis of 574 separate observations relating to potential displacement or diffusion of benefit. Of these observations, the study authors reported displacement in 26 % and diffusion of benefit in a further 27 %. Interestingly these figures varied slightly depending on the type of displacement considered. For example, temporal and target displacement were slightly more commonly observed, as was a spatial diffusion of benefit. However, the reliability of these variations is restricted by the low number of evaluations in some of the displacement categories. For 13 of the 102 studies, the study design and the information available allowed a more detailed independent quantitative assessment. Here analyses suggested that when spatial displacement did occur, it tended to be less than the treatment effect, suggesting that on balance the intervention was still beneficial. Moreover, a diffusion of benefit appeared to be the more likely outcome.

A further systematic review of the degree to which SCP measures displaced crime or lead to diffusion of benefit was conducted by Johnson et al. (2011) using procedures closely in line with those recommended by the Campbell Collaboration (see http://www.campbellcollaboration.org/). This found that for the 13 studies in which sufficient information was available for the quantitative calculation of effect sizes, geographic displacement was by no means inevitable for this type of intervention. It was rarely the case that crime increased in the potential displacement areas following intervention. In fact, crime was just as likely to decrease in the areas that surround an intervention area, demonstrating diffusion of benefit. This review was limited in scope by low numbers of suitable studies, and those that were included tended to involve SCP measures that aimed to increase surveillance. That said it is still encouraging that displacement does not appear to be the inevitable outcome of SCP that it has sometimes been made out to be.

Possible Controversies: Levels and Types of Evidence

One debated issue in SCP evaluation is the type of evidence that should be used in the assessment of overall levels of effectiveness of particular interventions. There are two schools of thought: those that consider that only high-quality evaluation research should be included in such an assessment and those that, while conscious of validity issues, believe it is more important to have an assessment which understands and accounts for differences in local circumstances which will ultimately provide more evidence on the context in which the measures will work and the mechanisms by which the schemes operate.

Those that advocate firm quality control on the evaluation research design often refer to the Maryland Scale of Scientific Evidence (Sherman et al. 1997). This grades research designs according to their internal validity and deems randomized control trials the highest standard, quasi-experiments with multiple control groups as the next and evaluations which have no control or no temporal aspect the lowest standard. Under this paradigm restrictions are put in place in terms of the evidence that should be used in reviewing effectiveness. Commonly this means discounting evidence that is not level 3 or above on the 5-point Maryland Scale. Some SCP reviews that adhere to this guidance include the CCTV and street lighting reviews conducted by Welsh and Farrington (2008a, b), the review of Neighborhood Watch conducted by Bennett et al. (2008), and the POP review conducted by Weisburd et al. (2008). All of these reviews were undertaken using the procedures outlined by the Campbell Collaboration.

The approach which is more focused on understanding the process by which situational measures work and how this is moderated by the particular context of an intervention is generally referred to as the realistic evaluation approach (see Pawson 2006 and Pawson and Tilley 1997). A realist review starts with a model based on program theory (a statement of how an intervention works) and adds layers of evidence onto this from, for example, empirical evaluations. As the layers of evidence are added, it is likely that the initial model will need refining in able to sufficiently absorb it. Hence, a review might find that an intervention works differently in a newly revealed context. A significant feature of this approach is that the evidence is not just limited to evaluations. Other bits of evidence that might be included are theoretical pieces on process, agency reports on implementation, or initial bids for funding to name a few. The focus is on taking the most relevant bits of evidence from a range of different sources. The realist approach does not rule out using evidence from a study because of a poor research design but looks instead for what can be seen as "gold nuggets" even if they are in otherwise problematic documents. A further consideration often raised under the realist approach is the degree to which implementation failure occurs. That is when an intervention does not operate as original planned or parts of the intended activity fail to materialize (see Knutsson and Clarke (2006) for a detailed account).

A third option of course is to take some sort of hybrid approach. A review by Van der Knaap et al. (2008) on prevention of violence in public domains planned to do just that. This review classified each of 48 publications describing 36 violence reduction interventions into one of the five levels on the Scientific Methods Scale. Most were level 3 or above and many were level 4 or above. The authors choose to use only those that were level 4 or above in a vote counting review. The results found that of those where sufficient evidence was available, nine appeared to be effective and six potentially effective, with the remaining approaches looking less promising. As well as undertaking this systematic enquiry, the authors also searched the documents for details of program theory. They found three overarching mechanisms for schemes deemed as effective: ones that were cognitive and focused on learning, teaching, and training; ones that used the social environment to reward or punish behavior; and ones that focused on risk reduction to victims by promoting protective factors. This example of an integrated approach is not specifically related to SCP intervention but demonstrates that there are strengths to reviewing in this way as it does not neglect either of the philosophies outlined above. However this approach has the distinct drawback that it is likely to be very time consuming in nature.

Open Questions

There are some remaining gaps in our knowledge concerning evaluation of situational prevention measures. It is obvious from the discussion above that these include a more extensive evidence based on the cost-effectiveness of different approaches and further research on the sustainability of interventions in terms of estimating the lifetime over which they are effective. Also useful would be further research into the extent of displacement of types other than the more traditionally measured spatial variety. This undertaking presents a number of data collection and methodological challenges. For example, more detailed information recording precisely what times an intervention was active between or precisely which individuals or properties within a scheme area were treated would be necessary to enable a reliable assessment of temporal or target switch displacement. A future methodological challenge would involve examining the interactions between potential types of displacement. For example, it is conceivable that an intervention caused a crime switch that involved both a different target and occurred at a different time?

A further useful line of enquiry would be in more systematically assessing the degree to which implementation intensity affects the level of reduction seen. Intensity refers to the dosage of intervention received per treated subject. It is usually expressed in financial terms and is represented, for instance, as pounds or dollars spent per head of the treated population. It is unclear whether the relationship between dosage/intensity and crime reduction is a linear one (Bowers et al. 2004b). It might be the case that there is a level of spending per head at which point cost-effectiveness is optimized and after which any further investment gives diminishing returns. Evaluations that look at the relationship between intensity and crime reduction remain few and far between (see, e.g., Ekblom et al. 1996).

There could usefully be more syntheses of evidence on effectiveness conducted from the perspective of both schools of thought outlined above. There are plenty of SCP interventions for which (to the authors' knowledge) there has not been a review. For example, there might usefully be a synthesis examining the degree to which controlling access to facilities, reducing anonymity, or reducing provocations can reduce crime. There should also be more syntheses that use mixed methods such as those illustrated by Van der Knaap et al. (2008).

Finally, to feed into the body of evidence on what works from a situational perspective, more individual evaluations are necessary. A framework for optimizing these in terms of the production of new research evidence rather than replication of already established effects would perhaps be the most desirable way forward. For example, the Welsh and Farrington (2008a) review of CCTV described results from 44 evaluations in a number of different contexts and across a number of different countries. This provides fairly solid evidence of the extent to which CCTV shows promise as a crime reduction measure and facilitates discussion of the factors that are likely to moderate its effectiveness. However, from a brief scan of the published academic literature, there appear to have been far fewer evaluations of the effectiveness of replacing beer glasses with safer alternatives (Warburton and Shepherd 2000; Anderson et al. 2009) and more evidence on how generalizable the findings of these studies would be useful, both from a qualitative and a quantitative perspective. Having said that more novel evaluations would be useful, it is important to balance this with a sufficient number of replications to increase the external validity of the evidence base.

Related Entries

- ▶ Crime Prevention Through Environmental Design
- ▶ Crime Science
- ▶ History of Randomized Controlled Experiments in Criminal Justice
- ▶ Place-based Randomized Trials
- ▶ Randomized Block Designs
- ▶ Randomized Experiments in Criminology and Criminal Justice
- ▶ Situational Approaches to Terrorism
- ▶ Situational Crime Prevention

Recommended Reading and References

Anderson P, Chisholm D, Fuhr D. (2009) Effectiveness and cost-effectiveness of policies and programmes to reduce the harm caused by alcohol. Lancet, 373(9682):2234–2246.

Barr R, Pease K (1990) Crime placement, displacement and deflection. In: Tonry MH, Morris N (eds) Crime and justice: a review of research, vol 12. University of Chicago Press, Chicago

Bennett T, Holloway K, Farrington D (2008) The effectiveness of neighborhood watch. Campbell Syst Rev 2008:18. doi:10.4073/csr.2008.18

Bowers KJ, Johnson SD, Hirschfield AFG (2004a) Closing off opportunities for crime: an evaluation of alley-gating. Eur J Crim Pol Res 10(4):283–308

Bowers KJ, Johnson SD, Hirschfield AFG (2004b) The measurement of crime prevention intensity and its impact on levels of crime. Brit J Criminol 44(3):1–22

Braga AA, Bond BJ (2008) Policing crime and disorder hot spots: a randomized, controlled trial. Criminology 46:577–607

Clarke RV (1992) Introduction. In: Clarke RV (ed) Situational crime prevention: successful case studies. Harrow and Heston, Albany

Clarke RV, Weisburd D (1994) Diffusion of Crime Control Benefits: Observations on the Reverse of Displacement. In RV Clarke (ed.) Crime prevention studies, Vol 2. NY: Criminal Justice, Monsey, pp. 165–182

Clarke RV, Eck J (2003) Become a problem solving crime analyst in 55 small steps. Jill Dando Institute of Crime Science, London. http://www.popcenter.org/library/reading/pdfs/55stepsUK.pdf

Clarke RV, McGrath G (1990) Cash reduction and robbery prevention in Australian betting shops. Secur J 1:160–163

Dubourg R, Hamed J, Thorns J (2005) The economic and social costs of crime against individuals and households 2003/04 Home office on-line report 30/05

Eck JE (1993) The threat of crime displacement. Crim Justice Abstr 253:527–546

Eck JE, Guerette RT (2012) "Own the place, own the crime" prevention: how evidence about place-based crime shifts the burden of prevention. In: Loeber R, Welsh BC (eds) The future of criminology. Oxford University Press, Oxford

Ekblom P, Law H, Sutton M, Crisp P, Wiggins R (1996) Safer cities and domestic Burglary. Home office research study no. 164. Home Office, London

Forrester D, Frenz S, O'Connell M, Pease K (1990) The Kirkholt Burglary prevention project: phase 2, crime prevention unit paper no. 23. Home Office, London

Gill M, Spriggs A (2005) Assessing the impact of CCTV. Home office research study, no. 292. Home Office, London

Guerette RT (2009) The pull, push and expansion of situational crime prevention evaluation: an appraisal of thirty-seven years of research. In: Tilley N, Knutsson ZJ (eds) Evaluating crime reduction initiatives crime prevention studies, vol 24. Criminal Justice, Monsey, pp 29–58

Guerette R, Bowers K (2009) Assessing the extent of crime displacement and diffusion of benefit: a systematic review of situational crime prevention evaluations. Criminology 47(4):1331–1368

Hesseling R (1994) Displacement: a review of the empirical literature. In: Clarke RV (ed) Crime prevention studies, vol 3. Criminal Justice, Monsey

Johnson SD, Bowers KJ, Jordan P, Mallender J, Davidson N, Hirschfield AFG (2004) Estimating crime reduction outcomes: how many crimes were prevented? Evaluation 10(3):327–348

Johnson SD, Bowers KJ, Guerette R (2011) Crime displacement and diffusion of benefit: a review of situational crime prevention measures. In: Welsh B, Farrington D (eds) The oxford handbook of crime prevention. Oxford University Press, Oxford

Knutsson J, Clarke RV (eds) (2006) Putting theory to work: Implementing situational prevention and problem-oriented policing. Crime Prevention Studies Vol 20. NY: Criminal Justice, Monsey

Knutsson J, Tilley N (eds) (2009) Evaluating crime reduction initiatives, vol 24, Crime prevention studies. Criminal Justice, Monsey

Painter KA, Farrington D (2001) The financial benefits of improved street lighting, based on crime reduction. Light Res Technol 33:3–12

Pawson R (2006) Evidence-based policy: a realist perspective. Sage, London

Pawson R, Tilley N (1997) Realistic evaluation. Sage, London

Roman J, Farrell G (2002) Cost-benefit analysis for crime prevention: opportunity costs, routing savings and crime externalities. Crime Prev Stud 14:53–92

Sherman L (1990) Police crackdowns: initial and residual deterrence. In: Tonry M, Morris N (eds) Crime and justice: an annual review of research. University of Chicago Press, Chicago, pp 1–48

Sherman L, Gottfredson D, MacKenzie D, Eck J, Reuter P, Bushway S (1997) Preventing crime: what works, what doesn't, what's promising. National Institute of Justice, U.S. Department of Justice, Washington, DC

Skinns D (1998) Doncaster CCTV surveillance system: second annual report of the independent evaluation. Faculty of Business and Professional Studies, Doncaster College, Doncaster

Smith MJ, Clarke RV, Pease K (2002) Anticipatory benefit in crime prevention. In: Tilley N (ed) Crime prevention studies, vol 1. Criminal Justice, Monsey, pp 71–88

Van der Knaap L, Leeuw F, Bogaerts S, Nijssen L (2008) Combining Campbell standards and the realist evaluation approach: the best of two worlds? Am J Eval March 2008(29):48–57

Warburton AL, Shepherd J (2000) Effectiveness of toughened glassware in terms of reducing injury in bars: a randomised controlled trial. Inj Prev 6(1):36–40

Weisburd D, Telep CW, Hinkle JC, Eck JE (2008) The effects of problem-oriented policing on crime and disorder. Campbell Syst Rev 2008:14. doi:10.4073/csr.2008.14

Welsh BC, Farrington DP (2008a) Effects of closed circuit television surveillance on crime. Campbell Syst Rev 2008:17. doi:10.4073/csr.2008.17

Welsh BC, Farrington DP (2008b) Effects of improved street lighting on crime. Campbell Syst Rev 2008:13. doi:10.4073/csr.2008.13

Electronic Ball and Chain

► Electronic Monitoring

Electronic Bracelet

► Electronic Monitoring

Electronic Incarceration

► Electronic Monitoring

Electronic Monitoring

Rita Haverkamp
Max Planck Institute for Foreign and
International Criminal Law, Freibur, Germany

Synonyms

Electronic ball and chain; Electronic bracelet; Electronic incarceration; Electronic tagging; Electronic tracking of offenders; Prison without bars; Virtual prison

Overview

Electronic monitoring (hereafter EM) was first introduced in the United States in 1983 and has continued to be used for almost 30 years since. In the meantime, EM has spread across the globe and is either in habitual use or has been on trial in many countries. The reasons for the implementation of EM are diverse. That said, claims of prison overcrowding and the cost of building new prisons are overwhelmingly relied on to support its introduction. Intimately connected with these arguments are expectations that see in EM an affordable instrument that is cheaper than imprisonment. From a treatment perspective, EM is regarded as a more humane approach to deal with offenders. Its main objective is to have offenders refrain from committing offenses and, in the long run, to decrease recidivism.

EM can fulfill the objectives of detention, restriction, and/or surveillance concerning the monitored offender. Hence, EM programs are diverse and may be implemented at different stages of the criminal justice system: pretrial, at sentencing, in prison, and postprison. Correspondingly, its application depends on the sophisticated technology required to address both low-risk and high-risk offenders. While the first generation of EM technology based on radio frequency (hereafter RF) transmissions only locates the offender's presence or absence at a single location (usually the place of residence), the next generation of Global Positioning System (hereafter GPS) technology enables the continuous tracking of an offender's movements and broadens the target group of EM to high-risk offenders. In the USA, GPS equipment seems to have now replaced to certain extent conventional RF equipment. Although the efficacy of EM has been discussed in numerous studies, a lack of evidence-based research is still apparent (see Nellis et al. 2013). The absence of reliable knowledge concerning its impact on recidivism is also caused by the diverse offender groups monitored and the variety of applications combined with burgeoning technologies. The theoretical foundation of EM also displays deficiencies. Various approaches see the electronic bracelet as a symbol of the rising society of control; as a means within a threefold model of "punishment-punitive," "rehabilitative-humanistic," and "managerial-surveillant" discourses; or as part of a global "corrections-commercial" complex. Moreover, not enough empirically based studies exist to place EM within a single theory.

Fundamentals

The origins of EM date back to the 1960s. In 1964, Ralph Schwitzgebel (nowadays Gable) developed the first functional system at Harvard University in Cambridge, Massachusetts (Burrell and Gable 2008). EM was piloted on juvenile offenders within prescribed geographical areas that were equipped with repeater stations. Each individual wore a 1 kg "radio telemetry device" which activated a modified missile tracking unit up to 400 m away. This repeater station transmitted the positioning signal to a base station that in turn determined the individual's location on a screen. Schwitzgebel's twin brother, Robert, improved the RF system by enabling it to send tactile signals in a two-way coded communication. However, neither the original nor the refined prototype was able to establish a foothold at this time, due mostly to economic and practical reasons (too expensive, too heavy, and too advanced) (Burrell and Gable 2008).

It was not until 1983 that the idea of tracking offenders was realized by the district court judge Jack Love in Albuquerque, New Mexico. A newspaper cartoon about Spiderman, in which the villain covertly attached a radar device onto Spiderman to track his movements, inspired Love to convince a company to create a monitoring bracelet (Burrell and Gable 2008). In April 1983, Love sentenced the first of three offenders to electronic home detention as a condition of probation. Two of the three offenders successfully completed their curfew but recidivated afterwards. In December 1983, another individual on probation was sentenced to a weekend of home detention in Monroe Country, Florida. In November 1984, a pilot scheme was carried out in Palm Beach, Florida. This evaluation lasted 5 years and included 415 cases from late 1984 to the end of October 1989. The researchers observed that 97 % of the monitored persons succeeded in completing electronic monitoring and almost 80 % passed their probation period (Lily et al. 1992). During the 1980s, EM rapidly spread throughout the USA; the number of offenders monitored increased from 826 in 1987 to approximately 12,000 in 1990 (Schmidt 1998). Nowadays, EM is conducted in all 50 states of the USA, with an estimated 100,000–120,000 persons monitored in 2006. However, comprehensive data on the EM caseload has not been available since the last relatively complete study in 1990.

The US model of EM aroused interest in several countries, and its use for various applications began to spread around the world. The first pilot projects after the USA were initiated by Canada (1987), Australia (1989), and England and Wales (first trial 1989). In South America, Argentina implemented EM in the late 1990s. New Zealand piloted EM in Auckland between 1995 and 1997 and adopted the measure nationwide in 1999. In Asia, Singapore (early 1990s), Israel (2005), and Taiwan (2006) now run EM programs with completely different applications and target groups. South Africa was the first African country to set up a trial in 1996. In many Western European countries, EM was introduced at the beginning of the twenty-first century (Pinto and Nellis 2011). Whereas the instrument was already established in England and Wales, Sweden, and the Netherlands, experiments have occurred at a national or regional level in Belgium (1998), France (2000), Germany (Hesse 2000), Italy (first pilot 2001), Portugal (2002), Switzerland (pilots in six cantons since 1999), and Spain (Catalonia 2000). Meanwhile, other Western and Eastern European countries have since introduced EM programs: Austria (2010, permanent), Denmark (2005), Estonia (2007), Czech Republic (2009–2010), Finland (2011, permanent), Germany (pilot study in Baden-Württemberg 2010, permanent federal state 2011), Ireland (2009), Luxembourg (2006), Norway (2008), Poland (2009), and Scotland (2006) (Table 1).

Electronic Monitoring, Table 1 Total number of participants electronically monitored in comparison with the population size

Jurisdiction	Population size (in thousands)	EM participants 2010
USA	308,746	100,000–120,000
Austria	8,350	97
Belgium	10,825	2,927 (2009)
Denmark	5,587	1,898
England and Wales	55,328	99,950
Estonia	1,340	140
France	57,300	12,124
Germany (Hesse)	6,067	206
Ireland	4,453	31
Luxembourg	507	45
Netherlands	16,600	843
Norway	4,853	1,001
Poland	38,500	616
Portugal	10,612	708
Scotland	5,222	2,626
Spain (Catalonia)	7,200	280
Sweden	9,045	3,739
Bern + Vaud (Switzerland)	1,612	190
Total	552,147	244,494

Origin: Pinto and Nellis 2011; own inquiry (Hesse)

The main reason for the rise of EM in the last decades has been prison overcrowding and the associated desire to reduce prison populations and avoid the construction of new prisons. The emergence of EM also occurred alongside the development of community-based sanctions to enhance their efficacy and increase public safety. Another important stimulus has been a push to reduce correctional system costs due to public finance crises and related budget restrictions. Despite its development as a cost-effective means, the introduction of EM is also related to the search for a humane alternative to incarceration, thereby supporting rehabilitation and reintegration of offenders into society. Though the motivations behind EM remain diverse, several additional objectives have been pursued since its appearance. Nonetheless, the variety of EM applications is based on three central rationales:

detention, restriction, and surveillance. Detention is achieved by house arrest schemes. The monitored offender has to stay at home during established curfew hours. Home detention programs were one of the first types of EM and continue to be the most prevalent. Restriction is achieved by preventing the offender from entering proscribed areas or from approaching prohibited persons (e.g., [potential] victims, co-offenders). Surveillance is achieved by continuously tracking the offender, without actually restricting their movements. The emphasis of EM lies within the framework of the enforcement of community-based sentences or other preventative measures. The function of EM is restricted to a technological means which ensures that the monitored offender complies with the imposed conditions. Therefore, EM can be widely implemented at all stages of the criminal justice and correctional system.

Key Issues

During the last decades, diverse EM technologies have been created; the new developments generally surpass their predecessors in functionality (more reliable technology) and portability (smaller and lighter devices). The equipment usually consists of a tamperproof and water-resistant device which is fastened around the ankle or sometimes the wrist of the wearer. In addition, specific technological systems require different additional items. Tracking applications for offenders can be divided into active and passive systems operated with RF or GPS technology. Whereas the active system is characterized by continuously signaling, the passive system is based on programmed contact. The nowadays rarely used passive RF systems contact the person monitored by a landline telephone to ensure their presence at a designated place (usually residence). Different means may verify the individual's identity like the inserting of a bracelet into a verification box, the submission of a password, or voice or biometric verification (fingerprint, retinal scan). The technology is limited to mere verification at random or specified times; it is not able to detect the individual's location or possible violations of conditions in the period between calls (Renzema and Mayo-Wilson 2005). In contrast, active RF systems have been the most commonly used technology since the introduction of EM. In essence, a transmitter worn by the individual continuously emits a signal to a receiver unit in the wearer's place of residence. The receiver unit is connected to the offender's conventional telephone which relays the signal to a computer at the monitoring station. The computer alerts the authorities to signal interruptions in the individual's schedule or to the possible tampering of the EM bracelet (Black and Smith 2003). Modern variations not only monitor the presence or absence of an offender at a single location (mostly place of residence) but also record an individual's position at a specified place via drive-bys (e.g., if the monitored person is at a specific location at a scheduled time, such as place of employment). Remote alcohol monitoring can be added to the EM programs of drunk drivers to measure alcohol consumption with a portable or stationary single use or continuous testing device. Passive RF systems are solely restricted to detention purposes, and active RF systems are also primarily used for this role.

In contrast, GPS EM enables authorities to pursue restriction and surveillance objectives. The application requires at least three different satellites, a network of ground stations, and a device worn by the offender (bigger than the RF bracelet, and the wearer must charge the battery). Active satellite technology enables the real-time monitoring of the wearer 24 h a day. In contrast, the more cost-effective passive GPS version of EM usually downloads location and movement data once a day and only allows for a retrospective verification. Therefore, active systems are favored to enhance the supervision of curfew hours, place restrictions, and contact to "forbidden" persons or (potential) victims. In comparison to RF technology, active satellite tracking signifies a major advancement in offender monitoring practices because continuous surveillance enables the enhancement of community-based measures and the supervision of further offender groups. Since the introduction of GPS monitoring in Florida, the number of RF placements has declined considerably (Bales et al. 2010). This suggests that GPS tracking might replace RF technology completely, highlighting the use of restriction and surveillance for all types of offenders.

As a technological innovation, EM is a versatile and feasible instrument which can be implemented across the entire criminal justice system (Black and Smith 2003). The variety of EM programs means that it can be used at the pretrial level, at the sentencing stage, within the prison regime, and during the postprison phase. In pretrial criminal proceedings, EM can be used either as an additional condition for release on bail or in lieu of traditional bail payments for poor individuals who would otherwise remain in custody (see the entry "▶ Alternatives to Pre-trial Detention"). The intention is to reduce the flight risk as well as the risk of committing new offenses. Further, EM is supposed to ensure that the defendant follows the imposed conditions. In a so-called front-door scheme (the door to prison), incarceration should be avoided, and the criminogenic environment of custodial settings should be prevented. In this way, EM serves as an intermediate sanction that is less harsh than prison but more severe than probation (Black and Smith 2003). These EM programs are usually of the home detention variety, with specified restrictions on the liberty of the offender. At the sentencing stage, the measure can be imposed as a primary sentencing option, as a court order to a suspended prison sentence, or as a condition to an imposed community sanction. EM may even be a viable alternative to imprisonment; after a court has sentenced an individual to imprisonment, he or she may apply, and be assessed as suitable, for an EM house arrest program. An "indoor scheme" has been introduced to the prison regime in some countries. Here, EM is used in open prisons (for inmate tracking) and outside during prison leave and work release. The "back-door scheme" (the door out of prison) represents the most common use of EM and can be divided into early release and post-release programs. For instance, EM can be used as a condition for early release (e.g., parole) to help ease the transition from prison back into the community (see the entry on "▶ Early Release from Prison"). Post-release programs are used especially for high-risk offenders, in order to monitor their movements and to deter them from committing severe crimes. Beyond the scope of the traditional justice system, EM can be found in a variety of other diverse areas. One well-known example is in court restraining orders to prevent a potential offender from approaching a potential victim. In cases of domestic violence, EM can function as a tool of victim protection: if a prohibited zone is entered, the authorities can contact the person monitored and the potential victim can be informed by a technical device at their home. A rarer application is the monitoring of foreign nationals who are awaiting deportation (Pinto and Nellis 2011). In the criminal justice system, EM is well established as a stand-alone measure or as a condition in numerous multimodal programs.

Various types of offenders (e.g., drug users, juveniles) are regarded as suitable target groups for EM, though the legal and technological application tends to depend on the risk level of the convicted (Renzema and Mayo-Wilson 2005). Pilot projects are often characterized by a cautious tendency concerning the selection of offenders; consequently, low-risk groups are mostly preferred for initial trials of EM (e.g., drunk drivers, unlicensed or suspended drivers, petty criminals, first-time offenders) (Gable and Gable 2005). Of course, it should be kept in mind that the technological limits of the first RF monitoring equipment also gave reason for its restrained implementation. Moreover, the original EM participants showed mostly favorable socioeconomic attributes such as a fixed residence, gainful employment, close family bonds, and relatively stable financial affairs (Gable and Gable 2005). At the end of the 1990s, several programs in the USA began to address moderate offenders (e.g., where the risk of recidivism existed) (Schmidt 1998). Since the advent of GPS surveillance, high-risk groups (e.g., sexual offenders, persistent and prolific offenders) have also been included. Whereas detention is pursued for low-risk populations, restriction and surveillance dominate the tracking of high-risk populations (Black and Smith 2003). Hence, the intervention level of a program varies according to the offender's risk assessment. While EM for low-risk offenders is frequently a sole measure or is combined with other forms of low-contact monitoring, EM for moderate and high-risk offenders is regularly embedded in a multimodal program to enhance human contact, supervision, drug treatment, or other services (Renzema and Mayo-Wilson 2005). Thus, the degree of interference is dependent on the offender. The higher the risk assessment, the more invasive EM will be, including the length of daily detention and the imposition of monitoring (continuous tracking or the implementation of restricted/prohibited areas). Indeed, several states in the USA have now enacted lifetime GPS monitoring for sexual offenders (e.g., Jessica Lunsford Act Florida 2005).

In home detention schemes, the monitored person remains at home according to an individually arranged and computer-programmed time schedule (Mayer and Haverkamp 2003). The place of residence may be left for the time spent at work or in treatment programs and even for leisure time. Provisions for RF monitoring are typically a suitable fixed residence, a functioning (landline) telephone, and an occupation such as work, study, or treatment. The monitored person and his or her fellow occupants usually consent to the measure. In several countries, the monitored person must contribute to the costs involved based on their individual income. Due to differences in legal frameworks, intended objectives, and methods of application, enforcement of EM is heterogeneous and may be performed by probation or correctional services and/or private security companies. Programs involving probation or correctional services are mainly of a reintegrating and rehabilitative nature and often refer to early release schemes. Assistance is in the form of practical aid with daily routines, through to treatment in order to support long-term behavioral changes. EM programs can also be divided into stricter or more lenient models. Whereas the stricter model includes unannounced home visits, alcohol/drug controls, compulsory treatment, and/or behavioral programs, the more lenient model allows the moderate consumption of alcohol, free time on weekends, and/or time for socially acceptable activities.

Controversies

In the twenty-first century, controversies on the normative and ethical issues that surround EM have played a minor role concerning the use of RF technology, especially when compared to the 1980s and 1990s (Albrecht 2005). The principal fear then was the infringement of the constitutional rights of the monitored person, especially the right to equality under the law and privacy (Lily and Ball 1987; but see also recently Nellis et al. 2013). Technical limits (monitoring restricted to a single location), the design of the programs (informed consent of the offender and fellow occupants), and efforts to avoid discrimination (provision of a place of residence, waiver of costs in the case of low-income earners) widely refuted these objections. It should also be remembered that EM often provides an alternative to prison and its punitive nature is distinctly less than that of incarceration. In different studies, the vast majority of interviewed offenders have confirmed this assumption (Martinovic 2010).

Hence, the debate on EM has shifted to an economic perspective, especially concerning the question of its cost-effectiveness when compared to imprisonment and community-based measures. However, the calculation of cost savings is a complicated matter, and numerous factors have to be taken into account. In comparison to prison, the monitored individual maintains his or her employment and fixed residence, pays taxes, and contributes to the costs for EM; yet such savings are not easy to determine (Black and Smith 2003). The simplest way to measure the economic efficiency of EM is to compare the daily costs for a monitored offender to the daily costs for an imprisoned offender which, according to many evaluative studies, generally supports the application of EM. Nevertheless, a comparison between different countries is difficult, as the daily costs are composed of different factors: e.g., some countries only calculate the costs of equipment and installation, other countries include the cost of social work, and yet other countries refer to savings in prison use (Pinto and Nellis 2011). In addition, this method of calculation has been criticized for its limited approach and because it disregards many of the real costs of EM. Consequently, unintended side effects of EM must also be taken into account, such as the net widening effect, which may actually increase financial expenditure. According to the concept of net widening, the use of EM may lead to the over-penalization of low-risk offenders who would otherwise be sentenced to a community-based order (offender net widening or front-end net widening) (Gable and Gable 2005; Nellis et al. 2013). Furthermore, technical violations of the monitored person can lead to a revocation of EM and to imprisonment (back-end net widening). In turn, this may lead to an increase in financial, staff, and technical resources for the correctional system (systemic net widening). That said, it is difficult to provide evidence concerning the emergence of net widening effects due to shortcomings in evaluative research and controversies among the current findings (Albrecht 2005).

As opposed to the use of RF technology, GPS surveillance arouses various concerns with regard to human rights and/or the replacement of social supervision. Human rights might be impinged by the intrusiveness of GPS EM due to the length of surveillance and the ability to monitor all aspects of a person's life. The observed growth of GPS monitoring for low- and moderate-risk offenders in Florida constitutes an added intervention which could entail net widening compared to the use of less intrusive RF technology. The blanket use of GPS technology for all types of offenders would violate the principle of proportionality, namely, the punishment must not exceed the seriousness of the offense and sentencing procedures must consider the extent of the offender's culpability.

The social impact of EM is another issue, especially with regard to its rehabilitative and reintegrative potential as a stand-alone measure or as an element in a multifaceted program. Results of evaluative studies show that permanent supervision can have a deterrent effect during the monitoring phase. The physical bracelet reminds the offender of his or her situation. This continuous confrontation may enhance compliance in a trust-, incentive-, and threat-based way (trust, no manipulation of the device; incentive, stay in familiar environment; threat, desistance) (Nellis 2006). Political hopes concerning long-lasting changes in offender attitudes have, however, not been fulfilled. Academic doubts often refer to the ability, or inability, of EM as a stand-alone measure to affect behavioral change in the long term. This is due to its diverse application to a wide range of offenders (Renzema and Mayo-Wilson 2005). However, rehabilitation-oriented experts favor the potential of EM when it is integrated in a multimodal program. According to evaluative findings, this type of EM might help offenders to successfully complete other rehabilitation programs (Bales et al. 2010). The continuous physical confrontation seems to offer a chance to affect the intrinsic motivation of the monitored person, whereby the person's biography as well as their past and (potential) future criminal behavior is reflected on (Huckelsby 2008). Therefore, a monitored offender may also benefit from participation in treatment interventions to support pro-social behavior (Gable and Gable 2005).

The question of rehabilitation is strongly connected with evaluative research on EM. Although EM is now a well-established instrument, the lack of evidence-based research is regrettable (Renzema and Mayo-Wilson 2005). The reason for this lack of evidence has to do with the methodological limitations of the vast majority of outcome evaluations that generate contradictory results on the effectiveness of EM. Attempts to measure the effect of EM on recidivism suffer from an absence of comparative control groups and the use of non-interpretable studies with randomly assigned or historical comparison groups. That said, during the last decade, some methodologically sound evaluations of EM's effectiveness have been carried out. Though they yielded dissimilar findings, this was likely due to differences in EM's application, the target group, and their cultural background. Whereas a Swedish early release study found that midrange offenders seem to benefit most from EM (Marklund and Holmberg 2009), no significant impact on recidivism could be observed in an English study on curfew orders (Renzema and Mayo-Wilson 2005).

Deficiencies also surround the theoretical foundation of EM. In the 1990s, Deleuze proclaimed that EM heralded a process of transition from disciplinary societies to societies of control (Deleuze 1992). This transformation advances the use of technology (e.g., EM), as control architectures unfold and alter penal practice. Contemporary penology identifies three interconnected discourses here: "punitive-repressive," "managerial-surveillant," and "humanistic-rehabilitative" (Nellis 2011). The "punitive-repressive" discourse emphasizes the populist call for more severe punishment and correspondingly the imposition of more prison sentences. The "managerial-surveillant" discourse highlights modern improvements in technology and the use of actuarial techniques that favor efficiency and constructive management (e.g., commodification) in the most cost-effective manner (Cotter and de Lint 2009). The "humanist-rehabilitative" discourse panders to the offender's responsibility. EM represents all three discourses, whereby the punitive-repressive discourse refers to the offender's perception of EM compared to incarceration. While the majority of interviewed offenders answered in various studies that RF technology is more lenient than prison, the punitiveness of GPS still needs to be measured due to its intrusiveness, duration, and surveillance intensity. Satellite monitoring also addresses the managerial-surveillant discourse to ensure the management of a deviant population. One crucial element here is the discussion on the economics of imprisonment, which sees EM as being representative of the "corrections-commercial" complex (Lily and Deflem 1996). However, the majority of academic literature in this area has little or no recourse to theoretical foundations or empirical research (Cotter and de Lint 2009).

Future Directions

In contrast to RF EM, the advent of GPS EM has been characterized by anxiety and apprehension on three separate grounds. The first ground has to do with disturbing technological improvements that can be made to GPS EM. On the one hand, the progressive miniaturization of the devices means that monitored persons are now better able to conceal their confinement. However, on the other hand, the development of implantable microchips in the near future might advance surveillance not only to the individual's location but also to physiological signs (offender's heart rate and blood pressure) which could serve as indicators for potential recidivism (Cotter and de Lint 2009). Other advanced devices could even incorporate a miniature video camera (Black and Smith 2003). Furthermore, the use of automated disciplinary action in the case of violations has also been contemplated; this would likely mean the use of immobilizing electric current pulses (Whitfield 1997). Convicted pedophiles and other sexual offenders have been suggested as possible target groups for these severe and invasive technologies. Implantation and removal procedures, the use of additive biochemical features, as well as the use of electric current pulses would definitely raise serious ethical concerns (e.g., human rights, data protection).

Closely connected with these ethical concerns is the second ground for anxiety and apprehension. As GPS EM proliferates, it is increasingly likely that it will be applied for the duration of a person's life, particularly in the case of so-called dangerous offenders, such as those who have committed sexual and/or serious violent offenses. The increasing application of GPS EM implies an over-reliance on technology, which overemphasizes surveillance and enforcement at the expense of rehabilitation (Payne and DeMichele 2011). However, risk assessments place prominence on background or static factors, while dynamic or changing offender characteristics take a backseat role. Moreover, the treatment perspective is neglected by a "standard" application of EM, especially one that does not differentiate between the heterogeneous types of sexual offenders. However, during the past decades, EM has consistently been applied homogenously to a diverse group of offenders. Paradoxically, whereas EM technology is embedded in an ongoing process of sophisticated development, a false sense of security is widespread due to the limited knowledge of its operating mode (Payne and DeMichele 2011). Despite misunderstandings about the functionality of GPS EM, misconceptions also exist about the typical manner of offending in sexual offense cases. For example, most children are sexually abused within their family or acquaintance context; in contrast, GPS monitoring addresses the minority of sexual offenders who commit anonymous sexual assaults.

The third ground for concern has to do with the crucial question of rehabilitation by means of EM. The rehabilitative potential of EM still remains unclear, and empirical knowledge in this field is in its infancy. Of those empirical findings that do exist, they hint at the positive structuring function of EM for the offender's life, especially with regard to the fulfillment of treatment obligations. However, in practical terms, a tendency can be observed towards the use of EM as a stand-alone sanction with the focus on punishment rather than treatment (Nellis 2011; Nellis et al. 2013). In theory, the use of EM solely as a form of punishment is to be strongly opposed, due especially to the negative aspects of surveillance and repression. From this perspective, EM should best be understood as one tool in a multimodal program of supervision, a program that should ideally be individually tailored for each and every offender.

Acknowledgments Special thanks go to Chris Murphy, who edited the final manuscript and contributed his translation services/skills.

Related Entries

- ▶ Alternatives to Pre-trial Detention
- ▶ Community Service in Europe
- ▶ Early Release from Prison

Recommended Reading and References

Albrecht HJ (2005) Electronic monitoring in Europe. A summary and assessment of recent developments in the legal framework and implementation of electronic monitoring. Freiburg im Breisgau, Germany, http://www.mpicc.de/shared/data/pdf/albrecht.pdf

Bales W, Mann K, Blomberg T, Gaes G, Barrick K, Dhungana K, McManus B (2010) A quantitative and qualitative assessment of electronic monitoring. Report submitted to the Office of Justice Program National Institute of Justice U.S. Department of justice. The Florida State University College of Criminology and Criminal Justice Center for Criminology and Public Policy Research

Black M, Smith RG (2003) Electronic monitoring in the criminal justice system. Australian Institute of Criminology, Trends & Issues in Crime and Criminal Justice No. 254, Canberra

Bottos S (2007) An overview of electronic monitoring in corrections: the issues and implications. Research Branch, Correctional Service of Canada, Ottawa

Burrell WD, Gable RS (2008) From B. F. Skinner to Spiderman to Martha Stewart: The past, present and future of electronic monitoring of offender. Probation Parole 46(Current Issues):101–118

Cotter R, De Lint R (2009) GPS-electronic monitoring and contemporary penology: a case study of US GPS-electronic monitoring programmes. Howard J 48(1):76–87

Deleuze G (1992) Postscript on the societies of control. MIT Press, Cambridge, MA, pp 3–7

Gable RK, Gable RS (2005) Positive intervention strategies. Fed Probat (69)1:21–25

Huckelsby A (2008) Vehicles of desistance? The impact of electronically monitored curfew orders. Criminol Crim Justice 8:51–71

Lily RJ, Ball RA (1987) A brief history of house arrest. North Ky Law Rev 20:505–530

Lily RJ, Deflem M (1996) Profit and penalty: an analysis of the corrections-commercial complex. Crime Delinq 42(1):3–20

Lily RJ, Ball RA, Curry GD, Smith RC (1992) The pride Inc. Program: an evaluation of 5 years of electronic monitoring. Fed Probat 56:42–47

Marklund F, Holmberg S (2009) Effects of early release from prison using electronic tagging in Sweden. J Exp Criminol 5:41–61

Martinovic M (2010) The complexity of punitiveness of electronically monitored sanctions. The western world's analysis. Lambert Academic Publishing, Berlin

Mayer M, Haverkamp R, Lévy R (2003) Will electronic monitoring have a future in Europe. Freiburg i. Br, Germany, Max Planck Institute for Foreign and International Criminal Law, Edition Iuscrim

Nellis M (2006) Surveillance, rehabilitation, and electronic monitoring: getting the issues clear. Criminol Public Policy 5(1):103–108

Nellis M (2011) The integration of probation and electronic monitoring – a continuing challenge. A reflective report for CEP. http://www.cepprobation.org/uploaded_files/EM%20Literature%20Research.pdf
Nellis M, Beyens K, Kaminski D (2013) Electronically monitored punishment. International and critical perspectives. Routledge, London/New York
Paterson C (2007–2008) Commercial crime control and the electronic monitoring of offenders in England and Wales. Social Just 34 (3–4):98–110
Payne BK, DeMichele M (2011) Sex offender policies: considering unanticipated consequences of GPS sex offender monitoring. Aggress Violent Behav 16:177–187
Pinto S, Nellis M (2011) 7th European electronic monitoring conference survey of electronic monitoring: analysis of questionnaires. http://www.cepprobation.org/uploaded_files/EM2011_Conference_Analysis-of-EM-Questionnaires.pdf
Renzema M, Mayo-Wilson E (2005) Can electronic monitoring reduce crime for moderate to high-risk offenders? J Exp Criminol 1(2):215–237
Schmidt AK (1998) Electronic monitoring: what does the literature tell us. Fed Probat 62(2):10–19
Whitfield RG (1997) Tackling the tag. The electronic monitoring of offenders. Waterside, Winchester

Electronic Tagging

▶ Electronic Monitoring

Electronic Tracking of Offenders

▶ Electronic Monitoring

Elite Deviance

▶ Women and White-Collar Crime

Emergency Response

▶ Preventing of Terror at Shopping Malls

Empathy and Offending

Darrick Jolliffe
School of Law, University of Greenwich,
London, UK

Overview

Empathy is an individual difference that has been theoretically linked with an increased likelihood of offending, particularly serious offending. In contrast to the strong theoretical relationship, however, the empirical evidence of the relationship between empathy and offending is lacking. This is problematic as many offender treatment programs are designed to address this supposed "lack of empathy." In order to further knowledge about empathy and offending, the measurement of empathy should be improved and studies should attempt to use more sensitive measures of offending (e.g., self-reports). Prospective longitudinal studies are needed to establish whether a lack of empathy is a correlate, a risk factor, or a causal risk factor for offending. Only if empathy is the latter would empathy enhancement be expected to reduce the likelihood of future offending.

Introduction

Empathy is an elusive construct, and as a result, clarifying its association with offending has challenged researchers for some time. Despite the definitional difficulties, empathy (or more accurately a lack of empathy) is viewed by criminologists as having relevance for understanding the causes of crime. For example, empathy is mentioned in influential theories such as Gottfredson and Hirschi's *General Theory of Crime* (1990, pp. 89–90), "a lack of empathy" is a key, and possibly the, defining characteristic of the concept of psychopathy (e.g., Blair 2007) and there is conceptual overlap between empathy and a host of factors that have been associated with explaining crime such as poor social cognition, lack of guilt, lack of emotional intelligence, poor perspective taking, and weak moral reasoning. Perhaps most importantly however, empathy enhancement (through role-play and various perspective taking exercises) is viewed as standard practice in many offender rehabilitation programs (Marshall et al. 2009).

There is also direct evidence of the perceived importance that criminologists place on empathy in understanding crime. In their 1999 postal survey of 147 criminologists from the register of the American Society of Criminology, Ellis and Walsh found that "a lack of empathy and concern for others" was rated as one of the most important factors associated with serious and persistent criminal offending (behind only "an economic system that frustrates efforts by some to participate"). Nine years later a replication with 1,218 criminologists (similarly recruited) found that "a lack of empathy and concern for others" was the most highly rated for explaining the persistence of criminal behavior (Ellis et al. 2008).

What is Empathy?

Although the concept of "sympathy" has a long history, the term "empathy" is relatively new, being in widespread use for less than 50 years (Hanson 2003). Originating from the German term "einfühlung," meaning "to feel into," the term was first used in the early part of the twentieth century to describe the tendency to feel emotions from works of art (Allport 1985). The concept of empathy soon shifted to the social realms, acquiring new meanings as it was used to explain the effective ingredients of psychotherapy, the early socialization process, and the accuracy of interpersonal perceptions.

Currently, there remain many definitions of the term "empathy" available in the literature, and these have tended to either emphasize the proposed emotional component of empathy (i.e., being able to experience the emotions of another person; e.g., Hoffman 2000) or the cognitive component (i.e., being able to understand the emotions of another person; Hogan 1969). However, a broad consensus seems to be forming around empathy containing both affective and cognitive components (e.g., Davis 1983; Jolliffe and Farrington 2006). This separation also appears to have support from neuroscientific explorations of empathy (with affective and cognitive components having different neuronal pathways and associated substrates; Shirtcliff et al. 2009).

The theoretical relationship between low empathy and offending is seductively simplistic. Individuals who have lower levels of empathy are more likely to offend. This is because low-empathy individuals are free to act without the constraints imposed by the vicarious experience or understanding of the emotional consequences of their actions on others (e.g., Feshbach 1975). The fear, distress, sadness, and other negative emotions experienced by others as a result of the individual's antisocial actions are not factored in as a cost or consequence of the transgression, and therefore not inhibitory for present (or future) antisocial behavior. Having the ability to share or understand another's emotions is like having an emergency handbrake to reduce the likelihood of antisocial behavior, and individuals with low empathy are proposed to be missing this.

A number of more detailed models of the relationship between empathy and antisocial behavior exist (e.g., Feshbach 1987; Hanson 2003; Hoffman 2000), each postulating various subcomponents (e.g., ability to recognize emotions in others), or ordering of constituent parts of empathy (e.g., understanding another's emotions leads to the affective sharing of emotions). In general, however, these models all have the same overall formula: Low empathy increases the likelihood of committing offenses. The theoretical relationship between low empathy and offending appears stronger for interpersonal offenses such as violence and sex offenses than for offenses where the victim is more abstract or less obvious (e.g., shoplifting, drug possession). The separation of affective and cognitive empathy is theoretically useful for explaining offending that might show evidence of sufficient cognitive empathy (i.e., to understand the emotions of another, in, for example, grooming for sexual offenses), but a clear lack of affective empathy (i.e., no sharing of victim's fear or distress).

Below are examples of some of the larger studies conducted looking at empathy and offending. Fisher et al. (1999) administered a measure of cognitive and affective empathy to a group of 140 child molesters (59 participating in community-based sex offender treatment, 81 incarcerated) and 81 prison officers in training. The two groups were similar in IQ but differed significantly in age (with the prison officers younger). Overall, the child molesters were found to have similar levels of cognitive empathy to the prison officers, but significantly greater affective empathy. Similarly, Goldstein and Higgins-D'Allessandro (2001) used questionnaires to measure the empathy of 184 offenders incarcerated in New York jails (117 male, 67 female) and 126 comparison individuals. The results suggested that offenders did not differ significantly from non-offenders on either cognitive or affective empathy.

Despite the strong theoretical relationship between low empathy and offending, overall, the empirical evidence for the relationship is less impressive. This is true among studies comparing general offenders to non-offenders (e.g., Goldstein and Higgins-D'Allessandro 2001) as well as those comparing sex offenders to non-offenders (e.g., Hayashino et al. 1995). The latter findings were contrary to expectation, especially as sex offenders have long been assumed to possess significant deficits in empathy which directly contribute to their chosen method of offending (Marshall et al. 2009).

The most comprehensive review of the relationship between empathy and offending was conducted by Jolliffe and Farrington (2004). This systematic review included 35 studies which compared offenders (defined as those who committed acts, that if detected could result in a criminal conviction) to comparison individuals on questionnaire measures of empathy. The results suggested that offenders had significantly lower cognitive and affective empathy than non-offenders, but the relationship was stronger for cognitive empathy. More importantly, the relationship between low empathy and offending was greatly reduced after controlling for intelligence and socioeconomic status, suggesting that these might be more relevant factors or that empathy might mediate the relationship between these factors and offending (Jolliffe and Farrington 2004).

In addition to these substantive findings, this study also highlighted a number of areas where the evidence regarding the relationship between low empathy and offending was clearly lacking. For example, the questionnaires that existed to measure empathy were not ideal, and studies of empathy and offending tended to use only official and usually incarcerated offenders to represent offenders. In addition, the studies tended to make minimal attempts to test for the causal influence of empathy on criminal behavior. These issues will be explored in turn.

Measurement of Empathy

While the systematic review of Jolliffe and Farrington (2004) suggested low cognitive empathy was strongly related to offending, there was evidence that the questionnaires which measured empathy might not have been adequately measuring cognitive empathy. That is, most of the studies which measured cognitive empathy did so using a scale that lacked face validity, making it unclear whether empathy was the actual covariate being compared between offenders and non-offenders. The other questionnaire measures of empathy that were predominantly used were also potentially problematic as they contained items that appeared to be assessing levels of sympathy as opposed to empathy. Sympathy and empathy are interlinked but separate constructs (Feshbach 1975), with empathy defined as the understanding or sharing of another's emotion (Davis 1983) and sympathy usually operationalized as the additional appraisal about how one feels about the emotions of another (Eisenberg and Strayer 1987). For example, the item from one of the more commonly used measures of empathy (the Questionnaire Measurement of Emotional Empathy) "I become more irritated than sympathetic when I see someone in tears" (Mehrabian and Epstein 1972) clearly assesses concerned feelings for others, rather than shared emotion. This overlap between empathy and sympathy in the questionnaire measurement of empathy makes it unclear to what degree empathy was actually being assessed, let alone the degree and magnitude of the relationship between empathy and offending. Additionally, there is evidence to suggest that items which assess sympathy are more open to social desirability bias than those which assess empathy (Jolliffe and Farrington 2006).

Prisoners as Offenders

All studies of empathy and offending identified for the systematic review compared the empathy of those incarcerated (to represent "offenders") to a comparison group of those variously recruited and matched. In one respect, prisoners are more likely to be serious offenders (e.g., Farrington 2003), so empathy deficits should be more apparent when this group is compared to those in the community. However, using prisoners as "offenders" actually provides very little information about the potential link between empathy and the likelihood of committing offenses as prisoners are a unique (and biased) sample of all those who commit offenses. Further complicating is the problem that empathy levels might be reduced by periods of incarceration (e.g., Cale 2006). This means that to the extent that prisoners have low empathy, this could be equally because low empathy is a characteristic of those who commit serious offenses or that being incarcerated makes people less empathetic.

Correlation or Cause: The Evidence for a Relationship Between Empathy and Offending

When considering the relationship between empathy and offending, it is important to determine the extent to which we can have confidence that empathy is a correlate (i.e., associated with offending), a risk factor (i.e., associated with a proceeds offending), or a causal risk factor for offending (i.e., associated with, and proceeds offending and changes in empathy result in changes in likelihood of offending (Kraemer et al. 2005; Murray et al. 2009). This is an important distinction because only if low empathy is a causal risk factor for offending would empathy enhancement be expected to reduce subsequent offending.

One method of assessing the degree to which we can have confidence that empathy is a correlate, a risk factor, or causal risk factor for offending is to score studies included in a systematic review using a series of checklists designed to assess the methodological quality of studies included (e.g., Murray et al. 2009; Jolliffe et al. 2012). For example, the Cambridge Quality Checklists were developed to draw attention to key features of the included studies so that an assessment could be made regarding the confidence that one can have that the factor of interest (in this instance empathy) is associated with an outcome (e.g., offending).

The correlate score ranges from 0 to 5 depending on the method of sampling, the response rate, the sample size, and the method and quality of measuring the correlate and the outcome. The risk factor score is ordinal (out of 3) depending on whether the data used in the study was cross-sectional, retrospective, or prospective. The causal risk factor score is scored out of 7 with the lowest level being "A study without a comparison group and no analysis of change" and the highest being "A randomized experiment targeting the risk factor" (Murray et al. 2009).

Applying the CQC to the empathy and offending systematic review makes for grim reading. Of the 35 studies included, the highest score achieved by any study on the correlated scale was a 2. Studies tended to be case–control with low sample sizes and limited measurement of offending. This means that based on the studies included in the review, we can have little confidence that empathy is a correlate of offending. In addition, all 35 studies used cross-sectional data. This means that, given the studies included in the review, low empathy could lead to offending, or equally, low empathy could be a consequence of offending. Also, the highest level achieved on the causal risk factor scale was 2. This was the lowest possible ranking given that, to be included in the review, the study needed to have measured empathy among both "offenders" and "non-offenders." This means that there was no evidence to suggest that empathy was a causal risk factor for offending.

In summary, the evidence regarding the relationship between empathy and offending is poor because of weak measurement of empathy, weak measurement of offending, and studies of empathy and offending employ relatively low methodological quality standards However, there have been recent attempts to address some of these issues. For example, Jolliffe and Farrington (2006) developed a new measure of empathy (The Basic Empathy Scale) which appears promising, especially as the factor structure has shown some cross-cultural validity (Albeiro et al. 2009; D'Ambrosio et al. 2009). Also, in studies that have related this measure of empathy to self-reported offending and bullying among a medium-sized group (n = 720) of secondary school students, an interesting pattern of results has emerged. Low affective empathy was found to be related to high levels of self-reported offending and bullying among both males and females. Furthermore, affective empathy was independently related to serious male bullying controlling for other important individual differences (e.g., high impulsivity, low intelligence) (Jolliffe and Farrington 2006, 2007, 2011). This fits well with evidence emerging from other areas of psychology which have suggested that it is deficits in affective empathy that characterize more serious offending (Blair 2007).

Empathy and Offending: The Way Forward

One of the first areas that should be addressed to better understand the relationship between empathy and offending is to improve the measurement of empathy. The BES empathy scale might be a good start, but it is unclear how well questionnaire measures capture empathy, especially affective empathy (Eisenberg and Fabes 1990). This is because responding to affective items not only requires a significant amount of insight into one's emotions, but also the ability and willingness to label and disclose these. This might be why questionnaire measures of empathy have been described as measures of how empathic one wants to be seen.

Ideally a multi-method study of empathy and potentially overlapping traits such as impulsiveness and intelligence could be undertaken, much like the multi-method multisource study of impulsivity carried out by White et al. (1994) who carried out as part of the Pittsburgh Youth Study. In this investigation, 430 boys (aged about 12) completed 11 different measures of impulsivity and these were then compared with measures of intelligence, socioeconomic status, and self-reported offending.

The assessment of empathy should involve the administration of a number of questionnaire measures of empathy (e.g., the Basic Empathy Scale), and verbal responses to movie vignettes, perhaps while monitoring galvanic skin response; a measure of autonomic arousal (as another form to assess the experience of emotions) to a large number of both males and females. The intercorrelations of these measures would be very informative, but it would also be essential that the measures of empathy were compared to measures of potentially overlapping constructs such as sympathy, perspective taking, emotional intelligence as well as established risk factors such as intelligence and impulsivity. This would be especially helpful in clarifying the extent to which cognitive empathy might overlap with intelligence and/or impulsivity so that empathy measures can be selected for future research which minimize this overlap. It might be also be interesting to compare the measures of empathy (and the other measures) to an outcome such as self-reported delinquency, but this would be more informative in terms of determining the direction of effect between empathy and behavior, if it involved a test of prospective validity using a longitudinal study.

Once a device or set of devices have been established to adequately measure empathy, high-quality prospective longitudinal studies are required. Assessing the levels of empathy of a large group of children from an early age, before the onset of antisocial behavior, would help to address whether a lack of empathy is a risk factor for offending. By measuring additional established explanatory risk factors, which have been identified important in predicting offending in prospective longitudinal researchers (e.g., intelligence, impulsivity, parental supervision, parental discipline, socioeconomic status, etc.), the independent relationship between low empathy and offending might be explored (e.g., Farrington 2003; Loeber et al. 2008). By regularly assessing levels of empathy and levels of offending, the extent to which changes in empathy are related to changes in offending could be used to evaluate whether low empathy might be a causal risk factor for offending.

This methodological approach is essential for being able to evaluate the mechanisms by which empathy might have an impact on offending. For example, empathy may be unrelated to later offending (i.e., both low empathy and offending are the result of low intelligence), empathy may measure an underlying cause of offending (i.e., low empathy is a proxy measure of high impulsivity), or empathy may be directly related to offending (i.e., low intelligence leads to low empathy which increases the likelihood of offending). Currently, each mechanism is equally plausible, but in only one mechanism could interventions for offenders designed to increase empathy hope to result in a reduction in later offending.

In conclusion, the research surrounding empathy and its potential relationship to offending is in its infancy. Greater emphasis on understanding the best methods of measuring empathy and higher quality studies are needed to create an empirical basis with which to properly evaluate a potential relationship.

Related Entries

- ▶ Establishing Causes of Offending in Longitudinal and Experimental Studies
- ▶ Psychopathy and Offending
- ▶ Sex Offender Treatment

Recommended Reading and References

Albeiro P, Matricardi G, Speltri D, Toso D (2009) The assessment of empathy in adolescence: a contribution to the validation of the Italian version of the basic empathy scale. J Adolesc 32:393–408

Allport GW (1985) The historical background of social psychology. In: Lindzey G, Aronson E (eds) Handbook of social psychology. Random House, New York

Blair RJR (2007) Empathy dysfunction in psychopathic individuals. In: Farrow T (ed) Empathy in mental illness. Cambridge University Press, New York

Cale EM (2006) A quantitative review of the relations between the “Big 3” higher order personality dimensions and antisocial behaviour. J Res Personal 40:250–284

D’Ambrosio F, Oliver M, Didon D, Besche C (2009) The basic empathy scale: a French validation of a measure of empathy in youth. Personal Individ Diff 46:160–165

Davis MH (1983) Measuring individual differences in empathy: evidence for a multidimensional approach. J Pers Soc Psychol 44:113–126

Eisenberg N, Fabes RA (1990) Empathy: conceptualization, measurement, and relation to prosocial behavior. Motiv Emot 14:131–149

Eisenberg N, Strayer J (1987) Critical issues in the study of empathy. In: Eisenberg N, Strayer J (eds) Empathy and its development. Cambridge University Press, Cambridge, UK

Ellis L, Walsh A (1999) Criminologists’ opinions about causes and theories of crime and delinquency. Criminologist 24:1–4

Ellis L, Cooper JA, Walsh A (2008) Criminologists’ opinions about causes and theories of crime and delinquency. Criminologist 33:23–25

Farrington DP (2003) Key results from the first 40 years of the Cambridge Study in Delinquent Development. In: Thornberry TP, Krohn MD (eds) Taking stock of delinquency: an overview of findings from contemporary longitudinal studies. Kluwer/Plenum, New York

Feshbach ND (1975) Empathy in children: some theoretical and empirical considerations. Counsel Psychol 5:25–30

Feshbach ND (1987) Parental empathy and child adjustment/maladjustment. In: Eisenberg N, Strayer J (eds) Empathy and its development. Cambridge University Press, Cambridge, UK

Fisher D, Beech A, Browne K (1999) Comparison of sex offenders to nonoffenders on selected psychological measures. Int J Offender Ther Comp Criminol 43:473–491
Goldstein H, Higgins-D'Allessandro A (2001) Empathy and attachment in relation to violent vs. nonviolent offense history among jail inmates. J Offender Rehabil 32:31–53
Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford
Hanson KR (2003) Empathy deficits of sexual offenders: a conceptual model. J Sex Aggress 9:13–23
Hayashino DS, Wurtele SK, Klebe KJ (1995) Child molesters: an examination of cognitive factors. J Interpers Violence 10:106–116
Hoffman ML (2000) Empathy and moral development: implications for caring and justice. Cambridge University Press, New York
Hogan R (1969) Development of an empathy scale. J Consult Clin Psychol 33:307–316
Jolliffe D, Farrington DP (2004) Empathy and offending: a systematic review and meta-analysis. Aggress Viol Behav 9:441–476
Jolliffe D, Farrington DP (2006) The development and validation of the basic empathy scale. J Adolesc 29:589–611
Jolliffe D, Farrington DP (2007) The relationship between low empathy and self-reported offending. Legal Criminol Psychol 12:265–286
Jolliffe D, Farrington DP (2011) Is low empathy related to bullying controlling for other factors? J Adolesc 34:59–71
Jolliffe D, Murray J, Farrington DP, Vanneck C (2012) Testing the Cambridge Quality Checklists on a review of disrupted families and crime. Crim Behav Mental Health 22:303-314
Kraemer HC, Lowe KK, Kupfer DJ (2005) To your health: how to understand what research tells us about risk. Oxford University Press, New York
Loeber Rl, Farrington DP, Stouthamer-Loeber M, White HR (2008) Violence and Serious Theft: Development and Prediction from Childhood to Adulthood. Routledge, New York.
Marshall WL, Marshall LE, Serran GA, O'Brien D (2009) Self-esteem, shame, cognitive distortions and empathy in sexual offenders. Psychol Crime Law 15:217–234
Mehrabian A, Epstein N (1972) A measure of emotional empathy. J Pers 40:525–543
Murray J, Farrington DP, Eisner MP (2009) Drawing conclusions about causes from systematic reviews of risk factors: The Cambridge Quality Checklists. J Exp Criminol 5:1–23
Shirtcliff EA, Vitacco MJ, Graf AR, Gostisha AJ, Merz JL, Zahn-Waxler C (2009) Neurobiology of empathy and callousness: implications for the development of antisocial behavior. Behav Sci Law 27:137–171
White JL, Moffitt TE, Bartusch A, Caspi JD, Needles DJ, Stouthamer-Loeber M (1994) Measuring impulsivity and examining its relationship to delinquency. J Abnorm Psychol 103:192–205

Environmental and Human Rights

Avi Brisman
School of Justice Studies,
Eastern Kentucky University,
Richmond, KY, USA

Synonyms

Environmental rights; Human rights

Overview

The term *environmental rights* can refer to *substantive rights* to live in a healthy environment, which entails access to unspoiled natural resources (e.g., air, land, water) necessary for survival; *procedural rights* allowing for access to environmental information, public participation in decision-making that affects (or has the potential to impact) the environment, and access to remedial procedures (i.e., redress before judicial authorities); or *ecological rights* – rights for the environment based on notions of nature's intrinsic value, rather than its utility to humans. *Substantive environmental rights* and *procedural environmental rights* both reflect an anthropocentric (or human-centered) perspective – one that views nonhuman nature instrumentally, that is, "as something to be appropriated, processed, consumed, and disposed of in a manner which best suits the immediate interests of human beings" (Halsey and White 1998:349). The notion of *environmental rights* as *ecological rights* reflects a *biocentric* perspective – one that holds that "non-human species have *intrinsic* value, that is, they possess a moral worth and will continue to have moral worth no matter how insignificant human beings conceive their existence or 'use value' to be" (Halsey and White 1998:352 (emphasis in original)).

Whether *environmental rights* include one or all of the above definitions (or some other conception of *environmental rights*) depends on how the notion of *rights* is understood.

Rights are ethical, legal, moral, or social principles of freedom or entitlement. They are interests or expectations guaranteed by an ethical, moral, or philosophical theory; a legal system; or social convention. Any discussion of *rights* – and any discussion of *environmental rights* – must begin with a consideration of the source of those *rights*.

An overview on sources of rights is provided below (Part II). This is followed, in Part III, by a consideration of sources of *environmental rights* (which depend on the denotation of *environmental rights* as *substantive rights*, *procedural rights*, or *ecological rights*). Part IV examines relationships between different "sectors" of rights, including examples where *environmental rights* overlap with or are encompassed by *human rights* (and vice versa), as well as instances where *environmental rights* conflict with other sets of rights. Part V explores bidirectional linkages between environmental degradation outside of/independent from *environmental rights* discourse and analysis.

Different Perspectives on Sources of Rights

Legal rights are rights guaranteed by law. In other words, they are rights that particular laws classify as rights. They can be considered rules pertaining to what a state's people are allowed or entitled to do and impose a corresponding duty on someone (usually, an agent of government) to ensure that the activity, interest, or expectation granted by the right is not infringed upon or otherwise abridged. *Civil rights* and *political rights* are types of *legal rights*. Usually set forth in a state's constitution or statutes, both *civil rights* and *political rights* protect individuals' freedom from unwanted and unwarranted encroachment by governments and private entities (such as freedom of assembly, religion, and speech). They also ensure one's ability to participate in the establishment or administration of the state without discrimination or repression (such as the right to vote and the right to hold public office). *Civil rights* and *political rights* that permit or oblige action are often referred to as *positive rights* (such as the right to counsel or rights to food, housing, and public education), whereas *civil rights* and *political rights* that permit or oblige inaction are frequently called *negative rights* (such as freedom of speech or freedom from slavery). *Acquired rights* – rights that a person procures (such as the right to own property, the right of first refusal, riparian rights, or a right of way) – constitute other forms of *legal rights*. Regardless of the category, sub-category, or type, *legal rights* are those that particular laws happen to classify as rights; therefore, the existence and facticity of all *legal rights* depend on the society in which they are enacted (Beirne 1999, 2007). As such, because *legal rights* depend on a specific societal context to have meaning and force, they are culturally and politically relative (although the extent of their relativity depends on whether the *legal rights* in question are the product of a state's constitution, statutes, or courts or a multilateral treaty, such as the International Covenant on Civil and Political Rights (ICCPR) and the International Covenant on Economic, Social and Cultural Rights (ICESCR)).

Natural rights are rights that are conceived of as principles that derive from universal conceptions of human nature or divine justice, rather than from legislation or judicial action. As such, *natural rights* are thought to apply to all people and to exist independently of rights created by a government or society (although these rights define the moral function of society and the protection of these universal principles is considered to be the purpose and touchstone of every society or government). Accordingly, *natural rights* (often referred to as *moral rights* or *inalienable rights*) are rights that cannot be transferred or surrendered (such as the right to liberty and the right to life).

Theories of rights can also find their base in ethics or moral philosophy. For example, *utilitarianism* holds that the proper course of action is the one that maximizes "happiness" – which is understood as the predominance of pleasure over pain. Under utilitarian doctrine, political/governmental decisions should be based on promoting the greatest happiness of the greatest number of persons, rather than on abstract individual rights. (Thus, the utilitarian principle of maximizing the total net sum of happiness stands in contrast to *natural rights* theory.) Utilitarians (such as the nineteenth-century Jeremy Bentham and the contemporary Peter Singer) argue that because nonhuman animals possess the capacity for suffering, their interests, like those of humans, must be considered; the theoretical framework of *rights* is not necessary for considering nonhuman animals' interests. Despite the fact that Singer rejects the idea of rights – arguing that the capacity to feel pain is sufficient to consider nonhuman animals' interests – his book, *Animal Liberation* (1975), is considered, by many, to be a (if not *the*) seminal text of the animal liberation movement.

Finally, it is important to recognize that these sources of rights are not mutually exclusive. For example, *human rights* – commonly understood as rights to which all human beings are inherently entitled – may exist as *natural rights* or as *legal rights* in international law (e.g., in the form of treaties, such as the ICCPR and ICESCR; as customary international law) or in national law (e.g., the United Kingdom's Human Rights Act of 1998). The relationship of *human rights* and *environmental rights* is explored further below.

Different Perspectives on Sources of Environmental Rights

As noted at the outset, the term *environmental rights* can be used to denote *substantive rights* (e.g., the right to a healthy, clean environment), *procedural rights* (e.g., freedom of information, the right to participate in the decision-making process, remedial measures), *ecological rights* (e.g., the right of species other than humans to survive), or some combination thereof.

The following three illustrations serve to demonstrate how the source of *environmental rights* will depend on the denotation of *environmental rights*:

1. *Substantive environmental rights* may exist as *natural rights* or as *legal rights*.

   The *right to life* can be conceived of as a *natural right* – one that is not contingent upon the beliefs, customs, or laws of any particular culture or government. If the *right to life* is conceived of as a *natural right* and if *the right to life* is interpreted to include the *right to a clean environment* (which, as noted above, is an example of a *substantive environmental right*), then *substantive environmental rights* or, at least, *this substantive environmental right* – the *right to a clean environment* – can be said to exist as a *natural right*.

   The *right to life* can be conceived of as a *legal right*. Article 21 of the Constitution of India guarantees every individual the right to life and liberty. The Supreme Court of India ruled in the case of *Subhash Kumar v. State of Bihar*, A.I.R. 1991 S.C. 420, 424, that the *right to life* in the Constitution included the *right to the enjoyment of pollution-free water and air*. If the *right to life* is conceived of as a *legal right* and if *the right to life* is interpreted to include the *right to the enjoyment of pollution-free water and air* (a *substantive environmental right*), then *substantive environmental rights* or, at least, *this substantive environmental right* – the *right to the enjoyment of pollution-free water and air* – can be said to exist as a *legal right*.

2. *Procedural environmental rights* are *legal rights*.

   On an international level, the United Nations Economic Commission for Europe Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters (usually known as the Aarhus Convention) grants rights to the public regarding access to information, public participation, and access to justice in governmental decision-making processes regarding matters pertaining to the local, national, and transboundary environment. Most US environmental law, to take a national example, grants administrative agencies significant discretion in setting standards and enforcing them, and provides interested parties with the opportunity to engage with these agencies to determine the implementation of the laws. This engagement may include critiquing environmental impact statements (that are required for actions or projects in which the federal government is involved), commenting on proposed regulations, participating in scientific advisory committees, providing data and information for agencies, and testifying at administrative hearings.

3. *Ecological rights* are (simply) rights, but can be implicated in issues regarding *legal rights*.

Ecological rights are rights of the ecosystem of the Earth apart from and beyond human purpose. Because *ecological rights* reflect the *biocentric* perspective that every form of life has value regardless of its worth to human beings and because most national and international environmental law utilizes an anthropocentric, utilitarian approach to nature wherein nature is considered for its instrumental value to humans (Halsey and White 1998:351; Lane 1998:244) – *ecological rights* cannot be considered as deriving from any of the sources listed above (e.g., from *natural rights*, which are universal, but only to humans). Rather, *ecological rights* are simply *rights* – "above all the right to a continued existence unthreatened by human activities – attached to non-human species, to elements of the natural environment and to inanimate objects" (Miller 1995:375).

Although *ecological rights* do not derive from *natural rights* or *legal rights*, issues regarding the rights of animals other than humans and ecosystems arise when the natural environment (or some component thereof) is threatened by human activity, but the harm or potential harm to humans is minimal or nonexistent. Consider the following:

(a) In the United States, animals other than humans and trees do not have *standing* (a right to make a legal claim in court). Despite arguments that animals other than humans, forests, oceans, and the natural environment as a whole should be bestowed with rights (see, e.g., Stone 1972, 1985, 2010), humans cannot "speak for the trees." Because nonhuman life cannot literally commence legal proceedings by itself, a human must act in its stead. But a person bringing a lawsuit on behalf of nature must, under US law, be able to demonstrate a tangible and particular harm to himself or herself (*see Lujan v. Defenders of Wildlife*, 504 U.S. 555 (1992)).

(b) Numerous countries have supported the Universal Declaration on Animal Welfare (UDAW), a proposed inter-governmental agreement to recognize that animals other than humans are sentient (i.e., they can suffer and feel pain), that the welfare of animals other than humans must be respected, and that cruelty to animals other than humans must end for good. If adopted by the United Nations as a nonbinding set of principles (like the 1948 Universal Declaration of Human Rights (UDHR), discussed below), the UDAW would serve to encourage and enable national governments to introduce and improve legislation and initiatives protecting animals other than humans. The Universal Declaration of Animal Rights (which proclaims, among other things, that all animals are equal, that animals should be respected and humans cannot harm animals, and that animal rights must be equal to those of human rights) has been offered as an alternative to the less radical, less expansive UDAW.

(c) The People for the Ethical Treatment of Animals filed a lawsuit in October 2011 in the United States District Court for the Southern District of California against SeaWorld, arguing that its wild-captured orcas are "enslaved" because they are being held in concrete tanks against their will and forced to perform (Crary and Watson 2011; Watson 2012a, b). The case, which named five whales as plaintiffs, was dismissed in February 2012 on the grounds that the 13th Amendment to the US Constitution, which prohibits slavery and involuntary servitude, applies only to humans (*Tilikum ex rel. People for the Ethical Treatment of Animals, Inc. v. Sea World Parks & Entertainment, Inc.*, 2012 WL 399214 (S.D.Cal.)). More broadly, some have argued that comparing the liberation struggles of humans to those of animals other than humans demeans and trivializes the quest for the emancipation of specific groups of human (see Benton 1998:159 for a discussion).

(d) Nature has few intrinsic rights in national and international environmental law (although laws preserving endangered species are expressed in less utilitarian terms); if law were to shift from an anthropocentric ecophilosophy to a biocentric ecophilosophy, where nature is valued in its own right and possesses *rights*, this would "criminalize previously acceptable behaviour, but also *liberate* behaviour that is currently seen as criminal" (Lane 1998:245).

Key Issues/Controversies: Relationships Between Rights: Tropes and Tensions

Conceptions of rights predate the twentieth century, but the "internationalization" of human rights protection did not occur until after the atrocities of World War II (Hunter et al. 2002:1288) with the 1948 Universal Declaration of Human Rights (UDHR), noted above, part of the International Bill of Rights, along with the previously mentioned ICCPR and ICESCR. The internationalization of environmental protection began 22 years later with the 1972 Stockholm Declaration of the United Nations Conference on the Human Environment. (The UN General Assembly convened the UN Conference on the Human Environment by G.A. Res. 2398 (XXIII) of December 3, 1968. The conference took place in Stockholm from June 5–16, 1972 (see United Nations Conference on the Human Environment, June 5–16, 1972, Report of the United Nations Conference on the Human Environment, U.N. Doc A/CONF.48/14/Rev.1 (1973)).) Although international law often develops as a response to specific challenges, resulting in "sectoral" approaches, concern for the environment can be found in the realm of international human rights law and, vice versa, concern for human rights can be found in international environmental law (Hunter et al. 2002). In the case of the former, the environment is not expressly referred to in the UN human rights system, but the system can be used to advance environmental rights (see Atapattu 2002; Dommen 1998; McInerney-Lankford et al. 2011). Similarly, while the environment is not referred to in the human rights guaranteed by the European Convention for the Protection of Human Rights and Fundamental Freedoms, 213 U.N.T.S. 222 (1950), the European Court of Human Rights has derived the "right to a safe environment" from the "right to life" (see Walters 2007:193; see also Messer 1993:236). In the case of the latter, concern for human rights can be found in international environmental law, such as the 1972 Stockholm Declaration, which provides that "Man has the fundamental right to freedom, equality and adequate conditions of life, in an environment of a quality that permits a life of dignity and well-being, and he bears a solemn responsibility to protect and improve the environment for present and future generations." Because different sectors of rights can overlap, *environmental rights* are intimately associated with other sets of rights or are part of another division of rights and can be conceptualized metonymically or synecdochally (see example #1 below). But because rights can be understood as inhering in individuals or in collective groups (see, e.g., Goodale 2006a, b; Messer 1993), as *descriptive* or *aspirational* – and in this way, more like "goals" (see Dworkin 1977:90-100; see also Hunter et al. 2002:91; Miller 1995:388; Young 1999:28) – and, at least in the international context, are often expressed in broad terms in order to ensure wider acceptance, support, passage, and enforcement, articles or principles of rights can conflict within and between rights regimes (see examples 2 and 3 below).

1. Environmental Rights as Human Rights

Scholars frequently divide human rights into "generations." "First-generation" human rights are *negative rights* – strongly individualistic civil and political rights that serve to protect the individual from private or governmental interference with the enjoyment of liberties. These rights, such as freedom of religion and freedom of speech, appear in places like the US Bill of Rights, in Articles 3–21 of the UDHR, and in the ICCPR (Michalowski 1998).

"Second-generation" human rights are *positive rights* – rights that require not restraint from the state (as "first-generation" human rights do), but "affirmative action on the part of the state" (Barak 1994:264) or the right to "the enjoyment of some substantive benefit such as employment, housing medical care, or education" (Michalowski 1998:524; see also Green and Ward 2000:103). These rights, which are cultural, economic, and social in nature, are embodied in Articles 22–27 of the UDHR, in the ICCPR, and in the European Social Charter, as well as in the constitutions of many nation-states.

"Third-generation" human rights, frequently promulgated by Third World nations, especially those in Africa, are "solidarity" or "development" rights. These largely aspirational (positive) rights are usually expressed in international "soft law" (i.e., non-binding instruments) – such as the 1972 Stockholm Declaration (see Dupuy 1991; Hunter et al. 2002:348–59).

Although *environmental rights* are considered by many to be "third-generation" human rights (see, e.g., Barak 1994:264; Campbell 1999:20; Messer 1993:223), some scholars argue that *environmental rights* do not fit neatly into any single "generation" of human rights. For example, Boyle (2007:471–72) explains that *environmental rights*

> can be viewed from at least three perspectives, straddling all the various categories or generations of human rights. First, existing civil and political rights can be used to give individuals, groups, and nongovernmental organizations (NGOs) access to environmental information, judicial remedies, and political processes. On this view, their role is one of empowerment: facilitating participation in environmental decision-making and compelling governments to meet minimum standards of protection for life, private life, and property from environmental harm. A second possibility is to treat a decent, healthy, or sound environment as an economic or social right, comparable to those whose progressive attainment is promoted by the 1966 United Nations (UN) Covenant on Economic Social and Cultural Rights. The main argument for this approach is that it would privilege environmental quality as a value, giving it comparable status to other economic and social rights, such as development, and priority over non-rights-based objectives. Like other economic and social rights, it would be programmatic and in most cases enforceable only through relatively weak international supervisory mechanisms. The third option would treat environmental quality as a collective or solidarity right, giving communities ("peoples") rather than individuals a right to determine how their environment and natural resources should be protected and managed.

Regardless of how *environmental rights* are conceptualized within the different "generations" of human rights, there has been significant attention to situating *environmental rights* in the human rights arena, with both support for an *environmental human rights jurisprudence* – "the indirect enforcement of environmental protection through human rights claims" – and a "self-standing" human right to an environment of a particular quality under international law (McInerney-Lankford et al. 2011:29), as discussed above in the context of *substantive environmental rights*. But there has also been concern about the tactical merit of litigating issues regarding environmental protection in human rights forums (see, e.g., Boyle 1996; McGoldrick 1996) and that efforts to achieve an autonomous human right to a decent environment could divert attention from more achievable (and potentially more effective) environmental and human rights objectives (Handl 1992). Finally, some have noted the difficulty of balancing (or reconciling, as the case may be) *environmental rights* with existing human rights and human rights objectives (see, e.g., du Bois 1996; McInerney-Lankford et al. 2011:24, 38). Thus, despite the connections between *environmental rights* and *human rights*, proponents of the former can – and often do – butt heads with those concerned about the latter, as the next two sections demonstrate.

2. Conflicts Between Environmental Rights and Property Rights

According to Hunter and colleagues (2002:113), "[o]ver-consumption of natural resources by the wealthiest nations is the single greatest cause of global environmental degradation." Attempting to curb environmental degradation by creating a *substantive environmental right* to a healthy environment or by interpreting a *right to life* as a *right to the enjoyment of pollution-free water and air* could present a tension with any national, regional, or international document that contains language similar to Article 17(1) of the UDHR: "[e]veryone has the right to own property alone as well as in association with others."

A number of commentators have observed that environmental protection, regardless of whether it is couched in the language of rights, is frequently at odds with cherished property rights. For example, McCright and Dunlap (2000:504) explain that "pursuit of environmental protection often involves government action that is seen as threatening core elements of conservatism, such as the primacy of individual freedom, private property rights, laissez-faire government, and promotion of free enterprise" (citation omitted). Maloney (1972:148) describes the tendency of courts in the United States "to overprotect the right to own and use private property and fail[ure] to recognize the ecological consequences of pollution" – a position supported by de Prez (2000:72), who remarks that in Britain, "the right to engage in legitimate business activities enjoys supremacy over considerations of environmental protection." Similarly, Miller (1995:386) observes that "a 'right to breathe clean air' cannot be achieved without prejudice to that most cherished modern expression of the right to liberty, namely, the right to drive wherever and whenever a car-owner wishes." In a different context, Beirne (2007:77) predicts that "in self-avowed property-owning democracies and in societies where individualism is energetically encouraged, the right of animals to be free from cruelty and abuse will ultimately be thwarted, if not altogether undermined, by human rights to privacy and to the enjoyment of private property." Indeed, Western society clings so inveterately to notions of private property rights that one commentator asserts that "people in a consumer society believed they now have an implicit *right to consume*" (Hayward 2004:161 (italics in original)). Despite increasing awareness of our global interconnectedness and interdependency, proponents of *environmental rights* may find that property ownership and consumption have become so intertwined with Western notions of identity and personhood that the distinctions between the right to property and the right to freedom of expression and the right to life have collapsed.

3. Conflicts Between Environmental Rights and Cultural and Indigenous Rights

Concern for property rights is not just a feature of Western individualism and consumption. The global South also possesses an interest in property-rights protection, although here, the tension is one between *environmental rights* and rights to political self-determination, sovereignty and management of indigenous lands, and control over socioeconomic development (see, e.g., Shelton 2009) – which are often subsumed within the categories of *cultural rights* and/or *indigenous rights*.

The UDHR contains some guarantees of cultural rights – which include, for example, "a group's ability to preserve its culture, to raise its children in the ways of its forebears, to continue its language, and not to be deprived of its economic base by the nation in which it is located" (Kottak 2008:198 (citing Greaves 1995)). Rights scholars disagree as to whether cultural rights, which are vested in groups, should be considered to be separate from, rather than a subset of, human rights, which are deemed inalienable, international, and vested in individuals (see, e.g., Glazer 2008:204-11; Michalowski 1985:27; Riles 2006; see also Cooper 2010). For example, Greaves (1995:3) explains that:

> [d]epending on how you look at them, cultural rights are either a special category of human rights, or are a different sort of right altogether. Either way, one can still say that most human rights are vested in individuals. One has the right to speak freely, to hold religious beliefs without prosecution, to not be enslaved or imprisoned without charge, and so on. Human rights protect an

> individual from specified, state-perpetrated injustices. They are rights that a person has vis-à-vis the state in which the person resides. Note that human rights are not ordinary laws; they are thought of as inalienable, metacultural, and to constitute the foundation for every just society. Cultural rights arguably are of the same eternal nature, but unlike human rights vested in individuals, cultural rights are vested in identifiable groups such as religious and ethnic minorities, and indigenous societies. Cultural rights accord specific protections to these groups, who live within a sovereign nation state dominated by others. The right of a group to perpetuate its culture, to raise its children in the way of its forebears, to continue its language, and not to be dispossessed of its economic base are cited as cultural rights.

If cultural rights are a subset of human rights, then when cultural rights and human rights conflict under, say, the UDHR, we are presented with tensions *within* a rights regime; if cultural rights are separate from human rights, then when a right codified under the ICCPR conflicts with one codified in the ICESCR, then we are presented with tensions *between* rights regimes.

To illustrate, environmental groups often share a common goal with indigenous peoples in protecting land from being exploited by developers and nonsustainable agrobusinesses. While Benton (2007:27) notes that "ecological 'enclosures' have been established in the name of biodiversity preservation in recent times, with associated exclusion of indigenous peoples," Reed (2009:239) points out that environmental groups are increasingly integrating indigenous peoples into the day-to-day management of conservation areas. But whereas the short-term objectives of the two groups may be similar (see Messer 1993:236), the long-term goals may be different: environmental groups frequently wish to protect forests and other resources in perpetuity, whereas indigenous groups are "fundamentally concerned with their own survival and see the forests as their primary resource in that struggle" (Reed 2009:239). As a result, environmental groups may discover that as they move to protect certain resources, they are in conflict not only with developers, but also with "the claims of primordial rights by indigenous residents" (Reed 2009:238). At the same time, indigenous groups sometimes find that they need to defend their ancestral lands from a new "developer" – the environmental protection group (Reed 2009:238; see also Messer 1993:236, 239). As Messer (1993:236) explains, "indigenous peoples increasingly come up against not only the development demands of states, transnational corporations, transnational religions, and NGOs, but sometimes also the anti-development demands of the environmental movement with which indigenous peoples might otherwise ally" (internal citations omitted).

To offer a more elaborate illustration, few would disagree that the "right to life" (Article 3 of the UDHR) includes (or should include) a "right to food" (see, e.g., Schwendinger and Schwendinger 1975:133–34). The right to food, in turn, has been increasingly used as "justification to protect the livelihoods of small-scale fishers [sic] around the globe" – "artisanal," "subsistence," or "traditional" fisheries that often exploit fishery resources and degrade the ecosystems on which they depend (Hauck 2007:272). According to Walters (2007:187), "eco-crime," which involves acts of environmental harm and ecological degradation, is "an act of violence and should be viewed as a human rights violation as citizens are deprived of freedoms and liberties." With this conception of eco-crime in mind and given that the European Court of Human Rights has identified the "right to a safe environment" (see above), it stands to reason that small-scale fisheries could be engaging in a human rights violation (ecological degradation) in the process of exercising their human rights (right to food) – or right to work (under Article 23 of the UDHR). While efforts could be undertaken in some situations to help subsistence fisheries maintain compliance with laws intended to ensure the sustainability of fisheries (such as closed seasons and restrictions on types of fishing gear), it is not readily apparent what solutions could be offered where the fisheries are already in serious decline or have collapsed.

A third illustration involves the Makah Indians – a tribe of fishermen and whalers (numbering about 1,500) that lives near the mouth of the Strait of Juan de Fuca on the Olympic Peninsula (in northwest Washington state). The Makah were guaranteed the right to hunt whales in an 1855 treaty with the United States (the Treaty of Neah Bay) – the only tribe with such an explicit treaty provision (see Kershaw 2005). The Makah decided to stop their whaling practices in the 1920s because the commercial whaling industry had depleted the gray whale species. Later, whale hunting would be regulated nationally (under the Marine Mammal Protection Act of 1972 (MMPA)) and internationally (under the International Convention for the Regulation of Whaling (ICRW)), and the USA would list the gray whale as endangered. And while several tribes of Alaska Natives are exempt from the MMPA, allowing them to engage in subsistence hunting of bowhead whale, the Makah were not granted an exemption (despite their treaty rights) (see Kershaw 2005), nor could they obtain any form of exemption under the ICRW. As Jenkins and Romanzo (1998:102) explain, "[i]t is a basic principle of international law that a party to an international treaty, such as the ICRW, may not invoke the provisions of its internal law as justification for its failure to perform under a treaty. Although the United States may consider the Treaty of Neah Bay to confer domestic legal rights upon the Makah Nation, the Makah Nation is not an international legal entity separate from the United States. Thus, as a matter of international law, the Treaty of Neah Bay is simply a domestic legal matter to be worked out between the US government and the Makah tribe: when the United States became a signatory to the ICRW, this act constructively abrogated any conflicting US obligations to the Makah Indians on the international plane."

When the gray whale population began to recover – it was removed from the Endangered Species List in 1994 (Endangered and Threatened Wildlife and Plants, 59 Fed. Reg. 31,094 (1994) (codified at 50 C.F.R. § 222.23)) – the Makah sought and received permission to hunt again. On May 17, 1999, the Makah conducted their first successful whale hunt in more than 70 years. A series of legal challenges, however, have prevented the Makah from further whale hunting. Nevertheless, on Saturday, September 8, 2007, five members of the Makah tribe harpooned and shot a 40-ft-long gray whale during a rogue hunt – one without a federal permit or the tribe's permission (Associated Press 2007a; Kennedy 2007). After suffering for hours, the whale eventually died and sank to the bottom of the waters in the Strait of Juan de Fuca. Tribal leaders denounced the killing (Associated Press 2007b). Two of the Makah tribal members were found guilty in federal court of misdemeanor violations of the MMPA and were sentenced to prison; the three others reached a plea deal with federal prosecutors that included community service and probation (Mapes 2008).

Under the ICRW, the International Whaling Commission (IWC) can grant an Aboriginal Subsistence Whaling (ASW) quota. The hurdle for receiving an ASW exemption is high – requiring both nutritional-subsistence and cultural needs. According to Jenkins and Romanzo (1998:75), "[t]he Makah cannot prove more than a weak (or nonexistent) cultural, or nutritional-subsistence need to whale for two reasons: (1) the Makah do not meet the stringent *cultural need* component of the IWC test for an ASW quota, because they lack a *continuing traditional dependence* on whaling, having voluntarily discontinued whaling over 80 years ago while successfully adapting their culture in the absence of such whaling; and (2) the Makah do not fulfill the *nutritional-subsistence* need component of an ASW quota, because they do not intend to engage in the traditional form of *opportunistic hunts* required for the exemption, and because they have no dietary reliance whatsoever on whale meat."

Although animal rights advocates oppose whaling on ethical and moral grounds, the Humane Society, which sued to stop the Makah from hunting, has distinguished between the Makah's and Alaskan tribes' desires to hunt – referring to the latter as "a true subsistence hunt" and the former as ceremonial or "cultural whaling" (Kershaw 2005) that is not essential to their diet and for which there are "nutritional-subsistence alternatives" (Jenkins and Romanzo 1998). All the while, the Makah continue to press their claims that they should be able to hunt gray whales under the provisions of the Treaty of Neah Bay and because of whaling's cultural significance (Hunter et al. 2002:992; see also <http://www.nwr.noaa.gov/marine-mammals/whales-dolphins-porpoise/gray-whales/makah-whale-hunt.cfm>).

Conclusion and Future Directions: Environmental Degradation, Protection and Human Rights

In summary, this entry has considered sources of environmental rights, the notion of environmental rights as human rights, and relationships between environmental rights and property rights and between environmental rights and cultural and indigenous rights. *Environmental rights* can be conceptualized as distinct rights or as part of or an extension of the broader realm of human rights. How *environmental rights* should be understood and/or realized with respect to human rights may be subject to further debate, but the fundamental point is that environmental degradation and environmental protection are intertwined with both human rights and environmental rights. This is likely to receive more prominent recognition in the future as, for example, human rights obligations (both substantive and procedural) are likely to be relevant to the design and implementation of effective responses to climate change and other global environmental harms.

Related Entries

- ▶ Crimes Against Animal Life
- ▶ Crimes of Globalization
- ▶ Green Criminology
- ▶ Ecoterrorism
- ▶ Environmental Regulation and Law Enforcement
- ▶ Forced Migration and Human Rights
- ▶ Green Criminology
- ▶ Human Rights Violations in Criminal Court
- ▶ Nuclear Weapons and State Crime
- ▶ Organized Crime and the Environment
- ▶ Victims and the International Criminal Court

Recommended Reading and References

Associated Press (2007a) National briefing: Northwest: Washington: charges in whale killing. The New York Times October 6:A10

Associated Press (2007b) Tribe did not authorize whale shooting. USA Today September 9. Accessed at http://www.usatoday.com/news/nation/2007-09-09-whale_N.htm

Atapattu S (2002) The right to a healthy life or the right to die polluted?: the emergence of a human right to a healthy environment under international law. Tulane Environ Law J 16:65–126

Barak G (1994) Crime, criminology, and human rights: toward an understanding of state criminality. In: Barak G (ed) Varieties of criminology: readings from a dynamic discipline. Praeger, Westport, pp 253–267

Beirne P (2007) Animal rights, animal abuse and green criminology. In: Beirne P, South N (eds) Issues in green criminology: confronting harms against environments, humanity and other animals. Willan, Cullompton, pp 55–83

Benton T (1998) Rights and justice on a shared planet: more rights or new relations? Theor Criminol 2 (2):149–175

Benton T (2007) Ecology, community and justice: the meaning of green. In: Beirne P, South N (eds) Issues in green Criminology: Confronting harms against environments, humanity and other animals. Willan, Cullompton, pp 3–31

Boyle A (1996) The role of international human rights law in the protection of the environment. In: Boyle A, Anderson M (eds) Human rights approaches to environmental protection. Oxford University Press, Oxford, pp 43–69

Boyle A (2007) Human rights or environmental rights? A reassessment. Fordham Environ Law Rev 18:471–511

Campbell T (1999) Human rights: a culture of controversy. J Law Soc 26(1):6–26

Cooper B (2010) New birth of freedom. The New York Times Book Rev September 26:16

Crary D, Watson J (2011) PETA lawsuit seeks to expand animal rights. Associated Press/Yahoo! News October 25. Accessed at www.news.yahoo.com/peta-lawsuit-seeks-expand-animal-rights-222219887.html

de Prez P (2000) Excuses, excuses: the ritual trivialisation of environmental prosecutions. J Environ Law 12:65–77

Dommen C (1998) Claiming environmental rights: some possibilities offered by the United Nations' human rights mechanisms. Georget Int Environ Law Rev 11:1–48

du Bois F (1996) Social justice and the judicial enforcement of environmental rights and duties. In: Boyle A, Anderson M (eds) Human rights approaches to environmental protection. Oxford University Press, Oxford, pp 153–175

Dupuy P-M (1991) Soft law and the international law of the environment. Mich J Int Law 12:420–435

Dworkin R (1977) Taking rights seriously. Harvard University Press, Cambridge

Glazer IM (2008) We're the bipeds: an introduction to anthropology, 5th edn. New Life Publishing, Rockaway Park

Goodale M (2006a) Introduction to "anthropology and human rights in a New Key". Am Anthropol 108 (1):1–8

Goodale M (2006b) Ethical theory as social practice. Am Anthropol 108(1):25–37

Greaves TC (1995) Problems facing anthropologists: cultural rights and ethnography. Gen Anthropol 1(2):1, 3–6

Green PJ, Ward T (2000) State crime, human rights, and the limits of criminology. Soc Justice 27(1):101–115

Halsey M, White R (1998) Crime, ecophilosophy and environmental harm. Theor Criminol 2:345–371

Handl G (1992) Human rights and protection of the environment: a mildly revisionist view. In: Cancado Trindade AA (ed) Human rights approaches to environmental protection. Oxford University Press, Oxford, pp 117–142

Hauck M (2007) Non-compliance in small-scale fisheries: a threat to security. In: Beirne P, South N (eds) Issues in green criminology: confronting harms against environments, humanity and other animals. Willan, Cullompton, pp 270–289

Hayward KJ (2004) City limits: crime, consumer culture and the urban experience. The GlassHouse Press, London

Hunter D, Salzman J, Zaelke D (2002) International environmental law and policy 2/e. Foundation Press, New York

Jenkins L, Romanzo C (1998) Makah whaling: aboriginal subsistence of a stepping stone to undermining the commercial whaling moratorium? Colo J Int Environ Law Policy 9:71–112

Kennedy MJ (2007) Tribal group kills whale off Washington. The New York Times September 11:A21

Kershaw S (2005) In petition to government, tribe hopes for return to whaling past. *The New York Times*, 19 Sept:A12

Kottak CP (2008) Window on humanity: a concise introduction to anthropology, 3rd edn. McGraw-Hill, New York

Lane P (1998) Ecofeminism meets criminology. Theor Criminol 2(2):235–248

Maloney FE (1972) Judicial protection of the environment: a new role for common-law remedies. Vanderbilt Law Rev 25:145–164

Mapes LV (2008) 2 Makahs to serve time for illegally killing whale. The Seattle Times July 1. Accessed at http://seattletimes.nwsource.com/html/localnews/20080267 95_makah01m.html

McCright AM, Dunlap RE (2000) Challenging global warming as a social problem: an analysis of the conservative movement's counter-claims. Soc Probl 47(4):499–522

McGoldrick D (1996) Sustainable development and human rights: an integrated conception. Int Comp Law Q 45(4):796–818

McInerney-Lankford S, Darrow M, Rajamani L (2011) Human rights and climate change: a review of the international legal dimensions. The World Bank, Washington, DC

Messer E (1993) Anthropology and human rights. Annu Rev Anthropol 22:221–249

Michalowski R (1985) Order, law and crime. Random House, New York

Michalowski R (1998) All or nothing: an inquiry into the (im)possibility of cause lawyering under socialism. In: Sarat A, Scheingold S (eds) Cause lawyering: political commitments and professional responsibilities. Oxford University Press, New York, pp 523–545

Miller C (1995) Environmental rights: european fact or english fiction? J Law Soc 22(3):374–397

Reed R (2009) Two rights make a wrong: indigenous peoples versus environmental protection agencies. In: Podolefsky A, Brown PJ, Lacy SM (eds) Applying cultural anthropology: an introductory reader, 8th edn. McGraw-Hill, New York, pp 233–240

Riles A (2006) Anthropology, human rights, and legal knowledge: culture in the iron cage. Am Anthropol 108(1):52–65

Schwendinger H, Schwendinger J (1975) Defenders of order or guardians of human rights? In: Taylor I, Walton P, Young J (eds) Critical criminology. Routledge & Kegan Paul, London, pp 113–146

Shelton D (2009) Environmental rights and Brazil's obligations in the inter-American human rights system. George Wash Int Law Rev 40:733–777

Singer P (1975/1990). Animal liberation. New York, HarperCollins

Stone CD (1972) Should trees have standing? – toward legal rights for natural objects. South Calif Law Rev 45:450–501

Stone CD (1985) Should trees have standing? Revisited: how far will law and moral reach? A pluralist perspective. South Calif Law Rev 59:1–154

Stone CD (2010) Should trees have standing?: law, morality, and the environment, 3rd edn. Oxford University Press, New York

Walters R (2007) Crime, regulation and radioactive waste in the United Kingdom. In: Beirne P, South N (eds) Issues in green criminology: confronting harms against environments, humanity and other animals. Willan, Cullompton, pp 186–205

Watson J (2012a) San Diego judge to decide future of whale case. Associated Press/Yahoo! News

February 6. Accessed at www.news.yahoo.com/san-diego-judge-decide-future-whale-case-201027887.html

Watson J (2012b) Slavery protections for animals? Judge to decide. Associated Press/Yahoo! News February 7. Accessed at www.news.yahoo.com/slavery-protections-animals-judge-decide-234713082.html

Environmental Crime

▶ History of Green Criminology

Environmental Criminology

▶ Geography of Crime and Disorder

Environmental Extremism

▶ Ecoterrorism

Environmental Regulation and Law Enforcement

Rob White
School of Social Sciences, University of
Tasmania, Hobart, TAS, Australia

Overview

This entry considers issues of environmental regulation and law enforcement, particularly in relation to violations of civil and criminal laws that are designed to prevent harm from occurring. The entry is concerned with describing official and unofficial responses to instances of environmental degradation. Therefore, the focus is on how to deal with offenders and offenses once specific harms or patterns of harm have been identified by the state and by non-state actors such as nongovernment organizations (NGOs). In essence, the concern is with what is to be done about environmental crime.

Global governance occurs within a complex legislative environment that incorporates many different laws, regulations, conventions, and guidelines that relate to local, national, regional, and international jurisdictions. Many different issues and trends are covered, relating to air, land, and water use; biodiversity; the transport and use of hazardous waste; and carbon emissions. Much of the formal state intervention involves regulatory engagement rather than criminal justice proceedings. Moreover, not all harms are criminalized or subject to state prohibition, leaving a space for non-state actors to partake in activities that challenge the lack of state intervention in protecting certain environments (e.g., campaigns against deforestation) or species (e.g., antiwhaling campaigns).

The entry discusses the limits and opportunities for enhanced environmental regulation and law enforcement in a period witnessing the continued and growing destruction of environmental well-being on a global scale. For this enhancement to occur, there is a need for political will in favor of environmental protection and ecological justice, the fostering of stronger enforcement capabilities, close collaboration across agencies and state/non-state sectors, and establishment of robust networks of environmental law enforcement officers and regulators. The central theme of the entry is the necessity for government and nongovernment agencies and actors to work together in a wide range of ways and contexts for the sake of better environmental governance generally.

Environmental Regulation

It is rare that the state uses coercion solely or even as the key lever of compliance in relation to environmental laws. Rather, many different measures are used, frequently in conjunction with each other, as a means to deal with environmental harm. Likewise, there is a range of agencies that are assigned the task of ensuring compliance and enforcing the law vis-à-vis environmental protection.

There are three broad approaches to matters pertaining to environmental criminalization and regulation (see White 2008). Each approach will be discussed separately below, although each is also intertwined with the other in varying ways and to different degrees. The first approach places the emphasis on *social regulation*, using many different means, as the key mechanism to prevent and curtail environmental harm. Attempts are made to reform existing systems of production and consumption through adoption of a constellation of measures, including enforced self-regulation and bringing nongovernment groups directly into the regulatory process. In the second approach the emphasis is on the use of *criminal law* as presently constituted. The focus here is on attempts to improve the quality of investigation, law enforcement, prosecution, and conviction of illegal environmentally related activity. The third approach is based upon *social action* and involves attempts to engage in social transformation through an emphasis on deliberative democracy and citizen participation, and support for radical social and economic change.

The environmental regulation approach emphasizes regulatory strategies that might be utilized to improve environmental performance, including "responsive regulation" (Ayres and Braithwaite 1992) and "smart regulation" (Gunningham and Grabosky 1998). These approaches attempt to recast the state's role by using nongovernment, and especially private sector, participation and resources in fostering regulatory compliance in relation to the goal of "sustainable development." Of great importance to these discussions is the perceived and potential role of third-party interests, in particular nongovernment environmental organizations, in influencing policy and practice (Braithwaite and Drahos 2000; Gunningham and Grabosky 1998).

The main concern of this kind of approach is with reform of existing methods of environmental protection. The regulatory field is made up of many different stakeholders and participants. These include, for example, businesses, employees, government agencies, communities, shareholders, environmentalists, regulators, the media, trade customers, financial institutions, consumers, and other relevant parties. The role and influence of various people and agencies is shaped by factors such as resources, training, information, skills, expertise, and legislation. These are also affected by the type of regulation that is the predominant model at any point in time.

It has been observed, for example, that the broad tendency under neoliberalism has been toward deregulation (or, as a variation of this, "self-regulation") when it comes to corporate harm and wrongdoing (Snider 2000). In the specific area of environmental regulation, the general trend has been away from direct governmental regulation and toward "softer" regulatory approaches. This places most of the regulation obligation and responsibility into the hands of the non-state sector.

The continuum of regulation ranges from strict regulation through to no regulation (see White 2008). Measures include environmental impact assessments (EIAs), which generally relate to project development and planning issues, and environmental management systems (EMSs), which relate to business practices and internal use of resources, through to voluntary adoption of good environmental practices. The continuum reflects varying degrees of government intervention and oversight, ranging from high levels of command and control through to scarcely any state involvement whatsoever (the "voluntary" end of the regulatory continuum). In practice, most systems of regulation involve a suite of measures across this continuum.

Two general models stand out when it comes to regulation in general and environmental regulation in particular. The first is Ayres and Braithwaite's notion of "enforced self-regulation" (1992). This is based upon a regulatory pyramid. The usual pyramid of sanctions has an extensive base with the emphasis on persuasion that rises to a small peak of harsh punishment. In the case of business transgressions, to take an example, the progression up the pyramid might include persuasion, a warning letter, a civil penalty, a criminal penalty, license suspension, and license revocation. By combining different forms of regulation, Ayres and

Braithwaite (1992) reconstitute the usual regulatory pyramid such that the bottom layer consists of self-regulation, the next layer enforced self-regulation (via government legislation), the next layer command regulation with discretionary punishment, and, at the top, command regulation with nondiscretionary punishment.

Building upon the insights of these and other writers, Gunningham and Grabosky (1998) argue that what is needed is "smart regulation." This basically refers to the design of regulation that still involves government intervention but selectively and in combination with a range of market and nonmarket solutions and of public and private orderings. The central thesis of "smart regulation" is that recruiting a range of regulatory actors to implement complementary combinations of policy instruments, tailored to specific environmental goals and circumstances, will produce more effective and efficient policy outcomes. Essentially this means incorporating into the regulatory field the full schedule of regulatory options, from direct regulation associated with command and control approaches through to voluntary schemes and economic incentive approaches (see Gunningham et al. 1998).

A number of issues arise in relation to how measures linked to the enforced self-regulation pyramid and smart regulation are utilized in practice. Questions can be asked regarding the following: the standards of what is deemed to be acceptable, the flexibility required in devising appropriate safeguards and strategies at local or site level, how to enact total management planning, what constitutes adequate monitoring, who is to do enforcement and compliance, what penalties and consequences are to consist of, how a plurality of instruments rather than a single approach is to be coordinated, how to deal with a culture of reluctance to use punitive measures against corporate misconduct, the general corporate immunity from prosecution and penalty, and why and how the extent of regulation varies according to size of firm.

What detailed examination of particular forms of regulation show, and what explorations of different approaches to environmental regulation acknowledge, is that how regulation is carried out in practice, and whose interests are reflected in specific regulatory regimes is basically an empirical question (see Stretesky 2006). That is, regulatory performance cannot be read off from an abstract understanding of regulation theory as such. Nevertheless, environmental regulation models directly influence the scope and possibilities of environmental regulation as it gets translated into practical measures at the ground level. The adoption of particular environmental regulation models thus helps to shape the methods and behavior of regulators. In ideal terms, the two key models of regulation discussed here would incorporate a range of actors and measures in order to "keep things honest," presumably in ways that would be to the advantage of all stakeholders.

However, the continuing degradation of the environment today is linked to the dominant regulation and enforcement framework that puts the stress on self-regulation and deregulation. This is reflected in state policies and practices. For instance, very often the preference on the part of state authorities is for education, promotion, and self-regulation rather than imposition of directive legislation and active enforcement and prosecution (White 2008; South 2011). Yet, to be effective, those in charge of regulation and enforcement must be willing to utilize the "big stick" and to monitor compliance systematically and diligently. For example, persistent and continuous inspections, accompanied by substantive operational powers (including use of criminal sanctions), can in fact lead to rapid positive changes in the prevention of polluting practices (see Commission for Environmental Cooperation 2001; White 2011).

Snider (2000) describes how in Canada, despite policy directives specifying "strict compliance," a permissive philosophy of "compliance promotion" has reigned. Given the tone of mainstream regulation literature (which offers a theoretical justification for enlisting private interests through incentives and inducements), it is hardly surprising that persuasion is favored at the practical level. Close examination of self-regulation models, however, finds evidence of regulatory failures, and this, in turn, indicates that governments cannot totally abdicate responsibility when a regulatory problem requires a state response (see Priest 1997–1998). Certain conditions are necessary if self-regulation, as such, is going to offer an effective form of regulation. The tendency, however, is for governments to shed regulatory functions and responsibilities and to rely upon the rhetoric and savings afforded by self-regulation (including at the international level, as illustrated by the powerful role of the private body, the International Standards Organization, in heavily influencing government policy responses to environmental regulation).

When considering environmental regulation, it is essential to consider the financial and political environment within which regulators are forced to work. For example, while never before in history have there been so many laws pertaining to the environment, it is rare indeed to find extensive government money, resources, and personnel being put into enforcement and compliance activities. Rather, these are usually provided in the service of large corporations, as a form of state welfare designed to facilitate and enhance the business climate and specific corporate interests. The fiscal crisis of the state, as manifest in massive budget cuts in Greece, Italy, Spain, Portugal, Ireland, Britain, and the United States, also bears with it a crisis in the regulatory field. Environmental protection agencies struggle with inadequate monies and demoralized officers as departmental belts are tightened and state priorities are placed elsewhere.

On top of all these difficulties, there are also special challenges for agency responses to transnational environmental harms in that many different jurisdictions have to be mobilized simultaneously around the same aims and objectives. Enforcement practices in these circumstances must be inclusive and comprehensive. This is achievable if there is enough consensus and political support among partner nation-states. Some of the issues that influence the manner and dynamics of global governance include the scale at which regulation and enforcement takes place (e.g., local council of a town or city, the nation-state, a regional grouping such as the European Union, or in the international sphere, Interpol); the type of collaboration, networks, and partnerships established in a particular area; the extent of harmonization of laws, enforcement practices, and communication strategies; and the sort of nongovernment organization involvement allowed, encouraged, and resisted in particular jurisdictional contexts.

Many contemporary regulatory approaches attempt to recast the state's role by using nongovernment, and especially private sector, participation and resources in fostering regulatory compliance around the goal of "sustainable development." Analyses of these new regulatory regimes, however, offer equivocal results in terms of effectiveness. For example, analysis of Canadian environmental law and policies reveals a patchwork of legislative and regulatory measures that fundamentally fail to protect the environment (Boyd 2003). At its broadest level, the ways in which regulation works or does not work are fundamentally shaped by systemic imperatives and philosophical vision. For instance, Boyd (2003) contrasts a model of regulation based upon an effort to mitigate the environmental impacts of an energy- and resource-intensive industrial economy, with that based upon ecological principles that are oriented to decreasing the consumption of energy and natural resources. However complex the laws and regulations in the first scenario, they cannot succeed in achieving sustainability because the system as a whole is inherently geared to growth in energy and resource consumption. In the latter case, the emphasis is on restructuring the economy to incorporate ecological limits and thus to reduce environmental harm over time.

Environmental Law and Criminality

The multiple demands placed upon specific environmental protection agencies by different sections of government, business, and community, and the varied tasks they are required to juggle (e.g., compliance, education, enforcement) may lead to a dilution of their enforcement capacities and activities in both the national sphere and the international arena. However, this by no means derogates the importance of a “bottom line” when it comes to compliance with environmental laws and rules. Environmental regulation ultimately rests upon the grounded activities of environmental enforcement officers and courtroom practices that add punch to regulatory compliance. A range of penalty types, approaches, and mechanisms can be identified in regard to environmental sanctions. These fall into the broad categories of civil, administrative, and criminal justice responses. Recent developments in this area include the following types of sanctions (White 2010):

- Prosecution as a central tool in enforcement and compliance activities, which means using the full application of criminal laws and criminal sanctions strategically and in proportion to the nature of the offense, including the use of imprisonment
- Alternative sentencing mechanisms which involve the compulsory contribution of offenders to an environmental project that requires restoration or enhancement of the environment
- Civil penalties for less serious breaches of environmental law, which ensure timely and efficient application of sanctions appropriate to the nature of the offense
- Imposition of stricter liability regimes (and use of nominated accountability) given the technical and resource difficulties in prosecuting large companies, which criminalizes actions in ways that allow courts to sidestep some issues of mens rea in cases of corporate crime
- Tailored enforcement approaches that take into account organization type, which means that sanctions such as fines are suited to the firm type rather than the offense committed
- Restorative justice and enforceable undertaking approaches that can involve the offender, victim, and community mutually discussing the nature of the offense and suitable remedies, as a prosecution alternative, which are aimed at repairing the harm at a substantive level

The sanctioning process for environmental offenses presently covers a broad range of strategies, with new possibilities on the horizon. Bell and McGillivray (2008), for instance, mention the use of cumulative penalties, as in the case of point systems in motoring offenses, so that a penalty infringement notice (PIN) does not become “routine” or permit wealthy operators the “right” to pollute. The more often you cause harm, the greater the penalty each time.

The recent trend in countries like Australia toward alternative sentencing options reflects both the difficulties of prosecution, especially in relation to corporations, and a shift in thinking away from the offender toward addressing the nature of the offense. For instance, the range of sentencing options around Australia, while they vary from jurisdiction to jurisdiction, includes an increasing number and type of orders (Preston 2007):

- Orders for restoration and prevention
- Orders for payment of costs, expenses, and compensation
- Orders to pay investigation costs
- Monetary benefits penalty orders
- Publication orders
- Environmental service orders
- Environmental audit orders
- Payment into environmental trust or for other purposes
- Order to attend training
- Order to establish training course
- Order to provide financial assurance

How this burgeoning range of sentencing options translates into particular sentencing outcomes warrants ongoing and close scrutiny. This is not only a matter of how the tension between compliance and facilitation is reconciled within specific institutional settings (such as an Environmental Protection Agency). It also goes to the heart of the processes of prosecutorial and judicial intervention on issues pertaining to environmental harm and the valuing of environmental harm in and by the criminal justice system (White 2011).

Research is necessary, for example, to investigate how “value” is perceived by magistrates and judges in relation to particular environmental offenses, as reflected objectively in sentencing outcomes (i.e., sentencing patterns over time in relation to various environmental offenses). What is the (ecological and economic) cost, for example, of illegal land clearance, and how do we ascertain a suitable penalty for the harm caused by this activity? Similar questions can be asked in regard to illegal fishing, illegal logging, and so on. These kinds of issues are being addressed in various ways. For example, Interpol provides information to support the work of prosecutors of environmental crimes, while in England a substantial tool kit has been prepared to guide magistrates in assessing the seriousness of environmental offenses, determining sentencing criteria for environmental offenses, and working through specific types of cases (Interpol Pollution Crime Working Group 2007; Magistrates Association 2009). These documents are underpinned by the idea that we need to take environmental crime seriously, and to do so, we need sanctions that reflect the seriousness of the crime.

Environmental Law Enforcement

Similar types of training packages are also starting to be produced in the specific area of environmental law enforcement. For example, in South Africa a recent training manual for law enforcement agencies provides an outline of the key legal principles of environmental law; the types of environmental crime, issues relating to criminal prosecution (including consideration of strict liability and vicarious liability), environmental inspection, and investigation; the gathering of evidence (including different types of evidence), factors relevant to prosecution; and the nature of the trial process (Akech and Mwebaza 2010). There are many different challenges for environmental law enforcement, not the least of which is the need to collaborate and network with state and NGO agencies across many different jurisdictions. Getting offenders into court in the first place is the role of environmental law enforcement officials.

Environmental harm, as construed by law enforcement agencies, is basically about the violation of national and international laws put in place to protect the environment. What is legally deemed to be "bad" or criminal, therefore, is the main point of attention, whether this is illegal trade in wildlife and plants, or pollution of the air, water, and land.

Many jurisdictions have specialist agencies to tackle particular sorts of crime. An Environmental Protection Agency (EPA), for example, may be given the mandate to investigate and prosecute environmental crimes. The police may play only an auxiliary role in relation to the work of these agencies. In other circumstances and for other purposes, members of the police service may be especially trained as law enforcement officers in defined areas of work.

Many different agencies, at many different levels, deal with environmental crimes. For some, the central mandate of the agency is driven by the specific type of crime. For example, environmental protection agencies (or their equivalent) often focus on "brown" issues pertaining to pollution and waste. Forestry commissions or national park authorities (or equivalents) tend to concentrate on "green" issues and so deal with matters of conservation, animal welfare, and land use. Bodies such as the Royal Society for the Prevention of Cruelty to Animals (or equivalent) are charged with the responsibility to intervene in cases of harsh treatment of, particularly, domesticated animals (such as companion animals or those destined to be food). Customs services typically are on the lookout for trade in illegal fauna and flora, as well as the international shipment of toxic wastes and banned substances. Police services may have a general duty to protect animals and monitor the environment, while in some cases being vested with the lead role in wildlife offenses. The regulation and policing of fisheries may involve specific fishery management authorities and specially trained fishery officers. Health departments may be the key authorities when it comes to disposal of radioactive and clinical waste. Park rangers could be tasked with the job of preventing the poaching of animals from national parks and private reserves. The list of agencies that have some role in environmental law enforcement is extensive and probably expanding.

Who is doing what at the domestic level is further complicated by the geographical proximity of countries, as in the case in Europe, which facilitates the transference and extension of crime problems across national borders. This can be compounded by political arrangements, such as a common European passport and currency that makes it easier to move within the boundaries of the EU as a whole. From a domestic policing perspective, this means that there is a need for both vertical connections within any particular national context (around particular crime issues that involve local, regional, and national interests) and horizontal connections with relevant agencies outside of that specific country (given the relative ease of crossover into other jurisdictions).

Environmental policing thus is carried out in the light of both considerable variations in policing functions and agencies and in relation to different levels of government. To put it differently, those who do "policing" work may not be *the* police, and the police are not necessarily involved directly in all types of environmental law enforcement work. Environmental law enforcement includes officials working for local municipal councils (and rural shire councils) through to those working on behalf of Interpol in Thailand. It also includes a wide range of NGOs that operate in various official and unofficial capacities. For instance, animal welfare may be deemed to be the official responsibility of organizations such as the Royal Society for the Prevention of Cruelty to Animals (RSPCA) who then investigate and prosecute cases of animal abuse. Other NGOs, such as Greenpeace, Amazon, may not have an "official" role per se, but nonetheless gather evidence of things such as illegal logging which can then be passed on to relevant police and judicial authorities. Who precisely is going to deal with which type of crime is partly a function of the alleged offense, since this will often dictate the agency deemed to be responsible for a particular area – whether this is disposal of hazardous waste, trade in endangered species, illegal fishing, or money laundering associated with trafficking of illegal forest products.

The composition of domestic police and the orientation of policing may differ from country to country. In some countries, local specialists or specially trained police can also wear more than one hat. For example, in the United States "conservation police" (a term that broadly refers to fish and wildlife officers, wildlife management officers, game wardens, park rangers, and natural resources police) have authority to deal with both conventional crime and environmental crime (Shelley and Crow 2009). This means that in addition to investigation and enforcement of laws relating specifically to fish and wildlife issues, their activities can incorporate more generalist policing concerns, including those involving things such as drug law enforcement and human trafficking. Environmental crime of course includes transnational crimes such as the illegal transborder movement of wildlife species.

The plethora of players and laws demand an approach to law enforcement and compliance that necessarily must be collaborative in nature and that acknowledges that the skill set of domestic police is expanding. Dealing with transnational environmental crime will demand extraordinary efforts to relate to each other across distance, language, and cultural borders; to understand specific issues; to coordinate actions; to enforce national and international laws and conventions; and to gather and share information and intelligence. This occurs within a domestic context as well as internationally. Knowing who to talk with, under which circumstances, requires sophisticated understanding of the mission and capacities of multiple allied law enforcement agencies. It also demands close attention to interagency protocols, such as those relating to the sharing of information and intelligence.

Police cooperation across national and international borders is increasingly important given the greater interaction between national police services in the course of dealing with transnational environmental crime. The local, regional, and global dimensions of transnational crime pose a number of challenges for effective policing. Such crimes may be difficult to detect (as in the case of some forms of toxic pollution that are not detectable to human senses, or radioactive emissions from nuclear plants after disasters). They may demand intensive cross-jurisdictional negotiation and intervention, and even disagreement between nation-states, in regard to specific events or crime patterns. Some crimes may be highly organized and involve criminal syndicates, similar to illegal drug trading or the illegal arms trade. Others may include a wide range of criminal actors, ranging from the individual collector of endangered species to the systematic disposal of toxic waste via third parties. These various dimensions of criminality and harm pose particular challenges for law enforcement, especially from the point of view of police interagency collaboration, the nature of investigative techniques and approaches, and the different types of knowledge required for dealing with specific kinds of crime. Moreover, many of the operational matters pertaining to transnational crime are inherently international in scope and substance.

From the point of view of international law enforcement, agencies such as Interpol provide an organizational platform to address major crimes such as terrorism; illegal trade in arms, drugs, and wildlife; and human trafficking. The role of personnel working within and in collaboration with Interpol tends to be advisory rather than substantive. That is, Interpol acts as a conduit for the exchange of information, training, intelligence, and operational knowledge, but does not necessarily carry out direct investigations per se. It provides an active forum in which criminal investigators from around the globe meet to discuss issues such as determining the role of organized crime in specific types of criminal enterprises (e.g., people smuggling) and developing training and enforcement actions to combat particular sorts of criminal activity (e.g., illegal oil pollution into oceans, seas, and inland waterways).

The transborder nature of illegal criminal operations – across state as well as international boundaries – means that often a local (such as Toronto Metropolitan Police) or state/provincial police service (such as Ontario Provincial Police) will necessarily have to work collaboratively with national agencies (such as the Royal Canadian Mounted Police), that, in turn, will have relationships with regional partner organizations (such as Interpol). In some instances, as with the Task Force on Organised Crime in the Baltic Sea Region (which includes representatives from Denmark, Estonia, the European Commission, Finland, Germany, Latvia, Lithuania, Poland, Russia, and Sweden), specific organizational structures are set up in order to share intelligence, in this instance, on environmental crime, and to develop cooperative enforcement structures to deal with offenders.

Dealing with transnational environmental crime demands new ways of thinking about the world, the development of a global perspective and analysis of issues and trends, the formation of formal and informal networks, and a commitment to particular crimes (e.g., toxic and e-waste) or thematic areas (e.g., illegal trade in endangered species) as a priority area for concerted police intervention. The challenges faced by police in affluent countries of the West will be different to their counterparts in third world countries, in countries undergoing rapid social and economic changes, and in countries where coercion and corruption are generally unfettered by stable institutional controls. Law enforcement practices and institutions in countries such as Brazil, Mexico, Indonesia, and the Philippines, for example, are beset by problems such as poor interagency cooperation; inadequate budgetary resources; and technical deficiencies in laws, agency policies, and procedures (Akella and Cannon 2004). These challenges are not unique to these countries: more investment in enforcement policy, enforcement capacity, and performance management is essential regardless of jurisdiction.

Specific forms of criminal law enforcement will require collaboration between different nation-states and different environmental law enforcement services. The development of capabilities in the specific area of transnational law enforcement is necessary and is inevitable given world trends. This includes the “soft skills” of interpersonal communication that enhance cooperation between groups. An important part of this process is the development of a shared consciousness of issues and a sense of what represents justice among enforcement personnel. Understanding the complexities of global issues is an important step in forging a transnational value system protective of collective social interests, ecological well-being, and human rights.

Fighting transnational crime will frequently demand a worldwide response. The role and capacity of domestic law enforcement agencies is an essential component in how responses to harms of a global nature will be framed and implemented. Most intervention occurs locally, even where national and international law enforcement agencies are called in and directly involved. In the end, to be effective, agencies need to be able to harness the cooperation and expertise of many different contributors and to liaise with relevant partners at the local through to the international levels. A “joined-up” approach also means that links can be made between different forms of crime as well as between different agencies and different parts of the world. For instance, illegal fishing (an important environmental crime) has been tied to trafficking of persons, smuggling of migrants, and the illicit traffic in drugs. This is due to the influence of transnational organized crime in the fishing industry worldwide (UNODC 2011).

Yet separate jurisdictions frequently have different laws, policies, and law enforcement practices. The notion of “harmonization” refers to attempts to provide uniformity across diverse jurisdictions. Attempts to harmonize laws, regulations, and practices are increasingly evident in countries with federal systems of governance (e.g., standardization of laws, policies, and practices across the states and territories of Australia), in regions striving for improved forms of integration (e.g., across the member countries of the European Union), and in relation to attempts to create global consensus around issues such as international terrorism, the illicit drug trade, and, for present purposes, environmental crime (e.g., adoption of common definitions and protocols within the context of United Nations’ agencies).

The impetus toward harmonization generally stems from a concern that there is often significant variation in cross-jurisdictional roles, strategies, and practices in relation to specific types of crimes. For transnational crimes such as trafficking in endangered species, harmonization aims to improve capacity and to ensure that police and other law enforcement officials are responding in similar fashion across the globe. In a nutshell, harmonization, at the global level, is informed by the observation that criminal networks operate with little concern for boundaries and borders, form relationships necessary to carry out crimes, and have the ability to act fluidly and adapt to external influences including enforcement activity. By contrast, law enforcement agencies generally operate strictly within regional or national boundaries, do not communicate effectively with one another, and take years to adapt to developing crime trends both tactically and legislatively.

Harmonization of legislation is viewed as a way of overcoming these types of enforcement limitations (Broodman 1991; House of Representatives 2006). It is also perceived as central to a consistent conceptualization of transnational environmental crime, to fostering more effective levels of cooperation among participating states, and contributing to a uniform legal framework. Perceived advantages include reducing complexity and increasing efficiency, and ensuring a consistent approach to particular criminal activities across jurisdictions.

Establishment of uniform laws is only one approach to harmonization endeavors. Such a legal approach may be accompanied by disputes over the content of substantive legislation and by difficulties in the administration and complexity of delivery based upon the legislation. Accordingly, at a practical level, a more productive strategy for harmonization is to focus on consistency in delivering regulatory and enforcement tasks, rather than focusing on uniform legislation as such. For instance, international networks of law enforcement officers (e.g., Interpol or organizations such as the International Network of Environmental Enforcement and Compliance) provide invaluable forums for the exchange of information and knowledge transfer about “best practice” and “what works” in which situations. Participation in common training programs and attendance at conferences and workshops provide opportunities to enhance overall law enforcement capabilities as well as contribute to shared understandings and values in regard to specific types of criminal activity. Importantly, the use of regional case studies and reference to local experiences both reaffirms the importance of acknowledging specific jurisdictional differences and creates opportunities for the adoption of a more balanced view of what constitutes the most productive law enforcement approaches and strategies.

Social Action and NGO Engagement

The main emphasis in regulation and criminal justice approaches is how to best utilize existing legal and enforcement mechanisms to protect environments and creatures within specific environments (e.g., illegal fishing). One can chart existing environmental legislation and provide a sustained socio-legal analysis of specific breaches of law, the role of law enforcement agencies, and the difficulties and opportunities of using criminal law against environmental offenders (Situ and Emmons 2000). This is the key focus of those with a particular interest in the role of the police and other official agencies in environmental law enforcement. For those who view environmental harm in a wider lens than provided by criminal law, this approach has clear limitations (White 2011). In particular, the focus on criminal law, regardless of whether or not the analyst is critical or confirming, offers a rather narrow view of "harm" that can obscure the ways in which the state facilitates destructive environmental practices and environmental victimization. In other words, a strictly legal definition of harm belies the enormous harms that are legal and "legitimate" but that nevertheless negatively impact on people, environments, and animals.

The dearth of adequate controls and regulatory actions within official criminal justice and state offices on matters pertaining to environmental harm is a problem of considerable proportions. To put it simply, not enough is being done to prevent, prosecute, and respond to environmental crime (see White 2010, 2011). Accordingly, it is very often transnational environmental activists who have stepped into the breach, exposing instances of ecological and species harm, providing details of poor regulation and enforcement practices, and contributing both formally and informally to crime reduction and prosecution processes. As increasingly important players in the world of environmental protection, conservation, and management, environmental activists frequently have to both confront powerful social, economic, and political interests and work with and alongside powerful groups, organizations, and state apparatus.

Environmental activism involves many different individuals, groups, and organizations, with diverse aims and missions, employing a wide variety of tactics and strategies. Key international nongovernment organizations (NGOs) include, for example:

- Friends of the Earth
- Greenpeace
- World Wide Fund for Nature
- Sea Shepherd
- Bird Life International
- Climate Action Network
- Biodiversity Action Network
- Humane Society International
- Sierra Club
- Environmental Investigations Agency
- Basel Action Network
- Environmental Justice Foundation

This list of named organizations extends into the hundreds (and indeed, the thousands) and includes local neighborhood action groups through to transnational or global NGOs.

The focus of activists varies greatly (White 2008). "Brown" issues tend to be defined in terms of urban life and pollution (e.g., air quality), "green" issues mainly relate to wilderness areas and conservation matters (e.g., logging practices), and "white" issues refer to science laboratories and the impact of new technologies (e.g., genetically modified organisms). There is generally a link between environmental action (usually involving distinct types of community and environmental groups) and particular sites (such as urban centers, wilderness areas, or seacoast regions). Groups are also demarcated by particular notions of justice, including those relating to environmental justice (e.g., specific human communities), ecological justice (e.g., protection and conservation of particular ecosystems), and species justice (e.g., animal rights and welfare).

Environmental activism deals with acts and omissions that are already criminalized and prohibited, such as illegal fishing or illegal dumping of toxic waste. But it also come to grips with events that have yet to be designated officially as "harmful" but that show evidence of exhibiting potentially negative consequences. It thus deals with different kinds of harms and risks, as these affect humans, local and global environments, and nonhuman animals. Not surprisingly, very often the target for action, and the object of change, is the state. In part this is because much environmental destruction globally is supported by particular nation-states in collusion with powerful corporations. This can take the form of acts of commission or of omission (see Kauzlarich et al. 2003). For example, some acts of harm are perfectly allowable and receive the approval of state authorities (e.g., clearfell logging). Other acts are illegal, but, without adequate state resources, are in effect allowed to occur as a matter of course (e.g., disposal of hazardous waste). Specific types of transnational environmental crime are basically linked in some way to the nature and extent of state intervention (or nonintervention), which in turn depends on the geographic location and political-economic importance of the specific activities in question.

The justification for legal and illegal actions around environmental and animal issues relates to perceptions that many presently legal activities in fact constitute a crime against nature (whether this be a forest or in relation to animals). Conversely, some of the types of actions taken to protest against these alleged crimes are themselves subject to considerable criticism on the basis of their present illegality (and, indeed, the harm they bring to others). Indeed, the use of both confrontational and conciliatory tactics is more than apparent when considering the actions of activists, states, and corporations (see White 2012). The point is that there are complex relations between activists, states, and corporations, relations that shift and change as individuals, groups, and agendas are transformed over time. This extends to issues of environmental governance and law enforcement as well.

At the level of practical interventions, activists engage in varying types of activity. Some organizations engage in militant and spectacular actions (e.g., Greenpeace and Sea Shepherd antiwhaling campaigns). Others focus on specific issues and work closely with governments and international regulatory bodies to enact change. For example, the Antarctic and South Ocean Coalition (ASOC) is an NGO established in 1976 to coordinate the activities of over 250 conservation groups on matters such as Patagonian toothfish management. In so doing, it works closely with governments in confronting issues associated with illegal, unreported, and unregulated fishing. So too do groups such as the Environmental Investigations Agency and the Freeland Foundation for Human Rights and Wildlife, which engage in independent investigations of illegal environmental activities, gathering evidence that is eventually handed over to local police authorities and which is suitable for prosecutions in relevant jurisdictions and courts.

Other groups, such as the Animal Liberation Front (ALF), use a variety of tactics that raise awareness about systematic animal cruelty. Breaking the law (such as illegal entry into animal laboratories or battery hen farms) is considered legitimate if it means that public consciousness is heightened and immediate harms to animals diminished through such actions. Some groups resort to tactics that have been described as ecoterrorism. Earth First!, for instance, has advocated a form of "strategic ecotage," that is, environmentally related sabotage, that has involved sabotaging the machines that destroy forests (e.g., monkeywrenching). Direct action by environmental activist groups has included blockades of logging roads, tree sit-ins, demonstrations, protests, and destruction of machinery including through tree spiking.

As indicated earlier, NGOs can have both formal and informal roles in environmental regulation and law enforcement. So-called "wildlife" NGOs such as the RSPCA may be granted official status and legal rights in regard to investigation and prosecution of animal abuse. Local environmental groups may be given a supplementary regulatory role, officially supported and partially funded by the state, to do such things as monitoring water quality at the regional or municipal level. For those agencies and groups contracted to do this kind of work issues of resources, specialist equipment and staff/volunteer training are important, as are questions concerning the effectiveness of NGO participation in securing good regulatory outcomes.

The role of nongovernment organizations in dealing cooperatively with police and other law enforcement officials to bring criminals to justice should not be underestimated. For example, convictions for illegal logging and for illegal trade in wildlife in places such as Brazil or Russia have been produced through the direct engagement of formal police services with NGO environmental activist organizations. This has involved collection of information and intelligence by NGO personnel that has then been forwarded to relevant authorities in the respective jurisdictions. While the NGOs may at one level be transnational organizations (e.g., Greenpeace), it is relationships forged with domestic police that make the difference when it comes to prosecution and sentencing.

Emerging issues in regard to the specifically regulatory or enforcement role of NGOs include matters such as:

- The nature of expertise (e.g., environmental forensic training, identification skills vis-à-vis endangered species)
- Investigatory skills (i.e., the lack of specialized training in detection, investigation, evidence collection, and case preparation)
- Displacement of formal authority role (e.g., where NGOs end up doing what should be done by formal state agencies)
- Cost-saving measures rather than intrinsic value adding to environmental law enforcement (e.g., EPA reliance upon volunteer groups to collect data, over which the volunteers have no control)
- "Western" agenda setting (e.g., protection of conservation reserves to the detriment of traditional users and owners of the land)

There are many ways in which transnational environmental activist groups can work with official agencies and personnel to achieve similar goals, including sharing of intelligence and joint efforts to gather evidence against wrongdoers. On the other hand, in countries and regions where legal and illegal logging is built into the fabric of state-corporate collusion, it may well be the policing of anti-logging activists that predominates. The specific status and role of NGOs will thus vary depending upon immediate social and political circumstances.

Conclusion

This entry has provided a brief outline of environmental regulation and law enforcement. Contemporary regulation theory stresses the importance of "third parties" as part of the regulatory process. As this entry has demonstrated, such considerations are vital when it comes to environmental crime. This is because of the plethora of "official" state agencies involved in many different types of environmental crime, and the active role of NGOs in likewise investigating and contributing to exposure of environmental harm and offender wrongdoing.

There are thus many different stakeholders engaged in environmental regulation and law enforcement. This makes issues of collaboration, coordination, and harmonization especially important. A common feature of much work in this area, however, is the lack of political will and financial resources in regard to environmental protection. Thus, while EPAs and the like struggle along with meager budgets and low staff numbers, and NGOs do what they can to highlight environmental degradation and species harm, there is still much that needs to be done.

For this to occur, the "environment" must come to the fore as an issue of considerable public interest and public standing. It is the politics of the environment that ultimately determines the extent and effectiveness of environmental regulation and law enforcement.

Related Entries

- ▶ Crimes Against Animal Life
- ▶ Crimes of Globalization
- ▶ Ecoterrorism
- ▶ Environmental and Human Rights
- ▶ Green Criminology
- ▶ Organized Crime and the Environment
- ▶ State-Corporate Crime

Recommended Reading and References

Akech M, Mwebaza R (2010) Enforcement of environmental crime laws: a framework training manual for law enforcement agencies. Institute for Security Studies, Pretoria

Akella A, Cannon J (2004) Strengthening the weakest links: strategies for improving the enforcement of environmental laws globally. Center for Conservation and Government, Washington, DC

Ayres I, Braithwaite J (1992) Responsive regulation: transcending the deregulation debate. Oxford University Press, New York

Bell S, McGillivray D (2008) Environmental law, 7th edn. Oxford University Press, London

Boyd D (2003) Unnatural law: rethinking Canadian environmental law and policy. UBC Press, Vancouver

Braithwaite J, Drahos P (2000) Global business regulation. Cambridge University Press, Cambridge

Broodman M (1991) The myth of harmonisation of laws. Am J Comp Laws 39(4):699–724

Commission for Environmental Cooperation (2001) Special report on enforcement activities: report prepared by the North American working group on enforcement and compliance cooperation. CEC, Montreal

Gunningham N, Grabosky P (1998) Smart regulation: designing environmental policy. Clarendon, Oxford

Gunningham N, Sinclair D, Grabosky P (1998) Instruments for environmental protection. In: Gunningham N, Grabosky P (eds) Smart regulation: designing environmental policy. Clarendon, Oxford

House of Representatives Standing Committee on Legal and Constitutional Affairs (2006) Harmonisation of legal systems: within Australia and between Australia and New Zealand. Commonwealth of Australia, Canberra

Interpol Pollution Crime Working Group (2007) Arguments for prosecutors of environmental crimes, Advocacy Memorandum, 5 June 2007

Kauzlarich D, Mullins C, Matthews R (2003) A complicity continuum of state crime. Contemp Justice Rev 6(3):241–254

Magistrates' Association (UK) (2009) Costing the earth: guidance for sentencers. Magistrates'Association, London

Preston B (2007) Principled sentencing for environmental offences – part 2: sentencing considerations and options. Crim Law J 31:142

Priest M (1997–1998) The privatization of regulation: five models of self-regulation. Ottawa Law Rev 29(2):233–302

Shelley T, Crow M (2009) The nature and extent of conservation policing: law enforcement generalists or conservation specialists? Am J Crim Justice 34(1):9–27

Situ Y, Emmons D (2000) Environmental crime: the criminal justice system's role in protecting the environment. Sage, Thousand Oaks

Snider L (2000) The sociology of corporate crime: an obituary (or: whose knowledge claims have legs?). Theor Criminol 4(2):169–206

South N (2011) Environmental offending, regulation and the legislative balancing act. In: Gobert J, Pascal A-M (eds) European developments in corporate criminal liability. Routledge, London

Stretesky P (2006) Corporate self-policing and the environment. Criminology 44(3):671–708

United Nations Office on Drugs and Crime (2011) Transnational organized crime in the fishing industry. United Nations, Vienna

White R (2008) Crimes against nature: environmental criminology and ecological justice. Willan, Devon

White R (2010) Prosecution and sentencing in relation to environmental crime: recent socio-legal developments. Crime Law Soc Change 53:365–381

White R (2011) Transnational environmental crime: towards an eco-global criminology. Routledge, London

White R (2012) Environmental activism and resistance to state-corporate crime. In: McCulloch J, Stanley E (eds) State crime and resistance. Routledge, London

Environmental Rights

▶ Environmental and Human Rights

Epidemiology of Cardiovascular Disease

▶ Imprisonment and Cardiac Risk

Equivocal Death Analysis

▶ Psychological Autopsy of Equivocal Deaths

Establishing Causes of Offending in Longitudinal and Experimental Studies

David P. Farrington[1] and Rolf Loeber[2]
[1]Institute of Criminology, University of Cambridge, Cambridge, UK
[2]Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA

Synonyms

Causes; Experimental; Longitudinal; Offending; Quasi-experimental

Overview

The causes of offending are often investigated in cross-sectional studies and/or between-individual analyses. However, these types of studies yield poor evidence of causality. If X is a cause, changes in X within individuals should be followed by changes in offending within individuals. Therefore, longitudinal data are needed to investigate causes. Causes can be demonstrated convincingly in studies that investigate whether a life event (e.g., getting married) is followed by an increase in offending or that investigate whether changes in a variable (e.g., parental supervision) are followed by changes in offending. Causes can also be demonstrated convincingly in randomized experiments, but few field experiments have attempted to study the causes of offending. Causes could also be demonstrated in prevention or treatment experiments, but it is difficult to disentangle the "active ingredients" in most of these, because the interventions are usually complex and multimodal. Both quasi-experimental analyses within individuals and randomized experiments can address threats to internal validity or valid causal inference such as selection effects (preexisting differences between "treated" and "control" persons). More of these kinds of studies, and of combined longitudinal-experimental studies, are needed to advance knowledge about the causes of offending.

Introduction

What are the causes of offending is a very important topic in criminology. Virtually, all criminological theories include statements about the causes of offending in order to explain, for example, why some individuals commit more offenses than others, why some individuals commit more offenses at some ages rather than other ages, and why some interventions to prevent or reduce offending are more effective than others. But what is the meaning of a cause, and how can causes be established in criminology? This is the topic of this entry (see also Farrington 1988; Farrington et al. 2010; Loeber and Farrington 2008).

The concept of a cause inevitably involves the concept of change within individual units. A factor X causes a factor Y if, with some specified degree of regularity, and after some specified length of time, changes in X are followed by changes in Y for each individual. For example, the death of a father may cause a decrease in the economic status of his family. As this example shows, the factors X and Y can be dichotomous (father living or dead), continuous (economic status), or of some intermediate kind. The individual unit can be the family rather than an individual person, although this entry will concentrate on changes within individual persons, in discussing how to study the causes of offending.

Causes are often inferred from variations between individuals rather than from changes within individuals. Many researchers draw causal conclusions from cross-sectional (correctional) data, accepting that one variable causes another if (a) the two variables are statistically associated, (b) one variable occurs before the other, and (c) the association holds independently of other (measured) variables. For example, a project might demonstrate that males were more likely to be convicted offenders than females and that this relationship held after controlling statistically for other measured variables. It might then be concluded that gender was a cause of offending. However, drawing conclusions about causes, or in other words about the effect of changes within individuals, on the basis of variations between individuals, involves a conceptual leap that may not be justifiable. For all practical purposes, one cannot investigate whether changing males into females would lead to a decrease in their offending.

Furthermore, the control of other causal factors that might influence an outcome is usually poorer in studies of variations between individuals than in studies of changes within individuals, where each person essentially acts as his or her own control. Nonexperimental studies of variations between individuals inevitably have low internal validity (a poor ability to establish causal influence convincingly) because of the impossibility of measuring and controlling for all possible factors that might influence offending. This is not true of randomized experiments on variations between individuals because – with large samples – the randomization ensures that the average individual in one condition is equivalent to the average individual in another, on all possible extraneous factors. However, it is difficult to study the causes of offending in randomized experiments (see later).

Our argument is that the causes of offending can be investigated most effectively in studies of changes within individuals. Such studies inevitably require longitudinal rather than cross-sectional data, and their internal validity can be increased by combining experimental or quasi-experimental designs with longitudinal data. The remainder of this entry essentially expands the arguments summarized in this introduction. This entry first discusses the investigation of causes in prospective longitudinal surveys, especially using quasi-experimental methods. Then this entry discusses the investigation of causes in randomized experiments and finally reviews the advantages of combining the two methods in longitudinal-experimental studies.

The contributions to the study of causes by Cook and Campbell (1979) have been outstanding. They stressed alternative causal explanations (also called threats to internal validity) that plague experimental and quasi-experimental studies. Among these alternative explanations are the following: (a) History: the observed effect is caused by other explanatory variables changing in the same time period; (b) maturation: the observed effect reflects a preexisting trend; (c) testing: the observed effect is caused by previous testing of the participants; (d) instrumentation: the observed effect is caused by changes in measurement techniques; (e) regression: the observed effect is caused by statistical regression of extreme scorers to the mean; (f) selection: the observed effect is caused by preexisting differences between the groups being compared; (g) mortality: the observed effect is caused by differential attrition from experimental and control groups; (h) instability: the observed effect reflects random variation; and (i) causal order: the true causal order is opposite to that hypothesized. When discussing inferences about causality in this entry, each of these possible threats can serve as a backdrop to the identification of causes. Putative causal effects that can be explained by one of the nine alternatives need to be excluded from further consideration.

The identification of causes is more difficult in some cases than in others. Of the known risk factors for delinquency, some concern discrete events, such as getting married, getting divorced, leaving home, or joining a gang. Establishing the causal effects of these life events tends to be more straightforward than establishing the causal effects of processes that often take place over months or years, such as poor communication between parents and children or poor child-rearing practices.

It should be noted, however, that even discrete causes may correlate with and sometimes operate through long-term processes. An example is becoming a single parent, which may be preceded by prolonged periods of conflict between partners and disagreements about parenting practices. This entry discusses both risk factors that predict a high probability of offending and "promotive" factors that predict a low probability of offending.

Prospective Longitudinal Surveys

The main focus of this entry is on investigating causes in prospective longitudinal surveys which involve repeated measures of the same people. Therefore, they involve at least two data collection points. The word "prospective" implies that risk and promotive factors are measured before outcomes. The most important surveys focus on community samples of hundreds of people, with repeated personal interviews spanning a period of at least 5 years. This entry focusses on community surveys (as opposed to surveys of offenders) because they are needed to study the natural history of offending and the effects of risk/promotive factors and life events. In order to avoid retrospective bias, it is important to measure risk and promotive factors before the development of offending and to calculate prospective probabilities. A minimum of a 5-year time period was set because such a period is needed to provide adequate information about the natural history of the development of offending. Interview data was a requirement because official record data cannot provide adequate information on offending, risk and promotive factors, and life events.

In criminology, the main advantage of these longitudinal surveys is that they provide information about the development of offending over time, including data on ages of onset and desistance, the frequency and seriousness of offending, the duration of criminal careers, continuity or discontinuity in offending, and specialization and escalation. They also provide information about developmental sequences, within-individual change, effects of life events, and effects of risk and promotive factors at different ages on offending at different ages. A great advantage of longitudinal compared with cross-sectional surveys is that longitudinal surveys provide information about time ordering, which is needed in trying to draw conclusions about causes. An investigation of causes leads to an emphasis on change. However, it should not be forgotten that longitudinal data are also useful in studying the opposite of change: the degree of consistency and continuity in behavior over time.

While prospective longitudinal surveys have many advantages, they also have problems. The main challenge in these surveys is to draw convincing conclusions about causal effects. Because of their focus on naturalistic observation, longitudinal surveys find it difficult to disentangle the impact of any particular variable from the effects of numerous others. It is particularly difficult to rule out selection effects; for example, child abuse may predict delinquency because antisocial parents are more likely to abuse their children and are more likely to have delinquent children, without there being any causal effect of child abuse on delinquency. Also, the infrequency of data collection often makes it difficult to pinpoint causal order.

Quasi-Experimental Approaches

A quasi-experimental analysis tries to isolate the impact of a naturally occurring presumed causal factor (e.g., joining a gang) by treating it as though it was experimentally manipulated and then trying to eliminate the plausible alternative explanations of observed effects discussed above. Prospective longitudinal data constitute the foundation for quasi-experimental approaches in studying the causes of offending.

Sometimes a catastrophe or a major beneficial event occurring to a population may trigger vast changes in criminality. For example, Costello et al. (2003) examined the impact of the opening of a casino on an American Indian reservation, which took place in the course of a longitudinal study by the authors (the Great Smoky Mountains Study of Youth). The revenue from the casino was shared by every adult and child tribe member. This led to a reduction in the number of Indian families with income below the federal poverty line, while non-Indian families did not benefit from the casino revenue. They found that children in Indian families showed a significant improvement in behavioral symptoms of oppositional/defiant and conduct disorder. This study is exemplary in that it takes advantage of a natural event that affected some families compared to other families, established temporal order, and shed light on mediating factors.

Almost all studies of the causes of offending have carried out analyses between individuals showing, for example, that unemployed people commit more crimes than employed people and that this relationship holds after controlling for measured extraneous variables. However, as mentioned, analyses within individuals are more relevant to the concept of cause, which suggests that changes within individuals in a causal factor (e.g., from employment to unemployment) tend to be followed by changes within individuals in offending (Farrington 1988). Similarly, analyses within individuals are more relevant to prevention or treatment research (which requires within-individual change). Quasi-experimental analyses within individuals control for individual factors that do not change over time (e.g., gender and race). For example, in the Pittsburgh Youth Study (Loeber et al. 2008), which is a prospective longitudinal survey of over 1,500 Pittsburgh males from age 7 to age 35, Gordon et al. (2004) found that the boys' offending increased after they joined a gang and decreased after they left a gang. The researchers controlled for selection effects, since the boys who joined gangs were more delinquent beforehand than those who did not join.

The Cambridge Study in Delinquent Development

Quasi-experimental analyses within individuals have been carried out in the Cambridge Study in Delinquent Development, which is a prospective longitudinal survey of over 400 London males from age 8 to age 48 (Farrington et al. 2009). One analysis attempted to test labelling theory (Farrington 1977). According to this theory, one effect of a conviction should be to amplify delinquent behavior, perhaps because of the effect of stigmatization on a person's self-concept. However, an alternative causal sequence is proposed by deterrence theory. According to this theory, the effect of a conviction should be to decrease delinquent behavior, because detected offenders become more afraid of the possible consequences of offending.

In comparing these theories, the basic design was to investigate the self-reported offending of boys before and after they were convicted for the first time (between ages 14 and 18). All boys were assigned percentile scores between 0 and 100 according to the number of their admitted offenses, and it was found that the average score of convicted youths increased significantly, from 59 before the conviction to 69 after. In this test, each youth acted as his own control, and it was clear that the convicted youths became relatively more delinquent. This was especially true when they were compared with an individually matched sample of unconvicted youths, whose average scores decreased from 59 to 51 during the same period. However, it was necessary to test a number of plausible alternative explanations of this effect (or threats to internal validity).

The results could not have been caused by maturation (processes within the individuals operating as a function of the passage of time), history (events occurring between the first and the second test), testing (the effect of taking one test on the scores in a second), or instrumentation (changes in the measuring instruments of scoring methods). Each of these factors would have been expected to affect the whole sample equally. The results could not have been caused by mortality (differential loss of individuals from the comparison groups) because the analysis was based only on youths interviewed at all three ages (14, 16, and 18). Furthermore, statistical regression to the mean could not explain the results because the scores of the convicted youths became even more extreme at 18 than they had been at 14.

The major plausible alternative explanation of the results was selection: the convicted youths could possibly have differed from the unconvicted ones in some factor unrelated to conviction that caused an increase in offending. However, no other factor was found that caused such an increase. The convicted youths were individually matched with unconvicted ones at age 14 not only on self-reported offending but also on a prediction score based on the best predictors of delinquency (troublesomeness, low family income, large family size, parental criminality, poor parental child-rearing behavior, and low intelligence). This did not significantly change the results. While these factors predicted high self-reported delinquency at age 18, they did not predict an *increase* in percentile scores between 14 and 18.

One key advantage of this kind of within-individual analysis is that all individual difference factors are controlled for. Another is that causal-order questions can be resolved more satisfactorily than in a between-individual analysis. The main problem of causal order here was whether the conviction preceded the increase in offending or whether the increase in offending preceded the conviction. The percentile self-report scores at age 16 showed that there was no sign of an increase in offending by convicted boys before the conviction occurred, and so it was concluded that the conviction preceded the increased offending.

Some remaining uncertainties about the interpretation of these results center on the theoretical constructs underlying the empirical variables. Did offending increase, or did the likelihood of admitting offending increase? A study of the dates of reported offenses suggested that both effects happened, since the convicted youths at age 16 were more likely to admit offenses committed before age 14 but not admitted at that age. What constructs intervened between convictions and increased offending? One construct that increased in the same way as self-reported offending was a hostile attitude to the police, and so this may have been one link in the causal chain. Finally, did the increase in offending reflect labelling or decreased deterrence? A later study by Farrington et al. (1978) was able to replicate the delinquency-amplifying effects of convictions between ages 18 and 21 and suggested that decreased deterrence was happening. The increase in offending was most marked for those who received the lightest possible sentences, whereas labelling theory might have predicted the opposite.

Developmental and life-course criminology aims to investigate the effects of life events on the course of development of antisocial behavior. In the Cambridge Study, going to a high delinquency-rate school at age 11 did not seem to amplify the risk of offending, since badly behaved boys tended to go to high delinquency-rate schools (Farrington 1972). The boys committed more offenses during periods when they were unemployed than when they were employed, suggesting that unemployment caused crime (Farrington et al. 1986a). However, there was only an increase in crimes leading to financial gain, such as theft, burglary, robbery, and fraud. There was no effect of unemployment on other offenses such as violence, vandalism, or drug use, suggesting that the link between unemployment and offending was mediated by a lack of money rather than by boredom.

It is often believed that marriage to a good woman is one of the most effective treatments for male offending, and indeed, Theobald and Farrington (2009) found that getting married led to a decrease in offending compared with staying single. Also, later separation from a wife led to an increase in offending compared with staying married, and the separated men were particularly likely to be violent (Farrington and West 1995). Another protective life event was moving out of London, which led to a decrease in self-reported violence (Osborn 1980). This was probably because of the effect of the move in breaking up delinquent groups.

One currently popular method of controlling for selection effects is to use propensity score matching. For example, in their investigation of the effect of marriage on offending in the Cambridge Study, Theobald and Farrington (2009) calculated propensity scores indicating the probability of getting married and matched married and unmarried men both on these and on prior offending. They found that convictions decreased after marriage, but only for men who married at relatively young ages (under age 25). Just as randomization equates the probability of each person receiving a treatment, matching on propensity scores aims to equate treated and untreated persons on their prior probability of receiving the "treatment" (here, marriage).

Comparing Changes Within Individuals with Variations Between Individuals

Only one article in criminology has compared whether causes identified by means of within-individual analyses were similar to or different from causes identified by means of between-individual analyses. Farrington et al. (2002) examined the course of offending of over 500 boys in the oldest sample of the Pittsburgh Youth Study over seven data waves between ages 14 and 18 on average. Putative causes were only examined if they were available at each data wave. These putative causes were hyperactivity-impulsivity-attention problems, low school achievement, depressed mood, poor parental supervision, low parental reinforcement, poor parent-boy communication, low family involvement, low social class, poor housing, and peer delinquency. The between-individual correlations were computed for each wave and then averaged across the seven waves. In contrast, the within-individual correlations were calculated for each boy (based on seven waves) resulting in 370–380 correlations for boys who had admitted at least one delinquent act.

All ten variables were significantly correlated with delinquency in the between-individual analyses. The within-individual correlations with delinquency were on average lower and statistically significant for only four variables: peer delinquency, poor parental supervision, low involvement in family activities, and poor parent-boy communication. To test whether the associations held prospectively, subsequent analyses investigated whether variables in one wave predicted delinquency in the next wave. Only poor supervision, low reinforcement, and low involvement predicted within individuals. In conclusion, although peer delinquency was correlated with offending between individuals, it was not a within-individual cause of offending. However, poor parental supervision, low involvement in family activities, and low parental reinforcement appeared to be causes in that as they rose or fell over time, the delinquency of most participants would subsequently rise and fall as well. Thus, temporal covariation between putative risk factors and offending is one of the most convincing demonstrations of causal status.

The results also showed that individuals varied considerably in their within-individual correlations between predictors and delinquency. For some, the correlation was negative, for others it was zero, and for others it was positive. Averaged across individuals, the within-individual correlation tended to be in the direction of the between-individual correlation. The point, however, is that causal factors that explain between-individual differences in offending are not necessarily operating in the same way for all individuals. Thus, the causal status of a particular variable may vary from individual to individual. For example, poor housing was positively related to delinquency for boys living in bad neighborhoods but not for boys living in good neighborhoods.

Randomized Experiments

An experiment is a systematic attempt to investigate the effect of variations in one factor (the independent or explanatory variable) on another (the dependent or outcome variable). In criminology, the independent variable is often some kind of intervention, and the dependent variable is some measure of offending. Most criminological experiments are pragmatic trials designed to test the effectiveness of an intervention rather than explanatory trials designed to test causal hypotheses. The independent variable is under the control of the experimenter; in other words, the experimenter decides which people receive which treatment (using the word "treatment" very widely to include all kinds of interventions).

The focus here is on randomized experiments, where people are randomly assigned to different treatments. Provided that a large enough number of people are assigned (e.g., at least 50 per condition), randomization ensures that the average person receiving one treatment is equivalent (on all possible measured and unmeasured extraneous variables) to the average person receiving another treatment, within the limits of small statistical fluctuations. Therefore, it is possible to isolate and disentangle the effect of the independent variable (the intervention) from the effects of all other extraneous variables (Farrington and Welsh 2005). However, it is also desirable to investigate intervening mechanisms or causal chains (mediators). The main strength of randomized experiments is in excluding selection effects (preexisting differences between persons in different conditions) as a possible explanation.

Few field experiments have been designed to investigate the causes of offending. However, in experiments on stealing, Farrington and Knight (1979, 1980) left stamped, addressed, apparently lost, unsealed letters on the street, each containing a handwritten note and also (except for control conditions) a sum of money. The experimenter, who was blind to the condition of each letter, observed the personal characteristics and behavior of each person who picked up the letter. Each person could honestly mail the letter and money to the intended recipient or could steal the money.

Behavior after picking up the letter predicted stealing. Almost all of the participants were observed to take out the note and read it. Those who then walked along holding the letter were likely to return it, whereas those who put the letter in a pocket or handbag were likely to steal it. This suggested that the decision to steal was made immediately. The prevalence of stealing varied remarkably, from about 20 % to 80 % in different conditions. This suggested that, depending on the experimental conditions, almost everyone would steal or almost no one would steal.

Our experiments were designed to test ideas of deterrence and were inspired by subjective expected utility theories (Farrington 1979). The study found that stealing increased as the amount of money that could be stolen increased, decreased when the apparent victim was an impoverished old lady (high cost) compared with an affluent young man (low cost), and decreased when the probability of detection was greater (with a postal order compared to cash). The study also found that younger people were more likely to steal than older ones, although in most cases (except when there was a large amount of money) there were few gender differences in stealing.

Experiments are usually designed to investigate only immediate or short-term causal effects. However, some interventions may have long-term rather than short-term effects, and in some cases the long-term effects may differ from the short-term ones (Farrington and Welsh 2013). More fundamentally, researchers rarely know the likely time delay between cause and effect, suggesting that follow-up measurements at several different time intervals are desirable. A longitudinal-experimental study deals with many of these problems.

Longitudinal-Experimental Research

More than two decades ago, Farrington et al. (1986b) in their book *Understanding and Controlling Crime: Toward a New Research Strategy* argued that:

1. The most important information about the development, explanation, prevention, and treatment of offending has been obtained in longitudinal and experimental studies.
2. New studies are needed in which these two important methods are combined, by embedding experimental interventions in longitudinal studies.

Longitudinal-experimental studies are needed to advance knowledge about the causes of offending.

There have been a number of longitudinal-experimental studies in criminology in which persons who did or did not receive an experimental intervention were followed up for several years (Farrington 2006). It is not controversial to argue for the desirability of adding a long-term follow-up to a randomized experiment. What is much more controversial is the desirability of embedding an experiment within an ongoing longitudinal survey, essentially because of concerns that the experiment might interfere with the aims of the longitudinal survey such as documenting the natural history of development.

Strictly speaking, every experiment is prospective and longitudinal in nature, since it involves a minimum of two contacts or data collections with the participants: one consisting of the experimental intervention (the independent variable) and one consisting of the outcome measurement (the dependent variable). However, the time interval covered by the typical experiment is relatively short. Farrington et al. (1986b) argued that longitudinal-experimental studies were needed with three elements: (1) several data collections, covering several years; (2) the experimental intervention; and (3) several more data collections, covering several years, afterwards. No study of this kind has ever been carried out on offending using interview data. A few experiments collected official record data retrospectively for a few years before an intervention and prospectively for a few years after the intervention, but these did not assess the effect of the intervention on criminal career trajectories or developmental sequences of offending.

An important advantage of a combined longitudinal-experimental study in comparison with separate longitudinal and experimental projects is economy. It is cheaper to carry out both studies with the same individuals than with different individuals. For example, the effect of interventions and the effect of risk or promotive factors can be compared on the same people. The number of individuals and separate data collections (e.g., interviews) is greater in two studies than in one (other things being equal).

The main advantages of longitudinal-experimental research have been summarized by Blumstein et al. (1988). The impact of interventions can be better understood in the context of preexisting trends or developmental sequences, which would help in assessing maturation, instability, and regression effects in before and after comparisons. The prior information about participants would help to verify that comparison groups were equivalent, to set baseline measures, to investigate interactions between types of persons (and their risk/promotive factors and prior histories) and types of interventions, to establish eligibility for inclusion in the experiment, and to estimate the impact of differential attrition from experimental conditions. The long-term follow-up information would show effects of the intervention that were not immediately apparent, facilitate the study of different age-appropriate outcomes over time, and make it possible to compare short-term and long-term effects and to investigate the developmental sequences linking them. The experimental intervention could help to distinguish causal or developmental sequences from different age-appropriate behavioral manifestations of the same underlying construct.

Conclusions

Longitudinal data make it possible to study both changes within individuals and variations between individuals separately. Because of this distinction, and because of the high internal validity of quasi-experimental analyses, longitudinal data are much more suitable than cross-sectional data for testing causal hypotheses. Furthermore, causal conclusions based on changes within individuals in longitudinal data could in principle have practical implications for preventive or rehabilitative treatment designed to change people, providing that the construct that varies within individuals is manipulable. No possible treatment implications could follow from the argument that a construct that varied only between individuals – such as gender or race – was a cause of offending, since one cannot change a person's gender or race. However, variables that are associated with gender or race could be causes of offending. Similarly, no possible treatment implications could follow from the argument that a non-manipulable construct that varied within individuals – such as age – was a cause of offending.

Treatment implications can readily be drawn from manipulable theoretical constructs. For example, it could be predicted that a decrease in school failure would cause a decrease in offending. This prediction was indeed supported in the Perry Program (Schweinhart et al. 2005). They found that a preschool intellectual enrichment program led to decreases in school failure and decreases in offending. In many respects, a good way to test a hypothesis about the causes of offending is to carry out a prevention or treatment experiment (Robins 1992). This clearly requires a manipulable theoretical construct. However, it is difficult to disentangle the "active ingredients" in most intervention experiments, because the "treatment" is usually complex and multimodal.

Criminologists should attempt to carry out more field experiments to investigate theories of offending, using a realistic measure of offending as the dependent variable. The most feasible dependent variables are probably stealing and vandalism; it is hard to imagine conducting an experiment with real violence as the dependent variable, although verbal aggression might possibly be studied. Most criminological experiments have investigated policing, early prevention, corrections, courts, or community treatment (Farrington and Welsh 2005). There is surely a need for field experiments that try to test theories of offending (Farrington 2008).

In reviewing causality problems arising in longitudinal studies several years ago, Loeber and Farrington (1994) stressed that the major issues were attrition, testing effects, the distinction between aging, period, and cohort effects, and establishing causes with high internal validity. All these issues remain crucial. The authors also stressed the need to combine longitudinal and experimental studies or, at a minimum, to turn experimental studies into longitudinal studies so that the long-term impact of change agents can be ascertained. In addition, there is an urgent need to reinvigorate the search for causes in at least the following ways:

1. A web-based inventory is needed of what is known about the causal status of the 60 or more known putative causes of delinquency. Such an inventory should not only summarize effect sizes but should also investigate which of the several crucial tests have succeeded in narrowing down risk factors into causes. Ideally, such a website should also contain information about promotive causes that foster nondelinquency and can aid in explaining desistance. The *Handbook of Crime Correlates* (Ellis et al. 2009) would be a useful source of references.
2. Systematic reviews and meta-analyses of putative causes of offending should be carried out. These reviews should investigate not only to what extent risk factors predict offending but also to what extent they predict offending after controlling for other putative causes (see, e.g., Ttofi et al. 2011).
3. A systematic survey is needed of the moderators and mediators of causes of delinquency. This will be of immense help in documenting processes that unfold over time.
4. Causes that operate within individuals should be more vigorously investigated so that eventually meta-analyses across different studies can be undertaken, leading to better generalizations about what is known about within-individual causes and their implications for interventions.

Unfortunately, there is very little solid evidence about the causes of offending. More within-individual quasi-experimental analyses in longitudinal surveys, and more randomized experiments, are needed to draw convincing conclusions and advance knowledge significantly.

Related Entries

- ▶ Age-Crime Curve
- ▶ Cognitive/Information Processing Theories of Aggression and Crime
- ▶ Desistance from Crime
- ▶ Labeling Theory
- ▶ Longitudinal Studies in Criminology

▶ Moffitt's Developmental Taxonomy of Antisocial Behavior
▶ Offender Decision Making and Behavioral Economics
▶ Social Learning Theory

Recommended Reading and References

Blumstein A, Cohen J, Farrington DP (1988) Longitudinal and criminal career research: further clarifications. Criminology 26:57–74

Cook TD, Campbell DT (1979) Quasi-experimentation: design and analysis issues for field settings. Rand McNally, Chicago

Costello EJ, Compton SN, Keeler G, Angold A (2003) Relationship between poverty and psychopathology: a natural experiment. J Am Med Assoc 290:2023–2029

Ellis L, Beaver K, Wright J (2009) Handbook of crime correlates. Elsevier, Amsterdam

Farrington DP (2008) Criminology as an experimental science. In: Horne C, Lovaglia M (eds) Experiments in criminology and law. Rowman and Littlefield, Lanham, pp 175–179

Farrington DP (1972) Delinquency begins at home. New Soc 21:495–497

Farrington DP (1979) Experiments on deviance with special reference to dishonesty. In: Berkowitz L (ed) Advances in experimental social psychology, vol 12. Academic, New York, pp 207–252

Farrington DP (2006) Key longitudinal-experimental studies in criminology. J Exp Criminol 2:121–141

Farrington DP (1977) The effects of public labelling. Br J Criminol 17:112–125

Farrington DP (1988) Studying changes within individuals: the causes of offending. In: Rutter M (ed) Studies of psychosocial risk: the power of longitudinal data. Cambridge University Press, Cambridge, UK, pp 158–183

Farrington DP, Knight BJ (1979) Two non-reactive field experiments on stealing from a "lost" letter. Br J Soc Clin Psychol 18:277–284

Farrington DP, Knight BJ (1980) Stealing from a "lost" letter: effects of victim characteristics. Crim Justice Behav 7:423–436

Farrington DP, Welsh BC (2005) Randomized experiments in criminology: what have we learned in the last two decades? J Exp Criminol 1:9–38

Farrington DP, Welsh BC (2013) Randomized experiments in criminology: what has been learned from long-term follow-ups? In: Welsh BC, Braga AA, Bruinsma GJN (eds) Experimental criminology: prospects for advancing science and public policy. Cambridge University Press, Cambridge, in press

Farrington DP, West DJ (1995) Effects of marriage, separation and children on offending by adult males. In: Hagan J (ed) Current perspectives on aging and the life cycle, vol 4, Delinquency and disrepute in the life course. JAI Press, Greenwich, pp 249–281

Farrington DP, Osborn SG, West DJ (1978) The persistence of labelling effects. Brit J Criminol 18:277–284

Farrington DP, Gallagher B, Morley L, St. Ledger RJ, West DJ (1986a) Unemployment, school leaving, and crime. Brit J Criminol 26:335–356

Farrington DP, Ohlin LE, Wilson JQ (1986b) Understanding and controlling crime: toward a new research strategy. Springer, New York

Farrington DP, Loeber R, Yin Y, Anderson SJ (2002) Are within-individual causes of delinquency the same as between-individual causes? Crim Behav Ment Health 12:53–68

Farrington DP, Coid JW, West DJ (2009) The development of offending from age 8 to age 50: recent results from the Cambridge study in delinquent development. Monatsschrift fur Kriminologie und Strafrechtsreform (J Criminol Penal Reform) 92:160–173

Farrington DP, Loeber R, Welsh BC (2010) Longitudinal-experimental studies. In: Piquero AR, Weisburd D (eds) Handbook of quantitative criminology. Springer, New York, pp 503–518

Gordon RA, Lahey BB, Kawai E, Loeber R, Stouthamer-Loeber M, Farrington DP (2004) Antisocial behavior and youth gang membership: selection and socialization. Criminology 42:55–87

Loeber R, Farrington DP (1994) Problems and solutions in longitudinal and experimental treatment studies of child psychopathology and delinquency. J Consult Clin Psychol 62:887–900

Loeber R, Farrington DP (2008) Advancing knowledge about causes in longitudinal studies: experimental and quasi-experimental methods. In: Liberman A (ed) The long view of crime: a synthesis of longitudinal research. Springer, New York, pp 257–279

Loeber R, Farrington DP, Stouthamer-Loeber M, White HR (2008) Violence and serious theft: development and prediction from childhood to adulthood. Routledge, New York

Osborn SG (1980) Moving home, leaving London, and delinquent trends. Brit J Criminol 20:54–61

Robins LN (1992) The role of prevention experiments in discovering causes of children's antisocial behavior. In: McCord J, Tremblay RE (eds) Preventing antisocial behavior: interventions from birth through adolescence. Guilford, New York, pp 3–18

Schweinhart LJ, Montie J, Zongping X, Barnett WS, Belfield CR, Nores M (2005) Lifetime effects: the high/scope Perry preschool study through age 40. High/Scope, Ypsilanti

Theobald D, Farrington DP (2009) Effects of getting married on offending: results from a prospective longitudinal survey of males. Eur J Criminol 6:496–516

Ttofi MM, Farrington DP, Lösel F, Loeber R (2011) The predictive efficiency of school bullying versus later offending: a systematic/meta-analytic review of longitudinal studies. Crim Behav Ment Health 21:80–89

Estimating the Effects of Incapacitation

Robert Apel[1] and Hilde Wermink[2]
[1]School of Criminal Justice, Rutgers University, Newark, NJ, USA
[2]Department of Criminology and Criminal Law, Leiden University, Leiden, The Netherlands

Overview

Among the manifold goals of penal confinement, incapacitation is intended to impose a period of "time out" from an offender's criminal career, by deliberate removal of the opportunity for the offender to commit crime in the community for the duration of his or her sentence. From a societal standpoint, therefore, the key question for incapacitation is the following: On average, how many crimes are prevented in society by incarcerating an individual for a given period of time? If the criminal justice system is operating in an efficient manner, then those who are incarcerated will be the most serious and prolific offenders in the population, and the incapacitation effect of prison will be substantial.

After a period of over 20 years of empirical stagnation, questions concerning the incapacitation effect of prison have resurfaced, as indicated by a 2007 issue of *Journal of Quantitative Criminology* especially devoted to empirical research on the topic. This recent attention has been spurred, in part, by the unparalleled growth in the use of incarceration in Western society over the last three decades. This growth is most stark in the United States, which presently has an incarceration rate (including jails and state and federal prisons) over 750 per 100,000 residents. Yet questions concerning the incapacitation effect of prison are not the sole province of American criminology, as prison growth has been a more widespread Western phenomenon. In Europe as well, two-thirds of 35 countries recently surveyed experienced growth in their incarceration rates during recent decades.

The goal for a study of incapacitation is to estimate the number of crimes that an incarcerated offender *would have committed*, were he free in the community rather than confined in prison. This is taken as an estimate of the "incapacitation effect," defined as the number of crimes averted by physically isolating an offender from society at large. This task would seem straightforward at first glance, but the problem on closer inspection is that this quantity is a "counterfactual" which can never, even in principle, be directly observed. Instead, the counterfactual outcome – how many crimes the offender would have committed if free – is unobserved and must be inferred from indirect sources. The challenge to providing a valid incapacitation effect on criminal offending, therefore, is one of overcoming this counterfactual problem.

In this entry, an overview of the first generation of incapacitation studies is first provided. These studies employ a within-individual design to solve the counterfactual problem. This is followed by a description of the most recent round of incapacitation research, with a focus on studies that employ a between-individual solution to the counterfactual problem. Details are provided on one specific between-individual approach that entails propensity score matching. This entry closes with a discussion of the ongoing empirical challenges to estimation of the incapacitation effect of prison.

Traditional Approaches to Estimating Incapacitation Effects

Spelman (2000) acknowledges two general kinds of incapacitation research. "Top-down" approaches use aggregate data to estimate the impact of growth in the prison population or in prison commitments on crime rates. Top-down estimates are often preferred by policy analysts for their ability to estimate the total effect of imprisonment on crime. Such estimates are, by

This chapter is largely based on an article that is published elsewhere (Wermink et al. 2012). Some parts are incorporated integral.

necessity, agnostic about the specific mechanism by which prisons influence behavior – specifically, through incapacitation, deterrence, or rehabilitation. Indeed, by treating the crime prevention mechanism as a black box, top-down studies are unable to discriminate between competing explanations for prison effectiveness. "Bottom-up" approaches, on the other hand, attempt to peer into the black box of prison effectiveness by using individual-level data to estimate incapacitation effects. These studies typically address the incapacitation question by using arrestee or inmate samples to obtain self-report estimates of offending frequency in the months prior to criminal justice intervention, a quantity referred to in the criminal career literature as "lambda" (see Blumstein et al. 1986).

The remainder of this section focuses on "bottom-up" studies of prison effectiveness (Spelman 2000), or empirical attempts to estimate individual criminal offending rates which can then inform estimates of incapacitation effects. Estimation of the incapacitation effect of prison in such studies is fundamentally an actuarial exercise. The empirical challenge is one of overcoming the *counterfactual problem*, that is, providing a credible estimate of behavior that is not observed because an individual is locked up in jail or prison rather than free in the community. The key quantity for incapacitation is known as lambda, λ, representing the annual offending frequency conditional on active offending, which can be taken as an estimate of the number of crimes avoided through incarceration. Zimring and Hawkins (1995, p. 81) observe that there are two approaches to estimating lambda:

> [T]o determine the level of crime that would have occurred if a particular group had not been confined, one must either study the criminal activity of the same group at a different time in their lives to estimate what that group would have done if not confined, or one must study the behavior of persons other than those confined to approximate the crimes avoided by imprisonment in the past.

In the former approach, the goal is to estimate a *within-person counterfactual*, while in the latter approach, the goal is to estimate a *between-person counterfactual*. The most influential incapacitation studies are universally of the first kind in Zimring and Hawkins' (1995) typology. These studies provide within-person counterfactual offending rates from self-report surveys of arrestees or prison inmates. With this approach, targeted individuals are questioned about their criminal activity during the months leading up to their arrest or confinement. These estimates are taken as the number of crimes they committed on an annual basis when they were free (prior to incarceration), and by implication, the number of crimes they would have committed per year during the time that they were incarcerated. Therefore, the counterfactual offending rate for incarcerated individuals is their own offending rate in the months preceding their confinement.

The best known source of a within-person counterfactual is the Rand Corporation's second inmate survey, a study of over 2,000 male inmates in California, Michigan, and Texas. The report by Chaiken and Chaiken (1982), for example, revealed that annual offense frequencies for ten different crimes were highly skewed, with active offenders at the median committing 15 offenses and offenders at the 90th percentile committing 605 offenses (these figures exclude drug offending). They estimated the mean offending rate to be either 187 or 278 crimes per year, depending on how ambiguous survey responses were treated. Based on these figures, incarceration was shown to have the capacity to substantially incapacitate criminal behavior. In fact, on the basis of this research, policy-oriented criminologists began to advocate "selective incapacitation" of high-rate criminal offenders as an explicit penal policy. In hindsight, however, research efforts using risk assessment tools to identify high-rate offenders and target them for longer prison sentences have been notoriously unsuccessful, calling into question the wisdom of a selective incapacitation policy.

These early bottom-up, within-person incapacitation studies had limitations, however. First and foremost is the question of the reliability of offender self-reports of their criminal behavior (Spelman 1994; Zimring and Hawkins 1995). It is unclear whether incarcerated offenders, particularly high-rate offenders, can accurately recall their prior criminal activity. Even if they can, they may not be motivated to report it honestly to interviewers. Spelman (1994), in fact, found evidence of both overreporting and underreporting in the second Rand inmate survey, but by individuals at different locations in the distribution of offending rates. Second, the presence and nature of "crime spurts" can introduce serious distortions in estimates of lambda (Blumstein et al. 1986) and, therefore, in estimates of incapacitation effects. Research has shown that offenders experience relatively short periods of high-rate offending immediately prior to incarceration, meaning that incapacitation effects will be severely overstated, especially if reporting windows are comparatively narrow. Moreover, a portion of the pre-incarceration crime spurt appears to be artifactual rather than behavioral, that is, a function of the way that individuals are filtered through the criminal justice system and selected for custodial sentences rather than reflective of genuine growth in offending frequency (Maltz and Pollock 1980).

Third, the early studies were possibly guilty of overly optimistic estimates of incapacitation effects. The Rand Corporation studies showed self-report offending rates to vary widely from one jurisdiction to another. For example, Chaiken and Chaiken (1982) reported the median total nondrug crime frequency among prisoners (not including jail inmates) in California, Michigan, and Texas to be 42, 17, and 9, respectively. Offending rates in other jurisdictions that attempted to replicate the Rand survey were comparably varied. Relatedly, the incapacitation estimates from this research tend to assume constant offending during the reference period, whereas Horney and Marshall (1991) demonstrated that most offenders were actively involved in crime only intermittently, and that failure to account for this intermittency would inflate estimates of the incapacitation effects of prison when offense rates are annualized.

Finally, existing bottom-up studies have employed samples of incarcerated offenders with questionable generalizability to contemporary circumstances (see Zimring and Hawkins 1995, for a similar critique). Prior studies base incapacitation estimates on inmates who were incarcerated near the beginning of an unprecedented expansion in US prisons. From the 1930s through the early 1970s, the incarceration rate hovered around 110 per 100,000 residents, then began a steady increase in the early 1970s and at present is in excess of 750 per 100,000. If the criminal justice system operates efficiently, officials will identify and incarcerate the most serious offenders – those who commit the most serious crimes and who do so at a high rate. Holding all else equal, however, rapid expansion in the prison population will also result in less active or less serious offenders entering prison. Contemporary incarceration thus might yield lower incapacitation effects than incarceration in earlier decades because of diminishing marginal returns as the criminal justice system reaches deeper into the offender queue. Therefore, incapacitation studies conducted in the 1980s and earlier may have limited utility for policy makers today, and may in fact overestimate the current incapacitation effect of prison.

Contemporary Approaches to Estimating Incapacitation Effects

Until recently, within-person counterfactual studies exhausted all empirical research in the incapacitation tradition. Two studies in the last 3 years have adopted the second approach in Zimring and Hawkins' (1995, p. 81) typology, that of estimating the offending rate of "persons other than those confined to approximate the crimes avoided by imprisonment." The studies by Owens (2009) and Sweeten and Apel (2007) represent only two of a new generation of incapacitation research. One strand of this research employs simulation techniques (Bhati 2007; Blokland and Nieuwbeerta 2007), while another strand takes advantage of a variety of aggregate "natural experiments" (Barbarino and Mastrobuoni 2008; Johnson and Raphael 2009; Kessler and Levitt 1999; Ramirez and Crano 2003).

Owens (2009) takes advantage of a change in the sentencing guidelines in Maryland that lowered the age at which an offender's juvenile record can be used to add "criminal history points" for the purpose of criminal sentencing. This policy change effectively reduced the recommended sentence faced by certain groups of young adults with a juvenile criminal history. In her study, individuals who were sentenced after the policy change accumulated 2.8 arrests per year (with an implied Index offending rate of 1.5 offenses per year) during the time that they would still have been confined under the old sentencing regime.

Sweeten and Apel (2007) rely on propensity score matching to select incarcerated and non-incarcerated individuals who closely resemble each other on a wide variety of background variables, including criminal history. Using data from a contemporary, nationally representative sample of American youth, they limit their attention only to individuals who were not incarcerated prior to the reference conviction. They estimate lambda among the non-incarcerated sample to be about 9 offenses per year among 16–17-year-olds (95 % confidence interval = 6.2, 14.1), and about 6 offenses per year among 18–19-year-olds (95 % confidence interval = 4.9, 8.4). In the next section, the logic and method of matched sampling, as used by Sweeten and Apel (2007), is described in more detail.

Details on Matched Sampling

Matched sampling is a quasi-experimental method of estimating the number of crimes averted through incarceration. The study by Sweeten and Apel (2007) is the only application of this approach thus far. Specifically, the offending rate of a "non-captive" but high-risk comparison sample is taken as the counterfactual offending rate for an incarcerated sample during the period of its confinement. In other words, the incapacitation effect is the offending rate of a contemporaneous sample of individuals who are not incarcerated.

This particular approach has appeal because it addresses the counterfactual problem in a straightforward and transparent way. The intuition underlying propensity score matching can be elaborated by drawing comparisons with an experimental approach. In a hypothetical experiment, individuals would be targeted for incarceration at random. This ensures that the experimental and control groups are *balanced*, meaning that they are indistinguishable (in expectation) on all possible sources of nonequivalence which might confound incapacitation estimates. By virtue of randomization, the control group serves as a *counterfactual* for the experimental group, or an estimate of what the experimental group's behavior would have been had it not been imprisoned. Randomization ensures that no additional adjustments are required, because all sampled individuals were "at risk" of being incarcerated with known probability; that is, at the start of the experiment, all were equally eligible to be assigned to the experimental group. In this scenario, the *assignment mechanism* is controlled by a statistically random process that renders it ignorable. The incapacitation effect of imprisonment then is simply the mean (or median) offending rate of the entire control group during the experimental group's confinement.

Propensity score matching is a quasi-experimental approach that attempts to approximate the conditions of a randomized experiment by creating "synthetic" experimental and control groups that are balanced on a wide variety of confounding variables. The technique is particularly advantageous in situations where randomization is not possible, but the assignment mechanism is nevertheless well understood and can be explicitly modeled from variables that are readily available to the analyst. A well-specified model increases confidence that the assignment mechanism is ignorable conditional on the propensity score, a situation that satisfies what is known as the *conditional independence assumption*.

The initial step of this approach is estimation of the propensity score, which in the present context can be defined as the predicted probability of incarceration conditional on criminal conviction. To meet the conditional independence assumption requires intimate knowledge of the incarceration process. For instance, research on criminal sentencing suggests that the key determinants of incarceration are the characteristics of the instant offense the offender's criminal history, and, to a much lesser degree, the offender's personal characteristics and life circumstances, characteristics of the sentencing judge, and jurisdictional factors.

Once the incarceration process has been adequately modeled and the propensity score estimated, attention turns to the matching process. The goal at this stage is to identify, for each incarcerated individual, a non-incarcerated counterpart who is observationally equivalent. A "suitable" match is therefore defined as a non-incarcerated individual who has a propensity score which is identical (or if not identical, at least within a minimum distance known as a caliper) to that of his or her referent, incarcerated peer. A property of propensity score matching is that not all incarcerated subjects will necessarily have a suitable match, and conversely, not all non-incarcerated subjects will necessarily be matched. A feature of the *common support condition* is that only those individuals who most closely resemble one another on the propensity score – and by implication, the variables indexed by the propensity score – are actually matched and therefore contribute information to the estimate of interest. With matched samples in hand, the final step is to estimate the counterfactual of interest.

Ongoing Challenges to Estimating Incapacitation Effects

Existing studies examine the incapacitation effect of prison incarceration, with virtually no attention devoted to sentences of incarceration less than 1 year. (A recent exception is Sweeten and Apel 2007, although they do not distinguish prison from jail incarceration in their incapacitation estimates.) This is a glaring omission, as jail incarceration in the United States is far more widespread than prison incarceration. The jail population at any given time tends to be about one-half the size of the prison population. However, the average daily jail population of about 750,000 individuals, from which the jail incarceration rate is calculated, underestimates by a very large margin the number of individuals who actually pass through the nation's jails in a year, which is estimated to be a shade under 13 million. Taking account of the fact that just over 60 % of these individuals are awaiting trial still yields about five million who are incarcerated in jail for a crime. So while the prison incarceration rate of about 500 per 100,000 provides a fairly accurate estimate of the number of people who spend time in prison during a year, the true jail incarceration rate is over 1,500 per 100,000. Short sentences of incarceration are clearly the norm, yet virtually nothing is known about their incapacitation potential.

Incapacitation effects are also likely to be highly sensitive to the inclusion of drug offenses, which presents two empirical problems – frequency and replacement. First of all, drug crimes are less serious violations that occur at very high frequency, far higher than other types of crime. For example, Chaiken and Chaiken (1982) report from the Rand Corporation's second inmate survey that among active drug dealers, the median offender sold drugs 100 times per year, while the offender at the 90th percentile did so on 3,251 occasions. The mean was between 880 and 1,299 crimes per year, depending on how ambiguous responses were handled. When drug offenses were excluded from the incapacitation effect, the average offender reported committing between 187 and 278 crimes per year (median = 15), but when drug offenses were included, the average offender committed between 614 and 933 crimes (median = 42). Thus, in samples which include large proportional representation of drug offenders, incapacitation estimates that include drug offending will be highly distorted.

A second problem with the inclusion of drug offenses in incapacitation estimates is that they are likely to be subject to much higher replacement rates than other kinds of crime. For instance, the removal of an active drug dealer from one street corner presents an opportunity for another drug dealer to take his or her place.

If replacement in this scenario is 1-for-1, and the replacement drug dealer sells drugs at the same rate as the newly incarcerated drug dealer, then the incapacitation effect is effectively 0 from a societal standpoint. In reality, the replacement rates for specific crimes, as well as the offending rates among replacement offenders, are unknown. Nevertheless, the very nature of drug offending – drug transactions can be conducted briefly and at high volume, as well as a replacement rate that is likely to be high – renders it highly problematic for use in generating estimates of the incapacitation effect.

Incapacitation effects are likely to be sensitive to co-offending patterns. Consider two offenders who prefer to only commit crime in each other's company, which means that their offending rates are equal. Incarceration of a referent offender in this co-offending dyad effectively removes the opportunity of both offenders to commit crime. In a manner of speaking, both offenders have been incapacitated, even though only the referent offender is imprisoned. On the other hand, if the co-offending partner continues committing crime at a rate which is unchanged, then the same number of crimes which would have been committed if the referent offender was not incarcerated will still continue unabated. In this latter scenario, the incapacitation effect is 0 from a societal standpoint (similar to the 100 % replacement rate discussed in the drug offending example above). The implication of co-offending networks for the magnitude of incapacitation estimates would thus be a very fruitful line of inquiry.

Conclusion

A growing literature has become attentive to the societal consequences of prevailing imprisonment policies. Increasingly, this literature cautions that the social costs of prison growth might outweigh the incapacitation benefits. One prominent line of research observes that confinement worsens an individual's capacity to adopt a law-abiding lifestyle upon returning to the community. Such "collateral consequences" of incarceration include labor market stigma, slowed wage growth, and marital disruption, among other consequences. These have the potential to exacerbate long-term criminal offending. Thus, while prison sentences will temporarily incapacitate criminal behavior, they might worsen offending over the long run relative to noncustodial sentences. It will thus be important to provide a full accounting of the impact of prisons on criminal behavior – both for better and for worse.

It will also be important to put incapacitation effects into historical context. Prison growth leads to the confinement of individuals who pose steadily lower risk to society, on the margin and all else equal. Consistent with the "law of diminishing returns," recent studies of prison expansion show substantial erosion in crime control over the last three decades as the scale of incarceration has grown (Johnson and Raphael 2009). To the degree that the marginal return with respect to crime does indeed shrink with unmitigated growth, incapacitation is likely to play an increasingly minor role in policy discussions concerning the use of prison.

Related Entries

- ▶ Crime Preventive Effects of Incapacitation
- ▶ Incapacitation
- ▶ Penal Philosophy and Sentencing Theory
- ▶ Theories of Punishment

Recommended Reading and References

Barbarino A, Mastrobuoni G (2008) The incapacitative effect of incarceration: Evidence from several Italian collective pardons. Unpublished manuscript. Research and Statistics Division, Federal Reserve Board, Washington, DC

Bhati AS (2007) Estimating the number of crimes averted by incapacitation: an information theoretic approach. J Quant Criminol 23:35–375

Blokland AAJ, Nieuwbeerta P (2007) Selectively incapacitating frequent offenders: costs and benefits of various penal scenarios. J Quant Criminol 23:327–353

Blumstein A, Cohen J, Roth JA, Visher CA (eds) (1986) Criminal careers and "career criminals", vol 1. National Academy Press, Washington, DC

Chaiken JM, Chaiken MR (1982) Varieties of criminal behavior. Report No. R-2814-NIJ. Rand, Santa Monica

Horney J, Marshall IH (1991) Measuring lambda through self-reports. Criminology 29:471–495

Johnson R, Raphael S (2009) How much crime reduction does the marginal prisoner buy? Unpublished manuscript. Goldman School of Public Policy, University of California, Berkeley

Kessler D, Levitt SD (1999) Using sentence enhancements to distinguish between deterrence and incapacitation. J Law Econ 42:343–363

Maltz MD, Pollock SM (1980) Artificial inflation of a delinquency rate by a selection artifact. Oper Res 28:547–559

Owens EG (2009) More time, less crime? Estimating the incapacitative effect of sentence enhancements. J Law Econ 52:551–579

Ramirez JR, Crano WD (2003) Deterrence and incapacitation: an interrupted time-series analysis of California's three-strikes law. J Appl Soc Psychol 33:110–144

Spelman W (1994) Criminal incapacitation. Plenum, New York

Spelman W (2000) What recent studies do (and don't) tell us about imprisonment and crime. In: Tonry M (ed) Crime and justice: a review of research, vol 27. University of Chicago Press, Chicago, pp 419–494

Sweeten G, Apel R (2007) Incapacitation: revisiting an old question with a new method and new data. J Quant Criminol 23:303–326

Wermink H, Apel R, Nieuwbeerta P, Blokland AAJ (2012) The incapacitation effect of first-time imprisonment: a matched samples comparison. J Quant Criminol DOI 10.1007/s10940-012-9189-3.

Zimring FE, Hawkins G (1995) Incapacitation: penal confinement and the restraint of crime. Oxford University Press, New York

Estimator Variables and Eyewitness Identification

Jennifer L. Beaudry[1], Christina L. Bullard[2] and Jennifer R. Dolin[2]
[1]Swinburne University of Technology, Melbourne, Australia
[2]University of South Carolina Beaufort, Bluffton, SC, USA

Synonyms

Event characteristics; Perpetrator characteristics; Witness characteristics

Overview

Eyewitness identifications can be either an asset or a detriment to a criminal investigation. Thus, it is vital that the individuals who are involved in the criminal justice system (e.g., police officers, attorneys, judges, and jurors) are aware of the myriad variables that affect eyewitness identification accuracy. These variables can be categorized as either system or estimator variables. The police have control over system variables, such as lineup presentation, biased instructions, post-identification feedback, lineup fairness, and wording of questions. In contrast, police have no control over estimator variables, which can, nonetheless, influence identification accuracy. This entry will examine estimator variables associated with the criminal event (presence of a weapon, crime seriousness, stress, exposure duration, and distance), perpetrator (race and disguise), and witness (age and psychological impairment). The impact of these variables on eyewitness memory and identification decisions should be considered when evaluating the reliability of eyewitness evidence.

Introduction

Gary Wells (1978) revolutionized researchers' approach to eyewitness identification issues by creating a distinction between system and estimator variable research. Both categories of variables can influence eyewitness identification accuracy; however, he distinguished between variables that are under the control of the criminal justice system (system variables) and those that are not (estimator variables). System variables include how the lineup is constructed, how the lineup is presented to the witness, and the instructions that accompany the lineup. This type of research informs recommendations regarding the current best practices for obtaining eyewitness identification evidence. Estimator variables include factors related to the criminal event, witness, and perpetrator, such as how long the perpetrator was in view, the age of the witness, and whether or not a weapon was present. Wells labeled these estimator variables because their impact in an actual case can only be estimated post hoc.

The courts generally rely on five factors to evaluate the reliability of eyewitness identification evidence (*Neil v. Biggers*, 1972; *Manson v. Braithwaite*, 1977): (a) opportunity to view the criminal during the event, (b) degree of attention, (c) accuracy of prior description of the criminal, (d) level of certainty demonstrated at confrontation/identification, and (e) time between crime and identification. The reliance on these limited variables is problematic, especially given the robust finding that providing confirming feedback to a witness (e.g., "Good, you got the guy!") significantly bolsters reports of the opportunity to view the criminal, degree of attention, and certainty in the identification decision (Wells and Bradfield 1998). Thus, relying on the Biggers/Manson criteria alone may lead triers of fact to make erroneous conclusions regarding the reliability of the evidence.

In a revolutionary decision, the New Jersey Supreme Court in *State v. Henderson* (2011) ruled that in order to determine the suggestiveness of an identification procedure, the court must consider all factors that may have influenced the witness's decision (e.g., blind administration, lineup construction, feedback). If the evidence is still admissible, the court should then consider the Biggers/Manson criteria and any additional applicable estimator variables (e.g., stress, weapon presence, exposure duration, age).

System variable research improves police procedures; estimator variable research, on the other hand, contributes to the understanding of how factors at the time of the crime influence eyewitnesses. When evaluating the reliability of identification evidence, both system and estimator variables should be considered. For instance, an identification made by a 3-year-old who viewed the perpetrator for 15 s in a dark alley from a distance of 250 ft should not be given as much weight as an identification made by a sober young adult who was face-to-face with the perpetrator for 30 min in a well-lit room. One cannot definitively state that the former identification is wrong and the latter is correct; however, awareness of the impact of age, exposure duration, lighting, and distance on identification accuracy should lead one to deem the latter identification as likely to be more reliable than the former.

State of the Knowledge of the Field

This entry maintains Wells' (1978) organization of estimator variables with regard to the division of characteristics associated with the criminal event, perpetrator, and witness. This non-exhaustive discussion reviews the most commonly researched variables in each category. Specifically, the section entitled "Characteristics of the Criminal Event" examines encoding conditions, the presence of a weapon, crime seriousness, and stress; "Characteristics of the Perpetrator" focuses on race and disguise; and "Characteristics of the Witness" covers age and psychological impairment.

Characteristics of the Criminal Event

Encoding Conditions

The memory process comprises three steps: encoding, storage, and retrieval. Any factor that reduces the information available for encoding reduces performance at retrieval – information (e.g., details of the person's face) cannot be stored nor later retrieved (e.g., used to distinguish between members of a lineup) if it was never encoded. This section addresses two of the conditions that restrict the availability of information at encoding: exposure duration and distance. Readers are directed to Narby, Cutler, and Penrod (1996) and Shapiro and Penrod (1986) for a review of additional encoding-related variables, such as perceptual salience, knowledge of recognition task, and number of faces presented during exposure.

Exposure Duration

Exposure duration refers to how long the witness is exposed to the target's face. In general, studies have found that as exposure time increases, identification accuracy increases. As discussed in Bornstein, Deffenbacher, Penrod, and McGorty's (2012) meta-analysis, shorter durations in research range from a few tenths of a second to 12 s, and longer durations typically hover around 45–60 s (although a few studies examined durations longer than 180 s). This updated meta-analysis confirmed findings from Shapiro and Penrod's (1986) meta-analysis. Specifically, longer durations produce higher hit rates and lower false identification rates compared to shorter durations, but exposure has a greater impact on hit rates than false alarms. In comparing the effect sizes from both meta-analyses, Bornstein et al. concluded that exposure duration is even more influential than researchers originally thought.

In real crimes, exposure durations may be longer than those typically used in eyewitness and facial recognition studies; however, exposure duration has a nonlinear effect, meaning that additional exposure beyond a certain point (approximately 30 s) has minimal impact on accuracy. Further evidence of the generalizability of laboratory findings to the real world is that research paradigm (facial recognition vs. eyewitness identification) did not significantly moderate the relationship; exposure duration had similar effects regardless of whether or not the witness's memory was studied under more ecologically valid conditions (Bornstein et al. 2012).

Exposure duration does not operate in isolation, but instead interacts with, and is affected by, other variables. For instance, as discussed in the next section, the presence of a weapon (or unusual object) is most detrimental to identification accuracy at intermediate durations, with a smaller effect at shorter and longer durations (Fawcett et al. 2013). One system variable that affects witnesses' reports of exposure duration is whether or not the police provide confirming post-identification feedback to the witness ("Good, you got the guy!"). This feedback leads witnesses to overestimate how long the perpetrator was in view (e.g., Wells and Bradfield 1998). Thus, in real cases, police and attorneys must be cautious when attempting to determine length of exposure. When possible, police should gather this information before any feedback is provided and attempt to corroborate the witness's report through other means.

Distance

When considering the distance between the eyewitness and the perpetrator of a crime, two factors need to be taken into account: (a) the actual distance, which affects identification accuracy, and (b) the witness's perceived distance, which impacts people's perceptions of that evidence. In a thorough investigation, Lindsay, Semmler, Weber, Brewer, and Lindsay (2008) examined the relationship between actual and perceived distance and how both influence identification accuracy. They found a relatively high level of correspondence between perceived and actual distances, but this relationship deteriorated significantly after an ecologically valid delay of one day, at which point witnesses underestimated the actual distance. Furthermore, participants' tendency to underestimate distance varied according to actual distance, with the largest underestimates at 36–41 m (~39–44 yards). Neither perceived nor actual distance influenced witnesses' choosing rates, but longer distances did reduce identification accuracy. The authors hesitated to state a distance beyond which an identification should be inadmissible in court, but instead argued that distance should be used to evaluate the reliability of the evidence.

Presence of a Weapon

Exposure to a perpetrator brandishing a weapon significantly impairs eyewitness identification accuracy because rather than focusing on the perpetrator, the witness' attention is drawn to the weapon itself – this phenomenon is known as the weapon focus effect. Original research in this area sought to understand how the presence of a weapon influenced witnesses' cognitive processes, which ultimately led to the development of the arousal/threat hypothesis. Easterbrook's (1959) cue-utilization hypothesis contends that arousal and attention interact to affect the cues people utilize. In a state of high arousal, people's attention is limited, and they can only utilize cues that are the focus of their attention. Meanwhile, all non-central, peripheral cues are underutilized. The arousal/threat hypothesis applies cue-utilization to the eyewitness scenario: The weapon increases physiological arousal due to its threatening nature and induces fixation as the source of that arousal. The weapon, then, becomes a central cue and the perpetrator's face becomes a peripheral cue, leaving the witness with a stronger memory for the weapon than the perpetrator. Tests of the arousal/threat hypothesis include manipulating threat or arousal, or examining the amount of arousal experienced in the presence of the weapon; these studies have produced mixed results suggesting that arousal/threat is not solely responsible for the weapon focus effect (Fawcett et al. 2013).

In contrast, other researchers have posited that the weapon focus effect is caused by the presence of an unusual or unexpected item (e.g., Pickel 1998). Supporters of the unusual item hypothesis contend that a witness's attention is drawn to an object that is unusual in a given context. An isolated object, like a pair of scissors, would be unusual in a yoga class, but not at all unexpected in a hair salon. Similarly, a whole, uncooked chicken would not be unexpected in a butcher shop, but extremely unusual in a hair salon. Pickel sought to separately examine the role of threat and unusualness by contrasting the two components. Unusualness, but not threat, significantly influenced witnesses' memory for non-object details. Specifically, compared to the low-unusualness object, a high-unusualness object significantly impaired witnesses' memory for peripheral details. The unusual item hypothesis has received considerable empirical support. The shifting view of the weapon focus effect is substantiated by the definition used in a recent meta-analysis: "weapon focus shall be defined as an object-related decrease in memory performance. . .for those elements of an event or visual scene coinciding with the presence of a weapon or unusual object" (Fawcett et al. 2013, p. 2).

Although the investigation into the underlying theoretical explanation of the effect continues, the effect itself is not in question. Fawcett and colleagues' (2013) meta-analysis reported an overall moderate effect of weapon presence on memory performance, with retention interval and exposure duration moderating the effect. Weapon presence impairs memory for both event-related details and identification accuracy; however, effect sizes are greater for recall than recognition, potentially because of measurement sensitivity. Both the arousal/threat and unusual item hypotheses received partial support in the meta-analysis. The weapon focus effect was greater when the stimuli were threatening rather than non-threatening. Likewise, compared to non-unusual objects, unusual objects reduced memory performance.

Crime Seriousness

The seriousness of a crime is generally evaluated by the level of violence, the level of risk for those involved, and/or the value of property stolen or damaged. In one of the earliest investigations, Leippe, Wells, and Ostrom (1978) isolated crime seriousness from physiological arousal by specifically staging a low-arousal criminal event in which a confederate stole a paper bag ostensibly left by a previous participant. Crime seriousness was determined by the contents of the bag (expensive item = highseriousness). Knowledge of the bag's contents was also manipulated, such that half were aware of the contents of the bag before the crime was committed and half learned this information afterward. Eyewitness identification accuracy was significantly higher when the crime was serious than not, but only when participants had prior knowledge of the bag's contents (and thus, the seriousness of the theft). Leippe et al. suggested that prior knowledge led the witnesses to pay more attention to the event and potentially devote higher levels of cognitive processing necessary to encode the culprit's facial features. Without prior knowledge, witnesses may not attend to features necessary to later accurately identify the target. Interestingly, and surprisingly, research has shown similar identification accuracy rates regardless of whether the witness was the victim of the crime or an uninvolved bystander to the event (e.g., Hosch et al. 1984).

Stress

The early literature examining stress and eyewitness memory is rife with speculation regarding whether increased stress would improve memory (up to a certain point, as in an inverted-*U* curve) or impede encoding (due to a reduction in available cognitive resources). Recently, in a meta-analysis, Deffenbacher, Bornstein, Penrod, and McGorty (2004) concluded that increased stress negatively affects eyewitness recall and identification accuracy. In terms of recall memory, stress has a greater impact on interrogative recall (specific questions) than narrative (free) recall. In terms of identification accuracy, target presence was a significant moderator. Increased stress reduces the likelihood that the witness will correctly identify the perpetrator from a target-present lineup, but does not increase the likelihood that the witness will falsely identify an innocent suspect from a target-absent lineup. Witnesses who did not select the perpetrator from the target-present lineup instead incorrectly chose a known-innocent lineup member, thereby discrediting them as reliable witnesses. Thus, although stress does not increase the likelihood that an innocent suspect will be selected, it does decrease the likelihood that the actual perpetrator will be prosecuted.

The type of research paradigm also significantly moderated the relationship between stress and eyewitness memory. Two research paradigms are generally used to examine memory for a face: facial recognition and eyewitness identification studies. Researchers examining theoretical questions about memory processes commonly use facial recognition studies in which participants are presented with numerous faces to remember, and then later asked to indicate whether each face from a larger pool of faces is old (previously seen) or new. Eyewitness identification studies, on the other hand, are more ecologically valid and generally used to examine applied issues regarding factors that could affect real-world eyewitness identification performance; typically, the witness is presented with one "perpetrator" and then later asked to identify him from a lineup. With regard to stress, participants' performance in the facial recognition paradigm was minimally affected by increased stress, but it did negatively impact their eyewitness identification accuracy. In addition, age was a significant moderator between stress and recall; increased stress negatively affected adults' recall, but had minimal impact on children's recall. Age, however, did not moderate the relationship between stress and eyewitness identification accuracy – both children and adults' recognition memory was debilitated by stressful conditions (Deffenbacher et al. 2004).

Related to stress is the physiological arousal associated with physical exertion, a common occurrence in the day-to-day activities of police, soldiers, and other emergency respondents. Physical exertion, however, is not limited to certain occupations – victims or witnesses may need to either fight with, or flee from, perpetrators. Hope, Lewinski, Dixon, Blocksidge, and Gabbert (2012) used a simulation study with police officers to determine how physical exertion influenced participants' memory. Compared to the control group, officers who physically exerted themselves made fewer correct identifications and recalled fewer details from information encountered before and after the physical exertion. Physical exertion, however, did not impact participants' reports of the presence of weapons. According to Hope et al., physically exerted participants had to allocate their cognitive resources differently than the control group. During the scenario, the physically exerted officers likely prioritized monitoring the immediate environment for anything that could be harmful (supported by the finding that their memory for presence of weapons was unaffected), while other environmental information was a lower priority. Notably, physical exertion not only dampened memory for the information encountered after the physical exertion, but also had a retroactive impact on memory for pre-exertion information, potentially due to an interruption in memory consolidation.

Taken together, the evidence suggests that there is no "optimal" level of stress when it comes to eyewitness memory; rather as stress increases, eyewitness memory decreases. The debilitating effects of stress are predicted to be even greater in real-world criminal situations, which are generally considered unethical to recreate in the laboratory.

Characteristics of the Perpetrator

Race

The race of a witness or perpetrator – in isolation – does not determine eyewitness recall or identification accuracy; however, varied combinations of witness and perpetrator race can significantly influence witnesses' identification decisions (Narby et al. 1996). In particular, individuals are more accurate when identifying someone of their own race rather than someone of a different race. This pattern of results is referred to as the own-race bias (also known as the cross-race or other-race effect). The majority of the own-race bias research has utilized facial recognition paradigms rather than eyewitness identification tasks; however, the effect persists across both paradigms (Meissner and Brigham 2001).

Most investigations into the own-race bias (ORB) seek to examine the possible social-cognitive processes responsible for the effect. Thus far, researchers have generally considered four theoretical explanations for the own-race bias (see Meissner and Brigham 2001 for more detailed explanations and additional citations). First, physiognomic homogeneity may play a role in the ORB. Group members of a certain race may have less physiognomic variability (and thus, greater facial feature similarity), which may reduce the ability of members of one race to distinguish between other-race faces. This theoretical explanation has received little support. Second, individuals with prejudiced attitudes may be less likely than non-prejudiced individuals to even attempt to differentiate other-race faces from one another. Racial attitudes are unrelated to memory performance when response bias is taken into account; however, prejudice is relevant to the extent that people who are more prejudiced report less contact with people of other races. This leads to the third theoretical explanation: interracial contact. That is, the more an individual interacts with someone from another race (in terms of quality or quantity), the less likely they are to demonstrate the ORB. In the most current meta-analysis, Meissner and Brigham found that interracial contact is a significant (but small) mediator of the ORB, explaining 2 % of the variability across participants. Fourth, perceptual learning may contribute to the ORB such that individuals use appropriate facial cues to discriminate between own-race faces; however, the appropriateness of these cues may not generalize to other-race faces. Using this general framework, researchers are exploring several possibilities including training people to discriminate between other-race faces, examining differences in processing the faces as a whole rather than featurally, and investigating the importance of how the face is represented in one's memory. Despite the unresolved questions regarding the social-cognitive mechanisms underlying the own-race bias, there is no question that this is a robust phenomenon.

The implications of the own-race bias have not gone unnoticed by the courts. Evaluating the reliability of cross-race identification evidence is an especially challenging task because there are a number of unanswered questions about how the ORB interacts with other variables. There is, however, evidence that factors such as short encoding times and long retention intervals exacerbate the own-race bias (Meissner and Brigham 2001).

Disguise

Although one might expect that disguise would always impair identification accuracy, the evidence is more equivocal. A few inconsistent findings notwithstanding (e.g., O'Rourke et al. 1989), disguises are generally detrimental to accuracy in eyewitness identification and facial recognition studies. Disguise may impair accuracy through four distinct (or possibly intertwined) processes. First, disguise prevents the witness from seeing, and thus encoding, specific and undistorted facial features. Second, disguise may alter the witness's metacognitions regarding the difficulty of the identification task. Third, disguise affects encoding specificity; that is, the discrepancy between exposure (encoding a disguised face) and test (recognizing an undisguised face) may interfere with identification accuracy. Finally, disguise may influence a witness's allocation of attention to the dangerousness of the situation, the disguise itself, or to the criminality of the perpetrator, thereby leaving fewer cognitive resources available to encode the undisguised features. At this point, researchers are unclear about the specific mechanism through which disguise functions but some or all of these mechanisms likely work in combination.

Research regarding different types of disguise has typically focused on a hat, sunglasses, or a stocking; in general, all three types impair identification accuracy. Until recently, investigations have focused on the presence or absence of disguise rather than varying degrees of disguise. To address this limitation, Mansour and colleagues (2012) systematically examined eyewitness identification accuracy across four levels of disguise (ranging from no disguise to fully disguised). Their first experiment investigated the impact of a hat and/or sunglasses, whereas the second examined a stocking disguise. The general pattern across both experiments was that correct identifications decreased as disguise increased; this supports Brewer, Weber, and Semmler's (2005) assertion that disguise affects the amount of information available for encoding. Notably, covering the face with a stocking to just above the mouth was as effective as covering the entire face; this supports the contention that disruptions to global facial configurations also play a role. Moreover, eyewitnesses' willingness to choose from the lineup decreased at a slower rate than their accuracy, suggesting that witnesses were not sensitive to the quality of their memory (ideally, a poorer memory resulting from greater disguise should be associated with lower choosing rates). At present, further research is needed to tease apart the different mechanisms through which disguise is effective.

Characteristics of the Witness

Age

Eyewitness researchers use adults (typically, between 18 and 59 years of age) as the baseline against which to compare the identification accuracy of children and older adults. Age has little impact on correct identifications – when presented with a target-present lineup, all age groups are equally likely to choose the target. Age becomes a concern, however, when the target is absent from the lineup – children and older adults have a greater tendency to choose, therefore, compared to younger adults, they are more likely to choose an innocent lineup member. Of course, researchers can manipulate whether the target is present or absent, but the police do not know if their suspect is guilty or innocent. Thus, the age of the eyewitness is a critical estimator variable that should be considered by triers of fact.

A substantial amount of research has examined children's eyewitness identification decisions from different lineup procedures; for a review of simultaneous and sequential lineup procedures see Dysart and Lawson ("▶ Eyewitness Research"). When presented with a simultaneous lineup, children under the age of 12 produce correct identification rates comparable to adults, but significantly fewer correct rejections than adults. The sequential lineup tends to reduce choosing for adults; however, the same reduction is not found with children, particularly when presented with a target-absent lineup (Steblay et al. 2011). Recent evidence suggests that children's decreased identification accuracy can be further exacerbated by an additional (and common) variable – presentation of the lineup by an officer in uniform. In this research, children were nearly six times more likely to make a false identification from a target-absent simultaneous lineup when it was administered by a uniformed officer compared to when the officer wore casual clothes (Lowenstein et al. 2010).

Researchers are working to devise new procedures to reduce children's tendency to choose regardless of whether the lineup actually contains the perpetrator. A promising (and on-going) line of research involves the development and investigation of the elimination lineup (Pozzulo and Lindsay 1999). Unlike the simultaneous or sequential lineup, the elimination lineup involves two distinct types of judgments. In the first judgment, the eyewitness is presented with all of the lineup members and is directed to select the lineup member who is most similar to the culprit. In the second judgment, the selected lineup member is isolated from the others, and the eyewitness must decide whether or not the selected person is in fact the culprit. This two-decision process is thought to be effective because the first decision maintains correct identification rates and the second increases correct rejection rates. At this point, the elimination lineup is the most encouraging solution to increasing children's identification accuracy.

At the other end of the age spectrum are older adults – normally considered to be 60 years of age and older. Like children, compared to young adults, older adults generally produce comparable correct identification rates and higher false identification rates from simultaneous lineups. When used with older adults, the sequential lineup decreases target-absent choosing rates, but also produces significantly lower correct identification rates compared to young adults (Steblay et al. 2011). As with children, researchers are examining ways to reduce older adults' tendency to choose from target-absent lineups. One line of investigation utilized a preidentification questionnaire, which had previously increased young adults' correct rejection rates. Before showing participants the lineup, the administrator asked participants the following questions: "How clear a memory do you have for the face of the criminal? How confident are you that you will be able to select the criminal if you see a photograph of him in a lineup? How confident are you that you will realize that the guilty person is not in the lineup if you are shown a lineup with only innocent people in it?" (p. 159). Unfortunately, the benefits of these questions did not generalize to older adults (Memon and Gabbert 2003). Another attempt involved modifying unbiased lineup instructions in order to increase the saliency that the lineup may not actually contain the culprit (Wilcock et al. 2005). Although the enhanced instructions made the older adults more aware that the culprit may or may not be present, they did not affect correct rejection rates. Taken together, the research demonstrates that children and older adults share the propensity to make more false identifications, and thus age should be considered by triers of fact.

Psychological Impairment

Police, judges, lawyers, and jurors are likely to question the credibility and reliability of a witness with a psychological impairment, whether it is temporary (e.g., intoxication) or a long-term medical disorder (e.g., schizophrenia). Our discussion centers on the most studied impairment in the eyewitness area: alcohol. For a thorough consideration of how mental retardation, autism, schizophrenia, and post-traumatic stress disorder affect memory see Soraci et al. (2007).

Alcohol may be a common estimator variable for crimes that occur at bars, parties, or late at night. Like many of the other estimator variables, alcohol can affect a person's attention, thereby influencing the cues encoded during the crime and those relied upon during recall and recognition. Yuille and Tollestrup (1990) found that, compared to sober witnesses, intoxicated witnesses recalled less information immediately after the crime, as well as in a follow-up interview 1 week later. While both groups had similar correct identification rates, intoxicated witnesses made more false identifications than sober witnesses. A similar pattern was found in a field study using showups (when the suspect is the only individual presented to the witness). Correct identification rates were similar regardless of blood alcohol level; in contrast, witnesses with higher blood alcohol levels were more likely to choose from a target-absent showup (Dysart et al. 2002). The authors attributed these results to the intoxicated witness's use of salient cues (e.g., hairstyle, clothing). These cues are appropriate when the suspect *is* the perpetrator; however, when the suspect is innocent, an intoxicated witness – instead of correctly rejecting the lineup – is more likely to choose an innocent lineup member if he/she is a close enough match to the witness' flawed memory.

Controversies in the Courtroom

Estimator variable research not only furthers people's understanding of the factors that influence witnesses' memory, but it also guides triers of fact who evaluate the reliability of eyewitness evidence. An examination of the opinions and beliefs of key players in the courtroom (i.e., judges, lawyers, and potential jurors) highlights the importance of this research. Take the weapon-focus effect as an example: 97 % of researchers surveyed indicated that the presence of a weapon can impair an eyewitness's ability to accurately identify the perpetrator's face (Kassin et al. 2001); in comparison, 68 % of judges (Magnussen et al. 2010), 50 % of prosecutors, 88 % of defense attorneys (Wise et al. 2009), and 53 % of jurors (Desmarais and Read 2011) agreed with this statement. Interested readers should refer to the aforementioned publications to better understand the discrepancies between empirical evidence and courtroom opinions.

Open Questions

Countless unanswered questions exist on the topic of estimator variables despite decades of eyewitness identification and facial recognition research. A few questions and controversies were discussed throughout this entry, most of which focused on theoretical explanations for various effects (e.g., own-race bias, weapon focus effect, disguise). Now, a few additional questions are highlighted.

Does the weapon focus effect occur in the real world? Neither archival analysis of actual crimes nor field studies have reported a significant effect of weapon presence on recall or recognition (see Fawcett et al. 2013 for a discussion). At least three differences between laboratory research and real-world conditions are relevant to discrepancy: (a) the lack of experimental control, (b) actual exposure conditions related to factors such as exposure duration and retention interval, and (c) exclusion of poor witnesses. The lack of experimental control is problematic because the archival researcher is unaware of what actually occurred (known as "ground truth"), and this could impact decisions regarding accuracy. For instance, it is flawed to evaluate description accuracy by examining the appearance of the suspect because the police may have picked up that suspect precisely because he fit the description provided by the witness. Moreover, time frames in laboratory studies do not adequately capture the real-world conditions in which a witness may be exposed to the weapon for several minutes or more, and may not have the opportunity to provide details to an officer until hours have passed, and worse, not be able to attempt an identification for days or weeks while a suspect is located (Fawcett et al. 2013). Finally, in real cases, witnesses who experience the weapon focus effect may realize that they did not get a good view of the perpetrator and may inform the police that they would not be able to identify him in a future lineup. Thus, witnesses with the poorest memory are likely never considered in the archival analysis. Further research is clearly needed to understand the role of weapon presence in the real world, and to resolve the current schism between laboratory and field research.

Questions about witness age remain unanswered: Why does age moderate the relationship between stress and recall, but not identification accuracy? Are similar cognitive processes responsible for increased choosing by children and older adults? If so, then perhaps the success of the elimination lineup with children will generalize to older adult eyewitnesses. With respect to crime seriousness, how does it interact with increased arousal? The limited research on crime seriousness has only examined low-arousal situations, but in the real world, a serious crime is likely to be associated with increased arousal. This area of research will be aided by the development of novel (and ethical) ways to investigate crime seriousness under high-arousal conditions. Finally, future research should investigate how alcohol and other drugs interact with system variables to influence the reliability of witness testimony.

Conclusions

Estimator variable research examines how the presence or absence of various factors – separately – influences identification accuracy. Minimal research, however, has examined the interactive effects of multiple estimator variables (a few notable exceptions: alcohol and cross-race identifications, exposure duration and weapon focus, and stress and age). Given the large number of variables that are likely to exist in any given criminal event, an examination of all possible main effects and interactions results in an unfathomable research design (Wells 1978). At present, the utility of estimator variable research is two-fold: (1) to further understanding of the factors that influence eyewitness identification accuracy, which will in turn help triers of fact evaluate the reliability of that evidence, and (2) to inform system variable research (e.g., although police cannot control the age of the eyewitness, they can reduce a child's tendency to choose from a lineup by altering *how* that lineup is presented).

Wells' (1978) distinction between system and estimator variables not only revolutionized how researchers and courts alike thought of the role of different variables, but it also (perhaps inadvertently) directed the field away from estimator variables and toward system variable research. This shift in the field led to major developments regarding best-practice identification procedures and provides the empirical support for widespread recommendations regarding the collection of eyewitness evidence. Perhaps the time has come for another shift – one toward a more balanced investigation of estimator and system variables. After all, estimator variables are always present and their impact is unlikely to be entirely eliminated even by best-practice identification procedures.

Finally, in this entry, and in the field in general, more attention is being paid to theory-driven estimator variable research. Greater reliance on theoretical frameworks is encouraged because it allows for integration of the extant body of work while informing and guiding future lines of research.

Related Entries

- ▶ Causes of Wrongful Convictions
- ▶ Eyewitness Evidence: Historical Background
- ▶ Eyewitness Identification Evidence
- ▶ Eyewitness Research
- ▶ Innocence Commissions
- ▶ Interviewing Eyewitnesses
- ▶ Jury Decision Making and Eyewitness Testimony
- ▶ Measuring Wrongful Convictions

Recommended Reading and References

Bornstein BH, Deffenbacher KA, Penrod SD, McGorty EK (2012) Effects of exposure time and cognitive operations on facial identification accuracy: a meta-analysis of two variables associated with initial memory strength. Psychol Crime Law 18:473–490

Brewer N, Weber N, Semmler C (2005) Eyewitness identification. In: Brewer J, Williams KD (eds) Psychology and law: an empirical perspective. The Guilford Press, New York, pp 177–221

Deffenbacher KA, Bornstein BH, Penrod SD, McGorty EK (2004) A meta-analytic review of the effects of high stress on eyewitness memory. Law Hum Behav 28:687–706

Desmarais SL, Read J (2011) After 30 years, what do we know about what jurors know? A meta-analytic review of lay knowledge regarding eyewitness factors. Law Hum Behav 35:200–210

Dysart JE, Lindsay RCL, MacDonald TK, Wicke C (2002) The intoxicated witness: effects of alcohol on identification accuracy from showups. J Appl Psychol 87:170–175

Easterbrook JA (1959) The effect of emotion on the utilization and organization of behavior. Psychol Rev 66:183–201

Fawcett JM, Russell EJ, Peace KA, Christie J (2013) Of guns and geese: a meta-analytic review of the 'weapon focus' literature. Psychol Crime Law 19:35–66

Hope L, Lewinski W, Dixon J, Blocksidge D, Gabbert F (2012) Witness in action: the effect of physical exertion on recall and recognition. Psychol Sci 23:386–390

Hosch HM, Leippe MR, Marchioni PM, Cooper DS (1984) Victimization, self-monitoring, and eyewitness identification. J Appl Psychol 69:280–288

Kassin SM, Tubb V, Hosch HM, Memon A (2001) On the "general acceptance" of eyewitness testimony research: a new survey of the experts. Am Psychol 56:405–416

Leippe MR, Wells GL, Ostrom TM (1978) Crime seriousness as a determinant of accuracy in eyewitness identification. J Appl Psychol 63:345–351

Lindsay RCL, Semmler C, Weber N, Brewer N, Lindsay MR (2008) How variations in distance affect eyewitness reports and identification accuracy. Law Hum Behav 32:526–535

Lowenstein JA, Blank H, Sauer JD (2010) Uniforms affect the accuracy of children's eyewitness identification decisions. J Investig Psychol Offender Profiling 7:59–73

Magnussen S, Melinder A, Stridbeck U, Raja AQ (2010) Beliefs about factors affecting the reliability of eyewitness testimony: a comparison of judges, jurors and the general public. Applied Cognitive Psychology 24:122–133

Mansour JK, Beaudry JL, Bertrand MI, Kalmet N, Melsom EI, Lindsay RCL (2012) Impact of disguise on identification decisions and confidence with simultaneous and sequential lineups. Law Hum Behav 36:513–526

Meissner CA, Brigham JC (2001) Thirty years of investigating the own-race bias in memory for faces: a meta-analytic review. Psychol Publ Policy Law 7:3–35

Memon A, Gabbert F (2003) Improving the identification accuracy of senior witnesses: Do prelineup questions and sequential testing help? J Appl Psychol 88:341–347

Narby DJ, Cutler BL, Penrod SD (1996) The effects of witness, target, and situational factors on eyewitness identifications. In: Sporer S, Malpass RS, Koehnken G (eds) Psychological issues in eyewitness identification. Lawrence Erlbaum Associates, Hillsdale, pp 23–52

O'Rourke TE, Penrod SD, Cutler BL, Stuve TE (1989) The external validity of eyewitness identification research: generalizing across subject populations. Law Hum Behav 13:385–395

Pickel KL (1998) Unusualness and threat as possible causes of "weapon focus". Memory 6:277–295

Pozzulo JD, Lindsay RCL (1999) Elimination lineups: an improved identification procedure for child eyewitnesses. J Appl Psychol 84:167–176

Shapiro PN, Penrod SD (1986) Meta-analysis of facial identification studies. Psychol Bull 100:139–156

Soraci SA, Carlin MT, Read JD, Pogoda TK, Wakeford Y, Cavanaugh S, Shin L (2007) Psychological impairment, eyewitness testimony, and false memories: Individual differences. In: Lindsay R, Ross D, Read D, Toglia M (eds) Handbook of eyewitness psychology: memory for events. Lawrence Erlbaum and Associates, Mahwah, pp 261–297

Steblay NK, Dysart JE, Wells GL (2011) Seventy-two tests of the sequential lineup superiority effect: a meta-analysis and policy discussion. Psychol Publ Policy Law 17:99–139

Wells GL (1978) Applied eyewitness-testimony research: system variables and estimator variables. J Pers Soc Psychol 36:1546–1557

Wells GL, Bradfield AL (1998) "Good, you identified the suspect": feedback to eyewitnesses distorts their reports of the witnessing experience. J Appl Psychol 83:360–376

Wilcock RA, Bull R, Vrij A (2005) Aiding the performance of older eyewitnesses: enhanced non-biased line-up instructions and line-up presentation. Psychiat Psychol Law 12:129–140

Wise RA, Pawlenko NB, Safer MA, Meyer D (2009) What US prosecutors and defence attorneys know and believe about eyewitness testimony. Appl Cognit Psychol 23:1266–1281

Yuille JC, Tollstrup PA (1990) Some effects of alcohol on eyewitness memory. J Appl Psychol 75:268–273

Ethnic Violence

Conrad Schetter[1] and Andreas Wimmer[2]
[1]Bonn International Center for Conversion (BICC), Bonn, Germany
[2]Department of Sociology, Princeton University, Princeton, NJ, USA

Overview

Public opinion finds ethnic violence a particular illegitimate form of violence. "Ethnic cleansing" in the form of mass shootings, systematic rape, or massacres on a genocidal scale is considered the quintessence of terror and a prime symbol of the "new world disorder." The systematic use of violence against civilian population is particularly repulsive, as is the random selection of candidates for liquidation based solely on the logic of categorical ascription, knowing that once thinking along ethnic lines has become firmly established, a murder victim will always be seen as a Serb or Croat, Tamil, or Singhalese.

This contribution is a fully revised and updated version of an article by Wimmer and Schetter (2003). It aims to outline the phenomenon of ethnic violence to clarify to what extent this can be understood as a particular type of violence with an inherent logic. Moreover, the article will address the macro-political changes which form the background conditions for the emergence of ethnic violence. The third task is to inquire into the more specific institutional and political conditions which make violence a feasible option in ethnic conflicts. Finally, this contribution will present an anatomy of the violent escalation of such conflict.

The Specifity of Ethnic Violence

Ethnic violence can be defined as actions aimed at physically harming persons on the basis of their ethnic background. Pogroms, riots, and massacres are forms of collective violence which occur frequently in the context of ethnicized conflicts but can also be observed in conflict constellations which are not ethnically tainted. Ethnic cleansing, ethnocide, and the extreme case of the Holocaust represent specific forms of ethnic violence (Mann 2005).

At the individual psychological level, ethnic violence has the effect of projecting (repressed) emotions such as desires and fears onto an out-group (Petersen 2002). The ego identity is thus freed from ambivalence and its identificatory alignment stabilized in the ethnic dimension. On a large scale, this mechanism leads to the reinforcement of ethnic boundaries. Hence, the ethnic group is perceived as a community sharing a common fate, and the boundaries between ethnic groups become increasingly rigid.

Whereas such group differentiation can also be observed with other forms of violence, the specificity of ethnic violence arises from the way groups are defined. Particular regarding ethnicity common origins and shared cultural traditions shape the focus of this form of self-identification (Smith 2003). When violence spreads, this belief in community can intensify and become a purity fetish which is often further strengthened by the "racialization" of the notion of descent: One's own community is now formed through the purity of blood and must not be polluted by the blood of "others." The body of one's group is to be purged of the otherness (Chasseguet-Smirgel 1996). This purity fetishism can lead to the exclusion or even the extermination of people of dubious or hybrid ethnic classification so as to allow clear borders to be drawn. In extreme cases, this obsession with racial purity becomes pathological hatred and prepares the ground for the systematic extermination of the ethnic out-group.

Against this background, women gain extraordinary significance as guarantors of the continuation and reproduction of the group. They symbolize its purity and represent its most vulnerable and precious possession. Rape, an expression of aggression and the will to power and domination which accompanies every war, can thus gain additional symbolic and strategic significance in ethnic conflicts (Nagengast 1994). Rape means not only the humiliation and dishonoring of the enemy, who has thus proven incapable of protecting "his women," but is also intended to weaken his reproductive capacity (Mac Kinnon 1994). In the Bosnian conflict, Serbian propaganda revived a perverted model of the "devshirme" or "child-tribute." In accordance with patrilineal modes of thought, Muslim women were now to bear Serbian children, thus increasing the number of Serbs while demographically decimating the Muslims (Allen 1996).

A second specific feature of ethnic violence is ethnic cleansing. It is also closely related to the idea of ethnic descent and purity. The spectrum extends from policies of enforced assimilation with threats of violence (Hayden 1996) to the expulsion of ethnic aliens from the territory claimed as one's "own" and culminates in ethnocide as practiced in Rwanda, Bosnia-Herzegovina, or Armenia. All these acts of violence are an assault not only on the life and limb of those classified as ethnic enemies but also on the symbols of their history and identity, for example, memorials, places of worship, and graves.

Ethnic violence is often described as being particularly brutal. There is a widespread opinion among journalists and politicians that "deep-seated, ancient feelings of hatred" handed down from generation to generation are responsible for this particular intensity (see Joras and Schetter 2004). There are arguments which speak against this theory. Members of different ethnic groups often lived in harmony and only became enemies in the course of an ethnicization of political conflicts (Harvey 2000: 43). Many ethnic groups were constituted in the recent past or only attained political significance in the course of recent developments. A prominent example is the ethnic difference between Hutu and Tutsi which became politically salient only by the (post-) colonial practice of divide and rule (Lemarchand 1994). Thus, the argument that ethnic hatred is handed down from one generation to the next often corresponds to the hostile parties' own perception, but not always to the actual historical development. However, this finding should not contest the fact that the memory of past violence and atrocities can be of great significance for current conflicts.

However, even among researchers, there is widespread support for the opinion that ethnic disputes are especially violent forms of conflict. Esser (1996) attributes this to the fact that ethnic conflicts revolve around the defense of "specific capital." This specific capital (e.g., a language, cultural traditions) is only of value within one's own ethnic group. Ethnic conflicts are thus a race declaring one's own specific capital the valid currency in a society and the constitution of a state. For this particular reason, ethnic conflicts often develop into zero-sum conflicts which only allow for a winner and a loser. As soon as it is a matter of "all or nothing," conflicts are fought out with particular ferocity (Wimmer 1997).

A further argument is that people can more easily fall victim to violence in ethnic conflicts because they do not need to be of any particular ideological persuasion but can be singled out on the basis of their belonging to an ethnic group. In many cases, the opponents can be identified by clearly perceptible attributes (Malkki 1995) – language (e.g., in Sri Lanka, in the Kurdish conflict), physical features (e.g., in Afghanistan), or an entry on one's ID card (e.g., in Rwanda) – represent simple distinguishing criteria which enable mass violence. Large-scale massacres in which thousands of people are murdered are only practicable when the victims' ethnic identity can be easily determined and the danger of accidentally murdering "one's own people" is kept to a minimum.

However, there are examples of similar dimensions of violence in ideological conflicts, and this speaks against the view that ethnic violence is particularly excessive (see Mann 2005; Kalyvas 2006). The events in Cambodia, in the Soviet Union under Stalin, or in Peru in the 1980s and 1990s show that mass murder can also be carried out without a predominant ethnic ideology. It must also be taken into account that ethnic attributes are not as unambiguous as some researchers have it: Individuals can avoid clear categorization by passing and code switching, and anthropological attributes (e.g., skin color, height) often allow only the broadest of distinctions. The struggle for the validity of specific capital is also present in ideological conflicts ("socialism or barbarism").

So it can sum up that the specificity of ethnic violence is expressed in particular forms of violence such as ethnic cleansing and ethnocide. However, it seems that in terms of its extent or intensity, ethnic violence cannot be distinguished from other conflict constellations.

Emergence of Ethnic Violence in History

Typical techniques for expelling ethnic-cultural others from one's own area – mass shootings, rape, pillage, and plunder – first arose toward the end of the *reconquista* in the fifteenth century with the expulsion of the Moors and Jews from the Iberian peninsula. The Peace of Westphalia (1648), on the basis of the principle of *cuius regio, eius religio*, which was first established by the Peace of Augsburg (1555), institutionalized the ideal of the congruence of political unit and ethnic-religious community which was to become one of the central legitimizing models of the impending world order of nation-states. Equals in ethnic, cultural, and religious terms were now to rule over their like, and "foreign rule" changed from being the normal state of affairs in dynastically ruled Europe to the expression of political injustice *par excellence*.

Ethnic violence acquired an almost systematic dimension as the various waves of nation-state formation rolled through the Western world and then went on to engulf the rest of the globe. Cultural homogeneity became the standard, and it was linked up structurally to the basic political principles of the modern age, that is, to the development of democratic procedures of legitimizing power and to the legal enforcement of equality, codified in the institution of citizenship. Political modernization thus meant the transformation of the mechanisms of integration and exclusion, more precisely the dissolution of the hierarchical, potentially global mechanisms of inclusion, characteristic of premodern empires, and their replacement by egalitarian and at the same time territorially defined mechanisms such as operate in modern national states. Political modernization and ethnic-national modes of integration and exclusion are interlinked by three mechanisms (Wimmer 2002):

Firstly, hierarchically structured society with its culturally defined stratification – which often overlaps with ethnic categories – is reconceived as an egalitarian community of equals. The external borders of this community are defined in ethnic-cultural terms: only those who are the same can be equal. The people in the sense of the community of citizens with equal rights and duties toward the state are identified with the people as a cultural community.

Secondly, the democratic principle of popular rule – institutionalized in the possibility of the citizens effecting a change of government through an expression of their will and in the political participation of the entire population in this process – is interwoven with a discourse on ethnic homogeneity. Since the sovereign and the source of legitimate power is no longer the prince by the grace of God but instead the national community, the rulers and the ruIed have to be alike in ethnic-cultural terms, that is, they have to speak the same language, be of the same religious denomination, and adhere to the same everyday practices which constitute their ethnic distinctiveness.

Finally, this isomorphism of a people as citizenry, sovereign, and nation leads to the coincidence of the three corresponding territorial delineations and thus to the intensification of a process which had begun when absolutist states emerged: The external borders of the state were no longer vaguely defined zones of transition between the spheres of influence of distant political centers as they had been in premodern empires but were now sharply drawn and guarded lines separating the homogeneous inside from the heterogeneous and hostile beyond.

These three mechanisms cause an ethnicization and territorialization of the principIes of inclusion and exclusion and thus lead to the development of various “minorities”: citizens of other states (“foreigners”) who live on one’s “own” territory, ethnic minorities whose “own” state is somewhere else, and religious or ethnic diasporas without their “own” state who live side by side with the dominant ethnic group. Because legal and political inclusion is linked to the principle of nationality, the relationship to these “minorities” becomes structurally problematic: Since only those who belong “to us” in cultural-ethnic terms can legitimately be included in the community of equals before the law, and since only those who belong to the “right” nation can be part of the electorate and be voted into political office, minorities are more often than not denied their full right to equality, and their loyalty is questioned. They are perceived as a political problem, a fifth column of foreign powers on own territory, a thorn in the flesh of the nation.

The past two centuries have seen myriad examples of ethnic violence: the wars of extermination against Indian “minorities” across America, the European “wars of national liberation and unification” in the second half of the nineteenth century, and the outbursts of violence in the course of the transformation of the Ottoman and Austro-Hungarian empires into a series of nation-states at the end of World War I, the wave of nationalist-motivated cleansing and expulsion at the end of World War II, the bloodshed accompanying many declarations of independence and the ensuing ethnic conflicts – for instance, the foundation of India and Pakistan in 1948 – and the wave of violence in the course of the foundation of nation-states after the collapse of the Soviet Union (Bell-Fialkoff 1996). All these manifestations of ethnic violence were driven by the endeavor to realize the same ideal – one people, one state, one territory – in a world which de facto is characterized by ethnic-cultural intermixture, overlapping, and ambiguity.

In terms of the scale and the systematic nature of the politics of extermination involved, the Holocaust represents the culmination of modern terror. Mann (2005) correctly emphasizes that reference to "the people," in whose name violence against members of the out-group is practiced, represents a perversion of the idea of democratic inclusion, and lends the appearance of modern legitimacy to even the most monstrous acts of extermination. The two last centuries witnessed a range of examples which show that democratic inclusion and exclusion on the basis of ethnic-nationalist ascriptions are historically and systematically interlinked and frequently lead to violent conflict constellations (Snyder 2000).

As late as in the first half of the twentieth-century practices of ethnic cleansing were still legitimized in some cases by international treaties and were played down with euphemistic terms such as "exchange of population," as in the Treaty of Sevres (1920) between Greece and Turkey. Today, different political conditions prevail, and the weight has shifted in favor of those normative components of the modern age which allow the protection of human life and the right to existence of minorities to be held up against the unitary impetus of the national idea. From this perspective, the episodes successfully repressed from the historical memory of the formation of one's own state are projected in the present onto the others and transmuted into an expression of their perceived primitiveness, as in the case of Rwanda, Burundi, and DR Kongo, or fundamentally violent character, as in the case of the Balkans.

One structural precondition for the occurrence of ethnic violence is thus political modernity which first gave legitimacy to ethno-national definitions of friend and enemy and corresponding strategies of violence. Premodern violence, however, sought its objects with a different logic whose constitutive elements were the dichotomies loyal-rebellious (violence of the imperial centers against rebellious peripheries), conformist-nonconformist (institutionalized violence against delinquents or witches), and believer-unbeliever (campaigns against heretics, anti-Jewish pogroms). It is interesting to note that none of these forms of violence explicitly intended the extermination of its victims; rather, they were to be reintegrated into the hierarchical structure of society. It would seem, as Bauman (1993) has elaborated, that programs of extermination spring from the genuinely modern incapability to tolerate ambivalence.

The context of political modernization reinforces the special feature of ethnic violence that we derived from the structure of ethnic categorization in the preceding section. In the framework of the logic of the nation-state, the strategy of ethnic-cultural "cleansing" of territories and their unambiguous allocation to a single state pays off. All attempts at de-escalation through external mediation are correspondingly difficult. The experiences of the past decade (in Chechnya, Bosnia, Liberia, or also in the Great Lakes region of east-central Africa) show that efforts to bring about "win-win" situations at the negotiating table often end in failure (Oberschall 2007): The logic of ethnic-territorial classification ordains that a particular village or region must always belong to one or the other project of nation-state formation.

Political-Institutional Conflict Constellations

Examples such as Switzerland, Belgium, or Cameroon show the limits to the explanatory power of the model delineated above. Evidently, democracy and a civic ethos can be introduced and linked to the national idea without this necessarily leading to the outbreak of ethnic violence. Political modernization represents a necessary but not a sufficient condition for ethnic violence – a structural prerequisite and not a cause with uniform effect. All in all, ethnic violence is relatively rare when seen in proportion to the possibility of its occurrence (Brubaker and Laitin 1998; Olzak 2006).

Whether or not this potential for ethnic conflict is activated depends on the structure of the political institutions and processes. Two fundamental remarks should be made at the outset: Firstly, ethnic violence occurs in all types of modern political systems, both in multiparty democracies and in one-party regimes, in consociational democracies, and in military dictatorships. Secondly – and contrary to the equally popular theory of violence as a pressure valve – violence does not represent the final stage of a buildup of pressure in a situation of pervasive conflict. Rather, violence can also be consciously employed as an escalation strategy and create the very conflicts which seem to be the cause of the phenomenon of violence (Brubaker and Laitin 1998). Four paths of escalation that depend on the institutional structure of the state and the constellation of actors involved can be distinguished.

In the first path of escalation, radicalized political groups employ a strategy of violence as a means of attaining goals which they cannot reach in any other way – for example, through the ballot box or through the political mobilization of large sections of the population. Examples of this are the terrorist activities of ETA, the Quebecois separatists, or the Shivsena in India. Sometimes, it is the violent action which first leaves an ethnic mark on a particular political constellation and forces the established actors to take a stand on the issue of the ethnic character of a state and to respond to the ethnicizing discourse of the groups with a propensity for violence.

Secondly, the holding of elections in an environment of politicized ethnicity can also lead to violence (Horowitz 1985; Sisk 1996). In constellations of this kind, the parties have already redefined themselves along ethnic lines so that the general perception of party A is that it stands for ethnic group X and party B for ethnic group Y. Under these circumstances, an election becomes a census establishing the size of each group, and the democratic principle of majority rule determines and cements the ethnic balance of power. For radicalized party militias, it is thus a conceivable and possibly even worthwhile strategy to use violence and intimidation to keep the other party – or rather its voters – from participating in the elections.

The third most widely discussed path to violence evolves by itself, as it were, from the very logic of the democratic process: In an ethnicized party system, the political positions frequently radicalize (Sisk 1996). In non-ethnicized party systems, politicians have to try above all to gain the support of floating voters in the middle of the political spectrum: Thus, they exercise moderation. An ethnic party, on the other hand, competes for votes only *within* a clearly demarcated segment of the population because in a climate of tension the affiliation of individuals to a particular group is beyond doubt. For the leaders of ethnic parties, it therefore makes sense to adopt radical positions, and in this way, they beat their rivals in representing the “true” interests of the group; in ethnically divided electorates, they can be assured of receiving the support of moderate voters. After all, where there are clear demographic majority conditions and open political competition, the political subordination of minorities is cemented, and a change can only be brought about through violence. However, the question as to whether this dynamics of escalation is inevitable, or whether the majority ethnic group can – or must – split into several moderate parties, so that in the end trans-ethnic coalition governments can be formed, is still a controversial issue (Horowitz 1985).

The fourth path to violence is taken by state authorities themselves when they organize state terror or even extermination campaigns against particular ethnic groups. It would make sense to distinguish between cases in which the state apparatus is controlled by an ethnic minority and those where the demographic majority ethnic group controls the state. Under modern political conditions – where the ethnic-cultural representativeness of the government is a central legitimizing principle – the status and position of minority elites are particularly precarious; for this reason, they tend to a drastic use of violence at any sign of protest against their ethnocracy. The events in Burundi in 1972, 1988, and 1993 were exemplary: No sooner had educated Hutu dared to question the dominance of the Tutsi than they were massacred in pogroms throughout the country (Lemarchand 1994). A different logic would seem to apply in cases where violence is used by a state elite belonging to the majority ethnic group to pre-empt or neutralize irredentist aspirations – for example, the case of the Armenian genocide in the course of the foundation of the Turkish state (Dadrian 1995). Such policies are also typically connected with the foundation of new nation-states upon the ruins of empires, for exampIe, after the collapse of the Soviet empire in the 1990s – national polarization within the state and the accompanying mobilization of the population at the very moment when the issue of the republics' borders was on the political agenda led to a spiral of violence (Toft 2003).

These four paths of escalation can lead to similar or different forms of ethnic violence, to phenomena characteristic of ethnic violence such as ethnocide or general ones such as terrorist acts against individual members of an enemy group. Irrespective of the institutional-political context discussed above, the very logic of ethnic-national category-formation implies certain dynamics of violence and counter-violence, a specific dramaturgy of terror.

The Anatomy of Ethnic Violence

Contrary to the image frequently encountered in the media, research has shown that the dramaturgy of terror follows certain rules (Brass 1996; Tambiah 1996). Often enough, a spiral of political radicalization is to be observed. The propensity to violence grows with increasing fear and deepening mistrust toward the political representatives of other groups and toward the state as the holder of the monopoly on the legitimate use of force and protector from arbitrary violence.

Groups with a propensity to violence play an outstanding role in the production of this kind of climate (Tambiah 1996; Gallagher 1997). They must depict ethnicity as the unquestioned principle of group formation in times of crisis and ethnic violence as a legitimate means of defending collective interests. In this context, particularly, the instrumentalization of mass media plays a crucial role: Calls to violence on radio or television often enough accompanied riots in Southeast Asia, as well as the massacres in Rwanda, and the ethnic cleansing in Yugoslavia.

Moreover, fear and mistrust give credibility to rumors and in turn are increased by credible rumors (Horowitz 1985): The misinterpretation of information causes the loss of credibility of the other ethnic group, which can ultimately lead to a security dilemma: Each group feels that the enemy is capable of the worst (Roe 2005). The thinner the network of relations between the groups, the more likely misinterpretation of information becomes (Varshney 1997). In this way, fear and misinterpretation can culminate in the conviction that discrimination, repression, or even destruction can only be avoided by taking recourse to armed offensives before the enemy does so (Lake and Rothchild 1996).

Additionally, the radicalization of ethnic stereotypes and the reinforcement of symbolic hierarchies fuel the expectation of and propensity to violence. One's own culture is stylized as the only acceptable form of human existence, and the past is idealized in terms of a history of the oppression or grandeur of one's own group (Smith 2003). The more comprehensive such a "shared social frame of reference" has become, the more readily a propensity to violence can be brought about by the propaganda of radicalized groups. The opposing group is demonized (Gallagher 1997) and viewed disparagingly as a horde of faceless creatures deprived of all individuality. As shown by the defamation of the Tutsi as "cockroaches" by the Hutu propagandists or of the Hazaras as "mules" by Tajik and Pashtun leaders, the disparagement of the enemy can go so far as to be condensed into one single pejorative word. The use of violence against groups denigrated and dehumanized in this way thus appears legitimate and even desirable.

When the fear of a demonized and dehumanized enemy spreads, every new event which fits into the ethnic friend-enemy scheme reinforces the validity of the scheme and the ethnicized interpretation of the situation. Even attacks on individuals in no way directly connected with the conflict can be interpreted as attacks on the ethnic group as a whole. Groups with a propensity to violence can use this mechanism to trigger off dynamics of escalating violence through selective assassinations and thus keep a conflict alive. In this context, Wright (1988) speaks of the representativeness of (ethnic) violence. The victims are not "selected" by reason of their personal characteristics but purely on the basis of their group membership. In this way, all members of the entire group become potential victims, and the actual victim represents them all. Interpreting violence as "ethnic" paves the way for further violence and can escalate into ethnic riots or even wars.

Riots and civil wars differ in their intensity and scale. Riots can mark the beginning of civil wars, while in some cases, wars are directly initiated by militias or armies without the involvement of civilians (Horowitz 2003). Ethnic civil war therefore does not necessarily represent an intensified form of ethnic riots, as implied by the popular idea that ethnic wars are an expression of increasing hatred between opposed ethnic groups. Rather, ethnic conflicts are usually characterized by an oscillation of violence such as in Sri Lanka, the Basque Country, or Northern Ireland, and less by linear dynamics of escalation. In contrast to their appearance as spontaneous, unbridled, and uncontrolled manifestations of hatred, riots and pogroms show some regularities and structures. Firstly, instigators, organizers, and manipulators play a major role (Tambiah 1996). Often pogroms have been triggered off by public or concealed calls to violence by prominent politicians; the attackers thus thought themselves safe from criminal prosecution and their acts of violence appeared legitimate, or even desirable. The massacres of Tutsi in Burundi were even initiated and partly carried out by government institutions.

Secondly, the cyclical occurrence of such riots can lead to a routinization and ritualization of ethnic violence. Brass (1996) speaks in this context of "institutionalized riot systems." Often ethnic identity is celebrated symbolically in specific locations (memorials, holy places, churches, etc.) and on particular holidays or days of mourning. These occasions often become the trigger and focal point of ethnic violence (Tambiah 1996). This is because localized festivals and processions offer an appropriate occasion for riots, or provoke the opposing group to commit acts of violence (for example on the occasion of the Orange Order procession). Attacks and pogroms clearly show that regardless of how an individual defines him or herself, it is ethnic classification that decides life and death. The representativeness of ethnic violence makes everyone a 'prisoner' of the groups with a propensity to violence. Anyone who rejects or even simply fails to actively support their 'own' protectors is branded a traitor or informer and runs the risk of being left completely at the mercy of the enemy. The survival of the individual then depends very much on the power relations between ethnic groups. In ritualized "riot systems," there is an increasing tendency toward formation of ethnic ghettos within which protection from the enemy is more readily afforded. Communication is limited to members of one's "own" ethnic group since those on the "other side" are no longer to be trusted. This lack of communication, in turn, fuels mutual mistrust.

Such "riot systems" can stabilize and have a life of their own, often in a localized context without direct repercussions on the macro-political constellations. In other cases, attacks of this kind trigger off – or provide the occasion for – large-scale violence which challenges the basic structures of the political system. In an "ethnicized climate" in which all political decisions are judged and justified from the angle of ethnic stereotypes, ethnic violence presents itself as an instrument for attaining long-term political objectives. The exact motives and violence strategies vary with the political-institutional constellations. However, all the actors in such ethnic civil wars remain tied to the logic of the nation-state – the idea that power can be justified by reference to an ethnically defined people, and that every people is entitled to a state which then merits recognition.

Ethnic civil war frequently begins with the occupation or conquest of territories which are considered strategically significant for the security of one's "own" ethnic group. The territorialization and ethnicization of concepts of belonging, characteristic of politics in the age of nation-states, also determines the further course of ethnic civil wars: the tendency toward expulsion or destruction of ethnic others from the territory of one's "own" future state, the use of expellees of one's own ethnic group for the colonization of such "cleansed" areas, the wanton destruction of remaining niches of interethnic coexistence, the particular brutality unleashed upon everything mixed and hybrid, and the destructive frenzy toward everyday signs of the existence of "foreign" groups – blowing up houses, filling in wells with concrete, the destruction of arable land, and razing temples, churches, statues, etc., which could serve as a memory of those who once "owned" the territory. Members of the enemy group – demonized and dehumanized – fall victim to the violent practices of marauding bands of young fighters: They are shot as an example to terrorize the survivors and make them flee, or they are herded together in camps or "resettled," or – as the ultimate form of ethnic violence – they are murdered en masse as in Rwanda or in Bosnia in minutely planned operations.

Related Entries

- ▶ Genocide
- ▶ Outdoor Serious Violence: The Role of Place
- ▶ Race and Ethnicity in Social Disorganization Theory
- ▶ Race, Ethnicity, and Youth Gangs

Recommended Reading and References

Allen B (1996) Rape warfare: the hidden genocide in Bosnia-Herzegovina and Croatia. University of Minnesota Press, Minneapolis

Bauman Z (1993) Modernity and ambivalence. Blackwell, Oxford

Bell-Fialkolff A (1996) Ethnic cleansing. Griffin, New York

Brass P (ed) (1996) Riots and pogroms. New York University Press, New York, pp 1–55

Brubaker R, Laitin D (1998) Ethnic and nationalist violence. Annu Rev Sociol 24:423–452

Chasseguet-Smirgel J (1996) Blood and nation. Mind Hum Act 7:31–33

Dadrian V (1995) The history of the Armenian genocide. Berghahn Books, Providence

Esser H (1996) Ethnische Konflikte als Auseinandersetzung um den Wert von kulturellem Kapital. In: Heitmeyer W, Dollase R (eds) Die bedrängte Toleranz. Ethnisch-kulturelle Konflikte, religiöse Differenzen und die Gefahren politisierter Gewalt. Suhrkamp, Frankfurt a.M. pp 64–99

Gallagher T (1997) My neighbour, my enemy: the manipulation of ethnic identity and the origins and conduct of war in Yugoslavia. In: Turton D (ed) War and ethnicity. Global connections and local violence. University of Rochester Press, San Marino, pp 47–75

Harvey F (2000) Primordialism, evolutionary theory and ethnic violence in the Balkans: opportunities and constraints for theory and policy. Can J Polit Sci 33(1):37–65

Hayden R (1996) Imagined communities and real victims: self determination and ethnic cleansing in Yugoslavia. Am Anthropol 23(4):783–801

Horowitz D (1985a) Ethnic groups in conflict. University of California Press, Berkeley

Horowitz D (1985b) The deadly ethnic riot. University of California Press, Berkeley

Joras U, Schetter C (2004) Hidden ties: similarities between policy and research approaches to ethnic conflicts. In: Wimmer A et al (eds) Facing ethnic conflicts. Towards a new realism. Rowman & Littlefield, Lanham, pp 315–332

Kalyvas S (2006) The logic of violence in civil war. Cambridge University Press, Cambridge

Lake D, Rothchild D (1996) Containing fear. The origins and management of ethnic conflict. Int Sec 21(2):41–75

Lemarchand R (1994) Burundi. Ethnocide as discourse and practice. Cambridge University Press, Cambridge

Mac Kinnon C (1994) Rape, genocide and women's human rights. In: Stiglmayer A (ed) Mass rape: the war against women in Bosnia-Herzegovina. University of Nebraska Press, Lincoln

Malkki L (1995) Purity and exile: transformations in historical-national consciousness among Hutu Refugees in Tanzania. Chicago University Press, Chicago

Mann M (2005) The darkside of democracy: explaining ethnic cleansing. Cambridge Univesity Press, Cambridge

Nagengast C (1994) Violence, terror, and the crisis of the state. Annu Revue Anthropol 23:109–136

Oberschall A (2007) Conflict and peace building in divided societies: responses to ethnic violence. Routledge, London

Olzak S (2006) The global dynamics of racial and ethnic mobilization. Stanford University Press, Stanford
Petersen R (2002) Understanding ethnic violence, fear, hatred and resentments in twentieth-century Eastern Europe. Cambridge University Press, Cambridge
Roe P (2005) Ethnic violence and societal security dilemma. Routledge, London
Rothschild J (1981) Ethnopolitics. A conceptual framework. Columbia University Press, New York
Sisk T (1996) Power sharing and international mediation in ethnic conflicts. United States Institute of Peace, Washington, DC
Smith A (2003) Chosen people. Sacred sources of national identity. Oxford University Press, Oxford
Snyder J (2000) From voting to violence. Democratization and nationalist conflict. Norton, New York
Tambiah S (1996) Leveling crowds. Ethnonationalist conflicts and collective violence in South Asia. University of California Press, Berkeley
Toft M (2003) The geography of ethnic violence: identity, interests, and the indivisibility of territory. Princeton University Press, Princeton
Varshney A (1997) Postmodernism, civic engagement, and ethnic conflict. A passage to India. Comp Polit 30(1):1–20
Wimmer A (1997) Who owns the state? Understanding ethnic conflict in post-colonial societies. Nat Natl 3(4):631–665
Wimmer A (2002) Nationalist exclusion and ethnic conflict. Shadows of modernity. Cambridge University Press, Cambridge
Wimmer A, Schetter C (2003) Ethnic violence. In: Heitmeyer W, Hagen J (eds) International handbook of research on violence. Kluwer, Dordrect, pp 247–260

Ethnoviolence

▶ Hate Crime

European Perspectives on Long-Term Imprisonment

Kirstin Drenkhahn
Department of Law, Freie Universitaet Berlin,
Berlin, Germany

Overview

Sentences of long terms of imprisonment seem to be a fashion in criminal politics that never passes. They are readily explained as necessary to protect the public from dangerous offenders (incapacitation) and as being deserved for very serious crimes (retribution). These simple ideas get more complicated with a closer look at what kinds of sanctions consist of long terms of detention in prison, what they are aimed at, who is sentenced accordingly, and what happens to long-term prisoners while detained.

A European perspective has the benefit of a wide range of sanctioning and prison systems with a common legal frame of reference, the European Convention for the Protection of Human Rights and Fundamental Freedoms. At the same time, European criminal justice systems disagree on many aspects that are relevant for long-term imprisonment, such as the specific sanctions, their aims, and the conditions of their execution. Thus, a European perspective offers the whole range of problems in a condensed form.

Fundamentals

European Prison Law

A European perspective on imprisonment that takes into account common legal requirements needs to go beyond the European Union. Although cooperation in penal matters and human rights protection are important political issues of the EU, it has not yet produced any official documents relating to long-term imprisonment (for developments: van Zyl Smit and Snacken 2009, p. 27 ff.). One reason is that all EU member states are also members of the Council of Europe, an international organization with 47 member states and even more criminal justice and prison systems. Since its founding in 1949, the Council of Europe has been a main promoter of human rights protection in Europe with the European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR) as the principal document. The ECHR is legally binding for the member states, and compliance may be assessed before the European Court of Human Rights (ECtHR, established in 1959) in an individual application process. In addition, the Council of Europe gives nonbinding recommendations that specify the ECHR in terms of minimum standards and are used by the ECtHR as well as national courts when interpreting the ECHR and domestic law, respectively. Two recommendations are particularly important for matters of long-term imprisonment, the Rec(2003)23 on the management by prison administrations of life sentence and other long-term prisoners and the Rec(2006)2 on the European Prison Rules. Another relevant body is the European Committee for the Prevention of Torture and Inhuman or Degrading Treatment or Punishment (CPT, established in 1989) that monitors the material conditions of and develops standards for specific situations of detention (van Zyl Smit and Snacken 2009, pp. 9–24). Within this system, it is agreed that imprisonment serves as punishment in itself, not as an occasion for punishment, and that prisoners "retain all rights that are not lawfully taken away by the decision sentencing them or remanding them in custody" (European Prison Rules, rule 2).

Definition of Long-Term Imprisonment

If a term of imprisonment is thought of as already long depends on the maximum length of sentences that a legal system offers and probably even more on the sentencing practice in a given jurisdiction. In Europe, there is a large variety in both regards: In Croatia, Norway, Portugal, and Spain, the law does not provide for life sentences as a punishment whereas, e.g., in Lithuania, there is the possibility of life without parole. Then, there are jurisdictions such as England and Wales, Germany, and Norway where criminal courts may issue orders of detention in prison after a prison sentence for long or even indeterminate periods as a means of protection of the public against recidivism. Regarding the length of sentences prisoners are actually serving, the Annual Penal Statistics of the Council of Europe (SPACE I) for 2010 show four jurisdictions where more than 10 % of the prison population were serving life sentences (England and Wales, Greece, Northern Ireland, Scotland) with a mean for all member states of 3.1 %. But there are also seven jurisdictions where more than a third of the 2010 prison population were serving a sentence of less than 1 year (mean: 16.4 %). Given this wide range, the Rec(2003)23 on long-term prisoners defines its target group as prisoners serving a life sentence or a prison sentence or sentences totaling 5 years or more.

Such long terms of detention in prison may be justified in several ways. The two basic ideas are, firstly, penalties as a response to past criminal behavior that are meant to be proportionate to the amount of guilt the offender has incurred and that carry moral censure (retribution) and, secondly, measures on occasion of an offense that are justified with the offender's alleged dangerousness and the need for public protection against future crimes (prevention, incapacitation). In both instances, at least diminished criminal responsibility is required. The implementation of these ideas in European criminal justice systems differs considerably (Padfield 2010, p. 25 ff.):

- Some systems provide only for commensurate penalties (Croatia, Slovenia).
- Some provide for possibilities to exceed such penalties for certain categories of offenders (e.g., recidivists, dangerous offenders) which are justified with either just deserts or public protection (Belgium).
- Some use life sentences for public protection after an individually or generally fixed term has been served. Release may then only be granted if there is no considerable risk of recidivism (England and Wales, Germany).
- There are two-track systems that distinguish between penalties and measures of public protection where such measures are imposed or reserved at the time of the sentence (Austria, Belgium, France, Germany, Spain, Switzerland).
- Some systems allow for the imposition of preventive detention at the time of release (Germany, Switzerland).

Prison Population in Long-Term Imprisonment in Europe

Table 1 provides data on prison population rates per 100 000 inhabitants and percentages of prisoners serving long prison sentences derived from SPACE I 2000, 2005, and 2010. For reasons of

European Perspectives on Long-Term Imprisonment, Table 1 Prison population rate (per 100 000 inhabitants, rounded) and percentage of long-term prisoners (5 years – life sentence) in member states of the European Union, 2000, 2005, 2010

	2000		2005		2010	
Country	Prison population rate	Long-term prisoners (%)	Prison population rate	Long-term prisoners (%)	Prison population rate	Long-term prisoners (%)
Austria	93	–	107	21.6	103	41.6
Belgium	85	45.0	90	59.8	105	52.5
Bulgaria	115	35.8	158	–	124	22.4
Cyprus	–	–	–	–	112	38.4
Czech Republic	219	–	186	18.4	209	21.0
Denmark	61	–	76	18.3	71	22.7
Estonia	328	51.5	327	40.1	259	37.1
Finland	52	–	73	12.9	62	29.2
France	80	42.0	92	37.6	103	25.7
Germany	–	–	96	12.9	88	12.5
Greece	76	–	87	70.4	106	73.0
Hungary	158	31.4	162	–	164	28.4
Ireland	76	–	–	–	97	39.7
Italy	93	47.3	102	46.2	113	42.1
Latvia	353	45.5	313	48.0	301	50.8
Lithuania	240	46.0	233	34.0	267	42.5
Luxemburg	90	45.0	152	35.4	137	43.1
Malta	–	–	74	42.5	141	–
Netherlands	90	18.8	134	19.7	71	17.4
Poland	169	–	216	16.5	211	14.0
Portugal	–	–	122	56.8	109	56.4
Romania	221	33.3	175	47.8	131	53.0
Slovak Republic	297	28.0	172	27.6	186	29.3
Slovenia	57	27.3	57	28.3	66	32.1
Spain (Catalonia)	–	–	–	–	143	48.8
Spain (State Admin.)	114	–	142	25.5	165	25.3
Sweden	64	21.6	78	26.2	74	–
UK: England and Wales	124	30.7	143	40.7	154	43.2
UK: Northern Ireland	–	42.4	77	43.0	82	55.3
UK: Scotland	–	44.1	133	37.3	151	30.5

Source: SPACE I; –: no data available

clarity, figures are shown only for EU member states and not the whole Council of Europe. Taking into account the huge differences in the size of general and prison populations between countries, absolute numbers would not be informative (explanation of these differences: Lappi-Seppälä 2012). The data show that both figures did not develop homogenously. There seems to be a trend in Western European countries with a growing prison population in general and an increasing percentage of long-term prisoners (Report accompanying the Recommendation (2003)23 on long-term prisoners, § 17). In 2010, there were only three countries with less than 20 % of their prison population serving 5 years or more, while there were six with a share of at least 50 %.

There are no European statistics that provide sociodemographic and offense-related data for specific groups of prisoners, and many member states do not have such national statistics. Thus, it is not possible to give a more detailed description. The report accompanying the Recommendation (2003)23 on long-term prisoners states "that murder, manslaughter, serious sexual offences, robbery, kidnapping, drug trafficking, involvement in organized crime, crimes against the security of the realm and against humanity commonly lead to long-term and life imprisonment" (§ 10). Still there are considerable differences in the specific legal requirements and the actual minimum – and maximum – sentences for such offenses (concerning life sentences: Griffin and O'Donnell 2012, p. 611 f.).

Research on mental health found that many European long-term prisoners suffer psychological distress at a clinically relevant level (Dudeck et al. 2011, with an overview of studies on prisoners' mental health in general p. 405; Grounds 2004). The prison systems in Europe are not prepared for adequate care during incarceration, and psychiatric aftercare for released persons is insufficient (Salize et al. 2007).

Key Issues and Controversies

Long-Term Imprisonment, Dangerousness, and the Prison Regime

The alleged dangerousness is one justification for long prison sentences. The term "dangerousness" is not self-explanatory but largely subjected to zeitgeist. If it is used as a rationale for a long prison sentence, it usually describes predicted future behavior of the convict/prisoner, meaning a certain risk of recidivism after release, although it is not clear to what kind of offenses or damage it relates, how certain this risk should be, and, most importantly, how this risk should be predicted (van Zyl Smit 2006, p. 415 ff.). "Dangerousness" may also describe a present risk for the physical integrity of others or the prisoner her-/himself (personal safety) or a risk for the external security, e.g., through violent escapes. Not all of these forms of dangerousness are inherent to all long-term prisoners – some may even not be dangerous at all – and these different forms of dangerousness should call for different precautions in prison.

The Rec(2003)23 on long-term prisoners (Snacken and van Zyl Smit 2009) refers to this problem most prominently in its general principles. The "security and safety principle" (no. 6) requires to differentiate between the risks posed to the external community and those to persons inside the prison including the risk of auto-aggressive behavior. It is recommended not to segregate long-term prisoners only because of their sentence (nonsegregation principle, no. 7). In addition, an individual risk and needs assessment should be used to offer appropriate interventions to address not only the risk of recidivism after release but also the present risk for him-/herself and the surroundings that a prisoner may pose (no. 10). In its section on security and safety in prison, the recommendation recognizes that segregation of individual prisoners might be necessary, but it should be avoided or at least reduced in time. Furthermore, maximum security units are described as a last resort. The regime in such units again should distinguish between the different types of dangerousness (no. 20 b). But even in maximum security units, there should be a relaxed atmosphere, relative openness, and freedom of movement, and security measures should only be used to the extent that is necessary and with respect to human dignity and human rights. According to the European Prison Rules, solitary confinement shall be used as a punishment only in very exceptional cases and for as short as possible (Rule 60.5).

If national prison administrations adhere to these rules is quite another question. Internationally comparative research on prison conditions is scarce; to date there seems to be only one cross-national study on prison conditions in long-term imprisonment in Europe (Drenkhahn et al. 2013). Information is provided in the reports of the CPT on its visits to member states of the Council of Europe. The Committee's members visit prisons and other places of deprivation of liberty and talk to prisoners in private. The observations are summarized in a report that will usually be published on the CPT's website. Based on these reports, the CPT has developed standards that serve as a rule for future visits (CPT 2011).

Already as early as 1991, the CPT paid special attention to solitary confinement. It stated that such a regime can amount to inhuman or degrading treatment and therefore should be as short as possible (CPT 2011, p. 20; Smith 2006). In 2011, the CPT pointed out the considerably higher rate of suicide among prisoners in solitary confinement than among the general prison population, concluding that solitary confinement on its own has the potential to amount to torture or inhuman or degrading treatment or punishment. According to the case law of the ECtHR, solitary confinement is only justified if it is proportionate to the risk; if it is lawful in domestic law; if it is accountable, meaning that all decisions must be documented; if it is necessary; and if it is nondiscriminatory. For the material conditions of solitary confinement, the same standards apply as for other prisoner accommodation, that is the European Prison Rules. The regime in solitary confinement should subject the prisoners to no more restrictions than necessary for their safe and orderly confinement (CPT 2011, p. 29 ff., 35 f.). These recommendations show that the CPT has found in the recent past that even basic requirements are not met.

A common practice is the transfer of troublesome prisoners to another establishment or even a chain of transfers. The CPT describes the counterproductive effects of this practice (CPT 2011, p. 20; van Zyl Smit and Snacken 2009, p. 271 ff.).

Another widespread problem is inter-prisoner violence. In this context, the CPT stresses the importance of a good quality of staff-prisoner relations and a good prison climate in general. It holds that it is most important that staff have appropriate interpersonal communication skills and use them (CPT 2011, p. 23) rather than weapons or other instruments of force. This is backed up by recent research (e.g., Liebling 2004; Snacken 2005). Liebling (2004, p. 154 f.) identified the relationship between prisoners and staff and social aspects of the regime as key elements of what really matters in prison. Snacken (2005, p. 335) found in her study on violence in Belgian prisons that although increased positive personal contacts would lead to more arguments and conflicts, the level of all forms of violence was lower than in prisons with a strict regime and less personal contacts.

The CPT also gives recommendations for high security units and life-sentenced and other long-term prisoners. Although high security risk prisoners will be segregated from the rest of the prison population, the regime in their units should be relatively relaxed, providing them with opportunities to meet fellow prisoners in the unit and with a variety of activities from which they can freely choose as well as adequate treatment programs. Humane treatment of the prisoners as well as effective control and staff security should be maintained with the help of positive relations between staff and prisoners. Again, the transfer to such a special regime should only last as long as necessary (CPT 2011, p. 27 f.).

With regard to long-term and especially life-sentenced prisoners, the CPT reports that “during some of its visits, the CPT has found that the situation of such prisoners left much to be desired in terms of material conditions, activities and possibilities for human contact. Further, many such prisoners were subject to special restrictions likely to exacerbate the deleterious effects inherent in long-term imprisonment; examples of such restrictions are permanent separation from the rest of the prison population, handcuffing whenever the prisoner is taken out of his cell, prohibition of communication with other prisoners, and limited visit entitlements” (CPT 2011, p. 28). The CPT points out that this indiscriminate application of restrictions to all such prisoners does not take into account the individual risk they may (or may not) present. So at least in some prisons in Europe, many non-desirable practices that are commonly justified with security considerations are part of the regular regime for long-term prisoners.

Effects of Long-Term Imprisonment

Much of the research on the effects of imprisonment in general has been conducted in prisons where long prison sentences are executed. Most of the early studies starting in the 1950s came from North America, but there is also a lot of research on a wide range of aspects of imprisonment in Europe (overview: Liebling and Maruna 2005; van Zyl Smit and Snacken 2009, p. 47 ff.; Zamble and Porporino 1988, p. 7 ff.; examples of European research: Chauvenet et al. 2008; Faugeron et al. 1996; Liebling 2004; Sparks et al. 1996).

In his classical study on a maximum security prison, Sykes (1958) coined the term "pains of imprisonment." These are:

- the loss of liberty,
- the deprivation of goods and services,
- the frustration of sexual desire,
- the deprivation of autonomy, and
- the deprivation of security.

Thus, the prison as a bureaucratic entity takes away prisoners' opportunities to express themselves as individuals and to exert control over their lives, while assuming this control and regulating the lives of prisoners with indifference to their individual needs. Prisoners respond to this situation with creating a "society of captives" with its own rules and values in order to overcome some of the deprivations and regain a certain level of control over their lives (see also Goffman 1961: "total institution").

Since then, a vast amount of research has been done. Although there are many studies that were able to show detrimental effects for psychological well-being such as diminishing self-esteem, a slower reaction time, and an increased hostility, there were others that found no evidence of psychological or intellectual deterioration or even identified desirable effects (van Zyl Smit and Snacken 2009, p. 50 ff.). These latter results led to a growing criticism from the 1970s onwards by psychologists of the mainly sociological studies on the "pains of imprisonment." This work was seen as methodologically unsound (qualitative vs. quantitative) and ideologically biased (contextual vs. individual). Much of the later research concentrated on coping with and adaptation to imprisonment and suggested that prisoners cope relatively well with their situation and at least do not change for the worse (Liebling and Maruna 2005, p. 10 ff.). Zamble and Porporino (1988, p. 149 ff.) described this as a "deep freeze." They found that prisoners remained mainly unchanged by the experience of imprisonment: in the face of problems, they would use the same coping strategies after release than before incarceration (Zamble and Porporino 1988, p. 152).

This finding has since been challenged by research that takes into account both the individual and the context (Haney 2005). The "deep freeze" itself might even be seen as a detrimental effect. In Jamieson's and Grounds' (2005) studies, ex-prisoners reported that after release they entered society on the very level of personal development that they had reached at the time of incarceration. This was felt to be quite disturbing, because it meant that in their late 30s or early 40s, they only had the life experience of someone in their early 20s. Other than their friends who had remained free, they had not had the opportunity to use this time for personal development and for settling down, finding a steady job, and having a family and children. Others (Irwin and Owen 2005; Murray 2005; Snacken 2005) describe aspects of prison life that work against the official aim of imprisonment (correction, rehabilitation, reintegration, preparation for the crime-free life of a responsible citizen) such as violence in prison, loss of agency (strict schedule organized around the shifts of uniformed staff, only small room for personal choice), the exposure to transmittable diseases (HIV/AIDS, hepatitis C, tuberculosis), loss of contact with family and caring others, and economic problems (debts, maintenance claims, exploitation through prison work). The research on prison climate and "what matters in prison" shows that social characteristics of prison life such as respect, trust, humanity, relations with staff and other prisoners, support, fairness, order, security, well-being, personal development, family contact, decency, the use of power, and giving meaning to the punishment (Liebling 2004) are major factors for surviving imprisonment. High levels of distress concerning security, family contact, fairness, and respect may even increase suicidality (Liebling et al. 2005).

To conclude, long-term imprisonment has the potential to be a very harmful experience – if and to what extent depends on characteristics of the individual prisoner, material conditions of the environment, and social interaction in prison. Thus, the prison experience not only depends on the official concept of a given system as, e.g., stated in prison laws, but much more on the microcosm that a prison and its population form.

Human Rights Issues of Long Prison Sentences

These findings hint at considerable infractions of human rights that go with imprisonment, but may be more severe for long-term prisoners given the special restrictions for dangerousness and security reasons. Living conditions in prison can add up to a violation of the prohibition of torture (art. 3 ECHR; e.g., ECtHR, Peers v Greece, 19 April 2001, application no. 28524/95; ECtHR, Kalashnikov v Russia, 15 July 2002, application no. 47095/99); the right to respect for private and family life (art. 8 ECHR) is seriously restricted, and there may be more restrictions of fundamental freedoms and rights. Even the right to life (art. 2 ECHR) may be touched in material conditions that foster the spreading of diseases and offer no adequate health care.

Apart from these human rights problems during the execution of a prison sentence, there are other human rights issues with long-term imprisonment. Firstly, there is the question if whole life sentences or even life sentences with the prospect of release are acceptable at all (Appleton and Grøver 2007; van Zyl Smit 2006). Since the death penalty has been abolished in all circumstances in most of the Council of Europe member states (44 of 47; Protocol No. 13 to the Convention for the Protection of Human Rights and Fundamental Freedoms, concerning the abolition of the death penalty in all circumstances; ECtHR [Grand Chamber], Öcalan v Turkey, 12 May 2005, application no. 46221/99, §§ 150 ff.), life imprisonment is the harshest criminal penalty at hand on the continent. But European criminal justice systems do not agree if life sentences are acceptable at all and, if so, whether the same is true for life without parole.

Some countries do not allow for life sentences at all: in Portugal, it is outlawed by the constitution, other countries just do not provide for it in the criminal code. The reasons for this are a humanitarian approach to punishment with a strong emphasis on human dignity and the view that the time of imprisonment has to be used for reeducation, i.e., preparation for a future crime-free life not by deterrence. In this perspective, a life sentence always bears the risk that the convict will never have the possibility to show that his efforts were successful (van Zyl Smit 2006, p. 410 f.).

Other criminal justice systems provide for life sentences, but every convict must have a chance for release. One example is Germany where the Federal Constitutional Court decided in 1977 (BVerfGE 45, p. 187, 21 June 1977) on the constitutionality of the mandatory life sentence for murder, using almost the same arguments as used against the life sentence in other countries. The Court held that this sanction is not per se unconstitutional, because the legislator has the discretion to choose the heaviest penalty at its disposal for the crime of murder as this crime violates the highest legal interest, the right to life. But at the same time, the constitutional "principle of resocialization" requires the state to offer life prisoners a concrete and fundamentally realizable expectation of release and a prison regime that prepares these prisoners for release (BVerfGE 45, p. 187, 245 f.; van Zyl Smit 2006, p. 408 f.). The practice of release through executive clemency did not meet the requirements of legal certainty, and therefore, Germany introduced rules for parole in life sentence cases into the Criminal Code in 1981. Other examples where constitutional courts adopted the same opinion are France and Italy. In England, on the other hand, whole life sentences are seen as in principle acceptable for the most heinous crimes for retribution and deterrence (R v Home Secretary, Ex parte Hindley [2001] 1 AC 410 [HL (E)]; Appleton and Grøver 2007, p. 603 ff.). The ECtHR has not had to decide yet if a whole life sentence is contrary to the ECHR and in particular if a life sentence without the formal prospect of release is an inhuman or degrading punishment in the sense of art. 3 ECHR. In the cases where this question was raised, the prisoners had some sort of prospect for release (van Zyl Smit 2006, p. 409 f.).

The second question concerns measures that are considered as nonpunitive but preventive by national legislation and that are in substance long-term or life imprisonment (preventive detention). The problem is how this infraction of the right to liberty (art. 5 ECHR) can be justified (Drenkhahn et al. 2012; Merkel 2011). For a long time, the ECtHR could easily define preventive detention as an exception of the right to liberty in the sense of art. 5(1)a ECHR, "the lawful detention of a person after conviction by a competent court", because the relevant domestic laws (e.g., in Belgium and Germany) required that the sanction was issued in the same judgment as the finding of guilt and the prison sentence. But from the end of the 1990s onwards, legislators in Europe introduced rules to broaden the scope of preventive detention. In 1998, Germany replaced the absolute limit of 10 years on the length of the first term of preventive detention by a provision that allowed for its indefinite extension even in cases where preventive detention had already been ordered. In 2004, the possibility to impose preventive detention at the time of release from the prison sentence was introduced. This possibility was also introduced in Switzerland in 2002, in force since 2007. In a series of judgments against Germany, the ECtHR decided that the retroactive extension of preventive detention over the 10-year limit was not justified by art. 5(1)a ECHR, because it was not causally linked to a conviction (M. v Germany, 17 December 2009, application no. 19359/04; Drenkhahn et al. 2012, p. 170 ff.). Consequently, the Court held that a preventive detention that was imposed at the end of the prison sentence on occasion of the same offense was not a detention "after conviction" (Haidn v Germany, 13 January 2011, application no. 6587/04; Merkel 2011).

The ECtHR considered "the lawful detention [. . .] of persons of unsound mind" (art. 5(1)e ECHR) as a justification in the above-mentioned cases but found that its requirements were not substantiated (Drenkhahn et al. 2012, p. 172; Merkel 2011, p. 973). Nevertheless, this ground seems promising to politicians and legislators who want to keep a wide scope of application for preventive detention, because there is a high prevalence of mental disorders, especially personality disorders, in persons serving such a sentence, although they did not lack criminal capacity (Habermeyer et al. 2010). As this ground for detention does not require a criminal conviction, it gives room for flexibility. The main objection though is that it sets the problem of what to do with dangerous criminals after they have served their sentence on a different track, from criminal justice and prison to psychiatry, and brings back to the front the old and outdated idea that criminality per se is some sort of insanity (Drenkhahn et al. 2012, p. 180 ff.; Merkel 2011, p. 975).

International Perspectives

Most of the issues raised in this article from a European perspective are also relevant in criminal justice systems outside of Europe. But when studying and comparing prison systems and their regimes for long prison sentences as well as the effects of imprisonment, it is crucial to remember that the official and unofficial rules might differ considerably between systems and even within one system. The same is true for material conditions and social interactions. In addition, prison systems might change considerably over time.

The human rights issues of life sentences are also discussed internationally and not only with reference to domestic law (van Zyl Smit 2006). With the establishment of the International Criminal Tribunals for Rwanda and the former Yugoslavia and the International Criminal Court, the question arose, what the heaviest penalty should be that these courts would be allowed to impose (Gumboh 2011). The discussion about life sentences is also relevant with regard to extradition to a country with a different set of rules, i.e., with sanctions for the case at hand that are not acceptable in the requested state (ECtHR, Soering v the United Kingdom, 7 July 1989, application no. 14038/88; R (Wellington) v.

Secretary of State for the Home Department [2008] UKHL 72, [2009] 1 A.C. 335 with a comment by Milanović 2009).

Future Directions

As one possible reaction to allegedly dangerous offenders, long-term imprisonment is a highly political issue. While national politicians support humanization of imprisonment on the somewhat clandestine European level, they might oppose to the same policies on the national level and call for more and longer imprisonment (Drenkhahn et al. 2012, p. 173). This means that research on the effects of long-term imprisonment is also highly political – not least because it may have results that lawmakers do not want to hear.

The number of long-term prisoners in Europe will most likely keep rising and add to prison overcrowding. This is a threat to all efforts of humanizing imprisonment in general, because it further curtails resources which are quite modest anyway due to the economic crisis. But it will also make life more difficult – or at least not easier – for long-term prisoners: for them, too, resources for modern, reintegrative regimes, treatment programs, meaningful pastimes, and humane living conditions will be scarce; it will be difficult to progress through the prison system and to keep reliable relations with family and friends. As Europe is diverse, this development will not occur in all countries and with the same impact. There is a chance for research in this, because cross-national research may not only describe the interactions of prisoners with a certain setting but also compare a range of settings and thus gain deeper insights into the impact of imprisonment and into its place in a society or a culture. Future research should also be comprehensive in the range of aspects it studies: it is not enough anymore to look at either individual characteristics of prisoners or the institutional environment, both have to be taken into account (Haney 2005, p. 86). In addition, future research needs a variety of methods to not only measure what we think we already know but to also identify new and different questions.

Some of these new questions have already been identified for prison research in general (Liebling and Maruna 2005, p. 12 ff.), and they are also valid for research on long-term imprisonment. The situation of particular groups such as young or female prisoners and vulnerable prisoners needs attention. Special groups in long-term imprisonment are aging and old prisoners – life sentences turn prisons into homes for the elderly. Other current issues are mental and physical health and health care in prison. The mental health status of prisoners and risk and protective factors are particularly important, because mental health may impact on relations to other prisoners and staff, it may influence a prisoner's willingness and ability to participate in treatment programs, and it might call for specialized care after release. Imprisonment itself is a risk factor for mental health: it can be a relevant trauma resulting in a posttraumatic stress disorder (PTSD; Dudeck et al. 2011). The whole problem of traumatization – not only by imprisonment – and PTSD in prisoners is still under-researched (Liebling and Maruna 2005, p. 15 f.).

The implementation of restorative justice mechanisms for prisoners and the victims of their crimes is another new task for research and practice. In cases of serious crimes that lead to long prison sentences, it may take a lot of time until the parties involved are ready to communicate openly. Therefore, it is worthwhile to introduce restorative justice in prisons to help victims and offenders come to terms with the crime and its consequences (PFI online).

Related Entries

- ▶ Deterrent Effect of Imprisonment
- ▶ Development of International Criminal Law and Tribunals
- ▶ Extent of Imprisonment: Global View
- ▶ Front Door and Backdoor Sentencing
- ▶ History of Restorative Justice
- ▶ Incapacitation
- ▶ Risk Factors for Prison Recidivism
- ▶ Suicide and Prisons
- ▶ Unintended Effects of Imprisonment

Recommended Reading and References

Appleton C, Grøver B (2007) The pros and cons of life without parole. Brit J Criminol 47:597–615

Chauvenet A, Rostaing C, Orlic F (2008) La violence carcérale en question. Presses Universitaires de France, Paris

CPT (European Committee for the Prevention of Torture and Inhuman or Degrading Treatment or Punishment) (2011) CPT Standards. CPT/Inf/E (2002) 1 – Rev. 2011, http://www.cpt.coe.int/en/docsstandards.htm (August 1, 2012)

Drenkhahn K, Morgenstern C, van Zyl Smit D (2012) What is in a name? Preventive detention in Germany in the shadow of European human rights law. Crim Law Rev 2012:167–187

Drenkhahn K, Dudeck M, Dünkel F (eds) (2013) Long-term imprisonment and human rights. Routledge, London, New York

Dudeck M, Drenkhahn K, Spitzer C, Barnow S, Kopp D, Kuwert P, Freyberger HJ, Dünkel F (2011) Traumatization and mental distress in long-term prisoners in Europe. Punish Soc 13:403–423

Faugeron C, Chauvenet A, Combessie P (eds) (1996) Approches de la prison. DeBoeck Université, Brussels

Goffman E (1961) Asylums. Penguin, London

Griffin D, O'Donnell I (2012) The life sentence and parole. Brit J Criminol 52:611–629

Grounds A (2004) Psychological consequences of wrongful conviction and imprisonment. Cana J Criminol Crim Justice 46:165–182

Gumboh E (2011) The penalty of life imprisonment under international criminal law. African Human Rights Law J 11:75–92

Habermeyer E, Passow D, Vohs K (2010) Is psychopathy elevated among criminal offenders who are under preventive detention pursuant to section 66 of the German criminal code? Behav Sci Law 28:267–276

Haney C (2005) The contextual revolution in psychology and the question of prison effects. In: Liebling A, Maruna S (eds) The effects of imprisonment. Willan, Cullompton, pp 66–93

Irwin J, Owen B (2005) Harm and the contemporary prison. In: Liebling A, Maruna S (eds) The effects of imprisonment. Willan, Cullompton, pp 94–117

Jamieson R, Grounds A (2005) Release and adjustment: perspectives from studies of wrongly convicted and politically motivated prisoners. In: Liebling A, Maruna S (eds) The effects of imprisonment. Willan, Cullompton, pp 33–65

Lappi-Seppälä T (2012) Explaining national differences in the use of imprisonment. In: Snacken S, Dumortier E (eds) Resisting punitiveness in Europe. Routledge, London, New York, pp 35–72

Liebling A (2004) Prisons and their moral performance. Oxford University Press, Oxford, New York

Liebling A, Maruna S (2005) Introduction: the effects of imprisonment revisited. In: Liebling A, Maruna S (eds) The effects of imprisonment. Willan, Cullompton, pp 1–29

Liebling A, Durie L, Stiles A, Tait S (2005) Revisiting prison suicide: the role of fairness and distress. In: Liebling A, Maruna S (eds) The effects of imprisonment. Willan, Cullompton, pp 209–231

Merkel G (2011) Case note – Retrospective preventive detention in Germany: a comment on the ECHR decision Haidn v. Germany of 13 January 2011. German Law J 12:968–977

Milanović M (2009) Extradition and life imprisonment. Cam Law J 68:248–251

Murray J (2005) The effects of imprisonment on families and children of prisoners. In: Liebling A, Maruna S (eds) The effects of imprisonment. Willan, Cullompton, pp 442–462

Padfield N (2010) The sentencing, management and treatment of "dangerous" offenders, PC-CP (2010) 10 rev 4, http://www.coe.int/t/dghl/standardsetting/prisons/PCCP documents 2010/PC-CP OJ 4 Bil – Dec 2010.asp (July 20, 2012)

PFI – Prison Fellowship International Centre for Justice and Reconciliation (no publication date) Prison cell. http://www.restorativejustice.org/prison (August 10, 2012)

Salize HJ, Dreßing H, Kief C (2007) Mentally disordered persons in European prison systems – needs, programmes and outcome (EUPRIS). Zentralinstitut für seelische Gesundheit, Mannheim. http://ec.europa.eu/health/ph_projects/2004/action1/docs/action1_2004_frep_17_en.pdf (August 7, 2012)

Smith PS (2006) The effects of solitary confinement on prison inmates: a brief history and review of the literature. Crime Justice 34:441–528

Snacken S (2005) Forms of violence and regimes in prison: report of research in Belgian prisons. In: Liebling A, Maruna S (eds) The effects of imprisonment. Willan, Cullompton, pp 306–339

Snacken S, van Zyl Smit D (2009) Europäische Standards zu langen Freiheitsstrafen: Aspekte des Strafrechts, der Strafvollzugsforschung und der Menschenrechte. Neue Kriminalpolitik 21:58–68

Sparks R, Bottoms A, Hay W (1996) Prisons and the problem of order. Clarendon, Oxford

Sykes G (1958) The society of captives. Princeton University Press, Princeton

van Zyl Smit D (2006) Life imprisonment: recent issues in national and international law. Int J Law Psychiatr 29:405–421

van Zyl Smit D, Snacken S (2009) Principles of European prison law and policy. Oxford University Press, Oxford, New York

Zamble E, Porporino FJ (1988) Coping, behavior, and adaptation in prison inmates. Springer, New York, Berlin

Evaluating Truthfulness

▶ Evaluating Truthfulness: Interviewing and Credibility Assessment

Evaluating Truthfulness: Interviewing and Credibility Assessment

Barry S. Cooper[1], Hugues F. Hervé[2] and
John C. Yuille[3]
[1]The Forensic Alliance, The University of British Columbia, Simon Fraser University and The Forensic Psychiatric Services Commission, Vancouver, BC, Canada
[2]The Forensic Alliance, Port Moody, BC, Canada
[3]The Forensic Alliance, The University of British Columbia, Salt Spring Island, BC, Canada

Synonyms

Credibility assessment; Evaluating truthfulness; Fact-finding interviews; Investigative interviewing; Lie/truth detection; Reading verbal and nonverbal behavior; Victim, witness, suspect interviewing

Overview

In the criminal justice context, the overarching goal of the investigative interviewer is to build a case. When facing a truth teller, the interviewer aims to gather enough information to validate that truth. When facing a liar, the interviewer's job is to allow the lie to surface in enough detail to be able to prove its falsehood. In each case, the task relies on a combination of effective interviewing, reading people, and credibility assessment. Are professionals adept at these tasks? If not, why? What does the science suggest that individuals should do to accurately evaluate truthfulness or, at least, improve their ability to evaluate truthfulness? In this entry, evidence-based and practical answers to these and other questions are offered. In the following sections, terms are defined and methodological problems affecting this complex area are explored, followed by a review of different approaches to evaluating truthfulness. The importance of proper interviewing when one is evaluating truthfulness is subsequently discussed. Finally, the results of evidence-based training in the area are reviewed followed by suggestions for future directions.

Evaluating Truthfulness: Definitions

Evaluating truthfulness requires distinguishing truths from lies. A lie represents a deliberate attempt to deceive without prior notification (Ekman 2001; Vrij 2008). Lies should not be confused with memory errors or other types of non-intentional inaccuracies, which can occur for a variety of reasons unrelated to deceit. There are different types of lies (e.g., falsification, concealment/omission, the incorrect inference dodge, telling the truth falsely) and different topics of deception (e.g., emotional, opinion, factual, intent). In contrast, a truth is an expression of an honestly held belief or the description of a memory that one believes to be the truth, irrespective of its historical accuracy (e.g., narrative vs. historical truth). Thus, in differentiating truths from lies, understanding the intent of the individual is key; while a liar intends to deceive or manipulate another person, the truth teller does not. The result is that the former may experience unique emotional and/or cognitive reactions that the latter does not experience (Cooper et al. 2009). It is these reactions that scientists are trying to understand and practitioners are attempting to identify.

The Difficulty of Evaluating Truthfulness

Most people lie, often a few times per day, and some people lie more than others. Although most lies are trivial in nature (e.g., "great haircut!"), some have significant personal, social, and legal consequences. Despite the prevalence of deception and the consequences of poor lie detection, most people are not adept at distinguishing truths from lies (Vrij 2008). Indeed, the research, which has typically relied on artificial laboratory paradigms (see below), has generally found that most individuals perform around the level of chance (Bond and DePaulo 2006), irrespective of the nature of their profession or their years of experience in their profession (Ekman and O'Sullivan 1991).

Methodological Problems

Although there is a large literature on evaluating truthfulness, it is marred by a number of problems that impact its generalizability to real-world settings. A major problem with this research is that it has been conducted predominately in the controlled setting of the laboratory. It is argued that, by relying almost exclusively on the laboratory, researchers have committed the offence of methodolotry (see Yuille 2013). Researchers' strong belief in the utility of controlled research has led them to rely on laboratory analogues to study truthfulness and deceit. More weight has been placed on methodological concerns than on issues concerning generalizability and applicability. In the modal experiment on deception, undergraduate research participants tell the truth or lie about some activity or opinion. The motives to fool others are usually weak (e.g., a small monetary incentive or course credit) and the consequences of being caught in such low-stakes lies have no significant personal or social consequences. The end result is that more is known about how to trigger effects using laboratory designs in undergraduate students instructed to lie under low consequence paradigms than how real-world deception and its detection takes place.

Another methodological problem is the overreliance on group designs. While group designs that compare truth tellers and liars meet stringent research requirements, the practice of evaluating truthfulness focuses on one individual, typically in the context of an interview, and therefore necessitates a single-subject design for analysis. Furthermore, the field has also been hampered by an overreliance on simplistic statistical procedures, particularly when comparing group averages (i.e., groups of liars are compared to groups of truth tellers) and a small number of indices (i.e., one or two behaviors that may differentiate truth tellers from liars). Using group averages washes out important individual differences in how people reveal their truths and lies – differences that are arguably more important in the real-world practice of evaluating truthfulness (Cooper et al. 2009). For example, if two individuals display their deception in contrasting ways (e.g., one increases and the other decreases a particular behavioral pattern while lying), the average finding would be that the behavior under investigation is unrelated to deception detection. Similarly, looking at only one or two behaviors may result in null findings simply because there are no universal signs of lying and truth telling (see below). That is, the particular behavior(s) may distinguish truths from lies in some participants and not in others. The reduced diagnostic specificity of the behavior(s) may then result in misleading null findings. Consequently, the science and practice of evaluating truthfulness would benefit greatly from more sophisticated methodological and statistical approaches that focus on within-subject methodologies and multivariate analyses.

It is not that laboratory research on evaluating truthfulness is unimportant; indeed, it is important to study how "most people" behave in contexts in which variables can be controlled and manipulated. However, it is also essential to know how to evaluate truthfulness in a particular person in a specified context when the stakes of lying are high (e.g., when the issues are of personal relevance and the consequences are of significance). For this reason, recent efforts have been made to conduct ecologically valid studies, including those examining high-stakes lies (Mann et al. 2002). Clearly, both laboratory and field research on evaluating truthfulness must be conducted to provide converging evidence of the factors revealing truthfulness and deception.

Why Lies Succeed

The current methodological problems of the field notwithstanding, research and practice suggest that lies succeed for a variety of reasons. For example, most lies are difficult to detect as they are often embedded in considerable truth, and some liars are simply too skilled to be detected. Errors in evaluating truthfulness are also made due to faulty assumptions held by the targets of lies (e.g., investigative interviewers), many of which are supported by common myths or through non-evidence-based training. The following is a description of some of the known errors that lead to poor deception detection (for more detail about these and other pitfalls, see Vrij et al. 2010). The most consistent error results from the widely held belief that there is a universal sign or signal diagnostic of deception (i.e., sometimes labelled the *Pinocchio error*; Ekman 2001). Decades of research in the area of evaluating truthfulness, however, have made it clear that there is no universal sign of lying that is displayed by all individuals in all contexts (DePaulo et al. 2003). For example, there has been no empirical support for the proposition offered by proponents of Neuro-Linguistic Programming (NLP) that eye movements are valid indicators of deception (e.g., Wiseman et al. 2012). The misattribution of the emotional state of others – the *Othello error* – is another source of error when evaluating truthfulness (Ekman 2001; e.g., misinterpreting fear of being disbelieved as detection apprehension). When an emotion is displayed, it only signals that the emotion is felt, not the reason for it. Only through effective interviewing can the cause of the emotion be uncovered (Yarbrough et al. 2013). Another error is when the recipient of the lie has little motivation to catch the liar (i.e., sometimes labelled the *ostrich effect*; Vrij 2008). Collusion would be an example of this effect (e.g., when the target wants to believe the liar).

Another error is the *Lie to Me error*, coined after the popular *Fox* television show, *Lie to Me*, wherein the lead character, played by Tim Roth, is a professed expert lie detector. There are several problems with the approach to evaluating truthfulness depicted in this show. First, Tim Roth's character presents with an enormous ego associated with overconfidence in his abilities – both of which are known to result in errors in differentiating truths from lies. Second, although he keenly and accurately detects changes in others' behavior – and while detecting change is indeed important to evaluating truthfulness (see below) – he comes to an automatic conclusion of the reason for the behavioral change without considering alternative hypotheses for the observed change (e.g., which also relates to the *Othello error*) or individual differences in behavior (note: the latter has also been called the *idiosyncrasy error*; Ekman 2001). A failure to consider multiple hypotheses for what is observed and heard coupled with impulsive decision making increases the chance of errors. Third, the fictional character believes that mere passive observation is a valid avenue to evaluate truthfulness accurately, shunning the notion that interviewing is important. As discussed below, however, evaluating truthfulness is heavily dependent on effective interviewing. The *Lie to Me error* concerns, in part, a combination of other errors and arguably represents a higher-order error. This error may be the most difficult to extinguish as, at some level, it involves a lack of core critical thinking skills.

Approaches to Evaluating Truthfulness

Currently, a variety of approaches to evaluating truthfulness have gained attention and/or empirical support. These approaches rely either on psychophysiological measures or the use of observational techniques. Each is discussed in turn.

Psychophysiological Techniques

A number of technologies based on psychophysiological assessments have been applied to the area of evaluating truthfulness. The polygraph is one of the most long-standing psychophysiological tools used to evaluate truthfulness in the modern era, and it has been the subject of an extensive amount of research (see Griesel and Yuille 2007). There are a number of different ways in which the polygraph is used, including the Directed Lie Test, the Control Question Test, the Guilty Knowledge/Concealed Knowledge Test, and the Relevant-Irrelevant Test. The polygraph is not a “lie detector,” although it has been referred to as such, often in the popular media. Rather, the polygraph detects changes in physiological arousal through measurements of heart rate, galvanic skin conductance, respiration, etc. Such changes may be due to lying but may also be due to other factors (e.g., stress, fear of being disbelieved, increased cognitive load). There are a number of concerns about the polygraph’s false-positive and false-negative rates, which, in most countries, precludes its use in contexts such as employee screening (National Research Council 2003). Similarly, concerns about the validity and reliability of the polygraph have also led to restrictions on its use and applicability in relation to matters of the criminal justice system. Arguably the strongest facet of the polygraph is the administrator. Anecdotally, effective polygraphers are excellent interviewers and critical thinkers – they treat the physiological results of the polygraph as simply one piece of data in the larger puzzle.

Recently, some companies have developed voice stress analyzers, which detect changes in verbal characteristics (e.g., pitch and tone of voice). Although these companies have marketed these tools as highly effective lie detectors, the research has shown these marketing claims to be false (Damphousse 2008). Although detecting change in behavior (e.g., via vocal clues) is essential to evaluating truthfulness (see below), such changes may be due to lying or to a host of other reasons (Cooper et al. 2009). The reason for the change can only be uncovered via testing alternative hypotheses through effective interviewing.

There has also been recent interest in the use of thermal imaging technologies, which detect body heat changes, in the evaluation of truthfulness. To date, there has been no strong support for thermal imaging as a valid method of evaluating truthfulness (Warmelink et al. 2011); however, the research has been sparse. Electroencephalography (EEG), a measure of neural activity, typically via event-related potentials (ERPs), has also been used in the study of evaluating truthfulness and also with limited success (Rosenfeld 2002). The past decade or so has seen greater efforts to identify the neural processes involved in deception via functional magnetic resonances imaging (fMRI) technology. Although gains have been made in this field, there is no empirical foundation for the notion that fMRI can be used to reliably evaluate truthfulness (Spence 2008). Indeed, although certain areas of the brain (i.e., within the frontal and parietal lobes) may be more active than others when deception occurs, research indicates that fMRI technology does not reveal neural processes in these areas that are unique to deceit (Monteleone et al. 2009).

There is a pattern that emerges from an examination of the various psychophysiological techniques proposed to evaluate truthfulness – as far as is currently known, there is no psychophysiological response that is unique to deception in all individuals and in all contexts. A failure to understand this fact leads individuals to commit the *Pinocchio error* (see above). Figuratively and, in some cases, literally, most of the psychophysiological measures detect *hot spots* (see below) – a change from baseline and/or inconsistencies across channels measured (Cooper et al. 2009) – not lies. Although detecting hot spots is important, such change is not diagnostic of lying. Evaluating truthfulness is a two-step process: first the observer needs to detect behavioral changes and/or inconsistencies; second, the observer needs to determine the reason the behavior occurred.

Observational Techniques: Reading Verbal and Nonverbal Behavior

Observational techniques for evaluating truthfulness rely on reading verbal and nonverbal behavior associated with truth telling and lying. Relative to psychophysiological technologies, these techniques – not dependent on any equipment – are low cost, transportable, generalizable, and noninvasive. There are two clear “camps” of researchers in the observational study of evaluating truthfulness: those in the cognitive camp (i.e., primarily examining verbal behavior) and those in the emotional camp (i.e., primarily examining nonverbal behavior). Most researchers in the cognitive camp ascribe to the theoretical assumption that, all factors considered equal, lying is more cognitively demanding than telling the truth because the liar has to fabricate details that do not exist, remember his/her lie, suppress information, evaluate the success of the lie, adapt accordingly, etc., while the truth teller simply has to recall and report details from memory (DePaulo et al. 2003; Vrij et al. 2010). Researchers from this camp believe that the increase in cognitive load (i.e., mental effort) associated with deception leads to behavioral changes that betray the liar, such as changes or particularities in verbal content, verbal style, and body language.

Most researchers in the emotional camp believe that lying may result in an emotional response (e.g., fear – detection apprehension; happiness – duping delight) and, as a consequence, emotional "leakage" – that is, the display of an emotion that may leak out and thus betray a lie (Ekman 2003). Micro and subtle behavior are viewed as especially important, with the former being a full behavioral display (e.g., a full two-sided shrug or a complete display of anger in the face) that occurs for only a fraction of a second and the latter being a partial behavioral display (e.g., a shrug on only one side; anger displayed only in the eyes) that can last a fraction of a second or longer. These types of behaviors are involuntary and therefore provide messages that the individual is trying to hide, be it to themselves or the interviewer. This is in contrast to macro behaviors, which are long-lasting full displays of behaviors that are more under control (e.g., can be faked by a deceptive person). To date, most of the research has focused on displays of emotional leakage in the face, particularly micro and subtle expressions, and changes in body language and vocal characteristics. This research focuses on identifying both changes in behavior, as change does not occur randomly, and inconsistencies between the various behavioral channels, such as claiming to love someone but showing disgust or contempt when talking about them.

In recent years, researchers from both camps have started to believe that their respective viewpoints and lines of research are complementary rather than contradictory. Most researchers and practitioners now agree that telling the truth or a lie may result in emotional and/or cognitive leakage, which itself may lead to changes in verbal and/or nonverbal behavior, and that the ability to trigger, identify, and make sense of these behavioral changes relies on effective interviewing strategies (Vrij et al. 2010). An integrated stepwise method to evaluating truthfulness which incorporates the evidence-based views of both camps and stresses the importance of effective interviewing during the collection of information has recently been provided (i.e., Cooper et al. 2009). First, according to this method, the interviewer must assess the interviewee's baseline behavior (i.e., the individual's "normal" way of behaving when not lying; e.g., via prior recordings, via the rapport stage of an investigative interview). Second, the interviewer must actively listen and observe for changes from baseline within a variety of observable behavioral channels (i.e., face, body, voice, verbal style, and verbal content; see Fig. 1), as well as inconsistencies across these channels. As depicted in Fig. 1, these changes and/or inconsistencies reflect cognitive and/or emotional load that has leaked out in behavior. This leakage should not be impulsively interpreted as a sign of deception given that there is no *Pinocchio* response. Rather, observed leakage should be viewed as a hot spot, that is, a behavior of importance that may be due to a number of causes (e.g., related to emotional or cognitive leakage) of which lying is only one possibility. In other words, the interviewer should consider multiple hypotheses for what she/he sees or hears. As shown in Fig. 1, the interviewer should also focus equally on signs of truth telling and lying, as focusing on either in isolation predisposes one to make the error of believing the liar or disbelieving the truth teller, respectively.

Finally, once a hot spot has been identified, the interviewer must test his/her hypotheses for the hot spot. This is accomplished by further exploring the topic that triggered the hot spot via effective interviewing (see below) and/or by employing specific interviewing techniques designed to distinguish truth tellers from liars.

Evaluating Truthfulness: Interviewing and Credibility Assessment, Fig. 1 Model for evaluating truthfulness (From Cooper et al. 2009)

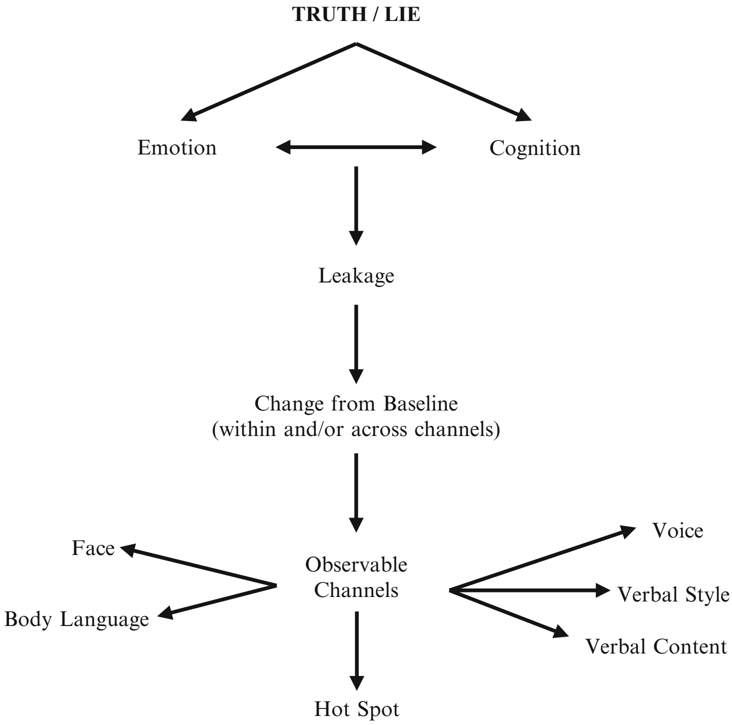

For example, the interviewer, after noting a hot spot, should allow the interviewee to continue with his/her statement. Once the statement is complete, the interviewer can return to the topic that led to the initial hot spot. If another hot spot arises, it provides more evidence that the hot spot was related to the topic as opposed to another reason. This final step is crucial as it permits the interview to rule out hot spots unrelated to lying, thereby allowing the truth teller to demonstrate his/her credibility; or it could result in further hot spots (and therefore evidence) in the liar that betrays his/her deceptiveness. An essential feature of this approach is that the interviewer must delay judgement until enough information has been collected and alternative hypotheses have been ruled out through effective interviewing. Prior to outlining the basics of effective interviewing, each behavioral channel noted above is briefly reviewed.

Face

Although a number of behavioral channels can display emotional leakage, the face is the clearest channel for this type of leakage; it is also the most researched behavioral channel concerning emotional leakage (Ekman 2001; Porter et al. 2012). As noted above, although identifying macro facial expressions can be informative (e.g., happiness, sadness, fear, surprise, anger, contempt), the detection of micro and subtle facial expressions is most revealing as these often signal information that the interviewee is trying to hide. This is especially true when they contradict what is said. For example, if a husband suspected of a domestic homicide claims to have had a loving relationship with his wife but concurrently shows a subtle expression of anger, this is an important hot spot that needs to be further investigated.

In addition to signalling emotional reactions, the face is a potential source of information about cognition (e.g., the pursing of the lips typically signals a mental search for information). For example, certain facial movements (e.g., the lowering of the eyebrows) are signs of concentration and, therefore, reflect increased cognitive load. If an interviewee shows these signs, the reason for the behavioral change may be obvious (e.g., responding to a relatively difficult question, attempting to recall a distant memory) or it may be a clue that an interviewee may be thinking too hard relative to the level of difficulty of the question (e.g., when asked his/her name or birthday), thereby suggesting that one hypothesis to consider is that the individual is being deceptive. Interviewers must entertain multiple hypotheses for the observed behavioral change in the face (e.g., the lowering of the eyebrows could be related to concentration or subtle anger; Ekman 2003). If the observed behavioral change is out of context, it is a hot spot (Cooper et al. 2009).

Body Language

Body movement or change is also an important behavioral channel to read when evaluating truthfulness. Indeed, certain activities (e.g., sweating, fidgeting, finger tapping) may provide clues to one's emotional state (e.g., stress, nervousness). There are also gestures that individuals display that are meaningful clues to other emotions or cognitions, particularly when viewed as a change from baseline. For example, the literature has identified three different types of gestures that are important to consider when reading people and evaluating truthfulness: emblems, manipulators, and illustrators (e.g., Friesen et al. 1979). Emblems are literally a type of body language (e.g., nodding one's head "yes") and are completely culturally determined. For example, the "thumbs up" sign has an entirely different meaning in North America than it does in some contexts in the Middle East (i.e., it is viewed as a rude gesture). Knowledge of the specific emblems of a culture is required to understand their meaning. The display of emblems is usually voluntary but emblematic slips may involuntarily leak information (e.g., when someone shrugs their shoulders – a sign of uncertainty – while attempting to convey a verbal message of certainty) – emblematic slips are thus, by definition, hot spots.

Manipulators, also referred to as self-manipulations, refer to the movement of one body part on another (e.g., scratching one's nose or ear, playing with one's hair; there are self-manipulators and object manipulators). Some research has shown that deception is associated with an increased use of manipulators (Porter et al. 2008). When evaluating truthfulness in practice, it is important to pay attention to individuals' baseline display of manipulators as a change from baseline is considered a hot spot (Cooper et al. 2009). Some individuals increase their use of manipulators when under increased emotional arousal, yet others may show an increase in manipulators when relaxed (e.g., an increase in grooming behavior).

Illustrators are hand movements used to illustrate speech as it is being spoken. Although some research (see DePaulo et al. 2003) suggests that individuals tend to decrease their use illustrators when being deceptive (e.g., as such may signify an increase in cognitive load), it is important to consider a person's baseline. Some people show an increase in illustrators when their cognitive load increases; others show a decrease. A change from baseline (i.e., an increase or decrease) would signify a hot spot. It is the interviewer's task to figure out the reason for the change.

Voice

The voice is another behavioral channel in which leakage can be observed as assessed via pitch, tone, loudness, etc. For example, a vocal change could be a cue to individuals' emotional state with the voice tending to have an edge when anger is experienced and to be higher in pitch when fear or anger is felt (Ekman 2003). It has also been shown that peoples' voices tend to soften when being deceptive; however, it has also been demonstrated that the voice tends to soften when people are uncertain or feeling sad. Thus, as with the other behavioral channels, a multiple hypothesis testing approach must be used to delineate the reason for the vocal change. The change is meaningful but should not be interpreted as a lie per se, as the companies that market voice stress analyzers suggest. For example, if an interviewee softens his voice relative to his baseline while being interviewed during an assessment of his suicide risk, the softening could suggest sadness (e.g., a potential risk factor for suicide), deception, or some other factor. Only through effective interviewing could the reason for the expressed vocal change be uncovered.

Verbal Style

Verbal style is an aspect of language that can leak both emotional and cognitive load. For example, relative to an individual's baseline verbal style, an increase in pauses or filled pauses could suggest an increase in cognitive load (Mann et al. 2002). The same would be the case for word repetitions, as well as an increase in stuttering or the use of jargon. It is the change in verbal style – relative to baseline – that is important, as opposed to the verbal style per se. For example, if an interviewee's baseline verbal style is the use of jargon (e.g., "you know what I mean," "you know") and the interviewee uses such jargon throughout an interview, the displayed jargon is relatively meaningless in terms of evaluating truthfulness. If, however, an interviewee does not typically use such jargon but does in relation to answering critical questions during an investigative interview, the change in verbal style should be considered a hot spot. Pronoun usage is another aspect of verbal style that has been applied to the area of evaluating truthfulness. For example, it has been empirically demonstrated that, all other factors considered equal, individuals tend to use less first person pronouns (e.g., I, my, me) when lying than when telling the truth, arguably in an attempt to distance themselves from their lies. As with the other behavioral channels, it is the change in pronoun usage – relative to baseline – that is important, not the rate of pronoun usage per se.

Verbal Content

A considerable amount of research has examined verbal content as a behavioral channel in which to evaluate truthfulness. A number of techniques have been developed to analyze the quality and quantity of verbal content including Scientific Content Analysis (SCAN; Sapir 1987, 2000), Reality Monitoring (RM; Johnson and Raye 1981), and Criteria-Based Content Analysis (CBCA; Yuille 1988). SCAN is a widely used technique developed by a former Israeli polygrapher. It involves eliciting a written statement about an event (e.g., an alleged crime) followed by an analysis of the statement with the use of certain criteria (e.g., out of sequence, social introductions). Anecdotal evidence of its utility notwithstanding, there has been no empirical support for SCAN (Nahari et al. 2012). Moreover, certain assumptions underlying SCAN are theoretically inconsistent with known memory/psychological processes.

RM is based on the theoretical assumption that memories of experienced events will contain more contextual information as well as more external-sensory information than memories that are the product of fabrication/imagination (Johnson and Raye 1981). The RM technique involves an interviewee rating their own memory (e.g., for vividness, detail) followed by an external rater who follows the same procedure after reading the interviewee's transcript. While RM is more firmly grounded in science than SCAN and there has been some evidence attesting to its utility, it is not without its problems, particularly when applied to the forensic arena (Colwell et al. 2013).

CBCA is one of the most researched techniques applied to the evaluation of the credibility of verbal content (Vrij 2005). It is one part of a larger system – Statement Validity Analysis (Yuille 1988). Of note, it is not possible to apply CBCA to an interview (or a transcript of an interview) unless the interview is of high quality (e.g., a free narrative is required for analysis). In some respects, the theoretical assumptions underlying CBCA are similar to those of RM. CBCA is based on the Undeutsch hypothesis that posits that memories of actual experiences differ in quantity and quality from memories based on invention (Undeutsch 1989). A number of versions of CBCA are in existence but most contain at least 19 criteria that are applied to an interview /statement/transcript. Depending on the version, the criteria are divided into the following five areas which are applied to a statement by a trained rater: general characteristics (i.e., coherence, spontaneous reproduction, appropriate amount of detail), specific content (e.g., contextual embedding, descriptions of interactions), peculiarities of the content (e.g., unusual details, related external associations), motivational features (e.g., spontaneous corrections, self-deprecation), and stylistic features (e.g., theme-related changes, rigid repetition). The use of CBCA has been empirically shown to be able to differentiate deceptive from truthful statements (i.e., concerning memories of past events) at a rate ranging from 55 % to 90 % in samples of children, adult witnesses, and adult offenders.

There are a number of misperceptions about CBCA that have implications for its applied utility. For example, CBCA is a complex qualitative procedure, not a quantitative tool (see Griesel et al. 2013). Indeed, it is not a matter of simply adding up the criteria present in a statement to make a judgement about its credibility as some research would suggest. Rather, as certain criteria arguably should have more weight than others (e.g., appropriate amount of detail vs. unusual details), the trained rater should apply CBCA in a qualitative manner. Further, training matters – in order to apply the procedure effectively, a considerable amount of training is required. This would include acquiring knowledge about the factors that impact memory (see Hervé et al. 2013) in order to know what is considered an appropriate amount of detail in a statement.

Summary of Observational Techniques

Each of the five behavioral channels can provide useful information for evaluating truthfulness. The channels must be considered together and not in isolation. Often, the most telling hot spots occur when there is an inconsistency between behavioral channels (e.g., when an interviewee says “no” but nods his head “yes”). These and other hot spots allow the interviewer to navigate the interview to identify areas in need of further enquiry. The interviewer’s job is to collect sufficient information to validate the credibility of the truthful person or elicit sufficient information to reveal the lie in the deceptive individual. Only when an interview provides a high quality and quantity of details can such tools as CBCA be applied. As no tool or technique is foolproof, it is imperative that the investigative interviewer corroborate his/her conclusion (e.g., verify alibis/whereabouts).

The Importance of Effective Interviewing When Evaluating Truthfulness

The importance of an effective interview when evaluating truthfulness cannot be overstated. In practice, lies are embedded in a great deal of truths and the ability to differentiate what is truthful from fictional relies heavily on effective interviewing. A poor interview only serves to further blur the distinction between truth and deception. The overarching goal of an effective interview is to gather uncontaminated information, that is, the interviewee’s version of events. In other words, the interviewer’s task is to cue memory, not lead memory. Leading questions run the risk of causing the interviewee to unintentionally provide inaccurate information, which could be misinterpreted as a sign of deception. Leading or suggestive questions can also unintentionally telegraph the interviewer’s goal or suspicions, thereby allowing deceivers to alter their deceptive tactic(s). In contrast, non-leading, open-ended questions allow the interviewee to provide what she/he knows in a spontaneous manner. Spontaneity is one of the most potent signs of truthfulness (Colwell et al. 2013). In order to accomplish the goal of eliciting uncontaminated information, it is suggested that interviewers adopt a memory-based approach to interviewing (Yarbrough et al. 2013), such as the Step-Wise Interview Guidelines (Yuille et al. 2009), the Cognitive Interview (Fisher and Geiselman 1992), or the PEACE model of interviewing (Milne and Bull 1999), all of which have the goal of cuing and exhausting memory without contamination.

A memory-based approach is crucial given that the typical investigative interview focuses on an interviewee’s memory about some past event. Accordingly, interviewers should be knowledgeable about how memory works, what factors impact memory, etc., in order to effectively navigate an interview and accurately evaluate the truthfulness of the memory received (see Hervé et al. 2013 for a review of memory patterns and of the biopsychosocial predisposing, precipitating, and perpetuating factors influencing memory). That is, the more one knows about memory, the easier it is to elicit it and evaluate its credibility. Indeed, the quality and quantity of details contained in a reported memory are key determinants to its credibility, and the "appropriateness" of these details in a particular case depends on the interviewees' predisposing, precipitating, and perpetuating memory factors.

While there are different memory-based interview models, they all share the following features. The first feature of an effective interview is preparation. This involves, among other factors, gathering as much background information about the interviewee and the precipitating event (s) as possible. The more the interviewer knows about the interviewee and the subject under investigation, the better she/he is equipped to develop rapport, assess baseline, identify and assess hot spots, entertain alternative hypothesis, and exhaust memory. For example, preparation may reveal culturally specific information that may help in the interpretation of body language, while knowledge of relevant crime patterns and the evidence in the case at hand may reveal inconsistencies between this information and the interviewee's statement. The second feature is rapport building. During the initial stages of an interview, the interviewer should develop a working relationship with the interviewee that assists in placing the interviewee at ease and encourages dialogue. During this part of the interview, the interviewer should assess the interviewee's baseline (a process that continues throughout the interview). The better the rapport, the clearer the baseline and the easier it is to detect deviations from baseline (or hot spots). For example, if the interviewer projects suspiciousness or dislike of the interviewee, this could cause stress or anxiety in the interviewee. This emotional load will leak out in behavior which could either be misinterpreted as a sign of lying or serve to mask leakage associated with deception to critical questions (i.e., the signal becomes lost in the noise). This is why it is recommended that interviewers adopt, at least initially, an information-gathering style of asking questions (i.e., as opposed to an accusatory style of interviewing – challenges are best left to the later stages of interviews).

The third feature of an effective interview is that it allows the interviewee every opportunity to provide his/her version of events. As noted above, this involves cuing the interviewee's memory via non-leading, open-ended questions. This typically involves a funnel approach to interviewing wherein the interviewer attempts to elicit an uninterrupted free narrative followed by open-ended W-H questions before asking specific questions as deemed necessary. Interruptions are avoided as they disrupt the reconstructive nature of memory. This also often serves to reduce the amount of questions being asked; the fewer the questions needed to exhaust memory, the better the quality of the information provided. Indeed, research has demonstrated that interviewees provide more correct information about past events during the free narrative aspect of interviewing than in response to specific questions. Further, a free narrative is the ideal aspect of an interview for the application of certain verbal credibility assessment techniques such as CBCA.

In addition to sharing the aforementioned features, memory-based interviewing approaches also include various strategies (e.g., narrative repetition, different perspective, backward recall) for enhancing memory when the provided memory appears to lack an appropriate amount of detail. There are many potential reasons why someone may provide a poorly detailed memory (see Hervé et al. 2013), some of which are valid (e.g., time-related memory decay) while others may be deceptive (e.g., making up an event to conceal the truth). Memory-enhancing techniques rely on the reconstructive and cued nature of memory, which dictate that one's narrative should appear somewhat different from one recall to the next (e.g., a few details added and/or a few details omitted), that additional details can be elicited given the right cue, and that the manner in which the information is recalled can be used to cue additional details (e.g., field vs. observer perspective; forward vs. backward recall). These memory-enhancing strategies are therefore important tools for an interviewer in his/her attempt to clarify the possible reason(s) for the observed lack of detail (i.e., test alternative hypotheses). On the one hand, the truthful person – having a real memory to rely on – should provide additional information when asked, for example, to repeat his/her narrative (irrespective of the perspective or direction of recall; note: this information could serve to validate his/her memory and/or provide additional case-relevant details). On the other hand, the liar who fabricated a story and therefore has no memory to rely on (unless the lie is based on a true experience) is predisposed to provide either a rigid repetition and/or a poverty of additional details, both of which are notable hot spots. Moreover, some of these techniques require additional cognitive resources, which liars find difficult to manage. For example, all other factors considered equal, it is easier for an interviewee to recall an event in chronological order in comparison to backward order. Research has found that the increase in cognitive load is more pronounced for liars than truth tellers and that the resulting leakage facilitates the task of differentiating truths from lies (Vrij et al. 2008). Recent research efforts have focused on defining additional interview strategies that create greater cognitive load in truth tellers than in liars, including asking unanticipated questions and the strategic disclosure of evidence by interviewers (Jordan et al. 2012; Vrij et al. 2009). Regardless of what interviewing strategies are used, the resulting leakage should be viewed as a hot spot. Multiple hypotheses must be entertained for the reason for the elicited hot spot, of which deception is only one possibility.

Evidence-Based Training on Evaluating Truthfulness

As indicated above, without any training, it has been shown that the average person evaluates truthfulness at the level of chance (Bond and DePaulo 2006; Ekman and O'Sullivan 1991; Vrij 2008). A number of researchers/organizations have provided evidence-based training (i.e., based on the principles of reading verbal and nonverbal behavior discussed above) to various professional groups (e.g., law enforcement, corrections) in the general area of detecting deceit/evaluating truthfulness. However, there exist only a few studies examining the effectiveness of such training. In general, these studies show that 1- to 2-day workshops can improve participants from about chance level at pretest to an accuracy rate of 77 % or higher at posttest (e.g., Porter et al. 2000; Colwell et al. 2009). Training gains, although smaller in magnitude, have also been demonstrated after 3-hour workshops (e.g., Colwell et al. 2012: 58 % at pretest to 72 % at posttest; Porter et al. 2010: 51 % at pretest to 61 % at posttest). While these findings are encouraging, it remains to be seen whether the gains made in these workshop generalize to the real world. Clearly, future research on the effectiveness and generalizability of evidence-based training is required.

The gains that individuals make in training notwithstanding, it is important to acknowledge that the field will never advance to the point that training will provide the knowledge and skills to allow judgements to be 100 % accurate. Evaluating truthfulness is far too complex of an enterprise. It is more realistic to be able to improve most individuals' abilities to evaluate truthfulness from chance level to some reliable level above chance. The goal should not be to produce *wizards* – rare individuals who appear to be able to reliably evaluate truthfulness at an exceptionally high rate (O'Sullivan 2013). That said, the continued study of these types of individuals will hopefully lead to new insights into the personality characteristics of effective evaluators, the process of differentiating truths from lies, and key features to consider.

Summary and Future Directions

This entry provided an overview of the nature and complexity of evaluating truthfulness. Common pitfalls and approaches were described, and the interaction between effective interviewing, reading verbal and nonverbal behavior, and assessing credibility was reviewed. It was shown that individuals can learn to improve their ability to evaluate truthfulness through evidenced-based training. Despite the gains made in the field through decades of social science research, more research is clearly needed, particularly research that uses single-subject designs in real-world settings and takes into consideration the perspectives of practitioners who evaluate truthfulness as a core aspect of their professional duties. With a single-subject design, an individual's baseline could be assessed and changes from baseline could be examined within and across all of the observable behavioral channels (i.e., face, body language, voice, verbal style, and verbal content). In this regard, more attention should be devoted to individual differences (e.g., personality, personality pathology, mental illness, cognitive functioning, emotional intelligence, culture) as these differences can affect each aspect of the process of evaluating truthfulness reviewed above.

It is recommended that high-stakes lies be examined in a variety of contexts. For example, a promising area of research in evaluating truthfulness concerns the study of computer (and related)-mediated communication (CMC; Hancock and Woodworth 2013). Future research should also examine the ability of individuals to evaluate truthfulness in the context of interviewing. Variables such as the length of the interview, the focus of the interview (e.g., memory for past vs. future actions), and the nature of the interview (e.g., overt vs. covert) should be examined in the laboratory and in the field (e.g., low- vs. high-stakes lies). As well, research should identify the type of individuals best geared to evaluate truthfulness – no doubt the research on wizards holds promise in this regard. Not everyone is suited to be an investigative interviewer and/or able to effectively evaluate truthfulness. Key characteristics are likely to include an information-gathering mind-set, critical thinking skills, and the capacity to develop rapport, read people, and make decisions about truthfulness holistically.

Finally, research attention should be devoted to deciphering the most effective way to deliver evidence-based training to ensure that the training results generalize to on-the-job performance. Training should focus on investigative interviewing as a foundation (e.g., how to gather information – first level training), followed by modules on reading people (e.g., identifying and probing for hot spots – second level training), and evaluating truthfulness (e.g., making sense of hot spots via hypothesis testing and verbal content analysis – third level training). Actual cases/recorded interviews should be used for demonstration and practice as part of the training. Ideally, the training would include a refresher component – which has shown promise in the interviewing training literature (Price and Roberts 2011). This could include literature updates and practice cases to avoid drift in training gains. Finally, post-training on-the-job performance should be assessed in the context of supervision/mentoring and the review of actual recorded interviews in which ground truth has been determined.

Related Entries

- ▶ Behavioral Investigative Advice
- ▶ British Police
- ▶ Criminal Investigative Analysis
- ▶ Detecting Deception with fMRI
- ▶ Evidence-Based Policing
- ▶ False Confessions and Police Interrogation
- ▶ False Memories
- ▶ Interview and Interrogation Methods Effects on Confession Accuracy
- ▶ Interviewing Eyewitnesses

Recommended Reading and References

Bond CF, DePaulo BM (2006) Accuracy of deception judgements. Personal Soc Psychol Rev 10:214–234

Colwell K, Hiscock-Anisman C, Memon A, Colwell LH, Taylor L, Woods D (2009) Training in assessment criteria indicative of deception (ACID) to improve credibility assessment. J Forensic Psychol Pract 9:199–207

Colwell LH, Colwell K, Hiscock-Anisman C, Hartwig M, Cole L, Werdin K, Youschak K (2012) Teaching professionals to detect deception: the efficacy of a brief training workshop. J Forensic Psychol Pract 12:68–80

Colwell K, Hiscock-Anisman C, Jacquelyn Fede J (2013) Assessment criteria indicative of deception (ACID): an example of the new paradigm of differential recall enhancement. In Cooper BS,Griesel D, Ternes M (eds) Applied issues in investigative interviewing, eyewitness memory, and credibility assessment. Springer, New York, pp 259–291

Cooper BS, Hervé H, Yuille JC (2009) Evaluating truthfulness: detecting truths and lies in forensic contexts. In: Bull R, Valentine T, Williamson T (eds) Handbook of psychology of investigative interviewing. Wiley-Blackwell, Chichester, pp 301–328

Damphousse KR (2008) Voice stress analysis: only 15 percent of lies about drug use detected in a field test. NIJ J 258:8–12

DePaulo BM, Lindsay JL, Malone BE, Muhlenbruck L, Charlton K, Cooper H (2003) Clues to deception. Psychol Bull 129:74–118

Ekman P (2001) Telling lies: clues to deceit in the marketplace, politics, and marriage. W.W. Norton & Co., New York

Ekman P (2003) Emotions revealed: recognizing faces and feeling to improve communication and emotional life. Henry Holt and Company, New York

Ekman P, O'Sullivan M (1991) Who can catch a liar? Am Psychol 46:913–920

Fisher RP, Geiselman RE (1992) Memory-enhancing techniques for investigative interviewing: the cognitive interview. Charles C. Thomas, Springfield

Friesen WV, Ekman P, Wallbott H (1979) Measuring hand movements. J Nonverbal Behav 4(2):97–112

Griesel D, Ternes M, Schraml D, Cooper BS, Yuille JC (2013) The ABC's of CBCA: verbal credibility assessment in practice. In Cooper BS, Griesel D, Ternes M (eds) Applied issues in investigative interviewing, eyewitness memory, and credibility assessment. Springer, New York, pp 293–323

Griesel D, Yuille JC (2007) Credibility assessment in eyewitness memory. In: Toglia MP, Read JD, Ross DF, Lindsay RCL (eds) Handbook of eyewitness psychology. Volume I, memory for events. Lawrence Erlbaum, Mahwah, pp 339–370

Hancock JT, Woodworth M (2013) An "Eye" for an "I": the challenges and opportunities for spotting credibility in a digital world. In Cooper BS, Griesel D, Ternes M (eds) Applied issues in investigative interviewing, eyewitness memory, and credibility assessment. Springer, New York, pp 325–340

Hervé HF, Cooper BS, Yuille JC (2013) Biopsychosocial perspectives on memory variability in eyewitnesses. In Cooper BS, Griesel D, Ternes M (eds) Applied issues in investigative interviewing, eyewitness memory, and credibility assessment. Springer, New York, pp 99–142

Johnson MK, Raye CL (1981) Reality monitoring. Psychol Bull 88:67–85

Jordan S, Hartwig M, Wallace B, Dawson E, Xhihani A (2012) Early versus late disclosure of evidence: effects on verbal clues to deception, confessions, and lie catcher' accuracy. J Investig Psychol Offender Profiling 9:1–12

Mann S, Vrij A, Bull R (2002) Suspects, lies and videotape: an analysis of authentic high-stake liars. Law Hum Behav 26:365–372

Milne R, Bull R (1999) Investigative interviewing: psychology and practice. Wiley, Chichester

Monteleone GT, Phan L, Nusbaum HC, Fitzgerald D, Irick J-S, Fienberg SE, Cacioppo JT (2009) Detection of deception using fMRI: better than chance but well below perfection. Soc Neurosci 4(6):528–538

Nahari G, Vrij A, Fisher R (2012) Does the truth come out in the writing? SCAN as a lie detection tool. Law Hum Behav 36(1):68–76

National Research Council (2003) The polygraph and lie detection. Committee to review the scientific evidence on the polygraph. The National Academic Press, Washington

O'Sullivan M (2013) Is *le mot juste*? The contextualization of words by expert lie detectors. In Cooper BS, Griesel D, Ternes M (eds) Applied issues in investigative interviewing, eyewitness memory, and credibility assessment. Springer, New York, pp 239–257

Porter S, Doucette NL, Woodworth M, Earle J, MacNeil B (2008) Halfe the world knowes how the other halfe lies: investigation of verbal and non-verbal signs of deception exhibited by criminal offenders and non-offenders. Leg Criminol Psychol 13:27–38

Porter S, Juodis M, ten Brinke L, Klein R, Wilson K (2010) Evaluation of a brief deception detection training program. J Forensic Psychiatry Psychol 21:66–76

Porter S, ten Brinke L, Wallace B (2012) Secrets and lies: involuntary leakage in deceptive facial expressions as a function of emotional intensity. J Nonverbal Behav 36:23–37

Porter S, Woodworth M, Birt AR (2000) Truth, lies and videotape: an investigation of the ability of federal parole officers to detect deception. Law Hum Behav 24:643–658

Price HL, Roberts KP (2011) The effects of an intensive training and feedback program on police and social workers' investigative interviews of children. Can J Behav Sci 43(3):235–244

Rosenfeld JP (2002) Event-related potentials in the detection of deception, malingering, and false memories. In: Kleiner M (ed) Handbook of polygraph testing. Academic, San Diego, pp 265–286

Sapir A (1987, 2000) The LSI course on scientific content analysis (SCAN). Laboratory for scientific interrogation, Phoenix, AZ

Schacter DL (1996) Searching for memory. Basic Books, New York

Spence SA (2008) Playing devil's advocate: the case against fMRI lie detection. Leg Criminol Psychol 13:11–26

Undeutsch U (1989) The development of statement reality analysis. In: Yuille JC (ed) Credibility assessment. Kluwer Academic Publishers, Dordrecht, pp 101–120

Vrij A (2005) Criteria-based content analysis: a qualitative review of the first 37 studies. Psychol Public Policy Law 11(1):3–41
Vrij A (2008) Detecting lies and deceit: pitfalls and opportunities, 2nd edn. Wiley, Chichester
Vrij A, Granhag A, Porter S (2010) Pitfalls and opportunities in nonverbal and verbal lie detection. Psychol Sci Public Interest 11(3):89–21
Vrij A, Leal S, Granhag PA, Mann S, Fisher RP, Hillman J, Sperry K (2009) Outsmarting the liars: the benefit of asking unanticipated questions. Law Hum Behav 33:159–166
Vrij A, Mann S, Fisher R, Leal S, Milne B, Bull R (2008) Increasing cognitive load to facilitate lie detection: the benefit of recalling an event in reverse order. Law Hum Behav 32:253–265
Warmelink L, Vrij A, Mann S, Leal S, Forrester D, Fisher R (2011) Thermal imaging as a lie detection tool at airports. Law Hum Behav 35:40–48
Wiseman R, Watt C, ten Brinke L, Porter S, Couper S-L, Rankin C (2012) The eyes don't have it: lie detection and neuro-linguistic programming. PLoS ONE 7(7):1–5
Yarbrough J, Hervé HF, Harms R (2013) The sins of interviewing: errors made by investigative interviewers and suggestions for redress. In Cooper BS, Griesel D, Ternes M (eds) Applied issues in investigative interviewing, eyewitness memory, and credibility assessment. Springer, New York, pp 59–95
Yuille JC (1988) The systematic assessment of children's testimony. Can Psychol 29(3):247–262
Yuille JC (2013) The challenge for forensic memory research: methodolotry. In Cooper BS, Griesel D, Ternes M (eds) Applied issues in investigative interviewing, eyewitness memory, and credibility assessment. Springer, New York, pp 3–18
Yuille JC, Cooper BS, Hervé HF (2009) The step-wise guidelines for child interviews: the new generation. In Casonato M, Pfafflin (eds) *Pedoparafile: psychological perspectives, forensic psychiatric*, Italy: Franco Angeli

Evaluation

▶ Bullying Prevention: Assessing Existing Meta-Evaluations

Event Characteristics

▶ Estimator Variables and Eyewitness Identification

Event Control: Criminal Event, Routines Activities, Presentness, Occasions

▶ Control Theory

Evidence-Based Corrections

▶ Examining the Effectiveness of Correctional Interventions

Evidence-Based Crime Policy

▶ Evidence-Based Policy in Crime and Justice

Evidence-Based Policing

Cynthia Lum and Christopher S. Koper
Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA

Overview

Evidence-based policing is a law enforcement perspective and philosophy that implicates the use of research, evaluation, analysis, and scientific processes in law enforcement decision making. In this entry, we review the nuances of this definition, the research underpinning evidence-based policing, and what agencies employing evidence-based approaches might look like.

Fundamentals of Evidence-Based Policing

Evidence-based policing is a law-enforcement perspective and philosophy that implicates the use of research, evaluation, analysis, and scientific processes in law-enforcement decision making. This research could cover a wide array of subject matters, from evaluations on interventions and tactics to analysis of police behavior, activities, and internal management. In his 1998 "Ideas in American Policing" lecture for the Police Foundation, Lawrence Sherman gave one of the most well-known articulation of evidence-based policing. He posited that "police practices should be based on scientific evidence about what works best" (p. 2). In particular, Sherman focused on two dimensions of a research orientation in policing: (1) using the results of scientifically rigorous evaluations of law-enforcement tactics and strategies to guide decisions, and (2) generating and applying analytic knowledge derived from an agency's analysis of its own internal issues and crime problems. Although using research and analysis was not a new concept in governance and social interventions when Sherman gave this lecture, what was innovative was his assertion that police should use research and analysis more frequently, substantively, and directly, discontinuing the use of tactics that were shown not to be effective.

The idea of using objective scientific information and criteria to inform public policy and agency decision making reflects a common value in modern liberal democracies: there must be evaluative and objective accountability for governmental actions and spending (Chalmers 2003; National Research Council 2004; Sherman et al. 2002; Sherman 2003). Sherman has argued that this is especially important in policing, given the important mandate the police have ensure the rule of law. In a lecture at the Royal Society for the Encouragement of Arts, Manufactures, and Commerce (RSA) in November 2011, Sherman emphasized this point when he said:

> The competence that we achieve today stems largely from the 18th century Enlightenment. The debt we owe to that era is the great transition in so many professions from customs to science, from opinions to evidence. That enlightenment idea of objective knowledge is also crucial to the success of our liberal democracy in which the rule of the majority protects individual liberty under a rule of law. No institution is more important to that success than the police, whose competence at ensuring the rule of law is constantly challenged by thousands of opinions on how the police should do their job. It is therefore essential that our society improves the competence of its police not merely with our opinions but primarily with a fact that is derived from objective knowledge.

This notion that evidence-based approaches are implicated in the values and ethics of modern democratic governance – particularly of government accountability and reducing harm (Chalmers 2003) – is mirrored in many social arenas, but especially in public health and medical practices. There are many requirements (and laws) stipulating that medical treatments and remedies must be supported by believable and rigorous scientific testing and replication and that those treatments must provide the least amount of harm or negative side effects (or at least report those side effects). Our demand for evidence-based treatment is so strong that doctors spend a large proportion of their income insuring themselves against lawsuits if they commit malpractice. However, when the police carry out an intervention that does not work, or that increases crime or recidivism, or worsens police-community relations, it is much less likely that they will be held similarly responsible. The ideology behind evidence-based policing, however, suggests nurturing similar expectations. Former National Institute of Justice Director Jeremy Travis went so far as to assert in his keynote address at the Sentencing Project's 25th anniversary celebration that "[w]e need a professional ethic that views failure to adopt those proven policies and practices as a form of justice malpractice." While this type of legal accountability is not likely to be soon adopted in American criminal justice practice, the idea behind evidence-based policing emphasizes that law enforcement should at least be held accountable to the knowledge already known about policing interventions and also to crime analysis generated in their own jurisdictions. Police should be deploying patrol officers, specialized units, detectives, supervisors and commanders in ways that can be shown to achieve results, whatever results are sought (i.e., crime or fear reduction; legitimate, fair, and respectful treatment; or responsiveness).

Adopting an evidence-based approach to police decision making may bring numerous benefits to the police. Most obvious are the rewards reaped from employing strategies and tactics that have been shown to reduce crime, increase legitimacy, reduce internal problems, address community concerns, or reduce fear (Lum 2009; National Research Council 2004; Sherman and Eck 2002). Policies deemed harmful or ineffective could be discarded (or at least questioned), potentially saving law enforcement agencies time, money, frustration, and blame. Using more objective judgments regarding deploying tactics also seems more ethically justifiable than other nonscientific methods such as best guessing or strategies based on anecdotes or personal preferences. This approach to policing may lead to greater transparency, legitimacy, and accountability in practice, which could improve police-citizen relations and trust.

There may be additional benefits of such an approach. Evidence-based policing requires agencies to access their information and data capabilities regularly (to carry out outcome evaluations or analysis). This may lead to improvements in managerial accountability and efficiency, better data recording, collection and analysis, and a push to improve information technology systems to accommodate these needs. Other decision-making perspectives in policing, including community-oriented policing, problem-oriented policing, and professionalism, could be strengthened through the inclusion and use of scientific information and analysis. Problem-oriented policing, for example (see the Braga entry for this encyclopedia; Eck and Spelman 1987; Goldstein 1979, 1990), demands analyzing crime problems, using interventions that have been shown to be effective, evaluating interventions against sought outcomes, and potentially discarding interventions that are not shown to be effective through rigorous evaluation. Community policing might also be strengthened from an evidence-based approach, given that many different types of community-based strategies have been developed and also evaluated, some which have not been shown to either reduce crime or improve legitimacy (see Sherman and Eck 2002; Weisburd et al. in progress). Research has also played – and can continue to play – an important role in developing professional policing, especially for police concerns such as the use of force, racial profiling, or internal corruption.

An evidence-based approach may also increase satisfaction in police work by providing creative ways to carry out the profession and also challenge the status quo. Many strategies shown to be effective run counter to professional mainstays, such as rapidly responding to 911 calls, randomly patrolling one's beat, or making arrests. Weiss (1980) found that challenging the status quo is one reason bureaucrats may be receptive to research in the first place. Related to this, applying critical, analytic, and proactive thinking in police work is also more akin to transformational, as opposed to transactional, leadership and deployment styles. Transformational approaches implicate more satisfaction among both supervisors and subordinates because more creative and proactive thinking is involved (Bass 1985; Burns 1978; McCardle 2011). In this manner, evidence-based policing, as with problem-oriented policing, could influence organizational and cultural forces that can inhibit both growth and a dynamic learning environment in policing. If using certain strategies, tactics, and internal practices lead to more positive results, this may then lead to greater motivation and job satisfaction.

Yet, despite these potential benefits, laments continue about the proverbial "gap" between research and practice (Bayley 1998; Lum 2009; Lum et al. 2012; Mastrofski 1999; NRC 2004; Weisburd 2008; Weiss 1980). Even in the field of medicine, where we might expect that practices are guided by the best scientific research, these gaps persist (Chalmers 2003; Sherman 2003). In policing, Weisburd (2008) points out that the best example of the disconnect between the evidence and its use in practice is the general failure of police agencies to regularly adopt hot spots, or place-targeted patrol, despite the strong evidence of the efficacy of this approach for crime and disorder reduction (NRC 2004) and the clear criminological support for the spatial concentration of crime (Weisburd 2008). Further, Koper (2008) found that while many agencies claim to be doing hot spot policing, much of the strategies agencies discussed appear to be consistent with more traditional beat- and neighborhood-based strategies. Other examples are the continued use of the DARE program (Drug Abuse Resistance Education), reactive arrests, rapid response to 911 calls, and gun buybacks – all strategies that have evidence showing ineffectiveness.

Reasons for the lack of evidence-based approaches in policing are many. Transforming more abstract and general research findings and experiences into tangible and specific law-enforcement tactics, practices, and strategies is a difficult and hard-to-measure venture, just as is applying one's education or training to any workplace task. Research knowledge often is not written in ways that make it straightforward for officers to receive or use in practice. Police chiefs often cite the lack of resources, political will, and the potential for police unions to object as challenges to adopting an evidence-based approach. There are also false expectations and beliefs about the role of researchers and research in policing. Despite what some have argued (see Sparrow 2011), researchers have rarely claimed that research or scientific processes can run a police department's daily operations or resolve law-enforcement concerns, just as problem-oriented policing, community-oriented policing, or professionalism cannot. However, incorrect beliefs and prejudices about research and researchers may lead to a widening communication gap between researchers and practitioners.

Further, implementing evidence-based policing is a challenge because its principles of decision making compete with an organizational culture in which decisions are made using other philosophies and processes. These include hunches and best guesses; traditions and habits; anecdotes and stories; emotions, feelings, whims, and stereotypes; political pressures or moral panics; opinions about best practices; or just the fad of the day (Lum 2009). Granted, sometimes these processes and best practices can be influenced by knowledge, information, and analyses. However, they are also highly vulnerable to personal opinions, ideology, and stereotypes. The difficulty of evidence-based policing lies not only in the transformation of research into practice but also in adjusting the culture to be more receptive to research and scientific processes.

Some have also argued that evidence-based policing needs to acknowledge that crime control research sometimes focuses on outcomes that are not as practical as initially believed. For example, Mastrofski and Willis (2010) suggest that police visibility and responsiveness are how the public judges effectiveness of policing, not necessarily some abstract notion of crime reduction. In turn, this may explain why police have not fully adopted an approach that targets crime concentrations. Citizens may want to see police patrolling their own neighborhoods, irrespective of where crime is most prevalent. And they may focus on other performance measures when judging the police, such as quick response and fairness. Whether this is the reason why police have been reluctant to move from a reactive to a more proactive patrol strategy should itself be subjected to study. Nonetheless, these are important debates about the utility of evidence-based policing and its focus. As Professor Stephen Mastrofski emphasized in personal correspondence to the authors, "for evidence-based policing to be useful in a democratic setting, it needs to measure success with sufficient sensitivity to the diversity of values that come to play in making choices about policy and practice. An evidence-based policing strategy that focuses on only a narrow range of values is more appropriately termed 'blinder-based policing'."

Thus, while the notion of evidence-based policing may in theory seem reasonable, rational, and even democratic, using research in police daily practice is much more complex and nuanced in reality. Building on Sherman's discussion, therefore, Lum et al. (2012) add three nuances into the definition of evidence-based policing, which in turn highlight the difficulty in the fulfillment of its own definition:

1. Evidence-based policing is a decision-making *perspective*, not a panacea.
2. It is grounded in the idea that policies and practices should be *supported by* research evidence and analytics, not blindly determined by them.
3. It suggests that research is *not ignored* and that it at least becomes a *part of the conversation* on what to do about reducing crime, increasing legitimacy, or addressing internal problems.

The Research Supporting Evidence-Based Policing

Given that evidence-based policing requires the use of research in practice, what then is the research evidence that supports it? There have been decades of policing and criminal justice research that can benefit police decision making across multiple areas of police practices and activities, as most recently reviewed by a special committee of the National Research Council (NRC 2004). Below, we focus on the research evidence regarding crime control and prevention of police interventions as only *one example*. We italicize "one example" because some critics view evidence-based policing as a "what works" movement, focused only on experimental evaluations of crime prevention interventions. However, we remind and caution the reader that there is much evidence in policing that is not about tactical interventions, which is also part of the research base in evidence-based policing, as Moore (2006) and Willis (2012) point out. Evidence-based policing is a *general* reform concept about the use of research in policing, not about the specific areas of research or the type of researcher that should be its focus.

But even though research can span the gamut of areas and types, it should also be mentioned that no matter what area of research is used to underpin law-enforcement decisions, a fundamental tenet of an evidence-based approach is that not *all* research in that area should be used. Research in policing uses a wide gamut of methods, including experiments, quasi-experiments, before-and-after designs with (and without) control groups, correlational research, case studies, ethnographies, and other qualitative studies. An evidence-based approach posits that agencies should make use of what Sherman (1998) calls the "best available" evidence – research that is of high quality, scientifically sound, believable, and therefore of high internal and external validity (Boruch et al. 2000; Campbell and Boruch 1975; Farrington 2003; MacKenzie 2008; Weisburd 2000). However, like the "what works" critique mentioned above, what constitutes high-quality research is also debated (Moore 2006). For example, with crime-control evaluation research, not all research evidence is created equal (Sherman et al. 2002). What is the threshold of methodological rigor by which an evaluation study should matter? Further, there are ranges of quality and fidelity in the implementation of research designs. Various sampling and measurement issues may also influence the validity of findings.

At the same time, we know there are many examples of criminal justice interventions that were deemed effective using scientifically weak assessments (or no assessment at all) and that were later discovered to be ineffective or even harmful (DARE, boot camps, reactive arrests, some community policing strategies, intensive supervised probation). As Weisburd et al. (2001) discovered, evaluation studies of lower quality in criminal justice, which may be less valid and reliable, are also more likely to report positive findings. While there are specific debates regarding the utility and ability of experimentation to evaluate crime prevention interventions (see Berk 2005), there is general agreement that an evidence-based approach posits that the rigor of the science behind an evaluation is an important consideration in the believability (and utility) of any particular study.

Thus, many reviews of research attempting to synthesize crime prevention and control research for practical use have been sensitive to the methodological rigor of studies. One of the more influential and recent reviews was conducted by Sherman as part of the University of Maryland Report to Congress on "What Works, What Doesn't, and What's Promising" in crime prevention, later updated by Sherman and Eck in Sherman and colleagues' (2002) study. Each policing evaluation was scored with regard to rigor of method used for evaluation, and then, based on evidence and the methodological rigor underpinning that evidence, a judgment was made about the effectiveness of policing interventions. Relying on the studies exhibiting strong internal validity, Sherman and Eck (2002) concluded in their update that directed patrols to hot spots, proactive arrests of serious repeat offenders and drunk drivers, arrest for employed suspects for domestic violence, and problem-oriented policing seemed to be evidence-based policing approaches. On the other hand, minor juvenile arrests, arrest for unemployed suspects of domestic violence, drug market arrests, some community policing approaches, or increasing police numbers did not seem evidence based.

In addition to – and prompted by – the Maryland Report were a number of subsequent reviews of the evidence of policing interventions promoted by the Campbell Collaboration, specifically its Crime and Justice Coordinating Group (see Farrington and Petrosino 2001). Similar to the Maryland Report, Campbell reviews emphasize that more weight should be given to research evidence with high levels of internal validity. In many reviews, only experiments and multi-subject quasi-experiments are considered in making final determinations about interventions in which multiple studies have been undertaken. Unlike the Maryland Report, Campbell reviews hone in on specific areas of policing and law enforcement. In policing and security, the most highly cited is likely the systematic reviews of the research evidence on hot spot policing (Braga 2007), which have shown place-based targeted patrol to be effective in reducing crime at places. The problem-oriented policing review (see Weisburd et al. 2010) indicates that specific and tailored approaches, in particular focusing on small places, can be effective. Bennett et al. (2008) neighborhood watch review found positive effects of neighborhood watch, although studies with lower internal validity were included in this review, and the conclusions of which counter Sherman and colleagues' (2002) more pessimistic findings. In their review of gun-carrying suppression, Koper and Mayo-Wilson (2006) found that police crackdowns on gun carrying are effective on reducing gun crime in crime hot spots.

Further, contrary to the Braga hot spots review, Mazerolle et al. (2007) systematic review of evaluations regarding street-level drug enforcement interventions found that problem- and community-oriented approaches were more effective in reducing drug calls and incidents than only focusing law enforcement at drug hot spots using standard tactics. In a review of second-responder programs for family abuse, Davis et al. (2008) found that such programs do not reduce the likelihood of future violence. Upcoming reviews include studies on pulling levers (Braga and Weisburd 2012) and community policing (Weisburd et al. in progress). These systematic reviews examine multiple studies on the same subject to draw conclusions and generalizations about the evidence, so that law-enforcement entities can better digest large amounts of research in an organized way.

Recently, a committee of policing scholars conducted a broader review of law-enforcement research for the National Research Council (NRC). Writing about the *Fairness and Effectiveness in Policing* (NRC 2004), the committee found that with regard to research on the "effectiveness of police activities in reducing crime, disorder and fear," the committee noted four key points about the state of research evidence in policing (summarized from NRC 2004: pp. 246–247):

1. The standard model of policing that emphasizes random patrol, rapid response to calls for service, follow-up investigations by detectives, and unfocused enforcement efforts has not been effective in reducing crime (see also Sherman 1997; Sherman and Eck 2002).
2. Some of the strategies falling under the umbrella of community policing have been effective in reducing crime, disorder, or fear of crime, while others have not (also see Bennett et al. 2008; Sherman 1997; Sherman and Eck 2002).

3. Police strategies that are more focused and tailored to specific types of crimes, criminals, and places are more effective (also see Braga 2007; Koper and Mayo-Wilson 2006; Mazerolle et al. 2007; Weisburd et al. 2010).
4. Problem-oriented policing – a strategy involving systematic analysis of crime and disorder problems and the development of tailored solutions (Goldstein 1979) – is effective (also see Weisburd et al. 2010).

Supporting many of the findings from the Campbell reviews and Sherman (1997) and Sherman colleagues' (2002) reviews, place-focused, hot spot policing strategies – that is, patrol, problem-solving, and/or other interventions focused on small areas or specific places of crime concentration – have proven particularly effective in several rigorous outcome interventions (Braga 2007). In the judgment of NRC, the research on hot spot policing constitutes the "strongest collective evidence of police effectiveness that is now available" (NRC 2004, p. 250).

Finally, and most recently, Lum (2009) and Lum et al. (2011) have developed the *Evidence-Based Policing Matrix* (http://cebcp.org/evidence-based-policing/the-matrix/) an interactive web-based tool which houses all police crime-control intervention research of moderate to high methodological quality. As with other reviews, the Matrix's intentions are to provide easy ways for the evidence base to be accessed and used by decision makers (i.e., practitioners, policy makers, and researchers). However, the goal of the Matrix is not just to show but also to refine knowledge and to facilitate its translation and implementation through the application of the online tool (Lum 2009; Lum et al. forthcoming). The idea of evidence-based policing translation tools suggest that such tools can draw out generalizations of tactics that are effective by examining a range of studies and then transform and apply that knowledge to specific problems an agency faces.

The Matrix uses a classification system for police interventions based on three very common dimensions of crime prevention strategies: the nature and type of target, the degree to which the strategy is reactive or proactive, and the strategy's level of focus (the specificity of the prevention mechanism it used). Using this three-dimensional "matrix," Lum (2009) and Lum and colleagues (2011) mapped all moderately rigorous to highly rigorous research evidence on police crime-control interventions according to how they might be characterized on these three dimensions (see Fig. 1, from Lum et al. 2011). By doing this, clusters of studies (and their findings) illustrate the distribution and concentration of evaluations and effective practices along intersections of dimensions. In other words, police might be able to better glean generalizations from a large body of research about what intersecting dimensions tend to characterize effective interventions.

In general, the Matrix indicates the following general principles of the evidence base of police crime-control interventions:

- A large majority (79 % at the time of writing) of successful interventions studied occur at "micro-places" or "neighborhoods."
- 64 % of successful interventions are "focused" or tailored strategies.
- 80 % of successful interventions are either "proactive" or "highly proactive."
- 53 % of interventions that show "no effect" or a "backfire effect" focus on targeting individual(s).

The evidence from the Matrix suggests that when police build strategies and tactics that are more proactive (or are based on using information-led and problem-solving approaches to reduce crime and related problems), when they target problems with greater specificity as to place and prevention mechanism, and when they shift from only a reactive/individual approach to incorporating more proactive place-based approaches, they will likely be more effective in reducing crime and disorder (see Lum et al. 2011). These findings are even more marked when only considering studies within the Matrix that are high-quality quasi-experimental and experimental research.

Although reviews of evaluation research by Campbell reviews, the Matrix, or the Maryland Report help to organize studies by rigor or type,

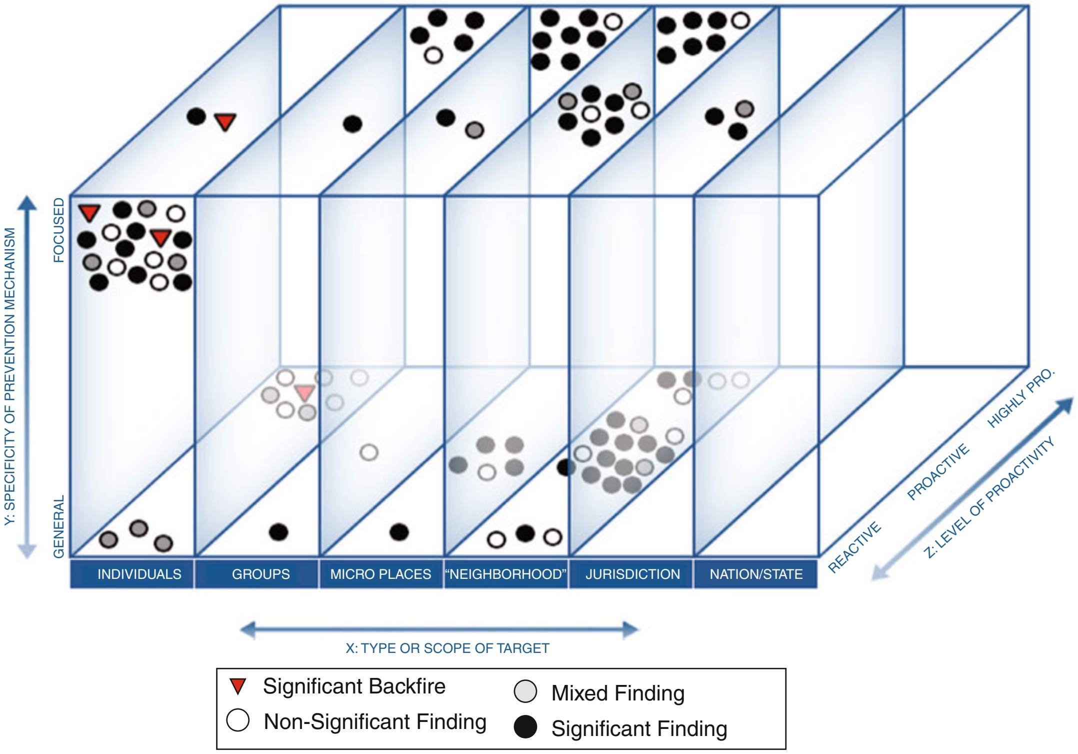

Evidence-Based Policing, Fig. 1 The evidence-based policing matrix (Lum 2009; Lum et al. 2011); www.policingmatrix.org

there are other sources that could also be used to support an evidence-based approach in policing. For example, Crimesolutions.gov, an Office of Justice Programs web site, reviews a number of studies across multiple criminal justice arenas. The Center for Problem-Oriented Policing (http://www.popcenter.org/) houses a number of guides, which emphasize not only the application of research knowledge but also systematic problem-solving approaches using crime analysis. Numerous centers and universities also house information and research regarding faculty areas of expertise.

What Would an Evidence-Based Policing Agency "Look Like"?

Additional insights into defining "evidence-based policing" can be gleaned from a discussion of what an evidence-based law-enforcement agency might look like given this specific research base. Again, the examples used here focus on the crime-control research base reflected in the Matrix, but other research categories could also be similarly applied (e.g., research on what strengthens police legitimacy). Given the evidence, at the most basic level, an evidence-based policing agency requires an effort to at least balance traditional and reactive deployment approaches with other types of patrol and investigative techniques that are *generally* supported by research evidence. But what are these more evidence-based patrol and investigative techniques? From the knowledge gained from the last three decades of research, an agency would be more "evidence-based" if it shifted its deployment strategy from one primarily spent on rapid response and random patrol to more proactive, directed, and problem-oriented targeting of crime hot spots. The research also indicates in its totality that a shift from vague community-oriented approaches to more targeted, problem-solving approaches at specific places involving multiple agencies may be more effective (Weisburd and Eck 2004; Weisburd et al. 2010). If agencies want to focus on individuals, focusing on repeat violent offenders and strengthening postarrest case enhancement and the use of DNA analysis may be important. Further, as Mastrofski commented above (see footnote 4), research on other values and priorities, such as due process, respectfulness, and fairness, would also have to be considered in developing these different patrol approaches.

Another basic requirement for an agency to be more research-oriented is that it has to have some way of assessing tactics and strategies and gathering knowledge about crime, the community, and internal issues. This involves having personnel assigned and committed to data collection, analysis, and evaluation, as well as having regular systems and procedures for using research and analytic evidence for various organizational functions (e.g., in managerial meetings, promotions, and assessments). This can include strengthening crime analysis or research and planning units, as wells as establishing standard operating procedures that ensure that units apply this information.

Finally, at the most basic level, knowledge from research, whether it comes from evaluations of policing interventions, research on racial profiling or traffic stops, or research regarding different ways to interact with the community, must be incorporated into academy and in-service training. This is a basic requirement that seems the easiest (and most obvious) to implement yet is likely furthest from reality. As indicated by Bureau of Justice Statistics data on academies (see Reeves 2009), it would not be surprising to find that many police academies and in-service systems do not seriously incorporate the latest information on what are the most effective ways police can use to reduce crime, increase legitimacy in the community, or reduce problem behaviors within the agency. Academies traditionally teach reactive skills such as applying police procedures and the law, writing reports and submitting evidence, or using firearms and motor vehicles. All of these focus on police reactive response to crime. Yet, a large portion of an officer's time is not committed to reactive procedural response but is uncommitted, leaving room for high levels of officer discretion (see Famega et al. 2005). Given this, the proactive skills and knowledge base that are needed to carry out effective and legitimate crime prevention during this discretionary time are often not taught. Yet a large portion of an officer's time is not committed to reactive procedural response but is open to use at the officer's discretion (see Famega et al. 2005). Further, knowledge disseminated about respectful and fair policing may build a more respective and legitimate police force.

Such efforts only reflect what an evidence-based approach would be *at the minimum*. Agencies that want to be more advanced in using analytic and research knowledge may pursue these with greater intensity and innovation. For example, mid-level achievement in evidence-based policing might be reflected in an agency having an active crime analysis culture (and specifically allocated resources) that constantly generates information for proactive enforcement and assessment of activities. In even more intense versions of evidence-based policing, these analysts become the criminologists, seeking to find underlying reasons for crime problems and patterns and also assisting the agency in evaluating both its strategies and tactics to reduce crime and its approaches to addressing internal concerns.

More intensive evidence-based approaches might include not only incorporating research and analytic knowledge into academy and in-service training but also using it to rewrite standard operating procedures (SOPs) to conform more to research knowledge. An example might be the SOPs related to preventive patrol and noncommitted time. Currently, SOPs may not include guidance or directives on what officers should be doing when they are not answering calls or writing reports. By building research knowledge into SOPs, such knowledge also becomes part of the information base used for decisions about issues such as promotions and transfers. Further, when police receive research information through familiar agency sources, it increases their receptivity to the information – (Lum et al. 2012). In these agencies, commanders, first line supervisors, and officers ideally would develop a more sophisticated understanding of these issues which become part of their technical expertise (and a requirement for promotion).

Agencies well entrenched in an evidence-based approach work regularly with researchers inside and outside of their agencies and use that knowledge and those relationships to mold both officer *and* citizen expectations of the law-enforcement role and function in society. The sheriff or chief might even take an active role in reminding his or her political counterparts in the city council or state governments about outputs from research as justifications for activities and resource allocation. Agencies committed to evidence-based approaches would have a portion of their budget specifically devoted to research and analysis of the agency's activities and behaviors.

These are only just a few examples of how research might be incorporated into police practices. Institutionalizing the use of research in practice involves many adjustments to organizational, personnel, incentive, and policy structures in policing. (For more in-depth discussion and examples of these ideas, see the *Matrix Demonstration Project* (http://cebcp.org/evidence-based-policing/the-matrix/matrix-demonstration-project/) as well as Lum et al. 2012 and Weisburd and Neyroud 2011.) Further, achieving evidence-based policing cannot rely only on the will of police leaders or the merits and volume of the research base. Police must be receptive to such an approach; the research must be useful, and there needs to be a demand for such knowledge in policing. The logic and rationality of evidence-based policy more generally in medicine, governance, or social interventions can be overshadowed by stronger human tendencies of habit, tradition, and culture that can block receptivity toward research use in policing. The example of the continued use of DARE (Drug Awareness Resistance Education) comes to mind (Birkeland et al. 2005). Despite research evidence showing DARE's ineffectiveness, police, schools, and parents support the program for reasons other than meeting its intended outcome. Overcoming a culture of reactivity, low use of analysis, weak supervision and accountability, and a suspicion of researchers, research, and scientific processes will be keys to successfully implement such an approach. This not only requires a sea change in law enforcement culture but also real changes in organizational infrastructure that allows for research to be better received and digested by members of the agency.

Related Entries

- ▶ History of Randomized Controlled Experiments in Criminal Justice
- ▶ Hot Spots and Place-Based Policing
- ▶ Intelligence-Led Policing
- ▶ Problem-Oriented Policing

Recommended Reading and References

Bass BM (1985) Leadership and performance beyond expectations. Free Press, New York

Bayley DH (1998) Policing in America: assessments and prospects, Ideas in American policing lecture series. Police Foundation, Washington, DC

Bennett T, Holloway K, Farrington D (2008) The effectiveness of neighborhood watch. Campbell Collaboration systematic review final report. http://campbellcollaboration.org/lib/download/248/. Retrieved 10 Jan 2012

Berk R (2005) Randomized experiments as the bronze standard. J Exp Criminol 1(4):417–433

Birkeland S, Murphy-Graham E, Weiss CH (2005) Good reasons for ignoring good evaluation: the case of the Drug Abuse Resistance Education (D.A.R.E.) program. Eval Program Plann 28:247–256

Boruch R, Snyder B, DeMoya D (2000) The importance of randomized field trials. Crime Delinq 46:156–180

Braga AA (2007) Effect of hot spots policing on crime. Campbell Collaboration systematic review final report. http://campbellcollaboration.org/lib/download/118/. Retrieved 10 Jan 2012

Braga AA, Weisburd DL (2012) The effects of focused deterrence strategies on crime: a systematic review and meta-analysis of the empirical evidence. J Res Crime Delinq. doi:10.1177/0022427811419368

Burns JM (1978) Leadership. Harper and Row, New York

Campbell DT, Boruch RF (1975) Making the case for randomized assignment to treatments by considering the alternatives: Six ways in which quasi-experimental evaluations in compensatory education tend to underestimate effects. In: Bennett CA, Lumsdaine AA (eds) Evaluation and experiment: some critical issues in assessing social programs. Academic, New York

Chalmers I (2003) Trying to do more good than harm in policy and practice: the role of rigorous, transparent, and up-to-date evaluations. Ann Am Acad Pol Soc Sci 589:22–40

Davis RC, Weisburd D, Taylor B (2008) Effects of second responder programs on repeat incidents of family abuse. Campbell Collaboration systematic review final report. http://campbellcollaboration.org/lib/download/233/. Retrieved 4 Jun 2010

Eck JE (2006) Science, values, and problem-oriented policing: why problem-oriented policing? In: Weisburd D, Braga A (eds) Police innovation: contrasting perspectives. Cambridge University Press, Cambridge

Eck JE, Spelman W (1987) Problem solving: problem-oriented policing in Newport News. Police Executive Research Forum, Washington, DC

Famega CN, Frank J, Mazerolle L (2005) Managing police patrol time: the role of supervisor directives. Justice Quarterly 22:540–559

Farrington DP (2003) Methodological quality standards for evaluation research. Ann Am Acad Pol Soc Sci 587:49–68

Farrington DP, Petrosino A (2001) The Campbell Collaboration crime and justice group. Ann Am Acad Polit Soc Sci 578:35–49

Goldstein H (1979) Improving policing: a problem-oriented approach. Crime Delinq 25(2):236–258

Goldstein H (1990) Problem-oriented policing. McGraw Hill, New York

Koper C (2008) The varieties and effectiveness of hot spots policing: Results from a national survey. Presentation at the annual meeting of the American Society of Criminology, St. Louis, November 2008

Koper C, Mayo-Wilson E (2006) Police crackdowns on illegal gun carrying: a systematic review of their impact on gun crime. J Exp Criminol 2:227–261

Lum C (2009) Translating police research into practice. In: Ideas in American policing lecture series. Police Foundation, Washington, DC

Lum C, Koper C, Telep CW (2011) The evidence-based policing matrix. J Exp Criminol 7:3–26

Lum C, Telep C, Koper C, Grieco J (2012) Receptivity to research in policing. Justice Research and Policy 14:61–95

MacKenzie D (2008) Reentry: examining what works in corrections. Keynote Address for the 16th Annual ICCA Research Conference, St. Louis, MO

Mastrofski SD (1999) Policing for people. In: Ideas in American policing lecture series. Police Foundation, Washington, DC

Mazerolle L, Soole DW, Rombouts S (2007) Street-level drug law enforcement: a meta-analytic review. Campbell Collaboration systematic review final report. http://campbellcollaboration.org/lib/download/123/. Retrieved 4 Jun 2010

McCardle B (2011) Activating officers: the relationship between leadership styles and officer engagement. (Unpublished master's thesis). Cambridge University, Cambridge, UK

Moore M (2006) Improving police through expertise, experience, and experiments. In: Weisburd D, Braga A (eds) Police innovation: contrasting perspectives. Cambridge University Press, Cambridge, UK

National Research Council (2004) Fairness and effectiveness in policing: the evidence. In: Skogan W, Frydl K (eds) Committee to review research on police policy and practices. Committee on Law and Justice, Division of Behavioral and Social Sciences and Education. The National Academies Press, Washington, DC

Reeves B (2009) State and local law enforcement training academies, 2006. Publication NCJ 222987. Bureau of Justice Statistics, Washington, DC

Sherman LW (1997) Policing for crime prevention. In: Sherman LW, Gottfredson D, MacKenzie D, Eck J, Reuter P, Bushway S (eds) Preventing crime: what works, what doesn't, what's promising. National Institute of Justice, U.S Department of Justice, Washington, DC

Sherman LW (1998) Evidence-based policing. In: Ideas in American policing lecture series. Police Foundation, Washington, DC

Sherman L (2003) Misleading evidence and evidence-led policy: making social science more experimental. Ann Am Acad Polit Soc Sci 589(1):6–19

Sherman LW, Eck JE (2002) Policing for crime prevention. In: Sherman LW, Farrington DP, Welsh BC, MacKenzie DL (eds) Evidence-based crime prevention. Routledge, New York, pp 295–329

Sherman LW, Farrington DP, Welsh BC, MacKenzie DL (eds) (2002) Evidence-based crime prevention. Routledge, New York

Sparrow M (2011) Governing science. In: New perspectives in policing. Harvard executive session on policing and public safety. National Institute of Justice, U.S. Department of Justice, Washington, DC

Weisburd D (2000) Randomized experiments in criminal justice policy: prospects and problems. Crime Delinq 46:181–193

Weisburd D (2008) Place-based policing. In: Ideas in American policing lecture series. Police Foundation, Washington, DC

Weisburd D, Bennett T, Gill CE, Telep CW, Vitter Z (In progress) Community-oriented policing to reduce crime, disorder, and fear and increase legitimacy and citizen satisfaction in neighborhoods. Campbell Collaboration systematic review

Weisburd D, Eck JE (2004) What can police do to reduce crime, disorder and fear? Ann Am Acad Pol Soc Sci 593:42–65

Weisburd D, Neyroud P (2011) Police science: toward a new paradigm. New Perspectives in Policing. Harvard University, Boston

Weisburd D, Lum C, Petrosino A (2001) Does research design affect study outcomes? Ann Am Acad Pol Soc Sci 578:50–70

Weisburd D, Telep CW, Hinkle JC, Eck JE (2010) Is problem-oriented policing effective in reducing crime and disorder? Findings from a Campbell systematic review. Criminol Publ Pol 9:139–172

Weiss CH (with Bucuvalas MJ) (1980) Social science research and decision-making. Columbia University Press, New York

Willis JJ (2012) The craft of policing. In: Ideas in American policing lecture series. Police Foundation, Washington, DC

Evidence-Based Policy and Practice

► Evidence-Based Policy in Crime and Justice

Evidence-Based Policy in Crime and Justice

Anthony Petrosino[1] and Robert Boruch[2]
[1]Learning Innovations, WestEd, Woburn, MA, USA
[2]Graduate School of Education, University of Pennsylvania, Philadelphia, PA, USA

Synonyms

EBP; Evidence-based crime policy; Evidence-based policy and practice

Overview

Evidence-based policy in crime and justice is defined here as a conscientious approach to intentionally use research to make decisions about programs, practices, and larger scale policies. Evidence-based policies, by and large, are those found to be effective in prior rigorous studies to reduce crime or improve justice. This article first summarizes the rationale for an evidence-based approach and the different types of evidence needed to respond to different policy questions. Exactly how much influence evidence should have on policy is then considered, and key mechanisms in the approach are illuminated. After providing a brisk history of evidence-based policy, the article concludes with some important by-products and limitations to the approach.

Introduction

In recent years, the term "evidence-based" has become the catch phrase to describe the intentional use of research evidence to make policy and practice decisions. The cornerstone of evidence-based policy would be a conscientious attempt by government and other decision-makers to adopt policies, practices, and programs (hereafter, policies) that have been shown to be effective in prior evaluation studies, usually rigorous randomized controlled trials or high-quality quasi-experiments. There are a few who would argue that policies are "evidence-based" if they have sound theoretical and conceptual footing based on prior research, even if the actual policy has not been previously tested. A commitment to evidence-based policy would be demonstrated, among other things, by decision-makers determining what prior research indicates about their current policies and whether they should be continued or revised and what prior evidence is on any new policies they are considering adopting.

Why Evidence-Based Policy in Crime and Justice?

Why has evidence-based policy become "all the rage" (Tilley and Laycock 2000), in the USA and elsewhere? The pursuit of evidence-based policy has several drivers. First, policies may have a range of effects, from strong positive impact (helping the situation) to no effect (basically a zero impact in the setting) to a large negative effect (hurting the situation). Policy makers who wish to maximize the possibility of having positive impacts with the future policies they implement may do so by adopting policies shown to be effective in prior research.

Second, governments are under pressure to make choices and commitments while simultaneously dealing with shrinking resources. Dollars spent and resources expended toward ineffective strategies are costly and hard to recoup. Allocating those resources to policies that, at least, have been demonstrably effective somewhere else is seen as a "safer bet."

Third, evidence is viewed as a more rational input for making decisions about policies than factors such as the special case or anecdote, ideology, politics, pressures from special interests, and tradition. Policies implemented on the basis of such input are sometimes criticized for not being based on rational consideration of the problem and potential solutions.

Finally, it is important to recognize that some policy makers and decision-makers do indeed want to make improvements in their venue and collaborate with others to help generate the evidence on what works and what does not. Indeed, some of these individuals are recognized annually through awards, made by George Mason University's Center for Evidence-Based Crime Policy, for instance, to police chiefs and others who have contributed notably to improvement through the generation and use of evidence.

What Evidence Should Answer What Question?

Evidence-based policy in crime and justice invites attention to the kinds of dependable evidence that exists, or can be generated, so as to inform decisions at the policy level. Five major evaluation questions invite efforts to generate dependable evidence.

The first is "What is the nature and severity of the problem? And what is the evidence?" No consideration of crime and justice policy can proceed without good understanding of where the problems are. Sherman (1998), in his seminal paper for the Police Foundation, argued that evidence-based policing would come to depend on large-scale monitoring databases which would help to empirically identify problems requiring law enforcement intervention. In recent years, such dependable evidence has been embodied in such information sources as COMSTAT, police administrative data (which is often used to identify trouble spots in the jurisdiction known as hot spots), and other sources in local jurisdictions. At the national level, answers have depended on sources such as the National Crime Victimization Surveys (NCVS) and Federal Bureau of Crime (FBI) police-captured crime reports. There are many standards for these kinds of data. The most critical and best articulated concern the statistical quality of the sample, the data's reliability currently and over time, and the data's validity in capturing what the researchers and policy or practice partners intend to measure.

The second set of questions fall under the general umbrella of etiology or causation, such as "Where did the problem come from?" and "what is the evidence?" There are many approaches to identifying potential causes of crime or other problems related to justice. However, longitudinal cohort studies are often recommended as a powerful form of evidence to answer questions about etiology of criminal behavior. This is because such studies usually include repeated measurements on individuals over substantial time periods. For example, longitudinal studies that included measurements on individuals from a very young age to adulthood can better identify factors that preceded the start or onset of delinquency or crime than cross-sectional studies that include only one measurement at the same point in time.

The third set of questions regarding evidence-based crime and justice policy is "What programs are deployed to address the problem, and how well are they deployed, and what is the evidence?" Before implementing new policies and programs to address a problem, it is useful to inventory what is currently being done about it. These data can be elusive, but funding agencies, organizations providing services, and schools or communities receiving services may be sources of learning what is going on. Evidence standards for such data are not as well thought out, but would likely include how comprehensive and reliable the data are, including how was current practice sampled. Some organizations will conduct "policy scans" to identify what is going on; this strategy may be simple, as in documenting state laws for sentencing sex offenders, or it may be more complex, as in documenting violence prevention strategies employed by schools, police, and community groups in a city. In any case, the quality of the policy or program's implementation is usually a crucial consideration whether the activity is ongoing or new and whether the program is tested in a controlled trial or not.

A fourth set of questions has to do with "What works and could work better?" Further, what is the evidence? This is often the crux of research contributions to evidence-based policy. As mentioned earlier, the initial policy idea may seem to be a good one, but it may go belly up after implementation. There may be unforeseen negative or backfire effects. Evaluations provide the evidence on answering questions about whether a policy has worked or not. But such impact evaluations are not all "created equal," and some types of studies have higher internal validity, i.e., they do a better job of convincing the skeptic that reasons for the outcome other than the program have been satisfactorily controlled.

In the context of this fourth set of questions, the randomized controlled trial is viewed by most researchers as providing the least equivocal evidence on what works. The aim of the randomized trial is to generate fair and unbiased evidence about whether one or another approach works better. The key theme is fair comparison. Randomized controlled trials, including place-based ones, on hot spot policing, mandatory arrest for domestic violence, Scared Straight programs, and others mounted since the 1960s which are described elsewhere in this Encyclopedia.

Because policies in crime and justice are often administered at federal or state level, they may be difficult to evaluate through randomized experimental design. Another class of statistical designs for impact evaluations that are viewed by some researchers as convincing (and having high internal validity) are quasi-experiments. These study designs do not use randomization to assign entities in the study to conditions. Instead, they use matching or other techniques to assess the impact of policies so as to approximate the results of a randomized controlled trial.

The final set of questions is "What is the cost effectiveness of alternatives? And what is the evidence?" Once there is some demonstrable evidence on effectiveness of a policy, it is also important to capture data on the costs of the policy and potential alternatives. This can be a very important consideration for policymakers, particularly if two alternatives are likely to achieve similar effects; the cheaper alternative may be the best option. To date, explicit standards for cost data are not well articulated or agreed upon. Cost-benefit analyses are demanding and often require researchers to make assumptions that are difficult to sustain and to estimate costs for phenomenon such as "psychic harm." Cost-effectiveness analyses are generally easier to generate as they estimate costs per unit of improvement for a program, i.e., every $1.00 spent under the policy results in a return on investment of $1.50. Data such as these permit researchers to compare economic data across interventions, thereby increasing the policy relevance of the research.

How Strong a Connection Between Evidence and Policy?

In some sense, evidence-based policy is simply a new label for a goal to which applied researchers have long aspired: a greater role for science in decision-making. Substantial research has been done, for instance, whether and why or why not dependable scientific evidence gets used in policy decision-making. The themes in this "knowledge utilization" literature are complex but indicate that the uses of research to make definitive decisions about policy, often referred to as "instrumental use," are not common. Carol Weiss, a leading scholar on the uses of research, discussed a different role for evidence, which she calls "conceptual use." Evidence may not be used to make clear-cut decisions about policies, for instance. But over time, in a context which she calls "the circuitry of enlightenment," the research results to provide new generalizations, ideas, or concepts that are useful for making better sense of the policy scene (Weiss et al. 2005).

The term, evidence-based policy, however, indicates a stronger role for research and evaluation than conceptual use. Many writers in the evidence-based arena argue for an evidence-informed approach that research and evaluation would, at the very least, "have a seat at the table," with other input criteria before a decision is made. Some writers have argued that decision-making should go beyond mere consideration of research (or being informed by studies) to actively using it in making choices about policies. In recent years, some have articulated a different role, in which policy makers implement a decision about a program and then seek evidence to post hoc "support" their decision. This has been referred to, with appropriate sarcasm, as *policy-based evidence*.

The Mechanisms for Evidence-Based Policy

Regardless of the exact nature of the connection, the current emphasis on evidence-based policy has created concern about the causal chain between research and evaluation evidence and the decisions and choices that are made about policy and practice in crime and justice. Enhancing the likelihood that the evidence will be used presents the challenges considered next.

First, the potential user of evidence must *know* of the existence of evidence. Evidence producers often assure that this happens by identifying potential users and developing a system that gets the information to them. Many organizations have been using passive mechanisms, such as having potential users self-subscribe to email newsletters and making specialized websites available. Active dissemination campaigns take time and money to develop properly even if one can identify the relevant target populations and the most appropriate vehicle for transmission of information to them. For example, the University of Maryland Report to the Congress on Crime Prevention (Sherman et al. 1997) was subsequently sent with summary and cover letter by the Jerry Lee Foundation to 500 leading US justice policymakers.

Another link in the evidence-based policy chain concerns the fact that the mere receipt of evidence does not assure understanding. How researchers can communicate the evidence is one aspect of this. Another aspect is the ability of the intended users to understand the evidence they receive. For example, legislators and legislative staff vary in their ability to understand what a randomized controlled trial is and why it is important. Lum et al. (2011) have generated an evidence policing matrix that portrays in brief prose, simple numbers, and graphs on the evidence for what works with what target populations and to what extent. New publications such as *Translational Criminology* have the aim of assuring that potential users can understand easily the evidence from dependable research. The US Department of Justice Office of Justice Programs has initiated a web-based effort with similar aims at http://crimesolutions.gov.

A major consideration is the organizational or political constraints on the use of evidence. The potential user may well be aware of the evidence and have an understanding of it (including limitations), but there may be serious internal and external constraints to its use. For example, on rotating shifts among the New York City Police Department officers was the subject of an evaluation conducted by the Vera Institute in the 1980s (Cosgrove and McElroy 1986). One set of precincts retained the rotating shift (where officers were continuously rotated every few weeks to different shift hours, such as 8–4, 4–12, 12–8), and the treatment precincts were provided with steady shifts. Although the experiment found that rotating shifts harmed officer health, increased sick days, and had other negative effects, the NYPD kept rotating shifts because of its traditional concerns with potential for officer corruption (i.e., shifting rotations makes it harder to know the "wrong people"). Although some actors in the NYPD asked for the study and welcomed the news, the more powerful in the bureaucracy were reluctant to make any changes.

Finally, there must be incentives to use dependable scientific evidence. Not everyone will want to come to the party or have incentives to do so. Weiss and colleagues (2008) have identified one major incentive, however, that has surfaced since the late 1990s in the United States, the "mandated use" of evidence. Mandated use means that grant funding is tied to the use of programs or policies from an accepted list or registry of evidence that identifies evidence-based programs. Jurisdictions sometimes will receive money to implement programs selected from the list. Weiss et al. (2008) raised concerns about this approach, particularly as the criteria for making the lists can also be questioned and can be politicized.

Often, there are disincentives to evidence-based policy, as decision-makers may worry about the consequences of adopting evidence-based programs if it means admitting that the ones they had implemented earlier were the wrong choices. For instance, high-quality evaluations may be eschewed by risk-averse executives who do not want to have to answer questions about their having earlier adopted "bad policy." Donald Campbell (Campbell 1969), one of the seminal writers on the evidence-based approach to policy decision-making, argued that policy makers should not be punished for using evidence when it showed their decision was wrong. Instead, they should be applauded – even rewarded – for following an approach of implementing, rigorously testing, and then modifying policy in light of this evidence.

A Brisk History of Evidence-Based Policy

Researchers have long focused on and written about the relationship between scientific research evidence and policy, although they did not specifically use the phrase "evidence-based." Indeed, in the seventeenth century, John Graunt, who produced the first serious statistical compendium of policy-related data for England, declared that one of his motives for doing so was "good, certain, and easy government." More recently, Donald T. Campbell (1969), for whom the international Campbell Collaboration is named, called for the US and other modern governments to adopt an experimental approach to developing and testing public policy:

> The United States and other modern nations should be ready for an experimental approach to social reform, an approach in which we try out new programs designed to cure specific social problems, in which we learn whether or not these programs are effective, and in which we retain, imitate, modify, or discard them on the basis of apparent effectiveness on the multiple imperfect criteria available. (p. 409)

Oakley (1998) called the period of the 1960s–1970s in the United States the "golden age of evaluation" because many government programs, rigorous experiments and quasi-experiments were influential in considerations about policy choices. As Oakley (1998) lamented, however, rigorous studies produced a number of null results, and the general critics' response was that there was something inherent in experimentation that could not detect those positive impacts that such interventions were undoubtedly having.

Although randomized experiments and quantitative systematic reviews of such evidence, called meta-analyses, have become common approaches to evidence generation since the 1960s, the more modern view of "evidence-based" policy was formulated in medicine and in the United Kingdom. A government researcher named Archie Cochrane argued that many health-care practices had already been shown to be ineffective in clinical trials but were still being used to the detriment of patients. Conversely, some medical procedures had been demonstrably effective in such rigorous studies but were being ignored in light of tradition, authority, education, clinical experience, and anecdote. Later, a group of researchers led by Iain Chalmers summarized the results of controlled trials in the area of childbirth and pregnancy, so as to identify what worked or did not based on rigorous studies. The success of that effort led Chalmers to begin working with other researchers to develop an international organization that would summarize trials in all other areas of health care. This was named the Cochrane Collaboration in honor of the man who inspired it (www.cochrane.org). Thus, an apparatus for summarizing research and getting that evidence into the hands of medical practitioners was born.

The early work summarizing the results of childbirth and pregnancy trials in Oxford and the success of the nascent Cochrane Collaboration led some to more closely connect prior evidence and current decision-making. A Canadian medical researcher, David Sackett, and his colleagues are often credited with articulating the first modern use of the term "evidence-based." Sackett and others (1996) called evidence-based medicine:

> . . .the conscientious, explicit and judicious use of current best evidence in making decisions about the care of the individual patient. It means integrating individual clinical expertise with the best available external clinical evidence from systematic research. (p. 71)

Since the early 1990s, the use of the term evidence-based has grown appreciably. In crime and justice research, Sherman's (1998) evidence-based policing report invoked the phrase probably for the first time. Sherman argued that large-scale police databases would serve as information to identify problems in need of innovation (new policies), and randomized experiments would be conducted to test the impact of the innovation on the problem. Police departments would be learning organizations that would refine their policies based on evidence from these RCTs.

The close coupling of evidence-based policy and systematic reviews began in the late 1990s. During that time, Chalmers approached the authors and others about establishing an analog to the Cochrane Collaboration that would prepare, update, and disseminate high-quality reviews of research in other areas outside of health care, such as crime and justice, education, and social welfare. These early discussions led to the official inauguration of the international Campbell Collaboration (www.campbellcollaboration.org) in January 2000. The Campbell Crime and Justice Group has overseen the development of systematic reviews relevant to crime policy and has organized annual conferences funded by the Jerry Lee Foundation to disseminate the results of these reviews to federal decision-makers and others on Capitol Hill. Collections of these reviews have also been published by notable journals including the *Annals of the American Academy of Political and Social Science* and the *Journal of Experimental Criminology*.

The advent of the C2 coincided with other major, cross-field evidence-based policy efforts, including the funding by the UK Economic and Social Research Council (ESRC) for several centers (called nodes) on various aspects of the interplay between research and decision-making, including the creation of a new journal called *Evidence & Policy*, published by the Policy Press. A US-based group called the Coalition for Evidence-Based Policy (http://coalition4evidence.org/wordpress/) that would advocate for federal government decisions to be closely tied to results of rigorous evaluations was created and has had considerable influence in getting legislation to mandate randomized controlled trials as the design of choice to evaluate these new policy initiatives (Wallace 2011). This has been mirrored in the UK by the creation of the Alliance for Evidence.

Evidence-based policy has now become a popular term throughout criminal justice, demonstrated by the creation of the Center for Evidence-Based Crime Policy at George Mason University in Virginia, USA (http://gemini.gmu.edu/cebcp/), directed by David Weisburd, and two new book series, on *Evidence-based Crime Prevention* and *Translational Criminology*, respectively, started by Springer Press.

Some By-Products of the Evidence-Based Policy Movement

Regardless of the long-standing interest in criminal justice and other fields on generating and using research in government decision-making, the more recent emphasis on evidence-based policy has led to a number of important by-products. For one, it has led to even more discussion and debate over the nature and quality of evidence that should influence policy decisions. This is seen most convincingly in the area of determining what evidence should be relied upon to determine whether an intervention works. There seems to be a consensus among scientists and policy makers that randomized controlled trials provide the least equivocal evidence on the effects of a policy intervention. But this is, by no means, unanimous, and challenges have been periodically raised to whether experiments are the best design for generating evidence on what works (e.g., see Derzon (in this Encyclopedia); Boruch 1975; Pawson and Tilley 1994). Despite these "common contentions," the emphasis on evidence-based policy has led to the recognition of the role of randomized experiments to evaluate justice policy when it is appropriate and possible to do so. For example, the Academy of Experimental Criminology was initiated in 2005 to recognize researchers who have conducted experiments as "Fellows" and to promote the use of experimental methods generally in the field of criminal justice (Weisburd et al. 2007). The *Journal of Experimental Criminology* was also initiated to provide a forum for rigorous evaluations and systematic reviews relevant to crime policy.

As aforementioned, because some policies cannot be evaluated using random assignment, such as when all of a population is receiving the intervention at the same time, the emphasis on evidence-based policy has led to proposals for using more sophisticated quasi-experimental approaches to evaluate crime policy (e.g., Henry 2009). Statistical impact evaluation designs such as regression discontinuity approaches and propensity score matching have been advocated and tried out in the social sciences (e.g., Campbell and Stanley 1966) and criminal justice (Berk and Rauma 1983; Minor et al. 1990) for several decades are now more commonly utilized. In fact, workshops for researchers to become better acquainted with the latest techniques have been offered since 2010 by professional associations such as the American Society of Criminology.

Besides putting the light on evidence, and highlighting the generation and use of results from randomized controlled trials and certain classes of quasi-experiments, the evidence-based policy focus has also led to increased attention to how the results of prior studies that bear on a policy are synthesized. Techniques for analyzing separate but related studies have been around for many decades (e.g., Mosteller and Bush 1954). Further, quantitative approaches popularized by the term "meta-analysis" have been extensively used in the social sciences since the late 1970s (see Smith and Glass 1977). "Systematic reviews," as they are now commonly referred to, have become the fulcrum of the evidence-based policy approach. This has also led to even more attention to how researchers locate relevant studies, how they can be pulled together and appraised for quality, and how they can be analyzed (Boruch and Petrosino 2010). For example, some studies are not published in peer reviewed journals, but if they meet quality standards, the evidence needs to be considered. The focus on evidence-based policy has brought with it attention on the rigorous and explicit methods needed to make sure that such evidence is located and included in syntheses designed for the decision-maker.

Another important by-product of the focus on evidence-based policy has been on developing systems that get the evidence on policies into hands of busy decision-makers. Many public agencies do not have staff that can spend the time necessary to do a systematic review, and they generally rely on external and trusted sources for evidence. The advent of electronic technology has meant that summaries of evidence from systematic reviews can be provided quickly so long as the intended user has access to the Internet and can download documents. Groups such as the Campbell Collaboration's Crime and Justice Group not only prepare and update

Evidence-Based Policy in Crime and Justice, Table 1 Evidence ratings for crime solutions

Evidence ratings			
	Icon		
Evidence rating	One study	More than one study	Description
Effective	✓+	✓+	Programs have strong evidence indicating they achieve their intended outcomes when implemented with fidelity.
Promising	✓	✓	Programs have some evidence indicating they achieve their intended outcomes. Additional research is recommended.
No Effects			Programs have strong evidence indicating that they did not achieve their intended outcomes when implemented with fidelity.

[a] A single study icon is used to identify programs that have been evaluated with only one study. A multiple study icon is used to represent a greater extent of evidence supporting the evidence rating. The icon depicts programs that have more than one study in the evidence base demonstrating effects in a consistent direction

reviews of evidence but make them freely available to any intended user around the world.

Campbell Collaboration reviews tend to be broad summaries of “what works” for a particular problem (e.g., gun violence) and classes of interventions (cognitive-behavioral programs). They are not usually focused on brand name programs or very specific, fine-grained definitions of an intervention.

Because decision-makers often need evidence on *particular* interventions, other approaches to providing evidence that is more fine grained have been developed. Some of these approaches fall under the rubrics of “best practice lists” or “evidence-based registries.” For instance, such a registry might summarize the evaluation evidence for a particular brand name of a program (e.g., D.A.R.E., Drug Abuse Resistance Education). Although the number of such registries and lists relevant to justice have grown since the late 1990s (Petrosino 2005), the largest effort in the field, Crime Solutions (crimesolutions.gov), was initiated in 2010. This project compiles the available evidence on more or less branded programs like “Functional Family Therapy” and “Aggression Therapy,” rates how rigorous the evidence is according to a schema (which prioritizes results from randomized controlled trials), and then provides a summary based on rigor and amount of evidence. Table 1 provides the schema on how evidence for these programs and policies are rated (Crime Solutions.gov 2012) (Table 1).

Limitations of Evidence-Based Policy

Despite the gains in the development and synthesis of rigorous evidence and the methods for rapid transmission into hands of policy makers (e.g., electronic distribution), there are some century-old barriers that limit and challenge those who wish to see evidence inform crime policy. As Lipton (1992) has written, the justice decision-maker at the policy level has many inputs into the choices made about funding, programs, allocations, and deployment. Research evidence is but one of the many “desiderata” of factors, and although evidence-based policy places a premium on scientific research to guide decision-making, choices are also influenced by ethics, fairness, resources, politics, ideology, special interests, tradition, experience, training, and many other factors. Researchers in criminal justice want a seat at the table, but they might find that the table is very crowded and that their voice may not be heard in the din.

There are many areas in US policy in which emotions are so charged that it is difficult for evidence to make an impact. For example, gun control has zealous advocates both for and against it, and impassioned discussions about evidence and their role in policy are often difficult if not impossible to have (Webster et al. 1997). A high profile murder, particularly if it is sexual in nature and involves a child victim, can cause knee jerk policy reactions and overreactions that have no basis or even run counter to research evidence (Petrosino 2000). In such charged situations, it is difficult for evidence to get a fair hearing, if any hearing at all.

As Boruch and Lum (2012:4) have written, "sturdy indifference to new and dependable evidence are a fact of life." Reluctance to trust evidence, even the most compelling, is true in the policy sector and other parts of society. As aforementioned, Campbell (1969) wrote about the reluctance of policy makers to embrace evaluation for fear that it will highlight failure of their decisions. Overcoming this to promote learning organizations that are not in fear of "upsets" has long been a challenge. Finckenauer and Gavin (1999) write that after the evaluation questioning the effect of California's Scared Straight program (known as SQUIRES) came out in 1983, the reaction by policy makers was to get rid of evaluation and keep the program. Instead of studying the program further, they evaluated it with "testimonials" from letters written to and by participating prisoners.

"Evidence destruction techniques" are easy to mount and often difficult to contend with. For example, in the case of Drug Abuse Resistance Education, or D.A.R.E., a long litany of evaluations questioned whether a police-led drug prevention program in elementary school reduced self-reported drug use by D.A.R.E. youth. D.A.R.E.'s leadership group, (D.A.R.E America) responded to each new evaluation with the same criticism: "D.A.R.E. has already revised its curriculum, the evaluation was of the old curriculum, therefore, the evaluation has no merit" (Petrosino 2005). Even minor tweaks of the curriculum were touted as major revisions and attempted to nullify evaluation evidence by making it appear obsolete and irrelevant. Vested interests often lead to the use of such techniques. A supporter of the "Campaign for Dark Skies" (Marchant 2004) criticized the methods in Farrington and Welsh (2002) systematic review of research on the effects of street lighting, motivated by fear of how their positive findings would affect the ability of professional and amateur astronomers to see the stars and planets in the UK.

Mears (2007) has highlighted some other challenges to evidence-based crime policy, including little questioning about whether the policies are needed in the first place, the pursuit of silver bullets or panaceas instead of a rational and holistic crime policy, the difficulties of implementing ideal evidence-based policy in the real world, the lack and misinterpretation of rigorous studies, and the challenge of getting good data on costs so that decision-makers can make rational choices.

Weiss and colleagues (2008) also make the point that in a democratic society, multiple inputs are expected and may be necessary. A strict adoption of evidence-based crime policy could translate into one small group (influential members of the scientific community) exerting too much influence on a process that, in a democratic society, should be representative of the majority of citizens.

Related Entries

- ▶ Crime Science
- ▶ Evidence-Based Policing
- ▶ Sex Offenders and Criminal Policy

Recommended Reading and References

Berk RA, Rauma D (1983) Capitalizing on nonrandom assignment to treatments: a regression discontinuity evaluation of a crime control program. J Am Stat Assoc 78(381):21–27

Boruch R (1975) On common contentions about randomized field experiments. In: Boruch, Reicken (eds) Experimental testing of public policy: the proceedings of the 1974 Social Sciences Research Council conference on social experimentation. Westview Press, Boulder, pp 107–142

Boruch RR, Lum C (2012) Eight lessons about evidence-based crime policy. Transl Criminol 2011. 1(Summer):4–5

Boruch RF, Petrosino A (2010) Meta-analyses, systematic reviews and evaluation syntheses. In: Wholey J et al (eds)

Handbook of practical program evaluation, 3rd edn. Jossey-Bass, New York, pp 531–553
Campbell DT (1969) Reforms as experiments. Am Psychol 24:409–429
Campbell DT, Stanley JC (1966) Experimental and quasi-experimental designs for research. Rand McNally, Skokie
Cosgrove CA, McElroy JE (1986) The fixed tour experiment in the 115th precinct: its effects on police officer stress, community perceptions, and precinct management. Executive summary. Vera Institute of Justice, New York
Crime Solutions.gov (2012) About Crime Solutions.gov. www.crimesolutions.gov/about.aspx
Dershowitz A (1996) Reasonable doubts: the criminal justice system and the O.J. Simpson case. Simon & Schuster, New York
Farrington DP, Welsh BC (2002) Improved street lighting and crime prevention. Justice Q 19:313–342
Finckenauer JO, Gavin PW (1999) Scared straight: the pan acea phe nomenon revisited. Waveland Press, Prospect Heights
Henry GT (2009) Estimating and extrapolating causal effects for crime prevention policy and program evaluation. In: Knutsson J, Tilley N (eds) Evaluating crime prevention initiatives. Criminal Justice Press, Monsey, pp 147–174
Lipton DS (1992) How to maximize utilization of evaluation research by policymakers. Ann Am Acad Polit Soc Sci 52(1):175–188
Lum C, Koper CS, Telep CW (2011) The evidence based policy matrix. J Exp Criminol 7:3–26
Marchant PR (2004) A demonstration that the claim that brighter lighting reduces crime is unfounded. Br J Criminol 44:441–447
Mears DP (2007) Towards rational and evidence-based crime policy. J Crim Justice 35(6):667–682
Minor KT, Hartmann DJ, Davis SF (1990) Preserving internal validity in correctional evaluation research: the biased assignment design as an alternative to randomized design. J Contemp Crim Justice 6(4):215–225
Mosteller F, Bush RR (1954) Selected quantitative techniques. In: Lindzey G (ed) Handbook of social psychology: vol. 1. Theory and method. Addison-Wesley, Cambridge, MA, pp 289–334
Oakley A (1998) Experimentation and social interventions: a forgotten but important history. BMJ 317(7167):1239–1242
Pawson R, Tilley N (1994) What works in evaluation research. Br J Criminol 34(2):291–306
Petrosino AJ (2000) How can we respond effectively to juvenile crime? Pediatrics 105(3 Pt 1):635–637
Petrosino A (2005) D.A.R.E. and scientific evidence: a 20-year history. Jpn J Sociol Criminol 30:72–88
Sackett DL, Rosenberg WMC, Gray JAM, et al (1996) Evidence based medicine: what it is and what it isn't. BMJ 312:71–2
Sherman LW (1998) Evidence-based policing. Ideas in American policing series. Police Foundation, Washington, DC
Sherman LW, Farrington DP (2002) In: Brandon W, Doris MK (eds) Evidence-based crime prevention. Routledge, London
Sherman LW, Gottfredson D, MacKenzie DL, Eck JE, Reuter P, Bushway S (1997) Preventing crime: what works, what doesn't, what's promising – a report to the attorney general of the United States. United States Department of Justice, Office of Justice Programs, Washington, DC
Smith ML, Glass GV (1977) Meta-analysis of psychotherapy outcome studies. Am Psychol 32:752–780
Tilley N, Laycock G (2000) Joining up research, policy and practice about crime. Policy Stud 21(3):213–227
Wallace JD (2011) Review of the coalition for evidence-based policy. Coalition for Evidence-based Policy, Washington, DC, http://coalition4evidence.org/wordpress/wp-content/uploads/Report-on-the-Coalition-for-Evidence-Based-Policy-March-2011.pdf. Accessed 5 Oct 2012
Webster DW, Vernick JS, Ludwig J, Lester KJ (1997) Flawed gun policy research could endanger public safety. Am J Public Health 87(6):918–921
Weisburd D, Lorraine M, Anthony P (2007) The academy of experimental criminology: advancing randomized trials in crime and justice. Criminologist 32:1–7
Weiss C, Murphy-Graham E, Birkeland S (2005) An alternate route to policy influence: evidence from a study of the Drug Abuse Resistance Education (D.A.R.E.) program. Am J Eval 26(1):12–31
Weiss C, Murphy-Graham E, Gandhi A, Petrosino A (2008) The fairy godmother and her warts – making the dream of evidence-based policy come true. Am J Eval 29(1):29–47

Evidence-Based Supervision

▶ Behavioral Management in Probation

Evidence-Led Policing

▶ Information Technology and Police Work

Evolution of Substance Abuse Treatment

▶ History of Substance Abuse Treatment

Evolutionary Perspectives on Crime Prevention

Jason Roach[1] and Ken Pease[2]
[1]Crime and Policing Group, University of Huddersfield, Huddersfield, West Yorkshire, UK
[2]Jill Dando Institute, University College London, London, UK

Overview

The Darwinian worldview, whereby organisms evolve by a process of natural and sexual selection, has come to permeate biological science and to varying degrees the social sciences, with anthropology, psychology, and economics the disciplines most influenced. Criminology has remained slightly touched by Darwinian thinking. This entry describes and refutes common criticisms of the application of evolutionary theory in the social sciences and sets out some possible points of contact between evolutionary theory and crime prevention. The policy recommendations which may be tentatively made by drawing on the meager volume of relevant research conducted to date are diverse and unlikely to offend liberal sensibilities.

Key Issues and Controversies

Evolution is the process whereby attributes of plants or animals, including people, change over generations as a result of selection pressures operating on random genetic mutations. Organisms best adapted to their environment (i.e., most equipped to survive and bring their offspring to reproductive maturity), thereby, come to feature in increasing numbers within the population. While the evidence base for evolution is seen as overwhelming by the great majority of scholars in the life sciences (see Coyne 2009), the theory remains controversial for three reasons. First, it may be rejected because its account of the origins of life is at odds with that of major religious texts. Second, the notion of "survival of the fittest" (not a quotation from Darwin) is seen as the theoretical undergirding for eugenics, the Holocaust and subsequent horrors in the name of "ethnic cleansing." The third reason is the erroneous ascription of biological determinism to Darwinian thought.

The first reason for rejection is not liable to challenge by evidence and will not be discussed, save to affirm the right of people to their own theology. The second reason stems from Herbert Spencer's perversion of Darwinian thought and should not preclude an open-minded consideration of the implications of evolutionary thinking for crime and criminality. Darwin's own view (Darwin 1871) was that the intentional neglect of the weak and helpless was a "certain and great present evil" (p. 169). The third reason is simply wrong, with large human brains offering diversity and flexibility of behavior consistent with Darwinian purposes (see Pagel 2012).

If evolution were to be accepted as a true account of the shaping of organisms' attributes, the implications for criminology are no less profound than they are for the other social sciences. Readers wishing to consider evolutionary implications for social policy generally are referred to the website of the Evolution Institute, brainchild of the biologist David Sloan Wilson http://evolution-institute.org/. The Institute's foci include risky adolescent behavior and other crime-relevant topics.

The key presumption of the theory of human evolution is that modern Homo sapiens represents the product of selection pressures operating on our forebears in the late Pleistocene era, known as the Era of Evolutionary Adaptedness (EEA). Since (excepting for a few remaining hunter-gatherer communities) our environment is now very different from that of the EEA, there is a mismatch between the setting for which evolution shaped us and the setting in which we are required to function. The plausibility of this view is best illustrated by current obesity and diabetes epidemics, the consequences of dietary preferences established at a time when food availability favored those ingesting protein and sugar-laden diets coming to prove disastrous in times of plenty (Gluckman and Hanson 2006).

Criminologists are used to confronting the most predatory and socially disruptive of human actions (plus those perhaps foolishly deemed by authorities of a particular culture and epoch as socially disruptive, hence, for example, legislation on sexual mores). But if one thinks of human behavior as lying on a continuum from cruel predation to heroic altruism, things (at least to the writers) look somewhat different. The bulk of the relevant evolutionary literature characterizes people as ultra-social, and "super-cooperators" (Nowak 2011). The term eusociality (deriving from the Greek prefix eu meaning good or real) represents a state of affairs in which some individuals reduce their own lifetime reproductive potential to raise the offspring of others. Eusociality underlies the most advanced forms of social organization and the ecologically dominant role of social insects and humans (see E.O. Wilson 2012). Wilson's view (in common with other scholars) is that communities require a majority of cooperators, with deviant "free riders" threatening community cohesion when constituting more than a small minority. Areas vary massively in rates of crime. Crime levels remain an important determinant of the decision to move home, especially for those with children and the affluent (see, e.g., Katzman 1980). The Wilson view would imply a highly skewed small area distribution of crime rates and a "tipping point" as those able to move away from an area do so.

To return to the central notion of Homo sapiens as a super-cooperator, it will come as a surprise to many criminologists, together with believers in Original Sin, that the default setting for human nature in maturity seems to be cooperation with, and weak altruism towards, other people. With that changed perception comes the suspicion that criminology ought not to form a distinct discipline. If the overarching aim of applied social science is to foster cooperative social relationships, then there is a case for ceasing to see the task as one of reducing the number of outlier behaviors labeled as crime and instead to see it as directing people towards the cooperative end of the continuum, *wherever on that continuum they started*, that is, to nudge altruists as well as the selfish further towards the heroic altruism pole. This would possibly be achieved by a rapprochement with the emergent discipline known as positive psychology. This perspective (bear in mind its origins in clinical practice) seeks to make normal life more fulfilling rather than exclusively to treat mental illness. It emphasizes the importance of using the scientific method to determine how things go right (Seligman and Csikszentmihaly 2000). In the same way, reconfiguring research so that the prosocial-antisocial continuum is treated as such, rather than antisocial outliers being hived off for study by a separate discipline called criminology, makes sense against the background of the raft of research on human eusociality. It should be stressed that evolutionary thinking is not the only route towards the incorporation of criminology within positive psychology. Indeed, religious and other obstacles mean that it may always be "the road less travelled." While the implications of what amounts to a paradigm shift in criminology are immense, that theme will not be pursued here. Rather, the bulk of what follows is restricted to the implications of evolutionary thought for criminology as currently conceived.

The Dunbar Number

Put crudely, if one accepts the Darwinian account, we evolved to act locally. How local is local? The Dunbar number refers to the maximum number of other people with whom a person can maintain stable social relationships (Dunbar 1992). Robin Dunbar found a relatively narrow band of group sizes with an average of 150. This number (plus or minus some) characterized group size among numerous contemporary hunter-gatherer groups and also the average estimated size of Neolithic farming villages. Modern comparisons were made with the size of military units. Whatever its evolutionary origins, the Dunbar number has major *unexplored* implications for designing conflict-minimal contemporary human settlements. We know that real differences in rates of crime are immense and that crime is a major reason for moving home among those who can afford to do so. We know that certain street types are associated with low levels of crime (Johnson and Bowers 2009), and the recent work of David Weisburd and colleagues at George Mason University have emphasized the need for small-scale, *street segment* analyses of crime levels (see Weisburd et al. 2012). What we have yet to do is to consider attributes of late Pleistocene settlements *alongside* contemporary variations in residential crime risk. For example, do sinuous cul-de-sacs of the kind identified by Johnson and Bowers as low in crime share line-of-sight or other attributes with hunter-gatherer settlements? The reader will think it unsatisfactory in an encyclopedia entry to say what should be done rather than what has been done, and indeed it is. However, it is inevitable given the lack of penetration of evolutionary thinking into contemporary criminology.

Child Abuse and Parenthood

Margo Wilson and Martin Daly have looked at violence (among other things) from a Darwinian standpoint for some 30 years. Their work provides an illustration of what was implicit in the foregoing text, namely, that an evolutionary mindset leads one to ask different kinds of question from those typically encountered in criminology.

For most, the killing of a child is the most heinous and distressing of all crimes (Adler and Polk 2001). It is mercifully rarer for children to be victims of homicide than adults. In England and Wales, there will be approximately 110 child homicides per year from an average total number of approximately 700 recorded homicides, roughly equating to 14 % of all homicide victims (Brookman 2005).

The notion that your own family represents the greatest danger to your well-being was conventional wisdom in the 1960s, and that undifferentiated view is still held in some quarters to this day. It was obvious to Daly and Wilson that biological and stepparents should be separated in any analysis of abuse (including lethal abuse) of children. When this was done, it was found that rates of child homicide by stepparents were *one hundred times or more* greater than rates of child homicide by biological parents. In a nod to the eponymous fairy story, they called this the Cinderella effect (Daly and Wilson 1998).

Daly and Wilson do not argue that killing the offspring of others is adaptive for humans, though it is for some other species, such as lions (see Packer and Pusey 1983) as a means whereby incoming males prevent females from investing in the care of the offspring of other males and eliminating the partial contraceptive effect of lactation, thus maximizing the period in which females may conceive their cubs. Rather, they argue that nurturance tends to be given preferentially to one's own offspring. Abuse and homicide are outliers. Most people don't keep their stepdaughters from going to balls or feed them poisoned apples. The Cinderella research has been much criticized, with Daly and Wilson defending their position persuasively (Daly and Wilson 2007). They assert that:

> "Abundant confirmatory research has followed, such that the disproportionate victimization of stepchildren is now the most extensively documented generalization in the family violence literature. This... raises further questions, such as what explains variability in the magnitude of Cinderella effects between maltreatment types and locales, and whether the individual level predictors of abuse are the same for fathers, mothers, stepfathers and step-mothers. Unfortunately, progress on these important issues has been hindered by a relentless distraction: the manufacture of "controversy" about whether Cinderella effects exist at all. We suspect that the reason for this nay-saying resides largely, though not entirely, in antipathy to the Darwinian world-view and its application to Homo sapiens." (pp. 383–384)

Daly and Wilson write thus of their critics:

> "From the perspective of two researchers on the receiving end of these attacks, a disturbing and sometimes perplexing element has been their incivility.... [our critics] are not just sceptical, they are angry, and we are still not entirely sure what they are angry about." (p. 396)

Future Discounting and Crime

Comparisons with ants and termites (the only other eusocial organisms) will be treated in some quarters with derision, but one is included here because it provides an arresting image to act as a mnemonic for the serious point subsequently made:

> "Old-aged termites go out with a bang, it appears. While aging, the insects brew a backpack of deadly chemicals, which they use to self-destruct when under attack, taking out any enemies with them… When the spotted termites were physically unable to defend themselves with their jaws, they would commit the ultimate sacrifice and burst a pouch on their backs, releasing a toxic liquid that quickly paralysed and killed any other termites it touched."(p. 14)

The moral is as follows: if you haven't got long to go, make the most of what you have. It has long been asserted that offenders are liable to high levels of future discounting, characterized by the preference for short-term rewards over larger long-term rewards (Daly and Wilson 2005). They apply an evolutionary lens to the issue. They interpret future discounting in terms of realistic anticipation of future life. They found homicide to be most frequent among those with least to lose, unemployed, and unmarried men, with divorced and widowed men reverting to higher rates of homicide. They interpret this in terms of the deployment of reproductive effort as a gamble. If one anticipates the nontrivial probability of an early death, or other form of removal from the ranks of the reproductively active, one chooses a strategy of attempting risky and frequent and if necessary coerced impregnation and the pursuit of short-term reputation and the acquisition of status symbols by acquisitive crime. Wilson and Daly argue that life expectancy itself may be a psychologically salient determinant of risk-taking.

Clearly linked to anticipations of future life is the decision known as "Dads or cads," referring to the choice of reproductive strategy among males (Cashdan 1993). Cads seek to impregnate as widely as possible, making no contribution to child rearing. Dads stay with the woman or women with whom they have conceived children. Cads will be the more effective survival vehicles if more of their children survive and reproduce. Otherwise, the genes of Dads win out in the next generation. If Wilson and Daly are correct in placing life expectancy at the center of strategy choice, the implications are clear in engendering hope of life and possible success in adolescents, especially adolescent males. This clearly has implications beyond the narrowly criminological.

Domestic Violence Against Women

According to the 2009/2010 British Crime Survey (BCS), 7 % of women aged between 16 and 59 years in England and Wales were victims of domestic abuse over the course of 1 year, the majority of the violence being "nonsexual" abuse by a partner. Although only 14 % of the total 2,087,000 violent incidents estimated by the BCS for that year were described as domestic violence, equivalent figures have been found to be up to five times higher wherein participant "self-completion" was used. Domestic violence is ubiquitous and as such if we are working to prevent violent crime than we can best make inroads by focusing on intimate partner violence. For example, a study in England and Wales (1995–2000) showed that 30 % of all the homicides were "femicides." Moreover, 57 % of these female victims were killed by an intimate (or ex-intimate) male sexual partner (Brookman 2005).

How can an evolutionary-informed violence prevention strategy help reduce domestic violence? The first step, as always, should be to try and understand what is going on.

Male "sexual jealousy" is the most frequently given explanation for intimate partner violence (Daly and Wilson 1988; Buss 2005) and so will only be briefly mentioned here. For example, research consistently points to male on female violence as being primarily motivated by "jealousy/control" on the part of the male. Put simply, men appear to use violence against women as a tactic to restrict their sexual behavior, primarily as means of enforcing sexual (i.e., reproductive) "exclusivity" (Daly and Wilson 1988). Fiona Brookman found that more than 80 % of femicides occurred where the female was either planning to leave her partner or where he perceived her at least to have been "unfaithful" with another sexual intimate, thereby compromising sexual exclusivity (Brookman 2005).

The foremost writer on this topic from an evolutionary perspective is Anne Campbell. Campbell makes the observation that in trying to explain the gender difference in crime, the male-centered approach has dominated evolutionary psychology, where there appears to be a broad consensus that the motivation to achieve status and "surplus resources" is more critical to male than female reproductive success (Campbell 2009). However, where a female-centered approach is taken, a different perspective on male on female violence is achieved, where her most important proximal goal is to stay alive for her present (and future) offspring (Campbell 2009).

Reproduction (and therefore sexual intercourse) poses more of a risk to female safety and survival than for males (Campbell 2009), a kind of male–female genetic arm race in which females must evolve defenses against the lethal potential of the sex drive of males. Intimate partner violence being one such defense, as when women kill, it is more often than not an intimate (or previously intimate) partner who is the victim (Daly and Wilson 1988).

One practical lesson which the writers take from the Campbell perspective is the significance of women's refusal to press charges or report violence committed against them by intimates. If their "wired-in" aim is survival with support (however flawed), this becomes more comprehensible. Perhaps the most common of police officer reactions to the victims of domestic violence who refuse to proceed down a criminal justice route is deep frustration, with a "why bother" bottom line. Understanding the survival perspective, alongside legal changes to maximize chances of conviction even when the victim has become unwilling to proceed, seem justified.

Epigenetics

Evolutionary history is written in our genome but does not determine our behavior. Surely all competent scholars now accept that both genetic (nature) and environmental (nurture) factors play crucial interacting roles in human development. With regard to crime and antisocial behavior, Moffitt and Caspi (2006) assert that familial inheritance is always both the result of genetic endowment and environment, but environments are made and are often correlated with the dispositions of those who inhabit them. Of particular current interest is the gene variant MOAO which lowers the expression of the enzyme monoamine oxidase A and which seems implicated in violence. This relationship is stronger among maltreated children (see Caspi et al. 2002). In brief, the effects of abuse are more pernicious in those with a particular genetic makeup. Rescue of all children from abuse is the only defensible child protection objective, but this is made more acute by the recognition that some of those children are genetically "primed" towards violence. Failure to rescue such children from abuse has a consequence in their and their victims' later misfortunes. Despite a sense of urgency for the work of the Moffitt-Caspi team to be incorporated into the child protection literature, as far as the writers have been able to determine after questioning relevant domain experts and practitioners, the implications of the Moffitt-Caspi work has not yet been translated into practice. So what are the implications for child protection? We may discard the notion of being indifferent to the parental practices imposed on those children whose genotypes afford them some protection against the acquisition of violent personalities! Rather it places an extra premium on ensuring as far as possible the quality of child care generally. Not to do so is effectively to collude in allowing preventable harm, which compounds the effects of the genetic lottery itself. A parallel may perhaps be drawn with phenylketonuria (PKU), where screening on neonates allows management and avoidance of the progressive mental impairment which otherwise ensues.

It has been speculated that the mechanism underlying the MAOA-abuse interaction may be epigenetic in nature. Epigenetics is concerned with the role that environments, including intrauterine environments, play in gene expression (see Francis 2011; Wortley 2011). This, both in the sense of responsivity to the cues which make primary crime reduction effective and in the particular sense of empathy (or its lack), will be a crucial point of departure for research which seeks to link applied criminology and evolution.

What has epigenetics got to do with evolution? The background to this question lies in very recent advances in genetics and related disciplines whereby much of what was hitherto regarded as “junk” DNA (i.e., DNA which does not code for proteins) in fact consists of a huge array of switches regulating gene expression. These switches can be thrown in pregnancy to yield enduring characteristics of the fetus, as was vividly revealed by the research following the Dutch “hungry winter” famine of 1944, in which the phase of pregnancy during which women went without food had long-term effects on the resulting child’s personality (Francis 2011).

Evolution requires many generations to yield change. It would aid a fetus’ prospects to get some idea of what awaits outside the womb. Epigenetic changes allow the possibility of switching genes on or off depending upon the clues which the environment gives. This makes evolutionary sense for both in utero and childhood clues. In the Caspi-Moffitt research, parental aggression is an early indicator of how aggressive one needs to be to prosper in the adult world. This form of gene regulation seems to be primarily under matrilineal control and has likely evolved partly to coordinate in utero development with maternal resource availability. The defensible next step will be to reanalyze longitudinal studies of criminality (the Cambridge study now incorporates data from three generations) to test epigenetic hypotheses. The literature on this topic continues to expand rapidly.

Evolution Without Tears: Natural Selection as Analogy

Assuming the reader is wholly unpersuaded by everything written here so far, the writer’s remaining hope is that evolution as analogy may be seen as an unthreatening heuristic device. The approach of selecting among varying designs on the basis of performance is an engineering technique known as genetic programming, the title brilliantly obscuring from those who would otherwise reject the approach on religious grounds the fact that it is in essence a simulation of biological evolution used in product design. In any event, evolution as analogy requires a user only to consider that it represents a way that improvement could occur, not necessarily the way it did occur. In this way, the theological teeth are drawn!

Perhaps the earliest proponent of this way of using Darwinian theory in relation to crime was Paul Ekblom, psychologist and polymath. He wrote in 1999:

> “But natural evolution is not simply a matter of ‘medieval warfare’ with increasingly better-armoured prey slogging it out against equally-improving armour-piercing capability of carnivores. The less dramatic struggle between plants and grazers is equally important (and may be a better model for property crime). The even longer struggle between *pathogens and immune system* has resulted in dynamic and adaptive strategies on each side. This has culminated in such sophisticated attackers as the HIV or smallpox viruses. Smallpox has about a hundred genes that interact with human defence mechanisms. In fact it has evolved *counter-countermeasures* to cope for example with a ‘virus alert’ chemical produced by infected cells, whose function is to warn nearby *uninfected* cells to activate their defences against virus attack.” (p. 29)

Ekblom details the links between other evolutionary struggles and specific crime types (see Table 1 below). The coevolutionary analogy has more recently been taken up by Raphael Sagarin and Terence Taylor in their book *Natural Security: A Darwinian Approach to a Dangerous World* (Sagarin and Taylor 2008) and applied to risk assessment in general and the threat of terrorism in particular.

Conclusions

Reluctance to apply an evolutionary perspective to problems of crime and criminality may be attributed to fundamentalist religious beliefs, the ascription of eugenic horrors of the past to Darwinian theory, and the notion of biological determinism erroneously laid at Darwin’s door. E. O. Wilson (1978) infamously wrote “genes hold culture on a leash” (p. 172). While containing a grain of truth, it turns out that there are multiple leads, very long and very twisted (see Pagel 2012).

Evolutionary Perspectives on Crime Prevention, Table 1 Some other evolutionary struggles

Realm	Struggle	Description and possible crime equivalent
The natural world	*Prey versus predators*	(Confronters, trappers, dupers), mainly resembling crimes against the person – assault, robbery, homicide
	Plant versus herbivore	Grazing- taking stored energy and materials from plants, resembling theft
	Host versus parasite	Parasitism by insects, tapeworms etc – resembling theft
	Host immune system versus pathogen	Infection by bacteria etc resembling robbery (overcoming host's defences)
	Host immune system versus viral pathogen	Infection by viruses, resembling fraud or embezzlement in misappropriation of resources for and control of production; computer hacking (breaking access and control codes), and computer viruses themselves
	Natural "*theft or robbery*"	Within or between species – eg birds taking each others' nest sticks, or robbing others' food in mid-air attacks
	Natural "*fraud*"	Birds taking nectar by pecking a hole in the side of the flower to avoid the effort required to pass on pollen, orchids pretending to be female wasps and cheating males of reproductive effort and opportunity
	Natural "*threat, assault*" *or killing*	Conflict over territory, mates, food
Humanity versus nature	*Disease control*	Hygiene, public health, inoculation, vaccination, antibiotics – resembling prevention of theft/robbery
	Pest control	Rats etc spoiling/stealing crops or livestock, spreading human diseases, acting offensively – resembling prevention of theft/damage. Disorder/ nuisance
The human world	*Military arms races and* (*counter*) *terrorism*	Arms versus armour, missiles versus electronic countermeasures, manoeuvrability – resembling assault and prevention of assault, homicide, disorder, theft of property, coercion, control of production
	War-games	Military training; evolution of new strategies in chess; computer-games of tactics and strategy
	Economic warfare	Outgrowing the enemy or disrupting their economy (shading into real crimes like forgery or extortion)
	Hacking	Shading into serious computer crime
	Espionage	Military/industrial, to steal information on resources, products, tactics and strategy, shading into theft of information/obtaining it in preparation for crime

There are enough results of interest produced by researchers operating within a Darwinian framework to offer recommendations for crime reduction policy, alongside social policy of other kinds. It turns out that these recommendations are by and large recognizably liberal. They include the need to nourish pregnant women well, to seek to ensure that adolescents have a realistically long life expectancy and some anticipation of later success in order to minimize early pregnancies and profligate and coercive sex. It emphasizes the absolute need to eliminate child abuse, given the genetic priming for violence of some so abused. It argues for recognition of the disincentive to report domestic abuse, given the preeminent family survival agenda of abused women. In the only clear enforcement item on the agenda, it argues for the particular monitoring of stepparents given their markedly greater prevalence of abuse against children within the home, always mindful of the fact that most stepparents behave well towards the children for whom they care. These recommendations represent the low-hanging fruit of the evolutionary perspective applied to crime, but given the marginality of such research within the discipline to date, it is surprising there is so much already. The wider question of whether the evolutionary perspective requires the abandonment of the discipline of criminology in favor of a perspective which encompasses the whole of the prosocial-antisocial continuum has been dodged here, but is probably crucial.

Related Entries

- ▶ Age-Crime Curve
- ▶ Domestic Violence
- ▶ Empathy and Offending
- ▶ Evolutionary Theories of Criminal Behaviors
- ▶ Genes, Crime, and Antisocial Behaviors
- ▶ Genetic Basis to Self-Control
- ▶ Innovation and Crime Prevention
- ▶ Moffitt's Developmental Taxonomy of Antisocial Behavior
- ▶ Prenatal and Postnatal Preventive Interventions Based on Risk Factors

Recommended Reading and References

Adler C and Polk K (2001) Child victims of homicide. Cambridge University Press, Cambridge

Brookman F (2005) Understanding homicide. Sage, London

Buss DM (2005) The murderer next door. Penguin, New York

Campbell A (2009) Gender and crime. In: Walsh A, Beaver KM (eds) Biosocial criminology: new directions in theory and research. Routledge, New York

Cashdan E (1993) Attracting mates: effects of paternal investment on mate attraction strategies. Ethol Sociobiol 14:1–23

Caspi A, McClay J, Moffitt TE, Mill J, Martin J, Craig I, Taylor A, Poulton R (2002) Evidence that the cycle of violence in maltreated children depends on genotype. Science 297:851–854

Coyne JA (2009) Why evolution is true. Oxford University Press, Oxford

Daly M, Wilson M (1988) Homicide. Aldine de Gruyter, Hawthorne

Daly M, Wilson M (1998) The truth about Cinderella: a Darwinian view of parental love. Orion, London

Daly M, Wilson M (2005) Carpe diem: adaptation and devaluing the future. Quart Rev Biol 80:55–60

Daly M, Wilson M (2007) Cinderella effect controversial? A case study of evolution-minded research and critiques thereof. In: Crawford C, Krebs D (eds) Foundations of evolutionary psychology. Lawrence Erlbaum, Mahwah

Darwin C (1871) The descent of man and selection in relation to sex. John Murray, London

Dunbar RIM (1992) Neocortex size as a constraint on group size in primates. J Human Evol 22(6):469–493

Ekblom P (1999) Can we make crime prevention adaptive by learning from other evolutionary struggles? Stud Crime Crime Prev 8(1):27–51

Francis RC (2011) Epigenetics: how environment shapes our genes. Norton, New York

Gluckman P, Hanson M (2006) Mismatch: the lifestyle diseases time bomb. Oxford University Press, Oxford

Griggs J (2012) Suicidal termites use chemical weapons to defend colony. New Scientist, July 26

Johnson S, Bowers K (2009) Permeability and burglary risk: are Cul-de-Sacs safer? J Quant Criminol. doi:10.1007/s10940-009-9084-8

Katzman MT (1980) The contribution of crime to urban decline. Urban Stud 17:277–286

Moffitt TE, Caspi A, Moffitt TE, Caspi A (2006) Evidence from behavioural genetics for environmental contributions to antisocial conduct. In: Wikström POH, Sampson RJ (eds) The explanation of crime: context mechanisms and development. Cambridge University Press, Cambridge

Nowak M (2011) Super Co_operators. Canongate, Edinburgh

Packer C, Pusey A (1983) Adaptations by female lions to infanticide by incoming males. Am Nat 121:716–728

Pagel M (2012) Wired for culture. Allen Lane, London

Sagarin R, Taylor T (2008) Natural security: a Darwinian approach to a dangerous world. University of California Press, Berkeley

Seligman MEP, Csikszentmihalyi M (2000) Positive psychology: an introduction. Am Psychol 55:5–14

Weisburd D, Groff E, Yang S (2012) The criminology of place: street segments and our understanding of the crime problem. Oxford University Press, Oxford

Wilson EO (1978) On human nature. Harvard University Press, Cambridge, MA

Wilson EO (2012) The social conquest of earth. Liveright, New York

Wilson J, Hernstein R (1985) Crime and human nature. Simon and Schuster, New York

Wilson M, Daly M (1997) Life expectancy, economic inequality, homicide, and reproductive timing in Chicago neighbourhoods. Brit Med J 314(7089): 1271–1274

Wortley R (2011) Psychological criminology. Routledge, London

Evolutionary Theories of Criminal Behaviors

Danielle Boisvert
College of Criminal Justice, Sam Houston State University, Huntsville, TX, USA

Overview

Rooted in Darwin's theory of natural selection, evolutionary explanations of antisocial and criminal behaviors have been studied extensively across various disciplines. This entry will focus specifically on the evolution of rape, spousal abuse, child abuse, and neglect, as well as homicide. In addition, a review of the most prominent evolutionary theories of antisocial and criminal behaviors, more broadly, is provided. These include cheater theory, r/K theory, conditional adaptation theory, alternative adaptation theory, staying alive hypothesis, and evolutionary neuroandrogenic theory. A review of this literature suggests that antisocial and criminal behaviors evolved as a way for our ancestors to overcome obstacles related to survival and/or reproduction. Incorporating evolutionary theories of antisocial behavior into mainstream criminological research has the potential to greatly enhance our current understanding of criminal behaviors.

Introduction

When studying criminal behavior from an evolutionary perspective, it is assumed that the behaviors considered "antisocial" today must have evolved because they had, at one point in time, served to increase our ancestors' likelihood of survival and chances of reproducing. Based on Darwin's (1859) theory of natural selection, certain selection pressures can lead to adaptations (physical, psychological, and behavioral) as a way to address a specific problem associated with survival and reproduction. Individuals who engaged in these adaptive behaviors were more likely to pass on their genes to future generations. Evolutionary theories of criminal behavior assume that the adaptive benefits to engaging in what we call "antisocial behavior" must have at some point outweighed the costs, at least in certain contexts and for certain people. While adaptations may have served to increase our ancestors' chances of survival and reproduction, not all adaptive behaviors from the past remain beneficial in today's environment. For example, the evolution of a "sweet tooth" may have been adaptive in ancestral environments as a way to recognize food that had valuable nutrients; today, however, this adaptation is associated with obesity, which is a leading cause of death (Wright 1994). Furthermore, adaptive behaviors from the past, such as aggressive/victimizing behaviors, may have increased an individual's chances of survival and procreation but are considered criminal and immoral by today's standards.

Several evolutionary theories have been developed to explain specific types of criminal behaviors, such as rape, spousal abuse, child abuse/neglect, and homicide. Other evolutionary theories have been proposed to explain criminal and antisocial behaviors more generally. These theories include cheater theory, r/K theory, conditional adaptation theory, alternative adaptation theory, staying alive hypothesis, and evolutionary neuroandrogenic theory. A review of these evolutionary theories and perspectives is provided next.

Rape

Perhaps one of the most controversial topics in evolutionary explanations of criminal behavior is rape. Research has clearly shown that males are more likely to be the perpetrator in rape cases compared to females. Similarly, rape (i.e., forced copulation) is not a human-specific behavior; it occurs in many other species, especially in those who are genetically similar to humans, such as orangutans (Ellis 1986). In 2000, Thornhill and Palmer presented an evolutionary explanation of rape in their book *A Natural History of Rape: Biological Bases of Sexual Coercion*. This book has received both positive and negative attention in the academic community. In discussing the ways rape could have evolved, Thornhill and Palmer present two distinct hypotheses: a rape-specific hypothesis and a by-product hypothesis.

First, the authors propose a rape-specific hypothesis, which states that males have evolved, through the process of natural selection, rape-specific behaviors. From this perspective, rape evolved in humans due to the tension between men and women in terms of their reproductive strategies. Specifically, males are able to reproduce with minimal to no parental investment and there is little restriction in the number of offspring they can produce. As a result, the male reproductive strategy is ultimately based on the number of fertile women he can mate with. Females, on the other hand, can only produce a limited number of offspring in their lifetime. Thus, their reproductive strategy is to choose mates carefully and to avoid mating with males who would provide little to no parental investment. Therefore, females are more selective in their mate choices (Thornhill and Palmer 2000).

Due to these sex differences in reproductive strategies, in some contexts, rape may have evolved as a conditional mating strategy for some men. For example, in the past, disadvantaged men who were unable to secure a voluntary mate, men who perceived a low cost of consequences (e.g., a woman who is alone and unprotected), and/or men who suspected that their partner was unfaithful may have resorted to forceful copulation (McKibbin et al. 2008). Overall, it is suggested that some males rape to increase their likelihood of reproductive success. As a result, their genes will be more represented in future generations compared to males who do not use force. If genes that promote rape have evolved to increase reproductive success, severe punishments would be needed to prevent their proliferation in future generations (Ellis 1998).

Next, Thornhill and Palmer (2000) propose a by-product hypothesis, which states that rape is a consequence of some other male sexual adaptations that served some other purpose. Specifically, a male's strong sex drive and desire to mate with several females may have accidentally led to rape. In other words, rape did not evolve as a specific reproductive strategy, but rather as a side effect of some other sexual adaptation (e.g., promiscuity, strong sex drive) that served to increase the chances of reproducing.

Although both hypotheses differ with regard to how/why rape evolved in humans, they both suggest that rapists are motivated by innate sexual desires. This perspective contrasts the feminist's view on rape as being motivated by a man's urge to control and dominate a woman and not by his need for sexual gratification (e.g., Brownmiller 1975). There are many critiques of Thornhill and Palmers's (2000) hypotheses, interpretations, and suggestions concerning the evolution of rape (see Coyne and Berry 2000; Lloyd 2001).

Spousal Abuse

Spousal abuse/assault can be viewed from an evolutionary perspective as an evolved form of aggressive behavior. While the research has predominantly focused on male abusers, females also commit a nontrivial amount of abuse toward their significant others (Dutton 2007). It has been suggested that spousal abuse/assault by women may have evolved as a way to discourage abandonment from a mate in an effort to retain his involvement in raising their offspring (Ellis 1998).

For men, one of the main motivators underlying spousal abuse and assault is sexual jealousy either as a result of infidelity or suspicions of infidelity (Daly et al. 1982). It has been suggested that there are at least two conditions under which it may have been adaptive, in ancestral environments, for men to abuse/kill their wives. These include cuckoldry as well as the possibility of losing a fertile partner to another man (Buss 2004). Cuckoldry refers to the parental investment in another person's offspring (Wilson and Daly 1993). Males are more likely than females to be cuckolded because they are not the ones who carry the child to term. Put differently, men are more likely to be cuckolded because, while a woman is always certain she is the parent, a man cannot be as sure. There are many consequences associated with being cuckolded including the following: (1) wasted time, effort, and resources in rearing another man's child; (2) the loss of time, effort, and resources invested in originally securing his sexual partner; (3) the loss of his partner's investment in any future children they could have produced together; as well as (4) reputational damage if others find out that he was cuckolded (Buss 2004).

It is clearly disadvantageous for a male to be parentally invested in an unrelated offspring as their own genes will not be passed on to future generations as a result of that investment. Due to uncertainty in paternity, males have evolved feelings of sexual jealousy and may take extreme measures to ensure that their mate is not sexually involved with other males (Wilson and Daly 1993). These extreme measures include physical abuse and violence as a way to either control a woman's sexual behavior, to intimidate a woman from sexually straying from the relationship or as a way to induce stress. Research has shown that stress can impact a woman's reproductive ability by lowering her chances of becoming pregnant as well as increasing her chances of miscarrying if pregnant. Thus, from an evolutionary perspective, it has been suggested that younger women with higher reproductive value are at greatest risk of domestic violence by males who are sexually jealous (Wilson and Daly 1993).

Child Abuse/Neglect

It would seem counterproductive, from an evolutionary perspective, to harm one's own child. There are, however, at least four conditions under which parental abuse, parental neglect, and/or even murdering one's own child could have evolved. First, if parents already have children who require care and resources, the addition of any new child will increase the likelihood of abuse and neglect as there is not enough time and/or resources to successfully care for multiple children. Therefore, if child abuse/neglect is an evolved adaptation, it would occur most often in larger families, with multiple births, and with few resources (Daly and Wilson 1988a; Ellis and Walsh 1997). These factors have in fact been found to be associated with child abuse/neglect. A second condition that could lead to child abuse/neglect is when one parent is not investing any time/resources into the care of their child. The absence of one parent can increase the likelihood that the remaining parent would also provide less resources and care toward their child (Daly and Wilson 1988a; Ellis and Walsh 1997). Research has shown that abuse is more likely to occur in single-parent households. Third, parents may be more likely to abuse/neglect their nonbiologically related children (e.g., stepchildren) compared to their biological children. This is because harming one's own biological child has greater repercussions by decreasing the likelihood of passing on familial genes to future generations (Daly and Wilson 1988a; Ellis and Walsh 1997). It is argued, however, that parents may also abuse/neglect their biological offspring if that child is perceived to be less likely to reproduce in the future, such as children who are ill or deformed (Daly and Wilson 1988a; Ellis and Walsh 1997). Overall, these four conditions are closely tied to circumstances where investing any time/effort into raising the offspring interferes with one's own potential of passing on genes to future generations.

Homicide

There are differing perspectives with regard to the evolution of homicide. Two in particular, set forth by Daly and Wilson (1988b) and Buss and Duntley (2002), have received quite a bit of attention in the evolutionary literature. First, Daly and Wilson (1988b) argue that homicide itself is not an adaptation, rather it is an unintended consequence, often referred to as an "overactive mistake," a "slipup," or a "dysfunctional by-product" of mechanisms evolved for some other nonlethal purpose (e.g., violence). For example, if a man kills his wife because of sexual jealousy, it is not because of an evolved psychological mechanism for murder, but rather it is a by-product of an evolved mechanism for violence. In this scenario, the man is intending to control his wife's sexual behavior through violence but it simply goes too far and results in death. This viewpoint, however, does not explain premeditated homicides. This led Buss and Duntley (2002) to propose that males have in fact evolved psychological mechanisms specifically designed to kill under certain circumstances.

Unlike Daly and Wilson (1988b) who suggest that homicide is a by-product of some other evolved response, Buss and Duntley (2002) state that homicide itself is an evolved response to solving adaptive problems and that the mechanisms underlying homicide are context specific. Their homicide adaptation/design theory acknowledges that, for the most part, the costs associated with killing another human far outweigh the benefits (Buss and Duntley 2002). However, they also recognize that in certain contexts, it may have been beneficial to kill another human. These circumstances are closely tied to situations where the killer may be (1) directly eliminating a competitor, which would result in greater reproductive advantages for the killer; (2) gaining access to the rival's material resources and fertile mates; (3) protecting not only himself but also his mates, offspring, relatives, and allies from being exploited, injured, raped, or killed by a rival; (4) gaining social status and the respect of others, thereby decreasing the chances of being exploited, injured, raped, or killed by others in the future; (5) protecting his territory, resources, shelter, and food from rivals; (6) eliminating the possibility of expending resources on individuals who are not genetically related, such as stepchildren; and/or (7) eliminating the possibility of expending resources on genetically related individuals who are incapable of themselves reproducing in the future, such as children who are ill or deformed. As such, under some circumstances, psychological adaptations for homicide may have evolved as the best strategy to solve adaptive problems (Buss and Duntley 2002).

Criminal Behavior in General

r/K Theory

MacArthur's (1962) r/K selection theory views reproductive strategies of organisms on a continuum with r- (growth rate) and K- (caring capacity) selection located at each end of the spectrum. First, those who reproduce in large quantities and provide little to no resources to raise their offspring are said to be r-strategists. These types of organisms can succeed in environments in which resources are plentiful and competition is low. On the other hand, organisms that follow a K-strategy will reproduce only a few times and will devote a great deal of time, effort, and resources into raising their offspring. This strategy is most often adopted in environments in which obtaining resources is competitive, and it serves to increase the likelihood that the few offspring will live on to reproduce (MacArthur 1962). In general, humans follow a K-reproductive strategy. However, there is a great deal of variation with regard to reproductive strategies within humans, with males being more r-selected than females due to their greater capacity to reproduce. There is also variation in reproductive strategies within individuals (Jolly 1985). Specifically, individuals tend to follow more of an r-strategy when younger but will eventually adopt a K-strategy as they age, especially if they have already produced children. This is because it is more beneficial to the parent, at a later point in their life, to devote their energy and resources to already existing children as opposed to putting effort toward gaining new mates in which to create new offspring with (Ellis 1987).

The r/K theory as it applies to criminal behavior suggests that criminals are more r-strategists compared to noncriminals who tend to follow a K-strategy. Ellis (1987) proposed that an r-selection approach to reproduction is highly intertwined with factors that are related to criminal behavior. Specifically, characteristics used to identify r-strategists are often the same ones associated with individuals who commit crimes against others. These factors include (1) large numbers of siblings within the home, (2) single-parent homes, (3) demographics (e.g., gender, age, and race), (4) prematurity and low birth weight, (5) short-life expectancy, (6) lower parental investment in child (i.e., child abuse and neglect), (7) early onset of sexual experiences, and (8) greater promiscuity. It is evident that these correlates of criminal behavior overlap with the description of an r-strategy approach to reproduction (Ellis 1987). r/K theory can be applied to explain criminal behavior in general as well as specific types of crimes, such as rape. For example, in the case of rape, an r-strategist may go to the extreme to achieve his goal of reproducing in large quantities with no or little investment in rearing offspring.

Cheater Theory

Cheater theory views criminal behavior as an evolved male strategy that is due to the distinct reproductive strategies employed by males and females (Machalek and Cohen 1991). For example, the mating strategy for males is based on the number of fertile females available to reproduce with. The more females they can have sex with, the greater their reproductive success. On the other hand, the number of times that females can produce offspring is limited by pregnancy, lactation, and menopause. This limitation makes females more selective when deciding on a sexual partner compared to males (Buss 1994; Wright 1994).

It is more beneficial for females to procreate with males who will provide resources and assistance in raising their offspring. However, males vary in terms of the amount of parental investment they will provide in raising their offspring, and their reproductive strategies have been termed "cads" and "dads" (Cashdan 1993). Cads are males who trick/force females to have sex with them and then leave to pursue additional women. This strategy is often used by adolescent males when pursuing a sexual partner and involves various techniques, such as (1) proclaiming love and commitment to a woman yet leaving once impregnated, (2) threatening and/or injuring rival males, (3) overstating their ability to provide resources by stealing and cheating others, and/or (4) resorting to the use of coercion/force to have sex. Cads tend to be self-gratifying and pleasure-seeking – characteristics that tend to overlap with antisocial individuals. The reproductive strategy used by cads will continue for males who are lacking in emotions (e.g., psychopaths) but for the most part will be replaced by strategies used by dads (Lykken 1995). Dads are the alternative male reproductive strategy to cads; these men help to raise their young by providing a great deal of parental involvement. This strategy is eventually adopted by most males in adulthood and is facilitated by social emotions, such as love (Walsh 1995). Overall, traits used to describe individuals who follow a cad strategy are similar to the traits used to describe antisocial individuals.

Conditional Adaptation Theory

While cheater theory focuses primarily on the reproductive strategies of males (i.e., cads and dads), conditional adaptation theory discusses female reproductive strategies as well (Walsh 2006). Specifically, conditional adaptation theory states that reproductive strategies vary based on environmental factors, such as the degree of paternal bonding that occurs early in the life course. The absence of a father in a child's life can have a profound effect on the reproductive strategy later employed by his offspring, particularly female offspring (Draper and Harpending 1982). For example, in unstable, resource-scarce environments where the quality of relationships is low and/or unpredictable, individuals are more likely to adopt an active and unrestricted reproductive strategy (Walsh 2006). For females, in particular, when early experiences with men are unstable (e.g., lack of attachment with father), she will be more likely to adopt a promiscuous sexual strategy. This is because her experiences with men tell her that there is a low probability that a man will invest much time in helping to raise their children (Hrdy 1999). Thus, in order to obtain short-term benefits from several men, a female is better off adhering to an unrestricted sexual strategy. On the other hand, individuals who have secure and stable relationships are more likely to adopt a restricted mating strategy and mate selectively. For young females, a strong relationship with their father shapes their perceptions of males as devoted caregivers. As a result, it is in the female's best interest to portray fidelity and to secure a sexual partner who is devoted to long-term care of her offspring (Walsh 2006). As it relates to criminal behavior, conditional adaptation theory proposes that when resources/attachments are unstable and low, there will be an increased likelihood that individuals will adopt a risky lifestyle.

Alternative Adaptation Theory

Alternative adaptation theory differs from conditional adaptation theory in that it incorporates individual genetic differences as a way to explain why some people choose one mating strategy over another (Rowe 1996; Walsh 2006). Similar to r/K theory, alternative adaptation theory views reproductive strategies on a continuum with one extreme representing mating efforts and the other extreme reflecting parenting effort. Mating effort is defined as the energy, time, and resources invested in the acquisition of a sexual partner (e.g., searching, courting, and procreating). Parenting effort, on the other hand, is defined as the time, energy, and resources invested in one's mate and offspring (e.g., caring, providing, protecting). Genetics, rather than the environment, will dictate where an individual falls along this continuum (Rowe 1996; Walsh 2006). Specifically, genetic differences that lead to increased mating efforts result in deception and criminality in the pursuit of mating opportunities. In other words, characteristics such as dishonesty, risk-taking, and self-gratification may help to secure sexual partners by putting effort into mating, and they are also useful in conducting criminal acts (Walsh 2006). On the other hand, traits such as compassion, selflessness, and caring are often found in individuals who devote much of their efforts toward parenting and who lead a prosocial lifestyle.

Alternative adaptation theory also incorporates the element of intelligence as an important factor in determining a person's reproductive strategy (Walsh 2006). Individuals who are more intelligent will be more likely to invest their time and effort into raising their children. These individuals are also better equipped at delaying gratification and living a prosocial life. Conversely, individuals who are less intelligent are more likely to devote their time and energy toward obtaining sexual partners. These individuals seek immediate gratification and give little to no thought to the consequences of their actions (Walsh 2006). Overall, from an evolutionary perspective, the genetically influenced traits used to characterize an individual who tends toward the mating end of the continuum are the same traits used to explain antisocial individuals.

Campbell's Staying Alive Theory

Campbell's (1999) staying alive theory focuses primarily on female criminal behaviors, or more accurately attempts to explain why women do not engage in as much criminal behavior as men do. Campbell (1999) proposed that females have evolved to avoid engaging in risky types of activities because their survival is critical to the survival of their offspring. In other words, greater female investment in their offspring leads them to avoid risky situations and therefore less likely to engage in criminal activities. By ensuring their own survival, women are increasing the likelihood that their children will survive and that their genes will be passed on to future generations. As such, a woman's reproductive success is dependent on her ability to "stay alive" by avoiding dangerous situations. The mechanism by which females have evolved to avoid risky situations is said to be through fearfulness (Campbell 1999). Specifically, even though women need resources to support their reproductive efforts (i.e., ovulation, pregnancy, lactation) and to provide for their developing child, fear counteracts the pressure on women to physically compete for resources themselves as the risks of injury/death are too high and costly. As an alternative, women tend to compete with other women to secure a partner who can provide for them and their children. Their success in securing a partner hinges primarily on the male's perception of her fidelity, which will impact his certainty in paternity and subsequently the degree to which he will invest in his offspring (Campbell et al. 2001). Overall, Campbell's staying alive theory hypothesizes that sex differences in criminal involvement are due to sex differences in fear threshold, which may have evolved due to the selected pressures on women to avoid risky situations in order to increase their chances of survival and the likelihood that their offspring will survive as well.

Evolutionary Neuroandrogenic Theory

Ellis (2000, 2005) developed his evolutionary neuroandrogenic theory (ENA) in order to explain the higher prevalence of competitive/victimizing behaviors displayed by men compared to women. While the evolutionary component of his theory seeks to explain why males tend to commit crimes at higher rates than females, the androgenic component focuses on how male sex hormones affect the male's brain to promote criminal behaviors, especially during the teenage years.

The first component of ENA theory involves the evolutionary process. Based on Darwin's theory of natural selection, biological and physical characteristics that increase the likelihood of survival also tend to increase the likelihood of reproducing (Darwin 1859). These characteristics are in turn passed on to future generations. For example, the strongest and most competitive organisms of a given species are more likely to survive and thus produce offspring, while the weaker and more timid organisms are more likely to die before they have the chance to procreate. Similar to natural selection, sexual selection also promotes the transmission of certain characteristics. For example, competitiveness in males can be considered a desired and sought after characteristic for females who are looking to reproduce. Therefore, males who can demonstrate this desired trait would be more attractive mating partners. This process is often referred to as "female choice."

Female mating preferences can explain, in part, why males typically engage in higher rates of aggressive behaviors. Specifically, women are more likely to mate with males who are reliable providers of necessary resources. In other words, a female is more likely to choose a sexual partner based on his ability to compete for resources and provide for both her and her children. In turn, this allows the woman to focus on caring for and raising children as opposed to spending her energy on obtaining resources. Therefore, the female choice phenomenon has the effect of encouraging competitive and victimizing behaviors in males. For example, males will aggressively compete with other males, even to the extent of victimization, in order to procure resources and increase their chances of being chosen by a female as a sexual partner (Ellis 2000, 2005). Since the majority of females prefer competitive males, genes which promote competitive/victimizing behaviors are more likely to be passed on from generation to generation. This theoretical perspective helps to explain why males are, in general, more prone to committing aggressive crimes compared to females.

The second component of ENA theory involves male sex hormones (i.e., androgens), particularly testosterone. Sex differences in testosterone levels may help to explain why males are more likely to engage in aggressive and victimizing behaviors compared to females (Ellis 2000, 2005). This component of ENA theory explains how increased levels of testosterone affect brain functioning in three ways: (1) suboptimal arousal, (2) susceptibility to seizures, and (3) rightward shift (Ellis 2000, 2005). First, exposing the brain to testosterone, both at the perinatal and pubertal stages, results in decreased brain sensitivity to incoming environmental stimuli. This reduction in sensitivity, often referred to as suboptimal arousal, increases the likelihood that an individual will exhibit criminal forms of competitive/victimizing behavior. Second, high levels of testosterone in the brain increase the susceptibility of the limbic system to experience seizures. Minor forms of seizures, known as episodic dyscontrol and limbic psychotic trigger action, occur in areas of the brain that regulate emotions, which can result in impulsive bursts of rage, especially when an individual is exposed to stressful events (Ellis 2000, 2005). Thus, increased levels of testosterone in the brain increases the brain's susceptibility to seizures, thereby resulting in aggressive outbursts. Finally, increased levels of testosterone shifts the brain's functioning away from the left hemisphere of the brain toward the right hemisphere (Ellis 2000). This results in individuals having better spatial reasoning abilities while simultaneously lacking in language-based reasoning, thereby increasing the probability of engaging in antisocial forms of competitive/victimizing behaviors.

While these three biological factors operate to promote competitive/victimizing behaviors, two neurological factors function to inhibit criminal forms of competitive/victimizing behaviors. These two factors include an individual's ability to learn (i.e., intelligence) as well as their ability to plan and foresee consequences (i.e., executive functioning). Specifically, greater mental capabilities and effective executive functioning capabilities can result in an increased likelihood of exhibiting more sophisticated and socially acceptable forms of competitive behaviors (e.g., becoming a competitive businessman)

(Ellis 2000, 2005). Conversely, males are more likely to engage in criminal activities when they fail, as the result of insensitivity to punishment and low IQ, to learn competitive strategies that avoid retaliation. Overall, ENA theory states that biological factors have evolved to promote competitive behaviors in males and, in turn, competitive behaviors have become pronounced within the human species because women prefer mating with men who are secure providers. The theory further assumes that most forms of criminality are "crude" expressions of male competitiveness.

Conclusion

Evolutionary theories have been used to explain a wide range of behaviors deemed antisocial by today's standards. A review of these theories and perspectives suggests that antisocial/criminal behaviors evolved as a way for our ancestors to overcome obstacles related to survival and/or reproduction. Whether the criminal behavior itself (e.g., rape, murder) is the adaptation or whether it is a by-product of another adaptation remains debatable. In a way, antisocial/criminal behaviors are said to be maintained over time through the process of natural and/or sexual selection. Future research should continue to investigate the evolved psychological mechanisms that motivate antisocial/criminal behavior. It is only through rigorous empirical studies that incorporate both ultimate and proximate explanations of antisocial/criminal behavior will we gain a greater understanding of these types of behaviors. In turn, a deeper understanding of the causes of antisocial/criminal behaviors can serve to better inform intervention and prevention efforts.

Related Entries

- ▶ Genes, Crime, and Antisocial Behaviors
- ▶ Hormones, the Brain, and Criminality
- ▶ Neurology and Neurochemistry of Crime
- ▶ Synthesizing Biological and Social Theorizing

Recommended Reading and References

Brownmiller S (1975) Against our will: men, women and rape. Simon Schuster, New York

Buss DM (1994) The evolution of desire. Basic Books, New York

Buss DM (2004) Evolutionary psychology. Allyn & Bacon, Boston

Buss DM, Duntley JD (2002) Murder by design: the evolution of homicide. Unpublished manuscript

Campbell A (1999) Staying alive: evolution, culture, and women's intrasexual aggression. Behav Brain Sci 22:203–252

Campbell A, Muncer S, Bibel D (2001) Women and crime: an evolutionary approach. Aggress Violent Behav 6(5):481–497

Cashdan E (1993) Attracting males: effects of paternal investment on mate attraction strategies. Ethol Sociobiol 14(1):1–23

Coyne JA, Berry A (2000) Rape as an adaptation: is this contentions hypothesis advocacy, not science? Nature 404:121–122

Daly M, Wilson M (1988) Evolutionary social psychology and family homicide. Science 242(4878):519–524

Daly M, Wilson M (1988b) Homicide. Aldine de Gruyter, New York

Daly M, Wilson M, Weghorst SJ (1982) Male sexual jealousy. Ethol Sociobiol 13(1):11–27

Darwin C (1859) On the origin of species. Greenwood Press, Westport

Draper P, Harpending H (1982) Father absence and reproductive strategy: an evolutionary perspective. J Anthrop Res 38(3):255–273

Dutton DG (2007) Female intimate partner violence and developmental trajectories of abusive females. Int J Men's Health 6(1):59–70

Ellis L (1986) Evolution and the nonlegal equivalent of aggressive criminal behavior. Aggress Behav 12(1):57–71

Ellis L (1987) Criminal behavior and r/K selection: an extension of gene-based evolutionary theory. Deviant Behav 8:149–176

Ellis L (1998) Neo-Darwinian theories of violent criminality and antisocial behavior: photographic evidence from nonhuman animals and a review of the literature. Aggress Violent Behav 3(1):61–110

Ellis L (2000) Theories of criminal/antisocial behavior. In: Ellis L, Walsh A (eds) Criminology: a global perspective. Allyn & Bacon, Boston

Ellis L (2005) A theory explaining biological correlates of criminality. Eur J Criminol 2(3):287–315

Ellis L (2006) Evolution and the nonlegal equivalent of aggressive criminal behavior. Aggress Behav 12(1):57–71

Ellis L, Walsh A (1997) Gene-based evolutionary theories in criminology. Criminol 35(2):229–276

Hrdy SD (1999) Mother nature: a history of mothers, infants, and natural selection. Politics Life Sci 20(2):107–252

Jolly A (1985) The evolution of primate behavior. Am Sci 73(2):230–239
Lloyd EA (2001) Science gone astray: evolution and rape. Mich Law Rev 99(6):1536–1559
Lykken DT (1995) The antisocial personality. Lawrence Erlbaum Associates, Hillsdale
MacArthur RH (1962) Some generalized theorems of natural selection. Natur Acad Sci 48(11):1893–1897
Machalek R, Cohen LE (1991) The nature of crime: is cheating necessary for cooperation? Human Natur 2(3):215–233
McKibbin WF, Shackelford TK, Goetz AT, Starratt VG (2008) Why do men rape? An evolutionary psychological perspective. Rev Gen Psychol 12(1):86–97
Rowe DC (1996) An adaptive strategy theory of crime and delinquency. In: Hawkins DJ (ed) Delinquency and crime. Cambridge University Press, New York, pp 268–314
Thornhill R, Palmer C (2000) A natural history of rape: biological bases of sexual coercion. MIT Press, Cambridge, MA
Trivers RL (1971) The evolution of reciprocal altruism. Q Rev Biol 46(1):35–57
Walsh A (1995) Parental attachment, drug use, and facultative sexual strategies. Biodemography Soc Biol 42(1–2):95–107
Walsh A (2006) Evolutionary psychology and criminal behavior. In: Barkow JH (ed) Missing the revolution. Oxford University Press, New York, pp 225–268
Wilson M, Daly M (1993) An evolutionary psychological perspective on male sexual proprietariness and violence against wives. Violence Vict 8:271–294
Wright R (1994) The moral animal: the new science of evolutionary psychology. Pantheon Books, New York

Evolving Judicial Roles

Victor Eugene Flango
National Center for State Courts, Williamsburg, VA, USA

Synonyms

Impartiality and fairness in judging; Job of a judge; Judicial temperament; Key skills involved in judging

Overview

The role of courts has evolved over time and the role of judges has adapted to match the changing role of courts. This entry discusses the role of the judge as originally conceived in the context of the adversary system and how that role has adapted to meet the needs of high-volume and problem-solving courts. The entry ends with a discussion of the importance of procedural fairness and how it comports with the three roles of a judge discussed here: the umpire, the adjudicator, and the problem solver.

Fundamental Issues

> . . .the way in which a judge conceives his judicial role is the most significant single factor in the whole decisional process. Kenneth Dolbeare

The role of courts has evolved over time and the role of judges has adapted to match the changing role of courts. Courts in the Common Law countries began using the adversary process, which shaped many of our expectations of what courts should be and therefore what judges should do. The adversary process with its emphasis on public trial is too slow and costly to be used to resolve the vast volume of ordinary cases, but it is still used for the “important” large-stakes cases as well as the “notorious” cases – mostly criminal cases with a large enough public interest to be covered in the news and perhaps televised. As case volume increased, a more streamlined and disposition-oriented process supplanted the adversary process for the ordinary cases. The role of the judge for these cases then becomes more of a decision-maker than umpire. The third stage of the evolution began relatively recently with the advent of the more treatment-oriented problem-solving court process and the consequent more active involvement of the judge as treatment team leader.

Henderson et al. (1984) identified three separate adjudication processes – procedural, decisional, and diagnostic – and noted that *all* three adjudicatory processes coexist and are used by contemporary courts in varying combinations. Flango and Clarke (2011) go a step further and suggest that court structures be modified to be more congruent with adjudication processes in use. It may be that the creation of different types of courts was an early attempt to separate the three functions – courts of general jurisdiction using the adversary process to resolve more serious cases, such as felonies; courts of limited jurisdiction using a more bureaucratic process to resolve the large volume of less serious cases, such as misdemeanors; and courts of special jurisdiction, such as juvenile courts, using a problem-solving adjudicatory process. In any event, different adjudication processes require different judicial attributes and skills – a confusion that becomes manifest when litigants come to court expecting an adversary process with an aloof judge and find themselves in a problem-solving court with an actively engaged judge.

This entry begins with the "classic" conception of a court – the neutral, aloof "umpire" judge using an adversary process to resolve cases – and then proceeds to elaborate on the temperament and skills required of judges using the more disposition-oriented and problem-solving court processes. This entry concludes by asking whether the new emphasis on procedural fairness can bridge the gap among the three roles.

The Umpire

The classic image of judge as neutral arbiter has its roots in the adversary system. The very conception of courts, and therefore the expectations we have of them, is derived from the adversary process, more particularly the criminal trial. Is the following image not what comes to mind when the word "court" is mentioned?

> There is a definite image about every aspect of the trial even what the courtroom is supposed to be like; the jury sits in its box, the judge sits on his or her high bench in a robe with an American flag in the background, the witnesses come in and sit on the witness chair; they raise their right hands and swear; the lawyers and the defendant sit at tables facing the judge. The trial begins with elaborate voir dire – the meticulous process of selecting a jury. The lawyers battle and squabble, trying to stack the jury with people they feel will vote the way they want. The trial itself gets going with opening arguments and statements from the lawyers. The trial itself is long, tense, and full of excitement. The lawyers joust with each other. There is clever and dramatic cross-examination. Lawyers jump up and cry, "I object";. . . . . .They end the trial with impassioned arguments. Then the judge instructs the jury, the jury retires to a locked room, and a spine-tingling period of waiting begins. Finally, the door opens, a hush comes over the crowd in the courtroom, and the jury comes in and announces its verdict. (Friedman 2004, p. 690)

This expectation of what a court is like, and what the role of the judge is, is derived much more from motion pictures and television programs than from court observations. These types of full-blown trials with all the trimmings do exist today, but for a very small percentage of cases. Yet this image provides the standard against which all courts are judged.

The adversary process assumes that there are two sides to the case – perhaps rooted in the medieval process of trial by combat (Strick 1977, p. 21). The core of the adversary system is the *form* of participation accorded to the parties. Lon Fuller (1960, p. 2) defines these as the "institutionally protected opportunity to present proofs and arguments for a decision in his favor." The government in a criminal case asserts a claim which the defendant denies. Each side has the opportunity to present arguments in his or her favor. Logically, the requirement for the participant to be able to provide proofs and arguments requires a neutral arbiter before whom to present the arguments and a set of standards or laws so that the litigants know the basis upon which the decision will be made.

Accordingly, the role of the judge in the adversary process is to preside over the proceedings and maintain order. During a trial, the judge rules on whether any of the evidence the parties want to use is illegal or improper. If the trial is before a jury, the judge gives instructions about the law that applies to the case; if the trial is before the court, the judge determines the facts and decides the case. After the trial, bench, or jury, the judge metes out the sentence to the convicted. Note also the role of the judge in this idealized conception – a very passive umpire enforcing the procedural rules of the game. In the language of modern confirmation hearing, the judge role is a passive referee who just “calls balls and strikes” (Roberts 2005).

Adversary proceedings are especially important when contending parties are unequal in power, e.g., the government versus an individual. The need for an impartial forum to equalize the contending parties is paramount. Lieberman (1981, p. 169) notes that “for certain types of problems, an adversary system is inescapable” and for a society with a commitment to political freedom,

> . . . putting the state to an extreme burden of proof and by guaranteeing defendants access to fiercely independent lawyers, we can in general prevent the state from imprisoning those whom it distrusts or fears.

For courts to be impartial, judges must be free to decide cases based upon the laws and facts of the case uninfluenced by *either external pressures or internal preferences*. Impartiality is impossible unless judges are independent – free from the external pressures of threats, intimidation, or fears of sanctions based upon the content of their decisions. In some places in the world, threats to impartiality can range from direct and drastic, such as threats on a judge’s life, to more subtle pressures, such as denial of salary increases, promotions, and staff or equipment needed to do the job.

Impartiality is also threatened to the extent that judges permit their personal conceptions of justice to enter into the decision-making equation rather than setting aside personal predispositions in deference to law as written. One object of law school is to socialize potential judges to defer to law rather than relying on their own conceptions of justice. This “precedent orientation” taught in law school makes judges feel bound by previous decisions made by other judges (Becker 1964). The role of the judge is to apply the law to different sets of facts raised in various cases and to rule accordingly. This “law applier” role is important to litigants because following precedent leads to consistency in decision-making, and therefore makes outcomes more predictable. About half of the judges responding to various questionnaires by different researchers consider themselves law appliers, implying that most cases can be decided by analogy to cases decided earlier (Flango et al. 1975, p. 285).

One way to safeguard judicial impartiality is the practice of rotating judges among judicial assignments – even among civil, criminal, and family law cases.

The Adjudicator

Role orientations are beliefs about “the kind of behavior proper for a judge” (Gibson 1978, p. 918). The role of the judge as umpire, primarily presiding over trials, gives way to reality for observers more familiar with the actual court processes. The criminal justice system would simply break down if most cases went to trial. According to the National Center for State Courts’ Court Statistics Project, there were over 21 (21.3) million criminal cases filed in state courts in 2008, and that is an undercount because of underreporting from some states. One look at these numbers confirms that a full-blown trial is not now, nor ever has been, a practical way to resolve most criminal cases – even most felonies. In 1976, only 7.6 % of criminal cases were resolved by trial, and that proportion has been declining steadily so that by 2008 that percentage dropped to 2 %. Trials are the last resort – the ultimate proceeding used when all other efforts at reaching an agreement have failed.

In most criminal cases, the prosecutor is the key de facto decision-maker because of the important role in screening cases in effect determines the resolution of most criminal cases. Prosecutors decide *whether* charges will be brought or dismissed, *which* charges will be brought, *how many* counts of each charge will be brought, and what will be offered in return for a guilty plea (Eisenstein 1973, p. 103; Blumberg 1967, p. 58). Plea agreements are how most cases are resolved now and have been since they replaced jury trials in the middle of the nineteenth century (Moley 1929, Chap. 7; Friedman 1993, p. 253; Fisher 2003). Before plea bargaining, most criminal cases were handled in summary fashion. Trials then did not resemble the full adversary process we expect today but were more “quick, slapdash” proceedings, where 12 people, impaneled without voir dire, sat for a series of "trials" (Friedman 2004, p. 692). Most defendants were not represented by counsel, there was little or no cross-examination and few objections, and the jury did not spend much time in deliberation.

The role of the judge in most criminal cases then is one of ratifying agreements reached between prosecutor and defense counsel. This is not a trivial function because it does provide a safeguard that the plea agreement was reached using a fair process, i.e., without coercion of the defendant.

If these aspects of a bureaucracy are true for state courts of general jurisdiction, they are even more applicable to courts of limited jurisdiction which resolve less serious, but more numerous, criminal offenses. Misdemeanor cases are now and have always been handled quickly and summarily without much technicality (Friedman 2004, p. 693). Approximately 80 % of criminal cases are misdemeanors, and most of them (more than 70 %) are handled by municipal judges, justices of the peace, or magistrates in courts of limited jurisdiction. (Even that 70 % is an underestimate because ten states plus the District of Columbia and Puerto Rico are unified and thus do not distinguish courts of general jurisdiction from courts of limited jurisdiction. Unified courts do, however, have a separate category of judge, e.g., associate judges, to handle misdemeanors and traffic cases.) In the sense that these lower criminal courts hear the bulk of criminal cases, including disorderly conduct, drunkenness, prostitution, petty theft, and simple assault cases, they *are* the courts with the most contact with offenders, and it is in these courts that the stereotype of "assembly line" justice was created (Feeley 1979). One Albany lawyer (Redlich 2007) describes in his blog the situation in the lower courts of New York:

> The biggest problem with our court system is the volume of cases. The volume is so large that the courts have to rely on assembly line justice. It really is an assembly line. The police officer prepares the initial papers and files them with the clerk. The clerk gives the papers to the prosecutor who reviews them and discusses the case with the lawyer or the pro-se defendant. The papers then go back to the clerk, who then hands them to the judge. The judge calls the case. There's a brief discussion at the bench. Then the papers go back to the clerk, who then processes the result (fine notice, schedule next date, etc.).
>
> Think about this: If a court has 100 cases on for a particular session (a typical number for courts like Colonie, Guilderland, Albany, etc), and each case takes 15 minutes, that would take 25 hours. That's not going to work. If each case takes only 5 minutes, it still takes 8 hours, so that's still not going to work. Most courts end up at about 1-2 minutes per case. That's assembly line justice.

These types of cases require facts to be established so that the law can be quickly applied. Sentences and financial penalties are limited so that dispositions can be expeditious (Henderson et al. 1984, p. 11). Glick (1971, pp. 30–34) divided this judicial role into two parts, which could be viewed as complementary sides of the same coin – the "adjudicator" who emphasized deciding cases and the "task performer" who emphasized processing litigation and maintaining smooth court operations. The "disposition-oriented" or "administrator" role of this judge is to skillfully apply judicial procedures to achieve the swift and consistent disposition of cases (Vines 1969; Ungs and Baas 1972). Clearing the docket then becomes very important, and the task becomes to process large number of individual cases, a more bureaucratic process not unfamiliar to the administrative agencies in the executive branch of government. Judges must decide large numbers of lower-stakes cases every day, rather than spending days or weeks making a decision in one case at trial, and so the procedures must be streamlined. Consequently, judges may take a more active role in all phases of case processing to move the case along while ensuring that the attorneys, many of whom may be court appointed, are devoting the proper attention to their clients. Judges simply cannot rely on parties to frame disputes. Overcrowded dockets and "overzealous litigants" lead judges to take a more "active, largely discretionary approach to pretrial case management" (Molot 2003, p. 29). Federal judges have described their involvement in pretrial management as "moderate" or "intensive," including "holding pretrial conferences, setting pretrial schedules and trial dates, setting limits on discovery and ruling on motions" (Lande 2005). Boyum (1979) found that judges who emphasized the "administrator" role did indeed take a shorter amount of time, on average, to resolve their cases. Timely resolution is a positive outcome if it does not inhibit litigants from seeking information, including clarification and follow-up questions, or make them feel that their concerns are not taken seriously.

The Problem Solver

In 1984, then Chief Justice Warren Burger said:

> The entire legal profession. . .has become so mesmerized with the stimulation of the courtroom contest that we tend to forget that we ought to be healers of conflicts. For many claims, trials by adversarial contest must in time go the way of the ancient trial by battle and blood. (Toohey 2010, p. 7)

Spurred by the perceived inadequacies of the adversary process, some legal leaders have promoted a more cooperative approach to dispute resolution. Particularly in family law, once a fertile source of trials, there have been calls to abandon adversarial proceedings "in favor of more informal approaches with the goal of encouraging parts to develop positive postdivorce co-parenting relationships" (Murphy 2010, p. 895).

Some types of cases, such as those involving juveniles, have never fit comfortably within the traditional law-court framework. Separate courts for juveniles were created first in Chicago in 1899 (Stevenson et al. 1996, pp. 5–6) and served as a model for the later development of specialized family courts, which have jurisdiction not only over juvenile cases but also of other cases affecting the family from divorce to domestic violence. More recent "problem-solving" courts originated from the efforts of "practical, creative, and intuitive judges and court personnel, grappling to find an alternative to revolving door justice, especially as dispensed to drug-addicted defendants" (Winick and Wexler 2003, p. 6). From the opening of the first drug court in Dade County, Florida in 1989, drug courts spread rapidly based upon anecdotal reports of success in reducing recidivism, as well as the infusion of federal dollars (Berman and Feinblatt 2001, p. 23). By the end of 2009, there were 2,459 drug courts and an additional 1,189 problem-solving courts in the United States, including courts for DWI, mental health, domestic violence, truancy, child support, homeless, prostitution, reentry, and gambling (Huddleston and Marlowe 2011, p. 1). The conventional term "problem-solving courts" has passed into the language even though most are not separate courts but separate dockets or calendars of larger courts or divisions. In most instances, they involve a single judge handling a single type of case on a periodic schedule.

Problem-solving courts require judges to be more active, less formal, and personally engaged with each offender, and this personal involvement creates a tension with the traditional role of the judge as a detached, neutral arbiter. Indeed, Hanson (2002, p. 10) sees the problem-solving judge going beyond Becker's "situation-oriented" judge by not only taking into account the impact of their decisions on the litigants and perhaps the public at large but by becoming a partner in the "therapeutic-oriented response" to ameliorate the underlying problems of litigants. This tension the "polar-opposite roles" provides the base for the charge that problem-solving judges need to become "social workers" or "therapists." One *New York Times* article (Eaton and Kaufman 2005) summarized:

> The judges often have an unusual amount of information about the people who appear before them. These people, who are often called clients, rather than defendants, can talk directly to the judges, rather than communicating through lawyers. And the judges monitor these defendants for months, even years, using a system of rewards and punishments, which can include jail time. Judges also receive training in their court's specialty and may have a psychologist on the staff.

The collaborative nature of the problem-solving-court approach sometimes raises questions about the impartiality of the judge. For example, problem-solving judges need to praise and sanction defendants, rather than remain aloof, but this active engagement could create the perception that they are not impartial. Some may also consider collaboration in "staffings," where the judge and treatment team meet in advance of hearings to discuss the offender's progress in treatment and to reach consensus about rewards and sanctions to be in conflict with the judicial role.

The counter position, as presented in the commentary to the ABA Model Code of Judicial Conduct, states that judges in problem-solving courts may be authorized by court rules to act in nontraditional ways. The new ABA Model Code states that a judge may "initiate, permit, or consider" ex parte communications expressly authorized by law, such as when serving on therapeutic or problem-solving courts. The collaborative decision-making process "does not violate the judge's duty of independent judgment so long as the final decision rests with the judge" (Freeman-Wilson et al. 2001, p. 3). Judicial ethics do not require disengagement but impartiality, and so a judge may show concern about recovery, even celebrate the successes, but must be equally concerned about the progress of each offender.

Problem-solving courts have raised related issues about the proper judge's role outside the courtroom and outside the context of dealing with an individual case. Judges presiding in drug courts and monitoring the progress of participants in those courts' programs may be authorized and even encouraged to communicate directly with social workers, probation officers, and others outside the context of their usual judicial role as independent decision-makers on issues of fact and law. Furthermore, judges are in a good position to advocate for court reform in general because of their respected position in the community. Section 4(B) of the Canons of Judicial Ethics permits judges to act as educators (Freeman-Wilson et al. 2001, p. 13). This is especially important for judges in problem-solving courts who want to build public support for treatment-oriented programs. Often education involves Illustrations using success stories, which is ethical as long as confidentiality is not breached or specific individuals identified.

Procedural Fairness

Are there any attributes of judges that transcend the three roles mentioned above? In an essay on "What Makes a Good Judge?" Sir Gary Higginbottom (2011) notes that the attributes of a good judge include professional, personal, and administrative components. Professional attributes include knowledge of the law, legal analytic skills, "good judgment," and intellectual concentration, whereas personal attributes include such qualities as integrity, objectivity, and temperament. Many other judges have provided summaries of the qualities of a good judge, and these usually include professional competence (legal abilities and intellect), integrity, and judicial temperament (neutral, decisive, respectful, and composed). Note that these qualities are ambiguous, amorphous, and hard to define.

Judges Burke and Leben (2007, p. 4) present a powerful case for the principles of "procedural fairness" as a construct to define the desirable qualities of a judge and to make them measurable.

> Judges must be aware of the dissonance that exists between how they view the legal process and how the public before them views it. While judges should definitely continue to pay attention to creating fair outcomes, they should also tailor their actions, language, and responses to the public's expectations of procedural fairness. By doing so, these judges will establish themselves as legitimate authorities; substantial research suggests that increased compliance with court orders and decreased recidivism by criminal offenders will result. Procedural fairness also will lessen the difference in how minority populations perceive and react to the courts.

Although the lists of characteristics are worded differently at times, the four key expectations of procedural fairness are:

1. *Voice*: The ability to participate in the case by expressing their viewpoint.
2. *Neutrality*: Consistently applied legal principles, unbiased decision-makers, and a "transparency" about how decisions are made.
3. *Respectful treatment*: Individuals are treated with dignity and their rights are obviously protected.

4. *Trustworthy authorities*: Authorities are benevolent, caring, and sincerely trying to help the litigants – this trust is garnered by listening to individuals and by explaining or justifying decisions that address the litigants' needs (Burke and Leben).

These principles are laudable. The principle of neutrality, however, suggests that they were developed with the image of a traditional trial court in mind and are therefore most applicable to courts using the traditional process of procedural adjudication. Procedural fairness meshes very well with the umpire role of judges.

Procedural fairness would appear to be tailor-made for the problem-solving courts as well, especially the principles of voice, respectful treatment, and trustworthiness. Clients often communicate with the judge directly rather than through an attorney, for example. Neutrality is more of a stumbling block in these courts, however. Problem-solving courts require judges to be more active, less formal, and personally engaged with each offender, and this personal involvement creates a tension with the traditional role of the judge as a detached, neutral arbiter. This collaborative venture means that problem-solving judges are not neutral – they are rooting that treatment succeeds. They need to praise and sanction defendants, rather than remain aloof, but this active engagement could create the perception that they are not impartial. Some may also consider collaboration in "staffings," where the judge and treatment team meet in advance of hearings to discuss the offender's progress in treatment and to reach consensus about rewards and sanctions, to be in conflict with the role of judge as umpire.

There is a way out of this dilemma; however, problem-solving methods could be used post-adjudication only. Especially in criminal cases with a substance abuse component, such as DWI cases, the full adversary process with all of its due process protections could be employed until guilt has been established. After guilt is established, problem-solving principles designed to prevent repeat offenses could be used to select the best sentencing options, whether they be therapeutic or punitive. If problem-solving processes were used primarily post-adjudication, procedural fairness is an amenable tool for problem-solving judges as well.

Procedural fairness principles are not compatible with the adjudicator role, however. When clearing the docket is the goal, processing large numbers of cases expeditiously is the key requirement. Consequently, judges must consider not only how to offer each client due process rights but how to do so and still resolve the cases quickly. Consequently, they must take a more proactive role in all phases of case processing. The point here is with that many cases to resolve in such a short time, can lower court judges really be expected to provide litigants with meaningful *voice* – the ability to participate in a case by expressing their viewpoint – and still keep ahead of their dockets? Is there time to express their caring to explain and justify their decisions? In sum, is procedural fairness possible in high-volume lower courts? Because judges have such short interactions with litigants in these high-volume courts, the way litigants are treated by court staff becomes more important. There is no reason why procedural fairness principles could not be employed by court staff as well as by judges. Court staff includes not only clerks but also security personnel, bailiffs, administrative staff, and any other staff members that interact with the public (Porter 2011).

This entry has suggested that the various processes by which courts resolve cases today be acknowledged explicitly and matched with the judicial role most appropriate for each process. All judges need not have the "umpire" role as the dominant model, but it is the role most appropriate for the adversary process. Similarly, "adjudicators" with case processing abilities and talents are most appropriate for courts handling the large bulk of the cases, and "problem-solving judges" are most appropriate for treatment-oriented courts. The principles of procedural fairness can improve the sensitivity of all judges, especially trial judges, but are least applicable to judges sitting in high-volume courts. In those courts, it is particularly important that court staff adopt a procedural fairness approach to litigants by answering questions, even those that are poorly phrased; by providing information about court rules, procedures, and resources even though they cannot provide legal advice; and generally by treating litigants with respect.

Related Entries

- ▶ Community Courts
- ▶ Comparative Courts and Sentencing
- ▶ Drug Courts
- ▶ Judicial Leadership and Performance
- ▶ Mental Health Courts
- ▶ Procedural Justice
- ▶ Procedural Justice, Legitimacy, and Policing
- ▶ Reentry Courts
- ▶ Scientific Evidence in Criminal Adjudication

Recommended Reading and References

Becker TL (1964) Political behavioralism and modern jurisprudence. Rand McNally, Chicago

Berman G, Feinblatt J (2001) Problem-solving courts: a brief primer. Law Policy 23:125–140

Blumberg A (1967) Criminal justice. Quadrangle, Chicago

Boyum K (1979) A perspective on civil delay in trial courts. Justice Syst J 5:170

Burke K, Leben S (2007) Procedural fairness: a key ingredient in public satisfaction (a white paper of the American Judges Association). Court Rev 44:4–25

Court Statistics Project (2010) Examining the work of state courts. National Center for State Courts, Williamsburg, http://www.ncsconline.org/d_research/csp/2008_files/2008_state_court_trial_sheets.html

Dolbeare K (1968) Trial courts in urban politics. Wiley, New York

Eaton L, Kaufman L (2005) In problem-solving court, judges turn therapist. New York Times. http://www.nytimes.com/2005/04/26/nyregion/26courts.html?pagewanted=2&ei=5070&en=17fc6df08c9c39e6&ex=1186632000. 26 Apr 2005

Eisenstein J (1973) Politics and the legal process. Harper & Row, New York

Feeley M (1979) The process is the punishment: handling cases in a lower criminal court. Russell Sage Foundation, New York

Fisher G (2003) Plea bargaining's triumph: a history of plea bargaining in the United States. Stanford University Press, Palo Alto

Flango VE, Clarke T (2011) Which disputes belong in court? Judge J 50:22–30

Flango VE, Wenner LM, Wenner MW (1975) The concept of judicial role: a methodological note. Am J Polit Sci 19:277–289

Freeman-Wilson K, Tuttle R, Weinstein SP (2001) Ethical considerations for judges and attorneys in drug court. National Drug Court Institute, Alexandria

Friedman L (1993) Crime and punishment in American history. Basic Books, New York

Friedman L (2004) The day before trials vanished. J Empir Leg Stud 1:689–704

Fuller LL (1960) Adjudication and the rule of law. P Am Soc Int Law 54:1–8

Gibson JL (1978) Judges' role orientations, attitudes, and decisions: an interactive model. Am Polit Sci Rev 72:911–924

Glick HR (1971) Supreme courts in state politics: an introduction to the judicial role. Basic Books, New York

Hanson R (2002) The changing role of a judge and its implications. Court Rev 38:10–16

Henderson TA, Kerwin C, Saizow H (1984) Structuring justice: the implications of court unification reforms. US Department of Justice, Washington, DC

Higginbottom G (2011) What makes a good judge? Spring conference of the Young Lawyers Division of the American Bar Association, Las Vegas Nevada, 13 May 2011. http://www2.americanbar.org/calendar/young-lawyers-division-spring-conference-2011/Documents/independentjudiciaryprogrammaterials.pdf

Huddleston CW, Marlowe DB (2011) Painting the current picture: a national report card on drug courts and other problem solving court programs in the United States. Bureau of Justice Assistance, Washington, DC

Lande J (2005) Replace "the vanishing trial" with more helpful myths. Altern High Cost Litigation 23:161

Lieberman JK (1981) The litigious society. Basic Books, New York

Moley R (1929) Politics and criminal prosecution. Minton, Balch, New York

Molot JT (2003) An old judicial role for a new litigation era. Yale Law J 27. 103

Murphy JC (2010) Revitalizing the adversary system in family law. Univ Cin Law Rev 78:895

Porter R (2011) Procedural fairness in California: initiatives, challenges, and recommendations. Center for Court Innovation, San Francisco

Redlich W (2007) Albany lawyer blog. http://albany-lawyer.blogspot.com/2007/04/assembly-line-justice.html. Retrieved 22 Nov 2011

Roberts JG (2005) Confirmation hearing on the nomination of John G. Roberts Jr. to be chief justice of the United States: hearing before the s. comm. on the judiciary, 109th Cong. 55 (statement of John G. Roberts, Jr.)

Stevenson CS et al (1996) The juvenile court: analysis and recommendations. Futur Child 6:4–28

Strick A (1977) Injustice for all. G.P. Putnam's Sons, New York
Toohey G (2010) The civil jury trial: going, going, gone. Precedent 7
Ungs T, Baas LR (1972) Judicial role perceptions: a Q-technique study of Ohio judges. Law Soc Rev 6:343–348
Vines K (1969) The judicial role in the American states: an exploration. In: Grossman J, Tanenhaus J (eds) Frontiers of judicial research. Wiley, New York, pp 461–485
Winick BJ, Wexler D (2003) Judging in a therapeutic key: therapeutic jurisprudence and the courts. Carolina Academic Press, Durham

Examining the Effectiveness of Correctional Interventions

Doris Layton MacKenzie
Justice Center for Research, The Pennsylvania State University, University Park, PA, USA

Synonyms

Cognitive transformation; Correctional interventions; Correctional treatment; Evidence-based corrections; What works in corrections

Overview

"Evidence-based corrections" has become a popular phase to explain how an ideal correctional system should function. From this perspective, research, evaluation, and scientific processes should be used to make correctional decisions. To accomplish this objective, scientific research and evaluation needs to be completed to determine which correctional strategies, policies, programs, and interventions effectively result in the desired outcome or outcomes. A body of research conducted over the past 40 years provides guidance to policymakers who are interested in the effectiveness of specific correctional interventions. While there are many potential outcomes desired by the public, policymakers, and correctional experts, perhaps the most consistent interest is in interventions and decisions that are successful in reducing future criminal activities or, in other words, reducing recidivism. Using an evidence-based perspective, this entry reviews the literature on "what works in corrections" and proposes a theoretical explanation, a theory of cognitive transformation, for why some interventions are effective in reducing the recidivism of delinquents and offenders while others are not. Despite the fact that we have some evidence of what works, overall the research on corrections and sentencing interventions fails to match the scientific rigor in methodology and research design that is evident in other fields. The entry ends with a caution about the complexity of transferring interventions from the original controlled setting to environments that may be completely different.

Introduction

Past behavior is the best predictor of future behavior, and, therefore, it is reasonable to attempt to prevent crime by preventing known offenders and delinquents from continuing their criminal and delinquent activities. Once these individuals are convicted or adjudicated, they come under the supervision of correctional systems and this provides an ideal period of time to attempt to change them so that they will cease their criminal activities. Reducing the future criminal activities of delinquents and offenders is of concern because many of these individuals are rearrested or convicted of new crimes a relatively short time after they are released from correctional supervision. Thus, while not the only reason for developing particular correctional strategies, reducing recidivism has remained a major goal of many correctional systems.

Decision makers, the public, and correctional personnel differ greatly in their perspectives on what to do with people who break the law, and their perspectives change dramatically over time. Some of the differences in viewpoints depend on political ideologies on what should be done to those who disobey criminal statutes. Equally as important are more utilitarian considerations related to what is believed to be effective in reducing recidivism and making communities safer. The new "evidence-based" paradigm is central to the utilitarian perspective of reducing recidivism.

The focus on evidence-based practices in criminal justice organizations began with a provocative entry by Sherman (1998). He argued that justice systems, like medicine, should adopt an evidence-based perspective to guide practice. In this way, approaches to crime prevention would be based on evaluation research demonstrating what is effective and this would be used to construct guidelines for effective practice. The approach emphasizes the need for accountability and continual improvement through the use of the most current research results. Thus, continual feedback from research would improve the effectiveness of the justice system in achieving the desired goals. In his opinion, the basic premise of evidence-based practice "is that we are entitled to our own opinions but not to our own facts" (Sherman 1998, p. 4).

For corrections, the idea of an evidence-based perspective is that correctional systems will be more successful in achieving their goals to the degree that science is used to assist in decision making (MacKenzie 2000, 2001). Policymakers, judges, and correctional administrators must make decisions about how to intervene in the lives of convicted offenders and adjudicated delinquents. Most often they do so with some specific goals or objectives in mind. Two major objectives are to increase public safety and to change offenders so that they do not continue committing crimes. From the perspective of evidence-based corrections, interventions intended to accomplish these objectives should be studied with rigorous research and evaluations to determine whether they are successful in achieving the desired outcome or outcomes, in this discussion a reduction in recidivism. Frequently, however, decisions are made with little knowledge of whether the intervention is effective in reaching the desired results. Or as Sherman argued, too frequently, practitioners come up with their own "facts" with little evidence to support their views.

In corrections, too often decisions are made based on "gut-feelings" about what will be effective. For example, despite the fact that a growing body of research demonstrated that correctional boot camps were not successful in achieving the desired reduction in future criminal activities, for many years, people continued to assert that the discipline and challenge in the camps was "just what young offenders needed." Policymakers referred to their experiences in the military and how this made them grow up. The correctional camps quickly proliferated in the USA; however, few jurisdictions conducted the meticulous research needed to determine if they were really accomplishing the reduction in recidivism that was the stated objective of the camps. Fortunately, some jurisdictions did provide support for research to examine whether the camps had an impact on criminal activity. For example, during the time when boot camps were proliferating across the USA, the California Youth Authority (CYA) investigated whether boot camps were something they wanted to initiate on a widespread basis in their facilities (Bottcher 1997). To answer the question of whether the programs would be successful in the desired objective of reducing recidivism, they conducted an experiment of a small pilot program. Researchers randomly assigned youth to the camp or the control group and compared the recidivism rates after youth were released from facilities. Their question about whether to start a large number of boot camps in their facilities was answered when they found releasees from the camps had no lower recidivism than those released from the traditional facilities. Decision makers decided not to start the boot camps in Youth Authority facilities. This is a good example of both how research can be used in decision making and the limitations of "gut-feelings."

In my experience, another reason science is not used more often in correctional decision making is due to differences between law and science in how they conceptualize "research." Many policymakers and court personnel are trained as lawyers and not as social scientists. The law profession looks to past precedence and authorities for guidance to answer questions, not empirical research. Legal research entails a search for past precedence and authorities in order to support one perspective. In law, one takes a position and then tries to find support for this position. Social scientists proceed from a very different perspective. The scientist goes into a research project to ask a question not to support a position and looks to the empirical world to provide the answer. The researcher in this case tries to stay as objective as possible so as not to influence the outcome of the research. Past experience, opinions of authorities, and past research helps the scientist formulate the question, but empirical data is what is used to answer the question.

A further reason given for not doing research to investigate whether an intervention is effective is that funding for research takes away from support for programs and interventions. This argument is frequently used by people who deliver services and treatment. So again, it is a difference in perspective with service providers believing that it is most important to use any funding available to actually deliver programs and services. In contrast, the scientist questions the wisdom of delivering programs before it is clear that the program really does achieve the desired goals or outcomes. For example, there has been little evidence that the programs designed for batterers are effective in changing their behavior. Yet much of the funding goes to providing programs while the quantity and quality of the research is very limited. A better use of funds, from the perspective of evidence-based scientists, is to conduct rigorous studies of the programs to identify what is effective in reducing future battering.

Overall, the expression "evidence-based corrections" has been widely accepted and the phrase has almost become a mantra; however, the number and quality of research in corrections is still relatively limited when compared with other fields. Despite the fact that there are factors that have contributed to limiting the research examining the impact of interventions on recidivism, a sufficient body of research currently exists on some interventions to draw conclusions about what works.

Examining the Effectiveness of Correctional Interventions

This entry focuses on what reduces criminal recidivism using an evidence-based perspective. Policymakers must consider many factors in sentencing and correctional decision making. Many and varied outcomes are important in these decisions including such things as public perceptions of safety, public acceptance of punishments, victim satisfaction, costs, and impacts on the offender like increased educational achievement and reduced substance abuse. These are all germane for policymakers to consider; however, the one factor that almost always is considered a central focus is the impact of decisions on preventing future criminal activity. That is, does the sentencing or correctional option have an impact on recidivism. Therefore, it is important to examine what strategies, programs, or interventions have been demonstrated in the research to be effective in reducing recidivism (MacKenzie 2006). The question becomes: What works in corrections to reduce recidivism?

Meta-analysis is a relatively new statistical technique for conducting systematic reviews of research literature to answer question such as whether an intervention has an impact on recidivism. These analyses are quantitative statistical methods for assessing the impact of a group of studies to determine if they are effective in achieving an identified outcome or outcomes (Lipsey and Wilson 2001 encyclopedia entry on meta-analysis). The studies are the units of analysis. Effect sizes are calculated to determine difference between treated groups who receive the target intervention and control or comparison groups who do not. Meta-analyses with recidivism as the outcome variable examine the impact of the intervention on recidivism and compare the recidivism rate of the treated group to the comparison. This quantitative data is used in the analysis to draw conclusions about effectiveness and to examine whether differences among studies vary as a function of participant characteristics, program components, or research design.

An examination of the literature using meta-analysis to examine the impact of various correctional options clearly demonstrates that certain types of programs are effective in reducing recidivism (MacKenzie 2006; see also the Campbell Collaboration 2012; Lösel 2012: 988 ff.). The following types of programs have been found to be effective in reducing future delinquent and criminal activities: academic and vocational education; cognitive skills programs (Moral Reconation Therapy; Reasoning and Rehabilitation; cognitive restructuring); cognitive behavior and behavioral treatment for sex offenders; Multi-Systemic Therapy (MST) for juveniles; drug courts; and, drug treatment for those in the community and in correctional facilities (MacKenzie 2006). All of these interventions have human service components that address individual deficits and programs that target the individual-level characteristics directly associated with criminal behavior, called "criminogenic needs" by correctional psychologists (Andrews and Bonta 2006). Furthermore, the deficits or problems targeted are dynamic, or changeable, and not static. From this perspective, effective programs address changeable deficits or problems directly related to criminal behavior; in other words, the programs address "dynamic criminogenic needs (Andrews and Bonta 2006)." In addition, the effective interventions address cognitive processes and appear to bring about cognitive transformations in many participants.

Ineffective Policies, Practices, and Philosophies

Correctional interventions that provide increased control or are based on the philosophies of deterrence and incapacitation have not been found to be effective in reducing recidivism. In the past 40 years, policymakers and many correctional system decision makers moved away from a philosophy of rehabilitation toward increased use of interventions focusing on incapacitation, deterrence, and control. Many of these interventions have been submitted to rigorous research to determine whether they are effective in reducing recidivism. The conclusion from the research is that neither increasing the onerousness of punishment to deter nor increasing control to limit individuals' ability to offend is effective in reducing recidivism. Deterrence, incapacitation, increased control, and other punitive approaches to corrections are not effective in reducing future criminal activities (MacKenzie 2006; Lösel 2012: 989 f.). The interventions do not address the criminogenic needs of those who offend. Investigations of popular correctional interventions reflecting this paradigm, such as correctional boot camps, intensive probation and parole supervision, electronic monitoring, and drug testing, have not been demonstrated to be effective in reducing recidivism. None of these interventions address the criminogenic needs of the offenders nor do they create cognitive change.

Lipsey and Cullen's (2007) review of meta-analyses of aversive sanctions and supervision provides further support to the findings of the failure of control, deterrence, and incapacitation interventions to reduce recidivism. They conclude that such sanctions do not inhibit subsequent criminal behavior, and they found evidence that many actually increase future criminal activities. One such program, Scared Straight, has consistently been found to increase recidivism. In addition, existing policies increasing the length of prison sentences do little to deter criminal behavior and such sentences may actually be criminogenic (Nagin et al. 2009). Unintended consequences of incapacitation and deterrence philosophies had other unplanned, negative impacts. The large number of young people being sent to prison in the USA decimated some urban communities and many of those suffering from these policies are young minorities from inner cities (Clear 2007).

Another philosophy about how to decrease recidivism comes from the traditional sociological perspective attributing crime to social conditions. Interventions developed from this perspective have not been found to be effective in reducing recidivism. From this point of view, the focus of interventions should be on changing peoples' social environments by giving them employment and housing, and not on individual differences such as thought processes and rehabilitation. Social bonds or ties to social institutions are assumed to be the keys to successful desistance from crime and, therefore, giving people opportunities to form these bonds was expected to reduce their criminal activity. Life course criminologists who promote this control theory position argue that the attachments or bonds to marriage and employment are critical events leading to desistance and, consequently, interventions should focus on giving offenders and delinquents environment opportunities to develop these times and bonds.

Interventions providing increased social opportunities have not been found to effectively reduce recidivism, a disappointing finding to many social bonds theorists and life course sociologists (see later section on the need for cognitive transformations prior to opportunities). Programs are designed to help offenders obtain employment or housing in order to help them adjust to community life. Reentry programs provide assistance in these areas for those leaving facilities. In-prison work programs and correctional industry are expected to reduce recidivism by giving offenders future employment opportunities. Life-skills program provide assistance in job searches, resume writing, check writing, and other general skills needed for successful adjustment to the community. Yet, these interventions have not been found to be effective. Work programs, multicomponent work programs, prison industries, and life skills have not been found to be successful in reducing recidivism. It is likely the reason these interventions have not been effective is because they are designed to give opportunities to people who may not be cognitively prepared to take advantage of the opportunities.

Another disappointing research finding is that programs designed for batterers or domestic violence perpetrators do not have an impact on recidivism. It is unclear why these programs are not effective (MacKenzie 2006). It may be that the attitudes and corresponding behavior of perpetrators are so deep-rooted that they are very difficult to change. Another possibility is that the relationship between the perpetrator and victim is more complex than realized and treatment does not address this in a manner that is effective. What is clear is that both the quantity and quality of scientific studies is inadequate, considering the large number of programs operating and the magnitude of the problem. One of the problems in this area of research is the hesitancy of those developing programs to test the effectiveness with rigorous research. They frequently take the position that any money used for research will limit the number and size of the programs. Of course, the response is that it is a waste of time and money to operate ineffective programs when research would help give us evidence of what is truly effective and provide evidence of this effectiveness.

Other ineffective programs are those that have little theoretical basis for expecting an impact on recidivism. For example, psychosocial sex offender treatment is based on a psychodynamic theoretical perspective that demands insight, self-reflection, and verbal abilities that are often beyond the skill level of the offenders. The therapy is not responsive to the particular learning styles of the offenders. Other ineffective programs are based on threats of punishment or additional control that do not have a theoretical basis. In general, residential placement for juveniles or adults does not reduce recidivism. Unless the residents receive some type of effective programming while they are in the facility.

Cognitive Transformations

As noted earlier, correctional interventions such as drug treatment, education, and cognitive skills programs are effective in reducing future recidivism. One consistency among the effective programs is a focus on human service. Furthermore, an important aspect of this programming, and the reason they may be successful, is that the interventions address cognitive processes. From this perspective, an individual-level change in cognition must occur before the person is prepared to form ties or bonds to social institutions (MacKenzie 2006). In order to form positive relationships with family, keep a job, support children, or form strong commitments to other social institutions, a person must be cognitively ready to move toward these ties and bonds, and, in many cases, for offenders or delinquents, this requires a change in cognitive reasoning and attitudes. Such a focus on individual change in thinking and information processing is critical to understanding what works in corrections and developing interventions that will successfully impact future criminal activity.

Before a person is prepared to move away from criminal and delinquent activity toward a noncriminal life style, a cognitive transformation is required. It is not the social conditions that have driven these people to become involved in crime. The way individuals think and process information influences whether they violate the law. Deficiencies in social cognition, problem-solving abilities, and sense of self-efficacy are all cognitive deficits or "criminogenic needs" found to be associated with criminal activity. According to the cognitive transformation theory, programs are effective to the degree they are capable of creating a cognitive change in thinking whether this is through changes in criminal thinking or criminogenic attitudes, improved executive functioning and problem solving, or increased maturity and moral development.

Various theories have been proposed to explain the cause of crimes and desistance from crime or the lack thereof. Many of these theories reflect the sociological tradition of criminology and focus on social bonds or ties as the causes for criminal activity. Recently, these theories have been called into question particularly when the literature on "what works" in corrections is examined. Programs based on providing opportunities have not been found to reduce recidivism. While marriage, employment, and school may be correlated with reduced criminal activity, they are not the explanation for why someone chooses to become involved in criminal activity nor do they provide a blueprint for the development of effective correctional programs. Effective correctional programs address cognitions and prepare people for a transition into a noncriminal lifestyle. A cognitive transformation must occur prior to the move away from criminal activities (MacKenzie 2006). Until this transition occurs, people will not be prepared to take advantage of environmental opportunities. Consistent with this perspective, desistance theorists, most notably Giordano and her colleagues (Giordano et al. 2002), Paternoster and Bushway (2009), and Maruna (2001), have proposed theories to explain how cognitive or self-identify changes occur and how these lead to a shift away from criminal activities.

In summary, the research demonstrates that correctional interventions can effectively reduce future criminal behavior. Effective programs address cognitions and prepare people for a transition to a noncriminal lifestyle. A cognitive transformation must occur prior to the move away from criminal activities. Until such a transformation transpires, people will not be prepared to take advantage of environmental opportunities for employment or housing. Consistent with this perspective are the new theories, proposed by several desistance theorists, to explain how changes in cognitions or self-identity can lead to a shift away from criminal activities. Empirical research has demonstrated the type of correctional programs that are effective in changing offenders and, given this information, theorists have developed hypotheses about how and why these programs are effective. The evidence-based correctional paradigm will need to incorporate these viewpoints into policy decisions if one of the goals of programming is to reduce recidivism.

Research Quality and Program Implementation

Approximately 40 years ago, Martinson and his colleagues reviewed the research literature on correctional interventions and concluded that few if any programs could be said to be effective (Martinson 1974; Lipton et al. 1975). Many people interpreted this to mean that nothing works in correctional management and treatment. However, the conclusion from the report was not that nothing could work but instead the researchers attributed the failure to find effective programs to two factors: (1) poor research design and methodology and (2) inadequate implementation of the programs that were studied. These two issues remain important in any assessment of the effectiveness of programs today. The first issue has to do with the quality of the research and the ability of the study design to rule out alternative explanations for the results. The second deals with the integrity of the program or how the program is designed and operated, the amount of time the participant spends in the program, and the training and oversight of the staff.

Since the time of the Martinson study, improvements have been made in research design and statistical techniques. Improved statistical techniques along with systematic reviews and meta-analyses provide evidence of what is effective in preventing and reducing criminal behavior. An increased number of studies have used experimental designs randomly assigning subjects to the intervention or control groups. For example, in a recent review, Farrington and Welsh (2005) found 83 experiments examining offending outcomes had been conducted since an earlier review in 1983. Meta-analytical techniques have increased researchers' ability to combine studies and statistically investigate the effectiveness of a group of studies.

The good news is that there is evidence that some correctional interventions do work and these interventions have specific characteristics consistent with certain principles that can be replicated in future program development. Scientific evidence exists that some correctional interventions are effective in reducing future criminal activity of delinquents and offenders.

Yet problems still exist. "Corrections" is far from other fields, such as medicine, in the use of randomized trials to examine effectiveness. In one study of correctional interventions, MacKenzie (2006) found only 14.8 % of the evaluations were successful randomized trials. In contrast, 23.2 % of the studies were evaluated as so low in the quality of the research methodology that they could not be used to determine effectiveness. The latter were considered to have very dissimilar comparison groups and, therefore, it was impossible to tell if the results were due to differences between the groups being compared or the intervention. That is, it was impossible to draw conclusions about the effectiveness of the program. MacKenzie's (2006) work only included studies that compared a group who received the intervention to a group that did not. The great majority of the published and unpublished manuscripts were excluded from the review because they were qualitative or descriptive studies without any comparison group, so it would be impossible to draw conclusions about the impact on participants. In the future, it will be critical to increase the rigor of the research designs if we are going to successfully employ evidence-based knowledge to correctional decision making.

The second issue Martinson and his colleagues believed limited their ability to draw conclusions about what was effective in reducing recidivism was the quality of the programs studied. We have made progress since that time and now have a body of research literature demonstrating what programs are effective in reducing recidivism. While research has successfully identified therapeutic programs that are effective, much less is known about how to insure these programs are delivered with fidelity and/or therapeutic integrity, or the extent to which interventions conform to the manner of service intended by the developers of the intervention. Insuring the quality of programs remains an issue today. Criminology has been criticized for its lack of focus on what goes on inside correctional programs (Petersilia 2004). Despite the increased knowledge and statistical precision, correctional research continues to fail to investigate what is in the "black box" of the program. If the program is going to be transferred to another jurisdiction with a different population of offenders, detailed information about the characteristics of the intervention is necessary. Martinson and his colleagues found that the programs studied were so poorly implemented that it was impossible to tell whether a well-implemented program would have had an impact. This remains an important element in the research today. Certainly, a program with well-trained staff, designed and operated according to plans in accordance with principles of effective programming, targeting cognitions, providing sufficient dosage appropriate for the risk level of participants, and having enough administrative oversight to afford quality control would be expected to be more effective than a program without such essentials. Consequently, an ideal component of any correctional program evaluation is a relatively thorough, yet unobtrusive assessment of program delivery, especially when a program is adopted from another jurisdiction or population. The portability of a promising program does not always go smoothly (Armstrong 2003). Some researchers have begun to study the implementation of programs by measuring the characteristics of programs, participants, and administration (Latessa and Holsinger 1998; Taxman and Friedmann 2009).

Conclusion

"Corrections" appears to be on the cusp of a new age with a focus on evidence-based programming and smart sentencing. The old paradigm emphasizing punitive punishments has failed and changes are on the way. A large number of delinquents and offenders who come under the supervision of correctional personnel continue to be involved in criminal activity while they are in the community or return to crime when they are released from a facility. How to change this behavior in order to prevent crimes is a critical question for correctional interventions. The research provides guidelines for the components of interventions that will be effective in reducing recidivism. The research clearly points in the direction of the need to develop programs that will address dynamic criminogenic characteristics that will bring about cognitive transformations. Interventions that were popular during the law-and-order state of US correctional philosophy have not effectively reduced recidivism nor have interventions that provide offenders with opportunities for housing or employment if these are not combined with programs addressing cognitive processes.

The emphasis on evidence-based corrections and smart sentencing appears to be heralding a new age. To be successful in these endeavors, we will need more research using rigorous research designs. Also important will be future efforts to understand the specific components of effective programs and how these can be successfully transferred to new populations and jurisdictions.

Related Entries

- ▶ Community Service in Europe
- ▶ Control Theory
- ▶ Correctional Education of Programs
- ▶ Deterrence
- ▶ Deterrent Effect of Imprisonment
- ▶ Domestic Violence
- ▶ Drug Abuse and Alcohol Dependence Among Inmates
- ▶ Drug Courts' Effects on Criminal Offending
- ▶ Early Release from Prison
- ▶ Effective Supervision Principles for Probation and Parole
- ▶ Electronic Monitoring
- ▶ Estimating the Effects of Incapacitation
- ▶ Evidence-Based Policy in Crime and Justice
- ▶ History of Boot Camps
- ▶ Incapacitation
- ▶ Juvenile Justice in the Get Tough Era
- ▶ Law, Diversion and Community Sanctions in Juvenile Justice
- ▶ Offender Change in Treatment
- ▶ Parole and Prisoner Reentry in the United States
- ▶ Punishment as Rehabilitation
- ▶ Rehabilitation
- ▶ Sex Offender Treatment
- ▶ Unintended Effects of Imprisonment

Recommended Reading and References

Andrews DA, Bonta J (2006) The psychology of criminal conduct, 4th edn. Lexis/Nexis, Newark

Armstrong T (2003) The effect of moral reconation therapy on the recidivism of youthful offenders: a randomized experiment. Crim Justice Behav 30:668–687

Bottcher J (1997) A boot camp and intensive parole program: the final impact evaluation (report to the California legislature). California Department of Youth Authority, Sacramento

Campbell Collaboration (2012) http://www.campbellcollaboration.org/resources/links/links_crime_and_justice.php. Search 3 Nov 2012

Clear T (2007) Imprisoning communities: how mass incarceration makes disadvantaged neighborhoods worse. Oxford University Press, New York

Farrington DP, Welsh BC (2005) Randomized experiments in criminology: what have we learned in the last two decades? J Exp Criminol 1:9–38

Giordano PC, Cernkovich SA, Rudolph, JL (2002) Gender, crime, and desistance: toward a theory of cognitive transformation. Am J Sociol 107(4):990–1064

Latessa EJ, Holsinger AM (1998) The importance of evaluating correctional programs: assessing outcome and quality. Correct Manage Q 2(4):22–29

Lipsey MW, FT Cullen (2007) The effectiveness of correctional rehabilitation: a review of systematic reviews. Ann Rev Law Soc Sci 3:297–320. http://lawsocsci.annualreviews.org

Lipsey MW, Wilson DB (2001) Practical meta-analysis. Sage, Thousand Oaks

Lipton DS, Martinson R, Wilks J (1975) The effectiveness of correctional treatment: a survey of treatment evaluation studies. Praeger, New York

Lösel F (2012a) Offender treatment and rehabilitation: what works? In: Maguire M, Morgan R, Reiner R (eds) The Oxford handbook of criminology, 5th edn. Oxford University Press, Oxford, pp 986–1016

Lösel F (2012b) What works in correctional treatmentand rehabilitation for young adults? In: Lösel F, Bottoms A, Farington DP (eds) Young adult offenders: lost in transition? Routledge, London/New York, pp 74–112

Mackenzie DL (2000) Evidence-based corrections: identifying what works. Crime Delinq 46(4):457–471

Mackenzie DL (2001) Corrections and sentencing in the 21st century: evidence-based corrections and sentencing. Prison J 81(3):299–312

MacKenzie D (2006) What works in corrections. Cambridge University Press, New York

Martinson R (1974) What works? Questions and answers about prison reform. Public Interest 35:22–54

Maruna S (2001) Making good: how ex-convicts reform and build their lives. American Psychological Association, Washington, DC

Nagin DS, Cullen FT, Jonson CL (2009) Imprisonment and reoffending. In: Tonry M (ed) Crime and justice: a review of the research. University of Chicago Press, Chicago

Paternoster R, Bushway S (2009) Desistance and the “feared self”: toward an identity theory of criminal desistance. J Crim Law Criminol 99:1103–1156

Petersilia J (2004) what works in prisoner reentry? Reviewing and questioning the evidence. Fed Prob 68(2):4–8

Sherman LW (1998) Evidence-based policing. Police Foundation, Washington, DC

Taxman FS, Friedmann PD (2009) Fidelity and adherence at the transition point: theoretically driven experiments. J Exp Criminol 5:219–226

Exchange

▶ International Policy Transfer

Executions

▶ Lawful Killings

Exonerations

▶ Prosecution and Wrongful Convictions

Experiential Effects

▶ Bayesian Updating and Crime

Experimental

▶ Establishing Causes of Offending in Longitudinal and Experimental Studies

Experiments

▶ Randomized Block Designs

Expert Evidence

▶ Scientific Evidence in Criminal Prosecutions

Expert Evidence and Criminal Trial Procedure

Paul Roberts
School of Law, University of Nottingham,
Nottingham, UK

Overview

This entry examines the main procedural rules and principles applied to scientific and other expert evidence in criminal trials, taking criminal adjudication in England and Wales as its primary illustration. English criminal trials are "adversarial" and, in serious cases, tried before lay juries, but these distinctive institutional design features do not determine the shape and content of criminal procedure rules to the extent commonly assumed. By considering a fuller taxonomy of criminal procedure law we can see that many types of procedural rules – such as those regulating the flow of information to the fact-finder and helping to structure inferential reasoning, as well as rules promoting normative conceptions of procedural integrity and due process – are common, in some form, to all modern legal systems. Turning to more detailed analysis of trial procedure in one particular legal jurisdiction, the admissibility and uses of expert evidence in England and Wales are governed, firstly, by general evidentiary principles applicable to all types of evidence, and secondly, by criminal procedure doctrines specifically tailored to receiving and evaluating expert witness testimony. Both strands of evidentiary doctrine are summarized in this entry, before concluding with some brief observations on possible reform options.

Institutional Contexts of Criminal Trial Procedure

Criminal trials are formal, public, carefully orchestrated, and closely regulated institutional events, quite unlike ordinary everyday forms of social interaction and distinct, even, from just about every other public occasion of major social, cultural, and political significance. If modern states can be said to "govern through crime" (Simon 2007), criminal adjudication might be viewed as a nodal point in the constitution and reaffirmation of political authority, when the administration of criminal justice is at its most visibly "spectacular" (cf. Levenson 2008). Yet criminal trials also perforce involve human participants, some of whom have specialist training as judges and lawyers and cumulative experience as "repeat players" in criminal litigation; others, including most complainants and witnesses and many accused, have neither. Seasoned practitioners and uninitiated laypeople alike, however, share broadly similar biologically endowed cognitive capacities and limitations. One pervasive challenge in the design and operation of criminal trial proceedings, then, is to accommodate these extraordinary public events to the prosaic realities of human cognition, reason, and motivation.

Every system of criminal adjudication confronts this challenge more or less self-consciously and answers it – to the extent that it does – in its own distinctive way. There is no such thing as "free" forensic proof in any literal sense. The question is always how, and to what extent, a particular legal system seeks to manage and regulate the ways in which relevant information ("evidence") is gathered, organized, presented, tested, and finally evaluated by the fact-finder. Various institutional tools, procedures, and mechanisms have been devised and adapted to this end. This entry focuses specifically on the law of criminal trial procedure as one prominent regulatory technique. Legal cultures differ markedly even in their basic concepts and terminology. Common lawyers might be slightly puzzled by the notion of the "law of criminal trial procedure," because they would generally refer to these trial rules as the law of criminal evidence and procedure, or simply as "the law of evidence." But the Law of Evidence, as an identifiable legal specialism, is alien to civilian jurisprudence, which normally treats criminal procedure and civil procedure as separate branches of procedural law. My improvised terminology is intended to cover all of the rules of procedure and evidence which regulate criminal trials, both in Anglophone common law and in continental civilian legal jurisdictions.

It is sometimes thought that trial rules of evidence emerged in response to the adoption in England and, later, in the USA, of trial by jury, but this is misconceived (Schauer 2006). The functional separation between the common law judge, as master of the law, and lay juries as triers of fact makes it easier, in practice, to withhold information deemed unsuitable from the fact-finder than it would be in trial systems where the tribunal (whether a single judge, panel of judges, or mixed panel of judges and jurors) determines both law and fact. In criminal trials in England and Wales or the USA inadmissible evidence can be excluded from the trial without the jury's ever knowing anything about it. But this functional convenience is not *the reason* for evidentiary exclusion. As John Jackson shows in his entry "▶ Scientific Evidence in Criminal Adjudication," evidence may be regarded as problematic quite irrespective of its validity or reliability – for example, because the way in which it was obtained or presented is regarded as being inconsistent with the normative requirements of a fair trial. Likewise, evidence obtained by torture is categorically inadmissible under Article 6 of the ECHR (*Gäfgen* v *Germany* (2011) 52 EHRR 1), irrespective of whether the fact-finder is a professional judge or a lay jury (or an institutional hybrid). So while it is right to say that *certain kinds* of exclusionary rule – chiefly, those predicated on assumptions about lay jurors' cognitive shortcomings – are characteristic of the common law world, rules of evidence and procedure, considered generically as tools for regulating criminal trials, are a universal feature of modern legal systems. There are important differences of degree and kind, but these distinctions can be drawn just as sharply within (cf. Roberts and Hunter 2012), as much as between, the idealized "adversarial" and "inquisitorial" procedural families.

Trials must always be viewed in procedural and temporal context. They are shaped to some extent by what comes after, and more particularly by what went before. Where juries give unreasoned general verdicts simply stating "guilty" or "not guilty," as they generally do in serious criminal cases in common law jurisdictions, opportunities for post-verdict review may be severely curtailed. Criminal appeals against conviction in England and Wales can succeed only if the accused identifies some serious procedural defect in the course of his trial. In the absence of genuinely new evidence or arguments, claims of "actual innocence" will be met with the stony-faced response that the jury has already considered, and rejected, the accused's protestations of innocence. The Court of Appeal is not empowered to substitute its own *ex post* evaluation of the evidence for the jury's verdict. In this procedural environment, where opportunities for reviewing the quality of the jury's output are limited, "input control" regulating the flow of information to the jury during the trial is at a premium. This plausibly accounts, in part, for the breadth, detail, and fastidiousness of evidentiary regulation that has evolved in common law jurisdictions (see Roberts and Zuckerman 2010), compared with the more freewheeling approach to receiving and evaluating evidence in Continental jurisdictions, where pretrial evidence gathering is routinely subject to much closer judicial supervision than it is in party-dominated adversarial systems and appeals against trial verdicts on points of law or fact are more freely available.

It is tempting to equate the law of evidence with exclusionary rules, and doing so tends to reinforce the widespread perception that detailed evidentiary regulation is a peculiarly common law preoccupation, while continental courts and jurists embrace "freedom of proof." But this is fatuous, as Murphy (2010) rightly insists, because legal proof is never a no-holds-barred forensic scrum in which just anything goes. Irrespective of local technical exclusions, if any, there must be legally regulated and epistemically defensible standards, first, for assessing the relevance of potential evidence, and subsequently for determining its probative value, in the context of the case as a whole and in the light of the applicable burden and standard of proof. Forensic science and other expert evidence provides an excellent illustration of these general features of criminal adjudication. The fact-finder always needs to determine, among other things: what does expert evidence go to prove? Has it been adequately tested? Can it safely be relied on? Does it alone, or in combination with other evidence, satisfy the requisite standard of proof for finding the accused guilty of the offense or offenses charged?

Taxonomy of Procedural Regulation of Expert Evidence

Procedural rules, principles, and related normative standards can be classified in different ways. One might distinguish, for example, (i) rules constituting the court and delimiting its jurisdiction; (ii) rules for the conduct of trial proceedings; (iii) evidentiary rules of admissibility; (iv) forensic reasoning rules; and (v) rules for rendering judgment (see Roberts 2011: 395 ff.). Every modern criminal trial system has rules falling under each of these five subheadings, though of course their number and detailed specification vary enormously across individual legal systems. If we further subdivide part (iii) of the suggested taxonomy we might well discover that certain kinds of admissibility rule, for example, those directed towards preempting lay fact-finders from drawing impermissible inferences or falling prey to prejudicial reasoning, are particularly associated with common law jurisdictions and are found less frequently, or not at all, in the civilian legal world.

Regarding part (ii) of the taxonomy, common law trial systems generally adopt adversarial criminal trial procedures. Adversarial trials proceed as a kind of contest or debate between two partisan factions, each urging a different conclusion on the fact-finder, who must somehow reconstruct an accurate portrayal of disputed events from the evidential debris of full-blooded forensic conflict. Adversarialism pervades the broader institutional structures and cultural milieu of criminal adjudication in England and Wales. In a procedural system in which expert witnesses are instructed by adversarial parties, it is likely that experts will experience diffuse, and perhaps occasionally more concentrated, pressures to tailor their evidence to lawyers' expectations (cf. Harris 2000). However, as independent advisors to the administration of justice whose advice is valued precisely because it is assumed to be scientifically valid and nonpartisan, there is only one type of advocacy that expert witnesses can afford to be drawn into, and that is advocacy promoting the integrity of their own, technical evidence. The potential for role-conflict and possible ethical dilemmas involved in being a party-instructed expert in an adversarial criminal trial are explored by Tony Ward in his *Encyclopedia* entry "▶ Expert Witnesses: Role, Ethics, and Accountability."

Another significant trait of common law systems' commitment to adversarial trial procedure is a reflex distrust of court-appointed experts. Although English judges notionally have a general common law power to instruct their own witnesses (*R v Roberts* (1985) 80 Cr App R 89, CA; *R v Cleghorn* [1967] 2 QB 584, CA), including experts (*R v Holden* (1838) 8 Car & P 606, 173 ER 638), this power is hardly ever exercised in criminal trials, and never to the exclusion of party-instructed expert witnesses. Depriving the accused of opportunities to develop the defense case by instructing scientific experts of his own choosing is presumptively unfair. Appearing to sanctify an expert's testimony with the imprimatur of the court might also be regarded as detracting from judicial impartiality and trenching on the jury's constitutional prerogatives. Yet without exclusivity, many of the vaunted advantages of court-appointed experts evaporate. For example, the fact-finder's task is hardly made any easier by being confronted with a three-cornered scientific dispute involving court-appointed as well as party-instructed experts.

Mirjan Damaška (1986) has argued that countries with inquisitorial criminal procedures tend to be characterized by relatively high levels of social trust in state bureaucracies and professional expertise, whereas adversarial cultures are generally more suspicious of state officials and tend to place greater store by lay involvement in public decision-making, even if this sometimes entails contradicting expert opinion. Though admittedly drawn in sweeping and somewhat indistinct brush-strokes, Damaška's suggestive parallels between a society's political culture and its design preferences for legal procedure provide a more-than-plausible explanation of the markedly contrasting fortunes of court-appointed experts on different sides of the English Channel. Court experts flourish where officials dominate legal process and authoritative expert pronouncements are seldom challenged (a theoretical hypothesis confirmed in practice by the deference typically extended to expert evidence by Continental fact-finders: van Kampen and Nijboer 1997). Where, by contrast, the parties develop their own cases and dictate the course of legal proceedings, while professionals, experts, and officials of all kinds incite popular ambivalence and even a measure of suspicion, the court expert will not be held in comparably high esteem and is unlikely to be an acceptable surrogate for party-instructed expert witnesses.

The remainder of this entry concentrates on parts (iii) and (iv) of my suggested taxonomy of criminal trial procedure rules. The next section on "General Principles of Admissibility" outlines general evidentiary doctrines applicable to all evidence, including scientific and other expert evidence. The following section then considers "Rules Specifically Regulating Expert Witness Testimony." Descriptions of the logic of inferential fact-finding apply, more or less, to all modern systems of criminal adjudication because they relate to the cognitive capacities and limitations of human fact-finders. Other features of the analysis are indicative of broad trends within the common law world or apply to any jurisdiction in which criminal trial procedure is broadly speaking adversarial (a classification incorporating an expanding number of legal systems, partly owing to the influence of the ECHR). Ultimately, however, each legal jurisdiction has its own unique set of procedural rules, institutions, and processes, suffused with a distinctive legal culture. Precision can therefore be achieved only at the cost of parochialism. The relevant legal jurisdiction for my purposes is England and Wales, and I will cite English law examples by way of illustration. Scotland, Northern Ireland, and a host of other, smaller legal jurisdictions within the United Kingdom and the British Isles have their own independent criminal justice systems. Even *within* the UK's legal jurisdictions there are major differences in criminal procedure, to say nothing of attempted comparisons with other common law systems of criminal adjudication in North America, Australasia, the Indian subcontinent, South East Asia, or the English-speaking parts of Africa. Generalizations about criminal adjudication in "common law" or "adversarial" systems may be useful signposts indicating likely assumptions and promising avenues for further inquiry, but they are always somewhat tendentious at the level of doctrinal detail or customary institutional practice.

General Principles of Admissibility

General principles of admissibility apply to scientific evidence and expert witness testimony *in addition to* bespoke procedural rules. Scientifically impeccable expert evidence might therefore still be inadmissible at trial, if it infringes one or more of the general legal principles of admissibility discussed in this section pertaining to: (i) relevance, (ii) fair trial, (iii) hearsay, and (iv) character.

Relevance

Only *relevant* evidence is admissible in criminal trials. Evidence is worthless to the legal process unless it might help to resolve the (legally defined) issues in dispute. Relevant evidence might still be excluded on other grounds, but irrelevant information falls at the first hurdle: it is categorically inadmissible. "[T]o be relevant," Lord Steyn observed in *R v A (No 2)* [2002] 1 AC 45, [31], "the evidence need merely have some tendency in logic and common sense to advance the proposition in issue." While the weight or "probative value" of any piece of evidence may vary, depending on how persuasively or conclusively it proves (or disproves) a fact in issue, evidence is relevant if it has *any* probative value whatsoever, and irrelevant if it has none.

The low threshold of relevance mandated by English law is normally easily satisfied by forensic science or other expert evidence. A fingerprint or blood sample placing the accused at the scene of the crime is patently relevant; as are the results of drug tests, toxicological screens, questioned document analysis, pathology reports, and so on, whenever such information tends to confirm (or disprove) that a crime has been committed or to identify particular perpetrators. Indeed, such evidence is not merely relevant, but often of immense and sometimes decisive probative value. There are, however, two types of situations in which scientific evidence might fail to surmount the relevance hurdle to admissibility. The first is where scientific evidence relates to matters no longer in dispute in the litigation, as where the accused in a rape case admits the fact of intercourse, but claims that the complainant consented. In these circumstances, DNA evidence establishing that the accused was the donor of semen recovered from the complainant is no longer strictly relevant to any contested fact (though in practice the DNA report would still be adduced unopposed or submitted to the court as agreed evidence). A striking example is provided by Barry George's successful appeal against his conviction of murdering the TV personality Jill Dando, partly on the basis of a single particle of firearms discharge residue (FDR) which was adduced at his trial to show that George had recently fired a gun. On appeal, it was conceded by forensic scientists that "the possibility that the FDR had come from the gun that killed Miss Dando was equally as remote as all other possibilities [including innocent contamination] and thus, on its own, entirely inconclusive": *R v George* [2007] EWCA Crim 2722, [51]. In fact, evidence which is incapable of distinguishing between the prosecution's case and the defense case, because it is equally likely under either hypothesis, would normally be regarded as irrelevant and should be excluded from the trial.

A second type of scenario conceivably raising relevancy objections concerns recent scientific discoveries or novel or idiosyncratic techniques. Since crack-pot theories, untested hypotheses, and unvalidated scientific techniques fail to satisfy even the threshold expectations of commonsense rationality, they are irrelevant in law, and consequently inadmissible in criminal trials. This is one basis on which, for example, the results of lie detector tests remain inadmissible in English (*R v Chapman* [2006] EWCA Crim 2545, [33]) and Scottish (*HM Advocate* v *Sheridan* [2011] HCJ 1) courts.

Fair Trial

Criminal justice cannot be achieved unless criminal proceedings, including trials, are conducted fairly. Dispensing justice demands due process of law, and justice must be seen to be done in order to sustain the moral legitimacy of criminal verdicts and foster their popular acceptability. Ensuring a fair trial has always been regarded as one of the judge's most fundamental common law duties (*R v Christie* [1914] AC 545, HL), but this traditional precept has been reinforced and greatly expanded in recent decades by judicial innovation against a backdrop of the growing influence of human rights law (Roberts and Hunter 2012; Emmerson et al. 2007). Article 6 of the ECHR guarantees the right to a fair trial and elaborates a number of specific procedural rights for criminal suspects and the accused (Trechsel and Summers 2005). Crucially, the "right to a fair trial" is not limited to what takes place *at trial*. It also regulates criminal investigations (e.g., *Teixeira de Castro* v *Portugal* (1998) 28 EHRR 101), evidence gathering (e.g., *Edwards and Lewis* v *UK* (2005) 40 EHRR 24), and pretrial criminal procedure (e.g., *Rowe and Davis* v *UK* (2000) 30 EHRR 1).

Fair trial standards bear on the way in which evidentiary material is first obtained, its handling and storage, and defense access to scientific testing, results, and data in the pretrial process. Thus, irrespective of the scientific validity of a particular testing procedure, scientific results establishing – say – a suspected substance's chemical composition or the identity of a perpetrator may still be ruled inadmissible at trial if the trace material on which the tests were conducted was obtained or retained illegally, if a proper chain of custody cannot be demonstrated to rule out the risk of accidental contamination (Pattenden 2008), or if test results were not disclosed to the defense when they should have been (ACPO/CPS 2010; *R v Ward* (1993) 96 Cr App R 1, CA). Proactive investigative methods can easily result in evidentiary exclusion if they are not carefully planned and monitored in order to keep criminal investigations within the bounds of propriety and preserve the integrity of their evidential product. Police "entrapment" is not a stand-alone defense in English criminal law (*R v Sang* [1980] AC 402, HL), but collusion by undercover officers or police informants which might be regarded as crossing the line into active incitement of criminality runs a serious risk that evidence of subsequent offending by a targeted suspect will be excluded (*R v Looseley; Attorney General's Reference (No.3 of 2000)* [2001] UKHL 53, [2001] 1 WLR 2060). In particularly egregious circumstances, the trial judge may effectively throw the case out of court by staying proceedings indefinitely as an abuse of process (Choo 2008: Chap. 5).

The unifying normative thread linking police illegality, unfair investigative tactics, and failures of prosecution disclosure is *procedural impropriety*. Justice demands due process and fair trial. Irrespective of internal standards of scientific validity and methodological rigor, scientific or other expert evidence tainted by procedural impropriety is flawed, from a legal perspective, and it is consequently vulnerable to be being sacrificed – ruled inadmissible at trial – in order to preserve the procedural integrity of criminal adjudication.

Hearsay

In the adversarial common law tradition, the oral testimony of witnesses speaking live, on oath and in the presence of the jury in the courtroom is regarded as the paradigm of judicial evidence. The rule against hearsay is the alter ego of this principle of orality (Roberts and Zuckerman 2010: Chaps. 7 and 9). Witnesses must testify only to what they themselves have personally observed, heard, or otherwise perceived through their senses, and be available for cross-examination on the veracity and accuracy of their reported sensory perceptions. Mere hearsay – information told to the witness by somebody else, other people's conversations overheard by the witness, and even the witness's own out-of-court statements – does not qualify as original evidence in English criminal trials.

The hearsay prohibition could, in theory, be a major irritant for forensic science evidence. A great deal of the informational content of expert witness testimony is technically hearsay. Scientific literature and experimental data are compiled over many decades, or even centuries, by countless scientific researchers scattered across the globe. Any individual scientist can, at best, claim first-hand, personal knowledge of only a tiny fraction of the theoretical underpinnings and empirical substance of their courtroom testimony. Likewise, the generation of scientific evidence in individual cases is typically a collaborative enterprise. Trace evidence is routinely collected by scenes of crime officers (SOCOs) or Crime Scene Investigators (CSIs) and submitted to the forensic laboratory where one scientist might undertake preliminary screening, subsequently distributing evidential material to a lab technician who prepares samples, for a third colleague to undertake scientific testing, ultimately reporting to a fourth, more senior forensic scientist who interprets the test results in a final written report and will later testify to its contents in court if called upon to do so. Strictly speaking, however, most of this report is hearsay, because the testifying expert did not personally collect the samples, conduct the scientific testing, or record the data produced in court. With the exception of direct observations or interpretations for which the scientist can personally vouch, courtroom testimony repeats only what the scientist has been told by his or her colleagues, and this is classically hearsay. The same objection theoretically applies to forensic databases of fingerprints, shoeprints, toolmarks, firearms, glass composition, DNA profiles, and so forth, which are likewise compiled over extended periods of time, possibly in multiple locations, by scores or hundreds of individuals who may not even be identifiable at a later date, but at any rate will not be in court to testify to the continuity and reliability of the processes for collecting, testing, and retaining relevant samples on the basis of their own first-hand, personal knowledge.

In practice, courts in England and Wales have on pragmatic grounds largely exempted expert evidence from the operation of the hearsay prohibition. Information derived from general scientific literature, reference works, experimental data, and the common store of scientific knowledge is admissible in court provided that it is incorporated into, or subsumed within, the "opinion" of an appropriately qualified expert witness (*R v Harris* [2006] 1 Cr App R 5; CJA 2003, s.118(1) rules 1 and 8). Evidence of "matches" between crime scene trace evidence and samples or data contained on large-scale criminal justice databases are also admissible (*R v Abadom* [1983] 1 WLR 126, CA; *R v Kempster (No.2)* [2008] 2 Cr App R 19, CA). It might be thought that any other conclusion would be idiotic, in view of the tried-and-tested probative value of statistical and experimental data. However, the hearsay rule is a doctrine of conceptual logic rather than common sense, and it sometimes leads to counter-intuitive results, including, in the past, the exclusion of apparently reliable database evidence which cannot be "laundered" through expert testimony (*Myers* v *DPP* [1965] AC 1001, HL). The position is notably far less flexible in the USA, where suspects enjoy a constitutional right of confrontation *in addition to* being able to object to prosecution evidence on traditional hearsay grounds. In *Melendez-Diaz* v *Massachusetts*, 129 S Ct 2527 (2009), the US Supreme Court ruled, by a 5–4 majority, that drug analysts must be available to testify in person and undergo cross-examination, even though drug testing is mostly routine and uncontroversial and its results could easily be summarized in a short written report. Considerations of cost and convenience had to give way to the Sixth Amendment right to confrontation, as the Justices comprising the majority interpreted it.

Notwithstanding English courts' evidentiary concessions, expert witnesses must still be cautious about their use of second-hand information relating to the current proceedings when writing their reports and testifying in court. For example, case details divulged to a forensic scientist by an investigating police officer, or factual information provided to the doctor who examines an injured complainant, or to a psychiatrist while interviewing the accused (e.g., *R v Miah* [2011] EWCA Crim 945, [60]; *R v Hurst* [1995] 1 Cr App R 82, CA) are all hearsay, and their legal status should be clearly flagged if experts make any reference to such alleged facts in their reports or testimony. Where an expert's conclusions are built upon particular factual assumptions, the expert's evidence is not even admissible, let alone persuasive, until the party calling the expert has laid an appropriate evidential foundation by proving the assumed facts through admissible, nonhearsay, evidence (Roberts 1996). If after the expert has already testified it turns out that the underlying facts cannot be established after all, the jury will be instructed to disregard the expert's evidence in its entirety. Scientific evidence which, although internally valid, has no demonstrable factual connection to the matters disputed in the trial is legally irrelevant, and therefore inadmissible.

Character Evidence

A criminal trial in the common law world is an investigation into particularized allegations of wrongdoing, not a wide-ranging evaluation of the accused's whole life and moral character. To the extent that character may be probative of conduct, however, there is room for admitting general character evidence at trial as material proof of the instant charge(s). This inferential logic applies to witnesses and complainants, whose character might be thought to shed light both on the events currently in dispute and on individuals' testimonial credibility as witnesses in court. It also applies to the accused, who might, for example, have committed similar offenses in the past or revealed distinctive criminal tendencies or patterns or techniques of offending (a signature "m.o.," *modus operandi*) through his previous – charged or uncharged – conduct.

Routine reliance by the prosecution on evidence of the accused's previous convictions or other extraneous misconduct has generally been regarded as unfairly prejudicial in common law jurisdictions. Such evidence is not normally admissible unless its probative value is compelling (*DPP* v *Boardman* [1975] AC 421, HL; *Makin* v *A.G. for New South Wales* [1894] AC 57, PC). The Criminal Justice Act 2003 relaxed English law's traditional aversion to character-based proofs, but evidence of the accused's extraneous misconduct still remains subject to a special evidentiary regime requiring more than bare relevance to secure its admissibility at trial. The general common law prohibition on evidence of the accused's bad character, and in particular on reasoning directly from propensity, still retains its traditional vitality in many parts of the English-speaking legal world.

It is conceivable that scientific evidence might be employed to link the accused to uncharged – or even previously acquitted (*R v Z (Prior Acquittal)* [2000] 2 AC 483, HL) – crimes in order to demonstrate the accused's propensity for committing the type of offense currently before the court, in which case the evidence must satisfy the additional admissibility criteria specified by ss.101–106 of the Criminal Justice Act 2003. If the evidence is excluded, this will not reflect any inherent scientific defect, but rather the court's judgment that revelations of extraneous misconduct would be unfairly prejudicial and disproportionate to the probative value of the evidence.

Fingerprint evidence and DNA matches potentially engage the law of bad character in more subtle ways. If it emerges during the course of a trial that the accused was first identified from a fingerprint left at another crime scene, or by a DNA profile derived from the National DNA Database (NDNAD), a savvy juror could easily infer that the accused must already have been "known to the police." The NDNAD currently contains profiles not only from offenders, but also from suspects who were ultimately acquitted or not even charged in the first place. Jurors might therefore wrongly infer from a matching profile that the accused has a criminal record when, in fact, he has none. In (somewhat tardy) response to the judgment of the European Court of Human Rights in *S and Marper* v *UK* (2009) 48 EHRR 50, English law is currently being amended so that only the DNA profiles and fingerprints of convicted offenders will be retained on the NDNAD (Protection of Freedoms Act 2012). Even if this will in future preclude criminal juries from *mistakenly* inferring that the accused has a criminal record because his profile was on the NDNAD, trial judges should still strive to avoid, or failing that seek to neutralise, any unfair prejudice that could arise from gratuitous disclosures of the accused's extraneous misconduct.

Character can be good as well as bad. Just as the prosecution may attempt to show that the accused has a propensity for engaging in distinctive types of criminality, the defense might try to invoke the accused's character and previous conduct in order to persuade the fact-finder that the accused is "*not* the sort of person" who would do what he is alleged to have done. Although the accused enjoys an antique dispensation to adduce evidence of good reputation (*R v Rowton* (1865) 34 LJMC 57; 10 Cox CC 25, CCR), however, the law has traditionally frowned on self-serving evidence of disposition (*R v Redgrave* (1982) 74 Cr App R 10, CA), which is normally regarded as easily manufactured and probatively trifling. Accordingly, the general rule is that a party is prohibited from leading self-serving character evidence to "bolster the credit" of the party's own witness. This prohibition operates, for example, to prevent a psychiatrist or psychologist who has examined the accused from advising the court that the accused is not a person disposed to violence, or that the accused is a truthful witness (*R v Robinson* (1994) 98 Cr App R 370, CA). Whether or not the accused was violent on the occasion in question, and whether his courtroom testimony should now be believed, are matters on which the jury is supposed to arrive at its own, common sense, judgments. It would be different if the accused or another witness was suffering from a clinical condition making him a pathological liar: the jury can always be informed of such expert medical or psychological diagnoses (*Toohey* v *Metropolitan Police Commissioner* [1965] AC 595, HL). In more recent pronouncements, the Court of Appeal has indicated somewhat "greater willingness to accept medical expert evidence on the issue of the credibility of a witness," in accordance with its professed general inclination to be "more generous towards the admission of expert evidence than was once the case" (*R v Pinfold and MacKenney* [2004] 2 Cr App R 32, [14]–[15]; see Roberts (2004) for discussion).

Rules Specifically Regulating Expert Witness Testimony

We have established that admissible expert evidence must be relevant to the matters in dispute, consistent with procedural due process and the right to a fair trial, and compatible with complex legal doctrines governing hearsay and character evidence. In addition, to secure its admissibility expert evidence must also be (i) helpful to the jury; and (ii) presented by a competent, properly qualified expert. Finally, if it succeeds in being admitted, scientific and other expert evidence is subject to (iii) a clutch of evidentiary doctrines seeking to guide or constrain its uses in fact-finding. These doctrines can be described, collectively, as "forensic reasoning rules" (Roberts and Zuckerman 2010: Chap. 15). They operate as instructions to the trial judge on how to direct the jury before jurors retire to deliberate on the evidence and produce a verdict. English criminal trial judges sum-up on the facts, as well as directing the jury on the applicable law; a practice not generally followed in US criminal trials. This gives the English trial judge a certain measure of practical influence over the jury's evaluation of the evidence, though the judge must always stress that the jury has the final word on questions of fact. In any event, the jury deliberates in secret and will presumably only follow those judicial directions and suggestions which jurors both understand and, broadly speaking, endorse as sensible and appropriate.

Helpfulness: The Primary Criterion

There is no comprehensive statutory framework to regulate the admissibility of expert testimony in England and Wales. The applicable rules are almost entirely derived from judge-made case-law precedents. In the leading case of *R v Turner* [1975] 1 QB 834, 841–2, 843, CA, Lawton LJ explained that the trial judge had correctly excluded psychiatric evidence tendered by the defense, on the grounds that it would not have assisted the jury's deliberations in this particular trial:

> An expert's opinion is admissible to furnish the court with scientific information which is likely to be outside the experience and knowledge of a judge or jury. If on the proven facts a judge or jury can form their own conclusions without help, then the opinion of an expert is unnecessary .... Jurors do not need psychiatrists to tell them how ordinary folk who are not suffering from any mental illness are likely to react to the stresses and strains of life .... [W]e are firmly of the opinion that psychiatry has not yet become a satisfactory substitute for the common sense of juries or magistrates on matters within their experience of life.

This passage has been taken as authority for the existence of a "common knowledge" rule, according to which expert evidence cannot be admitted if it concerns matters within the knowledge and experience of "ordinary folk" (Mackay and Colman 1991; Freckelton 1987: Chap. 3). Alternatively, or in addition, *Turner* is said to imply that expert evidence must be confined to mental abnormality as opposed to normal mental or psychological processes – an "abnormality rule" of admissibility (Sheldon and MacLeod 1991; Thornton 1995). More generally, Lawton LJ's observations are held up as a typical example of legal ignorance and hostility towards science in general, and mental health professionals in particular. As later decisions have tended to confirm, however, Lawton LJ probably meant to insist on nothing more provocative than this banal proposition: expert evidence should be admitted in criminal trials only when, and just insofar as, it is likely to be genuinely helpful to the jury.

Turner's critics rightly insist that expert testimony might still be helpful even if it concerns matters within jurors' broad general knowledge or questions of normal human psychology. Jurors are unlikely to share an expert's detailed, systematic knowledge of specific aspects of normal psychological processes, for example. Where "common sense" beliefs are a witches' brew of genuine knowledge, mangled half-truths, urban myth, and popular invention, an expert might be able to help jurors disentangle fact from fiction so that they can evaluate the disputed evidence more objectively. It is for these sound reasons that English courts do *not* routinely insist on any "common knowledge" or "abnormality" criterion: such tests are rarely more than rough-and-ready rules of thumb to guide the application of the helpfulness standard in particular cases.

If English law really did propound a "common knowledge" rule, for example, expert evidence of identification from a photograph could not have been received in *R v Stockwell* (1993) 97 Cr App R 260, CA. Looking at photographs is hardly an arcane, specialist activity, yet the Court of Appeal endorsed the reception of expert evidence where it was felt that jurors would benefit from further assistance in a matter well within their everyday experience. Likewise, courts in England and Wales have received expert evidence of child psychology (e.g., *DPP* v *A and BC Chewing Gum Ltd* [1968] 1 QB 159, DC), without any suggestion that children, as a class, are "abnormal" or beyond the experience of ordinary people. Helpfulness, however, is a composite standard encompassing costs as well as benefits, and this is something *Turner's* critics tend to overlook. Expert evidence may be misinterpreted or misunderstood, divert jurors' attention into collateral issues, or waste time with cumulative proof of points already well-established by other evidence. Some forms of expert evidence may be especially prone to misinterpretation. It is widely held, for example, that lay jurors have a poor grasp of probability and statistical evidence, though there is disagreement about whether fact-finders are overawed by statistics or more commonly underestimate their true probative value (Koehler 2001; Lempert 1991; Tribe 1971). Irrespective of its theoretical relevance or weight, expert evidence is likely to be excluded if the trial judge is persuaded that it is probably going to be more trouble than it could be worth to hear it.

Helpfulness is the only master-key necessary to unlock the logic of admissibility decisions in England and Wales. Whether expert evidence will actually be admitted on any particular occasion turns on a detailed evaluation of all the available evidence in the case viewed in light of the parties' respective litigation strategies, a factual matrix which will necessarily vary enormously from one trial to the next. Judicial assessments of the admissibility of expert evidence in the current proceedings may consequently be of limited value as legal precedents in subsequent cases.

Expert Qualification

English courts have demonstrated similar flexibility in their assessments of experts' scientific credentials. The legal test of competence for expert witnesses adds little to the threshold relevancy standard, and operates in practice to promote the reception of expert testimony.

There is no requirement of formal training or paper qualifications, professional practice, or even membership of a relevant organization or learned society. The well-known case of *R v Silverlock* [1894] 2 QB 766, CCR, decided that a witness's amateur interest and study were sufficient to qualify him as an expert in handwriting analysis. As Lord Russell of Killowen CJ explained, the question is simply whether or not the witness possesses expert knowledge relevant to an issue before the tribunal, irrespective of how he or she might have acquired it: "The question is, is he *peritus*? Is he skilled? Has he an adequate knowledge? .... There is no decision which requires that the evidence of a man who is skilled in comparing handwriting, and who has formed a reliable opinion from past experience, should be excluded because his experience had not been gained in the way of his business" (ibid. 771).

Case-law illustrations of English courts' catholic conception of forensic "expertise" abound. A drug user with no scientific training or knowledge was permitted to identify a substance in his possession as cannabis (*R v Chatwood* [1980] 1 WLR 874, CA), while a police officer with experience in drug offenses could testify as an expert on the street price of heroin (*R v Hodges* [2003] 2 Cr App R 247, CA). It is perfectly proper for an artist to give "facial mapping" evidence, even though he has "no scientific qualifications, no specific training, no professional body and no database" (*R v Stockwell* (1993) 97 Cr App R 260, 264, CA; confirmed in *R v Atkins* [2010] 1 Cr App R 8, [2009] EWCA Crim 1876). In another memorable case, membership of the Inner Circle of Magic qualified a witness as "a highly expert magician" (*Moore* v *Medley*, *The Times*, 3 February 1955, cited by Smith (1995: 113)). Just about anybody who can provide the jury with specialist information relevant to the proceedings qualifies as an "expert" in English law.

Bespoke Forensic Reasoning Rules for Expert Evidence

Once questions of admissibility have been determined, the focus shifts to how evidence may be used in the trial. English criminal procedure law's "forensic reasoning rules" come in various shapes and sizes. Some forensic reasoning rules are mandatory, notably including the trial judge's obligation to direct the jury on the burden and standard of proof. Other doctrines prescribe flexible directions drawing the jury's attention to the inherent infirmities of particular sorts of evidence, such as eye-witness identifications (*R* v *Turnbull* [1977] QB 224, CA) or the incriminating testimony of an erstwhile accomplice (*R v Beck* [1982] 1 WLR 461, CA) or warn the jury to resist slipping into enticing reasoning fallacies, such as misinterpreting the probative significance of a lie (*R v Lucas* [1981] QB 720, CA) and other forms of evidential "double-counting."

A number of forensic reasoning rules pertaining specifically to expert evidence have been crafted by common law judges over the years. The most fundamental precept is that the jury is the ultimate arbiter of disputed facts, and must be clearly instructed to form its own view, even in relation to uncontradicted scientific, technical, or other expert evidence (e.g., *R v Allen* [2005] EWCA Crim 1344). As Lord President Cooper famously put it in *Davie* v *Edinburgh Magistrates* [1953] SC 34, 40, the role of the expert witness is "to furnish the judge or jury with the necessary scientific criteria for testing the accuracy of their conclusions, so as to enable the judge or jury to form their own independent judgment by the application of these criteria to the facts proved in evidence." Legal commentators differ on the extent to which "juror education" is a practically feasible aspiration for legal adjudication (Allen 1994; Imwinkelreid 1997; Edmond 1998). It is implausible to imagine that a short, intensive course of juror education within the context of criminal trial proceedings could truly substitute for an expert's long years of education, training, and practical experience. But jurors can at least be expected to exercise their critical faculties in deciding to rely on expert evidence bearing on contested factual issues. Measured, reflective, rational deference to expertise should not be equated with slavish acceptance or abdication of deliberative responsibility. And the jury must emphatically retain its sovereignty in determining "factual" questions with a normative dimension, such as whether the accused was "grossly negligent" (e.g., *R v Misra and Srivastava* [2005] 1 Cr App R 21, [2004] EWCA Crim 2375) or acted with diminished "mental responsibility" (e.g., *R v Dietschmann* [2003] 1 AC 1209, HL).

Rational deference to expertise might be an attractive option when the jury is presented with uncontradicted expert evidence, but how should the jury proceed in cases where experts called by the adversarial parties disagree? The answer is that the jury should do its common sense best to interpret the experts' disagreement in the light of all the evidence in the case and its relevance to the charges and the facts in issue. Ultimately, in cases of enduring uncertainty or doubt, the criminal burden and standard of proof weigh in favor of an acquittal (e.g., *R v Cannings* [2004] 1 WLR 2607, CA). It is sometimes argued that cases involving disagreements between experts should attract a special judicial direction, but the Court of Appeal has resisted the notion that there is some ubiquitous form of words that must be used to cover this situation (*R v Hookway and Noakes* [2011] EWCA Crim 1989). The judge is merely required to do what trial judges must do in every case: to summarize fairly and fully any factual disagreement between the (expert) witnesses, indicate the potential salience of that disagreement for the questions that need to be determined, and leave the rest to the jury's good sense.

It should already be apparent from the preceding discussion that a firm presupposition of lay jurors' expertise in common sense reasoning is cemented into the institutional architecture of English criminal adjudication. Attempts to encourage the jury to depart from its ordinary reasoning processes are consequently frowned upon. Thus, in a case in which a defense expert encouraged jurors to apply Bayes' Theorem to contested evidence, the Court of Appeal thundered:

> [T]he attempt to determine guilt or innocence on the basis of a mathematical formula, applied to each separate piece of evidence, is simply inappropriate to the jury's task. Jurors evaluate evidence and reach a conclusion not by means of a formula, mathematical or otherwise, but by the joint application of their individual common sense and knowledge of the world to the evidence before them. (*R v Adams* [1996] 2 Cr App R 467, 481)

Moreover, the "ultimate issue" of determining the accused's guilt or innocence is always reserved to the fact-finder. The courts are assiduous in ensuring that "trial by expert," or even "trial by science," is never allowed to usurp the constitutional functions of trial by judge and jury. As Lawton LJ intoned in *Turner*, "We do not find that prospect attractive and the law does not at present provide for it" ([1975] 1 QB 834, 842).

The way in which juries are expected to reason about forensic science and other expert evidence has further implications for the way in which expert witnesses are expected to present their evidence in the courtroom. Expert testimony must be reasonably intelligible to lay people and communicated in a manner that successfully conveys its true probative significance. Thus, English courts have striven to ensure that DNA profiling evidence does not inadvertently encourage jurors (or trial judges directing them) to commit the well-known "prosecutor's fallacy," as reproduced by the Court of Appeal in *R v Doheny and Adams* [1997] 1 Cr App R 369, 372–373:

1. Only one person in a million will have a DNA profile which matches that of the crime stain.
2. The defendant has a DNA profile which matches the crime stain.
3. Ergo there is a million to one probability that the defendant left the crime stain and is guilty of the crime.

Deduction (3) is clearly fallacious, because "one in a million" is the frequency of the DNA profile in the suspect population, not the probability that the accused is the source of the profiled crime stain. If there were, say, four million people in the relevant suspect population, one would expect to find four individuals, plus the actual donor of the crime stain (who may or may not be the accused), with that matching DNA profile. Taking the profiling evidence in isolation, the probability that the accused is the source of the crime stain is actually 1-in-5 or 0.2. Glaring and pronounced though the fallacy seems when illustrated through simple hypotheticals, anecdotal evidence suggests that it may occur in practice with alarming regularity. English courts have attempted to insulate the jury from its distorting impact by insisting that a forensic scientist "should not be asked his opinion on the likelihood that it was the defendant who left the crime stain, nor when giving evidence should he use terminology which may lead the jury to believe that he is expressing such an opinion" (ibid. 374: generally, see Puch-Solis et al. 2012; Redmayne 2001).

More recently, in *R v T* [2011] 1 Cr App R 9, [2010] EWCA Crim 2439, the Court of Appeal went so far as to say that experts should not even employ likelihood ratios in their own case-work in the absence of a robust statistical basis to support further quantitative analysis. This was something of an overreaction (Redmayne et al. 2011), but the Court's chief concern was entirely legitimate. It feared that the expert, who testified regarding footwear marks, was building factual assumptions into his evidence which were being concealed from the jury at trial. Experts' own inferential reasoning needs to be transparent so that the jury can perform its designated role as fact-finder, rather than simply being presented with an expert witness's inferential conclusions as a forensic *fait accompli*. These procedural rules are perhaps best categorized as supplementary admissibility standards, regulating not so much *what* information can be adduced at trial but *how* it may be presented. But their rationale is evidently rooted in traditional conceptions of the methodology and status of jury fact-finding in English criminal adjudication, and particularly when expressly incorporated into the trial judge's summing-up to the jury, these procedural requirements are also aptly characterized as forensic reasoning rules.

Future Prospects: Reliability Through Reform?

Functional competence and "*Turner* helpfulness" constitute a flexible and relatively undemanding framework of bespoke admissibility standards, which in practice (and with the possible exception of certain kinds of behavioral science evidence) tends to militate in favor of receiving relevant expert witness testimony even when its probative value is sometimes questionable. Indeed, it might be said that English law has been rather *too* liberal in receiving scientific evidence in criminal trials, with too little scrutiny of its scientific credentials (Roberts 2008; Ormerod 2002). Some purported science is only "junk science" (Bernstein 1996; Huber 1991); not all forensic experts are properly qualified, and a few may be irresponsible charlatans. Even genuine experts might unwittingly propound false theories, make operational errors in their testing or calculations, or overstep the boundaries of their legitimate expertise. The question then arises as to whether criminal trial procedure is sufficiently well-insulated from fake science and bogus experts, and capable of distinguishing between valid and invalid scientific claims. Given that the whole system of criminal adjudication is predicated on lay fact-finders utilizing ordinary common sense techniques – principally listening to the testimony and observing the demeanor of witnesses who appear before them in the courtroom – ensuring scientific validity obviously poses a major design challenge for English criminal proceedings.

Promoting the validity and reliability of scientific and other expert evidence might be achieved in various ways. One might attempt to regulate forensic science providers directly, to subject crime labs to schemes of accreditation and performance management, or to compile registers of appropriately qualified experts certified court-safe. One might rely on internal professional regulation and codes of professional ethics to maintain standards. There is a sense in which criminal litigation itself exerts a kind of market pressure on expert witnesses, since experts judged poor by the adversarial party hiring them are unlikely to be instructed again in future cases. (However, this is a very imperfect market by any standards, and may be warped by perverse incentives: the most effective courtroom performers, who delight their instructing lawyers time after time, might be poor scientists or willing to indulge in unethical behavior in order to help "their" side win the case.) It is likely that an effective legal framework for promoting and monitoring the quality of forensic science evidence and expert witness testimony would incorporate several, and perhaps all, of these complementary strategies. A final question to consider is whether reform of existing trial procedure law should form part of a comprehensive regulatory package in England and Wales (or anywhere else).

Drawing on the experiences of other common law jurisdictions, and in particular the "*Daubert* trilogy" in the USA (Berger 2011), the Law Commission (2011) has recently recommended the introduction in England and Wales of an enhanced judicial gatekeeping function in relation to expert evidence. The basic idea, modeled on *Daubert* v *Merrell Dow* 113 S Ct 2786 (1993), is that, whenever the reliability of expert evidence proffered by the prosecutor or defense lawyers is challenged, the trial judge ruling on admissibility should be obliged to undertake a preliminary assessment of its validity and reliability, by reference to an open-ended list of statutory criteria. The proposed statutory criteria include "generic factors" like data quality and completeness, appropriateness of methodology, soundness of inferential conclusions, and extent of peer review, and make provision for supplementary “factors for specific fields” to be added by the Lord Chancellor through delegated legislation.

It remains to be seen whether these recommendations will find their way onto the statute-book. US courts’ mixed experiences with *Daubert* over the last two decades (Risinger 2007) clearly demonstrate that investing trial judges with a gatekeeping duty to scrutinize the validity and reliability of scientific and other expert evidence does not, at a stroke, solve all the practical problems associated with expert witness testimony in the courtroom. It could, indeed, create new difficulties. In its preliminary consultation paper the Law Commission (2009: 1.13) wisely cautioned that “reforming the law governing the admissibility of expert evidence would not provide a panacea.” Re-engineering formal admissibility rules does not necessarily alter judicial culture or practice, or to the extent that it does, adaptive behavior may defeat reformers’ expectations. There is potentially greater scope for effective intervention in pretrial proceedings, for example, through comprehensive pretrial disclosure and judicial encouragement to agree scientific evidence and narrow down the issues for trial. These are now prime objectives of proactive judicial trial management in England and Wales, within the framework of the Criminal Procedure Rules first promulgated in 2005 (see, e.g., *R v Reed and Reed* [2010] 1 Cr App R 23; [2009] EWCA Crim 2698). None of this amounts to an argument against reforming rules of admissibility, if reform is desirable and well-executed. It simply reaffirms the truism that the law of criminal trial procedure is only one among many influences on the conduct and outcomes of criminal adjudication, not least in trials involving scientific evidence or expert witness testimony.

Recommended Reading and References

ACPO/CPS (2010) Guidance booklet for experts: disclosure – experts’ evidence, case management and unused material. CPS Strategy and Policy Directorate, London. www.cps.gov.uk/publications/prosecution/index.html. Accessed 14 Feb 2013

Allen RJ (1994) Expertise and the *Daubert* decision. J Crim Law Criminol 84:1157–1175

Berger MA (2011) The admissibility of expert testimony. In: Federal Judicial Center and National Research Council, Reference manual on scientific evidence, 3rd edn. National Academies Press, Washington, DC, pp 11–36

Bernstein DE (1996) Junk science in the United States and the Commonwealth. Yale J Int Law 21:123–182

Choo AL-T (2008) Abuse of process and judicial stays of criminal proceedings, 2nd edn. Oxford University Press, Oxford

Damaška MR (1986) The faces of justice and state authority: a comparative approach to the legal process. Yale University Press, New Haven

Edmond G (1998) The next step or moonwalking? Expert evidence, the public understanding of science and the case against Imwinkelried’s didactic trial procedures. Int J Evidence Proof 2:13–31

Emmerson B, Ashworth A, Macdonald A (2007) Human rights and criminal justice, 2nd edn. Sweet & Maxwell, London

Freckelton IR (1987) The trial of the expert. Oxford University Press, Melbourne

Harris GC (2000) Testimony for sale: the law and ethics of snitches and experts. Pepperdine Law Rev 28:1–74

Huber PW (1991) Galileo’s revenge: junk science in the courtroom. Basic Books, New York

Imwinkelreid EJ (1997) The next step in conceptualizing the presentation of expert evidence as education: the case for didactic trial procedures. Int J Evidence Proof 1:128–148

Koehler JL (2001) The psychology of numbers in the courtroom: how to make DNA-match statistics seem impressive or insufficient. South Calif Law Rev 74:1275–1305

Law Commission (2009) The admissibility of expert evidence in criminal proceedings in England and Wales. LCCP No 190, TSO, London

Law Commission (2011) Expert evidence in criminal proceedings in England and Wales. Law Com No 325, TSO, London

Lempert R (1991) Some caveats concerning DNA as criminal identification evidence: with thanks to the Reverend Bayes. Cardozo Law Rev 13:303–341

Levenson LL (2008) Courtroom demeanor: the theater of the courtroom. Minnesota Law Rev 92:573–633

Mackay RD, Colman AM (1991) Excluding expert evidence: a tale of ordinary folk and common experience. Criminal Law Rev 800–810

Murphy P (2010) No free lunch, no free proof: the indiscriminate admission of evidence is a serious flaw in international criminal trials. J Int Criminal Justice 8:539–573

Ormerod D (2002) Sounding out expert voice identification. Criminal Law Rev 771–790

Pattenden R (2008) Authenticating “things” in English law: principles for adducing tangible evidence in common law jury trials. Int J Evidence Proof 12:273–302

Puch-Solis R, Roberts P, Pope S, Aitken C (2012) Assessing the probative value of DNA evidence. Royal Statistical Society, London. www.rss.org.uk/site/cms/contentviewarticle.asp?article=1132

Redmayne M (2001) Expert evidence and criminal justice. Oxford University Press, Oxford

Redmayne M, Roberts P, Aitken C, Jackson G (2011) Forensic science evidence in question. Criminal Law Rev 347–356

Risinger DM (2007) Goodbye to all that, or a fool's errand, by one of the fools: how i stopped worrying about court responses to handwriting identification (and "forensic science" in general) and learned to love misinterpretations of *Kumho Tire* v *Carmichael*. Tulsa Law Rev 43:447–475

Roberts A (2008) Drawing on expertise: legal decision-making and the reception of expert evidence. Criminal Law Rev 443–462

Roberts P (1996) Will you stand up in court? On the admissibility of psychiatric and psychological evidence. J Forensic Psychiat 7:61–76

Roberts P (2004) Towards the principled reception of expert evidence of witness credibility in criminal trials. Int J Evidence Proof 8:215–232

Roberts P (2011) Groundwork for a jurisprudence of criminal procedure. In: Duff RA, Green S (eds) Philosophical foundations of criminal law. Oxford University Press, Oxford, pp 379–408

Roberts P, Hunter J (eds) (2012) Criminal evidence and human rights: reimagining common law procedural traditions. Hart, Oxford

Roberts P, Zuckerman A (2010) Criminal evidence, 2nd edn. Oxford University Press, Oxford

Schauer F (2006) On the supposed jury-dependence of evidence law. Univ Pennsylvania Law Rev 155:165–202

Sheldon DH, MacLeod MD (1991) From normative to positive data: expert psychological evidence re-examined. Criminal Law Rev 811–820

Simon J (2007) Governing through crime: how the war on crime transformed American democracy and created a culture of fear. Oxford University Press, New York

Smith JC (1995) Criminal evidence. Sweet & Maxwell, London

Thornton P (1995) The admissibility of expert psychiatric and psychological evidence – judicial training. Med Sci Law 35:143–149

Trechsel S, Summers SJ (2005) Human rights in criminal proceedings. Oxford University Press, Oxford

Tribe LH (1971) Trial by mathematics: precision and ritual in the legal process. Harv Law Rev 84:1329–1393

van Kampen P, Nijboer H (1997) *Daubert* in the Lowlands. UC Davis Law Rev 30:951–995

Expert Witnesses

▶ Scientific Evidence in Criminal Prosecutions

Expert Witnesses: Role, Ethics, and Accountability

Tony Ward
School of Psychology, Victoria University of Wellington, Wellington, New Zealand

Overview

Expert witnesses make observations or draw inferences that lay or judicial fact finders would not be able to make for themselves. It is, nevertheless, the jury or judge and not the expert that is the ultimate fact finder. It is therefore central to the role of expert witnesses, and to their ethical obligations, that they give a clear account of their conclusions and reasoning that enables the fact finder to reach an informed and independent judgment as to how far expert evidence can be relied upon. In adversarial systems of trial, great reliance is placed upon cross-examination and the ability to call opposing experts to ensure that experts are held accountable for their methods and that fact finders are not misled. In practice, however, these methods have serious shortcomings and need to be combined with other forms of accountability such as oversight by professional bodies.

The Role of the Expert Witness

The role of expert witnesses differs from that of ordinary witnesses of fact in the scope they have to draw inferences from their own observations and those of other witnesses. In this respect, the expert is an exception to the general preference, at least in the common-world, for witnesses to stick as closely as practicable to an unvarnished account of what they saw and heard, leaving the fact finder – often the lay jury – to draw its own inferences from those facts (see ▶ Scientific Evidence in Criminal Prosecutions).

"Expert evidence" is a much wider category than "scientific evidence," although arguments about exactly which experts qualify as "scientists" are not particularly helpful. Gross and Mnookin (2003) offer a taxonomy which usefully captures the range of roles expert witnesses may play. Much expert evidence, they point out, is primarily descriptive. Experts may observe phenomena that lay people would miss, or know the meaning of terms (e.g., in a drug-using subculture) that a jury might not, or be able summarize information by means of a calculation ("the deceased's blood alcohol level was three times the drink-driving limit"). A second role for experts is that of instructing fact finders as to general propositions relevant to the case, such as the nature of mitochondrial DNA, the characteristic symptoms of autism, or the norms of acceptable medical practice. The role that gives rise to the greatest difficulty, however, is that of assessment of evidence. Experts draw upon their general knowledge and experience to offer opinions on diagnosis, causation, prognosis, identity – "this fingerprint matches that of the defendant" – or value (e.g., of drugs or stolen goods). In this role, experts go beyond merely providing information to the court, to interpret the factual or normative significance of the phenomena that they or others have observed. In doing so, they are undertaking a task that is normally reserved for the judge or jury.

Expert witnesses have to accept that their cognitive authority is, in principle, subordinate to that of the court: "forensic experts. . .agree to surrender ultimate control over the way scientific evidence becomes built into constructions of factual reality" (Smith 1989, p. 71). If challenged to justify lay fact finders' control over the "construction of reality," lawyers might point to the fact that legal disputes are rarely resolved by scientific evidence *alone* (Dwyer 2008), and that the "reality" that emerges is one that the community as a whole is asked to accept as a legitimate basis for social action – in particular, punishment. The question for the court is not what the experts think is true but what the community at large can accept as true, given such trust as it can justifiably place in the experts. The court might be able to assess the substantive merits of the expert's argument, but even if it cannot, judges or juries may have good reason to believe that the expert's conclusions are soundly based, albeit for reasons the court does not fully understand (Hardwig 1994). Even if they understand a good deal of an expert's reasoning, lay fact finders generally have to take some of what an expert says on trust. So an important part of the expert's role is to communicate information in a way that the fact finder will both understand and trust.

The dynamics of trust differ according to whether the system of trial approximates to an inquisitorial or an adversarial model (see ► Scientific Evidence in Criminal Adjudication). In inquisitorial, civil-law systems, where experts are normally appointed by the court, judges (or coroners in some common-law systems) are likely to trust "their" experts as a matter of routine, although that trust may be challenged in exceptional cases (Bal 2005). The adversarial model has been described, with some exaggeration, as one of "institutionalized pure mistrust" (Wynne 1989, p. 33). The expert is called, almost invariably, to support one side in a contest, so their evidence cannot be accepted as something like objective truth until the other side has had an opportunity to challenge it. Conversely, the absence or failure of a challenge is taken to indicate that the expert evidence is reliable. In a circular process, what is unchallenged appears unchallengeable, and therefore remains unchallenged, as was until recently the case with fingerprint evidence (Cole 2001; Fingerprint Inquiry Scotland 2011). Where the evidence is more contentious, the adversarial system puts a premium on the ability to persuade juries (or judges), and can be accused of distorting the expert's role by making presentational skills more important than scientific acumen (House of Commons 2005).

The potential for conflict between an expert's roles as servant of the court and as a member of the prosecution or defense team is exacerbated by the fact that experts are involved in the investigation as well as the prosecution of crime. The emphasis in forensic work appears, indeed, to be shifting away from the courtroom and towards what Barclay (2009, p. 339) calls "a completely integrated approach" to major criminal investigations, "with detectives, crime analysts, forensic scientists, pathologists and behavioural experts all part of the same team." This "paradigm shift," as Barclay describes it, requires that scientists be given much more information than would previously have been thought desirable about lines of enquiry being pursued by the police, and offer more "personal interpretations, soft facts, and opinions expressed in shades of grey" (Barclay 2009) than they would when giving evidence in court.

There is a tension between the usefulness for investigative purposes of interpreting evidence in context and the need for objectivity, which may be undermined by knowledge of contextual information. Such information may simply lead experts to see what they expect to see (Dror and Rosenthal 2008). Risinger et al. (2002, p. 28) argue that for a forensic scientist to rely on any information outside her own domain is to "abuse her warrant," since it is for the jury, not the scientist, to assess the nonscientific evidence of the "case circumstances." On the other hand, it appears legitimate for scientists to take contextual circumstances into account in deciding which hypotheses to test and how to test them.

Tensions between investigative and courtroom roles can also arise in civil law, inquisitorial systems. The leading case on expert evidence in the European Court of Human Rights, *Bönisch v. Austria* (1987) 9 EHRR 191, concerned a prosecution for food hygiene offenses in which the President of the institute who initiated the prosecution also served as court expert at the trial. The defendant was able to call another expert, but as a mere "witness" this person had a lower status than the ostensibly impartial "expert" appointed by the Court. This was held to be a breach of the principle of "equality of arms." Jakobs and Sprangers (2000, p. 383) report that in "almost all European countries" there is some provision for the state to bear the costs of appointing defense experts, "but this situation creates some kind of control by the state" since prosecutors, judges, or legal aid authorities have discretion whether to provide for a publicly funded expert.

Adversarial proceedings – and the "equality of arms" principle makes all European criminal proceedings adversarial to some extent – depend upon experts being available to testify for the defense as well as for the prosecution. Defense lawyers also need experts to advise them on possible weaknesses in prosecution evidence. In practice, access to such resources on the part of the defense may not be forthcoming – particularly in the USA, where judges routinely deny expert assistance to indigent defendants (Giannelli 2003; Garrett and Neufeld 2009). Both the "battle of the experts" that sometimes develops, and the situation where a prosecution expert faces little effective challenge, raise important ethical issues about the limits of appropriate persuasion.

The Ethics of Expert Witnessing

It is uncontroversial that the first duty of a witness is to tell the truth. However, because the roles of the expert include "instruction" and "assessment" as well as "description" (Gross and Mnookin 2003), telling "the whole truth" is not such a simple matter for an expert as it might appear to be for an ordinary witness of fact. Experts implicitly warrant not only that the factual content of their evidence is true, but that the opinions they express are based on good reasons. Given the reliance that others will place on their testimony, it "is unethical for expert witnesses to hold or express unjustified beliefs" (Sanders 2007, p. 1542).

In England and Wales, the Criminal Procedure Rules (CrimPR) and accompanying practice directions have recently codified the ethical duties of experts to ensure that their opinions are justified, and to provide the court with the information it needs in order to assess whether they are justified. According to the CrimPR, the duty of the expert is to "help the court to achieve the overriding objective" of dealing with cases justly "by giving objective, unbiased opinion on matters within his expertise" (r. 33.2).

One form of "objectivity" in expert evidence is that the expert states what she takes to be what any competent expert would observe or conclude given the same information (Smith 1989) – or if not, she explains why divergent views are possible. The duty to be objective and unbiased also entails that the evidence that experts give should not be colored by any interest they may have in the proceedings. Yet it is arguably unrealistic to expect experts involved in criminal cases to be entirely disinterested, for at least four reasons. First, as we have seen, expectations may create unconscious biases: Experts are only human, and human beings tend to see what they expect to see. Secondly, in an adversarial system there are inevitably pressures to identify with the party and lawyers by whom one is called. Thirdly, many experts are subject to the pressures of the market, either as individual professionals or as employees of commercial service providers. In such an environment, the probative value that experts can claim for their evidence, its power to discriminate between conflicting hypotheses, may be directly correlated with its market value (Lawless 2011). Fourthly, serious crimes arouse strong emotions to which experts are not immune, particularly if their work involves examining or treating the victim. It is questionable whether complete objectivity is possible in this situation (Candilis et al. 2007); at all events, it is hard to judge objectively whether one is objective or not! What is fairly uncontroversial is that experts should "strive for objectivity" (AAPL 2005, Section IV). This involves not only telling the court what one honestly believes, but explaining the reasons for *and against* those beliefs: what Candilis et al. (2007, p. 21) call the "duty of transparency."

Transparency involves stating the basis of one's opinions in a way that enables the court to make its own assessment of the reliance it can place on them. The ethical duty of transparency has at least four aspects.

Transparency as to the Basis of the Expert's Evidence

When an expert is testifying about matters beyond the knowledge of the court, the court may not be in a position to evaluate independently the reasoning on which the experts' conclusions are based. The court may then have little choice but to defer to the expert, assuming that – given their training and experience in the application of an established body of knowledge – experts probably have good reasons for what they assert (Hardwig 1994). Where the expert is relying not on specialized observations but on matters that the court is equally, or more, competent to assess, this kind of deference to expert authority has no place. The court needs to know when this is the case, both so that unwarranted deference can be avoided and to guard against "double-counting" the evidence.

This points to the particular importance of transparency on the part of experts who engage in the kind of contextual interpretation that Barclay's (2009) "new paradigm" of forensic science, and many kinds of medical or psychological diagnosis, require. For example, it is important for pediatricians reporting alleged signs of sexual abuse to make clear when they are relying on hearsay information or behavioral signs in interpreting ambiguous physical signs (*Lillie and Reid v Newcastle City Council* [2002] EWHC 600 (QB), [395]–[396]). Hence, the requirements in the CrimPR state that an expert's report must include the following:

(a) Give details of the expert's qualifications, relevant experience and accreditation
(b) Give details of any literature or other information which the expert has relied on in making the report
(c) Contain a statement setting out the substance of all facts given to the expert which are material to the opinions expressed in the report, or upon which those opinions are based
(d) Make clear which of the facts stated in the report are within the expert's own knowledge (r. 33.3)

Transparency as to Disagreement Within the Expert Community

One of the reasons that may induce a court to accept an inference drawn by an expert is that it derives from premises that the whole community of qualified experts accepts. The court, however, will usually be dependent on the expert witnesses themselves to tell it whether a consensus exists. Experts should "not give the impression that [they] speak for the community of experts when [they] do not. When the community of expert opinion is divided, there is an obligation to say that it is" (Hardwig 1994, p. 92). Or as the CrimPR put it, "where there is a range of opinion on the matters dealt with in the report," the report must include the following:

1. Summarize the range of opinion
2. Give reasons for his own opinion (r. 33.3(1)(f))

Transparency as to Areas of Uncertainty and Conflicting Evidence

The Criminal Procedure Rules require that "if the expert is not able to give his opinion without qualification, [his report must] state the qualification" (r. 33.3(1)(g)). This is particularly important where the limitations of research may afford grounds for reasonable doubt – even if the experts personally have no doubt about their conclusions. For example, one should not declare confidently that certain features of a young girl's vagina and hymen are abnormal and indicative of penetration, while failing to mention that very little is known about the incidence of similar characteristics in non-abused girls (*R v T* [2008] EWCA Crim 3229).

Transparency and Clarity in Explaining Degrees of Probability

Terms like "match" or "similar in all microscopic characteristics" may be highly misleading to juries (McQuiston-Surrett and Saks 2009). The phrase "consistent with" can make neutral evidence – evidence equally consistent with the prosecution or defense hypotheses – sound as if it is highly probative (e.g., *Barry George v R* [2007] EWCA Crim 2722). On the other hand, evidence couched in statistical terms but based on very rough estimates may confuse the jury and create an unwarranted "verisimilitude of mathematical probability" (*R v T (footwear mark evidence)* [2010] EWCA Crim 2439, [86]). The most candid way of expressing such testimony would appear to be to give a "subjective probability estimate" which makes clear that it is "a product of the examiner's subjective judgment – a guesstimate" (McQuiston-Surrett and Saks 2009, p. 438). Such estimates can be expressed in the form of a "likelihood ratio" – the relative likelihood of a particular finding given either the prosecution hypothesis (e.g., that the defendant's shoe made a certain mark) or the defense hypothesis (i.e., it was a different shoe). The Court of Appeal in *R v T* controversially ruled that likelihood ratios should not be used at all in the absence of satisfactory data. It is difficult to see how a bar on *thinking* in terms of likelihood ratios can be enforced, and to insist that any statistical "guesstimates" must not appear in the expert's report appears to defeat the Court's emphasis on transparency about the methods an expert has used. The use of Bayesian methods (in which likelihood ratios are a central element) may also have an ethical advantage in that it compels the expert to consider the likelihood of hypotheses favoring the defense as well as the prosecution; this appears to have been one reason behind the introduction of the Bayesian CAI (Case Assessment and Interpretation) model used in British forensic science (Lawless 2011, pp. 677–679).

Another controversial form of probabilistic evidence is the prediction of future violence. While some psychiatrists consider it inherently unethical to offer such evidence given the high rate of false-positives in such predictions, Grisso and Appelbaum (1992) argue that it is ethically permissible to give estimates of risk based on sound scientific research. The difficulty with this position, as pointed out by Litwack (1993), is that in many cases where psychiatrists would consider there was clearly a high risk if an offender were released (e.g., the patient has committed serious violence before when actively psychotic and is unwilling to take antipsychotic medication), actuarial data about the risk of reoffending may not be available because of the rarity of people in such supposedly high-risk categories being released.

An expert witness's duties of objectivity and transparency may be difficult to follow to the letter if the expert is also acting as a therapist to the victim (e.g., in a sexual assault case) or to the defendant. It may be difficult to combine the objectivity of a witness with the empathy of a therapist. A therapist's first duty is to the interests of the patient, a witness's is to the court, and these duties may conflict, especially if the witness gives evidence that is harmful or traumatic to the client

(Strasburger et al. 1997). Giving evidence may also conflict with the duty of confidentiality to a patient (e.g., Bal 2005). Such conflicts do not arise in every case, but a potential witness must always consider whether giving the most honest and objective possible evidence is compatible with therapeutic goals and with respect to patient confidentiality.

Accountability

The report of the searching *Inquiry into Paediatric Forensic Pathology in Ontario* defines accountability as "the obligation to answer for a responsibility conferred" by explaining and justifying one's decisions and actions (Goudge 2008, pp. 332–333). In this sense, the courtroom itself is one of the most important fora in which expert witnesses are called to account. In some respects, this can be effective. Timmermans (2006) found that it was the prospect of being called to account in court for every inference they drew that led to the cautious approach adopted by the medical examiners he observed. The care that is usually taken to document the "chain of custody" of DNA samples might also be attributed to the accountability exercised through courtroom cross-examination (Lynch et al. 2008). "History has shown" Aronson (2007, p. 211) concludes in his study of DNA profiling, "that a competent defense review of forensic evidence can reveal problems that are either consciously ignored or overlooked by forensic scientists and prosecutors." However, lack of knowledge, motivation, or ability among defense lawyers and lack of resources to pay for alternative experts allow many errors in forensic evidence to go undetected. Garrett and Neufeld (2009) studied 137 transcripts from US trials where defendants were convicted, partly on the basis of forensic evidence, and later exonerated by DNA evidence. In this admittedly unrepresentative sample of trials, they found that 60 % included scientific evidence that was "invalid" in some respect, usually either the misstatement of empirical data or the drawing of conclusions unsupported by the data. Invalid testimony was rarely subject to cross-examination, and the defense was rarely able to call its own experts, as courts declined to fund them. As Gross and Mnookin (2003, p. 158) observe, "If we are going to rely on the adversary system to guarantee competence and honesty in expert evidence we must actually have an adversary process. When one side is absent, the result may be disastrous."

In addition to competent and adequately funded defense lawyers, adequate disclosure of the basis of expert evidence is essential if the trial is to be an effective form of accountability. Lack of such disclosure has been central to several notorious miscarriages of justice in England. This issue is closely linked to the ethical duty of transparency discussed above. Giannelli (2007) cites several US cases where fraudulent or incompetent forensic scientists have disclosed only very terse and cryptic reports, thus increasing the difficulties the defense face in cross-examining them. The Federal Rules of Criminal Procedure do require that the summary of expert evidence disclosed to the defense must "describe the witness's opinions, the bases and reasons for those opinions, and the witness's qualifications" (r. 16(1)(G)), but this falls far short of the duties of transparency set out in the English Criminal Procedure Rules. The Fingerprint Inquiry Scotland (2011) called for fingerprint examiners to be "more transparent and ultimately more descriptive of their methodology, work practices and the specific factors involved in any particular conclusion" (para. 37.51) so as to facilitate more effective scrutiny of their work in court.

As Giannelli (2007) points out, even when criminal courts do expose shortcomings or misrepresentations in an expert's work, this does not necessarily prevent the same expert from testifying again in other cases. In this respect, civil-law systems with their panels of court-appointed experts have an advantage, since there is usually some form of disciplinary procedure by which experts can be removed from the panel (Jakobs and Sprangers 2000).

Given the uneven nature of the trial process as a form of accountability, it is often thought desirable that expert evidence given at trial should be scrutinized by other experts. The Ontario inquiry discussed a number of mechanisms operating in various parts of Canada for holding pathologists and forensic scientists accountable for evidence they give in court, including observation of a forensic scientist's testimony, or examination of the transcript, by a manager; asking prosecution and defense counsel to complete questionnaires on the expert's evidence; and having each forensic pathologist accompanied to court at least annually by another pathologist who observes and evaluates their testimony (Goudge 2008, p. 358). The report recommended an annual process of peer review for forensic pathology testimony, which "should be documented and encompass a process of discussion and feedback" (Goudge 2008, p. 359).

In England and Wales, the Forensic Science Regulator (2011) has issued a Code of Practice which includes provisions that providers whose staff give evidence in court should be able to demonstrate, inter alia, "that the principles, techniques and assumptions they have relied upon are valid," that the inferences they draw are sound and alternative hypotheses have been considered, that their methodology has been considered by other scientists and any objections have been addressed, and that they can cite specific instances of their methods resulting in demonstrably valid or misleading conclusions and explain how these support or undermine their claims (paras. 25.2.1–25.2.2). These are all matters that could appropriately be raised in cross-examination but often are not; they are also similar to the criteria for admissibility recently recommended by the Law Commission. At the time of writing, the Regulator has no statutory powers (although the Government has indicated that it may reconsider the position). The Code of Practice builds upon an ISO (International Standards Organization) standard, no. 17,025. EU Framework Decision 2009/905/JHA requires all member states to accredit DNA laboratories to this standard by November 2013 and fingerprint examiners by November 2015. With the publicly owned Forensic Science Service (FSS) having been closed down in 2012, private contractors bidding for forensic work are expected to be accredited by the UK Accreditation Service under ISO 17025 (House of Commons 2011, para. 97). As yet, there is no such requirement for police in-house forensic services which have taken over some of the work formerly done by the FSS (House of Commons 2011, paras. 101–109).

Other forms of accountability involve disciplinary or legal proceedings against experts who have failed in their duties. These have recently been the subject of two major cases in England. *Meadow v General Medical Council* concerned an eminent pediatrician who was struck off the medical register following his statistically illiterate evidence in the case of a mother convicted of murdering her two young children. At first instance ([2006] EWHC 146 (Admin)), Mr Justice Collins held that the common law immunity of witnesses from suit in respect of their evidence protected them not only from civil actions for damages but also from disciplinary proceedings by professional bodies. This ruling, based largely on policy considerations about the deterrent effect of possible complaints on the willingness of medical experts to undertake forensic work, was overturned on appeal. The Court of Appeal's view was consistent with the US authorities on this issue (Sanders 2007, pp. 1564–1565). Again the main ground was one of policy: the need to protect the public against unreliable experts. The majority of the Court, however, considered that Professor Meadow's misconduct in testifying erroneously on statistical matters outside his expertise had to be set in its "forensic context" ([2006] EWCA Civ 1390, [89], [205]). Defense counsel had failed to challenge the admissibility of this evidence or to expose Professor Meadow's fallacious reasoning in cross-examination, and the evidence they later called to refute his claims may have been insufficient to counter the "major effect" which the Court of Appeal suspected his vividly worded evidence to have had on the jury (*R v. Clark* [2003] EWCA Crim 1020, [178]). Since the lawyers carried some of the blame for what had gone wrong, the judges considered that Professor Meadow had not committed *serious* professional misconduct by straying beyond his expertise. He was therefore reinstated on the medical register.

In *Jones v. Kaney* [2011] UKSC 13, the Supreme Court decided that expert witnesses should not enjoy immunity from suit for breach of their duty to the party instructing them. The defendant had signed a joint statement by the experts in civil proceedings for personal injuries, which she later admitted did not reflect her true opinion and which undermined her client's claim for damages. Lords Phillips and Dyson both indicated, however, that civil liability would also extend to testimony by a defense expert in a criminal case (ibid. [60], [125]) – although a civil action by a convicted defendant would be stayed as an abuse of process if the conviction had not been quashed on appeal. As Lord Hope pointed out in his dissenting judgment, this produces the seemingly anomalous result that a prosecution expert continues to enjoy immunity but a defense expert does not. Lord Collins, who concurred with the majority, pointed to a number of US authorities holding that "friendly" experts are not immune to civil actions by their clients. These cases, however, all concern expert evidence in civil litigation. Witnesses in US criminal trials enjoy immunity from suit even if they perjure themselves (*Briscoe v LaHue* 460 US 325 (1982)).

Conclusion

Jones v Kaney brings us back to the ambiguity of the expert witness's role in an adversarial system. The defense expert's duty to the client is, in principle, identical to her duty to the court – a duty to give honest and impartial evidence in accordance with acceptable professional standards. Similarly the prosecution expert, particularly in the case of a private provider in the neoliberal economy of forensic science (Lawless 2011), is under a contractual duty to comply with standards of impartiality and scientific rigor. The idea that duties of impartiality are best enforced through legal accountability to the partisan purchasers of expert services may appear somewhat paradoxical.

On a more optimistic note, it may be that a model of accountability is gradually emerging under which courtroom processes and regulatory systems reinforce one another. If admissibility requirements (see ▶ Scientific Evidence in Criminal Prosecutions), the cross-examination of witnesses and the requirements of regulators are all informed by similar criteria of impartiality and reliability; if regulators can keep track of experts' performance in court; and if lawyers are aware of experts' compliance or noncompliance with regulatory requirements, we may develop an appropriate set of ethical standards for expert witnesses underpinned by an effective system of accountability for their breach.

Related Entries

▶ Behavioral Science Evidence in Criminal Trials
▶ Cognitive Forensics: Human Cognition, Contextual Information, and Bias
▶ Expert Evidence and Criminal Trial Procedure
▶ Forensic Science and the Paradigm of Quality
▶ Forensic Science and Miscarriages of Justice
▶ Legal Rules, Forensic Science and Wrongful Convictions
▶ Scientific Evidence in Criminal Prosecutions
▶ Scientific Evidence in Criminal Adjudication

Recommended Reading and References

AAPL (American Academy of Psychiatry and Law) (2005) Ethics guidelines for the practice of forensic psychiatry. http://www.aapl.org/ethics.htm. Accessed 11 Feb 2012

Aronson JD (2007) Genetic witness: science, law and controversy in the making of DNA profiling. Rutgers University Press, New Brunswick

Bal R (2005) How to kill with a ballpoint: credibility in Dutch forensic science. Sci Technol Hum Val 30:52–75

Barclay D (2009) Using forensic science in major crime inquiries. In: Fraser J, Williams R (eds) Handbook of forensic science. Willan, Cullompton

Candilis PJ, Weinstock RL, Martinez R (2007) Forensic ethics and the expert witness. Springer, New York

Cole S (2001) Suspect identities: a history of fingerprinting and criminal identification. Harvard University Press, Cambridge

Dror I, Rosenthal D (2008) Meta-analytically quantifying the reliability and biasability of forensic experts. J Forensic Sci 53:1–4

Dwyer D (2008) The judicial assessment of expert evidence. Cambridge University Press, Cambridge

Fingerprint Inquiry Scotland (2011) *The fingerprint inquiry report*. Available at www.thefingerprintinquiryscotland.org.uk/inquiry/3127.html. Accessed 12 Feb 2012

Forensic Science Regulator (2011) *Codes of practice and conduct for forensic service providers and practitioners in the criminal justice system*. Available at http://www.homeoffice.gov.uk/publications/agencies-public-bodies/fsr/codes-practice-conduct?view=Binary. Accessed 27 April 2012

Garrett BL, Neufeld PJ (2009) Invalid forensic science and wrongful convictions. VA Law Rev 95:1–99

Giannelli PC (2003) *Ake v. Oklahoma:* the right to expert assistance in a post-*Daubert,* post-DNA world. Cornell Law Rev 89:1305–1419

Giannelli PC (2007) Wrongful convictions and forensic science: the need to regulate crime labs. N C Law Rev 86:163–235

Goudge, Hon. Stephen T (2008) *Inquiry into paediatric forensic pathology in Ontario: report*. Available at www.attorneygeneral.jus.gov.on.ca/inquiries/goudge/index.html. Accessed 11 Feb 2012

Grisso T, Appelbaum PS (1992) Is it unethical to offer predictions of violence? Law Hum Behav 16:621–633

Gross SR, Mnookin JL (2003) Expert information and expert evidence: a preliminary taxonomy. Seton Hall Law Rev 34:141–189

Hardwig J (1994) Toward an Ethics of Expertise. In: Wueste DE (ed) Professional ethics and social responsibility. Rowman & Littlefield, Lanham, MD

House of Commons, Science and Technology Committee (2005) *Forensic science on trial: 7th report of the session 2004–5* (HC 96-i). The Stationery Office, London

House of Commons, Science and Technology Committee (2011) *Seventh report – the forensic science service*. Available at www.publications.parliament.uk/pa/cm201012/cmselect/cmsctech/855/85502.htm. Accessed 12 Feb. 2012

Jakobs LEMP, Sprangers WJJM (2000) A European view on forensic expertise and counter-expertise. Crim Law Forum 11:375–392

Lawless CJ (2011) Policing markets: the contested shaping of neo-liberal forensic science. Brit J Criminol 51:671–689

Litwack T (1993) On the ethics of dangerousness assessments. Law Hum Behav 17:479–482

Lynch M, Cole SA, McNally R, Jordan K (2008) Truth machine: the contentious history of DNA fingerprinting. University of Chicago Press, Chicago

McQuiston-Surrett D, Saks MJ (2009) The testimony of forensic identification science: what expert witnesses say and what factfinders hear. Law Hum Behav 33:436–453

Risinger DM, Saks MJ, Thompson RC, Rosenthal R (2002) The *Daubert*/*Kumho* implications of observer effects in forensic science: hidden problems of expectation and suggestion. Southern Calif Law Rev 90:1–56

Sanders J (2007) Expert witness ethics. Fordham Law Rev 76:1535–1584

Smith R (1989) Forensic pathology, scientific expertise and the criminal law. In: Smith R, Wynne B (eds) Expert evidence: interpreting science in the law. Routledge, London

Strasburger LH, Gutheil TD, Brodsky A (1997) On wearing two hats: role conflict in serving as both psychotherapist and expert witness. Am J Psychiatry 154:448–456

Timmermans S (2006) Postmortem: how medical examiners explain suspicious deaths. University of Chicago Press, Chicago

Wynne B (1989) Establishing the rules of laws: constructing expert authority. In: Smith R, Wynne B (eds) Expert evidence: interpreting science in the law. Routledge, London

Extended-Release Naltrexone

► Drug Abuse and Alcohol Dependence Among Inmates

Extent of Imprisonment: Global View

Roy Walmsley
World Prison Brief, The International Centre for Prison Studies, London, UK

Overview

The extent of imprisonment worldwide cannot be determined precisely, but a broad estimate can be achieved. The opportunity to compare the figures for individual countries provides the basis for assessing the need for policy changes. The first attempt at compiling a complete list of world prison populations was published in 1999; since then, information has been traced on the extent of imprisonment in almost all countries.

The estimated world prison population is now in excess of 10 million. The rate per 100,000 of the world's population (the prison population rate) is around 150. The country with the highest prison population rate is the USA with a rate of 730 per 100,000 of the national population. Fifteen other countries have rates over 400. By contrast, 15 countries have rates below 40; ten of these are in Africa. Prison population rates vary considerably between different regions of the world, and between different parts of the same continent.

Female imprisonment constitutes between 2 and 9 % of the total prison population in most countries. Six of the seven highest levels (between 15 and 22 %) are in Asian prison systems. Overall, more than 625,000 women and girls are held in penal institutions throughout the world.

Pretrial/remand prisoners constitute the vast majority of the prison population in some countries and only a small minority in others. Nine of the 12 highest levels (over 70 %) are in Africa. In 16 countries, the level is below 10 %. It is likely that there are now more than three million in pretrial/remand imprisonment.

Foreign prisoners generally constitute less than 30 % of the prison population, but in six countries more than 70 % of prisoners are foreign nationals.

Overcrowding in the world's prison systems is evidenced by the fact that the official capacity of the systems is exceeded in three fifths of the 196 countries on which information is available. In 23 countries, the occupancy level is more than twice the intended capacity.

Introduction

Imprisonment may be defined as custodial confinement as a result of being suspected or convicted of one or more criminal offenses.

It is not possible to know precisely the extent of imprisonment worldwide, but a broad estimate can be achieved.

There are several reasons why the precise level cannot be known, including the following:

(a) Some countries never reveal the number of people detained in prisons or other penal institutions as a result of being suspected or convicted of criminal offenses – e.g., the Democratic People's Republic of Korea (North Korea), Eritrea, and Somalia.

(b) Some countries only publish incomplete figures – e.g., the People's Republic of China only reveals the number of sentenced prisoners in Ministry of Justice facilities, while Maldives and Vietnam only reveal the number of sentenced prisoners; the number of pretrial/remand prisoners in these countries is not published.

(c) Some countries do not include in their prison population totals young people suspected or convicted of a criminal offense but held in a form of custody that is not known as a prison. Likewise, some do not include persons suspected or convicted of a criminal offense but held in custody in a police-run institution or a psychiatric facility.

(d) Some countries include in their prison population totals persons detained as illegal immigrants, even though they are not suspected or convicted of a criminal offense.

(e) People held in a form of custody that is not under the authority of the national prison administration are not usually included in national prison population totals.

(f) The information available as to the number of prisoners in each country does not relate to the same date.

Many of the above considerations do not generally have a large effect on the calculation of the extent of imprisonment worldwide. For example, in those countries where juveniles are held in custody in respect of a criminal offense but excluded from the published prison population total the number of such juveniles is generally not large, and the same is true of the number of persons held in psychiatric facilities as a result of criminal offenses.

It is (a) and (b) above that are the main obstacles to knowing the full extent of imprisonment worldwide. But by assuming that the prison population levels in the countries for which information is missing or incomplete are similar to the median figures for the subcontinental region in which they are situated, a broad estimate can be achieved.

The availability of information about the extent of imprisonment in the individual countries enables governments and prison administrations, and all those who have an interest in improving criminal justice, to compare the level of imprisonment in one country with that in another and to consider the reasons and justifications for the differences that are revealed. This provides the opportunity for the development of evidence-based policy changes.

Background of the Data Collection

Many countries have for a long time been recording the number of prisoners in their own prisons and detention centers, although others have only started doing so quite recently, and it is believed that there are still a small number of countries that do not keep such records on a regular basis even now.

But efforts to assemble prison population totals in a large number of countries have been very few. The Howard League for Penal Reform assembled figures (in respect of a date between December 1932 and June 1936) for 54 countries in "The Prisoner Population of the World," a document submitted to the League of Nations in August 1936.

The International Penal and Penitentiary Commission (now the International Penal and Penitentiary Federation) assembled figures (mostly in respect of 31 December 1936) for 55 countries, in response to a request by the League of Nations, in "Enquiry into the number of prisoners and measures taken to reduce it" a report submitted to the League of Nations in 1938.

In the 1970s, the United Nations commenced its series of surveys of "Crime Trends and the Operations of Criminal Justice Systems" which include a section that includes prison population data. These surveys (the most recent, the 12th, obtained data for 2009) generally receive prison data from 60 to 65 countries, about half of them in Europe.

In 1983, the Council of Europe commenced its annual series of prison statistics in its member states "Council of Europe Annual Penal Statistics" or, using its French language acronym, "SPACE" (Tournier 1983). Reflecting the increasing membership of the Council, the number of countries for which prison population data was received has grown from 19 in 1983 to 31 in 2000 and 47 (52 prison systems) in the most recent (2010) edition (Aebi and Delgrande 2012).

The first attempt at producing a complete list showing the number of prisoners in each of the world's countries was compiled in 1998 and published by the British Home Office (Walmsley 1999, 2001). Containing data for 180 countries, it acknowledged the limitations described in the above introduction and drew attention to the desirability of improving the reliability and comprehensiveness of the data and of including information on the proportion of females and of pretrial detainees in penal institutions and on occupancy levels.

The World Prison Population List has now reached its ninth edition (Walmsley 2011), which contains figures for 218 countries. From 2000, it has been complemented by the online World Prison Brief, which has substantially developed the information available on the extent of each country's imprisonment level along the lines indicated above, including the proportion of foreign prisoners, information on trends and lists ranking countries from the highest to the lowest on prison population data variables.

Overall Imprisonment Levels

Global overview. The world prison population rate, calculated on the basis of 10.1 million prisoners and a world population of 6.9 billion, is 146 per 100,000, or 156 per 100,000 on the basis of a world prison population of 10.75 million (including 650,000 in Chinese "detention centers") (Walmsley 2011). The median rate is 134 (World Prison Brief, accessed 10.8.12).

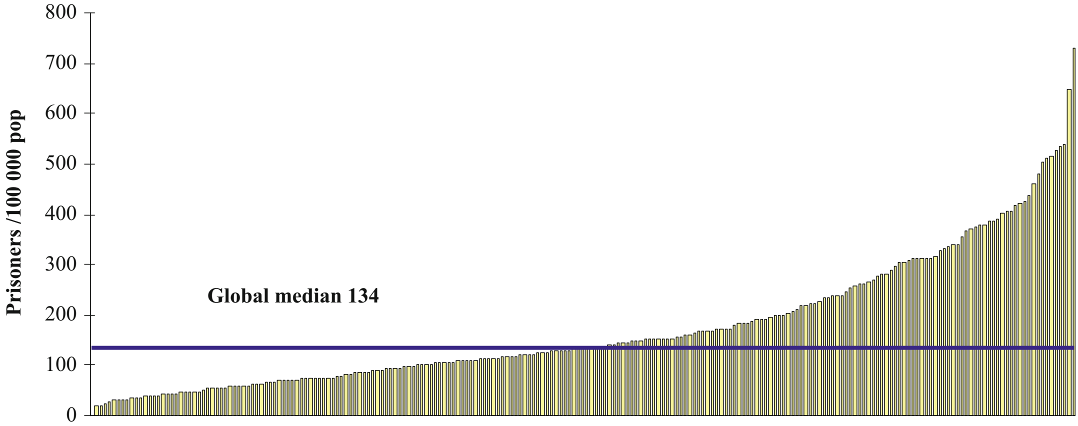

Extent of Imprisonment: Global View, Fig. 1 Prison population rates in 221 countries (Source: World Prison Brief (ICPS), Lappi-Seppälä 2013)

The country with the highest prison population rate is the USA with a rate of 730 per 100,000 of the national population. Fifteen other countries have rates over 400 per 100,000, seven of which are small Caribbean islands – Saint Kitts and Nevis (649), US Virgin Islands (539), Anguilla (480), British Virgin Islands (c. 460), Curacao (422), Bermuda (417), and Grenada (402); the other eight are Seychelles (535), Rwanda (527), Georgia (514), Cuba (510), Russian Federation (505), Belarus (438), and Azerbaijan and Belize (both 407) (World Prison Brief, accessed 10.8.12).

The countries with the lowest prison population rates are predominantly in Africa. This applies to ten of the fifteen countries that have rates under 40 per 100,000, namely, Liberia and Guinea (both 39), Mali (36), Chad (34), Republic of Congo and Democratic Republic of Congo (both 33), Nigeria (31), Burkina Faso (28), and Central African Republic and Comoros (both 19); the other five are Liechtenstein (39), Monaco (34), India (30), Faeroe Islands (21), and Timor-Leste (20) (ibidem).

Figure 1 displays the number of prisoners relative to 100,000 population (the prison population rate) in 221 countries. The figures range from below 20 to 730.

Prison population rates vary considerably between different regions of the world, and between different parts of the same continent. For example:

- In Africa, the median rate for western African countries is 47.5, whereas for southern African countries it is 219.
- In the Americas, the median rate for south American countries is 175, whereas for Caribbean countries it is 357.5.
- In Asia, the median rate for south central Asian countries (mainly the Indian subcontinent) is 42, whereas for eastern Asian countries it is 155.5.
- In Europe, the median rate for western European countries is 96, whereas for the countries spanning Europe and Asia (e.g., Russia and Turkey) it is 228.
- In Oceania (including Australia and New Zealand) the median rate is 135 (Walmsley 2011).

Figure 2 demonstrates the variation between different regions in Europe.

Nordic countries as a region have the lowest level of prisoners in Europe (with Slovenia and Switzerland on the same level). The Nordic figures vary from 47 to 74. The corresponding figures for Western European countries are 64–104, Southern Europe 108–168, Britain and Ireland 97–154, Central and Eastern Europe 146–221, Baltic countries 252–314, and countries of former Soviet Eastern Europe 338–514 (for explanations of differences, Lappi-Seppälä 2008, 2013).

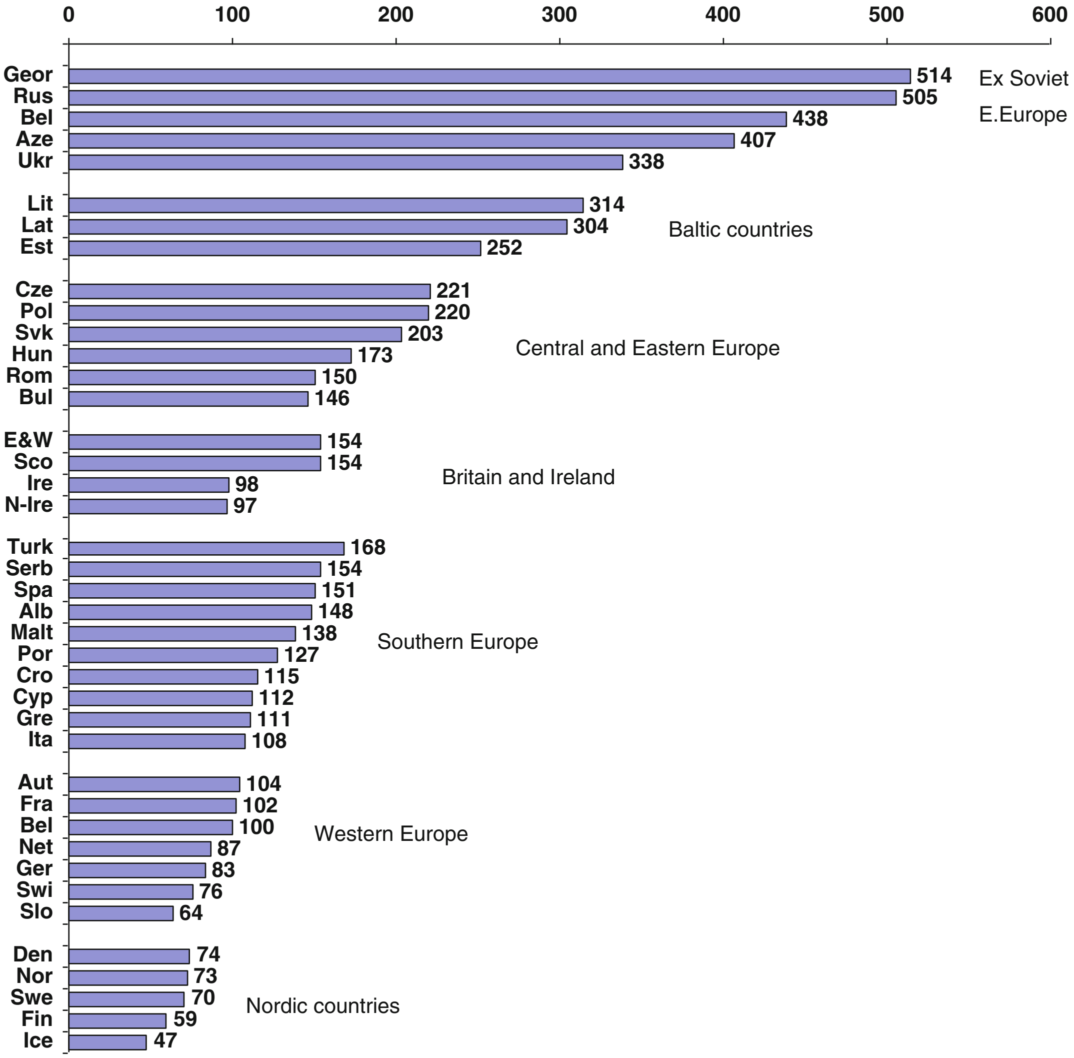

Extent of Imprisonment: Global View, Fig. 2 Prison population rates in 40 European countries (Source: World Prison Brief (ICPS))

Trends and changes. The estimated world prison population has risen from 8 million (based on the latest data available at November 1998) to more than 10.1 million (based on the latest data available at May 2011) – or 10.75 million if persons in Chinese "detention centers" are included (Walmsley 1999, 2011, for global trends 1997–2007 see Walmsley 2010 and Lappi-Seppälä 2013).

Figure 3 displays percentage changes in prison population rates from 1992 to 2012 in around 100 nations for which data is available.

More than four out of five countries have experienced increasing prison population rates since the early 1990s. Some have more than tripled their rates and the majority of countries have at least doubled their rates.

Imprisonment and Specific Groups

Female imprisonment constitutes between 2 and 9 % of the total prison population in most countries. The highest levels are in six Asian prison

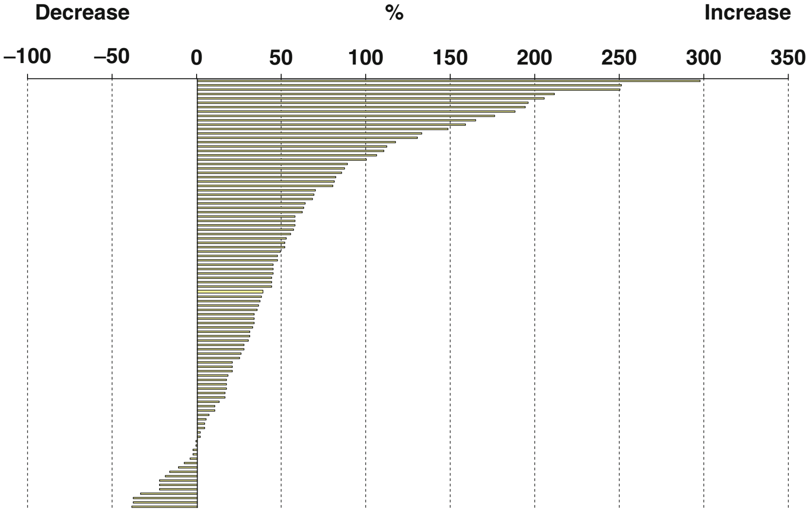

Extent of Imprisonment: Global View, Fig. 3 Percentage changes in prison population rates from 1992 to 2012 (Source: World Prison Brief (ICPS), Lappi-Seppälä 2013)

systems – Maldives (22 %); Hong Kong-China (20 %); Bahrain (18.5 %); Macau-China, Qatar, and Thailand (all 15 %); and also in Andorra (16 %) (Walmsley 2012).

There are continental variations in the prevalence of women and girls within the total prison population. The median level is 4.45 %, but in African countries, female prisoners constitute a much smaller percentage of the total (the median is 3.1 %) than in Asia where the median is almost twice as high (5.95 %). The median levels in the Americas, Europe, and Oceania are 5.15 %, 4.9 %, and 3.9 %, respectively (ibidem).

Overall, more than 625,000 women and girls are held in penal institutions throughout the world (ibidem). Nearly a third of these are in the USA (201,200). The next three countries in terms of numbers are China (with 84,600 plus women and girls in pretrial detention or "administrative detention"), the Russian Federation (59,200), Brazil (35,596), and Thailand (29,175). The other countries with more than 7,000 female prisoners are India (15,406), Vietnam (12,591 plus women and girls in pretrial detention), Mexico (10,072), Ukraine (9,697), and the Philippines (7,726).

There is evidence that the female prison population has increased by more than 16 % between 2006 and 2012. Further, the prevalence of women and girls within the total prison population has increased in recent years, especially in Asia and Europe (ibidem).

Pretrial/remand prisoners constitute the vast majority of the prison population in some countries and only a small minority in others. In 12 countries, nine of them in Africa, more than 70 % of prisoners are in this category. By contrast, in 16 countries pretrial/remand prisoners constitute less than 10 % of the prison population. The median level is 30 % (World Prison Brief, accessed 10.8.12).

In 2007, two and a quarter million persons were known to be held in pretrial and other forms of remand imprisonment, and it was estimated that a further quarter of a million were so held in countries on which such information was not available (Walmsley 2008). The subsequent increase in the total prison population would suggest that there are now at least three million in pretrial/remand imprisonment.

Foreign prisoners constitute less than 30 % of the prison population in most countries, but more than half the prison population are foreign nationals in 15 countries and in six of these foreign nationals make up more than 70 % of prisoners – United Arab Emirates and Monaco (both 92 %), Qatar (73 %), Saudi Arabia (72 %), and Liechtenstein and Switzerland (both 71 %) (World Prison Brief, accessed 10.8.12).

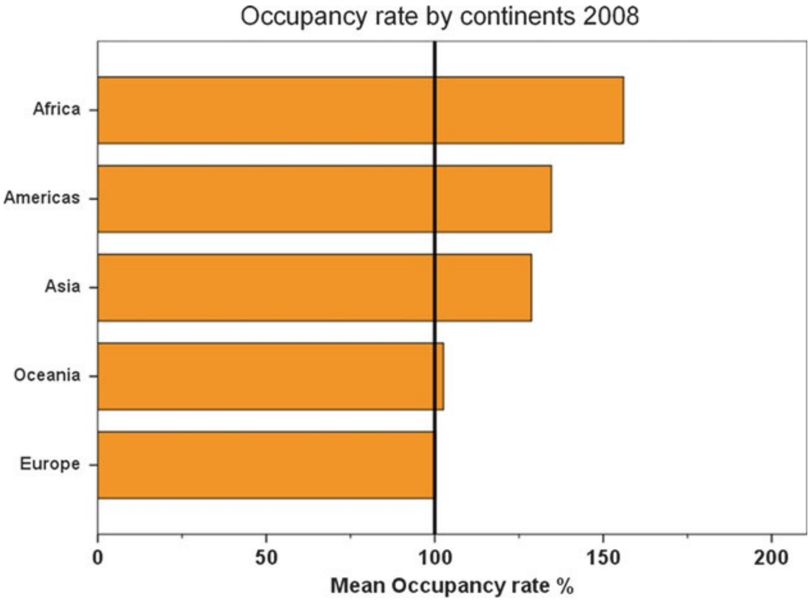

Extent of Imprisonment: Global View, Fig. 4 Occupancy rate by continents 2008

Overcrowding

High levels of imprisonment are invariably accompanied by overcrowding. If a country has a prison population that exceeds its official capacity (i.e., has an occupancy rate in excess of 100 %) it can safely be assumed that there is overcrowding and the higher the rate the more likely that the overcrowding is of a serious nature.

Three fifths of the 196 countries on which information is available (ibidem) have occupancy rates in excess of 100 %; 49 of these have rates over 150 %; and 23 of the 49 have rates over 200 %, which means that two prisoners are accommodated in the space intended for one. In the worst cases, the number of prisoners exceeds the number of prison places by three to one.

Eleven of the 23 nations with rates over 200 % are in Africa, a continent in which there are only a small number of countries where the prison system is not known to be over capacity. In the Americas, the system is over capacity in over three quarters of countries. Many prison systems are overcrowded also in Asia, where most countries in South Central Asia have rates over 150 %. The situation is less bad in Europe and Oceania; however, even in these continents about half the prison systems are over capacity.

Figure 4 shows the continental variations in occupancy levels in 2008.

Overcrowding is even worse than these figures suggest. This is because the official capacity of institutions is set in many countries at a level that provides prisoners with less space than international human rights bodies consider adequate. Furthermore, even in a prison system that is not overcrowded when taken as a whole, there are often individual prisons that are overcrowded – and sometimes severely overcrowded.

Recommended Reading and References

Aebi M, Delgrande N (2012) Council of Europe Annual Penal Statistics, SPACE I, Survey 2010, Strasbourg

Howard League for Penal Reform (1936) The prisoner population of the world, document submitted to the League of Nations

International Penal and Penitentiary Commission (1938) Enquiry into the number of prisoners and measures taken to reduce it, report submitted to the League of Nations

Lappi-Seppälä T (2013) Imprisonment and penal demands. Exploring the dimensions and drivers of systemic and attitudinal punitivity. In: Body-Gendrot S, Hough M, Kerézsi K, Levý R, Snacken S (eds.) European handbook of criminology. Routledge. Forthcoming 2013

Lappi-Seppälä T (2008) Trust, welfare, and political culture: explaining difference in national penal policies. In: Tonry M (ed) Vol. 37 of crime and justice: a review of research. The University of Chicago Press, Chicago, pp 313–387

Tournier P (1983) Council of Europe annual penal statistics. In: Prison information bulletin no. 2, Council of Europe, Strasbourg.

United Nations (1975) onwards (online) Surveys of ‘Crime Trends and the Operations of Criminal Justice Systems’ (see www.unodc.org)

Walmsley R (1999) World prison population list, 1st edn. Research findings no. 88, Home Office, London

Walmsley R (2001) World prison populations – an attempt at a complete list. In: van Zyl Smit D, Duenkel F (eds) Imprisonment today and tomorrow, 2nd edn. Kluwer, The Hague

Walmsley R (2008) World pre-trial/remand imprisonment list. International Centre for Prison Studies, London

Walmsley R (2010) Trends in world prison population. In: Harrendorf S, Heiskanen M, Malby S (eds.)International statistics on crime and justice. HEUNI Publication Series No. 64. Helsinki

Walmsley R (2011) World prison population list, 9th edn. International Centre for Prison Studies, London

Walmsley R (2012) World female imprisonment list, 2nd edn. International Centre for Prison Studies, London

World Prison Brief (2000) onwards (online). International centre for prison studies (www.prisonstudies.org)

Extortion Racketeering

► Racketeering

Eyewitness Evidence: Historical Background

James M. Doyle
Independent Researcher, Carney & Bassil,
Boston, MA, USA

Overview

The growing general recognition that eyewitness memory does not operate as a stable DVR or videotape record of experience but rather requires a complex multistage process of encoding, storage, and reconstruction that is vulnerable at every step to influence and contamination has intersected with a growing list of exonerations of wrongfully convicted individuals in eyewitness cases revealed by DNA comparisons. There is now a widespread perception that honest but mistaken eyewitnesses are the leading cause of wrongful convictions.

Law enforcement and legal practitioners have long recognized that their dangerous dependency on unreliable identifications of perpetrators by eyewitnesses in criminal cases is an important weakness in criminal justice systems (Borchard 1932). Since at least the end of the nineteenth century, social scientists have investigated the sources of eyewitness error and have attempted to persuade criminal justice practitioners to apply their findings in modernizing investigative and adjudicatory practices. The result has been a long and complicated relationship between researchers and criminal justice practitioners: a dynamic of reciprocal challenge and response that has influenced both the research agenda of the scientists and the daily routines of the practitioners. The history of the psychology/criminal justice interface in the eyewitness context parallels and may in fact have formed a template for interactions in other contexts such as the evaluation of interrogations leading to false confessions.

Concern with eyewitness unreliability is an international phenomenon that has attracted the attention of a polyglot array of researchers (Sporer and Cutler 2003; Roberts 2009), but the structure of the dialogue concerning eyewitnesses can be seen most clearly in common-law countries where investigative fact development is later subjected to adversarial end-stage adjudicative inspections. The collision between social science research and criminal justice system practice is starkly illuminated by experience in the United States, where the relationship of the science of eyewitness memory and the criminal justice system can be described as a chronicle of running battles that have stretched over more than 100 years.

This historical background exercises a significant influence over current discussions and contributes to the fact that while there is now a very substantial body of research and commentary addressing the eyewitness issue, “A significant shortcoming in much of the writing on the subject is the adoption of a parochial and procedurally atomistic approach” (Roberts 2009).

Early Encounters: Munsterberg and His Peers

The basic battle lines or the interaction of the psychology researchers and the criminal justice community was laid down in early controversies initiated by Hugo Munsterberg, a student in Wundt’s earliest psychology courses, who was brought to Harvard by William James to run Harvard’s experimental psychology laboratory. Munsterberg, an energetic promoter of his new discipline, had outlined in his best selling 1908 book, *On the Witness Stand*, a remarkably prescient series of insights concerning the operation of witness memory. He provided, for example, demonstrations that eyewitness mistakes were frequent, that memory was not permanent but malleable, and that witness confidence was not sure evidence of witness accuracy (Munsterberg 1908). These findings had obvious forensic applications, but Munsterberg, in his self-appointed role as publicist and controversialist, complained that although other disciplines and professions were eager to learn the lessons about eyewitness memory that his new field of experimental psychology was beginning to teach: “the lawyer alone is obdurate.” Munsterberg charged that the lawyers chose traditional primitive ignorance over scientific enlightenment.

This charge drew the attention of John Henry Wigmore, the Dean of Northwestern University School of Law and the preeminent evidence scholar of his era. Wigmore was an early proponent of the integration of scientific psychology into legal practice, and in his article responding to Munsterberg’s charges, he provided an appendix that compiled a full list of the extant psychological material and that indicated that Wigmore had actually reviewed that material more thoroughly than had Munsterberg himself (Doyle 2005).

The real purpose of Wigmore’s article was to illuminate the potential in a law and psychology relationship and to throw his prestige behind its inception. Wigmore agreed that the legal system had neglected psychologists’ research. Wigmore’s goal was to herald the day when the lawyers and psychologists could move forward in “a friendly and energetic alliance in the noble cause of justice” (Wigmore 1909). Even so, according to Wigmore, Munsterberg was premature in announcing the immediate utility of such an alliance; Wigmore thought that was still on the distant horizon. Wigmore pointed out that while Munsterberg’s findings on the reliability of witnesses were highly significant, the legal system’s concern was not with the reliability of *witnesses* in general but with the reliability of *verdicts* in individual cases. Munsterberg’s mistake, according to Wigmore, was his assumption that contemporary psychology could immediately help the criminal justice system to sort the correct eyewitnesses from the incorrect ones. Wigmore argued that day had not arrived. Still, Wigmore looked forward to its arrival, and he was confident that day would come. “When the psychologists are ready for the courts,” he announced in a subsequent piece, “the courts will be ready for the psychologists” (Wigmore 1937).

But it was not the substance of Wigmore’s response to Munsterberg that was best remembered by psychologists; it was the caustic satirical tone of his article. Wigmore’s criticism is still recalled as a paradigm of the kind of welcome social scientists should expect from the legal system and its practitioners (Doyle 2005).

In fact, Wigmore did want to issue a mild call to order: to correct Munsterberg’s overstatements and to address Munsterberg’s misapprehensions about the realities of legal practice. But Wigmore structured his article as a mock trial of a Munsterberg proxy for libeling the legal system and then proceeded to humiliate this fictional psychologist and, by extension, to humiliate psychology. Deployed in this starkly adversarial format, Wigmore’s Olympian status in the legal world and Wigmore’s corrosive sarcasm combined were devastating and drove Munsterberg from the field. After Wigmore’s article, although research continued in the laboratories, the question of the proper role of psychology in assessing eyewitness memory lay dormant as a public issue for close to 70 years.

The Expert Witness Era: Robert Buckhout and Elizabeth Loftus

Psychologist Robert Buckhout picked up Munsterberg's fallen banner in the politically charged atmosphere of the United States in the 1970s, and he relied on a modernized version of Munsterberg's demonstration approach (Doyle 2005). For example, Buckhout induced a New York television station to broadcast a staged crime and invite viewers to make choices from a staged lineup. The number of correct identifications this process yielded was lower than would have been achieved by random guessing (Doyle 2005). But while his method may have been similar to Munsterberg's, Buckhout's temperament was very different from his predecessor's. Munsterberg was an academic who retreated when faced with Wigmore's onslaught. Buckhout knew his science, but he was a cheerful agitator who carried the battle into the courts and into the popular media.

Buckhout published an accessible survey article on eyewitness error in the general audience magazine *Scientific American* (Buckhout 1974). He testified on the unreliability of eyewitness testimony in the trial of California radical Angela Davis and was instrumental in winning her acquittal. He seized every opportunity to comment in the media (e.g., opining on the case of a butcher identifying his own pork chops from a pork-chop lineup) where the lessons of eyewitness psychology could be taught. His science was aligned with his politics. Buckhout believed that criminal defendants, particularly poor and minority indigent defendants, were being abused by the legal system's complacent reliance on an antique view of how memory worked. He almost immediately rallied two groups of partners.

The first group was a cohort of idealistic younger psychologists, like Elizabeth Loftus, who were anxious to see that their science have an impact in the world. Loftus attacked the eyewitness issue in a radically different way: She "did science" in the form of rigorously controlled experiments, changing one variable while holding all others constant. The results she began to produce were striking. She showed, for example, that when questions about a white barn were introduced into interrogations of witnesses who had viewed a film of an auto accident, over 20 % of those viewers later reported seeing a white barn although in fact there had been no white barn in the film (Loftus and Palmer 1974).

This was a crucial finding for criminal eyewitness cases: It showed that eyewitness memory not only decayed but also *changed*. It showed that a witness could not only forget the right man but also – after being unknowingly influenced by viewing mug shots or showups (which operate as "post-event information" like the white barn in an interview question) – could remember the *wrong* man.

Loftus's findings and those of her colleagues mounted quickly, and they went to the heart of the eyewitness experience as it is encountered in the criminal justice system. Taken together they indicated that in an eyewitness case, the memory of the witness is for all practical purposes the scene of the crime. They showed that memory evidence was, in effect, "trace evidence": difficult to collect, easy to contaminate, but impossible to test for contamination after any contamination has occurred.

At the same time, Loftus's scrupulous scientific methods were winning her work admission to the blue-ribbon, peer-reviewed academic journals and encouraging younger academic psychologists to extend and challenge her research. Experimental findings such as Loftus's (unlike the demonstrations of Munsterberg and Buckhout) could be replicated or falsified. The number of published studies multiplied (Cutler and Penrod 1995). Forensically relevant findings on the effects of stress, own-race bias, and post-event information burgeoned.

By the late 1970s, the history of the eyewitness issue in the United States had become completely entangled with the peculiar features of common-law adversarial tradition. The United States

Supreme Court had ruled in a series of cases that identifications that had been so infected by police misconduct as to be irreparably unreliable would be excluded from trials on constitutional grounds (Manson v. Braithwaite 1977). But those rulings left open the question of how to assess the credibility of identifications that were not unconstitutional but may have been less reliable than jurors' "commonsense" understanding of memory as a stable DVR or videotape capacity suggested.

Tantalized by psychological research's potential for undermining confidence in that range of identifications, Buckhout's second group of recruits, the desperate criminal defense lawyers, joined in. Buckhout's testimony in the Angela Davis case had drawn their attention, and his *Scientific American* article quickly circulated through the defense bar. Elizabeth Loftus published a popular general audience account of eyewitness science, *Eyewitness Testimony*, at about this time, and that contribution was buttressed by an influential Stanford Law Review comment written by Frederick Woocher (a trained psychologist then in law school), which provided a blueprint for arguments for conveying psychological science through expert witnesses. Defense lawyers began to demand the admission of expert testimony before juries by Loftus, Buckhout, and their colleagues aimed at debunking faith in eyewitness evidence.

This point of entry was bad luck for anyone who hoped for Wigmore's "friendly and energetic alliance." Persistent litigation over admissibility of psychological testimony did help to keep the issue of eyewitness science alive in the courts, and feedback from skeptical courts did help to provoke new, better-targeted research. But these benefits came at a steep price.

The initial environment has affected discussions of eyewitness science in criminal justice ever since. Admissibility questions arise at the most acutely adversarial moments of the criminal process, and their resolution (at least in the eyes of the contending advocates) may determine who wins and who loses a trial. The US prosecutors – goaded by inflammatory rhetoric from Buckhout – quickly denounced eyewitness findings as enemy pseudoscience: a trick designed to let criminals go free by unnerving credulous lay jurors and sliming all eyewitnesses, most of whom were right and many of whom were crime victims. For some American prosecutors – then and now – eyewitness science is simply a shield for the guilty. For many judges, the cumulative price of the skirmishing over marginally interesting science the experts offered seemed enormous in terms of hours, dollars, and distended docket backlog.

The critics of expert psychological testimony garnered important support from within the research community. Well-respected psychologists took the position that the field's existing body of knowledge was crippled by problems of ecological validity and theoretical inadequacy and was insufficient to aid the accurate resolution of disputed identification questions. It was the position of this group – which was quickly found and utilized by prosecutors – that efforts to use experts' psychological testimony promised more harm than good: more juror confusion than enlightenment, more guilty people freed than innocent people protected. In this view, the testimony of Buckhout and his successors misappropriated the prestige of science to produce only unjustified and destructive general skepticism about eyewitnesses (McCloskey and Egeth 1983). Since the prevailing test for the admission of novel expert testimony in the United States at that time focused on the "general acceptance" of the science offered by the proposed experts within the relevant scientific community, adversarial battles over the admissibility of evidence focused tightly (for purely forensic reasons) on this fault line within the research community.

This dispute had the unexpected effect of acting as a goad to further research, as researcher-experts attempted to address the criticisms of skeptical adversaries and colleagues. A substantial body of knowledge began to accumulate, one peer-reviewed publication at a time. By 1989, Kassin and his colleagues, in a survey of psychologists, were able to generate a list of propositions – for example, the focus on a weapon degrades eyewitness reliability – with which 80 % of psychologist-experts agreed

(Kassin et al. 1989). In the view of some, the desire for pragmatic usefulness outran the substance of the research findings. For example, legal actors understandably sought to make use of psychologists' findings that post-event information had an impact on witness performance before scientists had determined whether that impact was to replace original memories, to alter memory traces, to lead to attribution of a memory to the wrong source, or to instigate a "memory blend" of old and new information.

New Focus on the Prevention of Error: "System-Variable" Research and Application

While the battles over admissibility of expert testimony ground on in American courts, another of Buckhout's recruits, Gary Wells, was engineering a shift in perspective (Wells 1978; Doyle 2005). Wells admired Loftus and accepted her findings as good science, but he also pointed out limits on their utility.

Because Loftus and her peers were scrupulous scientists, they had isolated and studied a single factor (e.g., the wording of a question, the stress of the event, the presence of post-event information) at a time. Wells noted that these studies yielded statistical results that could tell you what happened eight times out of ten, but could not tell you whether *this* case was one of the eight or one of the two. That second question, clinical and diagnostic (not probabilistic in nature as were the research results), was the challenge the legal system actually faced. Wells also noted that every criminal event incorporates many factors, not just one, and there was no science-based mechanism for combining these factors and assessing their interactions. From Wells' point of view, offering post hoc diagnosis of eyewitness error from the witness stand was the wrong way to mobilize the solid (but inherently probability-based) science that Loftus and a generation of their colleagues were producing.

Wells successfully called for the new orientation that has dominated criminal justice policy discussions about eyewitnesses for the past decade. He noted that some factors Loftus had studied (e.g., lighting, age of witness, stress of event) are not under the criminal justice system's control. He called these "estimator variables." But he also noted that there were other factors (e.g., lineup construction, lineup administration, witness interview technique) that the system's actors *do* have power over. If you understood how these "system variables" could be modernized, you could reduce the rate of error. Wells argued that preventing mistakes by identifying new "best practices" in investigation would be better than trying to catch mistakes from the witness stand after they happened. The task of psychological science in this conception was the prevention of eyewitness errors as evidence was being produced, not the retrospective inspection of eyewitness testimony to see if an error had occurred.

A torrent of research followed, exploring and refining new elements of "system-variable" design (Malpass 1981). In part, this was because the ethical barriers that bedevil research into the witness's time-of-event "estimator" experiences do not generally arise in "system-variable" investigations. Research-based designs for new techniques for interviewing witnesses without contaminating memories (Fisher and Geiselman 1989), for choosing the members for lineups and photographic arrays, and for performing identification procedures began to be developed and to gain momentum. That research coalesced around a photo-array and lineup protocol described by a number of leading researchers in an influential consensus "white paper" of the American Psychology-Law Society.

This document urged the adoption of four reforms to prevailing procedural routine. First, the authors recommended that a "blind" official who did not know the identity of the suspect administer the test. Second, the authors, adopting a line of defense pioneered by Malpass, aimed at reducing the likelihood that a witness might make a "relative judgment" that an innocent suspect "looks most like" the perpetrator and choose the suspect on that basis by urging that the eyewitness be told at every procedure that the real perpetrator "may or may not" be present in the lineup or photographic array (Malpass 1981). Third, the authors agreed that research indicated that the "fillers" in the procedure should be chosen to match the witness's verbal description of the perpetrator rather than the suspect. Finally, the "white paper" urged that the police take a statement of confidence from the witness immediately after any choice in order to minimize the influence of "feedback" which might later boost the witness's confidence artificially.

An alternative attack on the "relative judgment" problem, the "sequential" display of suspect and fillers one at a time was discussed but not included in the four core recommendations of the white paper, although it has since become an important (and controversial) focus of reform efforts.

Controversies over Developing "Systems" Applications

Criminal justice system efforts to harvest the lessons of social science findings in the eyewitness area received powerful new impetus from the release of the first lists of DNA exonerations of wrongfully convicted defendants: lists on which convictions driven by mistaken eyewitnesses predominated (Garrett 2011). Many systems have begun to institute "system-variable" procedural changes based on scientific findings.

These changes have been quite varied, and they have been arrived at by different routes in different systems. In the United Kingdom, for example, where the administration of criminal justice is highly centralized, the Devlin Commission and its successors assessed the problem and suggested changes that were made broadly applicable and supported by legislation (Roberts 2009). Innovations such as a preference for video lineups over traditional live identification parades have been adopted relatively seamlessly.

In the United States, where thousands of individual jurisdictions make independent decisions about eyewitness procedures, advances have not followed a straight line.

The origin of many changes in routine practice can be traced to an influential document produced by the United States Department of Justice's research arm, the National Institute of Justice, in 1999, in the immediate aftermath of the first DNA exonerations. The NIJ convened a mixed "Technical Working Group" composed of leading researchers, experienced police and prosecutors, and defense bar representatives (Technical Working Group 1999). That group debated the existing research in light of law enforcement operational concerns and then published and disseminated *Eyewitness Evidence: A Guide for Law Enforcement*, with the explicit goals of increasing the amount of eyewitness evidence collected (e.g., by employing research-generated "cognitive interview" techniques), enhancing the quality of eyewitness evidence (e.g., by avoiding memory contamination in interviews and decreasing "looks-most-like" pressures by providing pre-lineup instructions), and improving the justice system's ability to evaluate eyewitness evidence (e.g., by requiring recording of an immediate witness confidence statement at the time of any identification).

The *Guide* was not an expression of scientific consensus; it emerged from a long and contentious process during which operational realities, political compromises, and an awareness that US jurisdictions are diverse in their resources and attitudes all played a role. No scientist would have considered the *Guide* to be "cutting edge." For example, the *Guide* did not embrace a requirement that lineups be administered "double blind," and while it did include instructions for conducting the new "sequential" lineups in which the suspect and fillers are displayed individually, it did not announce a preference for that procedure. Nor did the *Guide* have any binding effect on anyone; it was purely advisory.

Nevertheless, the N.I.J. *Guide* seems to have inspired the creation in state and local jurisdictions of similar "all-stakeholders" working groups, task forces, and commissions which have wrestled over and then ultimately adopted versions of the "double-blind sequential" approach to lineup administration. This approach, often associated with Iowa State

University researcher Gary Wells, has at its core double-blind administration, fillers matched to descriptions, pre-lineup ("perpetrator may or not be there") instructions to witnesses, "sequential" presentation of lineup members, and an immediate confidence measurement.

The spreading application of the double-blind sequential approach has not been uncontested.

Even psychologists broadly supportive of system-variable reforms have expressed doubts about the empirical basis for particular elements of the new approach, especially the "sequential" approach to the display of suspects and fillers in a lineup or photographic array (McQuiston-Surrett et al. 2006).

Some feel that the double-blind sequential protocol imposes unacceptable burdens on frontline police investigators, although police in jurisdictions employing the method generally express satisfaction with its operability.

Others, both within and outside the academy, have objected on utilitarian policy grounds. They argue that the double-blind sequential format's successes in protecting the innocent are outweighed by the fact that it also seems to reduce the rate of correct identifications of the guilty (Clark 2012). While results vary on this point somewhat from study to study, it generally appears that the double-blind sequential technique reduces the number of false identifications when the actual perpetrator is not present in the lineup or array but at the price of reducing the number of accurate identifications when the actual perpetrator *is* in the group (Steblay 2011). Although the increase in the rate of "false misses" in the "target present" identifications is considerably smaller than the reduction in the rate of "false hits" when the actual perpetrator is not present, the debate over whether the costs of the traditional system (in which falsely identified suspects result in the guilty going free *and* innocents going to prison) are outweighed by the costs an aggravated rate of "false misses" in the double-blind sequential format is a point of contention in many discussions of reform (Clark 2012).

There have also been perennial challenges focused on the ecological validity of the entire body of eyewitness findings. Most study has occurred in laboratory settings, utilizing student subjects. Within that setting, certain conditions (e.g., the terror experienced by an actual sexual assault victim during a crime) simply cannot be duplicated for ethical reasons. Although, in general, such scant "real-world," archival data as has been compiled and studied (Berman and Davey 2001) conforms in general outline to the laboratory findings, there are exceptions, and the generalizability of eyewitness studies to real-world conditions can be expected to remain an ongoing issue.

Skepticism about the usefulness of laboratory studies in gauging the effectiveness of reformed "system-variable" practices has led to calls for "field tests" of the double-blind sequential technique in actual cases. An initial effort in this direction, designed and conducted by the Chicago Police, led to a substantial detour when it was offered as proof that the in the field the "double-blind sequential" protocol when it was claimed that it produced a higher rate of inaccurate identifications than did traditional "non-blind, simultaneous" methods. Examination of that study, however, showed that it suffered from fatal design flaws that rendered its results largely to interpret and useless (Schacter et al. 2007). A subsequent, more carefully designed field study appears to have vindicated the laboratory findings of the superiority of the double-blind sequential approach (Wells et al. 2011). Even so, it is important to remember that "superiority" here is used to mean "producing fewer mistakes" (Steblay 2011). Some commentators would put a higher value on freeing the innocent, others on identifying the guilty. It is still contested as a utilitarian matter what ratio of imprisoned innocents/freed perpetrators one finds "superior." Although most law enforcement agencies that have confronted the trade-offs required have been satisfied that the double-blind sequential technique strikes the best balance, the issue cannot be regarded as settled. Different agencies and jurisdictions may come to different conclusions. In the United Kingdom, for example, although video identification procedures display one suspect or filler at a time, individually, the witness controls whether the viewings terminate after one round or repeat, converting a procedure into the functional equivalent of a "simultaneous," multiple choice test (Roberts 2009).

All field studies are hampered by the fact that they can assess only the rate of *suspect* identifications, using the identifications or failures to identify suspects (who may or may not be factually guilty) as a proxy measure for perpetrator identifications. Although the most recent field study, conducted by an alliance led by the American Judicature Society, made strenuous efforts to develop an approximation of the "ground truth" of the identity of the real perpetrator who was (or was not) identified in the real-world cases under examination, there is no way to be certain in a field study whether the suspect identified (or missed) was, in fact, the person seen during the commission of the crime.

Future Directions

There is one sense in which the future direction of the psychology and criminal justice relationship will remain unchanged: The participants will continue to attempt to educate their interlocutors in the methods, findings, and practical requirements of their mutually unfamiliar fields.

But there will be change as well as continuity. The future development of eyewitness issues in criminal justice will advance along distinct but not unrelated paths.

Continuing "system-variable" experimentation in the laboratories and in the field will have direct impact on the fact-generating, investigations, stage of the criminal process. It can be expected that the prevention of eyewitness error through the design of better techniques will remain a primary focus and that inquiries into emerging "best practices" sparked by field experience will result in modifications. It can also be expected that adjustments in police operations – for example, the migration of identification tasks such as lineup and photo-array administration to digital media and mobile devices – will require additional scientific inquiry to guarantee that new utilities take an evidence-based approach to the gathering of memory evidence.

At the same time, weaknesses in the adjudicative, inspection, stage of the process when seen in light of the modern scientific findings must also be addressed. The DNA exonerations lists have provoked extensive commentary (Simon 2012) concerning the limited diagnostic capabilities of the traditional criminal adjudicative vehicles. Lawyers and courts will have to supplement their existing toolbox with others devices. In the aftermath of the extensive system-variable reforms to investigative interview, lineup, and photo-array routines, courts are being forced to confront the issue of how to manage the adjudication of cases where there are imperfections in the eyewitness evidence gathering that fall short of best practices but where identifications are still admissible because they do not reach the level of crude suggestion that would require their elimination from evidence altogether. This issue has reinvigorated the debate over the efficacy of expert testimony. It has led to calls to enhanced "expert-substitute" jury instructions. The question of the handling of imperfectly derived but not constitutionally excludable identification evidence has also focused renewed attention on the impact of "estimator variables," since no set of "system-variable procedures" can completely eliminate the dangers created by a confluence of damaging "estimator variables" like lighting, race, or acute stress. Besides, a suboptimal investigative practice (e.g., a failure to give appropriate pre-lineup instructions) matters less when a calm witness viewed a criminal for a long time in an extended nonviolent crime event than it does when the event was traumatic and brief. A more sensitive ability to evaluate these complex situations can be hoped for as the education of courts and lawyers advances.

Finally, there is broad agreement among psychologists – some of whom disagree with each other on many specific issues – that the field needs to develop a more robust interest in basic theory (Clark 2012; Charman and Wells 2007). Psychologists have gained the place that Munsterberg claimed for them: They are now intimately involved in the development and evaluation of evidence. But there is a consensus that to continue to contribute in that hard-won role, psychologists will have to diligently pursue not simply discrete individual applications responding to challenges from practitioners but a broad, theory-driven program of research (Brewer et al. 2007).

Related Entries

- ▶ Behavioral Science Evidence in Criminal Trials
- ▶ Expert Evidence and Criminal Trial Procedure
- ▶ Eyewitness Identification Evidence
- ▶ Eyewitness Research
- ▶ False Memories
- ▶ Interviewing Eyewitnesses
- ▶ Jury Decision Making and Eyewitness Testimony

Recommended Reading and References

Berman BW, Davey SL (2001) Eyewitness identification in actual criminal cases: an archival analysis. Law Hum Behav 25(5):475–491

Borchard E (1932) Convicting the innocent: errors in criminal justice. Yale University Press, New Haven

Brewer N, Weber N, Semmler C (2007) A role for theory in eyewitness identification research. In: Rod CL, Ross DF, Read JD, Toglia MP (eds) The handbook of eyewitness psychology, vol II, Memory for people. Lawrence Erlbaum, Mahwah

Buckhout R (1974) Eyewitness testimony. Sci Am 231:23–29

Charman S, Wells GL (2007) Applied lineup theory. In: Lindsay RCL (ed) The handbook of eyewitness psychology, vol II, Memory for people. Lawrence Earlbaum Associates, Mahwah

Charman S, Wells G, Lindsay RCL (2007) Applied lineup theory. In: Ross DF, Read JD, Toglia MP (eds) The handbook of eyewitness psychology, vol II, Memory for people. Lawrence Erlbaum, Mahwah, pp 219–254

Clark S (2012) Costs and benefits of eyewitness identification reform. Perspect Psychol Sci 7(3):238–259

Cutler B, Penrod S (1995) Mistaken identification: the eyewitness, psychology and the law. Cambridge University Press, New York

Doyle J (2005) True witness: cops, courts, science, and the battle against misidentification. Palgrave-MacMillan, New York

Fisher R, Gielselman RE, Amador M (1989) Field test of the cognitive interview: enhancing the recollection of actual victims and witnesses of crime. J Appl Psychol 4:772–731

Garrett B (2011) Convicting the innocent: where criminal prosecutions go wrong. Harvard University Press, Cambridge, MA

Kassin SM, Ellsworth P, Smith VL (1989) The "general acceptance" of psychological research on eyewitness testimony. Am Psychol 44:1098–1109

Loftus EF (1979) Eyewitness testimony. Harvard University Press, Cambridge, MA

Loftus EF, Palmer J (1974) Reconstruction of automobile destruction: an example of the interaction between language and memory. J Verbal Learn 13:585–591

Malpass R, Devine P (1981) Eyewitness identification: lineup Instructions and the absence of the offender. J Appl Psychol 66:482–496

McCloskey M, Egeth H (1983) Eyewitness Identification: what can a psychologist tell a jury? Am Psychol 38:550–556

McQuiston-Surret D, Malpass R, Tredoux C (2006) Sequential v. simultaneous lineups, a review of method, data, and theory. Psychol Public Policy Law 12(2):137–162

Munsterberg H (1908) On the witness stand. McClure, New York

National Institute of Justice, Technical Working Group on Eyewitness Evidence (1999) Research report: eyewitness evidence: a guide for law enforcement. National Institute of Justice, Alexandria

Roberts A (2009) Eyewitness identification evidence: procedural developments and the ends of adjudicative accuracy. Int Comment Evid 6(2):1–29

Schacter D et al (2007) Policy forum: studying eyewitness identification in the field. Law Hum Behav 32:3–5

Simon D (2012) In doubt. Harvard University Press, Cambridge, MA

Sporer SL, Cutler BL (2003) Identification evidence in Germany. In: van Koppen PJ, Penrod S (eds) Adversarial v. inquisitorial justice: psychological perspectives on criminal justice systems. Plenum, New York

Steblay N (2011) Seventy–two tests of the sequential lineup superiority effect: a meta–analysis and policy discussion, psychology. Public Policy Law 17:99

United States Supreme Court (1977) Manson v. Braithwaite, 388 U.S. 98

Wells G (1978) Applied eyewitness testimony research: system variables and estimator variables. J Pers Soc Psychol 30:1546–1559

Wells G, Steblay N, Dysart J (2011) A test of the simultaneous vs. sequential lineup methods: an initial report of the AJS national eyewitness identification field studies. American Judicature Society, Des Moines

Wigmore J (1909) Professor mueunsterberg and the psychology of testimony: being a report of the case of Cokestone v. Muensterberg. Il Law Rev 3:399–421

Wigmore J (1937) The science of judicial proof. Little Brown, Boston

Eyewitness Identification Evidence

Karen A. Newirth and Barry C. Scheck
The Innocence Project, New York, NY, USA

Overview

The admissibility and treatment of eyewitness identification evidence in criminal trials is governed in almost every US jurisdiction by the legal framework set forth by the United States Supreme Court in a series of cases decided between 1967 and 1977, when little was known about eyewitness memory and perception. This legal framework has remained firmly in place, despite the fact that in the intervening period there has been an explosion of research in the field of eyewitness identification, which now contains the largest and most rigorous body of scientific research of all the law-related social science fields (*State v. Henderson*, 27 A. 3d 872, 916 (NJ 2011)). The scientific research makes it abundantly clear that the current legal framework for evaluating eyewitness identification evidence is broken, a conclusion supported by the fact that eyewitness identification is the most common cause of wrongful convictions, playing a role in nearly 75 % of the 297 convictions overturned by post-conviction DNA testing. *See Eyewitness Misidentification*, Innocence Project, http://www.innocenceproject.org/understand/Eyewitness-Misidentification.php (last visited Jan. 12, 2012); *see also* Brandon L. Garrett, *Convicting The Innocent: Where Criminal Prosecutions Go Wrong* 48 (2011) (finding that 190 of the first 250 DNA-based exonerations in the United States involved eyewitness misidentification). This entry describes the current legal framework for evaluating eyewitness identification evidence and modifications to that framework imposed by lower courts, as well as suggestions for improvement based on scientific research.

Introduction

The Supreme Court has long recognized the significant and unique dangers that the admission of unreliable eyewitness testimony can pose for the criminal justice system. Forty-five years ago, long before the era of exculpatory DNA evidence, the Supreme Court held that the "vagaries of eyewitness identification are well-known; the annals of criminal law are rife with instances of mistaken identification" (*United States v. Wade*, 388 U.S. 218, 228 (1967)). The Court recognized that "a major factor contributing to the high incidence of miscarriage of justice from mistaken identification has been the degree of suggestion inherent in the manner in which the prosecution presents the suspect to witnesses for pretrial identification" (*Id.*). The work of the Innocence Project and affiliated entities that use post-conviction DNA evidence to exonerate the wrongly convicted have confirmed the Supreme Court's worst fears: mistaken eyewitness identifications have played a role in nearly 75 % of the 306 wrongful convictions identified by post-conviction DNA evidence, making eyewitness misidentification the most common contributing cause of wrongful convictions. The experiences of these exonerees, together with robust scientific research concerning eyewitness memory and perception, have identified fundamental flaws not only in traditional forms of pretrial identification procedures but also in the legal framework created by the Supreme Court and applied by courts to evaluate the admissibility of eyewitness identification evidence. Both the exonerations and the scientific research provide a useful roadmap toward an improved legal framework, which some state supreme courts have adopted in part in renovating aspects of the existing legal framework. *See*, e.g*., State v. Henderson*, 27 A.3d 872 (N.J. 2011); *State v. Lawson*, 291 P.3d 673 (Or. 2012).

The Current Legal Framework

A defendant's constitutional rights may be implicated by eyewitness identification procedures in one of two ways. First, and most commonly, due process is implicated when state-orchestrated identification procedures are impermissibly suggestive, resulting in a "very substantial likelihood of an irreparable misidentification" (*Simmons v. United States*, 390 U.S. 377, 384 (1968)). Second, a defendant's Sixth Amendment right to counsel is implicated if he is denied counsel at an identification procedure at which counsel has the right to be present. The Supreme Court has taken two distinct approaches to these different potential constitutional violations.

The Due Process Approach

Prior to 1967, when the Supreme Court heard *Stovall v. Denno*, 388 U.S. 293 (1967), the Court had never considered whether a pretrial identification procedure could implicate a defendant's right to due process. In *Stovall*, the Supreme Court recognized that the "totality of the circumstances" of an identification procedure may be "so unnecessarily suggestive and conducive to irreparable mistaken identification" as to deny a defendant due process of law (*Id.* at 302). *Stovall* involved a showup procedure where the defendant was brought to the victim's hospital room for a confrontation. The defendant was handcuffed to one of the seven police officers who escorted him and was the only black person in the room. The victim was asked by one of the police officers if the defendant "was the man," after which the victim identified the defendant as the person who stabbed him.

The Court set forth a two-part due process test, which directed that courts first ask whether the initial identification procedure was "unnecessarily" or "impermissibly suggestive" and, if so, to consider whether that suggestive procedure was "conducive to irreparable mistaken identification" (*Id.* at 302). While the Supreme Court noted that a showup identification, like the one at issue in *Stovall*, was "inherently suggestive" and "widely condemned," the Court acknowledged that exigent circumstances may make showups necessary. In *Stovall*, the "totality of the circumstances," including the victim's life-threatening injury and her inability to leave the hospital, led the Court to conclude that the showup was not impermissibly or unnecessarily suggestive as "an immediate hospital confrontation was imperative" (*Id.*).

One year later, in 1968, the Court applied the holding of *Stovall* to a case involving pretrial photographic identification procedures. In *Simmons v. United States*, 390 U.S. 377 (1968), the Court considered whether multiple, suggestive pretrial photographic identification procedures was so unduly prejudicial as to make the in-court identification at trial unreliable, thereby fatally tainting the conviction. The Court considered the "totality of the circumstances" and found that the in-court identification was not irreparably tainted by the suggestive pretrial identification procedure. The Court found that the suggestive procedures were necessarily given that a "serious felony had been committed;" the "perpetrators were still at large;" and the "inconclusive clues which law enforcement possessed led to [the defendants]" (*Id.* at 384–85).

The next year, the Court again considered whether a defendant's due process rights had been violated by pretrial identification procedures in *Foster v. California*, 394 U.S. 440 (1969). In that case, following the robbery of a Western Union, the police showed the witness a lineup that consisted of Foster and two others. Foster was at least six inches taller than the other two and was the only lineup member to wear a leather jacket, similar to the one the witness described the robber as having worn. The witness was unable to make a positive identification (stating that he "thought" that Foster was the robber) and asked for an opportunity to speak with him. The police allowed the witness to meet with and speak to Foster, after which the witness was still unsure whether Foster was the robber. A week later, the witness was shown a second lineup which consisted of Foster and four others; Foster was the only person to appear in both lineups. The witness reported that he was certain that Foster was the robber and testified to his out-of-court identification at trial, also making an in-court identification of Foster at trial. The Court reversed the conviction and held that Foster's right to due process had been violated because "the pretrial confrontations clearly were so arranged as to make the resulting identifications virtually inevitable" (*Id.* at 443). This was the first – and remains the only – time that the Supreme Court reversed a conviction on the grounds that a defendant's due process rights were violated as a result of a flawed identification procedure.

The next year, 1970, the Court heard *Coleman v. Alabama*, 399 U.S. 1 (1970), where petitioners appealed their convictions for assault with intent to murder. With respect to the identification procedures, Petitioners argued that the lineup they were subjected to was suggestive in several ways: first, the victim believed that the police had caught the assailants; second, Petitioners and their codefendant were the only lineup members required to utter specific words requested by the victim; and third, one of the Petitioners was the only lineup member wearing a hat like that worn by one of the assailants. The Court, considering the "totality of the circumstances," found that the identifications were "entirely based upon observations at the time of the assault, and not at all induced by the conduct of the lineup" (*Id.* at 5–6).

In *Neil v. Biggers*, 409 U.S. 188 (1972), the Court considered whether a showup identification procedure in a police station was so suggestive that the admission of the identifications violated the defendant's right to due process. The Court reviewed its earlier decisions from *Stovall* to *Coleman* and set forth the test for admissibility of eyewitness identification evidence that was then adopted by the Court in 1977 in *Manson v. Brathwaite*. 432 U.S. 98 (1977). It is this test that remains in place today. The two-part test first asks whether the procedure used was "unnecessarily suggestive," and, if so, whether the identification is nevertheless "reliable." The "reliability factors" set forth by the Court in *Biggers* and adopted in *Manson* are the following: (1) the opportunity of the witness to view the perpetrator, (2) the degree of attention paid by the witness, (3) the accuracy of the witness' prior description of the perpetrator, (4) the witness's level of certainty at the time of the confrontation, and (5) the length of time between the crime and the confrontation (*Id.* at 114). As the Court explained in *Manson*, "[r]eliability is the linchpin in determining the admissibility of identification testimony" (*Id.*).

In May 2011, for the first time since 1977, the United States Supreme Court agreed to consider a case involving a challenge to the admission of eyewitness identification evidence. *Perry v. New Hampshire*, 565 U. S. ____ (2012), involved the accidental (i.e., not police-orchestrated) viewing by a witness of a suspect who was the only civilian and only African American standing at the scene of a crime in the presence of uniformed law enforcement. In *Perry*, the Supreme Court considered a narrow question: does the Due Process Clause require an inquiry into the reliability of an eyewitness identification when the unnecessarily suggestive circumstances that produced the identification were not orchestrated by law enforcement. The Court held that "[w]hen no improper law enforcement activity is involved . . . it suffices to test reliability through the rights and opportunities generally designed for that purpose, notably, the presence of counsel at post-indictment lineups, vigorous cross-examination, protective rules of evidence, and jury instructions on both the fallibility of eyewitness identification and the requirement that guilt be proved beyond a reasonable doubt" (*Id.* at 1). Thus, while the Court once again recognized the "fallibility of eyewitness identifications" (*Id.* at 15) and the role of eyewitness identifications in miscarriages of justice, it concluded that the protections inherent in the trial process are sufficient to protect defendants when there has been no law enforcement involvement in the identification procedure, and a hearing that could result in suppression of eyewitness evidence was not constitutionally required. The Court explained its unwillingness to require a Due Process analysis where suggestion did not emanate from law enforcement as rooted in the Court's recognition that "[t]he Constitution . . . protects a defendant against a conviction based on evidence of questionable reliability, not by prohibiting introduction of the evidence, but by affording the defendant means to persuade the jury that the evidence should be discounted as unworthy of credit" (*Id.* at 6). The Court elaborated, "the jury, not the judge, traditionally determines the reliability of evidence" (*Id.* at 15).

The dissent, authored by Justice Sotomayor, disagreed that the traditional rights and opportunities available to defendants in the course of a trial are sufficient to protect against the possibility of "miscarriage[s] of justice from mistaken identification" (*Id.* at 1) (Sotomayor, J., dissenting) (*quoting Wade*, 388 U.S. at 228). Justice Sotomayor – relying on scientific research and the cases of the wrongfully convicted – found that the serious risk presented by suggestive circumstances exists no matter the source of the suggestion, and therefore, the protections of the Due Process Clause should apply whenever an identification is the product of suggestive circumstances:

> This Court has long recognized that eyewitness identifications' unique confluence of features – their unreliability, susceptibility to suggestion, powerful impact on the jury, and resistance to the ordinary tests of the adversarial process – can undermine the fairness of a trial

(*Id.* at 2) (Sotomayor, J., dissenting).

The dissent argued that the majority opinion represents a significant departure from the Court's jurisprudence, which had until *Perry* set forth "a clear rule: The admission at trial of out-of-court eyewitness identifications derived from impermissibly suggestive circumstances that pose a very substantial likelihood of misidentification violates due process" (*Id.* at 1) (Sotomayor, J., dissenting). In contrast, the majority opinion described the Court's jurisprudence as having a narrower procedural focus that permitted pretrial screening of identification evidence only in situations involving identification procedures arranged by law enforcement (*Id.* at 2). ("We have not extended pretrial screening for reliability to cases in which the suggestive circumstances were not arranged by law enforcement officers.") *See also Id.* at 2, n.1. ("Neither Perry nor the dissent, however, points to a single case in which we have required pretrial screening absent a police arranged identification procedure. Understandably so, for there are no such cases.").

The dissent further identified *Perry* as presenting a break from the Court's jurisprudence in that it "recasts the driving force of our decisions as an interest in police deterrence, rather than reliability" (*Id.* at 2) (Sotomayor, J., dissenting). Indeed, for the first time in any decision concerning the admissibility of eyewitness identification evidence, the Court identified police deterrence as "a primary aim" (*Id.* at 11). *Manson* cast deterrence as one of "several interests to be considered," together with reliability and the administration of justice (*Manson*, 432 U.S. at 112. *Accord Perry* at 10–11) (Sotomayor, J. dissenting).

Scientific and legal attacks on the validity of the *Manson* "balancing" test as a method for assessing the reliability of eyewitness evidence were not before the Court in *Perry*, the "all-or-nothing" aspect of the Manson legal architecture that provides no guidance to juries when suggestive procedures ("system variables") are used that increase the risk of error or "estimator variables" are involved that can also undermine the reliability of identifications. Arguments along these lines (arguments that were recently adopted in influential decisions of the New Jersey Supreme Court and the Oregon Supreme Court in *State v. Henderson* and *State v. Lawson*, respectively) were presented by certain *amici* including the Innocence Network and the American Psychological Association, but the Court did not choose to reach out to address them, not even in *dicta*. Since these arguments, based on decades of scientific findings, go more toward a "renovation" or improvement of *Manson* with respect to the way trial courts evaluate the reliability of eyewitness evidence and instruct juries when suggestive procedures by state actors occur, and not a new, lower standard for the exclusion of evidence, it should not be assumed that the *Perry* decision signals a hostility to the approach adopted by the New Jersey Supreme Court in *Henderson*. The *Perry* decision probably reflects, more than anything else, a predisposition against any expansion of the exclusionary rule and confidence in the ability of trial courts to use traditional procedures (jury instructions, motions in limine, or expert testimony) to assist juries in the evaluation of questionable eyewitness evidence. Indeed, the scientifically based framework adopted in *Henderson* is directed toward strengthening the traditional trial-based protections identified in *Perry* to ensure they are sufficiently flexible and robust to prevent a due process violation.

The Sixth Amendment Approach

On the same day in 1967 that the Court issued *Stovall*, the Court issued two other cases involving eyewitness identification procedures: *United States v. Wade*, 388 U.S. 218 (1967), and *Gilbert v. California*, 388 U.S. 263 (1967), which addressed whether violations of the Sixth Amendment right to counsel occurred during post-indictment identification procedures.

In *Wade*, the Court considered whether the failure to alert counsel that the client would be placed in a lineup post-indictment violated the client's right to counsel. Concluding that it did, the Court held that the Sixth Amendment guarantees an accused the right to counsel not only at his trial but at any "critical confrontation by the prosecution at pretrial proceedings where the results might well determine his fate and where the absence of counsel might derogate from his right to a fair trial" (388 U.S. at 223). The Court went on to find that a post-indictment lineup was just such a critical stage (*Id.* at 227). The remedy for this violation was to exclude the out-of-court identification from being admitted in evidence at trial, although the state would have "the opportunity to establish by clear and convincing evidence that the in-court identifications were based upon observations of the suspect other than the lineup identification" (*Id.* at 240). The Court identified factors to determine whether there exists an independent source for the in-court identification: the prior opportunity to observe the alleged criminal act, any discrepancy between any pre-lineup description and the defendant's actual appearance, any prior identifications or failures to identify the defendant, and the time between the crime and the confrontation (*Id.* at 241).

In *Gilbert v. California*, 388 U.S. 263 (1967), the Court applied the *Wade* test to determine whether Gilbert's rights were violated by a series of suggestive, post-indictment identification procedures conducted outside of the presence of counsel. The Court, concluding that Gilbert's right to counsel had been violated, announced a per se exclusionary rule for out-of-court identifications that are the product of procedures where the defendant had a right to counsel but that right was violated (*Id.* at 271). ("The admission of the in-court identifications without first determining that they were not tainted by the illegal lineup but were of independent origin was constitutional error.").

The Manson Balancing Test Is Flawed

Manson arose in an era when the exclusionary rule was being used by the Supreme Court as a remedy to deter police from violating citizens' constitutional rights. But, as the *Manson* court recognized, unnecessarily suggestive identification procedures themselves do not violate a suspect's constitutional rights because "[u]nlike a warrantless search, a suggestive pre-indictment identification procedure does not in itself intrude upon a constitutionally protected interest" (*Id.* at 113 n.13). The constitutionally protected interest at stake in eyewitness identification cases is the due process right to a fair trial.

Therefore, the *Manson* Court focused on the trustworthiness of the identification evidence and declared reliability to be the linchpin for its admissibility (*Id.* at 144). It feared that a per se rule suppressing unnecessarily suggestive out-of-court identification procedures "goes too far since its application automatically and peremptorily, and without consideration of alleviating factors, keeps evidence from the jury that is reliable and relevant" (*Id.* at 112). While conceding that a per se suppression rule has a "significant deterrent effect" in preventing the use of unnecessarily suggestive procedures, the Court still worried that the "rigidity" and "inflexibility" of a per se rule "may result, on occasion, in the guilty going free" and would make error by trial courts "more likely" (*Id.*).

The *Manson* Court was by no means refusing to acknowledge the "awful risks of misidentification" (*Id.* at 110), or the dangers posed by unnecessarily suggestive identification procedures, but believed that juries would understand that "[s]uggestive procedures often will vitiate the weight of evidence" and juries could be counted on to appropriately "discount" it (*Id.* at 112 n. 12). By developing what it believed was a flexible "totality of the circumstances" approach that stressed "reliability" and forced trial courts to make detailed, pretrial assessments of evidence, the *Manson* Court envisioned that its two-part balancing test would improve the "administration of justice" and produce more accurate verdicts (*Id.* at 112–113).

It is now clear that the *Manson* test, as currently configured, does not meet the objectives the Court set for it. Ironically, *Manson* was written the very year, 1977, that eyewitness identification research started to advance toward its current status as the "gold standard" for the reliable application of social science to the law (*State v. Henderson*, 27 A. 3d 872, 916 (NJ 2011)). Three decades of empirical findings demonstrate that the *Manson* framework suffers from five serious flaws that increase the chance of wrongful convictions based on eyewitness misidentifications: its balancing test is skewed by a scientific confound; it focuses solely on police misconduct; it limits trial courts to an inflexible, all-or-nothing suppression remedy; it does virtually nothing to deter unnecessarily suggestive identification procedures; and it fails to provide much needed "context" and guidance for jurors on how to evaluate eyewitness identification evidence.

The "Balancing" Test Is Confounded

Under *Manson*, courts are first supposed to balance the corrupting effects of unduly suggestive identification procedures against "reliability factors" and then decide whether to suppress in-court and out-of-court identification evidence if they find a "very substantial likelihood of an irreparable misidentification" (*Manson*, 432 U.S. at 122). The problem, of course, with such "balancing" is the undisputed scientific finding that both post-identification feedback and the use of unduly suggestive identification procedures tend to artificially inflate post-identification self-reports from witnesses about key reliability factors – opportunity to observe, the degree of attention paid, certainty, and description (Wells and Quinlivan 2009). The consequences of this confound are profound. It artificially inflates the apparent reliability of the eyewitness identification both for judges deciding admissibility and for jurors trying to evaluate the real weight of the evidence. This, in turn, brings about an unintended but deeply disturbing result: the improper use of a suggestive procedure tends to make it *more* likely that courts and juries will find the identification reliable. The *Manson* Court, of course, assumed exactly the opposite was true, that juries would realize that suggestive procedures "vitiate the weight of the [identification] evidence" and would, accordingly, "discount" it (432 U.S. at 112 n.12).

The confound also provides a perverse incentive to law enforcement who believe a suspect is guilty and hope an eyewitness can provide evidence to support their case – the more suggestive an identification procedure, the more likely an identification will be made, the more confirming feedback the witness will receive, and the more likely the witness will be certain about the identification itself, the opportunity to view, and the degree of attention paid. While the *Manson* Court recognized its approach would not "significantly" deter the use of suggestive police procedures, it still envisioned, misguided as its expectations were, that its two-part test would curtail police suggestion to some extent, and it certainly did not intend to create an impetus for law enforcement to conduct biased lineups (*Manson*, 432 U.S. at 112). ("Although the per se approach has the more significant deterrent effect, the totality approach also has an influence on police behavior. The police will guard against unnecessarily suggestive procedures under the totality rule, as well as the per se one, for fear that their actions will lead to the exclusion of identifications as unreliable.") Since the *Perry* Court labeled deterrence of police misconduct the "primary aim" of the Due Process analysis, it follows, whenever it reaches the issue, it should seriously consider "renovating" the Manson "balancing" test to prevent the confound from actually rewarding police misconduct.

Focus Is Exclusively on Police Misconduct

The seminal identification cases of the late 1960s and 1970s arose in the context of a contentious Supreme Court jurisprudence focused on the utility of the exclusionary rule as a remedy against misconduct by state actors. However, given our contemporary scientific knowledge that eyewitness memory is best understood as trace evidence susceptible to contamination from a wide spectrum of sources, it would be an error to scrutinize identification evidence only through the prism of police misconduct. Unlike the law, science does not differentiate between intentional and unintentional suggestion, since the good faith of a police officer is unrelated to the contaminating effects of police conduct on memory. Moreover, the linchpin of admissibility should be reliability, not whether the identification testimony was elicited by state or non-state actors.

To be sure, suggestive police procedures can taint the memory of an eyewitness and render any subsequent identification unreliable, but equally pernicious contamination of eyewitness memory is often brought about by sources unconnected to law enforcement – family members, friends, other witnesses to the same event, media reports, or simply the passage of time. Therefore, focusing solely on state action will surely miss non-state factors that contaminate eyewitness memory and fatally undermine the reliability of the identification evidence. State court decisions holding that non-state suggestion can contaminate eyewitness memory, as either a state constitutional matter or pursuant to their supervisory powers, are better suited to build on the growing, robust body of scientific research from experimental psychologists and neuroscientists about the malleability of memory than the more restrictive approach adopted by the Supreme Court in *Perry*. Indeed, in some cases, estimator variables alone could demonstrate that identification evidence was so weak it should be suppressed, or at least the jury should be instructed to treat it with great caution and distrust. Under the *Manson* framework, a defendant cannot claim that estimator variables so undermine the reliability of the eyewitness identification evidence as to render the admission of that evidence a violation of due process. Instead, the defendant's recourse will likely be limited to remedies available under the rules of evidence and procedure.

The All-or-Nothing Test Forgoes Helpful Intermediate Remedies, Such as *In Limine* Rulings and "Contextual" Jury Instructions, Based on Findings Made at Pretrial Judicial Assessments of Reliability

When courts apply *Manson*, their purpose is usually limited to answering one question: to suppress or not to suppress the identification evidence. Once courts decide that issue, they conceive of their mission as complete. The problem is that since courts rarely suppress identification evidence, they rarely see any purpose in identifying suggestive procedures that increase the risk of error or making findings about other factors relating to the event or the witness which tend to decrease the reliability of the identification.

Intermediate remedies short of suppression, including *in limine* rulings (such as limiting testimony based on artificially inflated self-reports about certainty and description) and "contextual" instructions (such as telling the jury that failure to comply with certain best practices can increase the risk of misidentification), would result in comprehensive assessments of reliability, which would require identifying and understanding key estimator and system variables. In this way, intermediate remedies would generate a more substantive judicial screening process, more reliable evidence, and more accurate verdicts.

Failure to Provide Jurors with "Context" and Guidance to Correct Misconceptions About Eyewitness Memory

After refocusing the analysis of eyewitness identification evidence on reliability and ensuring that juries would not be deprived of critical, if "flawed," evidence ("evidence with some element of untrustworthiness is customary grist for the jury mill"), the *Manson* Court was "content to rely upon the good sense and judgment of American juries" (432 U.S. at 116). The Court felt that "[j]uries are not so susceptible that they cannot measure intelligently the weight of identification testimony that has some questionable feature" (*Id.*).

Unfortunately, long-standing research studies show that jurors have great difficulty distinguishing between accurate and inaccurate eyewitnesses. *See* Tanja Rapus Benton et al., *Has Eyewitness Testimony Research Penetrated the American Legal System? A Synthesis of Case History, Juror Knowledge, and Expert Testimony*, in The Handbook of Eyewitness Psychology: Memory for People 453 (R.C.L. Lindsay et al. eds. 2007). Mistaken eyewitnesses are telling the truth as they believe it, and thus the cognitive faculties jurors usually deploy in making credibility judgments about lying witnesses do not work in this context. This also explains why cross-examination – the supposed great engine for uncovering truth – often sputters in the face of an honest but mistaken eyewitness, is insufficient, on its own, to guard against wrongful convictions based on mistaken identifications (as both the DNA exonerations and empirical study show), and serves as an inadequate substitute for expert testimony or jury instructions. *See* Jules Epstein, *The Great Engine That Couldn't: Science, Mistaken Identifications and the Limits of Cross-Examination*, 36 STETSON L. REV. 727 (2007); Wells (1995), *Eyewitness Identification Procedures: Recommendations for Lineups and Photospreads*, 22 LAW & HUM. BEHAV. 603, 609. ("Cross-examination, a marvelous tool for helping jurors discriminate between witnesses who are intentionally deceptive and those who are truthful, is largely useless for detecting witnesses who are trying to be truthful but are genuinely mistaken.") Research shows that jurors have fundamental misconceptions about eyewitness memory. *See*, e.g., Melissa Boyce et al., *Belief of Eyewitness Identification Evidence*, in THE HANDBOOK OF EYEWITNESS PSYCHOLOGY: MEMORY FOR PEOPLE 501 (R.C.L. Lindsay et al. eds. 2007); Elizabeth F. Loftus et al., EYEWITNESS TESTIMONY: CIVIL AND CRIMINAL (4th ed. 2007); Brian L. Cutler et al., *Juror Sensitivity to Eyewitness Identification Evidence*, 14 LAW & HUM. BEHAV. 185–191(1990); Richard S. Schmechel et al., BEYOND THE KEN? TESTING JURORS' UNDERSTANDING OF EYEWITNESS RELIABILITY EVIDENCE (2006); Tanja Rapus Benton et al., *Eyewitness Memory is Still Not Common Sense: Comparing Jurors, Judges and Law Enforcement to Eyewitness Experts*, 20 APPLIED COGNITIVE PSYCHOL. 115 (2006). Jurors tend to believe that memory works like a videotape; generally misunderstand the effect of confirming feedback on the self-reported factors in *Manson*; do not understand the effects of biased witness warnings; are not inherently sensitive to estimator variables such as disguises, weapon focus, violence during the event, retention intervals between the event and the identification procedure, foil bias, and cross-race identifications; tend to rely heavily on eyewitness factors that are not good indicators of accuracy (particularly the witness's confidence in her identification); overestimate eyewitness identification accuracy rates; and are not familiar with the trace evidence analogy. It is, therefore, critically important to correct these scientifically incorrect notions and provide jurors with "context" or guidance about eyewitness testimony that is firmly grounded in sound science.

Failure to Provide Significant Deterrence of Suggestive Identification Procedures

The *Manson* Court lamented the "inflexibility" and "rigidity" of the exclusionary rule and acknowledged that deterrence of improper police practices would be significantly diminished by not suppressing out-of-court identification procedures (432 U.S. at 114). The Supreme Court has now taken a different view of the importance of police deterrence, a view that supports the adoption of intermediate remedies that enforce best practices while informing juries accurately about the risks created by violations of best practices. In this way, courts can achieve a greater measure of deterrence while also providing much needed guidance to the jury.

State Law Challenges to the *Manson* Balancing Test

While the United States Supreme Court has not yet demonstrated an interest in reevaluating the existing legal framework for evaluating eyewitness identification evidence, state courts – which have firsthand experience with both the range of eyewitness identification evidence offered at trials and the human cost of "miscarriages of justice" resulting from eyewitness misidentifications – have indicated a willingness to reevaluate the existing legal framework.

The most comprehensive example of this is the New Jersey Supreme Court's decision in *State v. Henderson*, discussed below. While a full survey of the changing national landscape is beyond the scope of this entry, it is worth mentioning a few examples.

In September 2011, recognizing the role of eyewitness misidentification in wrongful convictions, the Massachusetts Supreme Judicial Court directed that a study group be convened to address all aspects of the handling of eyewitness identification evidence, from law enforcement practices to admissibility standards and the treatment of the evidence by courts (*Commonwealth v. Walker*, 460 Mass. 590, – n.16 (2011)). The court explained, "Our convening of a study committee on eyewitness identification reflects our willingness to revisit our jurisprudence regarding such evidence . . . and to consider, among other alternatives, the approach established in New Jersey . . ." Indeed, the study group ultimately created is comprised of attorneys, judges, and members of law enforcement and has been divided into subcommittees focusing on hearings, jury instructions, and police procedures.

The Oregon Supreme Court is considering a case, *State v. Lawson*, involving eyewitness identification. The court will consider several questions including whether any error in admitting eyewitness testimony in suggestive identification cases can be cured at trial by cross-examination, expert testimony, closing arguments, and jury instructions.

Other state courts (both trial and appellate) have stopped short of reconsidering the entire legal framework but have instead eliminated those parts of the *Manson* test not supported by science (such as the use of certainty as a reliability factor) and strengthened protections available to defendants against whom eyewitness identification evidence is offered. Typically, these protections have been in the form of expert testimony and jury instructions. The Texas Court of Criminal Appeals recently found that the trial court abused its discretion in excluding an eyewitness identification expert, which was "reliable, relevant evidence that would 'assist the trier of fact' by increasing the jurors' awareness of biasing factors in eyewitness identification." *Tillman v. State*, No. PD-0727-10, slip. op. at 35 (October 5, 2011) (internal citations omitted). In the past several years, the New York Court of Appeals has issued a number of rulings endorsing scientific research in the area and reversing lower courts' refusal to admit expert testimony. *See*, e.g.*, People v. LeGrand*, 8 NY3d 449, 452 (2007); *People v. Abney*, 13 NY3d 251, 267 (2009); and *People v. Santiago* 17 NY3d ___, 2011 NY Slip Op 07303 (2011).

Finally, courts that either lack the power to reconsider the entire legal framework or have yet to be presented with a case or controversy that raises the issue have indicated an interest in revisiting the existing legal framework. In a concurring opinion in *State v. Ferguson*, 804 N.W.2d 586 (2011), Minnesota Justice Anderson acknowledged the extensive science relating to eyewitness identification evidence, the problem of eyewitness misidentification, and discussed at length the New Jersey Supreme Court's decision in *Henderson*. Justice Anderson concluded:

> I hope that the district court will carefully consider whether the defects in the photo lineup procedure used here and the recent developments in social science require admission of eyewitness identification expert testimony and/or a cautionary jury instruction. Moreover, the court should look closely at New Jersey's safeguards and determine if those safeguards are appropriate here. Specifically, the court should consider the reliability of the eyewitness identification in light of the recent New Jersey Supreme Court decision, in addition to the factors our court has set out in *Miles* and *Helterbridle*. If the expert eyewitness testimony is not "otherwise appropriate," the court should consider alternative approaches to educating jurors on the variables that "can lead to misidentifications." As our courts and jurors grow to understand the science of eyewitness identification, we can better meet the "'twofold aim . . . that guilt shall not escape or innocence suffer.'"
>
> (*Id.* at 609–610).

State of New Jersey v. Henderson

In 2011, the New Jersey Supreme Court became the first court in the nation to reject the flawed *Manson* balancing test, finding that in “the 34 years since the United States Supreme Court announced a test for the admission of eyewitness identification evidence, which New Jersey adopted soon after, a vast body of scientific research about human memory has emerged. That body of work casts doubt on some commonly held views relating to memory. It also calls into question the vitality of the current legal framework for analyzing the reliability of eyewitness identifications.” *State v. Henderson*, 27 A.3d 872, 877 (NJ 2011).

The New Jersey Supreme Court’s decision was the direct result of the 1999 issuance of mandatory guidelines for conducting eyewitness identification procedures by New Jersey’s Attorney General. Office of the Attorney Gen., N.J. Dep’t of Law and Pub. Safety, Attorney General Guidelines for Preparing and Conducting Photo and Live Lineup Identification Procedures 1 (2001) (“Attorney General Guidelines” or “Guidelines”). Under New Jersey law, the Guidelines – which provided for the blind administration of identification procedures, the provision of instructions to witnesses, proper lineup construction (including minimum number of fillers, proper filler selection, one suspect per lineup), sequential presentation when possible, avoidance of feedback to the witness, and recording of the procedure, to include a witness confidence statement – were binding on all law enforcement agencies in New Jersey.

Henderson came to the New Jersey Supreme Court on appeal following a conviction for various crimes, including reckless manslaughter, aggravated assault, and various weapons charges, which arose out of a 2003 murder in a Camden, New Jersey, apartment. The primary evidence against Henderson at trial consisted of the statement of a single eyewitness and an investigating officer’s testimony about Henderson’s postarrest statement (*Id.* at 879–880). During the course of the investigation, the eyewitness – who, at the time of the crime, had been held at gunpoint in a dark, narrow hallway by a stranger – was presented with an eight-person photographic lineup by a blind administrator, consistent with the Attorney General’s Guidelines (*Id.* at 880). The witness was provided with instructions and viewed the lineup members sequentially, in keeping with the Attorney General’s Guidelines. The witness quickly eliminated five of the photographs, then eliminated a sixth, after which he said he “wasn’t 100 % sure of the final two pictures” (*Id*). At that point, the independent administrator left the room and the two investigating offers entered and conferred with the witness, thereby violating the Attorney General’s Guidelines that the identification procedure administrator be blind. Thereafter, the witness selected the police suspect’s photograph – that of Henderson.

When it reviewed Henderson’s appeal in 2009, the New Jersey Supreme Court declared that the trial record was inadequate to “test the current validity of [New Jersey] state law standards on the admissibility of eyewitness identification” and directed that a plenary hearing be held “[t]o consider and decide whether the assumptions and other factors reflected in the two-part *Manson/Madison* test, as well as the five factors outlined in those cases to determine reliability, remain valid and appropriate in light of recent scientific and other evidence” (*Henderson*, No. A-8-08, 2009 WL 510409, at *1–2).

The Court directed the parties, together with *Amici* Innocence Project and Association of Criminal Defense Lawyers of New Jersey, to participate in remand proceedings, which were presided over by Special Master Geoffrey Gaulkin, a retired New Jersey state appellate judge appointed by the New Jersey Supreme Court to handle the matter. Judge Gaulkin conducted the proceedings “more as a seminar than an adversarial litigation.” The parties submitted, and Judge Gaulkin considered, extensive scientific materials including more than 200 published scientific studies, articles, and books. Judge Gaulkin presided over 10 days of evidentiary hearings, at which seven expert witnesses – leading scientists in the field of eyewitness identification study – testified, and he received detailed proposed findings of fact and conclusions of law and heard oral argument. On June 18, 2010, based on his consideration of all of the information presented by the parties, Judge Gaulkin issued his report (the "Special Master's Report").

The Special Master's Report endorsed the remedy set forth by the Innocence Project in its proposed legal findings, "The Renovation of *Manson*: A Dynamic New Legal Architecture For Assessing and Regulating Eyewitness Evidence," as "wide-ranging, multifaced and highly detailed," and proposed that the current legal framework be modernized:

> to maintain the *Manson*/*Madison* principle that reliability is the linchpin of the inquiry, to expand that inquiry to include all the variables unaddressed by *Manson*/*Madison* and to assure that judges and jurors are informed of and use the scientific findings that bear on reliability

The Special Master's Report, at 84–86 (Citations Omitted and Emphasis Added)

The Special Master's proposed new design had two core elements: first, that courts treat eyewitness identification evidence as trace and scientific evidence, by placing the initial burden on the prosecution to demonstrate the reliability of the evidence, and second, to take "all available steps" to ensure that judges and juries are informed of and guided by scientific findings (*Id*).

In issuing its final decision in *Henderson*, 27 A.3d 872 (2011), the New Jersey Supreme Court adopted many of the recommendations set forth in the Special Master's Report. The New Jersey Supreme Court endorsed the scientific research considered by the Special Master, describing it as the "gold standard in terms of the applicability of social science research to the law" (*Id.* at 916).

Based on its review of the scientific research and the role of eyewitness misidentifications in wrongful convictions, the New Jersey Supreme Court took unprecedented action and became the first court in the nation to reject the *Manson* balancing test in favor of a totality of the circumstances test that would allow for the consideration of all relevant evidence – whether relating to system or estimator variables – when making a determination about the reliability of eyewitness identification evidence. (New York, Massachusetts, and Wisconsin follow a per se rule requiring suppression of out-of-court identifications that were the product of unnecessarily suggestive identification procedures. However, those courts then engage in a second step analysis to determine whether the in-court identification has a basis independent of the suggestive procedure for the identification. The independent basis test in these jurisdictions essentially mirrors the reliability factors set forth in *Manson*. Thus, it cannot be said that these jurisdictions have abandoned the *Manson* balancing test. *See*, e.g., *People v. Adams*, 423 N.E.2d 379, 384 (NY 1981); *Commonwealth v. Martinez*, 67 Mass. App.Ct. 788 (2006); *State v. Dubose*, 699 N.W.2d 582, 596 (Wise and Safer 2003) The revised legal framework set forth in *Henderson* has two principal features: pretrial hearings that are granted whenever a defendant can show "some evidence of suggestiveness that could lead to mistaken identification" *(Id.* at 920) and which will allow for the presentation of all system and estimator variables once the defendant has met this burden, and enhanced jury instructions that will seek to educate jurors about all aspects of eyewitness identification and memory, and may be given to the jury at the close of evidence as well as during the trial, at the time the witness testifies (*Id.* at 925–28). Additionally, the *Henderson* decision extended the court's prior decision in *State v. Cromedy*, 727 A.2d 457 (1999), in which the New Jersey Supreme Court held that jury instructions on the reliability of cross-racial identifications are necessary when "identification is a critical issue in the case" and there is no independent evidence corroborating the identification to require jury instructions on the reliability of cross-racial identifications "whenever cross-racial identification is in issue at trial." The *Henderson* court extended its ruling in *Cromedy* based on additional scientific research that emerged over the 12 years since the court had first considered the issue as well as the "more complete record about eyewitness identification in general" (*Id.* at 926). In considering remedies available at trial, *Henderson* also allows for the possibility of expert testimony at trial, provided it is otherwise admissible under New Jersey law (*Id.* at 878).

Finally, the *Henderson* opinion endorsed many of the police practices already required by the Attorney General's Guidelines and, through the court's supervisory powers, required additional police practices, including requiring that law enforcement inquire whether witnesses have had any communications with third parties concerning their observations and, if so, to take a written statement of the substance of those communications which must be disclosed to defense counsel (*Id.* at 909).

The *Henderson* decision has been hailed by the legal and scientific communities as a landmark ruling for its recognition that the *Manson* balancing test fails to achieve its objective and its attempt to offer a new approach that brings to bear the vast body of scientific research on the problem of assessing the reliability of eyewitness misidentification. Some have expressed concern with the court's reliance on enhanced jury instructions, as the scientific literature is sparse and somewhat mixed on the efficacy of jury instructions in educating jurors about complex aspects of memory and perception. The ultimate test of the success of the *Henderson* model will, of course, be seen in courtrooms throughout New Jersey in years to come.

Conclusion

More than 30 years of scientific research and more than 200 wrongful convictions confirmed by DNA make clear that the test for admissibility of eyewitness identification evidence in criminal trials cannot achieve its stated goal of ensuring reliability. Courts across the country, drawing lessons from scientific research and the cases of the wrongfully convicted, have begun to question the existing legal framework. Some have strengthened protections through expert testimony and comprehensive jury instructions. The New Jersey Supreme Court has rejected the *Manson* balancing test in its entirety and developed a new legal framework that holds enormous promise for improving both the collection of eyewitness identification evidence and the evaluation by courts and juries of the reliability of that evidence.

Acknowledgment The authors acknowledge and thank Ezekiel Edwards for his invaluable work in briefing *amicus* the Innocence Project's positions *State v. Henderson*, 27 A. 3d 872 (N.J. 2011), excerpts of which are included herein.

Recommended Reading and References

Bradfield, Wells (2000) The perceived validity of eyewitness identification testimony: a test of the five biggers criteria. Law Hum Behav 24:581

Cutler BL et al (1990) Nonadversarial methods for improving juror sensitivity to eyewitness evidence. J Appl Soc Psychol 20:1197, 1198–1200, 1202-06

Devenport et al (2009) Effectiveness of traditional safeguards against erroneous conviction arising from mistaken eyewitness identification. In Cutler BL (eds), Expert testimony on the psychology of eyewitness identification, 51:62

Epstein J (2007) The great engine that couldn't: science, mistaken identifications, and the limits of cross-examination. Stetson Law Rev 36:727

Epstein Irreparable misidentifications and reliability: reassessing the threshold for admissibility of eyewitness identification. Villanova Law Rev. Forthcoming

Monahan J, Walker L (2011) Twenty-five years of social science in law. Law Hum Behav 35:72

O'Toole and Shay (2006) Manson v. Brathwaite revisited: towards a new rule of decision for due process challenges to eyewitness identification procedures. Valparaiso University Law Rev 109

Penrod & Cutler (1999) Preventing mistaken convictions in eyewitness identification trials: the case against traditional safeguards. In Roesch R (eds), Psychology and law: the state of the discipline. 89

Wells GL (1995) Scientific study of witness memory: implications for public and legal policy. Psychol Pub Policy L 1:726, 727

Wells (2006) Eyewitness identification: systemic reforms. Wis L Rev 615

Wells GL, Murray DM (1983) What can psychology say about the neil v. biggers criteria for judging eyewitness accuracy? J Applied Psychol 68:347

Wells GL, Quinlivan DS (2009) Suggestive Eyewitness eyewitness id procedures and the supreme court's reliability test in light of eyewitness science – 30 years later. Law Hum Behav 33

Wise RA, Safer MA (2003) A survey of judges' knowledge and beliefs about eyewitness testimony. Court Rev.

Eyewitness Research

Jennifer Dysart[1] and Victoria Z. Lawson[2]
[1]Department of Psychology, John Jay College of Criminal Justice, New York, NY, USA
[2]Department of Psychology, John Jay College of Criminal Justice and The Graduate Center, The City University of New York, New York, NY, USA

Synonyms

Double-blind administration; Pre-lineup instructions; Sequential lineups

Overview

The following scene plays out in courtrooms around the world: from the witness box, an eyewitness is asked to look around the courtroom and indicate whether they see the person who committed the crime. The witness says she does, then outstretches her arm and points to the defendant. She is unequivocal, "That's the man. I'll never forget his face." Although the witness in this scene is absolutely certain that her identification of the defendant is accurate, a critical question remains: Does it also mean that she was right? Eyewitness researchers have studied this exact question for decades and have now identified many factors that influence both a witness' accuracy and confidence in their identification of the suspect. This entry will present the varying types of research methodologies that are conducted in the field of eyewitness identification, followed by a review of law enforcement procedures, known as system variables, that are utilized to secure the identification of police suspects and that have a direct impact on eyewitness accuracy and testimony: sequential lineup presentation, double-blind lineup administration, pre-lineup instructions, post-identification feedback, showups, the presentation of a suspect in multiple procedures, and co-witness contamination. Each identification procedure carries with it potential consequences (positive and negative) for an investigation and these will be discussed in turn.

Introduction to Eyewitness Research

In 1985, Kirk Bloodsworth was identified by five separate eyewitnesses in the murder investigation of a 9-year-old girl. Mr. Bloodsworth, a veteran with no prior criminal record, was swiftly convicted at trial and sentenced to death. After serving 8 years in prison, DNA testing on the forensic evidence in the case revealed that Mr. Bloodsworth was in fact innocent; all five eyewitnesses had been wrong. Kirk Bloodsworth's case was the first of 18 death penalty DNA exonerations in the United States, and it affirmed the harsh truth that eyewitness evidence is not infallible. In fact, an analysis of the more than 300 DNA exonerations in the United States reveals that mistaken eyewitness testimony played a role approximately 75 % of the time, making it the leading contributing factor of wrongful convictions. And like Kirk Bloodsworth's story where five eyewitnesses erroneously testified that he was the perpetrator, 35 % of the eyewitness exoneration cases had more than one eyewitness mistakenly identify the defendant at trial.

Although DNA exonerations show incontrovertible evidence of eyewitness error, they are only a small fraction of all identification cases, as it is rare to have DNA evidence in a typical eyewitness case. Therefore, eyewitness researchers turn to other sources of information, including archival and field studies, to help determine if eyewitness errors are common or rare occurrences in actual criminal investigations. In archival studies, police files from closed eyewitness cases are examined and the witnesses' identification decision from viewing a live or photo lineup are recorded. Archival research reveals a consistent and alarming pattern: approximately 20 % of all witnesses who view a lineup make a mistake and choose a lineup filler – one of the stand-ins who is known to be innocent of the crime. This statistic has also been replicated in field studies, where researchers gather data prospectively in ongoing criminal investigations.

Although the general public may be surprised by the high rate of identification errors in real cases, eyewitness researchers are not. Hundreds of eyewitness laboratory studies have also shown that nearly one in five identification decisions is an incorrect choice. Not only do the experimental laboratory studies confirm the actual data, but they also provide us with vital information that cannot be learned from actual cases and they allow us to make cause and effect conclusions about how identification procedures influence eyewitness accuracy. The accumulation of laboratory data over the last three decades has allowed eyewitness researchers to be in the position to make specific policy recommendations to law enforcement and legislators on how to increase the reliability of eyewitness testimony presented in court. To be sure, the "best practice" recommendations described in this entry have not been universally or equally embraced, but there has been a significant and enduring impact for those jurisdictions that have made the improvements to their identification practices.

Over a period of decades, researchers have established that when we experience an important event, we do not simply record it in our memory as a video recorder would. The situation is much more complex and most theoretical analyses of the memory process divide it into three major stages: encoding/acquisition, retention, and retrieval. First, the event is perceived by a witness and information is entered into the memory system. Next, some time passes before a witness tries to remember the event. Finally, the witness tries to retrieve the stored information. Psychologists who conduct research on eyewitness memory study and attempt to identify the factors that play a role in each of the three stages. In 1978, Professor Gary Wells, the leading researcher in the field of eyewitness identification, published an article that continues to guide the way in which eyewitness research is discussed and conducted. In that article, he introduced the terms "estimator and system variables" to refer to factors that typically relate to the encoding, retention, and retrieval stages of memory.

As it relates to law enforcement, research has shown that the procedures and practices that police use during the memory retrieval process can influence the reliability of an eyewitness identification and the witness' subsequent testimony. Examples of police procedures or system variables that can affect witness reliability include pre-identification instructions, the type of identification procedure administered (e.g., showup, simultaneous lineup, sequential lineup), whether the identification was conducted using a double-blind administrator, and the type of post-event information provided to a witness after their identification decision, to name a few. Each of the system variables listed above has received consideration in the eyewitness research community and will be reviewed below. Without question, however, two topics have received more attention and caused more controversy than the others combined: double-blind administration and sequential lineups.

Double-Blind Administration

At its core, the police lineup is akin to a scientific experiment, where a police officer tests the hypothesis that the suspect is the perpetrator. And thus, many of the principles that apply to conducting a research experiment similarly apply to the police lineup. Arguably the most important principle in this analogy is double-blind administration, where the person who is conducting the experiment does not know the "correct" answer and thus cannot consciously or unconsciously influence the participant or the outcome.

The impact of lineup administrator knowledge on witness decisions has been tested in laboratory settings and the results are consistent with the vast body of research on double-blind testing. One example is a study by Greathouse and Kovera (2009) which used witness-administrator pairs to test the hypothesis that an administrator could subtly influence the outcome of the procedure if they were aware who the suspect was. Administrators were given information about the case and told that they would receive a cash bonus if the witness selected the suspect, who was actually innocent of the crime. Half were then given the identity of the suspect. In this, non-blind, condition, 21 % of witnesses chose the innocent suspect, whereas only 9 % chose him in the double-blind condition. Greathouse and Kovera also surreptitiously video-recorded the interaction between the witness and the administrator and showed that even though the administrators believed they had not influenced the witnesses, their verbal and nonverbal behaviors suggested otherwise and were evident to those watching the videotape of the interaction.

As it pertains to police lineups, double-blind administration requires that the person who shows the lineup to a witness cannot know which person is the suspect and which are the fillers. This procedure eliminates the possibility that the detective hinted to the witness (consciously or not) which person to choose. This is a significant departure from the way in which law enforcement conduct lineup procedures and, not surprisingly, this recommendation has been met with resistance by many law enforcement agencies and prosecutors across the United States. A common (mis)interpretation of this recommendation is that detectives are being accused of deliberately leading witnesses and influencing the outcome of lineups. Although it is possible for deliberate influence to occur, eyewitness researchers believe this is rare. Instead, what is more likely to occur is the transmission of unconscious cues, such as nodding and smiling, when a witness is making their decision. The collective experience of many researchers is that after speaking with law enforcement and policy makers and educating them on the science and logic behind the double-blind administration recommendation, these groups are significantly more likely to embrace the reform. In addition to the International Association of Chiefs of Police (IACP) recommending double-blind administration, some states (CT, NC, TX) have changed their laws and now require that officers conducting identification procedures not know when the witness is viewing the suspect.

To be certain, there are, on the surface, reasonable objections to double-blind administration, including concerns regarding the additional cost of utilizing a separate person unfamiliar with the case or the logistics for smaller police departments who may not be able to find an officer who is not familiar with the case. In response to the latter concern, eyewitness researchers have recommended conducting the lineup in a "blinded" manner, where the officer is unable to see which photograph a witness is viewing at any particular time. This is typically done using the "folder shuffle method" where individual photographs are placed into unmarked folders. The folders are then shuffled, rendering the lineup administrator unaware of which lineup member is in each folder. The folders are then handed to the witness, one at a time, following the sequential procedure described in the next section. Blinded administration can also be achieved by using computer software to present the images in a random order and having the lineup administrator unable to see the screen. In the years to come, it is likely that more and more departments across the country will begin testing and using these methods.

Sequential Lineup Presentation

Perhaps equally controversial as the subject of double-blind administration is sequential lineup presentation. In sequential lineups, a witness views the suspect and fillers one at a time and makes a judgment about each person before the next is shown. In simultaneous or traditional lineups, the suspect and lineup fillers are presented at the same time and the eyewitness identifies which (if any) is the perpetrator. More than 20 years of laboratory research indicates that sequential lineups cut the rate of false identifications in half when compared to the traditional simultaneous lineup method (Steblay et al. 2011) and thus sequential lineups have become a recommended "best practice" (see Wells et al. 1998).

In addition to the results from laboratory studies, researchers have also compared (double-blind) simultaneous and (double-blind) sequential lineups using actual eyewitnesses to real crimes in ongoing criminal investigations. In a study sponsored by the American Judicature Society, Wells et al. (2011) found that sequential lineups produce fewer mistaken identification decisions than lineup procedures that present the photographs simultaneously. The study found that double-blind sequential lineups (compared to double-blind simultaneous lineups) as administered by police departments in four states across the country resulted in the same number of suspect identifications (27.3 % for sequential and 25.5 % for simultaneous) and fewer known-innocent filler identifications (12.2 % for sequential and 18.1 % for simultaneous).

One of the most notable results from the field study was that witnesses in these actual criminal cases who made positive identifications ("yes, that is the person I saw commit the crime") from a simultaneous lineup made an identification error 42 % of the time. That is, four out of every ten positive identifications that were obtained from double-blind simultaneous lineups were mistaken identifications of innocent lineup fillers. Even with the double-blind sequential procedure, three of every ten identifications were of an innocent filler. Thus, even when best practices are followed, identification errors are not entirely eliminated and witnesses can still be unreliable.

The dominant explanation for the difference in errors between simultaneous and sequential lineups is that witnesses who view simultaneous lineups – and do not immediately recognize anyone – are more likely to engage in a relative judgment process, whereby they compare the lineup members and choose the one who most closely resembles their memory for the perpetrator. Witnesses who view the images one at a time, however, are less able to engage in this comparison process and therefore are more likely to make an identification based on their memory rather than a combination of their memory and choosing the person who is the best option out of those presented.

The controversy over the use of sequential lineups arises from the finding that accurate suspect identifications are higher with simultaneous lineups than with sequential, and thus there is a concern that guilty suspects might "get away" if witnesses do not identify them. The concerns about sequential lineups were recently thoroughly vetted in several articles in the journal "Perspectives on Psychological Science," and the interested reader is directed to those articles for the full set of arguments. In summary, researchers who are cautious about recommending sequential lineups focus on the loss of correct suspect identifications (Clark 2012), and those who recommend sequential lineups ask "Why would a witness be unable to identify the suspect unless all of the lineup members were presented at the same time?" (Wells et al. 2012). In other words, if the witness has a great memory of the perpetrator and truly recognizes him, then she should be able to do so regardless if the lineup members are presented simultaneously or sequentially. Further, any loss in suspect identifications might be characterized as a loss of "lucky guesses" that resulted merely from relative judgments in the simultaneous procedure.

To date, there are several states and many jurisdictions that require the use of sequential lineup presentation. Thus, it appears that these groups who have considered the data and the "controversy" have been unmoved by the "loss of suspect identifications" argument made by those who question whether sequential lineups are superior to the simultaneous technique. Whether additional jurisdictions who are considering reform will be persuaded by these arguments is presently unknown. Whether new identification procedures can be developed to further reduce errors will likely be the subject of research over the next few decades. Until then, it is expected that more and more jurisdictions across the United States will adopt the sequential lineup procedure.

Pre-identification Instructions

Maybe the least contentious of all lineup reforms is the recommendation that witnesses be warned, prior to viewing a lineup, that the actual perpetrator may or may not be present in the procedure they are about to view.

Pre-lineup instructions have been studied by eyewitness researchers since the early 1980s (Malpass and Devine 1981). In this research, participants typically view a staged crime event and are then randomly assigned to an instructions condition. "Unbiased" instructions include wording to the effect that the actual perpetrator "may or may not be present," whereas biased (or standard) instructions provide no such warning. In the first study on this topic, Malpass and Devine found that witnesses who received biased instructions were 45 % more likely to make a false identification of an innocent person than were participants who were not told that the perpetrator may not be included in the array.

In 1997, Steblay published a meta-analytic review of the instruction bias literature and found, overall, a 25 % increase in eyewitness accuracy when unbiased instructions were provided to witnesses. Interestingly, Steblay also found that the mere failure to provide witnesses with a "may or may not be present" instruction had similar detrimental effects on accuracy as when witnesses were given some pressure to make a positive identification. Thus, failing to provide the warning is considered suggestive because it insinuates that the true perpetrator is in the lineup procedure. For these reasons, the National Institute of Justice Technical Working Group for Eyewitness Evidence (1999) included pre-lineup instructions in their research report titled *Eyewitness Evidence*: *A Guide for Law Enforcement*. In addition to discussing other eyewitness identification best practices, this report specifically recommended giving cautionary instructions to reduce the possibility that a witness would choose someone from the lineup merely because they expect that the police must have caught the perpetrator; otherwise, they would not be viewing an array. Some jurisdictions have gone further than the "may or may not be present" warning and now inform witnesses, among other things, that the investigation will continue even if the witness does not make an identification. These pre-lineup recommendations have been met with little resistance by law enforcement and others, and thus it is likely that this reform will continue long into the future.

Witness Confidence and Post-identification Feedback

Although there is a presumption among many actors in the legal system that there is a significant and positive confidence-accuracy relationship, serious questions about whether that is true have been raised by researchers within the law and psychology community. Decades of research show that there is only a small to moderate relationship between the accuracy of an eyewitness' identification and their confidence in that identification and that this relationship can be significantly affected by post-identification factors, making confidence highly malleable.

Confidence malleability refers to changes in a witness' confidence over time, generally as a result of events that take place after the identification. By the time a witness takes the stand at trial, their expression of confidence in their identification can result from multiple sources, including the strength of their memory of the person they saw but also a variety of system variables, such as learning that the person they identified has been charged with the crime or that other witnesses also identified the same person. For this reason, it is recommended that law enforcement ask witnesses, immediately after they make a positive identification, to state how confident they are in their decision. A statement of certainty at that point in time, unaffected by the passage of time and external influences, can be a more reliable indicator of accuracy than a statement taken months or years after the identification.

In their pioneering research on the topic of confidence malleability, Wells and Bradfield (1998) found that witnesses who were told that they had made a correct decision ("Good, you identified the suspect") were more confident than witnesses who were given no feedback. What was, and remains to be, surprising about this research, however, is that feedback has significant effects on a witness' recollection of other aspects of the event, including exaggerations in how much attention the witness paid to the perpetrator during the crime, how good a view they had, and so on. The results of this groundbreaking study have been replicated many times in both research labs (Bradfield-Douglass and Steblay 2006) and with actual witnesses in ongoing criminal investigations (Wright and Skagerberg 2007).

One of the explanations that have been proposed to explain the post-identification feedback effect, and its strong and pervasive influence on eyewitness confidence, is the theory of cognitive dissonance (Festinger 1957). In essence, this theory states that people are in a state of discomfort when they have inconsistent or contradictory beliefs, or when they have beliefs and behaviors that are inconsistent. As it relates to eyewitness identification, if a witness is confident in their identification, then they subsequently reason that they must have had a good opportunity to see the perpetrator. How else would they be able to "get it right"? A powerful example of cognitive dissonance is the DNA exoneration case of Dean Cage from Illinois. After Cage was exonerated in 2008, the victim refused to believe the accuracy of the DNA results and held on to her belief that Cage was guilty. Thus, cognitive dissonance was so powerful in that case that it was easier for the witness to believe that the DNA testing was flawed than to accept that she had made an error and identified an innocent person. Only after she was presented with independent results of the DNA testing did she come to accept that Cage was innocent and was not the man who had raped her in 1994.

The results of laboratory and field studies on post-identification feedback have led to the recommendation that law enforcement obtain a statement of witness confidence before any feedback about their decision is provided. Even then, it is not recommended that law enforcement tell witnesses that they were correct in their identification because of the pervasive influence feedback has on other memory-related judgments (attention, view, etc.) which could ultimately have an influence on a witness' credibility during their trial testimony.

Future research on post-identification feedback will likely examine whether there are ways to substantially mitigate the effect and protect the reliability of a witness' testimony while at the same time recognizing that actual witnesses will eventually receive information that the person they identified was charged with the crime and that the witness is needed to testify at trial. Steps have been made in this direction and the current recommendation of using double-blind administration – where the administrator cannot give meaningful feedback – appears to have some impact (Dysart et al. 2012).

Showups

The showup is an identification procedure in which law enforcement shows a single suspect – unaccompanied by fillers – to the eyewitness and asks them to decide whether the suspect is the perpetrator. Showups are generally conducted soon after the crime, and the suspect may either be viewed live (a "field" showup), the most common option, or the eyewitness may be shown a single photograph.

By definition, the showup is a suggestive procedure because only one person is presented to a witness and thus the identity of the police suspect is obvious. Showups can never be conducted double-blind; thus administrator bias is necessarily a concern with showups. Further increasing suggestiveness in police investigations is the fact that the suspect may also be handcuffed or seated in the back of a police car when the witness views him. Despite these reliability issues, showups are routinely used by law enforcement (between 30 % and 77 % of all identifications procedures in various areas of the country are showups, Dysart and Lindsay 2007) and largely accepted by the courts.

The research on showups is limited in contrast to lineups and much of the existing literature specifically compares accuracy between showups and lineups. A 2003 meta-analysis synthesized this research (Steblay et al. 2003) and highlighted a specific danger unique to showups: 100 % of identifications made from showups that contain an innocent suspect are false identifications. In lineups, however, filler identifications represent a measure of protection to a suspect. That is, if a witness was merely guessing in a lineup, they would have a one in six chance of guessing the suspect (in a six-person array). Innocent suspects presented in showups have little defense against a witness with a weak memory and a willingness to choose and are more likely to be falsely identified than if they were presented in an unbiased six-person array.

From a policy standpoint, there are two main reasons that showups continue to be used and that their use is likely to continue. The first reason is related to public safety. If police are not permitted to conduct showups, and yet have too little evidence to arrest a suspect, the guilty party may be set free to commit further criminal acts. The second reason is to allow innocent suspects to be cleared without undue delay. If the police have detained an innocent person, the sooner they present him to the eyewitness, the sooner he can be released, assuming the eyewitness correctly rejects him as the perpetrator.

Procedural issues and controversies regarding lineups have largely dominated the discussion about best practices in identification procedures, but, given the frequency with which showups are used by law enforcement, it is time for research and policy makers to pay greater attention to the showup procedure. A greater understanding is needed not only of how factors associated with reduced lineup accuracy affect showups but also how factors unique to the showup may affect the ability of eyewitnesses to make accurate identification decisions.

Co-witness Discussion

Co-witness discussion occurs when multiple witnesses to a crime speak with one another about what they viewed or heard. One of the reasons why co-witness discussion is of concern to researchers is that witnesses can incorporate information learned from other witnesses into their memory of the perpetrator and the event. As a result, the National Institute of Justice (1999) recommends that law enforcement, as part of their routine procedures, separate witness when they are being interviewed and when they view an identification procedure.

A particularly illustrative study on co-witness contamination was conducted by Hope et al. (2008), wherein participants viewed a videotaped event in pairs and then either discussed the event with the other person or answered questions about the video on their own with no discussion. The catch in the study was that only one member of each pair saw a theft take place because the event was filmed with two cameras from different angles and only one camera caught the critical theft. For participants in the study who did not see a theft take place, no one in the "no discussion" group reported seeing the theft but 50 % of those who had discussed the video with their partner reported a theft they in fact did not see. These witnesses were susceptible to misinformation from their co-witness and, as a consequence, produced less accurate recall accounts than participants who did not interact with another witness.

In another study, Zajac and Henderson (2009) showed that co-witness contamination can affect not only description accuracy but identification accuracy as well. In this study, research participants were paired with a research confederate whom they believed was just another participant in the study. Together, they viewed a video clip of a staged theft. Then, half of the participants were misinformed by the confederate that a target person in the video had blue eyes when in fact they were brown. Individually, participants then described the accomplice and viewed a lineup where the target person was not shown. In fact, all of the lineup members had blue eyes. Participants who had received the blue-eyed misinformation were several times more likely to describe the target as having blue eyes and twice as likely to falsely identify someone from the lineup.

In summary, the research literature on co-witness discussion/contamination shows that merely having a conversation with another person about an event can result in substantial changes to a person's memory and accuracy of that event. And thus, the recommendation that witnesses be separated and instructed not to speak with other witnesses is a critical element in assuring that the evidence a witness testifies to is from their own perceptions and has not been a result of contamination.

Multiple Presentations of the Same Suspect

The final topic that deserves attention within a discussion of system variables is the practice that law enforcement have of showing the same suspect to a witness in more than one identification procedure, such as mug book or a showup followed by a lineup. When this situation arises, it is important to consider the potential effects on reliability of having previously viewed or selected the suspect.

In research spanning three decades, investigators have examined the effects of repeated viewing of the same suspect and have identified two main effects: commitment and unconscious transference. Commitment occurs when a person who is selected from an earlier identification procedure is selected a second time in a subsequent procedure, regardless of if he or she is the correct person. Unconscious transference, on the other hand, occurs when an innocent person is merely viewed by a witness in one context and is only later selected as the target from a second procedure. The phenomenon of unconscious transference has likely plagued most people at one time or another as evidenced in the question "where do I know that face?" Witnesses that view a person in multiple identification procedures or in multiple contexts (e.g., in a photo array and then in court) are faced with a similar question. The correct answer is for the witness to say "I saw that face from several different contexts," but the erroneous conclusion is that the face is familiar **only** because it is the face of the perpetrator. The concern is that this sense of familiarity on the part of the witness may lead to an increased feeling of confidence in subsequent identification procedures. In fact, a meta-analysis on transference from viewing mug shot photographs confirms that witnesses are more likely to pick from a lineup a person previously viewed (Deffenbacher et al. 2006). Further, the 2006 meta-analysis confirmed that a person who has been identified in an earlier procedure is considerably more likely to be identified in a subsequent procedure.

Although it may seem obvious that a suspect should only be placed in one law enforcement identification procedure, the data from the United States DNA exoneration cases indicates that multiple identification procedures were used in a surprising number of those cases. It is also a relatively common practice for law enforcement to conduct a showup then a lineup, or a photo lineup followed by a live lineup procedure. Although this practice is common, research is very clear that one procedure can influence the reliability of the next and subsequent identification procedures should be viewed with great caution. Without a doubt, the practice of displaying multiple identification procedures to the same witness with the same suspect will be a subject of identification reform in the very near future.

Conclusions

Eyewitness errors are devastating. When an eyewitness makes an honest mistake and chooses an innocent suspect from an identification procedure, not only does that person suffer but the guilty person remains free, at liberty to commit additional crimes. However, research has shown that using modified techniques, such as the sequential double-blind lineup with pre-lineup warnings, will lead to fewer innocent people being convicted of crimes they did not commit. It is critical for eyewitness errors to be prevented in an investigation rather than litigated and disputed in court because both lab research and real-life cases have repeatedly shown that juries (and judges) are not able to distinguish between accurate and inaccurate witnesses. Instead, triers of fact are too often persuaded merely by a witness who appears confident but nonetheless has made a devastating error. Changes in law enforcement procedures with respect to eyewitness identifications, whether by choice or legislative mandate, are taking hold across the country with more reforms and changes to come.

Related Entries

- ▶ Causes of Wrongful Convictions
- ▶ Estimator Variables and Eyewitness Identification
- ▶ Eyewitness Evidence: Historical Background
- ▶ Eyewitness Identification Evidence
- ▶ Innocence Commissions
- ▶ Interviewing Eyewitnesses
- ▶ Jury Decision Making and Eyewitness Testimony
- ▶ Measuring Wrongful Convictions

Recommended Reading and References

Bradfield-Douglass A, Steblay N (2006) Memory distortion in eyewitnesses: a meta-analysis of the post-identification feedback effect. Appl Cogn Psychol 20:859–869

Clark S (2012) Costs and benefits of eyewitness identification reform: psychological science and public policy. Perspect Psychol Sci 7:238–259

Deffenbacher K, Bornstein B, Penrod S (2006) Exposure effects: retroactive interference, mug shot commitment, source confusion, and unconscious transference. Law Hum Behav 30:287–307

Dysart JE, Lindsay RCL (2007) Show-up identifications: suggestive technique or reliable method? In: Lindsay RCL, Ross DR, Read JD, Toglia MP (eds) The handbook of eyewitness psychology, vol II, memory for people. Lawrence Erlbaum, Mahwah, pp 137–154

Dysart JE, Lawson VZ, Rainey A (2012) Blind lineup administration as a prophylactic against the post identification feedback effect. Law Hum Behav 36:312–319

Festinger L (1957) Cognitive dissonance. Sci Am 207(4):93–107, 1962

Garrett BL (2011) Convicting the innocent: where criminal prosecutions go wrong. Harvard University Press, Cambridge, MA

Greathouse S, Kovera M (2009) Instruction bias and lineup presentation moderate the effects of administrator knowledge on eyewitness identification. Law Hum Behav 33:70–82

Hope L, Ost J, Gabbert F, Healey S, Lenton E (2008) 'With a little help from my friends...': the role of co-witness relationship in susceptibility to misinformation. Acta Psychol 127:476–484

Malpass R, Devine P (1981) Eyewitness identification: lineup instructions and the absence of the offender. J Appl Psychol 66:482–489

National Institute of Justice Eyewitness Technical Working Group on Eyewitness Evidence (1999) Eyewitness evidence: a guide for law enforcement. United States Department of Justice, Office of Justice Programs, Washington, DC

Steblay N (1997) Social influence in eyewitness recall: a met analytic review of lineup instruction effects. Law Hum Behav 21:283–298

Steblay N, Dysart J, Fulero S, Lindsay RCL (2003) Eyewitness accuracy rates in police showup and lineup presentations: a meta-analytic comparison. Law Hum Behav 27:523–540

Steblay N, Dysart J, Wells G (2011) Seventy-two tests of the sequential lineup superiority effect: a meta-analysis and policy discussion. Psychol Publ Policy Law 17:99–139

Thompson-Canino J, Cotton R, Torneo E (2010) Picking cotton. St. Martin's Press, New York

Wells G (1978) Applied eyewitness testimony research: system variables and estimator variables. J Pers Soc Psychol 36:1546–1557

Wells G, Small M, Penrod S, Malpass R, Fulero S, Brimacombe CAE (1998) Eyewitness identification procedures: recommendations for lineups and photospreads. Law Hum Behav 22:603–647

Wells G, Steblay N, Dysart J (2011) A test of the simultaneous vs. sequential lineup methods: an initial report of the AJS national eyewitness identification field studies. American Judicature Society, De Moines, IA

Wells G, Steblay N, Dysart J (2012) Eyewitness identification reforms: are suggestiveness-induced hits and guesses true hits? Perspect Psychol Sci 7:264–271

Wright D, Skagerberg E (2007) Post-identification feedback affects real eyewitnesses. Psychol Sci 18:172–178

www.InnocenceProject.org

Zajac R, Henderson N (2009) Don't it make my brown eyes blue: co-witness misinformation about a target's appearance can impair target-absent line-up performance. Memory 17:266–278

Eysenck's Personality Theory

Adrian Furnham[1] and K. V. Petrides[2]

[1]Division of Psychology and Language Sciences, Faculty of Brain Sciences University College London, London, UK

[2]Research Department of Clinical, Educational, and Health Psychology, Educational Psychology Group University College London, London, UK

Overview

This entry looks at the research on personality correlates and determinants of crime following the influential views of the late Hans Eysenck.

Eysenck, a differential psychologist, made a strong theory and empirically based claim for the idea that personality factors of individuals make them more likely to take part in delinquent and criminal acts. The theory is reviewed and critiqued.

Introduction

Hans Eysenck was perhaps the most prolific psychologist of his day and is still one of the most quoted in the world. Prof. Jeffrey Gray, student of and successor to Hans Eysenck, once famously described his personality theory as akin to somebody finding St. Pancras Railway Station in the jungle. The station is an extremely impressive piece of highly elaborate, complex, and beautiful Victorian architecture situated in central London. What Gray meant was that Eysenck's theory stands out dramatically from all around it.

It has been argued that Eysenck's approach to science was characterized by very specific principles (Netter 2007). First, he always argued, even at a time when this was deeply unfashionable, that there was a physical/biological basis to personality. He maintained that taxonomization was the beginning of science and that personality research could not proceed without it. He insisted on a hierarchical model with highly specific behavioral responses at the lowest level, leading up to broad habitual responses at the facet level (e.g., sociability, liveliness, and excitability) and culminating into three giant super-factors at the apex of the hierarchy. He was one of the earliest theorists to advocate a biologically based theory of personality and to promote a continuous theory refinement approach in order to link up specific stimulus properties with general personality functioning.

The Taxonomy of Personality

At present, the Eysenckian model still has three – and only three – super-factors: Psychoticism, Extraversion, and Neuroticism (known as P-E-N). Eysenck was quite convinced that these three conceptual and descriptive categories are necessary and sufficient for a thorough understanding of individual differences in personality.

Extraversion is perhaps the best known of the three Eysenckian dimensions and is at the very "heart" of the model. Eysenck's cortical arousal theory of Extraversion has been extensively described in numerous books (e.g., Eysenck 1973), book chapters, and articles. Revelle (1997) succinctly summarized the classic Eysenckian hypothesis of arousal underpinning his Extraversion factor. "The basic assumptions were: (1) introverts are more aroused than extraverts; (2) stimulation increases arousal; (3) arousal related to performance is curvilinear; (4) the optimal level of arousal for a task is negatively related to task difficulty; and (5) arousal related to hedonic tone is curvilinear. Assumption 1 was based upon many studies associating EPI-E with (low) physiological arousal Eysenck (1967). Assumptions 3 and 4 were based upon the Yerkes-Dodson Law (Yerkes and Dodson 1908) and the subsequent support for it by Broadhurst (1959). Assumption 5 was founded on Berlyne's (1960) discussion of curiosity and arousal. Based upon assumptions 1–4, it can be predicted that introverts should perform better than extraverts under low levels of stimulation but should perform less well at high levels of stimulation. Similarly, assumptions 1, 2, and 5 lead to the prediction that extraverts should seek out more stimulation than introverts" (p. 199).

Neuroticism is based on activation thresholds in the sympathetic nervous system or visceral brain. This is the part of the brain that is responsible for the fight-or-flight response in the face of danger. Neurotic persons have a low activation threshold: when confronted with even mild stressors or anxiety-inducing situations, they quickly and easily experience negative affect and become upset. These manifestations can range from physiological changes in heart rate, blood pressure, cold hands, sweating, and muscular tension to feelings of apprehension and nervousness, to the full effects of fear and anxiety. In contrast to Neuroticism (or emotionally unstable and labile persons), their emotionally stable peers have a much higher activation

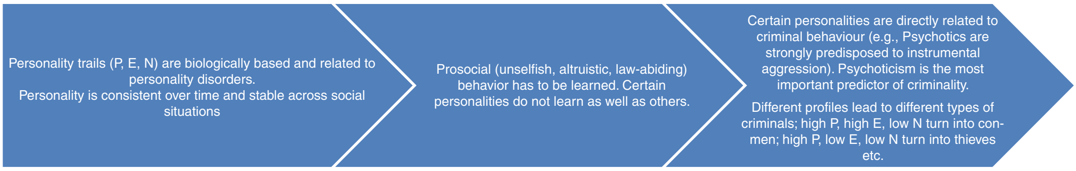
Personality trails (P, E, N) are biologically based and related to personality disorders.
Personality is consistent over time and stable across social situations
Prosocial (unselfish, altruistic, law-abiding) behavior has to be learned. Certain personalities do not learn as well as others.
Certain personalities are directly related to criminal behaviour (e.g., Psychotics are strongly predisposed to instrumental aggression). Psychoticism is the most important predictor of criminality.
Different profiles lead to different types of criminals; high P, high E, low N turn into con-men; high P, low E, low N turn into thieves etc.

Eysenck's Personality Theory, Fig. 1 Causal pathways in Eysenck's theory

threshold, and thus will experience negative affect only when confronted by major stressors. Such individuals tend to be calm even in situations that could be described as anxiety inducing or pressure laden. A full description of neurosis is provided by Eysenck in his 1977 book "You and Neurosis."

Psychoticism, the most underdeveloped and debated of Eysenck's three super-factors, was most fully described by Eysenck and Eysenck (1976) in the book "Psychoticism as a Dimension of Personality." High P scorers are predisposed to psychotic episodes but also tend to manifest a higher probability of engaging in aggression and demonstrating the kind of cold tough-mindedness that characterizes psychopathy and persons more likely to engage in crime (Eysenck 1977). While somewhat less fully described and with less empirical support than either E or N, the research that has been done has indicated that P also has a biological basis. The P factor is particularly relevant to crime and criminology. Indeed, the fact that criminals are overwhelmingly male was one of the key observations that Eysenck made in order to hypothesize that the biological basis of P may be related to androgens (e.g., increased testosterone levels).

The past 20 years has witnessed a protracted debate between pro-Eysenckians, who support various permutations of the Giant Three Eysenckian system, and those who support the various Big Five models. The two positions were clearly debated between proponents in conferences and journal articles. For those who were so fortunate to witness the debates between Eysenck and Costa, and others such as Goldberg (e.g., at the ISSID conference in Oxford University in 1991), these were truly akin to the "clash of the Titans." While the debate was always intense, each "proponent" bringing out more data to support their position, it was always respectful of each person's work and commitment as a scientist.

Costa and McCrae (1992) summarized the evidence for the validity of the five-factor model by stating the "four ways the five factors are basic." These were, *first*, that longitudinal and cross-sectional studies have shown five robust factors to underlie enduring behavioral dispositions; *second*, that traits associated with the five factors emerge from different personality data sources (self-reports versus other reports) as well as from studies of natural language (i.e., lexicon) conducted in many different countries; *third*, that the five factors are found in different age, sex, and ethnic groups; and *fourth*, that their heritabilities point to some underlying biological basis. Subsequently, further evidence demonstrated cross-cultural similarities and stabilities in the developmental and ageing trajectories of the five factors (Furnham et al. 2008).

Eysenck (1992) rigorously criticized five-factor models of personality. He suggested that the criteria set out by Costa and McCrae for accepting the five-factor model (some of which Eysenck had himself proposed and pioneered, such as heritability studies of personality) are necessary, but not sufficient for determining the essential dimensions of personality. He argued that Agreeableness and Conscientiousness are lower-level traits that can be conceptualized as facets of his higher-order factor Psychoticism. Additionally, he contended that Openness forms a part of Extraversion and (low) Conscientiousness a part of Neuroticism. Thus, O, A, and C were lower order, not super-factors. Incidentally, a new line of research focusing on the possibility of a General Factor of Personality has received its most emphatic support from robust intercorrelations between the supposedly orthogonal Big Five factors (Rushton et al. 2009).

Eysenck also pointed out the meta-analysis of factor analytic studies of personality, carried out by Royce and Powell (1983), which indicated a three-factor model similar to his own. Eysenck suggested that the five-factor model lacks a theoretical basis sufficiently grounded in empirical facts and is, therefore, arbitrary. He explicitly juxtaposed this arbitrariness against his conceptualizations of Neuroticism and Psychoticism, which were firmly rooted in mental health and clinical psychology research.

Eysenck (1992) argued that it is the nomological network in which a dimension is embedded that corroborates its validity. This network must include a detailed specification of a personality dimension's biological and psychophysiological origins, its cultural invariance, and its relationship to social behavior and psychological illness. Indeed, one of Eysenck's most significant contributions to personality research was his shifting of the field's agenda, from nebulous Freudian speculation to rigorous biopsychological integration (Eysenck 1967).

Eysenckian Measures

Eysenck devised and validated numerous inventories that are still widely in use. Furnham et al. (2008) suggested five reasons why these measures have stood the test of time:

(A) *Parsimony*. The P-E-N model offers a first-class conceptual foundation for the Eysenck Personality Questionnaire (EPQ), which is one of the most parsimonious and psychometrically robust personality inventories. It compares favorably with the 16 dimensions of Cattell's 16PF or the Big Five, Six, or Seven (there have been several attempts to expand Big Five models by adding allegedly new traits). In some ways, focusing on second-order factors is akin to the role of "g" in the measurement of intelligence. While some psychologists have argued through the bandwidth-fidelity debate for finer grain analysis at the facet (primary and lower-order) factor level, to some extent the Eysencks resisted this trend although the creation of measures such as the I_6 and I_7 did provide an assessment of several key variables including venturesomeness, impulsivity, and empathy. Recently, there has been a focus on primary factors and their acolytes started working with the EPP. More recently there has been considerable interest in the EPP and the facet/primary factor level of analysis (Jackson et al. 2000). While the Eysencks always measured super-factors or traits (i.e., the Gigantic Three), some researchers have argued that a finer-grained level of analysis (i.e., facet level) enables one to understand the nature of processes better (Petrides et al. 2003).

(B) *Explanation of Process*. More than any other test developers in the twentieth century, the Eysencks were not content to just describe and categorize traits: they constantly sought to escape tautological explanations. Eysenck's was the first counterintuitive theory of arousal to explain individual differences in Extraversion. Hans Eysenck showed, nearly a decade before the publication of the EPQ (Eysenck 1947), that his work and interests were solidly grounded in the biological basis of personality, which he set about exploring for the next 30 years (Eysenck 1967, 1982b). More than any other personality theorist, he was concerned with describing and explaining the mechanisms and processes which account for systematic individual differences. Eysenck's theory and model of individual differences is fully described in an earlier chapter by O'Connor and will not be repeated here; only a thumbnail sketch will be provided to provide the context for the development of the various personality measures developed by both Hans and Sybil Eysenck and colleagues.

(C) *Experimentation*: Eysenck was rigorously schooled in the experimental tradition at a time when behaviorism and psychoanalysis triumphed. He believed it essential that the correlational (individual differences) and experimental branches of psychology unite, so that personality effects will cease to be treated as error variance (Howarth and Eysenck 1968). This basic premise was the foundational underpinning for the creation of both the International Society for the Study of Individual Differences (ISSID) and also the Society's official journal, Personality and Individual Differences, arguably the leading journal in the field.

(D) *Wide Application*. The Eysencks were psychological pioneers as well as risk takers; they were eager to extend their research program to look at the significance of personality functioning in areas as disparate as sex, crime, parapsychology, astrology health, and behavioral genetics among many others. They were convinced that individual differences were systematically and predictably present in all facets of human functioning. To many, particularly those favoring sociological or more exclusively environmental perspectives (e.g., based on parenting styles or early childhood experiences), this was radical indeed.

(E) *Continuous Improvement and Development*. For nearly 40 years the Eysencks have engaged in a systematic research program aiming to improve, update, and validate their range of personality measures. Because their scales have been so widely used in research worldwide, they have been subject to detailed but, at times, ungenerous scrutiny by many people. Their observations and research studies, particularly those emphasizing a cross-cultural perspective, led to continuous improvements over time, with items being removed, changed, and replaced. Importantly, all of this psychometric development was guided not by commercial opportunities, but firmly by theoretical considerations and empirical evidence.

Personality Structure Cross-Culturally

Most personality theorists hypothesize that major individual differences variables, such as intelligence and personality, are universal constructs. While they may vary in their expression (phenotype) as a function of culture and other psychosocial factors, basic traits are common across humankind. Some psychologists would also argue that the measurement of their specifically described traits or constructs would be universal across all countries and cultures. Considerable research has been published over the years demonstrating the significant contribution that genetic, biological, neurological, and environmental factors make in shaping the expression of intelligence and personality.

For over 20 years, Sybil Eysenck published research comparing the results of studies on adults and children using the adult and junior versions of the EPQ. In 1981, the Eysencks summarized the results of 14 countries based on just under 15,000 subjects (Eysenck and Eysenck 1981). In 1985, they extended their analysis to 24 countries and in 1998 this was further extended to 34 countries (Barrett et al. 1998). As each national group was added, so inevitably the population sample increased, but more impressively so did the quality of the statistical analysis. Sybil Eysenck also published a dedicated book detailing specific studies of the Eysenck scales used in different countries with children and adolescents.

Personality and Crime

Perhaps the most robust and fecund of the theories of the criminal personality is that of Eysenck (1977). Feldman (1993) has noted: "One of the theory's great merits is that it makes predictions which are clear cut, testable and refutable" (p. 166).

The idea is beguilingly simple. Eysenck believed that sociological theory has little to offer society on the causes of crime, arguing that psychological theories have more explanatory power (Eysenck and Gudjonsson 1989). Criminal behavior is not the product of either environment *or* biology alone, but rather of a dynamic interaction between the two. This extends Eysenck's original hypothesis that biology played the largest part in determining criminality, expounded in his book "Crime and Personality" (Eysenck 1964). Eysenck suggested that some people are born with cortical and autonomic nervous systems that compromise their ability to learn from their environment and renders them susceptible to illegal and criminal acts.

Each of the three Eysenckian factors relates to some aspect of conditionability, notably the learning of rules of society.

Extraverts are social and impulsive. They are excitement seekers interested in novel experiences, which often leads them to be poorer learners than introverts at many tasks, including the acquisition of general social rules. They are somewhat more likely than introverts to become delinquents or criminals.

Neurotics are anxious, moody, restless, and rigid. They react strongly to threat often with great fear to painful stimulation. This means they too do not learn social rules well and are "inefficient learners" particularly with respect to punishment.

Psychotic (*tough-minded*) individuals tend to be aggressive, egocentric, insensitive, inhumane, uncaring, and troublesome, and they are far more likely than their tender-minded peers to engage in delinquent and criminal behavior.

Those who score high on the P-E-N end of the Eysenckian dimensions tend to be least socially conditioned and, hence, least socially restrained. They are aggressive, hedonistic, impulsive, and reckless. The Eysenckian view links personality and criminality with a simple mechanism: personality traits are linked to learning pro- and antisocial behaviors that are linked to delinquent and criminal behavior and recidivism.

Eysenck and Gudjonsson (1989) in summary note as follows:

1. There exists a general behavior pattern of antisocial behavior and criminality, marking the opposite end of a continuum that has prosocial and altruistic behavior at its other end.
2. Within the antisocial and criminal type of behavior, there is a certain amount of heterogeneity, marked particularly by the opposition between active and inadequate criminals (UNCLEAR), but probably also including differences according to type of crime committed.
3. Criminality is related to certain dimensions of personality, most especially Psychoticism, which is apparent across all age groups and conditions studied.
4. Extraversion is also linked to criminality, particularly in younger samples and among more active criminals; inadequate older criminals do not show high extraversion and may indeed be below average on this trait.
5. Most criminals are characterized by a high degree of Neuroticism, but this may not be found as markedly in children and youngsters.
6. Scores on the social desirability (L) scale are regarded in these studies as a measure of conformity (rather than of dissimulation) and tend to correlate negatively with antisocial and criminal conduct, in children, adolescents, and adults.
7. The criminality scale, made up of the most diagnostic items of the EPQ, discriminates reliably between criminals and noncriminals.
8. Primary personality traits, such as impulsiveness, venturesomeness, risk taking, and empathy correlate, in predictable directions, with antisocial and criminal conduct.
9. These relationships are observed also in research where self-reports of antisocial behavior are the criterion of interest. Thus, personality-criminality correlations are not confined to legal definitions of crime or to samples of incarcerated criminals.
10. The personality-criminality correlations have cross-cultural validity, appearing in different countries and cultures with equal prominence.
11. Personality traits characteristic of antisocial and criminal behavior also correlate with behavior that is not criminal per se, but is regarded as antisocial – whether legal or illegal (e.g., smoking and drug use, respectively). Studies show high P and N scores among drug users, although E scores are

elevated only among drug users who have also been convicted of other crimes.

> Eysenck and Gudjonsson (1989) concluded that "…psychological factors and individual differences related to the personality are of central importance in relation to both the causes of crime and its control. This does not mean to say that other factors, such as sociological and economic ones, are not important. Indeed, in many instances they are. We believe that sociological theories are particularly relevant in relation to victimless crimes and less so in the case of victimful crimes.
>
> Psychological factors in criminality, we argue, relate to genetic and constitutional causes and to personality and other sources of individual differences. This does not mean that some people are destined to commit crimes. Criminal behavior as such is not innate. What is inherited are certain peculiarities of the brain and nervous system that interact with certain environmental factors and thereby increase the likelihood that a given person will act in a particular antisocial manner in a given situation." (p. 247)

There are, of course, considered and extensive critiques of the Eysenckian approach to the understanding of criminals and crime frequently revolving around the points below:

- Self-report measures of both crime/delinquency *as well as of* personality are open to dissimulation, lying, denial, exaggeration, etc. In addition, their interrelationships may appear inflated in research investigations due to common method variance.
- Caught criminals are unrepresentative of the population. Most criminals are, alas, not caught.
- Criminals and delinquents are far from homogeneous (e.g., murderers are very different psychologically from con men).
- Incarceration may affect personality, notably increase Neuroticism and reduce Extraversion. Prisoners may change as a function of their imprisonment and, thus, there may be reciprocal links between personality traits and crime-related variables.
- Being nonexperimental in nature, personality research is unable to demonstrate unequivocal causal relationships between personality traits and outcomes of interest.
- Trait personality theories tend to underemphasize social and environmental factors, although recent advances in the field of multivariate statistics have allowed for more realistic conceptualizations combining all of these influences and countering objections by sociologists and educationalists.

Eysenck's position developed and changed over time. He gradually placed more emphasis on the role of Psychoticism and began to admit more freely the range of the social causes of crime. Nevertheless, he continued advocating a central role for personality traits, a position that has consistently received support over the years including from rigorous longitudinal (birth-cohort) studies. Thus, Caspi et al. (1994) and Wright et al. (1999) found replications in the Eysenkian hypothesized personality – crime relationships across country, gender, race, and method. As suggested by the theory, poor social control (impulsivity, hyperactivity, etc.) predicted weak social bonds, adolescent delinquency, and later criminality. In a large meta-analytic review, Miller and Lynam (2001) found that personality traits relate to antisocial behavior, in general, and to crime, in particular, in both distal and proximal ways. Thus, personality predicts how people react to situations, how other people react to them, and indeed the situations they find themselves in. Miller and Lynam concluded that individual differences variables can help account for the observed stability of antisocial behavior over the life-span.

Conclusion

It is perhaps still true that no psychologist before or after Hans Eysenck attempted to show and explain how personality variables relate to such a wide array of everyday social behaviors from aesthetic appreciation to sport and school success to sport. Among his interests was how personality, in terms of his Giant Three system, related to all aspects of crime. His first book, published nearly 50 years ago, set out a bold and clear theory linking personality to various aspects of crime. The book has been quoted over 1,200 times. It provoked criticism but galvanized differential psychologists into testing the theory. Both updates and reviews constantly appear (Blonigen 2010;

Levine and Jackson 2004). To some extent the early work rather overemphasized the role of personality variables in crime, probably in reaction to a domination of sociological and psychological theories of crime. His second book, coauthored with Gudjonsson – a clinical and forensic psychologist – was a more measured and better-referenced book on the topic.

Clearly personality variables must and do play a part in all aspects of criminal behavior. The question is how much of the variance they do explain apart from other demographic and social environmental factors.

Recommended Reading and References

Barrett PT, Petrides KV, Eysenck SBG, Eysenck HJ (1998) The eysenck personality questionnaire: an examination of the factorial similarity of P, E, N, and L across 34 countries. Personal Individ Differ 25:805–819

Berlyne DE (1960) Conflict, arousal and curiosity. McGraw Hill, New York

Blonigen D (2010) Explaining the relationship between age and crime: contributions from the developmental literature on personality. Clin Psychol Rev 30:89–100

Broadhurst P (1959) The interaction of task difficulty and motivation: the Yerkes-Dodson law revived. Acta Psychol 16:321–328

Caspi A, Moffit T, Sliva P, Kreuger A et al (1994) Are some people crime prone? Criminology 32:163–195

Costa PT Jr, McCrae RR (1992) Four ways five factors are basic. Personal Individ Differ 13:653–665

Eysenck HJ (1967) The biological basis of personality. CC Thomas, Springfield, Reissued in 2006 by New York: Transaction Publishers

Eysenck HJ (1947) Dimensions of personality. Praeger, New York, Reissued in 1998 by New York: Transaction Publishers

Eysenck HJ (1964/1977) Crime and personality. Routledge and Kegan Paul, London

Eysenck HJ (1973) Historical introduction. In: Eysenck HJ (ed) Eysenck on extraversion. Wiley, London, pp 3–16

Eysenck HJ (1977) You and neurosis. Maurice Temple Smith, London

Eysenck HJ (1982) Personality, genetics and behaviour. Praeger, New York

Eysenck HJ (1992) Four ways five factors are not basic. Personal Individ Differ 13:667–673

Eysenck HJ, Eysenck SBG (1976) Psychoticism as a dimension of personality. Hodder and Stoughton, London

Eysenck HJ, Eysenck SBG (1981) Culture and personality abnormalities. In: Al-Issa I (ed) Culture and psychopathology. University Park Press, Maryland

Eysenck HJ, Gudjonsson GH (1989) The causes and cures of criminality. Plenum Press, New York

Feldman P (1993) The psychology of crime. Cambridge University Press, Cambridge

Furnham A, Thompson J (1991) Personality self-reported delinquency. Personal Individ Differ 12:585–593

Furnham A, Eysenck H, Saklofske D (2008) The eysenck personality measures: fifty years of scale development. In: Boyle G, Matthrews G, Saklofske D (eds) Handbook of personality testing. Sage, California, pp 199–218

Howarth E, Eysenck HJ (1968) Extraversion, arousal, and paired associate recall. J Exp Res Personal 3:114–116

Jackson CJ, Furnham A, Forde L, Cotter T (2000) The structure of the eysenck personality profiler. Br J Psychol 91:223–239

Levine SZ, Jackson CJ (2004) Eysenck's theory of crime revisited. Leg Criminol Psychol 9:135–152

Miller J, Lynam D (2001) Structural models of personality and their relation to antisocial behaviour. Criminology 29:765–798

Netter (2007) Modern research on the neurochemical basis of individual differences rooted in Hans Eysenck's theory. Paper at the ISSID conference, July

Petrides KV, Jackson C, Furnham A, Levine S (2003) Exploring issues of personality measurement and structure through the development of a short form of the eysenck profiler. J Personal Assess 81:271–280

Revelle W (1997) Extraversion and impulsivity: the lost dimension? In: Nyborg H (ed) The scientific study of human nature: tribute to Hans J. Eysenck at eighty. Pergamon, New York, pp 189–212

Royce JR, Powell A (1983) Theory of personality and individual differences: factors, systems and processes. Prentice Hall, Englewood Cliffs

Rushton JP, Bons TA, Hoor Y-M, Ando J, Irwin P, Vernon PA, Petrides KV, Barbaranelli C (2009) A general factor of personality from cross national twins and multi-trait multimethod data. Twin Res Hum Genet 12:356–365

Wright B, Caspi A, Moffitt T, Silva P (1999) Low self-control, social bonds and crime. Criminology 37:479–513

Yerkes RM, Dodson JR (1908) The relative strength of stimuli to rapidity of habit information. J Comp Neurol Psychol 18:459–482

F

Fact-Finding Interviews

► Evaluating Truthfulness: Interviewing and Credibility Assessment

False Confessions and Police Interrogation

Mark A. Costanzo[1] and Marina L. Costanzo[2]
[1]Department of Psychology, Claremont McKenna College, Claremont, CA, USA
[2]University of Colorado, Colorado Springs, CO, USA

Overview

False confessions are a serious problem in the USA criminal justice system, playing a role in about a quarter of all known wrongful convictions. False confessions can be differentiated based on whether they are voluntary or coerced and whether they are authentic or instrumental. Although the "third degree" was once common in the United States, police now rely on psychological techniques to extract confessions from criminal suspects. During interrogation, police strive to convince the suspect that the evidence against him is compelling and that admitting guilt is the only sensible course of action. When innocent suspects are subjected to psychologically powerful interrogation techniques, some will falsely confess. The vulnerabilities of individual suspects can also play a role in producing false confessions. Methods of reducing the risk of false confessions and identifying false confessions once they occur include mandatory electronic recording of interrogations, allowing expert testimony at trial, modernizing interrogation training, and reforming the interrogation process itself.

Introduction

A confession can be true, false, or ambiguous in veracity. We can often be sure that a criminal confession is true because the information revealed by the confessor is strongly supported by other evidences in the case. For example, there is often medical evidence, or DNA evidence, or fingerprints, or video recordings, or computer records, or multiple, credible eyewitnesses that confirm the account of the crime provided by a suspect in the interrogation room. True confessions disclose accurate details about the crime, often lead to the discovery of new evidence (e.g., where a body is buried), and reveal facts and information only the true perpetrator would know.

False confessions involve people admitting to crimes they did not commit or misrepresenting or exaggerating their role in a crime. We can know for certain that a confession is false when DNA evidence proves that someone else committed the crime, when the account of the crime given by the confessor clearly contradicts the physical

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

evidence in the case or when it can be shown that it was physically impossible for the confessor to have committed the crime (e.g., the suspect was in a distant location at the time of the crime). False confessions are generally not supported by other reliable evidences in the case. Usually, an account of a crime given by an innocent false confessor will not match the facts of the crime, will contain substantial errors, and will not lead police to new or missing evidence.

Some confessions are ambiguous in veracity – they can be proven neither true nor false. If there is a confession, but no other compelling evidence of guilt, it may be impossible to determine with certainty whether a confession is true. Confessions are a powerful form of incriminating evidence. Often, people who confess are convicted and sent to prison. In a study of *proven* false confessions, researchers found that 80 % of people who falsely confessed and pled "not guilty" were convicted at trial (Drizin and Leo 2004). Prisoners convicted of crimes may continue to proclaim their innocence from behind bars, but unless strong exonerating evidence is later uncovered or it turns out that it was physically impossible for the confessor to have committed the crime, we may never be able to verify whether a confessor is innocent or guilty.

A significant problem in distinguishing true from false confessions is created by the process of *contamination*. Too often, interrogators leak key details of the crime to the suspect being interrogated (Leo et al. 2009). Prior to or during the course of a long interrogation, police may disclose several critical elements of the crime (e.g., how and where the victim was killed). When a suspect later incorporates these key details into his or her confession, the result is a contaminated false confession that has the appearance of veracity. Because of the process of contamination, many proven false confessions contain details that only the actual perpetrator should know (Garrett 2008). As discussed below, without an electronic recording of the entire interrogation, it is usually impossible to assess whether contamination occurred during the interrogation process.

It is currently impossible to calculate the frequency or rate of false confessions. Police departments and state governments do not keep precise records of the number of interrogations conducted annually, nor do they keep detailed records of which confessions were proven to be true or false. Researchers have not been granted access to the records that do exist. And, of course, confessions that are ambiguous in veracity cannot be proven true or false. What we do know is that false confessions are far more frequent that was commonly supposed prior to the development of DNA identification technology. The Innocence Project, an organization that seeks to exonerate wrongfully convicted prisoners using DNA evidence, has found that false confessions were a critical piece of evidence in 25 % of wrongful convictions (Innocence Project 2012). It is critical to note that because the biological evidence (blood, semen, skin cells, hair, or saliva) which allows for DNA testing is only available in a minority of criminal cases, a large number of false confessions can never be *proven* false.

The Historical Context of Police Interrogations

Historically, physical and psychological torture have been the basic tools of interrogators, and, in many countries, torture remains a standard interrogation technique (Costanzo et al. 2010). In the United States, at least through the early 1930s, police used the "third degree," regularly threatening and physically abusing criminal suspects in an effort to extract confessions. As a result of the Wickersham Commission Report in 1931 (*National Commission on Law Enforcement and Observance* 1931), and a number of Supreme Court cases beginning with *Brown v. Mississippi* (1936), physical abuse of suspects began to decline in the 1930s, and, by the 1960s, such abuse appeared to be rare in the United States (Leo 2008).

In 1966, the US Supreme Court ruled that if a suspect is in police custody, police are required to issue several warnings prior to interrogation. Suspects must be informed of their right to remain silent, told that they have the right to appointed counsel and to have that counsel present during questioning. Suspects must be also be warned that anything they say can be used against them in court (*Miranda v. Arizona*, 1966). Somewhat surprisingly, *Miranda* warnings have proven to be only a weak safeguard against coercion. A large majority of suspects (estimates range from 78 to 96 %) relinquish their rights and therefore appear to consent to interrogation (Leo 2001). Police have devised several strategies to elicit waivers to the *Miranda* warnings. These include reading the *Miranda* warnings in a quick, confusing, or perfunctory manner; de-emphasizing or minimizing the significance of the *Miranda* warnings; obscuring the adversarial nature of the police-suspect interaction; and telling the suspect that the interrogator can help him only if he first waives his *Miranda* rights (Leo and White 1999).

In addition to the requirements that police "Mirandize" suspects and refrain from using physical force during interrogations, police are forbidden from making explicit threats of punishment or explicit promises of leniency.

To replace abusive practices, police developed more sophisticated psychological techniques. In 1942, Fred Inbau published the first edition of his seminal interrogation manual, *Lie Detection and Criminal Interrogation*. With John Reid and others, Inbau revised and expanded this training manual through several revisions. The most recent edition – *Criminal Interrogation and Confessions* by Inbau, Reid, Buckley, and Jayne – was published in 2001. The Chicago-based firm, *Reid and Associates*, transformed these manuals into a training program and, in 1974, began offering seminars to police and security professionals across North America. More than a quarter million police have now been trained in the Reid technique of interrogation. Undoubtedly, many more police officers have either read the training manual or have learned Reid tactics by observing or working with other interrogators. Research confirms that these tactics are frequently used by detectives in the United States (Leo 1996; Feld 2006). The use and effects of the Reid technique will be more fully discussed in the section on "The Process and Tactics of Interrogation."

Types of False Confessions

Several classification schemes have been proposed to differentiate between types of false confessions. Costanzo and Leo (2008) distinguished between four types of false confessions by crossing two dimensions: instrumental or authentic and voluntary or coerced. Instrumental confessions are those offered as a means to an end to achieve some goal. The goal of a suspect may simply be to end an extremely stressful interrogation. In contrast, authentic false confessions are the result of a confessor's genuine but false belief that he or she may have actually committed the crime. Voluntary false confessions are given freely by the confessor, while coerced false confessions are produced by psychological pressure (and occasionally physical pressure) from the interrogators.

Crossing the two dimensions yields four types of false confessions. *Instrumental – coerced* false confessions occur when, as a result of a lengthy or intense interrogation, suspects confess to crimes they know they did not commit. During the course of such an interrogation, the suspect becomes convinced that the only way to end the interrogation or to receive more lenient treatment is to agree with the interrogators that he or she committed a crime. This appears to be the most common type of false confession. *Instrumental – voluntary* false confessions occur when suspects knowingly implicate themselves in crimes they did not commit in an effort to achieve some goal. A member of a criminal gang or organization may falsely confess to a crime to protect someone higher up in the gang hierarchy. There have also been cases where parents have intentionally given false confessions to protect their children from being convicted of a crime. Indeed, people have voluntarily confessed to achieve a variety of peculiar goals – to achieve fame, to impress a girlfriend, to provide an alibi for one's whereabouts while having an affair, or to make police look incompetent (Gudjonsson 2003).

An *authentic-coerced* false confession occurs when, in response to a long or stressful interrogation, suspects begin to doubt their own memories and become convinced – at least temporarily – that they may have actually committed the crime. At the suggestion of the interrogators, the suspect may come to temporarily believe that he "blacked out" or "repressed" his memories of the crime. In a few exceptional cases, intensely coercive interrogations have caused innocent suspects to create false but vivid memories of crimes they did not commit (Wright 1994). Finally, an *authentic – voluntary* false confession occurs when someone who may or may not be a serious suspect confesses to a crime with little or no pressure from interrogators. In high-profile, well-publicized cases, police investigators sometimes receive multiple, unsolicited confessions from delusional or mentally ill people who appear to believe they have committed the crime.

Suspect Vulnerabilities to Interrogative Pressures

The majority of false confessions have come from adults with no significant mental problems. However, there are several suspect characteristics that raise the risk of false confessions. Risk factors attributable to the suspect include youth, mental impairment, personality characteristics, and temporary impairments.

Many of the best known examples of false confessions have involved juveniles or young suspects. In a large sample of proven false confessions, 63 % of innocent confessors were under the age of 25 (Drizin and Leo 2004). Juvenile suspects waive their Miranda rights at a rate about 12 % higher than adult suspects, making juveniles more vulnerable to false confessions (Grisso and Schwartz 2000). Research in developmental psychology demonstrates that children, adolescents, and even young adults differ from full adults in mental and social maturity. Cognitive neuroscientists have shown that key areas of the brain such as the limbic system and the prefrontal cortex continue to develop into the early 1920s (Steinberg 2007). Compared to adults, adolescents are more impulsive, more emotionally volatile, and less able to regulate their emotions. They tend to engage in riskier behavior and tend to make decisions based on immediate rewards and punishments rather than long-term consequences. This combination of attributes puts juveniles at special risk for false confessions (Redlich 2007).

Various forms of mental impairment including low intelligence or mental retardation appear to increase the risk of false confessions. About 22 % of known false confessors could be classified as mentally retarded (Drizin and Leo 2004). Mentally retarded individuals tend to be more acquiescent, more likely to agree with illogical questions, more likely to change their answers (especially in response to positive feedback), and more easily confused by the tactics of interrogators (Finlay and Lyons 2002). People with memory impairments are likely to distrust their own recollections and are therefore more likely to believe the account of events suggested by interrogators. Similarly, mental illness can make suspects more vulnerable to false confessions. Only 10 % of suspects who falsely confessed had been previously diagnosed with some form of mental illness, but others may have suffered from an undiagnosed impairment of some kind (Drizin and Leo 2004). Like mentally retarded suspects, mentally ill suspects may suffer from an inability to foresee the long-term consequences of admissions made in the interrogation room.

People vary in their ability to resist coercion and persuasion. Although there may be several aspects of personality that make people vulnerable to false admissions (e.g., need for approval or social anxiety), two traits – high suggestibility and compliance to authority – have been specifically investigated because of their relationship to false confessions (Gudjonsson 2003). People with these traits tend to be more uncertain about answers to questions asked by interrogators, tend to be more likely to trust interrogators' claims that they want to help, and tend to want to please interrogators. In addition to these stable personality traits, a variety of transitory states may increase the risk of false confessions. Sleep deprivation lowers our resistance to coercion, and drug or alcohol use makes us less able to evaluate the consequences of our actions. Grief and other strong emotional states have also been implicated in false confessions (Leo 2008). All of these temporary impairments create mental confusion and appear to weaken resistance to pressure.

The Process and Tactics of Interrogation

Interrogations in the United States tend to be based on the Reid technique and have been characterized as manipulative, guilt presumptive, and confrontational. Although interrogations vary in length, content, and intensity, all are designed to induce a guilty suspect to confess. Analyses of proven false confessions and experimental research demonstrate that the tactics described below sometimes induce innocent suspects to confess to crimes they did not commit (Russano et al. 2005; Drizin and Colgan 2004).

The interrogation process is designed to break the resistance of a suspect who knows he is guilty and to persuade him that the benefits of confessing outweigh the costs. The process of interrogation in the United States can be reduced to four broad tactics (Costanzo and Leo 2008). First, there is almost complete *control* of the process by the interrogators. Suspects are typically interrogated in a cramped, sparsely furnished room. The physical characteristics of the room, and more importantly, the nature, length and pacing of the conversation are under the control of the interrogators. The goal is to remove the psychological comfort of familiar surroundings and to communicate to the suspect that he has lost control of the situation. This loss of control disorients the suspect, raises anxiety, and creates a desire to escape the interrogation room. The second broad tactic that sets the stage for a potent interrogation is *social isolation*. Whenever possible, suspects are interrogated alone. Isolation deprives the suspect of emotional support. The presence of a friend or ally might bolster the suspect's resistance to persuasion and might lead to additional challenges to the interrogator's version of events. The combination of control and social isolation enables interrogators to create the social reality of the situation and to reshape the decision-making process of the suspect.

The two remaining tactics, certainty of guilt and minimization of culpability, are used to convince the suspect that offering an admission of guilt is their best available option. *Certainty of guilt* refers to the message – communicated in a variety of ways – that the suspect committed the crime and that the evidence against him is irrefutable. Although an innocent suspect will respond to this accusation with denials, interrogators are trained to challenge, cut off, and dismiss all denials. To convince the suspect that denials are futile and that certainty of guilt is justified, police are permitted to use a variety of *false evidence ploys*. Put more directly, police in the United States are permitted to lie about evidence against the suspect. Examples of false evidence ploys include untrue claims that others have implicated the suspect ("eyewitnesses saw you do it," or "others have told us you did it") and false claims that various forms of physical evidence (fingerprints, DNA, videotape from a security camera) clearly point to the suspect as the person who committed the crime. If the suspect offers an alibi, the interrogator may attack it as inconsistent, implausible, contradicted by the evidence, or simply impossible. These ploys often cause the disoriented suspect to conclude that he bears the burden of proving his innocence, that conviction is inescapable, and that further denials are futile.

Certainty of guilt is paired with *minimization of culpability* to move the suspect from denial to admission. Minimization of culpability involves offering "face-saving" justifications or excuses for the crime. For example, it might be suggested to a murder suspect that he killed the victim by accident or in self-defense. It might be suggested to someone suspected of theft that the money was stolen to pay for an important family expense or suggested to someone suspected of rape that he engaged only in consensual sex. These minimizing scenarios serve to morally, psychologically, and legally justify, downplay, or excuse a suspect's actions. Supplying a minimizing or exculpatory reason for committing a crime clearly implies that the offense may not be that serious and that anyone making a judgment about the act (e.g., a prosecutor, judge, or jury) would be likely to recommend lenient treatment. The interrogator suggests that the suspect will be charged with a less serious crime and receive a shorter prison sentence if he accepts responsibility, admits guilt, and expresses remorse (Leo et al. 2009). Conversely, continued denial will result in a more serious charge and a longer prison sentence.

The cumulative effect of these tactics is to persuade the suspect that he will almost certainly be arrested, convicted, and punished. Only by accepting a minimizing scenario can he improve his otherwise hopeless situation. Police often present an admission as an "opportunity" for the suspect to "tell his side of the story." Further, it is often presented as a time-limited opportunity that will vanish once the interrogation ends. During the interrogation, police may misrepresent themselves as allies instead of adversaries and suggest that they can and will help the suspect minimize the consequences of his crime. The pretense that interrogators are the allies of the suspect is the foundational lie of the interrogation process. Interrogators may explicitly tell a suspect that they can help him by how they write up their report, by what they tell the prosecutor prior to charges being filed, or by how they testify before a judge or jury at trial. Interrogators may even offer to help a suspect in ways that extend beyond the criminal justice system, such as arranging for counseling or assistance from social services. Of course, the catch is that these benefits can only be realized if the suspect complies with the interrogators' demands for an admission.

Reducing the Risk of False Confessions

Perhaps the most important goal of research on interrogations and false confessions is to suggest "best practice" reforms that reduce the number of false confessions while increasing the number of true confessions. Scholars have proposed several such reforms based on research findings. Some reforms aim to prevent false confessions from occurring, while others increase our ability to detect false confessions after they have occurred.

The most fundamental and frequently recommended reform is to require video recording of interrogations (Kassin and Gudjonsson 2004). Electronic recording of interrogations is required in some states and common in felony cases. Mandatory recording of interrogations creates an objective, reviewable record of what occurred in the interrogation room. Jurors, judges, lawyers, and experts are able to see and hear for themselves not only what was said but also how it was said. However, to provide an optimally useful record, the interrogation must be recorded in its entirety, and both the suspect and the interrogator should be visible. Recording the entire interrogation ensures that we do not rely on partial, selective, or biased records of what transpired. To assess whether a confession is coerced or contaminated, we need to be able to examine the entire interaction that preceded a confession. The requirement that both the suspect and the interrogator be visible follows directly from research demonstrating that only an "equal-focus" camera angle allows viewers to evaluate both the suspect's confession and the pressures exerted by interrogators (Lassiter and Geers 2004).

A second important reform is the use of social science expert testimony in cases involving a disputed interrogation or confession (Costanzo and Leo 2008). Such testimony is routinely offered in many states, and there is a substantial body of scientific research that experts can rely upon. When a disputed confession is introduced at trial, jurors will want to know why a possibly innocent person could be made to confess to a serious crime. Expert testimony can displace common misconceptions – for example, the belief that an innocent person will not falsely confess unless he is physically tortured or mentally ill. Experts educate the jury about the reality of false confessions, the psychological dynamics of the interrogation process, and the factors that increase the probability of false confessions. The purpose of expert testimony at trial is to provide an overview of relevant research and to assist the jury in making a fully informed decision about what weight to place on the defendant's confession (Costanzo et al. 2010).

A third set of reforms focus on modifying the training of interrogators to give more emphasis to the problem of false confessions. When asked about the possibility of false confessions, a common response of those who train interrogators is, "we don't interrogate innocent people" (Kassin 2005). In theory, police screen out innocent suspects through initial interviews and evidence gathering. In actual practice, however, the list of proven false confessions is replete with cases in which innocent suspects were subjected to highly coercive interrogation techniques based on little more than intuition or a misreading of a suspect's behavior as suspicious or deceptive.

In particular, interrogation training needs to make far more modest claims about the ability of police to correctly interpret behavioral cues to deception. Research shows that it is extremely difficult for observers to make reliable judgments about whether or not someone is lying (Vrij 2008). Further, research indicates that many of the behavioral cues used by interrogation trainers do not differentiate liars from truth tellers. The general finding is that both untrained observers and people with relevant professional training (e.g., police and polygraph examiners) perform only slightly better than chance. Where chance accuracy is 50 %, average accuracy is only about 54 % (Bond and DePaulo 2006). Although training to improve accuracy does not reliably improve the ability to detect lies, it does have effects: after training, people are more confident about their judgments, and they list more reasons for their judgments (Costanzo 1992; Kassin and Fong 1999; Kassin et al. 2005). When faulty cues to deception are combined with overconfidence by interrogators, innocent suspects can be subjected to enormous pressure. Some will falsely confess.

The most controversial, but potentially most effective set of reforms focus on the interrogation process itself. Because interrogations culminating in proven false confessions are significantly longer than average, some researchers have suggested time limits on interrogations (e.g., no more than 3 h). Longer interrogations appear to increase the risk of false confessions by producing stress and fatigue and thus impairing a suspect's ability and motivation to resist police pressures. Other limits on the interrogation process have been proposed. Following the lead of the United Kingdom, the United States could prohibit police use of deception and trickery in the interrogation process. Training manuals continue to insist that police lying about evidence does not produce false confessions. Yet many false confessions are the product of interrogations in which police lied to suspects, used false evidence ploys, and pretended to possess damning physical evidence against the suspect (Ofshe and Leo 1997; Drizin and Leo 2004).

As noted earlier, the dominant form of interrogation used in the United States is strongly influenced by the Reid technique, which presumes that the person being interrogated is guilty and will only respond to a confrontational and deceptive style. An alternative to this approach was developed in England and Wales in response to several high-profile false confessions that led to wrongful convictions. The PEACE model of interrogation (an acronym for Preparation and planning, Engage and explain, Account and clarification, Closure, and Evaluation) has been in use since the mid-1990s and has been adapted for use in several other countries (Gudjonsson and Pearse 2011). This alternative model emphasizes careful preparation for the interrogation and thorough knowledge of the suspect and the crime. The process of interrogation stresses finding the truth rather than obtaining a confession. Lying to suspects, false evidence ploys, manipulation, and strong pressure are avoided. All interrogations are electronically recorded. Use of the PEACE model appears to dramatically reduce the rate of false confessions while still producing a high rate of true confessions (Bull and Soukara 2010).

Careful analysis of proven false confessions and a growing body of experimental studies provide us with new insights on how interrogations go wrong. We need to translate these insights into reforms that increase the likelihood of producing true confessions while reducing the likelihood of false confessions.

Related Entries

- ▶ Causes of Wrongful Convictions
- ▶ Compliance and Corporate Crime Control
- ▶ Methodological Issues in Evaluating Police Performance
- ▶ Investigative Psychology
- ▶ Law of Police Interrogation
- ▶ Measuring Wrongful Convictions
- ▶ Police Lying and Deception
- ▶ Predictors of Successful Investigations

Recommended Reading and References

Bond CF, DePaulo BM (2006) Accuracy of deception judgments. Pers Soc Psychol Rev 10:214–234

Brown v. Mississippi, 297 U.S. 278 (1936)

Bull R, Soukara S (2010) Four studies of what really happens in police interviews. In: Lassiter GD, Meissner CA (eds) Police interrogations and false confessions. American Psychological Association, New York, pp 81–95

Costanzo M (1992) Training students to decode verbal and nonverbal behavior: effects on confidence and performance. J Educational Psychol 84:303–313

Costanzo MA, Gerrity E (2009) The effects and effectiveness of using torture as an interrogation device: using research to inform the policy debate. Soc Issues Policy Rev 3:179–210

Costanzo M, Leo R (2008) Research findings and expert testimony on police interrogations and confessions to crimes. In: Costanzo M, Krauss D, Pezdek K (eds) Expert psychological testimony for the courts. Erlbaum, Mahwah

Costanzo MA, Shaked-Schroer N, Vinson K (2010) Juror beliefs about police interrogations, false confessions, and expert testimony. J Empir Leg Stud 7:231–247

Drizin SA, Colgan BA (2004) Tales from the juvenile confessions front. In: Lassiter GD (ed) Interrogations, confessions, and entrapment. Kluwer, New York, pp 127–162

Drizin SA, Leo RA (2004) The problem of false confessions in the post-DNA world. North Carol Law Rev 82:891–1007

Feld B (2006) Juveniles competence to assess Miranda rights: an empirical study of policy and practice. Minn Law Rev 91:26–100

Finlay W, Lyons E (2002) Acquiescence in interviews with people who have mental retardation. Ment Retard 40:14–29

Garrett B (2008) Judging innocence. Columbia Law Rev 108:55–142

Grisso T, Schwartz R (2000) Youth on trial: a developmental perspective on juvenile justice. University of Chicago Press, Chicago

Gudjonsson GH (2003) The science of interrogations and confessions: a handbook. Wiley, Chichester

Gudjonsson GH, Pearse J (2011) Suspect interviews and false confessions. Curr Dir Psychol Sci 20:33–37

Inbau FE, Reid JE, Buckley JP, Jayne (2001) Criminal interrogation and confessions, 4th edn. Aspen, Gaithersburg

Innocence Project (2012) http://www.innocenceproject.org/understand/False-Confessions.php

Kassin SM (2005) On the psychology of confessions: does innocence put innocents at risk? Am Psychol 60:215–228

Kassin S, Fong CT (1999) I'm innocent! Effects of training on judgments of truth and deception in the interrogation room. Law and Human Behav 23: 499–516

Kassin S, Gudjonsson GH (2004) The psychology of confessions: a review of the literature and issues. Psychol Sci Publ Interest 5:33–67

Kassin S, Meissner C, Norwick R (2005) "I'd know a false confession if I saw one". A comparative study of college students and police investigators. Law Human Behav 29:211–221

Lassiter GD, Geers AL (2004) Evaluation of confession evidence: effects of presentation format. In: Lassiter GD (ed) Interrogations, confessions, and entrapment. Kluwer, New York, pp 197–214

Leo RA (1996) Inside the interrogation room. J Crim Law Criminol 86:266–303

Leo RA (2001) Questioning the relevance of Miranda in the twenty-first century. Mich Law Rev 99: 1000–1029

Leo RA (2008) Police interrogation and American justice. Harvard University Press, Cambridge

Leo RA, White WS (1999) Adapting to Miranda: modern interrogators' strategies for dealing with the obstacles posed by Miranda. Minn Law Rev 84:397–472

Leo RA, Costanzo M, Shaked-Schroer N (2009) Psychological and cultural aspects of interrogation and false confessions: using research to inform legal decision making. In: Krauss D, Lieberman J (eds) Psychological expertise in court: psychology in the courtroom, vol 2. Ashgate, London, pp 25–56

Miranda v. Arizona, 384 U.S. 436 (1966)

National Commission on Law Observance and Enforcement (1931) Report on lawlessness in law-enforcement. Government Printing Office, Washington, DC, Vol. 11

Ofshe RJ, Leo RA (1997) The decision to confess falsely: rational choice and irrational action. Denver Univ Law Rev 74:979–1122

Redlich AD (2007) Double jeopardy in the interrogation room: young age and mental illness. Am Psychol 62:609–611

Russano MB, Meissner CA, Narchet FM, Kassin SM (2005) Investigating true and false confessions within a novel experimental paradigm. Psychol Sci 16:481–486

Steinberg L (2007) Risk taking in adolescence: new perspectives from brain and behavioral science. Curr Dir Psychol Sci 16:55–59

Vrij A (2008) Detecting lies and deceit: pitfalls and opportunities. Wiley, Chichester

Wright L (1994) Remembering Satan: a case of recovered memory and the shattering of an American family. Alfred A. Knopf, New York

False Memories

Eryn J. Newman[1], Steven J. Frenda[2] and Elizabeth F. Loftus[2]
[1]Victoria University of Wellington, Wellington, New Zealand
[2]University of California, Irvine, Irvine, CA, USA

Overview

Memory is indispensable to the criminal justice system. Countless criminal cases are built around the things victims and eyewitnesses remember: details, faces, and circumstances of events that occurred weeks, months, and sometimes even years earlier. In an ideal world, a person's memory would work something like a video camera – capturing details in high definition and saving them to a memory card so they can later be retrieved and trusted as reliable. As it turns out, memory is less like a video recording and more like an active reconstruction of the past. It can be altered by even the subtlest of influences, and in some cases, even memories held with great confidence can be completely wrong. Errors in people's memories can be inconsequential, but they can also cause profound complications, particularly in the legal system where a defendant's guilt or innocence can depend on the precise details of another person's recollection. This entry explores the small and large ways that memory can depart from reality – how people can misremember details about an event and how they can come to remember whole events that never occurred.

Misremembering Details

In early 2012, Florida resident George Zimmerman called 911 from his cell phone and reported that a young Black man (later identified as Trayvon Martin) was behaving suspiciously inside the gated community where Zimmerman lived. With police en route, Zimmerman confronted Martin. An altercation ensued and ended with Zimmerman fatally shooting Martin in the chest at close range. When the police arrived, Zimmerman appeared injured and claimed that Martin attacked him and that he, Zimmerman, shot in self-defense (Kovaleski 2012).

Numerous eyewitnesses came forward with their accounts, but strangely, several of their stories seemed to transform over time. For example, a few days after the shooting, a woman recounted to the police that she saw two men running and then brawling in the street. A few weeks later, she said she saw just one man running, but couldn't identify him because she was not wearing her contact lenses. Another witness initially said she saw two men on the ground after the shooting, but could not tell which man was on top. Days later, after seeing Zimmerman on television and learning new details about the case, she said she remembered that Zimmerman was the one on top. Another witness said he saw Martin pinning down Zimmerman and punching him while Zimmerman called for help, only to later claim that he wasn't sure he saw any punches or heard any calls for help. Finally, a man who saw Zimmerman after the shooting first described him as looking bloody and dazed and later said he seemed calm and collected.

The shooting of Trayvon Martin and Zimmerman's subsequent arrest was covered incessantly by the American news media in the days and weeks following the event, so it is not surprising that witnesses to the incident were exposed to a bevy of reactions, opinions, imaginings, and speculations. But the question remains, how is it possible that two starkly different accounts can emerge from the same eyewitness? One possibility is that the eyewitness is not being truthful in one version of events. Another possibility is that some kind of post-event information *changed* the witness' memory. Psychologists have now studied this phenomenon, known as the misinformation effect, for several decades (see Loftus (2005) for a review).

In trying to understand how people's memories can become contaminated and distorted, researchers developed a simple experimental paradigm mimicking the sequence of events that could happen in an actual case involving eyewitnesses. First, research subjects are shown some type of stimulus (e.g., a video or a photo slideshow of a simulated crime). Next, they are exposed to some misleading information about what they saw, usually embedded in a lot of true information. Finally, their memories for the original source material are tested. In one study, subjects were shown a series of photographs depicting a man interacting with a woman on a city street, stealing her wallet, and hiding it in his jacket pocket (Zhu et al. 2010a). Later, subjects read narratives that described the photographs. For some of the subjects, however, the narrative contained bits of inaccurate information. These subjects were told, for instance, that the man hid the stolen wallet in his pants pocket. All subjects were then asked a series of questions about the original event they saw (e.g., Where did the man hide the wallet?). In hundreds, perhaps thousands of studies like this one, subjects who are exposed to inaccurate information frequently incorporate these inaccuracies into their memories of the original event, rendering them much less accurate than witnesses who were never exposed to misinformation.

When Is Misinformation Most Effective?

In any one specific instance, it can be very difficult to disentangle whether memories are based on true experiences and when they have been contaminated by erroneous information. However, knowing about factors that tend to promote or inhibit memory distortion might help to illuminate whether someone's recollection has been altered by post-event information. With this in mind, psychologists have worked at identifying circumstances that appear to facilitate the distortion of people's memories. For instance, studies have clearly shown that a person's memory is more susceptible to misinformation effects when more time has passed since the original event occurred (e.g., Loftus et al. 1978). This may be because memories fade over time; following longer retention intervals, misleading information may be less recognizable as false and more easily incorporated into the original memory. Other research has shown that misleading information has a stronger influence on memory when people are in certain states (e.g., under hypnosis; Scoboria et al. 2002) or when people believe they are in certain states (e.g., when they think they are intoxicated; Assefi and Garry 2002).

Another factor that can influence the power of misleading information on memory is the source of the information. Studies have shown that people incorporate false information more readily into memory when they perceive the source of the information to be trustworthy. For example, in a series of experiments, Unkelbach and Stahl (2009) showed that people were more likely to identify false statements as true when they heard a trustworthy source make those statements earlier in the study. Misinformation is also more likely to change a person's memory when they are exposed to it repeatedly. For instance, one study investigating the effects of hearing other witness statements (Foster et al. 2012) showed that people are more likely to incorporate other people's errors into memory when they hear the errors repeated multiple times. Interestingly, the repetition effect occurred whether the misinformation came from multiple witnesses or from just one witness multiple times. Even our own actions with misinformation can affect its strength. For example, people are more misled when they elaborate on misinformation by considering what it looked like or how it contributed to the overall meaning of an event (Drivdahl and Zaragoza 2001; Zaragoza et al. 2011).

An Ounce of Prevention

Can anything be done to prevent misleading information from invading memory? One approach has been to study the effects of warnings. That is, when people are made aware that they might be exposed to inaccurate information, are they better able to preserve the accuracy of their memories? Results from studies that have investigated this question have been mixed. If warnings are given before misinformation is presented, then they can help people detect errors in post-event information and ward off misleading details (e.g., Butler et al. 2009). But other research shows that warnings given after the misinformation have no effect at all (Greene et al. 1982). That is, after people have been exposed to misinformation, it is difficult for them to unwind the effects of post-event suggestion. Still other studies have shown that under some circumstances (e.g., when subjects are given a very specific warning straight after encountering misinformation), warnings following misinformation can help subjects maintain more accurate memories (Eakin et al. 2003).

Instead of alerting people to the presence of misinformation, other research has focused on developing interview procedures that would help preserve the integrity of a witness' memory. The *Cognitive Interview* (CI) is one such procedure. In the CI, people participate in an initial free recall phase and make use of contextual cues, temporal reordering of events, and recounting events from the perspective of other people. Interviewers are encouraged to avoid suggestive questioning, to discourage guessing, and to develop rapport with the interviewee. One study found that compared to a free-response control interview, subjects who were given a CI produced more correct details about an event they had seen and were better able to maintain accurate memories in the face of suggestive questioning (Memon et al. 2009). However, the CI had an effect only when it came *before* the misinformation. Of course in the real world, it is hard to know when misinformation is going to invade memory. Thus, to best protect memory, both warnings and interview approaches like the CI should be delivered soon after the event in question.

Taken together, these studies demonstrate that some precautions can be taken to prevent the contamination of a person's memory. However, it seems as though the timing of these interventions is crucial, and once a person has been exposed to misinformation (particularly without being aware that the information might be wrong), the damage may already be done. Also of note, even under conditions that were designed to protect people from misinformation – when people are warned that they might encounter inaccurate details – many subjects still showed evidence of memory distortion.

Are Some People Especially Susceptible to Memory Distortion?

Are certain types of people more vulnerable to misinformation effects than others? While misinformation effects have been shown in a wide variety of subject samples, there is evidence that some groups of people are especially susceptible. For example, young children and the elderly are more likely to have their memories contaminated by misleading information than adolescents and adults (see Davis and Loftus (2005) for a review). One explanation for this finding is that these groups are less skilled at tracking and making accurate judgments about various sources of information – for example, discriminating between details that were actually seen firsthand and those that were described by another witness.

Other studies (Zhu et al. 2010a, b) have shown that people who score higher on measures of intelligence, perceptual ability, working memory capacity, and facial recognition ability are especially resistant to misinformation effects. And certain personality factors, such as fear of negative evaluation and cooperativeness, are associated with susceptibility to false memories, particularly in individuals with low cognitive ability. An interpretation of these results is that when people are high in cognitive ability, personality variables are less useful in predicting whether misinformation is likely to creep in to memory. However, for a person who is low in cognitive ability, traits like fear of negative evaluation and cooperativeness may motivate them to pay closer attention to the post-event information in an effort to buttress the accuracy of their memories, only to fall victim to misleading details. While these studies imply that some individuals may be more suggestible than others, a broad survey of the literature reveals that no one is immune to memory distortion. That recollections can be molded and changed by suggestion and misinformation is simply a fact of human memory. This body of work demonstrates how true memories can be distorted and altered, but other research suggests that memory can be molded in more dramatic ways and people can come to remember entire events that never took place at all.

False Memory for Entire Events

During the 1980s, while seeking therapy for depression, Elizabeth Gale came to remember that she had experienced horrific acts of satanic ritual abuse. But years later, Gale realized that the memories she recovered in therapy were actually entirely false, and in 2004, she won a malpractice lawsuit against the mental health professionals who led her to recover the traumatic memories of horrific acts that never happened (Napolitano 2004).

Unfortunately, Gale's case was just one among a number of other similar cases. During the 1980s and early 1990s, other cases emerged where patients had gone into therapy and come out with memories of traumatic events from their past, only to later learn that the memories they had recovered were in fact false. Cases like these often had devastating consequences. Family members were accused of horrific acts. In some instances people ended up in court and behind bars based on a memory recovered during therapy that turned out to be not only inaccurate but also completely false (see Loftus and Ketcham 1994).

How could people come to remember entire events that never actually occurred? A common thread in many of these cases was that the patients were participating in Recovered Memory Therapy (RMT). Clinicians practicing this therapy – popular during the 1980s – believed that their patients were harboring repressed memories of trauma (memories that had been hidden away from conscious awareness). The goal of RMT was to help patients uncover these buried memories. With the best intentions, therapists encouraged patients to pursue buried traumatic memories using repeated retrieval attempts, guided imagery, imagination, dream interpretation, and in some cases hypnosis. It was already well established in the cognitive psychological literature that some of these techniques such as repetition, imagination, and imagery could promote false memories. But in those studies people misremembered words, shapes, and stories, not entire autobiographical events. It wasn't until the 1990s that researchers developed a paradigm to examine how the same techniques used in RMT could lead people to develop wholly false memories for an experience that never actually took place.

In this paradigm, subjects read four descriptions about childhood events and repeatedly try to recall them. Three of the descriptions are of true events provided by parents, but one of the descriptions is created by the researchers (e.g., getting lost in the mall) and documents an event that never occurred. The false event is altered for each subject so that it is filled with personal details and looks just like the real events provided by the parents. For example, in one study, a subject read that "You, your mom, Tien, and Tuan all went to the Bremerton K-Mart. You must have been 5-years old at the time. . .and somehow lost your way in the store. Tien found you crying to an elderly Chinese woman. . ." (Loftus and Pickrell 1995, p. 721). The key question is whether a false suggestion like the one described above combined with RMT techniques would lead people to remember an entire event that never occurred.

Over a number of studies, using a variety of RMT-like techniques that ranged from guided imagery, repeated retrieval attempts, and imagination, on average 36 % of people came to remember a false event that never happened (Newman and Garry 2013). Moreover, people falsely remembered all kinds of events; some were common, relatively benign experiences like getting lost in a mall. Others were less common but still relatively unemotional events like spilling a punch bowl at a wedding. Still others were traumatic and emotional, such as being saved by a lifeguard after nearly drowning or even being viciously attacked by an animal (for a review of these studies, see Garry and Hayne 2006). In other work, researchers have planted memories for events that are impossible like having a skin sample taken as a child – a national medical test that never took place.

Moreover, these memories were filled with details – people, places, and sensory and temporal information – that is, they contained many of the characteristics one would expect to see in a memory for an event that really did happen. In recalling the punch incident, one subject said, "…I definitely think it's a friend of my mother's for some reason, and the people I spilled punch on…I just picture him getting up and being kind of irritated or mad, then the woman, I see her in a light coloured dress that has like flowers on it, I think I see flowers on it and stuff, and I see like a big red punch thing down the front of them…" (Hyman and Billings 1998, p. 10). This subject's false memory looks just like a true memory.

In experiments like these, researchers know which memories are true and which are false, but in the real world, it is much more difficult to decipher which memories represent a veridical record of the past (see Bernstein and Loftus (2009a) for a review). Even some biological measures reveal very few differences between true and false memories. In one study, people who believed they had been abducted by aliens had a similar physiological stress reaction to patients – who had current PTSD – when they recalled their (highly improbable) traumatic memories (McNally et al. 2004). Taken together, these studies fit with the earlier work using words, shapes, and stories and suggest that RMT techniques not only lead people to agree that an event might have happened, but they can lead people to develop memories layered with detail, emotion, and even physiological responses that make them feel true.

Routes to a False Memory

In these memory implantation studies, subjects typically read through narratives ostensibly provided by family members before engaging in certain RMT techniques. One might wonder whether that is the only route to a false memory. Put another way, could other kinds of suggestion or therapeutic practices promote vividly detailed memories for events that never happened? The answer is yes. One does not need to read a description of a childhood event in order to develop a false memory. Reviewing family photos, receiving false feedback, and even imagining or paraphrasing how an event might have occurred can have similar consequences.

Photos. Sometimes clients are invited to review family photo albums to help cue their memory during therapy. But while this retrieval cue can help people remember real events, it can also help people build memories for events that never happened. For example, one group of researchers asked people to examine old family photos and try to recall events depicted in each photo (Wade et al. 2002). Just like the studies described above, one of the events was false and a photo was doctored so it looked like the subject experienced an event that they did not (e.g., it showed the subject as a child in a hot air balloon with their parents). After three interviews, and repeated recall attempts, 50 % of subjects remembered the false event. But people rarely encounter photos like these; family photos are usually a reliable record of events from our past, they are not doctored to mislead us. Would real undoctored childhood photos lead people to remember events that never happened? The answer is yes: simply reviewing an old class photo can lead people to remember doing things they never did in their childhood, such as hiding a slime toy in a teacher's desk (Lindsay et al. 2004). Other work suggests that these effects extend to more recent events too. When researchers doctored video footage to make it look like someone was cheating in an experiment, subjects were more likely to sign a form corroborating that cheating behavior, even though they witnessed the event differently themselves only 5–7 h earlier (Wade et al. 2010). Together these findings suggest that photographs and videos – even old class photos that seem harmless – can facilitate the recall of events that never took place.

Feedback. One especially controversial source of feedback were self-help books that offered people advice about buried memories of trauma. Some of these books offered a list of symptoms that authors suggested would help people assess whether they were harboring a repressed memory. Readers were informed that they might have a buried traumatic memory if they displayed a certain set of these symptoms. What are the consequences of this kind of feedback? Could feedback like this from a trusted source lead people to develop fully fledged false memories? Subsequent research examining the power of feedback suggests that the answer is yes. In one line of work, researchers demonstrated the influence of feedback using a simple three-stage procedure. In initial session, people completed a food history survey. In session 2, people received a personalized food experiences profile that provided feedback on a critical childhood event, e.g., that the subject got sick eating strawberry ice cream – in fact there was no evidence that this event occurred. In the final session, the experimenters measured the subjects' belief in having experienced the critical event by asking them to complete the food history survey for a second time. Receiving the false feedback led people to believe they had a bad experience as a child: over a number of studies, people remembered getting sick eating a variety of foods including strawberry ice cream, dill pickles, and egg salad sandwiches (for a review of these studies, see Bernstein and Loftus 2009b). Feedback like this can not only change people's memories and beliefs about the past, but it can change their behavior too. Subjects who received the false feedback about getting sick eating egg salad sandwiches were more likely to avoid eating them when they encountered that food later on (Geraerts et al. 2008). Moreover, other kinds of feedback such as dream interpretation can have similar consequences. In one study, some subjects' dreams were interpreted to mean that they experienced an upsetting event as a child (e.g., being lost in a public space), while others received no interpretation (Mazzoni et al. 1999). Those who received feedback via dream interpretation were more confident that the upsetting event happened to them. Taken together, this research suggests that feedback and interpretations offered by self-help books or even a well-intentioned therapist could be dangerous and unintentionally plant false memories.

Imagination. Even without a therapist, or photos, certain cognitive acts can lead people to build false memories. A range of studies show that spending time imagining a childhood event can lead people to become more confident that the event really did happen to them. Simply explaining or paraphrasing how a childhood event could have occurred can lead to a similar boost in confidence (see Garry and Polaschek (2000) for a review). All these behaviors likely lead subjects to generate a rich mental representation of the hypothetical event, adding sensory details and characteristics that make it feel like a real memory. Imagination can also lead people to falsely remember more recent events – even ones that are bizarre. In one study, subjects went on a campus walk and performed a series of actions. Later they imagined performing some actions, but not others. Imagination led people to misremember – only 2 weeks later – having performed actions they did not, even when the events were highly unusual (e.g., proposing marriage to a Pepsi machine) (Seamon et al. 2006).

Although generated as a consequence of suggestive retrieval techniques, false memories feel like true memories and they have real consequences. In cases similar to Elizabeth Gale, real patients pursued real criminal cases and real families were broken apart. That is, outside of the laboratory when it is more difficult for us to tell what is true and what is not, false memories wield similar power to true memories. So how do people confuse false memories that they generate themselves with ones that are the product of real experience?

How Do People Confuse False Memories with Real Experience?

People don't walk around the world mystified about what is real and what is not. Indeed, people usually remember correctly that they drove a car to work, that they ate breakfast, and that they had a conversation with a friend on the phone. But as described earlier, memory does not work like a camera capturing and replaying everything in high definition. Instead, when people retrieve information, they make a decision using a variety of characteristics to decide whether a particular mental experience is a memory, something someone else told them (a suggestion or piece of misinformation) or something they simply imagined (see Lindsay 2008). Sometimes these memory classifications are easy and people can quickly and accurately determine the origin of mental events. But sometimes these decisions are hard; mental events come to mind with characteristics that make them seem like a real memory. And techniques that encourage elaboration about suggested false events like imagination and guided imagery can artificially boost the characteristics people rely on to accurately classify memories – leaving people especially vulnerable to errors.

Moreover, when making these memory classifications, people can be swayed by where information came from. Recall that people are more likely to be misled by details provided by someone who is trustworthy. One explanation for this effect is that people are less inclined to scrutinize information that comes to mind when it was suggested by a reliable source. Thus, when inaccurate suggestions come from someone trustworthy, it makes it easier for false information to invade memory.

Summary

Although many have probably marvelled at the information and detail that can be retrieved from memory, the research reviewed here suggests that it is not a perfect retelling of the past. In the context of the criminal justice system, the fallibility of human memory is especially worrisome. Errors – small and large – can influence the outcome of a case and inevitably an individual's life. Indeed, people are especially swayed by an eyewitness or victim that describes their memory of an event in the courtroom.

In light of this, what tools are available to diagnose the accuracy of someone's memory? Some researchers have examined the characteristics of true and false memories by comparing the content and detail in true and false memory reports from experiments. Others have considered individual difference variables that might help determine which people are more vulnerable to memory errors (see Bernstein and Loftus (2009a) for a review). But more recently, scientists have investigated whether measurements of brain activity can help distinguish true from false memories. For instance, in one recent study (Stark et al. 2010), subjects viewed a series of photographs and then heard an audio narrative about the photos that contained pieces of misleading information. Later, they were given a memory test while in an MRI scanner. Results indicated that true and false memories showed similar activation patterns, but that true memories (which came from visual information) were associated with more activation in the visual cortex, while false memories (which came from auditory information) were associated with more activation in the auditory cortex. The results suggest that when misinformation is delivered in a different sensory modality than the true memory information, there may be detectable differences in brain activation patterns. However, we are still a long way away from being able to take a single memory report and use some physiological measure to determine whether it reflects an authentic experience or a false one. The research on these tools for trying to accomplish this is still in its infancy.

A different approach to the problem of memory in the courtroom has been to implement procedures and policies that bring many of the issues described in this article to the attention of jurors when cases they are hearing involve eyewitness or victim memory. For example, in a 2011 ruling, the New Jersey Supreme Court issued new rules for dealing with cases that include eyewitness evidence. Moving forward in New Jersey, when defendants present evidence that an identification may have been corrupted, the judge must hold a special hearing to consider the evidence and must give detailed explanations to the jurors about the relevant factors that may have interfered with the accuracy of the eyewitness' memory (Weiser 2011).

People's memories will remain a crucial part of the criminal justice system for the foreseeable future. While the courts continue to find ways to ward off questionable memory evidence, researchers continue to develop new ways to study the factors and processes that lead to inaccurate memories.

Related Entries

- ▶ Eyewitness Research
- ▶ False Confessions and Police Interrogation
- ▶ Interviewing Eyewitnesses

Recommended Reading and References

Assefi SL, Garry M (2002) Absolute memory distortions: alcohol placebos influence the misinformation effect. Psychol Sci 14:77–80

Bernstein DM, Loftus EF (2009a) How to tell if a particular memory is true or false. Perspect Psychol Sci 4:370–374

Bernstein DM, Loftus EF (2009b) The consequences of false memory for food preferences and choices. Perspect Psychol Sci 4:135–139

Butler AC, Zaromb F, Lyle KB, Roediger HL III (2009) Using popular films to enhance classroom learning: the good, bad, and the interesting. Psychol Sci 20:1161–1168

Davis D, Loftus EF (2005) Age and functioning in the legal system. In: Noy YI, Karwowski W (eds), Handbook of forensic human factors in litigation. (pp 11-1–11-53). New York: CRC Press

Drivdahl SB, Zaragoza MS (2001) The role of perceptual elaboration and individual differences in the creation of false memories for suggested events. Appl Cognitive Psychol 15(3):265–281

Eakin DK, Schreiber TA, Sergent-Marshall S (2003) Misinformation effects in eyewitness memory: the presence and absence of memory impairment as a function of warning and misinformation accessibility. J Exp Psychol Learn Mem Cogn 29: 207–219

Foster JL, Huthwaite T, Yesberg JA, Garry M, Loftus EF (2012) Repetition, not number of sources, increases both susceptibility to misinformation and confidence in the accuracy of eyewitnesses. Acta Psychol 139:320–326

Garry M, Hayne H (eds) (2006) Do justice and let the sky fall: Elizabeth F. Loftus and her contributions to science, law, and academic freedom. Lawrence Erlbaum, Hillsdale

Garry M, Polaschek DLL (2000) Imagination and memory. Curr Dir Psychol Sci 9:6–10

Geraerts E, Bernstein DM, Merckelbach H, Linders C, Raymaekers L, Loftus EF (2008) Lasting false beliefs and their behavioral consequences. Psychol Sci 19:749–753

Greene E, Flynn MS, Loftus EF (1982) Inducing resistance to misleading information. J Verbal Learn Verbal Behav 21:207–219

Hyman IE Jr, Billings FJ (1998) Individual differences and the creation of false childhood memories. Memory 6(1):1–20

Kovaleski SF (2012) New details are released in shooting of Trayvon Martin. The New York Times. Retrieved from http://www.nytimes.com

Lindsay DS (2008) Source monitoring. In: Roediger HL III (ed) Cognitive psychology of memory. Learning and memory: a comprehensive reference, vol 2 (4 vols, Byrne J, ed). Elsevier, Oxford, pp 325–348

Lindsay DS, Hagen L, Read JD, Wade KA, Garry M (2004) True photographs and false memories. Psychol Sci 15(3):149–154

Loftus EF (2005) Planting misinformation in the human mind: a 30-year investigation of the malleability of memory. Learn Mem 12:361–366

Loftus EF, Ketcham K (1994) The myth of repressed memory. St. Martin's Press, New York

Loftus EF, Miller DG, Burns HJ (1978) Semantic integration of verbal information into a visual memory. J Exp Psychol Hum Learn Mem 4:19–31

Loftus EF, Pickrell J (1995) The formation of false memories. Psych Annal 25:720–724

Mazzoni GAL, Loftus EF, Seitz A, Lynn SJ (1999) Changing beliefs and memories through dream interpretation. Appl Cognitive Psychol 13(2):125–144

McNally RJ, Lasko NB, Clancy SA, Macklin ML, Pitman RK, Orr SP (2004) Psychophysiological responding during script-driven imagery in people reporting abduction by space aliens. Psychol Sci 15:493–497

Memon A, Zaragoza M, Clifford BR, Kidd L (2009) Inoculation or antidote? The effects of cognitive interview timing on false memory for forcibly fabricated events. Law Human Behav 34: 105–117

Napolitano J (2004) Ex-patient wins suit. The New York Times. Retrieved from http://www.nytimes.com

Newman EJ, Garry M (2013) False memory. In: Perfect T, Lindsay DS (eds) The SAGE handbook of applied memory

Scoboria A, Mazzoni G, Kirsch I, Milling LS (2002) Immediate and persisting effects of misleading questions and hypnosis on memory reports. J Exp Psychol Appl 8:26–32

Seamon JG, Philbin MM, Harrison LG (2006) Do you remember proposing marriage to the Pepsi machine? False recollections from a campus walk. Psychon Bull Rev 13:752–756

Stark CEL, Okado Y, Loftus EF (2010) Imaging the reconstruction of true and false memories using

sensory reactivation and the misinformation paradigms. Learn Mem 17:485–488
Unkelbach C, Stahl C (2009) A multinomial modeling approach to dissociate different components of the truth effect. Conscious Cogn 18:22–38
Wade KA, Garry M, Read JD, Lindsay DS (2002) A picture is worth a thousand lies: using false photographs to create false childhood memories. Psychon Bull Rev 9:597–603
Wade KA, Green SL, Nash RA (2010) Can fabricated evidence induce false eyewitness testimony? Appl Cognitive Psychol 24:899–908
Weiser B (2011) In New Jersey, rules are changed on witness IDs. The New York Times. Retrieved from http://www.nytimes.com
Zaragoza MS, Mitchell KJ, Payment K, Drivdahl S (2011) False memories for suggestions: the impact of conceptual elaboration. J Mem Lang 64(1):18–31
Zhu B, Chen C, Loftus EF, Lin C, He Q, Chen C et al (2010a) Individual differences in false memory from misinformation: cognitive factors. Memory 18:543–555
Zhu B, Chen C, Loftus EF, Lin C, He Q, Chen C et al (2010b) Individual differences in false memory from misinformation: personality characteristics and their interactions with cognitive abilities. Person Indiv Differ 48:889–894

Family Engagement Strategies to Reduce Crime

Silke Meyer and Lorraine Mazerolle
Institute for Social Science Research,
The University of Queensland, St. Lucia,
QLD, Australia

Synonyms

Networked policing; Nodal policing; Plural policing; Third-party policing (TPP)

Overview

This entry describes the principles of third-party policing (TPP) on the example of the Family Engagement Strategy (FES). The FES is a police-led partnership approach, tailored to reduce the need for police interventions among high-risk families and juveniles. Key principles of TPP are described. Their application is illustrated on the example of the FES, to highlight some of the key benefits and challenges inherent to TPP. This entry highlights the usefulness of TPP principles to both high-risk target populations and agencies involved in TPP models.

Introduction

The Family Engagement Strategy (FES) is a police-led intervention that focuses on high-risk families that generate a disproportionately large number of crime incidents, calls for police service, and arrests. The FES is a police response to the long known fact in criminology that a small number of people are accountable for a disproportionately large number of recorded criminal offenses and civil code violations. These offenses include various forms of antisocial and criminal behavior, youth deviance, and family violence (Sherman 1992; Wolfgang et al. 1972). Research consistently shows that less than 10 % of the juvenile offender population account for over 50 % of police contacts and court hearings (Braga et al. 2001; Wolfgang et al. 1972). These high-risk juveniles are often embedded in families with a complex range of risks and needs (Moffitt and Caspi 2001) that not only absorb a significant amount of law enforcement resources (see Farrington & Welsh 2003; Mulder et al. 2011), but also drain other regulatory and non-regulatory support agencies (Farrington et al. 2001; Wolfgang et al. 1972).

In 2009, the Brisbane Metropolitan North District of the Queensland Police Service (QPS) in Australia initiated the Family Engagement Strategy (FES), a partnership response aimed at reducing problems associated with high-risk juveniles and their families. QPS officers examined their annual statistics for police contacts associated with particular physical addresses in the Brisbane Metropolitan North Region and, like thousands of other police agencies throughout the world, they found that a small number of private family residences were associated with a disproportionately large number of police contacts. In the majority of these "hotspot" residences, the police contacts related to an adolescent member of the family household engaging in behaviors that warranted police attention, such as truancy, public nuisance, and curfew violations. While these initial contacts with police did not always attract criminal sanctions, they consumed a large proportion of police resources due to their frequent occurrence. Notably, many of the high-risk families also had records of past interventions with a range of other service agencies.

The police used the key ideas of third-party policing (TPP) to develop the FES partnership approach. At the core of the FES is an interagency working forum of agency representatives that have access to various legal levers and diverse responsibilities for reducing the deviant behavior of highly troublesome family members. In this way, the FES is modeled after existing multiagency intervention strategies which rely on a variety of partnerships, including partnerships between different law enforcement agencies (see, e.g., Bond and Gittell 2010), partnerships between clusters of social support agencies (see, e.g., Stern and Green 2005), and partnerships between law enforcement and other support agencies (see Penfold et al. 2004).

This entry outlines the Queensland Police Service (QPS) Family Engagement Strategy (FES) as an example of third-party policing (TPP) applied to policing the problems associated with deeply troubled families. It describes the underpinning philosophies and principles of TPP and demonstrates how some of the key components of TPP were applied to the FES. It begins with a description of TPP and plural policing strategies in general and the specific challenges relating to partnership formation and agency buy-in. It then summarizes the key components of the FES and draws some conclusions around how the FES dynamics might inform future TPP approaches.

Fundamentals: Third-Party Policing

Buerger and Mazerolle (1998) first introduced the idea of TPP in the late 1990s. As a model of policing partnerships, it involves the strategic collaboration between local police forces and civil regulatory agencies to address and prevent criminal activities (Buerger and Mazerolle 1998; Mazerolle and Ransley 2005). In some jurisdictions, governments have actively issued a mandate for police to form collaborative partnerships with other regulatory and non-regulatory agencies to pursue problem-solving policing approaches (see, e.g., the United Kingdom; Fleming 2006). In others, including Australia, there is a greater reliance on "voluntary" partnerships between police services and other regulatory agencies (Fleming 2006). In both cases, the focus is on forming effective partnerships between agencies concerned with similar social issues and target groups, who share a mutual interest in preventing and addressing crime.

In the case of TPP, the partnerships operate on a spectrum from invited or encouraged partnerships at one end, to forced or coerced partnerships at the other end (Mazerolle and Ransley 2006). At the latter end of the spectrum, police use their coercive powers to "force" partnerships with non-offending third parties that are able to address particular crime problems, but who would not usually see this as their role. These approaches generally focus on crime control rather than prevention. At the front end of the spectrum, police invite third parties that share a mutual interest in addressing certain forms of crime or antisocial behavior to contribute to policing strategies, which are usually prevention focused (Mazerolle and Ransley 2005). While partnerships can be formed under very different circumstances, the objective remains the same from a police perspective; namely, to form partnerships with third parties that are able to regulate criminal or deviant behavior and share in the responsibility of addressing and preventing crime within the community (Braga et al. 2001; Buerger and Mazerolle 1998).

Past research on TPP reveals that the majority of partnerships initiated at the coercive end of the spectrum are formed to target street-level criminal activities, including the street-level drug market, street prostitution, shop lifting, and other forms of low-level property crimes

(Buerger 2007; Mazerolle and Ransley 2005). Under the model of TPP put forward by Buerger and Mazerolle (1998), "partnerships" are formed with proximate targets (third parties) to minimize opportunities for crime perpetrated by ultimate targets (individual offenders). Proximate targets are those non-offending parties who police choose to approach because of their access to regulatory powers, or so-called legal levers, that can be used to address the situational circumstances of crime. Proximate targets may include, for example, property owners renting properties to known drug dealers, or shop owners serving known drug dealers who tend to hang out outside the shop and use the street level as their market platform (Mazerolle and Ransley 2005). The regulatory levers available to the proximate targets are applied to reduce the occurrence of crime in a location-specific manner. While these levers are not accessible to the police, the formation of coerced partnerships with a relevant third party enables the police to indirectly access them. At this end of the spectrum of TPP, non-offending third parties are "coerced" into the partnership through the application of another level of legal levers in the form of applicable civil laws that can be used against the third party should the third party refuse to collaborate with the police. An example of applicable civil law codes that can be used as "legal levers" to "coerce" partnerships is legislation and regulations around unsafe or substandard rental properties. In an attempt to control criminal activities (e.g., drug dealing) that take place in substandard rental properties, the police identify agencies or individuals with access to civil law codes that can be applied to indirectly regulate and control these criminal activities. The police approach the relevant party (in this case the property owner or housing company) and ask them to perform their third-party role by applying tenancy laws that they can access – to terminate the tenancy agreement of the ultimate target. Should the third party be reluctant to carry out the eviction (e.g., fearing a loss of rental income), the police can utilize the legal lever which defines their responsibility as a homeowner in relation to unsafe or substandard housing conditions. The legal lever available to police can be escalated from an initial written request outlining the possible consequences in case of noncompliance on behalf of the third party – at the lower end of the "coercive pyramid" (Ayres and Braithwaite 1992), to a formal report to the Housing Commission, which is likely to result in the property being boarded up. This approach "encourages" the selected third party to take responsibility for their property and partner with the police in crime control and community safety (Green 1996).

In addition to the street-level TPP approaches, which predominantly focus on crime and place in a context of crime control, more recent forms of TPP have incorporated aspects of crime control and crime prevention through partnerships that are able to generate individual behavior change rather than opportunities that converge in space and time (Braga et al. 2001; Corsaro et al. 2009). These forms of TPP target problems such as youth deviance, child safety, and domestic and family violence. These forms of crime and antisocial behavior often attract the attention of multiple regulatory agencies which focus on individual behavior change and regulation. Notably, TPP approaches that focus on problem families and individuals differ from the more place-specific TPP approaches in the type of agency approached for the third-party partnership as well as the nature and duration of the intervention. While place-specific approaches of TPP are often formed with parties who manage certain locations but share no primary concern in individual crime control or crime prevention, person-specific approaches utilize partnerships with agencies that share a similar agenda in reducing and preventing crime and other potentially harmful behaviors (see, e.g., Braga et al. 2001; Mazerolle and Ransley 2005). While these agencies still require encouragement to take on certain policing roles in addition to their general agenda, their engagement as a third party predominantly takes place within their organizational aims and objectives. This will be illustrated through the example of the FES, which is built on police partnerships with agencies concerned with the well-being of high-risk families.

In practice, partnerships that aim to address a range of social and criminal issues are often marked by tensions between agencies that operate within different practical and philosophical frameworks. Having a legislative framework that lends itself to the formation of partnerships around issues of mutual concern is therefore only the first step in forming effective third-party partnerships that follow the principles of crime prevention. Most parties involved experience significant challenges throughout the partnership formation process as well as the intervention development and implementation phases. Challenges mainly arise from the diverse backgrounds and philosophical underpinnings of the different agencies involved. While parties may share a mutual interest in reducing crime and antisocial behavior, for example, police priorities focus on the end result (i.e., reducing police resources invested in addressing the issue), the primary focus of the third party often centers on changing the underlying causes of the issue (e.g., school failure, family conflict) (Bond and Gittell 2010; Fleming 2006; Sloper 2004).

In addition to the potential mismatch in underlying philosophies that drive different partnership approaches, police-led partnerships are prone to experiencing struggles around sharing the power and entitlement associated with crime reduction and related interventions. As described above, certain civil laws offer an opportunity for police to "coerce" rather than "invite" or "encourage" partnerships. This can lead to hostile and reluctant cooperation from the third party, and as a result, so-called partnerships can be fragmented due to actual and perceived power imbalances and a lack of mutual trust and respect between agencies and organizations (Fleming 2006; Loader 2000). While fragmentation and power imbalances can be managed through the approach taken by the police during the partnership initiation phase, the success of these (usually short-term) interventions does not depend on the level of buy-in from the third party since cooperation can be "coerced" by the police.

In contrast, invited or encouraged partnerships require a sufficient level of buy-in on behalf of the third party to be effective. The level of buy-in is strongly influenced by the level of mutual trust and respect generated during the partnership initiation phase. While this initiation phase is more time consuming than establishing short-term "coerced" partnerships, it is necessary to build a suitable foundation for interventions that aim to achieve long-term change and behavior improvement rather than short-term changes to spatial opportunities (Bond and Gittell 2010; Braga et al. 2001; Fleming 2006).

Both forms of TPP partnerships share a common overarching goal, namely, the sharing of police responsibilities around crime control and crime prevention with third parties that are in a position to enforce citizen compliance through regulatory statutes not available to the police (Buerger 2007; Green 1996; Mazerolle and Ransley 2005, 2006). The ultimate goal is to minimize the use of police resources for matters that can be controlled or addressed by other regulatory agencies (Mazerolle and Ransley 2005, 2006). In the case of the FES, this goal is expected to be achieved through invited partnerships with agencies that are able to regulate individual behavior of high-risk family members through the application of regulatory levers.

Partnership Policing: A Matter of Shared Roles and Responsibilities

Partnership policing builds on the idea of shared responsibilities for controlling and preventing crime (Loader 2000; Stenning 2009) such that crime problems are no longer the sole responsibility of those empowered by the State to regulate public behavior through the use of force and punitive reactions (Loader 2000). The increasing use of pluralized approaches to policing over the past two decades has created an extended "policing family" in modern societies. Today, policing responsibilities are shared between a number of different agencies and organizations to address a broad range of social issues and concerns (Fleming 2006; Loader 2000). Some jurisdictions, including Australia, rely on voluntarily initiated partnerships (Fleming 2006), while others, such as the United Kingdom, have clear mandates around partnerships designed to provide crime prevention and crime control strategies (Florence et al. 2011). Whether mandated or optional, most contemporary Western societies now take a pluralized approach to policing problem people (Bond and Gittell 2010; Braga et al. 2001; Corsaro et al. 2009), problem places (Weisburd and Green 1995), or criminogenic situations (Mazerolle et al. 2000).

Pluralized policing can take a number of different forms, including handing over traditional public policing roles to private security companies (see Bradley and Sedgwick 2009; Sarre and Prenzler 2000), networked policing models that bring together police and community partners (Crawford 2006; Shearing and Stenning 1983), and third-party policing approaches that involve police co-opting other entities to share the crime control load (Ayling and Grabosky 2006; Mazerolle et al. 2000). Examples of these pluralized approaches include private security companies supporting federal police terrorism control and peacekeeping missions (Loader 2000), private security companies taking on the responsibility of crime and crowd control at public events (Sarre and Prenzler 2000), private security guards patrolling public places or private properties (Bradley and Sedgwick 2009; Stenning 2009), and police partnering with community-based neighborhood watch groups to share responsibilities of crime control with individual community groups (Ayling and Grabosky 2006; Bradley and Sedgwick 2009; Kempa and Johnston 2005).

The third-party partnership approach also follows the principle of police working with other agencies to establish a shared and pluralized approach to reducing crime and the associated police workload (Ayling and Grabosky 2006; Mazerolle and Ransley 2005, 2006). The TPP partnership approach, however, differs from other forms of pluralized policing in the rationale that drives this partnership formation. Partnerships forged through TPP are explicitly sought based on the legal lever accessible to the partner but otherwise unavailable to the police. Rather than simply sharing policing responsibilities by handing certain roles over to another entity or agency, TPP seeks access to the legal levers that are thought to facilitate police efforts to control and prevent crime (Buerger 2007; Mazerolle and Ransley 2005). Police identify suitable third parties with access to regulatory levers unavailable to the police (e.g., levers to regulate substandard housing arrangements or matters around family violence and child abuse). This approach enables police to harness help while remaining involved as the lead organization in the intervention efforts (see, e.g., Braga et al. 2001; Corsaro et al. 2009).

While third-party policing partnerships hold value for the police (see Mazerolle and Ransley 2006), they do not come without their challenges. As discussed above, third-party partnerships are formed on a spectrum from invited to coerced partnerships, the latter leading to a range of challenges around establishing a sufficient level of agency buy-in and collaboration (Buerger and Mazerolle 1998; Mazerolle and Ransley 2005).

Most police matters relating to high-risk juveniles and their families also constitute grounds for intervention by other agencies. Mandatory notification between the police and other regulatory agencies when dealing with matters, including truancy, family violence, concerns around supervision and suspected neglect, require cross agency communication and ideally collaboration (Ayling and Grabosky 2006). Formalized partnerships to jointly address these issues share the policing workload and reduce the use of resources at either agency end, therefore seem an all too obvious solution.

The Queensland Police Service Family Engagement Strategy (FES)

The FES is a crime control initiative that focuses on the cooperative aspects of TPP partnerships. The FES comprises the following three key components:

1. Focuses on long-term behavior regulation
2. Addresses a complex variety of presenting issues within high-risk families
3. Creates a multiagency working forum with access to a range of legal levers

The following sections provide an overview of the FES target population, the range and role of invited third-party partner agencies, and the processes associated with forming these partnerships.

Target Population

Based on the information derived from internal police data reviewed during the planning phase of the FES, families with adolescent children who fall into the "top 100 list" of police contacts linked to their residential address in the Metropolitan North region during the previous 12 months "self-select" as the target population for the FES. The presenting issues identified from official police records for contacts with one or more of the relevant family members in each high-risk family/household range from noncriminal risk-taking behavior, including truancy, substance misuse, and curfew violations, to mental health issues, child abuse, domestic violence, and a range of officially recorded criminal activities.

The QPS FES Team

The FES is led by the QPS team based in the Metropolitan North Region. The team consists of the leading Inspector for the Metropolitan North Region, an unsworn project manager (with a degree in human services), and three front line sworn police officers from three different districts within the Metropolitan North Region. The leading Inspector and project manager were actively involved in the partnership formation and project planning phase and coordinate all working forums. The project manager is further involved in all initial client engagement. All three frontline officers are involved in the identification of suitable families for the FES in their relevant districts and accompany the project manager during the initial client engagement.

Voluntary Third-Party Partnership Formation

With the absence of mandated policing partnerships in Australia (Fleming 2006) and the lack of applicable legal frameworks to "coerce" partnership interventions around long-term behavior regulations as opposed to short-term spatial interventions (Mazerolle and Ransley 2005), the FES requires the voluntary buy-in from potential third parties to ensure their cooperation in taking on some of the traditional policing responsibilities as they relate to regulating antisocial behavior displayed by high-risk families.

In late 2009, the QPS invited a range of regulatory agencies to become involved in a third-party partnership approach to reduce the disproportionately high number of police contacts accumulated by a relatively small number of high-risk families. The overarching goal was to form partnerships designed to address long-term behavior regulation at an individual level. Given the level of required buy-in and long-term commitment to a variety of interventions offered as part of the FES, the partnership formation phase required a lengthy negotiation phase marked by various stages of protocol negotiations and renegotiations between QPS and the approached third parties. Overall, the partnership development and project planning phase of the FES lasted for 18 months from the initial invitations extended by QPS in late 2009. This lengthy process was required to ensure the expectations of the lead agency (i.e., QPS) were aligned with the organizational guidelines and legislative frameworks of each agency when acting as a third party in a multiagency TPP approach.

The key challenges that arose throughout the partnership formation process were associated with initial differences in perceived and promoted aims and objectives between QPS as the lead agency and some of the third parties. Both police officials and third-party representatives had to develop an understanding for each other's underlying goals, values, and organizational philosophies to negotiate terms of references acceptable for all parties involved. Ongoing protocol negotiations between all parties involved ensured the development of a trusting and respectful partnership between agencies that often had to overcome a history of hostile working relations marked by power struggles and poor communication (see Meyer and White 2011). The challenges observed throughout the FES partnership formation phase are not uncommon. Several studies examining the processes and outcomes of partnerships between the police and other regulatory agencies suggest that the development of partnerships marked by mutual trust and respect can be a lengthy and sometimes impossible process where historically poor working relations are involved (Bond and Gittell 2010; Fleming 2006; Rosenbaum 2002).

Forum Driven

A key component of the FES was the forum-driven multiagency nature of this approach. Instead of relying on a dyad relationship with a single third party, the FES used a multiagency forum approach that involved multiple third parties within the same TPP approach. This approach was necessary to address the complexity of presenting issues and behaviors to be regulated under the FES. Partnership negotiations and protocol renegotiations had to take place between the police and a broad range of regulatory agencies to ensure access to a diverse range of regulatory levers applicable to the high-risk families targeted by the FES. This component further contributed to the lengthy partnership development phase.

The initial call for interest in a TPP approach to high-risk families went out to a broad range of regulatory agencies operating at a state level in Queensland (Qld), Australia. Following an opt-in approach, these agencies were asked to indicate their initial interest and later confirm their commitment to the FES by agreeing to the project's terms of reference and data sharing agreement. Regulatory agencies (or their respective Departments) approached to form the TPP partnership included the following:

- Qld Department of Education
- Qld Department of Health, represented by the following subbranches:
  - Primary and Community Health Services
  - Alcohol and Drug Service
  - Mental Health Service
  - Children and Youths Mental Health Service
- Qld Department of Communities, represented by the following subbranches:
  - Child Safety Programs and Partnerships
  - Child Protection Development
  - Housing and Homelessness Services
  - Youth and Family Support Services
- Qld Department of Employment, Economic Development and Innovation
- Qld Department of Corrective Services (Probation and Parole)

The FES working forum was formed through preliminary electronic and in-person conversations between QPS and the approached agencies. While all 11 regulatory agencies and departments approached remained involved beyond the initial working forum planning phase, only 8 opted into a formal TPP partnership through signing the data sharing agreement. Three of the four branches under the Department of Communities stream (namely, Child Safety, Child Protection, and Housing and Homelessness Services) chose to attend only some of the working forums throughout the project planning phase and maintained an informal involvement in the FES. Their involvement comprised the provision of informal feedback at the working forums they were able to attend. Over an 18 months planning phase, working forums were held every 2 months to negotiate processes and protocols between QPS as the leading agency, each regulatory agency formally involved as a potential third party, and the University of Queensland as the evaluator of the FES. Initial and ongoing involvement of the evaluation team was crucial to ensure a sound methodology for the initial process and a long-term impact evaluation.

Multiple working forums were required throughout the project planning phase to give representatives of all partner agencies the opportunity to be actively involved in the planning and development of the FES prior to its implementation. The lack of formal engagement by some of the regulatory agencies initially approached for a formal partnership is not unusual for TPP approaches built on invited partnerships, that is, where there is no legal lever available to the police to "coerce" collaboration. While approaches like the ones described on street-level interventions to housing and drug sales (Mazerolle and Ransley 2005, 2006) offer an opportunity for police to coerce collaboration of proximate targets through threats of adverse outcomes in case of non-compliance, invited partnerships require the full buy-in of invited third parties. Full buy-in can best be achieved in partnerships between agencies that share a certain level of mutual trust and respect and pursue similar goals in crime prevention (Rosenbaum 2002; Sloper 2004). Once partnerships had been established, the monthly working forum discussions were used to drive the development and later implementation of the FES. This collaborative forum approach gave involved third parties an opportunity to shape the FES through equal involvement and decision-making powers. Past research has identified equal input and decision-making powers as one of the key factors contributing to the effectiveness of police partnership projects (Bond and Gittell 2010; Huey 2008; Stern and Green 2005).

Regulatory Agencies and Their Involvement

Due to the complex nature of risks and needs prevalent in high-risk families, the FES regulatory agencies have to contribute a broad range of legal levers as part of their involvement. The following three scenarios illustrate some of the key presenting issues, identify suitable third parties, and describe their regulatory involvement and the role of proximate and ultimate targets as they relate to the FES.

1. *Presenting issue: truancy*. A young person and his/her parent(s) may have agreed to regular school attendance under a FES Action Plan. In the case of unexplained absence coming to the attention of the Department of Education, the department may choose to issue an initial warning to the young person and his/her parent(s) in their role as the responsible supervisors. This warning can later be escalated into a fine (Ayres and Braithwaite 1992; Mazerolle and Ransley 2005) should the young person and his/her parent(s) engage in further non-compliance with the Action Plan.
2. *Presenting issue: absconding*. A family may have child safety concerns issued against an adult caretaker for exposing the young person to domestic violence where this exposure may lead to the young person running away from home and engaging in antisocial behavior while unsupervised. In cases where further incidents of domestic violence come to the attention of the Department of Child Safety, the department may choose to issue an initial warning, subject the parent to mandatory parenting training, or escalate the lever to the removal of the young person from the violent family household.
3. *Presenting issue: public rental property damage*. A young person may engage in frequent and severe forms of aggression against other family members, leading to property damage and police attention. In cases where further incidents of escalating violence and property damage come to the attention of the Department of Housing, the department may choose to apply its legal lever by issuing an initial warning which can later be escalated to an eviction notice should the property damage continue.

In all three examples, the initial third-party partnership between the police and a selected regulatory agency was extended to a “secondary third party,” namely, the adult caretaker of the young person as a proximate target of the overall intervention. While QPS formed the initial partnership with a regulatory third party to share their workload as it relates to high-risk families, the selected third party was then in a position to apply the traditional concepts of TPP and “coerce” parents as proximate targets into a partnership that seeks to regulate the behavior of the ultimate target, namely, the young person whose behavior attracts a disproportionately high number of police contacts.

The Role of Additional Non-regulatory Partners

In addition to the regulatory third parties invited to participate in the FES, the established FES working forum felt it was important to go beyond the traditional TPP partnerships with regulatory agencies and invite further community-based agencies relevant to addressing the needs of high-risk families. Past research has documented the effectiveness of such “mixed” partnership approaches. The “Boston Pulling Levers” project (Braga et al. 2001) followed a problem-solving

TPP approach that invited additional non-regulatory agencies to combine the key deterrence component with a rehabilitative intervention component that was able to address the underlying risks and needs of the target population. The impact evaluation of this mixed partnership approach found strong support for its effectiveness, leading to the replication of its framework in other sites in the United States (see, e.g., Corsaro et al. 2009). In relation to the FES, the rehabilitative component is to be understood in a criminal justice context. This component therefore includes agencies that support individuals in developing pro-social skill sets and protective factors, including but not limited to rehabilitation in a medical sense (e.g., from substance misuse). The non-regulatory agencies invited to provide a rehabilitative component to supplement the regulatory work of the controlling partner agencies in the FES include the following NGOs:

- Salvation Army Youth Outreach Services
- Drug Awareness, Rehabilitation and Management
- Domestic Violence and Family Counselling Program (SANDBAG)
- Employment Service Inc (EPIC)
- Northwest Aboriginal and Islander Community Association (NWAICA)

All five invited, non-regulatory partner agencies signed the initial data sharing agreement. Except for the culturally specific service, NWAICA, all other non-regulatory partner agencies are represented at most working forums. While their intervention contribution is designed to be supplementary to the regulatory components of the FES, all non-regulatory agencies were also given the opportunity of active involvement and input during the planning phase of the FES.

Data Sharing

A crucial component of the forum-driven multiagency TPP approach of the FES is the data sharing agreement set up by QPS, which had to be signed by each partner agency at the beginning of their involvement. This agreement enables QPS as the lead agency to identify relevant legal levers matching the presenting problems identified for each target youth and his/her high-risk family through access to third-party records. The data sharing agreement with non-regulatory agencies is of similar importance because it provides a broader background on participants' proven or suspected civil law violations, behaviors associated with an increased risk in civil and criminal law violations along with past interventions participants had been exposed to. While non-regulatory agencies are not in a position to apply a legal lever to control subsequent violations (e.g., in form of domestic violence or substance misuse), their record sharing provides QPS with more detailed information on each family's key presenting issues along with past (unsuccessful) interventions. Along with the information received from all regulatory partner agencies, this information facilitates the selection of the most appropriate third party in a position to regulate these issues.

While both the police and their partner agencies had to overcome some perceived barriers and challenges around understanding the value and importance of data sharing for TPP and multiagency approaches to high-risk populations, all parties involved eventually negotiated a process they felt met their ethical requirements without compromising their clients' right to privacy. The benefits of such data sharing approaches have been evidenced in past research results, which suggest that the needs of vulnerable and high-risk populations can be met more effectively through transparent communication between relevant support agencies. It is, however, not uncommon for this to be a lengthy negotiation process (Fleming 2006; Florence et al. 2011).

Conclusion

This entry describes the emergence of different forms of police partnerships with a specific focus on TPP. It introduces a TPP program known as the FES and draws on the relevant literature to demonstrate the application of key components of TPP to the FES. It illustrates the contributions TPP is able to make in the area of crime control that requires individual behavior change rather than short-term spatial interventions. The FES is a novel approach to TPP, which further advances the original components of TPP through multilayered partnerships that move beyond the original dyad nature of TPP partnerships. Through inviting partnerships with a number of agencies, the FES constitutes a crime control strategy that is able to address a complex range of social and criminal issues. It is the first TPP approach to target the small proportion of high-risk juveniles and families that are responsible for the majority of police and other agency contact. Through the regulatory partnership approach, with a supplementary rehabilitative component, the FES offers an opportunity to instantly regulate high-risk behavior while at the same time addressing the underlying issues which influence long-term behavior change. Such an approach holds great value in addressing the needs of its target population while at the same time reducing the resources invested by each individual partner agency.

Related Entries

- ▶ Communities That Care
- ▶ Focused Deterrence and "Pulling Levers"
- ▶ Police Family Violence Services
- ▶ Problem-Oriented Policing
- ▶ Pulling Levers Policing
- ▶ Third Party Policing and School Truancy
- ▶ Third-Party Policing

Recommended Reading and References

Ayling J, Grabosky P (2006) Policing by command: enhancing law enforcement capacity through coercion. Law Policy 28(4):420–443

Ayres I, Braithwaite J (1992) Responsive regulation: transcending the deregulation debate. Oxford University Press, Oxford

Bond BJ, Gittell JH (2010) Cross-agency coordination of offender re-entry: testing collaboration outcomes. J Criminal Justice 38(2):118–129

Bradley T, Sedgwick C (2009) Policing beyond the police: a 'first cut' study of private security in New Zealand. Policing Soc 19(4):468–492

Braga AA, Kennedy DM, Waring EJ, Piehl AM (2001) Problem-oriented policing, deterrence, and youth violence: an evaluation of Boston's Operation Ceasefire. J Res Crime Delinquency 38(3): 195–225

Buerger M (2007) Third-party policing: futures and evolutions. In: Schafer JA (ed) Policing 2020: exploring the future of crime, communities, and policing. US Department of Justice, Federal Bureau of Investigation, Washington, DC, pp 452–486

Buerger M, Mazerolle L (1998) Third-party policing: a theoretical analysis of an emerging trend. Justice Quart 15(2):301–327

Corsaro N, Brunson RK, McGarrell EF (2009) Problem-oriented policing and open-air drug markets: examining the Rockford Pulling Levers Deterrence Strategy. Crime Delinquency. doi:10.1177/0011128709345955

Crawford A (2006) Networked governance and the post-regulatory state? Steering, rowing and anchoring the provision of policing and security. Theoretical Criminol 10(4):449–479

Farrington DP, Jolliffe D, Loeber R, Stouthammer-Loeber M, Kalb M (2001) The concentration of offenders in families, and family criminality in the prediction of boys' delinquency. J Adolesc 24(5):579–596

Farrington DP, Welsh BC (2003) Family-based prevention of offending: a meta-analysis. Aust NZ J Criminol 36(2):127–151

Fleming J (2006) Working through networks: the challenge of partnership policing. In: Fleming J, Wood J (eds) Fighting crime together: the challenge of policing and security networks. University of New South Wales Press Ltd, Sydney, pp 87–115

Florence C, Shepherd J, Brennan I, Simon T (2011) Effectiveness of anonymised information sharing and use in health service, police, and local government partnership for preventing violence related injury: experimental study and time series analysis. BMJ. doi:10.1136/bmj.d3313

Green L (1996) Policing places with drug problems, vol 2. Sage, Thousand Oaks

Huey L (2008) 'When it comes to violence in my place, I am the police!' Exploring the policing functions of service providers in Edinburgh's Cowgate and Grassmarket. Policing and Society 18(3):207–224

Kempa M, Johnston L (2005) Challenges and prospects for the development of inclusive plural policing in Britain: overcoming political and conceptual obstacles. Australian New Zealand J Criminol 38(2):181–191

Loader I (2000) Plural policing and democratic governance. Social Legal Studies 9(3):323–345

Mazerolle L, Ransley J (2005) Third-party policing. Cambridge University Press, New York

Mazerolle L, Ransley J (2006) The case for third-party policing. In: Weisburd D, Braga AA (eds) Police innovation: contrasting perspectives. Cambridge University Press, New York, pp 191–206

Mazerolle L, Price J, Roehl J (2000) Civil remedies and drug control: a randomized field trial in Oakland, California. Eval Rev 24(2):212–241

Meyer S, White J (2011) The family engagement strategy: pilot process report. The University of Queensland, Queensland

Moffitt TE, Caspi A (2001) Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. Dev Psychopathol 13(2):355–375

Mulder E, Brand E, Bullens R, van Marle H (2011) Risk factors for overall recidivism and severity of recidivism in serious juvenile offenders. Int J Offender Ther Comp Criminol 55(1):118–135

Penfold C et al (2004) Tackling client violence in female street prostitution: inter-agency working between outreach agencies and the police. Polic Soc 14(4):365–379

Rosenbaum DP (2002) Evaluating multi-agency anti-crime partnerships: theory, design, and measurement issues. Crime Prevention Studies 14:171–225

Sarre R, Prenzler T (2000) The relationship between police and private security: models and future direction. Int J Comparative Appl Criminal Justice 24(1):91–113

Shearing CD, Stenning PC (1983) Private security: Implications for social control. Soc Probl 30(5):493–506

Sherman L (1992) Policing domestic violence: experiments and dilemmas. Free Press, New York

Sloper P (2004) 'Facilitators and barriers for co-ordinated multi-agency services', child: care. Health Development 30(6):571–580

Stenning P (2009) Governance and accountability in a plural policing environment – the story so far. Policing 3(1):22–33

Stern R, Green J (2005) Boundary workers and the management of frustration: a case study of two Healthy City partnerships. Health Promot Int 20(3):269–276

Weisburd D, Green L (1995) Policing drug hot spots: the Jersey City Drug Market Analysis Experiment. Justice Quarterly 12(4):711–735

Wolfgang M, Figlio R, Sellin T (1972) Delinquency in a birth cohort. University of Chicago Press, Chicago

Fatigue

► Officer Safety, Health, and Wellness

FBI Influence on State and Local Police

William J. Vizzard
Division of Criminal Justice, California State
University Sacramento, Sacramento, CA, USA

Overview

This article reviews the history of Federal Bureau of Investigation (FBI) influence on US police agencies over the past 80 years. The FBI has exercised direct and indirect influence over American policing through a variety of activities, including the national fingerprint repository, the national crime lab, training for state and local police, the Uniform Crime Report system, the National Crime Information Center, direct investigative assistance, anti-crime task forces, and most recently, the national system for making investigative DNA comparisons. Perhaps of equal importance, the FBI was one of the leading agencies in the evolution toward professional policing, and was unparalleled in its focus on public relations and image management.

Before the Bureau

Formalized and structured policing did not develop in the United States until the mid nineteenth century. Initially only in a few larger cities, formal police departments spread rapidly to smaller cities after the Civil War. However, policing remained a city function throughout the nineteenth century. Although sheriffs and constables provided some level of law enforcement for rural areas, neither the states nor counties established formal police agencies during this period. Cities acted out of necessity, because rapid growth and immigration brought crime and disorder, and no other level of government was able or willing to assume the responsibility. American political culture and tradition combined with Constitutional restrictions to assure policing would be a local responsibility. The Tenth Amendment to the United States Constitution reserved all powers not specifically delegated to the federal government for the states or the people. Since policing is not even mentioned by the Constitution, this precluded the development of any type of national police agency. Even though the states were not precluded by any legal constraint from forming organized police forces, American preference for decentralized government and local control strongly favored local police.

The federal government did exercise some law enforcement powers. They had authority over crimes against the federal government, such as tax evasion and treason, and authority over crimes in the territories and on the high sea. United States Marshals were the first federal law enforcement officers followed by more specialized officers who protected the mail, enforce alcohol excise taxes and combated smuggling of goods to avoid tariffs. The first federal agency created solely for the purpose of investigating crimes was the United States Secret Service, established in 1865 to investigate and arrest counterfeiters. A bureau of the Treasury Department until 2002, the Secret Service was assigned the task of protecting the President after the assassination of President McKinley in 1901 because no other agency was available for the task.

The twentieth century brought widespread changes in American law enforcement and a greatly expanded role for counties, states and the federal government. By 1900 organized police forces were present in virtually every city of any size. As the century progressed two developments pushed states to begin organizing uniformed law enforcement agencies, the automobile and labor unrest. Rural areas still had little or no organized policing, so states established either state police or highway patrols. In the latter half of the century, most counties formalized their sheriffs offices into structured police agencies or formed county police to perform this function.

The federal government also expanded its role in law enforcement. Federal laws exercising control over narcotics and firearms were passed under the authority of the federal taxing authority in 1916 and 1934 respectively. These laws did not prohibit heroin or machine guns, they simply placed a tax that was higher than people were willing to pay. As tax laws, they avoided the restriction on federal authority (Vizzard 1997). In addition, from 1920 till 1933 federal law prohibited the manufacture and sale of alcoholic beverages as a result of the passage of the Eighteenth Amendment. Prohibition brought a wave of organized criminal activity and public corruption that generated demand for reform. Policing in the United States had been perceived as highly political and tainted by corruption from the beginning (Walker and Katz) and federal prohibition agents did not enjoy a better reputation (Okrent). Police reformers, such as Berkeley Police Chief August Vollmer had been campaigning for so called professional policing from early in the century. Their goals were the institution of civil service hiring, formal training and education, application of technology and science and independence from political control (Walker and Katz). The call for improved policing received national attention in 1931 as a result of the report from the Wickersham Commission, appointed by President Hoover to examine law enforcement in the United States (Walker and Katz).

J Edgar Hoover Assumes Control and Builds Mutual Dependence

In 1924, at the height of Prohibition, Attorney General Harlan Stone appointed John Edgar Hoover as Director of the Justice Department's Bureau of Investigation (BI). Originally formed in 1908, the BI was an obscure investigative unit with very limited jurisdiction and a poor reputation (Gentry). Hoover was a young lawyer who had worked at BI during World War I and the following years. Although very young and junior in the Justice Department, Hoover proved exceptional at bureaucratic organization. Over the next decade he survived two changes in administration while expanding both the size of the organization and its jurisdiction (Gentry). In 1935 the agency's name was finally changed to the current Federal Bureau of Investigation (FBI).

By aggressively pursuing real organizational reform and aggressive public relations, the FBI successfully developed a reputation for both honesty and efficiency that separated it from other law enforcement agencies from the 1930s through the early 1950s. Even though the law enforcement agencies of the Department of Treasury arrested and prosecuted far more offenders for violating federal law than did the FBI (Millspaugh), and local police continued to handle most crime, the FBI was seen by the public as what the FBI website describes as "America's premier law enforcement agency." This was accomplished by a combination of pursuing the most recent innovations in the applications of science and technology to criminal investigation, hiring well educated Special Agents, rigid discipline and rigid control of image. Hoover was obsessive in protecting the reputation of his agents and his bureau (Gentry; Powers 1983; Sullivan). This required that the FBI maintain a clear distinction from law enforcement in general, which did not enjoy a similar reputation. Yet, the FBI needed the support and assistance of local and state police. Police officials had considerable influence on politicians and local police had far more sources of information than the FBI.

The FBI pursued these goals in a number of ways. The Bureau of Investigation had assumed control of federal fingerprint files and consolidated them with a centralized file of prints taken by local police and maintained centrally by the International Association of Chiefs of Police (IACP) in 1924 (FBI). Although local police agencies had pioneered the use of fingerprints for identification, only a federal agency could provide the sort of central clearing house that would fully capitalized on their potential in a large and mobile nation. Contrary to the belief of many citizens, the FBI could not identify a person from latent prints left at a crime scene. A search of the files required a full set of rolled fingerprints and would continue to do so until computer technology developed to a level that allowed rapidly comparing a single latent fingerprint with the millions of individual prints in a file. But the centralized FBI file did allow the identification of persons in custody and provided a central library of prints from arrested persons. Once a potential suspect was identified, a latent print could be compared with the person's rolled prints. If the person had a local arrest record, the police could do this on their own. However, if the person had no such record, they were dependent on the FBI to make the comparison with the prints in the central file. This central finger print file created a symbiotic relationship between state and local law enforcement and the FBI unlike any relationship that had existed with a federal agency before. The FBI needed the cooperation of the police to build its fingerprint files by submitting fingerprints on all persons arrested, and the police needed FBI support in accessing the central file.

Technology and Training the Cornerstones

Although the fingerprint files created a mutual dependency, the FBI Laboratory did not. Initially established in 1932 to conduct research, the FBI Laboratory soon grew to become the largest and most diverse forensic laboratory in the country (FBI). The early technologies, such as fingerprint, tool mark and ballistic comparisons, had all been initiated by progressive local police agencies, such as Berkeley, California (Walker and Katz), but these technologies were not available to the majority of small or medium sized departments. Thus many of them came to depend on the FBI for support in the area of scientific investigation throughout the period from the 1930s through the early 1950s. The FBI laboratory served a variety of functions. First and foremost, it publicized and popularized the use of forensics in criminal investigation. Previously, police had to depend on eye witnesses or confessions to prove their cases. The dependence on confessions had produced a sordid history relating to police use of force during interrogations to extract confessions. In addition to popularizing "scientific investigation," the FBI provided direct service to police in the form of evidence examination as well as training and support for police personnel.

Training became a key component in the FBI support of local police. In 1935, the FBI established its National Academy (NA) to train local police officer. The NA would prove one of the FBI's most successful endeavors in building support among local police. Selected candidates attended classes in investigative techniques for several weeks at no expense to their parent agencies. Attendance at the National Academy soon became a status symbol for any ambitious police officer. The FBI focused on mid-level managers and followed up by making them members of the National Academy Associates. Special Agents were specifically assigned to facilitate and review applications. These same NA Coordinators arranged mixed training and social events that assured the NA graduates would remain connected to the FBI for their entire careers. Because police agencies routinely sent some of their most promising people to the NA and because NA attendance enhanced the graduates resume, the graduates routinely moved to the top positions in their respective agencies. Over time, this placed individuals with close ties to the FBI in charge of the majority of medium sized police agencies in the country. Although large agencies such as New York, Chicago and Los Angeles police departments tended to resist this trend, the NA and the accompanying follow up provided the FBI with substantial influence over and support from police. The NA also served to provide much needed training to key members of many police agencies in a time when such training was not available elsewhere. Although NA training may not seem advanced in the current era where universities and police agencies offer many options, it likely served as a key component of modernizing police before such options were available.

The FBI's training role did not stop with the National Academy. Over time, the FBI developed a cadre of skilled instructors, who provided training on a variety of subjects to local police. Up to the late 1950s only large police agencies and state police/highway patrols maintained formal training academies. Many officers learned on the job or received, at most, a short orientation upon employment. The FBI provided instructors for pre-service and in-service training, cost free. Surprisingly, the topics for instruction were not restricted to the areas where FBI expertise was greatest, such as law and investigation. Instruction ranged from unarmed combat to homicide investigation. For police agencies with limited budgets and staffing, this cost free training proved highly popular. Thus, cost free training was a critical element in maintaining FBI influence over local police.

The cost free nature of the FBI supplied training has been viewed as so fundamental to the FBI that the agency has even sought to undercut training offered by other federal agencies which involved cost sharing for fear that it would create a precedent. In 1992 this author was a member of the joint Justice Treasury Working Group to provide specialized training to local and state police at the Federal Law Enforcement Training Center. Although the FBI had a representative on the advisory board for the project and it was a Justice Department initiative, the FBI Assistant Director for Training openly urged members of the NA Associates to lobby Congress in opposition to the program.

Over time the FBI expanded its training role by developing both schools and field training cadres in a number of areas. One example is Special Weapons and Tactics (SWAT) tactical training. Although the SWAT concept was developed by the Los Angeles Police Department (LAPD) in the 1960s (Cannon), the FBI soon developed its own SWAT teams and SWAT school at the FBI Academy. By the early 1970s, the FBI became the dominate presence in SWAT training.

Field Operations and Investigative Support

Although these ancillary support functions constituted the core of the FBI program for supporting and gaining support from local police, some investigative functions also proved helpful. Federal criminal jurisdiction was greatly expanded during the 1930s as Congress began using the interstate commerce power granted by the Constitution to expand federal authority. Although the United States Marshals Service now takes the lead in apprehending federal fugitives, the FBI exercised this role for many years. While local agencies had limited interest in federal fugitives, they had much more interest in locating their own fugitives. Congress provided the FBI with a significant mechanism for assisting local police by passing the Unlawful Flight to Avoid

Prosecution (UFAP) statute. Although virtually never used to prosecute anyone, this law created federal jurisdiction for seeking and arresting violators of state law, who fled across state lines. Along with the Dyer Act making interstate transportation of stolen automobiles a federal crime, UFAP allowed the FBI to facilitate the return of both fugitives and stolen cars that would have otherwise presented a significant administrative problem for local police. In addition, the FBI routinely investigated bank robberies under federal jurisdiction because of government insurance and any kidnapping suspected of involving interstate movement. Practically speaking, the FBI often worked on kidnappings that involved no interstate movement. Many local agencies with limited resources were happy to turn most of these resource intensive cases over to the FBI.

Thus, three functions formed the basis for state and local police dependence on the FBI during the J. Edgar Hoover era: identification services, forensics and training. Concurrently, there were forces that generated competition and friction between the FBI and police. The FBI is charged with investigating violations of the federal civil rights laws. This sometimes places them in the position of investigating allegations against police officers. This became a particularly sensitive issue during the 1960s in the South when some police were directly involved in resisting racial integration (DeLoach). Another source of friction was Hoover's obsession with the publicity and fear of the FBI being blamed for the actions of other agencies. As a result, the FBI routinely announced the arrest of fugitives or other offenders arrested by local police and resisted working too closely with police investigators on crimes in which both agencies had jurisdiction. Many police officers tended to view the "feds" as either taking over their investigations or taking credit for their work. In addition, some prominent figures in policing, such as Chief William Parker of Los Angeles and Superintendent OW Wilson of Chicago, resented Hoover's high public profile as the "nations leading law enforcement officer."

The Police Pursue an Improved Image

Both the FBI and police reformers had adopted the ideas of the Progressives, who advocated de-politicizing governmental operations, recruiting the best personnel through competitive hiring, increasing efficiency of operations, standardizing procedures, adapting technology and providing managers authority to run their agencies. While the FBI received extensive laudatory publicity in magazines and books such as *The FBI in Peace and War* and films like *The FBI Story* (Powers 1983), local police reformers labored largely in obscurity. Throughout the first half of the twentieth century police were portrayed in the Boston Blackie and Thin Man films as well as print media as either irrelevant or incompetent in solving crimes. The literary tradition of private detectives being the solvers of crime the police were too dense or unimaginative dated back to Arthur Conan Doyle. The movies followed this pattern, but began to make exceptions for federal agents in movies like *G Men*, produced in 1935. In the period after World War II, FBI agents became the epitome of the clean cut, honest American hero in the popular media.

The first to break out of this pattern was William Parker. Parker assumed the position of chief in the LAPD in 1950. He instituted a program of reform reminiscent of Hoover's. Training and hiring standards were improved. Compliance with department rules, physical fitness and appearance were strictly enforced (Cannon). Unlike small suburban departments like Berkeley, LAPD was far more visible. This visibility was greatly enhanced by fortuitous geographic location, advancing communications technology and social change. In the early 1950s, America was shifting from radio to television as a primary source of home entertainment. One of the earliest scripted dramas to attain popularity was based on LAPD detectives. *Dragnet* was originally created by a young actor and producer, Jack Webb, as a radio show but soon made the transition to television (Cannon). Chief Parker provided support for the production on the condition of script review. The resulting series was the first of what would become known as police procedurals. Two LAPD detectives were portrayed as working diligently to solve crimes ranging from car theft to murder. As a result, Americans were presented with a very different view of local police than had been traditionally presented in films. LAPD went on to provide support and technical advice to numerous television series and films. Although the FBI's influence can only be implied, the pattern was clear. Law enforcement emulated the FBI's systematic efforts to shape an image of honesty, courage and technical efficiency through the popular media. This recognition of the importance of both media and public image, likely owe much to the successful model demonstrated by the FBI.

The Relationship Evolves

The relationship of the FBI with other law enforcement agencies underwent a significant change after the death of J. Edgar Hoover in 1972. The forces for change had already begun to reshape the relationship, but the rate of change greatly increased after Hoover died. Hoover had dominated the agency for almost 50 years. Virtually no agency head had every exercised such long or total control of a federal agency as Hoover (Lewis 1980). During the 1960s, local law enforcement had matured technically. Numerous states and local police agencies had opened their own forensic laboratories. In 1969 the Law Enforcement Assistance Administration (LEAA), an agency of the Justice Department, began granting money to states and local governments for the upgrading of police and other justice functions. As a result of LEAA support, states began to adopt police officer standards and training (POST) commissions, which established training requirements for new police officers (Walker and Katz). The POST model, developed in California during the late 1950s, (California POST) spread to all 50 states by the early 1970s. The combination of required training and selection standards and increased funding from LEAA radically accelerated the transition of police agencies to a more professionalized model. LEAA also provided forgivable loans for college expenses to those who entered law enforcement. Many police officers took advantage of this funding to return to school, and criminal justice programs in colleges and universities proliferated. Although the expansion of POST basic academies and advanced training increased the demand for FBI instructors, these changes also made the FBI only one of several entities engaged in training and professionalization of police.

Although Hoover had carefully built his relations with police and nurtured the dependence of police on the FBI, he had always kept his agents at arm's length enough from other law enforcement agencies to assure the FBI could retain full control over all its investigations and avoid any guilt by association from the actions of others. Although this approach served to protect the FBI from potential criticism and negative publicity, it also engendered a degree of mistrust and alienation. Police officers often expressed the view that FBI agents expected to receive information from local police, but would seldom share information. Although individual Special Agents often worked conscientiously to overcome this impression, it was wide spread.

After Hoover's death, power in the FBI became more decentralized, and procedures began to change. Hoover was replaced briefly by L. Patrick Gray, who resigned after failing to receive Senate confirmation as a result of accusations of aiding in the Watergate cover up. In 1993 Clarence Kelley was appointed director. Although Kelley spent 21 years as an FBI Special Agent and rose to the position of Special Agent in Charge, he subsequently served 12 years as the Kansas City, Missouri, Chief of Police. Under Kelley the FBI shifted its law enforcement emphasis to organized and white collar crime, began working undercover operations and began to work more cooperatively with other law enforcement agencies (FBI).

Hoover's death also marked a turning point in the FBI's image. Hoover had been largely successful in controlling any negative publicity regarding the agency and engineered substantial positive portrayals (Vizzard 2008). The combination of the Watergate investigations and later investigations by the Senate into intelligence gathering broke down the historic reticence of the media to criticize the FBI. Although this may have been viewed as negative within the agency, it removed the FBI from its pedestal; it likely reduced the police perception of the FBI as a privileged group somewhat separate from the rest of law enforcement. In subsequent years the FBI has been the subject of several serious exposés, including the conviction of Special Agents John Connolly for aiding Boston organized crime figure Whitey Bulger and Robert Hanssen for espionage, as well as criticism for numerous actions, including the Ruby Ridge and Waco sieges. However, the FBI has continued to thrive as an agency and build bridges to the police community.

A Partnership Evolves

Since the late 1960s the FBI has supported law enforcement through operation of the National Crime Information Center (NCIC). Although NCIC is governed jointly by state and federal representatives, its operation depends on the FBI. This system provides law enforcement with computerized, online indices of stolen property, wanted persons and criminal records. In addition, the system has been enhanced to provide computerized photographs and other identifying data. The FBI continues to maintain the national fingerprint files on arrested persons, and has established a centralized DNA file. The fingerprint file can be searched for a match with a single latent fingerprint. These centralized identification and information functions with records on millions of previously arrested persons, wanted felons, stolen property and missing children provide a critical link in the investigative and law enforcement functions. As has been the case from the inception of the FBI's fingerprint file, the FBI is as dependent on police agencies for submission of information as the police are on the FBI. In addition to continuing forensic science support, the FBI has provided police with investigative assistance from the National Center for Analysis of Violent Crime (NCAVC) which popularized behavioral analysis, or criminal profiling (FBI).

Although forensic advances such as DNA analysis and single fingerprint search have proven far more important in the solution of crime, behavioral analysis has captured the public imagination as criminal profiling. By applying inductive logic to evidence from the investigation using past offender patterns, behavioral profilers seek to describe identifying demographic characteristics and predict behavior of offenders based on their past patterns of activity. Because this process is expensive and only applicable in cases where significant unique information about the offender's actions can be developed from the investigation, most police agencies do not find it practical to employ behavioral analysts. The NCAVC also serves as a national clearing house for information on serial violent offenders and provides alerts to police via its Violent Criminal Apprehension Program (ViCAP). A number of localities have emulated these latter functions with regional centers that track violent serial offenses on a state or regional level (New Jersey State Police; New York State Police).

In addition to the National Academy, identification files and the laboratory, two other early FBI efforts at creating links to police have proven successful, the Uniform Crime Report (UCR) and the *FBI Law Enforcement Bulletin*. The UCR is a compilation of crime and arrest data from police agencies throughout the country. For many years the data in this report has constituted the "crime rate" for most of the public, policy makers and the media. Comparison of the UCR figures with those of the National Institute of Justice's National Crime Victim Survey (NCVS) clearly reveals that it does not fully count crime in the country, as many crimes are not reported to police (Clinton). However, it does reflect trends and has one significant advantage over the more accurate NCVS. The UCR data is reported by political jurisdiction; whereas the NCVS is national data drawn from sampling. Thus, the UCRs have long been a focus of police, who are organized by geography.

Neither police managers, local politicians nor the public are as concerned with national crime rates as they are with crime in their own community. Although the UCR does not fully report crime, it does allow for calculation of trends over time. These have had significant influence on police staffing, funding and administration, although the FBI is largely a passive agent in the process. Virtually all police agencies now track their own data, but are only able to compare that data to other political subdivisions by using the UCRs.

The *FBI Law Enforcement Bulletin* was begun in 1932 and is likely the most widely read publication among police. As with other FBI activities, it owes this largely to the fact that it has long been distributed widely to police agencies and officials at no cost to the recipient. Although the journal makes no pretense of being a scholarly journal, it provides readers with articles on law and police and investigative procedures. Articles are practical and applicable to daily police activities, thus likely to be read by working police officers.

Not all FBI influence on police results from the activities of FBI Headquarters. In the years since Hoover's death, the FBI has greatly increased its cooperative interactions with other law enforcement agencies. A variety of FBI task forces and fusion centers operate with FBI Special Agents and analysts working jointly with state or local police on investigations and information analysis. These task forces focus on areas such as violent gangs, computer crime, terrorism and organized crime and are often housed in a location away from the FBI office. Task forces allow the FBI to capitalize on the much wider coverage of a community enjoyed by police, who remain in one location and are far more numerous than FBI Special Agents. They also allow the FBI to largely shape the mission by providing resources and expertise not available to many police agencies. By working side by side with police officers, Special Agents build personal relationships which reduce the potential for police developing suspicion or resentment of the FBI. Although these joint operations have been widely welcomed by many police agencies, frictions have developed in some instances, particularly around the joint terrorism task forces (Levitt; Parascondola; Portland on Line; Yardley).

A final area in which the FBI has significantly influenced police is in the area of organized crime investigation. Until the late 1950s, the official position of the FBI was that the Mafia did not exist (Gentry; Sullivan). However, public disclosures of a disrupted 1957 conference of crime bosses in Apalachin, New York and the later testimony of former Mafia soldier Joseph Valachi resulted in a reversal of that position and the initiation of a major effort at investigating organized crime (Gentry). A combination of expanded federal statutes, significant resources and innovative investigative techniques has allowed the FBI to significantly disrupt organized crime, particularly in New York (Freeh). Only the FBI has the resources, jurisdiction and Justice Department support to pursue many of these cases. They have therefore had substantial influence over the investigation and prosecution of organized crime.

Thus the FBI has had a continuing influence on state and local policing in the United States, although the mechanisms and nature of influence have evolved over the past 80 years.

Related Entries

- ▶ Behavioral Investigative Advice
- ▶ Control of Police Misconduct
- ▶ Criminal Investigative Analysis
- ▶ Democratic Policing
- ▶ Development of the UCR and the NCVS
- ▶ DNA Profiling
- ▶ DNA Technology and Police Investigations
- ▶ Forensic Science
- ▶ Forensic Science Professions
- ▶ History of Criminal Investigation
- ▶ History of Organized Crime in the United States
- ▶ History of the Police Profession
- ▶ Managing Innovation of Policing
- ▶ Police Corruption
- ▶ Police Culture

- ▶ Police in Nazi Germany
- ▶ Police Integrity
- ▶ Police Leadership Styles
- ▶ Police Legitimacy and Police Encounters
- ▶ Police Lying and Deception
- ▶ Policing the Police
- ▶ Scientific Basis of the Forensic Process
- ▶ Sex Offenders and Criminal Policy
- ▶ Use of Social Media in Policing

Recommended Reading and References

California POST. http://post.ca.gov/

Cannon L (1997) Official negligence: how Rodney King and the riots change Los Angeles and the LAPD. Time Books, New York

Clinton R (1999) Victims and offenders in two crime statistics programs: a comparison of the national incident-based reporting system (NIBRS) and the national crime victimization survey (NCVS). J Quant Criminol 15:193–205

Collins F (1943) The FBI in peace and war. G.P. Putnam's Sons, New York

Coulson D, Shannon E (1999) No heroes: inside the FBI's secret counter-terror force. Pocket Books, New York

Cunningham D (2004) There's something happening here: the new left, the Klan and FBI counterintelligence. University of California Press, Berkeley

Deloach C (1995) Hoover's FBI: the inside story by Hoover's trusted lieutenant. Regnery, Washington

FBI, History http://www.fbi.gov/about-us/history/brief-history; Support http://www.fbi.gov/about-us/cirg/investigations-and-operations-support; LEB http://www.fbi.gov/stats-services/publications/law-enforcement-bulletin ;Partnerships http://www.fbi.gov/about-us/partnerships_and_outreach/; Organized Crime http://www2.fbi.gov/hq/cid/orgcrime/casestudies.htm

Felt M (1979) The FBI pyramid from the inside. G.P. Putnam's Sons, New York

Freeh L (2006) My FBI: bringing down the Mafia, investigating Bill Clinton and fighting the war on terror. St Martin's Griffin, New York

Gentry C (1991) J. Edgar Hoover: the man and his secrets. W.W Norton, New York

Jeffeys-Jones R (2007) The FBI: a history. Yale University Press, New Haven

Kelley C, Davis J (1987) Kelley: the story of an FBI director. Andrews, McMeel & Parker, Kansas City

Levitt L (2011) Kelly vs. the FBI: round 17,000. New York confidential: an inside look at the New York Police Department. http://nypdconfidential.com/columns/2011/110307.html

Lewis E (1980) Public Entrepreneurship: Toward a Theory of Bureaucratic Political Power: The Organizational Lives of Hyman Rickover, J. Edgar Hoover and Robert Moses. Indiana U. Press, Bloomington, Indiana.

Morn F (1995) Academic politics and the history of criminal justice education. Greenwood Press, Westport

New Jersey State Police. http://www.state.nj.us/njsp/divorg/invest/vicap.html

New York State Police. http://troopers.ny.gov/criminal_investigation/NYS_VICAP/

Parascandola R, A Gender (2011) Head of joint terrorism task force nearly fired for holding back info to NYPD superiors. New York daily news. http://articles.nydailynews.com/2011-03-13/local/29139713_1_nypd-fbi-brass-shea

Portland on line, joint terrorism task force review. http://www.portlandonline.com/mayor/jttfreview/. Accessed 21 Jan 2011

Portland city hall roundup: documents show how JTTF agreement morphed. http://www.portlandonline.com/leonard/index.cfm?a=350365&c=27435. Accessed 26 May 2011

Powers R (1983) G-Men: Hoover's FBI in American popular culture. Southern Illinois University Press, Carbondale

Powers R (2004) Broken: the troubled past and uncertain future of the FBI. Free Press, New York

Revell O, Williams D (1989) A G-Man's journal: a legendary career inside the FBI-from Kennedy to the Oklahoma city bombing. Pocket Books, New York

Smith IC (2004) Inside: a top G-Man exposes spies, lies and bureaucratic bungling inside the FBI. Nelson Current, Nashville

Sullivan W (1979) The bureau: my thirty years in Hoover's FBI. W.W. Norton, New York

Vizzard W (1997) Shots in the dark: the policy, politics and symbolism of gun control. Roman and Littlefield, Lanham

Vizzard W (2008) The FBI, a hundred-year retrospective. Public Adm Rev 68:1079–86

Yardley W (2011) Portland, Ore, votes to rejoin task force after terrorism scare. The New York times. http://www.nytimes.com/2011/05/01/us/01portland.html

Fear and Punishment

Helmut Kury[1] and Annette Kuhlmann[2]

[1]University of Freiburg, Freiburg, Germany

[2]University of Wisconsin, Baraboo, WI, USA

Overview

Since the 1960s fear of crime and attitudes towards punishment have become increasingly pertinent to the field of criminology and criminal policy. The availability of new methodologies in the social science, such as victim surveys, made these phenomena measurable and thus allowed for comparisons between groups and across time. In some ways these research methods played an important role in creating these topics in the first place. Especially in the USA, but also in other countries, politicians discovered the use of fear of crime to govern and control people. Crime policy can be simple in so far as people who are "fearful" generally ask for harsher punishments of the offenders, particularly after severe cases of crime. Politicians argue for harsher laws and sentencing policies as the "solution," especially for violent or sex crimes. Studies have shown repeatedly that measurements of fear of crime and of attitudes toward punishment tend to substantially overestimate both traits in the population.

Introduction

During the previous four decades, fear of crime and, subsequently, attitudes toward punishment have become increasingly important research topics in criminology. This research served as a backdrop for discussions of punitiveness in Western Europe. Imprisonment rates, development of penal law, sentencing behavior of penal courts, and public attitudes toward punishment are key elements of punitiveness which exert strong influences on criminal-political decisions (Dijk 2011). The increase in the number of prisoners in some countries, particularly the USA, has to be seen as political decisions influenced by media reports of citizens' fear of crime, their attitudes toward punishment, and their demand for more severe penal policies. Empirical criminological research has documented these developments, and some studies have found that fearful citizens are more punitive; consequently, in conjunction with other variables, fear of crime creates harsher attitudes toward offenders.

This section provides an overview of the important research results on fear of crime and attitudes toward punishment from different countries. Both of these phenomena are complex, undertheorized, and often measured only through few survey items (see Matthews 2005). In western countries in recent decades, the broad discussion of fear of crime in criminology, but also in the media, has led some scholars to the conclusion "that fear of crime is now recognized as a more widespread problem than crime itself" (Bannister and Fyfe 2001, p. 808). Since the beginning of the 1990s, the ongoing discussion of related practical issues, such as community crime prevention in Germany, occurred against the background of an increasing awareness of fear of crime in the public. This discussion has to include a consideration of the problems inherent in the methodologies used so far in the measurement of fear of crime and attitudes toward punishment. As shown, the measurement of fear of crime is vulnerable to political manipulation. This contributes, as some authors point out, to a "governing through crime," discussed particularly in regard to the USA, but also in Western European countries (Baker 2010).

Empirical Research About Fear of Crime and Attitudes Toward Punishment

As Lee (2001, p. 467) points out: "Since the late 1960s the 'fear of crime' has progressively become a profoundly engaging field of study for criminologists and other social researchers." Meanwhile, fear of crime has become a subdiscipline of criminology with an extensive body of literature, special chapters in college textbooks, and summarizing publications. The discussion began in the United States in the context of race riots of the 1960s and 1970s and increasing problems with violence in large metropolitan areas, particularly in neighborhoods with predominantly ethnic minority populations. In the 1960s social scientists began to develop a new research method, that is, population surveys, to collect information regarding people's perception and emotional responses toward criminality, fear of crime, and attitudes toward punishment. It was relatively easy to formulate questionnaire items and to find people willing to respond since crime was and continues to be an intriguing topic. An increasing number of surveys on crime, victimization, fear of crime, and attitudes toward punishment among the public brought these topics to the center of attention. The increasing discussion about victims of crime led to the establishment of a new subfield, victimology, "But it was the victim survey that, by providing data about the extent and severity of such fear, pinpointed a new area for criminological enquiry" (Zedner 2002, p. 425). The problem with the construct of fear of crime can be linked to a weak cognitively anchored attitude, meaning that respondents are likely to have only a rather vague perception of "their" personal fear of crime or none at all. Therefore, when asked to answer items in a questionnaire regarding such issues, these people are responding from a basis of limited awareness or personal reflection.

By the late 1960s an increasing number of organizations began to ask individual people about their experiences with crime victimization and possible effects such as fear of crime or attitudes toward punishment focusing on particular groups such as the elderly, young people, women, or foreigners (Lee 2001, pp. 475f). These surveys revealed that the dark figure of crime is higher than previously assumed, and at the same time, they provided extensive new information about the "effects" of crime. The number of victim surveys increased particularly in the United States and regularly included questions regarding fear of crime and attitudes toward punishment. Already in 1965 the National Opinion Research Center asked a sample of as many as 10,000 households, followed by similar surveys conducted by the Bureau of Social Science Research in Washington 1967. Ennis (1967) discovered in the first nationwide representative survey in the USA results which were confirmed on the international level in subsequent decades – namely, that women and the elderly express more fear of crime than men or younger people. The US President's Crime Commission also engaged in crime/victim surveys with the result that the USA and other countries, such as Great Britain and the Netherlands, carried out regular National Crime Surveys. "Through these surveys the scope of public concern about crime was 'discovered' empirically for the first time" (Lee 2001, p. 476). Zedner (2002, p. 425) points out: "Ironically, carrying out victim surveys increases sensitivity to the risk of crime. Situating questions about fear within a survey may consequently elicit higher levels of anxiety than would otherwise be the case." Vanderveen (2011, p. 40) points out for the Netherlands that "'fear of crime' is rooted in statistics and surveys by the Dutch government: without statistics, there would be no 'fear of crime' in the Netherlands as we know it today." So the problem of fear of crime – and punitive attitudes toward punishment – was created, at least partly, by the research itself and by the manner in which the survey items were framed. The results of these surveys were discussed in the media, and so people "learned" how fearful of being victimized they really are. The media realized that such reports interest people, so they presented this information often in a distorted, sensationalized manner.

At the end of the 1980s, Dijk et al. (1990) founded the first international victim survey, the "International Crime Victims Survey – ICVS." This new data set represents significant progress since we now have reliable international comparative data available that include the dark figure of crime as well as fear of crime and attitudes toward punishment. So far this survey has been carried out five times, in 1989, 1992, 1996, 2000, and 2005 in a number of different countries using the same basic questionnaire and research methodology (Dijk et al. 2007).

As a result of these different ICVS surveys, extensive data sets on fear of crime and attitudes toward punishment are now available not only from the western industrialized world, but also from developing countries. Nevertheless, Arnold (1991, p. 87) cautions: "Although fear of crime has interested social scientists for over two decades, and despite the fact that considerable research work has been done in the field of correlates of fear of crime, results have been inconsistent, inconclusive, and far from unequivocal." The main reason for the different and partly controversial results in the data on fear of crime as well as in attitudes toward punishment can be seen in the unclear definition of these concepts and the differences in measuring it. “Even the understanding of the term fear of crime and how it should be indicated or measured is not commonly accepted” (Arnold 1991, p. 88).

Zedner’s (2002, p. 425) observation that “… although fear of crime is closely related to levels of crime and tends to increase as crime rises, it cannot be seen as a mere function of crime rates” has to be further specified. In fact, the correlation between fear of crime and crime rates has been shown repeatedly as being relatively low, with the exception of countries or cities with high rates of (violent) street crimes. On the background of the worldwide results of the International Crime and Victimization Surveys, Zvekic and Alvazzi del Frate (1995, p. 28) point out that “there is some correlation between the rates of assaults/threats and robberies and fear of crime… the higher the crime rates are, the higher is the feeling of insecurity” (here in regard to Sub-Saharan Africa and Latin America). This is not surprising as, generally, people are informed about developments of criminality through media which tend to present distorted and sensationalized views, focusing on extreme and rare cases of violent and/or sex crimes. Only in regions or cities where people experience victimization as part of their daily lives do we find a strong correlation between crime rate and fear of crime. For instance, Zvekic and Alvazzi del Frate (1995, p. 27) have found such results for specific cities with high rates of street crime, such as Rio de Janeiro or Johannesburg.

Dijk and Vollaard (2012) analyzed data from different waves of the International Crime Victims Survey (ICVS) for 15 countries and compared the actual rate of victimization of burglary against the perceived risk of burglary, that is, a cognitive aspect of fear of crime, for that year. These data provide the foundations for a model of “the aggregate crime rate as the outcome of an interaction between potential offenders and potential victims,” focusing on domestic burglary. The results of the study show that fear of crime is strongly related to the actual crime rate for burglary. They found a strong positive correlation between the rate of victimization of domestic burglary and the perceived risk of victimization. This relationship can be shown both across the 15 countries and over time within the different countries. Increasing rates of burglary had the effect of an increasing investment of people in home security. These higher levels of home security, in turn, are related to lower percentages of successful burglary attempts in the future. Therefore, the authors conclude that “precautionary behavior is an important factor in understanding aggregate crime trends.” A reduction in officially registered crimes can be a result of more effective preventive measures by citizens. “The empirical evidence suggests that regulating built-in security is not only an effective policy but also a relatively cheap way of reducing crime compared to more traditional crime policies such as strengthening law enforcement or rehabilitation of offenders” (Dijk 2008).

The concepts of fear of crime and attitudes toward punishment, as operationalized in most research, are influenced by a number of different variables. A representative survey, for example, from East and West Germany in 1991/1992, that is, just after the reunification of both parts of the country, showed that the fear of crime rate in the Eastern part was substantial and significantly higher than the rate in West Germany. At the same time the official crime rate in East Germany was about one third to one half of that in the western part. The results show clearly the weak influence of crime rates on fear of crime. Over subsequent years the official, registered crime rate increased significantly in the Eastern part of Germany until it reached the rate of West Germany in 2000 – at the same time the fear of crime rate decreased noticeably. Obviously other factors besides crime rates are influencing the fear of crime.

The same case can be made for attitudes toward punishment. The former German Democratic Republic (GDR) was, like other former Soviet states, rather punitive in comparison to West Germany or other western countries; the state propaganda taught people that it was necessary to be harsh toward offenders to reduce crime. The death penalty was abolished in GDR in 1987, in West Germany 1949. In contrast to western countries, only few alternatives to harsh sanctions, such as crime prevention or victim-offender restitution, existed and were seldom used. As Niggli (2004, p. 192) points out, punitive attitudes in the general public are to a large extent shaped by the values in the society, the prevalent practices of sanctioning, and the character of the public debate about the problems in society surrounding crime. According to Niggli, the new punitive turn in many countries is a response to far-reaching changes in the context of globalization, particularly the opening of borders in Europe and increased migration. In recent decades new problems emerged that led to a discussion of western societies as "risk societies." These problems include changes in the employment structure, an increasingly extreme income disparity, and a decline of the social security system (p. 196).

After German reunification western penal law was introduced in former East Germany which implied more lenient sanctions and corresponding reports in the media. The people had new experiences. Although the GDR experienced a significant increase in crime rates in the beginning of the 1990s, a decade later, fear of crime indicators as well as attitudes toward punishment had changed and become more lenient. However, some interesting exceptions illustrate the salient influence of public debate. Over the last three decades, a heated public discussion about rape in matrimony emerged, and stricter laws were enacted as a result. This was not the case in GDR. So after reunification West Germans were more punitive than the East Germans in this regard. Subsequently, East Germans were influenced by the media about this discussion and became more punitive as well. This development stands in contrast to a comparable debate on illegal drugs. Possession had been punished severely in the former GDR, whereas West Germany had more lenient penal policies, especially in regard to soft drugs. After reunification, as a result of public debate, East German attitudes also became increasingly lenient in this issue.

In contrast to countries like the USA, Great Britain, or the Netherlands, Germany has no regularly conducted representative crime or victimization surveys. There are only individual surveys with different samples and methods, so the results are only partially comparable. But there is a private insurance company which has conducted such representative surveys on "Fears of the Germans" using the same methodology every year since 1991. The survey inquires about different fears, including fear of crime (R+V Infocenter 2012). The results show that since 1991 the fear of crime is higher in the former East Germany than it is in the west of the country and this fear has been increasing over the last two decades. In 1991 24 % of West Germans indicated any fears (like unemployment, political situation, illness, or crime), while the rate for East Germany was 28 %. In 2011 the values were 42 % in West and 47 % in East Germany.

It is striking that out of the 16 registered fears, fear of crime has, over time, moved into the background, while other fears, such as increases in costs for living, unemployment, insufficient funds for retirement, or politicians' inability to solve the country's problems, moved into the foreground. In 2011, fear of crime, out of all fears, placed on the second to last position on this scale. So while other fears increased, the significance of fear of crime, as part of an overall sense of anxiety, diminished. Yet this interpretation of these empirical results requires caution. Zarafonitou (2011, p. 61) discusses Greece which, according the EU-ICVS, had a particularly high rates of fear of crime over the last two decades compared to other EU states. She points out that "significant social changes have occurred" in this country, "the most important of which were the mass entry of immigrants as well as the recent economic crisis."

During the Soviet era, former Soviet states had, in comparison to western countries, a significantly lower (official) crime rate, more punitive penal laws, and more severe sentencing policies. Also, the population had more punitive attitudes toward offenders; at the same time, the fear of crime rate was regularly lower than in western countries. After the collapse of the Soviet Union, the (official) crime rate increased in most countries, while the economic situation worsened, that is, the unemployment rate in many of these countries increased while the helping – as well as controlling – function of the state declined. With this background the fear of crime rate also increased significantly.

These results indicate that fears are more focused on general fears about the economic and political situation in a country, the living conditions, etc. and that fear of crime, as we measure it in victim surveys, is a mixture of these different fears – only part of it really is a fear of crime. Against the backdrop of the dramatic changes in the former Soviet countries, these fears are higher in those countries than in western countries, while their actual crime rate might be lower. The fact that East Germans voted for harsher penal laws to fight crime might not necessarily be interpreted as a higher level of punitiveness in the population, but rather as an expression of general dissatisfaction with the political and social situation.

As Krajewski (2006, p. 490) points out, the punitive attitudes in postcommunist countries are based on the strict sentencing practices of the courts as well as the generally conservative, rigid, and intolerant attitudes of the former Soviet system, which the regime presented to the public via the mass media. It is often overlooked that the "will of the society," which continues to be cited by communist politicians as the reason for their partially absurd ideas regarding penal policies, is a result of the propaganda in the media regulated by them (Krajewski 2006, p. 490). Klaus et al. (2011, p. 267) point out, "that communist criminal law was exceedingly punitive and penalisation was a basic means to counteract undesirable social problems." Reports about crime were seldom published in the media of the former Soviet countries, and, if so, the reports were one-sided and biased. Such politics promote the development of punitive attitudes.

A more punitive attitude toward offenders in the former Soviet countries is also illustrated in a number of different articles in a collection by Kury and Shea (2011). Krajewski (2006, p. 490) points out that intolerance and punitive attitudes are discussed in the scientific literature only very broadly. With regard to Poland he shows that according to surveys in 1987, 74 % of the population felt their country was safe. In 2004, this rate declined to 33 % (p. 489). According to the results of the ICVS (Dijk et al. 2007), Poland participated regularly in different waves of these recurring surveys. Since 1992, the surveys revealed increasing punitive attitudes in the population. In 1992, 31 % supported sentencing a juvenile recidivist burglar to a prison term. In 1996 the value declined to 17 % but increased again in 2005 to 34 %. In 1964, 50 % of the population supported the death penalty for severe offenders; in 2002 this percentage increased to 77 %. The new penal law from 1998 abolished the death penalty, but since then this new law has been strongly criticized by the media and by members of different political parties. This widespread and harsh critique continues today in the context of very conservative politics. The opponents of the new regulation postulate more repressive responses toward offenders similar to the practices in the USA and Great Britain. Politicians and the media publicize these demands regularly which exert an influence on the population. Consequently, in 1993 67 % of the public supported a reintegration of offenders into society as a central theme of crime policy. But in 2002 the percentage declined to 17 %. Likewise, 1993, the protection of human rights for prisoners was supported by 40 % of the public, but declined to 6 % 2002. According to Krajewski (2006, p. 504), there is no doubt that today the Polish criminal justice system is among the most punitive in Europe. Danilov et al. (2011) report a punitive tendency in Russia. Saar (2011, pp. 345f) points to high levels of punitive attitudes in Estonia. More specifically, the actual crime rate in Estonia increased significantly between 1991 and 2001, but decreased again from 2002 to 2009. At the same time, the fear of crime rate in Estonia decreased from 1993, when 49 % of those surveyed felt unsafe walking alone in the street at night, to 28 % in 2009 (p. 353). These results show high rates of punitiveness in these former Soviet countries which do not correlate with the actual crime rates. The penal practices of a country, that which are considered "usual" in sentencing and frequently discussed in the media as being a necessary reaction to crime, have an effect on the attitudes of the public. A conservative and restrictive Polish government has been asking for harsh punishment on crimes. It requires time, new experiences, and valid information to change the attitudes of people.

Criminologists were able to show in their research that a more detailed understanding of the background of criminal behavior and of offenders reduces harsh attitudes toward felons. Hough and Park (2002) in their research project asked people about possible ways to reduce crime. After completing a preliminary survey, the research subjects participated in a 2-day seminar where underlying reasons for crime and criminal behavior were discussed. Subsequently, the participants retook the survey and did so again 10 months later. In both times their punitive attitudes had declined substantially compared to the initial survey. This illustrates again the conclusions that more knowledge about crime and criminals reduces punitive attitudes.

In spite of a worldwide trend to reduce the use of the death penalty, some countries still practice it, among western countries most notably the USA. In Europe, only Belarus continues to use this form of punishment with support of the public. In the USA 34 states allow the death penalty, but the number of persons executed has declined from 99 in 1999 to 43 in 2011. A 2010 poll by Lake Research Partners "found that a clear majority of voters (61 %) would choose a punishment other than the death penalty for murder." The majority supported alternatives to the death penalty: 33 % voted for capital punishment, 39 % for life without parole plus restitution, 13 % for life without parole, and 9 % for life with parole (6 % no opinion) (Death Penalty Information Center 2012, p. 4). That means 52 % of the participants in the US survey voted for life imprisonment without any chance of release. Critics point out that life without parole is also an exceedingly cruel punishment, comparable to the death penalty. Consequently, the USA has to be seen as a particularly punitive country, in spite of the reduction in the number of executions in recent years. A disproportionate number of those executed are poor and African American.

While at the beginning of the twentieth century nearly all countries used the death penalty, today only 58 of them have such legal provisions. In 2010 only 23 countries worldwide actually executed people. A Gallup International survey from 2000 asked samples of the population in 59 countries about their attitudes toward the death penalty. Only 52 % of all persons surveyed supported capital punishment. Citizens from regions where the death penalty was used more often also supported capital punishment at higher rates: 66 % in North America, 63 % in Southeast Asia, 60 % in Eastern Europe, 54 % in Africa, 37 % in Latin America, and 34 % in Western Europe where this practice is fully abolished.

In West Germany the death penalty was abolished 1949. At that time, according to a 1996 survey by the Institut fuer Demoskopie Allensbach (1996), 74.8 % of the population supported the death penalty. Over subsequent years the support for the death penalty among the West German population declined to about 50 % in the 1960s and dropped further to about 30 % in the beginning of the 1970s. In the context of the murders committed by the Red Army terrorists, support for death penalty peaked at nearly 45 % by end of the 1970s and then declined again to between one quarter and one third today (Kury et al. 2004, p. 65).

The ICVS presents comparable data about fear of crime and attitudes toward punishment from a number of different countries. Zvekic and Alvazzi del Frate (1995) discuss the results for developing countries. According to their data in the 1990s, the "safest regions are Asia and North Africa, while the highest levels of insecurity were registered in Sub-Saharan Africa… and Latin America…" (p. 27). They also found "some correlation between the rates of assaults/threats and robberies and fear of crime… the higher the crime rates are, the higher is the feeling of insecurity (Sub-Saharan Africa and Latin America). However, high levels of precautions adopted when going out after dark (avoiding certain areas that are considered dangerous and asking somebody for company) do not necessarily mean that respondents feel unsafe and are also adopted in the low-crime regions. They may just as well reflect the cultural pattern of an outgoing style which is not directly related to fear of crime" (p. 28). This finding again calls attention to methodological problems in measuring fear of crime.

In regard to attitudes toward punishment, Zvekic and Alvazzi del Frate (1995, p. 48) summarize that "there is a high degree of agreement among the population in the developing world that the most appropriate sanction is imprisonment: more than 50 % in all the regions, and even more than 70 % in Sub-Saharan Africa and Asia." Despite some countries' favor for community service, like in Argentina and Brazil, "Punitiveness prevails in the developing world" (p. 48).

An analysis of ICVS results for 18 EU countries since 1992 found an average of 26 % of the sample, across all surveys and all countries included, felt unsafe walking in the street after dark in their own residential area (Dijk et al. 2007, p. 66). In 2005 this rate was more or less the same at 28 %. But the differences between the countries are immense. The highest rate of fear was found in Greece (2005) where 42 % expressed fear of crime. Also very high rates of fear were registered among the population in Luxembourg (2005, 36 %), Italy (2005, 35 %), and Portugal (2005, 32 %). On the other end of the spectrum were Finland (2005, 14 %), Denmark (2005, 17 %), and the Netherlands (2005, 18 %). In Estonia, Poland, the Netherlands, and Finland, the fear of crime rate decreased starting in the 1990s, but in the United Kingdom it increased (2000, 25 %; 2005, 31 %); the same occurred in Belgium (1992, 20 %; 2005, 26 %) and Sweden (1996, 11 %; 2005, 19 %).

In regard to attitudes toward punishment, the average percentage of the public in the different countries opting for imprisonment of the young recidivist burglar in 2005 was 24 % and for community service 49 %. So the majority supported alternative sanctions for the offender. But the difference between the countries was again very high. In the United Kingdom 52 % of those surveyed agreed with a prison sentence, while this percentage was only 38 % in Ireland and 34 % in Poland (here including labor camp). On the other end of the scale, the percentage in France was only 13 %, the same as in Austria, and 15 % in Portugal, while Germany's was 19 % (see also Dijk 2011, p. 218). At the same time, the rates for community service were highest in 2005 for France (68 %), Portugal (68 %), and Luxembourg (68 %) and lowest for Greece (27 %), United Kingdom (29 %), and the Netherlands (37 %). The development of community service orders shows some shift over time. "For instance, those in the Netherlands in 2000 were less in favor of a community sentence than they were in 1989. In contrast, there was more support in Belgium and Finland in 2000 than in 1989. Between the 1996, 2000 and 2005 sweeps, though, there was little change" (Dijk et al. 2007, p. 87).

Dijk et al. (2007, pp. 89f) compare national attitudes toward sentencing with rates of imprisonment for different countries. "In the Western world, those countries where the public clearly favors imprisonment, such as the USA and the UK, tend to have comparatively higher prisoner rates." The authors found on the basis of the ICVS data within the EU context, "a very weak and statistically not significant relationship between public opinion on sentencing and the actual level of prisoner rates. The three new member countries, Hungary, and especially Poland and Estonia, stand out with prisoner rates far above the EU average while public attitudes in these countries are only slightly above the middle range. In these transitional countries public attitudes have shifted over the past 10 years away from imprisonment toward community service orders. Public opinion in these countries is now broadly in line with the EU majority point of view. Actual sentencing policies seem still to be comparatively punitive, although... levels of conventional crime are not excessively high in any of the three countries."

Criminological research shows clearly that fear of crime rates and attitudes toward punishment have different trajectories in different countries, and both are particularly influenced by the political debate, the economic situation, and media reporting on crime – but also by the way of measurement.

Measurement Problems

Fear of crime and attitudes toward punishment are exceedingly complex concepts, yet they are measured in many cases with one or a few items only. The “standard item” for fear of crime is: “How safe do you feel walking alone in your area after dark?” Here crime is not mentioned so the background for the response can include other causes such as darkness or aggressive animals. But other surveys used different forms of assessment (Farrall and Ditton 1999, p. 58) such as the following: “In your everyday life, are you afraid of someone breaking into your home?” When the questions are framed so differently, the results are hardly comparable.

Attitudes toward punishment are also often measured by asking about the level of acceptance of the death penalty or by presenting a special criminal case and asking about suggested reactions (see the ICVS, Dijk et al. 2007, p. 98). Empirical explorations of complex concepts such as attitudes toward punishment cannot lead to any degree of precision nor provide any adequate results when assessed with one or a few items only. Even if the results of such polls do tell us something about the attitudes toward punishment, or at least about certain of its aspects, it is not clear to what extent these results apply. They certainly do not mirror “the” attitudes toward punishment, even if they are often presented as such (Kury and Obergfell-Fuchs 2008, p. 280). These items only embrace single aspects of the unclearly defined concepts. Mostly the authors of studies in this field lament the methodological problems, but nevertheless use the old items more or less in a similar way. Usually the questionnaires have to be short, so one or a few items are used to measure the concepts. Under these circumstances it is not surprising that the results of these surveys are often contradictory.

Farrall et al. (1997) have pointed out that the customary standard polling techniques, based on standard, closed questionnaires, have led to a significant overrating of the fear of crime. In particular, comparative qualitative studies, providing more thorough analyses of the indicated fear values, have called into question the results of quantitative questionnaire surveys. Farrall et al. (1997, p. 658) remark that from the outset fear of crime as well as victimization experiences have been almost exclusively the domain of quantitative studies, leading to results that suggest “that fear of crime is an overriding social problem.” It is only in recent years that these results have been increasingly questioned by more differentiated studies. On the background of their own research, using quantitative and qualitative methods, Farrall et al. (1997, p. 3) conclude that closed inquiry systems as used by most polls, especially by big-scale surveys, consistently and substantially overestimate fear of crime. Based on a comparative survey in Germany, Kury and Obergfell-Fuchs (2008) came to the same conclusions and were able to validate the English results. In another research project they found similar results concerning attitudes toward punishment (Kury et al. 2009). About half of the sample which had been classified as punitive based on a survey with a standardized questionnaire appeared nonpunitive in the face-to-face interviews. The research subjects responded with higher levels of punitiveness on the questionnaire to express other problems, which were outside the focus of the survey.

Some authors found a “risk-fear paradox” indicating that some population groups, such as women and the elderly, have low victimization rates but high levels of fear of crime, while men and the young generally have relatively high levels of victimization, but low levels of fear of crime. Some authors explain this paradox as a consequence of measurement problems. “Historically, victim surveys have produced problematic data that suggested a weak correspondence between fear and risk” (Zedner 2002, p. 426). Instead, over the last several decades, researchers have increasingly “attempted to identify the social structural causes of people’s fear” (Zedner 2002, p. 426).

Increasing numbers of studies over the last several decades have shown that “fear of crime” is related to general fears or feelings of insecurity which generally are not assessed separately in surveys consisting of standardized questionnaires.

What we call "fear of crime" is obviously a mixture of different types of fear, feelings of insecurity, worries, anxieties, or perception of various risks. "Fear of crime" emerges not only through reports about or experiences of (severe) crimes, but also through the socioeconomic situation in the neighborhood, living conditions, or signs of increasing incivility. "Fear of crime may operate as a 'sponge,' absorbing all sorts of anxieties about related issues of deteriorating moral fabric, from family to community to society. People may use the language of 'worry' and 'crime' to express connecting conflicts, insecurities, and anxieties" (Jackson 2006, p. 261). As the German surveys by R+V Infocenter (2012) clearly show, since 1991 fear of crime has moved increasingly in the background and other "fears" related to social changes and living conditions in society increased. Zarafonitou (2011, p. 50) discusses the background of high rates of fear of crime in Greece in comparison with other EU countries and points out that the "research evidence in Greece reflects the association of citizens' insecurity with the perception of the quality of their everyday life as degraded as well as their dissatisfaction with the state services, and in particular with the police effectiveness, in this field. In this context, the interpretation of the examined phenomenon will be based on the fundamental assumption that a feeling of general social insecurity is expressed through fear of crime."

The same conclusion holds true for attitudes toward punishment. Here also the standardized polling of attitudes toward punishment, used in virtually all victim surveys, raises considerable methodological problems. Kury et al. (2004) proved in his experimental research project that survey results are influenced by the order and number of alternatives provided to a question about sentencing. In a randomized sample of three groups, each group was presented with the same questions about sentencing of different crimes. Only the possible options for responses to the questions differed. Group A was presented with five different sanction alternatives ranging from lenient to punitive and was asked to choose one of them (the options included private restitution, mediation by an official mediator, punishment with or without restitution). Group B was presented with the same questionnaire, but with the options listed in reverse order (from punitive to lenient). Group C was presented again with the same questionnaire as the first group, but with three additional options (no reaction to the crime, fine, imprisonment). The effects of the different methodologies on the results of the surveys were so pronounced that the results were not comparable, especially in criminal cases that were not severe. For example, in the case of theft, 72.4 % of group A voted for alternative sanctions, but only 20.8 % in group C did so; instead, 22.5 % of this group C voted for no reaction, an option not provided for the other two groups. In case of homicide 38.1 % of group A voted for alternatives as did 18 % of group B, but only 7.5 % did so in group C; again, their options for alternative measures were different. In the case of using public transport without paying for the ticket, 58.1 % of group A and 53.5 % of group B voted for alternative sanctions, in contrast to only 4.2 % in group C and of those 36.7 % opted for "no reaction." These findings show that the answers are influenced significantly by the way the items are presented; if we do not present "no reaction" as an option in the survey, the participant cannot inform us that he/she does not really suggest any alternative reaction, but to simply do nothing.

Political Relevance of Fear of Crime and Attitudes Toward Punishment

With fear of crime among the citizenry political actions can be initiated and justified, that is, the research results on this topic have tangible consequences. Fear of crime is, since we measured it, a very politicized topic. This can be seen in the development of the discussion and changing reactions to crime in the USA since the beginning of the 1970s (Lee 2001). Against the backdrop of a "perceived growing lawlessness in the USA," the public "was becoming anxious about the steadily climbing crime rate." Criminologists measured this development and "calls for tougher action in terms of policing, disciplining and punishing criminals began to grow" (Lee 2001, p. 476). All over the world and throughout history, the most favored reaction to crime has been (stricter) punishment. The USA had a dramatic increase of punitive reactions toward crime, and consequently, penal laws and sentencing policies became stricter with the effect that the incarceration rates climbed as did the costs of the criminal justice system. Meanwhile, the USA has the highest incarceration rate in the world.

In 1965 President Johnson set up the Commission on Law Enforcement and Administration of Justice (LEAA) "that was supposed to produce solutions to the crime problem in a non-political, non-partisan environment" (Lee 2001, p. 477). The commission was charged with investigating the reasons for the crime increase at the time and with discerning the societal factors responsible for it. The 1967 report of the commission, "The Challenge of Crime in a Free Society" (Harris 1969, p. 15), emphasized the "real" background of an increasing crime rate which was neglected in subsequent decades by politicians in favor of increasing stricter reactions to crime. "To speak of controlling crime only in terms of the work of the police, the courts, and the correctional apparatus is to refuse to face the fact that widespread crime implies a widespread failure by society as a whole" (Harris 1969, p. 16). The background of crime was seen here first and foremost as rooted in the social conditions, but over the following decades the causal attribution moved to the offenders themselves who were now seen as individually responsible for their behavior.

In the USA fear of crime, individual responsibility for (criminal) behavior and socioeconomic status, as well as penal policies are highly racialized. Young black males are particularly feared, yet they are also the most victimized group, that is, their homicide victimization rate is five times that of whites. Penal policies, particularly the war of drugs, have led to disproportionately high incarceration rates among this group; that is, it is currently seven times higher than that for whites, or one in three African Americans ages 20–29 is under some form of correctional supervision (Kuhlmann 2011). Sheldon (2010, p. 92) has called this new racial disparity in prisons "the New American Apartheid."

Fear of crime, combined with the historical mythology of the American West and corporate interests such as the National Rifle Association, contributed to constitutionally secured gun laws which allow individuals in many states not only to own and carry guns but to do so while concealing them. This mixture has proven particularly disturbing recently in the two shootings of black youths by private white citizens who felt threatened by the young men. It is unlikely that the white shooters will be convicted or even go to court.

Lee (2001) explains the development of the politics of crime over subsequent decades in the USA when politicians "learned" to use fear of crime as a tool for governing. It became popular to announce ever harsher punishment as a means to reduce the (severe) crime rate and to use the population's "fears of crime" to win elections. Thus, "governing through crime" became a pattern of penal politics not only in the USA but, as Garland (2001) demonstrates, also in Great Britain. The effect was that penal law became increasingly harsh in most western industrialized countries, a development which also applied to sentencing (the criminal justice systems in Eastern European countries were already much more severe) (see the chapters in Kury and Shea 2011). People ask for harsher punishment, politician creates stricter laws, and, as a result, people "learn" that they are right and the politicians have "understood"; this is a self-sustaining feedback loop (Lee 2001, p. 480). The resulting high levels of incarceration created economic and political interests groups, the prison-industrial complex, which includes not only construction businesses for the large number of new prisons but the producers of technologies, products, and services required to operate these. Defense industry giants like Westinghouse are retooling and lobbying Washington for their share of the domestic law enforcement market, for many rural area prisons are becoming the primary form of economic development, and advertising industries have found a new market. These businesses are so successful that they were identified as the sixth fastest growing industry in the USA in 2010.

Over the last few years, a debate erupted in Germany concerning the question of a "punitive turn," the possibility of a "new Punitiveness" in that country (Pratt et al. 2005). There are, in fact, a few signs for a "governing through crime," but overall, the attitudes toward punishment in Germany have been relatively stable over recent years; in comparison to previous decades, people express more lenient attitudes today (Kury and Obergfell-Fuchs 2011).

The discussion about more punitive attitudes toward crime regularly starts after particularly serious crimes, such as the cases in Oslo, Norway, or London, England, in 2011 or in Toulouse, France, in 2012. The Norwegian President reacted with a cautionary message of not allowing this crime to diminish the freedoms in the country and for the population and politicians to react quietly and rationally. In contrast, the prime reactions in Great Britain and France were to create stricter penal laws and to increase the penalties. Without considering if this reaction actually was the best way to reduce such types of crime, politicians assumed that this was what the people wanted to hear most – after all, the population generally is un- or misinformed and the politicians tell them what the right reaction is supposed to be.

Discussion

Beginning in the late 1960s, western countries have been involved in an expanding discussion about fear of crime. Evolving new research methodologies made comparisons between different groups and countries possible. Early on in this debate, politicians found ways to use this fear to further their own interests, especially to win elections. Crime policy in this context is simple: especially after the occurrence of a severe crime, a discussion starts about increasing punishment; people become more punitive and believe that crime can be reduced through more severe reactions to it.

Because fear of crime has such a strong impact on crime policy, it is important that criminological research results are valid and methodologically sound. The research experiences on this topic over the last decades reveal that the results for fear of crime and also for attitudes toward punishment are not always exact. Methodological studies show clearly that surveys with standardized questionnaires, often measuring fear of crime and attitudes toward punishment with single items, strongly overestimate the "fear of crime" rate and the desire for harsher punishment. These results are used to support wrong political actions, namely, harsher penal policies. What we measure with "fear of crime" is often more or less a mixture of different fears, insecurities, and worries about social conditions. If people are better informed about the underlying causes of crimes and the effects of punishment, as well as the resulting costs of imprisonment, fear of crime is reduced and attitudes toward punishment become less punitive. Citizens need to be informed that some level of crime is a "natural" aspect of each society and cannot be completely eliminated; instead, fears can also protect people from dangerous behavior. Criminologists have a responsibility to produce valid results to facilitate rational discussions about these topics. There needs to be more information available to the public as to the actual reasons for the prevalent crime rates as well as possible ways to reduce crime to rational levels. Stricter penal policies generally do not lead to this goal; instead people need to understand the connection between societal conditions and crime.

Recommended Reading and References

Arnold H (1991) Fear of crime and its relationship to directly and indirectly experienced victimization: a binational comparison of models. In: Sessar K, Kerner H-J (eds) Developments in crime and crime control research. Springer, New York/Berlin, pp 87–125

Baker E (2010) Governing through crime – the case of the European Union. Eur J Criminol 7:187–213

Bannister J, Fyfe N (2001) Introduction: fear and the city. Urban Stud 38:807–813

Danilov AP, Shestakov DA, Kury H (2011) Changes in criminal law and sentencing practices in Modern Russia: an assessment of new elements of punitiveness. In: Kury H, Shea E (eds) Punitivity – international developments, vol 2. Universitätsverlag Dr. N. Brockmeyer, Bochum, pp 509–550

Death Penalty Information Center (2012) Facts about the death penalty. DPIC, Washington, DC, http://www.deathpenaltyinfo.org/documents/FactSheet.pdf

Dijk JJM, Mayhew P, Killias M (1990) Experiences of crime across the world. Key findings from the 1989 international crime survey. Kluwer, Deventer/Boston

Dijk JJM, Manchin R, van Kesteren J, Nevala S, Hideg G (2007) The burden of crime in the EU. Gallup Europe, Brussels

Ennis PH (1967) Criminal victimization in the United States: a report of a national survey. GPO, Washington, DC

Farrall S, Bannister J, Ditton J, Gilchrist E (1997) Questioning the measurement of the 'fear of crime'. Findings from a major methodological study. Brit J Criminol 37:658–679

Farrall S, Ditton J (1999) Improving the measurement of attitudinal responses: an example from a crime survey. Int J Soc Res Methodol 2:55–68

Garland D (2001) The culture of control. Crime and social order in contemporary society. University of Oxford Press, Oxford

Harris R (1969) The fear of crime. Frederick A. Prager, New York

Hough M, Park A (2002) How malleable are attitudes to crime and punishment? Findings from a British deliberative poll. In: Roberts JV, Hough M (eds) Changing attitudes to punishment. Public opinion, crime and justice. Willan, Cullompton, pp 163–183

Institut für Demoskopie Allensbach (1996) Allensbacher Berichte Nr. 8. Institut für Demoskopie, Allensbach

Jackson J (2006) Introducing fear of crime to risk research. Risk Anal 26:253–264

Klaus W, Rzeplinska I, Wozniakowska-Fajst D (2011) Punitivity in polish law, public opinion, and penal policy. In: Kury H, Shea E (eds) Punitivity – international developments. Vol. 2: insecurity and punitiveness. Universitätsverlag Dr. N. Brockmeyer, Bochum, pp 245–268

Krajewski K (2006) Punitivität der polnischen Gesellschaft. In: Obergfell-Fuchs J, Brandenstein M (eds) Nationale und internationale Entwicklungen in der Kriminologie. Festschrift für Helmut Kury zum 65. Geburtstag. Verlag für Polizeiwissenschaft, Frankfurt/M, pp 485–506

Kuhlmann A (2011) Punitiveness in the US criminal justice system. In: Kury H, Shea E (eds) Punitivity – international developments. Vol. 3: punitiveness and punishment. Universitätsverlag Cr. N. Brockmeyer, Bochum

Kury H, Obergfell-Fuchs J (2008) Measuring the fear of crime. How valid are the results? In: Kury H (ed) Fear of crime – punitivity. New developments in theory and research. Universitätsverlag Dr. N. Brockmeyer, Bochum, pp 53–84

Kury H, Obergfell-Fuchs J (2011) Punitiveness – impacts and measurements. In: Kury H, Shea E (eds) Punitivity – international developments, vol 2. Universitätsverlag Dr. N. Brockmeyer, Bochum, pp 165–209

Kury H, Kania H, Obergfell-Fuchs J (2004) Worüber sprechen wir, wenn wir über Punitivität sprechen? Versuch einer konzeptionellen und empirischen Begriffsbestimmung. Kriminologisches J 36:51–88

Kury H, Shea E (eds) (2011) Punitivity – international developments, vol 3. Universitätsverlag Cr. N. Brockmeyer, Bochum

Kury H, Brandenstein M, Obergfell-Fuchs J (2009) Dimensions of punitiveness in Germany. Eur J Crim Policy Res 15:63–81

Lee M (2001) The genesis of 'fear of crime'. Theor Criminol 5:467–485

Matthews R (2005) The myth of punitiveness. Theor Criminol 9:175–201

Niggli MA (2004) Wie viel Strafe braucht der Mensch? Zur Strafdebatte in Europa. In: Stapferhaus Lenzburg (Hrsg) Strafen. Ein Buch zur Strafkultur der Gegenwart. hier + jetzt, Verlag für Kultur und Geschichte, Baden, pp 185–199

Pratt J, Brown D, Brown M, Hallsworth S, Morrison W (eds) (2005) The new punitiveness. Trends, theories, perspectives. Willan, Portland/Oregon

R+V Infocenter (2012) Die Ängste der Deutschen 2011. Eine Studie des Infocenters der R+V Vesicherung. R + V Versicherung, Wiesbaden, http://www.ruv.de/de/presse/r_v_infocenter/studien/aengste-der-deutschen.jsp

Saar J (2011) 20 years after: crime, crime control and fear of crime of the population in Estonia 1991–2010. In: Kury H, Shea E (eds) Punitivity – international developments, vol 2. Universitätsverlag Dr. N. Brockmeyer, Bochum, pp 343–356

Sheldon R (2010) Our punitive society: race, class, gender, and punishment in America. Waveland Press, Long Grove

van Dijk JJM (2008) The world of crime: breaking the silence on problems of security, justice and development across the world. Sage, Thousand Oaks

van Dijk JJM (2011) Trends in Dutch prisoner rates: regression to the mean or enduring exception? In: Kury H, Shea E (eds) Punitivity – international developments, vol 1. Universitätsverlag Dr. Brockmeyer, Bochum, pp 203–226

van Dijk J, Vollaard B (2012) Self-limiting crime waves. In: van Dijk J, Tseloni A, Farrell G (eds) The international drop in crime. New directions in research. Macmillan, Palgrave

Vanderveen G (2011) Fear of crime: its social construction in the Netherlands. Criminology, Athens 40–49

Zarafonitou C (2011) Fear of crime in contemporary Greece: research evidence. Criminology, Athens 50–63

Zedner L (2002) Victims. In: Maguire M, Morgan R, Reiner R (eds) The Oxford handbook of criminology. Oxford University Press, Oxford, pp 419–456

Zvekic U, Alvazzi del Frate A (1995) An overview of the main findings. In: Zvekic U, Alvazzi del Frate A (eds) Criminal victimization in the developing world. United Nations Interregional Crime and Justice Research Institute, Rome, pp 15–48

Fear of Crime and the Psychology of Risk

Jonathan Jackson[1] and Ioanna Gouseti[2]
[1]Department of Methodology and Mannheim Centre for Criminology, London School of Economics, London, UK
[2]Department of Methodology, London School of Economics, London, UK

Synonyms

Anxieties about crime; Psychology of risk; Risk perception; Worry about crime

Overview

Research on the fear of crime has developed slowly but surely since the early 1960s, providing a rich interdisciplinary literature of empirical, conceptual, and methodological interest. But studies often lack theoretical reach and ambition. In an attempt to stimulate interdisciplinary work that spans psychological and sociological modes of analysis, this entry tries to reframe the fear of crime literature according to psychological insights into social and risk perception.

Framing Fear of Crime and Risk Perception

The first section reviews psychological work on emotional reactions to the anticipation of victimization and risk perception.

Emotion

Emotional reactions and perceptions of risk lie at the heart of fear of crime, and one way to approach the link between them is to focus on "worry" not "fear." There are two reasons for this.

First, fear inadequately captures most people's emotions about the threat of becoming victims of crime (*see* inter alia Warr 2000; Gray et al. 2011). Fear is a strong physical response to an immediate and proximate threat, and while fear may be a good descriptor of some people's emotions in some situations (e.g., in the presence of immediate signs of danger), people's emotions about crime are often closer to some kind of rumination or anxiety about risk and uncertain threat (i.e., in the absence of immediate signs of danger). People can respond to an immediately threatening environment, but they can also worry about distant potential harm.

Worry about crime, by contrast, captures people's assessment of proximate *and* distal threat. Worry is both an evaluation of an immediate situation (one can worry about the threat of victimization indicated by close cues in the environment) and the rumination about future outcomes (an anxiety-producing concern about distal and future events yet to transpire). According to Berenbaum (2010, p. 963), worry can be described as repetitive and anxiety-producing thoughts that have three characteristics: "…(1) the repetitive thoughts concern an uncertain future outcome; (2) the uncertain outcome about which the person is thinking is considered undesirable; and (3) the subjective experience of having such thoughts is unpleasant."

Second, there is a good theoretical framework for the processes by which worry about crime starts, continues, and finishes. Berenbaum's (2010) *initiation-termination two-phase* model draws our attention not just to the links between perceived threat and emotion, but also to worry as a dynamic process that unfolds over time. On the one hand, worry is initiated by perceptions of threat which comprise the perceived probability of undesirable future outcomes, their cost, and the salience of risk and threat.

On the other hand, people continue to worry unless they can accept the uncertain future possibility of a threat, and have taken whatever efforts they can to prevent or cope with it. Conversely, they stop worrying if they become comfortable with the possibility that the threat might still be realized. In Berenbaum's (2010) model, this acceptance of threat – that often leads to the termination of worry – is linked to one's desire for certainty (i.e., the intolerance of uncertain situations and events), beliefs about the value of worrying (i.e., positive beliefs that some individuals hold about worry, such as perceiving it as an effective way of problem solving or an indicator of high levels of responsibility), a perseverative-iterative style (i.e., the tendency to focus on an object of concern by repeatedly thinking about the next possible step in a chain of connected outcomes), and a sense of closure in one's role to prevent or cope with the threat (i.e. the sense that every possible preventative or coping action has been taken).

Need for cognitive closure may be especially relevant to fear of crime (Jackson 2013). Kruglanski and Webster (1996, p. 278) define it thus:

> . . .a desire for definite knowledge on some issue and the eschewal of confusion and ambiguity . . . need for closure is presumed to exert its effects via two general tendencies: the urgency tendency, reflecting the inclination to attain closure as quickly as possible, and the permanence tendency, reflecting the tendency to maintain it for as long as possible

People with a high need for cognitive closure have difficulty accepting the prospect of the uncertain outcome; they exhibit a general intolerance of uncertainty and prefer predictability, structure, and quick decision-making. Imagine a group of people who all believe that falling victim of crime is a significant personal threat and all have a high need for cognitive closure. They are less able to live comfortably with non-negligible risk and cannot reach some sort of acceptance of the existence of possible harm, so they continue to worry in a persistent fashion.

Cognitive and Affective Assessments of Risk

The distinction between *risk as feeling* and *risk as analysis* (Loewenstein et al. 2001) may also be important to fear of crime. According to this perspective, there are two types of information processing that drive how people make sense of risk. On the one hand, *risk as analysis* involves a more formal, logical, and numeric style of reasoning. A style more applicable to quantitative and cognitive assessments of crime risk, *risk as analysis* typically involves a rather dispassionate assessment of likelihood and consequence via more logical and numerical assessments of risk. If people are using such a type of information processing regarding crime risk, then emotion is less likely to accompany and shape the resulting representation of potential threat.

On the other hand, *risk as feeling* involves more intuitive and experiential type of thinking in which emotion is particularly important. *Risk as feeling* brings in more qualitative aspects, like the vividness with which consequences can be imagined, one's mood, and prior experience with the event. For Slovic et al. (2004), the "experiential" system is affect-laden rather than formally logical like the "analytic system," involving rapid processing and the encoding of reality in images and metaphors rather than abstract symbols and numbers.

Importantly, affective reactions and appraisals of risk can often hold sway over more dispassionate perceptions (Loewenstein et al. 2001). For example, when the outcome is vivid and affect-laden, then an individual will be relatively insensitive to probability variations. Applied to fear of crime, people who process crime risk in affective and imagery-laden ways may be rather uninterested in the actual probability of victimization. They may focus instead on its imagined outcome.

Feelings about a risk object can also shape more cognitive and numeric appraisals of risk and benefit. The *affect heuristic* describes how a representation that is tagged with affect – a good or bad quality – can be influential in overall judgements of risk (Slovic et al. 2004). Vivid and emotion-laden representations of an outcome (i.e., a particular image of crime) may thus shape numerical assessments of probability (cf. Jackson 2011).

Availability Cascades

The *availability heuristic* (Tversky and Kahneman 1974) predicts that the probability of an event tends to be judged by the ease with which instances of it can be retrieved from memory. People tend to overestimate the frequency of very rare, spectacular events and underestimate the incidence of more frequent, less spectacular events. Applied to crime, the *availability heuristic* predicts that when individuals hold a particularly resonant and vivid image of a risk object (of a given crime, say), they judge its likelihood to be especially high: they substitute a relatively difficult question (e.g., how likely is it that I will become a victim of crime?) with a relatively easy question (e.g., how easy can I imagine becoming a victim of crime?).

Why can some people easily imagine becoming a victim of crime? Potentially relevant here is Sunstein's (2005) notion of "availability cascades," which posits that hearing about events via interpersonal communication and the mass media can have an impact on public perceptions of risk, partly through an interaction between availability and particular social mechanisms:

> [there may exist] social cascades, or simply cascades, through which expressed perceptions trigger chains of individual responses that make these perceptions appear increasingly plausible through their rising availability in public discourse. Availability cascades may be accompanied by countermechanisms that keep perceptions consistent with the relevant facts. Under certain circumstances, however, they generate persistent social *availability errors*—widespread mistaken beliefs grounded in interactions between the availability heuristic and the social mechanisms we describe. [emphasis in original] (Kuran and Sunstein 1999, p. 685)

Applied to fear of crime, circulating information about crime may shape people's emotional and cognitive representations of personal risk, in part by increasing its cognitive availability. Fear-inducing accounts of events – such as the Boston Marathon bombing in May 2013 and the incident in 2002 in Virginia when two snipers killed ten people – may be highly publicized, noticed, and repeated, leading to cascade effects as the event becomes available to an increasing number of people.

Sunstein (2005, p. 93) also speculates that existing predispositions can determine in large part what individuals pay attention to. One example he gives relates to genetic modification of food. Those who are predisposed to be fearful of this issue are more likely to seek out information about genetic modification. Furthermore, "group polarization" describes how, when individuals discuss with each other certain events and risks, they typically end up with a more extreme view (p. 98).

The subjective threat of crime may thus become enhanced in people's minds. Vivid and dramatic images may raise subjective probabilities, in turn initiating worry. Frightening images may then lead some people to process risk information in more rapid and basic ways that are sensitive to the vividness and associations that a given risk evokes. The emotional impact of the associated outcomes of victimization may shape cognitive assessments of probabilities, and emotional appraisal of risk may hold sway over cognitive assessments – when emotions and thoughts diverge – in shaping people's behavior. If individuals find it difficult to accept the threat, then worry will continue in a persistent fashion over time (cf. Berenbaum 2010).

Criminological Work on the Fear of Crime

This section reviews how criminological research has defined and measured emotion and risk perception, before then addressing how criminological studies have explored the links between the two.

Risk Perception, Vulnerability and the Fear of Crime

Criminological studies typically ask individuals about the intensity with which they worry about falling victim of crime (Hough 1995) or the extent to which they are fearful of falling victim of crime (Ferraro 1995) – although some continue to use single indicators of subjective safety in the home and streets. Asking individuals how "afraid" or "worried" they are about crime, surveys direct people to their emotions about future harm. But worry (or fear) may be rarer than one might think. In a study based on British Crime Survey (known as Crime Survey for England and Wales since April 2012) data, more people expressed concern about future harm (i.e., they said they were worried at the time of the interview) than had worried recently (Farrall et al. 2009). A good proportion of individuals who said that they were "very" or "fairly" worried also reported that they had not worried even once over the past twelve months. Asking about *frequency* may thus be a useful supplement to asking about the *intensity* of worry or fear (Gray et al. 2011).

Criminological research typically measures perceived risk as solely subjective probability. Survey respondents are asked how likely they think it is that they will become a victim of crime (e.g., in the next year). But a number of studies have widened the focus beyond perceived likelihood. Drilling into perceptions of the possible impact and seriousness of crime and victimization (Warr 1987; Jackson 2011) as well as the sense of control over the event (Jackson 2009); such work shows the relevance of people's perceptions of the seriousness of the consequences *if they were to fall victim* and people's sense of control *over its occurrence*.

This work suggests that worry about crime is partly generated and sustained by a sense of risk, threat and vulnerability, comprised of perceived likelihood, consequence and control (*see* Killias 1990; Gabriel and Greve 2003; Farrall et al. 2009; Jackson 2009, 2011). Consistent with Berenbaum's (2010) account of the initiation of worry, people start to worry about crime when they see significant threat. And critically, subjective threat may extend beyond the chances of an event happening, to include its controllability and possible impact. The heterogeneity of different types of crime with regard to "relevance, explanation and consequences" (Gabriel and Greve 2003, p. 6) may mean that the same crime could have a different anticipated resonance or impact for one individual compared to another.

Introducing the notion of "perceptually contemporaneous offences," Warr (1984) posited that some criminal events are viewed as more dangerous than others by particular segments of the population (typically women and the elderly) due to the crimes they imply. As such, "... circumstances or events that appear innocuous or comparatively minor to males or younger persons are apt to be viewed as more dangerous to females and the elderly because of the offences they imply or portend" (Warr 1994, p. 19). Similarly, Ferraro (1995, p. 87) argued that sexual harassment "... may shadow other types of victimization among women. Rape may operate like any other master offence among women, especially younger women who have the highest rate of rape, heightening fear reactions for other forms of crime."

Relatedly, Warr's (1987) model of "sensitivity to risk" predicts that the influence of likelihood on fear is moderated by perceptions of seriousness of the given type of crime. Analyzing US data, he found that when people judged crime to be especially serious in its effect, a lower level of perceived likelihood was needed to stimulate some level of personal fear. Individuals were more "sensitive" to a given level of perceived risk when they viewed the consequences of victimization to be especially serious. The risk sensitivity model suggests that once individuals associate a risk with high personal consequences and low personal control, only a relatively small level of perceived likelihood is needed to elicit a strong emotional response (Jackson 2011).

There is a clear overlap between the notions of perceived risk and vulnerability (or susceptibility) to crime. According to Killias (1990), there are physical, social and situational dimensions of vulnerability, related to three aspects of threat: "exposure to non-negligible risk," "loss of control," and "seriousness of consequences." Crucially, perceived risk may mediate the association between markers of vulnerability (typically sociodemographic variables, such as gender and age) and expressed emotion. For example, Jackson (2009) found that females tend to worry more frequently than males partly because (a) they feel less able to physically defend themselves, (b) they have lower perceived self-efficacy, (c) they have higher perceived negative impact, and (d) they see the likelihood of victimization as higher for themselves and for their social group. Those with fewer resources – typically posited to be females, the elderly, and those in lower socioeconomic groups – may feel less able to protect themselves and deal with the consequences, and therefore they may be more likely to report being worried or fearful over falling victim of crime.

Applying the work on *risk as feeling* and *risk as analysis* to risk perception and fear of crime, one individual may associate burglary with the risk of physical or sexual assault, while another individual may associate burglary with the loss of material goods and a great deal of inconvenience. For the first person, the risk of burglary may be appraised through the affective route (in part because of the severe consequences and sense of uncontrollability of the perceptually contemporaneous offence of assault) leading to a feeling of vulnerability and susceptibility to crime. Risk in this case may be constituted by the resonance of the consequences, the uncontrollability of the outcome, the vividness of the event, and the ease with which one can summon up a frightening image. Such imagery of consequence and control may then raise the subjective probability of burglary. But it may also interact with the perceived likelihood to produce a strong emotional reaction (cf. Jackson 2011). By contrast, the second person may think about burglary more through the cognitive route, with the subjective likelihood being relatively low, and less influential in shaping emotion than some resonant image of the impact of the event and its controllability.

Risk Perception and Circulating Media Representations of Crime and Danger

One might assume that personal victimization experience is the main source of people's personal sense of vulnerability and risk. People may feel threatened about crime in part because they have recently been victimized. It may be, for example, that people who have been burgled have a strong sense of the severe consequences, can easily imagine it happening, and know something about its controllability.

Yet victimization experience is not a strong predictor of fear of crime (*see* inter alia Hale 1996). This may be because of what Winkel (1998, p. 477) calls "active adaptational coping," whereby the experience of victimization simultaneously raises one's perception of the likelihood of it occurring and lowers their perception of its negative impact if it were to occur again (e.g., the reality of burglary may not be as frightening as one imagines). More work is clearly needed, however, and Berenbaum's (2010) model of worry may be a helpful orientating framework for research into the effect of victimization on fear of crime, capturing the process of its initiation and continuation or termination over time.

If it is not (primarily) personal experience of criminal victimization, what drives people's perceptions of risk? Why does crime become a vivid, accessible, and easy-to-imagine prospect for some people, involving a sense of severe personal consequences, uncontrollability and unpredictability?

Clearly, the media are a primary source of information about the extent, nature, and seriousness of crime in most countries across the world. The media tend to report shocking but rare criminal events that are intriguing, dramatic, and sensational (*see* inter alia Weitzer and Kubrin 2004; Surette 2011). In a study of the reporting of murder in the UK, for example, Gekoski et al. (2012, p. 17) found that journalists tend to "…believe that certain types of murder – involving 'ideal' victims, unusual features, extreme brutality and serial killers – are almost always of interest to the public, while certain other types – involving 'undeserving' victims in commonplace circumstances – are almost always not."

Crucially, the mass media and interpersonal communication may be an important backdrop for a widespread psychological proximity of the threat of crime. While the fear of crime literature lacks a clear and robust evidence based on the relationship between media coverage and risk perception (although *see* Chiricos et al. 1997), it does seem plausible that people develop personalized images of risk partly as a result of hearing about specific events of crime via the media. High levels of coverage – biased towards the dramatic and sensational – may make the events cognitively available to people. Personal images of risk may develop, and through the media's focus on the highly consequential and uncontrollable, crime reports may end up stressing certain attributes of the criminal event which only increase the vividness of the representations. If many people develop their own sense of risk, based on common imagery of particularly frightening crimes, then it may not be very surprising that they worry about crime: they receive dramatic images of risk that become cognitively available and are processed in a more intuitive and experiential manner, which stresses affect, images, and vividness.

People might, for instance, come to associate burglary with violence in part because they receive and process striking reports of especially serious burglaries from the newspapers, television and other media. Hearing about specific (and local) events of burglary – which are brought home in a very vivid way, "as if it could happen to me," that resonate with their own experience (*see* inter alia Chiricos et al. 1997) – may then explain why people develop these personalized representations of the risk of crime, with all the attendant sense of severe consequences, loss of control, and high likelihood (Jackson 2011).

Risk Perception and the Social and Physical Environment

What about the neighborhood in which someone lives and resides? Does the social and physical environment matter?

Here the fear of crime literature is relatively strong. First, a number of studies have used multilevel models to situate individuals in their specific neighborhood context (e.g., Wyant 2008; Brunton-Smith and Sturgis 2011). Focusing on people's expressed feelings of local safety, Wyant (2008) found significant neighborhood-level clustering in Philadelphia. Similarly, Brunton-Smith and Sturgis (2011) found significant area-level clustering in worry about crime in neighborhoods across England and Wales.

It clearly matters where one lives. But what explains this neighborhood-level clustering? Perhaps the most complete study in this regard is that of Brunton-Smith and Sturgis (2011). Examining neighborhood effects across a full national distribution of neighbourhoods in and around a range of different urban, rural, and metropolitan contexts, they found that recorded crime, neighborhood structural characteristics, and visual signs of disorder (derived from survey interviewer ratings rather than respondent assessments) all had direct and independent statistical effects on the expressed intensity of worry about crime. Additionally, neighbourhoods predicted crime-related worry in more subtle ways, moderating the statistical effects of its individual-level causes: between-group differences in expressed worry about crime were both exacerbated and ameliorated by the characteristics of the areas in which people live.

Second, one's neighborhood matters but so also does one's adjacent set of neighborhoods. People seem to draw on environmental cues not just from their own immediate environment, but also from neighboring local areas, basing their judgements of risk on both the characteristics of their local neighborhoods and the characteristics of the areas visited during their routine activities. In statistical terms, this manifests itself as spatial autocorrelation, with higher than average similarities amongst neighbourhoods in closer proximity to one another, signalling the existence of shared characteristics and joint influences.

Wyant (2008) incorporated this additional spatial structure in an assessment of neighborhood effects on fear of crime. Across 45 local neighborhoods in Philadelphia, Wyant reports significant spatial autocorrelation in his measure of neighborhood fear. Analyzing data from the Crime Survey of England and Wales, Brunton-Smith and Jackson (2012) found that worry about crime was weakly related not just to levels of crime and disorder in one's own neighborhood but also to crime and disorder in adjacent areas. By incorporating information about those areas that surround each neighborhood, it was possible to estimate the extent of these shared influences as well as the extent to which individuals base their judgements on the broader local environment.

Third, people's *perceptions* of their neighborhood are key. The aforementioned work used recorded crime levels, neighborhood structural characteristics and visual signs of disorder to help explain some of the (immediate and adjacent) neighborhood effects. But many more studies have addressed people's perceptions of the immediate environment, investigating how individuals variously make sense of neighborhood disorder and collective efficacy. Importantly, disorder is any aspect of the social and physical environment that indicates to the observer

(a) a lack of control and (b) the values and intentions of others that share the space (*see* inter alia Skogan and Maxfield 1981; Ferraro 1995). Ferraro (1995) analyzed data from a nationally representative survey of US citizens, finding that people's perceptions of incivilities predicted their perceptions of the chances of victimization, which in turn predicted people's expressed fear of crime.

Farrall et al. (2009) found similar results, this time analyzing data from the British Crime Survey. Although they did not use a multilevel framework, their analysis did link neighborhood-level characteristics to worries and anxieties about crime through two layers of mediation: individual perceptions (of disorder and collective efficacy) and risk perception (i.e., subjective probabilities). The statistical effects of social and structural characteristics of the neighborhood seemed to be mediated through people's reading of social organization and disorganization.

The Social Perception of Disorder

People thus seem to "read" the character and intentions of certain individuals and groups – and the presence or lack of local social control – from the level of social disorganization in their neighborhood. But importantly, two individuals living in the same environment can come to quite different conclusions about the same environmental cues. A good deal of work suggests that people's perceptions of their neighborhood are dependent not just on the social and physical cues in the local environment but also on the respondent's *relationship* to that environment and others who inhabit it (e.g., Sampson and Raudenbush 2004). Some individuals judge certain stimuli as "disorderly," while other individuals in the same environment might not, with people ascribing meaning to disorder they see in front of them (see Wickes et al. 2013).

An individual's existing attitudes, beliefs, or prejudices seem to provide a filter through which they experience and interpret their environment. Ambiguous cues need to be defined by the observer as "disorderly" *in the first place*. Exploring the basis on which individuals formed perceptions of disorder, Sampson and Raudenbush (2004) reasoned that citizens interpreted objective signs of disorder (measurable signs of litter, vandalism, graffiti, etc.) through existing and historical stereotypes of race and deprivation (see also Chiricos et al. 1997). Their work suggested that observers in Chicago not only infused disorderly cues with notions of race and deprivation; they also had existing cognitive representations or stereotypes that linked disadvantaged minority groups with "social images, including but not limited to crime, violence, disorder, welfare, and undesirability as neighbors."

Jackson (2004) also found that some people judged particular ambiguous stimuli as "disorderly" and representational of criminal threat, while other people in the same environment judged the same stimuli as benign and unthreatening. Respondents who held more authoritarian views about law and order, and who were concerned about a long-term deterioration of community, were more likely to perceive disorder in their environment. They were also more likely to link these physical cues to problems of social cohesion and consensus, declining quality of social bonds and informal social control (cf. Farrall et al. 2009). The symbolic nature of social order seemed to generate meaning in the context of their relationship to long-term social change and people's anxieties about cohesion and moral consensus.

In an important qualitative study, Carvalho and Lewis (2003, p. 791) explored how reactions such as fear, safety, and anger were dependent on personal distancing from issues of crime and disorder:

> The more distant the relationship, the less salient these problems, as dangers, are in the context of daily routines and the less they intrude in one's life. With close relationships, crime and incivilities occupy a central position in one's daily life. These problems appear isolated, detached from other domains of life, imposing themselves onto the person as dangerous (raising fear) or as bothersome (raising anger). With distant relationships, on the other hand, crime/incivilities are peripheral to the person. The dangers of these problems are part of life, and qualities other than their dangerousness are apparent. Peripheral relationships with crime/incivilities accompany a neutral reaction of safety (neither fear nor revolt).

Carvahlo and Lewis's (2003, p. 791) data suggested that individuals can neutralize dangers when they view crime/incivilities as "banal events . . . Dangers are contextualized temporally, referred to the past, with the idea that they have been there, they have existed, and they are ordinary." Dangers might also be neutralized when they are "delimited": "when they lose their random character (thus, the potential to affect just anyone) and become restricted to certain places of the neighbourhood and times of the day (physical delimitation), or to groups of people (social delimitation)" (p. 794).

Final Thoughts and Future Directions of Research

More work is needed on the links between emotion and risk perception in fear of crime. Studies need to address how people develop a sense of threat and psychologically try to manage that threat. The distinction between *risk as feeling* and *risk as analysis* may be especially important, exploring whether risk, circulated chiefly by the mass media and interpersonal communication, may be judged cognitively *and* emotionally. People may process information from the mass media (and interpersonal communication) using both analytical and associational routes. When the emotional appraisal of the threat of crime (using a more associationist rather than formal mode of information processing) differs to a cognitive sense of its likelihood, feelings may hold sway, and if crime is judged to have severe consequences and the outcome is vivid and affect-laden, then an individual may be insensitive to probability variations. That individual is thus unlikely to feel better if she or he is told that their chances of victimization are rather slight.

A feedback system might also mean that someone already emotionally animated by risk builds over time a more extensive and vivid image of the risk event, fleshing out effects, protagonists, and relevant causes and circumstances. This might then make the risk more substantial, structured, and relevant to that individual. Emotional systems may lead to the structuring and differentiating of risk images, personalizing them, fleshing them out, and bringing affect into the picture (Jackson 2011, 2012).

In addition, more work is needed to explore neighborhood context, social perception, and circulating images of crime in the context of the links between emotion and risk perception. Studies need to address the particular set of evaluative activities that links crime – in the eyes of observers – with certain individuals or groups that are judged to be (a) hostile to the local social order, (b) untrustworthy, and (c) representative of some sort of social breakdown. Such work should assess how the risk of crime is projected into a given environment and elaborated with a face (the potential criminal) and a context (the place it might happen). It should consider how emotion and risk perceptions disclose a whole host of subtle evaluations of and responses to the social world – a way of responding to variable levels of social order and control, a sense of unease in an unpredictable environment, and the association of particular individuals or conditions with deviance and hostile intent. Fear of crime may thus be both an expressive and experiential phenomenon (Jackson 2004).

In this respect, perceptions of crime risk may reveal processes of designation (where certain groups, behaviors or community conditions are labelled as potentially criminal and dangerous). Crime may be used as a way of articulating evaluations of people, community conditions, and social control, a lens through which people understand social order, low-level deviance, and diversity. The identification of dangerous individuals may operate to establish "moral communities" by locating "immoral communities." Stereotypes of particular groups may operate as distancing strategies for placing others, perpetuating normative boundaries of social conduct, roles and judgements, and strengthening one's own social identity. They may reinforce the boundaries and identity of a given community by identifying what that community is against, e.g., certain troublesome individuals or particular groups defined by their social class or their ethnicity. Social psychology shows that the identification of an out-group can act to strengthen in-group solidarity. Scapegoating may also arise, where one group comes to embody a particular social problem. Such evaluative activity, if this analysis is correct, reveals how people define social order and what they think is hostile to social order. It also means that risk is culturally conditioned: what one defines as dangerous depends on where one stands.

Criminological research shows that the fear of crime is a complex social phenomenon and a multi-faceted construct. Fear of crime clearly expands far beyond direct victimization experience and probabilistic evaluations of its occurrence, comprising multiple psychological and social phenomena – such as wider worrying mechanisms, perceptions of one's physical and social environment, and individuals' prejudices and stereotypes. But more work is needed to capture the complexity and multi-dimensionality of this important social and political phenomenon. Bridging sociological and psychological modes of analysis can open up new avenues of research, providing us with more powerful analytical perspectives to understand the causes, nature and effects of fear of crime.

Related Entries

- ▶ Anomie and Crime
- ▶ BCS
- ▶ Crime and the Racial Composition of Communities
- ▶ Defining Disorder
- ▶ Differential Social Organization
- ▶ Disadvantage, Disorganization and Crime
- ▶ Disorder: Observational and Perceptual Measures
- ▶ Fear and Punishment
- ▶ Informal Social Control
- ▶ International Crime Victimization Survey
- ▶ National Victimization Surveys
- ▶ New Media and Crime Images
- ▶ Public Representations of the Police
- ▶ Risk assessment, Classification, and Prediction
- ▶ Social Capital and Collective Efficacy
- ▶ Social Disorder and Physical Disorder at Places
- ▶ Understanding Victimization Frequency

Recommended Reading and References

Berenbaum H (2010) An Initiation-termination two phase model of worrying. Clin Psychol Rev 30: 962–975

Brunton-Smith I, Jackson J (2012) Urban fear and its roots in place. In Ceccato V (ed) Urban fabric of crime and fear. Springer, London, pp 55–82

Brunton-Smith I, Sturgis P (2011) Do neighborhoods generate fear of crime? An empirical test using the British crime survey. Criminology 49(2):331–369

Carvahlo I, Lewis DA (2003) Beyond community: reactions to crime and disorder among inner-city residents. Criminology 41(3):779–812

Chiricos T, Eschholz S, Gertz M (1997) Crime, news and fear of crime: towards an identification of audience effects. Soc Probl 44(3):342–357

Farrall S, Jackson J, Gray E (2009) Social order and the fear of crime in contemporary times. Oxford University Press, Clarendon Studies in Criminology, Oxford

Ferraro KF (1995) Fear of crime. Interpreting victimization risk. State University of New York Press, Albany

Gabriel U, Greve W (2003) The psychology of fear of crime. Conceptual and methodological perspectives. Br J Criminol 43:600–614

Gekoski A, Gray JM, Adler JR (2012) What makes a homicide newsworthy? UK national tabloid newspaper journalists tell all. Br J Criminol, First published online August 27 2012 10.1093/b2jc/azs047

Gray E, Jackson J, Farrall S (2011) Feelings and functions in the fear of crime: applying a new approach to victimisation insecurity. Br J Criminol 51(1):75–94

Hale C (1996) Fear of crime: a review of the literature. Int Rev Victimology 4(2):79–150

Hough M (1995) Anxiety about crime: findings from the 1994 British crime survey, Home Office research study no. 147. Home Office, London

Jackson J (2004) Experience and expression. Social and cultural significance in the fear of crime. Br J Criminol 44:946–966

Jackson J (2009) A psychological perspective on vulnerability in the fear of crime. Psychol Crime Law 15(4):365–390

Jackson J (2011) Revisiting risk sensitivity in the fear of crime. J Res Crime Delinq 48(4):513–537

Jackson J (2013) Cognitive closure and risk sensitivity in the fear of crime. Available at SSRN http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2112602

Killias M (1990) Vulnerability: towards a better understanding of a key variable in the Genesis of fear of crime. Violence Vict 5(2):97–108

Kruglanski AW, Webster DM (1996) Motivated closing of the mind: 'seizing' and 'freezing'. Psychol Rev 103(2):263–283

Kuran T, Sunstein CR (1999) Availability cascades and risk regulation. Stanf Law Rev 51(4):683–768

Loewenstein GF, Weber EU, Hsee CK, Welch ES (2001) Risk as feelings. Psychol Bull 127(2):267–286

Sampson RJ, Raudenbush SW (2004) Seeing disorder: neighborhood stigma and the social construction of 'broken windows'. Soc Psychol Q 67:319–342

Skogan W, Maxfield M (1981) Coping with crime. Sage, Beverly Hills

Slovic P, Finucane ML, Peters E, MacGregor DG (2004) Risk as analysis and risk as feelings: some thoughts about affect, reason, risk, and rationality. Risk Anal 24:311–322

Sunstein CR (2005) Laws of fear: beyond the precautionary principle. University Press, Cambridge

Surette R (2011) Media, crime, and criminal justice, 4th edn. Wadsworth, Belmont

Tversky A, Kahneman D (1973) On the psychology of prediction. Psychological Review 80(4): 237–251

Warr M (1984) Fear of victimization: why are women and the elderly more afraid. Soc Sci Q 65(3):681–702

Warr M (1987) Fear of victimization and sensitivity to risk. J Quant Criminol 3(1):29–46

Warr M (1994) Public perceptions and reactions to violent offending and victimization. In: Reiss AJ Jr, Roth JA (eds) Understanding and preventing violence offending and victimization. Volume 4: Consequences and control. National Academy Press, Washington, DC

Warr M (2000) Fear of crime in the United States: avenues for research and policy. Crim Justice 4:451–489

Weitzer R, Kubrin CE (2004) Breaking news: how local TV news and real-world conditions affect fear of crime. Justice Q 21:497–520

Wickes R, Hipp JR, Zahnow R, Mazerolle, L (2013) 'Seeing' minorities and perceptions of disorder: explicating the mediating and moderating mechanisms of social cohesion. Criminology. doi: 10.1111/1745-9125.12011

Winkel FW (1998) Fear of crime and criminal victimization: testing a theory of psychological incapacitation of the 'stressor' based on downward comparison processes. Br J Criminol 38(3):473–484

Wyant BR (2008) Multilevel impacts of perceived incivilities and perceptions of crime risk on fear of crime: isolating endogenous impacts. J Res Crime Delinq 45(1):39–64

Feminist Criminological Theory

Amanda Burgess-Proctor
Criminal Justice, Oakland University, Rochester, MI, USA

Overview

Informed by the observation that feminist criminology is *both* readily embraced by the discipline *and* still isolated from "core" theoretical knowledge, this entry examines two areas in which cutting-edge feminist scholarship is occurring: advancements in feminist criminological theory and feminist refinements of traditional criminological theory. The first half of this entry considers innovative applications of the intersectional theoretical framework to several areas of feminist criminological scholarship, with a particular focus on studies examining violent victimization among girls and women of color. The second half of this entry highlights efforts to infuse gender into mainstays of criminological theory from which gender has long been absent, including life course, social control, and social learning theories. Taken together, the exemplary studies highlighted in this entry demonstrate the important scholarly developments occurring within contemporary feminist criminological theory.

> If a prominent goal in much of the criminological enterprise is the creation of general theories. . ., feminist criminology has not been particularly successful. However, it is precisely feminist understandings of gender that require us to move beyond what broad, global explanations can provide. (Miller and Mullins 2006, p. 221)

Introduction

In her 1998 commentary on gender, crime, and criminology, Daly identified two distinct and divergent areas of theoretical development that served as sources of "field expansion" in criminology: criminological scholarship largely uniformed by feminist knowledge and feminist scholarship largely uninformed by criminological knowledge (p. 86). As Daly noted at that time, "Some feminist scholars in criminology straddle both sources of knowledge, and others may work from the second source to challenge criminological thought and to build from feminist theories" (p. 86). In the 15 years since that writing the gulf between criminology-proper and feminism-proper has narrowed, allowing contemporary feminist criminologists less time for straddling and more time for advancing knowledge. Indeed, recent theoretical innovations continue to push feminist criminology in new directions and into new applications, and this work is published in a previously unthinkable number of outlets (including the journal *Feminist Criminology*) to a seemingly ever-expanding and ever-receptive audience.

Though narrowed, the troublesome divide between criminological and feminist knowledge persists,and this disjuncture is driven more by perceived ideological differences than actual theoretical or empirical differences. Feminist criminological theories still routinely are positioned in opposition to "traditional" or "mainstream" theories. This is problematic because the characterization of feminist perspectives as "nontraditional" or "nonmainstream" risks giving the impression that they are nonnormative or lack the empirical rigor of those dominant theories, as some critics have (wrongfully) claimed. Still, the distinction between "traditional/mainstream" and "feminist" theories in criminology is perhaps best understood, to use an illustrative shorthand, as the difference between theories occupying the bulk of undergraduate theory textbooks and those relegated to a chapter or section at the end.

This delicate reality – that feminist criminological scholarship is simultaneously far more enthusiastically embraced than ever before, while still regularly positioned outside fundamental, core sources of theoretical knowledge – puts contemporary feminist criminology in a somewhat precarious position. Using the feminist "both/and" concept (Collins 2000; Daly and Stephens 1995), one may consider feminist criminology to be *both* a legitimate source of theoretical knowledge *and* auxiliary to core criminological theories – and, therefore, denied a position of theoretical privilege in the discipline.

Of course, feminist perspectives were developed to correct the absence of gender from the long line of traditional theories representing the pillars of criminology. Feminist criminologists have criticized the long-standing exclusion of girls and women – indeed, of *gender* – from nearly every theory developed to explain offending (Belknap 2007; Daly and Chesney-Lind 1988; Miller and Mullins 2006; Simpson 1989). Even today, "key contemporary schools of thought in criminology have yet to adequately incorporate gender" (Miller 2010, p. 136). Feminist criminologists' response to this oversight generally has taken two forms: (1) efforts to advance feminist, gender-specific criminological theories, and (2) efforts to refine "traditional" theories of offending using feminist insights (e.g., see Daly and Chesney-Lind 1988). These strategies raise an obvious question: "[g]iven that traditional criminological theories are posited to be gender-neutral (and universal) yet have been developed without consideration of how gender structures human behavior, what are the implications of theories that ignore gender in the face of real-life gender inequality and discrimination?" (Burgess-Proctor 2010, p. 433).

Informed by the observation that feminist criminology is *both* readily embraced by the discipline *and* still isolated from "core" theoretical knowledge and guided by the two approaches feminist scholars have taken to address the absence of gender from mainstream criminological theories, this entry examines two areas in which cutting-edge feminist scholarship is occurring: advancements in feminist criminological theory and feminist refinements of traditional criminological theory. Importantly, these two areas reflect Daly's (1998) conceptualization of efforts to advance feminist criminological scholarship from either its criminological or its feminist source.

Cutting-Edge Advancements in Feminist Criminological Theory

In their recent commentary offering an evaluation of feminist theory in criminology, Miller and Mullins (2006) draw upon Daly's (1998) conceptual schema to identify three core areas of feminist criminological inquiry: (1) *gendered pathways to lawbreaking*,

(2) *gendered crime*, and (3) *gendered lives*. First, *gendered pathways* research explores how men and women differentially initiate, continue, and terminate offending trajectories. Second, *gendered crime* research examines the types, contexts, and consequences of men's and women's offending behaviors, as well as the extent to which criminal justice processing of those offenses also is gendered. Third, *gendered lives* research examines how gender structures the broader lived experiences of girls and boys, which may facilitate (or protect against) victimization and offending opportunities. These three areas of inquiry are useful for evaluating cutting-edge work in contemporary feminist criminological theory.

Perhaps the most important advancement in feminist criminological theory has been the development of intersectional scholarship that simultaneously attends to the interlocking and interacting forces of race/ethnicity, class, gender, sexuality, age, and other systems of power (see Burgess-Proctor 2006 for a review). Both intersectional scholarship and feminist criminology share common goals: (1) to develop theoretical frameworks that account for structuring forces like gender, race/ethnicity, and class; (2) to utilize methodologies that give voice to groups whose experiences previously have been overlooked; and (3) to bridge theory and practice with an eye toward improving social justice for marginalized groups (Burgess-Proctor 2006). Indeed, intersectionality is particularly relevant to feminist criminologists because it considers how intersecting systems of race, class, and gender make marginalized groups more vulnerable to criminal legal system control (Sokoloff and Burgess-Proctor 2011).

Intersectional scholarship perhaps most closely aligns with the *gendered lives* area of inquiry, which Miller and Mullins (2006) note has been least empirically examined but may be most ripe for exploration. Contemporary feminist criminologists are performing intersectional analyses of a range of important criminological issues, including gender violence in the lives of urban girls of color, marginalized women's experiences with intimate partner abuse (IPA), and other areas beyond women and girls' violent victimization.

Girls of Color and Urban Gender Violence

Two recent works exploring how inner-city Black girls' victimization and offending experiences are simultaneously raced, classed, and gendered are good examples of cutting-edge, intersectional feminist criminological theorizing. First, Jody Miller's (2008) book *Getting Played* examines the lives of Black girls in impoverished St. Louis neighborhoods. Miller interviewed 75 youth (35 girls and 40 boys) ages 12–19 about their experiences with three types of gender-based aggression: sexual harassment, sexual coercion and assault, and dating violence. Summarizing her findings, Miller writes: "Urban African American young women face widespread gendered violence that is a systematic and overlapping feature of their neighborhoods, communities, and schools. In addition, while young women employ a variety of strategies to insulate themselves from such violence, they do so in a context in which ideologies about gender work against them at every turn" (p. 192). By virtue of their race and class status, and the structural inequalities that contribute to the impoverished communities in which they live, these girls "have limited support and few avenues – institutional or otherwise – for remedying the systematic nature of the gendered dangers present in their daily lives" (p. 192).

Second, targeted at the same population of girls in Philadelphia, Nikki Jones' (2010) book *Between Good and Ghetto* offers a similarly nuanced portrait of how race, class, and gender shape inner-city Black girls' lived experiences. The result of 3 years of ethnographic field work and research, Jones' analysis uncovers how deeply entrenched and intersecting systems of race, class, and gender shape the ways girls vacillate between conforming to ("good") and violating ("ghetto") expectations of Black femininity and respectability. Jones writes: "Inner-city girls who live in distressed urban neighborhoods face a gendered dilemma: they must learn how to effectively manage potential threats of interpersonal violence…at the risk of violating mainstream and local expectations regarding appropriate feminine behavior. This is a uniquely difficult dilemma for girls, since the gendered expectations surrounding girls' and women's use or control of violence are especially constraining" (p. 9).

Both studies uncover similar findings that significantly illuminate our understanding of how urban gender violence is simultaneously raced and classed, making them important contributions to feminist criminological theory. First, violence is endemic to the lived worlds of inner-city Black girls; they experience violent victimization and exposure to violence at home, in their neighborhoods, and in their schools. Second, the social mechanisms governing whether, why, and to what extent inner-city Black girls use and receive interpersonal violence are highly gendered, raced, and classed. Third – and this is perhaps the most distressing finding of both books – inner-city Black girls often are left to navigate, negotiate, and survive the violence in their lives by themselves. Finally, inner-city Black girls are forced to adopt survival strategies, including "situational avoidance" (Jones 2010) and isolating themselves in their homes (Miller 2008), that are damaging, disruptive, and likely to serve them poorly later in life. These maladaptive survival strategies may disadvantage girls who go on to experience intimate partner abuse (IPA) – a second source of cutting-edge intersectional feminist criminological theorizing.

Marginalized Women's IPA Victimization

An increasing body of intersectional scholarship in feminist criminology addresses marginalized women's experiences with IPA victimization. Feminist scholars who study IPA have advocated for intersectional theoretical models of both men's battering and women's responses to victimization that include, but do not prioritize, gender (e.g., see Sokoloff and Pratt 2005). For example, Potter's (2006) Black Feminist Criminology perspective aims to identify the structural, cultural, and familial forces that shape Black women's responses to IPA. Potter (2006) describes her perspective as moving "beyond traditional feminist criminology to view African American women (and, conceivably, other women of color) from their multiple marginalized and dominated positions in society, culture, communities, and families" (p. 107). This observation echoes findings from Miller's (2008) and Jones' (2010) research, which unsurprisingly documented high exposure to IPA among their samples. And, as Miller (2008) notes, even instances of IPA among adults outside youths' own families of origin (e.g., between neighbors, acquaintances, or even strangers) are still important modeling opportunities for children who witness them. Thus, the experiences of women – particularly marginalized women – with IPA victimization are most appropriately theorized using an intersectional framework, as recent feminist criminological scholarship demonstrates.

Beyond Violence: Other Applications of Intersectionality

Finally, contemporary feminist scholars are applying an intersectional race/class/gender framework to criminological topics that go beyond women and girls' violent victimization experiences. For example, Sokoloff and Burgess-Proctor (2011) use an intersectional lens to examine how contemporary US drug policy – specifically, the mass incarceration of drug offenders and restrictive ex-offender policies – disproportionately disadvantages poor women of color. Meanwhile, Erez and Berko (2010) draw upon an intersectional framework to explore the pathways to crime and imprisonment of Arab/Palestinian women living in Israel. Finally, Benson and Simpson (2009) use a race/class/gender perspective to consider a subject typically outside feminist purview: how opportunities for white-collar crime are differentially distributed across groups in society. As these examples reveal, contemporary feminist criminologists are using an intersectional framework to perform cutting-edge research and theorizing on a wide range of important criminological topics.

Efforts to Refine "Traditional" Criminological Theories Using Feminist Insights

Whereas the examples in the previous section are best thought of as existing within the feminist body of knowledge, cutting-edge feminist theorizing also is occurring in the criminological body of knowledge. Efforts to impart feminist insights into mainstream criminological scholarship are evident in recent feminist refinements of the life course theoretical model, as well as social learning and social control models.

Life Course Theoretical Model

In recent years, one framework in particular has been notably expanded and advanced by feminist scholars: the life course theoretical model. As its name suggests, the life course model examines the development of human behavior over time and in the context of historical and situational events and is characterized by "the notion that changing lives alter developmental trajectories" (Elder 1998, p. 1). Within criminology, the life course model is used to examine patterns of offending and desistance over time. The last two decades have seen a proliferation of life course criminological research. Perhaps the best known and most influential application of the life course model in criminology is Sampson and Laub's (2003, 1993) research on Boston-area delinquent boys using data originally collected by Harvard researchers Sheldon and Eleanor Gleuck (see Sampson and Laub 2005 for a review).

Broadly, feminist refinements of the life course theoretical model in criminology have taken two forms. The first type of feminist refinement is feminist pathways research. The Feminist pathways model is an extension of the life course criminological framework that examines women's and girls' offending behaviors in the context of their past victimization experiences. The feminist pathways and life course models share many similarities, including a focus on trajectories of behavior like offending and desistance and the role of transitions in altering or shaping those trajectories. While both models attempt to explain and predict offending over time, the feminist pathways model specifically considers how women's (and girls') prior victimization experiences initiate or facilitate offending pathways (Burgess-Proctor 2011; Gaarder and Belknap 2003; McDaniels-Wilson and Belknap 2009).

The second type of feminist refinement might be loosely termed "gendered pathways." In contrast to feminist pathways research that specifically considers the role of women's victimization experiences in their subsequent offending, gendered pathways analyses view life course processes through a gendered lens; in other words, these analyses seek to uncover how, why, and in what ways gender structures motivations and opportunities for both offending and desistance. (Of course, when gendered pathways studies examine only women's experiences without also including the experiences of men, the extent to which the processes they uncover are accurately termed "gendered" remains unclear.) While these analyses may be less likely to bear the "feminist" label than studies using the feminist pathways model, they nonetheless accomplish a primary objective of feminist criminology: exploring the meaning and nature of gender relations (e.g., see Miller and Mullins 2006).

Cutting-edge feminist pathways research in criminology continues to proliferate. Much of this scholarship highlights how women's trauma histories initiate subsequent offending pathways (Gaarder and Belknap 2003; McDaniels-Wilson and Belknap 2009; Simpson et al. 2008). For example, Simpson et al. (2008) found that girls who were sexually abused before sixth grade and who had consensual sex before age 13 were three and a half times more likely than other women to be childhood onset offenders. Another good example is Caputo's (2008) *Out in the Storm*, an ethnographic study of 38 drug-addicted women living as shoplifters and sex workers. In a succinct summation of the feminist pathways approach, Caputo notes of her participants: "Childhood trauma occurring within the home resonates as the first important part of women's journeys to drugs and crime... This [trauma] triggered a wave of negative reactions, a misdirected coming of age during the women's adolescence that would carry through life" (pp. 27–28). Finally, still another refinement comes from feminist efforts to apply the life course model to *non-offending* outcomes. For example, the feminist pathways model has been used to explain how battered women's past victimization experiences can initiate help-seeking trajectories that guide their response to IPA over time (Burgess-Proctor 2011).

Likewise, it is easy to find examples of cutting-edge gendered pathways research in criminology. At the forefront is Peggy Giordano's (2010) magnificent *Legacies of Crime*, a longitudinal study of children whose parents were incarcerated as teens. Giordano uses three waves of data collection: (1) initial interviews with teenage offenders at the time of their incarceration, (2) follow-up interviews with the now-adult participants 14 years later, and (3) subsequent interviews with participants and their children several years after that. The study explores both the gendered and "generic" family dynamics that foster the intergenerational transmission of offending, violence, substance abuse, and other antisocial behaviors over time.

Giordano's (2010) analysis constitutes cutting-edge feminist theorizing because it imposes a gendered lens on the traditional life course approach to studying the intergenerational transmission of violence and offending:

> If an essential theme within the mostly male tradition of life-course research is that of continuity and change, core concerns within the gender and crime literature are questions of similarity and difference. Thus, researchers continue to grapple with the issue of whether and to what degree theoretical perspectives, and even basic facts about crime that developed largely around male-focused studies, "fit" with the life experiences of women and girls...Focusing only on [generic, as opposed to gendered] features of the women's lives, however, ignores many of the realities that have been effectively depicted by feminist researchers. (pp. 19–24)

These gendered realities, Giordano notes, necessarily shape a young woman's family environment as well as her parenting practices – both of which have implications for the intergenerational transmission of antisocial behavior (p. 24).

There are other noteworthy examples of gendered pathways research. For example, several recent studies have attempted to predict women offenders' recidivism risk over time using a gendered pathways approach (Huebner et al. 2010; Reisig et al. 2006). Likewise, Block and colleagues (2010) do a masterful job of exposing the absence of gender in life course criminology by evaluating whether Farrington's (2003) "widely accepted conclusions" about criminal career research are applicable to, or reflect the experiences of, female offenders. Examining offending pathways of a nationally representative sample of offenders in the Netherlands, the authors found that women were significantly more likely than men to begin offending in adulthood (mean age of onset: 21.5 for men versus 29.3 for women). Noting that age of onset is a key indicator in criminal careers research, the authors caution: "the findings of significant gender differences in age of onset...not only challenge widely accepted conclusions of developmental life-span criminology but also require us to rethink life-span patterns in and out of crime for women" (p. 93). Finally, examining how the consequences of incarceration are gendered, Block et al. (2010) observe that imprisonment is not just damaging to the woman herself but also to the loved ones for whom she cares. This observation echoes Giordano's (2010) conclusions about the gendered impact of parental (and especially *maternal*) incarceration on young people's criminal propensity – both observations not easily made without a gendered pathways perspective.

Social Learning and Social Control Theories

The life course model is not the only criminological theory to have undergone feminist revision in recent years. Feminist scholars have offered refinements of two other titans of criminological theory: social learning theory and social control theory.

Throughout criminology, social learning explanations of offending have been rooted in "differential" models, including differential association and differential reinforcement. These theoretical models propose that individuals learn criminal behavior through exposure to some pro-crime, learning-facilitating mechanism. In Sutherland's (1924) differential association theory, that mechanism is definitions favorable to lawbreaking; in Akers' (1973; Burgess and Akers 1966) differential reinforcement model (often referred to in criminology simply as "social learning theory"), that mechanism is crime-reinforcing stimuli. In other words, differential exposure to these mechanisms distinguishes between people who learn to commit crime and people who do not (i.e., whose exposure was insufficiently high).

In contrast, social control theories start from the opposite approach: social control theorists assume that lawbreaking is a natural tendency and that, absent sufficient controls, people left to their own devices will become deviant. Therefore, social control models seek not to explain offending but rather to explain conformity – that is, what allows law-abiding people to resist their natural desire to offend. Arguably the most well-known social control theory in criminology belongs to Hirschi (1969), who theorized that strong bonds to pro-social others facilitate conformity. As any undergraduate theory student can no doubt recite, Hirschi (1969) articulated four elements of a social bond: attachment, commitment, involvement, and belief. For Hirschi – at least until the emergence of his later, self-control model (Gottfredson and Hirschi 1990) – people with attachment to law-abiding role models, commitment to pro-social goals, involvement in pro-social activities, and belief in good citizenship are predicted to experience sufficient social control so as to keep them from engaging in deviance.

As with the life course model discussed previously, both the social learning and social control models have undergone feminist reinterpretation in recent years. In fact, Giordano's (2010) aforementioned study is theoretically important for a second reason: it offers a gendered refinement of the traditional social learning model. While Giordano explicitly states that traditional frameworks like social control and social learning should not be abandoned, she recognizes that "attention to gendered processes" also is crucial for developing comprehensive, integrated, and fully realized theoretical explanations of girls' and women's criminal involvement (p. 21). Indeed, Giordano's revision of the traditional social learning model incorporates elements of symbolic interactionist and feminist theory to identify three concepts that are particularly relevant to girls' offending pathways: self and identity, agency, and emotion (p. 30).

Recent scholarship has likewise offered a gendered refinement of the traditional social control model. Notably, Zimmerman and Messner (2010) explore the effect of neighborhood disadvantage on the gender gap in adolescent violent crime. Using data from Chicago neighborhoods, the authors found that, as expected, exposure to violent peers increased (for girls and boys alike) as neighborhood disadvantage worsened, but the effect of violent peer exposure was more pronounced for *girls* than boys. Thus, their original social control explanation – that informal social control is a stronger buffer against violent peer influence for girls than boys – was not supported. Instead the data suggested a more "complex causal chain," leading the authors to "reaffirm the fundamental premise of feminist perspectives: explaining social phenomena requires an understanding of 'gendered lives'. . .Such variation likely reflects the different ways that gender organizes" daily life (p. 974). As with social learning theory, then, social control theory also has undergone refinement by contemporary feminist scholars.

Conclusion: Making Connections to Narrow the Theoretical Divide

The examples in this entry make clear the abundance of truly innovative and impressive theory development that is occurring in contemporary feminist criminology. While it is true that cutting-edge feminist theorizing may be broadly divided into efforts to *advance feminist criminological* *theory* and efforts to *refine mainstream criminological theory*, these initiatives do not operate entirely apart from one another's influence. Indeed, as these two bodies of scholarship share a common goal – the infusion of gender (and other social systems) into contemporary criminological theory – there are plentiful examples of cross-fertilization between the two. As each enterprise necessarily informs the other, a series of theoretical and empirical questions remain.

For example, given that Black girls living in distressed urban neighborhoods are highly likely to experience sexual victimization (Miller 2008), and that sexual victimization is an important predictor of women's later offending (McDaniels-Wilson and Belknap 2009; Simpson et al. 2008), how do we contextualize poor, Black women's offending? Moreover, given that their social marginalization sharply limits the survival strategies of inner-city Black girls to include primarily isolation and social withdrawal (Jones 2010; Miller 2008), how can we better understand their responses to intimate partner victimization later in adulthood (Burgess-Proctor 2011; Potter 2006)?

Additionally, delinquent peers are a demonstrated risk factor for youth-onset offenders (Simpson et al. 2008). How does this finding square with Zimmerman and Messner's (2010) findings about the effect of delinquent peer relationships on girls versus boys? How are the processes governing the effects of delinquent peer attachment gendered (and/or raced and classed)? Does delinquent peer attachment vary according to peer group sex composition (Zimmerman and Messner 2010)? Further, in light of evidence suggesting gender differences in adult-onset offending (Block et al. 2010), how do peer influences shift and change over time, and how does their relevance differ for boys and girls versus adult men and women?

Answers to these questions and other similar inquiries will no doubt be the focus of cutting-edge feminist criminological theorizing in the coming years. But what will be the status of feminist theory in the overall discipline a decade from now? Despite over 40 years of scholarship, "traditional" and "feminist" theoretical frameworks today still are cast in opposition to one another. Acceptance of this (false) dichotomy remains an entrenched problem in criminology.

For example, in the introduction to their comprehensive anthology assessing the current state of criminological theory, Cullen et al. (2006) remark: "Because they [were] invented by males, written about males, and originally tested on males, feminist scholars see traditional or core theories as 'men's theories' that do not capture the special circumstances inherent in women's criminality" (p. 15). Leaving aside the troubling assumption that only feminist scholars can "see" the problem with compulsory generalization to women of theories developed for and tested exclusively on men, the most vexing aspect of this statement is the positioning of life course, social control, and social learning theories as "core" frameworks to which feminist perspectives stand in opposition. As the examples offered in this entry reveal, cutting-edge work in feminist criminological theory engages and uncovers social mechanisms that cut to the very core of criminological and sociological theorizing, despite their continual categorization as derivatives of mainstream theory. Hopefully, as exemplary feminist scholarship continues to inform, refine, and reconceptualise both "traditional" and feminist criminological theories, the latter will come to be viewed by the discipline as "mainstream" explorations. Indeed, perhaps feminist criminological scholarship will truly have found its cutting edge when the theoretical and ideological divide between mainstream and feminist bodies of knowledge has disappeared altogether, and feminist perspectives are presented as core theoretical contributions in their own right.

Related Entries

- ▶ Feminist Theory in the Context of Sexual Violence
- ▶ Gendered Theory and Gendered Practice

▶ Gendering Traditional Theories of Crime
▶ Women of Color and Crime

Recommended Reading and References

Akers RL (1973) Deviant behavior: a social learning approach. Wadsworth, Belmont

Belknap J (2007) The invisible woman: gender, crime, and justice, 3rd edn. Wadsworth, Belmont

Benson ML, Simpson SS (2009) White-collar crime: an opportunity perspective. Routledge, New York

Block CR, Blockland AAJ, van der Werff C, van Os R, Nieuwbeerta P (2010) Long-term patterns of offending in women. Fem Criminol 5(1):73–107

Burgess RL, Akers RL (1966) A differential association-reinforcement theory of criminal behavior. Soc Probl 14:128–147

Burgess-Proctor A (2006) Intersections of race, class, gender, and crime: future directions for feminist criminology. Fem Criminol 1(1):27–47

Burgess-Proctor A (2010) Looking back, looking ahead: assessing contemporary feminist criminology. In: Copes JH, Topalli V (eds) Criminological theory: readings and retrospectives. McGraw-Hill, New York, pp 431–443

Burgess-Proctor A (2011) Pathways of victimization and resistance: toward a feminist theory of battered women's help-seeking. Justice Quart (in press)

Caputo G (2008) Out in the storm: drug-addicted women living as shoplifters and sex workers. Northeastern University Press, Boston

Collins PH (2000) Black feminist thought: knowledge, consciousness, and the politics of empowerment, 2nd edn. Routledge, New York

Cullen FT, Wright JP, Blevins KR (eds) (2006) Transaction Publishers, New Brunswick

Daly K (1998) Gender, crime, and criminology. In: Tonry M (ed) The handbook of crime and punishment. Oxford University Press, Oxford, pp 85–108

Daly K, Chesney-Lind M (1988) Feminism and criminology. Justice Quart 5(4):497–538

Daly K, Stephens DJ (1995) The "dark figure" of criminology: towards a Black and multi-ethnic feminist agenda for theory and research. In: Rafter NH, Heidensohn F (eds) International feminist perspectives in criminology: engendering a discipline. Open University Press, Philadelphia, pp 189–215

Elder J, Glen H (1998) The life course as developmental theory. Child Dev 69(1):1–12

Erez E, Berko A (2010) Pathways of Arab/Palestinian women to crime and imprisonment: an intersectional approach. Fem Criminol 5(2):156–194

Farrington DP (2003) Developmental and life-course criminology: key theoretical and empirical issues – the 2002 Sutherland Award address. Criminology 41(2):221–225

Gaarder E, Belknap J (2003) Tenuous borders: girls transferred to adult court. Criminology 40(3): 481–517

Giordano PC (2010) Legacies of crime: a follow-up of the children of highly delinquent girls and boys

Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Hirschi T (1969) Causes of delinquency. University of California Press, Berkeley

Huebner B, DeJong C, Cobbina J (2010) Women coming home: long-term patterns of recidivism. Justice Quart 27(2):225–254

Jones N (2010) Between good and Ghetto: African American girls and inner-city violence. Rutgers University Press, New Brunswick

Laub JH, Sampson RJ (2003) Shared beginnings, divergent lives: delinquent boys to age 70. Harvard University Press, Cambridge, MA

McDaniels-Wilson C, Belknap J (2009) The extensive sexual violation and sexual abuse histories of incarcerated women. Violence Against Women 14(10):1090–1127

Miller J (2008) Getting played: African American girls, urban inequality, and gendered violence. New York University Press, New York

Miller J (2010) Commentary on Heidensohn's "The deviance of women": continuity and change over four decades of research on gender, crime and social control. Br J Sociol 61(s1):133–139

Miller J, Mullins CW (2006) The status of feminist theories in criminology. In: Cullen FT, Wright JP, Blevins KR (eds) Taking stock: the status of criminological theory. Transaction Publishers, New Brunswick, pp 217–249

Potter H (2006) An argument for Black feminist criminology: understanding African American women's experiences with intimate partner violence using an integrated approach. Fem Criminol 1(2):106–124

Reisig M, Holftreter K, Morash M (2006) Assessing recidivism risk across female pathways to crime. Justice Quart 23(3):384–405

Sampson RJ, Laub JH (1993) Crime in the making: pathways and turning points through life. Harvard University Press, Cambridge, MA

Sampson RJ, Laub JH (2005) A life-course view of the development of crime. Ann Am Acad Pol Soc Sci 602:12–45

Simpson SS (1989) Feminist theory, crime, and justice. Criminology 27(4):605–632

Simpson SS, Yahner JL, Dugan L (2008) Understanding women's pathways to jail: analysing the lives of incarcerated women. Aust N Z J Criminol 41(1):84–108

Sokoloff NJ, Burgess-Proctor A (2011) Remembering criminology's "forgotten theme": seeking justice in US drug policy using an intersectional approach. In: Bosworth M, Hoyle C (eds) What is criminology? Oxford University Press, Oxford, pp 235–248

Sokoloff NJ, Pratt C (eds) (2005) Domestic violence at the margins: readings on race, class, gender, and culture. Rutgers University Press, New Brunswick

Sutherland E (1924) Principles of criminology. University of Chicago Press, Chicago

Zimmerman GM, Messner SF (2010) Neighborhood context and the gender gap in adolescent violent crime. Am Sociol Rev 75(6):958–980

Feminist Theory and Domestic Violence

Jo Dixon
Department of Sociology, New York University,
New York, NY, USA

Overview

For much of the past, US society turned a blind eye to violence between intimates. However, with the emergence of the progressive movement in the early 1900s and the women's movement in the 1960s, domestic violence began being constructed as a social problem worthy of attention by academics, policy makers, and the criminal justice system. Feminist constructions of domestic violence as a social problem vary and range from early constructions focusing on family preservation to later ones emphasizing mental illness, sex differences in intersubjectivity, male domination, family conflict, and survival. Portending to reduce and/or punish domestic violence, concomitant state social control mechanisms emerged. This entry traces the feminist constructions of domestic violence as a social problem as well as the relationship of these constructions to the social control mechanisms employed by state institutions, especially the criminal justice system.

Domestic Violence: Preserving Family and Disciplining Poor Families

Prior to the 1870s domestic violence in the USA was constructed as a private family matter, and the major social control mechanism was the interwoven fabric of family, church, and state constituted by Christian morality. The legal system of this period was a combination of English common law and the Holy Bible. While Puritan church courts tried domestic violence cases, the focus was on preserving the family rather than punishing wrongdoing. Moreover, most congregations accepted a sinner's confession, and only a few dozen cases per year involved any type of discipline (Pleck 1987). Although the Massachusetts Body of Liberties of 1641 and the 1672 law against spouse abuse made wife beating illegal, society accepted hierarchical relationships and the use of physical force by parents, husbands, and masters (Pleck 1987). By the 1700s the "rule of thumb" principle was established by an English judge and reinforced in the early 1800s by Southern appellate courts ruling that a husband could beat his wife as long as the beating tool was no thicker than a man's thumb.

In the late 1800s and early 1900s, the women's progressive movement lobbied for women's participation in the public sphere, but they did not place the private sphere and domestic violence as a high priority among their legislative and judicial agendas. However, there were indirect campaigns against domestic violence through the temperance, child-welfare, and purity movements aimed at disciplining minority, immigrant, and the working class families. One of the institutional legal developments during this period was the court of domestic relations, a tentacle of the early welfare state and the forerunner of today's family court. The domestic relations court personnel consisted of welfare officials, social workers, police, and prosecutors involved in the protection of dependent mothers and children. The court's main agenda was to discipline poor families and force husbands to be sober, work, and financially support their dependent wives and children (Willrich 2003). While police often encountered domestic violence, unless the violence was extremely severe, they would not arrest anyone. In some states, severe cases of domestic violence were punished via the whipping post. However, since forcing husbands to financially support wives and children was central, neither wives nor the state showed much interest in the imposition of jail or prison for domestic violence offenders. Throughout this period, some legislatures advanced making domestic violence illegal, but the criminal justice system turned a blind eye when it came to arresting and prosecuting batterers. Thus, domestic violence remained a private family affair to be ignored or resolved by informal dispute resolution in civil domestic relations courts.

Domestic Violence: Mental Illness

With the development of psychoanalytic theory in Europe and the USA, a construction of domestic violence as mental illness emerged in the 1940s. While male batterers were labeled as emotionally disturbed, the primary focus centered on female victims whose victimizations were attributed to sexual problems or a masochistic personality whereby victims provoked violence and subsequently remained in violent relationships. Only psychoanalytic therapy could uncover the unconscious motivations that led battered women to search for and stay with their batterers. Alternative psychological constructions challenged Freudian theories and linked domestic violence to cognitive or behavioral processes that could be diagnosed and treated by other forms of therapy. Feminists in the American Psychological Association opposed including the masochistic personality disorder in the DSM (Diagnostic and Statistical Manual of Mental Disorders) which was becoming the major diagnostic tool used by psychologists, psychiatrists, and insurance companies. While the masochistic disorder was blocked, an alternative category called the self-defeating disorder was included.

One of the major feminist opponents to the masochistic or self-defeating constructions of battered women was Lenore Walker (1979) who called the psychological effects of the trauma produced by domestic violence “the battered women’s syndrome.” Unlike earlier psychological perspectives arguing that women with particular psychological pathologies enter into battering relationships, Walker proposed a feminist theory arguing that psychological pathologies followed the trauma of battering. As an expert legal witness, Walker became instrumental in promoting feminist therapy and changing the image of battered women in the legal system. Nonetheless, supporters of psychological constructions of domestic violence as mental illness advocated individual and couples therapy as the appropriate mechanism for alleviating domestic violence.

Domestic Violence: Institutional Sex, Family and Legal Inequality, or Family Conflict

In the 1980s feminist theories of violence against women grew (Gilligan 1982; MacKinnon 1987), and in the 1990s feminist sociologists and criminologists began conducting empirical studies to identify the demographics of battering. Though domestic violence occurs across all race, class, educational levels, and age groups, sociological studies consistently showed that violence, whether it be in the public or private sphere, is more likely to occur among those who are black, lower class, less educated, and young (Fagan and Browne 1993; Kruttschnitt 1993). However, the most contentious debate focused on whether domestic violence, like crimes in the public sphere, was sex asymmetrical with men being more likely to engage in violence or whether men and women equally engage in violence in the private sphere.

Domestic Violence: Female Victims of Sex, Family, and Legal Inequality

During the 1980s radical and cultural feminist constructed domestic violence as a binary sex asymmetrical social problem caused by sex differences and sex inequality in the social institutions of sex, family, and law. Because both cultural and radical feminist theories of domestic violence constructed domestic violence in terms of group membership and focused on the patriarchal institutions of sex, family, and law, the issue of male victimization was ignored. Cultural feminists attributed violence against women to sex differences in intersubjective understandings of males and females (Gilligan 1982) with the males' voice being dominant in institutions, while radical feminists pointed to the role of violence and sex inequality in the male domination of women (McKinnon 1987). However, unlike those who focused on mental illness, both viewed cultural and structural changes in sex, family, and legal institutions as the social control mechanism for dealing with domestic violence. Cultural feminists argued for consciousness-raising about domestic violence and rallied for more sex equality by introducing the *female voice* in institutional reforms of sex, family, and law, while radical feminists lobbied for institutional reforms in sex, family, and law that created more sex equality by reflecting women's *interests*.

Along with grassroots organizations, cultural and radical feminists viewed domestic violence as a public rather than private family matter. Thus, they supported state aid to women's shelters where battered women could escape their battering, engage in female consciousness-raising about domestic violence, and increase their economic independence. A small number of feminists argued for more criminal justice involvement such as battering programs and rehabilitation *or* arrest and incarceration. However, most feminists were skeptical of criminal justice interventions, maintaining that victim empowerment would be better served by paying attention to the interests and voices of battered women. In addition, most believed that arrest policies without economic resources for victims would keep women in abusive relationships or lead them and their children to a life of poverty if they left their batterers.

Domestic Violence: Male and Female Victims of Family Conflict

While radical and cultural feminists continued to lobby for judicial and legislative reforms that would improve the status of women as a class, liberal feminists and researchers were busy arguing that domestic violence was individual and sex-neutral. Unlike radical and cultural feminist theories of domestic violence characterizing sex, family, and law as unequal patriarchal institutions where men use violence to dominate and control women, liberal feminist theories focused on individuals and family conflicts whereby both men and women used violence to deal with the conflicts and stresses associated with intense emotional relationships. Murray Straus (1980), the most vigorous proponent of the liberal feminist perspective, constructed domestic violence as a sex symmetrical family conflict problem rather than a sex asymmetrical woman's problem. He argued that the emotional intensity of marriage coupled with the stresses of family life leads to frustrations and aggressions for married men and women such that sex differences in the rates of violence found in the public sphere are absent in the private sphere. In 1975 Straus, Gelles, and Steinmetz interviewed a large sample of married and cohabitating couples to assess their use of violence in conflict resolution. Results from this study (and a later study in 1985) indicated that the number of acts of aggression committed by men and women was the same and that females were slightly more likely than men to be perpetrators of severe acts (Straus et al. 1990). The family violence construction of domestic violence turned attention away from a sex-specific focus on wife abuse to a sex symmetrical focus on family violence. The victim status reserved for women by cultural and radical feminists was extended to men, and the problem changed from a woman's problem to a family problem.

As family conflict constructions of domestic violence became more pervasive, the limited state support for battered women's shelters and victim's services declined and was replaced with informal conflict resolution intervention by police and family courts. As in the Progressive Era, the social control mechanisms employed for domestic violence based on liberal feminist family violence constructions focused less on welfare and services for women and, instead, on family conflict resolution in civil family courts.

Liberal feminist family violence theories of domestic violence gained momentum during the era of rehabilitation in criminal justice. Consequently, police crisis intervention and court diversionary programs became the predominant social control mechanisms for dealing with family violence. By 1980, most spousal assault cases brought to the attention of the court were handled in civil family courts (Dobash and Dobash 1992). Once again, domestic violence was regarded as a private family matter to be resolved through informal court resolution rather than a public matter to be adjudicated by criminal courts or resolved by welfare state shelters and victim services providing battered women with the economic resources to leave their batterers.

Domestic Violence: Female Survivors of Sex, Family, and Legal Inequality

For some feminists the radical and cultural feminist constructions of domestic violence as sex asymmetrical with women construed as passive victims without agency were uncomfortable. Equally uncomfortable was the liberal construction of domestic violence as conflict between individuals with agency in families or intimate relationships without regard for sex inequality in existing social institutions.

Taking both agency and institutionalized sex inequality in intimate relationships into account, some feminists constructed a view of domestic violence as sex asymmetrical with battered women being perceived as survivors (Hoff 1990). This type of construction led feminists to focus on the social context and survival strategies of battered women, especially women using self-defense against their batterers (Dobash and Dobash 1992; Maguigan 1991; Schneider 2000; Dixon 1995; Ferraro 1989). Another more nuanced position supported by research argued that a large number of families suffered from occasional outbursts of violence from either husbands or wives (common violence) while a significant number of other families experienced systematic violence from husbands engaged in patriarchal terrorism (Johnson 1995).

This genre of feminism presented a more complex construction of domestic violence which viewed women as active agents attempting to survive the inequality in sex, family, and legal institutions. This construction attempted to grapple with the issue of agency for battered women without denying the issue of inequality in social institutions. The feminist constructions of domestic violence that viewed battered women as survivors provided the basis for significant changes for cases involving battered women who killed their assailants. The sex symmetrical survival construction of battered women led to statutory and case law reforms that expanded self-defense options for battered women.

Domestic Violence: Intersection of Sex, Sexual Orientation, Race, and Class

As feminist scholarship and legal practice in the area of domestic violence expanded, criticisms began to emerge focusing on the lack of attention to the race, class, and sexual orientation contexts of domestic violence. Critical race theorists noted that existing feminist theories of domestic violence reflected the experiences of white battered women with little attention to the experience of black battered women or to the intersection between sex and race in battering (Crenshaw 1989). Similarly, feminists interested in issues of class inequality suggested that current feminist theories limited their analysis to battered women without recognizing social class variations or the intersection of sex and class in domestic violence. Lettie Lockhart's (1991) cross-racial perspective on domestic violence was one of the few empirical studies that examined the intersection of sex, race, and class in domestic violence.

Domestic Violence as Crime: Arrest, Prosecution, and Specialized Courts

Increasing crime rates in the USA during the 1970s and 1980s led to a situation whereby the criminal justice system began to criminalize many social problems once belonging to the private sphere such as drugs and domestic violence. Earlier debates over the merits of constructing domestic violence as a public offense adjudicated and punished in criminal courts or a civil offense handled by conflict resolution resurfaced.

After the results from the Minnesota Domestic Violence Experiment showed that arrest provided a significantly greater reduction in domestic violence than informal conflict resolution by police (Sherman 1992), some feminists joined hands with the criminal justice system to advocate mandatory arrest as the appropriate social control mechanism for dealing with domestic violence. Six replications of the original Minneapolis experiment were commissioned by the National Institute of Justice. Although some analyses yielded modest support for the deterrent effect of arrest found in the Minnesota study (Maxwell et al. 2002), other results suggested that deterrent effects varied across settings (Garner et al. 1995), and offender characteristics linked to stakes in conformity such as marital status and employment (Sherman and Smith 1992). The finding that arrest decreased future violence against women when offenders were married and employed but increased future violence when offenders were unmarried and unemployed produced a dilemma for feminists who construed women as a universal category. Some feminists also pointed to the absence of gays and lesbians in studies of domestic violence.

Research showing that female victims were more likely to use police in the future if offenders were arrested in accordance with the preferences of victims also called into question the degree to which mandatory arrest was empowering to women (Hickman and Simpson 2003). However, by this time the train had left the station. Mandatory arrest had become the predominant criminal justice policy in the USA. When research reported that prosecutors were dropping domestic violence cases even after mandatory arrest policies were initiated, many prosecutors followed with mandatory prosecution. Some feminists applauded mandatory prosecution (Wills 1997), while others argued that it empowered the criminal justice system while disempowering battered women (Ford 1991).

After the development of specialized problem-solving drug courts and the passage of the Violence Against Women Act (1994), the federal government provided money for the development of innovative criminal justice policies for dealing with domestic violence cases: specialized domestic violence prosecution bureaus, no-drop prosecution policies, specialized domestic violence court, victim advocacy programs, and court-community partnerships. Feminist debating the merits of mandatory arrest began debating the merits of mandatory prosecution and specialized domestic violence courts, i.e., criminal justice ownership of the domestic problem.

Domestic violence is no longer ignored or constructed as a private family problem. It is now constructed as a public crime problem solvable by criminal justice agencies. However, what is not clear is whether the specialized domestic courts will operate in a manner that empowers victims or, like the early domestic relations and later family courts, will operate more as an institution to discipline and control poor and minority family life by punishing poor minority men while limiting state resources for female victims.

Recommended Reading and References

Crenshaw K (1989) Demarginalizing the intersection of race and sex: a black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. Univ Chi Legal F 89:139–168

Dixon J (1995) The nexus of sex spousal violence and the state. Law Soc Rev 29(2):359–374

Dobash RE, Dobash RP (1992) Women, violence and social change. Routledge, London

Fagan J, Angela B (1993) Violence between spouses and intimates: physical aggression between women and men in intimate relationships. In: Reiss AJ, Roth JA (eds) Understanding and preventing violence, vol. 3, social influences. National Academy Press, Washington, DC, pp 115–291

Ferraro K (1989) Policing woman battering. Soc Probl 36:61–74

Ford DA (1991) Prosecution as a victim power resource: a note on empowering women in violent conjugal relationships. Law Soc Rev 25:313–334

Gilligan C (1982) In a different voice. Harvard University Press, Cambridge

Gordon L (2002) Heroes of their own lives: the politics and history of family violence. University of Illinois Press, New York

Hickman L, Sally S (2003) Fair treatment or fair outcome: the impact of police behavior on victim reports of domestic violence incidents. Law Soc Rev 37(3):607–634

Hoff LA (1990) Battered women as survivors. Routledge, London

Johnson M (1995) Patriarchal terrorism and common couple violence: two forms of violence against women. J Marriage Fam 57(2):283–294

Kruttschnitt C (1993) Gender and interpersonal violence. In: Reiss J, Roth A (eds) Understanding and preventing violence, vol. 3, social influences. National Academy Press, Washington, DC, pp 293–396

Lockhart L (1991) Spousal violence: a cross-racial perspective. In: Hampton RL (ed) Black family violence. Lexington Books, Lexington, pp 86–101

MacGuigan H (1991) Battered women and self-defense: myths and misconceptions in current reform proposals. Univ Mich Law Rev 140(2):379–486

MacKinnon C (1987) Feminism unmodified: discourses on life and law. Harvard University Press, Cambridge

Pleck E (1987) Domestic tyranny: the making of American social policy against family violence from colonial times to the present. Oxford University Press, New York

Renzetti CM, Miley CH (1996) Violence in gay and lesbian domestic partnerships. Harrington Park, New York

Schneider E (2000) Battered women and feminist lawmaking. Yale University Press, New Haven

Sherman L (1992) Policing domestic violence: experiments and dilemmas. Free Press, NewYork

Straus M (1980) The marriage license as hitting license: evidence from popular culture, law and social science. In: Straus MA, Hataling GT (eds) The social causes of husband-wife violence. University of Minnesota Press, Minneapolis, pp 39–50

Straus M, Gelles R, Steinmetz SK (1990) Behind closed doors: violence in the American family. Doubleday/Anchor, New York

Walker L (1979) The battered woman syndrome. Harper and Row, New York

Willrich M (2003) City of courts socializing justice in progressive era Chicago. Cambridge University Press, Cambridge

Wills D (1997) Domestic violence: the case for aggressive prosecution. Univ Calif Los Angel 7:173–199

Feminist Theory in the Context of Rape

► Feminist Theory in the Context of Sexual Violence

Feminist Theory in the Context of Sexual Violence

Callie Marie Rennison
School of Public Affairs, University of Colorado Denver, Denver, CO, USA

Synonyms

Feminist theory in the context of rape; Sexual assault and violence against females

Overview

Feminist theories offer significant contributions to the field of criminology, especially to the understanding of sexual violence since it was first introduced into the field approximately 40 years ago. Since their initial appearance, feminist theories have matured and grown. While some continuity is found among feminist theories, great variation exists in their underlying assumptions, approaches, and goals. This entry describes feminist theories in the context of sexual violence and is structured as follows. First, the background in which feminist theories emerged is described. Next, feminist theories of sexual violence are introduced. This entry then describes some specific integrative theories of sexual violence that are informed by the feminist paradigm. Subsequent sections address myths and controversies, as well as questions that remain unanswered in the literature. This entry concludes with a list of relevant references germane to feminist theories of sexual violence.

Introduction

Perhaps more than any other theoretical tradition, feminist theories greatly changed our understanding of sexual violence. Prior to their introduction, criminological research and theories were androcentric, having been shaped and informed solely by male understandings, perspectives, and experiences. Initial criminological understanding of victimization and offending focused almost entirely on male behavior and perceptions, and when attention was given to females, it was generally through the lens of sexism, misogyny, and the view of females as male property. Females who engaged in nontraditional behavior (i.e., deviance or criminal) were negatively characterized as morally corrupt, devious, manipulative, and even ill. And females who were victimized were similarly viewed as flawed, devious, and the sole cause of their victimization.

One of the most important contributions of feminist theories was to recast assumptions and understanding of criminal behavior and victimization, especially as it relates to sexual violence. Feminist theories introduced a more complete understanding of sexual offending and victimization because they started to take account of the actual experiences of females and highlight the considerable influence of gender (versus sex) as a construct on criminological behavior and victimization. However, feminist theories offer far more than a focus on the offending and victimization experiences of females and the effects of sexism. Feminist theories represent a reconceptualization of key concepts and ideas. They focus on how gender organizes the field of criminology, our understanding of offending and victimization, and the structure and functioning of criminal justice institutions. That feminists prioritize gender is widely established, though it is important to emphasize that "gender" is not equivalent to or interchangeable with "sex." Gender is the "sociocultural and psychological shaping, patterning, and evaluating of female and male behavior" (Schur 1984, p. 10). Sex is a biologically based and stable characteristic identified with two categories: male and female. Before discussing contemporary feminist theories of sexual violence, it is important to first define feminist theories, and to consider the theoretical context in which these theories emerged.

Background Description

Attempts to define a singular feminist theory is both naïve and incorrect because there is no "feminist theory" that is based on a single paradigm. Recent literature indicates at least 12 variants of feminist theory, though not all address sexual violence.

As such, the phrase "feminist theories" refers to a decentered and diverse collection of perspectives and methodologies based on various ideas about the basic assumptions regarding inequality, the role of gender and gender relations, the issues and problems requiring attention, and the methods needed to address these issues and problems. More succinctly, feminist theories are "woman-centered description[s] and explanation[s] of human experience[s] and the social world" and they are guided by the notion that "gender governs every aspect of personal and social life" (Danner 1991, p. 51).

There are common elements shared by the various feminist perspectives that distinguish them as a group from traditional criminology theoretical orientations (Daly and Chesney-Lind 1988). First, they view gender as a complex, social, historical, and culture product and not a natural fact. Second, gender and gender relations are fundamental organizers of social life and institutions. Third, gender relations in society and the role of masculinity and femininity are not equal or symmetrical. Rather, masculinity and femininity are used to organize and subordinate females based on male superiority. Fourth, traditional criminological theories are based on a gendered perspective that reflects solely a male's perspective of the criminological world. Fifth and finally, feminist theories agree that females should not be invisible, subordinate, peripheral, or viewed as appendages to males. Instead, feminist theories argue that females should be at the center of theory development, research, and intellectual inquiry.

Before turning to a description on how feminist theories inform sexual violence, it is important to describe the "traditional" criminological perspective as well as Liberal feminist theory that it led to. It is in this context that contemporary feminist theories focusing on sexual violence emerged.

Background Historical Context

Androcentric Theoretical Tradition

Traditional criminological theories virtually ignored the presence and experiences of females. Theories were based on male experiences and intellectual views, and these male-centered theories and understandings were generalized to the population at large – including females. When females were addressed in theories focused on criminology, it was assumed that male and female differences stemmed from biological differences. Evidence of this is found in Durkheim's work on homicide in which he described the importance of biological differences between males and females. He noted that females have smaller brains, smaller overall size, lesser strength, and were simply more primitive and less socialized than males. Further, Durkheim argued that these differences between males and females grew as societies evolved (DiCristina 2006). Pollak's (1950) views on the criminality of women also highlighted the importance of biological factors. He indicated that female criminals have an excess of male biological and psychological characteristics and make them more similar to male criminals than to female criminals or non-offenders. In general, this early work failed to consider external influences such as social or economic forces. In general, when these traditional perspectives discussed "gender differences," it was done in a way to highlight differences and deny the presence of gender inequality. For instance, disparities in strength and in aggression (in general) between males and females and the greater innate nurturing and care-giving behaviors (in general) found among females compared to males were seen as reflecting the natural order of things.

Given the belief that gender differences reflected the natural state of affairs and were biological imperatives, few scholars drew attention to strategies for social change until the introduction of Liberal feminist theory. Although this theory does not offer direct insight into sexual violence, it is important to understand the background role it plays in the emergence of contemporary feminist theories that do.

Liberal Feminist Theory

Liberal feminist theory challenges the notion of the natural state of biological imperatives that long kept males and females in separate spheres. This perspective holds that socially created, rather than biologically based, gender inequality underscores discriminatory social policies that maintain separate and distinct spheres for males and females, including the socialization of specific and distinct gender roles for males and females. Liberal feminist theory argues that both females and males are socialized and indoctrinated to believe that females are not capable of, or appropriate for, participating in intellectual activities, in physical activities, or in roles with a public presence; rather, a female's natural and proper role is one behind the scenes, subordinate to males, and outside of paid labor. Thus, Liberal feminist theory contends, it is not through a biological imperative that males and females are found in different spheres, but rather through the process of socialization and laws that this segregation and structural inequality takes place.

The work of Liberal Feminists was used as a way of organizing our understanding of *offending* and did not address sexual violence. This is not surprising given the timing of its emergence. Liberal Feminism emerged in the late 1960s into the 1970s during a time when little was known about sexual violence as it was widely considered a private or personal matter. On the rare occasion that sexual violence found its way into the public domain, the victim (generally female) was blamed for her victimization.

Liberal feminist theory does not hold that the *system* is inherently unequal, but rather, the policies by which it operates are. Thus, while not explicitly noted, proponents of this perspective would likely argue that sexual violence could be addressed through policy/legal changes (e.g., divorce laws, restraining orders, marital rape laws, and domestic violence laws) and structural changes which would address women's abilities to be socially and financially independent (e.g., equal opportunity laws, affirmative action, and equal rights).

While policy change approach with the goal of gender equality is an important advancement, it proved problematic in practice. First, a focus on equal treatment uses males as the yardstick by which females are treated; and in a practical sense, this approach has been shown to exacerbate problems and ultimately work against females.

Today, the approach for equality has been largely dismissed by newer more radical variants of feminist theory focused on parity. Parity allows for sex-based differences in treatments and policies. These newer more radical perspectives currently dominate research and theory within the feminist paradigm, and it is among these perspectives that a focus on sexual violence is found.

State of the Art: Feminist Theories in the Context of Sexual Violence

While sexual violence is heavily researched, it receives comparatively little theoretical attention even among feminist perspectives. This section addresses two theories that most often form the basis of inquiry into sexual violence. Following this presentation, the discussion moves to describing some feminist theoretical perspectives focused specifically on sexual violence.

Radical Feminist Theory

Radical feminist theory views "male power and privilege as the source of all social relations, inequality, and violence against women" (DeKeseredy 2011, p. 298). It identifies *patriarchy*, especially in terms of controlling and exploiting female sexuality, reproduction, and familial relations, as the root of gender inequality and ultimately sexual violence. Other characteristics such as class are of secondary importance. One very important and critical way in which Radical feminist theory differs from other feminist theories is that it rejects official notions and definitions of violence because extant definitions and ideas of violence are male centered and fail to incorporate the full range of female victimization experiences. Instead, Radical feminist theory conceives of a broad definition of sexual violence, which places this violence on a continuum ranging from sexual harassment (less severe) to violent completed rape (most severe). This sexual violence continuum does not reflect what males embedded in a patriarchal social structure view as sexual violence; sexual violence is based only on the experiences and notions of females. By reconceptualizating sexual violence in a broader sense, Radical feminist theory clearly indicates that sexual violence is a means of social domination and not merely random acts of aggression.

Patriarchy is the core concept in this perspective and refers to the belief that society views women as the property of males, and therefore, societal norms convey that females are subject to male control. Because males, in general, are larger and more powerful than females, males are able to use aggression and power to subordinate and control females in every sphere of life.

Radical feminist theory identifies *sex* as the arena in which females are controlled. It is not necessary that all females be sexually victimized, nor is it necessary that all males offend against females. It is only necessary that acts of sexual violence be well known so that fear of this violence keeps females in a subordinated position. Thus, rape, sexual assault, and intimate partner violence stem from patriarchy and represent a form of social control. The continual threat of sexual violence serves to perpetuate male domination over female minds and bodies by force.

Radical feminist theory contends that the end of patriarchal relations and the elimination of oppressive social structures are the only means to end the gender inequality and the resulting sexual violence. This can be accomplished by providing females sexual, reproductive, and familial autonomy.

An important early work representative of Radical feminist thought is Brownmiller's (1975) work on rape. This revolutionary work brought the Radical feminist perspective to the forefront of research and reframed our understanding of rape. Brownmiller's work recast rape as a crime of power and control and a part of the arsenal of tools used to control, dominate, and subordinate females. Previously, rape was considered a crime of passion or an erotically driven act. Her work identified the patriarchal system of gender inequality and how it controls and subordinates females, condones sexual violence, and encourages and justifies sexual coercion. By moving the conversation of rape as a sexual crime to rape as an issue of power and control, greater understanding about the motivation and consequences of rape and sexual assault were made. Contrary to traditional views, it is now established that rape and sexual assault are acts of violence, sex is the weapon, and offenders are motivated by power, control, and anger. Further, current understanding includes that victims of rape and sexual assault do not incite an offender, nor does this violent act occur only to "bad" females (or males). This shift in perspective changed policy as well. Until the late 1970s, most states did not consider spousal rape a crime. It was believed that a wife could not be raped by a husband as their marriage formed a contract meaning the wife had to submit herself to her husband. Currently, rape of a spouse is a crime in all states as well as the District of Columbia.

A second important example of research informed by Radical feminist thought is Kelly's (1988) seminal work on sexual assault. In *Surviving Sexual Violence*, Kelly interviewed 60 females to examine the range of sexual violence they experienced over their life course. The interviews provided groundbreaking information on how females defined their sexual violence, the strategies they used to resist this victimization, and finally how they coped with their experiences. Using this information, Kelly conceptualized sexual violence broadly and placed it on a continuum. She concluded that several changes were needed. First, females need to recognize their experiences as sexual violence. Second, sexually victimized females need to view themselves as survivors rather than victims. Further, Kelly argued that women need to work together to end sexual violence. Since the presentation of these early classics, a wide range of research on sexual violence, rooted in the Radical feminist school, has been conducted. Areas covered in this body of research include behaviors that disproportionately victimize females such as pornography, domestic violence, incest, and sexual harassment.

Marxist Feminist Theory

Marxist feminist theory posits that *class* is the primary basis of gender inequality and sexual violence and *gender*, or male-to-female relations, is a secondary basis for inequality and sexual violence disproportionately experienced by females. From this primary perspective, gender inequality is based on class differences stemming from private property ownership and the inheritance of property – social phenomenon enjoyed primarily by males. Because females as a sex class were historically excluded from working in the labor force (except as a reserve army of labor), owning property or inheriting private property, they were disadvantaged and placed in weak/powerless economic positions that increased their vulnerability to exploitation including sexual violence. Yet another way, Marxist feminist theory contends that class forms the base for female disadvantage is the master–slave relationship found between husbands and wives in traditional marriage. In a traditional marital relationship, a wife serves as unpaid and exploited laborer to the benefit of her husband.

Marxist feminist theory recognizes that violence against women in not equally prevalent in all societies and that high rates of rape and sexual violence are found among modern capitalist societies. This results directly from the history and culture found in exploitative class relations that characterize a capitalist society. A classic and cornerstone treatment of sexual violence rooted in the Marxist feminist tradition is found in Schwendinger and Schwendinger's (1983) *Rape and Inequality*. Their research indicates that rape is not present in all societies, but rather that it varies based on gender equality. Based on an historical, cross-cultural, and anthropological analysis, the authors find that rape rates are highest in capitalist societies precisely because of gender inequality. Specifically, the inequality bred by a capitalist society enhances the conditions for female subordination and sexual violence. In contrast, they find that in noncapitalist societies, rape is rare, and egalitarianism between males and females high.

To end this class inequality which disadvantages females and enhances the conditions for sexual violence, the Marxist feminist perspective stresses the need to transform the capitalist societies to noncapitalist societies. The elimination of a class-based society will remove the power and privilege enjoyed solely by males (i.e., wealthy white males) and benefit females who will be freed from economic dependency on males. In a society free from gender and class stratification, females would be able to participate equally in the public and private sphere activities including paid labor, housework and child rearing. And finally, this change argues for the abolition of traditional marriage which views females as the private sexual property of males. These changes designed to end class stratification will in turn end or greatly reduce sexual violence.

Feminist Theoretical Offerings on Sexual Violence

The previous section offers an overview of two major perspectives of feminist theory and how they address sexual violence. In reality, the parsing of feminist theories is not so neat and theoretical contributions in the literature are more often integrative. The next section presents several theoretical contributions by feminists valuable for our understanding of sexual violence.

A Feminist/Male Peer Support Model of Separation and Divorce Sexual Assault

Until the presentation of this model, DeKeseredy and Schwartz (2009) noted that there were no theories specifically designed to address why males sexually assault female partners who end, or wish to end, their relationship. While much work has produced important information on risk factors of sexual violence, none has offered a unified model to address this particular form of sexual violence.

This model is grounded in the notion that divorce/separation sexual assault is an expression of the broader societal forces including societal patriarchy and male proprietariness. Societal patriarchy is defined in the model as "male domination at the societal level" (DeKeseredy and Schwartz 2009, p. 35), and it is composed of structure and ideology. Based on the work of Wilson and Daly (1992, p. 85), male proprietariness is defined as "the tendency [of men] to think of women as sexual and reproductive 'property' they can own and exchange."

Exiting a relationship (whether physically or emotionally) is also a component of this model. Exiting a relationship cannot in and of itself be the cause of sexual violence, but it can and does contribute to the likelihood of sexual violence. Additional forces are necessary and this model indicates those forces are "patriarchal male peer support" and "threats to masculinity and patriarchal control." Male peer support is defined as the "multidimensional attachments men form to male peers who themselves sexually assault women, provide resources that perpetuate and legitimate such assault, or both" (DeKeseredy and Schwartz 2009, p. 41). Extant research demonstrates that male peer support is a powerful determinant of sexual violence: A connection with an all-male group tends to support and justify the commission of sexual violence against females. Further, an all-male support group can bolster the idea that a female who exits her relationship damages a male's masculinity. In an effort to restore threatened or damaged masculinity, males engage in sexual violence against their partners.

Masculinities

Messerschmidt's Masculinities (1993) theory is greatly informed by feminist theoretical perspectives and is valuable in understanding sexual violence. Building on earlier theoretical work, Messerschmidt contends that masculinity is not an inherent or fixed characteristic of males, but one that must be continually *accomplished* in every social role. Further, how a particular male accomplishes masculinity is directly based on his position in society (employment, stable family life, race, class, etc.). Thus, males are not uniformly afforded the same opportunities to accomplish masculinity. A wealthy white male may have opportunities unavailable to a poor, black male. His theory also contends that not all males share the same conception of what being male means.

For males who do not have legitimate means of accomplishing masculinity because they are positioned in the powerless underclass, criminal behavior and sexual violence against females is one way to accomplish gender and masculinity. Thus, violence – including sexual violence – is a method for some males to demonstrate their masculinity.

Though males do not share a singular method to demonstrate masculinity, all males do act within the confines of "hegemonic masculinity." Hegemonic masculinity emphasizes masculine practices toward control, independence, emotional restriction, self-reliance, active homophobia, aggressiveness, and the capacity for male perpetrated violence. Further, hegemonic masculinity socializes males into believing that they can accomplish masculinity by ridiculing, dominating, and subordinating all that is female. This includes sexual violence.

While a valuable approach in that it recognizes gender as a product of social action in everyday life, and that there exists a variety of masculinities (and femininities) shaped by structural positioning, it is not without criticism. Miller (2002, p. 437) argues that the perspective can be strengthened with careful attention to four issues: avoiding tautology; challenging gender dualism; accounting for stratification, hierarchy and power; and conceptualizing the complexities of agency and social practice.

Gendered Social Bond/Male Peer Support Theory of University Woman Abuse: New Insights

This particular feminist approach to understanding sexual violence integrates Hirschi's social bond theory (1969), feminist male peer support theory, masculinities, and other critical perspectives to present a theory of conformity used to explain sexual violence against university females. In this work, Godenzi, Schwartz & DeKeseredy (2001) invert Hirschi's social bond theory by asserting that it is precisely the social bonds with conventional institutions and peers that *encourage* sexual violence against females in university settings. This approach, in combination with the male peer support theory, offers an explanation about why men who are associated with patriarchal peer groups commit sexual violence against females.

Important in this explanation is the role that attachment plays. The greater the emotional ties males have with other male peers in a hypermasculine subculture, the greater the likelihood of sexual violence against females. A second critical element identified is commitment which suggests that males will conform to the patriarchal social order based on a strong loyalty. This can manifest itself by regular relations with females regardless of the circumstances involved. Through this activity, one can accomplish masculinity and demonstrate commitment to the patriarchal social order. Next the model addresses inverts Hirschi's proposition about the relationship between involvement and offending. Specifically, the more time spent with patriarchal peers – in activities like drinking and watching pornography – the more likely it is that men will perpetrate sexual violence against females. And finally, Hirschi's theory notes that those with a belief in the legitimacy of patriarchy and masculinity are most likely to follow it. In this case, following it can lead to sexual violence against females.

This integrative theory offers a rationale for why males engage in sexual violence against females. They are engaged in accomplishing masculinity and demonstrating attachment, commitment, loyalty, and involvement in a patriarchal system that both condones and encourages sexual violence. This integrative model offers several benefits to extant theories. First, it includes the role of gender making up for a flaw in social bond theory. Second, it includes the role of social group processes, also neglected in social bond theory. As noted by the authors, each perspective used makes up for weaknesses in the others.

Left Realist Gendered Subcultural Theory

Left realism is an example of critical criminology which has its roots in Marxist feminist theory. It however differs from traditional Marxist thought in that Left Realism addresses the Marxist shortcoming of being gender blind. In this particular perspective, DeKeseredy and Schwartz (2010) consider the role of economic policies, social exclusion, and sexual violence. This theoretical offering on sexual violence focuses on the criminogenic subcultural development and destructive economic policies that have characterized the last several decades, and continues today. Together, these two forces have increasingly made difficult male attempts to accomplish masculinity – at least in legitimate ways.

DeKeseredy and Schwartz argue that in the last several decades, employment opportunities have diminished greatly especially from the working class. This loss of employment opportunities for these males frustrates the accomplishment of masculinity as opportunities to be a breadwinner and demonstrates self-reliance is greatly diminished. A second result of this deindustrialization and loss of employment opportunities is an increase in male connection to peer groups that tout violent, macho, sexist, and patriarchal ideals. The growth of such peer groups is indicative of an increasing hegemonic masculine subculture. Thus, males without legitimate means of accomplishing masculinity come together with other similarly frustrated males to form subcultures. These subcultures then encourage and condone sexual violence against females, as it is one "legitimate" way to salvage their hegemonic masculinity.

The Influence of Feminist Theory in the Context of Sexual Violence

The influence of feminist theories in addressing sexual violence is widespread. Research informed by feminist theories is responsible for changes in the criminal justice system making it more responsive to the victimization of, and domination experienced by, females. Research based on feminist theories has resulted in important and rapid changes in policies and laws designed to address many forms of sexual violence. For example, it is responsible for the reformation of state rape laws including marital rape laws, and the adoption of and further inquiry into mandatory and pro-arrest intimate partner violence laws. Still, it is clear that much remains to be done given the problems that remain in the legal treatment of sexual violence.

Feminist approaches have also increased our understanding of the victim-offender overlap in sexual violence. Consider, for example, that research now routinely considers the influence of prior sexual and physical violence against women and how it influences future offending. Research routinely finds that many females, in escaping from victimization, turn to deviance and offending (e.g., running away) which in turns enhances victimization risk – especially sexual victimization risk. Research repeatedly indicates that adult and juvenile female offenders in the criminal justice system report higher rates of prior physical and sexual victimization than non-offenders. This series of behaviors – referred to as "criminalizing girls' survival" illuminated by feminist perspectives – offers better understanding of the experiences of females, especially as it relates to sexual violence.

Contemporary research also routinely considers the role of sexual inequality, prior sexual victimization, sexism, and patriarchy in sex work by females. A growing body of research indicates that a substantially greater proportion of prostitutes experienced childhood sexual abuse compared to non-prostitutes. Further, attention has recently been given to juvenile sex workers who traditionally have been framed as criminal offenders. In contrast, these juveniles – many of whom are being exploited and dominated by males – are now being seen as sexual violence victims.

Controversies in the Literature

Despite theoretical and research advances, considerable controversy remains over some very basic questions about sexual violence. For example, how much sexual violence is there? How should sexual violence be defined? Should a victim who believes she was not raped be counted as being raped? Does the victim or the surveyor determine who was raped? These simple, yet controversial, questions are best exemplified by a series of studies published by Koss (1988), Gilbert (1993), and others.

Koss (1988) directed a study on college campuses widely referred to as the "*Ms.* Report" given it was funded by *Ms.* Magazine. The research was grounded in the belief that rape is an extreme behavior on a continuum of normal male behavior and utilized a broad definition of rape. To capture the concept of rape, ten questions were used in the survey. For instance, one question asked respondents if they were physically restrained and forced to have sexual intercourse. Another question queried respondents about threats and/or forced sexual intercourse that occurred after they had been given drugs or alcohol by a male.

After interviewing more than 3,000 randomly selected college women in the nation, it was concluded that 15.4 % of females had been raped and an additional 12.1 % experienced an attempted rape. From the same study, Koss (1988) concluded that 27 % of the 15.4 % who had been raped did not consider what happened to them as rape. From this work, the *Ms.* Report reported that "one in four" women had been raped.

This statistic was widely reported and accepted in the media, and it was used to direct policy although it was substantially higher than official statistics and estimates found in previous academic research focused on rape. This discrepancy brought attention from many, including Gilbert who reviewed Koss' research and concluded that the numbers were exaggerated. For instance, he found that including females who answered "yes" to the question "have you had sexual intercourse when you didn't want to because a man gave you alcohol or drugs?" lacked rigor and exaggerated the problem. He contended that by including this question as a measure, Koss' definition of rape included regretted sex. Further, Gilbert was troubled by the fact that the majority of women Koss et al., classified as rape survivors did not believe they had been raped at all. Is it appropriate to reject the judgment of females who do not believe they were victimized? Should the law or a researcher's beliefs override a female's experience?

While some scholars and activists have "sided" with Koss' methodology, others find her estimates exaggerated and the definitions she used overly broad. In general, research generates controversy, but the controversy generated by this issue has been extreme. Researchers and experts have been called traitors, shunned from universities, and in the case of Gilbert, threatened with calls to "Kill Neil Gilbert," and for him to "cut it out, or cut it off." In one instance, the director of the California Women's Law Center stated that they wished that Gilbert would himself be raped.

Open Questions

As is true with any theoretical tradition, many open questions characterize feminist theory especially as it relates to sexual violence. The purpose of this section is to outline some of these open questions.

One open question focuses on the value of feminist theory for understanding criminological phenomenon outside of sexual violence. The connection between feminist theory and sexual violence has been present since at least the 1960s. However, the use of feminist theory continues to remain predominantly in the area of sexual violence. Calls for its widespread use in criminology have been made since at least 1988 when Daly and Chesney-Lind noted that criminology in general – with the exception of feminist treatments of rape and intimate violence – has remained untouched by feminist theory. Why does feminist theory continue to be found primarily in the domain of sexual violence? If feminist theory is a powerful explanation of violence, why is it not more prevalent in understanding other types of violence such as stranger robbery or homicide?

A second set of questions pertain to the preeminent role given patriarchy in some variations of feminist theory. First, does the emphasis on patriarchy mask the important influence of other phenomena? Second, how can one recognize and measure greater and lesser degrees of patriarchy? Third, how can patriarchy explain sexual violence among same-sex partners?

Can feminist theory focused on patriarchy explain same-sex and heterosexual relationship violence equally? And finally, how does patriarchy help explain female perpetrated violence against males?

Third, what is the most effective way to treat offenders? If as feminist theories indicate problems are systemic (e.g., patriarchy) what approach is best for addressing offenders? If the problem is structural, do individual-level treatments help? Can restorative justice be a safe and effective way of dealing with sexual violence? Do the potential benefits of such an approach exist only in theory? How does the role of an apology – a component of restorative justice – work within the context of intimate partner violence when an apology is considered a form of coercion in violent relationships? Some work offers some insight into these questions though additional research is warranted. Research by Daly and Curtis-Fawley (2006) concluded that a majority of victims in their study benefitted from a restorative justice proceeding instead of a court proceeding. While this approach is characterized by some problems (e.g., victim safety, manipulation of the process by offenders, pressure on victims, etc.), victims benefitted in that they felt heard, participated in a communicative and flexible environment in which the offender took responsibility for the offense.

And finally, could the more equal standing and greater opportunities available to females today, compared to decades gone by, be a basis for greater, not lesser, sexual violence? In the past, when it was more widely accepted that females were the domain and property of males, sexual violence was not needed to maintain the dominance over females. Today however, less subordination may give rise to some to demonstrate dominance over women via sexual violence. How do feminist theories address this possibility? Have changes in policy actually minimized what would have been a great increase in sexual violence? In other words, had policy changes offering greater opportunities to females never been enacted, would sexual violence rates be substantially greater?

Related Entries

- ▶ Feminist Criminological Theory
- ▶ Victimization, Gender, and the Criminal Justice System
- ▶ Women of Color and Crime

Recommended Reading and References

Brownmiller S (1975) Against our will: men, women, and rape. Simon & Schuster, New York

Daly K, Chesney-Lind M (1988) Feminism and criminology. Justice Q 5:497–538

Daly K, Curtis-Fawley S (2006) Restorative Justice for victims of sexual assault. In: Heimer K, Kruttschnitt C (eds) Gender and crime: patterns in victimization and offending. New York University Press, New York, pp 230–265

Danner M (1991) Socialist feminism: a brief introduction. In: MacLean B, Milovanovic D (eds) New directions in critical criminology. Collective Press, Vancouver, pp 51–54

DeKeseredy WS (2011) Feminist contributions to understanding woman abuse: myths, controversies, and realities. Aggress Violent Beh 16:297–302

DeKeseredy WS, Schwartz MD (2009) Dangerous exits: escaping abusive relationships in rural America. Rutgers University Press, New Brunswick

DeKeseredy WS, Schwartz MD (2010) Friedman economic policies, social exclusion, and crime: toward a gendered left realist subcultural theory. Crime Law Social Ch 54:159–170

DiCristina B (2006) Durkheim's latent theory of gender and homicide. Brit J Criminol 46:212–233

Gilbert N (1993) Examining the facts: advocacy research overstates the incidence of date and acquaintance rape. In: Gelles R, Loseke D (eds) Current controversies in family violence. Sage, Newbury Park, pp 120–132

Godenzi A, Schwartz MD, DeKeseredy WS (2001) Toward a gendered social bond/male peer support theory of university woman abuse. Crit Criminol 10:1–16

Hirschi T (1969) Causes of delinquency. University of California Press, Berkeley

Kelly L (1988) Surviving sexual violence. University of Minnesota Press, Minneapolis

Koss M (1988) Hidden rape: sexual aggression and victimization in a national sample of students in higher education. In: Burgess AW (ed) Rape and sexual assault, vol 2. Garland Publishing, New York

Messerschmidt JW (1993) Masculinities and crime: critique and reconceptualization of theory. Rowan & Littlefield, Lanham

Miller J (2002) The strengths and limits of ‘doing gender’ for understanding street crime. Theor Criminol 6:433–460

Pollak O (1950) The criminality of women. University of Pennsylvania Press, Philadelphia

Schur EM (1984) Labeling women deviant: gender, stigma, and social control. Temple University Press, Philadelphia

Schwendinger JR, Schwendinger H (1983) Rape and inequality. Sage, Beverly Hills

Wilson M, Daly M (1992) Till death do us part. In: Radford J, Russell DEH (eds) Femicide: the politics of woman killing. Twayne, New York, pp 83–98

Fencing/Receiving Stolen Goods

Mike Sutton
School of Social Sciences, Nottingham Trent University, Nottingham, UK

Overview

The Industrial Revolution, beginning in Britain and gathering momentum in the eighteenth Century, enabled widespread ownership of desirable, mass produced, identical goods. This changed the characteristics of general theft whereby it switched from being motivated predominantly by a desire to take often unique property for personal consumption to stealing to sell standardized goods once more destined for the personal enjoyment of others.

While weight and portability of items is considered by thieves (Felson and Clarke 1998), this most usually happens, at least where prolific thieves are concerned, only if they believe the goods will be saleable once removed (Sutton 1995); at which time, considerations regarding weight and portability, and even danger of removal, will be balanced against prices. Therefore, the issue of demand and supply by theft is important because the most valid predictor of items that most thieves will choose to steal is whether or not they believe they can be sold easily for a good price. For example, the recent meteoric rise in scrap metal theft is fuelled by a global metals shortage caused by the high demand for raw materials that are essential for the expanding industry, cities, and infrastructure of China. High prices motivate thieves to remove heavy lead flashings from the roofs of high buildings or risk electrocution stealing heavy and difficult to remove live copper cable. Globally, scrap metal copper prices have doubled since 2004, followed worldwide by a significant number of electrocution fatalities at substations, railways, oil wells, overhead power lines, industrial buildings, and other places by thieves attempting to steal live copper cable.

Understanding more about the various markets for stolen goods presents a challenge for criminology and the criminal justice system that is relatively overlooked by the conventional tight-focus upon only the thief and the act of theft. Yet, arguably, mankind cannot adequately understand the prevalence and incidence of theft of different goods without understanding how different types of stolen goods markets operate to influence demand and supply, who deals in them and why. Worldwide, societies that actively detect and punish thieves, seek to target harden property, or otherwise increase its capable guardianship all virtually ignore the large number of citizens who purchase stolen goods at bargain process, notwithstanding that all those who buy motivate thieves to supply by theft.

Background to the Problem

Thieves are seldom found living in, or otherwise owning, an Aladdin’s cave of stolen goods. A safe assertion, therefore, is that most prolific thieves seek to raise money by selling whatever they steal. And the type of goods most frequently stolen is determined by the level of demand for them. Prolific thieves are generally very good at gauging this because most sell stolen goods within 30 min of their theft (Sutton 2010).

Most burglars and other thieves steal because they want money, and Clarke (1999) explains that offenders have a hierarchy of goods that they prefer to take. Top of their list is cash, followed by items that can be sold easily for relatively high prices – such as jewelry and desirable high-technology equipment. Stolen goods markets and knowledge of what can be sold in them motivate many thieves. This explains why societies have experienced different crime waves comprising targeted theft of very specific items and commodities.

The prolific and successful thief must complete one or else two objectives while evading detection and apprehension. The first objective is to steal cash or else saleable commodities. If something other than cash is stolen, the second objective is to either sell or else trade the goods stolen. Where goods are stolen, typically, police, prosecutors, and criminologists view this behavior as comprising the two distinct crimes and actions of stealing and then of selling stolen goods. From the thieves' perspective, however, these one or two objectives need to be completed in order to achieve their main aim, which is usually to acquire whatever it is they initially needed or wanted to buy before embarking on the crime.

Criminologists and crime scientists focusing upon the act of theft typically seek to understand the causes of, and find solutions to, the first objective. But the wider aims and objectives of theft from the perspective of the thief include selling, fencing, and receiving stolen goods (the second objective). In effect, the offender is following a crime script that begins before and continues after the act of theft. Seeing this wider picture can help to increase depth of understanding of acquisitive offending and possibly reveal innovative and promising avenues for seeking to tackle it.

Dealers in stolen goods have most probably existed for as long there have been laws against theft and a demand for stolen goods. The fence is a middleman between the thief and the consumer of stolen goods.

Through the act of buying stolen goods as a trusted criminal middleman, the fence allows the thief to avoid the risk of being caught in the act of trying to sell their loot directly to untrusted strangers. Hence, the origin of the word fence is widely believed to stem from the shortening of "defense" during the seventeenth century as a common understanding of the dealer in stolen goods being the thieves' defense from detection.

The role of historical and contemporary fences as protectors of the secret of thieves' identities is perhaps most plainly highlighted by what is known about theft in the time of slavery. For example, Williams' (1963) translation of a letter written by the Archdeacon of Hispaniola to the Council of the Indies in 1542 reveals that trusted fences served as a safe and ready market, thereby protecting numerous slaves from being linked directly to their crimes through selling goods they stole in slower and more risky ways. Money from the fence was then paid by the thief to his "master" in order to buy off his otherwise callously enforced labor. Likewise, slaves who were fences bought their own way out of the same exploitation through trading in stolen goods:

> The Negros are already doing business and trading among themselves to an extent involving great value and cunning, and as a result, big and notable robberies are committed on all the farms in the country... Some steal to pay for the day's work which they have agreed to give their masters... Night and day they rob and steal anything in the country, including gold to be melted. These thefts are concealed with the assistance of two or three hundred Negros called "fences", who go about the city seeking to make profits as I have said...and to pay the daily wage in exchange for each day or month or year, that they are at large and travel about the island. They take away stolen goods for sale and carry and conceal all that they are accustomed to conceal...

As well as providing protection for thieves, fences are conveniently sited to provide information to assist their arrest and prosecution. Although it has long been suspected that disingenuous police officers might allow fences to continue trading in exchange for intelligence about the thieves who supply them, there is no published evidence that this actually happens. Perhaps the only published evidence of such things happening in the past can be found the case of Jonathan Wild. In the eighteenth century, Wild, the most notorious fence in history, presented himself as a public hero, arresting so many thieves that he earned the title Thief-Taker General. Yet all the while, Wild secretly led a gang of thieves, regularly received stolen goods, manipulated victims to offer a reward for their return, and then pretended to track down the goods he already illegally possessed to claim the reward. Unsurprisingly, the question of the degree of guilty mind of the receiver has, for centuries, occupied legal scholars (e.g., Colquhoun 1796; Hall 1952), but astonishingly few contemporary criminologists are concerned with criminal career ethnographies of successful fences (Steffensmeier 1986) or understanding the wider dynamics of the trade (Sutton 1998).

Stolen Goods Market Types: Guilty Minds and Offending Dynamics

The London magistrate Patrick Colquhoun (1796) was concerned with the guilty knowledge of people buying stolen goods at bargain prices. His threefold typology consists of:

1. Criminal receivers (professional fences who deliberately encourage theft)
2. Careless receivers (have a reckless disregard for the origin of the goods)
3. Innocent purchasers (believe good were legitimately purchased by the seller)

A new typology was created 156 years later by the jurisprudentialist Jerome Hall (1952), who set out to emphasize the role of the professional fence in marketing stolen goods with another threefold typology:

1. The professional receiver (fence who deals in stolen goods)
2. The occasional receiver (buys for resale but does so only infrequency)
3. The lay receiver (buyer and consumer of stolen goods)

Even professional fences operate at different levels, as can be seen in the following three-tier fencing level typology first outlined by Lewis (2006):

- *Level-1 fence*: The thief sells to a level-1 fence (often a storeowner such as a pawnbroker or jeweler), who then sells the goods in his store or else sells them to another fence.
- *Level-2* (wholesale) *fence*: This fence buys from a level-1 fence and then often cleans up and/or repackages the goods to make it look as though they came legitimately from the manufacturer. These are very clandestine operations that are perhaps most likely to be found by police officers working back from an investigation of a level-3 fence operation. Those who operate stolen car rings also fall within this subtype.
- *Level-3 fence*: A level-3 fence takes repackaged goods from a level-2 wholesale fence and diverts them to retailers. At times, major retailers find themselves innocently purchasing the very goods that were stolen from them. Level-3 fences have been known to sell perfume, cosmetics, razor blades, and shoplifted designer goods in this way.

Being human constructs, all typologies tell us as much about those who create them as they do about their subject matter. Hall's typology, for example, was purposely designed to emphasize the role of the professional fence in the marketing of stolen goods, because he sought to bring about a change in US law that would see professional fences treated with greater severity. Lewis's three-tier fence model is concerned only with describing the hierarchy and marketing dynamics of fences dealing in goods stolen in organized retail theft.

Seeking to classify the various ways that the most commonly stolen goods are sold, Sutton's original (1998) fivefold typology of handling dynamics, updated to six to incorporate later knowledge of eSelling (Sutton 2010), describes how thieves sell, dealers deal, and the public buys stolen goods. No one of Sutton's six market types is known to be more serious or important than any other in terms of the role it plays in promoting theft by motivating thieves, fences, and the buying public. Moreover, research suggests that thieves and dealers regularly use more than one type (Sutton 1998):

1. *Commercial Fence Supplies*. Stolen goods are sold by thieves to commercial fences operating out of shops such as jewelers, pawnbrokers, and secondhand dealers.
2. *Commercial Sales*. Commercial fences usually pose as legitimate business owners while secretly selling stolen goods for

a profit, either directly to the (innocent) consumer or more rarely to another distributor who thinks the goods can be sold again for additional profit.

3. *Commercially Facilitated Sales* (modified here from eSelling). This market type involves either the thief or a residential or commercial fence knowingly selling stolen goods through classified advertisements in traditional newspapers and magazines, through traditional auctions, or online through classified sales websites such as Craig's List or Preloved.co.uk. They may also sell stolen goods on Internet auction sites such as eBay.
4. *Residential Fence Supplies*. Stolen goods (particularly electrical goods) are sold by thieves to fences, usually at the fence's home. The fence may be drug dealer and may be a prolific dealer in stolen goods or may deal only occasionally.
5. *Network Sales*. Stolen goods are passed on, and each participant adds a little to the price until a consumer is found. This may involve a residential fence or commercial fence selling to other fences. Alternatively, the buyer may be the final consumer or may sell the goods on again through friendship networks.
6. *Hawking*. Thieves, or their friends, sell stolen goods directly to consumers on the street or in places such as bars and pubs or door to door in residential areas (e.g., shoplifters selling cigarettes, toiletries, clothes, or food).

Thieves tend to be flexible and may trade in several different markets depending upon where they are, the time of day, what they have for sale, and how quickly they need to sell it.

Thieves do not simply respond to requests to steal certain types of goods to order, because many *steal to offer* goods to total strangers, which can lead to the thief directly motivating members of the public to become regular customers in *Hawking* markets and tempting previously straight business owners to become fences by way of *commercial fence supplies* (Sutton 1998). Experienced commercial fences, in turn, use their respectable business *fronts* to recruit inexperienced thieves who come in to offer them stolen goods. Commercial fences sometimes mix stolen goods in with their legitimate stock. Somewhat perversely this helps to sell legitimate stock, because some people think they are getting a real bargain if the shop has something of a reputation for selling high-quality goods suspiciously cheaply as opposed to cheaply made legitimate merchandise.

Research suggests that stealing to order is not as common as *stealing to offer* (Sutton 2010). And thieves' knowledge of the "standing demand" for particular types of stolen goods influences the types of goods that are stolen, which can lead to crime waves when thieves target particularly hot products. Past crime waves of this kind have included theft of lead, copper, aluminum, bronze, silver, gold, color television sets, hi-fi equipment, videocassette recorders (VCRs), motor vehicle radio-cassette players, computer memory chips, mountain bikes, laptop computers, digital versatile disks (DVDs), games consoles, mobile phones, and in-car satellite navigation devices.

The Basic Principles of Fencing

According to the US Senate's comprehensive review of fencing operations, the overwhelming majority of fences in the North America operate legitimate businesses (Sutton 2010). To operate successfully and avoid detection, the criminal dealings of the fence must be much less visible than the offenses and offenders that initially supply stolen goods. To achieve this, the fence must coach thieves to avoid detection, conceal his own trading behind a legitimate front, remain willfully ignorant of the provenance of goods bought from other fences, avoid storing goods to avoid detection but know how to safely store them if necessary, be wary of working with police informants, and limit the number of people who know what he is up to. He must never admit to knowingly trading in stolen goods if questioned by detectives, and he must have money for a good lawyer in case of arrest (Steffensmeier and Ulmer 2005).

Thieves and dealers in the UK and USA at least operate a "two- and three-way split" whereby experienced thieves selling to fences ask for between half the wholesale price and a third of the fences' selling price. This tends to vary though depending upon whether or not items are in high demand as fast-moving consumer goods or high end luxuries. Thieves selling stolen cigarettes, for example, are generally paid between 30 % and 40 % of the retail price. Shoplifters selling stolen clothes, meat, and bottles of alcohol such as vodka, whisky, and brandy directly to consumers tend to receive half the retail price. Other stolen but used items, like electrical goods stolen from house burglaries, are usually sold by the thief for a third of the retail value. If a fence sells directly to someone who knows the goods are stolen, then they sell for half the retail price. If the fence is a businessperson selling stolen goods to innocent customers through a shop, then the goods are usually sold for two-thirds of the retail value. Gold jewelry is different, however, in that it is sold by thieves to jewelry shops for the going rate for scrap gold. Presumably that same rule applies for scrap metals of all kinds. Where other stolen goods are concerned, this two- and three-way split on prices appears to be cast in stone, not least because several writers have documented these pricing practices existing for well over 100 years (Quennell 1958; Steffensmeier 1986, Sutton 1998). Perhaps this is because it is simple to understand and operate by those motivated to make quick but regular profits in illicit markets.

The Seller's Dilemma

Devising and testing creative ways to increase the difficulties and dilemmas faced by those dealing in stolen goods might be a useful approach in designing theft prevention and control strategies. The stolen goods seller's dilemma, whether she is the thief or a dealer (the fence), is that to increase her profit, she has to increase her risks of getting caught. The seller can choose to sell only to people that she knows, which reduces her risks of being ripped off or detected but restricts her sales and buying opportunities. The thief can sell to strangers, which allows her access to more potential customers, but also increases her chances of being arrested or robbed. This dilemma applies to both the thief and the dealer, but the business-owning commercial fence must also simultaneously nurture the confidence of thieves with whom they deal while projecting a legitimate trading image.

These conflicting demands of access and security determine to a large extent the structure of local stolen goods markets. The small size of most fencing operations, uniformity in pricing, and limited profit making from theft should come as no surprise since competitive options are limited by virtue of there being only so many ways of doing things efficiently and effectively in any kind of illicit market. This is because advertising, marketing, warehousing, transportation, and expansion options are necessarily avoided, or else extremely limited, in order to avoid detection and regulation (see Reuter 1985).

Prevalence

Few estimates of the amount of trading in stolen goods markets exist, but an exercise conducted by the British Government to inform the UK National Accounts (1997) claimed that in 1995, thieves selling stolen goods within Britain cleared between £900 million and £1,680 million (net) and that fences cleared between £450 million and £875 million (net) through selling stolen property.

The 1994 British Crime Survey (Sutton 1998) found that over the previous 5-year period, 11 % of the population of England and Wales admitted buying stolen goods, which they knew or believed to be stolen; 70 % thought that at least some of their neighbors had stolen goods in their homes; and 21 % thought a lot of them had the same. Looking back at just the previous year, rather than the last five, the British Offending, Crime and Justice Survey (Sutton et al. 2008) found that 7 % of adults in England and Wales admitted buying stolen goods while 2.7 % admitted selling them. Furthermore, almost half of males aged 16–24 had been offered or else bought stolen goods. Comparing males with females, this research suggests that it is possible that more than twice as many males buy. In the poorest neighborhoods, 40 % of males bought stolen goods compared to 17 % of females. Even in the most affluent neighborhoods, 7 % of people bought stolen goods. Incredible as they are, these figures could be an underestimate if some respondents were reluctant to admit buying stolen goods and others forgot that they did so.

The importance of the stolen goods market problem is further highlighted by Graham and Bowling (1995) who found that handling stolen goods was the most prevalent crime admitted by their respondents, with 49 % of offenders admitting to having done it in the past year.

Responses to Stolen Goods Markets

A consistent theme in the social sciences and the multidisciplinary areas of criminology and criminal justice is the unintended, sometimes ironic, consequences of purposive action. This theme, which is central to Merton's (1949) self-fulfilling prophecy, is reflected in the rationale behind the market reduction approach to theft (Sutton 1998). Specifically, those who buy stolen goods unintentionally support a market supplied by theft for their own future victimization.

Since the existence of "safe" and ready stolen goods markets is a difficult to disentangle, downstream consequence of theft, as well as one causal factor behind the motivation for theft, knowing more about stolen goods markets, in order to seek to reduce them and make it more risky to deal in them with situational crime reduction approaches, might provide one potentially promising avenue. The market reduction approach (MRA) (Sutton 1998; Sutton et al. 2001) is designed to do this. Although it is recommended as promising practice on official websites supported by government agencies in the UK, USA, Australia, and New Zealand, the MRA remains as yet unproven (Hale et al. 2004) avenue for reducing acquisitive crimes.

A comprehensive review of promising policing and multi-agency partnership responses to stolen goods markets, including those known to have limited effectiveness, was commissioned by the US Department of Justice (Sutton 2010). The review examines various crackdowns on stolen goods markets that have been tried out in the UK and USA with various degrees of success. One important conclusion reached is that even the most apparently successful schemes prove difficult to maintain over time with the traditional focus on the theft act taking priority in the allocation of scarce resources (Walsh 1976). Moreover, limited research in this area reveals that what works and does not work in tackling stolen goods markets is complex and at times counterintuitive. Research findings – however limited – are particularly valuable, therefore, in helping police services avoid repeating past mistakes. Clandestine police storefront antifencing sting operations, for example, can have the unintended effect of generating theft in the surrounding area to meet the demand they have unintentionally created (Langworthy and Lebeau 1992). Furthermore, despite being a favorite crime reduction activity in many police services, property marking has never been proven to reduce theft largely because thieves steal both "invisibly" and clearly marked property and fences and citizens will buy it (Sutton 2010). Therefore, the oftentimes bold assertions made by commercial companies for the success of their expensive property-marking products have never been confirmed by independent academic research (Knutsson 1984; Sutton 1998; Harris et al. 2003; Hale et al. 2004).

Explaining the Relative Importance of Stolen Goods Markets: Why the Market, Not Opportunity, Makes the Thief

Property thieves, particularly prolific ones, are generally perceived to be "bent" offenders whose predation upon "straight society" can be explained by their relative poverty, subculture, wider cultural influences, poor socialization, substance addiction, or individual pathology acting alone as significant causes or else together as a combination of forces that interacts with so-called opportunities for theft. Here, current criminological understanding of the crime act has been shaped by the current criminological notion of “opportunity” that is classically defined in Crime Opportunity Theory as what happens when a relatively more capable and sufficiently motivated “likely” offender succeeds against a target or victim – thereby proving that they were capable offenders against relatively incapable or absent guardianship. However, this crime as “opportunity” explanation does not at the time of writing provide discoverable and measurable quantifiable values that would enable criminologists to predict and test individual or general victim or target vulnerability relative to actual or potential capable offender motivation or guardianship abilities (Clarke 1984). This same limitation applies in the area of repeat victimization and within various types of high crime environment or in crime hot spots. It is a truism therefore that capable and suitably motivated offender components of current notions of “crime opportunity” cannot be discovered and objectively measured in nature or society before a successful crime happens – only afterward. This limitation means that “crime opportunity,” as the “almost always, elements of a criminal act” described in the Routine Activity Theory (RAT) crime triangle by Felson and Boba (2010, pp. 28–40), and in the description of crime as opportunity that is classically outlined by Mayhew et al. (1976), is the essential data of a successfully completed crime in commission. In other words, it is always knowable – but only ever after the event – that the crime happened because the offender could successfully commit it or else failed because he could not.

Logically, therefore, “crime opportunity” cannot be a cause of crime because Crime Opportunity Theory merely describes whether or not offenders were in fact more capable than the crime target’s guardianship. It follows, therefore, that until a crime is successfully completed or fails in the attempt, the current notion of “crime opportunity” cannot be known by offenders to exist in advance of the crime being completed or failing. This is because no potential thief could know for sure that they would be more capable than any guardian or that guardianship would remain absent. After all, if that was possible, there would be no reason for so many failed criminal attempts.

The RAT “crime opportunity” incorporates the commonsense premise that before stealing, most offenders rationally weigh up what they perceive to be risks versus rewards. However, this explanation of perceptions in the so-called crime opportunity event does not include human guardian perceptions of their own relative capabilities. Offender and guardian perceptions aside, if the current notion of crime opportunity, which combines RAT with the known successful criminal outcome described in the classic crime triangle, could somehow exist in advance of the crime actually being accomplished, then it would logically follow that every successfully completed crime and every offense of attempt would be the cause of itself, which is clearly impossible according to the scientific laws that govern the universe above the subatomic level. Therefore, the widely cited claim that “opportunity makes the thief” (Felson and Clarke 1998) is undoubtedly logically flawed (see also Sutton 2012).

Consequently, this notion of opportunity is neither a rational framework nor model for theory building. Nor is it a stand-alone theory of crime causation (e.g., see Felson and Clarke 1998, p. 9; Tilley and Laycock 2002; Felson and Boba 2010, p. 47) nor any kind of *measure* of differential vulnerability (see Clarke 1984), because it does not quantify levels of vulnerability. What is more, this notion of “crime opportunity” does not even fit common understandings of opportunity because it does not describe any kind of realistic pre-crime situation formed by a juncture of circumstances favorable to crime. Logically, therefore, there is no opportunity in Crime Opportunity Theory.

The current classic RAT and situational crime prevention notion of “crime opportunity” (perhaps ratortunity is a better word for it) is an elegantly precise, perhaps perfect, post hoc description of any successfully completed criminal act, which makes it a veracious, and therefore important, description of what has happened. But descriptions, no matter how elegantly described, cannot explain the reasons for the data. It follows, therefore, that precise descriptions of the components present at every crime act cannot explain the cause of the crime. This is true for all descriptions and the data they describe (Shermer 1991). Just as a fossil embedded in the geological fossil record, no matter how precisely and elegantly described, cannot explain itself without a separate theory – such as Darwin's theory of evolution.

Criminology requires a crime theory to explain why, in all successfully completed crimes, offenders are sufficiently motivated to prevail against protective measures. Simply saying, for example, that successful offenders were sufficiently motivated to overcome levels of guardianship, perhaps even ones that had deterred them in the past, because they rationally reevaluated the risks and rewards (Felson and Clarke 1998) does not explain at what point and why the rewards and risks switched to make a once adequately protected target become sufficiently vulnerable. In order to do that, criminologists need to look for more promising explanations that are separate from the data and descriptions of criminal acts they seek to explain. One potentially promising avenue here is to focus upon precisely how variations in demand for particular stolen goods differentially influence the motivation of thieves and their perceptions of risks versus rewards. This is why the strangely neglected area of stolen goods markets deserves more attention.

In the case of theft, changes in technologies, cultures, consumption patterns, and the economy of a nation state can sometimes act as a spur for new crime motivating markets, leading to increased levels of theft of particular types of property and changes in ways and methods of offending (e.g., Mann and Sutton 1998). Future research in this area will most certainly build upon existing knowledge and seek to understand more about the role that stolen goods markets play in motivating people to begin and continue stealing. One thing waiting to be discovered is information about exactly how and why an increase in demand for particular things can change them from "warm" to "hot products" (Clarke 1999) and hence increase both licit and illicit prices (Sutton 1995). If future research could find and then attach a quantifiable value to the "sufficient motivation switching point" for offending (if indeed there is such a thing) as a sufficient condition for theft, then criminologists would be in a better position to predict acquisitive crime waves. Perhaps one day criminologists will be able to accurately predict the likelihood of the next otherwise unexpected crime wave before it becomes a crime harvest. Forewarned with such knowledge, it would be possible to take preventative action, rather than merely explaining why the crime wave happened and seeking to prevent individual repeat occurrences. Understanding more about the role of stolen goods markets in theft, therefore, offers further promising incremental advances and perhaps, potentially, a monumental breakthrough in criminological knowledge, crime reduction policy making, and policing.

Conclusions and the Way Forward

Little is known about the factors that influence demand for stolen goods, what impact ready markets have on potential and persistent property thieves, and what might be the most promising ways to tackle effectively the crime of knowingly buying stolen goods. One thing is certain though: if more goods are stolen from people on the street or from their houses and cars, then they are normally purchased by other people to enjoy on the street or in their own houses and cars.

Surprisingly little research has been conducted into who buys stolen goods and why. Accordingly, compared with other areas of criminology, such as ethnographic and subcultural analysis of various different types of theft and thieves, the subject of stolen goods markets is a weirdly neglected area. Despite the fact that fences and consumers create much of the demand for stolen goods that is met through supply by theft, policing and crime reduction initiatives remain for the most part heavily focused upon thieves alone. Given the pervasive nature of stolen goods markets and the fact that the stolen goods trade is, when compared with acts of theft, afforded far less resources and attention, a most telling question is whether it is evenhanded or particularly efficient criminal justice policy or policing practice to focus so much attention on property thieves, rather than those who buy from them.

Knowing what research reveals about adults who motivate young thieves to steal by fencing or otherwise criminally receiving stolen goods and considering the number of thieves occupying prison systems throughout the world reveals a telling question, namely, *why do so few receivers of stolen goods share prison time with their regularly incarcerated suppliers*? The answer lies partly in the fact that gathering sufficient evidence to prosecute fences is difficult because they conceal stolen goods trading behind legitimate business fronts. Professional fences are expert at this and can remain undetected for years. In addition, members of the general public who buy directly from thieves for their own consumption do not do so as prolifically as individual thieves tend to steal. Consequently, their risk of detection is lower. Furthermore, public tolerance toward those who deal out of their cars and houses, often using networks of contacts in the community, is high because these dealers are seen by their customers as providing a kind of community service by way of essential or expensive luxury items at bargain prices.

If fences and the general public who buy stolen goods are responsible for creating markets for everyone's potential victimization, then finding effective ways to reduce such markets appears to be a logical and compelling way to reduce theft. Detecting those engaged in handling stolen goods and applying legal sanctions against them ensures that thieves and handlers have less chance of profiting from the misery of victims of burglary and other thefts, which is arguably an important criminal justice end in its own right and perhaps one that criminologists should not lose sight of in pursuit of measurable theft reduction. That said, it is not possible to predict accurately how different populations might respond in the event they were significantly deprived of stolen bargains.

Since theft remains a problem to be solved, criminologists will continue to make progress in the area. The logical impossibility of "crime opportunity" being a cause of crime brings us to a nexus where the next fundamental breakthrough in understanding the causes of theft may possibly be a new hypothesis proposing an explanation for how "demand" for hot products interacts with human actors to cause a "switching point" in offender motivation with the effect that what was previously capable guardianship of valuable products becomes inadequate when their trading value increases to a certain level, which is exactly what happened when electricity first ceased to be a capable guardian of live copper cable when global copper prices rose significantly in 2004 and thereafter. If such a hypothesis is formulated, criminologists and economists will do their best to disprove it through a process of prediction and observable outcome. If disconfirming evidence is not forthcoming, then the issue of fencing and receiving stolen goods may no longer remain so strangely ignored because it might enable criminologists to better forecast crime waves. The theft reduction potential of adopting such a market reduction approach is huge.

Related Entries

- ▶ Crimes Against Animal Life
- ▶ Cultural Criminology
- ▶ Establishing Causes of Offending in Longitudinal and Experimental Studies
- ▶ History of Criminological Theories: Causes of Crime

- ▶ Innovation and Crime Prevention
- ▶ Modus Operandi of Sex Offenders
- ▶ Rational Choice, Deterrence, and Crime: Sociological Contributions
- ▶ Residential Burglary
- ▶ Routine Activities Approach
- ▶ Situational Crime Prevention
- ▶ Theft at Construction Sites
- ▶ Theories for Situational and Environmental Crime Prevention

Recommended Reading and References

Clarke RV (1984) Opportunity-based crime rates: the difficulties for further refinement. Br J Criminol 24(1):74–83

Clarke RV (1999) Hot Products: understanding, anticipating and reducing demand for stolen goods. Police Research Series Paper 112. Policing and Reducing Crime Unit. Research Development and Statistics Directorate, Home Office, London

Colquhoun P (1796) A treatise on the police of the metropolis; containing a detail of the various crimes and misdemeanours by which public and private security are, at present, injured and endangered: and suggesting remedies for their prevention, 3rd edn. C. Dilly, London

Felson M, Boba R (2010) Crime and everyday life. Sage, Thousand Oakes

Felson M, Clarke RV (1998) Opportunity makes the thief. Paper No. 98. Police Research Series. Home Office, London

Graham J, Bowling B (1995) Young people and crime. Home Office Research Study No. 145. Home Office, London

Hale C, Harris C, Uglow S, Gilling L, Netten A (2004) Targeting the markets for stolen goods: two targeted policing initiative projects. Home Office Development and Practice Report 17. Home Office, London

Hall J (1952) Theft, law and society, 2nd edn. Bobbs-Merrill, Indianapolis

Harris C, Hale C, Uglow S (2003) Theory into practice: implementing a market reduction approach to property crime. In: Bullock K, Tilley N (eds) Crime reduction and problem-oriented policing. Willan Press, Cullompton

Knutsson J (1984) Operation identification: a way to prevent burglaries? Report No. 14. National Council for Crime Prevention Sweden, Research Division, Stockholm

Langworthy R, Lebeau I (1992) The spatial evolution of sting clientele. J Crim Just 20(2):135–145

Lewis L (2006) Organized retail crime: retail's no. 1 security issue. California Grocer, pp 2–11

Mann D, Sutton M (1998) >>NetCrime: more change in the organisation of thieving. Br J Criminol 38(2):201–219

Mayhew P, Clarke RV', Sturman A', Hough JM (1976) Crime as opportunity. Her Majesty's Stationary Office, London

McCormack M (1995) The growing business of chip theft. Comput Fraud Secur Bull 1995(10):19–20

Merton R (1949) Social theory and social structure. Free Press, New York

Quennell R (1958) London's underworld. Spring Books, London

Reuter P (1985) The organization of illegal markets: an economic analysis. National Institute of Justice. US Department of Justice, Washington, DC

Shermer MB (1991) Science defended, science defined: the Louisianan creationism case. Sci Technol Hum Val 16(4):517–539, http://www.jstor.org/stable/689806

Steffensmeier DJ (1986) The fence: in the shadow of two worlds. Rowman and Littlefield, New Jersey

Steffensmeier D, Ulmer J (2005) Confessions of a dying thief: understanding criminal careers and illegal enterprise. Transactions Publishers, New Brunswick

Sutton M (1995) Supply by theft: does the market for second-hand goods play a role in keeping crime figures high? Br J Criminol 38(3):400–416

Sutton M (1998) Handling stolen goods and theft: a market reduction approach. Home Office Research Study 178. Home Office, London

Sutton M (2010) Stolen goods markets. Problem oriented policing guide No. 57. Department of Justice COPS Program, USA. http://www.popcenter.org/problems/stolen_goods/

Sutton M (2012) Opportunity does not make the thief: busting the myth that opportunity is a cause of crime. Best thinking. http://www.bestthinking.com/articles/science/social_sciences/sociology/opportunity-does-not-make-the-thief-busting-the-myth-that-opportunity-is-a-cause-of-crime

Sutton M, Schneider JL, Hetherington (2001) Tackling theft with the market reduction *approach*. Home Office Crime Reduction Research Series Paper 8. http://www.popcenter.org/problems/bicycle_theft/PDFs/Sutton_etal_2001.pdf

Sutton M, Hodgkinson S, Levi M (2008) Handling stolen goods: findings from the 2003 offending crime and justice survey. Internet J Criminol. Primary research paper. http://www.internetjournalofcriminology.com/Sutton_Stolen_Goods.pdf

Tilley N, Laycock G (2002) Working out what to do: Evidence based crime reduction. Crime Reduction Series Paper 11. London Home Office. https://www.ncjrs.gov/pdffiles1/nij/grants/193161.pdf

UK National Accounts (1997) Unpublished Office of National Statistics document. Cited in Sutton 1998. Handling stolen goods and theft: a market reduction approach. Home Office Research Study 178. Home Office, London, p 1

Walsh M (1976) Strategies for combating the criminal receiver of stolen goods. Organised crime anti-fencing manual. Office of Regional Operations, Law.

Enforcement Assistance Administration. US Department of Justice, Washington, DC
Williams E (1963) No 150. The Negro dancer in Hispaniola: a warning. Alvaro de Castro, Archdeacon of Hispaniola, to Council of the Indies. March 26, 1542. In: Documents of West Indian History 1492–1655: From the Spanish Discovery to the British Conquest of Jamaica. PNM Publishing, Port of Spain, p 156

Field Trials

► Multisite Trials in Criminal Justice Settings

Fines in Europe

Tapio Lappi-Seppälä
National Research Institute of Legal Policy,
Helsinki, Finland

Overview

Fine is the most used criminal sanction – both in history and in present. It is also one of the most neglected topics in textbooks and criminal policy analyses. The use of fines and other financial penalties has expanded during the last decades, as a result of several simultaneous changes. The growth of mass infractions in traffic and elsewhere has led to a differentiation of criminal fines into court-fines and summary fines (imposed either by the police or the prosecutor). Also the system of administrative monetary penalties has differentiated into fixed penalties for mass infractions and substantive penalties imposed for large-scale illegal economic activities and corporate fines.

Fines have several benefits which increase their potential as penal sanctions, also for middle rank offenses. This has been acknowledged especially by those jurisdictions that have adopted a day-fine (unit-fine) system. Under this system, the gravity of the offense is reflected in the number of imposed fine-units, while the monetary value of each unit is reflected by offenders' financial means. Thus, two offenders, sentenced for similar offenses, receive fines that are experienced as equally severe, despite the differences in wealth and income. The available (albeit fairly meager) evidence suggests that fines are generally followed by somewhat lower reoffending rates compared to more substantive sanctions.

Not all fines are paid voluntarily, and a backup system for fines is required. The benefits of monetary penalties are lost, if unpaid fines need to be converted to imprisonment. Unfortunately, many jurisdictions have resorted to default imprisonment as a backup sanction. However, there are signs of a wider adoption of new community alternatives to serve this purpose.

History

In tribal societies, injuries were sanctioned either by vengeance or by settlement between the injurer and the injured. The Barbarian Codes (from the fifth century onwards) fixed the amount which can buy off vengeance. The victim (his/her relatives) had the option either to accept the offer or to resort on retribution. These sums cannot be defined as punishment in the modern sense, as there was no third authority to compel the parties to agree with the proposed compensation. It was only through the formation of state and central power that made it possible to establish a public criminal justice system where the ruler had the powers to dictate the sanctions for offenses. In Europe, this period starts around the middle of twelfth century (the establishment of "King's Peace"). This was also the birth of fines as a criminal sanction.

Throughout the history, fines have retained a central position as criminal punishment. Fine was the principal punishment for most offenses in the Nordic codes, the earliest dating from mid-thirteenth century. The amount of fines was also fixed in detail according to the consequences of the acts, and in some cases also the mental state of the offender.

Still, the establishment of justice from above was a slow process. Private vengeance continued to be practiced throughout the later Middle Ages. In addition, fines, even after being authorized and confirmed by the state, served not only punitive but also private compensatory functions. This was also reflected on the way fines were divided in the Nordic codes into three parts: one third to the victim or his/her relatives, another to the King (central power), and one for the Church. Subsequently, punitive and compensatory elements were differentiated from each other, and fines became a matter of public criminal justice belonging to the central power alone. Alongside fines there grew a civil compensation to cover the damages and the losses of the victim.

European criminal codes grew gradually more punitive during the later Middle Ages. As a result, fines lost space to corporal punishments, especially in Central Europe and in England, where the administration of criminal justice came close to state-run terror in the sixteenth and seventeenth century. The eighteenth century witnessed the social and political revolutions across Europe, instigated and inspired by the writings of the enlightenment philosophers. Much of that criticism was a criticism against the brutality and severity of criminal justice of the ancient regime. The use of capital and corporal punishments was restricted and replaced by imprisonment, which came to form the backbone of the European criminal justice systems of the nineteenth century. But along with imprisonment, fines also retained a principal role as a basic sanction for minor and middle rank offenses, especially in the Nordic countries. Practical advantages of pecuniary penalties were recognized also by reformers like Montesquieu, who was among the first to recognize the need to grade the fines according to the financial situation of the offender.

> Our fathers the Germans admitted almost none but pecuniary penalties. These men, who were both warriors and free, considered that their blood should be spilled only when they were armed. The Japanese, by contrast, reject these sort of penalties on the pretext that rich people would escape punishment, But do not rich people fear for the loss of their goods? Cannot pecuniary penalties be proportionate to fortunes? And, finally, cannot infamy be joined with these penalties?
>
> Montesquieu, The Spirit of the Laws. Book VI, Chapter XVIII.

The practical use of fines was conditioned also by general economic conditions. According to Rusche and Kirchheimer (1939, p. 167), the poverty of lower classes remained as a central obstacle for the expansion of fines till the middle of the nineteenth century, whereafter their use has expanded, much caused by the decline of unemployment and the general rising of living standards. For example, in Finland and Sweden, over 90 % of cases dealt by the courts in the end of nineteenth century were punished by fines. Though often overlooked by historians and legal textbooks, fines remained the most used criminal sanction also during the twentieth century in several common offenses (see Table 1).

Fines in Europe, Table 1 The use of fines in 1930s (percent of all punishments)

	France 1933	Belgium 1930	Sweden 1934	Germany 1933
Assault (intentional)	68.9	79.3	97.9	62.2
Damage to property	63.1	79.3	97.9	62.2
Larceny	15.5	24.4	25.1	29.3
Fraud	10.5	10.7	46.9	34.2
Adultery	99.5	83.0	70.9	58.1

Source: Rusche and Kirchheimer 1939

The System of Financial Sanctions

Fines can be ordered by the courts or through simplified process by the prosecutor or the police. The expansion of mass-criminality in the form of new traffic offenses and the growth of petty property offenses (resulting from the change of opportunity structures) has led to an ever-widening adoption of summary proceedings. Today the majority of fines in most jurisdictions are imposed either by the police or the prosecutors.

Alongside "penal fines," there are other kinds of financial sanctions with punitive and preventive functions, and which are sometimes hard to distinguish from penal fines. Continental legal theory makes a clear distinction between criminal and administrative sanctions. This distinction has also far-reaching consequences for the imposition and enforcement of sanctions. For example, administrative matters are dealt with by administrative courts and criminal matters in criminal courts. The use of criminal punishments is tied by much stricter legal safeguards and procedural guarantees, including requirements of guilt, presumption of innocence, and the principle of proportionality.

The growth of administrative sanctions alongside is based on two distinct reasons. Criminal law with all its legal safeguards has proven to be too expensive and heavy measure to deal with bagatelle-offenses. For reasons of procedural expediency, many countries have developed own administrative sanction systems to deal with traffic-offenses, minor infractions against public disorder, etc. The German system of "Ordnungswidrigkeit" has served as a model for several jurisdictions in the continental Europe.

On the other hand, criminal law has proven to be ineffective and impractical when it comes to dealing with large-scale economic activities where the magnitude of possible gains from illegal activities may be counted in millions. The establishment of corporate liability and corporate fine to control large-scale illegal economic activities may have been arranged either inside criminal punishment (with the requirement of criminal guilt) or outside criminal law (as an administrative sanction based on objective responsibility).

In addition to fines based on national corporate liability, there are an increasing number of economic sanctions based on the need to protect supranational market interests (EU fines). The EU Commission has set up guidelines on the method of setting fines imposed on firms which infringe European Union (EU) rules, prohibiting cartels and other restrictive business practices and abuses of dominant position (article 23(2)(a) of Regulation (EC) No 1/2003). The basic amount is calculated as a percentage of the value of the sales connected with the infringement, multiplied by the number of years the infringement has been taking place. Additional factors in setting the fine include the gravity of the infringement and possible previous infringements. The maximum fine for each firm remains 10 % of its total turnover in the preceding business year (Regulation EC No 1/2003, http://europa.eu/legislation_summaries/competition/firms/l26118_en.htm). While traditional penal fines are counted in hundreds or occasionally even in thousands of Euros, EU-corporate fines are easily counted in millions.

In all, the present system of financial sanctions may be divided into three main groups (with considerable jurisdictional variations). (1) Fines defined as criminal punishments and imposed for natural persons, either by the courts or by other criminal justice agencies (prosecutors or the police). (2) Administrative penalties imposed for natural persons for mass infractions (parking tickets, etc.). (3) Substantial financial penalties/fees imposed for illegal economic activities for natural persons or for companies (corporate fine), either under national or supranational (EU fines) rules. The following part of the entry discusses mainly about penal fines imposed for natural persons (group 1).

The Benefits of Fines

The benefits of fines over other penal sanctions are generally known (see, for example, Gordon and Glaser 1991). Fines are generally less costly to administer and can provide revenue that, in some cases, exceeds their administrative costs. Secondly, monies paid to the court can be redirected to the victim of an offense in order to pay financial restitution for any physical or psychological harm incurred. Fines are a flexible penalty that can be adjusted to reflect both the severity of the crime and the offender's financial circumstances (see below). Finally, the use of financial penalties avoids many of the debilitating social costs that are attached to incarceration (including the loss of job and leaving dependents with reduced income and labeling and making it more difficult for the offender to successfully return to the community). Fines permit the offender to remain in the community and in employment, and, in doing so, reduce the need for social support. Much of these benefits are dependent on the way the setting and enforcement of fines is arranged.

The Day-Fine System

Equal treatment of offenders requires that penalties for the same offense should be equally severe for different offenders. The same sum of money means different things for the rich and the poor, and it is an elementary requirement of social justice that people should be treated in the same manner despite the differences in wealth and economic resources.

Most fining systems have tried to meet the demands of equal impact by developing counting rules for balancing the differences in economic wealth. The Scandinavian day-fine system has proven to be the most successful in this respect. The day-fine system was adopted first in Finland in 1921, with Sweden and Denmark to follow in 1931 and 1939.

The origins of the system go back to the discussions in the International Association of Criminalists and their Nordic counterparts in the shift of the century, while the term day-fines was coined by a Swedish law professor Johan C.W. Thyrén. The introduction of the day-fine system was motivated by three main reasons. Firstly, it provided a mechanism to ensure equal impact of sentence severity for the rich and the poor. Secondly, since the amount of fines was tied to the level of income, fines were more or less untouched by the changes in the value of money. And since this made it possible to maintain the level of fines high enough, fines could serve more effectively as alternatives for the much discredited short prison sentences. The Scandinavian day-fine system has served as an influential model for a number of European jurisdictions. During the last decades, countries such as Germany (1975), Austria (1975), France (1983), and Switzerland (2007) have adopted the system with success.

The main objective of the day-fine system is to ensure "equal severity" of the fine for offenders of different income and wealth. In this system, the number of day-fines is determined on the basis of the seriousness of the offense while the amount of a day-fine depends on the financial situation of the offender. The total sum of a fine is reached by multiplying the number of day-fines with their monetary value. The technical counting rules vary within each country, but the basic principle remains the same. It is also very much a matter of domestic penal (and social) policy whether income is counted before (brutto) or after (netto) taxes, and whether there exists a definite ceiling for the value of a single day-fine. In Finland, for example, the amount of the day-fine equals roughly half of the offender's daily income after taxes (and without any upper limit; in Austria, the limit is 5,000 Euros and in Germany 30,000 Euros). The exact amount results from a rather complicated calculation. However, the officials (police, the prosecutor, and the courts) have in their hand a handbook which makes it easy to count the amount of day-fines. Also the number of day-fines varies in different systems: in Finland from 1 to 120, but in Austria, Germany, and Switzerland from 1 to 360.

> *An example from Finland*: The typical number of day-fines for drunken driving with BAC of 1,0 o/oo would be around 40 df. The monetary value of one day-fine for a person who earns 1,500 euros/months would be 20 euros. For someone with a monthly income of 6,000 euros, the amount of one day-fine would be 95. Thus the total fine for the same offense would be for the former person 800 euros and for the latter 3,800 euros.

Not all countries have followed this line. Most common law countries, including the USA, implement tariff fines, which leave much less – if any – room for the adjustment of fines according to the financial means of the offender. England and Wales adapted the system of "unit-fines" based on the day-fine principle in 1991. However, after a heavy media-driven criticism, the government repelled the system in 1993. As noted by a leading legal commentator:

> [T]o sweep away the whole edifice and to return to a system that promises, on the Governments own view, less fairness, fewer fines, and greater difficulty to enforcing them, . . . is eloquent testimony to the power of short-term political gain over sound penal policy. (Ashworth and Gibson 1994, Criminal Law Review 1994, 101 ff)

The failure of the unit-fines system in England and Wales in 1991 was basically a political one. It can be attributed mainly to the UK government's inability to defend a sound system against ill-founded public pressure and misplaced criticism. As reported in surveys post the reform, courts that followed the unit-fines process were more consistent in their fining practice and they graduated the fines more proportionally according to the level of income. In short, their fine setting practice met the requirements of both equality and proportionality better, compared to courts that followed the old system (see Charman et al. 1996).

Despite the defeat in the UK, there have been ongoing efforts to expand the day-fine system in other common law countries. In the USA, this system has been renamed as "structured fines." The first structured fine project demonstration in the United States was designed and operated by the Vera Institute of Justice in Staten Island, New York, between 1987 and 1989. Soon after, several other US jurisdictions had started to implement similar programs. But as it was also noted, these programs required well thought-out policy formulation and program planning, as well as a strong collection system. As the situation stands to days, a growing number of US jurisdictions are experimenting with structured fines (see Bureau of Justice Assistance 1996: How To Use Structured Fines (Day Fines) as an Intermediate Sanction, https://www.ncjrs.gov/pdffiles/156242.pdf).

Institutional Requirements for Effective Enforcement

Effective enforcement of fines requires centralized and well-functioning enforcement organization. Fines need to be collected, which may sometimes need several consecutive acts by the state authorities and/or local municipalities. The authorities need to contact the debtor and, in case of nonpayment, enforce the payment, for example, by drafting a payment plan which allows payment by installments. The credibility of fining systems depends much on the effectiveness of this organization. Low payment rate undermines the credibility of the whole system. It reduces the public confidence of fining and it diminishes the willingness of the courts to use fines as sentencing options. The Nordic countries report fairly high collection rates (for example Sveri 2005, 376 reports that 90–95 % fines are paid in Sweden within 5 years). However, exact comparisons between different systems are complicated due to differences in counting methods.

As regards the day-fines system, key requirement for its effective implementation is that those responsible in setting the fines (whether the police, prosecutor, or the courts) have reliable information about the offenders' income and financial situation. This can be ensured in several ways. In most Scandinavian countries, the criminal justice authorities (including the police) have access to previous year's taxation files. More specifically, the police can receive on request last year total annual income via protected SMS data from the taxation officials. Should that option not be available, other sources for correct income data may be needed. In 1999, Finland criminalized giving false information of one's income for the police in order to reduce the amount of fines. (However, after the establishment of the SMS link, the criminalization of "fine deception" became obsolete, with only two or three convictions per year).

Default Penalties

The credibility of fines as criminal sanction requires a backup system, which will be enforced in the case of nonpayment. The standard default sanction for unpaid fines has traditionally been imprisonment. In the course of history, default imprisonment has – from time to time – become one of the major causes for locking up people. In the 1930s, half of the males and two thirds of females were in English prisons due to unpaid fines (Rusche and Kirchheimer 1939, 169). In Finland, during the last prohibition years in the late 1920s, four out of five admitted prisoners were taken in custody because of unpaid fines.

If fines are converted to custodial sanctions, the original benefits of monetary penalties are lost. One may also end up using prison for trivial offenses, and for offenders whose major fault is that they are too poor to be able to pay their fines.

Fines in Europe, Table 2 The use of default imprisonment for unpaid fines in Europe

	Fine defaulters (1 September 2009)			
	Abs	% of prisoners	/pop	/10,000 fines
Denmark	0	0.0	0.0	0.0
Sweden	0	0.0	0.0	0.0
France	6	0.0	0.0	0.2
Belgium	1	0.0	0.0	0.3
Scotland	4	0.0	0.1	0.7
England and Wales	113	0.1	0.2	0.9
Croatia	32	0.7	0.7	10.5
Finland	90	2.5	1.7	12.0
Latvia	28	0.4	1.2	28.5
Hungary	403	2.4	4.0	36.1
Netherlands	561	4.8	3.4	46.9
Germany	4197	5.7	5.1	72.0
Poland	3302	3.9	8.6	106.6

Source: Compiled from European Sourcebook 2010

The discriminatory nature of default imprisonment has been acknowledged for long, and most countries have made efforts to restrict the use of default imprisonment, however, with different results. Some jurisdictions have adopted community alternatives as default penalties (such as supervised attendance orders in Scotland, community service in Germany, or the newly adopted "fine service" in Norway). Other countries have just simply ceased to impose default penalties for unpaid fines, as was done in Sweden in the 1980s. More recently also Denmark has practically ended the use of default imprisonment in the late 2000s. Finland, with probably the highest number of imposed fines in Europe, has also had the largest number of fine defaulters in the Nordic countries. However, in 2006–2009, the average number of fine defaulters in prisons (at any given day) has been reduced from around 200 to 50, removing all prosecutors' fines outside the default system (Table 2).

As shown in comparative analyses, default imprisonment has practically disappeared in Denmark, Sweden, France, Belgium, and Scotland. In Poland, Germany, and the Netherlands, about 5 % of prisoners are in custody for unpaid fines. However, also in Germany, a considerable reduction of fine default imprisonment has been achieved by extending community service schemes in some federal states (see Dünkel 2004).

Fines in Practice

Criminal justice systems vary in their ways to deal with minor infractions. Some include even the most common traffic offenses (as is done in Finland) in criminal code, other exclude not only traffic offenses, but also many other minor infractions from their criminal justice statistics. With these caveats in mind, Table 3 makes effort to highlight the use of fines in European countries.

First column displays the number of all imposed sanctions (convictions) either by the court or prosecutor. All offenses and infractions are included. Comparison with column D (convictions excluding traffic cases) shows how much each nation's criminal justice statistics are affected by traffic violations. Finland has the highest overall conviction rate (4,158), but three out of four convictions are for traffic cases (non-traffic convictions: 1,413). England and Wales, in turn, reports only non-traffic cases, as is done also by the Czech Republic, Latvia, and Cyprus. One must therefore look at the overall number of imposed fines (B) with caution. They reflect for some countries mainly traffic offenses, for others something else.

Some of these difficulties may be overcome, once the use of fines is analyzed separately for different offenses. Table 4 lists imposed fines and other sanctions/pop by countries and regions.

The average number of sanctions imposed for theft in 21 European nations was 153, of which 34 (22 %) were fines and 119 other types of penalties. For assault, the corresponding figures were 72 (all) 22/31 % (fines) and 50 (other sanctions).

The Nordic countries impose the largest number (300–600) of sanctions for theft offenses (together with Scotland and Hungary). However, in Finland, 9 out of 10 theft offenses are dealt with fines. This leaves around 52 other sanctions than fines imposed for theft in Finland, which, in turn, is considerably lower than the European

Fines in Europe, Table 3 The use of fines in Europe 2006

	General application of fines (all offenses)			
	A All convictions/ pop	B All imposed fines/ pop	C Fines of all convictions %	D Non-traffic convictions/pop
Finland	4,158	3,659	*88*	1,413
England and Wales	2,455	1,768	*72*	2,445
Belgium	1,805	1,661	*92*	380
Hungary	1,419	412	*29*	1,120
Sweden	1,319	712	*54*	1,023
Scotland	1,318	659	*50*	1,164
Switzerland	1,308	497	*38*	703
Poland	1,214	231	*19*	804
Netherlands	1,025	318	*31*	722
Germany	945	756	*80*	709
France	940	357	*38*	522
Croatia	730	66	*9*	679
Slovakia	700	28	*4*	650
Czech Re	675	27	*4*	675
Portugal	623	473	*76*	306
Austria	489	191	*39*	489
Latvia	435	30	*7*	435
Slovenia	404	28	*7*	381
Romania	234	44	*19*	196
Albania	222	84	*38*	214
Cyprus	204	80	*39*	204

Source: Compiled from European Sourcebook 2010
Country order by column A
A, Convictions either by the court or prosecutor. All offenses and infractions; B, Imposed fines (both courts and prosecutor)/100,000 population; C, The share of fines (%) of all convictions (A); D, The number of non-traffic convictions/100,000 pop

average (119). Finland has the second highest number of convictions also in assault (207), after Scotland. More than half (112) of these are fines, which is five times the European average (22). This indicates that Finland imposes monetary sanctions for both offenses in a much larger scale than any other country (while the victimization rate for theft in Finland is well below the European average and for assault little above the average). The alternative explanation is that minor forms of property and assault offenses are in other countries dealt by sanctions that are not included in the European Sourcebook statistics.

There are marked differences between the countries and it is hard to discern clear patterns. However, with the exception of Hungary, fines are much rarer in the Eastern countries, while the overall level of imposed sanctions is fairly high in these countries, especially in theft offenses. Scotland imposes a large number of sanctions both for theft and assault offenses. Also fines are used in a high scale. Switzerland imposes a very low number of sanctions for both offenses, and uses practically no fines. Whether the use of fines is connected with the extent of imprisonment will be discussed next.

Fines as an Alternative to Imprisonment

Fines provide one alternative to short-term prison sentences. Finland has been, perhaps, more consistent with her efforts to expand the use of fines over other type of sanctions. After the adoption of

Fines in Europe, Table 4 The use of fines and other sanctions (/pop) in Europe 2006 in assault and theft offenses

		Theft (all)				Assault (all)			
		Fine/pop	%	Other sanctions/pop	All	Fine/pop	%	Other sanctions/pop	All
Nordic	Finland	530	*91*	52	582	112	*54*	95	207
	Sweden	125	*47*	142	267	18	*19*	79	97
Western 1	Switzerland[a]	0	*1*	37	37	1	*9*	11	12
	Germany	104	*71*	43	147	41	*61*	26	67
	Austria	29	*30*	69	98	38	*59*	26	64
Western 2	Netherland	31	*18*	143	174	25	*23*	82	107
	France	14	*12*	101	115	16	*19*	70	86
	Belgium	42	*53*	38	80	32	*68*	15	47
UK	England and Wales[b]	24	*13*	160	184	4[a]	*7*	56[a]	60[a]
	Scotland	87	*30*	204	291	136	*46*	160	296
South	Portugal	13	*37*	21	34	39	*77*	12	51
	Slovenia	5	*5*	92	97	2	*4*	36	38
East	Poland	11	*6*	165	175	6	*6*	97	103
	Slovakia	2	*2*	122	124	2	*6*	36	38
	Czech Rep	3	*2*	138	141	1	*4*	25	26
	Hungary	65	*21*	246	311	16	*24*	51	67
	Latvia	1	*1*	137	138	2	*9*	17	19
ALL (21)	Mean	34	*22*	119	153	22	*31*	50	72

Source: Compiled from European Sourcebook 2010

[a] The low number of fines in Switzerland does not match with the data from official sentencing statistics (see http://www.bfs.admin.ch/bfs/portal/de/index/infothek) which indicates that fines (especially suspended fines) play an a substantial role in sentencing (in 2010 total number of convictions 98,200, of which suspended fines 71,881, partly suspended 1,050, and unsuspended 11,554). It is possible that formerly used (prior the day-fine reform) lump-fines (Busse) have not been included in the table

[b] The low number of assaults in England and Wales compared to earlier editions suggests that the definition in the statistics has been changed in a manner that prevents reliable comparisons

day-fines in 1921, the next major reform took place in Finland in 1977 as a part of a general prison reduction program. The monetary value of day-fines was increased, thus making fines a more credible alternative for short-term prison sentences. As the follow-up studies show, this aim was largely achieved also in practice (Lappi-Seppälä 2011). Similar results were gained also in Germany in connection with the replacement of short-term prison sentences (up to 6 months) by fines in 1969 and the adaption of the day-fine system in 1975 (see Dünkel and Morgenstern 2010).

The association between the use of fines and national incarceration rates is examined in figure below (Fig. 1).

There is a clear inverse correlation with the overall use of fines relative to population, even though this correlation is much dominated by high rates from Finland and Denmark. The same applies to the share of fines and incarceration rates, with much more even distribution. Regional analyses reveal the exceptional position of the Nordic countries. They use much more frequently fines, and they have much lower incarceration rates. Eastern countries represent the other end of this scale with high incarceration rates and scarce use of fines.

Effectiveness

It is generally known that offenders receiving fines have the lowest recidivism rates. For example, in Finland, about 25 % of those receiving fines became reconvicted for some form of penalty during the following 5 years whereas the figure for those receiving prison sentence was as high as 75 %. But as the initial recidivism risk in these two groups is totally different, these figures

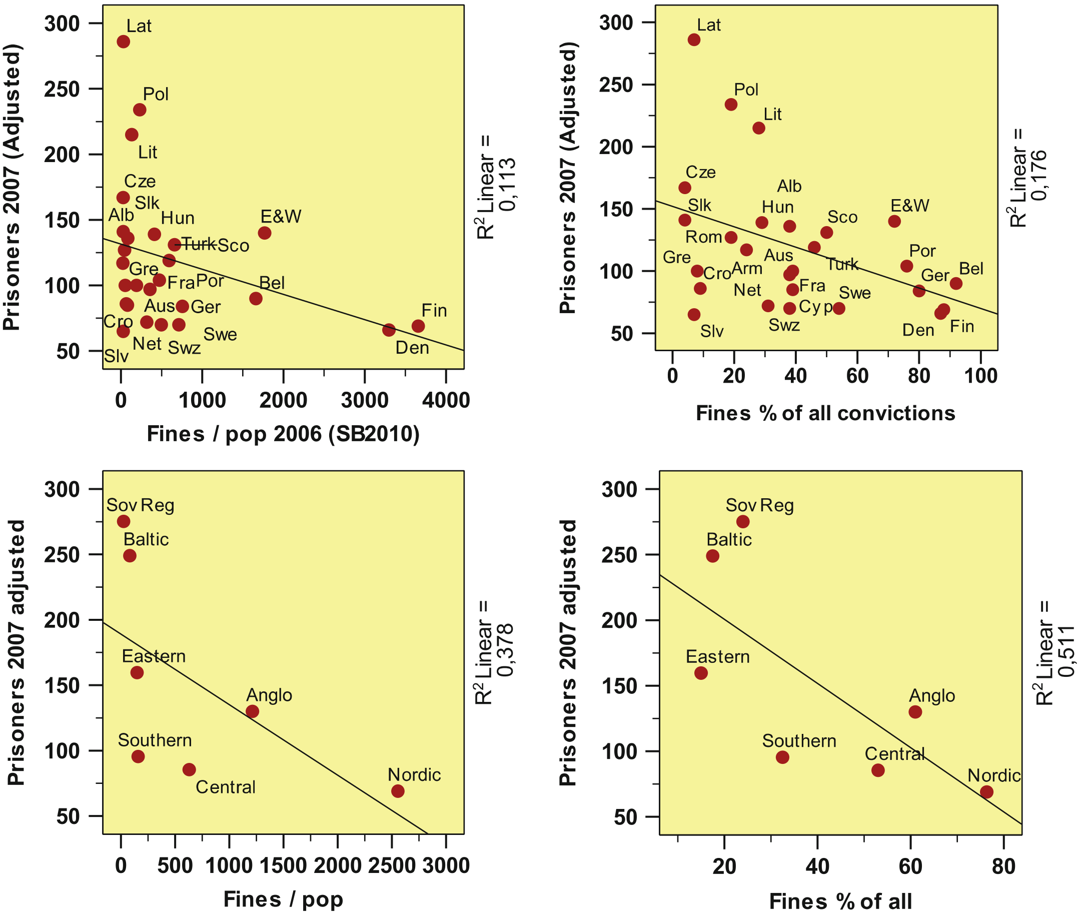

Fines in Europe, Fig. 1 The use of fines (/pop and as % of all convictions) and imprisonment rates by countries and by regions (Source: European Sourcebook 2010)

cannot be used as a basis for comparisons. Unfortunately, there is fairly little systematic research with control groups on the effects of fines on recidivism.

A study published in the USA in 1991 (Gordon and Glaser 1991) compared recidivism after different combinations of fines and other sanctions. Supervision with fines led to lower recidivism (25 %) than mere supervision (36 %). Equally short-term imprisonment with supervision and fines led to lower recidivism (37 %) than mere supervision and fines (50 %). However, the detected associations were statistically nonsignificant (Table 5).

A methodologically more advanced US study based on randomized experiment (Turner and Petersilia 1996) compared day-fines with flat-rate fines (no difference according to income). It turned out that the day-fine group had a lower recidivism rate (11 % vs. 17 %) and less technical violations (9 % vs. 22 %) than the control group (latter differences were statistically significant).

The third study is a meta-analysis covering 18 studies (Gendreau et al. 2000). The overall effects size in these 18 studies was −0.04. In other words, after groups had been matched, those receiving fines showed a 4 % lower recidivism rate than those receiving other types of penalties.

The impact of the size of fines on reoffending was examined by a large-scale study carried out

Fines in Europe, Table 5 Recidivism rates after fines

Study	Results
Gordon and Glaser (1991) USA	Supervision and fines (25 %)
No data on the matching of the groups	Only Supervision (36 %), nonsignificant
	Supervision, prison, and fines (37 %)
	Supervision and prison (50 %), nonsignificant
Turner and Petersilia (1996) USA	Arrests
Randomized experiment	Day-fines (11 %)
	Control group (17 %), nonsignificant
Poor data of treatment of control group	Technical violations
Consists of fixed fines	Day-fines (9 %)
	Control group (22 %), significant
Gendreau et al. (2000) Canada	Meta-analysis (18 studies)
	Effect-size −0.04 (= 4 % difference)
Moffatt and Poynton 2007	Reoffending was uninfluenced by the size of fines imposed for driving offenses
Marginal (specific) deterrent effect of the size of fines on recidivism	
Two-stage regression model	

in Australia by the NSW Bureau of Crime Statistics and Research, see Moffatt and Poynton (2007): 70,000 persons who received a court-imposed fine for a driving offense between 1998 and 2000 were followed for a period of 5 years to see whether they committed another driving offense. After controlling for a wide range of other factors likely to influence reoffending, a no relationship between the magnitude for the fine imposed and the likelihood of a further driving offense was found. The same negative result was obtained for drink-drive (PCA) offenses, drive while disqualified offenses, exceeding the speed limit and "other" driving offenses. The results are in line with the general findings that suggest that increased sentence severity does not go along with lower reoffending rates. However, the results leave the question of relative effectiveness of fines open.

Public Acceptance of Day-Fines: A Case Study from Finland

Fines are vastly the most extensively used criminal sanction. At least one out of ten Nordic citizens receives each year either a fine or a corresponding administrative sanction, usually for a traffic violation. As pointed out by a Swedish professor in criminology:

> There are many reasons why we should be more interested in the use of fine. We should never forget that more people are getting into contact with the legal system by being fined than by any other means. And because of this the way fines are handled by the representatives of the state will forever be remembered by the citizens. (Sveri 2005)

Smooth functioning of fining system is dependent on its legitimacy and public acceptance. Unfortunately, there is only limited research on the legitimacy and public perception of fines. The English government withdrew the unit-fine reform partly with reference to lack of public support – whether real or just printed in headlines. For this, the Finnish experience with another day-fine reform forms an interesting contrast.

In the shift of the millennium, the Finnish fining system received worldwide attention, thanks to extraordinary large traffic fines imposed for offenders with extraordinary high income. The Finnish law does not recognize any upper limit for a day-fine, and if the offenders' annual income is counted in millions, the fines may well be counted in thousands of Euros – or even more.

The other source for criticism was the worry expressed by the conservatives that a system based on gross-income instead of net-income leads to unjustly severe fines in general in higher income levels. Under the conservative coalition government, gross-fining was changed to net-fining in 1999 with the aim, "to introduce a more just fining system, whereby the size of the fine is perceived as fair among different income-groups" (Government Bill 74/1998).

A follow-up research was carried out in order to measure the degree of perceived fairness of the fining system. A total of 2,966 persons were interviewed at four different stages before and after the reform during the years 1999–2001. Key findings included the following (see Lappi-Seppälä 2004):

- In general just below 60 % considered traffic fines generally to be fair, about one fifth considered them too low, and about one sixth too severe.
- Three out of four considered that the system with day-fines was in general fair.
- When asked whether speeding fines should be graded according to the drivers' income, or whether the fine should be the same for all drivers, four out of five preferred differentiation according to income.

The assessments of fairness varied in different respondent groups. Not surprisingly, those who did not drive considered the system with day-fines to be fairer than those who drove (still, 70 % of drivers considered the system as fair). Men were more critical of the system with day-fines than women. Women would often be more apt to demand more severe traffic fines than men. As one could anticipate, respondents' critical attitudes toward the day-fine system grew along with income-level.

Several of the findings were in clear contradiction with general presumptions about the public support of the fining system. Evidently the fears of the perceived unfairness of the fining system were grossly exaggerated. The reform of the fining system introduced in 1999 did not as such bring about any significant change in public opinion, once measured before and after the reform. Neither seemed the stirring fines imposed in summer 2000 appear to have affected the public's confidence in the fining system for traffic offenses.

But as it also turned out, the general public was poorly aware of the rules concerning fines for traffic offenses. The fact, whether fines were counted on the basis on net-income or gross-income, turned out to be completely irrelevant for the perceived fairness of the fining system.

One major lesson was that the images of the public opinion, as expressed by the media and the politicians, may have very little to do with the views shared by the "silent majority." Not surprisingly, the perceptions of fairness and the proper amount of control depended heavily on the respondents' personal circumstances and financial situation: Those who do not drive wish more control and stiffer penalties, those who drive are less satisfied with the fining system, and those who have the highest income-level are the least satisfied with the day-fine system. Envy and self-interest are also essential elements in public opinion.

It seems that the legislators' attempt to "gratify" the well-off offenders by moving from "gross-fining" to "net-fining" did not succeed. Paradoxically, after the reform, the discontent of the fining system grew, especially among respondents in the highest income-level. In other words, those who benefited most from the new net-fining were the least satisfied with the changes brought by the reform. Either they had misperceived the content of the reform, or they were unhappy with the fact that giving false information to the police of their income had become a criminal offense (for the criminalization of fine deception, see above).

Surveys leave the door open for different interpretations. One fairly plausible conclusion could be that a clear majority of Finnish people consider the day-fine system and the fines imposed for traffic violations as reasonable fair and just. This was the case before and this was also the case after the reform (which, in fact, largely passed their attention). Another conclusion could be that the assumptions of the contents of "public opinion" may rest on a shaky basis – as may also lay the efforts to please this (assumed) general opinion.

Related Entries

- ▶ Community Service in Europe
- ▶ Examining the Effectiveness of Correctional Interventions
- ▶ Probation and Community Sanctions

Recommended Reading and References

Aebi M, Delgrande N (2012) Council of Europe annual penal statistics, SPACE I, survey 2010. Council of Europe, Strasbourg

Aebi M, de Cavarlay BA, Barclay G et al (2010) The European sourcebook of crime and criminal justice statistics – 2010, 4th edn. Boom Juridische uitgevers, The Hague, Cited as Sourcebook 2010

Ashworth A, Gibson B (1994) Altering the sentencing framework. Crim Law Rev February:101–109

Bureau of Justice Assistance (1996) How to use structured fines (day fines) as an intermediate sanction. https://www.ncjrs.gov/pdffiles/156242.pdf

Charman E, Gibson B, Honess T, Morgan R (1996) Fine impositions and enforcement following the Criminal Justice Act 1993, vol 36, Home Office Research findings. Home Office, London

Dünkel F (2004) Reducing the population of fine defaulters in prisons: experiences with community service in Mecklenburg-Western Pomerania (Germany). In: Council of Europe (ed) Crime policy in Europe. Good practices and promising examples. Council of Europe, Strasbourg, pp 127–138

Dünkel F, Morgenstern C (2010) Deutschland. In: Dünkel F, Lappi-Seppälä T, Morgenstern C, van ZyL Smit D (eds) Kriminalität, Kriminalpolitik, strafrechtliche Sanktionspraxis und Gefangenenraten im europäischen Vergleich, vol 1. Forum Verlag Godesberg, Mönchengladbach, pp 97–230

Gendreau P, Goggin C, Cullen F, Andrews D (2000) Effects of community sanctions and incarceration on recidivism. Forum Correct Res 12(2):10–13

Gordon MA, Glaser D (1991) The use and effects of financial penalties in municipal courts. Criminology 29:651–676

Greene J (1988) Structuring criminal fines: making an "Intermediate Penalty" more useful and equitable. Justice Syst J 13:37–50

Lappi-Seppälä T (2004) Public perceptions and the fairness of the day-fine system. An evaluation of the 1999 day-fine reform. Tidskrift utgiven av Juridiska Föreningen i Finland 3-4/2004. pp 386–397

Lappi-Seppälä T (2011) Changes in penal policy in Finland. In: Kury H, Shea E (eds) Punitivity. International developments, vol 1, Punitiveness – a global phenomenon? Universitätsverlag Dr. Brockmeyer, Bochum, pp 251–287

Moffatt S, Poynton S (2007) The deterrent effect of higher fines on recidivism: driving offences. Crime and Justice. Bulletin NSW Bureau of Crime Statistics and Research Contemporary Issues in Crime and Justice Number 106. March 2007. http://www.lawlink.nsw.gov.au/lawlink/bocsar/ll_bocsar.nsf/vwFiles/cjb106.pdf/$file/cjb106.pdf

Rusche G, Kirchheimer O (1939) Punishment and social structure. Russel & Russel, New York

Sveri K (2005) Incarceration for non-payment of fines. In: Bondeson U (ed) Crime and justice in Scandinavia. Forlaget Thomas, Kopenhagen pp 373–381

Turner S, Petersilia J (1996) Evaluation of day fines in Maricopa county, Arizona, 1991–1993. RAND, Ann Arbor

Fingerprint Identification

Stephen Bleay[1] and David Charlton[2]
[1]Centre for Applied Science and Technology, Home Office Science, Sandridge, Hertfordshire, UK
[2]Surrey and Sussex Police Forensic Services, School of Applied Sciences, Bournemouth University, Bournemouth, UK

Overview

This entry will provide an overview of the stages leading to fingerprint identification, including the enhancement of marks deposited at crime scenes, obtaining images of the enhanced marks that are suitable for identification, and the methodologies that are used during the identification process itself. The range of enhancement processes that can be used to reveal marks is described in detail, showing how different constituents of deposited marks can be progressively targeted to maximize the number of marks recovered. This entry also addresses some of the recent research associated with the fingerprint identification stage, including the possible influences of cognitive bias and the use of probabilistic approaches to describe fingerprint identifications in place of 100 % certainty.

Fundamentals

Definitions

In this section, the term "fingerprint" will be used to describe impressions of the ridge patterns on fingertips that have been taken under controlled conditions, for example, inked or digitally enrolled ten-print sets. The term "finger mark" is used to describe the marks recovered from articles or surfaces at crime scenes.

"Finger mark recovery" is defined as the process by which finger marks present on articles and surfaces are revealed and captured. The output of the finger mark recovery process is an image of the mark that can be submitted for comparison and identification.

"Fingerprint identification" is defined as the comparison of finger marks against sets of fingerprints, or two sets of fingerprints against each other, so that it can be established that they originate from a common source.

Finger Mark Recovery

Introduction

The finger mark recovery process begins with an initial examination of an article or surface where it is suspected that finger marks may be present. In some cases marks may already be visible during the initial examination, either because they have been deposited in a visible contaminant ("patent" marks) or because they have left permanent impressions in a soft surface ("plastic" marks). The majority of marks present on the surface are not visible during this initial examination, and some means of enhancement will be required to make them detectable by an eye or an imaging system. These marks are known as "latent" marks.

One of the principal differences between finger marks and other types of forensic evidence is the wide range of mark enhancement techniques that are available, and the fact that it is possible to use several of these techniques in sequence. The sequential application of enhancement techniques, commencing with nondestructive processes and progressing to techniques directly reacting with constituents of the mark, maximizes the chances of their recovery.

A wide range of optical, physical, and chemical techniques are available for the detection and enhancement of latent finger marks. The choice of the best sequence of techniques will depend on several factors that include the following:

- The nature of the surface and the presence of any particular contaminants
- Environmental factors
- The likely age of any evidential finger marks
- Investigative priorities
- Other forensic evidence

The techniques that are used for finger mark enhancement all have the same purpose, which is to increase the contrast between the finger mark ridges and the surface they have been deposited on to an extent where the mark can be readily seen and captured. Many different properties of the mark may be used to achieve this contrast, including optical properties, adhesive properties, chemistry of the mark, and electrical properties. The chemistry of finger marks is usually a complex one, and the chemicals that may be present in marks include constituents of eccrine, sebaceous, and possibly apocrine sweat, contaminants from any substances handled, and substances significant at crime scenes such as blood. If additional information can be obtained about the likely composition of the finger marks (e.g., contaminants that may be present), this will assist in the selection of the most appropriate enhancement techniques.

When considering which enhancement techniques should be applied to a particular article or surface, the nature of the surface and the environment it has been exposed to should also be considered. Surface characteristics such as porosity, texture, and color will all influence what happens to the mark post-deposition and, thus, will influence the selection of the most appropriate technique(s). This will be equally true of any known environmental conditions such as exposure to water and high temperatures. The initial examination stage should be used to obtain as much of this information as possible so that the selection of enhancement processes is optimized.

Finger Mark Enhancement Techniques

The sections below give an overview of some of the finger mark enhancement processes that are available, together with some indication of their potential applications. Many other processes have been reported, those outlined below cover a representative range.

Nondestructive Processes

Visual Examination Visual examination utilizes differences between the optical properties of the mark and the surface to provide contrast. The surface is illuminated with a light source, and the angle and proximity of the light source to the surface varied until optimal contrast is obtained. The process is most suited for greasy marks and marks in dust on nonporous surfaces.

Fluorescence Examination Fluorescence examination is performed by illuminating the surface with a high intensity light source and viewing any fluorescence through a filter that blocks the illuminating wavelength range (Menzel 1999). It can be used to detect latent finger marks on both porous and nonporous surfaces, or to enhance marks developed using other techniques. Contrast is obtained by either the mark or the background fluorescing, and the other component absorbing the incident light. When used to detect latent marks, any fluorescence observed in the mark is most likely to arise from the presence of contaminants although some natural constituents may fluoresce.

Ultraviolet Imaging Ultraviolet imaging is a specialized technique that requires an imaging system sensitive to short wave ultraviolet radiation. The surface is irradiated with short wave ultraviolet, and the outcome is viewed through the imaging system. On porous surfaces, marks may absorb radiation and the surface may reflect it, and on nonporous surfaces marks scatter radiation to a greater extent than the surface (Creer 1995). Both mechanisms may give sufficient contrast for the marks to be revealed. The short wavelength of the ultraviolet radiation is scattered more than visible wavelengths, meaning that fainter traces of marks can be detected. The technique should be used with caution because short wave ultraviolet is harmful to both skin and eyes, and can destroy DNA evidence.

Selective Deposition Processes

Powders Powdering is one of the oldest finger mark enhancement techniques. Finely divided powders are applied to the surface using a brush (or an equivalent delivery system such as a magnetic applicator), and some of the particles selectively adhere to the ridges of the mark, making it visible (Fig. 1). A variety of adhesive mechanisms may operate, including wetting by capillary action and physical adhesion due to the "sticky" nature of the mark. A wide range of powder types are available, and the most appropriate should be selected taking into account factors including the surface texture and color.

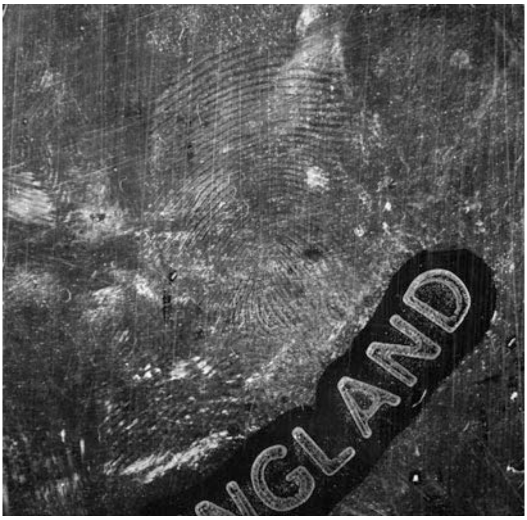

Fingerprint Identification, Fig. 1 A mark developed using aluminum flake powder and lifted using transparent tape

Small Particle Reagent Small particle reagent formulations utilize a dilute mixture of particulates (most commonly molybdenum disulphide) in a water/detergent solution. The detergent molecules form micelles around the particles, the lipophilic molecules in the micelles causing the particles to preferentially attach to the lipids in the mark. The article being treated is immersed in an agitated solution of the small particle reagent, and the particles are allowed to settle onto the article. Excess particulates are gently washed from the surface to reveal the mark. The technique is suited for use on nonporous surfaces and can be used on articles known to have been wetted.

Powder Suspensions Powder suspensions are very similar in nature to small particle reagent, but consist of a much more concentrated mixture of powder (either carbon, iron oxide, or titanium dioxide) in a detergent/water mixture. This is painted onto the surface using a brush and then washed off before it can dry. Some constituent of the mark (thought to be eccrine constituents held in an insoluble matrix) destabilizes the micelles and causes particles to preferentially settle on the ridges. The technique can be used on nonporous surfaces, adhesive surfaces (Bratton and Gregus

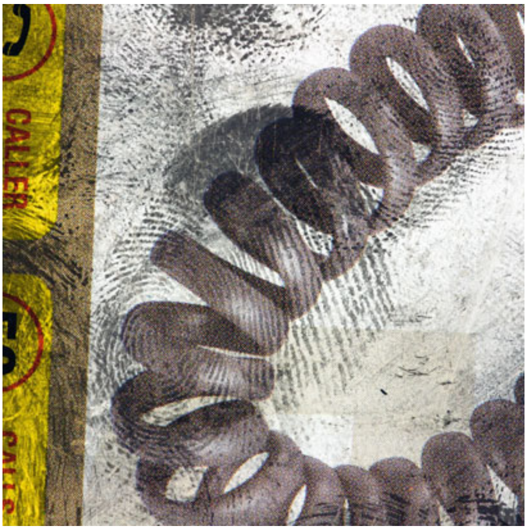

Fingerprint Identification, Fig. 2 Marks developed on a semi-porous printed cardboard box using black powder suspension

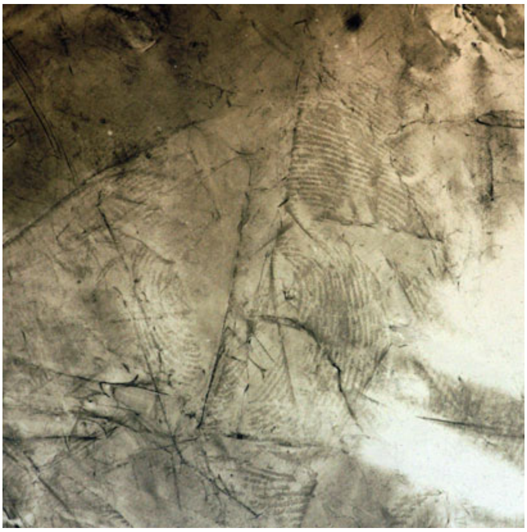

Fingerprint Identification, Fig. 3 A mark developed using vacuum metal deposition on a polyethylene bag and imaged using transmitted light

1997), and semi-porous surfaces, including those that have been wetted (Fig. 2).

Vacuum Metal Deposition Vacuum metal deposition reveals marks by selective deposition of metal onto the surface. Vacuum is required to reduce the metal vapor pressure to a point where it can be readily evaporated. The most commonly used process involves an initial evaporation of a small quantity of gold, which deposits all over the surface as small clusters. The size and distribution of these clusters is influenced by the nature of the surface, and often differs between the ridges and the background. A second evaporation of zinc is then conducted, zinc relying on the presence of gold clusters to initiate and grow as a film on the surface. Differences between the growth rate of zinc on the ridges and background regions make the marks visible (Kent et al. 1976). Vacuum metal deposition is most suited to nonporous surfaces but can also be used on semi-porous and porous surfaces including fabrics. It can be used to develop marks on nonporous surfaces that have been wetted (Fig. 3).

Multimetal Deposition Multimetal deposition is a liquid phase deposition process. The article is first exposed to an acidified colloidal gold solution. This causes constituents of the finger mark ridges to become charged, and colloidal particles are attracted by and bound to the mark. This is not sufficient for the mark to become visible, so the article is exposed to a second "physical developer" solution consisting of a silver-iron redox combination. Silver particles are preferentially deposited on the gold already bound to the ridges, and the mark becomes visible as a consequence (Schnetz and Margot 2001). Multimetal deposition can theoretically be applied to both nonporous and porous surfaces but in practice is most suited to nonporous surfaces, in particular plasticized polymers such as "cling film."

Cyanoacrylate Fuming Cyanoacrylate fuming reveals marks by selective polymerization of cyanoacrylate monomer vapor on finger mark ridges. Vapor can be produced in several ways, but is most commonly achieved by evaporation of liquid cyanoacrylate monomer from a dish placed on a heated stage. The resultant polycyanoacrylate deposit appears white and may give sufficient contrast for marks to be seen on dark surfaces. However, subsequent enhancement techniques are often used to increase contrast, such as staining the polymer deposit with

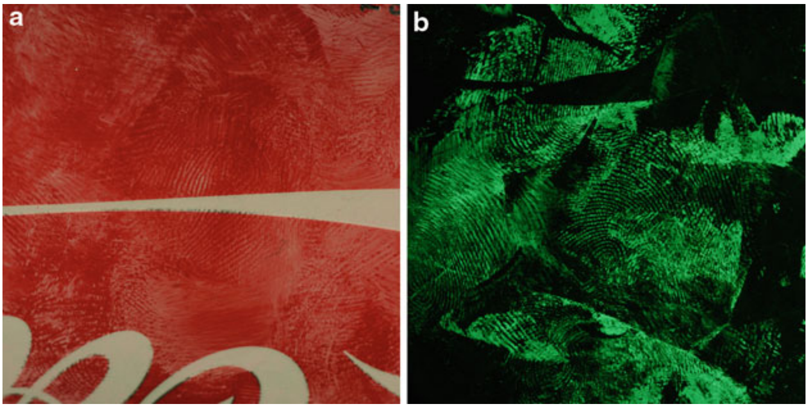

Fingerprint Identification, Fig. 4 Marks developed using cyanoacrylate fuming (**a**) marks on a drinks can viewed using white light, (**b**) marks on a black refuse bag, dyed with basic yellow 40 and viewed under fluorescence

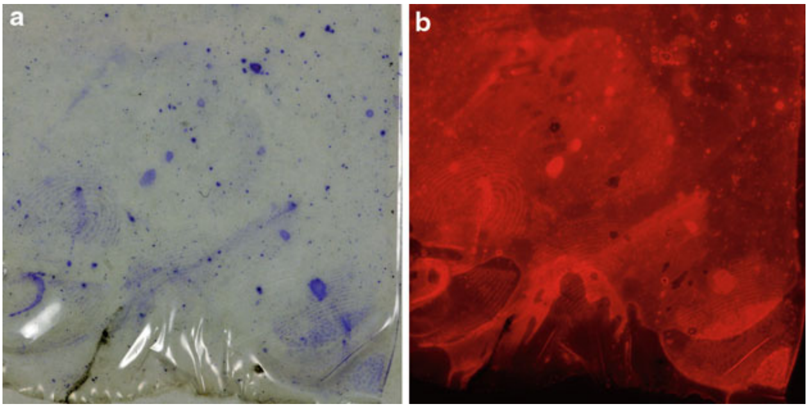

Fingerprint Identification, Fig. 5 Marks developed on adhesive tape using basic violet 3 (**a**) viewed using white light, (**b**) viewed under fluorescence

fluorescent dyes (Fig. 4). The polymerization reaction is initiated and catalyzed by the presence of finger mark constituents including salts (especially sodium chloride) and some other eccrine constituents in the presence of moisture (Wargacki et al. 2007). Fuming cabinets may be humidified to supply this moisture in a controlled manner. Cyanoacrylate fuming is generally used on nonporous surfaces, but is not suitable for use on wetted surfaces because the eccrine constituents that initiate polymerization are dissolved.

Staining Reagents

Basic Violet 3 Basic violet 3 is a dye that can be used to preferentially stain lipids, greasy contaminants, and epithelial cells sometimes shed within finger marks (especially on adhesive tapes). The dye molecules preferentially transfer from solution into any compatible constituents that may be present within the mark, staining them violet. Excess dye is then washed from the surface. The dye is also fluorescent and marks that are not stained sufficiently for them to be detected visually may become visible when examined using fluorescence. Basic violet 3 can be used on surfaces contaminated with grease, and on nonporous and adhesive surfaces. On nonporous and adhesive surfaces, it is generally used last in a sequence because it may cause heavy staining, but is capable of developing reasonable numbers of additional marks. The dye and solvents used do have health and safety issues associated with them and use should be restricted to small articles under controlled conditions (Fig. 5).

Solvent Black 3 Solvent black 3 is another dye that is used to selectively stain the lipid constituents of finger marks together with greasy contaminants. The article being treated is exposed to the dye solution until staining has occurred and then the excess washed from the surface, leaving the marks stained a deep blue/black. Its operational use tends to be confined to nonporous articles known to be contaminated with grease or drinks residue. It is nonfluorescent and nontoxic, although difficult to clean from scenes.

Amino Acid Reagents

Ninhydrin Ninhydrin is one of the mostly widely used finger mark enhancement techniques worldwide. It is most suited for use on paper articles but can be effective on other porous surfaces including matt painted walls and untreated wood. Ninhydrin is an amino acid reagent and reacts with the amino acids present in finger marks to give a purple reaction product (Oden and von Hofsten 1954). The surface or article to be treated is exposed to the ninhydrin solution by dipping or brushing, and is then allowed to dry. Small articles can then be exposed to elevated temperature (~80 °C) and humidity (~65 % relative humidity) which accelerates the reaction so that the product is visible in <10 min (Fig. 6). When ninhydrin is applied to large areas such as walls, this is not possible and marks are then allowed to develop at room temperature, which may take over a week.

1,8 Diazofluoren-9-one (DFO) DFO is another amino acid reagent, although in this case the reaction product formed between amino acids and DFO is fluorescent and fluorescence examination is required to visualize the marks (Pounds et al. 1990). Illumination in the green region of the spectrum is required, and the fluorescent marks appear orange (Fig. 7). DFO is more sensitive to ninhydrin, detecting lower concentrations of amino acids, but is less widely used because of the additional time taken to detect marks. DFO is applied in a similar way to ninhydrin, that is, by brushing or dipping, and exposure to temperatures >100 °C is required to develop marks in a reasonable time (20 min at 100 °C). For this reason, DFO cannot easily be used at scenes. DFO can be used in a sequential treatment routine before ninhydrin, which may develop additional marks when used after it.

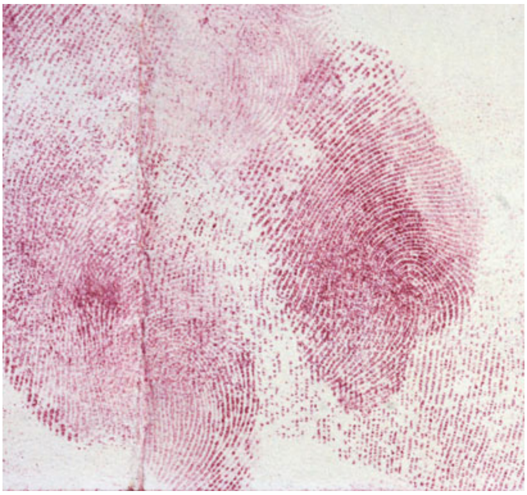

Fingerprint Identification, Fig. 6 Mark developed on white paper using ninhydrin

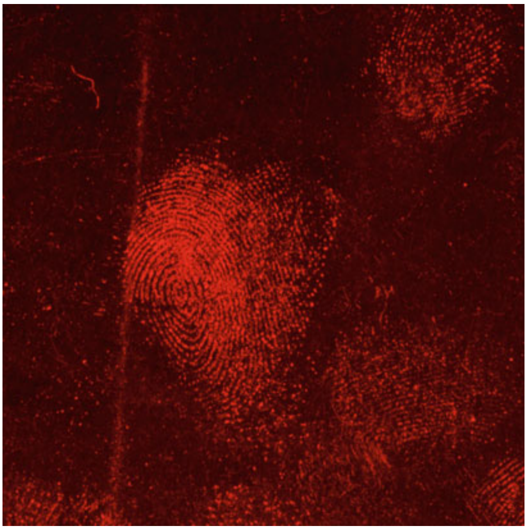

Fingerprint Identification, Fig. 7 Mark developed using 1,8 Diazafluoren-9-one on paper viewed under fluorescence

1,2 Indandione 1,2 Indandione is a more recently introduced amino acid reagent which is displacing DFO from some applications. The is now general consensus that 1,2 Indandione develops more, intensely fluorescent marks than DFO, but it has still not be determined whether the DFO-ninhydrin sequence remains more effective than a sequence of 1,2

Indandione-ninhydrin. 1,2 Indandione is applied and processed in the same way as DFO. Bright pink visible marks are occasionally developed, but the reagent is most effective in fluorescence mode where it is illuminated with green light and produces a yellow-orange fluorescence (Hauze et al. 1998).

Chloride Reagents

Silver Nitrate Silver nitrate is a long recognized enhancement chemical that targets the chlorides present in finger marks and is most suited to use on porous surfaces. A solution of silver nitrate is applied to the surface, which is then allowed to dry in the dark. The dried surface is then exposed to light (with ultraviolet giving most rapid development), and a reaction occurs which first converts to chlorides to silver chloride, and ultimately to silver. Developed marks appear dark brown and darken to gray/black with time. A disadvantage of the technique is that the background also progressively darkens and will ultimately obscure the marks. In common with the amino acid reagents, silver nitrate cannot be used on wetted surfaces because these constituents are water soluble.

Other Processes for Porous Surfaces

Physical Developer Physical developer is the most widely used reagent for porous surfaces that have been wetted. It consists of an unstable silver-iron redox solution, which certain constituents in the finger mark (thought to be eccrine constituents bound in an insoluble matrix) can locally destabilize resulting in the precipitation and growth of fine silver particles on the ridges. Because certain alkali fillers added to paper can also cause precipitation of silver, the paper is first immersed in a weak acid bath until all reaction with these fillers ceases, then immersed in the physical developer solution (Ramotowski 2000). The technique can develop addition marks if used sequentially after DFO and ninhydrin. It can also be used on other porous surfaces such as raw wood.

Oil Red O Oil Red O is a lipid stain that has been proposed for use on porous surfaces such as paper, in particular those that have been wetted. The article to be treated is immersed in a solution of Oil Red O until strong staining of the mark is achieved. This may take up to 90 min. The article is then transferred to a buffer solution, which stops the staining action and removes excess dye from the background (Beaudoin 2004). Because the process targets different constituents of the mark from DFO and ninhydrin, it can potentially be used in sequence after them and before physical developer. However, the lipids most strongly stained by Oil Red O break down with time and the effectiveness of the technique is significantly reduced on marks over 4 weeks old.

Iodine Iodine is possibly the oldest reported chemical enhancement technique for finger marks. The traditional means of applying iodine is as a fuming process, enclosing the article to be treated in a chamber with fumes of elemental iodine. The iodine fumes are preferentially adsorbed into the ridges of the mark, staining them a brown color. This coloration will gradually disappear unless a fixing agent is applied, the most commonly used being a solution of α-napthoflavone which stains the developed mark blue. Iodine can also be applied as a solution, with the fixing agent incorporated into the formulation. Blue marks are rapidly developed as the solution evaporates from the surface. Iodine is most commonly used on porous surfaces but can also be applied to nonporous surfaces, including those contaminated with grease. The principal constituent of natural secretions targeted by iodine is squalene, which is rapidly lost from the finger mark. As a consequence, iodine tends to be ineffective on marks more than a few days old.

Blood Dyes The reagents used to enhance marks in blood fall into two principal categories, protein stains that selectively stain the proteinaceous constituents of blood, and reactive dyes that are catalyzed by the heme constituent. Both types of dyes are outlined below.

Acid Dyes Acid dyes operate by selectively staining the proteins present in blood.

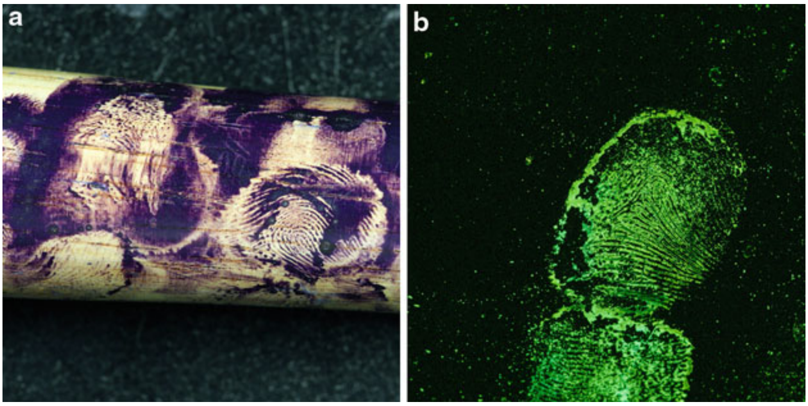

Fingerprint Identification, Fig. 8 Marks in blood enhanced using acid dyes (**a**) marks on a baseball bat handle enhanced with acid violet 17 and (**b**) marks on black plastic piping enhanced with acid yellow 7 and viewed under fluorescence

They require the blood to be first fixed, either by the application of heat, drying using methanol, or application of a fixing agent such as 5-sulphosalicyclic acid. Once the mark has been fixed, a staining solution containing the dye is applied, and the dye molecules become selectively bound to the proteins giving them a strong coloration. The excess dye is then washed from the surface using a destaining solution. A range of acid dyes have been used for this purpose, including acid black 1, acid violet 17, acid violet 19, and acid yellow 7 (Sears et al. 2005). All of these dyes are suitable for use on nonporous surfaces and most are suitable for nonporous surfaces although some dyes (e.g., acid yellow 7) cannot be easily washed from the background (Fig. 8).

Leuco Dyes Leuco dyes are the principal type of reactive dye that is used for enhancement of marks in blood. They are made into solutions in their colorless form, and when applied to marks in blood a reaction is catalyzed by the heme molecule which results in the formation of the colored form of the compound, thus making the mark visible. The two dyes most commonly used for this purpose are leuco crystal violet, which stains marks violet, and leuco-malachite green, which stains marks a green-blue. The reagents are most commonly applied in the form of a fine spray and are more suited to speculatively searching for marks in blood than the acid dyes.

Because there is more protein than heme present in blood, the acid dyes are expected to be more sensitive in this application. However, both types of reagents may give false-positive reactions with certain substances.

Enhancement of Developed Marks

Once marks have been enhanced, there may be sufficient contrast for them to be easily imaged. If this is not the case, then it may be possible to apply filtration or use more advanced imaging processes to make the mark sufficiently visible for imaging. Some of the means by which this can be achieved are described below.

Color Contrast Filtration If the mark is difficult to visualize, and either the mark or the background is colored, it may be possible to use colored filters (or a colored light source) to increase the contrast between the mark and the background. Where the mark is colored, but too faint to distinguish from the background, a filter of contrasting color can be used to increase visibility of the ridges. An example of this is the green filter sometimes used to improve the contrast of marks developed using ninhydrin. Color contrast filters can also be used to improve visibility of marks developed on light-colored surfaces using enhancement techniques such as white powder suspensions. In this case, the contrast filter darkens the background and the mark remains light.

Infrared Imaging Infrared imaging can be considered in cases where developed marks are obscured by patterned and/or multicolored backgrounds. Marks developed using a metallic or inorganic deposition process (e.g., powders, powder suspensions, physical developer) will remain visible when viewed in the infrared region of the spectrum, whereas many backgrounds printed in inks based on organic compounds become transparent. As a consequence, the effect of the background is suppressed, and the mark can be clearly seen. This technique requires an imaging system that is sensitive in the infrared region of the spectrum, a long pass filter blocking visible wavelengths and transmitting infrared placed in front of the imaging system, and a light source with significant intensity of output in the infrared.

Imaging of Developed Marks

Once marks have been enhanced sufficiently for ridge detail to be visible, and it is considered that sufficient ridge detail is present for a potential identification, it becomes necessary to capture an image of the mark.

To meet the requirements for fingerprint identification, it is essential that any image taken provides identification experts with all the information necessary to make decisions on identification/exclusion. The image of the mark should therefore ideally contain all features of significance captured at a resolution appropriate to that feature type, and have clear distinction between what constitutes a fingerprint ridge and what constitutes "background."

Imaging was traditionally carried out using film photography, although this has now almost entirely been replaced by digital imaging in many countries. Cameras and other digital imaging equipment such as scanners may be used for image capture. Alternatively, for marks developed using powders, a "lift" taken from the surface may occasionally be submitted to the bureau as the "image."

Digital images can be enhanced post-capture using image processing tools. If this is to be carried out before the image is submitted to the bureau, it is essential that a "master copy" of the originally captured image is retained, together with an audit trail of the modifications that have been carried out to the image submitted for identification.

The output of the finger mark recovery process is therefore an image of the mark that is submitted for comparison with fingerprints taken from suspects. This image may be in the form of an electronic image viewed on screen, a photograph printed from a negative, an image printed on a digital printer, or a physical lift from a powdered mark. In all cases, the identification specialist receiving the image should be given as much information as possible about the mark, including the location and surface it has been developed on, the technique(s) that have been used to enhance it, any modifications that have been made to the image, and any methods used to produce a hard copy. This information will all assist in the subsequent comparison and identification process, which is outlined in the following sections.

Future Trends in Finger Mark Recovery

There is a continued drive to increase the number of finger marks that are recovered from articles, whether it be to identify the optimum technique for "low yield" surfaces such as cling film and fabrics, or to investigate novel chemistries that can interact with constituents of the mark that are not currently targeted. Of particular note is ongoing research using nanoparticles, which can be chemically modified to attach themselves to specific finger mark constituents or suspect contaminants such as drug metabolites.

The compositional analysis of finger marks using advanced analytical techniques such as matrix-assisted laser desorption/ionization (MALDI) and attenuated total reflection Fourier transform infrared spectroscopy (ATR-FTIR) is also generating much interest, as this may offer a means of providing additional contextual information about the donor of a mark. Information that it may be possible to obtain includes the sex of the donor, the approximate age of the mark, and what the donor may have been handling (e.g., drugs or explosives). For these techniques to be truly effective, they will need to be integrated with conventional development techniques.

Fingerprint Identification

Introduction

The premise of fingerprint identification is that the fingerprints of each individual are unique and are persistent throughout their lifetime. It is therefore possible to identify an individual by comparing marks deposited at crime scenes with sets of prints obtained under controlled conditions. By finding sufficiency of agreement between the mark and the corresponding print, and an absence of any unexplained differences, identification is deemed possible. Although it is true that no two individuals have ever been found with the same fingerprints, the marks left at scenes often present partial, smudged, and distorted representations of the fingertip, and therefore, the role of the latent print examiner is to satisfy themselves that there is sufficient information present in the mark to individualize.

The fingerprints found on the tips of each finger consist of patterns formed by ridges in the surface of the skin. These were traditionally grouped into three general types of pattern, the whorl (where the central ridges complete a full revolution), the loop (where ridges enter at one side of the finger, curve back on themselves without touching, then exit at the same side), and arches (where ridges run from one side of the finger to the other without any backward turn). In practice, there are subdivisions of each pattern type, and other more complex pattern types may be found (Fig. 9).

Latent print examiners should assess holistically all features within a fingerprint from which to make conclusions as to identity. These features are broken down into three levels of detail. Level one refers to the basic pattern of the print outlined above, level two refers to the Galton details that consist of discontinuities in the ridge flow, and level three refers to the configuration of sweat pores and shapes of the ridge edges (edgeoscopy and poroscopy) (Fig. 10).

Under a magnifying glass, friction ridges are not uniform in configuration, nor do they necessarily flow in predictable direction or flow. Friction ridge skin has breaks and interruptions within the structure known as Galton details (Galton 1892; Ashbaugh 1999), also known as ridge characteristics or minutiae.

It is these characteristics that, when visible to the examiner, provide the principal means to make a determination of individualization for a latent crime scene finger mark against a suspect's ten-print card. There are two primary classifications of the characteristics, namely, ridge endings (A), where ridges stop abruptly, and bifurcations, where a ridge divides into two (B) (Fig. 11).

In this instance, the ridges running parallel to this feature diverge to accommodate the ridge division. There are variations of these primary characteristics. For example, a lake is where two bifurcations join together (C); an independent ridge is a short ridge that is divorced from any other ridge (D). A spur (E) is a combination of a small independent ridge and a bifurcation and a crossover (F) as the name suggests is a small ridge joined at each end to two parallel ridges.

Every feature and characteristic in friction ridge skin is readily identifiable and identification can be made so long as there are sufficient characteristics of acceptable quality present. It is possible to match very small amounts of friction ridge skin from a crime scene mark against the prints of a suspect, so long as there is sufficient clarity upon which to rely upon the permanence and reproducibility of visible features.

Identification Methodologies

When comparing detail between a crime scene mark and a ten-print exemplar, a generic methodology has been adopted known as ACE-V (Ashbaugh 1999). In order to make a value judgment on any identification, the examiner must first **assess** (or **analyze**) the friction ridge detail in the scene mark so that useful reference points can be identified in the next phase. Next, there must be a **comparison** of the mark features with that of features within the reference print. Then there must be an **evaluation** of the findings, from which conclusions as to identity must be derived. Finally, as in all scientific methods, there must be an element of peer review, known as **verification,** to repeat the ACE process independently to arrive at the same conclusions as that of a colleague.

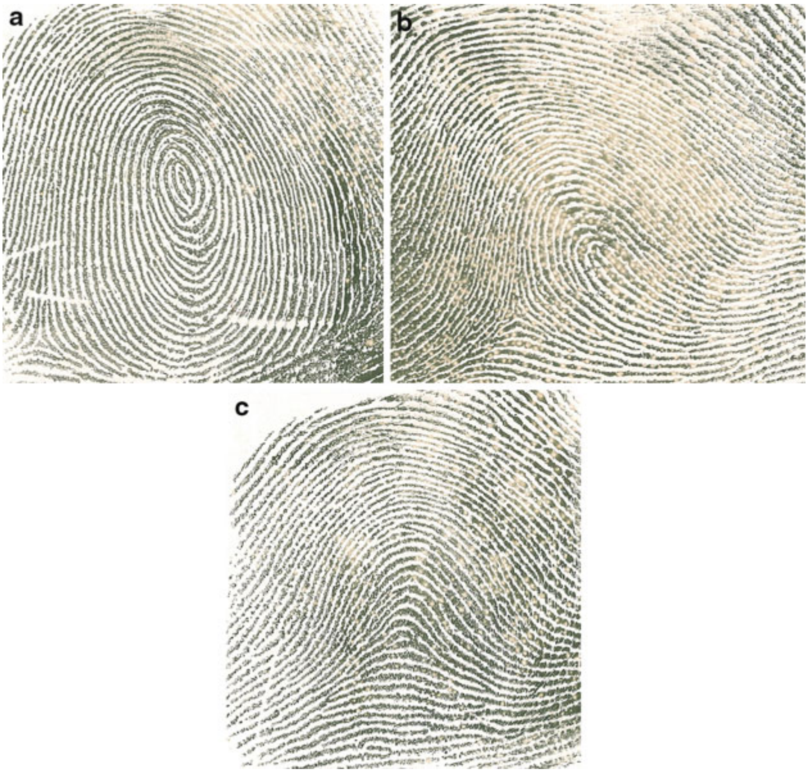

Fingerprint Identification, Fig. 9 Types of fingerprint pattern, (**a**) whorl, (**b**) loop, and (**c**) arch

During the **Assessment** (or **Analysis**) phase, the fingerprint examiner will look for the following:

- Distortion
- Development methods and medium used to visualize latent material (chemicals and powders, etc.)
- Deposition pressure to highlight areas of possible distortion
- Anatomical attributes (features)
- Clarity

During the **Comparison** phase, the examiner will look for similarities in the following:

- Pattern
- Ridge path
- Ridge shape
- Pore positioning

During the **Evaluation**, the examiner will look to form an opinion in the following:

- Can the mark be eliminated?
- Is there sufficient information available to individualize?

During **Verification**, an independent assessment of the casework is undertaken to see whether the ACE process has been correctly executed and that the conclusions reached are consistent with the original findings. In the UK, for a crime scene mark to be identified, the comparison must be undertaken three times. One check will be made by the original examiner, then verification by two further experts who must come to the same conclusions independently.

In some countries, there are a minimum number of features that must be found in agreement between the crime scene mark and the print for the identification to be confirmed. For example, in many European countries there is a requirement for 12 features to be found in agreement with none in unexplained disagreement for an identification to be confirmed. In other parts of the world, including the USA and the UK, there is no longer any “standard,” and it is the responsibility of the examiner to satisfy themselves that the holistic information present justifies the decision that is made.

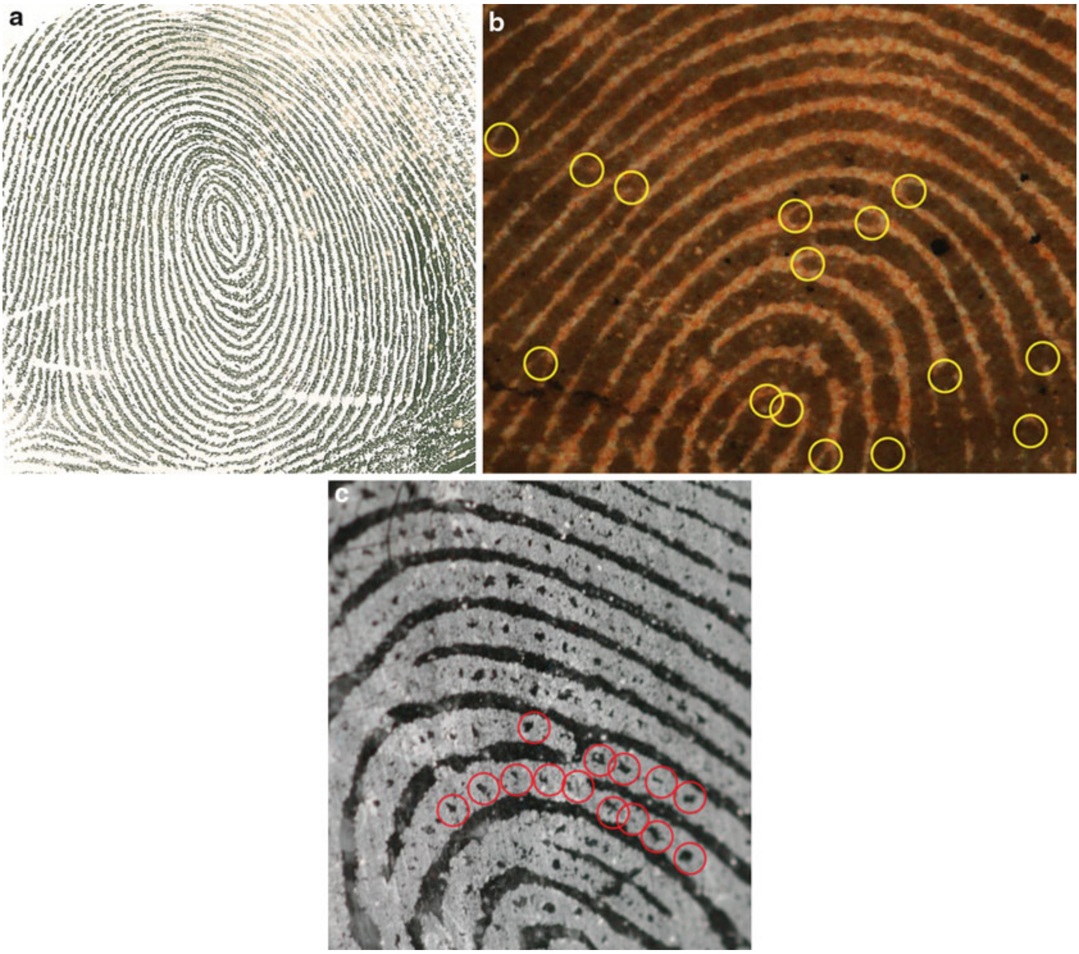

Fingerprint Identification, Fig. 10 Different levels of ridge detail within a fingerprint, (**a**) first-level detail, (**b**) second-level detail, (**c**) third-level detail

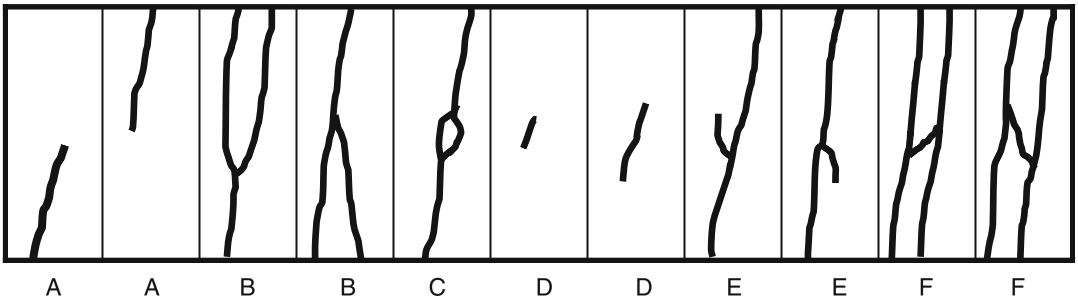

Fingerprint Identification, Fig. 11 Types of ridge detail

Automated Fingerprint Identification Systems

The role of automated fingerprint identification systems (AFIS) in the identification process is to provide the identification expert with a selection of candidate prints against which to compare the crime scene mark. It does not replace human opinion in the decision-making process. The algorithms used in AFIS convert images of fingerprints to wire drawings, identify the positions of level two features such as bifurcations and ridge endings, and store information relating to these locations and their spatial relationship to each other.

When a crime scene mark is input into an automated system, it is possible to "auto-encode" it, by allowing the database algorithms to go through the same conversion process and automatically mark the estimated position of ridge endings and bifurcations. In practice an expert will generally review this encoding and locate the positions of these level two features manually, removing any false features such as scratches in the image. The marked positions and spatial relationships of the level two features are then searched against those of the prints in the database, which displays the images associated with the most closely matching feature sets.

The examiner can then review the displayed images to select those that most closely correspond with the crime scene mark, and subject these prints to greater scrutiny in the ACE-V process. The human interpretation of the mark and print therefore remains essential to the decision-making process.

Scientific Validity of Fingerprint Examination and Human Error

Because the decision-making process during fingerprint examination depends on human opinion, the important question of the reliability of fingerprint examiners in making correct judgments when conducting fingerprint comparisons using ACE-V must be raised. Within the science of fingerprint examination, nobody really knows how often examiners make an erroneous decision. It is the examiners (and not necessarily the underlying biological science behind fingerprint identification) that are the potential weak link in the evidential chain.

Fingerprint examiners can make a false-positive judgment (an erroneous identification). However, there is also what is known as a false-negative decision (a missed identification). While the fingerprint profession prefers to weight the relative importance of such errors, preferring to insist that to have an "erroneous identification" is the most problematic mistake because a person has been wrongly brought into custody, one cannot ignore the "missed identification" errors, as they represent the majority of all known errors in the profession and as such are just as relevant because they relate to a failure in the human actor to apply methodology and skill to arrive at a correct conclusion. Whether the mistake is to erroneously identify a perpetrator of a crime or to erroneously exclude another person, the latter is potentially just as serious and no less a mistake. To wrongly exclude a perpetrator may leave him or her free to commit more crime, perhaps even to kill. One problem with knowing the true scale of the rate of error arising from a missed identification is that by the very nature of the error, a "miss" is only known about when circumstances permit that error being found. In so far as the erroneous identification has more potentially immediate and public ramifications, this has much to do with the rationale for the profession that regards errors in positive identification as being more serious than the missed identification. But in reality, both are technical mistakes arising from the same methodological failings.

The McKie case in Scotland (Campbell 2011) is perhaps one of the most notorious of all the recent controversies in latent print examination. Shirley McKie was arrested for perjury for stating under oath during a murder trial that a thumbprint, which was matched to her during routine elimination checks, was in fact not hers. McKie was vindicated after latent print examiners from other agencies around the world challenged the validity of the identification made by the Scottish Criminal Records Office (SCRO). To this day, there are experts in latent fingerprint analysis who still disagree as to whether it is a correct match or not. This inability to reach a consensus raises a number of fundamental issues associated with the reliability and repeatability of ACE-V as practiced by fingerprint examiners.

The erroneous fingerprint identification of Brandon Mayfield by FBI latent print examiners in response to the forensic investigation of the 2004 Madrid Bombings was another highly controversial case (Stacey 2004). Mayfield's fingerprints were alleged to have been identified against those found on a bag of detonators found in Spain after the bombings took place. However, subsequent reanalysis carried out after the Spanish authorities questioned the accuracy of the identification. The FBI fingerprint experts had to concede they had been incorrect in their original analysis.

These cases, as well as other errors and controversies, have resulted in fingerprint analysis coming under attack from both the judiciary and the academia. Some (Saks and Koehler 2005) have questioned the very underlying scientific assumptions made by fingerprint experts. The profession has built up over many decades a principle of infallibility around the methodology used to compare fingerprints that has delivered into the public domain a mantra that training and peer review through verification of findings are enough to ensure that erroneous identifications will not only be extremely rare, but that such occurrences will rarely if ever reach the courts.

Fingerprint examiners and law enforcement maintain that fingerprint evidence has unrivaled power within the judicial system. Fingerprint examiners also hold to a rigidity of thought around the clarity and unambiguousness of their work. In comparing fingerprints, there are limited choices an examiner can make. The only possible conclusions are it is a match, not a match, or there is insufficient detail upon which to make a conclusive decision (Bain 1985). However, the Evett/Williams report suggests that when many fingerprint experts use the term exact science (where conclusions are binary, it either is a match or it isn’t), this is a contradiction in terms, science is not exact (Evett and Williams 1995). This has now been endorsed by the recent findings of Sir Anthony Campbell who chaired the Fingerprint Inquiry (Campbell 2011) looking into the erroneous fingerprint identifications in the McKie case in which one of the key findings was that fingerprint experts should not use statements of fact over fingerprint identification, but rather should use statements of opinion instead. Without assessing every set of human prints in the world either alive now, or have ever lived, there can be no way of truly knowing whether fingerprints can ever be unique to the degree of having a categorical belief that two prints match one another.

Future Directions in Fingerprint Identification

Probabilistic Approaches

There has been a growing body of opinion that a statistical presentation of fingerprint evidence may be both valid and necessary (Champod and Evett 2001). For fingerprint evidence to be held up as an infallible science may be to overstate the power of the evidence and could, on occasions, misguide the pursuit of accurate and fair assessment of the case evidence by the jury.

Expressing fingerprint evidence in probabilistic terms would bring it in line with other types of evidence such as DNA, and may also enable probabilities of matches to be proposed for partial marks containing few features that would otherwise be disregarded. In probabilistic terms, examiners could testify that, for example, there was a one in a billion chance that the mark was left by someone other than the defendant.

This approach requires the position and spatial relationship of features to be marked on the crime scene mark, which would be compared to data on how often these feature combinations occur across large populations of fingerprints. Corresponding data is required on how the position and spatial relationship of these features is reproduced within a number of marks deposited by the same finger under different conditions. This enables the likelihood that the feature set in the crime scene mark was deposited by the finger of the suspect to be compared with the likelihood that it was deposited by any other person.

Probabilistic models have been developed by several different research groups, and work is underway to provide the supporting data sets to enable such analyses to be performed. It should, however, be noted that there is still much opposition from the fingerprint examiner community regarding the presentation of fingerprint evidence in this way.

Cognitive Bias

Examiners explore several different rhetorical strategies to appear to be demonstrating validity, while in fact demonstrating other things. These include "the fingerprint examiner's fallacy" (if the ACE-V methodology is followed, error cannot happen) and "the casework fallacy" (over 100 years of fingerprint examination only a handful of errors have been recorded). In essence, whenever an error is highlighted in the profession, it is invariably the case that the hierarchy blames the practitioner. It is the fault of the person doing the analysis and that the methodology is sound, even if human factors contributing to the error proved to be unavoidable and to a large extent unconscious to the individual. To date, it has been very rare for the profession to investigate possible factors other than practitioner incompetence for the occurrence of error. This is unfair to the practitioners and deflects away from the more serious underlying need to investigate the factors that contribute to error in the first place.

Validity research is rare in the fingerprint domain. Perhaps one of the reasons why validity studies are so rare is that there is a real fear of undermining the long-held core values and beliefs that underpin the dogma within the profession, and that to question the reliability of the method or to assess wider cognitive issues impacting upon practitioners is deemed both dangerous and by some, unnecessary. Fingerprint practitioners for the most part are simply not aware of the dearth of scientific support for their discipline's core assumptions.

There have been legal challenges for many years in the USA to the acceptance of fingerprint science as a valid methodology. In the landmark motion to exclude fingerprint testimony, (State of Maryland v. Bryan Rose K06-0545), the defendant contended that ACE-V was not a valid methodology which has been subjected to scientific testing and that the error rate in latent print identifications is unknown. It was asserted that in the absence of an error rate, the reliability of the methodology remained unproven. A fundamental problem, according to the defendant, was that the subjective comparisons in ACE-V involve psychological phenomena known as "confirmation bias." Further, the defendant argued that the "standards" for latent fingerprint identification are inadequately defined. The lack of error in certain fields of inquiry often derives from the nonexistence of methodological research into the problem and merely denotes a less advanced stage of that profession (Risinger et al. 2002).

There are examiners who are prepared to look at the issues of error rate and to investigate the methodology as well as human factors, and to suggest ways of mitigating potential weaknesses. Wertheim and Langenburg (Wertheim et al. 2006) showed in their research results that errors can and do occur. A study by Schiffer and Champod (Schiffer and Champod 2007) suggested that context may have an impact on practitioners dependent upon where in the ACE-V process the fingerprint analyzer found themselves. Latent print examinations are currently performed by human examiners, and human examiners are fallible, no matter in what domain they work. Therefore, the examinations are only as valid as the examiner performing the task and the conditions under which the tasks were conducted. Measuring and enhancing fingerprint expert performance is a challenging task. However, for any expert domain, be it medical experts, military fighter pilots, police officers, or forensic experts, one must study and understand the underlying expertise. Results obtained in law enforcement generally demonstrate that investigating police officers' judgments (this may include latent print examiners) are malleable and susceptible to influence from extralegal factors (Ainsworth 2005).

Enthusiasm to seek answers to these questions is gaining in strength (National Academy of Sciences 2009). The fingerprint community must be willing to subject established beliefs to verification and peer review (Busey et al. 2008).

There is now a substantial body of psychological research illustrating the tendency for human beings to engage in motivated reasoning or to be affected by contextual knowledge to a point where decision thresholds and even decision outcomes can be altered (Dror et al. 2006; Dror and Charlton 2007; Dror and Fraser-Mackenzie 2008). Such biases may be present in the work of latent print examiners. However, the extent to which cognitive bias creates the danger of erroneous interpretations in real-world circumstances has not yet received significant research attention. Continued research about the presence or absence of such biases in the latent print domain and the extent of any impact on accuracy and reliability is needed. However, given the decades-long research into the significant effects of cognitive bias in other domains, it would seem appropriate to minimize the potential for such biases in latent print interpretation now.

Given the genuine dangers of cognitive bias, examiners should be protected from inadvertent bias by shielding them from information that is clearly unnecessary and not relevant to their assessment. If examiners routinely receive extraneous, domain-irrelevant information, forensic service providers should examine whether they can modify their procedures to shield examiners from this unnecessary and potentially biasing information (Kaye et al. 2012).

Related Entries

- ▶ Biometrics and Border Control Policing
- ▶ Cognitive Forensics: Human Cognition, Contextual Information, and Bias
- ▶ DNA Profiling
- ▶ DNA Technology and Police Investigations
- ▶ Forensic Science
- ▶ Identification and the Development of Forensic Science
- ▶ Identification Technologies in Policing and Proof
- ▶ Philosophy of Forensic Identification
- ▶ Probability and Inference in Forensic Science

Recommended Reading and References

Ainsworth PB (2005) Psychology, law and eyewitness testimony. Wiley, West Sussex

Ashbaugh D (1999) Quantitative-Qualitative friction ridge analysis: an introduction to basic and advanced ridgeology. CRC Press, Boca Raton

Bain A (1985) The importance of fingerprints. Fingerpr Whorld 10:40

Beaudoin A (2004) New technique for revealing latent fingerprints on wet, porous surfaces: Oil Red O. J Forensic Identif 54(4):413–421

Bratton R, Gregus J (1997) Development of a black powder method to process adhesive tapes. Fingerpr Whorld 23(87):21–23

Busey T, Champod C, Langenburg G (2008) Amicus brief state of Maryland v Bryan Keith Rose. Case No. 03 K06000545

Campbell A (2011) The fingerprint inquiry Scotland. Crown copyright. Published on behalf of the fingerprint inquiry by APS Group Scotland. ISBN: 978-0-85759-002-2

Champod C, Evett IW (2001) A probabilistic approach to fingerprint evidence. J Forensic Identif 51(1):101–122

Creer KE (1995) The detection and enhancement of latent marks using specialised lighting and imaging techniques. Proceedings of the international symposium on fingerprint detection and identification, Ne'urim, Israel, 26-30 June

Dror IE, Charlton D (2007) Improving perception and judgment: an examination of expert performance. Fingerpr Whorld 33(129):232–234

Dror IE, Fraser-Mackenzie P (2008) Cognitive biases in human perception, judgment, and decision making: bridging theory and the real world. In: Rossmo K (ed) Criminal investigative failures. Taylor & Francis, Hoboken

Dror IE, Charlton D, Peron A (2006) Contextual information renders experts vulnerable to making erroneous identifications. Forensic Sci Int 156(1):74–78

Evett I, Williams RL (1995) A review of the 16 point fingerprint standard in England and Wales. Fingerpr Whorld 21:82

Galton F (1892) Fingerprints. Macmillan, New York

Hauze DB, Petrovskaia O, Taylor B, Joullie MM, Ramotowski R, Cantu AA (1998) 1,2-indandiones: new reagents for visualising the amino acid components of latent prints. J Forensic Sci 43(4):744–747

Kaye D et al. (2012) Latent print examination and human factors: improving the practice through a systems approach: the report of the expert working group on human factors in latent print analysis. NIST (In Print)

Kent T, Thomas GL, Reynoldson TE, East HW (1976) A vacuum coating technique for the development of latent fingerprints on polythene. J Forensic Sci Soc 16(2):93–101

Menzel ER (1999) Fingerprint detection with lasers. Marcel Dekker, New York

National Academy of Sciences (2009) Strengthening forensic science in the United States: a path forward. National Academy of Sciences, Washington, DC

Oden S, von Hofsten B (1954) Detection of fingerprints by the Ninhydrin reaction. Nature 173:449–450

Pounds CA, Grigg R, Mongkolaussavaratana T (1990) The use of 1,8-diazafluoren-9-one (DFO) for the fluorescent detection of latent fingerprints on paper. A preliminary investigation. J Forensic Sci 35(1):169–175

Ramotowski R (2000) A comparison of different physical developer systems and acid pre-treatments and their effects on developing latent prints. J Forensic Identif 50(4):363–383

Risinger DM, Saks MJ, Thompson WC, Rosenthal R (2002) The Daubert/Kumho implications of observer effects in forensic science: hidden problems of expectation and suggestion. Calif Law Rev 90(1):1–56
Saks M, Koehler J (2005) The coming paradigm shift in forensic identification. Science 309:825–892
Schiffer B, Champod C (2007) The potential (negative) influence of observational biases at the analysis stage of finger mark individualization. Forensic Sci Int 167:116–120
Schnetz B, Margot P (2001) Technical note: latent fingermarks, colloidal gold and multimetal deposition (MMD). Optimisation of the method. Forensic Sci Int 118:21–28
Sears VG, Butcher CPG, Fitzgerald LA (2005) Enhancement of fingerprints in blood – part 3 reactive techniques, acid yellow 7 and process sequences. J Forensic Identif 55(6):741–763
Stacey RB (2004) Report on the erroneous fingerprint individualization in the Madrid train bombing case. J Forensic Identif 54(6):706–718
Wargacki SP, Lewis LA, Dadmun MD (2007) Understanding the chemistry of the development of latent fingerprints by superglue fuming. J Forensic Sci 52(5):1057–1062
Wertheim K, Langenburg G, Moenssens A (2006) A report of latent print examiner accuracy during comparison training exercises. J Forensic Identif 56(1):55–93

Fingerprints

► Identification Technologies in Policing and Proof

Flogging

► History of Corporal Punishment

Focused Deterrence and “Pulling Levers”

Edmund F. McGarrell
School of Criminal Justice, Michigan State
University, East Lansing, MI, USA

Overview

The last few decades have witnessed considerable developments in both research and practice with respect to both offender- and place-oriented strategies aimed at reducing crime and violence. One such offender-oriented approach is based on the concept of focused deterrence and includes what is sometimes referred to as a pulling levers strategy. This entry describes the focused deterrence pulling levers strategy, reviews the research suggesting its capacity to affect levels of crime and violence, and poses a series of practical and research questions related to the model.

One of the most consistent findings of criminological research has been that a relatively small number of people and a relatively small number of places generate a disproportionately large amount of crime and calls-for-service to the police. The focus on recurring places such as street block segments and particular businesses has resulted in a variety of intervention strategies that research suggests have crime prevention and control efficacy. Similarly, recent years have witnessed the emergence of offender-oriented strategies that follow a focused deterrence logic model and demonstrate promise for violence and crime prevention and control. One particular strategy that has emerged has become known as the “pulling levers” approach.

Foundations

In the early to mid-1990s, officials in Boston began experimenting with a new approach to addressing the very serious issue of youth violence and youth gun violence in particular. Like many cities in the USA, the period coinciding with the crack cocaine epidemic generated elevated levels of homicide and gun crime, much of which involved youths in their teens and early twenties. To address this problem, Boston convened a multiagency working group consisting of federal, state, and local law enforcement; the US Attorney’s and the District Attorney’s Offices; probation and parole; street outreach workers; a coalition of faith leaders; and community leaders. Also included was a team of researchers from Harvard’s Kennedy School of Government. The working group agreed to follow a problem-solving approach that would begin with systematic analysis of the youth gun crime problem.

The analysis proved to be very insightful and ultimately shaped the offender-oriented pulling levers strategy. Specifically, gun violence in Boston was largely concentrated among a small group of chronically offending youth estimated at 1–3 % of the youth population. Further, these violence-involved youths were typically part of loosely organized gangs and street crews. Finally, these youthful offenders had long histories of involvement in the juvenile and adult criminal justice systems and homicide victims, and offenders often were mirror images of one another (Kennedy et al. 2001).

The strategy that emerged was based on these findings. First, the concentration of violence in a small subset of youth meant that a surgical intervention aimed at high-risk individuals would likely be more effective and generate fewer collateral costs than large-scale crackdowns affecting large proportions of the youth population. Second, extensive criminal histories meant that these individuals were subject to the sanctions and coercive power of the criminal justice system. For some this equated to incapacitative sentences; for others it meant that the criminal justice system had credible leverage over their behavior. Third, the group or network structure meant that it was possible that influencing a subset of the individuals involved in these social networks could have a broader impact on the larger group. Additionally, it might be possible to impose "group accountability."

The actual strategy involved delivering a highly focused deterrence message to youths involved in these violent street groups. Analysis would reveal groups currently believed to be involved in violence. Intelligence would identify members of the groups on probation and parole. These probationers and parolees would be ordered to attend what has come to be known as a "notification" or "call-in" meeting. At the meeting, a specific and focused deterrence message would be communicated to these individuals. Essentially, police, prosecutorial, and/or probation and parole staff would inform the attendees that officials know who they are, believe that they are at high risk for being involved in violence based on their behavior and their associations, and inform them that the violence must immediately stop or the task force will do everything possible, pull every lever, to try and get them off the streets. Attendees were reminded that as convicted felons they were prohibited from possessing or carrying a firearm and that there were severe sanctions in the federal system that would be imposed on felons in possession. For example, simple possession by a felon could carry a 5-year sentence in the federal system, along with no right to bail before trial and an anticipated time served of 85 % of the sentence. If the firearm was used as part of a drug trafficking operation or in the commission of a violent crime, much more substantial sentences were available in the federal system. Although there is significant variation across the states, typically the federal sanctions are more severe than those in the state court system. The message sought to gain credibility, and shift the perceived threat of sanction, by using examples of similar individuals often from the same network of street crews who had received stiff sentences.

An additional unique feature of the call-in meeting strategy was that the focused deterrence theme was coupled with two other key messages. The first involved the expression of the community voice. Typically community members would express their concern for the well-being of the attendees, their exasperation with the level of neighborhood violence and the loss of young people to violence and prisons, their hope that the attendees would go in another direction, and their support for the police and prosecutors. The community voice was considered important for "resetting" and clearly communicating community norms. The second key message was an offer of support from various social service providers. The social service providers and the community would typically describe various supports available (mentoring, housing, drug treatment, vocational and educational programs, work preparedness, etc.) should the meeting attendees seek assistance.

The community voice and the expression of social support were considered important in terms of increasing the perception of fairness and legitimacy of the overall stop the violence message. This was based on theory and research findings that compliance with the law increases with perceptions of fairness and legitimacy (Tyler 2003).

The Boston Ceasefire focused deterrence pulling levers approach has received significant attention for its apparent impact on youth homicide and shootings. A series of studies demonstrated that there was an approximate 63 % reduction in youth homicides and substantial declines in shots fired (−32 %) and gun assaults (−25 %) (Braga et al. 2001). These findings were contrasted with trends in other Massachusetts cities as well as major US cities, and Boston was distinct in terms of the timing and magnitude of the reduction.

The Diffusion of an Innovation

Although the peak of the homicide epidemic was ending, in many cities and for the country as a whole, homicide and gun homicide in particular remained at very high levels as news about the Boston Ceasefire program and its apparent impact began to spread. One of the first manifestations was US Department of Justice support for what became known as the Strategic Approaches to Community Safety Initiative or SACSI. SACSI provided technical support and research funding for ten participating cities. The sites received training in the Boston Ceasefire model including the research-based problem-solving process as well as the focused deterrence strategy involving offender notification meetings and the pulling levers approach. Overall, the SACSI evaluation team found that the SACSI cities experienced a reduction in violent crime compared to other US cities (Roehl et al. 2008).

One of the participating cities, Indianapolis, implemented an approach and strategy very similar to Boston Ceasefire. Employing a very similar analytic approach, the Indianapolis researchers found that after a crackdown on one violent street group and a series of offender notification meetings had occurred with probationers and parolees involved in gangs and violent street groups, homicide declined by approximately 34 %, and this decline was greater than that experienced by other similar size Midwestern cities (McGarrell et al. 2006). Subsequent analyses revealed that the reductions were highest for gang homicides (Corsaro and McGarrell 2009) and for the groups most at risk for homicide victimization, in the case of Indianapolis, this involved young African-American males in five specific hotspot areas (Corsaro and McGarrell 2010).

Another DOJ-funded site during the SACSI period, Los Angeles, followed a similar approach to address homicide and gun violence largely associated with gangs. Although the East Los Angeles site struggled with implementing all components of their strategy, they did implement a focused deterrence approach targeted at two rival gangs. Their evaluation found evidence of at least short-term declines in violence when comparing this target area to other parts of Los Angeles (Tita et al. 2003).

The positive findings from the SACSI jurisdictions seemingly had a policy impact as the Department of Justice developed a new initiative in 2001 that became known as Project Safe Neighborhoods (PSN). The PSN program incorporated elements of Boston Ceasefire and SACSI but also provided more flexibility to task forces operating out of each US Attorney's Office to adapt the strategy based on local problems and resources. Among the elements incorporated were the creation of multiagency teams, adoption of a problem-solving approach based on analysis of the local gun crime problem, and the incorporation of research partners into the PSN task forces. Two additional components related to Boston Ceasefire but also were influenced by an initiative developed in Richmond, Virginia, known as Project Exile. Whereas Boston had utilized selective federal prosecution of gun offenders as example for other potential felons, in the Exile program there was a commitment on part of the US Attorney's Office to prosecute as many illegal gun possession cases in the federal court as possible. Similarly, whereas Boston sought to communicate a deterrence message to at-risk individuals through call-in meetings, Richmond officials utilized a broad public education campaign involving multiple forms of media (radio, television, billboards, bus posters, etc.). PSN task forces were encouraged to use any and all of these strategies.

Several of the PSN sites followed the Boston Ceasefire model fairly closely and generated evaluation findings that suggested an impact. Although initiated before PSN, Stockton, California, worked with one of the original Boston researchers and followed a very similar approach focused primarily on gang homicide and violence. This study found a significant reduction in gun homicide that was much greater than that observed in eight comparable California cities (Braga 2008). Officials in Lowell, Massachusetts, followed the offender-oriented pulling levers approach to reduce youth violence. An interesting twist in this initiative was that Lowell officials applied an additional lever to adult offenders involved in illegal gaming activities with the goal of having the adults exert influence over youths involved in street violence. The results of an evaluation indicated a reduction in gang violence that was much greater than that observed in other Massachusetts cities (Braga et al. 2006). One of the most sophisticated evaluation designs was utilized in Chicago's PSN initiative that targeted specific geographic areas of the city. A twist on the pulling levers strategy was that Chicago included former inmates returning to the target areas in its offender notification meetings. The researchers found that homicide declined approximately 37 % in the PSN target areas. Further, they reported that the evidence suggested the most influential component of the strategy was the impact of the offender notification meetings (Papachristos et al. 2007).

Several additional jurisdictions, some involved in PSN and others distinct from PSN, also observed declines in homicide and/or gun crime following the implementation of the offender-oriented pulling levers strategy. Most of these studies employed limited designs involving pre-post trend analyses with either no comparison group or a weak comparison (e.g., remainder of city). However, the consistency of these findings from cities including Minneapolis, High Point, Baltimore, Greensboro, Omaha, and Winston-Salem was at least suggestive of impact on homicide and gun crime.

As would be anticipated given the latitude provided, there was considerable variation across PSN sites in terms of the implementation of the Ceasefire model generally and the pulling levers components specifically. All of the PSN task forces reported establishing new criminal justice partnerships although there was considerable variation in density and scope. All of the task forces received funds for a research partner. Officials from the majority of sites reported that the research partners brought value to the task force, though there were task forces where research was never integrated and it was difficult to discern a strategic problem-solving model. An estimated 45 % of the PSN task forces reported using offender notification meetings. The vast majority (98 %) reported increased federal prosecution of firearms cases and over 90 % utilized joint federal-local prosecutorial screening of gun cases, although the analysis revealed considerable variation across the 94 districts in terms of levels of federal prosecution of gun crimes. Over 40 % of the task forces reported other components of the Operation Ceasefire approach with respect to probation/parole home visits and firearm supply side interventions. Finally, approximately 58 % reported using directed patrols in hotspot locations and over 40 % utilized street-level firearms enforcement units, interventions that might be part of a "pulling levers" strategy (McGarrell et al. 2009).

Consequently, across the 94 PSN task forces, there were components of Boston's Operation Ceasefire's pulling levers strategy. The evaluators of PSN also found that there were significant differences across the task forces in the implementation dosage of three components: level of federal prosecution of gun crime, integration of research into the task force, and partnerships. Further, these implementation levels related to differences in the trend of violent crime with PSN target cities in high-dosage districts experiencing a significant decline in violent crime when compared to target cities at lower levels of dosage as well as with non-PSN target cities (McGarrell et al. 2010). Given that PSN was not implemented with a randomized design, the researchers were not able to exclude the possibility that some other characteristics of these high implementation cities produced the decline, but the results did appear to be consistent with the prior research on offender-based deterrence strategies.

As is evident in the above review, almost all of the research that has examined the offender-oriented pulling levers strategy has focused on the impact at an area level (either the entire city or particular neighborhoods or police districts). Two studies have included a focus at the individual level. The first is an ongoing study conducted by Papachristos and colleagues in Chicago. The results are not yet available. The second was an extension of the above-described Indianapolis initiative conducted by Chermak (2006). In 2002 Indianapolis conducted an experiment to assess the impact of two types of pulling levers meetings compared to traditional probation supervision. The first meeting was modeled on Boston Ceasefire and included the criminal justice deterrence message coupled with a message of opportunity provision. The second meeting was a type of meeting that had been developed by community groups who sought to deliver their own message urging participants to stop the violence and emphasizing the social support message. Probationers were randomly assigned to the three conditions. The population consisted of all felony probationers convicted of specific violent, drug, gun, and property offenses. There were a total of 540 probationers in the study with180 probationers in each group (Boston Ceasefire-style meeting, community support meeting, traditional probation). Interviews revealed that both types of meetings made an impression on the probationers and, consistent with the goal of the meetings, the Ceasefire meeting attendees were most likely to report that law enforcement was cracking down on violent crime and they were more likely to believe that illegal possession of a firearm was likely to result in stiff sanctions. Meeting attendees also reported that they discussed the message with their friends. Despite these differences in perception of message, no differences were found between the treatment groups and the control group on re-offending using both self-report and criminal history records. Caution was urged in interpreting these findings, however. There were two critical distinctions between the Boston Ceasefire and earlier Indianapolis pulling levers strategy and what occurred in the experiment. First, the selection of participants in the experiment was based on their probation status as opposed to their selection as being at high risk of violence based on their behavior and their group affiliations. Second, the study found that there was very little follow-up when violence occurred that may have involved meeting attendees and their associates. Similarly, there was very little follow-up in terms of delivery of service. Thus, although there was a threat of pulling levers and an offer of assistance, there was little actual implementation of either component.

Extensions of the Model Beyond Gun and Group Violence

As noted above, one of the first cities to implement the Boston Ceasefire pulling levers approach to addressing gun and gang violence was High Point, North Carolina. High Point followed virtually all of the components with a multiagency team of criminal justice, social service, and community partners, a research team, a data-driven problem-solving orientation, and the use of offender notification meetings. Having experienced what the police department and community partners considered success in addressing violent group crime, High Point extended the approach to addressing other problems. This, in turn, has resulted in innovative adaptation of the pulling levers approach in other communities as well.

On the basis of continuous monitoring and analysis of crime patterns, the High Point Police Department (HPPD) decided that the next major problem driving crime, violence, and disorder was overt drug markets. The analysis suggested that drug market activity was creating such problems in specific neighborhoods. Detailed analysis, street-level intelligence, and subsequent undercover operations revealed that a relatively small group of offenders were responsible for these discrete drug markets. Specifically, in the initial four markets targeted for the pulling levers approach, the number of identified dealers ranged from a low of 13 to a high of 26. This finding has since been replicated in over 20 communities where the number of individuals involved in the drug market has ranged from a low of 7 to a high of 59 and an average of 27 (Hipple et al. 2009). This manageable and specific network of street drug dealers suggested that the pulling levers approach had merit.

The main distinction between the drug market approach and the gang and group-based approach was that many of the dealers were not on probation or parole so the question arose of how to get them to attend the offender notification call-in meeting. The strategy that emerged involved a traditional undercover operation in which undercover buys would be made from all, or as many as possible, of the individuals believed to be involved in the specific market.

Following the undercover operation, police and prosecutors convened to decide who would be subject to aggressive prosecution and who would be called-in to a meeting and offered a second chance. The criteria agreed to by the High Point team were based on prior involvement in violence as well as extensive criminal histories for drug selling and other serious offenses. Individuals with histories of violence and gun use as well as chronic offenders were prosecuted as would normally occur. These individuals comprised the "A list." They would serve as examples for those diverted from prosecution. The diverted individuals, without histories of violence, were often the retail street-level sellers and lookouts who could be expected to continue to sell and to "backfill" for the drug-selling network as others were removed from the community. This group comprised the "B list."

Adopting the same approach as developed in Operation Ceasefire, the B list group was invited to attend a call-in session along with "influentials" in their lives. Here they were presented with the evidence against them, often including video of their drug sales, and informed that the drug market was closed. Continued involvement in drug selling and other crimes would result in arrest and prosecution similar to that of the A list. The meetings included opportunities for the community to express their exasperation with drug dealing and its impact on the neighborhood, their support for the criminal justice authorities, and their support for the individuals attending the meeting and being offered a second chance. The meeting also included social service representatives describing available services and offering support to the attendees.

Following the call-in meeting the police would provide an increased presence for a short period of time to send the message that dealing would no longer be tolerated. Substantial effort went into community engagement with the idea that the community would ultimately be responsible for asserting informal social control that would preclude the reemergence of the drug market (Kennedy 2009).

An evaluation of the High Point experience with the implementation of what has come to be known as the Drug Market Intervention (DMI) in the initial four neighborhoods found evidence of impact on violent crime. Using trajectory models that compared chronic high crime street blocks in the DMI target areas with comparable high crime street blocks in the remainder of the city, there was a significant decline in long-term violent crime trends in the DMI areas (Corsaro et al. 2012).

Given the frustration with traditional drug enforcement efforts, the High Point DMI has generated significant interest. The Bureau of Justice Assistance has supported training and technical assistance on the DMI model to 24 communities and at least 33 communities have implemented or are in the process of implementing the DMI approach. In addition to the High Point evaluation, studies have been completed in Rockford, Illinois, and Nashville, Tennessee. Both studies compared the pre-post trend in crime in the drug market target area with the trend in crime for the rest of the city. In the case of Rockford, there was a large decline in violent, property and drug crime following implementation. A careful evaluation utilizing growth curve modeling indicated that the target area had a statistically significant decline in property crimes relative to the rest of the city and a decline, though not statistically significant, in violent crimes (Corsaro et al. 2009). In Nashville, there were large declines in all offenses, although when compared to trends in the remainder of the city and controlling for preexisting trends, only the reduction on drug crime was found to be statistically significant (Corsaro et al. 2010). In both Rockford and Nashville, interviews with residents of the target neighborhood suggested perceived improvements in the quality of life, levels of crime, and support for the police. Given these promising findings the National Institute of Justice awarded the RAND Corporation a grant to carefully evaluate the DMI model. That study is ongoing.

High Point also extended this approach to the problem of street robbery when they experienced an increase in street robberies. Crime analysis and street-level intelligence identified groups involved in these robberies. Individuals involved in the groups under probation or parole supervision were ordered into very similar call-in meetings. Although no formal evaluation was conducted, HPPD officials report that the spike in street robberies declined soon after these call-in sessions.

An additional adaptation of the offender-oriented, focused deterrence model has involved former inmates returning to the community. Indianapolis was the first known community to implement this approach. The model was based on findings that a significant proportion of homicide victims and offenders had previous prison and parole experience. The meetings were organized geographically so that a group of returning former inmates from the same part of the city would be invited to a call-in meeting. A pilot evaluation of the project compared recidivism for former inmates attending a call-in meeting with similar former inmates in other neighborhoods who were not afforded this treatment. The results indicated a potential delay in the time to re-offending, although the results were not statistically significant in a multivariate analysis (McGarrell et al. 2003).

The most promising results suggesting the potential use of call-in meetings with returning former inmates comes from the previously mentioned Chicago PSN program (Papachristos et al. 2007). This program included returning parolees in the PSN target areas, and these areas experienced very significant declines in homicides and gun crime. As noted above, there is an ongoing analysis that is assessing the impact of the meetings on individual re-offending. A similar approach in Boston found very positive results in terms of reduced re-offending (Braga et al. 2009).

Finally, although it has not been evaluated, another variation of this approach in the reentry context comes from the PSN program in Dallas. Very large numbers of probationers and parolees (N = 400–500) from throughout the city are invited to an offender notification meeting. The message delivered is very similar to that of the smaller call-in meeting and includes the deterrent threat, the "felons cannot possess a gun" message, and the offer of social support.

Summary of the Research Findings

In the late 1990s when the results from Boston Ceasefire first came to the attention of the research community, the findings seemed too good to be true. A 60 % reduction in youth homicide and a city going two and a half years without a youth homicide caught the attention of researchers, policymakers, and law enforcement leaders. Yet it stood as one city's experience. Since that time a number of other studies of the group-based violence reduction strategy have been completed from cities including Baltimore, Chicago, Greensboro, High Point, Indianapolis, Los Angeles, Lowell, Minneapolis, Omaha, Stockton, and Winston-Salem. The evaluations associated with these efforts have varied in terms of approach and methodological rigor but have been consistent with an interpretation of the pulling levers approach having a desired impact on levels of homicide and gun violence. Although the research base is much more limited, the studies of the pulling levers extension to drug markets conducted in High Point, Nashville, and Rockford are also encouraging, at least for some types of offending. The results from the PSN research are also encouraging although more difficult to interpret given the multiple components of the interventions and the variation across cities.

Importantly, a systematic review of the focused deterrence model has been conducted as part of the Campbell Collaboration. This review was based on ten quasi-experiments and one randomized controlled trial. Although noting the need for more rigorous testing, the authors found an "overall statistically significant, medium-sized crime reduction effect" for these focused deterrence strategies (Braga and Weisburd 2012: 324). At a minimum, this series of studies suggests at least a short-term impact on lethal violence at the community level. Given the human, social, and fiscal costs associated with homicide and gun violence, these are important findings.

The research assessing the impact at the individual level is even more limited and leaves in question whether the pulling levers approach can significantly alter individual level patterns of behavior.

Limitations and Issues for Further Study

Despite this promising evidence, there remains uncertainty and perhaps skepticism that the pulling levers strategy has been the cause of the reductions that have been observed in these studies. One of the reasons for this skepticism is that the period during which this research has occurred has witnessed an overall decline in homicide and violent crime since the peak levels of the early 1990s. Given that the strategy was often implemented when a particular jurisdiction was experiencing a "homicide problem," the question arises whether there was a "regression to the mean" coupled with the long-term reduction in violence observed throughout the country. A number of the studies employed statistical techniques that are designed to minimize the impact of preexisting trends in the crime patterns but that cannot completely rule out the possibility that the decline would have occurred absent the intervention.

These concerns relate to the limitations of the evaluation designs that have been employed in the various studies. Although a number of the studies have employed quasi-experimental designs that provide a comparison group, they have not involved a true experimental design that can eliminate the threat that the decline in crime is based on some unmeasured factor rather than the pulling levers strategy.

Questions also arise due to the complex nature of the intervention. For example, one study comparing three strategies that emerged in the mid-1990s, Boston Ceasefire, Richmond Exile, and New York City's Compstat, found evidence that both Ceasefire and Exile were associated with reductions in violence (Rosenfeld et al. 2005). Similar results emerged in Exile-type strategies in Mobile and Montgomery, Alabama (Hipple 2010). Given that both Ceasefire and Exile included the threat of federal prosecution for illegal gun possession and use, it becomes difficult to disentangle the impact of the role of focused federal prosecution of gun crime from the impact of the pulling levers meetings.

Finally, a common threat to interpreting a series of studies like those described above is "publication bias." It is more likely that evaluations finding an impact for a strategy like pulling levers will be written up and published than evaluations that did not find an impact. Given the attention to the group-based pulling levers and Drug Market Intervention programs, it is possible that jurisdictions have implemented these strategies without an evaluation component or where the findings were ambiguous and did not show an impact and thus never made it to the research community.

In addition to these methodological questions, a series of theoretical issues arise. The theoretical foundations of the pulling levers strategy are multifaceted and include focused deterrence, procedural justice, social support, and community coproduction of order (Kennedy 2009).

The focused deterrence component is based on the idea of a small group of individuals being responsible for a large amount of crime and being subject to a direct message that may shift their perceived risk of sanction. The procedural justice component is based on research that finds that compliance with the law is most likely when individuals believe that they have been treated fairly and with respect (Tyler 2003). Social support links to the procedural justice component but also may include psychological, social, and tangible resources that may reduce strain or increase incentives to comply with the law. It may also involve informal social control as those most influential with the target population are included in the process. This aspect of informal social control includes both the network associated with individuals deemed at risk and the role of the broader community in coproducing order. This component involves collaboration with individuals, networks, and institutions in the community to increase the collective efficacy within the neighborhood (Sampson et al. 1997).

The logic model of the pulling levers approach builds on these multiple theories (see Rivers et al. 2012), yet there is very little research actually testing these theoretical components in the context of the offender-focused pulling levers strategy.

In addition to further research on the theoretical foundations of the offender-oriented, focused deterrence model, there are several areas deserving of further investigation.

As noted above, one of the lingering questions is the impact of the offender notification meetings on those selected to attend. A parallel question is the impact on their associates. From a human capital standpoint, the hope is that the meetings will result in reduced levels of offending, reduced levels of serious and violent offending, and reduced victimization among meeting attendees and their associates. At this point the evidence simply does not exist to answer whether this is the case. Having said this, if research continues to point to significantly reduced levels of community violence, then the approach may be considered successful even if there is minimal impact at the individual level of meeting attendees.

A second question relates to the sustainability of the violence reduction. Other types of police crackdowns have generated short-term declines but that were not sustained over time (Sherman 1990). Both Boston and Indianapolis experienced increases in homicide years after their initial findings of reduced violence. In contrast, High Point officials report sustained declines. At this point, there is insufficient research evidence to clarify whether implementation of the offender-oriented, focused deterrence strategy can yield long-term sustained reductions. There appear to be two sets of questions related to sustainability. The first is theoretical. Can this approach produce a long-term shift in perception about the relative costs of criminal justice sanctions and victimization risk, or are the findings to date driven by short-term impact that decays in the period after the offender notification meetings? The second question relates to the system capacity to sustain the efforts. By their very nature, the multiagency partnerships are fragile and can suffer from turnover at both the leadership and implementation levels. Responsibilities for intelligence gathering, undercover operations, organizing call-in meetings, and community collaboration are diffused across organizations and often are “add-ons” to the normal responsibilities of the various actors. Action may be driven by a sense of crisis, such as that generated by high levels of youth homicide in Boston and record levels of homicide in Indianapolis, but may be difficult to sustain when peak violence levels decline. Thus, research is needed both on the capacity of the model to have long-term impact and on the system capacity to sustain these strategic, data-driven, multiagency coalitions.

As noted above, prior studies found evidence that both Boston Ceasefire and Project Exile were associated with declines in homicide and violent crime (Rosenfeld et al. 2005). The PSN research found that case studies consistent with both approaches (Ceasefire and Exile) showed evidence of violence reduction. Both strategies include a focused deterrence message supported by the actual and threatened use of federal prosecution for very select behaviors (illegal gun possession and use). From this point the approaches diverge with Boston Ceasefire relying on direct communication of the message to at-risk groups and Exile using a broad public media campaign to communicate the message. Boston Ceasefire also placed greater emphasis on community engagement. The existing research, however, cannot distinguish which of these various components are critical for impact. Is the threat of federal prosecution a necessary ingredient? Is the direct communication of the message in face-to-face meetings the necessary ingredient? Does the community actually change in terms of increased collective efficacy and informal social control, and if so, is this a critical ingredient in violence reduction? Is it the combination of deterrence threat and social support that generates the impact, or can one or the other component have a similar impact alone? How do all of these components relate to the sustainability issue? Ultimately, multisite investigations with longer-term follow-up are needed to address these types of issues and to inform policymakers and professionals on the key elements needed to maximize violence reduction.

The focused deterrence approach is largely a people-based strategy. During the same years that these models have emerged, there has also been progress in the development of place-based strategies (Braga and Weisburd 2010). Both people- and place-based approaches are based on principles of situational crime prevention (Clarke 1997). Given that violence is highly concentrated in terms of both people and places, an extension of the offender-oriented focused deterrence model might attempt to link this people-based strategy with a variety of place-based interventions. In reality this may already be occurring. For example, Rockford, Illinois, included a place-based intervention focused on problem properties following its initial call-in meeting with individuals involved in street-level drug dealing. Similarly, Milwaukee, Wisconsin, and Roanoke, Virginia, conducted code enforcement surveys and landlord education sessions along with their offender call-in meetings. Although this further complicates evaluation issues, it does seem to offer a promising set of interventions based on research understanding of crime patterns (Tillyer and Kennedy 2008).

Summary

The National Academy of Sciences review of strategies intended to reduce violent crime noted that Boston Ceasefire represented a promising intervention that deserved further research (Wellford et al. 2005). The accumulation of evidence of the impact of the offender-oriented, focused deterrence model since the initial findings of Boston Ceasefire warrants at least a label of "promising," if not "evidence-based," practice. For police and criminal justice officials and their various community partners, these strategies offer an approach to reducing violence that offers significant promise for reducing high levels of violence, at least in the short-term. Further, there is some evidence that the greatest impact will be on those communities that suffer the most from high levels of violence. For scholars, it is an approach that also offers a host of unanswered research questions. Answering these questions will not only advance theoretical knowledge about violence and violence prevention but will also likely contribute to practice in ways that further maximize the likelihood of violence reduction.

Related Entries

- ▶ Certainty, Severity, and Their Deterrent Effects
- ▶ Community Policing
- ▶ Deterrence
- ▶ Deterrence: Actual Versus Perceived Risk of Punishment
- ▶ Gangs and Social Networks
- ▶ Police Use of Firearms

- ▶ Problem-Oriented Policing
- ▶ Procedural Justice, Legitimacy, and Policing

Recommended Reading and References

Braga AA (2008) Pulling levers focused deterrence strategies and the prevention of gun homicide. J Crim Just 36:332–343

Braga AA, Weisburd DL (2010) Policing problem places: crime hot spots and effective prevention. Oxford University Press, New York

Braga AA, Weisburd DL (2012) The effects of focused deterrence strategies on crime: a systematic review and meta-analysis of the empirical evidence. J Res Crime Delinq 49(3):323–358

Braga AA, Kennedy DM, Waring EJ, Piehl AM (2001) Problem-oriented policing, deterrence, and youth violence: an evaluation of Boston's operation ceasefire. J Res Crime Delinq 38:195–225

Braga AA, McDevitt J, Pierce GL (2006) Understanding and preventing gang violence: problem analysis and response development in Lowell, Massachusetts. Police Q 9:20–46

Braga AA, Piehl AM, Hureau D (2009) Controlling violent offenders released to the community: an evaluation of the Boston reentry initiative. J Res Crime Delinq 46:411–436

Chermak S (2006) Reducing violent crime and firearms violence: the Indianapolis lever-pulling experiment. Final Report Submitted to the National Institute of Justice

Clarke R (1997) Situational crime prevention: successful case studies. Harrow and Heston, Albany

Corsaro N, McGarrell EF (2009) Testing a promising homicide reduction strategy: re-assessing the impact of the Indianapolis 'Pulling Levers' intervention. J Exp Criminol 5:63–82

Corsaro N, McGarrell EF (2010) Reducing homicide risk in Indianapolis between 1997 and 2000. J Urban Health 87(5):851–864

Corsaro N, Brunson R, McGarrell EF (2009) Problem-oriented policing and open-air drug markets: examining the pulling levers deterrence strategy in Rockford, Illinois. Crime and Delinquency. doi:10.1177/0011128709345955

Corsaro N, Brunson RK, McGarrell EF (2010) Evaluating a policing strategy intended to disrupt an illicit street-level drug market. Eval Rev 34(6):513–548

Corsaro N, Hunt ED, Hipple NK, McGarrell EF (2012) The impact of drug market pulling levers policing on neighborhood violence: an evaluation of the high point drug market intervention. Criminol Public Policy 11(2):167–199

Hipple NK (2010) Project exile gun crime reduction. In: Klofas J, Hipple NK, McGarrell EF (eds) The new criminal justice. Routledge, New York, pp 51–58

Hipple NK, McGarrell EF, Bynum TS, Perez HA, Corsaro N, Garmo M (2009) Drug market intervention: implementation guide and lessons learned. Working Paper. School of Criminal Justice, Michigan State University, East Lansing

Kennedy DM (2009) Deterrence and crime prevention: reconsidering the prospect of sanction. Routledge, New York

Kennedy DM, Braga AA, Piehl AM (2001) Reducing gun violence: the Boston gun project's operation ceasefire. Part I: Developing and implementing operation ceasefire. U.S. Department of Justice, Office of Justice Programs, National Institute of Justice, Washington, DC

McGarrell EF, Hipple N, Banks D (2003) Community meetings as a tool in inmate re-entry. Justice Research and Policy 5(2):5–32

McGarrell EF, Chermak S, Wilson JM, Corsaro N (2006) Reducing homicide through a "Lever-Pulling" strategy. Just Q 23:214–231

McGarrell EF, Hipple NK, Corsaro N, Bynum TS, Perez H, Zimmermann CA, Garmo M (2009) Project safe neighborhoods: a national program to reduce gun crime: final project report. Submitted to the National Institute of Justice. Accessed 28 February 2009 from http://www.ncjrs.gov/pdffiles1/nij/grants/226686.pdf

McGarrell EF, Corsaro N, Hipple NK, Bynum TS (2010) Project safe neighborhoods and violent crime trends in U.S. cities: assessing violent crime impact. J Quant Criminol 26:165–190

Papachristos AV, Meares TL, Fagan J (2007) Attention felons: evaluating project safe neighborhoods in Chicago. J Empirical Legal Stud 4:223–272

Rivers L, Norris A, McGarrell EF (2012) Mental model of the drug market intervention. Journal of Drug Issues 42:59–81

Roehl J, Rosenbaum D, Costello S, Coldren J, Schuck A, Kunard L, Forde D (2008) Paving the way for project safe neighborhoods: SACSI in 10 U.S. Cities. National Institute of Justice Research in Brief April:1–17

Rosenfeld R, Fornango R, Baumer E (2005) Did ceasefire, compstat, and exile reduce homicide. Criminol Public Policy 4:419–450

Sampson RJ, Raudebush SW, Earls F (1997) Neighborhoods and violent crime: a multilevel study of collective efficacy. Science 277:918–923

Sherman LW (1990) Police crackdowns: initial and residual deterrence. Crime and Justice: A Review of Research 12:1–48

Tillyer MS, Kennedy DM (2008) Locating focused deterrence approaches within a situational crime prevention framework. Crime Prevention and Community Safety 10:75–84

Tita G, Riley KJ, Ridgeway G, Grammich C, Abrahamse A, Greenwood P (2003) Reducing gun violence: results from an intervention in East Los Angeles. RAND Corporation, Santa Monica

Tyler TR (2003) Procedural justice, legitimacy, and the effective rule of law. Crime and Justice: A Review of Research 30:431–505

Wellford C, Pepper JV, Petrie C (eds) (2005) Firearms and violence: a critical review. Committee to improve research information and data on firearms. National Academies Press, Washington, DC

Force

► Police and the Excessive Use of Force

Forced Migration and Human Rights

Wenona Rymond-Richmond
Department of Sociology, University of
Massachusetts Amherst, Amherst, MA, USA

Overview

Being forced to leave one's home or homeland constitutes a gross violation of human rights. Tragically, over 60 million people worldwide have been forcibly removed from their homes. The causes of forced displacement are diverse and include natural or manmade disasters, human trafficking, war and civil war, genocide, development projects, evictions, and political, social, ethnic, or religious persecution. According to recent figures published by the United Nations High Commission for Refugees (UNHCR 2010), over 43.3 million people worldwide have been forcibly displaced due to conflict and persecution.

The sheer enormity of these numbers should imply that forced displacement is central to discussions of human rights. Yet, forced displacement has not received the same attention as other human rights violations in public discourse in academia and "often seems to be accepted as a sad but inevitable consequence of war" (Hollenbach 2008: 2). Furthermore, the study of refugees is often overlooked in academia. Demonstrating the slow pace in which academics focused on refugees, the interdisciplinary field of "refugee studies" did not emerge until the early 1980s. Further, it was not until the late 1990s that a significant number of academic reviews of internally displaced people (IDP), individuals who are forcibly displaced and remain inside their original country, were published.

Criminologists should have much to contribute to the study of forced displacement and the relationship between crime, human rights, and social justice. However, the discipline of criminology has been remarkably unresponsive to examining the role of the state in acts of persecution and crimes committed by the state more generally. Inherent to the definition of refugee is recognizing either a state's crime against its citizens, the state's inability to protect its citizens, or the state's willingness to turn a blind eye on crimes committed against its citizens. For the most part, the discipline of criminology has failed to incorporate forced displacement into their research agenda. Instead, criminologists are primarily concerned with interpersonal acts of violence such as homicide, rape, and robbery. In other words, the focus of mainstream criminology is predominately the individual or possibly the neighborhood rather than the state as a unit of analysis. One notable exception is critical criminologists who are attentive to state crimes and crimes of the powerful more generally. Similar to criminologists silence on the topic of genocide is their silence on the topic of forced displacement (Hagan and Rymond-Richmond 2008).

Causes of Forced Displacement

According to recent figures published by the United Nations High Commission for Refugees (UNHCR 2010), over 43 million people worldwide have been forcibly displaced due to conflict and persecution. The causes of forced displacement are diverse and include natural or manmade disasters, human trafficking, war and civil war, genocide, development projects, evictions, and political, social, ethnic, or religious persecution. The focus of this entry is forced displacement due to conflict and persecution rather than forced displacement due to natural or manmade disasters.

While evictions might not initially register to some as a form of displacement resulting from conflict and persecution, human rights violations may occur during the process of urban redevelopment. Forced displacement and human rights violation can occur when individuals or families are forcibly evicted. This may occur because of urban redevelopment, "beautification" projects, gentrification, or as a side effect of development. Like other displaced people, those who are forcibly evicted may experience trauma from being forced out of one's home and or community. In recent years, the United Nations human rights program has become more attentive to examining human rights violations that occur in the practice of forced evictions. One example from the United Nations in regards to human rights violations and evictions comes from Miloon Kothari, the United Nation's highest-ranking expert on housing issues. After visiting a public housing development in Chicago that was slated to undergo demolition and redevelopment, Kothari stated that "evictions of public housing residents in the United States clearly violate international human rights, including the International Convention on the Elimination of All Forms of Racial Discrimination and the International Covenant on Economic, Social and Cultural Rights" (Hagedorn and Rauch 2007: 451). Under these guidelines, some public housing residents could qualify as internally displaced people (ibid).

Definition of Immigrants, Refugees, and Internally Displaced Persons

Defining and classifying who is and who is not a refugee has serious consequences for individuals and families. For some, being defined as a refugee can be a matter of life and death because with refugee status comes international protection. The legal definition of refugee was established in the 1951 United Nations Convention Relating to the Status of Refugees and its 1967 Protocol:

> Any person who owing to well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion, is outside the country of his nationality and is unable, or owing to such fear, is unwilling to avail himself of the protection of that country; or who, not having a nationality and being outside the country of his former habitual residence, is unable, or owing to such fear, is unwilling to return to it.

There are significant critiques and debates regarding the definition of refugee. First, is a dispute regarding whether immigrants and refugees are fundamentally distinct categories. Refugees and immigrants share common disadvantages including high levels of poverty and discrimination in their home country. Nonetheless, some argue that significant differences exist. One the one hand is the realist perspective, which suggests that the groups are distinguished by whether the migration was by choice or not. For refugees the migration is not by choice but due to social, cultural, or political pressures, which forced them out of their native country. For immigrants, the migration is by choice and typically for economic opportunities. Additional distinctions are made between the categories based on their likely ability to return to one's homeland and their relationship to the state (see Kunz 1973). On the other hand is the nominalist perspective, which suggests that the binary categories are blurry, are socially constructed, used to advance state interests, and mask the similarities between immigrants and refugees.

Among the debates regarding the definition of refugee is whether the definition is too broad or too narrow. Those who find the definition to be too broad emphasize that individuals defined as refugees have exceptionally diverse historical and political origins. Individuals' cultural, psychological, economic, spiritual, and personal differences may be minimalized as they are collectively subsumed under the label of refugee. Others assert that the definition is too narrow because the focus is on individual persecution rather than generalized persecution resulting from violence and war. Two legal documents that attempt to rectify this limitation are the 1969 Organization of African Unity Convention Governing the Specific Aspects of Refugee Problems in Africa and the 1984 Cartegena Declaration on Refugees. Both of these regional legal instruments broaden the definition of refugee to include persons who "have been threatened by generalized violence" (1984 Cartegena Declaration on Refugees).

The legal definition of refugee is further critiqued for being too narrow because it excludes individuals who were displaced from their homes for similar reasons as refugees, but have not left their country of origin and crossed an international boarder. The term used for this exceptionally large group of individuals is internally displaced persons (IDP). Approximately 25–30 million IDP's were displaced because of conflict. While there is no legal definition of IDPs, the Guiding Principles on Internal Displacement provides the following definition:

> internally displaced persons are persons or groups of persons who have been forced or obliged to flee or to leave their homes or places of habitual residence, in particular as a result of or in order to avoid the effects of armed conflict, situations of generalized violence, violations of human rights or natural or human-made disasters, and who have not crossed an internationally recognized State border.

Rather than adjudicate between these positions, the purpose here is to highlight controversies surrounding the definition of immigrants, refugees, and internally displaced persons and to examine human rights violations of forced migration.

Legal Violations and Rights of Forcibly Displaced People

The rights to freedom of movement, rights to choose one's residence, rights to security of the person, rights to life, and rights to freedom of expression are recognized in many international laws and national constitutions. In particular, forced displacement violates a number of international laws including the Universal Declaration of Human Rights (art.25, para.1), the International Covenant on Economic, Social and Cultural Rights (art. 11, para. 1), the International Covenant on Economic, Social and Cultural Rights, and the 1949 Geneva Conventions and the Additional Protocols.

Several international legal documents outline the rights of migrants, refugees, and internally displaced people. The 1951 United Nations Convention Relating to the Status of Refugees is the landmark legal instrument to protect and assist refugees. Additional documents supporting the human rights of migrants, refugees, and internally displaced people include the 1967 Protocol Relating to the Status of Refugees; the 1969 OAU Convention Governing the Specific Aspects of Refugee Problems in Africa; and the 1990 International Convention on Migrant Rights, which established that all migrant workers, whether they are documented or not, have fundamental human rights. Furthermore, the UN Pinheiro Principles detail the rights individuals have regarding returning home after a violent conflict. According to this document, individuals not only have the right to return home, but they also have the right to return to the same property. The aim of this principle is to ensure that no one profits from the violence and to reduce the amount of victimization that may occur after being forcibly displaced.

Status of Forcible Displacement: Refugees

Refugees

Accurately counting the number of refugees is difficult. According to the UNHCR, refugees totaled approximately 10.4 million refugee in 2009 and 10.55 million in 2010. The UNHCR's count of refugees is problematic because it does not include the large number of individuals displaced in urban cities around the world. Furthermore, the number does not include the more than 25 million individuals that are internally displaced.

Refugees are located throughout the world. The largest numbers are located in the Middle East and North Africa. There are approximately 2.7 million refugees in African, and 2.5 million refugees are in South and Central Asia. Refugees number nearly 650,000 in America and the Caribbean and nearly half a million in Europe. Pakistan hosted the highest number of refugees in 2010. By the end of 2010, the UNHCR estimates that the largest country of origin of refugees was

Afghanistan (3.05 million), followed by Iraq (1.7 million), Somalia (770,000), the Democratic Republic of the Congo (477,000), and Myanmar (416,000).

Approximately one-third of refugees currently live in a camp. Following larger population trends, since the 1950s, refugees are increasing settling in urban areas and mostly in developed countries. More than half of all refugees now reside in urban areas, and nearly half of refugees are women (47 %) and children (47 %).

IDP

In 2010, the number of people internally displaced by human rights violations, violence, and armed conflict totaled 27.5 million, the highest number in 10 years. Estimates are that 11.2–13.7 million children by the end of 2010 were internally displaced as a result of human rights violations, violence, and armed conflict. Approximately 80 % of all internally displaced individuals are women and children. Like refugees, internally displaced persons (IDPs) reside in countries throughout the world. The largest numbers are located in Columbia, Democratic Republic of Congo (DRC), Iraq, Somalia, and Sudan. In each of these countries, the IDP population was over one million. In 1 year alone, 2010, it is estimated that at least 2.9 million people in 20 countries were newly displaced. At least 200,000 IDPs in 2010 were displaced in the countries of Columbia, DRC, Kyrgyzstan, Pakistan, Somalia, and Sudan (IDMC 2010).

Current Issues and Controversies

1. Lack of Protection for Internally Displaced Individuals

   The UNHCR is responsible for protecting and assisting refugees; however, there is no comparable organization within the United Nations for internally displaced persons. With little to no international legal protection, internally displaced individuals often suffer greatly at the hands of their government. In 2010, attacks or threats against humanitarian workers occurred in Darfur, Afghanistan, Yemen, Somalia, and Columbia (IDMC 2010). In Darfur, the Sudanese government has prevented delivery of medical care and humanitarian aid to IDP camps (see ICC 2007; Totten 2006).

2. Refugees in Urban Areas

   More than half of all refugees currently reside in urban areas. Rather than be displaced in a refugee camp, many individuals select to be displaced in a city. UN High Commissioner for Refugees Antonio Guterres argues that their complex and unique challenges requires a radically new approach to providing protection and assistance. The primary concern is that internally displaced people receive fewer benefits if they live in urban areas than if they live in camps. Authorities often do not recognize them as displaced people and therefore do not provide them with services they are legally entitled to.

3. Protracted Displacement

   Returning home for a refugee is uncertain and typically involves a lengthy stay in a refugee camp. Displaced people may be displaced for years of even for decades. Nearly two-thirds of all refugees are in a protracted refugee situation, which is defined by the UNHCR as being in exile for more than 5 years. Like refugees, IDPs suffer from lengthy displacement. The Internal Displacement Monitoring Center (IDMC) found that at least 40 countries in 2010 contained IDPs living in protracted displacement. Protracted displacement violates numerous human rights contained in the 1951 Convention. Displacement affects all of nearly all aspects of life. Consequences of prolonged displacement includes poverty, loss of material possessions, kinship and village structure altered, and potential lack of ability to practice the culture and livelihood as they had in their homeland.

4. Loss of power and culture?

   Is culture lost when one is displaced? Should it be presumed that with loss of homeland come loss of cultural identity? To what extent is culture tie to homeland? Is culture portable or is it tied to homeland in such a way that it is lost when forcibly displaced?

The effect of displacement on culture and identity is a divisive topic. Our most nuanced theories of the relationship between displacement and culture comes from anthropological studies. Some anthropologists contend that while there is a strong tendency to equate displacement with loss of culture, this is an essentialist conception of place and emphasize that identity and place should be separated to show that when refugees are displaced, they do not necessarily lose their culture, identity, and power (Malkki 1995).

5. Crime, Perceptions of Crime, and Victimization

After being persecuted in one's homeland for reasons of race, religion, nationality, membership in a particular social group, or political opinion, refugees frequently continue to be discriminated against and victimized in the country that offers them asylum. In other words, refugees are forced to cope with discrimination, marginalization, and persecution in both their home and host country (see Pickering 2005). Refugees displaced due to genocide may continue to be victimized in much the same way as they were in their home country, and they may suffer from new forms of victimization such as ostracism, stigmatization, family disruption, domestic violence, and poverty.

Women in refugee camps from around the world report victimization that stems from gendered division of labor among some refugees. For example, collecting wood and water is often the delegated responsibility of women, and these activities make refugees more vulnerable to attacks because it requires travel to the edges and outskirts of the camps (see Hyndman 2004; Rymond-Richmond and Hagan 2012). In addition, raped women may continue to be victimized via physical trauma, sexually transmitted diseases, ostracism, and reduced likelihood of marriage.

One of the debilitating forms of discrimination is widespread negative stereotypes of refugees that render the refugee as undeserving, different, and marginalized. Of the stereotypes, perhaps none is more pervasive, damming, and incorrect than the perceptions of refugees as criminals. In fact, the perception of crime often extends beyond refugees to include all migrants. In the United States, the connection between immigration and crime has recently been in the spotlight, and it is here that members of the discipline of criminology have made valuable contributions. According to a poll conducted by the National Opinion Research Center in 2000, 73 % of Americans said they believed that immigrants are either "somewhat" or "very" likely to increase crime. Research by sociologists and criminologists find this perception is not only unsupported by data, but evidence demonstrates the opposite. Sampson and his colleagues examined more than 3,000 violent acts committed in Chicago from 1995 to 2003, census data, police reports, and a survey of more than 8,000 residents. They found that immigrants have lower rates of criminal involvement than nonimmigrants. In fact, first-generation Mexican immigrants were 45 % less likely to engage in violence than third-generation Americans. This pattern also extended to neighborhoods. Neighborhoods with large shares of immigrants tend to have lower levels of violence than similarly situated neighborhoods with few immigrants (Sampson 2008).

Coupled with the perception of migrants being criminal is a sentiment that all migration is illegal and undesired. In the USA, one of the results of this negative perception of immigrants is the passing of Arizona Senate Bill 1070. This controversial legislative act has received national and international attention and is considered the broadest and strictest anti-illegal immigration in US history. The result is discriminatory practices of police harassment and arrests including the right of police to demand immigration documents from any individual perceived as being undocumented. While other states are considering adopting similar legislation, some police departments discourage this practice because it could discourage immigrants from reporting crimes and cooperating with law enforcement. Further, a victimized group will likely be discouraged from reporting crimes committed against them, thereby leading to increased victimization of an already vulnerable population. Secretary of State Hillary Clinton reported to the United Nations Human Rights Council that SB1070 was evidence of human rights violations in the United States.

6. Human Rights of Migrants

   While the focus of this entry is human rights violations associated with forced migration, human rights violations are not limited to this particular type of migrant. In fact, "Evidence demonstrates that violations of migrants' human rights are so widespread and commonplace that they are a defining feature of international migration today" (Taran 2002:1). Yet, research on human rights of migrants has been minimal. Currently, there are approximately 214 million international migrants worldwide. Migrants represent 3.1 % of the world's population, or one out of every 33 individuals. 7.6 % of migrants are refugees.

   A few notable advancements in protecting migrant's human rights include the appointment of a UN Special Rapporteur on Human Rights of Migrants and 1990 International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families. The convention aims to prevent and eliminate exploitation of migrant workers. Further, the convention extends fundamental human rights to all migrant workers whether they are documented or undocumented.

7. "Refugee Studies" versus "Forced Migration Studies"

   Refugee studies emerged as an interdisciplinary field in the early 1980s. In more recent times, there has been scholarly debate concerning a paradigm shift from "refugee studies" in favor of "forced migration studies." The shift is largely motivated by the recognition that internally displaced people share much in common with refugees, yet technically would fall outside the scope of "refugee studies." Forced migration studies, on the other hand, would include the study of both categories of displaced people. For some, broadening the field of study signifies inclusion of a marginalized and invisible group of people. Others object with a primary concern that doing so could weaken attention on refugee concerns and detract from the distinct legal status and protection of refugees under international law.

8. Mental health

   Chronic mental health problems, which include depression, anxiety, and posttraumatic stress disorder in the aftermath of genocide and ethnic violence, affect a significant portion of refugees (De Jong et al 2001; Fazel et al. 2005). A survey of Darfur refugees in Cairo described depressive symptoms that include "hopelessness, tearfulness, apathy, decreased concentration, decreased or increased sleep, low appetite, weight loss, low mood, decreased energy, and guilt" (Meffert and Marmar 2009, 1841). Trauma for some forcibly displaced people may include witnessing extremely brutal killings and tortures. Darfurian refugees who survived the genocide provide a glimpse of the extreme brutality they witnessed and experienced. Several Darfurians interviewed as part of the ADS data reported perpetrators targeted pregnant women, male babies or fetuses, and children for torture and death (Rymond-Richmond and Hagan 2012). One female refugee witnessed extreme torture of young boys in the village, which included slitting their throat, cutting the foot open from the big toe to the ankle, cutting off hands and sexual organs, removal of the brain, and removal of skin. Another refugee from a small village in Darfur witnessed the death of seven men who were "dismembered alive. They cut off their penises and put them in their mouths. Cut off their tongues and cut them into pieces."

Future Directions and Suggestions

1. The discipline of criminology should expand beyond focusing on individual criminal actions to include crimes committed by the state.

Studying refugees requires this shift as the state is implicated as either the perpetrator of the crime, incapable of protecting its citizens from persecution or approving of the persecution.

2. Internally displaced persons are without a single agency clearly mandated to assist and protect IDPs. Assistance and protection of forcibly displaced people must include this marginalized population. The appointment of a representative on internally displaced persons (Walter Kälin as of 2004) by the United Nations Secretary-General is a step in the right direction, and significant progress will hopefully be made towards granting this group aid and protection.
3. The rights and protection of children in refugee camps and internally displaced camps must be upheld and prioritized. Children continue to be victimized and targeted for exploitation after being forcibly displaced from their homelands. In addition, their international human right to education is inhumanely comprised.

Related Entries

- ▶ Crime Investigations by the International Criminal Court
- ▶ Environmental and Human Rights
- ▶ Feminist Theory in the Context of Sexual Violence
- ▶ Genocide
- ▶ Marxist Criminology
- ▶ Political Crime
- ▶ Poverty, Inequality, and Area Differences in Crime
- ▶ Resistance to State Crime
- ▶ Restorative Justice and State Crime
- ▶ State Crime
- ▶ Trafficking in Human Beings
- ▶ Transitional Justice
- ▶ Victimization, Gender, and the Criminal Justice System
- ▶ Victims and Restorative Justice
- ▶ Victims and the International Criminal Court
- ▶ Victims of State Crime

Recommended Reading and References

De Jong JT, Komproe IH, Van Ommeren M, El Masri M, Araya M, Khaled N et al (2001) Lifetime events and posttraumatic stress disorder in four post conflict settings. JAMA 286:555–562

Fazel M, Wheeler J, Danesh J (2005) Prevalence of serious mental disorder in 7000 refugees resettled in western countries: a systematic review. Lancet 365 (9467):1309–1314

Hagan J, Rymond-Richmond W (2008) Darfur and the criminology of genocide. Cambridge University Press, Cambridge

Hagedorn J, Rauch B (2007) Housing, gangs, and homicide: what we can learn from Chicago. Urban Aff Rev 42:435–246

Hollenbach D (2008) Human rights as an ethical framework for advocacy. In: Hollenbach D (ed) Refugee rights: ethics, advocacy, and Africa. Georgetown University Press, Washington, DC

Hyndman J (2004) Refugee camps as conflict zones: the politics of gender. In: Giles W, Hyndman J (eds) Sites of violence. University of California Press, Berkeley

International Criminal Court (2007) Office of the Prosecutor, Situation in Darfur, the sudan: prosecutor's application under article 58 (7), 27 Feb 2007

Internal Displacement Monitoring Center (IDMC) (2010) Internal displacement: global overview of trends and developments in 2010. Accessed online 23 Mar 2012 http://www.internaldisplacement.org/8025708F004BE3B1/(httpInfoFiles)/F58CA5EF98ACEEB1C125785C00389FA8/$file/global-overview-2010-fact-sheet.pdf

Kunz EF (1973) The refugee in flight: kinetic models and forms of displacement. Int Migr Rev 7(2):125–146

Malkki L (1995) Refugees and exile: from 'refugee studies' to the national order of things. Annu Rev Anthropol 24:495–523

Meffert SM, Marmar CR (2009) Darfur refugees in Cairo: mental health and interpersonal conflict in the aftermath of genocide. J Interpers Violence 24:1835–1848

Pickering S (2005) Refugees and state crime. Federation Press, Annandale

Rymond-Richmond W, Hagan J (2012) Race, land, and forced migration in Darfur. In Kubrin CE, Zatz MS, Martinez R (eds), Collateral consequences of immigration policies: confounding effects and immigration responses. New York University Press, New York

Sampson RJ (2008) Rethinking crime and immigration. Context 7(1):28–33

Taran PA (2002) Human rights of migrants: challenges of the new decade. Int Migr 38(6):7–51

Totten S (2006) The U.S. investigation into the Darfur crisis and its determination of genocide: a critical analysis. In: Totten S, Markusen E (eds) Genocide in Darfur: investigating the atrocities in Sudan. Routledge, New York

UNHCR (2010) 2009 Global trends: refugees, asylum-seekers, returnees, internally displaced ad stateless persons. Country Data Sheets. 15 June 2010. Accessed online 17 Sept 2010. http://www.unhcr.org/4c11f0be9.html

Forecasting Crime or Crime Forecasting

▶ Predictive Policing

Forecasting of Shootings Using Risk Terrain Modeling

Joel M. Caplan and Leslie W. Kennedy
School of Criminal Justice, Rutgers University, Newark, NJ, USA

Synonyms

Risk terrain modeling (RTM); Spatial influence

Overview

This entry summarizes approaches that have been used to study patterns of shootings and empirically examines the environmental conditions that contribute to the emergence of these events. Geographic information systems (GIS) allow for the mapping of spatial influence which refers to the way in which correlates of crime, as features of a landscape, affect places throughout the landscape. Rather than just a feature's presence, its influence on space is important because context affects criminal behavior. Operationalizing the spatial influence of crime correlates allows analysts to move beyond just creating maps of points that coexist with crime to creating visual narratives of how settings become conducive to crime. This entry presents ways in which to operationalize criminogenic features of gun shootings to a geographic map. Then it presents the utility of this mapping technique for crime analysis through the use of risk terrain modeling (RTM) to produce a map of compounded criminogenic contexts for shootings in Irvington, NJ. RTM builds upon underlying principles of environmental criminology and problem-oriented policing to articulate and communicate criminogenic and vulnerable areas at the micro-level. By using theoretically grounded operationalizations of spatial influence from many correlates of gun shootings, the entry demonstrates how place-based risk analysis can be achieved by the application of some intuitively simple principles within the analytical reach of crime analysts using basic GIS software.

Introduction

At a time when overall crime rates have dropped, gun shootings continue to plague inner city areas tied to the illegal activity around gangs and drugs. Shootings are often seen as a sign of extreme disorder, occurring when offenders are at their most brazen and law enforcement at its least effective. Shootings are often used as indicators of crime sprees or uncontrollable gangs, and each event attracts special attention from the media. Shootings attract a lot of attention at the political level as well, prompting calls for special policing interventions to stop the violence. A well-known example of this is the multicity program "Operation Ceasefire" that has concentrated on reducing gun violence through a variety of initiatives, including arrest, assistance to offenders from criminal justice agencies, and active involvement in crime reduction by the affected community (for a complete review of these initiatives, see Koflas et al. 2011). Despite all of this attention, or perhaps because of the singular ferocity of the crime event, shootings are often considered in ways that focus more on the specific incident and not as much on the situational factors that might support and contribute to this violent outcome at certain locations. Academic research is beginning to turn more attention to these contextual affects. This entry summarizes approaches that have been used to study patterns of shootings and examines the environmental conditions that contribute to the emergence or disappearance of these events. The entry begins with an overview of the literature on place and crime and then turns to a discussion of spatial influence of place-based features on criminal outcomes, particularly gun shootings. It concludes with an example of a cartographic method for operationalizing spatial influence and then presents a composite model of place-based risk of shootings in Irvington, New Jersey.

State of the Art: Current Knowledge on the Concentration of Shootings

When considering gun shootings, questions have been raised about the extent to which this crime is, in fact, situational and nonrandom. Ratcliffe and Rengert (2008) make the case for the importance that previous shootings play in leading to near repeats – crimes that happen close to and soon after a previous similar incident: This research is modeled after similar studies on burglary. Ratcliffe and Rengert report that there is an elevated risk of about 30 % that a shooting will occur within 2 weeks and within one block of a previous incident. They suggest that these findings show shootings to be influenced by coercion, retaliation, and escalation of violence. So, the characteristics of the interaction between offender and victim will lead to an elevation in the likelihood of shootings in the vicinity of past events. But, why do the initial incidents happen where they do and are there similarities in the types of places that incidents occur? That is, are there environmental (i.e., contextual) factors that bring these crimes together at similar places?

Criminological research about the concentration of crime primarily focuses on spatial clusters or hot spots of crime – where there is a relatively high frequency of similar criminal behaviors at specific places. Foremost research by Sherman et al. (1989) studied the locations of requests for police assistance in the city of Minneapolis over a 1-year period. Spelman and Eck (1989) found that while 10 % of victims in the United States were involved in 40 % of victimization, 10 % of offenders were involved in 50 % of crime, and 10 % of places were involved in 60 % of crimes. More recently, Eck and Weisburd (1995), Johnson (2010), Johnson et al. (2008), Bowers et al. (2004), Braga and Weisburd (2010), and others reported similar clustering effects. Caplan (2011), for instance, reported that less than 1 % of the area of Irvington, New Jersey, accounted for more than 53 % of all shooting incident locations. In Houston, Texas, Wells et al. (2011) found gun assaults to cluster both temporally and spatially.

These findings related to hot spots and near repeats are important and address key situational aspects of violent crime. They provide evidence that these crimes connect to one another by being close to one another. However, they do not address the question of how these crimes, especially the initial instigator events, are influenced by the underlying conditions of the environment. Can one establish that there are spatial patterns that appear in these behaviors that are not just a function of individual decisions to be violent but are a function of underlying environmental factors that concentrate violent acts at specific locations? More simply, why do shootings happen in some places and not in others? This question leads one to consider how to use spatial analysis techniques to connect violent behavior to what Patricia and Paul Brantingham referred to as an environmental backcloth: a combination of qualities or characteristics of places that raises the risk of crime occurrence (Brantingham & Brantingham 1995). According to the Brantinghams, crime generators create settings that are conducive to particular types of criminal acts by creating particular times and places that provide appropriate concentrations of people and other targets. Their combined effects can heighten the risk of crime. The Brantinghams suggest that offenders find targets at these locations that are unlikely to be adequately protected through surveillance or policing. The high levels of crime that occur at certain places, then, come from people at these locations taking advantage of opportunities afforded them through placement of targets. The Brantinghams' conceptualization of crime generators was recently studied in detail by Bernasco and Block (2010) looking at robberies in Chicago. They indicate that certain features of the environment have a great influence on where crime locates. As a result, robberies, for instance, are not evenly distributed across the urban landscape but concentrate around generators. Similarly, Moreto et al. (2011) found that residential burglaries occur at micro-level places with risky environmental features and that near repeat burglary incidents were even more likely to occur at higher risk places than their instigator events. Studies such as these provide evidence that analyzing the risk heterogeneity of an area is especially useful for developing more complete understandings of crime problems. They also stress the reality that underlying (risky) environmental contexts are often present before either instigator or near repeat incidents.

It is clear from existing research that certain environments have characteristics that encourage or discourage the occurrence of shootings and, therefore, raise or lower criminogenic risk. Shootings tend to be spatially distributed according to the distributions of underlying generators, such as drug markets, or certain features of a landscape, such as bars, clubs, fast-food restaurants, or liquor stores; gang members; bus stops; schools; or public housing. Assigning risk levels to places with certain features of environments, such as these, requires an understanding (i.e., through past empirical research) of their relative importance in supporting shootings. The risk posed by each criminogenic feature throughout a landscape, as well as their confluence at the same place, contributes to a risk value that, when raised a set amount, increases the likelihood of shooting incidents. This risk value – a measure of the clustering of criminogenic factors – can be used to anticipate where shootings will occur and (possibly) cluster over a period of time. Stated another way, risk values are a measure of a place's vulnerability to crime.

People can live in vulnerable areas defined by an agreed-upon set of criminogenic features, but in the absence of motivated offenders, the existent risk of crime may be relatively low. Crime risks are both place-based and situational – affected by exposure to individual events that appear from location to location and by exposure to places where events cluster over time. It is important to build models of the environmental backcloth that consider different features of the landscape for different crime types because factors that influence the occurrence of one type of crime are likely to differ from those that are associated with other crime types. Cartographically modeling vulnerability as the clustering of risk factors, and then interpreting vulnerability in the context of exposure, permits a research strategy to identify, monitor, and control these environments.

Spatial Influence: The Future of Mapping Risk Factors

Opportunities for crime are not equally distributed across places, or "small micro-units of analysis" (Weisburd 2008, p. 2), and so the analytical approach to studying criminogenic places plays a critical role in the reliability and validity of efforts to assess vulnerabilities and future crime hot spots. Crime control and prevention activities must consider not only who is involved in the criminal events "but also the nature of the environments in which these activities take place" (Kennedy and Van Brunschot 2009, p. 129) because opportunity for crime is an attribute of all places. Risk assessment, defined by Kennedy and Van Brunschot (2009, p. 4) as "a consideration of the probabilities of particular outcomes," provides an efficient way to analyze crime opportunities. However, challenges appear in the operationalization of risk in a GIS. The way that criminogenic features have been modeled in a GIS is often contrary to how people experience and conceptualize their environments (Couclelis 1992; Frank and Mark 1991).

When assessing the risk of crime to occur at conceivably any location throughout a landscape, the use of finite points, lines, and polygons in a GIS is a poor representation of criminogenic features on a map because they bear no particular relationship to the dynamic environments of which they are a part (Couclelis 1992). "Points, lines, and polygons that define vector objects do not have naturally occurring counterparts in the real world," explained Couclelis (1992, p. 66). They are approximations of environmental features, but without any theoretical or empirical link to their geographies (Freundschuh and Egenhofer 1997). The way people conceptualize and operate in space is an important consideration for the mapping of crime risk throughout landscapes. Cartographically modeling these conceptualizations and the spatial influence of criminogenic features in a GIS in a way that reflects the actors' views is an important part of what Freundschuh and Egenhofer (1997) describe as "Naïve Geography, a set of theories of how people intuitively or spontaneously conceptualize geographic space and time" (Egenhofer and Mark 1995), and can yield more meaningful and actionable spatial intelligence for use by public safety professionals (Frank 1993; Mark 1993; Freundschuh and Egenhofer 1997).

The best way to map crime factors for the articulation of criminogenic backcloths (Brantingham and Brantingham 1981) is to operationalize the spatial influence of each factor throughout a common landscape rather than a-theoretically map the factors as points, lines, or polygons in a manner that keeps them disconnected from their broader social and environmental contexts. The concept of spatial influence refers to the way in which features of a landscape affect places throughout the landscape. The connection between criminogenic features and shooting incidents depends on one's ability to operationalize risk emanating from such features to all places throughout a landscape. Defining vulnerable places for shootings, then, is a function of the combined spatial influence of criminogenic features throughout a landscape that contribute to crime by attracting and concentrating on illegal behavior. Spatial influence can be cartographically modeled with a GIS to produce risk map layers for each criminogenic feature. So, while all places might carry some risk of shootings, some places are riskier than others because of the spatial influences of certain criminogenic features known to correlate with shooting incidents. Fortunately, decades of criminological research have identified a variety of independent variables to be significantly correlated with a range of crime outcomes that can be used to inform such expectations and to cartographically model them in a GIS, accordingly.

Operationalizing the spatial influence of crime risk factors addresses various theoretical and methodological issues concerning the use of GIS for crime forecasting and cluster analysis (Freundschuh and Egenhofer 1997). Though, crime analysis can yield different results when the same crime correlate is modeled in a GIS according to a variety of empirically justified spatial influences. To illustrate the issues at hand, presented in the next section is an illustration of the primary methods of operationalizing spatial influence of risk factors for shootings in one city.

Demonstrating Methods of Investigating the Concentration of Shootings

Figure 1 shows point symbols representing the locations of shooting incidents (the crime) and bars, clubs, fast-food restaurants, check-cashing outlets, and liquor stores (the crime generators) in Irvington, NJ. These crime generators, or criminogenic features of the landscape, are correlated with shooting incidents in several empirical research studies, both in Irvington and other settings (e.g., Caplan et al. 2011; Kennedy et al. 2011). There are 102 features and 60 shooting incidents from 2007 represented on the map in Fig. 1. Visual inspection of the map suggests that shooting incidents are distributed in a similar way as criminogenic features. However, only one shooting incident was at the exact same address as a fast-food restaurant.

Bars, clubs, fast-food restaurants, check-cashing outlets, and liquor stores are often not the exact locations where associated crimes happen. Rather, shootings occur at places that are in some way defined or influenced by them. As illustrated in Fig. 2, the spatial influence of these features on shooting incidents is related to their concentration at places throughout the

Forecasting of Shootings Using Risk Terrain Modeling, Fig. 1 Irvington, NJ: crime risk factors represented as points

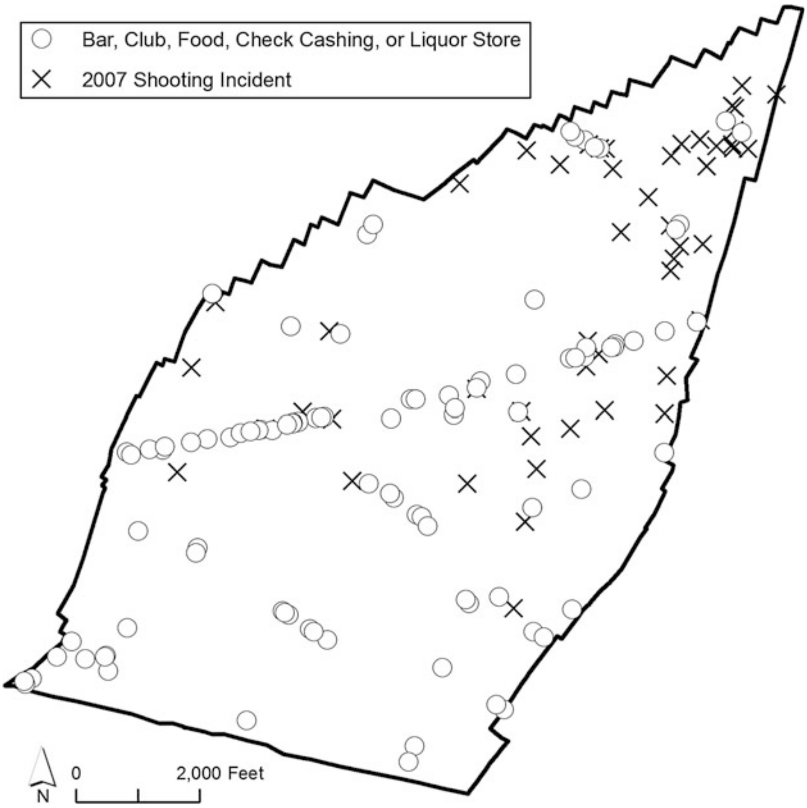

municipality and was operationalized as a raster density map. Shooting incidents are cartographically modeled as more likely to happen at places where bars, clubs, fast-food restaurants, check-cashing outlets, and/or liquor stores are most concentrated. That is, the area over which these features may have an influence on the occurrence of shootings is not modeled as limited to the precise location of the feature itself but is likely to diffuse to nearby places. "Places" are defined in the raster map as cells sized 100 ft by 100 ft, and the distal limits of nearby features (i.e., the bandwidth or search radius) used to define the density of each place were set at 1,480 ft (i.e., approximately four blocks; a meaningful sphere of influence for these criminogenic features as attributes of places within the radius). The density map is symbolized according to standard deviational breaks, with all places colored in darker gray having density values greater than +2 standard deviations from the mean density value – which statistically puts these places in the top 5 % of the most densely populated with criminogenic features. As shown on the map, places with greater concentrations of criminogenic features appear to be more frequented by shootings than the bars, clubs, fast-food restaurants, check-cashing outlets, or liquor stores themselves (compared to Fig. 1). In fact, 13 out of 60 (22 %) shootings during 2007 occurred at places with density values above +1 standard deviations.

Perhaps bars, clubs, fast-food restaurants, check-cashing outlets, and liquor stores are the venues where most suitable victims hang out or where the most likely and motivated offenders visit, become intoxicated, or lose self-control. However, due to increased police presence or other capable guardians such as bouncers, witnesses, or CCTV cameras, offenders do not shoot their victims inside or directly outside of such facilities. Rather, shootings are more likely to occur at certain distances away. Thought of in this way, the spatial influence of bars, clubs,

Forecasting of Shootings Using Risk Terrain Modeling, Fig. 2 Irvington, NJ: crime risk factors represented as density

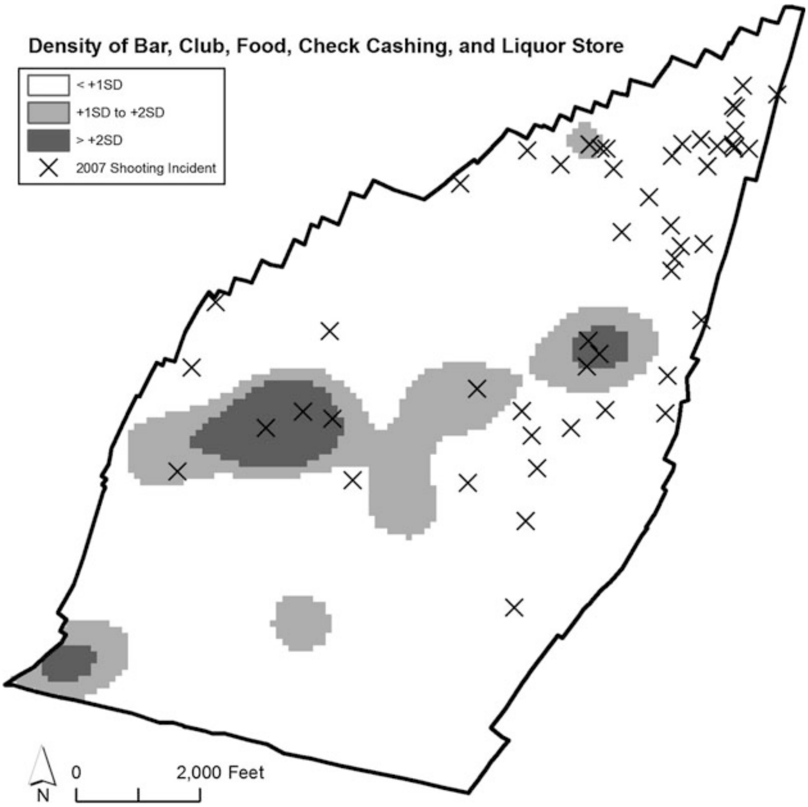

fast-food restaurants, check-cashing outlets, and liquor stores on shootings is more a function of distance from the closest feature rather than the presence or absence of the feature at the shooting incident location. Figure 3 presents a map that operationalizes this distal spatial influence with a 370 ft buffer (about 1 block) around all features. Thirty-one shootings occurred within one block from a bar, club, fast-food restaurant, or liquor store. The operationalized spatial influence of the highest risk created by these features to be all places within 370 ft – a definition that was grounded in theory and empirical research – identified places with more than half (52 %) of all shooting incident locations. These results are arguably due to identifying a larger catchment area to which shootings are aggregated. Compared to feature points themselves, this is true. However, the coverage area of places with density values above +1 standard deviations is 0.806 mile^2, and the coverage area of places within one block of a criminogenic feature is 0.725 mile^2. So, more shootings occurred in a smaller area that was deemed affected by nearby criminogenic features in a conceptually meaningful way.

The most meaningful cartographic model of places in Irvington that are at the greatest risk of shootings is one that operationalizes the spatial influence of bars, clubs, fast-food restaurants, check-cashing outlets, and liquor stores to be up to a certain distance away. This might not be the case in other jurisdictions or for other crime types, but it is corroborated by existing research, and it reinforces the importance of modeling these criminogenic features in commensurate ways on a map. This reinforces the notion that the study of crime concentration and approaches to crime analysis and forecasting cannot be a-theoretical and must be evidence based. They must be grounded in ways that account for the dynamic interaction of all criminogenic features throughout a landscape.

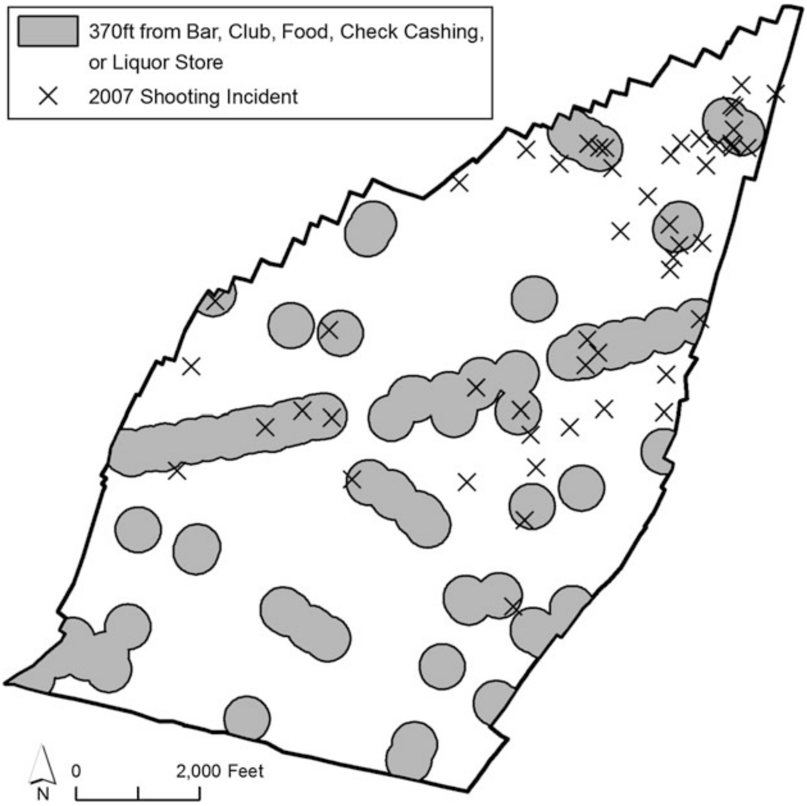

Forecasting of Shootings Using Risk Terrain Modeling, Fig. 3 Irvington, NJ: crime risk factors represented as distance

Combined Spatial Influences of Criminogenic Features: Risk Terrain Modeling

The inclusion of multiple risk factors for the place-based study of shootings allows one to consider the interaction between the offender and the environment. Particularly, the potential for crime occurrence as a function of the illegal behavior that one could expect from motivated offenders at these locations. For example, risk from gang membership could be measured as the addresses of all known gang members' residences and operationalized as a density map if the spatial influence of these features is understood as "areas with greater concentrations of gang members residing will increase the risk of those places having shootings." Risk from bus stops could be operationalized as a distance map up to 555 ft away if the spatial influence of these features is understood as "up to one and a half blocks away from bus stops are at greater risk for shootings because targeted victims are most vulnerable when they arrive at or leave these destinations." A school risk factor could be operationalized using evidence that "distances between one and three blocks (between 370 ft and 1,110 ft) are at the greatest risk for shootings"; public housing risk could be operationalized using knowledge that "distances up to one block (up to 370 ft) are at greatest risk for shootings."

Once shooting risk factors are identified and operationalized to maps according to their spatial influences, they can be combined in a geographic information system (GIS) to show their compounded presence, absence, or intensity at every location throughout the landscape. In such a composite map, crime explanations are accounted for by different factors that tie different components of risk together, rather than on their own, to explain environmental influences and impacts on shooting events. Clustering of illegal shootings at particular areas, then, is explained by the unique

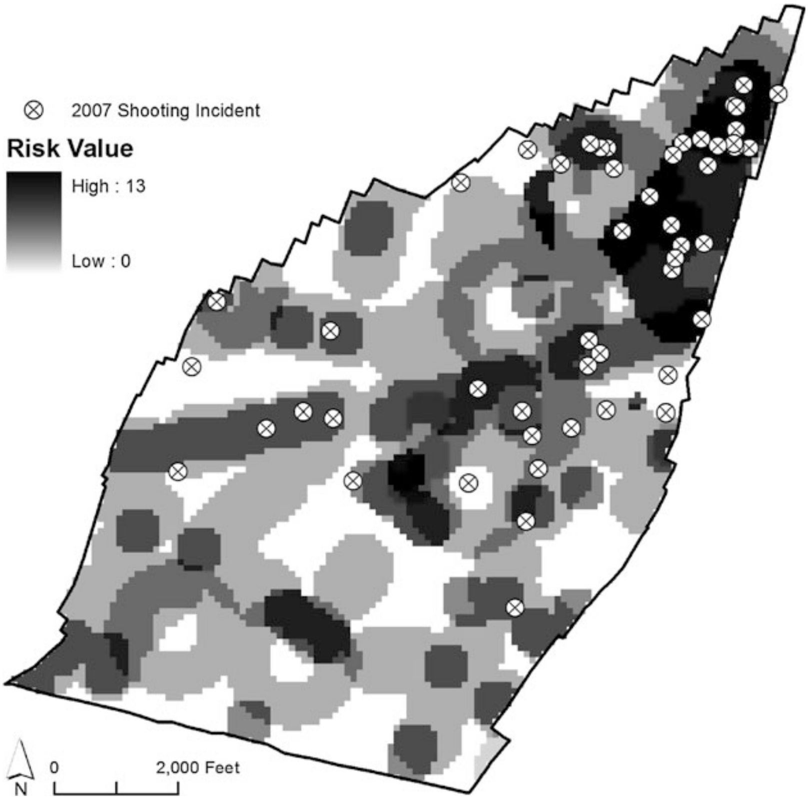

Forecasting of Shootings Using Risk Terrain Modeling,
Fig. 4 Irvington, NJ: weighted risk terrain of gun shootings

combination of factors that make these areas opportune locations for such crime.

Risk terrain modeling (RTM) is a method of spatial risk assessment that utilizes a GIS to attribute multiple qualities of the real world to places on a digitized map. RTM builds upon underlying principles of environmental criminology and problem-oriented policing and offers a statistically valid way to articulate and communicate criminogenic and vulnerable areas at the micro-level. It "paints a picture" of place-based environmental context for criminogenesis.

Figure 4 is a risk terrain for shootings that includes all of the aforementioned risk factors and that was produced in accordance with the steps described by Caplan and Kennedy's (2010) *Risk Terrain Modeling Manual*. Essentially, it is the product of the summed risk map layers of operationalized spatial influences of bars, clubs, fast-food restaurants, check-cashing outlets, and liquor stores; gang members' residences; bus stops; schools; and public housing, respectively. The co-location of the spatial influence of these risk factors at a place increases the place's risk value accordingly. It may be expected that higher risk places in Irvington were the most conducive for shootings to occur and, therefore, would be most likely to host shooting incidents in 2007, unless one or more risk factors were mitigated at these places. Logistic regression results for this risk terrain suggest that for every unit increase of a place's (i.e., 100 ft × 100 ft cell's; n = 3,975) risk value, the likelihood of a shooting occurring there in 2007 increased by 35 % (B = 0.30; S.E. = 0.045; Wald = 43.67; df = 1; Exp (B) = 1.35; $p < 0.001$). For places with one or more risk factors, we can be 95 % confident that if shootings happened in 2007 the likelihood of them happening at these places was between 23 % and 48 % greater than other places in Irvington ($p < 0.001$). According to results of a chi-squared test, more than 53 % of all shootings that happened in 2007 occurred at places which comprised about 15 % of the area (where shootings could have been geocoded) of Irvington (Pearson chi-square = 88.37, df = 12, p < 0.001, n = 3,975). That is, 15 % of the highest-risk places in Irvington accounted for more than 50 % of shooting incidents.

Open Questions for Further Research

Although shootings may occur at different places from one incident to the next, they are most likely to occur at places where features of the landscape better enable such illegal activity to happen. The risk terrain model presented above does not include many (if any) dynamic risk factors. That is, the locations of school, bars, etc., do not change drastically from year to year. Therefore, this "time-stable" risk terrain model, created with data current for the year 2007, is also likely to predict micro-level shooting locations during 2008. This spatial forecasting strategy for shooting incidents, based solely on environmental context and not on prior shooting event locations, should be valid even if shooting locations change over time. Results of a logistic regression suggest that for every unit increase of a place's risk value, the likelihood of a shooting occurring there in 2008 increased by 28 % (n = 3,975; B = 0.24; S.E. = 0.045; Wald = 29.37; df = 1; Exp(B) = 1.28; p < 0.001). So, the likelihood of a shooting happening at particular 100 ft-by-100 ft places in Irvington during 2008 significantly increases as each additional risk factor influences that place. Fifteen percent of the highest-risk places in Irvington in 2007 accounted for more than 43 % of future shooting incidents in 2008 (Pearson chi-square = 51.63, df = 12, p < 0.001, n = 3975).

Although practically meaningful and statistically significant, the predictive validity of the 2007 risk terrain model for 2008 shooting incidents is not as strong as that for 2007 shootings. This is likely because changes that occurred to the Irvington environment throughout 2007 as a result of other crime events, police activities, or social interventions are not accounted for in the risk terrain model and subsequent year-long risk assessment of place-based criminogenesis. These activities and characteristics of the environment could be added to the model to make it more representative of the criminogenic environment during 2008 and to improve its predictive validity. Even as is, the risk terrain model successfully demonstrates the substantial influence that environmental context has on illegal behavior and crime occurrence. Such information can help police commanders and other community stakeholders stay ahead of crime problems by assessing the risk for serious crime hot spots to emerge.

Discussion and Conclusion

This entry reviewed ways of using criminogenic features to explain the emergence and concentration of shooting incident locations. Risk terrain modeling is one method for modeling place-based risks in a GIS. The risk terrain model produced with thoughtfully operationalized criminogenic features, selected according to findings from previous empirical research, yielded a valid and reliable place-based forecasting model. Operationalizing the spatial influence of criminogenic features to maps was an important task for maximizing the construct validity of the impact these features had on crimes. Combining these maps in a GIS articulated environmental contexts of places that were most likely to attract, enable, and/or generate criminal shooting incidents as a function of the combined influences of criminogenic features in Irvington. Resulting maps can be used to efficiently allocate resources and target interventions.

Police officers are often well aware of the factors associated with shooting incidents in their jurisdictions, and their intuitions about such relationships tend to enjoy support from empirical literature. Gun shootings can occur at similar places to other types of illegal behavior, such as open-air drug markets, gang membership/territories, or armed robberies. There is also specialization in terms of where shootings occur – both as a function of the underlying conditions most suitable for related or precursor crimes and as a function of where certain opportunities exist (e.g., where victims are located). Take out eateries, for instance, are characterized by high foot traffic and late hours of operation, making it difficult to distinguish legitimate customers from those loitering for unlawful purposes. Due to the relative lack of guardianship and concentration of potential offenders and victims, such eateries can create ideal environments for shooting incidents (Cohen and Felson 1979). Including measures of environmental risks for shootings yields a better model of where future shooting incidents are likely to occur compared to predictions made with past violent shooting incidents alone. Risk heterogeneity of environments, as articulated by risk terrain maps, exists prior to the initial victimization and can be enduring without proper intervention efforts. Certain environments may also enable a contagion effect that promotes (re)victimizations in close proximity and soon after an original shooting occurs.

Crime mapping has become a standard function of law enforcement agencies throughout the United States. Mapping innovations directly contribute to the police function and have moved the field towards a place-based approach. This is because offenders occupy space and bring to certain locations characteristics that might threaten others and disrupt social interaction at these locations. They may also engage in illegal activities (such as gangs or drugs) that themselves change the risk character of these locations (and also attract attention from law enforcement). Most often a crime analyst's measure of the presence of offenders is designated as the number of crime incidents reported or arrests that are made and tabulated by police in crime reports. But, there are other types of measures to use that are more enduring than the crime event. This point alludes to human ecologists' interests in studying urban processes in which "the whole is greater than the sum of its parts" – where the focus is on the potential for, or risk of, crime in areas that comes as a result of the characteristics found in these areas. As Abbott (1997) states, "the Chicago School thought that no social fact makes any sense abstracted from its context in social (and often geographic) space and social time. . . .Every social fact is situated, surrounded by other contextual facts and brought into being by a process relating it to past contexts" (1152).

Contextual forecasting models are repeatedly found to be more accurate in predicting future crime locations than are models based only on previous crime occurrences (e.g., Johnson et al. 2008; Caplan and Kennedy's 2010; Kennedy et al. 2011; Groff and La Vigne 2002). Tying predictions of shootings and other crimes to geographic locations provides the basis for connecting attributes of space to actual behavior that occurs at these places. It also takes the police beyond a tactical reaction to crime occurrence to one that is more strategic – anticipating where resources will be needed and to prevent newly emerging crime problems using contemporary evidence-based and interdisciplinary practices.

Related Entries

- ▶ Crime Mapping
- ▶ Early Chicago School Theory
- ▶ Evidence-Based Policing
- ▶ History of Geographic Criminology Part I: Nineteenth Century
- ▶ Hot Spots and Place-Based Policing
- ▶ How to Make Crime Mapping More Inferential
- ▶ Intelligence-Led Policing
- ▶ Predictive Policing
- ▶ Risk Assessment, Classification, and Prediction

Recommended Reading and References

Abbott A (1997) Of time and space: the contemporary relevance of the Chicago. Soc Forces 75: 1149–1182

Bernasco W, Block R (2010, online first) Robberies in Chicago: a block level analysis of the influence of crime generators, crime attractors, and offender anchor points. J Res Crime Delinq 48(1):33–57

Bowers KJ, Johnson SD, Pease K (2004) Prospective hotspotting: the future of crime mapping? Br J Criminol 44(5):641–658

Braga AA, Weisburd DL (2010) Policing problem places. Oxford University Press, New York

Brantingham PJ, Brantingham PL (1981) Environmental criminology. Sage, Beverly Hills

Brantingham PJ, Brantingham PL (1995) Criminality of place: crime generators and crime attractors. Eur J Crim Policy Res 3:1–26

Caplan JM (2011) Mapping the spatial influence of crime correlates: a comparison of operationalization schemes and implications for crime analysis and CJ practice. Cityscape 14(3):57–83

Caplan JM, Kennedy LW (2010) Risk terrain modeling manual: theoretical framework and technical steps of spatial risk assessment. Rutgers Center on Public Security, Newark. Available at www.riskterrainmodeling.com

Caplan JM, Kennedy LW, Miller J (2011) Risk terrain modeling: brokering criminological theory and GIS methods for crime forecasting. Justice Q 28(2):360–381 [2010, online first]

Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44:588–608

Couclelis H (1992) People manipulate objects (but cultivate fields): beyond the raster-vector debate in GIS. In: Frank A, Campari I, Formentini U (eds) Theories and methods of spatio-temporal reasoning in geographic space. Lecture notes in computer science 639. Springer, New York, pp 65–77

Eck JE, Weisburd D (1995) Crime places in crime theory. In: Eck JE, Weisburd D (eds) Crime and place, crime prevention studies, vol 4. Willow Tree Press, Monsey, pp 1–34

Egenhofer M, Mark D (1995) Naïve geography, pp 1–15. In: Frank A, Kuhn W (eds) Spatial information theory – a theoretical basis for GIS, international conference COSIT ‘95, Semmering, Lecture notes in computer science 988. Springer, New York

Frank A (1993) The use of geographical information systems: the user interface is the system. In: Medyckyj-Scott D, Hearnshaw H (eds) Human factors in geographic information systems. Belhaven Press, London, pp 3–14

Frank A, Mark D (1991) Language issues for GIS. In: Maguire D, Goodchild M, Rhind D (eds) Geographic information systems: principles. Longman, London, pp 147–163

Freundschuh SM, Egenhofer MJ (1997) Human conceptions of spaces: implications for geographic information systems. Trans GIS 2(4):361–375

Groff ER, La Vigne NG (2002) Forecasting the future of predictive crime mapping. In: Tilley N (ed) Analysis for crime prevention, vol 13, Crime prevention studies. Criminal Justice Press, Monsey, pp 29–58

Johnson SD (2010) A brief history of crime concentration. Eur J Appl Math 21(4/5):349–370

Johnson S, Bowers KJ, Birks D, Pease K (2008) Predictive mapping: accuracy for different units of analysis and the role of the environmental backcloth. In: Weisburd D, Bernasco W, Bruinsma G (eds) Putting crime in it’s place: units of analysis in spatial crime research. Springer, New York, pp 171–198

Kennedy LW, Van Brunschot EG (2009) The risk in crime. Roman and Littlefield, New York

Kennedy LW, Caplan JM, Piza E (2011) Risk clusters, hotspots, and spatial intelligence: risk terrain modeling as an algorithm for police resource allocation strategies. J Quant Criminol 27:339–362

Koflas JM, Hipple NK, McGarrell E (2011) The new criminal justice: American communities and the changing world of crime control. Routledge, New York

Mark D (1993) Human spatial cognition. In: Medyckyj-Scott D, Hearnshaw H (eds) Human factors in geographical information systems. Belhaven Press, London, pp 51–60

Moreto WD, Piza EL, Caplan JM (2011) A plague on both your houses?: risks, repeats and reconsiderations of urban residential burglary. Paper presented at the American society of criminology annual meeting; Washington, DC

Ratcliffe JH, Rengert GF (2008) Near repeat patterns in Philadelphia shootings. Secur J 21(1–2):58–76

Sherman LW, Gartin PR, Buerger ME (1989) Hot spots of predatory crime: routine activities and the criminology of place. Criminology 27:27–55

Spelman W, Eck JE (1989) Sitting ducks, ravenous wolves, and helping hands: new approaches to Urban policing. Public Aff Comment 35(2):1–9

Weisburd D (2008) Place-based policing. Ideas in policing series. Police Foundation, Washington, DC

Wells W, Wu L, Ye X (2011) Patterns of near-repeat gun assaults in Houston. J Res Crime Delinq. Published online 12 May 2011

Forensic Anthropology

Roxana Ferllini
Institute of Archaeology, University College London, London, UK

Overview

Forensic anthropology, a subdiscipline of physical anthropology, is dedicated to the biological identification of human remains discovered within a legal context in advanced stages of decomposition, or in situations where the integrity of the body has been compromised, thereby thwarting a positive identification through conventional means. Since its inception during the late nineteenth century (Stewart 1979), forensic anthropology has matured as a science, and professionals in this field have become actively involved in domestic or international police investigations, human rights investigations, and disaster victim identification (DVI), forming an integral element of body handling within the mortuary and in some instances in the field during the search and location of victims.

The advent of global participation and communication among forensic anthropologists has led to the development of new methodologies, such as state-of-the-art 3D imaging technology, which assist with the biological identification of human remains within a variety of contexts and circumstances.

This entry aims to present the historical process from which the discipline has evolved, and sheds light upon the application and methods utilized when a biological identification is required during a legal investigation. Current challenges are expanded upon; the future of the discipline is discussed and also potential avenues for future research.

Fundamentals of Forensic Anthropology

Developmental Background

The origins of forensic anthropology commenced within the United States during the late nineteenth century, in connection with studies conducted by the anatomist T. Dwight (1843–1911), concerning among many elements, the determination of sex, age, stature, and variability in relation to the human skeleton; his contribution led to the title of *Father of American Forensic Anthropology* to be conferred upon him (Stewart 1979). Dwight's research constituted a catalyst for this discipline, and also served to inspire those who followed. When viewed in relation to modern standards, his studies may appear rather basic, but they provided the initial impetus for the promotion of a discipline that has aided law enforcement organizations with the identification of human remains through many decades and, within a more recent context, involvement in trauma identification and description related to cause and manner of death. Dwight's studies were continued by the American anthropologist G. Dorsey (1868–1931) into the early part of the twentieth century, at which time the latter offered a significant contribution to the development and refinement of techniques applied for human identification based on bone analysis; in addition, his intervention as an expert witness gave the discipline an initial profile within the crimninal justice system (Klepinger 2006; Stewart 1979).

The twentieth century can be seen as the era in which engagements with law enforcement agencies and the necessity to identify war casualties constituted decisive junctures which led to forensic anthropology becoming a valid discipline in its own right. During the last century, a particularly important figure which stands out is A. Hrdlička (1869–1943). Hrdlička, who was formally trained in medicine, was director of the Anthropology Department at the Smithsonian Institution, where he amassed an impressive collection of human remains, and conducted active research in the field of physical anthropology. Additionally, Hrdlička was the founder of the American Journal of Physical Anthropology (1918) and also the American Association of Physical Anthropologists (1930), of which both elements aided in providing a formal grounding to this discipline; within the area of forensic intervention, Hrdlička received skeletal cases from the Federal Bureau of Investigation (FBI) (Ubelaker 1999). Parallel to the work of Hrdlička was that of W. Krogman (1903–1987), who in 1939 published "Guide to the Identification of Human Skeletal Material" in the FBI Law Enforcement Bulletin, considered by some as the publication that formally recognized the specific importance of forensic anthropology (Tersigni-Tarrant and Shirley 2013; Stewart 1979; Ubelaker 1999). With the advent of Krogman's publication, a guide was available that provided a point of reference for further research, in addition to constituting a valuable reference guide for use in the field.

Upon Hrdlička's death, T. Dale Stewart (1901–1997) took over the post at the Smithsonian Institution; not only were investigative cases emanating from the FBI continued, but the valuable contributions that Stewart made to academic teaching and research improved and promoted new developments in bone analysis within forensic cases (Kennedy 2000).

In 1972 the physical anthropologist E. Kerley (1924–1998) founded the Physical Anthropology Section in the American Academy of Forensic Sciences (AAFS), and it is here where those involved in the analysis of human remains within the legal context were recognized formally within a prestigious association. This advance was followed by the possibility of members being formally certified through the American Board of Forensic Anthropology (ABFA), an organization created to foster and maintain standards within the profession and to confer professional weight and credentials to those practicing members (Tersigni-Tarrant and Shirley 2013).

These important events led to more in-depth research and initiatives, which were further strengthened when the US armed forces recognized the requirement for the identification of human remains of those who perished in combat. Such interventions occurred in World War II (1939–1945) and the Korean War (1950–1953) and during the Vietnam War (1965–1975) (Stewart 1979; Klepinger 2006), a challenge that has increased exponentially in relation to advancements in weapons technology. As a result of such interventions, the US Army Central Identification Laboratory, Hawaii (CILHI), was established in 1976; today it is called Joint POWMIA Accounting Command (Joint Prisoners of War Missing in Action Accounting Command) (JPAC 2013), with the mission of locating all of the Americans who are still missing from past wars. This is achieved by locating, recovering, and identifying the remains, a process whereby forensic anthropology is a key component for a successful outcome (JPAC 2013; Tersigni-Tarrant and Shirley 2013).

As a result of its evolution with respect to these major events, forensic anthropology became established as a specialized forensic field in its own right.

Definition and Functions

Forensic anthropology can be defined as a subdiscipline of physical anthropology (also referred to as biological anthropology), which conducts biological identifications on human remains within a legal context and in circumstances when they are in a poor state of preservation including advanced stages of decomposition, dismemberment, fragmented or possibly incomplete, burnt, skeletonized, saponified, or a combination of such factors (Nafte 2000; Klepinger 2006) (see Figs. 1 and 2). Within such scenarios, saponification occurs when the body, whole or in part, is covered with adipocere, a grayish waxlike substance formed by the hydrolysis of body fats when exposed to an environment which is moist and low in oxygen, such as the case with bodies held underwater; however, saponification may also occur in desertlike environments when the water content in the fat cells compensates for the lack of moisture within the environment (Gill-King 1997).

The process of identification is accomplished via biological profiling of the body in order to determine sex, age, stature, biological affinity (ancestry), and individual characteristics through bone analysis. In some individual cases, the forensic anthropologist will also be involved with trauma identification on bone in instances such as sharp (see Fig. 3) and blunt force trauma, gunshot wounds (see Fig. 4), strangulation and in circumstances in which a possible interpretation of events is required, such as cases where there is a requirement to determine if the event was peri-mortem (around the time of death) or postmortem (Byers 2008; Klepinger 2006; White and Folkens 2005). If the remains in question retain soft tissue of any kind, the forensic anthropologist will, as a matter of course, macerate the remains in order to access the bone tissue. However, in relation to the latter scenario, new and less invasive approaches through 3D imaging techniques are currently being developed (see Current Challenges and Future Directions section).

The profiling process is based on applying metric and nonmetric (morphological characteristics) assessment when appropriate, as both approaches are normally utilized in a complementary fashion. Ideally, all of the potential methodologies that can be applied should be used, as this will increase the possibility of a positive outcome.

- *Sex determination* is normally the first step taken when producing a biological profile of

Forensic Anthropology, Fig. 1 Skeletonized and fragmented human remains (work conducted in Syria through the Freie Universitat Berlin) (Photo by author)

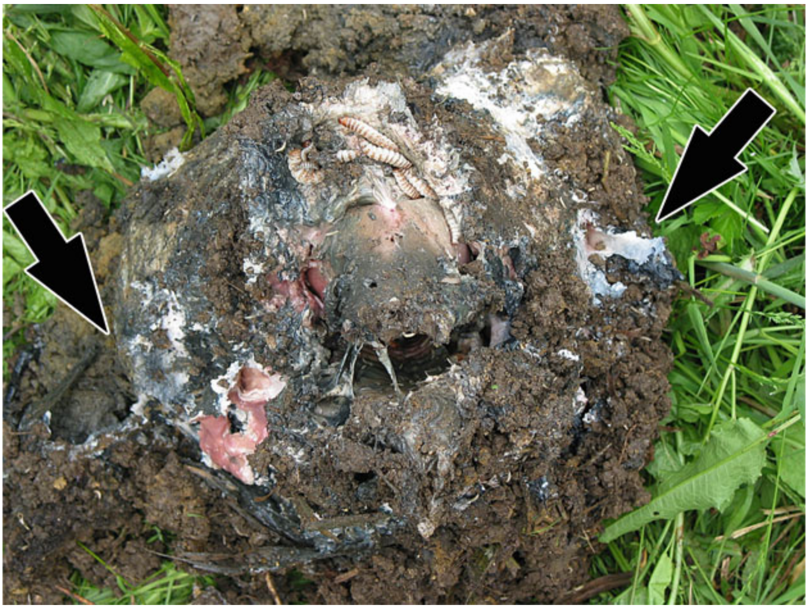

Forensic Anthropology, Fig. 2 Soft tissue in a state of decomposition with areas exhibiting adipocere formation, appearing as white/gray patches (experimental case using buried pig remains, *Sus scrofa*) (Photo by author)

human remains, as this element is also linked to the estimation of ancestry and the determination of stature (refer to appropriate sections below).

Determination of sex may be arrived at by morphological and metric analysis of adult skeletal remains (defined as those with finalized bone and dental development and which exhibit secondary sexual changes). Remains belonging to juveniles, also referred to as adolescents (characterized when secondary sexual changes are taking place) or younger individuals, are not normally sexed; in such cases, the skeleton is still going through a process of development, and sexually dimorphic characteristics (sexual dimorphism being the

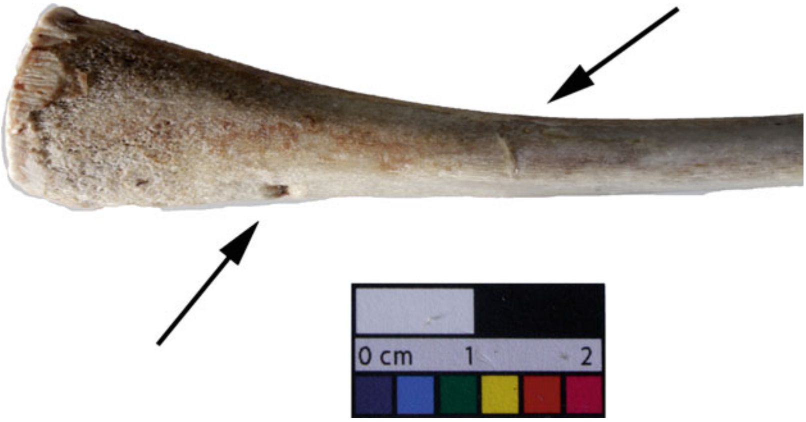

Forensic Anthropology, Fig. 3 Sharp force trauma caused by a straight edge knife wound to the thorax region (experimentation conducted on pig remains, *Sus scrofa*) (Photo by author)

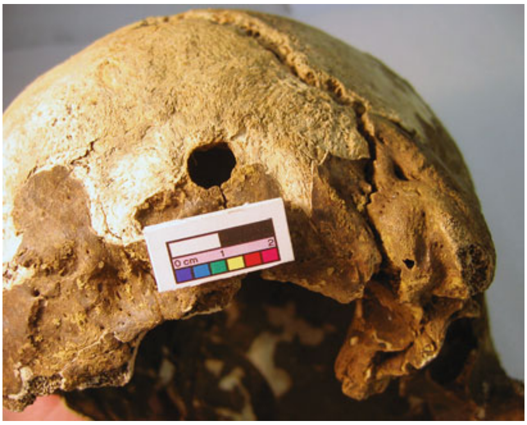

Forensic Anthropology, Fig. 4 Body within mass grave exhibiting a gunshot wound to the occipital region of the head (Courtesy of the Association for the Recuperation of Historical Memory, ARMH, Spain)

difference in size and shape between males and females) have not fully developed (Scheuer and Black 2000; Nafte 2000); when attempts at the determination of sex in adolescents are made, the results generally tend to be weak. Another consideration within the determination process is that traits may vary somewhat from one population to another, with some being more dimorphic than others; it is through accumulated experience that a forensic anthropologist gains insight into the specifics evident within any given group.

Various individual bones have been recently studied with the aim of developing new techniques pertaining to sex determination, with some approaches producing better results than others; however, the pelvis and the skull are still considered as the two areas of the body which are most reliable in this respect. In the case of the pelvis, composed of a sacrum and a right and left os coxa (hip bone), analysis through the Phenice's method, a morphological assessment of the pelvis, is reported to be as accurate as 96–100 % (Silva Braz 2009; Mays 2010), as this structure possesses well-defined sexually dimorphic traits associated with the ability of childbearing in females.

With reference to the skull, the major attributable differences are due to males being larger and with more robusticity, translating into noticeable muscle insertions upon bone, with competent analysis producing a rate of 92 % accuracy (Mays 2010).

Conversely, the metric approach uses a series of measurements on specified landmarks on bone. When using discriminant functions, the results are fed into a set formula that provides an ending score, which is subsequently compared against the sectioning point (determining the sex based upon the score obtained). A problem that is frequently encountered with available formulas is that they are population based, and as such, the results can be skewed if the remains do not belong to the relevant database. Moreover, the metric analysis can also be inputted into the computerized program Fordisc 3 (FD3), based on contemporary data gathered from the Forensic Data Bank at the University of Tennessee in Knoxville, USA

(Sauer and Wankmiller 2009; Tersigni-Tarrant and Shirley 2013). The negative side of using FD3 is that the said database is also population specific, with the possibility of the same problems arising as mentioned above.

- *Age determination* is a step that is applied to subadult and adult remains alike, in order to arrive at the age of death through morphological observations and metric analysis applied to dental and bone remains.
  - Subadults – subadult remains offer a wealth of information as the growth development follows a set sequence of events at the bone and dental level. These parameters have been cross-referenced with age and have been converted into what are referred to as growth standards, which are used by the forensic anthropologist in comparison with observations made on the final growth stage evident upon the remains being examined (Nafte 2000; Klepinger 2006). Because these processes of development tend to be standard in nature across populations, the overall results are highly accurate.

    Dental development may be appreciated through tooth mineralization and eruption and constitutes the most accurate method for aging as it is a constant and continuous process which has been well documented. Dental remains may be examined individually if found dislodged from the alveolar process (the socket which holds each dental piece in the upper and lower mandibles); furthermore, in cases when teeth have not erupted, they may be analyzed through x-rays or extracted and examined microscopically (Klepinger 2006). In the latter case, it is possible to determine if an infant was stillborn or a live birth as evidenced by the neonatal line, located upon the enamel surface on all deciduous dentition and first permanent molars, formed at birth or soon afterwards (Scheuer and Black 2000).

    Bone development can be assessed in subadults, and in the case of fetal remains, the bones can be measured from prescribed landmarks and compared with established growth standards that indicate the gestational age with relatively good confidence levels (Scheuer and Black 2000; Fazekas and Kósa 1978). It is after birth, however, that intrinsic and extrinsic factors play a strong role in skeletal development, creating variations in developmental rates at individual and population levels. As indicated by Byers (2008), age assessment in relation to the length of long bones may be applied up until the age of 10, at which time the correlation between age and length starts to widen, producing inaccurate age estimations.

    Another method utilized to establish age at death is through the identification of primary ossification centers (areas where bone is deposited), as well as their union with other primary centers or with secondary ossification centers (see Fig. 5). Because of their small size, the

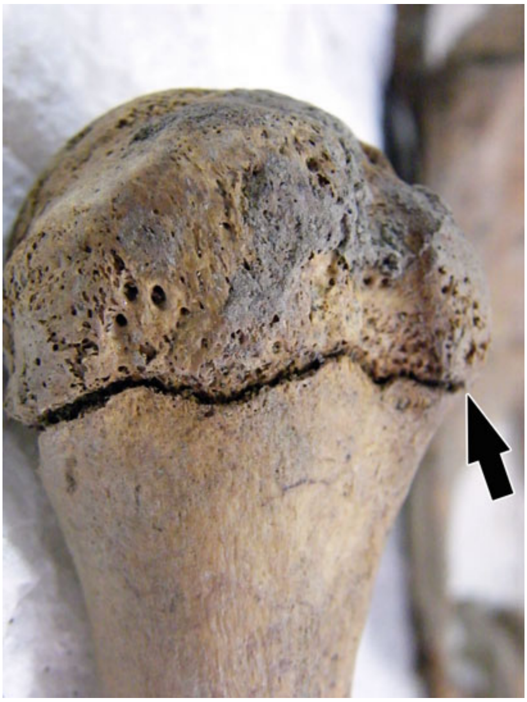

Forensic Anthropology, Fig. 5 Immature arm bone (humerus) with the epiphysis in process of ossification (Courtesy of the Association for the Recuperation of Historical Memory, ARMH, Spain) identification of primary centers may prove difficult to recognize and recover at a scene when dealing with skeletonized remains or bodies in which integrity has been compromised; this method is best employed in connection with examination of the skull, mandible, upper cervical vertebrae, and ribs. On the other hand, the union between the primary and secondary centers is frequently more productive within forensic anthropology, as they unite at a scheduled rate (as indicated earlier with individual variations and population trends) extending into early adulthood; such is the case of the clavicle, in which final fusion ends at approximately 29 years of age (Scheuer and Black 2000; White and Folkens 2005).

– Adults – in the case of adult remains, age is determined by morphological changes visible through the naked eye, which in some instances may be degenerative in nature, and by remodeling processes at the microscopic level. The main areas within the skeleton that aid in age determination through morphological changes are the pubic symphysis, the auricular surface of the ilium upon the os coxa, the sternal end of the fourth right rib, and the cranial sutures (White and Folkens 2005; Byers 2008; Klepinger 2006). Additionally, it is also possible to obtain an estimate of chronological age through dental analysis of the incisors, canines, and premolars as specified by the Lamendin's technique.

In the case of the first three skeletal areas, these are widely used, as a process of "metamorphosis" or changes in the bone morphology occurs progressively as a person ages, associated with sex, and also to a degree with respect to population affiliation; these changes possess set characteristics which are compared against a set of representative age stages. In the case of the pubic symphysis, various methods are available including the Suchey-Brooks model (White and Folkens 2005), which is widely utilized as it constitutes a multiracial database, but is, however, not wide enough to encompass all populations; as such, problems have been reported in the past in connection with skewed results in relation to populations that are not represented within its database.

With regard to cranial sutures, the skull is examined at specific points in order to assess the degree of closure, whereby a prescribed value is scored for each area examined; these values are then compared against an existing table to obtain an age range. Because closure is not a linear occurrence associated with age, the results may indicate wide age ranges making this method somewhat unreliable, such as within cases in which only the skull is recovered without dental remains, or present in poor condition and not suitable for ageing.

Dental remains are normally utilized with relation to subadults in order to obtain a chronological age at death in association with the developmental stage in evidence; however, the Lamendin's dental aging technique was designed to more accurately determine the age of adult dental remains. The latter technique consists of examining single rooted teeth based upon their degree of root translucency and the degree of periodontosis (recession of the gum line). Research indicates that results are linked to sex and, to a lesser degree, to ancestry; however, this method may not be desirable if applied to young adults. Results may also be influenced by additional variables such as dental hygiene and dental treatment intervention (Prince and Ubelaker 2002).

Finally, the histological method consists of examining changes in bones and the teeth at a microscopic level; however, this method is time consuming and requires the use of specialized equipment and specific training (Byers 2008), and as such it is not always a viable option in some cases.

- *Biological affinity* or ancestry is the most difficult task that a forensic anthropologist attempts to evaluate when biologically profiling human remains, producing results which may be somewhat tenuous; however, the estimation of ancestry within an investigation is a necessary element, as it provides an avenue of inquiry for the authorities. Additionally, this information is also required when calculating stature (see section on Stature).

More specifically, the challenge arises from the element of phenotypic traits (expression of the genetic makeup) reflected at the skeletal level in one given population being small in number, but which may also be shared in part by other populations with added variation, thereby creating the possibility of a continuum wherein boundaries are blurred, not to mention the inherent variability that exists at the interpopulation level. Furthermore, sex-determining traits can influence the traits associated with estimating ancestry (Klepinger 2006). Additionally, individuals within any given population may describe or view themselves not in biological terms but by device of ethnic/social classification. For example, many Latin Americans would describe themselves as white or Hispanic; however, their morphological traits would not necessarily be categorized as white, and Hispanic is a term stemming from an ethnic category, something that cannot be verified biologically. To further confound this issue, the high level of migrant influx and intermarriages within any given geographical zone will intensify the challenges with respect to identifying any given question of biological affinity.

At present, global populations are divided into different groups, deriving from genetic or skeletal research, whether referring to the global population as a whole or to a specific population within a particular country. At the skeletal level, four main ancestry categories have been created: Caucasoid (white American/European) (white American/European), Mongoloid (East Asian), Negroid (African black), and Native American (Byers 2008; Klepinger 2006; Tersigni-Tarrant and Shirley 2013). As it may be appreciated, this categorization does not embrace all of the global geographical regions.

Attempts to assess ancestry through traits present on bone are accomplished primarily through morphological or metric analysis of the craniofacial region, as this is where traits are most effectively expressed, mainly with reference to the shape of the head and nasal region and additionally the upper mandible. Within a given morphological analysis, a sum of characteristics is sought and subsequently compared to known samples to arrive at a determination (Byers 2008; Klepinger 2006), a task best accomplished by an experienced anthropologist. Because the estimation is based upon morphological observations, and not all traits may be strongly reflected within one particular individual, the results obtained may be argued over due to the element of subjectivity. Conversely, a metric analysis compiled between specific landmarks can be perceived as a more accurate, objective approach.

Metric analysis can also be fed into the FD3 software, which contains data for this purpose drawn from contemporary remains from the USA, in addition to world samples, and representative of various archaeological populations (Tersigni-Tarrant and Shirley 2013); as well, another computer programme available to forensic anthropologists is CRANID, with a data base comprised of 74 samples (3.163 skulls) from various regions of the world (Sauer and Wankmiller 2009; Wright 2012). However, the application may not provide adequate results when the affinity of the specimen in question has not been represented within the available databases.

Postcranial bones may be used, with the femur (leg bone) being considered as being the best choice as it possesses various traits and has been comprehensively studied (Byers 2008; Sauer and Wankmiller 2009). However, in general terms, the postcranial bones retain fewer traits which are not as clearly expressed as in the craniofacial region, thereby reducing confidence levels with respect to accuracy.

Research has been attempted upon dental material, but this particular avenue of inquiry has not been as comprehensive as that completed with respect to bone; additionally, dental traits are fluid, varying in degree, and not strictly present within any given population.

For the reasons stated herein, it is always recommended to collate results from all available methods and to be actively aware that a definite determination of ancestry from the remains is not always possible, but predominantly an educated estimation.

- *Stature* – calculating the stature at the time of death is accomplished by measuring a selected long bone from the upper or lower extremities, with the result computed into a regression formula that has been designed by bone type, sex, and biological affinity; a proportional correlation may then be compared between the bone used and the height of the unknown individual. The result obtained will constitute the mean height with a standard deviation. The femur and tibia are the most widely utilized reference points, as the results obtained from these bones provide more accurate outcomes (Byers 2008).

In cases in which the body is highly incomplete or fragmented, such as when dealing with dismemberment, air crashes, or the victims of terrorist acts, there are a series of methods available that allow estimation of the stature of any given individual through the use of small segments from the long bones (Byers 2008). Because of the nature of these measurements, the results are not as accurate, but are still helpful in cases where an approximation of height is required. When skeletal remains are dry, due to time lapse, shrinkage can occur, and adjustments can be necessary in order to compensate for such factors through preestablished data; the same approach may be taken with remains that are determined to be over 45 years of age, as stature starts to diminish beyond this period (Byers 2008). It should also be clearly stressed that all populations are not represented within all available formulas, and as such, adjustments must be made and considered when confronted with each particular case.

The results obtained from such calculations often prove difficult to verify, as accurate reference to stature is information that is not often available through family members and friends or via official records.

- *Individual characteristics* – individual traits can aid in an investigation for two reasons: when there are no indications of who it may be or when the authorities have a suspicion of who the person may be. Such traits may include evidence of a fracture or amputation that the person may have suffered in life, with time to heal through bone remodeling, including misalignment of segments which produce evidence of the event (Byers 2008).

Other traits include pathologies left on bone that can indicate past conditions, diseases, or habits, as in the case of osteoporosis upon the skull, associated with malnutrition and drug addiction (Klepinger 2006); an infected tooth may cause an abscess which produces a small opening on the jaw bone. Diseases may also affect bone, such as tuberculosis, a chronic infection caused by a bacteria which can leave marks in the vertebral column, and in other bones such the os coxae (hip bones). Furthermore, congenital anomalies due to improper formation can aid in the identification of a person and may be present in various parts of the body, such as the vertebral column, hands, and feet (Otner 2003; Klepinger 2006; White and Folkens 2005).

The human skeleton also possesses traits which can assist in the individualization process, as is the case of improper non-fusion of segments conforming a bone, missing bone sections, or characteristics within a particular bone which are not pathological in nature, but which can be identified and compared against an antemortem x-ray or computer tomography image (CT).

All of the above conditions may be used to guide investigations, and with cross-referencing the information gathered in tandem with medical records, which can aid in the identification process.

Facial reconstruction is a method which can aid in the identification process by applying a variety of methods from sculpting in clay to the use of sophisticated computer graphic programs (Wilkinson 2008); such avenues are attempted by individuals from a variety of disciplines, including those from purely artistic backgrounds to forensic anthropologists who specialize within this specific field.

The described steps that a forensic anthropologist takes in order to compile a biological identification can lead to a positive identification through DNA, odontological, and/or medical analysis, once the investigating personnel have a description with which to guide them.

Trauma Analysis

The determination of the cause and manner of death legally rests with the forensic pathologist; however, forensic anthropologists do participate within cases in which bone reconstruction is required due to fragmentation caused by high-velocity projectiles or in connection with blunt force trauma; in both instances, an interpretation and sequence of events can be indicated when possible, as in cases when more than one impact has been inflicted upon the skull. Interpretations concerning the circumstances in which the trauma was caused can be assessed; an example would be when, there is a need to determine whether an individual fell or jumped intentionally or to determine the angle of force in which a chest cavity was crushed, as would be evidenced in cases of severe stamping. Additionally, in the case of sharp force trauma, the type of weapon used, such as a knife (see Fig. 3), axe, hatchet, or machete, may be accurately identified, in addition to the angle with which the implement was used. In the case of dismemberment through sawing, the type of striations deposited upon the cut site can aid in a more specific identification of the tool which was utilized (Byers 2008; Galloway 1999; Nafte 2000).

In cases where strangulation was the cause of death, a close examination of the hyoid bone (located in the anterior side of the neck) may show signs of subtle breakage which can be interpreted by an experienced anthropologist. With reference to children, this bone is very resilient as it has not ossified; hence breakage may not occur (Byers 2008; Klepinger 2006).

Depending upon the condition in which a body is discovered, including burnt remains, there may be a requirement for the anthropologist to determine if a given alteration on bone is peri-mortem or postmortem, a detail which may produce ramifications within a legal investigation; in some cases this element may be difficult to assess, necessitating specialized analysis.

Investigations into Human Rights Violations

Investigations of human rights violations have given way to interventions in countries such as Guatemala, El Salvador, Colombia, Chile, Argentina, Democratic Republic of Congo, Rwanda, East Timor, Bosnia-Herzegovina, and Spain, where forensic anthropology has played an important part within multidisciplinary forensic teams. Among the many challenges encountered prior to deployment, one aspect that concerns work within forensic anthropology is to select the appropriate methods to be used as references for determining the biological profile of the victims, keeping in mind the genetic makeup of the population to be examined (as indicated above) (Ferllini 2007; EAAF 2013); as previously highlighted, inherent variations exist within populations and all references utilized should be directed at the required analysis in order to avoid the production of skewed results.

Within this arena, work is accomplished both within the field and at the mortuary. In the field, a comprehensive knowledge of bone anatomy is crucial during the search for human remains which may have been left on the surface and thereby exposed to environmental elements. It is also advantageous to be able to recognize small remains and distinguish them from animal bones within an outdoor setting. In situations when bodies are comingled and dispersed, an attempt is made to determine the number of individuals that are present while in situ, aiding in the implementation of proper lifting and bagging procedures prior to sending the remains to the assigned mortuary, which in some instances may be a temporary installation in close proximity to the area being surveyed.

In the case of buried remains, the forensic anthropologist is also directly involved with the exhumation process, whether dealing with the contexts of single or mass graves. With reference to the latter scenario, hundreds of bodies may be present, resulting from a single event, as was the case after the Rwandan genocide (1994), and where the remains may be comingled and skeletonized. In such cases the forensic anthropologist determines which bones pertain to any given individual by applying analysis of bone morphology and determining a general sex and age identity, elements which are verified at the mortuary (Ferllini 1999; Koff 2004). However, this is only possible in cases where the bodies have been interred with small amounts of clothing or have subsequently degraded, with the bones becoming free.

During mortuary work, it is customary that the bodies are subjected to fluoroscopy imaging in order not only to detect metal objects, including evidence of gunshot wounds, but also to determine whether the remains may have been booby-trapped by the perpetrators with the aim of dissuading others from recovering the remains, or to directly harm those attempting such work. The descriptors produced from work in the field are subsequently verified and a final biological profile is produced. With reference to cases in which discrepancies of any type arise, an open communication is maintained with the lead forensic archaeologist of the rescue operation (Ferllini 1999). As with police cases, identifying traits are also sought, and DNA testing is applied when possible. The ultimate objective is to be able to not only return the remains to the surviving relatives but also produce documented scientific proof of the atrocities committed against a given civilian population, to be used in future judicial proceedings, which may be held at the local or international level (Ferllini 2007).

Investigations into Disaster Victim Identification (DVI)

Disaster victim identification (DVI) is at present a serious concern, as mass disaster events are more severe in nature and in consequence, the numbers of victims may be in the thousands, including foreign nationals as with the World Trade Center disaster in 2001, and the Asian Tsunami of 2004. For the forensic anthropologist, their participation within such scenarios is mainly at the mortuary, involving the sorting of comingled remains and including the identification of small bone fragments, sometimes in a burnt state; determination of the number of individuals at hand, followed by a biological profile on each case when possible, and identifying any bone trauma constitute the eventual objectives (Blau and Briggs 2011; Byers 2008).

However, the anthropologist does not operate within a vacuum, but within a multidisciplinary team, in which each set of remains is sent to one or more stations to be examined by other forensic specialists who can also contribute to the identification process, in areas such as odontology and DNA analysis. The results are subsequently cross-checked with personal belongings associated within the context of each individual, with the aim of arriving at a positive identification of each of the victims (Byers 2008). The challenge increases within situations in which a floating population is present, meaning that the true number of victims may be difficult to determine, and unlike the scenario of an aircraft accident, there is often no list of victims available as a reference.

The techniques applied during the biological profiling process must be verified against standards from varied population databases due to the potential nature of the victim population. Furthermore, in many instances the remains may present soft tissue, and the resulting de-fleshing process may be extremely time consuming. Recently, new techniques with reference to digital imaging have been proposed for the purpose of the latter activity (Rutty et al 2009) (see section on Conclusions and Future Research).

Current Challenges and Future Directions

At present, due to the elevated level of influx of immigrants throughout the globe, whether via illegal clandestine means, in addition to localized increases in murder rates, potential legal investigations are often hindered by the lack of proper identification papers and antemortem medical and dental records and the possibility that no missing person's reports exist. This situation dictates that every forensic anthropologist must possess a working knowledge of the available morphological and metric data for various ancestry groups, as it is no longer tenable for a professional within this field to focus solely upon a specific population which relates only to his or her "normal" casework; additionally, all populations are highly influenced by secular, biological, and cultural trends.

Furthermore, constant warfare, the utilization of weapons with high destructive capabilities, the ever constant violations of human rights with associated mass graves, and increasing levels of mass fatality incidents dictate that the discipline further develops more efficient methods of processing high numbers of potential casualties, which in many cases may be fleshed. This is reflected in the approaches now taken within the present millennium with respect to the utilization of digital imaging technologies in order to examine and profile human remains; the examination of skeletal elements by such means can be performed within a shorter time frame, as images may be obtained from bodies contained in bags of various kinds, both clothed and fleshed. Therefore, time is not wasted in de-fleshing the remains (Rutty et al 2009; Blau et al 2008; Dedouit et al 2007), and the possibility of damaging bone and adding a pseudo trauma indicator is avoided. Furthermore, this technology is mobile and creates the opportunity of sending information on a global scale directly from the field via a remote basis (Rutty et al 2009).

Such digital imaging techniques originated from within clinical settings and are now being applied to forensic postmortem examinations. Postmortem computer tomography (PMCT) is rapidly gaining acceptance and is being used more on a global scale, thereby replacing the traditional use of x-rays, as the former technique is more interactive, allowing the images to be manipulated in response to the needs of the examiner; in turn, the profiling can be accomplished with a good degree of success and is comparable to that obtained through conventional methods (Verhoff et al 2008).

The use of a multi-slice computed tomography (MSCT), also known as multidetector computed tomography (MDCT), allows the body to be viewed without the requirement for maceration and to view any bone element within the skeleton in 2D and 3D for the purposes of profiling, including the taking of measurements which apply to traditional parameters, and the assessment of trauma analysis (Rutty et al 2009). Ante- and postmortem image correlation can also be gained in order to arrive at a positive identification; finally, facial reconstruction may also be achieved from 3D images (Verhoff et al 2008).

While such new technologies constitute positive additions to postmortem examinations, such types of infrastructure will not be readily available within every area of the globe when required; such methods are often costly and will not normally be obtained in many developing nations unless intervening countries aid with adequate equipment.

Conclusions and Future Research

Forensic anthropology has come a long way since the days when isolated bone cases were passed on to the physical anthropologist for analysis. At present, the discipline forms part of a wide range of scenarios, whether pertaining to the remains of one or more individuals within a variety of situations and conditions, and by applying techniques which have evolved from the traditional setting to the use of sophisticated modern imaging technology.

This is not to say that as a discipline forensic anthropology has reached its zenith and that it does not require constant refinement with respect to the techniques it utilizes. As with any science, improvements upon widely used techniques and the creation of innovative new approaches are always required to drive the process within a dynamic context, as in the case pertaining to population levels with diverse secular changes and the factor of immigration, in addition to the intense scrutiny that any forensic scientist now faces within a court of law.

Due to present global challenges, research is required pertaining to quantitative methods of determining human variation, encompassing the areas of sex, age, ancestry, and stature estimation, all of which are essential in producing biological identifications, not only to improve existing applications but to develop new methodologies (Ross and Kimmerle 2009), and encompassing work with fragmented and incomplete remains.

Additionally, the development and refinement of new and existing imaging techniques is desirable, as such elements not only speed up the profiling process and shield the participating investigators from potential contaminants, but also take into account religious and cultural beliefs in which body handling and avoidance of autopsy procedures are an issue.

The combination of quantitative methodologies with the additional use of CT technology can be of advantage for the reasons already stated here, but also to obtain knowledge of contemporary populations (as long as this complies with local legislation), in order to study nonmetric and metric traits within populations, as modern skeletal collections have become a thing of the past. Populations within Latin America, the Middle East, and Africa need to be addressed by developing their own standards, which will subsequently provide more accurate results.

Related Entries

► Cognitive Forensics: Human Cognition, Contextual Information, and Bias

Recommended Reading and References

Blau M, Briggs C (2011) The role of forensic anthropology in disaster victim identification. Forensic Sci Int 205(1–3):25–35

Blau S, Robertson S, Johnstone M (2008) Post mortem computed tomography. J Forensic Sci 53:956–961

Byers S (2008) Introduction to forensic anthropology, 3rd edn. Pearson, London

Dedouit F, Telmon N, Guilbeau-Frugier C, Gainza D, Otal P, Joffre F, Roug D (2007) Virtual autopsy and forensic identification – practical application: a report of one case. J Forensic Sci 52(4):960–964

Equipo Argentino de Antropologia Forense (EAAF). http://www.eaaf.org/. Accessed 15 Feb 2013

Fazekas G, Kósa F (1978) Forensic fetal osteology. Akadémiai Kiadó, Budapest

Ferllini R (1999) The role of forensic anthropology in human rights issues. In: Fairgrieve S (ed) Forensic osteological analysis: a book of case studies. Charles C. Thomas, Springfield, pp 287–302

Ferllini R (ed) (2007) Forensic archaeology and human rights violations. Charles C. Thomas, Springfield

Galloway A (ed) (1999) Broken bones: anthropological analysis of blunt force trauma. Charles C. Thomas, Springfield

Gill-King H (1997) Chemical and ultrastructural aspects of decomposition. In: Haglund W, Sorg M (eds) Forensic taphonomy: the postmortem fate of human remains. CRC Press, Boca Raton, pp 93–108

JPAC. http://www.jpac.pacom.mil/index.php?page=mission_overview&size=100&ind=0. Accessed 15 Feb 2013

Kennedy K (2000) Becoming respectable: T. Dale Stewart and the acceptance of forensic anthropology in the academic community. J Forensic Sci 45(2):253–257

Klepinger L (2006) Fundamentals of forensic anthropology. Wiley-Liss, Hoboken

Koff C (2004) The bone woman: among the dead in Rwanda, Bosnia, Croatia and Kosovo. Atlantic Books, London

Mays S (2010) The archaeology of human bones, 2nd edn. Routledge, London

Nafte M (2000) Flesh and bone: an introduction to forensic anthropology. Carolina Academic Press, Durham

Otner D (2003) Identification of pathological conditions in skeletal human remains, 2nd edn. Academic Press, New York

Prince D, Ubelaker D (2002) Application of Lamendin's adult dental aging technique to a diverse skeletal sample. J Forensic Sci 47(1):107–116

Ross A, Kimmerle H (2009) Contribution of quantitative methods in forensic anthropology: a new era. In: Blau S, Ubelaker D (eds) Handbook of forensic anthropology and archaeology. Left Coast Press, Walnut Creek, pp 479–489

Rutty G, Robinson C, Morgan B, Black S, Adams C, Webster P (2009) Fimag: the United Kingdom disaster victim/forensic identification imaging system. J Forensic Sci 54(6):1438–1442

Scheuer L, Black S (2000) Developmental juvenile osteology. Academic, London

Sauer N, Wankmiller J (2009) The assessment of ancestry and the concept of race. In: Blau S, Ubelaker D (eds) Handbook of forensic anthropology and archaeology. Left Coast Press, Walnut Creek, pp 187–200

Silva Braz V (2009) Anthropological estimation of sex. In: Blau S, Ubalaker D (eds) Handbook of forensic anthropology and archaeology. Left Coast Press, Walnut Creek, pp 201–207

Stewart T (1979) Essentials of forensic anthropology, especially as developed in the United States. Charles C Thomas Publisher, Springfield

Tersigni-Tarrant M, Shirley N (eds) (2013) Forensican anthropology: an introduction. CRC Press, New York

Ubelaker D (1999) Aleš Hrdlička role in the history of forensic anthropology. J Forensic Sci 44(4):724–730

Verhoff M, Ramsthaler F, Krähahn J, Deml U, Gille R, Grabherr S, Thali M, Kreutz K (2008) Digital forensic osteology – possibilities in cooperation with the Virtopsy® project. Forensic Sci Int 174(2–3):152–156

White T, Folkens P (2005) The human bone manual. Elsevier Academy Press, London

Wilkinson C (2008) Forensic facial reconstruction, 2nd edn. Cambridge University Press, Cambridge

Wright R (2012) CRANID. https://osteoware.si.edu/forum/osteoware-communityannouncenments/cranid-richard-wright-0. Accessed 19 Feb 2013

Forensic DNA Matching

► DNA Profiling

Forensic Environmental Evidence

Ruth M. Morgan[1] and Peter A. Bull[2]
[1]Security and Crime Science, University College London, London, UK
[2]Oxford University Centre for the Environment, University of Oxford, Oxford, UK

Overview

Forensic geoscience investigates the characteristics of sediments and soils and uses them to compare exemplar and specimen samples. They derive ultimately from both naturally occurring igneous, metamorphic, and/or sedimentary rocks, which maintain the characteristics of those parent materials, and also as admixtures of man-made constituents which in the main make up much smaller quantities within that sediment body. Some of these sediments are found in bulk form on beaches, outwash plains, etc., whereas others are found as trace particulates in both the natural and anthropogenic environments as dust, soil, and detritus. Soils represent sediments with organic additions, both floral and faunal, and so acquire not only the geological characteristics of the sediment but also incorporate materials such as palynomorphs, diatoms, mycological spores and fungi, and other micro- and macrofaunal and floral debris.

At a larger scale, an appreciation of the geomorphological and geological history of a region is pertinent for forensic characterization as is an understanding not only of the microscale characteristics of the forensic material but of the macroscale environment in which those materials exist (Ruffell and McKinley 2008). The macroscale landscape does not only have natural origins but also has anthropogenic modifications which are so important in forensic investigation. Macroscale landscapes can involve rural and urban environments, and indeed, investigations set purely within a building or dwelling can utilize geological evidence as well as anthropogenic additions incorporated within samples (Pringle et al. in press; Morgan et al. 2009).

Soils and Sediments

The dynamic nature of the natural world leads to the mixing of materials from very many provenances in discrete ways, and the addition of anthropogenic materials can characterize a particular forensic site to such an extent that it becomes highly distinctive and very useful when compared with samples derived from other pertinent locations. Thus, natural and anthropogenic soils and sediments are not mixtures of a myriad of provenances but are discrete and ordered micro-terrains and facies, worthy of the increased attention being given to these materials in the published literature (Morgan and Bull 2007a).

The concept of facies and sedimentary units is a keystone to the analysis of soils and sediments in forensic inquiry. A sedimentary facies has long been understood to be any spatially restricted part of a designated stratigraphic unit which exhibits discrete characteristics that are distinguishable from the other parts of the unit. The most important application of this is that sediments (both natural and anthropogenic) are not found to be in a melange of indiscrete units but rather the interrelationships of sedimentary environments, and hence, facies are not chaotic or random (Bull and Morgan 2006), but subject to controls such as climate, tectonics, and geological setting giving rise to discrete and ordered micro-terrains in both areal and profile contexts so that soils and sediments have great spatial variability at different scales of reference (Thomas 2001). As no two environments, however similar, are exactly identical, this characteristic of soils and sediments enables the use of comparison methods in the analysis of trace physical evidence to be used to great effect, and it is such comparison methods that are generally agreed to be central to forensic inquiry (Lane 1992). Due to the nature of the spatial, and indeed vertical, distribution and variation of discrete facies, soils and sediments have the ability to be compared with other samples and thus have the potential to be used as a predictive tool in forensic investigation. However, as the concept of the individualization of soils and sediments is problematic in the forensic context, we can only exclude and therefore state that a soil sample did not originate from the crime scene or indicate an inability to exclude samples taken from two discrete locations (i.e., the analysis results indicate that the two samples are consistent with having the same source).

In recent years, the importance of soils and sediments as physical evidence in forensic investigations has been more widely recognized. The writing of “Forensic Geology” by Murray and Tedrow (1975) brought the potential of soil and geology in forensic investigations to a wide audience. Textbooks on forensic science now increasingly make reference to soil evidence. However, it is very rare that more than a few pages are dedicated to the topic, and comparatively little is included about the analytical techniques that may be applied to such samples. It is a relatively new part of forensic science which has developed by means of resourceful geoscientists applying traditional geology in the forensic arena. While there are many examples of geologists and geomorphologists successfully aiding forensic investigations (at the microscale Murray and Tedrow 1975, and at the macroscale Ruffell and McKinley 2008), the use of soils and sediments has developed comparatively rapidly, and only recently has a specifically geoforensic philosophical framework been articulated and incorporated into the discipline (Morgan and Bull 2007b, c). The main problems encountered during the development of chemical analyses of forensic soil samples have been the complexity of the chemical nature of soils (which include complex organic acids and other additions). This issue has been compounded also by the use of automated chemical analysis techniques (e.g., ICP-MS) where false-positive or false-negative results can occur without recognition and therefore without correction, as a result of the necessary homogenization of the sample in preparation for the analysis.

Forensic Geoscience

Forensic geoscience utilizes the spatial and temporal variations found in soils and sediments to aid crime reconstructions. The distinctiveness of terrains at a variety of spatial scales reflecting the specific environmental and geological influences operating in particular places results in site-specific soils and sediment characteristics. Soil evidence has potential value in forensic investigations because of its prevalence at scenes of crime and also its propensity for transfer between crime scenes, victims, and perpetrators. A sample of soil from clothing, a vehicle, or a scene of crime can therefore be analyzed to identify its physical, chemical, and biological components, thereby enabling the indication of provenance within the forensic context.

Some common standard analyses can be seen in the published literature which distinguishes between soil and sediments derived from different source locations. These are outlined in Table 1.

The examination protocol for soil evidence will be dependent on the quantity of the sample available, the questions asked of the sample, and

Forensic Environmental Evidence, Table 1 Different forms of analysis documented in the published literature

Aspect of soil/sediment	Technique	Literature
Physical	Color	Guedes et al. (2011)
	Mineralogy	Bull et al. (2006)
	Particle size distribution	Morgan and Bull (2007a)
	Density gradient	Petraco and Kubic (2000)
	Heavy minerals	Palenik (2007)
	S.E.M	Fitzpatrick and Thornton 1975
	S.E.M. quartz surface texture analysis	Bull and Morgan (2006)
Chemical	Elemental chemistry	Rawlins et al. (2006)
	Ion Beam Analysis	Bailey et al. (2012)
	Chromatography	Siegel and Precord (1985)
	Organic content	Cox et al. (2000)
Biological	Pollen	Riding et al. (2007)
	Enzymatic classification	Thornton and McLaren (1975)
	Bacterial DNA	Horswell et al. (2002)
	Plant wax signatures	Dawson et al. (2004)
	Mycology	Hawksworth and Wiltshire (2011)
	Limnology	Cameron (2004)
	Microorganisms	Meyers and Foran (2008)
Anthropogenic	Lighter flint	Bull et al. (2006)
	Brick, glass beads	Morgan et al. (2007)
Combined	Palynology and mineralogy	Brown et al. (2002)
	Median particle size, modal class interval of particle size and organic matter	Wanogho et al. (1985)
	Mineralogy, particle size, and biological factors such as plant fragments, diatoms, and pollen.	Hunter et al. (2001)
	Sand, soil, building and road materials, vegetal fragments, paint, and fibers	Lombardi (1999)

the detection limits required. The general Palenik's sequence (outlined in Fig. 1) is often presented in the literature as a guideline for analysis that can be adapted for each separate case.

While such sequences are of great value in the geosciences, in forensic investigations, the quantity of the sample will often be a limiting factor, precluding the use of many steps included in such analytical sequences, and therefore, the forensic geoscientist needs to choose the relevant parts of the Palenik sequence that are appropriate for the samples being examined.

From a forensic (as opposed to a geological) perspective, the temporal scales involved are relatively short, even when the investigation of cold cases is considered. It is possible in some cases to return to a crime scene many years or decades after the crime took place and characterize the soils and sediments present and compare them with the original crime scene samples and indeed the relevant samples derived from the suspect. It has been demonstrated that certain aspects of soils and sediments may be subject to changes over these relatively short-time frames (such as the microbial communities present within soils (Meyers and Foran 2008)), but other biological components are often not subject to such significant changes over these time frames, such as the palynomorph and flora and fauna assemblages, and indeed similarly, the physical components such as mineral assemblages and characteristics.

Much of the early key literature in this field was based on case-based examples where forensic geoscience provides useful and pertinent intelligence for an investigation or ultimately evidence for a court hearing (Wiltshire and Black 2006;

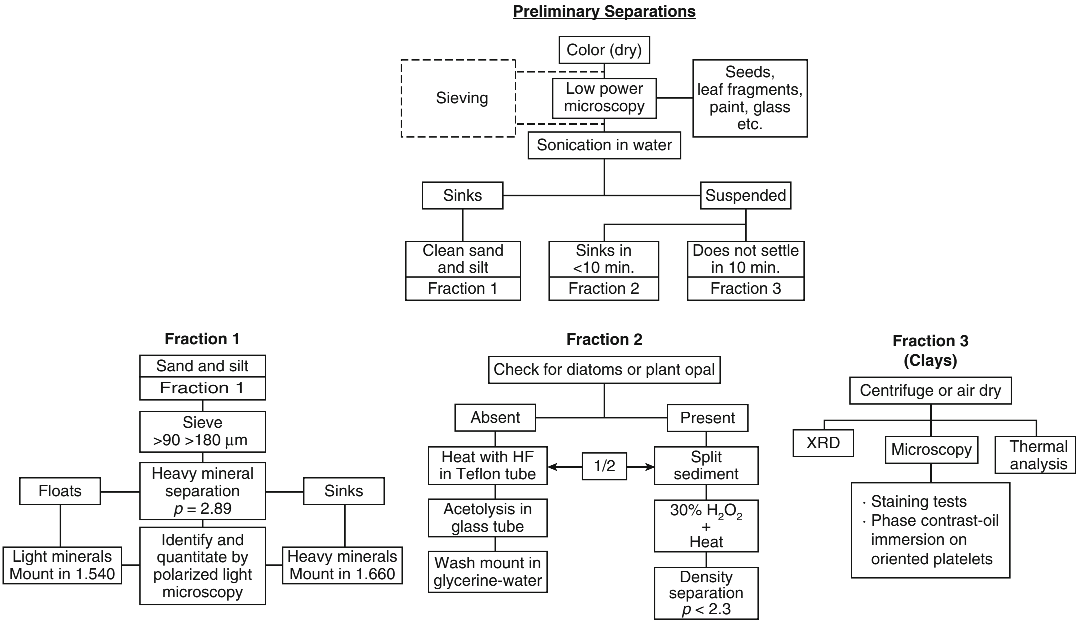
Preliminary Separations
Color (dry)
Sieving
Low power microscopy
Seeds, leaf fragments, paint, glass etc.
Sonication in water
Sinks
Suspended
Clean sand and silt
Fraction 1
Sinks in <10 min.
Fraction 2
Does not settle in 10 min.
Fraction 3
Fraction 1
Sand and silt
Fraction 1
Sieve >90 >180 µm
Floats
Heavy mineral separation p = 2.89
Sinks
Light minerals Mount in 1.540
Identify and quantitate by polarized light microscopy
Heavy minerals Mount in 1.660
Fraction 2
Check for diatoms or plant opal
Absent
Present
Heat with HF in Teflon tube
1/2
Split sediment
Acetolysis in glass tube
30% H2O2 + Heat
Wash mount in glycerine-water
Density separation p < 2.3
Fraction 3 (Clays)
Centrifuge or air dry
XRD
Microscopy
Thermal analysis
· Staining tests
· Phase contrast-oil immersion on oriented platelets

Forensic Environmental Evidence, Fig. 1 The Palenik examination sequence (Diagram adapted from Murray and Solebello 2002)

Bull et al. 2006). There has also been significant debate within the literature as to the best methods for soil and sediment analysis within a forensic context, from sampling protocols (McKinley and Ruffell 2007) to the various advantages and limitations of particular analytical techniques (Bull et al. 2008; Morgan and Bull 2007c). It is only comparatively recently that a philosophical framework for forensic geoscience has been incorporated into the methods and practices of this discipline (Morgan and Bull 2007b).

Essentially, four key principles exist, those of:

- Exclusion/uniqueness
- Independence of analytical techniques
- Issues of homogenization of samples of mixed provenance
- Exotic/ubiquitous components of soils/sediments

These fundamental tenets are fully outlined in Morgan and Bull (2007b), but perhaps the most important of these tenets is that of the principle of exclusion. Forensic geoscience adheres to the philosophy of falsification (Walls 1968) and is considered to be a probabilistic science. There can be no philosophically correct way to state or strongly infer a match between two soil samples. The characteristics of particular samples may be indistinguishable, but the strongest assertion that can be made is that it is not possible to exclude sample A from having derived from a similar provenance to sample B. The assumption of uniqueness has been widely discussed in a number of forensic contexts such as fingerprint analysis and DNA interpretation and has broadly been demonstrated to be a blind alley to be avoided (Saks 2010). Ironically, both the methodology of fingerprints and DNA forensic analysis were introduced as exclusionary techniques. Subsequently, they are often used to infer a “match” between exemplar and specimen samples.

In order to test the assertion that two samples have derived from different provenances, it is imperative that a number of analytical techniques are employed on the samples and that those analytical techniques are themselves independent of one another. The one-step indicator can be very problematic, particularly in the case of an appeal to higher courts. A full chemical assay of a sample utilizing elemental data, a mineralogical analysis of the same sample, and its color, pH, and conductivity may at the outset appear to be extensive independent techniques of analysis. However, they are all dependent upon the same characteristics of the sediment sample and therefore cannot be considered to be independent techniques. Such comparisons have much less weight than corroborative findings from a number of independent techniques.

In the majority of forensic contexts, mixing of soil/sediment material is likely to have occurred due to use over time of footwear, vehicles, clothing, etc. The forms of analysis undertaken, therefore, must be able to discern when difference between two samples is identified because the soil/sediment is genuinely derived from different provenances and when a difference is identified because either one or both of the samples are of mixed provenance which can thereby illicit a false discrimination (positive or negative). For example, soil/sediment recovered from vehicle footwells has been demonstrated to contain materials from multiple provenances that have been collected over time. If a sample from that footwell were compared with a sample from a crime scene by means of an analytical technique that requires the homogenization of both samples, it would be highly likely that given the multiple sources of material in the footwell, a sample from that footwell would be demonstrated to be different to the crime scene (even if one of the contributing sources of material to the footwell was in fact the crime scene in question) (Morgan et al. 2006). It is therefore vital that collection practices and the analysis that is undertaken on forensic samples can account for sample mixing so that false exclusions are avoided. A compounding problem is that false-positive or false-negative exclusions may not even be recognized after analysis, thus impacting on the interpretations made from the results:

> . . .this is evidence that does not forget. It is not confused by the excitement of the moment. It is not absent because human witnesses are. It is factual evidence. Physical evidence can not be wrong; it can not perjure itself; it cannot be wholly absent. Only in its interpretation can there be error. (Kirk 1974: 2)

The exotic or rare components of soil/sediment samples can be highly indicative of the provenance of that sample particularly when samples are compared to one another. The presence of a rare component in the soil at the crime scene can provide strong diagnostic evidence that two samples cannot be excluded from having derived from a similar source. Examples include rare palynomorphs or suites of palynomorphs, distinctive diatoms, or mineral assemblages (Cameron 2004). However, during such comparative analysis, it is very important that like is compared with like. Often crucial exhibits will only have very small trace amounts of soil/sediment that can be recovered, and this may preclude the inclusion of the exotic/rare component from the crime scene even if a direct contact and transfer has taken place simply due to the trace amounts of sample being analyzed. Most common (or ubiquitous) materials lack diagnostic capabilities in forensic analysis, but there are techniques which can utilize their abundance while presenting distinctive characteristics which aid provenance sourcing, such as environmental profiling and quartz grain surface texture analysis. For example, quartz is one of the most ubiquitous minerals within soils and sediments and is regularly recovered during forensic analysis of exhibits. However, while its presence is ubiquitous, the morphology of the individual grains and the surface features that can be identified on the surfaces of the grains has been demonstrated to be highly distinctive, and the combination of quartz grain types at particular locations is generally considered to be able to provide highly discriminatory conclusions (Bull and Morgan 2006). Thus, the balance must be reached between the value of exotic components and diagnostic capabilities of certain types of more ubiquitous trace components of soil/sediment samples.

Other Issues

Two further issues cut through all forms of soil/sediment forensic analysis, namely, the issues of interpretation and the necessity of empirical experimental studies to establish theoretical frameworks for the practice of forensic science.

The issue of interpreting the analysis of forensic soils and sediments is becoming widely acknowledged, particularly given the high-profile controversies surrounding the interpretation of other forms of forensic evidence such as fingerprints and DNA profiles. While some have called for the avoidance of statistics in the presentation of forensic geoscience evidence in court altogether, more recently there has been a resurgence in interest in assessing the best means of presenting forensic evidence to a court (Law Commission report 2011; Fenton 2011). The desire to be able to provide a measure of the weight of a particular form of evidence within a case can be argued to be part of the bigger question that relates to the philosophical approach to forensic science, and the nature of its position at the intersection of science and the law. Both the law and science are in the business of producing knowledge, but given their very different settings, the production of knowledge in each context necessarily serves different purposes (Jasanoff 2006). Table 2 outlines the different contexts within which the law and science operate. Science primarily concerns itself with testing hypotheses in order to identify general theories. Scientific findings therefore have the aim of advancing our knowledge in a particular field and enabling the future development of the discipline. In contrast the law is generally concerned with legally relevant facts which are by their nature specific to particular cases. It is thus most often observed in the practice of the law that facts relate to specific cases; previous cases are looked to in order to provide knowledge pertinent to the particular case at hand (Jasanoff 2006). *When science enters the courtroom, it should do so as an adjunct to the law's need for credible and meaningful story-telling. In a court of law, science cannot hold itself out as simply science, the source of transcendental truths; more modestly, and with appropriate caveats, it can be the source of just evidence* (Jasanoff 2006: 339). Furthermore, Jasonoff argues that the philosophical approach of science that has been carried out to serve the law should be distinct from the approach of science undertaken purely to advance the frontiers of science.

Forensic Environmental Evidence, Table 2 Contrasts between the law and the sciences

Law	Science
Specific to particular cases	General validity
Backward looking	Forward looking
Requires evidence	Acceptability through peer review
Short timescales	Long timescales

Other key areas of debate have concerned the way that science is presented in court: the measures of certainty employed and the statistical analysis carried out and the methods of its presentation to that court (Lynch 1998; Morgan and Bull 2006). There is significant interest currently in the Bayesian approach to interpret and present forensic evidence in general, and it is highly likely that as the methods and practices are established for other forms of evidence, it will become increasingly possible to apply those methods to forensic geoscience.

Finally, there is the issue concerning the need for experimental studies that are able to provide the foundation of forensic geoscience theory and thus inform its practice. Such a foundation is required so that the dynamics of trace geoforensic evidence can be more fully understood which in turn has great significance for enabling more accurate interpretations of such evidence. There is clearly the need for general-level empirical experimental work to be carried out in order to establish and affirm the general theories that underpin forensic geoscience. However, it is also important to acknowledge that there is the need for secondary-level experimental studies that pertain to particular cases in order for the evidence in that case to be collected, analyzed, interpreted, and presented in the most accurate and philosophically appropriate manner (Morgan et al. 2009).

International Perspectives

To date, this discussion has centered largely upon the discipline as it operates within the UK.

The primary tenets of forensic geoscience philosophy remain true outside of the UK, and much the same can be said for the collection, analysis, and interpretation of the results. There is some variation between the presentations of evidence in some countries; however, a primary distinction which can be mentioned herein involves the principle of the admissibility of geoforensic evidence. Frye and Daubert rulings within the USA influence some (but not all) US state courtrooms, a trend which is mooted more frequently in the UK and European arena now.

The UK has the advantage for the geoforensic investigator of having so many different geological/soil types on such a small series of islands, whereas in the United States, although there may be up to 3,000 different soil types (in comparison to approximately 300 in the UK), its area is so vast that there are large sedimentological terrains and facies of similar sediments and soils that spatial discrimination of the physical characteristics of those materials becomes problematic unless there are biological or anthropogenic additions which may provide higher resolution provenance indicators. Elsewhere specific areas of large global sand seas (sand deserts), for example, will have sediments comprised almost entirely of quartz with very little biological (floral/faunal) additions, with those sand grains exhibiting near identical transportational provenance origins. It is perhaps fortuitous that anthropogenic additions to even the most extensive sediment terrain or facies are so prevalent that discrimination is possible in very many if not all locations. There are after all many thousands of variations of biological, chemical, physical, and anthropogenic components of soils that discrimination will often be viable. Even at the largest spatial scale of investigation, microscale analysis can still enable discrimination.

Future Directions

It is widely acknowledged that the future of forensic geoscience lies in the growth of the empirical research base of the discipline. It will also be crucial that the discipline develops within the specifically forensic geoscience philosophical framework outlined above. Research in the forensic geosciences should incorporate as much as possible good scientific methods and experimental design but also acknowledge and develop research questions in a manner that incorporates the practice of forensic geoscience.

A number of areas are currently being developed that may give an indication of the future in forensic geoscience: firstly, the development of automation processes for evidence analysis. New capacity in this area would enable far more evidence to be analyzed and screened on an exclusionary basis, and this would reduce the amount of operator time required and ultimately lower the cost of such analysis making it more widely available. Recent examples of "proof of concept" work in this area would include the work of Newell et al. (2012) who demonstrate the capacity of computer recognition programs to be able to distinguish between quartz grain types that have been classified visually based on their surface textures. Secondly, the area of independent verification of particular forms of geoforensic analysis is being increasingly pursued in an effort to ensure that the analysis is admissible in court. While the interpretation of geoforensic evidence operates within the realm of probability, and thereby can only offer exclusionary intelligence and evidence, its weight can be enhanced by presenting the corroborative findings from two or more truly independent techniques. To take the example of quartz grain surface texture analysis, there has been preliminary work undertaken using PIXE and PIGE analysis using Ion Beam Analysis that identifies the potential for elemental screening of both the matrix and inclusions of the grain to offer independent verification of the quartz grain types identified morphologically using SEM (Bailey et al. 2012).

Related Entries

▶ Expert Evidence and Criminal Trial Procedure
▶ Forensic Anthropology

- ► Forensic Palynology
- ► Forensic Science
- ► Forensic Science and Criminal Investigation

Recommended Reading and References

Bailey, MJ, Morgan, RM, Comini, P, Calusi, S, Bull, PA (2012) An evaluation of particle induced X-Ray emission and particle induced gamma ray emission of quartz grains for forensic trace sediment analysis. Anal Chem (in press)

Brown AG, Smith A, Elmhurst O (2002) The combined use of pollen and soil analyses in a search and subsequent murder investigation. J Forensic Sci 47(3):614–618

Bull PA, Morgan RM (2006) Sediment fingerprints: a forensic technique using quartz sand grains. Sci Justice 46(2):107–124

Bull PA, Morgan RM, Sagovsky A, Hughes GJA (2006a) The transfer and persistence of trace particulates: experimental studies using clothing fabrics. Sci Justice 46(3):185–195

Bull PA, Morgan MR, Freudiger-Bonzon J (2008) A critique of the present use of some geochemical techniques in geoforensic analysis. Forensic Sci Int 178:e35–e40

Bull PA, Parker AG, Morgan RM (2006b) The forensic analysis of soils and sediment taken from the cast of a footprint. Forensic Sci Int 162:6–12

Cameron NG (2004) The use of diatom analysis in forensic geoscience. Geol Soc Spec Publ 232:277–280

Dawson LA, Towers W, Mayes RW, Craig J, Väisänen RK, Waterhouse EC (2004) The use of plant hydrocarbon signatures in characterizing soil organic matters. Geol Soc Spec Publ 232:269–276

Fenton N (2011) Improve statistics in court. Nature 479:36–37

Guedes A, Ribeiro H, Valentim B, Rodrigues A, Sant'Ovaia H, Abreu I, Noronha F (2011) Characterization of soils from the: algarve region (Portugal): a multidisciplinary approach for forensic applications. Sci Justice 51(2):77–82

Hawksworth DL, Wiltshire PEJ (2011) Forensic mycology: the use of fungi in criminal investigations. Forensic Sci Int 206(1–3):1–11

Horswell J, Cordiner SJ, Maas EW, Martin TM, Sutherland BW, Speir TW (2002) Forensic comparison of soils by bacterial community NDA profiling. J Forensic Sci 47(2):350–353

Hunter JR, Brickley MB, Bourgeois J, Bouts W, Bourguignon L, Hubrecht F, DeWinne J, Van Haaster H, Hakbijl T, De Jong H, Smits L, Van Wijngaarden LH, Luschen M (2001) Forensic archaeology, forensic anthropology and human rights in Europe. Sci Justice 41(3):173–178

Jasanoff S (2006) Just evidence: the limits of science in the legal process. J Law Med Ethics 34(2):328–341

Kirk PL (1974) Crime investigation, 2nd edn. Wiley, New York

Lane B (1992) The encyclopaedia of forensic science. Headline, London

Lombardi G (1999) The contribution of forensic geology and other trace evidence analysis to the investigation of the killing of Italian Prime Minister Aldo Moro. J Forensic Sci 44(3):634–642

Lynch M (1998) The discursive production of uncertainty: the OJ Simpson 'Dream Team' and the sociology of knowledge machine. Soc Stud Sci 28(5/6):829–868

McKinley J, Ruffell A (2007) Contemporaneous spatial sampling at scenes of crime: advantages and disadvantages. Forensic Sci Int 172(2–3):196–202

Meyers MS, Foran DR (2008) Spatial and temporal influences on bacterial profiling of forensic soil samples. J Forensic Sci 53(3):652–660

Morgan RM, Bull PA (2006) Data interpretation in forensic sediment geochemistry. Environ Forensic 7(4):325–334

Morgan RM, Bull PA (2007a) The use of particle size analysis of sediments and soils in forensic enquiry. Sci Justice 47(3):125–135

Morgan RM, Bull PA (2007b) The philosophy, nature and practice of forensic sediment analysis. Prog Phys Geography 31(1):43–58

Morgan RM, Bull PA (2007c) Forensic geoscience and crime detection. Identification, interpretation and presentation in forensic geoscience. Minerva Medicoleg 127(2):73–90

Morgan RM, Allen E, Lightowler ZL, Freudiger-Bonzon J, Bull PA (2009a) A forensic geoscience framework and practice. Polic: J Policy Pract 2:185–195

Morgan RM, Cohen J, McGookin I, Murly-Gotto J, O'Connor R, Muress S, Freudiger-Bonzon J, Bull PA (2009b) The relevance of the evolution of experimental studies for the interpretation and evaluation of some trace physical evidence. Sci Justice 49:277–285

Morgan RM, Freudiger-Bonzon J, Nichols KH, Jellis T, Dunkerley S, Zelazowski P, Bull PA (2009c) The geoforensic analysis of soils from footwear. In: Ritz K, Dawson L, Miller D (eds) Criminal and environmental soil forensics. Springer, Dordrecht, pp 253–269

Morgan RM, Wiltshire P, Parker AG, Bull PA (2006) The role of forensic geoscience in wildlife crime detection. Forensic Sci Int 162:152–162

Murray RC, Solebello LP (2002) In: Saferstein R (ed) Forensic science handbook, vol I. Prentice-Hall, Upper Saddle River/New Jersey

Murray RC, Tedrow JCF (1975) Forensic geology. Prentice-Hall, Englewood Cliffs

Newell AJ, Morgan RM, Bull PA., Griffin LD, Graham G (2012). Automated texture recognition of quartz sand grains for forensic analysis for forensic applications. J Forensic Sci (in press)

Palenik S (2007) Heavy minerals in forensic science developments in sedimentology, Chapter 37. vol 58, pp 937–961

Petraco N, Kubic T (2000) A density gradient technique for use in forensic soil analysis. J Forensic Sci 45(4):872–873

Pirrie D, Butcher AR, Power MR, Gottlieb P, Miller GL (2004) Rapid quantitative mineral and phase analysis using automated scanning electron microscopy (QemSCAN); potential applications in forensic geoscience. Geol Soc Spec Publ 232:123–136

Pringle JK, Ruffell A, Jervis JR, Donnelly L, McKinley J, Hansen J, Morgan R, Pirrie D, Harrison M, Jarvis K The use of geoscience methods for terrestrial forensic searches. Earth Sci Rev (in press)

Rawlins BG, Kemp SJ, Hodgkinson EH, Riding JB, Vane CH, Poulton C, Freeborough K (2006) Potential and pitfalls in establishing the provenance of earth-related samples in forensic investigations. J Forensic Sci 51(4):832–845

Riding JB, Rawlins BG, Coley KH (2007) Changes in soil pollen assemblages on footwear worn at different sites. Palynology 31:135–151

Ruffell A, McKinley J (2008) Geoforensics. Wiley Blackwell, Chichester, p 332

Saks MJ (2010) Forensic identification: from a faith-based “Science” to a scientific science. Forensic Sci Int 201(1–3):14–17

Siegel JA, Precord C (1985) The analysis of soil samples by reverse phase-high performance liquid chromatography using wavelength ratioing. J Forensic Sci 30(2):511–525

The Law Commission (Law Comm No 325) (2011) Expert evidence in criminal proceedings in England and Wales

Thomas M (2001) Landscape sensitivity in time and space – an introduction. Catena 42(2–4):83–98

Thornton JI, McLaren AD (1975) Enzymatic characterisation of soil evidence. J Forensic Sci 20:674–692

Walls HJ (1968) Forensic science. Sweet and Maxwell, London

Forensic Evidence

► Scientific Evidence in Criminal Prosecutions

Forensic Facial Analysis

Martin Paul Evison
Northumbria University Centre for Forensic Science, Northumbria University, Newcastle upon Tyne, UK

Overview

Closed-circuit television (CCTV) systems, digital cameras, webcams, and mobile devices are the source of a burgeoning number of facial images used in criminal investigations and prosecutions. Given the significance of facial identification to the courts – as well as to cases involving questioned identity documents and border control and immigration disputes – it is important that the strengths and weaknesses of methods used are properly understood.

Identification of an alleged offender is fundamental to the judicial process. Courts rely heavily on eyewitness evidence of identification, and they continue to do so where facial images are concerned. Evidence of identification, however, is widely acknowledged to be problematic. Procedures and processes intended to make identification more reliable – whether for use in investigation or in court – are perennial challenges.

Where facial image evidence is in issue, eyewitness evidence of investigators or other witnesses familiar with the suspect may be given. Alternatively, expert opinion of facial image analysis may be adduced. As well as being susceptible to the general weaknesses of eyewitness identification, the methods of facial image analysis are susceptible to biases known to affect other opinion evidence. The methods of facial image analysis are scientifically rudimentary and are as distinct from computational biometrics as they are from the other ideals of human identification: fingerprints and DNA.

Identification of Offenders

It is almost self-evident that for a criminal prosecution to take place, an accusation must be made against a particular individual. The prosecution must prove beyond reasonable doubt that the offense was committed and the accused committed it. Identification of the offender becomes a fact in issue when the accused denies they were the individual concerned. The onus is on the prosecution to prove it was.

Identification of Repeat Offenders

It is easy to suppose that before the advent of fingerprinting and forensic DNA profiling,

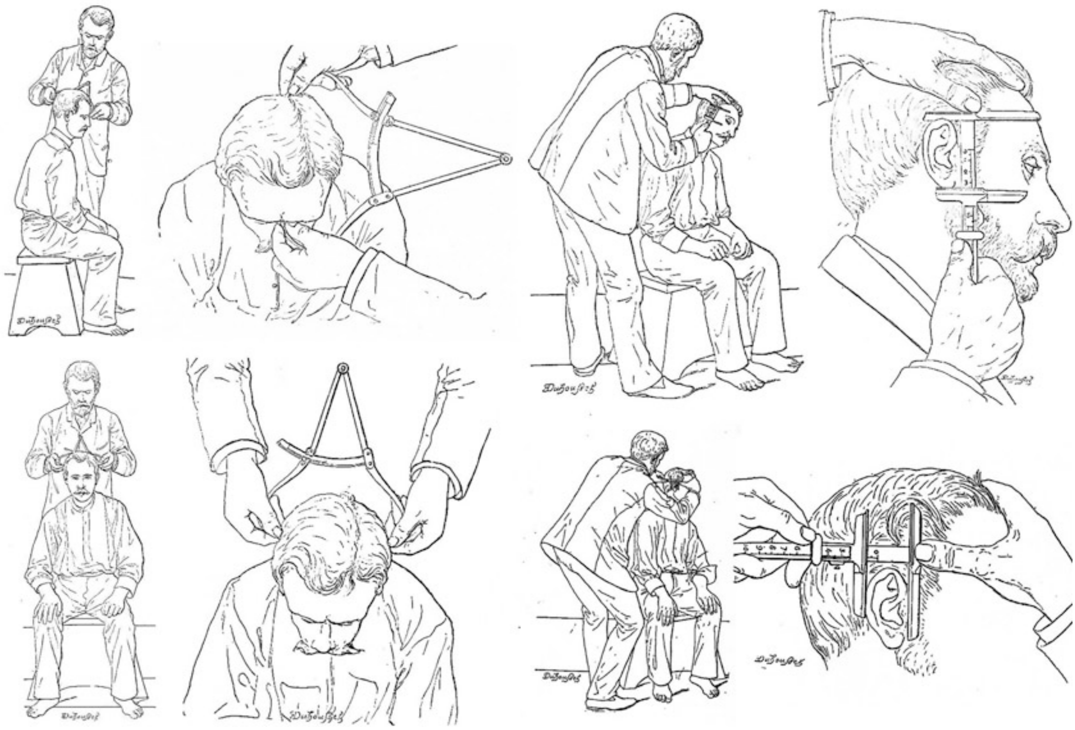

Forensic Facial Analysis, Fig. 1 Standard measurements of the head and ear adopted by Bertillon (1885)

the only evidence available to establish identity from physical appearance was eyewitness evidence, but this is not the case. In the late nineteenth century, Alphonse Bertillon (1853–1914) pioneered measurement of body shape – anthropometry – for the purposes of forensic human identification (Bertillon 1885). Bertillon was a pivotal figure in the development of forensic science, and facial comparison was a substantial component of his technique. The method – coined *Bertillonage* – involved measurement of a set of bodily dimensions, including those of the head and ear (Fig. 1). The method also incorporated classification of facial features, including the color of the iris (Fig. 2); the shape of the forehead, chin, and nose (Fig. 3); shape of the ear (Fig. 4); and style of hair and beard (Fig. 5). The nature and position of distinctive marks on the face were meticulously documented (Fig. 6) as part of a recording system which also established early standards for forensic photography.

Despite its widespread adoption, *Bertillonage* was abandoned in the early twentieth century in favor of dermatoglyphic fingerprinting. This method of human identification was faster, less error prone, required less equipment and training, and could be applied to marks left at a crime scene (Cole 2001).

Eyewitness Identification

If anthropometric measurement and systematic classification were discarded in favor of fingerprinting, then how are people identified from facial images and appearance in the criminal justice system?

Not surprisingly, individuals depicted in images are very often identified via eyewitness evidence of some kind. In the London Riots of 2011, for instance, many of suspects were recognized from CCTV footage. In these investigations, most suspects were identified by members of the public. Some police officers, however, are regarded by the Metropolitan Police

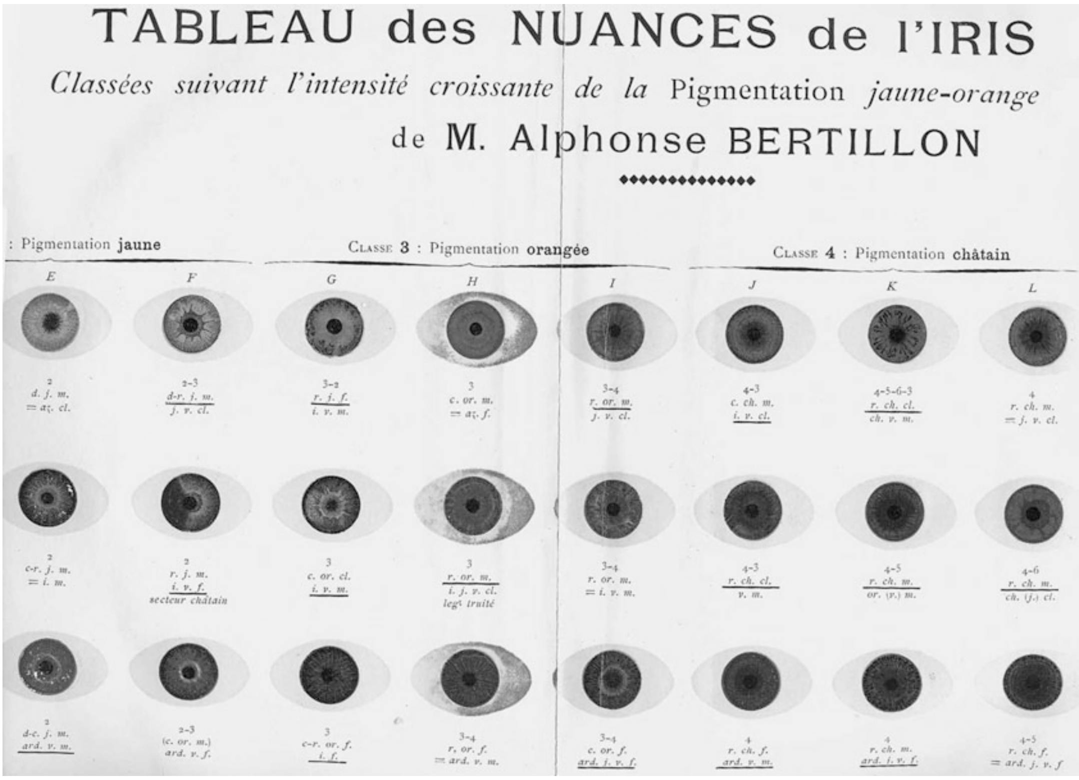
TABLEAU des NUANCES de l'IRIS
Classées suivant l'intensité croissante de la Pigmentation jaune-orange
de M. Alphonse BERTILLON
: Pigmentation jaune
Classe 3 : Pigmentation orangée
Classe 4 : Pigmentation châtain
E
F
G
H
I
J
K
L
2 d. j. m. = az. cl.
2-3 d-r. j. m. j. v. cl.
3-2 r. j. f. i. v. m.
3 c. or. m. = az. f.
3-4 r. or. m. j. v. cl.
4-3 c. ch. m. i. v. cl.
4-5-6-3 r. ch. cl. ch. v. m.
4 r. ch. m. = j. v. cl.
2 c-r. j. m. = i. m.
2 r. j. m. i. v. f. secteur châtain
3 c. or. cl. i. v. m.
3 r. or. m. i. j. v. cl. legt truité
3-4 r. or. m. = i. v. m.
4-3 r. ch. cl. v. m.
4-5 r. ch. m. or. (v.) m.
4-6 r. ch. m. ch. (j.) cl.
2 d-c. j. m. ard. v. m.
2-3 (c. or. m.) ard. v. f.
3 c-r. or. f. i. f.
3-4 r. or. f. = ard. v. m.
3-4 c. or. f. ard. j. v. f.
4 r. ch. f. ard. v. m.
4 r. ch. m. ard. j. v. f.
4-5 r. ch. f. = ard. j. v. f

Forensic Facial Analysis, Fig. 2 Standardized references of iris pigmentation adopted by Bertillon (1885)

as "super recognizers" on the basis of their ability to identify persons already known to them in CCTV footage using facial appearance and other cues. These apparent talents have been widely reported in the press (Grimston 2011; Jon 2009).

Forensic Facial Image Analysis

If eyewitness evidence is unavailable or inadequate, the methods of forensic facial image analysis may be applied. Three methods of facial image comparison are typically used in forensic investigation. They are photogrammetry or "facial mapping" (Fig. 7), image superimposition (Fig. 8), and anthroscopic or morphological analysis (Fig. 9).

Each of these methods requires an image of the offender – the questioned image – and a suspect image with the faces captured from the same or a very similar angle. The offender image may be obtained directly from digital closed-circuit television (CCTV), video or still cameras, cell phones, and similar devices. Offender images also continue to be extracted from legacy analogue videotape systems or scanned from photographs.

The offender image should ideally be the original. Alternatively, an unaltered copy can be used. There are a number of processes that have the potential to affect the fidelity of the copy, most notably digital image compression. Image compression may alter the image and may be "lossy," leading to a forfeiture of information that cannot be restored. Lowering the resolution of images leads to pixilation of points and features rendering them difficult to identify, classify, and measure. Digital capture or scanning of analogue images also has the potential to alter them. For these reasons, uncompressed copies at the highest possible resolution are desirable.

Since the pose angle of the offender in the questioned image cannot be altered, some effort

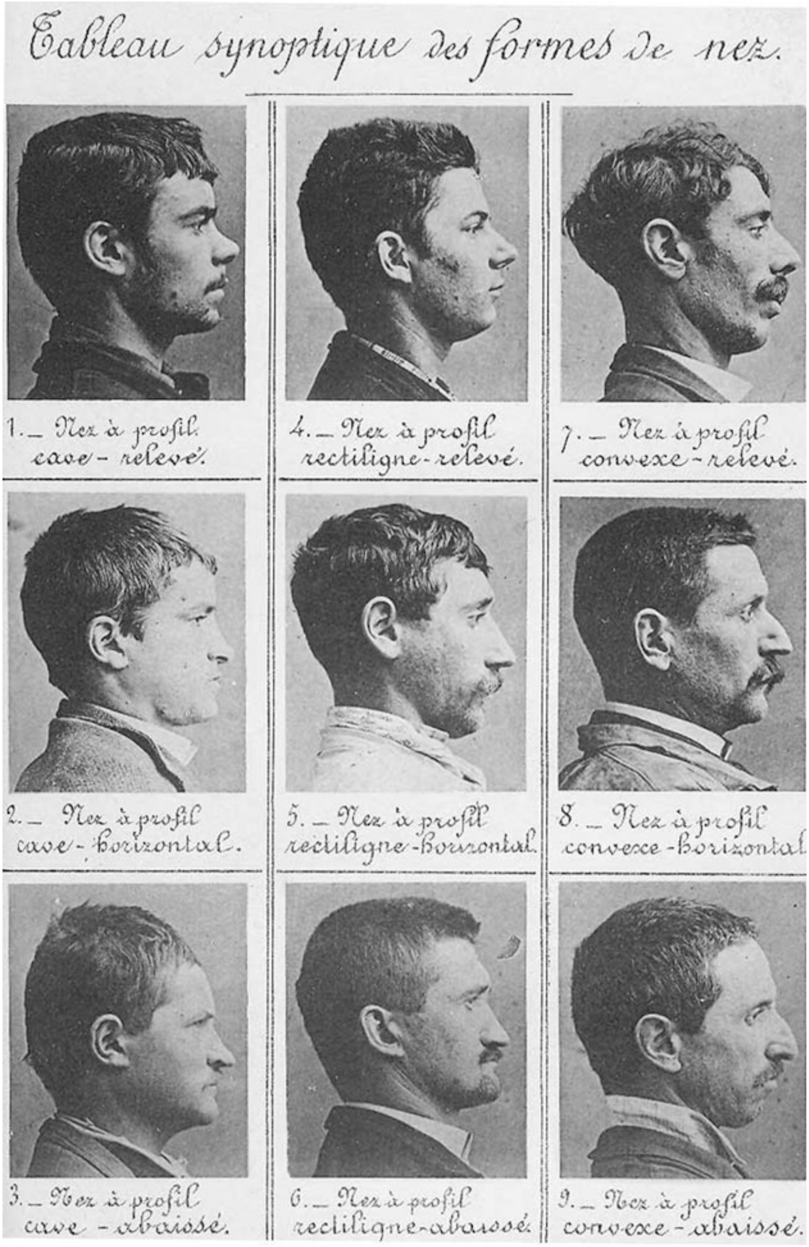
Tableau synoptique des formes de nez.
1. — Nez à profil cave - relevé.
4. — Nez à profil rectiligne - relevé.
7. — Nez à profil convexe - relevé.
2. — Nez à profil cave - horizontal.
5. — Nez à profil rectiligne - horizontal.
8. — Nez à profil convexe - horizontal
3. — Nez à profil cave - abaissé.
6. — Nez à profil rectiligne - abaissé.
9. — Nez à profil convexe - abaissé.

Forensic Facial Analysis, Fig. 3 Standardized references for the shape of the nose (Bertillon 1885)

can be made to capture an image of the suspect at the same pose angle – sometimes using the original imaging system in situ at the scene. Precision is very difficult to achieve, however, and pose angles of offender and suspect rarely match exactly.

The systems used to capture offender and suspect images may also affect the precision that can be achieved in a comparative analysis. The processes of digital and analogue recording are different, and in either case lens distortion, focus, and lighting may affect offender and suspect images differently.

Working from a presumption of innocence, if inexplicable differences between offender and suspect images are identified, then the

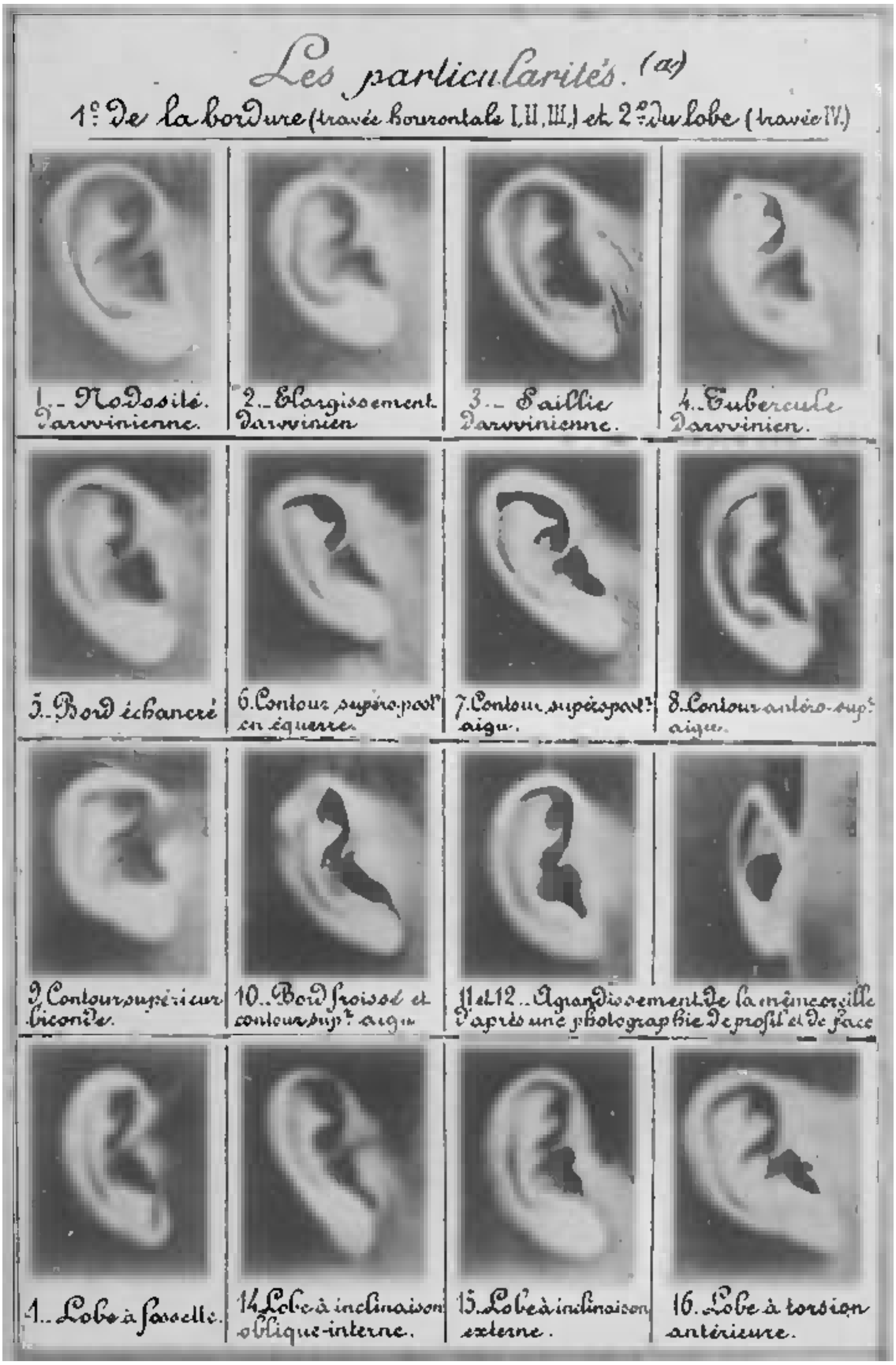
Les particularités. (a)
1º De la bordure (travée horizontale I, II, III,) et 2º du lobe (travée IV.)
1.- Nodosité. Darwinienne.
2.- Elargissement Darwinien
3.- Saillie Darwinienne.
4. Tubercule Darwinien.
5.- Bord échancré
6. Contour supéro-post.r en équerre.
7. Contour supéro-post.r aigu.
8. Contour antéro-sup.r aigu.
9. Contour supérieur biconde.
10. Bord froissé et contour sup.r aigu.
11 et 12.. Agrandissement de la même oreille d'après une photographie de profil et de face
1.. Lobe à fossette.
14. Lobe à inclinaison oblique-interne.
15. Lobe à inclinaison externe.
16. Lobe à torsion antérieure.

Forensic Facial Analysis, Fig. 4 Standardized references for the shape of the ear (Bertillon 1885)

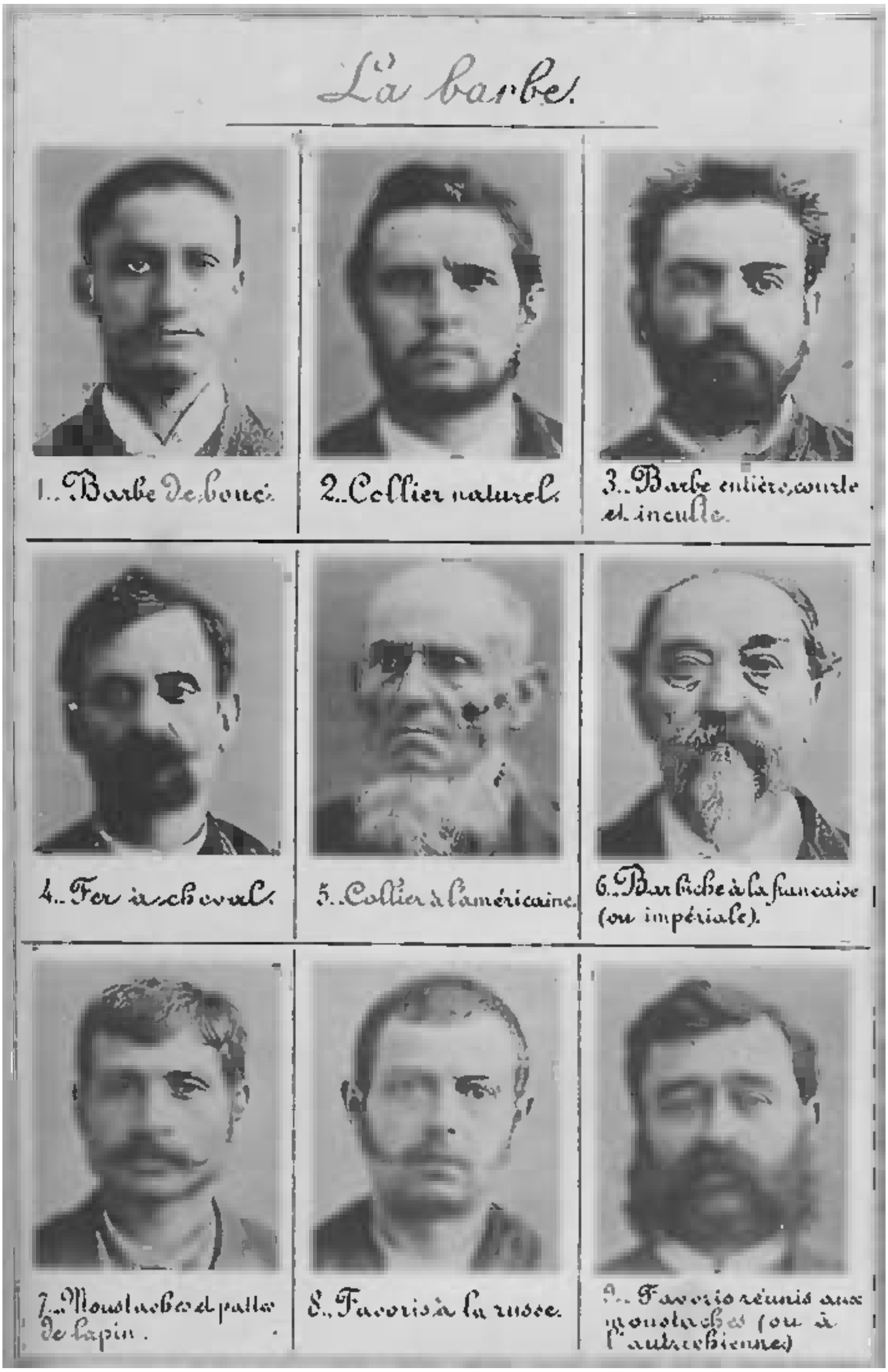
La barbe.
1.. Barbe de bouc.
2.. Collier naturel.
3.. Barbe entière, courte et inculte.
4.. Fer à cheval.
5.. Collier à l'américaine.
6.. Barbiche à la française (ou impériale).
7.. Moustaches et patte de lapin.
8.. Favoris à la russe.
9.. Favoris réunis aux moustaches (ou à l'autrichienne)

Forensic Facial Analysis, Fig. 5 Standardized references for the shape of the beard (Bertillon 1885)

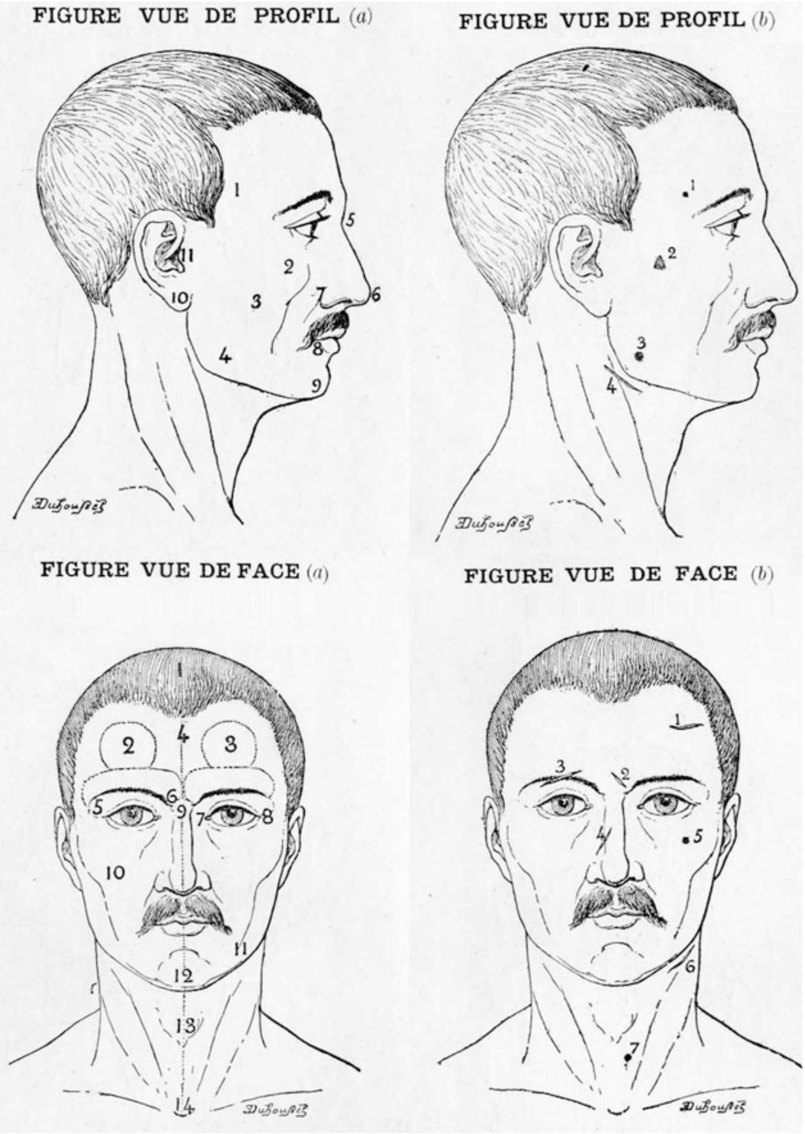
FIGURE VUE DE PROFIL (a)
FIGURE VUE DE PROFIL (b)
1
2
3
4
5
6
7
8
9
10
11
FIGURE VUE DE FACE (a)
FIGURE VUE DE FACE (b)
12
13
14

Forensic Facial Analysis, Fig. 6 Standard classification of distinguishing facial marks (Bertillon 1885)

facial images are assumed to originate from two different individuals, and an “exclusion” can be made. Explicable differences may be attributed to factors such as pose angle, lighting, lens distortion, focus, image resolution, and compression discussed above or due to other factors such as facial expression, aging, trauma or surgery, hair style, and clothing and jewelry. These are sometimes referred to as the “confounding factors” of forensic facial comparison.

If no explicable differences can be found, further comparisons using one or more of the methods of forensic facial image analysis may be undertaken.

Facial mapping and image superimposition require that the images be rotated into alignment and adjusted to the same scale. Good practice dictates that it is the suspect image that is adjusted in order that the offender image is unaltered. Changing the proportions of the image invalidates the process of comparison, however, as one of its aims is to assess

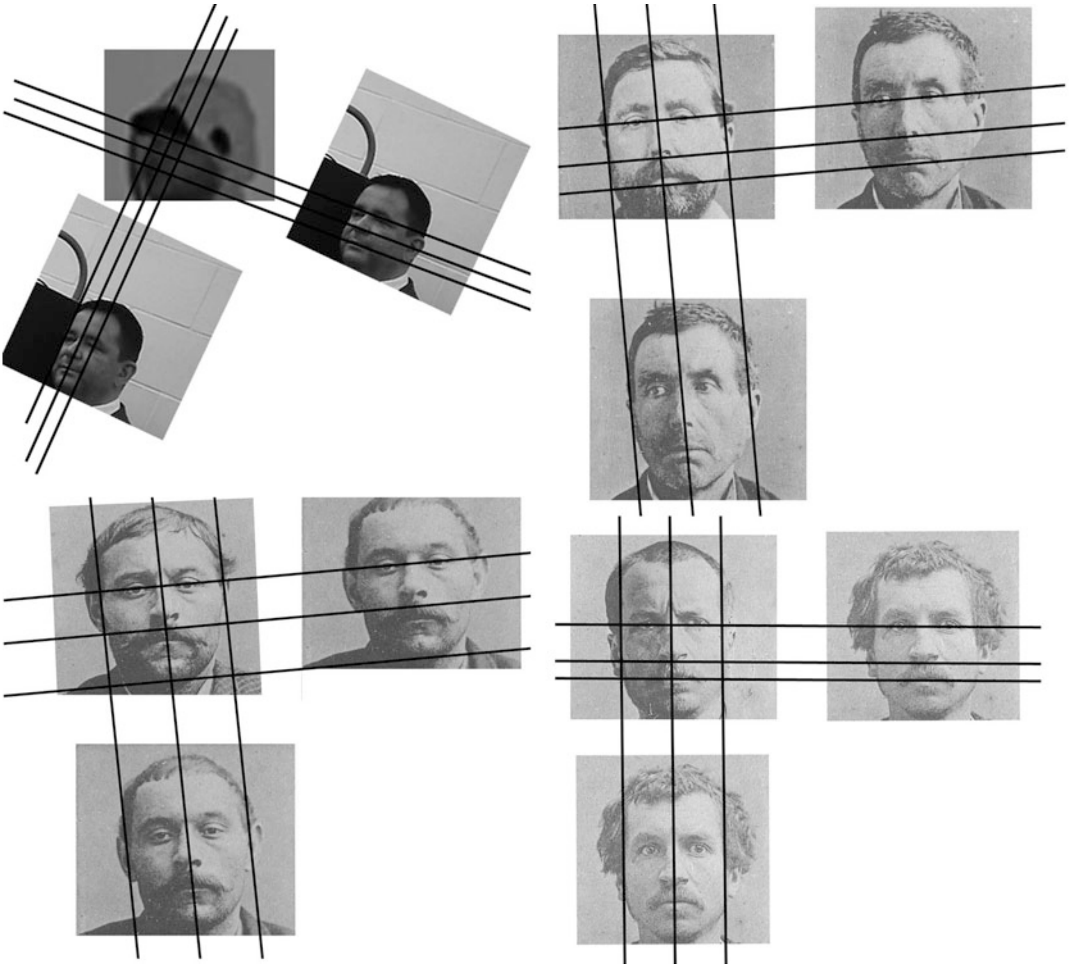

Forensic Facial Analysis, Fig. 7 Illustrative example of facial comparison by photogrammetry. Typically, a series of parallel lines are drawn between common features in offender and suspect images, such as the pupils, the commissure of the mouth, and the facial midline. If the lines are congruent and no explicable differences between the facial images can be identified, the expert may conclude the images come from the same individual or – more circumspectly – another individual of similar appearance

whether or not the two facial images share the same proportions. By implication, the scaling process would obscure any real difference in size between the heads of offender and suspect, should it occur.

Although "facial mapping" is known technically as photogrammetry, the implication that it involves detailed or complex mathematical analysis is misleading. The method simply consists of drawing parallel lines through points on features in the offender and suspect images, in order to compare their positions and the proportions of the face (Fig. 7).

Image superimposition involves the overlaying of offender and suspect images in order to perform a subjective assessment of similarity. The most common approach involves altering the visibility of the overlying image, gradually revealing the underlying one (Fig. 8). An alternative used in analogue systems is a "wipe," in which the overlying image is sequentially cropped from one side to gradually reveal

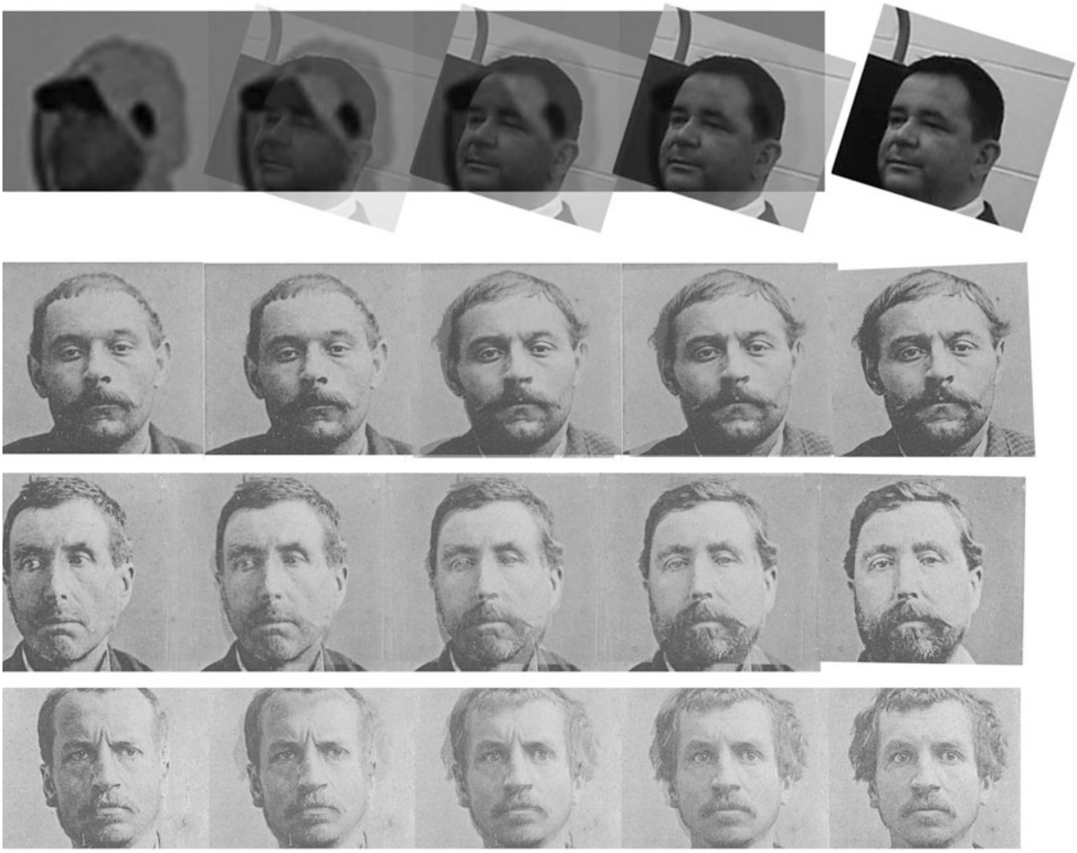

Forensic Facial Analysis, Fig. 8 Illustrative example of facial comparison by video superimposition. Images have been superimposed using Adobe® Photoshop® with the opacity of the superimposed image gradually reduced to reveal the underlying one. Again, if no explicable differences between the facial images can be identified, the expert may conclude the images come from the same individual. Video superimposition involves no explicit measurement, and its subjective nature is compounded by potential neurological (persistence of vision) and psychological (bias) influences on the examiner

the underlying one. A further alternative is a "blink," achieved by rapidly switching the overlying and underlying images. In each case, it is anticipated that the eye will detect subtle differences between the two images.

Morphological analysis or anthroscopy (Fig. 9) involves making a systematic assessment of the shape of the head, face, and facial features. Photographs or digital images can be used. A grid may be used to guide comparison.

If – after facial image analysis – no explicable differences can be found, the images may be said to "match." Normally, a further assessment is required to judge the extent to which the "match" supports the proposition that the faces in the offender and suspect images are from one and the same person.

The UK Forensic Imagery Analysis Group (FIAG) advocates a sliding scale of support (Bromby and Plews 2006), with criteria for assignment to each level on the scale (Table 1). The court is usually provided with a statement describing the analysis and the expert's view as to the strength of support it affords for the assertion that the facial images depicted are from the same individual. Recently, experts have qualified their opinion by stating matching evidence may arise from the defendant or *someone of similar appearance*.

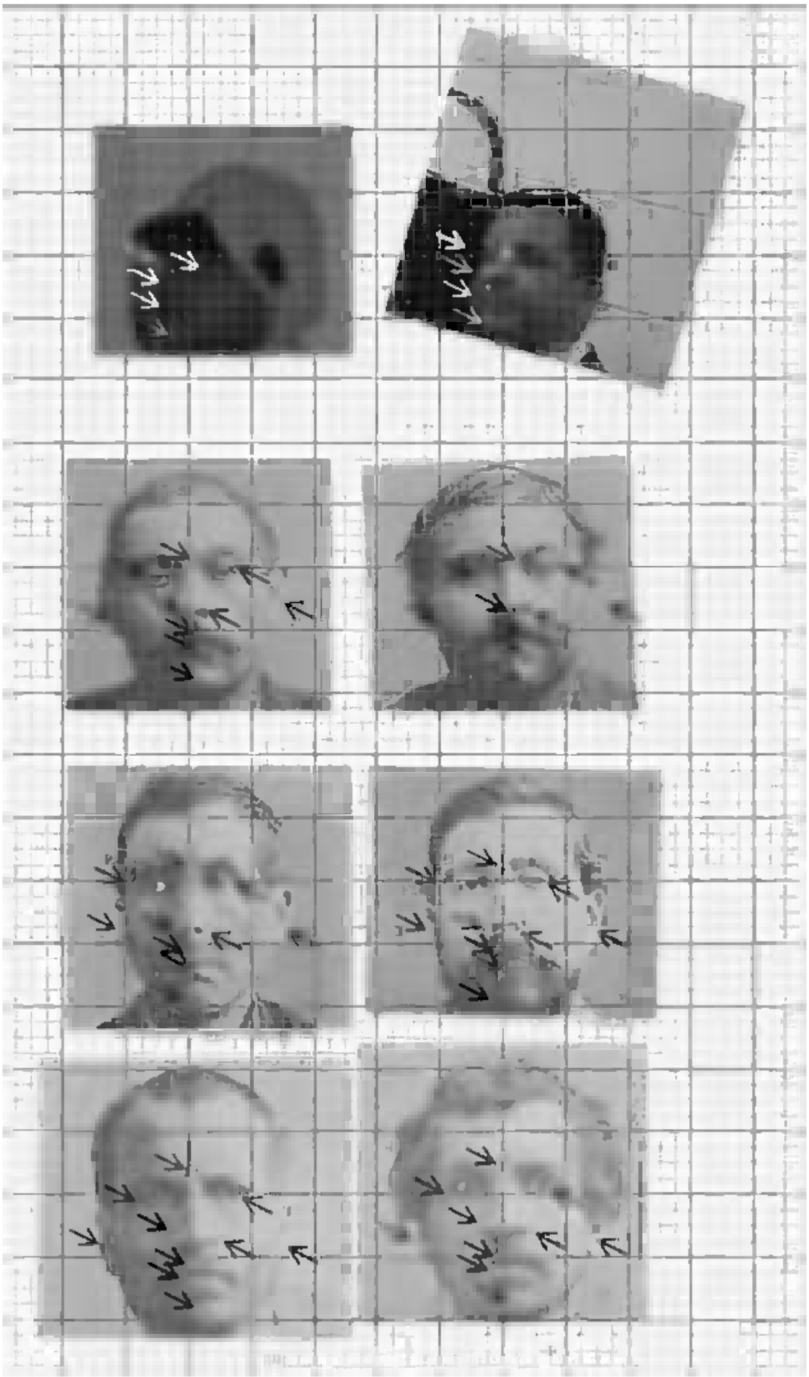

Forensic Facial Analysis, Fig. 9 Illustrative example of facial comparison by anthroscopy. Anthroscopy or morphological analysis involves the systematic comparison of the shape of the facial features and other distinguishing marks. Although subjective, anthroscopy can rely on the use of classical anatomical classification and descriptive terminology that is open to confirmation or dispute. Descriptive classifications may include, for example, the shape and position of the inner and outer canthi (commissures of the eyelids), chelions (commissures of the lips), alares (nostrils), attachment and shape of the ear lobes, and shape and prominence of the chin

Bertillon (1885) appears to have perceived this issue. Distinguishing dissimilar faces is frequently trivial. Distinguishing similar faces is the real problem. Other than the author's own image, the faces given in the examples in Figs. 7, 8 and 9 are those Bertillon chose to illustrate the problem of distinguishing similar faces. In two of the three examples, the images originate from the same individual, in one exceptional example they do not.

Forensic Facial Analysis, Table 1 Sliding scale of support (Bromby and Plews 2006)

No support	There are no significant differences observable, from which to eliminate, and there may be a number of similarities. The imagery evidence is too poor to permit observation in either direction
	There may be very general characteristics observable (i.e., race, gender, hair color), but these do not offer enough support towards specific identity
Limited support	There are a few general characteristics observable. In combination, these provide a low level of facial uniqueness
	The image quality does not permit observation of individual facial feature detail (or only very limited detail)
	The facial geometry may or may not be matching, but observation of identifiable landmarks is restricted by the image quality
Moderate support	There are a few general characteristics observable
	The image quality permits observation of a moderate amount of facial feature detail, that is, for each visible feature, one to a few descriptives can be provided. For example, the nose could be classified as having a narrow bridge and a straight ridge
	In combination, a moderate level of uniqueness is available for the facial features
	The facial geometry may be matching, but observable detail is moderate (observation of identifiable landmarks is somewhat restricted)
Support	The image quality permits observation of individual facial feature detail to some extent
	There are a few facial features observable
	There may or may not be matching geometry, but this is limited to one or two instances
	In combination (or alone), the facial features have an average amount of uniqueness; they are shared by an average number of people
	The facial similarities can be observed a few times. This may be in combination
	The face is observable from only a few differing viewpoints
Strong support	The image quality permits a good level of individual facial feature detail to be observed
	A high number of facial features are observable
	In combination (or alone), the features provide a reasonably high level of uniqueness
	The facial geometry may be matching, and multiple landmarks are observable from which to align
	The similarities can be observed a reasonably good number of times
	The face is observable from multiple differing viewpoints
Powerful support	A very high number of facial features are observable
	The image quality permits a very good level of facial feature detail to be observed
	In combination (or alone), the features provide a high level of facial uniqueness. In the case of the ear, multiple points can be observed to be matching
	The facial geometry is matching on several occasions
	The facial similarities can be observed alone or in combination a high number of times
	The face is observable from multiple viewpoints allowing for a pseudo-three-dimensional impression of the face

Current Issues and Controversies

Facial image analysis remains a subjective process sometimes supported by a technical method and then only in part. Psychological studies have raised issues concerning the reliability of face recognition and of cognitive bias in comparative forensic analyses. Reviews of the effectiveness of forensic science have questioned the scientific basis of many forensic science subdisciplines as well as the subjective paradigm from which the forensic identification sciences operate.

The Psychology of Face Recognition

While face perception and face recognition remain exciting subjects of academic inquiry, certain features of our ability to perceive and recognize faces are clear. People are very good at recognizing familiar faces and very bad at

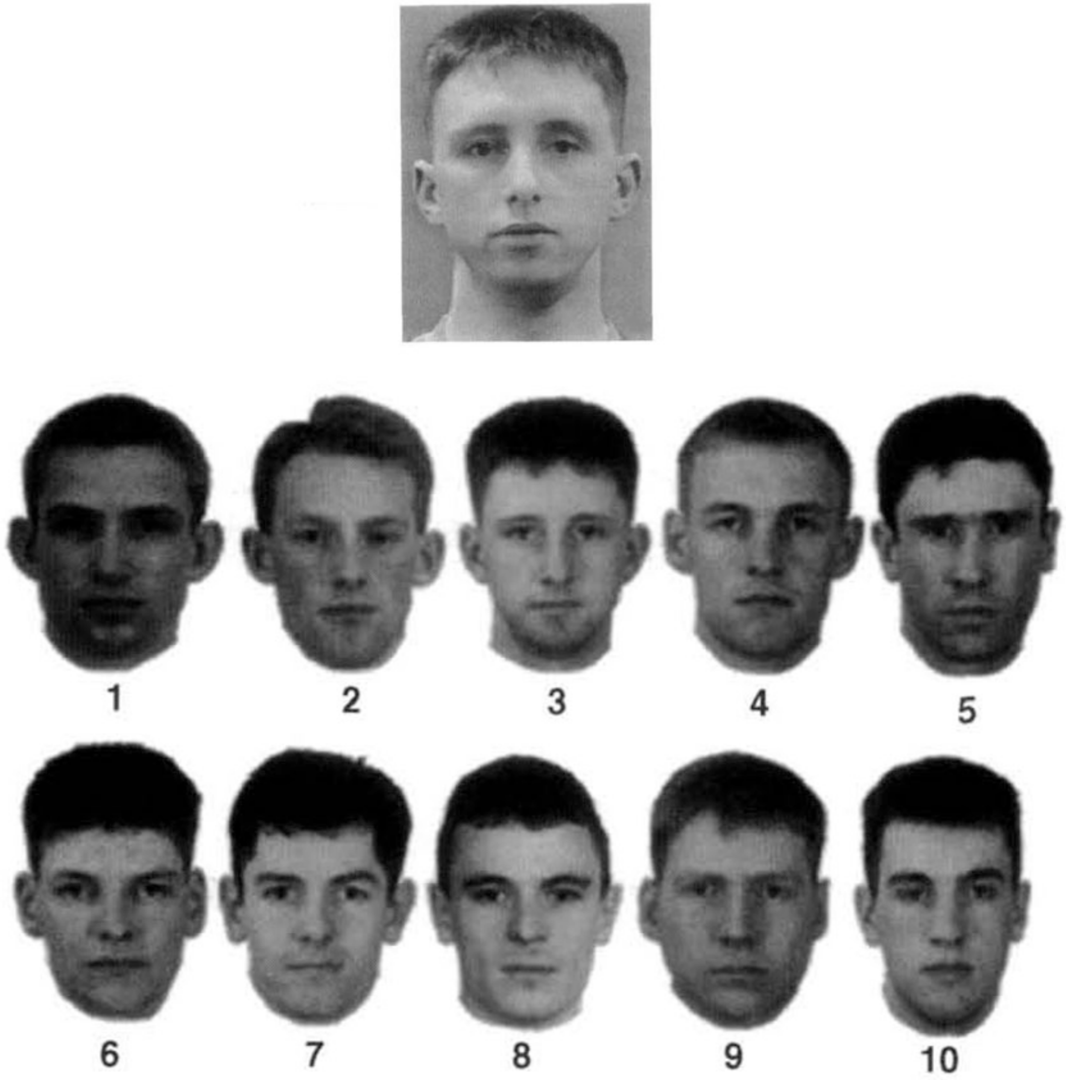

Forensic Facial Analysis, Fig. 10 Example of an experiment in facial recognition (Bruce et al. 1999 with permission) using a face pool. The subject is invited to establish whether or not the target face (*upper image*) is present in the array and, if so, which one it is

recognizing unfamiliar ones. Using standardized sets of photographic images called face pools (Fig. 10), researchers in psychology (Bruce et al. 1999) demonstrated the highly error-prone nature of recognition of unfamiliar faces. Kemp et al. (1997), for example, attached small photographs of shoppers to credit cards and asked cashiers to confirm that they belonged to the bearer. The subjects were able to detect fraudulent cards in only 36 % of trails when the photograph bore some resemblance to the bearer and in 66 % of trails when the photograph bore no resemblance.

Familiar faces are, in contrast, reliably recognized from face pools.

The same pattern is evident when reliability is assessed using CCTV footage of familiar and unfamiliar subjects. Burton et al. (1999) showed familiar individuals can be recognized quite reliably even from poor quality CCTV footage,

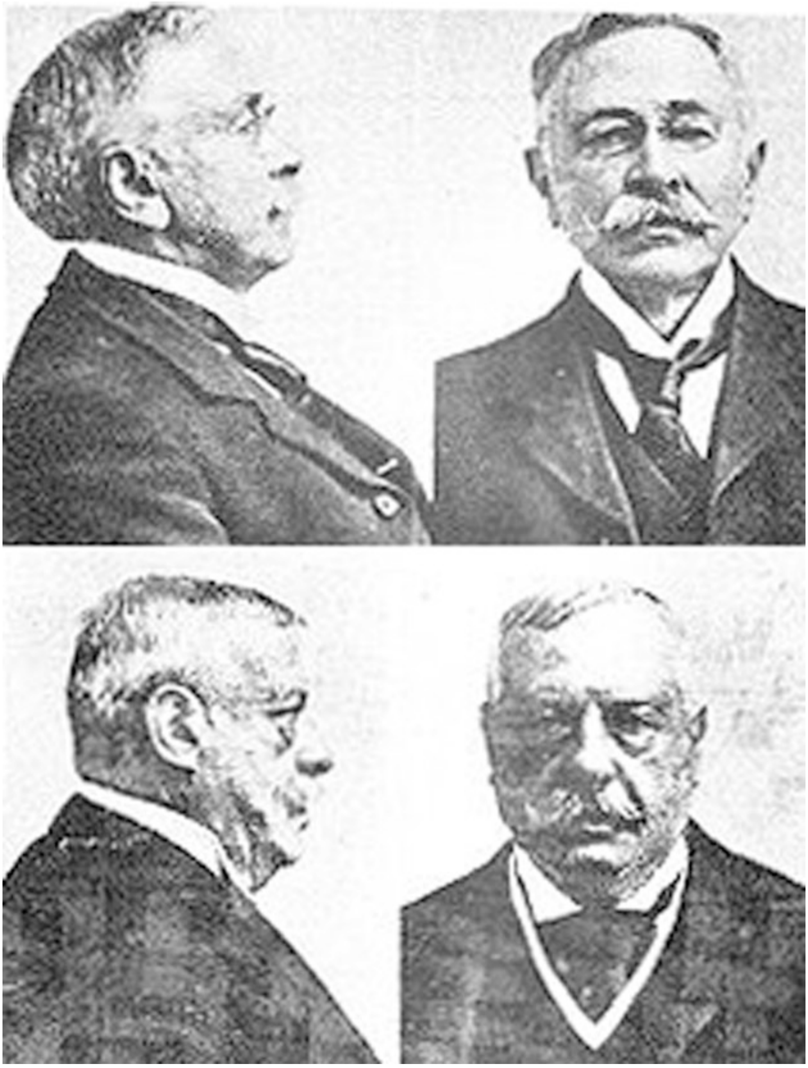

Forensic Facial Analysis, Fig. 11 Images of Adolph Beck (*upper*) and Wilhelm Meyer (*lower*). The Adolph Beck case was contributory to the development of the Court of Appeal in England and Wales

but unfamiliar subjects are frequently misidentified. The performance of students and police officers was similarly poor. Bruce et al. (1999, 2001) also demonstrated the highly error-prone nature of recognition of unfamiliar faces from CCTV images and video stream.

Error in Eyewitness Identification

The findings of Bruce and other psychologists of face perception offer scientific confirmation of the potential danger of eyewitness identification long recognized by the courts. Cases of mistaken identity – particularly of individuals unfamiliar to the witness – have become seminal in reforms of judicial procedure. In England and Wales, the *Turnbull* guidelines – arising from *R v. Turnbull* 1977 (QB224 CA) – govern the circumstances and nature of warnings to the jury of the risks inherent in eyewitness identification. Similar procedures are followed in US courts – see *Watkins v. Soders* 1981 (449 US 341; 101 SCt 654).

Both psychologists and the courts accept that eyewitness identification can be dangerous when unfamiliar faces are involved. Psychological evidence indicates that recognition of familiar faces is quite reliable, but it certainly is not perfect. A celebrated English case from the

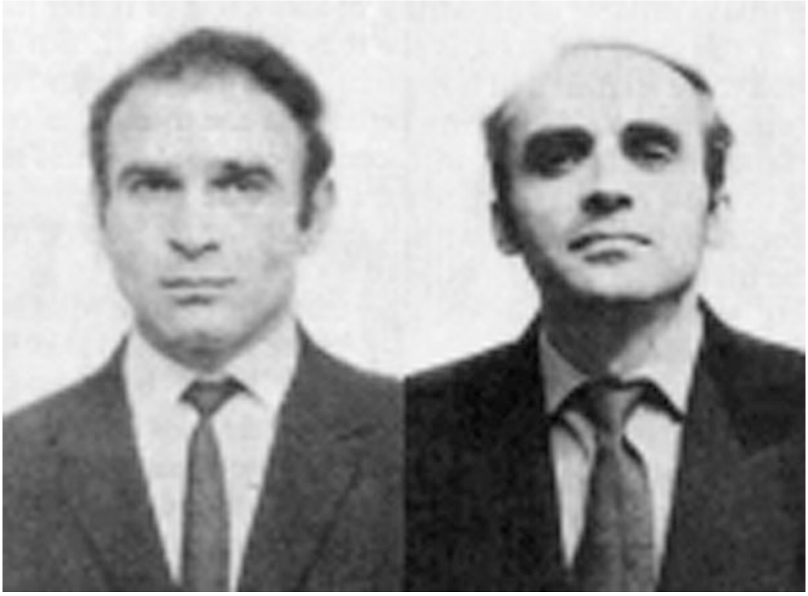

Forensic Facial Analysis, Fig. 12 Images of Laszlo Virag (*left*) and Roman Ohorodnyckyi or Georges Payen (*right*). The Virag case ultimately led to the adoption of the *Turnbull* guidelines in England and Wales

nineteenth century involves that of Adolf Beck (Anon 1999), a man who had the misfortune to resemble fraudster Wilhelm Meyer, who presented himself to his victims as Lord Willoughby. One of Meyer's victims mistook Beck for Meyer. Beck's misfortune was compounded when ten or 11 of Meyer's other victims also did so. Two police officers familiar with Meyer also identified Beck as Meyer. One, Ellis Spurrell, stated "There is no doubt whatever he is the man. I know what's at stake on my answer, and I say without doubt he's the man." Having been found guilty not once, but twice of crimes committed by Meyer, Beck was awaiting sentencing when Meyer himself was arrested. The mistaken identity – and the passing resemblance of the two protagonists (Fig. 11) – finally came to light, and Beck was pardoned.

A further case arose in England in the twentieth century, which involved the mistaken identity of Laszlo Virag as armed thief Roman Ohorodnyckyi – also known as Georges Payen (Fig. 12). In the Virag case (Devlin 1976), a - number of witnesses including police officers again mistook Virag for Payen and confidently so. One officer, Police Constable Smith, stated "his face is imprinted on my brain." Virag was convicted, but items subsequently found in Payen's possession, including the relevant firearm, led to Virag's exoneration, the Devlin Report (1976) on identification in criminal cases, and the adoption of the *Turnbulls guidelines*.

Cognitive Bias

The findings of Bruce and other psychologists of face perception offer scientific corroboration for the premise that humans are good at recognizing familiar faces but poor at recognizing unfamiliar ones. Other psychological studies support the proposition that some individuals are exceptionally good at face recognition (Russell et al. 2009). Humans are, unfortunately, susceptible to other cognitive biases that can lead witnesses, including police officers, to mistakenly – if honestly – make highly confident, but erroneous identifications from facial appearance.

The wrongful convictions of Adolph Beck and Laszlo Virag each led to changes in the English court system and – in order to prevent errors and preempt courtroom challenges – to changes in police procedures relating to identification from arrest photographs and identification parades.

The risks of context and confirmation bias in forensic comparisons have been well rehearsed by Dror (Dror et al. 2006) in relation to fingerprints, and similar dangers apply to identifications from CCTV and forensic facial images. Video superimposition is a particularly subjective technique, with the capacity to obscure differences rather than reveal them.

Scientific Basis of Forensic Facial Image Analysis

Forensic facial image analysis is scientifically rudimentary. There is a small literature (Porter and Doran 2000; Vanezis and Brierley 1996; Vanezis et al. 1996; Yoshino et al. 2000) supported by guidelines published in the USA (SWGIT 2012) and UK (ACPO 2009). Validation studies are almost unknown. There is no reported error rate for forensic facial image comparison, whether conducted by police "super recognizers" or "facial mapping experts."

There is no scientific basis for assigning the weight of a comparison against a sliding scale of support. This is a matter of subjective opinion. Two experts could disagree, and both be "right" – see Mallett and Evison (2013). This is a position that is hardly tenable in science. The limited scientific basis of forensic facial image analysis is discussed in further detail by Edmond (Edmond 2011) in relation to courtroom admissibility of facial comparison evidence. The UK Court of Appeal held in *R v. T* (ECWA Crim 2439; 2010) with regard to footwear impressions that the expert should not use the word "scientific" where it gives an impression to the jury of a degree of precision and objectivity that is not present. This caution is equally warranted with regard to facial image comparison and is all the more important giving the compelling nature of facial image evidence.

Conclusions and Future Research

Despite the ubiquity of the face as the primary mode of recognition in humans and the increasing prevalence of facial images, most forensic identifications are based on eyewitness evidence of some form. Forensic facial comparison is not scientific, and the scientific claims of facial mapping and image superimposition are tenuous. There are presently two significant drivers of potential change, however, that may lead to the development of a forensic science of facial image analysis. These are the systematic investigation of images from CCTV footage and the expectation for increasing scientific rigor in the forensic sciences.

Systematic Use of Images in Criminal Investigation

The investigation of the London Riots of 2011 is notable for the extensive implementation of a thorough and systematic approach to the investigation of images from CCTV and other sources – digital cameras and other mobile devices. This approach was based on preexisting Visual Images Identification and Detection Office (VIIDO) units introduced by Michael Neville of London's Metropolitan Police Service (Grimston 2011). Not only do these units control collection and analysis of image evidence, they ensure that it is supervised and actioned (Jon 2009). Online and mobile (APCO 2012) apps are offered to support dissemination of images captured from CCTV and other sources.

The Paradigm Shift in Forensic Science

The unparalleled success and reliability of forensic DNA profiling has led to a postulated "paradigm shift" in forensic science (Saks and Koehler 2005), which predicts that more of the forensic science subdisciplines will need to follow the model offered by DNA. Both the United States National Research Council (2009) and the UK Forensic Science Regulator (Rennison 2008) call for forensic science to be based on sound scientific methodology, logical and statistically supported interpretation methods, and known error rates. Biometric technology may offer some value to forensic facial comparison, but at present the value of automated systems is investigative and not probative (Dessimoz and Champod 2007; Evison 2011). Reliance on automated biometric systems may also lead to a new kind of

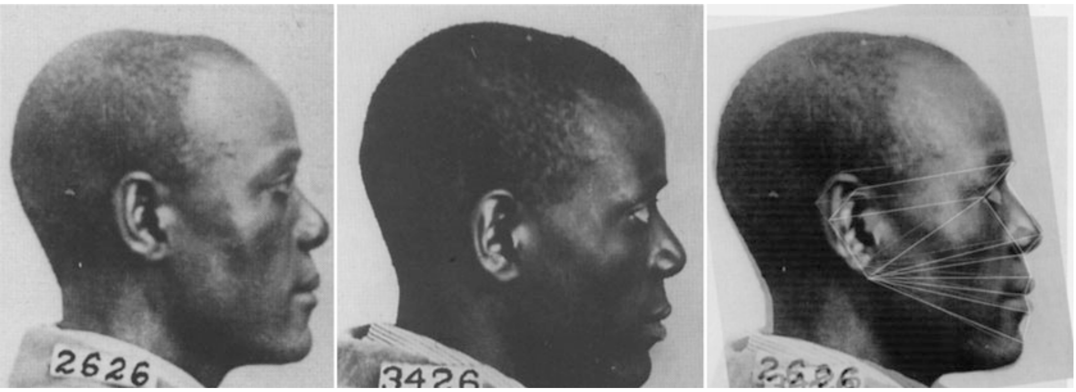

Forensic Facial Analysis, Fig. 13 Images of Will and William West, inmates of Leavenworth Penitentiary in the United States who shared the same name, had similar facial appearances, and could not be distinguished by the Bertillon system. Will and William West were distinguished using fingerprints, however. Note also that their facial proportions (see superimposition, *right*) and ear shapes are also different

complacency that promotes context bias (Dror and Mnookin 2010).

Some research, such as that of Mardia et al. (1996), Yoshino et al. (2000), and Evison et al. (Evison et al. 2010; Evison and Vorder Bruegge 2010), has explored the potential for measurement-based and probabilistic methods of facial comparison using features demonstrable to a lay person. This approach is provisionally able to distinguish the famously similar appearances of Will and William West (Fig. 13), an early case of persons distinguished and identified from their fingerprints by the FBI.

Presently, however, investigation relies on recognition from CCTV images and prosecution on the admissibility of eyewitness identification as evidence in court. Although some investigators coin CCTV image analysis the third forensic discipline (Jon 2009) after dermatoglyphic fingerprinting and DNA analysis, the allusion is somewhat misleading as there is no definitive forensic science of facial image comparison.

Related Entries

- ▶ Biometrics and Border Control policing
- ▶ CCTV and Crime Prevention
- ▶ Cognitive Forensics: Human Cognition, Contextual Information, and Bias
- ▶ Critical Report on Forensic Science
- ▶ Eyewitness Identification Evidence
- ▶ Eyewitness Research
- ▶ Forensic Science and Miscarriages of Justice
- ▶ Jury Decision Making and Eyewitness Testimony
- ▶ Philosophy of Forensic Identification
- ▶ Probability and Inference in Forensic Science

Recommended Reading and References

ACPO (2009) Facial identification guidance. Association of Chief Police Officers, London. http://www.acpo.police.uk/asp/policies/Data/Facial%20ID.pdf. Accessed 20 July 2012

Anonymous (1999) The strange story of Adolph Beck. Stationery Office, London

APCO (2012) Public encouraged to identify criminals with mobile app. Br Assoc Public Saf Commun Off J. http://www.bapcojournal.com/news/fullstory.php/aid/2239/Public_encouraged_to_identify_criminals_with_mobile_app.html. Accessed 30 July 2012

Bertillon A (1885) Identification anthropométrique: instructions signalétiques. Melun, Paris

Bromby MC, Plews S (2006) Guidance for evaluating levels of support. FIAG: Forensic Imagery Analysis Group. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1550752. Accessed 20 July 2012

Bruce V, Henderson Z, Greenwood K, Hancock PJB, Burton AM, Miller P (1999) Verification of face identities from images captured on video. J Exp Psychol Appl 5(4):339–360

Bruce V, Henderson Z, Newman C, Burton AM (2001) Matching identities of familiar and unfamiliar faces caught on CCTV images. J Exp Psychol Appl 7(3):207–218

Burton AM, Wilson S, Cowan M, Bruce V (1999) Face recognition in poor-quality video: evidence from security surveillance. Psychol Sci 10(3):243–248

Cole S (2001) Suspect identities: a history of fingerprinting and criminal identification. Harvard University Press, Cambridge, MA

Dessimoz D, Champod C (2007) Linkages between biometrics and forensic science. In: Jain AK (ed) Handbook of biometrics. Springer, Berlin

Devlin P (1976) Report to the secretary of state for the home department of the departmental committee on evidence of identification in criminal cases. H.M.S.O, London

Dror IE, Mnookin JL (2010) The use of technology in human expert domains: challenges and risks arising from the use of automated fingerprint identification systems in forensic science. Law Probab Risk 9:47–67

Dror IE, Charlton D, Peron AE (2006) Contextual information renders experts vulnerable to making erroneous identifications. Forensic Sci Int 156(1): 74–78

Edmond G (2011) The building blocks of forensic science and law: recent work on DNA profiling (and photo comparison). Soc Stud Sci 41(1):127–152

Evison MP (2011) Biometrics for forensics. In: van Tilborg HCA, Jajodia S (eds) Encyclopedia of cryptography and security. Springer, New York

Evison MP, Vorder Bruegge RW (2010) Computer-aided forensic facial comparison. CRC Press, Boca Raton

Evison MP, Dryden I, Fieller NRJ, Mallett XDG, Morecroft L, Schofield D, Vorder Bruegge RW (2010) Key parameters of face shape variation in 3D in a large sample. J Forensic Sci 55:159–162

Grimston J (2011) Eagle-eye of the yard can spot rioters by their ears. The Sunday Times, 20 Nov 2011

Jon D (2009) Closed circuit television's key role in modern policing. Doktor Jon's guide to CCTV and IP video surveillance. http://www.doktorjon.co.uk/pdf/DCI%20Neville%20interview0509.pdf

Kemp R, Towell N, Pike G (1997) When seeing should not be believing: photographs, credit cards and fraud. Appl Cogn Psychol 11(3):211–222

Mallett XGD, Evison MP (2013) Forensic facial comparison: issues of admissibility in the development of novel analytical techniques. J Forensic Sci (in press)

Mardia KV, Coombes A, Kirkbride J, Linney A, Bowie LJ (1996) On statistical problems with face identification from photographs. J Appl Stat 23:655–676

NRC (2009) Strengthening forensic science in the United States: a path forward. National Academies Press, Washington, DC

Porter G, Doran G (2000) An anatomical and photographic technique for forensic facial identification. Forensic Sci Int 114(2):97–105

Rennison A (2008) Manual of forensic science regulation. Home Office, Office of the Forensic Science Regulator, London. http://www.homeoffice.gov.uk/publications/police/operational-policing/Manual_of_Regulation_22.9.08.pdf. Accessed 20 July 2012

Russell R, Duchaine B, Nakayama K (2009) Super-recognizers: people with extraordinary face recognition ability. Psychon Bull Rev 16(2):252–257

Saks M, Koehler J (2005) The coming paradigm shift in forensic identification science. Science 309:892–895

SWGIT (2012) Best practices for forensic photographic comparison. International Association for Identification: Scientific Working Group on Imaging Technology. http://www.theiai.org/guidelines/swgit/guidelines/section_16_v1-0.pdf. Accessed 20 July 2012

Vanezis P, Brierley C (1996) Facial image comparison of crime suspects using video superimposition. Sci Justice 36(1):27–33

Vanezis P, Lu D, Cockburn J, Gonzalez A, McCombe G, Trujillo O, Vanezis M (1996) Morphological classification of facial features in adult Caucasian males based on an assessment of photographs of 50 subjects. J Forensic Sci 41(5):786–791

Yoshino M, Matsuda H, Kubota S, Imaizumi K, Miyasaka S (2000) Computer-assisted facial image identification system using a 3-D physiognomic range finder. Forensic Sci Int 109(3):225–237

Forensic Genetics

▶ DNA Technology and Police Investigations

Forensic Linguistics

Lorna Fadden[1] and Sandra Ferrari Disner[2]
[1]Department of Linguistics, Simon Fraser University, Burnaby, BC, Canada
[2]Department of Linguistics, University of Southern California, Los Angeles, CA, USA

Overview

Forensic Linguistics is a branch of Applied Linguistics involving the examination of language evidence in a criminal or civil matter and it can be carried out for two broad purposes. First, language analysis can be applied during investigations to assist in the identification of suspects or witnesses, or in determining the significance of utterances or writing to a case. Second, spoken or written language samples may be submitted as evidence in court, along with the testimony of a linguistic expert. Language evidence may bear heavily on the case itself, where the language in question constitutes a language crime: threats, coercion, bribery, hate speech, and hate literature, or language evidence might be more peripherally related to a case and it may necessitate a linguist to clarify the meaning of what is spoken or written, the manner in which it is delivered, and the role of context in the interpretation of the message.

This entry begins with some historical background, highlighting the first case to which a forensic linguistic analysis was applied. Four areas of practice are exemplified through case reports and the linguistic principles that underpin them: forensic discourse analysis, sociolinguistic profiling, authorship analysis, and forensic phonetics. A discussion of forensic linguists' roles in investigative and legal processes and the ethics surrounding practitioners' work concludes the entry.

Origins of Forensic Linguistics

The term "Forensic Linguistics" is said to have originated with Swedish professor of English, Jan Svartvik, who had examined a set of statements presumed to be a verbatim record of what was said by convicted murderer Timothy Evans during the 1949 investigation into the murders of his wife and infant daughter (Olsson 2004). Svartvik (1968) carefully scrutinized the language contained in four statements collected by police and demonstrated that a significant portion of Evans' words, particularly those drawn from the confession portions of the statements, were not likely to have been produced by Evans himself. Instead, Svartvik asserted, the investigators had likely added particularly damning information into his statements, which were then submitted as evidence. Svartvik highlights the compelling contrasts between the language likely to have originated with Evans, an illiterate, uneducated man with a low IQ, and language that would be consistent with someone versed in the discourse of law enforcement. The side-by-side comparison below illustrates the disparity.

Evans' words	Words falsely attributed to Evans
She **never** said **no** more about it.	She was **incurring debt**.
I **didn't** want **nothing** to do with it. . .	I **proceeded to**. . .
I **done** the usual thing.	I **made my way** to 93 Mount Pleasant
She asked me where Beryl and the baby **was**	I **then** fed the baby
My Guvnor used to pay her £50 **what** she used to sign for.	**In a fit of temper** I grabbed a rope.
	He handed me the money **which** I counted **in his presence**

Focusing on a few of the qualitative features that Svartvik uncovered, Evans' language samples contain several features consistent with nonstandard English, including double negatives: "she **never** said **no** more about it"; past tense verbs inconsistent with a more standard, educated variety: "I **done** the usual thing;" and relative clauses headed by relative pronoun "**what**," as opposed to "**that**" or "**which**." The words falsely attributed to Evans contained examples of vocabulary and phrasing likely out the range of someone like Evans, for example, "**incurring debt**," "**proceeded to**," "**made my way to**" and "**in his presence**"; and sentence structure containing hallmarks of a more educated speaking or writing style, including adverbs immediately after subjects, relative pronoun "**which**," rather than "**what**." The differences in sentence structure (syntax) and lexicon (vocabulary) comprised a significant portion of Svartvik's analysis that would demonstrate how Evans likely did not produce a confession, contradicting the police assertion that they had obtained from him a full confession.

Evans was executed for the murders of his wife and daughter, and would be posthumously pardoned after an inquiry showed that his neighbor, John Christie, had killed the mother and child, along with several others. The case today represents a milestone for the relevance of language evidence.

Forensic Discourse Analysis

The field of discourse analysis is vast, owing to the limitless purposes and contexts for human communication. Forensic Discourse Analysis is applied in cases where the meaning or mode of an utterance has some investigative or legal relevance to a case. Was something someone said a genuine threat? Does a racially tinged utterance constitute hate speech? Was the suicide note found beside a lifeless body written by the deceased or did his murderer write it? At what point in a series of written communications to a target (emails, letters, text messages) does the author exhibit language consistent with stalking? These and numerous other situations have called for linguistic expertise to shed light on language evidence, the meaning of which may be open to interpretation. The types of analyses in FDA touch on speech acts (Austin 1962; Searle 1969); speakers' cooperation (Grice 1975); turn-taking (Sacks et al. 1974); coherence and cohesion (Halliday and Hassan 1976); and institutional discourse (Drew and Heritage 1992). Two cases are shown below.

Hate Speech and Hate Literature

In many jurisdictions, the law prohibits people from making hateful statements toward a person or people on the basis of race, ethnicity, sexual orientation, religion, gender, or other demographic feature. It is not always certain, however, whether the spoken or written word meets the legal definition of hate speech or hate literature, at which point a linguist might examine the language in question and the context in which it occurs to assist in making that determination.

In Canada, hate speech and hate literature, including Holocaust denial, are illegal. In the 1990s, a case was brought before the Canadian Human Rights Commission, accusing Ernst Zündel of denying the Holocaust. Zündel, a German national who was living in Canada, disseminated anti-Semitic literature with language that was allegedly consistent with what the Canadian Charter of Rights and Freedoms deems hate language. Gary Prideaux, then a Professor of Linguistics at the University of Alberta, was consulted by the Commission and testified on the linguistic features that qualified Zundel's prolific work as anti-Semitic and holocaust denying (Prideaux 2011). A small sampling of his analysis is shown on the following passage, taken from Zündelsite (http://www.zundelsite.org/chrc/intervenors.html):

> To claim that World War II was fought by the Germans, as the Holocaust Lobby incessantly claims, to kill off the Jews as a group, is a deliberately planned, systematic deception amounting to financial, political, emotional and spiritual extortion. The "Holocaust," first sold as a tragedy, has over time deteriorated into a racket cloaked in the tenets of a new temporal religion - replete with martyrs to the Faith, holy shrines, high priests like Wiesel and Goldhagen, and theologians of the Faith such as Raul Hilberg, Deborah Lipstadt, et al.

Prideaux's analysis starts with the claim that WWII was fought "by the Germans. . .just to kill off the Jews as a group." The claimants here, the "Holocaust Lobby," and the presence of "just" suggests this group holds the simplistic belief that killing Jews was the Germans' only purpose in the war. Even the verb "claim" entails that the thing claimed is contentious, Prideaux indicates. "Holocaust Lobby" itself is a troubling label suggesting that there exists an organized, centralized group of Jews and Jewish sympathizers "advocating the reality of the holocaust" (p. 42) the reality of which is of course, what Zündel tenaciously rejects. That the Holocaust be "sold as a tragedy" suggests deception on the part of the "holocaust lobby" which then became an even more malicious movement in the form of a "racket." Rightfully, Prideaux asserts that "racket" entails a host of negative connotations: "shady, illicit, and negative commercialism, invoking hucksterism and racketeerting" (p. 43), all while cloaked in the accoutrements of religion, again pointing to deceptive, underhanded practices of the target group.

While there is no legislation in Canada equivalent to the United States' First Amendment guaranteeing freedom of expression, if the spoken or written word targets groups and individuals hatefully, then it is a criminal act, and the speaker or author can be prosecuted. As the passage above shows, the hostility that is apparent on the surface of Zündel's writing requires the precision of a linguistic analysis to demonstrate what, in the overall tenor, delivers the hatefulness, in this case Holocaust denial and anti-Semitism, that are both prohibited by law.

Undercover Recordings

Undercover recording is a tool used by law enforcement to aid investigation by capturing key evidence that arises in conversation. Assuming that such recordings are of adequate quality to hear any suspect and cooperating witness or undercover officer, it would seem that this sort of evidence would be very reliable. However, Georgetown linguist Roger Shuy illustrates how what is heard is not necessarily what was said or meant, nor is it necessarily even relevant to the purported criminal action (Shuy 2005).

In Texas v. T. Cullen Davis, the prosecution had relied upon covert recordings of Davis and a cooperating witness, David MCrory, whom Davis had allegedly hired to kill his wife and the judge presiding over their divorce. He also stood trial for the murders of his wife's lover, and her daughter, and the attempted murder of her as well. McCrory had been an employee in Davis' company, and he alleged that Davis confessed to the shootings, and that Davis wanted McCrory to kill his wife and the judge. McCrory cooperated with police, and during several interactions with Davis, he wore a hidden recording device. A portion of one conversation is shown below. Bracketed portions represent overlapping speech; italicized portions represent physical movements of the conversants as seen on video surveillance.

Davis	McCrory
I told him that, uh, to treat you like any other employee and, uh, so don't give me too much pressure in that regard. I can't uh, say you're gonna be gone a day or two every week or so.	
(*Davis gets out of the car and walks to the trunk*)	
	Well look, this fuckin' murder business—
You better—	
	—is a tough son of a bitch
alright	
	Now you got me into this
[give me]	[into this goddamn deal]
[give me]	[right?]
a little	
[noti-, advance notice]	[Now I got j-]
	I got Judge Edison dead for you
Good [inaudible]	
	I'll get the rest of them dead for you.
	You want a bunch of people dead, right?
alright but I	
[uh, you know]	[di-, di-]
	Help me too
[inaudible]	[okay?]
I got to have an alibi ready for Art when the-	
	Okay?
When the subject comes	
[up]	[alright]
(*Davis comes back to the front door of the car.*)	
So give me some advance notice	
	I will
warning	
	I gotta go
(*McCrory gets out of the car and leaves.*)	

Shuy (2005, pp. 46–47)

Using this side-by-side presentation, Shuy was able to demonstrate how Davis and

McCrory were speaking about two unrelated topics. Davis was talking about McCrory's having to account for his absences at work while McCrory was talking about the alleged murder-for-hire. That the two were not addressing the same topic is atypical of natural conversation, because speakers generally adhere to the topic at hand, and the turns are cohesive with one another, meaning that each turn bears some relevance to the turn before it (Halliday and Hassan 1976). The speakers here depart markedly from this principle, evidenced by each maintained his own topic. Davis himself does not in fact utter anything in relation to the killings, nor does he acknowledge what McCrory says about murder. Instead, he remains on the topic of McCrorys' work absences. He utters "good" and "alright" on the transcript, suggesting concurrence with McCrory's explicit reference to "getting people dead"; however, "neither was uttered in a response intonation, but rather as low-pitch discourse markers related to his own topic" (Shuy 2005, p. 48). It is not possible to be certain that Davis even heard McCrory, who was inside the vehicle by the time he was uttering what might have been the most damning words about the contract murder. Shuy refers to this as a "hit and run" in covert recordings, where the cooperating witness produces incriminating information while the suspect is either unaware or unable to respond to it. For language to be incriminating, it must be both cohesive with surrounding discourse, and if not from the mouth of the suspect, it must be evident that the suspect is in unambiguous agreement with what is said.

In both of the examples shown above, the linguists used linguistic principles to provide clarity in cases where language evidence was less straightforward than it appeared, and the importance of a thorough understanding of language samples cannot be overstated when the spoken or written word has some bearing in a legal matter. For a more thorough overview of Forensic Discourse Analysis, see Fadden (2012).

Authorship Analysis

In cases where the authorship of a document or set of documents is contentious, a linguist may be asked to opine on the authorship of the disputed writing. This typically entails comparing the disputed sample with samples known to have been written by a particular individual. There are two broad approaches to authorship analysis: linguistic and computational. Each is considered in this section.

Linguistic Approaches to Authorship

First, the linguistic approach, a more traditional approach to language evidence, typically involves a principled selection of qualitative and quantitative analyses comparing one piece of writing with another to rule in or rule out common authorship (see McMenamin 2002 for one systematic approach of this type). Underlying the linguistic approach is the notion of the *idiolect* (Bloch 1948, p. 7), meaning that "every native speaker has their own distinct and individual version of the language they speak and write" Coulthard (2004, p. 431) While the idea of a linguistic fingerprint unique to each language user is too strong, Coulthard assumes that individuals, with unique vocabularies and preferences for particular combinations and structures, typify the user's style. McMenamin puts it succinctly: "Unique markers are extremely rare, so authorship [attribution] requires the identification of an aggregate of markers, each of which may be found in other writers." (McMenamin 2002, p. 172) In other words, while we all have access to the same units and constructs of language, we as speakers have preferences for patterns and they cluster differently from one writer to the next.

Olsson (2004) walks the reader through several worked examples of authorship comparisons. Among his many techniques, Olsson applies the concept of markedness to his analyses. Markedness dates back to the 1930s commencing with Jakobson (see Henning 1989 for historical introduction), who described pairs of linguistic units as having an asymmetrical relationship to one another. The unmarked form is the most frequently occurring, or normal, ordinary form, and the unmarked form is the less frequently occurring, or unusual, irregular form. In the pair happy/unhappy, happy is unmarked, unhappy, with the derivational prefix un- is marked. Similarly, in the pair "He had often been seen with her" and "He often had been seen with her," the former is unmarked, with the adverb *often* being most commonly positioned between the two auxiliary verbs; the latter is marked, with the adverb *often* occurring in its less frequently found position before the first auxiliary verb.

In one of Olsson's authorship cases, anonymous, menacing letters were received by witnesses who were about to testify against the accused in a rape trail. A set of writings known to have originated with the accused were compared to the letters the witnesses turned over to police.

> Known Authorship 1: "…I was arrested and charged with the rape of Mary. I could not believe this as I am and was impotent."
>
> Known Authorship 2: "…when I was arrested, I was firstly charged with raping Mary, I stated to the Police that I could not rape anyone, as I was impotent, and had been since 1996."
>
> Unknown Authorship 1: "The good fortune is Joe was able through his solicitors, police forensics and medical reports to prove 100 % that he could not and did not rape or have unlawful sex, through his being Impotent and is and has been for five years or more."
>
> Unknown Authorship 2: "Neither would have known that Joe, was to enter hospital to have tests to Proving that he is and was impotent." (Olsson 2004, p. 49)

Given the ordering of the verb pairs on the basis of temporality, i.e., present + past (am/was; is/has been) or past + further past

am and was	(KA1)
was and had been	(KA2)
is and has been	(UK1)
is and was	(UK2)

Olsson concludes the known and unknown texts are likely to have been written by the same author owing to the fact that these verbs appearing in this temporal order represent marked structures, i.e., they occur much less frequently than past + present orderings. Olsson based his conclusion on a number of features in addition to the one described; however, these verb patterns provide a straightforward grammatical example of the type of elements that can be used to match writing samples.

Computational Approaches to Authorship

In recent years, those working at the interface of linguistics and computing science have begun to explore computational means for attributing authorship. Typically, their research focuses on the attribution of authorship in classic works in the humanities and in journalism, but it is not difficult to see the forensic application of computational approaches, in cases where the authorship of a piece of writing is contentious and of legal importance.

Whereas the traditional approaches to authorship involve manual, line-by-line analysis of linguistic features as shown in Olsson's example above, computing approaches involve mathematical and statistical treatments of the language data, making little or no use of linguistic elements (Juola et al. 2006). The leading approaches in the last decade reduce the documents to unpunctuated strings of letters that are then subject to linear discriminant analysis (Baayen et al. 2002; van Halteren et al. 2005), orthographic cross-entropy (Joula and Baayen 2005); common byte n-grams (Keselj and Cercone 2004), among others (see Juola et al. 2006 for more).

These methods have determined authorship accurately at rates much better than chance. In order to determine the most accurate and effective approaches to authorship attribution, contests are held where teams of researchers are given a set of documents (for example, sets of novels, sets of undergraduate terms papers, a set of letters, or other such corpora for which authorship is already known) and they compete to develop the most accurate system to attribute authorship. One such contest took place at a joint 2004 conference held by the

Association for Literary and Linguistic Computing and the Association for Computing in the Humanities, where the most successful team scored an average of 71 % across all of the sample types (see Juola et al. 2006 for an overview of the results and methods, including their own.) While these results are not yet high enough to satisfy the court for forensic determination of authorship, the trends and improvements are encouraging.

Linguistic or Computational?

The two approaches differ tremendously in that the first is wholly linguistic, where lexical features, syntactic constructions, and even rhetorical devices are examined and compared and the expert renders a conclusion regarding a match in the authorship of compared documents, a non-match, or inconclusive results. The second makes no use of linguistic features; rather letters or words are reduced to strings that are compared for similarity on the basis statistical measures.

Members of the International Association of Forensic Linguists, the professional association for linguists and law scholars concerned with language and law, have begun to address the merits of each approach (Solan 2011). Whereas traditional authorship attribution methods have been admitted in UK and US courts, Solan points out that at least for US jurisdictions, they do not meet the strict criteria set out by Daubert and Frye standards, specifically that the error rates of such methods are not known. Computational methods lend themselves well to meeting such criteria; however, current technology has not reached a satisfactory level for forensic identification.

Sociolinguistic Profiling

In the 1960s, William Labov introduced the linguistic variable, which accounts for why we have accents and why social factors like gender and cultural affiliation can contribute to our using language slightly differently from one another. As speakers, we belong to speech communities, meaning within a geographic area, we share many linguistic similarities, i.e., we share an accent. West coast English in the USA, for example, differs from East coast, northern varieties differ from southern varieties, and differences are found between rural and urban communities. The variation is not limited to geographic bounds. Ethnic and racial communities exhibit their own varieties: The English spoken by blacks and whites in the USA varies (Labov 1966; Shuy 1967). Aboriginal varieties of English are noted in Canada, the USA, and Australia (Fadden and LaFrance 2010; Scollon and Scollon 1979; Butcher 2008, etc.) There are differences found between young and old, male and female, and variation is distributed across socioeconomic lines, as well as linguistic variation noted between occupations, where specific vocabulary and phrasing serves the group in the commission of their jobs (see Wardhaugh 1986 for a comprehensive introduction in sociolinguistics). All of these noted variations can be useful clues to narrow down the suspect pool.

In a 2010 “poison pen letter” case, the board of directors and the executive team of a large Canadian company received a disturbing letter from someone claiming to be a staff member in the company. In the letter, the individual listed numerous complaints about the company, including nepotism, favoritism, and poor management. The company’s legal team asked one of the present authors to determine who, within the organization, could have written the letter.

While it is not possible to identify the precise individual on the basis of linguistic features, it was possible to assert that the author was positioned at the upper end of the company hierarchy rather than at the lower, service end.

The following excerpts show how the author had portrayed himself as part of a group of lower management and service staff addressing a variety of issues that would pertain to that portion of the company's workforce.

Claims to be low level management/staff	"Please accept this as a letter to the board on behalf of CompanyName managers and staff expressing our displeasure with the new extended hours."
Concerned with hours and scheduling	"…displeasure with the new extended hours"
	"These new hours have and will turn center manager's jobs into a full-time scheduling function."
Concern with morale	"…create poor morale"
"Us vs. them" stance	"They (executives) will all be home having supper with their families while we are dealing with the traffic that walked in the door at two minutes to six."
	"CEO's speeches about balancing home and work must only apply to him and his group."
No regard for service side staff	"The departments for each VP grow and grow and they care less and less about the front line staff."

The examples above might well have originated with a staff member who deals directly with customers. However, other features strongly suggest that the letter was written by someone who does not work at the service side of the business, but rather someone on the executive side. The following examples address concerns that would not fall within the domain of a service side employee.

Concern with company finances	"Determine how much Company has spent on the bank. How much has Company put into the bank accounts of other organizations and consultants during this project? What have been the returns? How much has been taken out of the wealth that was built by the traditional parts of Company?"
	"Why are they (one arm of the business) getting all their money for the constant spending without any regard to controlling expenses."
	"CEO makes sure the flow of cash…will eliminate accountability."
	"How much have their projects cost the other parts of the company with nothing to show for it but drained capital and resentment?"
Concern with business leadership	"[VPs] show unquestioning support to CEO's decisions and misguided image of himself as a business leader."
	"Like their leader, they (the VPs) are a group of people that are focused on their own self-promotion and have zero business acumen."

(*continued*)

It is unlikely that a low ranking staff member or even a low-level manger would have cause to comment on matters concerning the funding of projects, company cash flow, and capital. Nor does it seem likely that the same individual would make pointed criticism of the company executive because there would not likely be occasion for him to meet, much less work with VPs and the CEO to form opinions about their character.

The author further casts suspicion on his purported low-level position in the company in the following passage:

> Our managers are unwilling to deal with the worst of the staff knowing **they** will get little support form [sic] head office and will likely end up defending **themselves** from whatever accusation is thrown at **them.**

Personal pronouns (they, them) and reflexive pronoun (themselves) used in this sentence strongly suggest his exclusion from the group, and betray the author's claimed identity. The author of the letter was in fact a former vice president who had been forced to resign, and he had written anonymously to the board of directors and the executives to harass and embarrass the CEO.

Sociolinguistic profiling is carried out on language evidence and the linguist constructs an informed opinion on the basis of observable linguistic features. As with other types of profiling, it is an inexact science, because no subject will exhibit all features of their type, and subjects may exhibit features inconsistent with their type. The writer may also attempt to mask features, as the corporate letter case above demonstrates. However, the features to be found in our speaking or writing can in fact provide clues to our demographics. Alone or in combination with psych profiling, sociolinguistic profiling can be very useful when identifying suspects or witnesses at the outset of an investigation.

Phonetics in Criminal Cases

Where language evidence is spoken, phoneticians – that is, linguists specializing in the acoustics and articulation of the spoken word – may lend their expertise.

But forensic interest in the spoken word long predates the science of phonetics. Voice identification by earwitnesses has been admitted at trials dating back at least to the seventeenth century, when William Hulet was accused of having served as the executioner of King Charles I of England. Though masked on the scaffold, Hulet had been recognized by a witness solely "by his speech." This evidence – along with little else, mostly hearsay – resulted in Hulet being found guilty of high treason.

Another high-profile case involving voice identification, the Crime of its own Century, centered on the 1932 kidnapping of the young son of aviator Charles Lindbergh. Lindbergh, who had only heard six words uttered by his son's abductor, initially told a grand jury that "[i]t would be very difficult to sit here and say that I could pick a man by that voice". But at trial, nearly 3 years after hearing the voice of the abductor, Lindbergh went on to identify it as that of suspect Bruno Hauptmann. A year after Hauptmann was executed for this crime, the results of a voice-identification experiment shed light on some of the issues raised during the trial (McGehee 1937). Subjects were asked to pick out a previously heard target voice from a lineup of five similar voices. After an interval of 1 day, and even 1 week, subjects were able to identify the target voice with more than 80 % accuracy, but by 2 weeks that figure had dropped to 69 %, and by 3 weeks to 51 %. By 5 months, recognition had descended to chance level. This suggests some limitations to voice identification, unless a suspect is apprehended in short order, or the unknown voice is captured on a recording.

In 1962, an article in the journal *Nature* described a new and objective procedure for identifying voices with reportedly astonishing accuracy (Kersta 1962). The technology was not new, having been developed (under the name "sound spectrography") as part of the World War II effort to track German troop movements. The methodology – essentially pattern-matching among printouts of vocal energy – was based on a "theory of invariant speech," which claimed that each person's voice pattern is unique. (Little scientific support was ever offered for this theory, and it was soon undermined by studies involving vocal disguise.) Notably, the author of that article discarded the old name for the physical record of sound spectrography ("sound spectrograms"), and instead coined the word "voiceprints," which implied a level of reliability similar to that of fingerprints.

However, phoneticians (notably Professors Peter Ladefoged of UCLA and Harry Hollien of the University of Florida) argued strenuously that "voiceprints" of the highly mutable and continually shifting human voice are not nearly as reliable as fingerprints. The former are graphic representations of the energy produced by fluid articulatory gestures (such as the rounding of the lips and the raising of the tongue) which can occur simultaneously, sequentially, or not at all; the latter are direct physical impressions of digital topography, and are immutable.

Not long after the publication of the *Nature* article, the "voiceprint" technique began to be used by prosecutors in criminal cases. It was immediately challenged on the basis of not having gained general acceptance in the scientific community. Different jurisdictions reached different conclusions in this regard. However, within a few decades, Kersta's original methodology – mere pattern-matching, without even the need to listen – had been all but abandoned.

But a growing body of research has demonstrated that the acoustic information present in sound spectrograms, when interpreted in the light of phonetic principles, can indeed enhance the auditory identification of voices. Some robust results have been reported for lengthy voice samples. And even voice samples of moderate size can derive benefit from phonetic analysis.

Additionally, if two speech samples are lengthy enough to be compared in considerable detail, one may be able to evaluate the strength of the evidence in the form of a "likelihood ratio."

Numeric ratios of this sort can be assigned verbal values, such as "very strong evidence [to support the hypothesis of speaker identity]," "strong evidence," "moderately strong evidence," etc., and presented to the trier of fact.

Some of the phoneticians who had initially been most critical of the forensic use of sound spectrograms came to moderate their views, once those spectrograms were no longer associated with a superficial pattern-matching technique and no longer bore the unduly suggestive name "voiceprints." Ladefoged, for one, did so while advocating procedural safeguards such as sufficient length (far exceeding the FBI standard), clarity (a signal-to-noise ratio where the signal is higher by 20 dB), and frequency range (at least 3,000 Hz). Under good conditions, he later wrote, "virtually all phoneticians would agree that acoustic phonetic analyses have *some* evidential value" (Ladefoged 1982).

One example of a phonetic analysis that was considered to have evidential value was that presented at the trial of Paul Prinzivalli, who was accused of phoning in 25 separate bomb threats to Pan American World Airways in 1984. Prinzivalli, a disgruntled Pan Am employee, had been identified as the suspect by two senior supervisors – though his immediate supervisor disagreed, pointing to differences in the "inflection, accent, tone, depth, pronunciation, timbre" of his voice. The police nevertheless charged Prinzivalli with having made the bomb threats, and he faced a possible 6–8 years in prison.

Sound spectrograms of the would-be bomber's words and of Prinzivalli's exemplar of those same words displayed numerous differences. The lower vocal resonances, which mainly indicate which words are spoken, differed in a way that suggested the presence of two different accents. The upper vocal resonances, which can provide evidence of speaker identity, were attenuated, but still suggested the presence of two different individuals. In particular, the fourth resonance, where visible, did not match across conditions. And the third resonance closely tracked the rises and falls of the second resonance in the voice of the suspect, but remained high and steady in the voice of the would-be bomber.

These differences, presented in court by Ladefoged and by one of the present authors, were corroborated by a dialect study conducted by William Labov of the University of Pennsylvania, the author of numerous books on this topic, including the authoritative *Atlas of North American English*. Labov concluded that the unknown voice was a Bostonian, whereas the suspect was a lifelong New Yorker.

The judge found Prinzivalli not guilty, calling the linguistic evidence "objective" and "powerful."

Automatic Analyses

When an abundant amount of speech is available for analysis, the "auditory-acoustic" method described above may be supplanted by a computerized system that does not examine individual phonetic parameters at all. Instead, such "automatic" systems treat the speech sample as a single, continuously varying complex vibration, which can be modeled mathematically and then compared to another speech sample. "Automatic" methods of analysis are usually quite accurate, but they typically have a high rejection rate (otherwise associated with short samples, or with noisy samples of moderate length). In general, the optimal methodology will be dictated by the length and the quality of the speech signal.

Phonetics in Trademark Disputes

The phonetic methodology described above can also be employed in the resolution of trademark disputes. According to US law, trademark infringement may occur when two marks denoting similar products or services are sufficiently similar to cause confusion (or potential confusion) among consumers. When infringement is determined to have occurred, the newer ("junior") mark must give way to the senior mark. But there is often controversy over what constitutes "sufficient" similarity. What may seem misleadingly similar to one party may seem quite distinct to another.

The situation is difficult enough to resolve when it is limited to orthography and to related issues of trade dress such as the color and font of the mark. It becomes even more difficult for experts when sounds are taken into consideration.

A manufacturer may specify the intended pronunciation of a newly coined trademark directly on the package, or in a source such as the *Physician's Desk Reference*, or through advertising in broadcast media. But it is by no means clear whether the public will accept (or even notice) such specification, or instead fall back on the familiar rules of spelling. Such a dispute was, for example, at the heart of trademark litigation between the manufacturers of competing birth-control pills, OVRAL and B-OVal (Shuy 2002). The manufacturers of OVRAL maintained that their trademark name was to be pronounced with final stress (o-VRAL, to rhyme with go-ALL). However, final stress is not very common in English words. A reverse-order (or "rhyming") dictionary shows that, of almost 2,000 multisyllabic English words ending in –al, only about a dozen bear final stress.

A growing body of evidence shows that knowledge about such preferences, along with other spelling-to-sound relationships of the English language, is represented in the mind of the reader and is called upon when necessary, such as when words are encountered for the first time (Treiman et al. 2002). This can even help us to read nonce words, such as those in Lewis Carroll's poem *Jabberwocky*. But such rules can be a powerful counterweight to the trademark holder's intentions, and may indeed result in consumer confusion.

Litigants often commission consumer surveys to determine the likelihood of consumer confusion, but such surveys must be carefully crafted to avoid suggestiveness, and to adequately sample the marketplace.

As an alternative, or a supplement, to such surveys, litigants have called upon linguists to present any phonetic facts that might give rise to, or else rule out, the likelihood of consumer confusion. A linguist trained in phonetics can call attention to any articulatory or acoustic similarities that might exist between the sounds of two trademark names. The use of the International Phonetic Alphabet can override the vagaries of English spelling (such as orthographic "qu," which, despite appearances, is not a consonant-vowel sequence). And a table of documented misperceptions, a staple of the psycholinguistic literature, can suggest which of all possible pairs of sounds are most likely to be confused by the consumer.

All of the above arguments were made by one of the present authors in trademark litigation between two competing mortgage companies, Ameriquest and Americrest. Evidence from phonetics (articulatory lip rounding, acoustic overtone patterns, and the International Phonetic Alphabet) and from psycholinguistics (listeners' documented tendency to confuse the distinguishing letters of these trademarks under laboratory conditions – in this case a 28 % confusion rate, even without advertising jingles playing in the background) were offered as arguments for the potential confusability of these two marks. The case was promptly settled.

Forensic Linguistics and Ethics in the Legal System

As more linguists find themselves consulted by law enforcement and legal counsel on cases involving language evidence, the Linguistic Society of America compiled a four-part statement of ethics to which linguists are expected to adhere. First, a consulting linguist must act with professional integrity: He or she is to provide an objective analysis and compensation cannot be dependent on the outcome of any findings. Second, there must be transparency regarding the methodology used, as well as a clear account of any data, equipment, statistical tests, and software used. The linguist must also make clear any known limitations of the aforementioned and his or her own abilities. Third, the linguist must not disclose any confidential information relating to his or her casework without the consent of affected parties. Any research arising from casework should be anonymized such that there is no identifiable information appearing in presentations or publications. Furthermore, the linguist must not take on casework that can result in a conflict of interest, nor should she or he communicate informally or unofficially with an opposing party's linguist or legal counsel.

Fourth, linguists are strongly encouraged to communicate that they abide by the Code of Ethics passed by the Linguistic Society of America if practicing within the USA, although others are free to subscribe to this code as well. Furthermore, linguists are encouraged to abide by the codes agreed upon by any other professional organizations such as the International Association of Forensic Linguists and the International Association of Forensic Phoneticians.

Linguistics, like most other social sciences, is a wide field with a variety of subfields, and, as a discipline, is taught in many university campuses. Many research universities offer doctoral in linguistics. In American and UK courts, it is largely the decision of the trial judge to admit (or not) a linguist and his or her testimony, and it appears to be happening with greater frequency as the usefulness of linguistic analysis is understood.

Related Entries

- ▶ False Confessions and Police Interrogation
- ▶ Hate Crime
- ▶ Law of Police Interrogation
- ▶ Philosophy of Forensic Identification

Recommended Reading and References

Austin JL (1692) How to do things with words. Harvard University Press, Cambridge

Baayen RH, van Halteren H, Neijt A, Tweedie F (2002) An experiment in authorshihp attribution. In: Proceedings of JADT 2002, Université de Rennes, St. Malo, pp 29–37

Bloch B (1948) A set of postulates for phonemic analysis. Language 24:3–46

Butcher A (2008) Linguistic aspects of Australian Aboriginal English. Clin Linguist Phonet 22:625–642

Cambier-Langeveld T (2007) Current methods in forensic speaker identification. Int J Speech Lang Law 14 (2):223–243

Coulthard M (2004) Author identification, idiolect, and linguistic uniqueness. Appl Linguist 25(4):431–47

Coulthard M, Johnson A (2007) An introduction to forensic linguistics: Language in evidence. Routledge, London

Criminal Code, R.S. c. C-46 (1985) Retrieved from http://laws.justice.gc.ca/en/C-46/index.html. Accessed 11 June 2013, 11:14 PST

Drew P, Heritage J (eds) (1992). Talk at work: interaction in institutional settings (Studies in interaction sociolinguistics, 3 edn). Cambridge University Press, Cambridge

Fadden L (2012) Forensic discourse analysis. In Chapelle C (ed) The encyclopedia of applied linguistics. Wiley-Blackwell, Malden

Fadden L, LaFrance J (2010) Advancing aboriginal English. Can J Native Education 32

Grice HP (1975) Logic and conversation. In Jaworski A, Coupland N (eds) The discourse reader. Routledge, New York, pp 76–87

Henning A (1989) Markedness – The first 150 years. In Tomic O (ed), Markedness in synchrony and diachrony. Mouton de Gruyter, Berlin, pp 11–46

Halliday M, Hasan R (1976) Cohesion in English. Longman, London

Joula P, Baayen H (2005) A controlled-corpus experiment in authorship Identification by cross-entropy. Lit Linguist Comput 20(Suppl 1):59–67

Juola P, Sofko J, Brennan P (2006) A prototype for authorship attribution Software. Lit Linguist Comput 21:169–178

Kersta L (1962) Voiceprint identification. Nature 196:1253–12857

Keselj V, Cercone N (2004) CNG method with weighted voting. Ad-Hoc Authorship Attribution Contest ACH/ALLC. Gothenburg, Sweden

Kreiman J, Sidtis D (2011) Foundations of voice studies. Wiley-Blackwell, Malden

Labov W (1966) The linguistic variable as a structural unit. Washington Linguist Rev 3:4–22

Labov W, Ash S, Boberg C (2005) Atlas of North American English. de Gruyter, Berlin

Ladefoged P (1982) A course in phonetics. Harcourt Brace Jovanovich, San Diego

Ladefoged P (1984) Review: the phonetic bases of speaker recognition by F.J. Nolan. J Phonet 12:85–89

Ladefoged P, Disner S (2012) Vowels and consonants. Wiley-Blackwell, Malden

McGehee F (1937) The reliability of the identification of the human voice. J Gen Psychol 17:249

McMenamin G (2002) Forensic linguistics: advances in forensic stylistics. CRC Press, London

Morrison G (2010) Forensic voice comparison. In Freckelton I, Selby H (eds), Expert evidence (Ch. 99). Thomson Reuters, Sydney, Australia

Morrison G (2009) Forensic voice comparison and the paradigm shift. Sci Just 49:298–308

Olsson J (2004) Forensic linguistics: an introduction to language, crime, and the law. Continuum, London

Prideaux G (2011) Linguistic contributions to the analysis of hate language. Int J Law Lang Discourse 1(1):27–52

Sacks H, Schegloff, Jefferson G (1974) A simplest systematics for the organization of turn-taking for conversation. Language 50(4):696–735

Scollon R, Scollon S (1979) An ethnography of speaking at Fort Chippewyan, Alberta. Academic, New York

Searle J (1969) Speech acts: an essay in the philosophy of lanaguage. Cambridge University Press, Cambridge

Shuy R (1967) Discovering American dialects. National Counsel of Teachers of English, Champaign
Shuy R (2002) Linguistic battles in trademark disputes. Palgrave Macmillan, New York
Shuy R (2005) Creating language crimes: how law enforcement uses (and misuses) language. Oxford University Press, New York
Solan L (2011) Ethics and method in forensic linguistics. In Tomblin S, MacLeod N, Sousa-Silva R, Coulthard M (eds), Proceedings of the international association of forensic linguists' 10th biennial conference. Centre for Forensic Linguistics, Aston
Svartvik J (1968) The Evans Statements: a case for forensic linguistics. University of Gothenburg Press, Gothenburg
Treiman R, Kessler B, Bick S (2002) Context sensitivity in the spelling of English vowels. J Mem Language 47:448–468
van Halteren H, Baayen RH, Tweedie F, Haverkort M, Neijt A (2005) New machine learning methods demonstrate the existence of a human stylome. J Quant Linguist 12(1):65–77
Wardhaugh R (1986) Introduction to sociolinguistics, 2nd edn. Blackwell, Cambridge

Forensic Palynology

Andrew R. Laurence and Vaughn M. Bryant
Department of Anthropology, Texas A&M University, College Station, TX, USA

Synonyms

Forensic palynology (the study of pollen grains)

Overview

Forensic palynology is an underutilized, but important tool for obtaining trace evidence from suspects, items thought to be associated with a crime scene, or for determining the geolocation of a sample. First used more than 50 years ago in Austria, it is now often used in the United Kingdom, New Zealand, and more recently, in other countries of Europe and Asia, but is still underused in the United States. Forensic palynology has gained importance for its ability to provide information about pollen and spores trapped in clothing or other items of interest needed to help resolve criminal and civil cases, including instances of homicide, terrorism, genocide, bombings, forgery, theft, rape, arson, counterfeiting, manufacturing and distribution of illegal drugs, assault, cases of hit and run, poaching, and identity theft. The foundation for using pollen in forensic applications comes from the discipline of pollen analysis, which began a century ago as a way to search for clues about past environmental changes. A few examples demonstrating the utility of using pollen and spores in forensic applications during the last half century are discussed. Finally, some of the recent analytical techniques that are being used to try to increase the precision of pollen identifications related to a variety of criminal and civil cases are examined.

Introduction

Palynology (the study of pollen and spores) has historically been widely underutilized in forensic science. Traditionally, forensic palynology focuses on legal evidence obtained from the study of pollen and spores that are associated with a crime scene or other aspects related to situations involving the law. As a discipline, the first recorded use was a little more than one-half century ago, but even today the use of this technique is relatively unknown or underutilized in many regions of the world. Even in those areas where forensic palynology is routinely used as a forensic tool, such as the United Kingdom and New Zealand, its utility for forensic science is often questioned.

One of the earliest recorded and successful cases involving the use of forensic palynology occurred in Austria in 1959, where pollen was used to link a suspect with a murder and to the crime scene where a shallow grave contained the murdered victim's body (Erdtman 1969). A decade later, pollen from the suspected holy relic called the Shroud of Turin was recovered and then analyzed by Max Frei in the 1970s. Using pollen trapped on tapes that had been lightly pressed against the Shroud, Frei claimed the relic could be linked to its purported origin and thus suggested authenticity (Frei 1982). Over the next decade, there were a few additional attempts to use pollen as trace evidence in both civil and criminal cases, but the next use that gained widespread attention occurred in New Zealand (Mildenhall 1990). Dallas Mildenhall, a palynologist trained in geology was asked to help solve a case involving the suspected poaching of deer to obtain the velvet from their newly formed antlers. The recovery of pollen from the seized deer velvet did prove the deer had been killed illegally and thus led to a conviction of the poacher. Finally, forensic palynology in the United Kingdom began when Patricia Wiltshire (1993) first used pollen in helping to solve criminal cases in the United Kingdom. Since the early 1990s, the United Kingdom has led the world in the routine application and use of forensic palynology as part of many criminal investigations.

In the United States, there appeared to be little interest in following the examples set forth in the United Kingdom and New Zealand in using forensic palynology until after the destruction of the Twin Towers of the World Trade Center in New York City on September 11, 2001. Terrorists, using commercial airplanes flown into the two skyscrapers, claimed the lives of over 3,500 innocent victims and created an immediate need to learn more about the terrorists who committed that crime. In a search of information, authorities focused on using every known forensic technique, including the use of pollen as a forensic tool. Although the recovered pollen evidence associated with the terrorists did not produce any immediate useful clues, the addition and use of recovered pollen as a new forensic trace evidence tool was finally accepted as being potentially important for future uses in the United States.

Foundations of Palynology

To use forensic palynology effectively and to explain its potentials to others, one needs to understand factors that control the type and amounts of pollen and spore production and how those pollen and spores are dispersed. Among the cryptograms, spores are dispersed into either air or water and use those mediums as a transportation vector. Among the gymnosperms, all species produce wind-dispersed pollen. In angiosperms, the vast majority of species now rely on insects, birds, or mammals to vector the pollen from one plant to another. A limited number of angiosperms still rely on wind dispersion for its pollen.

All of the spore and pollen species that rely on wind dispersion generally produce large numbers of spores or pollen because of the inefficiency of wind dispersion. Conifers and some of the angiosperm species are capable of producing hundreds of millions of pollen grains per plant, and some spore-producing species, such as *Ganoderma* (Kadowaki et al. 2010), reportedly produce and disperse several billion spores during one annual dispersal period of about 6 months. The large number of produced pollen and spores cast into the air, and the inefficiency of that technique provides large numbers of essential clues that are captured as part of the pollen rain at each location. That unique mixture of airborne pollen and a few insect-pollinated pollen and spores at each location is what forensic palynologists primarily rely upon for clues related to a specific crime scene or to assign geolocation to some object of unknown origin (Fig 1). Some pollen grains are tiny, are lightweight, and are aerodynamically designed to travel easily in air currents, while other wind-pollinated types, which are larger and heavier, do not usually get dispersed far from their source. Some pollen grains have been recovered more than 1,000 km from their nearest known source, but as a general rule, the vast majority of airborne pollen and spores do not travel too far from their dispersal source, which in many cases means that much of the pollen falls to the surface within a radius of about 100 m (Tauber 1965) (Fig. 1).

Although the large number (about 75 %) of angiosperms that rely on some type of insect or animal vector to transport their pollen grains usually produces miniscule amounts of pollen, which often range in number from as few as 10

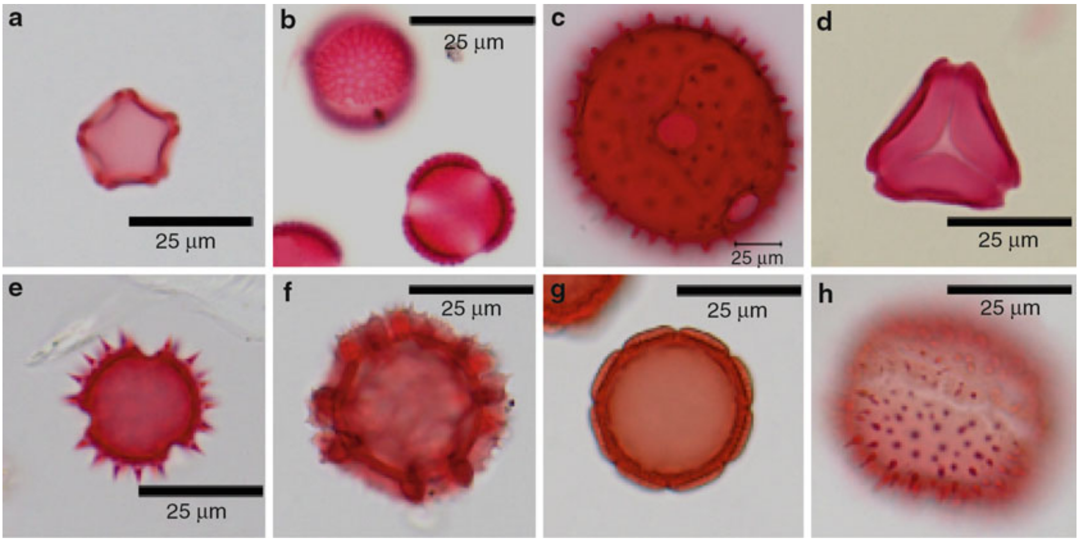

Forensic Palynology, Fig. 1 Pollen grains from (**a**) *Alnus crispa* (Aiton) Pursh., (**b**) *Brassica nigra* (L.) W.D.J. Koch., (**c**) *Cucurbita pepo* L., (**d**) *Eucalyptus calophylla* (Lindl.), (**e**) *Helianthus maximiliani* Schrad., (**f**) *Taraxacum officinale* F.H. Wigg., (**g**) *Spermacoce tenuior* L., and (**h**) *Nuphar luteum* (L.) Sm

up to about 1,000 pollen grains per anther, their pollen is often a key component in any forensic sample. By design, the non-wind-borne plant species produce pollen grains that are often highly ornamented, are thick walled, have a heavy mass, and are covered with sticky lipids and waxes. Those are all factors that make for the ideal transport of pollen by some animal or insect vector. The ornate surface ornamentation of these pollen types is designed to hold sticky lipids that will keep the pollen grains firmly stuck to the hairs or bodies of pollinators until those pollinators land on another flower. At that point, the sticky surfaces of the attached pollen grains adhere to the surface of the stigma of a new flower, and if a pollen grain falls in the right location, fertilization will occur. The thick pollen walls and heavy mass of these insect- and animal-pollinated types ensure that the fragile contents of the pollen grains remain protected from abrasion and sudden changes in humidity during the flight time from one flower to another. From the above reasons, only rarely do pollen grains from these types of plants become part of the normal pollen rain of a location. Nevertheless, it is that characteristic that becomes especially useful in many types of forensic situations.

When people or objects come in contact with insect-pollinated plants, especially ones that might be in flower, some of the sticky pollen can often be transferred from the plant's leaves or flowers to the person or object. Because most of these types of pollen become associated with people or other objects "only through direct contact with the plants," that type of evidence, when linked to similar plants at a crime scene, can infer evidence of direct association. The highly unlikely argument that such pollen grains were deposited on a person or object by accident from airborne sources as part of the normal pollen rain makes these types of pollen grains an ideal evidence for direct association. Another advantage of many of these insect- and animal-pollinated types is that their sticky surface, durable pollen walls, and ornate surface morphology enable them to adhere firmly to objects they touch. Even after persistent attempts to remove all such pollen traces from objects associated with a crime scene, some of these pollen types often remain and can help to place a suspect at the scene of the crime (Milne 2005).

Utility of Forensic Palynology

The analysis of pollen and spores (collectively called palynology) is recognized as an effective forensic tool for a number of reasons. First, many types of pollen-producing plants (angiosperms and gymnosperms) and spore-producing cryptograms (algae, fungi, ferns, mosses, liverworts, etc.) disperse vast quantities of pollen or spores as part of their reproductive process. For most of these, they rely upon wind currents to carry these single-celled pollen or spores from the dispersal source to another location where they can carry out part of the reproductive cycle. Because this process is haphazard, the vast majority of these dispersed cells eventually fall to the ground in a thin coating called the "pollen rain." In some regions, the amount of dispersed pollen and spores is so great that once deposited, they can coat exposed surfaces with a thin yellow layer or form a yellow scum on water surfaces (Faegri and Iversen 1989). Although not a precise measurement of the surrounding vegetation, the pollen rain at each location provides a snapshot of that area's vegetation. Even though there is not a one-to-one correlation between the percentages of each plant type in an environment with the precise percentages of each pollen type in the pollen rain, the resulting "pollen print" for each area becomes an effective way to identify each particular region of the world. The precision required to recognize a specific region, based on the pollen print from some forensic sample, often depends on the experience and knowledge of the pollen analyst and that person's awareness of how the presence and various percentages of each pollen type in a sample can often relate to the presence of certain types and quantities of plants.

A second reason why pollen and spores can become effective forensic clues pertains to their tiny, invisible size. All are microscopic in size and the sizes can range from only 1 or 2 μ (1,000 μm is equal to 1 mm) for some spores and up to over 200 μm in size for large pollen grains. Because of their tiny size, many are airborne and can then be deposited or become trapped on almost any type of surface. This means that at any geographical location, pollen or spores from local and even some regional vegetation, or in specific cases, the pollen and spores associated with a crime scene can become effective markers for those locations and can often link a suspect or some object with a precise geographical region or a specific crime scene.

Third, there are over 330,000 different pollen-producing plants and thousands more when adding the cryptogram species that produce spores. However, differences in the pollen and spores of closely related species, or in some instances related genera, may appear so similar that precise identification using only the resolution capacity of light microscopy (LM) is tenuous and sometimes unreliable. In those cases, when needed, greater precision of pollen and spore identification can often be achieved through detailed studies using the higher resolution capabilities of a scanning electron microscope (SEM) and/or the resolutions available with a transmission electron microscope (TEM). Nevertheless, obtaining these greater degrees of precision can be costly and time-consuming because they must rely upon comprehensive SEM and TEM pollen and spore reference micrographs for making direct comparisons. Because of the time and cost of using SEM and TEM, other analytical techniques are now being considered to speed the identification of both pollen and spores, which will be discussed later in this entry.

A fourth reason why most pollen and spores are useful as forensic tools is that the majority of them are highly resistant to destruction or decay. This is why the oldest geological evidence of organic material, related to plants, comes from what is believed to be very ancient spores (Faegri and Iversen 1989). This ability to remain preserved for hundreds or even millions of years means that pollen and spore evidence, if collected and stored properly, can still be used effectively many years after some event has occurred. In one such case, pollen and spore evidence collected more than 30 years ago from the clothing of a murdered victim was examined and then used to help identify where that person may have lived prior to being killed.

It should be noted that even though pollen and spores are resistant to decay, in certain types of oxidizing environments, they can degrade very rapidly and disappear completely in only a few months or years. In addition, studies reveal that in certain types of depositional environments, such as alkaline soils, certain fragile types of both pollen and spores can degrade almost immediately, while other types remain in pristine or recognizable conditions for many decades or hundreds of years (Traverse 2007).

In some regions, and in some countries, law enforcement personnel continue to expand the types of collected evidence they hope can reveal pertinent information based on the trapped pollen and spores. The list of items that can trap pollen and spores, and thus are being collected and examined, has become almost limitless (Mildenhall et al. 2006). During the past decade, a partial list of items that forensic palynologists have been asked to examine for pollen and spore evidence includes the vacuum contents trapped on the pages of books or on various documents, material trapped on the surface of weapons or inside homemade bombs, pollen and spores collected in air conditioning filters or in air filters from different types of vehicles, pollen and spores trapped on or in all sorts of transportation ranging from boats and airplanes to trucks and bicycles, briefcases, illegal and counterfeit drugs, laptops and cell phones, all sorts of clothing and the contents in suitcases, pollen and spores from victims that were buried, and even the pollen trapped on pieces of wire and shoelaces. Pollen and spore evidence examined by other forensic scientists has also been used to resolve situations involving forged documents, fake antiques, authentication of paintings by master artists, removal of artifacts from historic or archaeological sites, illegal poaching of animals or fish, and cases involving the illegal pollution of the environment (Mildenhall et al. 2006).

Although forensic palynology has been accepted as a proven tool by some, it does need to be pointed out that the reliability of using pollen and spore data as forensic trace evidence is often called into question by courts and law enforcement officials, particularly in the United States. Because forensic palynologists often present their results as relative percentages of various pollen types within the total pollen and spore composition of a sample, some question the uniqueness of such data, especially when a forensic palynologist attributed a sample to a single location or to an actual crime scene. Questions often arise about the precision of the overall pollen data and the precision of identification of individual pollen taxa.

Forensic pollen specialists generally present their conclusions in terms of probability of association and often cite the importance provided by the combination of many pollen and spore taxa in a single sample. Often when challenged to produce "statistical data" that could be used to support or negate the importance or relevance of a pollen sample, or the potential link of a pollen sample to either a suspect or to a crime scene, the palynologist cannot effectively do so. This inability to provide statistical confirmation results from an almost endless list of variables that influence the deposition of pollen types and their ratios within a single sample deposited on some surface or person. Even though some scientists have championed the limited use of statistical techniques, most types of statistical analyses cannot be adequately adapted in ways that will test the multiple unknowns or conclusions based on forensic pollen assemblages (Mildenhall et al. 2006).

One question that is frequently raised during investigations or in court focuses on whether or not two different locations will signal identical pollen spectra. The chance of that occurring is virtually impossible and thus far, an identical match in the pollen spectra at two different locations has yet to be proven. The unlikeliness that this possibility would occur depends to a great extent on how precise the attempt might be. Experience has shown. Minimal pollen counts of only a few 100 pollen grains from comparative samples collected at two separate locations might suggest that both locations appear nearly identical in terms of their pollen signatures. However, as the pollen signature of each sample is refined by counting larger numbers of pollen grains and spores from each sample, the potential of both samples providing nearly identical pollen spectra disappears. The list of variables at any given location is numerous, and each one contributes in some way to the formation of a pollen spectrum for that location, which in turn produces a pollen spectrum unlike those from similar locations.

Instead, the focus should be on the knowledge, skill, and research experience of the pollen analyst and on the discussion of which variables may have been a factor in producing the pollen assemblage in question. The reputation, previous work, and the acceptance of the person's skills by courts and by other forensic palynologists should serve as a guide to support the conclusions of competent scientists and also prevent evidence and testimony being accepted from less-skilled or under-qualified individuals.

Types of Forensic Palynologists

There are essentially two types of forensic palynologists, based upon the types of information or investigations that are being usually undertaken. Although some forensic palynologists are called upon to work in both areas, often each group tends to specialize and focus mostly on one type of needed information. The primary type of forensic pollen specialist becomes directly involved in the investigation of crimes as they pertain to situations involving victims, suspects, actual crime scenes, and items suspected of being associated with one of the previous situations. These individuals generally work directly with local, state, and national law enforcement agents, they usually work within a law enforcement agency or established forensic lab, and they are generally included as part of the forensic teams that are the first to reach crime scenes. This type of forensic palynologist relies on his/her botanical knowledge, previous experience, and an ability to evaluate a crime scene and/or other places of forensic interest. He/she should determine what types of comparative pollenreference samples need to be collected to properly reflect the pollen print of the crime scene and can also be used later for comparisons between the crime scene and evidence recovered from potential suspects and other items. These forensic palynologists need to have training in palynology, plant ecology, and plant taxonomy and should also understand the relationships of those botanical areas to other crime scene features such as, soils, past vegetational changes, and the topographic features of the landscape in and around the general region.

The other type of forensic palynologist worked mostly with questions related to geolocation. In other words, this type of individual is often called upon to examine items related to crimes, items of unknown origin, or items associated with threats to national security or terrorist activities. The goal in each of those types of studies is to determine the geographical location where such items originated or were used. For scientists working in this second area of forensic palynology, their background and training should include all of the areas and aspects needed to conduct the first type of forensic work, but in addition, they also need other related types of expertise. These individuals should have a comprehensive understanding of pollen and spore dispersal patterns, pollen productivity as it pertains to different taxa, knowledge about how topography and climatic conditions can affect the potential pollen rain in a region, pollen sinking speeds and potentials of key types of both pollen and spores being recycled, and access to an extensive pollen database covering many potential pollen and spore types found in different geographical regions from the Arctic to the equatorial tropics in all of the major continents.

One of the problems that also affect potential geographical interpretations, based on recovered pollen and spore data, is the never-ending process of expanded and new plant distributional ranges related to the introduction of agricultural or ornamental plant species into new regions of the world, the natural catastrophes such as volcanic eruptions and fires, or the widespread changes in landscapes created by human modifications. Before the Age of Discovery and the worldwide expansion by European nations, if pollen from key types of plants, such as maize (*Zea mays*) and chili peppers (*Capsicum*) (native to the New World), *Melaleuca* (native to Australia and some areas of Southeast Asia), and tamarind (*Tamarindus indica*) (native to Africa), were found in a forensic sample, it could be linked quickly to some specific geographical location. Today, however, those plants, and hundreds of other ones, have become almost pandemic and their pollen is now found in many regions far beyond where those plants were once restricted. Therefore, a major problem confronting these types of forensic analysts is an understanding of the potentials in each world region for the introduction of foreign ornamentals that are not native to those areas.

Collecting Forensic Samples

For those forensic palynologists, who work closely with law enforcement personnel, an early arrival at a crime scene is often essential. Pollen and spores at a crime scene are easily disturbed, removed, or contaminated with unassociated pollen and spores that arrive on the clothing of law enforcement personnel and forensic teams investigating the crime scene. Ideally, the forensic palynologist should be given early access to a crime scene so that careful studies of the plants and pollen sources can be collected, noted, and photographed. Normally, the forensic palynologist will need to collect undisturbed comparative samples of surface soils of various locations in and around the actual crime scene, which, when analyzed, form the basis for comparisons with future pollen evidence thought to be related with the crime scene. Unfortunately, potential comparative samples collected improperly or gathered by untrained personnel can often endanger the reliability of the pollen evidence or render samples useless as evidence. Such problems can also provide a basis for the dismissal of evidence in court.

One should collect many reference or comparative samples from a crime scene. Sometimes control samples need to be collected from both sides of a path or walkway or from a series of closely spaced areas in and around a crime scene. Without this type of precise knowledge, it becomes difficult to argue either for or against the confirmed association of a pollen assemblage found on a suspect's shoes, or car, or clothing, with the pollen types found at the actual crime scene. Therefore, even though a large number of comparative samples might be collected from a crime scene, not all of them may need to be examined. However, because the pollen spectrum in each comparative sample may vary slightly in reference to the pollen types and percentages of each taxon, the combined pollen spectra from a variety of comparative samples provide a potential range of pollen variation that can be expected in and around the actual crime scene. Once a crime scene has been trampled upon by law enforcement personnel, teams of forensic specialists, and eventually by others who might be curious about the events that took place, the value of collected pollen evidence is reduced and those samples become subject to suspect from post crime pollen contamination or removal.

An example that illustrates the importance of collecting adequate numbers of control samples focuses on samples collected from the ground where a sexual attack occurred. A woman was kidnapped at knifepoint while jogging in a New Zealand city park. She was then sexually assaulted and killed in a nearby wooded area. A man who lived near the park had been seen near the jogging trail in the park just before the woman disappeared. Later, when questioned by the police, the man admitted being in the park, but denied he had seen the victim and said he had never been in the nearby wooded area. A search warrant was granted and items of clothing were gathered from the suspect's apartment. Of particular interest were a dark woolen sweater and a pair of soiled pants, which might have been worn during the kidnapping and assault. Pollen studies of the sweater and pants revealed that each garment contained a rich combination of pollen and spores from pines and ferns. Comparison samples collected from the crime scene revealed a dominance of both pine pollen and fern spores. Likewise, the jogging shorts worn by the murdered woman also contained pine pollen and fern spores.

The pollen spectra from all three groups of samples (suspect's clothing, comparison samples, jogging shorts from the murdered woman) were consistent and revealed that the only way so much pine pollen and fern spores could have become trapped in the clothing was that both the victim and suspect must have struggled on the ground in the wooded pine forest where there was only one local area an understory covering of fern plants. Because pine trees are not native to New Zealand and are used as ornamentals in parks and grown for its timber, there was only one local area near the park where both pine trees and ferns grow together. Thus, the pollen evidence did not prove the suspect was guilty, but it did show he had been in the same wooded area where the crime scene occurred; an area he denied he had ever visited (Mildenhall 1992). As with all types of forensic trace evidence, it is essential to keep accurate records and photographs of how and where each comparison sample was collected, ensure that complete records of the chain of custody are maintained, and provide detailed information related to the method of analysis and the steps taken to provide for security of the samples when not being directly examined. When doubt can be raised about the possible contamination of a forensic pollen sample at any point between the time of collection and the time of final analysis, then that sample is compromised as evidence, even if it might clearly establish a link between a suspect and a crime.

Once all control samples have been collected, they should be stored in airtight, contamination proof containers that are clearly marked. Because comparative reference samples might contain moisture, which could encourage the growth of microbes, a microbe-killing solution, such as alcohol, needs to be added to each sample. Another alternative, which is effective, is to place samples in cold storage at or below 0° C, which will retard the growth of most microbes and will not damage the pollen or spores in the sample (Holloway 1989).

There are three potential areas where forensic pollen samples could be ruined or compromised. The first area is during the collection phase, the second area is during the extraction of pollen from the matrix material of collected samples, and the third area is in the identification and analysis of the collected sample.

Forensic palynologists often only have very small amounts of evidence material available for analysis; the exception is the collected comparative samples at a crime scene where ample amounts of soil can be collected for each sample. There is almost no limit to what types of pollen and spore samples one can collect as potential evidence because their tiny size enables them to become trapped on or in many different types of surfaces (Mildenhall et al. 2006). For instance, items sampled for pollen evidence include a tissue wipe of a tabletop, leather shoes and sandals, a shoelace from a boot, a piece of crumpled electrical wire from an explosive device, the vacuumed contents from pages in a book, dust vacuumed from the keyboard of a laptop computer, pollen trapped on tape used to wrap a package of illegal drugs, the lint trapped in the pockets of a pair of pants, and pollen and spores vacuumed from all sorts of clothing, backpacks, suitcases, rugs, and vehicles.

Samples used for forensic pollen evidence offer several major challenges during the extraction process. First, often there is very little material sample and thus the extraction procedures selected must be the right ones for the sample. Although some samples do not present much of a challenge, others do. Samples containing microscopic charcoal flecks or soot and samples collected from various types of oils and grease can present problems when trying to remove the pollen from those matrix materials. Other types of difficulties arise when examining samples vacuumed from polyester clothing or other synthetic materials, which contain many fibers and fiber fragments that resist most forms of chemical digestion or turn into partly digested globs that stick to pollen trapped in the sample. Second, because most samples cannot be replaced, great care must be taken not only to prevent contamination but also to prevent the loss of the pollen sample during the extraction process. Lab accidents must be avoided such as the accidental spilling or breaking of beakers or test tubes containing a sample, mixing of samples, using incorrect chemical reagents, or not keeping samples tightly sealed throughout each stage of processing. The primary objective in most cases is to remove as much debris in the samples as possible, thereby freeing the pollen grains for analysis and preventing remaining debris from obscuring part or all of the individual pollen grains during analysis. This means a forensic sample's matrix material will not be available later for other types of forensic testing. If planned studies of the DNA, isotopes, trace elements, hairs and fibers, or other sample contents are planned, they should be completed before pollen extraction occurs. In those types of situations, however, great care should be taken to avoid any pollen or spore contamination of the sample. There are a number of reliable extraction procedures that can be used to prepare forensic pollen samples for both light microscopy and for scanning electron microscopy analysis (Faegri and Iversen 1989). The choice on which procedures to use will be dictated by the type of sample and by the type of matrix materials in the sample.

Microscopy and Other Analytical Techniques

The primary analytical techniques used in forensic palynology is light microscopy because it is easy to use, and most of the identification keys and atlases (printed and on web sites) rely on LM images as guides to the identification of pollen and spore types. LM is also a fairly fast technique that, in most cases, is adequate for determining the ratio of each pollen or spore type in the total pollen and spore spectrum of a sample. Thus, it is often the combined identification of both the pollen and spore taxa and the abundance of each type that provides the critical data needed for analytical precision. The primary problem with LM is that the resolution capability is limited to about 0.3 μm even using the best optimal microscopic lenses available (Jones and Bryant 2007). That resolution ability typically limits the identification of pollen and spore in some groups to the family level (i.e., Cupressaceae, Poaceae, and Chenopodiaceae) and others to the generic level (i.e., *Pinus, Quercus, Artemisia, and Eucalyptus*). Unfortunately, when using LM analysis, it is rarely precise enough to identify most pollen types confidently to the species level. Therefore, when added precision is needed, the forensic palynologist must rely upon some other technique, such as the higher resolution ability provided by both SEM and TEM (Milne et al. 2005).

Some analysts have been asked to examine entire forensic pollen samples using only SEM in an effort to gain the added precision of pollen and spore identification to the species level. Analyzing samples using only SEM is time-cosuming, costly, requires access to expensive equipment, and assumes that the forensic palynologist has the expertise and familiarity with SEM usage to ensure accurate results. An added problem of conducting SEM analyses of forensic samples is that one must have access to a large databank of SEM images of all the potential species in a genus in order to ensure the correct identification, to the species level, of pollen grains found in samples. Although this can be done, and has been done (Jones and Bryant 2007), these techniques are rarely feasible based on the funding constraints facing most investigative agencies.

The search for new techniques that could increase the accuracy and identification of pollen and spores in forensic samples has focused on a number of analytical techniques often used with other types of forensic samples. One of these is the use of Raman spectroscopy.

Raman spectroscopy has gained praise for its use in forensics as a successful analytical tool. This praise is based on a number of advantages of Raman spectroscopy, which makes it ideal for applications in the field of forensics. First, the techniques used in Raman spectroscopy are rapidly changing as improvements broaden its potential uses. Second, Raman spectroscopy is a nondestructive technique meaning that samples do not have to be destroyed or changed in order to enable analysis. Third, Raman spectroscopy is quick because it does not require any special type of sample preparation prior to examination.

This means that forensic pollen samples can often be examined first using Raman spectroscopy rather than having to wait until other types of analyses have been completed, each of which might alter or contaminate the original sample. Fourth, only minimal amounts of material are required for a complete analysis, which means that tiny amounts of a sample, containing pollen and spore evidence, can be effectively examined. Fifth, Raman spectroscopy is highly sensitive to slight differences in the molecular and chemical composition of materials, which enables it to detect very small differences in samples. That type of high-resolution potential is important where minor differences are essential to distinguish one type of pollen or spore from all others (Ivleva et al. 2005).

A significant drawback of Raman spectroscopy for use in forensic pollen studies, however, is that the identification of pollen and spores down to the species level requires the construction of high-resolution spectral maps for each taxon. To construct such spectral maps for just 1 pollen or spore species requires an estimated 360–720 h of time gathering and storing the needed data. Therefore, the amount of time and cost expenditures required doing this type of analysis on samples containing a variety of pollen and spore species is usually considered unreasonable. An additional problem is that conducting and interpreting the signatures of Raman spectroscopy require a special expertise, which most palynologists may not have.

Although a few studies have shown that successful spectral signatures can be used to identify specific types of pollen using Raman spectroscopy (Manoharan et al. 1991), the use of this technology in forensic palynology is hindered by a number of problems. First, very few pollen species have spectral signatures that could be used for comparative purposes. Even if more pollen spectral signatures were desired, the very large investment of time and effort to obtain these signatures for each new species would be impractical. Second, a major problem with Raman spectroscopy is that for any given forensic sample, it might provide identification for some of the pollen and spores, but it could only offer statistical ubiquity rather than quantitative percentages of each taxon in the entire forensic sample. Finally, in some cases, it appears that Raman spectroscopy cannot identify some pollen types with certainty. After mounting fresh *Celtis* (hackberry) and fresh *Cannabis* (hemp, marijuana) pollen, it was noticed that some of the pollen grains in each group were folded, deflated, broken, and overlapped with other grains or were partly covered by pieces of debris. When setting the laser beam on specific targets within each group, the resulting Raman data for each group and between both pollen types was not conclusive enough to provide the certainty of identity that would be needed for accurate identification of those pollen types to the genus level.

Another technology that can aid in the identification of pollen grains in forensic applications is Fourier transform infrared spectroscopy (FTIR). This technique uses infrared rays (IR) to stimulate different chemical bonds in pollen grains that vibrate at certain known frequencies when excited. When the surface of a pollen grain is bombarded with an IR beam, the transmittance and reflectance of the IR beam occurs at different frequencies, which is measured and translated into an IR absorption plot. By matching the IR plot of an unknown pollen grain with IR signatures of known pollen taxa, the identity of the unknown type can be confirmed. The advantage of using FTIR technology is that it does not require complex sample preparations.

When targeting the surface of a pollen grain, an average depth of 2–3 μm is excited by the IR beam. Experiments using FTIR technology for pollen identifications note that individual pollen grains within a single genus and species will produce similar, yet sometimes slightly different, IR absorption plots depending on various factors such as slight chemical variations in the composition of the pollen wall, and the level of maturity of the pollen grain. These variations within a single pollen or spore taxon can be recognized by developing a FTIR database from multigrain spectra for each known pollen and spore species. Similar to Raman spectroscopy, there are several drawbacks associated with using FTIR technology for pollen identifications in forensics. The number of IR absorption plots for known pollen taxa in the current FTIR databases is limited, thereby decreasing the chance of successfully identifying individual pollen grains. Also, it has yet to be determined if pollen grains in various stages of degradation continue to send the same, or different IF absorption signatures, meaning it is unknown if degraded pollen grains will give the same signature as that of a pristine grain of the same species. Finally, there remains the problem of quantification of pollen and spore taxa in terms of the ratios of each type in a forensic sample to all others types present. Similar to Raman, FTIR can determine the "presence/absence" of individual taxa, but it cannot determine how many individual grains are present in a sample (Gottardina et al. 2007).

DNA analysis, a common technique used by many forensic laboratories, could potentially aid in pollen and spore identification. Recent evidence has demonstrated that a DNA signature can be recovered from a single pollen grain. DNA in pollen grains, however, can only be found in the cytoplasm inside the walls of a pollen grain. Once the cytoplasm degrades or is lost through rapid decay, the DNA signature is lost. Since this technique, as related to individual pollen grains, is still fairly new, the recovery methods, techniques to isolate individual pollen grains from the surrounding matrix materials, and adequate polymerase chain reaction (PCR) replication and identification database constructed from existing pollen DNA are still in the testing phase. Another problem concerns the type of pollen evidence that is used for forensic interpretations. Since DNA can only be recovered from viable pollen, using DNA signatures one could not identify nonviable pollen or degraded pollen in a sample (Zhou et al. 2007). Therefore, this type of data recovery would provide only limited results in terms of the total pollen present in a forensic sample. Another potential problem with using the DNA of individual pollen grains is that even with an adequate databank of information, it still may not be possible to identify individual pollen grains in a sample. That potential exists because certain species of plants can easily hybridize with other species of the same genus producing an F1 generation. Studies of the DNA signatures from other types of plant material, such as leaves and wood, reveal that the DNA of F1 plants is not identical to either of the parent plant sources (Schaal et al. 1998).

Unfortunately, DNA, Raman spectrometry, and FTIR analysis methods will provide only ubiquity types of data related to the pollen types in a sample. Instead, what are essential elements in forensic pollen studies are quantitative counts of the relative frequency of each pollen taxon. Finally, aside from the other already mentioned limitations created by the data obtained from the DNA of individual pollen grains, the ultimate cost and time needed for this type of study make it unsuitable for forensic pollen analyses.

Similar to DNA analysis, stable isotope analysis is another important analytical technique that is gaining added use in forensic science. The advantage of this principle is that isotope ratios in organic and inorganic materials retain a record, in their molecules, of conditions that existed when that material was created (West et al. 2006). Using that principle, studies reveal that pollen grains contain recoverable isotopic signatures that can be recovered and mapped. During the last decade, scientists have attempted to recover the isotope signatures recovered from individual fossil pollen grains and then use those data to provide information about the environmental and climatic variations that existed where the pollen grain was generated. Further studies have revealed that stable isotopes of individual pollen grains from different plant families and different genera, and in some cases even individual plant species, generate unique isotope signatures. However, the major problem those scientists have encountered when trying to do isotope studies of individual pollen grains is how to isolate and concentrate the pollen grains in a forensic sample and remove them from the various types of surrounding matrix material. Thus far, the various chemical treatments designed to concentrate pollen and remove non-pollen debris in forensic samples alter the molecular composition of the pollen wall (Loader and Hemming 2000, 2001).

Similar to Raman spectroscopy, DNA, and FTIR, this type of study using isotopes would provide only ubiquity data. Finally, isotope studies of individual pollen grains would be very costly, time-consuming, and extremely complicated, thereby limiting its effectiveness as an analytical technique in forensic palynology (Loader and Hemming 2004).

Case Studies

Forensic palynology has been used to assist in the solving of a variety of criminal cases. In one instance, a suspect with a large quantity of cannabis resin was arrested at a port in New Zealand. Authorities believed that the cannabis was grown outside of New Zealand and they wanted to test their hypothesis. Several subsamples of the cannabis were tested for pollen and spores, which were then analyzed to determine whether or not the cannabis came from New Zealand. All subsamples yielded a similar pollen and spore spectra that included plant sources that are not indigenous to New Zealand. Based on the pollen and spore evidence, authorities were able to confirm that the cannabis was imported from Asia and was not from a locally grown crop (Mildenhall 1990).

Forensic palynology can also help to resolve murder cases. In 1999, the body of a man was discovered in a mountainous region north of Wellington, New Zealand. An eyewitness claimed to have seen another person with the victim in the area where the murder took place. Based on evidence, a suspect was arrested and a search revealed that the suspect had the distinctive type of clothing, which has been described by the eyewitness. To verify the eyewitness' claim, the suspect's clothing, camping gear, and other accessories were analyzed for their pollen and spore content. The recovered pollen and spore assemblage from the suspect's clothing and backpack gear closely matched that of a mountainous region where the victim had been found. Although the pollen and spore evidence did not confirm the guilt of the suspect, it did show that the suspect had been in the immediate area where the crime had occurred and where he had been seen by the eyewitness (Mildenhall 2004).

In another case, a murder victim was found in an excavated depression wrapped in a bedcover in West Yorkshire, England. To determine whether or not the victim was murdered where the body was discovered, investigators sampled the bedcover and the victim's hair for pollen. Upon analysis, the recovered pollen from the bedcover and victim suggested that the murder had taken place in a household garden and not where the body was discovered. Wood ash was also recovered from the victim's hair, suggesting that the victim had been near a fire at the time of death. About a year later, a suspect was identified, and the recovered pollen spectra that had been recovered from the murdered victim and from the bedcover were compared with comparison soil samples collected from the suspect's garden. Multiple samples were taken from different locations around the suspect's garden and property. Not only did the pollen recovered from the bedcover and victim's clothing match that of the suspect's garden, but it also contained wood ash that matched an area next to a bonfire that had been constructed by the suspect. The pollen evidence did not confirm the suspect was the murderer, but it was important because it linked the victim and the bedcover to the suspect's garden and property (Wiltshire 2006b).

Summary

Forensic palynology is a predominantly underutilized field. Only recently have some governments and law enforcement agencies around the world begun to explore and appreciate the utility of pollen and spores in forensic cases. Among these, the United Kingdom and New Zealand still lead in terms of the number of cases and applications they have explored using pollen and spore evidence. The potential uses of forensic palynology are many and have already been used effectively in instances involving homicide, terrorism, genocide, bombings, forgery, theft, rape, arson, counterfeiting, manufacturing and distribution of illegal drugs, assault, cases of hit and run, poaching, and identity theft. Pollen and spores are microscopic in size and are often produced in large numbers. They can be found in almost any environment and can become attached to almost any type of surface; these are reasons why they have become useful as trace evidence of places where items originated or were used. Although forensic palynology has played an important role in helping to convict criminals and in some cases support the innocence of some suspect, most countries have been slow in adopting this technique. This hesitation often comes from the discipline's lack of notoriety and often an unfamiliarity with how the use of pollen and spores can contribute to as forensic evidence.

More law enforcement agencies should consider using pollen and spores in forensic cases and come to understand the importance of how to collect samples without contaminating them and whom to depend upon for help with the analyses. Pollen evidence has a bright future as one of the types of materials that can be useful in determining the geolocation of some person or object. Currently, there are limited numbers of scientists trained in the field of forensic palynology, mainly because there has not been a demand for their services. This, however, could change if the need to conduct these types of analyses was to broaden. Unfortunately, the training of these individuals cannot be accomplished overnight and thus when the time comes for new forensic palynologists, they will need to be able to meet the needs of law enforcement and others requiring their services.

Related Entries

- ▶ Forensic Anthropology
- ▶ Forensic Environmental Evidence
- ▶ Forensic Palynology

Recommended Reading and References

Bryant V, Holloway R, Jones J, Carlson D (1994) Pollen preservation in alkaline soils of the American southwest. In: Traverse A (ed) Sedimentation of organic particles. Cambridge University Press, London, pp 47–58

Bryant V, Jones G (2006) Forensic palynology: current status of a rarely used technique in the United States of America. Forensic Sci Int 163:183–197

Bryant VM, Mildenhall DC (1998) Forensic palynology: a new way to catch crooks. In: Bryant VM, Wrenn JH (eds) New developments in palynomorph sampling, extraction and analysis, American association of stratigraphic palynologists, contributions series # 33. American Association of Stratigraphic Palynologists, Foundation, Dallas, pp 145–155

Erdtman G (1969) Handbook of palynology. Hafner Publishing, New York, p 486

Faegri K, Iversen J (1989) Textbook of pollen analysis, 4th edn. Blackburn Press, Caldwell, p 328

Frei M (1982) Nine years of palinological studies on the shroud. Shroud Spectrum Int 3:3–7

Gottardina E, Rossi S, Cristofolini F, Benedetti L (2007) Use of Fourier transform infrared (FT-IR) spectroscopy as a tool for pollen identification. Aerobiologia 23:211–219

Holloway RG (1989) Experimental mechanical pollen degradation and its application to Quaternary age deposits. Tex J Sci 41:131–145

Ivleva N, Niessner R, Panne U (2005) Characterization and discrimination of pollen by Raman microscopy. Annu Bioanalytical Chem 381:261–267

Jones G, Bryant V (2007) A comparison of pollen counts: light versus scanning electron microscopy. Grana 46:20–33

Kadowaki K, Leschen RAB, Beggs JR (2010) Periodicity of spore release from individual *Ganoderma* fruiting bodies in a natural forest. Austral Mycologist 29:17–23

Loader NJ, Hemming DL (2000) Preparation of pollen for stable carbon isotope analyses. Chemical Geology 165:339–344

Loader NJ, Hemming DL (2001) Spatial variation in pollen δ^{13}C correlates with temperature and seasonal development timing. The Holocene 11:587–592

Loader N, Hemming D (2004) The stable isotope analysis of pollen as an indicator of terrestrial palaeoenvironmental change: a review of progress and recent developments. Quat Sci Rev 23:893–900

Manoharan R, Ghiamati E, Britton K, Nelson W, Sperry J (1991) Resonance Raman spectra of aqueous pollen suspensions with 222.5–242.4-nm pulsed laser excitation. Appl Spectrosc 45:307–311

Mildenhall DC (1990) Forensic palynology in New Zealand. Rev Palaeobot Palyno 64:227–234

Mildenhall DC (1992) Pollen plays part in crime-busting. Foren Focus 11:1–4

Mildenhall DC (2004) An example of the use of forensic palynology in assessing an alibi. J Forensic Sci 49(2):1–5

Mildenhall DC (2006) An unusual appearance of a common pollen type indicates the scene of the crime. Forensic Sci Int 163:236–240

Mildenhall DC, Wiltshire PE, Bryant VM (2006) Forensic palynology: why do it and how it works. Forensic Sci Int 163:163–172

Milne LA (2005) A grain of truth: how pollen brought a murderer to justice. Reed New Holland, Sydney, p 175

Milne LA, Bryant VM, Mildenhall DC (2005) Forensic palynology. In: Coyle H (ed) Forensic botany: principles and applications to criminal casework. CRC Press, Boca Raton, pp 217–252

Schaal B, Hayworth D, Olsen L, Rauscher J, Smith W (1998) Phylogeographic studies in plants: problems and prospects. Mol Ecol 7:465–474

Tauber H (1965) Differential pollen dispersion and interpretation of pollen diagrams. Danm Geol Unders IRR 89:1–69

Traverse A (2007) Paleopalynology, 2nd edn. Springer, Dordrecht

West JB, Bowen GJ, Cerling TE, Ehleringer JR (2006) Stable isotopes as one of nature's ecological recorders. Trends Ecol Evol 21:408–414

Wiltshire PEJ (1993) Environmental profiling and forensic palynology: background and potential value to the criminal investigator. Handbook for the national crime and operation faculty with the British association for human identification, pp 14–21

Wiltshire PE (2006a) Consideration of some taphonomic variables of relevance to forensic palynological investigations in the United Kingdom. Forensic Sci Int 163:173–182

Wiltshire PE (2006b) Hair as a source of forensic evidence in murder investigations. Forensic Sci Int 163:241–248

Zavada M, McGraw S, Miller M (2007) The role of clothing fabrics as passive pollen collectors in the northeastern United States. Grana 46:285–291

Zhou L, Pei K, Zhou B, Ma K (2007) A molecular approach to species identification of Chenopodiaceae pollen grains in surface soil. Am J Bot 94:477–481

Forensic Palynology (The Study of Pollen Grains)

▶ Forensic Palynology

Forensic Science

▶ Identification Technologies in Policing and Proof

Forensic Science and Bias

▶ Cognitive Forensics: Human Cognition, Contextual Information, and Bias

Forensic Science and Criminal Inquiry

▶ Forensic Science and Criminal Investigation

Forensic Science and Criminal Investigation

Olivier Delémont[1], Eric Lock[2] and Olivier Ribaux[1]
[1]Ecole des Sciences Criminelles, University of Lausanne, Lausanne, Switzerland
[2]Brigade de Police Technique et Scientifique, Police cantonale, Geneva, Switzerland

Synonyms

Forensic Science and Criminal Inquiry

Overview

Often narrowly depicted as the contribution of science to the production of evidence for courts, forensic science fundamentally relies on the extraction of information from physical traces that are remnants of a past action. By searching, scrutinizing, and testing physical traces, forensic scientists gain pieces of information that can bring significant contribution to different dimensions or phases of criminal inquiry.

The practice of this discipline alone does not really make any sense, as it is intimately dependent of the context and the expectations of the criminal inquiry. The decision to attend a scene as well as the process of physical traces detection and collection do not solely respond to technical considerations. They are influenced by different dimensions of knowledge available (such as the physical, immediate, and criminal environments) as well as by policing issues. Considering the distinct phases of a criminal inquiry reveals the manifold functions of the use of physical traces. The exploitation of physical traces provides intelligence or leads that support police investigation, participating for instance to the search for the perpetrator. It plays a crucial role in the reconstruction of the sequence of actions and the clarification of events by putting the information gained from the traces to context. Ultimately, the structuring and thorough interpretation of this information may also lead to the constitution of evidence to be presented and evaluated in court.

Beyond its use within the investigation of a specific case, physical traces hold a potential, still underexploited, to reveal or reinforce connections between offenses, unveiling the extent of a criminal activity or phenomenon. Moreover, the value of physical traces as vectors of information continuously bursts the context of criminal inquiry, expanding its contribution to issues related to security or criminology.

This perspective changes the traditional view on what is expected from forensic science. Requirements and needs are intelligence-led or the result of a trade-off between many parameters imposed by the actual inquiry, rather than based on the routine use of specialized techniques and methods essentially dedicated to the production of evidence for court. This dramatically impacts the way quality is perceived as forensic operations at the different stages of the judicial inquiry achieve different objectives following intelligence, investigative or evaluative processes.

As an event occurs and is recognized as a potential dispute or even an offense, legal systems, whatever their differences and specificities, usually converge to a response in the form of an inquiry. The formal terms as well as the scope of this inquiry may be defined differently according to legal systems, but it generally remains the duty of police forces proceeding under the responsibility or supervision of judicial authorities. The inquiry will have many aspects and examine the event from several angles. One of them whose importance has been growing in the last decades is the use of natural and computer sciences in order to gain information, sometimes also evidence, from physical traces produced during the course of the event. This aspect of the inquiry falls under the prerogative of forensic police units that use the principles and methods belonging to criminalistics in order to detect, recover, record, collect, decipher, analyze, evaluate, and present relevant physical traces (see entry *Forensic science culture*). For several years, this discipline of the police activity has been attracting lots of attention by the media. Movies, reports, and TV shows provide a very compelling, yet skewed, aspect of the use of forensic science, strengthening the illusion of the primacy of science and technology on crime solving and having deleterious effects on several levels of the judicial system (Durnal 2010; Brodeur 2010). On another hand, the utility and efficiency of forensic science in terms of its impact on judicial outcomes or crime solving is questioned (Baskin and Sommers 2010; Brodeur 2010). For instance, the effectiveness of DNA testing, often depicted as the most powerful and reliable physical evidence, has turned to be undisputable for high-volume crimes such as burglary, but much lower for more serious crimes, often violent crimes (Wilson et al. 2010). Besides these two antinomic views of the utility of forensic science to the judicial inquiry, its true contribution is not straightforward and unique, but may occur on different dimensions or levels in the context of the judicial response following an event as well as for other law enforcement activities.

Physical Trace as Fundamental Object of Forensic Science

Forensic science focuses its attention on physical traces that are produced during the course of an event, generally a litigious action. As first emphasized in the first half of the twentieth century by Edmond Locard, a French pioneer of criminalistics (see entry *Forensic science culture*), any criminal activity, considering the intensity it encompasses, implies the exchange of material and thus generates marks or what we call traces. The perpetrator of the action, more often unconsciously, leaves traces on the scene of the action (on the physical environment as on the victims), and he/she takes traces from the scene of the action. The intensity of the action influences the generation of physical traces as more violent or more severe events are prone to end up in massive exchange of traces. Traces then constitute the remnants of this action, its subsisting physical symptoms. By applying the scientific principles and methods of criminalistics, the forensic scientist aims at detecting physical traces, discriminating traces that are consecutive of the action from traces that are not relevant, and exploiting traces to gain information on the source – a person or an object – and/or the sequence of events that produced them. This approach is extremely challenging as the relevant traces may be dispersed on several, also often not known, locations and confused with "background" and/or "contamination" traces, i.e., traces produced by events anterior, respectively posterior, to the relevant action.

Physical traces are the vectors of information about a past action and, therefore, the fundamental element of forensic science scrutiny. But the information on the past carried by physical traces is more often than not fragmentary and does not enable to perfectly depict the sequence of events. Moreover, this information is not directly decipherable. Traces are symptoms of a specific event whose causes are not known and cannot be repeated, but has to be inferred from these symptoms. The detection and examination of physical traces lead to the construction of hypotheses about their causes (reasoning called "abduction" in Peirce's semiotic). With the information that is thus gained, physical traces become signs that provide indications, with variable degrees of reliability and certainty, that serve to the process of reconstruction of the relevant events. These clues provide information that feeds different processes of the criminal justice system and even beyond, e.g., gathering intelligence for the police inquiry, supporting the structuring of evidence for court decisions, or revealing the structure of repetitive criminal activities. Forensic scientists exploit physical traces to gain pieces of information in an approach of systematic questioning, trying to contribute to the fundamental questions the justice faces: who, what, where, when, why, and how. It is obvious that physical traces alone cannot answer comprehensively these questions. The part of the story that scientific clues help to tell is completed by the use of "traditional" human-based information gathered during the police inquiry. Physical traces may corroborate or invalidate hypotheses generated by police investigation; they are also able to bring out new hypotheses leading to unsuspected investigative paths. In the very first steps of the judicial inquiry, generally devoted to police forces, physical traces are used in combination with other clues, such as testimonies for instance, to generate intelligence about an action under investigation (Barclay 2009). A key aspect of the investigation is to implement a collaborative approach between scientists and investigators in order to keep a global view on the case and to develop, as comprehensively as possible, realistic alternative hypotheses, devise an experimental approach to test hypotheses, and search for the most relevant use of specialists and instruments. This is why the forensic generalist, as opposed to the forensic specialist, has a pivotal role during the investigation. It avoids favoring or ruling out premature conjectures. However tempting is the tendency to follow a lead that is favored, such approach must be banned as it encompasses the major risk of neglecting some traces and wrongly reinforcing the confidence in the selected hypothesis (confirmation or contextual bias) (see entry *Cognitive Forensics: Human Cognition, Contextual Information and Bias*).

Physical Traces Search and Collection: Crime Scene Investigation

The existence of physical traces is linked to the remaining presence of substrates or traits in the location and at the time of their search. As noted previously, traces are unconsciously exchanged during litigious actions and constitute remnants of these actions after their completion. But the use of physical traces in a judicial inquiry entails first that they are searched, seen, recognized, documented, and perhaps, collected. These different steps are undertaken during a so-called crime scene investigation that is not only performed when a crime is committed, but whenever a confirmed or presumed litigious action has occurred. The scope of the investigation encompasses then every location where traces relevant to the action under consideration are expected to remain: on the geographic sites where the action took place, at the place of a suspect's residence, in a vehicle used by a suspect, etc. While it is often essentially presented in technical terms in books and manuals, the problematic of the search and collection of traces on a crime scene calls for much wider competencies than solely technical skills (Ribaux et al. 2010a).

Different factors have an influence both on the decision to attend a scene, or not, once a potential litigious action is known (crime scene attendance) and on the manner the scene is processed and searched for physical traces (crime scene processing). First of all, the strategic policies of a law enforcement organization frame the decision-making process to attend a crime scene. Generally, the importance of a case, in reference to its criminal law qualification, and the priority given by organizations for specific security or policing issues rule the strategies and the attribution of resources. For example, cases assessed as important, such as homicides, will often result in a massive deployment of human, time, and material resources to the crime scene investigation, whereas less serious offenses, volume crimes, for instance, will rarely involve systematic crime scene attendance or intensive investigation. Beyond these strategic policies, which mainly rely on legal, managerial, and organizational issues, several knowledge dimensions contribute to the decision to attend a crime scene and to shape the investigative approach. A general knowledge of physical traces' taxonomy as well as of their potential exploitation in the different information processes of a judicial inquiry is a prerequisite to suitable crime scene investigation. The detection of physical traces on a crime scene represents a difficult yet decisive task that requires a solid knowledge of the multiple forms that traces can have. Crime scene investigation is never a passive process. Traces with blurred morphology, such as shoemarks or earmarks, or latent traces, such as certain fingerprints or DNA deposits, cannot be detected by someone who does not know what they look like or is not prepared to anticipate their presence. As emphasized by an aphorism well known to every forensic scientist, generally associated to Alphonse Bertillon who had this maxim displayed in his teaching room, "the eye only sees what it's looking at and it looks only at what is already in the mind." This quote also warns crime scene examiners to go beyond strict procedures and routine by keeping their mind open when processing each specific case. Thus, more than an awareness about physical traces, the process of searching and detecting traces relies on a structured education and training and based on the purchase of a forensic culture (Margot 2011). The preservation, recording, and collection of traces require then technical skills as well as knowledge of the persistence and recovery possibilities of the different types of traces in the given physical environment that is investigated. This specific knowledge is also often a significant parameter that is considered in the decision-making process of crime scene attendance: for instance, for burglary cases, if the chance of recovering relevant physical traces is weak on the basis of the available information, the decision not to investigate the scene will often be made.

Analysis of the immediate environment where the action occurred provides another kind of knowledge that influences the processing of crime scene. Sometimes epitomized as a determination of the modus operandi of the criminals, this analytical process relies on the attempt to reconstruct the sequence of activities in order to orientate the search for traces. As a starting point, the analysis of the immediate situation requires the integration of information about the environment and circumstances of the action. Taking into account this context, the processing of the crime scene can start progressing in the search for traces following an iterative hypothetico-deductive approach. Every piece of information or recovered trace leads to the generation of hypotheses regarding the events. These hypotheses then orientate the search and collection of new traces that contribute to invalidate, reinforce, or discriminate the hypotheses, maybe also suggesting new alternatives. Finally, the decision to undertake a crime scene investigation and the reasoning applied to detect and collect traces are definitely influenced by the knowledge of the general criminal environment. An analysis of the criminal activities informs about the current structure of the criminality and active criminal phenomena. Some litigious actions may belong to the general activity of a same perpetrator or a group of perpetrators, especially considering volume crime offenses that are highly repetitive. Therefore, knowledge about this activity and a priori detection that a given event may belong to a current criminal phenomenon provide valuable intelligence feeding both the decision to attend the scene and the management of scene search. As such, the practice of forensic science through crime scene investigation becomes a decisive instrument of intelligence-led policing (Ribaux et al. 2010b).

From Clue to Evidence: The Manifold Dimensions of Forensic Science' Contribution to the Criminal Inquiry

As an introductory overview, the tension between the omnipotent picture of forensic science constructed with science and technology and the dispute of its efficiency by empirical research was emphasized ► Forensic Science Effectiveness. A similar tension appears when considering the use of forensic science and its utility in the frame of the judicial inquiry. In the first stages of the judicial process, essentially conducted by police investigative units, the use of physical traces is continuously expanding. The types of traces that are searched, collected, and exploited increase, and the range of situations for the investigation of which traces are used extends. Some knowledge indicates that the major contribution of physical traces relies in refinement, checking, and preparation for the trial (Kind 1994), while other researches attribute them little value in influencing decisions during the later stage of the process (Baskin and Sommers 2010). In some aspects, the current contribution of forensic science can lead to harmful and counterproductive consequences as it is presented as a source of miscarriage of justice (National Research Council 2009). Beyond this apparent antagonism, the valuable benefit of forensic science appears in its potential to bring decisive contribution to the multiple dimensions of the judicial process. The criminal inquiry is not a continuous process but a succession of phases, or *times*, that are characterized by distinct aims, distinct reasoning, and distinct ways of using information. Being inspired by the decomposition of the criminal investigation in chapters suggested by Kind (1994) and by the taxonomy proposed by Brodeur (2010), it is possible to assess the multidimensional contribution of forensic science.

Physical Traces Serving the Investigations

The first phase of the criminal investigation is speculative and consists of searching. An action occurred which entails the starting of an investigation handled by police forces. Its main purpose is to clarify the events regarding possible further criminal prosecution (is this a crime?) and to identify the persons involved. In this perspective, the early investigation looks like a quest for information that serves as raw material to support initial generation of hypotheses about the events and the perpetrator. What is expected from forensic science at this time is thus dramatically different from services provided by traditional laboratories driven by court-oriented quality processes: rapid indications about the nature of a trace, even very uncertain, may orient decisively investigations.

The attendance of crime scene enables the detection of physical traces that, as vectors of information about past events, are exploited to gain intelligence. If the traces detected are relevant to explain the activity that is considered, they become clues that can assist in the clarification of the events. In this first phase of the investigation, the detection and collection of physical traces tend to be directed by the concern of the inquiry. The main, if not exclusive, issue to be answered at that time is often the question of "who?" with its two main subdivisions: the identification of the victim (e.g., when a dead body is discovered or when the victim is not known) and the search for the perpetrator. In certain cases, besides the question of "who?" the determination of the nature of the action ("what?") is also a main concern. For instance, the very first stages of the investigation of a dead person focus essentially on the issue of the cause of death (e.g., can the action of a third party be excluded?), the identity of the deceased, and the circumstances of the action.

As instruments serving the police inquiry, the detection and collection of physical traces are largely influenced by its concern. The approach of crime scene processing tends to favor the search for traces that could provide indications on the perpetrator or another participant of the action. This is why the detection of fingerprints and traces containing DNA is often promoted. Indeed, there is a direct association between those physical traces and the individual at the source of the trace. Thus, fingerprints and traces containing DNA are likely to provide information on the searched identity of a person. But the trace does not self-reveal its sources. Let's take the example of a trace left by the finger of an individual. The detection of this trace by applying the suitable enhancement technique and its identification as being produced by a finger (or a portion of friction skin showing papillary ridges) are not sufficient to provide the identity of the individual at the source of the trace. In order to gain this information, it must be compared with reference marks (e.g., ten print cards produced by a known person), and the results of this comparison must be evaluated. This mechanism operates thanks to the availability of databases holding ever-increasing amounts of reference fingerprints and DNA profiles. Obviously, the efficiency of this process relies firmly on how this reference material has been selected. This explains why major efforts are made to collate reference materials within databases and to increase exchanges between existing databases. In parallel, research activities and publications tend to focus on the potential for individualization of such traces, aiming at quantifying the output of the comparison process between a trace and a reference material. This does not however mean that without reference material, such traces are worthless.

Beyond their potential for identification or individualization traces provide above all elimination criteria. In a criminal inquiry process led by the progressive size reduction of a suspect population, the elimination potency of physical traces is a precious contribution (Barclay 2009; Kind 1994). Besides, it is worth to mention that the very first use of DNA in a criminal case in 1986 resulted in the exclusion of a suspect that the inquiry designated as guilty, the confirmation that the case was linked with a previous crime, and then a DNA-based manhunt founded on a progressive reduction of the population of potential suspects (Wong et al. 1987). Technically, the existence of DNA database opens also the perspective for alternative investigative possibilities, maybe more speculative and risky than the conservative quest for individualization. Partial profiles could thus be further exploited, giving a list of potential suspects than would then be reduced by traditional investigation. Familial searches in DNA databases are another example of investigative strategy that already proved to be decisive in some major inquiries, but which however can raise ethical issues that still need to be clarified.

Obviously, information obtained from physical traces detected and collected can also be misleading. As traces are remnants of the past history, the locations subject to scene investigation encompass traces related to the action of interest as well as nonrelevant traces produced either before or after the action. If rapid discovery of the action and efficient scene preservation may significantly limit the addition of "contamination" traces, there is no possible control or influence on vestiges from the history of the environment preceding the action of interest. Thus, on all physical traces that are detected and collected from the crime scene, some of them will inevitably convey pieces of information unrelated with the inquiry. Considered solely, these pieces of information materialize a risk of orientating on misleading paths or even lead to false exclusion. But considered with all other information gathered in the inquiry and after verification, the hypotheses they suggest can be disregarded. This typically occurs when characteristics of a trace do not match with a reference. In this situation, the general attitude would be to exclude the person of being involved in the case. However, this entirely depends on the circumstances of the case and how another origin of the trace is possible.

In the early times of the inquiry, the search and use of physical traces provide information that is neither perfect nor complete but which can be incorporated into the corpus of other pieces of information gathered from "traditional" inquiry. How the investigation integrates these *mute witnesses*, is able to generate relevant hypotheses as completely as possible, and devises experimental tests are central issues to properly use the whole information. Forensic scientists cannot be excluded from this process because they know the informational potential of physical traces and how to identify specialized fields of expertise needed. While this key dimension of the role of forensic science, as a corpus of scientific methods employed to gather intelligence from physical traces, is widely applied by forensic staff and crime scene units in police forces, it is often misjudged by scientific laboratories and the research community that are framed by a paradigm oriented to court concerns. Conversely, they create distance with the investigation in order to keep independence that is supposed to promote their impartiality in the perspective of the trial process.

Reconstruction of the Events: The Structuring of Evidence

As previously emphasized, the criminal investigation process encompasses multiple dimensions. A second phase that usually follows the searching consists in structuring the evidence against a suspect in order to form the decision to prosecute or not (Kind 1994; Brodeur 2010). At that time of the investigation, which begins with a fuzzy delimitation being the belief that an identified individual is the perpetrator of the action, this individual becomes the main focus of the inquiry and the reasoning scheme is significantly altered. All pieces of information accumulated during the first phase of the investigation are scrutinized under the assumption that the identified individual is the perpetrator, revealing elements that corroborate this assumption as well as those that invalidate it. There is a conspicuous risk, associated to this reasoning scheme, to fall in a cyclic argumentation consolidating the assumption of implication of the identified individual (Kind 1994).

This second phase of the investigation corresponds also to a shift of the main concerns of the investigation. As the inquiry tends to be focused on the identification of the perpetrator (and/or his accomplices) in the first phase, all the other issues of the systematic questioning (where, when, why, how, and what, if not already carefully examined) are considered in a process of reconstruction of the action. All pieces of information of the inquiry are sorted and integrated together reconstructing as precisely as possible the sequence of events and their degree of certainty.

This approach is often favorable to unfold new locations where physical traces can be looked for, extending the concept of "crime scene." Materializing the idea of exchange of materials and signs emphasized by Locard, forensic investigation is undertaken in the direct environment of the individual believed to be the perpetrator taking into account the context of the case: at his home, in his vehicles, on his clothes, his shoes, his tools, under his fingernails, etc. In practice, this search is generally not as speculative and open as the one undertaken during the first time of the inquiry. More often than not, it is directed towards the detection of physical traces related to the action or removed by the perpetrator. Such focussed investigation can be very successful if the individual suspected turns to really be the offender. However, it tends to neglect traces that could invalidate the putative link of the individual with the action or to underestimate the value of the absence of traces. Yet the fine reconstruction of the event or the extension of the concept of "crime scene" to new locations can reveal new hypotheses, new paths to investigate. The investigating process is not – or at least should not be – as directed as it seems but remains an iterative hypothetico-deductive process, open to hypotheses' generation and assessment.

The fine reconstruction of the action implies to consider all the available information within the context and situation of the case. It requires also to distinguish different dimensions of information that physical traces can convey (Cook et al. 1998). While the analytical process of the traces often aims at identifying and determining (if possible) their *source*, the reconstruction of the events entails the combination of the different traces as well as the study of their position. The information relating to the source of the traces is then considered within the context of the specific case, enabling to gain hypotheses or clues on the *activity*, i.e., the sequence of events that lead to the production of the set of traces.

The Provision of Evidence to Court

Once the whole inquiry conducts the investigators (or the judicial authority supervising them) to consider they have sufficient information to demonstrate that the perpetrator is identified and the sequence of the events can be reconstructed, they have to prepare a corpus of proofs from the collection of information. Physical traces can also significantly contribute to this third phase of the inquiry, the preparation of evidence to the court. Since they are remnants of the past whose availability in the present is not contingent upon the memory nor the interpretation of an individual, unlike testimonies for instance, physical traces hold a good potential to have the results of their analysis presented as evidence to a court. This requires interpreting the value of the results measured from the traces, assessing their level of uncertainty, and evaluating what this information brings to support the decision-making process of the judicial dispute. In this third time of the inquiry, only a small amount of traces are retained and presented as evidence, assuming they are informative and considering a specified and finite set of propositions. Under its simplest expression, the evaluation of the evidence takes the form of a balance between the probabilities to observe the evidence given two alternative propositions (usually referred as the proposition of the prosecution and the proposition of the defense). This "likelihood ratio" for the evaluation of forensic evidence (issued from the exploitation of traces) is expressed through a Bayesian framework that models how a new information (the physical trace) logically changes a priori beliefs on propositions (see entry *Probabilistic inference in forensic science*). This ratio indicates thus the strength of the evidence.

The use of traces as pieces of intelligence was presented as a collaborative process where some forensic scientists fully participate to the development of hypotheses. This contrasts with the evaluation scheme of evidence oriented to court, which restricts the forensic scientist to provide an opinion only by using absolutely necessary circumstantial information. While forensic scientists assess the strength of the evidence strictly in the eyes of the proposed assumptions, it is left to the judge's duty to integrate this assessment within the context of the case. Due to this fundamental difference, the practice of forensic science has been sometimes depicted as a confrontation of models: the first one promoting the use of traces to gather intelligence for the inquiry and the second centered on the preparation of evidence for the court. In fact, both of these paradigms adopt a narrow, and thus incomplete, view on the dimensions of information that can be generated by the exploitation of traces.

The Unifying Paradigm

But considering physical traces in a semiotic perspective reveals their intrinsic potential and allows to recognize their contribution to the production of information in a unifying paradigm embracing the two models that tend to confront each other. Instead of focussing on the finality of the use of information that can be extracted from traces (intelligence vs. evidence), it is much more relevant to place physical traces in the center of the model, recognizing them as vector of information about an action that belongs to the past. This also brings us to acknowledge the large scope of information traces can convey and to accept that they can be used for different purposes, feeding different reasoning processes: sometimes sustaining speculative and profuse argumentation and sometimes being the object of rigorous assessment based on strict interpretation and evaluation of uncertainties.

From the crime scene to the court, the role of the forensic scientist becomes thus progressively restricted. At the scene, he/she develops hypotheses and elaborates an experimental approach; he/she makes the essential of the decisions to investigate some places, engage technologies, and detect, record, and collect traces. During the inquiry, he/she shares the approach with experts and investigators in a more collective decision-making process. At the court, the decision is entirely the prerogative of the judge and the jury; the forensic scientist limits his role to show how the specific piece of evidence may change a priori beliefs on the propositions coming from prosecutor and the defense.

A Path Forward: Forensic Intelligence

This conception of traces, rather than representing a revolution, refers to the origin of forensic science, when traces were neither entrapped in predefined reasoning schemes nor confined to be used restrictively within the context of a specific case. Moreover, as pioneers of the discipline already identified it, physical traces can also be used beyond the frame of a given case. This additional dimension, referred as forensic intelligence, relies on physical traces to forge or corroborate connections between discrete actions perpetrated by the same offender, to draw an overall view of criminal phenomena, or to follow the evolution of certain types of crime (Ribaux and Margot 2007). The constitution of databases collecting massive amount of traces and information related to them increases the potential to find links between traces resulting from the same source. A systematic use of physical traces as indicators to establish potential links between actions, events, or crimes materializes an extension of the contribution of forensic science beyond the sphere of judicial inquiry, for security issues. While still underexploited, the potential of physical traces to act as semiotic marker of connections between crimes will indisputably gain more and more attention (Rossy et al. 2012). As it does not entail the transfer or divulgation of personal information, the exchange of traces escapes ethical or legal restrictions. Therefore, it will certainly become a major tool for the prevention and the repression of illegal activities occurring beyond administrative, structural, or political borders. Forensic intelligence may extend its scope of application to transregional or transnational volume crimes, to the comprehension of several forms of organized crime (such as to contribute to the fight against drug of abuse trafficking for instance), and to the identification of repeated offenses perpetrated in different jurisdictions.

Far away from the picture depicted in TV shows or in the media, the current needs expressed by law enforcement authorities, judicial authorities and other beneficiaries of the information conveyed by physical traces, lead to reconsider the role and the scope of contributions of forensic science. Forensic scientists must emancipate from the models and reasoning schemes they've applied so far to bring into focus the multiple and nonexclusive dimensions of information that can benefit from the detection, collection, and exploitation of physical traces.

Considering traces in criminal inquiry and intelligence necessitates a shift from the traditional court-oriented laboratory to a better integration in policing. Rather than focusing on the validity of specific techniques used mostly in routine processes, forensic science should much more engage in the resolution of criminal problems. The variety of the contribution and the logical processes, more offensive and risky by nature, have dramatic consequences in terms of the flexibility of standards, harmonized processes, attitudes, and globally on the way to envisage quality management for forensic operations. It necessitates combining quality of services for a court, with integrating to intelligence-led and investigative processes.

Related Entries

- ► Cognitive Forensics: Human Cognition, Contextual Information, and Bias
- ► Critical Report on Forensic Science
- ► Forensic Science

- ▶ Forensic Science and the Paradigm of Quality
- ▶ Forensic Science Culture
- ▶ Forensic Science Effectiveness
- ▶ Forensic Science Professions
- ▶ Organizing and Supplying Forensic Science Services
- ▶ Philosophical Basis of the Forensic Process
- ▶ Probability and Inference in Forensic Science

Recommended Reading and References

Barclay D (2009) Using forensic science in major crime inquiries. In: Willan Publishing Ltd (ed), Handbook of forensic science, Willan Publishing, Cullompton

Baskin D, Sommers I (2010) The influence of forensic evidence on the case outcomes of homicide incidents. J Crim Just 6:1141–1149

Brodeur J-P (2010) The policing web. Oxford University Press, Oxford

Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA (1998) A hierarchy of propositions: deciding which level to address in casework. Sci Just J Forensic Sci Soc 4:231–239

Durnal EW (2010) Crime scene investigation (as seen on TV). Forensic Sci Int 1–3:1–5

Kind SS (1994) Crime investigation and the criminal trial: a three chapter paradigm of evidence. J Forensic Sci Soc 3:155–164

Margot P (2011) Commentary on the need for a research culture in the forensic sciences. UCLA Law Rev 3:795–801

National Research Council (2009) Strengthening forensic science in the United States: a path forward. Committee on Identifying the Needs of the Forensic Sciences Community, The National Academies Press, Washington, DC

Ribaux O, Margot P (2007) La trace matérielle, vecteur d'information au service du renseignement. In: M. Cusson, B. Dupont, F. Lemieux (eds), Traité de sécurité intérieure, Éditions Hurtubise inc, Montréal

Ribaux O, Baylon A, Roux C, Delémont O, Lock E, Zingg C, Margot P (2010a) Intelligence-led crime scene processing. Part I: forensic intelligence. Forensic Sci Int 1–3:10–16

Ribaux O, Baylon A, Lock E, Delémont O, Roux C, Zingg C, Margot P (2010b) Intelligence-led crime scene processing. Part II: intelligence and crime scene examination. Forensic Sci Int 1–3: 63–71

Rossy Q, Ioset S, Dessimoz D, Ribaux O (2012) Integrating forensic information in a crime intelligence database. Forensic Sci Int http://dx.doi.org/10.1016/j.forsciint.2012.1010.1010

Wilson DB, McClure D, Weisburd D (2010) Does forensic DNA help to solve crime? The benefit of sophisticated answers to naive questions. J Contemp Crim Just 4:458–469

Wong Z, Wilson V, Patel I, Povey S, Jeffreys AJ (1987) Characterization of a panel of highly variable minisatellites cloned from human DNA. Ann Hum Genet 51:269–288

Forensic Science and Miscarriages of Justice

Simon A. Cole
Department of Criminology, Law & Society,
University of California, Irvine, Irvine, CA, USA

Synonyms

Criminalistics; Wrongful convictions

Overview

The relationship between forensic science and miscarriages of justice is complex and paradoxical. Miscarriages of justice are, in a sense, fundamentally unknowable. Forensic science, in the form of postconviction DNA testing, is the data source of much of the little we do know – and much of what we feel we know most securely – about miscarriages of justice. At the same time, forensic science has emerged from those very data as a significant contributor to miscarriages of justice.

Conceptual Framework

"Forensic science" is a broad term encompassing a variety of different techniques for using physical evidence in the investigation of crime. Forensic techniques include document examination, toxicology, pathology, drug analysis, print analysis, impression evidence, hair, fibers, paint, glass, soil, entomology, arson and explosives, gunshot residue, materials analysis, "jigsaw" physical fit matching, ballistics, blood spatter, crime scene reconstruction, computer forensics, serology, and DNA profiling.

"Miscarriages of justice" is an ambiguous term, more commonly used in the UK than elsewhere, that "can be defined in many different ways and nearly in whatever way one wishes" (Nobles and Schiff 2000). For scholars who construe the term broadly, a miscarriage of justice is any legal outcome in which the result is not just. Thus it may include both the conviction of factually (or "actually") innocent persons ("wrongful conviction") and the acquittal of factually guilty persons ("wrongful acquittals"). It may also include the conviction of legally, though not factually, innocent persons, those whom the state was not able to prove guilty to the appropriate legal standard. Arguably, it may also include much broader categories of injustice, such as excessive or insufficient punishments, unfair procedures, or unjust outcomes of nontrial procedures: pretrial detentions, plea bargains, failures to prosecute, dropped charges, closed investigations, and so on. In this sense, "miscarriage of justice" is a broader category than "wrongful convictions," a term with which it is often used almost interchangeably (Huff and Killias 2008). These ambiguities have inspired some commentators to propose alternative conceptual terminology such as "errors of justice," "false convictions," "actual innocence convictions," "unlawful convictions," and "truly innocent persons," none of which has really caught on. This entry primarily employs the generic term "miscarriages of justice," while occasionally referring to "wrongful conviction" or "wrongful acquittal" to designate particular types of miscarriage.

"Wrongful convictions" include at least three categories of cases: the conviction of the factually innocent, the conviction of the factually guilty but legally innocent, and the conviction of the factually guilty through procedures tainted by judicial error (Risinger 2007). Factual innocence (one did not commit the crime) is not the same as legal innocence (the state has not proved one guilty by the appropriate legal standard), though they are often conflated by the public (Nobles and Schiff 2000). The legal terms "convicted" and "acquitted" are not necessarily coterminous with the commonsense terms "guilty" and "innocent," and a trial, strictly speaking, is not concerned with determining "innocence," though its determinations do often come to be equated with innocence, or the lack of it, socially. For scholars with a narrower, legalistic definition of "miscarriages of justice," the procedurally tainted conviction of a factually guilty person would be a miscarriage of justice, but not a "wrongful conviction" (Naughton 2007). Most miscarriages of justice scholars, however, are primarily concerned with the conviction of the factually innocent.

A further complication is how to define "factual innocence." One approach is to rely on the courts, possibly supplemented by the findings of official inquiries or commissions, leading to the rather unsatisfying position that "miscarriages of justice are whatever appellate courts say they are" (Edmond 2002). The position is unsatisfying because courts are stingy about bestowing legal exoneration upon appellants, and many individuals claiming factual innocence are released from prison under procedural rulings, diversion from formal prosecution, or even guilty pleas – thus, without official findings of innocence. Another approach, then, is for external observers to try to make objective determinations of factual innocence. Inevitably, however, such determinations will be open to dispute by other observers. On very rare occasions, extra-legal events may provide strong epistemic authority for labeling something a "wrongful conviction," the classic example being the unexpected reappearance, alive and well, of a supposed homicide victim.

If we focus on the conviction of the factually innocent, a miscarriage of justice is the worst possible outcome of a legal procedure, producing the exact opposite of what it was intended and expected to yield: the awesome power of the state to punish has been deployed against a person undeserving of that punishment. In cases in which a crime undoubtedly occurred, the wrongful conviction is often accompanied by a collateral miscarriage of justice: the failure to punish the true perpetrator. Further adverse side-effects may include damage to the legitimacy of the courts and other criminal justice institutions. Because such outcomes are so manifestly unjust, they have often captured public attention. Such attention always operates on both a personal and a systemic level. In other words, miscarriages of justice are perceived as personal tragedies, but they are also generally perceived as symbolic of justice system failures (Nobles and Schiff 2000).

As an empirical topic, miscarriages of justice are notoriously difficult to investigate. Scholars, criminal justice system actors, and policymakers would especially like answers to two empirical questions about miscarriages of justice: (1) Prevalence: How frequent are they? (2) Causation: What proportion of blame for their occurrence should be assigned to what causes? Unfortunately, satisfying answers to these questions have been impeded by methodological stumbling blocks, of which two are paramount. First, miscarriages of justice are in a fundamental sense unknowable events since, by definition, they are events in which our primary determinant of truth, the criminal justice system, has produced falsehood and labeled it truth (Simon quoted in Gould and Leo 2010). As Gross (2008) notes, "We can't study an event if we can't tell when it happens. This is a severe problem for false convictions since, by definition, we don't know when they occur." What we can study, instead, is a small number of miscarriages of justice that have been *exposed* – been made known to the public. However, everything we know about exposure suggests that exposure is a fortuitous, rather than a systematic, process. Therefore, we do not know the proportional relationship between exposed and actual miscarriages of justice. A second major methodological headache is that the mechanisms typically exposing miscarriages apply to actual cases in a skewed, rather than a representative, fashion. Known miscarriages are skewed toward those serious crimes which attract the greatest legal, media, and public attention. Serious crimes, carrying longer prison sentences, present more time, as well as greater incentives, for the parties to pursue every avenue of redress. The most powerful exposure mechanism of all, post-conviction DNA profiling, is skewed toward a specific set of convictions, primarily rape-murders, that occurred during a specific historical period, disproportionately based on specific types of evidence (Gross 2008; Schiffer 2009; Simon quoted in Gould and Leo 2010; Natapoff 2012). It may be possible to make empirical generalizations about that set of cases, but whether and how to extrapolate from that data set to all relevant cases remains a contentious matter of judgment (Risinger 2007).

What, then, might these two phenomena, "forensic science" and "miscarriages of justice," have to do with one another? Forensic science is used as evidence in criminal prosecutions, and, as such, it may contribute to wrongful convictions. It may do so by erroneously implicating an innocent suspect, for a variety of reasons. Or, it may do so by failing to exculpate an innocent suspect. At the same time, forensic science may also be responsible for averting wrongful convictions by exculpating, or failing to implicate, innocent suspects when they fall under suspicion. Likewise forensic science might contribute to wrongful acquittals by failing to implicate, or even exculpating, a guilty suspect. Or, it may avert wrongful acquittals by implicating guilty suspects. Figure 1 summarizes these possibilities schematically by conceptualizing forensic science as an "independent check" on police investigators' theory of the crime. Of course, as noted above, "forensic science" is a general term encompassing a variety of techniques; the performance of different techniques in this scheme may well vary greatly.

Historical Relationship Between Forensic Science and Miscarriages of Justice

Historically, forensic science and miscarriages of justice were rarely discussed in concert. Certainly forensic science has been cited as a contributor to miscarriages of justice since as long ago as the Dreyfus case. But, until recently forensic science has tended to take a back seat in discussions of miscarriages of justice, compared to other issues like eyewitness identification, perjury, official misconduct, and interrogation practices (Roberts and Willmore 1993). Although the earliest US study of miscarriages of justice mentioned "[t]he unreliability of so-called 'expert' evidence" as

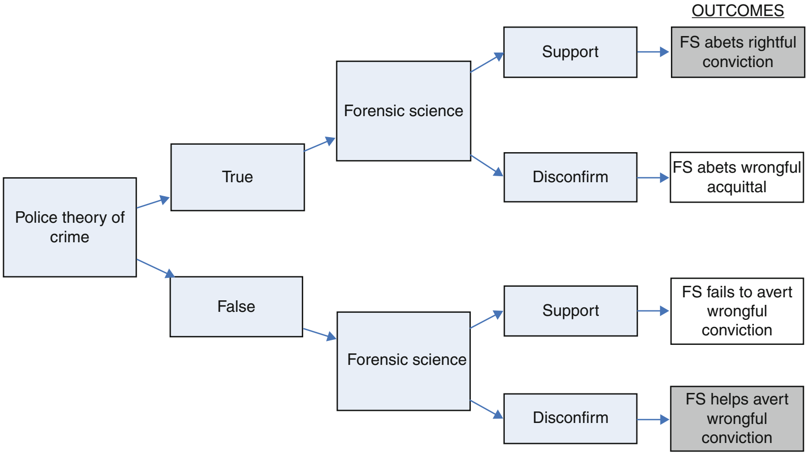
OUTCOMES
Police theory of crime
True
False
Forensic science
Support
Disconfirm
Forensic science
Support
Disconfirm
FS abets rightful conviction
FS abets wrongful acquittal
FS fails to avert wrongful conviction
FS helps avert wrongful conviction

Forensic Science and Miscarriages of Justice, Fig. 1 Conceptual model of role of forensic science in miscarriages of justice. Negative outcomes are in *white boxes*; positive outcomes in *dark gray*

a contributor to wrongful convictions (Borchard 1942), most of the early American studies which attempted to systematically identify causes of wrongful conviction discussed eyewitness identification, false confessions, police and prosecutorial misconduct, bad lawyering, race, failures of the discovery process, and public pressure for a conviction, but made scant mention of forensic science. Radelet et al. (1992) was a notable exception, discussing the use of misrepresented serology and hair evidence to leverage false confessions and misleading medical examiner testimony. As Schiffer and Champod (2008) observed, "forensic science (to convict and to exonerate) is underrepresented and often wrongly understood in research concerning wrongful convictions."

This disjunction between forensic science and miscarriages of justice made intuitive sense because the characteristics popularly associated with "science" would seem to be the antitheses of the characteristics of miscarriages of justice. Miscarriages of justice were thought to be caused by unclear, misguided, or fallacious reasoning, but science is supposed to embody clear, logical reasoning from valid, empirically demonstrable premises. Miscarriages of justice were thought to be caused by unjustified biases against people of certain races or classes, against persons with prior criminal records, or even simply against the police's preferred suspect, but science is supposed to be objective and free of bias. Miscarriages of justice were supposed to be caused by deceitful and otherwise unreliable information from witnesses, informants, co-conspirators, and the like, but science, goes the truism, "never lies." Miscarriages of justice were supposed to be caused by evidence that was less reliable than it appeared, like eyewitness identification evidence, but the very notion of science is associated, in the popular imagination, with high reliability, indeed often with certainty. For these reasons the notion that forensic science might contribute to miscarriages of justice is often treated as ironic because of the popular association of science with notions of "truth" and "certainty." Of course, any sober assessment should clearly understand that forensic techniques, like any other detection system, should be expected to yield errors – both "type I" and "type II" – at some rate (see Fig. 1). And yet, much discourse surrounding forensic science invokes popular stereotypes of science as "certain" in a way that other evidence is not.

The Rise of Forensic Miscarriages of Justice

The discourse on forensic science and miscarriages of justice changed dramatically during the 1990s. In large part this was due to the development of forensic DNA profiling in the mid-1980s, as will be discussed further below. However, even without DNA profiling, enough dramatic miscarriages of justice were exposed during the 1990s to generate a sense of a miscarriage of justice “crisis” (Nobles and Schiff 2000). In the UK, for example, this “crisis” was prompted by alleged miscarriages of justice in a series of Irish Republic Army (IRA) bombing cases, including the 1970s convictions of the “Guildford Four,” “Birmingham Six,” and “Maguire Seven,” some of which involved explosive residue evidence (Nobles and Schiff 2000). Two official inquiries prompted by these cases highlighted the role of forensic science in miscarriages of justice. The 1993 Runciman Report of the Royal Commission on Criminal Justice discussed a number of issues concerning forensic science, including failure to adhere to objectivity and impartiality, problems with interpretation of evidence, failure to communicate findings clearly, inequalities between defense and prosecution resources, defense access to samples, proprosecution bias, expert shopping, quality control issues, and the low accuracy of the residue detection techniques themselves (Edmond 2002). In 1994, the “May Inquiry” discussed the role of forensic science in the Guildford Four and Maguire Seven cases. The May Inquiry primarily blamed individual forensic scientists for the failings of forensic science (Edmond 2002). It has been observed that these cases could only be construed as miscarriages of justice by placing the same sort of faith in the *exonerating* scientific analyses that was – now, supposedly erroneously – initially placed in the *incriminating* forensic analyses (Edmond 2002). While these cases forged a connection between forensic science and miscarriages of justice, the most dramatic role in drawing attention to miscarriages of justice, especially in the US, came to be played by forensic DNA profiling.

Forensic DNA Profiling

The earliest use of forensic DNA profiling, in the investigation of two rape-murders in the English village of Narborough for which Colin Pitchfork was eventually convicted, arguably helped avert a miscarriage of justice in that the DNA evidence exonerated an individual who had emerged as the prime suspect and falsely confessed. Post-conviction DNA testing has exposed hundreds of miscarriages of justice in the US, beginning with the cases of David Vasquez, and then Gary Dotson, in 1989. These were both rape-murder cases in which physical evidence (hair and semen, respectively) presumed to derive from the assailants was recovered and implicated the defendants. However, post-conviction forensic DNA profiling on biological samples presumed to derive from the assailants excluded the convicted individuals. Because the prosecution theories of the crimes required that the defendants be the source of the samples, the two convicts were exonerated and released. Realizing the potential of post-conviction DNA testing to expose miscarriages of justice, the American attorneys Peter Neufeld and Barry Scheck founded the Innocence Project at Cardozo Law School in 1992 as a legal clinic dedicated to such testing. Over the next two decades, the Innocence Project and other independent efforts exposed more than 300 wrongful convictions in the US through post-conviction DNA testing. This set of wrongful convictions has taken on a significance beyond the parties involved in the cases themselves. These high-profile exonerations have drawn attention to the issue of miscarriages of justice, to flaws in the American justice system, and to capital punishment. In part, their significance derives from their ability to be packaged and conceptualized as a “data set,” and disseminated through reports, books, and the Innocence Project’s website. In addition, however, their significance derives from their ability to achieve supposed “scientific certainty” or “epistemological closure” (Aronson and Cole 2009). Alleged miscarriages of justice exposed through post-conviction DNA testing were less vulnerable to the sort of definitional disputes over whether alleged miscarriages of justice should be characterized as miscarriages of justice at all that had dogged previous scholarly analyses of miscarriages of justice. While some post-conviction DNA exonerations may be challenged, even the most determined innocence skeptics concede that the vast majority of post-conviction exonerations constitute genuine miscarriages of justice.

Post-conviction DNA exoneration has largely been an American phenomenon; other countries have not reported a proportionate spate of post-conviction DNA exonerations. Exposure of miscarriages in general seems to occur most frequently in the US and more often in the common law countries than in continental Europe (Schiffer 2009). However, it is unclear whether this difference represents a less frequent *occurrence* of miscarriages of justice because of different procedural safeguards and legal cultures or a less frequent *exposure* of miscarriages of justice perhaps because of less favorable policies on the retention of evidence or post conviction review (Huff and Killias 2008).

Forensic Science as Cause of Miscarriages of Justice

Post-conviction DNA exonerations emerged as a principal, and privileged, source of data about miscarriages of justice. A series of analyses of the post-conviction DNA exonerations were performed which treated the development of forensic DNA profiling as a sort of natural experiment that offered a window into flaws in the justice system. Each subsequent analysis treated a larger number of exoneration cases and was increasingly detailed (Connors et al. 1996; Scheck et al. 2000; Saks and Koehler 2005; Garrett 2008, 2011; Garrett and Neufeld 2009). These analyses were primarily concerned with identifying the major causes of wrongful convictions and roughly weighing their relative contributions. Among the most prominent causes identified were eyewitness identification, false confessions, perjury, police and prosecutorial misconduct, and ineffective counsel. Analyses of the post-conviction DNA exonerations, however, also revealed a paradox. Forensic science was not merely the engine for exposing miscarriages of justice: analyses of post-conviction DNA exonerations revealed that forensic science itself was ranked among the most prominent contributors to miscarriages of justice (Saks and Koehler 2005). Seemingly paradoxically, forensic science was little discussed as a cause of miscarriages of justice until its role was exposed by . . . forensic science.

Some analyses of exonerations have attempted to construct rank-ordered lists of contributory factors. The position of forensic science on such lists has varied. Saks and Koehler (2005) rated forensic science second only to eyewitness identification as a cause of miscarriages of justice, whereas Gross et al. (2005) hardly mentioned forensic science at all. A larger study by Gross and Shaffer (2012), however, lists forensic science fourth among the five leading causes of exposed wrongful convictions. Such measurements must be regarded as crude at best, and they have been challenged by defenders of forensic science (Collins and Jarvis 2009). Among the methodological problems that beset drawing inferences from post-conviction DNA exoneration data are: How should categories of causes be constructed? Should causes be coded for cases whenever present or only when contributing to the miscarriage of justice. If the latter, how would that be determined? How can we quantify the relative contribution of multiple causes to any individual miscarriage of justice when we do not know how much different items of evidence contributed toward the jury’s verdict? Finally, as noted above, post-conviction DNA exonerations are a manifestly unrepresentative data set. They can tell us something about the causes of the subset of miscarriages of justice susceptible to exposure through post-conviction DNA testing, but can tell us much less about the entire universe of all miscarriages of justice. While it seems reasonable to use such analyses as rough indicators of where in a criminal justice system the principal causes of miscarriages of justice might lie, the common practice of characterizing each cause as a proportion of the total number of exonerations risks media reporting that may fuel popular misconceptions. For example, the research finding that 60 % of miscarriages of justice exposed by post-conviction DNA testing were caused in part by eyewitness identification may be misinterpreted to imply that misidentification by eyewitnesses is responsible for 60 % of *all* miscarriages of justice; or even – far worse – that 60 % of eyewitness identifications result in miscarriages of justice.

Bearing these caveats in mind, analysis of post-conviction DNA exonerations clearly shows that forensic science contributed to exposed miscarriages of justice in two primary ways. First, serological evidence which ought to have been interpreted as either excluding the defendant or as having nothing useful to contribute to the fact finder's perception of the defendant's guilt was instead presented to the fact finder as inculpatory. This occurred in 67 of the first 250 post-conviction DNA exoneration cases. Second, microscopic hair comparison evidence that ought, if used at all, to have been conveyed to the fact finder only as failing to exclude the defendant, or perhaps as including the defendant among a very large population that could have contributed the hair, was instead presented to the fact finder as highly incriminating. This occurred in 29 cases. In Canada, meanwhile, the 1998 Kaufman Commission report discussed the role of microscopic hair comparison in contributing to the wrongful conviction of Guy Paul Morin for murder, a miscarriage of justice that was only exposed through post-conviction DNA testing (Kaufman Commission 1998; Roach 2009). Among other things, the report emphasized overstatement of the probative value of scientific findings and failure to disclose possible sources of contamination.

To be sure, these were not the only ways in which forensic science contributed to miscarriages of justice exposed through post-conviction DNA testing. Bitemark evidence, fingerprint evidence, shoe print comparison, voice analysis, and even DNA profiling all contributed to some miscarriages of justice. However, the frequency of such cases was small compared to those involving serology or hair comparisons (Garrett 2011).

Some have suggested that post-conviction DNA exoneration data point only to the failings of specific forensic techniques (serology and hair comparison), rather than to a problem with the larger institution of "forensic science" (Collins and Jarvis 2009). However, transcript analysis reveals that the issue was not merely these techniques' lack of discriminating power, but also repeated exaggeration of the probative value of the evidence by forensic expert witnesses (Garrett 2011). This suggests a general tendency among forensic scientists to exaggerate the probative value of evidence and a general failure of courts to control it. But to what extent is it valid to extrapolate from documented problems with serology and hair comparison to "forensic science" in general?

While serology is relatively undiscriminating and hair comparison may be a forensic technique with limited accuracy, at least part of the explanation for the prevalence of serology and hair cases in the post-conviction DNA exoneration data has to do with the skewed nature of the data set (Gross 2008; Schiffer 2009). Only a small subset of all miscarriages of justice is eligible to be exposed through post-conviction DNA testing. These are typically cases deriving from a specific historical period, in which preconviction DNA profiling was not performed, but biological evidence was preserved; in which biological evidence is recovered; and in which charges are serious enough for convicts and attorneys to make strenuous efforts to obtain post-conviction DNA testing. These cases will be skewed toward sexual assaults and rape-murder cases and away from crimes with lesser penalties (Risinger 2007; Gross 2008; Natapoff 2012). Such cases are quite likely to have relied upon serology – and to a lesser extent hair comparison – at the time of the original conviction. We may thus expect post-conviction DNA testing to be better at exposing miscarriages of justice associated with serology or hair comparison than miscarriages of justice precipitated by, say, fingerprint identification. Even further at the extreme, consider, for example, arson and medical evidence of unexpected infant death which are now suspected of being major contributors to miscarriages of justice (Science and Technology Committee 2005; Findley 2011; Plummer and Syed 2012). Miscarriages of justice involving these forms of medical and scientific evidence are rarely, if ever, susceptible to post-conviction DNA testing. Indeed, in most cases they are not susceptible to dispositive exonerating evidence in any form, in part because the material issue is not the identity of the perpetrator but whether a crime was committed at all (the alternative hypothesis being that the death was accidental) (Naughton 2007; Findley 2011). Usually, the best the convict can hope for is a finding that the court relied upon scientific evidence that is now in doubt and that the conviction, therefore, should be quashed.

Some commentators argue that the post-conviction DNA exonerations offer a window into more systemic flaws within forensic science as an enterprise (Thompson 2008; Garrett 2011). These flaws include biased interpretation of evidence; poor regulation of forensic laboratories (Giannelli 2007); vague, nonstandardized, and misleading reporting of scientific conclusions; and failure to validate forensic techniques. Overlaid upon these problems is the courts' extremely permissive stance in admitting forensic science evidence at trial, despite these documented problems (National Research Council 2009; Garrett 2011).

The exposed cases run the gamut from alleged forensic vigilantism to what appears to have been "honest error" (Schiffer 2009). These are obviously quite different problems which invite different remedies. Vigilantism suggests a sort of generic personnel problem that could affect any industry, whereas "honest errors" seems to indicate flaws in forensic procedures themselves. Assigning exposed miscarriages of justice to specific causes, however, is problematic. Often, it is difficult to determine through post hoc analysis whether a forensic analysis that contributed to a miscarriage of justice derived from malicious intent or honestly held belief. While thorough and transparent documentation of the reasoning behind a forensic conclusion may permit answering this question, many forensic techniques historically have required only the kind of rudimentary documentation of conclusions that would be of little help in determining the cause of an error. In addition, once a miscarriage of justice has been exposed, the forensic analysts involved will usually have retained their own lawyers and have little incentive to speak candidly with auditors.

In sum, while post-conviction DNA exoneration data may be helpful in drawing attention to systemic problems in forensic science and the courts' treatment of it, the two key issues, concerning (1) the prevalence of forensic miscarriages of justice and (2) the relative magnitude of forensic science as a contributing factor, remain matters of speculation and – sometimes heated – debate.

Broader Policy Impact

Despite these methodological limitations, analyses of exposed miscarriages of justice have exerted considerable influence on US public policy, notably in relation to the death penalty. Abolitionists have cited miscarriages of justice exposed – often fortuitously – through post-conviction DNA testing as clear evidence that the risk of executing an innocent person in the American capital punishment system was too great. One federal court even adopted this view before it was overturned by a higher court (United States v. Quinones 2002). Post-conviction DNA exoneration provided an appealing rhetoric in which "science" was seen exposing the unreliability of American capital punishment. However, this rhetoric proved to be a double-edged sword in that some politicians adopted the view that convictions that rested upon forensic science might be viewed as "certain," and thus impervious to the risk of being labeled miscarriages of justice. Forensic science, then, might render capital punishment certain and safe (Aronson and Cole 2009). Such claims obviously belie the understanding of forensic science as a detection system that should be expected to produce a certain number of errors, as a statistical truism (see Fig. 1).

Another area in which these analyses have had an impact is on the movement to reform forensic science itself. Although it may be argued that there are plenty of good reasons to reform forensic science independent of miscarriages of justice – lack of validation of techniques; lack of accreditation, certification, and regulation; lack of basic research; lack of standards; vague reporting practices; insufficient funding; insufficient education and training; insufficient independence from law enforcement; insufficient ties with mainstream science, and so on – high-profile miscarriages of justice have been important in supplying tangible causes célèbres and a *raison d'être* to propel forensic science reform. If it cannot be shown that acknowledged problems in forensic science actually result in miscarriages of justice, policymakers may wonder why forensic reform is necessary when apparently just results continue to be achieved despite acknowledged weaknesses. Such claims are sensitive to what might be called the "base rate of guilt" – if the police present forensic analysts with an extremely high proportion of factually guilty suspects, even very poor forensic analyses may yield very high rates of factual accuracy. Nevertheless, it seems that it is difficult to generate public and political momentum to improve forensic science as an end in itself, so that miscarriages of justice are invariably invoked in official reports urging reform of the forensic sciences. In this essentially presentational and rhetorical sense, recent miscarriages of justice have played an important role in giving forensic science reform more traction than it has been able to secure in the past (e.g., Kaufman Commission 1998; Science and Technology Committee 2005; National Research Council 2009).

Other Sources of Data about Forensic Science and Miscarriages of Justice

In view of the well-known methodological limitations of relying on post-conviction DNA exonerations as a measure of miscarriages of justice, an alternative approach attempts to preempt objections regarding representativeness by studying all relevant cases in which particular forensic techniques have been utilized (Cooley 2004; Giannelli 2007). Yet this alternative research methodology still suffers from the principal problem afflicting all miscarriages of justice research: the fortuity of exposure of miscarriages of justice. It is clearly statistically inadequate to estimate the rate at which a particular technique produces miscarriages of justice simply by treating exposed errors attributable to that technique as the numerator and all cases in which it figured as the denominator of a fraction (Gross et al. 2005; Naughton 2007; Gross 2008).

Another approach is to try to use experimental psychology to model the processes and "human factors" which might cause forensic science to contribute to a miscarriage of justice. There have been several studies of contextual bias in forensic science, developing the argument that biased scientific evaluations or expert opinions may be responsible for some miscarriages of justice (e.g., Schiffer 2009). Schiffer (2009) endeavored to study the relationship between forensic science and miscarriages of justice by interviewing forensic laboratory managers. Contradicting much received wisdom on what causes forensic science to go awry, Schiffer's interviewees suggested that the locus of error might be the crime scene as much as the crime laboratory. They also maintained that more coordination between forensic scientists and law enforcement might reduce forensic errors. This finding is in tension with the contextual bias literature, which tends to imply that forensic scientists should be shielded from "distorting" contextual information about the case provided by investigators and prosecutors.

Conclusion

The relationship between forensic science and miscarriages of justice has received greater attention over the last two decades, but that relationship remains complex. Forensic science, primarily in the form of post-conviction DNA profiling, has emerged as among the most powerful and persuasive exposers of miscarriages of justice. However, in the very process of exposing miscarriages of justice, post-conviction DNA profiling has implicated forensic science – including DNA profiling – as a contributor to acknowledged cases of wrongful conviction. Increasing awareness of miscarriages of justice has lent impetus to the growing clamor advocating reform of forensic science.

Generating empirical knowledge about the role of forensic science in miscarriages of justice poses methodological difficulties that render it difficult to draw firm conclusions. The evidence amassed thus far, however, does indicate that two particular forensic sciences, serology and microscopic hair comparison, played a major contributory role in generating that restricted and unrepresentative set of miscarriages of justice that were susceptible to exposure through post-conviction DNA testing.

The extent to which it is safe to extrapolate that finding to those unexposed miscarriages of justice that may have occurred in other cases or to other forensic science disciplines remain matters of ongoing debate.

Related Entries

- ▶ Causes of Wrongful Convictions
- ▶ Cognitive Forensics: Human Cognition, Contextual Information, and Bias
- ▶ Crime Science
- ▶ Critical Report on Forensic Science
- ▶ DNA Profiling
- ▶ DNA Technology and Police Investigations
- ▶ Expert Witnesses: Role, Ethics, and Accountability
- ▶ Forensic Science
- ▶ Forensic Science and Criminal Investigation
- ▶ Forensic Science and the Paradigm of Quality
- ▶ Forensic Science Culture
- ▶ Forensic Science Effectiveness
- ▶ History of the Death Penalty
- ▶ Innocence Commissions
- ▶ Legal Rules, Forensic Science and Wrongful Convictions
- ▶ Measuring Wrongful Convictions
- ▶ Organizing and Supplying Forensic Science Services
- ▶ Overturning Wrongful Convictions and Compensating Exonerees
- ▶ Prosecution and Wrongful Convictions
- ▶ Scientific Basis of the Forensic Process
- ▶ Scientific Evidence in Criminal Adjudication
- ▶ Scientific Evidence in Criminal Prosecutions

Recommended Reading and References

Aronson JD, Cole SA (2009) Science and the death penalty: DNA, innocence, and the debate over capital punishment in the United States. Law Social Inquiry 34(3):603–633

Borchard E (1942) Convicting the innocent: errors of criminal justice. Archon, Hamden

Collins JM, Jarvis J (2009) The wrongful conviction of forensic science. Forensic Sci Policy Manag 1(1):17–31

Connors E, Lundregan T, Miller N, McEwen T (1996) Convicted by juries, exonerated by science: case in the use of DNA evidence to establish innocence after trial. National Institute of Justice, Washington, DC

Cooley CM (2004) Reforming the forensic science community to avert the ultimate injustice. Stanford Law Policy Rev 15(2):381–446

Edmond G (2002) Constructing miscarriages of justice: misunderstanding scientific evidence in high profile criminal appeals. Oxf J Legal Stud 22(1): 53–89

Findley KA (2011) Defining innocence. Albany Law Rev 74(3):1157–1208

Garrett BL (2008) Judging innocence. Columbia Law Rev 108:55–142

Garrett BL (2011) Convicting the innocent: where criminal prosecutions go wrong. Harvard University Press, Cambridge

Garrett BL, Neufeld P (2009) Invalid forensic science testimony and wrongful convictions. Virginia Law Rev 95(1):1–97

Giannelli P (2007) Wrongful convictions and forensic science: the need to regulate crime labs. North Carolina Law Rev 86:163–235

Gould JB, Leo RA (2010) One hundred years later: wrongful convictions after a century of research. J Crim Law Criminol 100(3):825–868

Gross SR (2008) Convicting the innocent. Annu Rev Law Social Sci 4:173–192

Gross S, Shaffer M (2012) Exonerations in the United States, 1989–2012. National Registry of Exonerations. http://www.law.umich.edu/special/exoneration/Documents/exonerations_us_1989_2012_full_report.pdf. Accessed on 16 July 2012

Gross SR, Jacoby K, Matheson DJ, Montgomery N, Patel S (2005) Exonerations in the United States 1989 through 2003. J Crim Law Criminol 95:523–560

House of Commons Science and Technology Committee (2005) Forensic science on trial. London, TSO

Huff CR, Killias M (eds) (2008) Wrongful conviction: international perspectives on miscarriages of justice. Temple University Press, Philadelphia

Kaufman Commission (1998) Report on proceedings involving Guy Paul Morin

Natapoff A (2012) Misdemeanors. South Calif Law Rev 85:101–163

National Research Council (2009) Strengthening forensic science in the United States: a path forward. National Academies Press, Washington, DC

Naughton M (2007) Rethinking miscarriages of justice: beyond the tip of iceberg. Palgrave Macmillan, Basingstroke

Nobles R, Schiff D (2000) Understanding miscarriages of justice: law, the media, and the inevitability of crisis. Oxford University Press, Oxford

Plummer C, Syed I (2012) 'Shifted science' and post-conviction relief. Stanford J Civil Rights Crim Law 8:259–297

Radelet M, Bedau H, Putnam CE (1992) In spite of innocence: erroneous convictions in capital cases. Northeastern University Press, Boston

Risinger DM (2007) Innocents convicted: an empirically justified factual wrongful conviction rate. J Crim Law Criminol 97(3):761–806
Roach K (2009) Forensic science and miscarriages of justice: some lessons from comparative experience. Jurimetrics 50:67–92
Roberts P, Willmore C (1993) The role of forensic science evidence in criminal proceedings. HMSO, London
Saks MJ, Koehler JJ (2005) The coming paradigm shift in forensic identification science. Science 309:892–895
Scheck B, Neufeld P, Dwyer J (2000) Actual innocence: five days to execution and other dispatches from the wrongly convicted, 1st edn. Doubleday, New York
Schiffer B (2009) The relationship between forensic science and judicial error: a study covering error sources, bias, and remedies. PhD, University of Lausanne, Lausanne
Schiffer B, Champod C (2008) Judicial error and forensic science. In: Huff CR, Killias M (eds) Wrongful conviction: international perspectives on miscarriages of justice. Temple University Press, Philadelphia, pp 33–55
Thompson WC (2008) Beyond bad apples: analyzing the role of forensic science in wrongful convictions. Southwestern Univ Law Rev 37:1027–1050
United States v. Quinones (2002). 313 F.3d 49 (2nd Cir.)

Forensic Science and the Paradigm of Quality

Đurđica Hazard[1], Eric Stauffer[1,2] and
Pierre Margot[3]
[1]School of Criminal Justice, University of Lausanne, Lausanne, Switzerland
[2]Crime Scene Unit and Crime Laboratory, Fribourg State Police, Fribourg, Switzerland
[3]School of Criminal Justice, Institute of Forensic Science, University of Lausanne, Lausanne, Switzerland

Overview

Due to various contexts and processes, forensic science communities may have different approaches, largely influenced by their criminal justice systems. However, forensic science practices share some common characteristics. One is the assurance of a high (scientific) quality within processes and practices.

For most crime laboratory directors and forensic science associations, this issue is conditioned by the triangle of quality, which represents the current paradigm of quality assurance in the field. It consists of the implementation of standardization, certification, accreditation, and an evaluation process. It constitutes a clear and sound way to exchange data between laboratories and enables databasing due to standardized methods ensuring reliable and valid results; but it is also a means of defining minimum requirements for practitioners' skills for specific forensic science activities.

The control of each of these aspects offers non-forensic science partners the assurance that the entire process has been mastered and is trustworthy. Most of the standards focus on the analysis stage and do not consider pre- and post-laboratory stages, namely, the work achieved at the investigation scene and the evaluation and interpretation of the results, intended for intelligence beneficiaries or for court. Such localized consideration prevents forensic practitioners from identifying where the problems really lie with regard to criminal justice systems.

According to a performance-management approach, scientific quality should not be restricted to standardized procedures and controls in forensic science practice. Ensuring high quality also strongly depends on the way a forensic science culture is assimilated (into specific education training and workplaces) and in the way practitioners understand forensic science as a whole.

(Scientific) Quality

Generally speaking, quality represents the attribute of an element that will make this element different from the others; but it also refers to the notion of grade as a *"degree of excellence."* The American Society for Quality (ASQ 2011) defines quality as *"[a] subjective term for which each person or sector has its own definition. In technical usage, quality can have two meanings: 1. the characteristics of a product or service that bear on its ability to* *satisfy stated or implied needs; 2. a product or service free of deficiencies*." Indeed, reaching the highest degree of excellence is relatively symptomatic of the modern society, for which being efficient and effective is mandatory to ensure customers' satisfaction in a commercial competition. Quality management appears to be a key aspect in monitoring performances and in reaching preestablished goals.

Historically, the quality movement finds its roots in the Middle Ages (thirteenth century), where craftsmen started to organize guilds. These professional unions were dedicated to enacting rules to ensure product quality but also to exercise commercial control over their specific activities. As a sign of positive inspection, craftsmen were asked to mark flawless goods. With time, they added a new mark as an indication of origin that turned to be a pledge of quality related to the craftsmen's reputation. During the Industrial Revolution and throughout the twentieth century, the original marking practice of the craftsmen changed through manufacturing sector evolution while conserving an inspection dimension with the quality control of processes. Then, the total quality approach emerged from Japan, where quality assurance was no longer only focused on the manufacturing process, but encompassed the whole organizational process. In the twenty-first century, people are developing new quality control systems that involve not only the manufacturing sector but also areas such as healthcare or education, for instance. These quality systems have been developed to become a widely used marketing tool that controls industrial production, without taking into consideration the individual needs of each area.

Quality management is specific to any sector that wants to be competitive while being effective and efficient. For that matter, striving for quality calls for quality assurance and quality control. Quality assurance is the establishment of a program ensuring that quality standards are met through monitoring and evaluating the different stages of a process. Quality control consists of an aggregate of operational activities set up to meet the quality requirements.

The use of the "scientific" epithet before "quality" appears to give a stronger dimension of objectivity and robustness to that notion. This is often cited whenever science comes into the equation. Paraphrasing Jean Ladrière (1995), science is a critical mode of knowledge (or *"mode de connaissance critique"* in French), being characterized by specific principles and methods used through a reflexive and prospective reasoning process, which enables an extension of its body of knowledge. A way to strengthen a science is to challenge its principles and reasoning using a critical approach; this is achieved by establishing validation criteria used to test and control processes and every aspect of this science. Within that context, it can be assumed that the notion of scientific quality is also part of a critical approach of a science. It involves defining sound criteria of quality that require practitioners' agreement within the scientific discipline in order to evaluate and improve the degree of excellence of methods and processes. The implemented criteria and tools are found in quality management systems to improve performances and to ensure scientists' competencies in applying scientific methodologies. However, setting up quality criteria for a science implies a serious study of its dimensional aspects. It not only pertains to the science's application – this is where a suitable quality management system is of importance – but also to its foundations and philosophy, or more simply to its epistemology. In other words, scientific quality is often considered from the management's perspective in terms of performance when it may be more relevant to consider it from a cultural perspective of the given science (see ▶ Forensic Science Culture).

Within forensic science practice, quality is currently driven by a quality management culture (largely influenced by legal and public funding aspects), whereas scientific quality viewed from a forensic science culture perspective is largely missing.

Quality in Forensic Science

Since quality is defined differently depending on persons or activity sectors, scientific quality in forensic science should be determined by forensic science practitioners and researchers. Willis (2010) introduced this specific notion of quality but further added the need for competencies. She highlighted the Japanese two-dimensional vision of quality, which implies that quality must be adapted to specific needs while taking into account customers' considerations. This position is currently widely shared by the forensic science management communities. Quality notions in forensic science are oriented according to the business relationships between customers (defense attorneys, prosecutors, etc.) and (forensic science) service providers. This approach is clearly dictated by the quality movement coming from the manufacturing sector and driven by public administrations.

This vision was first applied to the laboratory examination stage. This is likely because this stage could largely benefit from methods that were already standardized and because it is the easiest one to get into a quality management system, at least for routine analyses, such as with drug or toxicological samples. Now, international forensic organizations tend to expand that vision to the pre- and post-laboratory examination stages, thus including the investigation scene as well as the opinion and report transmission (see ► Forensic Science and Criminal Investigation). This is where the forensic science culture truly becomes important: Although it was relatively straightforward to define and implement a quality system with laboratory procedures, it is a whole other problem with scene examination or opinion development. The forensic science community struggles today with the comprehensive quality approach because it failed to develop its own approach, instead, mistakenly copying the work from the manufacturing sector.

Quality System in Forensic Science Practice

Since the early 1990s, the quality triangle, presented by Lawrence Presley, a former quality assurance unit chief at the FBI laboratory, stands for the "paradigm of quality assurance" (Lentini 2006, 2009) for forensic science. Standardization, accreditation, and certification are the three sides of the quality triangle, which, when assembled together, represents the actual quality management system that should prevail. The evaluation process is used to test different aspects of the institution, such as its degree of expertise and the quality of the processes and results produced by it.

The forensic science quality movement traces its roots back to the mid-1970s when the Forensic Sciences Foundation (FSF), which was funded by the US National Institute of Law Enforcement Assistance Administration (LEAA), conducted a study about proficiency performances in American crime laboratories. The 3-year project had the following objective: "*to determine the feasibility of proficiency testing as a tool to uncover problem areas in laboratory performance*" (Peterson et al. 1978). Two hundred nineteen laboratories from the United States and Canada participated in the project and performed up to 21 controlled tests. Despite results showing a wide range of performances between laboratories, the authors found a positive outcome by stating that "*properly supported, laboratories [could] be extremely proficient*." Since then, crime laboratories and professional organizations, recognizing the heterogeneity of practices characterized by a lack of regulations and quality control aspects, had worked to implement a culture of quality in forensic science. Initially, the American Society of Crime Laboratory Directors (ASCLD) initiated the accreditation process through its ASCLD/LAB (Laboratory Accreditation Board) program. The four conditions shown in Fig. 1 are part of this culture that has expanded over the last 40 years.

Proficiency Tests

Proficiency tests appear to be the starting point of the forensic practitioners' interest for quality issue. Even though they are still performed today, they are no longer the focus point of current quality systems, which are rather based on accreditation programs and certifications. However, proficiency tests represent a valuable process that helps monitor the performances of the institution and, by extension, the robustness

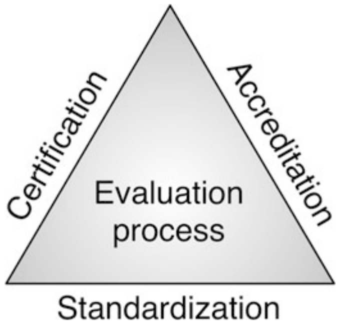

Forensic Science and the Paradigm of Quality, Fig. 1 The quality triangle as paradigm of quality assurance in forensic science

of their quality system. As such, they are part of a whole revolving around the quality triangle and cannot be dissociated from the other quality processes.

Proficiency tests are meant to evaluate different aspects of the service producer's practice. There are different types of proficiency tests. Blind proficiency tests are sent to a laboratory like regular casework, thus rendering the analysts unaware of the presence of an evaluation process. These tests are the most optimal for evaluating the entire process (administrative and scientific) of a laboratory without any bias. On the other hand, failing to pass the test does not enable to accurately identify where the problem lies. Known proficiency tests are usually utilized to evaluate a given part of the examination process. Some proficiency tests are designed to evaluate solely the interpretation process (e.g., a fingermark identification), while some others focus on the examination process (e.g.,, enhancement of fingermarks on a sheet of paper). However, since the analysts know they are working on a proficiency test, different resources may be attributed to the test than when working on regular casework.

As a result, proficiency tests are not limited to the evaluation of inter- and intra-laboratory/practitioners' knowledge, skills, and abilities. They (may) also include the evaluation of laboratory procedures and administration. In any case, they represent a useful tool to ascertain that the requirements for the accreditation process are met. In general, proficiency tests lead to the interlaboratory comparison of the results. This is a way of placing participants *"on a scale of agreements given by the consensus conclusion of the participants' population"* (CTS website). Pointing out potential differences in their conclusions when compared with colleagues and expected results provides an opportunity for confronting the robustness and reliability of methods and constitutes an interesting tool for analyzing what could be practically improved. However, foundations of quality systems are not based on interlaboratory testing, but on standard methods and practices (Lentini 2009).

Standardization

The standardization process consists of a normalization of (scientifically validated) procedures, considered as a valuable attribute of the quality system by forensic science practitioners and their partners (police, magistrates, public). This contributes to enhance the confidence of forensic science partners, presented as customers. However, it is (mistakenly) perceived as an assertion of reliability and validity of measurements and analyses.

Established norms for output format combined with the use of standard methods allow for interlaboratory comparisons and are necessary in databasing. In this regard, the use of a common language is the question that must first be addressed when a quality system is implemented.

Specific sources produce references within the forensic science field such as the following:

Scientific Working Groups (SWG), where forensic practitioners of a specific discipline meet to update their knowledge and, as mentioned in the framework for expert working group of ENFSI (European Network of Forensic Science Institutes), to promote quality assurance, develop professional standards, and harmonize methods.

ASTM International. A significant list of standard practices, test methods, and guides for forensic applications has been developed by ASTM International Committee E30 on forensic sciences. These standards are based on specific criteria accepted in the forensic science community.

Accreditation

The accreditation is defined as *“a procedure by which an authoritative body gives formal recognition that a body or person is competent to carry out specific tasks”* (ASTM E1732 2009). While ASTM International mentions accreditation for persons, the authors reserve the term “certification” for persons and the term “accreditation” for institutions.

This recognition process highlights the competencies of the accredited body to produce results using reliable and valid methods. Perceived to be of public benefit and as a quality safeguard, many institutions emphasized the implementation of accreditation as vital. Two examples are the National Academy of Sciences (NAS), which declared the implementation of accreditation and certification programs to be mandatory (see ► Critical Report on Forensic Science; NAS 2009), and the ENFSI, which requires pre-member institutes to be accredited (or in the process) to become members of its forensic science experts' network. The accreditation process is influenced by the business-oriented approach to the quality process, which is mainly focused on guaranteeing a high degree of customer satisfaction within a given economical environment (see ► Organizing and Supplying Forensic Science Services).

Management and professional entities support this culture of quality focused on tools applied to solve forensic science problems and, most particularly, laboratory processes. A substantial list of standards and guidance has been adopted and published, with the objectives of strengthening analytical procedures and ensuring more transparent actions in order to limit (or even cancel out) error risks. Within the collective perception, this analysis stage appears to be the place where problems occur; for most legal and forensic science communities, the forensic science process starts in the laboratory, without consideration that the process should be largely adapted to handle imperfect, incomplete, dirty trace specimens whose relevancy may also be questioned. This has become particularly evident with the increased sensitivity of techniques (such as DNA) and the problem of managing contamination issues. Pre- and post-analysis steps have become subject to difficult accreditation issues, requiring adaptation and specific guidance for the application of the norms.

The main accreditation norms introduced within forensic science are the following:

- *For crime laboratories*: ISO (International Organization for Standardization) 17025-05 *(General requirements for the competence of testing and calibration laboratories)*, associated with the guidance ILAC-G-19 (2002), specifically concerns forensic laboratories. This standard leaves proficiency tests as a minor activity within the quality assurance.
- *For scene investigation*: ISO 17020-98 *(General criteria for the operation of various types of bodies performing inspection)* is perceived to be more flexible than the previous norm leaving some scopes of adaptation to practitioners. Two specific standards comply with this norm: IAF/ILAC A4 (2004), guidance on its application, and the European standard EA-5/03 (2008) (*Guidance for the implementation of ISO/IEC 17020 in the field of crime scene investigation*). Interestingly, Australia has chosen to follow the stricter standard ISO 17025 (Horswell 2004).
- *For data evaluation and interpretation*: There is no accreditation norm emanating from ISO for the evaluation stage. However, the AFSP (Association of Forensic Science Providers) has proposed standards for the formulation of evaluative forensic science expert opinion in the UK (AFSP 2009; Brown and Willis 2009), although they are not compatible with most European continental legal systems.

In order to become accredited, an institution must go through an accreditation body, i.e., another institution that can verify that the accreditation standards are met and officially recognize it. In this regard, many accreditation bodies for forensic science have flourished around the world, such as ASCLD/LAB;

Forensic Specialties Accreditation Board (FSAB); National Association of Testing Authorities, Australia (NATA); and European co-operation for accreditation (EA).

Certification

Most norms ensure that procedures are followed according to rules of quality; they do not evaluate competences of the scientists behind the procedures. Certification is carried out as an independent process, established to monitor a forensic science practitioner's expertise and proficiency. Mostly conducted on a voluntary basis, certification is defined as the process that will recognize a minimum or certain level of competences evaluating the knowledge, skills, and abilities (KSA) of a scientist in a given field. Forensic professional organizations have to define the expected KSA in relation to their discipline. Setting up these programs is not an easy undertaking. There are as many definitions of KSA and conceptions of their evaluation, as there are professional certification programs. This situation shows a relative lack of homogeneity in assessing forensic scientists' competences. Every program has its specificity and requirements regarding the expected level of education (with or without academic degrees), experience (number of years), and examination type (written, oral, practical). Ideally, a certification program should attest to a forensic scientist's competence to carry out his/her duty, thus contributing to ensure a scientific quality within his/her practice. Unfortunately, in practice, certification does not always equate to competency.

Nevertheless, the expected KSA, the evaluation process of these competencies, and the conditions for maintaining certification should be well defined and harmonized within the whole forensic science community. The certification purpose should be specified, whether it is to ensure that a forensic scientist has reached a minimum level of competencies or to recognize a certain degree of expertise and excellence acquired through the practice (Stauffer and Schiffer 2007): two different perspectives that require their own approaches. It also means that the way forensic science is taught through various existing learning processes (education curricula, in-house training, continuing education) has sound and accepted foundations. For example, in the USA, certification programs do not require academic degrees as a condition to sit for the crime scene investigator certification exams, whereas university degrees, such as a Bachelor of Art (BA) or Science (BS) in a natural science, are required for analytical positions in a crime laboratory. Despite the fact there are many educational means of entering forensic practice, such discrepancies can be explained by a difference in the consideration of some forensic disciplines by forensic practitioners. For example, crime scene investigation is often not considered as part of the forensic science process by many practitioners, particularly laboratory ones. And yet, this is where the whole process begins. Surprisingly, many forensic science disciplines that require a strong and sound scientific basis suffer from a lack of relevant formal scientific education programs. This is the case with firearms/toolmarks examination or fingerprint examination.

Forensic science is a science with specific rules and a specific methodology that cannot be improvised or mastered through apprenticeships, but needs to be taught as a whole through a suitable and specific education program. It cannot be sufficiently emphasized that one must acquire the forensic science culture and formal education should not be mishandled. The needed harmonization of certification programs cannot be achieved unless the approach to forensic science education is fully reconsidered.

The Customer-Provider Relationship Approach

The standard terminology for forensic science (ASTM E1732-09) emphasizes that *"In quality management, consideration is given to economic aspects."* In reality, this approach has been taken to the extreme in the UK, where economy, effectiveness, and efficiency (named the 3 E's, coming from the circular 114/83 that had a strong impact on policing policy in UK in the 1980s) are linked and are part of a performance quest. Eventually, such a vision led to the closing of the

Forensic Science Services (FSS) in 2012 (see ► Organizing and Supplying Forensic Science Services, Lawless 2011).

A customer-provider relationship focus could limit and bias forensic scientists' vision in view of providing the best scientific information for their customers, the prosecution or defense rather than for justice or the courts (see ► Cognitive Forensics: Human Cognition, Contextual Information, and Bias). Quality assurance is important, but the way it is perceived and applied, i.e., through an economical perspective rather than a scientific one, leads to restraining the vision of forensic science from a complete process to an easily usable and high-performance tool for the sake of a client within the criminal justice system. Although forensic scientists provide results, which are considered valuable products, criminal justice system stakeholders should not be considered customers to satisfy, but partners to work and interact with in an open and effective way. Approaching the forensic science process through an economical perspective leads to fragmentation and compartmentalization that is detrimental to scientific inquiry (see ► Forensic Science and Criminal Investigation). As Willis and Brown (2009) stated, it is a *"complicated end-to-end-process,"* and market forces cannot rule it.

To cite Willis (2009), *"Quality assurance is increasingly seen as one of the safeguards of good science. It is sometimes mistakenly considered that quality systems guarantee good science and though this is definitely not the case, a quality system is vital in an organizational setting."* It is vital because it helps to harmonize methods and practices throughout forensic science practice. Furthermore, it helps to focus on the necessary KSA that forensic science practitioners must have to be good forensic scientists. Also, it forces crime laboratories to show some transparency in how things are done, a definite paradigm shift from the secrecy that these laboratories benefited from for decades. However, this race for standardization and accreditation should not be the first that forensic scientists should run.

Standardization is reached through peer consensus within a discipline and leads to processes of varying rigidities. In general, a standard puts up a structured and rigid frame, meaning that the analyst must not deviate at all from it. It is ideal to perform routine analyses, such as drug of abuse or toxicology analysis, since they can be ensured by calibration and comparisons with standards and accredited according to international quality norms. Alternatively, a standard could also offer some flexibility in its application, allowing the analyst to utilize other ways of achieving the desired result. Any standard can specify whether it is intended as a rigid frame or to offer some flexibility. The use of the wording "may," "could," "should," and "shall/must" usually defines the level of rigidity of the described action. However, the approach to standardization can only apply to specific and repetitive tasks on similar samples. As such, it does not allow for a view of the overall process that the forensic science practitioner has to face on a daily basis. This is where harmonization comes into play.

Harmonization, similarly to standardization, is based on agreement reached among expert groups such as ENFSI guides. Nevertheless, it is not meant to be mandatory or rigid, contrary to standards. Harmonization leads to recommendations and guidelines for which the scientist is able to choose fit-for-purpose methods to apply when faced with peculiar or less than perfect specimens, as it is the case with the investigation of a scene. Best practice manuals (based on a harmonized approach), edited by professional organizations such as ENFSI, should provide solid and sound tools for tackling every non-standardized specimen based on a scientific and critical approach. This perspective reinforces that a forensic specimen, or physical trace, does not have to adapt to practice but that practice has to adapt to the specimen or trace.

This is why the focus of forensic science practitioners should be on the harmonization process that leads toward quality developments rather than the standardization and accreditation of specific tasks that do not offer an overall view of a case.

Finally, independently of his/her KSA, a forensic science practitioner must adhere to a professional code of ethics. In this regard, professional associations, such as AAFS, American Academy of Forensic Sciences; ABC, American Board of Criminalistics; ANZFSS, Australian and New Zealand Forensic Science Society; and ENFSI, promulgate codes of ethical practice to which members must adhere. Without the ethical behavior of the forensic science practitioner, all the quality assurance and quality control systems in the world cannot offer any guarantee as to the validity of the examinations performed at the scene or at the laboratory. This is also where the process should sometimes start.

Strengths and Weaknesses of the Existing Forensic Science Quality Assurance Paradigm (Table 1)

The focus on quality in forensic science should differentiate various steps and applications in the whole process.

Strict controls and standards are necessary for routine applications that need interoperability (between laboratories and authorities), strict limits, or possibility to share databases, allowing comparisons between results originating from different laboratories that can be trusted along calibrated ranges. This means that data are of a comparable instrumental analytical nature and sufficiently routine to be used, as in a manufacturing plant. This is the case for DNA, some narcotic drug analyses, vehicle paints, blood alcohol concentration, etc.

However, standards and controls are clearly not adapted to deciphering information about criminal phenomena or activity-level information in forensic intelligence, investigation, or interpretation of an inferential nature (see ▶ Forensic Science and Criminal Investigation and ▶ Probability and Inference in Forensic Science). Yet, this is where forensic science is often the most useful and a valuable asset for the criminal justice system and for society (Ribaux et al. 2010a, b). Harmonization of the detection process here is essential.

One further difficulty in the way quality is perceived is the fragmentation and compartmentalization of data that feed linkage blindness. This phenomenon has been identified as the single most important factor in failed criminal/terrorist investigation (Egger 1984). A sound quality system for operational tools and methods is essential, but this is not where most problems lie when faced with crime-solving difficulties.

Forensic Science and the Paradigm of Quality, Table 1 Overview of the main strengths and weaknesses of forensic science performance management, based on the quality assurance system

Plus	Minus
Definition of international standards accepted through peer reviews	Overload with documentation
Standard operative procedures (SOP): anyone can rely on the documentation useful for new employees	Not so convenient and fit for purpose for some forensic science stages (crime scene and interpretation areas, for instance)
Transfer of procedural knowledge	Time and cost consuming
Ensure transparent actions by following standardized methods	Rigid frame that prevents initiative
Facilitation of data interchange between laboratories with feasible comparisons	Mechanical application of SOPs, *"Lack of focus on the purpose of the examination"* (Willis 2011)
Adequate to biological-chemical analysis and routine procedures	Does not ensure quality in the application of the forensic science as a science
Documentation of what has to be and was processed	Does not ensure a comprehensible implementation of the process by the forensic scientist
Systematic approach	Lack of scientific skepticism, '*brains are left outside the door and strict adherence to protocols replaces judgment*' (Willis 2011)
Legal and public trust in the forensic science process: robustness, validity, reliability	Overreliance upon the system

Concluding Remarks

Mainly attributed to laboratory management matters, quality issues are the engines for standardization and accreditation programs. All is related to evaluation processes through controlled and monitored actions based on standard operative procedures or specific attributes (criteria) considered as representative of what quality is. The criteria that are used are numerical, useful for lay people who need tools to evaluate and consider what seems to fulfill their high-quality requirements, and avoid errors and a lack of validity and reliability. Indeed, risks of miscarriages of justice are a reason that explains this quest for absolute quality assurance. Interestingly, perspectives diverge in perceptions of error sources and advanced solutions within forensic science circles and their partners (Schiffer 2009).

Quality within the forensic science process is a difficult question to answer to and a priority. However, as long as it is considered through the legal and economic perspectives, it is not consistent with case-oriented, strategic, and operational needs. To ensure the overall quality of instrumental processes and competences is a good start, but it is clearly insufficient, as it does not focus on the right issues: understanding criminal activities, diminishing linkage blindness, and solving crimes (Ribaux et al. 2010a, b). Legal and economical systems, rather than science, condition the way forensic science is approached as a profession. Forensic scientists should become the driving force in defining who they are and what they are working for. Ensuring and improving quality does not only find solutions in quality assurance and controls, where criteria are associated with performances management and market forces (see ► Organizing and Supplying Forensic Science Services, Lawless 2011), but also in the overall definition of the profession. The scientific foundations of forensic science define its values and performance as well as its position within a criminal justice system as an independent discipline. It provides a challenging perspective in this matter by emphasizing the whole process and the interdependence of the different parts of forensic science.

Related Entries

- ► Cognitive Forensics: Human Cognition, Contextual Information, and Bias
- ► Critical Report on Forensic Science
- ► Forensic Science and Criminal Investigation
- ► Forensic Science Culture
- ► Organizing and Supplying Forensic Science Services
- ► Probability and Inference in Forensic Science

Recommended Reading and References

ACPO (2008) Police reform green paper. The future of policing. Association of Chief Police Officers

AFSP (2009) Standards for the formulation of evaluative forensic science expert opinion. Sci Justice 49(3):161–164

ASQ (2011) http://asq.org/learn-about-quality/basic-concepts.html. Last checked March 2012

ASTM E1732 (2009) Standard terminology relating to forensic science

Brown S, Willis S (2009) Complexity in forensic science. Forensic Sci Policy Manag Int J 1(4):192–198

Egger S (1984) Working definition of serial murder and the reduction of linkage blindness. J Police Sci Adm 12(3):348–355

EN ISO/IEC 17025 (2005) Exigences génerales concernant la compétence des laboratoires d'étalonnage et d'essais

ENFSI-EA (2008) EA-5/03. Guidance for the implementation of ISO/IEC 17020 in the field of crime scene investigation. EA-5/03. ENFSI & European cooperation for Accreditation

Horswell J (2004) Crime scene investigation and third party quality systems accreditation. In: The practice of crime scene investigation. International Forensic Science and Investigation. CRC Press

IAF/ILAC-A4 (2004) Guidance on the application of ISO/IEC 17020

ILAC-G19 (2002) Guidelines for forensic science laboratories

Ladrière J (1995) Published: 2008 Encyclopedia universalis (on line). Sciences et discours rationnel

Lawless C (2011) The contested shaping of neo-liberal forensic science. Br J Criminol 51:671–689

Lentini JJ (2006) Liaison is a hallmark of committee E30 on forensic sciences. ASTM Standardization News. Feb. pp 22–25

Lentini JJ (2009) Forensic science standards: where they come from and how they are used. Forensic Sci Policy Manag Int J 1(1):10–16

NAS (2009) Strengthening forensic science in the United States: a path forward. Committee on identifying the needs of the Forensic Sciences Community, National Academy of Sciences

Peterson JL, Fabricant EL, Field K, Thornton JI (1978) Crime laboratory proficiency testing research program. National Institute of Law Enforcement and Criminal Justice, Washington, DC

Ribaux O, Baylon A, Roux C, Delémont O, Lock E, Zingg C, Margot P (2010a) Intelligence-led crime scene processing. Part I: forensic intelligence. Forensic Sci Int 195(1–3):10–16

Ribaux O, Baylon A, Lock E, Delémont O, Roux C, Zingg C, Margot P (2010b) Intelligence-led crime scene processing. Part II: intelligence and crime scene examination. Forensic Sci Int 199(1–3):63–71

Schiffer B (2009) The relationship between forensic science and judicial error: a study covering error sources, bias, and remedies. PhD Thesis, School of Criminal Justice, University of Lausanne, Switzerland

Stauffer E, Schiffer B (2007) La certification de forensicien: concepts de base et regard critique sur l'expérience américaine. Rev Int Crim Police Tech LX(4):461–478

Willis SM (2009) Forensic science, ethics and criminal justice. In: Fraser J, Williams R (eds) Handbook of forensic science. Willan Publishing Ltd, Cullompton

Willis S (2010) Power, process, people—a presentation on quality and competence in forensic science delivered at eafs2009. Sci Justice 50(1):23–25

Willis S (2011) The highs and lows of accreditation. Forensic Sci Policy Manag Int J 2(2):75–80

Forensic Science Culture

Đurđica Hazard[1] and Pierre Margot[2]
[1]School of Criminal Justice, University of Lausanne, Lausanne, Switzerland
[2]School of Criminal Justice, Institute of Forensic Science, University of Lausanne, Lausanne, Switzerland

Overview

Heretofore the issue of quality in forensic science is approached through a quality management policy whose tenets are ruled by market forces. Despite some obvious advantages of standardization of methods allowing interlaboratory comparisons and implementation of databases, this approach suffers from a serious lack of consideration for forensic science as a science. A critical study of its principles and foundations, which constitutes its culture, enables to consider the matter of scientific quality through a new dimension. A better understanding of what pertains to forensic science ensures a better application and improves elementary actions within the investigative and intelligence processes as well as the judicial process. This leads to focus the attention on the core of the subject matter: the physical remnants of the criminal activity, namely, the traces that produce information in understanding this activity. Adapting practices to the detection and recognition of relevant traces relies on the apprehension of the processes underlying forensic science tenets (Locard, Kirk, relevancy issue) and a structured management of circumstantial information (direct-indirect information). This is influenced by forensic science education and training. However, the lack of homogeneity with regard to the scientific nature and culture of the discipline within forensic science practitioners and partners represents a real challenge. A sound and critical reconsideration of the forensic science practitioner's roles (investigator, evaluator, intelligence provider) and objectives (prevention, strategies, evidence provider) within the criminal justice system is a means to strengthen the understanding and the application of forensic science. Indeed, the whole philosophy is aimed at ensuring a high degree of excellence, namely, a dedicated scientific quality.

Culture and Forensic Science

Culture is considered "in anthropology, [as] all knowledge that is acquired by human beings through their membership of a society. A culture incorporates all the shared knowledge, expectations and beliefs of a group." Besides, sociologists defined it as "human processes of meaning-making generating artefacts, categories, norms, values, practices, rituals, symbols, worldviews, ideas, ideologies and discourses" (Spillman 2007). If the first definition puts forward the acquisition of a specific knowledge within a limited group of persons, the second view specifies the nature of this extensive knowledge, which emanates from these meaning-making processes that have a signification or a reason to be only within a given society or community. Culture is not innate and is acquired through various cognitive processes including (continuing) education and (in-house) training.

Forensic science is characterized by heterogeneous conceptions regarding its position within criminal justice systems (CJS). It has its own culture, despite the fact that it is often perceived through the eyes of various specialized professional bodies and is hardly ever emphasized (Margot 2011a). The implementation of (accredited) standards and procedures according to the logic of market forces represents one of the meaning-making processes shared within the combined community; this does not stand for the core of the forensic science culture. Forensic science is neither chemistry nor biology or physics, but is in fact based upon a *historical* or *reverse engineering* reasoning surrounding *the physical residue of an event* (to paraphrase Kind (1987)), in the form of a physical or material trace.

This cultural background leads to investigate the meaning-making processes that generate the key characteristics of this science, like its fundamental principles or values. An epistemology of forensic science investigates its principles and information framework and the way this knowledge can be transmitted in the form of education, training, and research for a recognized profession. In this perspective, the notion of quality is not instrument or market based but rather positioned within a complex justice system, quality being the added value afforded by the extracted knowledge and its interpretation within the justice context.

The Physical Trace as a Primitive Source of Information

The Scientific Investigation of Crime

The need to introduce scientific foundations within the investigation of crime was understood and suggested by Gross in the 1880s (Gross 1899), followed by various authors in the twentieth century to the fundamental contribution of Kind. These authors emphasized this peculiar connection between forensic science and the investigation process—where police and magistrates are partners, who rely on a need for valuable information. Thus, various types of information, whether direct or indirect from human sources or investigative means, are traditionally considered in solving cases. These above authors introduced a further dimension with the physical information. It comes from the scientific study of the only physical remnants of the criminal activities, the physical traces. The latter are the essential components of evidence and intelligence-led policing due to the reliability of the information extracted from it, and it can turn into general policing knowledge.

In every science developing a critical relationship to time (such as history, archaeology), physical trace—improperly called evidence at the stage of its discovery—is considered the single most reliable record, seen as the relevant witness of past events, whereas forensic science still has the secondary role as far as understanding crime is concerned. It is mostly considered through its capacity as an evidential piece in case solving to support traditional law enforcement investigation. The quality of even fragmentary physical records is such that it often becomes evidence in desperately empty investigation due to its pure strength. It should take at least equal importance within the criminal justice framework since it is usually the only measurable resource. This task is complicated because current forensic scientists and partner communities do not give a great place to forensic science information other than the analytical end product for courts. There is little interest shown to determine the scientific quality outside the laboratory, because it is generally seen as a subaltern part of law enforcement or judicial administration with little or no control over any other processes but the laboratory process.

Nevertheless, physical records transmitted to crime laboratories are collected from scenes of investigation without, often, due consideration to the problem of detection, recognition, and diagnosis/interpretation that should be based on scientific and circumstantial observations. It is the starting point of the whole forensic science process. The quality of the information given to partners is mostly, not to say exclusively, conditioned by the work achieved at the scene of investigation: a sound reason to focus on this stage of the process whenever quality matters are discussed.

The scene of investigation is a complicated environment fraught with various uncertainties that constitute a risk forensic science practitioners have to confront, but also offers huge opportunities. Detecting and finding relevant physical records that will provide useful data is faced with the uncertainty of not knowing the causes that have created traces. The value of the ensuing analytical process may be highly controlled and calibrated; it remains useless when it comes to control the uncertainty pertaining to (criminal) causal events. The way forensic science practitioners deal with risk conditions decisions taken at this stage, which affect the whole process that follows the detection and collection of traces. From the collected observational data, a complex and continuous reasoning process starts at the scene through an abductive reasoning to determine/diagnose probable or possible causes/propositions of the resulting events that may be tested through a deductive phase, which should help question or strengthen those hypothetical explanations (Bonfantini and Proni 1988). Eventually, the selected hypotheses constitute a network that will lead to identify relevant information to be used for evidence-led inquiry and/or intelligence-led policing (Ribaux et al. 2010a, see ► Forensic Science and Criminal Investigation).

This management of uncertainties constitutes an interesting subject to study, approached by forensic science researchers through many ways: some focus on external/circumstantial bias effect advocating that information should be filtered to avoid contextual effect on decision making (see ► Cognitive Forensics: Human Cognition, Contextual Information, and Bias), others propose a Bayesian approach as an inference framework useful to consider objectively the whole spectrum of propositions within a given context (through the CAI (case assessment and interpretation) approach (Cook et al. 1999, see ► Probability and Inference in Forensic Science), or others carry out studies with sociological perception to determine the best conditions for improving process effectiveness (e.g., Julian et al. 2011, see ► Forensic Science Effectiveness). These approaches are not mutually exclusive but share a common perception that the most demanding step in the whole process occurs at the investigation scene and constitutes an essential quality issue, whereas the current debates are mostly focused on the end-product interpretation (after analyses) rather than the trace itself, the essential component of the forensic science discipline. The recognition of relevant information calls for a comprehensive vision and knowledge of what pertains to the apprehension of the trace, and a qualitative and quantitative appreciation of its information content (often incomplete) before any other issues.

The Physical Remnant of the Criminal Activity

Kind (1987) defined the trace as information whose origin was a material residue of the investigated event. More specifically, it is defined as a mark, a signal, or an object that is a visible sign (not always visible by naked eye) and a vestige indicating a former presence (source level information) and/or an action (activity level) of something where the latter happened. The physical trace is the common, elementary, and indispensable piece of the forensic puzzle. It is common since, to paraphrase Locard (1920), every action creates a trace on any kind of support as a consequence of interactions between the actors of the event and the support. It is elementary, because the trace represents the first essential entity onto which the reasoning process will be based at the scene. And it is indispensable by virtue of its potential content of information that feeds the investigative and intelligence processes. The trace can be viewed from three different angles via:

– The *genesis* (or creation): complex factors interact and lead to create relevant physical traces. The creation of the trace takes place in a specific context that is conditioned by the *activity* of the criminal (the *source*) who interacts with the *scene* according to basic forensic science *principles* (Locard, Kirk) within the *immediate environmental context* (according to the situational approach theory) in a certain period of *time*. The presence (or absence) of the physical element within such a framework allows inferences to be made about the criminal and his activity. Without its creation, there is little object of reasoning except the possible meaning of an absence of a trace.

- The *discovery*: this step corresponds to the intervention after the event when forensic science practitioners come into play. This implies successive reflections, decisions, and actions that will condition the latter stages of the forensic science process. The problem of finding, detecting, and recognizing relevant traces is not trivial; it requires a comprehensive study to understand the types and mechanisms of transfer. In any way, without the discovery of the trace (or a realization of an abnormal absence), there is also no object of analysis or reasoning.
- The *meaning-making process*: the information carried by the trace may be a strong indicator of source and/or activity. According to variable utilitarian dimensions and basic logical steps (such as trace-to-source, source-to-trace, trace-to-trace relationships), forensic science practitioners evaluate the potential information content of the trace. Among the various parameters used to decide its value, the utilitarian perception by the forensic science practitioner grossly influences the detection capability and the will to detect. Practitioners notoriously prefer highly discriminating traces or those presenting a direct relationship with their source, e.g., fingermarks and DNA, compared to those that offer an indirect relationship or considered as less discriminating, e.g., shoemarks and microtraces. This well-known preference for individualizing traces is due to the capacity to indicate the source, whereas indirect "evidence" indicates a probable source extended to a group or a specific activity, whose relevancy depends on the case's circumstances. This preference seems to be directly motivated by the strength it offers in the form of evidence in court, the main beneficiary of traditional forensic science. Practitioners may even decide to ignore some physical traces because of their expected complexity to be interpreted and/or used in court. In such situation, the selection at scenes is conditioned by an output concern "collected trace versus evidence easily interpreted in front of a Court" (Ribaux and Margot 2008). This view completely obliterates the rich information content (especially the activity dimension afforded by many traces) provided by physical data as shown in many studies. The multidimensional nature of relevant information for an extremely useful intelligence as well as evidence should be the norm and the cultural base of forensic science practice, because relevant physical traces provide information, whether in an intelligence process, in an investigative case solving perspective, or even to provide evidence for court.

The duality *criminal's activity–forensic science practitioner's activity* constitutes the elementary group conditioning the discovery and evaluation of relevant physical traces at scenes. If the former aspect is outside the control of the forensic scientist, the latter is entirely under her responsibility. Discussing scientific quality in this respect goes inevitably through a sound and introspective reflection of the role of forensic science and its activity as well as the perception of the profession.

The comprehensive vision and culture of the trace, as the essence of forensic science, emerges as a complex science whose principal role may not be the ultimate use by the courts, but the provider of essential data about crimes and their actors within the CJS that also helps public security strategies and measures. This leads to reflections about data collection and structured memories, and the necessity to implement sound and practical operating systems (e.g., relevant databases, data mining systems) to support and relay valuable and reliable information. This may strengthen the whole forensic science process only if it is part of a comprehensive model, such as the four-level model proposed by Ribaux et al. (2010a).

The Set of Inferences According to the Standpoint of "Clues" Paradigm

A culture of the physical trace exists in forensic science (Dulong 2004). It contributes to the representation of the discipline: a science ("a scientific approach and technical methods [based on] the study of physical traces" (Ribaux and Margot 2008)) fostering the ability to infer causes from observations. This model may be seen through the Ginzburg's *paradigma indiziario* or clues' paradigm (Ginzburg 1989), applied to the physical remnants of the criminal activity.

Ginzburg observed that society developed the view that knowledge about a whole could be obtained from physical details afforded by science and observations. He coined this *paradigma indiziario* as a model for a whole cultural movement in the nineteenth century. He traced the roots of this approach back into the cynegetics (i.e., the art to decipher traces and marks of animals decoded by hunters) back to Neolithic time when men were hunters and learned to use details through an elaborate, rapid, and complex thought mechanism to reconstruct events (Ginzburg 1989). The analogy between cynegetics and the forensic science process at the scene is astonishing: *the hunter* (*the forensic science practitioner*) *trails the animal* (*the criminal*) *using traces and marks* (*physical traces*) *left in the wood* (*at the investigation scene*) *by the animal* (*the author*) *following his passing* (*criminal activity*). The Ginzburg's paradigm has a distant origin and is found in various disciplines (e.g., divination, law, medicine) where the aim is to tell stories based on observations of traces, signs, and clues. This paradigm marked many detective novels in the nineteenth century. Writers, such as Voltaire, Poe, Gaboriau, and Doyle, turned their heroes into fine observers of traces who applied reasoning methods based on drawn inferences that can be linked to Peirce's pragmatic philosophy. More specifically, abductions are seen as a third fundamental form of logical inference, besides induction and deduction, that help devise logical explanatory hypotheses about observed results (traces/tracks), to go back to their potential probable causes.

Ginzburg draws a parallel between three specialists of importance: Morelli, doctor by training and art historian, who put in place an identification method relying on "negligible details" found on paintings in order to identify painters (method known today under the term "connoisseurship"); the second, Freud, who had an interpretative approach inspired from Morelli's method and used neglected details indicating the hidden presence of reality; and the third, Doyle, doctor and writer of the adventures of the famous detective gifted with an accurate sense of details. All three used traces within their reasoning processes, precisely symptoms for Freud, clues for Sherlock Holmes, and pictorial signs for Morelli. Besides their common medical training, they also shared the application of a reasoning model specific to medical semiotics, a medical discipline based upon sign study of the invisible disease in order to propose a diagnosis. The criminal activity is like the illness, leaving traces at scenes that can be visible or latent (like signs of illness); the trace is the indirect expression of this action that the forensic science practitioner investigates and tends to highlight by inferences drawn from discovered and collected traces (becoming clues), as the doctor who makes a diagnosis of a specific disease by listing and correlating the relevant symptoms. In the presence of an epidemic, the doctor may identify relevant signs just like forensic intelligence may help find hidden traces from serial crimes.

The trace may therefore be analyzed according to different viewpoints—which may also correspond to different specialized occupation from investigators and intelligence analysts, laboratory analysts to policy makers. The trace is considered through its information content to determine:

– Source/composition: it may be a pattern, a chemical composition, or a biological trait that will be identified (with the use of databases) and compared to establish potential common origin. Indeed, the instrumental analysis of the source is where current quality assurance is focused.
– Form/activity: it may be a distribution of the trace with a specific position—following a sequence, within a certain period of time, with a specific physical distribution—that provides information about the transfer and the persistence or the activity that took place.
– Frame: it may only be inferred from source and form that set up the limits within which

a reconstruction or understanding of the reasons for a current state is forming. This leads to hypothesize about the crime(s) and sets up a kind of window of opportunity.

– Aim/motive: considers the motive or a crime/event, which may influence both form and frame and may be seen as a result, which gives undue advantage to a party through fraudulent means.

Frame and motive may feed intelligence and need logical reasoning to integrate some form of reconstruction and linkage in organized memories. Form relies also on logical and transparent reasoning in the association of traces to the events. Determination of source/composition needs specific technical reasoning in relation to qualitative and quantitative analyses. In summary, everything is based on one basic object, the remnant of criminal activity in the form of a physical object: the trace, although it leads to a complex phase of reasoning and professional decisions.

Source/composition requires discrimination and sensitive analytical tools; but grouping in relevant classifications in order to have elements of a same class within the class and elements from different classes separated may often be more important than traditional analytical performances. For example, is it better to be able to discriminate each drug seizure or each counterfeit from each other, or is it better to determine that a whole market is dominated by a single mode of production that may be from one source? In most cases, the answer to the second question has much more strategic and operational dimensions needed in an inquiry than in the first option. This is a measure of intra- and inter-variability, which has a greater impact on the potential meaning of a trace than knowing the limit of detection (LOD) and limit of quantification (LOQ) of techniques that focus on analytical tools detailed in standard procedures.

Milestone #1. The trace is the primitive source of information available to forensic science. Found at the scene, it is the core of the forensic science practice. It finds an interesting parallel with Ginzburg's paradigm of "clues" viewed from an historian's perspective. The trace contributes to intelligence-led and evidence-led dimensions fundamental to resolving or understanding crime phenomena. Comprehension of the logical process information goes through—from the trace to the evidence/intelligence—should be an essential step toward scientific quality assurance. The cultural background of this process lies in its specific rules.

Fundamental Principles of Forensic Science

Locard's and Kirk's Exchange Principles

> There exists in the field of criminalistics a serious deficiency in basic theory and principles, as contrasted with the large assortment of effective technical procedures. (Kirk 1963)

Even in the days of undisputed DNA evidence, this quote finds its echo in the identified lack of sound scientific methodology in the application of DNA according to the latest Campbell systematic reviews (2011) and serious flaws detected in the way of reporting even within accredited laboratories. Improving procedures through standardization processes or pushing forward, selectivity, limits of detection, and quantification are the main focus of scientists. As long as the focus will remain on technical features, the scientific quality will suffer from deficient outcomes.

The well-known principles of Locard and Kirk constitute two fundamental and specific cornerstones of forensic science. Most inferences can be based on these fundamental principles using other established sciences such as chemistry (composition, structure), physics (dynamics, laws of nature), and biology (heredity, metabolism). Although attributed to Locard and Kirk, these are axiomatic dimensions of nature that a contact will in some way be the cause of some transfer of a physical nature that constitutes a kind of signature of this contact and that the shared uniqueness of the trace and its source may ultimately lead to the individualization of the trace to a specific source. There is a vital need to investigate the validity of Locard's and Kirk's principles, considered as forensic science's founding principles. Inseparable from its culture, they remain part of its foundations.

Interestingly, Crispino's recent contributions brought valuable scientific arguments to the critical investigation of those organizing principles (Crispino et al. 2011). He stated that:

1. Kirks' principle of individuality (Kirk 1974) can be considered as being a scientific principle upon which forensic science practices lie everyday through the comparison of entities, since it fulfills Popper's falsifiability criterion. Briefly, the whole point of comparison in forensic science is to ascertain the degree of concordances or discordances of measurable properties that might be found between two entities coming from allegedly the same source. Such process can be seen as an attempt to falsify the philosophical principle (Kirk's) stating the uniqueness of every object, making this principle scientifically acceptable according to the Popperian logic.
2. Locard's principle of exchange (Locard 1920) is introduced as the basis of forensic science. It refers to the trace as representative of the identity of a source that is the object of transfer during the questioned activity and that will be submitted to the effect of time. The whole point within forensic science is to be able to trace back the source or the activity, but the process comes with all the uncertainties related to the fact that causes remain unknown. This is where reasoning through probabilistic inferences (abduction to deduction) comes into play. Methodological solutions, such as proposed by Cook et al. (1999), with the CAI model, help focus on interpretation difficulties and solutions for forensic science practitioners. Crispino demonstrates that Locard's exchange principle is scientific because it is at the basis of forensic science reasoning. It is also corroborated pragmatically over centuries through the experiences of man as a hunter (cynegetic), where the reasoning process lies upon inferences drawn from observations of traces resulting from exchanges between animals and their environment. Investigating the scientific nature of Locard's principle comes back to questioning the scientific nature of forensic science, an epistemological approach that may prove very effective to strengthen forensic science.

The two principles are part of the forensic science paradigm, "both have the ability to be accepted as scientific laws. They add knowledge that can be measured and used in logic for the sake of the law (if not justice)" (Crispino et al. 2011). However, another fundamental issue for forensic practitioners is whether a detected trace is relevant, and this constitutes the essence of the forensic quest.

Relevancy and the Use of a Proper Terminology

The forensic scientist is not a witness of the events under investigation and can never be certain to observe, detect, and see relevant traces for collection and analyses; however, this is the challenging goal. From crime scenes to laboratories, working with relevant physical traces is a leitmotiv and belongs to the thought processes fed by Locard's and Kirk's principles. Relevancy can be seen as a basic tenet describing the everyday life and doubts of forensic science practitioners. Currently it is mainly perceived according to criminal laws. The concept of relevancy is a key issue in a legal perspective and is understood as the "the adequacy between the fact that has to be proven and the brought evidence or between an allegation and the law rule to apply" (Le Grand Larousse Universel 1995). Though it is a requested condition for evidence to be admitted in court, the legal concept is not well defined (Twining 2006). Understood differently depending on the legal structure and proceedings, it should not stand for the forensic reference, since it does not cover the essential discovery step involved at the onset of an investigation that should be part of forensic science practice.

Although preliminary, the concept of relevancy is not clearly defined, but implicit within the reasoning process in forensic science. At the scene of investigation, there is a need to take decisions as to what to collect and then within the laboratory as to what sequences of methods have to be followed. When structuring databases and memories or preparing expert's reports for

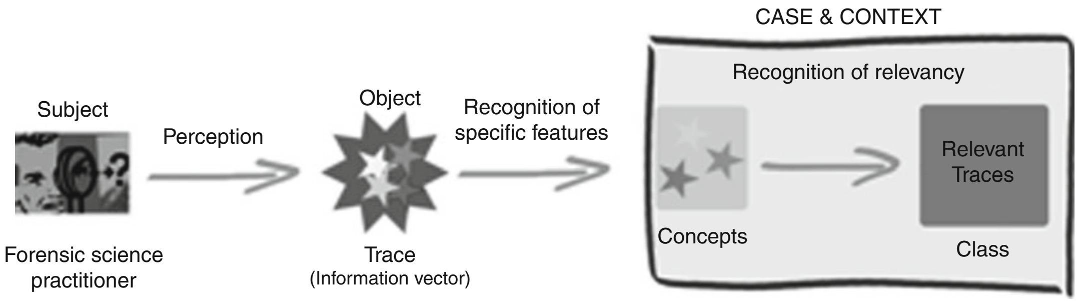

Forensic Science Culture, Fig. 1 Semiotic approach of trace objects when forensic science practitioners process the investigation scene (Hazard©)

intelligence or evidence, further decisions have to be made, and this is based on a sound appreciation of relevancy. The reasoning process can only be sound if pieces of the arguments are relevant.

Focusing on the relevancy concept in forensic science goes far beyond a simple question of definition. Precisely, Inman and Rudin (2000) did formalize the current conundrum that criminalists have to deal with: the most difficult challenge is *the recognition of relevant physical evidence*, although it may be questionable whether the capacity to recognize objects as clues would not have limits. Those limits fundamentally rely on the case-by-case approach: crime scenes are consecutive but not alike, being peculiar to every criminal activity, where resources are always limited (whether material, time, etc.). This forces forensic science practitioners to adapt to places and cases in order to find what is relevant.

A relevant trace carries information in relation to the event at its source and may be studied through basic logical relations (trace to source, source to trace, trace to trace, source to source, geographical, temporal relationships, etc.). The focus on the trace and the relevancy of meaning in relation to events is a semiotic quest that has to be studied.

The Recognition of Relevant Traces

> "L'oeil ne voit dans les choses que ce qu'il y regarde. Et il ne regarde que ce qui est déjà dans l'esprit" (i.e., "The eye only sees in things what it looks at. And it only looks at what it has in mind"), famous quote used by Bertillon that is attributed to the French journalist L. Peisse

Semiotics is the science that studies the sign—the very first act of communication—and its various forms; it questions and describes the mechanism of signification. This is a complex domain of research and study in both linguistics and medical diagnosis but has a direct importance in forensic science practice, since giving a meaning to observations is part of the process: whenever a relevant trace/object is recognized and perceived as such, a signification is attributed.

The semiotic standpoint of relevancy sets forth the importance of the processes of perception and recognition, which are conditioned by the practice *in regard to which the subject considers the object* in a given context (Sebeok 1986). The concept of relevancy is not an intrinsic quality; it only exists in a specific context (Fig. 1).

The semiotic key terms used in the previous figure are defined as such:

– A *class* consists of a set of equivalent objects regarding a specific use, i.e., the *class* of interest is the *relevant traces* that consists of any kind of traces that might be relevant to the investigation and/or intelligence process.
– A *concept* is a set of features that is specific to a *class*; *these features have to be recognized in the object to classify it*. The forensic science practitioner may perceive explicit or implicit features (e.g., *type of traces*, *quality*, *spatio-temporal localization*, *kind of information given once the trace object is used*) and recognize the concept they belong to, namely, the concept of source and action that are linked to the criminal event in this situation.

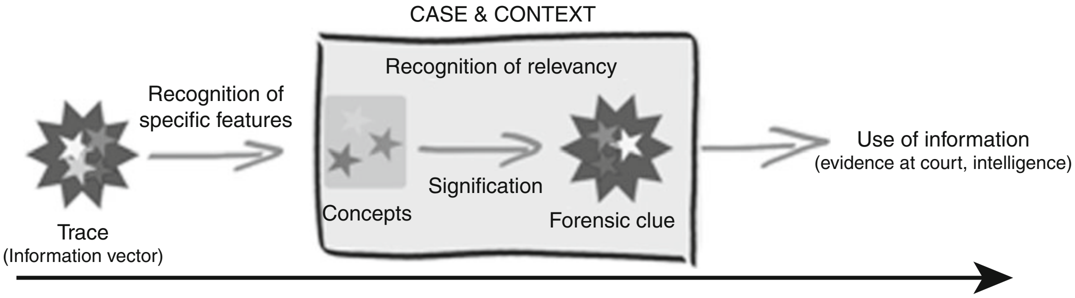

Forensic Science Culture, Fig. 2 Trace, clue, and relevancy: the following process from the scene to the beneficiaries (Hazard©)

– If a *concept* is recognized in an object, the latter becomes a part of the *class*, i.e., a trace object that realizes the *criminal source/action concepts* of the *class relevant traces* belongs to this *class*.

In summary, according to the semiotic view, relevancy is defined as a perception of trace objects conditioned by the *context* and by what the *forensic science practitioner* decides to *recognize* at investigative scenes and to use as features from (*relevant*) traces (i.e., *potential for discrimination*, *for case linkage*) found at the scene of investigation. These signification processes are ruled by cultural conventions; the whole reasoning is based upon the practice and the context. Relevancy is a conventional dimension specific to the practice the forensic practitioner evolves in; such a fact shows how important it is to understand what forensic science culture is and what it is made of, since it conditions forensic science practitioners' work.

Trace, Clue, and Evidence

Interestingly, the discovered physical trace is often called physical "evidence," whether at scene, laboratory, or court, although there is a huge difference in reality between these three settings. It is not because a trace has been discovered that it can be automatically classed as evidence: an indication of direction may or may not be relevant information. This information may be useful in directing search and in providing indications about frame, but it is not evidence (in a legal sense) until it is recognized in court to be an element on which the decision on the issue is made (judgment). Thus, evidence is solely the province of the court.

A trace exists in itself and does not have a meaning initially (although it can be measured), except that it is perceived as a support with an unexploited potential of information that might explain issues in the investigated cases. Once this potential is recognized, it is considered as a sign that potentially pertains to the class of *relevant traces*. The metric associated to the meaning gives some strength in view of probable causes. This measure, which may be expressed as the likelihood ratio in Bayes' theorem (see **Probability and Inference in Forensic Science**), may be a strong indicator or clue of something insofar as it is interpreted in a specific context. The value of this clue in deciding on the issue by a judge is the evidential value (Fig. 2).

As illustrated, the belief in the information content is built up between the discovered trace and the final outcome through measurements and a reasoning process. Therefore, the distinction of a trace from a clue and evidence are as follows:

– Trace and clue are two distinct states of a same object whose distinctive parameter is the recognition that the information it carries is relevant.
– Clue is therefore a trace with recognized and measured information content.
– Clues are gathered and constitute structured information about the case in issue. This helps the decision process (i.e., evidential process). Evidence consists of pieces of information that have been accepted to be relevant and useful in the decision process by a judge.

The study of relevancy is a key point between these entities and the understanding of it is an essentially cultural cornerstone of forensic science practice. Questioning the issue of relevancy in the forensic science process might be considered as a safeguard: it involves a meaning-making process where decisions and actions depend strongly upon the assessment of the relevancy of the features expected to be recognized as such. Significations come from the dynamic process of inferences set up by practitioners with regard to their personal stock-in-trade (knowledge, education-training, experience, called KEE model). Understanding this concept may help selection, training, and education of forensic scientists.

Milestone #2. Sound reflections regarding forensic science foundations enable it to be properly defined: "[...] a science on its own, defined as the science of identifying and associating traces for investigative and security purposes, based on its fundamental principles and the case assessment and interpretation that follows with its specific and relevant mode of inference" (Crispino et al. 2011). The study of principles is aimed to strengthen forensic science foundations and culture. Mastering the concepts and principles pertaining to forensic science is a way to ensure scientific quality.

Sound Education and Training for a Professional Judgment of Quality

The Roles of a Forensic Scientist

A real culture of the physical trace exists and is at the core of forensic science with two principles (Locard, Kirk) and the relevancy. The search, detection, recognition, and collection of relevant traces call for specific abilities for forensic scientists. Kirk (1974) described features of the forensic science practitioner's work: "The investigator...must understand (a) what physical evidence is [sic]; (b) how to collect and preserve it; (c) how to obtain from it the information it carries; and (d) how to interpret the information so obtained." This statement assumes some corollaries:

1. To know and master the way a physical trace is created, which calls for a sound understanding of transfer, persistence, selectivity through teaching, and experimental research, using common sense and logic.
2. To recognize the relevant trace with its potential of information (whether at the scene, the laboratory, in any kind of information process), which calls for a sound culture about the trace and the environment (recognizing situations as defined in criminological studies).
3. To be able to evaluate the discovered trace (to read the signs), which calls for a comprehensive vision of the forensic science process and the part it may play within the CJS. It requires a sound scientific methodology of interpretation.

The whole approach is based on cognitive processes such as experience and culture of the trace acquired through education and training but is also based on the thoughtful use of available information related to the case and the current series identified as crime phenomena (e.g., through forensic intelligence). The work is therefore not limited to running methods and techniques; it goes far beyond. As simple as it seems, a doctor is not good because of the quality of her tests (even if this is important), but because of her cognitive abilities used through sound observations of signs or symptoms and the application of a logic that leads to an accurate diagnosis; this statement is also true to qualify a good forensic scientist.

Emphasis is mostly directed to technological progress and implementation in many forensic science laboratories rather than promoting forensic science culture and development through education and training. New technologies are important to master but remain tools, even useful to provide answers to preestablished questions according to the given problem (DeForest 1997). Education has to focus on means to foster relevant questioning when forensic scientists are faced with a (crime) problem.

Surprisingly, in most countries, people with a primary degree in forensic science are not the first choice for employers. Brown and

Willis (2009) describe the typical route followed to become a forensic scientist within laboratories in the UK and Ireland. The preference is given to scientists with qualifications in another discipline rather than forensic science, who receive in-house training once hired. Despite a substantial number of university programs offering forensic science degrees popping up throughout the UK, but also elsewhere, such a statement is worrying about the quality of the forensic science education. It is reasonable to think that some of the forensic science programs do not educate students in a fit for purpose way to the reality of the field and in agreement with a forensic science culture, but responding to marketing interests instead (Lawless 2011), trying to attract students and associated financial rewards. There is a need for a sound and serious reappraisal of the way education of forensic science is conceived, and academic programs should be implemented considering a complete process of reasoning that starts from the perception of the trace at scenes to its interpretation as evidence/intelligence, a stage that represents a serious issue (currently the most controversial one).

Interestingly, the observed deficiencies in giving reliable and quality results of significance in forensic science laboratories participating in collaborative testing, even in the perceived "gold standard of DNA," is proving deficient as far as interpretation of forensic data is concerned. If such a discrepancy within results emanates from a lack of solid (academic) interpretation bases, it also shows that hired scientists may master technical skills but are not educated and trained to handle forensic issues.

Addressing the education-training issue is not an easy matter and requires a definition of roles and functions. A forensic scientist may be an investigator (who infers from direct observations based on discovered physical traces assessed as being relevant), an evaluator (who helps partners to make inferences with regard to the case and provide answers) (Jackson et al. 2006), or an intelligence provider (who provides information that will reduce linkage blindness).

The forensic scientist may take the different roles or specialize depending on organizational decisions. All roles must be emphasized and integrated into the education's vision in order to provide a foundation and a vision about the complexity of the forensic science process. It is, in all its actual dimensions, a true problem-solving process applied by the same or different people with a view to solve a problem (not to follow a process).

The Role of Education Programs

> "First we need to note that the educational field that underpins forensic science is heterogeneous and confusing." (Willis 2009)

Forensic science communities have, in some way, started to undertake this reflection, though organizations and structures are just as heterogeneous and confusing, perhaps because of an historical evolution that contributed to the current fragmented vision. Three visions coexist nowadays and find their roots with pioneers:

1. Gross—a lawyer—introduced the use of scientific methods as the basic methodology for an investigative magistrate. He called "Kriminalistik" the single basic methodology of criminal investigation. This became the French "criminalistique" used by Locard or the North American "criminalistics." The influence of Gross is still felt throughout Central and Eastern Europe and has seen the development of forensic science laboratories within justice departments.
2. Bertillon, an administrative assistant, introduced structured databases to identify recidivists within police structures. The methodological approach was developed from the 1880s, and its investigative methods became technical and identification-oriented at the base of technical police and identification bureaus.
3. Finally, the third approach was more about expertise with a method developed to express experts' opinions within specialized laboratories, which was emphasized by Locard (Crispino et al. 2011). Locard was a trained forensic medical doctor who directed his vision outside and complementary to medicine. Although he created what is now considered the first "police forensic science laboratory" (founded in 1910) in Lyon,

he was adamant that this should be within the justice department, and his first laboratories were within the Lyon tribunal!

Such fragmented evolution is reflected in the current education and training programs. The logic according to which the clue development relies on trace detection is not currently leading the scientific methodology and technologies (Crispino et al. 2011). The current education and CJS emphasizes efforts on an incomplete role perceived through the eyes of government and justice policies. Focusing on the trace and the science itself to reconstruct the whole puzzle might be a sound thread to follow in developing education programs.

Education is meant for developing critical scientific reasoning relying on a sound scientific basis, while training represents the acquisition of the command for specific tasks. Barnett (2000) emphasized that education programs should achieve three main goals by developing:

- Forensic science professors, to teach the forensic science "philosophy" with the methodology and logical reasoning process standing behind the words "forensic science." The knowledge specific to the discipline has to be fostered and transmitted through a more holistic vision.
- Practitioners dedicated to research, to improve and strengthen the science by investigating foundations of the science as well as leading research with a more pragmatic concern for the field.
- The core of the forensic science profession with practitioners who apply the developed knowledge and are expected to undergo further in-house training and continuing education.

To succeed in such an enterprise, relations between academia and practice have to be strengthened in order to establish a valuable and effective flow of knowledge (Willis 2009) and develop sound professional judgment to assess whether a particular piece of data is relevant.

Milestone #3. Education and training frameworks have to foster a holistic vision of forensic science as a science based on the trace, convey the culture of the relevant questioning, and emphasize the multiple roles of the forensic scientist within the CJS.

Only education, training, and research may guarantee quality and relevancy consistently within a forensic science profession.

An End to Think About

As a meaning-making process that defines the culture of forensic science is the one transforming traces into information and then into (forensic science) knowledge (see Fig. 3).

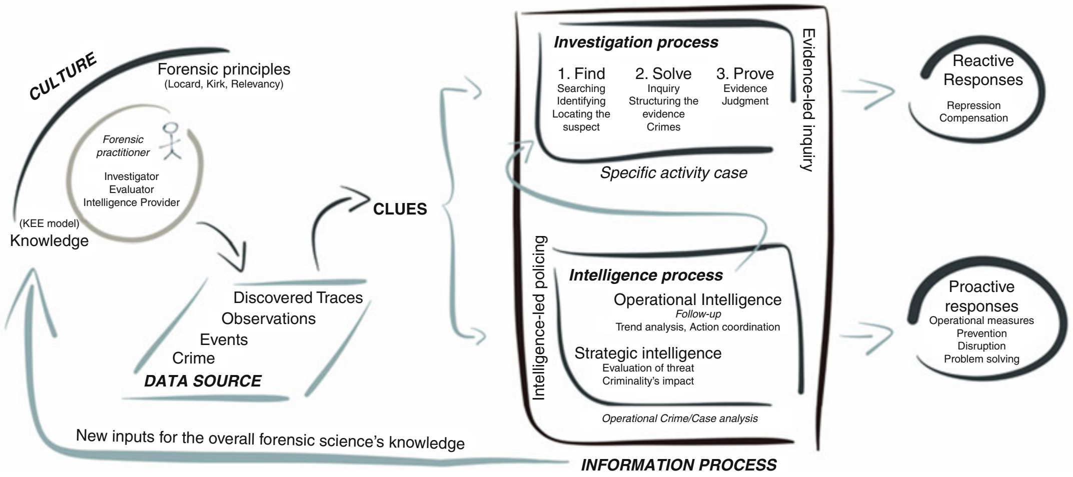
CULTURE
Forensic principles
(Locard, Kirk, Relevancy)
Forensic practitioner
Investigator
Evaluator
Intelligence Provider
(KEE model)
Knowledge
CLUES
Discovered Traces
Observations
Events
Crime
DATA SOURCE
New inputs for the overall forensic science's knowledge
Investigation process
1. Find
Searching
Identifying
Locating the suspect
2. Solve
Inquiry
Structuring the evidence
Crimes
3. Prove
Evidence
Judgment
Specific activity case
Evidence-led inquiry
Intelligence-led policing
Intelligence process
Operational Intelligence
Follow-up
Trend analysis, Action coordination
Strategic intelligence
Evaluation of threat
Criminality's impact
Operational Crime/Case analysis
INFORMATION PROCESS
Reactive Responses
Repression
Compensation
Proactive responses
Operational measures
Prevention
Disruption
Problem solving

Forensic Science Culture, Fig. 3 Forensic science as part of the overall security system (adapted from Ribaux and Margot 2008) (Hazard ©)

Think forensic consists of a thought methodology that lies on games of drawn inferences, which are based on critical observations of details (discovered traces), a sound scientific knowledge, a forensic science culture, and a capacity for adjustments. Details are made to "speak" and are used as clues for intelligence and/or evidence purposes. The latter are not mutually exclusive since the follow-up of information, part of the intelligence process, can provide support for the evidence-led inquiry or new inputs.

The forensic practitioner, who undertakes various roles, contributes to two processes, which diverge as to the effect of their outputs albeit they are part of the same system: the evidence-led inquiry provides reactive measures, like a response to a current case, while the intelligence-led policing allows proactive ones in terms of implementation of prevention, security, and operational matters.

Finally, ensuring scientific quality depends on the ability to recognize the impact and the limits of the forensic science knowledge and forensic scientists duty in the overall system, illustrated in this figure; by showing a comprehensive picture of their influence on the whole system, it enables to ask relevant questions that will improve the quality of the process. Such a conception of the forensic science culture fostered through an appropriate education and training will provide sound basis for forensic scientists, whose professional judgment would not be questioned as long as it relies on this basis, allowing the development of a sound trust in forensic scientists "expertise."

Related Entries

- ► Cognitive Forensics: Human Cognition, Contextual Information, and Bias
- ► Forensic Science and Criminal Investigation
- ► Forensic Science and the Paradigm of Quality
- ► Forensic Science Effectiveness
- ► Probability and Inference in Forensic Science

Recommended Reading and References

Barnett P (2000) The role of forensic science professional organisations in the new millennium of accreditation, certification, registration and standardization. Sci Justice 40(2):138–142

Bonfantini MA, Proni G (1988) To guess or not to guess? In: Dupin, Holmes, Peirce the sign of three. First Midland book edition. Bloomington, Indiana University Press

Brown S, Willis SM (2009) Complexity in forensic science. Forensic Sci Policy Manag: An Int J 1(4):192–198

Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA (1999) Case pre-assessment and review in a two-way transfer case. Sci Justice 39(2):103–111

Crispino F, Ribaux O, Houck MM, Margot P (2011) Forensic science—a true science? Aust J Forensic Sci 43(2):157–176

DeForest PR (1997) The 90th semi-annual seminar of the California association of criminalists. Recapturing the essence of criminalistics. Science & Justice, 39 (3):196–208

Dulong R (2004) La rationalité spécifique de la police technique. Rev Intern Crim Pol Techn Sci 3:259–278

Ginzburg C (1989) Traces. Racines d'un paradigme indiciaire. In: Mythes, emblèmes, traces. Morphologie et histoire. Flammarion, Paris, pp. 139–180

Girod A, Champod C, Ribaux O (2008) Traces de souliers. Presses Polytechniques et Universitaires Romandes, Lausanne

Gross H (1899) Manuel pratique d'instruction judiciaire à l'usage des procureurs et des juges d'instruction. Vol tome 1. Marchal & Billard, Paris

Inman K, Rudin N (2000) Principles and practice of criminalistics: the profession of forensic science. CRC Press LLC, Boca Raton

Jackson G, Jones S, Booth G, Champod C, Evett IW (2006) The nature of forensic science opinion—a possible framework to guide thinking and practice in investigations and in court proceedings. Sci Justice 46(1):33–44

Julian RD, Kelty SF, Roux C, Woodman P, Robertson J, Davey A, Hayes R, Margot P, Ross A, Sibly H, White R (2011) What is the value of forensic science? An overview of the effectiveness of forensic science in the Australian criminal justice system project. Aust J Forensic Sci 43(4):217–229

Kind SS (1987) The scientific investigation of crime. Forensic Science Society, Harrogate

Kirk PL (1963) The ontogeny of criminalistics. J Crim Law, Criminol Police Sci 54(2):235–238

Kirk PL (1974) Crime investigation, 2nd edn. Wiley, New York

Lawless C (2011) The contested shaping of neoliberal forensic science. Br J Criminol 51:671–689

Le Grand Larousse Universel (1995) Paris, Larousse

Locard E (1920) L'enquête criminelle et les méthodes scientifiques. E. Flammarion, Paris

Margot P (2011a) Forensic science on trial—what is the law of the land? Aust J Forensic Sci 2–3:89–103

Philip's (2008) Culture, World encyclopedia. Oxford Reference Online. Oxford University Press. University of Lausanne. 14 March 2012. http://www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t142.e2999

Ribaux O, Margot P (2008) La trace matérielle, vecteur d'information au service du renseignement. In: Cusson M, Dupont B, Lemieux F (eds) Traité de sécurité intérieure. Presses Polytechniques et Universitaires Romandes, Lausanne, pp 300–321

Ribaux O, Baylon A, Roux C, Delémont O, Lock E, Zingg C, Margot P (2010) Intelligence-led crime scene processing. Part I: forensic intelligence. Forensic Sci Int 195(1–3):10–16

Sebeok TA, Bouissac P, Umberto E, Pelc J, Posner R, Rey A, Shukman A (eds) (1986) Encyclopedic dictionary of semiotics (approaches to semiotics), vol. 2. Mouton de Gruyter, Berlin

Spillman L (2007) Culture. Blackwell encyclopedia of sociology. Blackwell Reference Online. http://www.blackwellreference.com/subscriber/tocnode?id=g9781405124331_chunk_g97814051243319_ss1-183, last visited 14th March 2012

Twining W (2006) Rethinking evidence. Exploratory essays, 2nd edn. University College London, UK, Law in Context

Willis SM (2009) Forensic science, ethics and criminal justice. In: Fraser J, Williams R (eds) Handbook of forensic science. Willan Publishing Ltd, Cullompton

Wilson DB, Weisburd D, McClure D (2011) Use of DNA testing in police investigative work for increasing offender identification, arrest, conviction and case clearance. Rapport 7. Campbell Syst Rev

Forensic Science Effectiveness

Claude Roux[1], Roberta Julian[2], Sally Kelty[2] and Olivier Ribaux[3]

[1]University of Technology, Sydney, Centre for Forensic Science, Broadway, NSW, Australia

[2]Tasmanian Institute of Law Enforcement Studies, University of Tasmania, Tasmania, Australia

[3]Ecole des Sciences Criminelles, University of Lausanne, Lausanne, Switzerland

Overview

This entry reviews the state of the art in the area of effectiveness of forensic science and ultimately attempts to address the central question: What do we actually know about the impact of forensic science on the criminal justice system in general and on criminal investigations in particular? It is concluded that, although research to date has shown interesting trends, there is a need to better understand and explain the broader contribution of forensic science, as a precondition to accounting for its effectiveness. Otherwise, conclusions drawn from limited studies may have unintended effects for the whole system. Research outcomes to date probably reveal more about the need for methodological improvements in how to undertake such research than providing a clear answer with respect to the effectiveness of forensic science. More research is needed.

Introduction

Over the last decade, forensic science has been heavily present in popular culture through a number of TV shows, fiction or documentary. It has been argued that this situation led to increased public awareness of, and interest in, forensic science. There is a widespread belief that science and technology have become a "magic bullet" against crime, partly fuelled by the media continuously reporting forensic science successes, in particular those obtained through DNA analysis. Not only is such analysis often described as having "revolutionized" the ability of law enforcement to protect the innocent from wrongful prosecution (Innocence Project 2012), but it is also quoted as having enhanced the capability of police to identify, arrest, and prosecute offenders. For example, in a number of countries, "cold cases" have been revisited, taking advantage of modern technologies, primarily DNA profiling and novel fingermark detection methods. This situation has reinforced the perception of the infallibility of science in solving crime.

Contrary to this view, recent studies have questioned the effectiveness of forensic science on actual crime solving and justice outcomes (Baskin and Sommers 2010; Brodeur 2010).

For example, the effectiveness of DNA testing was found to be questionable when applied to serious or violent crimes (Wilson et al. 2010) unlike in volume crimes such as break and enters. It also appears that the perception of scientists about their own contribution is not uniform; and such questions do not even seem to be of major interest to laboratory scientists (Crispino 2008). In fact, current knowledge regarding the application of forensic science, particularly its effectiveness in criminal investigations and judicial outcomes, is limited. Further, as quoted by Julian et al. (2011), "*to a large extent, the policing and forensic services community has been 'flying blind' in terms of the true impact of its work.*"

These discrepancies may be symptomatic of the fact that the contribution of forensic science is not straightforward to assess because it occurs on a number of different dimensions or levels depending on the context of the judicial response or police activity. In addition, the judicial response itself is not linear nor continuous but can be seen as successive phases with different aims, distinct reasoning, and distinct ways of using information (Delémont et al. 2013). This entry reviews the state of the art in this area and ultimately attempts to address the central question: What do we actually know about the impact of forensic science on the criminal justice system in general and on criminal investigations in particular?

Efficiency or Effectiveness?

The terms "efficiency" and "effectiveness" are often used interchangeably in attempts to assess the value and/or impact of forensic science in the criminal justice system. However, the terms have distinct meanings and should be clarified if being applied for the purpose of impact assessment. A lack of clarity when using these terms creates confusion when undertaking research and leads to difficulties in identifying any trends in the value of forensic science in the criminal justice system. In their research, Kelty and Julian (2012a) found that stakeholders attached five different meanings to the terms "effectiveness of forensic science" and how it could be better achieved:

- Higher quality science
- Broadening the role of forensic science
- Reflection on, and improvements to, the current process
- Having an influence on justice outcomes
- Ensuring a "bang for your buck"

It is the latter, with its emphasis on cost-effectiveness, that is meant by the term "efficiency." The meaning of the term "effectiveness" is broader, can be operationalized in a number of ways, and may often encompass "efficiency." Since the studies we discuss below assess the value of forensic science in a variety of ways, we will adopt the term "effectiveness" throughout, unless the research specifically incorporates a financial component in its assessment.

This entry reviews current knowledge on the effectiveness of forensic science in the criminal justice system by focusing on the stages of the forensic process: crime scene, triage, laboratory processes, databases, and from identification to charges and to court. This is followed by a review of studies that examine the whole process from crime to court. This entry concludes with a brief discussion of future directions in research that is attempting to measure the value and/or impact of forensic science.

Crime Scene

It is often quoted that good forensic science starts at the crime scene. Many well-publicized miscarriage of justice cases find their root in poor crime scene work. In general, crime scene processing cannot be reproduced. It predetermines the quality and quantity of the information available to the investigation and ultimately presented in court. From this angle, it is impossible to discuss the topic of effectiveness in forensic science without a closer examination of what is happening at the crime scene (Julian et al. 2012).

Despite the paramount importance of this work, it is interesting to note that, from an organizational standpoint, crime scene models vary greatly depending on the type of crime and across jurisdictions. Laboratory scientists often attend specific scenes requiring specialized capabilities. Examples include disaster victim identification cases, mass murders, and fire scenes. However, in most countries, the vast majority of crime scenes are attended by police personnel. And in many jurisdictions, a further distinction is made between volume crime (e.g., break and enter) and serious crime (e.g., homicide). The personnel attending the latter have access to more resources, including the latest technology, compared to those attending volume crime. The status of crime scene personnel also varies greatly: sworn police officer with minimal training in forensic science, fully trained sworn police officer, civilian with a science degree, etc.

Beyond these very general trends, police models are variable, and it is difficult to identify what the dominant model is (Tilley and Ford 1996; Bradbury and Feist 2005; Adderley et al. 2007; Girod et al. 2008; Roman et al. 2008). Regional or local structures and customs usually dictate the most appropriate implementation model. The time spent at the crime scene, the types of crime attended, or even the type of trace that is collected may sometimes be imposed by the investigation process in practice within a specific police force. Other peculiarities such as the size of the police force and the local topography are also factors to be considered. The following observations can be considered as symptoms of an apparent lack of harmonization between organizations:

- The number of crime scenes attended by crime scene examiners per annum can vary greatly depending on the crime scene unit. For example, Tilley and Ford (1996) noted a variation between 400 and 1,350 in their English study.
- The quantity and quality of traces collected at the scene can also vary greatly depending on the region or crime scene unit. For example, Girod et al. (2008) showed that shoemarks are collected in 60 % of domestic break and enter cases in a region and in only 10 % of the cases in another region in Switzerland. This picture converges with results obtained by Rix (2004) in England.

More puzzling is the variation which may exist within one crime scene unit. In his study on shoemarks, Girod (2002) showed that the collection rate could double within the same unit, depending on the individual crime scene examiner. Further, Adderley et al. (2007) found that some crime scene examiners are more prone to collect some types of traces than other types and that high performers collected higher quality evidence, attended more scenes, and submitted evidence more quickly, and the evidence they collected was more likely to lead to positive identifications. It appears that, in addition to individual knowledge, experience, and motivation, the quantity and quality of traces collected at the scene also depends on the quality of the relationship between the crime scene examiner and the investigator (Bradbury and Feist 2005; Adderley et al. 2007; Crispino 2008; Girod et al. 2008; Kelty et al. 2012b).

The trend identified by these studies is that crime scene work appears to encompass a major "human factor," which makes any study on its effectiveness at the team or unit level difficult to interpret or generalize. To this end, in their Australian study, Kelty et al. (2012b) attempted to profile "top-performing" crime scene examiners. They found that such examiners shared a number of qualities within the following seven key skill set categories: knowledge base, experience, work orientation, approach to life, communication, professional demeanor, and cognitive abilities.

Overall, forensic science at the crime scene shows significant variations in terms of quality and quantity of the work undertaken. It appears that there is no universally accepted role for forensic science at this stage, and, in turn, this role depends on the historical development of the hosting organization. Further, these variations cannot solely be explained by differences of organizational setting and/or personnel status, and dimensions that are individual to the crime scene examiner have to be factored in. It is fair to conclude that, with regard to the paramount significance of the work undertaken at the crime scene, we only have limited empirical data to assess the true effectiveness of forensic science in this area and further research is strongly needed.

Triage

Once the crime scene is completed, specimens are generally sorted, and a decision is taken as to whether they must be considered further and forwarded to a forensic science laboratory. This step, called triage, is essential for financial reasons and also to focus the forensic examination on relevant traces, that is, those traces that are more likely to bring valuable information to the investigation and intelligence or are the more promising in terms of evidence. Efficient triage is therefore necessary to avoid flooding the laboratory with unnecessary items which contribute to laboratory backlogs.

Delays in providing laboratory results strongly diminish the relevance of potential links or "matches." In homicide cases, laboratory results are often provided once the case is solved (Schroeder and White 2009). In some cases, traces are not sent to the laboratory because the investigator has not identified a suspect or assumes that results will not be obtained in time for the investigation to progress satisfactorily (Strom and Hickman 2010).

Interestingly, despite its crucial role, triage has been rarely studied. Some trends however exist. According to Bradbury and Feist (2005), the selection is based on the case urgency, whether the specimen is degradable and how serious the case is (e.g., violent crimes vs. property crime). Specimens coming from cases where a suspect is known or in cases where there is a direct judicial request tend to be selected and analyzed as a priority. Scientific aspects are also considered, for example, the probability of extracting a usable DNA profile from a given substrate and/or from a degraded specimen (Raymond et al. 2009). It becomes evident therefore that the best model for an efficient triage should integrate police (investigation) and scientific (laboratory) considerations with a strong focus on the trace and its relevance. However, it appears that, in reality, the triage function largely remains reactive in response to judicial pressures and does not follow a proactive strategy aimed at solving a problem. It is believed that sound triage management would go a long way to addressing the challenge of laboratory backlog, as well as providing value-added information to the criminal justice system.

Laboratory Processes

Once traces have been sorted, a number of exhibits and/or specimens are generally forwarded to one or several well-equipped forensic science laboratory or laboratories or otherwise examined using scientific equipment. This stage of the process, compared to crime scene, appears more codified and more in line with the work undertaken by traditional (i.e., non-forensic) scientific laboratories. However, does it mean that the work is more effective? Or even that its effectiveness can be more easily and accurately assessed? The studies reported so far give a rather confusing picture. However, it seems relevant to distinguish the results obtained in serious crime vs. volume crime.

Serious Crime

In their studies based on homicide cases, Mucchielli (2006) and Brodeur (2010) provided a rather skeptical picture with respect to the effect of modern technology on case resolution:

- In two thirds of the cases going to court, the perpetrator was arrested because of his or her behavior, and the enquiry did not contain elements of identification or localization (Mucchielli 2006).
- The trace found at the scene was crucial in charging a suspect in only 7 % of the cases (Mucchielli 2006).
- 71 % of homicides were solved within 24 h without the need to exploit traces collected at the scene (Brodeur 2010).
- DNA enabled the perpetrator identification in only 1 % of the cases Brodeur (2010) studied.

These results are supported by other studies which showed the minor contribution of DNA in solving homicides (Schroeder and White 2009) and the fact that serial murderers were primarily identified by means other than forensic traces, for example, traditional eyewitness or victim statements and confessions. However, Bradbury and Feist 2005 showed that, in serious crimes (e.g., homicide), traces are mostly useful to solve difficult cases. In other words, the potential for a positive contribution would exist in only 10–20 % of homicides approximately. This could explain the limited apparent contribution of forensic science in the studies described above.

Volume Crime

There is only a small number of studies focusing on volume crime (e.g., break and enter, vehicle thefts). It is also difficult to make generalizations. However, the trend regarding the impact of forensic science appears to be more positive than the results obtained considering serious crime. In particular it appears that a systematic use of DNA in such cases has a significant effect on solving such cases (Roman et al. 2008; Wilson et al. 2010).

Roman and coworkers compared, between 2005 and 2007 and across five police forces, cases where a DNA specimen was sent to the laboratory for subsequent search in a DNA database with similar cases where no specimen was analyzed and only a traditional method of investigation was applied. Using DNA led to more than double the number of suspects being identified, twice the number of suspects being arrested, and twice the number of suspects being accused, compared to the control group where no DNA was used. Multirecidivists were also in higher number in the group using DNA than in the group using traditional methods. This indicates that these methods of investigations are complementary as DNA databases and traditional methods tend to identify different suspects.

Wilson and coworkers included the study by Roman et al. in their systematic review of field studies on the effectiveness of DNA testing. This review included:

- A study that had examined the impact of a DNA database on rates at which crimes are cleared and/or offenders prosecuted and convicted in the context of prison inmates in New South Wales, Australia
- Studies of the effect of DNA on case outcomes
- An evaluation of DNA use in burglary investigations

Wilson et al. concluded that their findings "*generally support the value of DNA testing for police investigations, particularly for high volume crimes such as burglary, although most of the empirical evidence is methodologically weak. Additional work is needed, not only with respect to DNA testing but other forensic methods as well.*"

Databases

If forensic databases have been in existence for a long time, in some cases for more than 100 years (e.g., fingerprints, identification systems), the scope and computing power of these databases only really took off in the late 1990s with the development of systematic collection and collation of DNA profiles. Not only did the development of these databases change the way investigations are carried out, but it also changed the perception of the judiciary toward forensic traces (Walsh 2009). Such databases usually contain information (1) derived from traces found at crime scenes and (2) obtained from people. The use and success of these databases are constantly reported in the news. Anecdotal evidence about the effectiveness of such databases in identifying people is compelling. However, the question ought to be asked: In regard to their cost as well as the legal and broader societal concerns associated with their use, are forensic databases (and in particular DNA databases) really effective? What is their actual impact on solving crime and on other aspects of policing and security?

Bradbury and Feist (2005) confirmed that DNA databases lead to impressive results. An interesting finding, based on the experience in England and Wales, is that such databases seem to change the way forensic traces are perceived and used: From being mostly corroborative evidence previously, they are now more systematically used to identify suspects.

It is, however, difficult to make general and global conclusions in this area because the management of these databases and the processes involved vary greatly across jurisdictions and police and forensic organizations. Factors having an effect on the content and uses of DNA databases, and hence impacting on their effectiveness, include:

- Number of cases attended
- Number of organizations involved in the collection of specimens and their analysis
- Anonymization protocols
- Profile removal policy
- Technology used
- Crime scene and triage policies
- Training and qualifications of the people involved
- Legislative compliance

The ideal database would contain the entire criminal population that is active at a given time. This is obviously impossible in practical terms.

DNA databases performance can be measured in different ways (Walsh 2009). A common model is to measure the ratio between the number of "matches" and the number of people profiles stored in the database. Recent European data based on 19 available databases (ENFSI DNA Working Group 2011) show this ratio varies between 0.03 (France) and 0.26 (Denmark and England and Wales). This variation reflects variation in practices and usages rather than intrinsic differences between databases. For example, if a jurisdiction submits fewer profiles from traces than others for a similar number of people profiles stored in the database, there will be fewer "matches" and the performance ratio will, in turn, be smaller. Differences in the policy applied with respect to profile removal will also have a major effect on the calculated performance.

Another point to emphasize is that the discovery of a database "match" will have an influence all through the criminal justice chain. It is, however, difficult to know the global effect of this "match" on operations including the effects of saving time and linking cases on court outcomes (see below).

In summary, it can be accepted that the development of forensic databases such as DNA databases leads to spectacular results, especially when the aim is to identify links between a trace found at the crime scene and a person. However, their effectiveness in terms of linking information obtained from traces remains unclear. Further, accurate database performance evaluation is difficult and may be misleading without a complete understanding of the overall information treatment processes involved (Bieber 2006). As stated by Bieber (2006), "*a lack of integration between the DNA laboratories and the other components of the justice system responsible for following up on results is perhaps the biggest weakness, in that desirable outcomes have not been clearly defined or carefully researched*." In this regard, the integration of forensic information in a crime intelligence database should be seen as a positive development (Rossy et al. 2012). This approach seems particularly effective when the system in place enables the exchange and comparison of marks and traces across several jurisdictions.

From Identification to Charges and to Court

If the identification of the offender is one of the major aims of an investigation, the forensic science work goes well beyond the discovery of a fingerprint or a DNA "match." Identifying an individual does not necessarily lead to an arrest, let alone to a guilty outcome in court. The effectiveness of forensic science is therefore closely related to the conversion rate between "identifying an individual (through forensic means)" and "arresting this individual." Do we have an idea of what this rate could be? Only a small number of studies have been undertaken in this area and the results appear sometimes inconsistent. However, the following trends have been identified:

- In England and Wales, out of ten fingerprint and DNA "matches," approximately seven could be translated into an arrest in high-volume crime cases (Bradbury and Feist 2005).
- In a 1-year study across seven divisions of two police forces in England, it was found that 81 % of links identified on the basis of shoe marks, toolmarks, and biological traces led to

an arrest. Further, investigators interviewed in this study estimated that forensic traces played an essential or secondary role in the arrest in 91 % of the cases (Burrows and Tarling 2004).

Overall, there seems to be an attrition rate of 20–30 % between the identification and arrest stages. There are many reasons to explain this attrition. For example, the suspect may admit a contact with the scene or surface under investigation but challenges the context that would lead to the charges; for example, he or she can claim a legitimate contact. Further, secondary transfer, contaminations, uncertainty in the identification, and laboratory or clerical errors are well-known phenomena and can also play a role in some cases. For these reasons, a forensic "match" has little value per se. The investigator in charge must integrate the results obtained from forensic examination into the remaining information available in the case. In other words, the exploitation of a forensic result is strongly related to the investigation. It is therefore not surprising to see that the likelihood of an arrest once an identification is made depends on the quality of the investigator in charge (Julian et al. 2012). To this end, the investigator's work appears scientific in nature, not only because material traces are considered, but also because hypotheses are identified and tested in a structured and logical manner. However, some cultural issues (i.e., investigator vs. scientist) often remain and explain why investigators usually tend to favor human aspects (e.g., eyewitnesses, informants, and psychological profilers) over material traces. Integrating investigators and scientists, sharing the same goal, in a common (or at least closely related) structure appears to deliver improved outcomes, although this is difficult to capture and measure (Barclay 2009).

It is also interesting to note that a "match" may be identified once an individual has already been arrested. The forensic result in this case is added to the file, and it becomes difficult to measure its impact on the defendant's behavior and on the defense or prosecution strategies through the court process. There seems to be a difference between serious crime and volume crime. For example, Briody and Penzler (2005) found that DNA evidence was a positive predictor that prosecutors would pursue cases in court and had a powerful influence on jury decisions to convict. Incriminating DNA evidence demonstrated no significant effect on inducing guilty pleas from defendants for serious crimes against the person. However, it did correlate significantly with cases reaching court and with guilty pleas being entered for property offence cases. In serial cases involving DNA, offenders have the tendency to confess their involvement in additional cases, in which no forensic traces exist and no suspects have been identified (Bradbury and Feist 2005).

"End-to-End" Process

It is clear that the accurate assessment of forensic science effectiveness requires a consideration of the overall process involved in the practice of this discipline, from the actual crime to court or other contributions to policing. Such studies are difficult to undertake because the process by which forensic traces are discovered at the scene and then moved through the criminal justice system to court is very convoluted. As a result, while end-to-end studies provide a more holistic view on how forensic science contributes to the criminal justice system, there are many difficulties in generalizing any statement about forensic science effectiveness as practices vary tremendously and are almost never explicitly integrated with clear policing strategies. Examples of such studies reported to date are briefly described below.

- ***United Kingdom Scientific Work Improvement Model (SWIM) and Derbyshire Constabulary Study***

The SWIM project (Home Office 2007) built upon some work undertaken during 2002 and 2003 in the Derbyshire Constabulary, in which a direct correlation between the time taken from crime occurrence to forensic-led detection (lead time) and crime levels had been identified. In other words, reducing the lead time could reduce the level of crime. The SWIM program ran over a 2-year period and involved 41 forces looking at the police and scientific functions in England and

Wales. Focusing on burglaries and vehicle thefts and on DNA and fingerprints, the study examined the main stages of the overall process (attendance, submission, identification, and detection) and evaluated the lag time between each of the phases and the success of the case as it moved through to the next stage. At each stage, the result was calculated as the proportion of transactions that were transferred to the next stage. The lead time was calculated for each crime report as the earliest activity date at each forensic process stage. The success rate was calculated as the percentage of cases that successfully moved to the next stage.

The SWIM report made a large number of general and police force specific recommendations for improvement at all stages of the forensic process. In particular, it demonstrated the need to identify significant leakage points in the process and that systems should be developed to capture and compare relevant data and learn from top performers.

- ***Denver Colorado Study***

This study (Ashikhmin et al. undated) aimed to evaluate the effectiveness and cost of DNA technology on high-volume crimes such as burglary, auto theft, and theft from motor vehicles in 2006. The study found that aggressive use of DNA in such investigations and subsequent prosecutions resulted in a significant decrease in property crimes compared to similar metropolitan areas in the United States. It also found that, in burglary cases with hits in the Combined DNA Index System (CODIS), much harsher sentences were given to high-volume, habitual offenders whose criminal activity had a higher impact on society. The study also conducted a cost benefit analysis and found that the return on investment for every dollar spent with this approach was estimated to be $90 with an actual 2-year savings to the citizens and the city of Denver of more than $5 million in police costs and $36.8 million in property loss. In conclusion, this study recommended an expansion of DNA science in high-volume crimes based on the high success rate for prosecution and the value for money return on investment.

- ***Waikato District and Environmental Science and Research: Forensic (ESR Forensic) DNA Project***

In mid-2010, the Waikato Police District in New Zealand (NZ) undertook a 3-month study in collaboration with ESR Forensic, the main government forensic science laboratory in New Zealand (Coley 2010). The aims of this trial were to monitor the implementation of quicker turnaround times by ESR Forensic and police for DNA submissions from volume crime scenes and to assess the value of the forensic submissions being made in terms of their likelihood to produce a profile and the value of those links to investigators. Return on the investment was also considered.

The trial led to a significant improvement in turnaround times by ESR Forensic which significantly increased the value of the forensic results to police. It was also found that the trial made a significant contribution to volume crime reduction through better prioritization of district volume crime DNA collection and submission and through the actioning of forensic identifications. For example, it was found that ESR Forensic Volume Crime Laboratory averaged 5.4 days turnaround time from receipt to result in the laboratory for 78 % of Waikato submissions over the 3-month trial period, improving significantly on the previous 4-week turnaround. By ensuring attendance within the same day as a crime is reported, Scenes of Crime Officers (SOCO) were able to see the added value of their forensic results and the effect of their timely response on the current crime environment. Further, investigators identified the benefits of working with rapid identifications: (1) It has the potential to recover property and to prevent future offending and (2) it assists in identifying current "hot" offenders, which in turn allows the police to apply a targeted approach disrupting and influencing the current crime patterns.

- ***ANZPAA-NIFS End-to-End Forensic Identification Process Project***

The Australia New Zealand Policing Advisory Agency National Institute of Forensic Science endorsed this project in 2010 to *"review end-to-end forensic processes and develop* *a national framework for efficient crime scene analysis*" (Brown and Ross 2012). End-to-end processing was defined in this project as the time from the report of a crime through to the arrest of an offender. The process was broken into five distinct stages: attendance, submission, analysis, identification, and investigation. The study aimed to benchmark performance for the use of fingerprints and DNA in burglary cases. It was carried out in 2011 in partnership with the eight Australian police agencies and a number of relevant DNA laboratories. In total, 17 sites across Australia were considered for more than 8,000 burglaries reported over a 5-month period.

The findings and recommendations are still being discussed and reported to the participants and stakeholders at the time of writing. It is hoped that this project will not only provide an appreciation of the forensic performance within Australia but will also identify some directions with respect to (1) the scope of evaluation that could be carried out and (2) the possibilities that exist to improve forensic science service delivery. Further, a second snapshot study is envisaged to determine if recommended changes have a positive impact on performance for the end-to-end forensic process.

Conclusions: Future Directions

This entry presents a critical review of the topic of effectiveness of forensic science. In particular, it attempts to address the central question: What do we actually know about the impact of forensic science on the criminal justice system in general and on criminal investigations in particular?

It is, however, impossible to discuss the topic of effectiveness without knowing what the contribution of forensic science is expected to be. Similarly, it is impossible to assess effectiveness without contrasting outcomes and impact against what the forensic science community itself claims to be able to achieve. This discussion is challenging because, in general, expectations are poorly defined. Further, the traditional contribution of forensic science is primarily oriented toward the court's needs and requirements. How can we account for this contribution? In this context, it is impossible to define what "successes" and "failures" are, because, in this dimension, forensic science's role is to provide impartial information to the trier of facts. Forensic scientists never celebrate the outcome of a court process. Why would they be assessed on this basis then?

Beyond the court process, assessing the effectiveness of forensic science is equally challenging. Effectiveness cannot only be measured in terms of crime resolution. The contribution of forensic science is actually much broader (Barclay 2009). To this end, there is an emerging shift from the traditional focus on forensic science in the laboratory toward a more significant contribution of forensic science in intelligence-led policing and investigation (Ribaux et al. 2006, 2010a, b; Roux et al. 2012). This contribution, which is not trivial, must be accounted for in the overall assessment of the effectiveness of forensic science.

Research to date has shown interesting trends, such as that the best use of forensic science is reached when the collection and processing of forensic case data is promoted through close cooperation between forensic scientists and investigators. There are however major discrepancies in forensic policies across institutions and even across individuals, and they are poorly integrated into policing strategies. Thus, inference of any general rules about the effectiveness of forensic science drawn from individual studies may be invalid. It is argued that there is a need to better understand and explain the broader contribution, as a precondition to accounting for effectiveness. Otherwise, managerial decisions taken on the basis of limited studies and misunderstandings about what forensic case data can bring may have unintended consequences for the whole system.

Research outcomes to date probably reveal more about the need for methodological improvements on how to undertake such research than providing a clearer answer with respect to the effectiveness of forensic science. In other words, more research is needed in this area.

Related Entries

- ▶ Critical Report on Forensic Science
- ▶ Forensic Science
- ▶ Forensic Science and Criminal Investigation
- ▶ Forensic Science and the Paradigm of Quality
- ▶ Forensic Science Culture
- ▶ Forensic Science Professions
- ▶ Organizing and Supplying Forensic Science Services
- ▶ Philosophical Basis of the Forensic Process

Recommended Reading and References

Adderley R, Townsley M, Bond J (2007) Use of data mining techniques to model crime scene investigator performance. Knowl-based Syst 20:170–176

Ashikhmin S, Berdine SJ, Morrissey MR, LaBerge G (undated) Effectiveness and cost efficiency of DNA evidence in volume crime Denver Colorado site summary. Denver District Attorney's Office & Denver Police Department Crime Laboratory Bureau

Barclay D (2009) Using forensic science in major crime inquiries. In: Fraser J, Williams R (eds) Handbook of forensic science. Willan, Cullompton, pp 337–358

Baskin D, Sommers I (2010) The influence of forensic evidence on the case outcomes of homicide incidents. J Crim Justice 6:1141–1149

Bieber FR (2006) Turning base hits into earned runs: improving the effectiveness of forensic DNA data bank programs. J Law Med Ethics 34(2):222–233

Bieber FR, Brenner CH, Lazer D (2006) Finding criminals through DNA of their relatives. Science 312(5778):1315–1316

Bradbury SA, Feist A (2005) The use of forensic science in volume crime investigations: a review of the research literature. Online report, Home Office

Briody M, Prenzler T (2005) DNA databases and property crime: a false promise? Aust J Forensic Sci 37(2):73–86

Brodeur JP (2010) The policing web. Oxford University Press, New York

Brown C, Ross A (2012) End-to-end forensic identification process project—volume crime, ANZPAA-NIFS key project report

Burrows J, Tarling R (2004) Measuring the impact of forensic science in detecting burglary and autocrime offences. Sci Justice 44(4):217–222

Coley R (2010) Waikato district and ESR DNA project 2010. New Zealand Police

Crispino F (2008) Nature and place of crime scene management within forensic sciences. Sci Justice 1:24–28

Delémont O, Lock E, Ribaux O (2013) Forensic science and criminal enquiry, Encyclopedia of Criminology and Criminal Justice. Springer Verlag Publishers, London

ENFSI DNA Working Group (2011) ENFSI survey on DNA—databases in Europe 2011. http://www.enfsi.eu/page.php?uid=98. Last accessed 27 Jan 2013

Girod A, Champod C, Ribaux O (2008) Les traces de souliers. Presses Polytechniques Universitaires Romandes, Lausanne

Home Office (2007) Summary report of the Scientific Work Improvement Model (SWIM) package. http://www.innocenceproject.org/know/. Last accessed 27 Jan 2013

Julian RD, Kelty SF, Roux C, Woodman P, Robertson J, Davey A, Hayes R, Margot P, Ross A, Sibly H, White R (2011) What is the value of forensic science? An overview of the effectiveness of forensic science in the Australian criminal justice system project. Aust J Forensic Sci 43(4):217–229

Julian R, Kelty S, Robertson J (2012) "Get it right the first time": critical issues at the crime scene. Curr Issues Crim Justice 24(1):25–37

Kelty SF, Julian R (2012a) Success in forensic science research and other collaborative projects: meeting your partners' expectations. Forensic Sci Policy Manag Int J 2(3):141–147

Kelty SF, Julian R, Robertson J (2012b) Professionalism in crime scene examination: the seven key attributes of top crime scene examiners. Forensic Sci Policy Manag Int J 2(4):175–186

Mucchielli L (2006) L'élucidation des homicides: de l'enchantement technologique à l'analyse du travail des enquêteurs de police judiciaire. Déviance Soc 30(1):91–119

Raymond JJ, Van Oorschot R, Gunn P, Walsh SJ, Roux CP (2009) Trace evidence characteristics of DNA: a preliminary investigation of the persistence of DNA at crime scenes. Forensic Sci Int Genet 4(1):26–33

Ribaux O, Walsh SJ, Margot P (2006) The contribution of forensic science to crime analysis and investigation: Forensic intelligence. Forensic Science International 156(2–3):171–181.

Ribaux O, Baylon A, Roux C, Delémont O, Lock E, Zingg C, Margot P (2010a) Intelligence-led crime scene processing. Part I: forensic intelligence. Forensic Sci Int 195(1):10–16

Ribaux O, Baylon A, Lock E, Roux C, Delémont O, Zingg C, Margot P (2010b) Intelligence-led crime scene processing. Part II: intelligence and crime scene examination. Forensic Sci Int 199:63–71

Rix B (2004) The contribution of shoemark data to police intelligence, crime detection and prosecution findings. Home Office, Research, Development and Statistics Directorate, London, 236. http://www.homeoffice.gov.uk/rds/pdfs04/r236.pdf. Last accessed: 27 Jan 2013

Roman JK, Reid S, Reid J, Chalfin A, Adams W, Knight C (2008) The DNA field experiment: cost-effectiveness analysis of the use of DNA in the investigation of high-volume crimes. Urban Institute, Justice Policy Center, Washington, NCJ 222318. http://www.ncjrs.gov/pdffiles1/nij/grants/222318.pdf. Last accessed 27 Jan 2013

Rossy Q, Ioset S, Dessimoz D, Ribaux O (2012) Integrating forensic information in a crime intelligence database. Forensic Sci Int. http://dx.doi.org/10.1016/j.forsciint.2012.10.010

Roux CP, Crispino F, Ribaux O (2012) From forensics to forensic science. Curr Issues Crim Justice 24(1):7–24

Schroeder DA, White MD (2009) Exploring the use of DNA evidence in homicide investigations. Police Quart 12(3):319–342

Strom KJ, Hickman MJ (2010) Unanalyzed evidence in law enforcement agencies. Criminol Public Policy 9:381–404

Tilley N, Ford A (1996) Forensic science and crime investigation. In: Webb B (ed) Crime detection and prevention. Police Research Group, Home office, London, 73

Walsh SJ (2009) Evaluating the role and impact of forensic DNA profiling on key areas of the criminal justice system. PhD thesis, University Technology of Sydney, Sydney

White JH, Lester D, Gentile M, Rosenbleeth J (2011) The utilization of forensic science and criminal profiling for capturing serial killers. Forensic Sci Int 209(1–3):160–165

Wilson DB, McClure D, Weisburd D (2010) Does forensic DNA help to solve crime? The benefit of sophisticated answers to naive questions. J Contemp Crim Justice 4:458–469

Forensic Science Evidence

► Scientific Evidence in Criminal Prosecutions

Forensic Science in the Nineteenth and Twentieth Centuries

Douglas M. Lucas
The Centre of Forensic Sciences Ontario,
Burlington, ON, Canada

Overview

Medicine was the first science-based profession to be used for forensic purposes, and forensic medicine can be considered to be the mother of forensic science. In the sixth century, the Emperor Justinian recognized the special position of the expert witness when he declared that physicians were not ordinary witnesses but rather persons who gave judgments rather than testimony. In England, the Statute of Westminster in 1265 contained the first provision for the office of coroner, and, in 1507, the Bamberg Code in Germany required that medical evidence be presented in all cases of violent death. These provided a legal basis for the practice of forensic medicine and were the precursors of modern coroner and medical examiner legislation. In North America, the early English settlers brought the coroner system with them and inquests were held as early as 1635 in New England.

Prior to the nineteenth century, evidence of poisoning was purely clinical and circumstantial because there were very few chemical tests for poisons available. As knowledge of chemistry increased, it was applied to the detection of poisons, usually by university professors ensconced in traditional university departments. Other now commonly recognized forensic sciences – fingerprint identification, trace evidence examination, questioned document examination, firearms and toolmark examination, and serology were essentially unknown.

Toward the end of the eighteenth century, precursors of what would now be considered "Institutes of Forensic Medicine" began to appear as personal chairs in university departments. The first such was the *Chaire de médecine légale et de histoire de l'art de guérir* established in the University of Paris in 1794. Within a few years, similar chairs were established in the Universities of Strasbourg and Montpellier. Forensic science laboratories, however, were nonexistent.

Developments in the Nineteenth Century

Forensic Medicine

Although most of what we now refer to as the forensic sciences did not begin to develop until the latter half of this century, forensic medicine began to be recognized as a specialized branch of medicine, "legal medicine," early in the 1800s. In Austria, a Chair in Legal Medicine and Hygiene was established in the University of Vienna in 1804, while in France *l' Institut de médecine lé gale de Paris* was founded in 1868.

The first British Professor of Legal Medicine was Andrew Duncan Sr. in the University of Edinburgh. Duncan persuaded the University to establish the Regius (government-funded) Chair in Medical Jurisprudence in 1806; however, it was his son, Andrew Duncan Jr., who became the first incumbent in that Chair. This father/son succession pattern was repeated in the same Chair in 1906 when Henry Harvey Littlejohn succeeded his father Sir Henry Littlejohn. Similarly, at the University of Glasgow, John Glaister Sr. and Jr. combined to hold the Regius Chair of Medical Jurisprudence and Forensic Medicine, established in 1839, from 1898 to 1962. Edinburgh became the cradle of forensic medicine for the English-speaking world as graduates established forensic medical services in many other countries.

In England, the most prominent early name was Alfred Swaine Taylor who became Professor of Medical Jurisprudence at Guy's Hospital Medical School in London in 1834. Taylor was a prodigious author, and his "Principles and Practice of Medical Jurisprudence," first published in 1865 and continuously revised and updated by later authors, was long recognized as a standard work.

The first lectures on legal medicine in the USA were given in New York in 1804 by James S. Stringham, a graduate of the University of Edinburgh. A major development occurred in Massachusetts in 1877 with the abolition of elected coroners and the establishment of a medical examiner system which required that the medical examiner be a qualified pathologist.

Forensic Toxicology

As the science of chemistry developed, a few professors of forensic medicine began to use this new science for one of their more intractable problems, the detection of poisons. One of these was Mathieu Joseph Bonaventure Orfila, universally acknowledged to be the "Father of Toxicology" (Schuller 2003). Orfila studied at Valencia and Barcelona before graduating in medicine from the University of Paris in 1811. He soon became Professor of Legal Medicine and Chemistry at that University and, in 1814, published the first major work on toxicology "*Traité des poisons tirés des règnes minéral vé gétal et animal ou Toxicologie Générale*" which rapidly gained international recognition. Much of Orfila's research was devoted to the study of arsenic, and, in 1839, he was the first to extract it from human organs rather than from just stomach contents. Orfila used the results of this research in 1840 in his testimony at the trial of Marie Lafarge for the murder of her husband by means of an arsenic-laced cake. This trial focused worldwide attention on the new science of toxicology; *L'affaire Lafarge* was probably the first case in history in which convincing scientific testimony (other than medical) was presented.

The introduction of analytical methods for metallic poisons caused potential poisoners to turn their attention to the alkaloids that were beginning to be isolated from plants. These were considered to be undetectable in human organs; however, in 1850, this began to change with the development by Jean Servais Stas, a Professor of Chemistry at *L'École Royale Militaire* in Brussels and a former student of Orfila's, of a method for the extraction of nicotine from the organs of a murder victim. Stas' basic procedure (as modified by F.J. Otto in 1856) was based on the differing solubility of such drugs in acids and bases and is still in use. Identification of any extracted drugs was based on colors developed, and microscopic morphology of crystals formed, when the extract was reacted with other chemicals.

A trend toward specialization in toxicology began near the end of the nineteenth century with the emergence of, for example, the most prominent of the early American toxicologists, Rudolph August Witthaus who had studied chemistry at Columbia University and at the Sorbonne. He became a Professor of Chemistry and Physiology at New York University in 1876 and published one of the earliest American textbooks on chemistry and toxicology in 1879.

Fingerprints

In the late nineteenth century, the only accepted method for identifying habitual criminals was photography. While useful when the criminal's name was available, there was no way to file photographs other than by name. If the criminal changed his name, photograph files were of no assistance. In 1879, Louis Alphonse Bertillon, a junior records clerk in the headquarters of the *Sureté* in Paris, devised a detailed system of anthropometric measurements, such as the length and width of head and lengths of left middle finger, right foot, and left forearm, which would be unique and relatively constant for an individual. His system, which he called "anthropometry" and which the press dubbed "*Bertillonage*" was adopted by the *Sureté* in 1882. Bertillon went on to become the Director of the Police Identification Service in Paris.

Bertillonage was a cumbersome process with an unwieldy filing system and even Bertillon recognized that criminals do not leave their anthropometric measurements behind at a crime scene. The basis for a better technique had been established in a paper published in England in 1684 by Dr. Nehemiah Grew in which he referred to "innumerable little ridges" on the ends of the fingers. In 1686, these ridges were independently described by Marcello Malpighi, a Professor of Anatomy in the University of Bologna. Neither, however, suggested any potential of these for personal identification. It was not until 1823 that the patterns formed by these ridges were described by a Czech physiologist, Jan Evangelista Purkinje, of the University of Breslau. While Purkinje suggested a classification system based on nine major types, he too failed to recognize their potential for individualization.

In 1877, Thomas Taylor, a microscopist in the US Department of Agriculture, published an article in the "American Journal of Microscopy" in which he suggested that "markings on the palms of the hands and the tips of the fingers" could be used for identification in criminal cases. This suggestion was not pursued in the USA, and it was left to a Scottish physician working in Tokyo, Dr. Henry Faulds, to publish a letter in Nature in 1880 in which he suggested that they might be useful as a means of personal identification (Faulds 1880). His letter prompted publication of a second letter a month later in the same journal from William Herschel, a British officer working in the Indian Civil Service, reporting that he had been using thumb impressions to identify illiterate prisoners since 1856. Herschel, who later received a knighthood, also made the critical observation that the patterns of what he called "papillary lines" did not change with time.

Despite the fundamental importance of these observations, they were of limited practical value until Francis Galton, an English scientist and explorer, published a book titled "Fingerprints" in 1892 (Galton 1892). Galton corresponded with Herschel and they agreed that fingerprints were individual and permanent – a fundamental requirement for their value as a system for identification. Galton described a classification system, using the arches, loops, and whorls of fingerprint patterns, which came to be known as "dactyloscopy." Also in 1892, in Argentina a Croatian, Juan Vucetich, used the identification of a bloody fingerprint to solve the murder of two small children by their mother, the first use of fingerprint identification in a criminal case.

Edward Richard Henry, the Inspector General of Police in Bengal in the 1890s, became aware of Galton's work and, by 1897, had enhanced Galton's classification system into a more practical system that bears his name and is still used worldwide. Henry published "Classification and Use of Fingerprints" in 1900 (Henry 1900), and his system formally replaced anthropometry in England in 1901. Henry was by then Assistant Commissioner in charge of the Criminal Investigation Branch at Scotland Yard.

Trace Evidence

Two of the commonest requests made of a forensic science laboratory are "What is this material?" and "Could these two items have come from the same source?" The items can be almost anything but typically are paint, glass, soil, hairs, fibers, metals, gunshot residue, flammable liquids, or explosives. The science involved in trace evidence examination is principally chemistry and the early practitioners were usually professors of chemistry who accepted this type of request from local police investigators as a public service.

Remarkably, it was a lawyer, not a scientist, who first emphasized the assistance that police investigators could derive from the basic sciences. Hans Gustav Adolf Gross was a Professor of Law at the University of Linz in Austria. He also served as an examining magistrate which inspired him to spend 20 years studying science textbooks and developing principles of criminal investigation. Gross first published his classic book "*Handbuch für Untersuchungsrichter*" ("Handbook for Examining Magistrates") in 1893. It was later translated into English under the simple title "Criminal Investigation" and resulted in Gross becoming widely referred to as the "Father of Scientific Criminology."

Although chemistry was developing rapidly, it was microscopy which was the fundamental tool of the nineteenth-century forensic scientist and which piqued Gross's interest. Even today, the examination of trace evidence usually starts with a low-power stereo microscope, often just to find the evidence, and then may progress to higher power microscopes and, in the nineteenth century, to the basic "wet chemical" procedures such as color and crystal tests. One of the first types of trace evidence to be studied was human hair. In 1857, Jean Louis Lassaigne, a French chemist, published the results of his systematic microscopic studies of hair "*De l'examen physique des poils et des cheveux*" ("On the Physical Examination of Hair").

Firearms/Toolmarks

The marks left by a hard tool on a softer surface are known as "toolmarks" and their examination in an effort to, for example, associate a particular tool such as crowbar with a forced entry mark, is quite common. The striation marks made on a rifled firearm barrel by both the original rifling tool and subsequent use are impressed onto a bullet surface as it passes through the barrel and are a particular form of toolmark.

In 1835, Henry Goddard in London identified a lead ball which had killed a man as having been made in a mold belonging to a suspect on the basis of an unusual mold mark on the ball. During the Civil War in the USA, a number of shootings (including that of Confederate General Stonewall Jackson in 1863) were resolved by associating "class" characteristics of the bullets, for example, shape, caliber, number of lands and grooves, with the types of firearm suspected (but not with an individual weapon).

In 1889, Jean Alexandre Lacassagne, the founder of the Faculty of Legal Medicine at the University of Lyon in 1880, was probably the first to use a microscope to examine a bullet removed from a shooting victim. He observed seven longitudinal grooves on it and concluded that they must have been made by the rifling of the barrel. Further advances in firearms examination did not develop until the twentieth century.

Forensic Biology

Attempts to identify and individualize body fluids have been a major activity in forensic science since the early nineteenth century. In 1827, the above-mentioned Mathieu Orfila published his research on the microscopic identification of red blood cells based on their size and whether or not they were nucleated. This technique was quite impractical for dried blood in which the blood cells are no longer intact. In 1853, Ludwig Teichmann, a Polish Professor of Anatomy in Gottingen, Germany, described a microscopic crystal test for hemoglobin based on the formation of brownish crystals of hematin. A Dutch scientist, Izaac Van Deen, discovered in 1861 that when tincture of guaiacum and oil of turpentine were added to a bloodstain a blue color resulted. This "guaiac test" was the first of many color tests for hemoglobin; some such as luminol and phenolphthalein are still in use. These tests, although very sensitive, are not specific for blood and, therefore, are referred to as "presumptive tests."

Questioned Document Examination

Establishing the genuine or bogus nature of handwriting, and of documentary material in general, has been an issue in investigations for centuries. Early "handwriting experts" were teachers, bankers, etc., usually having little or no experience in forensic work. For example, Alphonse Bertillon, who had no qualifications in handwriting examination, testified in the infamous Dreyfus affair in 1894 in France and, incorrectly (as it later turned out), identified Captain Alfred Dreyfus' handwriting on the incriminating document.

By the end of the nineteenth century, although forensic medicine was frequently used in criminal investigations and trials, the other forensic sciences were just beginning to be recognized as having the potential to assist in such procedures. Some of the techniques (or at least the fundamental bases for them) had been established, but practitioners were relatively few. That was to change dramatically in the twentieth century.

Developments in the Twentieth Century

Forensic Science Institutes and Laboratories

Institutes of legal medicine began to increase in number in Europe in the early part of the century, but remained primarily based in universities. The Edinburgh influence continued to spread with the appointment of Sidney Smith, an Edinburgh graduate, as the Principal MedicoLegal Expert in Cairo, Egypt. There he persuaded a team of experts in other disciplines to devote some of their time to forensic science. One of these was Arthur Lucas of the Government Chemistry Laboratory who, in 1921, published "Forensic Chemistry" which was one of the first texts to cover forensic science topics other than medicine and toxicology. Smith himself developed expertise in the emerging field of firearms identification. He returned to Edinburgh to take up the Chair in the University in 1927 and received a knighthood in 1949.

In North America, there was minimal progress in forensic medicine until the establishment of the Office of the Chief Medical Examiner in New York City in 1918 under the leadership of Dr. Charles Norris. Norris' objective was to develop a medicolegal institute which would provide service and research equivalent to the leading institutes in Europe. The Office also became a leading teaching center after Milton S. Helpern became an integral member in 1931. "Alumni" of the Office went on to establish important institutes across the USA.

More broadly based forensic science laboratories gradually began to appear, the first being *L'Institut de police scientifique* in the University of Lausanne established in 1909 by Dr. Archibald Rodolphe Reiss. Within a year, it was followed by *Laboratoire de police technique* in Lyon under the leadership of Edmond Locard a native of Lyon who had studied under Lacassagne in the Faculty of Legal Medicine. Because he wanted to significantly expand the scope of Lacassagne's Institute, Locard persuaded the police of the *Rhône préfecture* to provide him with two rooms in an attic and two assistants to establish a police laboratory in Lyon, equipped initially with only a microscope and a simple spectrometer. This lab went on to become a major center for teaching and research, in addition to providing forensic science services to the police. Students came from around the world, many returning home to establish laboratories of their own. Over the ensuing 15 years, similar laboratories were established in Germany, Austria, Canada, Finland, Germany, Holland, and Sweden.

In Britain, it was not until 1935 that a forensic science laboratory was formally established by the Metropolitan London Police. The "Met Lab" went on to become one of the most respected such organizations in the world. In the rest of England, laboratories that had been started by local police forces gradually came under the wing of the Home Office Forensic Science Service (FSS) – as did the Met Lab in 1996. (Ironically, one of the reasons for these transfers was to enhance the appearance of impartiality and independence from the police. However, following the decision by the Home Office in 2010 to disband the FSS (see below), some police forces decided to again establish their own labs).

In 1966, the Home Office opened the Central Research Establishment (CRE) at Aldermaston. This laboratory had a major impact on forensic science not only within the FSS but worldwide. Most operational laboratories have neither the time nor the capability to perform systematic research on fundamental issues. This dearth was the subject of a classic essay by Paul Kirk of the University of California in 1963 (Kirk 1963). He wrote: "In short, there exists in the field of criminalistics a serious deficiency in basic theory and principles, as contrasted with the large assortment of effective technical procedures." CRE was staffed and funded to perform such studies and share the results internationally thus greatly enhancing the reliability and validity of scientific evidence. Unfortunately, CRE was closed in 1997.

All of the FSS Labs were funded from central and local government funds until 1991 when the FSS became an Executive Agency of the Home Office and changed to a fee for service basis. (This move toward "privatization" evolved into a full Government-Owned Contractor-Operated Company in 2005. The format was not successful financially in spite of the FSS's stellar scientific achievements and reputation and, in December 2010, the Home Office announced that the operation would be closed in March 2012. A fundamental issue is whether forensic science service delivery should be "commercialized" or should be recognized as an essential part of the justice system and thus a core public service funded by the state!).

In North America, the first government-funded forensic science laboratory was established by Dr. Wilfred Derôme in Canada in 1914. Derôme had gone to Paris in 1909 to study under Professor Victor Balthazard and, while there, became aware of the laboratory Locard was developing in Lyon. He determined to do the same in Canada and, on his return from Paris, Derôme persuaded the Attorney General of Quebec of the importance of this new type of service. As a result, in July 1914, the Premier of Quebec announced the establishment of *Laboratoire de recherches médico-légales* in Montreal (Côté 2003).

The first forensic science laboratory in the USA was established in the Los Angeles Police Department in 1923 by Chief August Vollmer, a great innovator in police work. After his brief term with LAPD, Vollmer returned to the Berkeley Police Department where he had established a School of Criminology in the University of California. Although located on the campus, the School had no official status until 1948 when the School of Criminology was formed and Professor Paul Kirk became the first Chair of the Criminalistics Department. Kirk became a legendary figure in forensic science in the USA and many of his students went on to establish forensic science laboratories across the state and the country. Unfortunately, the UC School, one of the few in the USA with a graduate program in forensic science, was closed in the 1970s.

Although the LAPD Laboratory was the first in the USA, its early days were overshadowed by the prominence of a laboratory established in Chicago in 1930. Following the St. Valentine's Day Massacre in 1929, a coroner's jury of prominent citizens was so impressed with the firearms evidence presented by Calvin Goddard (see below) that one of their members provided $125,000 to fund the creation of a full-service "crime laboratory" under Goddard's leadership. With the support of John Henry Wigmore, Dean of the Law School in Northwestern University, the "Scientific Crime Detection Laboratory of Chicago" was housed in the Law School building. At this lab, Goddard trained staff from many agencies around the USA. Following his retirement, the Laboratory was transferred to the Chicago Police Department in 1938 and then to the Illinois State Police in 1996.

What was to eventually become the largest and best known of the forensic science laboratories in the USA was not established until 1932 when J. Edgar Hoover created the FBI Laboratory in Washington, D.C., a modest facility with a staff of one who had studied firearms examination under Goddard in Chicago.

Beginning in the late 1960s, the number of forensic science laboratories in the USA virtually exploded, driven largely by "the war on drugs" and Federal Government funding through the Law Enforcement Assistance Administration (LEAA). The vast majority of these were established within law enforcement agencies, some staffed with police officers who had limited or no science backgrounds. As the century progressed, the trend was toward better qualified civilian scientific staff although sometimes still with police administrators.

Forensic Medicine

For much of the first half of the century, a giant in forensic medicine in England was Sir Bernard Spilsbury who was a Home Office Pathologist from1908 to 1948. It was he who convinced Scotland Yard Detectives of the importance of having the pathologist and other specialists at murder scenes. In the courtroom, he was able to exert a spell over judges and juries, sometimes with little more basis than confidence in his own vast experience of over 25,000 cases. His reputation was not based on his writings, however, since, unlike most of his predecessors and contemporaries, he published virtually nothing.

Following Spilsbury, forensic medicine in England was led by Keith Simpson at Guy's Hospital and Francis Camps at the London University Hospital. Camps set up the Department of Forensic Medicine in the London Hospital Medical College in 1945. By the time of his retirement in 1970, he had built this Department into one of the foremost of its kind in the world.

In the USA during much of the twentieth century, forensic medicine was dominated by a relatively small number of individuals. In addition to Charles Norris, there were Milton Helpern in New York, Alan Moritz in Cleveland, Theodore Curphey in Los Angeles, Russell Fisher in Baltimore, and Charles Petty in Dallas. Forensic pathology services in much of the rest of the country were provided by "part-timers" with varying degrees of qualifications and expertise. That began to change toward the end of the century with more and better training and with recognition of the need for research-based (as opposed to experience-based) testimony. Nevertheless, in large parts of the country, the quality of death investigation services remained variable – elected coroners (often funeral directors) presided and, for much of the century, autopsies had to be performed in makeshift facilities such as funeral homes.

Forensic Toxicology

It was not until Norris established a Toxicology Section led by Dr. Alexander Gettler in 1918 that improvements in toxicological service began to occur. Gettler had no model to follow so much of his early work was of a pioneering nature, and his laboratory became the birthplace of forensic toxicology in the USA.

Virtually all of the early analytical methodology was based on wet chemistry and microscopic procedures which required large samples – 500 g of liver was the typical starting point – as contrasted with the 1–2 μl of blood that is now the norm. In addition to his innovative analytical procedures, Gettler's principal contribution to modern forensic toxicology was a graduate course in toxicology at New York University which he started in 1935 and taught until his retirement in 1959. His graduates went on to become an outstanding "second generation" of forensic toxicologists who started laboratories across the country.

As the twentieth century proceeded, developments in the pharmaceutical industry made forensic toxicology much more complex. While the inorganic and vegetable poisons continued to be used, synthetic drugs which were much more potent became more common. Their detection required more sensitive analytical procedures. The traditional color and crystal tests began to be supplemented by ultraviolet spectrophotometry and paper chromatography in the late 1940s, gas chromatography in the late 1950s, thin layer chromatography in the mid-1960s, mass spectroscopy combined with gas chromatography (GC/MS) in the late 1960s, and by liquid chromatography and immunological techniques in the early 1970s. GC/MS became the predominant technique for identification; when it was introduced, the complexity of interpretation of the data was such that a Ph.D. was considered an essential qualification. Within a few years, interpretation of mass spectra became more a matter of "pattern matching" with databases which could be done by persons with lesser qualification. (A similar phenomenon later occurred with DNA analysis.)

Although not toxicology per se, identification of "street drugs" employs many of the same analytical techniques and this type of analysis, because of the "war on drugs" constitutes the largest volume of casework in most forensic labs in the USA.

Of all the drugs which forensic toxicologists deal with, by far the most significant in terms of its prevalence and impact on human behavior is ethyl alcohol. Dr. Erik Widmark, a Professor of Physiological Chemistry at the University of Lund in Sweden, was among the first to systematically study the absorption, distribution, and elimination of alcohol in the body (Widmark 1932). A prerequisite for such studies was a reliable microanalytical method for alcohol in blood. Widmark's method, published in 1922 (Widmark 1922), became the most commonly used procedure worldwide and continued to be used well into the 1960s and 1970s. His mathematical pharmacokinetic parameters are still in use.

Another major development with alcohol was the invention of breath alcohol testing devices. Rolla N. Harger, a Professor of Pharmacology and Toxicology in the Indiana University Medical School, perfected the first such device, the "Drunkometer," in 1937. Robert F. Borkenstein, of the Indiana State Police, improved substantially on this in 1954 with the "Breathalyzer®," an instrument which, although now superseded by more modern equipment, continued to be used in the USA and Canada for over 50 years.

Fingerprints

The first investigator to use fingerprints for criminal investigation in the USA was Detective Sergeant Joseph A. Faurot of the New York City Police Department in 1911. It was not until 1930 that the Director of the Federal Bureau of Investigation, J. Edgar Hoover, set up a national bureau for identification which became the largest repository of fingerprints in the world.

Inked fingerprint impressions compared with the known fingerprints of an individual are still considered to be the most positive form of personal identification, although other biometrics may be used for specialized applications. Latent fingerprints, on the other hand, present greater challenges. For much of the century, development of latents was primarily by dusting with fine powders which adhered to the moisture or sebaceous deposits retained in the impression. In the 1970s, it was recognized that the chemicals in perspiration, including amino acids and various salts, presented opportunities for chemical detection and some quite sophisticated chemistry and physics was applied to this challenge. These reactions, with other chemicals or induced by various forms of radiation including argon lasers (Dalrymple et al. 1977), can produce colored or fluorescent patterns distinguishing the ridges from the valleys. One interesting technique was discovered by accident in Japan in 1977. When a hair examiner in a crime lab was mounting hairs on a microscope slide using cyanoacrylate (Superglue), he noticed that his fingerprints were developing on the slides and brought this to the attention of a colleague who confirmed it to be a valuable latent fingerprint development process.

Latent prints left at the scene of a crime can be important investigative tools but only if they can be related to the known prints of a person. Until the late 1970s, finding the knowns for such comparisons was largely based on the memories for patterns of experienced latent print examiners. In 1977, the FBI introduced an automatic fingerprint recognition system for the computerized search of fingerprint databases. Commercial providers developed other systems over the next several years culminating in the introduction in 1999 by the FBI of the Integrated Automated Fingerprint Identification System (IAFIS) containing the fingerprints and criminal histories of millions of individuals. The ability of such systems to make "cold hits" has made fingerprints the important investigative tool that the pioneers in the field had originally envisaged.

Although the use of fingerprints for individual identification originated in the medical/scientific world, their widespread application became primarily a police function. In many areas in North America, the service is provided by police "identification bureaus" rather than by forensic labs. These bureaus are generally staffed by individuals with minimal scientific background. While there is now only slight dispute of the presumption that fingerprints (friction ridge patterns) are unique to an individual and persistent for life, there remains some valid criticism of the manner in which examinations are sometimes performed. Major efforts are being exerted in the twenty-first century to deal with this through enhanced training, research, and development of recognized standardized protocols.

Trace Evidence

The exchange principle that is the basis for trace evidence examinations – "every contact leaves a trace" – was first enunciated in 1920 by Edmond Locard in *L'enquete criminelle et les methodes scientifique*. When two objects, two persons, or an object and a person come into contact, dust, fibers, hair, paint, glass, soil, etc., may transfer between them and comparisons may establish the fact of the contact. Finding, identifying, and comparing such "trace evidence" introduced additional chemistry professors to the world of forensic applications of science during the first half of the twentieth century, and this type of examination later became a common part of the work of forensic science laboratories.

Although the microscope continues to be the basic tool of the trace evidence examiner, a variety of analytical instruments to supplement microscopy began to appear and these are now capable of providing almost unbelievable sensitivity. The first of these analytical instruments, introduced in the early 1940s, was the emission spectrograph which was ideally suited to trace evidence examination because it permitted the comparison of the trace element composition of a wide range of small samples. These results were often interpreted as indicating "batch" compositions of many types of materials, interpretations which are no longer considered feasible as a result of improvements in manufacturing processes. Later analytical techniques to appear include infrared (IR) spectrophotometry introduced in the early 1950s, gas chromatography (GC) in the late 1950s, atomic absorption spectrophotometry (AAS) in the early 1960s, gas chromatography linked with mass spectroscopy (GC/MS) in the late 1960s, scanning electron microscopy linked with X-ray spectrometry (SEM/XRS) in the early 1970s, and Fourier transform infrared spectrophotometry (FTIR) later in the 1970s. New developments such as the various forms of inductively coupled plasma spectroscopy (ICPS) in the 1980s/1990s continue to be adopted.

Ironically, the enhancements in analytical capability did not substantially change the answers that could be given to the typical question "Could these two items have had a common origin?" Because the evidence items are often common materials – either manufactured, for example, paint, glass, fibers, or natural, for example, soil, hair – the answers to the question were still typically limited to "Yes they could" or "No they could not." However, the answers were of much greater quality and reliability because of the significant increase in the amount and type of analytical data derived from the items. Toward the end of the century, with investigators and courts becoming used to the much more definitive answers provided by DNA analysis (although applicable to only about 10 % of cases received in a typical forensic lab), demands for this type of examination began to diminish despite the fact that the results can still often be of considerable value to an investigation.

Questioned Document Examination

In the early years of the twentieth century, Albert S. Osborne of New York City began to develop the examination of handwriting and questioned documents into a specialized forensic science discipline. Osborne, originally a teacher of penmanship, published several papers on document examination, including one on typewriting identification in 1901, but it was the first edition of his book "Questioned Documents" in 1910 that cemented his position as the leader of the profession (Osborne 1910). This work, with later revised versions, formed the cornerstone for much of what document examiners continue to do. Today, these types of examinations require knowledge about inks, paper, writing instruments, printers, copiers, and computers. Nevertheless, the majority of the work remains the comparison of handwriting.

In England, Wilson R. Harrison published his massive work "Suspect Documents, Their Scientific Examination" in 1958 and Locard in France published *Les Faux en Écriture et Leur Expertise* in 1959.

Following World War II, many new challenges began to develop as a result of advances in the technology of writing and printing instruments. In 1945, the ballpoint pen was introduced and by the mid-1950s had almost completely replaced the fountain pen. Around 1960, the fiber-tipped pen started to become popular and brought its own challenges to identification work. Typewriter technology also began to change with the introduction of the electric typewriter in the 1930s, proportional spacing typewriters in the early 1950s, the single element typeball machine in 1961, and the correcting "lift off" typewriter ribbon in 1973. The ability to easily change balls between machines and correct errors quickly presented another challenge to the examiner. The print-wheel typing unit which became common in word processing and computer systems was introduced in 1972 followed not long after by the dot matrix and laser printers that are so common today. All of these developments in "typewriting" reveal fewer of the individual anomalies which the document examiner relies on to identify the product of a particular machine.

As the office copier became common, it introduced a whole new set of challenges to the document examiner, including the identification of the machine on which a particular copy had been produced. The thermographic process, introduced in 1950, was quickly replaced by the electrostatic plain paper reproduction process in the late 1950s.

The tools available to the document examiner have also expanded. Electronic ultraviolet and infrared viewers for the differentiation of inks; electrostatic detection apparatus for revelation of pressure patterns and indented writing, specialized photographic procedures, and computerized digital imaging equipment have all been added to the examiner's arsenal. Improved training, enhanced quality assurance, and additional generally accepted protocols resulted in questioned document examinations becoming very reliable. (Although, as stated in the 2009 National Academy of Science Report (NRC 2009), "the scientific basis for handwriting comparison needs to be strengthened").

Firearms/Toolmarks

While interest in the identification of the firearm from which a bullet has been fired began before the nineteenth century, it was not until the beginning of the twentieth century that an important discovery was made – that microscopic marks ("striations") on bullets of the same caliber and type, fired through guns of different makes, varied in appearance. This critical observation was published in 1900 by Dr. Albert Llewellyn Hall in the Buffalo (New York) Medical Journal. One of the techniques for the examination of bullets used was to roll them on a plane surface of wax or lead foil to reproduce surface markings; however, only the gross markings could be reproduced.

The first use of striation matching did not involve bullets but rather knife marks in the wood of vandalized trees. R. Kockel, a Professor at the Institute of Legal Medicine in the University of Leipzig, published two papers in 1900 and 1903 describing a method for producing test marks with the knife and for using oblique lighting and photography for the comparisons.

The first case in the USA in which identifications of ammunition components were made occurred in what became known as the "Brownsville Affray" in which a civilian was killed in Brownsville, Texas, in 1906 by shots fired by several soldiers during a riot. In 1907, staff of the Frankford Arsenal in Pennsylvania were able to identify eleven of the fired cartridge cases to one of the soldiers' rifles, eight to a second, eleven to a third, and three more to a fourth. This pioneering work remained relatively unknown for many years because the only publication of it was in the Annual Report of the US Army Chief of Ordnance.

In 1912, Victor Balthazard, a Professor of Legal Medicine in the University of Paris, described the principles involved in linking a particular fired bullet to an individual firearm. His procedure involved making a detailed sequential series of photographs of the surface of the complete circumference of the bullets, enlarging the photographs and comparing the striations revealed in the process. In 1922, he published "*Identification des projectiles. Perfectionnement de la technique*" in *Annales de Médecine Légale*. Unfortunately, the system was too cumbersome to ever gain wide acceptance.

Between 1919 and 1923, Charles E. Waite, an investigator in the New York State Attorney General's Office, accumulated rifling data for almost all the firearms manufactured in the USA and Europe. In 1925, he persuaded John H. Fisher, a precision instrument designer, Phillip O. Gravelle, a microscopist at Columbia University, and Major (later Colonel) Calvin Goddard, to join him in establishing the "Bureau of Forensic Ballistics" in New York City. Goddard, a graduate in medicine from Johns Hopkins, had served in the Army Medical Corps but, following World War I, had transferred to the Ordnance Corps. Gravelle, who had used a comparison microscope in work with textiles, had the inspiration to suggest applying it to the comparison of fired bullets and cartridge cases. He therefore assembled a hybrid instrument suitable for examining ammunition components. Initial tests of the equipment by Goddard in April 1925 confirmed its value, and the comparison microscope subsequently became the universal tool of firearms/toolmark examiners worldwide. This technique continues to be used in much the same way that it was used by Goddard.

In 1925, Goddard published the first description of the use of the comparison microscope for firearms identification in the November/December issue of "Army Ordnance" under the unfortunate title "Forensic Ballistics" – "unfortunate" because, from then onward, "Ballistics" became the popular – yet incorrect – name for the work of the forensic firearms examiner. Over the next several years, Goddard was contacted by many experts, including Professor J. Howard Mathews of the Department of Chemistry in the University of Wisconsin, Sydney Smith (then still in Egypt), Derôme in Montreal, and Locard in France. Derôme published a book on the subject (Derôme 1929) and Professor Mathews produced a massive three-volume series "Firearms Identification," the first two volumes in 1962 and the third (posthumously) in1973.

Confirmation of Goddard's stature as the preeminent firearms expert in the USA came with the infamous "St. Valentine's Day Massacre" in Chicago in 1929. He was asked to examine the 70 fired .45 caliber cartridge cases and 14 bullets recovered from the scene. His conclusion was that 50 of the casings had been fired from one Thompson submachine gun and 20 from another. Later that year, two "Tommy" guns were located in a house owned by a member of the Al Capone gang and identified as having fired the bullets.

There has been little fundamental change in the examination of firearms and toolmarks since the early work of Goddard and others. Attempts to bring greater objectivity to identifications, such as that by Alfred Biasotti with his publication "A Statistical Study of the Individual Characteristics of Fired Bullets" (Biasotti 1959), were initially received coolly by practitioners. Toward the end of the century, however, Biasotti's concept of "Consecutive Matching Striations" began to receive greater acceptance by the discipline.

In 1991, Walsh Automation Inc. of Montreal developed the "Integrated Ballistics Identification System" (IBIS) for scanning the surface of bullets and cartridge cases using sophisticated laser and computerized digital imaging techniques. This system was adopted by the Alcohol, Tobacco, Firearms and Explosives (ATF) Bureau. In 1993, Mnemonic Systems, under contract to the FBI, developed a similar system "DRUGFIRE" for fired cartridge cases. The two systems were integrated in 1999 into the National Integrated Ballistics Information Network (NIBIN) operated by ATF. While highly valuable for developing and searching databases, final identifications continue to be made visually by examiners using the comparison microscope.

Similar to the situation with fingerprint identification, the delivery of firearms examination services developed in the latter half of the century, primarily within law enforcement agencies. Examiners frequently had little or no scientific background and were ineligible for membership in professional organizations such as the American Academy of Forensic Sciences (AAFS). They therefore established their own organization, the "Association of Firearm and Tool Mark Examiners" (AFTE), in 1969. AFTE has created valuable criteria for operational aspects such as training, examination protocols, and terminology. In 1972, the AFTE Journal was established and became the principal vehicle for publication of research results. Most of this research, however, relates to practical developments rather than the fundamental science of the process. Again, the NAS Report commented that "Individual patterns from manufacture or wear might, in some cases, be distinctive enough to suggest one particular source, but additional studies should be performed to make the process of individualization more precise and repeatable."

Forensic Biology

By the end of the nineteenth century, forensic scientists could determine that a stain was blood but could not answer the follow-up question – "Is it human blood?" That answer was not long in coming. In 1901, Paul Theodor Uhlenhuth, an Assistant Professor in the University of Greifswald in Germany, published "A Method of Investigation of Different Types of Blood, Especially the Differential Diagnosis of Human Blood" in the German journal *Deutsche Medizinishe Wochenschrift*. Uhlenhuth had observed that when chicken egg protein was injected into rabbits, the blood of the rabbit would subsequently react with the chicken protein, causing it to precipitate, but would not react with other types of bird protein, that is, it was specific for chicken. This reaction became the basis of the "precipitin test." Uhlenhuth soon expanded his research to develop antisera that would react with the protein of humans and other species. In his paper, Uhlenhuth modestly foresaw that his process "could prove to be of great importance in medical jurisprudence."

Although blood could now be identified as human, it might (theoretically) be from any human. Progress toward reducing the number of potential human sources began almost immediately. Later in 1901, Karl Landsteiner, an Assistant Professor at the Institute of Pathology and Anatomy in the University of Vienna, published a paper entitled "On Agglutination Phenomena of Normal Human Blood." It began with: "Some time ago I observed the fact that blood serum of normal persons can frequently agglutinate the red blood corpuscles of other healthy individuals." His eventual conclusion was that human blood was either Type A, B, O, or AB. The first step toward demonstrating the biological individuality of humans, as observable in blood, had been taken.

Landsteiner's test was based on antigens on the cells and, in 1915, Leon Lattes at the Institute of Forensic Medicine in the University of Turin in Italy, developed the first serum antibody test for the ABO blood groups. Lattes's procedure continued to be used by many forensic serologists for the remainder of the twentieth century.

Blood is not the only body fluid of importance in forensic science; semen and saliva are also common evidence materials. Identification of saliva is usually based on tests for its amylase activity. Semen can be identified by the microscopic detection of spermatozoa, a process first described by Henri-Louis Bayard in France in 1839. A presumptive color test for semen based on its high acid phosphatase content was developed in 1945 by Frank Lundquist in the University of Copenhagen. In 1925, a Japanese scientist, Saburo Sirai, discovered that the ABO system could also be detected in these other body fluids in about 80 % of the population who are referred to as "secretors."

Although the ABO system was widely used in forensic serology, its discriminating power (e.g., about 45 % of the population are group O) could establish exclusion of possible sources but was of only limited value for inclusion. Forensic serologists therefore continued to look for other "genetic markers" which could enhance their ability to discriminate between sources. This search led them in the 1950s and 1960s to a multitude of genetically controlled systems in blood including plasma proteins and variants of some of the enzyme systems. While the biological functions of these systems were important to medical researchers, genetic variants within any system were not. For this reason, identification and validation of these variants had to be performed almost exclusively by forensic scientists, primarily in the UK at the Metropolitan London Police Lab and the Home Office CRE.

The first of these polymorphic ("many forms") systems to be introduced was the phosphoglucomutase (PGM) system, discovered in 1964. Brian Culliford, of the Met Lab was able to apply it to bloodstains in 1967. PGM was particularly useful because it was subsequently shown to be present in semen, vaginal secretions, and hair roots while most of the later systems could only be applied to blood. Using Culliford's original procedure, three different PGM polymorphs could be identified. Later work in 1977 at CRE using a different process permitted the observation of ten different PGM types thereby vastly increasing its discriminating power. By the late 1970s, nine different polymorphic enzyme systems had been validated. In addition to the enzymes in red blood cells, some proteins in serum were also found to exhibit polymorphism. Not all of these systems became widely used, but they collectively provided a useful "toolbox" for the forensic serologist (Culliford 1971).

The incentive driving the development of additional blood group systems was the fact that each is inherited independently of the others. As independent variables, the population frequencies of each could be multiplied resulting in sometimes relatively low frequencies of occurrence for bloodstains. So, by the mid-1980s, forensic serologists had come a long way in their ability to exclude possible donors of a bloodstain.

As genetic variants, the blood group systems are manifestations of their genetic control by DNA. Why not, instead, look directly at the DNA? In forensic science, the ability to do so began in March 1985 when Alec J. Jeffreys (now Professor Sir Alec after being knighted in 1999) and two colleagues at the Department of Genetics in the University of Leicester in England published their seminal paper in Nature (Jeffreys et al. 1985). In that paper, they described regions along the length of the DNA molecule which were highly variable from one person to another. In the conclusion of that paper, Jeffreys wrote – in a massive understatement – "We anticipate that these DNA 'fingerprints' —— can be used in forensic applications ——" thereby coining a name which, although like "ballistics" before it was incorrect, became popular in the media.

The first forensic application of Jeffreys' process came in 1986 when a young man, accused of the rape/murder of two young girls near Jeffreys' laboratory in Leicester, was exonerated by his DNA and Jeffreys concluded that one other man, then unknown, was the source of the semen in both victims. After a massive investigation which came to be known as "The Blooding," Colin Pitchfork was eventually identified through his DNA and was convicted. As a result of this case, Peter Gill and David Werrett at the FSS became actively involved in the development of DNA profiling, and the FSS became a world leader in this field.

The analytical process Jeffreys used was "restriction fragment length polymorphism" (RFLP), which became the first technique used for "DNA profiling" in forensic laboratories in the late 1980s and 1990s.

In North America, the FBI quickly became interested in DNA and, through close cooperation with other forensic labs in the USA and Canada, developed standards of practice which became widely accepted. The first RFLP cases in North America were in 1987/1988.

The RFLP process, while highly discriminatory, was extremely labor intensive, very demanding in the size and condition of the sample, and required several weeks to carry out. In 1983, Kary Banks Mullis at the Cetus Corporation in California conceived the "polymerase chain reaction" (PCR) process which dramatically changed the face of molecular biology including forensic DNA profiling (Mullis et al. 1986). PCR uses small synthesized pieces of DNA (primers) to make a million or more copies of short but highly informative regions of the DNA molecule. As a result, much smaller samples in poorer condition can be profiled. Most forensic applications of DNA analysis are now based on PCR technology using "short tandem repeats" (STRs) of the molecule, introduced in1992 by Thomas Caskey of Baylor University in Texas.

DNA profiling represents what is unquestionably the most dramatic advance in forensic science of the twentieth century. In a period of only a few years, forensic science had gone from reporting frequencies of occurrence ranging from one in a few hundreds for blood, to frequencies of one in many millions or even billions – in effect individualization – for all body fluids. The dream of forensic serologists – individualization based on stains of body fluids – had become, virtually, a reality.

Other Developments

Until about the middle of the twentieth century, communication of developments in forensic science was primarily through publication in scientific journals, of which there were few devoted to forensic science. As a result, exchanges of information and discussion were limited and slow. In 1948, Dr. Rutherford B. Gradwohl, Director of the St. Louis Police Department Laboratory, invited 150 forensic scientists to attend a meeting in St. Louis which led to the formation of the American Academy of Forensic Sciences (AAFS). The Academy has since became the preeminent forensic science organization in the world and, in 2011, had over 6,000 members from 65 countries. The Academy's Journal of Forensic Sciences, established in 1956, has become a major peer review vehicle for the exchange of scientific data and research.

In the UK, Stuart Kind of the FSS led the creation of The Forensic Science Society in 1959. The Society's journal, now called "Science and Justice," is another important vehicle for communication among forensic scientists. Such organizations also facilitated communication by personal contact which was then dramatically enhanced with the development of the Internet.

When Dr. Briggs White became the Director of the FBI Lab in 1970, he recognized the need for closer cooperation with state and local labs. In 1973, he invited 30 US and Canadian lab directors to a meeting at the FBI Academy in Quantico, Virginia. This led in 1974 to the birth of the American Society of Crime Laboratory Directors (ASCLD) devoted to management and policy matters. ASCLD has gone on to play a leading role in the improvement of forensic labs in North America. In Europe, an organization with similar goals and objectives, the European Network of Forensic Science Institutes (ENFSI) was established in 1995.

ASCLD played a critical role in a major initiative which has contributed significantly to dramatic improvements in quality assurance in forensic science – laboratory accreditation. In the late 1970s, ASCLD recognized the need for objective standards for crime lab operations. A committee began studying the matter and drafting a series of objective criteria which lab directors could use to evaluate their laboratories. What emerged in 1981 was the American Society of Crime Laboratory Directors/Laboratory Accreditation Board (ASCLD/LAB). The first laboratories were accredited in 1982 and, in1988, ASCLD/LAB was incorporated as a totally separate entity from its parent. Recognition of the value of accreditation gradually expanded and there now are almost 400 forensic laboratories in the USA and around the world which have been accredited by ASCLD/LAB. Accreditation of forensic labs in other countries is done by national accrediting bodies such as UKAS in the UK, NATA in Australia, and the Standards Council of Canada.

Another major contributor to the enhancement of quality in forensic science is a "spin off" from the introduction of DNA profiling. The FBI recognized the enormous impact that this new process was likely to have and realized that strict standards of performance would be required to ensure acceptance of the evidence. This could best be accomplished through collaboration of many experts, and so it established the Technical Working Group on DNA Analysis Methods (TWGDAM) consisting of representatives from the fields of forensic, industrial, commercial, academic, and international communities. The name was subsequently changed to SWGDAM ("Technical" was changed to "Scientific"). The ongoing result of the development of these standards has led to much greater uniformity and consensus on the part of practitioners. SWGDAM was so successful that the concept was applied to other forensic science disciplines, and there now are about 20 "SWGs" developing recognized standards and protocols (Cormier et al. 2005).

Conclusion

Despite the many advances in forensic science in the past two centuries, there remain some fundamental limitations as described in the NAS Report. This, of course, does not mean that the procedures are invalid or unreliable but rather that there is a need to more adequately demonstrate their validity through conventional empirical data, peer review and publication. Much of the necessary data already exists because it is collected by every trainee during their training process to validate for themselves the reliability of the techniques and their ability to use them. Unfortunately, these data are scattered in thousands of widely distributed and uncoordinated laboratory training files and therefore not readily accessible; more systematized data collection is required to meet the required standards. This work is being done – in the twenty-first century!

Related Entries

- ▶ Automated and Manual Forensic Examinations
- ▶ DNA Profiling
- ▶ Fingerprint Identification
- ▶ History of Forensic Science in Policing
- ▶ Identification and the Development of Forensic Science

Recommended Reading and References

Ashbough D (1999) Quantitative-qualitative friction ridge analysis: an introduction to basic and advanced ridgeology. CRC Press, Boca Raton

Barnes JG (2011) History. In: McRoberts A (ed) The fingerprint source book. NIJ OJP, Washington, DC. Chap. 1

Biasotti A (1959) A statistical study of the individual characteristics of fired bullets. J Foren Sci 9:428–433

Cormier K, Calandro L, Reeder D (2005) Evolution of the quality assurance documents for DNA laboratories. Foren Mag 2:1–3

Côté J (2003) Wilfred derôme: expert en homicides. Les Editions du Boréal, Montréal

Culliford B (1971) The examination and typing of bloodstains in the crime laboratory. National Institute of Law Enforcement and Criminal Justice, Washington, DC

Dalrymple B, Duff J, Menzel E (1977) Inherent fingerprint luminescence – detection by laser. J Foren Sci 22:106–115

Derôme W (1929) Expertise en armes â feu. Beauchemin, Montréal

Faulds H (1880) On the skin – furrows of the hand. Nature 22:605

Galton F (1892) Finger prints. MacMillan, New York

Gunn A (2006) Essential forensic biology. Wiley, Chichester

Hamby J, Thorpe J (1999) The history of firearms and toolmark identification. AFTE J 33:266–284

Henry E (1900) Classification and uses of finger prints, 4th edn. George Routledge, UK, London

Houck M (2003) Trace evidence analysis. Academic, Burlington

Houck M, Siegel J (2010) Fundamentals of forensic science. Academic, Burlington

Jeffreys A, Wilson V, Thein S (1985) Hypervariable "minisatellite" regions in human DNA. Nature 314:67–73

Jickells S, Negruz A (2008) Clarke's analytical forensic toxicology. Pharmaceutical, London

Kelly J, Lindblom B (2006) Scientific examination of questioned documents. CRC Press, Boca Raton

Kirk P (1963) The ontogeny of criminalistics. J Crim Law Criminol Police Sci 54:235–238

Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H (1986) Specific enzymatic amplification of DNA In Vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol 51:263

National Research Council of the National Academies of Science (2009) Strengthening forensic science in the United States: a path forward. National Academies, Washington, DC

Osborne A (1910) Questioned documents. Lawyers' Cooperative Publishing, St. Paul

Schuller P (2003). Mateo orfila: a biography. National Library of Medicine http://www.ncbi.nlm.nih.gov/pubmed/15027707. Accessed 02 April 2012

Siegel J (ed) (2000) Encyclopedia of forensic science. Academic, London

Smith S (1951) History and development of forensic medicine. Brit Med J 24:599–607

Tilstone W, Savage K, Clark L (2006) Forensic science: an encyclopedia of history, methods, and techniques. ABC-CLIO, Santa Barbara

Widmark E (1922) Eine mikromethode zur bestimmungen von äthylalkohol im blut. Biochem Z 131:473–484

Widmark E (1932) Principles and applications of medicolegal alcohol determination. Translated from the 1932 German edition: (trans: Baselt RC). Biomedical Publications, Seal Beach, CA, 1991, p 1–163

Forensic Science Professions

James Robertson
National Centre for Forensic Studies (NCFS),
Faculty of Education, Science, Engineering,
Technology and Maths, University of Canberra,
Canberra, ACT, Australia

Overview

The use of forensic science used in criminal investigation brings the occupations of forensic science and policing into an immediate and interactive relationship.

Regardless of the organizational home of forensic science within or outside of policing organizations, the success or failure of forensic science in support of investigations depends on how well this interaction is managed.

This entry proposes that the tensions which can sometimes be seen in the current relationship would be reduced if both "occupations" more ideally met the characteristics of a true profession. The paper argues that in the modern world we need to look forward in developing a professional project which recognizes the increased role of the state in the regulation of professions.

Key Issues and Future Directions

Introduction

> *"We trust our health to the physician; our fortune and sometimes our life and reputation to the lawyer and attorney. Such confidence could not safely be reposed in people of a very mean or low condition. Their reward must be such, therefore, as may give them the rank in society which so important trust requires."*
> (Adam Smith, Wealth of Nations, BK I p. 10. As quoted in Macdonald 1999)

To the above should we add, we trust our liberty, safety and just outcomes to our police and law enforcement services, and if so, what role or contribution is made or should be made, by forensic science? Do we have confidence in policing and/or forensic science and do we accord them the "rank" which such trust requires? Adam Smith would not have thought of policing when he wrote Wealth of Nations in 1766, some 63 years before Robert Peel established the Metropolitan Police Force in the United Kingdom. Dating the "formal" start of forensic science will no doubt be dealt with elsewhere in this series but practically it is an invention of the late nineteenth and early twentieth centuries, or the modern world.

The purpose of this entry is to explore the relationship between policing and forensic science, focusing on how these two "occupations" have developed and what implications this has for that working relationship. The paper will propose that although both occupations will lay some claim to being professions they fall well short in critical areas of full professional status. It will further propose that "confidence" and "trust" of both policing and forensic science would be enhanced if both met more ideally a professional model. Macdonalds' (1999) concept of a professional project will be used to discuss how the latter might be achieved.

What Differentiates a Profession from an Occupation?

Although there is no single agreed definition of what constitutes a profession all dictionary and encyclopedia definitions of the word "profession" stress key common criteria which invariably include:

- An occupation or vocation requiring, or founded upon, specialist training or education, and,
- The body of persons in such an occupation.

In this sense any occupation has the potential to become recognized as a profession provided there is a specialist body of knowledge which can in some way be defined, codified and persons educated and trained.

The journey from occupation to profession, or "professional project," typically involves milestones being achieved such as:

- It becoming a full time occupation
- The establishment of formal training and education

- Local and national associations being formed, and,
- A code of professional practice being established.

The national association would commonly develop professional rules and structures aimed at establishing a high degree of autonomy and self-regulation. This concept of professional autonomy is based on three claims that:

- The work of professionals entails such a high degree of skill and knowledge that only fellow professionals can make accurate assessments of professional performance
- Professionals are characterized by a high level of selflessness and responsibility such that they can be trusted to work conscientiously, and,
- In those rare instances in which individual professionals do not perform with sufficient skill or conscientiousness, their colleagues may be trusted to undertake proper regulatory action.

(Velayutham and Perera 1995)

Included in the regulatory role of many professional bodies will be some measure of control over qualifications as a pre-requisite for professional practice and a code of practice. It is now well recognized that another hallmark of a profession is the recognition of the need for continuing development and lifelong learning. Finally, professions would "profess" to provide their services in a disinterested way and with an ethic of service to others.

Individuals may purport to deliver their services in a professional way, or meet agreed standards, but this in itself is not sufficient for an occupation to be considered a profession.

Forensic science is delivered by persons with science qualifications but also includes important roles for technicians. The US Department of Labor (Anon 2010) describes technicians as follows.

> *Science technicians use the principles and theories of science and mathematics to assist in research and development and to keep, invent and improve products and processes. However, their jobs are more practically oriented than these of scientists.*

The US Department of Labor also stress that technicians *"work under the direction of scientists"* and that an appropriate level of qualification is a 2 year or associate degree, although of interest is that it states biological and forensic science technicians usually require a bachelor's degree.

Other definitions of technician draw out the key differences between a scientist and technician as being

- Level of entry qualification (although this is increasingly becoming blurred!)
- Technicians will usually have a vocational qualification
- Technicians have a more practical understanding which will usually include a more in depth understanding of specific techniques than a scientist
- Technicians may have their own work structure and hierarchy but generally will come under the supervision of a scientist.

Technicians will often have a formal representative body or association which may even offer formal certification but these groups would not normally be considered professional groups although meeting many of the tenets of a profession.

Forensic science is characterized by its multidisciplinary nature and many commentators would consider some aspects of forensic science are in fact more technical than scientific. For example, are crime scene examination and some aspects of fingerprint examination technician work or scientific work? In the context of this entry *all* aspects of forensic work will be considered capable of either meeting the requirements for professional status or evolving towards this status.

The study of professions has attracted considerable attention from social scientists. It is beyond the scope of this entry to explore in any detail the history of the sociological study of the professions but the reader is referred to two key books on this subject by Larson (1977) and Macdonald (1999). In the context of policing Rohl (1990) suggested that Australasian police should adopt Witham's (1985) "attribute" approach which defines a profession around eight characteristics. These are,

1. Operates as an organized body of knowledge, constantly augmented and refined,
2. Involves a lengthy training/education period,

3. Operates so as to serve its clients best,
4. Operates autonomously and exercises control over members,
5. Develops a community of practitioners through professional standards,
6. Enforces a code of ethics and behavior,
7. Establishes uniform standards of practice, and,
8. Provides full professional mobility.

An alternative approach is that of Klegon (1978) who proposes two aspects to define a profession, an internal dynamic and an external dynamic. The internal dynamic relates to the strategies used by the profession such as establishing a code of practice whilst the external dynamic relates to the economic, political, social and intellectual influences which cannot be the exclusive domain of the "profession."

Allen (1991) has grouped Klegon's strategies as follows

1. *Formulation of a code of ethics:* once formulated, this code must be promoted as a symbol of the desire of the profession to serve the public.
2. *Delineation of the area of expertise:* exclusive domain of practice must be identified and protected from encroachment, the knowledge base of the profession validating the professional claim.
3. *Control of education and entry:* education should be university based and closely monitored by the professional association(s). Further education might be required before entrance to the association(s).
4. *Definition of competence levels:* promotion of differing classification of membership with reward in prestige and status for those who attain higher levels of expertise.
5. *Determination of standards:* utilizing their autonomy, the profession will determine their own standards of practice.
6. *Image building:* public promotion of positive image of the profession and by convincing the public of its professionalism, the occupation will be rewarded with professional recognition and status.
7. *Professional unification:* the profession must be united as factionalism undermines public confidence.
8. *Achieving a relationship with the State:* a balanced relationship with the State is required and achieved through legislative recognition and registration.

The final point above is an important contemporary issue for professions. In order to achieve a monopoly, or at least licensure and control, a profession must have a special relationship with the State, sometimes called the **regulative bargain**. Macdonald (1999) makes the point that even when a profession has an apparent monopoly it may still have to compete in the market against others who can provide similar, substitute or complimentary services. It must therefore at least defend and probably enlarge the scope of its activities or its **jurisdiction**.

This will be discussed further in relation to both policing and forensic science.

Medicine and the Law: Classical Professions: Can Policing and Forensic Science Learn from these Professions?

Medicine and the law are often quoted as examples of professions against which any occupation aspiring to achieve professional recognition should be compared. In this context it is important to more fully understand how medicine and the law achieved professional status as "classic" disciplines in a more modern era. The reality is that this is far from a simple story and often the simplified version is that of the development of the Anglo-American model. Macdonald (1999) analyses in depth the emergence of medicine and law as professions in Britain, the United States of America, Germany and France against a matrix of their history and different state systems. The striking, fact is that the evolution of both "professions" is very different indeed across even these four countries. Hence, it is a gross oversimplification to consider that there is a global shared professional model for either medicine or the law.

For example, Macdonald argues that the underlying canvass against which medicine and the law emerged in Britain goes back to an ancient and historically low level of state involvement and bureaucracy which meant the extension of state power over the institutions of civil society emerged relatively late in Britain. Although both professions are linked to the state, their high level of autonomy originates in practices and forms that go back to the Middle Ages and which had not been subordinated to the "civil service." Importantly medicine and the law in Britain were able to regulate their own education and training and, hence, control entry to their professions. The role of academic institutions by contrast has not been controlling but merely delivery of what the professions deem necessary as an "entry" to their professions.

The evolution of medicine and the law in the United States is more complex and heavily influenced by a populist climate and a desire to open knowledge based occupations post-independence to all. For example, Macdonald (1999) states that *"in response to these state pressures legislatures deprived the medical societies of their licensing powers and American medicine was reduced to a condition of factional disorganisation*." (at p. 83)

Because medicine and the law were deprived of their organizational base and traditional values seen in Britain they had to rely on their expert knowledge as a basis for claim to special recognition. The lack of an organization also meant the universities played a much more important role in passing on knowledge than was the case in Britain.

In the late 1880s as professional bodies reappeared *"the chief bases of their 'professional project' were their modern knowledge and the institution in which it was lodged, the university."* (Maconald 1999 at p. 84)

If the Anglo-American model had important differences then the evolution of the medical and law professions in France and Germany had even greater differences.

In France the state had a much higher degree of control and responsibility which reduced professional power and increased the influence of universities. Similarly in Germany universities were long established and status was focused on the "educated middle class" or Bildungsburgertum. This was in turn strongly associated with the civil service and academia rather than with the professions.

In summary, Macdonald concludes that "*the political ideology in America is a compound of nationalism, liberalism and populism while in Britain traditionalism replaces populism and colours the other two components. In France and Germany these elements are also to be found, but in different proportions and with the important addition of a centralised state, less penetrated by civil society*." (at p. 97)

Hence, it is probably unhelpful to refer back to the "classic" form of knowledge and development of professions for a "modern" society. Those occupations seeking to *now* achieve professional recognition need to understand that specialist knowledge is now more complex and more organized into disciplines (arguably more divided). There is now a general reluctance for the state to grant a special position. However, the state may see value in using professional bodies as an instrument of regulation.

Knowledge and the Professions

Adam Smith, quoted at the opening to this entry, was one of the leaders in the "enlightenment." A consequence of this movement was the exponential growth in knowledge and so called **cognitive growth** (Gellner 1988, p. 116). In premodern society knowledge was tightly controlled, owned and difficult (even dangerous) to question. Professional knowledge in the modern world is more complex. In one sense knowledge is now open to all, irrespective of status. Knowledge is no longer under the "guardianship" of a few. In this changed paradigm professionals increasingly will rely on certification and credentialing of higher level qualifications to underpin their claims to being a profession. Macdonald (1999, p. 163), proposes that to understand professions the starting point must be to define **professional work**. The context of professional work, the control of that work differentiating in types of work and the notion of jurisdiction, previously mentioned, are all important in the profession laying claim for its work. However, he argues that the quality that characterizes professional work at a knowledge level is the concept of **abstraction**. Abbott (1988) captures this in the following quote "*. . .abstraction is*

Forensic Science Professions, Fig. 1 A working theory of the professions: a conceptual outline (From Macdonald 1999)

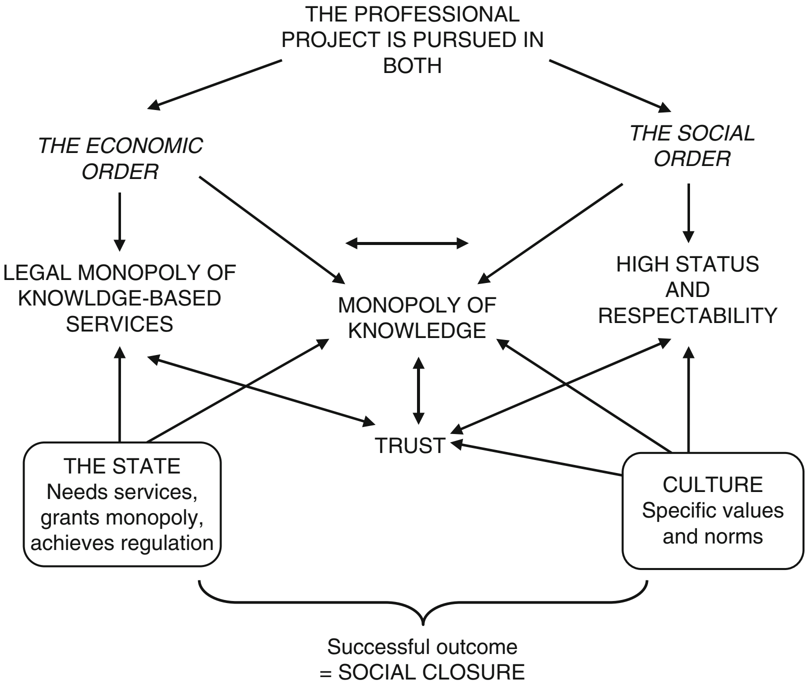
THE PROFESSIONAL PROJECT IS PURSUED IN BOTH
THE ECONOMIC ORDER
THE SOCIAL ORDER
LEGAL MONOPOLY OF KNOWLDGE-BASED SERVICES
MONOPOLY OF KNOWLEDGE
HIGH STATUS AND RESPECTABILITY
TRUST
THE STATE Needs services, grants monopoly, achieves regulation
CULTURE Specific values and norms
Successful outcome = SOCIAL CLOSURE

the quality that sets interprofessional competition apart from competition among occupations in general. Any occupation can obtain licensure (e.g. beauticians) or develop a code of ethics (e.g. real estate). But only a knowledge system governed by abstractions can redefine its problems and tasks, defend them from interlopers, and seize new problems – as medicine has recently seized alcoholism, mental illness, and hyperactivity in children, obesity, and numerous other things. Abstraction enables survival in the competitive system of professions." (1988, p. 9)

However, the concept of professional work needs to be more than about technique. In the modern world the possessor of knowledge and technique adds value through the exercise of **professional judgment**. The ever increasing move to standardization and codification has benefits and value and arguably, makes knowledge more accessible to a more educated public. However, taken to extremes (perhaps the Crime Scene Investigation or CSI effect?), this can undermine the role of professionals to apply appropriate cognitive processes encapsulated as professional judgment.

The Concept of a Professional Project

The professional project is the term Macdonald uses (1999) to present a conceptual model for the professions (Fig. 1).

This model brings together the matrix of factors which any occupation seeking to achieve recognition as a profession must consider. Policing and forensic science are assessed against this model and important differences identified which have implications for how these two occupations interact in a broader criminal justice system.

Toward a Policing Professional Project

The professionalization of police has been the subject of an enormous international literature and it is simply out of scope to review this literature in this entry. As was seen for medicine and law even the Anglo-American "story" of the development of these professions shows two quite distinct pathways. Similarly with two continental European countries they showed significant differences.

Hence, international interest in professionalization of policing will not necessarily translate into a universal model. In this regard the professional project is a more useful concept as it recognizes in particular the role of the state and culture as key drivers. These influences will be significant variables at a national level. For example, the role of police is influenced by the legal system and how it operates.

The starting point for any profession is defining a body of knowledge. It can be argued that there is sufficient specialist knowledge relevant to policing to constitute a body of knowledge although there is no broad agreement as to the academic level of that knowledge. It is certainly not the case that even a vocational qualification is a pre requisite for entry to policing in many countries. In Australia the then Australasian Police Ministers Council (APMC) in 1990 agreed to five major outcomes necessary to attaining full professional status. One of these was that national education standards and formal higher education qualifications were necessary for full professional status.

In 2001 the Australasian Police Professional Standards Council (APPSC) was tasked with advancing police professionalization. Suffice to say that although some jurisdictions in Australia have committed to graduate level entry, and others support an associate degree or vocational equivalent, there is still no general acceptance that police studies (*and* hence knowledge) are sufficiently distinct to form a necessary body of specialist knowledge. Stone and Travis (2011) in a paper titled "Toward a New Professionalism in Policing" state that "*for any profession to be worthy of that name, its members must not only develop transportable skills but also commit themselves both to a set of ethical precepts and to a discipline of continuous learning.*" Further, they observed that if the values of policing are really professional and not local, what is needed is a genuine national coherence in the skills, training and accreditation of police. (Stone and Travis 2011, p. 19).

The title of the above paper does send out a warning and that is not to confuse "professionalism" with developing a profession. As has been discussed previously it is entirely possible to deliver a professional service without meeting the tenets of a profession.

It is yet to be seen if the prediction of Stone and Travis (2011, p. 19) that "*the attraction of the new professionalism is likely to feed a flowering of specialist professional associations, bachelors and masters degree programs*" will be achieved.

Regrettably in 2012 the aspiration of the APMC stated in 1990, of formal higher education qualifications, have not been achieved in Australia. Certainly progress has been made in developing national educational standards but these have been set at a technician, vocational level.

By 2006 yet another report prepared for the Police Commissioners of Australasia identified six key objectives for the attainment of full professional status, these being:

1. Develop a definition of the profession of policing.
2. Implement university-based education for policing.
3. Develop a body of knowledge.
4. Propose ongoing professional development.
5. Develop registration and standards for policing.
6. Establish a professional body for policing.

(Lanyon 2007)

It is disturbing that even in an advanced country such as Australia there is still a need to develop a definition of the profession of policing. A further 6 years has seen little hard progress towards the realization of any of these goals.

It has also been observed that much of the discussion about professionalization of policing takes place at a senior manager and executive level and broad commitment by "rank and file" members is questionable.

Perceptively it has been said that "*occupations that fail to lead the agenda in advancing their own professionalization may well forfeit that role to other more influential agencies*" (Lanyon 2007, p. 122). Following a major review of police leadership and training in England and Wales (2011). Neyroud recommended the creation of a new professional body for policing. This recommendation has been accepted by the UK government who announced in December 2011 their intent to establish a new police professional body in 2012. Although the governance structure for this new body has not yet been made public it seems clear that there will be a greater degree of central government regulation and control of the professional aspects of policing. The Police Minister is quoted as saying that *"it's striking that while doctors, lawyers, teachers and nurses have their own professional bodies, police officers do not."* (Anon 2011a)

Indications are that this new body will have an independent chair from outside the police establishment and that it will eventually receive a royal charter similar to other professional bodies.

As discussed earlier in this entry state interest in professions is likely to increase with greater formal regulation. The lack of genuine professional member organizations in policing does not place the practitioners in a strong position to influence their journey towards full professional status.

Furthermore, policing as an occupation no longer has a monopoly of services with private security and other government enterprises sharing the "market."

It would probably be an exaggeration to say that there is a crisis in policing or with the public view from a status and respectability viewpoint. However, clearly this would vary widely at a global level. However, ongoing issues with police corruption and integrity do nothing to support the move to policing being accepted as a profession and police more often than not are increasingly subject to additional external oversight through independent integrity commissions. In this regard policing also falls short of the "trust" awarded through self regulation to professions.

A final aspect of a profession which is worthy of comment is that of full professional mobility. Implicit in this criterion is that this would be facilitated by mutual recognition of qualifications, training and rank or level. With some exceptions, and at senior police ranks, policing falls well short of full professional mobility.

Towards a Forensic Science Profession

Robertson (2011a) has discussed the issues relevant to assessing whether or not forensic science can meet the criteria of a true profession. In this discussion the professional project model will be used as the template against which this assessment will be made. This follows the same approach as for policing and enables a comparison of the status and progress of these two occupations towards achieving professional status.

The first criterion is that of a body of knowledge. In addressing this criterion it is important to consider what is the relevant body of knowledge. If forensic science is considered to be merely the application of core sciences such as biology, chemistry and physics to addressing science issues which may end up in a legal setting, then that body of knowledge is science knowledge. While this body of knowledge is always to an extent tentative and evolving, nobody would argue that science knowledge is not a sound basis underpinning all science based professions. It can be argued that the forensic dimension is merely context and in this respect forensic science is no different from any other occupation which relies on the application of science. The issue then becomes not one of underlying knowledge but rather one of how well that knowledge is applied and used. In the modern world, with a heavier focus on disciplines, forensic science can be considered to be simply a discipline or group of disciplines.

The US National Academies of Science (NAS) report on "strengthening Forensic Science in the United States: A Path Forward" (Anon 2009) commented that the: "*forensic science enterprise is hindered by its extreme disaggregation – marked by multiple types of practitioners with different levels of education and training and difference professional cultures and standards for performance.*" (at pp. 5–11)

The report also observed that "*of the various facets of under-resourcing, the committee is most concerned about the knowledge base,*" further commenting on the lamentable state of funding for forensic science research.

Crispino (2011) has argued that there is a unique scientific basis for recognizing forensic science as a distinct discipline. Although it can be argued that it is a "chicken and egg" situation the existence of numerous academic programs on forensic science could be put forward as further evidence in support of the proposition that there is a specific body of knowledge. At least some of these education programs pre date the "CSI" driven expansion in forensic programs within universities and some have been around for several decades. A pragmatic view of the world would recognize that forensic science is an academic specialization, is here to stay, and that there are sufficient unifying concepts for it to be considered a distinctive discipline. This is not to deny the need to improve the underpinning knowledge base and the need for a much enhanced research effort. As present there is no statutory requirement for entrants to the forensic profession to have a specific forensic qualification. Whilst there is an emerging trend toward accreditation of forensic education programs and toward individual certification, including commitment to ongoing education and training or continual professional development, these fall short of being mandatory requirements for professional practice. In this sense forensic science can be considered to be at best an immature profession with respect to its knowledge base (and any sense of a monopoly of knowledge) and the related education and academic/research structures.

Although a degree is a requirement of some forensic membership organizations for levels of professional recognition, many forensic occupations either have no educational pre-requisites for entry or these are set at a lower than degree level. In itself this does not preclude forensic science being considered a profession as it is possible to have a technician level within a professional structure.

Perhaps more worrying is that the culture of forensic science falls short of what would be expected of a profession, and specifically, the lack of acceptance, or commitment to research and development. (Robertson 2011b). It has also been observed that at least for some aspects of forensic science the cultural setting is closer to policing than science.

The question arises as to whether or not forensic science as an occupation should be part of a policing organization. The NAS report (Anon 2009) was clear on this point in recommending *"removing all public forensic laboratories and facilities from the administrative control of law enforcement agencies or prosecutor's offices."* (at pp. 5–17)

The factual situation is that there are numerous models and organizational structures at a global level through which forensic science is delivered (Robertson 2012). There is no ideal model and the "state" largely determines its needs, service delivery model and grants either monopoly or creates a market. An example of this is the current situation in England and Wales where the "state" has created a market for the provision of some aspects of forensic science, going as far as withdrawing from any public ownership of forensic laboratory services. However, even here it needs to be recognized that the "state" retains ownership of the majority of forensic science through its police services provision of crime scene, fingerprint and other forensic disciplines. One mechanism introduced in England and Wales to ensure scientific standards are met, has been the establishment of the Office of the Forensic Regulator (Anon 2011). At a level below the "state" level the third party accreditation of forensic services, assessed against relevant standards (most often ISO – 17025), goes someway to regulating levels of professional delivery if not the profession.

As a profession is the "body of persons" in that occupation the role of representative organizations is important. Forensic science societies and organizations play a useful and increasingly important role in "professionalising" the profession. However, these fall well short of the status and authority of similar groups in professions such as medicine, the law or pharmacy. As has been discussed earlier in this entry it is somewhat misleading to compare an evolving profession today, with the structures seen in the "Anglo" experience of pre-modern professions such as medicine and the law. In today's world the "state" is likely to relinquish less autonomy and more likely to demand greater regulation of emerging professions.

Notwithstanding, as with policing, forensic science would benefit from more mature member organizations focused on representing the professional interests of their members. It is perhaps surprising that forensic science has "escaped" the degree of formal regulation applied to more mature professions (see Robertson 2011a, discussing the regulation of Health professions in Australia).

Finally, there are limited barriers to the mobility of forensic science professionals and these relate more to employer barriers than to any science based barriers, provided the "employee" can meet the organizational educational entry requirements. The one exception to this is perhaps in some police services where there is still a requirement for the forensic employee to be a police member.

Conclusions

Forensic science and policing cannot escape the close interaction essential to the investigation of crime, whether or not, they belong to the same or different organizations. This "marriage" may be seen by some as a "marriage made in hell" and by others in a more positive light *but*, like any marriage, it will only work where both parties work together and where the relationship matures and evolves over time. Forensic science is still a relatively new occupation, far less a new profession. Policing is an older occupation but still largely a product of modern times. In this entry it is argued that the relationship between forensic science and policing would benefit from both occupations maturing into genuine professions. The professional project is a useful way to describe the journey ahead should forensic science and policing wish to evolve to full professional status. The practical working interactions between police and forensic science would surely be improved if both occupations more closely met the ideal requirements of a profession with shared core values.

Related Entries

▶ Comparing Police Systems Across the World
▶ DNA Technology and Police Investigations
▶ Expert Witnesses: Role, Ethics, and Accountability
▶ Forensic Science and Criminal Investigation
▶ Forensic Science Culture
▶ History of Forensic Science in Policing
▶ History of the Police Profession
▶ Organizing and Supplying Forensic Science Services
▶ Philosophical Basis of the Forensic Process
▶ Police Leadership Styles

Recommended Reading and References

Abbott A (1988) The system of professions. University Of Chicago Press, London

Allen K (1991) In pursuit of professional dominance: Australian accounting 1953–1985. Account Audit Account J 4(1):51–67

Anon (2009) Strengthening forensic science in the United States: a path forward. National Research Council of the US. National Academics of Science. At http://www.nap.edu/catalog.php?record_id=12589

Anon (2010) Bureau of labor statistics. At http://www.bls.gov/oco/ocos115.htm

Anon (2011a) Home secretary outlines plans for new police professional body. At www.homeoffice.gov.uk/media-centre/press-releases/police-professional-body

Anon (2011b) Codes of practice and conduct for forensic science providers and practitioners in the criminal justice system. Version 1.0. Pub. Crown copyright

Crispino F, Ribaux O, Houck M, Margot P (2011) Forensic science – a true science? Aust J Forensic Sci 43:157–176

Gellner E (1988) Plough sword and book. Collins Harvill, London

Klegon D (1978) The sociology of professions: an emerging perspective. Sociol Work Occup 5(3):259–283

Lanyon JJ (2007) Professionalisation of policing in Australia: the implications for police managers. In: Mitchell M, Casey J (eds) Police leadership and management. Federation Press, Annandale, pp 107–123

Larson MS (1977) The rise of professionalism. A sociological analysis. University Of California Press, London

Lawless C (2010) A curious reconstruction?. The shaping of 'Marketized' forensic science. Discussion Paper no. 63. Pub. LSE Centre for Analysis of Risk and Regulation

Macdonald KM (1999) The sociology of the professions. Sage, London

Neyroud P (2011) Review of police leadership and training. At http://www.homeoffice.gov.uk/publications/consultations/rev-police-leadership-training/

Robertson J (2011a) Forensic science – a true profession? Aust J Forensic Sci 43:105–122

Robertson J (2011b) Truth has many aspects. Sci Justice. doi:10.1016/j.scijust.2011.11.001

Robertson J (2012) Principles for the organisation of forensic support. In: Siegel JA, Saukko P (eds) Encyclopaedia of forensic sciences, 2nd edn. Elsevier, Oxford (at press)

Rohl T (1990) Moving to a professional status. Police J (S Aust) 71(6):6–9

Silverman B (2011) Research and development in forensic science: a review. At http://www.homeoffice.gov.uk/publications/agencies-public-bodies/fsr/forensic-science-review/

Stone C, Travis J (2011) Towards a new professionalism in policing. In New perspectives in policing, NIJ/Harvard Kennedy School. At http://www.ajp.usdoj.gov/nij/topics/law-enforcement/executive-sessions

Velayutham S, Perera H (1995) The historical context of professional ideology and tension and strain in the accounting profession. Account Historians J 22(1):81–102

Witham DC (1985) The American law enforcement chief executive: a management profile. Police Executive Research Forum, Washington, DC

Fraternal Organizations

► History of Police Unions

Fraud Victimization

Kristy Holtfreter
School of Criminology and Criminal Justice,
Arizona State University, Phoenix, AZ, USA

Overview

Generally defined, fraud victimization refers to a form of economic crime. A variety of crimes might be included in this category, but fraud victimization generally takes one of two forms: victimization of organizations, or victimization of individuals. There is considerable evidence indicating that fraud victimization continues to increase, and that losses experienced by victims of fraud greatly exceed those of street crime. Given that many fraud victims are reluctant to report their victimization to criminal justice officials, the problem is likely even more severe than official statistics suggest.

This entry examines key issues in the study of fraud victimization. Definitions and examples of occupational fraud and personal fraud are provided, with special attention to the nature and context of these offenses (e.g., the importance of fraud targeting and the methods of targeting victims). Theory and research on both forms of fraud victimization are presented. In addition to financial losses, other outcomes of fraud victimization (e.g., negative publicity, consumer confidence, reporting behavior, and risk perceptions) are considered. Directions for future research and implications for the prevention of both types of fraud victimization are discussed.

As noted at the outset, fraud victimization can take a variety of forms. Victimization of organizations, which includes occupational fraud, is primarily perpetrated by employees. Victimization of individuals, which includes consumer fraud, may be perpetrated by businesses or other individuals. Each is discussed in more detail in the sections that follow.

Victimization of Organizations

Occupational Fraud Defined

Occupational fraud, defined as “the use of one’s occupation for personal enrichment through the deliberate misuse or misapplication of the employing organization’s resources or assets,” takes place in organizational settings (Association of Certified Fraud Examiners, 2006, p. 6). This definition can be further divided into three categories of occupational fraud: asset misappropriation, corruption, and fraudulent statements. Examples of asset misappropriation include theft of cash or other inventory. Reports indicate that asset misappropriation is by far the most common form of occupational fraud. Examples of corruption include bribery and other forms of wrongful use of influence in business transactions. Finally, examples of fraudulent statements include falsification of an organization’s financial documents or other (i.e., nonfinancial) documents, such as an employee’s resume or credentials. Although the nature of these forms of occupational fraud varies, all are considered violations of trust because they go against an employee's fiduciary duties to the organization. Along these lines, occupational fraud is generally considered to be a form of white-collar crime (Holtfreter 2005).

Theory and Research

Offender Characteristics

The demographic characteristics of offenders who commit occupational fraud are similar to those who have been convicted of other forms of white-collar crime. For example, one study found that men committed slightly more than half of detected occupational frauds (Holtfreter 2005). The same research revealed that most offenders held employee-level positions, as opposed to managers or higher-level executives, were approximately 41 years old on average, and had greater than high school education. The results from this study are comparable to those reported by previous researchers studying white-collar crime in that the sample is best described as "middle class" (see e.g., Weisburd et al. 1991).

Motivation and Opportunity for Fraud

Early studies of white-collar crime identified non-shareable financial problems (e.g., involvement in gambling or other vices) as primary motivations for embezzlement and related offenses in the workplace (Cressey 1953). Many of the most well-known, early studies of fraud and other forms of white-collar crime were conducted when men made up the majority of the American work force. As a result, understanding of these offenses was based almost exclusively on the behavior of males. In perhaps the most comprehensive study on white-collar crime criminal careers, offender motivations mirrored those identified by Cressey, including alcohol addiction and a desire to finance higher education (Weisburd et al. 2001). Subsequent research found that women's rationalizations differed from those of their male counterparts in that women were often motivated by caring for family members (e.g., to pay medical bills for a sick child) (Zietz 1981). More recent research confirms that opportunities to commit fraud are linked to the offender's position within the organizational hierarchy (Holtfreter 2008a). In many large corporations, it is still the case that higher-level positions are held by men.

Victim Characteristics

The available data show that there is no set profile of a "typical" organization victimized by occupational fraud. The breakdown of victims in Holtfreter's (2005) study of 1,142 organizations included privately held businesses (32.1 %), publicly traded companies (30.8 %), government agencies (24.8 %), and nonprofit organizations (11.2 %). Organizational victims ranged in size, including subcategories ranging from 1–99 employees (36.7 %), 100–999 employees (20.3 %), 1,000–9,999 employees (25.8 %) to 10,000 or more employees (17 %). Apart from the employing organization itself, there may also be other, external victims of occupational fraud. For example, some types of fraud (e.g., corruption) may extend to consumers who purchase products from publicly traded companies or private businesses. External victims of occupational fraud are difficult to identify, as they may not even be aware that they have been victimized. On average, it is estimated that organizational victims lose approximately 5 % of their annual revenues to occupational fraud. Factors that predict the amount of losses experienced may vary, depending on the organization (Holtfreter 2008b). To be sure, the type of organization itself may influence the amount lost to occupational fraud victimization, as does the available security measures (e.g., internal audits).

Victim Reporting

Despite the losses experienced by organizational victims of fraud, many organizations remain reluctant to take formal legal action against their perpetrators, and instead choose to handle the offending employees internally. In Holtfreter's (2008a) study, initiating criminal proceedings against the offending employee was more likely among government agency victims. Both the age and education of the offender decreased the likelihood of reporting behavior. Although legal characteristics (e.g., criminal history and crime serious) are associated with legal decision making at later stages of criminal processing (e.g., sentencing), these factors did not appear to play a role in early-stage victim decisions. Among those victims who did not take legal action, fear of negative publicity was the primary reason cited, followed by reaching a private settlement, the perception that legal action was too costly, lack of evidence, and fear of a countersuit.

Future Research and Implications for Crime Prevention

Procedural Justice and Legitimacy

The research on fraud victimization of organizations suggests several directions for future research, as well as some important implications for crime prevention. In particular, more research on employee motivation for occupational fraud is needed. Although a variety of personal motivations outside the organization have been documented, additional research focusing on within-organizational factors would provide more direct insights for employers. For example, perceptions of unfair treatment and/or inequities in pay might contribute to employees' beliefs that fraud victimization is justified (Greenberg 1990; 1993). Recent research confirmed that perceptions of fair treatment and organizational justice are associated with lower levels of misconduct in the workplace (Wolfe and Piquero 2011). Given that this study focused on just one government organization – a police department – it will be important for future research to replicate the findings in other organizational settings.

To help prevent fraud, organizations should examine the effectiveness of their existing control mechanisms, such as audits and other forms of security. Such steps are consistent with routine activity theory's emphasis on target hardening and increasing the presence of capable guardians. To be more proactive in preventing fraud, organizations should also consider more aggressive hiring practices. Along those lines, careful background checks of employment histories, personal references, and criminal records should be implemented for all prospective employees (Holtfreter 2004).

Personal Fraud Victimization

Consumer Fraud Defined

Consumer fraud is a form of personal fraud victimization. Specifically, it consists of "intentional deception or attempted deception with the promise of goods, services, or other benefits that are non-existent, unnecessary, were never intended to be provided, or were grossly misrepresented" (Titus 2001:57). Although the nature of consumer fraud continues to evolve, there is evidence to suggest that this type of victimization has existed since biblical times (Holtfreter et al. 2005). According to the Federal Trade Commission (FTC), nearly one-third of adults in the United States report being targeted by fraud perpetrators on an annual basis, resulting in 11.2 % of the population being victimized (Anderson 2004). The FTC estimates yearly financial losses to be $680 million. The median financial loss experienced by personal fraud victims is approximately $220 per victim (Anderson 2004). These losses far exceed those experienced by victims of common or serious "street" crime. Taken together, these alarming statistics necessitate a closer look at this growing problem.

Before identifying directions for crime prevention, key issues in the study of consumer fraud victimization are discussed. As noted previously, this entry includes the nature of victimization, with special attention to context. Types of fraud and common methods of targeting are examined, followed by theory and research. In addition to financial losses, other outcomes of fraud victimization (e.g., consumer confidence, reporting behavior, and risk perceptions) are considered. This section concludes with a discussion of the implications for the prevention of consumer fraud victimization and the identification of avenues for future research.

Fraud Targeting

As noted above, contact between the victim and offender represents a necessary, but not sufficient, condition for personal fraud victimization to take place. In their attempts to target potential fraud victims, perpetrators use a variety of methods. Consumers are targeted via the same forms of media used by legitimate businesses.

For example, while many transactions still occur face-to-face, technological advances have increased the nature and availability of ways to reach consumers. These include old standbys of telemarketing "cold calls," and television infomercials, as well as via print media sources, such as mail, newspapers, periodical ads, and catalogs. As discussed below, the Internet has become a widely used mechanism for the perpetration of consumer fraud.

Internet Use and Fraud Targeting

More recently, the creation and expansion of the Internet has provided fraudsters with virtually unlimited opportunities to contact potential victims. Over 221 million Americans use the Internet on a daily basis. They do so for many reasons, such as sending e-mails, reading news, entertainment, and for financial purposes (e.g., to manage bank accounts or pay bills). Online shopping represents one of the most common forms of Internet activities. Although consumer fraud scams vary considerably, evidence indicates that perpetrators are increasingly using the Internet to approach their victims. According to the Federal Bureau of Investigation (2001), Internet fraud is: "any fraudulent scheme in which one or more components of the Internet, such as web sites, chat rooms, and e-mail, play a significant role in offering non-existent goods or services to consumers, communicating false or fraudulent representations about the schemes to consumers, or transmitting victims' funds, access devices, or other items of value to the control of the scheme's perpetrators." Governmental and media sources indicate that reports of Internet fraud are on the rise. Fraud targeting and victimization online have only recently begun to be explored within the context of criminological theory.

Theory and Research

Routine Activity Theory

Cohen and Felson's (1979) routine activity theory is one of the most prominent theoretical explanations for victimization. According to this perspective, aggregate changes in opportunity structures, combined with a lack of capable guardianship, will increase the convergence in time and space of motivated offenders and suitable targets. Early research in this tradition focused on the ways in which demographic variables increase risk of victimization. For example, in the context of street crime, males, minorities, and single people have higher risks of victimization. Subsequent studies have suggested that it is the behaviors of individuals, not demographics themselves, that put potential victims into contact with motivated offenders. Along these lines, deviant lifestyles that include activities such as drug use, prostitution, and other illegal behaviors increase exposure to victimization.

More recently, scholars have applied routine theory activity to the study of consumer fraud targeting and victimization (Holtfreter et al. 2008; Reisig et al. 2009; Pratt et al. 2010). Consistent with the theory, researchers have focused on identifying consumer behaviors that reflect "unguarded exposure." The first test of this theory in a fraud context revealed that remote-purchasing activities (e.g., responding to telemarketers, purchasing items from the Internet, ordering products after seeing an infomercial, and ordering products through the mail) increase consumers' risk of being targeted for fraud (Holtfreter et al. 2008). In this study, these activities did not, however, increase victimization. Rather, as discussed in more detail below, the relationship between routine activities and fraud victimization depends on targeted individuals' levels of self-control.

Building on Holtfreter et al.'s (2008) research, Pratt et al. (2010) examined fraud targeting in an Internet-specific context. This study drew on routine activity theory and consumer behavior research to understand how personal characteristics and online routines increased consumers' exposure to motivated offenders. The results indicated that the correlates of online fraud targeting differ from those regularly linked to street crime targeting. For example, young, minority males tend to be both targets and victims of violent crime. In comparison, individuals targeted by Internet fraud are younger and more educated.

Low Self-control

Self-control remains one of the strongest known correlates of crime. In Gottfredson and Hirschi's (1990) original formulation of self-control theory, they argued that individuals who lack self-control are more prone to engage in crime and a variety of crime-analogous acts. Those with low self-control are also shortsighted, impulsive, and tend to pursue their own self-interests without considering the potential consequences of their behavior. Schreck (1999) was the first to extend these ideas to the crime-analogous context of victimization. Along these lines, low self-control increases individuals' risks of victimization. Following Schreck (1999), subsequent research demonstrated considerable support for self-control as a theory of vulnerability for a variety of crimes. Initial applications of self-control theory focused on street crime, where indicators of low self-control are more readily visible to would-be offenders (e.g., someone who appears to be drunk in public) (Schreck et al. 2002).

In the context of fraud victimization, signs of low self-control (e.g., financial risk-taking behavior) are not so easily observed. With this in mind, Holtfreter et al. (2008) examined the relationship between low self-control, routine activities, and consumer fraud victimization. As discussed previously, consumers' routine activities increased their chances of being targeted by fraud perpetrators. Low self-control, on the other hand, did not influence targeting, which was expected in this particular victimization context.

Consumer Confidence

In addition to financial losses, fraud victimization may be associated with a variety of other, and potentially negative, life outcomes. Along these lines, researchers have also begun to examine the influence of fraud victimization on several unique outcomes. As is the case with any type of crime, it follows that the experience of fraud victimization may influence consumers' attitudes toward the police and other legal authorities. Throughout the United States, criminal justice system authorities frequently rely on measures of confidence, support, and trust as indicators of organizational performance and legitimacy (Tyler 2001). Citizen levels of legitimacy are important because they are highly correlated with cooperation with the police and related outcomes (e.g., reporting crime). Reisig and Holtfreter's (2007) study of 918 adult consumers in the state of Florida revealed that less than half of respondents had either "a great deal" or "quite a bit" of confidence in the ability of legal authorities to respond to consumer fraud victimization. This research also showed that younger consumers, the less educated, recent fraud victims, and financial risk-takers all had lower levels of confidence. These results suggest that public awareness campaigns and fraud prevention efforts should target these particular consumer groups.

Risk Perceptions

Reisig et al.'s (2009) study of 573 adult Internet users showed that socially vulnerable individuals and those who were financially impulsive perceived higher levels of risk associated with making online credit card purchases. Those with higher perceived risk levels altered their behavior in two important ways: (1) by spending less time online and (2) making fewer online purchases. However, despite perceiving higher risks, financially impulsive Internet users did not engage in similar risk-reduction strategies. The results of this study suggest that efforts designed to educate Internet users about online fraud victimization may achieve limited success when it comes to reaching financially impulsive consumers. Accordingly, crime prevention measures that are designed to harden targets (e.g., increased Internet security, pop-up blocking software, web encryption, and the like) will arguably have a greater impact on the reduction of fraud victimization (Holtfreter and Holtfreter, 2006).

Reporting Behavior

Despite the dollar losses experienced by fraud victims, there is some evidence to suggest that this form of victimization remains underreported. This is a trend that mirrors that of organizational fraud victims. Copes et al. (2001) examined these issues in a telephone survey of 400 Nashville, Tennessee residents. This research indicated that victims with higher levels of education, and those who were victimized by strangers, were more likely to report their experiences to criminal justice authorities. Older individuals and those who were married were also more likely to report their victimizations. Finally, the amount of money lost was the strongest predictor of victim reporting behavior. Consistent with previous studies, the authors hypothesized that victim self-blame may contribute to underreporting. Unlike the victims of many other forms of crime, personal fraud victims may attribute the cause of their victimization to their own behaviors (e.g., trying to make money on an investment "opportunity"). In turn, the shame associated with self-blame may make victims reluctant to report.

Future Research and Implications for Crime Prevention

General and Specific Population Studies

General population studies continue to be important for a number of reasons (Holtfreter et al. 2006). For example, they are particularly well suited to criminological theory testing, and also permit researchers to calculate estimates of fraud targeting and fraud victimization rates. Perhaps even more importantly, they allow for the identification of consumer segments that may be most at risk for victimization, which is particularly beneficial when it comes to implementing strategies for public awareness and crime prevention. Moving beyond these steps, more in-depth studies of specific, at-risk populations are also useful. Research utilizing college student samples provides insight into crime prevention for this segment of the population. However, less is known about fraud targeting and fraud victimization at the other end of the age distribution: the elderly. While reports from popular media often suggest that fraud perpetrators aggressively target elderly consumers, the research to date has been mixed. To improve understanding of the nature and incidence of fraud targeting and victimization against the elderly, and to better guide crime prevention policy, more comprehensive studies of this population are needed.

Cross-cultural Comparisons

With a few exceptions, the bulk of research on fraud victimization has been conducted in the United States. However, the international research to date tends to confirm many of the findings reported by American researchers. For example, both low self-control and routine activities have received empirical support in non-American countries (van Wilsem 2011; Van Wilsem in press). A recent study using data from 6,201 Dutch consumers found that people with low self-control had significantly higher risks of victimization (van Wilsem in press). In the same study, consumers who actively shopped online as well as those participating in online forums were also at greater risk of fraud victimization. In sum, these findings suggested an indirect relationship between low self-control and fraud victimization in that consumers with low self-control tend to be more involved in risky online activities. Another recent study authored by van Wilsem (2011) examined the links between online activities, offline activities, and different types of victimization. This survey of 6,896 Dutch residents in the general population found support for a reciprocal relationship between distinct routine activities and distinct forms of victimization. Put differently, the results suggested that it is possible for online routines to lead to traditional victimization, and for offline routines to lead to online victimization.

Given the global reach of this problem, this body of research could be extended by including fraud victims from multiple countries within the same study. Such comparative research would contribute to the growing support for a general, integrated theory of fraud targeting and fraud victimization based on self-control and routine activity frameworks.

Victim-Offender Overlap

A recent and promising trend in the fraud victimization literature is the examination of victim-offender overlap. This research suggests that a common underlying trait, low self-control, partially explains the overlap between fraud offending and victimization exposure. These findings are consistent with the research examining the relationship between low self-control and offending as a unique outcome, as well as studies supporting low self-control as a risk factor for victimization. Holtfreter et al. (2010b) study of college students found support for low self-control as a predictor of fraud offending, fraud victimization exposure, and the overlap between these two outcomes. Research by Holtfreter et al. (2010a) also confirmed a link between low self-control and fraud offending. In sum, individuals with low self-control tend to make impulsive decisions, engage in risky behaviors, and are more likely to act on opportunities promising quick returns for little investment. Future research should expand on this work by studying the relationships between low self-control,offending, and victimization in general population samples.

Victim-Offender Relationships

The available literature suggests that crimes of violence (e.g., assault) are most often perpetrated by offenders known to the victims. In comparison, the research suggests that the nature of many forms of fraud targeting and victimization (e.g., online fraud) does not require a personal relationship between victim and offender. Nonetheless, additional insights into the relationships between perpetrators, targets, and victims are clearly needed. Other forms of fraud victimization (e.g., credit card theft) may be more easily perpetrated by offenders who already have an established relationship with the victim (e.g., family members). Given that prior victimization is a risk factor for subsequent victimization, it is also possible that victims of violent crime may be at greater risk for fraud victimization. To be sure, additional research could shed light on these relationships. Future studies should attempt to identify the relationships between targets, victims, and offenders, and also consider whether other forms of victimization are linked to fraud.

Conclusion

It is clear that occupational fraud and consumer fraud result in significant financial losses to organizations, individuals, and in many instances, those who are not directly victimized by these offenses (e.g., the general public). The theoretically informed research literature on both forms of fraud victimization has advanced considerably in recent years, adding to our understanding of these crimes. As technology continues to advance, scholars, criminal justice system officials, and victim service providers, among others, will continue to face challenges as they study and respond to fraud victimization.

Related Entries

- ▶ Crimes of Globalization
- ▶ Lifestyle Theory
- ▶ Managerial Court Culture
- ▶ Monetary Strain and Individual Offending

Recommended Reading and References

Anderson KB (2004) Consumer fraud in the United States: an FTC survey. Federal Trade Commission, Washington

Association of Certified Fraud Examiners (2006) Report to the nation. Author, Austin

Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 52(1):170–183

Copes H, Kerley KR, Mason KA, Van Wyk J (2001) Reporting behavior of fraud victims and black's theory of law: an empirical assessment. Justice Q 18(2):343–363

Cressey DR (1953) Other people's money: a study in the social psychology of embezzlement. Free Press, Glencoe

Federal Bureau of Investigation (2001) Press release: internet fraud investigation "operation cyber loss." Retrieved 31 Oct 2011, from http://www.fbi.gov/pressrel/pressrel101/ifcc052301.htm

Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Greenberg J (1990) Employee theft as a reaction to underpayment inequity: the hidden costs of paycuts. J Appl Psychol 75(5):561–568

Greenberg J (1993) Stealing in the name of justice: informational and interpersonal moderators of theft reactions to underpayment inequity. Organ Behav Hum Decis Process 54(1):81–103

Holtfreter K (2004) Fraud in American organizations: an examination of control mechanisms. J Financ Crime 12(1):88–95

Holtfreter K (2005) Is occupational fraud "typical" white-collar crime? A comparison of individual and organizational characteristics. J Crim Just 33(4):153–165
Holtfreter K (2008a) The effects of legal and extra-legal characteristics on organizational victim decision-making. Crime Law Social Ch 50(4/5):307–330
Holtfreter K (2008b) Determinants of fraud losses in nonprofit organizations. Nonprofit Manag Leadersh 19(1):49–64
Holtfreter RE, Holtfreter K (2006) Gauging the effectiveness of U.S. Identity theft legislation. J Financ Crime 13(1):56–64
Holtfreter K, VanSlyke S, Blomberg TG (2005) Sociolegal change in consumer fraud: from victim-offender interactions to global networks. Crime Law Social Ch 44(3):251–275
Holtfreter K, Reisig MD, Blomberg TG (2006) Consumer fraud victimization in Florida: an empirical study. St Thomas L Rev 18(3):761–789
Holtfreter K, Reisig MD, Pratt TC (2008) Low self-control, routine activities, and fraud victimization. Criminology 46(1):189–220
Holtfreter K, Beaver KM, Reisig MD, Pratt TC (2010a) Low self-control and fraud offending. J Financ Crime 17(3):295–307
Holtfreter K, Reisig MD, Piquero NL, Piquero AR (2010b) Low self-control and fraud: offending, victimization, and their overlap. Crim Justice Behav 37(2):188–203
Pratt TC, Holtfreter K, Reisig MD (2010) Routine online activity and internet fraud targeting: extending the generality of routine activity theory. J Res Crime Delinq 47(3):267–296
Reisig MD, Holtfreter K (2007) Fraud victimization and confidence in Florida's legal authorities. J Financ Crime 14(2):113–126
Reisig MD, Pratt TC, Holtfreter K (2009) Perceived risk of internet theft victimization: examining the effects of social vulnerability and impulsivity. Crim Justice Behav 36(4):369–384
Schreck CJ (1999) Criminal victimization and low self-control: an extension and test of a general theory of crime. Justice Q 16(4):633–654
Schreck CJ, Wright RA, Miller JM (2002) A study of individual and situational antecedents of violent victimization. Justice Q 19(1):159–180
Titus R (2001) Personal fraud and its victims. In: Shover N, Wright JP (eds) Crimes of privilege: readings in white-collar crime. Oxford University Press, Oxford
Tyler T (2001) Public trust and confidence in legal authorities: what do majority and minority group members think? Behav Sci Law 19(2):215–235
van Wilsem J (2011) Worlds tied together? Online and non-domestic routine activities and their impact on digital and traditional threat victimization. Eur J Criminol 8(2):115–127
van Wilsem J (In press) 'Bought it, but never got it': assessing risk factors for online consumer fraud victimization. European Sociological Review
Weisburd D, Wheeler S, Waring E, Bode N (1991) Crimes of the middle classes. Yale University Press, New Haven
Weisburd D, Waring E, Chayet EF (2001) White-collar crime and criminal careers. Cambridge University Press, New York
Wolfe SE, Piquero AR (2011) Organizational justice and police misconduct. Crim Justice Behav 38(4):332–353
Zietz D (1981) Women who embezzle or defraud: a study of convicted felons. Praeger, New York

French Colonial Police

Emmanuel Blanchard
Centre de Recherches Sociologiques sur le Droit et les Institutions Pénales (CESDIP), Université de Versailles, Guyancourt, France

Overview

The question of policing is consubstantial to that of colonization and was raised at a very early stage, not just in diplomatic or military terms, but also with respect to the various manners of civility, regulation, and discipline that were to be expected from a good policing service. The notion of colonial policing, however, can hardly account for the many policing practices encountered in the various territories, with their diverse judicial and political statuses. It tends to foster a uniform vision of policing organization and practices, even though none of this was ever centralized, or even standardized, in this vast empire. Most importantly, it tends to conceal the fact that policing tasks were handled by multiple organizations and agents, many of whom were not, technically, actual police. Still, the very existence and usage of this phrase – colonial policing – suggests that, regardless of place and time, colonial authorities had their own set of policing practices and violent behaviors, which were partly distinct from what was prevalent in metropolitan France.

From a chronological point of view, the legacy of slavery and plantation power as well as the continued domination over various territories (e.g., some West Indies or Indian Ocean islands) must be considered: the hiatus between the first French colonial empire (built under the *Ancien ré gime*, during the seventeenth century in particular) and the second (refounded on the rubble of Napoleon's defeats, starting with the conquest of Algeria in 1830) should not be overstressed. Only in the late nineteenth century, however, considering the developments and reforms happening in the home country did colonial policing issues gradually and partially gain autonomous status, distinct from wider concerns about the frailty of colonial domination. Indeed, colonial domination only exceptionally reached the "pacification" stage, even though this goal was continually reasserted by civilian authorities, closely associated to their military counterparts.

State of Knowledge

Colbert complained as early as 1667 that "very little concern has been shown so far in New France for the policing and civil life of Algonquians and Huron." He considered this "long-standing neglect" a mistake that ought to be remedied promptly (Havard 2009). The word "*police*" was at the time highly polysemic in French. Indeed, the indeterminate nature of the term speaks volumes, insofar as police forces, under the *Ancien Régime* already, were associated to a civilizing mission with a particular focus on extending the reach of law. On its own, however, the "law and order" diptych cannot fully account for the nineteenth to twentieth-century colonial policing in the French empire – the period studied here.

The issue of policing organization and practices in the colonial context had to wait until the nineteenth century to be raised in concrete fashion. In the early days, the task of pacifying newly conquered territories was largely entrusted to the military. Only in areas specifically designed to receive European settlements was the creation of policing bodies seriously considered. A police commissioner (*commissaire de police*), for instance, was appointed in the immediate aftermath of the capitulation of the Dey of Algiers, on 5 July 1830. As evidenced by available documents, his meager staff spent many years focusing exclusively on managing and disciplining the "populace" and the "unwanted" freshly disembarked from Europe. Until the late nineteenth century, archives make scant mention of the "natives" ("*indigènes*"), who seldom appeared as a problem to be tackled by policing agents. The "gendarmerie of Africa," which was also created in the wake of the landing in Alger regency (June 1830) and was primarily conceived as a provost body for the supervision of locally stationed troops, occasionally took part in combats for the extension of colonized territories. Only under the Third Republic was it endowed with law and order enforcement responsibilities with regard to indigenous populations not governed by military administration.

Whenever it was felt that the action of armed forces should be supplemented by the creation of police units, the dominant method consisted in importing or adapting models or institutions that were prevalent in metropolitan France, two of which – the gendarmerie and the Paris prefecture of police – were frequent sources of inspiration. These flows of policing practices and frameworks from the home country to the colonies should not, however, lead us to underestimate other legacies that also contributed to the shaping of colonial policing: the settlers themselves but also privately owned companies, the Army, indigenous troops, country policemen and rangers, and colonial officials whose prerogatives could be extended with respect to the remoteness of their post (such as the *gardes-cercle*) have all been endowed with policing and penal capacities right until the decolonization era. Most importantly, in remote countryside areas, colonial authorities have endeavored to build upon local hierarchies and authorities, thus sparing themselves the task of having to establish specific policing bodies. While state order reigned supreme in "islands of domination" (in cities and along railways mainly), it was replaced in many places (plantations, concessions) by private authorities or by traditional social control mechanisms, so that the policing and judicial systems of settlers were only called upon as a last resort option or instrumentalized to pressure the promoters of local order.

Given this proliferation of policing agents and organizations, trying to distinguish between the "civilian" and the "military" makes little sense (Blanchard et al. 2011). The ongoing concern was the cost-effective, non-staff-intensive law and order maintenance. This perspective, however, left little room for insights going beyond the hackneyed remarks about natives being receptive only to the language of force. Hence, troubles were systematically interpreted as potential rebellions that might upset the entire edifice of colonial domination: despite many natives having been enrolled and various segments of the population having voluntarily rallied, while others had pursued arrangements and interests that they would be keen to stand for, colonial power remained exogenous. It was never able to take root in true cultural hegemony or in any political legitimacy reaching beyond the small circle of direct beneficiaries of this mode of government. A major indicator of the frailty of colonial power at the time is the way policing agents, administrative elites, and even the community of ordinary settlers tended to construe dreaded but simple attacks on private property as suggesting a general dissent against the balance of power – if not threats on their own existence.

Characteristics such as a lack of specialization of policing functions (which were closely linked to tax levying and statute labor) and a limited interest for crime repression, although not peculiar to the colonial setting, nevertheless do point to the principles of cheap government and under-management. The pattern has been well described by George Orwelle with respect to Burma (Orwelle 1998), but it also applied to most territories comprising the French empire. In an even more fundamental way, this pertained to some of the inner contradictions of the political project regarding settlements: enticed by separatism and rejecting any increase of tax pressure – which they felt should exclusively burden indigenous populations – settlers tended to impede the development of a specialized, professional policing apparatus. Thus, it is that in the years 1898–1900, their representatives in the *départements* of Algeria kept complaining of growing insecurity while simultaneously pressuring the authorities to obtain the withdrawal of the gendarmerie, which they considered too costly and too formal in its enforcement style.

The only positive aspects in their eyes were the reinforcement of the action of armed forces and of emergency powers (in particular the 1881 "*Code de l'indigénat*," a set of specific criminal rules for natives, which in fact legalized longstanding practices). Such tensions with civil authorities and police organizations could be found in many other places until the decolonization era. A case in point is the Kingdom of Cambodia under French protectorate, where the colonial administration, far from deferring to financiers and settlers, actually tended to moderate their excesses, providing a legal and practical framework that prevented extreme prejudice toward the natives, which might otherwise have triggered an uprising (Guérin 2008). The buffering role played by police forces in such circumstances should, however, not be overstated. Colonial policing did embrace the whole array of founding violence (land expropriation, physical constraint, attacks against local cultural and religious practice) traditionally associated to "despotic" colonial administrations that, as was the case, placed themselves under the auspices of the "sword" (Orwelle 1998).

The judicial inflation that can be observed in a colonial situation – "the empire of law" described by Emmanuelle Saada (2002) – was accompanied by the deployment of "coercive networks" (Sherman 2009), of which police forces were a mere segment. This removes a lot of interest from any attempts at reporting on policing in a colonial setting from an institutionnal perspective, since the coercive practices experienced by the colonized could hardly be accounted for. In the case of the French empire, the task is even made almost impossible by the fact that no centralized agency was ever put in charge of designing and evaluating such policies – i.e., there was no equivalent of the British *Colonial Police Service* nor was any formal, shared employment scheme ever deployed throughout the colonies. While some officers did circulate across colonies, they made up a tiny minority of the numerous police deployed in the empire. Furthermore, ambitious and thorough historical research focusing on the local variations and characteristics of an empire encompassing a great many different institutional frameworks is made impossible by the sheer scarcity of available studies: no overview, no synthesis, and no comparative research have been published so far, which is why the French case cannot be considered on par with the UK (Sinclair 2006) or the Netherland models (Bloembergen 2011). Indeed, the most recent synthetic works published on French police forces have almost nothing to say on the imperial dimension of policing, focusing instead on metropolitan France (Anderson 2011), even though the latter can hardly say to have been estranged from the overseas at the time.

Current Research

Recent or current research by young scholars and the abundance of studies on the gendarmerie (Luc 2010) do however allow us to outline what the mechanisms of policing have been in certain territories. On a general level, Florence Bernault's insights on colonial prisons, which were often administered by policing bodies, are a good starting point: "the colonial prison did not apprehend delinquents according to a single judicial register (. . .) it sought to maintain the racial and juridical hierarchies upon which colonial rule was founded, thus reproducing a violent, personalized, and circumscribed power (. . .) As the vanguard of these confining tactics, colonial prisons thus illuminate how European regimes in Africa survived as enterprises of unending conquest" (Bernault 2003).

Florence Bernault rightly insists that the conquest endeavor was an ongoing one. "Pacification" was never taken for granted, but remained a violent, uninterrupted process, which however varied in intensity according to place and time. But it is no less true that initial periods of military domination were generally followed by deliberate attempts at establishing civilian authorities, epitomized, among other symbols, by their police forces. In many territories gendarmes remained for a long time (sometimes as late as Word War II) the only actual, however furtive, colonial presence identified by the locals.

Using a – necessarily reductive – process approach, the timeline highlighted by John McCracken regarding the British empire, and more precisely Malawi (McCracken 1992), can be adapted to the French empire: an "era of coercion" encompassing the first decades of domination, followed by an "era of control," then one of "collaboration" (the interwar period), and finally a "crisis of authority" after World War II. Clearly, this pattern is not set in stone, and periods may overlap. Most importantly, it must be inscribed in singular colonial trajectories, which depend heavily on when the conquest was made as well as on the frequency and punch of violent actions of resistance to colonial order.

The gendarmerie has been the main colonial police force: its military police dimension (both as a provost body and because of the military status of gendarmes) made it a privileged tool in the conquest phase and during the "era of coercion." For instance, on 16 May 1843, about 30 gendarmes were involved in the battle of Taguin or more precisely in a raid against the *smala* of Abd El-Kader (Jaulin 2009), which resulted in the taking of more than 3,000 prisoners and major spoils (a "feat" that still adorns the flag of the French national gendarmerie). In Madagascar the gendarmerie's contribution to *pax gallica*, as imposed by Gen. Gallieni, was rather short lived (1901–1904, although a gendarmerie detachment was reestablished in 1931). But in the whole empire, the colonial gendarmerie did take part in numerous battles, regardless of the precise dates of territorial annexations. In 1918, it supported the French Army after France was granted a mandate on the Levant from the League of Nations. Incidentally, the Lebanon gendarmerie was on the front line – and targeted – during the 1925 Druze insurgency (Faisant de Champchesnel 2010). From this moment on and until the accession to independence, however, locals gendarmes tended to side with insurgents, and the resulting struggles, pitting gendarmes vs. gendarmes, highlighted the limits of the French policy of fueling ethnic and religious divides. Even where the colonization process was initiated in the diplomatic framework of trans-imperial competition instead of bringing about a war of conquest, the phrase "era of coercion" refers to those moments when French authorities were trying to tighten their grip. The kingdom of Cambodia, where a protectorate treaty had been signed between King Norodom and France in 1863, is a case in point: certainly, the 1885 uprising was stifled by military interventions, but it was used as an opportunity to reinforce the militiamen of the indigenous guard, who also acted as a police force. In particular, they were in charge of supervising tax collection and organizing statute labor days, up to and including the Highlands, an area where dissent toward central authority, whether Khmer of French, was at its highest (Guérin 2008).

The "era of control" should be understood as both (a) the moment when military head count decreases and powers are transferred to civilians (for instance, in 1871 in the main districts of Algeria) and (b) the period when new policing methods are being developed to control people, following technological innovations from metropolitan France. In Madagascar at the turn of the century, General Gallieni extensively resorted to photography for the census and racial classification of the population. Generally speaking, although colonial police forces tended to be understaffed and underfunded, they were at the cutting edge of anthropometric and identification techniques. These intellectual, technological, and human investments were markedly associated to specific representations of crime as pertaining to entire populations, which in many areas of the empire translated into collective fines and punishments.

From a technical and organizational point of view, the protectorate of Tunisia was a precursor, equipped, as early as 1890, with a judicial identification service similar to that created only a few years before at the Prefecture of police. Indeed, in the early 1890s, Alphonse Bertillon himself played a key part in the establishment of the Algiers judicial identification service. Similar bureaus were subsequently opened in the AOF (French West Africa), the AEF (French Equatorial Africa), Madagascar, and Indochina. But the most accomplished version of this anthropometric form of control – the best researched at least, thanks to the works of Ilsen About (2011) – occurred in Saigon. The anthropometric identification bureau was institutionalized in 1897, as part of the immigration services. Carding (forcing people to have an ID), fingerprints, and anthropometry were made instrumental in a vast census and registration project encompassing all natives and foreign Asians. Even though opposition from multiple sources (congregations, consular representatives, part of indigenous elite) eventually caused the shelving of this ambitious project, by 1917 the bureau had managed to gather some 700,000 records, complete with fingerprints. Indochina in this domain acted as a laboratory.

The emphasis on identification as a trait of colonial policing is further enhanced if one integrates metropolitan France into the global picture of the colonial empire (Rosenberg 2006). In Paris, after an identity card for foreigners had been established in 1917, French nationals of colonial origin were the next target (Blanchard 2011). A North African brigade was thus created at the Paris Prefecture of police in 1925, with multiple prerogatives (administrative and social in particular) aiming to feed a database that would list all colonized people originating from North Africa (90 % of whom were Algerian Muslims enjoying French citizenship, meaning that they needed no specific documents whatsoever to settle in metropolitan France). Yet these panoptic aspirations were never entirely fulfilled: given the rules governing the political representation of certain segments of the colonial world, from the 1930s onwards (especially after World War II), the colonized were sometimes able to oppose such designs, which in any case appeared to be limited by the sheer scope of their technical and financial requirements. Still, from the interwar period onwards, in metropolitan France as in the empire, political surveillance of the colonized became a policing priority. In 1922 in Dakar, the creation of a Central Service for the Policing and Safety of French Western Africa (*Service central de police et de sûreté de l'AOF*), whose structure was replicated in each of the component colonies (*Service central de police et de sûreté du Sénégal*, *Service central de police et de sûreté du Dahomey*), epitomized this concern. The general objective of controlling mobility and preserving disciplined, segregated urban spaces for the benefit of European minorities came on top of a hidden agenda: thwarting the ability of the better-educated colonized (the "evolved") to travel and act (Brunet-La Ruche 2012). These people were suspected – sometimes quite rightly so – of anti-colonialist or nationalist activism. As a matter of fact, during the interwar period, and especially in the AOF, this channeling of policing means and services into political surveillance contributed to relocating the fears and representations of insecurity from rural areas to the cities, where criminality was on the rise but did not seem to be eliciting any tailored policing response (Fourchard 2006).

The weight of political defiance toward indigenous elites partly explains why the notion of an "era of collaboration" does not apply to the French empire quite as it does to the British empire. No movement of indigenization proper has ever taken place in the higher ranks of the policing hierarchy. To a certain extent, policing management in the French empire was more reminiscent of the Belgian approach than the British method. In Belgian Congo, for instance, the *Force Publique* created in 1885 never entrusted any natives with leadership positions (Lauro 2011). The same applied to most of the French empire: with the exception of a handful of commissioners, "Africanization," and more generally indigenous access to managing positions, was only engaged in the very last years of colonization. Moreover, such developments were seldom considered as preparing a transition toward independence. In some cases, the requirements (educational, legal, etc.) were raised to such a level and the will to confront political and violent protests was so strong that any increase of staffing levels was matched by a corresponding growth in the ratio of European-born recruits (such was the case, for instance, for the gendarmerie in Algeria). None of the French police bodies ever surmounted the quandary of importing mainland operational methods and gradually implementing a shared legal framework without making it increasingly difficult to gather intelligence from the locals (Thomas 2008). Indeed, the chasm from local circles was such that intelligence services often had to resort to broad, hackneyed ideological patterns (the Communist threat, pan Arabism, the influence of foreign nations) which actually stood in the way of taking the full measure of protests or local separatist movements. In defense of these services, it should be said that nationalist, religious, and communist movements and motivations were so enmeshed – during rural uprisings in Vietnam (1930–1931), for instance – that deciphering the redistribution of anti-colonialism was an extremely difficult task.

Anti-colonialism was in fact eventually recognized from an institutional perspective – the African Democratic Rally (RDA), founded in 1946, is a prime example – and was sometimes able to weigh in on policy and policing developments. Political surveillance, which had been considerably reinforced all along the 1940s, was thus gradually loosened as the recognition of African states made progress: in 1958, both the Intelligence and the Judicial Identity services of the AOF were even officially disbanded, but the gendarmerie managed to grab some of their files and prerogatives (Papa Dramé 2010). Most importantly, in the wake of World War II, recruitment processes became more bureaucratic. The written word as well as legal proceedings took greater importance, both in daily tasks and in terms of career development: French-speaking, higher-educated colonized people massively took over the intermediate echelons of the hierarchy. Joël Glasman has convincingly shown how, in the case of Togo, such developments were both (a) inscribed in how part of the population adjusted to the colonizers standards and (b) loaded with contradictions and oppositions, not only within policing bodies but also among the local elites (Glasman 2010).

This standardization process, however, should not be overestimated: even in the *départements* of Algeria, where metropolitan rules and frameworks had supposedly been best integrated, the creation of the first constabulary academy was only decided in 1949, 1 year prior to that of the AOF in Dakar. Indeed, colonial police forces were never known to put a strong emphasis on disciplining and educating their personnel. Thus, restraining the use of force was never considered a priority, or even a reality. In the mid-1950s, Albert Memmi could write of the colonizer, with good reason, that “a machine-gun burst into a crowd of colonized causes him merely to shrug his shoulders” (Memmi 1965). Even during allegedly quiet periods – in terms of violent unrest against colonial domination – marked by a relative professionalization of policing by civil bodies instead of military ones, this reliance on shootings and killings was noticeable: in Tunisia in 1937–1938, both in the mining area of Gafsa (March 1937) and in Tunis (9 April 1938), union (CGT) and political (Néo-Destour) rallies and demonstrations triggered bloody repression. Massive use of gunfire by policemen, gendarmes, and colonial troops killed dozens.

During the so-called “crisis of authority” period, these shootings grew in both magnitude and frequency. In May 1945, for instance, in Sétif and Guelma (Algeria), vigilante European settlers sided with the police forces, who in turn had received support from the Army (Peyroulou 2009). They coordinated not just to suppress Algerian unrests but also to arrest, intern, and quite often kill or dispose of anyone suspected of belonging to the separatist movement. Parts of the casualties, which numbered in the thousands, were actually caused by shelling, from both sea and air. Ever since the interwar period, in the Maghreb and Middle East in particular, air policing had been used not only for land surveying but also to subdue outbreaks by the colonized: firepower was supposed to make up for scant police and military staffing on the field (Omissi 1990). Inspired by the Italian conquest of Libya (the 1912 bombings) as well as the action of the Royal Air Force in Afghanistan and Iraq in particular (1919–1920), the French also resorted to this ploy. After a few experiments in Morocco (the Rif War), they made the most of air domination to suppress the Druze rebellion of 1925–1926. The city of Damascus was massively shelled, and the residents of Druze and Muslim neighborhoods subjected to what can only be dubbed a terrorization policy. This particularly lethal use of air policing was designed to match the movements of Army troops and police forces, the gendarmerie in particular. The idea was to make up for their notorious and accepted understaffing.

Still, the “crisis of authority” period saw a rapid increase of police staffing: notwithstanding the decolonization wars (Indochina, Cameroon, Algeria), the growing intensity and attendance of protest demonstrations as well as the ever-increasing burden of the standards and regulations in metropolitan France left no other option. To develop but one example, the head count of the AOF gendarmerie (including Togo) rose from less than 300 in 1946 to almost 1,200 in 1957 (all of them “white,” bar two). Auxiliary gendarmes (black Africans) numbered 700 in 1946 and more than 2,625 in 1957. In Togo alone, from the 1930s to the 1950s, police staffing levels doubled but – more importantly maybe – remained quite low at less than 1,000 men at the end of the colonial era (Glasman 2010). Hence, even though the relative weight of uniformed service personnel (vigilantes, guards, gendarmes, constables) was particularly significant in the colonial state (from one third to one half of the entire personnel in Togo, a lower proportion whenever more Europeans were present), the overseas police-to-population ratio remained lower than in the home country. While in the 1950s, France had on average one police per 250 citizens (a ratio that was only matched in the most watched cities in Africa, such as Abidjan), the figure could drop down to less than one per 1,500 in certain cities in AOF (such as Kayes in French Sudan).

This lack of personnel is sometimes put forward as an explanation to the disproportionate use of force by policing authorities: gunfire in this case is interpreted as a means of compensating the powerlessness of outnumbered police doing their best to keep the crowd at a distance. Much like its British counterpart (e.g., with the 1919 Amritsar massacre), the French empire was marked by the permanence of this logic consisting of slaughtering protesting crowds of colonized people, as evidenced by the now well-documented case of the December 1952 repression of the Casablanca riot (House 2012). These casualties, however, which numbered in the hundreds each time, can hardly be attributed to any panic on the part of supposedly outnumbered, and therefore overwhelmed, personnel: not only was the police receiving support from colonial troops and vigilante settlers, but the same logic of massive, lethal gunfire against demonstrating colonized was also at work in metropolitan France at the same period (Blanchard 2011). While the Paris police had stopped using firearms to disperse demonstrations (even unauthorized ones) as early as the 1930s, gatherings of colonized people were treated differently. On 14 July 1953, for instance, six Algerian demonstrators were killed by police fire on the *place de la Nation*. In 1961, dozens of Algerians succumbed to the blows and bullets of the Paris police, who on 17 October conducted what has got to be considered a pogrom (House and MacMaster 2006). More than the format of policing or the actual behavior of demonstrators (police response hardly varied, irrespective of the peacefulness of demonstrations), it is the racial-political condition of the colonized that determined both the nature and degree of the policing techniques employed.

The last years of the colonization era had very little to do with a transition period. What entire segments of the police forces offered instead was a violent toughening of their stance. In Algeria, the police units were largely involved in the massive use of torture during the war of independence. The collusion between a majority of policemen and the most extreme supporters of French Algeria (the *OAS*, Secret Army Organization, in particular) contributed both to destabilizing legal authorities (in May 1958 and April 1961) and to protracting the fights beyond the cease-fire signed in March 1962. Aside from this well-known example, parts of Africa were also plagued by counterrevolutionary warfare, especially Cameroon, where the struggle against the partisans and maquis of the Union of the Peoples of Cameroon (UPC), sparked amidst extreme violence in the late 1940s, lasted a further decade after the independence had been proclaimed in 1960 (Mbembe 1996).

Postcolonial Legacies

The “racial coloration” of the colonial situation pointed out by Georges Balandier is undoubtedly one of the major characteristics of colonial policing (Balandier 1970). “Racial segregation” might actually be more to the point, insofar as local police were never able to impose the legitimacy of their prerogatives on European settlers: Amandine Lauro’s remark about Belgian Congo – black police were not even allowed to pick up drunk white people in the gutter – also applied in the French part of that colony (Lauro 2011). Conversely, and more tragically, Europeans never gave up their individual right to punish the colonized: whenever police action was considered not to be harsh enough, it would be complemented by punitive raids and the creation of vigilante militias. Beyond such paroxystic circumstances, urban police forces were the main tool for managing populations, and their primary task was to safeguard the feeling of close-knit community, the hygiene, and the tranquility of the “European city.” They were, however, never granted the wish to restrict indigenous access to European residential spaces to the sole servants. It is true that – as opposed to what was the case in Ghana or British Nigeria – no official segregationist policy (creating separate neighborhoods for Europeans, migrants, or minorities) was ever enforced in the cities of the French empire. Still, policing authorities – through vice squads in particular – did restrict and keep under surveillance the least movement across racial and colonial borders.

The question of prestige was so pressing for Europeans that many of them were actually not welcome in the colonies. Those who were allowed to settle there had to conform to a set of rules that worked as just as many markers of the superiority of French civilization. Close attention was paid to gender relations and roles: civil servants in particular were briefed on the subject and deprecatory discourses on the sexuality or virility/feminity of the colonized abounded. Police forces, in their surveillance reports for instance, contributed to spreading these stereotypes. More importantly, they played a direct part in the regulation of the sex trade. Studies by Christelle Taraud on North Africa (Taraud 2003) and Isabelle Tracol-Huynh on Vietnam have shown that prostitution regulationism in the colonies cannot be measured by the yardstick of hygienism, the primacy of the patriarchal family model, and the extension of police prerogatives, as was the case in metropolitan France. There was of course no lack of such dimensions, but the primary issue was the drawing of clear gender and racial lines in the government of the people.

In Vietnam, where marriages between colonizers and the colonized were more frequent than in North Africa and mixed-blood children more numerous, indigenous concubines of "Europeans" were even considered prostitutes (Tracol-Huynh 2010). Prostitutes, in turn, were *de facto* organized into a regulatory and penal hierarchy based on racial criteria. Other trade categories, loosely defined but highly visible, such as dancers or singers, were also under tight surveillance, which varied, however, depending on the means available to the police, the proportion of settlers in the overall population, and how the majority population responded to such regulatory and policing intrusions. In the Shanghai French concession for instance, eradicating or even regulating the massive prostitution – which predated colonial domination – was never really considered (Henriot 2001). Still, the fact that healthcare protective measures, however scarce, only catered for European clients clearly indicates that, even in this highly specific colony, the seeds of a racialized population policy had been sown.

This policy also reached metropolitan France, especially through the police or military institution consisting of organizing prostitution spaces dedicated to immigrants or North African troops (the military brothels aka "*bordels militaires de campagne*"). This was one among many aspects of the distinctive brand of policing applied to the colonized, whose condition was deeply affected – especially during the Algerian war of independence – by recurrent identity checks, frequent roundups, and administrative detention. They were also targeted by positive policies such as welfare (access to healthcare, job search support) and collective housing (accommodation facilities for migrant workers that were built starting in the 1930s, with a second wave launched in the late 1950s). While sometimes improving their living conditions and facilitating their understanding of mainland rules and bureaucracy, which differed significantly from what they had known overseas, these programs nevertheless followed a logic of control that did not vanish after the independences. To assess the weight of colonial legacies, one only has to look at the careers of former colonial administrators or officers, who started out managing agencies in charge of supervising the colonized (especially during the Algerian revolution) before heading immigrant support services. Such careers also epitomize the hybridization of policing repertoires, which became a blend of population control, political surveillance, and welfare.

Controversies on the colonial matrix governing the police control of immigrant or racialized populations did not, however, cease with the retirement of the last personnel to have started out at the time of the colonies or in the ranks of these particular agencies. Today, the colonial "difference principle" no longer is a founding principle of specific measures and programs. Sociological studies, however, unanimously demonstrate the existence of racialized gradations in terms of the ability to stand up for one's rights against policing authorities, police influence, and the use of force – especially lethal force. In France today, the best-known aspects of the racialization of policing practices are police "blunders" (often a euphemism for "killings") and the repeated, often illegal identity checks performed on young, nonwhite males: depending on their clothing habits, black or Arab-looking young men are 2-15 times more likely to have their identity checked than other categories of the population (Jobard and Lévy 2009). While certainly not all targeted persons can be linked to the former French colonial empire (or even its current remains, notably in the West Indies and Indian Ocean – the Mayotte and Réunion islands), it cannot be denied that the control of the colonized was one of the historical operators in the process of policing racialization. The weight of colonial policing's legacy has thus become a controversial issue, fiercely debated both in academic circles and in the media of contemporary France.

Colonial continuities are no monopoly of the former home country: many a policing institution has been passed on, almost "as is," to newly independent states. The few local police have often contributed to the training of younger generations, with the support of the French cooperation services (after the independences, as before, many police commissioners from West African states attended Saint-Cyr au Mont d'Or police academy, in the Lyon area). The most striking and symbolic legacy has got to be the flock of national gendarmeries in former French colonies: 150 years after Napoleon's conquests, the decolonization era acted as a second age for the spreading of the gendarmic model. From Ivory Coast or Mali to Syria and Lebanon, through Tunisia or Algeria in particular, the forces that stemmed from the French legions of colonial gendarmerie have played their part in several of the episodes of repression and political transition that are currently at the heart of geopolitical news. Beyond the racialized management of populations and the religious minorities policy that so deeply affected the Levant, for instance, only time and an intensification of research will tell whether these policing and gendarmic legacies (in their linguistic, symbolic, material, and organizational dimensions) have produced other continuities.

Related Entries

- ▶ Dutch Colonial Police
- ▶ Early Modern Police and Policing
- ▶ Police and the Excessive Use of Force
- ▶ Role and Function of the Police

Recommended Reading and References

About I (2011) Surveillance des identités et régime colonial en Indochine (1890–1912). Criminocorpus, revue hypermédia. http://criminocorpus.revues.org/417

Anderson M (2011) In thrall to political change. Police and gendarmerie in France. Oxford University Press, Oxford

Balandier G (1970) The sociology of Black Africa: social dynamics in Central Africa. Praeger Publishers, New York

Bernault F (ed) (2003) A history of prison and confinement in Africa. Heinemann, Portsmouth

Blanchard E (2011) La police parisienne et les Algériens, 1944–1962. Nouveau Monde, Paris

Blanchard E, Deluermoz Q, Glasman J (2011) La professionnalisation policière en situation coloniale : détour conceptuel et explorations historiographiques. Crime, Histoire & Sociétés 15(2):33–53

Bloembergen M (2011) The perfect policeman: colonial policing, modernity, and conscience on Sumatra's West Coast in the early 1930s. Indonesia 91:165–191

Brunet-La Ruche B (2012) "Discipliner les villes coloniales": la police et l'ordre urbain au Dahomey pendant l'entre-deux-guerres. Criminocorpus, revue hypermédia. http://criminocorpus.revues.org/1678

Faisant de Champchesnel H (2010) Les gendarmeries syriennes et libanaises: relais de la puissance mandataire ou instruments de l'indépendance des nouveaux Etats. In: Luc J-N (ed) Soldats de la loi. La gendarmerie au XXe siècle. PUPS, Paris, pp 203–210

Fourchard L (2006) Lagos and the invention of juvenile delinquency in Nigeria, 1920–1960. Journal of African History 47(1):115–137

Glasman J (2010) Penser les intermédiaires coloniaux : notes sur les dossiers de carrière de la police du Togo. History in Africa 37:51–81

Guérin M (2008) Paysans de la forêt à l'époque coloniale: la pacification des habitants des Hautes Terres du Cambodge. PUR, Rennes

Havard G (2009) "Les forcer à devenir Cytoyens". État, Sauvages et citoyenneté en Nouvelle-France (XVIIe–XVIIIe siècle). Ann Hist Sci SocAnnales. Histoire, Sciences sociales 64(5):985–1018

Henriot C (2001) Prostitution and sexuality in Shanghai. A social history 1849–1949. Cambridge University Press, Cambridge

House J (2012) L'Impossible contrôle d'une ville coloniale? Casablanca, décembre 1952, Genèses. Sciences sociales et histoire 91:78–203

House J, MacMaster N (2006) Paris 1961: Algerians, state terror and memory. Oxford University Press, Oxford

Jaulin E (2009) La gendarmerie dans la guerre d'Algérie. Lavauzelle, Panazol

Jobard F, Lévy R (eds) (2009) Profiling minorities. A study of stop-and-search practices in Paris. Open Society Justice Initiative, Paris

Lauro A (2011) Maintenir l'ordre dans la *colonie-modèle*. Notes sur les désordres urbains et la police des frontières raciales au Congo belge (1918–1945). Crime, Histoire & Sociétés 15(2):97–121

Luc J-N (ed) (2010) Soldats de la loi. la gendarmerie au XXe siècle. PUPS, Paris

Mbembe A (1996) La naissance du maquis dans le Sud-Cameroun, 1920–960: histoire des usages de la raison en colonie. Karthala, Paris

McCracken J (1992) Authority and legitimacy in Malawi: policing and politics in a colonial state. In: Anderson D, Killingray D (eds) Policing and decolonisation. Politics, nationalism and the police, 1917–1965. MUP, Manchester, pp 158–186

Memmi A (1965) The colonizer and the colonized. Beacon, Boston

Omissi D (1990) Air power and colonial control: the royal air force 1919–1939. Manchester University Press, Manchester

Orwelle G (1998) The complete works of George Orwelle, vol 10. Martin Secker & Warburg, London, pp. 172–180

Papa Dramé P (2010) Le service de gendarmerie au Sénégal à l'époque de l'Union française, 1946–1958 In: Soldats de la loi. La gendarmerie au XXe siècle. PUPS, Paris, pp. 223–232

Peyroulou J-P (2009) Guelma 1945: une subversion européenne dans l'Algérie coloniale. La Découverte, Paris

Rosenberg C (2006) Policing Paris. The origins of modern immigration control between the wars. Cornell University Press, Ithaca

Saada E (2002) The empire of law: dignity, prestige, and domination in the colonial situation. French Politics, Culture and Society 20(2):98–120

Sherman T (2009) Tensions of colonial punishment: perspectives on recent developments in the study of coercitive networks in Asia, Africa and the Carribean. History Compass 7(3):659–677

Sinclair G (2006) At the end of the line. Colonial policing and the imperial endgame, 1945–1980. Manchester University Press, Manchester

Taraud C (2003) La prostitution coloniale. Algérie, Tunisie, Maroc (1830–1962). Payot, Paris

Thomas M (2008) Empire of intelligence. Security services and colonial disorder after 1914. University of California Press, Berkeley

Tracol-Huynh I (2010) Between stigmatisation and regulation: prostitution in colonial Northern Vietnam. Culture, Health & Sexuality 12:73–87

Frequent Victimization

► Repeat Victimization

Front Door and Backdoor Sentencing

Nicola Padfield
Faculty of Law, University of Cambridge,
Cambridge, UK

Synonyms

Parole; Recall; Release; Sentencing

Overview

In the literature on criminology and criminal justice, much more appears to have been written on sentencing "law" than on sentencing "practice." And those who write on sentencing, law or practice, have tended to write about "front door" sentencing: the rules as well as the judicial decisions which impose the initial sentence. But most systems allow some form of "backdoor" resentencing: for example, judges may reduce the sentence on proof of cooperation and progress or administrators may recall to prison those who had been released "early." Most criminal justice systems allow some sort of "early" release from prison and also recall to prison offenders who somehow "fail" in the community. This entry argues that understanding "sentencing" involves understanding that it is not just a "one-off" event, but more of a continuous or ongoing process, which should be subject to regular judicial review or oversight.

Introduction

"Sentencing" is normally understood to be the process which happens in court, when an appropriate punishment is imposed on an offender, by a judicial authority. The way in which this happens varies across jurisdictions and within jurisdictions and across time. For example, common law jurisdictions tend to separate a criminal trial into two separate and distinct parts: one which focuses on the decision whether someone is to be found guilty or not of an offence and a separate sentencing stage or hearing which decides on the appropriate sentence. However, in many countries, particularly in continental Europe, the court may decide on conviction and sentence in one process (e.g., the sentence is part and parcel of the French concept of *condamnation*).

We are immediately faced with some definitional issues which make comparisons so difficult. When is a penalty a "sentence"? Are those who are diverted away from the criminal justice system by formal or informal cautions or warnings "sentenced" (see Padfield et al. 2012)? What about those who are compulsorily detained in mental hospitals, who may or may not have been convicted of their "crimes"? Is the outcome of mediation (increasingly common in a variety of jurisdictions) a "sentence"? What about fixed penalties and the increasingly popular but often misunderstood "restorative justice"? These are all arguably forms of "sentence," even when administered administratively and not by a court. But when is a court truly a court? Who is a "judge" or what is a "judicial authority"? None of these are easy questions to answer. It is not even clear when a sentence ends: public protection concerns have led many countries to continue to monitor offenders, either formally or informally, even after the completion of their sentence (see, e.g., the English MAPPA or multiagency public protection arrangements). Others allow civil "preventive detention" (e.g., Germany).

Then we reach the main difficulty, for current purposes. There is often a gulf between the sentencing law prescribed by law and what happens "on the ground." I have found it useful to borrow from the French and to identify three different "levels" of sentencing frameworks (see Delmas-Marty et al. 2003). I think the French language works more efficiently here than the English in creating neat labels:

- *La peine encourue* (to be found in the formal law, whether English statutes or continental *Codes Pénales*), which is the possible sentence, that which is permitted or prescribed by the law. Many of those (particularly many academic criminal lawyers from non-common law jurisdictions!) who look to study sentencing do not look beyond these formal legal rules.
- *La peine prononcée* is the sentence given by the judge(s) in court. There has been remarkably little work done around the world analyzing actual sentencing decisions. We may well know how many burglars, for example, are given a custodial sentence in a particular jurisdiction in a particular year, but often we know very little about how the judge(s) reached their decision. The recorded sentence alone gives no insight into the factors which affected the judge, into the way that aggravation and mitigating factors influence individual decisions. There is remarkably little good quality and systematic information about judicial practice (though there have been attempts to gather it: see Tata and Hutton (2002, especially Chap. 19) or the work of the Sentencing Council in England at http://sentencingcouncil.judiciary.gov.uk/docs/CCSS_Experimental_Release_web.pdf)
- *La peine appliquée or executée*. These are the sentences actually served by offenders. Particularly when we look at custodial sentences, it seems rare that offenders have to serve the precise sentence imposed in court. And even those serving sentences in the community may have their sentence modified during the course of its implementation, of course. It is even more difficult to gather reliable data on this area of sentencing. As we shall see, sentences may be extended or curtailed. They may be varied according to the offender's rehabilitative needs and/or his or her cooperation with the sentencing and administrative bodies. Short sentences of imprisonment are very often suspended if (but only if) the offender engages (adequately) with probation services. This entry seeks to argue that sentencing should be seen more as an ongoing process than a one-off event.

All three stages or levels of sentencing are authorized by law, but these labels are to my mind useful as they help us to distinguish law as applied in practice from the law found in statutes or in books. They allow us to contrast the formal rhetoric of statutory penalties with the law as it is implemented. I also adopt the distinction between “front door” and “backdoor” sentencing for similar reasons: it is important to remember that what the judges say when sentencing someone may not accord very closely to the sentence the offender will actually serve in practice. Unfortunately, too much political and academic attention has focused on what is often known as “front door sentencing,” but what happens at the “backdoor,” while often invisible, is just as important. Indeed, because it is often invisible, we must be even more vigilant to ensure due process (fair procedures and fair outcomes).

Front Door Sentencing

In Law

The formal rules of sentencing are laid down by law. Normally, the framework will be laid down by a Parliament or other elected body. This “framework” will vary enormously from jurisdiction to jurisdiction: many countries have a very detailed sentencing code, others a more confused (or confusing) collection of legal rules. The formal law normally lays down maximum sentences, but sometimes minimum or a range of possible sentences. Sometimes a sentencing regime is designed around a court regime: offences are categorized (felonies as opposed to misdemeanors or indictable as opposed to summary offences, crimes as opposed to delicts, etc.). Thus, offences may be subdivided according a court hierarchy or simply according to “seriousness.” The nature of the “judge” may vary: many jurisdictions involve lay judges, sometimes for less serious crimes (e.g., lay magistrates), sometimes for more serious cases (juries). In many jurisdictions the jury has absolutely no role in sentencing, but in others it may decide or in some sense affect the sentence, sometimes deciding with the judge, or sometimes alone, in isolation from the judge(s).

There is widespread agreement that the purposes of sentencing should be identified, but how this is done in law is less easily agreed. Thus, the English and Welsh judge must have regard to the following purposes (see s. 142 of the Criminal Justice Act 2003):

> the punishment of offenders
>
> the reduction of crime (including its reduction by deterrence)
>
> the reform and rehabilitation of offenders
>
> the protection of the public
>
> the making of reparation by offenders to persons affected by their offences

Would it be more “effective” to have one priority? Some constitutions grant a constitutional right to rehabilitation: for example, Article 27(3) of the Italian Constitution provides that punishments must aim at resocializing the convicted (Gualazzi et al. 2010). How this applies in practice of course needs to be monitored by empirical research.

Most people would agree that a key requirement is consistency. Consistent sentences are desirable, not least so that the system is fair, and perceived to be fair. But legislating consistency and indeed fairness is no easy business! Nor is it easy to monitor or assess how well a system is achieving consistency. Many jurisdictions have experimented with sentencing grids or frameworks or matrices and have created bodies (often “councils”) to advise either legislatures or the judiciary in the development of guidelines (see Tonry and Frase 2001; Tonry 2007). Appeal courts and academic treatises often offer detailed guidance. This topic is raised here simply to make the point that much time and effort has been extended on developing consistency in “front door” sentencing but much less, as we shall see, on consistent or principled “backdoor” sentencing decisions (though some of the first guidelines in the US system were parole guidelines, intended to correct for the wide differences and perceived arbitrariness of judicial sentencing).

Imprisonment remains the most common sentencing option for serious crime. Some countries, of course, continue to authorize the killing of offenders, implementing a lawful death penalty (see Hood and Hoyle 2008). From a European perspective, this contravenes human rights norms. For most legal systems, imprisonment is the main sentence for the more serious offences. It may be indeterminate or determinate. Some European countries (Croatia, Norway, Portugal, and Spain) have no provision for indeterminate or life sentences of any sort. Of those that do have life sentences, some have mandatory life sentences for some offences (UK, Turkey), while for many, life is only ever at the discretion of the sentencing judge or judges. Total life sentences, life without any possibility of parole, are rare. Some countries permit full or whole-life "tariffs" (there are, e.g., about 30 prisoners in England who know they are serving their whole life in prison) but most have a limit, fixing a minimum term which must be served before a prisoner can be considered for release. Many systems do not individualize this tariff, but simply specify a fixed minimum term to be served. What is important for our purposes is to recognize that without a clear understanding of the rules on getting into prison, we can have little understanding of the rules on getting out. Similarly, the rules on getting out, or on early release, make little sentence without a clear understanding of the rules on getting in. It may be easier to calculate release dates when sentences are of determinate or fixed term lengths. But as we shall see, this is by no means always the case.

Then there are sentences which are served in the community: often involving a financial penalty or the supervision of a probation service. The line between custodial and noncustodial sentences is not easy to understand. In many countries, a sentence of imprisonment may be suspended, with or without conditions, partially or fully. This does not mean that the sentence should not be categorized as a custodial sentence, but the imposition of such a sentence poses both theoretical and practical difficulties. In the Israeli context, Sebba (2001) explains that the sentence of community service work was introduced deliberately as a "prison" sentence, served in the community – in large measure to serve as an encouragement to the courts to confine the use of the sanction to cases sufficiently serious to have justified a sentence of imprisonment in its traditional sense, but doubtless it was also to make clear to a wider "audience" that a community sentence was not a soft alternative.

On what grounds should a court, having decided that an offence (or offender) crosses "the custody threshold," step back and suspend the sentence? I have argued that the concept of the "custody threshold," much used in England, is deeply unhelpful, maintaining the unnecessary fiction that even short custodial sentences are somehow "bigger" punishments than alternatives served in the community (Padfield 2011c). What should be the sanctions for minor breaches of conditions attached to noncustodial penalties? Not custody presumably, since the original offence was deemed not to deserve custody. Suspended and community sentences reflect the confusion of purposes: punitive, rehabilitative, and so on. Some systems combine fines and imprisonment. Others specify that they are only alternatives. And many allow "ancillary" penalties which are often the most burdensome components of the sentence: offenders may be deprived of certain rights (the right to drive, even the right to vote in some countries) or ordered to keep in contact with the authorities for many years. Compensation (to victims) and confiscation (of the profits of crime) may be ordered. Given the complexity of options available, it is perhaps not surprising that the practice of sentencing is difficult to understand.

In Practice

In most countries the penalty is "individualized." The judge or judges select an appropriate sentence or package of sentences for this particular offender. This sentence will normally be announced in open court and publically recorded. Judges are usually rightly required to explain their reasons – though the extent to which they do this varies from jurisdiction to jurisdiction and indeed within jurisdictions practice may vary over time ("moral homilies" come in and out of fashion). Understanding judicial reasoning is not always easy. Most systems require the sentencer to fix a sentence according to the culpability of the offender and the seriousness of the offence, and inevitably this requires some subjective assessment. The role played by aggravating and mitigating factors in reaching the "right" sentence is not well understood, in theory or practice (see Roberts 2011), and it is often difficult to obtain data which helps one understand the reasoning of judges in individual cases.

Here, it is important to explore the literature on discretion. Dworkin famously pointed out that "discretion, like the hole in a doughnut, does not exist except as an area left open by a surrounding belt of restriction. It is therefore a relative concept" (1977: 39). Many writers have sought to explore the use and abuse of discretion in sentencing (see Tata and Hutton 2002; Tonry 2007 for international examples). One interesting conclusion has been that, in jurisdictions which have sought to "control" the abuse of discretionary sentencing powers by grids or guidelines, there has often been (paradoxically) an expansion in discretionary powers elsewhere in the process: most obviously, prosecutors may initiate and control plea or sentence bargains, for example, to avoid rigid sentencing options. This entry seeks also to point out a different paradox: that while much attention has been paid in many jurisdictions to attempt to control or limit the discretionary power of the front door sentencer, "backdoor" sentencers still often have vast powers which may in effect undermine attempts to develop consistency. These powers often remain invisible to both the legal and academic communities.

Backdoor Sentencing

Noncustodial Sentences

What happens when a fine is not paid? This is more of a problem in some countries than others. The sanction for nonpayment may be a tougher penalty: a community punishment or even custody. This hardly seems to be a suitable response when the original offence only merited a fine. With community penalties, similar problems may arise in relation to noncompliance. Regular review may be better than allowing noncompliance to develop to such an extent that those involved in supervising the penalty feel that a custodial response becomes inevitable. Who supervises the payment of fines or the performance of community punishments? The answers to these questions vary enormously: fine enforcers may be administrative bodies or court-employed staff or private bodies acting under contract.

The front-line enforcers of community orders are in many countries called "probation officers," but this term is far from universal (they may be criminal justice social workers, justice assistants, or simply part of the prison administration: Padfield 2012a, b). What it is to be a "probation officer" also varies from jurisdiction to jurisdiction and over time within jurisdictions. One interesting issue is whether these "probation officers" should be part of the same service which runs prisons. In England, probation officers were clearly for many decades "officers of the court," but an increasing concern with the "silo mentality" which divided prison and probation services, thus failing to allow sentences to function "seamlessly" from custodial to community parts, led to the creation of an umbrella organization, the National Offender Management Service (NOMS). But rather than leading to a "seamless" sentence, we have seen the further fragmentation of services, exacerbated by shortage of resources in the community part of the system. The Americans use the term "corrections" to create an impression of coherence and continuity. The same arguments are taking place elsewhere: in France the *service pénitentiaire d'insertion et de probation* (SPIP), the equivalent of the English probation service, has long suffered as the poor relation of the prison estate. The SPIP or the probation officer may have delegated powers to enforce or "police" poor performance – more often, they will have to bring the nonconforming offenders back before a "court" (Herzog-Evans 2011a, b, 2012). Sanctions for poor performance can vary enormously. So do processes.

Custodial Sentences: Early Release

In reality, most custodial sentences are in effect partially suspended. Most prisoners leave prison before the full term of the sentence imposed in court has passed. In many countries there may be a system of remission, which is clearly fixed by law. Thus, in France, every prisoner gets an automatic *crédit de réduction de peine* (CRP), which can only be lost in the case of proven bad behavior in prison (the prison governor or prosecutor (*procureur)* can ask the supervising court to disallow the credit): 3 months remission is given for the first year of imprisonment and 2 months per year in subsequent years. The rules are more complex for the most serious offences when the credit can, curiously, be disallowed for medical reasons (see, Reuflet 2010). For short sentences the CRP is 7 days per month. As well as automatic remission, many systems allow a discretionary remission to those who make serious efforts at rehabilitation or to those who have helped the authorities.

Thus, the term "early release" is itself a deeply contested term (see van Zyl Smit and Snacken (2009, Chap. 8). Many authors prefer to speak simply of "release" or "conditional release." But not all releases are necessarily "conditional" (and the weight of conditions can vary enormously). "Early" release of course has both disadvantages and advantages. The fact that the public may see "early release" as a "cop out," a pragmatic and cynical way of letting prisoners escape paying the true price of the their crime, may fuel the political punitive rhetoric which (paradoxically) sees more and more people sentenced to longer and longer sentences. The advantages of "early" release to both the offender and the State are perhaps obvious: by limiting the length of sentences, the overall cost of sentences is reduced. These costs include not only the financial costs (which are enormous) but also the human and social costs. The negative effects of imprisonment are explored elsewhere in this encyclopedia. Where priority is given to the reintegration of prisoners into free society, it will often make sense to speed the process of release as much as possible. Early release rules can also help prison authorities maintain discipline in prison, by offering a "carrot" to encourage good behavior. But not all prisoners seek early release: in some jurisdictions prisoners may prefer to "max out"; in the USA and some European countries, prisoners can in effect refuse parole, serving all of their sentence in prison, or are released unconditionally, without any supervision post-release (see, e.g., Bauwens et al. (2012) on the situation in Belgium).

Who makes the decision to release a prisoner? In some countries, or for some prisoners, the rules may be fixed. The prisoner may have to serve a fixed proportion of their sentence. But many systems allow a certain flexibility: perhaps a decision on release is taken by the prison authorities or by a formal court. In some jurisdictions, the competent authority for conditional release is always a judge (a special chamber of the District Courts in Germany or the *tribunaux d' application des peines* in France). Elsewhere, the decision may be taken by a hybrid body composed of judges and experts or lay members. The Parole Board of England and Wales has been held not to be adequately "independent" to comply with the requirements of Art. 5(4) of the European Convention on Human Rights, which requires that "everyone who is deprived of his liberty by arrest or detention shall be entitled to take proceedings by which the lawfulness of his detention shall be decided speedily by a *court* and his release ordered if the detention is not lawful." But in fact, today, the Parole Board takes few release decisions, now focusing on the release of indeterminate sentence prisoners and those recalled to prison: first release will usually be fixed by law or taken by prison governors (operating a discretionary Home Detention Curfew scheme, which allows prisoners to be released up to 135 days before the midpoint in their sentence (the point at which release becomes automatic), subject to a curfew monitored electronically: see Padfield 2009). This decision should be taken by, or, at the very least, be closely overseen by, independent judicial authorities, especially, it might be argued, in prisons which are run by private sector companies.

The reality, of course, is that the "formal" decision is hugely constrained or may in reality have been taken by other decision-makers, earlier in the process. For example, many "parole boards" are unlikely to authorize release unless a "release plan" has been secured. If a plan has not been prepared, or unless the probation officer supports release, the court or board is unlikely to order or permit release. The extent to which a prisoner is involved in developing a release plan varies enormously: in many jurisdictions, prisoners feel that the "plan" is imposed on them with little discussion or debate. It is not only the presence or absence of a release plan which affects the decision-making. Danzigera et al. (2011) tested the common caricature that justice is "what the judge ate for breakfast." They looked at parole decisions made by experienced judges in Israel and, perhaps surprisingly, concluded that merely taking a food break may indeed lead a judge to rule differently in cases with similar legal characteristics. We should question not only who takes the decision but how the decision is taken.

Is a released prisoner still technically a prisoner? It may be useful to describe them this way in order not to give a false feeling of "freedom." Thus, in England and Wales, the moment of "release" can be confusing. Prisoners may be released at "half time" and yet then be required to live in "approved premises" where the rules may be more restrictive than they were in a prison (where the emphasis will have been more on perimeter security than on internal control). This raises, once again, the thorny question of the categorization of prison sentences and of prison regimes (well discussed in Cid and Tebar 2012).

Prisoners released "conditionally" may be subject to a very wide range of possible conditions. Probably a standard condition (which may or may not need to be articulated) is that the offender should not re-offend. German law distinguishes directives ("Weisungen") from obligations ("Auflagen"). Directives are imposed in order to influence the offender's behavior and to prevent reoffending, whereas obligations are meant to fulfill the victim's or society's desire for satisfaction, for example, by compensating the victim or by making a payment to a nonprofit organization or the state or of community service (Pruin 2012). In many jurisdictions, the conditions include:

– Meeting and keeping in touch with a probation officer
– A residence obligation, with possible curfew, which may or may not be electronically monitored
– Treatment by a psychiatrist/psychologist/medical practitioner
– A positive work obligation (or a requirement not to take work with certain groups, such as children)
– An obligation to make payments to victims
– A requirement not to reside in the same household as children
– A requirement not to approach or communicate with named people
– A requirement to avoid a particular area
– A requirement to attend courses for confronting addictions, improved driving, anger management, etc.
– A drug testing condition

In some US jurisdictions, the conditions may also be loss of civil rights or even a duty to pay supervision fees (Parole and Prisoner Reentry in the United States). Conditions are often complex, difficult to understand, and sometimes contradictory. The literature on life on license is small (but see Appleton (2010) for a detailed and complex picture of the lives of released discretionary life sentence prisoners in England). But when we come to consider the reasons why licenses are revoked, it is important to consider whether the practice of parole supervision might be more supportive (e.g., see Travis 2002). The reality of life for ex-offenders is often challenging, even precarious (see Standing 2011). They face many practical, health, and economic difficulties, including those associated with housing, unemployment, family, and substance misuse. Parole conditions can "get in the way" and make an already precarious lifestyle unworkable. While the conditions may make it easier to monitor the offenders, they can also make it more difficult for people to "go straight" (Maruna 2001). Further, academic research is needed which focuses less on reconviction outcomes and more on prisoners' perceptions of their lived experience (Padfield 2012a, b; Digard 2010, are an excellent small example).

Many systems of discretionary release rely on actuarial risk predictors. Others rely on more clinical assessments. Even those systems which allow "automatic" early release may continue to be used. The Risk Management Authority in Scotland maintains a detailed directory of approved assessment tools, including summaries of the published evaluations of each tool (see www.RMAscotland.gov.uk). Of course, actuarial assessments can only ever be predictors. It is very dangerous to move from these generalized risk assessments to the applications to the lives of real people. There will be false-positives (those who are predicted to re-offend and who don't) and false-negatives (those who are predicted not to re-offend, but who do so). As well, it can be immensely difficult for the individual to "live down" a risk assessment (see Padfield 2011b).

Custodial Sentences: Revocation of License/ Recall to Prison

While there is a small literature on the law and practice of release and parole (particularly in the USA and common law context and increasingly in Europe: see Padfield et al. (2010)), the international academic literature on parole or license revocations seems to be emerging very recently. It can sometimes appear as though academic interest ends with release: as though release is the end of the story. But the reality of release from prison is very different. Successful reintegration, or reentry, is notoriously difficult (see Maruna 2001). Desistance from a criminal lifestyle rarely happens overnight: prisoners serving sentences in the community may be returned to prison for "failures" on license. The costs of incarceration and failure are enormous, but how do we measure success and failure? It may be that insufficient attention is paid to what for some offenders are significant successes (e.g., turning up most of the time for appointments). It is all too easy to conclude that poor attendance is a failure!

Conditional release may end for a variety of reasons. Most common may be allegations of fresh offences. Of course these new offences do not necessarily have to be proved before revocation take place. So the "recall" may be initiated in effect by the police when they arrest a suspect. There are important questions about the status of offenders on license when rearrested: should they be unconvicted prisoners until the new offences are proved? Prisoners may remain in prison even after allegations are dropped or after they have been acquitted in court (see Padfield 2012a, b). Others will be brought back to prison because their behavior on license or on parole has caused significant concern to those who have the responsibly of supervising them in the community (see Collins 2007). These reasons may be very difficult for the prisoner to challenge. There are in many jurisdictions stark contrasts between the judicialized release process and the executive decision to recall to prison. The process for challenging recall may be administrative or judicial. Often it is very slow and impenetrable (see Padfield (2012a, b) and the special issue of the *European Journal of Probation* which carries articles from nine different European jurisdictions: Volume 4, Issue 1, 2012, freely available at www.ejprob.ro). In many jurisdictions the number and percentage of prisoners who are in prison as parole violators or for alleged breaching of their conditions is rising sharply. It may well be that this increasing number may have a significant impact on the stability of the prison population: angry and embittered prisoners are difficult to manage. Thus, there are practical as well as "human right" reasons for us to be concerned about this issue.

Why have scholars not followed developments in the law and practice of release and recall to prison more closely? Part of the explanation may well be the division between criminology and law in many countries. And within law, there is often a vast divide between criminal lawyers (whose role often seems to end when their client is sent to prison) and public or administrative lawyers (who are involved in challenging decisions taken by prison authorities). Yet there does seem to be a slowly growing recognition that backdoor sentencing is as "important" as front door sentencing. Might the time be coming when the same lawyer sees it as their responsibility to support their client throughout the whole of the criminal justice process? The provision (i.e., funding) of good quality legal advice and representation is of course also fundamental.

Another important consideration is the purpose of sentencing, which we started with in the introduction to this entry. Do the same purposes apply to backdoor sentencing as to front door sentencing? In the USA the concept of "reentry" seems to have won a certain symbolic or rhetorical importance (see Petersilia 2003; Travis and Visher 2005). In Europe, the concepts of rehabilitation or reintegration are more common: but all point towards an aspiration to include ex-offenders within the wider community. Yet governments are often more focused on excluding the “dangerous”: on public protection and more exclusionary means of reducing recidivism. The human rights discourse helps to remind us that offenders are individuals, not statistics. If we do not focus on this, we may regret that there is not just a front and a backdoor into the prison system, but simply a revolving door (see Weaver et al. 2012).

Conclusions

This entry has sought in particular to encourage a closer enquiry into the relationship between “front door” sentencing with the less studied “backdoor.” The danger of this brief entry is that it results in simple generalizations. Yet sentencing is a complex process. The decision-making process may be structured by the law, or unstructured, and some stages in the process may be less or more structured than others. A dilemma flows through all legal processes: the tension between the need for certainty and the need for flexibility. In order to be fair, some flexibility may well be necessary. For example, some prisoners will be released early on compassionate grounds: can the grounds for this be spelt out precisely in advance? Flexibility means discretionary powers, and discretion is the first cousin of discrimination. Dworkin’s doughnut (an American ring doughnut, not a British doughnut!), mentioned earlier, is interesting because of the surrounding belt of restriction: the socio-legal context of decision-making (for a fuller introduction to the literature on discretionary decision-making, see Gelsthorpe and Padfield 2003). In seeking to monitor the fairness of sentencing decisions, we need to be particularly vigilant about gender and racial discrimination (see Thompson 2009; Turnbull and Hannah-Moffat 2009; Barry and McIvor 2010; Hannah-Moffat and Yule 2011).

We need therefore to consider who is making these front door and backdoor decisions and in what conditions. We have noted that these “judges” are also severely constrained by decisions taken earlier in the process. Parole or probation officers often hold the key to release. Their role is vital yet confusing: often empowered simultaneously to monitor and to support, as well as to sanction (Lynch 1998). If sentencing is, as is argued here, a series of very different decisions, often made by very different “players” within the process, the importance of close oversight is obvious. Due process, the rule of law and the separation of powers suggest that judicial oversight is essential.

Lawyers, psychologists, psychiatrists, criminologists, sociologists, and anthropologists all have a vital role in developing the analytical tools which help us understand the whole sentencing process: from the moment of arrest perhaps or from charge and plea bargain, to release and recall, to successful reintegration. Vast differences exist between times and place, and political agenda and legal frameworks vary. Yet there are some common themes, which include an increasing literature on legitimacy and on fairness in sentencing. The right to individual liberty is fundamental; the power of the state should never be arbitrary. The rules on sentencing ought to be clear and to be enforced fairly and transparently.

Related Entries

- ▶ Evolving Judicial Roles
- ▶ History of Probation and Parole in the United States
- ▶ Parole and Prisoner Reentry in the United States
- ▶ Plea Bargaining
- ▶ Sentencing as a Cultural Practice
- ▶ Sentencing Research

Recommended Reading and References

Appleton C (2010) Life after life imprisonment. Oxford University Press, Oxford

Barry M, McIvor G (2010) Professional decision making and women offenders: containing the chaos. Probat J 57:27–42

Bauwens A, Robert L, Snacken S (2012) Conditional release in Belgium: how reforms have impacted recall. Eur J Probat 4:19–33

Cid J, Tébar B (2012) Revoking early conditional release measures in Spain. Eur J Probat 4:112–124

Collins H (2007) A consideration of discretion, offender attributes and the process of recall. In: Padfield N (ed) Who to release. Willan, Cullompton, pp 173–190

Danzigera S, Levavb J, Avnaim-Pessoa L (2011) Extraneous factors in judicial decisions. Proc Natl Acad Sci USA. www.pnas.org/content/early/2011/03/29/1018033108.full.pdf+html

Delmas-Marty M (ed) (2003) L'Harmonisation des Sanctions Pénales en Europe. Paris: Unité Mixte de Droit Comparé de Paris

Digard L (2010) When legitimacy is denied: offender perceptions of the prison recall system. Probat J 57:43

Dworkin R (1977) Taking rights seriously. Duckworth, London

Gelsthorpe L, Padfield N (2003) Introduction. In: Gelsthorpe L, Padfield N (eds) Exercising discretion: decision-making in the criminal justice system and beyond. Willan, Cullompton, pp 1–28

Gualazzi A, Mancuso C, Mangiaracina A (2010) Italy. In: Padfield N, van Zyl Smit D, Dunkel F (eds) Release from prison: European policy and practice. Willan, Cullompton

Hannah-Moffat K, Yule C (2011) Gaining insight, changing attitudes and managing 'risk': parole release decisions for women convicted of violent crimes. Punishm Soc 13:149–175

Herzog-Evans M (2009) French post custody law (2000–2009): from equitable trial to the religion of control. Eur J Probat 2:97–109

Herzog-Evans M (2011a) Droit de l'exécution des peines, 4th edn. Dalloz, Paris

Herzog-Evans M (2011b) Desisting in France: what probation officers know and do. A first approach. Eur J Probat 3:29–46

Herzog-Evans M (2012) Non-compliance in France: a human approach and a hair splitting system. Eur J Probat 4:46–62

Hood R, Hoyle C (2008) The death penalty: a worldwide perspective. Oxford University Press, Oxford

Lynch M (1998) Waste managers? The new penology, crime fighting, and parole agent identity. Law Soc Rev 32:839–870

Maruna S (2001) Making good: how ex-convicts reform and rebuild their lives. American Psychological Association Books, Washington, DC

McNeill F, Burns N, Halliday S, Hutton N, Tata C (2009) Risk, responsibility and reconfiguration. Punishm Soc 11:419–442

Padfield N (ed) (2007) Who to release? Parole, fairness and criminal justice. Willan, Cullompton

Padfield N (2009) Parole and early release: the Criminal Justice and Immigration Act 2008 changes in context. Crim Law Rev :166

Padfield N (2011a) An entente cordiale in sentencing. Crim Law Justice Wkly 175:239–242, 256–9, 271–4 and 290–293

Padfield N (2011b) The sentencing, management and treatment of 'dangerous' offenders: Final Report. European Committee on Crime Problems

Padfield N (2011c) Time to bury the custody "Threshold"?. Crim Law Rev 593–612

Padfield N (2012a) Editorial. Eur J Probat 4:34–45

Padfield N (2012b) Recalling conditionally released prisoners in England and Wales. Eur J Probat 4:34–45

Padfield N, Maruna S (2006) The Revolving door at the prison gate: exploring the dramatic increase in recalls to prison. Criminol Crim Justice 6:329–352

Padfield N, van Zyl Smit D, Dunkel F (eds) (2010) Release from prison: European policy and practice. Willan, Cullompton

Padfield N, Morgan R, Maguire M (2012) Out of court out of sight? Criminal sanctions and non-judicial decision making. In: Maguire M, Morgan M, Reiner R (eds) The Oxford handbook of criminology, 5th edn. Oxford University Press, Oxford

Petersilia J (2003) When prisoners come home: parole and prisoner reentry. Oxford University Press, New York

Pruin I (2012) Recalling released prisoners in Germany. Eur J Probat 4:63–72

Reuflet K (2010) France. In: Padfield N, van Zyl Smit D, Dunkel F (eds) Release from prison: European policy and practice. Willan, Cullompton

Roberts JV (ed) (2011) Mitigation and aggravation at sentencing. Cambridge University Press, Camabridge

Sebba L (2001) When is a prisoner not a prisoner? "service work" in Israel – and in Britain? Crim Law Rev :543–559

Standing G (2011) The precariat: the new dangerous class. Bloomsbury, London

Tata C, Hutton N (eds) (2002) Sentencing and society: international perspectives. Ashgate, Aldershot

Thompson A (2009) Releasing prisoners redeeming communities: reentry race and politics. New York University Press, New York

Tonry M (ed) (2007) Crime, punishment, and politics in comparative perspective. University of Chicago Press, Chicago

Tonry M, Frase R (eds) (2001) Sentencing and sanctions in western countries. Oxford University Press, Oxford

Travis J (2002) But they all come back: facing the challenges of prisoner reentry. Urban Institute Press, Washington, DC

Travis J, Visher C (2005) Prisoner reentry and crime in America. Cambridge University Press, New York

Turnbull S, Hannah-Moffat K (2009) Under these conditions: gender, parole and the governance of reintegration. Br J Criminol 49:532–551

Van Zyl SD, Snacken S (2009) Principles of European prison law and policy: penology and human rights. Oxford University Press, Oxford

Weaver B, Tata C, Munro M, Barry M (2012) The failure of recall to prison: early release, front-door and backdoor sentencing and the revolving prison door in Scotland. Eur J Probat 4:85–98

Gangs and Social Networks

Christian L. Bolden
Criminal Justice, Loyola University New Orleans, New Orleans, LA, USA

Synonyms

Hybrid gangs

Overview

Traditional depictions of gang organization have been erroneous or static or were unable to be measured. Relying on these traditional depictions, modern rhetoric regarding organization is that gangs have evolved into hybrid groups that participate in nontraditional behavior such as merging and cooperating with rival gangs. Examining this phenomenon through emic perspectives suggests an alternate explanation. Previous literature has indicated that gangs are very loosely organized, and more recent works have demonstrated that gangs are social networks rather than structured groups. The following essay comparatively examines gang organization through the social network lens.

Key Issues/Controversies

Gang organization has long been a topic of dispute in gang research. Outside of the academic realm remains a pervasive conceptualization of gangs in the American media and law enforcement rhetoric as one of national, highly organized, violent, drug-dealing entities. If this viewpoint were true, then it would indeed be worthy of the fear that it inspires. The problem is that academic research consistently shows that there is little to no validity to this conceptualization (Klein 1995). From the early gang literature, continuing to present day research, there are no strong indications that gangs are highly organized, but instead are loose, fluid affiliations.

If gangs are not the national entities they are believed to be, then what organizational structure do they have? After the millennium, police reports began to note "new" organizational structures that law enforcement agencies were unfamiliar with. This lead to further definitional problems and the introduction of groups referred to as "hybrid gangs." Believed to be significantly different than "traditional gangs" in composition and behavior, hybrid gangs are described as having:

> . . .members of different racial/ethnic groups participating in a single gang, individuals participating in multiple gangs, unclear rules or codes of conduct, symbolic associations with more than one well-established gang (e.g., use of colors and graffiti from different gangs), cooperation of rival gangs in criminal activity, and frequent mergers of small gangs. (Starbuck et al. 2004, p. 200)

More and more law enforcement agencies began reporting this phenomenon,

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

specifically as characteristic of late-onset gangs or gangs that appeared in cities post-1990 (Starbuck et al., 2004), and it began entering the collective knowledge of the mainstream. In 2008, the History Channel's series *Gangland* aired an episode called "Sin City." This episode, set in Las Vegas, Nevada, highlights hybrid gangs as an evolutionary advancement and emphatically states that multiple associations among gang members of different gangs are more deadly, but never explains why.

The sensationalism of claims that gangs have reached evolutionary advancement and presented increased danger called for further investigation, more so due to "hybrid gangs" not appearing in the literature beyond Thrasher's (1927) reference to gangs that included different racial/ethnic groups. However, under closer scrutiny, the fluid behaviors associated with hybrid gangs have been appearing in the literature since Yablonsky (1959) argued that gangs were near groups rather than solid organizations. More recently, some research has presented evidence that gangs are porous social networks intertwined with other gang networks and have been so for quite some time (Fleisher 2002). Differing viewpoints may be explained by the use of etic or emic methodology. Etic methodology is the imposition of an outsider's (i.e., law enforcement) interpretation of a phenomenon; in this case the behavior is believed to be a new gang structure. Alternatively, using emic methodology or understanding phenomenon from the respondent's point of view or, in this case, the gang member's point of view explains gang fluidity and nontraditional membership as elements of a social network system (Fleisher 2002; Weisel 2002).

Outside observers may be imputing meanings to gang behaviors that would be understood quite differently from the perspective of the gang member. Examining the literature more thoroughly may help clear up the disputed perspectives on the subject: Is the "hybrid gang" a new phenomenon? Are gangs more like social networks than organizations? Or are these social constructions that amount to "putting old wine into new bottles?"

Structural Typologies

Research has consistently presented gangs as marginally or very loosely organized and widely varied in the activities in which they engage. These findings have led to differing ways in how scholars have interpreted gang structures. For example, in a national survey of 385 police agencies, 45 % indicated that the typical gang was loose knit, and 47 % noted no formal structure in typical gangs (Weisel 2002). However, Weisel (2002) also reported regional differences in gang structure with police indicating more loosely structured delinquent gangs in the Southeast and Midwest, more violent gangs in the West, and more income-generating gangs in the Northeast. Smaller cities were more likely to have delinquent gangs, which engaged in criminal activity but had little involvement with drugs. Furthermore, violent gangs and drug selling gangs most often did not have consistent leadership or a highly structured organization.

Weisel (2002) takes an unusual step and includes in-depth interviews with members of the Black Gangster Disciples and Latin Kings in Chicago and Lincoln Park Piru and Logan in San Diego. Consistent with stereotypes, the Black Gangster Disciples and Latin Kings had extensive organizational structure. In accordance with most literature, the San Diego gangs had little structure, and members considered the groups to be friendship/kinship networks. While some gangs appear to be highly organized, an extensive amount of research indicates that most gangs are not as highly organized as generally believed (Klein 1995; Huff 1996; Decker et al. 1998; Fleisher 1998; Miller 2001).

Attempting to make systematic sense of gang organization has led to a plethora of gang typologies. Although these typologies may be useful in identifying characteristics of a gang at the time of the study, they present the assumption that the gang is static. Typologies focusing on particular criminal activity do not capture the social processes of the gang or a particular gang's relationships with others. Some of the typologies rely on precarious variables such as amount of drug use and type of crime engaged in or threat level, which have extreme within-group variation and are not stable over time. Klein's (2002) structural typology provides measurable variables and thus more utility, but does not take fluidity, permeability, or networking behavior into account.

Gang structures are anything but static, but the dynamic gang processes of changing, merging, or splintering have rarely been examined in considerable depth. Most authors have been cursory on the subject, and there has been little in-depth investigation. The information that has been gained concerning these processes was obtained, while scholars investigated gang structures and whether or not they had changed over time.

An intriguing contribution to the extant knowledge on the subject comes from Weisel, (2002) whose interviews portray mergers such as the Black Gangster Disciples forming from the combined gangs of the Black Disciples, Gangster Disciples, and High Supreme Gangsters. Interviews in San Diego also revealed that Logan splintered into "Logan Trece" and "Red Steps" as the gang grew larger and natural boundaries emerged. Unfortunately more detailed information about these mergers and splinter groups was not provided. Spergel (1990) argues that splintering develops from internal competition or if more criminal opportunity becomes available. Monti (1993) argues that it is the lack of control over larger gangs that causes them to split and that age-graded cliques are like gang building blocks that can merge, dissolve, and reassemble. Weisel (2002) views the process of merging and splintering through organizational theory, arguing that a path to organizational equilibrium explains why some groups dissipate and others break off from larger groups until a stable number of organizations are reached. This theoretical approach explicitly ignores the gang member worldview in favor of the assumption that gangs can be called organizations.

It is no surprise that relational behavior and emic perspectives are not often considered when discussing gang organization because doing so would recognize gangs as dynamic processes rather than static structures. Processes are much harder to define and subsequently eschewed by agencies that would like clearly labeled categories. Even though gang relationships are often sidestepped, they are nonetheless important. The gang as a criminological topic is predicated on the idea that individuals are committing crime in a group. The group as characterized by names, colors, symbols, and the like has become such a preoccupation that the relationship of "who" is committing crime together has been ignored. Not all organizational arguments have missed this dynamic. In Cloward and Ohlin's (1960) typology of delinquent subcultures, they recognized that criminal subcultures necessitated adult criminals to indoctrinate younger people into illicit activities or at the very least model illegitimate behavior.

Although limited by focusing only on Mexican-American gangs in San Antonio, Texas, Valdez (2003) provides a typology that further bridges gang structure and social network dynamics. One of the gang types is the "criminal-adult-dependent gang," which is a highly organized group that is focused on earning illegal income usually through drug sales. Adults outside of the gang provide the group with weapons, drugs, stolen merchandise, protection, and extensions to their criminal networks. Valdez (2003) notes two subtypes of this category. One of the subtypes is a family network in which the adult criminals are closely related to the gang members. The other subtype is dependent on a prison gang, which exerts control over the street gang. Either way, the importance of gang relationships is stressed. If members of different categorical groups work together, then at what point do organizational lines become blurred or nonexistent?

Hybrid Gangs or Social Networks?

Largely untouched by academic researchers, the "hybrid" gang phenomenon has been discussed primarily by law enforcement agencies such as Missouri's Kansas City Police Department (Starbuck et al. 2004). The "hybrid" gang is not a new term or idea. It was initially used by

Thrasher (1927) to describe gangs of mixed race/ethnicity, but in the modern era, the term encompasses many other characteristics as well. Different than most traditional depictions of gangs, which described gangs as being comprised mostly of lower class and minority males, police in many jurisdictions are starting to report hybrid gangs. These gangs in late-onset localities (i.e., post-1990) have a greater mix of race/ethnicity, with an increase of white youth, the presence of more females, and a larger proportion of middle-class teens (Howell et al. 2002). Additionally, it appears that members may switch gangs or participate in multiple gangs (Starbuck et al. 2004). For example, in San Antonio, 8 out of 15 former gang members interviewed had switched gangs or belonged to multiple gangs, and two gangs had switched their entire allegiance (Bolden 2012). Although San Antonio agencies have reported gangs since the 1950s, the gangs appeared to have the same "hybrid" characteristics that were being pointed out in emergent gangs.

Cooperation between gangs that are sometimes rivals is also noticed by law enforcement, as is the mixing of gang symbols from Chicago- and Los Angeles-based gangs. Just as these gangs are engaged in cafeteria-style offending, they are selecting characteristics of different established gangs. This gives further credence to the idea that gangs have now become popular brand names (Klein 1995). The use of the term "hybrid gangs" to describe groups that have more fluid membership and nontraditional membership (Starbuck et al. 2004) is an illustration of etic methodology,

In emic methodology, the idea of self-identifying oneself as a member of a particular gang becomes relevant. Fleisher (2002) argues that self-nomination as a gang member refers to both an attribute and a relational aspect of membership. Gang research tends to examine membership as an attribute and neglect the relational component. However, it is of great importance that self-nomination is a statement of having a particular relational status to other people. Fleisher (2002) used social network analysis to examine the nature of gang member relationships between females and how these affiliations affected gang participation. In explaining a gang member's ego network, which is the group of people that an individual directly interacts with on a regular basis, Fleisher (2002) found many members had relationships with people from other gangs than their own. Furthermore, while gang members certainly associated with other members of their gang, Fleisher found that none of the gang members knew all of the other members in their gang. The members who knew the most members in their gang only knew 10 % of the members in their gang. The gangs studied ranged from large Midwestern gangs such as the Gangster Disciples and Vice Lords to an independent gang called the Fremont Hustlers. It would seem that his conclusions may not be as relevant to smaller gangs, but Fleisher was revealing that individuals knew, interacted with, and spent time with several other gang members, in different gangs, rather than exclusively with people who identified themselves as being in the same gang.

The actual ego networks of each gang member, or the people they regularly interacted with, were fairly small and often included members of other gangs. Fleisher (2002) argued that the status of gang membership provided social capital and being included in the social networks of other gangs would further increase someone's social capital. While Bolden (2012) notes that gang members often referred to positive interactions with members of other gangs, Decker et al. (1998) provide one of the only studies that actually examined relationships among gangs. Studying 26 Gangster Disciples and 18 Latin Kings in Chicago, as well as 20 members of Logan Heights and 21 members of Lincoln Park Piru in San Diego, Decker et al. (1998) found that relationships with other gangs were very common. All of the San Diego gang members reported maintaining relationships with other street gangs, while 80 % of the Gangster Disciples and 75 % of the Latin Kings maintained these types of relationships. Furthermore, all of the Latin Kings and Logan Heights members maintained relationships with prison gangs. Eighty-seven percent of the Gangster Disciples and 75 % of the Lincoln Park Piru also maintained relationships with prison gangs. Although some would argue that gang alliances are brittle (Monti 1993), they do not deny that gangs assist each other in varying ways. Decker et al. (1998) provide us with evidence that in spite of often being overlooked, gang relationships are common and these networks may be an important part of the gang experience.

With data from a gang task force in New Jersey, McGloin (2007) also provides evidence that gangs are more aptly described as social networks and rather than being structured organizations, the gang boundaries are dynamic and opaque. Also using social network analysis, Papachristos (2006) did not find cohesion in gangs as a whole, but strong cohesion in subgroups of the gang. These ego networks were responsible for specific crimes and behaviors indicating that crimes are ego network related rather than gang related or motivated (Fleisher 2002). This is an alternate interpretation to the idea of organizational cooperation between rivals in hybrid gangs (Starbuck et al. 2004). Understanding gangs from a social network standpoint can help clear up the ambiguities in determining whether individuals or whole groups work together. It can also examine whether the "hybrid" label is a valid concept or if it is a misinterpretation of kinship/friendship networks.

The gang members in Weisel's (2002) study saw the gang as a friendship network, which is consistent with the findings of Fleisher (2002). Furthermore, though Yablonsky (1973) has been attacked for his depiction of the gang as sociopathic and violent, few have paid attention to his concept of the gang as near group. Yablonsky (1959) reported that gang members had no measurable number of members, no definition of membership, no specific roles, no understood consensus of norms, and no clear flow of leadership to action. Weisel (2002) placed the particular gangs studied in an organizational context because they portrayed orientation towards goals and organizational continuity. However if gangs lack the vast majority of organizational aspects as pointed out by Yablonsky (1959), can they really be considered organizations?

Viewing gangs in an organizational context forces categorical boundaries that may only exist in the mind of the outside observer. As gang characteristics noted by Yablonsky (1959) indicate, there is much more fluidity to gangs and gang members, and gang boundaries may be much more porous. The organizational viewpoint may also lead to an ecological fallacy of assuming that members of different gangs engaging in activity together means gangs are working together. Finally, this idea ignores the viewpoint of gang members that gangs are kinship/friendship networks. Confusing or misinterpreting the relationships of gang members may have led to the present label of "hybrid gangs." Viewing gang processes from a social network viewpoint clears up this confusion by distinguishing relationships among gang members as well as among gangs.

State of the Art

Examination of what law enforcement call hybrid gangs is in its infancy and as of yet has not received serious empirical investigation. Whether gangs have actually changed or these processes have previously existed and been overlooked, the "hybrid" gang label implies fluid membership between gangs, a lessening of violent events in joining and leaving gangs, more interaction among gangs, and a selective mix of identifying elements of well-known gangs. These gangs are also claimed to be late-onset (i.e., post-1990) gangs that are characterized as having an increase in females, an increase of whites, and an increase of middle-class youth. These apparent differences challenge many previously used assumptions of gangs as social islands with impermeable boundaries.

A considerable amount of literature informs us that gangs tend to not be highly organized structures but rather loose conglomerations of clique structures (Decker and Curry 2000; Fleisher 2002; Klein and Crawford 1967; Papachristos 2006). These conglomerations are not highly cohesive as a whole, albeit stronger cohesion occurs among particular cliques. These smaller cohesive cliques dilute the influence of the overall gang in favor of the immediate people that an individual interacts with. Lerman (1967, p. 71) describes the gang subcultural unit as "a network of pairs, triads, groups with names, and groups without names." What has not been discussed in the literature, with the notable exception of Fleisher (2002, 2006), is who is actually in these particular cliques.

If these smaller cohesive units are in networks, then the most viable framework for examining gang member relationships emerging from previous literature is social network analysis (SNA). Social network analysis is both a theoretical framework and a methodological approach. From a theoretical standpoint, people belong to intricate webs of social relationships that influence their lives in a myriad of ways and affect occupational chances, general opportunities, and perceptions of the world (Simmel 1955; Papachristos 2006). Social networks can be analyzed on the group level through degree centrality and density or by using individual ego networks as the unit of analysis. The ego network refers to the social ties/bond the individual has to other individuals. Until recently social network analysis has rarely been used to study gangs. Klein and Crawford (1967) and, more recently, McGloin (2005, 2007) and Papachristos (2006) have used this framework to examine cohesion of members within a gang, finding that there were cohesive subgroups or cliques but not strong cohesion in the gang as a whole. Fleisher (2006) used nomination of friends to identify ties between members of different gangs such as the Gangster Disciples, Vice Lords, and Stones, which are sometimes rivals. Fleisher argued that affiliation in the same categorical gang was not sufficient to foster sentiment between members. Even if members hung out with each other, they often indicated preference for other friends that were not a part of their own particular affiliation. Preference was also related to the social capital created by network relationships. Social capital in networks makes more actions and opportunities available (Papachristos 2006). Fleisher (2006) argues that even though belonging to a gang provides a level of social capital, gang member relationships are based more on the expanded social capital that a connection provides rather than affiliation with a particular group.

Fleisher (2006) explains that the use of other methods to study gangs has resulted in the concept of the bounded group. Although it makes obvious sense that gang members, like most other people in society, interact in many different social circles and utilize agency in choosing who to associate with, resulting in their not being bound by the gang per se, the conventional depiction of the gang member as the folk devil allows for easy disregard of viewing any behavior of the gang member as normal. Ironically, Cotterell (1996) comes to conclusion that interactional behavior between gang members is actually more fluid and less stable than other adolescent cliques. Cliques are groups of people who spend time together. Usually adolescents belong to many cliques with different sets of friends in varying contexts, such as sports teams, neighborhood friends, and school friends. In gangs, however, membership provides the individual with the social capital to more freely move between cliques. Cotterell (1996, pp. 33–34) describes gangs as "a series of changing microsystems. The individual joins one group for a time, then leaves, and rejoins or moves on to another."

Ayling (2009), who views gangs as organized criminal networks, theoretically argues that the weak links between gang network hubs or "loose couplings" make the gang functionally resilient against both law enforcement suppression and attacks from other groups. Damage done to one hub or clique will not destroy the entire network. Furthermore, the clique type network removes the sluggish and burdensome chain of command, allowing members enough freedom to instantly act and have improvisational responses to immediate concerns. Using police data, McGloin (2005) identified particular gang members as “cut-points” or the only connection or intermediaries between the different cliques within a gang.

Controversies and Questions

Methodology

Papachristos (2006, 2009) challenges Fleisher’s (2006) characterization of gang membership as relational attributes and argues instead that they are social groups based on the patterned actions that are caused by relational ties. Arguing that gangs are groups and not “pedagogical constructs,” Papachristos (2009) examines gang network relationships and demonstrates predictable patterns of homicide activity. Fleisher (2006) explains however that methodological choice will cause this discrepancy, and indeed Papachristos uses (2006, 2009) police data to examine gang networks.

Using emic methodology and qualitative interviews in San Antonio, Texas, and Orlando, Florida, Bolden (2010) expands on Fleisher’s work, which was conducted with female gang members, by not only identifying ties among male and female members of different gangs but also the nature and consequences of those network ties in regard to “hybrid” gang processes, such as belonging to multiple gangs, switching gangs, and fluidity of joining or leaving gangs. The social network of the gang member allows for expansion of social capital and expanded opportunities in the urban arena.

The following example is an ego-network depiction of a gang member in Orlando, Florida (see Fig. 1). **Caribe** self-identifies as belonging to Hoover Folk, yet his most common interactions were with members of different gangs. His social group that he spent time partying and getting high with consisted of two Bloods and a Latin Queen, all of which are the traditional primary enemies to his gang. The other members of Caribe’s ego network, a Crip and a member of a Jamaican Posse, are his business associates that he engages in illicit profit-oriented activity with.

Caribe-Hoover Folk	Progeny-ATF Compton Crip
Curly/Smokey-Nine-Trey Gangster Bloods	
Machete-Young Shottaz (independent)	Silk-Latin Queen

This example of one gang member’s ego network has profound implications for the study of gang organization, which include:

(a) This is not a categorical group (no shared name, signs, symbols, etc.).

(b) From an etic standpoint, this would be seen as one group and labeled as a hybrid gang. From an emic standpoint, this is not one group, but two separate segments of an ego network.

(c) Criminal activity is being committed by gang members together, but they are not in the same categorical gang. This becomes especially relevant to jurisdictions that define gang activity along the lines of categorical variables.

(d) Rather than gangs cooperating, this illustrates individual gang members cooperating.

(e) If the crime occurring in this network is considered gang crime, which bounded categorical group is attributed with the crime? If these actors commit a homicide together, which gang is credited for it?

Bolden (2010) found that all but 2 of the 48 current and former gang members he interviewed indicated that strong relationships along the lines of kinship, business, friendship, conflict-partners, sports, school, and romance existed between themselves and members of rival gangs. Many of these relationships were primary, in the sense that more time and activity, both criminal and benign, was spent with members of other gangs than members of their own gangs. Also of interest was that 22 of the interviewees were gang migrants from Chicago, Los Angeles, and New York, where they described the same dynamics occurring.

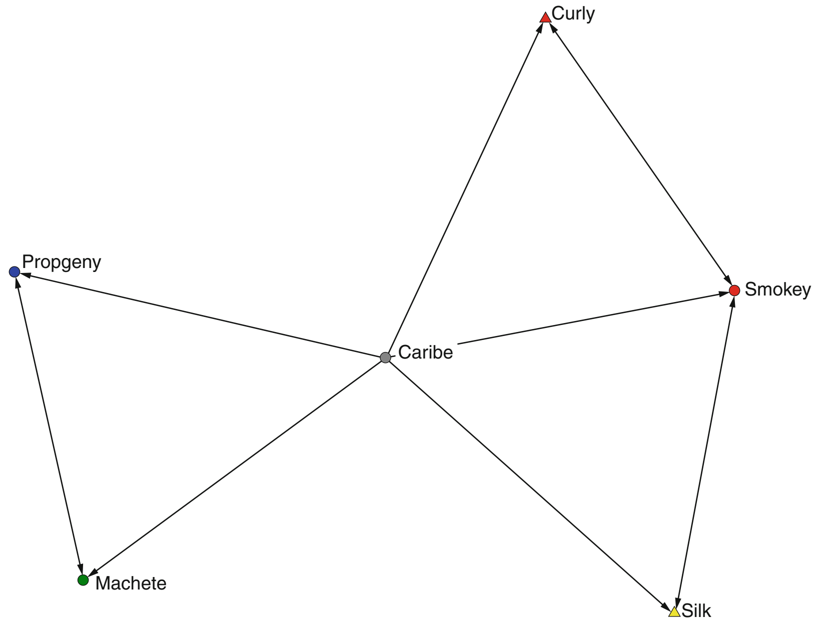

Gangs and Social Networks, Fig. 1 Caribe ego network

Evolution

Though there is more academic support for understanding gangs as social networks than there is for the "hybrid gang" label, the question of emergence still remains. Both Fleisher's (2002) and Bolden's (2010) studies were post-1990, clearly placing the networks studied in the late-gang onset era. This is not evidence that the networking phenomenon is new. It could be that the behavior was not previously noticed. Indeed, four of Bolden's (2010) interviewees participated in gang activity prior to the 1990s (one in the 1960s, three in the 1970s), and all indicated the same relational activities. Furthermore, Reymundo Sanchez (2000) chronicles his criminal activity as a Latin King in the early 1970s, which includes more criminal cooperation with members of other gangs including rivals than activity with the Latin Kings.

The evidence of preexistence, though not strong, indicates that to some extent the cooperation between rival gang members is not a new occurrence. This question remains open, but as it concerns history, it may be that we cannot ascertain the answer. What is more pertinent are the modern conceptions of gang organization. Will evidence that gangs are fluid, dynamic, network structures continue to be overlooked by all but a few in favor of static structural interpretations, or will we deign to entertain other arguments about gang organization?

Social network analyses of gangs are not new but have been relatively unused. Investigative hurdles in this regard are the lack of dissemination of SNA as a viable research option and training in this methodology. These analyses may require specific software programs such as UCINET that researchers have not had any training with. Furthermore, the data used in these analyses have to be very specific, and any quantitative data gleaned from police reporting would need some detailed component of gang member relational ties. Due to this major hurdle, researchers need to find agencies that already keep rich data or direct the data collection process from the onset of a study.

Key studies using social network analysis have been few and far between largely due to the difficulty of gaining access to the required information and the lack of dissemination regarding SNA as a viable option. Expansions of studies of this nature are extremely important because the question of organization has characterized the field of gang research for quite some time. The vast majority of literature indicates that most gangs are loose-knit pseudo-organizations, yet the idea of an organized group persists in the collective mind of the public and law enforcement. Examining the literature more carefully and more specifically from the gang member's perspective rather than an outsider's viewpoint indicates that the gang is a fluid network of connections between members of different "groups" that give said members the social capital needed for expansion and access in the street. Perceived gang boundaries seem to be more of an imposition of an outsider's insistence on categorical distinctions but may have nothing to do with the actual social activity of the gang member. Indeed, outsider insistence on categorical relationships ignores relational ties and their importance when it comes to who is actually committing crime together. Using alternate tools for study will help our field avoid the trap of adhering to outdated concepts and bring us closer to understanding the phenomenon at hand.

Related Entries

- ▶ Organized Crime, Types of
- ▶ Race, Ethnicity, and Youth Gangs
- ▶ Social Network Analysis of Organized Criminal Groups
- ▶ Social Network Analysis of Urban Street Gangs

Recommended Reading and References

Ayling J (2009) Criminal organizations and resilience. Int J Law Crime Justice 37:182–196

Bolden CL (2010) Evolution of the folk devil: a social network perspective of the hybrid gang label. Dissertation. The University of Central Florida, Orlando

Bolden CL (2012) Liquid soldiers: fluidity and gang membership. Deviant Behavior 33:207–222

Cloward R, Ohlin L (1960) Delinquency and opportunity. Free Press, New York

Cotterell J (1996) Social networks and social influences in adolescence. Routledge, New York

Decker S, Curry GD (2000) Addressing key features of gang membership: measuring the involvement of young members. J Crim Justice 28:473–482

Decker S, Bynum T, Weisel D (1998) A tale of two gang cities: gangs as organized crime groups. Justice Q 15(3):395–425

Fleisher MS (1998) Dead end kids: gang girls and the boys they know. University of Wisconsin Press, Madison

Fleisher MS (2002) Doing field research on diverse gangs: interpreting youth gangs as social networks. In: Huff CR (ed) Gangs in America III. Sage, Thousand Oaks, pp 199–217

Fleisher MS (2006) Youth gang social dynamics and social network analysis: applying degree centrality measures to assess the nature of gang boundaries. In: Short JF Jr, Hughes LA (eds) Studying youth gangs. Altamira Press, Oxford, UK

Howell JC, Moore JP, Egley A Jr (2002) The changing boundaries of youth gangs. In: Huff CR (ed) Gangs in America III. Sage, Thousand Oaks, pp 3–18

Huff CR (1996) The criminal behavior of gang members and nongang at-risk youth. In: Huff CR (ed) Gangs in America, 2nd edn. Sage, Thousand Oaks, pp 75–102

Klein MW (1995) The American street gang: its nature, prevalence, and control. Oxford University Press, New York

Klein MW (2002) Street gangs: a cross-national perspective. In: Huff CR (ed) Gangs in America III. Sage, Thousand Oaks, pp 237–254

Klein MW, Crawford LY (1967) Groups, gangs, and cohesiveness. J Res Crime Delinq 30(1):75–85

Lerman P (1967) Gangs, networks, and subcultural delinquency. Am J Sociol 73(1):63–72

McGloin JM (2005) Policy and intervention considerations of a network analysis of street gangs. Criminol Public Policy 4(3):607–635

McGloin JM (2007) The organizational structure of street gangs in Newark, New Jersey: a network analysis methodology. J Gang Res 15(1):1–34

Miller WB (2001) The growth of the youth gang problem in the United States: 1970–1998 (NCJ 181868). Office of Juvenile Justice and Delinquency Prevention, U.S. Department of Justice, Washington, DC

Monti DJ (1993) Origins and problems of gang research in the United States. In: Cummings S, Monti DJ (eds) Gangs: the origins and impact of contemporary youth gangs in the United States. State University of New York Press, Albany, pp 3–25

Papachristos AV (2006) Social network analysis and gang research: theory and methods. In: Short JF Jr, Hughes LA (eds) Studying youth gangs. Altamira Press, Oxford, UK, pp 99–116

Papachristos AV (2009) Murder by structure: dominance relations and the social structure of gang homicide. Am J Sociol 115(1):74–128

Sanchez R (2000) My bloody life: the making of a Latin King. Chicago Review Press, Chicago

Simmel G (1908/1955) Conflict and the web of group affiliations (trans: Wolf KH, Bedix R). The Free Press, New York

Spergel I (1990) Youth gangs: continuity and change. Crime and Justice 12: 171–275.

Starbuck D, Howell JC, Lindquist DJ (2004) Hybrid and other modern gangs. In: Esbensen FA, Tibbetts SG, Gaines L (eds) American youth gangs at the millennium. Waveland Press, Long Grove, pp 200–214

Thrasher FM (1927/1963) The gang: a study of 1,313 gangs in Chicago. University of Chicago Press, Chicago

Valdez A (2003) Toward a typology of contemporary Mexican American youth gangs. In: Kontos L, Brotherton D, Barrios L (eds) Gangs and society: alternative perspectives. Columbia University Press, New York, pp 12–40

Weisel DL (2002) The evolution of street gangs: an examination of form and variation. In: Reed WL, Decker SH (eds) Responding to gangs: evaluation and research (NCJ 190351). National Institute of Justice, U.S. Department of Justice, Washington, DC, pp 25–65

Yablonsky L (1959) The delinquent gang as a near group. Soc Probl 7(2):108–117

Yablonsky L (1973) The violent gang. Penguin Books, New York

Gendarmerie Policing

Clive Emsley
Department of History, The Open University,
Milton Keynes, UK

Overview

Gendarmerie is a form of policing that is primarily military in its structure, organization, and chain of command. The form and the name originated in France, but during the nineteenth century, the institution spread across Europe, the European world, and its empires. Generally speaking these institutions functioned alongside other police bodies that were more civilian in appearance and which answered to local government or to ministers whose portfolio was internal domestic or judicial affairs; in contrast the Gendarmeries usually were answerable to a minister of war/defense.

Origins

The contemporary French *Gendarmerie nationale* likes to trace its origins back to the Middle Ages, and the knights commissioned to protect the king's territories while he was absent on crusade. A more plausible beginning is the Edict of Paris (1539) by which François I extended the competence of the provosts of the marshals of the French Army (*les prévôts des maréchaux*) and of their companies (*maré chaussées*) to include anyone guilty of highway robbery within France. By the end of the reign of Louis XIV in 1714, the companies' jurisdiction had been extended to most serious offenses from burglary to murder, from rape to arson, and from popular disorder to coining. But, in keeping with the general complexities of the old regime, there was little uniformity in structure and the ranks. Moreover, many men had purchased their ranks for the various opportunities that such a position offered rather than through any desire to enforce the king's law (see, in general, Luc 2005).

The death of Louis XIV and the end of his incessant wars left a weak treasury and fears that a flood of mendicants and vagabonds would engulf France; the lack of money limited what could be done about these fears. However, in 1720 Claude Le Blanc, the minister of war, initiated a series of reforms designed to improve the *maréchaussées* and thus the internal security of provincial France. Henceforth, the companies were to have a uniform hierarchy and structure, and they were to be financed by the state. The jurisdiction of each individual company was to be tied to a *généralité*, the most meaningful unit of provincial administration; the men, recruited from former soldiers with good records, were to be stationed in bodies of generally four to six men in small barracks that were positioned on the main roads. This was essentially the structure that remained in operation until the French Revolution though there were minor amendments, such as the introduction of a system of patrolling in pairs carrying a *Journal de Service* that was to be signed by a notable such as the mayor or the *curé* in each community that the patrol visited; the *Journal* was inspected monthly by senior officers. In addition there was a gradual increase in the size of the force from around 3,000 men following Le Blanc's reforms to some 4,000 on the eve of the Revolution.

As with most police institutions, the *maréchaussée* had its critics and its supporters. There were those who protested that patrols were never to be found when they were required and that the men could be high handed, brutal, and corrupt or else too old, too infirm, or drunkards incapable of carrying out their duties. Local administrators disliked the fact that the men were not responsible locally but to the minister of war. The fact that the small brigades were required to supervise ballots for the militia, to check the passports of travelers, and to enforce order at local fairs and village festivals could provoke hostility among local communities. But there is also evidence that the men were accepted, well integrated, and often married into the community that they served. General inspections revealed a few men who were too old or infirm to carry out their tasks and that there was a sprinkling of total incompetents and drunks in the ranks. The guarantee of a pension in 1778 went some way towards removing the aged and physically unfit, and it seems fair to conclude that the institution improved as the eighteenth century wore on. Moreover, the justice provided by the *maréchaussée* through the *prévôtal* courts was generally quicker, and at least where beggars and the severity of punishments were concerned, it was arguably more balanced and moderate. Comments, both critical and favorable, were to be found in the *cahiers de doléances* that were drawn up at the beginning of the Revolution. There was some disquiet about *prévôtal* justice, but many of the *cahiers* argued for an expansion of the force to improve security (Emsley 1999, pp. 33–5).

The Revolution and Reorganization

A proposal for reform of the *maréchaussée* was on the table at the beginning of 1789, but it was rapidly overtaken by events. Over the next 2 years, there were debates about whether a police institution should be military; one of the principal opponents of a military police was Maximilien Robespierre. There were criticisms of the *prévôtal* courts, notably by the Comte de Mirabeau, and these were abolished in September 1790. The following February a law was passed reorganizing the force under the new name of the *Gendarmerie nationale*. This suggested a link with the past: the Gendarmerie had been an elite cavalry regiment in the royal household under the old regime and dated back to the fifteenth century. But the adjective "national" stressed its link to the nation and the new legal structure with its promise of equality before the law and the rights of men and citizens. The newly named force remained a military organization; its deployment in small barracks along the main roads was maintained together with its system of patrolling. But while the force was both military and national, appointments were devolved to the new provincial administrations – the *départements*. Moreover, the continuing power struggles of the Revolution and the demands of both civil and international war provided a decade of institutional disruption for the *Gendarmerie* in addition to major problems of law and order maintenance.

In the early years of the Revolution, and following the requests from the *cahiers* for a bigger and better police institution, the *Gendarmerie* was increased to 7,250 men; in April 1792 it was increased again to 8,700. But these were paper figures. Many *départements* failed to recruit to their quotas, and as the war required more and more soldiers to defend the frontiers, the state itself began to deplete the brigades by taking over half of the men back into the army. The reduction in numbers took no account of the country's internal difficulties. In addition to the policing tasks that had been undertaken by the old *maréchaussée*, the skeleton brigades of the *Gendarmerie* had to deal with counterrevolutionary activities, especially the west, with White terror gangs in the south, and also with bandit gangs swollen by deserters and, as successive governments made moves towards a comprehensive policy of conscription, by draft-dodgers. In addition, as with every institution in Revolutionary France, periodically men were purged as politically suspect. And, while gendarmes were supposed to be functionaries of the state and the nation, a gendarme serving in his region of origin or in his wife's native region faced problems of where his loyalty should lie. The gendarmes' military backgrounds probably ensured that they were less susceptible to such pressures than other police officers, but they were certainly not immune. Moreover, given the variety of languages and patois spoken in France at least until the end of the nineteenth century, it was always useful for a small *Gendarmerie* brigade (as had happened with the old *maré chaussée*) to have one or two local men, well versed in the local dialect.

But if there were problems for the *Gendarmerie* during the 1790s, the institution also ultimately showed itself to be invaluable in helping to reestablish order and imposing the state's concept of both law and the nation. The directory launched a fierce policy of military repression to suppress brigandage and disorder. The consulate continued this policy and took all the credit for its broad success. Once a greater degree of order had been established, it has been argued, the consulate and then the empire developed a security state in which the Revolution's democratic aspirations gave way to a tutelary administration and judicial apparatus that employed surveillance and regulatory control to maintain order (Brown 1997). The *Gendarmerie* played a key role in both the initial military repression and then in the new surveillance system, for which it also provided a small-scale coercive capability should the need arise.

As first consul, Citizen-General Bonaparte had his own ideas about the structure and deployment of the *Gendarmerie*. He favored fewer mounted gendarmes and more men on foot, particularly where the geography was unsuitable for horses. But he also knew how to choose capable subordinates, and in the case of the *Gendarmerie*, he chose tough, no-nonsense, plain-speaking army veterans. General Louis Wirion had experience of organizing the *Gendarmerie* in the 13 new *départements* of the northeast that had previously formed the Austrian Netherlands. Wirion was charged with reorganizing the force in the counterrevolutionary west. General Etienne Radet had commanded the 24th *Gendarmerie* Division that covered four of the most turbulent *départements* of the south. He was beginning to get a grip on the region when Bonaparte first met him on his return from the disastrous Egyptian campaign. As first consul Bonaparte summoned Radet to Paris to draw up plans for reorganizing the *Gendarmerie* as a whole.

Wirion stressed the image of the ideal gendarme as a courageous army veteran who was honest, moral, sober, and literate. He was to be a man that the local community looked up to and trusted; he was to know every part of his district intimately. Radet's ideas were formulated in an organizational order issued at the end of January 1801. This enlarged the total corps to 16,500 men. There were to be 2,500 brigades of six men deployed in the provincial *départements*; two-thirds of these were to be mounted. There were also to be four elite companies, two mounted two on foot, that were to be deployed to protect the first consul (and later the emperor) and his palaces. The officers were all to receive their commissions from the head of state; the other ranks were to be vetted by a *départemental* committee composed of the prefect and two officers from the force. Finally there was to be a headquarter staff under a general appointed as the *Premier Inspecteur Gé néral de la Gendarmerie*. Radet, by his own account, turned down the latter appointment and it went to General (later Marshal) Moncey. Wirion went off to establish the *Gendarmerie* in newly conquered Piedmont; Radet undertook similar tasks across the new French Empire. His toughness and aggression offended many civilian officials; unsurprisingly, when Napoleon needed a man to arrest the Pope in 1809, it was Radet who volunteered.

Moncey was a rather more refined individual than Wirion and Radet, but he could be prickly. Throughout his tenure as inspector general, he remained determined to maintain the *Gendarmerie*'*s* separation from other police institutions within Napoleon's France. But this, according to Joseph Fouché who served as minister of police for much of the period of the empire, suited Napoleon's intention to pursue a policy of divide and rule when it came to the police. There was a separate military police under General Duroc in his capacity as grand master of the palace. The Prefect of Police had responsibility for policing in Paris and reported to the minister of police, while civilian police, principally the *commissaires*, in the provinces also linked with the minister as well as with his fellows in the interior and justice. Moncey kept his command at a distance from the Ministry of Police; at the same time, he and Fouché confronted each other over a range of issues setting the tone for the "war of the polices" that continued in France for the next two centuries (Lignereux 2003).

The force that developed during the Revolutionary and Napoleonic period set the tone for the future in other ways. It remained proud of its standing as a military body. While it was sometimes useful to send individuals or small groups in civilian clothes to reconnoiter a town or village before a move to search out and apprehend deserters, it was formally stated in the various decrees regulating the institution that gendarmes always acted in uniform. Indeed, this became a matter of pride with many, and in 1857, Captain Frédéric de Bouyn requested an audience with Napoleon III concerned that the corps was being required to act in secretive ways and to investigate people's politics. Nothing in the force's behavior, he explained, "should excite suspicion, nothing should imply that its duties are mysterious and shadowy" (quoted in Emsley 1999, p. 127). The captain's open expression of his concerns stalled his career; others, however, who were less reticent in making political efforts on behalf of the government, tended to prosper.

The legacy of the Revolution and the Napoleonic adventure made it difficult for a state agency like the *Gendarmerie* to stand aloof from the politics of nineteenth-century France. Yet many gendarmes seem to have tried to be as honest, moral, and sober as Wirion, and the regulations insisted. Many individual brigades worked hard on behalf of the communities to which they were posted; they tracked down offenders, provided assistance in time of natural disaster, and so forth. But, in addition to the occasional demands for overt political action, they also had one other duty that could easily bring them into conflict with those communities; it fell to the gendarmes to ensure that when the demand was made, conscripts were collected and brought to the appropriate muster points. It also fell to the gendarmes to go in pursuit of refractory conscripts and deserters. Increasingly the *Gendarmerie*, together with other parts of Napoleon's bureaucratic state, wore down the opposition, but the enormous demand for men in 1813 sparked a new, furious wave of violence and disorder invariably directed at the gendarmes.

Continuity, Change, and Recurrent Issues in the Nineteenth Century

The French Revolution had proclaimed the Rights of Man and, of course, equality along with liberty and fraternity. Gendarmes were supposed to ensure a citizen's rights and to enforce the law equally among citizens, but the political turbulence that accompanied the various stages of the Revolution and the memory of Napoleon's empire that followed its fall left a legacy in which the different varieties of police in France were often seen as tied to, or acting at, the behest of those in power. At different stages of the Revolution, a few gendarmes were accused of disaffection and purged. When the Bourbons were restored for the second time, both Moncey and Radet faced courts-martial. *Départemental* committees, similar to those of the Revolution, were revived by the Bourbons to vet new recruits, and with varying degrees of determination, they also purged those who had shown, or continued to show, zeal for "the usurper." Yet in the various political upheavals of the restoration and the July Monarchy, while the *départemental* companies sent regular reports to Paris about the economy and political attitudes, most of the brigades appear to have aimed to fulfill their basic duties of maintaining order and protecting their local communities.

The elite *Gendarmerie* companies in the capital, whose functions mainly involved public order and ceremonies, earned the hostility of the Parisians for aiding Charles X's attempted coup in the July days of 1830. A *Gendarmerie mobile* was established in the wake of the 1830 Revolution, revived in 1848 and revived again under Napoleon III. But most of the small provincial brigades, whose daily patrols focussed on supervising and maintaining local tranquility, enforcing regulations and dealing with crime, sat tight during the revolutions and watched events.

As president of the Second Republic, Louis Napoleon feted the institution, but most provincial brigades saw discretion as the better part of valor and avoided confrontations with the peasant columns and the democratic-socialists who resisted his coup in 1851. Nevertheless, when the president became the Emperor Napoleon III, he maintained a close relationship with the *Gendarmerie* until the mid-1850s when, after the Imperial Decree reorganizing the force and outlining the essence of its service as a "continual and repressive surveillance" particularly in the countryside and on main roads, his interests moved more to reforming the police. The close relationship and the concept of "repressive surveillance" led to the concerns expressed by de Bouyn that the brigades were expected to comment on people's politics and to take an active role in elections.

Economic, political, and social life changed significantly in the French provinces during the nineteenth century, and the *Gendarmerie* was required to keep up both with the changes and with traditional sensibilities. The building of railways and movements of workers from different regions and countries led to increases in the workloads of some brigades and, occasionally, a decrease for others. Gendarmes provided coercive support for bailiffs in the eviction of workers from company housing or for local officials wanting to deal with unruly bars. At other times, however, they aided members of the working class and the poor. And this did not just mean acting in time of natural disaster or pursuing those responsible for criminal acts against poor people; a few gendarmes were known to give to charity the occasional monetary reward they had received, while others organized subscriptions to assist beggars. But an officious, self-important brigade commander who sought rigorously to enforce unpopular legislation especially, for example, on poaching or cabarets could provoke serious local hostility. The brigade commander at Bédarieux (Hérault) was one such, and his small barracks suffered the most violent and sustained attack following Louis Napoleon's coup leaving the brigadier, two gendarmes, and a gendarme's wife dead. Another of the Bédarieux brigade, however, was saved when a café proprietor intervened on his behalf; the gendarme had earlier broken regulations so as to permit the proprietor to attend a fellow radical's funeral. An informative statistical analysis of the attacks on gendarmes between 1800 and 1859 suggests a significant decline across the period and a corresponding growth in the legitimacy of the institution among the people in rural communities (Lignereux 2008).

The propertied classes generally showed approval of the force. Official literature painted the gendarme as a man cast in a heroic mold, honest, modest, and courageous. Descriptions of the small barracks sometimes suggested almost a monastic community, although any visit to a small barracks probably would soon have dispelled such a notion. The men, their wives and children, as well as their horses were all jammed in close proximity to each other. Moreover, while in some respects, his legitimacy grew among the plebeian classes; in various forms of popular culture, the gendarme was commonly portrayed as officious, stupid, and a figure of fun. In popular puppet plays, he might appear as an honest upholder of the law, but when up against shrewd working-class heroes like the Lyon weaver Guignol, he always came off second best. Perhaps the best known of the comic manifestations, however, are Gendarme Pandore and Brigadier Dussutour. Pandore, who first appeared in Gustav Nadaud's popular song written (and promptly banned) at the beginning of the Second Empire, feared God, was virile, but not very bright, and never questioned his superiors: "Brigadier, vous-avez raison!" Dussutour was the eponymous "Good Gendarme" of Léon Bloy's short story who confronts a division of invading

Prussians in 1871 with "I demand to see your papers" (see various articles in Luc 2003).

The close links that the *Gendarmerie* had enjoyed with Napoleon III meant that at the beginning of the Third Republic, the institution was regarded with suspicion by Republicans and the Left. There was general agreement that it should not have a political role and that it should stick to pursuing criminal offenders, maintaining order and enforcing various regulations. The *Gendarmerie mobile* was disbanded in 1885, and some advocated replacing the entire force with a *départemental* police while others urged that it be demilitarized. Such major changes were largely prevented because of the inability of the force's detractors to secure a majority among legislators for any one of their proposals during the early years of the new republic. At the same time, the gendarmes themselves could scarcely be faulted for the ways in which they accepted the role of defenders of the republic and enforced measures, such as the anticlerical laws, that critics of the regime opposed.

The Spread of the Model

In May 1806 Napoleon told his brother Joseph, who he had recently installed as King of Naples, that the Gendarmerie provided "the most efficient way to maintain the tranquillity of a country . . . it provides a surveillance half civil, half military spread across the whole territory together with the most precise information" (quoted in Emsley 1999, p. 56). As a part of Napoleon's empire, as a satellite state or as an ally, many regions of Europe had experience of the Gendarmerie model in the first decade or so of the nineteenth century. Even after Napoleon's fall, many of the rulers of these territories tended to agree with the emperor. Radet and Wirion had set up Gendarmeries in various parts of Italy. Under the restored Savoyard Monarchy, the restructured Piedmontese Gendarmerie, called the *Carabinieri Reale*, was to play a significant military role in the unification of Italy absorbing some of the other Italian Gendarmeries as it did so. Various German states had created Gendarmeries during the Napoleonic ascendancy and continued to deploy them thereafter. In Wùrttemberg, however, the German name of *Landjäger* was preferred, and in the Netherlands the former French name of *Marechaussee* was employed.

Even Napoleon's most consistent enemies created Gendarmeries as and when they thought it prudent. The Irish Constabulary emerged out of a series of initiatives taken by British administrators in Ireland and formalized by an Act of Parliament in 1822. The word "constabulary" was employed at least in part because of concerns about the French word Gendarmerie, but unlike its counterparts, the force was never linked to the Ministry of War (the War Office in British parlance). The Irish (from 1867 Royal Irish) Constabulary was seen by many administrators in the British Empire as a model for the kind of police that they required outside of the major centers of population in the colonies, and they often requested Royal Irish Constabulary officers and men to help create a new force. In Quebec what was originally the North West Mounted Police, and subsequently the Royal Canadian Mounted Police, is still known as *La Gendarmerie Royale du Canada*.

The Gendarmerie of Imperial Russia followed the track of its French counterpart beginning as a military unit established to police the Russian army of occupation in France following Napoleon's overthrow. It then acquired authority over the civilian population of Russia, but throughout the nineteenth century, it maintained a close link with the third section of the Tsar's Private Imperial Chancellery, the department responsible for political policing. Political turbulence in Spain during the 1830s led to the creation of the *Guardia Civil* in 1843, but in spite of the name, the force was militarized, under the direction of the minister of war and housed in small barracks virtually from its creation. In the Habsburg Empire, it was the Galician revolt of 1846 and the subsequent revolution of 1848 that convinced the government of the need for a Gendarmerie. States, like Greece and Romania, that emerged from the Ottoman withdrawal from Europe also created Gendarmeries as the best means of cementing the authority of the new rulers across their territories.

In addition to dealing with various forms of criminal activity, most gendarmes across continental Europe maintained a surveillance of the communities in which they served and collected information for the central government about the local economy, political attitudes, and any other apparently significant developments or events. They provided a first line of defense against popular disorder, and perhaps most significantly, they played a role in the evolving relationship between the nineteenth-century bureaucratic state and its people. Gendarmes in their barracks flying the national or imperial flag, celebrating national or imperial anniversaries and festivities, were the living manifestation of the state or empire for rural communities who rarely looked far beyond their immediate district. Across the continent they commonly assumed the same kind of attributes as those in France; they were portrayed in the official literature and their own corps literature as heroic, honest defenders of the law. The *Carabinieri* recruit, for example, was told how he was entering a family of men who depended on each other and who dedicated their lives to the good of others (Grossadi 1879). Gendarmes protected rural communities from bandits, brigands, and wild beasts; they pursued offenders and brought them to justice; they were the first to help communities in times of natural disaster. But there was an obverse side to these roles. If the state's gendarmes assisted the population, the population had to recognize its obligations to the state, and the gendarmes were also there to ensure that taxes were paid and that young men turned up as and when required for military service. Moreover, in the event of any form of labor or political unrest, the gendarmes were usually the first force available for the authorities to deploy against any demonstrators.

The First Half of the Twentieth Century

Although the French *Gendarmerie* carried out the orders of the Third Republic (1870–1940) loyally, in the 20 years or so before the First World War, it was often criticized as being inadequate for its tasks. The brigade structure meant that small disorders could be handled, but there was no *Gendarmerie mobile* available for dealing with significant industrial unrest and political demonstrations. There were a succession of discussions and proposals for reform put before the legislature, but they all foundered. The brigades, many of whom exchanged their horses for bicycles, were too isolated to deal with itinerant offenders or mechanized criminals such as the anarchist *bande à Bonnot*. The institution also suffered a major blow to its reputation during the First World War: for the first time since its creation it provided no front-line combat unit. Gendarmes were, however, deployed immediately behind the front line to enforce discipline and apprehend deserters. The *poilus* (soldiers) joked cynically about gendarmes, who were often noted for their heavy drinking, suppressing drunkenness among the front-line soldiers. They also scoffed that the front line ended where you met the first gendarme.

During the interwar period, the *Gendarmerie* like the rest of the French police, sought to modernize its fight against crime by developing its use of motor vehicles and telecommunications. Young officers appear also to have been keen to develop the corps role in criminal investigation (Haberbusch 2012). In addition to their traditional role in policing the countryside, and again like the civilian police, the *Gendarmerie* became increasingly preoccupied with the threat of communist subversion. A *Gendarmerie mobile* was reestablished in 1921 to deal with unrest, particularly in labor disputes. The right-leaning ministry that took the step had no desire to see the victorious army risking its reputation in such situations. But just over a decade later, the parliamentary committee appointed to investigate the Paris riots of 1934 concluded that at least on that occasion the *Gendarmerie mobile* was poorly led, inappropriately equipped, and lacking in useful intelligence. Yet in spite of the criticisms and the humiliation of its wartime involvement, the *Gendarmerie* remained loyal to the republic.

At the end of the First World War, the fear of communism and Revolution led to the creation of militarized police in both Germany and the Netherlands. During the summer of 1919 first in Prussia and then in other *Länder*, the nervous authorities set up “Security Police,” *Sicherheitspolizei* (or *Sipo*). These heavily armed police were viewed with suspicion by many in the SPD and by trade unionists. They were also suspected by the victorious allies as a means to circumvent the restrictions on the size of the German army. The *Sipo* did not outlive 1920; the Weimar Republic reverted to the more traditional policing structures, and the policing of rural areas by men designated as gendarmes continued through the Nazi period even as different police institutions were unified into the German Order Police (*Ordnungspolizei*). In the Netherlands rioting in Amsterdam in June 1919 led to the formation of the *Korps Politietroepen* which was incorporated into the army in 1922 and functioned alongside the *Marechaussee* and the *Rijksveldwacht* (state police) until the Second World War.

Elsewhere during the interwar years, other Gendarmeries found their loyalties tested by the turbulent politics of the period. In Italy the *Carabinieri*’*s* boasted first loyalty to the king and to Italy enabled it to ride out some of the problems following the Fascist takeover. However, it is at least arguable that a majority in the *Carabinieri* were more in sympathy with the radical Right than with the political Left and the corruption and clientelism of the Liberals. In Spain the *Guardia Civil* was torn over the formation of the Second Republic. Many in the guard saw their duty as maintaining order in Spain and judged a government on its ability to ensure that they were able to carry out this duty. When the Civil War broke out, a significant percentage of the guards went over to Franco, seeing him as a better hope for order in their beloved homeland – and they went over to Franco even if this meant shooting those among their officers who remained loyal to the republic.

The old rivalries between the Gendarmeries, primarily responsible to ministries of war, and the civilian police, responsible to the ministries of justice and the interior, also continued. Events in the Netherlands’ municipality of Oss during the 1930s provide a vivid example. The *Koninklijke Marechaussee* fought a successful campaign against serious crime in the district which prompted decorations from the queen for the commander and one of his staff but jealousy among the *Rijksveldwacht* and the municipal police. Subsequently, following a largely fruitless investigation of local municipal and religious authorities and the arrest of a factory owner and an insurance broker, the minister of justice prohibited further investigations and arranged to have the full brigade of the *Marechaussee* transferred elsewhere. Eventually a parliamentary enquiry was called to settle the matter which had turned into bickering between the ministers of justice and the interior as well as the minister of war and the *Marechaussee*. The civilian ministries’ dreams of creating a single, unified policing structure had to be shelved, and when reform came, it was imposed by the German occupiers who set out to amalgamate the *Rijksveldwacht* with the *Koninklijke Marechaussee* in March 1941. The aim was to create an institution modeled on the SS and loyal to the new German authorities.

The Second World War brought additional pressures. In France, determined to avoid the criticisms of its role in the previous war, the *Gendarmerie* deployed a combat unit. The combat unit drew on the younger, more energetic, and fitter gendarmes, and while experienced heads are important to policing, the drain on the provincial brigades appears to have had a negative effect. The recruits who came forward to fill the gaps during the German occupation were often well qualified, but many were using the institution as a way of avoiding *Service de travail obligatoire* in Germany and had no serious commitment to the job. In Vichy the *Gendarmerie* was purged of those elements that the government considered undesirable and a threat to France such as Jews and Freemasons. In both Vichy and the occupied zone, gendarmes were involved in enforcing the racial and political policies of the conqueror. In the early years of the occupation and Vichy, very few joined the resistance, but the Germans appear to have had suspicions about the *Gendarmerie* and occasionally picked on both officers and men, and by early

1944, many gendarmes appear to have been paying lip service to orders emanating from the occupying power. In France, Belgium, and the Netherlands, the different Gendarmeries, like other police institutions, faced issues of where their duty lay. The problem was especially acute in Belgium and the Netherlands where the vestiges of prewar governments had gone into exile in Britain. At the close of the war in all three countries, these became questions of the legitimacy for the different forces as members of the resistance and others sought revenge, and scapegoats and gendarmes were challenged and investigated as to whether they had put the emphasis on professional and patriotic duty (Campion 2011). The same problem arose in Greece where, in the wake of the German defeat, the country descended into civil war.

In Italy, loyalty to the king and to the concept of *Italia* enabled the *Carabinieri* to shift with wartime politics. When the war began, they loyally served the Fascist state; when Mussolini was overthrown, they were able to collaborate with the allies fighting their way up the peninsula. *Carabinieri* worked alongside allied military police in attempts to suppress the rampant black market and brigandage that appeared, especially in the south. In those areas that were occupied by the Germans, they collaborated but often in unique ways. The most notable and, as far as the *Carabinieri* itself is concerned, the most heroic example is the self-sacrifice of *vicebrigadiere* Salvo d'Acquisto who, in September 1943, voluntarily went before a German firing squad in order to save 22 innocent hostages. His last words, allegedly, were *Viva Italia*. But the allies were conscious of the *Carabinieri's* involvement with the Fascist state, and a mission, led by a senior English police officer and staffed by men from other British forces, set out to take the Fascist element out of all Italian police institutions. Similar British police missions were deployed in Austria and Germany; in the former the relaxed, some might say idle, attitude of the English commander ensured that Austrian Gendarmerie maintained the traditional characteristics of such a body. Paradoxically in the west of Germany, while the victorious allies were determined to denazify and demilitarize the police, the specifically military police of the Third Reich, the *Feldgendarmerie*, was the last military unit to be disbanded as its experience and discipline was considered too important in the struggle against the postwar crime wave. In postwar Greece, torn by civil war, another British police mission led by a former head of the Royal Ulster Constabulary (which had replaced, but closely resembled, the Gendarmerie-style RIC on the partition of Ireland) was keen to see the Greek Gendarmerie trained first and foremost as ordinary police and subject to civilian authority. Even so, the mission commander also recognized the virtue in a paramilitary Gendarmerie for the warring countryside.

To the Twenty-First Century

The rivalry and occasional friction between civilian police and Gendarmeries continued in the aftermath of the Second World War. Sometimes it sprang from the Gendarmeries' proud military tradition. In France, for example, during the last two decades of the century, morale in both the police and the *Gendarmerie* was periodically undermined by a variety of issues some of which resulted from shifting pressures in the job, such as the emergence of international terrorism and who should take precedence in handling it. There was also cultural change; more working wives and changes in the civilian working world challenged some of the traditional understanding of the gendarme's military commitment and subservience to old-style military discipline. The government could aggravate such concerns by suggesting a redefinition of ranks which brought a degree of unanimity between *Gendarmerie* and police but which also highlighted some better emoluments within the *Gendarmerie* as well as discrepancies between the responsibilities of different ranks in the different institutions. President Mitterand's creation of the *Groupe* *d'intervention de la Gendarmerie nationale* (*GIGN*) for his personal protection in the Elysée Palace with its additional responsibilities for anti-terrorism aggravated police-*Gendarmerie* relations, while the use of *GIGN* officers for investigating the president's opponents caused outrage beyond the ranks of the police and tarnished the *Gendarmerie's* image.

Elsewhere political involvement by senior gendarmes did their institutions little credit. Senior figures in the *Carabinieri* were suspected of taking their anti-communism further than was proper in a democracy. It was the refusal of members of the corps to leave their barracks that stifled an attempted coup led by a *Carabinieri* general in 1964. The Greek Gendarmerie was closely tied with the Colonels' junta that seized power in 1967. In the aftermath of the Colonels' fall, the Gendarmerie was united with the town police, yet the new police, while more closely tied to the civilian state, remains a military organization. Following the death of Franco, the *Guardia Civil* remained suspect in the eyes of many on the Left and the attempted coup of February 1981 in which a lieutenant colonel of the *Guardia* and his men seized the lower chamber of the Cortes confirmed such suspicions. The *Guardia Civil* survived the failed coup but came increasingly to resemble a civilian police, though it was not until 2009 that the traditional bicorn hat was replaced (except for ceremonial duties) with a more conventional baseball cap.

As the European Gendarmeries became more like civilian police institutions with, for example, the Belgian Gendarmerie's amalgamation with the country's civil police in 2000 and the bringing together of the *Gendarmerie nationale* and the *Police nationale* under the French Ministry of the Interior in 2009, so a new opportunity opened up for these corps in the wider world. It was argued that soldiers and marines lacked the necessary skills for establishing and maintaining basic law and order in failed states or states emerging from civil war or international conflict (Perito 2004). Police missions to such states as Bosnia, Kosovo, Iraq, and Afghanistan were commonly spearheaded by gendarmes. Even British Police missions under the auspices of the UK's International Police Assistance Board contained a disproportionate number of men from the old Royal Ulster Constabulary or its more civilian successor force (the Police Service Northern Ireland, PSNI) – but this successor force was itself well versed in crowd control and soothing the passions of rival communities (Sinclair 2012).

Globalization, concerns about organized crime on an international level, and the increase in demands for police missions like those to the new states emerging from former Yugoslavia combined to foster the creation of the European Gendarmerie (EGF) in 2006. The EGF, formalized by the Treaty of Velsen in October 2007, brought together gendarmes from corps of five of the EU's member states: France, Italy, the Netherlands, Portugal, and Spain. When Romania joined the EU, its Gendarmerie also became a member, while those of Poland and Lithuania were designated as partners. The intention was to have a force of up to 800 gendarmes available for deployment within 30 days of a request for assistance. The EGF was involved in NATO police missions, notably in Afghanistan, and it provided a small force to advise on security in Haiti following the earthquake of 2010. But the new institution raised fears, particularly about accountability, and there were even suspicions that EGF officers had been deployed to Greece during the disorders engendered by the crisis over Greek debt (see, e.g., www.golenxiv.co.uk/2011/10/foreign-riot-police-now-operating-in-greece/).

Related Entries

- ▶ British Police
- ▶ Comparing Police Systems Across the World
- ▶ Dutch Colonial Police
- ▶ French Colonial Police
- ▶ German Police Until 1918
- ▶ Policing of Peacekeeping

Recommended Reading and References

Anderson M (2011) In thrall to political change: police and gendarmerie in France. Oxford University Press, Oxford

Blaney G Jr (2013) The three-cornered hat and the tri-colored flag. The civil guard and the second republic, 1931–1936. Sussex Academic Press, Toronto

Brown HG (1997) From organic society to security state: the war on brigandage in France, 1797–1802. J Mod Hist 69:661–695

Campion J (2011) Les gendarmes belges, français et néerlandais à la suite de la Seconde Guerre mondiale. André Versaille Éditeur, Bruxelles

Emsley C (1999) Gendarmes and the state in Nineteenth-Century Europe. Oxford University Press, Oxford

Garrido DL (1982) La Guardia Civil y los origenes del Estado centralista. Editora Critica, Barcelona

Grossadi GC (1879) Galateo dei Carabinieri, 3rd edn. Candeletti, Turin

Haberbusch B (2012) Les gendarmes face au crime. Geste édition, La Creste

Lignereux A (2003) Gendarmes et Policiers dans la France de Napoléon: Le duel Moncey-Fouché. Service historique de la Gendarmerie nationale, Maisons Alfort

Lignereux A (2008) La France rébellionaire: Les résistances à la gendarmerie (1800–1859). Presses Universitaires de Rennes, Rennes

Luc J-N (ed) (2002) Gendarmerie, État et Société au XIXe siècle. Presses de l'Université Paris-Sorbonne, Paris

Luc J-N (ed) (2003) Figures de gendarmes, special edition of Sociétés et Représentations, Credhess, vol 16

Luc J-N (ed) (2005) Histoire de la Maréchaussée et de la Gendarmerie: Guide de recherche. Service historique de la Gendarmerie nationale, Maisons Alfort

Luc J-N (ed) (2010) Soldats de la Loi. La gendarmerie au XX siècle. Presses de l'Université Paris-Sorbonne, Paris

Malcolm E (2006) The Irish policeman, 1822–1922: a life. Four Courts Press, Dublin

Perito RM (2004) Where is the lone ranger when we need him? America's search for a postconflict stability force. United States Institute for Peace, Washington, DC

Sinclair G (2012) Exporting the UK Police "Brand": the RUC-PSNI and the international policing agenda. Policing: J Policy Pract 6(1):55–66

Gender and Race Differences

► Self-Reported Offending: Reliability and Validity

Gender-Based Violence

► Surveys on Violence Against Women

Gendered Theory and Gendered Practice

Marjorie S. Zatz and Heather R. Gough
School of Social Transformation, Arizona State University, Tempe, AZ, USA

Overview

Theoretical and empirical research on the relationships among gender, crime, and delinquency has grown in scope and sophistication in recent decades. Less apparent, though, is whether and to what extent these advances are reflected in current criminal and juvenile justice practices and policies. This entry briefly reviews major feminist and gendered theories of crime and delinquency and then turns to key laws, policies, and practices to assess whether these theoretical advances translate into improved policy and practice. We conclude with suggestions for how feminist scholarship can help shape gender-specific and culturally appropriate criminal and juvenile justice programming and training.

Introduction

Tough-on-crime and zero-tolerance laws and policies enacted in the United States in recent decades dramatically increased the number of women and girls processed through the criminal and juvenile justice systems. During this same era, another set of laws and policies sought to reduce violence against women and girls and to change the ways in which police, prosecutors, and judges respond to rape and intimate partner violence, in particular. To what extent do these legal reforms, as well as the practices of criminal and juvenile justice institutions, reflect theoretical advances and new knowledge about gender and crime? To address this question we, first, briefly summarize major feminist and gendered theories and perspectives. Second, we assess the extent to which recent criminal and juvenile justice policies and practices, as well as major legislative initiatives, reflect feminist perspectives. Third, we examine whether these reforms have successfully met feminist objectives or have been eroded and undermined and, if so, in what ways and to what effect. The entry concludes with suggestions for ways in which relevant theory and research might better inform criminal and juvenile justice policies and practices.

Feminist and Gendered Theories of Crime and Delinquency

This section briefly summarizes the principal theoretical perspectives on gender, crime and delinquency before turning to an assessment of their utility and visibility in criminal and juvenile justice policies and programs, as well as relevant legislation. Readers are referred to Chesney-Lind and Morash (2011), Miller and Mullins (2009), and Daly (1998) for more detailed reviews of cutting-edge theoretical work on gender, crime and delinquency, and compilations edited by Chesney-Lind and Pasko (2004), Heimer and Kruttschnitt (2006), and Zahn (2009) for examples of empirical research that has tested and refined these theories. It is also important to note that while much of this literature focuses on women and girls, a number of studies examine masculinities and crime to better understand how and in what contexts men engage in violence, both against women and against other men (e.g., Dobash and Dobash 1992; Messerschmidt 1993).

Most recent criminological research that explicitly examines gender can be categorized as studies of the "gendered ratio of crime," "gendered pathways," "gendered crime," and "gendered lives" (see Daly 1998:94–95). Studies of the *gendered ratio* of crime seek to explain the gender gap in self-reported and official rates of crime and delinquency. Research from this perspective often begins with what is known about male crime and delinquency and asks why women and girls engage in less crime. While some scholars examine individual level influences, feminist criminologists examining the gender gap tend to focus more on gendered social organizations and structural conditions, as these shape the types of offenses and contexts in which girls/women and boys/men commit crimes (e.g., Steffensmeier and Schwartz 2004).

Gendered pathways research examines the trajectories by which women and girls, and men and boys, move in and out of lawbreaking behaviors. Women's and girls' pathways often include histories of victimization and of strategies for coping with that victimization that may themselves be criminalized (e.g., substance abuse, running away, fighting back), further entangling them with the criminal and juvenile justice systems (Chesney-Lind and Pasko 2004; Daly 1992). Some research refers to the "blurred boundaries" between victimization and offending, but even when the boundaries are sharp, the focus here is on the most common paths to criminal and juvenile court and how those paths are structured by gender, race, class, and other sources of inequality.

Studies of *gendered crime* focus on the social context and situational qualities of offending by girls and women and by boys and men. Victimization may be a situational factor resulting in crime, as seen, for instance, when women kill their abusers. In other situations, though, crime and delinquency may be resources for gaining status, desired commodities, revenge, or simply a means of having fun. Much of this research emerges from sociological studies of "doing gender," reminding us that gender is performed in multiple ways, depending in large part upon the resources available, and those resources are themselves structured by gender, race, class, gender orientation, and age, among other dimensions (see, e.g., Miller 1998; Messerschmidt 1993).

Finally, research on *gendered lives* asks how gender structures the kinds of actions and identities women and girls, and men and boys, employ to survive. Examples are Maher's (1997) detailed analysis of sex work by women drug users and research on the differing types and effects of peer relationships among adolescent boys and girls (e.g., Giordano et al. 2003).

Regardless of which of these approaches is paramount, the most comprehensive theoretical and empirical studies of crime and delinquency attend to the multiple and intersecting ways in which gender, racial, economic, and other inequalities structure individuals' lives and choices in patterned ways. Borrowing from Miller and Mullins, we must study gender and gender inequality "at the macro level (overarching structural patterns of gender inequality and their effects on crime and delinquency), at the meso level (gender inequality within the context of social institutions such as family, school, and neighborhood) and at the micro level (interpersonal relationships within and across gender)" (2009:34).

Feminist Influences on Law, Policy, and Practice

The civil rights and feminist movements have led to major social and legal changes in the United States and, to varying extents, globally. They have provided social recognition for a variety of problems that had previously been defined as simply private harms. Yet empirical research and theorizing about gender, race, class, and other intersecting sources of inequality are not necessarily reflected in contemporary criminal and juvenile statutes, policies, and practices. Policies and programs that ignore how structural and institutional contexts shape and constrain individual lives are likely to be ineffective. As Miller and Mullins (2009) remind us, "taken-for-granted ideologies about gender" result in scholarship "that fundamentally misrepresents and misunderstands the nature of girls' delinquency and produces policies that are misguided at best, and quite harmful at their worst" (2009:31–32).

Legal Reforms

Since the 1970s, a number of significant legislative and policy initiatives have emerged in the areas of rape reform, intimate partner violence, and deinstitutionalization, among other issues. Many of these legal changes resulted from social protests and lobbying by feminist and civil rights organizations and reflect feminist thinking about crime and delinquency. Yet in most cases, the new laws were not as far-reaching as their advocates had hoped, and many were eroded further during the antifeminist backlash of the late 1980s and 1990s.

For example, the rape reforms that swept the United States in the 1970s were seen as a major victory at the time. They provided for rape shield laws, repealed requirements for corroboration, redefined resistance, and in some cases created the criminal statutes necessary for rape prosecutions, including the then novel concept of a crime of spousal rape. Yet as Caringella (2009) demonstrates, they have been largely educational and symbolic in value, with comprehensive rape reform still an unmet goal.

Similarly, public attention to intimate partner violence as a major social problem led in the 1980s to pro-arrest domestic violence legislation and policies in states across the USA and to passage of the Violence Against Women Act in 1994 (and its reauthorizations in 2000 and 2005), the Victims of Trafficking and Violence Protection Act of 2000 (and its reauthorizations in 2003, 2005 and 2008), and the Tribal Law and Order Act of 2010. These laws and policies succeeded in changing public understandings of rape and domestic violence, in particular, and brought these crimes out of the shadows. Yet they stopped short of the larger structural changes envisioned by their advocates and, in some cases, contributed to new problems. Most notably, as of 2000, nearly 60 % of states had passed either mandatory or arrest-preferred legislation, with the unanticipated and ironic consequence of increasing domestic violence arrests and prosecutions of women who were defending themselves against violent partners (Hirschel et al. 2007:265). Public attention to violence against women has also affected perceptions of women who ultimately kill their abusers. The "battered woman syndrome" legal defense helped judges and juries understand that in certain circumstances, killing one's abuser may be an understandable and even reasonable response to long-term abuse, yet women who did not fit stereotypic images of how battered women should behave did not receive the same level of concern and legal support. This was particularly problematic for women of color. More generally, the new laws and policies were a step forward in many ways, yet did not adequately address the experiences of racially and economically marginalized women, immigrants, lesbians, and transgendered persons who could not assume the same support from police and prosecutors as middle-class white women experienced.

Girls, as well as women, were impacted by these early reforms and by their often unanticipated consequences. The Juvenile Justice and Delinquency Prevention Act of 1974 called for deinstitutionalization of status offenders, the vast majority of whom were girls. Prior to that time, girls were incarcerated for lengthy terms, often longer than boys convicted of felonies, for minor acts such as truancy, running away, and other age-specific or "status" offenses that would not be crimes if the girls were over the age of majority. The intent of this legislation was negated, though, by zero-tolerance school policies, arrests of teenage girls for assault rather than incorrigibility when fighting with their parents, and other new laws and policies that expanded the pool of delinquent girls. These mechanisms, then, have come to be known as net-widening, relabeling, and upcriming (Chesney-Lind and Irwin 2008).

Policies and Practices

The greater willingness by police, schools, and other authorities to punish girls' misbehaviors as crimes, in combination with mandatory incarceration for a variety of drug and other offenses, has resulted in soaring arrest and incarceration rates for girls and women in recent years. However, while the percentage increases in these rates have been quite large, the total numbers of incarcerated females remain small compared to males. As a consequence of the vastly greater number of males in the system and the lack of knowledge about how best to meet the needs of female offenders, studies from the 1970s through the mid-1990s reported limited funds were available for at-risk and delinquent girls (figures range from 3 % to 8 %). In response, the 1992 reauthorization of the Juvenile Justice and Delinquency Prevention Act required that girls have adequate access to services, and federal agencies and research institutes began soliciting proposals for gender-responsive programs and staff training.

A review of the literature on programming for girls and women suggests three primary issues confounding efforts to better serve female offenders. First, as noted above, is the lack of adequate resources dedicated to the specific needs of women and girls. Even where programs have been developed, the funding is too often short term, preventing program sustainability and growth. Second, a number of studies have found that correctional and probation staff prefer working with males, characterizing women and girls as petty, whiney, and manipulative (see, e.g., Martin and Jurik 2007; Gaarder et al. 2004). It appears that the complexities of how and why women and girls become involved in the justice system, and the need to address a broad range of psychological, interpersonal, economic, and physical safety issues for rehabilitation and reentry, can overwhelm staff, especially in a context of inadequate resources. As a result, some probation and correctional officers come to see women and girls as simply overly emotional or needy and distance themselves from the individuals in their care. The third critical factor inhibiting better service provision is the shortage of gender-specific, culturally appropriate, and theoretically informed programming (Foley 2008). Effective programming must recognize and respond to the multiple and intersecting forms of inequality experienced by many of the girls and women in the system. These inequalities operate at structural, institutional, and interpersonal levels, functioning both as pathways into the justice system and as barriers to successfully stabilizing outside of the system.

In an effort to further our understanding of girls' offending and identify effective, evidence-based strategies for preventing and reducing girls' involvement in crime and delinquency, the Office of Juvenile Justice and Delinquency Prevention (OJJDP) convened and funded the Girls Study Group in 2004. Major outcomes of the Girls Study Group include publication of The Delinquent Girl (Zahn 2009), which is a comprehensive review of the state of the field; development of a curriculum to assist staff in addressing girls' unique experiences and recognizing how gender, race, culture, economic disadvantage, and other factors shape those experiences; and a review of assessment tools to ascertain their suitability for girls (see http://www.ojjdp.gov/programs/girlsdelinquency.html). OJJDP also compiled a Model Programs Guide, which lists programs from prevention through sanctioning to reentry, and they fund a number of gender-specific programs including substance abuse prevention inside and outside of correctional settings, as well as Girl Scouting in Detention Centers, which targets adjudicated or court-referred delinquent girls or girls who are wards of the state.

A second major resource for staff training and program development for delinquent girls was developed by the National Center for Crime and Delinquency (NCCD). Launched in 2008, the NCCD's Center for Girls and Young Women specifically addresses "advocacy, research, assessment services, staff training and evaluation to address juvenile justice and child welfare systems that are designed for boys and ill-equipped to meet the gender-specific needs of girls and young women" (http://justiceforallgirls.org).

The evolution of these programs demonstrates that policy, programming, and training are increasingly influenced by theoretical and empirical research elaborating the ways in which gender matters. Simultaneously, evaluations of these programs and interviews with incarcerated girls and women have advanced theorizing, especially with regard to theories centered on gendered pathways into delinquency and crime. Yet while feminist scholars seek to go beyond essentialized or simplistic models of how gender intersects with race, class, sexuality, gender identity, and other factors to shape the lived experiences of individual girls, the programs in place tend to be far less nuanced.

At the adult level, the National Institute of Corrections (NIC) funded and published a guide to gender-responsive strategies for female offenders (Bloom et al. 2003), maintains an online directory of programs for women involved with the criminal justice system (http://nicic.gov/wodp/), and offers a curriculum to support management and operations within women's prisons that draws at least in part on feminist theoretical and empirical research. Modules include women's pathways to prison, gender-responsive management principles, and methods of addressing staff sexual misconduct, among other topics.

Controversy continues to center, though, on the appropriateness of risk and needs assessment tools developed for use with male inmates – and especially the Level of Service Inventory-Revised (LSI-R) – when working with female offenders (see Morash 2009, for a summary of this debate). The NIC funded a multisite, multilevel (prison, probation, parole) Women's Classification Study to determine, first, whether assessment and treatment of gender-responsive needs pertaining to trauma, abuse, mental health, self-efficacy, self-esteem, parenting, and relationships are relevant to women's future offending and other adverse outcomes, and second, whether adding items tapping these gender-responsive needs improves the predictive validity of assessment tools such as the LSI-R. The research team affirmed that programming to address women's unique needs is relevant for criminal justice outcomes and that including gender-responsive items significantly improves the predictive validity of so-called "gender-neutral" tools (Van Voorhis et al. 2008). They caution, however, that the new instruments are intended for use in gender-responsive, evidence-based treatment centers where practitioners are skilled in addressing women's needs. Moreover, adequate resources must exist to support programming designed to "empower women, address and accommodate trauma, stabilize symptoms of mental health, accommodate family reunification, teach healthy relationships, facilitate communication with children, provide parenting classes, strengthen vocational, educational, and life skills, and provide gender-responsive substance abuse treatment" (Van Voorhis et al. 2008:19).

On the ground, though, the lack of resources and training significantly impedes efforts to offer gender-responsive programming. Perhaps the best-known program for women in state and county prisons and jails is Girl Scouts Behind Bars. This partnership between the Girl Scouts and the National Institute of Justice, begun in 1992, is designed to renew bonds between girls and their incarcerated mothers.

At the federal level, the Federal Bureau of Prisons' female offender educational, recreational, and job training programs are designed to be comparable to those available to men. While comparable programming is arguably better than no programming, as was the case for women for decades, the lack of gender-responsive programming in these core areas is problematic. Currently, the Bureau of Prisons' gender-specific programs are largely limited to medical and social services related to pregnancy and child birth, and to a 3-month community residential program (Mothers and Infants Nurturing Together) for female inmates who are pregnant at the time of commitment, low risk, and who meet other eligibility criteria. Aside from this program, women inmates in federal prisons may visit with their infants or older children only if the children are accompanied by an adult (http://www.bop.gov/inmate_programs/female.jsp).

Do These Policies and Programs Reflect Feminist and Gendered Theorizing?

While the legislative initiatives, policies, and programs discussed above broke new ground and, in some ways, reflect feminist concerns, they have also reinforced two dominant social constructions of women – first, as victims who must be protected and, second, as needing assistance in order to be "good mothers." Women are seen as rather passive in both of these constructions. There is little recognition that women have agency and are making choices that appear reasonable to them, given the structural conditions of their lives.

Taking steps to stop sexual and other forms of violence against women and recognition of the blurred boundaries between victimization and offending are of critical importance, but feminist and gendered theories remind us that this focus on women's victimization is insufficient for at least three reasons. First, it reinforces stereotypic images of women as weak and in need of protection. Second, it risks treating women as a monolithic, homogenous group rather than recognizing the patterned differences (based on race, ethnicity, class, gender identity, sexual orientation, age, physical ability, etc.) in their experiences and responses to victimization. For example, very few gender-specific programs incorporate culturally appropriate practices or address issues faced by lesbian and transgendered youth. And third, legislation and programming that focus on victimization tend to ignore the central problems of patriarchy, racial discrimination, and economic inequality that coalesce to make women vulnerable to men's violence in the first place. Thus, theoretical and empirical scholarship on gendered lives and gendered crimes reminds us of the complexity of women's and girls' (and men's and boy's) experiences, the multiple sources of inequality and disadvantage that may shape those experiences, and the diverse ways in which they respond to their circumstances (see similarly Hannah-Moffat 2010).

In addition to this focus on victimization, the second theme underscoring much of the programming for female offenders, and especially for those who are incarcerated, is based on their status as mothers or future mothers. These include traditional parenting classes and newer interventions such as prison nurseries, extended visitation for children, and programs such as Girl Scouts Behind Bars. This approach is beneficial, as studies confirm repeatedly that being able to spend time with their children is very important to the women and their children (e.g., Bloom et al. 2003; Chesney-Lind and Pasko 2004).

Nevertheless, this limited focus on women as mothers risks reinforcing gender, racial, and cultural stereotypes of women as defined by motherhood, hegemonic norms regarding what makes a woman a "good mother," and appropriate societal responses to "good" and "bad" mothers. Previous research has demonstrated that women who are perceived by court officials to be "good mothers" are granted some leniency and are less likely to be incarcerated, whether because their families are seen as capable of providing sufficient informal social control or because of the practical costs of caring for children of incarcerated mothers. Not surprisingly, perceptions of family roles and of appropriate sanctions when "good" and "bad" mothers transgress societal norms and laws have been found to intersect with race, class, and gender orientation. Contemporary legislation that mandates incarceration for a wide range of offenses and circumstances (e.g., probation violations) has resulted in incarcerating more women and, particularly, more poor women of color, than ever before. Yet a primary emphasis of much of the programming for incarcerated women focuses on one aspect of women's lives – trying to help them become better mothers – without adequately addressing other critical problems that they face (e.g., employment, housing, transportation, drug and alcohol addiction) or the systemic issues that shape women's choices.

Future Directions

There is a growing recognition among federal agencies such as the Office of Juvenile Justice and Delinquency Prevention and the National Institute of Corrections, as well as research centers such as the National Center for Crime and Delinquency, of the need for theoretically informed, evidence-based programs that are gender specific and culturally appropriate. Such programs will require sustained funding and ongoing evaluations to ensure that the most effective models are widely available. Lack of adequate resources to develop and maintain programming is especially problematic for incarcerated girls and women. Improved and enhanced staff training and support for probation officers, correctional and detention officers, and other court personnel is also of critical importance. This training should be designed to debunk cultural and gender stereotypes held by staff and provide them with a broader understanding of the multifaceted contexts and complexities of girl's and women's lives.

Increasingly, programming for girls and women incorporates recognition of gendered pathways and the blurred boundaries between victimization and offending. This must remain a central component of gender-responsive programming, but it is not adequate or sufficient to explain the varied and complex situations and structural conditions that result in women and girls – and men and boys – engaging in crime and delinquency. Attention to victimization histories must be augmented in a number of ways, including educational, employment, and interpersonal skill building. Peer and family relationships also must be explicitly addressed. Girls' delinquency often occurs within the context of their friendship and romantic relationships, and past and present family dynamics are often a factor in girls' and women's offending. Finally, programming for adolescents must recognize and address the ways in which concentrated disadvantage in neighborhoods and schools shapes girls' and boys' available options and choices.

In sum, curriculum development and staff training must continue to reflect differences in offending patterns (gendered ratios of crime and delinquency) but must also incorporate new understandings of gendered lives, gendered pathways, and gendered crimes if they are to be effective. Moreover, programming must take into account the structural, institutional, and interpersonal (i.e., macro, meso, and micro) forces and dynamics through which race, gender, and class collectively pattern, shape, and constrain individuals' choices and actions. Simultaneously, scholars are encouraged to partner with practitioners to develop, assess, and refine theoretically and empirically strong prevention and intervention models grounded in an understanding of the multifaceted ways in which gender matters.

Recommended Reading and References

Bloom B, Owen B, Covington S (2003) Gender-responsive strategies: research, practice, and guiding principles for women offenders. National Institute of Corrections, Washington, DC

Caringella S (2009) Addressing rape reform in law and practice. Columbia University Press, New York

Chesney-Lind M, Irwin K (2008) Beyond bad girls: gender, violence and hype. Routledge, New York

Chesney-Lind M, Morash M (eds) (2011) Feminist theories of crime. Ashgate, Surrey

Chesney-Lind M, Pasko L (eds) (2004) Girls, women, and crime: selected readings. Sage, Thousand Oaks

Daly K (1992) Women's pathways to felony court: feminist theories of law-breaking and problems of representation. S Cal Rev L Women's Stud 2:11–52

Daly K (1998) Gender, crime and criminology. In: Tonry M (ed) The handbook of crime and justice. Oxford University Press, Oxford, pp 85–108

Dobash RE, Dobash R (1992) Women, violence and social change. Routledge, New York

Foley A (2008) The current state of gender-specific delinquency programming. J Crim Just 36:262–269

Gaarder E, Rodriguez N, Zatz MS (2004) Criers, liars and manipulators: probation officers' views of girls. Justice Q 21:547–578

Giordano PC, Cernkovich SA, Holland D (2003) Changes in friendship over the life course: implications for desistance from crime. Criminology 41:293–327

Hannah-Moffat K (2010) Sacrosanct or flawed: risk, accountability and gender-responsive penal politics. Curr Issues Crim Justice 22:193–215

Heimer K, Kruttschnitt C (eds) (2006) Gender and crime: patterns in victimization and offending. New York University Press, New York

Hirschel D, Buzawa E, Pattavin A, Faggiani D (2007) Domestic violence and mandatory arrest laws: to what extent do they influence police arrest decisions? J Crim L Criminol 98:255–298

Maher L (1997) Sexed work: gender, race and resistance in a Brooklyn drug market. Oxford University Press, Oxford

Martin S, Jurik N (2007) Doing justice, doing gender: women in law and criminal justice occupations. Sage, Thousand Oaks

Messerschmidt JW (1993) Masculinities and crime: critique and reconceptualization of theory. Roman and Littlefield, Lanham

Miller J (1998) Up it up: gender and the accomplishment of street robbery. Criminology 3:27–66

Miller J, Mullins CW (2009) Feminist theories of girls' delinquency. In: Zahn M (ed) The delinquent girl. Temple University Press, Philadelphia, pp 30–49

Morash M (2009) A great debate over using the Level of Service Inventory-Revised (LSI-R) with women offenders. Criminol Public Policy 8:173–181

Steffensmeier DJ, Schwartz J (2004) Trends in female criminality: is crime still a man's world? In: Price BR, Sokoloff NJ (eds) The criminal justice system and women: offenders, victims, and workers. McGraw Hill, New York, pp 95–111

Van Voorhis P, Salisbury E, Wright E, Bauman A (2008) Achieving accurate pictures of risk and identifying gender responsive needs: two new assessments for women offenders. US Department of Justice, National Institute of Corrections, Washington, DC

Zahn M (ed) (2009) The delinquent girl. Temple University Press, Philadelphia

Gendering Traditional Theories of Crime

Stacy De Coster[1] and Karen Heimer[2]
[1]Department of Sociology and Anthropology, North Carolina State University, Raleigh, NC, USA
[2]Department of Sociology, The University of Iowa, Iowa City, IA, USA

Overview

The major theoretical paradigms in criminology developed with little to no consideration of sex or gender differences in offending. Insights from gender and feminist studies demonstrate, however, that ignoring sex and gender in explanations of crime is to the detriment of the field. In this essay, we discuss contributions of gender and feminist studies to criminology with an emphasis on gender research that developed within the major criminological paradigms as well as research that steps outside traditional paradigms to draw more explicitly from gender and feminist theory. These bodies of research have underscored the need to consider how gender influences male offending, female offending, and the gender gap in offending through shaping power arrangements, opportunities, definitions of the sexes, and identities. We propose that one of the most promising directions for future work in criminology lies in further integration of traditional criminological perspectives with emerging work on offending rooted in feminist paradigms and not yet integrated with some of the core insights from theories of offending.

Introduction

Research consistently demonstrates that males commit more crime than females. Perhaps because of the persistent gender gap in crime, the major criminological paradigms developed with little concern for understanding female offending. Indeed, early work often ignored females altogether or focused on explaining sex differences in behavior, but did not incorporate gender as a social process that might explain sex differences and similarities in criminal involvement. Insights from gender and feminist studies, however, have shown that ignoring gender prohibits full understanding of the sources not only of female offending but also of male offending and of the gender gap in offending.

In this entry, we discuss the contributions of gender and feminist studies to the major criminological paradigms – control, strain, and learning. In doing so, we discuss research that demonstrates that the variables and processes specified in traditional criminological theories are shaped by gender in ways that contribute to the gender gap in crime. Simply emphasizing that gender shapes these variables and processes, however, proves insufficient for understanding the social forces contributing to the gender gap in crime. As such, we highlight work that has drawn more heavily on feminist theorizing and gender studies to introduce new variables and processes – for example, power arrangements, gendered opportunities and constraints, and cultural meanings of gender – to traditional criminological theories. We maintain that these lines of work underscore limitations in traditional theories and suggest avenues for the expansion of criminological theory that are important not only for understanding the gender gap in offending but for understanding crime generally. We also discuss research on gender and crime that departs from traditional criminological theory to focus on relationships between masculinities, femininities, and intersectionalities and crime. This work has made important contributions to criminology beyond what has been achieved by work focused on traditional theories of crime. It is our view that more theorizing and research must integrate new gendered insights into traditional perspectives to provide more thorough understanding of the structural and social-psychological processes leading to male offending, female offending, the gender gap in offending, and gender/race/class differences in offending.

Theorizing Gender Vis-Vis Traditional Theories of Crime and Delinquency

Early theorizing in criminology provided only cursory explanations for the gender gap in illegality, positing that females have either less exposure to the processes and variables in traditional theories that promote illegal behavior or more exposure to those that control such behavior (Sutherland 1947; Gottfredson and Hirschi 1990). Some scholars have demonstrated that the concepts from traditional theories have gender-differentiated effects on crime and delinquency, proposing that these differences are linked to gendered processes that have remained largely unmeasured and under-theorized (e.g., LaGrange and Silverman 1999; Broidy 2001; Lonardo et al. 2009). There, of course, have been some important advances within each of the criminological traditions – control, strain, and learning – that have drawn heavily from feminist theory and gender studies (Hagan et al. 1987; Chesney-Lind 1989; Heimer 1996; Heimer and De Coster 1999). We discuss the development of each of these traditions and the contributions of feminist and gender studies to these traditions below.

Control Theories

According to control theories of crime, everyone would be tempted to break the law in the absence of social constraints. The prime focus in this category of theory, therefore, is not why people commit crime but why they refrain from illegal behavior. The most prominent perspectives in this tradition – Hirschi's (1969) social control theory and Gottfredson and Hirschi's (1990) self-control theory – were not formulated to address the role of gender differences in offending, although some research has moved in this direction. The power-control theory of Hagan and colleagues (1987), for example, was specifically formulated to address gender and sex differences in offending and, in so doing, contributed importantly to criminological thinking.

Briefly, Hirschi's (1969) social control theory suggests that bonds to society – attachment to conventional others, commitment to conventional pursuits, involvement in conventional activities, and belief in the single moral order – effectively explain offending regardless of the sex of the offender. Scholars subsequently have suggested that the process of social bonding may be gendered. For example, drawing on feminist theory about relationships to others, some scholars posit that females develop stronger attachments to others and that this is associated with lower rates of female than male offending (e.g., Heimer 1996). Although some scholarship concludes that family bonds influence female and male delinquency similarly (e.g., Kruttschnitt and Giordano 2009), other research supports the proposition that the process of bonding through attachments to others is gendered (e.g., Heimer and De Coster 1999). Additional research supports a gendered bonding perspective, reporting that conventional beliefs control delinquency more strongly for females than for males (Liu and Kaplan 1999). This research generally suggests that understanding sex differences in crime and delinquency may require thinking conceptually about gender and moving beyond the original statements of social control theory.

The same is true in the case of research based on Gottfredson and Hirschi's (1990) self-control theory. Proposed to be gender neutral, the theory posits that sex differences in low self-control – typified by impulsiveness, thrill-seeking, and an inability to delay gratification – are the primary cause of gender differences in law violation. More specifically, sex differences in offending result from the fact that females acquire higher levels of self-control because parents are more likely to monitor and punish misbehavior in daughters than sons (e.g., LaGrange and Silverman 1999). These differences established early in life are presumed to persist into adulthood. Beyond this, researchers have not often explored the possibility that the effect of self-control on illegality can vary across sex or that the processes of self-control may in fact be gendered. Pratt and Cullen's (2000) meta-analysis indicates, however, that the effect size of self-control is larger for females than for males. Theoretical discussion of the gendered reasons for such a difference has been minimal to date. This may reflect the fact that the theory is intended to be general and hence, in effect, gender neutral.

By contrast, power-control theory, developed by Hagan and his colleagues (Hagan et al. 1987), maintains that understanding sex differences in law violation requires a perspective that focuses explicitly on gendered power structures and control processes that relay gendered messages within families. Specifically, power-control theory maintains that gender differences in the control of sons and daughters, via emotional attachments and supervision, can be traced to power relations between mothers and fathers (Hagan et al. 1987). When fathers experience greater power than mothers in the workplace, this is said to translate into greater paternal than maternal power in the home (patriarchal households). When fathers and mothers have equal power in the workplace, they share more or less equally in power in the home (egalitarian households). These power dynamics shape the relative control of daughters and sons, through supervision and emotional attachments, which influences risk preferences and perceptions among boys and girls. These processes ultimately create a sex gap in delinquency that is larger in patriarchal households than in egalitarian households. In more recent formulations of the theory, risk orientations have been replaced with discussions of gender schemas related to appropriate behaviors and characteristics of the sexes

(Hagan et al. 2004). Empirically, research has supported the prediction that the sex gap in delinquency is larger in patriarchal families than in more egalitarian families because of gender-differentiated social control, gendered orientations toward risk, and gender schemas (e.g., Hagan et al. 1987, 2004). While some research fails to replicate some of these findings (e.g., Jensen and Thompson 1990), power-control theory has received a good deal of empirical support. This support demonstrates that traditional control perspectives that fail to consider the way in which gendered power structures shape controls and gendered messages within families offer an incomplete portrayal of the structural and social-psychological roots of law violation among both males and females.

Strain Theories

Strain theories maintain that crime is not an innate characteristic of individuals and instead focus on the social factors that promote offending. A central tenet of these perspectives is that illegality results from failure to achieve positively valued goals. Classic strain explanations focused on failure to achieve economic- or class-based status goals (e.g., Merton 1938; Cloward and Ohlin 1960), which some argue renders them entirely ineffective for explicating the sex gap in offending since females face more structural barriers to success in the economic realm than males (Chesney-Lind 1989). However, macro- and microlevel reformulations of strain theory coupled with insights from feminist and gender studies address this gendered challenge.

At the macrolevel, gender researchers have drawn on Cloward and Ohlin's (1960) extension of anomie/strain theory to emphasize that blocked legitimate opportunities do not always result in illegality because people also face barriers to economic success in the illegitimate world of crime. For instance, Steffensmeier (1983) proposes that the petty, nonserious, and non-lucrative nature of female crime can be explained largely by the fact that male-dominated illegal networks exclude or restrict women from organized criminal enterprises. This theme is one that finds prominence also in ethnographic work on economic street crimes, including robbery, burglary, and drug-trafficking (Miller 1986, 2001; Maher 1997 ; Mullins and Wright 2003). These studies demonstrate that the illegal activities associated with the highest pay, power, and status are largely the domain of males who have access to male-dominated street networks that operate to limit female participation in the most lucrative criminal enterprises in the streets. Daly's (1989) research shows that this pattern of exclusion applies also to white-collar crimes where women's opportunities are restricted by their occupational positions and limited access to both organizational resources and male-dominated criminal networks. Although these studies are not situated specifically in the strain tradition, they speak to strain theories by highlighting the importance of gender in shaping opportunities and constraints in the illegitimate world of crime. More generally, they demonstrate that understanding differences in crime across broad categories of not only gender but also of age, class, and race may require consideration of illegitimate opportunities that often are made accessible through social networks that tend toward homosocial reproduction.

At the microlevel, Agnew (2006) broadens the scope of crime producing strains beyond the economic realm by considering goals related to the pursuit of meaning, interpersonal relationships, and a desire for justice. His general strain theory posits that failure to achieve these goals produces negative emotions that promote crime when individuals lack social supports and legitimate coping resources.

In a gendered extension of this theory, Broidy and Agnew (1997) propose that gender socialization shapes the variables in general strain theory in ways that are relevant for understanding the gender gap in illegality. Specifically, they posit that the types of strains to which males and females are exposed, their emotional responses to these strains, and their coping mechanisms (including social support) are shaped by gender. For instance, they draw on feminist work to suggest that males and females experience different types of strains because they have gender-differentiated goals; females focus on maintaining relationships, nurturance, and the treatment of others in interaction, whereas males focus on goals related to economic success, personal achievement, and the outcomes of interaction. They propose that the gender-differentiated strains experienced by males and females are relevant for understanding the gender gap in offending because male-typical strains may be more conducive to illegality than female-typical strains. For instance, the male-typical strains of failure to achieve economic goals and criminal victimization are conducive to property and violent offending, respectively. Female-typical strains are less conducive to illegal behavior because such behavior often harms others or damages social relationships. Although research supports the notion that strains align with gender-differentiated goals, there is little evidence that male-typical strains are more likely than female-typical strains to produce crime and delinquency (see Agnew 2006).

Emotional responses to strain have been shown to be gendered in ways consistent with Broidy and Agnew's (1997) proposition that males and females both experience anger, but they do so in ways that are qualitatively distinct. That is, both males and females respond to strain with anger (Broidy 2001). However, female anger is more likely than male anger to be accompanied by depression, which Broidy and Agnew (1997) suggest tempers the impact of anger on their offending. This claim is not supported by empirical research, however (De Coster and Zito 2010). The proposition that coping and social support resources more effectively buffer the impact of strain on illegal behaviors among females than males also has not been supported empirically (see Agnew 2006).

Overall, strains and emotions are gendered in ways consistent with Broidy and Agnew's (1997) propositions. Empirical evidence indicates, however, that the gendering of strains and emotions is fairly unimportant for understanding the gender gap in offending. Instead, gender differences appear to emerge because males are more likely than females to respond to strain and negative emotions through illegal behaviors, suggesting that more emphasis should be placed on understanding the gendered messages males and females learn about appropriate methods for dealing with strain and expressing negative emotions.

The significance of Broidy and Agnew's (1997) research for understanding female offending may lay in the attention it draws to strains outside of the economic sphere. This insight has been at the center of some feminist research that focuses on physical and sexual abuse at the hands of family members as particularly relevant for understanding female crime because females are more likely than males to be the victims of such abuse in patriarchal society (see Chesney-Lind 1989). A common coping strategy for dealing with abuse at home – running away – has been criminalized, leading young girls to the criminal justice system or to the streets, where their opportunities for survival are limited, as discussed above. Given that females are valued as sexual objects in patriarchal society, prostitution becomes a viable survival strategy for these girls even though this strategy would be largely unavailable to similarly situated boys (Miller 1986; Chesney-Lind 1989). Although research shows that males and females are equally likely to respond to family abuse with delinquency (see Kruttschnitt and Giordano 2009), the theme underlying this work is that exposure to abuse, the types of abuse to which youths are exposed, and the reasons for abuse may be intimately informed by power relations in patriarchal society. That is, these studies highlight the importance of focusing on the unique problems and circumstances that lead females into crime and delinquency because they lead lives markedly different from males in patriarchal society. We believe this insight also highlights the importance of considering the ways in which patriarchy uniquely shapes males' lives in ways that may make them particularly crime-prone (e.g., Messerschmidt 1993).

Learning Theories

Learning theories contend that crime is motivated by social factors but move away from an emphasis on blocked goals and stressful circumstances to focus on the socialization and interactional processes by which individuals develop and internalize attitudes, behavior patterns, and definitions of situations that promote offending. The original statement of learning theory was Sutherland's differential association theory (Sutherland 1947), which was later reformulated as social learning theory (see Akers 1998). As with other traditional theories of crime, learning theories were not formulated to address sex differences in offending; however, because they focus on differences across groups in orientations toward crime and deviance, even the earliest versions included discussion of differences in the socialization experiences of females and males that are linked to differences in law violation (Sutherland 1947: 100–101). Following this observation, research on the gender gap has explored the ways that sex and gender shape the form and content of learning and interactions that lead to offending. The addition of concepts from feminist theory to this tradition emphasizes an important dimension of gender in society that has not been addressed in traditional learning theories – cultural definitions of gender, which like definitions of law are learned in interactions (e.g., Heimer and De Coster 1999).

A major tenet of learning theories of crime is that people learn definitions and behavior patterns that are both favorable and unfavorable to law violation in interactions with others; when a person's definitions and behavior patterns favorable to offending are in excess, law violation is likely to occur (Sutherland 1947; Akers 1998). Factors that affect interactional patterns – like sex and gender – can be expected to shape learning about crime and delinquency. Research shows that, regardless of sex, people learn law violation from interacting with criminal peers (e.g., Smith and Paternoster 1987) and deviant families (Lonardo et al. 2009). In terms of sex differences, research on juvenile delinquency shows that boys break the law more often than girls in part because boys have more deviant friends and, in some cases, because the impact (i.e., size of the effect) of peer influence is larger for boys (e.g., Heimer and De Coster 1999; Bottcher 2001).

Heimer's (1996; Heimer and De Coster 1999) reformulation of traditional differential association theory draws on the feminist argument that, traditionally, females have been more encouraged than males to focus on care, concern, and the maintenance of interpersonal relationships. Their research shows that close relationships with parents are more important for learning delinquent definitions among females and that a more direct parental control – supervision – is more relevant for boys' learning. These findings suggest that gender differences in learning crime cannot be understood in terms of only male-female differences in levels of supervisions and control, as emphasized by Sutherland (1947); rather, theory and research must also recognize that there are significant differences across sex in the mechanisms through which social groups (e.g., parents) communicate and affect the learning of definitions of crime and delinquency.

In addition, Heimer and De Coster (1999) maintain that it is not enough to consider the learning of definitions of the law; theory and research must also recognize that in gender-stratified societies, people are surrounded by strong cultural definitions of gender that become internalized and operate alongside definitions of the law to influence crime and delinquency. Because the dominant or hegemonic cultural definitions of femininity are more inconsistent with harming others (e.g., through theft or aggression) than are dominant definitions of masculinity, gender definitions would seem to be important for understanding criminal behavior. Indeed, research has shown that traditional conceptions of masculinity and femininity are important for understanding aggression and law violation (e.g., Simpson and Elis 1995; Heimer and De Coster 1999; Bottcher 2001). As such, learning theories cannot be limited to a focus on the cultural product of definitions of the law but also must acknowledge the role of definitions of gender.

Gender Studies, Feminist Theory, and Explanations of Offending

Other research on gender and crime – that is not clearly associated with the three dominant traditions in criminology – also emphasizes the importance of cultural definitions of gender for understanding crime. However, this work stresses the need to consider how race, class, gender, and sexuality shape persons experiences and definitions of gender in ways that deepen our understanding of law violation. We maintain that these developing lines of study are becoming important not for simply understanding gender differences but also for understanding variability in crime and violence among males and among females (De Coster and Heimer 2006).

There is a growing body of research that highlights the importance of hegemonic gender definitions – or culturally dominant ideals about gender. This work draws inspiration from gender studies, especially the theorizing of Connell (1995), which argues that societal definitions of masculinities represent a power hierarchy, with hegemonic masculinities being the idealized masculinity that is accessible to members of powerful groups, such as heterosexual white men from the middle and upper classes. By contrast, marginalized masculinity is constructed by members of oppressed groups as they try to navigate their disadvantages in terms of resources and power. The reaction of marginalized groups to hegemonic masculinity – unattainable to men who cannot demonstrate masculinity via legitimate avenues – can involve the use of violence and crime (e.g., Mullins et al. 2004). Messerschmidt (1993) argues that, over time, violence can become an accepted way to claim masculinity in communities that experience persistent blockage of legitimate avenues for "doing gender". In these settings, men can become culturally accountable to marginalized definitions of masculinity that are characterized by competition through the show of physical power, heterosexuality, and the use of violence (e.g., Mullins et al. 2004; Miller 2008).

Consistent with this, sociological studies suggest that claiming masculinity in extremely disadvantaged urban communities can require demonstrating "nerve" in various ways, including "throwing the first punch, getting in someone's face, or pulling the trigger" (Anderson 1999: 92). Research shows that these conceptualizations of masculinity have important implications for understanding law violation, including street crime and violence against women (e.g., Messerschmidt 1993; Mullins et al. 2004; Miller 2008). Although not born of traditional criminological theory, this line of work has very important implications for understanding variability in offending among males and across groups at different junctures of societal power. While it is clearly important for illuminating the "gender gap," research and theorizing on masculinities and offending has a good deal to offer to criminology *generally* – even to research that does not address sex differences. In short, it shows that "gender matters," in that constructing masculinities is a part of daily life that has implications for crime among men.

What has been less well developed is the relationship between femininities and offending. Some research proposes that normative conceptions of femininity are relatively incompatible with crime and violence (e.g., Simpson and Elis 1995; Heimer and De Coster 1999). This logic, however, has not accounted for variations in femininities but has focused on what gender studies scholars like Connell (1995) call "emphasized" or dominant femininities, which are most accessible to white, middle-class women (see Hill Collins 2004). These dominant femininities – which are inconsistent with the use of crime and violence to solve problems – support hegemonic masculinities and are sometimes not a cultural option for displaying gender among minority and poor women (Simpson 1991; Hill Collins 2004). Consequently, some researchers stress the need to focus on alternative forms of femininities that are more conducive to violence than the hegemonic form, resulting in discussions of "pariah" (Schippers 2007), "bad girl" (Messerschmidt 1997), and "racialized" (Hill Collins 2004) femininities.

Others suggest that females who break the law are not enacting alternative femininities but may be enacting masculinities or using law violation to protect feminine identities (Miller 2001). Indeed, ethnographic studies indicate that females in economically marginalized communities, like males, use violence instrumentally to garner respect and protect identities and reputations. Unlike males, who often are protecting identities as tough and masculine, females more often use violence to defend themselves against threats to their respectability or reputations as “respectable girls” (Miller 2001, 2008; Miller and Mullins 2006; Mullins et al. 2004). Additionally, it has been suggested that crime and violence by females can be a reaction to men displaying masculinity. For example, girls in gangs may act tough to avoid victimization (Miller 2001, 2008), and women may retaliate with violence and other crimes in response to prior and present abuse at the hands of men (Maher 1997). These examples may well be reactions to dangerous environments shaped by patriarchal power structures and not the enactment of hegemonic or alternative femininities.

Generally, there appear to be a variety of mechanisms through which gendered structures and identities influence crime and violence among females. Further articulation of these mechanisms is an important endeavor for the criminological community because theory and research on multiple masculinities and femininities may be particularly well suited for explicating the ways in which race, class, and gender intersect in the production of race-class-gender patterns of offending (De Coster and Heimer 2006).

Conclusions

This research shows that the variables and processes identified as important in the etiology of offending are gendered in ways that contribute to the sex gap in offending. This, of course, marks an important contribution to the understanding of crime and the sex gap in crime; however, a more comprehensive understanding of criminality by males and females requires consideration of variables and processes rooted specifically in gendered arrangements which have been highlighted in feminist and gender studies.

Researchers have incorporated feminist and gender studies into each of the theoretical traditions discussed, demonstrating the utility of wedding traditional criminology with gender and feminist insights by uncovering structural and social-psychological sources of offending by males and females that are shaped by gender – for example, power arrangements, illegitimate opportunities, and cultural definitions of gender – and that heretofore had been de-emphasized or ignored altogether (Hagan et al. 1987, 2004; Chesney-Lind 1989; Heimer and De Coster 1999). These feminist-informed studies are important not only for having reoriented traditional theories of crime but also for having laid the groundwork for the evolution of research on crime rooted more squarely in feminist and gender studies paradigms. This research, though not rooted in traditional criminology, proves promising for pushing traditional criminology forward as it becomes more fully incorporated into general theories of offending.

An important corrective to traditional criminology derived from emerging feminist scholarship on crime is that biological sex and gender socialization should be decoupled to allow for an emphasis on variations in gendered power, identities, and definitions within sex that give rise to offending differences within sex groups. Gender and feminist studies do not simply contribute to our understanding of the gender gap in crime; they significantly expand our understanding of the causes of crime. What we have learned is that even studies of male crime, for instance, are incomplete without consideration of how gender shapes everyday life. The incorporation of this insight into understanding crime is likely to be at the heart also of understanding variations in offending across groups and individuals situated at various junctures in race, class, and gender hierarchies (see De Coster and Heimer 2006).

An essential direction for future work in criminology lies in the integration of traditional criminological perspectives with emerging work on offending rooted in feminist paradigms. Miller and Mullins (2006) recently advised that the most promising direction for the future of feminist criminology resides in work that enriches its analyses of the gendered lives of males and females with work from the broader field of criminology. We add to this that the most promising direction for the future of criminological endeavors generally resides in drawing further insights from feminist studies that underscore the importance of considering how gender structures everyday life.

Related Entries

- ▶ Feminist Criminological Theory
- ▶ Gendered Theory and Gendered Practice
- ▶ History of Criminological Theories: Causes of Crime

Recommended Reading and References

Agnew R (2006) Pressured into crime: an overview of general strain theory. Roxbury, Los Angeles

Akers RL (1998) Social learning and social structure: a general theory of crime and deviance. Northeastern University Press, Boston

Anderson E (1999) Code of the street. W.W. Norton, New York

Bottcher J (2001) Social practices of gender: how gender relates to delinquency in the everyday lives of high-risk youths. Criminology 39:893–932

Broidy LM (2001) A test of general strain theory. Criminology 39(1):9–36

Broidy L, Agnew R (1997) Gender and crime: a general strain theory perspective. J Res Crime Delinq 34(3):275–306

Chesney-Lind M (1989) Girls' crime and woman's place: toward a feminist model of female delinquency. Crim Delin 35(1):5–29

Cloward RA, Ohlin LE (1960) Delinquency and opportunity: a theory of delinquent gangs. Free Press, New York

Connell RW (1995) Masculinities. University of California Press, Berkeley

Daly K (1989) Gender and varieties of white collar crime. Criminology 27:769–793

De Coster S, Heimer K (2006) Crime at the intersection: Gender, race, and violent offending. In: Peterson R, Krivo L, Hagan J (eds) The many colors of crime: Inequalities of race, ethnicity and crime in America. New York: New York University Press, pp. 138–156

De Coster S, Zito RC (2010) Gender and general strain theory: the gendering of emotional experiences and expressions. J Contemp Crim Just 26(2):224–245

Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Hagan J, Simpson J, Gillis AR (1987) Class in the household: a power-control theory of gender and delinquency. Am J Sociol 92(4):788–816

Hagan J, Boehnke K, Merkens H (2004) Gender differences in capitalization processes and the delinquency of siblings in Toronto and Berlin. Br J Criminol 44(5):659–676

Heimer K (1996) Gender, interaction, and delinquency: testing a theory of differential social control. Soc Psychol Q 59(1):39–61

Heimer K, De Coster S (1999) The gendering of violent delinquency. Criminology 37(2):277–318

Hill Collins P (2004) Black sexual politics: African Americans, gender, and the new racism. Routledge, New York

Hirschi T (1969) Causes of delinquency. University of California Press, Berkeley

Jensen GF, Thompson K (1990) What's class got to do with it? A further examination of power-control theory. Am J Sociol 95(4):1009–1023

Kruttschnitt C, Giordano P (2009) Family influences on girls' delinquency. In: Zahn M (ed) The delinquent girl. Philadelphia: Temple University Press, pp. 107–126

LaGrange TC, Silverman RA (1999) Low self-control and opportunity: testing the general theory of crime as an explanation for gender differences in delinquency. Criminology 37(1):41–72

Liu X, Kaplan HB (1999) Explaining the gender difference in adolescent delinquent behavior: a longitudinal test of mediating mechanisms. Criminology 37(1):195–216

Lonardo RA, Giordano PC, Longmore MA, Manning WD (2009) Parents, friends, and romantic partners: enmeshment in deviant networks and adolescent delinquency involvement. J Youth Adolesc 38(3): 367–383

Maher L (1997) Sexed work: gender, race, and resistance in a Brooklyn drug market. Clarendon Press, Oxford

Merton RK (1938) Social structure and anomie. Am Sociol Rev 3(5):672

Messerschmidt JW (1993) Masculinities and crime. Rowan and Littlefield, Totowa

Messerschmidt JW (1997) Crime as structured action. Sage, Thousand Oaks

Miller EM (1986) Street woman. Temple University Press, Philadelphia

Miller J (2001) One of the guys: girls, gangs and gender. New York, Oxford

Miller J (2008) Getting played. New York University Press, New York

Miller J, Mullins CW (2006) Feminist theories of crime. In: Cullen FT, Wright J, Blevins K (eds) Taking stock: The status of criminological theory. Piscataway, NJ: Transaction Publishers, pp 217–250

Mullins C, Wright R (2003) Gender, social networks, and residential burglary. Criminology 41:813–840

Mullins C, Wright R, Jacobs BA (2004) Gender, streetlife and criminal retaliation. Criminology 42(4):911–940

Pratt TC, Cullen FT (2000) The empirical status of Gottfredson and Hirschi's general theory of crime: a meta-analysis. Criminology 38(3):931–964

Schippers M (2007) Recovering the feminine other: masculinity, femininity, and gender hegemony. Theor Soc 36:85–102

Simpson S (1991) Caste, class, and violent crime: explaining difference in female offending. Criminology 29:115–135

Simpson SS, Elis L (1995) Doing gender: sorting out the caste and crime conundrum. Criminology 33(1):47–81

Smith DA, Paternoster R (1987) The gender gap in theories of deviance: issues and evidence. J Res Crime Delinq 24(2):140–172

Steffensmeier DJ (1983) Organization properties and sex segregation in the underworld: building a sociological theory of sex differences in crime. Soc Forces 61:1010–1032

Sutherland EH (1947) Principles of criminology. J. B. Lippincott, Oxford

General Patrol

► Preventive Patrol

General Strain Theory

Robert Agnew
Department of Sociology, Emory University,
Atlanta, GA, USA

Overview

General strain theory (GST) states that strains increase the likelihood of crime, particularly strains that are high in magnitude, are seen as unjust, are associated with low social control, and create some pressure or incentive for criminal coping. Examples include parental rejection, criminal victimization, a desperate need for money, and discrimination. These strains increase crime for several reasons; most notably, they lead to a range of negative emotions, which create pressure for corrective action. Crime is one possible response. Crime may be used to reduce or escape from strains (e.g., theft to obtain money, running away to escape abusive parents), seek revenge against the source of strain or related targets, or alleviate negative emotions (e.g., through illicit drug use). A range of factors, however, influence the response to strains. A criminal response is more likely when people lack the ability to cope in a legal manner, are disposed to crime, and the costs of crime are low and the benefits are high. This entry describes the current status of GST, and is organized around a set of propositions dealing with (a) the nature of those strains conducive to crime; (b) the reasons why these strains increase crime; (c) the factors influencing or conditioning the effect of these strains on crime; and (d) efforts to apply GST to new areas, such as group differences in crime.

General Strain Theory: Key Propositions

General strain theory was first proposed in 1992 and has since inspired hundreds of research reports (see Agnew 1992, 2006; Agnew and Scheuerman 2011; Hoffmann 2010). This research has tested the theory, proposed revisions in it, and applied it to new areas. It is therefore important to describe the current state of GST. Further, it is an opportune time to do so, given that 2012 was the twentieth anniversary of the theory. This entry draws on the research to present an updated version of GST, stated in the form of several core and secondary propositions. Each proposition is followed by a definition of key terms, the rationale for the proposition, summaries of the relevant research, and suggestions for further research. Given the large body of research on GST, it is only possible to cite previous reviews of GST and certain of the more recent research. And what follows is of course the author's view of the current state of GST; others may argue for additional revisions in and extensions of the theory. In fact, the chapters on GST in this volume by Baron, Slocum, and De Coster propose several major extensions in GST.

As the propositions below make clear, the basic form of GST remains intact. Certain strains increase the likelihood of crime, in part through their impact on negative emotions, with a range of factors influencing the likelihood of a criminal response. But the original statement of theory has been revised and extended in numerous ways. As noted below, the theory now better specifies the types of strain most likely to result in crime, more fully describes why these strains increase crime, and lists additional factors that may condition the effect of strains on crime. Also, the theory has been extended to explain group differences in crime, offending over the life course, and a broader range of crimes and deviant acts. At the same time, the theory is in need of further development in several areas. Most notably, the theory needs to better explain why some individuals are more likely than others to cope with strains in a criminal manner. Related to this, there should be an effort to better link GST to biological factors and larger social forces – both of which influence the exposure and reaction to strains. And efforts should be made to apply GST to the control of crime and the analysis of the criminal justice system.

The Nature of Those Strains Conducive to Crime

1. *Certain strains increase the likelihood of crime*. Strains refer to events and conditions that are disliked by individuals (Agnew 1992, 2006). Strains are similar to "stressors," but the term "strain" is used to emphasize the fact that GST is derived from prior strain theories in criminology and that it does not focus on all stressors, but rather on a subset of stressors conducive to crime. Strains may fall into one *or more* of three broad categories: (a) the inability to achieve valued goals (e.g., monetary, status goals), (b) the loss or threatened loss of valued stimuli (e.g., material possessions, the death of family members), and (c) the presentation or threatened presentation of negative stimuli (e.g., verbal and physical abuse). For example, an insult may involve the failure to achieve a valued goal (respect), the loss of valued stimuli (respectful treatment), and the presentation of a negative stimulus (the insult itself). The three categories of strain are presented not because they are necessarily distinct from one another, but rather to ensure that researchers consider a broad range of strains.

1a. *Those strains most conducive to crime (a) are high in magnitude; (b) are seen as unjust; (c) are associated with low social control; and (d) create some pressure or incentive for criminal coping*. These characteristics increase the likelihood that strains will impact those intervening mechanisms that lead to crime (see below). For example, strains that are high in magnitude and seen as unjust, such as an unprovoked physical attack, are more likely to make individuals angry than strains without these characteristics, such as an accidental bump. Agnew (2001, 2006) provides further information on these characteristics, including strategies for their measurement. Strains that are high in magnitude are high in degree (e.g., a large versus small financial loss), are frequent, recent, of long duration, and expected to continue into the future. Also, such strains are high in centrality; that is, they threaten the *core* goals, needs, values, activities, and/or identities of the individual. Strains that are seen as unjust typically involve the voluntary and intentional violation of a relevant justice norm or rule. Strains associated with low social control involve little direct control by others, weak attachment to conventional others, and/or little investment in conventional institutions. An example is parental rejection. Juveniles experiencing this strain are usually poorly supervised by and weakly bonded to their parents. Finally, strains that create some pressure or incentive for criminal coping are easily resolved through crime and/or involve exposure to others who encourage or model crime. For example, that strain involving a desperate need for money is readily resolved through crimes such as theft, drug selling, and prostitution. That strain involving the inability to achieve educational success, however, is not so easily resolved through crime. To give another example, strains involving criminal victimization expose individuals to others who model crime.

Several specific strains tend to possess the above characteristics, including parental rejection; erratic, excessive and/or harsh discipline; child abuse and neglect; negative secondary school experiences (e.g., low grades, poor relations with teachers); abusive peer relations; work in the secondary labor market (jobs with low pay, few benefits, poor working conditions); unemployment, especially when chronic and blamed on others; certain marital problems (frequent conflicts, verbal and physical abuse);

the failure to achieve selected goals (thrills/excitement, autonomy, masculine status, the desire for much money in a short period of time); criminal victimization; residence in economically deprived communities; homelessness; and discrimination based on race/ethnicity, gender, and religion. Research suggests that these strains increase the likelihood of crime, with some being among the most important causes of crime (see Agnew 2006; Agnew and Scheuerman 2011; Baron 2009).

Researchers, however, need to devote more attention to certain of these strains, such as discrimination (e.g., Eitle 2002; Simons et al. 2003). Also, researchers should explicitly measure characteristics such as the magnitude and perceived injustice of strains, determining if variation in them affects crime (see Agnew 2006; Rebellon et al. 2009). Unfortunately, the research on GST sometimes employs measures of strain that are rather general; so it is not clear whether the strains examined possess the above characteristics. For example, researchers often measure the experience of strains by employing checklists that ask such things as whether a family member has died. But when explaining crime, it makes a great deal of difference whether the death involved a great-grandmother who died of natural causes (moderate magnitude, little injustice) or a brother who was shot by rival gang members (high magnitude, high injustice). Related to this, there is a need for *qualitative* research on the nature of and response to strains; such research will allow for the detailed measurement of the nature of strains and perhaps suggest other characteristics that should be considered when explaining crime.

1b. *Strains may be objective or subjective in nature, with subjective strains having a greater effect on crime*. Objective strains refer to events and conditions disliked by most people in a given group. Subjective strains refer to events and conditions disliked by the people experiencing them (Agnew 2006). Research indicates that individuals often differ a good deal in their subjective evaluation of a given objective strain, such as divorce. Some individuals view their divorce as the worst thing that ever happened to them, while others view it as a cause for celebration. Subjective strains should have a greater impact on crime since they are more likely to trigger negative emotional reactions and other of the intervening processes conducive to crime (see below). At the same time, objective strains may increase crime even when not subjectively disliked, since these strains may still have criminogenic effects (e.g., they may lower social and self-control). Most studies employ objective measures of strain, focusing on events and conditions assumed to be disliked by most people. The few studies that have examined both objective and subjective strains have produced mixed results; only some find that subjective strains have a stronger effect on crime (e.g., Froggio and Agnew 2007). More research is needed in this area, including research that measures the subjective interpretation of strains in more detail. For example, are strains perceived to be high in magnitude and are they blamed on others (a key component of perceived injustice). Baron (2008) provides an excellent illustration of the potential value of such research, finding that the manner in which unemployment is interpreted has a major impact on whether it affects crime.

1c. *Strains may be experienced, vicarious, and anticipated, with experienced strains having a greater effect on crime*. GST focuses on the individual's personal experiences with strains, but certain vicarious and anticipated strains may affect crime (Agnew 2006; Baron 2009). Vicarious strains refer to the individual's awareness of strains experienced by others. Such strains are more likely to affect crime if they involve close others, such as family and friends; the individual feels some responsibility for the welfare of these others; the strains have the characteristics listed in 1a; and the strains signal a threat to the individual. An example involves the shooting death of a fellow gang member by those in a rival gang. Anticipated strains refer to the individual's expectation that current strains will continue into the future or that new strains will be experienced. Anticipated strains are more likely to result in crime when individuals believe that they have a high probability of occurring in the near future and they have the characteristics listed in 1a. Experienced strains should be more likely to affect crime than vicarious and anticipated strains, since they pose a more direct and immediate problem for individuals. At the same time, research on criminal victimization indicates that the victimization of close others and the anticipated victimization of oneself may increase crime (Agnew 2002; Baron 2009). Research on other types of vicarious and anticipated strain is needed (e.g., financial hardship experienced by close others, the anticipation of financial difficulties by corporate executives).

1d. *Strains are more likely to affect crime when they are clustered together in time*. The temporal clustering of strains is more likely to result in the criminogenic effects described below, such as the taxing of legal coping resources (see Agnew 2006). And certain research suggests that crime is more likely when several strains are experienced at the same time (Slocum et al. 2005). This clustering effect implies an interaction: The effect of a given strain on crime will be stronger when other strains are present. It is often difficult, however, to examine the many interactions between multiple strains. But researchers can examine both the separate and the cumulative effects of strains, with the cumulative effect being measured in terms of the number of criminogenic strains experienced at the same time or close together in time. Crime should become more likely as this number increases. More research is needed in this area, however. Among other things, the nature of this cumulative effect needs to be better specified. For example, it may be the case that the cumulative strain measure has a nonlinear effect on crime, such that further increases in the number of strains beyond a certain threshold point have little effect on crime.

Why the Above Strains Increase Crime (Intervening Processes)

2. *The strains listed above increase the likelihood of crime for several reasons; most notably, they lead to negative emotional states, which create pressure for corrective action – with crime being one possible response*. Recall that strains refer to disliked events and conditions, with the most criminogenic strains being high in magnitude and perceived as unjust. Not surprisingly, these strains lead to negative emotions such as anger, frustration, and/or depression. These emotions create pressure for corrective action: Individuals feel bad and want to do something about it. Crime is one response. Research suggests that state anger partly explains the effect of strains on crime, especially violence. And research has begun to focus on the mediating role played by other emotions, such as depression (Agnew 2006; Agnew and Scheuerman 2011). It is beginning to appear that some types of strain may be more conducive to certain negative emotions than others, and that some negative emotions may be more conducive to certain crimes than others. For example, anger may be more conducive to externalizing behaviors such as aggression, while depression may be more conducive to internalizing behaviors such as drug use. More research is needed in this area. Research is also needed on the reasons why negative emotions increase crime; for example, anger may increase aggression because it energizes individuals for action, creates a desire for revenge, reduces concern for the consequences of one's behavior, and/or limits the ability to engage in many legal coping behaviors – such as negotiation (Agnew 2006).

2a. *Strains reduce the ability to legally cope, making crime seem like a rational option*. A large financial loss or chronic unemployment, for example, may exhaust one's financial resources, leaving few legal options for obtaining money. Consequently, crime may appear to be the most effective way to respond to strain. As such, strains may sometimes lead to crime in the absence of negative emotions (see Hoffmann 2010).

2b. *Strains may also increase crime because they foster traits conducive to crime, particularly negative emotionality/trait anger and low constraint/low self-control*. This is especially true of chronic strains (Agnew 2006; Colvin 2000). Chronic strains tax legal coping resources, such that individuals are more easily upset when they experience new strains (a key characteristic of trait anger and negative emotionality). Also, certain strains – such as harsh and erratic discipline – may foster low constrain or low self-control.

Individuals must be taught to exercise self-restraint, and this occurs when they are consistently sanctioned for misbehavior in a reasonable manner. There is limited support for these arguments, with several studies finding that strains reduce self-control and, especially, increase trait anger/negative emotionality (Agnew 2006; Colvin 2000). Further, these traits partly mediate the effect of strains on crime. More research is needed here, however, particularly longitudinal research.

2c. *Strains may increase crime because they reduce social control*. Strains frequently involve negative and unjust treatment by others, including parents, spouses, teachers, employers, and those in the criminal justice system. Such strains may reduce attachment to these others, as well as commitment to school, work, and the criminal justice system. Individuals experiencing negative treatment may also limit their contact with these others: in some cases, running away from parents, dropping out of school, divorcing spouses, and quitting jobs. This reduces direct control, since these others are less able to monitor and sanction criminal behavior. Further, these effects may weaken beliefs condemning crime, since ties to those others who teach conventional beliefs are undermined. Beyond that, certain strains – such as chronic unemployment and homelessness – overlap with low social control. Researchers usually do not examine the effect of strains on social control, but several studies – some of which are longitudinal – have found that strains reduce the major types of control (Agnew 2006).

2d. *Strains may increase crime because they foster the social learning of crime*. Strains may increase the likelihood of association with criminal peers, a point emphasized by classic strain theorists (see Agnew 2006). Strains increase association with criminal peers by weakening social control, which frees individuals to associate with criminal peers, and by increasing the appeal of criminal groups. In particular, criminal groups are often seen as a solution to the strains one is experiencing. For example, criminal groups often provide status, protection from others, and opportunities to make money. In additions, strains foster beliefs conducive to crime. Individuals being treated in a negative and unjust manner by others are more likely to develop justifications and excuses for crime (e.g., peer abuse justifies violence, chronic unemployment excuses theft). Several studies provide support for these arguments (e.g., Agnew 2006).

2e. *Strains may have both contemporaneous and lagged effects on intervening variables and crime*. The experience of particular strains most often has a contemporaneous effect on intervening variables and crime. For example, imagine that a parent ridicules a child. This ridicule likely has a fairly immediate effect on the child's anger and attachment to the parent, and may lead to delinquent acts in the near term – such as striking the parent or running away from home. But with time, the child's anger and dislike of the parent fades, along with the likelihood of delinquency. In certain cases, however, strains may have lagged as well as contemporaneous effects on delinquency. This is the case with strains that are very high in magnitude, such as sexual abuse, and with chronic or persistent strains. Such strains may lead to negative emotional *traits*, to long-term reductions in the ability to legally cope, to long-term reductions in social control, and to long-term associations with criminal others and beliefs favorable to crime. For example, parents who regularly discipline their children in a harsh manner may lead the children to develop the trait of negative emotionality, may permanently reduce levels of parental attachment, and may foster the belief that violence is an appropriate response to certain problems. As a result, this chronic strain is likely to have both a contemporaneous and lagged effect on delinquency. When longitudinal data are available, researchers should attempt to estimate the contemporaneous effects of strains on crime and, where appropriate, the lagged effects as well.

Factors that Influence the Effect of Strains on Crime (Conditioning Variables)

3. *There are numerous ways to cope with strains, with a range of factors influencing the likelihood of criminal coping*. Criminal coping is more likely among those who (a) have poor conventional coping skills and resources (e.g., poor

problem-solving and social skills, low self-efficacy, limited financial resources), (b) have criminal skills and resources (e.g., criminal self-efficacy), (c) have low levels of conventional social support, (d) are low in social control, (e) associate with other criminals, (f) hold beliefs favorable to crime, (g) have traits conducive to crime (e.g., negative emotionality, low constraint), and (h) are more often exposed to situations where the costs of crime are low and the benefits high (Agnew 2006). Researchers have added to the list of conditioning variables since GST was introduced in 1992, pointing to the importance of environmental variables such as religious involvement and traits such as low self-control (Agnew 2006; Jang and Johnson 2005). The research on conditioning effects, however, has produced mixed results – and this constitutes perhaps the greatest puzzle regarding GST. Agnew (2006) lists several possible reasons for these mixed results, including methodological problems that make it difficult to detect conditioning effects in survey research, problems in the measurement of certain conditioning variables, and the fact that researchers usually consider the conditioning variables in isolation from one another. Criminal coping should be most likely among those who possess all or most of the factors conducive to such coping, an argument that has received some support (Mazerolle and Maahs 2000).

Research is needed to better investigate these possibilities (see Agnew 2006). Also, researchers should attempt to better draw on the bio-psychological research, which suggests that there are substantial differences between individuals in their sensitivity and reaction to stressors. Biological markers for these differences have been identified, and researchers should examine whether they condition the effect of strains on crime (Moffitt et al. 2011). Further, researchers should draw on the coping literature, which describes a range of strategies that individuals employ to cope with stressors (Agnew 2006). Some strategies appear to be more effective than others, although the effectiveness of particular strategies may vary by the type of stress. For example, religious involvement may be an effective way to cope with the death of a family member, but not with peer abuse. Researches should determine if certain strategies are more often linked to crime and, if so, they should examine the factors affecting the use of these strategies. Related to this, researchers should also examine whether past experiences with strains *reduce* the effect of current strains on crime, particularly in cases where the past strains were successfully resolved. Past experiences with strains may improve one's coping skills and increase one's tolerance of current strains.

Extending General Strain Theory

4. *GST can help explain group differences in crime, including socio-demographic, community, and societal differences.* Certain groups have higher crime rates than others because they are (a) more exposed to the criminogenic strains described above; (b) more likely to react to these strains with negative emotions; and/or (c) more likely to cope with these strains and negative emotions through crime, because they differ in their standing on one or more of the conditioning factors listed above. GST has been used to explain gender, race/ethnic, age, social class, community, and societal differences in crime (e.g., Agnew 2006; Bao and Haas 2009; Brezina et al. 2001; De Coster and Zito 2010; Froggio and Agnew 2007; Jang and Lyons 2006; Kaufman et al. 2008; Perez et al. 2008; Simons et al. 2003; Warner and Fowler 2003). Research suggests that GST applies to a range of groups and that it often helps explain group differences in offending. However, it is somewhat difficult to generalize from these studies. They frequently examine different strains, emotions, and conditioning variables; employ different measures; and examine different types of samples. So at present, there is still some uncertainty about how groups differ in the extent and nature of the strains they experience, as well as in their reaction to these strains. More research is needed here. Related to this, criminologists should attempt to link GST with macro-level theories – which can shed light on why and how groups differ in strains and the reaction to them. GST is a social-psychological theory, but it is compatible with a range of

macro-level theories, including most critical theories, feminist theories, and Institutional Anomie Theory (see Agnew 2006). Such theories discuss group differences in the experience of certain strains (e.g., types of gender oppression) and often provide some insight into group differences in the reaction to strain (e.g., gender differences in values and levels of control).

5. *GST can help explain patterns of offending over the life course, including "adolescence-limited" and "life-course-persistent" offending* (see Agnew 2006; Slocum 2010). Offending increases during adolescence because the social environment and biology of adolescents are such that they experience more criminogenic strains and associated negative emotions, and are more likely to cope with these strains/emotions through crime. Some individuals offend at high rates over much of their lives because (a) they possess relatively stable traits that increase their exposure to criminogenic strains and their likelihood of criminal coping; (b) they are part of the urban underclass, which increases their exposure to criminogenic strains and their likelihood of criminal coping, with an amplifying loop being set into motion – their criminal and other negative responses to strains increase the likelihood of further strains; and/or (c) they experience "stress proliferation," wherein the experience of certain strains leads to further strains (e.g., chronic unemployment leads to family conflict). Several studies have found that the level of strain over time is associated with the level of crime, and recent research has begun to test the explanations for the patterns of offending just described (e.g., Agnew 2006; Eitle 2010; Slocum 2010).

6. *GST can help explain situational variations in crime, with such variations partly due to situational variations in strain – particularly provocations*. Most crime research focuses on the factors that create a general predisposition for crime, but even highly predisposed individuals only engage in crime in certain situations. The routine activities perspective has dominated the explanation of situational variations in crime, arguing that crime is most likely in situations where motivated offenders encounter attractive targets in the absence of capable guardians. But qualitative data suggest that crime is also likely in situations where individuals encounter much strain; most notably, violent crime often results when individuals are provoked by others, with such provocations usually involving verbal and physical abuse. And property crime appears more likely in situations where individuals have a desperate need for money (Agnew 2006; Hoffmann 2010). More research is needed here, particularly survey research that builds on the qualitative studies (see Slocum et al. 2005).

7. *GST can explain a range of crimes and deviant acts beyond street crimes, although the theory should be customized to maximize explanatory power*. GST has been applied to the explanation of white-collar and corporate crime, terrorism, states crimes such as genocide, self-harming behaviors such as suicide attempts and eating disorders, cyber-bullying, police deviance, and other criminal/deviant acts (e.g., Agnew et al. 2009; Agnew 2010a; Hay et al. 2010; Maier-Katkin et al. 2009; Piquero et al. 2010). It is often the case that "customized" versions of GST are developed for each type of crime and deviance, with each version pointing to those strains and conditioning variables that are especially relevant. For example, the strains that prompt corporate crime or terrorism differ somewhat from those that prompt street crimes, although many strains are relevant across a range of crimes.

8. *GST can guide efforts to control crime, with such efforts reducing the exposure to criminogenic strains and the likelihood of criminal coping*. A variety of approaches may be employed here, including (a) reducing those strains conducive to crime (e.g., reducing child abuse, raising the minimum wage); (b) altering strains to make them less conducive to crime (increasing the perceived justice of criminal sanctions through the restorative justice approach); (c) removing people from criminogenic strains (e.g., from abusive homes or schools); (d) equipping people with the traits and skills to avoid criminogenic strains (e.g., teaching people to behave in a less provocative manner); (e) altering the perceptions or goals of people to reduce subjective strain (e.g., reducing

the emphasis on consumerism); (f) improving coping skills and resources (e.g., teaching problem-solving skills); (g) increasing social support (e.g., mentoring programs); and (h) reducing the disposition for criminal coping (e.g., altering beliefs conducive to crime). There have not yet been any attempts to use GST to reduce crime, although GST is compatible with a range of successful crime control programs (see Agnew 2006, 2010b).

GST can also be used to shed light on the nature and operation of the criminal justice system. In particular, strains may influence the focus of the criminal justice system (e.g., high levels of strain in the general population, such as unemployment, may contribute to more punitive criminal justice policies). Related to this, strains may influence public attitudes toward criminal justice (e.g., individuals experiencing certain strains may place more emphasis on punitive approaches). Strains may also influence the attitudes, behavior, and effectiveness of law enforcement, court, and corrections workers (e.g., strains may contribute to police deviance and turnover among prison staff). And strains may influence the attitudes and behavior of those processed by the criminal justice system (e.g., prison strains may affect inmate behavior and recidivism). A few recent studies have applied GST to the analysis of the criminal justice system (e.g., Blevins et al. 2010), and hopefully more work will be done in this area.

Conclusion

General strain theory has much support and has established itself as one of the leading theories of crime. In particular, there is much evidence that the strains identified by the theory impact crime and that they do so partly through negative emotions. GST is also increasingly being applied to new issues, such as the explanation of group differences in crime and offending over the life course. At the same time, certain parts of the research on GST have produced mixed results, particularly research on those factors said to condition the effect of strains on crime; several areas of the theory are in need of further exploration; and numerous opportunities for applying the theory to new areas remain.

Related Entries

- ▶ Anomie and Crime
- ▶ Group Characteristics and General Strain Theory
- ▶ Monetary Strain and Individual Offending
- ▶ Poverty, Inequality, and Area Differences in Crime

Recommended Reading and References

Agnew R (1992) Foundation for a general strain theory of crime and delinquency. Criminology 30:47–87

Agnew R (2001) Building on the foundation of general strain theory: specifying the types of strain most likely to lead to crime and delinquency. J Res Crime Delinq 38:319–361

Agnew R (2006) Pressured into crime: an overview of general strain theory. Oxford University Press, New York

Agnew R (2010a) A general strain theory of terrorism. Theor Criminol 14:131–154

Agnew R (2010b) Controlling crime: recommendations from general strain theory. In: Barlow HD, Decker SH (eds) Criminology and public policy: putting theory to work. Temple University Press, Philadelphia

Agnew R, Scheuerman H (2011) Strain theories. Oxford bibliographies online at http://oxfordbibliographiesonline.com/

Agnew R, Piquero NL, Cullen FT (2009) General strain theory and white-collar crime. In: Simpson SS, Weisburd D (eds) The criminology of white-collar crime. Springer, New York

Bao W, Haas A (2009) Social change, life strain, and delinquency among Chinese urban adolescents. Sociol Focus 42:285–305

Baron SW (2008) Street youth, unemployment, and crime: is it that simple? Using general strain theory to untangle the relationship. Can J Criminol Crim Justice 50:399–434

Baron SW (2009) Street youths' violent responses to violent personal, vicarious, and anticipated strain. J Crim Justice 37:442–451

Blevins KR, Listwan S, Cullen FT, Johnson CL (2010) A general strain theory of prison violence and misconduct: an integrated model of inmate behavior. J Contemp Crim Justice 26:148–166

Brezina T, Piquero AR, Mazerolle P (2001) Student anger and aggressive behavior in school: an initial test of Agnew's macro-level strain theory. J Res Crime Delinq 38:362–386
Colvin M (2000) Crime and coercion. St. Martin's Press, New York
De Coster S, Zito RC (2010) Gender and general strain theory: the gendering of emotional experiences and expressions. J Contemp Crim Justice 26:224–245
Eitle DJ (2002) Exploring a source of deviance-producing strain for females: perceived discrimination and general strain theory. J Crim Justice 30:429–442
Eitle D (2010) General strain theory, persistence, and desistence among young adult males. J Crim Justice 38:1113–1121
Froggio G, Agnew R (2007) The relationship between crime and 'objective' versus 'subjective' strains. J Crim Justice 35:81–87
Hay C, Meldrum R, Mann K (2010) Traditional bullying, cyber bullying, and deviance: a general strain theory approach. J Contemp Crim Justice 26:13–147
Hoffmann JP (2010) Contemporary retrospective on general strain theory. In: Copes H, Topalli V (eds) Criminological theory: readings and retrospectives. McGraw Hill, New York
Jang SJ, Johnson BR (2005) Gender, religiosity, and reactions to strain among African Americans. Sociol Q 46:323–357
Kaufman JM, Rebellon CJ, Thaxton S, Agnew R (2008) A general strain theory of racial differences in criminal offending. Aust N Z J Criminol 41:421–437
Maier-Katkin D, Mears DP, Bernard TJ (2009) Toward a criminology of crimes against humanity. Theor Criminol 13:227–255
Mazerolle P, Maahs J (2000) General strain theory and delinquency: an alternative examination of conditioning influences. Justice Q 17:753–778
Moffitt TE, Ross S, Raine A (2011) Crime and biology. In: Wilson JQ, Petersilia J (eds) Crime and public policy. Oxford University Press, New York
Perez DM, Jennings WG, Gover AR (2008) Specifying general strain theory: an ethnically relevant approach. Deviant Behav 29:544–578
Piquero NL, Fox K, Piquero AR, Capowich G, Mazerolle P (2010) Gender, general strain theory, negative emotions, and disordered eating. J Youth Adolesc 39:380–392
Rebellon CJ, Piquero NL, Piquero AR, Thaxton S (2009) Do frustrated economic expectations and objective economic inequity promote crime. Eur J Criminol 6:47–71
Simons RL, Chen Y, Stewart EA, Brody GH (2003) Incidents of discrimination and risk for delinquency: a longitudinal test of strain theory with an African American sample. Justice Q 20:827–854
Slocum LA (2010) General strain theory and the development of stressors and substance use over time: an empirical examination. J Crim Justice 38:1100–1112
Slocum LA, Simpson SS, Smith DA (2005) Strained lives and crime: examining intra-individual variation in strain and offending in a sample of incarcerated women. Criminology 43:827–854
Warner BD, Fowler SK (2003) Strain and violence: testing a general strain theory model of community violence. J Crim Justice 31:511–521

General Theory of Crime

Travis C. Pratt and Jillian J. Turanovic
School of Criminology and Criminal Justice,
Arizona State University, Phoenix, AZ, USA

Synonyms

Low self-control; Self-control theory

Overview

The self-control-crime/deviance link has been well established empirically, with over two decades of studies indicating that self-control is a robust predictor of a host of criminal and analogous behaviors under an equally wide array of methodological conditions. This pattern appears to be such a "given" that the field has largely moved on to other areas of self-control research, such as assessing the other harmful consequences of self-control, like criminal victimization, to testing the degree to which self-control is or is not stable within individuals over time, and to examining the "causes" of low self-control. This entry takes stock of these more recent developments within the self-control tradition.

Introduction

Gottfredson and Hirschi's (1990) self-control theory has stimulated a considerable amount of research and discussion regarding the influence of low self-control on criminal and analogous behaviors. As originally formulated, Gottfredson and Hirschi's (1990) self-control theory predicts that individuals with low self-control are more predisposed to engage in a host of criminal and analogous acts. This is because those lacking self-control tend to pursue their own self-interest without consideration of the potential long-term consequences of their behavior. Criminal behavior therefore becomes attractive to those with low self-control because it can be exciting and immediately gratifying.

The self-control-crime/deviance link has been well established empirically, with over two decades of studies indicating that self-control is a robust predictor of a host of criminal and analogous behaviors under an equally wide array of methodological conditions (Pratt 2009; Pratt and Cullen 2000). To be sure, research has revealed that those with low self-control are significantly more likely to be a general offender, a versatile offender, a cheater on exams, a software pirate, and even a drunk-dialing user of profanity in public spaces (see Pratt and Cullen 2000; Reisig and Pratt 2011). These findings are sufficiently robust that criminologists have turned their attention to other areas of self-control research, such as the link between self-control and criminal victimization, the stability of self-control over time, and to the biological, familial, and contextual sources of low self-control. Each of these more recent developments is discussed here.

Self-Control and Victimization

It has been well established that, above all, individuals with low self-control are thrill seekers. In the pursuit of self-pleasure, such individuals feel the need to frequently engage in dangerous, reckless, and risky behaviors that provide them with a sense of excitement. Such individuals are unlikely to consider how their behavior may impact others nor would they be likely to spend time contemplating how engaging in such activities might put them at risk for numerous negative outcomes – one of the most serious of which may be victimization. Indeed, a large body of work has demonstrated that those who engage in deviance are more likely to be victims of crime, even when the degree of deviance is minor (e.g., Lauritsen et al. 1991; Turanovic and Pratt 2013a). Researchers have recently expanded upon these conclusions and have started to explore the relationship between self-control and victimization.

Traditionally, the study of victimization has been limited to situational approaches that emphasize the role of opportunity, such as routine activity (Cohen and Felson 1979) and lifestyle theories (Hindelang et al. 1978). Routine activity explanations suggest that individual patterns of behavior increase victimization through the space-time convergence of likely offenders, suitable targets, and the absence of capable guardians, while lifestyle theory similarly puts forth that individuals may engage in daily behaviors that expose them to crime and risky circumstances. In the past decade or so, scholars have broadened the scope of this inquiry by exploring other avenues that may explain victimization risk and looked further into why certain individuals are more likely to find themselves in high-risk situations. In particular, Schreck (1999) extended and reformulated Gottfredson and Hirschi's (1990) conceptualization of self-control into a theory of vulnerability. Schreck argued that the processes that determine individuals' lifestyle choices and exposure to risk could be explained by their levels of self-control. Put simply, because individuals with low self-control are shortsighted and take part in impulsive, unsafe behaviors, they may differentially place themselves in dangerous situations and might be less likely to take the precautions necessary to avoid being a victim of crime.

Specifically, Schreck suggested that each dimension of low self-control – lack of future orientation, risk taking, lack of empathy, low tolerance for frustration, lack of diligence, and preference for physical over mental activity – has the potential to increase individuals' risks of being victimized. In short, each element of low self-control is associated with an aspect of vulnerability, increasing the attractiveness of these individuals (and their belongings) to motivated and opportunistic offenders. For example, since individuals with low self-control tend to have a low tolerance for frustration and are quick to anger, Schreck argued that these individuals may be more likely to behave in a belligerent manner that can provoke a criminal attack. And since those with low self-control prefer physical over mental tasks, these individuals may be less likely to use their cognitive skills to assess and diffuse hostility. Instead, they may be more likely to react to threat in a defensive or aggressive manner, which may make the situation worse. Having low self-control may also seriously impede a person's ability to carefully and consistently protect his or her belongings. Due to shortsightedness and lack of diligence, such individuals may chronically fail to set alarms, lock doors, and become impatient with procedures associated with arming complex security devices.

Schreck's (1999) extension of self-control theory has received considerable empirical support, both in predicting violent and nonviolent victimization, especially among individuals who also engage in offending. While this area of research is still relatively new, existing studies have consistently found low self-control to be among the most important predictors of criminal victimization – a finding which has held up across a variety of contexts, even when controlling for other robust criminogenic risk factors. Specifically, empirical examinations have demonstrated that self-control increases vulnerability to personal, property, violent, and sexual victimization, as well as the likelihood of fraud and internet theft (Franklin 2011; Holtfreter et al. 2008; Schreck et al. 2006). It is important to note, however, that while self-control can strongly predict victimization in many forms, it has not been found to fully moderate the effects of routine activity/lifestyle variables on victimization. Indeed, self-control has been found to exert both direct *and* indirect effects on victimization. For instance, associating with deviant peers and engaging in "risky lifestyles" (such as offending, substance use, risky sexual behaviors) have also been found to increase the probability of victimization, independent of one's level of self-control (Franklin 2011; Schreck et al. 2006; Turanovic and Pratt 2013b). What these findings have brought to light is that neither self-control nor routine activity/lifestyle perspectives *alone* can fully account for victimization. As a result, scholars have been urged to integrate self-control and opportunity theories together in order to develop a more accurate and comprehensive understanding of victimization and vulnerability (see Holtfreter et al. 2008).

While empirical findings demonstrate the importance of combining self-control and routine activity/risky lifestyle perspectives to predict victimization, there has been substantial variation in how these routine behaviors are captured – most likely because a vast array of actions can potentially be deemed as "risky." For example, in their assessment of low self-control on disaggregated personal and property victimization, Schreck et al. (2006) conceptualized risky lifestyles as violent, property, and drug delinquency in addition to having delinquent friends, while Franklin (2011) incorporated such acts as risky sexual behaviors, illegal drug use, alcohol consumption, and pornography consumption in her assessment of self-control on violent sexual victimization. It is important for scholars to begin to systematically identify which types of risky lifestyles put those with low self-control most at risk for particular forms of victimization.

Because this area of research is still developing, relatively little is known about the effects of low self-control on types of victimization that are largely directed towards women, such as sexual assault and domestic violence. To be sure of the aforementioned studies conducted since Schreck (1999) proposed his vulnerability hypothesis, only a few controlled for gender (e.g., Holtfreter et al. 2008; Schreck et al. 2006), and less have looked at the impact of self-control on the victimization of women independent of men (e.g., Franklin 2011). While the research is in short supply, the findings of these studies are promising enough to invite further explorations into how self-control influences the unique victimization experiences of women.

Additionally, as research in this area progresses, scholars will likely devote more attention to the longitudinal consequences of self-control on victimization. Schreck et al. (2006) demonstrated that those with low self-control are more likely to experience repeated victimization, but the roles of routine activity/lifestyle factors in this process need to be explored further. Those with low self-control may be less willing, and less able, to restructure their "risky" behavioral routines following an initial victimization incident. For some, victimization may be a mere price to pay for the pleasure of engaging in deviant activities. All told, self-control is not only paramount to the study of criminality – it is also gaining support as a vital component to the study of victimization. While findings have been largely consistent in demonstrating that self-control is a robust predictor of victimization, additional research is still needed to help strengthen and clarify the preliminary work in this area.

The Stability (or Not) of Self-Control

One of the more controversial propositions regarding the general theory of crime concerns the "stability thesis" specified by Gottfredson and Hirschi (1990). They argued that self-control is developed early in childhood, usually by around the age of 10 or so, primarily through parental socialization efforts, and is relatively stable throughout the life course. This proposition is potentially problematic, of course, because of the well-known "age-crime curve," where individual rates of criminal behavior generally rise and peak in the late teen years and decline steeply from the early 20s and beyond. How, then, can the key "cause" of crime remain stable within individuals even though their participation in criminal behavior can vary considerably over time?

Scholars have examined this issue longitudinally with mixed results – where some studies indicate high levels of stability in self-control over time, at least in the short term (e.g., Beaver and Wright 2007), others have found evidence of more moderate correlations between measures of self-control taken at repeated points in time (e.g., Burt et al. 2006). Perhaps most telling is Hay and Forrest's (2006) analysis of data from the National Longitudinal Survey of Youth (NLSY) – a nationally representative longitudinal dataset. While their analyses revealed fairly strong stability in self-control overall, there was a nontrivial portion of the sample (16 %) whose levels of self-control changed substantially between ages 7 and 15.

Of course, an argument could be made that Gottfredson and Hirschi (1990) were really only referring to "relative" stability in this context; that is, the relative ranking between individuals should remain fixed over time even if absolute levels of self-control can change with age. Hirschi and Gottfredson (2001: 90) later clarified this point when they stated that "the *differences* [in self-control] observed at ages 8 to 10 tend to persist. . .Good children remain good. Not so good children remain a source of concern to their parents, teachers, and eventually to the criminal justice system" (emphasis added). Even so, Gottfredson and Hirschi (1990) were never explicit about *why* absolute levels of self-control would increase with age or what would explain the variation between individuals concerning how much or how little self-control would be gained over time.

A more fundamental problem, however, might have to do with the conceptualization of self-control itself (see, e.g., the discussion by Pratt 2009). In particular, criminologists' focus on individuals' *levels* of self-control has caused them to miss another key component within this theoretical tradition: within-individual *variability* in self-control. Indeed, recent research regarding changes in self-control over time (Hay et al. 2006), the disaggregation of the desire to exercise self-control versus the ability to exercise self-control (Tittle et al. 2004), Hirschi's (2004) own recent revision of self-control theory, and especially the recent research on the "depletion" of self-control under stressful conditions (Muravin et al. 2006) all hint at – to a greater or lesser degree – the importance of within-individual variation in self-control from one situation to the next. Since research into this question has only begun to emerge, it appears that the "stability thesis" of self-control is far from settled.

Part of the problem is that getting to this question empirically is no easy task. It requires either one of two things, the first of which would be longitudinal data with uniform indicators of self-control measured repeatedly over time, so one could assess directly changes that do or do not occur. In this scenario, it would be necessary to have a wide age range; that is, knowing how self-control changes (either absolutely or relatively) between, say, age 13 and 50, would be more empirically and theoretically important than knowing how it changes from age 13 to 15. A second approach would entail the use of experimental data where situational characteristics could be manipulated to induce self-control depletion under a variety of contextual conditions *that might actually occur in "the real world"* (e.g., conditions of economic deprivation and the reality of human interactions that follow from it, including the presence of, e.g., cultural values associated with the "code of the street" as identified). Neither of these approaches come cheap or easy. Nevertheless, pursuing them may be the next necessary step in self-control-crime research. And if additional research uncovers that self-control is, in fact, much more fluid and malleable over the life course than Gottfredson and Hirschi (1990) originally contended, the theory itself may need to be revised.

The Causes of Self-Control

While the link between self-control and crime/deviance has been consistently demonstrated empirically, what is less clear at this point is how self-control is established within individuals. The primary explanation regarding the genesis of self-control in the criminological literature is Gottfredson and Hirschi's (1990) parenting thesis. In short, Gottfredson and Hirschi contend that self-control will develop in children through effective parenting, where parents who monitor their kids' behavior recognize deviant behavior when it happens and punish such behavior consistently will produce in their children the internal control mechanisms necessary for resisting the temptations that criminal and deviant behavior provide.

Support for this proposition is certainly present. Polakowski's (1994) analysis of data from the Cambridge Youth Study; Feldman and Weinberger's (1994) assessment of 81 sixth-grade boys; the student samples analyzed by Cochran et al. (1998) and by Gibbs et al. (1998); and Hay's (2001) survey of 197 urban high school youth have all explored the dynamics of parenting and self-control. Others have followed suit as well (see, e.g., Pratt et al. 2004; Turner et al. 2005; Unnever et al. 2003), with Perrone et al. (2004) analysis of the data from the first wave of the National Longitudinal Study of Adolescent Health (a nationally representative sample of over 13,000 youth) providing some of the most convincing evidence. Indeed, with the exception of Cochran et al. (1998) study of self-control and academic dishonesty, the research conducted thus far generally lends credence to the notion that, net of statistical controls, parental efficacy is important to the process of developing self-control in children.

Nevertheless, empirical evidence has emerged indicating that the processes that establish individuals' levels of self-control are more complex than those specified by Gottfredson and Hirschi. Specifically, research has begun to emerge that examines alternative sources of low self-control. For example, research has found that indicators of biological predisposition (e.g., ADHD, indicators of neuropsychological deficits such as low birth weight and low cognitive ability; please see Kevin Beaver's contribution to this volume concerning the biological and genetic sources of low self-control) are significantly related to levels of self-control independent of measures of effective parenting (McGloin et al. 2006; Unnever et al. 2003). In addition, controls for such biological/neuropsychological factors tend to partially mediate – and in some cases fully mediate – the effect of parenting on the development of self-control (see, e.g., Wright and Beaver 2005). Taken together, this research indicates that certain biological and neuropsychological risk factors need to be considered in the formation of self-control.

Furthermore, criminologists have begun to focus on how different types of neighborhoods influence parenting behavior and, in turn, the development of self-control in children. The first study in this tradition was Pratt et al. (2004) analysis using data drawn from the National Longitudinal Survey of Youth (NLSY), which found that conditions of neighborhood deprivation significantly influenced measures of parental monitoring and socialization. Furthermore, such neighborhood conditions directly affected the development of self-control in children independent of measures of parental efficacy. A subsequent study by Hay et al. (2006) went a step further and found a significant interaction term between neighborhood conditions and parental efficacy on the development of self-control. As such, this work clearly indicates that community context is yet another factor that must be seriously considered by scholars with regard to the development of self-control in children.

Finally, Gottfredson and Hirschi (1990: 105) suggested that "like the family, the school in theory has the authority and the means to punish lapses in self-control." And as Denise Gottfredson (2001: 48) also observed, "schools have the potential to teach self-control and to engage informal social controls to hold youthful behavior in check." Empirical work has recently emerged that has tested these various propositions. Turner et al. (2005) analysis of the NLSY data revealed two conclusions along these lines. First, indicators of "school socialization" (which closely resembled typical parenting measures associated with the monitoring and supervision of children) were significantly related to the development of self-control independent of parental efficacy. Second, the effects of school socialization on youths' levels of self-control varied according to (i.e., interacted with) levels of parental efficacy, as well as conditions of neighborhood deprivation. In particular, the effect of school socialization on children's development of self-control was strongest when parental efficacy was low and when neighborhood conditions were criminogenic. These results therefore highlight the ability of social institutions – in this case the school – to "pick up the slack" for instilling self-control in children when other mechanisms, such as parents and the community, break down. Put simply, based on the body of empirical research presented above, it is clear that the causes of how and why self-control develops within individuals are far more complex than the simple parenting thesis offered by Gottfredson and Hirschi (1990).

Conclusions

When it comes to Gottfredson and Hirschi's (1990) general theory of crime, there is a lot that we know. For example, we know that the core theoretical proposition specified by Gottfredson and Hirschi – that a wide array of criminal and deviant behaviors are the result of low self-control – is extremely well supported in the criminological literature. Yet in the process of laying out why low self-control should lead to criminal behavior, they also made a number of corollary assumptions to support the self-control-crime thesis, such as those assumptions surrounding the link between self-control and victimization, the stability thesis, and the cause of low self-control.

Accordingly, we seem to know less about these corollary assumptions. On the one hand, it is true that there is convincing evidence of a link between self-control and victimization, that self-control is somewhat stable over the life course, and that parenting "matters" when it comes to instilling self-control in children. On the other hand, there seems to be equally convincing evidence that self-control is quite fluid and flexible within individuals and that the sources of self-control are far more varied and complex than Gottfredson and Hirschi said they should be. On balance, however, they were certainly on the right track.

And there are, of course, a number of other corollary assumptions regarding the general theory of crime that still remain relatively unaddressed. For example, Gottfredson and Hirschi also noted that self-control should predict crime and deviance equally well across various populations and subpopulations of people and that the relationship between self-control and problematic behavior should be roughly similar across racial and gender categories (known as the "invariance thesis") and that one's level of self-control should influence heavily the nature of one's peer group. Aside from a small handful of recent studies, none of these propositions have been adequately addressed in the literature, so how they end up shaking out for Gottfredson and Hirschi remains to be seen. Even so, from what we do know about the nature and consequences of self-control, it appears that the empirical attention and respect that have been afforded Gottfredson and Hirschi's theory is certainly deserved.

Related Entries

▶ Genetic Basis to Self-Control
▶ Measurement of Self-Control
▶ Social Control and Self-Control Through the Life Course

Recommended Reading and References

Beaver KM, Wright JP (2007) Stability of low self-control from Kindergarten through first grade. J Crime Justice 30:63–86

Burt CH, Simons RL, Simons LG (2006) A longitudinal test of the effects of parenting and the stability of self-control: negative evidence for the general theory of crime. Criminology 44:353–396

Cochran JK, Wood PB, Sellers CS, Wilkerson W, Chamlin MB (1998) Academic dishonesty and low self-control: an empirical test of a general theory of crime. Deviant Behav 19:227–255

Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44:588–608

Feldman SS, Weinberger DA (1994) Self-restraint as a mediator of family influences on boys' delinquent behavior: a longitudinal study. Child Dev 65:195–211

Franklin CA (2011) An investigation of the relationship between self-control and alcohol induced sexual assault victimization. Crim Justice Behav 38:263–285

Gibbs JJ, Giever D, Martin JS (1998) Parental management and self-control: an empirical test of Gottfredson and Hirschi's general theory. J Res Crime Delinq 35:40–70

Gottfredson D (2001) Schools and delinquency. Cambridge University Press, Cambridge

Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Palo Alto

Hay C (2001) Parenting, self-control, and delinquency: a test of self-control theory. Criminology 39:707–736

Hay C, Forrest W (2006) The development of self-control: examining self-control theory's stability thesis. Criminology 44:739–774

Hay C, Fortson E, Hollist D, Altheimer I, Schaible L (2006) The Impact of community disadvantage on the relationship between the family and juvenile crime. J Res Crime Delinq 43:326–356

Hindelang MJ, Gottfredson MR, Garofalo J (1978) Victims of personal crime. Ballinger, Cambridge

Hirschi T (2004) Self-control and crime. In: Baumeister RF, Vohs KD (eds) Handbook of self-regulation: research, theory, and application. Guilford, New York

Hirschi T, Gottfredson MR (2001) Self-control theory. In: Paternoster R, Bachman R (eds) Explaining criminals and crime. Roxbury, Los Angeles

Holtfreter K, Reisig MD, Pratt TC (2008) Low self-control, routine activities, and fraud victimization. Criminology 46:189–220

Lauritsen JL, Sampson RJ, Laub JH (1991) The link between offending and victimization among adolescents. Criminology 29:265–292

McGloin JM, Pratt TC, Piquero AR (2006) A life-course analysis of the criminogenic effects of maternal cigarette smoking during pregnancy: a research note on the mediating impact of neuropsychological deficit. J Res Crime Delinq 43:412–426

Muravin M, Pogarsky G, Shmueli D (2006) Self-control depletion and the general theory of crime. J Quant Criminol 22:263–277

Perrone D, Sullivan C, Pratt TC, Margaryan S (2004) Parental efficacy, self control, and delinquent behavior: a test of a general theory of crime on a nationally representative sample. Int J Offender Ther Comp Criminol 48:298–312

Polakowski M (1994) Linking self- and social control with deviance: illuminating the structure underlying a general theory of crime and its relation to deviant identity. J Quant Criminol 10:41–78

Pratt TC (2009) Reconsidering Gottfredson and Hirschi's general theory of crime: linking the micro- and macro-level sources of self-control and criminal behavior over the life course. In: Savage J (ed) The development of persistent criminality. Oxford University Press, New York

Pratt TC, Cullen FT (2000) The empirical status of Gottfredson and Hirschi's general theory of crime: a meta-analysis. Criminology 38:931–964

Pratt TC, Turner MG, Piquero AR (2004) Parental socialization and community context: a longitudinal analysis of the structural sources of low self-control. J Res Crime Delinq 41:219–243

Reisig MD, Pratt TC (2011) Low self-control and imprudent behavior revisited. Deviant Behav 32:589–625

Schreck CJ (1999) Criminal victimization and low self-control: an extension and test of a general theory of crime. Justice Q 16:633–654

Schreck CJ, Stewart EA, Fisher B (2006) Self-control, victimization, and their influence on risky lifestyles:

a longitudinal analysis using panel data. J Quant Criminol 22:319–340

Tittle CR, Ward DA, Grasmick HG (2004) Capacity for self-control and individuals' interest in exercising self-control. J Quant Criminol 20:143–172

Turanovic JJ, Pratt TC (2013a) The consequences of maladaptive coping: integrating general strain and self-control theories to specify a causal pathway between victimization and offending. J Quant Criminol (onlinefirst)

Turanovic JJ, Pratt TC (2013b) "Can't stop, won't stop": self-control, risky lifestyles, and repeat victimization. J Quant Criminol (onlinefirst)

Turner MG, Piquero AR, Pratt TC (2005) The school context as a source of self-control. J Crim Justice 33:327–339

Unnever JD, Cullen FT, Pratt TC (2003) Parental management, ADHD, and delinquent involvement: reassessing Gottfredson and Hirschi's general theory. Justice Q 20:471–500

Wright JP, Beaver KM (2005) Do parents matter in creating self-control in their children? A genetically-informed test of Gottfredson and Hirschi's theory of low self-control. Criminology 43:1169–1202

Genes, Crime, and Antisocial Behaviors

Kevin M. Beaver[1,2] and Joseph A. Schwartz[1]
[1]School of Criminology and Criminal Justice, Florida State University, Tallahassee, FL, USA
[2]Center for Social and Humanities Research, King Abdulaziz University, Jeddah, Saudi Arabia

Overview

When it comes to explaining the causes of crime, criminals, and antisocial behavior, there is virtually an endless supply of ideas. Some of these explanations are based on mounds of data, such as theories that focus on the assumption that antisocial behavior is produced by peer pressure or exposure to maltreatment early in life. Other explanations, however, teeter on the absurd, such as the belief that all crime is due to the pursuit of economic gain. The point is that virtually any factor that could be linked to crime has, in some way, been spun into a criminological theory. A perusal of any introduction to criminology textbook is quite revealing by showing the tremendous breadth of factors and explanations that have been advanced to explain crime and that have yet to be falsified by the academic community. Interestingly, the factors that do not seem to get very much coverage by criminologists – at least not in a balanced format – are genetic factors. In fact, most criminologists argue that genetic factors have absolutely nothing to do with the development of criminal involvement and that only environmental factors matter (Wright et al. 2008a). Recent empirical-based research has called into question the assumption that genes have no effect on criminal involvement and instead have revealed that genetic factors are the dominant etiological influence for crime and antisocial behaviors (Ferguson 2010; Moffitt 2005; Rhee and Waldman 2002).

Genetic Effects and Antisocial Behaviors

One of the major obstacles to studying genetic effects is trying to provide accurate and reliable estimates of both genetic and environmental effects on crime and antisocial behaviors. One of the more common ways of accomplishing this goal is by making use of a naturally occurring experiment: twinning. There are two types of twins: monozygotic (MZ) twins and dizygotic (DZ) twins. MZ twins share 100 % of their DNA (i.e., they are essentially genetic clones of each other), and they also share the same environmental upbringings. DZ twins, in contrast, share only 50 % of their distinguishing DNA, but they too share the same environmental upbringings. By comparing the similarity on measures of crime/antisocial behavior of MZ twins to the similarity of DZ twins, it is possible to quantify the proportion of variance that is the result of genetic factors and the proportion of variance that is the result of environmental factors. The more similar MZ twins are to each other in comparison with DZ twins, the greater the genetic effect. Why? – since MZ twins and DZ twins are exposed to similar environments, the only reason that MZ twins should be more similar to each other than DZ twins is because they share twice as much genetic material.

Twin-based research thus provides an estimate of both genetic and environmental effects on variance in antisocial behaviors (Beaver 2008). Technically speaking, the proportion of variance accounted for by genetic factors is known as heritability. Unlike social science research that treats all environments as the same, twin-based research has helped to make the distinction between two types of environments: shared environments and nonshared environments. Shared environments are environments that are the same between twins (or siblings) and that systematically increase their similarity. Nonshared environments consist of any environmental factors that are different between twins and that make them dissimilar (Beaver 2008). For example, as it applies to crime, shared environments might include such factors as being reared in poverty or living in a disadvantaged neighborhood. Examples of nonshared environments might consist of factors such as having different peer groups or being treated differently by parents. When summed together, the effects of heritability, shared environmental factors, and nonshared environmental factors explain 100 % of the variance (Beaver 2009).

There has been a great deal of research using twin-based studies to estimate genetic, shared environmental, and nonshared environmental effects on an assortment of antisocial behaviors, such as crime, violence, aggression, and other types of antisocial behaviors (Beaver 2008; 2011; Miles and Carey 1997). The results of these individual-level studies tend to provide slightly different point estimates of heritability, shared environmental effects, and nonshared environmental effects depending on the sample analyzed and the precise measure of antisocial behavior investigated. When aggregated and averaged together, these studies tend to suggest that genetic factors explain about 50 % of the variance in antisocial behaviors, shared environmental factors explain between 0 % and 10 % of the variance, and nonshared environmental factors explain about 40 % of the variance (Beaver 2009; Ferguson 2010; Moffitt 2005). Instead of genetic factors having no effect like most criminologists assume, genetic factors appear to have the dominant effect on antisocial behaviors.

While twin-based research has been quite valuable in the quest for estimating genetic and environmental effects on antisocial behaviors, there are some limitations to this research design. Opponents of genetic research argue that these limitations are fatal flaws that bias the results and thus any findings generated from twin-based studies are not believable. Fortunately, there are a number of other research designs that can also be used to estimate genetic and environmental effects (Beaver 2009). These alternative research designs, moreover, are not host to the same limitations as twin-based studies and thus they can act as “checks” on the twin-based research designs. As long as the genetic and environmental estimates are relatively consistent across all research designs, then the likelihood that the results are being affected by limitations is low. However, if the genetic and environmental effects vary significantly across different research designs, then this pattern of findings would tend to indicate that the results are being systematically affected by the choice of research designs.

One of the alternative research designs that has been used in place of the traditional twin-based research design capitalizes on a relatively rare occurrence wherein MZ twins were separated at birth, adopted by different families, and raised without even knowing they had a co-twin (known as MZAs [monozygotic twins who were reared apart]). Later in life, they are often told by family and friends or discover through serendipitous events that they were born as part of a twin pair. Upon reuniting, researchers are presented with one of the most effective ways to estimate genetic effects. Since MZAs were reared in different environments and by different families, the only reason that MZAs can be similar to each other is because they share 100 % of their DNA. MZAs obviously are quite rare, but scientists at the University of Minnesota have established the Minnesota Study of Identical Twins Reared Apart (MISTRA) and have located more than 100 pairs of MZAs to interview and study (Bouchard et al. 1990). Upon learning of an MZA, investigators on the MISTRA project invite the twins to participate in their study where the twins undergo a weeklong series of tests designed to measure virtually every human characteristic. Overall, the results of the MZA studies have converged with those of traditional twin-based studies, indicating that genes are highly influential when it comes to crime, aggression, violence, and other antisocial characteristics (Beaver 2009).

Even though twin-based studies and MZA studies reveal results that are relatively consistent, there are still those opponents to genetic research that argue that MZA studies also are host to a number of different flaws. According to such critics, the genetic effects reported in MZA studies are also not believable. Luckily, there is yet another way to estimate genetic effects on antisocial behaviors: adoption studies (Beaver 2009). Adoption studies represent another accurate way to estimate genetic and environmental effects on behaviors and traits. To do so, adoptees are compared to their biological parents (with whom they have had little to no contact) and their adoptive parents (with whom they share no genetic material). The only reason that adoptees should resemble their biological parents on measures of antisocial behaviors is because of the genetic material they share with them. And, the only reason that adoptees should resemble their adoptive parents on measures of antisocial behaviors is because of the environment that they share with them. As with twin-based studies and MZA studies, adoption-based studies have revealed that crime and other types of antisocial behaviors are affected by genetic factors (Beaver 2010; Rhee and Waldman 2002).

Molecular Genetics and Antisocial Behaviors

When the results culled from twin-based studies and adoption-based studies are viewed simultaneously, they paint a very detailed and accurate picture indicating that crime, delinquency, and antisocial behavior in general are influenced in large part by genetic factors. While these studies have provided very compelling evidence establishing the genetic foundations to criminal involvement, they do not provide any information as to the specific genes that are involved in creating crime. The next logical step, therefore, is to uncover the particular genes that are associated with crime and antisocial behaviors. In order to do so, it is first necessary to be exposed to some of the basics of molecular genetics.

Genes are inherited on 23 pairs of threadlike structures called chromosomes. One pair of chromosomes is inherited maternally and the other pair is inherited paternally. Of these 23 pairs of chromosomes, there are 22 pairs of autosomes and one pair of sex chromosomes. For all genes located on the autosomes, there are two copies (one on the maternal chromosome and one on the paternal chromosome). When it comes to the sex chromosomes, there is a slightly different pattern of inheritance. Females have two X chromosomes and thus they have two copies of all genes located on the X chromosomes. Males, however, have one X chromosome (always inherited maternally) and one Y chromosome (always inherited paternally). As a result, males have only one copy of each gene located on the X chromosome and one copy of each gene located on the Y chromosome.

Most of the genes in humans do not vary from person to person which is why, structurally and anatomically speaking, humans look very similar to each other (e.g., two arms, two eyes, two legs, a heart). A small percentage of genes, somewhere around 1–10 % depending on how genetic differences are measured, do vary from person to person (Beaver 2009). Genes that can vary from person to person are known as genetic polymorphisms. For example, genes that are involved in creating height are genetic polymorphisms, which is what produces variation in the height of humans. Alternative versions of genetic polymorphisms are known as alleles (e.g., alleles that make someone tall or alleles that make someone short). Criminologists interested in examining the molecular genetic basis to crime and other antisocial behaviors focus on genetic polymorphisms because only genes that vary can explain variation in antisocial behaviors and traits (otherwise it would be analogous to trying to explain a variable (i.e., crime) with a constant).

Historically there has been much confusion among criminologists and social scientists over how genes ultimately affect criminal behaviors (Wright et al. 2008b). The overarching belief is that there is a single gene that decides whether someone will become a criminal; persons who have this gene will always become a criminal and persons lacking this gene will never become a criminal. This type of deterministic thinking is preposterous. Complex behaviors like criminal involvement are multifactorial meaning that they are produced by a combination of genetic and environmental factors. Moreover, criminal and antisocial behaviors are considered polygenic phenotypes. Polygenic phenotypes refer to human behaviors and traits that are affected by many genes, with each gene having only a small effect on the propensity to engage in crime. Seen in this way, genes work in a probabilistic way, where the presence of certain alleles increase or decrease the probability of an antisocial outcome (Beaver 2009). Each genetic variant, however, would only increase the odds of a crime and antisocial behavior by a relatively small margin, typically less than 10 %. Clearly there is nothing deterministic with this contemporary view of the ways in which genetic factors ultimately affect crime, delinquency, and antisocial behaviors.

During the past decade, there has been a considerable amount of molecular genetic research examining whether the alleles of certain genes predispose someone to engage in crime and violence. Although the results of this line of research have not always been entirely consistent, the main theme running across these studies is that genes that are involved in neurotransmission are the genes most likely to affect criminal involvement. Neurotransmission refers to the process by which neurons communicate with each other. Neurons are brain cells and in order for information to be processed across neuronal networks, adjacent neurons must communicate with each other to pass information from neuron to neuron. Neurons, however, are not physically wired together as there is a small gap – known as a synapse – that exists between neurons. So transmitting information between neurons requires that the synapse be bridged in some capacity. This is accomplished with neurotransmitters, such as serotonin and dopamine, which are chemical messengers that are released from the presynaptic neuron where they cross the synapse and lock into receptors on the postsynaptic neuron.

After the neurotransmitter has locked into the postsynaptic neuron, the neurotransmitters need to be removed from the synapse. There are two main ways that neurotransmitters are purged from the synapse. The first is through a process called reuptake. With reuptake, transporter proteins are released into the synapse where they seek out neurotransmitters, remove them from the synapse, and return them to the vesicles of the presynaptic neuron. The second way that neurotransmitters are removed is through enzymes that degrade neurotransmitters into inactive particles that are then flushed from the synapse. Both reuptake and enzymatic degradation work simultaneously to regulate levels of neurotransmitters. If something interferes with either of these processes or if either of these two processes is not working efficiently, then levels of neurotransmitters may deviate from normality. Importantly, there is a good deal of research linking variation in neurotransmitter levels to a range of psychopathologies, including aggression, violence, suicide, and crime (Brunner et al. 1993; Beaver et al. 2007).

Genes are involved in coding for the production of transporter proteins that are central to the process of reuptake. This is especially important because some genetic polymorphisms related to transporter proteins are functional, meaning that different alleles correspond to differences in the activity level of the transporter protein. For example, different alleles of a genetic polymorphism that codes for the production of the serotonin transporter protein (5HTTLPR) are associated with different transcriptional efficiencies (functional differences that may ultimately produce different levels of serotonin). Similarly, a gene that codes for the production of the monoamine oxidase A (MAOA) enzyme, which is responsible for breaking down neurotransmitters, also has a functional polymorphism (in the promoter region of the gene). Some alleles of this MAOA polymorphism code for the production of low MAOA activity, meaning that the MAOA is not as efficient at mopping up neurotransmitters from the synapse, while other alleles code for the production of high MAOA activity, which is much more efficient at removing neurotransmitters from the synapse.

Overall, molecular genetic research has provided some evidence tying variants of certain genes, such as dopaminergic genes (e.g., DAT1, DRD2, and DRD4), serotonergic genes (e.g., 5HTTLPR), and genes coding for the production of enzymes (e.g., MAOA and COMT) to various antisocial behaviors (Beaver et al. 2007; Caspi et al. 2002; Brody et al. 2011). These genes tend to have small effects, a finding which is consistent with polygenic explanations of human behavior. Perhaps partially as a result of the small effects associated with genes, genetic association studies are often plagued by an inability to replicate the original results. What this means is that after a study first reports a statistically significant association between a genetic polymorphism and an outcome, replication studies often fail to detect that same association in independent samples. There are a number of potential explanations for a failure to replicate, including the possibility that the original report was a methodological or statistical artifact. Consequentially, replication studies are of utmost importance when attempting to figure out whether a putative candidate gene is indeed related, perhaps causally, to a human trait/behavior.

Gene-Environment Interplay and Antisocial Behaviors

Although molecular genetic research has revealed that some genes are related to a variety of antisocial outcomes, the most cutting-edge genetic research examines the interplay between genetic factors and environmental factors (Moffitt 2005). There are two main types of gene-environment interplay that have been tied to antisocial behaviors: gene-environment interactions and gene-environment correlations. Gene-environment interactions capture the processes by which genetic effects are moderated by environmental factors and/or by which environmental effects are moderated by genetic factors. In short, gene-environment interactions can explain why two people who encounter the same stimuli may turn out quite differently. An example will help to clarify what is meant by a gene-environment interaction. Suppose that a study was examining a group of adolescents who were raised in a disadvantaged neighborhood with high rates of crime, poverty, and homelessness. Exposure to this neighborhood is certainly a criminogenic risk factor that heightens the propensity for engaging in crime and disrepute. Even so, most of the adolescents exposed to such conditions will not turn out to be criminal, but a handful of adolescents will develop into criminals. The million-dollar question, of course, is what accounts for these differential outcomes? The logic of gene-environment interactions can easily answer this question by drawing attention to the fact that for a criminal to develop they must (1) be exposed to a criminogenic environment *and* (2) have a sufficient genetic predisposition for crime. If either of these two factors is absent, then their independent effects are either muted or attenuated significantly. To summarize, then, gene-environment interactions refer to the fact that genetic effects are strongest when paired with environmental liabilities (and vice versa).

Much of the contemporary genetic research examining genetic effects on antisocial behavior and psychopathologies has been guided by gene-environment interactions. What this line of research has revealed is that the effects of specific polymorphisms on antisocial behaviors are structured in part by exposure to environmental liabilities and stressors. For example, the MAOA gene has consistently been found to be related to violence and aggression in males (Caspi et al. 2002; Haberstick et al. 2007; Kim-Cohen et al. 2006). Consistent with the logic of gene-environment interactions, however, the effect of MAOA only surfaces among males who have been abused and maltreated as children. For males who have not been abused as children, there is no effect of MAOA on antisocial behaviors. Similar findings have been observed with other genetic polymorphisms linked to antisocial propensities. These findings underscore the mutual interdependence of genes and the environment when trying to understand the etiology of criminal and antisocial behaviors.

Gene-environment interactions may also be integral to understanding why there are some problems with replicating molecular genetic findings. Molecular genetic research is often based on clinical, non-nationally representative samples. This necessarily means that the samples that are used to test for genetic associations are differentially exposed to environmental conditions. The respondents from one sample, for instance, may be exposed to severe poverty, whereas respondents from another sample may have been drawn from a wealthy population. If poverty acts as a trigger for a genetic effect to surface (i.e., a gene-environment interaction), then the genetic effect would be detected in the first sample, but not in the second. So, failing to recognize differential exposure to environmental liabilities may be one additional reason for why there is a failure to replicate some genetic associations with criminal behavior.

The second type of gene-environment interplay that has direct bearing on criminology is known as gene-environment correlation. Gene-environment correlation captures the effects that genes have on predicting and explaining variance in environmental measures (Beaver and Wright 2005; Scarr and McCartney 1983). To most criminologists and social scientists, even suggesting that an environment could be affected by genetics seems a bit odd. There is, however, a great deal of empirical evidence indicating that almost all environments are influenced, at least in part, by genetic factors (Beaver and Wright 2005). To understand how it is possible to estimate genetic effects on environmental measures, it is essential to revisit the twin-based methodology. Recall that with the twin-based methodology, the similarity of MZ twins on some behavior trait is compared with the similarity of DZ twins on that same behavior/trait. If MZ twins are more similar than DZ twins, there is evidence of a genetic effect on the behavior/trait. This same twin-based methodology can be employed to estimate genetic effects on environmental measures; the only difference is that an environmental measure is used in place of the measure of the behavior/trait. A pool of empirical research has used the twin-based study to estimate genetic effects on environmental measures, and the results have provided strong support in favor of gene-environment correlations for environments related to crime and delinquency (Beaver and Wright 2005; DiLalla 2002). For example, variance in measures of parental socialization (Beaver and Wright 2007; DiLalla 2002), exposure to antisocial peer networks (Beaver et al. 2009), stressful life events (Dick et al. 2006), and various dimensions of family life have all been found to be affected, to varying degrees, by genetic factors. While the precise heritability estimates ebb and flow across the environmental measures, on average, genetic factors account for around 25 % of the variance in most environmental measures (Kendler and Baker 2006). Some environments, such as exposure to delinquent peer groups, have much higher heritability estimates with genetic factors explaining about around 60 % of the variance (Cleveland et al. 2005).

In addition to twin-based studies, a small number of studies have also explored the possibility that certain genetic polymorphisms might explain variance in measures of criminogenic environments. While this line of research is still in its infancy, there is some emerging evidence linking specific genetic polymorphisms, such as dopaminergic polymorphisms, to negative parental socialization, to contact with delinquent peers, and even to the probability of getting married (Dick et al. 2006). As more and more social science datasets begin to include genotypic information, the number of genes that are found to be associated with specific environmental conditions will likely increase.

Establishing a link between genetic variance and environmental variance is important in unpacking the ways in which genes and the environment work together to produce human phenotypic variation. Nonetheless, simply establishing a link reveals nothing about the underlying processes that account for gene-environment correlations. Existing research, however, has delineated three different types of gene-environment correlations, each of which explains a different mechanism by which genes account for environmental variance (Scarr and McCartney 1983). These three types of gene-environment correlations are known as passive gene-environment correlation, active gene-environment correlation, and evocative gene-environment correlation.

Passive gene-environment correlation is grounded in the fact that parents pass along two entities to their children: genes and an environment. Given that the child is unable to pick and choose their genes and/or their environment, they passively receive these. Moreover, because both genes and the environment are traced to the same source – that is, parents – the two are likely to be correlated (Beaver 2009; Scarr and McCartney 1983). For example, highly aggressive parents are likely to pass along a genetic predisposition for aggression and violence to their children. At the same time, parents who are highly aggressive are also statistically more likely to abuse their children, live in low SES areas, and not supervise their children closely. All of these environments have been shown to increase the probability of antisocial behaviors. With passive gene-environment correlation, then, children are hit with a "double whammy" of risk factors, wherein they have the genetic predispositions for antisocial behavior and they are also born into an environment that also contributes to antisocial behavior. This process thereby captures part of the reason that criminogenic environments are affected by genetic factors.

The second type of gene-environment correlation is active gene-environment correlation. Active gene-environment correlation avers that genotype pushes or nudges people into certain environments that are compatible with their genetic predispositions (Beaver 2009; Scarr and McCartney 1983). To illustrate, a person who is highly aggressive and violent is likely to seek out environments that are conducive to these predispositions. So, they might join a gang or befriend other people who are violent, too. In this example, choosing a gang is not a random occurrence, but rather is structured in part by the genetic predisposition for violence and aggression. It is important to underscore the fact that there are not genes that are "for" any type of environment; rather, genes operate indirectly, such as via their effects on personality traits. Continuing on with the gang example, a part of the reason a person might choose a gang is because they are genetically predisposed to be violent, not because there is a single gene that tells the person to join a gang.

The last type of gene-environment correlation is known as evocative gene-environment correlation. With evocative gene-environment correlation, genotype elicits certain responses from the environment, which are ultimately correlated with their genotype (Beaver 2009; Scarr and McCartney 1983). A person with a bad temper (a genetically influenced phenotype), for instance, is likely to evoke negative responses from their environment. Evocative gene-environment correlations are virtually synonymous with child-effects models except that the effect is thought to flow directly from genotype. Consider two siblings: one who is genetically predisposed to be unruly and one who is genetically predisposed to be passive and compliant. The unruly child will likely evoke negative and harsh parenting, such as being spanked, while the compliant child will likely escape such discipline. In this case, parental discipline is differentially invoked against the two siblings, but the reason for this difference is the result of the genetically influenced propensities and temperaments of the two siblings.

The saliency of these three types of gene-environment correlation waxes and wanes over different sections of the life course. Passive gene-environment correlation is thought to be most evident early in life, especially during infancy and childhood. During this time period, children are under the control of their parents and have their parents' environment imposed on them. As children develop into adolescence, they begin to gain more autonomy and thus are able to follow their genetic predispositions without as much guidance and control by their parents. By the time of early adulthood when most children have moved out of their parents' home, they are able to follow their genetic predispositions without interference. Evocative gene-environment correlation tends to have relatively equal effects throughout life. In childhood, for instance, genetically influenced antisocial traits may elicit negative responses from parents; in adolescence, genetically influenced antisocial traits may elicit negative responses from peers and teachers; and in adulthood, genetically influenced antisocial traits may elicit negative responses from employers and from potential mates.

Taken together, the two types of gene-environment interplay – gene-environment interaction and gene-environment correlation – draw attention to the very real possibility that genes and the environment do not represent simple dichotomies (Beaver et al. 2009). Instead, a wealth of scientifically rigorous scholarship has shown that there is a close interdependence between genes and the environment and the way to understand the causes of all human phenotypes – including antisocial ones – is to examine these two influences simultaneously and to model directly the various types of gene-environment interplay.

Conclusions and Future Directions

During the past century, the field of criminology has made some enormous gains in terms of identifying the causes and correlates of crime, delinquency, and other types of antisocial behaviors. For the most part, however, all of these advancements have been on discovering the environmental underpinnings to criminal involvement. With the recent mapping of the human genome and with empirical studies strongly implicating genes in all human phenotypes, the time is ripe for criminology to begin to examine the dual influences of genes and the environment in the genesis of antisocial behaviors. Such an approach does not mean that all existing theories need to be abandoned. Instead, such an approach allows for a fuller integration of findings from biological sciences into mainstream criminological theories. This type of theoretical integration would likely pave the way for an explosion of knowledge about the developmental pathways to crime and violence. The knowledge flowing from this research could then be used to open up newer and perhaps more effective avenues for the treatment and rehabilitation of offenders that would ultimately increase public safety and reduce criminal involvement.

Related Entries

- ▶ Evolutionary Theories of Criminal Behaviors
- ▶ Hormones, the Brain, and Criminality
- ▶ Neurology and Neurochemistry of Crime
- ▶ Synthesizing Biological and Social Theorizing

Recommended Reading and References

Beaver KM (2008) Nonshared environmental influences on adolescent delinquent involvement and adult criminal behavior. Criminology 46(2):341–369

Beaver KM (2009) Biosocial criminology: a primer. Kendall/Hunt, Dubuque

Beaver KM (2010) Genetic influences on being processed through the criminal justice system: results from a sample of adoptees. Biol Psychiatry 69(3): 282–287

Beaver KM (2011) The effects of genetics, the environment, and low self-control on perceived maternal and paternal socialization: results from a longitudinal sample of twins. J Quant Criminol 27(1):85–105

Beaver KM, Wright JP (2005) Biosocial development and delinquent involvement. Youth Violence Juv Justice 3(2):168–192

Beaver KM, Wright JP (2007) A child effects explanation for the association between family risk and involvement in an antisocial lifestyle. J Adolesc Res 22(6):640–664

Beaver KM, Wright JP, Delisi M, Walsh A, Vaughn MG, Boisvert D, Vaske J (2007) A gene × gene interaction between DRD2 and DRD4 is associated with conduct disorder and antisocial behavior in males. Behav Brain Funct 3(1):30–37

Beaver KM, Delisi M, Wright JP, Vaughn MG (2009) Gene–environment interplay and delinquent involvement: evidence of direct, indirect, and interactive effects. J Adolesc Res 24(2):147–168

Bouchard T, Lykken D, McGue M, Segal N, Tellegen A (1990) Sources of human psychological differences: the Minnesota study of twins reared apart. Science 250(4978):223–228

Brody GH, Beach SRH, Chen YF, Obasi E, Philibert RA, Kogan SM, Simons RL (2011) Perceived discrimination, serotonin transporter linked polymorphic region status, and the development of conduct problems. Dev Psychopathol 23(02):617–627

Brunner H, Nelen M, Breakefield X, Ropers H, van Oost B (1993) Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A. Science 262(5133):578–580

Caspi A, McClay J, Moffitt TE, Mill J, Martin J, Craig IW, Taylor A et al (2002) Role of genotype in the cycle of violence in maltreated children. Science 297(5582): 851–854

Cleveland HH, Wiebe RP, Rowe DC (2005) Sources of exposure to smoking and drinking friends among adolescents: a behavioral-genetic evaluation. J Genet Psychol 166(2):153–169

Dick DM, Agrawal A, Schuckit MA, Bierut L, Hinrichs A, Fox L, Mullaney J et al (2006) Marital status, alcohol dependence, and GABRA2: evidence for gene-environment correlation and interaction. J Stud Alcohol 67:185–194

DiLalla LF (2002) Behavior genetics of aggression in children: review and future directions. Dev Rev 22:593–622

Ferguson CJ (2010) Genetic contributions to antisocial personality and behavior: a meta-analytic review from an evolutionary perspective. J Soc Psychol 150(2):160–180

Haberstick BC, Timberlake D, Hopfer CJ, Lessem JM, Ehringer MA, Hewitt JK (2007) Genetic and environmental contributions to retrospectively reported DSM-IV childhood attention deficit hyperactivity disorder. Psychol Med 38(7):1057–1066

Harris JR (1998) The nurture assumption: why children turn out the way they do. Touchstone, New York

Kendler KS, Baker JH (2006) Genetic influences on measures of the environment: a systematic review. Psychol Med 37(5):615

Kim-Cohen J, Caspi A, Taylor A, Williams B, Newcombe R, Craig IW, Moffitt TE (2006) MAOA, maltreatment, and gene–environment interaction predicting children's mental health: new evidence and a meta-analysis. Mol Psychiatry 11(10):903–913

Miles DR, Carey G (1997) Genetic and environmental architecture of human aggression. J Pers Soc Psychol 72:207–217.

Moffitt TE (2005) The new look of behavioral genetics in developmental psychopathology: gene-environment interplay in antisocial behaviors. Psychol Bull 131(4):533–554

Raine A (1993) The psychopathology of crime: criminal behavior as a clinical disorder. Academic, San Diego

Rhee SH, Waldman ID (2002) Genetic and environmental influences on antisocial behavior: a meta-analysis of twin and adoption studies. Psychol Bull 128(3):490–529

Rowe DC (1994) The limits of family influence: genes, experience, and behavior. Guilford, New York

Scarr S, McCartney K (1983) How people make their own environments: a theory of genotype environment effects. Child Dev 54:424–435

Walsh A (2009) Biology and criminology: the biosocial synthesis. Routledge, New York

Walsh A, Beaver KM (2009) Biosocial criminology: new directions in theory and research. Routledge, New York

Wright JP, Beaver KM, Delisi M, Vaughn MG, Boisvert D, Vaske J (2008a) Lombroso's legacy: the miseducation of criminologists. J Crim Justice Educ 19(3):325–338

Wright JP, Tibbetts SG, Daigle LE (2008b) Criminals in the making: criminality across the life course. Sage, Los Angeles

Genetic Basis to Self-Control

Kevin M. Beaver[1,2] and Joseph L. Nedelec[3]
[1]School of Criminology and Criminal Justice, Florida State University, Tallahassee, FL, USA
[2]Center for Social and Humanities Research, King Abdulaziz University, Jeddah, Saudi Arabia
[3]School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA

Overview

In 1990 Gottfredson and Hirschi advanced a parsimonious and provocative theory of crime that was designed to explain all types of antisocial behaviors. Unlike most dominant criminological theories that focus on social/environmental factors as causes of crime, their theory focused on an individual-level factor – levels of self-control – as the key causal agent for crime, delinquency, and other forms of disrepute. More specifically they argued that persons with relatively low levels of self-control who have a criminal opportunity will, on average, commit more crime and antisocial acts in comparison with people who have relatively high levels of self-control. An impressive amount of empirical research has assessed the potential association between levels of self-control and a host of antisocial behaviors. The results of these studies have been remarkably consistent in showing that measures of self-control are among the most consistent and strongest predictors of a wide range of antisocial behaviors, including crime, delinquency, and drug use. Moreover, the link between levels of self-control and antisocial behaviors has been detected between genders, across racial/ethnic groups, and within samples collected from different countries (Pratt and Cullen 2000).

Given the robust criminogenic effects associated with levels of self-control, there has been a wave of research attempting to uncover the factors that are associated with causing variation in levels of self-control. Much of this research has been guided by Gottfredson and Hirschi's thesis that levels of self-control are largely the result of parental socialization. According to these theorists, parents who wish to raise offspring with high levels of self-control must engage in three intertwined parental management techniques. First, parents must supervise their children. Second, parents must recognize when their children are engaging in socially taxing or antisocial behavior. Third, parents must consistently correct their children's antisocial behaviors. Parents who engage in these three parenting techniques and who are attached to their children should, on average, raise offspring with relatively high levels of self-control. According to Gottfredson and Hirschi, parents do not have much time to inculcate self-control because by around the age of 10, levels of self-control are formed and remain relatively stable over the remainder of the life course.

A line of research has examined the merits of the parental management thesis and the results of these studies have provided some support in favor of Gottfredson and Hirschi's argument. For example, most studies reveal that measures of parental management techniques predict levels of self-control, with parents who supervise their children, recognize their children's transgressions, and punish their children, tending to raise offspring with relatively high levels of self-control (Gibbs et al. 1998; Hay 2001; Polakowski 1994). However, Gottfredson and Hirschi likely overstated the effects that parents have on sculpting levels of self-control. Numerous studies indicate that measures of parenting only have very small effects on levels of self-control and leave an overwhelming amount of variance unexplained. As a result, there are likely other salient factors that are causing variation in self-control that were not identified by Gottfredson and Hirschi. An emerging pool of research has pointed to the very real possibility that biological and genetic factors may be the dominant source of variation in self-control.

Genetic Influences on Levels of Self-Control

In articulating their theory, Gottfredson and Hirschi were very clear that they did not believe that genetic factors played any role in the development of self-control. However, research from multiple lines of inquiry, including developmental psychology, behavioral genetics, molecular genetics, and neurobiology, points to a very different conclusion – namely, that self-control, self-regulation, and impulse control problems are all affected in large part by genetic factors (Beaver 2009). Studies exploring the extent to which genetic variation accounts for variance in measures of self-control, and related disorders, typically employ samples that consist of twin pairs. There are two types of twins: monozygotic (MZ) twins who share 100 % of their DNA and dizygotic (DZ) twins who share, on average, 50 % of their DNA. Both types of twins, however, are assumed to share environments that are approximately comparable to each other. In other words, the environments of MZ twins are assumed to be no more similar than the environments of DZ twins. As long as this assumption is met (known as the equal environments assumption [EEA]) – and there is good evidence to indicate that it is – the only reason that MZ twins should be more similar to each other than DZ twins is because they share twice as much genetic material. So, genetic effects are detected if the similarity of MZ twins is greater than the similarity of DZ twins. And the greater the similarity of MZ twins in comparison to the similarity of DZ twins, the larger the magnitude of the genetic effect. The proportion of variance accounted for by genetic factors is known as the heritability estimate.

Although twin-based research designs provide a great deal of information regarding the genetic foundations to behaviors and traits, they also provide some of the most compelling evidence for the role of environmental factors. The reason for this is because the variance that is not explained by genetic factors is explained by environmental factors (plus error). Unlike most social science research that pools together all types of environments, twin-based research designs make the distinction between two different types of environments: shared environments and nonshared environments. Shared environments are those environments that are the same between twins and therefore make twins more similar to each other. Some of the common examples of shared environments are neighborhood conditions, family socioeconomic status, and common parental socialization factors. Nonshared environments, in contrast, consist of environments that are different between twins and that make twins dissimilar from each other. (Nonshared environments also capture the effects of measurement error.) Examples of nonshared environments include peer groups, parenting practices unique to each twin, and even different prenatal environments (especially for non-twin siblings). Together, the heritability, the shared environment, and the nonshared environment account for 100 % of the variance in any measure being studied in a twin-based research design.

Since criminological research focuses almost exclusively on the role of the environment in relation to criminal and antisocial behaviors, they have overlooked and ignored the potential role of genetic factors. As a consequence, much of the evidence that has examined the genetic basis to levels of self-control has to be culled from fields of study outside of criminology. Fortunately, an impressive number of studies have employed twin-based research designs (or variants of them) to examine the genetic origins to traits/disorders that overlap with Gottfredson and Hirschi's conceptualization of self-control. For example, studies have investigated the heritability of self-regulation, impulsivity, and attention-deficit/hyperactivity disorder (ADHD). The results of these studies have consistently revealed that genetic factors explain between 50 % and 90 % of the variance, with the remaining variance being attributable to nonshared environmental factors (Rietveld et al. 2003; Mick et al. 2002; Wright and Beaver 2005). In most studies, shared environmental factors explain none or very little of the variance. Only recently have criminologists begun to explore the genetic underpinnings to self-control and the findings have yielded results that parallel those reported by non-criminologists. Overall, genetic factors explain approximately 40–70 % of the variance, nonshared environmental factors explain between 30 % and 60 % of the variance, and shared environmental factors explain none of the variance. When taken together, the empirical research unequivocally indicates that levels of self-control are scripted, in part, by genetic factors along with nonshared environmental factors (Beaver 2009).

Although twin-based research designs represent a rich analytical tool that can be used to estimate genetic and environmental effects, they are limited in the ability to identify the specific genes and the specific environments that are involved. For instance, knowing that genetic factors account for 40–70 % of the variance in self-control is only part of the puzzle; finishing the rest of the puzzle requires searching for the particular genes which account for variation in self-control. Molecular genetic studies are useful in this regard.

Molecular genetic studies attempt to link variation in specific genetic markers to variation in certain behaviors, traits, and other characteristics (e.g., levels of self-control). To understand the underlying logic to these types of studies, it is necessary to provide some elementary background information about genes. The human genome is comprised of approximately 25,000 genes which are located on 23 pairs of chromosomes (22 pairs of autosomes and 1 pair of sex chromosomes). Because genes are located on pairs of chromosomes, all genes (except those located on the sex chromosomes for males) consist of two copies: one inherited maternally and one inherited paternally. The information encoded into genes determines virtually every physical characteristic (e.g., eye color, hair color) and influences many other unobservable traits (e.g., personality, cognitive skills). Most genes consist of only one version, meaning that all humans have the same "type" of gene. However, for a small percentage of genes, there exist at least two versions in the population. Genes that vary across people are known as genetic polymorphisms and alternative copies of the gene are known as alleles. To illustrate, suppose there was a gene for height and there were two different versions of the height gene: a short allele and a tall allele. It would be possible to inherit two short alleles (one maternally and one paternally), two tall alleles (one maternally and one paternally), or a short allele (either maternally or paternally) and a tall allele (either maternally or paternally). A person's height, therefore, would be partially a function of the alleles they inherited for this hypothetical polymorphic height gene. In reality, height is affected by hundreds of genes along with environmental factors (e.g., nutrition).

A great deal of research has examined the potential association between certain genetic polymorphisms and ADHD, impulsivity, and attention problems. The results of these studies have provided some evidence tying certain genes – especially those of the dopaminergic system – to these disorders and traits. There is some limited evidence, moreover, that genes of the serotonergic system may be involved in the development of self-control (Beaver 2009). The available evidence suggests that specific genetic polymorphisms tend to have their strongest effects on levels of self-control when they are paired to certain environments. What this necessarily means is that genetic polymorphisms are associated with levels of self-control, but the effects of these genetic polymorphisms are even more pronounced when they are coupled with adverse and criminogenic environments. These types of relationships are called gene-environment interactions, and they illustrate the incredible complexities underlying the causes of human behavior. Given that molecular genetic research involving self-control is still in its infancy, much more research needs to be undertaken to provide a more complete picture of which genetic polymorphisms may ultimately be responsible for affecting variation in levels of self-control. But, as the results of the twin-based studies reveal, genes tend to be the dominant force in structuring the development of self-control.

Neurobiology and Levels of Self-Control

The findings generated from genetic research often produce a great deal of confusion concerning what they actually mean. The media, for example, frequently talks about the discovery of a gene *for* X or a gene *for* Y (e.g., the crime gene, the low self-control gene). In reality, though, genes are not "for" any particular behavior, trait, or other human characteristic; rather, they are responsible for coding for the production of proteins. So, if genetic polymorphisms continue to be discovered that explain variance in levels of self-control, then how do these genetic polymorphisms actually produce varying levels of self-control? Stated differently, if genes do not directly determine levels of self-control, then what is the mechanism that ultimately links genetic variance (i.e., different alleles for specific genetic polymorphisms) to variance in levels of self-control? While there is not a definitive answer to this question, recent research from neurobiology sheds some light on this issue.

A proliferation of neuroimaging techniques have been developed, including functional magnetic resonance imaging (fMRI), positron emission tomography (PET), and single photon emission computed tomography (SPECT), that allow neuroscientists to examine the functioning of the brain when certain tasks are being performed. If a certain region of the brain is activated during a specific task, then it is quite likely that that area of the brain is involved – to some degree – in the skills needed for the successful completion of the task. A line of neurobiological research has examined brain activity in response to tasks designed to tap self-control, impulse control, judgment, and attention. Very generally, these studies have converged to show that the area of the brain partially responsible for these tasks is the prefrontal cortex (Beaver et al. 2007).

The prefrontal cortex is situated directly behind the forehead and is often divided into three different regions to help delineate the location, direction, and function of each. The first region, the dorsolateral prefrontal cortex (DLPFC), is located in the lateral of the prefrontal cortex and has been shown to be responsible for behavioral modulation, information processing, and memory formation. The second region is referred to as the orbitofrontal cortex and is located directly above the eyes and is interconnected with the DLPFC. The orbitofrontal cortex has been shown to be involved in maintaining goal-oriented behaviors, the regulation of emotions, and to affect decision-making processes. Finally, the medial prefrontal cortex (MPFC) is housed deep within the brain and is connected with the DLPFC. The MPFC is involved in promoting tasks that necessitate a significant amount of concentration.

The coordinated functions of the DLPFC, the orbitofrontal cortex, and the MPFC are typically referred to as executive functions. Although definitions of executive functions vary, they can be generically defined as a suite of functions that are responsible for judgment, decision-making, the ability to delay gratification, the ability to anticipate the consequences of actions, and the ability to modulate behaviors and emotions. Importantly, the definition of executive functions employed by psychologists and neurobiologists overlaps considerably with the conceptualization of self-control as set forth by Gottfredson and Hirschi. In fact, there is both empirical and theoretical evidence indicating that self-control is one of the many tasks that fall within the parameters of executive functions (Ishikawa and Raine 2003).

If self-control is indeed an executive function, then what accounts for variation in executive functions and self-control? As indicated previously, brain science research has revealed that the prefrontal cortex is largely responsible for executive functions. Variation in the structure and functioning of the prefrontal cortex therefore is largely the driving force behind variation in executive functions and thus levels of self-control. For example, some people's prefrontal cortexes are highly active, and, perhaps as a result, they score very high on measures of self-control. Other people, however, have relatively underactive prefrontal cortexes, and they are, on average, more likely to score lower on measures tapping self-control. The point is that executive functions and levels of self-control appear to be tied directly to the structure and functioning of the prefrontal cortex.

Being able to identify what causes variation in the structure and functioning of the prefrontal cortex is thus the key to identifying what causes variation in executive functions and levels of self-control. Neuroimaging research has been instructive in this endeavor, wherein twin studies have been conducted to estimate genetic and environmental effects on variation in brain structure and functioning. The results of these methodologically rigorous and highly influential brain science studies have revealed that upwards of 80 % of the variance in brain structure and functioning in the brain, including the prefrontal cortex, is the result of genetic factors (Toga and Thompson 2005). Consequently, it is now possible to provide a more detailed and scientific answer to why there is variation in levels of self-control. The evidence reviewed above indicates that self-control is an executive function. Executive functions are housed in the prefrontal cortex of the brain and variation in executive functions (thereby including levels of self-control) appear to be largely affected by the structure and the functioning of the prefrontal cortex. Variation in the structure and functioning of the prefrontal cortex is the result of genetic factors and to a lesser extent nonshared environmental factors. It should also be noted that approximately 60 % of the 25,000 genes in the human genome are in some way related to coding for the brain and/or brain development. As a result, there are likely thousands of genes that affect the structure and functioning of the prefrontal cortex with each of these genes tending to have relatively small effects.

A Biosocial Critique of the Parental Management Thesis

Even though there is a vast amount of empirical research indicating that levels of self-control are under strong genetic influence, the vast majority of criminological research ignores the potential genetic influence on self-control. Furthermore, much of this research also neglects the important contributions of neurobiology. Instead, most criminological studies only examine the role that social factors – especially parenting – have on the development of self-control. Critics of genetic research frequently point out that most of the criminological research shows that social factors are critically important to all types of antisocial behaviors, including levels of self-control. These findings, in short, appear to be at odds with those garnered from genetic and neurobiological research, leading to the question of which body of research should be believed.

In order to address this issue, it is essential to explore the most common type of methodology used in criminological research: standard social science methodologies (SSSMs). With SSSMs, researchers analyze data collected about one focal child (or adolescent) per household. The child's parents and teachers may also be interviewed to collect more detailed information, but the key concern is the behavior and/or traits of the focal child. As it applies to self-control, SSSMs would collect information about the focal child's level of self-control and certain social factors, especially parental management techniques. Statistical analyses are then conducted to determine whether levels of self-control covary with the social factors. If there is a statistically significant association between levels of self-control and the variation in the social factors, then most criminologists and social scientists often interpret the results as being in line with a causal explanation (e.g., Gottfredson and Hirschi's parental management thesis).

There is a serious problem with research that is based on SSSMs – namely, that SSSMs are unable to take into account genetic factors. Recall that twin-based research designs are used to estimate the relative effects of genetic and environmental factors. In order to do so, there must be at least two siblings (e.g., twins) included in the analysis. Unfortunately, criminological research rarely includes more than one sibling in the sample and, in fact, great pains are taken to include only one child per household for statistical reasons (i.e., to preserve independence in observations). By including only one child per household, it is not possible to estimate genetic effects and thus any research that employs an SSSM is making the assumption that genes have absolutely no effect on the behavior or trait being studied (e.g., levels of self-control).

The methodological limitations of SSSMs become all the more exacerbated for studies that examine the effects of parenting on childhood and adolescent levels of self-control. Recall that according to Gottfredson and Hirschi parents must supervise their children, recognize their children's misbehaviors, and consistently punish such waywardness. These three parental management techniques require a substantial amount of time, dedication, attachment, and self-control on the part of the parents. Parents who lack these prosocial traits are unlikely to follow the parental management suggestions offered by Gottfredson and Hirschi. Indeed, there is ample evidence to suggest that parents who lack self-control, parents who have a criminal record, and parents who engage in antisocial behaviors are not the most responsible parents and thus are unlikely to engage in effective parenting tactics.

All of this is critically important because as was discussed previously, levels of self-control are highly heritable. So, if a parent has low levels of self-control they are, statistically speaking, at risk for raising offspring who also have relatively low levels of self-control purely because the parent passed on the genetic material that influences levels of self-control. Research that fails to control for genetic factors (i.e., all research using SSSMs) when examining the potential nexus between parenting techniques and their offspring's level of self-control and detects a statistically significant association between parenting and self-control may erroneously attribute that association to a parenting effect, when it is really the result of genetic transmission.

In methodological parlance, this type of an association is referred to as a spurious relationship. One common example of a spurious relationship is the well-noted association between ice cream sales and violent crime. As ice cream sales increase, so does violent crime. It would be absurd to think that the consumption of ice cream causes an increase in violence; instead, there is likely a causal factor common to both ice cream sales and violence: temperature. As the temperature increases, people are more likely to consume ice cream and they are also more likely to engage in routines that place them in contact with violent criminals. A spurious relationship therefore can be defined as an association between two variables (e.g., ice cream sales and violence) that disappears after a third variable (e.g., temperature) is taken into account. In social science research, it is nearly impossible to eliminate the possibility of spuriousness, but most research attempts to minimize spuriousness through the use of control variables (or random assignment). Importantly, the only time a control variable needs to be included in a study to prevent spuriousness is when that variable is associated with both the independent variable (e.g., parenting) and the dependent variable (e.g., levels of self-control). If it is only related to one variable or the other, then there is no need to include the control variable to prevent spuriousness.

Now it is quite easy to see how criminological research using SSSMs and examining the association between parenting and self-control is likely biased. Genetic factors are likely affecting the way a parent raises their offspring, and these genetic factors are also passed on to their children in the form of genetic material. So, a parent who is abusive and neglectful likely has low levels of self-control, and the genetic predisposition for low levels of self-control is inherited by their children. Consequently, there is no way to know for certain whether a relationship between parenting and levels of self-control is a true "parenting" effect or whether it represents a spurious relationship when an SSSM is used.

To rule out spuriousness owing to genetic factors, a twin-based research design (or variant thereof) must be employed. Only a handful of studies have used twin-based research designs to examine the link between parenting and levels of self-control. The results of these studies have revealed that once genetic factors are taken into account, there is no association between parental socialization and offspring levels of self-control (Beaver et al. 2009a; Wright and Beaver 2005). These findings have serious implications for criminological research because they indicate that research using SSSMs is likely biased in favor of finding an environmental effect when such an effect is either likely upwardly biased or completely spurious. The use of SSSMs also helps to reconcile the different findings generated by behavioral genetic studies and by criminological studies. Behavioral genetic studies use appropriate research designs that make no assumptions about the role of genetics and the environment. Criminological studies that rely on SSSMs only produce unbiased findings when the assumption of no genetic effect is met. Unfortunately, all of the available evidence strongly suggests that this assumption is violated thereby casting doubt on the accuracy of criminological studies purportedly showing a link between parenting and levels of self-control.

Policy Implications

One of the major attacks leveled against biological explanations of antisocial behaviors and traits, including levels of self-control, is that the policy implications flowing from such research will result in oppressive and inhumane practices. According to this line of reasoning, since DNA is immutable, there is no way to change levels of self-control. In other words, people are born with a fixed level of self-control.

Those who are genetically endowed with low levels of self-control are destined for a life of crime and antisocial behaviors whereas those born with high levels of self-control are likely to lead prosocial and productive lives. This logic is erroneous and is based on the confusion between the physical structure of DNA and the effects emanating from DNA.

While it is true that the physical structure of DNA is nearly impossible to change, it is not true that genetic effects are always constant. Contemporary genetic research has shown that genetic effects change over different developmental time periods, with some genes being switched on at certain periods in the life course and switched off at other times. In addition, the effects of some genes appear to be controlled, at least in part, by exposure to environmental conditions. What this necessarily means is that a gene may have a strong effect in one environment, but no effect in another environment. One gene that was found to be related to levels of self-control was shown to reduce levels of self-control when it was paired with a criminogenic environment, but it was shown to be unrelated to levels of self-control when the criminogenic environment was absent (Beaver et al. 2009a).

The finding that genes are moderated by environmental conditions holds particular promise for prevention and intervention programs designed to change antisocial traits and behaviors, such as levels of self-control. For instance, it is quite possible that intervention and prevention programs could be individually tailored to each person's unique suite of genes to create a more effective treatment program. There is some empirical evidence emerging indicating that such an approach is quite effective. A recent study, for example, revealed that a program designed to reduce externalizing behavioral problems among children was effective for children with a certain genetic variant, but ineffective for children lacking that particular genetic variant (Bakermans-Kranenburg et al. 2008). This type of research remains very new and exploratory and so the true extent of its impact remains to be determined, but it does provide some promise for increasing the effectiveness of prevention and treatment programs.

Biological and genetic research has already had a profound impact on the juvenile justice system. In a landmark Supreme Court decision, the death penalty was abolished for juveniles. The decision to abolish the death penalty was based largely on the findings generated from academic research studies. These studies, however, were not produced by criminologists showing that environmental factors contributed to violence and aggression. Rather, the studies that ultimately swayed the Court's decision were produced by neuroscientists showing that the prefrontal cortex of the brain – the area of the brain that houses the executive function – is not fully developed until the mid-20s. Because the brains of adolescents are structurally immature in relation to the brains of adults, the Court ruled that adolescents should not be held as culpable for murder. Thus, in one of the most progressive policy decisions affecting the criminal justice system, it was biological research, not criminological research, that was the driving force (Beaver 2009).

Conclusions and Future Directions

The 1990s are known as the decade of the brain largely because of the tremendous amount of research that was conducted examining how neurobiological factors are related to different types of human behaviors and traits. Thus far in the 2000s, there has been an immense amount of genetic research trying to tie specific genetic variants to human behaviors and traits. Out of these two intertwined lines of research has emerged incontrovertible evidence indicating that self-control is a largely genetic trait that is housed in the prefrontal cortex of the brain. Despite these insights into the etiology of self-control, much remains unknown about the origins of self-control. Future research, for instance, needs to explore a wider swath of genetic polymorphisms to see which ones affect variation in self-control. At the same time, it is critically important to study the environmental factors that may moderate the effects of such genes. If this endeavor is successful, then it would likely provide a rich framework from which to create programs designed to promote levels of self-control among at-risk children, youth, and adults. In order to do so, however, criminological research must go beyond the sociological factors first purported by Gottfredson and Hirschi and integrate biosocial data, methods, and findings in order to come to a greater understanding of low self-control and its influence on criminality.

Related Entries

▶ Measurement of Self-Control
▶ Social Control
▶ Social Control and Self-control Through the Life Course

Recommended Reading and References

Bakermans-Kranenburg MJ, van Ijzendoorn MH, Mesman J, Alink LRA, Juffer F (2008) Effects of an attachment-based intervention on daily cortisol moderated by dopamine D4: a randomized control trial on 1- to 3- year-olds screened for externalizing behavior. Dev Psychopathol 20:805–820

Barkley RA (1997) ADHD and the nature of self-control. Guilford, New York

Beaver KM (2009) Biosocial criminology: a primer. Kendall/Hunt Publishing Company, Dubuque

Beaver KM, Wright JP, DeLisi M (2007) Self-control as an executive function: reformulating Gottfredson and Hirschi's parental socialization thesis. Crim Justice Behav 34:1345–1361

Beaver KM, Ratchford M, Ferguson CJ (2009a) Evidence of genetic and environmental effects on the development of low self-control. Crim Justice Behav 36:1158–1172

Beaver KM, Shutt JE, Boutwell BB, Ratchford M, Roberts K, Barnes JC (2009b) Genetic and environmental influences on levels of self-control and delinquent peer affiliation. Crim Justice Behav 36:41–60

Flint J, Greenspan RJ, Kendler KS (2010) How genes influence behavior. Oxford University Press, New York

Gibbs JJ, Giever D, Martin JS (1998) Parental management and self-control: an empirical test of Gottfredson and Hirschi's general theory. J Res Crime Delinq 35:40–70

Goldberg E (2001) The executive brain: frontal lobes and the civilized mind. Oxford University Press, New York

Gottfredson M, Hirschi T (1990) A general theory of crime. Stanford University Press, Palo Alto

Harris JR (1998) The nurture assumption: why children turn out the way they do. Touchstone, New York

Hay C (2001) Parenting, self-control, and delinquency: a test of self-control theory. Criminology 39:707–736

Ishikawa SS, Raine A (2003) Prefrontal deficits and antisocial behavior. In: Lahey BB, Moffitt TE, Caspi A (eds) Causes of conduct disorder and juvenile delinquency. Guilford, New York, pp 277–304

Mick E, Biederman J, Prince J, Fischer MJ, Faraone SV (2002) Impact of low birth weight on attention-deficit hyperactivity disorder. Dev Behav Pediatr 23:16–22

Pinker S (2002) The blank slate: the modern denial of human nature. Viking, New York

Plomin R, DeFries JC, McClearn GE, McGuffin P (2008) Behavioral genetics (5th edition). Worth Publishers, New York

Polakowski M (1994) Linking self- and social control with deviance: illuminating the structure underlying a general theory of crime and its relation to deviant identity. J Quant Criminol 10:41–78

Pratt TC, Cullen FT (2000) The empirical status of Gottfredson and Hirschi's general theory of crime: a meta-analysis. Criminology 38:931–964

Raine A (1993) The psychopathology of crime: criminal behavior as a clinical disorder. Academic, San Diego

Rietveld MJH, Hudziak JJ, Bartels M, van Beijsterveldt CEM, Boomsma DI (2003) Heritability of attention problems in children: cross-sectional results from a study of twins, age 3 to 12 years. Neuropsychiatric Genetics 1176:102–113

Rowe DC (1994) The limits of family influence: genes, experience, and behavior. Guilford, New York

Rutter M (2006) Genes and behavior: nature-nurture interplay explained. Blackwell, Malden

Siegel DJ (1999) The developing mind: how relationships and the brain interact to shape who we are. Guilford, New York

Tancredi L (2005) Hardwired behavior: what neuroscience reveals about morality. Cambridge University Press, New York

Toga AW, Thompson PM (2005) Genetics of brain structure and intelligence. Annu Rev Neurosci 28:1–23

Walsh A (2009) Biology and criminology: the biosocial synthesis. Routledge, New York

Wright JP, Beaver KM (2005) Do parents matter in creating self-control in their children? A genetically informed test of Gottfredson and Hirschi's theory of low self-control. Criminology 43:1169–1202

Wright JP, Tibbetts SG, Daigle LE (2008) Criminals in the making: criminality across the life course. Sage, Los Angeles

Genetic Expression

▶ Neurology and Neurochemistry of Crime

Genetic Profiling

▶ DNA Profiling

Genocide

Wenona Rymond-Richmond
Department of Sociology, University of
Massachusetts Amherst, Amherst, MA, USA

Overview

Genocide, the definitive "crime of crimes," has produced hundreds of millions of victims who were murdered, raped, sexually assaulted, forcibly displaced, kidnapped, robbed, and mutilated. The most severe violations of human rights are committed during genocidal violence. The sheer number of victims and crimes should imply that genocide is central to the discipline of criminology. Yet, genocide is neglected, marginalized, and undertheorized by criminologists. For the most part, criminologists have been remarkably indifferent to the crime of genocide and failed to incorporate genocide into their research agenda. The unresponsiveness of criminologists to the crime of genocide is part of a broader pattern of collective denial.

Criminologists should have much to contribute to the study of genocide as deviant behavior and social group conflict is at the heart of the legal definition of genocide and a central focus of our discipline. Understanding how and under what conditions people commit deviant acts and why particular groups or behaviors become victimized are undeniably criminological questions. Yet, criminology has been slow moving, unresponsive, and nearly silent towards incorporating genocide within its disciplinary boundaries. This neglect is part of a larger pattern of the discipline's near failure to incorporate any form of international war crimes into their research agenda. Mainstream criminology is preoccupied with interpersonal and intranational criminal acts of violence such as homicide, rape, and robbery leaving the role of the state in acts of crimes underexplored. Far too often, criminologists consider the state as a bulwark of crime rather than the perpetrator of crime. One notable exception is critical criminologists who condemn mainstream criminology for not considering the role of the state as a criminal actor.

A sociological approach to criminology can provide crucial insights, evidence, and theories about genocidal processes by explicitly addressing the collective dynamics of state-organized criminal victimization. Yet, if the past is an indication of the future, then the lack of application and extension of traditional criminological theory to genocide either means that we cannot expect much from mainstream theories or that latent racism and collective denial have hampered attempts to make connections between intranational and interpersonal crimes with international-, group-, and state-sponsored crimes. We may well need to develop new historically grounded theories that better account for state-sponsored collective violence.

The failure of criminologists to speak about genocide is not only a missed opportunity, but it also brings the validity and ethics of the discipline into question. Using the present-day genocide in Darfur as a case study, this entry will discuss the crime of genocide, criminologists silence on the topic, the ways in which criminology can contribute to the study of genocide, current issues and controversies, and future directions and suggestions for criminologists.

Definitions

Determining whether an atrocity is labeled as genocide, crimes against humanity, or ethnic cleansing has serious legal, social, historic, and symbolic consequences. As illustrated in the ongoing debates on whether the violent conflict in Darfur is labeled genocide is contentious. On one hand is a diverse group that includes Sudanese President Omar al-Bashir and scholar Mamdani (2009) who deny genocide is occurring in Darfur. On the other hand are academics including Eric Reeves (2007) and Hagan and Rymond-Richmond (2008a, b) and the Chief Prosecutor for the International Criminal Court (ICC) who has issued warrants of arrest on the charge of genocide. Defining a conflict is more than merely semantic as each term implies different legal and symbolic consequences and can influence the international community's response to the atrocities. For example, if a violent conflict is labeled a crime against humanity rather than genocide, the evidence needed for conviction is likely reduced, and this naming of the events will lack the symbolic force and probably mean less in the collective memory than would a legal determination of genocide (Savelsberg and King 2011). In fact, all genocides by definition are crimes against humanity, but not all crimes of humanity are elevated to the symbolic significance of genocide. Similarly, the term ethnic cleaning may rightfully describe the intentions of the perpetrators, yet this does not carry the same legal meaning and recourse as a determination of genocide.

Genocide

The term genocide came into existence in 1944 in response to the Holocaust. Raphael Lemkin, a Polish-Jewish lawyer, coined the word genocide by combining the Greek word for race or tribe, geno-, with the Latin word for killing, -cide. In large part due to Lemkin's efforts, on December 9, 1948, the United Nations approved the Convention on the Prevention and Punishment of the Crime of Genocide (Genocide Convention). The Genocide Convention established genocide as an international crime. The legal definition of genocide is found in Articles II and III of the Genocide Convention. Article II defines genocide as any of the following five acts committed with intent to destroy, in whole or in part, a national, ethnic, racial, or religious group:

(a) Killing members of the group
(b) Causing serious bodily or mental harm to members of the group
(c) Deliberately inflicting on the group conditions of life calculated to bring about its physical destruction in whole or in part
(d) Imposing measures intended to prevent births within the group
(e) Forcibly transferring children of the group to another group

For a charge of genocide, only one of the five acts described in Article II Sections a, b, c, d, and e needs to be met. Article III of the Convention describes the following five acts as punishable:

(a) Genocide
(b) Conspiracy to commit genocide
(c) Direct and public incitement to commit genocide
(d) Attempt to commit genocide
(e) Complicity in genocide

Incorporating genocide into international law represents a historic and momentous advancement in recognizing the act of genocide as criminal. Yet, tragically, establishing genocide as a crime has not eliminated its occurrence. Millions of individuals have been the victim of genocide since it was legally established. Victims include but are not limited to approximately 400,000 civilians in the Vietnam War, over 1 million Bengali in Bangladesh, 100,000 Hutu in Burundi, 1.7 million Cambodians, 200,000 Bosnian Muslims and Croats in the Former Yugoslavia, over 200,000 in Ethiopia, 100,000 Mayan Indians in Guatemala, 50,000–200,000 Kurdish in Iraq, 9,000–30,000 deaths in what is referred to as the Dirty War in Argentina, 8,000 in the Bosnian genocide, and 800,000 Tutsi in Rwanda, and over 400,000 Black Africans have been murdered in Darfur.

While the Genocide Convention provides a legal definition of genocide, legal and social scientific definition may differ. Many scholars have critiqued the legal definition for being too narrow and thereby omitting political groups and social classes from legal protection. Some have extended the meaning of genocide beyond the legal definition to embrace atrocities left outside the meaning of genocide. Examples of altering the definition include nonlethal acts that threaten the security of members of a group (Lemkin 1946), emphasizing the role of the state (Horowitz 1980), highlighting the one-sided mass killing by the state or other authority (Chalk and Jonassohn 1990), and specifying genocide as "politically motivated mass murder" (Chirot and Edwards 2003: 15).

Crimes Against Humanity

Crimes against humanity are defined differently in the Rome Statute and the statutes of the International Tribunals for the Former Yugoslavia (ICTY) and Rwanda (ICTR). The International Criminal Court (ICC) defines crimes against humanity as acts such as murder, enslavement, torture, rape, enforced prostitution, sexual violence, enforced disappearance, crime of apartheid, and "other inhumane acts of a similar character intentionally causing great suffering, or serious injury to body or to mental or physical health" when they are "committed as part of a widespread or systematic attack directed against any civilian population, with knowledge of the attack."

Collective Denial

Why do so few criminologists study genocide? And why has the work of the few that studied genocide, such as Ralph Lemkin and Sheldon and Eleanor Glueck, been largely forgotten? What can account for a discipline that specializes in crime, virtually ignoring the crime of genocide? Violence, murder, rape, property destruction, and victimization provide the foundation for the vast amount of research conducted by criminologists, yet when these crimes are perpetrated in their most extreme form, as is the case in genocides, criminologists fain interest.

A review of presentations at criminology annual conferences and journal publications exposes the failure of criminologists to speak about genocide. An examination of presentations at two annual crime-centered conferences and articles published between 1990 and 1998 in 13 top criminology journals reveals that out of 19,304 presentations, only 18, or .001 %, addressed genocide and out of 3,138 published article, only 1 was devoted to the crime of genocide (Yacoubian 2000: 12–13). Considering the fact that approximately one million people were murdered in the genocide in Rwanda and the genocide in Former Yugoslavia and two international tribunals were established during this time frame, the silence of criminologists is even more shocking and shameful.

Hagan and Rymond-Richmond (2008a) argue that sociologists and criminologists should incorporate genocide in their research agenda and end their silence on the "crime of crimes." The failure of criminology to engage in the study of genocide is a critique also applicable to its parent discipline, sociology. For example, Fein (1979) reviewed introductory sociology texts from 1947 to 1977 and found that few acknowledged genocide. A survey of texts in anthropology produced a similar pattern of neglect on the topic of genocide (Shiloh 1975).

Methodological Challenges

There are methodological challenges associated with studying genocide; however, they are not insurmountable and should not be used as an excuse for the discipline's lack of research on the crime of crimes. In addition to several methodological challenges, reasons for the limited research on genocide by criminologists include a lack of empathy for the crime because genocide and its victims appear distant since the crime was not likely experienced firsthand, an unwillingness to examine the United State's genocidal origins, perceived low-status topic, and latent racism.

Initial methodological difficulties include the catastrophic and heartbreaking possibility that entire groups and places may be eradicated because of massive killings. Documenting this bloodshed is imperative, but difficult without survivors. Additionally, entering into conflict areas might not be possible. The states' participation in genocide, as perpetrator, silent bystander, or ineffective intervener, as well as their role in covering up the atrocities can make data collection arduous. As is the situation in the genocide in Darfur, states have denied entry to outsiders, which may include researchers, humanitarian workers, journalists, and even security monitors. In addition to the difficulties of entering into a conflict zone for research purposes is the reality of potential physical and mental harm occurring while conducting research during or in the aftermath of genocide. Further methodological challenges include insufficient education on the topic of genocide in criminology. In fact, Yacoubian (2000: 8) found that only 3 out of 21 criminal justice, criminology, or justice studies programs that offered doctoral degrees provided a course on international crime. Finally, potential genocide researchers may be deterred by the claim that each genocide is so unique and particular that it is impossible to make comparisons or general claims.

Methodological challenges makes studying genocides difficult, yet one solution includes "extending time demands beyond neat packaged and predictable schedules based on survey-research or other easily amenable source of data" (Fein 1990: 6). Research methods available to criminologists and social scientists more broadly have inherent limitations and/or difficulties. Yet, these limitations can potentially be overcome and may lead to methodological advancements. Furthermore, "imperfect" data should not be used as an excuse for not researching genocide or a tactic use to shame scholars into participating in and thereby perpetuating the silence on genocide. "Perfect" data is hard to obtain and potentially impossible as each method suffers from some bias or limitations. Rather than dismiss "imperfect" data, the more challenging and intellectually engaged activity is to make an assessment on the quality of the data and, if appropriate, analyze the data and be transparent about the limitations in publications. Good scholarly work addresses these issues anyway.

Latent Racism

Hagan and Rymond-Richmond (2008a) speculate that genocide has been a marginalized topic of research for criminologists because of latent racism and a belief that the study of genocide is overly emotional and not worthy of academic research. Three individuals devoted a part of their career to writing books and articles about genocide and war crimes, yet few criminologists today acknowledge their efforts. Sheldon Glueck and his wife Eleanor are best known for their contributions to developmental criminology and their research on genocide is nearly forgotten, and Lemkin is barely known to criminologists at all. Laub and Sampson (1991: 1408) speculate that the Glueck's may have suffered from institutionalized anti-Semitism and sexism in American academia. This may account for why they and their research on genocide were marginalized. Horowitz (1980: 3) shares the speculation that research on genocide is limited due to its perceived low academic status and states that sociologists feel "a studied embarrassment about these issues, a feeling that intellectual issues posed in such a manner are melodramatic and unfit for scientific discourse."

Criminological Approach to Genocide

Historically orientated scholars have dominated the study of genocide. While their contributions have been indispensable, theory testing and development has been underdeveloped (Hagan and Rymond-Richmond 2009). Public health researchers have also made important contributions and the benefits their research has provided to genocide victims is commendable and life saving. Yet, they have a distinct approach, which typically only examines deaths in the refugee camps or IDP camps and does not specifically examine the issue of violence death. Nor is there typically an attempt by public health researchers to understand the causes of genocide and to asses blame. In addition, race scholars and political sociologists typically advance the literature and theories of ethnic and national violence. Yet criminologists may have as much to contribute on ethnic violence as sociologists who specialize in race and political issues.

Criminologists should have much to contribute to the study of genocide as deviant behavior, and social group conflict is at the heart of the legal definition of genocide and a central focus of our discipline. Criminologists are fundamentally concerned with understanding the relationship between victim and the perpetrator; characteristics of perpetrators; social, economic, and communal costs of victimization; effects of sanctions on deterring criminal behavior; and the context in which criminal actions are facilitated. Criminologists should explore if current theories such as strain, social-psychological, control, social learning, social disorganization, collective efficacy, conflict, techniques of neutralization, and feminist theories that are commonly used to explain intranational crimes can also account for the criminal behavior of genocide perpetrators. Alternatively, is there a need to develop entirely new theories? For example, techniques of neutralization may be a crucial process individuals experience as a way to rationale not only interpersonal crimes but also genocidal crimes (Sykes and Matza 1957 and Alvarez 1997). Social-psychological theories of obedience to authority and group conformity, developed from well-known studies such as the Stanford prison experiment, may be essential to understanding collective violence.

Additional questions tailored to the crime of genocide, yet firmly rooted in mainstream criminological concerns, include: Can rehabilitation, reintegration, and or restorative justice reduce violence as it has been shown to do for crimes that are more "traditional?" What effect does being a victim of genocide have on one's likelihood of becoming a perpetrator? What are the financial, emotional, communal, and household effects of being a genocide survivor? Can "bystander effects" be extended to include not only individuals but also states and the international community at large? If there were desistors to the crime, what where their characteristics and the context in which the desistance occurred? What are the physical, biological, economic, social, economic, and regional characteristics of perpetrators? Is there an unequal distribution of genocidal violence in particular locations, and if so why? What role, if any, does concentrated disadvantage and social disorganization have in genocide? Do genocide victims and perpetrators know each other, as is typically the pattern when there is a single perpetrator? Does formal and informal social control deter criminals and reduce future criminal opportunities? Can collective efficacy in the form of mutual trust and cohesion, traditionally theorized to reduce neighborhood crimes, be extended to account for the crime of genocide? Do parallels exist between hate crimes and genocide with the underlying connection being racism and dehumanization?

Competing Understandings of Genocide

Among the established explanations of genocidal victimization are the following six. First is a state insecurity approach that focuses on justifiable reactions to insurgent threats (e.g., Posen 1993). Second is a primordial explanation that emphasizes hatreds so long standing that they are considered exogenous (e.g., Kaplan 1993). Third is the population-resource perspective where competition for life-sustaining resources is considered (Diamond 2005). According to this perspective, opportunities and incentives are greatest, and resources most strained, in densely settled areas. Fourth is the instrumental perspective that emphasizes state-based ethnopolitical entrepreneurs who advance their interests by cultivating public fear and disrespect of subordinate groups (see Hardin 1995; Valentino 2004). Fifth is the constructionist approach that emphasizes racial symbols and identity manipulation by elites (e.g., Kaufman 2001). Finally, the sixth approach is a cognitive framing approach that identifies the shifts that appear during emerging conflicts as ranging from "normal" to "crisis" scripts or frames (Oberschall 2000).

Hagan and Rymond-Richmond (2008a, b) synthesize these six approaches into what they call a critical collective framing approach to explain the atrocities in Darfur. In addition, the theory builds on Coleman's (1986) social action theory and draws on criminological theories including Sampson's (2006) and Matsueda's (2007) concepts of collective and social efficacy and Sutherland's (1947) differential social organization theory. The theory helps to explain the link between microlevel social actions transforming into macro-level systems leading to organized genocidal victimization.

Rape and Genocide

Most people associate genocide as massive killing for the purpose of wiping out an ethnic group in whole or in part. Rape and sexual violence is not as frequently associated with genocide, despite its similar ability to be used as a tool of ethnic cleansing and genocide. As recognized in the groundbreaking *Akayesu* judgment by the International Criminal Tribunal for Rwanda, rape may "constitute genocide in the same way as any other act as long as they were committed with the specific intent."

Horrifically, rape and sexual violence are widespread in genocidal conflicts around the world. Rape is intended to terrorize women, dehumanize victims, destroy families and communities, and or to control the biological and cultural reproduction of women through impregnating victims. The genocide in Darfur exemplifies rape committed as a means to control the biological and cultural reproduction of women. Due to the patriarchal structure of Darfurian society, lineage is determined by patrilineal descent, thereby determining children of Black-African rape victims as Arabic. Intergroup rape is a means of controlling biological and cultural reproduction through "changing the race" and is a powerful weapon of destruction. In the Darfur genocide, there are numerous reports of perpetrators using racial epithets during the rapes and stating their intentions of impregnate female victims. For example, a Black-African genocide survivor reported that her Arabic attacker raped her and said, "You will have Arab babies" (Hagan and Rymond-Richmond 2008a). Another survivor reported that her Arabic attacker screamed, "We will kill all men and rape the women. We want to change the color. Every woman will deliver red. Arabs are the husbands of those women" (Hagan and Rymond-Richmond 2008a).

Estimating the number of rape victims during genocide is exceptionally challenging due to stigma, deaths resulting from health injuries suffered by rape victims, and because their assailants kill some victims after the rape. In the Rwandan genocide, estimated rapes total 500,000. Many of the rape victims were killed shortly after being raped. In Congo, rape estimates vary from 15,000 to 40,000. A study of Darfurian survivors residing in refugee camps in Chad analyzed by sociologists and criminologists Hagan et al. (2009) found that 4 % reported personal sexual victimization. Examining sexual victimization by gender demonstrates that women were sexually victimized at a greater rate than men were. Seven percent of the women in the full sample reported personal sexual victimization.

The stigma and legal ramifications Darfurian rape victims suffer is severe. Raped women are frequently disowned by their family and ostracized by their community. Adding to raped women's pain is a commonly held belief by the Sudanese that unwanted sex cannot make a woman pregnant. Therefore, if a woman becomes pregnant, it is believed that she was not raped. Making matters worse, under Public Order 1991/Criminal Order 1991, sex outside of marriage, premarital sex, and prostitution are prohibited. Rape in Sudan is considered sex outside of marriage or premarital sex done without the consent of the victim. Therefore, survivors of rape can be convicted for sex outside of marriage (called zina) unless they can prove that they did not consent to intercourse. Unmarried women convicted of zina can receive 100 lashes and married women can be stoned to death.

As a result of detrimental social and legal consequences, reports of personal sexual victimization are drastically underreported and women are more likely to report sexual victimization of others than they are to report sexual victimization of self. For example, when Darfurian genocide survivors were asked in the same survey discussed above to report on sexual victimization of other villagers rather than personal sexual victimization, the percent increased dramatically from 4 % to nearly 30 % (Hagan et al. 2009).

Racism and Genocide

Racially specific intent is central to the legal definition of genocide. However, the Genocide Convention has been criticized for being vague on how to prove intent. The ICTY and the ICTR have provided some clarity for determining perpetrator's intent. The use of racial epithets during an attack has been established in international law as an indicator of motivation and intent. The *Akayesu* decision in Rwanda (UN 1998), the *Jelisi* decision in Bosnia (UN 1999), and the Trial Chamber in *Kayishema* and *Ruzindana* emphasize the importance of spoken language as evidence of genocide.

To understand the motivations and intention of the perpetrators in the genocidal violence in Darfur, Hagan and Rymond-Richmond (2008a, b) relied on the testimony of the surviving victims and witness. The data reveals that the predominately Arabic attackers were yelling racial epithets as they killed, raped, abducted, and destroyed the homes of Black-African villagers. As exemplified below, these epithets involved tropes of slavery and dehumanization:

> "You donkey, you slave; we must get rid of you."
>
> "You blacks are not human. We can do anything we want to you. You cannot live here."
>
> "You blacks are like monkeys. You are not human."
>
> "Black prostitute, whore; you are dirty—black."

The documentation of racial epithets used during the attacks provides evidence that the violence was racially motivated and the intent was racially specific. Hagan et al. (2005) further demonstrate that perpetrators use of racial epithets significantly affected the degree of total victimization during the attack.

Current Issues, Controversies, and Debates

Defining Genocide

As discussed above, defining genocide is a controversial topic. Criticisms include the assertion that the legal definition of genocide is too narrow, thereby excluding groups from protection. Steps forward may include altering the definition of genocide by revisiting sections of the Convention that are frequently critiqued for being ambiguous, namely the issue of intent and what qualifies as "partial" group destruction. Criminologists should seriously consider expanding upon the legal definition of genocide by utilizing insights from their discipline to develop a criminological definition of genocide.

Genocide in Darfur

Among the global atrocities occurring at the time of this writing is a genocide occurring in the Darfur region of Sudan. The genocidal conflict began in February 2003 and is still disastrously ongoing. More than 400,000 Black-African Darfurians have been killed, and 2–3 million have been forcibly displaced (Hagan and Rymond-Richmond 2005). The perpetrators are the Sudanese government and the Janjaweed, who are almost exclusively Arab. The victims are Black Africans. Unlike Southern Sudan, where religious differences are frequently attributed as the cause of conflict, in Darfur, the Arabs and Black Africans practice the Muslim religion. The root cause of the genocidal conflict in Darfur is racial and ethnic hatred (Hagan and Rymond-Richmond 2005, 2008a, b).

A number of significant changes have occurred since the genocide began in 2003. First, the violence that started in Darfur has spread to South Kordofan region. Second, the county of Sudan split into two separate nations. In July 2011, South Sudan officially declared its independence from Sudan. Third, the International Criminal Court (ICC), which became operational on July 1, 2002, has issued warrants of arrest for President Omar al-Bashir, Hussein, Harun, Kushayb, Garda, Nourain, and Jamus for their participation in the genocide in Darfur. The establishment of ad hoc and permanent international criminal tribunals makes it possible to resolve international crimes such as genocide. The International Criminal Tribunals for the Former Yugoslavia (ICTY), the International Criminal Tribunal for Rwanda (ICTR), and the International Criminal Court (ICC) have jurisdiction for the crimes of genocide, crimes against humanity, and war crimes. The most historically significant warrant of arrest by the ICC is for Sudanese President Omar al-Bashir. This is the first time a sitting head of a state was issued a warrant by the court. Charges filed by Chief Prosecutor Luis Moreno-Ocampo against President Bashir include war crimes, crimes against humanity, genocide, rape, and mass murder as genocide. Of these charges, rape as genocide is the most groundbreaking. Prosecuting the crime of rape as genocide is unprecedented for the ICC and relies on two lesser-known ways of destroying a people: "causing serious bodily or mental harm to members of the group" or "deliberately inflicting on the group conditions of life calculated to bring about its physical destruction in whole or in part." Prosecuting President Bashir with genocide including using rape as a form of genocide will provide a legal precedence for the International Criminal Court to pursue rape as a form of genocide in the future.

Additional legal means to reduce crime and inequality include enforcing Sudanese laws against rape and eliminating Public Order 1991/Criminal Order 1991. Flogging, amputating, and formalizing the death penalty for a wide range of offenses including adultery, embezzlement, and "corruption" were restored under this order. The order includes numerous criminal acts, which are vaguely defined permitting those in power to select who is criminal and what to call a crime. Women in particular have been targeted, victimized, and controlled under Public Order 1991.

Conclusion and Future Research

The boundaries of criminology must expand beyond focusing on individual criminal actions to include crimes committed by the state. Within the discipline, we can look at the subdiscipline of critical criminology for guidance. The inclusion of genocide into criminological research should be accompanied by revisiting the pioneering work of Glueck, Lemkin, and criminologists who worked on this marginalized topic as well as interdisciplinary appreciation for research on mass atrocities and war crimes conducted by anthropologists, historians, political scientists, social theorist, psychologists, and philosophers.

A second recommendation for future research is to theorize about the ways in which perpetrators of genocide and war crimes more broadly resemble and depart from "ordinary" delinquents more commonly studied by criminologists. Are they a vastly different type of criminal? Can their behaviors be explained through mainstream criminological theories or is there a need to develop new theories to account for war criminals? To what extent can mainstream criminology theories that frequently view criminals as engaging in delinquent behavior account for genocidal violence in which the perpetrator is potentially committing a crime of obedience? In what ways can we account for the role of the state in genocidal violence?

Third, criminologists can make value contributions to understanding the consequences of genocidal victimization. Indeed, one of the most resilient findings in victimology is that victimization does not end after an attack. Crime victims typically continue to suffer socially, economically, and psychologically after an attack. Similarly, victimization does not end after a genocidal attack (Rymond-Richmond and Hagan 2012). Fourth, researchers must study the behaviors, motivations, and emotions of perpetrators. Most of what we know about the motivations of perpetrators comes from the perspective of the victim. While telling and important, augmenting this data with the perspective of the perpetrator could provide vital insights into the causes, process, and preconditions and quite possibly illuminate potential solutions to eliminating genocide.

Fifth, while the Genocide Convention contains two provisions that obligate signatories to the Convention to intervene to halt genocidal violence, it has invoked fairly weak international efforts to intervene. The concept of laws on the books and laws in action has been studied by criminologists and other social scientists to illuminate and theorize about this disjuncture. Criminologists may be able to contribute to our understanding of why this disjuncture exists between states obligation to intervene in the genocide and their minimal efforts to do so. Moreover, and potentially more importantly, criminologists may be able to identify mechanisms that move from recognizing genocidal occurrences to international intervention.

In addition, the importance of genocide research, the unique contributions criminologists can make, and the increased financial necessity when conducting research abroad should be recognized by funding agencies such as the National Institute of Justice (NIJ). A review of NIJ-funded research projects during the years 1995, 1996, and 1997 demonstrates that $140 million dollars was awarded to 529 different research projects and only one, or .002 %, was tenuously associated with international crime (Yacoubian 2000: 14). Since this review does not address how many of the proposed research projects focused on genocide, it is difficult to know whether the virtual nonexistent NIJ funding is due to a lack of proposals on the topic or whether NIJ also engages in silencing genocide.

The terms "activists" and "scholar" have often been pitted against each other as if one cannot be both. Hagan and Rymond-Richmond (2009) insist that criminologists can conduct methodologically strong, peer-reviewed research that embraces activisms. Outside of criminology, there is a little more acceptance of merging the two together. In sociology, this is referred to as public sociology, and in anthropology, it is called action anthropology. In criminology, the term used to describe Hagan and Rymond-Richmond's research is activists' criminology. As good, reflexive researchers acknowledge, each researcher has their own bias, experience, race, class, gender, and even theoretical inclinations that affect the topics we choose and the lens in which we view our data. Acknowledging positionality does not negate responsible, good research.

Criminologists have a scholarly and moral obligation to build a "new criminology of genocide" (Matsueda 2009). Understanding the context of genocide, the motivations of the criminals, and the cumulative effects of victimization are important steps criminologists can make towards preventing their future occurrence. Furthermore, through the collection and analysis of data, criminologists can play a role in criminal prosecution of the perpetrators by providing information on the participation of the state, the degree of the destruction, and the racial nature of the genocide.

The building of a new criminology of genocide may well require criminologists to move beyond the familiar and comfortable mainstream. This is not completely uncharted. Feminist, Marxism, peacemaking, and postmodernism theory have modeled how this may be accomplished. Sutherland's work nearly 50 years ago on white-collar crime is another example of expanding the boundaries of criminology. It is time Raphael Lemkin's concept of genocide is incorporated into the research agenda of criminologists. If for no other reason, neglecting to study the crime of genocide is to miss an opportunity to make significant contributions to the discipline.

Related Entries

- ▶ Crime Investigations by the International Criminal Court
- ▶ Environmental and Human Rights
- ▶ Feminist Theory in the Context of Sexual Violence
- ▶ Genocide
- ▶ History of Restorative Justice
- ▶ Human Rights Violations in Criminal Court
- ▶ Marxist Criminology
- ▶ Monitoring and Evaluation of Restorative Justice
- ▶ Political Crime
- ▶ Resistance to State Crime
- ▶ Restorative Justice and State Crime
- ▶ State Crime
- ▶ Transitional Justice

▶ Victimization, Gender, and the Criminal Justice System
▶ Victims and Restorative Justice
▶ Victims and the International Criminal Court
▶ Victims of State Crime

References and Further Readings

Alvarez A (1997) Adjusting to genocide: the techniques of neutralization and the Holocaust. Soc Sci Hist 21(2):139–178

Chalk F, Jonassohn K (1990) The history and sociology of genocide: analyses and case studies. Yale University Press, New Haven

Chirot D, Edwards J (2003) Making sense of the senseless: understanding genocide. Context 2(2):12–19

Cohen S (1993) Human rights and crimes of the state: the culture of denial. Aust N Z J Criminol 26:97–115

Coleman JS (1986) Social theory, social research, and a theory of action. Am J Sociol 91(6):1309–1335

Diamond J (2005) Collapse: how societies choose to fail or succeed. Penguin, New York

Fein H (1979) Accounting for genocide: national responses and Jewish victimization during the Holocaust. Free Press, New York

Fein H (1990) Social recognition and criminalization of genocide. Curr Sociol 38(1):1–7

Fein H (2007) Human rights and wrongs: slavery, terror, genocide. Paradigm, Boulder

Fox JA, Levin J (1998) Multiple homicide: patterns of serial and mass murder. In: Tonry M (ed) Crime and justice: a review of research, vol 23. Chicago University Press, Chicago

Glueck S (1944) War criminals: their prosecution and punishment. Knopf, New York

Hagan J, Rymond-Richmond W (2008a) Darfur and the criminology of genocide. Cambridge University Press, Cambridge

Hagan J, Rymond-Richmond W (2008b) The collective dynamics of race and genocidal victimization in Darfur. Am Sociol Rev 73(6):875–902

Hagan J, Rymond-Richmond W (2009) Criminology confronts genocide: whose side are you on? Theor Criminol 13:503–511

Hagan J, Rymond-Richmond W, Parker P (2005) The criminology of genocide: the death and rape of Darfur. Criminology 43:525–561

Hagan J, Rymond-Richmond W, Palloni A (2009) Racial targeting of sexual violence in Darfur. Am J Public Health 99(8):1386–1392

Hardin R (1995) One for all: the logic of group conflict. Princeton University Press, Princeton

Horowitz IL (1980) Taking lives: genocide and state power. Transaction Books, New Brunswick

Kaplan R (1993) Balkan ghosts: a journey through history. St. Martin's Press, New York

Kaufman S (2001) Modern hatreds: the symbolic politics of ethnic war. Cornell University Press, Ithaca

Kuper L (1981) Genocide: its political use in the twentieth century. Penguin, New York

Laub JH, Sampson R (1991) The Sutherland-Glueck debate: on the sociology of criminological knowledge. Am J Sociol 96(6):1402–1440

Lemkin R (1946) Genocide. Am Scholar 15:227–230

Mamdani M (2009) Saviors and survivors: Darfur, politics, and the war on terror. Verso, London

Matsueda R (2007) Differential social organization, collective action, and crime. Crim Law Soc Change 46(3):33

Matsueda R (2009) Toward a new criminology of genocide: theory, method, and politics. Theor Criminol 13(4):495–502

Oberschall A (2000) The manipulation of ethnicity: from ethnic cooperation to violence and war in Yugoslavia. Ethn Racial Stud 23:982–1001

Posen B (1993) The security dilemma and ethnic conflict. Survival 35:27–47

Reeves E (2007) A long day's dying: critical moments in the Darfur genocide. The Key Publishing House, Toronto

Rymond-Richmond W, Hagan J (2012) Race, land, and forced migration in Darfur. In: Kubrin CE, Zatz MS, Martinez R (eds) Punishing immigrants: policy, politics, and injustice. New York University Press, New York

Sampson R (2006) How does community context matter? Social mechanisms and the explanation of crime rates. In: Wikstrom PH, Sampson R (eds) The explanation of crime: crime, mechanisms, and development. Cambridge University Press, Cambridge, UK, pp 31–60

Savelsberg J, King R (2011) American memories: atrocities and the law. Russell Sage, New York

Shiloh A (1975) Psychological anthropology: a case Study in culture blindness? Cult Anthropol 16(4):618–620

Sutherland EH (1947) Principles of criminology. J.B. Lippincott, Philadelphia

Sykes G, Matza D (1957) Techniques of neutralization: a theory of delinquency. Am Sociol Rev 22:664–670

Valentino BA (2004) Final solutions: mass killing and genocide in the 20th century. Cornell University Press, New York

Yacoubian GS (2000) The (In)significance of genocidal behavior to the discipline of criminology. Crim Law Soc Change 34:7–19

Geo-forensic Analysis

▶ Applied Geographical Profiling

Geographic Profiling

D. Kim Rossmo
School of Criminal Justice, Texas State
University-San Marcos, San Marcos, TX, USA

Overview

Geographic profiling is a criminal investigative methodology for analyzing the locations of a connected series of crime to determine the most probable area of offender residence. Its major function is suspect prioritization in investigations of serial crime. The technique is based on the theories, concepts, and principles of environmental criminology. Crime pattern, routine activity, and rational choice theories provide the foundation for understanding the target patterns and hunting behavior of criminal predators.

These theories suggest a method for describing the mathematical relationship between offender travel and likelihood of offending. This relationship can be described by a buffered distance-decay function that looks in cross-section rather like a volcano with a caldera. The function is encoded in a computer algorithm. Geographic profiling uses specialized crime-mapping software based on this algorithm to determine the most probable area of offender residence from the spatial pattern of the crimes.

Geographic profiling has turned out to be a robust and versatile methodology. Originally developed for analyzing serial murder cases, it was subsequently applied to rape, arson, robbery, bombing, kidnapping, burglary, auto theft, credit card fraud, and graffiti investigations. A number of innovative applications outside law enforcement also exist, with geographic profiling being used in military operations, intelligence analysis, biology, zoology, epidemiology, and archaeology.

Introduction

Geographic profiling is a criminal investigative methodology for analyzing the locations of a connected series of crime to determine the most probable area of offender residence (Rossmo 2000). Its major function is suspect prioritization in investigations of serial crime. A criminal investigation involves two tasks – finding the offender and proving guilt. Guilt can only be established by a confession, a witness, or through physical evidence. The task of finding an offender, a particular challenge in a "whodunit" investigation of a stranger crime, involves collecting, prioritizing, and evaluating suspects. High profile cases often have thousands of suspects and consequent problems of information overload. In such situations, geographic profiling can assist in case information management.

Geographic Profiling

History

Geographic profiling is typically employed in cases of serial violent or property crime, though it can be used in investigations of single crimes involving multiple locations or significant aspects of geography (Davies and Dale 1995; Knabe-Nicol and Alison 2011). Police have long used ad hoc mapping efforts to support certain criminal investigations. The first recorded use of investigative spatial analysis was during the Hillside Stranglers investigation in 1977. The Los Angeles Police Department analyzed the sites where the murder victims were abducted, their bodies dumped, and the distances between these locations, enabling them to identify the area where the killers were based. A similar analysis, using spatial means and distance-time factors, was conducted in 1980 during the Yorkshire Ripper inquiry in England (Kind 1987). More sophisticated models emerged from research conducted at Simon Fraser University's School of Criminology in Canada (Rossmo 1995). The technique was first implemented operationally in the Vancouver (British Columbia) Police Department's Geographic Profiling Section in 1995.

Theory

Geographic profiling is based on the theories, concepts, and principles of environmental

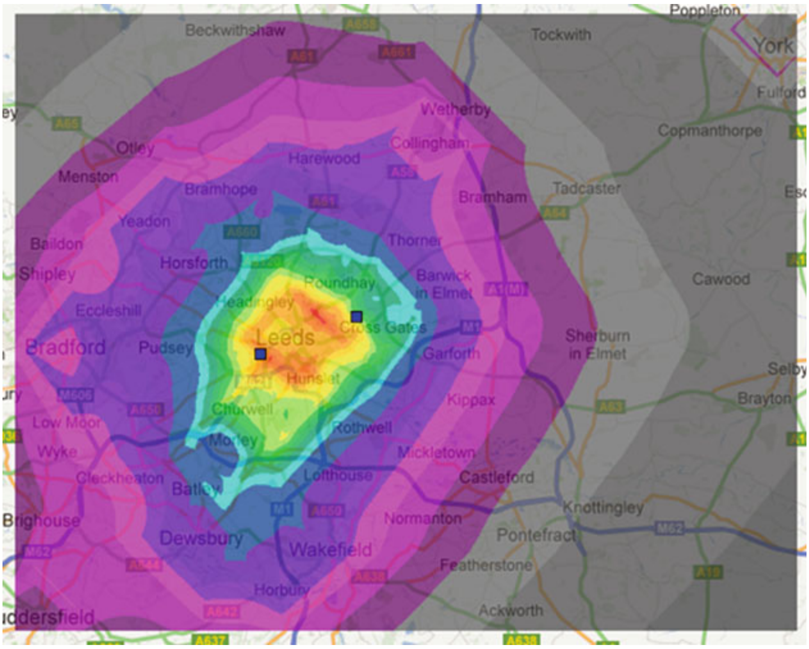
Poppleton
York
Beckwithshaw
A658
A61
A661
Tockwith
Fulford
Wetherby
Otley
A65
Collingham
Copmanthorpe
Harewood
Menston
Bramhope
A58
Bramham
Tadcaster
A61
A64
Yeadon
A660
Baildon
Thorner
Horsforth
Shipley
Barwick in Elmet
A1(M)
Cawood
Roundhay
Headingley
Eccleshill
Cross Gates
Leeds
M1
Bradford
Pudsey
Garforth
Sherburn in Elmet
Selby
Hunslet
M606
Kippax
Brayton
A650
Churwell
A63
Low Moor
Rothwell
Morley
Wyke
Mickletown
Cleckheaton
Lofthouse
Castleford
Batley
Knottingley
Brighouse
M1
A650
Normanton
M62
Dewsbury
Pontefract
M62
Wakefield
A644
Featherstone
A19
A636
Horbury
Ackworth
uddersfield
A642
A637
A638

Geographic Profiling, Fig. 1 Jeopardy surface (top 10 %) for Operation Lynx

criminology. Crime locations are not distributed randomly in space but rather are influenced by the road networks and features of the physical environment. This focus on the crime setting – the "where and when" of the criminal act – offers a conceptual framework for determining the most probable area of offender residence. Environmental criminology is interested in the interactions between people and their surroundings and views crime as the product of offenders, victims, and their setting (Brantingham and Brantingham 1981, 1984). The three theories underlying geographic profiling – crime pattern (Brantingham and Brantingham 1981, 1993), routine activity, and rational choice (Clarke and Felson 1993) – provide the foundation for understanding the target patterns and hunting behavior of criminal predators.

Computer Systems

Geographic profiling uses specialized crime-mapping software to determine where an offender most likely lives. The mathematical relationship between offender travel and probability of offending can be described by a buffered distance-decay function that looks, in cross-section, rather like a volcano with a caldera. The function is encoded in a computer algorithm. Working from the point pattern of the crime locations, the computer software produces a jeopardy surface – a three-dimensional probability surface – outlining the most probable area of offender residence (see Figs. 1 and 2 below). Those positions higher on the jeopardy surface (colored red) are more likely to contain the offender's residence, those lower, less likely. A police investigator can then prioritize suspect addresses or other locations by their position on this probability surface.

A geographic profiling program includes an analytic engine, geographic information system (GIS) capability, database management, and powerful visualization tools. *Rigel*, based on the Criminal Geographic Targeting (CGT) algorithm developed in 1991 at Simon Fraser University, was the first commercial geographic profiling system. Two similar geographic profiling programs, *CrimeStat* and *Dragnet*, primarily used for research rather than police operations, were later released in 1999.

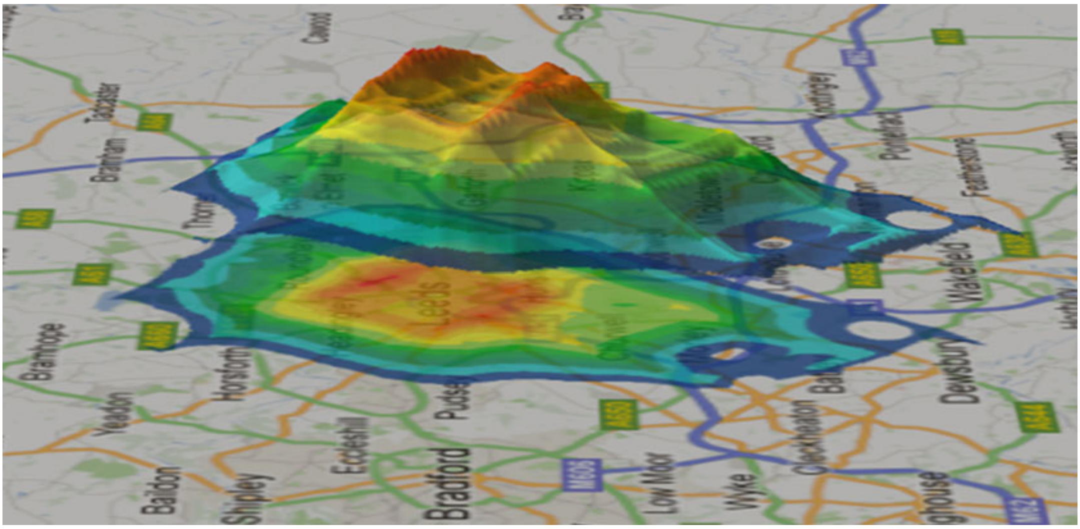

Geographic Profiling, Fig. 2 Geoprofile for Operation Lynx

Law enforcement agencies can use geographic profiling software to optimize their limited resources. The performance of a geoprofile is determined by a measure called the hit score percentage (HS%), defined as the ratio of the size of area that has to be searched, following the geoprofile prioritization, before the offender's base is found, to the total hunting area. For example, if the crimes in the series covered 10 square miles, and the geoprofile located the offender in 1.5 square miles, then the HS% would be 15 %. The smaller the HS%, the more precise the geoprofile's focus and the better its performance. An evaluation of geographic profiles prepared for operational cases that were eventually solved showed that the mean HS% was 5 % and the median 3 % (Rossmo 2011).

Process

The process of generating a geographic profile involves a number of steps, beginning with determining which crimes are connected in a series, evaluating the case information, creating the geoprofile, and finally recommending investigative strategies. This process is described in more detail below.

Linkage Analysis

The investigation of a crime series starts with determining which specific offenses are connected together. This procedure is known as linkage analysis or comparative case analysis. Every crime in the pattern can be considered a piece in a jigsaw puzzle; the more pieces, the more information; the more information, the more detailed the overall investigative picture. Linkage analysis identifies case similarities and common suspects, leading to information sharing between detectives and different police jurisdictions. When a case is solved, more crimes may be cleared and the courts can sentence convicted offenders more appropriately.

A linkage analysis requires the comparison of similarities versus differences for both connected and unconnected crimes; connected crimes should show more similarities than differences and unconnected crimes, more differences than similarities. There are three main methods used to link crimes: (1) physical evidence; (2) offender description; and (3) crime scene behavior (Rossmo 2000). If present, physical evidence such as DNA or fingerprints can establish crime linkage with certainty. In contrast, the other two methods are probabilistic in nature; witness descriptions vary in their accuracy, and offenders do not always exhibit the same modus operandi.

Signature – unique behaviors not required for the commission of the crime, such as certain fantasy-based sexual routines in a rape series or the inscription of “FP” (for “fair play”) by the New York Mad Bomber – provides a solid link if present; unfortunately, signature is rare. Proximity in time and place between crimes significantly increases their likelihood of being connected.

After a linkage analysis has been completed, the connected crime locations, distinguished by site type, are entered into the system by street address, latitude and longitude, or digitization. These optional entry methods reflect the reality that crime can happen anywhere – houses, parking lots, back alleys, highways, parks, rivers, and even mountain ravines.

Considerations

An exploratory analysis of the data is next conducted. Several different crime factors and environmental elements need to be considered in the construction and interpretation of a geographic profile:

1. Crime sites – While the locations of the crimes are essential to a geographic profile, the types of crime sites are also important. A homicide, for example, can involve separate or combined encounter, attack, murder, and body disposal sites; each site type has a separate analytic meaning.
2. Temporal factors – When the crimes occurred (date, day of week, and time of day) and their chronological order provide valuable context for understanding the crime sites. Temporal information may also provide insight into the offender’s routine activities.
3. Hunting style – A criminal’s hunting method (defined as the search for, and attack on, a victim or target) influences his or her crime site pattern (Beauregard et al. 2011). Hunting style is therefore an important consideration in geographic profiling.
4. Target backcloth – The target backcloth is an opportunity surface representing the availability of potential targets or victims in a given area. An offender’s choice options may be limited if the target backcloth is constrained or patchy (e.g., a criminal preying on street sex workers in a red light district). This may reduce the importance of certain types of crime sites (victim encounters) in the preparation of a geographic profile.
5. Arterial roads and highways – People, including criminals, do not travel as the crow flies. They follow street layouts and are more likely to travel along major arterial routes, freeways, or highways. Crimes often cluster around freeway exits.
6. Bus stops and rapid transit stations – Offenders without vehicles may use public transit or travel along bicycle or jogging paths. The locations of these routes and their access points may be an important consideration for understanding the crime patterns of such offenders.
7. Physical and psychological boundaries – Movement is constrained by physical boundaries such as rivers, lakes, ravines, and highways. Psychological boundaries, resulting from socioeconomic, ethnic, or race differences, also influence movement.
8. Zoning and land use – Zoning (e.g., residential, commercial, industrial) and land use (e.g., stores, bars, businesses, transportation centers, major facilities, government buildings, military institutions) provide important keys as to why an offender may have been in a particular area.
9. Neighborhood demographics – Some sex offenders prefer victims of a certain racial or ethnic group. These groups may be more common in certain neighborhoods than in others, affecting spatial crime patterns.
10. Displacement – Media coverage or police patrol presence can cause spatial displacement, affecting the locations of subsequent crime sites. Any displacement issues have to be compensated for in a geographic profile.

Creating the Geoprofile

Once all these various factors are considered, a scenario, involving a subset of crime locations most relevant for determining the offender's residence, is created (e.g., non-independent crime sites will be excluded). The next stage is the actual generation of the geographic profile. Conducting hundreds of thousands of iterations of its criminal hunting algorithm, the software assigns probability values (depicted with a color spectrum) to each pixel in what is typically a 40,000-pixel grid overlaid on a map of the crime sites. The final output is a color two-or three-dimensional map that shows the most likely area of offender residence (see Figs. 1 and 2 below). The geoprofile can then be used as the basis for a number of police investigative strategies.

Investigative Strategies

The function of a geographic profile is to focus a criminal investigation. Police agencies have employed a number of strategies over the past 20 years that take advantage of this spatial prioritization. The development of these approaches has been an ongoing interactive process involving investigator input and operational experience (Daniell 2008). Geographic profiling investigative strategies can be broadly divided into suspect-based and area-based approaches, depending on whether individuals or locations are being prioritized. It should be emphasized that a geographic profile is only one of many techniques in the detective's repertoire. However, it can increase effectiveness and efficiency and, in some situations, make possible an investigative approach that would otherwise not be feasible.

Suspect-Based Strategies

Suspect prioritization involves the assessment of individuals, including suspects, persons of interest, and known criminals. One of the benefits of geographic profiling is the ubiquity of address-based record information (estimated to be as high as 85 %). Potential suspects and investigative leads can be found in a variety of databases: police dispatch, record management, and jail booking systems, sex offender registries, and parolee and predatory violent criminal lists.

Data banks are often geographically based, and parole and probation offices, mental health clinics, social services offices, schools, and other agencies located in prioritized areas may provide information of value. Several commercial companies offer law enforcement agencies the ability to search multiple personal information databases. Their systems (e.g., Accurint and AutoTrack) use proprietary data-mining algorithms to sample and select large quantities of data electronically and assign them to individual profiles.

A department of motor vehicles (DMV) record search for a suspect vehicle can be focused by cross matching an offender's description from driver's licenses files and prioritizing the results using geographic profiling. The combined search parameters act as a linear program to produce a manageable list of records.

Area-Based Strategies

Area prioritization involves the allocation of police resources for such activities as surveillance, canvassing, and directed patrolling (for an interesting case example involving a serial burglar, see Rossmo and Velarde 2008). It has also been used to focus intelligence-led DNA screens ("bloodings") in which individuals are prioritized based on geography, criminal record, age, and other relevant criteria. In certain missing person cases that are suspected homicides, geographic profiling can help determine probable body disposal sites or burial areas if a suspect has been identified.

Operation Lynx

Operation Lynx was the name of a major police operation that investigated a series of five brutal rapes in central England from 1982 to 1995. The first victim was attacked in December 1982 in a parking lot in Bradford by a man with a Scottish accent. The offender forced his way into her car and then drove to a deserted airport where he raped her. A month later, the second victim was abducted in a similar manner from the parking lot of a Leeds hospital. Afterward, the offender abandoned her in an industrial area in the central part of the city. He continued this pattern, raping a woman in Leicester in May 1984 and another in Nottingham in May 1993. The last victim, a student, was attacked in a multistory parking garage in Leeds in July 1995. The offender put crazy glue over her eyes so she could not see him.

These crimes spanned multiple police jurisdictions, resulting in significant investigative problems. Finally, in 1996, the crimes were officially linked. The various police agencies formed Operation Lynx, which eventually became, with over 180 police officers from five different police forces, the largest manhunt in England since the Yorkshire Ripper inquiry. Investigators recovered DNA from semen and blood found at two of the crime scenes, but unfortunately, the offender was not on the UK National DNA Database. They also had a partial fingerprint from one of the victim's vehicles, though it lacked a sufficient number of points for an automated fingerprint identification system (AFIS) comparison. Investigators then decided to try a manual search of the fingerprint files of West Yorkshire Police – a jurisdiction of two million people. A prioritization scheme, one element of which was geography, was developed to focus the search.

In 1997, the Vancouver Police Department was asked to prepare a geographic profile for the case (Rossmo 2000). Unfortunately, even though each rape involved multiple locations, the crime series was spread over 13 years and several areas, suggesting the offender had operated from different bases. Generating separate geoprofiles for each city would have resulted in the critical loss of information. Investigators, however, had linked a stolen Ford Cortina to the second attack. The owner's credit card had been left inside the vehicle, and someone had used it to make numerous purchases throughout Greater Leeds. If this person was the rapist, then the geographic profile could also include these locations.

Proceeding on this basis, a geographic profile was prepared from the 20 locations of the Leeds crimes and credit card purchases. The result focused on two neighborhoods in central Leeds – Millgarth and Killingbeck. Consequently, the manual fingerprint search was narrowed by age (35–52 years), criminal record (minor offenses), residence area (Millgarth or Killingbeck), and Scottish origin, among other parameters. In March 1998, after 940 h of examining more than 7,000 prints, a match was made to a man named Clive Barwell. DNA subsequently confirmed he was the rapist. Barwell resided in Killingbeck, and his address was in the top 3.0 % of the geoprofile; his mother, who used to beat him when he was a child, lived in Millgarth. In October 1999, after Barwell pled guilty in court, he was sentenced to eight life terms in prison. He is still a suspect in the murder of a young woman.

Barwell may have been found sooner, but it turned out he was not Scottish. He had faked an accent in order to mislead police. He was also listed as being in prison during the Nottingham attack; an undocumented release gave him the opportunity to rape again.

In the hunt for Barwell, detectives engaged in 24,324 actions, knocked on over 14,000 doors, tested the DNA of 2,177 men, and reviewed an additional 9,945 suspects. A total of 33,628 names were entered in the inquiry's computer system, more than in any other case in British policing history. Operation Lynx is a dramatic example of the importance of suspect prioritization. Given the multijurisdictional nature and time span of the crimes, the manual fingerprint review would never have been successful without narrowed search criteria and a geographic focus.

Figure 1 shows the top 10 % of the three-dimensional jeopardy surface created with *Rigel* for Operation Lynx in 1997. Figure 2 shows the full two-dimensional geoprofile; the blue square in the southwest of central Leeds marks Barwell's address, and the one in the northeast, his mother's address.

Training

The Vancouver Police Department (VPD) implemented the first geographic profiling training program in 1997. The Royal Canadian Mounted Police (RCMP), Ontario Provincial

Police, British National Crime Faculty (now part of the Serious Organised Crime Agency or SOCA), and the Bureau of Alcohol, Tobacco, Firearms, and Explosives (ATF), among other agencies, have all had personnel trained in the methodology. The National Law Enforcement and Corrections Technology Center-Southeast Region (NLECTC-SE), working with the VPD, expanded geographic profiling training to property crime investigations in 2001 (Rossmo and Velarde 2008). Training is now available for police investigators and crime analysts internationally through various universities and police agencies. Over 600 people, representing 264 law enforcement, intelligence, and military agencies from 14 countries, including Australia, Canada, England, Wales, Korea, The Netherlands, South Africa, Sweden, Switzerland, Thailand, Turkey, and the United States, have now been trained in geographic profiling.

Future Directions

New Applications

Geographic profiling has turned out to be a robust and versatile methodology. Originally developed for analyzing serial murder cases, it was subsequently applied to rape, arson, robbery, bombing, kidnapping, burglary, auto theft, credit card fraud, and graffiti investigations. It has also been used to refine probability calculations in familial searches of DNA databanks (Gregory and Rainbow 2011). Some of the more interesting applications have included the geoprofiling of payphone locations in a murder case, cellular telephone switch tower sites in kidnapping cases, stores where bomb components were bought, locations of credit card purchases and bank ATM withdrawals in a rape case, and, in a historical analysis, the locations of anti-Nazi postcards left on the streets of Berlin, Germany, during the early 1940s. Geographic profiling has also found a number of innovative applications outside law enforcement, including uses in military operations, intelligence analysis, biology, zoology, epidemiology, and archaeology (Rossmo 2012).

Counterinsurgency and Counterterrorism

Traditional military responses to insurgency attacks in Iraq and Afghanistan, often quasi-criminal in nature, are usually not possible because of the civilian nature of the surrounding population. Counterinsurgency operations therefore require intelligence analysis and some form of investigative response. These attacks have underlying spatial and temporal patterns, enabling the use of geographic profiling by military analysts to determine the most probable locations of enemy bases (Brown et al. 2005; Kucera 2005). For example, urban and countryside insurgency problems have included attacks from improvised explosive devices (IEDs), vehicle bombs, land mines, rocket-propelled grenades (RPGs), mortars, and snipers. Insurgents typically obtain their heavier armaments and munitions from supply centers – homes, mosques, warehouses, and various other buildings. For geographic profiling purposes, the insurgent attack locations are equivalent to crime sites, while their supply centers are equivalent to offender bases.

Terrorism is a covert threat, and important patterns can be lost in the large volume of data collected by counterterrorism and intelligence agencies. Geographic profiling models can be used to prioritize suspects, tips, and leads. While it has seemed to some that terrorists, with transnational structures and decentralized networks, lack a geographic structure, it turns out that they do. Many minor terrorist actions are ordinary crimes, such as robbery, theft, and credit card fraud. Even major terrorist attacks of targets specifically selected for their symbolism require the establishment of terrorist cells in the areas of operation. In both situations, a geographic relationship exists, whether it is the target determining the locations of the terrorist cell sites or the terrorist cell sites determining the location of the target. Bennell and Corey (2007) retrospectively applied geographic profiling to terrorist bombings in France and Greece. They concluded that when appropriate assumptions are met (an area requiring further research), geoprofiles of terrorism can be accurate. Rossmo and Harries (2011) analyzed the site patterns of

Marxist and Jihadist terrorist cells in Turkey and Spain and found they possessed internal geospatial structures that could be quantified.

Biology and Epidemiology

Biologists have adopted geographic profiling models to the study of animal predation. The technique has been used to describe foraging patterns of different colonies of bats in Scotland (Le Comber et al. 2006), discriminate spatial search processes and predict nest locations of bumblebees (Raine et al. 2009; Suzuki-Ohno et al. 2010), and investigate the nonrandom nature of white shark attacks in South Africa (Martin et al. 2009).

Geographic profiling has also been used in epidemiology research to locate the origins of infectious diseases, including contaminated water sources for cholera and mosquitogenic breeding pools for malaria (Le Comber et al. 2011). Source populations of invasive species have been identified from geographic profiles of their current locations (Stevenson et al. 2012). The expansion of geographic profiling to these other domains demonstrates the reach and power of the environmental criminology approach.

Future Improvements

Future improvements in geographic profiling require an integration of both scholarly research and operational experience. One area that requires more study is the journey to crime. While many studies have measured the distance between offenders' homes and their crime sites, only a few have examined the exact nature of their crime journeys. Rossmo et al. (2011) mapped and analyzed the spatial-temporal patterns of a group of reoffending parolees on an electronic monitoring program with a global positioning system (GPS). Their research revealed the characteristics of actual crime trips and provided a more nuanced understanding of offenders' spatial patterns. Bernasco (2010) studied the spatial influence of offenders' residential history on their crime locations. He found past residences still influenced where a criminal offended provided he or she had recently moved and had lived in the prior residence for a period of time. Summers et al. (2010) used maps in interviews of convicted property offenders, gaining insights into how they view space and search for criminal targets.

Combining geo-demographics and other area-based information with the point pattern analysis of geographic profiling is another approach with potential (Rossmo et al. 2004). Levine and Block (2011) developed a Bayesian approach to geographic profiling that integrates historic offender residence data with journey-to-crime estimations. However, while Bayesian models may be useful for prioritizing geographic areas, they cannot be used for suspect prioritization as their calibration is based on known offender residences.

Conclusion

The stranger nature of serial crime creates challenges for police investigations. Geographic profiling can help detectives prioritize suspects and manage information in such cases. It is only one of several available tools and is best employed in conjunction with other police methods. As addresses are a common database element, geographic profiling can be used as a decision support tool in a variety of contexts. The overall geographic pattern of a crime series is just as much a clue as any of those found at an individual crime scene.

Related Entries

▶ Biological Geographical Profiling

Recommended Reading and References

Beauregard E, Rossmo DK, Proulx J (2011) A descriptive model of the hunting process of serial sex offenders: a rational choice approach. In: Natarajan M (ed) Crime opportunity theories: routine activity, rational choice and their variants. Ashgate, Surrey

Bennell C, Corey S (2007) Geographic profiling of terrorist attacks. In: Kocsis RN (ed) Criminal profiling: international theory, research, and practice. Humana Press, Totowa, pp 189–203

Bernasco W (2010) A sentimental journey to crime: effects of residential history on crime location choice. Criminology 48:389–416

Brantingham PL, Brantingham PJ (1981) Notes on the geometry on crime. In: Brantingham PJ, Brantingham PL (eds) Environmental criminology. Sage, Beverly Hills, pp 27–54

Brantingham PJ, Brantingham PL (1984) Patterns in crime. Macmillan, New York

Brantingham PL, Brantingham PJ (1993) Environment, routine and situation: toward a pattern theory of crime. In: Clarke RV, Felson M (eds) Routine activity and rational choice. Transaction, New Brunswick, pp 259–294

Brown RO, Rossmo DK, Sisak T, Trahern R, Jarret J, Hanson J (2005) Geographic profiling military capabilities. In: Final report submitted to the Topographic Engineering Center, Department of the Army, Fort Belvoir

Clarke RV, Felson M (eds) (1993) Routine activity and rational choice. Transaction, New Brunswick

Daniell C (2008) Geographic profiling in an operational setting: the challenges and practical considerations, with reference to a series of sexual assaults in Bath, England. In: Chainey S, Tompson L (eds) Crime mapping case studies: practice and research. Wiley, Chichester, pp 45–53

Davies A, Dale A (1995) Locating the stranger rapist (Special Interest Series: Paper 3). Police Research Group, Home Office Police Department, London

Gregory A, Rainbow L (2011) Familial DNA prioritization. In: Alison L, Rainbow L (eds) Professionalizing offender profiling: forensic and investigative psychology in practice. Routledge, Abingdon, Oxfordshire, pp 160–177

Kind SS (1987) Navigational ideas and the Yorkshire Ripper investigation. J Navig 40:385–393

Knabe-Nicol S, Alison L (2011) The cognitive expertise of geographic profilers. In: Alison L, Rainbow L (eds) Professionalizing offender profiling: forensic and investigative psychology in practice. Routledge, Abingdon, Oxfordshire, pp 126–159

Kucera H (2005) Hunting insurgents: geographic profiling adds a new weapon. GeoWorld 37(30–32):37

Le Comber SC, Nicholls B, Rossmo DK, Racey PA (2006) Geographic profiling and animal foraging. J Theor Biol 240:233–240

Le Comber SC, Rossmo DK, Hassan AN, Fuller DO, Beier JC (2011) Geographic profiling as a novel spatial tool for targeting infectious disease control. Int J Health Geogr 10:35–42

Levine N, Block RL (2011) Bayesian journey-to-crime estimation: an improvement in geographic profiling methodology. Prof Geogr 63(2):1–17

Martin RA, Rossmo DK, Hammerschlag N (2009) Hunting patterns and geographic profiling of white shark predation. J Zool 279:111–118

Raine NE, Rossmo DK, Le Comber SC (2009) Geographic profiling applied to testing models of bumble-bee foraging. J R Soc Interface 6:307–319

Rossmo DK (1995) Place, space, and police investigations: hunting serial violent criminals. In: Eck JE, Weisburd DL (eds) Crime and place: crime prevention studies, vol 4. Criminal Justice Press, Monsey, pp 217–235

Rossmo DK (2000) Geographic profiling. CRC Press, Boca Raton

Rossmo DK (2011) Evaluating geographic profiling. Crime Map J Res Pract 3:42–65

Rossmo DK (2012) Recent developments in geographic profiling. Policing J Policy Pract 6:144–150

Rossmo DK, Harries KD (2011) The geospatial structure of terrorist cells. Justice Q 28:221–248

Rossmo DK, Velarde L (2008) Geographic profiling analysis: principles, methods, and applications. In: Chainey S, Tompson L (eds) Crime mapping case studies: practice and research. Wiley, Chichester, pp 35–43

Rossmo DK, Davies A, Patrick M (2004). Exploring the geo-demographic and distance relationships between stranger rapists and their offences (Special Interest Series: Paper 16). Research, Development and Statistics Directorate, Home Office, London

Rossmo DK, Lu Y, Fang T (2011) Spatial-temporal crime paths. In: Andresen MA, Kinney JB (eds) Patterns, prevention, and geometry of crime. Routledge, London, pp 16–42

Stevenson MD, Rossmo DK, Knell RJ, Le Comber SC (2012) Geographic profiling as a novel spatial tool for targeting the control of invasive species. Ecography 35:704–715

Summers L, Johnson SD, Rengert GF (2010) The use of maps in offender interviewing. In: Bernasco W (ed) Offenders on offending: learning about crime from criminals. Willan Publishing, Cullompton, Devon, pp 246–272

Suzuki-Ohno Y, Inoue MN, Ohno K (2010) Applying geographic profiling used in the field of criminology for predicting the nest locations of bumble bees. J Theor Biol 265:211–217

Geographical Analysis

► Crime Mapping
► Inferential Crime Mapping

Geographical Offender Profiling

► Applied Geographical Profiling

Geography of Crime

▶ Crime Mapping
▶ Inferential Crime Mapping

Geography of Crime and Disorder

Anthony Bottoms
University of Cambridge, Cambridge, UK
University of Sheffield, Sheffield, UK

Synonyms

Ecology of crime; Environmental criminology; Socio-spatial criminology

Overview

The study of the geography of crime has a long history, stretching back to the early nineteenth century. After a quiescent period in the middle of the twentieth century, it has now again become a field of strong research activity. However, the interests of scholars in this field have often been divergent: for example, some have been motivated by very practical (crime-preventive) concerns, while others are more interested in explanatory theory; some have mainly focused on neighborhoods, but others see greater merit in studying micro-locations. Not infrequently, work in these different traditions has been pursued in relative isolation, leading to some fragmentation of the field. However, there is now a growing interest in developing a more integrated understanding of geographical criminology (Bottoms 2012; Weisburd et al. 2012; Wikström et al. 2012). This chapter is written from that integrative perspective, with a special focus on arguably the three most important strands of research in the recent geography of crime, namely, the "hotspots approach," the "neighborhood effects tradition," and the "signal crimes perspective." The chapter is organized into three main sections, focusing respectively on space, social structure, and social action.

Three Research Approaches in the Contemporary Geography of Crime

The "hotspots" approach within criminology effectively began with a seminal paper by Sherman et al. (1989), which showed that 50 % of the crime-related calls made to the police in Minneapolis related to just 3.3 % of the micro-locations in the city. This degree of concentration of crime seemed to offer significant potential for targeted crime prevention programs, and such programs have subsequently been successfully developed by scholars and practitioners working in the hotspots tradition (Braga and Weisburd 2010; Weisburd et al. 2010). Indeed, it has been recently argued that "the police can be more effective if they shift the primary concerns of policing from people to places" (Weisburd et al. 2010, p. 7).

The "neighborhood effects tradition" is older and can be traced back to the work of the Chicago School of Sociology in the 1930s. The argument of the Chicagoans was that "socially disorganized neighborhoods" produced more juvenile delinquents than other residential areas with similar populations; in other words, that aspects of the social structures and the culture of certain neighborhoods could lead to the development of criminality in some individuals who would otherwise have remained crime-free. For a time, the existence of such "neighborhood effects" (as opposed to "individual effects" or "family effects") was challenged, but in light of further research evidence, they can now be regarded as clearly established (for a review see Bottoms 2012). For example, in a paper based on data from the Pittsburgh Youth Study, Wikström and Loeber (2000) showed that, after controlling for various individual "risk factors" for offending (such as impulsivity, poor parental supervision, and low school motivation), neighborhood factors remained causally important in generating adolescent criminality, at least in the poorest neighborhoods.

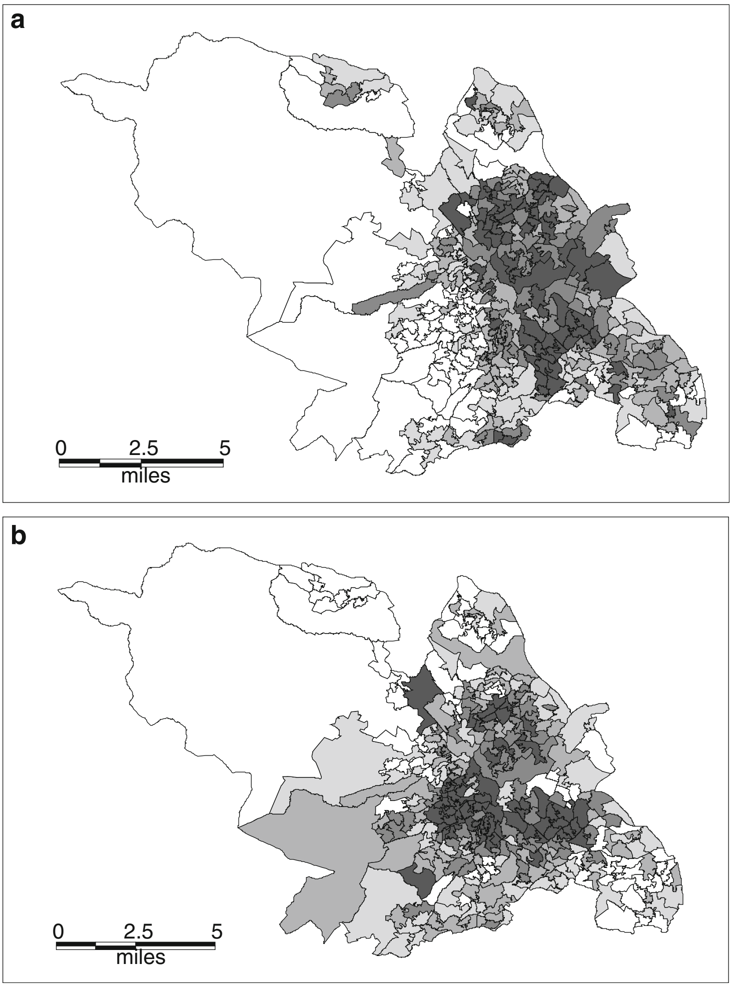

Geography of Crime and Disorder, Fig. 1 City of Sheffield: geographical distribution of offender residences and burglary victimizations, September 2004 – August 2006. (**a**) Map of offender residences (in quintiles, *highest rate darkest*) (**b**) Map of victimizations for offenses of residential burglary (in quintiles, *highest rate darkest*). Note: both maps are based on UK Census lower level super output areas (Source: Bottoms 2012, p. 459; maps prepared by Dr. Andrew Costello)

There are two important differences of empirical focus as between the hotspots approach and the neighborhood effects tradition. First, as their names imply, there is a difference in geographical scale: hotspots are micro-locational, while the study of neighborhoods entails a locationally wider research focus. Secondly, the dependent variable for hotspots researchers is the *criminal event located in geographical space*, while the neighborhood effects tradition, being concerned with potential neighborhood influences on the criminality of individuals, focuses particularly on the *geographical location of offenders' homes*. Yet despite these differences, in real-life research contexts the two approaches are empirically interconnected, because one of the best-established and durable findings in the geography of crime is that – for reasons to be discussed later – *many offenses are committed close to offenders' homes*. This is illustrated in Fig. 1 which displays data on the location of offender residences and of recorded residential burglary victimizations in Sheffield, England, a city with a population of half a million. Sheffield is well known as a "city of two halves," with the population of the eastern half being substantially poorer, and having a lower life expectancy, than those living in the western half. Figure 1a shows that offender residences are very disproportionately concentrated in eastern Sheffield. One might therefore suppose that many offenders intending to burgle might travel the short distance from east to west Sheffield to take advantage of the (on average) much more lucrative theft opportunities available in the houses in the west. However, although there is clearly some movement in this direction (see Fig. 1b), most of the areas with high victimization rates for residential burglary are, as with offenders' residences, found in the east. Given data of this kind, it is now well established that an important risk factor for a high area victimization rate is that the location lies close to an area of high offender residence. That simple fact is one of the most important reasons why a better integration of research from the "hotspots" and "neighborhood effects" traditions is desirable.

The third strand of research, "the signal crimes perspective," is the most recent. It draws in part from the field of study known as *semiotics* ("the science of signs and symbols"), which focuses especially on the operation of such symbols in communication. Semiotics is relevant to geographical criminology because the signs emitted at particular locations can send messages that significantly affect behavior; for example, research in the Netherlands has shown that signs of disorder in public places (such as uncleared litter) tend to encourage further breaches of norms (including theft), while norm-compliant behavior (e.g., someone sweeping up litter) can encourage other norm-compliant behavior (e.g., helping a stranger who has dropped fruit in the street) (Keizer et al. 2008; Keizer 2010). According to Martin Innes, the originator of the signal crimes perspective (SgCP), criminology lacks "a coherent explanation of the public understanding of crime and disorder and how such understandings are imbricated in the wider symbolic construction of social space." Substantively, SgCP proposes that "*some crime and disorder incidents matter more than others to people in terms of shaping their risk perceptions*" in visiting defined locations (Innes 2004, p. 336, emphasis added). Thus, for example, three spouse murders in a month in a medium-sized town would be very unusual, but would not necessarily create widespread fear, or a sense of threat, in the community at large, because they would be seen as "private matters." By contrast, the abduction and murder of a local schoolgirl on her way to school would almost certainly generate much more fear, and a sense of threat, in the area, because of the *signal* it would transmit about potential risks in the community. In light of this signal, ordinary people might freshly consider as "risky" certain places, people, or situations that they might encounter in their everyday lives; hence signals are seen in SgCP as social-semiotic processes by which particular crimes and disorders might have a disproportionate effect in terms of fear and perceived threat, often in relation to specific locations. It will accordingly be noted that the focus of this third theoretical approach is – by contrast with the first two approaches – on neither area crime rates nor area offender rates *as such*, but rather on the symbolic meanings that people might attach to specific locations as a result of certain acts occurring in or near them or the characteristics of people believed to be present in the location.

From the above account, it will be clear that the three research traditions described, while distinct from one another, are in no sense necessarily contradictory. Coherent synthesis of the approaches, within defined contexts, is therefore in principle certainly possible, and aspects of this complementarity will hopefully become more evident in the ensuing sections of this chapter. In these sections, research results on the geography of crime and disorder will be considered within the framework of the three core dimensions of research in human geography: *space*, *social structure*, and *social action*.

Space

Human beings exist only within physical bodies. Although, in the Internet age, an increasing number of activities can be undertaken in virtual space, physicality remains important for many human activities, notably travel and the use of public space. This first section therefore considers physical places and spaces and their relevance within geographical criminology.

The physicality of humans is, however, not their only characteristic. Uniquely within the animal kingdom, humans possess the gift of language, and the ability to be self-reflexive about the situations in which they find themselves. In considering humans' use of space, therefore, it is necessary to address *perceptions* of locations, as well as the physical context. Within this dual framework, three topics will be discussed, namely, crime opportunities, hotspots, and the meaning of space.

Crime Opportunities

It is now well established in psychological research that much behavior is situation specific, and that people often act differently if some aspect of a particular type of situation is altered. A striking demonstration of these points was provided by Ronald Clarke and Pat Mayhew (1988), who showed that the suicide rate had dropped suddenly in Britain in the 1970s when the source of domestic gas supplies was switched from coal gas (toxic) to natural gas from the North Sea (nontoxic) and that there was no plausible explanation of this decline other than the change in the nature of the gas supply. In other words, even people desperate enough to try to end their own lives often did not turn to alternative methods when the immediate situational context was altered. Given this and other similar evidence, Ronald Clarke and others have made a special study of the specific situations in which different offenses are committed, with a view to modifying those situations to produce a reduction in crime – an approach known as *situational crime prevention* (StCP). Typically, this might involve strategies such as making goods harder to steal by strengthening security at potential targets, reducing the availability of the means to commit certain crimes (e.g., by restricting sales of guns or knives), or environmental management (e.g., separating the fans of opposing teams at football matches). Such strategies have often led to beneficial results, and there are now many reports of successful StCP initiatives (Felson and Boba 2010: ch. 10). Some of these initiatives are explicitly locational, and a good example is the strategy known as "alley-gating." In working class residential areas in British industrial cities, a favored style of pre-Second World War housebuilding was "back-to-back terraces," that is, rows of terraced houses opening directly on to the street, with small yards or gardens at the rear, backing on to an identical set of terraces in the next street. Normally, a "back alley" is located between the yards/gardens of the two rows of houses. Crime analysts noticed that the entry point for house burglaries in such areas was very often from the rear, and this led to the prevention strategy of "gating" the entrances to the back alleys. Evaluations of such initiatives have shown them to be successful in reducing burglary; moreover – and congruently with the evidence from the gas suicide example – there was no evidence of displacement of burglaries to nearby residential areas (Bowers et al. 2004).

Closely related to the StCP approach is *routine activities theory* (RAT), whose central proposition is that "the probability that a [criminal] violation will occur *at any specific time and place* might be taken as a function of the convergence of likely offenders and suitable targets in the absence of capable guardians" (Cohen and Felson 1979, p. 590; see also Felson and Boba 2010). This approach, which incorporates a specifically spatial dimension (see the italicized phrase), usefully divides the concept of "opportunity" into two component parts ("targets" and "guardianship"), but more importantly, it draws attention to the fact that the routine activities of a population can – often unwittingly – create or restrict criminal opportunities in defined locations. To take an example from a British study, multistory parking lots at train stations used by commuters (where few owners visit the lot during the working day) were shown to have a higher victimization rate than similar parking lots attached to shopping malls (where shoppers are leaving and returning to their vehicles all day, providing a flow of natural surveillance or guardianship). Many examples of a similar kind are provided by Felson and Boba (2010); as they memorably phrase the matter, these examples show how "noncriminal realities give rise to criminal events" (p. 205).

It is therefore a serious mistake to see criminal opportunities as existing in a social vacuum. This point is further illustrated through Patricia and Paul Brantingham's seminal contribution to geographical criminology, first formulated 30 years ago (Brantingham and Brantingham 1981), and now incorporated within a broader *crime pattern theory* (Brantingham and Brantingham 2008). These authors postulated that most offenders, like most people, feel much more comfortable in areas that they know reasonably well; in consequence, even when – for example – potential burglars are engaging in a search pattern for a suitable target, it is hypothesized that they will usually not wish to venture into residential areas that are completely unknown to them. In other words, the suggestion is that offenders will usually commit crimes in areas already known to them through their routine activities. (Oddly, routine activities theorists until recently paid little attention to this point, being more interested in how targets and guardianship are influenced by routine activities.) Subsequent empirical research evidence has strongly supported the Brantinghams' hypothesis (see especially Wikström et al. 2012: ch. 7), and this body of work clearly helps to explain why many offenses are committed not far from offenders' homes (see above), as well as in other locations in which they regularly spend time, and where there are criminal opportunities (such as downtown areas and leisure outlets). These social realities therefore illustrate the point that a so-called opportunity for crime in a given area is, in the full sense, only an opportunity when it is perceived as such by someone who (i) might enter the area and (ii) is willing to consider turning the opportunity into an act of crime. Otherwise identical physical "opportunities," if located in different social contexts, might differ greatly in the number of passersby who would consider committing a crime at the micro-location, and such differences will almost certainly be reflected in their aggregate victimization levels.

Sometimes, even more complex narratives emanate from discussions of "opportunities for crime"; street lighting is an important example. Improved street lighting was originally advocated as a situational crime prevention strategy in the belief that better lighting would increase the number of people using the location at night (thus providing improved guardianship of the area) and would also make potential offenders more visible (thus making them more easily detectable if they committed a crime). Both these effects were thought likely to deter potential offenders from committing crimes in the relevant area, because the opportunities for successful lawbreaking had been reduced. However, the actual research evidence is more complex. A Campbell Collaboration review of relevant empirical studies has concluded that, while "improved street lighting significantly reduces crime," it was also the case that, in areas with improved lighting, "night-time crimes did not decrease more than daytime crimes" (Welsh and Farrington 2008, p. 2). This finding presents problems for a simple opportunity theory, because if enhanced deterrence were the only mechanism in play, one would expect crime reductions in better-lit areas during hours of darkness, but no difference in crime rates during daylight hours. Accordingly, the Campbell reviewers concluded that the daytime effect was possibly occurring because "improved lighting signals community investment in the area and that the area is improving, leading to increased community pride, community cohesiveness and informal social control." This interpretation therefore seems to embrace not only "opportunities" but also the more normative language of the signal crimes perspective (note the word "signal" in the sentence quoted above, although the SgCP approach is not explicitly mentioned by the reviewers). However, the temporal sequencing of possible "deterrent" and "signal" effects has been insufficiently studied, so – as the authors make clear – further research is needed. But the juxtaposition of the two approaches in the Campbell Review is of great interest.

Hotspots

Like those who have focused on crime opportunities, criminologists who have developed the "hotspots approach" have concentrated on micro-locations. Indeed, they have very usefully shown that the neighborhood level of analysis can be misleading in the geographical study of crime events, because – for example – what might seem overall to be a "crime-prone neighborhood" often contains micro-locations with very varied levels of crime, including some low-crime locations.

These findings have led to a justifiable emphasis on the desirability of using small units of analysis when studying crime events (Weisburd et al. 2009). Accordingly, in a major recent research study in Seattle, to be described later, Weisburd and colleagues (2012) adopted as their unit of analysis the "street segment," that is, both sides of a street between two intersections in a standard North American street grid.

A main motivating reason for the development of the hotspots approach has been crime prevention. However, before researchers could confidently recommend to policymakers a preventive strategy based on hotspots, several issues beyond the simple demonstration of crime concentrations needed to be explored. These included the following:

(i) Perhaps the identified hotspots were ephemeral, so that a different set of hotspots would appear if one were to conduct the analysis for a later year or a series of years.

(ii) Perhaps crime prevention initiatives targeted on specific places would not actually reduce crime at the hotspot.

(iii) Even if the crime prevention effort was successful at the hotspot, perhaps those committing crimes in that location would simply move their activities elsewhere ("crime displacement").

These were reasonable questions, but subsequent research has convincingly dispelled doubts. Briefly, longitudinal analyses have now shown that hotspots in given cities are usually enduring, not ephemeral; rigorous evaluative studies have demonstrated that "hotspot policing" crime prevention strategies can be effective; and displacement is rarely a problem (Braga and Weisburd 2010; Weisburd et al. 2010). This body of work constitutes an impressive contribution to scholarship by hotspots researchers. But in addition – and of great interest to the quest for integration under discussion in this essay – hotspots researchers have begun to pursue the question of the *legitimacy* of crime prevention initiatives at crime hotspots. (On legitimacy and criminal justice, see Bottoms and Tankebe 2012). For example, Braga and Weisburd (2010: ch. 6) have set out in schematic form three different models by which police could in principle seek to control crime at hotspots: these are the "traditional" (or "reactive") model, in which police respond to incidents as they occur; a proactive "strong enforcement" model (such as the widely discussed "zero tolerance" approach to policing); and a proactive "situational prevention" model, focusing crime prevention efforts on targeted micro-locations identified through prior crime analyses. The third of these is the authors' preferred model, and a principal reason given for rejecting the "strong enforcement" model is its potential for legitimacy deficits: "overly aggressive and indiscriminate arrest-based strategies are more likely to generate community concern and poor [police-public] relations" (Braga and Weisburd 2010, p. 229). On this view, therefore, when assessing the policy-relevant features of specific locations, it is necessary to consider not only objective matters such as the concentration of crime events but also more normative and perceptual issues such as the way in which law enforcement agents are viewed in the locality (which will include the historical legacy of past law enforcement, still present in the memories of current residents). Such matters will, however, very often relate not simply to the hotspot itself, but to the wider socio-structural context in which it is set (e.g., a particular policing approach might have been adopted for a whole neighborhood, not simply a specific hotspot). In such circumstances, it becomes necessary to widen the spatial scale of the issues to be addressed when considering policy responses to crime in the hotspot.

Despite the unquestioned importance of the hotspots approach as a development within geographical criminology, early research in this field was subject to some criticism, particularly concerning two alleged omissions: first, that (apart from the discussion of legitimacy) hotspots researchers had paid little attention to hotspots from the potential offender's point of view and, secondly, that they had largely failed to analyze a key explanatory question, namely, *why* do some micro-locations have high crime, while others do not? Fortunately, both of these omissions have been addressed in more recent research.

An important study of hotspots from the offender's perspective was undertaken by Peter St. Jean (2007) in a socially very deprived high-crime neighborhood in Chicago. In qualitative interviews, the researcher elicited persistent offenders' views about optimum micro-locations for drug dealing and robbery. A strong "opportunity" dimension to offenders' locational preferences was shown, and the most favored street blocks for committing crimes were busy road intersections containing commercial premises such as grocery stores, liquor stores, and check-cashing outlets, all of which bring together "people who can be clandestine clients of drug dealers or easy targets for robbers" (p. 197). In more residential areas, however, a further "opportunity" variable seemed to be in play, namely, the vigilance or otherwise of residents, since "watchful neighbors…sometimes serve as effective constraints against neighborhood robberies" (p. 208).

The issue of watchful neighbors brings into focus the theory known as *collective efficacy*, particularly developed by Robert Sampson. "Collective efficacy" is defined as *the institutional ability to achieve what a group or community collectively wishes to achieve*; and in criminological research it has been principally utilized in research in small residential areas. It is normally measured with survey items that tap into two linked matters: first, the willingness of residents to intervene for the common good in certain defined situations (such as children spray-painting graffiti on a building) and, secondly, the existence or otherwise of conditions of cohesion and mutual trust among neighbors (on the assumption that one is unlikely to take action in a neighborhood context where the rules are unclear and residents mistrust one another). Various empirical studies have confirmed that, in many local contexts, there is evidence for the crime-reductive effectiveness of collective efficacy (Sampson 2006).

The second main omission in early hotspots research (i.e., why do particular micro-locations have high crime?) has been principally addressed through two major research studies using different methodologies, one in the USA and one in England.

The American research was a multivariate correlational study using a unique 16-year dataset for Seattle for the years 1989–2004 inclusive (Weisburd et al. 2012). As previously noted, the unit of analysis in this study was the micro-locational "street segment." Congruently with earlier research, the study found (i) a strong concentration of crime in "hotspots" (5–6 % of street segments each year accounted for c.50 % of crime incidents) and (ii) with some exceptions, a "tremendous stability of crime at street segments over a long period of time" (p. 56). Eight "trajectory [crime] patterns" were identified. Four of these (comprising 84 % of street segments) were stable over time, ranging in level from "crime-free" to "chronic crime"; the remainder showed increasing (4 %) or decreasing (12 %) trajectories. A multivariate analysis was then undertaken, in an attempt to uncover the factors that would predict a particular trajectory pattern in any given street segment. A principal feature of this analysis was an assessment of the comparative explanatory strength of (i) variables measuring the "opportunity perspective" (roughly, the kind of variables utilized in StCP and RAT) and (ii) variables measuring the "social disorganization perspective" (roughly, variables traditionally used in "neighborhood effects" research). Principal findings in the Seattle research were (i) that both these sets of variables were important contributors to explaining differential crime trajectory patterns in street segments, (ii) that opportunity variables had greater explanatory power than social disorganization variables in predicting variations in crime between micro-locations at any one point in time, and (iii) that social disorganization variables had greater explanatory power than opportunity variables in predicting changes in crime levels in micro-locations over time.

The first author of the Seattle study (David Weisburd) has been one of the leading researchers within the "hotspots" tradition. It is therefore noteworthy that in this study he and his colleagues deliberately set out to "broaden and expand the [traditional] theoretical foundations of the criminology of place" (p. 13), particularly by including variables representing the "social disorganization" (or "neighborhood effects") tradition – a tradition that, as they observe, had previously been "virtually ignored" by hotspots researchers (p. 43). Moreover, the results of the Seattle analyses strongly support the view that both "opportunity" and "neighborhood effects" dimensions need to be included within any full answer to the question "why are certain micro-locations hotspots, and others not?"

The second piece of recent research that aims to explain "why hotspots are hotspots" is the Peterborough Adolescent Development Study (PADS+) (Wikström et al. 2012). This study, which was conducted in the medium-sized city of Peterborough, England (pop. c. 165,000), differs in three main respects from the Seattle study: it uses a different methodology, it is significantly more theory-driven, and – because it reports data from the early stages of a longitudinal study of criminal careers – it is restricted to juveniles. Despite these differences, its results in no way conflict with the principal conclusions of the Seattle study.

PADS+ was conducted within the theoretical framework of Per-Olof Wikström's "Situational Action Theory" (SAT). SAT postulates that to understand how criminality arises, one needs to consider two matters: first, an individual's *propensity* to commit crime (defined as consisting of both the person's morality and his/her ability to exercise self-control) and, secondly, his/her *exposure to situations of varying criminogenic potential* (where *criminogenic exposure* is defined as "encounters with settings in which the (perceived) moral norms and their (perceived) levels of enforcement...encourage breaches of rules of conduct...in response to particular opportunities or frictions": Wikström et al. 2012, p. 17). The theory therefore envisages a possible *interaction* as between criminal propensity and environmental (geographical) conditions.

Two main types of "criminogenic exposure," both with a strong locational component, were considered in the Peterborough study, namely, that a young person spent time "unsupervised with peers in an area with high public entertainment" or "unsupervised with peers in a residential area with poor collective efficacy." Both of these conditions, for different reasons, provide increased opportunities for offending. Using a detailed interview-based "time budget" methodology, and measuring crime by self-reported offending on the days for which time budgets were compiled, the researchers found that, overall, a substantially greater number of crimes were committed when young people were in conditions of hypothesized criminogenic exposure than when they were not. For those with a "strong" criminal propensity, the difference in the rate of crime commission when "in" or "not in" a criminogenic environment was particularly marked (for both exposure conditions). On the other hand, those with a low crime propensity committed no crimes whether or not they were in criminogenic exposure conditions; for them, their strong anti-crime dispositions were able to override any temptations offered in criminogenic exposure conditions. Thus, the results show a genuine interaction effect as between the propensity of individuals and the environmental contexts of these locations.

Interestingly, the Peterborough results also showed that those with a high criminal propensity tended to spend more time in conditions of high criminogenic exposure than other young people. This finding therefore emphasizes – as the research of the Brantinghams had suggested in a different way many years before – the particular importance for the geography of crime of the routine activity patterns of those with a high criminal propensity.

The Symbolic Meanings of Places

In developing the "signal crimes perspective" (see above), Martin Innes and his colleagues conducted detailed qualitative interviews in 16 neighborhoods in England and Wales, asking representative respondents what they would identify as the key *potential threats to neighborhood safety* in their area (described in the theory as *signal crimes* and *signal disorders*). Not surprisingly, there was a good deal of local variation in the perceived threats identified in different localities. Nevertheless, a striking and largely unexpected result (although see not dissimilar concerns raised by Wilson and Kelling 1982) was that various kinds of *physical and social disorders*, not all of which were criminal acts, featured particularly strongly as perceived threats in the public responses *in all areas*. (Examples of such disorders included youths hanging around in groups, drug detritus, litter/graffiti, vandalism, and public drinking.) Indeed, such items were, in almost all areas, perceived as a greater threat to the area than residential burglary. These results appeared to be mainly explicable by the fact that the listed disorders are *events occurring in public space*, often with a strongly repetitive dimension: such events therefore seemed to send a powerful message to residents that "my area is out of control." As one respondent put it to Innes (2004, p. 348):

> Yes, it is daft, it is almost daft, but graffiti is the thing that sort of bothers me more, because it is in my face every day. I mean obviously rape and murder are more horrendous crimes, but it is graffiti that I see.

Thus, even quite minor incivilities in public space can, especially if persistent, be perceived as significant threats to peaceable daily living. This evidence is consistent with more general social scientific research results showing that disruptions of people's everyday routines can be perceived as significantly threatening to their sense of ontological security. Such results accordingly suggest that some earlier scholars in geographical criminology have neglected an important dimension of lived experience by focusing exclusively on crimes, rather than on both crimes and disorders.

Social Structure

The most important social-structural topic of relevance for geographical criminology is that of *social disadvantage*. In some countries, including the United States, this topic has a strongly ethnic dimension, but for reasons of space the discussion here will focus on disadvantage more generally.

A consistent finding in geographical criminology is that offenders' residences are very disproportionately located in more socially disadvantaged areas of cities (see, e.g., Fig. 1a). Given that many offenses are committed close to offenders' homes, it is also the case that socially disadvantaged residential areas usually have higher crime victimization rates than richer areas (see Fig. 1b), and, at least in England, national crime surveys also show that they have the highest rates of repeat disorders. These facts seem to most geographical criminologists to require a serious engagement with issues of social structure, although this view is not universally shared (see Felson and Boba 2010, p. 206, final sentence).

Of these various indicators, the resident offender rate normally shows the starkest contrast between rich and poor areas, and this raises the possible existence of a "neighborhood effect" on offending in such areas. In considering this issue, great care in interpretation is required, as may be seen from the recent meticulous research on juvenile delinquency in the Peterborough study (Wikström et al. 2012). On the one hand, social disadvantage is not explicitly part of the explanatory model tested and affirmed in this research (see above: the model, based on Situational Action Theory, is focused on criminal propensity and criminogenic exposure). Moreover, by no means all of the young people with a high criminal propensity lived in deprived neighborhoods. Nevertheless, in this study the researchers found that "population social disadvantage was a particularly important predictor of the number of resident young offenders" in a given area and that "social disadvantage and its influence on the efficacy of key social institutions such as the family and the school are clearly implicated in the understanding of why some young people develop a stronger crime propensity" (Wikström et al. 2012, p. 239). Clearly, some complex social dynamics lie behind these important results, and the authors intend that future analyses from the Peterborough study will explicate the issues more fully.

From results such as these (see also Wikström and Loeber 2000), it seems a reasonable hypothesis that moving children and young people from very socially deprived areas to less deprived areas should reduce their criminality, as well as providing other benefits to their families. These matters were tested in a large-scale randomized controlled trial known as the "Moving to Opportunity" (MTO) experiment, which was conducted in five major cities in the United States (Baltimore, Boston, Chicago, Los Angeles, and New York). So far as juvenile criminality was concerned, the results were unexpectedly mixed – on the best available evidence, the hypothesis of reduced misbehavior was confirmed for girls/young women, but boys/young men moving to less disadvantaged areas appeared to become *more* delinquent (Kling et al. 2007; Briggs et al. 2010; Clampet-Lundquist et al. 2011). Possible reasons for this gender difference, and in particular the counterintuitive result for boys, are discussed at the end of this entry.

In considering issues of social disadvantage, it should not be forgotten that patterns of both informal and formal social control can vary in areas of differing social status and that these are of potential criminological significance. This point was drawn starkly to criminologists' attention in the pathbreaking book by Lawrence Sherman (1992) analyzing the results of a series of experiments on police responses to domestic violence. In these randomized controlled trials, six American police departments responded to nonlife-threatening domestic violence incidents (where the assailant was still at the home) *either* by arresting the suspect *or* by some measure short of arrest. The results were more complex than had been anticipated. In summary, Sherman characterized three cities (Minneapolis, Colorado Springs, and Miami) as having shown a deterrent effect following arrest, while in three other cities (Omaha, Charlotte, Milwaukee) arrest appeared to have produced what was called a "backfiring effect" – that is, it seemed to *increase* subsequent violence among arrested suspects, by comparison with controls. This pattern of results naturally raised the question: "what…factors might explain the differences between the deterrent and backfiring effect cities?" The data showed that the strongest single area difference was to be found "in the proportion of black suspects, which averaged 63 % in the three backfiring cities…compared to 36 % in the three deterrent effect cities" (Sherman 1992, p. 18). In other words, differing *community characteristics* (including social disadvantage) in the six cities seemed to be at least partially relevant to the explanation of the results. On an individual (as opposed to areal) basis, the study also found that arrest was most likely to suppress subsequent violence if the suspect had high "stakes in conformity" (= employed and married) and least likely to suppress it when he had no such stakes (= unemployed and unmarried). Aspects of the interpretation of these results remain tentative, but it does seem from this research that the familial and communal relationships within which a person is embedded can significantly affect the manner in which he responds to an arrest (in principle, from deep shame to resentment against unfair police practices), and that such differences could affect the probability of subsequent violence against a partner. Referring back to an earlier discussion, one might add that the extent to which the policing in a neighborhood appears to its residents to be *legitimate* also seems potentially relevant – and this also might vary by social-structural context.

Before leaving the topic of social structure, a word must be added on the subject of *housing markets*. Housing markets are of great importance for the study of crime and criminality in residential areas, because the mechanisms of such markets strongly frame both who enters a given area as a resident, and how easy or difficult it is to leave the area if and when one wishes to do so. Of course, housing markets are often strongly influenced by economic factors (the rich buy large houses in desirable areas; the poor do not). However, economic factors are certainly not the only variables in play in shaping the operation of a housing market in a given area, for example, cultural questions or the reluctance of older residents to move might be very important in influencing exit decisions, and in some countries the allocative rules of public (or social) housing authorities are also relevant. In consequence, it is even possible for two adjacent areas with nearly identical economic and demographic characteristics to have radically different victimization and resident offender rates, flowing from the direct and indirect consequences of housing market processes (Bottoms et al. 1989). In short, as Robert Sampson (2009a, p. 90) has rightly commented, “because housing markets act as a mechanism of allocation, . . .they. . .need to be better integrated into sociological and criminological theory.”

Social Action

The third and final conceptual dimension to be considered, *social action*, is by its nature more dynamic than the other two – social actions are, after all, usually taken in an attempt to achieve certain consequences. Accordingly, this section contains more discussion of longitudinal theorization and research than previous sections.

The most famous longitudinal theory in contemporary geographical criminology is the *broken windows hypothesis*, originally formulated by Wilson and Kelling (1982) albeit in a somewhat informal and discursive manner. Taylor (2001, p. 98) has usefully formalized the proposed sequential stages of the Wilson-Kelling thesis; slightly modified, these stages are:

(i) Unrepaired low-level signs of incivility (e.g., broken windows, graffiti) appear in an area.
(ii) The local residents tend to withdraw from public areas and become fearful.
(iii) Emboldened antisocial locals commit more petty crimes, signs of incivility grow.
(iv) Local residents become more fearful and withdraw more.
(v) [Frequently but not inevitably]: Serious offenders from other areas note the lack of guardianship in the area and move in.

Thus, according to Wilson and Kelling (1982, p. 32), “the citizen who fears the ill-smelling drunk, the rowdy teenager or the importuning beggar is not merely expressing his distaste for unseemly behavior, he is also giving voice to a bit of folk wisdom that happens to be a correct generalization – namely, that serious street crime flourishes in areas in which disorderly behavior goes unchecked.” As regards policy responses, for Wilson and Kelling (p. 36) “the key is to identify neighborhoods at the tipping point- where the public order is deteriorating but not unreclaimable.” But others have gone further, linking the “broken windows” approach to zero tolerance policing, as in the following statement by former British Prime Minister Tony Blair (2010, p. 493): “if you tolerate the low-level stuff, you pretty soon find the lawbreakers graduate to the high-level stuff. So cut it out at source: tolerate nothing, not even painting a street wall or dropping litter.”

Empirically speaking, the evidence about broken windows is mixed. On the one hand, the Dutch research previously cited shows that a breach of a social norm (such as uncleared litter) does indeed, in the short term, encourage the breach of other norms. On the other hand, those empirical studies that have attempted to evaluate the broken windows hypothesis on a long-term basis have found little supporting evidence for the proposition that “disorder leads to serious crime” (Sampson and Raudenbush 1999; Taylor 2001; Harcourt and Ludwig 2006; Sampson 2009b).

Although further research is needed, on present evidence two matters seem particularly likely to explain the apparent tension between the short-term and long-term effects of “broken windows.” The first is that the final stage of the proposed Wilson-Kelling sequence (“serious offenders from other areas move in”) might often fail to materialize. St. Jean’s (2007) Chicago research (see above) showed that such offenders were usually more interested in nonresidential than residential areas as attractive sites for their activities; moreover, physical (as opposed to social) incivilities in residential neighborhoods were of no special interest to them. Secondly, the suggestion made by Tony Blair and others that one must “cut [incivilities] out at source,” or a downward spiral to “the high-level stuff” will inevitably occur, is incorrect. On the contrary, the literature contains several case examples of communities acting successfully at a later stage to ameliorate a downward spiral.

One such example is described in research by Taub, Taylor, and Dunham (1984). These researchers showed that two Chicago neighborhoods *with high crime rates* not only received positive "satisfaction with safety" scores in a residents' survey but also had rapidly appreciating residential property values. How did this unusual combination of factors arise? In both the areas concerned, "there [were] highly visible signs of extra community resources being used to deal with the crime problem" (Taub et al. 1984, p. 172). This observation has to be set within the context of another finding in the study, namely, that, in evaluating local areas, people make a generalized, *gestalt* judgement, taking into account a range of positive and negative factors, of which crime (and, by implication, disorder) is one. Within such a framework, the injection of "visible signs of extra community resources" is clearly a potentially positive factor.

The term "visible signs," used by these authors, is highly congruent with the later development, within the signal crimes perspective, of the concept of "control signals" (used in SgCP alongside the concepts of "signal crimes" and "signal disorders" – see above). "Control signals" are acts (particularly those taken by officials or by informal community leaders) that communicate ("send signals") to the general public, in a way that helps to promote the general sense of order in a neighborhood. An example within the Taub, Taylor, and Dunham research concerned the neighborhood that includes the University of Chicago. Anxious about decline in the area, university managers invested heavily in the local urban infrastructure and helped to introduce various initiatives that directly addressed citizens' worries about safety in public space – such as a private security force, 24-h "safety buses" and emergency telephones (Taub et al. 1984, pp. 99–102). This whole package of measures seemed to send a strong "control signal"; for while crimes such as burglary remained high (pp. 21–22), as previously noted, the area was perceived by the residents as safe, with rapidly appreciating property values. Thus, utilizing the trilogy of concepts around which this chapter is structured (space, social structure, social action), this was a powerful example of how *social action* can help to shape the future of a neighborhood. It seems likely – although we do not yet have fully confirmed evidence of this – that less dramatic but not dissimilar kinds of social action are not infrequently taken in other apparently declining local areas, with positive effects; and perhaps the apparent effect of improved street lighting on daytime crime levels (see above) results from similar social mechanisms. If these suggestions are correct, they could of course help to explain the lack of long-term evidence in support of the broken windows hypothesis.

It is not accidental, however, that in the above example the University of Chicago is a *corporate actor*. It is much easier for a corporate actor than an individual resident to send successful control signals; indeed, as Taub and his colleagues point out, for an individual household in a declining area, the instrumentally rational course of action will usually be to try to make a fast exit from the area. Thus, effective social action to halt social decline will probably normally require a lead to be taken by a public agency, a major private institution, or an active community organization. Thereafter, however, it is likely that local citizens will respond positively to the lead and feel emboldened to play their part in neighborhood renewal: recall that Keizer's (2010) experimental research in the Netherlands suggests that actions to reinforce prosocial norms, as well as actions that breach such norms, can have a "contagious" effect on the behavior of others in a locality.

Yet a final caveat is required. Sherman's (1992) study of domestic violence showed that police actions such as arrest can sometimes "backfire"; and the same, it would seem, is true also of residents' social action. For various technical reasons, it is not easy to provide a definitive explanation of the strikingly different results for boys and girls in the MTO experiment (see more fully Bottoms 2012). The most plausible account of the counterintuitive result for MTO boys is to be found in a detailed qualitative study in two cities (Baltimore and Chicago: see Clampet-Lundquist et al. 2011). According to this research, one of the factors in operation in these cities was that the adolescent boys who moved to less poor neighborhoods continued, in their new environments, to practice the dominant leisure activity they had learned in their previous home area, namely, "hanging out" with one another in public places. But, in the changed context, they found themselves less accepted by the local residents (possibly because these residents were wary of adolescent males who were known to have moved from stigmatized public housing projects). Moreover, the new areas had higher collective efficacy than the baseline neighborhoods and therefore – see earlier discussion – more adult interventions with teens. In consequence, "a negative side of collective efficacy" seemed to be apparent, that is, adult interventions were made, but they led to resentment from the boys (Clampet-Lundquist et al. 2011, p. 1171). This example therefore once again illustrates the very complex social dynamics of neighborhoods – dynamics that need to be taken fully into account in the study of the geography of crime and disorder.

Conclusion

Within the broad field of the geography of crime, a great deal of excellent research has been completed in the last two decades. However, much of this work has been conducted in relative isolation from those working in other parts of the field. A strong case can therefore be made for the adoption of a more integrated approach. Such an approach should in particular seek to synthesize more fully the work in the "hotspots," "neighborhood effects," and "signal crimes" traditions, each of which has contributed powerfully to our overall understanding of geographical aspects of crime and disorder. Such a synthesis should lead to improved understanding of the very complex social dynamics in play within this field. Potentially, also, it could have an important payoff in improved crime prevention, given the persuasive case that has recently been made for "the importance of place in policing" (Weisburd et al. 2010).

Related Entries

- ▶ Agent-Based Models to Predict Crime at Places
- ▶ Broken Windows Thesis
- ▶ Crime Location Choice
- ▶ Crime Mapping
- ▶ Criminal Careers of Places
- ▶ Criminology of Place
- ▶ Disadvantage, Disorganization and Crime
- ▶ Early Chicago School Theory
- ▶ Geographic Profiling
- ▶ History of Geographic Criminology Part I: Nineteenth Century
- ▶ History of Geographic Criminology Part II: Twentieth Century
- ▶ Hot Spots and Place-Based Policing
- ▶ Law of Crime Concentrations at Places
- ▶ Neighborhood Effects and Social Networks
- ▶ Public Housing and Crime Patterns
- ▶ Reassurance Policing and Signal Crimes
- ▶ Routine Activities Approach
- ▶ Situational Action Theory
- ▶ Situational Crime Prevention

Recommended Reading and References

Blair T (2010) A journey. Hutchinson, London

Bottoms AE (2012) Developing socio-spatial criminology. In: Maguire M, Morgan R, Reiner R (eds) The Oxford handbook of criminology, 5th edn. Oxford University Press, Oxford

Bottoms AE, Mawby RI, Xanthos P (1989) A tale of two estates. In: Downes D (ed) Crime and the city. Macmillan, London

Bottoms AE, Tankebe J (2012) Beyond procedural justice: a dialogic approach to legitimacy in criminal justice. J Crim Law Criminol 102:119–170

Bowers KJ, Johnson SD, Hirschfield A (2004) Closing off opportunities for crime: an evaluation of alley-gating. Eur J Crim Policy Res 10:285–308

Braga AA, Weisburd DL (2010) Policing problem places. Oxford University Press, Oxford

Brantingham PL, Brantingham PJ (1981) Notes on the geometry of crime. In: Brantingham PJ, Brantingham PL (eds) Environmental criminology. Sage, Beverly Hills

Brantingham PL, Brantingham PJ (2008) Crime pattern theory. In: Wortley R, Mazerolle L (eds) Environmental criminology and crime analysis. Willan Publishing, Cullompton

Briggs X de S, Popkin SJ, Goering J (2010) Moving to opportunity: the story of an American experiment to fight ghetto poverty. Oxford University Press, Oxford

Clampet-Lundquist S, Edin K, Kling JR, Duncan GJ (2011) Moving teenagers out of high-risk neighborhoods: how girls fare better than boys. Am J Sociol 116:1154–1189

Clarke RV, Mayhew P (1988) The British gas suicide story and its criminological implications. In: Tonry M, Morris N (eds) Crime and justice: a review of research, vol 10. University of Chicago Press, Chicago
Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44:588–608
Felson M, Boba R (2010) Crime and everyday life, 4th edn. Sage, Thousand Oaks
Harcourt BE, Ludwig J (2006) Broken windows: new evidence from New York City and a five-city social experiment. Univ Chic Law Rev 73:271–320
Innes M (2004) Signal crimes and signal disorders: notes on deviance as communicative action. Br J Sociol 55:335–355
Keizer K (2010) 'The spreading of disorder', proefschrift, University of Groningen. http://irs.ub.rug.nl/ppn/329349163
Keizer K, Lindenberg S, Steg L (2008) The spreading of disorder. Science 322:1681–1685
Kling JR, Liebman JB, Katz LF (2007) Experimental analysis of neighborhood effects. Econometrica 75:83–119
St Jean PKB (2007) Pockets of crime: broken windows, collective efficacy and the criminal point of view. University of Chicago Press, Chicago
Sampson RJ (2006) Collective efficacy theory: lessons learned. In: Cullen FT, Wright JP, Blevins KR (eds) Taking stock: the status of criminological theory. Transaction Publishers, New Brunswick
Sampson RJ (2009a) Analytic approaches to disorder. Br J Sociol 60:83–90
Sampson RJ (2009b) Disparity and diversity in the contemporary city: social (dis)order revisited. Br J Sociol 60:1–31
Sampson RJ, Raudenbush SW (1999) Systematic social observation of public spaces: a new look at disorder and crime. Am J Sociol 105:603–651
Sherman LW (1992) Policing domestic violence: experiments and dilemmas. Free Press, New York
Sherman LW, Gartin PR, Buerger ME (1989) Hot Spots of predatory crime: routine activities and the criminology of place. Criminology 27:27–55
Taub R, Taylor DG, Dunham JD (1984) Paths of neighborhood change: race and crime in urban America. University of Chicago Press, Chicago
Taylor RB (2001) Breaking away from broken windows: Baltimore neighborhoods and the nationwide fight against crime, grime, fear and decline. Westview, Boulder
Weisburd D, Bernasco W, Bruinsma GJN (eds) (2009) Putting crime in its place: units of analysis in geographic criminology. Springer, New York
Weisburd D, Groff ER, Yang S-M (2012) The criminology of place: street segments and our understanding of the crime problem. Oxford University Press, Oxford
Weisburd D, Telep CW, Braga AA (2010) The importance of place in policing: empirical evidence and policy recommendations. Swedish National Council for Crime Prevention (Brå), Stockholm
Welsh BC, Farrington DP (2008) Effects of improved street lighting on crime. The Campbell Collaboration, Oslo
Wikström P-OH, Loeber R (2000) Do disadvantaged neighborhoods cause well-adjusted children to become adolescent delinquents?: a study of male serious juvenile offending, individual risk and protective factors, and neighborhood context. Criminology 38:1109–1142
Wikström P-OH, Oberwittler D, Treiber K, Hardie B (2012) Breaking rules: the social and situational dynamics of young people's urban crime. Oxford University Press, Oxford
Wilson JQ, Kelling G (1982) Broken windows. The Atlantic Monthly (March), pp 29–38

German Police Until 1918

Herbert Reinke
Bergische Universität Wuppertal, Wuppertal, Germany

Overview

Federalism is, except for periods during the twentieth century (Nazi Germany and communist East Germany after 1945) a key characteristic of political systems in Germany throughout the nineteenth and twentieth centuries. Due to this focal pattern of the German political system, there had been polices of the respective federal units during these periods, but no national German police. Early nineteenth-century administrative and political reforms established the police in many German states as a state prerogative, which manifested itself by the establishment of Gendarmerie forces. In Prussia, this Gendarmerie was to police the countryside, but in relation to major cities, the Prussian state commissioned municipalities with the establishment and the maintenance of an urban police, establishing state police forces proper in a few cities only. A comparable mixed system was established in some other German states as well. A characteristic, which was shared by almost all the polices in Germany throughout the nineteenth century and even during the first decades of the twentieth century consisted of the continuity of a comprehensive notion of police and policing, deriving from the Ancien Régime notion of "Polizey," beyond a modern, on law, and order maintenance focused understanding of police. Attempts were made by different actors (the police, the courts, members of the legal sciences) to narrow this comprehensive notion, but did not really touch upon the reality of police and policing in nineteenth-century Germany.

Due to "order" being a core issue of policing in Germany during the nineteenth century, crime and crime control was not a major concern of the police during that period. Until the First World War, only a minority of the major German cities had specialized crime investigation units among their municipal police forces. Because of this situation, the crime investigation department of the Berlin police headquarter became a key institution, as it professionalized increasingly its expertise for investigating complex and spectacular criminal cases.

The emergence of a political police and of political policing at the beginning of the nineteenth century was a correlate of the continuity of monarchist, bureaucratically framed regimes after the Napoleonic wars. Political policing regained momentum with the rise of the organized labor movement during the second half of the nineteenth century, when industrialization did set in Germany. Although political policing was seen by contemporaries as a significant policing pattern, the application of a comprehensive notion of police remained a key strategy for policing the emerging industrial society.

Police and Policing During the First Half of the Nineteenth Century

During the first decade of the nineteenth century, the Prussian government, because of the necessities to modernize Prussia on the background of its defeat in the war against Napoleonic France, a number of institutional reforms and governmental reorganizations were inaugurated, which were to bring the Prussian state into a position of meeting the challenges posed by Napoleonic France. These reforms touched upon the police in Prussia as well. During the Ançien Régime period, the police had either been a matter of urban authorities or had been carried out in the countryside by noble estate owners as part of their privileges. The early nineteenth-century reforms established the police as a state prerogative. This state prerogative was put into practice by the establishment of a – quantitatively – large-scale Gendarmerie, which, in the Prussian case, was to police the countryside, while, as far as the cities were concerned, the Prussian state commissioned municipalities with the establishment and the maintenance of an urban police, establishing state police forces in a few cities only. During the first half of the nineteenth century, the Prussian state established state polices only in those cities, which were seen as hotspots of political opposition and dissent. While the state polices were on the budgets of the Prussian state, it were the municipalities themselves which had to pay for their municipal police. During the 1860s, the Prussian state withdrew its state polices from a number of cities, leaving it to the municipalities to police them with municipal polices, the reason for this withdrawal being a mixture of budgetary considerations and changed security evaluations by the Prussian government.

The installation of the Gendarmerie, following the French example, was meant as introducing the state monopoly of power in the Prussian countryside, where the police until the reforms of the early years of the nineteenth century had been in the hands of the estate-owning nobility. Gendarmerie forces were created – except for Austria and the Habsburg Lands – in most of the German states. The reasons were similar across the states: The central governments in these states, often with new boundaries as a result of the Vienna congress, wanted a penetration of their authority into the whole of the state, were seeing the Gendarmerie as an appropriate instrument for creating some sort of identity among the inhabitants of the respective state and were using this force for policing the vagrant rural underclass, which was seen by contemporaries as a major threat. In Prussia, the introduction of the Gendarmerie was answered by the estate-owning rural nobility with (in the end successful) large-scale protest and opposition, resulting in a massive decrease of the Gendarmerie force and a restitution of the Ançien Régime model of policing the countryside by combining property titles (estate ownership) and police functions. In Prussia, the figures for the Gendarmerie forces were, in comparison to Gendarmerie forces in other German states, extremely low: In 1848, there were 8 gendarmes per 100,000 of population in Prussia, 52 per 100,000 of population in Baden, 52 in Bavaria, 40 in 1839 in Brunswick, 23 in Hanover, 10 in Saxony, and 25 in Württemberg (Jessen 1991). When looking at these figures and when taking into account, that the Prussian state left the policing of cities in most cases to communal or municipal police forces, then it might look as if Prussia had been underpoliced during the first half of the nineteenth century. But that does not take into account the massive presence of the military in very many Prussian cities. The Prussian military, having garrisons in quite a number of Prussian cities, did have in the respective cities policing prerogatives. In 1840, 53 % of the urban population of Prussia lived in a city, which held a garrison. State and communal or municipal officials continuously relied on the military in any case they considered an emergency for public order and safety. The interventions of the military into everyday urban policing matters have been interpreted as a significant militaristic impact on policing in Prussia during the nineteenth century (Lüdtke 1989). Additionally, the Berlin police headquarters became a key institution for the policing of politics in Germany (Fricke 1962).

Political Policing in Nineteenth-Century Germany

The emergence of a political police and of political policing in Germany at the beginning of the nineteenth century is a correlate of the continuity of monarchist, bureaucratically framed regimes after the Napoleonic wars. During this so-called period of "restauration" after 1815, the monarchist regimes of the German states aimed at defending their regimes against the growing demands of the bourgeoisie for political participation. In order to control political opposition, the Prussian government established after 1815 a number of government commissions. A similar model for supervising the political opposition was adopted by the German Federation, which did set up interstate government commissions not only for controlling opposition movements, but for disseminating information among the governments of the German Federation. The strategies of these government commissions were rather retrospective, focusing on political incidents, which had happened already, although elements of proactive control patterns were introduced as well already during this period. One of the Prussian government commissions became charged with examining the political reliability of teachers, which were to be employed, and with examining the political positions of candidates for the clergy and the judicial professions. This cooperation was enforced after the defeat of the revolutionary movements in the German states in 1848/1849. During the 1850s, when it became visible, that the revolutionary initiatives of 1848/1849 would not gain ground and significance again, these control activities were gradually reduced.

Political policing regained momentum with the rise of the organized labor movement. Together with the introduction of repressive legislative instruments for suppressing the organized labor movement after the foundation of the German Empire, a specific state police commissioners were introduced, which were to supervise the organized labor movement and organized groups of foreigners, such as members of the Polish minority. Municipalities were reluctant to establish political police branches within their municipal police forces, but the Prussian government circumvented these problems by establishing this nucleus of a political police strictly as a state police, thus creating the first political police proper in Germany.

Policing Nineteenth-Century Industrialization and Urbanization

Prussia and other parts of Germany have been, in comparison with countries and regions in Western Europe, latecomers as far as industrialization and urbanization were concerned. In Germany, large-scale industrialization and urbanization did not set in but after the beginning of the second half of the nineteenth century. In Prussia, the major industrial areas with their rapidly growing urban conglomerations were mostly covered up to the first years of the twentieth century by communal or municipal police forces or by a few gendarmes only. It was only in 1909, that in some towns of the Ruhrgebiet area in the West of Prussia, the municipal police forces were transformed into proper state polices (Funk 1986; Jessen 1991; Spencer 1992). In the cities with municipal police forces, the police remained principally part of the state prerogative on the monopoly of power. In these cities, the mayor represented this prerogative by being the official head of the local or municipal police. While the mayor was the head of the municipal administration, elected by the city council he was at the same time, with respect to his duties as head of the police in the respective city, executing state functions and had thus to obey state orders. Due to these specific patterns of police organization, the Prussian state could theoretically intervene in everyday municipal matters, although in practice, the cities had many possibilities for diverting or impeding such attempts. But more important was that municipalities had police competences of their own. Prussian legislation in 1850 and legislation on the municipalities had given the municipalities in Prussia ample competences regarding police matters, as it had provided the legal frame for issuing police ordinances concerning local matters (Jessen 1991). The municipalities operated within this frame extensively by issuing thousands of police ordinances until the First World War.

The overall appearance and organization of municipal police forces in Imperial Germany, not only in Prussia, was connected to Ançien Régime model of the police, the "Polizey." Under this model, policing covered virtually the entire complex of local administration, control, and regulation. This global police model was most visible in the towns that had municipal police forces. By the middle of the nineteenth century, the police in Elberfeld, which had been by that time one of the early industrial towns in Prussia, comprised among others the night watch; the police of strangers and the police of pubs and inns; the health-police and the life insurance-police; the medical police; the police of the religious cults; the educational police; the police of morals and order; the trade's and business' police; the police of measures and weights; the building-police; the fire-police; the police of the roads; the market-police; the police of the hunt; the forest-, field-, and agricultural police; the river-police; and the police of the dogs (Reinke 1992, 1993; Spencer 1992). Shortly before the First World War, communal policemen still spent much of their time carrying out routine duties relating to these duties, such as distributing tax and registration forms, enumerating buildings and animals, keeping lists of children to be vaccinated, and men liable to military service, providing a variety of identity papers needed by citizens, making certain various fees were paid, and licensing and inspecting a wide range of enterprises and activities (Spencer 1992). By the beginning of the twentieth century, numerous older police responsibilities had either lapsed or been transferred to newly emergent, specialized agencies. Fire fighting and street lighting, for example, were removed from police budgets. As urban service providers became more diverse and specialized, social welfare workers began to assume some functions policemen had performed in the past. In some cities, female welfare workers were taking over from the police the placement and supervision of foster children as well as the counseling of "morally endangered" women and girls and of unmarried couples living together, hoping that this approach would prove more beneficial than a police intervention only. At the same time, however, that the police cooperated with the new social workers in regulating the behavior of those who were poor, transient, or otherwise outside the limits of respectable society, determining what behavior was permissible and what forbidden. All this meant that despite reassignment of duties, little overall narrowing of police functions occurred in the cities (Spencer 1992). The shedding of some older police assignments was counterbalanced by the imposition of new ones. Communal policemen remained deeply involved in many aspects of the daily lives of the city's inhabitants, especially when they were members of the working class. Among police responsibilities were those arising from official concern with public health and sanitation. Before the First World War already, more and more health-related supervising was transferred to technically trained specialized personnel, but much remained for the police to do. Policemen inspected latrines and monitored the removal of wastes. Continued police responsibility for the control of stray dogs was also in part a health measure, linked to prevention of rabies. Furthermore, concern with public health could be used as one justification for the inspection of housing, another major task for the police, even where specialized urban housing inspectors had been appointed. In Düsseldorf in 1909, the police inspected 16,828 dwellings, whereas the city's housing inspectors examined only 2,891. The municipal administration described inspections by the police as a service to tenants, making it possible for renters to bring pressure on their landlords to correct deficiencies (Spencer 1992). At the same time, however, inspections of rental housing provided the authorities with an excuse for increased police observation of the private lives of poorer citizens. Special attention was paid to households including boarders, since worried bourgeois observers of working-class life saw the presence of outsiders in a family setting as fraught with potential for increased immorality. However, efforts to prevent overcrowding and thereby presumably to lessen temptations to promiscuity often proved impossible to enforce because of the lack of alternative accommodations in rapidly growing industrial cities. Representative of the catchall nature of many responsibilities assigned to the police was the decision to entrust them with the maintenance of lost-and-found services. In Elberfeld, the police also took the initiative in establishing and staffing facilities enabling the public to call taxis, with operating costs being paid by the taxi owners. In their expanding regulation of markets, local transport, and insurance contracts, the police cast themselves as the guardians of consumer interests, often arousing the resentment of small traders, carters, and cab drivers in the process (Spencer 1992).

Among the most troublesome of requests for intervention that came to the police were those linked to master-servant relations. Until 1918, police remained responsible for monitoring contracts entered into by domestic servants. When difficulties arose, both sides to a dispute might turn to the police for support. The police could try to mediate or impose fines to force compliance with their decisions, but they could not really make someone work in a household who refused to do so. Disputes between landlords and tenants and among neighbors often proved equally intractable.

With requests for intervention and assistance coming from so many sources, the communal police tried to free themselves from some of their most menial and irksome tasks. But their successes remained limited. In Elberfeld in 1906, for instance, the police informed a local school that they could not, without detracting from "real police responsibilities," repeatedly delegate patrolmen to conduct truants to school. In that same year, however, Düsseldorf policemen accompanied 567 reluctant boys and girls to schools of various kinds (Spencer 1992). During the decades, which preceded the First World War, statistical figures for issuing fines against parents and for accompanying boys and girls to schools showed that these tasks were a main activity of municipal patrolmen (Reinke 1991).

As in the past, many new tasks were transferred to the police simply because they represented a widely dispersed body of public servants, available on an around-the-clock, 7-days-a-week basis, and with closer contact to the daily life of the general populace than most public employees had. In addition, utilization of the communal police for a wide variety of highly visible services helped legitimate their costly and intrusive presence in the eyes of city councilmen and local taxpayers. A considerable number of the nineteenth-century police functions remained effective until the end of the Empire in 1918, and Elberfed was not unique in this among the major industrial cities in the West of Prussia. However, the continuity of the "Polizey" approach should not be misconstrued simply as conservative administrative behavior. Often, the various police were the nuclei of service-orientated departments of the city administrations. Attempts to narrow and to concentrate police functions on the maintenance of law and order were launched in the decade before the beginning of the First World War, but they became decisive only at the beginning of the Weimar Republic.

Policing the Industrial Society

Policing industrial society implied two approaches for the Prussian police during the Empire: on the one hand, in the tradition of the old "Polizey" model of policing, it meant control of the consequences of industrialization, in terms of environmental damage, control of the conditions of labor, etc., with some of these police activities being executed by the so-called trade and business police ("Gewerbepolizei"). But in practice, the police restricted their activities in this field to controlling and supervising traditional crafts and enterprises rather than intervening in the structures and the development of industrial enterprises.

Policing the industrial society meant of course on the other side the control of the working-class population with its unions and political organizations, mainly the Social Democrats: In the industrial society of Imperial Prussia, they were the main object of the control and repressive activities of the police. The legal system of the German Empire allowed for an extensive control of the working class by the police. Repressive legislation against the Social Democrats ("Sozialistengesetz") constituted the basis for a real police war against the unions and the Social Democrats, which took place from the end of the 1870s to the end of the 1880s. But when this legislation did not find a parliamentary majority any more, other measures were applied: When used against the activities of the organized working class, the exercise of extensive police discretion was very often approved by court decisions; legislations on associations ("Vereinsgesetz") allowed all kinds of police interventions and repression of the political mobilization and organization of the working class; the "Gewerbe-Ordnung," the law that regulated business activities and labor conditions, provided possibilities for breaking strikes (Saul 1974).

The early 1870s after the Franco-Prussian war were the period, when industrialization accelerated its pace in Prussia, resulting in a period of cumulative developments. The economic boom immediately after the Franco-Prussian war was followed by a deep depression with all its economic and social consequences. During these years, there was not only industrial activity, but also the spread of moral panics, and the rise of moral entrepreneurs launching their campaigns. Fears of danger to law and order and the increase of criminality were major themes of public discussion in this period. Although not all the higher public officials in the Prussian state administration took the moral panics of this time at face value, the state police administration seized the opportunity for promoting an increase in the number of royal policemen at the beginning of the 1880s. Starting with a ratio of one policeman to 1,500 inhabitants in the 1870s, the Ministry of the Interior improved this ratio gradually to 1:700 by the beginning of the new century. The state administration put pressure on the municipal police forces to come up with an improved police-population ratio as well, and by the 1880s a ratio of 1:1,500 was recommended for municipal police forces (Funk 1986). The symbolic representation of police forces, which were to expand were often given expression by impressive new police headquarters buildings. Significantly, the police headquarters in Berlin was the second largest building after the royal palace. The rise in the figures of the police personnel did not match the figures for the growth of the population, in particular in the industrialized West, along the Rhine and in the Ruhr area, were the figures for police personnel per 100,000 of population fell or stagnated during the 1880s before rising steadily again in the 1890s after strike actions by the miners in the Ruhr area (Jessen 1991).

Another measure for coping with the threats, which were seen as endangering society during this period resulted in institutionalizing a separate crime investigation branch within the police, the Kriminalpolizei. The Berlin Kriminalpolizei engaged itself intensively into the scientific policing during the later years of the nineteenth century and became not only in Germany as police institution a celebrity. But the impact of the institutionalization of a separate crime investigation branch remained limited. It remained very much Berlin based with some follow-up establishments in other very large German cities. But only few police headquarters in other Prussian cities institutionalized similar crime investigations in their departments or did so shortly before the First World War only (Funk 1986).

A major test of the effectiveness of police control and intervention was the first big miners' strike in 1889. This strike, whose scale was without precedent in German history, caught the municipal and state administrations completely unprepared. The Prussian Army was called upon; the intervention of the military finally led to a breakdown of the strike movement. But because of the strict military logic, the Prussian army applied against the strike movement and because of complex and difficult Prussian state authority – Prussian military relations, Prussian state authorities refrained until the First World War from using extensively the military as police in industrial disputes (Johansen 2005). Instead, strategies for enhancing the presence of state police in Ruhr area were enhanced. One measure was the establishment of state police administrations in three cities in the Ruhr area in 1909 (Funk 1986; Jessen 1991). Up to that time, these towns had been policed by their own municipal police forces. To this point, the state administration had pursued very ambivalent policing strategies in the Ruhr area, the "Wild West" of the Empire. Before the outbreak of the strike in 1889, policing in this area was carried out at a very low quantitative level, as compared to other parts of Prussia. While the industrial and urban development in the Ruhr area was already progressing rapidly, the Prussian state administration maintained the previously appropriate rural-cantonal administrative organization of this region, with a state representative at its head and a few policemen and gendarmes for policing it. This led to administrative patterns such as industrial villages with more than 100,000 inhabitants. To cope with this insufficient situation, a second measure, besides introducing a state police in three Ruhr area cities, was discussed, which consisted of deploying more Gendarmerie in the area. This measure, which was never put into practice, included an increase of the gendarm-population ratio per 100,000 of population to 28 and the establishment of Gendarmerie barracks. But Prussian state administrators nevertheless headed for turning the Gendarmerie gradually into Prussia core police force for policing the industrialized society. Shortly before the First World War, 45 % of Prussia's Gendarmerie was based near to Prussia's major industrial centers, in the East in Silesia, around the heavily industrialized capital Berlin, and in the West in the Ruhr area and along the Rhine (Funk 1986).

The Recruitment and the Training of Nineteenth-Century Police Personnel

Who could become a member of the police force in Prussia? There are some resemblances to patterns to be observed in France during the nineteenth century. In theory, previous service in the Prussian army for 9–12 years was an absolute prerequisite for being recruited into the uniformed Police in Prussia. Thus, the ranks of the uniformed policemen on the beat were supposed to be filled with non-commissioned officers (NCOs). Military service as a prerequisite for admission to the police served several functions: the NCO policeman was supposed to represent king and state in the everyday life of the citizen. His superiors expected him to show his derived authority, and, if necessary, to compel compliance to this authority from the public. On the other hand, long training within the military hierarchy was supposed to have made him an obedient servant to his superiors and a reliable instrument for maintaining law and order. But in reality, this recruitment pattern worked partially only. Although the civil service did have a high social ranking, the position of a uniformed patrolman had a low ranking on this list. Getting a job in the police was not what an army NCO necessarily looked for after years of military service. For many NCOs, the job of a policeman was only a transitory phase on the way to a quieter existence in the civil service than the police service could offer. In Berlin, for example, whose royal police was to serve as an example for other polices in Prussia, more than 2,000 policemen left the police force during the 1890s. The Berlin police force totaled about 4,000 men in the middle of the 1890s. A quarter of those who left the police went into other positions within the civil service. Due to these turnover rates, a large number of police posts were permanently vacant. Not only in Berlin, but other Prussian cities also had to cope with this situation. During the 1890s and again during the first decade after the turn of the century, the Prussian state administration tried to solve the turnover problem by reducing the military service requirements for employment in the police force. As a result, more men were drawn into the police service. The 9–12-year's military service remained nevertheless the ideal requirement looked for when conceiving a real Prussian patrolman on the beat. Those policemen who had acquired their post on the basis of the reduced requirements were often considered as some sort of second-class policemen (Reinke 1991).

Vacancies in the civil service, the police included, were announced and advertized by state and city administrations in public lists and journals. But the police posts offered regularly outnumbered applicants from the NCO's ranks. Cities with municipal police forces therefore reduced their employment requirements by recruiting men for the police service who had only fulfilled their obligatory military service after conscription, rather than service as regulars. By the 1880s, and much more so after 1890, cities in the West of Prussia, along the Rhine and in the Ruhr did find it increasingly difficult to recruit candidates with the adequate military background. While some cities reported during the 1890s that they were still managing to staff the municipal police positions accordingly, other cities found the supply of candidates with the requisite military experience falling ever more short of their needs. By 1911, the police department in Düsseldorf, one of the largest cities in the Prussian West, with 46 patrolmen's positions to fill, reported that of 500–600 applicants, only 22 had the adequate military background. Thus, despite persistent pressure from the Prussian state to seek out NCOs for police service, the cities in the Prussian West turned increasingly to the local wage-earning population for recruits. Urban administrators usually did so reluctantly, sometimes continuing to see the ideal policeman not just as someone shaped by long years of military discipline but also as someone from the outside, preferably of rural or small town origin. But the reality was that most applicants were local residents with only minimal military experience. As a consequence, the possibility of recruiting of policemen having personal contacts with organized workers loomed ever more threatening (Spencer 1992).

The training of policemen was carried out on the job and was oriented along the practical requirements of the man's on the beat everyday activities. The patrolman had to get acquainted with his precinct and with the things he was supposed to look after when on the beat. A number of "formalities" were also briefly taught, such as some basic ideas about the penal code of the Empire, the police ordinances that had been issued for the respective city or community, etc. Apart from that, the writing of dictations was part of the training. Often this was essential because the men's ability to write seems to have suffered considerably during the long years of military service. But all in all, serious training did not take place. Around the turn of the century, the Prussian state government as well as municipal police administrations acknowledged an urgent need for an improvement in the policeman's qualifications. In order to ensure better-qualified personnel, police schools were set up. And additionally, the military habits of the ordinary policeman were no longer deemed sufficient for the handling the everyday problems the police encountered in the growing urban contexts. The more Prussian cities grew and the more complex urban society became, the more qualifications were required from the policeman beyond his authoritarian and military attitude. In 1899, the first police school was established in Prussia, as a school for the Gendarmerie. The Gendarmerie took the initiative, since for this force, which was still part of the military, the problem was the most urgent. In 1901, the first police school for municipal police personnel was set up in Düsseldorf. The costs of the schools were covered by those municipalities who sent their police men there. But usually these schools were relatively small and few policemen were sent to there. In general, municipal administrations accepted the necessity for improving the qualifications of their police personnel, but for financial reasons, they kept down the number of men they sent to these schools.

Similar institutions were being established in other Prussian districts as well at about the same time, beginning with Berlin in 1895. Police schools were intended not only to impart necessary knowledge and approved attitudes but also to raise police prestige. Increased formal training (whatever its content) would bridge the gulf between policemen and respected representatives of the Prussian state. In Düsseldorf, the course for patrolmen lasted 2 months and that for senior officers for 3 months. Students were required to live at the school so that its influence could prevail around the dock. The cities paid the costs for the patrolmen, fearing that if the men were required to use their own resources they would fall into debt, a situation the policemen were strongly encouraged to avoid. The senior officers, expected to come from somewhat more well-off families, had to pay for their own instruction. Cities tried to protect their investment in the candidates by stipulating that those who left their departments after less than 5 years had to repay all or part of the costs of their schooling. Also, attendance at the school was typically reserved for those recruits who had already completed 6–12 months of service. As justification for this fiscally prudent move, police administrators argued that schooling was more meaningful if it followed a substantial period of practical experience. By 1906, the state district administration in Prussia had stipulated that in cities of 10,000 or more, patrolmen either had to attend the police school or pass an examination before being confirmed in their posts. Supporters of the Düsseldorf school were dismayed to find that many communal police departments, to save the cost of instruction, either tried to hire recruits who had already attended a police school elsewhere or else encouraged the taking of the examination. To make certain that the Düsseldorf school had enough students to pay for itself, the Prussian provincial administration before the war was contemplating eliminating the examination option. As a step toward making municipal policemen more credible as rule enforcers by increasing the likelihood that they knew and understood the rules and what they were doing and for what purposes, the Düsseldorf police school represented only a hesitant beginning (Spencer 1992).

A Note on the Historiography of Nineteenth-Century German Policing

Police history had a late start in Germany. Except for a few legal history studies, police history was not part of a mainstream in German historiography but rather the exception until the end of the 1970s (for the exceptions see Koselleck 1975; Maier 1986). In the context of new thematic issues arising during the 1970s and 1920s within German historiography, emerging studies on the history of nineteenth-century police focused on the functions and the role of the police in eminent political issues, which were at the core of politics in nineteenth-century German states, such as the relationship between the police and the military in Prussia (Lüdtke 1989; Johansen 2005) or the emergence of political policing during the first half of the nineteenth century (Siemann 1985).

The first German study, aiming at a comprehensive historical analysis of the police within the economic, political, and social structures of this period, has been published in the mid-1980s by a political scientist (Funk 1986), relating organizational patterns of the police, its everyday practices included, and describing police development in Prussia as a correlate of the increasing economic, social, and political cleavages within Prussia. The author analyzes how the ruling old and the new Prussian elites sought to cope with the threats to the economic, the social and the social order, deriving from what was seen as an overall change, by trying to increase the quantitative and qualitative strength of the state police and the Gendarmerie. The working class and its organizations were among these threats, but urbanization created a major moral panic as well.

While these early studies of a renewed look at the police concentrated very much at the political systems level or did put a focus on Berlin (e.g., Funk 1986), more recent studies, focusing on the local level, in particular on urban policing in other Prussian and German cities, have described the complexities and the contradictions of nineteenth-century policing in Germany and the attempts to modernize and professionalize urban policing (Reinke 1991, 1992, 1993, 2000, 2000a; Spencer 1992; Jessen 1991; Roth 1997). This modernization and professionalization has been seen by the research as strategy for adjusting the police to the growing control requirements, originating from industrialization and urbanization patterns. But urban policing during this period meant not only the enhancement of what could be seen as modern control strategies: a major characteristic of urban policing during this period was the persistence of large-scale welfare functions as part of police functions, thus placing police practices at the turn of the nineteenth to the twentieth century to some extent into the continuity of Ancien Régime *Polizey* models. These patterns resulted in an overpolicing, being abolished only gradually. This continuity created a legacy, which lasted until after the Second World War, when "police" and "welfare" became separated during the occupation of Germany by the victorious Allied forces (Reinke and Fürmetz 2000b).

Related Entries

- ▶ Conceptualizing of Police
- ▶ History of Police Unions
- ▶ History of the Police Profession
- ▶ High Policing
- ▶ National and Local Policing
- ▶ Order Maintenance Policing
- ▶ Police in the Police State
- ▶ Police School Services
- ▶ Role and Function of the Police

Recommended Reading and References

Evans JE (ed) (1989) Kneipengespräche im Kaiserreich. Stimmungsberichte der Hamburger Politischen Polizei 1892–1914. Rowohlt, Hamburg

Fosdick RB (1972) European police systems. Patterson Smith, Montclair (first published 1915)

Fricke D (1962) Bismarcks Prätorianer. Die Berliner Politische Polizei im Kampf gegen die deutsche Arbeiterbewegung (1871–1990). Rütten & Loening, Berlin

Funk A (1986) Polizei und Rechtsstaat. Die Entwicklung des staatlichen Gewaltmonopols in Preußen 1848–1914. Campus, Frankfurt am Main

Jäger J (2006) Verfolgung durch Verwaltung. Internationales Verbrechen und internationale Polizeikooperation 1880–1933. UVG Verlagsgesellschaft, Konstanz

Jessen R (1991) Polizei im Industrierevier. Modernisierung und Herrschaftspraxis im westfälischen Ruhrgebiet 1848–1914. Vandenhoeck & Ruprecht, Göttingen

Johansen A (2005) Soldiers as police. The French and the Prussian armies and the policing of popular protest, 1889–1914. Ashgate, Aldershot

Knöbl W (1998) Polizei und Herrschaft im Modernisierungsprozeß. Staatsbildung und innere Sicherheit in Preußen, England und Amerika 1700–1914. Campus, Frankfurt am Main

Koselleck R (1975) Preußen zwischen Reform und Revolution. Allgemeines Landrecht, Verwaltung und soziale Bewegung von 1791 bis 1848, 2nd edn. Klett, Stuttgart

Lindenberger T (1985) Straßenpolitik. Zur Sozialgeschichte der öffentlichen Ordnung in Berlin 1900 bis 1914. Verlag J.H.W. Dietz Nachfolger, Bonn

Lüdtke A (1989) Police and State in Prussia, 1815–1850. Cambridge University Press, Cambridge

Lüdtke A (ed) (1992) "Sicherheit" und "Wohlfahrt". Polizei, Gesellschaft und Herrschaft im 19. und 20. Jahrhundert. Suhrkamp, Frankfurt am Main

Maier H (1986) Die ältere deutsche Staats- und Verwaltungslehre, dtv, 2nd edn. München

Reinke H (1991) "Armed as if for a war": the state, the military and the professionalization of the Prussian police in Imperial Germany. In: Emsley C, Weinberger B (eds) Policing Western Europe. Politics, professionalization and public order, 1850–1940. Greenwood Press, New York, pp 55–73

Reinke H (1992) "hat sich ein politischer und wirtschaftlicher Polizeistaat entwickelt". Polizei und Großstadt im Rheinland vom Vorabend des Ersten Weltkrieges bis zum Beginn der zwanziger Jahre. In: Lüdtke A (ed) "Sicherheit" und "Wohlfahrt". Polizei, Gesellschaft und Herrschaft im 19. und 20. Jahrhundert. Suhrkamp, Frankfurt am Main, pp 219–242

Reinke H (ed) (1993) "… nur für die Sicherheit da …"? Zur Geschichte der Polizei im 19. und 20. Jahrhundert. Campus, Frankfurt am Main

Reinke H (2000a) "Großstadtpolizei". Städtische Ordnung und Sicherheit und die Polizei in der Zeit des deutschen Kaiserreiches (1871–1918). In: Dinges M, Sack F (eds) Unsichere Großstädte? Vom Mittelalter bis zur Postmoderne. UVK Universitätsverlag, Konstanz, pp 217–239

Reinke H, Fürmetz G (2000b) Polizei-Politik in Deutschland unter alliierter Besatzung. In: Lang H-J (ed) Staat, Demokratie und Innere Sicherheit in Deutschland. Leske und Budrich, Opladen, pp 87–86

Roth A (1997) Kriminalitätsbekämpfung in deutschen Großstädten 1850–1914. Ein Beitrag zur Geschichte des strafrechtlichen Ermittlungsverfahrens. Erich Schmidt Verlag, Berlin

Saul K (1974) Staat, Industrie, Arbeiterbewegung im Kaiserreich. Zur Innen- und Außenpolitik des Wilhelminischen Deutschland 1903–1914. Bertelsmann Universitätsverlag, Düsseldorf

Siemann W (1985) "Deutschlands Ruhe, Sicherheit und Ordnung". Die Anfänge der politischen Polizei 1806–1866. Max Niemeyer, Tübingen

Spencer EG (1992) Police and the social order in German cities. The Düsseldorf District, 1858–1914. Northern Illinois University Press, De Kalb

Wilms R (1992) Politische Polizei und Sozialdemokratie im Kaiserreich. Zur Tätigkeit der Politischen Polizei in der Provinz Hannover von der Zeit der Reichsgründung bis zum Ende des Sozialistengesetzes. Pater Lang, Frankfurt am Main

Wirsing B (1991) Die Geschichte der Gendarmeriekorps und deren Vorläuferorganisationen in Baden, Württemberg und Bayern 1750–1850. Unpublished dissertation, Konstanz University

GLM

► Good Lives Model

Good Lives Model

Tony Ward[1], Svenja Göbbels[2] and
Gwenda M. Willis[3]
[1]School of Psychology, Victoria University
of Wellington, Wellington, New Zealand
[2]Maastricht University, Maastricht,
The Netherlands
[3]University of Auckland, School of Psychology,
Auckland, New Zealand

Synonyms

GLM

Overview

The Good Lives Model (GLM) is a strengths-based theory of offender rehabilitation. The purpose of this essay is to (i) briefly overview the desistance literature; (ii) describe the risk management approach; (iii) highlight its limitations, including its weak fit with desistance theory and research; and (iv) provide a detailed description of the GLM. The GLM incorporates the advantages of the risk management approach at the same time as addressing its limitations. It can be argued that it provides social workers and other professionals with a more comprehensive framework to guide their work with clients in the criminal justice system.

Introduction

The rehabilitation of offenders is a complex process and involves reentry, and ultimately reintegration, into social networks and the broader society. While offenders need to work hard at modifying their personal characteristics that are related to their offenses, the community also has a responsibility to support this personal work with social capital and resources. Once individuals have begun to serve their sentence, they are entitled to have a chance at redemption and reconciliation. The presumption of human beings' equal value is a cornerstone of a decent and just society. It applies to offenders as much as to all humans. Furthermore, the rehabilitation of offenders is a normative and strength-building process, and therefore, from a practice perspective, both science and ethics are equally important. It is only legitimate to inflict significant harm (i.e., compulsory treatment) upon others when all human beings are regarded as equal in dignity and moral standing (Laws and Ward 2011).

Practitioners need rehabilitation theories to help them navigate through the various challenges and problems that materialize when working with offenders. These rehabilitation theories function as a conceptual map and facilitate effective intervention. Ideally, these maps will provide guidance on important issues such as the overall aims of intervention, what constitutes risk, what the general causes of crime are, how best to manage and work with individuals, and how to balance offender needs with the interests of the community. In recent years, strengths-based or "restorative" approaches to working with offenders have been formulated as an alternative to the very popular Risk-Need-Responsivity model of offender rehabilitation. In short, risk management approaches primary practice focus is on the detection and modification of dynamic risk factors (i.e., criminogenic needs), while strengths-based perspectives aim to create competencies in offenders and reduce risk more indirectly.

Offenders *are* "people like us" (Laws and Ward 2011). It is necessary to start relating to them in ways that reflect this attitude to improve correctional outcomes and reduce reoffending rates. The desistance research is clear that offenders respond well to practitioners who show an interest in them and believe in their capacity to turn their lives around (McNeill et al. 2005). Treating offenders with respect and decency rather than as sources of contamination to be quarantined (not cured) is likely to reduce the risk of practicing an ineffective confrontational therapeutic style (RNR; Andrews et al. 1990). Most information in this text is drawn on the sexual offending literature; although equally applicable to general offending, the GLM was initially developed in reference to the sexual offending literature.

Desistance from Crime

In contrast to the forensic psychology literature's focus on individual factors implicated in offending and reoffending, the desistance literature seeks to understand the lifestyle change process associated with cessation of crime (e.g., Laws and Ward 2011; Serin and Lloyd 2009). To suggest that a reduction in dynamic risk factors solely explains desistance seems to be unconvincing. Such an explanation is arguably somewhat simplistic and does not account for the very nature of human beings. Rather than being passively determined by external circumstances, humans actively seek outcomes that are personally meaningful and valued. The desistance literature disentangles *how* offenders change in regard to dynamic risk factors. Therefore it provides a richness not captured by the forensic psychology literature (Laws and Ward 2011). Available evidence indicates that there are a number of social and psychological factors that assist the desistance process (Laws and Ward 2011). These factors were named, for example, as "turning points" (Laub and Sampson 2003), "hooks for change" (Giordano et al. 2007), a "change in narrative identity" (McNeill et al. 2005), or "making good" (Maruna 2001).

Perhaps the most significant contributions to the desistance literature in recent years are those of Laub and Sampson (2003) and Maruna (2001). Laub and Sampson conducted an extended and comprehensive follow-up of men from Sheldon and Eleanor Glueck's landmark research (Glueck and Glueck 1950). They were interested in factors that differentiated serious and persistent delinquent boys from a matched group of nondelinquent boys. Laub and Sampson found that conventional adult social bonds such as marriage and employment explained variations in crime. Other variables, like childhood adversity, did not predict these. Specifically, they found that strong social bonds, for example, strong marital attachment and job stability, could facilitate the lifestyle change required for criminal desistance. Their findings have been replicated throughout the desistance literature (e.g., Maruna 2001; see below), and similar findings have been reported in the forensic psychology literature (e.g., Hanson and Harris 2000). Laub and Sampson also replicated the long-standing finding in criminology that frequency of offending decreases with age. They furthermore acknowledged the role of human agency, noting that men who desisted played an active role in their desistance process through making explicit choices to disengage from crime. Maruna (2001) replicated Laub and Sampson's findings regarding the significance of social bonds but also found that human agency or cognitive transformation (i.e., creation of a new, conventional, more adaptive narrative identity) was a key element in desistance. In sum, both external factors (e.g., social support, access to employment opportunities) and internal factors (e.g., making a conscious decision to want a different life) are required to facilitate the lifestyle change process associated with desistance. Desistance is the central aim of offender rehabilitation. In the following paragraphs, the risk management approach and its failure to account for the desistance literature will be contrasted with a recent strengths-based rehabilitation model, the Good Lives Model of Offender Rehabilitation.

The Risk Management Approach to Offender Rehabilitation

The risk management approach to offender rehabilitation emerged from Andrews and Bonta's influential book, *The Psychology of Criminal Conduct* (PCC; Andrews and Bonta 2010). The PCC sought to explain criminal behavior through empirically derived predictors of recidivism using what Andrews and Bonta termed a *general personality and social psychology perspective*. The PCC provides three empirically based principles aimed at reducing offenders' risk of recidivism: *risk*, *need*, and *responsivity* (Andrews and Bonta 2010). Hence, an underlying assumption of the risk management approach is that offenders are primarily containers of risk for recidivism, and the sole aim of offender rehabilitation is to reduce this recidivism risk through adherence to the RNR principles. The risk principle states that the dosage or intensity of interventions should match an offender's risk level. Therefore, low-risk offenders should receive less intense or no intervention, whereas high-risk offenders should be subjected to very intensive treatment. The needs principle incorporates that interventions should target criminogenic needs, also known as *dynamic risk factors*. Those factors are causally related to offending and are changeable. Dynamic risk factors include antisocial attitudes and antisocial associates (Andrews and Bonta 2010) and in the case of sexual offending, deviant sexual interests and self-regulation difficulties (Laws and Ward 2011). The aim of treatment is to reduce dynamic risk factors and, according to the need principle, directing intervention efforts at non-criminogenic needs will prove ineffective. For example, non-criminogenic needs such as low self-esteem and a history of victimization should not be targeted in treatment, given they have not been linked with recidivism (Andrews and Bonta 2010). Finally, the responsivity principle informs the actual delivery of interventions in order to maximize their efficacy. The responsivity principle involves matching the style and mode of intervention to the offender's learning style and abilities (Andrews and Bonta 2010). General responsivity advocates structured cognitive behavior therapy (CBT) interventions, given their general acceptance as the best treatment currently available for sex offenders (e.g., Hanson et al. 2002). Enhancing specific responsivity requires considering cognitive ability, learning style, personality profile, culture, and other characteristics of individual offenders and delivering treatment accordingly. The RNR has been hugely influential in offender rehabilitation initiatives internationally, forming the basis of correctional treatment since its inception in the early 1990s.

Although meta-analyses have found support for the efficacy of RNR-based treatment programs in reducing recidivism among general and sexual offenders (e.g., Hanson et al. 2009, 2002; Lösel and Schmucker 2005), some researchers argue that the available evidence is insufficient to conclude current treatment programs are in fact efficacious (e.g., Marques et al. 2005). Research suggests that a considerable amount of treated sexual offenders recidivates (e.g., Hanson et al. 2002). Thus, substantial scope remains for improving sex-offender rehabilitation and reintegration initiatives. In this entry it is argued that the Good Lives Model (GLM) offers exciting promise for further enhancing the effectiveness of current efforts by addressing limitations of the risk management approach, which are expanded on in the following entry.

Limitations of the Risk Management Model

The most heavily cited criticism of the RNR model revolves around its failure to motivate and engage offenders in the rehabilitation process (e.g., Mann 2000). Attrition from sex-offender treatment programs is particularly high with reported rates as high as 30–50 % (e.g., Ware and Bright 2008), which have been attributed to poor treatment engagement (e.g., Beyko and Wong 2005). Consistent evidence shows that men who drop out of treatment are more likely to reoffend compared to treatment completers (e.g., Hanson et al. 2002; Marques et al. 2005) as well as untreated comparison groups (Hanson et al. 2002). Without addressing the problem of treatment attrition, current treatment programs fail to deliver to groups of sex offenders most requiring treatment (Beyko and Wong 2005) and therefore fail to adhere to the RNR risk principle. Thus, although empirically derived, in reality the risk principle is difficult to adhere to.

What is behind the failure of the risk management approach to engage clients in rehabilitation? At the outset, the risk management approach differs substantially from therapeutic models used with other client populations (e.g., in the treatment of mental health problems) in the orientation of treatment goals, limited collaboration between client and therapist, and limited attention to problems not causally related to the problem behavior (i.e., in the case of offending – non-criminogenic needs such as self-esteem or personal distress). Addressing the first issue, risk management interventions rely profoundly on *avoidant* goals. These treatments try to encourage hypervigilance to threats of relapse and the reduction of dynamic risk factors (Mann 2000). By contrast, *approach* goals provide an individual with direction toward his or her goal. It has been suggested that individuals focused on approach goals concentrate on positive outcomes and thus persevere longer than people motivated by avoidance goals, who tend to focus on threats (Ward et al. 2007). Reframing the overarching goal of treatment (i.e., reducing risk of reoffending) as an approach goal might be “to become someone who lives a satisfying life that is always respectful of others” (Mann 2000, p. 194). This approach goal remains consistent with avoiding relapse. The final goal of avoidance of reoffending can be separated into personally meaningful subgoals that provide offender clients with direction in life, for example, increasing confidence in socializing with adult partners. Thus, by using approach goals, treatment can help offenders live a better, more satisfying life, not just a less harmful one, in ways that are personally meaningful and socially acceptable – and risk reducing (Mann 2000). In fact, it is likely that the combination of approach and avoidance goals is implicated in successful desistance. A balance between something that is hoped for – a better life – and what is feared for, i.e., recidivism, is thought to be more effective in reaching one’s ultimate goal (Paternoster and Bushway 2009). The resulting motivation can be seen as additive, because it incorporates avoidance and approach goals. Indeed, Mann et al. (2004) showed that an approach-goal-focused intervention with sex offenders was associated with increased treatment engagement compared to a traditional avoidant-goal-focused intervention.

Secondly, treatment goals in the risk management approach are enforced upon offenders rather than mutually agreed upon in therapy (Mann 2000). This can compromise the therapeutic relationship. Marshall and his colleagues (e.g., Marshall et al. 2003) demonstrated that confrontational therapeutic styles had a negative impact on attitude and behavior changes. Displays of empathy, warmth, encouragement, and some degree of directiveness facilitated treatment change. This suggests that careful attention to the therapeutic relationship might increase treatment engagement. The didactic, strictly formalized nature of the risk management approach, however, allows limited scope for enhancing the therapeutic relationship.

Third, some researchers have convincingly argued that a sole focus on criminogenic needs obstructs treatment engagement. Sometimes attention to non-criminogenic needs is necessary to establish enhanced well-being and quality of life. This in turn is likely to enhance treatment engagement (Laws and Ward 2011). More specifically, targeting non-criminogenic needs might be a necessary precursor for targeting criminogenic needs through enhancing the therapeutic alliance (Ward and Stewart 2003). For example, attempting to tackle criminogenic needs in the context of personal distress or financial crisis (both non-criminogenic needs) will likely prove fruitless if the more acute issues are not sufficiently addressed.

Another general limitation of the risk management approach is its minimal consideration of reentry and reintegration issues (beside identifying and then actively avoiding high-risk situations). The desistance literature emphasizes the decisive role of environmental systems such as close, supportive relationships and employment in ceasing offending (e.g., Laub and Sampson 2003). Thus, building and strengthening environmental opportunities, resources, and supports should be as central to offender rehabilitation and reintegration endeavors as it is to psychological treatment. Moreover, in the case of treated offenders, environmental factors can facilitate or impede the continuation of treatment-related change to dynamic risk factors. Treatment effects can only be generalized to parole if the environment supports and reinforces newly learned concepts, such as the restructuring of offense-supportive beliefs. If an offender, for instance, is embedded in a criminal subculture, which endorses offense-supportive beliefs, it is unlikely that the effects of cognitive restructuring will last long.

The failure of the risk management approach to engage offender clients in the rehabilitation process is derived from its theoretical underpinnings, or mainly a lack thereof (for a detailed discussion, see Laws and Ward 2011), which ignore the nature of human beings as value-laden, goal-directed beings. The risk management approach seems to be too mechanistic and reductionist – that is, there is an implicit assumption that through fixing a malfunction offenders are (hopefully) restored to their optimal functioning state. Humans, on the other hand, are arguably not simply clusters of mechanisms but also persons with an array of values. Therefore, it is not simply enough to correct personal deficits, or reduce criminogenic needs, and expect individuals who have committed crimes to be rehabilitated. It is important to acknowledge that these crimes were often committed in a misguided pursuit of subjectively valued outcomes. In other words, the theoretical grounding in managing risk, rather than improving the lives of offenders, compromises client engagement and their capacity for change.

In summary, critics argue that the RNR approach commonly current in offender rehabilitation and reintegration endeavors constitutes a necessary, but not sufficient foundation for effective interventions (Ward and Stewart 2003). It is essential subjecting offenders to interventions that are empirically supported; however, there is still much to be done in the arena of correctional practice and that desistance theory and research can offer those working with offenders numerous good ideas and practices. It has been convincingly argued that offender rehabilitation endeavors require a dual focus: reducing risk but also promoting human needs and values through approach goals, thereby engaging offenders in the treatment process. The GLM was developed as an alternative, more comprehensive approach to rehabilitation which is able to accomplish a dual focus. In other words the very nature of the GLM addresses limitations of the risk management approach, by including motivating offenders to engage in treatment, by addressing desistance issues, and by considering offenders' environmental contexts (Ward et al. 2007; Ward and Stewart 2003). Although developed independently on a theoretical level, the GLM is a natural ally of desistance theory on a practical level. This is because of the overlapping nature of both perspectives' theoretical assumptions and their common stress on the importance of both offender agency and social resources (for a more detailed discussion of this point see Göbbels et al. (2012)).

The Good Lives Model

The Good Lives Model (GLM), first proposed by Ward and Stewart (2003) and further developed by Ward and colleagues (e.g., Ward and Gannon 2006), is a strengths-based approach to offender rehabilitation. It is a strengths-based rehabilitation theory, because it is responsive to offenders' particular interests, abilities, and aspirations. It also directs practitioners to explicitly construct intervention plans that help offenders to acquire the capabilities to achieve the things that are personally meaningful to them, however in a socially appropriate manner. It assumes that all individuals have similar aspirations and needs and that one of the primary responsibilities of parents, teachers, and the broader community is to help each individual to acquire the tools required to make our own way in the world. Ideally, this is achieved through socialization in childhood and adolescence. Criminal behavior results when individuals lack the internal and external resources necessary to satisfy their values using pro-social means. In other words, criminal behavior represents a maladaptive attempt to meet life values (Ward and Stewart 2003). Rehabilitation endeavors should therefore equip offenders with the knowledge, skills, opportunities, and resources necessary to satisfy their life values in ways that do not harm others. Inherent in its focus on an offender's life values, the GLM places a strong weight on offender agency. That is, offenders, like all human beings, actively seek to satisfy their life values through whatever means available to them. The GLM's dual attention to an offender's internal values and life priorities and external factors such as resources and opportunities give it practical utility in desistance-oriented interventions. In addition, the GLM as a theory has the conceptual resources to incorporate desistance ideas because it also stresses agency, interdependency, and development. In other words, there is natural resonance between desistance theory and the GLM because of their overlapping theoretical ideas and broad way of conceptualizing the relationship between human beings and their social world.

The GLM is a theory of offender rehabilitation that contains three hierarchical sets of assumptions: general assumptions concerning the aims of rehabilitation, etiological assumptions that account for the onset and maintenance of offending, and practical implications arising from the first and second sets of assumptions. Each set of assumptions will be detailed, followed by a summary of empirical research investigating the utility of the GLM.

General Assumptions of the GLM

The GLM is grounded in the ethical concept of human dignity (see Ward and Syversen 2009) and universal human rights. The GLM perceives human beings as agents, rather than passive recipients who are determined by external circumstances only. That is, the GLM is concerned with individuals' ability to formulate and select goals, construct plans, and to act freely in the implementation of these plans. A closely related assumption is the basic idea that offenders, like all humans, aspire to certain states of mind, personal characteristics, and experiences. These are defined in the GLM as *primary goods*. Following an extensive review of psychological, social, biological, and anthropological research, Ward and colleagues first proposed ten classes of primary goods. In more recent work (e.g., Ward and Gannon 2006; Ward et al. 2007), they separated the goods of friendship and community to produce 11 classes of primary goods: (1) life (including healthy living and functioning),

(2) knowledge, (3) excellence in play, (4) excellence in work (including mastery experiences), (5) excellence in agency (i.e., autonomy and self-directedness), (6) inner peace (i.e., freedom from emotional turmoil and stress), (7) friendship (including intimate, romantic, and family relationships), (8) community, (9) spirituality (in the broad sense of finding meaning and purpose in life), (10) happiness, and (11) creativity (Ward and Gannon 2006, p. 79). While it is assumed that all humans seek out all the primary goods to some degree, the weightings or priorities given to specific primary goods reflect an offender's values and life priorities. Moreover, the existence of a number of practical identities, based on, for example, family roles (e.g., parent), work (e.g., nurse), and leisure (e.g., rugby player), means that an individual might draw on different value sources in different contexts, depending on the normative values behind each practical identity.

Instrumental goods, or s*econdary goods*, provide concrete means of securing primary goods and take the form of approach goals. For example, completing an apprenticeship or university degree might satisfy the primary goods of knowledge and excellence in work, whereas joining an adult sports team or cultural club might satisfy the primary good of friendship. If offenders engage in certain personally valued activities, it is likely that dynamic risk factors are reduced in an indirect manner. Thus, the GLM targets approach goals directly and avoidance goals indirectly.

Etiological Assumptions of the GLM

According to the GLM there are two primary routes that lead to the onset of offending: direct and indirect (Ward and Gannon 2006). The direct pathway is implicated when an offender actively attempts (often implicitly) to satisfy primary goods through his or her offending behavior. For example, an individual lacking skill to relate to potential partners may try to acquire the good of intimacy by committing date rape. The indirect pathway is implicated when primary human goods are frustrated over and over again. This can lead to a ripple or cascading effect that culminates in a criminal offense. For example, conflict between the goods of intimacy and excellence in work might lead to the breakup of a relationship and subsequent feelings of loneliness and distress. Maladaptive coping strategies such as the use of alcohol to alleviate distress might, in specific circumstances, lead to a loss of control and culminate in sexual offending (Ward et al. 2007).

Four types of difficulties in offenders' attempts to secure primary goods have been proposed. First, and most common in the direct route to offending, is the use of *inappropriate strategies* (secondary goods) to achieve primary goods. For instance, a preferential child sexual offender might not be able to achieve the good of intimacy with an adult partner and consequently tries to secure the good by molesting a child. Second, an individual's implicit good lives plan might suffer from a *lack of scope*, meaning that a number of goods are neglected in his or her life plan. For example, an offender might neglect the good of excellence in work and may feel incompetent because of his lack of achievement. These feelings might accumulate to a high degree of life dissatisfaction and can cumulate to sex offending. Third, *conflict* in the pursuit of goods might result in acute psychological stress and unhappiness. Fourth, an individual might lack *internal and external capabilities* to satisfy primary goods in the environment he or she lives in. Internal capabilities include relevant knowledge and skill sets, while external capabilities include environmental opportunities, resources, and supports.

Empirically identified criminogenic needs are conceptualized in the GLM as internal or external obstacles that interfere with the acquisition of primary goods. Indeed, as outlined by Laws and Ward (2011), each of the primary goods can be linked with one or more criminogenic needs. Taking the primary good of agency as an example, impulsivity might obstruct good fulfillment. Similarly poor emotional regulation might block attainment of inner peace.

Practical Implications of the GLM

To reiterate, the aim of correctional intervention according to the GLM is the promotion of primary goods, or human needs that, once met, enhance psychological well-being and functioning. In applying the GLM, assessment begins with mapping out an offender's good lives conceptualization by identifying the priorities given to the various primary goods. This is achieved through (i) asking increasingly detailed questions about an offender's core commitments in life and his or her valued day-to-day activities and experiences and (ii) identifying the goals and underlying values that were evident in an offender's offense-related actions. Once an offender's good life conceptualization is understood, future-oriented secondary goods aimed at satisfying an offender's primary goods in socially acceptable ways are formulated collaboratively with the offender and translated into a good lives treatment plan. Treatment does not have a one-size-fits-all approach but is individually tailored to assist an offender implement his or her good lives intervention plan and simultaneously address criminogenic needs that might be blocking goods fulfillment. Accordingly, intervention might include building internal capacity and skills and maximizing external resources and social supports to satisfy primary human goods in socially acceptable ways.

Ward et al. (2007) outlined a group-based application of the GLM based on seven modules typical of current best-practice sex-offender treatment programs: establishing therapy norms, understanding offending and cognitive restructuring, dealing with deviant arousal, victim impact and empathy training, affect regulation, social skills training, and relapse prevention (RP). They highlighted that most modules were associated with an overarching primary good, consistent with the notion that dynamic risk factors can be considered maladaptive means of securing primary goods. For example, an overarching good in the understanding offending and cognitive restructuring module is that of knowledge, attained through providing offenders with an understanding of how their thoughts, feelings, and actions led them to offend. The social skills training module is associated with the overarching goods of friendship, community, and agency. Offenders' individual good lives plans should inform the nature of interventions provided in this module. Some offenders, for example, may value other primary goods such as excellence in play and work over the good of friendship; thus, basic social skills training will likely suffice. Other offenders, however, may highly value intimate relationships; thus, intensive therapeutic work on intimacy and relationships might be required. This illustrates that the GLM promotes tailoring or treatment to individual offenders.

Willis et al. (in press) give specific recommendations on how to integrate the GLM successfully into cognitive-behavioral and risk management treatments. The basic assumption of humans as active agents implies that intervention planning should be collaborative. In addition, clients should be informed about their risk assessment results and should be explained to them. The GLM promotes individualization of treatment. In treatment, appropriate secondary goods should be acknowledged, reinforced, and incorporated together with future-oriented approach goals. For instance, an offender may be particularly ambitious and hardworking and thereby be very successful in his or her profession. In contrast, he or she might not be able to meet the good of relatedness with appropriate secondary goods. The offender might express the desire to reconnect with his or her family, make more friends, and find an intimate partner, and so on. Due to the holistic orientation of the GLM, also these non-criminogenic needs are addressed. The GLM has a dual focus. It capitalizes offender well-being *and* reduction of recidivism risk. Therefore, the offender does not introduce himself with only with offenses, sentencing information, and criminal history. The offender as a person is acknowledged by the therapist and other group members, rather than being reduced to his criminal offense. Offenders are treated as "people like us" (Laws and Ward 2011). As a consequence, offenders are also informed about the GLM and links between treatment modules and the acquisition of a good life. In contrast, hostility, negative labeling of the client, and the use of confrontation are all inconsistent with the GLM approach. In addition, the therapist should not be seen as a teacher or as superior to the clients. Clients are fellow human beings, who are active participants in the therapeutic process. Another crucial aspect is the GLM's emphasis of the client's environment. The offenders learn to attain valued goods with appropriate secondary goods to assure an offense-free life. This involves the work of a multidisciplinary team (correctional workers, nurses, healthcare workers, therapists, etc).

Empirical Research Supporting the Utility of the GLM

The most commonly cited criticism of the GLM is its lack of empirical support (Andrews and Bonta 2010). However, the GLM is not a *treatment* theory, but is rather a rehabilitation framework that is intended to supply practitioners with an overview of the aims and values underpinning practice. It functions as a broad *map* which needs to be supplemented by specific theories smaller in scope concerning concrete interventions such as cognitive-behavioral treatment techniques (Laws and Ward 2011). Thus, the criticism that the GLM (itself) has not been empirically supported misses the point. Rather, it is intended to provide a more comprehensive framework for offender practice than currently exists. However, programs can be – and are – constructed that reflect GLM assumptions and these can (and should) be evaluated. But in this case they are best construed as GLM consistent programs and are not the GLM itself (Laws and Ward 2011). To reiterate, the GLM provides an overarching rehabilitation framework, but does not prescribe specific intervention content (Willis et al. in press). Importantly, treatments that are developed within the framework of the GLM have to adhere to its basic and etiological assumptions. If there is a lack in fidelity with the GLM, the treatment might not work or might be, in the worst case, not beneficial. Due to its individual, flexible, and holistic focus, the GLM does not provide clinicians with rigidly structured treatment manuals. However, structure is needed to assure that central treatment targets are addressed and risk of reoffending will be reduced.

Keeping these general points in mind, recent programs have incorporated principles of the GLM with RP-based treatment, with positive results. For example, Ware and Bright (2008) recently reported preliminary results following the incorporation of GLM principles into their sex-offender treatment program, concurrently with the introduction of open treatment groups, meaning offenders work through treatment modules at their own pace (in contrast to closed treatment groups whereby group members start and finish together). Since the implementation of these changes, the treatment attrition rate has reduced, and staff have reported feeling more effective and positive in their work, likely benefiting their therapeutic relationship with clients. In another study, Lindsay et al. (2007) demonstrated the incorporation of GLM and RP principles with sex offenders using two case examples. They reported the dual focus on improving quality of life and managing risk-enhanced treatment engagement and provided offenders with a pro-social and personally meaningful life focus. Both offenders remained offense-free 5 years following their referrals for treatment. Consistent with reports of the GLM's effectiveness with sex offenders, the GLM has also been successfully applied with a high-risk violent offender (Whitehead et al. 2007) reported that the implementation of GLM principles facilitated treatment readiness and promoted long-term reintegration goals.

Other studies have empirically examined the underlying assumptions of the GLM. Willis and Grace (2008) retrospectively coded child molesters' release planning and found that the presence of secondary goods (i.e., socially acceptable approach goals relating to one or more primary goods) was a protective factor against any type of recidivism (i.e., sexual, violent, or general recidivism), again implicating the importance of goods fulfillment in the desistance process.

Barnett and Wood (2008) investigated how imprisoned sex offenders had operationalized the primary goods of agency, relatedness, and inner peace at the time of their offending. A lack of scope in good lives conceptualizations (e.g., through neglecting inner peace) and problems and/or conflict in the means used to pursue each good were evident. This supports the notion that difficulties fulfilling primary goods are implicated in offending. More recently, the GLM was applied to a sample of released child molesters and showed that the majority of primary goods were endorsed with high importance, supporting the premise that the GLM primary goods represent a set of universally sought after human values (Willis et al. 2010). In addition, it was found that positive reentry experiences at 1 and 3 months post-release (in terms of accommodation, social support, and employment) were associated with increased primary goods attainment 6 months post-release, suggesting that positive reentry experiences provided external capabilities for the implementation of good lives plans and eventual realization of life values.

In sum, the GLM has demonstrated preliminary effectiveness in addressing key limitations of the risk management approach to offender rehabilitation. This was done by enhancing treatment engagement, fostering desistance, and paying increased attention to environmental contexts. Moreover, a growing body of research supports the GLM's underlying assumptions.

Conclusion

Individuals with a history of criminal offending are more than bearers of risk, and as such, rehabilitation and reintegration endeavors require more than managing risk. The risk management approach has been hugely influential, and the primary RNR principles should not be rejected. Rather, the principles of risk, need, and responsivity should be integrated within a broader, strengths-based rehabilitation theory, the GLM. Through acknowledging that offenders are people like all human beings, the GLM engages offenders in the process of desistance, thereby bettering their lives and the lives of people they come into contact with. A problem with risk management practice models is that they tend to be overly focused on individual offenders and lack sufficient theoretical and ethical resources to enlarge their vision to the broader social and cultural vista. In other words, if the ultimate aim is to help individuals to cease offending and stay on the straight road, it is necessary to have a just, caring, and mutually accountable society.

Recommended Reading and References

Andrews DA, Bonta J (2010) The psychology of criminal conduct, 5th edn. Anderson, Cincinnati

Andrews DA, Bonta J, Hoge RD (1990) Classification for effective rehabilitation: rediscovering psychology. Crim Just Behav 17(1):19–52. doi: 10.1177/0093854890017001004

Barnett G, Wood JL (2008) Agency, relatedness, inner peace, and problem solving in sexual offending: how sexual offenders prioritize and operationalize their good lives conceptions. Sex Abus J Res Treat 20:444–465

Beyko MJ, Wong SCP (2005) Predictors of treatment attrition as indicators for program improvement not offender shortcomings: a study of sex offender treatment attrition. Sex Abus J Res Treat 17:375–389

Bouman YHA, Schene AH, de Ruiter C (2009) Subjective well-being and recidivism in forensic psychiatric outpatients. Int J Forensic Ment Health 8:225–234

Giordano PC, Schroeder RD, Cernkovich SA (2007) Emotions and crime over the life course: a neo-median perspective on criminal continuity and change. Am J Sociol 112:1603–1661

Glueck Glueck S, Glueck E (1950) Unraveling juvenile delinquency. The Commonwealth Fund, New York

Göbbels S, Ward T, Willis GM (2012) An integrative theory of desistance from sex offending. Aggress Violent Behav 17(5):453–462. doi: http://dx.doi.org/10.1016/j.avb.2012.06.003

Hanson RK, Harris AJR (2000) Where should we intervene? Dynamic predictors of sexual assault recidivism. Crim Justice Behav 27:6–35

Hanson RK, Morton-Bourgon KE (2005) The characteristics of persistent sexual offenders: a meta-analysis of recidivism studies. J Consult Clin Psychol 73:1154–1163

Hanson RK, Gordon A, Harris AJR, Marques JK, Murphey W, Quinsey VL et al (2002) First report of the collaborative outcome data project on the effectiveness of psychological treatment for sex offenders. Sex Abus J Res Treat 14:169–194

Hanson RK, Bourgon G, Helmus L, Hodgson S (2009) The principles of effective correctional treatment also apply to sexual offenders: a meta-analysis. Crim Justice Behav 36:865–891

Laub JH, Sampson RJ (2003) Shared beginnings, divergent lives: delinquent boys to age 70. Harvard University Press, Cambridge

Laws DR, Ward T (2011) Desistance from sex offending: alternatives to throwing away the keys. Guilford, New York

Lindsay WR, Ward T, Morgan T, Wilson I (2007) Self-regulation of sex offending, future pathways and the good lives model: applications and problems. J Sex Aggress 13:37–50

Lösel F, Schmucker M (2005) The effectiveness of treatment for sexual offenders: a comprehensive meta-analysis. J Exp Criminol 1:117–146

Mann RE (2000) Managing resistance and rebellion in relapse prevention intervention. In: Laws DR, Hudson SM, Ward T (eds) Remaking relapse prevention with sex offenders: a sourcebook. Sage, Thousand Oaks, pp 187–200

Mann RE, Webster SD, Schofield C, Marshall WL (2004) Approach versus avoidance goals in relapse prevention with sexual offenders. Sex Abus J Res Treat 16:65–75

Marques JK, Wiederanders M, Day DM, Nelson C, van Ommeren A (2005) Effects of a relapse prevention program on sexual recidivism: final results from California's sex offender treatment and evaluation project (SOTEP). Sex Abus J Res Treat 17:79–107

Marshall WL, Serran GA, Fernandez YM, Mulloy R, Mann RE, Thornton D (2003) Therapist characteristics in the treatment of sexual offenders: tentative data on their relationship with indices of behaviour change. J Sex Aggress 9:25–30

Maruna S (2001) Making good: how ex-convicts reform and rebuild their lives. American Psychological Association, Washington, DC

McNeill F, Batchelor S, Burnett R, Knox J (2005) 21st century social work. Reducing reoffending: key practice skills. Scottish Executive, Edinburgh

Paternoster R, Bushway S (2009) Desistance and the "feared self": toward an identity theory of criminal desistance. J Crim Law Crim 99(4):1103–1156. http://www.jstor.org/stable/20685067

Serin RC, Lloyd CD (2009) Examining the process of offender change: the transition to crime desistance. Psychol Crime Law 15:347–364

Ward T, Gannon TA (2006) Rehabilitation, etiology, and self-regulation: the comprehensive good lives model of treatment for sexual offenders. Aggress Violent Behav 11:77–94

Ward T, Stewart CA (2003) The treatment of sex offenders: risk management and good lives. Prof Psychol Res Pract 34:353–360

Ward T, Syversen K (2009) Human dignity and vulnerable agency: an ethical framework for forensic practice. Aggress Violent Behav 14:94–105

Ward T, Mann RE, Gannon TA (2007) The good lives model of offender rehabilitation: clinical implications. Aggress Violent Behav 12:87–107

Ware J, Bright DA (2008) Evolution of a treatment programme for sex offenders: changes to the NSW custody-based intensive treatment (CUBIT). Psychiatry Psychol Law 15:340–349

Whitehead PR, Ward T, Collie RM (2007) Time for a change: applying the good lives model of rehabilitation to a high-risk violent offender. Int J Offender Ther Comp Criminol 51:578–598

Willis GM, Grace RC (2008) The quality of community reintegration planning for child molesters: effects on sexual recidivism. Sex Abuse J Res Treat 20:218–240

Willis GM, Ward T (2010) Striving for a good life: the good lives model applied to released child molesters. Manuscript submitted for publication

Willis GM, Yates PM, Gannon TA, Ward T (in press) How to integrate the good lives model into treatment programs for sexual offending: an introduction and overview. Sex Abus J Res Treat. doi: 10.1177/1079063212452618

Governance

▶ Judicial Leadership and Performance

Graft

▶ Corruption

Green Criminology

Rob White
School of Social Sciences, University of Tasmania, Hobart, TAS, Australia

Overview

The term "green criminology" emerged in the 1990s to describe a critical and sustained approach to the study of environmental crime (Lynch 1990; South 1998). This chapter provides an outline of the distinctive features of green criminology, its main concepts and foci of analysis, and the continuing debates that mark its further and continuing development as a bona fide perspective within criminology.

Generally speaking, green criminology takes as its focus issues relating to the environment (in the widest sense possible) and harm (as defined in ecological as well as strictly legal terms). Much of this work has been directed at exposing different instances of substantive social and ecological injustice. It has also involved critique of the actions of nation-states and transnational capital for fostering particular types of harm and for failing to adequately address or regulate harmful activity. Given the pressing nature of many environmental issues, it is not surprising that many criminologists are now seeing environmental crime and environmental victimization as areas for concerted analytical and practical attention.

From a criminological perspective, any attempt to take up the challenge offered here will require rigorous and sophisticated analysis of the social dynamics that shape and allow certain types of activities harmful to the environment (including humans and animals) to take place over time. This sort of analysis, in turn, demands that environmental issues be framed within the context of a sociological and criminological imagination (White 2003). That is, study must appreciate the importance of situating environmental harm as intrinsically socially and historically located and created. Interpretation and analysis thus has to be mindful of how current trends reflect the structure of global/local societies, the overall direction in which such societies are heading, and the ways in which diverse groups of people are being affected by particular social, economic, and political processes.

The Foundations of Green Criminology

Green criminology refers to the study by criminologists of environmental harms (that may incorporate wider definitions of crime than that provided in strictly legal definitions), environmental laws (including enforcement, prosecution, and sentencing practices), and environmental regulation (systems of civil and criminal law that are designed to manage, protect, and preserve specified environments and species and to manage the negative consequences of particular industrial processes) (White 2008).

The key focus of green criminology is environmental crime. This is conceptualized in several different ways within the broad framework of green criminology. For some writers, environmental crime is defined narrowly within strict legal definitions – it is what the law says it is. For others, environmental harm is itself deemed to be a (social and ecological) crime, regardless of legal status – if harm is done to environments or animals, then it is argued that this ought to be considered a "crime" from the point of view of the critical green criminologist.

Specific types of harm as described in law include things such as illegal transport and dumping of toxic waste, the transportation of hazardous materials such as ozone depleting substances, the illegal traffic in real or purported radioactive or nuclear substances, the proliferation of "e"-waste generated by the disposal of tens of thousands of computers and other equipment, the safe disposal of old ships and airplanes, the illegal trade in flora and fauna, and illegal fishing and logging.

However, within green criminology there is also a more expansive definition of environmental crime or harm that includes (White 2011):

- Transgressions that are *harmful to humans, environments, and nonhuman animals*, regardless of legality per se
- Environmental-related harms that are facilitated by *the state*, as well as *corporations and other powerful actors*, insofar as these institutions have the capacity to shape official definitions of environmental crime in ways that allow or condone environmentally harmful practices

The definition of environmental crime is, therefore, contentious and ambiguous. Much depends upon who is defining the harm and what criteria is used in assessing the nature of the activities so described (e.g., legal versus ecological, criminal justice versus social justice) (Beirne and South 2007; White 2008).

For many green criminologists the biggest threats to environmental rights, ecological justice, and nonhuman animal well-being are system-level structures and pressures that commodify all aspects of social existence that are based upon the exploitation humans, nonhuman animals, and natural resources and that privilege the powerful over the interests of the vast majority. It is for this reason that assessment of environmental injustice requires critical scrutiny of how states themselves intervene with regard to specific environmental harm issues.

An eco-justice perspective refers to the broad orientation of green criminology which is largely directed at exposing different instances of substantive social and environmental injustice. From an eco-justice perspective, environmental harm is best seen in terms of justice, which in turn is based upon notions of human, ecological and animal rights, and broad egalitarian principles. A key issue is the weighing up of different kinds of harm and violation of rights that may involve stretching the boundaries of conventional criminology to include other kinds of harms than those already deemed to be illegal.

Most green criminology is informed by at least one of the three approaches that collectively make up an eco-justice perspective. These include *environmental justice* (where the main focus is on differences within the human population: social justice demands access to healthy and safe environments for all and for future generations), *ecological justice* (where the main focus is on "the environment", as such to conserve and protect ecological well-being, e.g., forests, is seen as intrinsically worthwhile), and *species justice* (where the main focus is on ensuring the well-being of both species as a whole, such as whales or polar bears, and individual animals, which should be shielded from abuse, degradation, and torture).

A major factor that influences the study of environmental harm, therefore, relates to the specific interests that count the most when conceptualizing the nature and seriousness of the harm. For example, when criminalization does occur, it often reflects human-centered (or anthropocentric) notions of what is best (e.g., protection of legal fisheries, legal timber coups) in ways that treat "nature" and "wildlife" simply and mainly as resources for human exploitation. The intrinsic value of specific ecological areas and particular species tends to be downplayed or ignored.

Nevertheless, recent years have seen greater legislative and judicial attention being given to the rights of the environment per se and to the rights of certain species of nonhuman animal to live free from human abuse, torture, and degradation. This reflects both the efforts of eco-rights activists (e.g., conservationists) and animal rights activists (e.g., animal liberation movements) in changing perceptions, and laws, in regard to the natural environment and nonhuman species. It also reflects the growing recognition that centuries of industrialization and global exploitation of resources are transforming the very basis of world ecology – for example, global warming threatens us all, regardless of where we live or our specific socioeconomic situation.

Studying Environmental Harm

Green criminology provides an umbrella under which to theorize and critique both *illegal* environmental harms (i.e., environmental harms currently defined as unlawful and therefore punishable) and *legal* environmental harms (i.e., environmental harms currently condoned as lawful but which are nevertheless socially and ecologically harmful). How harm is conceptualized is thus partly shaped by how the legal-illegal divide is construed within specific research and analysis.

There are a number of intersecting dimensions that need to be considered in any analysis of environmental harm (White 2008). These include consideration of who the victim is (human or nonhuman), where the harm is manifest (global through local levels), the main site in which the harm is apparent (built or natural environment), and the time frame within which harm can be analyzed (immediate and delayed consequences). Many of the main features pertaining to environmental harm are inherently international in scope and substance.

Indeed, the categorization of environmental harm is varied in that there are different ways in which environmental crimes have been conceptualized and sorted. For instance, Carrabine et al. (2004) discuss environmental crimes in terms of primary and secondary crimes. Green crimes are broadly defined simply as crimes against the environment. Primary crimes are those crimes that result directly from the destruction and degradation of the earth's resources, through human actions. Secondary or symbiotic green crime is that crime arising out of the flouting of rules that seek to regulate environmental disasters. The first set of crimes relates to the harm as being bad in itself; the second relate to breaches of law or regulation associated with environmental management and protection.

In recent years researchers have studied a wide range of environmental harms associated with both "green" issues and "brown" issues. In the former the work has been motivated by either a concern with species justice or an interest in conventional environmental crimes such as illegal fishing. For instance, work over the past decade has been carried out in respect to crimes such as lobster poaching (McMullan and Perrier 2002), animal abuse (Beirne 2009), the illicit trade in endangered species (Wellsmith 2010), and deforestation (Boekhout van Solinge 2010).

In regard to "brown" issues, the production and disposal of waste is a matter of significant concern to academic researchers interested in questions of environmental harm. Relevant examples of such research include the role of organized criminal syndicates in the dumping of waste (Ruggierro 1996), inequalities associated with the location of disadvantaged and minorities' communities near toxic waste sites (Pellow 2007), and the global trade in electronic waste (Gibbs et al. 2010a).

The range of substantive topic areas that green criminology is presently investigating is growing. So too, the complexities involved in studying environmental harm are likewise being acknowledged. For example, the detection and origins of some types of environmentally related harm may be unclear due to significant time lags in manifestation of the harm. Here it is important to acknowledge the notion of cumulative effects. For example, this could refer to the way in which dioxins accumulate in fish flesh over time. It could also refer to the cumulative impact of multiple sources of pollution as in cases where there are a high number of factories in one area (such as places along the US-Mexican border). Diseases linked to asbestos poisoning may surface many years after first exposure, and this, too, provides another example of long-term effects of environmental harm. Persistent use of pesticides in particular geographical areas may also have unforeseen consequences for local wildlife, including the development of new diseases among endemic animal species (as has been suggested has occurred in the case of facial tumor disease now rampant among the Tasmanian devil population in Australia).

As extensive work on specific incidents and patterns of victimization demonstrates, it is also the case that some people are more likely to be disadvantaged by environmental problems than others. For instance, studies have identified disparities involving many different types of environmental hazard that especially adversely affect people of color, ethnic minority groups, and indigenous people (Bullard 2005). There are thus patterns of differential victimization that are evident with respect to the siting of toxic waste dumps, extreme air pollution, chemical accidents, access to safe clean drinking water, and so on (Williams 1996). It is the poor and disadvantaged who suffer disproportionately from such environmental inequalities.

The kinds of harms and crimes studied within green criminology include illegal trade in endangered species (e.g., trade in exotic birds or killing of elephants for their ivory tusks), illegal harvesting of "natural resources" (such as illegal fishing and illegal logging), and illegal disposal of toxic substances (as well as pollution of air, land, and water). Wider definitions of environmental crime extend the scope of analysis to consider harms associated with legal activities such as clear-felling of old-growth forests and the negative ecological consequences of new technologies such as use of genetically modified organisms in agriculture (e.g., reduction of biodiversity through extensive planting of GMO corn) (see, e.g., Walters 2011). Recent work has considered the criminological aspects of climate change (see White 2012), from the point of view of human contributions to global warming (e.g., carbon emissions from coal-fired power plants) and the criminality associated with the aftermath of natural disasters (e.g., incidents of theft and rape in the wake of Hurricane Katrina in New Orleans).

Differences Within Green Criminology

There is no green criminology *theory* as such. Rather, as observed by South (1998), there is what can loosely be described as a green "perspective." Elements of this perspective generally include things such as a concern with specifically environmental issues, social justice, ecological consciousness, the destructive nature of global capitalism, the role of the nation-state (and regional and global regulatory bodies), and inequality and discrimination as these relate to class, gender, race, and nonhuman animals. The green criminology perspective, therefore, tends to begin with a strong sensitivity toward crimes of the powerful and to be infused with issues pertaining to power, justice, inequality, and democracy.

Green criminology has emerged in the last 20 years as a distinctive area of research, scholarship, and intervention. It is distinctive in the sense that it has directed much greater attention to environmental crime and harm than mainstream criminology and has heightened awareness of emergent issues such as the problems arising from disposal of electronic waste (e-waste) and the social and ecological injustices linked to the corporate colonization of nature (including biopiracy and imposition of GMO crops in developing countries). Within the spectrum of ideas and activities associated with green criminology, there are nonetheless several different kinds of analytical framework.

While the link between and among green criminologists is the focus on environmental issues, important theoretical and political differences have become more apparent over time. For example, some argue that green criminology must necessarily be anticapitalist and exhibit a broad radical orientation (Lynch and Stretesky 2003). Others, however, construe the task as one of conservation and natural resource management, within the definitional limits of existing laws (Herbig and Joubert 2006; Gibbs et al. 2010b). Still others promote the idea that the direction of research should be global and ecological and that new concepts need to be developed that will better capture the nature and dynamics of environmental harms in the twenty-first century (White 2011).

Typically there are important differences within green criminology around issues pertaining to the distinction between "harm" and "crime." These differences do not stem solely from disputes over the legal/illegal divide however. There are also profound disagreements with regard to victimization and varying conceptions of justice. For instance, there may be differences *within* a particular area of work, such as debates over "animal rights" versus "animal welfare" in the case of concerns about species justice (Francione 2010). There are also disagreements in terms of priorities, values, and decision-making *between* particular areas of green criminology (Beirne 2011). This is evident, for example, in debates over multiple land-use areas. This kind of dispute can involve those who argue that human interests should come first (from the perspective of environmental justice), or that specific ecological niches be protected (from the perspective of ecological justice), even if some animals have to be killed or removed from a specific geographical location. From the point of view of species justice, however, big questions can be asked regarding the intrinsic rights of animals and the duty of humans to provide care and protection for nonhuman species.

The hallmark of green criminology, regardless of diversity of opinion and the plurality of views, is that proponents argue that criminology ought to take seriously environmental crimes, and in doing so to rethink how it does what it does, and how it might conceptualize the relevant issues. It is interesting in this respect that a number of prominent criminologists are now utilizing their expertise from mainstream areas of criminology (e.g., situational crime prevention, general strain theory) to study specifically environmental issues such as illegal trade in elephant tusks and social problems arising from climate change (Agnew 2012; Mesko et al. 2010; Lemieux and Clarke 2009). Green criminology is not only expanding in its own right but simultaneously there is a greening of criminology more generally.

Differences within green criminology are not only apparent at the level of theoretical focus and orientation. They are also manifest when it comes to responding to environmental crime or harm. From a critical green criminology view, for example, environmental harm is related to exploitation of both environments and humans by those who control the means of production. Analysis of global capitalism provides answers to questions such as why it is that human societies simultaneously respect and protect certain creatures (especially animal companions such as dogs and cats) while allowing and even condoning the dreadful treatment of others (as in the case of factory farming of battery hens to produce eggs) (Beirne 2009). It also allows us to better understand why it is that we strive to preserve some environments (via creation of national parks) while at the same time ensuring the devastation of particular ecosystems (such as clear-felling of old-growth forests).

Environmental harm takes place within the overarching context of a distinct global political economy. Most writers within the green criminology perspective concentrate on exposing specific types of criminal or harmful environmental actions or omissions. In doing so they have provided detailed descriptions and analyses of phenomena such as the illegal trade of animals, illegal logging, dumping of toxic waste, air pollution, and threats to biodiversity. In many cases, the corpus of work identified within this field has highlighted issues pertaining to social inequality, speciesism, ecological and environmental injustice, and crimes of the powerful. What is less common, however, are examples of study that locate these harms, crimes, injustices, and corrupt practices within the context of an explicit theoretical understanding of the state or economic relations. In other words, it is rare to find a sustained political economy of environmental harm.

Yet, analysis of broad trends indicates that it is systemic imperatives and historical transformations associated with global capitalism that, in today's world, ultimately shape what it is that individuals do with their lives and their environments. Even a cursory examination of dominant world political economic trends reveals the close link between capitalism as a system and environmental degradation and transformation. The sphere of production is dominated by the production of commodities, the advance of technology and biotechnologies, and the exploitation of labor (particularly in so called Third World countries) in the service of mass production of goods and services that, in turn, demand a high turnover rate. Extensive and intensive forms of consumption are essential to the realization of surplus value – that is, profit depends upon a critical mass of buyers purchasing the mass-produced commodities. The link between production and consumption is found in the form of specific kinds of distribution processes (e.g., transportation of goods and services, retail outlets, storage, roads, railways, bridges, and ships) and exchange mechanisms (e.g., finance capital, credit availability) that sustain and contribute to extensive use of natural and human resources. Economic efficiency is measured in how quickly and cheaply commodities can be produced, channeled to markets, and consumed. It is a process that is inherently exploitive of both humans and nature.

In essence, the competition and the waste associated with the capitalist mode of production have a huge impact on the wider environment, on humans and on nonhuman animals (e.g., in the form of pollution and toxicity levels in air, water, and land). These same processes pose major threats to biodiversity and the shrinking of the number of plant and animal species generally. This is related both to the legal and illegal trade in species, as well as to mass industrial production and extensive use of genetically modified organism (GMO) technologies.

Differences in opinion over the nature of global political economy, and over the tactics and strategies most likely to bring about desired social and ecological transformations, manifest in different approaches to how responses to environmental harm are construed (White 2008). One approach is to chart up existing environmental legislation and provide a sustained socio-legal analysis of specific breaches of law, the role of environmental law enforcement agencies, and the difficulties and opportunities of using criminal law against environmental offenders. Another approach places emphasis on social regulation as the key mechanism to prevent and curtail environmental harm, including attempts to reform existing systems of production and consumption through a constellation of measures and by bringing nongovernment and community groups directly into the regulatory process. A third approach presses the need for transnational activism, with an emphasis on fundamental social change. What counts is engagement in strategies that will challenge dominant authority structures and those modes of production that are linked to environmental degradation and destruction, negative transformations of nature, species decline, and threats to biodiversity. Social movements are seen to be vital in dealing with instances of gross environmental harm.

By its very nature, the development of green criminology as a field of sustained research and scholarship, will incorporate many different approaches and strategic emphases. For some, the point of academic concern and practical application will be to reform aspects of the present system. Critical analysis, in this context, will consist of thinking of ways to improve existing methods of environmental regulation and perhaps to seek better ways to define and legally entrench the notion of environmental crime. For others, the issues raised above are inextricably linked to the project of social transformation. From this perspective, analysis ought to focus on the strategic location and activities of transnational capital, as supported by hegemonic nation-states on a world scale, and it ought to deal with systemic hierarchical inequalities. Such analysis opens the door to identifying the strategic sites for resistance, contestation, and struggle on the part of those fighting for social justice, ecological justice, and animal rights.

There are major political divisions within the broad spectrum of green criminological work (and indeed within green political movements), and these have major implications for whether action will be taken in collaboration with capitalist institutions and state authorities, or whether it will be directed toward radically challenging these institutions and authorities. Similarly, there are significant tensions between ecological justice and species justice approaches, as indicated in the following observation:

> The [green environmentalists] rarely champion the sites of their concerns with rights talk, whereas for [animal rights advocates] their very focus is the criterion for moral standing and holding of rights. This crucial deep-seated difference is already present in green criminology in environmentalist notions such as ‘fisheries’ and ‘harvests’ and ‘conservation’, all of which are the stuff and fodder of animal welfare and sustainability but mostly anathema to animal rights. (Beirne 2011: 354)

To put it differently, some green criminologists view nature instrumentally, and harm is viewed through the lens of legality; others view the exploitation of nature, particularly in relation to animals, as intrinsically bad and harmful. How or if this “moral fissure” can be overcome is of major interest to many currently working within the broad area of green criminology.

Conclusion

Green criminology has many different substantive contributions and theoretical dimensions. Debates will continue over how best to define concepts such as harm, crime, and victim; over the moral calculus that weighs up human, ecosystem, and animals interests and rights; and over which interventions will achieve what kinds of intended and unintended outcomes. Dialogue around these issues will ensure lively and healthy deliberations over environmental matters now and into the future.

The development of green criminology has led to new interests, new conceptualizations, and new techniques of analysis. This is because there is increasing acknowledgement of environmental problems and the relevance of this to traditional criminological concerns with social injury and social regulation. There is also greater awareness of the interconnectedness of social and environmental issues. For example, matters relating to poverty, health, indigenous people's rights, exploitation of nonhuman nature, corporate business wrongdoing, state corruption, and so on are seen in many instances to be inseparable. As well, there is recognition of the need for multidisciplinary approaches to the study of environmental harm, involving cooperation between different "experts," including those with traditional and experiential knowledge associated with culture and livelihood (such as indigenous peoples and farmers and fishers), as well as sensitivity to ideas and research generated in intellectual domains such as law, police studies, political science, international relations, zoology, biology, philosophy, sociology, and chemistry.

These kinds of observations and interrelationships are forcing a rethink of the social and natural universe and a reconceptualization of the relationship between humans and nature in ways that accord greater weight to the nonhuman when it comes to assessing issues such as environmental harm. In practical terms, this translates into new and overlapping domains of consideration within green criminology itself: hence, the concern with transgressions against humans, environments, and animals.

Related Entries

- ▶ Crimes Against Animal Life
- ▶ Crimes of Globalization
- ▶ Ecoterrorism
- ▶ Environmental and Human Rights
- ▶ Environmental Regulation and Law Enforcement
- ▶ History of Green Criminology
- ▶ Organized Crime and the Environment
- ▶ State-Corporate Crime

Recommended Reading and References

Agnew R (2012) It's the end of the world as we know it: the advance of climate change from a criminological perspective. In: White R (ed) Climate change from a criminological perspective. Springer, New York

Beirne P (2009) Confronting animal abuse: law, criminology, and human-animal relationships. Rowman & Littlefield, New York

Beirne P (2011) Animal abuse and criminology: introduction to special issue. Crime Law Soc Change 55:349–357

Beirne P, South N (eds) (2007) Issues in green criminology: confronting harms against environments, humanity and other animals. Willan Publishing, Devon

Boekhout van Solinge T (2010) Deforestation crimes and conflicts in the Amazon. Crit Criminol 18:263–277

Bullard R (ed) (2005) The quest for environmental justice: human rights and the politics of pollution. Sierra Club Books, San Francisco

Carrabine E, Iganski P, Lee M, Plummer K, South N (2004) Criminology: a sociological introduction. Routledge, London

Francione G (2010) Abolitionist approach to animal rights. In: Bekoff M (ed) Encyclopedia of animal rights and animal welfare, Volume 1. Greenwood Press, Santa Barbara

Gibbs C, McGarrell E, Axelrod M (2010a) Transnational white-collar crime and risk: lessons from the global trade in electronic waste. Criminol Publ Pol 9(3):543–560

Gibbs C, Gore M, McGarrell E, Rivers L III (2010b) Introducing conservation criminology: towards interdisciplinary scholarship on environmental crimes and risks. Brit J Criminol 50:124–144

Halsey M (2004) Against "Green" criminology. Brit J Criminol 44(4):833–853

Halsey M, White R (1998) Crime, ecophilosophy and environmental harm. Theor Criminol 2(3):345–371

Herbig F, Joubert S (2006) Criminological semantics: conservation criminology – vision or vagary? Acta Criminologica 19(3):88–103

Lane P (1998) Ecofeminism meets criminology. Theor Criminol 2(2):235–248

Lemieux A, Clarke R (2009) The international ban on ivory sales and its effects on elephant poaching in Africa. Brit J Criminol 49(4):451–471

Lynch M (1990) 'The greening of criminology: a perspective on the 1990s'. The Critical Criminologist 2(3): 1–4 and 11–12

Lynch M, Stretesky P (2003) The meaning of green: contrasting criminological perspectives. Theor Criminol 7(2):217–238

McMullan J, Perrier D (2002) Lobster poaching and the ironies of law enforcement. Law Soc Rev 36(4):679–720

Mesko G, Dimitrijevic D, Fields C (eds) (2010) Understanding and managing threats to the environment in South Eastern Europe. Springer, Dordrecht

Pellow D (2007) Resisting global toxics: transnational movements for environmental justice. MIT Press, Cambridge

Ruggierro V (1996) Organized and corporate crime in Europe: offers that can't be refused. Dartmouth, Aldershot

South N (1998) A green field for criminology? A proposal for a perspective. Theor Criminol 2(2):211–233

Walters R (2011) Eco crime and genetically modified food. Routledge, New York

Wellsmith M (2010) The applicability of crime prevention to problems of environmental harm: a consideration of illicit trade in endangered species. In: White R (ed) Global environmental harm: criminological perspectives. Willan Publishing, Devon

White R (2003) Environmental issues and the criminological imagination. Theor Criminol 7(4):483–506

White R (2008) Crimes against nature: environmental criminology and ecological justice. Willan Publishing, Devon

White R (2011) Transnational environmental crime: toward an eco-global criminology. Routledge, London

White R (ed) (2012) Climate change from a criminological perspective. Springer, New York

Williams C (1996) An environmental victimology. Soc Justice 23(4):16–40

Group Characteristics and General Strain Theory

Stacy De Coster
Department of Sociology and Anthropology,
North Carolina State University, Raleigh,
NC, USA

Overview

Although general strain theory was developed to explain differences in offending across individuals (Agnew 1985, 2006), recent scholarly efforts suggest that the theory offers significant insight into group differences in offending. These efforts suggest that group differences in exposure to strain, emotional reactions to strain, and access to resources for dealing with strain and negative emotions contribute to different rates of offending across various groups. In applying this argument to sex differences in offending, the theoretical emphasis has been on how patriarchal structures inform gender socialization and gendered roles, which ultimately shape stress exposure, emotional and behavioral responses to stress, and the efficacy of legitimate coping resources for dealing with strain (Broidy and Agnew 1997). Explications of race differences in offending emphasize the role of additional structural arrangements and race discrimination in shaping the relationship between strain and offending (Kaufman et al. 2008). Cross-national research guided by general strain theory has been somewhat sparse, but explanations for research findings on samples of non-United States residents have emphasized that cultural values shape the strain process in ways that may contribute to understanding differences in offending cross-nationally. A particular strength of general strain theory for understanding group differences in offending is that it allows for consideration of how structural arrangements and cultural values coalesce to create group variations in crime and delinquency. An important direction for future research is to consider the ways in which the structural arrangements emphasized in the race literature and cross-national studies coalesce with the cultural beliefs about gender at the heart of the literature on sex differences to provide insight into variations in offending across groups and individuals situated at various junctures in gender-race-class hierarchies and race-gender-nationality locations.

Introduction

General strain theory is among the most prominent individual-level explanations of offending (Agnew 1985, 2006). The theory posits that strain leads to negative emotions – including anger, frustration, and depression – that promote offending when resources for legitimate coping are limited. Although the theory was developed to explain differences in offending across individuals, recent scholarly efforts indicate that the theory also offers significant insight into group differences in offending. Generally, theoretical and empirical efforts suggest that group differences in offending can be attributed in part to structural and cultural forces that shape group differences in exposure to strain, emotional reactions to strain, and access to resources for dealing with strain and negative emotions. This broad argument has been applied most commonly to understanding differences in offending across sex, race, and neighborhood groups but likely applies equally well to societies and to groups situated at various junctures at the intersections of race-gender-class and race-gender-nationality.

Sex Differences

A gendered extension of general strain theory offers that the theory can explain why males are more likely to engage in illegal behaviors than are females, positing that the strains most frequently experienced by females and males, emotional and behavioral reactions to strain, and the availability of legitimate coping resources are shaped by gendered roles and gender socialization (Broidy and Agnew 1997). The first hypothesis in this gendered perspective centers on sex differences in the types of strains most often experienced by females and males.

One of the crime-producing strains emphasized by general strain theory is failure to achieve positively valued goals. Drawing on gender research, Broidy and Agnew (1997) note that gender socialization emphasizes different goals for females and males. Females in patriarchal society traditionally are socialized to be concerned with maintaining relationships, finding meaning in life, and how people are treated in interactions; by contrast, males are taught to focus on economic success, personal achievement, and outcomes of interactions. As a result, females and males are likely to experience different types of goal-related strains. Relational strains – such as conflict in close relationships, network events, and suicide attempts of loved ones – align with feminine concerns, whereas, agentic strains – including economic failure, academic failure, mistreatment by others, and criminal victimization – are linked with masculine concerns. Consistent with these gendered concerns, adults often occupy gendered roles that require more emotion work and caretaking from females and emphasize the primacy of economic responsibilities for males (e.g., Kessler and McLeod 1984). Like gender socialization, the gendering of social roles in patriarchal society informs the types of strains – relational and agentic – to which males and females are most exposed and those they are most prone to experience as stressful (e.g., Kessler and McLeod 1984).

Broidy and Agnew (1997) propose that the differentiation of relational and agentic strains is pertinent for understanding sex differences in offending because agentic strains are more likely than relational strains to promote crime. For instance, criminal victimization and mistreatment may be particularly conducive to violent crime; failure to achieve economic goals can provoke property crime. Relational strains are less likely to encourage illegal behavior because such behavior can threaten social relationships or harm others.

Although some empirical evidence documents that females report more exposure to relational strains and males report more agentic strains (e.g., Kessler and McLeod 1984; De Coster 2005), there is little evidence that agentic strains are more criminogenic than relational strains (see Agnew 2006). This lack of evidence can be attributed to the fact that empirical assessments of general strain theory most often use composite scales of strain that do not separate relational and agentic strains. Studies that differential between types of strain typically include measures of either relational *or* agentic strains or include measures of both but do not examine which is more criminogenic.

The second hypothesis offered in Broidy and Agnew's (1997) gendered perspective proposes that emotional reactions to strain are shaped by gender socialization. Agnew (1985) has given primacy to the negative emotion of anger, positing that anger is the emotion most likely to lead to offending because it energizes individuals for action and leads to a desire for retaliation. Given that males and females are equally likely to respond to strain with anger, Broidy and Agnew (1997) emphasize that understanding sex differences in offending requires consideration of qualitative distinctions in the experience of anger among males and females. Specifically, they propose that female anger is more likely than male anger to be accompanied by depression, anxiety, and sadness because females are socialized to view anger as inappropriate and also may worry that their anger may jeopardize valued social relationships.

This distinction in how males and females experience anger is relevant for understanding why males are more likely to offend than females, according to Broidy and Agnew (1997), because the depression that accompanies female anger may mitigate anger's impact on law violation. Although research demonstrates that female anger is more likely than male anger to be accompanied by depression and sadness, limited evidence suggests that depression actually exacerbates the impact of anger on offending (De Coster and Zito 2010). Since the feminine experience of anger – concurrently with depression – is more conducive to offending than the masculine experience, the key to understanding links between gender, emotions, and illegality in general strain theory may reside in gendered expressions of emotional responses to strain rather than in gendered experiences of emotions.

An emphasis on expressions of emotional responses to strain is consistent with a large body of work on emotional displays and with Broidy and Agnew's (1997) proposition that illegitimate responses to strain and negative emotions may be shaped by gender socialization, gendered roles, and cultural beliefs about gender and emotional displays in patriarchal society. Consistent with much prior theorizing, they note that femininity is the antithesis of crime and violence but that masculinity is consistent with and may even promote illegality. In addition, feminine roles are more likely to limit access to crime as a coping strategy than are masculine roles. As such, males with limited access to legitimate coping resources are likely to respond to strains and negative emotions with illegal behaviors; similarly situated females may respond with eating disorders, suicidal ideation, distress, or other mental health problems. Indeed, empirical evidence demonstrates gender distinctions in the expression of problems (e.g., Horwitz and White 1987; Kaufman 2009).

The final hypothesis in Broidy and Agnew's (1997) gendered perspective offers that males are more predisposed to illegality and have fewer legitimate coping resources. That is, males are more likely than females to engage in crime without thinking (predisposition) and are more likely to lack the social skills necessary for legitimate coping and the maintenance of socially supportive relationships. These differences are linked once again to the fact that gender socialization emphasizes the development of relational concerns for females and agentic concerns for males. The expectation is that legitimate coping resources, social support, and noncriminal predispositions – buffering factors – protect females to a greater degree than males from offending in the face of strain and negative emotions. Although research often supports the claim that females possess more coping and social support resources than males (Thoits 1995), these resources do not buffer the impact of either strain or negative emotions on offending more for females than for males (e.g., Morash and Moon 2007; Jennings et al. 2009). However, criminal predispositions – captured with measures of aggressiveness and low self control – exacerbate the impact of strain on illegal behaviors more for males than for females (e.g., Liu and Kaplan 2004; Cheung and Cheung 2010).

Overall, Broidy and Agnew (1997) propose that general strain theory can effectively explain why males engage in more law violation than do females. Their overarching framework emphasizes that males and females occupy gendered roles in society and are socialized to have concerns and goals that are consistent with traditional expectations for their gender. This shapes their experiences of strain, emotional reactions to strains, and their coping resources, social supports, and criminal predispositions. Empirical studies support the propositions that the types of strains to which males and females are exposed, emotional responses to strain, and expressions of negative emotions are gendered. However, current evidence does not support the claim that the gendering of strains and emotional responses to strain are relevant for understanding sex differences in illegal behaviors. Instead, sex differences appear to emerge because males are more likely than females to *express* negative emotions illegally and because the effect of strain on offending is more likely to be exacerbated by criminal predispositions for males than for females.

Perhaps the greatest limitation of the empirical literature to date is the failure to differentiate sex from gender. That is, tests of the propositions specified by Broidy and Agnew (1997) typically have assumed that gender socialization is perfect, ignoring variability within groups of females and males with respect to how much traditional gender socialization and gendered expectations shape their goals, emotions, predispositions, and reactions to strain and negative emotions. Research that includes direct measures of the extent to which individuals embrace traditional definitions of gender would provide a more accurate assessment of Broidy and Agnew's (1997) hypotheses than what has been offered to date.

Race and Neighborhood Differences

Theory and research on race and offending in criminology most often emphasizes the broad structural forces that shape offending. Consistent with this, discussions of race and offending from a general strain perspective emphasize the role of discrimination and neighborhood disadvantage in the etiology of race patterns of offending.

Strain theorists highlight several mechanisms through which race discrimination can produce elevated rates of offending among Blacks in particular. Perhaps the most obvious argument is that discrimination is a form of strain – failure to achieve the valued goal of justice or presentation of a negatively valued stimulus – that produces negative emotions and criminal coping (see Kaufman et al. 2008). Indeed, Agnew (2001) defines race discrimination to be among the strains most likely to produce crime because it is likely to be seen as unjust and is high in centrality/magnitude because it threatens core values and identities. Consistent with this, research demonstrates a link between race discrimination and offending among Blacks that is mediated at least partially by anger and depression (Simons et al. 2003).

Race discrimination can also influence race patterns of offending by situating Blacks in neighborhoods where strains proliferate, they are exposed to angered/frustrated individuals, and access to legitimate coping resources is limited (Kaufman et al. 2008). Much research demonstrates that the racial segregation of disadvantaged urban neighborhoods arises in part from economic constraints (often produced by discriminatory practices in the labor market and discriminatory practices in housing markets). Given this, Agnew's (1999) general strain theory of neighborhood crime rates provides important insights into elevated rates of crime among Blacks, who often are concentrated in the most disadvantaged neighborhoods.

One mechanism through which disadvantaged neighborhoods create elevated crime rates, according to general strain theory, is through the generation of strain (Agnew 1999). Consistent with this, research shows that residents of disadvantaged areas are more likely than those in more advantaged areas to experience a wide array of strains, including harassment and threats, criminal victimization, and witnessing violence (e.g., Warner and Fowler 2003; Kaufman 2005). Importantly, studies demonstrate that the disproportionately high offending rates among Blacks can be attributed partially to their greater exposure to major events, such as criminal victimization and witnessing serious violence (Eitle and Turner 2003).

The concentration in disadvantaged areas of individuals who experience a wide array of strains translates into a high concentration of angered/frustrated individuals in these areas, which increases the chances of interacting with angry/frustrated people in what can be thought of as a "charged environment" (Bernard 1990; Agnew 1999). This, of course, is an additional source of crime-provoking strain for residents in disadvantaged neighborhoods. When confronted with this and other forms of strain generated by disadvantage and race, individuals are likely to blame their situation and their angry feelings on external factors, which increases the chances of illegitimate coping (Bernard 1990; Agnew 1999).

The likelihood of illegitimate coping is also increased by the fact that disadvantaged neighborhoods provide individuals and groups with few legitimate coping resources. For instance, retaliatory crime is a means through which groups can deal with victimization of self or loved ones, particularly when models of effective coping and access to police, court, and psychiatric resources are largely unavailable. Retaliatory crime, in fact, may become a necessity in these areas to help avoid the strains of future victimization and identity threats (Mullins et al. 2004).

In sum, general strain theory offers that elevated offending rates among Blacks may be explained both directly and indirectly by discrimination experiences. Race discrimination has been shown to be a source of strain that leads to negative emotions and offending (Simons et al. 2003), but it impacts crime also through the role it plays in concentrating minorities in disadvantaged urban neighborhoods. These neighborhoods are characterized by high rates of offending in part because they expose residents to strains – such as victimization of self and loved ones, mistreatment/harassment, and hostile interactions with angry people – that are particularly criminogenic. Limited access to legitimate coping resources in disadvantaged areas only exacerbates the criminogenic nature of these strains.

Although research supports the argument that strains emanating from disadvantaged neighborhoods help mediate the impacts of race and neighborhood disadvantage on offending (e.g., Kaufman 2005), much less emphasis has been placed on assessing whether legitimate coping resources interact with strain in the production of race and neighborhood crime patterns. Perhaps this is because some studies find that race differences in levels of exposure to strain are relevant for understanding racial patterns of offending, but differences in the likelihood of responding to strain illegally are negligible in explicating these patterns (Eitle and Turner 2003). This might imply that differential access to coping resources is not of primary importance for understanding race and neighborhood patterns of offending. Given that research on neighborhoods, race, and strains is relatively sparse, however, the role of coping resources – particularly at the neighborhood level – requires further exploration.

National/Societal Differences

Although cross-national research from a general strain perspective has been relatively sparse (see Agnew 2006), there is sufficient evidence to suggest that the theory has meaningful insights to offer for understanding variability in rates of offending across societies or nations. For instance, cross-national comparative studies show that nations with high rates of economic inequality – a structural source of strain – have higher homicide rates than nations characterized by less inequality (e.g., Messner 1989; Pratt and Godsey 2003). Importantly, the impact of economic inequality on homicide rates is tempered by national-level social support resources (see Pratt and Godsey 2003). This suggests that cultural values – the value of being supportive of citizens – may significantly shape the extent to which strain impacts offending across nations.

This suggestion bears out most prominently in studies of general strain theory on non-United States samples, which often invoke cultural arguments to explain why certain strains may be more or less criminogenic in other nations than in the United States. Research findings support the applicability of general strain theory in samples of Chinese, Philippine, South Korean, Israeli, and Ukrainian samples (Landau 1997; Maxwell 2001; Bao et al. 2004; Morash and Moon 2007; Botchkovar et al. 2009; Cheung and Cheung 2010). However, the conclusions of these studies generally suggest that cultural values shape the types of strains that may be most relevant for understanding offending within various nations. For instance, mistreatment by teachers appears to be a particularly important source of crime-producing strain among South Korean adolescents. Morash and Moon (2007) propose that the importance of this strain may be rooted in both South Korean cultural norms that emphasize emotional and physical punishment as achievement motivators and Confucian values that emphasize academic achievement as a prerequisite for success in social and economic realms.

Studies that fail to replicate links between specific strains and offending that have been established in American samples most often conclude that cultural differentiation provides the key to understanding why some strains are relevant in some nations and not others. For example, the strain of coercive parenting leads to delinquency among American adolescents but not among Chinese adolescents (compare Hay 2003 and Cheung and Cheung 2010). An explanation offered for this difference is that coercive parenting is consistent with the collectivist ideology of China, which encourages the subordination of the individual to family and community. In American society, where individualism is lauded, coercive parenting is likely to be experienced as a stressful infringement on individuality, thereby resulting in delinquent responses (Cheung and Cheung 2010). Youths from the Philippines respond to aggression between caretakers with delinquency but are not vulnerable to delinquency when faced with physical aggression by their caretakers. Maxwell (2001) explains this by offering that the general cultural acceptance of physical punishment in the Philippines may buffer the relationship between caretaker aggression and delinquency often reported in samples of American youths (e.g., Smith and Thornberry 1995). Finally, Botchkovar and colleagues (2009) conclude their study of general strain theory in Ukrainian, Greek, and Russian samples by noting that their general strain scale may have proven ineffective in the prediction of offending decisions among their Greek and Russian samples because it failed to consider that cultural values may shape the types of strains that are most relevant for understanding criminal intentions (Botchkovar et al. 2009).

Generally, these studies suggest that general strain theory has much to offer for understanding rates of offending across nations. To date, the majority of studies have provided ad hoc cultural explanations for why various strains are particularly relevant or irrelevant for explicating offending in different nations. Research on the applicability of general strain theory across societies would be strengthened by a theoretical framework that takes into consideration structural and cultural variations between nations and derives hypotheses about how these variations might shape the types of strains, emotions, and coping resources most relevant for understanding variations in offending across societies.

Conclusions

Despite the fact that general strain theory was developed to explain individual offending, recent scholarship demonstrates the theory's relevance for illuminating sex, race, neighborhood, and societal rates of offending. Explanations for sex differences in offending have focused primarily on gender socialization and gendered roles rooted in patriarchal structural arrangements; race differences have been discussed as emanating from neighborhood structural arrangements that are shaped by race discrimination and economic factors; and societal-level arguments have focused on economic structures and cultural values. A strength of general strain theory for understanding group differences in offending is in its ability to consider how structural arrangements and cultural beliefs coalesce to predict offending rates across groups. Given the emphasis on intersectionalities and crime in the broader literature on offending, an important next step for general strain theorists may be to further unite some of the cultural and structural arguments that have been offered in the literature to help explicate patterns of offending across race-sex-class or race-sex-nation groups.

One avenue for pursuing this goal would be to consider that some of the structural factors discussed as relevant for understanding race and neighborhood patterns of offending may shape the gendered roles and socialization that Broidy and Agnew (1997) propose shape strains, emotions, coping resources, and offending across sex groups. Intersectionality theorists propose, for instance, that race discrimination and economic disadvantage have made it such that the feminine roles and ideals lauded in patriarchal society have never applied to Black girls and women. As such, Black families socialize their daughters in ways that differ from how White families socialize their daughters. This means that Black females are less likely than their White counterparts to embrace the cultural form of femininity and feminine roles highlighted in Broidy and Agnew's (1997) discussion of sex and crime. This insight may prove relevant for understanding how race discrimination, community disadvantage, and gender socialization operate simultaneously in a general strain theory framework to help explicate patterns of offending across gender-race-class. One can also envision consideration of how variability across nations in cultural beliefs about gender may shape the strain process differently for males and females in different nations, thereby providing insight into variability in the size of the sex-gap in crime across nations.

Related Entries

► General Strain Theory

Recommended Reading and References

Agnew R (1985) A revised strain theory of delinquency. Soc Forces 64(1):151–167

Agnew R (1999) A general strain theory of community differences in crime rates. J Res Crime Delinq 36:123–155

Agnew R (2001) Building on the foundation of general strain theory: specifying the types of strain most likely to lead to delinquency. J Res Crime Delinq 38:319–361

Agnew R (2006) Pressured into crime: an overview of general strain theory. Roxbury, Los Angeles

Bao WN, Haas A, Pi Y (2004) Life strain, negative emotions, and delinquency: an empirical test of general strain theory in the People's Republic of China. Int J Offender Ther Comp Criminol 48:281–297

Bernard TJ (1990) Angry aggression among the 'truly disadvantaged'. Criminology 28:73–96

Botchkovar EV, Tittle CR, Antonaccio O (2009) General strain theory: additional evidence using cross-cultural data. Criminology 47:131–176

Broidy LM, Agnew R (1997) Gender and crime: a general strain theory perspective. J Res Crime Delinq 34:275–306

Cheung NWT, Cheung YT (2010) Strain, self-control, and gender differences in delinquency among Chinese adolescents: extending general strain theory. Sociol Perspect 53:321–345

De Coster S (2005) Delinquency and depression: gendered responses to gendered stresses. Sociol Perspect 48:155–187

De Coster S, Zito RC (2010) Gender and general strain theory: the gendering of emotional experiences and expressions. J Contemp Crim Justice 26:224–245

Eitle DJ, Turner RJ (2003) Stress exposure, race and young adult crime. Sociol Q 44:243–269

Hay C (2003) Family strain, gender and delinquency. Sociol Perspect 46:107–135

Horwitz AV, White HR (1987) Gender role orientations and styles of pathology among adolescents. J Health Soc Behav 28:158–170

Jennings WG, Piquero NL, Gover AR, Pérez DM (2009) Gender and general strain theory: a replication and exploration of Broidy and Agnew's gender/strain hypothesis among a sample of southwestern Mexican American adolescents. J Crim Justice 37:404–417

Kaufman JM (2005) Explaining the race/ethnicity-violence relationship: neighborhood context and social psychological processes. Justice Q 22:224–251

Kaufman JM (2009) Gendered responses to serious strain: the argument for a general strain theory of deviance. Justice Q 26:410–444

Kaufman JM, Rebellon C, Thaxton S, Agnew R (2008) A general strain theory of racial differences in criminal offending. Aust N Z J Criminol 41:421–437

Kessler RC, McLeod JD (1984) Sex differences in vulnerability to undesirable life events. Am Sociol Rev 49:620–631

Landau SF (1997) Crime patterns and their relation to subjective social stress and support indicators: the role of gender. J Quant Criminol 13:29–56

Liu X, Kaplan HB (2004) Role stress and aggression among young adults: the moderating influences of gender and adolescent aggression. Soc Psychol Q 67:88–102

Maxwell SR (2001) A focus on familial strain: antisocial behavior and delinquency in Filipino society. Sociol Inq 71:265–292

Messner SF (1989) Economic discrimination and societal homicide rates: further evidence of the cost of inequality. Am Sociol Rev 54:597–611

Morash M, Moon B (2007) Gender differences in the effects of strain on the delinquency of South Korean youth. Youth Soc 38:300–321

Mullins C, Wright R, Jacobs BA (2004) Gender, streetlife and criminal retaliation. Criminology 42:911–940

Pratt TC, Godsey TW (2003) Social support, inequality, and homicide: a cross-national test of an integrated theoretical model. Criminology 41:611–643

Simons RL, Chen Y, Stewart EA, Brody GH (2003) Incidents of discrimination and risk for delinquency: a longitudinal test of strain theory with an African American sample. Justice Q 20:827–854

Smith C, Thornberry TP (1995) The relationship between childhood maltreatment and adolescent involvement in delinquency. Criminology 33:451–477

Thoits PA (1995) Stress, coping and social support processes: where are we? What next? J Health Social Behav (Extra Issue):53–79

Warner BD, Fowler SK (2003) Strain and violence: testing a general strain theory model of community violence. J Crim Justice 31:511–521

Group Crime

► Co-offending and Offender Attributes

Group-Based Trajectory Modeling

► Group-Based Trajectory Models

Group-Based Trajectory Models

Gary Sweeten
Arizona State University, Phoenix, AZ, USA

Synonyms

Group-based trajectory modeling; Latent class growth modeling; Latent trajectory modeling; Semi-parametric group-based method

Overview

Group-based trajectory modeling is a powerful and versatile tool that has been extensively used to study crime over the life course. The method was part of a methodological response to the criminal careers debate but has since greatly expanded in its applications. The current state of the art of group-based trajectory modeling is complex, but worth becoming familiar with so as to recognize the variety of uses to which this tool can be applied. The method has attracted unusually robust criticism for a statistical tool. Researchers should aim to use it and other statistical tools as effectively as possible.

Introduction

Criminologists have long been interested in studying crime as a longitudinal phenomenon. Recent interest stems from vigorous debate surrounding the interpretation of the fact that criminal behavior bears a robust curvilinear relationship with age, quickly ramping up to a peak in the late teen years, and declining thereafter. Part of the criminal careers debate of the 1980s concerned the interpretation of this aggregate relationship. The peak around age 17 may reflect patterns of offending among all offenders, or it may reflect an influx of offenders with relatively short criminal careers around their teen years. In short, the question is whether individual criminal careers resemble the aggregate age-crime curve. This is one of the many questions that group-based trajectory modeling – a general statistical tool for uncovering distinct longitudinal patterns in panel data – can answer.

Group-based trajectory modeling can best be understood in contrast to its alternatives. The most common alternative to group-based trajectory modeling is hierarchical linear modeling, also known as general growth curve modeling. This alternative identifies the average trend over time and quantifies the degree of variation around this average. It is relatively parsimonious because it assumes a specific error distribution around the overall average, typically Gaussian. Group-based trajectory modeling, in contrast, makes no assumptions about the distribution of parameters. Rather, it assumes that the distribution can be approximated by a finite number of support points. Group-based trajectory modeling is related to hierarchical linear modeling in the same way a histogram is related to mean and standard deviation. The mean and standard deviation quickly provide information about the distribution of a variable, but a histogram can provide more detailed information on the shape of the distribution. Both are approximations of a more complex reality. Given a specified number of groups, group-based trajectory modeling attempts to discover a set of longitudinal paths that best represents the raw data. Group-based trajectory modeling is a powerful tool for uncovering distinct longitudinal paths, particularly when significant portions of the population follow a path that is very different from the overall average.

Group-based trajectory modeling is a versatile tool, not limited to revealing distinct longitudinal paths. Once the best group-based trajectory model has been chosen, the researcher may proceed by: (1) describing characteristics of the groups (i.e., characteristics of their growth patterns and covariates); (2) predicting group membership based on preexisting characteristics; (3) evaluating the group-specific effects of time-varying covariates and testing whether such effects vary by group; (4) using groups as independent variables to predict later outcomes; and (5) incorporating measures of group membership as control variables in regression models or propensity score matching protocols. Attesting to its utility and the ease of interpreting its results, although group-based trajectory modeling was developed to answer specific questions posed by the criminal careers model (Nagin and Land 1993), the method has been used to model trajectories of abdominal pain symptoms in pediatric patients (Mulvaney et al. 2006), fungal growth on the forest floor (Koide et al. 2007), and of course, extensive use in the study of longitudinal patterns of criminal activity (Piquero 2008).

This entry will proceed by first describing some of the theoretical background for the emergence of group-based trajectory modeling. The bulk of the entry will be devoted to describing the state of the art of the method including the basic components of the model, how to choose the number of groups, and several model extensions. Next, it will cover a number of controversies surrounding the method as well as some possibilities for further development.

Background

A central fact of criminology is that a small portion of the population commits a large share of crime. Goring (1913) found that just 3.4 % of the population accounted for about half of all convictions in England. This finding was repeated in a Philadelphia birth cohort: 5 % of the cohort accounted for over half of all police contacts (Wolfgang et al. 1972). This rediscovery of the concentration of offending occurred around the same time that the rehabilitative focus of punishment gave way to rationales of deterrence and incapacitation, creating demand for theoretical accounts and statistical tools that identified high-rate or “chronic” offenders as a distinct group

Alfred Blumstein and colleagues (1986) promoted a criminal careers framework that could be used to describe the volume of crime committed over an individual lifespan, including age of onset, rate of offending, age of termination (desistance), and career length. The criminal careers paradigm suggested that each of these parameters warranted investigation and, possibly, distinct theoretical explanations. It also provided a way to distinguish “career criminals” from others, through their long criminal career length. In opposition to this perspective, Travis Hirschi and Michael Gottfredson (1983) claimed that the criminal careers model did not provide any special insight into age and crime. They claimed the age-crime curve was invariant across social context and no sociological variable was able to account for it. They deemed longitudinal research a waste of resources since the correlates of crime are consistent across age. These two positions formed the basis for the criminal careers debate.

These developments set the stage for a number of theoretical and statistical advances in the 1990s. On the theoretical side, Robert Sampson and John Laub (1993) drew from Glen Elder’s life-course framework (1994) to develop an age-graded theory of informal social control. The life-course framework presents a way of thinking about individual development. Anything that can be measured once can be measured repeatedly, and stringing these measurements together over time creates a *trajectory*. Embedded within, and providing meaning to these trajectories are life *transitions*, which, if they redirect the trajectory onto another path, are called *turning points*. In contrast to Sampson and Laub's age-graded theory, Moffitt's (1993) typological theory of antisocial behavior posited that the age-crime curve is made up of two groups with distinct trajectories of antisocial behavior and distinct etiologies.

The criminal careers debate engendered several methodological advances as well. Rowe and colleagues (1990) used Rasch modeling to arbitrate between Hirschi and Gottfredson's parsimonious latent trait theory and the contention of the criminal careers model that each component of the criminal career may require distinct theoretical explanation. And in 1993, Nagin and Land introduced the group-based trajectory model to address several theoretical puzzles posed by the criminal careers model. This paper modeled longitudinal patterns of individual offending using a finite mixture model of time-varying rates of offending. This established the groundwork for a long line of statistical advances that have profoundly affected the field of criminology.

State of the Art

Overview

Any phenomenon or characteristic that evolves over age or time in an identifiable population may be an appropriate candidate for group-based trajectory modeling. This tool has the potential to draw out common patterns and help the researcher to make sense of the data. Group-based trajectory modeling produces a parsimonious summary of a complicated distribution of longitudinal developmental trajectories using a finite set of developmental trajectories and their estimated population prevalence. In contrast, hierarchical linear modeling and latent curve analysis summarize this complicated distribution even more parsimoniously, with a single developmental trajectory representing the average, and some characterization of the nature of variation around this average, obtained by imposing fairly strict distributional assumptions.

As opposed to these strict distributional assumptions, group-based trajectory modeling is agnostic on the nature of the distribution of growth parameters. It allows for the possibility for meaningful subgroups within the population that follow distinct developmental paths. The nature and definition of these "subgroups" has fueled a considerable amount of debate. It should be recognized that these groups may only be identified *after* the developmental paths that define them have already unfolded. That a distinct group is identified by the model does not speak to the issue of prospective identification of the group. Groups must be understood, like simple regression parameters, as useful simplifications and summaries of a complex reality.

Group-based trajectory models can be put to a variety of purposes. Once groups are obtained, description of each group's developmental paths and antecedents can be very useful. For example, if high-rate stable, medium-rate decreasing, and low-rate offending groups are identified by a particular model, the proportion of each group possessing certain risk variables can provide information of etiological significance. More generally, group membership may be treated as a dependent variable, something to be predicted as accurately as possible. Covariates may be added to the model to assess group-specific effects of certain conditions or life events on the developmental outcome of interest as well as whether these effects statistically differ between groups. Group membership can also be treated as an independent variable in a number of different ways. Variables representing group membership can be used in regression models to parsimoniously control for complicated developmental histories. Groups may also be used as the basis for matching or incorporated into a propensity score matching framework. These and other uses will be described below after first presenting data requirements, parameterization, and practical issues to address in choosing the "correct" model.

Practical Requirements

The group-based trajectory model requires panel data. There must be repeated measurements of a particular dependent variable within some unit of analysis. The unit of analysis depends on the purposes of the researcher and availability of data. To give a sense of the range of questions this method can be applied to, all the following units of analysis would be appropriate and several have been used in published research: person (delinquent, police officer, and academic), residence, street segment, journal article, sports team, gang, prison, terrorist organization, state, country.

There are currently three pieces of software able to estimate group-based trajectory models: MPLUS, the PROC TRAJ add-on to SAS, and the traj add-on for Stata. Estimation specifics in this entry will refer to PROC TRAJ, the most complete implementation of the method. Using this software, panel data must be structured in wide format: one line for each unit of analysis, with repeated measures of the dependent variable, time, and time-varying covariates represented using distinct variable names.

There are three broad distributional categories permissible for the dependent variable in PROC TRAJ: logit, censored normal, and Poisson. The censored normal distribution is flexible enough to incorporate cases of upper censoring, lower censoring, both, and neither. And the assumptions of the Poisson model can be relaxed using a zero-inflated Poisson parameterization. Categorical dependent variables and other distributions not listed here are not supported in PROC TRAJ.

Balanced panel data is not required. Any unit with one or more observations will be used in estimating the final model. Generally, at least three repeated measures within each unit are preferred to estimate a stable and meaningful model. The independent variable can be represented in units of age, time (years/months/ etc.), or other sequence. It is often desirable to create a meaningful zero value in the independent variable. For example, if estimating trajectories in young adulthood, age can be transformed so that zero represents age 18, and other ages are interpreted relative to age 18. This allows the estimated intercept of each group to quickly be interpreted as the predicted value at age 18. Centering the independent variable and scaling it down by dividing by 10 or 5 also helps the model to converge more readily.

It is worth thinking carefully about the scale of the independent variable, as transforming it can result in very different trajectory models. Consider, for example, the implications of a group-based trajectory model of marijuana use in adolescence with three types of age scales: untransformed age, age relative to the age of onset for marijuana use among those who ever reported using, and age relative to high school dropout among those who ever dropped out. The first scaling of age provides a model that characterizes longitudinal patterns of marijuana use in the population of interest. The second represents desistance/persistence patterns after initial use, and the third reveals marijuana use patterns relative to a potential turning point event.

Finally, it should be noted that the unit of time need not proceed in lock-step for each observation. For example, the model will estimate ideal group-based trajectories from age 14 to 22 even if one-third of the sample is observed only from ages 14 to 18, another third from 16 to 20, and the final third from 18 to 22. The final groups would have to be interpreted with great caution however, since they represent 8-year trajectories when no single individual is observed for more than 4 years.

Estimation

Equations used in this entry draw heavily from Nagin's (2005) detailed presentation in *Group-Based Modeling of Development*. Three subscripts will be used throughout: i refers to the unit of analysis, t to time or age, and j to group. Corresponding to these, N refers to the number of units in the model, T to the number of time periods, and J to the total number of groups in a particular model. The general form of the data used for group-based trajectory models is a set of longitudinal observations for each unit i:

$$Y_i = \{y_{i1}, y_{i2}, \ldots, y_{iT}\}$$

The set of parameters to be estimated for groups 1 through J represents a growth polynomial with respect to the unit of time. The order of the polynomial is specified prior to estimating the model, and can vary across groups. In SAS's PROC TRAJ implementation of group-based trajectory modeling, the polynomial order of each group's growth curve can range from 0 to 5. Below, for the sake of example, I present a cubic growth curve for each group j.

$$f(y_{it}) = \beta_0^j + \beta_1^j Age_{it} + \beta_2^j Age_{it}^2 + \beta_3^j Age_{it}^3 + \varepsilon_{it}$$
$$\pi_1, \pi_2, \ldots, \pi_J$$

$f(y_{it})$ refers to a link function for the outcome of interest. This can either be the censored normal, Poisson, or logit link. In all cases, we are modeling a polynomial on age/time that defines the shape of the trajectory for each group 1 through J. Although the TRAJ procedure lists group membership percents (π_j) among its parameter estimates, it actually estimates $J - 1$ parameters (θ_j) that are used to produce these probabilities. While group membership percents are bounded between 0 and 100 and must sum to 100, theta parameters are completely unbounded. This is useful in estimating the likelihood function. The two sets of parameters are related as follows:

$$\pi_j = \frac{e^{\theta_j}}{\sum_{j=1}^{J} e^{\theta_j}}$$

Because J − 1 group membership percents define the remaining group membership percent, only J − 1 parameters have to be estimated, and θ_j is set to 0, so that it equals 1 when exponentiated. This part of the estimation procedure is the same no matter which type of dependent variable is used.

The goal of estimation is to obtain a set of parameters that maximize $P(Y_i)$, the probability of observing $Y_{i,}$ the full set of individual developmental trajectories. The likelihood for each individual is obtained by summing the product of the estimated size of each group and the likelihood that the individual belongs to each group, J products for each individual. Estimates are obtained using maximum likelihood estimation with a quasi-Newton algorithm. As a by-product, if the assumptions of the model are sound, it benefits from known properties of maximum likelihood estimates.

Importantly, the model assumes conditional independence: conditional on membership in group j, deviations from the group-specific growth curve at time t is not correlated with deviations from the group-specific growth curve at time $t - 1$. In other words, conditioning on group membership, there is no serial correlation in the residuals. This is a stronger assumption than that made by standard growth curve models (CIA at the individual-level), but growth curve models also assume a multivariate distribution of parameters, which group-based models do not.

The three types of link functions entail additional parameters and modeling decisions. Censored normal models allow for censoring at either the upper or lower bounds and can also accommodate uncensored normal distributions. Lower and upper censoring limits must be provided to model this kind of distribution. If limits are set that are far outside the observed range of data, these limits do not figure heavily in the likelihood function and essentially the uncensored normal distribution is estimated. Otherwise, the model assumes the existence of a latent trait y^* that is observed only if the dependent variable is within the upper or lower censoring bounds. When y^* is outside the bounds, the observed value, y, is equal to the upper or lower bound. Additionally, censored normal models have the option of estimation of random intercepts and slopes, allowing one to estimate a hierarchical linear model with one group or growth mixture models with multiple groups. Other link functions do not provide this option in PROC TRAJ.

The logit distribution is appropriate for dichotomous outcomes such as arrest, conviction, incarceration, gang membership, high school dropout, etc. Similar to the censored normal distribution, the link function for the logistic distribution assumes the existence of a latent trait y* such that $y = 1$ if $y^* > 0$ and $y = 0$ if $y^* \leq 0$. If left untransformed, the growth curves from the logistic model are in terms of y* which is the log odds that $y = 1$.

When the dependent variable is count data such as number of arrests, the Poisson distribution may be appropriate. This distribution requires that the dependent variable take only nonnegative integer values. This type of model was in fact the first presentation of group-based trajectory models because "lambda" in the criminal careers model corresponds directly to lambda in the Poisson model: the rate of offending. Nagin and Land (1993) added a parameter for intermittency that doubles as a way to account for more zero counts than expected from the Poisson distribution. This zero-inflated Poisson (ZIP) model breaks up the probability of a zero count into two components: the probability of a zero count because lambda actually equals zero, and the probability of a zero count by chance when lambda is greater than zero. There are several options for how to specify this model using the "iorder" command in PROC TRAJ. The simplest is to estimate a single parameter for zero inflation across all groups. This can be relaxed in two ways: by estimating group-specific zero-inflation parameters, and by estimating higher order polynomials for the zero-inflation function. Additionally, exposure time can be incorporated into the model to account for different intervals between waves or across units.

Further details on the likelihood functions for these three distributions can be found in Nagin (2005, pp. 28–36).

Choosing the Final Model

PROC TRAJ does not directly reveal the best number of groups. Rather, given a set of data, a model specification, a predetermined number of groups, and starting values (optional), it obtains parameter estimates. There is a danger that these estimates represent local solutions, particularly with very complicated models. The only way to guard against this is to try different sets of starting values to determine if a better solution is obtained. But even the best solution for a specified parameterization and number of groups may not be the best overall solution. A major part of finalizing a group-based trajectory model is choosing the number of groups, tied up with this is the choice of polynomial order and other characteristics of the model.

The model choice set consists of the set of models considered as candidates for the final model. Nagin (2005) recommends a hierarchical decision-making process to reduce the number of models in this set, since it is impossible to try every possibility. The first step is to choose the optimal number of groups. Holding constant other modeling options, vary only the number of groups and assess model adequacy to determine the optimal number of groups. In the second step, specifications for each group are altered until the best solution is obtained, holding constant the number of groups. For example, in Poisson models, should zero inflation be modeled, and if so, should it be general or group-specific and should it be constant or vary over time? The optimal number of groups with general zero inflation may be different from the optimal number of groups with group-specific zero inflation, so this decision point may need to be included in the model choice set at Step 1.

There are a variety of criteria for arbitrating between models to choose the best solution. The most commonly used criterion is the BIC score, based on the log likelihood. The model with the higher BIC score (less negative) is preferred. Given a set of models, the probability that model j is the "correct" model is given by:

$$\frac{e^{(BIC_j - BIC_{MAX})}}{\sum_{1}^{J} e^{(BIC_j - BIC_{MAX})}}$$

While this may seem to provide a clear guideline for choosing the right model, it is sometimes ignored or downplayed, particularly when the optimal solution according to the BIC score is a very large number of groups. Other quantitative criteria can provide evidence of model adequacy as well.

After model estimation, the probability that each unit belongs to each group can be calculated using the likelihood function. These posterior probabilities are part of the PROC TRAJ output dataset. Each unit's maximum posterior probability is used to assign units to groups. The maximum certainty for group membership is 1, and the minimum is anything larger than 1/*J*. The mean posterior probability of group assignment conditional on assignment to that group (AvePP) is indicative of how well that group was identified. Nagin (2005, p. 88) provides a rule of thumb that these average posterior probabilities should be above .7 for each group. However, there is no formal test using this measure that indicates whether a model should be rejected. The minimum group-specific average posterior probability can be compared across models to get a sense of better-performing models. Additionally, an overall average posterior probability can be obtained by averaging every unit's maximum posterior probability, equivalent to a weighted average of group-specific average posterior probabilities, and a rough correlate of entropy, estimated in MPLUS.

Second, combining group-specific average posterior probabilities and estimated group sizes, one can calculate group-specific odds of correct classification. This essentially compares the ratio of the odds of correctly classifying individuals into group j based on the AvePP value, to the odds of correctly classifying individuals into group *j* based solely on the estimated proportion of the sample that belongs in group *j*. It is an odds ratio:

$$OCC_j = \frac{AvePP_j/(1 - AvePP_j)}{\hat{\pi}_j/(1 - \hat{\pi}_j)}$$

where AvePPj is the group-specific average posterior probability. Nagin's (2005, pp. 89) rule of thumb is that each OCC should be above 5. Again, however, this is not a formal test, but useful for comparing fit across models.

Nagin also suggests that estimated group probabilities can be compared to the proportion of the sample assigned to the group. Models with more divergent figures are less preferred. No hard limit or even rule of thumb is suggested however, other than "reasonably close correspondence" (2005:89).

PROC TRAJ also has the capability to produce confidence intervals for group membership probabilities and group-specific trajectories. These can also be produced using traditional or parametric bootstrapping. In general, as these confidence intervals widen, models are less preferred.

Finally, solution *J* + 1 may be rejected because the additional group is not substantively different from any of the groups in the *J* group solution, for example, when one group is split into two parallel groups. Sometimes a solution is rejected because it identifies an additional group estimated to comprise only a very small fraction of the population, which has limited external validity.

Focusing exclusively on the BIC score can lead to models with little utility. For example, when using a very rich dataset with large N and large T, the upper limit on the number of groups may outstrip their empirical and/or theoretical utility. A more reasonable approach takes into account a variety of model diagnostics to choose the best solution. Unfortunately, there is no ironclad rule for any of these diagnostics. Even the BIC, which seems to offer some certainty to the model choice problem, is highly influenced by the model choice set, and can lead to models which seem to capitalize on random noise. In the end, professional judgment must be exercised in choosing the best model. Taking all diagnostics into consideration, the most defendable and useful model should be retained.

Post-Estimation

As noted in the previous section, after the model is estimated, for each individual and each group, a posterior probability of group membership is estimated. In addition, a categorical variable is created based on posterior probabilities that classifies each unit in a group based on the maximum posterior probability. These are useful in many ways besides model diagnostics. First, they can be used to create group profiles. Second, they can be used as control variables in regression models. Finally, they can be incorporated into matching frameworks.

For each application, there are two options: classify and analyze vs. expected value methods. To illustrate, if we wish to determine the proportion of each group that are male, using the classify-analyze method, we would simply cross-tabulate the categorical group variable with the gender variable. The main shortcoming of the classify and analyze method is that it does not take into account uncertainty in group assignment. When average probability of group membership in a particular group is 75 %, for example, it does not seem advisable to take categorization into the group at face value. Using posterior probabilities takes this uncertainty into account. The expected value method requires a few extra steps to estimate the proportion male in each group. To estimate the proportion of group 1 that is male, calculate the mean for the "male" variable using posterior probabilities for group 1 as a weight. This is repeated for each group. Each person contributes to this group-specific mean proportional to the probability of membership in that group.

The same options are available when incorporating measurements of group membership into regression or matching models. One danger in this type of application, however, is that regardless of which method is chosen, estimation error in group classification is not taken into account. Failing to take this estimation error into account results in biased standard errors. This measurement error can be accounted for using bootstrapping methods.

Model Extensions

Since introducing the PROC TRAJ procedure (Jones et al. 2001), Nagin and colleagues have added a number of modifications and extensions to the model (Haviland et al. 2011; Haviland and Nagin 2005, 2007; Jones and Nagin 2007; Nagin 2005). The discussion here will focus on bootstrapping to obtain confidence intervals, testing equality of coefficients, trajectory covariates, and group membership predictors. Dual trajectory analysis, multitrajectory analysis and incorporating group-based trajectory modeling into matching protocols will be briefly touched upon as well, with reference to other sources for a full treatment.

We can place confidence intervals around any single parameter estimated via group-based trajectory modeling since standard errors are provided for every estimated parameter. However, certain values of interest are functions of multiple estimated parameters. Group membership probabilities are a nonlinear function of $J - 1$ theta parameters and the growth curves themselves are products of polynomials sometimes transformed through a link function or otherwise modified through extra parameters of the Poisson and censored normal specifications. There are several approaches that can be taken to address this problem and place confidence intervals around these estimates.

A typical approach to bootstrapping would create a large number (500, for example) of samples by resampling from the original dataset with replacement, and reestimate the trajectory model on each of the samples to generate a distribution of parameters, group membership proportions, and/or levels of each group's developmental trajectory to obtain a confidence interval free of distributional assumptions. While this approach is possible using PROC TRAJ, it can be very time intensive. Further, it's not clear that trajectory groups have consistent meaning across bootstrap samples.

Nagin (2005) suggests using parametric bootstrapping to obtain confidence intervals. Instead of resampling and reestimating the model, parametric bootstrapping simulates the exercise by using the means of the parameter estimates and the variance/covariance matrix. These are taken as values defining a multivariate normal distribution, for which a large number of draws are estimated, and then confidence intervals are obtained the same way they usually are, by appropriate percentile ranks. Because the multivariate normal distribution is well defined, taking even 10,000 replications is much easier than even 500 replications via traditional bootstrapping.

Jones and Nagin (2007) propose a third alternative for generating confidence intervals around growth curves. They introduced a "ci95m" option in PROC TRAJ that provides 95 % confidence intervals around growth curves using Taylor polynomial expansion to approximate the standard errors of combinations of parameters. These confidence intervals can be plotted using the "%trajplotnew" macro. Regardless of how confidence intervals are obtained, they are useful for establishing the precision of a group-based trajectory model. Distinct groups become less interesting if their confidence intervals overlap. Similarly, we change our assessment of a 5 % high-rate offending group if we learn that the confidence interval for the group's size ranges from 1 % to 20 % versus 4 % to 6 %.

Jones and Nagin (2007) also introduced an important macro add-on for PROC TRAJ that allows one to easily conduct Wald tests for equality of coefficients (%trajtest). This can be used to test equality of growth parameters across groups and has a number of important applications in combination with other extensions of PROC TRAJ described below.

The life-course paradigm organizes the study of human development around longitudinal trajectories, life transitions, and turning points. If a life transition shifts an individual's developmental trajectory onto a new path, it is a turning point. This suggests a counterfactual type of analysis, answering the question of what would have happened for individuals in situations which they did not experience. Group trajectories, because they represent a group of individuals that are similar with respect to their developmental path and related covariates, provide an important source of plausible counterfactuals to test for these kinds of effects. Group-specific trajectories can easily be generalized to accommodate *m* time-varying covariates (x_1 through x_m):

$$f(y_{it}) = \beta_0^j + \beta_1^j Age_{it} + \beta_2^j Age_{it}^2 + \beta_3^j Age_{it}^3 + \alpha_1^j x_{1it} + \alpha_2^j x_{2it} + \ldots + \alpha_m^j x_{mit} + \varepsilon_{it}$$

Trajectory covariates can test a wide range of hypotheses from life-course criminology. The specific interpretation of these covariates depends on how they are entered into the model. The simplest strategy is to enter a series of dummy variables reflecting being in a particular life state such as being married, in a gang, employed, etc. This results in group-specific estimates that are easy to interpret, but does not capture the dynamic element of turning points. It does allow for an easy test of whether the effects of that life state depend on group membership, using the %trajtest macro. It assumes constant effects over time/age and over years in the state. If marriage is introduced in this manner, it imposes an assumption that the effect of marriage depends neither on when a person gets married nor length of marriage. A more flexible method consistent with the life-course paradigm would enter a dummy variable for being in a particular state, age interacted with this dummy variable, as well as a counter variable for years in the state. This would allow the effect of marriage to vary according to when one gets married and to increase or decay depending on how long one stays married.

While posterior probabilities and group membership categories can be used to create group risk profiles, these profiles are of limited use because of error estimates biased due to lack of incorporation of group estimation error. But these risk variables can be incorporated directly into trajectory models, allowing all parameters to be estimated at the same time and with correct standard errors.

The trajectory model changes in several important ways when risk variables are incorporated into it. The shape of the trajectories in the optimal solution may change, and posterior probabilities depend not only on levels of the dependent variable but also on risk variables. Group membership probabilities are estimated conditional on a set of risk variables (x_1 through x_r, for *r* risk variables). Multiple thetas are estimated for each group. These are used to generate predicted group probabilities for specified sets of risk variables.

$$\theta_j = \sum_{r=1}^{R} \theta_{jr} x_r = \theta_{j0} + \theta_{j1} x_1 + \theta_{j2} x_2 + \ldots + \theta_{jR} x_R$$

With risk variables, the intercept theta estimates can be used to calculate group membership probabilities when all risk variables are equal to zero. If all risk variables are centered at their means before estimating the trajectory model, then the intercepts would represent theta estimates at the average. The differential impact of risk variables on group membership probabilities can be tested using the %trajtest macro, and confidence intervals around theta estimates or group membership probabilities for specific sets of risk variables can be calculated using bootstrapping.

When more than one dependent variable is of interest, either because of comorbidity of developmental trajectories, heterotypic continuity or a developmental predictor of a developmental outcome, a couple options are available in PROC TRAJ. Dual trajectory analysis (Jones and Nagin 2007; Nagin 2005, Chap. 8) is a very flexible extension of standard trajectory analysis. The basic requirement is that two dependent variables are meaningfully linked via the unit of analysis. Each unit should have repeated observations for both dependent variables. The dependent variables need not be distributed the same way, need not be measured at the same time nor have the same number of measurements, and the estimated number of groups need not be identical. The two dependent variables are linked through a shared variance-covariance matrix and a set of group membership probabilities for one series that are conditional on group membership in the other series. These allow the full set of unconditional, conditional, and joint group membership probabilities to be estimated. Conditional group membership predictors can be included as well.

When more than two dependent variables are of interest, the parameter space for extending the dual trajectory analysis quickly becomes unwieldy. But when the focus is on only two or three dependent variables, multitrajectory modeling is available in PROC TRAJ using the "MULTGROUPS" option (Jones and Nagin 2007). This type of model is constrained in that no cross-classification is estimated across dependent variable types. Instead, each group is defined by a set of trajectories across two or three dependent variables. This simplification greatly reduces the number of parameters that need to be estimated while still providing a rich portrayal of group heterogeneity. As with dual trajectory analysis, the distribution of the dependent variables, the number of observations (T) need not match, and the timeframe need not overlap. The main restriction is that the number of groups must be the same, and of course no cross-classification is estimated.

Recent work incorporates group-based trajectory modeling into a matching framework (Haviland and Nagin 2005, 2007). The basic insight of this work is that developmental histories can serve as important proxies for more complicated processes, and matching within developmental history group serves to balance numerous characteristics – not only the dependent variable of the trajectory model, but numerous related developmental trajectories and risk variables correlated with these trajectories. Haviland and Nagin (2005, p. 4) think of these groups as "latent strata in the data." Of course, the power of this method depends quite a bit on the extent to which the developmental history being modeled is linked to the selection process for the treatment of interest.

If treatment is random conditional on group membership, this method can provide a convincing measure of turning point effects, where estimated effects are group specific, but can be aggregated through a weighted average to obtain population average effects. Haviland and Nagin's method establishes developmental paths preceding potential turning points. Within each path, there are two possibilities, one might experience the turning point event, or not. Those individuals within the developmental group who do not experience the turning point event serve as natural counterfactuals for those that do. Assessing differences longitudinally after the turning point event allows us to see whether differences between the two groups emerge, and whether those differences increase, remain stable, or decay over time. To the extent that they persist, evidence for the existence of a turning point is shown. Life-course research also suggests that the response to a turning point event may depend on prior developmental histories. This method allows one to explicitly test that hypothesis. To the extent that groups do not balance important treatment predictors, the group-based trajectory model can be incorporated into a more general propensity score framework, either with propensity score models embedded within each trajectory group, or posterior probabilities of group membership incorporated into a single propensity score model.

Controversies

Group-based trajectory modeling has attracted an unusual amount of criticism of a surprisingly harsh tone. Objections to group-based trajectory models center around the meaning of "group" and the relative merits of alternative statistical models.

Since the concentration of offending was publicized by Wolfgang and colleagues (1972), the search for the high-rate offending group *before* they engage in the bulk of their crimes has taken up a great deal of resources. In fact, reliably prospectively identifying the chronic offender has been something of a white whale of criminology for decades. Blumstein and Moitra (1980) showed that retrospective identification of groups in Wolfgang et al. did not hold up prospectively, so that the "chronic offender" label was not useful for selective incapacitation of "career criminals." More recently, Sampson and Laub (2005) have shown that although retrospective offending groups can be identified in their Boston data, even very rich sets of theoretically based predictors cannot reliably distinguish between the groups prospectively. The central point is that even though high-rate offending "groups" can be identified retrospectively either through simple decision rules as in Wolfgang et al. (1972), or through more statistically sophisticated methods such as group-based trajectory models, clustering or other finite mixture models, prospectively, for policy and theoretical purposes, these "groups" do not exist. This is not to say that these groups have no significance, but that their identification in a group-based trajectory model is just the starting point. If the researcher wants to make the case that a specific group is important, that argument must be grounded in theory (Brame et al. 2012) and there must be some demonstration that the group exists outside a single group-based trajectory model.

Critics of this method either over-interpret the significance of groups or imply that because of the nature of the model, other less sophisticated researchers will reify the groups. But of course, the search for the "chronic offender" that Blumstein and Moitra (1980) debunked demonstrated that sophisticated models are not needed in order for reification of a high-rate offender typology to occur.

Many critiques of group-based trajectory modeling are premised on the notion that another statistical tool is superior. Generally speaking, there are three options for summarizing longitudinal developmental patterns: group-based trajectory models, hierarchical linear models, and growth mixture models. Each of these models simplifies a more complex reality in a different way. The differences boil down to how heterogeneity in developmental patterns is characterized. Hierarchical linear models characterize heterogeneity as a jointly normal distribution of parameters centered around the overall average. Group-based models characterize heterogeneity using a mixture of a finite number of growth curves that are support points of a complex distribution. Growth mixture models combine the two strategies, modeling jointly normal distributions of parameters around each group's average trajectory. Arbitrating between these choices depends on the nature of the research problem and the purposes of the analysis. Nagin (2005) points to complexity of the continuous distribution as an important guide when choosing between group-based and hierarchical linear models.

Future Directions

To date, although group-based trajectory modeling has been extensively employed in the criminological literature, its uses have usually been limited to descriptive exercises. More sophisticated applications such as Haviland and Nagin's matching method and correct modeling of standard errors using some form of bootstrapping have been less often employed. It is hoped that future work using the method will realize its full capabilities.

Despite the lack of uptake of these more sophisticated features, further development of the model has the potential to greatly benefit the field. The range of dependent variable distributions, although sufficient for most applications in criminology, could be expanded. For example, multinomial, categorical, and quantile group-based trajectories could have immediate applications. Advances such as multitrajectory modeling could be generalized to any number of dependent variables, allowing for complex characterizations of developmental history typologies and perhaps an even stronger basis for identifying latent strata in the data. And the problem of local solutions could be tackled and better understood by incorporating an automated start value algorithm, which is currently offered in the MPLUS implementation of the model.

Conclusion

Group-based trajectory modeling is a powerful tool that can be put to a very wide variety of uses alongside hierarchical and growth mixture models. Through careful, thoughtful use, it can test a number of important theoretical hypotheses (Brame et al. 2012).

Related Entries

- ▶ Age-Crime Curve
- ▶ Career Criminals and Criminological Theory
- ▶ Desistance from Crime
- ▶ Group-Based Trajectory Models and Developmental Change
- ▶ Identification Issues in Life Course Criminology
- ▶ Longitudinal Crime Trends at Places
- ▶ Longitudinal Studies in Criminology
- ▶ Moffitt's Developmental Taxonomy of Antisocial Behavior
- ▶ Onset of Offending
- ▶ Optimizing Longitudinal Studies in Offending
- ▶ Social Control and Self-Control Through the Life Course

Recommended Reading and References

Blumstein A, Moitra S (1980) The identification of "career criminals" form "chronic offenders" in a cohort. Law Policy 2:321–334

Blumstein A, Cohen J, Roth JA, Visher C (eds) (1986) Criminal careers and "career criminals". volume 1. National Academy Press, Washington, DC

Brame R, Paternoster R, Piquero AR (2012) Thoughts on the analysis of group-based developmental trajectories in criminology. Justice Q 29:469–490

Goring C (1913) The English convict. His Majesty's Printing Office, London

Haviland AM, Nagin DS (2005) Causal inferences with group based trajectory models. Psychometrika 70:557–578

Haviland AM, Nagin DS (2007) Using group-based trajectory modeling in conjunction with propensity scores to improve balance. J Exp Criminol 3:65–82

Haviland AM, Jones BL, Nagin DS (2011) Group-based trajectory modeling extended to account for nonrandom participant attrition. Soc Methods Res 40:367–390

Hipp JR, Bauer DJ (2006) Local solutions in the estimation of growth mixture models. Psychol Methods 11:36–53

Hirschi T, Gottfredson MD (1983) Age and the explanation of crime. Am J Soc 89:552–584

Jones BL, Nagin DS (2007) Advances in group-based trajectory modeling and a SAS procedure for estimating them. Soc Methods Res 35:542–571

Jones BL, Nagin DS, Roeder K (2001) A SAS procedure based on mixture models for estimating developmental trajectories. Soc Methods Res 29:374–393

Jr Elder GH (1994) Time, human agency, and social change: perspectives on the life course. Soc Psychol Q 57:4–15

Koide RT, Shumway DL, Bing X, Sharda JN (2007) On temporal partitioning of a community of ectomycorrhizal fungi. New Phytol 174:420–429

Laub JH, Sampson RJ (2005) Shared beginnings, divergent lives: delinquent boys to age 70. Harvard University Press, Boston

Mulvaney S, Warren Lambert E, Garber J, Walker LS (2006) Trajectories of symptoms and impairment for pediatric patients with functional abdominal pain: a 5-year longitudinal study. J Am Acad Child Adolesc Psychiatry 45:737–744

Nagin DS (2005) Group-based modeling of development. Harvard University Press, Cambridge, MA

Nagin DS, Land KC (1993) Age, criminal careers, and population heterogeneity: specification and estimation of a nonparametric, mixed Poisson model. Criminology 31:327–362

Piquero AR (2008) Taking stock of developmental trajectories of criminal activity over the life course. In: Liberman A (ed) The long view of crime: a synthesis of longitudinal research. Springer, New York

Sampson RJ, Laub JH (1993) Crime in the making: pathways and turning points through life. Harvard University Press, Boston

Wolfgang ME, Figlio RM, Sellin T (1972) Delinquency in a birth cohort. University of Chicago Press, Chicago

Group-Based Trajectory Models and Developmental Change

Robert Apel
School of Criminal Justice, Rutgers University,
Newark, NJ, USA

Overview

The widespread availability of panel data on unlawful behavior has stoked criminological interest in statistical methods that are capable of quantifying developmental change. One such method is the group-based trajectory model of Nagin (1999, 2005; Nagin and Land 1993), which has become an essential tool for quantitative criminologists who conduct analysis with longitudinal data. It has been widely utilized to study the development of delinquent and criminal behavior in a variety of American and cross-national contexts.

The appeal of the group-based trajectory model stems from the fact that it explicitly allows for heterogeneous developmental pathways, without imposing any assumptions about the distributional properties of the heterogeneity. Instead, heterogeneity of unknown form is approximated using a discrete mixture of two or more distributions. This is in contrast to other conventional panel data estimators, which require the analyst to impose untestable (ergo, unfalsifiable) assumptions about the form of the heterogeneity, namely that it varies continuously in the population (usually in the form of a normal distribution).

This essay proceeds as follows. First, some background on the group-based trajectory model, as used in criminology, will be provided. Second, details on the estimation of the model will be given by way of an empirical illustration. Third, more flexible uses of the model, beyond quantifying developmental change, will be described. The essay closes with some concluding remarks.

Background of the Group-Based Methodology

The rationale of the group-based trajectory model is provided by Heckman and Singer (1984), in an analysis of the determinants of unemployment duration. They argued that insufficient attention had been devoted to *population heterogeneity*, or temporally persistent unobservables which underlie relative differences across individuals in spell length. The problem is that failure to adequately control for population heterogeneity produces bias in structural parameters of interest. They observed further that theory rarely provides guidance about the distribution of unobservables: “The choice of a particular distribution of unobservables is usually justified on the grounds of familiarity, ease of manipulation, and considerations of computational cost” (p. 272). In a parametric analysis of single-spell duration data, they demonstrated that structural estimates (for age, education, marriage, tenure of the previous job, unemployment benefits, unemployment rate, and unemployment duration) were extremely sensitive to the distribution chosen to model unobservables, each of which was *ex ante* theoretically plausible. Indeed, they concluded that "there are as many different 'structural models' as there are distributions of heterogeneity" (p. 276).

In an econometrically dense exposition, Heckman and Singer (1984) then proposed a non-parametric estimator intended to minimize the impact of (arbitrary) distributional assumptions on parameter estimates. Their method entailed the specification of a finite number of underlying distributions or "points of support" for the distribution of unobserved heterogeneity, each with unique location (mean) and shape (variance) parameters. Importantly, no single parametric distribution produced estimates that harmonized fully with the non-parametric estimator. Furthermore, their simulation showed that, while it did not reliably estimate the underlying mixing distributions, the non-parametric estimator was quite successful at recovering structural parameters and reproducing sample duration times.

In criminology, two developments had a more immediate influence on the origin of the group-based methodology: the criminal career paradigm of the 1980s and taxonomic theories of unlawful behavior of the 1990s. First, the 1980s gave rise to the criminal career perspective, which partitions the aggregate age distribution of crime into two distinct phenomena: the fraction of the population that is criminally active at any given age (participation), and the number of crimes committed per year by any given active criminal (frequency) (Blumstein et al. 1986). Crime prevention policies could then be classified as to whether their impact on the crime rate was through participation or frequency. This perspective led to the parameterization of the essential features of an individual's "criminal career," including age of initiation (A_0), age of termination (A_N), length of the criminal career ($T = A_N - A_0$), and offending frequency (λ, or lambda), or the number of crimes committed per year while criminally active.

The second major influence on the development of the group-based trajectory model was taxonomic theorizing about the unfolding of delinquent and criminal behavior over the life span, most prominently in the work of Moffitt (1993). She theorized that the population is comprised of two distinct groups or "taxons" that differ in meaningful ways in the causes, consequences, and developmental courses of unlawful behavior. The "life-course-persistent" pathway characterizes a comparatively small fraction of the population, but is distinguished by early onset, temporal persistence, and involvement in a large class of antisocial behaviors, originating in neuropsychological deficits that interact with environmental disadvantages. The "adolescence-limited" pathway, on the other hand, characterizes the lion's share of the population, and is distinguished by pubertal onset, acquisition from mimicry, maintenance through social reinforcement, and discontinuity as the maturity gap is successfully bridged. In short, the life-course-persistent pathway is pathological, whereas the adolescence-limited pathway is normative.

A key empirical requirement of both the criminal career paradigm and taxonomic theory is availability of long-term panel data on individual offending, along with a grouping procedure capable of identifying distinct classes of offenders. With panel data and latent class methodology, then, it becomes possible to answer a variety of unresolved questions: Is λ constant or age graded over the duration of the criminal career? Do all individuals follow the same basic criminal career, perhaps differing only in degree, or are there distinct groups in the population? How many groups exist in the population? Do different groups of offenders have distinct etiological pathways?

With the foregoing econometric and criminological developments as background, Nagin and Land (1993) introduced the group-based trajectory model to criminology. They applied the non-parametric approach of Heckman and Singer (1984) to panel data on criminal behavior, in the process providing preliminary answers to questions of interest for the criminal career paradigm and taxonomic theories. In their analysis, they used data on criminal conviction frequency spanning ages 10–30, from the Cambridge Study in Delinquent Development. They characterized their method as a non-parametric, mixed Poisson model. Their key findings are summarized below, but statistical details on their model are reserved for the next section.

Nagin and Land (1993) settled on a four-group solution, producing a number of important findings. The first result concerned the form of λ over time, and revealed that conviction frequency has a pronounced relationship with age. When they constrained the sample to have a common age trajectory (i.e., a common peak age) that differed only in level, conviction frequency was shown to reach a single peak at age 18 and then decline by a considerable margin by age 30. The second result addressed heterogeneity in criminal careers, and revealed substantial differences across groups in the essential features of the age distribution of criminal conviction. Peak ages ranged from 14 to 22, and the groups varied considerably in the shape of the age-conviction curve, exhibiting a pronounced peak at an early age followed by a rapid decline ("adolescence limited"), slow acceleration and deceleration but the most frequent convictions at all ages ("high-rate chronic"), and a flatter, more chronic age profile ("low-rate chronic"). The third result concerned the differential predictability of group membership from a number of background risk factors, revealing a variety of differences in both degree and kind. For example, the high-rate chronic offenders were distinguished by the presence of delinquent siblings and more widespread involvement in antisocial behavior (e.g., lying, truancy, heavy drinking, marijuana use), while the low-rate chronic offenders were distinguished by unusually low intelligence. The adolescence-limited offenders, on the other hand, were distinguished from the chronic offending groups by more popularity among peers and better school performance.

The Nagin and Land (1993) study was closely followed by further efforts to explore the utility of the group-based trajectory model for studying the essential features of criminal careers (D'Unger et al. 1998; Land et al. 1996; Land and Nagin 1996; Nagin et al. 1995). Across numerous datasets, these studies have found that there is a general tendency to extract four–five trajectory groups, with some commonality in the basic contours of the trajectory profiles. One group can be classified as non-offenders or intermittent offenders, one group can be described as adolescence-limited offenders, and one group can be classified as chronic offenders. Oftentimes, the latter can be further stratified into two subgroups: a high-rate chronic and low-rate chronic group.

Estimation and Evaluation of the Group-Based Trajectory Model

The group-based trajectory model is a semi-parametric approach to modeling longitudinal outcomes. Described simply, the method groups subjects into more homogeneous latent classes based on their longitudinal sequence of behavior. It is in the class of finite mixture models discussed at length by McLachlan and Peel (2000). The group-based model is parametric by specifying a probability distribution governing the realization of the response variable (e.g., normal, logistic, Poisson), as well as by specifying the functional form of the response as a polynomial in age or time. The group-based model is non-parametric in terms of specifying the population as a multinomial mixture of two or more such distributions ("points of support"), with the parameters of each component of the mixture estimated freely. The group-based trajectory model has been shown to have desirable properties in applications where the components in the mixture are well separated (Brame et al. 2006; Loughran and Nagin 2006).

To walk readers through an empirical application, data on self-report offending from the National Longitudinal Survey of Youth 1997 (NLSY97) will be analyzed. The response variable is a variety scale composed of 20 dummy indicators, ranging in seriousness from minor offenses such as petty larceny (less than $50) and destruction of property, to more serious offenses such as aggravated assault and robbery. For each offense, respondents are coded "1" if they reported engaging in the behavior since the last interview (or ever prior to the initial

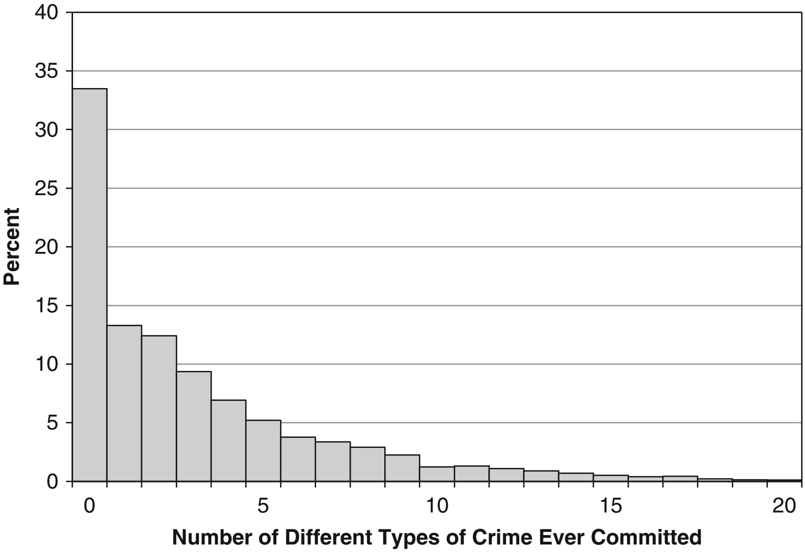

Group-Based Trajectory Models and Developmental Change, Fig. 1 Histogram of cumulative self-report crime variety (Note: $N = 8{,}984$. Percentages are weighted. This figure represents the distribution of delinquent/criminal behavior across the first seven waves of the NLSY97, using all available information from respondents. Source: National Longitudinal Survey of Youth 1997, Rounds 1–7)

interview) and coded "0" otherwise. These indicators are then summed at each interview, yielding a measure of the number of different kinds of criminal acts committed. A histogram of the cumulative distribution of the variety scale is provided in Fig. 1. To create this figure, a lifetime indicator (using information from all available interviews) for each of the 20 criminal acts was constructed for each respondent, and was then summed. Two-thirds of the NLSY97 sample (66.5 %) reports involvement in unlawful behavior at some point during the first seven interview waves.

Before proceeding with the analysis, Fig. 2 provides an illustration of the age-graded nature of delinquency/crime in the NLSY97 data. It exhibits the classic features of the age-crime curve, with a peak of almost 1.3 different types of crime per year during the 14–16 age range, followed by a steady decline through the late teens and a leveling out at under 0.3 in the early 20s. Basically, the group-based trajectory model will be used to ascertain whether this average developmental pathway characterizes all youth in the sample (to a greater or lesser degree, at least), or whether qualitatively distinct pathways are evident in the data.

To formalize the group-based trajectory model, Y_{it} represents a count of the number of different kinds of criminal behaviors (i.e., crime variety) reported by subject i (i = 1,. . .,N) in time t (t = 1,. . .,T). It is natural to presume that Y_{it} is a Poisson random variable with density:

$$f(Y_{it}) = Pr(Y_{it} = y|\lambda_{it}) = \frac{e^{-\lambda_{it}}\lambda_{it}{}^{y}}{y!}$$

wherein λ_{it} represents an individual's underlying rate of offending, or the number of different kinds of crimes committed per unit of time (e.g., per year). An individual's (logged) offending rate, given membership in trajectory group j, is then modeled as a polynomial function in age:

$$ln\left(\lambda^{j}_{it}\right) = \beta^{j}_{0} + \beta^{j}_{1}Age_{it} + \beta^{j}_{2}Age^{2}_{it} + \beta^{j}_{3}Age^{3}_{it}$$

Here, age is assumed to follow a cubic functional form, although other polynomials are obviously possible. The parameters define the shape of the trajectory, and the j superscripts denote group-specific parameters for j = 1,. . .,M trajectory groups. So if a one-group model is specified, four shape parameters are estimated; if a two-group model is specified, eight shape parameters are estimated; and so on.

Estimation of the group-based trajectory model proceeds by way of maximum likelihood.

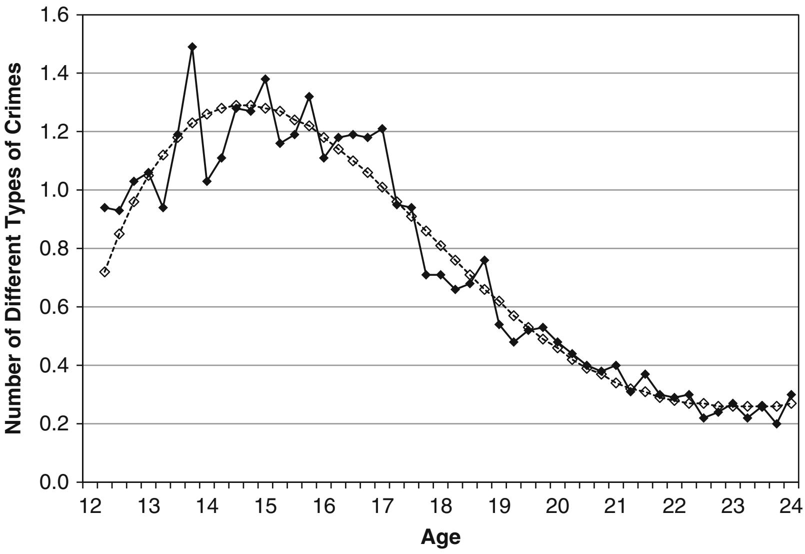

Group-Based Trajectory Models and Developmental Change, Fig. 2 Age distribution of self-report crime variety (Note: $N = 8{,}984$. Estimates are weighted but are not exposure adjusted. The *dashed lines* are predicted values fit from a fourth-order polynomial in a linear model with random effects. Source: National Longitudinal Survey of Youth 1997, Rounds 1–7)

Define a vector $\mathbf{Y_i}$ to represent the longitudinal sequence of individual i's self-report crime variety:

$$\mathbf{Y_i} = \{Y_{i1}, Y_{i2}, \ldots, Y_{iT}\}$$

Then let P($\mathbf{Y_i}$|j) represent the conditional distribution of $\mathbf{Y_i}$ given membership in group j, in other words, the probability of observing individual i's crime "trajectory" conditional on his being in group j. In the current example, this is constructed from the Poisson probability distribution. An important assumption of the group-based trajectory model concerns the conditional independence of the sequential realizations of Y_{it}, which allows us to write the conditional distribution of $\mathbf{Y_i}$ as:

$$P(\mathbf{Y_i}|j) = \prod_{t=1}^{T} f(Y_{it}|j) = \prod_{t=1}^{T} \frac{e^{-\lambda_{it}^{j}} \lambda_{it}^{j\ y}}{y!}$$

where

$$\lambda_{it}^{j} = e^{\beta_0^j + \beta_1^j Age_{it} + \beta_2^j Age_{it}^2 + \beta_3^j Age_{it}^3}$$

Now, let π_j represent the probability of membership in group j. Thus formalized, the unconditional probability of observing individual i's longitudinal sequence, or the marginal density of $\mathbf{Y_i}$, can then be recovered by aggregating across M conditional distributions:

$$P(\mathbf{Y_i}) = \sum_{j=1}^{M} \pi_j P(\mathbf{Y_i}|j)$$

Finally, in a sample of N independent individuals, the sample likelihood is formed by the product of N such marginal densities:

$$L\left(\pi_j, \beta_0^j, \beta_1^j, \beta_2^j, \beta_3^j | Y_{it}, Age_{it}\right) = \prod_{i=1}^{N} P(\mathbf{Y_i})$$

A SAS-based procedure, Proc Traj, has been developed by Jones et al. (2001; Jones and Nagin 2007) to estimate the group-based trajectory model as parameterized above. One practical challenge to estimation of group-based trajectories is model selection, specifically, decision making about the optimal number of trajectory groups. Nagin (1999) advocates use of the Bayesian Information Criterion (BIC) rather than the log likelihood, because trajectory models are not formally nested in the way required to conduct a likelihood ratio test. The BIC is estimated by:

$$BIC = ln(L) - \frac{1}{2} K ln(N)$$

Group-Based Trajectory Models and Developmental Change, Table 1 Model summaries from trajectory group solutions

Model summary	2-group solution	3-group solution	4-group solution	5-group solution
Parameter estimates	9	14	19	24
Log likelihood	−66,247.15	−62,187.58	−60,521.25	−59,388.47
BIC	−66,288.11	−62,251.30	−60,607.74	−59,497.71
2 × (ΔBIC)	–	8,073.62	3,287.12	2,220.06
Probability correct model	0.00	0.00	0.00	1.00

Note: $N = 8{,}984$. The models use information on self-report crime variety from the first (1997) to the seventh (2003) interviews. Although the coefficients are not shown, a cubic functional form in age is specified for each trajectory group. The first row from the bottom is the log Bayes factor approximation, and contrasts a model with j + 1 trajectory groups to a model with j trajectory groups. Values in excess of 10 are regarded as very strongly supportive of a model with j + 1 trajectory groups. The bottom row provides the posterior probability that the referent model is the correct model, when equal weight is placed on the prior probability that each model is the correct one

where ln(L) is the log likelihood, K denotes the number of parameters, and ln(N) is the log sample size.

Table 1 provides the BIC and other quantities of interest from models assuming two–five Poisson mixtures for the NLSY97 data. Computing twice the difference in BIC's between two models provides a useful guide to decision making about the number of components in the mixture (Jones et al. 2001), as does the posterior probability that any given model is "correct" when more than two models are being compared and equal weight is placed on their prior probabilities (Nagin 1999). The latter is estimated by the formula:

$$\frac{e^{BIC_j - BIC_{max}}}{\sum_j e^{BIC_j - BIC_{max}}}$$

where BIC_j is the BIC from the referent model, and BIC_{max} is the maximum BIC of all of the models under consideration. By both of these criteria, Table 1 shows that the preferred solution is the model with five trajectory groups. (Note that this example is not intended to be a formal analysis of the optimal number and shape of the crime trajectories. It is strictly intended to serve as an illustration. A formal analysis for this author would require sequentially adding more components to the mixture until the model can no longer converge, adjusting the order of the polynomials for the trajectory groups to identify the optimal fit, examining the utility of including an inflation parameter in the Poisson model, and experimenting with a variety of start values to evaluate the robustness of the final solution).

Results from the five-group solution are provided in Table 2. The output from Proc Traj includes the parameter estimates characterizing the shape of the trajectory for each group and the proportion of the population that is estimated to follow group j's pathway, as well as estimates of the probability that a subject belongs to each trajectory group. The latter are known as "posterior probabilities" (not to be confused with the posterior probability that a model is correct, as estimated from the formula above). To aid interpretation of the parameter estimates characterizing the shape of the trajectories, they are plotted against age in Fig. 3.

One of the first things to notice about the parameter estimates and plotted trajectories is just how heterogeneous the developmental pathways of delinquent/criminal behavior are in the NLSY97. The model suggests that about one-half of youth – Group 1 (49 %) – follow a non-offending pathway, or at least a highly intermittent one, as the offending rate is non-zero but quite low during the teenage years. One-third of youth – Groups 3 and 4 (25 % and 8 %, respectively) – follow an adolescence-limited pathway that peaks at 14–15 years of age, but at very different levels. Specifically, Group 3 peaks at about 1.5 different crimes while Group 4 peaks at almost 4.5 different crimes. (Substantively, a value of 1.5 on the variety scale translates into about 3.5 total crimes per year – one criminal act every 3.5 months – while a value of 4.5 on the

Group-Based Trajectory Models and Developmental Change, Table 2 Parameter estimates and model diagnostics from the five-group trajectory model

	Group 1	Group 2	Group 3	Group 4	Group 5
Coefficients					
Constant	13.999 (11.4)	30.438 (11.0)**	−56.527 (13.1)***	−65.660 (14.9)***	−27.751 (8.39)***
Age	−1.861 (2.06)	−5.885 (1.87)**	10.124 (2.34)***	11.082 (2.57)***	4.251 (1.40)**
Age Squared	0.055 (0.12)	0.365 (0.11)***	−0.575 (0.14)***	−0.587 (0.15)***	−0.199 (0.08)**
Age Cubed	0.000 (0.00)	−0.007 (0.00)***	0.010 (0.00)***	0.010 (0.00)***	0.003 (0.00)*
Diagnostics					
Population (%)	46.0%	13.9%	26.1%	9.0%	4.9%
Sample (%)	48.8%	12.0%	25.4%	8.2%	4.7%
Mean post. prob.	0.91	0.87	0.83	0.86	0.93
OCC	11.9	41.5	13.8	62.1	257.9

Note: N = 8,984. Estimates are exposure adjusted and weighted. The age profiles for this model are displayed in Fig. 2
* p < .05, ** p < .01, *** p < .001 (two-tailed tests)

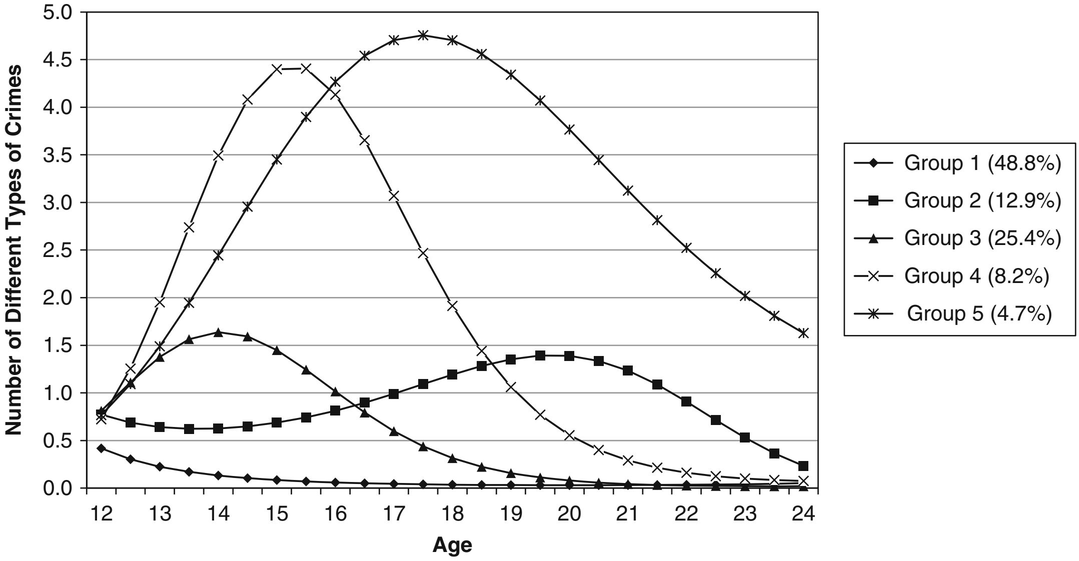

Group-Based Trajectory Models and Developmental Change, Fig. 3 Age distribution of self-report crime variety, by Trajectory Group (Note: Parameter estimates for this trajectory model are shown in Table 2)

variety scale represents about 23.2 total crimes per year – about two criminal acts per month. This was determined by mapping mean crime frequency onto crime variety.) A bit less than one-fifth of youth – Groups 2 and 5 (13 % and 5 %, respectively) – appear to follow a chronic pathway, but again at very different levels. Group 2 peaks at about 1.5 different crimes while Group 5 peaks at more than 4.5 different crimes. However, both groups continue offending well into their 20's, when all other youth have, for all intents and purposes, desisted from crime.

Having selected the optimal model and described the five trajectory groups, an important next step is to diagnose how well the model fits the data. Nagin (2005) outlines a variety of diagnostic criteria, including the discrepancy between the proportions of the population and sample in each trajectory group, the mean posterior probabilities of group assignment, and the odds of correct classification (OCC). Each of these criteria is shown in the bottom half of Table 2.

One diagnostic involves a comparison of the proportion of the population that is estimated to follow a given trajectory, with the proportion of the sample that is “assigned” to the same trajectory. The former proportion, π_j, is a component of the likelihood function:

$$\pi_j = \frac{e^{\theta_j}}{\sum_{j=1}^{M} e^{\theta_j}}$$

where the normalization, $\theta_1 = 0$, ensures identifiability (note, $e^0 = 1$). The sample counterpart is derived from the posterior probabilities, or the model-predicted probability that an individual is assigned to each trajectory group. Formally, the probability that any given sample subject is assigned to trajectory group j, given her longitudinal sequence of criminal behavior, is estimated by appealing to Bayes’ theorem:

$$P(j|\mathbf{Y_i}) = \frac{\pi_j P(\mathbf{Y_i}|j)}{\sum_{j=1}^{M} \pi_j P(\mathbf{Y_i}|j)}$$

The proportion of the sample assigned to each group is determined using the maximum posterior probability assignment rule, in which each subject is classified into the trajectory group with the highest posterior probability. The population and sample proportions shown in Table 2 suggest no major discrepancy between these two proportions.

A second useful diagnostic is an inspection of the mean posterior probabilities. These are determined by assigning subjects to trajectory groups using the maximum posterior probability assignment rule, and then computing the mean posterior probability among the individuals assigned to each group. Nagin (2005) suggests that a mean in excess of 0.7 is evidence of good classification. In the current example, shown in Table 2, the lowest mean posterior probability is 0.83 (Group 3), indicating good model fit by this criterion.

A third diagnostic is the “odds of correct classification” (OCC), which uses the mean posterior probability as well as the population proportions. It is computed for each trajectory group in the following manner:

$$OCC = \frac{\overline{\mathrm{P}}(j|\mathbf{Y_i})/(1-\overline{\mathrm{P}}(j|\mathbf{Y_i}))}{\widehat{\pi}_j/(1-\widehat{\pi}_j)}$$

Nagin (2005) indicates that an OCC in excess of 5.0 indicates good model fit. Among the estimates in Table 2, the smallest OCC is 11.9 (Group 1), confirming that the model provides good classification of sample subjects to trajectory groups.

Extensions of the Group-Based Methodology

There are manifold uses for the group-based trajectory model, beyond its clear descriptive capability as shown in the previous section. For example, the model can be used to quantify the degree to which the developmental history of crime modifies the causal effect of a nonrandom “treatment” on subsequent criminality (e.g., employment, school dropout, marriage, gang membership, incarceration). There are at least two pernicious problems in the estimation of treatment effects on crime from nonexperimental data, which the trajectory methodology is well positioned to remedy. The first problem is that treatment assignment may not be independent of the response variable of interest. The second problem is that the impact of treatment may not be uniform across the population of interest. Manski (1995) refers to the first as the *selection problem*, and to the second as the *mixing problem*.

The selection and mixing problems were demonstrated in a study by Apel et al. (2007), who were interested in quantifying the effect of first-time, formal employment at age 16 on delinquent behavior. They estimated trajectories of delinquency from ages 11 to 15 among those who had not yet transitioned into the formal labor market and identified four distinctive clusters of individuals – conformists, decliners, low-level risers, and high-level risers. The selection problem was evidenced by the fact that the low- and high-level risers were more likely to be employed in excess of 20 h per week at age 16, confirming the differential selection of highly delinquent youth into “intensive” employment. The mixing problem was illustrated by the fact that, whereas for most youth the treatment effect was zero, the high-level risers experienced a significant decline in delinquent behavior when they began working intensively at age 16. In this study, then, the developmental history of delinquency was correlated with selection into treatment and also modified the existence and magnitude of the treatment effect of employment on delinquency.

A study by Haviland and Nagin (2005) assessed the effect of first-time gang membership at age 14 on violent delinquency. Their trajectory model identified three groups of youth based on their history of violence from ages 11 to 13 – low, declining, and chronic. Consistent with the selection problem, youth in the chronic trajectory were much more likely to join a gang at age 14 compared to their non-chronic peers. Consistent with the mixing problem, the effect of gang involvement exacerbated violent behavior among all three trajectory groups, but did so particularly among the chronic group.

There are many more uses of the group-based trajectory model than estimation of treatment effects than can be fully described here. Many of these are outlined in Jones and Nagin (2007). First, time-invariant regressors can be introduced into a model of group assignment, which involves introducing covariates into the model for π_j. Second, time-varying regressors can be modeled, in addition to age, in order to allow the impact of such covariates to vary by trajectory group. Third, distinct but related (i.e., co-morbid) behaviors can be modeled using the dual trajectory methodology, in order to examine how such behaviors jointly unfold. Fourth, Haviland and colleagues (Haviland and Nagin 2007; Haviland et al. 2007) have demonstrated the utility of the group-based trajectory model in combination with propensity scores to study the impact of non-random treatments on antisocial behavior.

Conclusion

In criminology, the group-based trajectory methodology arose in response to a desire to provide answers to unresolved questions that arose from the criminal career tradition, and to evaluate the empirical validity of taxonomic theories of criminal offending over the life span. It has since evolved into a general analytical approach capable of answering a variety of questions of interest to behavioral scientists. Its use is widespread not only in criminology (Piquero 2008), but in other disciplines such as clinical psychology (Nagin and Odgers 2010).

The strength of the group-based trajectory model is the flexibility gained by approximating unobserved heterogeneity of unknown density using a discrete distribution rather than the traditional approach of assuming a single continuous density, usually the normal or gamma. Yet the model is not without controversy (see Eggleston et al. 2004; Sampson and Laub 2005; Skardhamar 2010). The view of this author is that most of this controversy stems from how the groups produced by the model are used in applied settings, namely, the tendency to label and reify the groups rather than treat them as the imperfect approximations that they are.

One particularly promising use of the group-based trajectory model is as a way to control flexibly for age in panel models of criminal behavior, a recent application of which is provided by Ezell and Cohen (2005). They showed that valid inferences about the impact of time-varying regressors on criminal behavior depend crucially on how one controls for age-graded developmental change. Models which imposed a uniform age distribution for their sample resulted in substantial overestimates of the impact of time-varying regressors. Their solution was two-pronged. In the first step, they estimated a group-based trajectory model. In the second step, they entered the posterior probabilities into a panel model (except one, for the group which served as a contrast), and interacted each posterior probability with age (as well as age squared and age cubed).

There is no doubt that the group-based trajectory model is an indispensible tool in the panel data expert's toolbox. At a minimum, Heckman and Singer (1984) advocated use of non-parametric estimators – the group-based trajectory model is one such estimator – to evaluate the plausibility of estimates from more conventional, parametric specifications of the distribution of unobservables. It is difficult to find fault with such advice. Panel researchers who are serious about causal models of criminal behavior would be well advised to follow it.

Related Entries

- ▶ Age-Crime Curve
- ▶ Career Criminals and Criminological Theory
- ▶ Criminal Careers
- ▶ Desistance from Crime
- ▶ Group-Based Trajectory Models
- ▶ Moffitt's Developmental Taxonomy of Antisocial Behavior
- ▶ Pathways to Delinquency

Recommended Reading and References

Apel R, Bushway S, Brame R, Haviland AM, Nagin DS, Paternoster R (2007) Unpacking the relationship between adolescent employment and antisocial behavior: a matched samples comparison. Criminology 45:67–97

Blumstein A, Cohen J, Roth JA, Visher CA (eds) (1986) Criminal careers and "career criminals". National Academies Press, Washington, DC

Brame R, Nagin DS, Wasserman L (2006) Exploring some analytical characteristics of finite mixture models. J Q Criminol 22:31–59

D'Unger AV, Land KC, McCall PL, Nagin DS (1998) How many latent classes of delinquent/criminal careers? Results from mixed poisson regression analyses. Am J Sociol 103:1593–1630

Eggleston EP, Laub JH, Sampson RJ (2004) Methodological sensitivities to latent class analysis of long-term criminal trajectories. J Q Criminol 20:1–26

Ezell ME, Cohen LE (2005) Desisting from crime: continuity and change in long-term crime patterns of serious chronic offenders. Oxford University Press, New York

Haviland AM, Nagin DS (2005) Causal inference with group based trajectory models. Psychometrika 70:1–22

Haviland AM, Nagin DS (2007) Using group-based trajectory modeling in conjunction with propensity scores to improve balance. J Exp Criminol 3:65–82

Haviland AM, Nagin DS, Rosenbaum PR (2007) Combining propensity score matching and group-based trajectory analysis in an observational study. Psychol Methods 12:247–267

Heckman J, Singer B (1984) A method for minimizing the impact of distributional assumptions in econometric models for duration data. Econometrica 52:271–320

Jones BL, Nagin DS (2007) Advances in group-based trajectory modeling and an SAS procedure for estimating them. Sociol Methods Res 35:542–571

Jones BL, Nagin DS, Roeder K (2001) A SAS procedure based on mixture models for estimating developmental trajectories. Sociol Methods Res 29:374–393

Land KC, Nagin DS (1996) Micro-models of criminal careers: a synthesis of the criminal careers and life course approaches via semiparametric mixed Poisson regression models, with empirical models. J Q Criminol 12:163–191

Land KC, McCall PL, Nagin DS (1996) A comparison of poisson, negative binomial, and semiparametric mixed poisson regression models: with empirical application to criminal careers data. Sociol Methods Res 24:387–442

Loughran T, Nagin DS (2006) Finite sample effects in group-based trajectory models. Sociol Methods Res 35:250–278

Manski CF (1995) Identification problems in the social sciences. Harvard University Press, Cambridge, MA

McLachlan G, Peel D (2000) Finite mixture models. Wiley, New York

Moffitt TE (1993) Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. Psychol Rev 100:674–701

Nagin DS (1999) Analyzing developmental trajectories: a semiparametric, group-based approach. Psychol Methods 4:139–157

Nagin DS (2005) Group-based modeling of development. Harvard University Press, Cambridge, MA

Nagin DS, Farrington DP, Moffitt TE (1995) Life-course trajectories of different types of offenders. Criminology 33:111–139

Nagin DS, Land KC (1993) Age, criminal careers, and population heterogeneity: specification and estimation of a nonparametric, mixed Poisson model. Criminology 31:327–362

Nagin DS, Odgers CL (2010) Group-based trajectory modeling in clinical research. Annu Rev Clin Psychol 6:109–138

Piquero AR (2008) Taking stock of developmental trajectories of criminal activity over the life course. In: Liberman AM (ed) The long view of crime: a synthesis of longitudinal research. Springer, New York, pp 23–78

Sampson RJ, Laub JH (2005) Seductions of method: rejoinder to nagin and tremblay's "developmental trajectory groups: fact or fiction?". Criminology 43:905–913

Skardhamar T (2010) Distinguishing facts from artifacts in group-based modeling. Criminology 48:295–320

Growth Curve Models with Categorical Outcomes

Katherine E. Masyn[1], Hanno Petras[2] and Weiwei Liu[3]
[1]Harvard Graduate School of Education, Cambridge, MA, USA
[2]Research and Development, JBS International, North Bethesda, MD, USA
[3]NORC at the University of Chicago, Bethesda, MD, USA

Overview

Motivated by the limited available literature on the treatment of longitudinal binary and ordinal outcomes in a growth modeling framework, the goal of this entry is to provide an accessible and practical introduction of this topic for a criminological audience. The parameterization of categorical latent growth models is explained by integrating aspects of the more familiar conventional latent growth models and generalized linear models. Emphasis is placed on the process of model building, evaluation, and interpretation. The entry contains an elaboration of how to include predictors of developmental change in the model for covariate-related hypothesis tests along with remarks regarding of the importance of auxiliary information for assessing model validity and utility. Finally, several model extensions including nonlinear change, generalized growth mixture modeling, and longitudinal latent class analysis are discussed.

Introduction

Criminologists typically encounter data on crime and deviance which is skewed and discrete, thus violating the assumptions of ordinary least square (OLS) regression models, which require that the outcome variable is continuous and is (conditionally) normally distributed. Many criminological studies involve binary outcomes, such as arrest versus no arrest, or unordered categorical outcomes, such as judge and jury consensus or disagreement on conviction versus acquittal. Other outcomes consist of categories, which represent a natural ranking or ordering, such as offense severity or self-reported attitudinal items measured on a Likert scale ranging from strongly disagree to strongly agree. Finally, some outcomes in criminological inquiries consist of counts of a particular event, such as the number of police contacts or the number of arrests. Generalized linear models (GLM), originally formulated by Nelder and Wedderburn (1972), represent a flexible generalization of OLS regression to accommodate skewed and discrete outcomes in a regression framework. In addition to classic textbooks (e.g., Agresti 2002; Long 1997), ample guidance exists on direct applications of GLMs for cross-sectional binary, unordered, and ordered categorical outcomes (e.g., Britt and Weisburd 2010) as well as for count outcomes (e.g., MacDonald and Lattimore 2010) in criminological research.

Beyond the explanation and prediction of cross-sectional outcomes, describing and predicting the developmental course of individuals' involvement in criminal and antisocial behavior is a central theme in criminological inquiries. It is well-known that only repeated observations of individuals across time allow for explicit modeling of intra individual change processes and enable the charting of interindividual differences in intra individual age-crime curves including the manifest features of onset, continuation, and cessation in criminal activity (Piquero et al. 2007). In particular, longitudinal data permits proper inferences about stability and change in individual trajectories, differences across individuals with respect to their trajectories, and the predictive effects of time-invariant risk and vulnerability factors as well as time-dependent life events on those trajectories (Piquero 2008). Given the ever-growing interest in not only describing but testing hypotheses related to individual differences in criminal behavior across the life course, more researchers have endeavored to collect longitudinal data on samples of individuals and are making use of statistical models that effectively and appropriately utilize those repeated measures data. In the last two decades, latent growth modeling (LGM; also known as growth curve modeling, latent trajectory analysis, hierarchical linear modeling, linear mixed models, etc.) has emerged as the preferred analytical choice. This preference is in part due to the fact that LGM is more flexible than repeated measures analysis of variance or observed change score analysis in dealing with missing data, unequally spaced time points, complex nonlinear developments, and, importantly, non-normally distributed and discretely scaled repeated measures (Curran et al. 2010).

Along with their increasing popularity, vast resources have accumulated to instruct researchers in the application of LGMs with longitudinal continuous (e.g., Bollen and Curran 2006; Muthén 2004; Petras and Masyn 2010) and count outcomes (Kreuter and Muthén 2008; Nagin and Land 1993). With the exception of Feldman et al. (2009); Mehta et al. (2004); Muthén (1996); and Skrondal and Rabe-Hesketh (2004), the discussion regarding the treatment of binary and ordinal outcomes in a growth modeling framework is, by comparison, quite limited, especially for the applied criminology audience. Thus, the goal of this entry is to provide an accessible and practical introduction to the study of change using binary and ordinal repeated measures. The remainder of the entry is organized as follows: First, the conventional growth model for continuous outcomes is briefly introduced followed by a presentation of generalized linear models with binary and ordinal outcomes. Then the process of modeling repeated binary and ordinal measures in a latent growth modeling framework is advanced. This entry is concluded with a summative discussion including an overview of model extensions and alternatives.

Conventional Latent Growth Models

Modern growth models generally treat longitudinal outcomes in one of two ways: (1) as multilevel outcome data, where time or measurement occasions at "Level 1" are nested within persons at "Level 2"; or (2) as multivariate outcome data, where the repeated outcome measures are multiple indicators for latent growth factors values for each individual. Taking a multivariate approach, intra individual change is captured by the measurement model for the growth factors, describing the relationship between individual growth factor values and the observed outcomes over time, and interindividual differences are captured by the structural model, i.e., the mean and variance-covariance structure of the growth factors, describing the distribution of the growth factors in the population of individuals.

Although it is possible to specify analytically equivalent unconditional and conditional growth models across the multilevel and (multivariate) latent growth modeling frameworks, estimated via full-information maximum likelihood (FIML), utilizing the latent variable approach affords access to a variety of modeling extensions not as easily implemented in other frameworks, e.g., models that simultaneously include both antecedents and consequences of the change process, higher-order growth models with multiple indicators of the outcome at each assessment, multiprocess and multilevel growth models, and models that employ both continuous and categorical latent variables for describing population heterogeneity in the change process (for more on growth modeling in a latent variable framework, see, e.g., Bollen and Curran 2006; Muthén 2001, 2004). Given this greater flexibility, it is the multivariate approach to longitudinal data in a latent variable modeling framework that we focus on herein.

The latent growth model specification is a restricted form of a more general structural equation model (SEM; Kline 2010). In the SEM formulation of a latent growth model, there are T repeated measures, $y_t(t = 1, \ldots, T)$, that serve as the indicators or manifest variables, where T is the number of time points or waves during which study participants were assessed. For a *linear* latent growth curve model, there are two latent factors: an intercept growth factor, η_0, and a slope growth factor, η_1. The measurement and

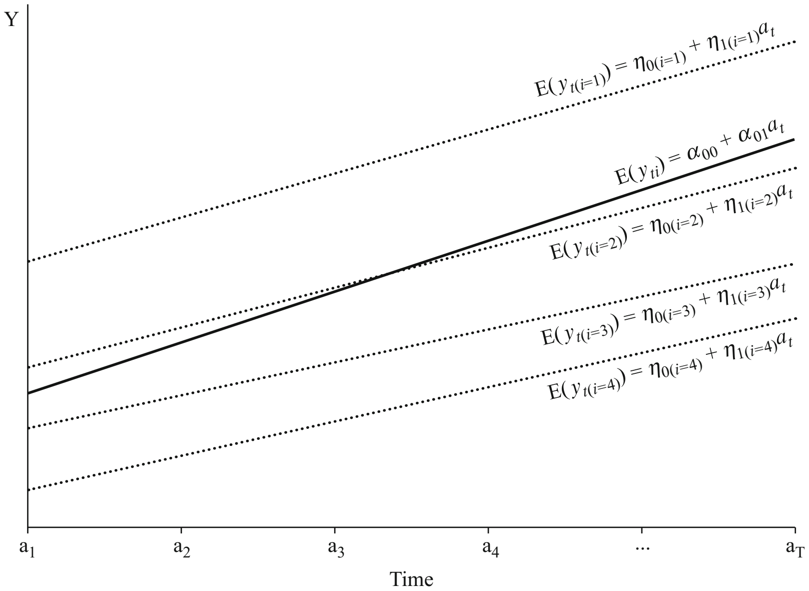

Growth Curve Models with Categorical Outcomes, Fig. 1 Individual trajectories (*dotted lines*) for a hypothetical random sample of four individuals ($i = 1, 2, 3, 4$) drawn from a population with a mean growth trajectory given by the *solid line*

structural portions for an unconditional linear latent growth model are given by

$$\begin{aligned}
&\text{Measurement model:}\\
&\qquad y_{ti} = \eta_{0i} + \eta_{1i}a_t + \varepsilon_{ti},\\
&\text{Structural model:} \qquad\qquad\qquad\qquad (1)\\
&\qquad \eta_{0i} = \alpha_{00} + \xi_{0i},\\
&\qquad \eta_{1i} = \alpha_{10} + \xi_{1i}.
\end{aligned}$$

Here, y_{ti} is the observed outcome y for individual $i(i = 1, \ldots, n)$ at time $t(t = 1, \ldots, T)$, a_t is the time score at time t, η_{0i} is the random intercept factor (i.e., the expected outcome on y for individual i at time score $a_t = 0$), and η_{1i} is the random linear slope factor (i.e., the change in the expected outcome on y for individual i for a one unit increase in time, on the scale of a_t). The values for a_t are fixed to define the slope factor as the linear rate of change in y on the observed time metric; for example, in a panel study for which participants were assessed annually for T years, we might use $\mathbf{a} = (0,1,2, \ldots, T - 1)'$ so that one unit on the time metric defined by $\mathbf{a}$ is one year. Typically, the first time score, a_1, is fixed at zero so that the intercept factor can be interpreted as the expected response at the first time of measurement. The ε_{ti}s represent independent and identically distributed measurement and time-specific errors on the y_{ti}s at time t, and the ε_{ti}s are usually assumed to be uncorrelated across time. In the structural model, α_{00} is the population mean of the individual intercept factor values, α_{10} is the population mean of the individual slope factor values, ξ_{0i} is the deviation of η_{0i} from the population mean intercept, α_{00}, and ξ_{1i} is the deviation of η_{1i} from the population mean slope, α_{10}. The distribution of individual intercept factor values and slope factors values is assumed to be multivariate normal, as is the distribution of the ε_{ti}s; the growth factors are assumed to be uncorrelated with the errors. Figure 1 displays the expected individual trajectories (dotted lines) of a hypothetical random sample of four individuals drawn from a population with an overall mean growth trajectory given by the solid line. Intra individual change is represented by each of the individual-specific trajectories (each with person-specific intercept and slope values), and interindividual differences are represented by the variability in individual-specific intercept and slope values relative to the overall mean intercept and slope values.

Extending now to the conditional latent growth model, hypothesized predictors of the interindividual differences can be included in both the measurement model (for time-varying predictors and time-invariant predictors with unrestricted time-varying effects) and the structural model (for time-invariant predictors of the intercept and slope factors) as given by

Measurement model:

$$y_{ti} = \eta_{0i} + \eta_{1i}a_t + \pi_{1t}w_{ti} + \varepsilon_{ti},$$

Structural model: (2)

$$\eta_{0i} = \alpha_{00} + \alpha_{01}x_i + \xi_{0i},$$
$$\eta_{1i} = \alpha_{10} + \alpha_{11}x_i + \xi_{1i},$$

where π_{1t} is the difference in the expected outcome, y_t, corresponding to a one unit difference in the time-varying covariate, w, specifically at time t; α_{01} is the difference in the mean of the intercept factor corresponding to the one unit difference in the time-invariant covariate, x; and α_{11} is the difference in the mean of the slope factor corresponding to the one unit difference in the time-invariant covariate, x. The conditional growth model with a predictor for the growth factors is depicted in path diagram form in Fig. 2 using the following diagramming conventions: latent variables are represented by circles, observed variables are represented by rectangles, linear directional relationships are represented by single arrow paths, correlational relationships are represented by double arrow paths, and error terms are represented by unanchored single arrow paths.

Now that we have provided a brief overview of conventional latent growth modeling in a latent variable framework, we next introduce the foundations of binary and ordinal logistic regression followed by discussion of how repeated measures of a categorical outcome can be analyzed in this same latent variable framework.

Generalized Linear Models for Binary and Ordinal Outcomes

As mentioned in the introduction, categorical and limited dependent variables are quite common in criminology research. These types of outcome variables, be they cross-sectional or longitudinal, violate most if not all the assumptions of the linear models in standard use for continuous outcomes. Generalized linear models (GLMs) are a family of regression models that extend OLS regression to accommodate noncontinuous outcomes while still working with outcome predictors in a linear regression framework. There are different (often equivalent) approaches for parameterizing and estimating GLMs (Long 1997; Skrondal and Rabe-Hesketh 2004). For the purposes of this entry, we will introduce a somewhat less standard specification, known as the *latent response variable* (LRV) formulation for ordinal outcome variables. Our choice of the LRV formulation is based on the ease with which it enables the extension of the latent growth model specification given in the previous section to include categorical longitudinal outcomes.

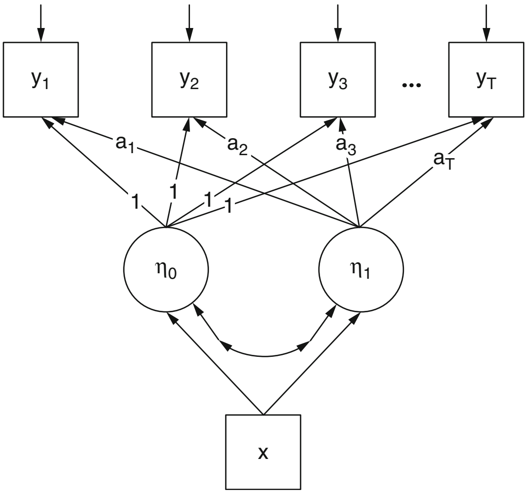

Growth Curve Models with Categorical Outcomes, Fig. 2 Path diagram for a conditional linear latent growth model with continuous outcomes and a time-invariant predictor of the growth factors

In the latent response variable formulation, it is assumed that the observed ordinal outcome, y, is a discretized form of an underlying continuous latent response variable, y^*. For example, consider the binary (0/1) outcome of clinical depression, *depress*. One could imagine an underlying continuum of depression, $depress^*$, such that individuals whose values of $depress^*$

exceeded a certain level, or threshold, would all be observed with binary outcome *depress* = 1. In general, the relationship between a binary outcome, y, and the latent response variable, y^*, is given by

$$y_i = \begin{cases} 1 \text{ if } y_i^* > \tau_1 \\ 0 \text{ if } y_i^* \leq \tau_1, \end{cases} \qquad (3)$$

where τ_1 is the *threshold* for y^* and $\Pr(y = 1) = \Pr(y^* > \tau_1)$. As another example, consider the four-category ordinal outcome, *oppose*, measuring opposition to or disapproval of the death penalty for first degree murder with response categories on a Likert scale: *strongly approve/support* (0), *approve/support* (1), *disapprove/oppose* (2), and *strongly disapprove/oppose* (3). One could imagine an underlying continuum of opposition or disapproval, *oppose**, such that four different ranges of *oppose** defined by three cut points or thresholds map onto the observed values of *oppose*. In general, there are $J - 1$ thresholds that define the relationship between a latent response variable, y^*, and its J-category ordinal form, y, such that

$$\text{y}_\text{i} = \begin{cases} 0 \text{ if -}\infty < y_i^* \leq \tau_1 \\ 1 \;\text{ if } \tau_1 < y_i^* \leq \tau_2 \\ \vdots \\ J-1 \;\text{ if } \tau_{\text{J-1}} < y_i^* \leq \infty \end{cases} , \qquad (4)$$

where τ_j is the j^{th} *threshold* for y^*, delineating responses $j - 1$ and j on the scale of y. In this LRV formulation, $\Pr(y \leq j) = \Pr(y^* \leq \tau_{j+1})$ and $\Pr(y = j) = \Pr(y^* \leq \tau_{j+1})\text{-}\Pr(y^* \leq \tau_j)$, where $j \in \{0,1,\ldots,J\text{-}1\}$, $\tau_0 = -\infty$, $\tau_J = \infty$, and $\tau_0 < \tau_1 < \ldots < \tau_{J-1} < \tau_J$. The continuous latent response variable, y^*, is then expressed as the sum of a mean and error term given by

$$y_i^* = \mu_i + \delta_i. \qquad (5)$$

The distribution and scale of the error, δ, must be specified a priori by the analyst; the two most common distributions for δ are the standard normal distribution and the standard logistic distribution. Figure 3 provides a visual representation of the latent response variable formulation for an ordinal outcome with four response categories. The bottom portion of Fig. 3 depicts a standard logistic distribution for y^* with three corresponding thresholds, (τ_1, τ_2, τ_3). All individuals in the population with y^* values between τ_1 and τ_2 will manifest the response value $y = 1$, and the shaded area under the probability density curve for y^* between τ_1 and τ_2 is equal to the probability of the $y = 1$ outcome, as depicted in top portion of Fig. 3. Figure 3 also displays the path diagram representation of the relationship between y^* and y. The three solid squares at the point where the path from y^* meets y indicate the deterministic relationship specified between given values of y^* and resultant observed values on y defined by the three thresholds, (τ_1, τ_2, τ_3).

The conditional model for ordinal outcomes is specified so that the observed predictors of the categorical outcome are related to cumulative response probabilities via the latent response variables as follows:

$$\begin{aligned} y_i^* &= \mu_{i|x_i} + \delta_i, \\ \mu_{i|x_i} &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki}. \end{aligned} \qquad (6)$$

Figure 4 displays a plot of linear regression model of latent response variable, y^*, underlying a four-category ordinal outcome, versus a single predictor, x. As with a standard linear regression model, the expected values of y^* given x falls along the line, $\beta_0 + \beta_1 x$, and the distributions of y^* values in the population at each given value of x are the same in shape and variance (as shown in Fig. 4 by the probability density curves at the x-values, $x = x_1$, $x = x_2$, and $x = x_3$). For model identification, the intercept, β_0, is suppressed, i.e., fixed at zero. The linear model implies that the difference in the expected value of y^*corresponding to a one unit difference in x is equal to β_1 across the entire range of x. The conditional model given in Eq. 6 also makes the assumption of *threshold invariance*, that is, assuming that the thresholds defining the

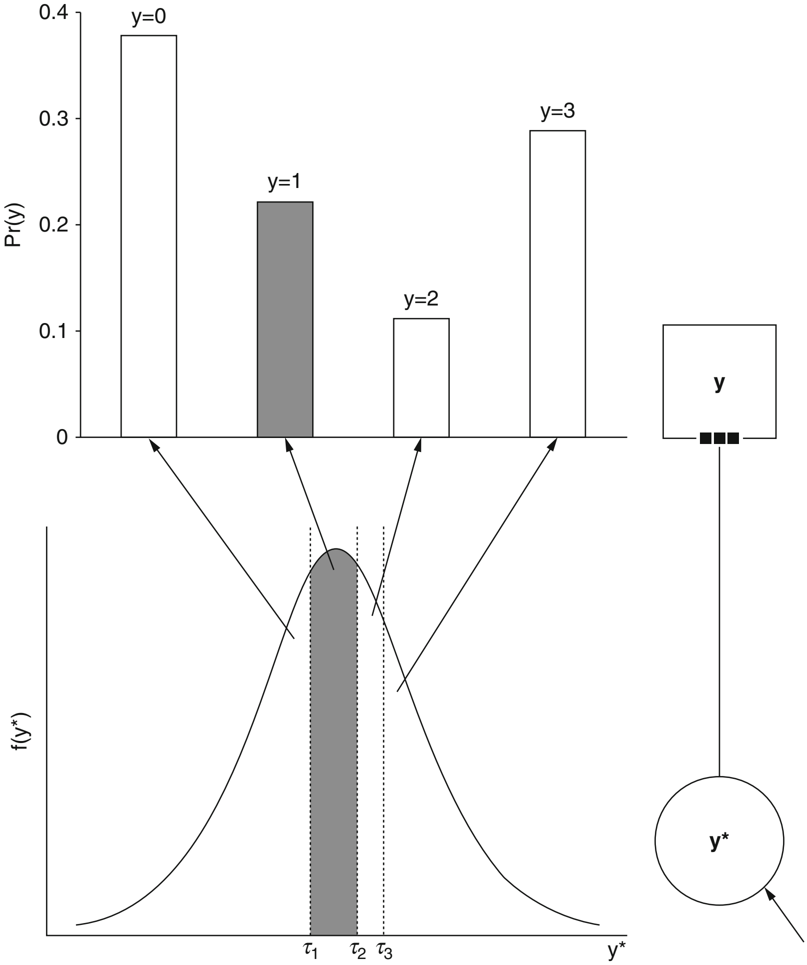

Growth Curve Models with Categorical Outcomes, Fig. 3 Graphical and path diagram representations of the latent response variable formulation for an observed ordinal outcome with four response categories

relationship between y^* and y do *not* depend on (i.e., are invariant) relative to x. To assist in visualizing how the linear functional relationship between y^* and x and the assumption of threshold invariance translate to the relationship between observed response probabilities of y and x, the three thresholds for y^*, which do *not* depend on x, are drawn as horizontal lines in Fig. 4. The area under the probability density function for y^* between τ_1 and τ_2, corresponding to $\Pr(y = 2|x)$, at $x = x_1$, $x = x_2$, and $x = x_3$, is shaded making it easy to see that even though the mean of y^* shifts linearly across the range of x, the response category probabilities do *not* change linearly or even monotonically.

Using a standard logistic distribution for δ, the linear regression for y^* on a single predictor (easily replaced with a linear combination of multiple predictors as given in Eq. 6) and the assumption of threshold invariance translates to the following relationship between the response category probabilities and the predictor:

$$\Pr(y_i \le j|x_i) = \frac{1}{1 + \exp\left(-\tau_{j+1} + \beta_1 x_i\right)}, \tag{7}$$

or, equivalently,

$$\log\left(\frac{\Pr(y_i \le j|x_i)}{\Pr(y_i > j|x_i)}\right) = \tau_{j+1} - \beta_1 x_i. \tag{8}$$

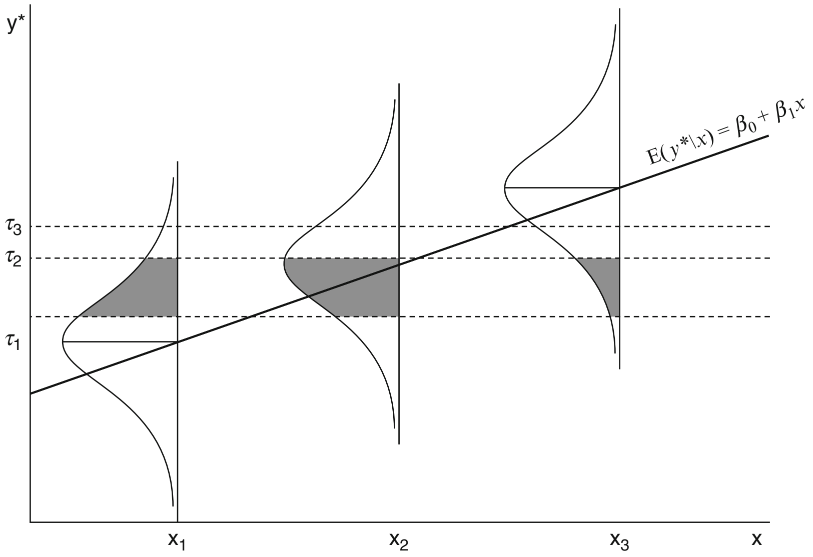

Growth Curve Models with Categorical Outcomes, Fig. 4 Plot of the linear regression line for a latent response variable, y^*, underlying an observed four-category ordinal outcome variable, y, versus a single predictor, x. Probability density curves depict the distribution of y^* values at three different values of x. Dashed horizontal lines show the threshold values for y^* delineating the four value ranges mapping on to the four response categories of y. Shaded areas of the probability density show the changing Pr ($y = 2|x$) across the three values of x

For a binary variable (with $J = 2$ categories), Eq. 8 reduces to the familiar logistic regression equation:

$$\log\left(\frac{\Pr(y_i = 1|x_i)}{\Pr(y_i = 0|x_i)}\right) = -\tau_1 + \beta_1 x_i. \qquad (9)$$

In Fig. 5 we can see the representation of the relationships between the cumulative response category log odds and the predictor as described in Eq. 8 – that cumulative log odds for the different response categories all vary linearly as a function of x and that the lines for each of the three cumulative log odds are *parallel* with the distance between them determined by the differences in the threshold values.

The functional form of each of the three lines in Fig. 5 and the fact that they are all linear is a direct consequence of linearity assumption for the relationship between y^* and x. Assuming a linear relationship between y^* and x constrains the relationship between y and x such that cumulative response category log odds differ identically and linearly for every one unit difference in x. Consequently, β_1 is interpreted not just as the difference in the expected value of y^* corresponding to a positive difference of one unit on x but also as the log odds ratio for responding at or below a given response category (rather than above) corresponding to a *negative* difference of one unit on x, i.e., the odds for a response less than or equal to category j differ by a factor of exp $(-\beta_1)$ for every one additional unit on x, for all $j = 0$, 1, or 2. Thus, the cumulative odds for each response category differ identically and proportionally for every one unit difference in x. The linearity assumption is also known as the *proportional odds* assumption.

The equidistance between each pair of lines in Fig. 5 across the full range of x is a direct consequence of assuming that the thresholds defining the relationship between y^* and y do not depend on x. Thus, the differences in the cumulative log odds across the response categories (i.e., the vertical distance between the lines) at any given value of x are constant across all values of x. The assumption of *threshold invariance*, meaning that the thresholds do not depend on x, is also known as the *parallel regression* assumption. Both the proportional odds assumption and the

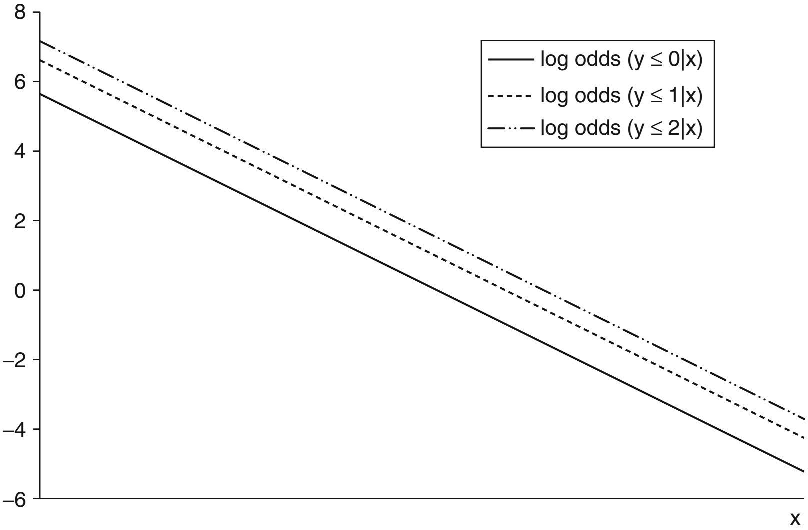

Growth Curve Models with Categorical Outcomes, Fig. 5 Plot of the cumulative response log odds for *y* versus *x* based on a linear regression model of y^* on *x* with threshold invariance

parallel regression assumption can be relaxed (and tested), and this will be mentioned again in the growth modeling context in the next section.

Building upon the overviews of both conventional latent growth models and cross-sectional latent response variable models for binary and ordinal dependent variables, the intersection of these two modeling approaches that enables the study of change in binary and ordinal longitudinal outcomes is introduced next.

Latent Growth Models for Binary and Ordinal Outcomes

The specification for the latent growth model for binary or order categorical outcomes is built on the same latent response variable formulation used with cross-sectional data. It is assumed that there is a continuous latent response variable, $y_{ti}{}^*$, underlying the observed response on the J-category ordinal outcome, y_{ti}, for individual $i(i = 1, \ldots, n)$ at time $t(t = 1, \ldots, T)$ with the relationship between $y_{ti}{}^*$ and y_{ti} given by

$$y_{ti} = \begin{cases} 0 \text{ if } \; -\infty < y^*_{ti} \le \tau_{1t} \\ 1 \text{ if } \; \tau_{1t} < y^*_{ti} \le \tau_{2t} \\ \vdots \\ J-1 \;\text{ if } \tau_{(j-1)t} < y^*_{ti} \le \infty \end{cases}, \quad (10)$$

where τ_{jt} is the j^{th} *threshold* for $y_t{}^*$, delineating responses $j - 1$ and j on the scale of y_t. Usually, in the model building taxonomy, it is common to begin with models that make the assumption of *longitudinal threshold invariance*, meaning that the set of $J - 1$ thresholds, $(\tau_{1t}, \ldots, \tau_{(J-1)t})$, are the same at each wave; i.e., $\tau_{jt} = \tau_j, \forall t$. The longitudinal threshold invariance assumption can then be evaluated by testing the improvement in model fit when threshold invariance is relaxed.

The continuous latent response variable, $y_{ti}{}^*$, at each wave, t, is expressed as the sum of a mean and error term given by

$$y^*_{ti} = \mu_{ti} + \delta_{ti}. \quad (11)$$

As before, the distribution and scale for the error at each wave, δ_{ti}, must be specified a priori by the analyst – for the purposes of this entry, the standard logistic distribution for the error terms will be used. A latent growth model is then specified for the individual μ_{ti} values in a similar way as for observed continuous repeated measures. A linear latent growth model expresses the expected value on the latent response variable for individual i at time t as a function of the intercept and growth factors values; that is,

$$\mu_{ti} = \eta_{0i} + \eta_{1i} a_{\,t}, \quad (12)$$

or, equivalently, combining Eqs. 11 and 12,

$$y_{ti}^{*} = \eta_{0i} + \eta_{1i}a_t + \delta_{ti}. \qquad (13)$$

The structural portion of the latent growth model is identical to the specification for observed continuous repeated measures; that is,

$$\begin{aligned}\eta_{0i} &= \alpha_{00} + \xi_{0i},\\ \eta_{1i} &= \alpha_{10} + \xi_{1i},\end{aligned} \qquad (14)$$

where $\xi_{\bullet i}$ s are specified to have a multivariate normal distribution. Since the location and scale of the latent response variable is indeterminate, an additional restriction must be placed, fixing $\alpha_{00} = 0$, similar to fixing $\beta_0 = 0$ for y^* in the cross-sectional model. The path diagram representation of an unconditional linear latent growth model for observed ordinal outcomes is depicted in Fig. 6. Although the above specification expresses the change in the latent response variable over time as a linear function of the time metric, the same approaches can be used as with observed continuous outcomes to investigate interindividual differences in curvilinear and other forms of nonlinear trajectories of change.

To gain a better understanding of what specifying a linear latent growth model for the latent response variable with longitudinal threshold invariance implies with respect to interindividual difference in intra individual change in the observed ordinal outcome, consider Fig. 7 which displays expected individual growth trajectories (dotted lines) on the latent response variable underlying an observed four-category ordinal variable for three hypothetical individuals, $i = 1$, 2, and 3, and the population overall mean growth trajectory (solid line). The time-invariant thresholds are shown by the horizontal dashed lines. Each individual in the population has their own expected y^* trajectory as given by Eq. 12. Each individual's cumulative response probabilities at each point in time are determined by the thresholds and the distribution of y_{ti}^{*} centered at μ_{ti}, as depicted by the separated density curves for each of the three hypothetical individuals at time points a_2, a_3, and a_4 shown in Fig. 7. Assuming that δ_{ti} are independently and identically distributed standard logistic, the relationship between the observed ordinal response and the growth factors is given by

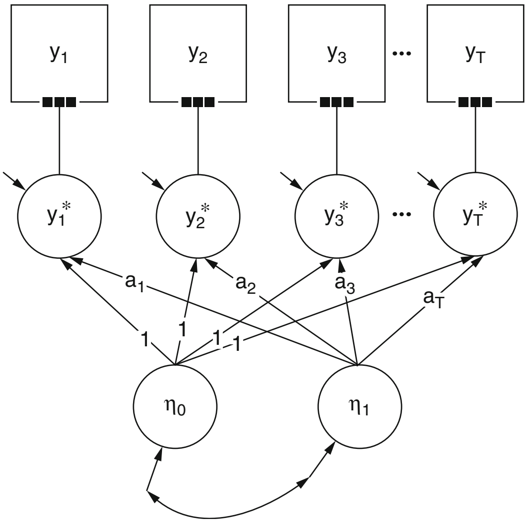

Growth Curve Models with Categorical Outcomes, Fig. 6 Path diagram for an unconditional linear latent growth model for the latent response variables underlying a set of observed longitudinal ordinal outcomes

$$\log\left(\frac{\Pr(y_{ti} \le j)}{\Pr(y_{ti} > j)}\right) = \tau_j - (\eta_{0i} + \eta_{1i}a_t). \qquad (15)$$

Thus, $\tau_j - \eta_{0i}$ is the cumulative log odds for response category j for individual i when $a_t = 0$, and $-\,\eta_{1i}$ is the change in the cumulative log odds (or cumulative log odds ratio) for response category j for individual i corresponding to a one unit increase in the time metric of a_t.

It can be seen from Eq. 15 that the longitudinal threshold invariance assumption (i.e., $\tau_{jt} = \tau_j, \forall t$) implies that the difference in cumulative log odds for any two response categories will be the same at any given fixed point in time, a_t, across the entire time span for a given individual i. Assuming a linear function for the intra individual change process in the latent response variable under the threshold invariance assumption implies that the change in the cumulative log odds for a given individual and for a given response category corresponding to a one unit difference in the time metric is the same across the entire time span and across all response categories. As mentioned

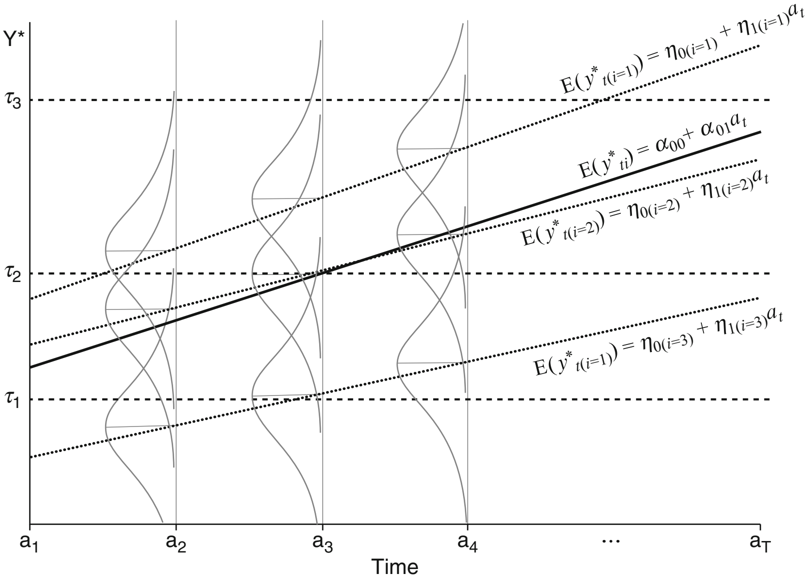

Growth Curve Models with Categorical Outcomes, Fig. 7 Individual trajectories (*dotted lines*) of latent responses to repeated measure of an observed four-category ordinal outcome variable for a hypothetical random sample of three individuals ($i = 1, 2, 3$) drawn from a population with a mean growth trajectory for y^* given by the *solid line* with *dashed horizontal lines* show the time-invariant threshold values for y^* delineating the four value ranges mapping on to the four response categories of y

before, both the linearity and threshold invariance assumptions can be relaxed and tested. However, in order for the structural portion of the latent growth model to be identified, *at least one threshold* must be held invariant across time. For a binary longitudinal outcome with only one threshold at each time point, complete threshold invariance must be imposed.

It is important at this point to make the reader aware of one critical difference between the latent growth model for observed continuous outcomes and the LGM for observed ordinal outcomes. For continuous outcomes, the mean growth trajectory for the population is just a linear function of the population mean intercept factor and population mean slope factor. This is only the case at the latent response variable level of the categorical LGM – it is *not* the case for the mean population responses for the observed ordinal variable across time. In other words, it is incorrect to just plug the means of η_0 and η_1 into Eq. 15 to get the population mean cumulative response probabilities because the average of the individual cumulative response probabilities is not the same as the cumulative response probability corresponding to the mean growth trajectory for the latent response variable. That is,

$$E[\Pr(y_{ti} \leq j)] \neq \frac{1}{1 + \exp\left(-\tau_j + (\alpha_{00} + \alpha_{10}a_t)\right)}. \tag{16}$$

To compute the mean cumulative response probability for the population at a given point in time requires calculating and then averaging the individual cumulative response probabilities for all members of the population, that is,

$$E[\Pr(y_{ti} \leq j)] = \lim_{n\to\infty}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{1+\exp\left(-\tau_j+(\eta_{0i}+\eta_{1i}a_t)\right)}\right)\right]. \tag{17}$$

The above expression can also be written as a double integral equation, integrating over both η_0 and η_1. Obtaining model-estimated mean cumulative response probabilities for a specific response category at a specific measurement occasion requires numeric integration based on the model-estimated distribution of the latent growth factors which is a post-estimation option in some modeling software, such as Mplus V6.12 (Muthén and Muthén 1998–2011a).

It is straightforward to extend the unconditional ordinal LGM to include time-invariant and time-varying predictors, just as is done with observed continuous repeated measures. Time-invariant predictors can be included as predictors of the growth factors, or direct predictors of the expected latent response variable across time and time-varying predictors can be included as direct predictors of the expected latent response variable across time, as given by

$$\begin{aligned} y_{ti}^{*} &= \mu_{ti} + \delta_{ti}, \\ \mu_{ti} &= \eta_{0i} + \eta_{1i}a_t + \pi_{1t}w_{ti}, \\ \eta_{0\mathrm{i}} &= \alpha_{00} + \alpha_{01}x_i + \xi_{0i}, \\ \eta_{1i} &= \alpha_{10} + \alpha_{11}x_i + \xi_{1i}, \end{aligned} \tag{18}$$

where δ_{ti} are assumed to be *i.i.d.* standard logistic and α_{00} is fixed at zero for identification.

The unconditional and conditional ordinal latent growth models presented in this section can all be estimated using maximum likelihood estimation. These growth models can be more computationally intensive than the cross-sectional models for observed ordinal outcomes as maximizing the likelihood function with continuous latent predictors (i.e., the latent growth factors) for the ordinal responses requires numeric integration with one dimension of integration for each growth factor. Full-information maximum likelihood (FIML) is available for these models which allows the inclusion of cases with incomplete data on the longitudinal outcomes under the missing-at-random (MAR) assumption (for more about missing data, see, e.g., Enders 2010).

Discussion

This entry has discussed the application of latent growth curve modeling to categorical outcomes by integrating aspects of the conventional latent growth model with the generalized linear model. The unconditional growth model with respect to parameterization and estimation was presented, and concise guidelines were provided for how to conduct the model building and evaluation process. It was then elaborated on how to include covariates to evaluate hypotheses about the relationship between predictors and the developmental change process.

There are several interesting augmentations of and alternatives to these models that are currently available to applied researchers. Although in Eq. 18 the change in the expected latent response outcome was expressed as a linear function of the time metric, it is possible (with an adequate number of repeated observations on each subject) to investigate interindividual differences in curvilinear and other forms of nonlinear trajectories of change. The two most common approaches are (1) to freely estimate $T - 2$ of the time scores loadings for η_1 (fixing one loading at zero to define the intercept location and fixing one loading at unity for identification and to set the slope factor metric) or (2) to use additional growth factors (beyond the intercept and slope factors) to accommodate curvilinear polynomial functions of times (e.g., adding a third, quadratic growth factor with loadings fixed at the values ${a_t}^2$). Alternative specifications of time can also be easily accommodated, including piece-wise linear growth models as well as exponential and sinusoidal models of change (see, e.g., Blozis et al. 2007; Bollen and Curran 2006).

In this entry, the application of a latent growth model to repeatedly measured categorical outcomes was discussed. In this model intra-individual change over time is estimated by person-specific growth parameters (e.g., intercept and slope), and interindividual differences are modeled by allowing for individual variation around the estimated growth factor means. Notably, this model is somewhat restrictive in that it assumes population homogeneity in the growth trajectories, i.e., that the growth factors for all persons in the sample are identically distributed. However, etiological as well as prevention and intervention evidence exists to suggest that criminal and antisocial behavior in the overall population may be better represented by a mixing of unobserved heterogeneous subgroups of individuals characterized by differently distributed developmental trajectories, differential risk factors, and differential responses to behavioral and policy interventions. Fortunately, it is relatively straightforward to extend the categorical latent growth model to a generalized linear growth *mixture* model (Feldman et al. 2009). As is the case for continuous outcomes, different latent trajectory classes can be characterized by class-varying mean and variance structures for the growth factors. For example, the developmental change process for one latent subgroup may be perfectly captured by an intercept and a linear slope, while for another subgroup, an additional nonlinear slope is needed.

In addition to modeling nonlinear change and developmental heterogeneity, it is possible that developmental differences cannot be captured by continuous or discrete individual variability around a structured function of time (i.e., intercept, linear, and nonlinear slope). Longitudinal latent class analysis (LLCA: Vermunt et al. 2008; also known as repeated measure latent class analysis, RMLCA: Collins and Lanza 2010) can be used in such situations. LLCA is the application of latent class analysis to repeated outcomes, and unobserved heterogeneity in the developmental response profiles is captured exclusively by a categorical latent variable. In comparison to a growth model which models time-scaled change, LLCA models the longitudinal patterns of discrete states (see, e.g., Feldman et al. 2009; Liu et al. 2010; Liu et al., in press).

In the above-discussed three modeling extensions (nonlinear change, generalized linear growth mixture model, and longitudinal latent class analysis), the inclusion of antecedents and distal outcomes plays an important role. The use of such auxiliary information, potentially derived from substantive theory, is highly relevant to determine the concurrent and prognostic validity of specific growth factors and developmental trajectory profiles derived from a particular data set (Kreuter and Muthén 2008; Petras and Masyn 2010). That is to say, the inclusion of auxiliary information in these models is a necessary step in understanding as well as evaluating the fidelity and utility of the resultant trajectory profiles from a given study. In the simplest case, auxiliary information can consist of observed univariate or multivariate variables measuring predictors or distal outcomes. However, the auxiliary information could itself be a latent variable with its own measurement model and can consist of repeated measures which are observed sequentially or concurrently and modeled simultaneously with the change process of the outcome.

Clearly, these models hold great potential for aiding empirical investigations of developmental theories of normative and non-normative behaviors and risky outcomes across the lifespan. In no way is this more evident than in the marked increase in their use among applied researchers in criminology and other behavioral sciences. However, there is still much opportunity in the realm of methods development to capitalize on the potential of these models and extensions to better accommodate the complexities of developmental theories. And, as with any statistical tool, the research question, along with previous theoretical and empirical work, should guide these models' application in a particular study, with thoughtful and purposeful choices for model specification, selection, and interpretation.

Acknowledgment Dr. Weiwei Liu received support through a training grant from the National Institute of Mental Health while working on this entry (T32 MH18834). Correspondence concerning this article should be addressed to Hanno Petras, Ph.D., at JBS International, Inc., 5515 Security Lane, Suite 800, North Bethesda, MD 20852-5007, USA. Phone: (240) 645-4921. Email: hpetras@jbsinternational.com.

Related Entries

- ▶ Crime Specialization, Progression, and Sequencing
- ▶ Desistance From Crime
- ▶ Pathways to Delinquency
- ▶ General Theory of Crime
- ▶ Group-Based Trajectory Models and Developmental Change
- ▶ Group-Based Trajectory Models
- ▶ Interrupted Time Series Models
- ▶ Longitudinal Crime Trends at Places
- ▶ Longitudinal Studies in Criminology

Recommended Reading and References

Agresti A (2002) Categorical data analysis, 2nd edn. Wiley, New York

Blozis SA, Conger KJ, Harring JR (2007) Nonlinear latent curve models for longitudinal data. Int J Behav Dev 31:340–346

Bollen KA, Curran PJ (2006) Latent curve models: a structural equation perspective. Wiley, Hoboken

Britt CL, Weisburd D (2010) Logistic regression models for categorical outcome variables. In: Piquero AR, Weisburd D (eds) Handbook of quantitative criminology. Springer, New York, pp 649–682

Collins LM, Lanza ST (2010) Latent class and latent transition analysis with applications in the social, behavioral, and health sciences. Wiley, Hoboken

Curran PJ, Obeidat K, Losardo D (2010) Twelve frequently asked questions about growth curve modeling. J Cogn Dev 2:121–136

Enders CK (2010) Applied missing data analysis. Guildford Press, New York

Feldman BJ, Masyn KE, Conger RD (2009) New approaches to studying problem behaviors: a comparison of methods for modeling longitudinal, categorical adolescent drinking data. Dev Psychol 3:652–676

Kline RB (2010) Principles and practice of structural equation modeling, 3rd edn. Guilford Press, New York

Kreuter F, Muthén B (2008) Analyzing criminal trajectory profiles: bridging multilevel and group-based approaches using growth mixture modeling. J Quant Criminol 24:1–31

Liu LC, Hedeker D, Segawa E, Flay BR (2010) Evaluation of longitudinal intervention effects: an example of latent growth mixture models for ordinal drug-use outcomes. J Drug Issues 40:27–43

Liu W, Lynne-Landsman S, Petras H, Masyn K, Ialongo N The evaluation of two first grade preventive interventions on childhood aggression and adolescent marijuana use: a latent transition longitudinal mixture model. *Prev Sci* (in press)

Long JS (1997) Regression models for categorical and limited dependent variables. Sage, Thousand Oaks

MacDonald JM, Lattimore PK (2010) Count models in criminology. In: Piquero AR, Weisburd D (eds) Handbook of quantitative criminology. Springer, New York, pp 683–698

Mehta PD, Neale MC, Flay BR (2004) Squeezing interval change from ordinal panel data: latent growth curves with ordinal outcomes. Psychol Methods 9(3):301–333

Muthén BO (1996) Growth modeling with binary responses. In: Eye AV, Clogg C (eds) Categorical variables in developmental research: methods of analysis. Academic, San Diego, pp 37–54

Muthén B (2001) Second-generation structural equation modeling with a combination of categorical and continuous latent variables: new opportunities for latent class/latent growth modeling. In: Collins LM, Sayer A (eds) New methods for the analysis of change. APA, Washington, DC, pp 291–322

Muthén B (2004) Latent variable analysis: growth mixture modeling and related techniques for longitudinal data. In: Kaplan D (ed) Handbook of quantitative methodology for the social sciences. Sage, Newbury Park, pp 345–368

Muthén B, Muthén LK (1998–2011). Mplus (Version 6.12) [Computer software]. Muthén & Muthén, Los Angeles

Muthén LK, Muthén B (1998–2011) Mplus user's guide, 6th edn. Muthén & Muthén, Los Angeles

Nagin DS, Land KC (1993) Age, criminal careers, and population heterogeneity: specification and estimation of a nonparametric, mixed Poisson model. Criminology 31:327–362

Nelder J, Wedderburn R (1972) Generalized linear models. J R Stat Soc 135(3):370–384

Petras H, Masyn K (2010) General growth mixture analysis with antecedents and consequences of change. In: Piquero A, Weisburd D (eds) Handbook of quantitative criminology. Springer, New York, pp 69–100

Piquero AR (2008) Taking stock of developmental trajectories of criminal activity over the life course. In: Liberman AM (ed) The long view of crime – a synthesis of longitudinal research. Springer, New York, pp 23–78

Piquero AR, Farrington DP, Blumstein A (2007) Key issues in criminal career research: new analyses of the Cambridge study in delinquent development. Cambridge University Press, Cambridge

Skrondal A, Rabe-Hesketh S (2004) Generalized latent variable modeling. Multilevel, longitudinal, and structural equation models. Chapman Hall, London

Vermunt JK, Tran B, Magidson J (2008) Latent class models in longitudinal research. In: Menard S (ed) Handbook of longitudinal research: design, measurement, and analysis. Elsevier, Burlington, pp 373–385

GST

▶ Individual Characteristics and General Strain Theory

Guilds

▶ History of Police Unions

Handwriting as Evidence

S. C. Leung
Scientific Consultancy Limited, Hong Kong, China

Synonyms

Handwriting examination: the art and science

Overview

Handwriting is produced by finely coordinated and precisely articulated movements of the hand, finger, and arm, manipulating a tool which leaves a trace on a surface with the ultimate control being responsible – the brain. It is a very complex activity requiring intelligence that distinguishes human being from the rest of the animal species. By way of writing, the author can convey and preserve ideas or information. With the advancement of technology, the writing instrument changes considerably, but handwriting, being the most personal and immediate means of graphic communication, and the movements associated with it remain virtually the same since the creation of the strokes and the alphabets.

Ever since the invention of the "written language," the need for the verification of authorship started to exist. Nowadays, there is a widespread demand for the examination of handwriting, not only in criminal cases and civil litigations but also in historical and archaeological investigations, and handwriting examination is an important specialist subject in forensic sciences. Basic principles in handwriting identification are given in a number of books authored by authorities in the field (Harrison 1966; Osborn 1978; Hilton 1982; Ellen 1993), while forensic scientists continue to conduct research in the scientific examination of handwriting.

Writer Identification

Physiologically the ability to write is a "motor-perception" skill, involving psychological aspects, perception, and muscular coordination on the part of the author who is constantly under the influence of a variety of factors. These comprise both internal (or personal) and external (or circumstantial) influences. Because of the large number of variables affecting the writer, handwriting from one person on different occasions cannot be expected to be identical in each and every aspect of line quality, speed, pen pressure, shading, pen emphasis, slant, arrangement of strokes, proportionality, chronological sequence, and connecting components. This phenomenon of a "nonuniformity" in the handwriting of an individual is commonly referred to as natural variation.

That handwriting varies does not present serious problems to the task of identification. The skill to write is developed and practiced over many years and is largely a subconscious

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

and automatic activity of human beings. Thus, the handwriting of a person, despite not being 100 % precise, conforms to a distinctive and predictable pattern. The verification of handwriting, as to whether or not it was written by a certain individual, is based upon the fact that handwriting embodies various qualities and attributes which in combination are sufficiently unique to be individualized. The joint probability of a compound event comprising a number of mutually independent simple events is the product of the probabilities of the simple events. Hence, if the disputed handwriting contains a variety of personal characteristics that lie within the range of natural variation of the known writing and at the same time not even a single fundamental, or unexplainable, difference from the genuine specimens can be found, then the combined probability of occurrence of this set of characteristics present in the handwriting of another person selected at random from a population will be so small that it can be ignored. The argument is that it would not be reasonable to assume that the same combination of writing attributes, in sufficient numbers and weight, would accidentally occur in the handwriting of another person without showing even a single idiosyncrasy from the control specimens. This hypothesis forms the basis of identification.

Handwriting characteristics are the observable features generated by the movement of the author's hand. The act of writing involves not only two-dimensional movements of the pen but also variations in the pressure of the writing tip on the substrate. Minutiae of the landing movement when the writing tip approaches and finally "touches down" on the surface of the paper generate a large variety of shapes at the start of a stroke. Likewise the rehearsed motion at the conclusion of a stroke provides specific calligraphic attributes. Details of the habitual manners of utilizing the pen to write contribute toward much of the visual appearance which can be generalized by the terminology of line quality. The spatial relationship of strokes, the formation of the basic components of words or characters, the relative dimensions of two or more alphabets, and the ways of connecting adjacent strokes, in aggregate, constitute the writer's calligraphic profile. The determination of authorship therefore depends on the discovery, in the disputed script, of personal characteristic features that fall within the range of natural variation of writing habits of the author of the control material.

In practice, the process of identification must include a determination of the extent, kind, and significance of this resemblance as well as of the variation. It is necessary to determine whether the divergence is due to the operation of a different individual or is only the expected and inevitable variation found in the genuine writing of the same writer. It is also necessary to decide whether the resemblance is the result of a skilful imitation or is the habitual and characteristic resemblance which naturally appears in a genuine writing. When these two questions are correctly answered, the whole problem of identification is solved (Osborn 1978). Figure 1 gives a diagrammatic presentation on the process of handwriting examination.

For many years, document examiners have been practicing the empirical theory of handwriting examination. The US Supreme Court case, Daubert et al. v. Merrell Dow Pharmaceuticals, alongside with several other cases, namely, US District Court ruling, United States v. Starzecpyzel; US Supreme Court ruling, General Electric Co. v. Joiner; US Supreme Court ruling, Kumho Tire Co. v. Carmichael; and US 11th Circuit Court of Appeals, United States vs. Paul, have prompted the so-called Daubert Standard which refers to the following guidelines for admitting scientific expert testimony: (1) *Judge is gatekeeper*: The task of "gatekeeping," or assuring that scientific expert testimony truly proceeds from "scientific knowledge," rests on the trial judge. (2) *Relevance and reliability*: The trial judge is required to ensure that the expert's testimony is relevant to the task at hand and that it rests on a reliable foundation. The judge must find it more likely than not that the expert's methods are reliable and reliably applied to the facts at hand. (3) *Scientific knowledge = scientific method/methodology*: A conclusion will qualify as scientific knowledge

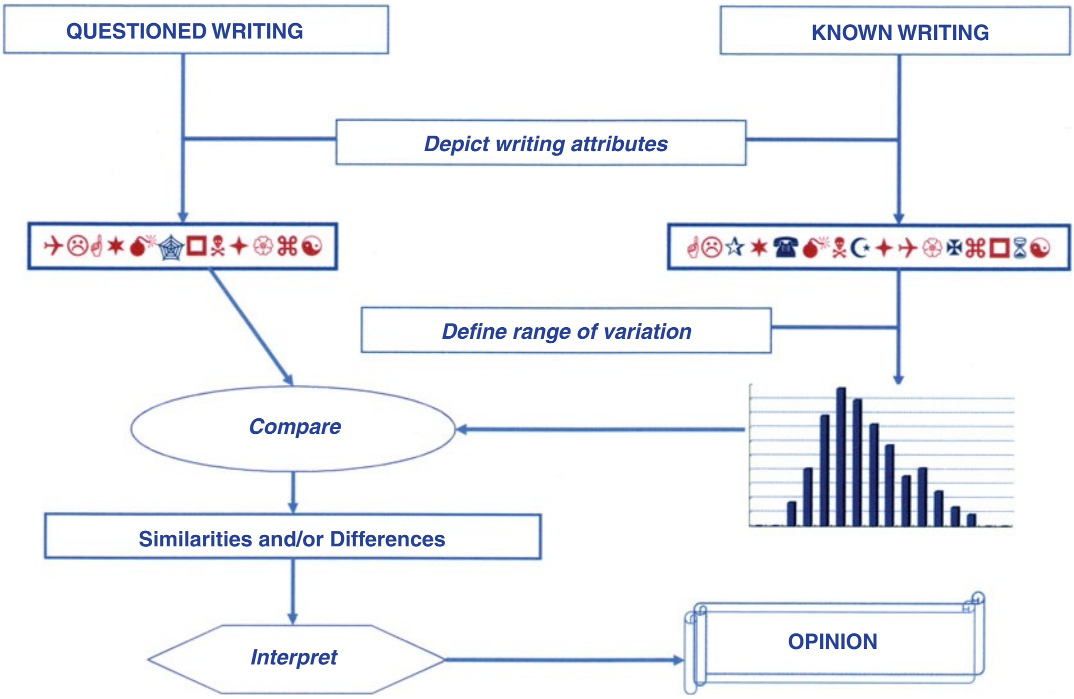
QUESTIONED WRITING
KNOWN WRITING
Depict writing attributes
Define range of variation
Compare
Similarities and/or Differences
Interpret
OPINION

Handwriting as Evidence, Fig. 1 Diagrammatic representation of the handwriting comparison process

if the proponent can demonstrate that it is the product of sound "scientific methodology" derived from the scientific method. (iv) *Factors relevant*: The Court defined "scientific methodology" as the process of formulating hypotheses and then conducting experiments to prove or falsify the hypothesis and provide a test for establishing its "validity":

- Empirical testing: The theory or technique must be falsifiable, refutable, and testable.
- Subjected to peer review and publication.
- Known or potential error rate.
- The existence and maintenance of standards and controls concerning its operation.
- Degree to which the theory and technique is generally accepted by a relevant scientific community.

In 2000, the above rule was amended in an attempt to include the additional provisions which state that a witness may only testify if:

- The testimony is based upon sufficient facts or data.
- The testimony is the product of reliable principles and methods.
- The witness has applied the principles and methods reliably to the facts of the case.

Although the Daubert Standard is concerned with legal requirements for scientific testimony in the United States, it is nevertheless important to assess the mandate of handwriting examination as a specialist forensic science subject.

Scientific Method and Handwriting Examination

Science is a branch of knowledge conducted on objective principles involving the systematized observation of, and experiment with, phenomena especially concerned with the material and function of the physical universe (Allen 1990). It stems from the curiosity of an observant mind, which generates impetus to the quest for an explanation. Observation is the primary scientific procedure of comprehending things; based on observation, scientists attempt to provide logical deductions that can plausibly explain the incident. A hypothesis is thus formulated.

The proposed explanation is tested by way of further observations and experimentation. The latter simulates the observed events under controlled conditions and the designer of the experiment attempts to draw logical inferences from the results thus obtained. Similar experiments, often using a variety of approaches and methodology, are performed by different workers to verify the results reproducibility. When a hypothesis cannot live up with the standard or is considered unsatisfactory, it is either modified or discarded. On the other hand, a hypothesis passing stringent scientific tests will be adopted into the framework of a theory which not only explains the observations in a reproducible way but also makes meaningful predictions.

Application of the scientific method often involves measurement which facilitates a more accurate description of the observed phenomenon in numerical terms and quantifiable units. Phenomena presented in the form of numerical data can be compared and correlated, and their numerical representations treated by mathematical and/or statistical procedures to generate more information which objectively explains the observations in a more detailed and profound manner. Handwriting embodies many parameters that can be measured, and in the Latinized writing – slant or slope – the angle of inclination of the axis of letters relative to the baseline (Hilton 1982) is an important measurable parameter. It varies between writers from backward slanting to a sharply inclined forehand slope. The wide range of variation in slant gives rise to good discriminating power among individual writers.

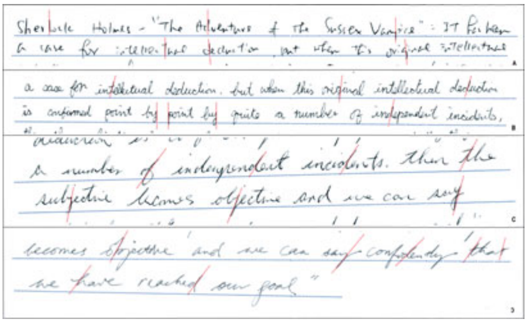
Sherlock Holmes – "The Adventure of the Sussex Vampire" : It has been a case for intellectual deduction, but when this original intellectual

a case for intellectual deduction, but when this original intellectual deduction is confirmed point by point by quite a number of independent incidents,

a number of independent incidents. then the subjective becomes objective and we can say

becomes objective and we can say confidently that we have reached our goal"

Handwriting as Evidence, Fig. 2 Variability of slant in English handwriting

Figure 2 illustrates the variability of slant in the handwriting of four individuals: Writer A exhibits a habit of occasionally writing alphabets with a backward slant, for example, the *l* of *Sherlock*, the *p* of *Vampire*, the first *d* of *deduction*, and the *g* of *original*. But many of the other letters exhibit a near-vertical or slightly forward slant. On the other hand most, if not all, of the characters in sample B are forward slanting, for example, the *d* and the *t*, although a few of the letters are more vertically oriented, for example, the *p* of the two *point*s and *independent*, and the *y* of the *by*'s. Author C produces consistently forward-slanting alphabets in many of the *d*'s and *t*'s. While displaying natural variation, specimen D appears to be more consistently forward slanting. The diversity of slant in the handwriting of different individuals can therefore be utilized for the identification of authorship. Additionally, the slant of different alphabets in a sample of handwriting can vary considerably in magnitude, for example, as evidently shown in the writing of C, the *j* of *subjective* and *objective* and the *y* of *say* are less forward slanting than the other letters. It is generally observed that (1) while the slant of an alphabet in the handwriting of an individual generally conform with one another, different alphabets of the writings of the same individual can vary to a greater extent; (2) in some writings, the position of an alphabet in different words can display different slanting angles, for example, the initial *t* of the word *that* in sample D is much more slanting than the terminal *t* of the same word; and (3) the slant of an alphabet can also vary in accordance with the letters adjacent to, or flanking, it. Leung (Leung et al. 1987) noted the same phenomenon in Chinese handwriting and found that measurable parameters of the handwriting of a writer do not conform to a uniform value but may vary in accordance with the structure of, and the position of the strokes in, the character. The concept of "internal variability" was suggested which is attributed to the divergence of writing habits in relation to the position of the writing element and the overall structure of the written words or characters.

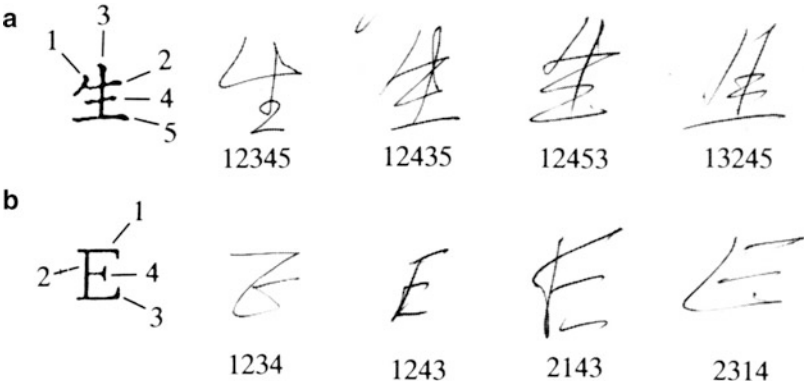

Handwriting as Evidence, Fig. 3 Classification of stroke sequence of the Chinese character *Live* and application of the concept to characterizing the capital letter *E*. (Reprinted with permission from Journal of the Forensic Science Society 1993; 33:9–19)

It was maintained that this calligraphic property can be utilized as an additional tool for handwriting identification (Leung et al. 1987; Leung 1994). To conclude, the numerical expression of depicted features enables the document examiner to characterize handwriting objectively and accurately and is especially useful when the similarities and/or differences are inconspicuous.

Another important facet of the scientific method is classification which simplifies the approach to a given problem by narrowing its broader elements and ramifications on the basis of categorical distinctions. Classification methods can be effectively applied to find out the relationship, if any, of certain observations so that general principles can be derived. It forms the basis of natural sciences such as biology, chemistry, and taxonomy. Apart from writing attributes which can be characterized quantitatively, there are other qualitative features that are equally useful for identifying the writer. Eldridge reported a scheme for the classification of a selected set of cursively handwritten letters comprising *d*, *f*, *h*, *k*, *p*, and *t* (Eldridge et al. 1984). These letters were treated in terms of three major structural components: the *staff*, the *base*, and the *crossbar*. Twelve features were chosen for classification, although not all features are applicable to all letters. They are as follows: (1) *construction order*, the order of construction of the base and the staff; (2) *number of strokes*; (3) *base direction*, the direction of the curvature from the uppermost part of the base to the beginning of the upstroke to the next letter; (4) *base closure*; (5) *top of staff closure*; (6) *top of staff direction*; (7) *bottom of staff closure*; (8) *bottom of staff direction*; (9) *preceding letter join*; (10) *following letter join*; (11) *crossbar position*; and (12) *crossbar curvature*. Each feature has at least two, and possibly as many as seven, distinct forms. It was found that the letter *k* best discriminates between writers in the sample of 61 subjects.

A systematic classification scheme not only provides valuable information on the variability of handwriting but also inspires objective interpretation on examination results. Leung conducted statistical studies into 98 selected Chinese characters written by 438 subjects (Leung et al. 1993). A classification system on the qualitative features of pen emphasis, sequence of strokes, stroke connections, and structural detail was established. Diversities of writing habits depicted among the subjects confirmed the belief that many of the classifiable handwriting attributes contribute toward the science of authorship verification. Figure 3 illustrates four chronological writing sequences in the Chinese character *Live* which comprises one slanting stroke, one vertical stroke, and three horizontal strokes. The possibility of applying the concept of classification by means of stroke sequencing on the capital letter "E" is demonstrated.

Individuality of Handwriting

The principles of handwriting examination have been laid down in classical treatise that the identification of authorship depends on the discovery in the questioned script of a combination of personal characteristics which lie within the range of variation of the control specimens accompanied by the absence between them of any idiosyncrasies that cannot be satisfactorily explained. However, as pointed out by Sugiyama and Kurauchi, the problem of the methods in practice relies on the basic task of determining and extracting a sample's distinguishing features, which have often been questioned for dependence of subjective judgment and intuition rather than on objectivity and empirical evidence (Sugiyama and Kurauchi 1986). The rapid advances in computer technology enable the practical application of statistical classification techniques to handwriting examination. Many forensic scientists reported positive results; to name just a few, Hardcastle, Thornton, and Totty reported a computer-based system for the classification of block handwriting on checks (Hardcastle et al. 1986), and Eisermann and Hecker described the Forensic Information System for Handwriting (FISH), which utilized an image-processing technique with statistical classification analysis for the identification of handwriting (Eisermann and Hecker 1986).

Motivated by several rulings in US courts concerning expert testimony in general, and handwriting testimony in particular, Srihari undertook a study to objectively validate the hypothesis that handwriting is individual (Srihari et al. 2002). They maintained that there are two variances of concern in handwriting comparison, within the handwriting of the same individual and between the handwritings of two individuals, and that the within-writer variance is less than the between-writer variance. The goal of their study was to objectively verify this common observation. A source document in English of 156 words capturing all characters (letters and numerals) and certain character combinations of interest was designed for each participant to copy three times. Handwriting samples of 1,500 individuals, representative of the US population with respect to gender, age, ethnic groups, etc., were obtained. Analyzing differences in handwriting was done by using computer algorithms for extracting features from scanned images of handwriting specimens. Two types of features characteristic of the handwriting were obtained: conventional and computational. Conventional features are the handwriting attributes that are commonly used by forensic document examiners which comprise 21 discriminating elements, namely, (1) *arrangement*, (2) *class of allograph*, (3) *connections*, (4) *design of alphabets and their construction*, (5) *dimensions (vertical and horizontal)*, (6) *slant or slope*, (7) *spacings (intraword and interword)*, (8) *abbreviations*, (9) *baseline alignment*, (10) *initial and terminal strokes*, (11) *punctuation*, (12) *embellishments*, (13) *legibility or writing quality*, (14) *line continuity*, (15) *line quality*, (16) *pen control*, (17) *writing movement*, (18) *natural variations or consistency*, (19) *persistency*, (20) *lateral expansion*, and (21) *word proportions* (Huber and Headrick 1999). Macro- and micro-computational features are those that can be determined algorithmically by software operating on scanned image of handwriting. The following set of 11 macro-computational features were used: (1) *entropy of gray values*, (2) *gray-level threshold*, (3) *number of black pixels*, (4) *number of interior contours*, (5) *number of exterior curves*, (6) *number of vertical slope components*, (7) *number of horizontal slope components*, (8) *number of negative slope components*, (9) *number of positive slope components*, (10) *slant*, and (11) *height*, while micro-computational features are computed at the allograph, character, or shape level.

Statistical analysis was conducted using the identification model and the verification model. In the identification model, given a questioned handwriting sample x and samples of handwriting of n known writers, the classification task was to pick out the writer of x among n writers. In verification model, given two unknown handwriting samples x_1 and x_2 and samples of handwriting of n writers, the objective was to determine whether x_1 and x_2 were written by the same person or by two different people among the n writers. It was found that writer identification accuracy was close to 98 % for two writers while the verification accuracy was about 96 %.

Using global attributes of handwriting and very few characters in the writing, the ability to determine the writer with a high degree of confidence was thus established.

Another study was made to show that the handwriting of an individual is so unique that by applying discriminant analytical methods on the parameters of width-height ratio, symmetry factor, slant, and tilt, the writer could be objectively indentified (Cheung and Leung 1989). In the above experiment, 25 Chinese characters were selected. A total of 376 handwriting specimens were collected from 21 subjects and after 3 months, 6 of the subjects were randomly selected to write once again the 25 characters which were then used as the "questioned samples"; each of the handwriting specimens was characterized by one or more of the 4 measurable parameters. Applying linear discriminant analysis, although the six "questioned" handwritings could not be correctly associated with their corresponding authors by using only one of the parameters, a perfect recognition rate was recorded using all four parameters. The findings are consistent with the hypothesis that verification of authorship is based upon the discovery of a number of divergent characteristic features which in combination become sufficiently personal for an identification to be made.

The above-mentioned two pieces of research have clearly demonstrated the uniqueness of handwriting, and it can be inferred that consideration of both measurable and qualitative parameters in handwriting gives rise to a powerful theoretical foundation for the scientific identification of authorship. Using this theoretical framework, identification of handwritings can be made, with objectivity occupying the center position.

The Art and Science of Handwriting Examination: An Epilogue

Successes in computational classification experiments on handwritings provide scientific support for the long-established (but not vigorously tested) practices of handwriting examination. However, these experiments were conducted on natural handwritings but were not directed to forgery, or any other examinations involving such complications as disguise, handedness, and influence of health and drugs. It is envisaged that for a long time, handwriting identification would have to be made by specialist document examiners instead of relying on computational software for verifying authorship. As such, the entailment of a certain degree of subjectivity would be inevitable in handwriting examination. The objective component of handwriting examination provides the most easily acquired skills of a document examiner, while the subjective reasoning processes require years of training to fully develop (Whiting 1986).

The initial stage of handwriting analysis is basically a comparison of morphological features between the questioned writing and the control exemplars. It depends on the ability, on the part of the document examiner, to find out the more subtle, personal writing attributes in the script. This exercise is *by and large* objective. Some degree of subjectivity is involved because the procedure relies on the expert's cognitive selection of handwriting features which he or she regards as *important* personal characteristics. While it is the author's personal experience that well-trained document examiners generally pick out similar assortment of characteristics from the same handwriting specimens, there are deviations, both in terms of numbers and the ways of how the depicted characteristics are described and comprehended.

The second stage of authorship determination is the interpretation of the depicted writing characteristics. The document examiner has to evaluate the relative significance of the writing attributes which he or she has discovered and consider the information related to the case, for example, the physical or mental condition of the alleged writer, the presence or absence of any external influences, and the possibility of disguise. This is a mental exercise which encompasses subjectivity to a larger extent. Handwriting experts can diverge considerably, especially for difficult cases, in their assessment of the relative importance of the depicted handwriting features. It is during this stage of the examination process that divergence of opinion among handwriting experts sometimes arises.

An element of subjectivity also relates to the expression of an opinion in that after assessing the evidence and forming an opinion, one has to transform the conceptual inclination into written words. Any findings or observations will be useless unless they are comprehended and their significance appraised. Based on the assessment of the examination results, the expert will have to come up with an opinion on the authorship. Inaccuracy and variability can occur. While a *wholly subjective* opinion will be the personal view on the subject matter without any evidential support which is undoubtedly inacceptable, an *entirely objective* opinion would be restricted to facts and figures only. In handwriting examination, a 100 % objective "opinion" will not meet the expectations of the parties concerned because the so-called opinion will only be a list of the similarities or differences that the expert has discovered without any interpretation as to what these findings indicate.

Opinion expressions used by handwriting experts vary among different schools of document examiners, both in terms of opinion expressions and meaning (Leung and Cheung 1989). Some of the more popular opinion expressions being adopted are as follows: *wrote or written by, highly probable or very high probability, probable or high probability, very likely, likely, could well have, consistent with, could have, inconclusive, no evidence, consistent with (not being written by), unlikely, highly unlikely, probably not, could not have, and did not write*. While not all of the above-mentioned expressions are used by a single document examiner, the choice of terminology depends on the reasoned *preference* of the individual expert. The levels of opinion also vary among different practitioners. A few document examiners, which can be described as the "all-or-none" school, believe that only three classes of opinion, namely, identification, elimination, and inconclusive, should be given (McNally 1979). However, the majority of document examiners who maintain that since facts of handwriting examination will vary in number and significance, the language selected to summarize different sets of facts must vary, some being in definite terms and some in qualified terms (Cole 1980). Nowadays, it is widely accepted that in order to reflect the *degree of certainty or belief*, a scale consisting of different levels of expressions should be used. Ellen described a scale of five ranges of opinion and proposed that a small number of divisions of a scale are advantageous in that the significance of each point on the scale can be easily defined (Ellen 1979), but an 11-point opinion scale was also suggested (Leung 1994). In 1991, the Questioned Documents Section of the American Academy of Forensic Sciences and the American Board of Forensic Document Examiners have adopted recommended guidelines in report and testimony which consist of a 9-point opinion scale (identification, strong probability, probable, indications, no conclusion, indications did not, probably did not, strong probability did not, elimination) based on the concept that probability, as used in handwriting opinions, is not a statistical measurement but a measurement of the examiner's confidence based on scientific principles and experienced judgment (McAlexander et al. 1991).

The divergence of the number of opinion scales and terminology used by handwriting experts clearly show that the process of expressing opinions does involve a certain degree of subjectivity on the part of the expert although the expressed opinion is based on scientific observations and experienced judgment. After examining the handwritings or signatures, the expert has to assign the opinion in ordinary language terms that he or she believes to be the most appropriate, and subjectivity comes into play during the transformation from abstract thinking – a degree of certainty or confidence in one's mind – into written words which is susceptible to vagueness and controversy when the readers of the report attempt to convert in their minds these same written words back to the abstract certainty of belief. The inherent difficulty in handwriting and signature examination is due to the fact that contrary to DNA profiling, basic statistical data in relation to frequencies of occurrence of handwriting characteristics, especially for signatures which are unique calligraphic entities of their owners, are unknown. By comparison, the chance match of DNA grouping of, say, one in a billion reported by two laboratories embodies the same meaning because the *entirely* objective opinions from both laboratories are the results derived from statistical data. Handwriting experts *cannot* provide explicit probability figures routinely appearing in DNA grouping reports but instead have to use such ordinary language terms as "highly probable," "probable," or "may" to represent the degree of confidence, or certainty.

The scientific validity of the methodology for handwriting examination has been well demonstrated and from the "utilitarian" point of view, testimonies in handwriting analysis do help jurors and judges make their own decisions, although occasionally expert opinions are criticized for containing a certain element of subjectivity. In this connection, Evett makes an interesting point ". . . that a so-called 'objective test' can only exist within a framework of assumptions and the validity of those assumptions in an individual case is a matter for subjective judgment. The objectivity is an illusion" (Evett 1996). A good enlightenment about the relationship between objectivity and subjectivity, which is aptly applicable to handwriting evidence, can be found in *The Adventure of the Sussex Vampire* where Sherlock Holmes said ". . . It has been a case for intellectual deduction, but when this original intellectual deduction is confirmed point by point by quite a number of independent incidents, then the subjective becomes objective and we can say confidently that we have reached our goal." Hence, the probabilistic principles of handwriting identification can, perhaps, be treated from a philosophical perspective.

Related Entries

- ▶ Automated and Manual Forensic Examinations
- ▶ Cognitive Forensics: Human Cognition, Contextual Information, and Bias
- ▶ Forensic Science and Miscarriages of Justice
- ▶ Forensic Science Effectiveness
- ▶ Philosophy of Forensic Identification
- ▶ Probability and Inference in Forensic Science
- ▶ Scientific Basis of the Forensic Process

Recommended Reading and References

Allen RE (ed) (1990) The concise Oxford dictionary of current English, 8th edn. Oxford University Press, Oxfor

Cheung YL, Leung SC (1989) A comparative approach to the examination of Chinese handwriting part 4 – identification by statistical classification techniques. J Forensic Sci Soc 29:77–8

Cole A (1980) The search for certainty and the use of probability. J Forensic Sci 25:826–83

Eisermann HW, Hecker MR (1986) FISH-computers in handwriting examinations. In: Proceedings of the 44th annual meeting of the American Society of questioned document examiners, Savannah, Georgi

Eldridge MA, Nimmo-Smith I, Wing AM, Totty RN (1984) The variability of selected features in cursive handwriting measures. J Forensic Sci Soc 24:179–21

Ellen DM (1979) The expression of conclusions in handwriting examination. Can Forensic Sci J 12:117–12

Ellen D (1993) The scientific examination of documents, 1st edn. Ellis Horwood, Chicheste

Evett IW (1996) Expert evidence and forensic misconceptions of the nature of exact science. Sci Justice 36:118–12

Hardcastle RA, Thornton D, Totty RN (1986) A computer-based system for classification of handwriting on cheques. J Forensic Sci Soc 26:383–39

Harrison WR (1966) Suspect documents. Their scientific examination, 2nd edn. Sweet & Maxwell, Londo

Hilton O (1982) Scientific examination of questioned documents, revised edition. Elsevier North Holland, New York

Huber RA, Headrick AM (1999) Handwriting identification: facts and fundamentals. CRC Press, Boca Raton

Leung SC (1994) The scientific examination of Chinese handwriting. Forensic Sci Rev 6(2):97–145

Leung SC, Cheung YL (1989) On opinion. Forensic Sci Int 42:1–13

Leung SC, Chung MWL, Tsui CK, Cheung WL, Cheung SYL, Mok MMC (1987) A comparative approach to the examination of Chinese handwriting part 2 – measurable parameters. J Forensic Sci Soc 27:157–173

Leung SC, Cheung WL, Fung HT, Cheung YL (1993) A comparative approach to the examination of Chinese handwriting part 5 – qualitative parameters. J Forensic Sci Soc 33:9–19

McAlexander TV, Beck J, Dick RM (1991) The standardization of handwriting opinion terminology. J Forensic Sci 36:311–319

McNally JP (1979) Certainty or uncertainty in expert testimony. J Police Sci Admin 7:26–27

Osborn AS (1978) Questioned documents, 2nd edn. Patterson Smith, Monclair

Srihari SN, Cha SH, Arora H, Lee S (2002) Individuality of handwriting. J Forensic Sci 47(4):856–874

Sugiyama T, Kurauchi H (1986) Identification of handwriting in Chinese characters using discriminant analysis. Behaviormetrika 19:55–71

U.S. 11th Circuit Court of Appeals, United States vs. Paul. (97–9302), 999

U.S. District Court ruling, United States v. Starzecpyzel. 880 F. Supp. 1027 (S.D.N.Y.), 1995

U.S. Supreme Court ruling, Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 1993, 584–587

U.S. Supreme Court ruling, General Electric Co. v. Joiner. (96–199) 78F.3d 524

U.S. Supreme Court ruling, Kumho Tire Co. v. Carmichael. (97–1709) 131 F.3d 1433

Whiting F (1986) The reasoning process: an integral component of questioned document examinations. In: Paper presented at the 44th annual meeting of the American Society of questioned document examiners, Savannah, Georgia

Handwriting Examination: The Art and Science

▶ Handwriting as Evidence

Harassment, Threat Assessment

▶ Stalking

Hate Crime

Barbara Perry
Faculty of Social Science and Humanities,
University of Ontario Institute of Technology,
North Oshawa, ON, Canada

Synonyms

Bias crime; Bias/Hate motivated crime; Ethnoviolence

Overview

This entry aims to lay out some of the key controversies developments in hate crime scholarship and practice in recent years. Specifically, it speaks to the dilemmas of defining and legislating hate and to the harms associated with hate crime. In addition, it enumerates “what’s hot” and “what’s not” in terms of current scholarship, identifying those areas that are well developed and those that have received little attention. It concludes with a discussion of emerging nonpunitive means by which to address hate crime in practice.

Key Issues

Sadly, the phenomenon we have come to call hate crime has a lengthy history in the west. History does indeed repeat itself as similar patterns of motivation, sentiment, and victimization recur over time. Just as immigrants in the 1890s were subject to institutional and public forms of discrimination and violence, so too were those of the 1990s; likewise, former black slaves risked the wrath of the Ku Klux Klan when they exercised their newfound rights in the antebellum period, just as their descendants risked violent reprisal for their efforts to win and exercise additional rights and freedoms in the civil rights era; and women who demanded the right to vote on the eve of the twentieth century suffered the same ridicule and harassment as those who demanded equal rights in the workplace later in the century. While the politics of difference that underlie these periods of animosity may lie latent for short periods of time, they nonetheless seem to remain on the simmer, ready to resurface whenever a new threat is perceived – when immigration levels increase, or formerly powerless groups are suddenly empowered, or when relationships between groups shift for other political, economic, or cultural reasons. Anders Breivik’s appalling attacks in Norway attest to this, as does the surge of Islamophobic violence in the aftermath of the September 11 attacks.

Nonetheless, it has only been in the past two or three decades that practitioners and academics have begun to take hate crime seriously. Most scholars trace the origins of the concept now known as hate crime to the 1980s. The first legislative recognition of hate crime in the United

States came in the form of a federal hate crime statute in 1969 and as state act in California in 1978. The real "flurry" began, however, with the signing of the Hate Crime Statistics Act of 1990. By the end of the 1990s, virtually all US states had hate crime legislation. Most other western nations seem to have followed suit, introducing their own variants during the late 1980s and early 1990s. Interestingly, while most explicitly used the umbrella term hate crime, the UK did not, opting instead for more directed legislation around racial hatred, and later religious hatred.

Oddly, scholarship in the field did not seem to keep pace with legislation. We were much slower to recognize this emerging construct (Shively 2005). Certainly, there were no major works with "hate crime" in the title prior to the mid-1980s. In fact, few references of any kind are to be found in the scholarly literature, until 1989 and 1990, which seemed to be watershed dates. The field really took off in the opening years of the 1990s, with the publication of such works as Kelly's (1991) *Bias Crime: American Legal Response*, Herek and Berrill's (1992) *Hate Crimes: Confronting Violence Against Lesbians and Gay Men*, and Levin and McDevitt's (1993) *Hate Crimes: The Rising Tide of Bigotry and Bloodshed*. This entry seeks to assess the current state of the art in the field.

The entry opens with traditional definitional dilemmas, moving then to identify emerging approaches to the consideration of the impacts of hate crime for the individual as well as the community(ies) at large. This is followed by a discussion of the tensions associated with legislating and policing hate crime. Focus is then turned to a consideration of "what's hot" and "what's not" in the context of hate crime scholarship. The entry ends with a call to consider the value of nonpunitive approaches to hate crime. The aim is to draw attention to areas in need of further development and to inspire a dialogue on how to move forward both academically and practically.

Defining Hate Crime

One of the greatest challenges facing the prevention and response to hate crime is the lack of a common definition and understanding of what constitutes hate crime. There are legal definitions which have generally followed the lead of the model legislation as set out by the Anti-Defamation League:

> A person commits a Bias-Motivated Crime if, by reason of the actual or perceived race, color, religion, national origin, sexual orientation or gender of another individual or group of individuals, he violates Section ______ of the Penal code (insert code provisions for criminal trespass, criminal mischief, harassment, menacing, intimidation, assault, battery and or other appropriate statutorily proscribed criminal conduct).

Slightly different are "hate incidents" which are

> expressions of bias, prejudice and bigotry that are carried out by individuals, groups, organizations and states, directed against stigmatized and marginalized groups or communities, and intended to affirm and secure existing structures of domination and subordination. (Hate Crimes Community Working Group (hereafter HCCWG) 2006: 18)

It is important to include these noncriminal incidents in our understanding of hate crime, since what appear to be relatively petty incidents are very common and thus wear cumulatively on the victim. That means that, from a scholarly point of view, legalistic definitions are too exclusive, relying as they do on violations of criminal law. They are also significantly limited in their ability to breathe life into the term. They provide little insight into the power dimensions that are inherent in the act of hate crime. Consequently, the preferred definition that has come to be used widely in the literature accounts for the politicized nature of hate crime and the role of hate crime in constructing the relative identities and subject positions of both the victim and the offender, individually and collectively:

> Hate crime, then, involves acts of violence and intimidation, usually directed toward already stigmatized and marginalized groups. As such, it is a mechanism of power and oppression, intended to reaffirm the precarious hierarchies that characterize a given social order. It is intended to simultaneously recreate the threatened (real or imagined) hegemony of the perpetrator's group, and the "appropriate" subordinate identity of the victim's group. It is a means of marking both the Self and the Other in such a way as to re-establish

> their "proper" relative positions, as given and reproduced by broader ideologies and patterns of social and political inequality. (Perry 2001)

Here, we see context, intent, and impact. Hate crime is embedded in relations of power and intended to reinforce extant relations. And its aim and effect – typically – is to intimidate a group of people who "hold in common a single difference from the defined norm – religion, race, gender, sexual identity" (Pharr, cited in Wolfe and Copeland 1994: 203). As such, hate crime is a "message crime," intended to remind not just the immediate victim but also his/her community of their "rightful" place in carefully crafted hierarchies of race, religion, gender, and sexuality, inter alia.

What's the Harm?

This notion of "message crime" has more recently given rise to more attention to the issue of the impacts of hate crime. In particular, scholars are beginning to assess the ways in which collective harm might derive from bias-motivated violence. Indeed, running through much of the hate crime literature – even through court decisions on hate crime – is the assumption that such offenses are, for that same reason, qualitatively different in their effects, as compared to their non-bias-motivated counterparts. Specifically, Weinstein (1992) identifies three potential levels of harm: "that racial violence causes injury to the victim above and beyond physical damage, that racial violence causes injury not only to the immediate victim but also to the victim's racial or ethnic group, and that racial violence has particularly pernicious ramifications for society as a whole." The first of these has been the subject of considerable scholarly attention. The empirical findings in studies of the physical, emotional, psychological, and behavioral impact of hate crime are beginning to establish a solid pattern of more severe impact on bias crime victims, as compared to non-bias victims (see, e.g., McDevitt et al. 2001).

When we move beyond the experiences of the immediate victim to the broader *in terrorem* effects, we enter the realm of speculation. Many scholars point to the "fact" that hate crimes are "message crimes" that emit a distinct warning to all members of the victim's community: step out of line and cross invisible boundaries and you too could be lying on the ground, beaten and bloodied. Consequently, the individual fear noted above is thought to be accompanied by the collective fear of the victim's cultural group, possibly even of other minority groups likely to be victims. Yet, there have been few studies that explicitly survey members of victims' reference communities to determine the veracity of this assumption (Lim 2009; Noelle 2002; Perry and Alvi 2011).

Another related area that has received scant attention has been the impact of hate crime on perceptions of national ideals. Hate crime throws into question national commitments to tolerance and inclusion. Speaking specifically of Native Americans over 50 years ago, legal scholar Felix Cohen noted that mistreatment – legal or extralegal – of minorities "reflects the rise and fall of our democratic faith." In other words, it is possible that the persistence of hate crime is a challenge to democratic ideals. It reveals the fissures that characterize its host societies, laying bare the bigotry that is endemic within each. As such, it may very well be the case that bias-motivated violence is not just a precursor to greater intergroup tension, but is an indicator of underlying social and cultural tensions. In this interpretation, hate crime is but one indicator that enshrined ideals of freedom and equality are illusory.

Alternatively, one possibility is that hate crime acts as a catalyst to positive change. That is, patterns of persistent violence, or highly publicized cases – like the 1998 Matthew Shepard or James Byrd cases – often have the unintended effect of mobilizing victim communities and their allies. For example, the racially motivated murders of Michael Griffith in Howard Beach in 1996 and Yusuf Hawkins in Bensonhurst in 1989 both inspired widespread demonstrations condemning the racism of the perpetrators' communities, as well as the racist culture of New York City generally. On a smaller scale, we've seen some evidence of this in Ontario in response to the alleged incidents of hate crime against

Asian anglers north of Toronto. Indeed, community outcry resulted in the establishment of an inquiry into those events (Ontario Human Rights Commission 2007).

Legislating Hate: Persistent Challenges

Debate continues to swirl over whether, in fact, we should even attempt to legislate "hate" crime as a distinct entity. For some, this springs from practical concerns with identifying motive. The argument is that it is exceedingly difficult to determine what may have motivated any particular incident, especially in the absence of offensive language or other direct indicators. Thus, officers are likely to be required to gather additional background information "proving" motive – such as ties to racist groups, or previous expressions of homophobic attitudes. It is this approach that raises the thorny issue of "hate thought."

The most strident critics of hate crime legislation are those who equate the regulation of hate crime with the regulation of hate "thought." Jacobs and Potter (1998) have been especially articulate in their charges in this respect. Writing of sentencing enhancement provisions, for example, they question how increased punishment for a bias motive could be interpreted as anything other than punishment for beliefs, that is, thoughts, rather than simply the actions. Often, this argument is expressed in constitutional terms, in so much as restrictive legislation is perceived as a violation of the freedom of expression.

Nonetheless, most western nations have responded to public concerns about bias-motivated violence with an array of policy interventions. As early as 1981, the Anti-Defamation League (ADL) developed model legislation aimed at confronting hate crime. The framework has proven popular not just in the United States but internationally as well, with nations adopting similar legislative structures.

The most momentous piece of explicit hate crime legislation at the federal level in the USA was the Hate Crime Statistics Act in 1990, according to which states and their respective law enforcement agencies were "required" to report hate crime statistics on an annual basis. With the passage of the act, the federal government appeared to have committed itself to the task of collecting "accurate" information on hate crime nationwide. The Act mandated that all law enforcement agencies report annually on reported and recorded hate crime in their jurisdictions. As of the 2008 reporting year, however, less than 20 % of relevant agencies were in compliance.

The limits of the US federal legislation are immediately apparent in S(b)(1) itself, especially with respect to the narrow definition of both the protected groups and the enumerated offenses. Only five grounds for motivation and eight offenses are to be counted in the UCR. This leaves a lot of ground uncovered. Other criminal offenses and equally injurious noncriminal offenses are left uncounted. Similarly, victimization on the basis of gender or political orientation, for example, is excluded. Moreover, this brings to mind the problem of inconsistency between reporting agencies. Not all states recognize the same categories of bias in their legislation. Some states do not include gender in their hate crime legislation; some don't include sexual orientation; yet others include such anomalous categories as "whistle blowers." In short, we have yet to resolve, either philosophically or legally, the question of the "boundaries" of protection.

This problem is, of course, exacerbated in the EU context where member states have diverse pieces of legislation, covering a wide array of different offenses, protected categories, and/or sentencing responses. Reports from the Office for Security and Cooperation Europe (OSCE 2009), for example, consistently bemoan the lack of standardization across EU states. In fact, some states have no specific provisions on hate crime (e.g., Greece, Italy and Portugal). A 2009, OSCE report on hate crime laws includes an overview of the different protected categories among member states, including the most commonly protected classes (e.g., race), frequently protected classes (e.g., gender), and rarely protected classes (e.g., political affiliation). Similarly, states vary on the nature of the legislation, ranging from genocide, to sentencing enhancement, to hate speech provisions. The UK is an interesting case. Policy guidance is largely derived from the Association of Chief Police Officers (ACPO) which holds a relatively inclusive definition of hate crime as an offense motivated by race, sexual orientation, faith, and disability. Yet prior to 2003, concrete legislation revolved around racially motivated crime (Crime and Disorder Act, 1998) and religiously motivated crime (Anti-terrorism, Crime and Security Act, 2001). With the passage of the Criminal Justice Act 2003, provisions were extended to include sexual orientation and disability. Protection is thus still limited to only four categories.

Policing Hate

The weaknesses associated with legislative efforts have had the effect of similarly limiting law enforcement's ability to investigate and respond to hate crime. In addition, Bell (2009: 35) identifies an array of structural limitations on police recording of hate crime including the presence of departmental hate crime policies, level of officer training, and administrative commitment to the issue. As with statutory provisions, there is wide variation among police departments in the extent to which any of above factors prevails. In the USA, for example, it appears that very few departments are effective in identifying or investigating hate motivated crime. On the contrary, very few acknowledge hate crime when it occurs. For instance, in 2005, only 16 % of agencies reported any hate crime; most departments in the Deep South reported no hate activity for the year, which many would find incredible.

In addition to the limitations imposed by law enforcement agencies are those presented by trends in public underreporting. In fact, some argue that hate crimes are even more dramatically underreported than other UCR offenses (Berrill 1992). Gay victims, for example, may fear that the admission of their victimization is concomitantly an admission of their sexual orientation. Reporting an anti-gay crime to the police is tantamount to "outing" themselves – an event for which they may not be prepared. Similarly, the undocumented laborer may fear the repercussions of his or her status being revealed. Moreover, victims may well fear secondary victimization at the hands of law enforcement officials. At the very least, they may perceive that police will not take their victimization seriously.

Most western nations have attempted to enhance their efforts around hate crime through the development of viable and engaged bias crime units, or in the case of smaller agencies, at least one or two personnel devoted to leading the effort to prevent and investigate hate crime. Most large regional police agencies in Canada, for example, Toronto, Edmonton, and Vancouver, have such dedicated units, generally comprised of 2–4 officers and occasionally civilian researchers and/or analysts.

What's Hot...

If there is a single field within hate studies that can be said to be highly developed it is in the area of violence against lesbian and gay communities, and especially the latter. Indeed, among the first book length publications in the 1990s were two that dealt explicitly with violence against gays and lesbians (Comstock 1991; Herek and Berrill 1992). Herek would follow this with a series of papers further detailing the "sequalae" of hate crime against gays and lesbians. Indeed, these two books seemed to open the floodgates for work on anti-gay violence, in that it dominated the substantive literature in the broader field of hate crime for most of that decade.

However, two key pieces were missing from the literature on sexualities and on gender. To this day, there has been very little scholarship on the experiences of lesbians specifically. It is as if scholars assumed that the dynamics and impacts of violence of gay men and women alike were equivalent. Indeed, simply glancing over journal titles makes this elision clear, in that they typically include the phrase "lesbians and gay men," rather than one or the other. Moreover, these pieces do not typically separate the differential experiences, dynamics, or impacts according for gay men and lesbians. This dismisses the fact that gender – and gender expectations – as well as sexualities play a significant role in motivating these crimes.

In the current lexicon, we generally speak of "the" LGBT community. However, the "T" – transgender community – has largely been invisible to hate crime scholars. While "trans violence" is on the surface concerned with gender identity, the emerging scholarship makes it clear that it, too, is often multifaceted, insomuch as sexuality comes into play, at least from the perpetrators' perspective. Shelley (2008) shares a number of victim narratives in which it is difficult to disaggregate the motivating factors in their assaults. He observes that "trans people are frequently subject to homophobia and heterosexism, whether or not they themselves are heterosexual in orientation" (Shelley 2008: 96). This suggests, then, that trans people are being punished for what is perceived to be inappropriate performance of both their gender and sexual identities. Available evidence is beginning to point to the fact that trans violence may be even more brutal than homophobic violence and more likely to involve sexual assaults. So, too, is it likely to be more dramatic in its impacts. This is perhaps part of the reason why trans people have higher rates of suicide than even their homosexual peers. Given the nascent stage of scholarship in this area, there is still much to learn.

Another long ignored class of targeted violence that has more recently come to the fore is what is typically referred to as disablist hate crime directed toward those with physical or mental disabilities. In the United States, this was the last protected class to be included in the reauthorization of the Hate Crime Statistics Act (HCSA) in 1996. With sexual orientation, it was only added to UK legislation in 2003. Disability has been a protected category in the Canadian sentencing enhancement provision since its inception. Regardless of its legal status, disablist hate crime has long been neglected by hate crime scholars. Hannah Mason-Bisch (2010: 62) suggests that this is because policy makers and scholars alike place disability at the bottom of the "victim hierarchy." In recent years, a number of scholars have turned their attention to documenting the extent, intensity, and impacts of anti-disability violence. Thus, it is becoming increasingly clear that people with disability are subject to the full spectrum of emotional and physical violence that characterize other communities (Sherry 2010; Lane et al. 2009). What distinguishes this group from others, however, is that they are much more likely to be victimized by people they know – especially caregivers – than is the case for others. Sadly, this dynamic, coupled with a common inability to recognize hate crime, means that disablist violence is dramatically undercounted. It is, thus, an area in need of deeper empirical examination.

The last decade has, for obvious reasons, seen increased attention to Islamophobic violence. Since the terrorist attacks in New York City and Washington DC and above Pennsylvania on September 11, 2001, ethnic minorities associated with Islam in most western countries have experienced increased negative attention from the media, from police and security forces, and indeed from agitated citizenry. There has been a concomitant increase in all such countries in the extent of anti-Muslim or "Islamophobic" hate crime, racial vilification, and discrimination (Poynting and Noble 2004; Welch 2006). This has been exacerbated by subsequent terrorist events, notably Bali in October 2002 and October 2005, Madrid in March 2004, and London in July 2005.

In addition to scholarship on particular victim categories, several other issues have captured the imagination of scholars and practitioners. Since the early days of the recognition of hate crime as a public concern, emphasis has been placed on the activities and rhetoric of organized hate groups. In recent years, more attention has been focused on the online activities of these groups. It is likely that this will only increase in the aftermath of Anders Breivick's attacks in Norway. He had made extensive use of electronic media to share ideas and especially to disseminate his manifesto. This was not something new, but rather something that a natural extension of the growth of the "Web of hate" initiated as early as 1985 when Donald Black created *Stormfront*, which is generally considered to have been the first white supremacist Web site.

The proliferation of hate-oriented and "White Power" sites has occurred at such a rate that it is virtually impossible to estimate the number of such Web sites. Such counting is made even more difficult by the fact that sites move and shut down on a frequent basis, often resurfacing on other servers with other names. Nonetheless, monitoring organizations have tried to quantify the problem. One 2005 estimate, from the Simon Wiesenthal Center, suggested that there were more than 5,000 Web sites advocating hatred at that time. A 2005 International Network Against Cyber Hate publication reported that "during the last 4 years, INACH members handled about 15,000 cases of online hate" (p. 7) among its network of 14 member organizations. It is difficult to assess the accuracy of such estimates. What monitors and scholars can agree on, however, is that the number of such sites has vastly increased over the past decade. Today, it is probably not an exaggeration to say that there are thousands of domestic and international Web sites marketing various brands of hatred and intolerance. For example, as of this writing (December 2012), the Hate Directory runs to nearly 200 pages and thousands of sites.

What's Not...

Generally speaking, hate crime scholars have not paid very close attention to the concrete experiences of discreet categories of hate crime victims. Comments above about new directions in victim research nothwithstanding, it remains the case that we have done insufficient empirical work with vulnerable communities. In light of this shortcoming, it is hardly surprising, then, that we have failed to take seriously the implications of intersectionalities for how hate crime is experienced. We simultaneously occupy multiple subject positions or identities. Kimberlé Crenshaw (1994) was among the first scholars to highlight academic neglect of the "intersection of racism and patriarchy." We might add other categories of difference to this equation, such that we attend to the intersection of racism and patriarchy, but also heterosexuality, class, religion, and disability. All black people do not experience their race in the same ways, but through the prisms of the other elements of their identity. Similarly, to be gay is to experience life differently depending on one's ethnicity, or religion, or sex. Our identities and life chances are dramatically affected not by one element of who we are, but multiple, cross-cutting, and sometimes contradictory positions. Yet this is not something that has taken up much space in the hate crime literature. A powerful exception to this is the work of Doug Meyer (2010). In a series of articles, he has explored empirically and theoretically how gay men and lesbians experience hate crime differently depending upon their class, race, and gender identities. This is a fruitful beginning to a new area of inquiry.

Even in light of the above assessment of the limited depth of research on vulnerable communities, it is clear that we still know more about victims of hate crime than we do about offenders. There are myriad reasons for this. First and foremost is that few such crimes are reported by victims. Where there is no documentation, there is no way to know about the offender. Additionally, even where police are involved, it is very likely that no suspect will be identified. Hate crimes are often random, anonymous events. Especially in incidents of vandalism or graffiti or other property crimes, there is no direct contact between the victim and offender that would allow identification.

Beyond these investigative problems lie academic limitations as well. We frequently speak to victims of hate crime, but rarely do we engage with offenders. In part, this is due to the difficulty in soliciting volunteers for interviews or surveys. Generally, offenders are not necessarily anxious to admit their wrongdoing. A dramatic exception involves members of organized hate groups – these typically young men are often quite happy to share their perspectives on why people of color, gays, or Jews, for example, "deserve" their victimization. However, the proportion of hate crime committed by such individuals remains a very small part of the population of offenders. Consequently, very little research has delved into the explicitly expressed motivations of the "typical" hate crime offender.

Across nations, it is clear that the typical hate crime offender is a young white male who is generally performing among a group of peers. Yet while perpetrators act in a group, it is nonetheless a myth that they are associated with organized hate groups. It is generally thought that less than 5 % of offenders have strong ties to white supremacist or other hate groups. Who, then, are the remainder? It is perhaps troubling to admit that, for the most part, they are probably "normal" young men, acting out cultural scripts of difference and superiority. Very early in the history of hate crime scholarship, Levin and McDevitt (1993), for example, argued that "hate crimes are more often committed under ordinary circumstances by otherwise unremarkable types – neighbors, a coworker at the next desk, or groups of youngsters looking for a thrill" (5).

Levin and McDevitt (1993) were also early to try to explicitly discern the motives provoking hate offenders. They created the now-classic typology of hate crime motives that included thrill seeking, defensive violence, mission/moral agenda, and retaliatory violence. While there have been attempts to replicate and refine Levin and McDevitt's categorization of offenders (Fisher and Salfati 2009), little work has been done with the typical hate offenders – rather the emphasis, noted above, continues to be on members of organized hate groups in spite of their relative underrepresentation as actual perpetrators of violence. This is especially the case in North America. In contrast, there have been a handful of powerful studies in the UK in recent years. Among these are Sibbit's (1997) study of racist offenders on two south London housing estates and Gadd and Dixon's (2011) study in the English North Midlands. Both provide important insights into the cultural and biographical histories that contextualize perpetrators attitudes and behaviors and are thus valuable starting points. Yet they may be limited geographically and culturally. We would do well to explore such violence in other contexts as well.

This suggests another limitation to our current understanding of hate crime. It might be said that we have been myopic, if not blatantly ethnocentric, in our study of hate crime. Attention is focused almost exclusively on the "western" experience of hate crime, with particular emphasis on the USA, the UK, Australia, and Western Europe. We know virtually nothing about racial or religious violence in Japan, or homophobic violence in Saudi Arabia, for example. There is some extant literature on the genocides of Rwanda, or the ethnic cleansing of the Balkan states, but we have not mined these tragic conflicts for lessons about the evolution of hate, or what the ADL might refer to as the pyramid of hate.

In 1998, two "international" anthologies on hate crime were published almost simultaneously. As strong as these were in many respects, they did little to expand our scholarly borders. Kaplan and Bjorgo's (1998) *Nation and Race* focused almost exclusively on European issues. Kelly and Maughan's (1998) *Hate Crime: The Global Politics of Polarization* strayed a little further afield, to address hate crime in India, ethnic cleansing in Colombia, and Palestinian persecution in the Arab world alongside the standard European and American fare. Subsequent to these two anthologies, however, little else has been done to examine bias-motivated violence outside of Europe and North America. And where attention is paid to racialized or other forms of bigoted violence, as in the Rwandan genocides, for example, this is not typically connected to the continuum on which hate crime falls. Interestingly, an exception appears to be emerging and that is consideration of anti-gay violence in the Caribbean (White and Gerke 2007).

Recent events, such as the 9/11 attacks and even Breivek's assaults in Norway, have raised the question of the links between terrorism and hate crime. Rosga (1999) and Hamm (1994) have separately articulated arguments that hate crime might, in fact, be understood as a form of terrorism, whereby "an attack motivated by prejudice targets not only its individual victim, but by its symbolic weight, effectively targets a whole group of marked individuals . . .. It functions, in other words, to reduce complex, multi-faceted individuals into one-dimensional, victimized identity categories" (Rosga 1999: 145).

Similarly, Mark Hamm (1994) champions the use of the term "domestic terrorism," by which he means violence – generally perpetrated by organized extremists – intended to reinforce a "putative norm." Typically, that "putative norm" refers to existing hierarchies along raced or gendered lines.

For obvious reasons, "terrorism" – especially now – is a powerfully evocative word. Both uses of the term capture the dramatic collective impact of bias-motivated violence. The term demands the recognition that "hate crime" is intended to reinforce the subordination – perhaps even the invisibility – of the targeted groups. The two phenomena are at least close cousins (Perry 2009). One might say they run along the same continuum of intimidation and fear, differing in scale more than kind. Yet how is this assessment useful? For one, it encourages us to take hate crime seriously. It could easily evolve into terrorism, as it does among organized hate groups, for example. Yet this point is typically obscured "when attacks on non-government targets, such as abortion clinics or members of unpopular or stigmatized religions, ethnic, racialized, sexual minorities, immigrants or gypsies are defined as 'hate crimes,' thus suggesting that such attacks are no more than expressions of personal prejudice" (Jaggar 2005: 204). As most would agree, however, the intent of hate crime is not only or even primarily intimidation of an individual, but of their community.

Countering Hate Crime

The majority of western nations have responded to hate crime punitively, opting for harsher sentences. However, we've learned in the last decade or so that harsher sentences don't necessarily make safer communities. Purely punitive responses have the potential to be counterproductive, in that they may "actually increase the social divisions they are designed to ameliorate" (Franklin 2002: 154). Moreover, hate crime offenders who go to prison often find themselves among peers who will reinforce their racist, or homophobic, or religious biases. Most offenders are youth, especially young men who are responding to what they see as a threat – to their community, to their neighborhood, or to their self-esteem. Often, these threats are more imagined than real. It has proven to be more effective, then, to challenge those myths and to thus "humanize" the victims and their communities. Incidents of hate crime can be taken as a starting point for education and healing rather than simply punishment. Consequently, community-based responses represent valuable alternatives. For the most part, however, these remain underdeveloped both conceptually and in practice. Indeed, Iganski and Smith (2011) recently canvassed a number of western nations, seeking evidence of programs for young hate crime offenders specifically. The findings were meager indeed, with nations like Australia, New Zealand, and Canada lacking any such programs, while promising US initiatives had been largely defunded. Germany, Sweden, and the UK seemed to be the only states boasting a commitment to anti-hate programming. While limited in number, the programs that were identified did seem to point to the importance of nonpunitive approaches. Alternative initiatives can be grouped into a number of categories.

Education. Rabbi Steven Moss, creator of New York City's Stopbias program stated that "I find most defendants are not bigoted in their hearts but are acting out, using hate words but often not knowing why they are hateful. You want to create an environment in which the [student] can grow from this." With this in mind, one option is to recommend an educational opportunity for hate crime offenders. This will make an impact on both the community and the offender. At the simplest level, the court could be asked to require that offenders take college or university courses on diversity, or on the community that has been victimized. Another, that is widely used in the United States, is for offenders to engage in community service within the victim's community. These alternatives are intended to deter further bias-motivated violence by teaching offenders the effects of their actions and by putting a human face on the victims.

Offender Counseling. Related to this are more formalized anti-bias programs that might be available in local communities, either through the justice system or through nonprofit agencies. Nearly two-thirds of all known perpetrators of hate crimes are teenagers or young adults. When appropriate, a victim-offender restitution program or offender counseling program can be an effective sanction for juveniles. Like the educational initiatives noted above, these are grounded in the understanding that young offenders, especially, often act out of ignorance and uncritical adoption of community norms of racism, homophobia, etc.

Counseling initiatives provide an opportunity to educate hate crime offenders about those who are the object of their violence. The goal is to alter and expand the offenders' perceptions of other cultures. These programs can consist of such activities as visits to places of worship or community centers, listening to guest speakers from other cultures, and listening to stories from victims of bias crimes. It might also include lessons on human and civil rights law, racism, anti-Semitism, and other forms of prejudice and a short history outlining the legacy of discrimination against minorities in America.

Victim-Offender Mediation. Recently, victim-offender mediation has been used in an increasing number of contexts, including hate crime. This option allows victims and offenders enter into a dialogue intended to enhance their understanding of one another: the motives of the offender and the effects on the victim. Part of the dialogue – which must be freely entered by both parties – is also the resolution of their conflict, generally by the offender admitting guilt and offering a sincere apology. Additionally, however, the two parties are meant to come to a mutually acceptable agreement about "what's next," typically, some kind of restitution. The goal in victim-offender mediation, then, is to empower the victim while evoking compassion and understanding on the part of the offender, so as to minimize recidivism.

Restorative Justice. In some cases, restorative justice may also be viable, particularly for low-level bias-motivated offenses and for juvenile offenders. The restorative justice model goes beyond victim-offender mediation to promote involvement of the victim, the offender, *and* the community in the justice process. In particular, restorative justice interventions help to restore victims' and communities' losses by holding offenders accountable for their actions by making them repair the physical and emotional harm they have caused. Such interventions also focus on changing the behavioral patterns of offenders so that they become productive and responsible citizens. The restorative justice model places emphasis on everyone affected by the crime – the community and the victim as well as the offender – to ensure that each gains tangible benefits from their interaction with the criminal justice system. This alternative, then, specifically addresses the community impact of hate crime, allowing any affected party a place at the table.

It is not enough to intervene at the level of the individual offender. Perpetrators are largely reflections of the communities within which they live, where sentiments are shaped by popular opinion as well as political rhetoric. We can expect little popular change in attitude without a corresponding – or even precedent – shift in government policies and rhetoric. In light of the rise of viable right-wing political parties across the west, this battle will not be an easy one. Hate-motivated vilification and violence can only flourish in an enabling environment. In most western nations, such an environment has historically been conditioned by the activity – and inactivity – of the state. State practices, policy, and rhetoric often provide the formal framework within which hate crime – as an informal mechanism of control – emerges. Practices within the state, at an individual and institutional level, which stigmatize, demonize, or marginalize traditionally oppressed groups legitimate the mistreatment of these same groups on the streets.

If political rhetoric fuels the flames of hatred, then it is also clear that a positive politics of difference expressed at the level of the state is vital if we are to temper, if not extinguish, those flames. First and foremost, politicians must assume a leadership role in condemning rather than embracing hate crime, organized hate groups, and other blatant expressions of intolerance. As leaders, they must express a commitment to social justice grounded in justice rather than injustice, inclusion rather than exclusion, respect for rather than resentment of difference. Just as the hate movement has piggybacked on the reactionary politics of the right, so too might a progressive movement exploit the windows opening up within more progressive parties.

Giving oppressed communities an opportunity to have a voice in such conversations is a first step in realizing social justice. Social justice revolves around participation and democratic representation in the home, in the workplace, and in political arenas. In other words, it consists of the ability to "do difference" without fear of violent reprisal. It frees women to make household decisions without fear of being beaten; it frees people of color to pass through or live in any neighborhood without fear of attack; it frees gay men and lesbians to demand equal treatment and recognition without fear of violence.

Related Entries

- ▶ Communities and the Police
- ▶ Ethnic Violence
- ▶ Genocide
- ▶ Impacts of Community-Oriented Policing
- ▶ Victimization, Gender, and the Criminal Justice System

Recommended Reading and References

Bell J (2009) Policing and surveillance. In: Lawrence F (ed) Responding to hate crime. Praeger, New York, pp 31–50

Berrill K (1992) Anti-gay violence and victimization in the United States: an overview. In: Herek G, Berrill K (eds) Hate crimes: confronting violence against lesbians and gay men. Sage, Newbury Park, pp 19–45

Comstock G (1991) Violence against lesbians and gay men. Columbia University Press, New York

Crenshaw K (1994) Mapping the margins: intersectionality, identity and violence against women of color. In: Albertson Fineman M, Mykitiuk R (eds) The public nature of private violence. Routledge, New York, pp 93–118

Fisher C, Salfati CG (2009) Behavior or motivation: typologies of hate-motivated Offenders. In: Barbara P, Randy B (eds) Hate crime series, Hate crime offenders. Praeger Perspectives, vol 4. pp 103–136

Franklin K (2002) Good intentions: the enforcement of hate crime penalty-enhancement statutes. Am Behav Sci 46:154–172

Gadd D, Dixon B (2011) Losing the race: thinking psychosocially about racially motivated crime. Karnac, London

Hamm M (1994) Hate crime: international perspectives on causes and control. Anderson, Cincinnati

Hate Crimes Community Working Group (2006) Addressing hate crime in Ontario, Attorney General and Minister of Community Safety and Correctional Services, Toronto

Herek G, Berrill K (eds) (1992) Hate crimes: confronting violence against lesbians and gay men. Sage, Newbury Park

Iganski P, Smith D (2011) Rehabilitation of hate crime offenders. Report to the Equality and Human Rights Commission, Scotland

International Network Against Cyber Hate (2005) Annual Report, INACH, Amsterdam

Jacobs J, Potter K (1998) Hate crime: criminal Law and identity politics. Oxford University Press, Oxford

Jaggar A (2005) What is terrorism, why is it wrong, an could it ever be morally permissible? J Soc Psychol 36(2):202–217

Kaplan J, Bjorgo T (eds) (1998) Nation and race: the developing Euro-American racist subculture. Northeastern University Press, Boston

Kelly R (1991) Bias crime: American legal responses, 2nd edn. Office of International Criminal Justice, Chicago

Kelly R, Maughan J (eds) (1998) Hate crime: the global politics of polarization. Southern Illinois University Press, Carbondale

Lane FJ, Shaw LR, Kim M (2009) Hate crimes committed against persons with disabilities. In: Perry B (ed) Victims of hate crime. Praeger, Westport, pp 173–198

Levin J, McDevitt J (1993) Hate crimes: the rising tide of bigotry and bloodshed. Plenum, New York

Lim HA (2009) Beyond the immediate victim: understanding hate crimes as message crimes. In: Iganski P (ed) Hate crimes: the consequences of hate crime. Praeger, Westport, pp 107–122

Mason-Bisch H (2010) Future challenges for hate crime policy: lessons from the past. In: Chakraborti N (ed) Hate crime: concepts, policy, future directions. Willan, Cullompton, pp 58–77

McDevitt J, Balboni J, Garcia L, Gu J (2001) Consequences for victims: a comparison of bias- and non-bias motivated assaults. Am Behav Sci 45(4):697–711

Meyer D (2010) Evaluating the severity of hate-motivated violence: intersectional differences among LGBT hate crime victims. Sociology 44(5):980–995

Noelle M (2002) The ripple of effect of the Matthew Shepard murder: impact on the assumptive worlds of members of the targeted group. Am Behav Sci 46(1):27–50

Office for Security and Cooperation Europe (2009) Hate crime laws: a practical guide. OSCE, Warsaw

Ontario Human Rights Commission (2007) Fishing without fear: report on the inquiry into assaults on Asian Canadian anglers. Ontario Human Rights Commission, Toronto

Perry B (2001) In the name of hate. Routledge, New York

Perry B (2009) The continuum of fear: hate crime/terrorism. Justice Canada discussion paper. Justice Canada, Ottawa

Perry B, Alvi S (2011) "We are all vulnerable:" the *in terrorem* effects of hate crime. Int Rev Victim 18(1):57–72

Poynting S, Noble G (2004) Living with racism: the experience and reporting by Arab and Muslim Australians of discrimination, abuse and violence since 11 September 2001. Report to the Human Rights and Equal Opportunity Commission. Published at: http://www.humanrights.gov.au/racial_discrimination/isma/research/index.html

Rosga A (1999) Policing the state. Georgetown J Gender Law 1:145–171

Shelley C (2008) Transpeople: repudiation, trauma, healing. University of Toronto Press, Toronto

Sherry M (2010) Disability hate crimes: does anyone really hate disabled people? Ashgate, Surry

Shively M (2005) Study of literature and legislation on hate crime in America. National Institute of Justice, Washington, DC

Sibbit R (1997) The perpetrators of racial harassment and racial violence. Home office research study. Home Office, London, p 176

Weinstein J (1992) First amendment challenges to hate crime legislation: where's the speech? Crim Justice Ethics 11(2):6–20

Welch M (2006) Scapegoats of September 11th: hate crimes and state crimes in the war on terror. Rutgers University Press, New Brunswick

White R, Gerke D (2007) Gendered experiences of sexuality-based discrimination in Jamaica. In: Badgett M (ed) Sexual orientation discrimination: an international perspective. Routledge, New York, pp 153–163

Wolfe L, Copeland L (1994) Violence against women as bias-motivated hate crime: defining the issues in the USA. In: Davies M (ed) Women and violence. Zed Books, London, pp 200–213

Health and Safety Law

► Criminalization and Occupational Health and Safety

Health and Wellness

► Officer Safety, Health, and Wellness

Health Care Reform and Criminal Offenders (US)

Alison Evans Cuellar[1] and Jehanzeb Cheema[2]
[1]George Mason University, Fairfax, VA, USA
[2]University of Illinois at Urbana-Champaign, Urbana, IL, USA

Synonyms

Affordable Care Act

Overview

The 2010 Patient Protection and Affordable Care Act (ACA), health care reform, comes at a particularly unique time in the criminal justice system – a time when mass incarceration policies are being scaled back and more prisoners are being released to the community. The ACA attempts to tackle weaknesses in the existing health insurance system and in the treatment delivery system through a variety of mechanisms. These changes could dramatically improve the health and justice outcomes of over five million US adults in the justice system.

The relative poor health of jail and prison inmates is well documented. Inmates have high rates of many chronic medical conditions, mental health disorders, and substance abuse disorders (Wilper et al. 2009). In addition, their rates of hypertension, asthma, and arthritis are higher than the general population even after controlling for race, education, and other socioeconomic differences (Binswanger et al. 2007, 2009), and rates of infectious diseases including current or chronic hepatitis C and HIV/AIDS are also high (National Commission on Correctional Health Care 2002; Maruschak 2010). Finally, sexual victimization rates in jail are higher than reported in community (Beck and Harrison 2008).

Justice-involved individuals are less likely than others to have insurance coverage. Working-age men in general have high rates of uninsurance. And, offenders' health status and constrained employment prospects after conviction and incarceration make obtaining health insurance less likely for them than for other adults.

Because they have relatively high rates of substance abuse disorder and low rates of insurance, offenders must rely on a part of the health care delivery system that is reliant on scarce and shrinking public financial support (Buck 2011). Offender surveys have found that addiction treatment services are underutilized; few community justice agencies integrated their services with substance abuse providers; and the quality of community treatment is frequently quality (Taxman et al. 2007).

This entry provides an overview of health insurance and challenges in securing coverage for offenders today. It then addresses the Affordable Care Act, including how prisoners and jail inmates are likely to be affected by the new federal health care reform law, as well as potential future impacts, challenges, and opportunities.

Fundamentals

Having health insurance means treatment will be less costly to individuals when they seek health care. Thus, insurance removes a financial barrier to care. Numerous studies have found that having health insurance is associated with obtaining more health services and having better health outcomes (Manning et al. 1987; Card et al. 2008). However, health insurance coverage can present a challenge for the justice population. For working-age adults in the United States, the primary source of health insurance is through employer-provided insurance offered to full-time employees (DeNavas-Walt et al. 2012). Very few nonelderly adults purchase coverage on their own. The justice population, particularly those reentering the community, may be disconnected from the workforce and need time to find housing, retool, and secure jobs. If their work is part-time, temporary, or low pay, then employer-based health coverage may not be available or affordable.

Medicaid is another important source of health insurance coverage for adults, primarily for adults who have disabilities and who have low incomes. In 40 states, poor, disabled individuals who receive Supplemental Security Income (SSI) also receive Medicaid. Medicaid has not traditionally covered childless adults who do not have disabilities. Only in 12 states does Medicaid coverage include poor adults, regardless of SSI. However, individuals who enter prison or jail may have their SSI benefits suspended, meaning they lose their Medicaid coverage as well. Both income and health benefits must be reinstated upon reentry into the community. Lacking health insurance criminal justice populations have traditionally relied on a web of safety net providers to meet their complex health care needs.

Health Benefits Covered by Insurance Today. Private insurance typically covers a range of medical services, but often with a share of the cost borne by the patient through deductibles, meaning the amount paid by the member before insurance begins, or through cost-sharing, meaning the amount paid by the patient when treatment is sought. Newer, so-called high deductible heath plans with a deductible of at least $1,000 ($2,000 for a family) have become more common in the United States and included 13 % of the privately insured in 2011 (Fronstin 2011).

Coverage for behavioral health care is often less generous for behavioral health than for general medical care, a consequence of longstanding concerns about the overuse of mental health services (Glied and Cuellar 2006). Consequently, behavioral health coverage traditionally has required greater financial contribution on the part of the patient and has placed limits on how much treatment would be covered. Efforts over many years to place mental and medical benefits on more equal footing ultimately lead to the federal Mental Health Parity and Addiction Equity Act of 2008. This Act required that mental health and substance abuse care be covered like other medical care, at least in large group health insurance plans. Under parity mental health and substance use, disorder coverage – if such benefits are offered – must be equivalent along co-payments and deductibles, maximum visits, annual and lifetime dollar limits, and out-of-network coverage. However, the law does not require that health insurance include any behavioral health benefits, only that they be equivalent if offered. The law also does not affect Medicaid unless Medicaid is offered to beneficiaries through privately managed care plans.

Research has found parity legislation to increase use of mental health and substance abuse services by removing significant financial barriers (Frank et al. 2001; Goldman et al. 2006; Dave and Mukerjee 2008). In light of high rates of mental health and substance abuse disorder, this is a potentially very important development for offenders.

Medicaid play s a unique role in coverage for low-income individuals with complex chronic conditions, as it covers many services that private insurance does not. Medicaid covers services such as targeted case management, personal care, and wrap-around services that are important for people with complex chronic conditions, such as HIV/AIDS or severe mental illness (Kaiser Family Foundation 2011). The same is not true for substance abuse coverage where coverage under Medicaid is frequently limited in amount and by type of provider (Kaiser Family Foundation 2011).

One particular challenge for Medicaid beneficiaries is the cumbersome application and renewal process. For inmates in particular, this process imposes onerous documentation, ongoing renewal, and identification requirements. Inmates who receive Medicaid by virtue of their SSI benefits have their SSI benefits are suspended while they are incarcerated or even in halfway houses that are under the auspices of the department of corrections. Additional restrictions apply to individuals with outstanding felony warrants or violators of parole or probation (Cuellar and Cheema 2012). Individuals who are incarcerated for longer than a year have a greater task, as they must reestablish the existence of their disability and wait for SSI determinations before accessing Medicaid.

Health Care Delivery Today. The justice system has the largest concentration of people with substance abuse disorders, and 50 % of those with substance abuse disorders have co-occurring mental health conditions (Taxman et al. 2007). Yet, the general substance abuse treatment system is arguably less robust than other parts of the delivery system, including mental health, in part because funding from insurance and Medicaid has been lacking. Of the 20 million individuals with alcohol and drug abuse disorders, only 10 % received treatment in the past year (Center for Behavioral Health Statistics and Quality, 2012). Among those not treated, 44 percent cited lack of insurance and cost as impediments to care and one quarter of the insured reported lack of coverage or affordability.

The general substance abuse delivery system is characterized by many small providers who are frequently understaffed. One third of addiction treatment settings do not contract with a physician or have one on staff and fewer than a half employ masters-level counselors or above (Buck 2011). Furthermore, substance abuse treatment is not well integrated with other medical care, such as HIV (Altice et al. 2011; Berg et al. 2011). Only 30 % of community substance abuse treatment programs offer HIV treatment and counseling. Lack of insurance coverage, staff training, and patient acceptance have been cited as barriers to the integration of HIV care into substance abuse treatment programs (Bini et al. 2011).

In contrast, mental health delivery has benefitted from expanded insurance coverage through Medicaid coverage of people with disabilities, insurance parity, and technology changes. Unlike substance abuse disorders, mental health disorders of sufficient severity are qualifying impairments under the SSI program which conveys eligibility for Medicaid in many states. This coverage, combined with state decisions to cover evidence-based mental health treatments, has led to greater insurance-based financing for mental health care.

Access to care by specialists remains a challenge in many communities (Cunningham 2009). The 2008 Community Tracking Survey of primary care physicians, e.g., family practice or internal medicine doctors, found that 42 % had trouble accessing high-quality specialty care for their patients and 59 % were unable to obtain specialty mental health services they thought were medically necessary for their patients. The relatively lack of mental health specialists, paired with innovations in pharmaceuticals and screening tools for common mental disorders, has led to more mental health care being provided in primary care settings than ever (Frank and Glied 2006).

Changes Under Health Care Reform. Health care reform, specifically, the Patient Protection and Affordable Care Act (ACA), has the potential to improve care for offenders, particularly those who live in the community, in several ways: (1) Most individuals must obtain health insurance. (2) Medicaid will be expanded significantly, particularly for childless adults who have income below 138 % of the federal poverty level. (3) Insurance exchanges will be created through which individuals can receive federal subsidies to purchase private coverage. (4) Numerous initiatives are created to address fragmented care for individuals with chronic conditions.

The ACA requires nearly every resident of the United States to obtain health insurance by January 1, 2014. Before the Supreme Court decision on the constitutionality of the ACA, the Congressional Budget Office estimated there would be 30–33 million newly insured individuals, with 17 million newly covered through Medicaid and 23 million covered through new health insurance exchanges, and a small reduction in employer-provided coverage (Congressional Budget Office, March 2012). These initial estimates assumed all states would expand Medicaid. However, the Supreme Court decision in National Federation of Independent Business v. Sebelius, 132S. Ct. 2566 (2012) held that states could choose whether or not to expand eligibility for coverage under their Medicaid program pursuant to the ACA and, despite strong financial incentives, not all states likely will elect to expand (Congressional Budget Office, July 2012). Further, some states may delay expansion until after 2014 and some may seek federal approval for partial expansions. As a result, the Congressional Budget Office scaled back its estimates concluding that seven million fewer individuals will be in states that expand Medicaid coverage by 2016 and four million more will be uninsured relative to prior estimates.

The Medicaid expansion is particularly important to criminal justice populations. Medicaid will become the primary source of health insurance for most low-income individuals, including childless adults. Medicaid is a shared federal-state financial responsibility, and states that expand coverage will pay proportionately less for the newly eligible Medicaid beneficiaries. The fact that states may not expand Medicaid will particularly affect criminal justice populations as they overwhelmingly represent low-income men. And, in those states where Medicaid is not expanded, there will be less financial support to bolster a treatment system that is fragmented and where low-income individuals frequently cannot find care when it is needed.

Cuellar and Cheema (2012) estimated that 21–34 % of prisoners released each year would be enrolled in Medicaid. Applying a similar methodology for jail inmates implies that a total of 2.4–3.6 million prison and jail inmates who are released would be eligible for Medicaid. The estimates are based on the average incomes of individuals with demographics similar to those of released prison and jail inmates and take into account a possible wage penalty from having a criminal record and the fact that some inmates will not have lived in the United States sufficiently long to qualify for benefits. In light of the more recent Supreme Court decision, these estimates should be scaled back by as much as 60 %.

Historically, states have found that fewer people actually enroll in Medicaid than are eligible, as few as 50–75 % of eligibles. This makes health care reform an opportunity only. It must be matched with substantial efforts to reach out and enroll individuals who are newly eligible for Medicaid and involved in the criminal justice system. To date, such outreach has not been particularly salient because so few nondisabled individuals met eligibility requirements. Going forward, outreach will require a new type of collaboration between criminal justice and Medicaid.

The criminal justice population that will be eligible for Medicaid is only a fraction of the estimated ten million newly covered Medicaid beneficiaries. However, their high health care needs and complex social conditions could pose new challenges for Medicaid health plans. Outreach, enrollment and education of providers, and health plans are key steps to improving health for the criminal justice population under reform.

Individuals not eligible for Medicaid can receive assistance obtaining health insurance through new health insurance exchanges under the ACA. The new health insurance exchanges will offer subsidies, on a sliding scale, to those with income up to 400 % of the federal poverty level who buy private coverage. The amount of the subsidy will depend on the enrollee's income relative to the federal poverty level and the cost of local plans. Cuellar and Cheema (2012) estimate that 50–54 % of released offenders would be eligible for a subsidy, based on their income if plan premiums are sufficiently high. In light of the Supreme Court decision, these estimates should be increased slightly.

Systematic outreach and simplified enrollment are emphasized under health reform, both in Medicaid and in exchanges. Federal assistance is available for states toward the design and implementation of the necessary enrollment technology. All states must engage in consumer information and outreach and related activities must be reported to the federal government. From the perspective of the criminal justice population, a strong enrollment system is one that can incorporate relevant corrections information when determining eligibility. Eligibility is determined based on vital, employment and income information and, therefore, correctional data can fill important data gaps for inmates. Inmates might also be handed key documents, including incarceration dates, to ease income verification and identification. Correctional facilities could also communicate release dates to eligibility offices. To smooth the transition from prison to post-release health care, community providers could be identified in advance. However, in some instances these steps may not be feasible as the average jail has a weekly turnover rate of over 60 %. Nonetheless, states can involve corrections, probation, and parole representatives in system design and include them in subsequent training programs.

Whereas health insurance today can vary considerably in terms of covered services and coverage limits, under health care reform the newly insured population will have insurance that covers "essential health benefits." The essential benefits must include mental health and substance abuse disorder treatment, as well as preventive and wellness services and chronic disease management, and behavioral health parity provisions will apply.

An increase in funding for behavioral health services is expected under health care reform, stemming from private insurance and Medicaid coverage. Providers of behavioral health services, including those who deliver care to individuals at risk of incarceration, must adapt to an environment where funds are tied to insurance, rather than grants from states or the federal government. This means greater adoption of financial information systems, professional credentialing to meet insurers' network requirements, and the ability to track the amount and quality of care delivered.

Over and above expanding health insurance, the health care reform law provides incentives to change how treatment providers coordinate care. Several provisions in the ACA promote new care delivery models for individuals with chronic conditions, including medical or health homes models. These models have enormous potential. Medical homes receive financial support specifically to manage chronic conditions and coordinate care across a range of specialists. For criminal justice populations, these medical homes will need strong ties to mental health and substance abuse treatment providers. Ideally, these medical homes will coordinate with corrections and probation departments around benefit enrollment, continuity of treatment from prison to community treatment, and have a basic understanding of probation terms and court orders.

Significant changes are likely for addiction treatment providers. Some anticipate that in the wake of the ACA, care delivery for substance abuse will shift away from traditional substance abuse providers (Buck 2011). Because the ACA emphasizes and rewards primary care–based models, more care may shift to primary care. This offers the promise that both the medical and behavioral health needs of offenders could be more comprehensively addressed and coordinated. Whether care improves will depend in part on how rapidly evidence-based care can be delivered in the traditional medical sector for this population. Another delivery model is to have the traditional behavioral health system integrate medical and chronic disease care. The success of either approach has yet to be determined on a wide scale, but considerable investment is being made, financial incentives are in place, and the intuitive appeal is strong.

Much of the burden of enrolling newly insured individuals through Medicaid and through health exchanges will fall on states. States face the practical challenges of orchestrating health insurance coverage for 30 million individuals, a fraction of whom are in criminal justice and have not only complex eligibility circumstances, but also complex health needs. Actively assisting inmates to enroll in health plans will remove financial barriers to treatment. Developing medical and health home models that can address offenders' complex health and behavioral health could be beneficial to both public health and public safety.

Recommended Reading and References

Altice FL, Bruce RD, Lucas GM et al (2011) HIV treatment outcomes among HIV-infected, opioid – dependent patients receiving buprenorphine/naloxone treatment within HIV clinical care settings: results from a multi-site study. J Acq Immun Def Synd 56(Suppl 1):S22–S32

Beck AJ, Harrison PM (2008) Sexual victimization in jails reported by inmates, 2007. NCJ-221946. U.S. Department of Justice, Office of Justice Programs, Washington, DC

Berg KM, Litwin A, Li Z et al (2011) Directly observed anti-retroviral therapy improves adherence and viral load in drug users attending methadone maintenance clinics: a randomized controlled trial. Drug Alcohol Depend 113(2–3):192–199

Bini EJ, Kritz S, Brown LS et al (2011) Barriers to providing health services for HIV/AIDS, hepatitis C virus infection, and sexually transmitted infections in substance abuse treatment programs in the United States. J Addict Disorder 30(2):98–109

Binswanger IA, Stern MF, Deyo RA, Heagerty PJ, Cheadle A, Elmore JG et al (2007) Release from prison – a high risk of death for former inmates. N Engl J Med 356:157–165

Binswanger IA, Krueger PM, Steiner JF (2009) Prevalence of chronic medical conditions among jail and prison inmates in the USA compared with the general population. J Epidemiol Community Health 63:912–919

Buck JA (2011) The looming expansion and transformation of public substance abuse treatment under the affordable care act. Health Aff 30:1402–1410

Card D, Carlos D, Nicole M (2008) The impact of nearly universal insurance coverage on health care utilization: evidence from medicare. American economic review 98(5):2242–58

Center for Behavioral Health Statistics and Quality (2012) National survey on drug use and health: detailed tables. Substance abuse and mental health services administration, U.S. department of health and human services Rockville, MD

Congressional Budget Office (2012a) Updated estimates for the insurance coverage provisions of the affordable care act. CBO, Washington, DC

Congressional Budget Office (2012b) Estimates for the insurance coverage provisions of the affordable care act updated for the recent Supreme Court decision. CBO, Washington, DC

Cuellar A, Cheema J (2012) As roughly 700,000 prisoners are released annually, about half will gain health coverage and care under federal laws. Health Aff 31:931–938

Cunningham P (2009) Primary care physicians' perspectives on access to mental health care. Health Affairs 28 (3):w490–w501

Dave D, Mukerjee D (2011) Mental health parity legislation, cost-sharing and substance abuse treatment admissions. Health Econ 20(2):161–183

DeNavas-Walt C, Proctor BD, Smith JA (2012) Income, poverty, and health insurance coverage in the United States: 2011 current population reports. U.S. Census Bureau, U.S. Department of Commerce, Washington, DC, pp 60–243

Frank RG, Glied SA (2006) Better but Not Well: mental health policy in the United States since 1950 Baltimore, Johns Hopkins University Press

Frank RG, Goldman HH, McGuire TG (2001) Will parity in coverage result in better mental health care? N Eng J Med 345(23):1701–1704

Friedmann PD, Zhang Z, Hendrickson J, Stein MD, Gerstein DR (2003) Effect of primary medical care on addiction and medical severity in substance abuse treatment programs. J Gen Intern Med 18(1):1–8

Fronstin P (2011) Findings from the 2011 EBRI/MGA consumer engagement in health care survey EBRI (Issue Brief No 365). Employee Benefit Research Institute, Washington DC

Glied S, Cuellar A (2006) Better behavioral health care coverage for everyone. N Eng J Med 354:1415–1417

Goldman HH et al (2006) Behavioral health insurance parity for federal employees. N Eng J Med 354(13):1378–1386

Kaiser Family Foundation (2011) Medicaid benefits: online database. [Internet] Available from at Medicaidbenefits.kff.org/index.jsp. 22 Apr 2011

Manning WG, Newhouse JP, Duan N, Keller EB, Leibowitz A, Marquis MS (1987) Health insurance and the demand for medical care: evidence from a randomized experiment. Am Econ Rev 77:251–277

Maruschak LM (2010) HIV in prisons, 2007–08. U.S. Department of Justice, Office of Justice Program, Bureau of Justice Statistics, Bureau of Justice Statistics Bulletin, Washington, DC

National Commission on Correctional Health Care (2002) The health status of soon-to-be-released inmates. National Commission on Correctional Health Care, Washington, DC

Taxman FS, Perdoni ML, Harrison LD (2007) Drug treatment services for adult offenders: the state of the state. J Subst Abuse Treat 32(3):239–254

Wilper AP, Woolhandler S, Boyd JW, Lasser KE, McCormick D, Bor DH et al (2009) The health and health care of US prisoners: results of a nationwide survey. Am J Public Health 99(4):666–672

Health Issues in Prison Reentry Models

Dora M. Dumont[1], Landon B. Kuester[2] and Josiah D. Rich[1]
[1]The Center for Prisoner Health and Human Rights, The Miriam Hospital/Brown University, Providence, RI, USA
[2]Centre for Research on Drugs and Health Behavior, London School of Hygiene and Tropical Medicine, London, UK

Overview

Incarceration has been used all too often as a solution for shortcomings in health and other social services. Medical and public health researchers are among those recommending that correctional facilities be used as a venue to improve individual and population health, since correctional facilities serve as health-care providers of last resort for the medically underserved (Freudenberg 2001; Rich et al. 2011). Reentry likewise constitutes a public health opportunity that is both often lost and problematic. Prison and jail inmates confront a range of challenges upon reentry into the community. The incarcerated carry a higher disease burden than the general population, and their health status and health care upon reentry have significant repercussions not only for their own post-incarceration lives but for public health as well. Correctional facilities can both serve as sentinels for the health of a disadvantaged population and implement care and transition services that ensure the continuity of medical care from prison into the community upon release, but they often fail to do so. There are complex challenges to the successful implementation of discharge planning and reintegration into society, but capitalizing upon such services to address mental health, substance dependence, and other medical needs has important consequences for both individual and community health.

Health Profile of the Incarcerated

Incarceration is concentrated in low-income communities, and health is strongly correlated with socioeconomic status. In addition, risk factors for incarceration and for poor health frequently overlap. As a result, inmates tend to have far more health problems than the general public. The United Nations and other international reviewers have documented that prisons and detention facilities in many world regions are not in compliance with international standards like the Standard Minimum Rules for the Treatment of Prisoners, established by the UN High Commissioner for Human Rights (UNHCR). Facilities are often overcrowded and unsanitary and thereby may have additional adverse health effects for an already ill population (Jurgens et al. 2011).

Public health practitioners have long been concerned with correctional facilities as repositories of infectious diseases, especially tuberculosis, HIV, and hepatitis C virus (HCV) (Hammett 2006). Despite gains since the 1990s, HIV prevalence in the USA remains about five times higher in state/federal correctional facilities than in the general public: 1.6 % among male inmates and 2.4 % among female inmates, compared with 0.36 % in the total US population for 2006 (Hammett 2006). Internationally, studies have found wide estimate ranges for HIV prevalence in prisons, especially in the former Soviet republics and Southeast Asia. For instance, a 2007 review found between 0 % and 26 % HIV prevalence in Ukraine and 4–22 % in Indonesia, reflecting the difficulties of establishing an accurate health profile even within individual countries. HIV prevalence could not be identified at all for 76 of 152 low-income or transitioning countries (Dolan et al. 2007).

Chronic diseases among incarcerated populations are gaining attention, given the simultaneous emergence of chronic conditions like diabetes among younger people and the aging of the incarcerated population. In the USA, between 39 % and 43 % of inmates are diagnosed with diabetes, hypertension, asthma, or another chronic condition, rates consistently higher than among the general population (Binswanger et al. 2009; Wilper et al. 2009). Moreover, comorbidity (i.e., the coexistence of two or more medical conditions) is high in this population. The risk behaviors for HIV – especially injection drug use (IDU) – put the same individuals at high risk for HCV, which is 9–10 times more prevalent among the incarcerated than among the non-incarcerated (Freudenberg 2001). Because patients coinfected with both HIV and HCV have also been found to have more comorbidities than HIV mono-infected patients, these inmates are likely to carry a greater overall burden of coexisting chronic and/or infectious diseases.

These physical conditions are frequently complicated by the mental health profile of the incarcerated. The emergence of prisons and jails as the largest institutions in the United States housing the mentally ill reflects the de facto criminalization of mental illness. Well over half of inmates at any given time have a DSM-IV (Diagnostic and Statistical Manual of Mental Disorders) mental disorder. In US jails, 63 % of blacks and 71 % of whites in a 2002 national survey self-reported symptoms or diagnoses of mental illness, with rates slightly lower in state and federal prisons (James and Glaze 2006). Women, who constitute a small minority of inmates, generally have much higher rates of mental illness, and posttraumatic stress disorder (PTSD) is common among incarcerated women, about one-third of whom were subject to physical abuse and one-third to sexual abuse prior to incarceration (Lewis 2006; Baillargeon et al. 2010). Because of differing research methodologies, it is difficult to determine whether the significantly higher rates of mental illness among US female inmates reflect a distinct US experience or whether they merely reflect different data collection practices. A systematic review of studies in 12 western nations also found disproportionately high rates of mental illness among inmates. Rates of personality and antisocial disorders were lower among women than among male inmates but were similar for women and men for psychotic illnesses and major depression (Brooker et al. 2007). Rates for other world regions are difficult to obtain, in part because both medical exams and surveys are not consistently conducted.

An important component of prisoners' mental health profile is substance dependence and addiction. Estimates of the number of the incarcerated meeting DSM-IV criteria for drug dependence or addiction vary widely but are well above 50 % and again substantially higher among female inmates (Freudenberg 2001; James and Glaze 2006; Binswanger et al. 2009). Comorbidity of substance abuse and mental illness is common among inmates. Addictions and other mental illnesses both increase the probability of engaging in behaviors associated with infectious disease transmission and complicate treatment for conditions such as HIV and HCV, which require a sustained adherence regime and careful monitoring for adverse drug interactions.

While incarceration may improve health conditions for some (e.g., by providing access to health care, stable shelter, and regular meals), it can worsen health outcomes in other respects. This may be the case especially with mental illness and addiction. Those with mental health issues are more vulnerable to placement in isolation, an environment particularly hazardous to mental health. Most facilities adopt a "cold turkey" approach to substance addiction, leaving the addicted subject to withdrawal during incarceration and more vulnerable to overdose upon release. Although drug use is technically forbidden in nearly all facilities, it remains common and the shared use of needles also contributes to the transmission of infectious diseases, especially HCV. Needle and syringe exchange programs exist in several European and Asian countries, but few facilities to date participate (Jurgens et al. 2011).

A number of national and international mandates exist regarding the provision of health care to inmates. In the United States, two key Supreme Court decisions have upheld the finding that correctional facilities must provide health care comparable to what might be reasonably attained in society. In 1976, the *Estelle v. Gamble* decision declared that the deliberate failure to provide health care constituted unconstitutional cruel and unusual punishment. Since *Estelle v. Gamble*, litigation or the threat thereof has greatly improved correctional health care, and in 2011, *Brown v. Plata* ordered the state of California to reduce overcrowding on the grounds of its interference with the due provision of health care. Despite these legal advances, a lack of standardized reporting and quality measures has made it difficult to know the quantity or quality of health care actually provided on a national basis. However, studies have indicated that many inmates in need of health care do not receive it or receive suboptimal care, whether for HIV, HCV, chronic conditions such as diabetes, mental illness, or substance abuse. There are some correctional facilities that are recognized as public health models of successful diagnosis and treatment of inmate health conditions. For the most part, though, there is still room for improvement in prison health services. Jails and pretrial detention centers diagnose and treat even smaller percentages of inmates, in part because of higher turnover and shorter stays. Since 95 % of inmates are eventually released, the inmate health profile follows them to reentry, with considerable implications for both their own reentry experience and community health.

Transitioning to the Community

Incarceration is a temporary situation for most inmates, and as recidivism rates are high, many people cycle repeatedly between the community and detention. In the USA, over seven million people leave jail and an additional 600,000 exit state and federal prisons annually (US Bureau of Justice Statistics 2003). They are largely a high-need, low-resource demographic: in addition to more than usual health needs that may interfere with daily functioning, as a group they have less education and fewer economic and social resources than the general population. Releasees also frequently return to the community in highly unstable circumstances, struggling to reestablish housing, employment, and relationships. In this context, attention to health needs often falls by the wayside, and high-risk behaviors such as unsafe sex and relapse to drug use are common. The overwhelming challenges of reentry are evident in mortality trends. Data in the USA and other nations have documented a spike in mortality in the first 2 weeks after release: in the USA, releasees were 12.7 times more likely to die in that period than the general public and 129 times more likely to die of an overdose (Binswanger et al. 2007). Similar postrelease mortality rates have been identified in other countries.

Correctional authorities are increasingly addressing this problem with discharge planning, a term that broadly refers to the process of helping prisoners to prepare to make the transition from incarceration back into the community, but the resources allocated to transition services have been steadily reduced in the last 20 years even as the correctional population has increased fourfold (Re-entry Policy Council, 2005). As a result, only about 10 % of releasees from US state prisons in need of discharge planning actually receive it, far less than was the case in earlier decades

(Mellow and Greifinger 2005). In general, the high turnover rates and short stays of jail inmates prevent discharge planning in jails altogether.

To date, there has not been any large-scale research examining the patterns of implementation or results of discharge planning, either across the USA or internationally. However, collaborative work focusing on HIV provides one basis for such research. A multicenter, Health Resources and Services Administration (HRSA)-sponsored Special Project of National Significance ("Enhancing Linkages to HIV Primary Care and Services in Jail Settings Initiative") was established in 2007 to test and evaluate models for linkage to care for people with HIV leaving jail. One linkage project under way makes use of new federal reporting requirements to link correctional and clinical data, thereby constructing an empirical basis for evaluating why some linkage succeeds and some fails. This HIV-specific work may be creating paradigms that can be used to evaluate the transition from correctional health care to community-based health care for other health conditions as well. It is clear that to achieve successful linkage to community-based health care, discharge planning must extend beyond the provision of information and even initial free services. A widely noted study in Texas found that only 5 % of HIV-positive releasees filled free prescriptions in time to prevent an interruption of treatment, suggesting the critical role of factors beyond the ability to pay (Baillargeon et al. 2009).

Discharge planners increasingly understand the need to view health within a broader spectrum of reentry needs. Because inmates are so often undereducated and lacking both financial and social resources, discharge planning is often unsuccessful unless it provides sustained, holistic assistance. On the one hand, untreated medical conditions will reduce the probability of successful reintegration into the community; on the other, the failure to address other reentry needs (especially employment and stable housing) significantly undercuts the probability of resolving health needs. Thus, medical discharge planning must occur in coordination with other transitional services targeting complex social needs.

Numerous guidelines recommend that discharge planning requires an ongoing, two-stage approach in order to be effective. Ideally, the process should begin shortly after admission into the correctional facility, though in general it is set in motion only a few weeks or months before the inmate's anticipated release. There is broad consensus among guidelines that the second stage should proceed to ongoing case management in the community upon reentry (Re-entry Policy Council, 2005). Randomized controlled trials have not yet confirmed the benefit of case management service; however, studies found an association between case management and positive health or behavioral outcomes (Springer et al. 2011). As a result, discharge planning requires the kind of sustained cross-agency collaboration that is frequently difficult for state and community agencies to coordinate and maintain. Only in rare cases do state health departments have any jurisdiction over correctional health care, so discharge planning requires Departments of Correction (or their non-US counterparts such as Ministries of Justice) to reach out to partners they are not ordinarily organized to work with.

Despite these organizational challenges, there are programs that provide more successful discharge planning results for health care and health behaviors, and there are examples of discharge planning being addressed by collaborations between corrections officials, cross-agency coalitions, and local community agencies. Many of the existing programs aim to build multi-sectorial collaboration with community-based organizations in order to prepare individuals for release and provide them with health care within a spectrum of other resources such as housing, employment, and social benefits. One such program is the Hampden County Jail in Massachusetts, which adopted a public health paradigm for its inmate health-care services (Conklin et al. 2002). The county sheriff's office partnered with the public health department, regional medical centers, and community health centers to provide continuous health-care delivery by assigning new inmates to care based on their zip code, in order to facilitate continuous care delivery by linking inmates to community providers prior to their release. Staff from community health centers provide health services inside the facility during incarceration and develop individualized discharge plans before release. Program evaluations are incomplete but indicate improved inmate and community health, reduced recidivism, and cost savings. Even in this closely coordinated system incorporating community providers into prerelease correctional care, however, releasees are frequently lost to follow-up care.

In light of generally insufficient resources, even where correctional authorities are concerned to support discharge planning, such services are often only provided for inmates with dedicated resources such as those with HIV or severe mental illness. HIV-centered reentry programs have been implemented in a number of states. Transitional services may be facilitated by two mediators: a small geographic range to work within and a close-knit, locally respected health professional community that works effectively with corrections authorities. (Thus, guidelines and best practices for discharge planners often warn not to "lift" programs from other communities, since what works in one setting may not work in another without adequately taking account of local circumstances.) In the Community Partnerships and Support Services for HIV-Infected People Leaving Jail (COMPASS) program, for instance, the Rhode Island Department of Corrections agreed to have jail staff notify community outreach workers when inmates with HIV were admitted or when inmates were newly diagnosed. Outreach workers then met with inmates in the jail to assess their postrelease needs and provided sustained case management upon reentry (Nunn et al. 2010).

Qualitative research with COMPASS care recipients revealed another complicating feature of transition services: their success relies to a great extent upon sustained personal relations. Successful initiation of community-based care upon reentry and ongoing treatment adherence was strongly associated with intensive case management, including home calls to remind participants to keep doctor's appointments, providing transportation to medical and other services, and assisting with processes like registering for Social Security or disability benefits. As the smallest state in the USA, RI has a highly centralized and compact service area and close ties among professional communities, advantages that may be difficult to replicate elsewhere. Participants also highlighted the importance of having an engaged and perceptibly caring doctor, the absence of which was identified as a barrier to treatment adherence in other studies (Nunn et al. 2010).

Because patients with HIV are six times more likely than those with other chronic conditions to identify a regular source of care after discharge, non-HIV postrelease continuity of care may be even worse. This is particularly relevant for those with addictions and mental illnesses who received insufficient treatment during incarceration, many of whom are incarcerated in the first place because community-based treatment was not available to them. Although more likely to receive discharge planning (66 % of state prisons provided some form of postrelease support to access community-based mental health services in 2000 (Baillargeon et al. 2010)), the proportion of releasees receiving needed mental health services is further reduced by incomplete identification of mental health needs among inmates prior to release. A 2008 study in Texas and Ohio found that self-reported mental illness was about one-third more common postrelease than during incarceration, which the authors hypothesized was due to inmates withholding information about mental health diagnoses from correctional staff (Mallik-Kane and Visher 2008). It is also possible that mental health deteriorates for some inmates as a result of incarceration itself.

For those inmates with mental health needs who are identified in time to arrange transition services, several sets of guidelines and best practices have been set out by the American Psychological Association, the Council of State Governments, and other entities. Even with such resources available to discharge planners, releasees with mental health needs often fail to link to community-based care and, even when they do, experience frequent treatment relapses, especially when there is an addiction or substance-dependence comorbidity. Mentally ill releasees are more likely to be homeless and usually exhibit patterns of health-care utilization for acute episodes (e.g., emergency department use or psychiatric hospitalizations) rather than ongoing mental health and medical treatment. Even though inmates with mental illnesses are generally given a supply of medications upon release, medication maintenance has been found to decline 8–10 months postrelease (Mallik-Kane and Visher 2008). The successful transition to community-based care upon reentry may be especially difficult when mentally ill inmates complete their sentences and are released without community-based supervision by a parole officer. Parolees may be subject to oversight and sanctions for failing to complete treatment (though this gives rise to possible ethical concerns about enforcing undesired medical treatment, which has more generally been found unconstitutional in the USA and elsewhere). Releasees who have completed their sentence, on the other hand, are not required to follow the recommendations of discharge planners; this scenario also eliminates the parole officer as a potential resource, albeit a coercive one, for releasees.

The specific mental health needs of people with substance dependence or addiction are subject to similar difficulties in effective discharge planning. Releasees with addictions are particularly prone to relapse and recidivism; many are returning to families or social groups with similar substance abuse habits, increasing the risks of relapse (Mallik-Kane and Visher 2008). In addition, enforced abstinence during incarceration contributes to both relapse and overdose upon release. Medication-assisted treatment (MAT) for opiate addiction has been found effective in reducing heroin use, criminal behavior, HIV-risk behaviors, and overdose deaths. For example, methadone, a synthetic opioid, has been used for several decades to treat opiate addiction, as it reduces the craving for shorter-acting opioids and with proper dosing blocks their euphoric effects. However, correctional authorities are generally reluctant to employ it. As a result, inmates with substance dependence or addiction are less likely to have methadone or other MAT as part of their discharge planning (Nunn et al. 2009). Interventions that follow in-prison drug treatment programs with postrelease treatment have been shown to reduce post-reentry drug use and its associated recidivism. Since the most recent developments in behavioral science emphasize the role of environmental triggers over individual motivation, though, such treatment is often undermined by a return to the original environment upon reentry.

Reviews of existing transition services suggest that discharge planning can provide critical assistance in linking releasees to continued health care upon reentry, with improved outcomes for both health and recidivism. However, discharge planning remains available to only a fraction of reentering prisoners with health needs. In many facilities, concerned staff receive little institutional support from administrators who may view continuity of care as extraneous to the facility’s function of punishment and/or public security. Even in facilities with supportive administrators, budgetary constraints frequently limit the amount of services available. It is also clear that even well-planned and well-funded transition services are undermined by other structural barriers that releasees encounter upon reentry.

Foremost among these is the problem of access to care. Unlike most other developed nations, the uninsured constitute a substantial percentage of the US public. Almost 80 % of former prisoners in a 2008 survey were without private or public insurance upon reentry (Mallik-Kane and Visher 2008). Because unemployment is high among releasees, Medicaid (state/federal coverage for those under 65 with income, family, or health qualifications) is particularly important. Although federal regulations require only that states suspend Medicaid during incarceration, most states terminate it altogether. As a result, releasees lack financial coverage and thus access to most health care during the critical reentry period. Since the reapplication process is cumbersome and often overwhelming for a low-resource population (e.g., low literacy hampers successfully completing paperwork, and the lack of stable housing interferes with the reception of notices from the Social Security Administration), resuming coverage can take months or occur not at all. This problem was exacerbated after 1996, when the Social Security Administration implemented a system of financial incentives to encourage correctional facilities to notify them upon incarceration of a Medicaid recipient, resulting in prompt termination of benefits including medical coverage.

Starting in 2014, the 2010 Patient Protection and Affordable Care Act (ACA) will expand eligibility for Medicaid for most reentering offenders, especially by targeting the low-income adult males that constitute the majority of inmates. Even the expansion of coverage under the ACA, however, will leave about one-third of the currently uninsured without coverage; thus, insurance can be expected to remain a critical factor in postrelease access to health care.

In addition to low rates of coverage, releasees are hampered by poor access to health care and often rely on emergency rooms for care. The successful transition to community-based care upon reentry is further hindered by poor communication among the various health-care providers involved. It is seldom possible for patients even among the general population to conduct their own medical histories from provider to provider, making inter-provider communication all the more critical for patients with complex health needs and low medical literacy. While discharge planning is far more frequent in prisons than in jails, state and federal prisons are also much more likely to be geographically distant from inmates' home communities and thus from the health-care providers to whom releasees must transition. Since a minority of US providers have transitioned to electronic health records (EHR), the successful communication of medical histories and treatment plans from correctional facilities to community-based providers is even more difficult to coordinate. As a result, information regarding health needs and service history can be fragmented and care correspondingly poorly coordinated.

Finally, releasees with health needs frequently reenter a social context that is poorly equipped to support their health needs and can undermine even the best discharge planning. Regardless of intent, releasees often have little choice but to return to networks and associations that support criminal activity and risky health behaviors. Family dynamics are especially critical upon reentry. Inmates with physical health needs are as likely to receive material support from their families as inmates without physical health problems, but they also appear to be more likely to be involved in domestic violence. In a 2008 survey, men with health problems were more likely to be victims of domestic violence than other men upon reentry, though their female counterparts were more likely to perpetrate domestic violence than women without physical health problems (Mallik-Kane and Visher 2008). Releasees with substance dependence or addiction are especially likely to return to families engaged in substance use and criminality; here too, men were more likely to suffer domestic violence than male inmates without substance dependence.

Reentry and Community Health

Correctional facilities do not often live up to their potential to link the medically underserved to health care. In a widely cited 2008 study of releasees in two US states, most of the prisoners returning to the community had received some health care for their physical conditions but one-third of men and a quarter of women had not (Mallik-Kane and Visher 2008); moreover, no studies are available that measure the sufficiency of care that is provided. As a result, most inmates return to their communities with unresolved health problems. This has repercussions not only for the postrelease health of prisoners themselves but for the community as well.

Research into incarceration's effects on community health has focused particularly on the transmission of infectious disease upon release. In the USA, the majority of HIV infections are acquired prior to incarceration, not while in prison or jail (Beckwith et al. 2010). However, incarceration makes a significant contribution to the spread of infectious diseases by means of the behavior it fosters upon reentry (Lichtenstein 2009). Because incarceration in the USA is concentrated in specific communities rather than being randomly distributed, it removes large numbers of young males from impacted neighborhoods and contributes to high-risk sex by leading to increased sexual relationship concurrency. Other studies have found that incarceration is associated with sexually transmitted diseases and teenage pregnancy by similarly affecting neighborhood social characteristics such as the ratio of males to females (Thomas et al. 2008). In addition, because HIV treatment is interrupted upon release in over 90 % of cases, releasees' heightened viral loads carry with them a greater infectiousness and likelihood of developing drug resistance, creating in turn the potential for reservoirs of drug-resistant HIV in communities most affected by incarceration (Baillargeon et al. 2009).

A second field of research has emerged regarding incarceration and family health, examining both the direct effects on the mental health of inmates' families and mediated effects on their physical health. For some, the imprisonment of a family member may represent a reprieve from a family member prone to substance abuse or domestic violence (Comfort 2007), and reentry can combine relief at the return of a family member with new stressors. Since the prison environment may foster psychological orientations like suspicion and aggression, release may have adverse effects on already-existing marriages, in addition to reducing marriage prospects in some communities by affecting the supply of marriageable men (London and Myers 2006; Massoglia 2008). These patterns are of public health concern because marital status is recognized as a key marker of health status, particularly for poor men. Moreover, committed relationships protect against sexual risk taking, so that the incarceration-based disruption of relationships contributes further to the spread of infectious disease.

There is increasing attention among US researchers to the effects of incarceration and reentry upon the children of inmates as well. Although some children's well-being has been found to benefit from the incarceration of a parent (i.e., where there is a history of domestic violence or violent crime), these cases are outnumbered by the adverse effects on mental health when the parent was incarcerated for a nonviolent offense, as is the case in the substantial majority of the increase in incarceration since the 1970s (Wakefield and Wildeman 2011). Parental incarceration is associated with both internalizing mental health problems (e.g., anxiety or depression) and externalizing problems (e.g., aggression or delinquency) (Wildeman and Western 2010). The magnitudes of these associations are small, but on a population level, they contribute 25–45 % of the disparities in behavioral problems between black and white children (Wakefield and Wildeman 2011).

Finally, it is important to recognize the health consequences of the economics of reentry. In addition to decreasing employment opportunities, in most states, a prison record eliminates eligibility for public assistance such as food stamps, public housing, and student loans. These restrictions have repercussions on life trajectories, health behaviors, and health outcomes. State and national policy also contributes to the community health effects of incarceration and reentry. The US Census counts prisoners as residents of the largely white, nonurban communities where prisons and jails are increasingly relocated. As a result, the public funding and political redistricting that utilize census data for distribution favor the latter over the communities from which prisoners come, and both political weight and funding for public health initiatives are diminished in the low-resource communities most releasees return to.

Conclusions

Both incarceration and reentry offer opportunities to improve the treatment and health of high-needs individuals who often lack financial and social resources to address their health needs independently. At the same time, incarceration has been associated with poor long-term health outcomes, and because people of color are incarcerated far more frequently than whites in the

USA, the experience may ultimately exacerbate rather than mitigate health disparities (Massoglia 2008; Wildeman 2011). Similar trends are visible elsewhere, for example, in the disproportionate incarceration of Australian aboriginals. Correctional authorities work regularly with the public health, medical, and social services professions to establish transitional services for releasees with health needs, but the highly complex individual and social challenges of reentry coupled with scant resources have made effective discharge planning rare. Although *Estelle v. Gamble* and similar international mandates do not extend to reentry, significant improvements in postrelease individual and community health are possible with substantial intervention and advocacy by the public health and medical communities.

Acknowledgement Dr. Rich's work was supported by the following NIH grants: NIDA K24DA022112, NIH/CFAR P30-AI-42853, NIDA R01DA030778, and NIDA R01DA027211.

Related Entries

- ▶ HIV in the Correctional System
- ▶ Mentally Disordered Offenders Under Community Supervision
- ▶ Offender Change in Treatment
- ▶ Re-entry
- ▶ Treating Mentally Ill Offenders
- ▶ Unintended Effects of Imprisonment

Recommended Reading and References

Baillargeon J, Giordano TP, Rich JD, Wu ZH, Wells K, Pollock BH, Paar DP (2009) Accessing antiretroviral therapy following release from prison. JAMA 301:848–857

Baillargeon J, Hoge SK, Penn JV (2010) Addressing the challenge of community reentry among released inmates with serious mental illness. Am J Community Psychol 46:361–375

Beckwith CG, Zaller ND, Fu JJ, Montague BT, Rich JD (2010) Opportunities to diagnose, treat, and prevent HIV in the criminal justice system. J Acquir Immune Defic Syndr 55(Suppl 1):S49–S55

Binswanger IA, Stern MF, Deyo RA, Heagerty PJ, Cheadle A, Elmore JG, Koepsell TD (2007) Release from prison – a high risk of death for former inmates. N Engl J Med 356:157–165

Binswanger IA, Krueger PM, Steiner JF (2009) Prevalence of chronic medical conditions among jail and prison inmates in the USA compared with the general population. J Epidemiol Community Health 63:912–919

Brooker C, Sirdifield C, Gojkovic D (2007) Mental health services and prisoners: an updated review. Centre for Clinical and Academic Workforce Innovation, Lincoln, UK

Comfort M (2007) Punishment beyond the legal offender. Ann Rev Law Soc Sci 3:271–296

Conklin T, Lincoln T, Wilson R (2002) A public health manual for correctional health care. Hampden County Sheriff's Department and Massachusetts Public Health Association, Ludlow

Dolan K, Kite B, Black E, Aceijas C, Stimson GV (2007) HIV in prison in low-income and middle-income countries. Lancet Infect Dis 7:32–41

Freudenberg N (2001) Jails, prisons, and the health of urban populations: a review of the impact of the correctional system on community health. J Urban Health 78:214–235

Hammett TM (2006) HIV/AIDS and other infectious diseases among correctional inmates: transmission, burden, and an appropriate response. Am J Public Health 96:974–978

James DJ, Glaze LE (2006) Mental health problems of prison and jail inmates. Bureau of Justice Statistics, Washington, DC

Jurgens R, Nowak M, Day M (2011) HIV and incarceration: prisons and detention. J Int AIDS Soc 14:26

Lewis C (2006) Treating incarcerated women: gender matters. Psychiatr Clin North Am 29:773–789

Lichtenstein B (2009) Drugs, incarceration, and HIV/AIDS among African American men: a critical literature review and call to action. Am J Mens Health 3:252–264

London A, Myers N (2006) Race, incarceration, and health: a life-course approach. Res Aging 28:409–422

Mallik-Kane K, Visher C (2008) Health and prisoner reentry: how physical, mental, and substance abuse conditions shape the process of reintegration. Urban Institute, Washington, DC

Massoglia M (2008) Incarceration, health, and racial disparities in health. Law Soc Rev 42:275–305

Mellow J, Greifinger R (2005) Successful reentry: the perspective of private correctional health care providers. J Urban Health 84:85–98

Nunn A, Zaller N, Dickman S, Trimbur C, Nijhawan A, Rich JD (2009) Methadone and buprenorphine prescribing and referral practices in US prison systems: results from a nationwide survey. Drug Alcohol Depend 105:83–88

Nunn A, Cornwall A, Fu J, Bazerman L, Loewenthal H, Beckwith C (2010) Linking HIV-positive jail inmates to treatment, care, and social services after release:

results from a qualitative assessment of the COMPASS program. J Urban Health 87:954–968
Re-entry Policy Council (2005) Report of the Re-entry Policy Council: charting the safe and successful return of prisoners to the community. Council of State Governments, New York
Rich JD, Wakeman SE, Dickman SL (2011) Medicine and the epidemic of incarceration in the United States. N Engl J Med 364:2081–2083
Springer SA, Spaulding AC, Meyer JP, Altice FL (2011) Public health implications for adequate transitional care for HIV-infected prisoners: five essential components. Clin Infect Dis 53:469–479
Thomas JC, Levandowski BA, Isler MR, Torrone E, Wilson G (2008) Incarceration and sexually transmitted infections: a neighborhood perspective. J Urban Health 85:90–99
US Bureau of Justice Statistics (2003) Reentry trends in the United States: inmates returning to the community after serving time in prison. US Department of Justice, Washington, DC
Wakefield S, Wildeman C (2011) Mass imprisonment and racial disparities in childhood behavioral problems. Criminol Public Policy 10:793–817
Wildeman C (2011) Invited commentary: (mass) imprisonment and (inequities in) health. Am J Epidemiol 173:488–491
Wildeman C, Western B (2010) Incarceration in fragile families. Futur Child 20:157–177
Wilper AP, Woolhandler S, Boyd JW, Lasser KE, McCormick D, Bor DH, Himmelstein DU (2009) The health and health care of US prisoners: results of a nationwide survey. Am J Public Health 99:666–672

Hepatitis C

► Drug Abuse and Alcohol Dependence Among Inmates

Heroin-Assisted Treatment

► Drug Substitution Programs and Offending

Hidden Crime Survey

► History of the Self-Report Delinquency Surveys

High Policing

Gary T. Marx
Massachusetts Institute of Technology,
Cambridge, MA, USA

Overview

What is "high policing?" The term does not refer to the euphoria police may feel after an adrenalin generating challenge is met, nor does it refer to policing while high, although both might accompany the activities falling within the concept's broad meaning. In its original meaning, it referred to the use of political intelligence to preserve the power of the ruler, in particular as this involved stealth, spying, espionage, and intrigue. Yet, like barnacles that become attached to a ship, over time the concept has evolved and layers of meaning have been added. The original ship is long gone, but parts of it endure in new forms and settings throughout society.

High policing can refer to the *location* of an agency such as those attached to the highest levels of governance, to an *ethos* involving intelligence collection and prevention of threats to what is now called national security, and to *methods* of information discovery and subsequent action swathed in secrecy and deception, some of which are at the edge or over the bounds of conventional morality and are supported by an end-justifies-the-means perspective (and sometimes the reverse). The latter is illustrated by a police official's statement – "if we have the technology, why not use it?"

High policing grew out of the late seventeenth- and eighteenth-century ideas and practices associated with the appearance of the modern policed society and police state. The modern police state has high policing at its core. Yet it also is a legitimate and (sometimes illegitimate) feature of democratic societies. Elements of high policing are found in all states, and the use of intelligence (as broadly defined) to protect the organization and to further its goals (and to varying degrees the private goals of leaders) is a characteristic of all modern organizations.

The tools of high policing can thus be seen as either an overarching mushroom cloud or as a rainbow. Society must possess a police force governed by law, but the police force must not possess the society, nor be beyond the law, nor create the law. As ideal types, states vary along multidimensional continua from full-fledged, totalitarian police states that are a law unto themselves where citizens have no rights to democratic societies under the rule of law where citizens in principle have full rights. In democratic societies, the activities historically associated with high policing involve a challenge central to any hierarchical social organization: how to grant authorities the resources to effectively govern and serve the public or the organization's interest while ensuring that the use of those resources is not abused for the private interests of those governing. In democratic societies, this requires differentiating those who wield power from the social institutions they are supposed to serve. Contemporary leaders would never say as popular myth claimed the absolutist French king Louis XIV did, "I am the state" (L'etat, c'est moi").

This entry considers the emergence of high policing as a developing form in Europe over recent centuries concomitant with modernization; following the work of Jean-Paul Brodeur, its key components are noted and some elaborations of these are suggested. The entry concludes with a discussion of some enduring trade-offs and tensions of high policing in a democratic society.

Fundamentals

Emergence of the Policed Society and the Police State

In 1970, political scientist Brian Chapman offered a succinct primer defining the characteristics of the police state over several centuries. He observes that the words police, politics, and policy share common roots in *politeia*, the Greek word for tribe or community. In contemporary usage, polity refers to a politically organized You +1'd this publicly. Undo unit. Such units are administered and regulated by designated government agents.

Chapman discussed but did not offer a formal definition of high policing. But from his historical examples, it involves police directly attached to the head of government with a goal of maintaining the sovereign or a single political party in power above all else. Some of the agents and the means they use may be moral and supported by laws developed through genuinely democratic procedures. However, pragmatism, rather than morality, was historically the guiding principle. For a series of reasons, deception and what, in contemporary terms, are called intelligence and counterintelligence "dirty tricks" – whether involving coercion or deception – were major tools. The ironically named high police of European empires operated in the shade (Brunet 1990) and often used shady practices, although not always for nefarious purposes. From another perspective, the tactics were anything but high. Livres

In the fifteenth century, given the intertwining of the state and the church, religious surveillance of the public was the prominent form. Indeed, it could be indistinguishable from political surveillance. It built on the watchful and potentially wrathful eye of the biblical God. This involved the search for heretics, devils, and witches, as well as the more routine policing of religious consciousness, ritual, and religiously based rules such as those involving adultery and wedlock. Religious officials kept records involving births, marriages, baptisms, and deaths.

However, as individuals gained new rights and the demand for religious orthodoxy weakened, religious surveillance gradually declined. The state took over keeping traditional records on the population (such as birth and death), additional data was recorded, and identity documents and dossiers became more prominent.

In the sixteenth and seventeenth centuries, with the appearance and growth of the embryonic nation-state, which had new needs as well as a heightened capacity to gather and use information, political surveillance gained in prominence. New concerns over political loyalty appeared, along with more strident demands by the public for democratic citizenship rights. The search for political dissidents gained in prominence relative to the search for religious dissidents. Not surprisingly, monarchs sought more sophisticated means of ruling – including dealing with threats to their regimes, at first from their aristocratic peers and later from the public. A silent (and sometimes not so silent) Praetorian Guard developed to protect the sovereign.

Wider domestic political surveillance is a natural corollary of the centralization of power and the increased heterogeneity, scale, and mobility associated with the emerging modern society. Concerns over possible links between internal and external security become more evident.

Over the next several centuries, we see a gradual move to a broadly "policed" society in which agents of the state, industry, and commerce come to exercise control over ever-wider social and geographical areas. This was done in a world rapidly becoming more economically and socially interdependent and in which societal scale and complexity markedly increased. The tactics of high policing served to usher in the seemingly evermore powerful, centrally administered state. Modern, presumably rational, even supposedly scientific, ideas were applied to the administration of an emerging "policed society." In building on the work of Bentham, Marx, Nietzsche, and Weber, many later scholars have elaborated on these developments (e.g., Silver 1969, Shils 1975; Foucault 1977; Nisbet 1977, and Cohen 1985). A brief history of high policing forms will be considered next.

In prerevolutionary France, Cardinal Richelieu (1585–1642) created an explicitly political police unit with a very broad mandate to find conspiracies among elites. This helped consolidate royal power and squashed opponents. Steeped in secrecy and the mystique that can be associated with it, they sought to create the impression that they were omniscient (which they weren't) and omnipotent (which they pretty much were).

The "high police" (high because they were the direct agents of the sovereign and also gave special attention to threats posed by rival aristocrats) were among the first modern intelligence agencies. They looked to the future and to the larger picture. This contrasted with the more visible "low police" who came to regulate everyday affairs such as those involving markets, street crime, and sanitation.

Chapman locates the first modern police state in the highly militarized society of Prussia under Frederick William (1688–1740) and Frederick II (1744–1797) and in Austria under Emperor Joseph II (1741–1790). Joseph Fouche (1759–1820), Napoleon's ruthless, innovative, and intelligence-driven minister of police, made extensive use of high policing tactics, building on the early French efforts of Cardinal Richelieu. He reorganized and centralized the French police and made extensive use of criminals as informants. His police kept order but also operated as a domestic intelligence service spying on citizens to keep the emperor informed. His legacy has endured both internally and internationally. The French political intelligence system is one of the world's most developed and extensive.

The omnipresent police system that Emperor Joseph created watched government officials to be sure they carried out the reforms he sought and enforced the state's expanding rules. His police magistrates had a responsibility to report on, and regulate, a wide range of public behavior. But he also created a less (or non-bureaucratic) secret police group that made extensive use of informants and investigated potentially, or actually subversive groups (defined as being in opposition to the emperor). These police reported directly to the leader.

While hardly beacons of freedom or the city on the hill, the eighteenth-century societies of "enlightened absolutism" were not the draconian, tyrannous places described by Aristotle nor those of twentieth-century fascist and communist societies. Drawing on Plato's idea of benevolent despotism that filtered through to Hobbes and Rousseau, leaders believed that a strong state was needed to protect the emperor and the public (their interests were assumed to be synonymous). In complicated matters of governing, leaders were presumed to know what was best. It was not so much that the ideas they applied were new but that the capability of acting on them had improved and the perceived need to do so in the face of a growing urban society and more internationalized world was felt to be great. Intermittent reliance on the military for policing was no longer sufficient.

A more penetrating and comprehensive societal regulation from a central government source was necessary relative to the more decentralized, episodic, laissez-faire control regimes of the feudal period. As Max Weber noted, the state developed a self-perpetuating, hierarchal bureaucracy to administer an ever expanding set of minutely defined rules. These were applied to the economy, public welfare, and morals and were dependent on the systematic collection and management of information about the populace.

In the seventeenth and eighteenth centuries, threats to those in power were most likely to come from other elites – sometimes presumed to have ties to foreign countries. There was less concern with political surveillance of the unruly and sometimes riotous masses. However, after the French Revolution and with the coming of the nineteenth-century industrialization and increased resources for challengers (including modern means of communication and new political rights), the high police mandate broadened to greater concern with average citizens and threats from other countries.

Over the last several centuries, police intelligence agencies became much more systematic, organized, and specialized. Their resources vastly expanded and continued to expand, their tools became evermore sophisticated (from agents in balloons to satellites), and new goals and forms of organization appeared. An expanded census, improved record keeping, police registers and dossiers, identity documents (including those based on crude biometrics such as the shape of the head), and inspections offered data. Personal information came to be collected for taxation, conscription, law enforcement, and border control (both immigration and emigration) uses and also to determine citizenship and eligibility for democratic participation and to improve social planning and public health. And as a corollary, the line blurred between direct political surveillance and a more modern and in some ways more benign or at least neutral governance or administration, even while containing the seeds of twentieth-century totalitarianism and beyond.

The Work of Jean-Paul Brodeur

In a classic article, with the memories of the 1960s and the early 1970s protests and police actions in the USA and Canada still vivid, along with the Watergate scandal, Canadian criminologist-philosopher Jean-Paul Brodeur (1983) expanded on the concept of high policing. He continued to work on the topic (as both a scholar and government advisor) over the remaining decades of his all too short life (Manning 2012). This work helped open up an interdisciplinary and international field (culminating in his monumental book *The Police Web 2011*).

Brodeur saw high policing as consisting of political surveillance and low policing as law enforcement (holding apart the overlaps between the two). In 1983, he identified four features of the former:

1. It was absorbent in the sense of casting a wide net (what it absorbed was information).
2. It conflated legislative, judiciary, and executive or administrative powers.
3. The preservation of the existing political regime was its primary goal.
4. A major tool was the use of informants from throughout the society.

   In 2011, he noted five additional characteristics implicit in the above:

5. The state as intended victim
6. The utilization of criminals
7. Secrecy
8. Deceit
9. Extralegality

This comprehensive list can be expanded to note that such policing is not only absorbent like a sponge soaking up all information it encounters but is also laser-like in its ability to focus. These attributes may be combined as the agent moves from a general search (often using automated data) to a focus on specific profiles suggesting subjects and suspects of interest. The reliance on the new surveillance (e.g., computer surveillance, matching and profiling, satellites, drones, video cams, DNA, and other biometric measures) is central here. The ratio of human intelligence (humint) to signals intelligence (sigint) has shifted markedly toward the latter in recent decades.

Extralegality can be expanded to include gray policing and intelligence (Hoogenboom 2010) which blur the borders between the public and the private and takes creative advantage of value conflicts and the ambiguity of laws and policies, sometimes avoiding restrictions by delegating activities to organizations or countries not bound by restrictions and by working in areas that are not (yet) regulated but should be, given the standards of the society.

A significant change involves the emergent, complex, breaking, reconfiguration, and hybridization of a variety of high policing borders. Some of the traditional distinctions between "us" and "them" and inside and outside of a society change with globalization under the sway of new threats, technologies, and economic arrangements.

Organizations that were previously more distinct show increased commonalities – whether in the means used, intelligence-based and risk assessment goals, and in the shifting of personnel among federal and local police, police, among military, and national security agencies, between national and international agencies, and between government agencies and the private sector. There are new fusion centers where different agencies work together and an increase in task forces for specific investigations. There is some merging or convergence of military, national security, domestic police, and private sector security into a new multifaceted professional field (Bigo 2006). In the struggle with stateless terrorism and global crime, high policing mixes with traditional law enforcement as well as the military.

There are also increased overlaps between groups engaged in drug dealing and other crimes, violent political groups, and failed states. To varying degrees, national security agencies in democratic societies also covertly work with such groups. The secrecy and the multiplicity of agencies sometimes mean they work at cross-purposes.

There is increased cooperation between state and private sector actors and significant growth of national and international private sector agencies. Private police groups in pursing the goals of their corporate sponsors, or being delegated piece work by government agencies, are increasingly proactive rather than reactive and are far less accountable even than the public sector in its use of secret tactics. Lubbers (2012), for example, has documented many cases of private sector intelligence and provocation activities against protest groups.

For North America, empirical documentation on efforts to harm movements is widely available (e.g., Donner 1990, Cunningham 2004, Greenberg 2011). Such efforts were intensified and took new forms after 9/11 and with the worldwide anti-globalization protests (O'Harrow 2005, Chesterman 2011; Starr et al. 2011 and Earl 2011).

Beyond trying to hurt opponents, significant resources go into covertly helping allies such as by providing funds and propaganda. The Iran-Contra affair is an example. In a domestic context, Marx (1979) identifies a number of actions including giving allies intelligence, building leaders, creating a positive image of a movement, aiding recruitment, and providing money and supplies. In the United States during the 1960s, the CIA provided funds to the National Student Association, a group seen to be supportive of its goals, while the FBI and local police through programs such as Cointel sought to damage other student groups that were protesting. Similar actions on the part of the Royal Canadian Mounted Police under its Disruptive Tactics program were seen in Canada.

Brodeur emphasizes secrecy and deceit as means of piercing informational borders. Given the importance of finding information that is itself often secret, the turn to stealth is to be expected. Creating the myth of an omnipresent and omnipotent surveillance can itself be a form of control. But secrecy and deception may also be used on behalf of coercion and violence because rulers wish to hide their responsibility for the actions taken and to avoid reciprocal actions by opponents. Subterfuge may mask the identity of those sponsoring and carrying out violence. New forms based on secret intelligence collection such as armed drones and cyber-attacks became prominent in the twenty-first century.

Whether it is called the policing of politics, political intelligence, or regime (or societal) protection, high policing with respect to both its goals and its means is a pillar of social control in modern society – nationally and, even more so, internationally (although not the only pillar). It involves organizations, an ethos, and a set of activities that are most commonly associated with, and developed by, national security (as broadly defined) agencies but in fact goes beyond this wide domain to many social control settings in which organizations protect their flanks, try to anticipate or orchestrate events, and attend to actions, persons, and contexts seen as threats.

Security and intelligence agencies engage in wide sweeping information gathering activities for strategic planning and for preventive and preemptive purposes. Depending on the kind of society, this may be to protect the public interest by protecting the state's interest (including but going far beyond protection of the sovereign), to protect the sovereign or the ruling party alone, or even to only serve the interests of the police agency.

High police remind us that the police are above all political actors. Two meanings of this can be noted. The low police on the street were concerned with protecting individual victims and with order maintenance. High police, in contrast, were concerned with protecting a corporate entity – the political regime. A major question is whether in protecting the sovereign, police are also protecting the broader (civil) society. In a democratic setting when police protect the leader, they ideally are also protecting the society's democratic constitution and institutions. In an authoritarian or totalitarian setting, high police act as a private police ensuring the power of the rulers (whether a king or political party) apart from the broader society. That does not preclude disingenuous legitimating claims about acting on behalf of the people.

Policing that puts knowledge at its core often has a macro and longer-time frame focus. In addition to specific groups, persons, and behavior, sources of possible insecurity are sought far beyond the traditional cabal. The CIA, for example, has a Center on Climate Change and National Security.

A country's economic well-being and the economic activities of rival countries are also presumed to relate to national security. National intelligence agencies routinely gather open source economic information on their rivals and are sometimes suspected of industrial espionage in order to help their domestic economy. France's DST (Direction de la Surveillance du Territoire) through its unit for the Security and Protection of National Assets was well known for the attention it devoted to supporting French economy and technology in an array of fields.

Given the number of variables associated with high policing, there is no single theme that characterizes all activities and agencies. Tasks such as intelligence collection, analysis, and operations may be internally divided or allocated among different agencies. Organizations may be centralized or decentralized. The work may involve a single or (most often) multiple agencies, have responsibility for foreign and/or domestic matters (as with Britain's MI6 and MI5), be national or local (as with the FBI and urban police), be distinct or attached to other branches of government, and if the latter be part of the military, state, or justice departments, emphasize technical (as in the USA) or human (as in France) information collection sources. But however the activity is organized; accountability themes are an enduring concern in democratic societies.

While some activity is overt and involves open source information (see the CIA's World Factbook on its webpage www.cia.gov), agencies and agents whose identity, location, and activities are clearly known much remain secretive and hidden. These activities are buffeted to varying degrees from the conventional accountability of government and direction and control by legislatures, the judiciary, and even the executive. High police may be a law unto themselves, partly because some activities are not controlled by law (being neither required nor prohibited by a legislature but tolerated if done discretely). There will likely be different degrees of control, oversight, and review at different stages for domestic and international actions and for citizens and noncitizens.

A vital question is what the limits should be in an age of globalization where traditional borders are weakened – even as new borders appear and weapons of unimaginable horror are present and formal war has not been declared. Should (or can) there be common standards of law and ethics for these methods across national borders? Should noncitizens within a country have the same rights and protections as citizens? Should citizens have any special protections from their own government if they are in another country and behave as enemies of their government? The United Nations Universal Declaration of Human Rights suggests that basic rights must transcend national borders or citizenship status. But can standards be enforced internationally when mechanisms for compliance are weak or absent, when adversaries do not honor these standards, and when another standard argues for respecting differences among local cultures that give less emphasis to human rights in the western tradition?

In democratic societies, there is an ebb and flow of efforts to rein in or unleash high police activities (and even to acknowledge them), depending on the gravity of perceived threats, crisis, and scandals. Some data gathering techniques are simply too new to have been given much legislative control such as drone and satellite surveillance. The laws that do regulate can be vague and provide for maximum discretion and for states of exception. Plausible deniability may also be offered by the delegation of prohibited or controversial tactics to private contractors or to countries with fewer inhibitions (a practice known as a "functional equivalent").

Even when the laws and regulations are adequate, the scale of intelligence gathering, the shrouds of secrecy protecting the behavior, and the "national security justification" work against implementing the kind of accountability found with other government agencies. In the United States, for example, in 2011, there were 16 agencies concerned with national security with a budget of over 75 million dollars, 200,000 employees, and extensive private contracting. The activities of these secret agencies (including specialists in misinformation) are classified and often exist within an insulated culture of incipient crisis distrustful of outsiders. Secrecy feeds on itself, whether out of caution or self-protection, and there is a point at which it becomes highly dysfunctional and makes democratic dialogue impossible.

Such factors make it difficult for citizens and researchers to know what is going on, let alone most specialists. In 2012 in the United States, an estimated 4.2 million persons had access to classified documents. But information access was highly compartmentalized and limited to the narrow task at hand. Given such factors relatively little is known about the agencies, often even by government oversight committees, let alone by citizens and researchers.

Claims about high policing must be carefully assessed given the deception and secrecy surrounding the topic and the vested interests of the parties involved (police do not wish to reveal operational tactics or behavior that might create bad public relations or harm a prosecution and opponents often have an interest in painting agents in a negative light). Yet information is sometimes available as a result of government hearings and court records, first person accounts from agents who publicly disavow their actions, leaks, Freedom of Information Act requests, archives, police training materials, and investigative journalists.

Ideal types call attention to broad differences yet obscure crosscutting dimensions that may unite what seems different and separate what seems similar. As an ideal type, covert high policing organizations usefully contrast with the overt low policing organizations largely involved with the immediate policing of everyday life such as traffic control, property crimes, or domestic disputes. In these settings, uniformed personnel have a narrow focus on local incidents that occurred, or are occurring, and that involve order maintenance or law enforcement. Law plays a more significant role in their activities, and their actions are distinct from (but not independent of) those of the judiciary.

For some purposes, broad contrasts between high and low police organizations are helpful, but the variables built into the ideal type can also be unpacked and seen as dichotomous or continuous factors. The latter take account of degrees (just how secret is the organization, to what extent is it controlled by law and policy?). Apart from any particular organization, high policing can be defined as an ethos and a set of activities found to varying degrees in any control organization – national or local, police or nonpolice, and public or private.

The elements of high policing can be used to analyze social control settings. Viewing high policing in terms of its component parts provides a systematic way to identify variation and to think about explanations and consequences of the different ways of doing policing. This approach emphasizes policing as a process or function rather than as an organizational structure. It also calls attention to cultural themes which may be widely shared in a society such as systems thinking and an interest in dealing with root causes, rather than only with consequences and in technical developments associated with the information age which provide new resources for analysis and planning.

Thus, many of the elements of high policing are seen in current policies such as intelligence-led (database-based) and community policing and specialized police units such as those involving intelligence, undercover, internal affairs, and planning. Detectives, and even uniformed police, make use of traditional (informers) advanced tools of surveillance (thermal imaging). The reverse is also the case – high policing units are involved in activities associated with law enforcement – reacting to events rather than only trying to prevent them locating suspects. Over centuries, the mission of high policing, at least in democratic societies as an aspect of modernization, has in principle broadened beyond protecting those in power to ensuring the security of civil society and protecting citizen's rights. As noted, this is part of modernization. Secret police may occasionally even help create a democratic state in opposition to the wishes of a country's leaders, as to some degree was the case of Poland's velvet transition in 1989 (Los and Zybertowicz 2000). There can also be significant formal limits in the discretion lower level operational units have in doing intelligence investigations (Lowe 2011).

Some Ethical and Practical Issues

Those using the less morally elevated tactics of high policing offer many justifications. Police cannot be everywhere, and the frequent resort to secrecy and deception on the part of their adversaries creates informational borders. Self-defense – whether by states or individuals – is in general morally and legally defensible. A US judge said, "the Constitution is not a suicide pact." Yet fighting fire with fire may burn and even destroy the fighters.

Simmel observed that in conflict settings, adversaries come to resemble each other. The secrecy, discretion, and minimal accountability often found with the policing of politics can result in abuses by the state that are difficult to document. The uneasy alliances between intelligence agents and criminals may result in criminals who are even less subject to restraints coming to control the agents. Such associations and the use of tactics perceived as dirty may taint the agents. As the history of the British police suggests, police are moral actors and their behavior has symbolic meaning.

What happens to this elevated image of police (and of government more generally) when they resort to the dirty tactics of deceiving, lying, spying, infiltrating, and relying on informers and criminals? Is the state then indistinguishable from its genuine enemies who may use the same tactics indiscriminately?

Using citizens who are not criminals as informers (apart from alliances with criminals) greatly extends the eyes and ears of police and avoids some problems. From one standpoint, this is even democratic and honors the British tradition that police are simply citizens with no powers beyond those of other citizens. The fact that the first modern British police were unarmed reflects this community orientation in which police were dependent on the good will of those they served. Citizens would obey the law out of respect for it, not fear, coercion, or deception. Citizens were seen to have a responsibility to contribute to the policing of their communities. This "low" police force was self-consciously created in opposition to the high police of the European continent who served the sovereign not the people. Yet in widely cultivating informers among citizens (as was done in the former East German Republic where over 2 % of the public were registered informers and many more unofficially participated), privacy, trust, and community are difficult to sustain.

High police efforts emphasize prevention. This impetus has been increased by the astoundingly destructive potential of modern weapons, whether nuclear, chemical, and biological or cyber-attacks, in an evermore interdependent world. Prevention can be approached from a standpoint of big picture hard or soft engineering efforts to avoid problems (e.g., improved infrastructure protection as with enhanced security of bridges or computer networks or institution building in an effort to win hearts and minds), through deterrence via means of identification and severe sanctions, or through direct interventions to thwart planned attacks.

The latter efforts often draw upon undercover means as when agents substitute fake for real dynamite or gather evidence to prove charges of conspiracy. Such efforts may also backfire as agents lose control of events. There is also a trade-off here since in preventing an unwanted event agents risk giving away their sources of information and other tactics. There is often a conflict between those involved in intelligence collection who wish to protect their sources and those involved in operations. There also is the hoarding and failure to share information among rival intelligence agencies as was the case with the FBI and the CIA in the 9/11 case (9/11 Commission 2004).

Information collection can feed upon itself generating the need for evermore information and to wait to use it until the last moment. The idea of risk management on a broad scale requires large amounts of data across many cases (Ericson and Haggerty 1997) and lends itself to profiles and statistical judgments, rather than emphasizing the unique individual. The latter is a condition of due process which assumes that there will be some grounds for suspicion before a citizen is subject to state intervention.

The advantages of prevention are obvious. Yet resources should not go into preventing things that would be unlikely to occur or things which are a result of agent provocation. The line between speech and action is easily blurred; speech is of course a kind of action, but it does not offer the kind of definitive evidence of intent, nor the harm, that results from actually carrying out a destructive act. Yet to wait until after the fact can mean failing to protect the society. But under conditions of law, the need to prevent may conflict with the need to have adequate evidence to proceed and that may require waiting and involvement with suspects. Prevention and facilitation may conflict.

Yet assuming that the internal and external threats that high police tactics are directed against are genuine and not orchestrated by police themselves through dissimulated media campaigns and agent provocateurs or based on vindictive or erroneous informers, some distinctions can be noted. The legality and desirability of both means and ends and the procedures followed need to be considered. Using the tactics to protect a legitimate democratic government is very different from using them to overthrow such a government or to protect the private interests of those with power (as was the case in much of nineteenth-century Europe and with former president Nixon and Watergate).

Attention as well needs to be given to the relationship between dissent and crime. A robust democratic society encourages questioning and guarantees citizens the right to protest. At the same time, pressures to criminalize dissent are always present, and some laws, such as those involving conspiracy and sedition, lend themselves to elastic interpretations.

When the goal is legitimate, questions still need to be asked about the availability of other more desirable means and the consequences of not acting, the creation of a precedent, the ability to control events, the impact on agents and third parties, and the symbolic message of using a tactic. Using tactics that are legally sanctioned through a democratic process and subject to independent review and challenge (whether by a judge, an oversight committee, an inspector general, or supervisors) brings a degree of accountability and is distinct from their use in an unregulated fashion – whether by agents of the state or opponents. Oversight by the mass media and interest groups also has an important role here.

Yet even when legal and procedurally wise, the decision for a democratic state to sanction use of the tactics must be done reluctantly given the risks and an understanding of how easily secret power can be abused. There are no easy answers to what Klockars (1980) called, after the film of the same name, the "Dirty Harry" problem regarding means and ends. Awareness and caution are continually required.

When police states fail as in Eastern Europe in 1989, among the first demands is for the abolition of the secret police. However, while particular leaders may be punished, their organizations destroyed, and extreme tactics prohibited, secrecy and many of the information gathering tactics endure as a functional necessity, and agents pop up elsewhere. No modern state is without security agencies concerned with internal and external threats, and in a global environment with geographical mobility, the lines between these can be blurred given sleeper agents and the recruitment of citizens on religious, ethnic, or ideological grounds to violent international causes. Few societies will favor moral purity over self-protection more than once.

But how should a democratic society control high police tactics? Most states prefer to keep these means in the shadows – neither unleashing nor prohibiting them. They are seen to fit within the category of necessary evils – means which may occasionally be justified, but certainly not preferred. When present, the laws, regulations, and reviews that legitimate them are often shrouded in secrecy, leave considerable wiggle room, and change as conditions do. Thus, in the United States, many of the reforms limiting surveillance that followed the Watergate scandal and earlier abuses were undone following 9/11 with the Patriot Act and other legislation. Similar loosening occurred in Canada and in Europe.

Some domestic tactics may be banned outright, e.g., torture, lengthy detention without due process. Others such as wiretapping require a warrant, review by a secret court, or are permitted only temporarily under emergency conditions (the challenge is to permit such states of exception from becoming permanent and not to keep policy discussions about them secret as with a tendency to overclassify in the name of security). Other tactics such as undercover means in the United States are subject to internal organizational policies and reviews rather than legislation, in contrast to much of Europe (Ross 2004).

Regulating (rather than outright banning) the tactics via the law and policy may lessen problems but, ironically, can also lead to their expanded use since they are officially sanctioned. In contrast, without a formal mandate legitimating and acknowledging such tactics, agents may be unduly passive and hesitate to use them because of uncertainty about where to draw the lines and whether they will be sanctioned for going too far. Yet to formally ignore the tactics is an invitation to abuse. Careful supervision and subordination to the executive is necessary, but that power can ironically lead to the executive misusing police powers. Critics often do not realize that police committed to the rule of law may face political pressure from the executive to break the law.

What kind of regulation then is appropriate? Democratic societies walk an uneasy and ambivalent line here, particularly domestically. Regulation and transparency can bring accountability, but if too detailed and formulaic, these can become straightjackets that eliminate organizational discretion, flexibility, and innovation and stifle personal questioning, honest communication, and experimentation. On the other hand, if they are too open-ended and permissive of discretion, they invite abuse. Too much visibility regarding the limits and operations of security agencies can be a resource for those who would destroy the state.

In an important sense, a democratic police is a politically neutral police even as it may have legitimate resort to secrecy, coercion, and deception. In democratic societies, police are to behave in a universalistic fashion and strive for equal law enforcement. Should their personal attitudes depart from the demands of the role they are playing, this must not affect their behavior. They are not to act in an explicitly political fashion, such as by spying on, or disrupting groups they disagree with, or failing to enforce the law against groups they support. Nor are they to serve the partisan interests of the leader or the party in power (or the party they would like to see in power). Their purpose must not be to enforce political conformity. Police show neutrality if they simply enforce the rules in equivalent contexts regardless of the beliefs or characteristics of the persons or group involved (e.g., their race, gender, age, ability-disability status, or social class). With this ethos, police are better seen as referees or umpires than as conventional political actors pursuing self-interest. From this perspective, given freedom of the press and association and the legal toleration (indeed even protection) of dissent, it is sometimes asserted that in democracies there is no policing of politics, only of crime.

But apart from the ideal in the above paragraph, there is a second sense in which police are not neutral and *all* policing is political. Police are agents of a particular state and enforce the laws of that state. In that sense, any policing is "political" (in and of itself that does not imply that it is right or wrong). Police have a mandate to legally use force and to deprive citizens of their liberty. To those who disagree with those laws, police tactics – whether high or low – will not appear neutral since they are on behalf of the regime in power and the status quo. As a result, police, even in a democratic society, are likely to be much more controversial than other agencies of government. Yet the line between speech and action can be hazy, and definitions of crime can be elastic.

There is an enduring unease in sustaining a democratic society partly through coercion and secrecy in the face of civil liberties protections and in the absence of traditional criminal violations. While national security efforts may be legally sanctioned and ultimately draw on criminal law, prosecution plays a very minor role, and the implementation of legal oversight is a continual challenge – both because of agency resistance and resources and, not infrequently, the reluctance of regulators to probe. The unease may be dealt with by outright denial, obfuscation, illusion, and rules and policies that maximize discretion. The public is often told by agency leaders, "trust us because we carefully choose, train and supervise agents and rigorously review actions to be taken." Less often heard is the refrain, and besides, "you don't want to know."

Conclusion

A thread running through all systems with an ethos of total control – whether totalitarianism in a state, a prison, or a demanding cult – is for the authority to deny individuals the right to control information about themselves and to restrict their liberty. At the same time, in such societies, those with power seek to maintain full control over information about themselves with few or no restrictions on what they do. The importance given to the end of staying in power and otherwise pursuing their interests overwhelms concern with the morality of the means used and full concern with the public good in which the end of official power is to serve citizens.

It has been said that a civilization's nature can be seen in how it treats its prisoners. It might also be seen in the respect it shows for the informational borders and liberty of the person and the ability to form groups independent of the state. Central here is the kind of social structure and culture within which power is embedded.

As de Tocqueville argued (after his nineteenth-century visit to the United States), a democratic society requires that individuals be free to come together in associations apart from the domination of the state or other all-powerful organizations and that there be significant sharing of power and the acknowledgment of the legitimacy of different, often conflicting, interests. When restrictions upon liberty are appropriate, this must be under conditions of law, due process, and accountability.

In his argument, Tocqueville drew on the classic insights of Aristotle regarding how tyrannical regimes maintained their power. The latter's characterization anticipates the modern police state (not to mention the society George Orwell and other dystopian novelists depicted). Tyrannical regimes seek to perpetuate themselves through the creation of atomistic populations, the blocking of private communication, a prohibition on the forming of private associations, the elimination of borders to visibility, and the sowing of suspicion and discord among their subjects.

Aristotle, drawing from his knowledge of the great tyrannies of the ancient world, argues that to remain in power the tyrant, "...must not allow common meals, clubs, education, and the like; he must be upon his guard against anything which is likely to inspire either courage or confidence among his subjects; he must prohibit literary assemblies or other meetings for discussion, and he must take every means to prevent people from knowing one another (for acquaintance begets mutual confidence). Further, he must compel all persons staying in the city to appear in public and live at his gates; then he will know what they are doing: if they are always kept under, they will learn to be humble..... A tyrant should also endeavor to know what each of his subjects says or does, and should employ spies..." Eavesdroppers should be sent to "any place of resort or meeting; for the fear of informers prevents people from speaking their minds, and if they do, they are more easily found out. Another art of the tyrant is to sow quarrels among the citizens; friends should be embroiled with friends, the people with the notables, and the rich with one another."

In contrast, the democratic leader must do the opposite. The state must ensure that its rules are fairly enforced and that citizens are protected from each other and from internal and external threats (who defines these, and how, are vital questions). It must guard its own guards to be sure they act appropriately. Finally, the democratic state has an obligation to protect itself as the representative of the civil society it serves. That contrasts with using its vast resources to protect the narrow, personal interests of those who are in power at the moment. The police (and government more broadly) must belong to the society. The society must not belong to the police. Any complex democratic society must be a policed society within a framework defined by law for reasons of administration and protection.

A supportive culture, appropriate checks and balances, accountability, and visibility are central to countering the abuses that unequal power and secrecy invite. Unfortunately, no complex society can do without police secrecy – a factor that can be independent of a secret police. There is an eternal tension between protecting the public from itself (as well as from other country's publics) and from those entrusted with doing the protecting. The "Dirty Harry problem" of means and ends and related questions of "good people and dirty work" (Hughes 1962) can never be resolved. But they can be acknowledged and must be continuously wrestled with. Within appropriate limits, legitimate authority must have the power to act coercively and deceptively. Societies are messy affairs awash in moral dilemmas, value conflicts, and trade-offs. Desirable as well as dastardly deeds can occur under cover of darkness; sunlight can illuminate as well as blind; those in positions of authority sometimes do the wrong thing in order to do the right thing. While Dr. Faustus may sometimes be granted a place at (or under) the table, he must be invited reluctantly and only with adequate oversight. It is vital to confront, rather than deny, the ironies and challenges this brings. Therein lies the central paradox of government, and indeed of any authority – so well put by James Madison in the Federalist (Paper 51), "You must first enable the government to control the governed and in the next place, oblige it to control itself."

Related Entries

- ▶ Communities and the Police
- ▶ Conceptualizing of Police
- ▶ Democratic Policing
- ▶ Intelligence-Led Policing
- ▶ Police in the Police State
- ▶ Police Lying and Deception
- ▶ Political Crime

Recommended Reading and References

Aristotle, Politics (book 5). http://classics.mit.edu/Aristotle/politics.html. Accessed 01 July 2012

Bamford J (2008) The shadow factory; the ultra-secret NSA from 9/11 to the eavesdropping on America. Doubleday, New York

Bigo D (2006) Globalized (in)security: the field and the Ban-Opticon. In: Bigo D, Tsoukala A (eds) Illiberal practices of liberal regimes: the (in)security games. L'Harmattan, Paris

Brodeur J (1983) High policing and low policing: remarks about the policing of political activities. Soc Probl 30:507–520

Brodeur J (2011) The police web. Oxford University Press, Oxford

Brunet JP (1990) La Police de l'Ombre. Seuil, Paris

Chapman B (1970) Police state. Pall Mall, London

Chesterman S (2011) One nation under surveillance: a new social contract to defend freedom without sacrificing liberty. Oxford University Press, New York

Cohen S (1985) Visions of social control. Polity Press, Cambridge

Cunningham D (2005) There's something happening here: the new left, the Klan, and FBI counterintelligence. University of California Press, Berkeley

Donner F (1990) Protectors of privilege. University of California Press, Berkeley

Earl J (2011) Political repression: iron fists, velvet gloves, and diffuse control. Annu Rev Sociol 37:261–284

Ericson R, Haggerty K (1997) Policing the risk society. Oxford University Press, Oxford

Foucault M (1977) Discipline and punish: the birth of the prison. Pantheon, New York

Greenberg I (2010) The dangers of dissent: the FBI and civil liberties since 1965. Lexington Books, Lanham

Hoogenboom R (2010) The governance of policing and security ironies, myths and paradoxes. Palgrave Macmillan, London

Hughes EC (1962) Good people and dirty work. Soc Probl 10(1):3–11

Klockars C (1980) The dirty Harry problem. Ann Am Acad Polit Soc Sci 452:33–48

Loew D (2011) The lack of discretion in high policing. Polic Soc 21:233–247

Los M, Zybertowicz A (2000) Privatizing the police state. MacMillan, Basingstoke

Lubbers E (2012) Secret manoeuvres in the dark. Pluto Press, London

Manning P (2012) Jean-Paul Brodeur on high and low policing. Nouvelle Rev Int Criminol 11

Marx GT (1974) Thoughts on a neglected category of social movement participant: the agent provocateur and the informant. Am J Sociol 80:415–421

Marx GT (1979) External efforts to damage or facilitate social movements: some patterns, explanations, outcomes, and complications. In: Zald M, McCarthy J (eds) The dynamics of social movements resource mobilization, social control and tactics. Winthrop, Cambridge, MA

Nisbet R (1977) The twilight of authority. Oxford University Press, New York

Priest D, Arkin W (2011) Top secret America. Little Brown, Boston

Ross J (2004) Impediments to transnational cooperation in undercover policing: a comparative study of the United States and Italy. Am J Comp Law 52:569–624

Shils E (1975) Center and periphery. University of Chicago Press, Chicago

Silver A (1967) The demand for order in civil society: a review of some themes in the history of urban crime, police and riot. In: Bordua DJ (ed) The police. Wiley, New York

Starr A, Fernandez L, Scholl C (2011) Shutting down the streets: political violence in the global era. New York University Press, New York

The 9/11 Commission Report (2004) USGPO, Washington, DC

High-Frequency Victimization

► Repeat Victimization

Historical and Comparative Perspectives on Incarceration

Matthew DeMichele[1] and Janet P. Stamatel[2]
[1]Justice Center for Research, Department of Sociology, Penn State University, University Park, PA
[2]Department of Sociology, University of Kentucky, Lexington, KY, USA

Overview

The US incarceration rate is exceptional in a number of respects. First, the USA has the highest number of people imprisoned (over two million) and the highest imprisonment rate in the world (730 per 100,000 national population) as of 2010. The number of people imprisoned in the USA is almost double that of China, which is ranked second on this measure, and more than ten times that of most Western countries. Other countries ranked highest in the number of people incarcerated include Russia, Brazil, India, and Iran (International Centre for Prison Statistics 2013). There is certainly something paradoxical about the USA, a democratic country that is a global hallmark of freedom and civil liberties, imprisoning more citizens than authoritarian regimes that are often criticized for human rights abuses.

This US exceptionalism has received a great deal of attention in the criminal justice literature as researchers have tried to explain mass incarceration in the USA in terms of political, economic, and cultural conditions specific to the USA (Christie 2000; Garland 2001; Tonry 2011). However, much less research attention has been given to comparative analyses of imprisonment that can situate the US experience within a broader global context. Such analyses bring greater attention to macrolevel historical, cultural, and institutional factors that are often overlooked in case studies.

One criminological axiom is that punishment trends are a reflection of structural forces and not simply a function of crime rates (Christie 2000; Garland 2001; Tonry 2011). Despite the research supporting this claim, the misperception persists that incarceration is driven by crime levels. This does not suggest that crime is completely unrelated to incarceration, but rather that there are several steps between crime commission and incarceration at the microlevel and numerous macrolevel factors that influence incarceration levels. For example, if you compare homicide rates internationally, which is often argued to be the most reliable indicator of cross-national crime differences, the US homicide rate is much lower than many other countries that have lower imprisonment rates (United Nations Office of Drugs and Crime 2011). This is because punishment is shaped by historically embedded preferences, values, and norms that are durable, yet malleable. Comparative research offers unique insights into the historical, structural, and cultural contexts of punishment. This entry illustrates how historical and comparative research on incarceration can shed light not only on US exceptionalism but more broadly on understanding why countries differ in their types and levels of punishment.

The Historical Context of Incarceration as Punishment

Formal punishment serves multiple purposes, including the direct punishment of an individual (i.e., inflicting pain), as well as the indirect demonstration of public definitions of acceptable social behavior. For much of human history, rulers relied on overt forms of brutal physical punishment to control populations, including whipping, ear cropping, branding, and other excruciating practices, as well as execution. Punishing the body left a physical mark (e.g., cut ear, missing finger, brand) so that others could easily identify and stigmatize the criminal. Additionally, physical punishments took place in public to send a message to other potential law violators that they too would receive a similar torture.

The brutality, corruption, and arbitrariness of early modern punishment fueled an egalitarian backlash demanding fairer treatment of individuals. Changes to the labor market also contributed to shifts in social control mechanisms. Specifically, brutal physical punishment no longer met emerging nineteenth-century economic demands for a large, able-bodied workforce that was needed to fill the factories emerging throughout the Western world (Rusche and Kirchheimer 1968). This new economic reality fostered the development of houses of corrections that were responsible for more than just punishing inmates. They were designed to correct and shape modern citizens through a reflective incapacitation process.

The USA was an early adopter of imprisonment, and the first prisons of the nineteenth century in New York and Pennsylvania emphasized this reflective incapacitation approach. Silence, separation, and introspection were seen as tools to teach pro-social values and compensate for the poor upbringings and lack of discipline that were believed to be the causes of criminal behavior.

However, structural changes in the late twentieth century brought about another shift in penal strategies. Many Western countries experienced changes in economic, social, and cultural frameworks that served as the *cultural preconditions* necessary for the development of a "culture of control" characterized by a growing reliance on criminal justice responses to more behaviors and longer prison sentences. These preconditions were accompanied by increasing crime rates that fostered a new perception of crime as a common occurrence that residents need to take proactive steps to prevent. The state was seen as incapable of protecting citizens from everyday predations because crime was now an embedded feature of the late modern social fabric. In the face of deindustrialization, growing unemployment rates, and inflation, the once flourishing central business districts of many cities experienced capital and human flight. These departures left only the most disadvantaged and marginalized groups behind. Instead of strengthening the social safety net, the US criminal justice system stepped in to regulate the resulting social disorganization and poverty (Garland 2001).

This historical context explains how incarceration became the primary means of punishment and social control in the USA, and it illustrates how social structural and cultural forces shape policy decisions regarding punishment. However, it does not explain how the USA became such an extreme outlier in the implementation of this form of punishment. How did the USA come to punish more people for longer periods of time? Comparative scholars began to recognize that there are unique legal institutional differences that shape different penal outcomes. Whitman (2003) provided an erudite historical analysis of punishment in France, Germany, and the USA to explain why three democratic, egalitarian, and capitalist nations sharing numerous cultural values could produce such different types of punishment. Interestingly, Whitman (2003) was not concerned with explaining the differences in the amount of punishment; rather, he focused on the differences in the types of punishment. The USA had distinctively harsh forms of punishment with many jurisdictions now using chain gangs, military-style boot camps, and double-bunking inmates, not to mention the forms of civil and social exclusion related to a criminal status. US inmates, for the most part, face felony voting restrictions, employment bans, and limited educational opportunities, such as being ineligible to receive financial aid to pursue a college degree. Whitman (2003) contrasts these exclusions with the Continental experience in which inmates maintain their privacy and dignity as they are housed in single-occupancy cells where guards must knock on doors before entering.

The crux of Whitman's argument is that French and German punishment mechanism consciously – yet slowly – moved away from stigmatizing responses. Criminal justice officials recognized the need to reduce the emotionally laden approach to crime control. Although there is much rhetoric among US policy makers of adopting evidence-based models of punishment, there are structural differences that make such attempts difficult. French and German prosecutors and judges are civil servants; they are experts that possess knowledge of human behavior and principles of behavior change. They have extensive and specialized training specifically related to their position, and the judiciary must act within a tightly bound system of accountability checks to ensure that punishments prepare inmates for normal life, not strip them of humanity. In contrast, prosecutors and judges in the USA have broad discretionary power regarding whom to charge, what to charge these individuals with, and what sentences to apply. The US plea bargaining system is relatively unthinkable in other countries due to the amount of prosecutorial discretion that it involves. Further, the US adversarial system is one that provides the judiciary with few boundaries during the sentencing phase of a trial; no doubt, sentencing guidelines do exist but these are presumptive. These distinctions were summarized by Weber (1967) as differences between types of legal thinking in which the Continental law relied on formal rational mechanisms and the common law relied on substantive rational mechanisms. To summarize these differences briefly, formal rational law attempts to eliminate individual discretion and subjectivity, whereas substantive rational law encourages legal actors to inject morality and ethics in legal decisions.

Tensions Between Modern Capitalism and Incarceration

The shift to an industrialized, capitalist economy not only changed labor market demands but also increased inequality and poverty, creating a marginalized class that could be potentially disruptive to social order. Governments have used a variety of tools to maintain order under such conditions. According to Bourdieu (1999), the left hand of the state controls the citizenry through less overt forms of coercion (e.g., education, welfare, and public healthcare), and the right hand of the state delivers direct control through police, courts, and prisons. At least since the end of World War II, Western nations have developed welfare programs to protect citizens from the inherent negative effects of capitalist production and have simultaneously bolstered criminal justice apparatuses to control the poor.

Beckett and Western (2001) argued that within the USA "penal and social welfare institutions comprise a single policy regime aimed at the problems associated with deviance and marginality" (Beckett and Western 2001, p. 44). The tremendous growth in US incarceration necessitates pondering who is occupying these prisons. Western (2006) provides an illuminating empirical analysis that shows how incarceration is driven by poor, uneducated, African-American males. This punishment trend is situated within a context of deindustrialization, stigmatizing subsistence-oriented welfare mechanisms and limited opportunities for the growing underclass left behind. Changing structural realities were embedded within a criminal justice culture lacking any real institutional blocks to increasing punishment. That is, US criminal justice actors did not push for rational and proportional sentences; rather, austere measures were the typical response because crime and criminals were seen as social pariahs, misfits, and incorrigible individuals. This is contrasted with Continental and Nordic perceptions of punishment in which criminals are a part and product of their social landscape – not disembodied from it. Criminality and crime control mechanisms fit within an inclusive context that provides a safety net from inequality as well as placing institutional barriers on criminal justice actors that prevent the reliance on overly emotional sorts of punishment. This is not to say that punishment lacks an emotive element. Rather, punishment takes a scientific, rational, and bureaucratic form. This form differs from the US response by recognizing that unpleasant features of social disorganization cannot be eliminated with incarceration nor can the rare, sensational, and highly publicized criminal events be prevented with new policies or longer sentences.

Comparative researchers demonstrate that the context in which governments, employers, and workers negotiate is consequential for incarceration rates. In particular, the ways in which governments set wages and determine labor arrangements to buffer the negative effects of capitalism – deindustrialization, unemployment, and underemployment – affect incarceration rates. Comparative research has shown that countries with corporatist bargaining, strong unions, and powerful left-oriented political parties have significantly lower incarceration rates than countries lacking these policies, such as the USA (Cavadino and Dignan 2006; Jacobs and Kleban 2003; Sutton 2000, 2004). Sutton (2000, 2004) concluded that across the advanced capitalist world criminal justice and welfare policies are historically intertwined as subthemes within a larger policy discourse about the management of social marginality. Furthermore, countries with a collectivist orientation have political-economic institutions in place that keep popular punitivism in check. Decentralized governmental structures increase the likelihood that the public's emotional desires for more punishment will creep into legal actor decision making (Whitman 2003).

Strong centralized efforts to protect labor interests keep incarceration rates down in at least two ways. First, the presence of collective bargaining and other political-economic frameworks that protect the working class from unemployment, underemployment, and universalizing welfare demonstrates interclass solidarity. Countries with such characteristics have a smaller gap between classes, an emphasis on full employment, and a social philosophy focused on diminishing inequality. Second, political-economic frameworks that protect the working class from economic uncertainty provide actual monetary benefits that prevent the development of the durable underclass that many have shown are driving the US mass incarceration movement.

This research clearly demonstrates that incarceration is not simply a response to crime. Instead, incarceration must be situated within a broader fabric of social control. This fabric is complex and varied and sometimes merges contradictory elements. Governments have varying preferences regarding diminishing gaps in inequality with wage bargaining and worker solidarity situated within left-political power, producing serious reductions in the reliance on incarceration as a social control mechanism in some countries. The USA, conversely, is a context in which policy makers, the public, and the media interact in a cyclical process of defining crime as something committed by the underclass, and criminals are in need of the state's "iron fist" through crime control policies centered on longer sentences, harsh treatment, and civic exclusions upon release. This approach fosters a path-dependent process creating policy inertia that makes it difficult to change course to more rational crime control responses that actually might reduce social inequality.

The Influence of Legal Traditions on the Use of Incarceration

The emphasis on state-labor relations in the comparative incarceration literature needs to be situated within historical legal traditions. The three legal families of the Western world – common law, Roman law, and Nordic law – codify and define enforcement strategies in different ways that reflect social preferences about state intervention that affect variations in the use of incarceration. Common law is a legal system that emphasizes common knowledge of facts brought to the court by private individuals. The contemporary adversarial system emerged from a private form of law that excluded defense attorneys from trials, lacked a public prosecutor, and did not allow judges to decide on guilt or innocence (Langbein 2005). The Roman legal system dates back to the Roman Emperor Justinian during the fifth century in which he commissioned legal scholars to devise a "gapless legal system" (Weber 1967). This system was institutionalized as an inquisitorial, code-based approach to eliminate legal actor discretion and lay influence in legal matters. The Nordic law combines elements of substantive (i.e., common law) and formal (i.e., Roman law) types of legal thinking with a cultural imperative focused on social solidarity and limited inequality. The shared histories and unique development of each legal family delineate legal theories, philosophies, and varying levels of discretion to legal decision makers.

Countries fitting into a specific legal family will adopt similar institutions to define and enforce state authority. This is not to suggest that all common, Roman, or Nordic law countries will be identical to one another; rather, these countries will have a high degree of interfamily similarity and high degree of intra-family variation. Simply, the USA and Australia, as common law countries, have similar legal practices that differ from their Roman law counterparts, say France and Germany, or from Sweden and Norway as Nordic law countries. It may seem out of place to discuss legal family or legal origin when explaining contemporary punishment outcomes, but this approach builds upon comparative economic research that shows how macroeconomic outcomes vary across countries according to their type of legal family (Hayek 1960).

Legal families define legal actor roles and structure financial markets in the form of rules of corporate ownership, stockholder protection, and private property rights that influence contemporary economic outcomes (Glaeser and Shleifer 2002, p. 1193). Legal institutions vary by legal family to reflect a country's use of law to protect property owners, enforce private contracts, and regulate the relationship between employers and employees. For example, Roman legal families use institutions that diminish legal actor discretion, rely on codified law, and prefer government intervention in financial markets. The common law approach, alternatively, uses institutions that favor judicial flexibility, case-based law, limited regulation, and private market solutions (Glaeser and Shleifer 2002). Comparative economic researchers find that the judicial flexibility of common law situates corporate actors within a context in which laws can be changed quickly according to social and economic preferences. This, essentially, shapes individual's willingness to rely on legal mechanisms of conflict resolution as there is a high expectation that legal decision makers will rule on the side of corporate interests.

This macroeconomic research is discussed because it – along with ample sociological literature (e.g., Weber 1967) – suggests that the rules defining and providing meaning to legal actor roles are consequential for punishment outcomes. Recently, (DeMichele & Janoski 2010) argued that similar legal institutions important for macroeconomics are consequential for criminal justice policies as well. Analyzing data from 1960 to 2002, he found that the common law countries had the highest amount of incarceration with average rates hovering between 125 and 150 per 100,000 in the population. The Roman law countries, composed of Continental European nations, had average incarceration rates between 80 and 100 per 100,000 in the population. And the Nordic law countries had the least incarceration with average rates ranging between 40 and 70 per 100,000 in the population. The unique blend of legal actor discretion, lay participation in the legal system, sentencing rules, and alternatives to incarceration coalesced to produce varying levels of incarceration. This research concluded that just as there are three types of capitalist welfare systems (Esping-Anderson 1990), there are also three types of Western punishment. Different legal systems produce environments that make possible the adoption of specific legal practices that reflect dominant definitions of the relationship between the state and individuals.

What Does Historical and Comparative Incarceration Research Tell Us About US Exceptionalism?

Historical research has put current incarceration patterns and trends in a larger context to understand how a confluence of economic, political, and cultural factors encouraged the use of incarceration as an important form of punishment and social control. Additionally, comparative research has shown how modern capitalism changed labor relations and exacerbated poverty and inequality in ways that also favored incarceration as a tool of social control. But why was this tool used more extensively in the USA relative to other advanced capitalist societies in the West? Criminologists have searched for institutional variation that differentiates the USA from other common law countries to point to the peculiarity of using popular elections to appoint many criminal justice and legal actors. That judges, prosecutors, and others are elected by the public may seem to be an inconsequential fact. However, it is well-known that during election cycles prosecutors and judges are found to implement longer sentences. Therefore, the common law relies on substantive practices that are more susceptible to public punitivism, and in the USA legal decision makers lack the bureaucratic protection found in other countries from these public desires for punishment. For example, consider the emotional reactions to sensational cases and how laws are created in the name of victims to gain public support. These institutional arrangements are situated within a US political-economic culture that favors individual responses to inequality, right-oriented politics, limited union strength, and decentralized wage bargaining – features known to correlate with increased incarceration. Mass incarceration, essentially, resulted from a social landscape that relies on formal punishment as the primary response to inequality embedded within a type of legal thinking predicated on the insertion of individual subjectivity and public emotionalism.

Future Directions

Although the literature explaining cross-national differences in the use of incarceration has not been extensive, it has certainly provided some important insights to explain US exceptionalism and also to delineate macrolevel forces that shape decisions regarding the tools of social control. Nonetheless, several questions regarding cross-national variations in incarceration rates remain unanswered. In particular, the prevailing focus on the use of incarceration in Western European countries provides a fairly narrow view of the question of why countries vary in their use of incarceration as punishment. The focus on the West is a common pattern in cross-national crime and justice research (Stamatel 2006) that has been exacerbated in the punishment literature because of the extreme US exceptionalism with respect to incarceration.

While Western European nations are natural points of comparison for the USA due to similarities in political and economic development, other countries outside of this region could provide interesting points of comparison for other reasons. For example, consider the fact that many countries with the highest incarceration rates per 100,000 population are former colonies, particularly British colonies, such as St. Kitts and Nevis, Anguilla, the Virgin Islands, or Bermuda (International Centre for Prison Statistics 2013). Are there commonalities in the colonization experiences of the countries, and the USA, that would explain similarly high usage of incarceration as punishment? Are their historical trajectories with respect to social control and punishment different from other formerly colonized countries with lower incarceration rates?

Another factor that is noticeably missing from the historical and comparative research on incarceration is religion. Tonry (2011) argues that Protestant fundamentalism is key to understanding harsh punishments in general in the USA. To what extent is this a common feature among other countries with high imprisonment rates? Is the type of dominant religion important for explaining cross-national differences in punishment or is fundamentalism the relevant explanatory factor regardless of type of religious foundation? The moral dimension of punishment that was so central in research on nineteenth-century practices seems to have been put on the back burner in the twenty-first century.

Lastly, the lack of diversity in comparative incarceration samples means that some types of legal systems have not been considered, most notably the socialist legal system that still has a strong influence on many Eastern European countries despite the fall of communism in that region or the hybrid legal systems that encompass much of the world, particularly countries with a history of colonization. These countries are interesting comparatively for two reasons. First, several countries with a socialist legacy have high incarceration rates, such as Russia, Belarus, the Ukraine, and Kazakhstan. Second, most socialist and hybrid systems, particularly in Eastern Europe and South America, have been undergoing significant changes in their legal systems (Newman 2010), raising questions about whether these changes will bring about more or less punitiveness.

Historical and comparative research is uniquely equipped to answer these types of questions. By contextualizing the historical, structural, and cultural contexts of punishment, we can identify and explain the macrolevel forces that affect punishment decisions and ultimately create cross-national differences in incarceration rates. Existing research has already shown how macroeconomic changes in labor markets, the creation of a permanent underclass, political configurations regulating state-labor relations, historical legal traditions shaping degrees of state intervention, and accompanying cultural perspectives regarding state strength, social control, and emotional gratification have contributed to different levels of incarceration across countries. This line of inquiry is still fairly new and many questions remain unanswered. Future research in this area has the potential not only to further explain US exceptionalism with respect to incarceration but also to situate the primacy of incarceration as punishment in a global context. This knowledge would be beneficial to other areas in criminology and criminal justice as it speaks to a number of important issues in the discipline, including social control, state control, cultural values, and the interrelationships among the criminal justice system and other social institutions.

Recommended Reading and References

Beckett K, Western B (2001) Governing social marginality: welfare, incarceration, and the transformation of state policy. Punishm Soc 3(1):43–59

Bourdieu P (1999) Rethinking the state: genesis and structure of the bureaucratic field. In: Steinmetz G (ed) State culture: state formation after the cultural turn. Cornell University Press, Ithaca

Cavadino M, Dignan J (2006) Penal systems: a comparative approach. Sage, London

Christie N (2000) Crime control as industry. Routledge Press, London

DeMichele M, Janoski T (2010) A theory of cross-national incarceration rate variation: three worlds of western punishment, 1960–2002. American Society of Criminology, San Francisco, California

Esping-Anderson G (1990) The three worlds of welfare capitalism. Princeton University Press, Princeton

Garland D (2001) The culture of control: crime and social order in contemporary society. University of Chicago Press, Chicago

Glaeser EL, Shleifer A (2002) Legal origins. Q J Econ 117:1193–1229

Hayek F (1960) The constitution of liberty. University of Chicago Press, Chicago

International Centre for Prison Statistics (2013) World prison brief. http://www.prisonstudies.org/info/worldbrief/wpb_stats.php?area=all&category=wb_poprate. Accessed 17 May 2013

Jacobs D, Kleban R (2003) Political institutions, minorities, and punishment: a pooled cross national analysis of imprisonment rates. Soc Forces 82:725–755

Langbein J (2005) Origins of the adversary criminal trial. Oxford University Press, New York

Newman GR (2010) Introduction. In: Newman GR (ed) Crime and punishment around the world, vol 4. ABC-CLIO, Santa Barbara

Rusche G, Kirchheimer O (1968) Punishment and social structure. Columbia University Press, New York

Stamatel JP (2006) Incorporating socio-historical context and temporal variation into quantitative cross-national criminology. Int J Comp Appl Crim Justice 30(2):177–208

Sutton J (2000) Imprisonment and social classification in five common law democracies, 1955-1985. Am J Sociol 106(2):350–386

Sutton J (2004) The political economy of imprisonment in affluent western democracies, 1960-1990. Am Sociol Rev 69(2):170–189

Tonry M (2011) Punishing race: a continuing American dilemma. Oxford University Press, New York

United Nations Office of Drugs and Crime (2011) 2011 global study on homicide: trends, contexts, data. http://www.unodc.org/documents/data-and-analysis/statistics/Homicide/Globa_study_on_homicide_2011_web.pdf. Accessed 17 May 2013

Weber M (1967) Max Weber on law in economy and society. Edited and annotated by Max Rheinstein. Simon and Schuster, New York

Western B (2006) Punishment and inequality in America. Russell Sage, New York

Whitman J (2003) Harsh justice: criminal punishment and the widening divide between American and Europe. Oxford University Press, New York

Historical Criminology

Paul Knepper
School of Law, University of Sheffield,
Sheffield, UK

Synonyms

History of crime

Overview

Historical criminology brings methods and concepts from history to the study of crime and criminal justice. This entry covers the differences between criminologists and historians, leading theoretical frameworks in historical criminology, methods of historical research, and the role of time in criminology theory and research.

Key Issues/Controversies

Criminologists and Historians

There are well-established areas of research that overlap history and criminology.

Manuel Eisner (2003) and Hans van Hofer (2011) analyze levels of crime over time. The study of crime trends makes intuitive sense to criminologists, particularly when these trends are examined from the Second World War. Trends in crime and imprisonment during the late twentieth century are readily interpreted using criminological models, so readily that criminologists do not regard this as "history." Criminologists are also interested in the history of criminology. This research examines the founding assumptions, theoretical traditions, and methods of criminological inquiry. Mary Gibson and Nicole Rafter have led the examination of Cesare Lombroso and biological criminology. They have explored the emergence of biological positivism in nineteenth-century Italy (Gibson 2002) and the popularity of criminal anthropology in the United States (Rafter 1992). Their work has advanced our understanding of Lombroso as more (and less) than the originator of the "born criminal" concept and raised new questions about criminology as a science of the human condition.

Nevertheless, there are real differences between criminologists and historians. For criminologists, historical research must do more than reveal the past. To be of use, history must be relevant to the present, meaning, there must be a clear statement about a present concern. In a sense, there is a commitment to "futurism." Criminologists try to build models that are good enough at explaining the present to anticipate what is likely to happen, and this provides a basis for advising crime policy (Lawrence 2012, p. 8). Historians are reluctant to comment on the present. They are more concerned with understanding the significance of past events against wider currents of human experience. They are happy to correct misrepresentations of the past (by politicians or others), but do not feel the need to offer policy recommendations.

Criminologists are committed to theory-driven research. They are particularly troubled by "empty empiricism," an assemblage of facts presented outside a clear conceptual framework (Loader and Sparks 2005, p. 14). While there is ample room for theory construction, and new theories of crime are always welcome, leading journals appreciate research developed within a theoretical tradition of criminology. Mathieu Deflem's (2002) investigation of international police cooperation deals with history, except that as a sociological criminologist, he constructs his book different than an historian would. He declares from the beginning that his work draws on the framework of Max Weber and that he will explore several propositions. Historians rely on theories to organize their facts, but do not specify them in the same way. Theories are not stated in the tradition of scientific research, but are rather weaved into the fabric of the narrative. Historians refer to concepts in the course of presenting their research material. They engage the arguments of other historians along the way: sharpening arguments, questioning evidence, and weighing interpretations.

Because the research training criminologists receive does not include historical research, most criminologists are not in a position to provide historical background into their topic of interest. The curriculum may include an option concerning the history of crime, but methods' requirements, even at advanced levels, do not include the craft of historical research and writing. Further, as social scientists, criminologists are suspicious of anecdotal data and wonder whether a few historical events can be stretched far enough to become social processes. In contrast, historians can choose crime as a topic of study rather more easily, and in fact, many have turned to crime in recent years. In Britain, there has been an "explosion" of scholarship concerning the history of crime (Godfrey et al. 2010, p. 18). The British Crime Historians have met every 2 years since 2008. Further, the American Social Science History Association and the European Social Science History Association operate networks for historians engaged in the study of criminal justice and legal history. Historians of crime, who are familiar with the aims and techniques of social history, are receptive to social science concepts.

And then there is the matter of style. Criminologists question the value of description. Historical material makes sense provided it is organized around criminologists concepts.

The sequence of events is secondary to the explanation of concepts and their connection to the criminological literature. Historians are more comfortable with narrative. Not all historians rely on chronology to organize their work, but they adhere to the belief that the sequence of events matters. Historians would not frame a research question as a "test of a hypothesis," so their writing does not follow the familiar sequence of literature review, research questions, data analysis, findings, and discussion. The technique of historical argument can be compared to constructing a dry stone wall. The builder fits all the available stones into a shape that holds together. In the same way, the historian must fit all the known evidence into a coherent pattern.

Theoretical History

Historical criminology could not have developed without the contribution of several theoretical frameworks which have built bridges between history and criminology. In the 1970s, the work of several British historians crossed over to criminology. The British Marxist historians, as they became known, pursued a "history from the bottom up" or "history from below," that is, the experiences of working class and ordinary people. *Primitive Rebels* (1959) and *Bandits* (1969) by Eric Hobsbawm, *Albion's Fatal Tree* (1975) by Douglas Hay and others, and *Whigs and Hunters* (1977) by EP Thompson examined the response to crime in the context of class structure and the emergence of industrial capitalism. Because these works engaged a conceptual vocabulary drawn from Marxist social theory, they found a ready reception among criminologists pursuing "radical" and "critical" approaches. Criminologists brought historical arguments about the rule of law and maintenance of social order to the contemporary critique of criminal justice policies. The British Marxist historians had excavated deep into the foundations of Anglo-American legal heritage. They offered insights into the administration of English criminal law in the eighteenth century that could apply to the policing of street crime in American cities of the late twentieth century.

About this same time, another influential book appeared: the English version of Michel Foucault's *Discipline and Punish* (1977). It was, ostensibly, a book about how Europe had come to rely on imprisonment. It inspired a great academic discussion about the "birth of the prison." But it was also more than this. In this book, and other work that followed, Foucault wielded an arsenal of images that captured the imagination of criminologists. Criminologists retold his examples (the execution of Damiens, Bentham's panopticon) and added his words ("discourse," "surveillance," and many more) to the lexicon of criminological terms. Historical studies framed as a: "history of the present" or "genealogical account" began to appear in criminology books and journals. Foucault made "appalling" mistakes about what happened, when, and where. But none of this mattered to criminologists who, particularly when it comes to history, prefer an entrée of ideas with only a small serving of historical details, perhaps as the apéritif. Foucault did not write a history of criminal justice; he turned criminal justice history into a worthy intellectual project.

Foucault even inspired further research by what he did not say. Foucault did not address the confinement of women, nor, for that that matter, did other widely-read historians of the prison. Nicole Rafter (1985) pointed out that it was a mistake to regard the history of women's imprisonment as an extension of the history for men. She presented the women's reformatory movement, the rise of custodial prisons for women, the legal and social characteristics of imprisoned women, and the effects of "race" and class on treatment of women (Rafter 1992). Work by Zedner (1998), and others, confirmed that gender could not be explained away as a lighter or softer version of men's experience. Historical research into institutions for women opened the doors to a parallel world operating according to its own logic, moving at its own pace, and leading to its own consequences.

David Garland took the "history of the present" to another level. In a series of influential books and articles, Garland produced an intricate history of changes in crime policy from the late nineteenth century. He examined institutional practices within wider social and cultural contexts with a particular eye for “moments of transition.” In *Punishment and Welfare* (1985), he examined how science entered legal culture before the First World War, and in *The Culture of Control* (2001), he analyzed the displacement of rehabilitation by security as the characteristic response of governments in the late twentieth century. Garland’s research opened up new issues for analysis: changes in prison practice, the origins of criminology, the role of the state, and convergence of welfare policy and penal policy. He generated a new conceptual vocabulary for historical study in criminology, such as “penal welfarism,” “mass imprisonment,” “the crime complex,” and “the culture of control,” and in the process, opened up new tracts of real estate for development by historical criminology.

The social theory of Norbert Elias has also been influential. In *The Civilizing Process* (1939), Elias recovered the transformation of European society from barbarism to civilization. Drawing on an interesting set of archival documents, commentary about manners, he charted the emergence of polite society. Through a process of cultural change so gradual as to seem invisible, control of violence shifted from external force to internal restraint. Elias’s theory of social change, sometimes called “process sociology” has been especially important in studies of penal policy and practices. Pieter Spierenburg (1984) and John Pratt (2000), two of Elias’s leading interpreters, have brought the civilizing process to punishment practices in the Netherlands and New Zealand. Mucchielli (2010) points to recent crime trends in France as evidence of an overall decline in violence. Crime figures in recent decades suggest Elias’s concept of the “civilizing process” at work, a fall in violent crime owing to the stigmatization or delegitimization of violence.

Archival Methods

Historical criminology borrows the methods historical research. Research in historical criminology relies on primary sources, documents, and materials that originate in the time period of interest, such as prison registers, police reports, and court records. Digitization has made significant material accessible for the historical study of crime such as the Old Bailey online project. The proceedings of the Old Bailey, London’s central criminal court, are available via the internet for 1674– 1913. Internet access has also made available an increasing number of books and periodicals, government documents, and newspapers from the nineteenth century and earlier. The vast majority of archival material, however, remains accessible only through traditional search strategies.

One of the challenges is avoid a narrow administrative history. Making a historical portrait from documents left by prison managers, police authorities, and so on presents a particular view, that is, a managerial or institutional view. The “usual suspects” – the poor, immigrants, the crowd – did not produce their own written records. To interrogate the deeper purpose and wider impact of institutions and practices, it is necessary to get beyond official records. It is true that much of the institutional record is “limited for anything other than administrative history” (Bosworth 2001, p. 434). Diaries of prisoners, letters from the accused, and the like are rare. But the official view can also be reframed with reports of visitors, commissions of inquiry, courtroom testimony, newspaper accounts, and even novels.

Another challenge is moving beyond the nation state. Criminal justice is a national or local project, and not surprisingly, much of the historical record concerns the response of national governments and city authorities. For the history of human trafficking, terrorism, drug trafficking, and other cross-border crimes, the records and materials left by nongovernmental organizations are particularly useful. These include organizations such as the Jewish Association for the Protection of Women and Girls, founded in 1885 in response to concerns that women emigrating from the Pale of Settlement were vulnerable to the white slave traffic (Knepper 2009) Other intergovernmental organizations, such as the League of Nations, also provide valuable material. The League of Nations led major campaigns concerning human trafficking and drug trafficking in the early twentieth century.

The archives in Geneva contain minutes of meetings, correspondence, field reports, photographs, and other materials (Knepper 2011).

Recent interest in "globalization" has invited new questions about the past, including an awareness of the legacy of empires. On paper, the British Empire represented one of the largest criminal justice systems in world history, with colonies, territories, and possessions extending from Australia to India, Ireland to Singapore. The Colonial Office attempted to coordinate police, prisons, and legal measures, and the influence was felt not only in the colonies but within Britain itself. Looking back yields insight into the diffusion of police and practices, and opens up new lines of inquiry, including the conditions for successful "policy transfer" and the influence of colonial practices on domestic institutions (Godfrey and Dunstall 2005).

Criminologists can gain a sense of history from published work by historians, but they cannot answer the questions they pose through secondary sources alone. Historians are interested in different questions, and the nature of historical research means that choices need to be made about sources and interpretation. Criminologists need to conduct historical research for themselves; they cannot proceed as if theory is all that really matters. Imposing a particular theory on the past yields a dull and pointless project, something like bringing one's own food to eat while on a holiday tour of several countries. As in social science, there is an interaction between theory construction and empirical findings. The historical researcher needs a theory to judge the significance of documents; the documents lead to new evidence, leading in turn to revising theories (Knepper and Scicluna 2010).

The Criminology of Time

Before Foucault, another French intellectual challenged social scientists to think about history. In the 1950s, Fernand Braudel encouraged social scientists to situate the objects of their inquiry in time. It was a mistake to believe that the significance of time could be gauged with a questionnaire or a few months of observation. Each social activity, he proposed, operates according to its own particular time, so that any historical moment is moving at multiple speeds.

In his history of the Mediterranean, Braudel presented three levels of historical time. He saw the natural environment, the mountains, deserts, and the sea, as the source of constraints on behavior that persist across generations. These behaviors produce the "almost immobile history" of geographical time. He described the "slowly rhythmic history…of groups and groupings," shaped by long-term trends in trade, transport, empire, and forms of war. This he referred to as social time. And, he narrated "history of short, nervous, rapid oscillations" brought about by political and military leaders. The "history of events" unfolded in individual time; individual time "has the dimension of their anger, dreams and delusions." (Braudel 1980, pp. 3–4) The fact that Braudel came up with his idea of multiple time while in prison during the Second World may suggest a particular relevance for criminology.

Criminologists have been interested in discontinuities, so much so that "paradigm shift" has become a cliché. Continuity, as suggested by geographic time, may be more important than we have realized. Media and crime affords one example. Given the development of electronic media since the Second World War, and the internet, satellite communication and so on in recent decades, it is easy to believe that the media has a powerful effect public understanding of crime, leading to particular crime policies. The influence of media on policy making about crime has been described as part of the "challenge of our times" (Garland and Sparks 2000). But as Robert Shoemaker points out with a look back at the eighteenth century, crime was a recurring topic in every form of print. As print become more available, popular understandings of crime were shaped more by what people read than by what they knew from personal experience or reports from friends and relatives. Pamphlets from social reformers exaggerated the extent of crime to justify particular projects. Newspapers overemphasized crime both in the frequency of reports and tone of coverage. Printed reports of trial proceedings distorted understanding of crime and justice, as much by what we left out as what was said. Edited versions of trials portrayed English court procedures as more coherent than they actually were. Eighteenth-century readers approached this surfeit of information with a skeptical eye; they knew that crime problems were more complicated than reformers, journalists, and the authorities made it appear (Shoemaker 2009). So, aside from the technology, what has really changed?

Braudel was interested in trends or cycles that occur over several decades, or what he referred to as "conjunctures." This is what LaFree had in mind when he urged criminologists to pursue historical analysis. There should be more emphasis on historical data and historical interpretation because theoretical insights cannot be obtained through cross-sectional designs that compare within a single moment of time. Crime events, such as the rise in crime during the 1960s and fall in the 1990s (in the United States), can be understood by recognizing that crime occurs within "distinct historical periods," and one of the goals of criminology is to understand factors behind the transition from one era to another (LaFree 2007). A recent example can be seen in theory of "crime waves." Periodization is a familiar idea to historians, and criminologists, who are accustomed to thinking about the importance of social context, are well-positioned to pursue it. Martin Killias (2006) examined fluctuations in crime rates across Europe with reference to "breaches," the way in which social developments surrounding technological advances open up new opportunities for crime. He combined the historical idea of periodization with the criminological idea of opportunity structure.

One of the criticisms of Braudel's approach, and particularly his emphasis on "structural history" grounded in continuity, is the absence of human agency. The agency-structure problem is well known to social science, and in examining biographies, criminologists have provided an interesting solution. In what must be the "longest longitudinal study," John Laub and Robert Sampson (2003) interviewed men some five decades after they had been selected for a study of delinquency. The interviews confirmed how individuals construct their own lives through the actions they take within the constraints of historical and social context. Criminal activities did not correspond with historical events, such as economic downturn, or social conditions, such as substandard schooling. Rather, in looking back over the course of their lives, the men saw the significance of choices they had made. Sampson and Laub (2005) encourage a view of time that balances structural constraints with human agency. Some historical events affect people very deeply. Others pass over, like a cloud across a quiet lake.

Concluding Thought

Criminologists have produced significant historical scholarship over the past few decades and, in recent years, have taken several steps toward incorporating historical criminology as a subfield of inquiry. Two presidents of the American Society of Criminology – John Laub in 2003 and Gary LaFree in 2006 – have used the annual presidential address to emphasize the value of historical perspectives (Laub 2004; LaFree 2007). In 2011, the Crime, Criminal Law and Criminology in History network was established with support from the Scandinavian Research Council for Criminology, and in 2012, the European Historical Criminology working group was established within the European Society of Criminology. It may be a step too far to suggest that historical criminology represents a recognized subfield. Criminologists and historians ask their own questions of the past, cultivate expertise in their own methods of inquiry, and organize their inquiries in different ways. But, from what has been done so far, one thing is clear enough: The most interesting criminology arises at the point that history and criminology meet.

Related Entries

- ▶ Historical and Comparative Perspectives on Incarceration
- ▶ History of Corrections

- ▶ History of Criminological Theories: Causes of Crime
- ▶ History of Geographic Criminology Part I: Nineteenth Century
- ▶ Quantitative Methods Used by Criminologists in the Nineteenth Century

Recommended Reading and References

Bosworth M (2001) The past as a foreign country? Some methodological implications of doing historical criminology. Br J Criminol 41:431–442

Braudel F (1980) On history. University of Chicago Press, Chicago

Deflem M (2002) Policing World Society: historical foundations of international police coperation. Oxford University Press, Oxford

Eisner M (2001) Modernization, self-control and lethal violence: the long-term dynamics of European homicide rates in theoretical perspective. Br J Criminol 41:618–638

Eisner M (2003) Long-term historical trends in violent crime. Crime Justice A Rev Res 30: 83–142

Garland D, Sparks R (2000) Criminology, social theory and the challenge of our times. Br J Criminol 40:189–204

Gibson M (2002) Born to crime: Cesare Lombroso and the origins of biosocial criminology. Praeger, Westport

Godfrey B, Dunstall G (2005) Crime and empire 1840-1940. Willan, Devon

Godfrey B, Lawrence P, Williams C (2010) History and crime. Sage, London

Killias M (2006) The opening and closing of breaches: a theory on crime waves, law creation and crime prevention. Eur J Criminol 3:11–31

Knepper P (2009) The invention of international crime: a global issue in the making, 1881-1914. Palgrave, London

Knepper P (2011) International crime in the twentieth century: the league of nations era, 1919-1939. Palgrave, London

Knepper P, Scicluna S (2010) Historical criminology and the imprisonment of women in nineteenth century Malta. Theor Criminol 14:407–424

LaFree G (2007) Expanding criminology's domain: The American Society of Criminology 2006 presidential address. Criminology 45:1–31

Laub J (2004) The life course of criminology in the United States: The American Society of Criminology 2003 presidential address'. Criminology 42:1

Laub J, Sampson R (2003) Shared beginnings, divergent lives: delinquent boys to age 70. Harvard University Press, Cambridge, MA

Lawrence P (2012) History, criminology and the 'use' of the past. Theor Criminol 16(3):313–328

Loader I, Sparks R (2005) For an historical sociology of crime policy in England and Wales since 1968. In: Matravers M (ed) Managing modernity: politics and the culture of control. Routledge, Oxford, pp 5–32

Mucchielli L (2010) Are we living in a more violent society? A socio-historical analysis of interpersonal violence in France, 1970s-present. Br J Criminol 50:808–829

Pratt J (2000) Civilization and punishment. Aust N Z J Criminol 33:183–201

Rafter N (1985) Gender, prison and prison history. Soc Sci Hist 9:233–247

Rafter N (1990) Partial justice: women, prisons and control. Transaction, New Brunswick

Rafter N (1992) Criminal anthropology in the United States. Criminology 30:525–546

Sampson R, Laub J (2005) A life-course view of the development of crime. Ann Am Acad Pol Soc Sci 602:12–45

Shoemaker R (2009) Print culture and the creation of public knowledge about crime in 18th century London. In: Knepper P, Doak J, Shapland J (eds) Urban crime prevention, surveillance and restorative justice: effects of social technologies. CRC Press, Boca Raton, pp 1–22

Spierenburg P (1984) The spectacle of suffering. Cambridge University Press, Cambridge

Von Hofer H (2011) Punishment and crime statistics in Scandinavia, 1750-2008. Crime Justice A Rev Res 40:33–108

Zedner L (1998) Women, crime and custody in Victorian England. Clarendon, Oxford

History of Bail

Amanda L. Russell and Robert G. Morris
The University of Texas at Dallas, Richardson, TX, USA

Synonyms

Pretrail detention; Pretrial release

Overview

Although not traditionally viewed as a function of corrections, bail is an integral mechanism within the system that can allow a defendant to avoid the confines of jail in the early stages of the criminal justice process (i.e., before trial). However, the bail system – particularly in the United States – is frequently viewed as punitive for those who are denied bail as well as for those who cannot secure bail through financial means. In particular, individuals who are denied bail are immediately thrust into the correctional system despite an entitlement to a presumption of innocence. Meanwhile, indigent defendants are penalized for their inability to pay the requisite bail amount. In both scenarios, the punishment (of pretrial detention) occurs before guilt has even been determined.

Bail is commonly defined as the temporary release of an arrestee secured by a bond or promise to appear at future court hearings. Bail, as it is utilized today in the United States and other countries, is rooted in the long-standing practice of using personal sureties as an alternative to pretrial detention.

The term is a derivative of the French word *baillier* – a denotation of the verb "to deliver." In historical practice, it refers to the pretrial custody transference of an arrestee, from the court, to others who then become responsible for that individual's fulfillment of his responsibilities to the court ("Indemnification Contracts" 1949). These individuals are typically referred to as sureties because they function as guarantors of the arrestee's appearance. In some cases, financial stipulations are attached that must be satisfied by the arrestee, a personal surety, or third party. These monetary provisions may be required up front before bail is granted or may be later required from the arrestee or surety in the event that the individual fails to appear in court. Despite the many nuances of bail systems from one country to another, the practice has not changed significantly since its reported inception in the seventh century as a mechanism of pretrial release. A review of the extensive history of bail, for example, reveals that the theoretical basis underlying bail has essentially remained the same even into the twenty-first century. In other words, bail continues to be used as a deterrent to pretrial flight. Sureties, in particular, are believed to be motivating elements that ensure a person's appearance in court. What has changed, however, are the conditions in which bail is granted, the types of bail and the means to obtain it, and the purposes to which bail is intended to serve.

In the United States, specifically, the bail system has evolved into one that must reconcile conflicting objectives. In particular, the system must protect the presumption of innocence while also ensuring that individuals who are released return to court, abstain from crime, and maintain the sanctity of the case (Devine 1991; Goldkamp and Gottfredson 1979). Bail in the United States has also developed into an institution that emphasizes financially based means to obtain pretrial release. These are commonly referred to as secured bonds. In cases such as these, the judge or magistrate sets a bail amount to be paid by the arrestee. Here, the assumption is that the individual will opt to return to court rather than to forfeit the money or collateral posted to obtain release before his trial (Ryan 1967). Secured bonds, in fact, are one of the most widely used mechanisms of pretrial release in the United States. A report provided by the Bureau of Justice Statistics (BJS) indicates that of those released between 1997 and 2004, 48 % of felony defendants in state courts were released through secured bail, including surety, deposit, and collateral bonds. The remaining 52 % were released without the requirement of financial conditions (e.g., release on own recognizance, conditional release, unsecured bond) (Cohen and Reaves 2007). While the extensive use of secured bonds in the United States is contrary to how bail systems operate in other countries, such as England (Bottomley 1968), Norway, Denmark, and Germany (Foote 1965), the exchange of money or property for pretrial release is not unique to the United States or to modern-day systems of bail. Court reports from medieval Elton, Huntingdonshire (England) suggest that many arrestees provided payments to their sureties, or pledges, in return for their services (Pimsler 1977). Moreover, countries like England and Canada may impose financial conditions of bail in certain cases. The United States, however, is particularly exceptional in terms of its use of commercial bail bonds (Devine 1991), in which bail is secured through an arrangement between a professional bail bond agent and the arrestee for a fee – a development that reportedly occurred in the nineteenth century and has since become a standard of the American bail system.

This increasing reliance on secured bonds in the United States has raised a number of concerns, of which a majority has been voiced in the last century by individuals seeking bail reform. Critics suggest that the American bail system has strayed from the intended purpose delineated by the country's founders and has deviated from the criminal justice system's traditions of the presumption of innocence and equal protection under the law. Yet, despite its opponents, bail in this country continues to drift from its long-standing principles, whereas bail systems in other countries remain focused on nonfinancial factors and the use of personal sureties.

Early Origins of Bail

Bail is considered one of "the most ancient of Anglo-American criminal justice traditions" (Advisory Commission on Intergovernmental Relations [ACIR] 1983, p. 54). Its roots in Anglo-Saxon history serve as a common denominator that brings together the laws and systems of several countries, including England, the United States, and Canada (Metzmeier 1996). Although the precise origins of bail are unknown, many scholars concede that it originated at some point in medieval England. Some have traced it as far back as to the time of Charlemagne in the eighth century and the practice of hostageship. Hostageship was used as a tactic of war in which a person was held captive until the "principal" appeared. If he did not show up, the hostage suffered the fate intended for the principal ("Indemnification Contracts" 1949). DeHaas (1940), however, suggested that bail stems from Anglo-Saxon laws during the reigns of King Hlothaere and Eadric in the late seventh century. During this time, the accused had to pay a "bohr" to the victim's family; the money was returned, however, if he was found innocent of the crime. The use of this early form of pretrial release was not the consequence of forward-thinking sovereigns, but rather it stemmed from the realization that imprisonment was an expensive and problematic way to deal with accused criminals (Duker 1977). Others believed that the modern system of bail began in the ninth or tenth century due to the problems experienced in England's early criminal court system. At that time, the sheriff assumed primary responsibility for the detention of the suspect after the individual's arrest. It is likely that in some cases, the suspect had to stay in jail for considerable periods of time – even years – because of geographical issues and the judicial district assignments of the traveling judges ("Bail: An Ancient Practice" 1961). As such, the sheriff would allow the accused to pay a cash bail or utilize personal sureties in return for the suspect's promise to appear for trial, and, in some instances, he would release the accused on his own recognizance (Pollack and Maitland 1898).

The use of sureties, as noted previously, was reported to have existed some time before the Norman Conquest of 1066 in which individuals agreeing to be sureties were subjected to a formal contract of payment, in which the price was set at the value of a free man's life (Hazeltine 1910). Sureties were also used in the medieval practice of pledging. In this case, pledges were responsible for the assurance of other behaviors aside from appearance at trial, such as payment of court fines and repayment of a debt. Even still, it appears that the typical function of pledges was the equivalent of the modern-day bail bondsmen; in return for their services, it is likely that the pledgees paid their sureties (Pimsler 1977).

Generally, most writings on the history of bail attribute its modern-day traditions to the time of William I after the Norman Conquest. Prior to the thirteenth century, sureties were bound "body for body" by the court ("Indemnification Contracts" 1949). This meant that if the defendant did not appear at trial, the surety was then subjected to the punishment that would have been inflicted upon the suspect (e.g., enslavement or execution). In later years, however, this requisite was relaxed, such that the surety was responsible for a financial penalty if the accused did not appear in court, thus combining the older tradition of the responsibility of the surety with a different method to impose it (Pollack and Maitland 1898). Bail, therefore, developed into a formal bond by which the accused became the custody of the surety, and the surety promised to pay the penalty if the accused fled.

Statute of Westminster, 1275

Perhaps the most notable turning point in the evolution of modern-day bail was the Statute of Westminster in 1275. Prior to the thirteenth century, sheriffs had sovereign powers in the release or detention of accused individuals. Any dealings between the sheriff and the accused were not circumscribed by law. Consequently, they could use whatever factors they wanted in making this determination of pretrial release. The Statute of Westminster, therefore, endeavored to institute a standardized practice of bail. Yet, rather than providing a general schema to guide all decisions about bail, Parliament defined which crimes were bailable and which crimes were not subject to bail (Bottomley 1968). The statute also limited the authority of the sheriff by delineating under what conditions release was sanctioned. Interestingly, sheriffs' authority to determine bail amounts remained intact.

The provisions of the Statute of Westminster persisted for nearly five centuries thereafter. Although several statutes were passed between that time, they did not bring about any significant modifications to the fundamental principles of when or when not bail should be granted. Instead, the statutes subsequent to that of Westminster affected the processes and procedure underlying the granting of bail (Bottomley 1968). For instance, following the Statute of Westminster, a law was passed in 1487 due to issues in the release of nonbailable suspects. In response, Parliament required the use of two magistrates in bail granting proceedings. This law, however, did little in terms of correcting the problem, as it was realized that one judge was releasing suspects without the second judge's knowledge of the case. Therefore, in 1554, Parliament reaffirmed the previous statute's requirement but mandated that the bail decision be made in open session with both magistrates present and that the evidence considered in bail decisions be documented on paper.

Although these events document the historical use of bail, it is unlikely that bail was extensively utilized because pretrial detention itself was used infrequently (Pollack and Maitland 1898). As mentioned previously, imprisonment was an expensive and troublesome method of dealing with suspects, and to avoid responsibility of the accused, the sheriff would often entrust him to a friend or relative. Even when individuals were initially jailed, but could not produce the necessary surety, there is evidence that the accused was released. In short, there was a likely disconnect between what was written in law and what was actually practiced (Foote 1965).

Beginnings of Major Bail Reform in England

Although progress had been made in terms of bail reform, the restrictions placed upon sheriffs via the Statute of Westminster 1275 were not applicable to judges until the late 1600s in "Darnell's Case." In 1626, King Charles I sought and failed to obtain money from Parliament, thereby demanding loans from his subjects. After refusing to provide him with funding, numerous individuals were imprisoned without notification of the charges by the King. Of these, five knights responded by filing a *habeas corpus* petition, claiming that their imprisonment was unlawful and violated not only their right to bail but also their right to trial. Citing "sovereign prerogative," the court upheld the King's decision (Duker 1977). Darnell's Case, consequently, served as the impetus for the Petition of Rights later passed by Parliament in 1628. The Petition acknowledged the rights guaranteed under the Magna Carta and attempted to provide assurances that individuals could not be imprisoned without due process of law nor could they be detained before trial without being informed of the specific charges. Despite Parliament's efforts, problems persisted, eventually leading to the passage of the Habeas Corpus Act of 1679. The act was intended to protect an individual from unlawful detention; it did not, however, address the issue of excessive bail. Thus, judges were using undue financial conditions of release to detain individuals for indefinite periods of time. As a result, Parliament enacted the English Bill of Rights of 1689, famously mandating that "excessive bail ought not to be required." Together, the Petition of Rights, Habeas Corpus Act, and the English Bill of Rights formed the modern-day safeguards against pretrial imprisonment (Foote 1965). These concepts were later translated into American colonial law and eventually adopted by the Founding Fathers in the United States Constitution.

Bail in Colonial and Early America

English bail laws of the seventeenth century were considered "important matrices" that greatly informed subsequent American criminal law and procedures (Foote 1965, p. 973). As such, it is hardly surprising to find that the issue of bail was not one of contention at the time of the American Revolution. In fact, the prohibition of pretrial imprisonment was already in place in several colonies even before the English Bill of Rights. Massachusetts, for example, had enacted the state's Body of Liberties in 1641, considered by Foote (1965) as a "major benchmark in American bail history" (p. 975). This statute was the first in colonial America to delineate a number of individual rights and liberties, many of which were later referenced in the US Bill of Rights. Regarding bail and pretrial detention, the Body of Liberties states that an individual should not be detained by a judge "before the law hath sentenced him thereto, if he can put in sufficient security, bail, or mainprise, for his appearance and good behavior in the meantime, unless it be in capital crimes...." Therefore, unlike in England, where the Statute of Westminster determined which crimes were bailable, this statute provided the right to bail to all individuals not accused of capital crimes.

The value of liberty and the presumption of innocence in early America were evident not only through the Massachusetts Body of Liberties but also through the Frame of Government of Pennsylvania and the New York Charter of Liberties and Privileges of 1682 and 1683, respectively. The Pennsylvania law, in particular, had a major impact in later years as it reemerged in North Carolina in the late eighteenth century and largely found its way into many state constitutions in subsequent decades. This law was particularly important given its advisement of the right to bail in all noncapital cases (Foote 1965). Pennsylvania's Frame of Government also was the archetype for the bail provisions delineated in the Northwest Ordinance of 1787. With only slight modifications in verbiage, the ordinance stated, "All persons shall be bailable, unless for capital offenses, where the proof shall be evident, or the presumption great." By 1789, the absolute right to bail movement was well under way (Foote 1965), and by 1850, most state constitutions included the right to bail (Metzmeier 1996).

The significance of early English and American colonial law is observable not only in state constitutions but also in the US Constitution. Interestingly, the federal provisions of bail were addressed in "separate legislative packages" in 1789 (Foote 1965, p. 971). Following the English Bill of Rights and the Virginia Declaration of Rights in 1776, the excessive bail clause prohibiting high bail amounts was proposed as a part of the Eighth Amendment of the Constitution, whereas the right to bail was addressed in Section 33 of the Judiciary Act. The Eighth Amendment references only the use of excessive bail amounts and the prohibition thereof; it does not, however, grant an absolute right to bail in all cases nor does it define which crimes are bailable. In Foote's (1965) extensive analysis of the historical context of bail, he concluded that this omission was accidental and one that went unnoticed by Congress. Nevertheless, the Judiciary Act accounted for the omissions in the Constitution regarding bail. Specifically, the act established which offenses were nonbailable (e.g., capital crimes). In capital cases, judges were left to determine whether the suspect must be detained prior to his trial. Through the act, Congress also limited judicial discretion in setting bail. These collective efforts of the first Congress remained largely intact for nearly 150 years thereafter.

Bail in the Nineteenth and Early Twentieth Centuries

Though major developments occurred in eighteenth-century America in terms of bail reform, little changed occurred to the bail laws in England after the Bill of Rights of 1689, that is, until 1826. At this time, a law was enacted that mandated the right to bail in cases where the likelihood of the suspect's guilt is slight, but that bail should be avoided in cases where the probability of guilt is significant. Only 9 years later, this statute was repealed in place of a new law that granted bail for any crime, regardless of the presumption of guilt. This change in statute was particularly important because it reduced bail decisions to the sole determination of risk of nonappearance (Bottomley 1968). The court in *R. v. Robinson* (1854) not only substantiated this purpose of bail but also provided a guideline from which determinations of pretrial flight risk could be made. This included three questions that pertained to the nature of the offense, the likelihood of conviction, and the punishment that would be required if convicted. The "Robinson factors" were used in making bail release decisions until bail reform efforts in 1976 (Metzmeier 1996). The nature of the offense was also determined to be an important factor in bail decisions through the ruling in *United States v. Lawrence* (1835). In this case, the Supreme Court contended that the judge should consider the seriousness of the offense and the accused's ability to post bail. This statement of the court also reveals two other important elements: judicial discretion in bail decision making and consideration of the individual's capacity to obtain bail.

The Emergence and Proliferation of Commercial Bail

Before the nineteenth century, bail in both the United States and England was obtained primarily through the use of sureties with lesser concerns about cash bail. In England, personal relationships were considered to be of great importance in assuring the appearance of the suspect ("Indemnification Contracts" 1949). In the United States, however, cash bail later became the primary means of obtaining release. It is likely that several events converged to bring about the extensive use of this type of bail. First, the social context of that time had changed. Urbanization and the mobility of the population increased, thereby leading to a weakening of community ties. The American frontier presented issues as well. Personal sureties were not easy to come by for a suspect in a new location with few acquaintances and people willing to become sureties (Ryan 1967). Furthermore, it is likely that the expansive and transitory aspects of the frontier provided an incentive to flee, especially if the likelihood of conviction was great ("Bail: An Ancient Practice" 1961). Overall, the old surety system in the United States was rendered impractical for a mobile population. Collectively, these factors, therefore, gave rise to the role of the commercial bail bondsman as an alternative to the personal surety.

In general, commercial bail is an arrangement between a bail bondsman and the accused. For a fee, usually between 5 % and 10 % of the bail amount, the bondsman acts as a surety on behalf of the accused. In theory, if the suspect fails to appear, the bondsman forfeits the bail amount required by the court. Eventually this type of bail all but replaced alternative forms of bail (e.g., surety bond and release on recognizance), such that even well-connected individuals were being released in this manner (Ozanne et al. 1980). The prevalence of money bail and the bondsman were further realized after insurance companies began to serve as underwriters to the bondsmen ("Bail: An Ancient Practice" 1961). The changing climate of the bail system was noted by Justice Holmes in *Leary v. United States* (1912) who stated, "The distinction between bail and suretyship is pretty nearly forgotten. The interest to produce the body of the principle in court is impersonal and wholly pecuniary."

Not long after its inception, the commercial bail bond system was the subject of numerous accusations, including collusion, bribery of public servants, and conducting illegal business (Paulsen 1966). Moreover, some individuals grew concerned over the power and authority readily bestowed upon bondsmen. Although the judge sets the bail amount, often times it is the bondsman who ultimately makes the release decision. In other words, he decides whether or not he will "act as surety" for those considered a good risk (i.e., individuals who are able to pay the requisite fee); thus, those who cannot pay the fee remain in jail (Paulsen 1966, p. 115). Given the power of the bondsman, therefore, it is argued that suspects can effectively thwart the purpose of bail (Ozanne et al. 1980). Additionally, the commercial bail bond system was criticized for its ineffectiveness in ensuring that the accused was returned to court. Even though it would appear to be in the best interests of the bondsman to ensure his client's appearance as a means to avoid forfeiture of the bail amount, it was found that in some cases, the court waived the bondsman's forfeiture or even refunded a portion of the forfeited amount. In these instances, the court claimed that the bondsman used "reasonable diligence" in his efforts to produce the suspect ("Bail: An Ancient Practice" 1961, p. 968). Yet, in spite of the numerous complaints of the commercial bail bond system, this system's use was substantiated in *Nicholls v. Ingersoll* (1810) and *Taylor v. Taintor* (1873) in which the powers of the bail bond agent and bounty hunters were delineated by the court.

Inequities of Bail

As the reliance on money bail began to take shape, injustices throughout the bail system also began to emerge. Indigent defendants, in particular, were perhaps hardest hit – oftentimes detained before trial for periods of time that likely would have exceeded their imposed sentence had they been convicted (Metzmeier 1996). Three early studies on bail exposed the inequity in the bail system, demonstrating that bail decisions were based more often on financial means as opposed to the individual's risk of flight. Consequently, poorer individuals were more likely to be detained than wealthier individuals (see Beeley 1927; Foote 1954; "A Study of the Administration of Bail" 1958). Paulsen (1966) later described the bail system in the United States as "a costly failure, pressing hardest on the poor" (p. 125).

Bail Reform in the 1960s

Prior to the 1960s, several scholars had already begun to question the efficacy and fairness of the bail system in the United States. In the 1920s, Pound and Frankfurter (1922) protested the use of commercial bail bonding, arguing against its tendency "to prostitute the administration of justice in the inferior courts" (p. 292). Yet, the use of commercial bonds persisted. Shortly thereafter, Beeley's (1927) study of bail in Chicago revealed the extensive use of surety bail as well as a number of problematic issues, particularly in terms of bail decision making – which often excluded financial and social factors. In their place, judges were making decisions based on the nature of the crime. This study also shed light on bail-setting amounts, which appeared to be largely indiscriminate. Follow-up investigations into the bail system were conducted in Philadelphia and New York in the mid- to late 1950s. Researchers found similar results to the Chicago study, despite these studies taking place decades later. Moreover, they observed differences in the outcomes between pretrial detainees and individuals who were able to obtain release. Specifically, they found that pretrial detainees were more likely to be found guilty compared to those released on bail; in addition, they were less likely to receive lighter sentences, such as probation or a suspended sentence (see Foote 1954; "A Study of the Administration of Bail" 1958).

Despite these findings, little change occurred. Friedman (1976) suggested that the lack of change was due partly to the nonexistence of major research organizations working as part of a coordinated criminal justice system; thus, there was no central agent or organization to put together a plan of action or to effectively lobby for change. Equally problematic was the "wholly passive role" of the courts regarding bail despite the mounting evidence of the system's inequities (Foote 1965, p. 959). However, in the Supreme Court case of *Stack v. Boyle* (1951), Justice Jackson addressed this issue, stating that "the spirit of the procedure is to enable them to stay out of jail until a trial has found them guilty." Further, he wrote, "In allowance of bail, the duty of the judge is to reduce the risk by fixing an amount reasonably calculated to hold the accused available for trial and its consequence."

Nevertheless, it was perhaps due to the changing social climate in the 1960s that a concerted effort toward bail reform began to emerge. During this time, scholars, particularly those in the legal realm, grew skeptical of the bail system (Kennedy 1980). Not only were they concerned with the prevalence of money bail and the role of the bondsman, they also voiced concerns in regard to the seemingly disparate applications of bail and the detrimental effects of pretrial detention upon the indigent (e.g., the individual's loss of freedom, loss of job, and inability to help with his defense). Although bail was originally intended to serve as a means of release, it was argued that the bail system worked in such a manner that it proved more effective in detaining suspects rather than in freeing them (Foote 1965). The reality was that many bail decisions were based on an incomplete, if not absent, assessment of the defendant, despite claims that the following factors be considered in the bail decision making process: a person's ties to the community, likelihood of guilt, and financial ability to post bail. Instead, these decisions were frequently based on the nature of the offense. Consequently, a significant number of individuals were detained before trial regardless of their risk of flight.

The reality of this situation served as the impetus for the Vera Foundation and the Manhattan Bail Project and the emergence of pretrial service agencies in the 1970s. The Vera Foundation and the Manhattan Bail Project, specifically, were intended to address the issue of pretrial detention of impoverished defendants in New York City (Friedman 1976) but eventually inspired the major bail reforms that occurred during this decade. At a cocktail party in 1960, Louis Schweitzer – a chemical engineer and businessman – was surprised when he discovered that thousands of people in New York City were imprisoned before they had even gone to trial. After a tour of the jail facilities in the city and discussions with key players about jailed suspects who could not afford to post bail, Schweitzer offered his friend, Herbert Sturz, the task of developing a strategic plan (Friedman 1976). Thus, the Vera Foundation was created in 1961, followed shortly that same year by the commencement of the Manhattan Bail Project. Underlying this project was the assumption that nonfinancial factors were an effective alternative to money bail in terms of deterring the accused from flight (Ryan 1967). Therefore, the Vera Foundation, along with assistance from New York University law students, implemented one of the first control group experiments conducted in an American court. The Manhattan Bail Project was devised to first gather information about the individual's background and community ties from the accused and, second, to provide the judge with recommendations as to who would be a good candidate for release on personal recognizance. By the study's end in 1964, the court had granted release on personal recognizance to 60 % of the defendants recommended by the Vera Foundation. Only 1 percent of these defendants failed to appear.

The success of the Manhattan Bail Project sparked national interest and prompted a number of subsequent bail projects in other cities across the country. In light of the study's findings, Attorney General Robert Kennedy formed the Allen Committee to serve as an advisory board to the United States Department of Justice. In 1963, the Allen Committee published a report, recommending that the federal system broadens its use of release on personal recognizance. Yet, the report failed to include decision making guidelines addressing who should be released in this manner (Friedman 1976). Later that same year, efforts were made to conduct the first National Conference on Bail and Criminal Justice. Attended by over 400 policymakers, criminal justice officials, attorneys, and practitioners in May 1964, the conference focused on specific alternatives to money bail, keeping in mind the findings from the Manhattan Bail Project. According to Friedman (1976), there was a significant correlation between the conference and the rate at which other bail agencies grew. After the bail study in New York, 89 bail projects had been implemented.

The impact of the Manhattan Bail Project also spreads to other countries, including Canada and England. Inspired by the project, Canadian reformer Martin Friedland took to task the study of bail in Toronto's court system. Additionally, reform efforts in England eventually led to the Bail Act of 1976 (Metzmeier 1996). In the United States, Congress enacted its own bail reforms through the Bail Reform Act of 1966. This act reaffirmed what had been addressed in previous statutes but went a step further to "order" that the accused in a noncapital case be released under minimal conditions (e.g., personal recognizance or personal bond) unless it was decided by the judge that either method would be insufficient in guaranteeing the accused's return appearance. The statute also advised against the consideration of community safety in noncapital cases. Instead, this element should only be considered in capital cases or after the individual has been convicted.

Additional Reform Efforts

Although not as popular, the Manhattan Summons Project borrowed largely from its predecessor. This study, however, focused on an earlier stage in the process, particularly at the stage of arrest. In this case, release was to be determined by the police. The Summons Project commenced in the spring of 1964 at the New York City Police Department's 14th precinct. The study was designed in such a way that an individual arrested for a summonable offense was first interviewed by a member of the research team. Information, similar to that of the bail study, was collected and verified, and then recommendations were made to the precinct's desk officer. It was recommended that either a summons be issued (with instructions to appear in court 14 days later) or that the individual proceeds through the traditional criminal justice process (i.e., arraignment). The results of the three-month study were so successful that the project was eventually expanded to three precincts and then expanded to all of Manhattan. In the first 2 years of the project, over 36,000 summons were issued. Of those, only 5.3 % of the individuals who received a summons failed to appear in court (Friedman 1976).

Additional efforts to reform the bail system were undertaken in Illinois. In 1963, Illinois enacted the state's Ten Percent Deposit Plan as an alternative to financial surety bonds. Rather than going through a third party to gain release, individuals pay ten percent of the bond amount to the court, thus eliminating the need for bond agencies. The Illinois Code of Criminal Procedure was also amended to emphasize that criminal sanctions, rather than financial penalties, be used to guarantee a suspect's appearance (Paulsen 1966). Since that time, three other states have followed suit and eliminated commercial bonding (e.g., Kentucky, Oregon, and Wisconsin); others have experimented with similar deposit bail systems, such as in Massachusetts.

Setbacks in Bail Reform

While strides were being made in terms of bail reform to (1) reduce the reliance on money bail and (2) make bail decisions based only on flight risk, other events were taking place that contradicted these reforms. In *Hinton v. United States* (1965), for example, the court provided justification for the use of money bail, stating that "it has been the Court's experience that indigent defendants are often in the community on bond, having obtained the means therefor [sic] from family or friends." The court further argued that the bond provided by the suspect's kinship served as a deterrent to flight because the accused has an "obligation" to those ties as well as to the court.

Setbacks were also experienced in 1969, with efforts to amend the Bail Reform Act of 1966. Despite observations by the 1967 President's Commission on Law Enforcement and Administration of Justice that too many people were being detained and that release decisions were not based on fact, the United States Department of Justice submitted a proposal to Congress to amend the Bail Reform Act of 1966. The proposal recommend that federal courts be permitted to detain suspects for up to 60 days, particularly individuals accused of certain violent crimes and those who are deemed dangers to the community despite the strictest conditions of release imposed. Furthermore, for individuals not subjected to this provision, it was recommended that the court be allowed to consider the suspect's dangerousness to the community when making decisions regarding pretrial release conditions (Mitchell 1969).

The proposal by the Department of Justice was only one of many attempts in 1969 to amend the 1966 Bail Reform Act to include a provision of preventive detention. None of these bills were passed by Congress (Kennedy 1980). However, growing concern was being voiced regarding the possibility of dangerous suspects being released into the community under the act. Congress, in turn, allowed for the inclusion of a preventive detention provision in the District of Columbia Court Reform and Criminal Procedure Act of 1970. Suspects meeting any of the three criteria for pretrial detention were subject to a pretrial detention hearing. At the hearing, the judge must make a determination that no release conditions will "reasonably ensure" the community's safety. He must also find that there is a significant likelihood of the suspect's guilt. Finally, if pretrial detention is imposed, the case must be tried within 60 days.

The "Get Tough" Movement in the 1980s

Despite the bail reform efforts in the 1960s, the 1980s "get tough" movement in crime control stalled its progress. This time was marked by increased crime rates and the public's diminishing confidence in the justice system's rehabilitative model. As a result, policymakers and criminal justice officials responded with a number of sentencing reforms and statutes in an effort to curb the crime problem. The bail system did not remain unaffected for long, however. Members of Congress grew concerned about the possibility that suspected felons out on bail were committing crimes. Additionally, they recognized that judges were setting high bail amounts and imposing excessive conditions of bail for individuals they believed were dangerous (Metzmeier 1996). In light of these events, Congress repealed the Bail Reform Act of 1966 through the passage of the Bail Reform Act of 1984 – a statute that allowed for the consideration of the accused's perceived dangerousness to the community and permitted preventive detention. Critics argued that a preventive detention provision, such as the one in the Act of 1984, "pits the fundamental common law principle of the presumption of innocence against the desire by police and judges to protect the public from purportedly dangerous defendants" (Metzmeier 1996, p. 399–400). In addition, some scholars believed that it violates the suspect's constitutional rights and undermines a system based on the presumption of innocence. Finally, others recognized the issues preventive detention poses because of the difficulty in predicting who will endanger the community (Kennedy 1980). Nevertheless, in *United States v. Salerno* (1987), the Supreme Court upheld the statute, claiming that "Congress formulated the detention provisions not as punishment for dangerous individuals, but as a potential solution to the pressing societal problem of crimes committed by persons on release." Therefore, they believed that this was a "legitimate regulatory goal."

Criminology's Increasing Role in the Study of Bail

Up until the 1980s, a majority of the literature on bail in the United States was generated and advanced by legal scholars. Legal scholars played a particularly active role in the discussions on bail reform due to the constitutional issues surrounding the bail system. In the 1980s, however, more criminologists began to conduct studies akin to the research carried out decades before in the early part of the twentieth century. Of interest were the factors related to bail release decisions, risk of flight, and the outcomes of pretrial detention (e.g., conviction). One such study found that due to the timing of bail hearings, which oftentimes occur shortly after arrest, judges are not equipped with sufficient information on which to base their decisions. Consequently, there are many inaccuracies in predicting which subjects will commit crimes during their pretrial release (Gottfredson and Gottfredson 1988). In the examination of the factors related to release, some researchers found that the offense charge and type of counsel affected the likelihood of a suspect's release (e.g., Fleming 1982; Holmes et al. 1987). Although the study of bail is an area that has remained relatively underresearched since that time, the role of the criminologist has increased in subsequent decades.

Bail in the Present Day and Future Trends

Goldkamp (1993) contended that the American bail system has experienced at least two generations of major bail reform in the United States. The first generation occurred in the 1960s, as substantial efforts were made to move the system toward bail processes that emphasized pretrial release through nonfinancial factors. This generation was followed by major bail reform efforts in the 1980s, which in many ways, propagated the principles of bail that reformers of the 1960s were attempting to eliminate. Nevertheless, both generations of reform were marked by conflicting ideas as to what should be the underlying purposes and functions of the bail system specifically, the presumption of innocence versus community safety – a conflict that has pervaded the American criminal justice system as a whole.

The conflicts and issues of the American bail system in the past continue to persist today. More specifically, dissension remains as to the factors weighted in bail decisions and the types of bail that are utilized. Regardless, the current trend in the bail system is one that has reverted back to a reliance on surety bonds provided by a bail bondsman and a system that still considers community safety as an important requisite for release. The former is reflected in national-level statistics of felony defendant pretrial releases in state court which indicate that approximately 33 % of the releases between 1990 and 2004 were obtained through surety bonds (Cohen and Reaves 2007) and that this trend has continued upward through 2006 (Cohen and Kyckelhahn 2010). Drawing from data collected by the Bureau of Justice Statistics (BJS), Clark (2010) suggested that the increased use of money bail is inversely related to the reduction of individuals on pretrial release. In fact, the number of pretrial detainees in jail has been on the rise since 1998, thereby further contributing to the jail crowding problem. The latter is manifested in statutes such as the 2001 Anti-Terrorism Act and the 2002 Patriot Act. Both statutes have extended the crimes for which pretrial detention is applicable in the federal system. In addition, most states have amended their laws to reflect the consideration of public safety in the decision to grant bail.

The present state of the bail system in the United States, therefore, has sparked the attention of politicians, scholars, lobbyists, and the public once again, with staunch proponents on both sides of the bail issue, including efforts by some to privatize the system. Lobbying groups, such as the American Legislative Exchange Council, are increasing their efforts to further privatize the criminal justice system through a post-conviction release strategy that mimics the commercial bail in pretrial release. Other groups, such as the American Bar Association and the Pretrial Justice Institute, however, continue to urge that the bail system should impose the least restrictive conditions and rely on personal recognizance and other nonfinancial forms of bail (e.g., conditional releases and unsecured bonds). Interestingly, there remains a lack of empirical research as to which approach works best in terms of appearance rates and cost-effectiveness. Even still, it has been suggested that a third generation of the bail system is underway (Schnacke et al. 2011), concerned primarily with the use of empirically based screening tools in making pretrial release decisions and evidence-based decision making in release decisions through the use of pretrial service agencies. Assessments of some pretrial screening tools have already produced positive results (e.g., Lowenkamp et al. 2008). However, such initiatives remain arguably highly nuanced compared to the more common and generalized commercial bond practice. In fact, one rather robust empirical study found that defendants from 75 of the largest counties in the USA who were release via a surety bond (bail bondsman) were considerably less likely to fail to appear in court (28 % less likely) compared to similar defendants released on personal recognizance (Helland and Tabarrok 2004). Further, absconders released via a bondsman were 53 % to be at large for long periods of time compared to absconders released on personal recognizance. Moreover, national efforts reminiscent of those in the 1960s, including the National Symposium on Pretrial Justice, have taken place as recently as 2011 with the intent to further advance America's system of bail. Whatever the course, the objective remains clear – finding the balance between community safety and minding the long-standing constitutional presumption of innocence. In the end, there is a paucity of quality research on "what works" in regard to pretrial release practices. Few quantitative studies are generalizable to the criminal justice system as a whole, and scholars have only recently begun to such important questions with solid criminal justice data and advanced empirical research designs. Qualitative inquiries into the issue are altogether absent from the scholarly literature surrounding bail release in the United States.

Related Entries

► Alternatives to Pre-trial Detention

Recommended Reading and References

A study of the administration of bail in New York City (1958) Univ Pa Law Rev 106:693–730

Bail: An ancient practice reexamined (1961) Yale Law J 70:966–977

Beeley A (1927) The bail system in Chicago. University of Chicago Press, Chicago

Bottomley AK (1968) The granting of bail: principles and practice. Mod Law Rev 31:40–54

Clark J (2010) The impact of money bail on jail bed usage. American Jails 47–54

Cohen TH, Kyckelhahn T (2010) Felony defendants in large urban counties, 2006. US Department of Justice, Washington, DC

Cohen TH, Reaves BA (2007) Pretrial release of felony defendants in state courts. US Department of Justice, Washington, DC

DeHaas E (1940) Antiquities of bail: origins and historical development in the criminal cases to the year 1275. Columbia University Press, New York

Devine FE (1991) Commercial bail bonding: a comparison of common law alternatives. Praeger Publishers, New York

Duker WF (1977) The right to bail: a historical inquiry. Albany Law Rev 42:33–120

Fleming R (1982) Punishment before trial: an organizational perspective of felony bail processes. Longman, New York

Foote C (1954) Compelling appearance in court administration of bail in Philadelphia. Univ Pa Law Rev 102:1031–1079

Foote C (1965) The coming constitutional crisis in bail: I. Univ Pa Law Rev 113:959–999

Friedman LS (1976) The evolution of a bail reform. Policy Sci 7:281–313

Goldkamp JS (1985) Danger and detention: a second generation of bail reform. J Crim Law Criminol 76:1–74

Goldkamp JS (1993) Judicial responsibility for pretrial release decision making and the information role of pretrial services. Federal Probation 57:28–35

Goldkamp JS, Gottfredson (1979) Bail decision making and pretrial detention. Law and Human Behavior 3:227–249

Gottfredson MR, Gottfredson DM (1988) Decision making in criminal justice: toward the rational exercise of discretion, 2nd edn. Plenum Press, New York

Hazeltine HD (1910) The formal contract of early English law. Columbia Law Rev 10:608–617

Helland E, Tabarrok A (2004) The fugitive: evidence on public versus private law enforcement from bail jumping. J Law Econ 47:93–122

Holmes MD, Daudistel HC, Farrell RA (1987) Determinists of charge reductions and final dispositions in cases of burglary and robbery. Journal of Research in Crime and Delinquency 24:233–254

Indemnification contracts in the law of bail (1949) V Law Rev 35:496–504

Kennedy EM (1980) A new approach to bail release: the proposed federal criminal code and bail reform. Fordham Law Rev 48:423–436

Lowenkamp CT, Lemke R, Latessa E (2008) The development and validation of a pretrial screening tool. Federal Probation 72:2–9

Metzmeier KX (1996) Preventive detention: a comparison of bail refusal practices in the United States, England, Canada and other common law nations. Pace Int Law Rev 8:399–438

Mitchell JN (1969) Bail reform and the constitutionality of pretrial detention. V Law Rev 55:1223–1242

Ozanne MR, Wilson RA, Gedney DL (1980) Toward a theory of bail risk. Criminology 18:147–161

Paulsen MG (1966) Pre-trial release in the United States. Columbia Law Rev 66:109–125

Pimsler M (1977) Solidarity in the medieval village? The evidence of personal pledging at Elton, Huntingdonshire. J British Stud 17:1–11

Pollack F, Maitland FW (1898) The history of English law: before the time of Edward I, 2nd edn. Cambridge University Press, Cambridge, UK

Pound R, Frankfurter F (1922) Criminal justice in Cleveland. The Cleveland Foundation, Cleveland

Ryan JV (1967) The last days of bail. J Crim Law Criminol Police Sci 58:542–550

Schnacke TR, Brooker CMB, Jones MR (2011) The third generation of bail reform. Denver University Law Review Online. Retrieved 10 Feb 2011, from http://www.denverlawreview.org/practitioners-pieces/2011/3/14/the-third-generation-of-bail-reform.html

History of Boot Camps

Ojmarrh Mitchell
Department of Criminology, University of South Florida, Tampa, FL, USA

Synonyms

Boot camps; Shock incarceration; Intermediate sanctions

Overview

Correctional boot camps are programs modeled after military basic training. Just like basic training, boot camps emphasize drill and ceremony, and physical activity. Generally, boot camps target young, nonviolent offenders with limited criminal history. Boot camps are almost always short-term programs lasting 90–180 days. Inmates who successfully complete these programs are released under supervision back to the community; however, inmates who drop out or are dismissed from the boot camps often are required to serve longer terms of incarceration in traditional correctional facilities.

Correctional boot camps in the United States emerged, proliferated, and receded with remarkable speed. In the early 1980s, boot camps emerged in two states. Soon thereafter, boot camps became a national phenomenon with at least one boot camp operating in the majority of states by the mid-1990s. Near their peak popularity, the boot camp bubble burst and their popularity plummeted in the new millennium. Thus, in a span of a little more than 25 years, boot camps went from programs in a few states to being widespread back to obscurity. All the while, boot camps remained controversial.

This entry traces the history of the rise, fall, and controversies of boot camps in the United States. In short, while many factors fueled the rise of boot camps, undeniably boot camps' growth was driven by growing correctional populations and boot camps' fit with prevailing political sensibilities. Conversely, their demise was driven by a combination of research inconsistent with the goals of boot camp and prominent cases of abuse. These factors served to erode boot camps' political support. In all of this, there is a larger cautionary tale that warns against latching on to the latest criminal justice fad without some reasonably rigorous empirical support for that intervention.

The Boot Camp Model

The term "correctional boot camp" is a generic term referring to various correctional programs that utilize a militaristic style and atmosphere. Boot camps programs are sometimes called "shock incarceration" programs, "leadership camps," "accountability programs," and so forth. While program names may vary, the military style that is at the core of each makes them a boot camp.

The environment of boot camps resembles military basic training. Inmates typically live in barrack-style housing, are dressed in military uniforms, and are supervised by correctional officers with military titles (e.g., drill sergeant, corporal, captain). Inmates ("cadets") refer to correctional staff as “sir” or “ma’am.” Cadets enter the boot camp in groups called “platoons.” Each platoon lives and works together as a group throughout the period of confinement.

The daily routine at a boot camp is filled with planned activities and idle time is rare. Cadets are required to wake early, typically 5 am. They dress quickly and then march to an exercise area. There, they engage in strenuous physical exercise (“physical therapy”) and perform military drill and ceremony. After an hour or two of physical therapy and drills, boot camp cadets march to a dining area for breakfast. Breakfast and all meals are orderly ceremonies. Cadets stand with their meals in hand until commanded to sit. Meals are eaten quickly and with minimal conversation. Generally, when breakfast is complete, cadets either attend school or leave the camp to engage in community service such as road cleanup. Afterwards, cadets are required to complete more physical therapy and military drills.

Cadets progress through three or more stages. Program activities vary by stage and there is usually decreasing emphasis on physical therapy and increasing emphasis on performing manual labor, work in the community (e.g., road cleanup), and treatment activities. The total length of stay in a boot camp is 90–180 days but can be extended by misbehavior that causes an inmate to repeat a stage.

At all times, boot camps require inmates to adhere to a strict code of conduct. Deviations from these rules are met with verbal reprimands, punishments involving physical exercises (e.g., push-ups), or the removal of privileges. Serious cases of misbehavior can result in an inmate being required to repeat a stage, and continued serious misbehavior can lead to expulsion from the program. Those expelled are usually placed in other correctional facilities and required to serve longer periods of incarceration. Conversely, those who complete boot camp programming are honored at a formal graduation ceremony which family members are encouraged to attend.

Those eligible for boot camp programs are almost always young and most often nonviolent offenders. Because of the physically demanding nature of boot camps, inmates are required to be youthful; thus, boot camp inmates are juveniles or young adults (typically, less than 30 years old, and very rarely more than 35 years old). Further, those eligible for boot camp participation are required to complete health screenings to ensure that they are capable of engaging in taxing physical activities.

Outside of this military atmosphere, boot camps vary widely. Some programs have little to no time allotted for treatment activities, while others devote considerable portions of the day to these activities. Some programs require offenders to volunteer for the programs; others allow judges and/or corrections officials to mandate boot camp participation. Another important variation is in the manner and intensity of post-release community supervision; some programs offer offenders limited community supervision, while other boot camp programs offer intensive supervision.

It is also important to note that boot camp programs appeared to evolve over time towards a rehabilitative focus. Early boot camps heavily emphasized military training. Later boot camps still kept the military model as a core component; however, these later programs included high levels of rehabilitative programming, especially drug treatment (Gransky et al. 1995).

Above and beyond these programmatic differences, at its core, the boot camp model is premised on the notion that an austere correctional environment combined with military style training will reduce crime through two mechanisms. First, the boot camps’ harsh environments deter crime by showing young, nonviolent offenders that crime is met with strict punishment. Second, boot camps instill discipline in program participants, which presumably is lacking, and this self-discipline allows participants to resist criminal temptations in the future.

The Rise of Boot Camps

Modern correctional boot camps emerged in the early 1980s. Boot camps, however, have clear predecessors. Perhaps, the earliest forerunner is New York’s Elmira Reformatory, which implemented military training in 1888 under

Zebulon Brockway's leadership. This militaristic approach was adopted in an effort to instill discipline and to keep inmates active – the same goals as modern boot camps. The latter goal was particularly important at the time, as recent legislation restricted the use of inmate labor for commercial purposes, which had been the norm. This military atmosphere remained in place at Elmira until 1920 (Anderson et al. 1999). Thereafter, the "rehabilitative ideal" reached its peak popularity and the military training used at Elmira was adopted only sporadically at other correctional institutions until the 1980s.

Military training was most systematically and thoroughly implemented in the modern boot camps that emerged in the early 1980s. Oklahoma opened the first correctional boot camp in 1983. Later in the same year, Georgia opened a boot camp at the Dodge Correctional Institution. Thereafter, boot camps spread rapidly across the United States. By 1993, just 10 years after their emergence, 59 camps were in operation in 29 states. In 1995, this number expanded to 75 state-operated boot camps for adults and 30 more for juveniles were in operation, as well as another 18 adult boot camps operating in county jails. Boot camps even spread to the federal correctional system, where three correctional boot camps were in operation. These numbers surely underestimate the actual number of boot camps in operation, as many boot camps, especially those for juveniles, were privately run. Thus, at one point before the turn of the new millennium, the majority of states, the federal system, and some counties had at least one boot camp in operation.

This explosive growth in boot camp programs was driven by many factors but three factors appear most important. The first of these factors is the dramatic growth of correctional populations. Correctional populations increased rapidly in the 1980s and 1990s, which caused overcrowding in many prisons and jails. Some of the offenders caught in this tidal wave of offenders were marginal who were deemed too serious for standard probation but too minor for prison. Boot camps, along with other "intermediate sanctions" (i.e., correctional sanctions in between standard probation and prison), provided courts with options suitable for these marginally serious offenders.

A second factor driving the popularity of boot camps was their purported ability to simultaneously reduce prison populations and costs, while reducing recidivism. As discussed above, boot camps involved short-term confinement. Proponents argued that these short periods of confinement were less costly than prison sentences, and consequently, boot camps would reduce prison costs and prison populations. Further, the harsh environment of boot camps were touted as effective in reducing recidivism via specific deterrence and perhaps more broadly via general deterrence. Thus, boot camps were advertised to be lower cost and more effective than traditional prisons.

The third, and perhaps most important, factor was that boot camps were a perfect fit to the "get tough on crime" political sensibilities of the time. In this era, increasingly policy-makers rejected correctional interventions based on "soft on crime" rehabilitative approaches and instead asserted the effectiveness and appropriateness of punitive sanctions based on just deserts, deterrence, and incapacitation. Policy-makers recognized boot camps' fit to this new correctional paradigm and responded by allocating millions of dollars for boot camps. Specifically, Congress in the Violent Crime Control and Law Enforcement Act of 1994 allocated $24.5 million for boot camps, and this money was used to fund 44 grants for boot camp planning and construction in 1995. This allocation was relatively small in monetary terms, but it served as an important, symbolic endorsement of the boot camp model.

Controversies

While boot camps became popular in the mid- to late 1990s, they were not without controversy. Proponents touted the utility of boot camps as intermediate sanctions, as well as their effectiveness in reducing recidivism, prison populations, and prison costs. These claims proved controversial as empirical research addressing these issues was in its infancy prior to the period of boot camp growth.

Most fundamentally, critics contended that because boot camps did not address the underlying individual and contextual risk factors for criminal behavior, such programs would be ineffective. According to these critics, the boot camp model fails to address the known correlates and causes of offending and instead focuses on factors such as self-discipline and physical fitness, which are not strong correlates of offending. The gap between the attributes targeted by boot camps and the attributes known to predict criminal behavior is so vast that Latessa et al. (2002) offer boot camps as an example of "correctional quackery."

Further, given that history demonstrated most intermediate sanctions draw offenders from probation – not prison – boot camps were unlikely to reduce prison populations and costs. Other critics went a step further and argued that boot camps may actually be harmful. For example, Mathlas and Mathews (1991) argue that boot camps not only widen the net (i.e., lead to more offenders being incarcerated) but also the offenders drawn into boot camps are more likely to be ill-suited for these environments and, as a consequence, they are more likely to be harmed by the experience.

Still other critics contended that boot camps confrontational and unsupportive environments would cause damage to young offenders, especially offenders with histories of abuse. These critics predicted that youths with histories of abuse would have difficulty adjusting to the boot camp environment.

Likewise, Morash and Rucker (1990) argue that boot camps are abusive and dangerous environments. These authors provide one of the earliest, most critical, and, in some regards, most prescient critique of the boot camp model. They argue that boot camps are steeped in images of masculinity and power. These "ultramasculine" environments encourage aggression by staff and inmates, and thus boot camps are dangerous.

In short, few criminal justice-based interventions have been simultaneously as popular and as controversial as boot camps. These controversies stimulated research on boot camps and the popularity of these programs facilitated this research.

Analyzing the Controversy

Central to the disagreements between boot camp proponents and critics is the assumption that boot camps and traditional correctional facilities are vastly different. Thus, a fundamental issue concerns the similarity between these two kinds of correctional facilities. Research studying the environments of these facilities finds that boot camps and traditional facilities do in fact differ in many regards. The most prominent research comparing the environments of boot camps and traditional facilities is the work of Doris MacKenzie and colleagues (MacKenzie et al. 2001a). These authors surveyed juveniles and correctional staff at a large number of boot camps and traditional correctional facilities. These survey data revealed that both staff and juveniles indicated that boot camps are more structured/regimented and boot camps keep juveniles more active than traditional facilities. Interestingly, both staff and juveniles indicated that boot camp environments exhibited greater safety in regard to inmate-to-inmate aggressive behaviors but boot camps exhibited somewhat greater danger to juveniles from correctional staff. This latter finding is explained by the fact that boot camp correctional staff typically used more confrontational disciplinary practices such as verbal reprimands and physical therapy as punishment. These confrontational and sometimes aggressive disciplinary tactics, according to Lutze and Brody (1999), raise concerns about Eighth Amendment violations and make boot camps susceptible to civil, and perhaps criminal, litigation. The above findings are largely consistent with popular perceptions of boot camps and were largely confirmed in later research conducted in adult facilities.

The research, however, does not support certain popular perceptions about boot camps. Most notably, boot camps exhibit more rehabilitative programming than is typically assumed. In fact, research finds that boot camps have similar or greater amounts of rehabilitative programming as traditional correctional facilities. This finding contradicts the critique that boot camps will be ineffective because they fail to address the underlying risk factors for criminal behavior, as boot camps contain the same kinds and amounts of rehabilitative program as non-boot camp facilities.

Another important, but much less well-known, finding is that when asked about their perceptions of overall correctional experience, boot camp inmates reported more positive perceptions about their experiences than inmates in traditional facilities (Lutze 2001; MacKenzie et al. 2004). In other words, inmates confined to boot camps found their experiences less onerous than inmates confined in traditional facilities. Why this is so is unclear, but some portion of this finding appears to be related to the high activity levels found in boot camps. One of the negative aspects of correctional confinement is the sheer boredom of having lots of idle time. Boot camps solve this problem for inmates, and, therefore, instead of perceiving boot camps negatively, inmates actually preferred their experiences to those confined in traditional facilities! This finding undermines the notion that boot camp inmates will be deterred by the harshness of these facilities' correctional environments; in that, one would expect boot camp inmates to have more negative perceptions of their correctional experience in comparison to inmates confined in other institutions, if the boot camp environment were truly aversive.

Considerable research also has been given to describing and assessing offender adjustment to the environments of boot camps. These studies generally find that boot camp inmates adjust to the correctional environment as well or better than inmates in traditional facilities (Lutze 2001; MacKenzie and Shaw 1990). However, juveniles confined in boot camps with histories of abuse adjusted less favorably to these environments than similar kids in traditional facilities (MacKenzie et al. 2001b). Somewhat similarly, other research finds that inmates who perceive the boot camp environment as more masculine adjusted more poorly than other inmates. And almost universally, research finds that a substantial minority of offenders cannot either conform to boot camps' strict rules or handle the physical rigors, and as a result, they either drop out or are dismissed. These offenders tend to perceive the boot camp environment more negatively, have greater indicators of criminal tendency, and have lower self-efficacy than other participants. Thus, once again, the research fails to fully support the caricature of boot camps in that inmates typically adjust to their environments quickly – contrary to a common criticism of boot camps. However, there is a nontrivial minority of boot camp inmates for whom adjustment is difficult, if not impossible.

Without a doubt, the most controversial issue surrounding boot camps is whether or not they reduce recidivism. Boot camp proponents assert that boot camp inmates will have lower rates of recidivism than other inmates, whereas some boot camp critics contend that boot camp inmates will have higher rates of recidivism. Taken as a whole, the research supports neither of these claims. The research comparing recidivism rates of inmates confined at boot camps and other correctional institutions can be summarized accurately and succinctly with just two words: no difference. While there are a small number of evaluations that find statistically significant differences between these two groups' recidivism rates, the great majority of such evaluations find highly similar recidivism rates. In fact, a large and sophisticated meta-analytic review of boot camp evaluations found no statistically significant difference in the average recidivism rates from these evaluations (Wilson et al. 2005). This finding held regardless of the evaluations' methodological rigor and sample features (e.g., adult or juvenile sample). Interestingly, they found that boot camps that had greater emphasis on treatment or had an aftercare component did significantly better than their comparison groups. This conclusion was further supported by a recent evaluation by Kurlychek and Kempinen (2008) that found boot camp participants who received a mandatory aftercare component have significantly lower recidivism rates relative to the control group.

Despite the fact that boot camps in many jurisdictions were adopted at least in part to ease prison overcrowding and reduce costs, relatively few studies have addressed these issues. Most studies focuses on these issues conclude that boot camps have minimal effects on correctional populations and costs. This finding is attributable to the fact that most boot camp inmates appear to be drawn from probation, not prison populations. Further, a substantial proportion of boot camp inmates are expelled for noncompliance, and as a result, they are sent into the very incarcerative settings from which they were supposed to be diverted. As an example, the work of Parent is instructive (Parent et al. 1999). These authors analyzed prison populations from three jurisdictions. They found that even if boot camp participants are drawn from prison-bound offenders, boot camp graduates are given large reductions in sentence length, and few boot camp participants are kicked out of the program, cost savings are likely small. Thus, even in the ideal set of circumstances, boot camps save few prison resources.

Taken as a whole, the empirical research fails to fully support boot camp proponents or critics. Contrary to claims of proponents, boot camps do not lower correctional costs or prison populations, and boot camp inmates do not perceive their experiences as particularly aversive in comparison to inmates at other institutions. Contrary to the claims of boot camp critics, most inmates quickly adjust to boot camps, boot camps have as much or more treatment programming than traditional institutions, and boot camps do not experience greater inmate-to-inmate aggression than traditional institutions. Perhaps, if the controversies discussed above were the only factors affecting boot camps' popularity, these programs' popularity may have tapered off slowly or perhaps remained stable. Instead, however, the popularity of boot camps was dramatically altered by growing evidence and dramatic cases of staff-to-inmate abuse.

The Fall of Boot Camps

Between 1995 and 2000, the number of state-operated boot camps dropped by more than 25 % and the number of inmates confined in boot camps dropped 30 %. Thereafter, national figures do not exist, but it is clear that the number of boot camps continued to decline sharply in the new millennium.

Boot camps' popularity diminished rapidly in the face of three factors: (1) lack of empirical support for boot camps' effects on recidivism and prison costs, (2) numerous and prominent cases of inmate abuse, and (3) the erosion of their political support due to the above factors. The lack of empirical support for boot camps' claims of effectiveness in reducing recidivism and costs undermined their reason to exist. However, given policy-makers historical reluctance to modify criminal justice policy and practice in response to empirical research, it seems unlikely that boot camps would have fallen out of favor so rapidly if this were the only issue.

A more important factor leading to the abandonment of boot camps was the growing evidence of inmate abuse at the hands of boot camps' correctional staff. As a collective, boot camps were plagued by complaints of abuse. These complaints received media attention beginning in the late 1990s and continuing into the new millennium when several deaths of juveniles confined in boot camps occurred. In 1998, Nicholaus Contreraz died at a privately run boot camp in Arizona from an untreated infection in his lungs. Despite Contreraz's reports of feeling ill, correctional staff did not seek medical attention for him. Instead, correctional staff, believing that Contreraz was faking his illness, reportedly ridiculed him and made perform exercise as punishment. The infection in his lungs worsened and eventually led to his death. An investigation of this death revealed approximately 100 previous complaints about the facility. As a result of this investigation, the facility and several similar facilities in Arizona lost their licenses to operate and were closed in 1998.

Similar cases of abuse occurred in other boot camp facilities in the following years including the deaths of Gina Score, who died after a forced march in a juvenile boot camp located in South Dakota; Anthony Hayes, who died of dehydration in an Arizona juvenile boot camp; and Roberto Reyes, who died from an untreated case of rhabdomyolysis, probably caused by a spider bite.

In all of these cases, despite complaining of feeling ill, these juveniles were forced to perform physical exercise until they died. Further, the autopsies in many of these cases revealed widespread bruising – suggesting that these juveniles were beaten prior to death.

The videotaped case of Martin Lee Anderson's death at a juvenile boot camp in Florida is the most well-known and infamous case. Within hours of arriving, Anderson was forced to exercise including a run. When he complained of fatigue, he was physically forced to continue to run and denied medical treatment. Eventually, Anderson collapsed. He died one day after being admitted to the juvenile boot camp. The force used against Anderson was captured on video, which eventually became public and created a public outcry. Just as in the cases discussed above, there had been numerous complaints of abuse made against the boot camp. As a result of the public outcry concerning Anderson's death and the ensuing investigation, the Florida Legislature closed all of the state's juvenile boot camps.

The closure of Florida's juvenile boot camps is a sure sign of the third factor that leads to the demise of boot camps: eroding political support. Florida had once been the nation's leader in juvenile boot camps. Currently, Florida has none. Thus, even in a famously "tough on crime" state like Florida, boot camps are no longer a correctional option for juveniles (the state still operates at least one boot camp for youthful adults). This is just one prominent example of the eroding political support for boot camps. Another sign is the fact that no federal funds dedicated to the planning and implementation of boot camp have been made available after the mid-1990s. In fact, in 2007, Congress held a hearing concerning abuse at boot camps and other juvenile residential treatment facilities. This hearing and an accompanying GAO report detailed many cases of abuse at boot camps, and the witnesses featured in this hearing uniformly testified against the boot camp model. The overall message was clear: federal policy-makers no longer endorse boot camps and the boot camp model is obsolete.

Boot Camps: A Cautionary Tale in Criminal Justice Interventions

Boot camps burst onto the scene touted as effective interventions for offenders in the early 1980s. Their popularity was fueled by their fit to the political sensibilities of the era, and the Violent Crime Control Act of 1994 endorsed boot camps as worthy crime control programs by providing federal funds for these programs. All of this occurred without any reasonably sound empirical support for their effectiveness and safety. In short, politics trumped science.

The moral to the history of boot camps is obvious: proceed with caution. Policy-makers should not have invested millions of dollars in boot camps. Federal policy-makers should not have spurred state and local jurisdictions to develop boot camps. At the time, the effectiveness and safety boot camps were not established. Instead of proceeding with caution, policy-makers latched onto the latest criminal justice fad. A more reasonable approach would have been to fund and carefully evaluate a handful of boot camp demonstration projects before endorsing them by making funding available for such programs.

Given the history of corrections, the cautionary tale of the rise and fall of boot camps is likely to be forgotten. Likewise, the common sense advice of proceeding with caution in promoting and adopting correctional interventions is likely to go unheeded. As a consequence, the next criminal justice fad of the day is only a short series of serendipities away.

Recommended Reading and References

Anderson JF, Dyson L, Burns JC (1999) Boot camps: an intermediate sanction. University Press of America, Lanham

Austin J (2000) Multisite evaluation of boot camp programs: final report. George Washington University, Institute on Crime, Justice, and Corrections, Washington, DC

Gover AR, MacKenzie DL, Styve GJ (1999) Boot camps and traditional correctional facilities for juveniles: a comparison of the participants, activities, and environments. J Crim Justice 28:53–68

Gransky LA, Castellano TC, Cowles EL (1995) Is there a “second generation” of shock incarceration facilities? The evolving nature of goals, program elements, and drug treatment services in boot camp programs. In: Smykla JO, Selke WL (eds) Intermediate sanctions sentencing in the 1990s. Anderson, Cincinnati, pp 89–111

Kurlychek M, Kempinen C (2008) Beyond boot camp: the impact of aftercare on offender reentry. Crim Public Policy 5:363–388

Latessa EJ, Cullen FT, Gendreau P (2002) Beyond correctional quackery: professionalism and the possibility of effective treatment. Fed Probation 66(2):43–49

Lutze F (1998) Are shock incarceration programs more rehabilitative than traditional prison? A survey of inmates. Justice Q 15:547–563

Lutze FE (2001) The influence of a shock incarceration program on inmate adjustment and attitudinal change. J Crim Justice 29:255–267

Lutze FE, Brody DC (1999) Mental abuse as cruel and unusual punishment: do boot camp prisons violate the Eighth Amendment? Crime Delinquency 45:242–255

MacKenzie DL, Armstrong GS (2004) Correctional boot camps: military basic training or a model for corrections? Sage Publications, Thousand Oaks

MacKenzie DL, Hebert EE (1996) Correctional boot camps: a tough intermediate sanction. United States Department of Justice, Washington, DC

MacKenzie DL, Parent DG (1992) Boot camp prisons for young offenders. In: Byrne JM, Lurigio AJ (eds) Smart sentencing: the emergence of intermediate sanctions. Sage Publications, Newbury Park, pp 103–122

MacKenzie DL, Piquero A (1994) The impact of shock incarceration programs on prison crowding. Crime Delinquency 40:222–249

MacKenzie DL, Shaw JW (1990) Inmate adjustment and change during shock incarceration: the impact of correctional boot camp programs. Justice Q 7:125–150

MacKenzie DL, Brame R, McDowall D, Souryal C (1995) Boot camp prisons and recidivism in eight states. Criminology 33:401–430

MacKenzie DL, Gover AR, Armstrong GS, Mitchell O (2001a) A national study comparing the environments of boot camps with traditional facilities for juvenile offenders. NIJ research in brief (NCJ# 187680). National Institute for Justice, Washington, DC

MacKenzie DL, Wilson DB, Armstrong GS, Gover AR (2001b) The impact of boot camps and traditional institutions on juvenile residents: perceptions, adjustment, and change. J Res Crime Delinquency 38:279–313

MacKenzie DL, Mitchell, O, Bierie D, Van Brakle M, O’Neill L, Franke D, Mitchell F (2004) A randomized study of the Maryland correctional boot camp for adults: final report. Unpublished manuscript, College Park

MacKenzie DL, Bierie D, Mitchell O (2007) An experimental study of a therapeutic boot camp: impact on impulses, attitudes and recidivism. J Exp Criminol 3:221–246

Mathlas RES, Mathews JW (1991) Boot camp programs for offenders: does the shoe fit? Int J Offender Ther Comp Criminol 35:322–327

Morash M, Rucker L (1990) A critical look at the idea of a boot camp as a correctional reform. Crime Delinquency 36:204–222

Parent DG, Snyder RB, Blaisdell B (1999) Boot camps’ impact on confinement bed space requirements: Final report to National Institute of Justice. Cambridge, MA: Abt Associates, Inc.

Simon J (1995) They died with their boot on: the boot camp and the limits of modern penality. Soc Justice 22:25–48

Stinchcomb JB (1999) Recovering from the shocking reality of shock incarceration: what correctional administrators can learn from boot camp failures. Correct Manag Q 3:43–52

Styve GJ, MacKenzie DL, Gover AR, Mitchell O (2000) Perceived conditions of confinement: a national evaluation of juvenile boot camps and traditional facilities. Law Hum Behav 24:297–308

Wilson DB, Mackenzie DL, Mitchell FN (2005) Effects of correctional boot camps on offending: a Campbell collaboration systematic review, Campbell Collaboration. http://www.campbellcollaboration.org

History of Capital Punishment

► History of the Death Penalty

History of Corporal Punishment

G. Geltner
University of Amsterdam, Amsterdam,
The Netherlands

Synonyms

Flogging; Physical punishment

Overview

In the gallery of penal practices, corporal punishment, or the dispensing of bodily harm in response to or as a deterring measure against crime, occupies a stable position as a marker of cruelty, especially when condoned by a central authority such as a state. From Cesare Beccaria to Émile Durkheim to Max Weber, and especially under the more recent and diverse influences of social philosophers Norbert Elias and Michel Foucault, modern students of punishment have construed cultures that allow physical pain to be legitimately, let alone publicly, inflicted as out of step with the process of civilization and as retaining a relic of an unenlightened past. Corporal punishment, however, has a far more complex history than a long and steady fall from grace, an inverse trajectory as it were to the progress of humanity (Scott 1938; Yelyr 1941). For, fluctuations in frequency aside, the past uses of corporal punishment were never devoid of reason, at least in the sense that it was mostly meted out proportionately, gradually, and with a view to achieving social goals far beyond individual suffering, such as shaming and paving the way to an offender's reintegration. And, on the other, it is still used today: openly in certain milieus and surreptitiously in others, with some indications that it is nowhere near to being abolished. Indeed, some scholars have recently argued for an expansion of corporal punishment as a solution to the crisis of modern penology (Newman 1983; Moskos 2011).

Fundamentals

For the purposes of this discussion, corporal punishment will encompass the licit infliction of intentionally nonlethal bodily pain, temporary or lasting, including physical impairment or aesthetic alteration meant to reduce capacity and create social disability. So defined, and observed from a broad geographical and historical perspective, corporal punishment appears to be a common measure that, moreover, rarely serves as an exclusive penalty typifying any particular civilization, be it modern or premodern; usually, it was one option among several practiced in a given context, and different cultures made recourse to it in different judicial environments and with changing frequency and goals.

Antiquity

Corporal punishment, like most current penal measures, can be traced back to the earliest civilizations whose records have come down to us, such as the early law "code" of Ur-Namma, ruler of the Sumerian city of Ur (2112–2095 BCE) and founder of its third dynasty. This casuistic text addresses a host of cases, running the gamut from labor to civic to criminal law, and in doing so not only attests the degree of judicial sophistication attendant upon a complex society but also denies the ubiquity of ancient corporal punishment specifically and of physical *lex talionis* ("eye for an eye") in general. The vast majority of the text's 85 surviving promulgations prescribe monetary penalties or rewards, and a mere handful order the death penalty. On three occasions alone is the infliction of nonlethal physical harm considered meet: "If a man fractures another man's skull in a fistfight, they will flog him 180 times" (§22); "If a freedman beats a slave, they will whip him sixty times with a strap and sixty times with a belt" (§26); and "If a slave girl insults someone who is acting as her mistress they will rub her mouth with one *sila* of salt" (§30). To be sure, legal texts do not always convey penal realities, but at least at the normative level it appears that Ur-Namma's policies more often than not aimed at sparing the rod. Even in cases amenable to the employment of the *lex talionis*, monetary fines are often preferred. Thus, for instance, "[i]f a man breaks another man's nose with his fists, he will pay forty shekels of silver" (§19), and "[i]f a man causes the loss of another man's eye, he will pay thirty shekels" (§23).

Contrary to the common view that sees punishment as becoming physically milder over time, the subsequent and more famous and elaborate Laws of Hammurabi (*c.* 1780 BCE) illustrate a "reverse" trend of increased brutalization. Whereas physical and even capital punishments in Ur were scarcely decreed, in Babylonian law they become all but ubiquitous. Yet, rather than expressing a knee-jerk response to any and all violations, Babylonian prescripts for corporal punishment were undergirded by several distinct rationales, which would remain influential for centuries to come. The first was to target the "offending limb," an approach already underlying the fate of the misbehaved slave girl mentioned in the previous paragraph. Thus, if the son of a paramour or a prostitute disowns an adoptive parent, he would have his tongue cut off (§192); a nurse concealing the death of an infant by obtaining and suckling another child will have her breasts cut off (§194); a reckless surgeon would lose his hands (§218); and so forth (see also §§193, 195, 226, 253). A further model prescribed retaliations according to the injury inflicted ("classic" *lex talionis*), as when a man putting another man's eye out or breaking his bone or teeth would suffer the same fate (§§196, 197, 200). Last, offenders were to be maimed in body parts less immediately tied to their crime, but in ways that nonetheless branded them as offenders of a particular sort: a false accuser was to be marked on his brow (§127), and a slave disowning his master would have his ear cut off (§282), as would a slave striking the body of a freedman (§205). Beating, by comparison to these three penal modes, was uncommon in Babylonian law. Only on one occasion is it prescribed, namely, when a man strikes someone of a higher social rank, in which case he is to be publicly subjected to sixty blows "with an ox-whip" (§202).

In Ancient Egypt, by contrast, beating appears to have been less common as a form of summary justice already during the Early Kingdom (2686–2181 BCE), while a papyrus from the reign of the Pharaoh Thutmosis IV (late fourteenth century BCE) records the case of a soldier by the name of Mery who was subjected to 100 blows for false litigation. Here as elsewhere in the ancient Near East, time brought more rather than less frequent recourse to corporal punishment (VerSteeg 2002). A royal decree issued under Seti I (late thirteenth century BCE) orders the infliction of 200 blows, five open wounds, and a monetary fine against civil administrators who illicitly requisition free personnel or slaves, or detain a boat or its crew belonging to a foundation for Osiris of Abydos. Anyone encroaching upon the boundaries of the same foundation or stealing an animal from it was to be punished by having his or her nose and ears cut off and then being put to work as a cultivator (Lorton 1977). Brutalization there may well have been, yet here as well as in other contemporaneous documents, there is an explicit proportionality between the offense, the number of blows, and the monetary fine or status change to be imposed upon the offender. Maiming, to take another example, is clearly meant to go beyond the infliction of pain: it is used to deter and shame by associating mutilation with and thus defining what is a major transgression. Practically, moreover, such disfigurement made escape more complicated.

These and other strategies were associated with physical punishment in early Vedic India (1700–1100 BCE), although here disabling the culprit appears to have been paramount. The range of punishments is also somewhat broader as well as more clearly gradated. Measures included whipping, beating with a broken bamboo cane, driving iron nails into the body, dripping hot oil into orifices, starvation (also an extremely painful form of capital punishment), branding, and various forms of maiming. More consistently than elsewhere in the ancient world, dismemberment in India was linked functionally and symbolically to the offending limb. For instance, someone who urinated in a forbidden place would have his penis cut off, a blasphemer would have his lips cut out, a thief his hand, and so forth. Most of these penalties, however, were never used summarily or gratuitously, but rather had to follow proper admonition, reprove, and fines. At least in principle, only when these failed to deter or prevent recidivism could a judge resort to corporal punishment. And even then certain and telling limitations were put in place, which served to reinforce social hierarchies and boundaries. For example, the more flexible (and thus potentially more lenient) punishment of whipping was meant by default for women, infants, the mentally ill, the poor, and the sick, and hot oil and iron could only be used against (and in turn defined) major offenses, such as blasphemy by a Sudra against a Brahmin or a king. A distinct feature of ancient Indian branding, finally, was that it was usually carried out on one's forehead and indicated the type of crime committed: female genitalia signified incest, a dog's foot the theft of gold, and so on. A branded person was often also banished from his or her city (Doongaji 1986).

While Egypt, India, and other early civilizations employed various principles to regulate the use of corporal punishment, no uniformity in its actual application seems to have emerged. A collection of some 200 early Hittite prescripts from the thirteenth century BCE, for instance, contains but a single unambiguous reference to penal amputation, namely, the practice of cutting off the nose and ears of a slave who burglarized a house (Hoffner 1997, §95). One other reference, which may or may not allude to beating a thief with a spear (§101), mentions it as a past practice, to be replaced by a monetary fine. Ancient China is another case in point. While Chinese law had no qualms about using physical torture to obtain confessions, and despite the common use of capital punishment, attitudes toward corporal punishment throughout that long period were discerning and even ambiguous. Under the Han Dynasty (206 BCE–220 AD), mutilations such as tattooing and cutting off the nose and feet were thought to be cruel and were commuted to extensive beatings ("bambooing") and various forms of hard labor. The shift from permanent to temporary harm had unexpected consequences, however, as the abuse of bambooing could and apparently often did end with death. To amend this situation, later Han ordinances repeatedly reduced the number of blows and defined the specific manner in which they were to be administered in order to ensure culprits' survival (Hulsewé 1955). The marginalization of corporal punishment and especially of maiming continued under later dynasties. The *T'ang-Yin-Pi-Shih*, a popular compendium of 144 legal cases stretching from *c.* 300 BCE to *c.* 1100 CE (van Gulik 2007), contains only three instances of flogging, one following a son's unintentional killing of his mother in the context of an attempt to prevent her from hurting a thief just apprehended (48A), another after the death of a food thief killed by his victim (48B), and a final one in response to a youth's entry into another's house at night (70A). The farther we look in time and space, then, the more diverse and complex the use of corporal punishment in antiquity becomes. Nor can we trace a steady decline over time in the application of this group of punishments or its commutation into seemingly milder measures such as incarceration or fines.

Greece and Rome

The quantity and variety of evidence surviving from Ancient Greece allows us to explore new or perhaps heretofore-hidden aspects of premodern crime and punishment. Authors such as Isocrates (436–338 BCE), in his *Areopagitus* (7.39–42), and Plato (d. 347 BCE), especially in Book 9 of *The Laws*, attained a high degree of theoretical abstraction in discussing the goals of punishment and even challenged its very legitimacy as a social institution. As far as corporal measures are concerned, the process we can detect in Classical Antiquity is not so much a reduction in usage as a sublimation and gradual disappearance of such penalties from the public sphere, accompanied by their proliferation in the private sphere, especially as regards non-free members of society. The same process characterizes Roman penal history to an even greater extent, as we shall see.

Written around the eighth century BCE, and possibly reflecting an earlier tradition, Homer's *Iliad* and *Odyssey* portray a society heavily reliant on self-help (private vendetta), where often harsh punishments were determined and meted out quite simply by the party that could. Subsequent centuries saw a gradual transition from private vendetta; to ad hoc, third-party arbitrators; to judicial assemblies; and finally to quasi-professional courts, which decided cases on the basis of increasingly detailed protocols, the most celebrated instances of which were produced under Draco in 621 BCE and Solon in 594 BCE (Saunders 1991). Within this process, corporal (but not capital) punishment largely fades from the public domain as it becomes associated with summary justice carried out in the domestic sphere, especially against offending slaves. Free citizens, the audience of most new laws and legal procedures, considered corporal punishment as particularly demeaning. According to Demosthenes (384–322 BCE), “the body of a slave is made responsible for all his misdeeds, whereas corporal punishment is the last penalty to inflict on a free man” (*Speeches* 24.167; and see 8.51; 22.55). Indeed, nothing could portray a ruler in more shocking terms than reporting that he had flogged free men. Accordingly, Thucydides (*c*. 460–400 BCE) observed how efficiently one exiled soldier incited his comrades by reporting (or in this case, misreporting) that under a certain regime “all were punished with stripes” (*History* 8.74.3). And in the *Athenian Constitution* (35), Aristotle (384–322 BCE) ominously describes the oligarchic Thirty’s rise by pointing out that they began ruling the city with three hundred “lash-bearers” before violently crushing their opponents.

Thus, the relative absence of flogging in particular and corporal punishment in general from most contemporary records, including legal and narrative sources, is not necessarily an indication of these measures’ decline (Robinson 2007). Under the radar of officialdom, but with the latter’s consent, corporal punishment crystallized into a highly effective mechanism for social othering, a measure particularly suitable for foreigners and non-free members of society. By and large this remained the case under Roman law, which consciously drew on ancient Greek law here as in numerous other matters. The Twelve Tables (449 BCE), a foundational text in Roman jurisprudence, prescribes beating prior to enslavement for anyone caught stealing in daylight. Slaves perpetrating the same offense are also to be scourged and then thrown off the Tarpeian Rock (II.5). Beatings are to precede execution for intentional destruction of property or grain and public verbal abuse or insult (VII.6 and 8). Last, a child shall be beaten at the discretion of the praetor if caught destroying or appropriating another’s crop at night and then forced to pay double the value of the damage (VII.4). Yet, as a sole punishment to be meted out to free adult citizens, beating became taboo: as in Greece, so in Rome, accusing a ruler of causing bodily harm as a form of punishment against free men was a sure way to taint his reign as especially, indeed extraordinarily, vicious (see Suetonius, *Caligula* 27). A similar point was implied in Late Antiquity by several Christian authors who described the bodily suffering of their brethren during the anti-Christian persecutions, which entailed beating as well as forced prostitution (*Acts of the Christian Martyrs*, Agape 5-6 and Pionius 7).

Status continued to serve as a determining factor in the application of corporal punishment throughout Roman history. Rome’s transition from kingdom to republic to empire, and the immense territorial expansion accompanying this process, culminated with the formal extension of citizenship to all free men under Roman rule. Yet while the Edict of Caracalla (212 CE) grew the political base of Rome, it also prompted a closing of the ranks among members of the elite citizenry. In the third and fourth centuries CE, a distinction between a class of *honestiores* (nobler) and *humiliores* (baser) men restricted the latter’s access to numerous legal privileges, also in the realm of penal law. In theory, flogging remained a licit penalty only as a form of summary justice reserved for foreigners, slaves, and children, and various offenses by these groups would have been punished in private by a *paterfamilias* or in public by the organ known as the *tresviri capitales* (Cascione 1999). Yet there is strong evidence to suggest that *humiliores* could be collapsed into the latter group for penal purposes. Claudius Saturninus, a Roman jurist probably writing in the early third century CE, authored many of the passages relevant to corporal punishment now preserved in Justinian’s *Digest* (early seventh century CE). Among his durable prescripts is one declaring: “All those whom it is not permitted to punish by whipping are persons that should have the same respect shown them that decurions have” (48.28.5). In other words, it was not or rather no longer enough to be a free citizen in order to avoid this humiliating form of punishment; the massive expansion of Roman citizenship prompted at best only a minor reduction in the use of corporal punishment.

Judaism, Islam, and Christianity

In and beyond the physical and chronological boundaries of Greek and Roman civilizations, three distinct religious traditions developed their own approaches to social control, including the application of corporal punishment: Judaism, Islam, and Christianity. Tracing these religions' pertinent ideas and practices, while not exhaustive of long-term trends, nonetheless offers a palpable link between antiquity and the medieval period, and their joint and individual legacies at present.

Mosaic and, later, Jewish law espoused corporal punishment via the early adoption of the *lex talionis*: "Eye for eye, tooth for tooth, hand for hand, foot for foot, burning for burning, wound for wound, stripe for stripe" (Exodus 21:24–25; and see Leviticus 24:19–20). Yet it is with particular respect to the latter measure – flogging – that Jewish jurists stand out in terms of the attention they paid to it and the high degree to which they promoted its practice (Goldin 1952). Operating under changing political fortunes and a generally declining autonomy, and eventually spread across the Near East and Europe, Jewish legislators turned flogging into a staple penalty, despite the measure's minor presence in the Old Testament. Indeed, flogging is explicitly prescribed only once in the Bible, in a passage specifying a limit of 40 blows to which a culprit may be subjected for a single offense (Deuteronomy 25:1–3). However, by the time of its completion in the early third century CE, the *Mishnah* expanded the use of penal whipping as applicable to some 168 offenses, including transgression of dietary, agrarian, and Passover laws (Makkot 3.1). Nor, apparently, and in contrast to earlier and contemporaneous cultures, was flogging directed at any specific social group or class, despite the claim by Flavius Josephus (*c*. 37–100), possibly influenced by a Greco-Roman tradition, that the measure "is the most ignominious for a freeman" (*Antiquities* 4.8.21 and 23). While corporal punishment proliferated, however, it did so at the expense of measures such as capital punishment and *kerait* or death by divine intervention. Perhaps reflecting a greater need for social cohesion under gentile rule, the authors of the *Gemara* (*c*. third to fifth centuries CE) and later commentators offered numerous exemptions from whipping. Still, at least at the normative level, corporal punishment remained common throughout and beyond the Middle Ages. The highly influential medieval Jewish physician and legislator, Maimonides (d. 1204), dedicated three whole chapters in his *Mishneh Torah* (Sefer Shoftim, Sanhedrin 17–19) to enumerating offenses punishable by flogging and to defining the terms of the measure's licit use (Jacob and Zemer 1999). And by and large this discussion was integrated into Jewish Halacha through its reception by Joseph Karo's *Shulchan Aruch* (1563), which remains until this day the most authoritative prescriptive text in and beyond Orthodox Judaism.

Unlike their Jewish counterparts, Islamic jurists dealt with punishment in general and corporal punishment in particular mostly within the context of a hegemonic culture; and in contrast with their Christian counterparts (see below), they did so in the absence of an identifiable separation between church and state. The Koran (mid-seventh century CE) enjoins the use of corporal punishment, and especially flogging, on several occasions as either a mandatory (*ḥadd*) or a discretionary (*ta'zīr* and *siyāsa*) penalty for a range of offenses, including slander and sexual misconduct. Subsequent oral and written traditions, however, glossed and at times supplemented such verses, giving rise to different schools of interpretation. For instance, while Koran 24:2 states that a "woman or man found guilty of sexual intercourse – lash each one of them with a hundred lashes," a later *ḥadīth* or oral tradition attributes to Mohammed a sentence of death by stoning against the woman and 100 lashes and banishment for one year against the man (Ibn Rushd, *Bidāyat Al-Mujtahid* 56.6.2). And while Koran 5:90 prohibits the consumption of wine, a corporal penalty for drinking alcohol was only devised much later, also on the basis of an oral tradition (Ibn-'Asqālani, *Bulūgh Al-Maram* 1064–66). Islamic law's adoption of the *lex talionis* facilitated the use of dismemberment in several cases, although in some instances, especially under Shiite law, the penalty could be commuted into a fine. Dismemberment was also prescribed to fulfill the social goals of incapacitation and deterrence, as in the case of theft: "[As for] the thief, the male and the female, amputate their hands in recompense for what they committed as a deterrent [punishment] from Allah" (Koran 5:38). Last, Koran 5:33 lists cutting off the opposite hand and foot (cross-amputation) as one among several penalties that can be imposed upon anyone found guilty of corruption. Beyond these contexts, corporal punishment in any form could be applied as a discretionary punishment whenever a defendant was awarded *ta'zīr*. Depending on the inclinations of a particular legal school, this could in theory be frequent, but in practice was subject to Islamic law's strict rules of evidence. As in Judaism, so in Islam, corporal punishment consolidated its status as a staple penalty in the Middle Ages, and so it remained throughout the early modern period, for instance, under the Ottoman Empire. In some Muslim countries, colonization by Western powers as well as independent state-building brought pressure upon legislators and judges to mitigate corporal punishment by its commutation into fines, especially in the case of women, and by replacing penal amputation with incarceration. Today corporal punishment continues to be practiced in several Muslim countries, especially wherever Shari'a law prevails as state law, for instance, in Saudi Arabia, Yemen, Sudan, northern Nigeria, Iran, and Pakistan (Peters 2005).

Early Christianity's blanket rejection of capital and corporal punishment (*sine vi humana*) in favor of, for instance, abstinence and incarceration, was to some extent at odds with hitherto prevailing notions of retributive justice, both Jewish and pagan. Yet physical pain did not disappear from the ecclesiastical penal landscape altogether. Church leaders and, later, canon lawyers promoted the use of, inter alia, severe fasts, flagellation (also self-inflicted), and exposure to extreme weather as ways to foster discipline and "medicate" against sin, first within ascetic and monastic milieus and later among laymen (Hillner 2009). From the perspective of a Greco-Roman class-based penal tradition, there surely was something deeply humiliating in free people's willingness to submit to physical pain, and this perspective was not lost on early church leaders, who, on the one hand, construed such afflictions as a "sweet inversion" of a pagan cultural taboo and, on the other, emphasized continuity with a biblical tradition (albeit one concerning the discipline of children rather than the punishment of criminals). Either way, throughout the Middle Ages, corporal punishment within the church appears to have been mostly limited to the regular clergy (monks) and rarely practiced among the secular clergy (priests) unless their punishment also involved cloistering. The church's capacity to apply its own range of penalties among lay people in this period varied widely across space and time, but some records suggest that corporal punishment was occasionally used as a main or additional measure (Plöchl 1955). From a broader cultural perspective, the notion that spiritual purgation involves physical pain continued to have wide purchase, as did the perceived causal relation between sin and divinely inflicted physical suffering, both of which were articulated and popularized throughout this period reaching an apogee of sorts in Dante's *Comedy*, published in the early fourteenth century.

The end of Roman Catholic hegemony in sixteenth-century Europe, along with the rise of the absolutist state, led not, as is often thought, to the overall contraction of religious law, including the regulation of corporal punishment. For, while the early modern period saw an increased policing of public morals by secular powers (the so-called criminalization of sin), it also witnessed an intensification of congregationally based discipline, especially (but not exclusively) where state churches were absent, as in the Dutch Republic (Schilling 1987). Neither process entailed a significant change in the role of corporal punishment, despite criminal law's renewed commitment to bloodless penalties under Protestant influences. Flogging and the stocks, for instance, remained minor and thus largely stable measures employed in a Christian penal context (Pihlajamäki 2006).

Premodern Europe

While Europe grew increasingly more homogeneous since Late Antiquity in religious terms, it also came to be characterized by political fragmentation and a concomitant legal pluralism. Different polities, occasionally conquering one another, operated under an array of traditions, including Roman and customary laws, which, juxtaposed with a more centralized church, shaped a diverse legal and penal landscape (Berman 1983). Contrary to popular opinion on the matter, the role of corporal punishment among these regimes was quite limited (Dean 2001); nor, once again, is it possible to trace a general decline in its use as a legal punishment throughout this period. The early medieval statutes contained in *Rothair's Edict* (643), for instance, display a typically strong preference for monetary compensation according to social rank (*wergild*) and functional damage, while later collections of German and French customary laws such as the *Sachsenspiegel* (*c.* 1220) and Louis IX's *Etablissements* (*c.* 1254) appear to have no qualms with dismemberment and branding. Among the Italian city-states in the fourteenth and fifteenth centuries, on the other hand, branding, amputations, and flogging were rare, and punishments tended to be overwhelmingly pecuniary (Dean 2007; but see Piasentini 1992), while in England, by the mid-thirteenth century, mutilations practiced under Norman rule had all but disappeared, only to rise again to prominence under the Tudors in the late fifteenth and sixteenth centuries (Bellamy 1973). The latter trend found a chronological parallel in the Netherlands, where branding and, more commonly, whipping were carried out both secretly and in public, and often accompanied by exposure on the scaffold (Spierenburg 1984). Last, throughout the Habsburg Empire, the enacted *Constitutio Criminalis Carolina* (*c.* 1532) paved the way for a broader use of judicial torture and corporal punishment (§§101, 104, 196–98), as was the wont of ever more powerful monarchs in France (Langbein 1974).

Modernity

It was thus perhaps more in response to the penal realities of the early modern period, rather than to any perennial premodern or even medieval state of affairs, that Enlightenment figures such as Cesare Beccaria (1738–1794), Jeremy Bentham (1748–1832), and others developed their revisionist views on punishment. Overall, these reformers rejected physical pain as a licit form of retribution or as a means to affect real psychological and behavioral change in favor of the more malleable punishment of incarceration. In the long run, the development of penal incarceration in Europe, and later in the US, was accompanied by dramatically reduced recourse to overt corporal punishment (Andrews 1994), although imprisonment itself entailed (and continues to pose) numerous physical risks and hardships. Ironically, while these reforms gradually contributed to the reshaping of Europe's penal landscape, Western colonial powers abroad remained quite tolerant and often even promoted the use of corporal punishment among indigenous populations across Africa, Asia, and the Americas, whether or not such measures predated their arrival (Benton 2002; Peters 2005). Judicial caning in present-day Singapore, Malaysia, and Brunei, for instance, can be traced back to British rule, although the measure has since become both a mandatory and a discretionary penalty, and its application has accordingly experienced a sharp rise. By the early nineteenth century, certain forms of corporal punishment such as amputation and branding had become a marginal measure in Western penology, as against a rise in the use of, on the one hand, imprisonment and, at least for a while, capital punishment (Weisser 1979). From a global perspective, however, flogging and beating still occupy a prominent role as a main or additional penalty in numerous countries, whereas in others they have been only recently banned. The last documented public whipping that took place in Delaware, for instance, dates to 1952, but the law allowing it was cancelled as recently as 1972.

There are other indications for the longevity of corporal punishment, which, as we have seen, for thousands of years has been a staple of summary justice and discipline in the semipublic and professional spheres, to say nothing of the private and domestic ones. Child psychologists and psychiatrists employ the term corporal punishment mostly to the chagrin of crime and punishment scholars (Ellison and Sherkat 1993; Straus and Donnelly 1994), who argue that transgressions in these environments rarely constitute crimes and consequently their desserts cannot be construed as punishments. Whipping a child for uttering profanity is a disciplinary measure meant to foster a certain behavioral norm that may or may not be strongly frowned upon outside the school or home, and slapping a spouse for sexual infidelity has less to do with enforcing a law than with expressing shame or frustration over a given situation. Conversely, and by comparison with the treatment of slaves in earlier periods, the law in many countries does not consider such measures as a form of summary justice. However, in certain domains such as the home and the school and in the religious and professional spheres, disciplinary action reflects (and in the past has certainly informed) mechanisms, procedures, and power structures in the world outside. In this sense nonpublic disciplinary action, including the use of corporal punishment, offers a political education and helps shape new generations' ideas of licit and illicit penalties. Moreover penalties meted out by the state are often in tune with those employed in domestic, religious, and professional contexts. Thus, a broad consensus regarding the need to protect children from domestic violence, for instance, is more apparent than real. As of January 2012, only 32 countries offer children full legal protection from corporal punishment. Despite no lack of local and global advocacy efforts, societies where corporal punishment is not applied to children remain the exception rather than the rule (Montague 1978).

From this perspective, the still prevalent use of corporal punishment (spanking, belting, slapping, whipping, etc.) in many milieus is significant. Rather than reduce it to a sign of immigrant maladjustment or attribute it to psychopathologies, it can also be interpreted as an enduring toleration and in some cases growing appreciation of physical pain as a legitimate disciplinary measure. In many present-day societies, the use of physical force by authority figures remains a battleground for competing truth claims about discipline and cultural identity. Despite consistent and growing evidence of spanking's negative short-, medium-, and long-term effects on children's mental health (McCord 1995; Afifi et al. 2012), parents, guardians, and community leaders continue to advocate it for two main reasons: first, it is irresistibly effective; children generally stop performing an undesirable action with alacrity after being struck. Secondly, it is condoned by certain authorities who lay claim to superiority on such matters at least in the domestic, academic, or religious sphere (Pearl and Pearl 1994). For instance, among those drawing inspiration from the Old Testament, some are reluctant to deny that "He who spares the rod, hates his child" (Proverbs 13:24). Cultural but not necessarily religious conservatives exhibit a similar attitude: the ban on corporal punishment in UK private schools was effected as recently as 1998 and by a narrow majority at that. And in the US, spanking is still the default prerogative of schools in a number of states. To these persistent forces countering the abolition of corporal punishment in homes and schools – and by implication in the public sphere – one must add an even wider faith in violence as a legitimate means to resolving disputes, the right to privacy, and religious and cultural autonomy (Straus and Donnelly 1994). As we edge into the twenty-first century, the high regard in which numerous societies continue to hold each of these principles suggests the durable attractiveness of corporal punishment. Modern penology may have encouraged a rethinking of the link between physical pain and punishment, but it has hardly been able to sever it (Garland 2011).

Recommended Reading and References

Afifi TO et al. (2012) Physical punishment and mental disorders: results from a nationally representative US sample. Pediatrics 130:1–8

Andrews RM (1994) Law, magistracy and crime in Old Regime Paris, 1735-1789. Cambridge University Press, New York

Bellamy JG (1973) Crime and public order in England in the later Middle Ages. Routledge and Kegan Paul, London

Benton L (2002) Law and colonial cultures: legal regimes in world history, 1400–1900. Cambridge University Press, Cambridge, UK

Berman HJ (1983) Law and revolution: the formation of the Western legal tradition. Harvard University Press, Cambridge, MA

Burns N (1993) Literature review of issues related to the use of corrective force against children. Department of Justice, Research and Statistics Directorate, Ottawa

Cascione C (1999) Tresviri capitales. Storia di una magistratura minore. Editoriale Scientifica, Naples

Dean T (2001) Crime in medieval Europe, 1200–1550. Longman, New York

Dean T (2007) Crime and justice in late medieval Italy. Cambridge University Press, Cambridge, UK

Doongaji D (1986) Crime and punishment in ancient Hindu society. Ajanta, Delhi

Ellison CG, Sherkat DE (1993) Conservative protestantism and support for corporal punishment. Am Sociol Rev 58:131–144

Garland D (2011) The problem of the body in modern state punishment. Soc Res 78:767–798

Goldin HE (1952) Hebrew criminal law and procedure. Mishnah: Sanhedrin – Makkot. Twayne, New York

Hillner J (2009) Monks and children: corporal punishment in Late Antiquity. Eur Rev Hist (Revue eurpéenne d' histoire) 16:773–791

Hoffner HA (1997) The laws of the Hittites: a critical edition. Brill, Leiden

Hulsewé AFP (1955) Remnants of Han law. Brill, Leiden

Jacob W, Zemer M (1999) Crime and punishment in Jewish law: essays and response. Berghahan, New York

Langbein JH (1974) Prosecuting crime in the renaissance: England, Germany, France. Harvard University Press, Harvard

Lorton D (1977) The treatment of criminals in the ancient near east through the New Kingdom. J Econ Soc Hist Orient 20:2–64

McCord J (1995) Coercion and punishment in long-term perspective. Cambridge University Press, New York

Montague A (1978) Learning non-aggression: the experience of non-literate societies. Oxford University Press, New York

Moskos P (2011) In defense of flogging. Basic Books. New York

Newman G (1983) Just and painful: a case for the corporal punishment of criminals. Macmillan, London

Pearl M, Pearl D (1994) To train up a child. NJG Ministries, Pleasantville

Peters R (2005) Crime and punishment in Islamic law: theory and practice from the sixteenth to the twenty-first century. Cambridge University Press, Cambridge, UK

Piasentini S (1992) 'Alla luce della luna': I furti a Venezia (1270–1403). Il Cardo, Venice

Pihlajamäki H (2006) Executor diviniarum et suarum legum: criminal law and the Lutheran reformation. In: Mäkinen V (ed) Lutheran reformation and the law. Brill, Leiden, pp 171–204

Plöchl W (1955) Geschichte des Kirchenrechts II: Das Kirchenrecht der abendländischen Christenheit. Herold, Vienna, pp 1055–1517

Robinson OF (2007) Penal practice and penal policy in ancient Rome. Routledge, New York

Saunders TJ (1991) Plato's penal code: tradition, controversy, and reform in Greek penology. Clarendon, Oxford

Schilling H (1987) 'History of crime' or 'history of sin'? Reflections on the social history of early modern church discipline. In: Kouri EI, Scott T (eds) Politics and society in reformation Europe. Macmillan, London, pp 289–310

Schmidt E (1965) Einführung in die Geschichte der deutsen Strafgeschichte. Vandenhoeck & Ruprecht, Göttingen

Scott GR (1938) The history of corporal punishment. T. Werner Laurie, London

Spierenburg PC (1984) The spectacle of suffering. Cambridge University Press, Cambridge, UK

Straus MA, Donnelly DA (1994) Beating the devil out of them: corporal punishment in American families. Lexington Books, New York

Van Gulik RH (2007) Crime and punishment in ancient China: T'ang-Yin-Pi-Shih, 2nd edn. Orchid Press, Bangkok

VerSteeg R (2002) Law in ancient Egypt. Carolina Academic Press, Durham

Weisser MR (1979) Crime and punishment in early modern Europe. The Harvester Press, Hassocks

www.corpun.com

www.endcorporalpunishment.org

Yelyr R (1941) The whip and the rod: an account of corporal punishment among all nations and for all purposes. Gerald G. Swan, London

History of Corrections

Logan M. McBride
The Graduate Center, The City University of New York, New York, NY, USA

Synonyms

Imprisonment; Reformation; Rehabilitation

Overview

With more than two million adults serving time in American prisons and jails, another five million adults under forms of correctional supervision such as probation or parole that threaten imprisonment for violators, and an incarceration rate of more than 700 people per 100,000 population, the United States leads the world in incarceration (Glaze, 2011). Although imprisonment has come to dominate American corrections, the prison as a form of punishment is a relatively recent historical development, emerging as part of a transnational conversation in the United States and Europe in the late eighteenth and early nineteenth centuries, and later in various forms throughout the rest of the world. To be sure, while neither the use of confinement nor the loss of liberty as punishment was invented in the United States, American penal innovation greatly influenced corrections on a global scale.

The history of corrections in America is one of failed reform. In response to the economic, political, and social concerns of each historical moment, reformers, politicians, and the American public have embraced various models of corrections emphasizing deterrence, retribution, incapacitation, and/or rehabilitation. But the reforms implemented to reduce crime, solve prison overcrowding, or reduce recidivism, for example, often generated additional problems, which combined with other contemporary concerns yielded further reforms. What follows is an overview of these shifting practices and goals in American corrections, from the colonial period through the present and in a transnational and global context. The entries in this volume examine methods of corrections that have at different times in history been endorsed, sometimes instead of, but often as a supplement to, imprisonment.

Corrections in Colonial America

Punishment in English colonial America reflected the legal and cultural traditions of the mother country, with adaptations appropriate to local conditions and needs. Like their counterparts across the Atlantic, colonial courts meted out a variety of public, corporal, and humiliating punishments as a way of deterring crime. Depending in part on the type of offense, but more often on the status and history of the offender, punishments might include the stocks or pillory, the whipping post, flogging, ear cropping, the wearing of signs, branding, or standing in the gallows. Fines were commonly imposed for lesser crimes and first offenses and were sometimes accompanied by, or substituted with, a shaming or corporal punishment. The most serious crimes incurred sentences of banishment or even the gallows. The aim was not to reform the individual offender, but to produce such a grim spectacle to deter a fearful public from sinful acts of crime (Kealey 1986).

The types of corporal and shaming punishments employed throughout colonial America derived from the English criminal code. But although the harsh “bloody code,” as it later came to be known, was in force in the colonies, local conditions and British salutary neglect resulted in its less frequent and less intense application in America. The English criminal code enumerated 50 crimes that were punishable by death when it was introduced in 1688, and over the next century the number of capital crimes increased to more than two hundred. However, a practice of royal and clerical pardons resulted in only a fraction of those condemned in England being hanged, and in seventeenth-century colonial America, a shortage of labor discouraged the frequent use of maiming punishments or the gallows, resulting in an even lower rate of execution than in Britain (Morgan 1985; McLennan 2008; Rothman 1971).

As in England, criminals in the American colonies were not typically sentenced to indefinite or extended confinement in workhouses or jails, nor were such institutions equipped for long stays. Jails were used to ensure an individual’s whereabouts either until his trial or sentence was complete or until any monies owed might be recuperated, and workhouses were utilized most often for vagrants. For the most part, colonists saw little use for imprisonment as punishment, holding that it served neither as a public deterrent nor as a means of securing labor, but instead created a class of dependents that developing colonial economies were ill equipped to sustain.

Some colonists had come to view English punishment as sanguinary and unjust, advocating instead the use of confinement and hard labor. In Quaker Pennsylvania, William Penn's 1682 "Great Law" reduced capital offenses to murder and designated imprisonment in the workhouse for most criminal offenses. The law and the experiment with imprisonment were short-lived, however; it was repealed in 1718 and the English criminal code was once again reinstated.

Britain's system of pardons provided many condemned prisoners with reprieve in the form of banishment. These men joined the masses of convicted young larcenists on a journey across the Atlantic to the British colonies in America or the West Indies. With the passage of the Transportation Act of 1718, banishment became the leading punishment for property offenses, the most common form of crime in Britain. In some jurisdictions, banishment made up more than half of the sentences for all convicted criminals. Between 1718 and 1775, as war approached, more than 50,000 British convicts were transported to American colonial shores. Transported convicts made up roughly one quarter of all British immigrants to the American colonies during this period, and one half of all English immigrants. Convict transportation, some in England argued, not only solved the problem of crime at home, but also helped to solve the problem of labor in the colonies. On the other side of the Atlantic, especially in the Chesapeake where the majority of convicts debarked, many colonists resented serving as England's dumping ground for rogues, anxious about the crimes that these felons might – and did – commit. More enterprising colonists saw an opportunity for cheap labor. Shippers carried convicts and indentured servants in the same hold and auctioned their labor to the highest bidders in the colonies for profit and to defray the cost of transport. Like indentured servants, though commonly at a fraction of the price, convicts sold into penal servitude were contracted to labor roughly 7 years before being freed. But whereas indentured servants were more easily absorbed into colonial society upon expiration of their contract, penal servants faced persistent stigma. Colonists tended to identify penal servants more closely with their black slaves than with white servants, passing legislation that restricted penal servants' legal rights to testify in court, vote, and even receive freedom dues, a customary land grant provided upon expiration of the term of indenture. As war with the American colonies loomed, British convict transportation ceased in 1775, leaving both the British and the fledgling United States to seek new solutions to the problem of corrections (Ekirch 1985; Grubb 2000; Morgan 1985).

The Birth of the Prison

Historians have offered several explanations for the birth of the prison in Europe and America in the late eighteenth and early nineteenth centuries. Perhaps most famously, Michel Foucault situated the shift from corporal punishment focusing on the body to disciplinary punishment centered on the soul within the context of a growing population and increasing class tensions, an industrializing economy, and the emergence of human sciences. While the humanitarian concerns of reformers worked to gain public acceptance of the prison, Foucault argued, it was the demands of the modern industrial age for disciplined bodies that led imprisonment to become the dominant form of punishment in the west (Foucault 1977).

Foucault emphasized the increase in access to the internal life of the citizen implied by this shift. While early colonists and Europeans left determinations about the soul to God, their nineteenth-century successors believed they could impel internal change through control of the prisoner's body. With convict transport cut off during the American Revolution and increased social and economic dislocation as a result of industrialization, the promise of reforming delinquents into productive laboring classes through imprisonment was appealing to England's upper classes who grew anxious about rising crime and a growing population (Ignatieff 1978). Reformers, such as William Allen, Elizabeth Fry, John Howard, and Jeremy Bentham, supported the use of imprisonment on religious, humanitarian, and utilitarian grounds, as a means of correction and reformation. The Penitentiary Act of 1779, influenced heavily by Howard's 1777 report on *The State of the Prisons in England and Wales*, authorized construction of two new prisons, which would attempt to reform convicts through the use of solitary confinement, disciplined labor, and religious instruction. The Act went largely unenforced, however, especially as by 1787 convict transportation had resumed, this time to colonies in New South Whales (Australia) and Van Deiman's Island (Tasmania). Similarly, in France, legislation drafted in the wake of the French Revolution providing for the construction of prisons largely went unfulfilled. Just as in the United States, it would be another several decades before Europe embraced a penitentiary system.

The American Penitentiary

In the United States, the 50 years following independence marked both the rise and the institutionalization of the penitentiary. Although the forms and aims of corrections changed dramatically in only half a century, it is important to remember that the shift from public corporal punishment to a reliance on imprisonment was gradual and incomplete. Legally sanctioned public executions were carried out well into the twentieth century, and corporal punishment remained a mainstay of prison discipline throughout the nineteenth and twentieth centuries. Moreover, calls for the reform of American corrections had many different motivations, from religious and humanitarian concern to intellectual and cultural shifts, class anxiety, and social and economic changes. Imprisonment was one of many efforts to bring moral and practical order to society, including schools, almshouses, and asylums.

The first wave of reform came in the 1780s, as many states, led by Pennsylvania, revised their criminal codes to reduce the number of capital offenses. By the eighteenth century, first-degree murder remained the only crime punishable by death in most northern states. America's ruling intellectual and elite classes were guided by Enlightenment concepts of rationalism, humanity, and the social contract, and these ideas also influenced their notions about effective corrections. As in Britain there was anxiety among the elite over social instability and the unruly lower classes, but part and parcel of the work of forging a new nation was establishing a system of punishment that aligned with the new republican ideals. Many reformers who opposed capital punishment did so not only on humanitarian grounds, but argued that it was ineffective, unchristian and even dangerous for society. The alleged effectiveness of shaming punishments in the colonial era relied on a close-knit, face-to-face society where the scorn of neighbors and friends was humiliating to the offender and, crucially, onlookers could imagine themselves in the convict's place. But with the demographic changes and the increase in mobility and size of the population that took place after the American Revolution, shaming lost its effectiveness. Further, capital punishment, Thomas Jefferson wagered, did little to maintain loyalty between a citizenry and its government. Moreover, Benjamin Rush argued, such public displays of violent execution had a brutalizing effect on spectators, spurring further criminal activity. Early American juries often failed to convict defendants who faced punishments they deemed too harsh for the crime, even when they believed the defendant to be guilty. Thus, reformers opposed the widespread and public use of capital punishment, embracing instead the proportionality advocated by Italian philosopher Cesare Beccaria as most efficacious. Beccaria's *On Crimes and Punishments*, translated into English in 1767, was widely read by America's intellectual classes (Masur 1989; McLennan 2008).

But as states turned away from the use of public corporal punishment, it was still unclear how exactly convicted criminals should be punished. Pennsylvania embarked on a brief experiment in restriction of liberty, sentencing convicts to hard labor on public roads. These work gangs were meant to serve as a warning to onlookers of both the penalty for crime and a reminder of the power of the state to punish. In fact, the laboring gangs often drew mocking mobs and crowds of sympathetic spectators who sometimes provided convicts alms. Concerned about the corrupting influence of the convicts on the public and chagrined by the jeering crowds, local leaders scrambled for an alternative, revising the state's criminal code once again in 1789 to provide for sentences of imprisonment at hard labor in Philadelphia's Walnut Street Jail. Opposition to imprisonment as punishment was relatively minor but was made both on religious and republican grounds. Strict Calvinists believed that only biblical punishments could be justly imposed. Others argued that imprisonment threatened Americans' property rights because it would require raising taxes to maintain convicts in confinement – a breach of the social contract. Still, critics of the prison were relatively few (McLennan 2008; Meranze 1996; Masur 1989).

In 1790, the Pennsylvania legislature established Walnut Street as the state prison and ordered the construction of the penitentiary house on the grounds of the jail to include 16 solitary cells for the most hardened criminals. Most convicts and those still awaiting trial were held in the jail's large rooms at night and spent their days in the yard, shoemaking, nail making, weaving, or stone sawing to defray the cost of their imprisonment and to earn small amounts of money for use upon release. Conditions were stark; inmates frequently grew ill, and as the jail's population swelled, so too did tensions, idleness, and violence. New York City's Newgate Prison, built in 1796, faced similar troubles. Inmates at Walnut Street, Newgate, and other early American prisons protested their forced labor and squalid living conditions through disobedience, work slowdowns and stoppages, destruction of tools and machinery, and even open rebellion.

By the second decade of the nineteenth century, overcrowding, inmate violence and escapes, as well as administrative and financial disputes forced authorities to acknowledge the failure of these early institutions and spurred the second wave of reform. Evolving and competing penological philosophies, coupled with economic, social, and political concerns in Jacksonian America, led to the development of two distinct institutions and systems of imprisonment at Eastern State Penitentiary in Philadelphia and Auburn Prison in upstate New York. Both prisons became models of a new type of penitentiary, replicated across the country and around the world. Perhaps more important than the differences between the systems was the fact that they marked the institutionalization of incarceration as the American response to crime.

Classification and segregation of convicts at Walnut Street was legislated in 1790. However, the implementation of this practice was limited by the physical structure of the prison. With only 16 individual cells, the remaining convicts were housed together, in groups of 20 to 30 per cell. As Walnut's inmate population gradually swelled, the state legislature succumbed to public pressure and authorized the construction of two penitentiaries – Western State near Pittsburg in 1818 and Eastern State in Philadelphia in 1821 – that would be large enough to accommodate every inmate in solitude. Designed by the famous architect John Haviland, the impressive Eastern State Penitentiary opened its gates in 1829. It was built on the theory that removing the fallen individual from all potentially negative influences, functionally isolating him completely, would furnish him with the necessary conditions for reflection, repentance, and finally, redemption. To facilitate the reformation of the criminal in this introspective manner, Eastern housed inmates individually for the duration of their sentence, providing only a Bible, occasional handicraft work, and his conscience to occupy his time. As such, each of the initial 250 cells boasted plumbing, heating, and an adjacent private exercise yard surrounded by 10-foot high walls. The vaulted ceilings in each cell were meant to encourage spiritual reflection and the thick walls between cells served to prevent communication. Visitors and correspondence were strictly forbidden, the only exception being calls from the minister. (Teeters and Shearer 1957).

Using Bentham's panopticon as a model, Haviland designed the prison with a central rotunda from which one jailer could monitor activity on all of the seven radiating cellblocks. An imposing stone wall surrounded the prison, 12-ft thick at the base and 30 ft high to keep the public out and the convicts in. The thick-walled design of the prison's cells and the administration of the facility all served to enforce a strict rule of silence. Even so, inmates subverted authority by tapping messages to one another on pipes. Although such a focus on the redemption of the soul promised a more humanitarian form of punishment, to keep order, jailers again turned to the body, this time behind the prison's thick walls (Meranze 1996).

Construction began on New York's Auburn Prison in 1816. Like Newgate, it was initially patterned on Walnut Street Jail. Reformers and legislators believed that the problems at these early prisons stemmed from overcrowding and that more space would solve the matter. By 1819, however, the situation had not improved at Newgate and Auburn had developed its own problems. In response to a particularly heinous riot at Newgate in 1819 the legislature took action, authorizing corporal punishment in the prison and ordering the grading of prisoners by offense, housing the most hardened criminals in solitary confinement. That same year, construction began on what would become Auburn's distinctive cellblock: two rows of individual cells, measuring 7 ft long, 3 ½ feet wide, and 7 ft high, were placed back to back and stacked five-stories high inside of a larger building. Construction was completed in 1821. The first 80 men selected for the new punishment of solitary confinement were locked in their cells on Christmas Day of that same year. There they were meant to remain for the duration of their sentence. Over the first year, jailers were alarmed by high rates of illness, injury, death, and insanity among these men. In 1823, New York's Governor pardoned all survivors and New York abandoned the sentence of solitary confinement.

A new system of penology began to develop at Auburn during the 1820s. Prison administrators and reformers found value in silence and seclusion but regarded the complete and total isolation of inmates as unnatural and inhumane. Thus, the new Auburn boasted common dining and work facilities. Still isolated from the outside world and forbidden to have contact with their families, prisoners left their cells during the day to work and eat in silence among their fellow inmates. They marched in lockstep, a form of walking developed at Auburn to prevent communication among inmates during movement, with each man holding his arms just under those of the man in front of him, and turning his heads to one side, to face the guards while marching in unison. The Auburn system incorporated the advantages of convict labor – inmates learned skills and the value of hard labor, vital to their reform, while contributing to the cost of maintaining the prison – into a silent system. Indeed, an important part of Auburn's promise to the state was the reform of inmates at little or no cost to taxpayers. Because inmates worked and dined in communal spaces, cells were only used for sleeping so could therefore be made smaller, thus costing less to build and maintain. The construction cost in the 1820s and 1830s for a prison modeled on Auburn was $91 per cell, whereas in an institution patterned on the Pennsylvania system cost an astronomical $1,650 per cell (Conley 1980; Lewis 1965).

From the beginning, inmates worked to supply prison needs. In addition to providing physical labor for the construction of the prison itself, inmates worked on the institution's farm to supply food, in the kitchen to cook meals, and did basic general maintenance. Auburn's very small female population, first admitted in 1825, performed much of the sewing, ironing, and laundry. Before long, private industry had infiltrated the prison and convict labor was contracted to the highest bidder. Inmates worked in the production of barrels, hats, chairs, tools, locks, shoes and boots, clothing, carpets, buttons, combs, harnesses, brooms, clocks, buckets and pails, wagons, and even rifles. The introduction of mulberry trees in the late 1830s to begin silk production speaks to the diversity of industry in the prison and the extent to which prisoners were now considered part of the low-skilled pool of cheap labor that also included women, children, immigrants, and African-Americans (McLennan 2008).

Both the Pennsylvania and the Auburn systems were considered reformist and modern, and the merits and disadvantages of both were hotly debated by reformers and penologists. Thousands of visitors from across the country and around the world flocked to the two prisons; emissaries from Europe and Latin America toured the institutions, taking note of the system best suited for their nations. Notable travelers included Gustave de Beaumont and Alexis de Tocqueville in 1831 and Charles Dickens in 1842. Blueprints and notes on the American penitentiaries could be found as far flung as China, Japan and, Mexico, and Peru (Botsman 2005; Salvatore 1996).

Europe embraced the Pennsylvania system perhaps because it offered a clearer alternative to the workhouse model that Auburn more closely resembled; Holland and England had employed workhouses in corrections in the seventeenth and eighteenth centuries (Spierenburg 1987). The Pennsylvania system was also met with some popularity in Latin America, where, however, lack of funds and local racial and ethnic ideologies resulted in only its partial implementation. But it was the Auburn system that came to dominate American penology, in no small measure because it fell in line with the new ethic of production ushered in by industrialization. The Auburn system allowed for not only a self-supporting prison system but one that could turn a profit.

As the century wore on, the drive for profit and the strict imposition of silence at Auburn and other prisons built on the Auburn model resulted in the widespread use of physical punishment. Inmates were whipped; given “baths” (bound hand and foot, with necks braced, and placed in a barrel while a steady stream of ice water dropped from a great height onto their heads); forced to wear the 40-lb iron “yoke” across their necks; or were “bucked” (hung upside down in their cells until they passed out). Just as reformers in the early nineteenth century railed against the brutish sanguinary punishment of the colonial days, reformers in the late nineteenth and early twentieth centuries clamored for the abolition of these inhumane practices.

An Era of Reform

In the antebellum South, prisons and other state-supported institutions were less common than in the North. Where they did exist, state prisons were typically smaller models of the Auburn system that embraced contract labor, usually in agriculture rather than in workshops. Their inmate population was almost exclusively white. Masters commonly punished slaves in private and on the rare occasions when slave punishment was carried out by the state it was usually corporal and administered in public. Even after the Civil War and the abolition of slavery and despite the rewriting of legal codes such that the numbers of convicts in Southern states increased dramatically, the Southern penitentiary remained largely the domain of whites. Racist pseudoscience supported popular beliefs that blacks were inherently deviant, beyond reform, and thus not well served by a stint in prison. Instead, white southerners argued, lacking the institution of slavery for their management, blacks needed to be taught labor discipline. A combination of this type of paternalism, anxiety, and racism coupled with pressing economics drove the development and institutionalization of a system of convict leasing in the South. Moreover, the language of the Thirteenth Amendment that abolished slavery in the United States and its territories made a provision for involuntary servitude for parties duly convicted of a crime. A post-Reconstruction backlash against black economic and political gains resulted in the criminalization of large numbers of African-Americans, many of whom were then rented out to private firms or individuals for farming, mining, or other hard labor, with all revenue returned to the state. The lease system offered little economic motivation for the renter to protect the health or life of the prisoner, and convicts faced grueling work under brutal conditions. Death rates among leased convicts in the South were extremely high, leading reformers to argue that the lease system was worse than slavery. Yet convict leasing persisted in most Southern states well into the twentieth century. Eventually convict leasing gave way to the chain gang; rather than renting labor to private hands, the state retained it (Curtin, 2000). So horrific were conditions even into the second half of the twentieth century that some inmates engaged in self-mutilation, the most common form of which involved severing the Achilles tendon, or heel stringing, so as to avoid forced labor (Perkinson 2010).

In the North, the inhumanities of the prison system and convict labor in the late nineteenth and early twentieth centuries were more understated, tucked behind the thick stone prison walls and out of the public eye. State budgets for prison building and maintenance in the post-Civil War period were tight; new prisons were built with small, cheap cells and older institutions fell into disrepair. Plagued by overcrowding, prisons often slept two to three inmates per cell, far from the reform ideals of silence and isolation espoused in the early nineteenth century. Dense and decrepit living conditions had severe health and psychological consequences for inmates. Moreover, with little centralized oversight for most state prisons, wardens often relied on corporal punishment to keep order and drive labor. Reverend Enoch Cobb Wines and Theodore Dwight's 1867 *Report on Prisons and Reformatories of the United States and Canada*, commissioned by the Prison Association of New York, highlighted the deplorable conditions in American prisons and called for an overhaul of the system such that prisons might more genuinely work to reform the convicts sentenced within. The report served as the catalyst for an historic meeting of the First National Prison Congress in 1870 in Cincinnati. There, delegates from more than 24 states as well as from Canada and South America adopted a Declaration of Principles that viewed crime as a moral disease to be treated through revised sentencing that could provide rewards for good behavior, education and industrial training for prisoners, the abolition of contract labor, the scientific collection of criminological data, and the professionalization of prison staff. Influenced by innovations in penal reform in Australia and Europe, namely, William Crofton's "Irish System," the Declaration of Principles called for the provisions of early release, or parole, and indeterminate sentencing. The Principles made clear that it was in society's interest to rehabilitate the convict, and it was the goal of the prison to remake the delinquent as a productive, self-respecting citizen (Rotman, in Morris and Rothman, eds., 1995).

The Declaration of Principles stood as the cornerstone of the progressive prison reform movement in the late nineteenth and early twentieth centuries, guiding American sentencing and corrections policy over the century. It reflected the effort of Progressive Era reformers who, whether motivated by humanitarian concerns or social anxieties relating to mass immigration and the rapid industrialization and urbanization of the period, sought to improve society's moral fabric by making it more efficient and less corrupt. Embracing a scientific, positivist approach to criminology, these progressives sought to model the prison on secular society, focusing on positive learned behavior rather than religion as the vehicle for reform. Progressive "new penology," as it was called, left its mark on American corrections with innovations such as probation, parole and indeterminate sentencing, reformatories, and the juvenile court system (Rothman 1980; Sullivan 1990).

Penologist Zebulon Brockway, commonly credited with first implementing parole and indeterminate sentencing in the United States, is viewed as the father of the reformatory movement. Although "good time" laws that reduced time imprisoned for good behavior were in existence as early as 1817 in New York, such legislation was employed more as a method to control overcrowding rather than a tool to encourage rehabilitation of the offender. With the shift in Progressive Era penology toward reform based on learned behavior, parole took on new meaning and came into wider use in the late nineteenth century.

In 1876, Brockway was appointed warden at New York State's brand new Elmira Reformatory for youthful male offenders, aged 16–30. After successfully lobbying the state legislature in 1877 for a law permitting indeterminate sentencing with parole, Brockway was able to implement the Declaration of Principles' ideas for individualized, reward-based reform through a graded merit system at Elmira.

Inmates were classified according to their educational and work performances as well as general conduct and psychological makeup. Good behavior allowed promotion to a higher grade with greater privileges and eventually early paroled release. Poor behavior, on the other hand, would result in demotion to a lower grade, the loss of privileges, and longer periods of confinement. Young men in the lowest grade were subject to the older penal practices, such as wearing bright uniforms, being prohibited from receiving visitors or correspondence, and being required to walk in lockstep. Similar provisions for parole and indeterminate sentencing were soon widely adopted. In 1910, legislation enacted parole for federal prisoners; and by 1927, Florida, Mississippi, and Virginia were the only states without a parole system.

Through a program of (gendered and racialized) vocational training and an appropriate system of rewards, Brockway and other "new penologists" believed the criminal would be reformed and society would be well served. Progressive wardens across the country implemented programming that they hoped would inspire a sense of responsibility, self-worth, and citizenship in their inmates. Elmira's inmate-written newspaper, the first of its kind, began publication in 1883. This period saw the introduction of inmate sports teams, musical concerts, lecture series, and later, films to break the monotony of the prison routine. At Sing Sing Prison, just north of New York City, warden Thomas Mott Osborne famously instituted an inmate self-governing organization, "The Mutual Welfare League," which was responsible for organizing entertainment, managing the prison newspaper, and even maintaining discipline. For good work and behavior, inmates earned fake money with which to pay for the cost of living, including renting cells – the more desirable cells being more costly. By simulating free society in the interest of the prisoner, Osborne argued, the convict could be transformed into a responsible citizen (Rothman 1980).

In addition to these inmate-focused programs, Progressive prison reformers encouraged the professionalization of a prison staff and relied on the analysis of new penological experts. The late nineteenth and early twentieth centuries witnessed the professionalization of many fields, including medicine, law, and education. The burgeoning fields of sociology and psychology provided scientific and medical interpretations of criminality that reinforced reformers' notions about rehabilitation. Extending the medical analogy, "new penologists" viewed the convict as a sick patient who would be released once cured. The prison thus functioned as a hospital, and its staff made up the nurses and doctors. Classification of inmates was based on psychological and sociological categories and allowed for individualized treatment. Progressive trends toward specialization and scientific management were apparent in American corrections with the creation of separate facilities for mentally disabled offenders, for women, and for juveniles and in the classification of men's facilities as either maximum or minimum security.

Unsurprisingly then, progressive penology and the reformatory movement brought major changes to the administration of criminal justice for women. For most of the nineteenth century, female prisoners were so few in number that they garnered little attention from prison administrators. Moreover, because Victorian sensibilities regarded women as the moral sex, a female criminal was considered depraved and incapable of reform, much like black convicts in the South.

Viewed by prison administrators as nuisances, female convicts were usually housed together, tucked away in an attic or area adjacent to the men's institution. There the women performed cost-saving labor such as sewing, laundry, and ironing but were generally left to their own devices, receiving little exercise or fresh air for their trouble. Neglect and abuse were common.

In the late 1860s, middle-class women reformers in the Northeast and Midwest became concerned with improving conditions for the growing number of incarcerated women. To protect them from abuses, reformers argued, female offenders should be housed separately in reformatories and looked after by female matrons. The reformers did little to account for female felons, who would remain in custodial settings. Rather their interests lay with "fallen women" such as prostitutes and unwed mothers. Not previously subject to state punishment, women charged with such moral crimes were seen as ripe for betterment and given indeterminate sentences by the courts, the length of which would be figured by their progress at the reformatory. Thus, reformatories focused on a "redeemable" population of young, almost exclusively white women convicted mostly of misdemeanant public order crimes. Custodial prisons, on the other hand, tended to hold older, more hardened, and disproportionately African American female criminals, women convicted of felony, violent, and property crimes.

Reformers sought to feminize the female reformatory to suit what they saw as women's more passive nature, believing that fallen women could be rehabilitated given the proper conditions and training. Unlike male prisons, which held convicts behind metal bars and high walls, women's reformatories tended to house small groups of inmates in cottages in rural settings. Progressives believed that in a familial arrangement, with exposure to the outdoors and fresh air, and under the guidance of morally exemplary matrons, fallen women could be redeemed. Reformatories reinforced traditionally accepted gender roles through the threat of incarceration in state penitentiaries, parole revocation, and the transfer of intractables – women who refused to reform – to asylums for the feebleminded. These institutions were often run by well-educated, professional women, such as Katharine Davis at Bedford Hills in New York, who later went on to serve as the first female Commissioner of New York City's Department of Correction (Freedman 1981; Rafter 1990).

The Reformatory Movement, finally established by the early twentieth century, faded in the 1930s as support for Progressive reforms waned. During this period, national concern with prostitution diminished, feminist momentum slowed, and funding was sparse. As states closed women's units in state prisons, the female felon population was transferred to reformatories, displacing the petty offenders who had been the focus of the Movement. With felons filling reformatories, misdemeanants were once again sent to local jails, and the system turned to a custodial model with remnants of the Movement: separate female facilities and differential treatment that locked women into gender roles (Rafter 1990).

The progressive efforts at classification and individualization had long-lasting effects on the application of criminal justice for youth. Prison reformers had long been concerned with the corrupting influence that hardened criminals might have on youthful offenders. But until the late nineteenth century, only piecemeal efforts were made to accommodate young people in the criminal justice system. All of that changed with the establishment of the juvenile court in 1899 in Chicago. This court created a completely separate system for juveniles charged with offending the law, and it attempted to decriminalize youth by removing them from criminal court. The separate court promised a more individualized approach to dealing with juvenile delinquency and with it a rehabilitative ideal that utilized probation over incarceration. The Chicago model quickly spread throughout the United States. By 1912, 22 states had juvenile courts, and by 1928, only two states lacked them (Binder 2001).

The juvenile court system also helped facilitate the more rapid adoption of probation for youth and adults. In the mid-nineteenth century, the practice of probation was born through the actions of religious and wealthy Bostonian named John Augustus. Between 1841 and his death in 1859, Augustus provided bail for more than 1,000 people in the Boston courts, convincing the judge to release offenders into his charge temporarily, during which time they could persuade the courts that they had been reformed. Massachusetts adopted probation for juveniles in 1878, and, following the establishment of a juvenile court system, other states quickly followed suit. New York was the first state to adopt probation for adults in 1901. Probation, along with parole and indeterminate sentencing, furthered the Progressive Era aims of individualizing punishment.

Despite these gains, prison reform of the late nineteenth and early twentieth centuries fell far short of the Progressives' professed goals. Their efforts to model prison after society were moderately successful. They succeeded in abolishing the rule of silence and the use of striped uniforms as well as in allowing for greater communication with family and in bringing educational and vocational training and some entertainment into prisons. But behind prison walls, physical abuses continued. Even at Elmira, a state investigation revealed in 1899 that inmates were whipped (Pisciotta 1994). Overpopulation and lack of funds persisted and were compounded by resistance to changes from poorly paid prison staff. In a bid for efficiency, Progressives created large prisons managed by a centralized professional bureaucracy (Rotman, in Morris and Rothman, eds., 1995).

During this period, the federal prison system expanded rapidly. Prior to 1891, federal offenders were mostly held in state facilities. Ground was broken on the first federal prison in Leavenworth, KY, in 1891 (though construction lasted 30 years) and the first federal prison for women was established in 1902 in Alderson, WV, modeled on the cottage-style reformatory.

Post-WWII Corrections

By the 1950s, long years of tight budgets from the Depression into WWII had left prison facilities in poor condition. Overcrowding, lack of medical facilities, bad food, poor management, unprofessional staff, and endemic brutality led to a string of riots across the country in the 1950s. Experts charged that the problem stemmed partly from a lack of rehabilitative programming. Postwar optimism, economic prosperity, low crime rates, and a faith in psychological treatment encouraged the embrace of rehabilitation by both the public and correctional professionals. Prisons in the late 1950s and 1960s saw great changes in their physical environments. Common areas became more plentiful and more spacious; paint colors were selected to brighten the facility; gyms, libraries, and classrooms were added to prison grounds. The quality of food improved, educational and vocational programs were increased, and discipline was more relaxed, giving inmates an opportunity to form study groups and attend counseling and training sessions.

At the same time, demographic shifts, urban crisis, and new legislation and law enforcement practices changed the face of America's prison population. As the number of incarcerated Americans grew, so too did the disproportionately high numbers of black, Latino, and Native American men, women, and youth who spent time behind jail and prison bars.

But despite the apparent improvements, mistreatment of inmates and poor conditions persisted. Led by the Black Muslims, a prisoners' rights movement gained momentum in 1965 with the Supreme Court ruling in *Cooper v. Pate*. The ruling gave inmates standing in the courts and initiated federal involvement in state corrections to protect the rights of prisoners. Change in American prisons frequently depended on inmate initiative – through the courts or by other means. Inspired in part by tactics of protest employed in the civil rights movement, inmates across the country engaged in collective and individual acts of resistance including hunger strikes, work stoppages, and prison takeovers to call attention to their inhumane treatment. Inmate writing in the form of published memoirs, poetry, literature, and letters also helped to successfully garner the attention of the media, lawyers, and the general public. Facing public and governmental scrutiny, corrections staff felt under attack from inside and outside of the prison and clamored for greater protections and improved working conditions. Tensions mounted. Although for a variety of very different reasons, the general consensus was that American prisons were failing (Gottschalk 2006).

The crisis of corrections was manifest in the riots in the late 1960s and early 1970s, perhaps most famously at Attica in 1971. The political and public response was on the one hand concern over treatment of prisoners, but on the other, an urgent and loud call to stop coddling inmates and, fueled by fear of growing crime rates, to get tough on crime. The 1973 Rockefeller Drug Laws established mandatory minimum sentencing that eliminated judicial discretion and set a pattern followed in many other states (most famously in California in 1994 with the "Three Strikes Law") and on the federal level in the 1980s. The 1974 Martinson Report echoed this sentiment, arguing that rehabilitation did not work.

By the late 1970s, tougher laws, enhanced policing, and stricter sentencing had resulted in the rapid increase of the American prison population that showed no signs of slowing. Between 1970 and 1980 the prison population doubled, then more than doubled again over the next 15 years. This despite crime rates during the same period remaining stable or declining. Racialized fears of crime, shifting public sympathies to emergent victims' rights groups, and strengthened and politicized law enforcement and corrections union organizing worked together by the 1990s to guard against effective political challenges to the tough on crime politics that maintained high incarceration rates, especially among low-income and black and Latino Americans (Gottschalk 2006). With more and more Americans receiving longer and longer sentences, the prison population grew at a seemingly exponential rate. To accommodate the ballooning numbers of inmates, the country witnessed a rash of prison building in the mid-1990s. The politics around deindustrialization, urban crisis, government spending, the cold war, and the string of recessions beginning in the 1960s expanded and entrenched the use of incarceration as the dominant form of corrections, despite the general consensus based on extensive bodies of research that prisons do not work to reduce crime.

Related Entries

- ▶ Deterrence
- ▶ Deterrent Effect of Imprisonment
- ▶ Historical and Comparative Perspectives on Incarceration
- ▶ History of Bail
- ▶ History of Boot Camps
- ▶ History of Corporal Punishment
- ▶ History of Criminological Theories: Causes of Crime
- ▶ History of Fines
- ▶ History of Juvenile Justice
- ▶ History of Probation and Parole in the United States
- ▶ History of Substance Abuse Treatment
- ▶ History of the Death Penalty
- ▶ Rehabilitation
- ▶ Traditional Retributivism

Recommended Reading and References

Binder A (2001) Juvenile delinquency: historical, cultural and legal perspectives, 3rd edn. Anderson Publishing Company Cincinnati

Botsman D (2005) Punishment and power in the making of modern Japan. Princeton University Press, Princeton

Conley J (1980) Prisons, production and profit: reconsidering the importance of prison industries. Journal of Social History 14:2

Curtin M (2000) Black prisoners and their world, Alabama 1865–1900. University Press of Virginia, Charlottesville

Ekirch A (1985) A profile of British convicts transported to the colonies, 1718–1775. The William and Mary Quarterly 42(2):184–200

Foucault M (1977) Discipline and punish: the birth of the prison. Trans. Alan Sheridan. Vintage, New York

Freedman E (1981) Their sisters' keepers: women's prison reform in America, 1830–1920. University of Michigan Press, Ann Arbor

Gilmore R (2007) Golden gulag: prisons, surplus, crisis, and opposition in globalizing California. University of California Press, Berkeley

Glaze L (2011) Correctional populations in the United States, 2010. Bureau of Justice. http://bjs.ojp.usdoj.gov/content/pub/pdf/cpus10.pdf. Accessed 16 Feb 2012

Gottschalk M (2006) The prison and the gallows: the politics of mass incarceration. Cambridge University Press, New York

Grubb F (2000) The transatlantic market for British convict labor. The Journal of Economic History 60(1):94–122

Ignatieff M (1978) A just measure of pain: the penitentiary in the industrial revolution, 1750–1850. Macmillan, New York

Kealey L (1986) Patterns of punishment: Massachusetts in the eighteenth century. The American Journal of Legal History 30(2):163–186

Kunzel R (2008) Criminal intimacy: prison and the uneven history of modern American sexuality. University of Chicago Press, Chicago

Lewis W (1965) From Newgate to Dannemora: the rise of the penitentiary in New York, 1796–1848. Cornell University Press, Ithaca

Linebaugh P (2003) The London hanged: crime and civil society in the eighteenth century, 2nd edn. Verso, London

Masur L (1989) Rites of execution: capital punishment and the transformation of American culture, 1776–1865. Oxford University Press, New York

McLennan R (2008) The crisis of imprisonment: protest, politics, and the making of the American penal state, 1776–1941. Cambridge University Press, New York, 2008

Meranze M (1996) Laboratories of virtue: punishment, revolution, and authority in Philadelphia, 1760–1835. North Carolina Press, Chapel Hill

Morgan K (1985) The organization of the convict trade to Maryland: Stevenson, Randolph and Cheston, 1768–1775. The William and Mary Quarterly 42(2):201–227

Morris N, Rothman D (eds) (1995) The Oxford history of the prison: the practice of punishment in Western Society. Oxford University Press, New York

Perkinson R (2010) Texas tough: the rise of America's prison empire. Metropolitan Books, New York

Pisciotta A (1994) Benevolent repression: social control and the American reformatory-prison movement. NYU Press, New York

Rafter N (1990) Partial justice: women, prisons and social control, 2nd edn. Transaction Publishers, New Brunswick

Rothman D (1971) The discovery of the asylum: social order and disorder in the New Republic. Little, Brown & Co, Boston

Rothman D (1980) Conscience and convenience: the asylum and its alternatives in progressive America. Little, Brown and Company, Boston

Salvatore R, Aguirre C (eds) (1996) The birth of the penitentiary in Latin America. University of Texas Press, Austin

Schlossman S (2005) Transforming Juvenile justice: reform ideals and institutional realities, 1825–1920. Northern Illinois University Press, DeKalb

Spierenburg P (1987) From Amsterdam to Auburn an explanation for the rise of the prison in seventeenth-century Holland and nineteenth-century America. Journal of Social History 20(3):439–461

Sullivan L (1990) The prison reform movement: forlorn hope. Twayne Publishers, Boston

Teeters N, Shearer J (1957) The prison at Philadelphia, Cherry Hill. Columbia University Press, New York

History of Crime

► Historical Criminology

History of Criminal Investigation

Steven G. Brandl
Department of Criminal Justice, University of Wisconsin-Milwaukee, Milwaukee, WI, USA

Overview

During the course of history, police organizations have responded to a variety of external forces that have caused changes in how organizations perform the investigative function and the methods used to investigate criminal occurrences. This entry describes the evolution of the criminal investigation function in the United States. The predecessors of modern detectives (informers and parliamentary reward, thief-takers, thief-makers) are discussed, the tactics and technologies of early investigations in the USA are examined, the development and importance of private detectives and state and federal law enforcement agencies are discussed, and changes in the investigative function through the political, reform, and community problem-solving eras of policing are chronicled.

The purpose of this entry is to provide an overview of the history of the criminal investigation function in American law enforcement. An understanding of history, the history of criminal investigations in particular, is important for at least three reasons. First, history provides a context in which to place modern criminal investigation and thus allows for an appreciation of how much or how little things have changed over time. Second, as the adage goes, those who do not remember the past are condemned to repeat it. To move forward, we must understand from where we have come. Accordingly, an understanding of history may provide insight into previous attempts and new methods of solving persistent problems. Third, if history is cyclical, if it repeats itself, then we may be able to predict the future by knowing the past (see Brandl 2003). It is with these understandings that attention is turned to the history of criminal investigations.

The Evolution of the Investigative Task: English Developments

American policing and criminal investigation has its roots in English policing systems. As English society became more complex, and as traditional mechanisms (e.g., family, community) for controlling individuals' behaviors began to lose their effectiveness over time, there was a need for an increasingly sophisticated system of social control. Policing was first an individual responsibility, then a responsibility of the group. With the Frankpledge Model of 1100 AD, all males older than the age of 12 years were designated as responsible for the protection of the group. Then, the Parish Constable system of the 1200s required a constable to organize and supervise watchmen, who would protect members of the town or village. Variations of the watchmen arrangement lasted until formal police departments were created in the early 1800s. It was only then that the modern detective was born.

Informers and Parliamentary Reward

A system of parliamentary reward operated during the 1700s and early 1800s in England. With this system, a reward was offered by the government to anyone who brought criminals to justice or provided information that led to the apprehension of criminals; the more serious the crime, the larger the reward. Although this system may sound like a historical equivalent of a modern-day tip line, there were major differences, one of which was the legal context in which parliamentary reward operated. During the time of parliamentary reward, more than 200 offenses were punishable by death. These offenses included theft, vagrancy, forgery, and even cutting down a tree without permission. The methods of execution included hanging, burning, and drawing and quartering. Many referred to the laws of the time as the *bloody code*. Most people, however, did not support the legal system nor did they believe the legal code was just. As a result, victims were often unlikely to pursue charges, witnesses often refused to testify, and juries were often not willing to convict. There was public sympathy for petty criminals who faced the possibility of execution (Klockars 1985). The problem was that by benefiting from providing information that would lead to the apprehension of petty criminals, informers were viewed with the same contempt as the legal system. Informers were not the answer; they were part of the problem.

Thief-Takers

In the early 1800s, *thief-takers* came to be. A thief-taker was a private citizen who was hired by a victim to recover stolen property or to apprehend the thief. The fee that the thief-taker charged was most often based on the value of the property recovered, and thief-taker only received compensation when the property was returned. The most desirable crimes to investigate, therefore, were ones that involved considerable property and ones for which the property was easily recovered. As such, thief-takers were not likely to spend much time on crimes for which the property was not likely to be recovered or on thefts that involved small amounts of property (Klockars 1985). In essence, the thief-takers most often worked on behalf of the rich, not the poor. In addition to this "selective attention" problem, there was an even more serious problem with thief-takers: Thief-takers often worked in cooperation with thieves. Some thief-takers even employed thieves (Klockars 1985). No question, the thief-taker arrangement was extraordinarily corrupt. The thief would steal from the victim, the victim would hire a thief-taker, the thief would sell the property to the thief-taker, and the thief-taker would then "sell" the property back to the victim. Everyone prospered at the victim's expense.

Agent Provocateur and Thief-Makers

Along with thief-takers there were *thief-makers*. A thief-maker was an individual who tricked another person into committing a crime and then would turn that person in for the parliamentary reward. Thief-makers were often thief-takers who resorted to deception, seduction, trickery, and entrapment to apprehend criminals and receive the monetary rewards (Klockars 1985). These people essentially created criminals for their personal benefit. Not surprisingly, the methods used by these individuals were frequently viewed by citizens as outrageous and unacceptable.

London Metropolitan Police Department

With the 1800s came the Industrial Revolution and the dramatic and rapid increase of the population of cities, where people lived to be in close proximity to where they worked. Factory production was the basis of the new economy. With Industrial Revolution also came an increase in wealth among some people and poverty among others. "Urban" problems were born – sanitation and health issues, ethnic conflict, and crime. With all the changes came political pressure for the government to institute a more formal, more sophisticated, and more effective system of property protection. It was in this context that in 1829, the London Metropolitan Police Department was established.

Introduced early in the London Metropolitan Police Department was the concept of the "plain clothes police officer" – a detective to some, a police spy to others. In designing the job of detective, tremendous public resistance had to be overcome. The resistance was caused, in large part, because of the problems associated with earlier investigative arrangements – the problems associated with parliamentary reward, thief-takers, and thief-makers. To overcome these obstacles, and to allow detectives to be accepted by the public, certain features were incorporated into the design of the detective position (Klockars 1985).

First, to address the problems of parliamentary reward (when petty criminals faced unjust punishment because of the actions of informers), detectives were linked to the crime of murder. There was no public sympathy for murderers. The "architects" of the detective position capitalized on stories of murder and offered "detectives" as a way to combat this horrible crime. In addition, detectives were to play a dual role. Not only were they to help bring punishment to the worst of criminals, they were also supposed to save the innocent from the worst of punishments (Klockars 1985). One can see very clearly the direct association between detectives and murder in early detective fiction (e.g., Poe's *Murder in the Rue Morgue*, Doyle's *A Study in Scarlett*), and this likely helped sell the idea of the police detective to a skeptical public.

Second, to address the problems associated with the thief-taker arrangement, the most significant of which was that thief-takers only worked on behalf of the rich, detectives were to be given a salary (Klockars 1985). If detectives were given a salary, it was argued, they could work on behalf of the rich and the poor alike. Ideally, they could investigate crimes for which the property loss was small. In addition, given the profitability of working on crimes for a fee in a private arrangement, detectives were to be paid more than police officers.

Third, to address the problems associated with thief-makers, particularly the practice of thief-makers tricking people into committing crimes for the thief-maker's benefit, detectives were made "reactive" and were assigned "cases" (Klockars 1985). Only after crimes occurred were detectives given the responsibility for investigating them. As a result, there was limited opportunity for thief-maker trickery. Detectives were to be evaluated in terms of their success in solving crimes. As a result, detectives were given more control over how they were to spend their working time and more discretion in determining how to investigate the cases they were assigned. These features – being responsible for the most serious of crimes, receiving a salary, and being reactive – eventually neutralized public resentment toward detectives and paved the way for their incorporation into police operations.

The Evolution of the Investigative Task: American Developments

At the time of the ratification of the US Constitution, there were few federal laws and, accordingly, the policing function was almost exclusively a responsibility of local government. Policing was quite informal and consisted most often of volunteers assigned to the "watch" who would guard the village or town at night and later during the day. Local control of the police function was a desirable feature of American policing because, ideally, it allowed residents (and politicians) to influence more easily how policing was conducted in their community. The desire for local control also helped explain why the creators of the Constitution were resistant to the idea of an all-powerful national police force.

The First American Police Departments and Detectives

It was not until the mid-1800s that formal municipal police departments were created. These institutions were primarily located in the large and rapidly growing cities of the eastern United States (e.g., Boston, Philadelphia, New York). The Industrial Revolution created similar problems in America as in England. Of particular significance were the violent labor protests and the rioting that stemmed from clashes between immigrants and native-born Americans (Conti 1977). Municipal police detectives, those with primary responsibility for criminal identification, did not appear until later in the 1800s, and this development occurred largely in response to public concern about the increasing amount of crime. For example, in most years of the early to mid-1800s, there were no homicides recorded in Suffolk County (Boston), Massachusetts. Between 1860 and 1869, however, 70 homicides occurred. During the 1870s, 107 homicides were reported (Lane 1967). Detectives in America had much in common with those in England – they were responsible for the most serious of crimes (at least in image), they were to receive a salary, and they were primarily reactive.

The mid-1800s to the early 1900s has been characterized as the political era of policing (Kelling and Moore 1988). As the name suggests, policing at the time was all about politics; the police organization was an arm of the political machine. Politicians, particularly mayors and ward politicians, controlled virtually every aspect of policing including who got hired, what work officers performed, and who got fired. Besides political connections, there were few selection standards. There were tremendous opportunities for corruption. Police supervisors were few, and their influence over beat officers was minimal because there was no mechanism to provide for supervision. There was little ability for citizens to summon the police when needed because there was no means of communication. Officers patrolled on foot. The police made few arrests. According to Lane (1967), more than half of all arrests made at this time were for public drunkenness. This was an offense that beat cops could easily discover and no investigation was necessary. The police simply did not have the capability to respond to and investigate crimes. When an arrest was made, it was usually as a last resort. Making an arrest in the late 1800s usually involved a lot of work; officers would literally have to “run ‘em in” to the police station, or, when arresting a drunk, the officer would use a wheelbarrow and wheel him into the station (Haller 1976). “Curbside justice” with a hickory baton was often seen as an easier and more effective alternative by officers.

The political era of policing did not provide a large role for police detectives. Like the beat cops, detectives had limited capabilities in responding to investigating crimes. During the late 1800s, Boston’s politicians actually disbanded the police department’s detective bureau because their contributions were viewed as minimal (Lane 1967). Although important qualities for beat cops were size and fighting ability, the most important quality for detectives was a familiarity with criminals and their tactics. Many detectives were selected from the ranks of prison guards, and some were reformed criminals (Lane 1967). With this specialized knowledge, detectives received more pay than beat cops. Detectives also received extra compensation through “witness fees,” compensation for providing testimony in court. Detective work was often a clandestine activity. Detectives were sometimes considered to be members of a secret service (Kuykendall 1986). Detectives depended heavily on criminals for information to solve crimes. Detectives often worked in an “undercover” capacity to collect this information. Detectives never wore uniforms. Rather, they often wore disguises, even in court, to protect their identities. Sometimes detectives submitted their court testimony in writing so as not to reveal their identity (Kuykendall 1986).

It was at about this time that identification systems began to be developed and applied to criminal investigations. The first technology that was used for this was photography. By 1858, the New York City Police Department had on file photographs of known criminals. This collection of photographs was known as a rogues gallery (Dilworth 1977). Although photographs were commonly used in "wanted posters" and assisted in the apprehension of criminals, photographs were extremely limited in their usefulness because criminals could alter their appearance either deliberately or simply over time. Of course, to be useful, authorities first needed to have a photograph of the wanted person.

The most famous identification system of the time was the one developed by Alphonse Bertillon, a French criminologist who lived from 1853 to 1914. His system was known as Bertillonage, and it was considered a major improvement over the use of photographs. The premise of the system was that the bone structure of an adult did not change over the course of a lifetime. Bertillon identified 11 measurements (e.g., length and width of the head, length of the left foot, the length of the left middle and little fingers), and these measurements, it was suggested, could be used to identify people and to differentiate one person from another (Muller 1889). Bertillon estimated that the probability of two persons having the same 11 measurements was more than four million to one (Rhodes 1968). Instruments and instructions were developed by Bertillon to make the measurement process as precise as possible. In addition, an elaborate filing system was developed to organize the classification of individuals from whom measurements were taken. Because it was difficult for the police to take measurements of criminals on the street, Bertillon also developed a scaled-down version of his system. Although the technique enjoyed initial success in confirming the identity of suspected and known criminals and was used by police departments in many countries, by the early 1900s, the deficiencies of the system were obvious. It was simply too cumbersome, error prone, and limited in its applicability to be viable as an identification strategy.

Along with the use of these identification methods, detectives at the time also used various "investigative" tactics to deal with crime and criminals. One common strategy was the dragnet roundup of suspects. When informed of a crime, the police would find and arrest all suspicious persons and would keep these people in custody until it could be determined that they did not commit the crime. In essence, the police would often resort to "rounding up the usual suspects."

Another commonly used investigative tactic at the time was the "third degree," which was also known as shellacking, messaging, breaking the news, or giving him the word (Lavine 1930). The origin of the expression "the third degree" is not clear, although some have speculated that the first degree was the arrest, the second degree was being transported to the police station, and the third degree was the interrogation (Kuykendall 1986; Skolnick and Fyfe 1993). Common methods of administering the third degree included beatings with a rubber hose (Haller 1976), placing a suspect in a sweat box for hours or days under constant questioning (Kuykendall 1986), drilling teeth, burning with lit cigars or cigarettes, and beating with blackjacks or batons (Lavine 1930). Many accounts suggest that the use of the third degree to obtain confessions was commonplace into the 1930s and beyond (Kuykendall 1986). However, it is important to note that in 1936, the US Supreme Court ruled in Brown v. Mississippi that prolonged beatings used to extract confessions were no longer a legally acceptable police practice. After the physical abuse of suspects abated, psychological coercion persisted, and in many situations, this coercion has been approved by the courts.

Sheriff, State Police, US Marshal, and the Bureau of Investigation

While police departments were being developed in the major cities in the eastern portion of the country, other areas were most likely to be served by sheriffs and marshals. In the western portion of the country, US marshals were often the sole police power (Ball 1978). Marshals often employed deputies who also served as sheriffs, deputy sheriffs, or constables. With the appearance of automobile and in the face of corrupt and ineffective municipal police agencies and sheriffs' departments, state police agencies were created to assist. In 1905, Pennsylvania created the first state police agency. It was designed to provide a police presence throughout the state, to assist the local police, and to provide police services in less populated, rural areas of the state (Conti 1977).

Also of significance at this time was the development of the Bureau of Investigation, to be later known as the Federal Bureau of Investigation. President Theodore Roosevelt initially asked Congress to create a federal detective force in 1907. Congress opposed President Roosevelt's idea on the official grounds of the long-cited public disdain for an all-powerful federal law enforcement agency. However, unofficially, it was significant that in 1906 two members of Congress had been prosecuted for fraud, the investigation of which was conducted by the Justice Department using agents from the Department of the Treasury. As a result, many members of Congress were concerned about giving the executive branch of government more investigative power (power perhaps to initiate more investigations of congressmen). Along with denying President Roosevelt's request, Congress passed legislation that prohibited the Justice Department from using investigators from other agencies. Not to be stopped by Congress, in 1908 President Roosevelt created a Bureau of Investigation by executive order and directed Attorney General Charles Bonaparte to develop the agency within the Department of Justice. Twenty permanent and eighteen temporary investigators were hired (Murray 1955).

With the turmoil surrounding its creation, it is not surprising that during the first years of its operation (1910s), the Bureau of Investigation was entrenched in scandal. However, at the same time, it was slowly becoming accepted as a law enforcement agency and was assigned law enforcement responsibilities. For example, in 1910 Congress passed the Mann Act, which prohibited the transportation of women across state lines for immoral purposes. Responsibility for the enforcement of the law was given to the Bureau of Investigation. Other statutes followed, prohibiting the transportation of stolen goods, vehicles, and obscene materials (Murray 1955).

In 1916, with 300 agents, and in the face of war in Europe, the Bureau was given power to conduct counterintelligence and antiradical investigations. In 1919, the country experienced a series of bombings with targets ranging from police departments to banks (and the residence of Attorney General A. Mitchell Palmer). These actions were believed to be the responsibility of communists and others who were "un-American." The bombings and their aftermath became known as the Red Scare. In response to the bombings, Attorney General Palmer established the General Intelligence Division (GID) within the Justice Department to increase significantly the ability to store information on radicals and those suspected of being sympathetic to radicals. John Edgar Hoover was named the head of GID. In 1920, using information from the GID, Attorney General Palmer authorized a series of raids (to be known as the Palmer raids) in cities across the country that resulted in thousands of arrests of people believed to be un-American or communists. The problem was that most of the people arrested were not radicals at all. The courts ordered many of those arrested to be released (Murray 1955).

Private Detectives

In the mid-1800s and early 1900s, private detectives also played an important role in criminal investigations. In addition, many corporations (e.g., railroads, iron and coal mines) hired their own police forces for the primary purpose of dealing with their labor strikes (Conti 1977). With regard to private detective agencies, the most prominent was Pinkerton's agency. Allen Pinkerton was a barrel maker turned Chicago deputy sheriff, turned Chicago detective. In 1850, he quit his job in the Chicago Police Department and established his own private detective agency. At first, most of the work of the agency involved protecting several midwestern railroads and railroad bridges from sabotage of the confederates as well as striking laborers. Pinkerton and his associates' preferred method of operation was to mingle with known rebels and criminals in taverns, hotels, and brothels to learn of their plans. In 1861, Pinkerton learned of a plan to assassinate the president-elect Lincoln and was able to persuade Lincoln to alter his travel plans to disrupt the plot. Pinkerton was also hired to spy on the confederacy, to collect information on their strengths and weaknesses, and to apprehend enemy spies. Also at this time, the Justice Department, having no investigators of its own, used agents from the Pinkerton agency. Pinkerton had investigative and operational advantages over governmental agents; in particular, the agency operated without concern for cumbersome political jurisdictional lines. This capability made Pinkerton ideal for pursuing mobile criminal such as train robbers.

Pinkerton also had a well-developed system of internal communication, record keeping, and files on criminals. Police departments often relied on this information to learn what criminals were working in their area. By the turn of the century, the agency had a system in place to share information with the investigative services of foreign nations: Pinkerton operated on a global scale (Conti 1977).

The Reform Era

With the problems of the political era policing system well noted, there were efforts made to reform the police, namely, to get the police out from the control of politicians. To do so required a new way of thinking about policing. This effort took the form of police professionalism. Forward-thinking police leaders such as O.W. Wilson and August Vollmer advocated a new philosophy and system of policing. This new way of thinking about policing was in direct reaction to the ugly politics of before. According to Kelling and Moore (1988), the system of policing from the early 1900s to the 1960s was known as the reform era.

The reform era was all about police professionalism and anti-politics. The police presented themselves as experts who had the specialized knowledge and capabilities to control crime. It was argued that if the police were able to distance themselves from citizens and politicians (i.e., professional autonomy), they would be the most efficient and effective. Crime control and criminal apprehension were viewed as the primary functions of the police. The new technology of the time contributed to and supported the ideals of the new way of thinking about policing. Specifically, the police patrol car allowed the police to increase their mobility and to respond quickly to crime scenes, random motorized patrol was viewed as a means of deterring criminals, and the patrol car provided separation between police officers and citizens. The two-way radio allowed police supervisors to be in constant communication with, and to have supervision, over beat officers. The two-way radio also allowed patrol officers to be directed to places where they were needed. The telephone turned police departments into 24 hour-a-day agencies that were just a phone call away. With the telephone, citizens could easily summon the police when needed.

During the reform era, detectives became an important tool in police departments' efforts to enhance their professionalism and to deal with crime. Detectives were the ultimate professionals. They were well paid and highly trained. The media at the time portrayed detectives as efficient and effective crime solvers. Similar to the police style in general, detectives often went about their work in a professional, aloof manner. *Dragnet*, a popular television show during the 1960s, captured this style well. The show was about two Los Angeles Police Department detectives and the investigations they conducted. They cut through the emotion of their work and became famous for their line, "Just the facts ma'am."

As a continuing attempt to provide organizational control over officers and detectives, detective work became much more removed from interactions with criminals. With scientific advances, more emphasis was placed on getting information from science (and from victims and witnesses) as opposed to from criminals. The rise of science was led in large part by the FBI. Through the 1920s and 1930s, several initiatives were embarked on by the Bureau, each of which helped solidify its reputation as the top law enforcement agency in the country. First, the

Bureau took the lead in the development of fingerprints as a method of criminal identification. In 1924, the Bureau began a campaign to collect fingerprints from every American. Representatives of the Bureau even went door-to-door in an effort to collect prints (FBI 1990). Fingerprints were recognized to have numerous advantages over the earlier identification (e.g., photography, Berrillonage). It was understood that fingerprints did not change and could not be changed, from birth to death. In addition, fingerprints were unique. Furthermore, there was a uniform method of collection and classification, and this methodology was relatively simple. Fingerprint printing and storage systems were relatively inexpensive. Finally, fingerprints were useful not only for identification but also for evidence, as they could be left by a suspect at a crime scene. Fingerprints could provide a clue regarding the individual who committed the crime. Photography and Berrillonage provided no such advantages. The major problem with fingerprints as a method of criminal identification was that matching a print of an unknown suspect to a potentially vast number of prints on file was a daunting and unproductive task. It was not until the development of Automated Fingerprint Identification Systems (AFIS) in the 1980s that this limitation was addressed.

The second initiative of the Bureau was the development of a crime laboratory. In 1932, with a borrowed microscope and a few other pieces of equipment, the Bureau's laboratory opened. During its first year of operation, the laboratory conducted 963 examinations – nearly all of which involved examinations of handwriting in extortion cases (FBI 1990).

Finally, in 1935 the Bureau began operation of a National Police Academy (to be known later as the FBI National Academy) to train select local police officers in investigative methods. Selection for and graduation from the National Academy was, and continues to be, a prestigious law enforcement accomplishment.

With these initiatives of the FBI, the agency took the lead in reform and crime fighting. It became the role model for other law enforcement agencies. Calvin Coolidge was elected president in 1923 and, in the aftermath of the Palmer raids, one of his first tasks was to reform the Justice Department, the Bureau of Investigation in particular. In 1924, J. Edgar Hoover was appointed the director of the Bureau of Investigation. In 1924 the Bureau had 441 agents. One of Hoover's first tasks was to upgrade the personnel standards of the FBI agents. Hoover fired about one quarter of all agents and mandated additional training for others. Hiring standards were raised, and training in law or accounting was required. A training school was established for various skills and for learning the procedures of the Bureau. According to Hoover, promotion was to be based on performance, not seniority. Control and standardization were the themes that reflected his management style.

During the late 1920s through the 1930s, numerous high-profile crimes and criminals cook center stage with J. Edgar Hoover and his G-men. In particular, gangsters became larger than life, capturing the imagination of millions of Americans. With the assistance of Hollywood and detective fiction writers, he was able to portray his agents (and himself) at the same mythical levels as gangsters. Gangsters were portrayed as public enemies, the Bureau of Investigation's G-men were cop-heroes, and J. Edgar Hoover was the top cop.

In 1932, the nation was transfixed over the news of the kidnapping and murder of the infant son of Charles and Anne Lindbergh. Kidnapping laws were passed by Congress, and the Bureau of Investigation was made responsible for this and other high-profile investigations.

World War II brought new challenges. With the rise of totalitarian abroad, a concern with internal enemies developed. By request of President Franklin D. Roosevelt in 1936, the FBI resumed with increased enthusiasm their information collection activities of domestic communists and radicals. By1939, Hoover had reestablished the GID. The war years provided the FBI with a powerful rationale for monitoring political radicals. In addition, the passage of the Smith Act in 1940 provided a legal basis for FBI domestic security investigations. The Smith Act made it a crime to advocate or conspire to advocate the forceful overthrow of the government. During the early 1940s, the FBI underwent dramatic growth, with the number of agents doubling to nearly 6,000 (FBI 1990).

The 1950s saw a decreasing concern with domestic communism and an increasing concern with organized crime; new laws strengthened the Bureau's jurisdiction in organized crime cases. Aggressive use of wiretaps continued unabated (FBI 1990).

The Community Problem-Solving Era

The 1960s was a troubling time for many Americans and the police. In the 1960s, America was in the grip of the Vietnam War. There were war protests across the country. It was the time of the civil rights movement and the related demonstrations, marches, and riots. The police became viewed by many as an "occupying army" in the low-income, minority ghettos of urban cities. The police were called "pigs." In the1960s, President John F. Kennedy was assassinated, as was Senator and presidential candidate (and former Attorney General of the United States) Robert Kennedy and civil rights leader Martin Luther King, Jr. American society was in turmoil. Fear of crime was increasing dramatically. Actual crime was also increasing; the crime rate doubled from 1960 to 1970. Helter Skelter and Charles Manson were front-page news. The Beatles sang *Revolution*. The police were experiencing a crisis, yet they were supposed to have the knowledge and capabilities to control crime successfully. If the situation was not bad enough for the police, the US Supreme Court rendered several landmark decisions (e.g., Mapp v. Ohio, Miranda v. Arizona) that were seen as "handcuffing" the police. In addition, in the late 1960s and early 1970s, several major research studies were conducted to examine the effectiveness of police operations. The Kansas Preventive Patrol Experiment (Kelling et al. 1974) concluded that random motorized patrols did not deter crime. The RAND study on detectives (Greenwoord et al. 1977) concluded that detectives contributed little to solving crimes and that many detectives could be replaced with clerical personnel.

In the face of this multifaceted crisis, the police realized that the old ideas of professionalism no longer worked. The police needed to get closer to the community to enlist their support and assistance in fighting crime. With this new realization came the community problem-solving era of policing (Kelling and Moore 1988), for which community policing is of central importance. A cornerstone of community policing is that police and citizens must be "coproducers" of crime prevention. The reform era emphasized police-citizen separation, the community era emphasized police-citizen cooperation.

Skolnick and Bayley (1988) identified four definitional elements of community policing. First and most important for the investigative function, community policing involves the community in efforts of criminal investigation and crime prevention. Second, community policing consists of a reorientation of patrol activities. It is realized that the patrol car is not the only means by which police services can be delivered. Community policing initiatives allow for the placement of police personnel in mini-stations, provide for home visits, or provide for foot patrols (as well as patrol via bicycle, scooter, horse, roller blades, and so forth). Third, community policing requires increased accountability through citizen input into police practices and policy. Citizen satisfaction is a legitimate goal according to the community policing philosophy. Finally, with community policing, the police understand that different communities have different priorities and problems. Accordingly, community policing requires the decentralization of command and provision of services to individual neighborhoods.

All these features are supposed to result in several desired ends. First, such initiatives are supposed to reduce crime. Because community policing programs allow the police to have more interaction with the public, there are likely to be increased opportunities for the police to receive information that leads to crimes being solved (Skolnick and Bayley 1988). Crime may also be reduced through citizen education and involvement in crime prevention programs. Crime may be reduced by paying attention to the conditions, the disorder, which may lead to an environment in which crime can flourish (Wilson and Kelling 1982). Second, community policing programs are often expected to reduce fear of crime. This goal may also be attained through citizen education and involvement in crime prevention activities, as well as through the increased emphasis placed on reducing signs of disorder. Finally, because community policing programs enable citizens to have input into policing in their neighborhoods and make the police "easily accessible, frequently visible, and caring in their relationships with citizens" (Goldstein 1987, pp. 8–9), community policing programs are often expected to improve citizen attitudes toward the police.

Although the research that has examined the effects of community policing has shown that it often does not work as expected (see Rosenbaum 1994), the means and ends of community policing, in principle, seem to be quite congruent with the task of criminal investigation. The basic task of the police in a criminal investigation is to collect information that will lead to the identification, apprehension, and conviction of the subject who perpetuated the criminal act. Much of the research on the investigative function highlights the role of the public as suppliers of information to the police (Greenwood et al. 1977). Simply stated, the police are dependent on the public, and community policing makes this dependence explicit.

Strategies that provide an opportunity for community residents to share information with the police in solving crimes are particularly relevant in the era of community policing. For example, tip lines are quite common in criminal investigations today. The Crime Solvers, Crime Stoppers, and WeTip programs are supported by many communities across the country. In addition, police departments frequently establish designated tip lines for major investigations.

Along the same line, school liaison officers are located in a setting where they are available not only to assist students with questions or problems that they may have but also to obtain information about crimes from them. Similarly, police involvement with community watch groups provides a public service and also, arguably, makes it easier for residents to contact and provide information to the police that may assist in investigations. These strategies make police dependence on the public explicit and are congruent with the ideals of community policing.

Along with an explicit dependence on community residents for information, other developments in criminal investigation have occurred during the community problem-solving era of policing. Chief among these is DNA analysis as a method of identification. DNA analysis represents an extraordinary advance in science and in identification methods as applied to criminal investigations. The science of DNA, along with the introduction of computer technology to store, record, and match DNA prints across individuals, has the potential to revolutionize criminal investigative methods. In addition, other technology in the form of computer networks and databanks are also changing criminal investigations in dramatic ways. COMSTAT, an operational approach to policing which is based on "gathering accurate and timely intelligence, designing effective strategies and tactics, the rapid deployment of personnel and resources, and relentless follow-up and assessment" (Dabney 2010, p. 34), also has the potential to dramatically affect how criminal investigations are managed and performed. In addition, empirical research continues to be conducted on the criminal investigation process, the contribution of detectives in solving crimes, and the impact of forensic evidence on crime solving (e.g., Eck 1983; Brandl and Frank 1994; Baskin and Sommers 2010).

As before, the FBI is often considered at the forefront of technological changes in the criminal investigation process. Today, the FBI crime laboratory is the most scientifically advanced and well-funded in the world, although it has been subject to criticism for its "sloppy" work in several cases (Kelly and Wearne 1998). The FBI also operates the NCIC – a computerized network and storage system of crime information. The FBI continues to operate the National Academy and it provides many other types of operational assistance to federal, state, and local law enforcement agencies (see FBI 1999), including psychological profiling.

During the course of history, police institutions and organizations have responded to a variety of external forces that have caused changes in their structure and function. From these changes emerged the criminal investigation function and investigative methods. Most people would argue that much progress has been made in criminal investigations. Ultimately, that is for the future to decide.

Related Entries

- ▶ COMPSTAT
- ▶ Crime Mapping
- ▶ Criminal Investigative Analysis
- ▶ DNA Profiling
- ▶ DNA Technology and Police Investigations
- ▶ FBI Influence on State and Local Police
- ▶ Fingerprint Identification
- ▶ History of Technology in Policing

Recommended Reading and References

Ball LD (1978) The United States Marshals of New Mexico and Arizona Territories. University of New Mexico Press, Albuquerque, pp 1846–1912

Baskin D, Sommers I (2010) The influence of forensic evidence on the case outcome of homicide incidents. J Crim Justice 38:1141–1149

Brandl SG (2003) Back to the future: the implications of September 11, 2001 on law enforcement practice and policy. Ohio State J Crim Law 1:133–154

Brandl SG, Frank J (1994) The relationship between evidence, detective effort, and the disposition of burglary and robbery investigations. Am J Police 13:149–168

Conti PM (1977) The Pennsylvania State Police: a history of service to the commonwealth, 1905 to present. Stackpole, Harrisburg

Dabney D (2010) Observations regarding key operational realities in a compstat model of policing. Justice Q 27:28–51

Dilworth DC (1977) Identification wanted: development of the American criminal identification system 1893–1943. International Association of Chiefs of Police, Gaithersburg

Eck JE (1983) Solving crimes: the investigation of burglary and robbery. Police Executive Research Forum, Washington, DC

Federal Bureau of Investigation (1990) FBI: facts and history. U.S. Department of Justice, Washington, DC

Federal Bureau of Investigation (1999) Your FBI. U.S. Department of Justice, Washington, DC

Goldstein H (1987) Toward community oriented policing: potential, basic requirements, and threshold questions. Crime Delinq 33:6–30

Greenwood PW, Chaiken JM, Petersilia J (1977) The criminal investigation process. D.C. Heath, Lexington

Haller MH (1976) Historical roots of police behavior: Chicago, 1890–1925. Law Soc Rev 10:303–323

Kelling GL, Moore MH (1988) The evolving strategy of policing, vol 4, Perspectives on policing. National Institute of Justice, Washington, DC, pp 1–15

Kelling GL, Pate T, Diekman D, Brown CE (1974) The Kansas City preventive patrol experiment. Police Foundation, Washington, DC

Kelly JF, Wearne P (1998) Tainting evidence: inside the scandals at the FBI crime lab. Free Press, New York

Klockars CG (1985) The idea of the police. Sage, Newbury Park

Kuykendall J (1986) The municipal police detective: an historical analysis. Criminology 24:175–200

Lane R (1967) Policing the city: Boston 1822–1885. Harvard University Press, Cambridge

Lavine E (1930) The third degree: a detailed and appalling expose of police brutality. Garden City, New York

Muller G (1889) Alphonse Bertillon's instructions for taking descriptions for the identification of criminals. American Bertillon Prison Bureau, Chicago

Murray RK (1955) Red scare. McGraw-Hill, New York

Rhodes HTF (1968) Alphonse Bertillon: father of scientific detection. Greenwood Press, New York

Rosenbaum D (1994) The challenge of community policing: testing the promise. Sage, Thousand Oaks

Skolnick JH, Bayley DH (1988) Theme and variation in community policing. In: Tonry M, Morris N (eds) Crime and justice. University of Chicago Press, Chicago

Skolnick JH, Fyfe JJ (1993) Above the law: police and the excessive use of force. Free Press, New York

Wilson JQ, Kelling GL (1982) Broken windows: the police and neighborhood safety. Atl Mon 249:29–38

History of Criminological Theories: Causes of Crime

Gerben Bruinsma
Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, The Netherlands
VU University Amsterdam, Amsterdam, The Netherlands

Synonyms

Early Chicago School Theory; Historical criminology; History of Geographical Criminology

Overview

For thousands of years, mankind has developed and applied vague or specific thoughts about what behavior should be labeled as criminal, who the criminals are, and what the causes of crime may be. Philosophers in the eighteenth and early nineteenth century put forward the rational criminal who is free to make choices and criticized the cruel and unjust criminal justice system. The scientific study of crime started in the beginning of nineteenth century with mathematicians searching for empirical regularities between countries and areas in administrative governmental data. Sociologists, psychologists, and physicians introduced societal and individual causes of crime in a comprehensive way. Later, new developments in other disciplines triggered new thoughts on the causes of crime, notably strain, control, and cultural theories. These have been dominant for about many decades, although critical theories emerged in the 1960s and 1970s of the twentieth century. Recent developments of theoretical criminology point at including disciplines that have not been very involved in the field of crime like neurosciences and econometrics so far.

The First Rudimentary Ideas

For thousands of years, mankind has developed and applied vague or specific thoughts about what behavior should be labeled as criminal, who the criminals are, and what the causes of crime may be. In the heydays of the Egyptian, Chinese, Greek, Inca, or Roman empires, criminal laws were introduced to criminalize certain behaviors and how to deal with criminals (Kerner 1998). These written (in stone, on papyrus, walls, or clay tablets) or oral laws (UK) were introduced to protect the social cohesion of the empire, the existence of primary groups, or the power of those who were in charge. Till that time conflicts were settled by private parties; the state was not involved.

In the Middle Ages, the search for the causes of crime in Europe was more than ever regulated by religious motives: those who committed crimes "must" offend God's will and "therefore" must be possessed by the devil. Draconic tortures were applied to release the man or the woman from that evil. Until that time scientific knowledge as we have nowadays did not exist. Monks copied unique written texts about religion or other affairs that were stored in monasteries. Spiritual explanations were at that time the dominant thought about the causes of crime. Desiderius Erasmus (1466 (?)–1536) was one of the first philosophers who wrote about the "natural" causes of crime. He related the occurrence of criminal behavior not to the "spirit" of the individual but to poverty in general (although without any empirical evidence). The physician G. Battista della Porta (1535–1615) was the first who systematically was searching for observable causes of crime. He opened the brains of hanged criminals in order to study irregularities in their shape, color, and structure. In the age of Enlightenment, new vague thoughts about crime were developed, of which the rudiments can still be found in criminological theory of today. During Enlightenment intellectuals promoted the idea of ratio and philosophy and in general the upcoming scientific work. They challenged widespread superstition and abuses of rights of residents by those in power and by the church. Combined with the upcoming Protestantism, the individual became more and more central in the new philosophies (as they called science those days). This was an important period in the study of crime, because many of our contemporary fundamental assumptions about the causes of crime originated in the seventeenth, eighteenth, or nineteenth century (although the words may be different and specifications may have been very well elaborated during the last 300 years).

The Upcoming Scientific Search for Causes of Crime

What happened? In the seventeenth and eighteenth centuries, science in general begins to expand and to make progress in explaining many (natural) phenomena. New instruments made it possible to explore in more detail the human body, several illnesses, and nature in general. The fast growth of mathematics laid the foundation of many new natural sciences like physics, chemistry, medicines, biology, anthropometry, and physiognomy. Social philosophers stood near the cradle of the political and social sciences and of early psychology. The cruel and unfair criminal justice system of their time inspired many philosophers to suggest improvements of criminal law. Cesare Beccaria (1738–1794) and Jeremy Bentham (1748–1832) focused on the issue of the criminal justice process and how to deter individuals from crime (Marsh 2004). They came up with the precursor of rational choice paradigm in which individuals are assumed to assess the benefits and costs of criminal behavior. If the benefits are higher than the costs, they will commit a crime. It was assumed that people have a free will to make decisions to commit crimes. That was revolutionary in a period when spiritual and religious thoughts dominated the thinking of people. It took about 200 years of silence until an economist and Nobel Prize winner Gary Becker again suggested rational choice decision making as an explanatory model of why individuals commit crimes (Becker 1968). Today, the rational choice perspective is popular in criminology, especially in geographic criminology. Routine activity theory and crime pattern theory that explain why crimes cluster at certain places are, for instance, based on the rational choice perspective as well as the basic principles of crime prevention (Clarke and Felson 1993; Cornish and Clarke 1986; Felson 2008).

Long before the name of criminology was introduced by Raffaele Garofalo in his 1885 published *Criminologia*, researchers empirically studied crime under the label of moral statistics or criminal science. The foundation of empirical criminology was laid in the beginning of that century when Belgian, French, and British mathematicians and statisticians like Guerry (1832), the astronomer and first PhD of mathematics of Ghent University in Belgium Quételet (1984 [1831]), Ducpétiaux (1827), and Rawson (1839) and Fletcher (1848) of the Royal Statistical Society of London tried to uncover regularities in the social and geographic distributions of crime and criminals between counties, areas, cities, and countries (Beirne 1987; Weisburd et al. 2009). The publication of the *Comptes Générales de l'Administration de la Justice Criminelle en France* in 1825/1826 inspired them to explore official data on crime in more detail (Weisburd et al. 2009, p. 6). These scholars were the first who studied crime as a social phenomenon that is unequally distributed over society (Guerry 1832, 1833, 1864). Quételet (1847) discovered the normal distribution in statistics in order to assess the "normality" of empirical regularities he demonstrated in research. These scholars were inspired by the reform movement in their countries of which fighting illiteracy and illness among the poor was the priority political issue. That is the reason why they studied the empirical relationships between illiteracy and wealth and personal and property crimes in France, England, and the low countries (now the Netherlands and Belgium). They did not systematize their thoughts into structured theorizing. Instead, they made up inventories of all kinds of correlates of possible causes of crime and criminals (good examples of that kind of inventories are Afschaffenberg (1906) and Parmelee (1918/1919)).

Later in the nineteenth century, Charles Darwin formulated his evolution paradigm on the origins of species (and like many original thinkers encountered conflicts with religious powers of his time). Darwin's revolutionary thoughts completely changed the human perspective on the roots of mankind. The physician Cesare Lombroso (1835–1909) was fully impressed by Darwin's theory. He elaborated Darwin's innovative way of thinking and applied it to the explanation of crime (Lombroso 2006 [1878]). As physician he studied sentenced criminals as if they suffered from a drawback in the evolutionary process, believing that "those" people can be recognized by outer stigmata. Scholars in his time had an optimistic and naive view on science (positivism) assuming that scientific knowledge could change the world in a better place to live in. Lombroso later studied in his career many other topics as well (mostly social aspects of crime), among others the spatial variations in the distribution of crime across areas, provinces, and countries (Lombroso 2006 [1878]).

A landmark in the study of crime causation was the work of the sociologist Emile Durkheim (1858–1917). He built on the analyses of social philosophers and the first sociologists who observed in the nineteenth century macro-social processes in the speedy developing societies in Europe (Bierstedt 1978). Saint-Simon (1760–1825) can be viewed as the first who envisaged the new industrial society, and together with the younger Auguste Comte (1798–1857), he analyzed the disruptive consequences of the many European political revolutions. The abrupt changes in political structure and society that they observed led to a bigger gap between the state and the families and individuals when intermediate institutions disappear. Also Georg Simmel (1858–1918) was very critical at the new technical, urbanized materialistic culture that emerged in Europe. Ferdinand Tönnies (1855–1936) focused more on the social changes in society. In his view, societies transformed from *Gemeinschaft* into *Gesellschaft*, leaving behind the small-scale agricultural villages with strong social cohesion and informal social control. In modern societies of a *Gesellschaft* type, trade and industry took the lead; laborers recruited from rural areas and ties between people were weakened. People in Europe became more individualized, and sociologists feared the negative consequences for society and for crime rates. Durkheim built on their thoughts and focused on the relationships between people and the impact of societal changes on them (Durkheim 1951 [1897], 1964 [1893]). He argued that relationships between people are based on cohesion, division of labor, religion, (scientific) knowledge, and morality, and when societies become more heterogeneous, their collective conscience is diminished. Anomic societies lose their power to regulate social relations between people, and increased levels of crime are a consequence. People with less ties feel free to break the moral rules more easily. Central is the idea that external forces from society (measured by structural features) explain why people commit crime. Durkheim demonstrated empirically that in societies with lower levels of social integration, people tend to commit more suicides and more crimes.

Further Developments in Early Twentieth Century

Durkheim had a profound influence on the emergence of two important theoretical perspectives in criminology: the *anomie* paradigm and *control* paradigm (Cullen and Messner 2007). Growing up in the relative deprived neighborhoods of South Philadelphia as a son of a Ukraine immigrant family, the sociologist Robert Merton (1910–2003) experienced personally at young age that the American society offered unequal opportunities to different social groups (Coser 1977; Cullen and Messner 2007). He published at the age of 27 his famous article on anomie and social structure in which he stated that the American society promoted central material ends (American dream) to pursuit (cultural) goals to all citizens, but certain social groups – especially the lower classes – encounter structural closures to reach these ends in a legally prescribed way (Merton 1938). The discrepancy between culturally defined goals and legal opportunities create tensions or strains in society that people try to reduce by applying adaptive behavior patterns. Merton's anomie theory aims to explain why aggregates (social groups) withdraw from society, conform, become rebellious, ritualists, or innovative. According to Merton, innovative people try to reach cultural accepted goals by committing crimes. Among others, Merton inspired many criminologists to study the role of subcultural values (Albert Cohen (1918–)) and poverty on the overrepresentation of the lower classes in the official crime statistics. Richard Cloward (1928–2001) and Lloyd Ohlin (1918–) focused on the differential opportunities of people having to use legal or illegal means to reach material goals and developed a typology of kinds of subcultures by merging differential association theory (see later) and anomie theory. Robert

Agnew (1953–) elaborated Merton's anomie theory in an individualistic way for his strain theory (Agnew 1992, 2006). In that theory he stated that three kinds of strain (negative relationships and stressful life events when other people prevent a person from achieving valued goals, take away something valued, or impose on the person that is unwanted) may lead to criminal behavior as a kind of response pattern. Steve Messner and Richard Rosenfeld elaborated the idea of American dream as cultural goal that leads to more competition between citizens (Messner and Rosenfeld 1997). And more competition between citizens in the context of a relatively dominant economic institution vis-à-vis family, education, and other social institutions leads to more crime in societies, an idea that has also been raised earlier at the beginning of twentieth century (Bonger 1969 [1905 and 1916]).

Durkheim on the other side inspired other influential criminologists to focus on the effects of social control on crime. Travis Hirschi (1935–) is the best known protagonist of this perspective. In line with Durkheim's reasoning, Hirschi stated in *Causes of Delinquency* that if an individual is not tied to society, he is free to deviate from the norms, e.g., free to commit crimes (Hirschi 2006 [1969]). That view was opposite to the mainstream and dominating cultural deviance theories of his time. Hirschi was especially critical of the differential association theory, arguing that individuals need a certain level of commitment to society and social ties with other people to refrain them from crime. The family (especially parents) plays a central role in establishing ties.

Hirschi was certainly not the first to introduce the concept of social control as a condition for crime. A number of predecessors can be mentioned. Edward Ross (1866–1951), one of the most influential sociologists of the United States, was the mastermind of the concept of social control. Around the turn of the twentieth century, he pointed at internal and external kinds of social control that regulate criminal behavior (Ross 1918 [1901]). Walter Reckless (1899–1988) focused later on social control as exercised by parents on their children: parental supervision (Reckless 1940, 1961). When parents do not succeed in supervising their spouses, the children will develop weak self-control and weak other internal controls. If they succeed, kids can resist the worst neighborhood influences that normally lead to more crime (containment). Albert Reiss Jr. (1922–2006) specified kinds of social control into personal and social control that both refrain youngsters from criminal behavior (Reiss 1951); Jackson Toby introduced the concept of "stakes in conformity" referring at how much a person has to lose when he breaks the law (Toby 1957), later followed by F. Ivan Nye who distinguished four dimensions of social control (direct social control, internalized social control, indirect social control, and access to legal means (Nye 1958)). All of their work was integrated by Hirschi and specified in testable propositions.

The idea of social control has been expanded in several new directions. Travis Hirschi, collaborating with Michael Gottfredson, proposed a general theory of crime in which was stated that human's low self-control was the key explanatory mechanism in the explanation of crime (Gottfredson and Hirschi 1990). Low self-control is developed by psychological deficits or by parents in unorganized families by "inconsistent" parenting styles on their offspring. Blumstein and his colleagues (1988) started more empirical research on criminal careers, putting aside static views on offenders. The idea of a small group of active offenders was revealed by Wolfgang and his colleagues in their Philadelphia study (Wolfgang et al. 1972). Later research found out that offenders are influenced by the number and nature of risk factors they encounter during their life course (Farrington 1988, 2002; Loeber and Wikström 1993). This life-course perspective has been developed by several scholars. Dan Nagin developed a method to analyze empirically various trajectories of offending (Nagin 2005). John Laub and Robert Sampson (Laub and Sampson 2003; Sampson and Laub 1993) initiated the developmental criminology and explored the idea of varying social controls during the life course that may be responsible for the variations in crime when people coming of age. They based their analysis on the data of two matched groups of 500 delinquents and 500 nondelinquents collected by Sheldon and Eleanor Glueck for their pioneering empirical study on juvenile crime in the 1950s (Glueck and Glueck 1950). Rolf Loeber in Pittsburgh (Loeber 1990), Marc LeBlanc in Montreal (Leblanc 1997), and Farrington with his longitudinal study of male laborers in London (Farrington 2002, 2003) have brought up new dynamic views on the causes of crime. Terry Moffitt's distinction between adolescent-limited and persistent criminals also inspired many present studies on criminal careers, recidivism, and redemption (Moffitt 1993). Although her distinction between two groups of offenders is probably too simple, she inspired may researchers to study criminal careers with a variety of explanations for distinct career trajectories (Cullen 2011). In the 90s of the last century, new promising horizons emerge from the idea that perhaps neurological defectives play a role as cause of crime (Rafter 2008; Raine 1993). In the future, when scientists know more about DNA and the human gnomes, new kinds of causal theories will be developed and tested.

The City of Chicago and Its Scholars

The city of Chicago played a special role in criminology. It is probably the most researched city in the world. From 1850 till 1930, the city grows from 4,470 to almost four million residents from all continents longing for a new and hopefully prosperous future. Various scholars in sociology, economy, psychology, and criminology carried out innovative theoretical and empirical studies on the structural changes of their city in the newly established University of Chicago. Inspired by former journalist Robert Park (1863–1944), pragmatic academics like Ernest Burgess (1886–1966), Clifford Shaw (1896–1957), and methodologist Henry McKay (1899–1980) started quantitative and qualitative geographic criminological studies on the impact of urbanization processes on human behavior and institutions (without being aware till 1933 that in Europe there already existed a long tradition of this kind of research) and on the spatial distribution of crime in cities. A very important contribution stems from William Thomas (1863–1947) who introduced the concept of social disorganization in criminology (based on his ideas on social change (Thomas 1966: p. 3)). Later Louis Wirth (1897–1952) elaborated that concept (Thomas 1966; Thomas and Znaniecki 1918/1920; Wirth 1938). George Mead (1863–1931) laid the foundation for his symbolic interactionism, inspiring social control theory and labeling theory (Mead 1974 [1934]). Thorsten Sellin (1896–1994) developed his culture conflict theory of crime (Sellin 1938), and Edwin Sutherland (1883–1947) elaborated in different versions his differential association theory (Cohen 1968; Sutherland 1924, 1934, 1939, 1947), while Burrhus Skinner (1904–1990) created his operant version of behaviorism (Ronald Akers based his social learning theory partly on his thoughts) (Skinner 1976). The concept of social disorganization is still in the center of many etiological and geographic studies in criminology (Kubrin and Weitzer 2003; Sampson 2010). Robert Bursik worked in the 1980s in the Chicago tradition and elaborated that concept to study the effects of *changing* neighborhoods on crime concentrations in that city (Bursik 1984, 1986, 1988). With colleague Harold Grasmick (1993), Bursik also highlighted how different types of community controls (private, parochial, public) were related to neighborhood crime, disorder, and citizen's fear of crime. Robert Sampson (1995) and his colleagues continued that line of investigations by putting forward the idea of collective efficacy as one of the explanatory characteristics of neighborhoods (Morenoff and Sampson 1997; Sampson 1987, 1995, 2010, Sampson and Groves 1989; Sampson et al. 1997). The Chicago school was and still is not a united framework of thought but an umbrella under which various brilliant scholars have been developing divergent theories, especially on the causes of crime. Various scholars of all over the world are studying the spatial clustering of crime in their cities. New specialized developments are to be observed in the field of crime pattern theory

(Brantingham and Brantingham 1991), of crime location decision models, or of the mobility of offenders, based on ideas that have been developed many decades earlier. One of the most popular theories on spatial crime distributions is the routine activity theory (L. Cohen and Felson 1979; Felson 2008) that has influenced spatial and environmental criminology for the last 30 years now. New methodologies and better administrative data and surveys to include measurement of social processes (Raudenbush and Sampson 1999) allow to study smaller units of analysis, elevating clusters of crime and victimization, even in neighborhoods where normally there is a low level of crime (Eck and Weisburd 1995; Oberwittler and Wikström 2009; Sherman and et al. 1989; Weisburd et al. 2009). A new criminology is emerging to address the issue of unequal crime places in cities, not large neighborhoods anymore.

Symbolic interactionism as developed by Mead was important not only for sociology but also for criminology in two respects. Edwin Sutherland (1883–1947) built his famous differential association theory in the 1930s and 1940s on the theoretical roots of symbolic interactionism (Sutherland 1947). Criminal behavior is learned in a social process within social groups by an excess of positive definitions to law violation above negative definitions towards law violation. He also combined the macro perspective of Chicago sociology on urbanization, social disorganization, and social groups with the micro theory of Mead on how people learn attitudes and behavior in intimate groups. Sutherland not only studied the crimes of young people in this respect but applied his theory also on homeless people as well as professional thieves and managers of powerful corporations (Sutherland 1937, 1983 [1949]). Sutherland had a great number of students who dedicated their academic life in exploring, defending, and extending the differential association theory (Cressey 1964, 1969; Glaser 1972). The differential association theory inspired many empirical studies on the impact of peers on the criminal behavior of adolescents.

After the Second World War: The Rise of New Criminological Theories

In the roaring 1960s of previous century, a young generation opposed the authority of states and the power of social and legal institutions on several ways, leading to various nontraditional criminological theories: labeling theories, critical criminology (of all different schools of thought), and feminist criminology (the same). Howard Becker (1928–) put forward the labeling perspective and focused on the role of moral entrepreneurs in society to criminalize people and behavior (Becker 1963). As proclaimed by symbolic interactionism, people can change their identity because of the reactions of other people on their behavior. Becker studied the culture of so-called deviant groups in society and how law enforcement pushes them to deviate more often and seriously from the norms of conventional society. Becker's legacy is also partly built on the studies of others. Frank Tannenbaum (1893–1969), sent to prison for a protest demonstration when he was a young man, focused attention on the negative consequences of the criminal justice system, especially the prison system on individuals (Tannenbaum 1938). Erving Goffman (1922–1982) became famous in psychiatry and in criminology for his study on total institutions, like prisons, (mental) hospitals, or army camps, that are in full control of those who were forced to live in it (Goffman 1961). Edward Lemert (1912–1996) distinguished between primary and secondary deviance (broader than crime) and demonstrated that official reactions force people to change their identity into a deviant one (Lemert 1951, 1967). Stigmatizing criminals or mentally ill people forces them to continue their deviant behavior or as Becker called it, the state (especially the police) amplifies deviancy in society. Inspired by the roaring 1960s and the 1970s, more critical criminologists focused the attention to all kinds of sources of conflict in society that promote crime. Richard Quinney (1970) argued that crime emerges in a politicized society when those who are in power control the lower class by criminal law; William Chambliss stressed the power of interest groups to influence the legal system and studied state crimes instead of crime of those without power (Chambliss 1969; Chambliss and Seidmann 1986), and Turk elaborated the conflict perspective by proposing a theory of criminalization that attempted to explain that societal conflicts between authorities and individuals will result in criminalizing the individual (having less power) (Turk 1969). Critical criminology became more influential when Walton, Taylor, and Young published their criticism on traditional theoretical thinking in criminology, including labeling and conflict traditions (Taylor et al. 1973), later followed by (Taylor et al. 1975). The emergence of feminist criminology was related with the second-wave emancipation and liberation of women in the western world in the 1960s and 1970s. Criminology was (and mostly still is) a male-dominated scientific discipline focusing on male problematic behavior. Klein was one of first critics of male criminological theories (Klein 1973), stating that traditional theories largely explained the criminal behavior of men and the few theories that explain criminal behavior of female offenders used simplistic and stereotypical assumptions about women. Smart also raised serious questions about the assumptions of many criminological theories on the nature of women in general and on the female offender in particular as law abiding by nature (Smart 1976). Feminists also emphasized gender differences in treatment by the criminal justice system as offenders and as victims. Often female offenders are punished more harshly than males and often psychiatrically by the justice system "because their criminal behavior is against their law-abiding nature," and women who are victimized are not "true" victims but rather seduce the "innocent" male offender. Freda Adler's *Sisters in Crime* argued that women, once liberated, will play a role in the male world and consequently (have to) act more aggressive and competitive like men (Adler 1975). Because of the progressing emancipation of women, they will have higher crime rates in the future. By 1980 feminist criminology could be classified into different schools of thought: Marxist, socialist, liberal, and radical schools (Jagger and Rothenberg Struhl 1978). However, these classifications were soon replaced with uniquely feminist approaches (e.g., doing gender; sexed bodies, intersectionality (Daly and Maher 1998; Simpson 1989)).

Integrated and Elaborated Theories

John Braithwaite (1951–) started from an anomaly between labeling theory and other theories like strain and control theories predicting that the criminal justice system is aiming at reducing crime and at the same time stimulates crime (Braithwaite 1989). Braithwaite introduced the idea that the outcome of justice processing depends on the kind of shame that is exploited by the system. Disintegrative shaming leads to more recidivism and integrating shaming reduces future criminal behavior; reintegrative shaming will reduce crime in the future because offenders will stay integrated in society. David Garland (1955–) extended the sociology of punishment by opening the discussion between punishment and control in an area of expanding culture of control in which the media, politics, and moral entrepreneurs play their part (Garland 2001; Garland and Sparks 1995). His inspiration comes from classic philosophers and sociologists as Durkheim. By creating fear and anger about their safety among the citizens, control can expand far beyond the limits of civil rights in the contemporary world we are living (Garland 2001).

The Current State of Theoretical Criminology

One of the conclusions of this brief overview is that during the last centuries, a great number of causal ideas have been developed in criminology to address the explanation of criminal behavior or crime rates. Many of these ideas are still presented and discussed in full length in the textbooks of our discipline to teach new generations of students in criminology. In most cases, after the presentation of the theories, criticism is presented as well, leaving the freshmen in doubt whether there might be one theory in criminology that meets the criteria of the discipline. From a distance the discipline looks like a battlefield of masses of rival and conflicting ideas about the causes of crime. Is it social class or mental deficits that cause crime? Is it the (inherited or socialized) propensity of crime or is the social and physical environment responsible for humans to commit crimes? Or are the laws of states the cause of crime?

Some may argue that this "wealth of ideas" is a blessing for the discipline; others would argue that its presence demonstrates the impotence of theoretical criminology. There is no superior theory (despite the claims of individual scholars) that can explain more and better than other theories, let alone can predict that a person will commit a crime in the future given the conditions of that theory. In the past, criminologists and others have also heavily and regularly criticized the state of causal theory. In the 1930s of last century, the mathematicians Michael and Adler (1933) critically evaluated criminology for the government and concluded that criminological theories are no more than empirical generalizations that do not meet any mathematical criteria of theory formation. Forty years later a German sociologist discussed the most important criminological theories of that time and judged them as vague, with unclear concepts and unspecified relations between them, without any formal structure and low explanatory power (Opp 1974). Wolfgang and his colleagues drew similar conclusions after evaluating criminology between the years 1945 and 1972 and observing empirical progress but hardly any theoretical improvements (Wolfgang et al. 1978). Bernard (1990) repeated earlier critics and questioned whether criminological theories have any explanatory power. He stated, "*Despite the enormous volume of quantitative research, there has been no substantial progress in falsifying the criminology theories that existed 20 years ago*" and that "*None of these theories has been falsified in the past 20 years*" (Bernard 1990, p. 147). Finally he concluded: "*Current controversies in criminology indicate that there is fundamental disagreement on even the most basic facts about crime*" and "*That being the case, the failure to accumulate verified knowledge in the context of specific theories should not be surprising. Even on simple correlations like those between age and crime, social class and crime there is no agreement*" (Bernard 1990, p. 329).

What might be going on with the discipline? Before condemning theoretical criminology, we must first try to disentangle the existing causal theories whether they are rivals or not to assess the state of a discipline (Lakatos 1970). However, that is a very difficult task to accomplish for criminology. *First*, we have to acknowledge that not all criminological theories can be compared because they are about different units of analysis. A majority of causal theories have individuals as subject of analysis, but an increasing number of others are about geographical areas, like street segments, neighborhoods, or countries, or social and ethnic groups, business corporations, states, organized groups, institutions like prisons or probation systems, police, or laws. In order to compare whether theories are rivals, they should at least have the same unit of analysis, or else one is comparing apples with pears. *Second*, in case of theories about individuals, the dependent variables (i.e., the behavior to be explained) differ too much prohibiting comparisons. Most theories are about criminal behavior in general, others are about a specific kind of crime like burglary or violence, some are about deviant behavior including mental disturbances or sexual or psychological abnormalities, some include all aggressive behaviors, and some even do not mention the kind of behavior that is aimed to be explained! *Third*, criminology is a product of intellectual influences of many other disciplines, notably sociology and psychology but also from medical sciences and law. As a consequence conceptual frameworks of these disciplines are introduced in criminology, without further reference to other, already existing theoretical concepts. The complex interdisciplinary discourse must be unraveled to assess whether we have distinct independent causal factors or not. *Fourth*, theories of which the propensity to crime is the object of explanation cannot be compared with theories in which criminal behavior is the outcome. For instance, in Hirschi's bonding theory, the outcome of the bonding process is that the individual is free to deviate, not the frequency or seriousness of the criminal offense. *Lastly*, most causal theories explain only little variance in the criminal behavior of humans as Weisburd and Piquero recently demonstrated (Weisburd and Piquero 2008). They pointed at the underdeveloped way of theory testing in criminology and the practice to use seldom primary data for that purpose. But that is only part of the story. Probably the fact that we have to do with only fragmented and limited explanations anyway is important as well.

Acknowledgement The author wants to thank professor Sally Simpson (Maryland) and professor Lieven Pauwels (Ghent) for their helpful comments on a previous draft.

Related Entries

- ▶ Differential Association Theory
- ▶ History of Geographic Criminology Part I: Nineteenth Century

Recommended Reading and References

Adler F (1975) Sisters in crime. McGraw-Hill, New York

Afschaffenberg G (1906) *Das Verbrechen und seine Bekämpfung. Einleitung in die Kriminalpsychologie für Mediziner, Juristen und Soziologen; Ein Beitrag zur Reform der Strafgesetzgebung*, 2 edn. Carl Winter's Universitätsbuchhandlung, Heidelberg

Agnew R (1992) Foundation for a general strain theory of crime and delinquency. Criminology 30:47–87

Agnew R (2006) Pressured into crime. An overview of general strain theory. Roxbury, Los Angeles

Becker GS (1968) Crime and punishment: an economic approach. J Polit Econ 76:169–217

Becker H (1963) Outsiders. Studies in the sociology of deviance. The Free Press, New York

Beirne P (1987) Adolphe Quételet and the origins of positivist criminology. Am J Sociol 92(5):1140–1169

Bernard TJ (1990) Twenty years of testing theories: what have we learned and why. J Res Crime Delinq 27(4):325–347

Bierstedt R (1978) Sociological thought in the eighteenth century. In: Bottomore T, Nisbet R (eds) A history of sociological analysis. Heineman, London, pp 3–38

Blumstein A, Cohen J, Farrington DP (1988) Longitudinal and criminal career research: further clarifications. Criminology 26(1):57–74

Bonger WA (1969 [1905 and 1916]) Criminality and economic conditions (abridged edition ed.). Indiana University Press, Bloomington

Braithwaite J (1989) Crime, shame and reintegration. Cambridge University Press, Cambridge

Brantingham PJ, Brantingham PL (eds) (1991) Environmental criminology. Waveland, Prospect Heights

Bursik RJ (1984) Urban dynamics and ecological studies of delinquency. Soc Forces 63:393–413

Bursik RJ (1986) Ecological stability and the dynamics of delinquency. In: Reiss AJ Jr, Tonry M (eds) Communities and crime. Crime and justice. A review of research, vol 8. University of Chicago Press, Chicago, pp 35–66

Bursik RJ (1988) Social disorganization and theories of crime and delinquency: problems and prospects. Criminology 26(4):519–552

Bursik RJ, Grasmick HG (1993) Methods of studying community change in the rate and pattern of crime. In: Farrington DP, Sampson RJ, Wikström P-OH (eds) Integrating individual and ecological aspects of crime. National Council for Crime Prevention, Stockholm, pp 241–258

Chambliss WJ (ed) (1969) Crime and the legal process. McGraw-Hill, New York

Chambliss WJ, Seidmann R (1986) Law, order, and power. Addison-Wesley, Reading

Clarke RV, Felson M (eds) (1993) Routine activity and rational choice. Advances in criminological theory, vol 5. Transaction Books, New Brunswick

Cohen AK (ed) (1968) International encyclopaedia of the social sciences, vol XV. Macmillan, New York

Cohen L, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44(4):588–608

Cornish DB, Clarke RV (eds) (1986) The reasoning criminal. Rational choice perspectives on offending. Springer, New York

Coser LA (1977) Masters of sociological thought. Ideas in historical and social context, 2nd edn. Harcourt Brace Jovanovich, New York

Cressey DR (1964) Delinquency, crime and differential association. Martinus Nijhoff, Den Haag

Cressey DR (1969) Epidemiology and individual conduct. In: Cressey DR (ed) Delinquency, crime and social process. Harper & Row, New York, pp 557–577

Cullen FT (2011) Beyond adolescence-limited criminology: choosing our future – The American Society of Criminology 201 Sutherland Address. Criminology 49(2):287–330

Cullen FT, Messner SF (2007) The making of criminology revisited. An oral history of Merton's anomie paradigm. Theor Criminol 11(1):5–37

Daly K, Maher L (1998) Criminology at the crossroads: Feminist readings in crime and justice. Oxford University Press, New York

Ducpétiaux E (1827) De la Justice de prévoyance, et particulierement de l'influence de la misere et de l'aisance, de l'ignorance et de l'instruction sur le nombre des crimes. J.J. de Cautaerts, Bruxelles

Durkheim E (1951[1897]) Suicide: a study in sociology (trans: Simpson G, Spaulding JA). The Free Press, New York

Durkheim E (1964[1893]) The division of labor in society. Free Press, New York

Eck JE, Weisburd D (eds) (1995) Crime and place, vol 4. Criminal Justice Press, New York

Farrington DP (1988) Studying changes within individuals: the causes of offending. In: Rutter M (ed) Studies of psychosocial risk. Cambridge University Press, Cambridge, pp 158–183

Farrington DP (2003) Developmental and life-course criminology: key theoretical and empirical issues – the 2002 Sutherland Award Address. Criminology 41(2):221–255

Farrington DP (ed) (2002) Human development and criminal careers. Clarendon, Oxford

Felson M (2008) Routine activity approach. In: Wortley R, Mazarolle L (eds) Environmental criminology and crime analysis. Willan, Devon, pp 70–77

Fletcher J (1848) Moral and educational statistics of England and Wales. J Stat Soc Lond 11:344–366

Garland D (2001) The culture of control: Crime and social order in contemporary society. Oxford University Press, Oxford

Garland D, Sparks R (eds) (1995) Punishment and modern society: A study in social theory. Clarendon, Oxford

Glaser D (1972) The differential association theory of crime. In: Rose AM (ed) Human behavior and social processes. An interactionist perspective. Routledge/ Kegan & Paul, London, pp 425–442

Glueck S, Glueck ET (1950) Unraveling juvenile delinquency. Harvard University Press, Cambridge

Goffman E (1961) Asylums: Essays on the social situation of mental patients and other inmates. Penguin, Harmondsworth

Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Guerry A-M (1832) La Statistique Comparée de l'Etat de l'Instruction et du Nombre des Crimes. Revue Encyclopédique 55:414–424

Guerry A-M (1833) Essai sur la Statistique Morale de la France. Crochard, Paris

Guerry A-M (1864) Statistique Morale de l'Angleterre Comparée avec la Statistique Morale de la France. Bailliére, Paris

Hirschi T (2006 [1969]) Causes of delinquency, 5th edn. Transaction, New Brunswick/London

Jagger AM, Rothenberg Struhl P (1978) Feminist frameworks: Alternative theoretical accounts of the relations between women and men. McGraw-Hill Book Company, New York

Kerner H-J (1998) The global growth of criminology. Annales Internationales de Criminologie 36(1–2): 27–42

Klein D (1973) The etiology of female crime: a review of the literature. Issues Criminol 8:3–30

Kubrin CE, Weitzer R (2003) New directions in social disorganization theory. J Res Crime Delinq 40(4):374–402

Lakatos I (1970) Falsification and the methodology of scientific research programmes. In: Lakatos I, Musgrave A (eds) Criticism and the growth of knowledge. Cambridge University Press, Cambridge, pp 91–196

Laub JH, Sampson RJ (2003) Shared beginnings, Divergent lives. Delinquent boys to age 70. Harvard University Press, Cambridge

Leblanc M (1997) A generic control theory of the criminal phenomena: the structural and dynamic statements of an integrative multilayered control theory. In: Thornberry TP (ed) Developmental theories of crime and delinquency. Transaction, New Brunswick, pp 215–286

Lemert EM (1951) Social pathology. McGrawHill, New York

Lemert EM (1967) Human deviance, social problems and social control. Prentice Hall, Englewood Cliffs

Loeber R (1990) Development and risk factors of juvenile antisocial behavior and delinquency. Clin Psychol Rev 10:1–41

Loeber R, Wikström P-OH (1993) Individual pathways to crime in different types of neighborhoods. In: Farrington DP, Sampson RJ, Wikström P-OH (eds) Integrating individual and ecological aspects of crime. Allmänna förlaget, Stockholm, pp 169–204

Lombroso C (2006 [1878]) Criminal man (translated and with a new introduction by Gibson M, Han Rafter N). Duke University Press, Durham & London

Marsh I (2004) Criminal justice an introduction to philosophies theories and practice. Routledge, London/ New York

Mead GH (1974 [1934]) Mind, self & society from a standpoint of a social behaviourist, Vol 1 (ed. and with an introduction by Morris CW (19 edn). The University of Chicago Press, Chicago/London

Merton RK (1938) Social structure and anomie. Am Sociol Rev 3(5):672–682

Messner S, Rosenfeld R (1997) Crime and the American dream, 2nd edn. Wadsworth, Belmond

Michael J, Adler MJ (1933) Crime, law and social science. Harcourt Brace, New York

Moffitt TE (1993) Adolescence-limited and life-course persistent antisocial behavior: a developmental taxonomy. Psychol Rev 100(4):674–701

Morenoff JD, Sampson RJ (1997) Violent crime and the spatial dynamics of neighborhood transition: Chicago, 1970–1990. Soc Forces 76:31–64

Nagin DS (2005) Group-based modeling of development over the life course. Harvard University Press, Cambridge

Nye FI (1958) Family relationships and delinquent behavior. Wiley, New York

Oberwittler D, Wikström P-OH (2009) Why small is better: advancing the study of the role of behavioral contexts in crime causation. In: Weisburd D, Bernasco W, Bruinsma GJN (eds) Putting crime in its place: Units of analysis in geographic criminology. Springer, New York, pp 35–59

Opp K-D (1974) Abweichendes Verhalten und Gesellschaftsstruktur. Luchterhand, Neuwied

Parmelee M (1918/1919) Criminology. MacMillan, New York

Quételet A (1847) Statistique morale de l'influence du libre arbiter de l'homme sur les faits sociaux, et particulierement sur le nombre des marriages. Bulletin de la Commission Centrale de Statistique III:135–155

Quételet A (1984 [1831]) Research on the propensity for crime at different ages (translated and with an introduction by Sylvester SF). Anderson Publishing, Cincinnati

Quinney R (1970) The social reality of crime. Little Brown, Boston

Rafter NH (2008) The criminal brain: Understanding biological theories of crime. New York University Press, New York

Raine A (1993) The psychopathology of crime: Criminal behavior as a clinical disorder. Academic, San Diego

Raudenbush SW, Sampson RJ (1999) Ecometrics: towards a science of assessing ecological settings, with appliance to the systematic social observations of neighbourhoods. Sociol Methodol 29:1–41

Rawson RW (1839) An inquiry into the statistics of crime in England and Wales. J Stat Soc Lond 2:316–344

Reckless WC (1940) Criminal behavior. McGraw-Hill, New York/London

Reckless WC (1961) The crime problem, 3rd edn. Appleton, New York

Reiss AJ Jr (1951) Delinquency as the failure of personal and social controls. Am Sociol Rev 16(2):196–207

Ross EA (1918 [1901]) Social order: a survey of the foundations of order. Macmillan, New York/London

Sampson RJ (1987) Communities and crime. In: Gottfredson MR, Hirschi T (eds) Positive criminology. Sage, Newbury Park, pp 91–114

Sampson RJ (1995) The community. In: Wilson JQ, Peterselia J (eds) Crime. Twenty-eight experts look at the most pressing problems of our time. ICS Press, San Francisco, pp 193–216

Sampson RJ (2010) Collective efficacy theory. In: Cullen FT, Wilcox P (eds) Encyclopedia of criminological theory. Sage, Thousands Oaks, pp 802–812

Sampson RJ, Groves WB (1989) Community structure and crime: Testing social -disorganization theory. Am J Sociol 94(4):774–802

Sampson RJ, Laub JH (1993) Crime in the making: Pathways and turning points through life. Harvard University Press, Cambridge

Sampson RJ, Raudenbush SW, Earls F (1997) Neighbourhoods and violent crime: a multilevel study of collective efficacy. Science 277:918–924

Sellin T (1938) Culture conflict and crime. Social Science Research Council, New York

Sherman LW, Gartin PR, Bueger ME (1989) Hot spots of predatory crime: routine activities and the criminology of place. Criminology 27:27–56

Simpson S (1989) Feminist theory, crime, and justice. Criminology 27(4):605–631

Skinner BF (1976) About behaviorism. Vintage Books, New York

Smart C (1976) Women crime and criminology. A feminist critique. Routledge/Kegan Paul, London

Sutherland EH (1924) Criminology. Lippincott, Philadelphia

Sutherland EH (1934) Principles of criminology. Lippincott, Philadelphia

Sutherland EH (1937) The professional thief. The University of Chicago Press, Chicago

Sutherland EH (1939) Principles of criminology. Lippincott, Philadelphia

Sutherland EH (1947) Principles of criminology. Lippincott, Philadelphia

Sutherland EH (1983[1949]) White collar crime. The uncut version. Yale University Press, New Haven/London

Tannenbaum F (1938) Crime and the community. Ginn, Boston

Taylor I, Walton P, Young J (eds) (1975) Critical criminology. Routledge/Kegan Paul, London/Boston

Taylor I, Walton P, Young Y (1973) The new criminology. For a social theory of deviance. Harper and Row, New York

Thomas WI (1966) On social organization and social personality. Selected papers (edited and with an introduction by Janovitz M). The University of Chicago Press, Chicago/London

Thomas WI, Znaniecki F (1918/1920) The polish peasant in Europe and America. Badger, Boston

Toby J (1957) Social disorganization and stake in conformity: complementary factors in the predatory behavior of hoodlums. J Crim Law, Criminol Police Sci 48:12–17

Turk A (1969) Criminality and legal order. Rand McNally, Chicago

Weisburd D, Bruinsma GJN, Bernasco W (2009) Units of analysis in geographic criminology: historical development, critical issues, and open questions. In: Weisburd D, Bernasco W, Bruinsma GJN (eds) Putting crime in its place: units of analysis in geographic criminology. Springer, New York, pp 3–31

Weisburd D, Piquero AR (2008) How well do criminologists explain crime? Statistical modelling in published studies. In: Tonry M (ed) Crime and justice. A review of research, vol 37. The University of Chicago Press, Chicago, pp 453–502

Wirth L (1938) Urbanism as a way of life. Am J Sociol 44(1):1–24

Wolfgang ME, Figlio RM, Sellin T (1972) Delinquency in a birth cohort. University of Chicago Press, Chicago

Wolfgang ME, Figlio RM, Thornberry TP (1978) Evaluating criminology. Elsevier, New York

History of Fines

Tomer Einat
Bar-Ilan University, Ramat-Gan, Israel

Overview

Criminal fines are monetary penalties paid to the state by offenders (Hillsman 1992). This sanction has several advantages over other types of penalties (Hillsman 1988; Zender 2004: 206). For one, it incorporates the penal functions of retribution, deterrence, and rehabilitation (Ashworth 2000). In addition, the amount of the fine can be tailored to the seriousness of the particular crime and to the specific individual's criminal history and resources (Uglow 1995). Furthermore, it allows the offender to remain in the community, maintain economic and social bonds (Radzinowicz and Hood 1996), and avoid secondary effects of incarceration, such as loss of employment and forcing dependents to rely on public assistance (Ruback and Bergstorm 2006). Moreover, the criminal fine is inexpensive to administer and can be enforced by existing criminal agencies (Klein 1998). Finally, criminal fines provide income for the state (Einat 2008).

Accordingly, it shouldn't be a surprise that the criminal fine has become practically one of the most important and frequently ordered sanction by many Western and non-Western criminal and civil courts. In some cases, it has even been accorded statutory priority over short imprisonment (Grebing 1982).

Curiously, despite the importance and frequent use of the criminal fine, little theoretical, empirical, and critical attention has been paid by modern criminology and sociolegal disciplines to its nature, underlying rationale, characteristics, and historical development (O'Malley 2009). Nonetheless, due to the relatively high number of historical, sociolegal, and criminological variables affecting the development, usage, enforcement, and impact of the criminal fine (Einat 2004 ; O'Malley 2009; Weisburd et al. 2008) as well as their complexity, the current review will incorporate only a few of them. A comprehensive review and analysis of such variables and aspects should, however, be the subject of a separate work.

Historical Development of the Criminal Fine

Monetary sanctions were known in a basic form to various ***ancient cultures*** that had passed through the stage of self-help and of private and blood vengeance while retaining the underlying principles of retaliation (lex talionis; talion). The purely material talion, namely, revenge through harm of equal severity, was restricted in favor of a discharge of the talion through the payment of money and through the performance of services having a monetary value. Such discharge corresponding with the developing notion of equivalence associated with barter and sale. Thus, monetary sanctions became the most common punishments in the early days of many states. Even if the compensatory character of these payments was to some extent still predominant, there were, however, cases of authentic fines.

For example, according to the criminal law of ***Hammurabi*** (2250 BC), stealing was punished by payment of five times the value of the stolen object. In the case of theft of domestic animals, the sanction amounted to the tenfold value, and in cases of theft of royal property or temple, the penalty extended to thirty times of the value.

This exponentiation principle frequently characterized other, less ancient, criminal law systems: The laws of ***Manu*** in India (1200–500 BC) prescribed numerous fines and also provided for confiscation. Offenders, unable to pay their fines, could have discharged their duties to pay through the performance of services, thus becoming "slaves of their punishment." In ***Egypt***, fines and confiscation were imposed for a variety of offenses, viewable as "breaches of public order," often, to the point of abuse. Under ***Mosaic*** law, monetary sanctions served the purpose of satisfying and compensating the victim. These penalties were either calculated in accordance with the exponentiation principle and in proportion to the harm caused or fixed by law. Following the exponentiation principle, the amount of monetary penalties was two to five times in cases of fraud, breach of trust, or theft. Additionally, offenses against the person could have been sanctioned by monetary sanctions, thus avoiding vengeance. Under ***Greek*** law, monetary sanctions appeared as original punishments and played a significant part in relief from severe punishment. In this manner, maliciousness, disregard of official duties, insult, and rape were punishable per se with monetary payments, either to the state or to the victim. Finally, monetary sanctions were of significant importance in ***Roman*** law: The monetary penalty went as *Multa* into the public pursue and as *Poena* to the injured party. Monetary sanctions imposed against offenses of property were a multiple or quadruple of the value of the harm caused. Capital punishment was accompanied by full or partial confiscation of property, which, being a source of revenue, later served the state well.

Germanic law exerted a special influence on the genesis of the criminal fine in Europe. It gradually restricted the dispute over the right of revenge in favor of compensation through performance of valuable services. Furthermore, an arrangement system was developed with different types of penalty. With the emergence of state administration of criminal justice, monetary sanctions finally lost their compensatory character and took the form of actual fines in most countries.

The Criminal Fine in the Modern Era

The criminal fine has not followed a stable and continual course in its development down to the present day, but has been subject to two major turning points.

The first crossroad came around the thirteenth century, with the development of severe form of physical punishment, which limited the imposition of fines to the domain of petty offenses and breaches of public order. Nonetheless, relief from such harsh sanctions was possible through the payment of fines.

The second decisive moment came around the beginning of the nineteenth century with a considerable growth in the importance of the fine, due to its appreciation by the authorities as a source of revenue, successive decline in the severity of physical punishment, and the fact that prisons were not generalized institutions for punishment (King 1996). Oddly, and in contrast to the legal discourse, these historical processes brought the fine into complete discredit in all Western European countries. The fine was generally regarded as a mediocre penalty, customarily imposed against misdemeanors, breaching of public order, greediness, and petty offenses (Briggs et al. 1996; Seagle 1948). For example, in seventeenth-century England, the criminal fine was frequent and typically light, usually from a few pence to a few shillings. The reason why fines were so small was that they were imposed on ordinary working people, people whose incomes and resources were minimal (Briggs et al. 1996). Between 1760 and 1820, close to 80 % of convictions in England resulted in the imposition of a fine, most frequently of only small amounts of up to a shilling (King 1996). Simultaneously, the proportion of offenders fined in England for assault drop from nearly 80 % to only 26 % (King 1996: 48–50). Corresponding with this, the proportion imprisoned (mainly for short terms) rose from less than 4 % to over 50 % (*Ibid.*). In short, between the thirteenth century and the nineteenth century, criminal fines were a default sanction for minor offenses, and short terms of imprisonment temporarily displaced them from their historically ancient dominancy.

Crucial international and national reconsideration of the criminal fine in most modern Western sanctioning systems occurred at the last third of the nineteenth century, mainly due to the strong criticism razed against the ineffectiveness of short-term imprisonment (Grebing 1982). The short prison term was seen to have neither reformative nor deterrent effects and was criticized for having extremely harmful, stigmatizing, and criminogenic effects, especially on first and occasional offenders. Garofalo (1885/1968), for example, argued that imprisonment for first and minor offenders was pointless because “if the occasion has been an exceptional one, if there is little likelihood of its future occurrence, there is no need of employing any means of elimination" (226). Furthermore, short-term imprisonment was perceived as costly to the family and administratively burdening. This criticism was fortitude by additional claim suggesting that short terms of imprisonment convert casual offenders into confirmed criminals (O'Malley 2009). Leading criminologists at that time – Garofalo, Rosenfeld, Walberg, Von Liszt, and Bonnevile De Marsangy – argued strongly for the abolition or, at the very least, general restriction of short-term imprisonment and for the fine's restitution. Their disapproval of short-term imprisonment and their view that the criminal fine could be used as substitutes for such sanction found ready acceptance in various European and Scandinavian countries. Movements and fights against short-term imprisonment were held in England, Italy, and Austria by leading academic criminal lawyers. In the Netherlands and Belgium, Prins Van Hamel propagated the program of the modern school of criminal law – pushing to the elimination of short-time imprisonment and the extensive use of fine and conditional sentencing. In addition, the problems arising from short imprisonment and the question of its replacement by fines (or other alternatives to imprisonment) appeared repeatedly on the agendas of international congresses. For example, the International Penitentiary Congress, held in London in 1872, in Rome in 1885, and in St. Petersburg in 1890, as well as the meeting of the IKV in Brussels in 1889 and in Oslo in 1891, dealt with the question of avoiding both short imprisonment and imprisonment in default of payment through the imposition of noncustodial forced labor, judicial reprimands, and suspended sentences.

Though the criticism against short terms of imprisonment and the significance of the criminal fine were well understood in principal by academicians, philosophers, lawyers, and several sanctioning systems, fines could not be installed to any degree in all European jurisdictions before the end of the nineteenth century, because of the extent of unemployment and poverty (Rusche and Kirchheimer 1939).

As of the end of the nineteenth century and into the twentieth century, due to increase in employment and real living standards that provided the conditions under which the fine could have become a more generally applied sanction (Rusche and Kirchheimer 1939), many European countries, legally and practically, began restricting the use of short-term imprisonment replacing it by criminal fines and suspended sentences (Grebing 1982). For example, in Germany, between 1882 and 1932, the percentage fined increased steadily from 11 to 47.5 %, having peaked at 50.5 % in 1930 (Rusche and Kirchheimer 1939: 167). Over the same period, the proportions sentenced to terms of over 3 months imprisonment remained stable, indicating that short periods of imprisonment were being substituted by fines (O'Malley 2009: 38). On the basis of rather thin evidence, Rusche and Kirchheimer (1939) concluded that "…In general therefore the application of fines in the first half of the nineteenth century was infrequent because the necessity of commuting the punishment into imprisonment would have unduly complicated criminal procedure…The decline in unemployment and the rising living standard in the second half of the nineteenth century, however, introduced a fundamental change…many of the difficulties lying in the way of a fine system lost their force" (168). Four decades later (1969 and 1975), the Federal Republic of Germany passed the first and second (respectively) criminal law reform acts declaring that short imprisonment under 6 months is to be the last solution and the fine was given statutory priority in this sanctioning area. Similarly, the Austrian and Brazilian legislatures followed this example and passed the rule that the fine must take precedence over a short prison term. Other countries such as Belgium, the Netherlands, Switzerland, Denmark, Norway, Finland, and Sweden pursued the objective of restricting short-term imprisonment – mainly through the imposition of fines and suspended sentences. Nonetheless, these countries have neither enacted a statutory minimum term of imprisonment nor introduced a priority in the selection of penalties in favor of the fine. Consequently, in spite of a noticeably large number of fines imposed, they exhibited (and still exhibit) a vast proportion of short-term imprisonment sentences especially in regard to medium-range crimes.

In contrast to the legislative and sanctioning policies mentioned above, the criminal fine was perceived negatively by several East European jurisdictions during the postrevolutionary years (Grebing 1982). It was regarded as inferior to reformative sanctions and, thus, almost completely neglected from many East European criminal codes. In this manner, the criminal fine was omitted from the leading principles of the 1919 Soviet Union criminal legislation, displaced by short periods of "correctional labor."

During the 1920s, the criminal fine tiptoed back only for a few trivial and financial-oriented offenses. This remained true until the late 1980s (O'Malley 2009) where fines seen as incompatible with some of the basic ideological premises of socialism. It was maintained that the fine was a criminal measure typical of the capitalist system in which everything, even the criminal justice system, was profit oriented. Consequently, fines made up only about 3–4 % of sanctions in the USSR, about 5 % in Czechoslovakia, about 10 % in Bulgaria, and 18 % in Poland (Frankowski and Zelinska 1983: 39).

While in most East European socialistic jurisdictions the fine "was held in considerably bad scent" (O'Malley 2009: 35), it could have been logically presumed that the United States, one of the most capitalist and wealthiest of nations that glorifies market processes, would be devoted to imposing fines for criminal offenses (*Ibid*.). In practice, this assumption appears to be and to have been solely true as regards to punishment against businesses and corporations. Individual offenders were hardly punished by criminal fines, and the few who had were usually fined due to petty financial-oriented crimes. For example, in the 1950s, over 90 % of antitrust violations and offenses against labor standards legislation, and almost the same proportion of offenses against food and drug legislation, were sanctioned by fines at the federal level. Nonetheless, when the focus shifted to assault, the proportion sanctioned by fines dropped dramatically to 10 %; in the case of theft dropped to only 4 %; and with juvenile delinquency dropped to a less than a 1/3 % (The University of Pennsylvania Law Review 1953). Furthermore, a broad repertoire of offenses attracting criminal fines more than half the time in Britain and more than three quarters of the time in Germany attracted sentences of fines in only 5 % of cases in USA at the federal level (Gillespie 1981). Even by the mid-2000s, at the federal US level, only 4 % of all offenders were ordered to pay a criminal fine (only 3.5 % received a sole fine), and this was true for only 1 % of those convicted of felonies. In addition, the proportion fined for misdemeanors was 29 %. In comparison, criminal fines were imposed in the same years against nearly 25 % of all criminal offenses in England and Wales (Home Office 2002) and against 70 % of offenders accused for misdemeanors. Lastly, at the state level, the criminal fine has hardly been used as the sole sentence in relation to felonies but was employed as an additional sanction in about 25 % of all cases – most frequently as an add-on to a sentence of imprisonment (Bureau of Justice Statistics 2004). Some states (e.g., Philadelphia) and federal guidelines explicitly limit the criminal fine to a supplemental role, something to be imposed in addition to the real sentence.

The practice of fines in America, both in federal and in state levels, thus suggests that the criminal fine has been losing ground over the past decade. In contrast to Europe where criminal fines increased in proportion and in money value precisely at the time they came to be regarded as a replacement for short periods of imprisonment, short terms of imprisonment did not decline in the United States in the same way (O'Malley 2009). Fines in the United States were not seen as comparable to short periods of incarceration and, consequently, did not attract significant penalties in dollar value. Hypothetically, this existing pattern stems either from the United States high commitment to the "Tough-on-Crime" punitive agenda and low commitment to reform – leading approximately 75 % of all sentenced offenders at the federal level to imprisonment (Bureau of Justice Statistics 2004) – or from doubts as to the correctional and/or rehabilitative value of the fine (American Bar Association 1971; The National Commission on Reform of Federal Criminal Laws 1971).

Recommended Reading and References

American Bar Association (1971) Standards relating to sentencing alternatives and procedures. American Bar Foundation, Chicago

Ashworth A (2000) Sentencing and Criminal Justice, 3rd edn. Butterworths, London

Briggs JC, Harrison A, Vincent D (1996) Crime and punishment in England. UCL Press, London

Bureau of Justice Statistics (2004) State court sentencing of convicted felons 2004. US Department of Justice, Washington DC

Einat T (2004) Criminal fine enforcement in Israel: administration, policy, evaluation and recommendations. Punishm Soc 6(2):175–194

Einat T (2008) Sentencing rationales, judicial discretion, and the practice of criminal fines in Israel. Journal of Criminal Justice 36:444–452

Frankowski S, Zelinska E (1983) Non-custodial penal measures in European Socialist countries. Int Rev Crim Policy 36:38–46

Garofalo R (1885/1968) Criminology. Patterson Smith, Montclair

Gillespie R (1981) Sentencing traditional crimes with fines: a comparative analysis. Int J Comp Appl Crim Justice 5:195–204

Grebing G (1982) The fine in comparative law: a survey of 21 countries. Institute of Criminology Occasional Paper No. 9, University of Cambridge, Cambridge

Hillsman S (1988) Fines and fine administration. Justice Syst J 13:5–16

Hillsman S (1992) Fines and day-fines. In: Morris M, Tonry M (eds) Crime and justice: a review of research. University of Chicago Press, Chicago, pp 49–98

Home Office (2002) Criminal statistics England and Wales 2002. Home Office, London

King P (1996) Punishing assault: the transformation of attitudes in the English courts. J Interdiscip Hist 27:43–74

Klein A (1998) Alternative sentencing, intermediate sanctions and probation. Anderson Publishing, Cincinnati

National Commission on Reform of Federal Criminal Laws (1971) Final report. National Commission on Reform of Federal Criminal Laws, Washington, DC

O'Malley P (2009) The currency of justice: fines and damages in consumer societies. Routledge-Cavendish, Abington, pp 78–92

Radzinowicz L, Hood R (1996) A history of English criminal law and its administration from 1750. Stevens, London

Ruback RB, Bergstorm MH (2006) Economic sanctions in criminal justice: purposes, effects, and implications. Crim Justice Behav 33:242–273

Rusche G, Kirchheimer O (1939) Punishment and social structure. Columbia University Press, New York

Seagle W (1948) Fines. In: Seligman E, Johnson A (eds) Encyclopedia of the social science. Macmillan, New York

Uglow S (1995) Criminal justice. Sweet & Maxwell, London

University of Pennsylvania Law Review (1953) Fines and finding – an evaluation. Univ Pa Law Rev 101:1013–1030

Weisburd D, Einat T, Kowalski M (2008) The miracle of the cells: an experimental study of interventions to increase payment of court ordered financial obligations. Criminol Public Policy 7(1):9–36

Zender L (2004) Criminal justice. Oxford University Press, Oxford, pp 204–209

History of Forensic Science in Policing

Simon A. Cole
Department of Criminology, Law & Society,
University of California, Irvine, Irvine, CA, USA

Synonyms

Criminalistics

Overview

The nascent development of institutional forensic science as we know it today, in the late nineteenth century, was a crucial component in the larger process of the professionalization of policing. In the twentieth century, forensic science transitioned from a location outside of police agencies, primarily in universities, and an auxiliary role assisting police investigators toward greater integration with, and in some cases even location in, police agencies. The location of forensic science in police agencies brought many benefits in terms of coordination and control of the crime scene and potentially in the development of "intelligence-led policing" but also brought costs in terms of their perceived impartiality and its links to "mainstream" scientific institutions. The capabilities of forensic science have developed enormously over the course of the twentieth century, but research suggests that it continues to be underused by police.

Introduction

Forensic science may reasonably be viewed as having played an important role in the professionalization of policing that began in the late nineteenth century (Jones 1994; Laybourn and Taylor 2011). While there are histories of forensic science (generally written by practitioners) and histories of forensic medicine, expert witnesses, and the professionalization of policing (generally written by historians), the intersection of these developments remains underexplored.

Forensic science is, of course, quite old and predates the development of professional police and, indeed, police altogether. The origins of forensic science and medicine have been traced as far back as the ancient Egyptians, Greeks, Chinese, and Aztecs (Bell 2008). However, forensic science as an institution as we know it today, with dedicated laboratories, infrastructure, professional organizations, and so on, only began developing in the late nineteenth century. While this development began outside of police agencies, later some aspects of institutional forensic science were incubated within police agencies.

Institutional forensic science as it exists today generally traces its history to late nineteenth-century physicians who were interested in what was called at the time "legal medicine." The Medico-Legal Institute was founded in Paris in 1868 (Watson 2011) and "during the half-century following 1870, state- or municipally-sponsored institutes of forensic medicine were established in Paris, Berlin, Vienna, Budapest and many other principal cities of Europe" (Ambage and Clark 1994, p. 293). Late nineteenth-century pioneers included Alexandre Lacassagne in Lyon; Joseph Bell in Edinburgh (who has been identified as the Conan Doyle's model for the fictional detective Sherlock Holmes); John Glaister, Sr. in Glasgow; Bernard Spilsbury in England; Hans Gross in Austria, who published what has been called "the first textbook on criminalistics" (Jones 1994) in 1893; and R.A. Reiss in Lausanne. The primary site of this activity was at universities, specifically medical schools, in France and Britain, or law schools, in the German-speaking countries. As police agencies professionalized and became increasingly interested in solving crimes, these physicians were called upon to assist with investigations. Forensic science, however, was external to the police; there was no question of locating forensic science within the police agencies themselves.

Some have suggested that it was what we might call the "identification disciplines" – photography, anthropometry, and fingerprinting – that occasioned the crossover of forensic science from medicine and the university to police agencies (Valier 1998; Roux and Robertson 2009; Margot 2011; cross-reference: ▶ Identification and the Development of Forensic Science). These three techniques were all developed not for forensics, but for criminal record-keeping purposes: for linking offenders held in custody to their past criminal histories with greater accuracy without being thwarted by the use of aliases or the adoption of false identities. These technologies were adopted in divisions of police agencies known as "identification bureaus" and were manned by "identification clerks" generally drawn from the ranks of the police. One of these techniques, fingerprinting, happened to also lend itself to a forensic application: the linking of accidental prints (or "latent prints" or "marks") left at crime scenes to specific suspects. Although some early research on fingerprints was performed by physicians and scientists, it seemed natural to use the identification clerks, who classified fingerprint patterns for record-keeping purposes, as "experts" to analyze latent prints for forensic purposes. Thus, police identification clerks began to share the title "forensic scientist" with physicians and chemists (Cole 2001).

Fingerprinting became perhaps the earliest technique whose use was so routine that it could be delegated from physicians to "specially trained policemen as its use became widespread amongst police forces" (Crowther and White 1988).

The development of institutions fully devoted to forensic science, as opposed to treating forensic science as an aspect of, say, public health or pathology, took two general routes (Margot 2011). The first was the university. For example, in 1909, Reiss established the first stand-alone School of Forensic Science at Lausanne, associated with the faculties of both law and medicine (Margot 2011). In 1908, August Vollmer, the Police Chief of Berkeley, founded a Police School in association with the University of California which included forensic science among other topics intended to make policing more professional and generally "scientific" (Carte and Carte 1975). And, Calvin Goddard established the Scientific Detection Crime Laboratory at Northwestern University in 1930 (Tilstone et al. 2006).

The other route was outside the university, in institutions dedicated to criminal justice. Edmond Locard, a former student and assistant of Lacassagne, left the University of Lyon to found the Police Scientific Laboratory, the world's first police laboratory, in Lyon in 1912 (Watson 2011). Significantly, the laboratory was controlled by the court, not the police (Margot 2011). Other early "criminalistics" or "police science" laboratories were attached to government chemical or toxicological laboratories, as in Sweden (Ambage and Clark 1994). In England, many police agencies used the services of private laboratories (Laybourn and Taylor 2011). However, most crime laboratories outside universities soon fell into the control of police agencies. The Federal Bureau of Investigation crime laboratory opened in 1932 (Walker 1977). "By 1930, almost every European capital and many other principal cities has its own medico-legal institute, while France, Italy, Germany, Switzerland, Sweden and... Finland all had at least one 'police science' laboratory" (Ambage and Clark 1994, p. 293), as well as Egypt and Ceylon. The exception was England. The Metropolitan Police "rebuilt and reorganized" (Laybourn and Taylor 2011, p. 94) its old laboratory into a modern laboratory at Hendon in 1935, and the Home Office established the Forensic Science Service and opened its first laboratory in Nottingham in 1936 (Watson 2011). Neither of these was an institute of legal medicine, however – a proposal to house such an institute at the University of London failed, signaling the separation of forensic science from forensic medicine and the "apparently inexorable decline of English forensic medicine" (Ambage and Clark 1994, p. 307).

These two routes seemed to cause some tension between the development of forensic science outside and within police agencies (Margot 2011; cross-reference: ▶ Identification and the Development of Forensic Science). Many forensic scientists located at universities seeking to develop, expand, or maintain stand-alone institutes dedicated to forensic science were undermined somewhat by the steady assumption of routine forensic work by police agencies (Crowther and White 1988). Thus, forensic laboratories today tend to be located in law enforcement agencies, although there are also national laboratories and laboratories located in universities, prosecuting authorities, or judiciaries, especially in Europe (Tilstone et al. 2006). By the post-World War II period, even some university-based forensic laboratories, such as the pioneer laboratory at the University of Glasgow, came to be staffed in large part by the police (Crowther and White 1988), and the police won control of the Police Scientific Laboratory in Lyon (Margot 2011). Some have suggested that the university-police partnership that seemed to be emerging after 1920 never really came to fruition. Instead, nearly all technical developments in forensic science have come from mainstream academic or industrial science, not from either the police laboratories or the university forensic science programs (Tilstone et al. 2006). The US National Research Council (2009) also recently noted forensic science's "lack [of] strong ties to our research universities and national science assets."

Although forensic science began to be used early in the twentieth century in isolated places and cases, it was only in the 1930s "that detective and scientific work became more truly uniform and modernised and increasingly subject to government guidelines and direction (Laybourn and Taylor 2011). Innovation in forensic science and technology and its integration into police work provided clear benefits in terms of solving crimes, one, but by no means the only, function of the police. However, there have been recurrent complaints that forensic science was underused by police investigators. Many forensic techniques were useful in linking known suspects of crime scenes, but until the development of searchable databases, they were of little help in generating suspects when none existed. The development of computerized, searchable databases, therefore, was a crucial development, perhaps even more important than the remarkable innovations in forensic technologies themselves.

As Peterson et al. note (2010), "there is little published empirical data identifying the types of evidence routinely collected, and the extent to which this evidence is submitted to and examined in forensic laboratories." However, what research exists suggests that forensic science has been utilized far less than one might have imagined given its seemingly utility for investigating crimes. For example, Peterson et al. (1987) found relatively low uses of forensic science in a sample of criminal investigations in American cities. Ethnographers of detective work found police investigators far more focused on nonforensic activities like interviewing witnesses, informants, and suspects than on forensic science. Nonetheless, the utilization of forensic science appears to be increasing. Almost completely absent in Ericson's 1981 ethnographic report, it appears somewhat more in more recent detective ethnographies (Innes 2003; Jackall 2005). Nonetheless, researchers have continued to find forensic evidence, including DNA evidence, grossly underused and poorly understood by police (Raymond et al. 2004; Pratt et al. 2006; Strom and Hickman 2010). Strom and Hickman assert that in addition to the better-known backlog of forensic evidence at crime laboratories – where samples are not processed because of insufficient resources – there is a less well-known backlog of evidence that is collected by the police but never submitted to the laboratory for analysis. Consistent with Pratt et al.'s findings, Strom and Hickman's qualitative survey reported that the primary reason for underuse of DNA profiling is that it is conceived by police investigators as a tool for building evidence against a suspect identified by other means rather than as a means of generating a suspect by treating existing archives of genetic information as what have been called "DNA intelligence databases" (Walsh and Buckleton 2005). Peterson et al. (2010) likewise found forensic evidence to be submitted to laboratories at low rates. They concluded, "it is clear that criminal justice officials external to the laboratory screen much of the forensic evidence and have a major influence on evidence examination priorities and practices" (Peterson et al. 2010).

Whether police agencies' historic preference for non-forensic methods is desirable is an interesting question. As forensic capabilities, capacity, and efficiency increase and costs decrease, important questions will arise about whether greater reliance on forensic science in the investigation will be preferable from the perspectives of security, social control, social justice, civil and human rights, and economic rationality. Such questions have hardly been explored in criminology (but see, e.g., Duster 2004).

The tendency to locate forensic laboratories in police agencies undoubtedly facilitated close coordination between forensic science and police investigators and enabled coordinated control of crime scenes. However, this coordination apparently came at a cost. Recently, the location of the bulk of forensic science infrastructure in police departments has emerged as a major point of criticism of forensic science (e.g., Giannelli 1997). The criticism is generally twofold: First, philosophers and sociologists of science have generally cited independence from institutional, and especially state, control as crucial to allowing scientists to operate as if their sole allegiance is to scientific truth. This is generally thought to require the location of scientists in institutions where such freedom is permitted and encouraged. Universities, with their notion of "academic freedom," are the most common models, of course, but even industrial (especially the celebrated industrial research laboratories of the twentieth century, now largely an extinct species) and national laboratories have been conceived as bestowing varying degrees of such freedom and independence. If forensic science is to be considered a scientific endeavor (and this itself has been the subject of some debate), its overwhelming location within state institutions would seem to violate this requirement. Second, the *telos* of all forensic analysis is a report or testimony that a legal party (usually the state) can at least threaten to use (even if it rarely is actually used) in a criminal trial. Under an adversarial system of justice, the trial is seen as a contest between two sides; vigorous (though fair) pursuit of victory by each side is supposed to generate the best chance of reaching an outcome corresponding with justice. In such systems, the association, whether administrative, financial, or emotional, of forensic science institutions with one of those sides (the state) to the exclusion of the other (the suspect) seems irredeemably problematic. Even in inquisitorial systems of justice, the association of forensic science with the state seems at least as problematic, if not even more so.

For these reasons, the tendency for forensic science institutions to be located within state law enforcement agencies is generally cited as perhaps the leading deficiency of contemporary forensic science among those interested in the reform and improvement of forensic science today (e.g., National Research Council 2009). But the prospects for change seem daunting. The major points of resistance are probably political and resource driven; forensic science organizations are understandably skeptical that state governments would fund forensic science to even its current levels if its mission was construed as driven by science rather than public safety.

Removing forensic laboratories from police agencies would also entail some costs in terms of coordination, such as allowing forensic scientists to use the nature of the case to decide what items should be tested and how they should be tested. It is also argued that forensic evidence can only be understood in the context of the whole case (e.g., Margot 2011). Looking forward, it is argued that integration of forensic science into police agencies will facilitate the trend toward emerging "intelligence-led policing" (Ribaux et al. 2006; Margot 2011). Some scholars, however, argue that forensic science would improve with greater distance from police agencies. Forensic scientists would focus more on the "pure" results of their scientific analysis, and their interpretations of the evidence would be less tainted by contextual information about the case fed to them by investigators (Risinger et al. 2002; Dror and Charlton 2006; Krane et al. 2008). With these compelling principles at odds, the issue of whether the current arrangement of locating forensic science in police agencies is likely to be hotly debated in coming years. As this historical review shows, the current arrangement is not eternal: It is not as old as it may seem, and it is the product of specific contingent decisions and circumstances early in the history of interactions between newly professionalizing police forces and the nascent, hybrid field called "forensic science."

Related Entries

- ▶ Cognitive Forensics: Human Cognition, Contextual Information, and Bias
- ▶ Critical Report on Forensic Science
- ▶ DNA Profiling
- ▶ DNA Technology and Police Investigations
- ▶ Fingerprint Identification
- ▶ Forensic Science and Criminal Investigation
- ▶ Forensic Science and Miscarriages of Justice
- ▶ Forensic Science and the Paradigm of Quality
- ▶ Forensic Science Culture
- ▶ Forensic Science Effectiveness
- ▶ Forensic Science Professions
- ▶ History of Criminal Investigation
- ▶ History of Technology in Policing

- ▶ Identification Technologies in Policing and Proof
- ▶ Information Technology and Police Work
- ▶ Intelligence-Led Policing
- ▶ Legal Rules, Forensic Science and Wrongful Convictions
- ▶ Organizing and Supplying Forensic Science Services
- ▶ Philosophical Basis of the Forensic Process
- ▶ Philosophy of Forensic Identification
- ▶ Probability and Inference in Forensic Science
- ▶ Scientific Basis of the Forensic Process
- ▶ Transnational Exchange of Forensic Evidence

Recommended Reading and References

Ambage N, Clark M (1994) Unbuilt Bloomsbury: medico-legal institutes and forensic science laboratories in England between the wars. In: Clark M, Crawford C (eds) Legal medicine in history. Cambridge University Press, Cambridge, pp 293–313

Bell S (2008) Crime and circumstance: investigating the history of forensic science. Praeger, Westport

Carte GE, Carte EH (1975) Police reform in the United States: the era of August Vollmer, 1905–1932. University of California Press, Berkeley

Cole SA (2001) Suspect identities: a history of fingerprinting and criminal identification. Harvard University Press, Cambridge

Crowther MA, White B (1988) On soul and conscience: the medical expert and crime. Aberdeen University Press, Aberdeen

Dror IE, Charlton D (2006) Why experts make errors. J Forensic Ident 56(4):600–616

Duster T (2004) Selective arrests, an ever-expanding DNA forensic database, and the specter of an early-twenty-first-century equivalent of phrenology. In: Lazer D (ed) DNA and the criminal justice system: the technology of justice. MIT Press, Cambridge, pp 315–334

Ericson RV (1981) Making crime: a study of detective work. Butterworths, Toronto

Giannelli P (1997) The abuse of scientific evidence in criminal cases: the need for independent crime laboratories. Va J Soc Policy Law 4:439–478

Innes M (2003) Investigating murder: detective work and the police response to criminal homicide. Oxford University Press, Oxford

Jackall R (2005) Street stories: the world of police detectives. Harvard University Press, Cambridge

Jones CAG (1994) Expert witnesses: science, medicine, and the practice of law. Clarendon, Oxford

Krane DE, Ford S, Gilder JR, Inman K, Jamieson A, Koppl R, Kornfield IL, Risinger DM, Rudin N, Taylor MS, Thompson WC (2008) Sequential unmasking: a means of minimizing observer effects in forensic DNA interpretation. J Forensic Sci 53(4):1006–1007

Laybourn K, Taylor D (2011) Policing in England and Wales, 1918–39. Palgrave Macmillan, Houndmills

Margot P (2011) Forensic science on trial – what is the law of the land? Aust J Forensic Sci 43(2–3):89–103

National Research Council (2009) Strengthening forensic science in the United States: a path forward. National Academies Press, Washington, DC

Peterson JL, Ryan JP, Houlden PJ, Mihajlovic S (1987) The uses and effects of forensic science in the adjudication of felony cases. J Forensic Sci 32(6):1730–1753

Peterson J, Sommers I, Baskin D, Johnson D (2010) The role and impact of forensic evidence in the criminal justice process. National Institute of Justice, Washington, DC

Pratt TC, Gaffney MJ, Lovrich NP, Johnson CL (2006) This isn’t ’CSI’: estimating the national backlog of forensic DNA cases and the barriers associated with case processing. Crim Justice Pol Rev 17(1):32–47

Raymond JJ, Walsh SJ, Van Oorchot RA, Gunn PR, Roux C (2004) Trace DNA: an underutilized resource or Pandora’s Box? A review of the use of trace DNA analysis in the investigation of volume crime. J Forensic Ident 54(6):668–686

Ribaux O, Walsh S, Margot P (2006) The contribution of forensic science to crime analysis and investigation: forensic intelligence. Forensic Sci Int 156:171–181

Risinger DM, Saks MJ, Thompson WC, Rosenthal R (2002) The *Daubert/Kumho* implications of observer effects in forensic science: hidden problems of expectation and suggestion. Cal L Rev 90(1):1–56

Roux C, Robertson J (2009) The development and enhancement of forensic expertise: higher education and in-service training. In: Fraser J, Williams R (eds) Handbook of forensic science. Willan, Uffculme, pp 572–601

Strom KJ, Hickman MJ (2010) Unanalyzed evidence in law enforcement agencies: a national examination of forensic processing in police departments. Criminol Public Policy 9(2):381–402

Tilstone WJ, Savage KA, Clark LA (2006) Forensic science: an encyclopedia of history, methods, and techniques. ABC-Clio, Santa Barbara

Valier C (1998) True crime stories: scientific methods of criminal investigation, criminology and historiography. Brit J Criminol 38(1):88–105

Walker S (1977) A critical history of police reform: the emergence of professionalism. Lexington Books, Lexington

Walsh S, Buckleton J (2005) DNA intelligence databases. In: Buckleton J, Triggs CM, Walsh SJ (eds) Forensic DNA evidence interpretation. CRC Press, Boca Raton, pp 439–469

Watson KD (2011) Forensic medicine in western society: a history. Routledge, London

History of Geographic Criminology Part I: Nineteenth Century

Gerben Bruinsma[1,2] and David Weisburd[3,4]
[1]Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, The Netherlands
[2]VU University Amsterdam, Amsterdam, The Netherlands
[3]Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA
[4]Faculty of Law, The Hebrew University, Mt. Scopus, Jerusalem, Israel

Synonyms

Early Chicago School; Geography of crime and disorder; Historical criminology; History of criminological theories: Causes of crime; History of geographic criminology part II: Twentieth century

Overview

Criminologists have had a long enduring interest in the geography of crime and the explanation of spatial variations of crimes and offenders. In this entry, we describe the first century of empirical research starting in Europe in the late 20s and early 30s in Europe, especially in Belgium and France and in England (London). These studies contain the first analyses of registered figures on crime and prisoners. The results were also displayed on maps. There existed a vivid discussion between mathematicians, statisticians, and politicians during several meetings in London. The unit of analyses was in the beginning larger administrative areas and countries. The research agenda was inspired by new administrative data that became available in France in 1825 and by the political agenda of reformism, a political movement that after the Napoleon wars aimed at the improvement of education and health for the poor and illiterates. After 1860, the interest in geographic criminology in Europe diminished. Criminologists in Chicago took over the lead in geographic research in the beginning of the twentieth century.

The First Steps in France and Belgium

Geographical criminology begins with the publication of statistics on the French population by the French Home Office in the 1820s. The publication of the *Comptes Générales de l'Administration de la Justice Criminelle en France* in 1825 inspired many statisticians and other scholars to explore in more detail registered data on crime. Among the very first was Baron Charles Dupont who spoke for the first time about statistics on morality of people and the nation in a meeting of the *Conservatoire des Arts et Métiers* in November 1826 (Beirne 1987). As a cartographer, he published in 1816 tables on the "distribution of illiterates" across the regions of France. During this meeting, he promised to publish a map on crime and criminals soon. However, the Belgian Edouard Ducpétiaux was the first to publish a table with crime and suicide figures of regions in Spain, France, Italy, and England (Ducpétiaux 1827). He observed remarkable differences between regions and countries and concluded that the morality of nations differed (Ducpétiaux 1827, pp. 11–12). The Lower Lands (the Netherlands) were especially praised for their low crime rates and for a low number of homeless people (Ducpétiaux 1827, p. 29).

In 1829, the first geographical map of crime was published. Partly based on the *Comptes Généraless de l'Administration de la Justice Criminelle en France*, Michel-André Guerry and the Venetian cartographer Adriano Balbi published, on one large sheet, three maps on the distribution of crime in France in the years 1825–1827 (Balbi and Guerry 1829). It was a novelty in the new field of criminology that they made use of a cartographic method of presenting statistical material. They concluded from their work that (1) in certain regions in France (departments) with higher numbers of personal crimes (violent crimes), there was less property crime; (2) the area above the line of Orléans and Lyon showed the highest rate of property crimes in France; and (3) in urban areas, especially in the capital of Paris, the highest numbers of property and personal crimes could be observed. Later, when Guerry became head of the Crime Statistics Unit of the France ministry of Justice, he continued his work on mapping crime. In 1833, his influential *Essai sur la Statistique Morale de la France* was published (Guerry 1833). Inspired by the Reform Movement of the nineteenth century (in France and England, the Reform Movement focused its policy on public health and education for the poor), Guerry examined whether poverty and density of population might lead to higher crime rates. He observed however an empirical complexity. The rich north *departements* were confronted with higher property crime rates than the poor *departements* in the south of France. He concluded that the level of poverty was not the direct cause of crime. Similarly, his data suggested that population density was not a cause of crime.

The studies by Guerry soon became the subject of a heated debate between proponents and opponents of the Reform Movement, especially in England. A British Member of Parliament defended the work of Guerry, but the industrialist William Greg criticized Guerry's conclusion that education or poverty might not be a cause of crime (Greg 1835). In 1864, Guerry published again a comparison of crime rates between England and France (Guerry 1864). We can hardly imagine today the work he had to carry out collecting and analyzing the data, which included over 226,000 cases of personal crime in the two countries over 25 years and, for France only, over 85,000 suicide records (Friendly 2007). The results of this effort reinforced the findings he had published 30 years before.

His friend Alexandre Parent-Duchâtelet published in 1836 an empirical study containing maps on the distribution of prostitution from 1400 till 1830 in Paris (Parent-Duchâtelet 1837 [1836]). Because of the official control of brothels by the Paris authorities, systematic data were available on prostitutes, especially from the years 1817–1827. Even information regarding the *departements* they came from was collected. Not surprisingly, the center of the city had the highest number of prostitutes, especially *quartier 6*. He used neighborhoods, as defined by administrative boundaries as units of analyses.

The French scholar Michel-André Guerry is often bracketed together with the Belgium statistician and astronomer Adolphe Quételet who discovered the normal distribution in statistics with which deviations can be observed and calculated (Landau and Lazarsfeld 1968). In 1828, Quételet (1984 [1831]) examined the French *Comptes Gé nérales de l'Administration de la Justice Criminelle en France* in a lecture for the Academy of Science to show how crime rates vary with the seasons of the year. He demonstrated that June had the highest numbers of violent crimes of the year. He also examined crime data in each *arrondissement* and added: "It appeared to me that these numbers were able to give a sufficiently satisfactory idea of the state of knowledge in each department and especially for the inferior classes where most crimes are committed" (Quételet 1984 [1831], p. 30). He used provinces and countries as units of analyses. Groningen, South Brabant, Anvers, Limburg, and Drenthe had the highest crime rates of all provinces of the Lower Lands (including Flanders). He discussed these observations as was common in the period, in the context of social factors such as poverty, heterogeneity, composition of the population, and attractiveness of cities. He explained the higher rates of property crimes of the richer provinces, as we noted in our introduction, by the unequal distribution of wealth: A great number of people possess nothing compared to the relatively few rich citizens (Quételet 1984 [1831], p. 38).

The French and Belgian scholars were the first who scientifically analyzed crime geographically. These scholars focused on the administrative and political borders of their time in their spatial crime analyses. Nations, regions, counties, provinces, departments, and *quartiers* were the units of analyses, and they were used as a unit for systematic comparisons of crime figures. They were fully dependent on official crime data and other data that the government supplied arranged within these larger geographic units. While the early French and Belgian researchers concerned with crime and place also examined some variability of crime within cities, their overall focus was generally on larger administrative units. Importantly, these early criminologists in their focus on crime rates and official statistics also helped to encourage the more general development of a positivist empirical criminology.

The First Steps in England

France and Belgium were not the only countries where geographical studies on crime were carried out in the nineteenth century. Members of *the Statistical Society of London* also regularly published on crime topics in their statistical journal (first published in 1838). Two articles on crimes were included in the second volume. One was by a prison chaplain who wrote on *Criminal Statistics of Preston* and the other was a short article about the distribution and kinds of robberies in London and Liverpool (mentioned in Morris 1957, p. 53). In this and later papers of members of the Society, the influence of the work of Guerry can be observed.

In the year 1839, a lecture by William Greg on the spatial distribution of population density, fertility, education, and crime in the Netherlands (in those days, parts of Belgium and Luxembourg were also part of the Netherlands) for the *British Association for the Advancement of Science* of August 1835 was published by Ridgway and Sons (Greg 1835). He compared the crime figures on property crimes, violent crimes, and on serious crimes like rape, murder, and manslaughter in the Netherlands with those of England and France. The overall crime figures showed remarkable differences: For the year 1826, the data indicated that the Netherlands had less crime than the other two countries: 1:28.900 inhabitants, in England 1:23.400, and in France 1:17.570 (Greg 1835, p. 15). Greg also studied in more detail large areas of the Lower Lands. He counted the most crimes for the province of Overijssel, followed by South Brabant, Groningen, Drenthe, and West Flanders (like Quételet observed 25 years later). The lowest overall crime rates were found in the province of Friesland. For that period, the highest position of Overijssel in the crime ranking was striking because this province also showed the highest education level for all of Europe (Greg 1835, p. 24). Serious crime was more likely to be present in the provinces of South Brabant, Liege, Groningen, and East Flanders.

Rawson R. Rawson, then secretary of the Society, correlated variables such as age and sex with crime but also presented the number of crimes for the districts of England and Wales (Rawson 1839). He discussed the variability in crime across the districts and concluded that climate and ethnic differences in the population could not be causes of crime. He rejected the conclusions of Guerry on the relationship between education and crime. He assumed that the type of labor could be a cause of crime and divided regions in England into (1) rural areas with farms, (2) industrial areas, (3) mining areas, and (4) big cities. Rawson's research illustrates the critical role that the unit of analysis can play in the development of our understanding of crime at place. He is the first scholar to go beyond the usual administrative and political borders (Morris 1957, p. 55). Based on official data, he found that large cities had the highest crime rates and mining areas the lowest.

His successor at the Society, Joseph Fletcher, continued the work of Guerry and Rawson (Fletcher 1848, 1849). He studied for many years the relationship between education and crime by producing maps showing the levels of crime and illiteracy of England and Wales (later published in Fletcher 1850). According to Fletcher, there were four causes of the level of "immorality" of populations: (1) population density, (2) the distribution of property across societies, (3) the number of people earning their own income, and (4) the level of illiteracy of the population (indicated by the number of crosses instead of signatures in the registers of marriage). He indicated the level of immorality of the population by the number and nature of crimes committed, the number of marriages with a man younger than 21 (indicating forced marriage because of pregnancy), the number of illegitimate children, and the number of bank accounts in the population. He argued (1850) that not only were differences between regions important in the explanation of crime rates, but so was the speed with which these regions changed over time economically and demographically. With these ideas, Fletcher can be seen as a precursor of the French sociologist Emile Durkheim who introduced the concept of anomie to explain the impact of societal changes on people and society (Durkheim 1964 [1893]).

John Glyde (1856) was the first to question the validity of the research findings when large areas were chosen as units of analysis in geographic criminology. In his paper *Localities of Crime in Suffolk*, he showed very clearly that larger units of analysis hide underlying variations in crime. When smaller units than districts or *departements* were taken into account, significant differences in crime rates across smaller areas appeared. As Morris (Morris 1957, p. 58) notes: "Of the regional studies, a major criticism is that the county was the smallest territorial unit considered, but Glyde, by breaking Suffolk down into its seventeen Poor Law Unions was able to demonstrate that the 'County Aggregate' masked considerable differences between the smaller geographical units of which it was composed." Glyde also observed that middle-sized cities situated along main roads had higher crime rates than the mean of the large area they were part of. Jelinger Symons (2000 [1857], p. 281) also examined the relationship between urbanization and crime through ecological analyses of crime in Liverpool, Bristol, and Cardiff. In his view, it was the speed with which the population increases that explains higher crime figures.

In studies of crime and place in England in the nineteenth century, the work of Henry Mayhew cannot be neglected. He is well-known in criminology (and cited therefore) for his descriptive studies of the underworld of London in the middle Victorian Age (1851/1950). However, his detailed studies on the distribution of crime in England and London are also a rich source for those studying crime. He was an excellent observer of his time, describing in four voluminous books precisely and in detail the daily life of the Londoners, their habits, their cloths, their quarters, and streets in the Victorian era (Mayhew 1865, 1950 [1851/1862]). Mayhew also tried to uncover patterns in the distribution of crime in the city of London combining ethnographic methods as well as statistical data. Mayhew's was the precursor of Robert Park who advocated 70 years later a similar methodology. He interviewed prostitutes, criminals, and other citizens about alcoholism, poverty, housing conditions, and economic uncertainty. He was the first scholar who focused on small areas like squares, streets, and buildings as a unit of analysis in criminological research, predating modern interests in micro crime places (see later) by over a century. Mayhew also used police data of the seven Metropolitan Police Divisions and revealed that two of these divisions produced about 65 % of all the suspects in London.

The Decline of European Geographic Criminology

After 1870, the interest of French and English academics in geographic and statistical analyses diminished. At the same time, there are important exceptions. In one of them, Cesare Lombroso studied the geographical distribution of homicide, infanticide, parricide, and suicide across very large geographic areas in Italy (Lombroso 2006 [1878]). He explained the different violent crime rates between the north and the south of the country in reference to the "racial inferiority" of the southern population. The French criminologist Jean-Gabriel Tarde studied the relationship between urbanization and crime from a different perspective than earlier scholars (Tarde 1912 [1890], p. 338). By comparing the crime levels of larger areas, he argued that cities were exporting crime to the rural areas. He wrote: "... today we can see crime spreading from the great cities to the country, from the capitals to the provinces, and these capitals and great cities having an irresistible attraction for the outcasts and scoundrels of the country, or the provinces, who hasten to them to become civilized after their own manner, a new kind of ennobling" (Tarde 1912 [1890], p. 338).

According to Brantingham and Brantingham (1981, p. 12), the nineteenth-century geographic studies in the spatial distribution of crime have revealed important insights: (1) Crime rates vary substantially between geographical areas; (2) these variations are found at different levels of aggregation; (3) these variations persist over (long) periods of time; (4) high crime rates are highly correlated with the presence of other social problems like poverty, alcoholism, high population density, and illiteracy; and (5) lastly, that locations with high number of crimes correlate with the presence of attractive targets and potential victims.

Recommended Reading and References

Balbi A, Guerry A-M (1829) Statistique comparée de l'état de l'instruction et du nombre des crimes dans les divers arrondissements des Académies et des Cours Royales de France. Unknown Publisher, Paris

Beirne P (1987) Adolphe Quételet and the Origins of Positivist Criminology. Am J Sociol 92(5): 1140–1169

Brantingham PJ, Brantingham PL (eds) (1981) Environmental criminology. Sage, Beverly Hills

Ducpétiaux E (1827) De la Justice de prévoyance, et particulierement de l'influence de la misere et de l'aisance, de l'ignorance et de l'instruction sur le nombre des crimes. J.J. de Cautaerts, Bruxelles

Durkheim E (1964 [1893]) The Divison of Labor in Society. Free Press, New York

Fletcher J (1848) Moral and educational statistics of England and Wales. J Stat Soc Lond 11:344–366

Fletcher J (1849) Moral and educational statistics of England and Wales. J Stat Soc Lond 12:151–335

Fletcher J (1850) Summary of the moral statistics of England and Wales. Private distribution of the author, London

Friendly M (2007) A.-M. Guerry's moral statistics of France: challenges for multivariable spatial analysis. Stat Sci 22:368–399

Glyde J (1856) Localities of crime in Suffolk. J Stat Soc Lond 19:102–106

Greg WR (1835) Social statistics of the Netherlands. Harrison and Crosfield, London Ridgway

Guerry A-M (1833) Essai sur la statistique morale de la France. Crochard, Paris

Guerry A-M (1864) Statistique morale de l'Angleterre comparee avec la Statistique Morale de la France. Bailliére, Paris

Landau D, Lazarsfeld PF (1968) Quételet, adolphe. Int Encycl Soc Sci 13:247–257

Lombroso C (2006 [1878]) Criminal man (translated and with a new introduction by M. Gibson & N. Han Rafter). Duke University Press, Durham/London

Mayhew H (1865) London labour and the London poor, vol 4, Those that will not work. Charles Griffin, London

Mayhew H (1950 [1851/1862]) London's underworld. Being selections from "Those that will not work", 4th vol. of "London labour and the London poor" (ed Quennell P). Spring Books, London

Morris T (1957) The criminal area. A study in social ecology. Routledge & Kegan Paul, London

Parent-Duchâtelet A (1837 [1836]) De la prostitution dans la ville de Paris: considéré sous le rapport de l'hygiène publique, de la morale et de l'administration. Baillière, Paris

Quételet A (1984 [1831]) Research on the propensity for crime at different ages (translated and with an introduction by Sylvester SF). Anderson Publishing, Cincinnati

Rawson RW (1839) An inquiry into the statistics of crime in England and Wales. J Stat Soc Lond 2:316–344

Symons J (ed) (2000[1857]) On crime and density of population. Copperhouse Publishing, Incline Village

Tarde G (1912 [1890]) Penal philosophy (trans: Howell R). Little, Brown, Boston

History of Geographic Criminology Part II: Twentieth Century

Gerben Bruinsma[1,2] and David Weisburd[3,4]
[1]Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, The Netherlands
[2]VU University Amsterdam, Amsterdam, The Netherlands
[3]Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA
[4]Faculty of Law, The Hebrew University Mt. Scopus, Jerusalem, Israel

Overview

After the turn of the century, the locus of geographic research on crime moved to the United States, and especially to the city of Chicago. A group of American sociologists, among them Robert Park, William Thomas, Louis Wirth, Ernest Burgess, Clifford Shaw, and Henry McKay, took a leadership role in the development of the criminology of place, in contrast to the statisticians, criminal lawyers, or psychiatrists who dominated criminology more generally in Europe (see entry on the ► History of Geographic Criminology Part I: Nineteenth Century). New theoretical concepts were introduced (notably social disorganization) and innovative methods and techniques to study empirically the distribution of crime in cities. In the 1950s, methodological critiques and developments in other areas of criminology diminished the research attention to geographic criminology. In the 1980s, young scholars reemerged the study of spatial issues studying the dynamics of neighborhoods, elaborating the concept of social disorganization. Recently, the focus is more laid on micro places like blocks, squares, and street segments, because empirical research revealed great variations in crime within neighborhoods.

Chicago and the Study of Dynamics in Cities: Variations in Crime and Offender Rates Between Neighborhoods

After the turn of the century, the locus of geographic research on crime moved to the United States, and especially to the city of Chicago. At the University of Chicago, a group of sociologists took the initiative to undertake new research on urban problems, which centered, in part, on crime (Beirne and Messerschmidt 1991; Bulmer 1984; Burgess and Bogue 1964b; Faris 1967; Harvey 1987). They also moved the action of crime and place research from broad comparisons across large geographic areas to more careful comparisons within cities. Interestingly, the Chicago School scholars were either not aware of or ignored till 1933 the work of nineteenth-century crime and place researchers in Europe (Elmer 1933; Levin and Lindesmith 1937).

American cities grew in the second part of the nineteenth century and first part of the twentieth century faster than ever before in history, with all the social problems associated with such growth. Chicago itself played an important role in the integration of large numbers of Italian, Irish, German, Chinese, Polish, Jewish, and Scandinavian immigrants. The city evolved from a very small settlement in 1840 (with 4,470 inhabitants) to a city with a half-million inhabitants in 1880. Ten years later, the population had increased to one million, and in 1930, to about 3.5 million. Crime was perceived as one of the most important urban problems:

“After World War I (1914–1918), Chicago sociologists turned their ecological attentions to a variety of social problems. Exacerbated by the severe hardships of the Great Depression, Prohibition, and by the well-publicized rise of gangland warfare and union racketeering, crime itself came to be seen as a major social problem in Chicago. Crime, therefore, was one of the chief topics studied by members of the Chicago-School.” (Beirne and Messerschmidt 1991, p. 362)

Now, a group of American sociologists, among them Robert Park, William Thomas, Louis Wirth, Ernest Burgess, Clifford Shaw, and Henry McKay, took a leadership role in the development of the criminology of place, in contrast to the statisticians, criminal lawyers, or psychiatrists who dominated criminology more generally in Europe (Vold et al. 2002).

William Thomas contributed to the criminology of place by introducing the important concept of *social disorganization*, referring to "a decrease of the influence of existing social rules of behavior upon individual members of the group" (Thomas 1966, p. 3). The concept naturally focused on neighborhoods or communities. Robert Park (1864–1944), who was recruited by Thomas, was the initiator of urban social research on crime places, shifting the unit of analyses from countries and large areas to cities and their neighborhoods (Park and Burgess 1967 [1925]). The city in his opinion was more than "…a congeries of individual men and of social conveniences – streets, buildings, electric lights, tramways, and telephones, etc; something more, also, than a mere constellation of institutions and administrative devices – courts, hospitals, schools, police and, civil functionaries of various sorts. The city is, rather, a state of mind, a body of costumes and traditions, and of the organized attitudes and sentiments that inhere in these costumes and are transmitted with this tradition. The city is not, in other words, merely a physical mechanism and an artificial construction. It is involved in the vital process of the people who compose it; it is a product of nature, and particularly of human nature" (Park and Burgess 1967 [1925], p. 1). Park argued that urban life must be studied in this context in terms of "its physical organization, its occupations, and its culture" *and* especially the changes therein (Park and Burgess 1967 [1925], p. 3). Neighborhoods in his view were the elementary form of cohesion in urban life.

His younger colleague, Ernest Burgess, drawing from an inventory of price changes in housing values and other real estate in Chicago by a real estate agent developed a concentric zone model of the distribution of social problems and crime for cities (especially for Chicago) (Burgess 1967 [1925]). Burgess suggested that Chicago included five concentric zones, each containing various neighborhoods, four of them situated around "The Loop" (the business center of the city): "the typical processes of the expansion of the city can best be illustrated, perhaps, by a series of concentric circles, which may be numbered to designate both the successive zones of urban extension and the types of areas differentiated in the process of expansion" (Burgess 1967 [1925], p. 50). Burgess' unit of analyses was a series of neighborhoods within cities that share similar characteristics. He assumed that depending on the distances to the center and the special features of these zones, the levels of crime would vary.

Clifford Shaw was one of the first Chicago sociologists to carry out extensive empirical research on the geographical distribution of crime on the basis of Burgess' zone model (Shaw et al. 1929). This study can be seen as a landmark in the history of geographic studies because of its detailed data collection, advanced methods, and innovative statistical tools. Based on the concentric zone model of Burgess, the authors studied the distribution of truancy of young people, juvenile delinquents, and adult offenders in Chicago. Assisted by young researchers like Henry McKay, Frederick Zorbaugh, and Leonard Cottrell, Shaw took natural areas as units of analyses (Abbott 1997) but in more detail than ever before in these kinds of studies. Shaw and his colleagues introduced new units of analyses. First, they introduced *spot maps* by plotting the home address of thousands of offenders on a map of Chicago. Second, he combined the offender address data with census data to created *delinquency rate maps* of square mile areas. And finally, he constructed *radial maps* and *zone maps*, which displayed delinquency rates at regular distances from the city center (Snodgrass 1976).

For further analyses on the distribution of crime across Chicago, Shaw et al. divided the city into 431 census tracts in 1910 and 499 in 1920 (Shaw et al. 1929). Each census tract included convenient age and sex groups of the population. Subsequently, these census tracts were combined into square mile areas with a minimum population of 500 residents. The technique was to allocate delinquents to their place of residence, and to divide the number of delinquents by the number of boys of juvenile court age, in order to compute rates for small areas. These rates were used by Shaw and his associates to make shaded maps or compute correlations.

In the same year, Shaw's research assistant Harvey Zorbaugh published his PhD in which he compared a slum neighborhood (the Lower North Side) with a wealthy area (Gold Coast) in Chicago, both situated in close proximity (Zorbaugh 1929). In this more qualitative study, Zorbaugh presented only a few maps, all of them less detailed in information than Shaw's study. However, his research demonstrated clearly that two areas in close physical proximity did not illustrate that physical and social distances coincide. As Park wrote in his foreword of Zorbaugh's study: "a situation in which people who live side by side are not, and – because of the divergence of their interests and their heritages – cannot, even with the best good will, become neighbors" (1929, p. ix), pointing to the invalid assumption of policy makers (and criminologists) of the time that people living in the same locality shared the same backgrounds and interests. This conclusion is very important for the choice of the unit of analysis in criminological research. This PhD study made explicit that administrative and political areas and social spaces are not identical. Depending on the size of the area, a variety of social communities with different identities can exist. Importantly, he concluded that the smaller the unit of analysis, the greater the chance of a homogeneous community.

In 1942, Clifford Shaw and Henry McKay published their magnum opus *Juvenile Delinquency and Urban Areas* in which they presented their geographical and etiological analyses of crime rates not only in the city of Chicago but also those of other cities: Philadelphia, Boston, Cincinnati, Cleveland, and Richmond. In principle, they used similar analytic tools as introduced in their study of 1929 (Shaw and McKay 1942). Again, they employed spot maps, rate maps, radial maps, and zone maps to illustrate concentrations of crime and offenders in the city of Chicago over a long period of time (up to 60 years). In all of the studied cities, they found similar patterns in the geographical distribution of crime. However, the units of analysis differed a good deal between the various cities. These differences were due to the lack of detailed official crime data in cities other than Chicago. The rapid changes in the city of Chicago over a long period of time enabled them also to study the effects of the dynamics of the city on crime and other phenomena. One of their findings was that: "The data on trends also demonstrate with equal sharpness the rapid rise in rates of delinquents in certain areas when a population with a different history and different institutions and values takes over areas in a very short period of time" (Shaw and McKay 1942, p. 382).

The Decline of Geographic Criminology

The Chicago studies inspired other criminologists to carry out empirical crime and place research in other cities (e.g., see Burgess and Bogue 1964a) like the relatively unknown policy report of Edwin Sutherland (1883–1950) on the geographical distribution of juvenile delinquency of the city of Bloomington, Indiana (Sutherland 1937). Also inspired by the work of Shaw and McKay and using the zone model of Burgess, Cyrill Burt (1883–1971) revealed there, like in Chicago, certain delinquent neighborhoods with high numbers of adult and juvenile offenders. Like Shaw and McKay, he used the location of the home addresses of delinquent boys and girls in the years 1922 and 1923 in London. Following the Chicago School findings, he noted that the highest concentrations in crime were found in three neighborhoods situated closely to the city center of London at the time: Holborn, Finsbury, and Shoreditch (Burt 1944 [1924]). According to Burt, these oldest but not poorest neighborhoods of London were for offenders of strategic importance, because they were situated closely to attractive crime targets in the inner city and – if necessary – they could function as a place to hide from the police.

On the other hand, as the decades passed, empirical and methodological critics of the Chicago approach began to emerge (Lander 1954). First, it was argued that Shaw (1929) and Shaw and McKay (1942/1969) could not distinguish between the dwelling place of the offender and the location where he or she committed a crime, neglecting the variability in the mobility of offenders. Secondly, by relying on official crime figures, their research was seen as biased because offenders of the lower class had (and still have) a greater chance to be processed in the criminal justice system (for instance, Beirne and Messerschmidt 1991; Chilton 1964; Chilton and Dussich 1974; Gordon 1967). Thirdly, delinquency rates after 1945 in Chicago did not conform to the distribution patterns of Shaw and McKay's early assumptions (Bursik 1984, 1986). European studies also showed contradicting results. Morris (1957) examined the offender rates of the county of Croydon (UK), but could not confirm the zone model of Burgess. Twenty years later, Morris' findings were replicated in the city of Sheffield (Baldwin 1975; Baldwin and Bottoms 1976). In Europe, the direct and indirect consequences of the operation of housing markets confounded the results of the geographical distribution of crime in American cities.

Another criticism that is key to our concern with units of analysis for crime place studies is that brought by Robinson (1950) who discussed the use of ecological correlations in geographical studies like that of Shaw and McKay (1942/1969). According to Robinson, the object of an ecological correlation is a group of persons, not a person: "…the individual correlation depends on the internal frequencies of the within-areas individual correlations, while the ecological correlation depends upon the marginal frequencies of the within-areas individual correlations" (Robinson 1950, p. 354). He concluded that ecological correlations cannot validly be used as substitutes for individual correlations. Such an ecological fallacy leads to meaningless conclusions. Looking back, these empirical and methodological critics diminished the attention of criminologists in studies of crime and place for almost 20 years.

Reemerging Interest in Communities and the Emergence of Study of Micro Crime Places

In the 1980s, Albert J. Reiss Jr. was to encourage a group of younger criminologists to return to the interests of the Chicago School where he had received his PhD in 1949. Reiss (1986) saw the criminological tradition as including two major theoretical positions, one that focused on individuals and a second that focused on crimes. Communities and crime was a main focus of the latter tradition and he sought to rekindle criminological interest in understanding variability of crime within and across communities. Editing an early volume in the *Crime and Justice* series, Reiss and Michael Tonry sought to bring *Communities and Crime* (1986) to the forefront of criminological interests (Reiss and Tonry 1986).

Reiss did not see the new interest as simply mimicking the insights of the Chicago School criminologists. Rather, he sought to raise a new set of questions about crime at place that had been ignored in earlier decades: "Recent work on communities and crime has turned to the observation that Shaw and McKay neglected: not only do communities change their structures over time but so often do their crime rates (…), a recognition that communities as well as individuals have crime careers" (Reiss 1986, p. 19). Many of the contributors to the Reiss and Tonry volume would become leaders of a new generation of criminologists, once again suggesting the important and enduring role of crime and place in advancing the criminological enterprise more generally. Among the contributors were Wes Skogan, Robert Sampson, Douglas Smith, Robert Bursik, Ralph Taylor, Stephen Gottfredson, and Lawrence Sherman.

This volume and other work developed in this period drew upon the identification of neighborhoods and communities to expand insights about the development of crime (Brantingham and Brantingham 1981, 1991; Bursik and Webb 1982; Clarke 1983; LeBeau 1987; Rengert 1980; Roncek and Bell 1981; Sampson 1985; Sampson and Groves 1989). Smith (1986), for example, identified neighborhood variation in the behavior of the police, suggesting the importance of place in understanding not only the etiology of crime but also the etiology of criminal justice. Skogan brought new insights not only to our understanding of the interaction of community characteristics and policing (Skogan 1987) but also more generally to the developmental processes that led to the emergence of crime and disorder in urban communities (Skogan 1990). More recently, scholars led by Robert Sampson have used a focus on the community to draw new insights into developmental crime patterns, arguing that social cohesion within communities and shared expectations of community members combine to affect both crime and social disorder (Raudenbush and Sampson 1999; Sampson 2010; Sampson et al. 1999; Sampson et al. 1997).

Consistent with Reiss' call for investigation of the criminal careers of communities, Bursik (1986; Bursik and Webb 1982) revisited crime in Chicago neighborhoods over time and challenged earlier views of the stability of crime within neighborhoods and communities, arguing that stability in crime patterns was a result of long-term stability in the social characteristics of places, and that instability in such patterns would also lead to instability in crime rates. Similarly, Schuerman and Kobrin (1986) identified stability and variability in criminal careers of places, focusing on the residences of juvenile delinquents as had Shaw and McKay (1942). Using the number of residential addresses of officially known delinquents by census tracts in Los Angeles as an indicator of aggregate crime, they found three general patterns that led to high crime rates in 1970. The first pattern they termed "emerging" and referred to those clusters that were relatively crime-free in 1950 but had moderate to high crime in 1960 and 1970, respectively. The second pattern, "transitional," refers to those clusters that had moderately high crime in 1950, a higher level in 1960, and an even higher level in 1970. The last pattern is referred to as "enduring" and refers to those clusters that had persistently high crime rates at all points in time. The vast majority of census tracts within the clusters were designated as having enduring crime rates over the time span, with fewer census tracts in the transitional and emerging categories.

Interestingly, though the approach of the Chicago School called for the identification of units of geography that would not be drawn from administrative data collection, but from the social units that defined neighborhoods or communities, this new generation of scholars concerned with communities and crime have generally used officially defined units for drawing their data and conclusions. In this case, the US Census definitions, most often census tracts or the smaller census block groups, have become the main source for defining the units of geography that are the focus of research in the USA, despite the fact that the goals of the census in creating physically contiguous geographic units are often inconsistent with the goals of community and crime researchers. Often, such studies will simply assume that census units such as census tracts reflect actual community boundaries (Hipp 2007), though some scholars in this area combine census units with the idea of creating boundaries of communities that are more consistent with the theoretical interests of researchers (e.g., see Sampson et al. 1997). Importantly, this new focus on communities and crime often led to study of much smaller geographic units of analysis than had drawn the interests of the early Chicago School scholars.

While a reemergence of interest in communities and crime had been one important source for renewed study of crime and place in recent decades, the 1980s produced a more radical reformulation of the unit of geography that should form the basis of crime place studies, continuing to push the unit of geographic analysis to a more micro level. Traditional criminological interest in place has focused on higher level geographic units such as regions, cities, communities, or neighborhoods. One reason for this focus on macro levels of geography is simply that data were often not available at geographic levels lower than the standard administrative or census divisions. But even when data were available, statistical and analytic tools were not readily available for linking crime easily with micro units of geography.

Certainly, the difficulty of mapping crimes to specific places and of analyzing geographic data were factors that prevented study of crime at micro units of geography, but another barrier was the lack of consistent theoretical interest in micro places as contrasted with research on individual criminality, or crime across macro geographic units (Weisburd et al. 2004; Weisburd and McEwen 1997). Such theoretical interest was not to emerge until the late 1970s and 1980s, about the time that computerized crime mapping and more sophisticated geographic statistical tools were to emerge (Weisburd and McEwen 1997). A new group of theorists challenged traditional criminological interests and began to focus more on the "processes operating at the moment of the crime's occurrence" (Birkbeck and LaFree 1993, p. 114). One influential critique that was to have strong influence on the development of interest in micro units of geography was brought by Lawrence Cohen and Marcus Felson (1979). They argued that the emphasis placed on individual motivation in criminological theory failed to recognize the importance of other elements of the crime equation. They argued that crime rates could be affected by changing the nature of targets or of guardianship, irrespective of the nature of criminal motivations. The "routine activities" perspective they presented established the spatial and temporal context of criminal events as an important focus of study.

Canadian criminologists Patricia Brantingham and Paul Brantingham (1993) made the connection between routine activities and place even more directly in their development of "crime pattern theory." Crime pattern theory focuses directly upon places by asking how targets come to the attention of offenders and how that influences the distribution of crime events over time and across places. Like Cohen and Felson, Brantingham and Brantingham routine human social and economic activities as a critical feature of the crime equation, but in this case, the place is made an explicit rather than implicit part of this equation, providing a "backcloth" for human behavior.

Drawing upon similar themes, British scholars led by Ronald Clarke began to explore the theoretical and practical possibilities of "situational crime prevention" in the 1980s (Clarke 1983, 1992; Cornish and Clarke 1986; Wikström et al. 1995). Their focus was on criminal contexts and the possibilities for reducing the opportunities for crime in very specific situations. Their approach turned traditional crime prevention theory on its head. At the center of their crime equation was opportunity. And they sought to change opportunity rather than reform offenders. In situational crime prevention, more often than not, "opportunity makes the thief" (Felson and Clarke 1998). This was in sharp contrast to the traditional view that the thief simply took advantage of a very large number of potential opportunities. Importantly, in a series of case studies, situational crime prevention advocates showed that reducing criminal opportunities in very specific contexts can lead to crime reduction and prevention (Clarke 1992; 1995).

One implication of these emerging perspectives is that micro crime places were an important focus of inquiry. Places in this "micro" context are specific locations within the larger social environments of communities and neighborhoods (Eck and Weisburd 1995). They are sometimes defined as buildings or addresses (e.g., see Green (Mazerolle) 1996; Sherman et al. 1989), sometimes as block faces, "hundred blocks," or street segments (e.g., see Taylor 1997; Weisburd et al. 2004), sometimes as clusters of addresses, block faces, or street segments (e.g., see Block et al. 1995; Sherman and Weisburd 1995; Weisburd and Green Mazerolle 1995). Research in this area began with attempts to identify the relationship between specific aspects of urban design (Jeffery 1971) or urban architecture and crime (Newman 1972) but broadened to take into account a much larger set of characteristics of physical space and criminal opportunity. In 1989, Sherman and colleagues coined the term the "criminology of place," to describe this new approach that drew its theoretical grounding from routine activities and situational crime prevention to emphasize the importance of micro crime places in the etiology of crime.

Recent studies point to the potential theoretical and practical benefits of focusing research on micro crime places. A number of studies, for example, suggest that there is a very significant clustering of crime at places, irrespective of the specific unit of analysis that is defined (Oberwittler and Wikström 2009; Roncek 1981; Weisburd et al. 2009). The extent of the concentration of crime at place is dramatic. In one of the pioneering studies in this area, Sherman, Gartin, and Buerger (1989) found that only three and a half percent of the addresses in Minneapolis produced 50 % of all calls to the police. Fifteen years later in a study in Seattle, Washington, Weisburd et al. (2004) reported that between 4% and 5% of street segments in the city accounted for 50 % of crime incidents for each year over 14 years.

Geographic criminology not only resurrected in the USA but also in Europe. In England, studies have been carried out in Glasgow, Liverpool, London, and Poplar to the spatial distribution of crime (Baldwin 1975). In Sheffield, Antony Bottoms established a research line on the geographic distribution of crime and victimization (Baldwin and Bottoms 1976), later with Paul Wiles (Bottoms 2007; Bottoms and Wiles 1992, 2002). At the University College London (former Jill Dando Institute) in London, spatial criminology is one of the main fields of research, and Ken Pease at the Loughborough University (Farrell et al. 2007). In the Netherlands, some studies have been undertaken in the past (for an overview, see Bruinsma 2007), and at the NSCR, geographic criminology is one of the main research themes. At Griffith University, Australia, a group of researchers study the spatial aspects of crime and guardianship. Leading centers in geography of crime are situated in Vancouver, Canada; George Mason University, Washington, USA; Temple University, Philadelphia, USA; University College of London (former Jill Dando), UK; the Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, the Netherlands; and at Griffith University in Brisbane, Australia. Many other universities and research institutes from all over the world are more or less active in the field of criminology that started almost 200 years ago in Belgium, France, and London.

Related Entries

▶ History of Geographic Criminology Part I: Nineteenth Century

Recommended Reading and References

Abbott A (1997) Of time and space: the contemporary relevance of the Chicago School. Soc Forces 75(4):1149–1182

Baldwin J (1975) British areal studies of crime: an assessment. Br J Criminol 15(3):211–227

Baldwin J, Bottoms AE (1976) The urban criminal. Tavistock Publications, London

Beirne P, Messerschmidt J (1991) Criminology. Harcourt Brace Jovanovich, San Diego

Birkbeck C, LaFree G (1993) The situational analysis of crime and deviance. Ann Rev Sociol 19:113–137

Block CB, Dabdoub M, Fregly S (eds) (1995) Crime analysis through computer mapping. Police Executive Research Forum, Washington

Bottoms AE (2007) Place, space, crime, and disorder. In: Maguire M, Morgan R, Reiner R (eds) The Oxford handbook of criminology. Oxford University Press, Oxford, pp 528–574

Bottoms AE, Wiles P (1992) Explanations of crime and place. In: Evans DJ, Fyfe NR, Herbert DT (eds) Crime, policing and place: essays in environmental criminology. Routledge, London, pp 11–35

Bottoms AE, Wiles P (2002) Environmental criminology. In: Maguire M, Morgan R, Reiner R (eds) The Oxford handbook of criminology. Oxford University Press, Oxford, pp 620–656

Brantingham PJ, Brantingham PL (eds) (1981) Environmental criminology. Sage, Beverly Hills

Brantingham PJ, Brantingham PL (eds) (1991) Environmental criminology. Waveland, Prospect Heights

Brantingham PL, Brantingham PJ (1993) Environment, routine, and situation: toward a pattern theory of crime. In: Clarke RV, Felson M (eds) Activity and rational choice. Advances in criminological theory, vol 5. Transaction, New Brunswick

Bruinsma GJN (2007) Urbanisation and urban crime. Dutch geographical and environmental research. In Tonry M, Bijleveld C (eds) Crime & justice in the Netherlands, vol 35. pp 453–502

Bulmer M (1984) The Chicago School of sociology. Institutionalization, diversity, and the rise of sociological research. University of Chicago Press, Chicago/London

Burgess EW (1967 [1925]) The growth of the city. An introduction to a research project. In: Park RE, Burgess EW (eds) The city: suggestions for the investigation of human behaviour in the urban environment. The University of Chicago Press, Chicago/London, pp 47–62

Burgess EW, Bogue DJ (1964a) The delinquency research of Clifford R. Shaw and Henry D. McKay and Associates. In: Burgess EW, Bogue DJ (eds) Contributions to urban sociology. The University of Chicago Press, Chicago/Londen, pp 591–615

Burgess EW, Bogue DJ (1964b) Research in urban society: a long view. In: Burgess EW, Bogue DJ (eds) Contributions to urban sociology. The University of Chicago Press, Chicago/Londen, pp 1–14

Bursik RJ (1984) Urban dynamics and ecological studies of delinquency. Soc Forces 63:393–413

Bursik RJ (1986) Ecological stability and the dynamics of delinquency. In Reiss AJ Jr, Tonry M (eds) Communities and crime. Crime and justice. A review of research, vol 8. pp 35–66

Bursik RJ, Webb J (1982) Community change and patterns of delinquency. Am J Sociol 88:24–42

Burt C (1944 [1924]) The young delinquent. University of London Press, London

Chilton RJ (1964) Continuity in delinquency area research: a comparison of studies for Baltimore, Detroit, and Indianapolis. Am Sociol Rev 29(1):71–83

Chilton RJ, Dussich JP (1974) Methodological issues in delinquency research: some alternative analysis of geographical distributed data. Soc Forces 53(1):73–82

Clarke RV (1983) Situational crime prevention: its theoretical basis and practical scope. In: Tonry M, Morris N (eds) Crime and justice: an annual review of research, vol 14. The University of Chicago Press, Chicago, pp 225–256

Clarke RV (ed) (1992) Situational crime prevention: successful case studies. Harrow and Heston, New York

Clarke RV (1995) Situational crime prevention. In: Tonry M, Farrington DF (eds) Building a safer society: strategic approaches to crime prevention Crime and justice: a review of research, vol 19. The University of Chicago Press, Chicago, pp 91–150

Cohen L, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44(4):588–608

Cornish DB, Clarke RV (eds) (1986) The reasoning criminal. Rational choice perspectives on offending. Springer, New York

Eck JE, Weisburd D (eds) (1995) Crime and place, vol 4. Criminal Justice Press, New York

Elmer MC (1933) Century-old ecological studies in France. Am J Sociol 39(1):63–70

Faris REL (1967) Chicago sociology 1920-1932. Chandler, San Francisco

Farrell G, Bowers KJ, Johnson SD, Townsley M (eds) (2007) Imagination for crime prevention: essays in honour of Ken Pease. Criminal Justice Press, Monsey

Felson M, Clarke RV (1998) Opportunity makes the thief: practical theory for crime prevention. Home Office, London

Gordon RA (1967) Issues in the ecological study of delinquency. Am Sociol Rev 32(6):927–944

Green (Mazerolle) L (1996) Policing places with drug problems. Sage, Thousand Oaks

Harvey L (1987) Myths of the Chicago School of sociology. Avebury, Hong Kong

Hipp JR (2007) Block, tract, and level of aggregation: neighborhood structure and crime and disorder as a case in point. Am Sociol Rev 72:659–680

Jeffery CR (1971) Crime prevention through environmental design. Sage, Beverly Hills

Lander B (1954) Toward an understanding of juvenile delinquency. Columbia University Press, New York

LeBeau JL (1987) The methods and measures of centrography and the spatial dynamics of rape. J Quant Criminol 3:125–141

Levin Y, Lindesmith A (1937) English ecology and criminology of the past century. J Crim Law Criminol 27:810–816

Morris T (1957) The criminal area. A study in social ecology. Routledge & Kegan Paul, London

Newman O (1972) Defensible space: crime prevention through urban design. Macmillan, New York

Oberwittler D, Wikström P-OH (2009) Why small is better: advancing the study of the role of behavioral contexts in crime causation. In: Weisburd D, Bernasco W, Bruinsma GJN (eds) Putting crime in its place: units of analysis in geographic criminology. Springer, New York, pp 35–59

Park RE, Burgess EW (1967 [1925]) The city. Suggestions for investigation of human behavior in the urban environment. The University of Chicago Press, Chicago

Raudenbush SW, Sampson RJ (1999) Ecometrics: towards a science of assessing ecological settings, with appliance to the systematic social observations of neighbourhoods. Sociol Methodol 29:1–41

Reiss AJ (1986) Why are communities important in understanding crime? In: Reiss AJ, Tonry M (eds) Communities and crime. Crime and justice: a review of research, vol 8. The University of Chicago Press, Chicago, pp 1–34

Reiss AJ, Tonry M (eds) (1986) Communities and crime, vol 8. The University of Chicago Press, Chicago

Rengert GF (1980) Spatial aspects of criminal behavior. In: Georges-Abeyie DE, Harries KD (eds) Crime: a spatial perspective. Columbia University Press, New York, pp 47–57

Robinson WS (1950) Ecological correlations and the behavior of individuals. Am Sociol Rev 15(3):351–357

Roncek DW (1981) Dangerous places: crime and residential environment. Soc Forces 60(1):74–96

Roncek DW, Bell R (1981) Bars, blocks, and crimes. J Environ Syst 11(1):35–47

Sampson RJ (1985) Neighborhood and crime: the structural determinants of personal victimization. J Res Crime Delinq 22(1):7–40

Sampson RJ (2010) Collective efficacy theory. In: Cullen FT, Wilcox P (eds) Encyclopedia of criminological theory. Sage, Thousands Oaks, pp 802–812

Sampson RJ, Groves WB (1989) Community structure and crime: testing social -disorganization theory. Am J Sociol 94(4):774–802

Sampson RJ, Raudenbush SW, Earls F (1997) Neighbourhoods and violent crime: a multilevel study of collective efficacy. Science 277:918–924

Sampson RJ, Morenoff JD, Earls F (1999) Beyond social capital: spatial dynamics of collective efficacy for children. Am Sociol Rev 64(5):633–660

Schuerman L, Kobrin S (1986) Community careers in crime. In: Reiss AJ, Tonry M (eds) Communities and crime, vol 8. University of Chicago Press, Chicago, pp 67–100

Shaw CR, McKay HD (1942) Juvenile delinquency and urban areas. University of Chicago Press, Chicago

Shaw CR, Zorbaugh FM, McKay HD, Cottrell LS (1929) Delinquency areas: a study of the geographic distribution of school truants, juvenile delinquents, and adult offenders in Chicago. University of Chicago Press, Chicago

Sherman LW, Bueger ME, Gartin PR (1989) Repeat call address policing: the Minneapolis RECAP experiment. Crime Control Institute, Washington, DC

Sherman LW, Weisburd D (1995) General deterrent effects of police patrol in crime ‘hot spots’: A randomized, controlled trial. Justice Q 12(4):625–648

Skogan WG (1987) Disorder and community decline. Center for Urban Affairs and Policy, Northwestern University, Research

Skogan WG (1990) Disorder and decline: crime and the spiral of decay in American neighborhoods. Free Press, New York

Smith DA (1986) The neighborhood context of police behavior. In: Reiss AJJ, Tonry M (eds) Communities and crime. Crime and justice: a review of research, vol 8. The University of Chicago Press, Chicago, pp 313–341

Snodgrass JD (1976) Clifford R. Shaw and Henry D. McKay: Chicago criminologists. Br J Criminol 16:1–19

Sutherland EH (1937) Report on ecological survey of crime and delinquency in Bloomington, Indiana: U.S. National Youth Administration & Indiana University: Institute of Criminal Law and Criminology

Taylor RB (1997) Social order and disorder of street blocks and neighborhoods: ecology, micro ecology, and the systemic model of social disorganization. J Res Crime Delinq 34(1):113–155

Thomas WI (1966) On social organization and social personality. Selected papers (ed and with an introduction by Janovitz M). The University of Chicago Press, Chicago/London

Vold GB, Bernard TJ, Snipes JB (2002) Theoretical criminology, 5th edn. Oxford University Press, New York

Weisburd D, Green Mazerolle L (1995) Policing drug hot spots: the Jersey City DMA experiment. Justice Q 12(4):711–735

Weisburd D, McEwen T (1997) Introduction: crime mapping and crime prevention. In: Weisburd D, McEwen T (eds) Crime mapping and crime prevention. Criminal Justice Press, Monsey, pp 1–23

Weisburd D, Bushway S, Lum C, Yang S-M (2004) - Trajectories of crime at places: a longitudinal study of street segments in the city of Seattle. Criminology 42(2):283–321

Weisburd D, Bernasco W, Bruinsma GJN (eds) (2009) Putting crime in its place: units of analysis in geographic criminology. Springer, New York

Wikström P-OH, Clarke RV, McCord J (eds) (1995) Integrating crime prevention strategies: propensity and opportunity. Stockholm National Council for Crime Prevention, Sweden

Zorbaugh HW (1929) The gold coast and the slum. A sociological study of Chicago’s near north side. The University of Chicago Press, Chicago

History of Geographical Criminology

► History of Criminological Theories: Causes of Crime

History of Green Criminology

Nigel South[1], Katja Eman[2] and Gorazd Meško[2]
[1]Department of Sociology, University of Essex, Colchester, Essex, UK
[2]Faculty of Criminal Justice and Security, University of Maribor, Maribor, Slovenia

Synonyms

Eco-crime; Environmental crime

Overview

This entry provides an overview of work in the developing field of "green criminology" (other terms and definitions will be described below). It discusses the breadth of topics covered in the literature, reviews some key contributions from writers in the field, and draws attention to some future directions for research.

Global extensions of environmental degradation are slowly but persistently breaking into every sphere of human life. Throughout the twentieth century and particularly following changes brought about by the First World War and accelerating after the Second, advanced societies have experienced complexity in their progress and development that has shaped changes in socioeconomic systems and in corresponding and linked criminal activities. Within this process, impacts on the environment have tended to be overlooked until relatively recently. Across the natural and social sciences, the preservation and sustainability of our environment is now recognized as a major issue for the survival of humanity and all species, and within criminology, it is now also being recognized that concerns about environmental harms, crimes, and damage should be given a more prominent place in the field. The consequences of environmental pollution, for example, are dramatic and drastic, including the extinction of plants and animals, poor health outcomes for some population groups, damage to food chains, depletion of natural resources, exacerbation of climate change, and natural disasters stimulated or worsened by human action. Environmental harms have become a topic of public discussion and frequent media reporting, as well as the subject of international conferences and scientific symposia, debates, and meetings of political leaders. In this context, a relevant branch of criminology has emerged and has been referred to as a "green criminology."

Fundamentals

"Green criminology" has grown internationally and is a good example of how the English-language research literature can fail to register and acknowledge pioneering work published in other languages. Opinions about the beginnings and formation of an environmentally sensitive criminology differ. Clifford (1998: 10) suggests environmental criminality is a relatively new field of research within criminology, beginning in the period around the 1970s, while Koser Wilson (1999: 155) states that criminologists showed an interest in environmental crime at a point when the momentum of environmental politics brought such crime within the orbit of criminal sanctions as a part of the criminal-legal system. Pioneering work from Slovenia by Pečar (1981) put forward early warnings about new ecologically damaging forms of criminality (Eman et al. 2009: 584) and offered an etiology of forms of environmental criminality in Slovenia, defining the role of criminology and sciences related to this and setting out tasks necessary for the control of offending institutions. However, despite his forecasts and prescriptions, there was little subsequent interest in such a green criminology in Slovenia, and with no English-language translation, this entry made no wider international impact. This is one of the "moments" when a green criminology was "born." Pečar's work reflected emerging global interest in the green agenda, and unsurprisingly other writers in other countries also began to contribute, quite independently of each other, to such "moments" in the birth of a green criminology.

Pečar (1981) argued that criminology has a double role to play in the case of deviance against the environment: criminology explains the phenomena that have already been criminalized and against which the society reacts, and criminology also responds to social phenomena and events that are becoming much more important than traditionally considered. Criminology can make valuable proposals to society about appropriate reactions. From the point of view of this critical perspective, it follows that traditional criminality is not the most dangerous social deviance, because, as Kanduč (1992) wrote, ecological, political, and enterprise criminality are far more dangerous. A green perspective fits with a radical turn in criminology that focuses more

History of Green Criminology, Table 1 Timeline of environmental harms and degradation and development of a green criminology

1760–1945	*Industrialization, urbanization, population expansion, and colonization* establish momentum of pollution of air, water, and soil, and loss of huge woodland areas; first extinction of animal and plant species due to human actions and signs of global temperature change; pro-environmental sentiments become movements and expand, affinities with liberal, left, and Marxist viewpoints
1900–	*Development of environmental protection law*
	The end of nineteenth and twentieth century in Europe
	Between the First and the Second World War in the United States of America
1970–	The *beginnings of sociological research* on threats to the environment and the early origins of an environmental criminology. At the end of 1960s, critical criminology warns about threat to a healthy natural environment, for example, via work on corporate environmental criminality
	The *beginnings of environmental justice via* the Warren County case in North Carolina
1990	*Forming of green criminology*
	Lynch (1990) first uses the term *green criminology*, which becomes the synonym for an environmental criminology concerned with the natural world. Later terms such as conservation criminology and eco-critical criminology emerge
1998	*Growth of scientific publications in green criminology and of international network of criminologists*
	Beirne and South (1998) is first collection of essays on such green issues (e.g., environmental justice, animal abuse, green criminology, crimes against the environment) in a special issue of Theoretical Criminology. Halsey and White (1998) write on both sociological and philosophical viewpoints concerning threats to the environment and the relationship between humans and nature. Henceforth, international cooperation and development of green criminology at the national and international level begins
2004	Wider interest in environmental issues in the social sciences, especially criminology, in Serbia and Former Yugoslav Republic of Macedonia
	Halsey (2004) initiates a debate *pro* and *against* "green" criminology, based on discussion of Lynch and Stretesky (2003) about "the meaning of green" from various criminological perspectives
2006	Introduction of conservation criminology by Herbig and Joubert (2006)
2008–2010	New development of criminological investigation of environmental criminality in Slovenia and first scientific publications in the field of environmental criminology
2008	Introduction of ecoglobal criminology by White (2008)
2009	A beginning of development of green criminology in Bosnia and Herzegovina
	Beirne (2009) develops animal rights and protection as a green criminology interest field
2010	Introduction of the conservation criminology framework by Gibbs et al. (2010)
2011–2012	Wider interest in transnational environmental crime (White 2011; Ellefson et al. 2012)

Summarized after Rock (1994), Eman and Meško (2009)

on sociological and economic conditions and less on individuals and their characteristics.

Lynch (1990: 4) first introduced the term "green criminology" to criminological discussions as a new subfield that nonetheless, empirically and theoretically, still reflects existing traditions while fostering emergent directions. Lynch (1990: 3) acknowledged that a green criminology does not represent "an entirely new perspective or orientation within criminology" for "a number of criminologists have examined various environmental hazards and crimes," and studies of white-collar and corporate crime have made significant contributions in this regard (South and Beirne 2006). Nonetheless, the power and utility of a green umbrella or perspective (South 1998a) lies in loosely covering or unifying the study of environmentally related harms and crimes. As examples of what might be studied, Lynch suggested a non-exhaustive list as follows:

> environmental and wildlife laws and regulations; . . . the social harms associated with chemical and pesticide manufacturing on a local or global scale;. . . international treaties devoted to environmental protection; . . . environmental politics and power . . .; . . . unsafe working conditions and hazards created by pesticides, both in the field and in the factory; global political and economic structures responsible for the exportation of

> environmental hazards from industrialised, core nations to the periphery etc.; and [the] political, economic and class relationships that structure these outcomes (*ibid*.).

In addition, and pioneered in particular by Beirne (1995), the green criminology approach has also encompassed the increasing study of animal abuse, speciesism, and crime or harms against or involving animals (e.g., wildlife trafficking, see South and Wyatt 2011; Sollund this volume). Included here are diverse human actions that may cause suffering or death to nonhuman species or damage or destroy their habitats. For some writers, a wide range of animal welfare matters is at stake, and abuse of nonhuman species could entail physical, psychological, and/or emotional actions and harms, or indeed inaction or passive neglect (Cazaux and Beirne 2006). Varieties of environmental degradation and threat damaging to human populations (e.g., air and water pollution, soil erosion, oil spills, climate change) are, of course, also likely to be equally or even more harmful to animals and their habitats.

All this gives a sense of the subject matter content of a green criminology, but what of the conceptual approach that characterizes green criminology? Eman et al. (2009) have written of the development of green criminology research in Slovenia and South-Eastern Europe, noting along the way the challenges of definition and theory that arise. Thus, they observe "Different irregularities, indecencies, violations and crimes against nature are addressed among . . . criminologists as green crimes" (2009: 581) but then note that it has been hard to identify or develop a single theoretical approach in this area.

> The theoretical frame of green criminology is hard to define. White (2008: 14) is clear and brief: "There is no green criminological theory as such." Closest to [a] theory is South (1998a: 212), although he talks only about the perspective or ecological viewpoints and not about theory. . . . Clear definitions are rarely presented making successful research into green crimes more difficult. Halsey (2004) criticises green criminology because of its lack of a (suitable) definition and challenges criminologists to reduce ambiguity with a clear definition of green criminology (Eman et al. 2009: 581, footnote 9).

In her Presidential speech to the American Society of Criminology in 1998, Zahn (1999) wrote of the far-reaching impacts of pollution and biodiversity loss and indicated that in the future she expected to see more criminological focus on environmental crime, this in turn would bring "a new definition of victims to include species other than humans and a definition of offenders to include those who pollute for convenience . . . [and] for profit." Importantly, Zahn observed that "Just as Sutherland's white-collar crime expanded our crime paradigm (1949), . . . environmental crime will change it in the future." (Zahn 1999). South's (1998b: 445–453) typology reflects both new directions as well as classic core concerns in criminology such as breaching laws, making laws, and reactions to lawbreaking and law bending. Thus South proposed three foundations for a green criminology:

1. Studies of breaches of regulations, disasters, and corporate and state violations

   The themes of the relevant literature here concern (a) the study of regulation, for example, the positive and negative features, and the consequences, of different regulatory models; (b) pollution, disasters, and liabilities as "single-event" case studies; and (c) corporate and state misconduct or crimes which have environmental consequences, with the focus bearing on the perpetrators, culpability, and the nature of serial offending.

2. Legal frameworks, criminalization, and "shaming"

   Is concerned with the legal issues raised by environmentally damaging acts and how these should be classified (violations? crimes?) and responded to (by regulation? criminalization? Inspectorates? police?). Such cases may be difficult to prosecute because of lack of evidence and proof, blurring of the lines between the "willful criminal violator" and the "legal risk-taking entrepreneur," and/or corruption in the system. The problem of how to respond to such offenses in an effective legal manner through criminalization has been subject to debate and has embraced arguments against criminalization as well as powerful arguments in favor. Even if a prosecution is brought and is successful, penalties are usually modest relative to the damage done: if the corporation is fined, it will absorb such costs and/or simply pass them onto consumers. Attempts to identify and sanction key, responsible individuals have had only rare success and alternatives such as Braithwaite's (1989) notion of "shaming," and approaches based on principles of restorative justice seem promising based on the proposition that corporate image is a more vulnerable target for censure and sanctions than corporate assets and that "making good" may be more beneficial than imposing immaterial fines or trying to identify a "culprit" to receive a prison sentence.

3. Social movements, "green politics," and policy futures

   Numerous political and pressure groups debate, champion, question, and dispute matters related to the environment. These include anti-environmentalist organizations supported by corporate and state interests and pro-environment groups which may be extreme (e.g., employing terrorism) or controversial (championing animal rights as more important than human rights), as well as those engaging in "new social movements" and alternative lifestyles, or those of the middle-range political left, center, and right who support campaigns to "protect and preserve" and resist construction development projects. The importance of the feminist critique of masculine violence against the environment and, in the USA, the emergence of networks of black activists working against environmental damage to their communities (Bullard 1990) are expressions of protest that should also be noted.

Most "green" criminologists would probably agree that it is now quite clear what kind of subject matter is relevant but might also debate whether it would be possible or even desirable to attempt to impose a single theoretical proposition or framework on this area of work. There is therefore some variety in the positions outlined in the literature. Lynch and Stretesky (2007) refer to an eco-critical criminology as a contrast with the more familiar critical criminological approach that excludes nonhuman nature from its analysis and also (Lynch and Stretesky 2003: 218) "take the view that green crimes, like other crimes, are social constructions influenced by social locations and power relations in society." White (2003: 484) takes this point further, arguing that:

> investigating environmental issues from a criminological perspective requires an appreciation of how harm is socially and historically constructed. In turn this necessitates understanding and interpreting the structure of a globalising world; the direction(s) in which this world is heading; and how diverse groups' experiences are shaped by wider social, political and economic processes. Thus this area of criminology is at once basic and exceedingly complex.

More recently, White (2010: 6) refers to ecoglobal criminology as a framework or paradigm within criminology "that is informed by ecological considerations and by a critical analysis that is worldwide in its scale and perspective." The research agenda of an ecoglobal multidisciplinary approach can be used for the study of environmental harm. Furthermore, ecoglobal criminology defines different forms of environmental harm as "criminal," though they may not be considered illegal in conventional terms.

Herbig and Joubert (2006) have proposed the idea of a conservation criminology. Conservation criminology is concerned with environmental protection, adopting interdisciplinary and multidisciplinary research approaches to understanding environmental criminality, environmental threats, and related risks. This approach therefore offers the integration of criminology with natural resource disciplines, other social science fields, and risk and decision sciences. It encourages the examination of political, cultural, economic, and social influences on the definition of environmental crime. The parameters of conservation crimes can be seen as falling within existing and commonly used crime categories such as white-collar crime, organized crime, invisible crime, and property (economic) crime. Gibbs and colleagues (2010: 125–126) concur with Herbig and Joubert (2006) that conservation criminology identifies core themes in the environmental field of study and suggest that the focus of conservation criminology begins with assessment and study of environmental risks at the nexus between humans and natural resources that involve issues of crime, compliance, and social control (Gibbs et al. 2010: 128). Conservation criminology encompasses the study of wildlife, pollution, human populations, and other areas often considered to be mutually exclusive.

State of the Art

A green perspective in criminology offers an open framework rather than ties to a closed theory and lends itself well to interdisciplinary insights and collaborations. Research covers subjects such as environmental crimes and criminality, harm and victimization, legislation and regulations, protection measures, and public responses to violations. Green criminology is much more than just a discussion about environmental issues and reflects a critical criminological position on the need to defend the environment and to uphold the rights and safety of both humans and nonhuman species. The research agenda for green criminology is based on the objective of studying known forms (and uncovering unknown forms) of deviant and criminal behavior committed against the natural environment and those living beings dependent upon it.

Environmental crime is a now widespread problem across modern societies and may be defined as crime in formal terms where temporary or permanent acts or activities damaging the environment and species are determined and defined as criminal by (inter)national legislation. However, there are numerous forms of harm which lie outside the formal labels of judicially defined offenses but which have immediate or long-term consequences of a nature harmful or prejudicial to well-being, health, safety, and life. For example, artificial changes, degeneration, or destruction affecting air, water, soil, mineral materials, plant species, bacteria, and viruses can harm the natural environment or interrupt its natural changes with consequent and interactive impacts on human and animal species. These latter events may not be defined technically as crimes but are serious in their harmful outcomes and for this reason are considered within the framework of a green criminology. In any of these cases, the violator could be any or every one of us (corporations, companies, groups, individuals, etc.). Environmental crimes have multiple victims because besides, or directly through, the affected environment (biotic and abiotic natural elements), human and other species are harmed as well. However, as with many other forms of deviance and damage, not all are legally defined as crime but are nonetheless viewed as significant harms deserving censure, control, and remedy.

Basic human conditions for existence are essential and include air, water, and fertile soil. For a green critical criminology, it is obvious that these are resources to which all must have access and that fall within the notion of universal rights. Denial of such access represents the point of conflict between environmental rights and property rights. In this critical view, both the wealthiest individuals (or members of the wealthiest social class) and the wealthiest nations are guilty of overconsumption of natural resources and hence are the most significant contributors to environmental degradation and ultimately the impoverishment of other people and the planet. A rights-based approach to addressing such inequalities is complex but at the least might promote a substantive right to a healthy and pollution-free environment (Brisman this volume).

Climate change is expected to be a significant criminogenic force in future decades creating shortages of food and water that may lead to illegal markets and conflicts. In times and locations of impoverishment, human populations are more ready to exploit the resources of the natural world without regard for consequences and conservation, and they may also be more likely to move to other places that are known or believed to be more abundantly resourced. Both voluntary or forced migration and displacement can provide solutions but also simply lead to the reproduction of the original circumstances as money, labor market opportunities, or other resources prove to be unavailable. New cycles of vulnerability, victimization, offending, and perhaps yet further migration can be set in motion (South 2010; White 2008).

As well as having a criminological foundation, work in this area arises from socio-legal studies and environmental law (see, e.g., Farrall et al. 2012). However, Stallworthy (2008: 4–5) has argued that "...there can be a sense that environmental law discourse is ultimately shackled by a dependent, satellite status, a repository of greener values, but for the most part swimming against a distinctively ungreen tide of prevailing legal priorities." Obviously, this is not the end of the matter, and there can be few other areas of law with such a claim to significance for the future of the planet and its inhabitants. There is considerable scope to broaden legal discourse to incorporate "long unasked questions as to the ecosystem and biodiversity protection, as well as appropriate conditions for access and use of natural resources" (ibid: 3). This "broadening of the discourse" is undoubtedly happening. Today, international environmental policy is at the forefront of many progressive developments prefiguring more general trends in public international law. For more than a century, nation-states have adopted international environmental instruments that aim to have the force of law. They are of various types: global, multilateral, regional, and bilateral treaties; court-made and customary law; and soft law; and vary in the extent to which they affect national sovereignty and establish jurisdiction (DiMento 2003: 56). International environmental protection law is constantly changing with the high number of accepted international legal instruments (about 300 at the end of 2010) clearly pointing to the seriousness of the problem of violations of environmental protection. Attempts at meetings (in Rio de Janeiro, The Hague, Nairobi, Stockholm, Montreal, Kyoto, Buenos Aires, Johannesburg, and Copenhagen) to construct an international law of the environment have witnessed an intense and active but often confusing drama. DiMento (2003) questions whether these sessions, involving very different voices from high state representatives to indigenous people to NGOs and observers, can actually produce commonly agreed and understood effective law.

Methodologically, both qualitative and quantitative social science methods are appropriate and employed. Croall (2011: 360) notes that "as for corporate crime, there is no centralized source of information about environmental crime, and a wide range of sources, most from outside criminology, have to be used. Offences rarely appear in victim surveys, not being considered as 'criminal', and often being invisible. Like corporate or state crime, research is further hampered by difficulties of access and obtaining information, much of it regarded as confidential, as it is not widely regarded as part of the criminological agenda, funding tends to be scarce." Nonetheless, research in some areas such as environmental justice and environmental victimology does lend itself to survey methods but also to sensitive interview methods and to investigative data gathering, analyzing news coverage, or scientific reports, for example. In this respect, some green criminology studies lie close to epidemiological public health research and support the development and use of science in the interests of those who are victimized but frequently excluded from recourse to redress or protest. The victimized are paradoxically often excluded from debates and decisions most affecting them because greater economic power can mobilize expert opinion and promote particular positions. A green criminology can play an advocacy role using science and data on behalf of the victimized to tell the stories otherwise ignored. It is also felt that the use of a comparative approach and methods should be particularly fruitful and informative within green criminology. The use of comparative methods in criminology and criminal justice studies is necessary to ascertain social, cultural, political, economic, and other impacts on any differences in patterns of offending and attitudes towards environmental crime, in environmental protection laws, and in law enforcement responses to the violation of laws. For this reason, comparative studies should be more frequently used in comparing forms of environmental crime and environmental justice responses between two or more countries. However, it will always remain true that the identification of the particularities and specifics of each country must not be ignored and may require separate treatment. Heckenberg and White have provided valuable overviews of research methodology appropriate to this field and argue that the different dimensions of environmental harm need to be broken down as these "pose particular challenges for researchers insofar as different types of knowledge are required for dealing with specific kinds of environmental harm. Moreover, analysis needs to take into account considerable diversity in terms of: Who the victim is (human or non-human); Where the harm is manifest (local through to global); Main site where the harm is apparent (built or natural environment)" (Heckenberg and White 2013). The authors also advocate the compilation of data from across a wide range of possible and available sources, international, interdisciplinary, and cross-culturally, as the basis for analysis of contemporary issues but also for purposes of extrapolation and horizon scanning to identify future challenges and crises.

Key Controversies and Open Questions

At present, it seems that late-modern societies remain set on a path of continuing behavior seriously detrimental to the environment and, according to some commentators, worthy of description as "ecocide" and requiring legally recognized instruments of protection and response (Higgins 2010; South 2010). Social and economic conflicts arising from resource scarcity or exploitation, as well as climate change phenomena, present major challenges for human rights, civil society, and species preservation in many parts of the world. Climate change will be both criminogenic and a stimulus to control and security measures. The global flow of people and goods is intimately bound up with impact on the environment, and there is scope for protest, both peaceful and violent, about, for example, unnecessary transport, production of waste, and inequalities in access to food. Future conflicts and crimes may increasingly center on environmental resources, with global warming and accumulating resource poverty raising the temperature of existing tensions and accelerating processes of exploitation, whether between social classes, different ethnic groups, or regions and states competing for or claiming ownership of resources. Conflict and crime are likely to follow if legally sanctioned, but politically unpopular – as well as illegal – migration is stimulated by climate change-induced impacts on living conditions and the sustainability of agriculture, water availability, or other vulnerable systems. In such circumstances, the balance of rights such as freedom of movement versus measures that impose restrictions on mobility will come under scrutiny, and nationalist politics and national security preoccupations could lead to the creation of labels of "criminal" and "unwelcome threat" applied to those who should be seen as "victims" and "refugees."

Future Directions

The extension of this work in countries which have similar or very different profiles emphasizes one of the fundamental requirements for a future global green criminology – a program of comparative work. A research program to help reduce environmentally damaging activities and hence consequences could form the basis for a prospectus for green–crime-prevention research as well as propositions for remedial actions and responsive policies. In all of this, it has to be recognized that variations in definitions of offenses and differences of approach between jurisdictions mean that not only is enforcement hampered but data gathering and comparison are difficult – a further reason why we need comparative work and some harmonization of understanding and approach in areas of law and enforcement. Environmental issues affect different countries in different ways. Creating and enforcing national and international laws can be a very difficult process (Michalowski 1998: 328), particularly when dealing with environmental destruction and degradation issues. Environmental crime has no regard for national borders, so such problems often cannot be solved within the scope of national powers or policies. In cases such as safeguarding of international waters, seas, and air, or international transport by planes (noise and emissions), or transportation of hazardous materials and waste, or destruction of forests (acid rain), then interstate or international law is necessary. Such transnational action should lead to the determination of (international) standards which support the objectives of achieving environmental protection, and relevant research will be required (Meško et al. 2011).

Increasing efforts to regulate and control could create new forms of criminal offense or at the least widen the embrace of those found guilty of breaches of expectations, rules, or local laws. Insofar as this is part of a process generally set in train to protect the common interest and defend victims of uncontrolled and irresponsible actions, the principles are not dissimilar to the traditional operation of criminal justice; however, as much that needs to be regulated or curtailed is currently acceptable and legal behavior, then new groups of offenders and victims may be created. This may occur in relation to, for example, controls to mitigate climate change or requirements to conserve energy or water. This set of circumstances will create a set of issues for future research.

As criminality changes through time, so does criminology. The discovery of patterns of environmental crime in any region and the sharing of findings about similar green issues will enable cooperation between countries and support their responses to environmental issues.

Conclusion

In the twenty-first century, a relevant criminology needs to be aware of the meaning of the environment for the survival of the planet and its inhabitants (South 1998a; Eman et al. 2009: 581), and therefore, its intellectual breadth must accommodate crimes and harms relating to the interdependencies between environments, humans, and animals. Human beings share the planet with other animals, plants, and living beings (such as microorganisms and microbes) in a natural cycle, without which it would be impossible to survive (Eman et al. 2009: 581). The development of a green (or ecoglobal or conservation) criminology as a perspective or approach has been a product of a period when environmental issues have become more prominent on public and political as well as scientific agendas. Such a new direction in criminology has emerged independently in different countries but with a broadly shared and similar main objective – to study and critique crimes and harms against the environment on a regional, national, or global level. As this green field grows, one of its main strengths is the increasing international joining together and networking of these various branches.

Related Entries

- ▶ Environmental and Human Rights
- ▶ Environmental Regulation and Law Enforcement
- ▶ Green Criminology
- ▶ Organized Crime and the Environment

Recommended Reading and References

Beirne P (1995) The use and abuse of animals in criminology: a brief history and current review. Soc Justice 22:5–31

Beirne P (2009) Confronting animal abuse: law, criminology, and human-animal relationship. Rowman & Littlefield Publisher, Maryland

Beirne P, South N (1998) Editor's introduction. Theor Criminol 2:147–148

Beirne P, South N (2007) Issues in green criminology: confronting harms against environments, humanity and other animals. Willan Publishing, Portland

Braithwaite J (1989) Crime, shame and reintegration. Cambridge University Press, Cambridge

Bullard R (1990) Dumping in Dixie: race, class and environmental quality. Westview Press, Boulder

Cazaux G, Beirne P (2006) Animal abuse. In: McLaughlin E, Mucie J (eds) The Sage dictionary of criminology. Sage, London

Clifford M (1998) Environmental crime: enforcement, policy, and social responsibility. Aspen Publisher, Gaithersburg

Croall H (2011) Crime and society in Britain. (2nd edn), Harlow: Longman Pearson

DiMento JF (2003) The global environmental and international law. University of Texas, Austin

Ellefson R, Larsen G, Sollund R (2012) Eco-global crimes: contemporary ad future challenges. Ashgate, Farnham

Eman K, Meško G (2009) Ekološka kriminologija – poskus opredelitve in razprava. Varstvoslovje 11:357–381

Eman K, Meško G, Fields CB (2009) Crimes against the environment: green criminology and research challenges in Slovenia. Varstvoslovje 11:574–592

Farrall S, French D, Ahmed T (2012) Climate change: legal and criminological implications. Hart, Oxford

Gibbs C, Gore ML, McGarrell EF, Rivers L (2010) Introducing conservation criminology: towards interdisciplinary scholarship on environmental crimes and risks. Br J Criminol 50:124–144

Halsey M (2004) Against 'green' criminology. Br J Criminol 44:833–853

Halsey M, White RD (1998) Crime, ecophilosophy and environmental harm. Theor Criminol 2:345–371

Heckenberg D, White R (2013) Innovative approaches to researching environmental crime. In: South N, Brisman A (eds) The international handbook of green criminology. Routledge, London

Herbig FJ, Joubert SJ (2006) Criminological semantics: conservation criminology – vision or vagary? Acta Criminol 19:88–103

Higgins P (2010) Eradicating ecocide: laws and governance to prevent the destruction of our planet. Shepheard-Walwyn, London

Kanduč Z (1992) Radikalna kriminologija – kriminologija vsakdanjega življenja. Revija za kriminalistiko in kriminologijo 43:103–115

Koser Wilson N (1999) Eco-critical criminology – an introduction. Crim Justice Policy Rev 10:155–160

Lynch M (1990) The greening of criminology: a perspective for the 1990's. Crit Criminol 2:11–12

Lynch ML, Stretesky P (2007) Green criminology in the United States. In: Beirne P, South N (eds) Issues in green criminology: confronting harms against environments, humanity and other animals. Willan Publishing, Portland

Lynch MJ, Stretesky PB (2003) The meaning of green: contrasting criminological perspectives. Theoretical Criminology 7(2):217–38

Meško G, Dimitrijević D, Fields CB (2011) Understanding and managing threats to the environment in South Eastern Europe. Springer, Dordrecht

Michalowski R (1998) International environmental issues. In: Clifford M (ed) Environmental crime: enforcement, policy, and social responsibility. Aspen Publisher, Gaithersburg

Pečar J (1981) Ekološka kriminaliteta in kriminologija. Revija za kriminalistiko in kriminologijo 34:33–45

Rock P (1994) History of criminology. Dartmouth: Aldershot

South N (1998a) A green field for criminology?: a proposal for a perspective. Theor Criminol 2:211–233

South N (1998b) Corporate and state crimes against the environment: foundations for a green perspective in European criminology. In: Ruggiero V, South N, Taylor I (eds) The new European criminology. Routledge, London/New York

South N (2010) The ecocidal tendencies of late modernity: trans-national crime, social exclusion, victims and rights. In: White R (ed) Global environmental harm: criminological perspectives. Willan Publishing, Cullompton

South N, Beirne P (2006) Green criminology. Dartmouth, Aldershot/Brookfield

South N, Wyatt T (2011) Comparing illicit trades in wildlife and drugs: an exploratory study. Deviant Behav 32:1–24

Stallworthy M (2008) Understanding environmental law. Sweet & Maxwell, London

White RD (2003) Environmental issues and the criminological imagination. Theor Criminol 7:483–506

White RD (2008) Crimes against nature: environmental criminology and ecological justice. Willan Publishing, Cullompton

White RD (2010) Prosecution and sentencing in relation to environmental crime: recent socio-legal developments. Crime Law Soc Changes 53:365–381

White RD (2011) Transnational environmental crime. Willan Publishing, Cullompton

Zahn M (1999) Presidential address – thoughts on the future of criminology. Criminology 37:1–16

History of Juvenile Courts

Megan C. Kurlychek
School of Criminal Justice, University at Albany, State University of New York, Albany, NY, USA

Overview

The juvenile court of today, often referred to as a quasi-criminal court, is seen primarily as a slightly less severe version of the adult criminal court. It exists to hold youth accountable for offenses committed while still offering some semblance of rehabilitation or treatment. Like adults, we speak of these juveniles as being arrested and sentenced. However, their sentences are typically less onerous than those imposed on more mature and culpable adult offenders by the criminal court. In reality, however, there are many more differences between the two systems of justice than a mere mitigation of punishment, not the least of which are their underlying purposes and origins. Indeed the seeds of the juvenile court were sown in social welfare and not the criminal justice system. The catalyst of its creation was not to merely modify the adult court, but rather to provide an entirely separate system of justice for juveniles that was not based on punishment at all, but rather treatment and rehabilitation. Youth were viewed as being different than adults because they were but "potential" adults. It was to be the experiences of these formative years that would determine who they would only later become as adults. Thus, keeping adolescents, even those committing serious offenses, away from more hardened adult criminals and providing them with education and treatment services was seen as the moral obligation of society to produce the next generation of upstanding citizens. It was from this sense of civic responsibility and "moral" obligation that a juvenile court was to be born.

The following sections will present the historical facts leading up to the establishment of the nation's first juvenile court, trace the evolution of the court over time and space, and place the events of modern times in their historical context.

Before the Juvenile Court

In early America the colonies were settled by groups of people who shared similar ethnic origins and homogenous religious beliefs. In these small communities, it was easy to maintain order as everyone had a role to play and shared values that assisted the community in enforcing norms without the need for much outside formal intervention. However, as America and its cities began to grow, large influxes of immigrants changed the landscape of America. Known as the "melting pot," the historical image often presented is one of these groups coming together and melting into a harmonious society which benefited from the unique contributions of its various members. In reality, cities grew in pockets as settlers from different nations huddled together for security and protection from the unknown others (e.g., in New York City still today there exist areas such as Chinatown and Little Italy). In addition, with the diversity brought by the new immigrants, American cities also became a host for other social ills at the center of which was abject poverty.

As the immigrant families struggled to find the land of opportunity they were promised, their children also suffered. There was not a formal public educational system or compulsory attendance laws as there are today, so many adolescents simply gathered together in groups to form what might be known as the earliest youth gangs. Because of the difficulty finding jobs and the low wages paid to many of these workers when they did find jobs, it became more common for these youth to beg for money, play instruments, or perform on street corners for change and even pickpocket or steal. Violent crime, while less prominent, still existed as mistrust bred conflict between citizens from various ethnic and religious backgrounds.

When a youth was arrested for one of these behaviors, the most typical blanket charge was "disorderly conduct." Brought before an adult court judge, the choice was then for the court to simply do nothing with the youth or place him in an adult jail/lock-up facility. Both options proved problematic. To do nothing was to send a message to the youth that he/she could simply get away with it. This would then promote further deviant and criminal behavior because no consequences were experienced for negative behaviors. On the other hand, to take a youth and place him/her in adult jails with hardened criminals was to expose that youth to more criminogenic ideas and behaviors, also leading to further criminality.

This problem was first formally addressed by the Society for the Prevention of Pauperism. This Society, which was established in 1817 by two prominent Quakers, Thomas Eddy and John Girscom, indeed defined the problem not as juvenile crime, but rather as pauperism, that is, people who were poor because they never learned a good work ethic or were ridden with other social ills such as alcoholism that prevented them from becoming productive adults. Because such people were often confined in poor houses or prisons, the first report of this Society issued in 1818 focused on the conditions in New York City's penitentiaries. A second report released a year later drew attention to the fact that there were no separate facilities for juveniles in these penitentiaries. This was followed by a third report, written by Cadwallader Colden, mayor of New York City, who concluded that such confinement of youth with adult criminals was harmful because the penitentiary was “a fruitful source of pauperism, a nursery of new vices and crimes, a college for the perfection of adepts in guilt” (Pierce 1969).

The fourth report, issued in 1822, took one step further, recommending the establishment of an entirely separate penitentiary for juveniles. In addition, since this facility was envisioned by the society for the *prevention* of pauperism, it was designed so that its focus was not upon punishment, but rather the reform of these young offenders through the provision of a good education and a strong work ethic.

Early Institutions for Juvenile Offenders

In 1825, New York opened the first such institution called the House of Refuge. As noted, this was not to be a place of punishment, but rather of reform. Youths sent here were to find refuge from the negative influences in their homes or on the streets that were leading them down the wrong path, that of an adult pauper or criminal. Based on the concept that delinquency was related to poverty, these institutions thus adopted several practices from poor houses as well. First, these juveniles were not sentenced for a certain length of time proportionate to their offense, as in criminal courts. Rather, boys were committed until their 21st birthday and girls until their 18th birthday (this was later amended to be the same as the boys) (Bernard and Kurlychek 2010). Second, commitment to the institution did not require a criminal conviction. A city alderman could simply issue an order admitting the youth, or a parent could apply to the board of the House of Refuge. As such, it is not surprising that this institution received few, if any, youths who would have been sent to an adult penitentiary. For example, in its first year of operation, the New York House of Refuge received 73 children (Fox 1970). Of these, 63 had been sent for either vagrancy, running away from a poorhouse, or stealing. If the House of Refuge did not exist, these youth would most likely have had no intervention because of the adult court judges’ reluctance to lock up these children with adult offenders (Pierce 1969). Finally, also consistent with the practice in poorhouses was the heavy emphasis on work. Commentators on the children in the House of Refuge emphasized the connection between idleness and temptation.

The idea of the House of Refuge spread quickly. A second house was established in Boston in 1826 (1 year after the first one in New York) and a third in Philadelphia the following year. As these houses grew so did questions about their operation, particularly the practice of holding children who had committed no offense other than being poor. Once such challenge came before the Pennsylvania State Supreme Court concerning a young girl, Mary Ann Crouse. Like many others, she had not committed any criminal offense, but was a poor child who appeared to be in danger of growing up to become a pauper. On the complaint of her mother, she was brought into court and committed to the House of Refuge. Her father objected to this and filed a writ of *habeas corpus* (a legal demand for the state to explain why it is holding someone). This case raised a crucial legal issue: that there should be no punishment unless a crime has been committed. Thus, this case argued that the House of Refuge was a form of punishment, and therefore, it was not right to hold a child in its grasp who had committed no such offense.

In 1838 when the Court handed down its final decision in this case, it would, however, disagree. The Court rejected the father’s arguments and held that sending Mary Ann to the House of Refuge was perfectly legal because it was not punishment. The Court reasoned as follows:

> The object of the charity is reformation, by training its inmates to industry; by imbuing their minds with principles of morality and religion; by furnishing them with means to earn a living; and, above all, by separating them from the corrupting influence of improper associates. To this end, may not the natural parents, when unequal to the task of education, or unworthy of it, be superseded by the parens patriae, or common guardian of the community?... The infant has been snatched from a course which must have ended in confirmed depravity; and not only is the restraint of her person lawful, but it would be an act of extreme cruelty to release her from it. (Mennel 1973, 22–23)

This case is important for two primary reasons. First, it held that the purpose of the institution was in fact charity, thereby confirming its roots in social welfare rather than criminal justice. Second, it borrowed from an English doctrine used in chancery rather than criminal courts to validate the role of the state in holding such children. This doctrine, *parens patriae*, allowed the state to act as the parent or guardian when the natural parent was unable to do so. Originally intended for use only when children were orphaned, the Pennsylvania Supreme Court clearly extended the reach of such a doctrine to hold to children whose parents were still alive, but not raising their children in a manner seen as fit by the state.

This was not to be the last word on the matter, however. The House of Refuge movement continued to spread with over 40,000–50,000 youths having been confined in these houses by 1868. As the Houses became more crowded, other problems ensued as well. It was harder to control the behavior of inmates, which led to strict disciplinary practices. It also became more and more difficult to find jobs for these children upon release, leading them to simply be sent back to their homes. As such, the practices of the Houses, rather than merely their charitable roots, came into question, leading to a second challenge to their power. In this second case, the parents of Daniel O'Connell, similar to the father of Mary Ann Crouse, filed *a writ of habeas corpus* challenging the right of the state to hold their son, who had committed no criminal offense, in such an institution. This case, originating in Chicago, went before the Illinois Supreme Court. In a finding contrary to that of the earlier Crouse case, this Court found that commitment to the House of Refuge did constitute punishment as follows:

> Can the State, as parens patriae, exceed the power of the natural parent, except in punishing crime? These laws provide for the "safe keeping" of the child; they direct his "commitment" and only a "ticket of leave" or the uncontrolled discretion of a board of guardians, will permit the imprisoned boy to breathe the pure air of heaven outside his prison walls....The confinement may be from one to fifteen years, according to the age of the child. Executive clemency cannot open the prison doors for no offense has been committed. The writ of *habeas corpus*, a writ for the security of liberty, can afford no relief, for the sovereign power of the State as parens patriae has determined the imprisonment beyond recall. Such restraint upon natural liberty is tyranny and oppression. (Platt 1977, 103–104)

Thus, this case ended with the Illinois Supreme Court ordering Daniel released to his parents. According to the argument presented by Bernard and Kurlychek (2010), the different finding was not based on different practices or circumstances between the two cases, rather it was based upon the court's decision to focus on actual practice in the later case rather than "good intentions" as in the former case.

The First Juvenile Court

Despite this seemingly final decision by the state court, the sentiment behind the creation of the Society for the Prevention of Pauperism and these early institutions had not changed. This sentiment was that a growing portion of the nation's children, particularly those in urban centers, were in need of help.

While the first juvenile institutions reviewed above had been founded by men, the first juvenile court was to be largely the result of the work of women. Although women typically did not have political power at this time, the fact that these women were addressing the care of children, a typical female concern, allowed then entrance into an otherwise male realm of influence. Also, these women were the wives and daughters of prominent and powerful men who ran Chicago's political and business establishments – in other words, they "had connections." Like their husbands and fathers, these women were wealthy, white, Anglo-Saxon, and Protestant. According to Anthony Platt's classic work, "The Child Savers" (1977), these women had the sense that their culture and way of doing things was indeed superior to other cultures and races and wanted to instill their values into the next generation of youth.

Despite their desire to "help," these women had no clear path to do so. For the greater part of the previous century, youth were able to be removed from their homes without question and put in institutions designed to "help" them. Now, thanks to the O'Connell decision such a practice was illegal. While juveniles who committed serious crimes could still be brought before the adult court and put in institutions, for most juveniles who were committing minor acts of delinquency, the choice was again to do something that seemed too harsh (lock them up with adult criminals) or something too lenient (to let them off scot free).

For example, Julia Lathrop, who was later Chief of the Federal Children's Bureau, reported that in the 6 months prior to the establishment of the first juvenile court, 332 boys between the ages of 9 and 16 were sent to the city prison in Chicago (Lathrop 1912). Of that number, 320 were sent on a blanket charge of "disorderly conduct," which could represent a variety of offending behaviors from very trivial to fairly serious. About one-third of these were pardoned by the mayor, meaning the ultimate outcome was to again do nothing. So it was that the officials in Chicago were faced with the same dilemma that led to the creation of the House of Refuge only 75 years earlier. Not surprisingly then, much of the rhetoric of the time mirrored that put forth by the Society for the Prevention of Pauperism. For example, the previous report by Julia Lathrop went on to argue the following:

> There are at the present moment in the State of Illinois, especially in the city of Chicago, thousands of children in need of active intervention for their preservation from physical, mental and moral destruction. Such intervention is demanded, not only by sympathetic consideration for their well-being, but also in the name of the commonwealth for the preservation of the State. If the child is the material out of which men and women are made, the neglected child is the material out of which paupers and criminals are made. (Lathrop 1912, 4)

Such advocacy by the Child Savers ultimately resulted in the Illinois Conference of Charities voting to recommend the establishment of a separate court for juveniles. The idea, suggested by Frederick Wines, one of the most prominent penologists in the country, quickly gathered support, and on January 1, 1899, a juvenile court law was enacted unanimously by the Illinois Legislature. This new court, built on charity rather than principles of justice, was designed to serve the best interests of the child and society. For these reasons it did not look like a criminal court, but more closely resembled a chancery court. Because chancery court was intended to help children who lacked proper parental care, this model fit the goals of the child savers well. It allowed them to intervene again in the lives of "needy" children even if a crime was not committed and without the due process protections provided in adult court. Indeed, drawing on the early Crouse decision handed down from the Pennsylvania Supreme Court, the new juvenile court was to be based on the parens patriae doctrine and to serve "in the best interests" of the child.

To establish a juvenile court as a chancery court required two separate steps: Jurisdiction over juveniles had to be (1) removed from criminal court and (2) established in the new juvenile court. The process of the new court was then to be confidential, informal, and ultimately beneficial to the child. Because it was not a criminal court, not only did due process protections not apply, but neither did standard criminal justice terminology such as being tried, found guilty, and sentenced. Instead these juveniles were "taken into custody" by the state. A social history was to be prepared by a case worker, a role originally taken by social workers that would later evolve into juvenile probation. This social history was not to focus on a specific offense committed, although that was indeed part, but included interviews with family members, friends, educators, and the like. If it was determined that official court intervention was needed, a petition was filed asking the court to hear the case. At this stage, however, many youth were handled informally and services given without further need for official declaration of delinquency. In some cases, however, the youth would be adjudicated delinquent by the juvenile court judge.

Similar to being found guilty in adult court, adjudication indicates that there was evidence that the youth was involved in some act of delinquency or a status offense. However, being a civil court, the standard of proof was preponderance of the evidence (51 %) instead of proof beyond a reasonable doubt which is the criminal court alternative. Also, an adjudication of delinquency implied something else: that the youth was in need of care and treatment. This later part often outweighed consideration of evidence that any offense was committed. After all, if the ultimate goal of this new court was to help needy youth, it would be much better to provide services to a youth in need before things were to become worse. Upon adjudication, the juvenile court judge was then to determine a disposition crafted in the best interests of the child. Thus, it was not a punishment for the offense committed (if any) but a treatment plan. This disposition stage in juvenile court was to be very different from sentencing in adult court, where punishments were administered that were proportional to the offense committed.

To keep the process informal, no lawyers were present at proceedings, just the juvenile, the probation officer or case worker, the juvenile's parents, and the juvenile court judge. All records were confidential to protect the youth from being stigmatized or labeled as a delinquent in the community. Moreover, to avoid the youth internalizing any feelings of being a hardened criminal, youth were not fingerprinted, photographed, or otherwise processed in the formal ways of an adult offender (see ▶ Labeling and Deviance).

The juvenile court was seen as a success, and variations of this model were adopted in many jurisdictions across the nation, with all 50 states and the District of Columbia enacting official juvenile court systems by the mid 1900s.

Challenges to the Court

As the juvenile court evolved, considerable focus was placed on keeping these juveniles away from adult offenders in placements similar to the earlier House of Refuge movement. Many jurisdictions enacted cottage systems which were to provide small homelike environments for these youth. But as the court grew, so did its clientele, and again, overcrowding led to the development of large state-run institutions that more resembled prisons than a "home." As the "beneficial" nature of these institutions came into question, so eventually did the informal nature of the court and its seemingly unyielding power to confine youth for extended periods of time with little to no evidence of any wrongdoing by the child.

The time period of the 1960s was characterized by great social change. Many challenges were brought before the nation's courts addressing violations of citizens' rights and liberties. The new juvenile court was not to be immune. It was during this same time that the first formal challenges to this new system of justice were to be heard by the nation's highest court, the United States Supreme Court. The case of Morris Kent was the first in a series of five cases often lumped together under the terminology "the due process movement." In this case, Kent was accused of very serious charges including several counts of breaking and entering, burglary, and rape. Based on the serious nature of the charges and Kent's history with the juvenile system, the juvenile court judge had decided to waive the case to adult court without a formal hearing. It was the informality of the juvenile court judge's procedure that was brought into question before the Supreme Court. After reviewing the facts of the case and the juvenile court statute of the District of Columbia (the court of original jurisdiction), the Court determined that waiving a juvenile to adult court without a formal hearing and the right to an attorney was indeed a violation of the due process protections provided by the United States Constitution. While this decision was rather narrow, focusing on the juvenile court statute of the District of Columbia and addressing only one decision point – the waiver decision – its message was clear: The Supreme Court was now willing to address issues of due process in the nation's juvenile courts.

The next case to reach the Court, In re Gault, was to delve much deeper into the core of the juvenile court and its informal practices. In this case a young boy of 15, Gerald Gault, had been accused of making a prank phone call. Without notice of charges or notice to his parents, he was taken into custody. Gerald admitted to dialing the number but indicated that his friend had made the lewd comments on the phone. The victim, Mrs. Cook, did not appear in court or give testimony against Gerald. Yet, he was adjudicated delinquent and committed to the state's Industrial School until his 21st birthday. While he probably would not have been held nearly this entire time, it is still interesting to contrast this outcome to what would have been a possible penalty for a prank phone call if Gerald had been in adult court: imprisonment of not more than 2 months and a fine of $5–50. Similar to the cases reviewed before of Mary Ann Crouse and Daniel O'Connell, Gault's parents filed a writ of habeas corpus saying that their son's due process rights had been violated. The specific rights in question were the right to a timely notice of charges, the right against self incrimination, the right to cross-examine witnesses (remember Mrs. Cook did not appear), the right to an attorney, the right to a transcript of the proceedings, and the right to an appeal. All of these rights were not granted under the state's juvenile court law.

In its decision, the Supreme Court granted the first four of these rights stating that:

> Under our Constitution the condition of being a boy does not justify a kangaroo court. (In re Gault 1967)

Another change to the juvenile court was to be close on the heels of this decision. In the case of In re Winship a youth accused of shoplifting was challenging the standard of proof used in the juvenile court. In this case the youth was adjudicated delinquent, and his attorney specifically questioned the judge on whether there was proof beyond a reasonable doubt that he was guilty of the offense. The judge replied that there was not, but that there was a preponderance of evidence and that was the standard for juvenile court. Upon review of this case, the Supreme Court decided that the same standard of proof (beyond a reasonable doubt) should hold in juvenile as in adult court.

In 1971 the case of McKeiver v. the Commonwealth of Pennsylvania was to similarly make its way through the state courts and all the way to the United States Supreme Court. Here, the question at hand was whether to afford juveniles a right to a trial by a jury of their peers. Contrary to the previous two decisions which clearly extended due process rights to juveniles, this case was to mark what is known as "the end" of the due process movement. In this decision the Court held that while the decisions in Gault and Winship focused on insuring accurate fact finding (admission of guilt, cross-examination of witnesses, standard of proof), there was no evidence that a jury was more successful than a judge at determining the facts of a case (see ▶ Juries, Lay Judges, and Trials). Moreover, the plurality for the first time in this series of decisions noted the beneficial nature of the juvenile court and suggested that this particular right would most interfere with the underlying purpose of the court, which was to rehabilitate and not punish the youth. While this reasoning may sound contrary to the prior decisions, it is important to note that individual judges did not for the most part change their vote. That is, those who voted to extend rights to Gault and Winship voted to extend the right to a jury trial as well and vice versa. The difference seemed to be in the changing nature of the Court. Most specifically, Justice Fortas, who had written the majority opinion in both the Gault and the Winship decisions, was gone, replaced by President Nixon's conservative replacement, Justice Blackman. Justice Warren, who had served as the Chief Justice throughout most of the civil rights movement, was now replaced as well by a more conservative Chief Justice Burger.

This shift from liberal to conservative politics was to effect more than the makeup of the Supreme Court but criminal justice policies across the nation. While the Due Process movement was credited with providing protections to juveniles, it also did something more subtle – it redefined them from needy youth to defendants in court in need of rights. It is only a small step from envisioning these youth as defendants in court to envisioning them then as full-blown criminals. Thus it may come as no surprise that the next major reform to the juvenile justice system was to target its very core: its social welfare roots.

Now referred to as the "get tough movement," the period of the 1980s and 1990s introduced the most prolific changes to the juvenile court since its inception (see ▶ Juvenile Justice in the Get Tough Era). Rather than being based on national court cases however, this movement was spearheaded by a media campaign of fear promoting the view of juveniles not as needy and dependent youth but as vicious and hardened criminals in need of punishment (Zimring 1998) (see ▶ Juvenile Violence). As the changes took place at the state level, the flavor of juvenile justice still differs somewhat across states, with some being identified as more punitive (e.g., Texas and Florida), while others maintain the more traditional child-saving orientation (Massachusetts). Many states, however, now fall somewhere in the middle trying to balance the historical treatment focus of their juvenile system with a recognition that there are some juveniles from whom society needs to be protected. This new mission of juvenile justice was grown from a model of juvenile probation, the Balanced Approach, proposed by Dennis Maloney, Troy Armstrong, and Dennis Romig (1988). It has since been combined with principles of restorative justice and is promoted by the federal Office of Juvenile Justice and Delinquency Prevention as Balanced and Restorative Justice (BARJ) (see ▶ Restorative Justice and State Crime).

In addition to the purpose of juvenile justice, its practice was altered as well. According to a report by the National Center for Juvenile Justice (Torbet et al. 1996), the get tough movement also changed the confidentiality provisions of many states, allowing for the name of arrested juveniles to be released to the media and to their schools for certain types of offenses. In addition, many states now routinely fingerprint and photograph juveniles accused of felonies, thereby making the process more similar to adult court than a social welfare agency. Every state also readdressed practices for transferring certain juveniles to adult court for processing and created new and innovative ways to transfer more and more juveniles to adult court. For example, in addition to the discretionary waiver hearing set forth by the Kent decision discussed earlier, many states enacted tough new statutory exclusion provisions that exclude from juvenile court jurisdiction either certain offenses (violent offenses) or certain youth based on combinations of their age, prior record, and current offense. Other states awarded both juvenile and adult court concurrent jurisdiction over such classifications of youth and offenses and placed the decision of which court to file charges within in the hands of the prosecutor. This practice is often referred to as "direct file" or "prosecutorial waiver." Fifteen states have also enacted blended sentencing laws that allow juvenile court judges to impose adult court sentences and/or adult courts to impose juvenile sentences, thereby further blurring the boundaries between juvenile and adult courts.

Although the legacy of the due process movement led to the redefinition of juveniles as "criminal defendants" rather than needy and dependent children, unfortunately, many of its original provisions did not sell as well. According to a series of state-level reports conducted through the American Bar Association's Juvenile Justice Center, at the turn of the twenty-first century attorneys are still not present at many adjudication hearings. Furthermore, when attorneys do represent youth, they are most often overworked public defenders with little to no training in juvenile justice procedures (Brooks and Kamine 2003; Calvine 2003; Miller-Wilson 2003; Puritz and Sun 2001; Puritz and Brooks 2002). Moreover, Bernard and Kurlychek (2010) also propose that most cases never even reach the trial stage because juveniles are waiving their right to an attorney and confessing to acts before understanding the consequences of such a confession. Because historically the juvenile court was seen as there to help youth, parents often encourage children to confess to crimes, and without the advice of an attorney, such confessions are often offered. Once a youth has confessed, there is little need for an adjudication hearing, so the potential benefits of all other due process protections are negated.

The Juvenile Court Today: Purpose and Practice

The juvenile court of today retains some of the basic principles of its predecessor mixed with a blend of adult criminal court terminology, practices, and punishments. Some of these changes are indeed a direct response to what was seen as a more violent and dangerous type of juvenile crime originating in the inner-city drug markets of the 1980s. Others have been referred to as more "knee-jerk" reactions to an imagined new breed of juvenile offender that simply does not exist. In harsh reality, however, what has happened is that more of the punitive changes of the get tough movement have been embraced than the civil rights protections offered during the due process movement. That is, many juveniles now find themselves before a court focused on punishment without the benefit of full, if any, due process protections.

What does the juvenile court of today then look like? While again, the details will vary drastically state to state, in general, a youth will come into contact with the system via an arrest or a referral through a school (often for truancy) or a parent. Once referred, the juvenile will meet with an intake officer who will conduct a social history investigation and most likely attempt to have the juvenile confess to any charges. Upon confession, a diversion option might be offered (see ▶ Juvenile Diversion) such as an informal probation or community service upon the completion of which all charges will be dropped, still allowing the youth to proceed into adulthood with a clean slate. However, most youth will be processed formally in the juvenile court, where they will most often be adjudicated delinquent (usually without the representation of an attorney). A disposition hearing will then be held. The disposition hearing will most likely still result in a "sentence" of probation. However, about 23 % of adjudicated youth will find themselves placed out of home.

For most of those processed formally, their offense will follow them to adulthood and can count against them for educational, employment, and criminal justice purposes. For example, most college applications now ask if one has had a juvenile adjudication. Such questions are routine on employment and housing applications as well. Many youth also do not understand that the juvenile adjudication can count as a prior offense in adult court even though they were not convicted with all the due process protections offered in adult court. For example, in California a juvenile felony counts as a "first strike" in the state's three strike law.

In addition to concerns over lack of due process protections and lifelong consequences of system involvement, the system has also been greatly criticized for racial bias. For example, while African-Americans make up only about 16 % of the juvenile population of the nation, this group accounts for 27 % of all juvenile arrests. Furthermore, about 70 % of white arrestees as compared to 82 % of African American arrestees are referred to court for formal processing. These disparities continue throughout the system with African American youth being more likely than white counterparts to be detained, adjudicated delinquent, and placed out of home, and also waived to adult court (Snyder and Sickmund 2006). While some argue that due to social and other circumstances it may be that minority youth simply commit more crime, studies suggest that such factors do not account for the vast levels of disparity in the system. Moreover, the fact that the disparity continues throughout decision-making points in the system further points to an institutionalized bias against minority youth (Feld 1999; Bishop and Frazier 1988). Indeed the problem has gained national attention with amendments to the 1974 Juvenile Justice and Delinquency Prevention Act calling for states to directly address disproportionate minority contact with the juvenile justice system in order to be eligible for federal funding.

The Future of the Juvenile Court

With the more formal processes of juvenile court, the lifelong consequences, and the noted racial bias, some question whether there is even a need for a juvenile court at all. For example, noted juvenile justice scholar Barry Feld accuses the court of providing juvenile "injustice" and calls for the possible demise of the court altogether. According to Feld (1993) there are three possible solutions to the current problem. One is to restructure the juvenile courts to fit their original therapeutic purpose. The second is to admit that the new juvenile court is focused on punishment and not treatment and to provide juveniles within this system full due process protections including the right to a jury trial. The final, and in his view perhaps the most promising, is to abolish the juvenile court altogether and try young offenders in criminal court with certain modifications to both the procedure and punishment to account for youths' immaturity of judgment (see ▶ Hormones, the Brain, and Criminality).

Others, not yet willing to give up on the juvenile court, prefer the first option. Indeed, even after the get tough movement, public opinion surveys showed that when given the choice between spending tax dollars to imprison youth or to rehabilitate them, most still support the concept of a rehabilitative system (Nagin et al. 2006). Also, states that once embraced a get tough philosophy are beginning to relent on this position and reintroduce leniency for juveniles. For example, Connecticut recently raised its maximum age of juvenile court jurisdiction from 16 to 18, and a recent report put forth by Florida's Blue Print Commission (2008) indicates it is time to stop getting tough on juveniles and time to start getting smart.

This return to a rehabilitative philosophy fits well within an argument first made by Tom Bernard in 1992 and reaffirmed by Bernard and Kurlychek in 2010. According their book, *The Cycle of Juvenile Justice*, society fluctuates between thinking juvenile crime is "caused" by lenient programs for juveniles (e.g., rehabilitation) or by policies that are too harsh, thereby forcing officials to choose between doing nothing or imposing detrimental sanctions. Because these ideas clearly place the "cause" of juvenile crime within the system, efforts to reduce crime then also fluctuate between toughening up the system at times when it is thought to be lenient and introducing more lenient policies when it is thought to be too hard (e.g., the current status of the system after the get tough movement). Unfortunately, this misses the point that while the juvenile justice system can respond to delinquency, it does not, nor ever did, "cause" the problem. Juvenile delinquency developed as a consequence of complex modern society and its various social ills. Therefore, every new policy will eventually "fail" to solve the problem it did not create, thus pushing us on to the next turn in this cycle.

However, these authors do not leave us in a sense of perpetual hopelessness but acknowledge that although delinquency will exist as long as the societal conditions that created it exist, there still can be identified "best practices" in juvenile justice policy that help to alleviate the hardships faced by many youths. The federal Office of Juvenile Justice and Delinquency Prevention is a key leader in this endeavor, providing funding to state agencies, program providers, and research entities as they strive to find programs that truly rehabilitate youth into productive adult citizens.

Whatever the future holds, today the reality is that every state in our nation retains some form of a juvenile court that is neither entirely true to its child-saving roots nor fully implemented as a criminal court.

Recommended Reading and References

Bernard T (1992) The cycle of juvenile justice. Oxford University Press, New York

Bernard T, Kurlychek M (2010) The cycle of juvenile justice. Oxford University Press, New York

Bishop DM, Frazier CE (1988) The influence of race in juvenile justice processing. J Res Crime Delinq 25:242–263

Blue Print Commission (2008) Getting smart about juvenile justice. Florida Department of Juvenile Justice, Tallahassee

Brooks K, Kamine D (2003) Justice cut short: an assessment of access to counsel and quality of representation in delinquency proceedings in Ohio. American Bar Association, Criminal Justice Section, Juvenile Justice Center, Washington, DC

Calvine EM (2003) Washington: an assessment of access to counsel and quality of representation in delinquency proceedings. American Bar Association, Criminal Justice Section, Juvenile Justice Center, Washington, DC

Cauffman E, Steinberg L (2000) Immaturity and judgments in adolescence: why adolescents may be less culpable than adults. Behav Sci Law 18:741–769

Feld B (1990) The punitive juvenile court and the quality of procedural justice disjunctions between rhetoric and reality. Crime Delinq 36:443–466

Feld B (1993) Juvenile (in)justice and the criminal court alternative. Crime Delinq 39:403–424

Feld B (1999) Bad kids: race and the transformation of the juvenile court. Oxford University Press, New York

Fox S (1970) Juvenile justice reform: an historical perspective. Stanf Law Rev 22:1187–1192

Grisso T (1981) Juveniles' waiver of rights. Plenum, New York

Grisso T, Schwartz R (2000) Youth on trial: a developmental perspective on juvenile justice. University of Chicago Press, Chicago

In re Gault (1967) 387 U.S. 1

In re Winship (1970) 397 U.S. 358

Kent v. United States (1966) 383 U.S. 541

Lathrop J (1912) Introduction. In: Breckinridge SP, Abbott E (eds) The delinquent child and the home. Arno, New York

Maloney D, Armstrong T, Romig D (1988) Juvenile probation: the balanced approach. Juv Fam Court J 39:1–57

McKeiver v. Pennsylvania (1971) 40 U.S. 528

Mennel RM (1973) Thorns and thistles: juvenile delinquents in the United States, 1925–1940. University Press of New England, Hanover

Miller-Wilson S (2003) Pennsylvania: an assessment of access to counsel and quality of representation in delinquency proceedings. American Bar Association, Criminal Justice Section, Juvenile Justice Center, Washington, DC

Nagin DS, Piquero AR, Scott ES, Steinberg L (2006) Public preference for rehabilitation versus incarceration of juvenile offenders: evidence from a contingent valuation survey. Criminol Public Policy 5(4):623–672

Pierce BK (1969) A half century with juvenile delinquents. Patterson-Smith, Montclair

Platt A (1977) The child savers. University of Chicago Press, Chicago

Puritz P, Brooks K (2002) Kentucky advancing justice: an assessment of access to counsel and quality of representation in delinquency proceedings. American Bar Association, Criminal Justice Section, Juvenile Justice Center, Washington, DC

Puritz R, Sun T (2001) Georgia: an assessment of access to counsel and quality of representation in delinquency proceedings. American Bar Association, Criminal Justice Section, Juvenile Justice Center, Washington, DC

Singer SI (1996) Recriminalizing delinquency. Cambridge University Press, New York

Snyder H, Sickmund M (2006) Juvenile victims and offenders: a national report. National Institute of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention, Washington, DC

Torbet P, Gable R, Hurst H, Montgomery I, Szymanski L, Thomas D (1996) State responses to serious and violent juvenile crime. Office of Juvenile Justice and Delinquency Prevention, U.S. Department of Justice, Washington, DC

Zimring F (1998) American youth justice. Oxford University Press, New York

Zimring FE (2005) American juvenile justice. Oxford University Press, New York

History of Juvenile Justice

Guadalupe Mendiola-Washington and
Traqina Q. Emeka
University of Houston-Downtown, Houston,
TX, USA

Overview

The creation of the juvenile justice system began with the efforts of benevolent individuals and institutions who responded to the evolving social, economic, and political issues in early America. The juvenile justice system reflected a change in society that addressed and recognized the distinct differences between adults and juveniles. The first juvenile court sought to avoid the adversarial system of the adult court and emphasized individual treatment and rehabilitation. The juvenile court addressed issues concerning those children who allegedly committed delinquent offenses as well as those neglected and abused children. The parens patriae doctrine gave the court authority over the child if the parent was deemed unable and/or unwilling to care for the child. The operation of the juvenile court remained relatively uncontested until the 1960s when the US Supreme Court intervened and questioned the practices of the juvenile court. Prior to US Supreme Court intervention, the juvenile court had complete discretion and authority over the treatment and placement of juveniles. The US Supreme court cases extended many due process protections offered adults to juveniles. Thus, began a shift towards punitive sanctions for juveniles. Although the juvenile justice system has tried to balance both rehabilitation and appropriate sanctions for delinquency, policy changes made in the late twentieth century tended to favor punitive sanctions and the criminalization of juveniles. However, it can be argued that the current juvenile justice system is again moving towards rehabilitation.

Introduction

Prior to the seventeenth century, there was no concept for childhood and thus juveniles who committed crimes were treated as adults. England's 1600 Chamberlain's Court, a court that handled disputes of English children who served as apprentices, is the forerunner to the current juvenile court of the United States. England attempted to address the problem of delinquency by developing ages of culpability, between the ages of 8 and 14, and by establishing havens to commit delinquent youth. In 1756, the English Marine Society, which also served to provide soldiers for war, was created for poor and delinquent boys and the House of Refuge for Orphan Girls was created in 1758 and in 1788, England established the first private juvenile prison.

The history of juvenile justice in the United States is based on English laws and began as early as the seventeenth century. According to Platt (1977), the plight to create a system in the United States that is unique to juveniles can be traced to the seventeenth century. The Stubborn Child laws were established in Massachusetts in the 1600s. This law was the first to distinguish the behavior of children from that of adults. The law granted the authority to punish children for behavior or actions deemed inappropriate. The Puritan ideals, which rested primarily in religion, emphasized the family, community, education, and the church to teach children how to be morally responsible law-abiding adults. The stubborn child law basically made disobeying a parent an act punishable by death. This law was the first of many that used the law to govern the behavior of children and to outline the expectations of parents. The enactment of the stubborn child law reflected political values of the Commonwealth as a means of maintaining social control by way of the law. The basic Puritan ideals of 1646 have, to a certain extent, withstood the test of time as politics and ideologies continue to guide the laws of today. The Puritans maintained that no individual was exempt from sin and it was the responsibility of the moral institutions (family, church, and community) to train and educate children in the proper manner. Social institutions were held responsible for molding children.

The term "delinquency" resulted from modernism and was coined in the 1700s. However, the issue of juvenile delinquency heightened during the nineteenth century. The rapid industrialization that took place during the 1800s in the United States brought about changes for which people and society may not have been prepared. Prior to urbanization, the parent and the community handled crimes by juveniles informally. Other methods used for controlling juvenile delinquency included indentured servantship, corporal and capital punishment, and transport to other countries, respectively. The means of social control for juveniles progressed to housing juveniles in adult prisons. Urbanization resulted in increased crime, as there was an influx of immigrants and a population boom.

Platt (1977) describes the rapid growth experienced by many towns, and how city populations grew at seven times the rate compared to the doubling of rural populations during the same time frame. Thus, cities grew at unexpected rates. As the population grew by drastic numbers, so did the competition for jobs and housing. The populations of cities, such as Chicago and New York, grew because of immigration and the migration of people from the southern part of the United States to the northern part of the United States.

Migrant families left the spacious surroundings of rural life, collective familial assistance, and cohesiveness for overcrowded urban life where the competition for jobs and housing was fierce. The design of the family unit was forever changed. The family lost fathers as authoritarian figures during industrialism because individualism became more important. People no longer felt a sense of community and the role of the father seemed to diminish. The move from communal life to individuality in the city brought light to the issue of juvenile delinquency. City life was viewed by reformers and the middle class as a breeding ground for a life of crime primarily for immigrant and migrant youth. Social reformers saw the need to redirect negative behavior and made attempts to transform the poor, immigrant youth into productive citizens.

The Reformers and Juvenile Institutions

The idea of helping children was a focal point of the progressive movement and the fight for children's rights taking place in the United States at the time (Hawes 1991). Social reformers viewed children as the innocent victims of the changing culture and industrial advances. The vulnerability of the children was a great concern to the child advocates of the time. Some of the better known reformers are known as the Chicago-based *child savers* (Platt 1977). Upper- and middle-class individuals stepped in to try and rescue poor, neglected, and mistreated urban children.

Before the reform attempts of the Chicago-based child savers, there existed other out-of-home placement options for children across the county. One of the first known homes for youth in the United States was the House of Refuge, which opened in 1825. As early as 1819, a report came out that mentioned a lack of separate penitentiaries for juveniles. The *Society for the Reformation of Juvenile Delinquents* established the New York House of Refuge for children who were in danger of growing up to be paupers and delinquents. The opening of the House of Refuge was the catalyst for the opening of at least 16 similar type homes in several other cities throughout the United States. The New York home opened with the intention of keeping youth apart from adults when confined. Youth confined to the House of Refuge could be committed with indeterminate sentences; youth did not have to engage in delinquent conduct to be committed and parents could ask for direct confinement.

Regardless of the attempts of reformers to rehabilitate juvenile offenders and poor youth, the House of Refuge settings did not alleviate juvenile delinquency. While the founders of the House of Refuge masked their good deeds for poor, immoral children by financially supporting the home, their true intentions were to keep members of the upper class in power. The wealthy supporters of the House of Refuge wanted to maintain control over the impoverished to protect their own status and maintain power over them using religion as a justification. The Houses of Refuge were known for enforcing corporal punishment, solitary confinement, and other physical punishments against children. Additionally, the children in these homes were used as cheap labor. A movement from privately run homes to state-run institutions was on the horizon. There was a disconnect between the promise of proper care and the abuse suffered by males and females at many of the Houses of Refuge; children suffered brutal beatings and physical abuse.

Although some may have the impression that the House of Refuge and reformatories are the same, they are discussed as distinct placements for juveniles. There was US transition from House of Refuge placements to reformatories. Reformatories replaced Houses of Refuge in the late nineteenth century because Houses of Refuge were deemed to be too harsh; yet, reformatories stressed training and subjected children to drills and constant supervision. Reformatories, unlike the Houses of Refuge were state-supported entities and strived to portray a parent–child setting for the juvenile. This did not prove the case because reformatories operated in a prison-like fashion with a coercive, rigid setting.

Reformatories began in the state of Massachusetts in 1848. Reformatories opened across the United States just as quickly as the Houses of Refuge. The best known of reformatories is perhaps the Elmira Reformatory established in 1876 in New York. In New York after the Civil War, there was an increase in crime which led to a lack of trust in the courts, police officers, and the overall justice system. The general public did not believe the justice system was doing enough to eliminate juvenile crime. As a result of the distrust, penal reformers favored a movement towards reformatories. Elmira was built to provide youthful offenders with structure and discipline to become obedient members of society. The core reasons for establishing reformatories were to segregate children from adults, remove children from the ills of society, minimize court intervention, indeterminate sentencing, enforce the cottage approach, make youth productive, and introduce compulsory education. Enoch C. Wines was the authority on reformatories prior to the twentieth century because he was an advocate of youth reformatories and an expert on crime and penology. Reformatory advocates assumed that hard work, discipline, and education could reverse poverty, poor parenting, and a dangerous environment and produce respectable members of society. Reformatories remained popular for about 40 years, and this foundation of rehabilitation can still be found in present-day juvenile correctional settings. Despite the names of these out-of-home placements, providing youth with educational, spiritual, and vocational enrichment appeared to be the goal. It was hoped that giving youth a solid foundation would prevent further delinquency. Children were held in reformatories without being convicted of a crime, and concern regarding this issue prompted the establishment of a juvenile court.

Child advocates were located throughout the United States; while reformers included men, the majority was women who had an interest in child care reform. The *child savers* were a group comprised of Chicago women, primarily of middle and upper class, who wanted to make an impact on shaping the lives of youth. The agenda of the child savers tied in to the progressive reforms taking place during the time. Many of the women who were fighting for reforms for children were also advocating for the rights of women and other social changes. The social reformers, club women, and politicians pushing for change were made up mostly of middle class, white Protestants hoping to change the lives of poor immigrants. While most reformers may have come from the middle class, a few of the women, like Louise de Koven Bowen, were quite wealthy. Ms. Bowen was independently wealthy even though she married a well-to-do banker. While the duties carried out by most of these women were done in their spare time, so as not to interfere with their family obligations, it seems that the tasks were more than hobbies. In addition to trying to reform city children, the women of the child savers movement also looked into the conditions of the jails where they found some children housed with adults. As a result of some of these findings, it seems the child savers preferred having youth, under 17 years of age, sent to reform schools instead of jails to save them from adult criminal labels.

Although many of the child savers were educated, they had limited choices for careers. Nonetheless, these women used their influence, money, and political connections to reform the way young people were handled in the justice system. Jane Addams, another woman born to privilege and wealth, founded Hull House in 1889. Jane Addams was a leader in many other Progressive Era groups that tackled organized labor and women's rights issues. As women like Bowen and Addams pushed for improving the way juveniles were handled in the criminal justice system and the foster care system, they utilized their maternal instincts, networking skills, and political influence. The women influenced the development of professional social workers. The serious issues highlighted by the child savers and their desire to have jurisdiction over children eventually led to the push for establishing the first juvenile court. Prior to the establishment of separate courts for juveniles, many youth were sent to poor houses, homes, reformatories, and industrial schools for indefinite periods of time or until the age of majority.

The men and women involved in the movement of children and youth reforms of the 1800s created, shaped, and impacted professions in the juvenile justice system. Just like the early reformers who founded the House of Refuge, the reformers of the latter nineteenth century wanted out-of-home placements that would have greater emphasis on education with the possibility of learning a trade or earning an apprenticeship. The juvenile court movement was a push to change punitive reformatories into humane schools where children could acquire skills in certain trades.

Preventing juvenile delinquency and reforming wayward children was based on the popular beliefs of the reformers. The people in charge of improving the social conditions of children did so by imparting their moral and spiritual values. Religion was used as the foundation. Members of Christian groups would enter city jails and teach Sunday school lessons to boys in hopes of imparting their own religious values and beliefs.

In addition to a strong religious foundation, the child savers believed strongly in the benefits of rural life. They believed rural living would decrease juvenile delinquency in urban areas; therefore, children were transported to western parts of the United States. Even before the Houses of Refuge was established, children were relocated to rural parts in the early 1800s; youth were sent to work on farms in other states. As the number of incarcerated youth in cities grew, supporters agreed to relocate many of the Houses of Refuge to rural areas. There was a belief that there was a purity in rural living which was a sharp contrast with the corruption of city life. A lack of confidence in Houses of Refuge led post-Civil War child savers to open more cottage-style homes for youth.

Cottage-like reformatories were different from traditional youth reformatories because fewer children were housed together; each cottage had its own dining, schooling, and recreational schedule. Additionally, each cottage was run by a married, Christian couple who served as the mother and father figure to the detained children. The cottage-plan reformatory was used to provide delinquent and neglected youth with a wholesome and loving home with the hope they would escape the temptation and corruption of the city. The new penology sought intensive supervision and to create an atmosphere of family life with the cottage plan. The cottage-plan reformatory would, theoretically, provide a natural, peaceful environment that was conducive to a healthy and productive life.

The Establishment of the Juvenile Court

The first juvenile court was established in Chicago, Illinois (Cook County), in 1899. However, before the bill to establish the court was signed into law, there was much debate and controversy. In addition to the women previously mentioned, there were other men and women who played key roles in getting the "*Juvenile Court Act*" passed (Anderson 1988). The Governor of Chicago at the time, John P. Altgeld, was an advocate for child reform. Governor Altgeld's advocacy days began when he was an attorney in the 1880s; he studied and wrote about the treatment of children processed through the criminal justice system. His writings mentioned that the handcuffing of children, housing them with adult offenders, and sending them to houses of correction prepared them for a life of crime.

As Governor, Altgeld appointed Julia Lathrop, a social reformer in her own right, to the Board of State Commissioners of Public Charities in 1892. Ms. Lathrop was a very close friend of Jane Addams. She visited all 102 counties in the state of Illinois. Julia Lathrop was also part of the Chicago Woman's Club, where she worked closely with Lucy Flower, the club president. Ms. Flower returned to Chicago in 1895 after spending time in Boston studying 25 years of the probation system. Upon returning to Chicago, Ms. Flower urged the women to propose legislation to have separate hearings for young offenders; however, the group's legal advisors did not feel comfortable moving forward with the proposal as it was. A year earlier in 1894, Mrs. Perry Smith, a volunteer with the Protective Agency for Women and Children also asked for a separate court for youth. And while some felony youth offender cases were heard on Saturday mornings, this involved only a small number of cases. A few years later, the push to establish the juvenile court continued with greater force.

The full title of the "*Juvenile Court Act*" that called for separate hearings for youth was "*An Act to Regulate the Treatment and Control of Dependent, Neglected, and Delinquent Children*." While the earlier attempts to push the legislation forward came primarily from the Chicago Woman's Club and perhaps even a handful of men, the women decided that the new legislation should be introduced by men, more specifically, male attorneys (Anderson 1988). The women would still play a vital role in getting the bill passed; however, the lawyers would lead the discussions and debates. One must keep in mind that many of these women were also involved in the ongoing fight for women's rights due to being treated as second-class citizens to men. Their status in the fight for children's rights was no different, but they were bright enough to place men in the forefront of the struggle. The bill was reviewed by Judge Harvey Hurd and it was determined that the passing of the bill, as it was written, would also go against the Illinois constitution because of a uniformity clause. Ongoing tensions between Chicago and other parts of Illinois meant that not everyone in the state would support a mandated, statewide initiative. As a result, Judge Hurd recommended the language in the bill be changed to be *permissive* rather than *obligatory*. Thus, cities could opt out of creating a separate juvenile court if they so chose.

The main opposition to the proposed bill came from those who supported the industrial schools for youth, perhaps, because of a potential loss of state funding. Oscar Dudley, a strong opponent to the bill, believed that there was still a place for industrial schools because not all children were equipped to live at home. The authors of the bill decided to bargain with the industrial school supporters and allowed them to retain the power to release youth or place them in foster care without the court's consent.

After much negotiation and some amendments, the bill would be signed into law and incarcerated youth under the age of 16 would be kept separate from adults, except in the cases of capital felony offenders. Some argue that the juvenile court was established to benefit law enforcement, judges, and prosecutors. Nevertheless, child care reformers seemed to gain a huge victory and the juvenile justice system in the United States was born. The creation of the juvenile court did not come without opposition. Although the creation of a separate court for children was a milestone, there were opponents who believed the juvenile judges had too much power over youth and their families.

While the United States was engaged in a war in Europe in the early part of the twentieth century, the struggle for reform continued at home. The United States Children's Bureau was created in 1912 with reformer Julia Lathrop named its first director by President Taft. A new concept of childhood developed as more and more children were encouraged to obtain an education. The creation of the Cook County Juvenile Court spurred the creation of other juvenile courts throughout the nation in the early part of the twentieth century. All but two states had some form of a juvenile court by 1927; and by 1932, there were over 600 juvenile courts in the United States. Shortly after the end of World War II, all states had a juvenile court.

In regards to juvenile justice, the interconnectedness of social influence and policies peaked in the 1960s. Although arguments and objections had been brought against the juvenile court since its inception, the first juvenile case did not reach the US Supreme Court until the 1960s. Five Supreme Court Decisions relating to juveniles will be discussed along with the social and political context of the time. Additionally, the policies surrounding treatment for offenders through the remainder of the twentieth century will be mentioned because treatment is also dependent on the social climate of the day.

The 1960s

During the 1960s and 1970s, states began to distinguish between treatment, punishment, and confinement of juvenile delinquents. Prior to the establishment of the first juvenile court, many of the offenses for which youth came to the attention of court were not criminal in nature. Youth viewed as poor, incorrigible, and disobedient could be incarcerated indefinitely or until the age of majority (18 or 21). Several decades after the establishment of the juvenile court, some youth were still being detained for committing status offenses. Often status offenders were punished more severely than other delinquents in the 1960s.

During the 1960s, the issues surrounding the labeling of youth came to light. Indeed, an entire theory on labeling was formulated by various individuals. Moreover, less imposing rehabilitative measures were promoted. Only the most serious of offenders were to be placed in institutions and detention centers. This stance towards rehabilitation followed the original premise of the juvenile court: the parens patriae doctrine. One of the best aspects of the original juvenile court was the care and treatment it could provide to children. Judge Ben B. Lindsey, the first juvenile court judge in Denver, believed the court was to be an extension of the community and society. Others refer to the juvenile court as being a provider of services to treat causes of misbehavior in order to restore the youth to be a responsible member in the family and the community at large. Despite the mindset of politicians and practitioners of the time, a number of juvenile court matters made their way to the US Supreme Court.

Supreme Court Decisions of the 1960s and 1970s

The social unrest of the 1960s and 1970s would bring with it challenges to the juvenile court process. Along with minority groups and women fighting for equal rights, certain juvenile rights, or lack thereof, were called into question. There are five landmark decisions that contributed to our current juvenile justice system.

The first juvenile landmark case to make it to the Supreme Court took place in 1966. *Kent v. United States* brought into question the transferring of juveniles to adult criminal courts. A District of Columbia 16-year-old was judicially waived to the adult court system without a hearing. The youth would subsequently be convicted and sentenced to 99 years in prison. Kent's attorney was not allowed to view his file and was denied psychiatric testing for his client. The outcome of the case called for hearings for juveniles being transferred to adult court. Additionally, District of Columbia youth were granted the right to counsel in a transfer hearing with the attorney also given the right to access a juvenile's social records. This court ruling would lay the foundation for the next Supreme Court case the following year.

Perhaps the most transforming of landmark cases relating to juvenile matters is *In re Gault* in 1967. A 15-year-old, Arizona resident Gerald F. Gault was accused by his female, adult neighbor of making an obscene telephone call. Gault's parents were not called about their child's detainment, and in two separate hearings, the complaining witness was not present. Moreover, the youth had no legal representation and was never advised of any rights. There was even a question as to whether the youth or his friend made the obscene remarks. Regardless, the youth was committed to the State Industrial School where he was to remain until he turned 21. The case went to the US Supreme Court. Upon hearing the case, the Supreme Court ruled in favor of Gerald Gault and granted all juveniles the following due process rights: (1) Right to timely notice of actual charges, (2) Right to an attorney in hearings, (3) Right to confront accuser, and (4) Protection against self-incrimination. In essence, juveniles were afforded many of the rights given to adults.

The mid-1970s brought a movement against detaining status offenders because they had not broken any actual criminal laws. President Ford's signing of the 1974 *Juvenile Justice and Delinquency Prevention Act* called for the decriminalization and deinstitutionalization of status offenders. Some status offenses are "redefined" so they can be considered delinquent offenses. Thus, in a matter of years, a shift towards a more punitive form of justice developed. There are three more important Supreme Court Decisions that took place in the 1970s.

The *Winship Case* of 1970 was also of great importance. In New York, 12-year-old Samuel Winship was charged with breaking into a woman's locker and stealing over $100 from her purse. Winship was found guilty because the state's burden of proof was based on civil proceedings "preponderance of the evidence" and not the "beyond a reasonable doubt" burden of proof used in adult criminal cases. A "preponderance of the evidence" means that more evidence exists that a person committed a crime than evidence that a person did not commit a crime. The standard of "preponderance of the evidence" was used in juvenile cases because they had always been seen as civil court matters, as opposed to criminal. The Supreme Court ruled that juveniles accused of crimes that may result in a possible loss of freedom through confinement should be afforded the same burden of proof of beyond a reasonable doubt.

A year later in 1971 came the case of *McKeiver v. the State of Pennsylvania*. Two teenagers charged with engaging in delinquent conduct, including 16-year-old Joseph McKeiver, were denied the right to a jury trial. The Supreme Court upheld the decision that while adults may have the right to a trial by jury, juveniles do not. States could choose to extend the right to a jury trial to juveniles, but it was not a federal mandate. The rationale behind the ruling was to preserve the foundational principles of the juvenile court system. Justice Blackmun and his colleagues were concerned that a jury trial would invite delays and possible publicity endured by adults. Nonetheless, some states do afford juveniles the right to a trial by jury.

In 1975, a fifth case came before the Supreme Court: *Breed v. Jones*. Seventeen-year-old Gary Jones was adjudicated for armed robbery and two other offenses involving a weapon. Jones committed the offense on February 8, 1970 and was adjudicated delinquent on March 1, 1970. The dispositional phase of the case was set 2 weeks later. At disposition, the court waived jurisdiction to the adult system. Jones' attorney appealed his case, arguing he had been subjected to double jeopardy. The Supreme Court upheld the ruling that Jones was subjected to double jeopardy and that transfer of jurisdiction hearings needed to take place prior to adjudication. Additionally, once any evidence is heard at a juvenile hearing, a youth cannot be subjected to adult criminal proceedings.

The five Supreme Court cases challenged and called for a change in how juvenile court matters are handled. While legal proceedings for youth remained civil in nature, the landmark decisions added an element of criminal court proceedings. The protests and outcries of the 1960s and 1970s would lay the groundwork for equal treatment across various sectors in society. As women and other minority groups pressed for reform measures, the treatment of juvenile delinquents was equally in the judicial spotlight. Although considerable strides were achieved, the decade to follow would bring a conservative political shift that would change treatment options from rehabilitative to punitive.

1980s to Present Day

The 1980s brought a different conservative, hardened perspective towards crime and offenders. Perhaps the media coverage during the 1980s and 1990s increased the public's concern regarding how criminals were being handled by the justice system. A punitive versus rehabilitation mode of dealing with offenders impacted the adult and juvenile justice systems. The penetration of "crack" cocaine and the increase in gun-related youth killings ignited the "war on drugs." The so-called war on drugs fueled by the conservative right during the 1980s brought forth stiffer punishments for drug offenders. As a result, there was an increase in drug offense–related arrests. This ushered in the conservative "tough on crime" agenda of the 1980s. Williams and McShane (2009) point out that by the time the Reagan administration was beginning the "war on drugs," drug use was actually on the decline. Regardless, an era of "deterrence and punishment" became the reality for juvenile offenders. It appears that the juvenile justice system abandoned its founding principle of rehabilitating juveniles.

The shifts in treatment echo Bernard's (1992) cyclical patterns in juvenile justice. The public is always viewing juvenile crime as being on the rise, either because we are too lenient or too harsh with our rehabilitation efforts. The latter part of the 1990s saw the emergence of a different perspective to the treatment of offenders. Some believe that even during the "tough on crime" era, practitioners were developing new strategies that focused on prevention, early intervention, and rehabilitation.

The "renewed" interest in rehabilitation was mainly due to the costs associated with an increase in the prison population. A tripling of drug offenders during the 1980s and 1990s meant an increase in incarceration rates. The cost of incarcerating a drug offender is significantly higher than the cost of treatment. Providing treatment for youthful offenders remains a popular response to delinquency in the early part of the twenty-first century. Modern-day child savers continue the fight to keep juvenile delinquents in the community and out-of-home placement as a response of last resort. Additionally, specialized courts like drug courts and mental health courts seem to provide delinquents with hearings that are reminiscent of the original juvenile court (Mendiola-Washington 2012).

The parens patriae doctrine seems to remain alive in the juvenile justice system. However, the question of how to apply the doctrine justly and adequately remains constant. Effective means of dealing with offenders will always remain a topic of debate. Ensuring youth receive the necessary services will depend on contemporary juvenile justice reformers. Indeed, the history of juvenile justice in the United States is fascinating, multifaceted, and forever evolving. The next 100 years will determine if rehabilitation, parens patriae, and the best interest of the child remain integral components of the juvenile justice system.

Related Entries

- ▶ Bullying
- ▶ Communities That Care
- ▶ Juvenile Justice in the Get Tough Era
- ▶ Juvenile Justice System Responses to Minority Youth
- ▶ Juvenile Victimization
- ▶ Juvenile Violence

Recommended Reading and References

Anderson PG (1988) The good to be done: a history of juvenile protective association of Chicago, 1898–1976. The University of Chicago, ProQuest dissertations and theses, 1. https://ezproxy.uhd.edu/login?url=http://search.proquest.com/docview/303584309?accountid=7109

Bernard TJ (1992) The cycle of juvenile justice. Oxford University Press, New York

Colomy P, Kretzmann M (1995) Projects and institution building: judge Ben. B. Lindsey and the juvenile court movement. Soc Probl 42:191–215

Feld B (1999) A funny thing happened on the way to the centenary. Punishm Soc 1:187–214

Ferdinand TN (1991) History overtakes the juvenile justice system. Crime Delinq 37:204–224

Hawes JM (1991) The children's rights movement a history of advocacy and protection. Twayne, Boston

Krisberg B (2005) Juvenile justice: redeeming our children. Sage, Thousand Oaks

Mendiola-Washington G (2012) The history of Juvenile Justice and the formation of specialized Courts. Unpublished masters thesis, University of Houston-Downtown, Houston

Pisciotta AW (1994) Benevolent repression: social control and the American reformatory-prison movement. New York University Press, New York/London

Platt AM (1977) The child savers: the invention of delinquency, 2nd edn. University of Chicago Press, Chicago

Tanenhaus DS (2004) Juvenile justice in the making. Oxford University Press, Oxford

Watkins JC (1998) The juvenile justice century-a sociological commentary on American juvenile courts. Caroline Academic Press, Durham

Williams FP, McShane MD (2009) Youth, drugs and delinquency. In: Benekos PJ, Merlo AV (eds) Controversies in juvenile justice and delinquency. Mathew Bender, Newark, pp 85–106

Williams FP, McShane MD (2010) Criminological theory, 5th edn. Pearson Education, Upper Saddle River

History of Organized Crime in Mexico

Elaine Carey
Department of History, St. John’s University,
Queens, NY, USA

Overview

Due to Mexico’s 1,969-mile-long shared border with the United States, transnational organized crime networks have long existed. Moreover, Mexican organized crime has long been tied to its modern political system. In fact, the Mexico’s modern political system emerged alongside the rise of organized crime. This entry examines the century-long organized crime from the period of the Mexican Revolution (1910–1921) to the present.

Guns and Drugs: The Mexican Revolution

Organized crime in Mexico is as diverse as in the United States because of Mexico’s proximity to the United States and due to the flows of people from other parts of the world to Mexico during the nineteenth and twentieth century. In Mexico, Chinese organized crime has played a role since the 1800s. With the passage of the US Chinese Exclusion Act of 1882, Chinese migrants sought to gain entry into Mexico in order to pass through its unpatrolled border. Chinese tongs operated in Mexico beginning in the 1800s. Even prior to 1882, the Mexican government approached the Chee Kung Tong in San Francisco to recruit and acculturate Chinese workers to work in the “hot lands” of northern Mexico. With the passage of the US Chinese Exclusion Act, tongs based in San Francisco operated a human smuggling network that moved people from China to the United States, Mexico, and Cuba due in part to a continued presence of Chinese in Mexico.

Mexico’s organized crime emerged before the Mexican Revolution and at times with the consent of the Mexican government. During the Revolution, military figures saw forms of crime as a means to accumulate wealth and materials to supply their armies but also as a means to create political power structures that endured after the fighting had ceased.

Even prior to the outbreak of the Mexican Revolution (1910–1921), antidrug crusaders recognized the logistical importance of Mexico to the international drug trade. Beginning in 1908, Canadian and US officials criticized Mexican port security that allowed an array of commodities to enter the country to be transshipped north to the United States. Significantly too, Mexican revolutionary figures had tapped into illicit flows of firearms that moved guns from north to south. Pancho Villa was one such smuggler who emerged as a northern cacique (boss). After his 1916 raid of Columbus New Mexico, Villa became a criminal in the United States, but in retrospect his criminal activities were far milder than some of his contemporaries. Villa was a and bandit, but he was also one of the most successive military generals of the Mexican Revolution.

Other revolutionary leaders, such as Colonel Esteban Cantú, sought profits through racketeering. Cantú extorted poppy growers and opium producers in the state of Baja California and subsequently went into the business himself. Another military leader, General Abelardo L. Rodríguez, displaced Cantú and took control of the border state just as Prohibition began in the United States. Rodríguez made a fortune in the vice boom that emerged to serve the thirsty and bored Americans who came south looking for drink and distraction. In the 1920s, the *jefe maximo* President Plutarco Elías Calls, from the state of Sonora, placated US demands regarding the control of narcotics, while his son-in-law was intimately involved in drug trafficking. Calles waged a battle against Catholicism as well as against Chinese. In both cases, these battles were for power. To Calles, from the state of Sonora, the church undermined modernization and national control. Sonoroan Chinese undermined Mexican-owned development.

The Anti-Chinese movement of the 1920s portrayed Chinese Mexicans as drug-addicted traffickers. While Calles officially left the government in 1928, he controlled it until 1935. Rodríguez led the state of Baja California until 1929 until he became a cabinet minister in the administration of Pascual Ortiz Rubio until he abruptly became president of Mexico in 1932.

In other Mexican states, the profitability of narcotics allowed for the consolidation of power into political dynasties, and the generals who emerged as political bosses and governors continued to be involved in organized crime such as prostitution as well as the narcotics industry. In Chihuahua, for example, General Rodrigo M. Quevedo created a political dynasty that held power for decades. Quevedo was part of a group of *cacique* governors who engaged in political power plays supported by illicit activities such as drugs, smuggling alcohol, prostitution, or extortion. President Manuel Ávila Camacho (1940–1946) worked closely with the United States in poppy crop eradication in states such as Sinaloa, Durango, Jalisco, and Baja California. During WWII, he arrested drug traffickers due to the pressure of the United States and jailed them without due process. All this took place while his brother Maximino was one of the *cacique* governors of the state of Puebla who also dabbled in the drug trade.

During prohibition, Mexicans in various positions of power recognized the opportunity to make money by supplying alcohol to the north. Mexicans on the US border resented Anglo laws that interfered with their of use of and profits from alcohol consumption and some went into the moonshine business in cities such as San Diego, Albuquerque, and Houston. Others smuggled alcohol along the US-Mexico border. In Texas, tequila smugglers gained a romantic foothold – they are remembered in song and popular culture on the border. The *tequileros* brought tequila from the agave producing areas of central Mexico, while *Tejanos* (Texans of Mexican heritage) served as guides in Texas. The *tequileros* sold to bootleggers who distributed the tequila to cities throughout the USA – the profits were enormous for the time.

For the political elites in Mexico, prohibition in the USA combined with the vice districts in Mexico created great profits. Bars, cantinas, and hotels as well as sex shows and prostitution became features of Mexican border towns. The governors of border states offered thirsty Americans a place to drink and indulge corporal pleasures. Profits lined the pockets of governors, narcotics agents, and police. Citizen groups routinely complained about the vice districts, but they were complaining to the very people who had the most interest in those industries.

The First War on Drugs

With the end of prohibition, it appeared that Mexican organized crime might suffer irreparable harm. But the routes created to supply bootleggers were converted to offer greater supplies of other commodities in demand: marijuana, morphine, and heroin. During the 1930s, struggles ensued between native Mexican drug traffickers and Chinese organized crime syndicates for control of the lucrative trade. Chinese Tongs also bribed the police to target other tong competitors for deportation. Chinese involvement in drug trafficking provided the Mexican government with a scapegoat for the country's role in the international narcotics trade. This led to raids of Chinese businesses and deportations. Nevertheless, the drug industry flourished and then grew exponentially into the 1940s.

In the 1930s, Jewish organized crime groups operated in Mexico – Mexico City had a Jewish immigrant community that dated from colonial times. Mexican Jewish groups transshipped Asian and European drugs through Mexico. Frequently, Jewish criminals repackaged foreign morphine and heroin supplies in Mexico. This counterfeit medicinal morphine or heroin was then smuggled across the US-Mexico border. The Mexican branches of Jewish organized crime supplied their distributors in cities throughout the USA. As in the USA, Mexican Jewish organized crime demonstrated cross-class and cross-ethnic alliances.

In striking contrast to the USA, women emerged as recognized bosses in cities such as Mexico City and Ciudad Juárez. María Dolores

Estevez Zuleta, known as "Lola la Chata," became a subject for American beat writer William Burroughs. And she had her husband Enrique Jaramillo, an ex-police officer, built a successful transnational trafficking organization that had labs and distribution sites in three Mexican states. Ignacia Jasso, la viuda de González, who some scholars argue created the first cartel, lived and worked in the border town of Ciudad Juárez but had fields of poppy and processing labs in multiple Mexican states. Her location on the border gave her direct access to the US market across the Río Grande.

The drug trade from Asia and Europe was disrupted by World War II, creating greater opportunity for Mexican traffickers whose heroin was found beyond its traditional domains of the US Southwest. Mexican narcotics in cities such as Chicago, Detroit, Toronto, and New York revealed alliances between Mexican and US organized crime.

Post-WWII

During the 1940s, the Federal Bureau of Narcotics (FBN, predecessor to the Drug Enforcement Administration (DEA), routinely requested names and information about narcotics traffickers and dealers operating in Mexico. Officials pressured the Mexican government to arrest traffickers and at times demanded their extradition. In 1947, the Mexican government established the *Direccion Federal de Seguridad* (Federal Security Directorate; DFS) in part as a response to growing US pressure. Organized similar to the FBI, the DFS was responsible for dealing with political insurgents as well as criminals. For much of its almost 40 years of existence, the agency developed a reputation for security failures and egregious human rights violations. Though the DFS was to fight organized crime, throughout the 1940s and 1950, Mexican organized crime families flourished. In Tijuana, for example, "Big Mike" Barragan and his wife Ismelda Galindo controlled heroin flows into San Digeo and supplied dealers as far away as Houston. They controlled bordellos, the dog racing track, and a taxi business. In 1955, Mexican traffickers such as Jasso and the Barragans became the subject of US senate hearings.

Emerging New Players

The 1960s and increased drug use led to greater wealth and new players entered the business. While many drug traffickers had enriched themselves, the 1960s and early 1970s marked a change in the levels of wealth, corruption, and transnational ties. Drugs continued to flow, and the ties between policing agents, government officials, and organized crime increased. Top commanders of the DFS were found to have become involved with the criminal drug organizations offering protection and information.

In the 1970s, it was apparent that the drug rings were becoming highly sophisticated. Cuban national Alberto Sicilia Falcón, a successful marijuana and cocaine trafficker, built a highly sophisticated organization that moved in elite circles and across borders. The US senate investigation into Silicia Falcón's organization revealed information regarding the US arms traffic to Mexico. He had become so wealthy that he had attempted to gain the rights to manufacture certain US firearms in Mexico in clear violation of the country's laws. The only reason that deal fell through was that Silicia Falcón was arrested. The senate hearings demonstrated that Falcón's wealth gave him access to business people on both sides of the border.

Investigations into the 1985 murder of DEA agent Enrique Camarena, who was stationed at the Guadalajara regional office, implicated a major trafficker, Miguel Angel Félix Gallardo and his lieutenant Rafael Caro Quintero. Evidence suggested that Caro Quintero had planned Camarena's murder. When he was arrested, Caro Quintero possessed a DFS identification card.

Transnational Organized Crime

In the 1980s and 1990s, Mexican organized crime diversified: In addition to heroin, cartels controlled marijuana, methamphetamines, and cocaine, and some branched into the human trafficking working with Salvadoran gangs such as Mara Salvatrucha or MS 13. In the 1990s, further evidence of ties between *narcotraficantes* and the political elite emerged. During his presidency, Carlos Salinas' family members were indicted on drug trafficking charges. In the mid-1990s, governors of the states of Morelos, Sonora, Chihuahua, and Tamaulipas were acknowledged to be involved in drug trafficking. Jesús Gutiérrez Rebollo, director of the National Institute to Combat Drugs, was arrested on corruption charges. He protected Amado Carrillo Fuentes, the "Lord of the Skies" and his Juárez cartel while pursuing their competitors, the Tijuana (Arellano Félix) cartel. Rebollo was sentenced to a prison sentence of 40 years.

Mexican drug trafficking organizations control most of the US drug market and have established varied transportation routes, advanced communications capabilities, and strong affiliations with gangs in the United States, overseeing drug distribution in US cities.

In 2007, President Felipe Calderón declared a war on drugs with the military playing a key role. He extradited Mexican drug kingpins to the United States and purged police departments of corrupt officers or disarmed them leaving little protection for citizens. The Mexican government's military response has resulted in fierce gun battles with cartel gunmen who have refused to surrender and have ambushed soldiers and police officers.

The Calderón and G.W. Bush administrations signed the Merida Initiative, a partnership between the two countries to fight organized crime by offering equipment and training to Mexican law enforcement. In 2008, Mexico received 400 million to combat the cartels. With Calderon's war on drugs and the Merida Initiative, the drug violence continued to escalate growing more brutal and violent with public displays of decapitated bodies, brazen recruitment efforts by cartels, and strategic killings of journalists, nongovernmental activists, and politicians. The 2012 election of Enrique Peña Nieto confounded US observers when the candidate routinely sidestepped direct questions regarding how he would handle Mexico's most vexing problems: drugs and organized crime.

Related Entries

- ▶ Careers in Organized Crime
- ▶ History of Organized Crime in the United States
- ▶ Organized Crime, Types of
- ▶ Russian Mafia
- ▶ Triads and Tongs

Recommended Reading and References

Abadinsky H (2010) Organized crime. Cengage Publishers, New York

Andreas P (2006) Border games: policing the U.S.-Mexico divide. Cornell University Press, Ithaca

Andreas P, Nadelmann E (2006) Policing the globe: criminalization and crime control in international relations. Oxford University Press, Oxford

Astorga L (2003) Drogas sin fronteras: los expedientes de una guerra permanente. Grijaldo, México

Astorga L (2005) El siglo de las drogas: el narcotráfico, del porfiriato al nuevo milenio. Plaza y Janés, México

Bailey J, Godson R (eds) (2000) Organized crime and democratic governability: Mexico and the U.S.-Mexican borderlands. University of Pittsburgh Press, Pittsburgh

Campbell H (2008) Female drug smugglers on the US-Mexico border: gender, crime, and empowerment. Anthropol Q 1(1):233–267

Campbell H (2009) Drug war zone. University of Texas Press, Austin

Carey E, Marak A (eds) (2011) Smugglers, brothels, and twine: historical perspectives on contraband and vice in North America's borderlands. The University of Arizona Press, Tucson

Carey E (forthcoming) Selling is more of a habit: women and drug trafficking, 1900 to 1980. University of New Mexico Press, Albuquerque

Chao Romero R (2010) The Chinese in Mexico, 1882–1940. University of Arizona Press, Tucson

Kenny P, Serrano M (eds) (2012) Mexico's security failures: collapse into criminal violence. Routledge, New York

Quintana A (2010) Maximino Avila Camacho and the One-Party State. Lexington Books, Lantham, MD

Walker W III (1996) Drugs in the western hemisphere: an odyssey of cultures in conflict. Scholarly Resources Press, Wilmington

History of Organized Crime in the United States

Howard Abadinsky
St. John's University, New York, USA

Overview

Organized crime in America has a history that stretches from 1920 into the twenty-first century. From its beginnings in urban America to the global connections that characterize the phenomenon today, organized crime has proven to be a dynamic enterprise. This entry provides a summary examination of the nearly 100 years that organized crime has existed in the United States beginning with the critical role of Prohibition.

The Role of Prohibition

Prior to the Prohibition era (1920–1933) vice – gambling and prostitution – insofar as it was organized, was dominated by big-city political machines aided by street gangs that proliferated among the urban poor. The gangs were local entities serving as the military wing of corrupt political organizations. They engaged voter fraud – voting early and often – and intimidation of political opponents. In short, they were violent errand boys at the bottom of an organizational pyramid that featured political bosses at the top. Then came Prohibition. The "Great Experiment" was a catalyst for organized crime, especially violent forms, to blossom into an important force in American society. Prohibition led to the mobilization of criminal elements in an unprecedented manner, and the ensuing competitive violence turned the power structure upside down. It also led to a new level of criminal organization.

Domestic production of alcoholic beverages is relatively large scale, requiring raw materials, production facilities, warehouses, and trucks, while liquor importation requires ships, warehouses, and trucks, and both need associated personnel. Above all was the need to protect the liquor businesses from predators who could hijack trucks and otherwise exploit a business that being illegal could not depend on law enforcement. Physical protection from the unparalleled surge of violence that accompanied Prohibition was essential for anyone seeking his fortune in the alcohol business.

The organizational pyramid that characterized pre-Prohibition crime was turned upside down – gang leaders pushed the politicians aside and became dominant players in many urban areas. Al Capone, for example, a Brooklyn-born thug became the single most important person in Chicago. In New York, a cabal of Italian and Jewish gangsters ruled a vast underworld and influenced city politics. In urban areas throughout the country, Prohibition changed the relationship between politicians, vice entrepreneurs, and gang leaders. Before 1920, the political boss acted as a patron for the vice entrepreneurs and the gangs: He protected them from law enforcement and they gave him financial and electoral support. The onset of Prohibition, however, unleashed an unsurpassed level of criminal violence, and violence is the specialty of the gangs. Physical protection from rival organizations and armed robbers was suddenly more important than was protection from law enforcement. Prohibition turned gangs into empires.

Prohibition encouraged cooperation between gang leaders from various regions – *syndication*. Domestic and imported liquor shipments were trucked intra- and interstate to warehouses at distribution points. At each juncture the shipment required political and physical protection. Only the criminal organization dominant in the local area could provide such protection. Syndication arose out of these needs, the result of meetings between important organized figures from throughout the United States.

Post-Prohibition Organized Crime

With the onset of the Great Depression (1929) and the subsequent repeal of Prohibition (1933), the financial base of organized crime narrowed considerably. Many players dropped out: Some went into legitimate enterprises or employment; others drifted into conventional criminality. Bootlegging required trucks, drivers, mechanics, garages, warehouses, bookkeepers, and lawyers – skills and assets that could be converted to noncriminal endeavors. For those who remained in the business, reorganization was necessary. When Prohibition ended, bootleggers were still young men – generally in their thirties – yet with wealth and nationwide contacts that had grown out of their liquor-related enterprises. In addition, they had substantial investments in restaurants, nightclubs, gambling, and other profitable businesses. In the 1930s and 1940s, they used their national contacts, diverse interests, and available capital to cooperate in a variety of entrepreneurial activities, legal and illegal.

One of these endeavors was labor racketeering. Leaders of organized crime engineered the domination of international labor unions by combining their control over union locals in their respective territories. Profits from Prohibition were jointly invested in the development of luxury casino hotels in Nevada and Cuba. When Prohibition was repealed in 1933, the United States (and the rest of the world) was experiencing the Great Depression. Ironically, a cash-starved society had a cash-rich criminal class. This led to the wholesale entry of gangsters into "banking": They became major loan sharks and, as a natural extension of this enterprise, assumed (often hidden) ownership of a variety of legitimate businesses.

Organized crime with roots in Prohibition remained a formidable entity into the 1960s, but the next decade was characterized by an irreversible decline. Although organized crime in the United States has a history that encompasses cities in every region of the country, outside of the Northeast and Chicago, there are few or no vestiges of the formidable criminal organizations that evolved out of Prohibition. The Irish, Jewish, and Italian communities from which these criminal organizations drew the bulk of their personnel failed to provide a continuous supply of replacements necessary for their survival. The Irish found success in politics and, although Jewish immigrants and their progeny played a vital role in the development of organized crime, by the third generation Jews had moved into business and the professions. Italian-American communities were sufficiently large enough to provide adequate replacement personnel in only a few cities, such as New York and Chicago. Even there, those aspiring membership in the American Mafia are a tiny fraction of the Italian-American community now solidly part of America's middle class and whose members are well represented in the ranks of professionals and public officials.

With the enactment of the Racketeer Influenced and Corrupt Organizations (RICO) statute, part of the Organized Crime Control Act of 1970, prosecutorial success reached unparalleled level of success. The long prison sentences of RICO encouraged disloyalty and American Mafia members routinely became government witnesses testifying against former colleagues, leading to a wholesale breakdown in discipline. In the meantime, new criminal organizations have emerged, eclipsing the remnants of the American Mafia. Some are domestic entities while many more are part of or have ties to criminal organizations outside the United States.

The New Players

In the 1960s, Colombians began immigrating to the United States in numbers sufficient to establish communities in Miami, Chicago, Los Angeles, and New York. Among them were criminals experienced in cocaine trafficking. Violence between rival Colombian "cocaine cowboys" brought terror to South Florida. When the gun smoke cleared units affiliated with the two major cartels in Colombia, Medellín and Cali, were dominant. Colombian cocaine traffickers eventually partnered with their Mexican counterparts who had long-established heroin networks in the United States. Both supplied Dominican organizations that wholesaled drugs to other groups further down the supply chain such as MS-13, a primarily Salvadoran organization with members organized into *clikas* active in different parts of the United States, particularly places with large numbers of Latino immigrants.

Since the collapse of the Soviet Union, Russian criminal groups have flourished in the United States in the absence of the corrupt political machines that provided a protective incubator for the Irish, Jewish, and Italian criminal organizations of an earlier era. They typically have military experience and many are college educated.

African American criminals dominated the numbers (illegal lottery) racket in cities such as New York, Philadelphia, and Chicago until they were overpowered by violent white gangsters such as Dutch Schultz (New York) and Sam Giancana (Chicago) who had superior police/political connections. Although blacks have traditionally been locked out of many activities associated with organized crime (e.g., labor and business racketeering) by prejudice, *dope is an equal opportunity employer*. African American criminal groups made important strides in the heroin business when the Vietnam War exposed many black soldiers to Far Eastern sources. Black criminal groups – African American, Jamaican, and Nigerian – use the drug trade as much as their Irish, Jewish, and Italian predecessors did bootlegging.

Added to the pantheon of organized crime are outlaw motorcycle clubs. The outlaw motorcycle club phenomenon dates to the years after World War II, when many combat veterans, particularly those residing in California, formed a motorcycle club that eventually became the Hell's Angels. From the fun-long, hell-raising clubs of the post-World War II era, the Hell's Angels and several other outlaw clubs, such as the Bandidos and the Outlaws, have parlayed their reputation for violence into self-perpetuating and disciplined organizations whose major source of income is from criminal activity, in particular, drug trafficking.

New players continue to emerge, prison-based gangs, Albanian organizations, and Chinese groups, as the history of organized crime in America continues to evolve.

Related Entries

- ▶ Careers in Organized Crime
- ▶ History of Organized Crime in Mexico
- ▶ Motorcycle Clubs
- ▶ Organized Crime, Types of
- ▶ Russian Mafia
- ▶ Social Network Analysis of Organized Criminal Groups
- ▶ Triads and Tongs

Recommended Reading and References

Abadinsky H (2013) Organized crime, 10th edn. Wadsworth, Belmont

Albini JL (1971) The American Mafia: genesis of a legend. Appleton, New York

Anderson AG (1979) The business of organized crime: a Cosa Nostra family. Hoover Institution Press, Stanford

Block AA (1979) East side–west side: organizing crime in New York, 1930–1950. Christopher Davis, Swansea

Coen J (2009) Family secrets: the case that crippled the Chicago Mob. Chicago Review Press, Chicago

Cressey DR (1969) Theft of the nation. Harper & Row, New York

Fox S (1989) Blood and power: organized crime in the twentieth century. William Morrow, New York

Haller MH (1990) Illegal enterprise: a theoretical and historic interpretation. Criminology 28(May): 207–235

Jacobs JB (2006) Mobsters, unions, and feds: the Mafia and the American labor movement. New York University Press, New York

Landesco J (1968) Organized crime in Chicago. University of Chicago Press, Chicago, Originally published in 1929

Maltz MD (1976) On defining organized crime. Crime Delinq 22(July):338–346

Moore MH (1987) Organized crime as a business enterprise. In: Edelhertz H (ed) Major issues in organized crime control. GPO, Washington, DC, pp 51–64

Nelli HS (1976) The business of crime. Oxford University Press, New York

Raab S (2005) Five families: the rise, decline, and resurgence of America's most powerful Mafia Empires. St. Martin's Press, New York

Reuter P (1995) The decline of the American Mafia. Public Interest (Summer) 120:89–99

Smith DC Jr (1982) Paragons, pariahs, and pirates: a spectrum-based theory of enterprise. Crime Delinq 26:358–386

History of Police Unions

Joanne Klein
Department of History, Boise State University,
Boise, ID, USA

Synonyms

Associations; Fraternal organizations; Guilds; Police federations; Societies

Overview

Giving police officers a voice within police organizations raised vexing questions. In many nations, police officers were to "police by mutual consent" with the public but often did not experience mutual confidence between senior officers and the rank and file. They needed a satisfactory outlet for their grievances, an advocate to protect their interests, and input into their occupation. A union could hinder arbitrary power from above, but authorities worried that unions undermined the organization. However, officers without an authorized union worked at the mercy of superior authorities and were more prone to strike.

The earliest organizing efforts in nineteenth-century Europe, America, and Australia aimed not only at pay, pensions, and regular days off, but also at the "right to confer." They insisted that officers had the right to be consulted by police authorities on policy and conditions of service. Much of western Europe, Scandinavia, and Australia moved toward legalizing police unions during the early twentieth century. However, they remained prohibited in Britain and America. World War One put the police under immense economic and work stress, ending in strikes in Cape Town, London, Liverpool, and Boston. After 2 years with strikes, the British government responded by creating a police federation and with a comprehensive Police Act (1919). The United States went in the opposite direction, banning unions into the 1950s. American police had to rely on mutual aid and fraternal organizations to protect their interests. This adversarial attitude toward police unions persisted well after the unions began to be legalized after World War Two.

The Civil Rights era revived existing and new police unions in response to calls for reform and integration. Police union rights were linked to broader organizing rights. While still concerned with pay and conditions, unions began lobbying policy makers to improve internal conditions and address external issues. They used a variety of strategies, both old and modern. "Work to rule" and slowdowns continued. Media campaigns won over public and political support. This created concerns that unions were threatening police political neutrality, particularly after scandals where police unions and party leaders made secret deals promising support in return for specific policies. Recently, more positive directions for police unions included working for affirmative action across the African continent, and pioneering shared leadership programs in the United States. Challenges for police unions persisted. They earned a reputation for being reactionary and inflexible, though unions took steps toward cooperating in reforms and community issues in constructive ways. Research on police unions is sparse, and mostly focuses on recent decades. Scholars tend to have pessimistic assessments, but lately have explored linkages between workplace rights and civil rights. The picture, however, remains incomplete.

Fundamentals

Labor unions developed to protect basic worker rights such as fair pay, safe work conditions, and the right to confer, as well as address concerns unique to each occupation. While no universal model of police union existed, they typically focused on (1) pay, pensions, and allowances, (2) conditions of service, including discipline and promotion, (3) the right to confer, and (4) legal reforms and policies. Police officers first joined together not just for improved pay but also to end the tyranny of superior officers and authorities who refused to listen to them. The earliest unions already focused on the need to replace arbitrary power and corruption with fair standards and procedures for promotion and discipline. Always uncomfortable about the idea of the police "downing tools," they typically adopted "no strike" policies only abandoned under acute stress. World War One sparked police strikes in cities where police officers were overwhelmed by war-time inflation, and police unions were banned. In the twentieth century, police forces applied themselves to becoming a profession. Unions became more like white-collar associations that faced the similar challenge of trying to be professional but also address typical union concerns. Lobbying became more central after World War Two as police forces responded to Civil Rights movements and demands for police reform and integration. Forces slowly became more diverse with more women and minorities, yet minority groups continued to see a need for their own advocacy groups. Some unions became highly political, creating challenges for police organizations that insisted that they were politically impartial. These changes in police unions reflected changes in police officers themselves; nineteenth-century working-class policemen in strictly regimented departments gradually became in the twentieth century better educated and more diverse men and women in more complex, though still hierarchical, organizations.

Every police union had its own trajectory. They were diverse in size, goals, and power, over time and space. Police forces could have unions, but many had federations, associations, guilds, fraternal organizations, or societies that might carry out similar functions as a union and had greater or lesser authority depending on local conditions. Groups could be local, national, or international. Multiple groups coexisted. What kind of organization existed depended on laws and circumstances. For simplicity's sake, the terms "police union" and "union" will be used except when a specific group is mentioned. Australian, Dutch, French, and Scandinavian police unions were legal throughout the twentieth century but varied in size and style. When police officers were not allowed to unionize, they often set up societies and clubs. In 1918, Cape Town police had a committee that organized a successful strike, and in 1919, the Boston Social Club transformed itself into the Boston Police Union. After 1919, British police could not join unions, particularly not the National Union of Police and Prison Officers; its official replacement, the Police Federation, attained negotiation powers in the late 1940s. New Zealand followed a similar path, attempting to unionize and ending up instead with an official Police Association in 1935. In the United States, the Fraternal Order of Police was allowed while police unions and collective bargaining remained illegal until the 1960s. Today, US unions are highly local, Australia has state unions, and the Republic of South Africa has two national unions. Unions might be associated with a national federation or association, or remain unaffiliated. Only recently did police officers attempt to set up unions in authoritarian nations or nations transitioning from dictatorships to more open regimes, such as in Argentina and Swaziland where such unions were declared illegal. In recent decades, unions affiliated directly or indirectly with the International Council of Police Representative Associations (ICPRA), formed in 1996 to promote professional standards and police civil rights.

The key function of these unions was to give a voice to the rank and file who often felt they had no control within their occupation. Strains existed between upper and lower ranks over promotion and discipline, which could seem arbitrary and corrupt. Relations between unions and police authorities could be confrontational, friendly, or anywhere in between, affecting cooperation and communication. At the very least, unions provided a safety valve where officers could air grievances and blow off steam. Unions generally had a decent record on improving pay and protecting gains, though they were still at the mercy of economic downturns. They fought to create and protect fair procedures for promotion and discipline. They provided support during disciplinary cases, varying from advice to legal services. However, a simple idea such as improving conditions for senior patrol officers demonstrated how entangled issues could become. Suggesting rewards such as “master patrol officer” or favorable transfers created new problems balancing seniority, pay equity, and education. Unions sometimes got bewildered over how to improve education standards and attract quality recruits without undermining the pay and status of older officers. As unions got stronger in recent decades, police authorities and the public raised concerns that they had become too insular, defensive, and resistant to change. On the other hand, examples of dynamic, constructive unions existed. The question is whether this is the direction for the future or a few exceptions.

Most unions included the lowest two or three ranks so they represented the majority of, but not all, police officers. Higher ranks had their own organizations. Some unions included retired officers and a few included civilian staff. The New Zealand Police Association was unusual in including civilian staff and all ranks except the very highest. Where legal, joining the union was fairly automatic. Where official bodies existed, such as the Police Federation of England and Wales, joining the police meant joining the Federation. How active members were depended on how useful they considered their union to be. Alongside unions, societies existed to represent specific groups. Police authorities banned or tolerated them depending on their behavior and local attitudes. Initially, these tended to be linked to religion, such as Catholic or Protestant associations which could get themselves into trouble if they became too fractious. As forces diversified in the twentieth century, women’s and ethnic minority associations appeared, and more recently, groups for gay police officers. Religious groups persisted, but with greater variety. These organizations paralleled unions and often prodded them to address entrenched sexism, racism, and other forms of discrimination. Their relationships with both police forces and police unions could be acrimonious given often defensive police attitudes.

The legal status of unions depended on national and local laws. Many western nations authorized police unions early in the twentieth century. Police unions in the Netherlands, Scandinavia, and parts of Australia were permitted before World War One. French police unions began in 1905, to be legalized in 1926. Australian unions formed state by state until every region had one in 1945, the same year that the Police Federation of Australia and New Zealand was formed. German police unions formed during World War One and were active under the Weimar government, only to be shut down under the Nazis. Some nations barred unions but allowed “federations” or “guilds” that fulfilled similar or reduced roles. After police strikes in 1919, police unions were banned in Britain and America. The British set up the Police Federation of England and Wales while in America, the Fraternal Order of Police (FOP) provided a limited alternative. Despite deep antagonism from police chiefs who viewed unions as sources of corruption, tiny unions persisted, some affiliated with the FOP or the National Conference of Police Associations. Finally, American police unions were legalized in the 1960s though not all gained collective bargaining rights. They remained local and diverse, similar to their police forces, with most acting independently to protect local interests. Unlike the older unions in Europe and Australia, American unions tended to be more reactionary and antireform. To the north, provincial Canadian police had unions with collective bargaining rights, but as recently as 1999, the Canadian Supreme Court ruled that the Royal Canadian Mounted Police did not have the right to unionize. Instead, they had official staff relations representation which loosely resembled the British Federation but without its authority. Finally, unions appeared in emerging democracies such as the Republic of South Africa. The Police and Prisons Civil Rights Union (POPCRU) formed in 1989, and it continued to focus on civil rights issues both at home and across the continent. The South African Police Union was more moderate and took an apolitical approach. In 2003, two Argentine police groups asked the International Labour Organization (ILO) for support to address poor working conditions and tyrannical superior officers. However, the ILO limited the rights of police officers to unionize if such unions were banned by national law and did not back them (Marks and Fleming 2006). Attempts by Swaziland police to unionize met similar barriers. Police authorities continued to argue that unions undercut their authority, but forces with a long history of legal unions had better morale and less insubordination than forces without unions.

Early Unions and Police Strikes

The first police unions appeared in the later nineteenth century and their forms roughly reflected unionism generally. Early efforts focused on pay and conditions including training, uniforms, and work shifts, and internal questions such as discipline, tyrannical bosses, and basic fairness. They dealt with local and sometimes national governments since changes could require new laws. In areas without unions, mutual aid societies, such as the New York City Patrolmen's Benevolent Association (1892), and rank-and-file newspapers, such as Britain's *Police Review and Parade Gossip* (1893), provided police forums. With or without a union, police officers rarely supported the idea of striking, so other tactics had to be explored. Early actions included petitioning authorities, "work-to-rule" protests, and refusals to parade. An early protest in 1853 Manchester, where about half the force resigned over low pay, had shock value but had clear strategic flaws. More commonly, officers staged what amounted to "sit-down" strikes, refusing to leave the station until their grievances were heard. These could be effective though risky. London officers staged a protest demanding pensions in 1890. The 1890 Police Act did give police officers pensions, but protest leaders and 39 protesters were dismissed (Allen 1958). By the early 1900s, informal efforts became unions, police officers learning that unions were more likely to be taken seriously. If the unions were recognized, actions that might threaten dismissal were unlikely to be required. However, areas that resisted police unions left police officers with few useful options to get their voices heard.

The most prominent time of police strikes was during and immediately after World War One when war-time responsibilities and inflation put police officers under severe stress. The police officers resorted to strikes as a last resort after months of having their concerns ignored. Bad pay was a factor but abuse of power as well. None had recognized unions. The 1918 strikes in London and Cape Town, South Africa, had an advantage of taking place during the war, making governments eager to resolve the crises quickly. In January 1918, white policemen in Cape Town organized a strike over pay, understaffing, and overwork after months of being ignored by the Minister of Justice. Organized by a committee, a small group of younger, single officers struck, both protecting married men and keeping enough men on duty to cover beats. The strikers were imprisoned, provoking more men to strike. Eventually 100 out of 500 men struck, but order continued. The public sympathized with the strikers who acted with restraint, and at least in part after learning that patrolmen's pay had eroded so much that they had to live in "respectable" colored neighborhoods (Nasson 1992). The strike lasted 2 weeks when all the men returned to duty and their war bonuses were increased. In 1919, a commission was formed to improve their conditions of service. In August 1918, the London Metropolitan and City of London Police also went on strike. The strike began for similar reasons, with the addition of concerns over "victimization," and after being similarly ignored. The London police, however, had an unrecognized union, the National Union of Police and Prison Officers (NUPPO), and the firing of a union member sparked the strike. Effectively both forces went out, yet order did not break down. The government settled within a week, with police commissioners forced to admit ignorance of the dire circumstances of their own men and the public perceiving the police as having just grievances. Similar to Cape Town, no men were fired, pay and war bonuses increased, and a commission was called to examine police conditions. These two strikes ended successfully for the police officers. The war forced governments fearful of Bolshevism to respond quickly, and the quiet streets kept the public's support.

By 1919, with governments no longer constrained by war, strikes became less successful and more violent. Victory in the 1918 London strike spread NUPPO branches throughout Britain. Police officers elected union members to the Desborough Committee, the commission on police pay and conditions. However, union leadership made excessive demands, forgetting that with war over, the government had no reason to bend. While many British police were content with improved pay and willing to hear the Desborough Committee's recommendations, Liverpool police feared entrenched favoritism and internal corruption which persisted unabated. In August 1919, half the force went on strike, leading to 3 days of rioting. Nearly 1,000 strikers were dismissed. The NUPPO was suppressed and membership declared illegal. While the Liverpool strike failed, it did result in important additions to the 1919 Police Act regarding conditions of service, promotion, and training. At the same time, Boston policemen were arguing with their mayor over pay and conditions, and getting nowhere. Inspired by the English strikes, they formed a union recognized by the American Federation of Labor. Other American forces took similar actions. In September 1919, the Boston Police Commissioner banned their union, leading to a strike remarkably similar to Liverpool. After rioting and looting, over 1,000 strikers were dismissed. President Wilson condemned the strike as a "crime against civilization" (Russell 1975). Both the Liverpool and Boston strikes turned the public against police unions because of the violence. However, British and American authorities responded to the strikes quite differently. The Boston strike seemed to confirm US police chiefs' worst fears of Bolshevism infiltrating the police and left them wholly against police unions for decades. Other police forces that had been forming unions gave up the idea. Britain created the Police Federation, which for all its weaknesses recognized that officers needed representation and an outlet for grievances.

The 1918–1919 strikes highlighted a central question of police unions – police officers demanding the right to confer and police authorities fearing that unions undermined their authority. In the United States particularly, police chiefs adamantly opposed any kind of police unions. They argued that public servants could not be linked to outside authority, unions undermined discipline, and unions obstructed change. Yet ironically, police officers barred from organizing were more likely to strike while unionized officers were more willing to negotiate. The reality was that police officers were going to organize, so allowing them to organize openly and legally was a better choice. By banning police unions and portraying them as threats, US police forces may have created an acrimonious relationship that continued to echo after unions were legalized in the 1960s. Police authorities were correct to worry that unions threatened their personal discretion. Unions were more likely to be illegal in nations that stressed a strong, top-down management style of policing. Police officers organized out of a sense of powerlessness against arbitrary superior officers and worked to create due process. Later unions often became reactionary to protect these gains and remained fearful of arbitrary authority.

Unions and the Professional Police

During the twentieth century, police forces focused on earning recognition as a profession despite a lack of clear direction. Policing did not require a professional degree, forces often lacked mutual standards, and transferring between forces was difficult. The role of police unions in this shift remained ambiguous. Some maintained a union identity while others became more like professional associations. When tangible gains such as pay were more secure, the status of being professional could be appealing. Unions often shared the confusion over their identity, wanting to be perceived as professional organizations yet with the collective bargaining powers of unions. More important than semantics, the rank and file argued that they had positive contributions to make to policy and service conditions. In Europe and Australia, police unions were included in these deliberations while in the United States, typically they were not. When the Civil Rights era brought police forces under criticism, the union model revitalized and unions became more assertive. Nevertheless, professionalism remained important, stressing efficiency and high standards.

In the 1960s, police unionism surged, with existing unions becoming more active and new unions forming, particularly in the United States. But unions gained a reputation as reactionary and defensive, seeing calls for reform as attacks on their profession. They viewed change as an admission of failure, even as giving in to disorderly and rebellious protestors. The 1960s and 1970s were a low point for American and British unions, made up largely of white, conservative men who not only resisted reform but framed their refusals to diversify as defending their rights. The Fraternal Order of Police opposed affirmative action, requiring the Afro-American Patrolmen's League to protect black officers. In Atlanta, Georgia, in the 1970s, the two groups filed lawsuits and countersuits over promotion and job assignment issues. After 3 years of stalemate, all sides agreed on a policy of treating all officers on an equal basis (Lewis 1989). Bringing women into police forces presented unions not only with challenges to their masculine image, but more importantly, unions feared that women could be used as an excuse to erode pay gains. Most unions opposed women though some supported equal pay for women to protect the pay of men. The domination of police unions by white men explained the importance of minority organizations to protect the interests of minority officers and to prod the majority toward fairer policies.

Clearly, police unions deserved many of the complaints that they blocked innovations and calls for change. From the unions' perspective, they used collective bargaining and similar tools to protect hard-won gains. From outside and from above, they looked intransigent and outdated. In the 1970s, some US unions supported education incentive plans, but many opposed them for violating prized "equal pay for equal work" standards. Overall, US unions focused more on pay issues and less on the larger picture that a better education was better for policing (Juris and Feuille 1973). Union resorted to denying all problems in response to criticism for ineffectiveness, discrimination, or corruption. The standard approach was to launch public relations campaigns, present the "bad apple" idea, and even inhibit investigations. If problems were found, unions insisted that they be dealt with by the police. The London Metropolitan Police refused to acknowledge that they had problems with racism despite numerous commissions, investigations, and court judgments in the 1970s and 1980s. Only in 1999 when the Stephen Lawrence Inquiry Report accused them of "institutional racism" did the force admit that they needed to make changes. Support or opposition to change tended to rest on how successfully the rank and file was included in discussions. Both civilian reformers and superior officers could be patronizing and rude to police officers, viewing them as poorly educated and unable to make thoughtful decisions. When unions responded defensively, that confirmed the stereotype and perpetuated the problem.

Police unions became acutely worried whenever civilians got too close to policing's internal affairs. Replacing police officers with civilian staff could look like efficiency to reformers but like diluting the work force and a cost-cutting measure to police officers. The greatest threat was civilian review boards which sprang up in the 1970s in response to scandals as well as civilian activism. Unions loudly objected to civilian boards, seeing them as ignorant nuisances. The Police Association in Victoria, Australia, managed to shut down a Police Complaints Authority in the 1980s, and police unions in the United States fought hard against ombudsmen and civilian boards (Finnane 2008). Nevertheless, civilian review spread and unions had to adjust. Community policing programs, where civilians augmented the police, could also be problematic when local police forces were not adequately included. On one side, police unions resisted involving the community, equating such ideas with budget cuts and interference. On the other hand, civilians lacked creativity in including the police as partners in community policing (Sklansky and Marks 2008). Where police unions were included in setting up community policing programs, police fears of outside intrusions could be allayed. Chicago incorporated key terms regarding management and pay into its program. The program was a success and the local FOP endorsed it (Walker 2008). Broken Arrow, Oklahoma, set up a police shared leadership program in cooperation with the police union, a hint of what was possible when communications were open.

Police unions developed strategies to get their views across without disrupting order and without big strikes. These fell into four categories of on-duty action, off-duty action, media, and lobbying with two basic goals of affecting contract negotiations and government policy. Sometimes, a specific issue related to a particular incident provoked union action, such as when Pittsburgh police officers caught the "blue flu" to protest how the force handled transferring white and black officers in and out of inner-city neighborhoods in April 1970 (Juris and Feuille 1973). Work actions such as sick-outs and mass car servicing were planned not to endanger the public. Other options were "go-slows" and work-to-rule campaigns. Traffic ticket slowdowns reminded authorities that the police provided important revenue and were popular with the public. Detroit staged one in 1967 and Brisbane, Australia, in 1976, and both made their point quickly. Following a shift by trade unions generally, police officers began using brief 1- or 2-day strikes designed to force negotiations or policies. In May 1970, Rochester police staged an 8-h strike which resulted in negotiations resuming though a similar strike in Milwaukee in January 1971 failed in part due to public apathy (Juris and Feuille 1973; Fleming and Lafferty 2001). But more short strikes worked than failed, and they continued into the twenty-first century. Brazilian police in 2001, Dutch police in 2007, and Slovenian police in 2010, all staged short strikes over pay. The threat of major walkouts remained if police authorities misread their officers. In 1969, Montreal police and firefighters struck over pay but also over job hazards created by separatist terrorists resulting in looting, bank robberies, and violence. Unions preferred not to strike, but kept the option in reserve.

Off-duty union efforts ranged from public meetings and marches to letter-writing and media campaigns. When Dutch police were unhappy with pay negotiations in 2007, for example, thousands of off-duty officers staged noisy protests outside the Ministry of the Interior (Berry et al. 2008). Such efforts could work though could take patience to bring results. In the 1950s, the New York City Police Benevolent Association (PBA) evolved into a union role through a variety of media and legal actions. In 1958, when the mayor refused to negotiate a grievance procedure, the PBA published an "Open Letter to the Mayor of the City of New York" as a newspaper advertisement. They petitioned the city Department of Labor and finally appealed to the US Supreme Court. The PBA lost their appeal but proved to be tough advocates. In 1961, New York established a police grievance system. The political environment shaped what a union could expect. The 1950s New York Police operated in a city hostile to unions, where the mere threat that a union might be forming convinced the city to improve conditions (Levi 1977).

In theory, most police forces declared themselves to be politically neutral. Despite this, increasingly police unions took part in the political campaigns of candidates or parties with policies favorable to the police. Issues included improving police numbers, modernizing equipment, and reforming laws. This could be positive or negative on numerous levels with unions protecting their interests, working for reforms, and risking conflict of interest. An early example of a campaign came when the *Police Review* helped lobby Parliament to give the police 1 rest day in seven which resulted in the Police (Weekly Rest Day) Act in 1910. After World War Two, unions focused on getting sympathetic candidates elected at the local and national levels. They risked their apolitical reputations despite frequent protestations that their actions were impartial.

Dutch unions proved good at campaigning without losing respect, but Australian unions made some embarrassing miscalculations. In 1996, the Queensland Police Union was determined to unseat the local Labor government and campaigned against them successfully in a by-election. However, a "memorandum of understanding" between the union and opposition leaders came to light, promising policies favorable to the police. This provoked outrage and damaged the image of the union which was forced to reconsider both its strategies and its democratic values (Finnane 2008). Unions did not necessarily commit to one party. In the late 1970s, the British Police Federation conducted a law and order campaign that matched the policies of the Thatcher administration. However, by the 1990s, the Federation rebuked the Conservatives for saying that the police's primary aim was fighting crime, insisting that their first priority was protecting life and preventing crime (Berry et al. 2008). Unions continued to grapple with their declared political neutrality, yet with widespread involvement in political lobbying.

Openly political police unions existed, typically with multiple unions for police officers to choose from based on ideology. These could be motivated by one idea, such as POPCRU which spotlighted civil rights. The French police had competing unions aligned with local and national political parties. Since the 1970s, French governments consulted with unions over reform and policy, causing many police officers to wonder if their unions had become just another factor in corruption and favoritism problems. Then again, French police unions spoke out against police abuses of power, and supported creative reforms and modernized equipment (Horton 1995). They provide an alternate model to the apolitical one that could often be unrealistic. However, becoming too political had its own dangers. In India, ruling elites tended to use police as an arm of their governments, and allowed police abuses of power. In the 1980s, the police union in Gujarat turned the tables, becoming too powerful and blackmailing the local government. The union threatened not to police during confrontational annual holidays unless its demands were met. In 1988, the government refused to be coerced. The union called a strike, but its leaders were quickly dismissed and the union disbanded. Only in 2006 did Indian courts allow the union to reorganize. However, the problematic relationship between governments and police remained unresolved (Joshi 2000; Aggarwal 1988).

Police unions remained divided over the advantages of local organizations versus a more national focus. American unions often operated in isolation from each other while Scandinavian police unions supported nationalizing. Nevertheless, increasingly unions at least talked to each other, sharing information on pay and conditions to help with negotiations and working to improve professional standards. In 1996, acknowledging an increasingly global society, the International Council of Police Representative Associations (ICPRA) was founded with momentum from the Canadian Police Association. Its basic goals reflected a persistence of nineteenth-century concerns – open communication, consistent and fair procedures, proper training and equipment, political neutrality, and the right to freedom of association. Gaining full civil rights for police officers remained an issue; many police officers could not vote in the nineteenth century and many nations denied them suffrage in the twenty-first century. Their objectives included expanding collective bargaining and helping police forces establish their own associations. Overall, their agenda was practical, including improving qualifications of officers, standards of police methods, and professionalism. In 2012, ICRA was helping the Swaziland Police's fight to unionize and supporting Guardia Civil to gain rights being limited by the Spanish government. Membership was limited to national police associations, in part to encourage local unions to work together. Members in 2012 included Australia and New Zealand, Canada, the Fraternal Order of Police, Ireland, the United Kingdom, Portugal, Denmark, and the European Confederation of Police. The two Republic of South Africa unions were the only ones from Africa. No members came from Latin America, the Middle East, or Asia. However, IPRCA and other organizations were supporting efforts in those areas to unionize police forces.

Historiography and Future Directions

The historiography of police unions is unexpectedly thin. Official publications, such as the American *Police Unions and Other Police Organizations*, issued in 1919 and reissued in 1944, took a negative view of police unions as undermining policing's core values. With the exception of V.L. Allen's "The National Union of Police and Prison Officers" in 1958, the first substantial scholarship came out in the 1970s. Robert Reiner's sociological study, *The Blue-Coated Worker*, was based on interviews of British police officers, including their views on the Federation and police unions. In the United States, studies such as Hervey Juris and Peter Feuille's *Police Unionism* explored unions from the perspective of industrial relations and collective bargaining. Only in the 1990s and 2000s did historians begin focusing on police unions, notably Mark Finnane's *When Police Unionise* on Australia. Books on police strikes were more common but they did not focus on unions and they tended to be for a popular audience, such as Francis Russell's *A City in Terror* (1975) on the Boston strike which focused largely on Calvin Coolidge. Numerous articles came out, but they tended to focus on contemporary issues rather than history, with a few exceptions such as Nasson's 1992 article on the 1918 police strike in Cape Town. Articles on Swedish and Belgian police unions were not readily accessible to English readers. Unions were covered as part of larger works but typically only for a few pages. A large gap in the scholarship exists for history focused specifically on police unions.

Scholars of police unions can roughly be divided into pessimists, pragmatists, and optimists. Pessimists viewed unions as reactionary, conservative, and undermining authority. Most historians tended to take this stance, including many US police historians as well as sociologist Robert Reiner. Recently, pragmatists, notably Mark Finnane and Samuel Walker, reminded readers that the first priority of unions was to serve their members. The optimists were a new group, arguing that unions could move the police out of stale models and cooperate with the community in constructive ways. In 2006, many police scholars came together at an international, interdisciplinary conference on "Police Reform from the Bottom Up" at the University of California, Berkeley. It resulted in special issues of *Policing & Society* (March 2008) and *Police Practice and Research* (May 2008) which focused on ways that police unions could participate in reform, as well as be obstacles to such a role. The authors made an effort to move away from the dominant pessimistic view and explore more nuanced interpretations of police unions, as well as explore why the negative view of unions existed. Even here, though, the articles focused more on recent decades than on before the 1960s.

Union scholarship was often cast as a struggle for control between police authorities and the rank and file. One question explored was whether police unions weakened or strengthened police forces. They could hurt forces by undermining discipline or help them by breaking down a military mentality. More pessimistic historians argued that police unions became part of the military mentality, resisting civilian oversight and outside investigation. They pointed out the reactionary and insular behavior of unions against diversity and innovation. However, scholars remained uncertain on whether unions challenged top-down control or reinforced it (Marks and Sklansky 2008). In 1996, a study found that police officers were more satisfied with their jobs and had a better relationship with the public with less heavy-handed management, and that unionized officers reported better community relations. They concluded that shedding the "top-down" approach could be a good idea for both the police and the public (Magenau and Hunt 1996). For all their flaws, Samuel Walker pointed out the importance of unions in shared governance and the move away from authoritarian models of policing in America (Walker 2008). The tendency to view police unions negatively needed to be corrected and contextualized. Mark Finnane was also less epic in his expectations of police unions either for good or evil. Their primary task was to look out for their members' pay and pensions, and to be advocates in discipline and promotion cases.

In the 2000s, a democratic theory of police unions was advanced which remains controversial. Monique Marks and Jenny Fleming favored the argument that police unions needed to shift from the military model to a modern civilian, democratic model, and become agents of positive change. Fleming especially argued that police unions helped break down the militaristic attributes of police culture. Their idea was that unions helped police officers understand the importance of workplace democracy and so became willing to protect democracy more generally. Unions provided rights such as association, free speech, and voting, and supported work rights like collective bargaining, fair pay, and impartial discipline. Police officers who valued their own rights would protect the rights of others. The theory leaned on examples from the developing world, particularly POPCRU in the Republic of South Africa. Its applications tended to be more for recent and future developments. Dutch and Swedish police unions supported POPCRU in the 1990s, and in 2007, POPCRU began supporting Swaziland and Mauritian police officers working to unionize. Similarly, police unions from Australia and New Zealand worked with police associations in the Cook Island, Papua New Guinea, and Fiji (Marks and Fleming 2006). Whether unions helped police officers link workplace rights to broader human rights historically will require additional scholarship.

More research remains to be done on police union history to build a more complete picture of their role in police history. Clearly police officers will organize regardless of its legality. Why parts of Europe and Australia were comfortable giving the police rank and file a voice while Britain and the United States resisted is a comparison requiring further investigation. Australia suggests that having a long history of legal unions did not necessarily make a difference in community relations and openness to innovation and reform, but within Australia, this varied from state to state. American forces condemned unions for so long that they may have created the pent-up frustrations that burst out in the 1970s. However messy unions could be, having police unions decreased acrimony and improved morale. What is perhaps most startling is that the early concerns of police unions from the nineteenth century share so much with the 2004 goals of the International Council of Police Representative Associations. Unions continue to complain about poor relations with their superiors, just like unions in the nineteenth century. The central concern remains the right to confer.

Related Entries

- ▶ Communities and the Police
- ▶ Community Policing
- ▶ Comparing Police Systems Across the World
- ▶ Democratic Policing
- ▶ Managing Innovation of Policing
- ▶ Minorities Within the Police Workforce
- ▶ Women Police

Recommended Reading and References

Aggarwal Y (1988) Gujarat police strike. Econ Pol Wkly 23(32):1611–1612

Allen VL (1958) The National Union of Police and Prison Officers, 2nd series. Econ Hist Rev 11(1):133–143

Berry J, O'Connor G, Punch M, Wilson P (2008) Strange union: changing patterns of reform, representation, and unionization in policing. Police Pract Res 9(2):113–130

Finnane M (2002) When police unionise: the politics of law and order in Australia. Institute of Criminology, University of Sydney, Sydney

Finnane M (2008) No longer a 'workingman's paradise'? Australian police unions and political action in a changing industrial environment. Police Pract Res 9(2):131–143

Fleming J, Lafferty G (2001) Police unions, industrial strategies and political influence: some recent history. Int J Employ Stud 9(2):131–140

Furuhagen B (2012) The police as a municipal or a state agency? The Swedish police in a Scandinavian comparative perspective 1900–1965. Unpublished paper, European social science history conference, April 2012. Uppsala University, Sweden

Horton C (1995) Policing policy in France. Policy Studies Institute, London

International Council of Police Representative Associations. http://icpra.org

Joshi R (2000) Democratising the police: lessons from UK. Econ Pol Wkly 35(40):3589–3596

Juris HA, Feuille P (1973) Police unionism: power and impact in public-sector bargaining. Lexington Books, Toronto

Klein J (2002) Blue-collar job, blue-collar career: English policemen's perplexing struggle for a voice in the early twentieth century. Crime, Histoire & Sociétés/Crime Hist Soc 6(1):5–29
Levi M (1977) Bureaucratic insurgency: the case of police unions. Lexington Books, Lexington
Lewis WG (1989) Toward representative bureaucracy: blacks in city police organizations, 1975–1985. Public Adm Rev 49(3):257–268
Magenau JM, Hunt RG (1996) Police unions and the police role. Hum Relat 49(10):1315–1343
Marks M, Fleming J (2006) The untold story: the regulation of police labour rights and the quest for police democratisation. Police Pract Res 7(4):309–322
Marks M, Sklansky D (2008) Voices from below: unions and participatory arrangements in the police workplace. Police Pract Res 9(2):85–94
McGill D (1992) No right to strike: the history of the New Zealand police service organizations. Silver Owl Press, Wellington
Nasson B (1992) 'Messing with coloured people': the 1918 police strike in Cape Town, South Africa. J Afr Hist 33(2):301–319
Reiner R (1978) The blue-coated worker: a sociological study of police unionism. Cambridge University Press, Cambridge
Russell F (1975) A city in terror: Calvin Coolidge and the 1919 Boston police strike. Beacon, Boston
Sklansky DA, Marks M (2008) The role of the rank and file in police reform. Polic & Soc 18(1):1–6
van der Wal R (2007) De geschiedenis van de Nederlandse politie: De vakorganisatie en het beroepsonderwijs. Boom, Amsterdam
Walker S (2008) The neglect of police unions: exploring one of the most important areas of American policing. Police Pract Res 9(2):95–112

History of Probation and Parole in the United States

James T. McCafferty[1] and
Lawrence F. Travis III[2]
[1]Department of Sociology and Criminal Justice, Kennesaw State University, Kennesaw, GA, USA
[2]School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA

Overview

Probation and parole are two types of community supervision that form the "bookends" of corrections in America. Probation is a community supervision sanction imposed in lieu of incarceration, while parole involves supervision of individuals released from incarceration. Together, these two broad types of supervision take on the lion's share of criminal clients under the formal control of the criminal justice system. Designed to surround incarceration, these correctional settings offer offenders conditional release into the community. Offenders receive formal, written supervision conditions, and a supervision officer monitors their progress and behavior in the community. An offender who commits a new crime while on community supervision or violates the supervision conditions may be placed into an institutional setting. Within probation and parole, offenders may be managed differently. Offenders who commit more serious crimes and are also considered high risk may be placed on higher levels of supervision that have stricter conditions, as opposed to less serious offenders who may not be monitored as closely.

The current systems of probation and parole developed in America during the mid-nineteenth century; however, the historical purposes and use of these systems of community management have European origins. As an alternative to incarceration, justice officials have traditionally used probation and parole for offenders that do not need to be incarcerated. By monitoring these criminals outside of prison, the community supervision agencies can help alleviate prison crowding and keep criminal justice costs down. Supporters believe that parole can also have a positive benefit by influencing prison behavior by acting as an incentive for good behavior. Scholars, politicians, and the public have attacked community supervision for being ineffective or "soft" on crime. Research suggests that with proper identification of risk and appropriately matched correctional interventions, community supervision can be effective at keeping offenders from recidivating and reducing correctional costs.

Like any other area in criminal justice, probation and parole present issues of financing, management, and newer issues concerning the role of technology in supervision. In response to greater numbers of people under correctional control and strained budgets, community supervision is most likely at the beginning of the next phase of its evolution: increasing the effectiveness of supervision through proper identification of risk and needs of offenders, while matching clients to the appropriate level of supervision and keeping criminal justice costs to a minimum.

Introduction

Today, imprisonment is an accepted and frequently used criminal sanction; however, the criminal justice system supervises a majority of criminals in the community. Nearly 7.1 million people were under correctional authority at the end of 2005. Community supervision monitored nearly five million of those people (Glaze and Bonczar 2006). For as long as statistics on the correctional population have been tracked, about two-thirds of those under correctional control have been supervised on probation and parole.

Taken together, probation and parole are the "bookends" of incarceration in America because these systems "surround" incarceration. Probation is a community supervision sanction imposed in lieu of incarceration, while parole serves to supervise individuals released from incarceration. Offenders placed on formal supervision must abide by specific conditions and must report to an officer responsible for tracking their progress and behavior in the community. Should the offender violate these conditions, the court or paroling authority can revoke the community supervision status. If found to be in violation of supervision conditions, the resulting violation (or revocation) hearings often lead to stricter community supervision conditions or incarceration.

Many people consider incarceration, especially the prison, to be the cornerstone of correctional intervention in the United States. While America's prison population has drastically risen over time, probation and parole are still responsible for a majority of offenders under formal control. As prisons become increasingly crowded, it is important to address alternative sanctions to incarceration because community corrections will receive many criminals that otherwise might have been sent to prison. This entry will discuss the historical origins of community supervision and their function within the community. It will also address important contemporary issues confronting community corrections including effectiveness, appropriate identification of criminals' needs, financing, and the role of new technology in supervision. This entry will conclude with a discussion of the future of probation and parole.

The Development of Probation and Parole

Probation and parole developed in the United States during the nineteenth century; however, these systems are not uniquely American inventions. During the Middle Ages and early Modern Era, feudal law would often require a strict sentence of corporal or capital punishment for criminal behavior. Over time, due to changing societal values and the reemergence of unified nation states, English justice officials were in need of an alternative to corporal and capital punishment and early forms of incarceration. Justice officials began to explore community supervision for several reasons, including early forms of prison crowding and as an altruistic recognition that the behavior of some offenders did not warrant imprisonment or capital punishment.

The development of probation can be traced to formal statutes and church intervention in England. During the medieval period, clergy members accused of crimes could have their cases moved from a criminal court to a church court. As a form of asylum, the "benefit of clergy" would allow the accused to avoid hanging as a penalty for their crimes. By reading the 51st Psalm, the criminal could claim the status of being a member of the clergy. The 51st Psalm was recognition of the offenders' remorse and repentance but also a literacy test to identify true clergy members. By the early nineteenth century, English statutes would end this practice due to misuse by illiterate, common criminals that would commit the verse to memory and misrepresent themselves as clergy members (Travis 2012). In its place, the English established statutes to allow court officials to exercise a form of official judicial reprieve. This formalized practice allowed sentencing officials to delay sanction under the conditions of good behavior for a set amount of time. At the same time, the courts developed release on recognizance (ROR), a practice that released accused persons awaiting hearings before circuit magistrates. Court officials had the discretionary power to release accused offenders into the care of a sponsor or by the offender posting some form of collateral. This practice was also an early forerunner to the system of bail in the United States. Together, judicial reprieve and ROR influenced alternatives to incarceration that were more formal process functions that directly led to the creation of probation and parole.

Early use of ROR in America first occurred in Massachusetts in 1830. In *Commonwealth v. Chase*, a woman was freed after pleading guilty without sentence being passed. In her case, the court released her, allowing her "upon her recognizance for her appearance in this court whenever she should be called for, to go at large" (as cited in Cromwell et al. 1985). Today this type of release is common for criminals with minor charges.

John Augustus is considered the "father of probation" and the creator of the most widely used correctional sanction in criminal justice. Originally a shoe cobbler in Boston, Massachusetts, Augustus became a volunteer probation officer in an unofficial capacity, when in 1841 he persuaded local court officials to release nonviolent drunkards into his custody in lieu of incarceration. Augustus would conduct a rudimentary background investigation of these individuals (a forerunner to presentence investigations), and if he chose to sponsor an individual, he would post a small fee for their release and supervise the individual in the community where he expected certain conditional behaviors. Augustus expanded his work to sponsoring men accused of other crimes, women, and some children. Over nearly two decades, he managed almost 2,000 criminals in the community. The Commonwealth recognized his work by codifying "probation" as an official sanction and the law made John Augustus the first paid probation officer. Later, other states adopted the Massachusetts model.

Parole developed in a similar manner around the same time as probation. Like Augustus, who became dissatisfied with blanket incarceration, other justice officials began to recognize the shortcomings of imprisonment. Early proponents of parole-like actions did not believe that simple incarceration in prisons had the power to "reform" criminals. Instead of a determined period of incarceration served by an offender, they developed systems to allow prisoners the opportunity to earn early release from their sentence. For example, Sir Walter Crofton developed "tickets of leave" for inmates in Ireland. After an early period of solitary confinement, prisoners could improve their stage of treatment through hard work and good behavior. After passing through an intermediate stage of treatment in prison, inmates could secure their release with a "ticket of leave." Under this system, Crofton gave himself the discretion to release an inmate early from their sentence and report to local authorities that an offender would be returning home. Under Crofton's system, there was no formal supervision of the offender after his release.

Captain Alexander Maconochie is considered the "father of parole." As an administrator of a British penal colony, Maconochie developed a "mark system" that allowed inmates to improve their classification. When entering the prison, all prisoners were at the lowest level of classification – the "penal stage" which was defined by hard labor and intense supervision. After a period of time with good behavior, prisoners could earn enough "marks" to move to the "social" stage. At this level, prisoners would have better living conditions in a smaller group setting with less supervision. Prisoners reached the final stage ("individual") through even more "marks" and earned the privilege of living in a cottage while working towards their conditional release (Travis 2012). Maconochie's system also called for the loss of "marks" and status. If offenders were to have poor behavior or laziness, they could be penalized and demoted. This carrot-and-stick system gave the prisoners incentives for good behavior in the prison, which helped control bad behavior in the institution.

In 1870, at the meeting of the American Prison Association in Cincinnati, prison reformers pushed for the creation of parole and a system of reformatory discipline for United States prisons. A prison in Elmira, New York, was the first where prisoners could be sentenced to terms of indeterminate incarceration. Parole in America would develop two main components: (1) discretionary early release from prison for worthy prisoners and (2) a period of supervision in the community that follows release (Travis 2012). Parole could also be used to support discipline within prison.

Current concerns about community supervision focus on how offenders are supervised, the types of services they receive, and whether these types of community management can curb new crime. These concerns have spurred the creation of specialized courts, such as reentry courts. The purpose of these courts is to increase the courts' supervision of an offender and allow for better coordination of community-based services (e.g., treatment programs, job placement). Reentry courts may represent a new paradigm in community corrections, which have been largely unchanged in purpose or design for over a century. Much in the vein of drug courts and mental health courts, reentry courts serve a population of released offenders, with a goal of improving offenders' transition from an institutional setting to a community setting.

Legal Issues in Community Corrections

The Progressive Era of corrections in the early twentieth century survived legal challenges to the discretionary powers inherent in probation and parole. In fact, these legal challenges helped formalize the systems of how offenders secure an early release and how they are supervised in the community and violation procedures. For example, *Morrissey v. Brewer* (1972) required that parolees accused of violations be granted certain due process protections. Offenders had the right to (1) written notice of violation, (2) disclosure of evidence, (3) the right to a hearing and the right to present evidence at that hearing, (4) a limited right to confront witnesses, (5) a hearing before a neutral body, and (6) a written statement of the decision that addressed the evidence upon which it was based. *Gagnon v. Scarpelli* (1973) extended these rights to probation violators, stating that there was no substantial difference between probationers and parolees. In sum, the court combined these two different forms of supervision because they supported the interests of probation and parole clients in maintaining their conditional liberties and, by doing so, granted them certain due process rights during revocation hearings.

Note that probation and parole are restrictive forms of supervision for criminals. Even though criminals in the community enjoy greater freedoms than those incarcerated, they are still held to a higher standard of behavior than those not on supervision. Formal conditions restrict simple behaviors that most Americans take for granted. For example, unrestrained citizens can move across state lines without official permission from a court or paroling authority, but most of those under supervision must gain this permission before performing an otherwise legal behavior.

The Organization and Function of Community Supervision

While both are forms of community supervision, the level of organization of probation and parole can be very different. Generally, probation is organized at the county and municipal level, while parole is typically a state function. On the other hand, in some states, the same officer may be charged to act as both a probation and parole officer.

If an offender is placed on probation, that offender is under the jurisdiction of the court and has conditions imposed by the sentencing judge. The probation officer is an officer of the court, responsible for completing a presentence investigation for the court, holding the probationer to the set of standards specified by the judge, and acting as a fact finder for the court if there is a violation of these conditions. Parolees are under the jurisdiction of the paroling authority. The paroling authority will set the conditions of release into the community, while parole officers are responsible for holding the parolee to the set of standards specified by the authority. The parole officer operates much like a probation officer and has similar duties of monitoring clients and reporting any wrongdoing.

In its most basic form, probation and parole represent an agreement between the offender and the state. In order to gain his or her release into the community, the offender must agree to abide by a set of rules dictated by the sentencing judge or paroling authority. Violation of these conditions can be the basis for increased sanctions or the loss of liberty and subsequent incarceration. There are two broad types of violations. The first type covers new crimes committed by an offender while on supervision status. The second type covers technical violations that do not require any “criminal” behaviors. Due to the restrictions placed on probationers and parolees, these offenders must abide by a higher set of standards than those not under correctional control. Offenders can violate their conditions of supervision with such simple behaviors as unauthorized movement across state lines, using alcohol, or failing to report their new address to their supervising officer.

The status of probationer or parolee is sometimes known as being “on paper.” Probationers and parolees typically receive written rules or conditions of supervision. This listing of the conditions is the “paper.” Conditions serve many purposes. One purpose of written probation and parole conditions is to ensure that supervision can occur. That is, these rules are in place to define the behavioral expectations on the part of the offender and to define any further restrictions or treatments expected of them while in the community. These conditions also define how a probation or parole officer will interact with an individual. For example, if the offender is a known drug user, the conditions may require the officer to conduct drug tests with the client a defined number of times per month. Another purpose of written conditions is to place the offender on notice. Some conditions notify offenders that they may owe a certain amount of money in fines or may have to perform a certain amount of community service hours before he or she can be considered for outright release. Finally, this notice may be a simple time commitment, clearly defining that an offender must spend a specific period on supervision before discharge. Other conditions require the offender to seek or complete treatment. Finally, some conditions are designed to be punitive, such as requirements that offenders serve some jail time or write “punish lessons.”

Generally, probation and parole conditions are not universal across jurisdictions. There is a set of standard conditions that are expected of everyone in that specific jurisdiction (e.g., follow up with all subsequent court hearings and probation meetings, do not commit a new crime); however, while they are often similar, these standard conditions vary slightly in tone and expectation across jurisdictions. Probationers and parolees also receive special conditions. Instead of general expectations imposed on everyone on community supervision, these conditions are typically tailored to the individual (e.g., completion of treatment programs – substance or behavioral, restrictions on drug or alcohol use). General and specific conditions reflect the purpose of modern community corrections that offenders are formally controlled in the community while receiving necessary services to keep them from recidivating.

Issues in Community Supervision Effectiveness

Community supervision has come under attack from critics as being ineffective and “soft” on crime. Those that measure “effectiveness” as a total reduction in recidivism may never be satisfied with community supervision. The evidence reported in this section suggests that community supervision can be cost-effective and may have a positive impact on reducing criminal recidivism with certain types of offenders. The key to effective community supervision is embedded in proper identification of offenders that can behave successfully in the community with less state restriction. As the science of identification, rehabilitation, and controlling risk continues to improve, so will the effectiveness of community corrections. This section examines the effectiveness of the system according to research on the risk of offenders in the community, the cost-effectiveness of community supervision, and the successful types of community interventions.

Does Community Supervision "Work"?

The answer to this seemingly simple question may depend upon the yardstick used to define what "works" in terms of community corrections. Put differently, to define and measure what "works" may depend upon the point of view of the person asking the question. For example, politicians who campaign on crime policies may use a reduction in recidivism as evidence of a "working" system, while budget officials or economists may use a cost-benefit ratio to explore evidence of an effective system.

Community corrections should serve a basic purpose of community protection. Probationers should be appropriately identified as the type of people who do not need to be incarcerated and can operate pro-socially in the community. However, due to the increase in prison populations, community corrections are taking on more clients, some of whom may not be appropriate for this type of intervention. This is particularly a concern for probation since violent offenders are making up a larger portion of probationers (Auerhahn 2007), which can be attributed largely to a lack of bed space in prisons and poor identification of risk. Adding to this concern is the evidence that simple probation supervision may not be appropriate for certain types of offenders. Research indicates that probation has failed to prevent overall crime and is specifically ineffective with violent criminals and criminals guilty of fraud or forgery (MacKenzie et al. 1999). But probation can effectively prevent new crime for certain types of offenders such as drug dealers and people that engage in property crime. This illustrates an important point in moving forward: probation can "work" if clients are appropriately matched to the correct form of intervention and supervision. If evidence suggests that violent criminals are not appropriate for community control, policies should be reevaluated to make sure that these criminals do not receive a probation sentence. Alternatively, supervision practices might be changed to better monitor and control these types of offenders.

Other scholars have questioned the appropriateness of probation with certain types of offenders. Petersilia found that due to risks to public safety, felons in California were not appropriate for probation supervision. However, other scholars came to different conclusions when studying different states, citing that her results may be an artifact of California supervision. The concept of effectiveness was measured similarly across these studies, since each study was interested in *new* crimes committed by offenders, yet the researchers reached different results. The researchers could have defined and measured the concept of effectiveness differently and come to different conclusions altogether. For example, they could have examined whether the probation period was successfully completed by felons without any type of technical violations or time to recidivism.

Parole has also been harshly criticized by some scholars who have called for the abolition of parole in the United States. This argument hinges on ethical issues (e.g., supervision after incarceration adds extra punishment) and the lack of evidence supporting parole effectiveness. Indeed, measurement of parole effectiveness also suffers from the same measurement problems that plague probation research. Conflicting research defines effectiveness as curbing new crimes by parolees or keeping clients out of prison due to technical violations of their supervision conditions. The latter point is particularly important given that one-third of all inmates in state prisons were parole violators as of 2005 (Sabol et al. 2007). Some scholars have suggested that counting these types of offenders as failures of the system makes it impossible to adequately measure parole effectiveness (Flanagan 1985). In other words, it is important to identify how parole effectiveness should be determined: reduction in new crimes by the offenders, reduction in technical violations, or both. If parole keeps offenders from committing new crimes, then it may be successful. Research that measures technical violations in addition to new crimes may not be indicative of higher rates of crime.

The use of different outcome measures will lead to different conclusions about community supervision effectiveness. Generally probation and parole work to the degree that they (1) control risk of new crime, (2) reduce incarceration and correctional costs, and/or (3) match punishments with offenders. These three areas represent particular goals of the system and appear to suggest community supervision can be an effective form of intervention.

Controlling Risk

Attempts to identify and manage risk through probation and parole are difficult and imperfect but necessary for the system to operate within its means. A critical question for risk control of offenders in the community is a pure cost-benefit question: how safe is safe enough? Supervision of known criminals in the community will always present the risk of some offenders committing future crimes, but is there a threshold of new criminality below which officials would deem community supervision safe enough? If the widespread use of community supervision were to lead to an increase in overall crime, these types of interventions would be unacceptable. Minimizing new crime by these offenders would be optimal; however, if a level can be reached that maintains current crime rates, then the risk presented by community supervision would be no greater than that posed by incarceration.

The development of risk assessments and their appropriate use by probation and parole officers allow the risk of recidivism among correctional populations and specific types of offenders to be validly predicted (Holsinger et al. 2006). If an offender presents a higher risk of recidivism, justice officials may consider a higher level of supervision in the community or a period of incarceration. Risk can be eliminated entirely if the system had the means to incarcerate every criminal sentenced. However, even if the system had the means for blanket incarceration, an overwhelming majority of inmates would receive a simple discharge into the community.

Assessments of recidivism suggest that community supervision of offenders does not place the community at large under a greater threat of crime. Evaluations of parolee recidivism suggest that more offenders could be placed on parole with little to no change in the rate of parole failures (Wilson 2005). This assessment also noted that most of the failures could be attributed to technical violations, not new crimes by the offenders.

Cost-Effectiveness

Estimates suggest the cost of each new prison bed to be approximately $80,000, with an annual maintenance fee of $20,000 or more for each inmate's house. Even though the costs of incarceration are high, prisons continue to be built and the increased use of incarceration puts bed space at a premium. Yet, in reality the system does not have the means to incarcerate all offenders. Community supervision was born in part from a desire to reduce overcrowding and costs associated with corrections. Clear and Byrne (1992: 321) flatly say, "The frank bottom line for the intermediate sanction movement must be whether it is able to reduce overcrowding in corrections."

Though always dated, most cost evaluations suggest that community supervision of an offender will cost far less than incarceration. An early study measuring direct and indirect costs of different sanctions suggested a state could save nearly six million dollars in 1 year if 3,000 inmates were diverted from prison and placed on probation instead (as cited in Latessa and Allen 1997). More recent evaluations have attempted to examine this cost-benefit analysis with greater rigor, specifically, by accounting for recidivism rates of probationers versus parolees. These studies also support the conclusion that community supervision is more cost-effective than incarceration.

Matching Offenders with Effective Interventions

Properly identifying clients' needs and matching them to appropriate interventions may be key to determining greater effectiveness of community corrections. Since there is a wide variety of offender needs and crimes committed, the system must sustain a wide range of interventions. Matching offenders to the appropriate punishment begins with identifying a proper penalty for the seriousness of the crime committed. Thus, the first basic decision point is whether the offense merits prison or consideration for community supervision. Of course, under probation supervision, there exists a myriad of levels and types of supervision. Next, justice officials must identify the essential needs of the client and match him or her with the correct intervention. It would be pointless, for example, to require a forger with no history of substance abuse to attend Alcoholics Anonymous meetings.

Correctional rehabilitation has been continuously attacked for being "soft" and easy on crime. However, there is substantial evidence that when we correctly identify the needs of offenders and match them to a program with a strong treatment modality, the client will be less likely to recidivate. Gendreau found that 64 % of offender rehabilitation studies showed that appropriate correctional intervention can reduce recidivism, and if these programs are using the principles of effective intervention, their effectiveness can be strengthened (as cited in Latessa and Allen 1997). Assessment tools have been created that help correctional staff identify risk and needs. These assessment tools have changed the manner in which probation and parole officers conduct their jobs.

Other Issues in Community Supervision

This section examines broad current issues relevant to community supervision. As the science behind effective criminal control advances and government budgets become more strained, the issues of financing, management, and technology have become more relevant.

Financing

Several states have implemented incentive programs for counties that support the use of community supervision and community-based intervention programs. This type of legislation encourages counties to decrease the number of commitments to state prisons, while also helping to subsidize community corrections programs. In 1965, California was the first state to pass such legislation when it codified the Probation Subsidy Act. This law used a formula that estimated the number of offenders expected to be sentenced to prison by each county and then paid $4,000 to the county for each additional offender below that estimate who was not sentenced to state prison. Other states followed suit by passing subsides tied to the development and expansion of community intervention programs.

Some states and counties have also imposed supervision fees to help subsidize the cost of community supervision. Supervision fees can be found in the form of a one-time flat fee for supervision or a monthly fee paid by the offender for the privilege of maintaining community supervision. Criminal sentences increasingly include restitution as a sanction that orders the offender to pay money. Restitution can compensate the victim for any money lost during the criminal act or for subsequent medical bills. Some states also have general victim-compensation funds that each offender must pay into as part of their conditions.

Supervision fees and restitution can offset the costs of community supervision, but they also present new problems for the system. First, the collection of money puts probation officers in a situation where they become bill collectors, a role that these officers generally dislike. Secondly, while supervision fees are minimal (e.g., $10 per month), how should the system proceed if the client does not or cannot pay that fee? If the offender is employed, wages can be garnished through order of the court, but the time and money spent on securing this garnishment may offset the minimal fee collected. On the other hand, if the client is not employed or simply does not pay, this client could be violated for failure to pay these conditional fees. This will likely be counterproductive to managing finances since incarceration is more expensive than community supervision.

Management

Several models of organization of caseload work have been proposed and adopted, more recently focused on specialty supervision. The traditional casework model expects a single officer to provide needed services and supervision for a wide range of offender types. Sometimes referred to as "line officers," these supervision officers supervise the bulk of offenders in the community and have a wide, general range of responsibilities. New innovations in management have created specialty positions, such as team supervision and differential case management.

Team supervision is an approach that attempts to capture officers' strengths. This type of supervision would use an officer who might be particularly good at helping people find jobs as an employment officer. However, this type of group supervision can become confusing if the offender experiences a varied continuum of supervision. Another development in supervision involves tailored case management for specialty offender populations. For example, a few probation officers may be assigned specifically to supervise sex offenders.

The model case management system, developed by the National Institute of Corrections, has three components: (1) classification, (2) case planning, and (3) the assignment of workload units (Travis 2012). Classification is accomplished through risk and needs assessments that are completed by the probation officer and offender. Through these assessments, distinct categories of risk and program needs are created.

Technology

Innovations in technology have also influenced community supervision. As community supervision agencies adopt these innovations, they changed the role of the probation and parole officer. While some innovations have strengthened the ability of officers to execute their duties, some have added responsibilities that traditionally did not exist.

Mandated drug and alcohol testing are heavily used supervision conditions. Many scholars, policy makers, judges, and correctional staff believe that these types of testing will deter substance use and in turn reduce new crime. Testing can be in the form of blood tests sent to a lab to determine if an offender is using alcohol or drugs. More instant tests involve single or panel urine screens conducted on site by the probation or parole officer. If the offender is considered to be a habitual marijuana user, the supervising probation officer can use a single dipstick to test for THC present in the urine. If the supervising officer wishes to screen for a variety of drugs, a panel of tests contained on one stick can be conducted on urine. Drug tests have shaped the work of the modern day probation and parole officer. Officers now spend a good amount of time collecting and testing urine samples.

Electronic monitoring devices can be affixed to an offender's body to allow the supervising officer to monitor his or her whereabouts. Traditional electronic monitoring involves an ankle bracelet monitored by radio frequency through a transmitting device placed in the offender's home using a land phone line. The monitoring system will either randomly dial the offender's phone to ensure they were home or utilize a passive transmitter to notify the supervising officer when the offender is outside of the transmitting zone (e.g., a distance of 50 yards). If the client is not on house arrest, regular "windows of leave" may be entered into the system by supervising officers so that offenders could leave their home for work, school, treatment sessions, or basic necessities like grocery shopping or doctor visits. Newer advances in electronic monitoring involve the use of global position systems (GPS) to give supervising officers instant, geographically specific updates on the whereabouts of their clients.

Finally, in some jurisdictions, probation officers are required to collect DNA samples from clients. For example, Pennsylvania requires that certain offenders submit to a DNA test upon conviction. Originally collected through blood by a nurse, samples can now be collected through a cheek swipe conducted by probation officers. The DNA sample is processed by the probation officer and sent to the state repository, adding an additional responsibility to officers' role.

Discussion: Open Questions Concerning the Future of Probation and Parole

Having considered the historical ideologies, modern operation of probation and parole, and current issues facing these systems, this section addresses the future of community corrections. As discussed throughout this entry, the criminal justice system continues to increase the amount of offenders under correctional control – both in prison and in the community. Overcrowding in prison continues to be a concern for justice officials, but a less visible problem is bulging caseloads for community officers. This is highlighted by *Brown v. Plata* (2011), a recent United States Supreme Court ruling that orders California to release nearly a quarter of its prison population for reasons including inadequate physical and mental health care and prison overcrowding. The release of offenders into the community will have an immediate impact on the caseload sizes of community supervision officers in California. Yet this decision will also have longer-term effect on caseload size as judges may tailor future sentencing decisions to comply with overcrowding regulations set by the court. That is, judges may place even more offenders in the community that previously might have been incarcerated.

The larger issue for community corrections in the future is saturation: at what point will community corrections, like prisons today, become overburdened? Can similar logic used for defining overcrowding in prison be applied to community services? Prisons are designed to house a specific number of inmates, and should a prison reach and exceed its design capacity, it can be considered overcrowded. A review of the research on the relationship between caseload size and a positive probation outcome (e.g., lower recidivism rates) has not provided justice actors with any baseline number for an overburdened caseload. What research has suggested is that generally large caseloads are associated with an increase in property offenses and, in contrast, smaller caseloads are criticized for being too intense and have an unintended outcome of net-widening within the system. Other recent research has indicated that a reduced caseload size in addition to sound evidence-based treatment has the potential to reduce offender recidivism (Jalbert et al. 2010). Nevertheless, the balance between the two extremes is vague. More community-based offenders under correctional control will also place a larger burden on community-based mental health and drug treatment services. It may be unreasonable to assume that the inadequacy of social services in prison will be improved in the community when larger numbers of individuals are placed into the community under correctional control.

In response to overburdened caseloads and understaffed departments, some agencies have turned to new technology to help supervision. For example, in some states certain low-risk offenders are "supervised" by a kiosk. That is, the individual under supervision checks in periodically with a computer kiosk to pay fines or restitution and update any information with the court. The use of the computer kiosk (and its less engaging predecessor – the filing cabinet) raises ethical questions concerning the purpose of community-based punishment. As noted previously, probation and parole are less restrictive forms of supervision than incarceration. But these forms of supervision involve human interaction, checks on behavior and adherence to supervision conditions. If judges have identified a portion of offenders that do not need to be supervised because of their low-risk categorization or minor offense, why are these individuals on supervision? Furthermore, these offenders are not receiving any treatment or rehabilitation services from the court as part of their community supervision status. Therefore, the use of non-reporting supervision or kiosks does not appear to serve any clear purpose associated with community supervision.

The future of probation and parole will be dictated by the amount of resources given to these community agencies. If policy makers, court officials, and researchers continue to advocate for increased use in community corrections, more money must be funneled towards these official agencies and community agencies that lend social support. However, if probation and parole services reach a saturation point that result in the increased use of non-reporting supervision or inadequate services, these sentences become supervision and punishment in name only. Ultimately, for probation and parole to continue to increase their efficiency in case management and to optimize the effect on recidivism, these correctional systems should continue to utilize and expand evidence-based practices such as supervision and treatment strategies that adhere to the principles of effective intervention.

Related Entries

▶ Effective Supervision Principles for Probation and Parole
▶ Intensive Probation and Parole
▶ Parole and Prisoner Reentry in the United States
▶ Probation and Human Rights
▶ Probation and Parole Practices
▶ Probation Officer Decision-Making

Recommended Reading and References

Auerhahn K (2007) Do you know who your probationers are?: using simulation modeling to estimate the composition of California's felony probation population, 1980–2000. Justice Q 24(1):121–142

Brown v. Plata 563 U.S. ____ (2011)

Camp C, Camp G (1996) The corrections yearbook, 1996. Criminal Justice Institute, South Salem

Clear TR, Byrne J (1992) The future of intermediate sanctions: questions to consider. In: Byrne J, Lurigio A, Petersilia J (eds) Smart sentencing: the emergence of intermediate sanctions. Safe, Beverly Hills, pp 319–331

Cromwell PF Jr, Killinger GC, Kerper HB, Walker C (1985) Probation and parole in the criminal justice system, 2nd edn. West Publishing Company, St. Paul

Flanagan T (1985) Questioning the other parole: the effectiveness of community supervision of offender. In: Travis LF III (ed) Probation, parole, and community corrections: a reader. Waveland, Prospect Heights, pp 167–183

Gagnon v. Scarpelli 411 U.S. 778 (1973)

Glaze L, Bonczar T (2006) Probation and parole in the United States, 2005. Bureau of Justice Statistics, Washington, DC

Holsinger A, Lurigio A, Latessa E (2001) Practitioners' guide to understanding the basis of assessing offender risk. Federal Probation 65(1):46–50

Jalbert SK, Rhodes W, Flygare C, Kane M (2010) Testing probation outcomes in an evidence-based practice setting: reduced caseload size and intensive supervision effectiveness. J Offender Rehabil 49:233–253

Latessa EJ, Allen HE (1997) Corrections in the community. Anderson Publishing, Co., Cincinnati

Latessa E, Smith P (2011) Corrections in the Community, 5th edn. Elsevier, Burlington

MacKenzie DL, Browning K, Skroban SB, Smith DA (1999) The impact of probation of the criminal activities of offenders. J Res Crime Delinq 36(4):423–453

Morrissey v. Brewer 408 U.S. 471 (1972)

Sabol W, Minton T, Harrison P (2007) Prison and jail inmates at midyear 2006. Bureau of Justice Statistics, Washington, DC

Travis LF III (2012) Introduction to criminal justice, 7th edn. Anderson Publishing, Co., Newark

Wilson J (2005) Bad behavior or bad policy? An examination of Tennessee release cohorts, 1993–2001. Criminol Public Policy 4(3):485–518

History of Randomized Controlled Experiments in Criminal Justice

Cynthia Lum[1] and Lorraine Mazerolle[2]
[1]Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA
[2]Institute for Social Science Research, The University of Queensland, St. Lucia, QLD, Australia

Overview

This entry examines the history of randomized controlled experiments in criminology and criminal justice. First, fundamental characteristics of randomized controlled experiments are briefly described, emphasizing the connection between this research method and evidence-based crime policy. Then, historical trends of experiments in criminal justice are reviewed, highlighting David Farrington's work in this area. The authors continue by connecting the history of experiments as well as their characteristics and the debates surrounding their use in the context of the evidence-based crime policy movement. Specifically, the authors suggest that the history of experiments in criminal justice and their relative rarity compared to other evaluation research cannot be divorced from the broader discussions and realities about the "what works" movement in criminal justice. Finally, this entry provides thoughts about the future of experiments in criminal justice and the research infrastructure being built to support them.

Fundamentals of Randomized Controlled Experiments

A randomized controlled experiment is a research method used to examine the effect that a treatment, program, intervention, or condition has on a particular outcome. In criminology, experimental methods are used to examine whether a crime policy or law, an offender treatment, a crime prevention program, or a new organizational process can reduce or prevent crime, lower recidivism, or temper fear of crime. Experiments can also be used to examine the impact that interventions have on legal processes and justice externalities, including differential treatment, legitimacy, employee job satisfaction, or corruption of justice agents. Although the terms *randomized controlled experiment* and *randomized controlled trial* are often used interchangeably, a trial implies the implementation of the experimental design.

Experiments are differentiated from other evaluation methods, such as "quasi-experiments" and before-after designs, because the treatment or condition is randomly allocated to some subjects (persons, places, situations, etc.) and not others. By randomly allocating an intervention to control and treatment groups, a researcher can ensure that no systematic bias contributes to group differences before or after an intervention is applied (Campbell and Stanley 1963). Specifically, random allocation allows for the assumption of equivalence between treatment and comparison groups, which allows the researcher to rule out confounding factors that might explain differences between groups after treatment (Weisburd 2003). By randomly allocating treatment, an appropriate counterfactual is developed in the control group, which shows what would happen if a program or treatment was not administered (Cook 2003).

Randomized controlled experiments have been called – not without debate – the "gold standard" of evaluation research, as it is seen as one of the best ways that evaluators can maximize a study's internal validity, thereby generating findings that are believable (Boruch et al. 2000a; Farrington 2003a; Weisburd 2000). Internal validity refers to an evaluator's ability to determine whether the intervention – as opposed to other effects – caused a change in the outcome measured (Shadish et al. 2002). This determination is also facilitated by good design and fidelity of implementation of the experiment. However, the label of "gold standard" relies on its strength in establishing internal, rather than external validity, a weakness of randomized controlled designs. Other criticisms are presented in more detail below.

Historical Trends in Experimental Criminology

Farrington (1983) and Farrington and Welsh (2006) have examined the history of the use of randomized controlled experiments (RCEs) in criminal justice and criminology in detail, and we summarize their efforts here. As they describe, the use of RCEs in criminology has focused on crime prevention and delinquency treatment programs, whose beginnings are marked by the Powers and Witmer (1951) Cambridge-Somerville Youth Study. This program was initiated by Richard Cabot in 1939 and involved over 500 delinquent boys, half of whom were randomly assigned to a treatment group and the other half to a control group. The Cambridge-Somerville Youth study is widely considered one of the earliest and most influential experiments in criminal justice research (Weisburd and Petrosino 2004). This experiment and McCord's (1978) follow-up research found that the treatment may have backfired, causing the boys who received the treatment to commit more crimes and have a lower quality of life.

After the Cambridge-Somerville Youth study, early experimentation in criminal justice did not necessarily show optimistic results, and many foundational experiments and reviews created a pessimistic environment for criminal justice interventions. Using a threshold of 100 or more subjects, Farrington (1983) identified 37 experiments in criminal justice in the first 25 years of this history (1957–1982), many of which did not show positive effects. The four policing experiments he reviewed focused on juvenile diversion schemes, two of which reported positive results of youth diversion programs and two reporting backfire effects of the programs. Six crime prevention experimental evaluations were also not optimistic (which included the Cambridge-Somerville Youth Study and its follow-ups). It appears the only experiment with significant positive findings in this crime prevention group was the one conducted by Maynard (1980), which examined the effects of a multi-site community-based intervention operated by the Manpower Demonstration Research Corporation. In that study, highly structured work experiences and job placement led to significantly lower arrest rates in the treatment than control group, but only in the short term.

Farrington (1983) also found 12 correctional experiments satisfying his criteria, focusing primarily on counseling offenders. Only three resulted in significant reductions in offending. Three court-related experiments were also discovered, with only one showing reductions of truancy by bringing children to court and using adjournments (delaying supervision by social workers) versus immediately assigning them to supervision. The authors concluded that the effect may have been caused by the type of worker handling adjournment versus supervision, the former more likely to be concerned with reduction of truancy than the latter. And finally, Farrington (1983) reports the results of 12 community-oriented interventions evaluated with experimental methods. Only one, he discovered, could be said to show significant reductions in offending.

An updated review conducted by Farrington and Welsh in 2006 proved more positive. They found 85 additional experiments between 1982 and 2004, many showing positive effects of the interventions studied. With regard to policing, the updated review reported 12 experiments, largely focusing on the area of domestic violence and repeat offenders, but some experimenting on police interventions at places. Of these 12 studies, six reported significant effects in the intervention reducing recidivism and crime. The place-based experiments were an innovation in the history of experimentation, which previously was focused on individuals. The Minneapolis Hot Spots Experiment (Sherman and Weisburd 1995) was significant in this regard, which was then followed by numerous place-based experiments in criminology. Overall, the policing research in Farrington and Welsh's new review had a more optimistic outlook, showing that police can indeed prevent crime among people and places.

Prevention experiments also grew from six in Farrington's first review to 14 in the updated review. Of the 14, six showed that interventions had significant positive effects, and some of the interventions have been highly influential in the discipline. In particular, the work of Olds and his colleagues (1998) on home visitation nurses showed that non-visited mothers tended to abuse and neglect their children more. His work, which subsequently contributed to his receipt of the Stockholm Prize, led to a movement for improvement in postnatal care using home visitation. Other developmental experiments involving sending kids to preschool, early child and parent skills training, multi-systemic therapy (as opposed to supervision and probation), and employment services showed positive effects. Unlike the vague counseling strategies covered in Farrington's earlier review, these promising interventions were more specific, targeted, and focused on developmental stages.

Farrington and Welsh also included 14 correctional experiments with interesting findings that challenged widely held beliefs. In particular, popular interventions like scared straight and juvenile boot camps showed backfire or nonsignificant effects on juvenile delinquency, while therapeutic communities, drug treatment, and cognitive behavior therapy proved promising (see also MacKenzie's review in 2002). Farrington and Welsh also located 22 court-based experiments, such as court-mandated treatment, restitution, intensive supervised probation, restorative conferences, and pretrial methods. Like corrections studies, many of these studies and their replications helped show that a number of traditional criminal justice programs were not as effective as initially believed (again, we point the reader to Farrington and Welsh 2006, for detailed discussions of these experiments).

Finally, Farrington and Welsh also found more community-based experiments (23 studies) than Farrington's initial review, which offered mixed results, especially in the area of intensive supervised probation. Petersilia and Turner (1993) found that, in some cases, community-based supervision led to increased offending, while in others, it led to decreased recidivism. This was highly connected to the reporting of probation violations and revocations, not necessarily to the frequency of crime commission, which is a much harder aspect to measure. MacKenzie's (2002) assessment of the corrections experiments was more nuanced, given these mixed findings. It appears that more tailored and structured community supervision programs that involve the right therapy, treatment, and employment opportunities can have an effect on recidivism, even if intensive supervised probation alone may not be able to do this. On the other hand, experiments in the corrections and community area have also shown that shock probation, scared straight, unstructured and vague rehabilitation programs, correctional boot camps, urine testing, community residential programs, and juvenile wilderness programs are not effective (MacKenzie 2002).

Randomized Experiments and the Evidence-Based Crime Policy Movement

The reviews by Farrington, Welsh, and MacKenzie show a general growth in experimental research since the 1950s and more optimistic findings about specific crime and justice relevant interventions. However, the history of experimental criminology is still a "meager feast" (Farrington 2003b) centered on American criminology (Farrington and Welsh 2006). Despite its gold standard label, the use of experiments has been relatively rare compared to other methods used in evaluating interventions (Weisburd 2003). Indeed, Farrington and Welsh (2006) still only discovered 122 experiments between 1957 and 2004, compared to the many more hundreds of evaluations conducted using other methods.

Others have also confirmed this finding of a dearth of experiments (see Boruch et al. 2000a; Palmer and Petrosino 2003; Petrosino et al. 2003; Sherman et al. 1997, 2002). Lum and Yang (2005), in a review of Sherman and colleagues' (2002) updated review of 657 evaluations of criminal justice programs, found that experiments accounted for 16 % ($n = 102$) of the evaluations reviewed. In contrast, the vast majority of studies in Sherman's updated review used before-after designs or compared non-comparable groups with and without treatment. Baird's (2011) review of the population of all 1,025 articles published in our field's leading journal, *Criminology*, from 1980 through August 2011, reveals only 14 articles (1.4 %) that used an experimental design. Moreover, Baird's (2011) review of the 14 experimental studies published in *Criminology* finds that even for those published in our leading journal,

> ...the descriptive validity of statistical power in many of the studies was low. Six of the studies did not report they were experimental in their title. Five studies did not report enough information to conduct power analysis. Errors such as failing to report directional hypotheses, sample sizes of groups, and alpha levels are critical omissions. For those studies that did provide enough information to calculate power, the power levels were generally high for large and moderate effects and were generally low for small effects. (2011, p. 15)

There have been numerous reasons given for the scarcity of experimental research in criminology and social science, and practical and ethical concerns are cited most often (see Boruch et al. 2000b; Clarke and Cornish 1972; Cook 2003; Farrington 1983; Lum and Yang 2005; Petersilia 1989; Shepherd 2003; Weisburd 2000). Some also believe that using experiments is challenging when studying the complexities of social relationships and issues (Burtless 1995; Clarke and Cornish 1972; Heckman and Smith 1995; Pawson and Tilley 1997), or that it is difficult for experiments to maintain high fidelity when implemented (Berk 2005). Lum and Yang (2005) also found, when surveying researchers about their choices to use or not use experimentation, that the lack of funding was often more a concern than ethical or even practical reasons. Indeed, Palmer and Petrosino (2003), when examining the history of the California Youth Authority (CYA), found that when funding switched from the experimentation-friendly National Institute of Mental Health to the Law Enforcement Assistance Administration, which rarely funded such studies, the focus of the CYA also moved away from experiments.

The practical, ethical, and funding challenges of randomized controlled experiments do not exist in a vacuum; they are also influenced by the context of the role and history of experimental evaluation in crime prevention more generally. In particular, the low proportion of experimentation may be a result (or symptom) of the messy quest for understanding and implementing "what works" in criminal justice, or the "evidence-based crime policy" movement. In general, this movement suggests that criminal justice practice should be guided by the best available research and analysis about interventions (Sherman 1998; Sherman et al. 2002), rather than by anecdote, feelings, habit, or opinions about best practices (Lum 2009). However, the goal to translate and use research in practice requires a keen sensitivity to practicality, ethicality, and costs. These core values of evidence-based crime policy can be, as previously mentioned, challenges to experimental criminology. Thus, while the demand for knowledge about effective interventions in the evidence-based crime policy movement certainly encourages more use of high quality evaluation designs, other core values can also work against that effort.

Two discussions may better illustrate the ironically difficult relationship between experiments and the evidence-based crime policy movement. The first involves the story of the Martinson Report (See Lipton et al. 1975; Martinson 1974) and the subsequent University of Maryland Report to Congress (Sherman et al. 1997, 2002). In 1975, Lipton, Martinson, and Wilks published a review of 231 evaluations of rehabilitative programs, to which Martinson publicly asserted that such programs had "no appreciable effect on recidivism" (Martinson 1974, p. 25). As described by Pratt et al. (2011), the report, which became known as the "Martinson Report," may have been one reason (in the context of rising crime rates and changing social cultures) for the decline of the rehabilitative ideal. At the same time, however, this shift from rehabilitation to punishment and the pessimistic assertions by Martinson arguably sparked the modern-day, evidence-based crime policy movement. Not only did evaluators wish to find what *did* work for rehabilitation and corrections, but they also wanted to increase the believability of such studies by using stronger evaluation methods (i.e., experiments). As Doris MacKenzie (2008) stated in a keynote address, an evidence-based approach to crime prevention "rejects the nothing works philosophy and instead examines what works to change criminal and delinquent activities. It is a belief that science can be used to inform public policy decisions about which programs or interventions should be used to change offenders."

Sherman and colleagues' (1997) University of Maryland Report to Congress, and the update to this report (Sherman et al. 2002) could arguably be viewed as an attempt to challenge the consequences of the Martinson Report. Not only did they take into account the many new evaluation studies generated between 1974 and 2002, but they also weighed and ranked the methods used in those evaluations to make judgments about how much we should believe specific study findings. Developing a Scientific Methods Scale (SMS), they rated studies in criminal justice according to the following scheme:

1. Correlation between a crime prevention program and a measure of crime or crime risk factors
2. Temporal sequence between the program and the crime or risk outcome clearly observed, or a comparison group present without demonstrated comparability to the treatment group
3. A comparison between two or more units of analysis, one with and one without the program
4. Comparison between multiple units with and without the program, controlling for other factors, or a nonequivalent comparison group has only minor differences evident
5. Random assignment and analysis of comparable units to program and comparison groups (Sherman et al. 1997)

Sherman and colleagues' methodological ratings were purposeful. The ratings implied that not all research could be considered equal. Conclusions of methodologically stronger studies were arguably more reliable than weaker ones in their assertions about "what works, what doesn't and what's promising" (the title of the report) in crime prevention interventions. The SMS scoring system solidified the importance of randomized controlled experiments in this endeavor, which were rated with the highest score of "5." Rating the believability (and usability) of evaluations in a document intended to influence criminal justice practices arguably helped to link experimental criminology with evidence-based crime policy more generally.

Given that the Maryland Report and the SMS ratings became part of the more general evidence-based crime policy movement, what has been its connection to the history of experiments in criminal justice? Analysis using the SMS ratings of studies may prove helpful in better understanding this complex relationship between experimental criminology and evidence-based crime policy. Weisburd et al. (2001) and Lum and Yang (2005) have studied the evaluations chosen by Sherman et al. (1997) and Sherman et al. (2002). Using their data, we present Tables 1 and 2. Table 1 shows that despite the increase of evaluations between 1997 and 2002, the proportion of evaluations using experiments remains consistently low. Of the 308 and 657 studies in Sherman et al. (1997) and Sherman et al. (2002), respectively, only 15–16 % of the studies employed a randomized controlled trial to evaluate the effects of interventions. Interestingly, there was an increase in the number of quasi-experimental studies in criminal justice between the 1997 report and the 2002 update (from 9 % to 25 %), and a decline in the proportion of lower quality studies (SMS = 1, 2 or 3). While this shows a general improvement in the state of criminal justice evaluation research, the proportion of experiments have remained the same.

History of Randomized Controlled Experiments in Criminal Justice, Table 1 Distribution of SMS scores across the Maryland Report (and update) studies

	Number of studies (with %)	
SMS score	Sherman et al. 1997	Sherman et al. 2002
1.00	10 (3.2 %)	15 (2.3 %)
2.00	94 (30.5 %)	151 (23.0 %)
3.00	130 (42.2 %)	224 (34.1 %)
4.00	28 (9.1 %)	165(25.1 %)
5.00	46 (14.9 %)	102 (15.5 %)
Total	308 (100 %)	657 (100 %)

Adapted from data from Weisburd et al. (2001) and Lum and Yang (2005)

Further, as Table 2 shows, the use of RCEs varies across different sectors of criminal justice research. Using the studies in Sherman et al.'s (2002) review, we find experimentation is more common in interventions related to family or employment arenas than in policing and corrections, two of the largest justice institutions.

Thus, while evaluation research has increased generally in criminal justice, and while the evidence-based crime policy movement arguably gained momentum between 1997 and 2002, the proportion of experiments generally, and within two large areas of criminal justice, remain consistently low. Further analysis of the initial Maryland Report by Weisburd et al. (2001) also revealed another relationship between experiments and evidence-based crime policy. They discovered an inverse relationship between study findings and research design. At least in

History of Randomized Controlled Experiments in Criminal Justice, Table 2 Scientific methods scores across different areas of criminal justice research in Sherman et al. (2002)

	Scientific methods score (1–5)					
	1.00	2.00	3.00	4.00	5.00	Total
Community	0 (0.0 %)	9 (31.0 %)	14 (48.3 %)	3 (10.3 %)	3 (10.3 %)	29 (100 %)
Corrections	0 (0.0 %)	56 (37.8 %)	53 (35.8 %)	19 (12.8 %)	20 (13.5 %)	148 (100 %)
Family	0 (0.0 %)	0 (0.0 %)	0 (0.0 %)	9 (26.5 %)	25 (73.5 %)	34 (100 %)
Labor	5 (9.6 %)	7 (13.5 %)	18 (34.6 %)	9 (17.3 %)	13 (25.0 %)	52 (100 %)
Place	4 (4.0 %)	37 (37.0 %)	51 (51.0 %)	5 (5.0 %)	3 (3.0 %)	100 (100 %)
Police	6 (5.4 %)	39 (34.8 %)	39 (34.8 %)	13 (11.6 %)	15 (13.4 %)	112 (100 %)
School	0 (0.0 %)	3 (1.6 %)	49 (26.9 %)	107 (58.8 %)	23 (12.6 %)	182 (100 %)
Total	15 (2.3 %)	151 (23.0 %)	224 (34.1 %)	165 (25.1 %)	102 (15.5 %)	657 (100 %)

Adapted from data from Weisburd et al. (2001) and Lum and Yang (2005)

the crime and justice evaluation arena, it appears that studies that employ more rigorous methodological designs (i.e., experiments) are more likely to find that interventions are not effective, or even may create "backfire" or harmful effects (McCord 2003). Their analysis emphasized the practical (and perhaps ethical) importance of using the best methods available when evaluating interventions to achieve the goals of evidence-based crime policy. Specifically, conclusions about "what works" based on less rigorous evaluations may lead to false optimism about the effectiveness of interventions in reducing crime and recidivism. This in turn might lead to bad decisions based on such research, clearly not a goal of the evidence-based crime policy movement. At the same time, this might also explain why experiments aren't used as often- findings are often less optimistic when researchers use this method.

The SMS ratings about research quality also raised important debates and questions about whether experiments should be rated so highly and have such a major influence on policy (see discussions by Hough 2010; Pawson and Tilley 1997). Although ranking experiments as the highest quality evaluation method available recognizes their scientific value, such rankings also make value judgments about the types of research that should be used to influence policy. The evidence-based crime policy movement not only emphasizes using good science (such as RCEs), but also the translation of that science into practice (Lum 2009; Lum et al. 2012; see also the "Evidence-Based Policing" entry in this encyclopedia). Some have argued these two goals may conflict with each other. In particular, Berk (2005) points out the difficulty in maintaining the fidelity of implementation in experimenting on justice interventions and the consequences for using results in practice (see also Petersilia 1989). Pawson and Tilley (1997) have discussed how experiments may not take into account the realities of interventions and their contexts and, therefore, may not always be the most optimal way of evaluating programs. Further questions have emerged in this debate: What should be the methodological threshold by which evaluation research could be used to inform policy? Are experiments the "best" way to guide policy and practice? The reviews of the research and the rating of studies have brought forth these questions and show that the pursuit of experiments within the evidence-based crime policy movement is complicated.

A second discussion that may further illustrate how the history of randomized controlled trials is entangled in the evidence-based crime policy movement involves policing evaluations within this broader movement. Starting with the early experiments in Kansas City (Kelling et al. 1974), Newark (Police Foundation 1981), and Minneapolis (Sherman and Berk 1984), evaluation research on police crime control interventions has accumulated over the last four decades. These evaluations – too numerous to list here – run the range of police crime control and prevention efforts, and use many different types of

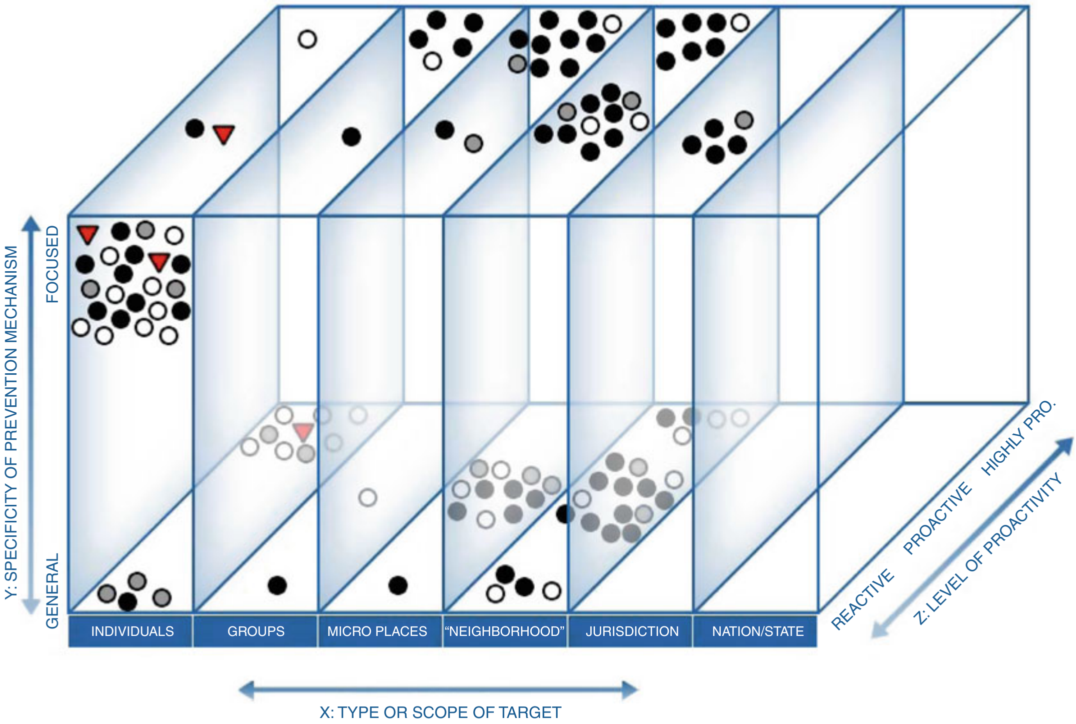

History of Randomized Controlled Experiments in Criminal Justice, Fig. 1 The Evidence-Based Policing Matrix (Lum et al. 2009, 2011). ▼ Significant backfire (increased harm, recidivism, or crime), ○ Nonsignificant finding, ◍ Mixed finding (both significant and nonsignificant findings), ● Successful intervention (statistically significant finding)

research methods, from simple correlation studies to randomized controlled experiments. There have also been a variety of systematic reviews, many conducted for the Campbell Collaboration (see www.campbellcollaboration.org) in addition to the Maryland Report that have organized this research to glean generalizations on "what works." In 2004, the National Research Council also conducted a larger review of the gamut of police research (see National Research Council 2004).

Currently, the most comprehensive and consistently updated systematic review of evaluations of the range of police crime control interventions is the *Evidence-Based Policing Matrix* (Lum et al. 2011; see also www.policingmatrix.org). The Matrix (shown in Fig. 1) is a web-based tool that houses all police crime control intervention research of moderate to high quality. For studies to be included in the Matrix, evaluations must at least include a comparison unit of analysis that did not receive an intervention – similar to a Maryland Report SMS score of "3" or higher. The Matrix classifies all police intervention researches on three very common dimensions of crime prevention strategies: the nature and type of target, the degree to which the strategy is reactive or proactive, and the strategy's level of focus (i.e., the specificity of the prevention mechanism it used). The Matrix also color-codes studies by the outcome and significance of their findings. By "mapping" studies using these three dimensions, generalizations regarding the effectiveness of interventions with these characteristics can be visualized more easily. The visualization allows for multiple aspects of an evaluation to be seen simultaneously including the type of intervention

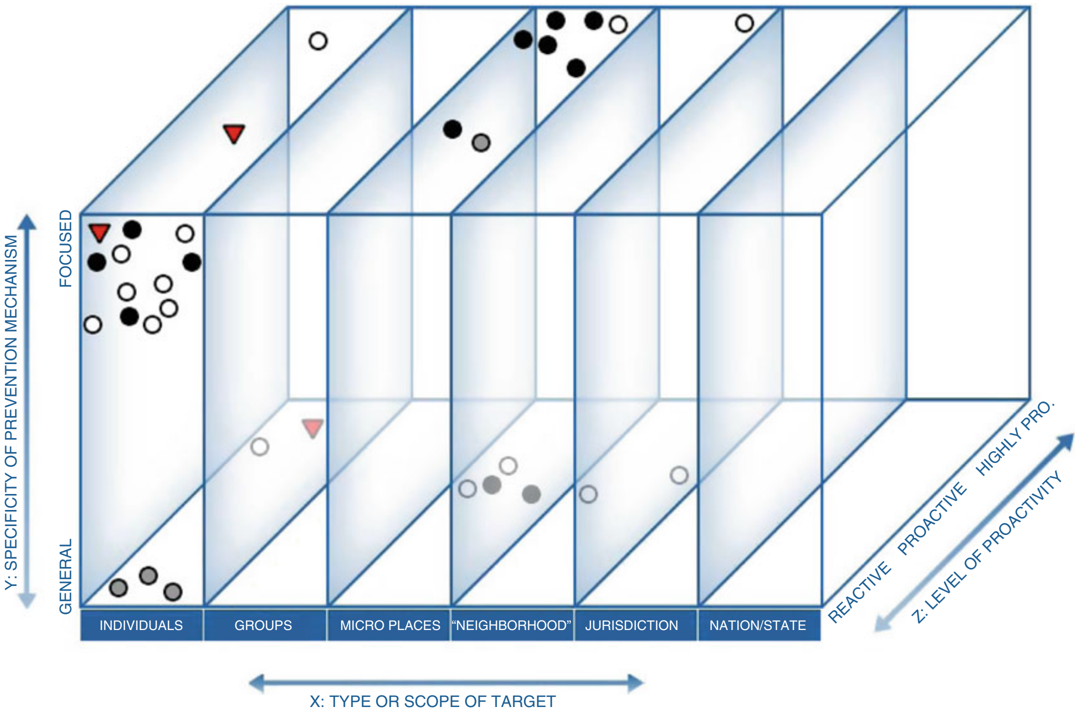

History of Randomized Controlled Experiments in Criminal Justice, Fig. 2 The Evidence-Based Policing Matrix – experiments only. ▼ Significant backfire (increased harm, recidivism, or crime), ○ Nonsignificant finding, ◍ Mixed finding (both significant and nonsignificant findings), ● Successful intervention (statistically significant finding)

studied, the finding of the intervention, and the methodological rigor of the study. The visualization also allows for practitioners to see groupings of research findings quickly, thereby allowing for generalizations from the totality of the research.

The explanation of the Matrix is covered in detail elsewhere (see Lum et al. 2011; also, see Lum and Koper's entry in this encyclopedia under "Evidence-Based Policing"). However, it is used here to illustrate the relationship between experiments and evidence-based crime policy. Because the Matrix provides for generalizations from the large body of police evaluation research through its visualization, it is primarily intended as a research translation tool for practitioners (Lum 2009). This speaks to a core value of evidence-based crime policy – that research should be meaningful and useful to decision makers, not just other researchers. Given this, the Matrix creators decided to include in the visualization studies that were not randomized controlled trials (studies that at least had a comparison group present were also included). Had only experiments been included, the Matrix would appear as it does in Fig. 2.

While Fig. 2 can show a few important ideas for police deployment, and while it also can show where we need more high quality research, some usefulness is also lost from its original form in Fig. 1. As in Fig. 1, we can see that highly proactive, micro-place, focused approaches seem to work well in reducing crime, and that reactive strategies geared toward individuals, even when focused, are not promising (and may cause more harm). However, the sparsely populated Matrix in Fig. 2 may be less practical to police departments trying to move toward an evidence-based approach to developing deployment options because of the scarcity of information caused by the filtering of studies of weaker designs. At the same time, including non-RCTs in the Matrix may lead to false optimism about, for example, the impact of neighborhood-based approaches on crime, of which there are many in Fig. 1. While studies evaluating neighborhood approaches are promising, we know from Weisburd et al. (2001) that there is an inverse relationship between methodological ratings and positive study findings. The authors' choice in including non-RCTs in this translation tool emphasizes their own compromise between pursuing evidence-based crime policy and upholding the importance of experiments in evaluating police interventions. Granted, this does not mean that evaluators should not strive toward using high quality research designs like experiments to evaluate interventions. But the comparison of Figs. 1 and 2 does show if we just focus on experiments, researchers may not have much to offer the policing community with regard to making deployment more evidence based (which could be discouraging for those advocating for evidence-based policing).

Both examples of the Maryland reviews and the Matrix indicate that the dearth of experimentation in criminal justice may not only result from difficulties in doing experiments. Experiments are part of the larger context of the evidence-based crime policy movement, whose proponents are still debating (and reconciling) the role of experiments in criminal justice policy.

The Infrastructure for Experimental Criminology

Amid this healthy debate about the role of experiments in evidence-based crime policy, there have also been efforts to create infrastructure to support experimental criminology that have become part of, and catalysts for, the history of the RCE. For example, the Campbell Collaboration (see http://www.campbellcollaboration.org) was formed – in part by efforts of those involved in the Maryland Report and by philanthropist Jerry Lee – to advocate for more rigorous evaluations of social interventions. Modeled after the Cochrane Collaboration for medical and public health interventions, the Campbell Collaboration sponsors systematic reviews and meta-analyses on specific areas (one major area being crime) to draw conclusions from the totality of evaluation research on a specific topic. The purpose of these reviews is to succinctly inform decisions makers about what the totality of the research says regarding a specific intervention. For example, if five studies examine the impact of pre-textual car stops to reduce gun violence and each has different findings, what can we conclude from all five studies? Like the Maryland Report, implied in this question is the need to rank studies on methodological rigor so that the most believable studies are relied upon more to draw such conclusions (Petrosino et al. 2003). The Campbell Collaboration, like the Maryland Report, again emphasizes the importance of ranking evaluations based on methods, thus emphasizing the importance of experimentation and its connectivity to evidence-based practice and policy. However, the Collaboration also is significant to the history of RCEs because it created organizational infrastructure and social networks around this idea, in not just criminal justice, but social interventions more generally.

Individuals involved in the Campbell Collaboration, especially Joan McCord, David Farrington, Lawrence Sherman, Jerry Lee, David Weisburd, Lorraine Green Mazerolle, Anthony Braga, Joan Petersilia, Frederick Lösel, Doris MacKenzie, Denise Gottfredson, David Olds, Jonathan Shepherd, Anthony Petrosino, Robert Boruch, Mark Lipsey, David Wilson, and others, also helped to build the infrastructure to support experimentation by forming the Academy of Experimental Criminology (AEC) in 1999. The AEC was established to recognize and encourage the efforts of scholars who had conducted experiments in the field through its Academy Fellows distinction. At the time of writing, there were 60 AEC fellows and 13 honorary fellows, all prominent members of the criminology community. In 2009, members of the AEC petitioned the American Society of Criminology membership to form the Division of Experimental Criminology (DEC). The DEC is now one of eight divisions within the American Society of Criminology and, in 2012, combined its efforts with the AEC through a joint web site and mutual activities (for more information on the DEC, see www.cebcp.org/dec). The further prominence and importance of experiments was reflected in the rising prestige of *The Journal of Experimental Criminology (JEC)*. Since its inception (2005) through the end of 2012, 86 of the 164 main articles have been experiments, while another 27 of the articles are meta-analyses or systematic reviews of research.

In addition to these organizations, a number of university centers related to evidence-based crime policy and experimental criminology, as well as more graduate training and presentations at conferences, helped build the infrastructure to support experimentation. Both authors' home centers, the Center for Evidence-Based Crime Policy (see www.cebcp.org) and the Australian Research Council funded Center for Excellence in Policing and Security (see www.ceps.edu.au), are examples of these efforts. Jerry Lee, the philanthropist mentioned above, also established Centers at the University of Pennsylvania and the University of Cambridge to carry out more experimental evaluations. Mr. Lee also helps fund the Stockholm Prize in Criminology, in which a number of experimentalists, including David Weisburd, Jonathon Shepherd, David Olds, and Friedrich Lösel, have been recipients.

The infrastructure supporting experimentation has also been facilitated by the federal funding arena, through in requirements in solicitations for funding by the U.S. Department of Justice Office of Justice Programs and the National Institute of Justice (NIJ). Numerous NIJ directors have supported more experimentation for evaluating crime prevention interventions following the efforts by James "Chips" Stewart who served from 1982 to 1990 and led more recently by Assistant Attorney General Laurie Robinson. In current funding solicitations by NIJ, for example, the following wording appears:

> Within applications proposing evaluation research, funding priority will be given to experimental research designs that use random selection and assignment of participants to experimental and control conditions. When randomized designs are not feasible, priority will be given to quasi-experimental designs that include contemporary procedures such as Propensity Score Matching or Regression Discontinuity Design to address selection bias in evaluating outcomes and impacts.
>
> *Research and Evaluation in Justice Systems* (NIJ solicitation number SL000981)

The Office of Justice Programs, under the leadership of Laurie Robinson, has also developed *CrimeSolutions.gov*, which, like the Maryland Report, judges studies of interventions by their methodological rigor and internal validity. Again, such rankings consistently highlight the importance of experimentation in evidence-based crime policy despite continuing debates.

The Future of Experiments

The history (and future) of randomized controlled experiments in criminal justice and criminology does not occur in a vacuum and cannot simply rely on the scientific merits of experimentation. Not only must RCEs answer to methodological challenges about their ability to establish causation, but perhaps an equally large obstacle to the future of experimentation is whether it can survive criticisms of evidence-based crime policy (and vice versa).

On a more fundamental level, the interaction and history of experimentation and evidence-based crime policy reflects a broader debate and discussion about the role of science in modern democracies. The ideology that science is important in governance and public health seems more accepted in the realm of medicine than social interventions. We hope that medical treatments have been tested in rigorous ways so that we can be assured that treatments will help us. The debates over the merits of experimentation in criminal justice interventions continue because science does not enjoy such acceptance in this realm.

Related Entries

- ▶ Cambridge-Somerville Youth Experiment
- ▶ Differences-in-Differences in Approaches
- ▶ Evidence-Based Policing
- ▶ Longitudinal Studies in Criminology
- ▶ Randomized Experiments in Criminology and Criminal Justice
- ▶ Rational Choice Theory

Recommended Reading and References

Baird J (2011) Descriptive validity and statistical power: a review of experiments published in criminology from 1980–2011. Unpublished manuscript. George Mason University, Fairfax

Berk R (2005) Randomized experiments as the bronze standard. J Exp Criminol 1(4):417–433

Boruch R, Snyder B, DeMoya D (2000a) The importance of randomized field trials. Crime Delinq 46(2): 156–180

Boruch R, Victor T, Cecil JS (2000b) Resolving ethical and legal problems in randomized experiments. Crime Delinq 46(3):330–353

Burtless G (1995) The case for randomized field trials in economic and policy research. J Econ Perspect 9(2):63–84

Campbell D, Stanley J (1963) Experimental and quasi-experimental designs for research on teaching. In: Gage NL (ed) Handbook of research on teaching. Rand McNally: American Educational Research Association, Chicago, pp 171–247

Clarke R, Cornish D (1972) The controlled trial in institutional research: paradigm or pitfall for penal evaluators? vol 15, Home office research studies. Her Majesty's Stationery Office, London

Cook T (2003) Resistance to experiments: why have educational evaluators chosen not to do randomized experiments? Ann Am Acad Polit Soc Sci 589:114–149

Farrington D (1983) Randomized experiments on crime and justice. In: Tonry M, Morris N (eds) Crime and justice: an annual review of research, vol IV. The University of Chicago Press, Chicago, pp 25–308

Farrington D (2003a) Methodological quality standards for evaluation research. Ann Am Acad Polit Soc Sci 587:49–68

Farrington D (2003b) A short history of randomized experiments in criminology: a meager feast. Eval Rev 27(3):218–227

Farrington D, Welsh B (2006) A half-century of randomized experiments on crime and justice. In: Tonry M (ed) Crime and justice, vol 34. University of Chicago Press, Chicago, pp 55–132

Heckman J, Smith J (1995) Assessing the case for social experiments. J Econ Perspect 9(2):85–110

Hough M (2010) Gold standard or fool's gold: the pursuit of certainty in experimental criminology. Criminol Crim Justice 10:11–32

Kelling G, Pate A, Dieckman D, Brown C (1974) The Kansas city preventive patrol experiment: summary report. The Police Foundation, Washington, DC

Lipton D, Martinson R, Wilks J (1975) The effectiveness of correctional treatment: a survey of treatment evaluation studies. Praeger, New York

Lum C (2009) Translating police research into practice, Ideas in American policing lecture series. Police Foundation, Washington, DC

Lum C, Yang S (2005) Why do evaluation researchers in crime and justice choose non-experimental methods? J Exp Criminol 1(2):191–213

Lum C, Koper C, Telep CW (2011) The evidence-based policing matrix. J Exp Criminol 7:3–26

Lum C, Telep C, Koper C, Grieco J (2012) Receptivity to research in policing. Justice Res Pol 14(1):61–96

MacKenzie D (2002) Reducing the criminal activities of known offenders and delinquents: crime prevention in the courts and corrections. In: Sherman LW, Farrington DP, Welsh BC, MacKenzie DL (eds) Evidence based crime prevention. Routledge, London, pp 330–404

MacKenzie D (2008) Examining what works in corrections. In: Proceedings of the 16th annual ICCA research conference keynote address, St. Louis

Martinson R (1974) What works? Questions and answers about prison reform. Pub Interest 35:22–54

Maynard R (1980) The impact of supported work on young school dropouts. Manpower Demonstration Research Corp, New York

Mazerolle L, Price J, Roehl J (2000) Civil remedies and drug control: a randomized field trial in Oakland, CA. Eval Rev 24(2):212–241

McCord J (1978) A thirty-year follow-up of treatment effects. Am Psychol 33(3):284–289

McCord J (2003) Cures that harm: unanticipated outcomes of crime prevention programs. Ann Am Acad Pol Soc Sci 587:16–30

National Research Council (2004) Fairness and effectiveness in policing: the evidence. Committee to review research on police policy and practices. In: Skogan W, Frydl K (eds) Committee on Law and justice, division of behavioral and social sciences and education. The National Academies Press, Washington, DC

Olds D, Henderson C, Cole R, Eckenrode J, Kitzman H, Luckey D, Powers J (1998) Long-term effects of nurse home visitation on children's criminal and antisocial behavior: 15-year follow-up of a randomized controlled trial. J Am Med Assoc 280:1238–1244

Palmer T, Petrosino A (2003) The "experimenting agency": the California youth authority research division. Eval Rev 27:228–266

Pawson R, Tilley N (1997) Realistic evaluation. Sage, London

Petersilia J (1989) Implementing randomized experiments – lessons from BJA's intensive supervision project. Eval Rev 13(5):435–458
Petersilia J, Turner S (1993) Intensive probation and parole. Crime Justice Rev Res 17:281–335
Petrosino A, Boruch R, Farrington D, Sherman L, Weisburd D (2003) Toward evidence-based criminology and criminal justice: systematic reviews, the Campbell collaboration, and the crime and justice group. Int J Comp Criminol 3:42–61
Powers E, Witmer H (1951) An experiment in the prevention of delinquency. Columbia University Press, New York
Pratt T, Gau J, Franklin T (2011) Key ideas in criminology and criminal justice. Sage, Los Angeles
Shadish W, Cook T, Campbell D (2002) Experimental and quasi-experimental designs for generalized causal inferences. Houghton-Mifflin, Boston
Shepherd J (2003) Explaining feast of famine in randomized field trials. Eval Rev 27(3):290–315
Sherman L, Berk R (1984) The Minneapolis Domestic Violence Experiment. Police Foundation, Washington, DC
Sherman L (1998) Evidence-based policing, Second invitational lecture on ideas in policing. Police Foundation, Washington, DC
Sherman L, Weisburd D (1995) General deterrent effects of police patrol in crime hot spots: a randomized controlled trial. Justice Quart 12:625–648
Sherman L, Gottfredson D, MacKenzie D, Eck J, Reuter P, Bushway S (1997) Preventing crime: what works, what doesn't, what's promising: a report to the United States Congress. National Institute of Justice, Washington, DC
Sherman L, Farrington D, Welsh B, MacKenzie D (eds) (2002) Evidence based crime prevention. Routledge, London
Sparrow M (2009) One week in Heron city. In: Harvard executive session on policing and public safety. National Institute of Justice, U.S. Department of Justice, Washington, DC
The Police Foundation (1981) The Newark foot patrol experiment. Police Foundation, Washington, DC
Weisburd D (2000) Randomized experiments in criminal justice policy: prospects and problems. Crime Delinq 46(2):181–193
Weisburd D (2003) Ethical practice and evaluation of interventions in crime and justice: the moral imperative for randomized trials. Eval Rev 27(3): 336–354
Weisburd D, Green L (1995) Policing drug hot spots: the Jersey city drug market analysis experiment. Justice Quart 12:711–736
Weisburd D, Petrosino A (2004) Experiments, criminology. In: Kempf-Leonard K (ed) Encyclopedia of social measurement. Academic, San Diego
Weisburd D, Lum C, Petrosino A (2001) Does research design affect study outcomes? Ann Am Acad Pol Soc Sci 578:50–70

History of Restorative Justice

Karen Heetderks Strong

Overview

In the final decades of the twentieth century, an approach to criminal justice, called restorative justice, emerged in response to concerns about offender recidivism, corrections costs, and crime victims' rights, along with encouraging outcomes in the use of victim-offender dialogue and restitution. Rather than defining criminal justice in terms of prosecution and punishment only, restorative justice focused on identifying and repairing the harms caused by criminal acts, holding offenders accountable for making things right with those harmed, and empowering victims and communities to participate in the response. In addition, restorative justice considered the proper roles and responsibilities of both government and communities for a just society.

Roots of Restorative Justice

Much modern legal structure and popular thinking about criminal justice views crime as if the government and its law are the injured parties. Violation of a government-imposed law calls forth a series of government-initiated responses: investigation and prosecution, followed by imposition of a sentence proportionate to the offense. The sentence frequently involves a fine (payable to the government) and some period of supervision or incarceration to express the legal and social condemnation of the proscribed behavior. By serving their sentence, criminal offenders are said to "pay their debt to society" but this usually only abstractly relates to the harm experienced by the actual victims of a crime. The idea that offenders have wronged individuals and violated communities, and thus must make reparation to those directly harmed, is at the heart of restorative justice. Its roots can be found in the legal codes of the ancient Near East, Rome, Europe, and Britain.

Examples of this direct connection between the harm to victims and the sentence may be found in the Code of Hammurabi (c. 1700 B.C.E.) and the Code of Lipit-Ishtar (1875 B.C.E.), which prescribed payment of restitution to the victims of property offenses. Other codes, such as the Sumerian Code of Ur-Nammu (c. 2050 B.C.E.) and the Code of Eshnunna (c. 1700 B.C.E.), provided for restitution even in the case of violent offenses. The Roman Law of the 12 Tables (449 B.C.E.) required thieves to pay double restitution unless the property was found in their houses, in which case they paid triple damages; for resisting the search of their houses, they paid quadruple restitution. The Lex Salica (c. 496 C.E.), the earliest existing collection of Germanic tribal laws, included restitution for crimes ranging from theft to homicide. The Laws of Ethelbert (c. 600 C.E.), promulgated by the ruler of Kent, contained detailed restitution schedules that distinguished the values, for example, of each finger and its nail. Each of these diverse legal codes represent a mandate that offenders and their families should make amends to victims and their families—not simply to ensure that injured persons received restitution but also to restore community equilibrium and avoid repeating cycles of revenge. Although actual practice may not have lived up to this intent, the presence of restitution as a legal obligation is significant.

In addition to these examples, there is evidence that practices in precolonial African societies showed a willingness to forgo punishment of criminal offenders in order to address the consequences to their victims and the community. Many sanctions were compensatory rather than punitive, intended to restore victims to their previous position. Contemporary Japanese culture exhibits a similar emphasis on compensation to the victim and restoration of community peace. The traditions of indigenous peoples, for example, the Maoris in New Zealand and First Nations in Canada, allowed for working out an appropriate response to crime in the context of the affected community. Offenders were involved in the process of determining guilt and harm and taking necessary steps toward restoration of those harmed. Practices associated with restorative justice such as group conferencing or circle processes have their roots in these indigenous ways.

Validated in part by these historical and indigenous roots, several twentieth-century social movements or critiques fed into the emergence of restorative justice theory. In addition to what may be generally called the informal justice movement, there were explorations and proponents of restitution, victims' rights, prison abolition, and alternative theories of social justice.

Informal Justice Movement

In the 1970s, amid a growing crisis of confidence in the legitimacy of formal, government structures, legal anthropologists noted that laws and legal processes are specific to their social and historical context and that, in virtually all societies, justice is pursued using both formal *and* informal proceedings. A series of proposals emerged for informal alternatives to government-controlled responses to crime, emphasizing greater participation of the parties, de-professionalization of the processes, and lessening both stigmatization and coercion for participants. Two leading proponents of informal justice were Jerold S. Auerbach and Nils Christie. Christie has been frequently cited in restorative justice literature because of seminal themes in his work. In his article, "Conflict as Property," Christie (1977) made the case that the criminal justice system reflects the government's theft of the dispute that is properly owned by the victim and offender. In his subsequent book on punishment, *Limits to Pain* (1981), Christie drew a connection between this "theft" and the use of punishment. In criminal law, values are clarified by graduated punishment. Rather than being established through conversation among the participants—the rightful "owners" of the conflict—values are communicated by the state through the infliction of degrees of pain. He proposed participatory justice as a better response to crime, a response characterized by direct communication between the owners of the conflict leading to compensation that satisfies the claims established between them.

Restitution Initiatives

Rediscovery of the historical notion that paying back the victim could be an appropriate modern criminal justice sanction led to *restitution initiatives* beginning in the 1960s. Several rationales were offered for the use of restitution: (1) the victim is the party harmed by criminal behavior, (2) restitution offers an alternative to restrictive or intrusive sanctions such as imprisonment in some cases, (3) there may be rehabilitative value in requiring the offender to pay the victim, (4) restitution is relatively easy to implement, and (5) this might lead to a reduction in more costly retributive sanctions when the public observes the offender actively repairing the harm done. Evaluations conducted in the 1970s and 1980s raised questions about whether restitution programs had adequately lived up to those expectations, but restitution has been increasingly ordered (and less frequently collected) in many countries nonetheless.

One proponent of restitution was Stephen Schafer (1970). He described the era of compensatory justice prior to the development of centralized governments in Europe as the "golden age of the victim" because it was a time in which the victim's interests and freedom of action were given greatest deference. He proposed that compensation could once again become a means of sanctioning offenders, either in conjunction with or as an alternative to imprisonment.

Abel and Marsh (1984) argued that restitution offers an approach to punishment that is ethically, conceptually, and practically superior to the status quo. In their model, most offenders would live outside prison under varying degrees of supervision as necessary, working and paying restitution to their victims. Imprisonment should be used only as a last resort for truly dangerous offenders and even they should be given the opportunity and obligation to earn wages and compensate their victims as well as paying the state to offset the cost of their incarceration. This recognition of the obligation of offenders for redress to their victims flowed into the formation of restorative justice theory.

Victimology and the Crime Victims' Movement

Beginning as early as the mid-1950s, some scholars took an interest in the needs and rights of crime victims and the deficiencies of an offender-oriented justice system. The notion that victims, not just offenders, ought to be considered in the study and response to crime began to gather momentum. In 1963, the first crime victim's compensation law passed in New Zealand. In the following years, similar initiatives were established in England and parts of the United States, Canada Australia, and Europe. In 1972, the United States federal government began the *National Crime Survey* (renamed the *National Criminal Victimization Survey* in 1990), signaling a desire to discover the extent of victimization beyond what was officially reported to law enforcement agencies.

The *First International Symposium on Victimology* was held in 1973, primarily dealing with research on societies' attitudes toward crime victims. Two years later, an *International Study Institute on Victimology* went further and introduced discussion of the treatment of victims and their standing in the criminal justice system. In 1975, Frank Carrington published *The Victims*, asserting that victims' "current sorry status in the criminal justice system need not be so and that something can and must be done to enhance the rights of the victim." Attention began to be focused on ways crime victims' interests should be specifically accommodated in criminal justice and community processes. The *crime victims' movement*, thus, rallied around victims' experiences and applied victimology research; it organized advocacy coalesced in three broad areas: (1) appropriate services to victims in the aftermath of the crime, (2) financial compensation for the harm done, and (3) victims' rights to information and intervention during the course of the criminal justice process.

Basic rights and protections for victims of crime began to be included in statutes in the late 1970s, and in 1982, the State of California became the first in the USA to pass a constitutional amendment guaranteeing crime victims' certain enforceable rights. Other jurisdictions followed suit. Specific provisions vary, but generally victims are granted the right to be treated fairly and with dignity, to attend the trial or other proceedings, to be awarded restitution, to be informed of certain events such as the offender's release and, further, to be heard at critical stages in the case such as sentencing or parole hearings.

All of these efforts and outcomes have served to make it explicit that crimes violate persons, not just laws; that victims have a legitimate and necessary role beyond simply giving evidence; and that the harms to victims ought to be acknowledged and addressed in a society's response to crime. These tenets have been incorporated into the idea of restorative justice.

Prison Abolition Movement and Alternatives to Incarceration

Questions about the use and validity of prisons were also gaining attention and during the late 1960s and into the 1970s, an informal reform initiative emerged in Europe and North America calling for the *abolition of prisons*. It garnered support from a number of political and philosophical perspectives around the common theme that prisons not only fail to rehabilitate, they are in fact places of acute and morally indefensible suffering for prisoners. Although the Society of Friends (Quakers) helped invent the "penitentiary" in the late 1700s, they joined the prison abolition movement in urging that the use of prisons be significantly curtailed or abolished and replaced by other responses to crime. In part, this was due to prison abuses—abuses that seemed to be inherent in the institution of prison itself and consequently not amenable to reform.

Some proponents of prison abolition called for prisons to be done away with completely. Others sought to decrease the use of prisons dramatically. Still others campaigned for a moratorium on construction of new prisons. In place of prisons, they suggested that restitution, compensation, community service, and reconciliation programs be established in local communities so that the response to crime could be decentralized and, to some extent at least, humanized.

This notion was appealing to jurisdictions facing budget crises due to economic recession in the 1980s. Community corrections programs providing alternatives to expensive incarceration under the auspices of probation and parole agencies included such things as day reporting centers, halfway houses, work release, and community service. These initiatives were organized with multiple goals including offender accountability, rehabilitation, surveillance, and fiscal efficiency. Proponents built public support for these alternatives by emphasizing that they impose a higher level of accountability on offenders. Rather than sitting out their time in a cell, community corrections required them to take an active role such as doing community service or working in order to make restitution to their victims.

Alongside these developments, programs emerged involving mediated conversations between victims and offenders with the aim of a mutually agreed and enforceable reparative outcome. Initially, such programs focused on juveniles and diverted them from the formal criminal justice system, with judges' cooperation in accepting and helping enforce reasonable agreements from the mediation. Rather than keeping victims and offenders away from each other, their interests represented principally by attorneys, these programs included them in informal processes that emphasized safety, dialogue, and agreement on appropriate redress and next steps. Outcome research found that the parties generally valued the opportunity to speak and be heard and had a higher level of satisfaction that justice had been served in the process, compared with those who simply went through the standard criminal justice approach. These experiments fed directly into the theoretical and practical development of restorative justice.

Social Justice Critiques

A conviction that justice could not be achieved in an unjust society found its voice in publications and advocacy alongside the prison abolitionists' practical concerns about prison abuses. One influential expression of this concern was *Struggle for Justice*, the report from a working group assembled by the American Friends Service Committee (1971). The report articulated concerns about sentencing practices that were disproportionately adverse for poor defendants and people of color and also the coercive nature of so-called rehabilitative programs in corrections. Similar concerns about disproportionate penalties and disparate sentencing practices contributed to the rise of the "just deserts" model for corrections. Andrew von Hirsch, in *Doing Justice* (1976), proposed that it is just for society to punish someone for a criminal offense but unjust to tailor the punishment based on considerations of future offending (such as deterrence or rehabilitation). Rather, penalties must be proportionate to the severity of an offense and uniformly administered so that all similar offenses carry the same sentence. This important development, however, did not address the concerns expressed in *Struggle for Justice* about harsh penalties, degradation of offenders, and the brutality of prisons.

Gerald Austin McHugh's *Christian Faith and Criminal Justice* (1978) argued that although Western penal models grew out of a medieval religious view of sin and retributive punishment, this was not the only relevant motif inherent in Judeo-Christian theology. Its profound themes of mercy, relationship, restoration, forgiveness, reconciliation, and hope ought also to be applied in response to crime. McHugh suggested that application of these values to criminal justice policy would result in very different structures and processes from those now in place. Daniel W. Van Ness' *Crime and Its Victims* (1986) and Howard Zehr's *Changing Lenses* (1990) also made the case that biblical justice is highly concerned with the needs and rights of victims, as well as with the worth of offenders, and that this has implications for the foundations and policies of a just society.

From a feminist perspective, scholar Harris (1987) called for a fundamental restructuring of criminal justice to reflect the equal value of all human beings, the social precedence of harmony and well-being over power and possession, and the assertion that the personal is the political. The criminal justice status quo, she asserted, is driven by the values of control and punishment. She suggested that a preoccupation with legitimate rights blinds parties to the need for a caring and interdependent response and that justice has a broader scope than individual rights. Citing Gloria Steinem's principle that the means you use are reflected in the ends you achieve, Harris called for a recognition and response to the need and opportunity for participation by all parties in order to justly address the needs of all.

Criminologists Pepinski and Quinney (1991) also explored the idea that crime has a personal dimension in which society has a stake. They examined the factors that positively contribute to peace and safety and cited these as the proper focus of criminology, as opposed to a negative focus ("what causes deviance and criminal behavior?," and "how do we win the war on crime and violence?"). Others echoed this positive, peacemaking focus for criminology, asserting that the goals of peace and safety require more than restraint of wrongdoing and its punishment. They need to be achieved through transformation of persons and social institutions, affirming individual responsibility for wrongdoing and also society's contribution to it and implementing effective, fair, humane, and peace-building community responses.

All of these movements influenced and contributed to the development of restorative justice and their themes are evident in its theory and practice. Johnstone and Van Ness (2007) have identified three main concepts that typically are incorporated (in some way) in definitions of restorative justice. The first is *encounter*. The parties involved (e.g., victim, offender, community members) are engaged together in discussing the offense, its effects, and what should be done about it. Rather than being done for them by professionals in a court of law, the parties themselves engage in a dialogue that is safe and fair for all concerned. The second concept is *reparation*, meaning that the response to an offense should involve redress by the offender to those harmed.

Preferably, the redress would be decided with the involvement of all parties. But if that is not possible, reparation in some form could and should still be made in order for the outcome to be considered at all restorative. The third concept is *transformation*; it is more expansive than the others. It recognizes social and structural issues that either are themselves unjust or create impediments to fair, harmonious, and healthy relationships among people and their social and physical environments. So beyond the restoration of individual instances of harm, restorative justice aims for a more just society and world.

The Meaning of the Term "Restorative Justice"

The first use of the term "restorative justice" was by Albert Eglash in several 1958 scholarly articles in which he suggested that there are three types of criminal justice: (1) retributive justice, based on punishment; (2) distributive justice, based on therapeutic treatment of offenders; and (3) restorative justice, based on restitution. Both the punishment and treatment models, he noted, focus on the actions of offenders, deny victim participation in the justice process, and require merely passive participation by the offender. Restorative justice, on the other hand, focuses on the harmful effects of offenders' actions and actively involves victims and offenders in the process of reparation and rehabilitation. Eglash's ideas and terminology were rediscovered decades later.

From its more popular emergence in the early 1990s up to the present day, the term has come to have a variety of meanings and nuances. Some definitions focus more on outcomes, others on processes. Howard Zehr (1990) described restorative justice in this way: "Crime is a violation of people and relationships. It creates obligations to make things right. Justice involves the victim, the offender, and the community in a search for solutions which promote repair, reconciliation, and reassurance." Martin Wright (1991) wrote that restorative justice would respond to crime in such as way as "not to add to the harm caused by imposing further harm on the offender, but to do as much as possible to restore the situation. The community offers aid to the victim; the offender is held accountable and required to make reparation. Attention would be given not only to the outcome, but also to evolving a process that respected the feelings and humanity of both the victim and the offender." Tony Marshall (1999) described it as "a process whereby all the parties with a stake in a particular offense come together to resolve collectively how to deal with the aftermath of the offense and its implications for the future." Van Ness and Strong (2010) define it as "a theory of justice that emphasizes repairing the harm caused or revealed by criminal behavior. It is best accomplished through cooperative processes that include all stakeholders."

The term, restorative justice, today is used throughout the world to describe processes and programs that offer alternatives to simply punitive approaches to handling offending and sanctions. Over time, the term has even gone beyond its original scope as a response to crime and its ideas are finding applications in handling neighborhood disputes, school discipline, and even workplace problems. All of these are built on foundations laid by some key developers of the theory of restorative justice.

Developments in Restorative Justice

Howard Zehr is sometimes called the "grandfather" of restorative justice. Zehr's early experience with victim-offender reconciliation programs led to articles, speeches, books, and teaching that have drawn others into theoretical and practical conversations about justice. In his 1990 book, *Changing Lenses*, Zehr consolidated and advanced his critique of criminal justice as failing to meet the needs of victims or offenders. He suggested that the current criminal justice "lens" views crime as lawbreaking and justice as allocating blame and punishment. He contrasted that with restorative justice, which views crime as a violation of people and relationships, which in turn leads to obligations to "make things right." Justice, then, is a process in which all parties search for reparative, reconciling, and reassuring solutions.

Another pioneer is Martin Wright, who contributed to the development of restorative justice thinking and practice, particularly in Europe. In his 1991 book, *Justice for Victims and Offenders*, Wright drew from his experiences as an advocate for victims and for prison reform. He argued that criminal justice should be restorative both in its process and its effect by expanding compensation, restitution, and mediation opportunities to permit greater participation by both victims and offenders. Wright suggested two governmental departments to facilitate this: one for crime prevention, emphasizing deterrence through enforcement rather than through punishment; the other to ensure a just response to crimes when they do occur, including victim support, mediation, and reparation, as well as courts that emphasize restitution.

In 1992, Virginia Mackey wrote an evocative discussion paper on restorative justice to facilitate conversation in the faith community on the problems of current approaches to crime and on biblically reflective alternatives. She proposed a restorative model built on six practical principles: (1) safety should be the primary consideration for the community, (2) offenders should be held responsible and accountable for their behavior and the resulting harm, (3) victims and communities harmed by crime need restoration, (4) the underlying conflicts that led to the harm should be resolved if possible, (5) there must be a continuum of service or treatment options available, and (6) there must be a coordinated and cooperative system in place that incorporates both public and private resources.

Whereas some discussions presumed that formal criminal justice processes are inherently counter to the foundational values of a restorative approach, Cragg (1992), a philosopher and a longtime volunteer with a prisoner advocacy and prison reform organization, thought otherwise. He found insupportable an emphasis on punishment, whose aim is to cause suffering. However, Cragg underscored the importance of formal processes that should provide within their frameworks the opportunity for informal resolution and offender acceptance of responsibility. Formal justice, in his view, need not be punishment-driven, but may incorporate restorative virtues such as forgiveness, compassion, mercy, and understanding.

Another theoretical framework related to restorative justice is "reintegrative shaming," a term used by Braithwaite (1989)—an influential critic of the just deserts movement and contributor to restorative thinking about justice. The term refers to positive ways for offenders to acknowledge the wrongness of acts for which they should rightly be ashamed. The point of the process is not shame but rather reintegration—concrete steps to take the offender from shame to being a fully accepted individual within the community. But it was not until a family group conferencing program was being organized in Wagga Wagga, a city in New South Wales, Australia, that reintegrative shaming and restorative justice were intentionally linked. In 1993, David Moore reflected on this approach to crime and reintegration from the perspective of moral psychology, moral philosophy, and political theory. He concluded that reintegrative shaming offered a framework for theoretical analysis and evaluation of the restorative potential of conferencing programs. It is important to note that the notion of "shaming" has been quite controversial among restorative justice advocates. Some believe shaming is inherently dehumanizing and thus is an illegitimate means even if it leads to good outcomes. Others distinguish between stigmatizing and reintegrative shaming, wherein the former makes shame an end in itself that is attached to the offender's identity and the latter is part of a humane process leading to removing the shame and restoring the individual.

Restorative justice includes victims as primary stakeholders who should play a direct role in the justice process. Some assert that there is a historical basis for questioning the criminal-civil separation in Western legal systems and allowing victims to have full standing to advocate on their own behalf for redress as a part of criminal proceedings. A scholarly debate on this point was published in 1993 in the *Criminal Law Forum*. Daniel Van Ness made the case for the idea that victims could be fully and procedurally included in criminal justice processes aimed at addressing the harms experienced by all stakeholders. Andrew Ashworth supported victims' interests in certain outcomes but warned that it is important to distinguish between the needs of victims for assistance and any legal rights that they might have in criminal courts. He also cautioned about attempting to accomplish larger criminal justice goals through sentencing policy. Together with von Hirsch, Ashworth has defended the objectives of just deserts as preferable to those of restorative justice for both just processes and outcomes.

An attempt to root restorative criminal justice ideas within a broader conceptual framework was offered in Burnside and Baker (1994), *Relational Justice*. Noting erosion in the quality of relationships in Western cultures, these authors considered whether "relationalism" might provide an antidote to problems including criminal justice. Although not specifically referring to restorative justice theory, contributors presented victim-offender mediation and family group conferences as examples of relational justice and suggested ways the activities of police, probation, and prison authorities might be evaluated by their capacity to strengthen relationships.

From the beginning, restorative justice's main focus was how to respond to crime. A number of writers have noted in passing that restorative justice principles may have relevance to crime prevention also. One of the more comprehensive proposals on that aspect of criminal justice policy was offered by Marlene A. Young, Executive Director of the US-based National Organization for Victim Assistance, in her paper *Restorative Community Justice*: *A Call to Action* (1995). She proposed a series of program elements that might constitute a model of restorative justice as a community responsibility with potential to develop more effective problem solving and foster healthier, less crime-prone communities. The elements of this model included community policing, community prosecution, community courts, and community corrections. The first, community policing, involves police officers actively building strong community bonds within the neighborhoods in which they function. Although none of these elements were new ideas, Young envisioned their integration into an overall structure that would shift from reactive to proactive engagement and increase the level of meaningful victim and community participation in prevention, adjudication, and corrections.

Victim advocates have often been concerned that restorative justice is actually offender-focused, suspecting that its programs are primarily designed to provide better conditions for offenders, with victims a secondary concern. Some claim that victims are being used as agents to achieve offender-focused goals. While this may not be accurate in theory, studies of actual practices of restorative justice show that there is validity in these concerns in at least some instances. For example, those who work with victims in some restorative justice programs are volunteers or professionals who lack sufficient expertise and understanding of the victims' needs and perspectives. The very methods designed to give victims a voice, such as mediation or circles, are sometimes compromised by pressuring victims to participate or giving them more of a sideline role. Efforts to inform and contact victims for participation are sometimes short-changed in the interest of moving to a quicker and less complicated process.

So-called restorative programs occur in the midst of social contexts and are affected by such variable factors as public mood, government acceptance, and available funding. Evaluations of practices and outcomes help monitor the extent to which the idea of restorative justice meets the reality of its implementation. Advocates naturally highlight a wide variety of studies that underscore the good outcomes achieved through restorative justice processes. Critics point out that the same attention must be paid to studies that reveal problems since the good intentions of restorative justice practitioners and proponents may not necessarily result in processes and outcomes that are, in fact, restorative.

Summary

Restorative justice is now a global idea. Not only has it been influenced by people and developments in many parts of the world, it has also been adopted in diverse cultures, economies, political systems, and legal structures. In fact, one of the remarkable features of restorative justice is that elements of it emerged almost simultaneously in different regions of the world. In some cases, the program or theory was developed prior to contact with the ideas of restorative justice being discussed elsewhere.

One way to sum up the history of restorative justice is to consider several stages of its growth and reach:

Stage One: *Restorative justice as a community-based alternative to the criminal justice system*. Early on, particularly in common law countries, restorative processes were viewed as an alternative, community-based response to crime that diverted some offenders out of the government-run process in favor of a more personalized and participatory one. That was certainly true of the first victim-offender reconciliation programs in Canada, the United States, and England. In countries in continental Europe, where the legality principle limits the discretion of police and prosecutors to divert cases out of the justice system, some legislative changes were needed, but the purpose of the changes was to allow pilot projects for demonstration and research.

Stage Two: *Restorative justice as a resource for public policy*. In the latter half of the 1980s, countries and organizations began exploring the policy implications of what was being learned from restorative processes. For example, Justice Fellowship, a nonprofit in the United States, embarked on a multiyear project to translate restorative vision into principles of public policy; Canada's Parliamentary Standing Committee (known as the Daubney Committee for its chairman) affirmed that sentencing law changes in that country should take into consideration what was being learned in restorative processes. New Zealand incorporated elements of Maori practices in revising its juvenile justice laws.

Stage Three: *Restorative justice as a viable part of the criminal justice system*. The final decade of the twentieth century saw remarkable innovation and expansion in restorative justice. New models, and variations on those models, were developed as conferencing and circles took their place with victim-offender mediation. Governments took steps to encourage the use of restorative processes through grants, research, legislative change, and, in some instances (such as the Thames Valley Police and the Royal Canadian Mounted Police), through offering restorative processes themselves. Organizations dedicated to promoting and expanding restorative processes emerged, offering practitioners and researchers the opportunity to network with others and to learn from their experiences. This information sharing has expanded dramatically through the Internet and the growing number of excellent sites dedicated to restorative justice.

Stage Four: *Restorative justice as an international reform dynamic*. The twenty-first century has seen multinational bodies such as the United Nations, the Council of Europe, and the European Union strongly endorse the potential of restorative justice and urge their member states to introduce and then expand their use of restorative processes. Restorative concepts have contributed to the development of Truth and Reconciliation Commissions to help address the deep social and personal consequences of violence and systemic injustice. Where early work with restorative processes was done by grassroots organizations, now they have been joined by multinational organizations in calling on governments to use these new approaches to criminal justice.

The future of restorative justice as an idea is intriguing. Although it has become an accepted and fairly common term, its specific meaning is far from uniform. In some ways it has become a catchall applied to a wide range of practices and policies that have one or another element typically associated with restorative justice. In addition to applications in the criminal justice area, restorative approaches are being considered and applied to conflict and discipline in schools, workplaces, and mass social conflicts. All these realities present a challenge to the idea of restorative justice. Some scholars have called for clear distinctions in terminology between theoretical, values-based foundations for restorative justice and their expressions in practice. Others have suggested a set of evaluative criteria or standards for what constitutes a restorative approach. Van Ness and Strong (2010) offer a framework for assessing a system to determine whether it is minimally, moderately, or fully restorative or if it is restorative at all regardless of its claims. In any case, the idea of restorative justice continues to develop, along with a proliferation of programmatic expressions that shows no sign of waning.

Related Entries

- ▶ Community Courts
- ▶ Community Service in Europe
- ▶ Development of the UCR and the NCVS
- ▶ Empathy and Offending
- ▶ Feminist Criminological Theory
- ▶ Historical and Comparative Perspectives on Incarceration
- ▶ History of Corrections
- ▶ History of Fines
- ▶ Informal Social Control
- ▶ International Responses to Victims in Criminal Justice
- ▶ Juvenile Diversion
- ▶ Law, Diversion and Community Sanctions in Juvenile Justice
- ▶ Monitoring and Evaluation of Restorative Justice
- ▶ Penal Abolitionism
- ▶ Penal Justice and Social Injustice
- ▶ Penal Philosophy and Sentencing Theory
- ▶ Preventive Order: Civil Law
- ▶ Punishment and the State
- ▶ Rehabilitation
- ▶ Restitution as a Penal Aim
- ▶ Restorative Justice and Practice
- ▶ Restorative Justice and State Crime
- ▶ Sentencing as a Cultural Practice
- ▶ Sex Offenders and Criminal Policy
- ▶ Shame in Criminological Theory
- ▶ Theories of Punishment
- ▶ Traditional Retributivism
- ▶ Unintended Effects of Imprisonment

Recommended Reading and References

Abel CF, Marsh FA (1984) Punishment and restitution: a restitutionary approach to crime and the criminal. Greenwood Press, Westport

American Friends Service Committee (1971) Struggle for justice: a report on crime and punishment in America. Hill and Wang, New York

Auerbach J (1983) Justice without law? Oxford University Press, New York

Braithwaite J (1989) Crime, shame, and reintegration. Cambridge University Press, Cambridge

Burnside J, Baker N (eds) (1994) Relational justice: repairing the breach. Waterside Press, Winchester

Carrington FG (1975) The victims. Arlington House Publishers, New Rochelle

Christie N (1977) Conflict as property. British Journal of Criminology 17(1):1

Christie N (1981) Limits to pain. Universitetsforlaget, Oslo/Bergen/Tromsø

Clear T, Karp DR (1999) The community justice ideal: preventing crime and achieving justice. Westview Press, Boulder

Cornwell DJ (2006) Criminal punishment and restorative justice: past, present and future perspectives. Waterside Press, Winchester

Cragg W (1992) The practice of punishment: towards a theory of restorative justice. Routledge, New York

Harris MK (1987) Moving into the new millennium: toward a feminist vision of justice. The Prison Journal 67(2):27–38

Hudson J, Galaway B (eds) (1977) Restitution in criminal justice. D.C. Heath, Lexington

Johnstone G (ed) (2003) Restorative justice reader: texts, sources, context. Willan, Portland

Johnstone G, Van Ness DW (2007) Handbook of restorative justice. Willan, Cullompton

Marshall T (1999) Restorative justice: an overview. Home Office Research Development and Statistics Directorate, London

McHugh GA (1978) Christian faith and criminal justice: toward a Christian response to crime and punishment. Paulist Press, New York

Moore D (1993) Shame, forgiveness, and juvenile justice. Crim Justice Ethics (Winter/Spring):3

Pepinski HE, Quinney R (eds) (1991) Criminology as peacemaking. Indiana University Press, Bloomington

Schafer S (1970) Compensation and restitution to victims of crime. Patterson Smith, Montclair
Shapland J (2008) Justice, community and civil society: a contested terrain. Willan, Devon
Strang H, Sherman LW (2003) Repairing the harm: victims and restorative justice. Utah Law Review 1:15–42
Sullivan D, Tifft L (eds) (2006) Handbook of restorative justice: a global perspective. Routledge, London
Van Ness DW (1986) Crime and its victims: what we can do. InterVarsity Press, Downers Grove
Van Ness DW, Strong KH (2010) Restoring justice: an introduction to restorative justice, 4th edn. Matthew Bender, New Providence
Von Hirsch A et al (2005) Restorative justice and criminal justice: competing or reconcilable paradigms? Hart, Oxford
Walgrave L (2008) Restorative justice, self-interest and responsible citizenship. Willan, Cullompton
Wright M (1991) Justice for victims and offenders. Open University Press, Philadelphia
www.restorativejustice.org
Young MA (1995) Restorative community justice: a call to action. National Organization for Victim Assistance, Washington, DC
Zehr H (1990) Changing lenses: a new focus for crime and justice. Herald Press, Scottdale
Zehr H, Toews B (eds) (2004) Critical Issues in restorative justice. Criminal Justice Press, Monsey

History of Staff Skills in Probation

Ioan Durnescu
Faculty of Sociology and Social Work,
University of Bucharest, Bucharest, Romania

Overview

In the last few years, the idea that "nothing works" has been systematically confronted, and more and more knowledge is accumulating to demonstrate that a lot could be done to reduce reoffending and improve public safety. An extensive body of literature now emphasizes what is known in terms of evidence-based practice and interventions based on different theoretical perspectives (risk/needs/responsivity, desistance, good lives model). Increasingly, research uncovers other aspects that influence the effectiveness of interventions, like treatment readiness and motivation for change, organizational structures, neighborhood characteristics, and so on. From time to time, a focus on the skills and characteristics of probation and parole staff appears in the literature and then disappears again. At the origins of probation, staff skills seemed to be quite crucial. Later they became important but not essential, and then they almost disappeared in the years of "nothing works." As a new body of evidence on "What Works" and desistance is accumulating, staff attributes make their way back into the mainstream of effective practice. Indeed, empirical research conducted mainly after the 1990s seems to support the idea that not only program content or offender attributes but also the characteristics and style of the personnel in charge of the implementation of probation interventions matter for effective practice. With this optimistic view a new line of research has developed in recent years on "who works" in probation service. However, some aspects remain unexplored, including the factors that explain differences between professionals in terms of skills and characteristics and how particular skills are associated with different supervision results.

In this chapter the concept of staff skill is understood as a behavior that could be learnt and is aimed at supporting the probation officer in doing his/her job. In contrast, others consider staff characteristics to be inner traits of the probation officer that are relevant to his/her job (e.g., morality, reliability). Obviously between these two categories, there is not a clear-cut but a fluid line that can be crossed sometimes, depending on the circumstances. One typical example is empathy. Some authors consider empathy as inherited or a moral quality, while others emphasize the idea that empathy could be taught and therefore could be considered as a professional skill (Trotter 2006).

Men of God

A number of scholars (Jarvis 1972; Vanstone 2004) suggest that the origin of probation was in the work of the Church of England Temperance

Society (in England and Wales) and the Washingtonian Total Abstinence Society (Boston, United States). The activity of early probation officers was almost exclusively missionary and was based on "Christian charity and humanitarian concern" (Jarvis 1972, p. 1). The main duty of the probation officer was to "save the souls" of the "drunkards" (Jarvis 1972, p. 2, citing the letter of Frederick Rainer from 1876). To achieve this goal, missionaries were expected to visit prisons and police courts; speak at temperance meetings; deal with "individual drunkards" (Jarvis 1972, p. 4); and provide practical help in terms of accommodation, employment, and emigration. Being mostly sustained by their religious beliefs, the missionaries had to have faith in every single individual and treat them all as holders of the "divine spark." For instance, to become a missionary at the London Police Court Mission, one was expected to be a "man of God, a man with vocation, a man of character, a man with experience and tact and full of the milk of human kindness" (Jarvis 1972, p. 8).

Thus, the early probation officer was expected to be "a man of God" and possess the right vocation and personality animated by the desire to make good. Going back to the distinction between skills and characteristics, note that a "suitable" probation officer was defined in the literature by his moral characteristics and not by his skills. What was primarily important was what the probation officer should be like and not what exactly he or she should do in offender interactions. In most of the historical literature, the probation officer was presented in a descriptive manner with no reference to effectiveness. The ideal probation officer's characteristics were taken for granted and described as if they emanated logically from the missionary of humanistic philosophy.

The Caring Officer

After the first Probation Acts were passed in United States and England and Wales (in 1878 and 1907 respectively), the courts were asked to consider the offender's age, offence, and character before making a sentencing decision. Punishment had to fit not only the offence but also the personality of the offender. In the process of determining "true responsibility" (Saleilles 1911), probation officers were asked to contribute in separating those offenders who deserved to be saved from those considered incorrigible. This new approach of punishment claimed indirectly a certain level of specialism. Not everybody could provide the court with a reliable and robust assessment. In this context some officials and scholars started to think about probation as a profession.

Trought (1927), for example, stated that probation as a profession "involves training and a technique based on scientific principles. It requires not only personality – the principal asset – but training on and off the job" (p. 193). However, probation officers continued even after the First World War to consider in their daily practice evangelical principles and the spirit of God as well as factors such as education, employment, social structures, and so on (Vanstone 2004). The tension between science and religion was mentioned several times in the literature (see for instance Le Mesurier 1935). Dr. Selbie (cited by Trought 1927) even went so far to define the probation officer as "the great teacher and the good shepherd" (p. 194). This tension is very well emphasized in the following: "knowledge without love is almost as useless as love without knowledge…" (p. 193). Therefore both dimensions – science and religion – were required to make a good probation officer. How these two factors worked together was an important source of ambiguity.

As defined in the Probation of Offender Act 1907, Section 4(d), the duty of the probation officer is "to advise, assist, and befriend [the client], and, when necessary, to endeavor to find him suitable employment." In order to implement these duties, a probation officer had to acquire or possess certain types of characteristics and knowledge:

> Apart from any question of personal qualities, such as tact, patience or common sense, it is evident that to deal effectively with all these different

> applications definite knowledge of two kinds is required: The officer must possess a sound knowledge of the law governing the grounds on which various summonses may be granted by the court . . . And it will be equally vital for him to have a practical working acquaintance with the charitable organizations . . . (Le Mesurier 1935, p. 58).

According to the same author, while Probation Committees were still interested in considering personality and vocational call in the hiring of probation officers, their training was key. This observation is even more obvious in the job requirements in England and Wales at the time:

(a) Education: Preferably graduation from college or its equivalent, or from a school of social work
(b) Experience: At least one year in casework under supervision
(c) Good personality and character: tact, resourcefulness, and sympathy

(Probation and Supervision, Juvenile Courts at Work, cited in Trought 1927)

In the process of professionalization, the link between probation and social work played an important role. Not only were significant parts of the probation responsibilities common with social work, but also the technicality borrowed by probation service from social work enhanced the professional status of probation. Even the casework approach comes from the social work tool kit, including the idea that the process should follow the so-called ASPIRE framework (assessment, planning, intervention, review, and evaluation). At the same time, religious values such as "love," "charity," and so on were still present in the probation discourse but apparently not on the front page of job descriptions.

The Caseworker

The link between probation and social work was further strengthened after the Second World War. For many years in England and Wales, most probation officers were educated in social work. In the context of significant progress in the social sciences, new methods of working with offenders were developed. One of these new methods was social casework, which was defined as "the creation and the utilization, for the benefit of an individual with personal problems, of a relationship between that person and a trained social worker" (Home Office 1962, p. 24). The same idea was emphasized by Perlman (1957) ". . .[the] casework process, like every other process intended to stimulate growth, must use relationship as its basic means" (p. 64).

Therefore, the principal medium of intervention was the relationship between the specialist and the client. The importance of this relationship was even greater in the context of working with "unwilling clients":

> If the relationship becomes one which [the client] values and which is of help to him, he is likely to be willing to use it. If it provides an experience which is unpleasant, he will be less willing to try to use it. (Monger 1964, p. 45)

Thus, in the mid-twentieth century it was believed that a significant part of the skills and characteristics that the probation officer should possess related to the imperative of creating a positive relationship. King (1958), for instance, stressed the capacity for "a warm and sincere concern" (p. 73). Describing the abilities of a caseworker, Heywood (1964) mentions that he or she needs to have the ability to be a "person of warmth and likeability" (p. 45) and a "quality of imaginative sympathy, allied to temperament" (p. 45). There are also multiple nuances that influence the relationship: the caseworker's "demeanour, attitude to life, optimism, pessimism, courtesy, thoughtfulness, warmth, coolness, perhaps the things revealed by his moral qualities" (pp. 45, 46). Furthermore, Heywood considers that the basis of casework consists of two elements: the application to supervision of casework principles of acceptance, self determination, client participation, and confidentiality, and "the ability and mastery of his subject which the worker brings and which give hope to the client so that he goes forward confidently with the worker to try to do something about his problem" (p. 46).

Increasingly during the postwar period, probation officers were required to develop certain skills in order to diagnose and treat offenders as well as develop positive relationships with them. As McWilliams states:

> The gradual movement from the religious, missionary ideal to the scientific, diagnostic ideal, depending, in part, on notions of professionalism, required that probation work should be something for which people were trained to enter rather than called to follow (quoted in May 1991, p. 12)

It appears that after the Second World War, probation officers were expected to acquire and develop technical skills specific to a scientific approach. At the same time their personal characteristics, although present, ceased to represent an important aspect of the recruitment and selection process. For instance, qualities like reasonableness, maturity, openness, warmth, and so on, although mentioned, were not high up in the requirements list. On the contrary, candidates were expected to attend different classes or to demonstrate some skills in order to be successful.

The aim of the probation officer was quite clear: to "heal" the offender through casework and other rehabilitative interventions. The offender was perceived as ill or not fully aware of their inner thoughts and therefore in need of treatment. It comes as no surprise that this phase of probation development is also known as the "medical stage." A large number of concepts and procedures employed by probation come from the medical world via the social work profession.

The Helper

In the early 1970s, at the request of the US Government, Martinson and colleagues evaluated a significant number of parole interventions and concluded that "with few and isolated exceptions, the rehabilitative efforts that have been reported so far have had no appreciable effect on recidivism" (1974, p. 25). The arguable conclusion was therefore that "nothing works," which meant, among other consequences, the abandonment of the rehabilitation ideal in public discourse. If rehabilitation does not work, what is left for the probation service to aim for?

As an answer to this question, Bottoms and McWilliams proposed the "non-treatment paradigm" (1979). This model prescribed a new role for the probation service, taking into account the structural characteristics of the service in England and Wales (e.g., proximity to prisons, a relatively large geographical coverage) and also the research on effectiveness. According to this paradigm, the probation service is left with four "primary aims": "(1) the provision of appropriate help for offenders, (2) the statutory supervision of offenders, (3) diverting appropriate offenders from custodial sentences and (4) the reduction of crime" (p. 168). Unlike the previous model of practice, the notion of "help" is defined by the client and not by the so-called expert, the caseworker:

> The caseworker does not begin with an assumption of client-malfunctioning; rather he offers his unconditional help with client-defined tasks, this offer having certain definite and defined boundaries (Bottoms and McWilliams 1979, p. 172).

Under this new paradigm the probation officer was expected to deliver help in client-defined terms. Based on a number of parolee studies, "help" was mainly understood as practical support aimed at assisting the parolees to overcome the obstacles of reintegration.

The "non-treatment paradigm" transformed the role, and therefore the profile, of the probation officer from treatment provider to a help provider, with the main role played not by an expert (as in the previous "medical model") but by the client himself. Furthermore, the probation officer was expected to select and supervise those offenders that could be diverted from custody. As Smith (2005) puts it, "The purposes of the service were thus re-written to stress its value as a vehicle for reducing, or at least controlling, the prison population" (p. 626).

The probation officer was expected to practice a new form of casework that maximized the areas of choice for the offender within the limits of the probation order. "Within-officer" skills such as collaborative assessment, task setting, and so on were still important, but the probation officer's personal characteristics were rarely mentioned. From the literature it seems that the shift towards a more “law enforcement”-oriented practice did not call for any special staff characteristics – at least, not the ones typical for social workers. The “nothing works” era was characterized not just by client-centered help but also a focus on crime control and surveillance. Although the control element existed previously to a limited extent embedded in the concept of supervision, during this era it became as important as the care dimension, if not more important.

The New Rehabilitator

At the beginning of the 1990s, Raynor (1988) and others in England and Wales and Ross et al. (1988) in Canada started to question Martinson’s conclusions. For example, Ross and his colleagues (1988) reviewed 40 years of experimental work with offenders and noticed that persistent offenders “suffer” from a range of thinking skills deficiencies (e.g., lack of awareness of consequences, lack of rigorous planning skills, insufficient negotiation skills). On this basis they designed the Reasoning and Rehabilitation program, which aimed to reduce crime by dealing with these deficits. Most of these initial studies mentioned different qualities that an effective probation officer should have, including the ability to motivate, model reasoning and problem solving, and identify cognitive distortions; creativity; and so on (Ross and Fabiano 1985). However, the role of staff skills in effective programming was unfairly overlooked at the policy level (Raynor et al. 2010).

Several meta-analyses (e.g., Andrews and Bonta 1998; Dowden and Andrews 1999) have suggested that interventions based on the principles of risk, need, and general responsivity (RNR) are associated with important reductions in reoffending. In short, the principle of risk suggests that the intensity of a program or intervention should match the level of offender risk. The need principle states that an effective intervention should target the offender’s criminogenic needs (e.g., substance misuse, peer associations, antisocial attitudes). The responsivity principle suggests that the style and mode of interventions should match the learning style of the participating offenders. Cognitive–behavioral and social learning strategies proved to be most effective in reducing reoffending. Other meta-analysis indicated that the integrated application of these three principles could reduce reoffending by 26–30 % (Dowden and Andrews 2004). However, most of these observations focus more on the program content and less on the way they are implemented by staff. At the beginning of the What Works movement, staff skills and characteristics were neglected to a certain extent. “[P]rogramme fetishism” (Smith 2005) replaced the focus on the attributes that a good probation officer should have.

Trotter (1996) was probably one of the first modern researchers to put staff skills back on the agenda. In his experimental study of probation services undertaken in Victoria, Australia, he concluded that prosocial modeling and reinforcement by probation officers were significantly correlated with a lower reoffending and imprisonment rate; the use of problem solving and role clarification was related to reduced reoffending; and empathy together with prosocial modeling was associated with reduced reoffending. In their meta-analysis, Dowden and Andrews (2004) completed the picture of the core correctional practices (CCP) essential in reducing reoffending. In their analysis they found that five particular skills and characteristics were associated with significant reduction of reoffending: effective use of authority, anticriminal modeling and reinforcement, problem solving, use of community resources, and quality of interpersonal relationship between staff and client.

Bonta and colleagues (2010) demonstrated that probation officers trained in using risk–need–responsivity principles are more effective than others. In that study, probation officers who were trained within the Strategic Training Initiative in Community Supervision (STICS) increased adherence to the need principle and also had a more clearly defined structure of the intervention (teaching focus), had better relationship skills, and employed more cognitive techniques than the probation officers that did not attend the training. These techniques were also correlated with reoffending rates after 2 years and the conclusion was very optimistic: the difference in reconviction rate between the control group and the treatment group was 15 % (25.3 % for the treatment group and 40.5 % for the control group).

Most of these conclusions were confirmed in the Jersey study by Raynor and colleagues (2010). This research was based on analyzing 95 videotapes involving 14 probation officers with a total of 384 offenders under supervision. The skills employed by the probation staff during supervision were evaluated using a checklist with nine dimensions: setup, nonverbal communication, verbal communication, use of authority, motivational interviewing, prosocial modeling, problem solving, cognitive restructuring, and the overall interview structure. Raynor and colleagues (2010) concluded that there are significant differences between individual probation officers in terms of "structuring" skills intended to bring about change (e.g., problem solving, prosocial modeling, cognitive restructuring), which might influence offenders' progress. They also suggest that offenders supervised by probation officers who scored above the average on the skills scale had significant improvements in their LSI-R scores. Another important preliminary finding is that "use of authority" and "problem solving" are also strongly associated with improvements in LSI-R score (Raynor, personal communication). As Raynor suggests, all these preliminary findings need to be investigated in larger studies.

Another group of researchers (see McNeill 2006 for a review of the literature) reformulated the question from what makes offenders reoffend (specific to the risk/needs/responsivity paradigm) into what makes ex-offenders desist from offending? This shift of interrogation generated a whole new body of research contributing to the desistance paradigm (McNeill 2006). From a desistance perspective, change is understood as a process that a probation officer needs to promote and support. Desistance is placed at the intersection between structure, agency, and reflexivity, emphasizing at the same time the importance of human and the social capital and also the motivation and the subjective identity of the offender.

One of the great merits of this new theoretical framework is that it stresses even more the importance of staff skills and characteristics in order to create a "legitimate and respectful relationship" (McNeill 2006, p. 55) and provide help to offenders in a more collaborative approach. This observation is in line with Farrall's (2002) conclusions based on the experience of a group of 199 probationers. After interviewing the probationers in three waves, Farrall (2002) concluded that desistance could be attributed to probation interventions in only a few cases where a probation officer helped with finding a job or reconciling an offender with his family. Apart from those isolated cases, desistance took place away from the probation service, in the probationer's personal and social context where different obstacles to desistance were addressed. The importance of overcoming obstacles to achieving desistance was stated again in (Farrall 2002), casting the role of the probation officer as supporting desistance by helping individuals to overcome barriers within their own personal and social context. Instead of "managing" the offender, an effective probation officer should practice a more collaborative approach in the "co-designing [of] interventions" (McNeill 2006, p. 57).

The importance of the mutual relationship between probation officer and offender was further explained in Rex's (1999) study of 21 probation officers and 60 probationers. The study was based on semi-structured interviews that were made in 1994. From the interviews, Rex (1999) concludes that probationers who attributed change to the probation experience benefited from an active and participatory supervisory experience. In terms of professional skills that probationers would like to see in the probation officer, Rex (1999) noted that the probation officer needs to be experienced, knowledgeable, reasonable, and also to display "expert qualities" (p. 371) that would encourage probationers to disclose sensitive information and feel confident that they are listened to and taken seriously. The same study stated that in order for probationers to feel committed and positively engaged in the supervisory relationship, the probation officer needs to demonstrate empathy and have the capacity to listen and show interest and understanding, enabling probationers to talk. Although it is not known whether or not these probationers were "desisters" and there is no demonstrated causal relationship between staff skills variables and desistance, the contribution of this study is significant for understanding the "user's perspective" of an effective supervision experience.

A new body of research focuses on the question of how to assist offenders to become constructive members of society (Ward and Maruna 2007). Together with others, Tony Ward developed a new theoretical paradigm of rehabilitation – The Good Lives Model (GLM) – that looks at the offender as a potentially useful member of the family, community, and society rather than a "risk carrier." Within this positive psychology framework, based on the human dignity principle, it is crucial to treat the offender with respect in their endeavor towards a better life.

The concept of "good lives" has a few interrelated features. The most important assumption is that humans are intrinsically seeking to live a fulfilling life. In order to do that, they look for basic goods that are worthwhile in themselves, such as knowledge, health, intimate relationships, and excellence in play and work. These are defined as primary goods and are considered objective and based on human nature. The secondary goods are instrumental to securing these goods (means). Most often crime is described as an illegitimate and inefficient way to secure the primary goods. Therefore, the treatment plan should incorporate the primary goods relevant to the person and provide conditions to legally secure them. It seems that using GLM enhances treatment engagement, promotes positive working alliances, and contributes to long-term desistance (Purvis et al. 2011).

Discussion

As demonstrated above, the role of staff skills and characteristics in probation services has had a sinuous career. Staff characteristics used to be very important in the early days of probation but faded away as professionalization and crime control paradigms advanced. As effectiveness studies became more prevalent, staff skills were initially neglected but have increased in importance over time. Within the new theoretical frameworks, the focus on staff attributes finds its way back to the forefront, especially as they relate to developing effective or positive relationships with offenders. Research is replete with evidence that some skills are better than others in reducing reoffending or desistance. The danger is now that the practice may take on board only those skills and attributes that are evidence-based (demonstrated by research that they are effective). By doing that, probation practice would suffer from being too restrictive and not allowing innovation and creativity. Probably a balance between evidence-based and practice-informed approaches would promote a more nuanced practice that would have as a core element the so-called human touch (see Partridge 2004). Although research has advanced a lot in studying probation skills and characteristics, more should be done in order to validate these findings in real life situations. Sometimes the "technology transfer" (Bourgon et al. 2010) into the real world is quite problematic and results do not always confirm what was discovered in experiments. Moreover, more research should be invested into producing explanatory models of how different characteristics and skills develop in probation officers. As most of the changes have an important subjective dimension, qualitative methodologies should be employed more and more. Probably, ethnographic methods would be more appropriate to capture the individuality and also the psychological features of the change process. Realistic designs described by Pawson and Tilley (1997) represent an approach that rejects the mechanical feature of the experiment and focuses the analysis on the person(s) in a given concrete situation. Thus, researchers can obtain a richer, a more sensitive, and complex picture of the social and psychological realities.

Acknowledgement This work was supported by The Romanian Council for Scientific Research (CNCSIS) under the contract no. 29/02.08.2010.

The author is grateful to Prof. Brian Stout for his advice and comments.

Recommended Reading and References

Andrews D, Bonta J (1998) The psychology of criminal conduct. Anderson, Cincinnati

Bottoms AE, McWilliams W (1979) A non-treatment paradigm for probation practice. Br J Soc Work 9(2):159–202

Bourgon G, Bonta J, Rugge T, Gutierrez L (2010) Technology transfer: the importance of ongoing clinical supervision in translating 'what works' to everyday community supervision. In: McNeill F, Raynor P, Trotter C (eds) Offender supervision. New directions in theory, research and practice. Willan, Abingdon

Dowden C, Andrews D (1999) What works for female offenders: A meta-analytic review. J Res Crime Delinq 45:438–452

Dowden C, Andrews D (2004) The importance of staff practice in delivering effective correctional treatment: a meta-analytic review of core correctional practice. Int J Offender Ther Comp Criminol 48(2):203–214

Farrall S (2002) Rethinking what works with offenders: probation, social context and desistance from crime. Willan, Cullompton

Heywood J (1964) An introduction to teaching casework skills. Routledge & Kegan Paul, London

Home Office (1962) Report of the Departmental Committee on the Probation Service. Cmnd 1650, HMSO

Jarvis F (1972) Advise, assist and befriend. a history of the probation and after-care service. National Association of Probation Officers, London

King JFS (ed) (1958) The probation service. Butterworth, London

Le Mesurier L (1935) A handbook of probation. The National Association of Probation Officers, London

Martinson R (1974) What works? – questions and answers about prison reform. Publ Interes 35:22–54

May T (1991) Probation: politics, policy and practice. Open University Press, Buckingham

McNeill F (2006) A desistance paradigm for offender management. Criminol Crim Justice 6(1):39–62

McWilliams W (1983) The mission of the English police courts 1876–1936. Howard J Crim Justice 22:129–147

Monger M (1964) Casework in probation. Butterworths, London

Partridge S (2004) Examining case management models for sentences. Home Office Online Report. London, Home Office

Pawson R, Tilley N (1997) Realistic Evaluation. Sage, London

Perlman H (1957) Social casework. University of Chicago Press, Chicago

Purvis M, Ward T, Willis G (2011) The good lives model in practice: offence pathways and case management. Eur J Probation 3(2):4–28

Raynor P (1988) Probation as an alternative to custody. Aldershot, Avebury

Raynor P, Robinson G (2009) Rehabilitation, crime and justice. Palgrave Macmillan, Basingstoke

Raynor P, Vanstone M (1994) Probation practice, effectiveness and the non-treatment paradigm. Br J Soc Work 24(4):387–404

Raynor P, Ugwudike P, Vanstone M (2010) Skills and strategies in probation supervision: the Jersey study. In: McNeill F, Raynor P, Trotter C (eds) Offender supervision. New directions in theory, research and practice. Willan, Abingdon

Rex S (1999) Desistance from offending: experiences of probation. Howard J Crim Justice 36(4):366–383

Ross RR, Fabiano EA (1985) Time to think: a cognitive model of delinquency prevention and off ender rehabilitation. Institute of Social Sciences and Arts, Johnson City

Ross RR, Fabiano EA, Ewles CD (1988) Reasoning and rehabilitation. Int J Offender Ther Comp Criminol 20:165–173

Saleilles R (1911) The individualization of punishment. Translated from french by Rachel Szold Jastrow. William Heineman, London

Smith D (2005) Probation and social work. Br J Soc Work 35(5):621–637

Trotter C (1996) The impact of different supervision practices in community corrections. Australian and New Zealand Journal of Criminology 29(1):29–46

Trotter C (2006) Working with involuntary clients. Allen & Unwin, Crows Nest

Trotter C, Evans P (2010) Supervision skills in juvenile justice. In: McNeill F, Raynor P, Trotter C (eds) Offender supervision. New directions in theory, research and practice. Willan, Abingdon

Trought T (1927) Probation in Europe. Basil Blackwell, Oxford

Vanstone M (2004) Supervising offenders in the community. A history of probation theory and practice. Ashgate, Aldershot

Vanstone M (2008) The international origins and initial development of probation. An early example of policy transfer. Br J Criminol 48:735–755

Ward T, Maruna S (2007) Rehabilitation: beyond the risk paradigm. Routledge, London

History of Substance Abuse Treatment

Alana Henninger[1] and Hung-En Sung[2]
[1]John Jay College of Criminal Justice, New York, NY, USA
[2]Department of Criminal Justice, John Jay College of Criminal Justice, New York, NY, USA

Synonyms

Evolution of substance abuse treatment

Overview

Efforts at combating the negative health and social consequences of substance abuse and dependence have always existed in the United States. Often swinging between the rival contexts of moralistic and positivistic discourses, these efforts have led to the articulation of the major therapeutic paradigms in the field of substance abuse treatment. The earliest interventions were grassroot interventions focusing on individuals with drinking problems whose goals shifted from moderation to abstinence over time. As the patterns of substance use and abuse quickly diversified along the processes of immigration and urbanization, a wider variety of substances and a more diverse assortment of users became targeted for an even richer array of therapeutic experiments. The gradual involvement of the state in the planning and administration of substance abuse treatment has resulted in the growing use of institutionalization and coercion to trigger and maintain the recovery process. The emerging consensus that substance addiction is a chronic and relapsing brain disease represents a redefinition of an old problem and will determine the direction of the science and art of substance abuse treatment in the years to come.

Introduction

Mind-altering substances have been a part of the human experience since the beginning of civilization (Katcher 1993; Saah 2005). Many of the psychotropic effects enjoyed by recreational drug users today were sought by ancient civilizations as a means of increasing their chances of survival (Saah 2005). Ancient civilizations also realized, however, that it was possible to enjoy and become addicted to certain substances. Five-thousand-year-old Egyptian records report that individuals suffering from alcohol addiction were often cared for in the private homes of people who provided treatment. Records from ancient Greek and Roman sources recommended that alcoholics receive treatment in “public or private asylums” (White 1998). This chapter will provide an overview of American substance abuse treatment from the 1700s to the present.

The Early Years of America: 1600 to 1900

Americans have consumed alcohol since colonial times, making alcohol the earliest use of an intoxicant by the settlers. Alcohol was an integral part of culture at the time, but public drunkenness was not tolerated (Levine 1978; Stolberg 2006). The production and consumption of alcohol increased steadily between colonization and the late eighteenth century. By 1810, the number of distilleries in the United States had increased exponentially to a staggering 14,191, with the average person drinking four and a half gallons of alcohol per year (White 1998). The level of alcohol consumption had reached its zenith by 1830, with the average American consuming approximately five to seven gallons of alcohol per year (Stolberg 2006). Alcohol had become easily accessible to, and affordable for, the general population, resulting in deleterious effects on society (especially the family) due to increased alcohol abuse (Katcher 1993; White 1998; Stolberg 2006). Treatment of addiction in the United States, therefore, began with alcohol addiction (Lemanski 2001).

Early Institutional Care for Addiction

In the late eighteenth and early nineteenth centuries, individuals suffering from an addiction to alcohol were grudgingly housed in a variety of locations. Individuals who were considered drunkards were frequently incarcerated in local jails (Rosenberg 1995; White 1998; Rothman 2001), which is especially interesting because alcohol was often served in early American jails (Rothman 2001). Hospitals were an option for medical care but were limited in number during the eighteenth and nineteenth centuries. Both almshouses (places for the poor, sick, or destitute to seek refuge) and asylums for the mentally ill also provided a place for alcoholics to stay but, like jails and hospitals, failed to provide treatment for addiction (Rosenberg 1995).

Benjamin Rush, a prominent physician and activist beginning in the latter half of the eighteenth century, proposed that "sober houses" be used to treat alcohol addiction through medical treatment as well as "religious and moral instruction" (Levine 1978; Baumohl 1990; White 1998). Rush, who initially advocated for moderation rather than abstinence, went on to describe the symptoms and potential social consequences of alcohol abuse in pamphlets that were distributed to the public in 1784 (Levine 1978; Katcher 1993; Lemanski 2001; Stolberg 2006). Early treatments of alcoholism used by Rush ranged from cold baths, vomiting, and aversion therapy to the practices of bleeding, blistering, and sweating the patient (White 1998).

Physicians in the early nineteenth century made significant contributions to the identification of key physical consequences of alcoholism, such as liver, stomach, muscle tissue, and nerve tissue damage. Delirium tremens (or the "D.T.s") were first described prior to 1819 as "alcohol-induced brain fever," and then later articulated by Dr. Walter Channing and Dr. John Ware between 1819 and 1831. The two physicians provided the first American documentation of a variety of symptoms of delirium tremens, such as hand tremors, hallucinations, vomiting, and loss of appetite. The term "alcoholism" was also coined in the mid-1800s by Dr. Magnus Huss, a Swedish physician who chronicled the development in global knowledge of alcohol addiction (White 1998).

The discovery of the physical side effects and consequences of chronic alcohol use, as well as the ineffective drive for abstinence during the temperance movement, played a role in the push for the institutional treatment of alcoholism (Baumohl 1990; White 1998). The American Association for the Cure of Inebriation (AACI) provided the first attempt at the professionalization of treatment services, arguing that addiction was a disease that should be treated rather than a vice or criminal offense that should be punished. In 1870, six managers of inebriate homes and inebriate asylums joined together to form the AACI. Inebriate homes were small facilities that provided some treatment for alcohol addiction but functioned primarily as a shelter for alcoholics. Inebriate asylums, on the other hand, were often large private or state-sponsored facilities that focused on the provision of medical treatment for alcohol addiction. Private sanitaria (often referred to as lodges or retreats) were also developed at the end of the nineteenth century to provide addiction treatment services to the wealthy (White 1998).

Types of services provided at the different types of addiction facilities included inpatient treatment, short- or long-term stays for detoxification, and the first attempts at outpatient services. The first attempts at a "continuum of care" were also made during the inebriate asylum movement. A local physician would first refer a patient suffering from alcohol addiction to an institution for the purposes of detoxification. After further evaluation, the patient would be transferred to a facility that provided long-term treatment for addiction. After the completion of long-term treatment, the patient would be returned to the care of their local physician, who would then provide any necessary aftercare (White 1998).

Specific treatment methods within individual institutions varied. The most important feature of early institutional treatment was isolating the inebriate from society so that the inebriate would no longer face the temptation of alcohol. Once isolated, many institutions began the process of detoxification. Detoxification could be done abruptly and without the aid medication (often referred to as “cold turkey”) or gradually with medications prescribed to alleviate the symptoms of withdrawal. Other types of psychotropic substances, such as cannabis, coca, chloral hydrate, and belladonna, were commonly prescribed for detoxification during the nineteenth century. Detoxification was often followed by an attempt to restore the patient to a healthy physical state, through the provision of medical treatment, healthy meals, fluids, vitamins, and exercise. It was also believed that massages, sunlight, and electrotherapy would aid in the recovery process. Music, counseling, aversion therapy, and religious instruction were other methods that were occasionally offered, but were not the norm (White 1998).

Construction of institutions for the treatment of alcohol addiction in the United States expanded to a total of one hundred facilities by 1902. This trend of expansion, however, was short-lived. The majority of the inebriate homes and asylums had either disappeared or been reallocated for correctional or psychiatric purposes by the 1920s. Contributing factors to the demise of the inebriate homes and asylums were lack of funding, technology, and professionalization as well as a waning of the optimistic belief that alcoholism could be cured. Although early attempts at the treatment of alcohol addiction were not successfully sustained, they conceptualized addiction as a disease rather than a personality disorder, revolutionized treatment of addiction, and provided the foundation for the continued development of future methods of substance abuse treatment (White 1998).

The Temperance Movement

The temperance movement played a pivotal role in the evolution of substance abuse treatment by sparking a new way of viewing addiction and recovery. The temperance movement began in 1808 as a response to the social problems resulting from the rapid increase in alcohol abuse among the American population (White 1998; Lemanski 2001). Alcohol abuse was initially thought to be a moral affliction rather than a disease. The original goal of the temperance movement, therefore, was to reduce alcohol consumption. Early attempts at reducing alcohol intake usually consisted of replacing spirits with beer or wine (Hall 2010; Stolberg 2006).

Beginning in 1831, a variety of temperance societies were created throughout the country in an effort to promote a communal recovery process rather than an isolated attempt at sobriety. The Washingtonian Society, founded in 1840, was one of the more well known of these temperance societies. The purpose of the Washingtonian Society was to provide a place for alcoholics to meet, hear temperance lectures, and share their experiences of alcoholism and recovery. Their method of recovery required members to publicly acknowledge their addiction and commit to recovery, share their experiences with other members, and engage in leisure activities without alcohol (White 1998). The Washingtonians also attempted to assist their fellow members in times of need, which is an aspect of the society that made it different from other temperance societies operating at the time. Efforts were made, through charitable donations, to offer legal assistance, shelter, food, and clothing to fellow Washingtonians as needed (Lemanski 2001).

The rapid growth of the Washingtonian Society, however, was met with an equally rapid demise. The Washingtonian Society had failed to establish a cohesive ideology, as well as an organized standard of practices for meetings, necessary to support continued expansion of the movement (Lemanski 2001). Perhaps more importantly, however, pledging abstinence was not a sufficient method of treatment for individuals suffering from alcoholism. By 1840, the goal of the temperance movement had shifted from moderation to abstinence (Katcher 1993). This shift in goals was largely a result of the growing recognition that alcoholism could not be cured through simple will power. Advocates in the temperance movement began to argue that the best way to fight alcoholism was to prevent future generations from drinking alcohol by prohibiting the consumption of alcohol (Lemanski 2001; Stolberg 2006).

The temperance movement continued to flourish until the Civil War began in 1861. Although the temperance movement stalled during the Civil War, the movement regained its momentum as a political endeavor after the war ended (Lemanski 2001; Stolberg 2006). Women's groups, such as the Women's Christian Temperance Union, began to push harder not only for temperance but also for prohibition. In 1869, the Prohibition Party was formed in an effort to make the distribution and transportation of alcohol illegal in the United States (Lemanski 2001). This policy initiative came to fruition with the adoption of the Eighteenth Amendment to the Constitution in 1920. As will be seen in the next section, prohibition had a powerful impact on substance use, abuse, and treatment among the American population in the early twentieth century (Katcher 1993; Lemanski 2001; Hall 2010).

Early Drug Treatment

While addiction treatment services focused predominantly on alcohol in the nineteenth century, there was a burgeoning interest in the treatment of other drugs (White 1998; Stolberg 2006). It is important to note that the use of psychotropic drugs was legal in the United States until the enactment of the Harrison Act in 1914. The availability of drugs (especially opiates and cocaine) greatly increased during this time period, resulting in the first drug epidemic. Physicians with experience in treating patients with drug addictions (usually to opiates) began discussing addiction in terms of disease rather than moral failure (White 1998).

More women than men used psychotropic drugs during this time period, most likely because opiates were the popular treatment for "female" maladies such as menstruation and hysteria. Opiate addicts in the nineteenth century were, more often than not, educated white women of a higher socioeconomic status. Due to the stigmatization of substance abusers, especially during the female-driven temperance movement, women often hid their substance use and abuse from their family and friends. When women did seek out treatment, the reported ailment was usually of a physical or psychological nature (White 1998).

Ironically, however, early treatments for physical, psychological, and substance abuse disorders often included the prescription of other drugs (Courtwright 2001). Opiates were often prescribed for physical and psychological maladies, while cocaine was often prescribed for alcohol and opiate addiction. Additionally, cocaine was frequently used as an anesthetic during surgical procedures. Unfortunately, the treatment of narcotics prior to the twentieth century predominantly focused on easing the suffering of withdrawal symptoms during the detoxification process through the prescription of other highly addictive psychoactive substances (e.g., codeine, hypnotics, bromides, and chloral hydrate). Early experimentation with maintenance programs began at this time, but the programs were largely ineffective because the narcotics used to maintain the patient continued to provide a high (unlike agonists that are used today, such as methadone). The cycle of addiction, therefore, was often sustained (White 1998).

The Miracle Cure

During the second half of the nineteenth century, several coinciding circumstances provided the perfect milieu for the sale of "miracle cures" for addiction. Miracle cures were usually in the form of a medication that could be taken for a short period of time for the successful treatment of alcohol (e.g., Mickey Finn Powders, White Star Secret Liquor Cure, and the Hay-Litchfield Antidote), hangovers (e.g., Alka-Nox, Wink, and Sober-Up), tobacco (e.g., Nicotol, Tobacco Redeemer, and Gustafson's Tobacco Remedy), or other psychotropic substances (e.g., Mrs. Baldwin's Home Cure for Cocaine, St. Anne's Morphine Cure, and Weatherby's Opium Antidote). A variety of tonics and syrups consisting of various mixtures of alcohol, opium, morphine, and cocaine promised to cure any addiction in the privacy of one's own home at a fraction of the cost of institutional treatment (White 1998).

Miracle cures could be purchased over the counter, through mail order services, and were also provided in treatment facilities (e.g., inebriate houses). Aggressive marketing strategies were used to advertise miracle cures to the public through the use of newspapers, magazines, billboards, and speaking events. Gimmicks, such as the use of the testimonial (e.g., 100,000 satisfied customers), the money-back guarantee, and discreet packaging, became very popular during this time period (White 1998). Some of these miracle cures, such as Leslie Keeley's Double Chloride of Gold Cure, actually became treatment franchises (Baumohl 1990; White 1998). Doctors and charlatans alike frequently attached their names to the production and promotion of a new, top-secret formula (e.g., Dr. Meeker's Antidote) that would purportedly revolutionize the treatment of addiction. The sale of miracle cures continued well into the mid-twentieth century (White 1998).

1900 to the 1950s

The goal of prohibition was to make the distribution and transportation of alcohol illegal so that the consumption of alcohol would become impossible. If the consumption of alcohol could be prevented, it was reasoned, then alcohol addiction would cease to be a social problem necessitating treatment (Hall 2010). Unfortunately, alcohol abuse rose while treatment options disappeared. The idea that alcohol was a physical disease necessitating medical treatment was once again replaced with the belief that alcohol was a lapse in moral character that could be cured by prohibiting inebriety. As a result, religious groups such as the Salvation Army continued to provide treatment services to alcoholics, but the institutional treatment of addiction completely disappeared until the late 1940s (White 1998; Lemanski 2001).

The Effect of Criminalization of Drugs on Addiction Treatment

Although it is widely believed that the "war on drugs" began with President Nixon in the 1970s, it could be argued that the "war" began with the passage of the Harrison Act in 1914. Prior to the Harrison Act, access to opiates, cocaine, and other drugs was unregulated. This lack of regulation, combined with the overprescription of both drugs, resulted in high rates of addiction. In 1912, the first morphine maintenance clinics opened in an attempt to treat the many men and women who had become addicted to opiates. Outpatient treatment was provided through the prescription of decreasing amounts of morphine for the purpose of weaning patients off of drugs at a pace that would allow them to avoid withdrawal symptoms (White 1998).

When the Harrison Act was passed, the possession of illicit substances was criminalized. It was at this point in history that the treatment of substance use and abuse shifted from a public health model to a criminal justice model. The passage of the Harrison Act limited the access to opiates and cocaine to physicians, who were then prohibited from prescribing the drugs to addicts for the purposes of detoxification or maintenance because the government believed that too many addicts were remaining on maintenance programs indefinitely rather than being weaned off of drugs (Levine 1978; White 1998; Courtwright 2001). By 1925, all of the morphine maintenance clinics had been closed (White 1998).

Thus, although the Harrison Act was intended to address opiate and cocaine addiction, it made it very difficult for addicts to receive any treatment whatsoever. Shortly after drugs were criminalized, the mandatory detention of substance users and abusers began. Many individuals suffering from addiction were sent to psychiatric hospitals or special "colonies" for mandatory treatment. Jails and federal penitentiaries were also flooded with addicts who had violated the Harrison Act. Between 1915 and 1929, the number of drug-related incarcerations increased from 63 to 1,889. Overcrowding of correctional facilities led to the creation of two federally funded and managed "narcotics farms," which served as separate facilities for drug offenders in need of long-term treatment (White 1998).

Drug abuse and dependence treatments during the first half of the twentieth century were often experimental in nature. For example, treatments for alcoholism were often applied to drug users. Other extremely invasive methods were also used, such as insulin-induced comas, electroconvulsive therapy (ETC), aversion therapy, psychosurgery (i.e., lobotomy), and serum therapy (where serum was withdrawn from blisters that were raised on the abdomen of a patient and then reinjected). Psychological approaches for the treatment of alcohol and drug addiction were also developed during the early nineteenth century, with psychoanalytic approaches being the predominate format. Although the prohibition of alcohol was eventually deemed a failed social experiment, the criminalization of other psychoactive substances remains to this day (White 1998; Lemanski 2001).

The Modern Alcoholism Movement

After the enactment of the Twenty-first Amendment to the US Constitution in 1933, prohibition ended and a new phase in the treatment of alcohol addiction began (White 1998; Lemanski 2001). A third paradigm shift in the treatment of addiction occurred between 1933 and 1955. This period is often referred to as the "Modern Alcoholism Movement" (p. 178). During this time, alcoholism was once again reconceptualized as a disease and public health issue rather than a moral weakness. Professionals and members of society alike began to believe that alcoholism could be treated. This new belief in the merits of treatment opened the door for the allocation of funds for new research on addiction to be conducted by professionals in the medical, psychological, and social work fields (White 1998).

The Modern Alcoholism Movement was not, however, a unified process. The activities of a variety of institutions and interest groups, often with very different agendas, worked separately to achieve social change. Three of the most prominent professional groups to exact change during this period were the Research Council on Problems of Alcohol, the Yale Center of Alcohol Studies, and the National Committee for Education on Alcoholism. These three groups collectively succeeded in redefining alcoholism; changing policies and treatment practices relating to addiction; convincing government and private sources to invest money in research, education, and intervention; and initiating alcohol treatment in the workplace. Another interest group, Alcoholics Anonymous, used a grassroots approach to become an especially integral part of the Modern Alcoholism Movement (White 1998; Lemanski 2001).

Alcoholics Anonymous

Treatment options for alcoholics had all but disappeared in the United States by 1930. In 1935, Bill Wilson and Dr. Robert Smith developed a self-help recovery group called Alcoholics Anonymous (A.A.) in an effort to fill this treatment void (Hubbard et al. 2009). A.A. evolved out of a nondenominational group, known as the Oxford Group, who sought to encourage positive societal change through individual spirituality. Although alcoholism was not the primary concern of the Oxford Group, many group members believed that religion could provide a strong foundation for individuals seeking sobriety. A small subgroup of alcoholic members, led by Wilson and Smith, formed within the Oxford Group to provide support to one another during their struggle for sobriety. The nonalcoholic members of the Oxford Group disagreed with this focus on sobriety, which led to the separation of this subgroup in 1937 (White 1998; Lemanski 2001; Alcoholics Anonymous 2012). The group was christened Alcoholics Anonymous in 1939 (White 1998; Alcoholics Anonymous 2012).

While the early years of A.A. consisted of relatively unstructured group meetings, Wilson and Smith soon began to focus on building a financially independent group with a guiding ideological framework. In 1938, Wilson began writing a book (called *Alcoholics Anonymous*) that outlined how he wanted the program to function and recorded the stories of the recovery of many of the group members. What began as six guiding principles eventually turned into the twelve-step program that A.A. is known for today. The twelve-step program provides an outline for the course of recovery that an alcoholic will follow as a member of A.A. (e.g., the member must admit that they have a problem and that they need help, put their faith in God, make amends, acknowledge that sobriety is a lifelong commitment). In this way, the focus was not on how or why the individual became an alcoholic. The focus has always been on how the individual will achieve sobriety, and the identity reconstruction that must occur for the individual to remain sober (Alcoholics Anonymous 2012).

After the publication of their book in 1939, A.A. gained a great deal of press coverage, which resulted in enormous popularity. Membership quickly grew to 100 members within that first year and then skyrocketed to 8,000 members by the end 1941 (White 1998; Lemanski 2001; Alcoholics Anonymous 2012). A.A. then began to play a pivotal role in the medical treatment of alcoholism. Hospitals had always grudgingly treated (or refused to treat) alcoholics because it was believed that alcoholism was the result of a moral weakness rather than a disease. Alcoholics also tended to be very difficult and frustrating patients, in already overcrowded facilities, who did not pay their bills. Members of A.A. made it their mission to change the perception of alcoholism from an untreatable moral weakness to a disease with the potential for successful treatment, and they succeeded (White 1998).

Through tireless advocacy, A.A. (1) convinced local hospitals that alcoholism could be treated, (2) showed hospitals how to treat alcoholism, (3) suggested that separate wards be established for the sole treatment of alcoholism, (4) suggested that these wards be managed by A.A. members to take the stress off of hospital employees, and (5) promised to pay the bills of any patient that A.A. had sponsored for alcohol treatment. A.A. then began to extend their treatment model to private, psychiatric hospitals, and prisons, thus playing a major role in the return to the disease model of addiction treatment (White 1998).

In 2011, it was estimated that there are 2,057,672 members worldwide (Alcoholics Anonymous 2012). A.A. has evolved into “the most fully developed culture of recovery that has ever existed- a culture with its own history, mythology, values, language, rituals, symbols, and literature” (White 1998, p. 162). The founders of A.A. chose to use the personal experiences of group members who have experienced alcoholism to provide support and guidance in the recovery process rather than utilize the services of professionals in addiction treatment. Nevertheless, Alcoholics Anonymous revolutionized addiction treatment in the first half of the twentieth century and has continued to influence social policy well into the twenty-first century (White 1998).

1950s to the Present

The Minnesota Model

The Minnesota Model of addiction treatment began with the establishment of a new method of alcoholism treatment in three Minnesota centers in the late 1940s: Willmar State Hospital, Pioneer House, and Hazelden. Willmar State Hospital functioned as an inebriate hospital beginning in 1912 but was converted into a psychiatric facility as a result of prohibition. Pioneer House opened its doors in 1948 in an effort to offer a different type of treatment option. Using A.A. as a program model, Pioneer House provided residential treatment for men with alcohol addiction for two to four weeks (White 1998). Hazelden Farm, a former retreat for alcoholic men, was then established a year later to offer A.A.-based residential treatment program (Hazelden 2012).

What is known today as the “Minnesota Model” evolved throughout the 1950s. One of the primary tenets of the Minnesota Model that emerged during this period was respect. The Willmar team argued that patients would greatly benefit from a recovery process built on mutual respect rather than degradation, which was a radical idea during this period. The staff at Willmar then defined alcoholism “not as a symptom of underlying emotional problems, but as a primary, progressive disease” (White 1998, p. 203) that should be treated using a multidisciplinary and holistic approach. They also believed that professionalization of addiction treatment was an integral part of this approach. Medical treatment was to be provided by doctors and nurses, counseling was to be provided by psychologists and social workers, and spiritual guidance was to be provided by the clergy. Perhaps more importantly, however, recovered alcoholics were also professionalized through training and the provision of credentials.

The Willmar team believed that recovered alcoholics were often the best counselors, but their position as a part of the addiction treatment team needed to be legitimized. The radical idea of professionalizing recovered alcoholics was met with a great deal of resistance but has become routine in current treatment models (White 1998).

The Minnesota Model continued to flourish and solidify during the 1960s at both Willmar and Hazelden, with several core elements emerging during this period (White 1998; Hazelden 2012). A 28-day stay became the standard treatment period, and patients were first prescribed medication to aid in the detoxification process (Lemanski 2001). A counselor, who often times was a recovering alcoholic, was assigned to each patient. Patients and counselors were usually matched according to age and gender in an effort to establish a supportive environment in which the patient would feel comfortable engaging in self-disclosure (White 1998). Individual counseling was supplemented by group counseling (which combined support and confrontation), lectures, and working through the twelve steps of A. A. Patients also received aftercare upon their return to the community (Quinn et al. 2004).

The Minnesota Model was immensely popular and had been adopted by 114 state agencies that provided addiction services by 1971. Presentation of the Minnesota Model at medical conferences took the treatment approach nationwide. Individuals suffering from alcohol addiction, as well as practitioners who wanted to be taught the treatment method, began pouring in from around the country. This innovative approach of treating chemical dependency is still followed today for both alcohol and drug addiction (White 1998; Lemanski 2001; Quinn et al. 2004).

The Resurgence of Drug Treatment

Treatment models for alcohol addiction and drug addiction evolved in vastly different ways due to the early criminalization of drug use. While treatment options for alcoholics continued to evolve, drug addicts began seeking treatment in psychiatric hospitals after the morphine maintenance clinics closed in the 1920s because there was literally nowhere else to go. Rapid withdrawal, and perhaps some psychiatric treatment, was the only help that drug addicts received in these facilities. A slight resurgence in drug treatment then began to occur in New York City in the 1950s, however, partly as a response to surging rates of adolescent heroin addiction. A few twenty-eight-day residential treatment programs were established inside of hospitals. Upon admission, patients were detoxified and then participated in individual and group therapy. After discharge, outpatient aftercare services were provided. Hospitals were still resistant to treating individuals suffering from drug addiction, so many individuals were committed for twenty-eight-day cold turkey detoxification on Riker's Island Penitentiary. Correctional detoxification continued until the practice was banned in the early 1960s. Treatments options were bleak, but both drug legislation and methods of treatment were about to be revolutionized in the 1960s (White 1998).

Narcotics Anonymous

In an effort to fill the treatment void, attempts to use the philosophies of Alcoholics Anonymous began in various locations in the late 1940s and early 1950s. The leaders of A.A. did not want to extend membership to drug addicts, so recovering addicts created their own version of A.A. The first of these meetings, which were called "Addicts Anonymous," took place in 1947 inside the federal narcotics farm located in Lexington, Kentucky. Addicts Anonymous meetings soon spread to other treatment facilities and hospitals, but when the organization established community-based meetings, the name was changed to "Narcotics Anonymous (N.A.)" to avoid confusion with Alcoholics Anonymous. The two programs are almost identical, with only slight differences in wording. Although N.A. got off to a very shaky start, almost disappearing in the late 1950s, it is now a thriving international organization (White 1998).

Community-Based Treatment

The first attempts at community-based support services for recovering addicts returning home after institutional care began in the late 1950s (De Leon 2000). Cities across the country attempted to provide outpatient treatment via social workers to individuals who had recently completed inpatient detoxification. These early attempts were largely unsuccessful, however, because follow-up appointments were often missed by patients. Continuing care for drug-addicted patients predominantly fell on private physicians, who often treated patients who had overdosed or contracted a sexually transmitted disease. Religious organizations continued to provide outreach services for drug addiction that often included medical and psychological treatment. N.A. also attempted to fill the void in community treatment, but it was not until Synanon, therapeutic communities, and new legislation permitting the use of agonist and antagonist treatment in community facilities that community-based treatment really took off (White 1998).

Therapeutic Communities

In mid-1958, a member of Alcoholics Anonymous named Charles Dederich began holding meetings at his home in California during which he experimented with different methods of group recovery for individuals addicted to both alcohol and drugs. These methods of group recovery were often confrontational, bordering on verbally abusive, and were intended to break down the defense mechanisms and excuses of addicts. His experimentation led to the idea for a therapeutic commune where drug addicts could live together during the recovery process. Dederich left A.A. to pursue his new treatment modality (which was essentially a residential treatment program), which he called Synanon (National Institute on Drug Abuse 2008; Janzen 2011). This first phase of development, known as Synanon I, became immensely popular with both the public and practitioners. By 1969, however, Dederich decided to take Synanon I in a new direction. While the goal of Synanon I was to rehabilitate addicts, the goal of Synanon II was to form an alternative lifestyle and to promote societal change. Synanon II evolved into Synanon III in 1974, at which time Synanon was declared a religion (Janzen 2011).

Synanon may have come to a decidedly bizarre end, but it did initiate a movement towards other therapeutic communities. Five hundred therapeutic communities using the early philosophies of Synanon were operating in the United States by 1975 (De Leon 2000). These new therapeutic communities focused on the rehabilitation and sustained recovery by adjusting the lifestyle, relationships, and even the personality of individuals suffering from drug addiction. Aggressive, confrontational, and even humiliating methods of group recovery were often used within a hierarchical system used to motivate patients to work for higher-status positions within the community (De Leon 2000; Inciardi and McElrath 2008; National Institute on Drug Abuse 2008). Most early therapeutic communities progressed their patients through three phases over a period of 18 months to 3 years. Patients first received isolated treatment within the therapeutic community, were then employed outside of the therapeutic community, and finally engaged in activities at the therapeutic community while living and working on their own (White 1998).

Therapeutic communities continued to evolve into the model that is used today. Residential stays are similar but tend to be shorter, with treatment methods that are not humiliating or as confrontational and aggressive. Today, therapeutic communities are also often linked to the criminal justice system and outpatient centers. Many jails and penitentiaries have therapeutic communities operating within their walls (De Leon 2000; National Institute on Drug Abuse 2008). The continued acceptance of therapeutic communities and other evolving treatment options, however, were contingent upon the wealth of new legislation that was occurring simultaneously throughout the 1960s.

Civil Commitments and Other Legislations

By the late 1950s, individuals who were suffering from drug addiction were viewed as people who were immersed in a destructive lifestyle that they were not willing, or able, to change. It was believed that addiction was a social disease that could be transmitted to other people; therefore, the family and friends of drug-addicted individuals were considered to be in harms way. Civil commitment, the court-ordered institutionalization of a mentally ill individual, provided a solution to these problems at the state level (Inciardi et al. 1996).

Involuntary and voluntary civil commitment to psychiatric institutions was extended to drug addicts in the early 1960s as a way of ensuring that substance abusers were treated rather than incarcerated. It was thought that addicts would be forced into the treatment that they may not be ready to seek for themselves while also preventing the spread of addiction to the community (Maddux 1988; White 1998). Patients were typically committed for 1 to 3 years, during which time they were detoxified and then provided with medical treatment and counseling. After discharge, the patient was returned to their community and received outpatient treatment (White 1998).

Civil commitment was not particularly successful at initiating sustained abstinence. With psychiatric facilities overburdened and patients continuing to relapse, the government became motivated to find different methods of treating patients in noninstitutional settings. The Joint Committee of the American Bar Association and the American Medical Association on Narcotic Drugs released a report in 1961 that recommended treating patients in community clinics. The more radical recommendation of this report, however, was the suggestion that investigations into maintenance programs be resumed. This step towards a disease model of addiction was further supported by several Supreme Court decisions in the early 1960s that prohibited laws that made addiction a crime, framed addiction as a disease deserving treatment rather than punitive action, and cautioned that civil commitment should be used sparingly. This paradigm shift made it possible for researchers claiming that immediate and sustained abstinence is not a realistic outcome for all drug addicts to be taken seriously (White 1998).

Pharmacotherapy

Using narcotics to medically maintain individuals suffering from opiate addiction was prohibited in the United States by the Harrison Act of 1914, but research concerning the practice continued to evolve. German researchers began developing a synthetic narcotic (White 1998; Joseph et al. 2000; Inciardi and McElrath 2008), called Dolophine, in the late 1930s. The use of Dolophine was discovered by the United States in the 1940s and was then marketed under several names by Eli Lilly, Inc. beginning in 1947 (White 1998). What is now known in the United States as methadone is an agonist that mimics the effects of other opiates but without the associated high. Methadone, therefore, can be prescribed in decreasing doses to individuals suffering from opiate addiction as way of weaning them off of the opiate without suffering withdrawal symptoms (Inciardi and McElrath 2008; U.S. National Library of Medicine 2009).

Methadone was used successfully for detoxification in a few US Public Health Hospitals in the late 1940s, but the use of narcotics for maintenance was prohibited until the addiction legislation of the 1960s called for research of new treatment methods (Courtwright 2001; Hubbard et al. 2009). In 1963, Dr. Marie Nyswander and Dr. Vincent Dole argued that methadone could be used not only to detoxify but also to maintain heroin addicts while they acclimate to the metabolic changes that occur during the recovery process. They argued that stabilizing the patient while their body underwent drastic physical change could prevent relapse (White 1998; Joseph et al. 2000; Courtwright 2001; Dole and Nyswander 2008).

Early experimentation occurred in intense inpatient settings. Patients received methadone along with counseling, while dosage levels were explored. Experimentation with dosage levels finally paid off in 1964 when Nysander and Dole discovered that one high dosage of methadone prevented intoxication, craving, euphoria, dysphoria, and withdrawal for 24 hours or more (White 1998; Joseph et al. 2000). Patients were able to function normally with very few side effects. Methadone became the wonder drug, and outpatient clinics were constructed all over the country. During the 1970s, federally funded methadone programs were established at record rates as a response to the return of heroin-addicted Vietnam veterans and a surge in urban crime. Today, methadone is provided in outpatient clinics, usually in conjunction with other addiction services (White 1998; Joseph et al. 2000; Courtwright 2001).

Experimentation with other drugs with the potential for maintaining opiate-addicted patients during recovery also occurred during the 1960s. An agonist called LAAM was another viable option discovered for maintaining patients with opiate addiction programs, but methadone remained the more popular of the two options. The effects of another group of medications, known as antagonists, were also explored. Antagonists, (e.g., nalorphine, naltrexone, and naloxone), block rather than mimic the effects of opiates (White 1998; Nutt 2010). These drugs work just as well as agonists (like methadone), but have no street value because they are not effective pain medications, which is a significant advantage. Antagonists are not popular among patients, however, because they do not suppress cravings and they induce withdrawal if taken when a patient has opiates in their system. Researchers have also been experimenting with partial agonists (like buprenorphine) and the combination of agonists with antagonists (like Suboxone®) as a way getting the best of both worlds (Nutt 2010).

Criminal Justice-Mandated Treatment

Prisons have always been major substance abuse treatment facilities, and criminal justice referrals have historically been the largest source of the publicly funded drug treatment admissions in local communities, accounting for 40–50 % of referrals to community-based treatment programs (Farabee et al. 2004). These community referrals and those admissions to prison- or jail-based treatment programs are generally known as coerced treatment. Coercive treatment approaches for drug addiction have been applied consistently throughout the twentieth century, beginning with the morphine maintenance clinics in some communities during the 1920s. The 1930s marked the establishment of the federal narcotics treatment facilities in Fort Worth, Texas, and Lexington, Kentucky. During the 1960s broad-based civil commitment, procedures were implemented in the federal system, as well as in New York and California. The present system, beginning in the 1970s, relies less on formal civil commitment procedures and emphasizes community-based treatment as an alternative to incarceration or as a condition of probation or parole. The inception of the drug court model in 1989 and of the prosecutor-led drug treatment alternative to prison (DTAP) model effectively propel the use of the threat of incarceration as the leverage and the application of graduated sanctions as the calibrator in the recovery process (Sung and Belenko 2006).

Some earlier researchers (Hartjen et al. 1982; Platt et al. 1988) have argued that little benefit can be derived when a drug user is forced into treatment by the criminal justice system because treatment can be effective only if the client is truly motivated and ready to change. The allocation of limited treatment slots to drug abusers who do not really want to be treated is perceived as ethically unjust and clinically unwise. Other researchers, on the other hand, posit that few chronic drug users will enroll and stay in treatment without some external pressure and that legal coercion is as justifiable as any other motivation for treatment entry (Anglin and Maugh 1992; Salmon and Salmon 1983). As a matter of fact, most clients begin their treatment under some type of pressure from their families or employers.

Legal coercion represents a range of options of varying degrees of severity across the various stages of criminal justice processing. It can be used to refer to such actions as a probation officer's recommendation to enter treatment, a prosecutor's offer of a choice between treatment and jail, a judge's requirement that the offender enter treatment as a condition of probation, or a correctional policy of sending inmates involuntarily to a prison treatment program in order to fill the beds. In all these instances, it is hoped that the external motivation (i.e., legal pressure or fear of punishment) may transform into internal motivation (i.e., desire to change and stop using drugs) and therapeutic engagement during the treatment process (Sung et al. 2001).

Decades of evaluative efforts have concluded that coercion in substance abuse treatment is both a therapeutic factor that can be planned and manipulated for recovery purposes (Young 2002; Young and Belenko 2002) and that it is particularly useful when administered as structured, incremental responses to treatment noncompliance (Taxman et al. 1999). Studies after studies have attested the effectiveness of criminal justice-based treatment (National Institute on Drug Abuse [NIDA], 2012). However, and most importantly, the fundamental lesson learned from evaluation research remains this: Coercion can create incentives to enroll and remain in treatment, but coercive tactics by themselves contribute very little to a drug-free lifestyle if they are not supported by evidence-based therapeutic and service components.

Conclusion

Substance abuse treatment in the United States has followed a spectacular trajectory over the past 300 years that was largely influenced by public perception of addiction, as well as social and political movements. The popularity and acceptance of medical models of addiction treatment have ebbed and waned according to strict moral climates, culminating in the interdisciplinary system of treatment that we have today. With the growing consensus that addiction to psychoactive substances is "a chronic, relapsing brain disease that is characterized by compulsive drug seeking and use, despite harmful consequences" (NIDA 2010, p. 5), treatment programs currently attempt to use a holistic approach that addresses the many facets of this public health challenge. Research continues to explore the biological, psychological, and sociological causes of addiction in an effort to provide new interventions that can treat addiction more successfully.

Recommended Reading and References

Alcoholics Anonymous (2012) Estimates of A.A. groups and members. Retrieved May 24, 2012 from http://www.aa.org/lang/en/subpage.cfm?page=74

Anglin MD, Maugh TH (1992) Overturning myths about coerced drug treatment. Calif Psychol 22:19–22

Baumohl J (1990) Inebriate institutions in North America, 1840–1929. Br J Addict 85:1187–1204

Courtwright DT (2001) Dark paradise: opiate addiction in America before 1940, 2nd edn. Harvard University Press, Cambridge, MA

De Leon G (2000) The therapeutic community: theory, model, and method. Springer, New York

Dole VP, Nyswander M (2008) Methadone maintenance: a theoretical perspective. In: Inciardi JA, McElrath K (eds) The American drug scene: an anthology, 5th edn. Oxford University Press, New York, pp 435–439

Farabee D, Shen H, Prendergast M, Cartier J (2004) The effectiveness of coerced admission to prison-based drug treatment. In: Knight K, Farabee D (eds) Treating addicted offenders: a continuum of effective practices. Civic Research Institute, Kingston, pp 1–12

Hall W (2010) What are the policy lessons of national alcohol prohibition in the United States, 1920–1933? Addiction 105:1164–1173

Hartjen CA, Mitchell SM, Washburne NF (1981) Sentencing to therapy: some legal, ethical, and practical issues. J Offender Couns Serv Rehabil 6(1–2):21–39

Hartjen CA, Mitchell SM, Washburne NF (1982) Sentencing to therapy: some legal, ethical, and practical issues. J Offender Couns Serv Rehabil 6(1–2):21–39

Hazelden (2012) Hazelden history. Retrieved May 24, 2012 from http://www.hazelden.org/web/public/hazelden_history.page

Hubbard R, Simpson DD, Woody G (2009) Treatment research: accomplishments and challenges. J Drug Issues 39(1):153–165

Inciardi JA, McElrath K (eds) (2008) The American drug scene: an anthology, 5th edn. Oxford University Press, New York

Inciardi JA, McBride DC, Rivers JE (1996) Drug control and the courts. Sage, Thousand Oaks

Janzen R (2011) The rise and fall of Synanon: a California utopia. The Johns Hopkins University Press, Baltimore

Joint Committee of the American Bar Association and the American Medical Association on Narcotic Drugs (1961) Drug addiction, crime or disease? Interim and final reports of the Joint Committee of the American Bar Association and the American Medical Association on Narcotic Drugs. Indiana University Press, Bloomington

Joseph H, Stancliff S, Langrod J (2000) Methadone maintenance treatment (MMT): a review of historical and clinical issues. Mt Sinai J Med 67(5):347–364

Katcher BS (1993) Benjamin Rush's educational campaign against hard drinking. Am J Public Health 83(2):273–281

Lemanski M (2001) A history of addiction and recovery in the United States. Sharp Press, Tucson

Levine HG (1978) The discovery of addiction: changing conceptions of habitual drunkenness in America. J Stud Alcohol 39(1):143–174

Maddux JF (1988) Clinical experience with civil commitment. In: Leukefeld CG, Tims FM (eds) Compulsory

treatment of drug abuse: research and clinical practice. National Institute on Drug Abuse, Rockville
National Institute on Drug Abuse (2008) Advances in therapeutic communities. In: Inciardi JA, McElrath K (eds) The American drug scene: an anthology, 5th edn. Oxford University Press, New York, pp 421–434
National Institute on Drug Abuse (2010) Drugs, brains, and behavior: the science of addiction (Revised edition). National Institute on Drug Abuse, Bethesda
National Institute on Drug Abuse (2012) Principles of drug abuse treatment for criminal justice populations: a research-based guide (Revised edition). National Institute on Drug Abuse, Bethesda
Nutt DJ (2010) Antagonist-agonist combinations as therapies for heroin addiction: Back to the future? J Psychopharmacol 24(2):141–145
Platt JJ, Buhringer G, Kaplan CD, Brown BS, Taube DO (1988) The prospects and limitations of compulsory treatment for drug addiction. J Drug Issues 18(4):505–525
Quinn JF, Bodenhamer-Davis E, Koch DS (2004) Ideology and the stagnation of aoda treatment modalities in America. Deviant Behav 25:109–131
Rosenberg CE (1995) The care of strangers: the rise of America's hospital system. The Johns Hopkins University Press, Baltimore
Rothman DJ (2001) The discovery of the asylum: social order and disorder in the New Republic, 2nd edn. Little Brown, Boston
Saah T (2005) The evolutionary origins and significance of drug addiction. Harm Reduction J 2:8–14
Salmon RW, Salmon RJ (1983) The role of coercion in rehabilitation of drug abusers. Int J Addict 18(1):9–21
Stolberg VB (2006) A review of perspectives on alcohol and alcoholism in the history of American health and medicine. J Ethn Subst Abuse 5(4):39–106
Sung H-E, Belenko S (2006) From diversion experiment to policy movement: a case study of prosecutorial innovation in the United States. J Contemp Crim Justice 22(3):220–240
Sung H-E, Belenko S, Feng L (2001) Treatment compliance and problem incidents among criminal justice clients. J Subst Abuse Treat 20(2):153–162
Taxman FS, Soule D, Gelb A (1999) Graduated sanctions: stepping into accountable systems and offenders. Prison J 79(2):182–204
U.S. National Library of Medicine (2009) PubMed health: methadone. Retrieved May 24, 2012 from http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0000591/
White WL (1998) Slaying the Dragon. Chestnut Health Systems, Bloomington
Young D (2002) Impacts of perceived legal pressure on retention in drug treatment. Crim Justice Behav 29(1):27–55
Young D, Belenko S (2002) Program retention and perceived coercion in three models of mandatory drug treatment. J Drug Issues 32(1):297–328

History of Technology in Policing

Mathieu Deflem and Stephen Chicoine
Department of Sociology, University of South Carolina, Columbia, SC, USA

Overview

Technology has fundamentally changed the face of society, and the effects that technological advances have had on the form and function of policing is no exception. In response to the changing conditions of society brought about by new developments in technology, law enforcement agencies have in turn adopted new technological advances in order to fulfill their function in the areas of crime control and order maintenance. This has in turn led to the historical development of the bureaucratization of law enforcement, which has transformed the police from a political enforcement apparatus to an autonomous professional institution. Technology has not only facilitated this transformation of police within their respective nations, but has also facilitated international police cooperation. The institutional autonomy and the technological advances in communications and transportation enabled international cooperation between police organizations which addressed, ironically, the new forms of crime that were enabled by those very same technological advances. Harnessing new forms of technology would become a constant theme in the efforts of international police cooperation, with new initiatives and meetings calling for the advancement of the means for law enforcement agencies of different nations to exchange information on criminals, suspects, and new means of policing.

Technological changes have occurred for the better part of a century, tracing back to, at least, the early nineteenth century, when police were highly connected with the political sphere and walked their beats with whistles and batons, up until the beginning of the modern era shortly after World War 2. By this time, the police had become recognizably modern, despite all the advancements that still had yet to occur in more recent decades. The changes in the institution of policing have not been limited to the means by which police pursue their goals, but have also altered the organization and goals of policing from political concerns to criminal objectives.

Fundamental Policing Technologies: Weapons and Communications

Early forms of policing bear only partial likeness to modern-day law enforcement. In terms of technology, police agents were traditionally outfitted only with a truncheon or baton and a rattle for raising the alarm (Johnson 1981; Manwaring-White 1983). These early police were largely on their own upon leaving the police station since the technologies of the time did not allow for communication between officers patrolling their beats and their command structure. As a result, officers in the police station were rarely aware of criminal situations on the streets unless a citizen lived close enough to walk to the station or until an officer returned from patrol (Johnson 1981). Due to the lack of communication technology at the time, there was also little ability to coordinate officers on patrol or to keep account of the officers' actions while on the streets. While the means of physical repression of the police have remained largely unchanged since the institution's early inception, the means of communication and transportation have transformed the form and function of policing in rather considerable ways.

Historically police agents were armed only with a nightstick in the United States or a baton or truncheon in Europe. The adoption of firearms by police was once a contentious issue in terms of the power of the state to arm against its people, and first occurred in the United States, arguably as a result of the embedded nature of firearms in American culture (Johnson 1981). The initial standing orders given to officers prohibited the carrying of firearms while on patrol, and the public was largely opposed to arming the police (Johnson 1981). At this time during the formation of the United States, many within the United States carried firearms, and conflicts among citizens sometimes escalated to the point that shots were regularly exchanged. The police of the time were charged with resolving such conflicts, but officers would often be shot while on duty. After several shootings of police officers had taken place, it became an informal practice among the police to carry a firearm despite the orders against this, and this informal practice eventually came to be institutionalized (Johnson 1981). In the United States, the first agency to adopt the multi-shot pistol officially was the Texas Rangers in the 1850s, and other agencies across the nation adopted firearms shortly after (NCJ 1998). In England the firearm was adopted considerably later on, and the police there were not visibly armed until a shoot-out took place involving two anarchists in 1911 (Manwaring-White 1983).

Major advances in the history of policing and technology have taken place with respect to communications technology. Police in the 1830s relied on the rattle, whistle, or loud shout in order to alert other officers in the area to come their assistance (Stewart 1994; Johnson 1981). These police were essentially isolated while on patrol, and the police station was unable to communicate with them while they were on patrol. As a result, the station could not coordinate nor direct its officers on the streets, the officers could not report on situations on the streets until they returned, nor ask for assistance. Citizens were likewise limited in their access to police assistance unless they lived within walking distance to the station or until they could find an officer along his route. Several technological advancements were pursued and instituted in the efforts to facilitate communications between the police, the citizenry, and the station.

From the late nineteenth century onwards, important advances in communications technology began to be adopted by the police. The first police telephones were installed in Albany, New York, in 1877 and were used to connect the five districts with the mayor's office. Similarly, in the United Kingdom telephones were first used to connect police in 1880, and by 1886 the emergency phone system had connected all of the police and fire stations in Glasgow in the United Kingdom. While these earlier advances enhanced the ability of police to communicate with other stations, the mayor, and other emergency services, the police call box allowed for communications between police on patrol, citizens, and police stations.

While today the police call box has been subsumed by the widespread use of the radio and telephone, and is perhaps more recognizable as the transportation of the time-traveling Doctor in the BBC series *Dr. Who*, it was once a major advancement in police communication (Stewart 1994). In 1884, the first police call boxes were installed in Chicago and Boston, and in 1891 the call boxes were also adopted by the Glasgow police (Stewart 1994). Other police agencies began using the call boxes extensively during the 1890s (Johnson 1981). The call boxes served as a line of communication between beat patrols and headquarters, as well as a rallying point where citizens could call the police. Police call boxes were placed at the intersections of several police officers' patrol routes with a light at the top. This light would be turned on to alert the officers if the station needed to contact them, or they could be activated by citizens in need of police assistance. Officers could therefore be signaled from a distance, via the signal, that their attention was needed at the call box. Inside the box was a telephone for communication with the station, creating some degree of communication between the officers, the station, and the citizenry (Reiss 1992; Stewart 1994; Johnson 1981). The police call boxes increased the response rate and efficiency of officers responding to emergencies and added to the efficiency of the police by reducing the number of patrolling officers needed and operational costs (Stewart 1994). Despite these advantages, the police call box would eventually become replaced by the radio, just as the police foot patrol would eventually give way to the use of the automobile (Reiss 1992).

As early as 1920 it was becoming apparent that the foot patrol method of policing was handicapped in light of the growing prevalence of the automobile (Walton 1958). The increasing size of residential areas in large cities were also beginning to strain police stations by making the costs of foot patrols prohibitive (Walton 1958). The foot patrols of the previous century were therefore ill-suited to the new social conditions, requiring a new approach to policing. Under these conditions, police forces developed rapid-response methods with the aid of the telephone, two-way radio, and automobile. In the United States, these practices began to develop during the 1930s, with the first radio patrol car used in Detroit as early as 1929 (NCJ 1998; Reiss 1992; Battles 2010; Johnson 1981). This system of officers patrolling in automobiles, coordinated through the use of two-way radios responding to calls made by citizens, created a more efficient, effective, and responsive police force (Johnson 1981). This development was strongly connected to the specific conditions of 1930s America. The radio police car was not used until much later in other national settings such as the United Kingdom which did not widely adopt automobiles in policing until the 1960s (Manwaring-White 1983).

The developments in the history of police and technology have been largely influenced by the changing conditions that the new technologies created, as well as the bureaucratic drive of the police towards adopting standards of efficiency in policing. The advances in technology facilitated the development of the institutional autonomy of police institutions from politics, since the development of communications technology led to greater reporting to headquarters and thereby created superior professional officers instead of politically loyal servants (Johnson 1981). Variable social conditions influenced the adoption of technologies among police across countries. Firearms were adopted earlier by police in the United States, for instance, because they were also more prevalent among the population. The radio patrol car was also adopted earlier there than in the United Kingdom, which did not face the same issues in terms of geographic expansion. Finally, offenders themselves facilitated technological changes of policing by possessing firearms or driving automobiles. The response of the police has often been to adopt these technologies in the form of a criminal-law enforcement arms race.

This arms race has not been confined to the character of local or national policing, but has also influenced the development of criminal identification technologies and their adoption.

Advances in Criminal Identification

In premodern times there was no issue in criminal identification since the size of the communities virtually ensured that everyone knew everyone else (Cole 2001). As societies expanded and geographic mobility increased due to advances in transportation technology, such as the train, it became easier for criminals to hide from their previous history by moving to a different area or even by merely adopting a different name. Problems with linking criminals to their previous convictions became particularly problematic when, during the 1850s, a movement towards harsher punishments for recidivists began (Cole 2001). Two central problems faced the courts and law enforcement officials attempting to link a suspect with a criminal past: the ability to access the information stored, and the ability to positively identify a recidivist or other suspect.

In the beginning of the nineteenth century, efforts in criminal identification were limited to descriptions entered in prison registers which did not contain a uniform language and were not necessarily specific enough to distinguish an inmate. Further, such records did not differentiate between individuals with the same name and had no ability to prevent the use of aliases (Cole 2001). One of the first improvements upon this system was the introduction of record-keeping technologies developed by Arnould Bonneville de Marsangy in 1844. Bonneville suggested the use of an alphabetical system to facilitate the search for inmate record and the use of index cards which could be added to and expanded without disrupting the order of the information (Cole 2001). This advance did not alleviate the problems with identifying recidivists, since criminals needed only to fabricate a false name in order to fool the system.

Other technological advances that facilitated criminal identification included photography, anthropometry, and fingerprinting. The earliest reports of photography come from British and French police in the 1840s (Cole 2001). The first picture of a convict was taken in Brussels in 1843. Police departments began to collect these photos and the Parisian police were the first to set up a picture collection of criminals in 1874 (Deflem 2002a). By the end of the nineteenth century, all major police forces had photographic identification services. The photographs themselves provided pictures to the names that convicts gave the police, but in making positive identifications it was the photograph's use in the anthropometric system that had the largest impact.

Two technologies developed around the late nineteenth century which would bring about drastic advances in the area of criminal identification. The first to be implemented was the Bertillonage, or anthropometry, a system which categorized criminals by the measurements of their bodies (Cole 2001; Deflem 2002b; Johnson 1981). The anthropometric system was developed by Alphonse Bertillon in 1870 and was adopted by the French government in the 1880s. The anthropometric system involved specific and precise measurements of the criminal body, a written description, and two photographs, one profile and one full face, which served as the predecessor to today's mug shots. Bertillon emphasized the consistency and accuracy of these measurements and descriptions, laying out a precise process by which to take measurements and developing a specific terminology by which to describe the criminal (Cole 2001). Bertillon also developed a filing system which grouped criminals by 11 key measurements grouped into three general categories: body, head, and appendages (Johnson 1981). Bertillon's final advancement was the "*portrait parlé*" or spoken portrait, which reduced the criminal's description to a language which could be easily transmitted via telegraph or telephone (Cole 2001). As a result of the anthropometric system, a criminal's identity could be confirmed, regardless of the name they presented, by matching their bodily measurements with those on file. However, this process required a trained Bertillon operator and took several man hours to make an identification since measurements often had to be taken twice in order to make a positive identification (Cole 2001). Furthermore, as the anthropometric system spread from France and was adopted by other countries, modifications to the process and instruments led to a loss in the efficacy of the system and problems in transmitting data of the *portrait parlé* to the police of other nations (Cole 2001). The anthropometric system would become dominant in most nations as it was initially regarded as more scientific than fingerprinting (Cole 2001).

Fingerprint technologies were explicitly developed to develop a classification system useful for law enforcement. In 1877, Sir William J. Herschel proposed using fingerprinting for criminal identification in India and Dr. Henry Faulds came to a similar conclusion independently in 1880 (Cole 2001; Johnson 1981). Most importantly, in 1892 Sir Francis Galton published his findings that fingerprints did not change over the course of a lifetime and that each individual's fingerprints were unique (Johnson 1981; Cole 2001). This led to the incorporation of fingerprints in addition to the anthropometric system in 1894 by Scotland Yard (Johnson 1981). At the same time, Juan Vucetich in Argentina also adopted fingerprints for criminal identification (Deflem 2002a). The United States was slower to adopt fingerprints, with St. Louis being the first to use them after a Scotland Yard official introduced the technique but it was not instituted widely nationally until the 1930s when J. Edgar Hoover made the FBI available for identifying fingerprints and training local officers (Johnson 1981). Similar to the *portrait parlé*, fingerprints were described numerically on the basis of certain specified characteristics in order to facilitate their transmission via telegraph (Deflem 2002b).

In 1899, the colonial legislature passed the Indian Evidence Act, which was the first endorsement of fingerprints as legal evidence, not only identifying criminals and recidivists, but also for finding and convicting them (Cole 2001). Originally, fingerprints were not conceived of as a forensic tool, but merely as a record-keeping system. The earliest forensic identification occurred in Europe in 1902, when Alphonse Bertillon, developer of the anthropometric system, matched a bloody fingerprint at a crime scene with a fingerprint available in a criminal file (Cole 2001). Scotland Yard began using fingerprints for forensic identification in the beginning years of the twentieth century. At the time, fingerprints in the courts were faced with the same hesitancy as DNA testing later would be. The accuracy and precision of the method had to be demonstrated to the courts in oftentimes theatrical displays until the method eventually had become more solidified (Cole 2001). By the beginning of World War II, fingerprint examiners had transformed fingerprints from a novel technique into the most credible and unassailable form of criminal identification evidence.

The anthropometric system developed by Bertillon remained dominant for several years, but eventually fingerprinting would gradually overtake this system as the primary means for identifying criminals. Fingerprinting had several advantages to the anthropometric system in that they were quicker and did not require skilled labor or specialized instruments (Cole 2001). Fingerprints were advantageous since they offered industrial-style speed, efficiency, productivity, and economy over the anthropometric system (Cole 2001). The appeal of the anthropometric system was one of quality and scientific rigor, but these arguments faded in part due to the human error involved in the measurement and description process as well as the growing confidence in fingerprints (Cole 2001). Furthermore, fingerprinting had the additional advantage that it could be used not only for criminal identification, but for forensic identification as well. Fingerprints could not only help identify a recidivist, but help convict a criminal suspect as well.

The methods of fingerprinting and anthropometry both included an organizational classification by which they could be categorized and retrieved, addressing the early problem of accessing the information that was stored. The anthropometric system was particularly time consuming in accessing its records since two measurements had to be taken in order to effectively find a single file (Cole 2001). As the police bureaucracies grew, so did their records, and efficiently finding matching records became a steadily more daunting task. By the 1930s, the FBI had amassed the largest volume of fingerprints. However, to successfully access them and compare them required other identification information, and it still required a great deal of time to go through the records (Cole 2001). As early as the 1920s, identification bureaus began turning to automated data processing technologies such as IBM punch card sorters to handle searching the large volume of records. In the 1930s, the FBI and New York State Division of Criminal Identification began using such IBM card sorters (Cole 2001). However, these methods remained technically awkward and still required human examination of the selected cards, though they did represent an advancement in the technologies of data management which would develop drastically during the digital age.

Technologies of criminal identification initially developed as a means to identify recidivists. However, in the course of their development, one of the most well-known forms of forensic identification, fingerprinting, also came to be developed. These technologies had to be revised and expanded in view of the adaptive behavior of offenders. Attempting to avoid identification because of prior convictions, offenders would give false names to avoid the harsher penalties that were given to recidivists. This development in itself was a function of the increasing size and mobility of society, moving away from the small community wherein everyone was easily recognizable. This societal mobility, as well as the newly developed forms of criminal identification, were key elements to the growing internationalization of policing which developed to confront the growing class of criminals without borders.

Technologies of International Policing

Many characteristics of society were fundamentally changed by advances in technology, and as a result, the practice and organization of policing changed as well. In order to address the changing times, the police had to adopt new technologies in order to confront the realities of crime control. Police gained a greater level of technical superiority in the course of pursuing their organizational goals as the means of crime control developed in technical respects. This increasing technical superiority was a key factor in the development of professional police bureaucracies (Deflem 2002a). As the institution of law enforcement developed as a bureaucracy, importantly, police institutions gained autonomy from the political institutions of the state. And this autonomy facilitated the international cooperation between law enforcement organizations. Autonomy from the political sphere was a necessary prerequisite for international police cooperation since otherwise political agendas would disrupt the goal of purely criminal-law enforcement. These changes in the structure of law enforcement institutions were necessary to enable international cooperation, which developed directly to address the growing connectedness across societies as a result of advances in communications and transportation technologies.

Technology was central to the development of international policing (Deflem 2002b). First, it enabled the development of autonomous and professional law enforcement agencies by contributing to the bureaucratization of law enforcement through the necessary advances in technical superiority that accompanied the adoption of new technologies. Second, advances in technology created new forms of crime and enabled the internationalization of crime since advances in transportation and communications enabled criminals to coordinate internationally as well as providing a new level of mobility between nations. The internationalization of crime provided the motivation for law enforcement agencies to collaborate with foreign counterparts, and the same technologies that enabled international criminal behavior also enabled the police to communicate and cooperate on a scale that was never before possible.

The initial efforts for international police cooperation were motivated by a period of revolutionary activity during the 1840s throughout Europe. It was widely believed by police that the internationalization of protest was aided by the technological developments since the mid-1800s (Deflem 2002b). The press and printed writings could orient public opinion and could be distributed beyond national borders. As a result, the means to publish and publications of "subversive" groups were targeted by the police of the time. Furthermore, railway lines across Europe grew exponentially from the mid to late nineteenth century, thereby drastically increasing the ability of individuals to move from one nation to the next. This development was seen by police as presenting greater opportunities for political dissidents to cooperate across national boundaries in order to overthrow the state and, later, for offenders to flee from the jurisdictions in which they had committed a crime (Deflem 2002b). These threats, made possible through the printed word and railways, motivated law enforcement agencies to pursue international cooperation.

The earliest technologies to influence international policing were those that enabled the exchange of information (Deflem 2002a, b). During the 1840s, Europe experienced a revolutionary period which led to the adoption of a political objective by the police who had not yet developed an institutional autonomy from the political sphere. This led to the creation of the Police Union of German States in April 1851 with the express purpose of sharing information to track down political radicals. The Police Union focused particularly on developing faster means of exchanging information. Union members held regular meetings and exchanged printed magazines with information on political opponents. Police bureaus of Vienna and Berlin, both members of the Union, would also print bulletins containing information on wanted suspects and extradited foreigners. In the late nineteenth century, these bulletins began to be sent across Europe. The Police Union of German States would eventually disband due to political conflicts between its member states, but later efforts to develop international police cooperation would become more successful, particularly in facilitating the spread of police technologies and the use of communication technology to exchange information.

The initial forms of information exchange among the law enforcement agencies of different nations relied on printed media that contained information and pictures of criminals and wanted suspects. This method of exchange was particularly developed among the political police in Europe, but it was also used for criminal purposes, and the distribution of published materials on criminals remained a favored means of information exchange throughout the twentieth century (Deflem 2002b). In the 1880s, the New York City Police Department would develop a collection of photographs of international criminals which it shared with European authorities. The anthropometric, or Bertillon, system and fingerprints were the most important technological innovations to the internationalization of law enforcement.

Internationally, the same technologies of communications and transportation that facilitated international crime made not only the exchange of information among police possible, but also the diffusion of new means of criminal identification. In December 1898, Italian authorities coordinated an international police meeting, the International Conference of Rome for the Social Defense against the Anarchists, in order to develop a system to combat anarchism (Deflem 2002a). Although the meeting failed to develop a system to address anarchism, it did result in enhanced cooperation among police on certain practical matters of police technology. In particular, with the exception of the authorities from France and Britain, all of the countries agreed to introduce the *portrait parlé* system that was developed for the anthropometric system for the identification of criminals. The system enabled the transmission of information on criminals among nations since the system reduced the information to a series of numbers easily transmittable by telegraph or telephone (Deflem 2002b). A synthesis of technologies therefore began to be used, the method of criminal identification, its system for communication, and the technologies of communication, in order to facilitate the exchange of information among the police departments of different nations. This focus on developing better means of exchanging information among the police of various nations has remained a constant theme in the history of international policing until this day.

Criminal identification and the means of information exchange among police agencies would also be the central theme at the First International Police Congress in Monaco in 1914 (Deflem 2002a). The meeting was ineffective in developing an organization for international police cooperation, but it did include discussions and resolutions relevant to the adoption of technology by police agencies across national boundaries. The Congress discussed ways to enhance the exchange of information among police agencies of different nations, and a resolution was passed calling for the nations in attendance to grant free use of postal, telegraphic, and telephonic means of communication in regards to international criminals (Deflem 2002b). At the meeting, also, a universal system of identification including fingerprints and photos was proposed as well as an international publication containing search warrants.

Technologies of criminal identification were particularly important for the further development of international information exchange among police. The anthropometric system developed by Bertillon had spread to several other nations, but the system was not employed universally by all agencies. Police agencies in several nations augmented the system, which produced a problem in the exchange of information since the measurements and descriptions could not be relied upon to be internationally uniform (Cole 2001). Fingerprints presented the same problem with respect to the means by which the characteristics of the print were described and which fingerprints were taken. To address this problem, a so-called Distant Identification system was introduced at the Monaco Congress as a means of harmonizing the codes of different countries (Cole 2001). In 1923, at the International Police Congress in Vienna, the system was eventually adopted.

The 1923 International Police Congress in Vienna led to the development of the International Criminal Police Commission (ICPC), the organization known today as Interpol. There is no better example of the role of technology in the development of international law enforcement than the ICPC (Deflem 2002a). The facilities and institutionalized means of information exchange had expanded since the organization's inception, and by 1934 the ICPC headquarters located in Vienna included specialized divisions on fingerprints and photographs, the falsification of currencies and other documents, and various other forms of information on suspects and convicts. The ICPC also established systems of technologically advanced means for international communication, including a telegraphic code, a system of radio communications, and printed publications in addition to meetings. In addition to a strong focus on the most efficient use of new technologies, the ICPC was primarily preoccupied with a new growing class of offenders that committed crimes that were intrinsically made possible by technological advances, such as train robberies and hotel theft. Other international police initiatives would similarly focus on the use of new technologies to carry out the function of crime control, including international conferences as well as international exchanges of information between law enforcement agencies regarding methods of policing.

Conclusion

Technology remains a central focus in the organization and practices of policing. In large part, the increasing reliance on technology by police is the product of the way society at large has changed as a result of technological advancements. Modern society has expanded and become more interconnected, producing a new context within which law enforcement must operate. In part, these changes in addition to old challenges have motivated the adoption of new technology by police. These advancements have also facilitated international cooperation among law enforcement, with technologies enhancing the institutional autonomy needed for police agencies to cooperate across national borders. These technologies include not only advances in communication and transportation methods, but also pertain to the means of criminal identification which evolved into various forms that can be communicated among police of different jurisdictions. Technological advancements have not only provided police the tools by which to engage in collaboration, they have also provided an impetus for police collaboration by influencing the internationalization of crime.

The means of communication among police regionally and internationally, the means of criminal identification, and the means of transportation have been central to the transformation of policing. These advancements began as steps towards addressing the everyday issues in policing. What was once the lone patrolman armed with a whistle and baton, walking his beat without oversight or coordination and often strongly connected with political power, has become a professional police institution that endeavors to achieve its goal of crime control efficiently. Thus developed a police that organizes patrol with automobiles and two-way radios, and that relies on identification methods using photographs and fingerprints and can share information across jurisdictions with the use of telegraphs, telephones, and mail. Changes in the means of policing have also directly addressed technologically enabled changes in society, specifically the growing societal mobility which made individuals harder to identify, the larger geographic space which made beat patrols cost prohibitive and ineffective, and the inability of the early police to coordinate their actions. Since that time, even more advancements in technology have further changed the institution of law enforcement. Undoubtedly, policing will continue to adapt to the changing circumstances that are brought about by technological advancements.

Related Entries

- ▶ DNA Technology and Police Investigations
- ▶ Fingerprint Identification
- ▶ Identification and the Development of Forensic Science
- ▶ Identification Technologies in Policing and Proof
- ▶ Information Technology and Police Work
- ▶ Police Use of Firearms

Recommended Reading and References

Battles K (2010) Calling all cars: police dragnets and the technology of policing. University of Minnesota Press, Minneapolis

Cole SA (2001) Suspect identities: a history of fingerprinting and criminal identification. Harvard University Press, Cambridge, MA

Deflem M (2002a) Technology and the internationalization of policing: a comparative-historical perspective. Justice Quart 19(3):453–475

Deflem M (2002b) Policing world society. Oxford University Press, Oxford

Johnson DR (1981) American law enforcement: a history. Forum Press, Saint Louis

Manwaring-White S (1983) The policing revolution: police technology, democracy and liberty in Britain. Harvester Press, Brighton

NCJ (1998) Evolution and development of police technology. NCJ 173179

Reiss AJ Jr (1992) Police organization in the twentieth century. Crime Justice 15:51–97

Stewart RW (1994) The police signal box: a 100 year history. Eng Sci Educ J 3(4):161–168

Walton FE (1958) Selective distribution' of police patrol force, history, current practices, recommendations. J Crim Law Criminol Police Sci 49(2):165–171

History of the Death Penalty

Michael H. Reggio

Synonyms

History of capital punishment

Overview

From the beginning of time, the death penalty has been enacted. However, throughout history, it has decreased more and more until today the world has the least number of people put to death for crime. This entry describes the historical development of capital punishment through the centuries and provides an account of the forms and methods used to execute convicted offenders. The entry further discusses the behaviors and crimes that led to the ultimate punishment, while presenting the controversies and legal aspects of capital punishment. Violent crimes and murder rates are also discussed in lieu of the ultimate punishment, and its potential effect on current crime rates.

Lethal injection, electric chair, firing squad, hanging, poison, drowning, drawn and quartering, immersed in boiling oil¸ stoning, crucifixion, sawing asunder, pressing, tearing apart by wild beasts, beheading – all of these and whatever more the mind could conceive have been used to impose the death penalty. As far back as the Ancient Laws of China, the death penalty was established as a punishment for crimes. In eighteenth Century BC, the Code of King Hammurabi of Babylon codified the death penalty for 25 different crimes, although murder was not one of them. The first death sentence historically recorded occurred in the sixteenth century BC Egypt where the wrongdoer, a member of nobility, was accused of magic, and ordered to take his own life. During this period, non-nobility was usually killed with an axe.

In the fourteenth century BC, the Hittite Code also prescribed the death penalty. The seventh century BC Draconian Code of Athens made death the penalty for every crime committed. In the fifth century BC, the Roman Law of the 12 Tablets codified the death penalty. Again, the death penalty was different for nobility, freemen, and slaves. Punishment was for crimes such as the publication of libels and insulting songs, the cutting or grazing of crops planted by a farmer, the burning of a house or a stack of corn near a house, cheating by a patron of his client, perjury, making disturbances at night in the city, willful murder of a freeman or a parent, or theft by a slave. Death was often cruel and included crucifixion, drowning at sea, burial alive, beating to death, and impalement (often used by Nero). The Romans had a curious punishment for parricides (murder of a parent): The condemned was submersed in water in a sack, which also contained a dog, a rooster, a viper, and an ape (Laurence 1960, pp. 1–3). The most notorious death execution in BC was – about – 399 BC when the Greek philosopher Socrates was required to drink poison for heresy and corruption of youth (Kronenwetter 1993, p. 71).

Mosaic Law codified many capital crimes. In fact, there is evidence that Jews used many different techniques including stoning, hanging, beheading, crucifixion (copied from the Romans), throwing the criminal from a rock, and sawing asunder. The most infamous execution in history occurred approximately 29 AD with the crucifixion of Jesus Christ outside Jerusalem by the Romans. About 300 years later, the Emperor Constantine, after converting to Christianity, abolished crucifixion and other cruel death penalties in the Roman Empire. In 438, the Code of Theodosius made more than 80 crimes punishable by death (Kronenwetter 1993, p. 72).

Britain influenced the colonies more than any other country and has a long history of punishment by death. About 450 BC, the death penalty was often enforced by throwing the condemned into a quagmire. By the tenth Century, hanging from gallows was the most frequent execution method. William, the Conqueror, opposed taking life except in war, and ordered no person to be hanged or executed for any offense. However, he allowed criminals to be mutilated for their crimes. During the middle ages, capital punishment was accompanied by torture. Most barons had a drowning pit as well as gallows, and they were used for major as well as minor crimes. For example, in 1279, 289 Jews were hanged for clipping coin (cutting coins into pieces for their worth). Under Edward I, two gatekeepers were killed because the city gate had not been closed in time to prevent the escape of an accused murderer. Burning was the punishment for women's high treason and men were hanged, drawn, and quartered. Beheading was generally accepted for the upper classes. One could be burned for marrying a Jew. Pressing became the penalty for those who would not confess to their crimes. The executioner placed heavy weights on the victim's chest. On the 1st day he gave the victim a small quantity of bread, on the second day a small drink of bad water, and so on until he confessed or died. Under the reign of Henry VIII, the number of those put to death is estimated as high as 72,000. Boiling to death was another penalty approved in 1531, and there are records to show some people boiled for up to 2 h before death took them. When a woman was burned, the executioner tied a rope around her neck when she was tied to the stake. When the flames reached her she could be strangled from outside the ring of fire. However, this often failed and many were literally burnt alive (Kronenwetter 1993, p. 72; Laurence 1960, pp. 4–9).

In Britain, the number of capital offenses continually increased until the 1700s when 222 crimes were punishable by death. These included stealing from a house in the amount of 40 shillings, stealing from a shop the value of five shillings, robbing a rabbit warren, cutting down a tree, disturbing the peace, and counterfeiting tax stamps. However, juries tended not to convict when the penalty was great and the crime was not. Reforms began to take place. In 1823, five laws passed, exempting about a 100 crimes from the death [penalty]. Between 1832 and 1837, many capital offenses were swept away. In 1840, there was a failed attempt to abolish all capital punishment. Through the nineteenth and twentieth centuries, more and more capital punishments were abolished, not only in Britain, but also all across Europe, until today only a few European countries retain the death penalty (Laurence 1960, pp. 9–14). Worldwide, this trend to abolish the death penalty also holds true. While in the past, the number of capital punishments was in the 1,000 each year, between 2007 and 2010, only 5,541 executions took place (excluding China where information is sparse, but it is estimated that 1,000 are still executed, in fact more than all other countries combined). The major exception today is with some Greater Middle Eastern countries. For example, during the time this entry was written, these countries had numerous executions: Iran – 1663, Saudi Arabia – 423, Iraq – 256, Pakistan – 171, Yemen – 152, Libya – 39, Afghanistan – 34, Syria – 33, Bangladesh – 28, Egypt – 12. A few other countries notable for their high execution rate during this time are: China with 1,000, United States – 220, North Korea – 105, Madagascar – 39, and Vietnam – 58 (Death Sentences & Executions 2007–2010, October 2012). It is documented that capital punishment was carried out in 23 countries in 2010 and only 20 countries in 2011. It is estimated that executions increased from 527 in 2010 to at least 676 executions in 2011. This increase was mainly due to significant increases in judicial killings in the three large Middle Eastern countries of Iran, Iraq, and Saudi Arabia (Death Sentences and Executions 2011, October 2012). The trend is clear. In 1977, only 16 countries had abolished the death penalty. By 2012, 141 countries had abolished by practice or by law (Death Penalty Trends, March 2012).

The first recorded execution in the English American colonies was in 1608 when officials executed George Kendall of Virginia for supposedly plotting to betray the British to the Spanish. In 1612, Virginia's governor, Sir Thomas Dale, implemented the Divine, Moral, and Martial Laws that made death the penalty for even minor offenses such as stealing grapes, killing chickens, killing dogs or horses without permission, or trading with Indians. Seven years later, these laws were softened because Virginia feared that no one would settle there (Kronenwetter 1993, pp. 72–73).

In 1622, the first legal execution of a criminal, Daniel Frank, occurred in Virginia for the crime of theft (Bedau 1982, p. 5). Some colonies were very strict in their use of the death penalty, while others were less so. In Massachusetts Bay Colony, the first execution was in 1630, but the earliest capital statutes do not occur until later. Under the Capital Laws of New-England that went into effect between 1636 and 1647, the death penalty was meted out for premeditated murder, sodomy, witchcraft, adultery, idolatry, blasphemy, assault in anger, rape, statutory rape, manstealing, perjury in a capital trial, rebellion, manslaughter, poisoning, and bestiality. Early laws were accompanied by a scripture from the Old Testament. By 1780, the

Commonwealth of Massachusetts only recognized seven capital crimes: murder, sodomy, burglary, buggery, arson, rape, and treason (Bedau 1982, p. 7).

The New York colony instituted the so-called Duke's Laws of 1665. This directed the death penalty for denial of the true God, premeditated murder, killing someone who had no weapon of defense, killing by lying in wait or by poisoning, sodomy, buggery, kidnapping, perjury in a capital trial, traitorous denial of the king's rights or raising arms to resist his authority, conspiracy to invade towns or forts in the colony, and striking one's mother or father (upon complaint of both). The two colonies that were more lenient concerning capital punishment were South Jersey and Pennsylvania. South Jersey had no death penalty except for murder and treason, both punishable by death (Mackey 1976 pp. xi–xii).

Under the direction of the Crown, all the colonies included the death penalty for major crimes, especially treason and premeditated murder, but William Penn's Great Act (1682) changed this in Pennsylvania and made the penalty imprisonment. By 1776, most of the colonies had roughly comparable death statutes which covered arson, piracy, treason, murder, sodomy, burglary, robbery, rape, horse-stealing, slave rebellion, and often counterfeiting. Hanging was the usual sentence. Rhode Island was probably the only colony which decreased the number of capital crimes in the late 1700s.

Some states were more severe. For example, by 1837, North Carolina required death for the crimes of murder, rape, statutory rape, slave-stealing, stealing bank notes, highway robbery, burglary, arson, castration, buggery, sodomy, bestiality, dueling where death occurs, hiding a slave with intent to free him, taking a free Negro out of state to sell him, bigamy, inciting slaves to rebel, circulating seditious literature among slaves, accessory to murder, robbery, burglary, arson, or mayhem and others. However, North Carolina did not have a state penitentiary and, many said, no suitable alternative to capital punishment (Bedau 1982, p. 7).

The first reforms of the death penalty occurred between 1776 and 1800. Thomas Jefferson and four others, authorized to undertake a complete revision of Virginia's laws, proposed a law that recommended the death penalty for only treason and murder. After a stormy debate, the legislature defeated the bill by one vote. The writing of European theorists such as Montesquieu, Voltaire, and Bentham had a great effect on American intellectuals, as did English Quaker prison reformers John Bellers and John Howard (Mackey 1976, pp. 7–8).

On Crimes and Punishment, published in English in 1767 by the Italian jurist Cesare Beccaria, whose exposition on abolishing capital punishment was the most influential of the time, had an especially strong impact. He theorized that there was no justification for the taking of life by the state. He said that the death penalty was "a war of a whole nation against a citizen, whose destruction they consider as necessary, or useful to the general good." He asked the question what if it can be shown not to be necessary or useful? His essay conceded that the only time a death was necessary was when only one's death could insure the security of a nation – which would be rare and only in cases of absolute anarchy or when a nation was on the verge of losing its liberty. He said that the history of using punishment by death (e.g., the Romans, 20 years of Czaress Elizabeth) had not prevented determined men from injuring society and that death was only a "momentary spectacle, and therefore a less efficacious method of deterring others, than the continued example of a man deprived of his liberty...." (Beccaria 1963).

Organizations were formed in different colonies for the abolition of the death penalty and to relieve poor prison conditions. Dr. Benjamin Rush, a renowned Philadelphia citizen, proposed the complete abolition of capital punishment. William Bradford, Attorney General of Pennsylvania, was ordered to investigate capital punishment. In 1793, he published *An Enquiry How Far the Punishment of Death is Necessary in Pennsylvania*. He strongly insisted that the death penalty be retained, but admitted it was useless in preventing certain crimes. In fact, he said the death penalty made convictions harder to obtain, because in Pennsylvania, and indeed in all states, the death penalty was mandatory and juries would often not return a guilty verdict because of this fact. In response, in 1794, the Pennsylvania legislature abolished capital punishment for all crimes except murder "in the first degree," the first time murder had been broken down into "degrees." In New York, in 1796, the legislature authorized construction of the state's first penitentiary, abolished whipping, and reduced the number of capital offenses from 13 to 2. Virginia and Kentucky passed similar reform bills. Four more states reduced its capital crimes: Vermont in 1797, to three; Maryland in 1810, to four; New Hampshire in 1812, to two; and Ohio in 1815, to two. Each of these states built state penitentiaries. A few states went in the opposite direction. Rhode Island restored the death penalty for rape and arson; Massachusetts, New Jersey, and Connecticut raised death crimes from six to ten, including sodomy, maiming, robbery, and forgery. Many southern states made more crimes capital, especially for slaves (Mackey 1976, pp. xvi–xvii).

The years 1833–1853 saw the next great reform era. Public executions were attacked as cruel. Sometimes tens of thousands of eager viewers would show up to view hangings; local merchants would sell souvenirs and alcohol. Fighting and pushing would often break out as people jockeyed for the best view of the hanging or the corpse! Onlookers often cursed the widow or the victim and would try to tear down the scaffold or the rope for keepsakes. Violence and drunkenness often ruled towns far into the night after "justice had been served." Many states enacted laws providing private hangings. Rhode Island (1833), Pennsylvania (1834), New York (1835), Massachusetts (1835), and New Jersey (1835), all abolished public hangings. By 1849, 15 states were holding private hangings. This move was opposed by many death penalty abolitionists who thought public executions would eventually cause people to cry out against execution itself. For example, in 1835, Maine enacted what was in effect a moratorium on capital punishment after over 10,000 people who watched a hanging had to be restrained by police after they became unruly and began fighting. All felons sentenced to death would have to remain in prison at hard labor and could not be executed until 1 year had elapsed and then only on the governor's order. No governor ordered an execution under the "Maine Law" for 27 years. Though many states argued the merits of the death penalty, no state went as far as Maine. The most influential reformers were the clergy. Ironically, the small but powerful group which opposed the abolitionists were also clergy. They were, almost to a person, members of the Calvinist clergy, especially the Congregationalists and Presbyterians who could be called the religious establishment of the time. They were led by George Cheever (Mackey 1976, pp. xix–xxv).

Finally, in 1846, Michigan became the first state to abolish the death penalty (except for treason against the state), mostly because it had no long tradition of capital punishment (there had been no hanging since 1830, before statehood) and because frontier Michigan had few established religious groups to oppose it as was the case in the east. In 1852, Rhode Island abolished the death penalty led by Unitarians, Universalists, and especially Quakers. In the same year, Massachusetts limited its death penalty to first-degree murder. In 1853, Wisconsin abolished the death penalty after a gruesome execution in which the victim struggled for 5 min at the end of the rope, and a full 18 min passed before his heart finally quit (Mackey 1976, pp. xvii–xviii).

During the last half of the century, the death penalty abolition movement was cut in half with many of its followers moving into the slavery abolition movement. At the same time, states began to pass laws against mandatory death sentences. Legislators in 18 states shifted from mandatory to discretionary capital punishment by 1895, not to save lives, but to try to increase convictions and executions of murderers. Still, abolitionists gained a few victories. Maine abolished the death penalty, restored it, and then abolished it again between 1876 and 1887. Iowa abolished the death penalty for 6 years. Kansas passed the "Maine Law" in 1872 which operated as de facto abolition (Mackey 1976, pp. xxx–xxxi).

Electrocution as a method of execution came onto the scene in an unlikely manner. Edison Company with its DC (Direct Current) electrical systems began attacking Westinghouse Company and its AC (Alternating Current) electrical systems as they were pressing for nationwide electrification with alternating current. To show how dangerous AC could be, Edison Company began public demonstrations by electrocuting animals. People reasoned that if electricity could kill animals, it could kill people. In 1888, New York approved the dismantling of its gallows and the building of the nation's first electric chair. It held its first victim, William Kemmler, in 1890, and even though the first electrocution was clumsy at best, other states soon followed the lead (Mackey 1976, p. 15).

The Second Great Reform era was 1895–1917. In 1897, US Congress passed a bill reducing the number of federal death crimes. In 1907, Kansas took the "Maine Law" a step further and abolished all death penalties. Between 1911 and 1917, eight more states abolished capital punishment (Minnesota, North Dakota, South Dakota, Oregon, Arizona, Missouri, and Tennessee – the latter in all cases but rape). Votes in other states came close to ending the death penalty.

However, between 1917 and 1955, the death penalty abolition movement again slowed. Washington, Arizona, and Oregon in 1919–1920 reinstated the death penalty. In 1924, the first execution by cyanide gas took place in Nevada, when Tong war gang murderer Gee Jon became its first victim. The state wanted to secretly pump cyanide gas into Jon's cell at night while he was asleep as a more humanitarian way of carrying out the penalty, but, technical difficulties prohibited this and a special "gas chamber" was hastily built. Other concerns developed when less "civilized" methods of execution failed. In 1930, Mrs. Eva Dugan became the first female to be executed by Arizona. The execution was botched when the hangman misjudged the drop and Mrs. Dugan's head was ripped from her body. More states converted to electric chairs and gas chambers. During this period of time, abolitionist organizations sprang up all across the country, but they had little effect. There were a number of stormy protests against the execution of certain convicted felons (e.g., Julius and Ethel Rosenberg), but little opposition against the death penalty itself. In fact, during the anti-Communist period with all its fears and hysteria, Texas Governor Allan Shivers seriously suggested that capital punishment be the penalty for membership in the Communist Party (Mackey 1976, pp. xxxii–xxxiv, xli). The movement against capital punishment revived again between 1955 and 1972.

England and Canada completed exhaustive studies which were largely critical of the death penalty and these were widely circulated in the United States. Death row criminals gave their own moving accounts of capital punishment in books and film. Convicted kidnapper Caryl Chessman published *Cell 2455 Death Row* and *Trial by Ordeal*. Barbara Graham's story was utilized in book and film with *I Want to Live!* after her execution. Tim Robbins' produced and directed the movie *Dead Man Walking* based on Sister Helen Prejean's book about Matthew Poncelet, a prisoner in Louisiana State Prison. This Catholic nun became one of the leading proponents for the abolition of the capital punishment. Not only movies, but television shows were broadcast on the death penalty.

Hawaii and Alaska ended capital punishment in 1957, and Delaware did so the next year. Controversy over the death penalty gripped the nation, forcing politicians to take sides. Delaware restored the death penalty in 1961. Michigan abolished capital punishment for treason in 1963. Voters in 1964 abolished the death penalty in Oregon. In 1965 Iowa, New York, West Virginia, and Vermont ended the death penalty. New Mexico abolished the death penalty in 1969 (Mackey 1976, pp. xlvii–xlix). Some states reestablished capital punishment and some states

History of the Death Penalty, Fig. 1 Death row

abolished it. Some reestablished it and then abolished it again. Today, 33 states have the death penalty and 17 do not (2 of these, Connecticut & New Mexico, have outlawed the death penalty, but only for new convictions and still have people on death row). In the last 5 years, five states have outlawed capital punishment Fig. 1.

Trying to end capital punishment state-by-state was difficult at best, so death penalty abolitionists turned much of their efforts to the courts. They finally succeeded on June 29, 1972, in the case *Furman v. Georgia*. In nine separate opinions, but with a majority of 5–4, the US Supreme Court ruled that the way capital punishment laws were written, including discriminatory sentencing guidelines, capital punishment was cruel and unusual and violated the Eighth and Fourteenth Amendments. This effectively ended capital punishment in the United States. Advocates of capital punishment began proposing new capital statutes which they believed would end discrimination in capital sentencing, therefore satisfying a majority of the Court. By early 1975, 30 states had again passed death penalty laws and nearly 200 prisoners were on death row. In *Gregg v. Georgia* (1976), the Supreme Court upheld Georgia's newly passed death penalty and said that the death penalty was not always cruel and unusual punishment. Death row executions could again begin. Another form of execution was soon found. Oklahoma passed the first death by lethal injection law, based on economics as much as humanitarian reasons. The old electric chair that had not been used in 11 years would require expensive repairs. Estimates of over $200,000 were given to build a gas chamber, while lethal injection would cost no more than 10–15 dollars "per event" (Bedau, 1982, p. 17). Since 1976, there have been 1,131 lethal injections (35 states plus the US government use lethal injection), 157 electrocutions [last in 2010], 11 in gas chambers [last in 1999], 3 by hanging [last in 1996], and 3 by firing squad [last in 2010]) (Facts about the Death Penalty, September 2012) Fig. 2.

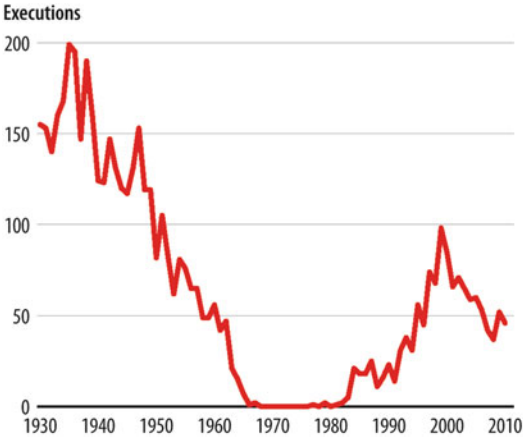

History of the Death Penalty, Fig. 2 Number of persons executed in the United States during 1930–2010. (Source: BJS, National Prisoner Statistics Program)

In 2011, all state executions were carried out by lethal injection. Today, the United States government lists 41 offenses punishable by death. They include espionage, treason, and death resulting from aircraft hijacking. However, they mostly comprise various forms of murder such as murders: related to the smuggling of aliens, committed during a drug-related drive-by shooting, during a civil rights offense, of a Congressperson or Supreme Court Justice, during a hostage taking, by use of a weapon of mass destruction, retaliating against an immediate family member of a law enforcement official, etc. (41 Federal Capital Offenses, October 2012).

According to Bureau of Justice Statistics, "Homicide Trends" (issued in 2011) murder rates recently declined to levels last found in the 1960s. In the late 1960–2000, the rate trended upward. It doubled from the 1960s to the late 1970s. In 1980, it peaked at 10.2 and fell to 7.9 in 1985. It rose again in the late 1980s and early 1990s to another peak of 9.8 in 1991 (FBI Uniform Crime Report 1992, October 2012). In late 1990s, it began to trend down. By 2000, it was 5.5 and by 2010 it was 4.8 (Homicide Trends in the U.S., November 2011, October 2012).

The controversy over the death penalty continues today. During the 1990s, there was a strong movement against lawlessness propelled by citizens' fears for their security. Politicians at the national and state levels were taking the floor of legislatures and calling for more frequent death penalties, death penalties for more crimes, and longer prison sentences. But, correlating with the drop of homicides, the annual death sentences in the United States have dropped dramatically since the year 2000. In the last 4 years, the number of death sentences has been lower than any time since reinstatement of the death penalty in 1976. Executions have declined as well, from a high of 98 in 1999 to just 37 in 2008; there were 46 executions in 2010, and 43 executions in 2011.

The battle lines are still drawn and the combat will probably always be fought. The arguments range from morality, to deterrence, to economics, to socioeconomic level to attorney ability. More than half of Americans today look toward alternative consequences for homicide rather than use the death penalty. However, this could change, especially if the homicide rate goes up. The death penalty seems positively correlated to this rate.

A number of important capital punishment decisions have been made by the Supreme Court. The following is a list of the more important ones along with their legal citations:

Wilkerson v. Utah 99 U.S. 130 (1878) – Court upheld execution by firing squad, but said that other types of torture such as "drawing and quartering, emboweling alive, beheading, public dissection, and burring alive and all other in the same line of. . .cruelty, are forbidden."

Weems v. U.S. 217 U.S. 349 (1910) – Court held that what constitutes cruel and unusual punishment had not been decided, but that it should not be confined to the "forms of evil" that framers of the Bill of Rights had experienced. Therefore, "cruel and unusual" definitions are subject to changing interpretations.

Louisiana ex rel. Francis v. Resweber 329 U.S. 459 (1947) – On May 3, 1946, convicted 17-year-old felon Willie Francis was placed in the electric chair and the switch was thrown. Due to faulty equipment, he survived (even though he was severely shocked). He was removed from the chair and returned to his cell. A new death warrant was issued 6 days later. The Court ruled 5–4 that it was not "cruel and unusual" to finish carrying out the sentence since the state acted in good faith in the first attempt. "The cruelty against which the Constitution protects a convicted man is cruelty inherent in the method of punishment," said the Court, "not the necessary suffering involved in any method employed to extinguish life humanely." He was then executed.

Tropp v. Dulles 356 U.S. 86 (1958) – The Court Ruled that punishment would be considered "cruel and unusual" if it was one of "tormenting severity," cruel in its excessiveness or unusual in punishment "must draw its meaning from the evolving standards of decency that mark the progress of a maturing society."

Furman v. Georgia 408 U.S. 238 (1972) – The Court looking at three cases struck down the death penalty in many states and set up the standard that punishment would be considered "cruel and unusual" if any of the following were present: (1) it was too severe for the crime; (2) it was arbitrary (some get the punishment and others do not, without guidelines); (3) it offends society's sense of justice; (4) it was not more effective than a less severe penalty.

Gregg v. Georgia 428 U.S. 153 (1976) – [The] Court upheld Georgia's newly passed death penalty and said that the death penalty was not always cruel and unusual punishment.

Tison v. Arizona 481 U.S. 137 (1987) – [The] Court upheld Arizona's death penalty for major participation in a felony with "reckless indifference to human life."

Thompson v. Oklahoma 108 S. Ct. 2687 (1987) – The Court considered the question of execution of minors under the age of 16 at the time of the murder. The victim was Thompson's brother-in-law, whom he accused of beating his sister. Thompson and three others beat the victim; shot him twice; cut his throat, chest, and abdomen; chained him to a concrete block; and threw the body into a river where it remained for 4 weeks. Each of the four participants was tried separately and all were sentenced to death. In a 5–3 decision, four Justices ruled that Thompson's death sentence violated the Eighth Amendment's prohibition of cruel and unusual punishment. The fifth, O'Connor, concurred but noted that a state must set a minimum age and held out the possibility that if a state lowers, by statute, the minimum death penalty age below 16, she might support it. She stated, "Although, I believe that a national consensus forbidding the execution of any person for a crime committed before the age of 16 very likely does exist, I am reluctant to adopt this conclusion as a matter of constitutional law without better evidence that [sic] we now possess." States with no minimum age have rushed to specify a statute age.

Penry v. Lynaugh 492 U.S. 302 (1989) – The Court held that persons considered retarded, but legally sane, could receive the death penalty. It was not cruel and unusual punishment under the Eighth Amendment if jurors were given the opportunity to consider mitigating circumstances. In this case, the defendant had the mental age of approximately a 6-year-old child.

Atkins v. Virginia 536 U.S. 304 (2002) The Court held in that it is a violation of the 8th Amendment's prohibition of cruel and unusual punishment to execute defendants with mental retardation.

Roper v. Simmons 543 U.S. 551 (2005) The Court struck down executions of juveniles on the basis of 14th Amendment due process and 8th Amendment cruel & unusual punishment, overturning previous decisions allowing death penalty for minors.

Related Entries

- ▶ Sex Offenders and Criminal Policy
- ▶ Designing Products Against Crime
- ▶ Deterrence
- ▶ History of Corrections
- ▶ History of Geographic Criminology Part I: Nineteenth Century
- ▶ History of Geographic Criminology Part II: 20th Century
- ▶ International Responses to Victims in Criminal Justice
- ▶ International Sentencing
- ▶ Lawful Killings
- ▶ Proportionality
- ▶ Punishment and the State
- ▶ Theories of Punishment

Recommended Reading and References

"41 Federal Capital Offenses," ProCon.org, Web, deathpenalty.procon.org/view.resource.php?re sourceD =004927

Beccaria C (1963) On crimes and punishment, (trans. Henry Paolucci). Bobbs-Merrill, Indianapolis

Bedau HA (1982) The death penalty in America. Oxford University Press, New York

"Death Penalty Trends," Amnesty International, Web, March 26, 2012 www.amnestyusa.org/our-workssues/death-penalty/us-death-penalty-facts/death-penalty-trends

"Death Sentences and Executions 2011," Amnesty International, Web, March 26, 2012 www.amnestyusa.org/research/reports/death-sentences-and-executions-2011

"Death Sentences & Executions 2007-2010", docs.google.com/spreadsheet/ccc?key=0AonYZs4MzlZbdGJiUzRwTVhlM25DWDlP djBmNURjOUE&hl=en#gid=0

"Facts about the Death Penalty - *Updated: September 20, 2012*)," Death Penalty Information Center, Web, www.deathpenaltyinfo.org/documents/FactSheet.pdf

FBI Uniform Crime Report 1992; The Sentencing Project.

"Homicide Trends in the U.S., 1980–2008" November 2011, NCJ 236018, Web, bjs.ojp.usdoj.gov/index.cfm?ty=pbdetail&iid=2221

Kronenwetter M (1993) Capital punishment: a reference handbook. ABC-CLIO, Santa Barbara

Laurence J (1960) A history of capital punishment. The Citadel Press, New York

Mackey PE (1976) Voices against death: American opposition to capital punishment, 1787–1975. Burt Franklin, New York

History of the Dutch Crime Victimization Survey(s)

Paul R. Smit[1] and Jan Van Dijk[2]
[1]Ministry of Safety and Justice, Research and Documentation Centre, The Hague, The Netherlands
[2]Human Security and Safety International Victimology Institute, Tilburg University, Tilburg, The Netherlands

Overview

The series of Crime Victim Surveys started in the early 1970s in the Netherlands. Changing policy needs, stakeholders, and available interview modes have repeatedly led to important adjustments in the methodology, compromising comparisons of results over the years. Nevertheless, interesting trends spanning four decades can be constructed for some crime types. In this entry, an overview is given of the eight subsequent versions of the Dutch Victim Survey and their main characteristics. Next, longtime trends are presented for victimizations by burglary, bicycle theft, violent crime, and total crime. Trends in feelings of unsafety are shown from the early 1990s onwards.

Introduction

Already in the early 1970s, the need for a survey to measure crime by asking the general population about their experiences instead of relying on police records only was felt in the Netherlands. In 1973, the Ministry of Justice funded a pilot study by Jan Fiselier who experimented with different interview modes and different wordings of key questions (Fiselier 1978). In 1974, the Ministry commissioned the first round of what would develop into an annual victimization survey (van Dijk and Steinmetz 1980). The results were initially published in the statistical paragraph of the annual report of the Dutch Prosecution Service as a complement to statistics of recorded crime. The Dutch victimization survey has, as most other similar surveys, undergone many changes in its organizational setup and methodology (questionnaire, sampling, sample size, interview modes). Even the name of the survey has changed no less than seven times over the years, the current name being the “Veiligheidsmonitor” (Security Survey). During several years, two or more different surveys were conducted parallel to each other. The Netherlands has also participated in all rounds of the International Crime Victim Surveys (ICVS).

The many transformations which the Dutch survey has undergone can be explained by three main factors. The most important reason for changing the existing survey was that leading government agencies continued to formulate new information needs related to their topical policy agendas. For instance, in the Netherlands, the interest shifted from just measuring the trends in real crime to monitoring perceptions of the public regarding safety issues and the performance of the police. From the early years of the surveys, questionnaires included sets of questions on varying themes such as the use of crime prevention measures and victim needs. An important driver of adjustments in the survey’s methodology in the 1990s was the emerging need to monitor crime trends, safety perceptions, and police performances in all 25 police districts in the country. This new policy objective required much larger sampling sizes, sometimes of up to 200,000 households, with obvious costs implications having an impact on the choice of interview techniques.

The second factor driving the changes was the involvement in the survey or surveys of three different national government agencies: the Ministry of Justice, the Ministry of the Interior and Statistics Netherlands (CBS), and a changing coalition of the larger cities. All these agencies acted at times as co-funders of the survey with a stake in its contents and costs.

The third reason for the changes in the Dutch survey is common to all survey research, namely, the ever-changing nature of possibilities, costs, and benefits of various modes of data collection. As known, state-of-the-art modes of drawing samples and modes of interviewing have limited time spans. A telling example is the rise and decline of computer-assisted telephone interviewing. This technique offered a window of opportunity when fixed telephone penetration surpassed 70 % around 1985 in many Western countries, including the Netherlands, but has run into severe problems 20 years later due to the rise of mobile phones. The Dutch survey has since 1973 moved from postal questionnaires, piloted by Fiselier, to face-to-face interviewing, computer-assisted telephone interviewing, and finally mixed-mode interviewing, comprising web-based interviewing, postal questionnaires, and telephone interviewing.

Obviously, these changes in content and methodology, unavoidable as they may have been for either policy or technical reasons, have compromised the comparability of the results over time. It is well established that even minor changes in any aspect of the methodology can have a serious impact on the results (Van Dijk et al. 2010). In recognition of the known sensitivity of the survey to even minor changes, efforts have consistently been made to monitor such effects and to correct results for them. To support such analyses, existing surveys have sometimes solely for this purpose been repeated with smaller samples simultaneously with a newly designed survey.

Next to the victimization surveys among citizens, several business surveys have been conducted over the years, starting in 1977 with a limited retail business survey on shoplifting by Statistics Netherlands. In 2004, a standardized business survey was launched with has been repeated regularly since (MCB: Monitor of Crime in the Business Sector, *Monitor Criminaliteit Bedrijfsleven*). In this survey, information is collected on the victimization of businesses by burglary, theft, bribe taking, destruction of property, and violent crimes.

Characteristics and Methodology of the Surveys

Table 1 presents an overview of all Dutch surveys and the ICVS with their main characteristics, such as period covered, key organizing agency, mode of data collection, and sample size. Except for the two business surveys just mentioned, all surveys relate to victimization of individuals or households. To explain the characteristic of the surveys in more detail, a short description of each of the Dutch surveys is given in a commentary to the table.

Table 1 shows that there has been a distinct shift away from expensive personal, face-to-face interviewing towards the cheaper mode of telephone interviewing. In recent years, a shift has been made to the even cheaper mode of collecting data via self-completed Internet questionnaires. Another development is the incremental growth of the sample size, resulting from the policy need to monitor the performance of regional or local police units. Not reflected in the table is the changing content of the questionnaires with increasing numbers of items on police performance and perceptions of safety.

WODC Victim Survey (1973–1980)

The WODC Victim Survey (*WODC Slachtofferenquête*) was the first regular victim survey in the Netherlands. It was a direct answer to the political need to measure the real extent and trends of crime independent from police administrations of recorded crime in a period of rising concerns about crime.

The WODC surveys were carried out by a commercial polling company, NIPO, associated with Gallup/USA. Samples were drawn from a list of addresses maintained by the agency responsible for collecting radio and television license fees. There was some oversampling for the four largest cities in the Netherlands which allowed for the publication of city rates. A personal interview with a paper questionnaire was held with one person per household in the sample, older than 16 (or for the surveys before 1978, 18 years or older). The net sample varied somewhat but increased to 6,000 in 1980, with a response rate of 42 %. Victimization in the last calendar year was asked for (the interviews were all held in the first months of the following year).

Different crime types were asked for, mainly property crimes like burglary, bicycle or moped theft, and theft from cars. Starting in 1975,

History of the Dutch Crime Victimization Survey(s), Table 1 The Dutch crime victimization surveys

Survey name	Period (reporting years)	Responsible organization(s)[a]	Modes	Net sample size
Persons/household surveys				
WODC Victim Survey	1973–1980; yearly	MoJ[b]	Personal interview	6,000
ESM	1980–1984; yearly 1986–1992; every second year	CBS	Personal interview	5,000–10,000
ERV	1992–1996; yearly	CBS	Personal interview	5,000
PMB	1992–1998; every second year 2000–2004; yearly	MoJ, MoIA	Telephone interview	25,000–75,000
POLS	1997–2004; yearly	CBS	Personal interview	5,000–10,000
VMR	2004–2010; yearly	MoJ, MoIA, CBS	Mainly telephone interview; some personal interviews	5,000 (2004) 20,000 (2005–2007) 6,000 (2008–2010)
IVM	2008–2011; yearly	MoJ, MoIA, CBS, local police, city councils	Both Internet questionnaire and telephone interview; some personal interviews	About 40,000–220,000
VM	2012 onwards; yearly	MSJ, CBS, local police, city councils	Mainly Internet questionnaire	At least 90,000
Business surveys				
Shoplifting statistics (retail business)	1977–1983	CBS	Paper and pencil questionnaire	300–400 (1977–1981) 3,000 (1982–1983)
MCB	2004 onwards; yearly	MoJ, MoIA	Telephone interview	30,000–38,000
International surveys				
ICVS	1988–2005, irregular	MOJ	Telephone interview	2,000

[a] *MoJ* Ministry of Justice, *MoIA* Ministry of Interior Affairs, *CBS* Statistics Netherland, *MSJ* Ministry of Safety and Justice

[b] The research department of the Ministry of Justice WODC was responsible for this survey

questions on some violent crimes and vandalism were added as well. Also follow-up questions to identified victims concerning reporting behavior were always part of the questionnaire (including reasons for not reporting) and some other information on the crime such as damages, injuries, and weapon use. Questions on the use of prevention measures and some background characteristics of the respondent were also included. Feelings of safety were not covered by this survey.

ESM (1980–1992)

In 1980, it was decided that a regular victim survey should become a part of the program of Statistics Netherlands, responsible for providing statistics. This was the start of the ESM (Crime Victim Survey, *Enquête Slachtoffers Misdrijven*) which was conducted yearly in the period 1981–1985 and thereafter every second year.

The sampling method changed somewhat over the years. In the first years, it was primarily a person-based sample (persons older than 15), but for some cities, an address-based sample was taken. After 1982, the sample was completely taken from an address register. Data were collected with personal interviews using a paper questionnaire. The net sample was about 10,000 (1980–1986); this was reduced to about 5,000 in the last three surveys for reasons of costs (1988, 1990, 1992). As with the WODC Victim Survey, victimization in the last calendar year was asked for (the interviews were held in the first months of the following year).

The questionnaire was very similar to the WODC Victim Survey, covering the same types of crime and other subjects. Also, questions on neighborhood characteristics, prevention behavior, going out, and safety feelings were asked in the ESM. Nevertheless, when comparing the results of the WODC Victim Survey and the ESM, clear differences surfaced compromising any straightforward analysis of the longer time trends.

ERV (1992–1996)

In 1992, Statistics Netherlands decided to cut costs by integrating the victim survey in their ongoing program of population surveys. This, together with some technical improvements, led to considerable changes in the survey's design, resulting into a really new survey, the ERV (Justice and Security Survey, *Enquête Rechtsbescherming en Veiligheid*) which was held yearly in the period 1992–1996.

As with the ESM, the sample was based on an address register. If possible, two persons (15 years or older) living at the same address were interviewed, again through a personal interview. The results were directly introduced in a computer. The net sample was about 5,000 persons from about 3,500 households. The survey was held continuously over the year, and victimization in the preceding 12 months was asked for.

Although many crime types asked for in the survey were the same as in the ESM (burglary, bicycle theft, theft from a car, etc.), particularly the questions on violent crimes were changed. Some new questions on feelings of (un)safety were introduced.

PMB (1992–2004)

In 1992, another survey was started parallel to the ESM/ERV surveys of Statistics Netherlands: the PMB (Police Monitor Public, *Politiemonitor Bevolking*). At that time, a reorganization took place within the Dutch police, abolishing both the existing national state police and the many municipal police forces. Instead, the Netherlands was divided into 25 police regions with their own forces. Both the Ministry of Justice and the Ministry of Interior Affairs felt that information on victimization, safety feelings, and public opinion on the police was needed at the level of police regions. This was not possible with the limited sample sizes of the surveys from Statistics Netherlands. In addition, the questionnaire was judged to be insufficiently tailored to the policy and planning needs of both ministries, jointly responsible for policing. For this reason, a new survey was started. The PMB was held in the period 1992–2004, first every second year and after 2000 each year.

Samples were taken from a fixed-line telephone directory. Net response was at least 1,000 respondents for each of the 25 police regions. Individual police regions had the opportunity to buy in with additional samples in order to obtain information on an even lower level than the region (city or even district or neighborhood level). This resulted in a net sample varying from 25,000 (in 2002) to 90,000 (in 2003). Apart from 2002, the net sample was always at least 50,000. Interviewing was done by telephone. Fieldwork was done in the first months of the year, and victimization in the previous calendar year was asked for.

The questionnaire was not too different from the Statistics Netherlands questionnaire regarding victimization. More or less the same types of crime were covered. However, more emphasis was laid on vehicle-related crimes for all household members. Apart from victimization, a large part of the questionnaire was devoted to safety feelings, preventive measures, and opinions on police performance. The results were not easily comparable with the other surveys. One of the main reasons was that victimization rates of vehicle-related crime were not relative to the population but – for example, in the case of bicycle theft – relative to the bicycle owners. Another reason was the different interview mode (telephone vs. face-to-face interviews).

POLS (1997–2004)

In 1997, the ERV survey from Statistics Netherlands was fully incorporated in a more general survey, the POLS (permanent survey on living conditions, *Permanent Onderzoek Leef Situatie*). The POLS survey was a large modular survey with modules on different aspects of social living conditions. Victimization and feelings of safety were asked in one or more modules within this survey.

For the whole survey, a sample was taken from the general Dutch persons register (GBA, *Gemeentelijke Basisadministratie*) consisting of all persons living in the Netherlands (so not necessarily restricted to the Dutch nationality). The net response for the whole survey was 50,000 but only 10,000 for the modules including victimization and 5,000 for the module including more detailed questions on safety feelings. As with the ERV, the survey was held continuously, through personal interviews. Respondents were asked about victimization in the preceding 12 months.

The questionnaire was identical to the ERV questionnaire, with some minor changes introduced in 1999. Because only the sampling method was different, the results were more or less comparable to those of the ERV.

VMR (2004–2010)

Due to budget cuts, Statistics Netherlands – although realizing that informing the general public on safety issues was one of their responsibilities – could no longer maintain an expensive large survey based on personal interviewing. At the same time, both the Ministry of Justice and the Ministry of the Interior relied more and more on the results of these surveys (POLS and PMB) for the purpose of planning and evaluation. Another relevant factor was that many cities, most notably Rotterdam, had started to launch their own “safety monitors”, monitoring safety issues and living conditions at a local level. Driven by economical needs, both ministries and Statistics Netherlands decided to launch a joint national and police regional survey (the VeiligheidsMonitor Rijk), with the option to expand into a survey catering to local information needs as well.

As with the POLS survey, sampling was based on a person’s register (GBA). For the majority of persons thus selected, a phone number (fixed or mobile) could be found, and interviewing was done by telephone. The persons with no known telephone number were interviewed personally at home. The net response in 2004 was 5,000 (basically this was a pilot survey for comparison with both the PMB and POLS), about 20,000 in 2005–2007, and about 6,000 in later years (again mainly to compare the VMR with the subsequent survey). Like the PMB (but different from POLS), fieldwork was done in the first months of the year, and victimization in the previous calendar year was asked for.

The questionnaire was redesigned, but crime types asked for and other information were not too different from POLS and PMB. However, the changes in the questionnaire and in particular the other methodological changes caused a considerable break in the series.

IVM (2008–2011)

As was already foreseen at the start of the VMR, steps were made to fully integrate the VMR with the running local safety monitors. Both for financial reasons and to have uniform national and local safety indicators, the methodology of all surveys was standardized. The resulting IVM (Integrated Safety Monitor, *Integrale Veiligheidsmonitor*) started in 2008 as a joint venture between the Ministry of Justice, the Ministry of Interior Affairs, Statistics Netherlands, local police organizations, and city councils. The IVM was held every year until 2011.

The organization and the methodology of the IVM survey are rather complicated. The basic idea is that there is a nationwide sample resulting in a net response of at least 18,750 (i.e., at least 750 respondents in each of the 25 police regions) each year. The fieldwork for this sample is carried out by Statistics Netherlands. Besides, all local parties can add their local samples employing local polling companies. In practice, this resulted in total net samples of between 40,000 (2010) and 220,000 (2011). Sampling itself is done centrally, by Statistics Netherlands, in the same way as with the VMR (through the GBA). The persons in the sample are approached with a letter and are invited to fill in the questionnaire either through the Internet or on paper. Remaining nonrespondents are approached again and if possible interviewed by telephone (as in the VMR for most persons in the sample, a telephone number is known). If a telephone number is not found, respondents can, as a final resort, be interviewed personally at home, but this has rarely happened. The fieldwork takes place in the period September–December; victimization in the previous 12 months is asked about. To ensure uniformity in the methodology – in particular, the fieldwork – the IVM is coordinated by a central agency.

The questionnaire had to be adjusted considerably to serve the needs of all parties involved. To improve comparability of national and local data, the IVM questionnaire had a standardized modular design with some obligatory modules (such as the victimization module) and some optional modules (as on police contacts and attitudes towards the police). To improve comparability, the order in which the modules were placed was fixed; in particular, all obligatory modules were positioned at the beginning of the questionnaire. In the national and police regional samples, all modules were asked; local participants were free to add one or more optional modules (in the prescribed order) or not or to add some additional "free questions" at the end of the questionnaire. In the 3 years, 2008–2010 the VMR was held parallel to the IVM in order to analyze the differences between both surveys and repair the breaks in the series.

VM (2012 Onwards)

In the 4 years, during which the IVM was conducted, two problems became apparent. Firstly, the IVM is a relatively expensive survey because of the large number of interviews by telephone. And secondly, the large variation in the sample sizes caused methodological problems because the proportion of different interview modes differed over the years. Weighting of the results could not fully compensate for these differences as became evident from comparisons with the results of the VMR. Therefore, the decision was taken to redesign the survey in order to correct and stabilize its results: the VM (Safety Monitor, Veiligheids Monitor).

To prepare for the new design, a large-scale experiment was conducted with random assignment of respondents to four different modes (web, telephone, face-to-face, and paper). The experiment aimed to disentangle mode-specific selection and measurement bias in the crime survey (Buelens et al. 2012). The results will be used to improve a future mixed-mode approach. Although details are still under discussion, the central idea of the VM is that there will be a large net sample size (about 90,000), almost exclusively using web-based self-completion. Fieldwork is partly done by Statistics Netherlands and partly by one selected commercial company. To reduce costs further, the questionnaire is made somewhat shorter, although more or less the same crimes are asked for in the victimization module as before. This VM will be implemented in the fall of 2012.

MCB (2004 Onwards)

In addition to the surveys where victimization and other safety issues were asked from citizens and after a limited survey on shoplifting in 1977–1983, starting in 2004 a business survey was carried out. This survey is organized by the ministry of Justice and the ministry of Interior Affairs. The net sample size of this survey was 30,000 in 2010. The survey has been carried out in the following five sectors of the business community: construction, retail, hotel and restaurant, transport, and commercial services. Interviewing is done by telephone.

Victimization by burglary, theft, bribe seeking, vandalism, and violence is included. Also total losses and damage by crime, the use preventive measures, reporting behavior, and experience with the police are covered.

Some Results, 1973–2011

Due to the different methodologies used, it is hazardous to directly compare victimization levels (and other results) from the different surveys. Straightforward comparisons do indeed often reveal striking differences between the results. Having said that, interesting findings regarding trends in victimization can be obtained for a selection of crimes and for some other

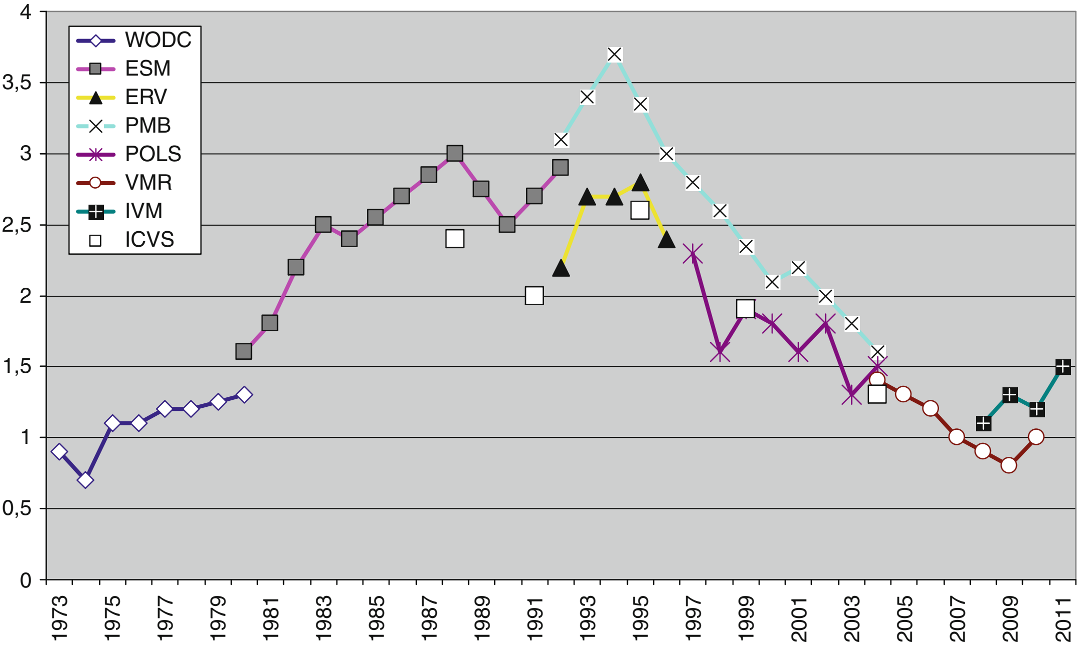

History of the Dutch Crime Victimization Survey(s), Fig. 1 Victimization rates for burglary, 1973–2011. Percentage of persons having been victim of burglary once or more in the preceding year

indicators. Trends in victimization rates – defined as the percentage of respondents that has been victim at least once – regarding two specific crime types will be presented here, i.e., burglary and bicycle theft. Both types of crime were included in the questionnaires of all surveys using more or less the same definitions. Next some trends in rates of victimization by violent crime and total victimization rates will be presented as well. Methodological problems will be more visible here because questions about these types of victimization were not formulated in the same way in the different surveys. Also we will look at trends in feelings of unsafety, starting in 1992.

Burglary

Data on burglary victimization are available for the period 1973 up to 2011. In several years, more than one survey has been conducted simultaneously. The data collected by the various surveys on victimization by burglary is surprisingly consistent across the eight different surveys (see Fig. 1). Level differences can be observed, but they are not too large, certainly not when realizing that the *statistical* 90 % confidence interval for the samples used and percentages found for burglary will typically be in the order of 0.5 %. Most of the level differences seen here are smaller than this confidence interval.

There is a clear upward trend from 1973 until about 1993–1995. Thereafter an almost continuous downward trend can be seen until 2008–2009. After that, it remains to be seen if the victimization rate is stabilizing or starting a new upward trend.

Bicycle Theft

As can be seen in Fig. 2, the level differences for bicycle theft are larger than those for burglary. Apparently the wording of the question, sampling method, and interviewing mode have had more influence on the measurement of this crime than was the case with burglary. For example, in some surveys such as the ICVS, respondents are asked whether any of the bicycles of the household have been stolen, whereas other surveys ask about the theft of one’s own bicycle only.

Although it is less clear than with burglary, the trends over the last 40 years are not too different.

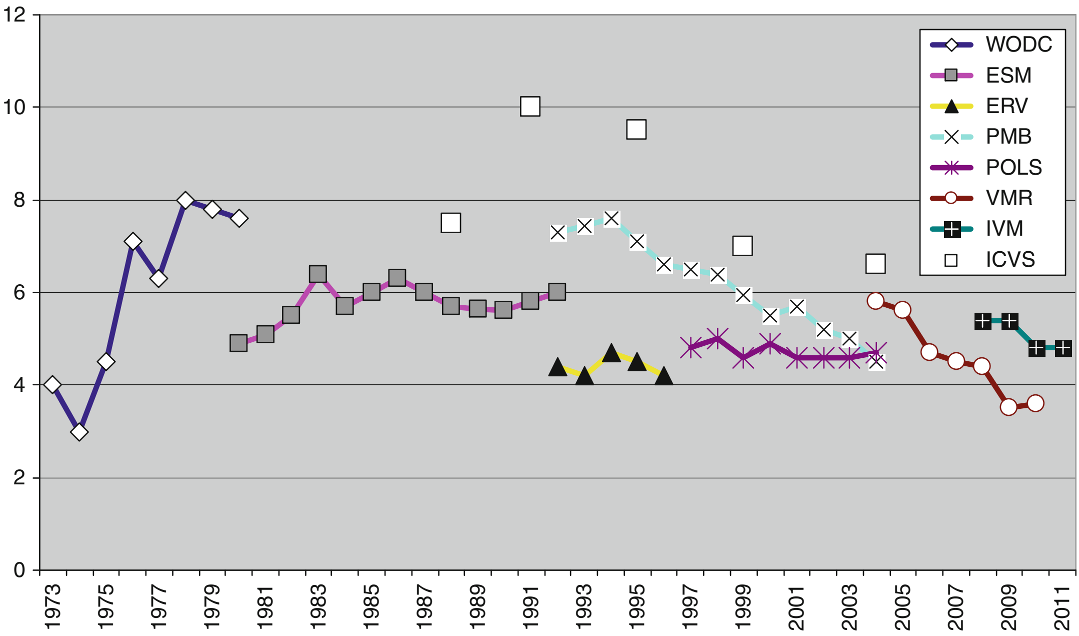

History of the Dutch Crime Victimization Survey(s), Fig. 2 Victimization rates for bicycle theft, 1973–2011. Percentage of persons having been victim of bicycle theft once or more in the preceding year

An increase can be seen until about 1983; however, it is not obvious what the trend was in the next 10 years until about 1993. After 1993, almost every survey shows a downward trend, just like with burglary.

Violent Crime

It is very problematic to look at trends in victimization of violent crime for the whole period 1973–2011 due to the varying definitions employed. In the first WODC survey and the subsequent ESM, the question on assaults/threats was limited to incidents in the public domain. In the later surveys, these questions also included incidents in the private sphere. In Fig. 3, the results from the WODC and ESM surveys are represented by a dotted line describing trends in assaults/threats on the street or in other public places. Therefore, the level differences between these two first surveys and the later surveys are largely caused by the use of a broader definition in the latter. For the other surveys, level differences seem to be reasonably small.

After an initial increase in the 1970s, as shown by the WODC survey, victimization for violent crimes seems to have remained stable until about 1993. Hereafter an increase can be seen which continued until about 2002 (with the ERV figure for 1996 as a notable exception). From 2002 onwards, all surveys show an almost continuous decrease.

Total Crime

Victimization by total crime is defined in the Dutch surveys, as in the ICVS, as the rate of persons victimized by any of the types of crime included in the questionnaire. The WODC survey did initially cover a smaller range of types of crime than the later surveys. As a proxy of the total crime rate, the sum of the limited number of crimes in the WODC survey (at least until 1978) is taken. The level of victimization by total crime is clearly dependent on the list of crimes covered. For example, the inclusion or noninclusion of items on various types of vandalism (destruction of property) can make a big difference in the total rate. Considerable level differences can indeed be seen, in particular with the PMB which used the most comprehensive list of crimes (Fig. 4).

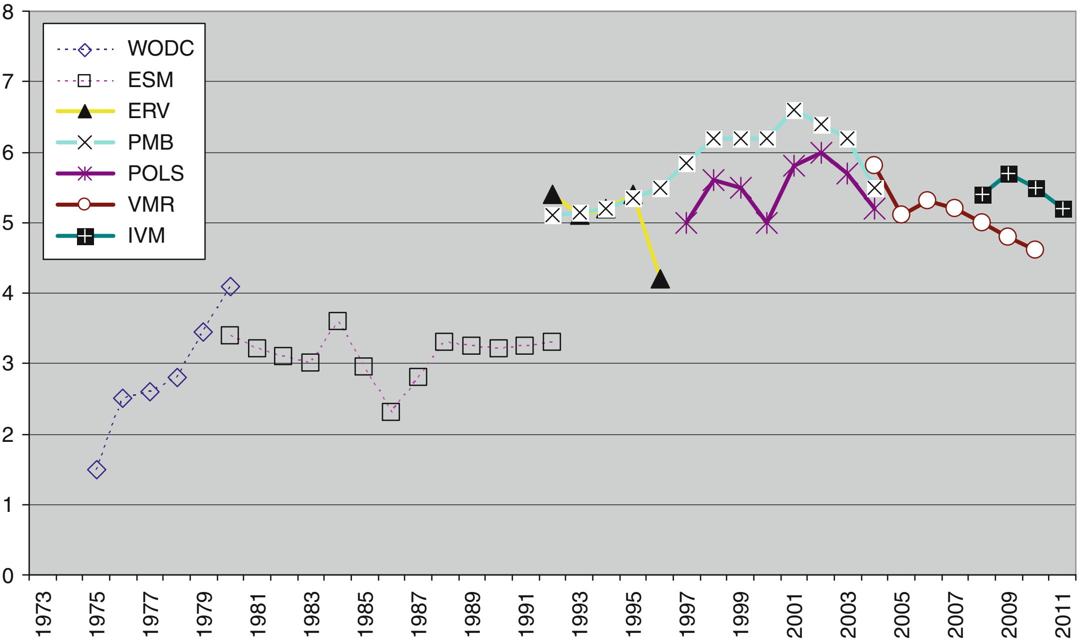

History of the Dutch Crime Victimization Survey(s), Fig. 3 Victimization rates for violent crime, 1975–2011. Percentage of persons having been victim of an assault/threat (for WODC and ESM only victim of incidents in public places) once or more in the preceding year

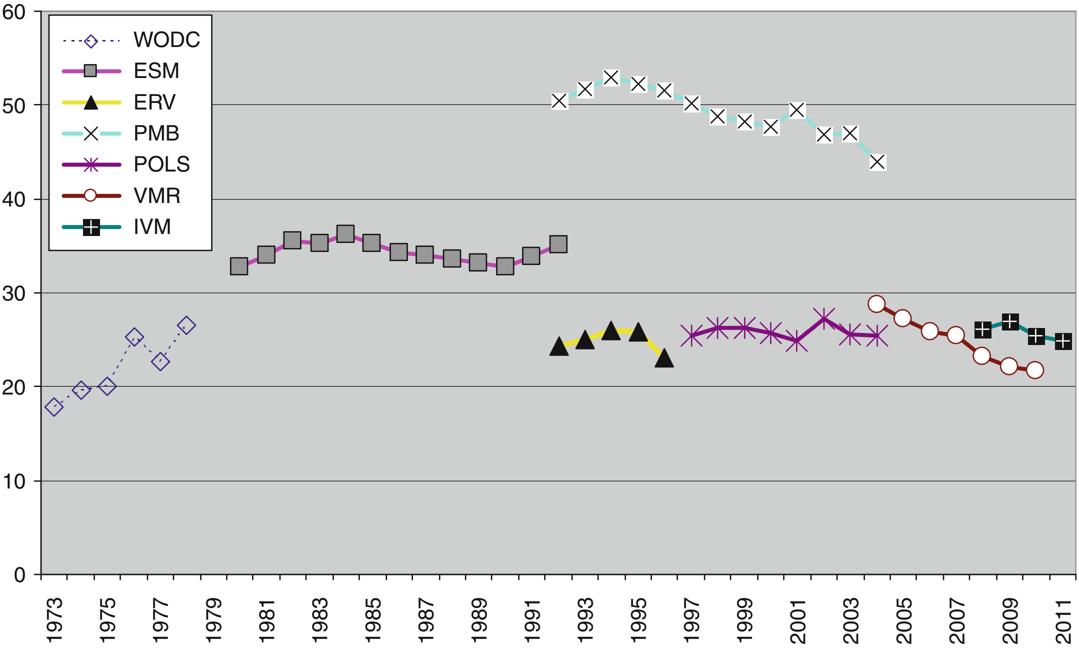

History of the Dutch Crime Victimization Survey(s), Fig. 4 Trends in victimization by any crime, 1973–2011. Percentage of persons having been victim of any crime (for WODC victim of a few crimes only) once or more in the year reported on

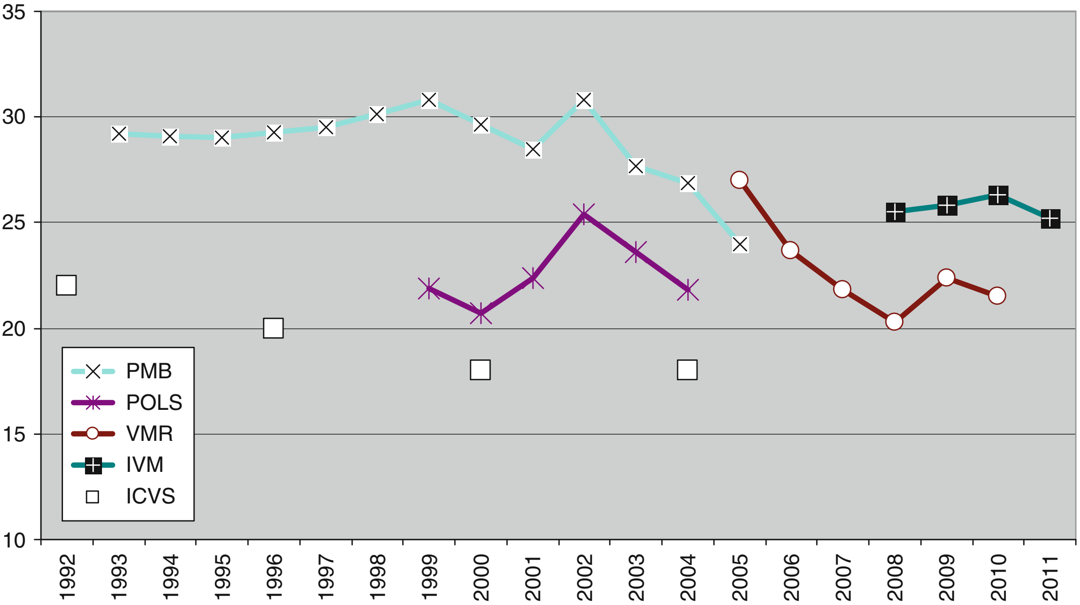

History of the Dutch Crime Victimization Survey(s), Fig. 5 Feelings of unsafety, 1992–2011. Percentage of persons feeling unsafe sometimes (for ICVS feeling unsafe after dark on the street)

The results of the first two surveys suggest an upward trend in total crime between 1973 and 1985. From 1994, a downward trend can be seen, particularly in the PMB, VMR, and IVM surveys.

Feelings of Safety

Starting in 1993 (or in the ICVS in 1992), respondents were asked how often they feel unsafe. In Fig. 5, the percentage of respondents “feeling unsafe sometimes” are given. Again, some clear-cut level differences can be observed which point at effects of the methodologies used. The lower rates found in the ICVS can be explained by the use of a different question: in the ICVS, unsafety feelings were restricted to experiences on the street and after dark.

However, trends appear to be fairly similar across surveys (e.g., looking at results of PMB and POLS).

By and large, feelings of unsafety have declined since 1992 with the exception of a brief spike around the turn of the century. This could well have been caused by the 11/9/01 attacks and the murder in May 2002 of a high-profile Dutch politician.

Conclusion

The Dutch national crime victimization surveys belong to the longest running and largest exercises in the field of crime surveying. The considerable investments made in the Netherlands in this line of research must be understood against the background of a continuous political debate about the problems of crime since the early 1970s and a strong infrastructure for policy-oriented criminological research. Over the years, the WODC (the Research and Documentation Centre of the Ministry of Justice and since 2010 the Ministry of Security and Justice) has been in a position to secure the survey’s continuation as its guardian angel. The popularity of the survey has been much enhanced by the growing political interest in its potential to provide performance indicators for local police forces and municipal security programs since the 1990s. The survey’s successes, though, seem to have come at a cost. The involvement of different stakeholders and executing agencies has impacted on organizational setups and methodologies. For this reason, the methodology of the Dutch survey seems to have been even more in flux than similar national surveys such as the NVCS and the BCS. The choice for web-based interviewing in future rounds of the survey seems propelled by cost considerations imposed by the use of relatively large samples required for the calculation of local rates. Although the mixed-mode approach is based on experimental studies, its rapid introduction may give rise to renewed debates about the survey's methodology in the coming years.

The available datasets of the Dutch surveys allow for attempts at longtime trend analyses. Results presented here provide unique survey-based evidence of a rise in different types of crime in at least one European country in the 1970s. Over the years, the datasets have been used for numerous secondary analyses by researchers (Wittebrood and Junger 2002; Van Wilsem 2003; Goudriaan et al. 2006; Vollaard and Van Ours 2011). The surveys have been one of the drivers of the current boom in Dutch criminology.

Related Entries

- ► BCS
- ► Development of the UCR and the NCVS
- ► History of the Statistics of Crime and Criminal Justice
- ► International Crime Victimization Survey

Recommended Reading and References

Buelens B, Van der Laan J, Schouten B, den Brakel V, Burger J, Klausch T (2012) Disentangling mode-specific selection and measurement bias in social surveys, discussion paper. Statistics Netherlands, Heerlen

van Dijk JJM, Steinmetz CHD (1980) The RDC victim surveys, 1974–1979. Ministry of Justice, The Hague

van Dijk JJM(2009) Approximating the truth about crime; comparing crime data based on population surveys with police figures of recorded crimes; synthetic report, CRIMPREV, Groupe Europeen de recherches sur les Normativites (GERN/CNRS)

van Dijk JJM, Mayhew P, Van Kesteren J, Aebi M, Linde A (2010) Final report on the study on crime victimisation (commissioned by Eurostat). INTERVICT, University of Tilburg, Tilburg

Fiselier JPS (1978) Slachtoffers van Delicten; een onderzoek naar verborgen criminaliteit. Ars Aequi, Utrecht

Goudriaan H, Wittebrood K, Nieuwbeerta P (2006) Neighbourhood characteristics and reporting crime. Effects of social cohesion, confidence in police effectiveness, and socio-economic disadvantage. Br J Criminol 46:719–742

Vollaard B, Van Ours J (2011) Does regulation of built-in security reduce crime? Evidence from a natural experiment. Econ J 121(May):485–504

Wilsem van J (2003) Crime and context: the Impact of individual, neighborhood, city and country characteristics on victimization. Katholieke Universiteit Nijmegen, Nijmegen

Wittebrood K (2008) Comparing crime data in The Netherlands, paper CRIMPREV WP 7, GERN/CNRS

Wittebrood K, Junger K (2002) Trends in violent crime: a comparison between police statistics and victimization surveys. Soc Indic Res 2(59):153–173

History of the Police Profession

Herbert Reinke[1] and Margo De Koster[2]
[1]Bergische Universität Wuppertal,
Wuppertal, Germany
[2]Vrije Universiteit Brussel, VU University,
Brussels, Belgium

Overview

Policing as a professional métier emerged in many European countries and in Northern America throughout the nineteenth century, although in many countries gradually only. The move toward professional policing came from two sides: The authorities governing the police (the state, municipalities) strived for a more efficient and above all for a more accountable law-and-order maintenance, while members of the (uniformed) police perceived and used the enhancement of professionalism within the police a strategy for promoting career prospects, pay, and status. Police schools and police unions were part of this development. The strive for professionalism remained limited throughout most of the nineteenth century, because most of the uniformed policemen were rather a sort of general factotum than professional agents for law-and-order maintenance. The late nineteenth century saw in many countries the advent of specialized police branches, specializing on crime investigation, detection, and control (crime detections units, when rising fears of crime and disorder in the cities enhanced the demands for professional crime control). The social defense modus of crime control during the decades before and after the turn of the nineteenth to the twentieth century, focusing on perpetrators threatening "property," served a very specific notion of crime. In addition with the high level of public prominence allotted to crime investigation work, the (societal) status and prestige of the work of uniformed policemen decreased, whereby crime investigation acquired the status of "real" police work, although, seen from a quantitative perspective, crime detection is a minor part of overall police activity.

Key Issues and Concepts

The police understand themselves as a profession and the members of the police see themselves as professionals. Since the emergence of the concept "profession" in the social sciences, the literature has come forward with a number of theoretical approaches, proposing characteristics for describing "profession" and "professionals." There is a narrower conceptualization, which sees university trained occupations as professionals, but a wider concept is nowadays used quite often as well, which also includes institutions such as the police among the professions and includes the police personnel among the professionals. The characteristics describing professions and professionals allow for variation, but a number of core elements have come to stand out, depending on the respective theoretical approach, which allows summarizing definitions of the concept (Reiner 1978). There is a functionalist approach, which sees professions as generally characterized by the exercise of very particular skills, acquired through very specific training processes. This training and these skills generally involve the acquisition of corresponding specialized knowledge, which only the members of the respective profession have at their disposal. Within the functionalist context of this approach, professions and professionals are serving basic societal needs and values, appreciated by society. Another characteristic of "profession", which the literature refers to, concerns the existence of specific interest politics, which professions are pursuing. These interest politics are often indicated by the establishment of organizational structures, serving the promotion of the respective profession (income, status, etc.). In this context, unions are not only seen as representations of workers but of professions as well. Promotion means as well, that these organizational structures quite often enhance a sort of exclusiveness of the profession (skills, knowledge), which is strived at to clearly set it apart from other professions. The literature labels this understanding of "profession" and "professionals" as the control approach, seeing professionalism as a strategy for controlling the labor market. The control approach avoids the ideological layer to be found in the functionalist understanding of "professions" and "professionals," which has been criticized for not giving answers to questions about those within society, which are able to define and to enforce social need and values.

Critics have argued that applying the concept of "profession" to the police implies a basic contradiction. Due to the police's strict hierarchical structures and due to (at least throughout the nineteenth century) its military features, it is or has been rather a "bureaucracy" than a "profession" (Reiner 1978, 2000; Johansen 2005). But applying elements of the functionalist understanding of professionalization, which include the emergence of professions as a follow-up to societal, economic, and political changes and demand, allows for asking questions about the professionalization of the police during the nineteenth and the early twentieth centuries: Controlling rioting crowds and controlling disorder and crime turned out to be the challenges the police was to meet during this period.

The explosive growth many European cities experienced, and its problematic consequences, as well as upsurges of social and political unrest during the decades of the nineteenth century, led to heightened anxieties about crime and disorder.

The conditions under which people lived in the cities were seen as prerequisites per se for political unrest, immorality, and flourishing crime. These fears led to demands for police efficiency, followed by intense political and public debates about the organization and functioning of the police, and accomponied by considerable attempts all across Europe to professionalize the police.

Toward Professional Policing: Uniformed Polices in Europe During the Nineteenth Century

It was not before the last decade of the nineteenth century that professionalization became a significant concern among police forces in Europe. It not only became a concern for the state and for the municipal authorities directing police forces, it became a concern for the policemen themselves as well. Being a policeman in uniform meant in most European countries throughout the nineteenth century a combination of long-working shifts, a low pay, no training for the job, and low prestige. Most men in the uniformed police had a military background, even in those police forces which were supposed to be a civil police such as the English police (Emsley 1991). Being trained for the job was not part of the uniformed patrolman's work experience. Members of the Gendarmerie forces with their military background did have a military training, but lacked specific training for their policing duties as well. For the patrolmen in uniform, who emerged in European cities throughout the first half of the nineteenth century, wearing a uniform in a proper way, and walking the beat accordingly were in some way already the essentials to be taught to them. Until the First World War, walking the beat or being on patrol in the countryside remained for the patrolmen their core police activity.

European countries, although different in terms of political systems, political structures, and cultures during the nineteenth century, showed *grosso modo* similar lines concerning approaches for improvements of the recruitment and above all the training of its police personnel. During the latter years of the nineteenth century, these attempts accelerated, not only by establishing schools for improving the training of the uniformed police, but also by attempts and initiatives of the police officers themselves to improve their situation and their status. France, Belgium, England, Prussia, and the Netherlands (Meershoek 2007; van der Wal 2007) provide examples of these developments.

In France, it was the Third Republic which began paying attention to the recruitment and training of the police by taking first steps toward a professionalization of the police service. There were a number of reasons, why the new republican regime gradually began enhancing a professionalization of the police. Firstly, within the political system and the political culture of the Third Republic, an electorate based on universal suffrage could hardly be expected to remain indifferent to the standards and achievements of the police service; secondly, a republican government could not simply take over the police bequeathed to it by the authoritarian regimes which preceded it, which it had always denounced as dictatorial prior to its own accession to power; and thirdly, the French police, at least the one in Paris and in other major cities of France became confronted with the consequences of urbanization and industrialization, which complicated the policeman's lot significantly, raising the issue of how the police could adapt to these changes (Berlière and Lévy 2011).

This move towards a professionalization met difficulties due to the complex and conflict-loaded structures of the French police. The French police since the first decades of the nineteenth century consisted of a number of different polices. One of these polices, the *sûreté générale*, was rather an institutional roof, expressing the state's aspiration to cover police-related matters as "state." In reality, the French government ran its own police, besides the Gendarmerie, throughout the nineteenth century in Paris and in two other major cities only (Lyon, Marseille). Quantitatively, much more significant were the municipal and local polices. Within these structures, the criminal or judiciary police, responsible for crime investigation, was dependent on four police authorities. This institutional layout was in itself already complex, but it became even more complex, as the Third Republic never succeeded in putting forward a clear view as to who was responsible for the powers of the municipal police – torn as it was between a tradition that these police powers should be seen as emanating from the local authority and by centralizing tendencies. As a result, the municipal police found themselves dependent on the Ministry of the Interior in some matters, and on the local authority in others. The Ministry of Interior was, for example, responsible for the appointment of the *Commissaire de Police* (obligatory post in towns of over 5,000 inhabitants), and for fixing the size of the police force and that of the municipal police budget in county towns of 40,000 or more inhabitants. Towns of over 5,000 inhabitants thus had their own police (in towns of less, the *gendarmerie* took on the functions of the police), who theoretically came under the authority of a commissaire appointed by the Ministry of the Interior, but who were paid by the municipality. Apart from the post of commissaire, all the other municipal police appointments were made by the mayor, who also decided on their salary levels. They were recruited, paid, disciplined, and punished by the local authority. This system had several unfortunate consequences, among them conflicts between mayors and commissaires as the commissaire was placed in an impossible position. It required a great deal of balancing to keep the police between state and municipal authorities going, while his situation became totally untenable, if the municipality was hostile to the central government of which the commissaire was the representative (Berlière 1996).

The state of the municipal police was very insufficient: In towns with less than 40,000 inhabitants, the mayor and municipal council alone decided on how large a police force was needed, and in order to save money, kept these as small as possible. In certain towns, the Commissaire de Police – perhaps with one or two assistants – might be the only one acting as policeman. The standard of men and means in many municipal police forces was ridiculously low, and only Lyon and Marseille with its state polices had a more developed police system. The rules that existed under previous regimes were hardly altered during the period of the Third Republic. Instead, it was the military laws of 1872, 1889, and 1905 which had overriding effect on the recruitment of police: In order to encourage a further 5-year service after the period of conscription, a civilian post was ensured to men after they had completed their military service. Thus, for a long time, the great majority of French policemen were former soldiers. Only a fifth of the police employed at the turn of the nineteenth to the twentieth century were civil candidates. The way in which candidates were assigned within the police service to different types of posts was completely random, as candidates were assigned to places and functions according to vacancies and not according to their capabilities (Berlière and Lévy 2011).

The Third Republic thus started its police regime with a police that lacked professional training completely, but the last two decades of the nineteenth century and the years until the beginning of the First World War were marked by efforts to improve training patterns. Recruitment patterns and conditions of employment changed partially as well during these decades, leading to higher wage levels; to a system of allowances for clothing, housing, retirement, and children; to a month's holiday per annum; and for a retirement pension. Since it rather quickly proved to be impossible to fundamentally change and improve patterns of recruitment, attention from the 1880s onward focused on police training. A school for the municipal police was established in Paris in 1883. The courses there included dictation, instruction on how to deal with practical day to day police tasks, how to telephone and send telegrams, memorization of the streets of one's beat, the rudiments of first aid, and how to disperse a crowd, and instructions on how the police were to treat the public, and on how they were themselves to behave. The setting up of this school demonstrated a real desire to improve the level, training, and professional qualification of the police, but it remained for a decade and half the only institutionalized full time training facility within the French police. During the years before the First World War, particularly the Paris police started to upgrade its technical capacities and to enhance the professional specialization of the force by establishing a training school for all police in May 1914, where training was much longer, including new courses with a obligatory diploma course in descriptive studies for all higher ranks, by linking promotion with higher professional qualifications, and by creating a number of special services, in order to adapt the police to the transformations brought about by accelerating urbanization and industrialization (Berlière and Lévy 2011).

In neighboring Belgium, similar debates were going on and comparable initiatives were developed in the late nineteenth century. Critics of the police were widespread; they also came from within the police itself: Senior police officers openly admitted that both in terms of manpower and quality of their personnel, all municipal police forces had extremely low standards. Improvements turned out to be difficult to realize, however, as each municipality in Belgium had its own police force, financed and managed by the local authorities. As a result, municipal police forces were very different from one another, not just in terms of manpower strength but also in many other respects, such as the recruitment and training of police officers, their wages, equipment and weapons, accommodation and material resources, and the way in which they carried out police duties. The standard of men and means in many municipal police forces was very low, one of the consequences being that, even though so-called municipal autonomy in politics was frequently and fiercely adulated, many local authorities had to call out the army at the slightest excuse to maintain law and order. Only Brussels and Antwerp, the two largest Belgian cities, had more developed municipal police forces. The efficiency of the Belgian municipal police was not only undermined by their limited numbers, but also by the poor qualifications of their personnel, which were the result of inadequate recruitment and a complete lack of any professional training. There was a total lack of professional selection criteria for candidates applying for a post as constable; in general, only physical qualities (meeting the age and height requirements, being physically fit, and looking tough), behavior, and morality were taken into account. The candidate might be subjected to a low level type of examination chiefly concerned with writing and arithmetic, but the results were of little significance for eventual admission to the force. Nominations in the higher ranks, on the other hand, often depended on whether the candidate had personal relations with senior police officers and enjoyed local political support. Unsurprisingly, this recruitment pattern negatively affected the quality of the municipal police. The inability of the police authorities to recruit sufficient suitable or well-qualified personnel is further illustrated by the fact that all the forces were confronted with rapid turnover. This only aggravated problems of manpower strength, because another result was that a large number of police posts were permanently vacant. To make things worse, newly recruited constables were not given any serious training before letting them loose on the city streets. The training of policemen was carried out "on the job" and was oriented along the practical requirements of the police's everyday activities: The police constable had to get acquainted with his precinct and with the things he had to look after when on the beat. A number of "formalities" were briefly taught, such as some basic ideas about the Belgian penal code and the police ordinances that had been issued for the respective city or municipality, etc. In short, "real" professional training did not take place. Finally, professional specialization of the police through the creation of special services was still far off (Keunings 2009).

These problems and shortcomings of the Belgian municipal police did not escape the notice of contemporary observers, police reformers, and responsible politicians, and the whole of the period of 1890–1914 was marked by efforts to expand police forces and professionalize police work. In the major urban centers (Brussels, Antwerp, Ghent, and Liège), the municipal police grew considerably in per capita strength: The overall figures for the police personnel doubled. This increase occurred primarily in the lower ranks: manned with the plain constables; in Antwerp, for example, their size almost tripled between 1883 and 1914. By 1914, the proportion of police to population in the largest Belgian cities was not too far from that in other major European centers. Thanks to a subsidy of the central government, Brussels had even become the "best-policed" European city in terms of manpower strength (Keunings 2009; De Koster 2008).

The growth in per capita strength of Belgium's municipal police forces did result in an improvement in the conditions of service, more particularly in shorter working hours and a greater number of days of leave. In Antwerp, this evolution started off in 1898, when the City mayor announced that the city inhabitants had approved a municipal tax raise for the enlargement of the police force in a referendum, which permitted the recruitment of a hundred extra constables, a reorganization of the force, including new schedules, and the introduction of a "nicer" uniform. At that time, the average constable still had very long and irregular hours on the street, with two shifts a day of about 5 h and 20 min, except after a night shift (day of 4 h and a half). Moreover, police assignments and schedules reveal that policemen were often reassigned from one post to another and always rotated from one shift to another. At the same time, Antwerp policemen could only take 1 day of rest and 3 days of leave a year at the end of the nineteenth century, and lost them in case of punishment or illness. In 1907, however, this last rule was withdrawn and the number of days of leave was increased by four. By 1912, working hours had been reduced to an average of 8 h and 25 min a day. The most definite improvement in the conditions of service before the First World War was realized through the gradual introduction of higher wage levels and a retirement pension, paid by the city (from 1886). It has to be noted, however, that the rise in wage rates achieved by 1914 was considerably accentuated by the fact that initially, at the end of the nineteenth century, wages had simply been ridiculously low. By 1914, the pay of Antwerp constables had come close to that of a concierge, while wage rates in Brussels slightly exceeded those of the lower civilian personnel of the city (De Koster 2008).

Although they were relatively limited, these improvements in the conditions of service had an immediate effect on recruitment. Belgian municipal police forces experienced much less troubles in finding candidates for the jobs: A greater number of applications followed and rates of dismissals and resignations fell, which raised the standard among the recruits. Overall, the municipal police gradually became a more permanent place of employment for a growing number of people and the composition of these forces gradually shifted to a corps of more experienced policemen. However, the adaptation of a growing number of officers to the new institution was accompanied by what the police still considered to be a high departure rate. Scores of young policemen left the municipal police every year to seek their fortune elsewhere, and although those who left were generally soon replaced, the authorities were not always happy with the general quality of the new recruits. Possibly, the police were not able to attract suitable personnel in sufficient numbers to offset the annual wastage and meet the high manpower demands of the force, and had to accept recruits below the prescribed norms (Keunings 2009). Antwerp provides a good example for the subsequent attempts by Belgian municipal authorities and police administrations to improve the policeman's qualifications during the first decade of the twentieth century. In order to obtain better qualified personnel, a "police school" was set up, the cost of which was covered by the city. It was relatively small, however, and few policemen were sent to it. Despite these efforts, there remained a considerable lack of professional training (De Koster 2008).

However, there were other significant signs of police professionalization in Belgium. The turn of the twentieth century saw a formidable increase in the number of police manuals and handbooks being published. A set of standardized examinations was devised, which candidates for promotion were obliged to pass at almost every stage in their career. Promotion to superior ranks was thus made dependent on climbing the rungs of the police hierarchy. In terms of equipment, the municipal police also knew considerable improvements, especially in Brussels, where the police disposed of not only a telegraph (1873), straitjackets (1883), and solid helmets (1897), but from the early twentieth century onward also revolvers (1902), bicycles and the famous white traffic batons (1910), a police dog brigade (1907) – copied from the Ghent police (1899), who was the first in the world to introduce this – and even police cars (1914). A parallel development, spurred to a large extent by the first police unions ("associations," from 1881 onwards), was the creation of specialized units within the largest urban police forces: antigambling brigades, vice squads, traffic police, port police, and criminal investigation departments. All these developments went hand in hand with a militarization of police culture, most police officials being drawn from commissioned army officers (Keunings 2009).

In Prussia, the institutional layout of the police resembled the one in France although major differences existed. During the so-called Prussian reform period at the beginning of the nineteenth century, Prussia established the police as a state prerogative, allowing the Prussian state to establish state polices in Prussian cities, but allowing the Prussian state as well to commission cities with the establishment and the direction of communal or municipal police forces, establishing state police forces in a few cities only. The Prussian state established during the first half of the nineteenth century state polices only in those cities, which were seen as hotspots of political opposition and dissent. While the state polices were on the budgets of the Prussian state, it were the municipalities themselves which had to pay for their municipal police. During the 1860s, the Prussian state withdrew its state polices from a number of cities, leaving it to the municipalities to police them with municipal polices, the reason for this withdrawal being a mixture of budgetary considerations and changed security evaluations by the Prussian government. In these cities, the mayor represented the prerogative by being the official head of the local or municipal police. While the mayor was the head of the municipal administration, elected by the city council, he was at the same time, with respect to his duties as head of the police in the respective city, executing state functions and had thus to obey state orders. Due to these specific patterns of police organization, the Prussian state could theoretically intervene in everyday municipal matters, although in practice, the cities had many possibilities for diverting or impeding such interventions (Funk 1986; Jessen 1991; Reinke 1993; Spencer 1992).

Who could become a member of the police force in Prussia? There are some resemblances to patterns to be observed in France during the nineteenth century. In theory, previous service in the Prussian army up to 12 years was an absolute prerequisite for being recruited into the uniformed Police in Prussia. Thus, the ranks of the uniformed policemen on the beat were supposed to be filled with noncommissioned officers (NCO). Military service as a prerequisite for admission to the police served several functions: The NCO policeman was supposed to represent king and state in the everyday life of the citizen. His superiors expected him to show his derived authority, and, if necessary, to compel compliance to this authority from the public. On the other hand, long training within the military hierarchy was supposed to have made him an obedient servant to his superiors and a reliable instrument for maintaining law and order. But, in reality, this recruitment pattern worked partially only. Although the civil service did have a high social ranking, the position of a uniformed patrolman had a low ranking on this list. Getting a job in the police was not what an army NCO necessarily looked for after years of military service. For many NCOs, the job of a policeman was only a transitory phase on the way to a quieter existence in the civil service than the police service could offer. In Berlin for example, whose royal police was to serve as an example for other polices in Prussia, more than 2,000 policemen left the police force during the 1890s. The Berlin police force totaled about 4,000 men in the middle of the 1890s. A quarter of those who left the police went into other positions within the civil service. Due to these turnover rates, a large number of police posts were permanently vacant (Funk 1986). Not only Berlin, but other Prussian cities also had to cope with this situation (Funk 1986; Roth 1997; Spencer; 1992). During the 1890s and again during the first decade after the turn of the century, the Prussian state administration tried to solve the turnover problem by reducing the military service requirements for employment in the police force. As a result, more men were drawn into the police service. The nine- to twelve-year military service remained nevertheless the idea requirement looked for when conceiving a real Prussian patrolman on the beat. Those policemen who had acquired their post on the basis of the reduced requirements were often considered as some sort of second-class policeman.

Vacancies in the civil service, the police included, were announced and advertized by state and city administrations in public lists and journals. But the police posts offered regularly outnumbered applicants from the NCO's ranks. Cities with municipal police forces therefore reduced their employment requirements by recruiting men for the police service who had only fulfilled their obligatory military service after conscription, rather than service as regulars (Jessen 1991; Spencer 1992). By the 1880s, and much more so after 1890, cities in the West of Prussia, along the Rhine and in the Ruhr, did find it increasingly difficult to recruit candidates with the adequate military background. While some cities reported during the 1890s that they were still managing to staff the municipal police positions accordingly, other cities found the supply of candidates with the requisite military experience falling even more short of their needs. By 1911, the police department in Düsseldorf, one of the largest cities in the Prussian West, with 46 patrolmen's positions to fill, reported that of 500–600 applicants, only 22 had the adequate military background. Thus, despite persistent pressure from the Prussian state to seek out NCOs for police service, the cities in the Prussian West turned increasingly to the local wage-earning population for recruits. Urban administrators usually did so reluctantly, sometimes continuing to see the ideal policeman not just as someone shaped by long years of military discipline but also as someone from the outside, preferably of rural or small town origin. But the reality was that most applicants were local residents with only minimal military experience. As a consequence, the possibility of recruiting of policemen having personal contacts with organized workers loomed ever more threatening (Jessen 1991; Spencer 1992).

The training of policemen was carried out on the job and was oriented along the practical requirements of the man's on the beat everyday activities. The patrolman had to get acquainted with his precinct and with the things he was supposed to look after when on the beat. A number of "formalities" were also briefly taught, such as some basic ideas about the penal code of the Empire, the police ordinances that had been issued for the respective city or community, etc. Apart from that, the writing of dictations was part of the training. Often this was essential because the men's ability to write seems to have suffered considerably during the long years of military service. But all in all, serious training did not take place. Around the turn of the century, the Prussian state government as well as municipal police administrations acknowledged an urgent need for an improvement in the policeman's qualifications. Police schools were set up, and additionally, the military habits of the ordinary policeman were no longer deemed sufficient for the handling of the everyday problems the police encountered in the growing urban contexts. The more Prussian cities grew and the more complex urban society became, the more qualifications were required from the policeman beyond his authoritarian and military attitude. In 1899, the first police school was established in Prussia, as a school for the gendarmerie. The gendarmerie took the initiative, since for this force, which was still part of the military, the problem was the most urgent. In 1901, the first police school for municipal police personnel was set up in Düsseldorf. The costs of the schools were covered by those municipalities who sent their policemen there. In general, municipal administrations accepted the necessity for improving the qualifications of their police personnel, but for financial reasons, however, they limited the number of men they sent to these schools (Jessen 1991; Reinke 1993; Spencer 1992).

Similar institutions were being established in other Prussian districts as well at about the same time, beginning with Berlin in 1895. Police schools were intended not only to impart necessary knowledge and approved attitudes but also to raise police prestige. Increased formal training (whatever its content) would help lessen the gulf between policemen and respected representatives of the Prussian state. In Düsseldorf, the course for patrolmen lasted 2 months, that for senior officers for 3 months. Students were required to live at the school so that its influence could prevail around the dock. The cities paid the costs for the patrolmen, fearing that if the men were required to use their own resources, they would fall into debt, a situation policemen were strongly encouraged to avoid. The senior officers, expected to come from somewhat more well-off families, had to pay for their own instruction. Cities tried to protect their investment in the candidates by stipulating that those who left their departments after less than 5 years had to repay all or part of the costs of their schooling. Also, attendance at the school was typically reserved for those recruits who had already completed 6–12 months of service. As justification for this fiscally prudent move, police administrators argued that schooling was more meaningful if it followed a substantial period of practical experience. By 1906, the state district administration in the Prussian had stipulated that in cities of 10,000 or more, patrolmen either had to attend the police school or pass an examination before being confirmed in their posts. Supporters of the Düsseldorf school were dismayed to find that many communal police departments, to save the cost of instruction, either tried to hire recruits who had already attended a police school elsewhere or else encouraged the taking of the examination. To make certain that the Düsseldorf school had enough students to pay for itself, the Prussian provincial administration before the war was contemplating eliminating the examination option. As a step toward making municipal policemen more credible as rule enforcers by increasing the likelihood that they knew and understood the rules and what they were doing and for what purposes, the Düsseldorf police school represented only a hesitant beginning (Jessen 1991; Spencer 1992).

"Professional" Police Knowledge

Throughout most of the nineteenth century, a coherent and exclusive stock of police knowledge did not exist. This stock of police knowledge emerged only gradually, acquired empirically, and further elaborated during the last decades of the nineteenth century via the huge data collections, which police headquarters in European capitals (Berlin, London, Paris, etc.) started assembling during this late nineteenth century period. There rather existed different layers of police knowledge, which spread according to the functions and tasks different strata of policemen fulfilled. At the turn of the nineteenth century, a comprehensive stock of knowledge had been assembled, but this stock incorporated a pronounced ideological bias, focusing – in the wording of the contemporary discourse – on a "social defense" modus, i.e., focusing on those threatening the values of bourgeois society and "property" as its *Leitmotiv*.

During the first half of the nineteenth century, uniformed policemen started with no knowledge at all. They simply were looking at people, things, and circumstances, when on the beat or on patrol. It has been argued that a specific empiricist police gaze developed gradually, which categorized people as disorderly, deviant, or as potential criminal perpetrators. Parallel to this emergence to a empiricist police gaze existed a flourishing publication market, consisting of police handbooks. Especially in Central Europe, these handbooks were connected significantly to the Central European continuity of "Polizey," thereby conceiving police beyond law-and-order maintenance in a narrower sense. These handbooks were very much meant to give guidelines for protecting the state against those threatening it, thereby mixing the handling and the control of the economic and social challenges developing already during the first half of the nineteenth century with the control of political dissent and opposition, threatening the political regimes of the nineteenth century postrevolutionary period. Both layers, the results of the patrolman's gaze when on the beat and guidelines of the handbooks, were connected loosely only – if at all.

As patrolmen on the beat remained some sort of general factotum until the beginning of the First World War, specialized police knowledge could not develop on that level. The small handbooks and notice books, which were brought on the market, which were published parallel to the establishment of police schools since the end of the nineteenth century, mirrored this, since they were rather designed for guiding in – what we would see nowadays – general administrative matters than in law-and-order maintenance. But another layer became more significant during this period. As a follow-up to rising anxieties about crime and disorder, especially since industrialization and urbanization accelerated its pace since the middle of the nineteenth century, immense data collections on those threatening bourgeois society and particularly "property" were assembled. These data collections, especially the data collected about the body or parts of the body of suspects or perpetrators (bertillonage, dactylography) and immensely growing number of photos taken, brought the policeman's gaze close to contemporary scientific methods and adopted them. But the data collections, or rather the way these data collections were organized, mirrored the threats the police was to handle. The data collections at the Berlin police headquarters reflected this: The core collection there consisted of an extremely detailed collection on theft, indicating not only contemporary anxieties but urban opportunity structures as well (Roth 1997; Wagner 1996).

Professional Policing and Crime Investigation

The second half of the nineteenth century saw in many European countries the rise of specialized criminal investigation departments, which has been perceived as an answer to urban growth and problems related to urbanization patterns during that period. The Berlin police headquarters established a separate and specialized crime investigation branch at the beginning of the 1870s. There had been police officers investigating crime before, but now crime investigation became a specifically institutionalized branch. In 1878, the Metropolitan Police in London established the Criminal Investigation Department/CID, and the Police Judiciaire in Paris underwent significant reform efforts during that period as well. Detectives, Commissaires, and Kommissare had operated in these countries before (Emsley and Shpayer-Makov 2006; Kalifa and Karila-Cohen 2008; Roth 1997; Shpayer-Makov 2002), but henceforth, the establishment of new or reformed police branches, specialized on crime detection and investigation, was a response to similar patterns in very many European countries. Massive urbanization, occurring all across Europe, brought about significant changes in the social fabric of the cities. While urbanization was read on the one side as an indicator of progress during the nineteenth century, it nevertheless led to the anxieties and fears.

The complexities, created by these massive economic and social transformations, were in parts translated into fears of rising crime, to which the establishment of specific crime investigation units was a response. Similar secular trends are observable here in some European countries. Already contemporary nineteenth century observers argued from time to time that the rise in the fear of crime exceeded the rise of crime, but crime investigation departments expanded continuously during the nineteenth century, accompanied by a growth of its personnel. When recruiting for these new or restructured crime investigation units, the police in some European countries followed different recruitment patterns than those followed for the uniformed police. The Berlin police headquarters during the second half of the nineteenth century provide a good example for this trend: The recruitment and the career patterns at the crime investigation unit of Berlin police headquarters, the *Kriminalpolizei*, were increasingly based on expertise and specialized training than on military background. This expertise became – not only in Berlin – the more urgent the more the complexity of the expanding cities grew, accompanied by a growing diversification of opportunity structures allowing for theft. Some governments and some police forces in Europe answered to this growing complexity by diversifying its recruitment criteria when hiring men for the crime investigation branch. In order to catch up with this, the Berlin Kriminalpolizei became exempt to military recruitment obligation, to be observed when hiring men for the uniformed police. Among the members of the Berlin Kriminalpolizei were officers with very different backgrounds. There were former army officers of course, but academics as well, whose academic background comprised theology, the humanities, jurisprudence, medicine, and other disciplines. As these men were better paid than their counterparts in the uniformed police, the number of candidates for the detective force soon exceeded the vacancies (Funk 1986; Wagner 1996).

Police Associations and Police Unions

Between 1882 and the beginning of the Nazi regime in 1933, men from different polices in a number of German states founded approximately 100 police associations, representing the interests of its members. Very many of these associations (Berufsvereine) were very ephemeral, but they indicate the attempts of the police officers to organize and to engage in interest politics, to fight for better pay, for better work conditions, and increasingly for an improved training. Due to the local character of very many polices in Prussia, most of the foundations of these associations before the turn of the century were accordingly of local origin. It was quite often before the First World War only that the police officers managed to organize themselves nationwise or statewise. In quite a number of European countries (Belgium, England, France, the Netherlands, Prussia, and other countries), first attempts were made by police officers to organize. These associations mostly took the form of benevolent societies, as policemen were not allowed to create syndicates, i.e., formal unions or to join them. Policemen in most European countries were forbidden to organize these associations or societies as unions (Keunings 2009; Emsley 1991, 2000; Shpayer-Makov 2002; Berlière and Lévy 2011; van der Wal 2007; Jessen 1991). That is one of the reasons, that, although strike and protest action had been carried out in England in particular in London during the 1870s and the 1880s already, attempts to unionize did fail around 1890. This was taken up again shortly before the First World War, when a police union had established itself firmly within the London Metropolitan Police (Emsley 1991, 2000; Reiner 1978; Shpayer-Makov 2002). But this organization had to be clandestine, as policemen were still not allowed to organize in unions. Police officers had to join secretly. The establishment at the Metropolitan police was probably very much inspired by the successful efforts of members of the Paris police, who had founded a professional association in 1912 (Berliére and Lévy 2011). The growth, from 1905 onward, of trade unionism among the police, observable in many European countries was not only meant as a strategy for improving the conditions of pay and work, but included a more general, professional purpose: The more the police perceived themselves as professionalized, the more they regarded themselves as autonomous vis-à-vis nonprofessionals, vis-à-vis the politicians, for example. This purpose gained momentum within many European polices in particular during the 1920s.

Concluding Remarks

The scope of the development of the police as a profession and of members of the policed as professionals remained limited throughout the nineteenth century. Recruitment patterns for uniformed patrolmen, moving away from military recruitment to civil recruitment, changed gradually only. Institutionalized training facilities (police schools) were established, but the number of these facilities remained low in most European countries before the First World War, indicating nevertheless a beginning. Besides the question, whether the police is rather a "bureaucracy" than a "profession," a question today's research is still dealing with (Reiner 1978; Reiner 2000), is another feature of early nineteenth and twentieth century policing which can be seen as a major obstacle for conceiving "police" as "profession," up until today: Uniformed policemen remained throughout the nineteenth and well into the twentieth century a general factotum (Emsley 1991), serving a vast array of functions beyond law-and-order maintenance in a more narrow sense. The late nineteenth century saw in many countries the advent of specialized police branches, specializing on crime investigation, detection, and control. This branch acquired a public prominence already during the last decades of the nineteenth century, when rising fears of crime and disorder enhanced the demands for professional crime fighting. The social defense modus of crime fighting during the decades before and after the turn of the century, focusing on criminal perpetrators threatening "property," served a very specific, ideologically loaded notion of crime. In addition with the high level of public prominence allotted to this ideologically focused crime investigation work, the (societal) status and prestige of the work of uniformed policemen decreased, whereby crime investigation acquired the status of "real" police work, although, seen from a quantitative perspective, crime detection is a minor part of overall police activity.

A final remark: The development of professional police work is quite often seen as a linear and positive process, where professionalism within the police moves from very small beginnings to "real" and successful professionalism throughout the decades of the twentieth century. There is a remarkable a-historical bias in this conceptualization of professionalization along the time axis. A significant number of European policemen, which saw themselves as professionals during the 1930s already, participated actively and very much driven by own initiatives in mass murder or in the preparation of mass murder during the years of dictatorship and war in the 1930s and 1940s.

A Note on the Historiography of Police as a Profession

Early historical research on the history of the police, in particular research carried out by English-language historians during the 1970s, has started by conceptualizing the police as the strong and effective arm of the state. This conceptualization of the police as "state" was very much guided by an understanding of state, quite often orientated toward Marxist notions of "state," as efficiently controlling, supervising, and directing. This approach left little room for asking questions about the quality of police work, which qualifications policemen had for doing police work, and what the quality of police work meant for the control aspirations of the state.

This top-down-perspective, the research had taken when analyzing the police, has changed for a bottom-up perspective only gradually. Within the field of the history of the police as a profession, this change occurred when the historical police research broadened its scope and started to examine in a more detailed way: How police practices were carried out in local and urban contexts, who the men had been, where they came from, how they were trained and paid, and how effectively they performed their control functions in terms of the expectations attached to the police as an efficient control instrument, the expectations attached to the police as law-and-order-maintainers, as professional crime fighters or as "thief-takers." This research demonstrated not only how contradictory and how slow these patterns developed during the nineteenth century, but also shed light on the extent, to which policemen engaged themselves in efforts to increase the performance of police work and that of individual police officers, and to improve the overall status of the police force. This bottom-up perspective not only contextualized the policemen's lot as workers, striving for the status as professionals, but also offered a multifaceted account of the police force as an institution of professionals. The available literature gives a detailed account of how different approaches and strategies (training, expertise, unionization, and status improvement) were applied as professionalization devices until the First World War. Some periods and contexts are still greatly underexplored, however. Only a limited stock of knowledge is available for the interwar years and the impact of their multifold crises on the police's strive for professionalization. Equally ignored by historians for a long time and becoming a research topic only very recently is the issue of what professional police expertise and performance – or the police's understanding of this – had meant for the police's participation in colonial regimes, for its role as being an essential part of dictatorships in many European countries since the 1920s, and for its active participation in murderous and genocidal practices, such as in the Soviet Union in the 1930s or such as under the Nazi regime in Germany after 1933 and in the countries occupied by the Germans between 1933 and 1945.

Inspired by a more recent model of professionalism (which however never replaced that of the professional crime-fighter or "thief-taker"), in which police serve, learn from, and are accountable to the community, historians have drawn on older strands of sociological participatory research on police practices to shift their attention more to police relationships with the public, and to one of the basic issues with regard to the role of the police in Western urban society: the conflict between professional autonomy and popular control. In other words, the bottom-up perspective gradually went hand in hand with less police-centered research, whereby the profession and notions of professionalism are examined in their interactions with society and (the demands and attitudes of) the citizen.

In a similar way, inspired by a shifting emphasis in attempts at increased police professionalization from the late twentieth century onward from a primacy of personnel reform (recruitment, training, pay, etc.) to one of technological managerial reform (whereby ultimately the drive for new efficiency and technology has become dominant over the search for a new policeman), historians have increasingly begun to examine the historical antecedents of "scientific" and "intelligence-led" policing and the use and impact of new technologies on professional practices.

Related Entries

- ▶ British Police
- ▶ Conceptualizing of Police
- ▶ Forensic Science Professions
- ▶ German Police Until 1918
- ▶ History of Police Unions
- ▶ History of Technology in Policing
- ▶ Intelligence-Led Policing
- ▶ Organizational Change and Police Legitimacy
- ▶ Police Culture
- ▶ Police in Nazi Germany
- ▶ Police Selection
- ▶ Police Self-Legitimacy
- ▶ Policing in Britain c.1750–1829
- ▶ Role and Function of the Police

Recommended Reading and References

Berlière M (1996) Le monde des polices en France. XIXe–XXe siècles. Editions Complexe, Paris

Berlière JM, Lévy R (2011) Histoire des polices en France de l'Ancien Régime à nos jours. Nouveau Monde éditions, Paris

Curilla W (2006) Die deutsche Ordnungspolizei und der Holocaust im Baltikum und in Weißrussland 1941–1944. Ferdinand Schöningh, Paderborn

De Koster M (2008) Routines et contraintes de la police urbaine à Anvers, 1890–1914. In: Berlière JM, Denys C, Kalifa D, Milliot V (eds) Etre policier: Les métiers de police en Europe, XVIIIe–XXe siècle. Presses Universitaires de Rennes, Rennes, pp 345–362

Deluermoz Q (2012) Policiers dans la ville. La construction d'un ordre public à paris (1854–1914). Publications de la Sorbonne, Paris

Emsley C (1991) The english police: a political and social history. Harvester Wheatsheaf, Hemel Hempstead

Emsley C (2000) The policeman as worker: a comparative survey c. 1800–1940. Int Rev Soc History 45:89–110

Emsley C (2009) The Great British Bobby: a history of British policing from the 18th century to the present. Quercus, London

Emsley C, Shpayer-Makov H (2006) Police detectives in history. Ashgate, Aldershot
Emsley C, Weinberger B (1991) Policing Western Europe: politics, professionalism, and public order, 1850–1940. Greenwood Press, New York
Fijnaut C (2009) A history of the Dutch police. Boom, Amsterdam
Funk A (1986) Polizei und Rechtsstaat. Die Entwicklung des staatlichen Gewaltmonopols in Preußen 1848–1914. Campus Verlag, Frankfurt/Main
Jessen R (1991) Polizei im Industrierevier. Modernisierung und Herrschaftspraxis im westfälischen Ruhrgebiet 1948–1914. Vandenhoeck & Ruprecht, Göttingen
Johansen A (2005) Soldiers as police: the French and Prussian armies and the policing of popular protest, 1889–1914. Ashgate, Aldershot
Kalifa D (2007) Histoire des detectives privés en France (1832–1942). Nouveau Monde éditions, Paris
Kalifa D, Karila-Cohen P (2008) Les commissaires de police au XIXe siècle. Publications de la Sorbonne, Paris
Keunings L (2009) Des polices si tranquilles. Une histoire de l'appareil policier belge au XIXe siècle. Presses Universitaires de Louvain, Louvain-la-Neuve
Klein J (2010) Invisible men: the secret lives of police constables in Liverpool, Manchester, and Birmingham, 1900–1939. Liverpool University Press, Liverpool
Meershoek G (1999) Dienaren van het gezag. De Amsterdams politie tijdens de bezetting. Van Gennep, Amsterdam
Meershoek G (2007) De geschiedenis van de Nederlandse politie. De Gemeentepolitie in een veranderende samenleving. Uitgeverij Boom, Amsterdam
Reiner R (1978) The blue-coated worker: a sociological study of police unionism. Cambridge University Press, Cambridge
Reiner R (2000) The politics of the police. Oxford University Press, Oxford
Reinke H (1993) '. . . nur für die Sicherheit da. . .'? Zur Geschichte der Polizei im 19. und 20. Jahrhundert. Campus Verlag, Frankfurt/Main
Roth A (1997) Kriminalitätsbekämpfung in deutschen Großstädten 1850–1914. Ein Beitrag zur Geschichte des strafrechtlichen Ermittlungsverfahrens. Erich Schmidt, Berlin
Shpayer-Makov H (2002) The making of a policeman: a social history of a labour force in metropolitan London, 1829–1914. Ashgate, Aldershot/Burlington
Shpayer-Makov H (2011) The ascent of the detective: police sleuths in Victorian and Edwardian England. Oxford University Press, Oxford
Spencer EG (1992) Police and the social order in German cities: the Düsseldorf District, 1848–1914. Northern Illinois University Press, DeKalb
van der Wal R (2007) De geschiedenis van de Nederlandse politie. De vakorganisatie en het beroepsonderwijs. Uitgeverij Boom, Amsterdam
Wagner P (1996) Volksgemeinschaft ohne Verbrecher. Konzeptionen und Praxis der Kriminalpolizei in der Zeit der Weimarer Republik und des Nationalsozialismus. Hans Christians, Hamburg
Williams CA (2011) Police and policing in the twentieth century. Ashgate, Aldershot

History of the Self-Report Delinquency Surveys

Janne Kivivuori
Criminological Unit, National Research Institute of Legal Policy, Helsinki, Finland

Synonyms

Hidden crime survey; Offender-based crime survey; Self-report delinquency survey

Overview

Criminologists use survey methods in order to study the full extent of crime, including both unrecorded crimes and crimes recorded by the criminal justice system and related official agencies. The self-report delinquency survey is one of the two major survey instruments currently used by criminologists to that effect (the other one being the victim survey). This entry presents a history of the offender-based self-report delinquency survey. The invention of the research method is described. In addition, the rise of the self-report survey is explained by reference to internal and external factors influencing criminology.

The main scientifically *internal* factor was the problem of unrecorded crime which haunted the age of moral statistics (early criminology) from 1820s to 1930s. The researchers of this era despaired over the impossibility of breaking the official control barrier of crime measurement. The barrier was finally broken by the groundbreaking self-report delinquency survey studies published in 1943 by Austin Larimore Porterfield, professor of sociology at the Texas

Christian University. This moment signifies the end of the moral statistics era and the beginning of scientific criminology because with the survey method criminology became independent of the official control and medical agencies.

While the current use of the self-report method is based on research showing its reliability and validity, the initial rise of the method took place in a specific historical context. Its rise was influenced by the following *external* factors: religious practices and ideologies, populist perception of hidden depravity among the middle classes, the legislative activism of the Progressive Era, suspicion of control biases in an ethnically diverse society, and the largely successful attempt of sociology to take over the field of criminology from the previously hegemonic psy-sciences. Inspired by this historical context of discovery, the first generation of self-report researchers sought to normalize crime by redescribing it as normal, thus giving empirical support for Émile Durkheim's notion that "crime is normal." In addition to these cultural factors, there were also a number of necessary conditions for the widespread deployment of the quantitative and typically self-administered delinquency survey, such as mass literacy and compulsory basic education.

Of the three basic quantitative methods of measuring crime independently of the criminal justice or clinical statistics, the *offender-based self-report survey* (asking offenders to tell about their crimes) preceded the rise of the *victim survey* (asking victims to tell about their experiences) but came after various *proxy report* methods (asking someone who knows offenders) were tried. The modern self-report delinquency survey has three basic features. First, it is based on standardized questions which are identical for the respondents. Second, the instrument is used to produce quantitative data in samples systematically selected by the researcher according to a clearly defined procedure; the samples are not self-selected (as in clinical data), selected by control agents (as in official control data), or unsystematically selected (as in qualitative studies or journalism). Third, the self-report survey is based on confession: the respondents are asked to report, typically anonymously, about the crimes they have committed. This entry presents the history of the modern self-report delinquency survey (for a more detailed description, see Kivivuori 2011).

Moral Statistics and the Official Control Barrier

Within criminology as a discipline, the history and invention of the self-report delinquency is intimately related to the problem of unrecorded crime. That problem was born in the early nineteenth century when European states started to publish the so-called moral statistics, including crime statistics. It was immediately recognized that a major problem with such crime statistics was that they excluded undetected and unrecorded offenses. The immediate solution was the so-called *constant ratio doctrine*. Formulated by Adolphe Quetelet, this doctrine stated that the ratio of hidden crime to known crime was stable, meaning that official statistics was a valid proxy measure for real crime (Beirne 1993, p. 80).

The moral statisticians were never naïve about the problem of hidden crime; they intensely felt that reliance on official statistics was a serious limitation. The concept of "dark figure," coined by the Japanese lawyer Shigema Oba in 1908 (Oba 1908), was thus a continuation of long self-critical reflection in the then almost century-old moral statistics tradition. What ensued was a gradual shift of emphasis in the kind of official statistics which was seen as most valid: the craft of moral statistics moved down the ladder of criminal justice institutions in search for more valid indicators of criminal behavior: from the courts and prosecutors to the police, and within the police statistics, from ascertained to reported crimes and to arrests. Thus, the researchers of the moral statistics era were moving closer to the offense as it looked for means to measure the extent and patterns of crime more precisely. Towards the 1920s and 1930s, the police statistics were prevailing as crime indicators, but this appeared to be the end of the road. There appeared to be no way of proceeding even closer to the offense. By the early decades of the twentieth century, it was typical of crime researchers to despair about the impossibility of direct measures of crime. In recent historical scholarship (Kivivuori 2011), this obstacle and the associated sense of despair has been described as the *official control barrier* of crime measurement. That barrier was almost like the Heisenberg uncertainty principle in physics: empirical observation seemed to be fundamentally warped by it.

Non-survey Responses to the Problem of Hidden Crime

There were however no shortage of attempts at breaking the official control barrier. The self-report survey was not the only (or first) attempt to that effect. Various methods which were subsequently eclipsed by the survey were tried. Often these methods were based on *proxy reports* by informants who for some reason knew about crime; such informants included the police, philanthropic services, medical sources, insurance companies, banking associations, and private firms as sources of information. Some of the proxy report solutions were developed very early; thus, in the 1880s Charles Booth conducted the first "social survey" of the London poor by employing school board visitors as proxy reporters; the social survey tradition was also interested in the morals of neighborhoods. In the early part of the twentieth century, multiple local crime surveys were conducted in various areas of the USA. It is important to observe that in these early stages, the word survey (as in "social survey movement" or "Illinois crime survey," etc.) stood for local fact-finding missions using multiple data sources but not anything like the modern self-report or victim surveys. They were thus non-survey methods of crime measurement. Some proxy report studies used the reporting propensity of institutional victims to derive a multiplier which was then used to estimate the true crime rate from the number of recorded crimes. Other methods included direct observation as in the UK Mass Observation project during the 1930s (Willcock 1949). Some of the early qualitative sociology in the Chicago School style was similarly a means of studying deviance directly, without any intermediaries, following the example of earlier social reports by continental European intellectuals giving first-hand descriptions of "underworlds."

Early Attempts at Systematic Use of Self-Reports

The above-described early solutions to the riddle of the official control barrier often resulted in an anecdotal style of reporting. In contrast, the distinctive feature of the modern criminological self-report delinquency survey is that data collection is based on standardized (identical) questions and clearly defined target group selected by the researcher. Before the first deployment of the full-blown modern self-report survey in 1940, there were occasional incidents were such standardization was attempted.

The surveys of W. E. B. Du Bois (1868–1963) in Atlanta are among the "anticipators" of criminological surveys. In the 1904 Atlanta survey, he asked young people about their contacts with the courts and the police (in any role). While there was no explicit or stated aim of measuring the extent of unrecorded crime, the Du Bois survey represents an early attempt to bypass the criminal justice system in gathering criminological data (Du Bois 1904). At the same time, important developments were taking place in Europe. The surveys of homosexual behavior conducted in Amsterdam (1901) and Berlin (1903–1904) by Lucien von Römer (1873–1965) and Magnus Hirschfeld (1868–1935), respectively, are among the more important anticipations of the discovery of the systematic self-report delinquency survey. Today of course homosexuality is not a crime in most developed nations, but a century ago it was, making the von Römer and Hirschfeld surveys count as among the earliest systematic self-report delinquency surveys. Both the Dutch and the German surveys were connected to the activities of Hirschfeld's Scientific-Humanitarian Committee, a liberal pressure group which aimed at normalizing sexual deviance by showing its huge prevalence.

Römer and Hirschfeld were both trained in medicine. Interestingly, Hirschfeld referred to the Catholic practice of confession as a source of inspiration in the development of the self-report survey. Other sources of inspiration included early US educational sociologists making surveys in schools. Typically for self-report delinquency research, both pioneers used student samples, but Hirschfeld additionally surveyed blue-collar workers by mail survey. Fifty years later, the Kinsey research team observed that Hirschfeld deserves "considerable credit for having tried on a larger scale than anyone had before to ascertain the facts on a matter that has always been difficult to survey" (Kinsey et al. 1948, p. 620).

Within psychology, the 1920s witnessed the development of what is today called the social desirability scale. The social desirability scales are used to control for the validity of survey responding. However, the development of the social desirability measures remained largely separate from the work that led to the modern self-report delinquency survey.

Approaching Takeoff: The Sutherland Experiments

In 1931, Thorsten Sellin formulated the axiom which later became known as Sellin's law, stating that "Due to a number of variable elements represented by changes in administrative policies and efficiency, the value of a crime rate for index purposes decreases as the distance from the crime itself in terms of procedure increases" (Sellin 1931–1932, p. 346). At about the same time, a century of reliance on official criminal justice statistics climaxed in a sequence of local crime surveys in various US cities and states. These were surveys in the old meaning of the word (fact-finding missions), but they were all "moving towards the offense" as stipulated by Sellin. In many ways, the most sophisticated old-style local delinquency survey was the one conducted in New York by Sophia Moses Robison, published as *Can Delinquency Be Measured* (1936). Her idea was to collect statistics from all kinds of control and philanthropic institutions, not only the police. She thus shifted the boundary of recorded and unrecorded crime by including more records, a strategy that still remained within the paradigm of moral statistics. Sophia Moses Robison (1888–1969) was however aware of these limitations; she suggested that delinquency data could be collected by asking people about it; indeed, it was known to her that Edwin H. Sutherland was trying to do something like that.

It is apparent that Sutherland (1883–1950) was conducting self-report experiments among his students. Throughout the mid-1930s his writings and speeches occasionally refer to efforts at breaking the official control barrier of crime measurement. Speaking to the Milwaukee police in 1936, he described one such experiment as follows: "Of a class of forty, ten did not hand in papers, and since I had no signatures, I could not tell whether they were lily whites or had so many and so serious thefts that they refused to tell. Of those who did hand in reports, only one insisted that he could remember no thefts" (Sutherland 1936, in Snodgrass 1972, p. 222). This experiment had most aspects of the standard SR survey: standardized question, researcher-selected population, and confessional logic. The results revealed a very high lifetime prevalence of theft (97 %, N = 30). Sutherland used these survey experiments to press the point that crime was very prevalent and therefore normal (see also Sutherland 1973 [1936], p. 143 and 1972 [1937], p. 213).

Breaking the Official Control Barrier of Crime Measurement

Austin Larimore Porterfield

The first modern criminological self-report delinquency survey was conducted in 1940–1941 by the Texan sociologist Austin L. Porterfield (1896–1979). The year 1940 can thus be seen as the moment when the official control barrier of crime measurement was finally broken. The result of the survey was first published as the article "Delinquency and Its Outcome in Court and College" (Porterfield 1943) but became more widely known as part of his monograph *Youth in Trouble* (1946). Porterfield's discovery marks the moment when the discipline of moral statistics was transformed into the science of criminology, because at that point the empirical study of crime became independent from data created by agencies of the state.

Porterfield tried to make his questions about crime as similar as possible to the repertory of the offenses defined by the law. In two consecutive data collections in three Texan educational institutions (1940–1941 and 1941–1942), he targeted both precollege and college students about their crimes; the total number of respondents was 337. The results showed a "universal prevalence of past delinquency among college men and women" (Porterfield 1946, p. 38). More than one in five males had driven when drunk, while 62 % had participated in ordinary fighting. Ten percent had shoplifted. At the time, these figures where shockingly high. Like Hirschfeld, Porterfield interpreted his findings by underscoring the high prevalence and hence *normality* of crime. Delinquents were "not a sub-species of Homo Sapiens"; they were like all of us (Porterfield 1946, p. 45). Anticipating problems related to this interpretive frame, he however noted that incidence measures of crime might show different results, marking essential differences between nondelinquents and delinquents (Porterfield 1946, p. 42). Otherwise his take-home message was that delinquency was normal. He rejected the notion that delinquents are fundamentally different from the "law-abiding" citizen. He was a strong critic of what he called the "we-they" fallacy, the tendency to see a black-and-white distinction between supposedly "abnormal" criminals and a (putatively) law-abiding majority (Porterfield 1957, p. 46). The great importance of Porterfield's work was not immediately recognized, but already in the 1950s his work became known as the initiator of the self-report delinquency survey tradition.

Porterfield's style of writing and analysis was satirical. He critiqued the hypocrisy of the local elites by first exposing how their punitivity was compromised by their hidden delinquency. Characters such as store managers, middle-class parents, drunken police chiefs, and members of the Ku Klux Klan were satirically exposed as hypocrites. The breakthrough phase of the self-report delinquency survey thus combined quantitative analysis and social critique.

The Randen Foundation Crime Survey

Apart from Porterfield's groundbreaking studies, the second most important early self-report delinquency survey was Randen Foundation Crime Survey, the results of which were published in 1947 by James S. Wallerstein (1910–1990) and Clement J. Wyle (1903–1991). The title of their seminal paper was typically satirical, like many other early self-report studies: "Our Law-abiding Law-breakers" (Wallestein and Wyle 1953 [1947]). This study expanded the application of the newly invented method to adult population in contrast to student samples used by Sutherland and Porterfield. The data collection took place in the state of New York, with 1,698 respondents in the final dataset (response rate not known). As in other pioneering self-report surveys, the main finding was that lifetime participation in criminal activity was much more prevalent than typically expected. Nearly half of the male respondents reported "assault," while one in six admitted to burglary. Wallerstein and Wyle duly noted that some of the high figures may have reflected people's broad conceptions as to what "assault" means. However, the main finding was that unrecorded criminality was very prevalent in the population. Like other pioneers of the self-report survey, Wallerstein and Wyle underscored the high prevalence and therefore the normality of (occasional) crime. They also developed the random control argument, suggesting that being registered for crimes might reflect chance events more than criminality as such. Like Porterfield, they also critiqued the punitive mind-set of the putatively law-abiding people. During the late 1940s, Wallerstein published several articles using the normal-because-prevalent argument to critique punitivist attitudes.

Cambridge-Somerville Youth Study

When the early tradition of self-report delinquency surveys was shaped, a third important study was the Cambridge-Somerville Youth Study (CSYS). Its main goal was to asses in an experimental setting the efficacy of social work in the prevention of crime (Powers 1949). During the course of the project, much first-hand information was collected from the boys who participated in the experiment. In an article published in 1946, the CSYS workers reported the very high prevalence of unrecorded crime among the project subjects (Murphy et al. 1946). They concluded that because of the great frequency of undetected crime, changes in the efficacy of social control could drastically influence crime trends based on official statistics. In so arguing they thus presented a strong empirical critique of the constant ratio doctrine. Later on these findings were also interpreted from the point of view of normal-because-prevalent argument (Powers 1949, p. 82). While the Murphy et al. study was and is often cited as one of the first hidden crime studies, it remains a borderline case with respect to the modern self-report delinquency survey. Based on the psychiatric case worker data collection, its level of standardization remained comparatively obscure. The inclusion of boys who were known to come from disadvantaged backgrounds was also unique for this project. This was important because it enabled the CSYS researchers to conclude that boys with higher intensity of offending were more likely to be prosecuted and convicted. This finding was a crack in the moral framework based on the alleged normality of offending, something that was immediately recognized by Porterfield (1947, pp. 425–426).

External Factors Influencing the Rise of the Self-Report Survey

From the perspective of the internal development of criminology as a science, the self-report survey was a way of breaking the official control barrier in crime measurement. The survey enabled the systematic quantitative study of offenses which are not known to criminal justice or medical authorities. The invention of the self-report survey in 1940 thus marks the moment when moral statistics was transformed into scientific criminology because at that time criminology became independent of the control system of society.

Like all scientific innovations, the invention of the self-report delinquency survey took place in a specific historical context involving a constellation of social, cultural, technological, and policy-related influences. In science studies, the difference between internal scientific developments and external influences is often conceptualized with the so-called context distinction which separates the context of justification and the context of discovery when explaining an innovation in science (Schickore and Steinle 2006). The context of discovery captures all external influences which are not strictly internal to science. In what follows, the major external influences defining the context of the rise of the self-report delinquency survey are described.

Religious Influences

Previously, it has been suggested that the rise of the self-report survey was linked to the punitivist mind-set of the American Puritan tradition, which would have made scholars very interested to study minor infractions. Recent historical research (Kivivuori 2011) has definitely refuted this thesis by showing that the first-generation self-report researchers were typically liberals attacking punitivist attitudes by empirically showing the high prevalence and hence normality of crime. But it is correct that there is a religious substratum in the relevant paradigm shift. Hirschfeld's early work was inspired by confession as a religious practice. But more importantly, the generation of US scholars who pioneered the self-report method came from backgrounds influenced by the so-called Christian Sociology and the Social Gospel movement. From the end of the nineteenth century, these traditions sensitized scholars to the notion that seemingly law-abiding persons could be hidden offenders. In the beginning, this was associated with corporate offenders and "criminaloids" who were later renamed as white-collar offenders. Books such as *Sin and Society* (1907) by the sociologist Edward Alsworth Ross were influential bridges between religious and secular types of populism (see below). More generally, the critique of hypocrisy was based on the latent religious notion that since we are all offenders, it is wrong to judge harshly those who are detected offenders. Thus, many first-generation self-report researchers were inspired by the Biblical dictum "He who is without sin among you, let him throw the first stone" (John 8:3–7). This is pronounced in writers such as Porterfield, Wallerstein, and especially among the Nordic pioneers of the self-report survey.

Populist Traditions

The US populist tradition, which can be seen as a secular offshoot of the Social Gospel movement, was also an important part of the context which sensitized a generation of sociologists to see hidden criminality. The critique of "robber barons" transformed into the study of (the largely unrecorded) crimes of the white-collar offenders. It is no coincidence that Sutherland continued this tradition in his classic *White Collar Crime* (1948) while simultaneously experimenting with the self-report method. Both endeavors were based (or intended to show) that crime was committed by persons who were psychologically perfectly normal. Indeed, in its early phases, the self-report delinquency survey was regarded as part of the more general subfield of "white-collar crime studies", and this lasted at least until the 1950s. The US populist and Progressive Era also impacted criminology by initiating a period of intense legislative activism. As new behaviors were defined as crimes, the need to replace official statistics as the most important data source of criminology became very acute.

Suspicion of Biased Control in Ethnically Diverse Society

One of the contextual factors that sensitized US scholars to the artificiality of the official crime data as a social construction was the treatment of minorities. Scholars such as W. E. B. Du Bois wanted to know whether rising figures of recorded crimes reflected the increasing "efficacy of the judicial system in ferreting out and punishing crime" and concluded that crime statistics are "too mingled with extra-moral causes to be trustworthy" (Du Bois 1904, p. 10; Du Bois and Dill 1914, p. 11). Thorsten Sellin was also motivated by the same problem in his efforts to improve crime indexes.

Disciplinary Conflict

A very important external factor in the rise of the self-report survey was the disciplinary warfare between sociology and the psy-sciences (psychiatry and psychology). The abnormality paradigm of the psy-sciences had been hegemonic in criminology since the times of Lombroso. In contrast, the self-report survey was invented and supported by sociologists. They used this new empirical instrument to show that criminal behavior was so prevalent that it was impossible to see it as a sign of abnormality. This gave empirical boost to Émile Durkheim's early notion of crime as a "normal" social phenomenon. Sutherland especially was known as a long-standing critic of psychiatry and psychology; his link to normal-because-prevalent argument is part of that picture. Key sociological mandarins of the mid-twentieth century period, such as Robert K. Merton, used the earliest self-report surveys in arguing against psychology and psychiatry. The high cultural salience of Freudian psychoanalysis during the Post-War era in the USA probably helped to disseminate the notion that deep down, every person is deviant. This may have supported the acceptance (though not the invention) of the self-report delinquency survey. Indeed, it can be argued that the normality notion was the single most important part of the context of discovery in which the self-report delinquency survey was invented. This was so especially in the original discovery and in the second phase when the Nordic countries applied the method from the late 1950s.

Attack Against Punitivity

The first generation of self-report researchers used quantitative social analysis to attack notions which would today be described as "right wing" or "punitivist." Porterfield, Wallerstein and Wyle, and especially the Nordic self-report pioneers strongly critiqued harsh punishments. They argued that since everyone commits crimes, it is unjust to be harsh towards the few who are detected. This argument was intimately connected to subterranean religious traditions, and it was also possible to combine it with populism. From the perspective of disciplinary conflict, the empirical proof of the normality of crime was used both against the psy-sciences (crime is normal) and lawyers (punishments are unfair and ineffective). In critiquing outdated normative and repressive regulation with reference to high prevalence of "deviant" behaviors, the first generation of self-report delinquency researchers spearheaded the kind of argumentation which was also evident in the more famous Kinsey report of sexual behavior (Kinsey et al. 1948).

Other External Preconditions

The self-report survey was a means of breaking the official control barrier of crime measurement. Its success in this task was context dependent and socially embedded in multiple ways. First, the deployment of the survey in large populations required the rise of mass literacy. The historical link of the survey with educational institutions is thus very strong: the rise of basic education created the conditions of the survey by disseminating literacy and by additionally constituting an institutional basis for the survey as much of the early work was targeted at students. The crucial importance of the rise of the state-sponsored literacy is testified by Hirschfeld's ability to conduct a self-report mail survey among German blue-collar workers in the early years of the twentieth century.

Later on, the closely related development of newspaper publishing, opinion survey techniques, and improvements in sampling theory further helped in the creation of techniques that would later be applied in criminology as well (Converse 2009). An important precondition was also the existence of a critical mass of empirically oriented social science researchers and sufficient research funding. These factors were probably decisive in explaining why the official control barrier was broken in the United States as opposed to Europe.

In sum, the slow development of crime indicators during the long century of moral statistics from 1820s to 1930s cannot be attributed to a lack of insight among the moral statisticians. For much of that period, the necessary external preconditions of quantitative and literacy-based mass surveys were simply not in existence. When the required conditions emerged, cutting-edge scientists were quick to break the official control barrier. The time and place where this happened (the USA in the 1930s) was additionally influenced by the cultural movements described above.

Developments After Initial Discovery

The above narrative describes how the self-report delinquency survey was initially invented as response to an internal problem of crime measurement. The breakthrough took place roughly between 1930 and 1950 in the United States. While there were multiple external factors which influenced this process, it is important to underscore that the continued scientific use of the method was soon disconnected from the heavily morally driven context of its original historical discovery. Today criminologists use the self-report delinquency survey because its reliability and validity is ascertained in independent methodological research. Indeed, recent historical research has shown that much of the post-1950 development of the survey ironically produced results which were not expected nor anticipated by the generation that first started using the method. The notion of normality is a case in point. When the incidence measures of the surveys were improved, the small minority of "abnormal" or "chronic" offenders made a spectacular comeback, and this time his existence could not be denied by the claim that he was a social construction created by biased control. This shows that it is important to use the analytic distinction of context of discovery and context of justification in the historical analysis of science.

Mainstream Departs from the Normality Framework

The Post-War era witnessed what has been described as the Americanization of social sciences (a process that took place in the USA and elsewhere, see Haney 2009). This process involved a detachment from moral- and politics-oriented research and a move towards a more positivist style research based on incremental accumulation of knowledge. The development of the self-report delinquency survey since the 1950 was no exception to this trend. The work of James F. Short and F. Ivan Nye concentrated on the scale development of self-report instruments and then moved towards analysis of delinquency causation. Harrison Gough, Walter C. Reckless, Simon Dinitz, and others also shared these goals. Starting from this period, validation studies increasingly suggested that official control may be biased but not random with respect to criminality.

Nordic Revival of the Normality Framework

The first geographical dissemination phase of the self-report delinquency survey involved the Nordic countries. For a long time and until the early 1980s, it was possible for observers to comment that "most studies of self-reported delinquency have been American or Scandinavian" (West 1982, p. 20). Ultimately, this reflects the close ties of Nordics with the USA after the Second World War. The first pilot studies were initiated in 1959 by the Norwegians. Soon after that the Nordic Research Council of Criminology created an ambitious all-Nordic effort known as the Nordic Draftee Research project. The NDR project was globally the first internationally comparative self-report delinquency survey and kept this pioneering position for a very long time. The findings of the NDR were similar to the US precedents, and the Nordic scholars enthusiastically underscored the normal-because-prevalent argument, using it to critique the psy-sciences and the old-school lawyers for ungrounded punitivism. Today, self-report delinquency research remains one of the focal areas of Nordic criminology (Kivivuori and Bernburg 2011).

Subsequent Developments

During the 1960s and early 1970s, the self-report delinquency survey started to spread to areas outside the US and the Nordic area. The sheer volume of research increased as its flexibility in testing theories was fully recognized (Krohn et al. 2010). The 1960s witnessed several countries making efforts towards nationally representative youth samples, and self-report measures were being incorporated to longitudinal projects such as the *Cambridge Study in Delinquent Development* (West and Farrington 1973). After that, there was a period of slower dissemination and, in some countries, even stagnation (Aebi 2009, pp. 31–32). This partially reflected the reemerging interest in register-based studies. After all, the self-report delinquency survey had itself corroborated that the frequent offender existed and that, furthermore, the frequent offender was more likely to be arrested than others. This result supported the feasibility of register-based data in crime analysis. The classic monograph *Delinquency in a Birth Cohort* by Wolfgang et al., published in 1972, is a case in point. Additionally, for some time since the late 1960s, the energies of survey-based researchers were directed to the development of the victim survey as distinct from the offender-based self-report survey. These developments may partially explain, for instance, why the first full-blown international self-report project became reality only in the early 1990s when the *International Self-Reported Delinquency Survey* was launched at the initiative of Josine Junger-Tas. With the partial exception of the Nordic experience in the 1960s, all these developments took place in the neutral framework of crime measurement. The offender-based survey had become a regular and objective methodological tool of criminology with no connections to the morally loaded period of the initial discovery phase.

Further Readings and Research Needs

The history of the self-report delinquency survey method in criminology has been recently described in greater detail in the monograph *Discovery of Hidden Crime* (Kivivuori 2011) which covers both the original discovery phase in the US and the Nordic sequel. Krohn et al. (2011) provide an insightful overview of how the self-report delinquency survey has influenced theoretical developments in criminology. An in-depth description of the first internationally comparative self-report delinquency survey, the Nordic Draftee Research program, is also available (Kivivuori and Bernburg 2011). While the history of criminology has evoked more scholarly interest over the recent years, several important research problems call for further scrutiny. The prehistory of the self-report delinquency surveys probably goes further back in time than the von Römer and Hirschfeld studies and remains to be charted. Studies of survey history would also benefit from a closer look at the relationship of sexology and criminology because before the historically recent relaxation of moral legislation, sex surveys may have contained questions on criminalized behaviors. It is also known that the Kinsey survey influenced US criminologists. While the Kinsey report came after (1948) the key breakthroughs of criminological self-report studies, there were earlier sex surveys which may have influenced criminologists. The criminological work also influenced sex research as scholars like Porterfield were active in both fields.

Furthermore, more historical and archival work is needed on the period which immediately precedes the Porterfield breakthrough. Furthermore, it might be of some interest to know more about the strange role of the polymath and novelist James S. Wallerstein whose article “Our Law-abiding Law-breakers” (1947), coauthored with Clement J. Wyle, was very influential in criminology. On a more general level, an in-depth history of the victim-based survey remains to be written; such an enterprise might benefit from explicit comparison with the (historically earlier) rise of the offender-based self-report delinquency survey. The history of the self-report delinquency survey provides an important lesson generally for science studies: while the discovery of the method was strongly embedded in specific moral and policy framework (normality of crime), its deployment yielded empirical results which partially contradicted that very framework. Observation was initially theory laden, but empirical findings forced researchers to modify their theoretical framework. This also means that “theory history” and “methods history” cannot be separated.

Related Entries

- ▶ History of the Dutch Crime Victimization Survey(s)
- ▶ History of the Statistics of Crime and Criminal Justice
- ▶ National Victimization Surveys
- ▶ Self-Reported Offending: Reliability and Validity

Recommended Reading and References

Aebi M (2009) Self-reported delinquency surveys in Europe. In: Zauberman R (ed) Self-reported crime and deviance studies in Europe. Current state of knowledge and review of use. Brussels University Press, Brussels

Converse JM (2009) Survey research in the United States. Roots and emergence 1890–1960. Transaction Publishers, New Brunswick

Du Bois WEB (1904) Notes on Negro crime particularly in Georgia. Atlanta University Press, Atlanta, Atlanta University Publication No. 9

Du Bois WEB, Dill AG (1914) Morals and manners among Negro Americans. Atlanta University Press, Atlanta, Atlanta University Publication No. 19

Haney DP (2009) The Americanization of social science. Intellectuals and public responsibility in the post-war United States. Temple University Press, Philadelphia

Kinsey AC, Pomeroy WB, Martin CE (1948) Sexual behavior in the human male. Indiana University Press, Bloomington

Kivivuori J (2011) Discovery of hidden crime. Self-report delinquency surveys in criminal policy context. Oxford University Press, Oxford

Kivivuori J, Bernburg JG (2011) Delinquency research in the Nordic countries. In: Tonry M, Lappi-Seppälä T (eds) Crime and justice in Scandinavia. Crime and justice – a review of research. University of Chicago Press, Chicago

Krohn MD, Thornberry TP, Gibson CL, Baldwin JM (2010) The development and impact of self-report measures of crime and delinquency. J Quant Criminol 26:509–525

Murphy FJ, Shirley MM, Witmer HL (1946) The incidence of hidden delinquency. Am J Orthopsychiatry 16:686–696

Oba S (1908) Unverbesserliche Verbrecher und ihre Behandlung. Inaugural-Dissertation der juristischen Fakultät der Friedrich-Alexanders-Universität zu Erlangen. Buchdruckerei Rober Noske, Borna, Leipzig

Porterfield AL (1943) Delinquency and its outcome in court and college. Am J Sociol 49:199–208

Porterfield AL (1946) Youth in trouble. Leo Potishman Foundation, Forth Worth

Porterfield AL (1957) The “we-they” fallacy in thinking about delinquents and criminals. Fed Probat 21:44–47

Powers E (1949) An experiment in prevention of delinquency. Ann Am Acad Pol Soc Sci 261:77–88

Ross EA (1973 [1907]) Sin and society. An analysis of latter-day iniquity. Harper & Row, New York

Schickore J, Steinle F (eds) (2006) Revisiting discovery and justification. Historical and philosophical perspectives on the context distinction. Springer, Netherlands

Sellin T (1931–1932) The basis of a crime index. Am Inst Crim L Criminol 335:335–356

Snodgrass J (1972) The American criminological tradition: portraits of the men and ideology in a discipline. A Dissertation in Sociology, University of Pennsylvania

Sutherland EH (1972 [1937]) The professional thief. By a professional thief (Annotated and Interpreted by Sutherland EH). University of Chicago Press, Chicago

Sutherland EH (1973 [1936]) Juvenile delinquency and community organization. In: Edwin H, Schuessler K (eds) Sutherland: on analyzing crime. University of Chicago Press, Chicago

Sutherland EH (1983 [1949]) White-collar crime. The uncut version. Yale University Press, New Haven

von Römer L (1905) Het uranisch gezin. Wetenschappelijk onderzoek en conclusies over homosexualiteit. G. P. Tierie, Amsterdam

Wallerstein JS, Wyle CJ (1953 [1947]) Our law-abiding law-breakers. In: Clyde R, Vedder, Konig S, Clark RE (eds) Criminology. A book of readings. Dryden Press, New York

West DJ (1982) Delinquency. Its roots, careers and prospects. Heinemann, London

West DJ, Farrington DP (1973) Who becomes delinquent? Second report of the Cambridge study in delinquent development. Heinemann, London

Wolfgang ME, Figlio RM, Sellin T (1972) Delinquency in a birth cohort. University of Chicago Press, Chicago

History of the Statistics of Crime and Criminal Justice

Michael D. Maltz[1,2] and Kathleen Frey[3]
[1]Department of Sociology, Criminal Justice Research Center, The Ohio State University, Columbus, OH, USA
[2]Emeritus, Department of Criminal Justice and Department of Information and Decision Sciences, University of Illinois at Chicago, Chicago, IL, USA
[3]Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA

Overview

This entry provides an account of the history of statistics and its relation to criminal justice and criminology. It describes how the "Law of Errors," initially based on astronomical observations, was adopted to explain "social physics" and in particular, the relative constancy of crime. Analytic methods that were developed to uncover patterns in data, which were not apparent prior to the utilization of the new methods, are described. Frequentist and Bayesian views of statistics are discussed, as well as issues surrounding some of the analytical methods used in criminology and criminal justice, including the concept of "statistical significance"; these issues are often glossed over by those who apply them.

Introduction

When a researcher in criminology and criminal justice does statistical analysis nowadays, all too often it goes like this. S/he:

- Gets permission to access a data set – for example, the National Longitudinal Survey of Youth (NLSY), the National Longitudinal Study of Adolescent Health (AdHealth), and the Project on Human Development in Chicago Neighborhoods (PHDCN)
- Then develops a "model" [*about which* Freedman (1985: 348) *noted: In social-science regression analysis, usually the idea is to fit a curve to the data, rather than figuring out the process that generated the data. As a matter of fact, investigators often talk about 'modeling the data.' This is almost perverse: surely the object is to model the phenomenon, and the data are interesting only because they contain information about that phenomenon. Whatever it is that most social scientists are doing when they construct regression models, discovering natural laws does not seem to be uppermost in their minds*]
- Calls it into one of the four Ss (SAS, SPSS, Stata, Systat) or some other collection of canned statistical routines
- Tweaks the model
- Looks for an encouraging p-value and starts writing.

This process was not always the case. First, data were hard to come by, and the researchers themselves often had to laboriously collect them. Second, and the topic of this entry, there were very few methods available to perform the analyses. The early statisticians had to develop these methods – and perform the calculations without the aid of even a calculator.

This entry describes the history of statistics and its relation to criminal justice, the analytic methods that were developed to uncover patterns in the data, and issues surrounding the analytical methods. The first section describes the early history of statistics used in demography. Section 2 describes the development of the "error law." Section 3 then details how the "error law" was misapplied to criminal justice data. Sections 4 and 5 describe the development of new analytic methods in the early twentieth century as well as the development of statistical analysis in the field of criminology. Section 6 discusses how these methods lead to hypothesis testing and statistical significance.

Early History

The practice of searching extant records for patterns seems to have been initiated over three centuries ago by John Graunt (Hacking 1975: 102). Church records, not state records, were the original source of demographic information. While examining London's church records of births and deaths (called Bills of Mortality), Graunt generated mortality tables that calculated the likelihood of an individual surviving to any age. These statistics were then used to estimate the number of London males eligible for military duty. As a result of this use of the Bills of Mortality, a number of European governments mandated that they be collected and provided to them (Hacking 1975: 102). One of Graunt's contemporaries, William Petty, "was a man who wanted to put statistics to the service of the state" (Hacking 1975: 105); he termed this use of data "political arithmetic" (Porter 1986: 19). The word "statistics" (*statistik* in German) was coined by Gottfried Achenwall in 1749 (Hacking 1990: 24), and soon replaced Perry's invention.

The first to use the Bills of Mortality for analytic purposes appears to be John Arbuthnot, who gathered data on birth cohorts at the end of the seventeenth century. He found that the ratio of male to female births was stable across years and location and wrote a monograph entitled "An argument for Divine Providence, taken from the regularity observed in the birth of both sexes" (Stigler 1986: 225). In other words, he equated this particular year-to-year regularity of an unequal number of births between males and females as proof of the existence of God. Since men lived lives that likely led to an early death, more men needed to be born each year compared to women.

[Of course, a statistician would point out that tossing a coin provides the same regularity, but perhaps s/he would not attribute this regularity to divine providence. Not only that, but ironically it took a Czech monk, Gregor Mendel, to show (in 1865) that even in genetics, a probabilistic model would suffice.]

Even earlier, in 1713, Jacob Bernoulli sought to explain such regularities in more detail. He determined that as the number of trials increased, the uncertainty about an event's true proportion of successes would decrease. Bernoulli tested his propositions by considering an urn that contained a number of two different colored balls in regard to the chance of pulling a certain colored ball out of the urn. He later applied these ideas to real-life examples (Stigler 1986: 65). Instead of ascribing this regularity to divine providence, Bernoulli showed that as the number of births (or other binomial phenomena) increases, the true proportion will be approached, which is why there was stability in the proportion of births over 80 years of data (Stigler 1986: 64ff).

Twenty years later, Abraham De Moivre made a significant contribution to Bernoulli's work on binomial statistics. He developed a formula that approximated the binomial probability distribution for a fair coin ($p = 0.5$), which resembles the symmetrical normal distribution of today. He showed that as the number of observations increases, the binomial distribution becomes a smooth curve which is highest at the true proportion (p) and whose logarithm drops off proportional to the square of the deviation from the mean. This led him to develop a concept similar to the central limit theorem, which was later refined by Pierre-Simon Laplace in 1781 (Stigler 1986: 136). These advances eventually led to the consideration of the concept and characteristics of statistical distributions since, although a particular value of p was the most likely outcome, values near p were also very likely.

A parallel effort in matters probabilistic was undertaken by Thomas Bayes in 1763. His eponymous theorem showed the way to update an estimate of a probability distribution or set of statistical parameters after one has obtained new data. His formula,

$$\text{posterior probability} = \text{likelihood} \times \text{prior probability}$$

stands in contrast to the frequentist view of statistics, based on sampling theory, that provides a probability estimate of the *data* rather than of the parameters. As Lindley (1992: 360) noted,

> [T]he concept of a significance level or a confidence coefficient, although probability statements, are [sic] not about the hypotheses nor the parameters, but about the data. It is this failure to use probability for hypotheses and parameters that distinguishes the frequentist, sampling-theory approach to statistics from the fully-coherent Bayesian approach.

The prior probability refers to the initial belief of the probability of an event occurring, while the likelihood is the probability that other observations will occur given some new data, and finally the posterior probability is the new belief based on the prior and the likelihood. Over time, Bayes' theorem has been advanced by many mathematicians and statisticians and had undergone serious critique but yet never disappeared (McGrayne 2011). In criminology it has been used in estimating recidivism parameters (Maltz 1984: 150) and in studying rational choice and deterrence (Anwar and Loughran 2011).

The "Law of Errors"

[This section relies heavily on Stigler (1986).] Concepts related to statistical distributions were also found when considering measurement error. When repeated observations are made, for example, of the angle between the horizon and a star for navigational purposes, they rarely produce the same number. Dealing with the variation in these observations was an important statistical issue in the eighteenth century, especially for seaborne navigation out of sight of land. Building on De Moivre's work, Thomas Simpson decided not to concentrate on the distribution of the *observations* but rather on the *deviations* between the observations and the true value or the errors (Stigler 1986: 90). While this may seem obvious now, at the time this was a major step forward since it moved away from the actual event, the observation itself, to an abstract quantity, the error, and its distributional properties.

Simpson made a number of simplifying assumptions about the error distribution and about the law of large numbers but was able to provide useful advice to astronomers that improved the way they dealt with error. But he (and Laplace after him) showed that the error curve was relatively symmetrical, and as the size of the error increases, the probability of it being that large should decrease monotonically away from zero.

Subsequent work by Carl Friedrich Gauss and Adrien-Marie Legendre showed that the mean of a series of observations minimized the sum of squares of the errors from the mean. Further, Gauss posited that the normal distribution (also known as the Gaussian distribution) was the appropriate distribution to use. His argument

> was essentially both circular and non sequitur. In outline its three steps ran as follows: The arithmetic mean (a special, but major, case of the method of least squares) is only 'most probable' if the errors are normally distributed; the arithmetic mean is 'generally acknowledged' as an excellent way of combining observations so that errors may be taken as normally distributed (as if the 'general' scientific mind had already read Gauss!); finally, the supposition that errors are normally distributed leads back to least squares. (Stigler 1986: 141)

The Error Law, Crime, and "Social Physics"

In the nineteenth century, a Belgian physicist, Adolphe Quetelet, began to apply these new statistical tools to social data. France had begun to collect criminal justice statistics in 1825, and Quetelet noted the relative constancy in the number of persons accused of crime and the percent convicted (Stigler 1986: 179). Of course, there may have been other reasons for this constancy – police, prosecutorial, or judicial manpower limitations, arrest quotas, and a limited prison capacity – but these were not considered.

Instead, Quetelet saw in this statistical regularity "signs of a deeper social reality" and wrote of the "average moral man (*l'homme moyen morale*)." While considering how the repeated flipping of a coin with probability p would produce a distribution of outcomes centered on p, he posited that the same would hold for humans; there was a fixed value for social phenomena as well, deviations from this value being error. He surmised that "[I]f an individual at any given epoch of society possessed all the qualities of the average man, he would represent all that is great, good, or beautiful" (quoted in Stigler 1986: 171). And based on a mistake in the data, he noted a gradual decrease in the annual conviction rate and assumed that the population was getting more moral over time.

Simeon-Denis Poisson had a different view of the conviction process. Using corrected data, he modeled jury decision-making as a binomial process and found that the variation in rates over time was "not so large as to support a belief that there has been a notable change in the causes" (Stigler 1986: 190). In other words, the year-to-year variation was not, in current terminology, "statistically significant."

Back to Quetelet essentially inverted the logic behind the "error law." That is, he implicitly assumed that if errors are generally normally distributed, then whenever we find a distribution that is normally distributed we can safely assume that there is an underlying "truth." This "truth" is a trait common to all persons but corrupted by errors. Quetelet called this truth a population's *penchant au crime*: he wrote "Society prepares the crimes and the guilty person is only the instrument," implying that "free will exists only in theory" (Hacking 1990: 116).

This idea that a population had fixed social characteristics, and that individual differences were due only to natural variation, continued to be believed. Emile Durkheim also concluded that there were social laws that acted on all individuals "with the same inexorable power as the law of gravity" (Hacking 1991: 82; see also Duncan 1984: 99). For example, he wrote of "suicidogenetic currents" in society, as if there was an innate propensity of the same magnitude for suicide within every individual (Hacking 1990: 158).

Others, especially Francis Galton (1889), criticized this approach. He saw variation as natural and desirable and criticized those who focused on averages as being "as dull to the charm of variety as that of the native of one of our flat English counties, whose retrospect of Switzerland was that, if its mountains could be thrown into its lakes, two nuisances would be got rid of at once" (quoted in Porter 1986: 129). Yet Galton's view was not shared by most of the social scientists of the time. In fact, there was widespread belief that those far from the mean were deviant in the pejorative sense of the word and that they were "errors." Statistics was not seen then so much as the science of variation (Duncan 1984: 224) as it was the science of averages.

Vestiges of this belief continue to this day (Savage 2009). In criminology, the General Theory of Crime (Gottfredson and Hirschi 1990) attributes all criminality to be generated by a single cause: lack of self-control. A lack of self-control may indeed lead a pedophile, a person engaged in insider trading, a serial killer, a drug dealer, and a pickpocket to perform criminal acts. However, making that assertion is no more useful than attributing all cancer to the same cause, uncontrolled cell growth: bone, breast, colon, lung, and prostate cancer have different etiologies, and each should be dealt with differently. They should not be looked upon as variations on a single theme.

Early Analytic Methods

Following in Galton's footsteps, Karl Pearson made further improvements in statistical methods and helped to bring statistics into the social sciences. Pearson, in 1900, embraced the centrality of the normal curve in statistical theory by asserting that any measurement that was three standard deviations from the mean was significantly different from the mean. This new concept became known as the "rule of three." Significance levels led way to early conceptions of hypothesis testing, which would eventually lead to the multitude of tests we have today (Ziliak and McClosky 2008).

In the 1920s Ronald Fisher introduced the concept of random sampling and controlled experiments. He accounted for individual variation and argued that control and experimental groups were needed to test treatments. The two groups needed to be generated by random assignment to one or the other from the study population, since each subject should have an equal and independent chance of being selected for either group. This would make it very likely (but not certain) that a representative sample populated both experimental and control groups.

Fisher also placed guidelines on determining a sample size. It is necessary to have a certain sample size based on characteristics of the groups in order to get reliable results. Nowadays power analysis software is used to determine the appropriate sample size for each group.

Fisher also promoted formal hypothesis testing and generated statistical tables for use by researchers to determine the level of "significance" of a finding. The "rule of 3" suggested by Pearson was not sufficient, and Fisher recommended setting an alpha level (often at 0.05) to determine "statistical significance." The alpha level is the probability that the hypothesis (of "no difference" between, say, experimental and control groups) is correct. If this hypothesis is rejected, then it is concluded that there is a difference.

There is a controversy over significance testing that continues to this day (e.g., Cohen 1990, 1994; Loftus 1993; Maltz 1994a; Savage 2009; Wilkinson et al. 1999; Ziliak and McClosky 2008), due in part to the use of the term "significance." Alfred Blumstein has suggested that the word "significance" be replaced with "discernibility," since the nonstatistical meaning of the word "significance" is "importance"; many *statistically* significant findings are far from *substantively* significant, and what the tests actually show is the extent to which one measurement is discernible from another.

Experiments can be costly, and one cannot always generate a large enough sample to invoke the central limit theorem and the assumption of normality. William Gosset found this to be the case while working at the Guinness brewing company. Since his employers did not want their competitors to see how his statistical findings could benefit them, Gosset published his findings under the name "Student", in 1908. He found that he could not use the standard z-score tables, which were based on the normal distribution. Instead, he developed a distribution useful for hypothesis testing with a small sample. Later on, Fisher would adjust the formula, and the Student's t-distribution was created (Ziliak and McClosky 2008).

The accomplishments of these statisticians should not be underestimated. They had to deal with data that needed to be collected, often by themselves, with no computing aids except for statistical tables compiled with an enormous expenditure of time and patience. They were positing crude models that they hoped would give them insights into patterns of complicated processes. Their focus on means and variances was doubtless due to the fact that in order to develop findings of any value, they had to use summary statistics to characterize the distributions. That they succeeded so well in these endeavors is a testament to their perseverance and ingenuity.

The Twentieth Century and the Birth of Criminology

In the 1920s, however, another way of looking at data was developed by University of Chicago sociologists (the "Chicago school" of sociology).

Rather than base their analyses on a data set's distributional properties (mean, standard deviation), they focused on its geographical variation. Two types of data emerged from this era: neighborhood data and ethnographic data.

Building on the work of Park and Burgess in *The City* (1925), Shaw and McKay (1942) collected data on the variation in juvenile delinquency rates in different neighborhoods in Chicago over several time periods and mapped the data. They then looked at the distribution of these rates and noticed that neighborhoods nearest the city center had the highest rates of delinquency, and as one moved away from the center, crime rates declined. In order to understand the reasons for the varying crime rates, they also gathered data on residential mobility, racial heterogeneity, and socioeconomic status. Shaw and McKay found that in areas with high crime rates, there was a large amount of residential turnover, racial heterogeneity, and low socioeconomic status, all of which decreased in places farther from the city center. Their findings supported the ecological perspective proposed by Park and Burgess. The methods of Shaw and McKay were highly influential and opened up the use of qualitative analysis in criminology (Noaks and Wincup 2004). It set the stage as well for geographical analyses in criminology, including computer-based crime mapping (Maltz et al. 1991) and hot spot analysis (Sherman and Weisburd 1995).

Another method of analysis used by the Chicago school was ethnographic analysis. In particular, the Chicago school is known for its use of life histories. In this method, a person is interviewed about his/her own personal history to understand the events in life that may have triggered crime and delinquency. One well-known life history study is *The Jack Roller* (1930) by Shaw, who interviewed a boy that had been institutionalized. Using official records, Shaw conducted the interview to fill in the story around each police interaction. Eventually an in-depth story was put into a book that followed this boy's delinquent career. This type of method allows the criminologist, or any researcher, to reach new causal hypotheses than can be further tested with data. Although ethnographies are qualitative and not statistical analyses, they can help derive new theoretical insight that can be studied with quantitative methods (Sampson and Laub 1993).

Hypothesis Testing and Statistical Significance

There has been a growing discontent with the (over)use of hypothesis testing and the search for statistically significant findings in the social sciences (e.g., Cohen 1990, 1994; Loftus 1993; Maltz 1994a). As a consequence, the American Psychological Association formed a committee, the Task Force on Statistical Inference (TFSI), to deal with it (Wilkinson et al. 1999), and various books (Lieberson 1985; Savage 2009; Ziliak and McClosky 2008) have taken up the cause as well.

In addition, there is growing evidence that many such findings may not be replicable, and often the initial finding turns out to be less "significant" due to regression to the mean. Moreover, journals are not interested in insignificant findings since they "show nothing." Unfortunately, many interesting articles go unpublished that could have useful findings, since a null finding may have its own policy implications (Lehrer 2010).

These hypothesis tests are often based on a comparison of mean values. Using the mean assumes that there is a unimodal distribution based around the average of the sample; however, it is often the case that more than one type of behavior is present. [For example, there are many different homicide syndromes (Block and Block 1995: 29), each of which would have its own set of characteristics.] Nowadays there are a number of methods and techniques to deal with this, but these tests were first proposed when the costs of data collection and analysis, both in time and money, were substantially greater than today. Analyses that took days to accomplish can now be done in microseconds.

Moreover, the data analysis landscape has changed over the past few decades. We now have oceans of data to analyze, collected for us by criminal justice agencies and other organizations. Additional techniques, based on exploratory data analysis (Tukey 1977), have been brought to bear on data. A wealth of computer programs is at our disposal to facilitate the use of scores of analytic and visualization methods. As Tukey (1992: 444) noted, "If one technique of data analysis were to be exalted above all others . . ., there is little doubt which one would be chosen. The simple graph has brought more information to the data analyst's mind than any other device."

These innovations have brought data analysis into the mainstream of social science, while at the same time, they have made it so easy to deal with data that the assumptions that undergird the statistical methods are too often ignored. As described above, there is now a great deal of data available for analysis using tried-and-true (as well as tired and not-quite-true) methods.

Summary

As noted above, the data landscape has changed considerably over the past few decades. Courtesy of the US Department of Justice, we now have the FBI's Uniform Crime Reports (UCR, from 1960) and the Bureau of Justice Statistics' (BJS) National Crime Victimization Survey (NCVS, from 1973), each of which has provided researchers with useful information about the study of crime, criminal behavior, and official (and unofficial) responses to these acts. And courtesy of other federal agencies and other grant-giving organizations, we also have other survey-based data sets (NLSY, AdHealth, PHDCN) to test hypotheses and find relationships. Having these data sets publicly available provides researchers with the raw material on which to apply their statistical tools.

The statistical/methodological picture has also changed over the same time period, although not as quickly nor as much. Many still cling to methods developed long ago when reliable data were hard to come by. As can be seen by perusing the table of contents of this encyclopedia, a host of new ways of analyzing data has been developed more recently.

It is important to understand the premises and assumptions that undergird whichever method one uses in research. Although very useful and innovative when they were first developed, hypothesis testing and its partner statistical significance may have outlived (most of) their usefulness. Their limitations, described in the cited papers and books by Cohen, Leiberson, Loftus, Maltz, Wilkinson, and Ziliak and McCloskey, are too often ignored by those who apply them. It is hoped that this review of their origins will lead to more their careful application and that greater (and appropriate) use is made of the methods described in other entries.

Acknowledgments The authors thank Laura Dugan for her helpful comments. This entry borrows liberally from Maltz (1994a, b).

Related Entries

▶ Development of the UCR and the NCVS

Recommended Reading and References

Anwar S, Loughran T (2011) Testing a Bayesian learning theory of deterrence among serious juvenile offenders. Criminology 49:667–698

Block C, Block R (1995) Trends, risks, and interventions in lethal violence: proceedings of the third annual spring symposium of the Homicide Research Working Group. National Institute of Justice, US Department of Justice, Washington, DC

Cohen J (1990) Things I have learned (so far). Am Psychol 45:1304–1312

Cohen J (1994) The earth is round (p <.05). Am Psychol 49:997–1003

Duncan OD (1984) Notes on social measurement: historical and critical. Russell Sage, New York

Freedman DA (1985) Statistics and the scientific method. In: Mason WM, Fienberg SE (eds) Cohort analysis in the social sciences: beyond the identification problem. Springer, New York, pp 343–366

Gottfredson M, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Hacking I (1975) The emergence of probability. Cambridge University Press, Cambridge

Hacking I (1990) The taming of chance. Cambridge University Press, Cambridge

Hacking I (1991) How shall we do the history of statistics? In: Burchell G, Gordon C, Miller P (eds) The Foucault

effect: studies in governmentality. University of Chicago Press, Chicago, pp 181–195
Lehrer J (2010) The truth wears off. New Yorker. 13 Dec 2010; available online at http://www.newyorker.com/reporting/2010/12/13/101213fa_fact_lehrer
Lieberson S (1985) Making it count: the improvement of social research and theory. University of California Press, Berkeley, California
Lindley DV (1992) Introduction to Good (1952) rational decisions. In: Kotz S, Johnson NL (eds) Breakthroughs in statistics, volume I: foundations and basic theory. Springer, New York, pp 359–364
Loftus GR (1993) A picture is worth a thousand p-values: on the irrelevance of hypothesis testing in the computer age. Behav Res Meth Instrum Comp 25:250–256
Maltz M D (1984) Recidivism. Academic, Orlando. Internet version published in 2001; available at http://osu.academia.edu/MichaelMaltz
Maltz MD (1994a) Deviating from the mean: the declining significance of significance. J Res Crime Delinq 31:434–463
Maltz MD (1994b) Operations research in studying crime and justice: its history and accomplishments. In: Pollock SM, Barnett A, Rothkopf MH (eds) Operations research and the public sector (Chapter 7), Volume 6 of the handbooks in operations research and management science, edited by GL Nemhauser and AHG Rinnooy Kan. Elsevier, Amsterdam/Netherlands, pp. 201–262
Maltz MD, Gordon AC, Friedman W (1991) Mapping crime in its community setting: event geography analysis. Springer, New York. Internet version published in 2000; available at http://osu.academia.edu/MichaelMaltz
McGrayne SB (2011) The theory that would not die: how Bayes' rule cracked the Enigma code, hunted down Russian submarines & emerged triumphant from two centuries of controversy. Yale University Press, New Haven
Noaks L, Wincup E (2004) Criminological research: understanding qualitative methods. Sage, London
Park R, Burgess E (1925) The city. University of Chicago Press, Chicago
Porter T (1986) The rise of statistical thinking. Princeton University Press, Princeton
Sampson RJ, Laub JH (1993) Crime in the making: pathways and turning points through life. Harvard University Press, Cambridge
Savage S (2009) The flaw of averages: why we underestimate risk in the face of uncertainty. Wiley, New York
Shaw C (1930) The jack roller. University of Chicago Press, Chicago
Shaw C, McKay H (1942) Juvenile delinquency and urban areas. University of Chicago Press, Chicago
Sherman LW, Weisburd D (1995) General deterrent effects of police patrol in crime hot spots: a randomized, controlled trial. Justice Quart 25:625–648
Stigler SM (1986) The history of statistics: the measurement of uncertainty before 1900. Harvard University Press, Cambridge
Tukey JW (1992) The future of data analysis. In: Kotz S, Johnson NL (eds) Breakthroughs in statistics, volume II: methodology and distribution. Springer, New York, pp 408–452
Tukey JW (1977) Exploratory data analysis. Addison-Wesley Publishing Company, Reading, Massachusetts
Wilkinson L, Task Force on Statistical Inference (1999) Statistical methods in psychology journals: guidelines and explanations. Am Psychol 54:594–604
Ziliak S, McClosky D (2008) The cult of statistical significance: how the standard error costs us jobs, justice, and lives. University of Michigan Press, Ann Arbor

HIV

► Drug Abuse and Alcohol Dependence Among Inmates

HIV in the Correctional System

Anne C. Spaulding, Shawnta L. Lloyd and
Lauren C. Messina
Rollins School of Public Health, Emory
University, Atlanta, GA, USA

Overview

The United States has witnessed a significant rise in the number of incarcerated men and women since the 1970s. In the more recent years of this "epidemic of incarceration," the epidemic of the human immunodeficiency virus (HIV) has hit the correctional system especially hard (Spaulding et al. 2002). Many persons first learn of their HIV infection through jail- or prison-based screening (de Voux et al. 2012). Among Americans living with HIV, about one in five are unaware of their HIV status (Centers for Disease Control and Prevention 2010; Centers for Disease Control and Prevention 2012). As the HIV prevalence among US prisoners is more than three times higher than that of the general population (Maruschak 2008), implementing more HIV testing in correctional facilities would provide an opportunity to identify new cases of HIV.

Integrating HIV care into standard medical care in all healthcare venues "normalizes" the disease and may encourage persons with known HIV diagnoses to disclose their status. Additional programs are needed to link HIV-positive individuals to HIV treatment and other medical care, as well as to HIV/AIDS education, both in correctional facilities and upon release. The criminal justice system has the opportunity to not only promote public safety, but to lessen health disparities found in the populations passing in and out of correctional facilities.

Dual Epidemics: Incarceration and HIV

The United States has the highest incarceration rates in the world (International Centre for Prison Studies), accounting for approximately 2.3 million Americans serving time behind bars (Glaze 2010). In comparison to the general population, individuals in correctional facilities are more likely to be living with mental illness, tuberculosis, viral hepatitis, substance use disorders, and/or sexually transmitted diseases (STDs) (Rich et al. 2011). Incarcerated individuals are considerably more likely to be living with HIV disease at all stages, including late stage and symptomatic HIV [otherwise known as acquired immunodeficiency syndrome (AIDS)] compared to the general population; approximately 17 % of all Americans living with HIV pass through correctional facilities at least once in a given year (Spaulding et al. 2009).

Minorities are over-represented in both the correctional population and the HIV-positive population. Non-Hispanic Blacks and Latinos are represented disproportionately among those inmates held in prisons and jails. In addition to exceedingly high incarceration rates, Blacks and Latinos are disproportionately affected by HIV. Although Blacks represent only 14 % of the total US population, they account for 46 % of people living with HIV. Latinos represent 16 % of the total US population and account for 17 % of people living with HIV (Centers for Disease Control and Prevention 2012). HIV infection affects those incarcerated as well as the communities to which these people return.

HIV can be transmitted through sexual fluids or blood, placing individuals who participate in unprotected sex, particularly unprotected anal sex, or injection drug use at an increased risk of becoming infected with HIV in comparison to those who do not. When the HIV epidemic began in the United States, gay, primarily white, men were the first persons infected with the virus. As time went on, the epidemic spread to those who participated in injection drug use and unprotected heterosexual sexual intercourse. Additionally, HIV infections were identified among blood transfusion recipients and newborns with an HIV-positive mother. Many injection drug users lived in urban communities and were well acquainted with the criminal justice system. During the time of the HIV transition into a new population, the United States criminal justice system was ballooning, due to the declared "war on drugs" (Spaulding et al. 2002). Drugs users were incarcerated at rising rates, increasing the number of HIV-infected persons in correctional facilities.

HIV Screening

Why Is Screening Necessary?

Most incarcerated men and women who are infected with HIV acquired their infection while in the community, before entering a correctional facility (Jurgens et al. 2009; Centers For Disease Control and Prevention 2009). High-risk behaviors such as unprotected sexual encounters and drug use and lifestyles linked with HIV are common among persons who subsequently enter a correctional facility (Braithwaite and Arriola 2003). Substance abuse besides injection drug use can lead to disinhibiting behavior which promotes high-risk behavior, making alcohol, cocaine, and methamphetamine abusers as vulnerable to HIV infection as injection drug users. Social factors epidemiologically associated with high HIV risk include poverty and homelessness (Braithwaite and Arriola 2003). HIV testing of asymptomatic persons has the ability to identify cases that may otherwise go undiagnosed and untreated. Since 18–20 % of Americans are unaware of their HIV status (Centers for Disease Control and Prevention 2010; Centers For Disease Control and Prevention 2012) and the incarcerated population are known to participate in high-risk behaviors associated with HIV infection, there may be a large number of undiagnosed HIV-positive individuals in correctional facilities.

After learning their status, HIV-positive persons tend to decrease unsafe sexual risk behavior and needle sharing; the prevalence of unprotected anal and vaginal intercourse can drop by more than 50 % after HIV diagnosis (Marks et al. 2005). The same phenomenon has been observed among inmates who tested HIV positive during incarceration; receipt of an HIV diagnosis is associated with lower rates of anal intercourse among male inmates (Jafa et al. 2009). HIV testing in correctional facilities, either during or after entry, not only benefits those incarcerated, but also those in the community as well. Partner Services conducted by the local health department (Altice et al. 2010) may be able to locate former sexual and injecting drug partners of incarcerated HIV-positive individuals, allowing those in the communities to seek testing; if positive, these individuals can obtain clinical care and HIV/AIDS education. This is an opportunity to decrease the transmission of HIV/AIDS and to increase HIV/AIDS knowledge in the communities, extending the potential benefits of HIV testing in the correctional system.

Testing Procedures

The Center for Disease Control and Prevention (CDC) published the *HIV Testing Implementation Guidance for Correctional Setting*s in 2009. The CDC recommends the adoption of a universal, routine, opt-out testing approach to improve HIV detection among the incarcerated population. This approach includes incorporating consent for HIV treatment into the general consent form for medical services within the institution, meaning that all inmates should be tested for HIV unless an inmate refuses to have the test completed. Studies of jail detainees newly diagnosed with HIV reveal that the majority had no risk for HIV acquisition other than heterosexual sexual relations (Macgowan et al. 2009). Thus, limiting HIV screening to those who report injection drug use or men who have sex with men is less effective in identifying HIV cases compared to universal testing. A large proportion of individuals moving through the correctional system may be offered screening and tested before being released into the community or transported to another facility undiagnosed. Inmates should be given information about HIV/AIDS before testing occurs, and the decision to test must be noncoercive (Centers For Disease Control and Prevention 2009). Medical staff should note when an inmate declines HIV testing because he or she may accept a repeat offer in a subsequent medical encounter. Reasons why inmates may refuse testing include fear of needles, conflict with another activity, lack of knowledge of risk behaviors, and/or an upcoming release date (Altice et al. 2010). Routine opt-out testing has the ability to standardize HIV testing in correctional facilities, reducing the stigma and fear associated with HIV testing. Universal opt-out testing may not be perceived as an option in some facilities when budgets are tight or where custody staff are unable to provide security services for a prolonged medical encounter at entry. Alternatives to the optimal approach of universal testing by nursing at entry include risk-based screening or screening during subsequent medical encounters. Risk-based screening constitutes offering HIV testing to inmates who have had high-risk characteristics (injection drug use, men who have sex with men, sex with an at-risk partner, transactional sex, multiple sex/drug partners, and STDs) within the last 12 months. These approaches allow correctional facilities to test those individuals that may be at the highest risk of HIV infections. However, as noted above, an important study showed that the most common risk factor for undiagnosed HIV among those entering a correctional facility is heterosexual sexual behavior (Macgowan et al. 2009). Thus, targeted testing will miss most infected persons.

Traditional testing consists of an HIV enzyme immunoassay (EIA). When traditional HIV testing is performed, the specimen is processed in an external laboratory; it can take several days for the result to be returned to the staff of the correctional facility. It is still a practical option for inmates with a longer length of stay such as those found in prisons. As of mid-2012, any positive HIV screening test is followed by a second confirmatory test using a Western blot assay (Centers for Disease Control and Prevention 2009). The algorithm for confirming positive tests may be changing in the near future.

In correctional facilities with rapid population turnover such as jails, new "rapid HIV tests" can offer several advantages. Most rapid tests have been approved to be used at the "point of care"; processing does not require sending the specimen to a laboratory. The median length of stay in many jails is 48 h or less (Spaulding et al. in press-a), which means half of the population may have left the facility before conventional HIV tests return. Training for conducting rapid, point-of-care tests is not extensive. With an oral swab or a finger stick for a rapid test, medical staff can complete the HIV tests in as little as 10–30 min (Altice et al. 2010). Medical staff are therefore able to quickly inform inmates of their HIV status shortly after the testing, rather than bringing the person back to the healthcare setting at a later time. Rapid testing offers the advantage of producing a preliminary result quickly; further testing can take place during the same visit, at the clinician's discretion.

Some facilities do not have access to resources to offer opt-out testing and linkage to appropriate medical care. Partnerships with local health departments and/or community-based organizations can aid in carrying out HIV testing and counseling in correctional facilities where further resources are needed. Additionally, academic institutions have the capacity to assist with testing and clinical care. Professionals with expertise in various disciplines such epidemiology, sociology, and criminology can provide the criminal justice system with data collection and analytic skills to aid in the improvement of testing and clinical care (Rich et al. 2011).

Testing Permits Treatment and Prevention

HIV in the incarcerated population is diagnosed at a younger age and earlier stage of disease than in the general population, factors leading to better HIV/AIDS health outcomes (Spaulding et al. 2002; de Voux et al. 2012). Prompt diagnosis after infection and treatment has the ability to save lives. HIV testing is crucial in the correctional system to identify HIV-positive cases or allow individuals to reveal their HIV-positive status, if previously diagnosed. Upon diagnosis, treatment options can be explored and safer behaviors can be implemented to slow disease progression and prevent transmission to others, leading to longer, healthier lives.

Once an HIV-positive status is known, the first step to treatment is HIV education to ensure that the inmate is aware of necessary lifestyle changes and the importance of adherence to treatment (U.S. Department of Health and Human Services 2011). Before initiating antiviral medications, untreated infected persons should have their virus tested to determine what drugs would be most likely to suppress it. This test of susceptibility should be reviewed with a healthcare provider experienced in treating HIV to determine the appropriate therapeutic regimen. In some cases, HIV-infected persons may need to be referred to an outside HIV specialist depending on the complexity of the inmate's medical issues. Medical staff in some correctional facilities may lack HIV specialists or persons with adequate HIV experience and education on site; these facilities can develop links to HIV specialists in the community and contract with providers to provide the necessary consultative care (Centers for Disease Control and Prevention 2009). Those inmates who disclose their known HIV-positive status and have been prescribed antiretroviral medication should continue or resume treatments under the direction of a provider with HIV experience. Previous medical records should be consulted.

The two methods of administering HIV medications in facilities are directly observed therapy (DOT) and keep-on-person (KOP) therapy. The DOT system requires staff members to administer medications; patients are known to have better adherence using this system, but an inmate may fear the loss of confidentiality. Conversely, KOP therapy requires an inmate to be responsible for administering his/her own medication: the inmate is not monitored by the medical staff and adherence to medical treatment is often unknown. However, this method can aid in adherence of HIV treatment after release (U.S. Department of Health and Human Services 2011).

Since many inmates have co-occurring illnesses, the treatment of disease can become complex. Adherence to HIV medication is essential to decrease viral loads and increase CD4 lymphocyte (CD4+) counts. It is common for inmates, including HIV-positive inmates, to have multiple medical conditions such as substance abuse or mental health conditions. Inmates should receive treatment for co-occurring issues while receiving HIV medication to ensure that they are in a state of mind to understand how to properly take the medication and maintain a healthy lifestyle. Additionally, medical staff need to learn how to identify side effects to HIV treatment and schedule follow-up visits during the first weeks of therapy to monitor adherence and potential side effects (U.S. Department of Health and Human Services 2011).

Treatment: Specific Issues for Prisons

Correctional facilities have become a significant provider of HIV medical care and treatment due to the high number of HIV-infected persons transitioning through the criminal justice system. Additionally, HIV-positive persons incarcerated in prisons have sentences longer than a year unlike those in jails, providing adequate time to receive testing, results, and antiretroviral therapy. The introduction of highly active antiretroviral treatment (HAART) in correctional facilities has aided in the suppression of viral loads and increased CD4+ counts in HIV-infected inmates (Springer and Altice 2005). Sustaining HIV treatment allows HIV-infected individuals to achieve undetectable viral loads, which can greatly decrease the risk of HIV transmission to sexual or drug injection partners. An increased CD4+ count in an HIV-infected person allows preservation of the immune system to help the infected person avoid life-threatening medical conditions. Accordingly, AIDS-related deaths among all deaths in state prisons decreased from 34.2 % to 4.6 % between 1995 and 2006 (Maruschak 2008). This massive decrease in AIDS-related deaths mirrors the decrease of HIV-/AIDS-related deaths nationwide.

Incarceration can be an opportunity for prisoners to receive quality healthcare from medical staff; adherence to medical treatment may be facilitated in a controlled environment. When individuals in correctional facilities are provided with adequate care, health outcomes can be comparable to those in the community setting; however, due to the variation of healthcare services currently offered in correctional facilities, health outcomes vary from facility to facility. A study completed in Texas found that only one-third of HIV-infected inmates who were eligible to be treated based on the Department of Health and Human Services criteria for initiation of HAART were on HAART therapy (Baillargeon et al. 2000). While treatment of HIV infection in correctional facilities can be successful, inmates have the right to refuse treatment and not all inmates take advantage of the possible medical treatments for reasons such as adverse side effects or poor prognosis.

Diagnosis and Treatment Is Feasible in Jails

HIV testing and treatment rates in jails tend to be lower than those in prisons (Culbert 2011). However, due to the success rates observed in some prison settings, jails have the potential to produce successful declines in HIV complications. On a given day, 750,000 persons are incarcerated in jails, most only staying for a short period of time (Spaulding et al. 2009). Jails provide a setting to offer HIV testing and treatment, if necessary, to thousands of persons daily. Inmates may not have access to adequate healthcare in the community and may not pursue testing on their own. Without testing, undiagnosed HIV-positive individuals will continue to engage in risky behaviors and may transmit the virus while incarcerated and/or in the community.

Given the rapid turnover of inmates and chaotic environment of jails, traditional HIV testing was challenging, but rapid testing has made HIV testing feasible in jail settings. However, if a rapid test produces a positive result, results of confirmatory testing may take 1–10 days to arrive at a correctional facility. Some inmates may be released from jail by the time the test result returns. However, community-based organizations can aid correctional facilities in contacting former inmates regarding their HIV status and offer linkage to medical treatment, if needed. An oral fluid HIV test can offer an alternative to blood tests, especially for persons who fear needles; confirmatory results will arrive in 3–5 business days (Altice et al. 2010).

In 2003, the states of Florida, Louisiana, New York, and Wisconsin received funding from the CDC to conduct HIV testing in jails. These four programs successfully performed 33,211 tests and identified 269 new HIV cases. With the use of rapid HIV testing, 99.9 % inmates received their test results. These testing programs also demonstrated the feasibility of universal opt-out testing. Of those who received a reactive rapid HIV test, rates were particularly high among persons who did not report risk behaviors (3.1 %), including injection drug use, male-to-male sex, transactional sex, sexual assault, STDs, heterosexual behavior, or sex with an at-risk partner. Incarcerated men who did not report risk behaviors were more than four times likely to be diagnosed with HIV than incarcerated men who reported heterosexual behavior. Therefore, new cases may be missed if inmates were only offered testing based on risk behaviors. A total of 440 known and newly diagnosed HIV-positive inmates with a reactive rapid HIV test were referred to medical care, treatment, and prevention services (Macgowan et al. 2009), demonstrating that HIV testing and linkage to medical care and treatment is feasible in the fast pace environment of jails.

In a jail setting, inmates may be released soon after being diagnosed as HIV positive; therefore, inmates may not have the opportunity to receive a pretreatment evaluation and begin antiretroviral treatment in jail. Linkage to HIV medical care and treatment in the community is critical to ensure that these HIV-positive individuals begin treatment as soon as possible after release. For those individuals with a longer length of stay in jail, pretreatment evaluation can be completed and the appropriate antiretroviral treatment can begin. Many newly diagnosed HIV-positive inmates start HIV treatment while in a correctional facility, receiving HIV medical care for the first time. Additionally, inmates and former inmates may be able to resume HIV therapy previously interrupted in jail and through linkage to care after discharge. Considering the short length of stay for most jail inmates, the cost of treatment for those with a longer length of stay should be minimal (de Voux et al. 2012).

Trust and Confidentiality

Although HIV/AIDS diagnosis in the correctional system has its benefits, some inmates may delay or refuse HIV/AIDS-related services in correctional facilities for reasons such as mistrust of staff and/or fear of discrimination from peers. Routine and opt-out testing has the capacity to reduce the stigma associated with HIV/AIDS. Inmates can be provided with literature on various health topics to deter others from learning about their medical conditions. In many facilities, HIV testing is provided only upon request. In a harsh environment such as a correctional facility, inmates may not feel comfortable requesting an HIV test. Inmates may feel more comfortable releasing medical information and completing care during medical assessments when the testing is common practice in correctional facilities (Centers for Disease Control and Prevention 2009).

HIV treatment in correctional facilities can lead to issues regarding confidentiality, as inmates have little privacy while incarcerated. It is important for facilities to create methods to guarantee inmates' medical confidentiality during HIV treatment delivery. In a setting where HIV may be stigmatized and the risks of violent acts are high, inmates may not feel safe while disclosing and receiving medical care from the medical staff. To receive medications, inmates may have to be escorted to another area in the correctional facility. When HIV-positive inmates are segregated, adherence to medication may be reduced and confidentiality compromised. Methods as small as allowing the doors to be closed while medical staff are disclosing information to the inmates may decrease the amount of information overheard by other staff and inmates as well as gain the trust of the inmates. All medical records should be under lock and key with access to only select personnel. The trust between the staff and inmates is vital to encourage the treatment of HIV and persuade HIV-positive inmates to initiate and adhere to the recommended treatments during their sentence and when released back into the community (Centers for Disease Control and Prevention 2009; Wakeman and Rich 2010). Ultimately, the medical staff and the custody staff must work as a team to keep inmates being tested for HIV safe.

HIV Prevention Inside Correctional Facilities

More risk behaviors usually take place in the community than while incarcerated (Braithwaite and Arriola 2003); however, the risk behaviors and potential to transmit HIV and other infectious agents to others during incarceration should not be overlooked. A study completed in Georgia found new cases of HIV during incarceration among inmates who tested negative during entry HIV testing (Jafa et al. 2009). There is evidence that transmission of HIV occurs in correctional facilities; however, the frequency of its occurrence and method of transmission is not absolutely known. The incidence of HIV transmission may vary in different facilities.

Existing data and research provide evidence that high-risk behaviors associated with HIV, including sexual intercourse and injection drug use are present in correctional facilities, even though these activities may be banned. The World Health Organization and the Joint United Nations Programme on HIV/AIDS strongly recommend the use of harm reduction strategies in correctional facilities. Condom distribution in correctional facilities has the capacity to reduce HIV transmission during penetrative sexual intercourse. Additionally, distribution of clean needles and syringes or diluted bleach along with instructions for cleaning injection equipment has the ability to reduce HIV transmission during drug injection, tattooing, and skin piercing (The World Health Organization and The Joint United Nations Programme on HIV/AIDS 2006). Most US correctional facilities have resisted adopting HIV prevention policies and practices such as condom distribution, needle and syringe programs, and decontamination strategies for injection equipment to reduce the transmission of HIV, because adoption of the UN standards is perceived as condoning such behavior. Items such as condoms and needles are viewed as contraband in jails and prisons, and inmates are not allowed to possess them while incarcerated (Tucker et al. 2007; Jurgens et al. 2009). Condoms are available in state prisons in only two states, Vermont and Mississippi, and in jails in five large urban jails (New York City, Philadelphia, San Francisco, Los Angeles, and Washington D.C.) (Braithwaite and Arriola 2003). Controversy surrounds the issue of providing inmates with harm reduction strategies such as condoms and clean needles due to the conflict between risk behaviors and rules in correctional facilities. Many administrators of correctional facilities do not want to acknowledge that risk behaviors continue to take place in their facilities since rules have been established in an attempt to prevent such occurrences. Even though some correctional officials may be aware of the needs for HIV prevention and medical care among the incarcerated population and wish to provide these populations with harm reduction equipment and knowledge, limited resources and decreased funds may replace the health of inmates with custody and security issues on the list of priorities (Braithwaite and Arriola 2003; Tucker et al. 2007). To decrease the concern that some HIV prevention programs may contradict the rules and policies of the correctional system, it has been recommended that messages and education to reduce HIV transmission be presented in a way as it applies to HIV prevention once released from a correctional facility (Braithwaite and Arriola 2003). In this way, correctional officials would not have to deviate from the rules in their facilities, and inmates would have the opportunity to learn how to protect themselves and sexual and/or drug injection partners from HIV transmission.

Unintended Consequences of Incarceration: Concurrency in the Community

The criminal justice system removes individuals away from the community, friends, family, and intimate partners. Once sexual or injection drug partners are separated, new partners may be introduced into the relationship; partners left in the community may engage in sexual intercourse and injection drug use with new partners. When the number of sexual or drug injection partners are increased, the risk of becoming infected with HIV and other sexually transmitted diseases increases. While incarcerated, inmates may engage in sexual activities with new partners, either by choice or coercion. Injection and tattooing equipment may be shared with other inmates. Due to the large number of incarcerated black males, black women are disproportionately represented in the community. This increases the opportunity for men in the community to have relationships with multiple women and decreases the number of potential long-term male partners for women (Pouget et al. 2010). Men and women in the community and those in correctional facilities are placed at a higher risk becoming infected with HIV.

Among recently incarcerated males, high-risk sex partnerships are likely to be reported. A study of men, aged 15–44 years old, found that incarceration may impact sexual behavior between both users and nonusers of illicit drugs. Multiple and concurrent partnerships were strongly associated with recently incarcerated men; however, the introduction of illicit drugs heightened risk behaviors with multiple or concurrent partners. Drugs enable individuals to engage in risky sexual intercourse and further drug use without sound judgment (Khan et al. 2009). The correctional system has a significant impact on those incarcerated and those left behind in the community.

Linkage to Medical Care After Release

After being released from a correctional facility, individuals need to be integrated back into the community. Former inmates may encounter a host of social and medical challenges upon release. In a population with an exceedingly high number of substance abusers and individuals suffering from mental illness, relapse to addiction and untreated medical illness is common upon release, respectively. Though there is a particularly high prevalence of mental illness in the correctional system, including 60 % in state prisons, 45 % in federal prisons, and 64 % in jails, only 25 % of inmates with psychiatric conditions received medications. Mental conditions left untreated can lead to a low adherence to antiretroviral therapy and reentry into a correctional facility (Rich et al. 2011). In states with a high rates of opiate abuse, drug overdose has been noted as a leading cause of death in the immediate post-release period (Binswanger et al. 2007). Homelessness and poverty often burden releases. It is possible for some inmates to lose Medicaid benefits during their sentence, resulting in an average interruption of 3 months in insurance coverage while the individual reapplies (Wakeman et al. 2009). Individuals convicted of drug-related charges are banned from receiving food stamps or federal assistance under the Welfare Reform Act and can be denied public housing under the Anti-Drug Abuse Act of 1988 (Pogorzelski et al. 2005). Antiretroviral therapy in correctional facilities has been successful in a few states; however, the improvements in an inmate's health observed during incarceration may quickly diminish if adequate care and treatment is not continued (Rich et al. 2011). Medical services offered during incarceration may be the individual's first link to primary healthcare and HIV/AIDS care. It is common for incarcerated persons to return to poor communities with little or inadequate access to HIV medical care. As there are a considerable number of obstacles presenting themselves, adhering to HIV treatments may not be a high priority for many releasees while trying to gain stability in the community. Release from a correctional facility has been associated with poor HIV treatment adherence: only 5.4 % of released HIV-positive inmates filled their antiretroviral prescriptions within 10 days, 17.7 % after 30 days, and 30.0 % after 60 days (Baillargeon et al. 2009). Poor adherence can lead to disease progression and emergence of drug–resistant viral HIV strains. Therefore, prerelease discharge planning, case management, and linkage to care are crucial to ensure that former inmates initiate and continue HIV/AIDS care and treatment. Non-adherence to HIV treatment upon release from a correctional facility may also be linked to inadequate discharge planning (Culbert 2011).

Prerelease discharge planning allows inmates to develop a plan to access medical care and treatment while incarcerated to reduce the burdens faced by inmates upon reentry into the community. Many agree that discharge planning is a legally and ethically mandated standard of care for those in the criminal justice system. During discharge planning, inmates can be provided with (a) a list of medical providers in the community, (b) an appointment with a community care provider, (c) education about the importance of adherence to medications, (d) transfer of medical records, (e) assistance with insurance applications, and (f) linkage to HIV case management services. Upon release, former inmates should be provided with an adequate supply of medication to avoid interruption in treatment until the initial appointment in the community (Centers for Disease Control and Prevention 2009).

HIV case management services have the ability to assist with discharge planning to link recent releases to healthcare and other resources in the community; case managers provide a link between former inmates and healthcare facilities to make access to healthcare easier. Inmates who participate in adequate discharge planning may be more likely to sustain their health and avoid reentry into correctional facilities (Culbert 2011). Although case managers can be beneficial to recent releasees, there are shortcomings that come with using just case management programs to ensure that adherence to HIV treatment and non-detectable viral loads are continued upon release, and that linkage to medical services is established. Discharge planning and HIV case management should not only involve linkage to HIV medical services and treatment, but also to organizations that can alleviate social burdens upon release as well such as health insurance, housing, and substance abuse treatment to reduce the number of burdens that releasees face. In the face of social burdens, HIV-positive releasees may disregard adherence to medication.

To create an effective discharge planning program, a considerable amount of money and face-to-face time is required in addition to the development and continuance of partnerships within the communities in which former inmates reside upon release. However, budget cuts and transfer of correctional healthcare services into the private sector have reduced partnerships with community-based organizations (Culbert 2011). Although there are challenges and resources may be limited in some facilities, an effort can be made to increase adequate discharge planning and case management programs in the correctional system.

Starting in 2006, the Rollins School of Public Health at Emory University and Abt Associates, Inc. coordinated the evaluation of the Enhancing Linkages to HIV Primary Care and Services in Jail Settings Initiative (EnhanceLink). This initiative successfully provided HIV testing in 20 US jails, resulting in 822 new diagnoses out of 212,464 inmates who agreed to HIV testing. Services were provided to not only address linkage to medical care but also to social needs upon release such as mental health and substance abuse treatment and linkage to housing. Transitional services were accepted by 82 % of HIV-positive persons in the program offered services (Spaulding et al. in press-a). At 6 months follow-up, 26 % of releasees had suppression of their HIV viral load (Spaulding et al. in press-b). In a jail setting, HIV testing and linkage to care after release are feasible. With adequate program planning, successful testing and transitional programs can be implemented in jails and prisons throughout the United States.

Conclusion

The prevalence of HIV is high in the correctional population, but HIV disease needs to be understood in context. Men and women entering the correctional system are also disproportionately affected by low socioeconomic status, violence, mental illness, substance abuse, chronic disease, and infectious disease. With a large proportion of HIV-positive individuals passing through the correctional system, these facilities have the opportunity to play an instrumental role in the diagnosis and clinical care of HIV/AIDS patients. Known HIV-positive persons would be able to receive HIV medication and avoid interruption of therapy. Undiagnosed HIV-positive persons could learn of their HIV status and begin treatment; access to HIV testing may be suboptimal after release. Treatment of HIV with antiretroviral therapy is needed to lower viral loads and restore the immune system, allowing HIV-infected persons to live a healthier life. Correctional facilities are settings where HIV-positive persons might control their HIV for the first time.

Furthermore, due to the poorer health status and higher incarceration rates of people of color, the criminal justice system has the opportunity to significantly decrease health disparities. If HIV medical services are not provided to the incarcerated and released population, poor health outcomes could continue to devastate communities plagued by poverty and high crime rates. The incarcerated have a right to adequate medical care. Many individuals transitioning through a correctional facility do not have access to adequate healthcare, prior to receiving healthcare in a correctional facility. An effort has to be made to develop more partnerships between the correctional system and the community-based organizations.

Medical care in the criminal justice system provides medical and public health professionals the opportunity to work with a vulnerable and underserved population. In order to continue to have a supply of providers sensitive to the needs of inmate patients, medical schools and residency programs should consider providing mentored training opportunities in correctional facilities. The correctional setting provides opportunities in a challenging environment to reach a large underserved population, while developing skills in areas such as working with patients with co-occurring medical conditions and dealing with a high turnover rate of patients. There are many practice and research opportunities available in correctional facilities to address and aid in the needs of the incarcerated. The criminal justice system has the capacity to be a wonderful training ground for future health professionals.

Correctional healthcare, if administered well, can have a positive impact on the lives of HIV-positive inmates. There is an opportunity to enhance HIV medical care in correctional facilities. Correctional healthcare can mitigate the disparities among a poor, underserved population. Improved HIV educational programs, HIV testing, and linkage to care is needed to ensure that better health outcomes are observed in the incarcerated population and after release to the community. Failing to address health disparities is an additional punishment.

Recommended Reading and References

Altice F, Sylla L et al (2010) Jail: time for testing, institute a jail-based HIV testing program. Yale University School of Medicine, New Haven

Baillargeon J, Borucki M et al (2000) Antiretroviral prescribing patterns in the Texas prison system. Clin Infect Dis 31:1476–1481

Baillargeon J, Giordano T et al (2009) Accessing antiretroviral therapy following release from prison. JAMA 301(8):848–857

Binswanger I, Stern M et al (2007) Release from prison – a high risk of death for former inmates. N Engl J Med 356(2):157–165

Braithwaite RL, Arriola KRJ (2003) Male prisoners and HIV prevention: a call for action ignored. Am J Public Health 93(5):759–763

Centers for Disease Control and Prevention (2010) Monitoring selected National HIV prevention and care objectives by using HIV surveillance data – United States and 6 U.S. dependent Areas – 2010, Part A. HIV Surveill Rep Suppl Rep 17(3):22

Centers for Disease Control and Prevention (2012) HIV in the United States: at a glance http://www.cdc.gov/hiv/statistics/basics/ataglance.html.

Centers for Disease Control and Prevention (2009) HIV testing implementation guidance for correctional settings. Atlanta GA. pp 1–38. Available at: http://www.cdc.gov/hiv/topics/testing/resources/guidelines/correctional-settings.

Culbert GJ (2011) Understanding the health needs of incarcerated men living with HIV/AIDS: a primary health care approach. J Am Psychiatric Nurses Assoc 17(2):158–170

de Voux A, Spaulding AC et al (2012) Early identification of HIV: empirical support for jail-based screening. PLoS One 7(5):e37603

Glaze L (2010) Correctional population in the United States, 2010. U.S. Department of Justice, Bureau of Justice Statistics, Washington DC. NCJ 236319

International Centre for Prison Studies. World prison brief: entire world – prison population rates per 100,000 of the national population http://www.prisonstudies.org/info/worldbrief/

Jafa K, McElroy P et al (2009) HIV transmission in a state prison system. PLoS One 4(5):e5416

Jurgens R, Ball A et al (2009) Interventions to reduce HIV transmission related to injecting drug use in prison. Lancet 9:57–66

Khan MR, Doherty IA et al (2009) Incarceration and high-risk sex partnerships among men in the United States. J Urban Health 86(4):584–601

Macgowan R, Margolis A et al (2009) Voluntary rapid human immunodeficiency virus (HIV) testing in jails. Sex Transm Dis 36(Suppl):S9–S13

Marks G, Crepaz N et al (2005) Meta-analysis of high-risk sexual behavior in persons aware and unaware they are infected with HIV in the United States: implications for HIV prevention programs. J Acquir Immune Defic Syndr 39(4):446–453

Maruschak L (2008) HIV in prisons. US Department of Justice, Bureau of Justice Statistics, Washington DC. NCJ-222179

Pogorzelski W, Wolff N et al (2005) Behavioral health problems, ex-offender reentry policies, and the “Second Chance Act”. Am J Public Health 95:1718–1724. doi:10.2105/AJPH.2005.065805

Pouget E, Kershaw T et al (2010) Associations of sex ratios and male incarceration with multiple opposite-sex partners: potential social determinants of HIV/STI transmission. Public Helath Rep 125(Suppl 4):70–80

Rich JD, Wohl DA et al (2011) HIV-related research in correctional populations: now is the time. Curr HIV/AIDS Rep 8(4):288–296

Spaulding A, Stephenson B et al (2002) Human immunodeficiency virus in correctional facilities: a review. Clin Infect Dis 35(3):305–312

Spaulding AC, Seals RM et al (2009) HIV/AIDS among inmates of and releasees from US correctional facilities, 2006: declining share of epidemic but persistent public health opportunity. PLoS One 4(11):e7558

Spaulding AC, Booker CA et al. Jails, HIV testing and linkage to care services: an overview of the enhanceLink project. AIDS Behav (in press-a)

Spaulding AC, Messina LC et al. Planning for success predicts virus suppressed: results of a non-controlled, observational study of factors associated with viral suppression among HIV-positive persons following jail release. AIDS Behav (in press-a, b)

Springer S, Altice F (2005) Managing HIV/AIDS in correctional settings. Curr HIV/AIDS Rep 2(4):165–170

The World Health Organization Geneva, Switzerland and The Joint United Nations Programme on HIV/AIDS New York (2006) HIV/AIDS prevention, care, treatment and support in prison settings

Tucker JD, Chang SW et al (2007) The catch 22 of condoms in US correctional facilities. BMC Public Health 7(1):296

U.S. Department of Health and Human Services (2011) This is an on-line document. URL: http://aidsetc.org/pdf/p07-cg/CM_Jan2011.pdf. Guide for HIV/AIDS clinical care

Wakeman SE, Rich JD (2010) HIV treatment in US prisons. HIV Ther 4(4):505–510

Wakeman S, McKinney M et al (2009) Filling the gap: the importance of Medicaid continuity for former inmates. J Gen Intern Med 24(7):860–862

Home Burglary

► Residential Burglary

Homegrown Terrorism in the United States

Joshua D. Freilich[1,3], Steven M. Chermak[2,4] and Roberta Belli[5]
[1]Jay College of Criminal Justice, New York, NY, USA
[2]School of Criminal Justice, Michigan State University, East Lansing, MI, USA
[3]National Consortium for the Studies of Terrorism and Responses to Terrorism (START)
[4]National Consortium for the Studies of Terrorism and Responses to Terrorism (START), University of Maryland, College Park, MD, USA
[5]UNDSS Somalia SIOC, Springette, Spring Valley, Nairobi, Kenya

Synonyms

Domestic terrorism; Political violence; Terrorism inside the USA

Overview

This entry focuses on homegrown terrorism in the United States. There are many terrorism definitions. Indeed, Schmid and Jongman (1988) identified over 100 definitions that differed over whether they included 1 of 22 elements. Nonetheless, Freilich et al. (2009a, b) demonstrate that almost all definitions require a violent, ideologically motivated illegal act committed by a political extremist, and this definition is used here. The Congressional Research Service (Bjelopera and Randol 2010, p. 1) defines homegrown terrorism as a "terrorist activity or plots perpetuated within the United States or abroad by American citizens, permanent legal residents, or visitors radicalized largely within the United States."

A review of the empirical literature demonstrates that homegrown terrorists currently pose a threat to public safety. These threats are dynamic however and change over time. In addition, policy interventions and strategic tactics often have unintended consequences. Sometimes these interventions have created backlash effects that led to greater numbers of crimes. This entry first discusses how homegrown extremists pose a threat and then illustrates how it has changed over time. Second, the entry outlines the nature and scope of the current danger posed by homegrown terrorists and extremists. It then discusses how policy makers, law enforcement personnel, and crime analysts should carefully monitor the homegrown threat and consider the unintended consequences of their proposed responses to these extremists. Finally, note that this entry includes new material and draws from materials published in two 2009 articles (Freilich et al. 2009a, b).

Key Issues

Homegrown terrorists and extremists pose a risk to the American community. Recently, attention has focused on American supporters of Al-Qaeda and related ideologies, because of the devastating attacks committed on September 11 and its desire to strike again within the United States. However, LaFree et al. (2006) found that the United States also faces other homegrown threats. Hewitt (2003; see also Smith 1994) found that between 1954 and 2000, over 3,000 terrorist incidents were perpetrated in the United States, many by non-Islamic jihadist homegrown terrorists such as far-rightists, Puerto Rican nationalists, Jewish extremists, Black nationalists, and animal and ecoterrorists. These acts claimed over 700 lives and caused hundreds of millions of dollars in damage. More recent research indicates that the number of incidents occurring in the United States might be higher (Freilich et al. 2012).

A review of the accumulated empirical evidence finds that homegrown terrorist threats are not static. Smith's (1994) study of terrorist strikes in the United States found homegrown far-left groups to be the most active in the 1970s. Carlson (1995) surveyed police chiefs in large-sized American cities (>100,000) in the early 1990s. He asked the chiefs to rank the top four groups that, in their opinion, were most likely to commit a terrorist act within the next 2 years. Despite the celebrated nature of the 1993 Trade Center bombing, police chiefs did not think that Middle Eastern terrorists posed the greatest threat. Instead, homegrown antiabortion perpetrators and white supremacists were rated the two greatest threats. Riley and Hoffman's (1995) RAND survey of state and local agencies over 2 years in the early 1990s similarly found that issue-specific homegrown organizations (e.g., antiabortion) and right-wing groups posed the greatest terrorist threat. Freilich et al. (2009) more recent survey of state police agencies found that almost all respondents strongly agreed that Islamic jihadists and their American supporters were the top national security threat and most strongly agreed or agreed that they were a significant threat to state security. Importantly, far-right and animal/environmental rights homegrown extremist movements were also seen as representing a severe threat to both national security and to the security of individual states.

These results demonstrate that homegrown terrorism threats change over time. Carlson's research from almost 15 years ago identified homegrown antiabortion activists as the greatest threat. Freilich et al. (2009) found that state police agencies, however, do not currently rate antiabortion activists as posing a great threat to either national security or their state's security. The attacks of September 11 have clearly impacted concerns about the potential threat of Islamic extremists in the United States. While Smith's (1994) domestic terrorism study found homegrown far-left groups to be the most active in the 1970s, currently state police agencies do not rate far-left groups as a major threat on either the national or state level.

Relatedly, the strength of individual groups and the larger homegrown movements they belong to also evolves over time. These extremist and terrorist groups and their external environment are dynamic. For example, research finds that the homegrown far-right militia movement was impacted by the Oklahoma City bombing, forcing many individuals to leave the movement, others to join, and some to go underground or seek other extremist organizations in fear of government infiltration (Chermak 2002). Extremist organizations and terrorist groups must adapt to changes occurring in their external environment. Research (Freilich et al. 2009; Rapoport 1992) shows that most terrorist and extremist groups are unable to adapt and therefore fail in a relatively short time.

Empirical evidence also finds that the rise and fall of homegrown extremist organizations are linked to critical events and organizational-level variables that enhance or disrupt the group. Freilich et al. (2009) identified factors that produced growth and decline within four homegrown far-right extremist organizations (the National Alliance, the Aryan Brotherhood, Public Enemy Number 1, and the Oklahoma Constitutional Militia). The groups that grew did do because they (1) had able leadership that set forth a clear ideological message and goals, (2) undertook concrete actions to further their ideology and goals and had the finances necessary, (3) took advantage of political opportunities, and (4) were internally cohesive. However, the groups declined because of organizational instability and/or responses by law enforcement and nonstate actors, such as watch groups. The fall/decline of a particular group likely affects other groups or supporters in the same movement. For example, the decline of one far-right group often provides political openings for another extremist organization to grow and prosper. Another example is law enforcement operations that eliminate violent extremists. Such actions could unintentionally outrage and further radicalize other extremist groups to turn to violence (Freilich et al. 2009).

Current Threat

The United States Extremist Crime Database (ECDB) is a unique, open-source, relational database. The ECDB is funded by the Department of Homeland Security through the National Consortium for the Study of Terrorism and Responses to Terrorism (START). It tracks ideologically motivated and routine/nonideological violent and financial crimes committed by homegrown movements seen as representing the greatest threat to American public safety: far-rightists, Al-Qaeda supporters, and animal and environmental rights extremists (Freilich et al. 2009). Definitions/descriptions of these movements are set forth in Freilich et al (in press) work and are available upon request. The ECDB collects information on multiple units of analysis. The ECDB's inclusion criteria are based upon a criminological approach, and the ECDB is different than other terrorism databases because it includes a broader array of illegal acts. For an incident to be included in the ECDB, it must satisfy a two-prong test. The first prong is behavioral and requires that an illegal violent act or financial scheme must have been committed inside the United States. The second prong is attitudinal and requires that, at the time of the incident or scheme, at least one of the suspects who committed the act must have subscribed to a far-right, Al-Qaeda-influenced, or extremist animal or environmental rights belief system (Freilich et al. in press).

The ECDB has identified thousands of criminal incidents committed by homegrown far-right extremists between 1990 and 2010. These crimes range in important ways, such as the level of violence imposed on victims, the number of suspects involved, and the motivational circumstances underlying each incident. Far-right extremist crimes include hate crimes, attacks on abortion providers, and foiled terrorist plots, as well as financial crimes and cases involving illegal firearms and other weapons.

According to the ECDB, homegrown far-right extremists committed over 370 homicide incidents in the period spanned by the data. Over 40 % of these homicide events were ideologically motivated (and thus most akin to terrorist attacks), and another 20 % were nonideological but were still movement related (such as an internal killing over drugs or a woman). These homicides varied both spatially and temporally. Homegrown far-rightists also killed close to 50 law enforcement officials in the line of duty between 1990 and 2010. These incidents involved federal, state, or local police officers, correctional officers, private security guards, and a judge (Freilich et al. in press).

Smith and Damphousse's (2009) American Terrorism Study (ATS) documents the tremendous property damage and financial losses that environmental and animal rights attacks have caused in the United States. Indeed, one group "The Family" committed 21 arson and "ecotage" attacks between 1995 and 2001. The group's most serious act was an arson attack at the Vail ski resort in the late 1990s that caused over $25 million dollars in damage.

The ECDB was still involved in data collection on crimes committed by eco – and animal rights extremists when this entry was being prepared, and thus the results are not finalized. Nonetheless, a preliminary examination indicates that these homegrown extremists have committed over 230 violent (i.e., bombing and arsons) incidents between 1995 and the present (Freilich et al. 2012). In addition, the ECDB is investigating a few violent acts committed by possible supporters of extremist animal and environmental rights movements that targeted persons. This is important because many academics and others conclude that this type of homegrown extremist has intentionally avoided harming persons (Amster 2006; Cothren et al. 2008; Taylor 1995). These incidents that may have involved animal and environmental rights extremists who targeted persons include the Unabomber's letter bomb attacks, the murder of a Red Bluff police officer by a self-proclaimed anarchist who also criticized corporate malfeasance including their harmful actions against the environment, a car bomb attempted murder of a veterinary officer, and the firebombing of a car belonging to a University of California, Santa Cruz scientist that resulted in injuries to both of his feet.

Importantly though, a review of the empirical literature thus far indicates that research on homegrown terrorists has focused mostly on Al-Qaeda supporters and other similar extremists (see below) or far-right extremists. Gruenewald et al. (2009) identified over 320 studies on far-right-wing extremism and political crimes in the fields of criminology, political science, sociology, and terrorism, and most of the current studies focus on Al-Qaeda supporting networks. Conversely, a review found many fewer studies on animal/environmental rights extremists. Future research needs to focus on eco – and animal rights homegrown extremists and study them more systematically.

Again, the ECDB also tracks crimes committed by supporters of Al-Qaeda and groups involved in local jihads across the globe. Many of the perpetrators of these crimes are homegrown extremists. Here too, the ECDB is still involved in data collection. But, a preliminary examination indicates that these extremists have committed over 25 fatal (i.e., the suspect killed others, was killed by police, or committed suicide) incidents since 1990. In addition, in this period, the ECDB identified a few honor-killing incidents, as well as scores of foiled plots in the United States (Freilich et al. in press).

In addition to ideologically motivated violent crimes, homegrown extremists have also committed nonviolent crimes that have harmed US security. Far-right homegrown extremists committed over 430 "financial schemes," while Al-Qaeda and other similar extremists committed over 100 financial or material support schemes since 1995. These schemes were worth over 600 million dollars. Since data collection and coding is ongoing, this number may change (Freilich et al. in press).

A preliminary descriptive analysis of ECDB data of financial schemes prosecuted in federal court in 2004 reveals that far-right and Al-Qaeda-influenced and similar extremists in the USA engage in different types of financial schemes, follow distinct modus operandi, and are prosecuted for a variety of both financial and nonfinancial crimes (Belli 2011). Despite these differences, a common trend was identified, that is, the involvement of non-extremist accomplices who provided useful resources for the crime commission process.

With respect to the far-right, the findings suggest that these may be individuals with specialized skills and know-how who provided professional services and expertise on a variety of issues (e.g., tax preparation, business ventures, financial and legal counseling). Al-Qaeda-influenced and similar extremists seem to have relied on non-extremist accomplices for a diverse set of reasons, for example, as middlemen to conduct wire transfers and manage bank accounts, couriers for cash smuggling, and nominees to set up storefront businesses (Belli 2011).

A comparison of financial schemes, crimes, and techniques revealed significant differences in the financial criminality of far-rightists and Islamic extremists. Far-rightists engaged mostly in tax avoidance schemes as a form of antigovernment protest. These schemes were perpetrated using techniques that require both proactive and passive behaviors. Proactive behaviors included filing false tax documents, underreporting income, marketing antitax packages, using bogus trusts, and setting up offshore "shell" corporations. Passive behaviors included ideologically motivated tax refusal, that is, a popular crime of omission that involves the failure to file an income return or to pay employment taxes on the basis of misinterpretations of tax laws and the constitution, usually rejected as "frivolous arguments" by US courts.

The majority of Al-Qaeda-influenced and similar extremist-related schemes involved "money-dirtying" and money-laundering strategies for self-financing purposes. Money dirtying (the reverse of money laundering) refers to the process of using money raised legitimately (e.g., through donations or legitimate investments) for illicit purposes (e.g., terrorism financing). Although there are some similarities between money-laundering and money-dirtying strategies, there are also at least two important differences: (1) the source of the money, that is, criminal in the former (e.g., drug dealing, tax evasion, cigarette smuggling) as opposed to either legitimate or criminal in the latter (e.g., legitimate business revenues, charitable donations), and (2) the goal, that is, transforming "dirty" money into expendable income as opposed to channeling funds of any origin to individuals or groups to enable acts of terrorism. To carry out these schemes, Al-Qaeda-influenced and other Islamic extremists made ample use of traditional money transfer mechanisms both domestically and internationally (e.g., wire transfers, cash deposits, checks). They also used simpler methods that avoid leaving paper trails, like conducting transactions in cash and physically transporting money across borders.

There were also differences in the way far-right and Islamic extremist financial crime cases were handled by federal prosecutors. Comparing criminal charges in federal indictments revealed that far-right cases were usually prosecuted for typical financial crime offenses (e.g., tax evasion, mail fraud, money laundering), whereas suspects in alleged terrorism financing cases were indicted for a variety of criminal offenses, ranging from money laundering to providing material support to a foreign terrorist organization (FTO), immigration fraud, RICO conspiracy, and false statements.

Finally, a recent report using ECDB data compared the organizational differences between violent and nonviolent hate groups (Chermak et al. 2013). This study used the Southern Poverty Law Center's (SPLC) annual Intelligence Report and Klan Watch publications to produce a list of hate groups in the United States (over 6,000 hate groups) and then focused on organizations that existed for at least 3 consecutive years (over 500 groups). Two hundred seventy-five (275) groups were then randomly selected to study in depth. Groups were categorized as violent or nonviolent.

The research revealed that 21 % of the 275 far-right hate groups included in the study had members who had committed at least one violent criminal act. Based on findings from a number of statistical models, there appear to be several indicators that were related to a group's propensity for violence even when controlling for other significant predictors. Age and size were related to a group's propensity for violence. That is, as groups increased in the number of years in existence or in the number of their members, the likelihood of them being involved in violence increased. This result makes sense as groups have an opportunity to learn over time. The significance of group size may be that simply having more members increases the odds that at least one individual will be linked to a violent act. Larger organizations also have a more diverse body of members who bring different skills and expertise, and this diversity may allow them to evade capture for a longer period of time and thus provide the opportunity to commit more violent crimes.

Groups that published ideological literature, such as newsletters or pamphlets, were significantly *less likely* to be involved in violence. Such literature is used to attract potential members to the organization, and perhaps these groups realize that the publication of their rhetoric will bring increased attention to their group and thus decrease the likelihood of being involved in violence. Groups that had some specific conflict with another group, had charismatic leaders, advocated for leaderless resistance tactics, and were headquartered in the West and Northeast (compared to groups in the South) were significantly more likely to be involved in violence. *Policy makers must carefully monitor political extremists and consider the unintended consequences of their proposed interventions.*

These findings suggest that law enforcement intelligence and analysis have an important role to play in countering terrorism. First, the ECDB's preliminary financial findings have important policy implications. They provide evidence that political extremists today engage in criminal behaviors that have traditionally been associated with profit-driven crime for various reasons and using different methods. In addition, the differences in the modus operandi and goals pursued highlight the importance of conducting systematic analyses of crime patterns using an "all-crimes/all-threats" approach to improve threat assessment and prioritization strategies. Lastly, they highlight the need to investigate the role of non-extremists as well as the nature and scope of their interactions with political extremists, as they may provide crucial resources in the form of knowledge, skills, and contacts.

Again, another important obstacle for law enforcement and analysts to overcome is to understand that identifying the homegrown threats that should be prioritized is like hitting a moving target. Analysts must engage in dynamic analyses (Freilich et al. 2009b; Smith 1994). Changes in contextual and organizational factors may cause a homegrown movement to increase or decrease in strength and potential to commit violent acts. Paying attention to such changes will allow law enforcement to better assess a movement's threat level as it changes, differentiate between truly dangerous groups and less dangerous groups, differentiate between immediate versus potential future threats, and develop policies to best address these threats (Chermak et al. 2009; Duffy and Brantley 1997; McGarrell et al. 2007).

Presently, different types of counterterrorism strategies have been implemented to reduce the number of terrorism attacks in the United States and abroad. Most strategies are "hard control" measures, such as retaliatory attacks, increasing punishments and using law enforcement measures to "crack down" on certain active groups. Lum et al. (2006, p. 22) conclude in their systematic review that there are few positive results: "Many effects are close to, or cannot be statistically discerned from, a zero effect."

A review of recent empirical research supports two conclusions. First, previous studies have found that responses by both the state and non-state actors may contribute to the decline of *specific* homegrown terrorist or extremist groups. Smith (1994, p. 91; see also Hamm 1993) concludes that homegrown far-right skinhead groups declined in the United States in the 1980s because of a "double barreled blast of aggressive federal criminal prosecutions and private civil rights attorneys." Hamm (2007) reached similar conclusions about the demise of the homegrown Arizona Patriots (see also Hewitt 2003).

Importantly, such responses, although contributing to declines in these groups, may also have unintended consequences. For example, Hamm's (2007) important analysis examining terrorist involvement in traditional crime discusses how terrorist groups seek traditional means of funding until such funding is unavailable. Groups then adapt and seek other sources of revenue. Hamm examines the effects of a crackdown on the IRA's efforts to raise money. In the 1970s, law enforcement shut down the IRA's fund-raising efforts in the United States, but this success increased their reliance on other sources of revenue. This shift may have also increased the harm caused by the organization as it moved to more traditional criminal and organized crime activities to finance operations, including bank robberies, smuggling goods and weapons, and extortion.

Lum et al. (2006, p. 26) systematic review evaluates several types of counterterrorism measures and also finds that such interventions often produce unintended and counterintuitive findings. For example, the implementation of metal detectors in airports were effective in decreasing airport hijackings, but caused displacement effects and increased the use of other types of terrorism, such as kidnappings, assassinations, and bombings.

Similarly, Freilich et al. (2009a) found that after the homegrown far-right National Alliance and the Aryan Brotherhood imploded for various reasons after the deaths of their well-known leaders, other groups, such as skinhead organizations, seemed to benefit. They concluded that groups within a larger movement are interrelated. Strategies that might mitigate activities against only one part of a movement may result in unanticipated consequences to other parts of that movement.

Relatedly, it seems obvious that successful law enforcement operations can eliminate a group or individuals that have engaged in crimes. This appears to bolster deterrence and incapacitation views that call for harsh responses to extremist groups. Many watch groups call for firm government actions against extremist organizations (Chermak 2002; Freilich 2003).

A review of empirical research in this area also suggests that there are sometimes unintended consequences to law enforcement responses. Perceived harsh government and police actions may lead to consciousness raising (Dobratz and Shanks-Meile 1997) that outrages the wider homegrown movement and causes a backlash. Other extremist groups and the movement at large may grow, and individuals and groups may become further radicalized and turn to violence. A growing number of studies find that harsh government responses to terrorism often have no deterrent impact and sometimes lead to a backlash (LaFree et al. 2009; Pridemore and Freilich 2007). For example, Lum et al. (2006) systematic review of the impact of counterterrorism strategies concluded that retaliatory strikes after a terrorist incident (i.e., the United States attack on Libya in 1986 in response to a bombing in Berlin) actually led to additional attacks.

Similarly, Freilich et al. (2009a) found that the racist Aryan Nations group capitalized on the killings of Gordon Kahl and Robert Mathews by law enforcement in the 1980s and the organization's popularity increased. In addition, Mathews partially created the notorious group "The Order" – the premiere domestic terrorist group in the 1980s (responsible for five murders and spectacular armored car robberies) because of anger over Kahl's death. The National Alliance, the militia movement, and other far-right groups grew in reaction to the government's excesses at Waco and Ruby Ridge in the early 1990s.

Alternative Policies

Based upon a review of the empirical findings in this area, we conclude that the authorities must strike a careful equilibrium. The government must balance a carefully calibrated campaign that eliminates dangerous homegrown groups and stresses that violence will not be tolerated, with responses that avoid outraging and possibly increasing crime from the larger movement. There are several options to consider. First, law enforcement should only use deadly or harsh responses as a last resort. For instance, harsh government actions at Waco and Ruby Ridge in the early 1990s that led to the death of homegrown cult members and far-rightists outraged the antigovernment far-right. Most observers conclude that these harsh policies led to a backlash and fueled the growth of the paramilitary right (Chermak 2002; Freilich 2003). Conversely, the ADL (2003) explains that law enforcement's nuanced and patient response that allowed far-right extremists to save face led to a peaceful resolution of sieges in Montana in 1996 and in Texas in 1997.

The government should consider accompanying harsh law enforcement actions against violent groups with "outreach" programs to nonviolent wings of extremist movements. These strategies should encourage members of the movement to forsake illegal behaviors in favor of participation in the political system or other legal activities. Jones and Libicki's study found that 43 % of 268 terrorist groups worldwide – the largest category – ended because of a transition to the political process. Significantly, F.B.I. agents charged with monitoring domestic extremists have made similar arguments. Agents Duffy and Brantley (1997) created a typology of four categories of homegrown far-right militia groups that ranged from noncriminal to organizations that conducted serious – for example, homicide and bombings – crimes. Duffy and Brantley (1997, p. 2) urged local police to open a dialogue with leaders from the two nonviolent categories "so that the two sides can voice their concerns and discuss relevant issues in a non-confrontational way."

These strategies should be used more frequently, especially when law enforcement is involved in crackdowns on specific movements. These outreach policies could be modeled after law enforcement efforts to build bridges in immigrant communities that are suspicious of the government (Clarke and Newman 2006; Freilich and Chermak 2009). Newman and Clarke (2008) recently set forth steps that police agencies could take to reduce tension with migrant communities that have come under suspicion because a few terrorists, such as the 9/11 hijackers, used them to fit in. These steps included assigning community police agents to solely work with migrant communities, taking advantage of ethnic media outlets to communicate with the larger community, and clearly publicizing the police agency's mission and policies. Similar strategies could be undertaken with the more moderate portion of homegrown extremist movements. In addition to initiating dialogue with nonviolent groups, agencies could assign officers to appear on extremist media outlets. Such campaigns should explain the government response and note that it is only directed at those who commit crimes. Authorities should reassure noncriminal members of the movement that their rights will be protected and encourage them to focus on lawful activities while stressing that violence will not be tolerated.

While there are hard-core individuals who have no interest in pursuing change through legitimate processes, it is also true that most extremist individuals and groups do not engage in violence or terrorism themselves. Further, some who end up engaging in terrorism are reluctant participants and are not strongly committed to an extremist ideology. McGarrell et al. (2007) discuss how most (nearly 70 %) of the individuals who were detained as terrorists or linked to a series of bombings were not ideologically committed.

Finally, backlash effects are usually found when a government takes action *after* a crime or terrorist act has been committed. Strategies that prevent these crimes from happening by such methods as using situational crime prevention to prevent groups from raising funds and furthering their objectives (Clarke and Newman 2006; Freilich and Newman 2009; Hamm 2007; Newman and Clarke 2008) may preempt the harsh police responses to these crimes and thus avoid backlash effects.

Conclusion

This entry provides an overview of the threat of homegrown terrorism in the United States. Despite the considerable investment in the study of the terrorism that has occurred following the

September 11 attacks, there is still little known about attacks occurring in the United States and the extent to which group dynamics might function similarly or differently when comparing at-home to abroad attacks. What has been documented provides evidence that there is a significant threat, that it is a constantly evolving threat, and that extremists of different ideologies are involved in a significant number of violent and financial crimes. Similarly, although there have been some lessons learned about how to best respond to such terrorism, there is much that is unknown and needs to be discovered. The need for additional research examining the etiology of homegrown terrorism is critical to understand the successes and failures of different types of responses.

Related Entries

▶ Ecoterrorism
▶ Hate Crime
▶ Policing Terrorism and Legitimacy
▶ Political Crime
▶ Situational Approaches to Terrorism
▶ Social Disorganization and Terrorism
▶ Strategies of Policing Terrorism
▶ Terrorist Organizations
▶ Use of Social Media in Policing

Recommended Reading and References

Amster R (2006) Perspectives on ecoterrorism: catalysts, conflations and casualties. Contemp Justice Rev 9(3):287–301

Belli R (2011) When political extremists and greedy criminal meet: a comparative study OF financial crimes and criminal networks in the United States. PhD dissertation, Graduate Center, The City University of New York

Bjelopera JP, Randol MA (2010) American Jihadist Terrorism: combating a complex threat. Congressional Research Service, Washington, DC, 7 Dec 2010. http://www.fas.org/sgp/crs/terror/R41416.pdf

Carlson JR (1995) The future terrorists in America. Am J Police 14(3/4):71–91

Chermak SM (2002) Searching for a demon: the media construction of the militia movement. Northeastern University Press, Boston

Chermak SM, Freilich JD, Suttmoeller M (2013). The organizational dynamics of far-right hate groups in the United States: Comparing violent to non-violent organizations. Studies in Conflict and Terrorism 36(3):193–218

Chermak SM, Freilich JD, Shemtob Z (2009) Law enforcement training and the domestic far-right. Crim Justice Behavior 36(12):1305–1322

Clarke RV, Newman GR (2006) Outsmarting the terrorists. Praeger Security International, New York

Cothren J, Smith BL, Roberts P, Damphousse KR (2008) Geospatial and temporal patterns of preparatory conduct among American terrorists. Int J Comp Appl Crim Justice 32(1):23–41

Dobratz BA, Shanks-Meile SL (1997) White power, white pride! The white separatist movement in the United States. Twayne Publishers, New York

Duffy JE, Brantley AC (1997) Militias: initiating contact. Federal Bureau of Investigation. http:/?www.fbi.gov/library/leb/1997/July975.htm

Freilich JD (2003) American militias: state-level variations in militia activities. LFB Scholarly LLC, New York City

Freilich JD, Chermak SM (2009) Preventing deadly encounters between law enforcement and American far-rightists. Crime Prev Stud 25:141–172

Freilich JD, Newman GR (eds) (2009) Reducing terrorism through situational crime prevention, vol 25, Crime prevention studies. Crim Justice Press, Monsey

Freilich JD, Chermak SM, Caspi DJ (2009a) Critical events in the life trajectories of domestic extremist white supremacist groups: a case study analysis of four violent organizations. Criminol Public Policy 8:497–530

Freilich JD, Chermak SM, Simone J Jr (2009b) Surveying American state police agencies about terrorism threats, terrorism sources, and terrorism definitions. Terror Polit Violence 21:450–475

Freilich JD, Chermak SM, Belli R, Gruenewald J, Parkin WS (in press) Introducing the United States Extremist Crime Database (ECDB). Terror Polit Violence (forthcoming)

Gruenewald J, Freilich JD, Chermak SM (2009) An overview of the domestic far-right and its criminal activities. In: Perry B, Blazak R (eds) Hate crimes: hate crime offenders. Praeger, Westport, pp 1–21

Hamm MS (1993) American skinheads: the criminology and control of hate crime. Praeger, Westport

Hamm MS (2007) Terrorism as crime: from Oklahoma City to Al Qaeda and beyond. New York University Press, New York City

Hewitt C (2003) Understanding terrorism in America: from the Klan to Al Qaeda. Routledge, New York

LaFree G, Dugan L, Fogg H, Scott J (2006) Building a global terrorism database. Final report to the national institute of justice. United States Department of Justice, Washington, DC

LaFree G, Dugan L, Korte R (2009) The impact of British counter terrorist strategies on political violence in

Northern Ireland: comparing deterrence and backlash models. Criminology 47:501–530
Lum C, Kennedy LW, Sherley AJ (2006) The effectiveness of counter-terrorism strategies: a campbell systematic review. Available from the WWW at: http://db.c2admin.org/doc-pdf/Lum_Terrorism_Review.pdf
McGarrell EF, Freilich JD, Chermak SM (2007) Intelligence-led policing as a framework for responding to terrorism. J Contemp Crim Justice 23(2):142–158
Newman GR, Clarke RV (2008) Policing terrorism. An executive's guide. U.S. Department of Justice: Office of Community Oriented Policing Services, Washington, DC
Pridemore WA, Freilich JD (2007) The impact of state laws protecting abortion clinics and reproductive rights on crimes against abortion providers: deterrence, backlash or neither? Law Hum Behav 31:611–627
Rapoport D (1992) Terrorism. In: Mary H, Maurice K (eds) Encyclopedia of government and politics, vol II. Routledge, New York
Riley JK, Hoffman B (1995) Domestic terrorism: a national assessment of state and local preparedness. Rand Corporation, Santa Monica
Schmid AP, Jongman AJ (1988) Political terrorism: a new guide to actors, authors, concepts, databases, theories and literature. North Holland, Oxford, UK
Smith BL (1994) Terrorism in America: pipe bombs and pipe dreams. State University of New York, Albany
Smith BL, Damphousse KR (2009) Patterns of precursor behaviors in the lifespan of an american eco-terrorist group. Criminology and Public Policy 8(3):475–496
Taylor B (ed) (1995) Ecological resistance movements: the global emergence of radical and popular environmentalism. State University of New York Press, Albany

Homicide Perpetration

► Youth Homicide in the United States

Homicide Rates/Trends

► Youth Homicide in the United States

Honor Killings

► Lawful Killings

Hormones, the Brain, and Criminality

Lee Ellis
Department of Anthropology and Sociology,
University of Malaya, Kuala Lumpur, Malaysia,
Malaysia

Overview

"Is biology really relevant to criminology? After all, criminal behavior is learned, right? People aren't just born criminals; that's ridiculous." Critics of the biosocial perspective still use this sort of reasoning to raise doubts about the role of biology in crime causation. However, such thinking overlooks evidence that many personality traits and even human interests and preferences are genetically *influenced* to a substantial degree (Bouchard and McGue 2003; Kreek et al. 2005; Schermer and Vernon 2008). There is also considerable evidence that all human learning boils down to neurochemical processes that vary from one individual to another (Lawson 2003). Consequently, even though criminality is learned, biology still affects *how* we learn. To set the stage for examining this line of reasoning further, this entry will describe the role of a class of biochemicals called *hormones* in learning criminal behavior. But before doing so, a little background on biosocial criminology is in order.

Criminology's Biosocial Perspective

The first scientist to make the case for biological influences on criminality was an Italian physician named *Cesare Lombroso*. Writing in the mid-1800s, he is most famous for having argued that persistent criminals were *atavistic*, meaning that they were "throwbacks" to an earlier stage in human evolution.

Today, criminologists' opinions about the relevance of biology to understanding criminal behavior have gone far beyond Lombroso's concept of *atavism*. These opinions can be organized into three rather extreme categories:

1. Criminal behavior is almost completely determined by biological forces.
2. Biology and the social environment interact to influence criminal behavior.
3. Criminality is determined almost entirely by social factors.

According to surveys conducted in recent years, no criminologists subscribe to the first statement, although critics of the biosocial perspective often imply otherwise. About 20 % of criminologists subscribe to the second option, i.e., biological and social factors *interact* to affect criminal behavior. The remaining 80 % of criminologists still believing that biology has little or nothing to do with people's involvement in crime (the third option). So, the biosocial perspective remains a minority position among criminologists, although it does seem to have gained a little in popularity in recent decades (Cooper et al. 2010).

As other entries in this *biosocial section* illustrate, a great deal of progress has been made in recent decades deciphering biological influences not only on criminality but also on several related behaviors (e.g., aggression, risk taking, and susceptibility to alcohol and drug addiction). What has emerged from this research is *not* the conclusion that people are "hard wired" to be murderers, thieves, robbers, or rapists. Instead, the following four interrelated principles have become part of how biosocial theorists reason:

1. All human actions are controlled by the functioning of the brain.
2. Genes and other biological factors profoundly affect how the brain functions in response to whatever environment it encounters.
3. By affecting brain responses to the environment, biological factors can alter the probability of individuals running afoul of the law.
4. Hormones are among the most important biological factors influencing brain functioning.

About Hormones

Hormones are substances produced primarily in one part of the body (such as various glands) that travel to other parts of the body (usually by way of the blood system) where they have their greatest effects. The brain is one of the main targets of hormones.

This entry focuses on two types of hormones. One type is called *sex hormones* because they exist in greater quantities in one sex than in the other. The other is known as *stress hormones* because they tend to be produced in greater quantities when individuals are experiencing stress and anxiety.

Androgens: The Hormones from Hell

Those familiar with farm animals – be they cattle, horses, hogs, sheep, or even chickens – know that males are nearly always more belligerent and combative than their female counterparts. What "barnyard culture" is responsible for such a sex difference? Of course, culture is *not* the culprit; biology is. Even though this is a rather obvious conclusion, think about how quickly people attribute male-female differences in belligerency and combativeness to cultural learning in *humans* while still recognizing that culture is not responsible for the same sex difference in other animals.

Scientists and farmers alike have learned that sex differences in belligerency and combativeness have hormonal underpinnings partly because they have identified how to eliminate the sex differences using a procedure called *castration*. This simple operation involves removing the testicles, glandular organs that specialize in producing hormones known as *androgens* (which collectively refers to all sex hormones that males produce in higher quantities than do females). Within a few months after being castrated, males nearly always become much more passive, like their female counterparts.

Studies have shown that the most influential androgen regarding physically aggression is *testosterone*. While most androgens are produced by the male testicles, smaller amounts are also synthesized by the adrenal glands (that both sexes have) as well as by the female ovaries. Therefore, even though androgens are called "*male* sex hormones," one should keep in mind that females also produce them, only in lower quantities.

Hormones, the Brain, and Criminality, Fig. 1 Crime according to age and gender

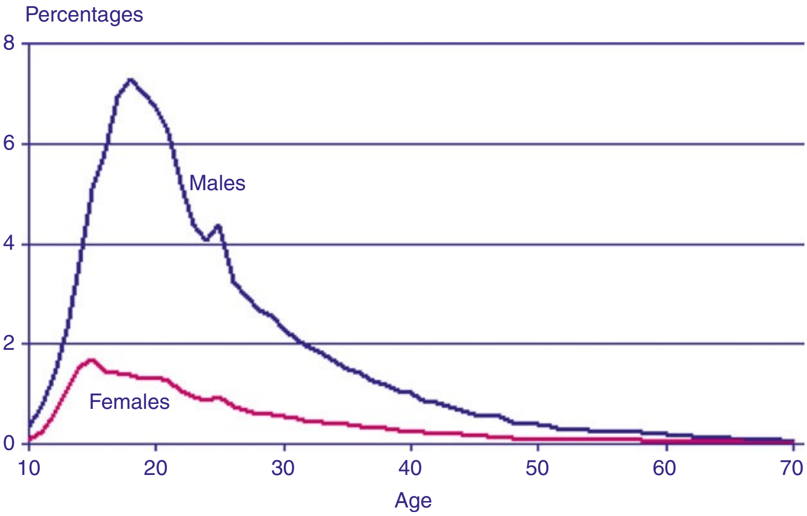

Linking Androgens to Criminality. If testosterone and other androgens are responsible for average sex differences in belligerency and combativeness, could these hormones also contribute to criminality? Nearly all biosocial criminologists believe the answer is yes. After all, many crimes have distinctive aggressive elements, especially those crimes committed primarily by males.

Before going further, it is important to note that no one seriously argues that because androgens contribute to criminality, the "solution" to the crime problem is castration! Even so, the fact that castration can dramatically reduce aggression in farm animals provides additional evidence that androgens deserve research attention by criminologists.

One way to think about the potential influence of androgens on criminality involves tracking crime rates over the course of a human lifetime. In this regard, not only are males much more criminal than females, but most male offenses occur between the ages of 13 and 30. The relationships between sex, age, and criminality are illustrated in Fig. 1. This particular graph is based on persons convicted of serious offenses in England and Wales during 2000, but almost exactly the same pattern can be found anywhere else in the world (Ellis et al. 2009, pp. 17–20).

Keep in mind this sex-age-crime relationship shown in Fig. 1 and then look at Fig. 2. This second graph represents the average levels of testosterone in the human body throughout life (adjusted for body weight). This second graph traces testosterone all the way back to conception (for reasons to be explained shortly), but for now simply notice how the rise in testosterone closely parallels a rise in the probabilities of criminal convictions for males as well as for females. These two graphs along with research on how androgens affect the brain provide support for the argument that androgens contribute to criminality. Some of this research will be briefly reviewed.

How Androgens Work. As already noted, androgens appear to affect behavior by altering the brain. But one should add that these alterations primarily occur in two developmental stages, not just one. The first, called the *perinatal* (or *organizational*) *stage*, takes place before and shortly following birth. Refer back to Fig. 2, and notice how males are producing considerably more testosterone than females throughout gestation and for a month or two after birth.

It is during this perinatal stage that the brain becomes irreversibly *sexed*, i.e., made to be either masculine or feminine, although to widely varying degrees. In the simplest of terms, if androgens are present in low amounts, the brain will be left in its "neutral" or "default" form, which in mammals is feminine. On the other hand, if androgens are present in high amounts, the brain (along with

Hormones, the Brain, and Criminality, Fig. 2 Lifelong testosterone

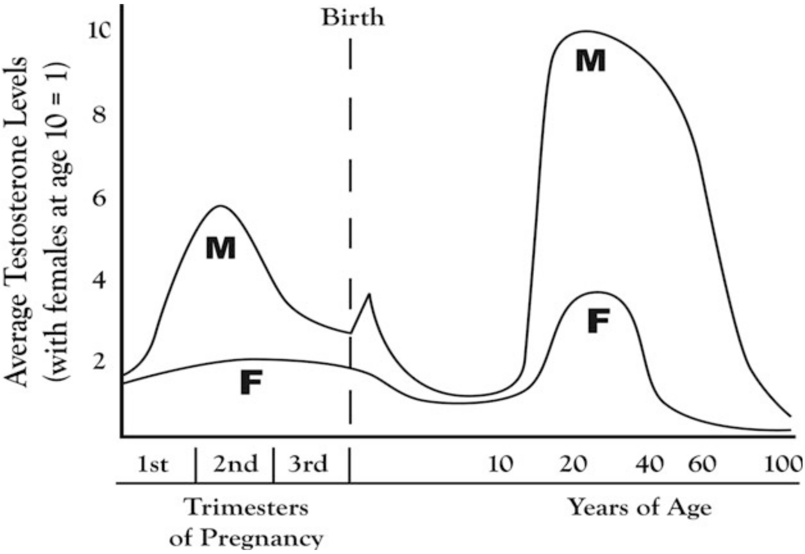

the rest of the body) will be masculinized. (As an aside, the fact that the "default sex" is female explains why males have nonfunctional nipples, testes that are really modified ovaries, and penises that are essentially enlarged clitorises.)

The second stage of brain sexing begins at puberty and continues throughout adulthood as shown in Fig. 2. Basically, to the extent that the male brain is exposed to androgens following the onset of puberty, it begins to function in a fully masculine mode. This second state is referred to as the *postpubertal* (or *activational*) *stage* (Sisk and Zehr 2005). However, it is important to realize that the *potential* for fully masculine behavior was laid down during the first stage; the second stage merely "turns on" the behavior.

Because female brains receive lower exposure to androgens than do the brains of males, their behavior tends to be much less "androgen influenced." Nevertheless, there is no simple straightforward relationship between high testosterone at a given point in time and the probability of someone robbing a bank or assaulting a neighbor. This fact has been demonstrated by studies that have correlated postpubertal testosterone levels in the blood or saliva with involvement in crime. The relationships tend to be quite modest (Ellis et al. 2009, pp. 208–210). However, one should keep in mind that neither the blood nor saliva provides a very good indication of how much testosterone is in the brain. Overall, the cause-and-effect relationship between androgens and criminality is complex even though it is still quite real.

Some Detailed Behavioral Effects of Androgens

Let us consider the behavioral consequences of high androgen exposure in greater detail because it is not just aggression that is affected. In other words, what types of behavior besides those that are illegal and violent are promoted by exposing the brain to androgens?

Competitive-Victimizing (CV) Behavior. Numerous animal experiments have carefully manipulated both perinatal and postpubertal influences of androgens on behavior. These experiments have shown that animals with high exposure exhibit more aggressive and combative tendencies, especially when their access to resources or mates is challenged or blocked (Archer 1994). Regarding the timing of the androgen exposure, it appears that perinatal exposure is actually more important than postpubertal exposure (Romeo et al. 2003).

Studies suggest that the effects of exposing the brain to androgens are similar to the effects in other mammals. Collectively, the behavioral effects of androgen exposure can be described as *competitive-victimizing (CV) behavior*

(Ellis et al. 2008). CV behavior primarily consists of elevated aggression and competitiveness, even in the face of considerable personal risk (Booth et al. 2006; McDermott et al. 2007). In humans, one manifestation of CV behavior is criminality, especially regarding violent and property offenses. The fact that males have higher levels of androgens than females offers a scientific explanation for why males are much more criminal than females throughout the world (Ellis et al. 2009; McDermott et al. 2007; van Bokhoven et al. 2006).

The idea that brain exposure to androgens is a major cause of sex differences in criminality and for the dramatic rise in male offending in the teenage years is visible when one compares Figs. 1 and 2. However, there is one interesting discrepancy revealed when comparing these two graphs. Notice that after age 30 or so, male offending rates drop off dramatically (Fig. 1) even though testosterone levels remain quite high (Fig. 2). How can this inconsistency be explained?

One proposal has been that the discrepancy reflects the influence that learning has on criminality (Ellis 2005). In particular, to the extent that males learn how to compete for resources and mates in ways that minimize the risk of retaliation by victims, their offending rates will rapidly decline. Thus, males with difficulty learning will have longer criminal "careers" than males who learn quickly.

Imagine two hypothetical males whose brains have been equally androgenized. Assume that one of these boys is able to learn quickly and the other one requires much more time. Theoretically, they should both be involved in some degree of delinquency during adolescents. However, the quick-learning boy should transition within a year or so into a socially responsible citizen who is able to compete for resources without violating the law. The slow learner, on the other hand, may take decades to finally "settle down."

One criminologist made the essential distinction between two types of offenders. She termed one "adolescent limited" and the other "life-course persistent" (Moffitt 1993). Both types engage in delinquency during their teenage years, but only the latter continues to offend as full adults. This distinction can be merged with the idea that androgens contribute to criminal activity by noting that both types of offenders will still be predominantly males. The males who become life-course-persistent offenders will be mainly those who have the greatest difficulties learning (nearly always due to some underlying neuropathology). This line of reasoning is supported by numerous studies documenting lower IQ scores, more learning disabilities, and higher school dropout rates among the most serious and persistent offenders (Ellis et al. 2009, pp. 150–159).

Did Criminal Behavior Evolve?

Another aspect of biosocial criminology has involved exploring the possibility that criminality has evolutionary roots. This means that the commission of crime could be part of how many individuals, particularly males, pass their genes on to subsequent generations (Ellis 2008; Rowe 1996).

At the heart of the argument is that over countless generations, females who have chosen mates who are unusually ambitious and competitive have left more of their (and their mate's) offspring in subsequent generations than is true for females who use other criteria when choosing mates. As a result, males may have evolved tendencies to victimize others, including in ways that most people deem criminal.

Hormones and Stress

Not all hormones are involved in sexual differentiation. Some hormones are released into the blood system primarily during times of stress. These are known as *stress hormones*. One stress hormone, called *cortisol*, has been repeatedly found associated with persistent criminal behavior (Ellis et al. 2009, pp. 210–213). However, unlike androgens that tend to be positively correlated with offending rates, cortisol levels are inversely correlated. In other words, criminals usually exhibit lower levels of cortisol levels than do persons in general.

How might the relationship between low cortisol and criminality be explained? Most proposals have reasoned as follows: Having elevated levels of cortisol tends to be relatively unpleasant. Individuals with high cortisol levels, or with cortisol levels that rise rapidly whenever they begin to experience stress, will be more easily deterred from engaging in whatever behavior is associated with elevated cortisol.

Consistent with such reasoning, low levels of cortisol have been found associated not only with criminality but also with childhood conduct disorders, persistent childhood aggression, and attention deficit hyperactivity disorders (Pajer et al. 2001). It is not very surprising to learn that all of these childhood traits have been found to be statistically associated with involvement in criminal behavior later in life (Ellis et al. 2009, p. 213).

Overall, persons who are most likely to exhibit these behavior patterns have lower “resting” levels of cortisol, or, when they experience stressful events, their cortisol levels rise very little (Gordis et al. 2006). Consequently, individuals who happen to be “hyporesponsive” to stress in terms of cortisol levels are more likely to be involved in crime.

Why do people vary in cortisol levels, especially in response to stress? Studies based on twins have revealed that genetic factors are important (Bartels et al. 2003; Wüst et al. 2000). There is also evidence that cortisol may actually interact with testosterone to affect offending behavior (Terburg et al. 2009). Ultimately, no matter what causes low cortisol levels and high androgen levels to be associated with criminality, their relationships provide further evidence that criminology without biology is incomplete. Biosocial criminologists are in a unique position to help bring these two important disciplines closer together.

Conclusions

Unlike the majority of criminologists who still argue that *only* learning and social variables are needed to explain criminal behavior, biosocial criminologists incorporate biological factors into their theories and research. This chapter has provided a thumbnail sketch of how two types of hormones appear to be relevant to understanding criminal behavior – sex hormones and stress hormones.

Regarding sex hormones, androgens appear to be central to the sexual differentiation of both the brain and the rest of the body. By affecting the brain, androgens increase the chances of many forms of competitive/victimizing behavior. Several extreme types of such behavior have been criminalized in all societies. Because male brains are exposed to higher levels of androgens, their contribution to crime is considerably greater than that of females.

At least one type of stress hormone – cortisol – has also been found associated with involvement in criminal behavior. Unlike androgens, it is unusually low levels of cortisol that are linked with criminality.

Despite the simplicity of the idea that both sex hormones and stress hormones are associated with criminality, one should appreciate that there are complex chains of events that still need to be considered as one attempts to understand the connections. For instance, most effects of hormones on behavior appear to be mediated by neurotransmitters (chemicals in the brain that help transmit signals from one nerve cell to another) (Birger et al. 2003).

Will biosocial criminology ultimately help reduce the crime problem? In all likelihood, yes (Archer 1994; Ellis 2008; MacGillivray 1997). Some critics of biosocial criminology fear that this perspective could reopen the door to Nazi-style “solutions” to crime and other social problems (Mitchell and Richard 1998). However, the risks are low in democratic societies where people insist on balancing concerns about public safety with respect for individual freedom and liberty. Also, most people understand that human behavior is much too complex to ever be easily manipulated exclusively at biological levels.

Related Entries

▶ Evolutionary Theories of Criminal Behaviors

Recommended Reading and References

Archer J (1994) Testosterone and aggression. In: Hillbrand M, Pallone NJ (eds) The psychobiology of aggression: engines, measurements, control. Hawthorn, New York, pp 3–26

Bartels M, de Geus EJC, Kirschbaum C, Sluyter F, Boomsma DI (2003) Heritability of daytime cortisol levels in children. Behav Genet 33:421–433

Birger M, Swartz M, Cohen D, Alesh Y, Grishpan C, Kotelr M (2003) Aggression: the testosterone-serotonin link. Isr Med Assoc J 5:653–658

Booth A, Kivlighan KT, Mazur A, Granger DA (2006) Testosterone and social behavior. Soc Forces 85:167–191

Bouchard TJ, McGue M (2003) Genetic and environmental influences on human psychological differences. J Neurobiol 54:4–45

Cooper JA, Walsh A, Ellis L (2010) Is criminology moving toward a paradigm shift? Evidence from a survey of the American Society of Criminology. J Crim Justice Educ 21:332–347

Ellis L (2005) Theoretically explaining biological correlates of criminal behavior. Eur J Criminol 2:287–315

Ellis L (2008) Reducing crime evolutionarily. In: Duntley JD, Shackelford TK (eds) Evolutionary forensic psychology. Oxford University Press, New York, pp 249–267

Ellis L, Das S, Buker H (2008) Androgen-promoted physiological traits and criminality: a test of the evolutionary neuroandrogenic theory. Personal Individ Differ 44:701–711

Ellis L, Beaver K, Wright J (2009) Handbook of crime correlates. Elsevier, Amsterdam

Gordis EB, Granger DA, Susman EJ, Trickett PK (2006) Asymmetry between salivary cortisol and [alpha]-amylase reactivity to stress: relation to aggressive behavior in adolescents. Psychoneuroendocrinology 31:976–987

Kreek MJ, Nielsen DA, Butelman ER, LaForge KS (2005) Genetic influences on impulsivity, risk taking, stress responsivity and vulnerability to drug abuse and addiction. Nat Neurosci 8:1450–1457

Lawson AE (2003) The neurological basis of learning, development, and discovery: implications for science and mathematics instruction, vol 18. Springer, Amsterdam

MacGillivray L (1997) California’s mandatory chemical castration program for repeat sex offenders: an analysis of the legislation under German and American constitutional law. Suffolk Transnatl Law Rev 21:143–178

McDermott R, Johnson D, Cowden J, Rosen S (2007) Testosterone and aggression in a simulated crisis game. Ann Am Acad Polit Soc Sci 614:15–33

Mitchell WJ, Richard A (1998) Taboo until today? The coverage of biological arguments in criminology textbooks, 1961 to 1970 and 1987 to 1996. J Crim Justice 26:1–19

Moffitt TE (1993) Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. Psychol Rev 100:674–701

Pajer K, Gardner W, Rubin RT, Perel J, Neal S (2001) Decreased cortisol levels in adolescent girls with conduct disorder. Arch Gen Psychiatry 58:297–302

Romeo RD, Schulz KM, Nelson AL, Menard TA, Sisk CL (2003) Testosterone, puberty, and the pattern of male aggression in Syrian hamsters. Dev Psychobiol 43:102–108

Rowe DC (1996) An adaptive strategy theory of crime and delinquency. In: Hawkins JD (ed) Delinquency and crime: current theories. Cambridge University Press, Cambridge, pp 268–314

Schermer JA, Vernon PA (2008) A behavior genetic analysis of vocational interests using a modified version of the Jackson Vocational Interest Survey. Personal Individ Differ 45:103–109

Sisk CL, Zehr JL (2005) Pubertal hormones organize the adolescent brain and behavior. Front Neuroendocrinol 26:163–174

Terburg D, Morgan B, van Honk J (2009) The testosterone-cortisol ratio: a hormonal marker for proneness to social aggression. Int J Law Psychiatry 32:216–223

van Bokhoven I, van Goozen SHM, van Engeland H, Schaal B, Arseneault L, Séguin JR et al (2006) Salivary testosterone and aggression, delinquency, and social dominance in a population-based longitudinal study of adolescent males. Horm Behav 50:118–125

Walsh A, Beaver K (eds) (2008) Biosocial criminology: new directions in theory and research. Sage, Thousand Oaks

Wüst S, Federenko I, Hellhammer DH, Kirschbaum C (2000) Genetic factors, perceived chronic stress, and the free cortisol response to awakening. Psychoneuroendocrinology 25:707–720

Hot Dots

▶ Prediction and Crime Clusters

Hot Products

▶ Designing Products Against Crime
▶ Prediction and Crime Clusters

Hot Smudges

▶ Prediction and Crime Clusters

Hot Spots

► Criminal Careers of Places
► Prediction and Crime Clusters

Hot Spots and Place-Based Policing

Cody W. Telep[1] and David Weisburd[2,3]
[1]School of Criminology and Criminal Justice, Arizona State University, Phoenix, AZ, USA
[2]Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA
[3]Faculty of Law, The Hebrew University, Mt. Scopus, Jerusalem, Israel

Overview

Over the past two decades, a series of rigorous evaluations have suggested that police can be effective in addressing crime and disorder when they focus in on small units of geography with high rates of crime (see Braga et al. in press; National Research Council [NRC] 2004; Weisburd and Eck 2004). These areas are typically referred to as hot spots, and policing strategies and tactics focused on these areas are usually referred to as hot spots policing or place-based policing. This place-based focus stands in contrast to traditional notions of policing and crime prevention more generally, which have often focused primarily on people (see Weisburd 2008). For example, police work often begins with a response to citizens who call the police, and police are very focused on identifying and arresting offenders who commit crimes. While hot spots policing does not ignore the offenders found within crime hot spots, the focus is very much on the places where crime is occurring.

Police, of course, have never ignored geography entirely. Police beats, precincts, and districts determine the allocation of police resources and dictate how police respond to calls and patrol the city. With place-based policing, however, the concern is with much smaller units of geography than the police have typically focused on. Places here refer to specific locations within the larger social environments of communities and neighborhoods, such as addresses or street blocks. Crime prevention effectiveness is maximized when police focus their resources on these micro-units of geography.

The evidence suggesting the effectiveness of place-based policing efforts is detailed in the sections that follow. A definition of crime hot spots is first discussed followed by a review of research suggesting that crime is highly concentrated at micro-units of geography, and these concentrations remain stable over time. Hot spots policing is then defined, and the strategies place-based policing can entail are discussed. The extent to which hot spots policing has diffused across American policing is also examined. The empirical evidence for hot spots policing and the different approaches to addressing high crime places are then reviewed. Finally, some suggestions for future research are presented.

Defining Crime Hot Spots and the Concentration of Crime at Place

Defining Crime Hot Spots

Crime hot spots are small units of geography with high rates of criminal activity. The specific geographic area that makes up a hot spot varies across studies, ranging from individual addresses or buildings (e.g., Sherman et al. 1989), to single street segments (i.e., both sides of a street from intersection to intersection; e.g., Sherman and Weisburd 1995), to small groups of street segments with similar crime problems such as a drug market (e.g., Weisburd et al. 2006). Hot spots are smaller than the units that police departments typically use for dividing up patrol resources such as patrol beats, zones, or sectors. Hot spots can be viewed as micro-places, to differentiate them from larger geographic units such as communities and neighborhoods that have traditionally been of interest to criminologists discussing crime and place (see Weisburd 2008).

There is also no firm rule as to how much crime must be found in a micro-place before it can be categorized as a hot spot. This will vary across studies and contexts, based on the overall rate of crime in a jurisdiction. In the studies of hot spots policing described below, hot spots are typically defined by drawing upon a rank ordered list of the highest crime locations (e.g., addresses or streets) in the city based on calls for service or crime incidents. Thus, places are not defined as crime hot spots by reaching an absolute threshold of crime, but instead because of extremely high levels of criminal activity relative to other places in the city.

Crime Concentration at Hot Spots

Regardless of how hot spots are defined, a number of studies over the past 20 to 30 years have found that a relatively small number of micro-places are responsible for a significant amount of total crime in a city (e.g., see Weisburd and Mazerolle 2000; Weisburd et al. 2004). The hottest spots in a city, regardless of the specific unit of analysis chosen, are responsible for a large chunk of a jurisdiction's crime problem. Perhaps the most influential of these studies was Sherman et al.'s (1989) analysis of emergency calls to street addresses over a single year. Sherman et al. found that only 3.3 % of the addresses in Minneapolis, Minnesota, produced just over 50 % of all calls to the police. If crime were randomly distributed across the 115,000 addresses in Minneapolis, one would not expect any places to have 15 or more calls in a single year. Instead, Sherman et al. found 3,841 such addresses, indicating that crime is far more concentrated than would be expected by chance.

A study conducted by Weisburd et al. (2012) not only confirms the concentration of crime, but also the stability of such concentrations across a long time span. Weisburd et al. examined street segments in the city of Seattle from 1989 through 2004. They found that each year, 50 % of crime incidents occurred on between 4.7 and 6.1 % of the street segments. These data overall illustrate a kind of "law of concentration" for crime, suggesting that crime is heavily clustered in cities with fewer than 5 or 10 % of addresses, street segments, or small clusters of addresses and street segments accounting for a majority of crime in a city each year (Weisburd et al. 2012). Thus, police can target a sizable proportion of citywide crime by focusing in on a small number of places.

Sherman (1995) argues that such clustering of crime at places is even greater than the concentration of crime among individuals. Comparing the concentration of calls in Minneapolis to the concentration of offending in the Philadelphia birth cohort study, he notes that future crime is "six times more predictable by the address of the occurrence than by the identity of the offender" (1995: 36-37). Sherman asks, "Why aren't we thinking more about wheredunit, rather than just whodunit?"

The Stability of Crime Hot Spots

Concentration by itself has important implications for the effectiveness of hot spots policing. However, it is also important to examine the stability of crime hot spots over time. If hot spots of crime shift rapidly from place to place each year, it makes little sense to focus crime control resources at such locations, because they would naturally become free of crime without any criminal justice intervention (Spelman 1995).

Data on crime stability suggests that hot places tend to remain hot over time. Spelman (1995), for example, examined calls for service at schools, public housing projects, subway stations, and parks and playgrounds in Boston. He found evidence of a very high degree of stability of crime at the "worst" of these places over a 3-year period. Spelman concluded that it "makes sense for the people who live and work in high-risk locations, and the police officers and other government officials who serve them, to spend the time they need to identify, analyze and solve their recurring problems" (1995: 131).

The most comprehensive examination of the stability of crime at place over time was conducted by Weisburd et al. (2012) in their study of crime incidents at street segments in Seattle. Using group-based trajectory analysis to categorize street segments according to similar developmental patterns, they identified 22 specific trajectory groups. The most important finding in their study was that crime remained fairly stable at places over time, particularly in a small group of high crime segments they refer to as the "chronic group." These street segments remained among the hottest in the city throughout the 16-year study period. Representing just 1 % of the street segments in Seattle, the chronic streets accounted for 22 % of the crime in the city. Hot spots of crime thus appear to remain hot over long periods of time. This can be contrasted with developmental studies of individual offending where there is often tremendous change across relatively short periods, especially for high-rate offenders (see Horney et al. 1995), and a great deal of variability across the life-course as many high-rate "hot" offenders in adolescence age out of crime and cool off in adulthood.

The Distribution of Crime Hot Spots

How are hot spots of crime distributed geographically across cities? This question is important to explore, because it gets at the issue of whether it is necessary for police to focus in on micro-places for hot spots policing. For example, if all the crime hot spots were concentrated in only one or two neighborhoods in a city, then neighborhood or beat-level initiatives might be just as effective as hot spots approaches in addressing crime concentrations.

Evidence to date suggests that crime hot spots can be found throughout a city. Weisburd and Mazerolle (2000), for example, identified 56 drug markets in Jersey City, New Jersey. Although the drug markets were more concentrated in socially disadvantaged areas, they could even be found in areas that were generally seen as more established and better off. Weisburd and Mazerolle argued that even "good" neighborhoods can have "bad" places. Importantly, most places even in very disadvantaged neighborhoods were relatively free of serious drug problems.

Groff et al. (2010) examined the distribution of all crime hot spots in Seattle. While hot spots may be more likely in high activity places in the city like the central business district, they tend to be found in a variety of different neighborhoods, and even neighborhoods with higher concentration of hot spots tend to have a sizable proportion of crime free locations. Groff et al. (2010) conclude that much information about crime would be missed by focusing on larger units such as neighborhoods. These findings overall suggest the usefulness of a policing approach that narrows the focus to hot street segments or micro-places as opposed to hot neighborhoods or hot communities.

Place Risk Factors

Research by Weisburd and colleagues (2012) provides new insights into the key risk factors that help explain whether places become hot spots or not. They overall find that social and opportunity factors are very highly concentrated at the street block level, and these concentrations are generally very closely linked to concentrations of crime. This suggests that certain factors may contribute to street segments remaining hot spots of crime over time, and hence could be viewed as criminogenic risk factors for a street becoming (and remaining) a crime hot spot.

Some of these factors present opportunities for the police in addressing crime at chronic hot spot streets. For example, the number of employees working on a block is a strong predictor of a street segment being a chronic hot spot. While the police have little control over the number of people working on a block, they can control the level of guardianship they provide on each block and how they approach the potential crime attractors and generators on each block (e.g., working to install CCTV in parking areas on streets with many employees to address car break-ins). As another example, higher numbers of high-risk juveniles (defined as those with high rates of truancy or poor school performance) on a block substantially increased the likelihood a street would be a chronic hot spot. Increased enforcement from police could reduce levels of truancy and limit the amount of time these juveniles spend on their street unsupervised.

Current State of Hot Spots Policing

What Is Hot Spots Policing?

Hot spots policing, also sometimes referred to as place-based policing (see Weisburd 2008),

Hot Spots and Place-Based Policing, Table 1 Percentage of police agencies using computers for specific tasks by population served (Adopted from Tables 22 and 25 of Reaves 2010)

Population served	Crime analysis	Crime mapping	Hot spot identification	Providing crime maps to officers in-field
1,000,000+	100	100	92	31
500,000–999,999	100	100	100	29
250,000–499,999	100	100	80	48
100,000–249,999	96	94	66	32
50,000–99,999	88	82	56	31
25,000–49,999	69	60	31	16
10,000–24,999	53	41	19	11
2,500–9,999	37	23	9	9
Less than 2,500	21	11	5	8
All sizes	38	27	13	11

covers a range of police responses that all share in common, a focus of resources on the locations where crime is highly concentrated. Just as the definition of hot spots varies across studies and contexts, so do the specific tactics police use to address high crime places. There is no one way to implement hot spots policing. As Weisburd (2008) notes, approaches can range rather dramatically across interventions.

For example, the strategies of place-based policing can be as simple as dramatically increasing officer time spent at hot spots patrol, as was the case in the Minneapolis Hot Spots Policing Experiment (Sherman and Weisburd 1995). But place-based policing can also take a much more complex approach to the amelioration of crime problems. In the Jersey City Drug Market Analysis Project (Weisburd and Green 1995), for example, a three-step program (including identifying and analyzing problems, developing tailored responses, and maintaining crime control gains) was used to reduce problems at drug hot spots. In the Jersey City Problem-Oriented Policing Project (Braga et al. 1999), a problem-oriented policing approach was taken in developing a specific strategy for each of the small areas defined as violent crime hot spots. The results of these interventions are discussed more in the next section.

Are Agencies Using Hot Spots Policing?

To what extent are police agencies using hot spots policing in practice? To answer this, data from the 2007 Law Enforcement Management and Administrative Statistics (LEMAS) data were examined (see Reaves 2010), as well as survey data that asked police agencies about hot spots policing. While data on specific tactics are limited in the LEMAS survey, there are a number of questions related to police technology (see Table 1). Crime analysis and crime mapping, in particular, are very important for the successful implementation of hot spots policing. The results overall suggest widespread use of these technologies in larger departments but not in smaller ones. For example, while over 90 % of the largest agencies are using computers for hot spot identification, just 13 % of departments overall are. Even in moderately sized cities (population 100,000–249,999), just 56 % of departments use computers to identify hot spots. When it comes to using computers for crime analysis and crime mapping, results are similar. One hundred percent of the largest departments make use of computers for such tasks, but only 38 % of agencies overall use computers for crime analysis, and 27 % use computers for crime mapping. In terms of patrol officer's access to crime data, just 31 % of the largest agencies and 11 % of agencies overall provide officers access to crime maps in their patrol cars.

The use of computers for crime mapping and analysis has increased since the 2003 LEMAS survey. For example, the percentage of officers working in a department that uses computers for crime mapping jumped from 57 % in 2003 to 75 % in 2007 and for hot spot identification the increase was from 45 % in 2003 to 58 % in 2007 (Reaves 2010). Research by Weisburd and Lum (2005) suggests that the use of crime mapping diffused quickly across policing from the mid-1990s through 2001. They note that the adoption of crime mapping was closely linked to the use of hot spots policing. When agencies were asked why they developed crime mapping in the department, the largest response category was "to facilitate hot spots policing." Kochel (2011: 352) cites similar statistics, noting "Within 15 years, hot spots policing had diffused almost completely throughout large U.S. police departments." These latest LEMAS data suggest that the use of crime mapping continues to increase, but smaller agencies are lagging behind.

Data on the adoption of hot spots policing also come from a Police Executive Research Forum (PERF) survey of 176 policing agencies of various sizes on their efforts to reduce violent crime (Koper 2008). Responses showed close to two-thirds (63 %) of agencies used hot spots policing to reduce violent crime. This was by far the most popular response to the question of how to address violence. When asked what sort of places the agency defines as a hot spot, the majority of respondents noted addresses or intersections (61 %) or clusters of addresses (58 %). However, a majority of respondents (57 %) also identified neighborhoods as potential hot spots and a sizable minority pointed to patrol beats (41 %) as the sort of place that would be defined as a hot spot. Crime is not as highly concentrated at these larger units of geography, so focusing on them may not produce the same crime control benefits as using more micro-units of analysis.

The PERF survey also provides data on what tactics agencies are using to respond to homicide, robbery, assault, gang violence, and drug crime hot spots. While agencies identified a number of different potential responses to hot spots, the top responses in each category tended to be some combination of problem solving and analysis, directed patrol, community policing/partnerships, and targeting known offenders (Koper 2008). These strategies and the empirical evidence supporting them are discussed more below, focusing, in particular, on problem solving and directed patrol.

It is important to note that while the data discussed above provide an estimate of the extent to which police agencies have adopted hot spots policing, they do not make clear how frequently agencies are actually using hot spots policing in practice. In other words, while it appears that a substantial portion of larger agencies are using computers to identify crime hot spots, what percentage of officer time is actually being spent on hot spots policing in these agencies? Data on the level of implementation of hot spots policing is not available at the department level, but such data would be important to better understand to what extent hot spots policing is becoming a primary tactic in police agencies.

Empirical Evidence on the Effectiveness of Hot Spots Strategies

The evidence base for the effectiveness of hot spots policing in reducing crime and disorder is particularly strong. As the NRC (2004: 250) review of police effectiveness noted, "studies that focused police resources on crime hot spots provided the strongest collective evidence of police effectiveness that is now available." The Braga et al. (in press) systematic review came to a similar conclusion; although not every hot spots study has shown statistically significant findings, the vast majority of such studies have (20 of 25 tests from 19 experimental or quasi-experimental evaluations reported noteworthy crime or disorder reductions), suggesting that when police focus in on crime hot spots, they can have a significant beneficial impact on crime in these areas. In Braga and colleagues' (in press) meta-analysis, they found an overall mean effect size of 0.184, suggesting a significant benefit of the hot spots approach in treatment compared to control areas. As Braga (2007: 18) concluded "extant evaluation research seems to provide fairly robust evidence that hot spots policing is an effective crime prevention strategy."

Importantly, there was little or no evidence to suggest that spatial displacement was a major concern in hot spots interventions. Spatial crime displacement is the notion that efforts to eliminate specific crimes at a place will simply cause criminal activity to move elsewhere, thus negating any crime control gains. Braga et al. (in press) found significant evidence of spatial displacement in only one study (Ratcliffe et al. 2011) and even here the amount of displacement was far less than the main crime prevention benefit of the intervention. Thus, in nearly every study, crime did not simply shift from hot spots to nearby areas (see also Weisburd et al. 2006). Indeed, a more likely outcome of such interventions was a diffusion of crime control benefits (Clarke and Weisburd 1994) in which areas surrounding the target hot spots also showed a crime and disorder decrease. Displacement is not inevitable, in part, because hot spots tend to have specific features that make them attractive targets for criminal activity, and these same features may not exist on neighboring blocks. For example, in the Weisburd et al. (2006) study, the prostitution hot spot targeted by police had few homes and many vacant buildings, making it an attractive site for prostitution activity. In contrast, one of the catchment areas near the target site had many more residences, making it more likely that the police would be called when prostitution occurred. In this context, prostitutes could not easily move their illegal activity to areas nearby.

The experimental evidence on hot spots policing is discussed below, focusing when possible on what tactics seem to be most effective in addressing crime hot spots. The literature has not provided the same level of guidance on what specifically police officers should be doing at hot spots to most effectively reduce crime. As Braga (2007: 19) notes "Unfortunately, the results of this review provide criminal justice policy makers and practitioners with little insight on what types of policing strategies are most preferable in controlling crime hot spots." The update to Braga's (2007) review by Braga and colleagues (in press) does provide some additional guidance, suggesting that problem-oriented hot spots interventions may be somewhat more effective than simply increasing police presence, although the authors caution these comparisons are based on a small number of studies.

The first randomized experiment on hot spots policing, the Minneapolis Hot Spots Patrol Experiment (Sherman and Weisburd 1995), used computerized mapping of crime calls to identify 110 hot spots of roughly street-block length. Police patrol was doubled on average for the experimental sites over a 10-month period. Officers in Minneapolis were not given specific instructions on what activities to engage in while present in hot spots. They simply were told to increase patrol time in the treatment hot spots, and the activities they engaged in ranged a good deal from more proactive problem-solving efforts to simply sitting in the car parked in the center of a hot spot street segment. The study found that the experimental as compared with the control hot spots experienced statistically significant reductions in crime calls and observed disorder. The significant decline in calls for service was driven largely by a decline in soft crime calls (e.g., disturbances, drunks, noise, and vandalism). While soft crime calls declined 7.2–15.9 % in the treatment group relative to the control group, depending on the time period examined, hard crime calls, (e.g. burglaries, auto thefts, and assaults) declined a non-statistically significant 2.6–5.9 %.

The results overall though suggested increased police presence could have a significant effect on crime, particularly disorder and less serious crime. This marked a major change from the conventional wisdom about the impact the police could have on crime. The most influential prior study of police patrol at the time was the Kansas City Preventive Patrol Experiment (Kelling et al. 1974), which found that neither doubling nor removing preventive patrol in a beat had a significant impact on crime or victimization. The finding that police randomly patrolling beats is not an effective crime deterrent makes sense based on the review of the crime concentration literature above. Since crime is very concentrated across cities, it makes little sense from an effectiveness and efficiency standpoint to respond with a strategy relying on the random distribution of police resources across large geographic areas. The Minneapolis study and the other hot spots studies described below suggested police could have an impact on crime by appropriately focusing their resources on the locations where crime was highly concentrated.

Significant crime and disorder declines are also reported in five other randomized controlled experiments that tested a more tailored, problem-oriented approach (see Braga and Weisburd 2010) to dealing with crime hot spots. In the Jersey City Problem-Oriented Policing in Violent Places experiment (Braga et al. 1999), the exact response varied by hot spot, but the responses all included some aspect of aggressive order maintenance and most included efforts to make physical improvements to the area (e.g., removing trash, improving lighting) and drug enforcement. Strong statistically significant reductions in total crime incidents and total crime calls were found in the treatment hot spots relative to the control hot spots. Social and physical observation data showed improvement in visible disorder in 10 of the 11 treatment areas compared to the control sites after the intervention.

In the Jersey City Drug Market Analysis Program experiment (Weisburd and Green 1995), a step-wise problem-solving model was compared to generalized enforcement in drug hot spots. The treatment group received a three-stage intervention. In the planning stage, officers collected data on the physical, social, and criminal characteristics of each area; in the implementation stage, officers coordinated efforts to conduct a crackdown at the hot spot and used other relevant responses to address underlying problems; and finally, in the maintenance stage, officers attempted to maintain the positive impact of the crackdown. The experimental sites had significantly smaller increases in disorder calls compared to the control sites. In particular, the project had a positive impact on calls related to public morals and suspicious persons.

In the Oakland Beat Health study, Mazerolle and colleagues (2000) evaluated a civil remedies initiative to deal with high crime addresses and blocks. The intervention involved a police officer and police service technician visiting a site to identify and analyze the problem and to make contact with the property owner or place manager to try to address the problems. The intervention typically involved pressuring third parties (e.g., the landlord of a problem apartment building) to make changes to improve property conditions or face the possibility of civil action. Results showed significant reductions in drug calls for service at the treatment hot spots relative to the controls.

In Lowell, Massachusetts, a problem-oriented policing intervention to address disorder was associated with a statistically significant 20 % reduction in crime and disorder calls for service at the treatment hot spots compared to the control hot spots (Braga and Bond 2008). Systematic observations confirmed the official crime data results, showing significant reductions in social and physical disorder at the treatment places relative to the control places.

The Braga and Bond (2008) experiment also included a mediation analysis to assess which hot spots strategies were most effective in reducing crime. Results suggested that situational prevention strategies had the strongest impact on crime and disorder. Situational strategies are concerned with the physical, organizational, and social environments that make crime possible, and they focus on efforts to disrupt situational dynamics that allow crime to occur by, for example, increasing risks or effort for potential offenders or reducing the attractiveness of potential targets. Such approaches are often a prominent part of hot spots interventions, particularly those involving problem solving and include things like securing lots, razing abandoned buildings, and cleaning up graffiti. Increases in misdemeanor arrests made some contribution to the crime control gains in the treatment hot spots, but were not as influential as the situational efforts. Social service interventions did not have a significant impact on crime and disorder. These findings suggest not only the importance of situational crime prevention as a strategy for addressing crime facilitators in hot spots, but also that aggressive order maintenance through increases in misdemeanor arrests may not be the most effective way of addressing high disorder places. The potential negative consequences of intensive enforcement in the form of decreased perceptions of police legitimacy are discussed below (see Braga and Weisburd 2010).

More recently, in an important experimental study in Jacksonville, Florida, Taylor et al. (2011) were the first to compare different hot spot treatments in the same study with one treatment group receiving a more standard saturation patrol response and the second receiving a problem-oriented response that focused on officers analyzing problems in the hot spot and responding with a more tailored solution. Results showed a decrease in crime (though not a statistically significant decrease) in the saturation patrol hot spots, but this decrease lasted only during the intervention period and disappeared quickly thereafter. In the problem-oriented policing hot spots, there was no significant crime decline during the intervention period, but in the 90 days after the experiment, street violence declined by 33 % (a statistically significant decline). These results offer the first experimental evidence suggesting that problem-oriented approaches to dealing with crime hot spots may be more effective than simply increasing patrols in high crime areas. They also suggest that problem-solving approaches may take more time to show beneficial results, but any successes that come from a problem-oriented framework may be more long-lasting in nature.

Braga and Weisburd (2010) detail the desirability of problem-oriented policing as a strategy for long-term crime reduction in chronic hot spots. They recognize that even when agencies use problem-solving approaches, they often tend to prioritize more traditional, enforcement-oriented responses. Drawing upon the results of Braga and Bond (2008) and other hot spots studies, they argue that "situational" problem-oriented policing is not only more innovative but also more likely to produce significant crime control benefits. As they conclude "based on the available empirical evidence, we believe that police departments should strive to develop situational prevention strategies to deal with crime hot spots. Careful analyses of crime problems at crime hot spots seem likely to yield prevention strategies that will be well positioned to change the situations and dynamics that cause crime to cluster at specific locations" (Braga and Weisburd 2010: 182–183).

Two other recent experiments shed new light on the effectiveness of particular approaches to dealing with crime hot spots. While the initial Minneapolis study did not include a systematic examination of officer activities in the hot spots, subsequent analyses by Koper (1995) provide some insight into how much time officers should be spending in hot spots to maximize residual deterrence. Koper analyzed observational data on nearly 17,000 instances when police drove through or stopped at a hot spot and examined the time from when the officer(s) left the location until the next occurrence of criminal or disorderly behavior. Using survival analysis techniques, he found that each additional minute of time officers spent in a hot spot increased survival time by 23 %. Survival time here refers to the amount of time after officers departed a hot spot before disorderly activity occurred. The ideal deal time spent in the hot spot was 14 to 15 min; after about 15 min, there were diminishing returns and increased time did not lead to greater improvements in residual deterrence. This phenomenon is often referred to as the "Koper curve" as graphing the duration response curve shows the benefits of increased officer time spent in the hot spot until a plateau point is reached at around 15 min (see Koper 1995, Fig. 1). As Koper (1995: 668) notes "police can maximize crime and disorder reduction at hot spots by making proactive, medium-length stops at these locations on a random, intermittent basis." Koper (1995) argues for an approach in which police travel between hot spots, spending about 15 min in each hot spot to maximize residual deterrence, and moving from hot spot to hot spot in an unpredictable order, so that potential offenders recognize a greater cost of offending in these areas because police enforcement could increase at any moment.

These recommendations suggest a potential reason why an experimental hot spots intervention at crack houses in Kansas City had quickly decaying effects (Sherman and Rogan 1995). The crackdowns on drug locations led to significant but modest improvements in the experimental street blocks, but these crackdowns were only one-time events and hence, residual deterrence was limited. Within a few weeks, crime on the streets where crack houses were located returned to pre-crackdown levels.

Koper's (1995) recommendations were recently applied to the design of a hot spots policing experiment in Sacramento, California. Officers were explicitly instructed to rotate between treatment group hot spots and to spend about 15 min in each hot spot. Results suggest the Koper (1995) approach to hot spots policing had a significant impact on crime. Treatment group hot spots had significantly fewer calls for service and Part I crime incidents than control group hot spots when comparing the 3-month period of the experiment in 2011 to the same period in 2010 (Telep et al. in press).

Ratcliffe and colleagues (2011) evaluated the impact of foot patrol in crime hot spots in a randomized experiment in Philadelphia. Foot patrol has traditionally been viewed as an effective strategy for reducing fear of crime, not actual crime. Results suggested significant declines in violence in the treatment hot spots compared to the control sites. The intervention was particularly effective for hot spots that reached a threshold of violence (i.e., the hottest hot spots). Overall, the 12-week intervention prevented a net total of 53 violent crimes in the target areas. These findings suggest the importance of reconsidering the effectiveness of foot patrol when intensive foot patrol is used in high crime micro-places.

Hot Spots Policing and Police Legitimacy

In sum, the empirical research is highly supportive of the effectiveness of hot spots policing in reducing crime and disorder. The effectiveness of policing, however, is also dependent on public perceptions of the legitimacy of police actions (NRC 2004; Tyler 1990). The police need the support and cooperation of the public to effectively combat crime and maintain social order in public spaces. Legitimacy here refers to the public belief that there is a responsibility and obligation to voluntarily accept and defer to the decisions made by authorities (Tyler 1990). A number of scholars have recently argued that intensive police interventions such as hot spots policing may erode citizen perceptions of the police (e.g., see Kochel 2011; Rosenbaum 2006). Rosenbaum (2006), for example, argues that enforcement-oriented hot spots policing runs the risk of weakening police-community relations. Aggressive tactics can drive a wedge between the police and communities, as the latter can begin to feel like targets rather than partners. This is particularly relevant in high crime minority communities where perceptions of the police already tend to be negative. This has implications for the crime control effectiveness of hot spots policing as Tyler (1990) has argued that legitimacy is an important predictor of long-term compliance with the law. If hot spots policing interventions weaken perceptions of legitimacy, then the short-term crime control gains from the intervention might be offset by long-term increases in criminal offending.

Despite arguments that intensive interventions such as hot spots policing will have negative impacts on police legitimacy, the limited evidence available from such interventions tends to suggest that citizens living in targeted areas welcome the increased police presence. Recent research from three cities in San Bernardino County, California, for example, found that a broken windows style intervention in street segments had no impact on resident perceptions of police legitimacy (Weisburd et al. 2011). It would also be useful to assess the views of residents of areas nearby hot spots and target sites to assess whether such interventions have spillover effects (either positive or negative) on legitimacy perceptions. Additionally, research to date has not attempted to measure how the individuals who were stopped and searched by the police perceive such programs. Ideally, a study would compare the perceptions of individuals stopped as part of a hot spots intervention to those stopped under standard routine preventive patrol. Braga and Weisburd (2010) describe a community-oriented approach to hot spots policing focused on community consultation on the tactics used in hot spots and efforts to ensure that hot spots policing strategies do not reduce levels of police legitimacy. The knowledge base on this important topic of how police fairness and effectiveness intersect in hot spots policing remains limited.

Conclusions and Future Research

Overall this entry suggests that significant crime control benefits can result from the use of hot spots policing and place-based policing. Basic research suggests that the action of crime is at very small geographic units of analysis, such as street segments or small groups of street blocks. Such places also offer a stable target for police interventions, as contrasted with the constantly moving targets of criminal offenders. Crime at place is not simply a proxy for larger area or community effects; indeed, the basic research evidence suggests that much of the action of crime occurs at very small geographic units of place. This basic research is reinforced by a strong body of experimental evidence for the effectiveness of place-based policing in reducing crime and disorder without simply displacing crime to areas nearby. Additionally, rigorous evaluation research in policing suggests that place-based interventions, including hot spots policing, tend to be more effective than interventions focused on people (Weisburd and Eck 2004).

While hot spots policing already has a larger evidence base than many police interventions, there are still areas where future research can shed new light. First, as noted above, additional research is needed on how hot spots policing impacts perceptions of police legitimacy in hot spot locations and the implications this has for the long-term effects of intensive police efforts. In particular, studies to date have not adequately examined the perceptions of offenders arrested or stopped in hot spot areas. Second, nearly all hot spots studies have taken place in highly populated urban areas. While smaller departments with only a few officers are unlikely to have a significant quantity of street-block size hot spots, future research should better examine the applicability of hot spots policing across varying contexts. Third, there are no studies that examine the long-term effects of hot spots interventions. In some respects, this makes sense. Hot spots interventions that focus on increasing patrol time to maximize deterrence are not expected to have an impact long after the police have left. But more problem-oriented hot spots approaches are designed to have longer-term impacts by changing the dynamics of places and so longer follow-up periods would be desirable in future projects. Fourth and in a related way, there should be more research that examines how police can disrupt long-term crime concentrations. In other words, can police address certain criminogenic factors that would stop the same places from appearing in the chronic trajectory group year after year as they did in Seattle? Fifth, further study on the importance of time in understanding crime hot spots is needed. To date, researchers have typically focused on annual data to generate hot spot locations, but the year may not be the most appropriate temporal unit. Police agencies must find a balance between targeting areas that have shown some stability in crime rates over time, while also allowing for flexibility in shifting resources when data suggest the movement of hot spots or the emergence of new hot spots. Sixth, future research should consider the cost-effectiveness of hot spots policing. How much is crime reduced per dollar spent or unit of officer time devoted to hot spots policing? How does the cost-effectiveness of hot spots policing compare to other policing strategies such as random preventive patrol? Assessing cost-effectiveness is particularly important in times like the present when many police agencies are facing budget cuts.

Finally, there is a need for better data on what agencies are doing in the field. LEMAS, for example, suggests the largest agencies have many of the capabilities and technologies needed for properly implementing hot spots policing, but many smaller agencies still lag behind. LEMAS is limited though in providing data on what departments are engaged in day-to-day. Because it is a massive national survey, there are limits to the number of questions that can be asked about daily practices, and even with additional questions, the LEMAS data reflect the survey responses of only certain individuals in the department, who may not always have sufficient knowledge about daily practices to provide reliable data. Qualitative studies could potentially provide greater insight into what policing looks like at the street level.

Qualitative or mixed methods approaches could help better assess to what extent departments are actually engaging in hot spots policing.

Related Entries

- ▶ Crime Mapping
- ▶ Effectiveness of Situational Crime Prevention
- ▶ Place-Based Randomized Trials
- ▶ Problem-Oriented Policing
- ▶ Procedural Justice, Legitimacy, and Policing

Recommended Reading and References

Braga AA (2007) The effects of hot spots policing on crime. Campbell collaboration systematic review final report. http://campbellcollaboration.org/lib/download/118/. Accessed 15 Apr 2012

Braga AA, Bond BJ (2008) Policing crime and disorder hot spots: a randomized controlled trial. Criminology 46:577–608

Braga AA, Weisburd D (2010) Policing problem places: crime hot spots and effective prevention. Oxford University Press, New York

Braga AA, Weisburd DL, Waring EJ, Mazerolle LG, Spelman W, Gajewski F (1999) Problem-oriented policing in violent crime places: a randomized controlled experiment. Criminology 37:541–580

Braga AA, Papachristos AV, Hureau DM (in press) The effects of hot spots policing on crime: An updated systematic review and meta-analysis. Justice Quart

Clarke RV, Weisburd D (1994) Diffusion of crime control benefits: observations on the reverse of displacement. In: Clarke RV (ed) Crime prevention studies, vol 2. Criminal Justice Press, Monsey, pp 165–184

Groff ER, Weisburd D, Yang S-M (2010) Is it important to examine crime trends at a local "micro" level?: A longitudinal analysis of street to street variability in crime trajectories. J Quant Criminol 26:7–32

Horney J, Osgood DW, Marshall IH (1995) Criminal careers in the short-term: intra-individual variability in crime and its relation to local life circumstances. Am Sociol Rev 60:655–673

Kelling GL, Pate AM, Dieckman D, Brown C (1974) The Kansas city preventive patrol experiment: technical report. Police Foundation, Washington, DC

Kochel TR (2011) Constructing hot spots policing: unexamined consequences for disadvantaged populations and for police legitimacy. Crim Just Pol Rev 22:350–374

Koper CS (1995) Just enough police presence: reducing crime and disorderly behavior by optimizing patrol time in crime hot spots. Just Quart 12:649–672

Koper CS (2008) The varieties and effectiveness of 'hot spots' policing: results from a national survey. Presentation at the annual meeting of the American Society of Criminology, St. Louis

Mazerolle LG, Price JF, Roehl J (2000) Civil remedies and drug control: a randomized trial in Oakland, California. Eval Rev 24:212–241

National Research Council (2004) Fairness and effectiveness in policing: the evidence. Committee to review research on police policy and practices. In: Skogan W, Frydl K (eds) Committee on law and justice, division of behavioral and social sciences and education. National Academies Press, Washington, DC

Ratcliffe J, Taniguchi T, Groff ER, Wood JD (2011) The Philadelphia foot patrol experiment: a randomized controlled trial of police patrol effectiveness in violent crime hotspots. Criminology 49:795–831

Reaves BA (2010) Local police departments, 2007. Bureau of Justice Statistics, U.S. Department of Justice, Washington, DC

Rosenbaum DP (2006) The limits of hot spots policing. In: Weisburd D, Braga AA (eds) Police innovation: contrasting perspectives. Cambridge University Press, New York, pp 245–263

Sherman LW (1995) Hot spots of crime and criminal careers of places. In: Eck JE, Weisburd D (eds) Crime and place, crime prevention studies, vol 4. Criminal Justice Press, Monsey, pp 35–52

Sherman LW, Rogan DP (1995) Deterrent effects of police raids on crack houses: a randomized controlled experiment. Just Quart 12:755–782

Sherman LW, Weisburd D (1995) General deterrent effects of police patrol in crime "hot spots": a randomized, controlled trial. Just Quart 12:625–648

Sherman LW, Gartin PR, Buerger ME (1989) Hot spots of predatory crime: routine activities and the criminology of place. Criminology 27:27–56

Spelman W (1995) Criminal careers of public places. In: Eck JE, Weisburd D (eds) Crime and place, crime prevention studies, vol 4. Willow Tree Press, Monsey

Taylor B, Koper CS, Woods DJ (2011) A randomized controlled trial of different policing strategies at hot spots of violent crime. J Exp Criminol 7:149–181

Telep CW, Mitchell R,Weisburd D (in press). How much time should the police spend at crime hot spots? Answers from a police agency directed randomized field trial in Sacramento, California, Justice Quart

Tyler TR (1990) Why people obey the law: procedural justice, legitimacy, and compliance. Yale University Press, New Haven

Weisburd D (2008) Place-based policing, Ideas in American policing. Police Foundation, Washington, DC

Weisburd D, Eck JE (2004) What can the police do to reduce crime, disorder, and fear? Ann Am Acad Pol Soc Sci 593:42–65

Weisburd D, Green L (1995) Policing drug hot spots: the Jersey City drug market analysis experiment. Just Quart 12:711–736

Weisburd D, Lum C (2005) The diffusion of computerized crime mapping policing: linking research and practice. Police Prac Res Int J 6:419–434

Weisburd D, Mazerolle LG (2000) Crime and disorder in drug hot spots: implications for theory and practice in policing. Police Quart 3:331–349

Weisburd D, Bushway S, Lum C, Yang S-M (2004) Trajectories of crime at places: a longitudinal study of street segments in the city of Seattle. Criminology 42:283–321

Weisburd D, Wyckoff LA, Ready J, Eck JE, Hinkle JC, Gajewski F (2006) Does crime just move around the corner? A controlled study of spatial displacement and diffusion of crime control benefits. Criminology 44:549–592

Weisburd D, Hinkle JC, Famega C, Ready J (2011) The possible "backfire" effects of hot spots policing: an experimental assessment of impacts on legitimacy, fear and collective efficacy. J Exp Criminol 7:297–320

Weisburd D, Groff ER, Yang S-M (2012) The criminology of place: street segments and our understanding of the crime problem. Oxford University Press, New York

Hot Targets

▶ Prediction and Crime Clusters

Housing Choice Vouchers and Escaping Neighborhood Crime

Ronald E. Wilson and Brent D. Mast
Office of Policy Development and Research,
U.S. Department of Housing and Urban
Development, Washington, DC, USA

Overview

The Housing Choice Voucher Program (HCVP) provides relocation opportunities for low-income families to find better places to live. In examining why HCVP, participants want to relocate research generally focuses on the "pull" factors of why low-income families want to relocate. Most reasons for relocation are related to some benefit the new neighborhood offers, such as access to better schools, jobs, housing stock, and other amenities to improve their lives. Often overlooked are the "push" factors that drive residents from a neighborhood. Much research on these push factors has focused on life-course changes in family structure, career options, or some financial gain. One significant push factor that uniformly effects residential satisfaction is how safe families feel with respect to witnessing or being victims of crime in their neighborhood. Families who use a voucher to relocate are no exception to this concept and have stated across numerous studies that the primary reason for wanting to leave their neighborhood is to escape chronic problems with crime. Numerous studies and reports have documented low levels of residential satisfaction by families in impoverished neighborhoods, primarily driven by violence, drugs, and gangs in the neighborhood. Research across several disciplines is consolidated in this entry about why families in impoverished neighborhoods want to leave their current residences for new neighborhoods, many of which use vouchers to escape crime.

Introduction

Many factors known to cause delinquency, social exclusion, antisocial behavior, violence, poor health, and low academic performance in youth come from living in distressed places (Buck 2001; Overman 2002; Friedrichs and Blasius 2003; Musterd et al. 2003; Sharkey 2008). Project-based public housing communities are usually located in places that exhibit a confluence of these factors. Families in these neighborhoods often have little means to escape and must withstand regular exposure to these factors reducing their neighborhood satisfaction in a multitude of ways. While there are family and peer factors that contribute to the prevalence of these factors, simply relocating out of these places can help diminish the negative effects of living in distressed places.

In 1974, the US Department of Housing and Urban Development (HUD) launched the Housing Choice Voucher Program (HCVP) to financially assist low-income families to relocate to neighborhoods of their choice. In 2010 alone, approximately 2.1 million families received assistance through the program (U.S. Department of Housing & Urban Development 2010). The HCVP is a "tenant-based" residential assistance program resulting from the Community Development Act (CDA). Other HUD assistance programs are "project based" and require participants to move to campuses, high-rise complexes, or other centrally located housing. Participants in the HCVP use a voucher to seek rental housing anywhere within the local housing market and pay at least 30 % of their own income toward rent and utilities; the remainder is paid by HUD through a local housing authority. HCVP participants can come from any neighborhood, not just project-based sites. Prior to the HCVP, residential choice for assisted families was limited to a small number of centrally located project-based sites.

Success of the HCVP is contingent on participants' searching for and finding housing that satisfies program requirements as well as their own preferences. The program's main intent is to provide access to safe and sanitary housing that must pass housing quality inspections. But HCVP descriptions commonly emphasize neighborhood as much as dwelling. This language implies a dual objective of finding both a dwelling and a neighborhood. With the HCVP, choice of where to live becomes more than just about the quality of housing stock and is inexplicably tied to the neighborhood the dwelling is located within. A key outcome of the HCVP design is that households have a maximized spatial mobility choice in selecting where to live (Winnick 1995). Place, then, is a central factor in the voucher recipient decision-making process, and (Ross 2011) found evidence that HCVP tenants were more likely to look at neighborhoods before housing units.

The growth and popularity of the HCVP is fueled by its distinct advantage in allowing a participant's locational needs and preferences to become the driving factor in choosing a place to live. As such, low-income families have an impetus for applying to the program to escape the often impoverished places they currently live, particularly when crime is a constant threat to their safety and their children.

Establishing the Idea of Place for Understanding the HCVP

Place is a central concept to understanding the HCVP. In housing studies, the term "place" often refers to project-based sites. Other social science disciplines also tend to geographically limit the term "place" to an address or nothing larger than a street block, which in geography is usually associated with the concept of "location." When limited to such small units, the concept of place is constrained to understanding why individuals, families, and households select certain environments to live and leave based on what occurs in that household or on that street. Decisions to move from environmental conditions are seldom limited to that scale. In geography, place is a scalable term that represents a delineated area that is distinguishable from other areas. Under this definition, a "place" could be a room, building, street block, neighborhood, section of a city/county, metropolitan area, or even a region. Place, whatever the scale, can be used as a framework that is inclusive of a location and its surrounding environment in which to examine individual, peer, and family interaction within a larger geography as part of a complete context.

Locations are focal points within a larger geography that situates *where* something is positioned or an event occurs. Locations exhibit specific characteristics that make them relatively fixed entities, represented accurately with an X and Y coordinate pair. Environments exert influence on the characteristics of that location contributing to its condition. Change the environment and the characteristics of a location will likely change. Changing a location's characteristics can also cause the surrounding environment to also change. As such, using locations as unit of analysis can constrain a more complete measurement of individuals, families, or households interactions. A "place" is representative of the scale that captures the influence of conditions and events occurring between a collection of locations in their larger environment.

Neighborhoods form an appropriate definition of place in studies of urban crime as opposed to locations. Neighborhoods matter as "places" that frame the interactions between individuals, peer, and families while accounting for the impact the environment has on them and them on it. Most crime clusters within a larger area than at a single address or set of addresses. Limiting place to such a small area, then, reduces the opportunity to understand the full impact of the neighborhood crime on residential satisfaction. People situate their locations within the larger neighborhood they live in and reference the features, events, and changes that occur within that space on how they feel. As such, separating the location from the neighborhood does not permit the ability to fully measure the factors that makes a resident satisfied with the neighborhood, i.e., the research cannot disentangle internal characteristics from external influences.

Crime is a deeply personal experience for victims. The traumatization associated with crime can stem from both personal victimization but also with place. Residents who witness or hear about crime in their neighborhood are also affected through the transmission of conditions and other events that occur in the neighborhood. Whether victim, witness, or simply "hearing" about crime in the neighborhood negative feelings about place can lead to behavior modification – of which leaving the neighborhood is one – based on anticipated fear of being victimized. This fear does not manifest itself from just incidents in buildings, addresses, street blocks, or a public housing sites, but emerges as product of the location and the larger environment. As such, location will not suffice in understanding crime in place for those using a voucher to relocate. In this entry, the term "place" refers to the dwelling and the neighborhood unless otherwise specified.

Wanting to Leave the Neighborhood for a Better Place

There are many reasons families relocate. Most reasons are brought on by life-course turning points in family status that entails changes in job opportunities, school quality, housing needs, or desirable neighborhood amenities. But decisions to move can result from factors related to satisfaction with neighborhood conditions (Wolpert 1965; Rossi 1980; Galster 1987; Lu 1999; Basolo and Strong 2002). For low-income families, neighborhood conditions may be more important than life-course events when deciding to apply to the HCVP to relocate. Participants applying to the HCVP program are looking to more often escape negative "life-altering" events that result from living in neighborhoods characterized as socially disorganized (Shaw and McKay 1942).

Social disorganization is the result of concentrated economic disadvantage and social isolation in place. These places are bleak geographies that exhibit higher levels of residential turnover, unemployment, single-headed households, uneducated populace, low quality businesses, poverty, and poor quality housing stock. Social networks and local organizational ties in these places are weak because of collapsed informal neighborhood relationships that leave residents unable to actualize common values – maintenance of a healthy, safe, and attractive residential environment – that contribute to higher levels of crime (Wilson, and Kelling 1982; Sampson and Groves 1989; Bursik and Grasmick 1993; Krivo and Peterson 1996; Sampson et al. 2002; Fagan and Davies 2004; Weisburd et al. 2004). Many HCVP applicants live in these types of neighborhoods.

Across the 1960s, 1970s, and 1980s, the quality of life in many public housing communities severely declined in the form of concentrated poverty (Banks and Banks 2004), resulting in socially disorganized places. Residents of these communities repeatedly cited crime as one of their greatest concerns among numerous neighborhood problems. Crime concerns grew so prominent that in 1989, Congress created the National Commission on Severely Distressed Public Housing (NCSDPH) to identify the national extent of the problem. This effort was the result of an earlier HUD study (Jones et al. 1979) that identified a set of indicators to measure the severity of social problems residents in public housing reported. The Commission used four broad areas to measure distress levels in public housing. Violent crime was one of those areas and was the first direct national-level acknowledgement that crime was a critical component of social and economic disadvantage in specific places.

Following the Commission's report, Lawrence Vale, a prominent scholar on housing issues, wrote a critical piece (1995) on the NCSDPH final report with several discussion points about crime and place. Vale made a key point about the spatial connection public housing developments share with their surrounding environment, noting they were at risk of becoming future problems partly due to their neighborhoods they were situated within. Responding to Vale's commentary, Jeffery Lines (1995), a consultant on the NCSDPH report, acknowledged crime was one of the most pervasive problems that made public housing residents perceive their neighborhood was severely distressed. These negative feelings toward the neighborhood appeared to be the cornerstone of residential dissatisfaction by most public housing residents and not just within their housing project. Vale's main argument was that a focus on too small of geographies limited the understanding, measurement, and interventions of crime in public housing. Both Vale and Lines stated that if analyses were limited to such small geographies, the likelihood of revitalizing public housing would not be fruitful because the problem must be solved inclusive of the surrounding neighborhoods.

In 1998, the Bureau of Justice Statistics (BJS) produced a report that provided more evidence that neighborhood crime was one of the most important aspects to residents with respect to neighborhood satisfaction. Using six waves of American Housing Survey (AHS) data between 1985 and 1995, BJS statisticians examined resident perceptions of neighborhood crime (DeFrances and Smith 1998). Across all six waves, 25 % of black and 13 % of white public housing respondents cited crime as a problem in the neighborhood where they lived. For overall households, 5 % white and 14 % black reported crime to be a problem. Of the public housing respondents, 18 % (24 % black and 13 % white) perceived crime so "objectionable" that it caused them to consider moving. The authors also used the 1998 National Crime Victimization Survey (NCVS) and found that even though crime victimization was the same in public housing as it was outside of public housing (23.8 % vs. 23.4 % respectively), the percentage of residents who expressed the desire to relocate was substantially higher for those in public housing communities. The analysis also revealed that, comparable to crime, black residents had equal concerns about other neighborhood problems, such as disliking neighbors, too many people, and loitering. Central city white households 15 % reported these "people" problems as a neighborhood problem whereas 13 % reported crime as the most pressing problem. While not being able to determine if these reasons were related to concerns of safety, other researches have noted these minor incivilities contribute to perceptions of safety in the neighborhood.

In 1999, Dugan used four waves of the National Crime Victim Survey (NCVS) between 1986 and 1990 to examine if victimization prompted neighborhood residents to relocate. Dugan found a 12 % increase in the probability a resident will move after the first victimization is experienced or known. Equally important was the finding that knowledge of criminal offenses outside the neighborhood had no effect on prompting residents to move, indicating the immediate neighborhood affected perceptions of safety that would prompt a consideration to move.

In examining relocation preferences from the Chicago Area Survey Project, Harris (2001) measured black and white preferences in neighborhood relocation and found crime and disorder to be of most concern to neighborhood choices. Specifically, the results showed that residents' perceptions of "big" crime problems lead to a significant decrease (.40 and .65 standard deviation unit reduction) in residential satisfaction for both whites and blacks, respectively. School quality and neighborhood deterioration were also found to be significant determinants of resident satisfaction.

Basolo and Strong (2002) surveyed residents of low-income neighborhoods in New Orleans as part of a revitalization initiative evaluation and found crime to be the greatest concern for potential home buyers. The authors examined questions that determine both neighborhood and housing satisfaction along several dimensions. Respondents stated their decision to purchase a house in a neighborhood was based on housing condition but safety was the strongest factor, the absence of crime was the most important factor toward being satisfied with the surrounding neighborhood. The authors found that 40 % of respondents who were renters cited poor housing conditions and crime as the primary reasons they would not buy a house in a neighborhood if they had the opportunity.

Hipp (2009) used the American Housing Survey data between 1987 and 1999 in 24 metropolitan areas to identify the determinants of neighborhood satisfaction. He extended the earlier BJS study by conducting a multilevel analysis for individual and neighborhood effects and found that perceived disorder and other social disorganization factors were related to increases in neighborhood dissatisfaction. Overwhelming, perceived crime was an accelerant to being dissatisfied with a neighborhood. A key aspect of Hipp's study was that he linked the elements of social disorganization with the desirability of neighborhoods.

In 1994 and 1995, several thousand residents in Chicago were surveyed for Project Human Development in Chicago Neighborhood (PHDCN) to examine how families, schools, and neighborhoods affected child and adolescent development (Earls et al. 1995). Numerous survey questions revolved around residential satisfaction, many questions related to crime. One question asked respondents to indicate the reasons they might move. Sixty four percent of respondents listed living in a safer neighborhood as a reason they might move, 62 % cited crime, and 60 % cited drug problems. Wanting to get their children away from bad influences (50.4 %) and lower housing/rental costs (49.9 %) were cited as the next top reasons. All other reasons amounted to less than 35 %. These results indicate crime is a driving factor in residential satisfaction and wanting to relocate.

Surveys of HCVP participants also showed crime was a major determinant of tenant satisfaction. Buron and Patrabansh (2008) examined data from an HCVP customer satisfaction survey and found that respondents rated crime as the most important quality of life factor. Fifty percent of respondents who rated their neighborhood as low quality cited crime as "a big problem," while other neighborhood conditions were of much less concern. Residents who rated their neighborhood of higher quality also cited crimes as their greatest concern. The correlation between crime and neighborhood rating was −0.62, indicating that perceived neighborhood quality declined with increased perceptions of crime problems. The Voucher Homeownership Assessment (VHA) program also showed that proximity to high-crime areas was among the two main negative features reported to the surveyors (Turnham et al. 2003).

Studies have shown that when given the opportunity to relocate, publically assisted families tend to move into neighborhoods with lower poverty and less crime (Austin et al. 2002; Fauth et al. 2004). Two particular housing initiatives examined the effects of low-income families in severely distressed public housing using HUD vouchers to relocate. The first was the *Gautreaux* program, a court ordered initiative to remedy racial discrimination in Chicago's public housing communities. The second was the *Moving to Opportunity* (MTO) experiment in five cities by the US Department of Housing and Urban Development. The intent of Gautreaux and MTO was to give public housing residents the option to move to new neighborhood of their choice for the purpose of improving access to opportunity. Participants in both programs stated that crime was one of the most prominent reasons they wanted to relocate (Popkin et al. 2003; Goering and Feins 2003). With MTO an average of 53 % of participants across all five cities cited getting away from drugs and gangs as the most important reason for wanting to move (Goering and Feins 2003). The second most important reason was to get their children into better schools (30 %). Follow-up research on Gautreaux and MTO consistently showed that participants saw improvements in the conditions of their neighborhoods with regard to health, economy, and safety

(Varady and Preiser 1998; Katz et al. 2001; Orr et al. 2003; Fauth et al. 2004; Kling et al. 2004; Gilderbloom et al. 2005; Keels et al. 2005; and Briggs et al. 2010). These studies indicated that even after relocation, crime was still a primary determinant of residential satisfaction for HCVP tenants. Long-term positive effects on the participants, however, have shown to be less than desired.

Varady et al. (2001) examined survey and neighborhood-level data across four US cities to determine if counseling led to more successful housing searches by HCVP participants. The authors found that perceived neighborhood safety was the strongest driver (coefficient for safety was 0.315) of housing satisfaction compared to 12 other factors. Distance also played a significant role, with participants wanting to remain in close proximity to their previous neighborhood while simultaneously wanting to move to a new place offering a greater sense of safety. Their findings were consistent with previous studies that found that HCVP participants tended to feel safer after using a voucher to relocate.

Conclusion

Questions about the impacts of the HCVP on neighborhoods are an emerging policy issue. Sadly, little attention is paid to the crime aspect of housing vouchers and voucher mobility even though the evidence presented in this entry shows it is a major factor in the decision to move. The findings across these studies clearly show that for low-income families, regardless of receipt of rental assistance, crime is a central determinant of participant residential satisfaction.

Fear of crime is often a driving force behind residential dissatisfaction and the need for safety (Gibbs and Hanrahan 1993) that manifests itself in behavior modification. Studies from criminology have shown that the anxiety generated from anticipated victimization prompts neighborhood residents to take protective actions, such as altering daily routines to minimize chances of being victimized, fortifying their homes (Baumer 1980), limiting children's activities (Kling et al. 2004), avoidance of walking outside (Roman and Chalfin 2008), and moving to a safer place (Skogan 1990; Rountree and Land 1996; Xie and McDowall 2008; Hipp et al. 2009). These behavior modifications clearly connect to the neighborhood and not just the location of their dwelling or street.

The combination of immediate location and the surrounding area is what constitutes place utility, upon which people make decisions to stay or move based on residential satisfaction (Wolpert 1965). Thus, using "place" offers an advantage over the separate use of "location" or "neighborhood" toward understanding of crime and place. The objectives of the HCVP emphasize the place utility concept and exemplify how the program is codependent on both people and place, with residential satisfaction as the link. Success for the HCVP participant is having a better place to live and work and for their children to grow, less fearful of crime. Success for the neighborhood is having residents that contribute to its well-being so that it is a desirable geography for people to live, work, attend school, socialize, and visit. Without place satisfaction, a resident may not contribute to the neighborhood and the neighborhood may not contribute to the resident's well-being.

The housing literature clearly shows a nexus with the study of crime and place. The thoughts and language used by residents in housing studies reveal that residents see their homes situated within a larger environment that contributes to their sense of satisfaction and safety. The housing literature identifies and discusses many of the same factors and conditions of social disorganization (residential turnover, unemployment, single-headed households, uneducated populace, low quality businesses, poverty, and poor quality housing stock) in project- and tenant-based programs that are used in the study of crime and place in criminology. There is a constant reference to "in and around" in the housing literature and indicates that residents and scholars understand there is an inseparable link between a home and its neighborhood. Housing scholars also use variables that go beyond the demographic structure that environmental criminologists have discussed and used, such as building conditions and values, land use, family structure changes, or business types. These housing scholars, though, do not often include social disorganization factors relevant to the understanding of crime in place. Some criminologists have been using some of these variables to inform the understanding of crime and place (see the Geographic Criminology entry of this encyclopedia). Both criminology and housing scholars could benefit from drawing upon evidence from both disciplines – and in particular, the theories and evidence from geography – regarding the effect of place on safety perceptions and the effect of safety perceptions on choice of place.

Mayor Robert Sabonjian of Waukegan, IL (a town north of Chicago), summed up the common sentiment about Section 8 housing in a local newspaper that highlights the unconscious understanding in policy-making that place is a key aspect to the success of the HCVP. Mayor Sabonjian stated that the problems associated with public housing "are not going to go away by getting rid of one problem building – it's going to have to be a systemic answer across the city" (Moran 2012). This statement demonstrates the recognition that a wide range of geographies are crucial to any solution to assisted housing. Briggs et al. (2010) stated that the success of relocation programs "hinge" on holders having proximal access to jobs and resources, i.e., the geography of opportunity model (Galster and Killen 1995). Research findings have made it quite clear that residential satisfaction includes the neighborhood, although there remains a debate about what constitutes the geography of a neighborhood (Wilson 2012). The studies presented in this entry present evidence of the larger role of neighborhoods in residential place satisfaction. Residents of high-poverty neighborhoods and public housing communities are often stuck in places exhibiting classic indicators of social disorganization in which crime is an everyday part of life. The HCVP plays a crucial role in promoting access to places with greater chances for residential satisfaction and socioeconomic opportunity, which are inextricably linked to the absence of crime.

Acknowledgments We would like to thank John Markovic from the Community Oriented Policing Services Office (COPS) of the US Department of Justice, Paul Joice of the US Department of Housing & Urban Development, Mark Stallo of the Dallas Police Department, and Emma Wilson of Montgomery College for providing valuable comments toward improving the clarity of this entry. We would also like to thank Sue-Ming Yang for giving us the opportunity to include this work.

The views expressed in this entry are those of the authors, and do not necessarily represent the official positions or policies of the US Department of Housing and Urban Development or the US Government.

Related Entries

- ▶ Crime and the Racial Composition of Communities
- ▶ Criminal Careers of Places
- ▶ Desistance from Crime
- ▶ Disadvantage, Disorganization and Crime
- ▶ Disorder: Observational and Perceptual Measures
- ▶ Early Chicago School Theory
- ▶ History of Geographic Criminology Part I: Nineteenth Century
- ▶ Lifestyle Theory
- ▶ Longitudinal Crime Trends at Places
- ▶ Neighborhood Effects and Social Networks
- ▶ Public Housing and Crime Patterns
- ▶ Social Capital and Collective Efficacy
- ▶ Social Disorder and Physical Disorder at Places

Recommended Reading and References

Austin DM, Furr AL, Spine M (2002) The effects of neighborhood conditions on perceptions of safety. J Crim Justice 30:417–427

Banks JG, Banks P (2004) Chapter 9, The decline of public housing. In: The unintended consequences: family and community, the victims of isolated poverty. University Press of America, Lanham, pp 63–74

Basolo V, Strong D (2002) Understanding the neighborhood: from residents; perceptions and needs to action. Hous Policy Debate 13(1):83–105

Baumer TL (1980) Personal protective behaviors and the threat of crime: a comparative analysis. Final report, U.S. Department of Justice, National Institute of Justice, Washington, DC

Briggs X, Comey J, Weismann G (2010) Struggling to stay out of high-poverty neighborhoods: housing choice and locations in moving to Opportunity's first decade. Hous Policy Debate 20(3):383–427

Buck N (2001) Identifying neighbourhood effects on social exclusion. Urban Stud 38(12):2251–2275

Buron L, Patrabansh S (2008) Are census variables highly correlated with housing choice voucher holder's perceptions of the quality of their neighborhoods? Cityscape: J Policy Dev Res 10(1):157–183

Bursik RJ, Grasmick HG (1993) Neighborhoods and crime: the dimensions of effective community control. Lexington Books, New York

DeFrances CJ, Smith SK (1998) Perceptions of neighborhood crime, 1995. Special Report. Bureau of Justice Statistics, Washington, DC

Dugan L (1999) The effect of criminal victimization on a household's moving decision. Criminology 37(4):903–929

Earls FJ, Brooks-Gunn J, Raudenbush SW, Sampson RJ (1995) Project on human development in Chicago neighborhoods: community survey, 1994–1995. ICPSR-2766: individual-level data codebook. Inter-university Consortium for Political and Social Research [Producer and Distributor], Ann Arbor, 2007-10-29. doi:10.3886/ICPSR02766.v3

Fagan J, Davies G (2004) The natural history of neighborhood violence. J Contem Crim Justice 20(2):127–147

Fauth RC, Leventhal T, Brooks-Gunn J (2004) Short-term effects of moving from public housing in poor middle-class neighborhoods on low-income, minority adults' outcomes. Soc Sci Med 59:2271–2284

Friedrichs J, Blasius J (2003) Social norms in distressed neighbourhoods: testing the Wilson hypothesis. Hous Stud 18(6):807–826

Galster G (1987) Identifying the correlates of dwelling satisfaction: an empirical critique. Environ Behav 19(5):539–568

Galster GC, Killen SP (1995) The geography of metropolitan opportunity: a reconnaissance and conceptual framework. Hous Policy Debate 6(1):7–43

Gibbs JJ, Hanrahan KJ (1993) Safety demand and supply: an alternative to fear of crime. Justice Q 10(3):369–394

Gilderbloom J, Brazley MD, Pan Z (2005) HOPE VI: a study of housing and neighborhood satisfaction. J Environ Sustain 11:1–26

Goering J, Feins JD (2003) Choosing a better life: evaluating the moving to opportunity social experiment. The Urban Institute Press, Washington, DC

Harris D. R. (2001) Why are whites and blacks averse to black neighbors? Soc Sci Res 30:100–116

Hipp JR (2009) Specifying the determinants of neighborhood satisfaction: a robust assessment in 24 metropolitan areas over four time points. Soc Forces 88(1):395–424

Hipp JR, Tita G, Greenbaum RT (2009) Drive-bys and trade-ups: the impact of crime on residential mobility patterns in Los Angeles. Soc Forces 87:1777–1812

Jones R, Kaminsky D, Roanhouse M (1979) Problems affecting low rent public housing projects: a field study. Final report, U.S. Department of Housing and Urban Development, Washington, DC

Katz LF, Kling JR, Liebman JB (2001) Moving to opportunity in Boston: early results of a randomized mobility experiment. Q J Econ 116:607–654

Keels M, Duncan GJ, Deluca S, Mendenhall R, Rosenbaum J (2005) Fifteen years later: Can residential mobility programs provide a long-term escape from neighborhood segregation, crime, and poverty? Demography 42(1):51–73

Kling JR, Leibman JB, Katz L, Sanbonmatsu L (2004) Moving to opportunity and tranquility neighborhood effects on adult self sufficiency and health from a randomized housing voucher experiment. Working paper #481, Industrial Relations Section, Princeton University

Krivo LJ, Peterson RD (1996) Extremely disadvantaged neighborhoods and urban crime. Soc Forces 75:619–650

Lines JK (1995) Comment on Lawrence J. Vale's "Beyond the problem projects paradigm: defining and revitalizing 'severely distressed' public housing. Hous Policy Debate 4(2):183–198

Lu M (1999) Determinants of residential satisfaction: ordered logit vs. regression models. Growth Chang 30:264–287

Moran J (2012) Waukegan Officials Rap Section 8 Housing Units. Lake County News-Sun. Chicago-Sun Times Publishers, January. http://newssun.suntimes.com/news/9813284-418/waukegan-officials-rap-section-8-housing-units.html

Musterd S, Ostendorf W, de Vos S (2003) Neighbourhood effects and social mobility: a longitudinal analysis. Hous Stud 18(6):877–892

Orr L, Feins JD, Jacob R, Beecroft E, Sanbonmatsu L, Katz LF, Liebman JB, Kling JR (2003) Moving to opportunity for fair housing demonstration program: interim impacts evaluation. Moving to opportunity interim impacts evaluation. U.S. Department of Housing and Urban Development by Abt Associates, Inc., and the National Bureau of Economic Research. U.S., Washington, DC

Overman HG (2002) Neighbourhood effects in large and small neighbourhoods. Urban Stud 39(1):117–130

Popkin SJ, Harris LE, Cunningham MK (2003) Families in transition: a qualitative analysis of the MTO experience. Final report, U.S. Department of Housing and Urban Development by Abt Associates, Inc., and the National Bureau of Economic Research. U.S. Department of Housing and Urban Development, Washington, DC

Roman CG, Chalfin A (2008) Fear of walking outdoors: a multilevel ecological analysis of crime and disorder. Am J Prev Med 34(4):306–312

Rosenbaum J (2001) Moving and changing: how places change people who move into them. Working paper, Institute for Policy Research, Northwestern University

Ross LM (2011) The impact of housing vouchers on renters neighborhood satisfaction – understanding the

perceptions and constraints among assisted and unassisted renters. Selected paper for the American housing survey user conference, Washington, DC, March
Rossi PH (1980) Why families move, 2nd edn. Sage, Beverly Hills
Rountree PW, Land KC (1996) Perceived risk versus fear of crime: empirical evidence of conceptually distinct reactions in survey data. Soc Forces 74(4):1353–1376
Sampson R, Groves WB (1989) Community structure and crime: testing social disorganization theory. Am J Sociol 94:774–802
Sampson RJ, Morenoff JD, Gannon-Rowley T (2002) Assessing neighborhood effects: social processes and new directions in research. Annu Rev Sociol 28:443–478
Sharkey P (2008) The intergenerational transmission of context. Am J Sociol 113(4):931–969
Shaw C, McKay HD (1942) Juvenile delinquency in urban areas. University of Chicago Press, Chicago
Skogan WG (1990) Disorder and decline: crime and the spiral of decay in American neighborhoods. Free Press, New York
Turnham J, Michlin N, Locke G, Wood M, Baker M (2003) The voucher homeownership program assessment, vol I. Final report, U.S. Department of Housing and Urban Development by Abt Associates, Inc., and the National Bureau of Economic Research. U.S. Department of Housing and Urban Development, Washington, DC
U.S. Department of Housing & Urban Development (2010) Public and Indian housing tenant-based rental assistance: 2012 summary statement and initiatives. http://portal.hud.gov/hudportal/documents/huddoc?id=Transform_R_Assis_2012.pdf
Vale LJ (1995) Beyond the problem projects paradigm: defining and revitalizing "severely distressed" public housing. Hous Policy Debate 4(2):147–174
Varady DP, Preiser WFE (1998) Scattered-site public housing and housing satisfaction. J Am Plann Assoc 64(2):189–207
Varady DP, Walker CC, Wang X (2001) Voucher recipient achievement of improved housing conditions in the US: do moving distance and relocation services matter? Urban Stud 38(8):1273–1304
Weisburd DL, Bushway S, Lum C, Yang S-M (2004) Trajectories of crime at places: a longitudinal study of street segments in the city of Seattle. Criminology 42(2):283–321
Wilson RE (2012) Point of contention: defining neighborhoods. Cityscape: J Policy Dev Res 14(2):219
Wilson JQ, Kelling GL (1982) Broken windows: the police and neighborhood safety. Atl Mon 249:29–38
Winnick L (1995) The triumph of housing allowance programs: how a fundamental policy conflict was resolved. Cityscape: J Policy Dev Res 1(3):95–121
Wolpert J (1965) Behavioral aspects of the decision to migrate. Pap Proc Reg Sci Assoc 15:159–169
Xie M, McDowall D (2008) Escaping crime: the effects of direct and indirect victimization on moving. Criminology 46(4):809–839

How to Make Crime Mapping More Inferential

▶ Information Technology and Police Work

Human Immunodeficiency Syndrome

▶ Drug Abuse and Alcohol Dependence Among Inmates

Human Rights

▶ Environmental and Human Rights

Human Rights Violations in Criminal Court

Joachim J. Savelsberg
Department of Sociology, University of Minnesota, Minneapolis, MN, USA

Synonyms

Crimes against Humanity; Genocide; International Criminal Court; International Tribunals; War Crimes

Overview

Grave violations of human rights have become increasingly criminalized in recent decades. Chances that perpetrators will be charged in criminal courts, including international criminal courts, have grown. These developments culminated in the creation of the first permanent international criminal court, the International Criminal Court in 2002. This entry first provides a summary of the history of criminal justice intervention against perpetrators of human rights crimes. A brief discussion follows concerning the societal conditions that advanced this trend. A further section examines challenges of criminal justice intervention against human rights violators and issues of debate: the building of functioning international criminal courts, their institutional incentives and constraints, the scope of their jurisdiction, the question of individual versus collective liability, tension between political reason versus legal procedure, and finally actual chances to effectively reduce grave human rights violations. Selective examples from a growing and at times impressive new body of literature are cited.

Historic Roots and Fundamentals

Human rights, those civil, political, social and economic rights that are granted humans independently of their citizenship, are codified in the Universal Declaration of Human Rights (UDHR). The United Nations General Assembly passed this declaration on December 10th, 1948, in response to the atrocities committed by Nazi Germany. Human rights are associated with the right and obligation of the international community to intervene against offenses by national governments. They thus weaken the principle of national sovereignty, established by the Peace Treaty of Westphalia that ended the Thirty Years War in 1648. While most violations of human rights do not constitute crimes and while the principle of national sovereignty is still prominent, considerable change has occurred. It has been argued that many previous centuries can compete with the 20th with regard to the level of human rights abuses, but that the twentieth century is distinct as serious attempts have been undertaken to intervene in atrocities, to slow or prevent them (Minow 1998). Some even diagnose a "justice cascade" (Sikkink 2011).

The process began during the late nineteenth century when humanitarian law, or the law of wars, was first codified in a series of international conventions, known as the Geneva and Hague Conventions. The conventions were drafted by multinational conferences that initially sought to establish rules regarding the treatment of wounded soldiers and of prisoners of war. In 1949, rules were added against the deportation of individuals or groups, the taking of hostages, torture, collective punishment, "outrages upon personal dignity," the imposition of judicial sentences without due process, and discriminatory treatment. The Geneva Convention has garnered broad support across the world's nations as is indicated by the large number of 180 signatories.

While these conventions set binding international standards for situations of international conflict, they continue to respect national sovereignty. Yet, two subsequent Protocols of 1977 marked a weakening of sovereignty. Protocol I extended the protections of the Hague/Geneva Conventions to persons involved in wars of "self-determination," typically liberation wars which former colonies fought against colonial powers. Thus, violations of humanitarian principles could no longer be considered internal affairs of colonizers. Going one step further, Protocol II extended humanitarian protections to persons involved in severe civil conflicts, prohibiting collective punishment, torture, hostage taking, acts of terrorism, slavery, humiliating and degrading treatment, rape, and enforced prostitution. These Protocols have also found broad support, albeit weaker due to their more interventionist nature, with 150 and 145 signatories respectively.

Humanitarian law provided a foundation for judicial intervention. The International Military Tribunal in Nuremberg (IMT) against leading Nazis drew legitimacy from the conventions, as did the Tokyo trials against Japanese war criminals. Yet, the history of Nazism had shown that even the treatment of domestic populations should not remain unchecked by the international community. Unbearable in itself, mistreatment at home may prepare yet more horrendous offenses against foreign peoples. In Nazi Germany, for example, the gassing of mentally retarded German children and adults served as a training ground for the SS's work on the "final solution" against Europe's Jews and other groups

(Schmidt in Heberer and Matthäus 2008). The idea was thus born to extend rights that many countries guaranteed their own citizens (civil rights) to all humans (human rights), independently of their citizenship, and to certify these rights in a covenant, the UDHR, and a series of international treaties. Sovereign states would now be bound by a new and growing system of international law that was not limited to times of armed conflict.

In reaction to the Nazi crimes, The Convention on the Prevention and Punishment of the Crime of Genocide came into force on January 12th, 1951. Article 2 of the Convention defines genocide as: "…any of the following acts, committed with the intent to destroy, in whole or in part, a national, ethnical, racial or religious group, such as: Killing members of the group; causing serious bodily or mental harm to members of the group; deliberately inflicting on the group conditions of life calculated to bring about its physical destruction in whole or in part; imposing measures intended to prevent births within the group; forcibly transferring children of the group to another group." Genocide, "whether committed in time of peace or in time of war, is a crime under international law which… [the contracting parties] undertake to prevent and to punish" (Article 1). Threatened with punishment are "constitutionally responsible rulers, public officials or private individuals" (Article 4).

Other conventions sought to protect the human rights of women (1979), children (1990) and indigenous peoples (1991), but only the Convention against Torture and Other Cruel, Inhumane and Degrading Treatment or Punishment (1987) applies standards of criminal liability, like the genocide convention and humanitarian law before it and the Rome Statute of 1998 that established the ICC.

Despite such progress, a recent review of the history of international human rights law and its enforcement cautions us. Enforcement of the growing number of international treatises tended to be weak at least through the 1980s (Sikkink 2011). Treaty bodies created by the UN General Assembly, such as the Human Rights Council, focus on states' legal accountability. They monitor violations and work toward solutions with accused governments, but with few enforcement powers. New regional courts, such as the European Court of Human Rights, the African Court of Human Rights, and the Inter-American Court of Human Rights also apply a state accountability model. Treaty Bodies and courts ask states to provide remedies when violations are recorded, including changes in policies and/or reparation payments to individuals victimized by past policies.

This situation partially changed during the 1990s. A new model of criminal liability began to supplement state accountability, even if preceded by the London Agreement through which the victorious powers of World War II had created the IMT. The IMT had marked the serious beginning of the construction of international criminal courts as central institutions in the fight against grave humanitarian law and human rights violations. The occupying powers had arrested and the IMT had charged the defendants with crimes against the peace (conducting illegal, aggressive war), war crimes and crimes against humanity. Yet, as the trial focused on offenses committed during the war it shied away from challenging the notion of national sovereignty. Pre-war offenses of the totalitarian Nazi regime against German citizens were not covered. Eventually, 12 defendants were sentenced to death and executed, 7 received prison terms, and 3 were acquitted (on related trials, including the famous "doctors'" and "lawyers' trials" see Heberer and Matthäus 2008).

After a long hiatus caused by the Cold War between the two super powers, only the breakdown of the Soviet Union allowed for new initiatives. Different from Nuremberg or Tokyo, these new courts did not grow out of military victory. Instead, the UN Security Council established both the ICTY in 1993 and the ICTR in 1994; and an international agreement, the Rome Statute of 1998 created the ICC, the first permanent international criminal court. The Rome Statute entered into force in 2002 when 60 countries had ratified. As of 2008 there were only 108 ratifications, with several powerful members of the international community opposed, including the US. The creation of new legal institutions corresponds with a quadrupling of trial activity, at the domestic and international levels, in only two decades from the early 1980s to the early part of the twenty-first century (Sikkink 2011). In addition to international courts, a diversity of other types are involved.

Hybrid courts, emerging from agreements between the UN and national governments, are staffed with groups of domestic and international judges applying domestic and international law. Recent examples include courts in Sierra Leone, East Timor, Kosovo, and Cambodia, the latter a much delayed response to the genocidal mass killings of the 1970s under the Khmer Rouge regime. Such trials, typically conducted in the countries where the crimes were perpetrated, allow for easier access to victims and witnesses than international courts, while international participation may reduce the risk of partisan abuses of trials as revenge mechanisms by post-transition regimes – or, in the alternative, obstruction by past perpetrators who managed to hold on to political power under the new regime. *Foreign courts* became famous through the Jerusalem trial against Adolf Eichmann in 1961. More recently, under the new doctrine of universal jurisdiction, a Spanish judge charged Chilean General Augusto Pinochet and (unsuccessfully) requested his extradition from the UK

Cases tried in foreign courts remain rare though, while *domestic courts* are a prominent part of international justice when they apply international humanitarian and human rights law, often in combination with domestic law. Importantly, domestic courts now operate under the shadow of the ICC. Its mere existence is likely to encourage domestic enforcement, as most countries prefer cases to be handled at the domestic level (Sikkink 2011; Roht-Arriaza 2005). Crucial here is the doctrine of complementarity that governs the ICC. The court can only take up cases if domestic courts are unable or unwilling to do so. Domestic courts, however, tend to try top leaders only after regime changes (e.g., Argentina, Chile, Iraq), while low-level defendants are targeted in cases of regime continuity. Examples are the cases of My Lai and the abuse and torture committed in the Abu Ghraib or Guantánamo Bay prisons. In addition, penalties in these cases tend to be either mild or greatly reduced after initial sentencing.

In short, the growing openness toward international intervention into domestic affairs comprises three processes:

- A universalization or globalization of human rights law, its institutionalization in international doctrines and organizations, including courts of law and its application by many nations in domestic courts
- An individualization of human rights law, allowing for grievances to be directed not just against governments or entire countries, but also against individuals acting on behalf of these countries (and filed by individuals) and
- A partial criminalization of human rights law, supplementing compensatory mechanisms with penal responses

Societal and Global Conditions of Change

Recent research explores structural and cultural conditions for these trends (Savelsberg 2010). On the cultural side scholars such as Emile Durkheim and Erving Goffman, in their classic works, alert us to the status of an individual's dignity in modern society. While it sets limits to the intensity and types of punishment, it simultaneously encourages the punishment of killers, including dictators who radically disregard the "sacred" status of individuals in modern society. Such reactions are intensified by the growing sensitivity to physical violence that accompanied the pacification of domestic life to which Norbert Elias's classical arguments about the civilizing process alert us.

In light of such cultural trends, the outrages of the Holocaust further advanced global consensus regarding the dignity of individuals. Through symbolic extension of the Shoah and psychological identification with the victims, members of a world audience became traumatized by an experience that they themselves had not shared

(Alexander in Alexander et al. 2004). In these terms, the punishment of leading Nazi perpetrators by the IMT and by subsequent trials was performative. It provided, consistent with a semiotic model of social life, images, symbols, totems, myths, stories, and it thus contributed to the formation of a collective memory of evil to which we shall return below.

Once it was established as universal evil, the Holocaust served "analogical bridging" to reinterpret later events in light of this earlier trauma (Alexander in Alexander et al. 2004: 245–249). Examples are the treatment of minorities in the United States or the victimization of millions in the Balkan wars during the 1990s. In the latter case, analogical bridging occurred famously through the image of an emaciated Bosnian concentration camp inmate behind barbed wire, published on the front pages of most international newspapers and magazines. Thus building a bridge from the Holocaust to the cruelties committed in Bosnia advanced diplomatic and military intervention and the establishment of the ICTY, with great potential at contributing to new international criminal law (Hagan 2003).

Structural changes also contribute to the growth of international criminal justice against human rights perpetrators. Especially shifts in the balance of power at the international and national levels opened the way to the prosecution of national leaders. Globalization with its internationalization of economic ties and modern technologies of communication and transportation created a new dependency of nation states, and their leaders, on the international community. This shift is supported by the establishment of international governmental organizations (IO) such as the UN and its many sub-organizations and by the multiplication of International Non-Governmental Organizations (INGOs), representing a form of civil society at the global level. Five times as many organizations worked on human rights in the 1990s, than in the 1950s. Specifically, Transnational Advocacy Networks (TANs) prove effective, especially when they raise issues involving bodily harm to vulnerable populations; when responsibility can be attached to specific actors; when networks are dense, involving many actors and providing reliable information flow; and finally when target actors show material or normative vulnerabilities (Keck and Sikkink 1998).

While structural changes in the balance of power have opened up opportunities for international criminal justice intervention, continuing imbalances of power still leave their mark. Powerful countries such as China, Russia and the United States, and their rulers, have been spared criminal justice intervention when they offended against human rights law. Further, for the most part, only *formerly* powerful leaders have been indicted, and their numbers are few (e.g., Charles Taylor and Saddam Hussein; but see Slobodan Milošević and Omar al-Bashir). Mostly, when current officials are prosecuted, typically in situations of regime stability, they tend to be low ranking officers, internationally and nationally.

Challenges and Debates

Criminal justice responses to grave human rights violations face several challenges and controversies, some of which are specific to international interventions while others apply at all levels. A short section is devoted to each of these themes.

Functioning of Courts, Forms of Capital, and Competition: The Case of the ICTY

Formal Security Council decisions or international treaties are only first and necessary steps toward creating international courts. Their actual functioning depends on multiple social actors and forces. Sociologist-criminologist John Hagan (2003) identifies, for the ICTY, a multitude of innovative actors in the judicial field, all with different strengths (or forms of capital) and exposure to various national legal traditions. These actors compete against and cooperate with others inside the court and in the world of diplomacy and international politics to make the court succeed. Two brief examples must suffice.

After the UN Security Council had established the ICTY, it appointed Richard Goldstone as chief prosecutor, a South African judge with impeccable human rights credentials but little criminal law experience. Goldstone first used his international contacts and continued media presence, to secure a $30 Million budget for the court. Later he gained, through his diplomatic contacts, access to CIA aerial images of mass graves around Srebrenica, allowing ICTY investigators to advance their massive exhumation project. Still, after 2 years and 70 indictments, only 6 suspects were in custody, and the court resorted to legally problematic hearings against defendants who were not in the custody of the court ("in absentia hearings"). Change came only after a new prosecutor with different forms of capital took over from Goldstone. Canadian jurist Louise Arbour's substantial expertise in criminal law then became crucial in transforming the "virtual tribunal" with its questionable "in absentia" hearings into a "real time tribunal" (Hagan 2003: 93–131). Arbour linked traditional tools of criminal law with the UN mandate by introducing sealed (secret) indictments and surprise arrests through NATO military forces. These strategies resulted in a substantial increase in the number of defendants in custody. In short, Security Council resolutions alone do not determine the success of a newly founded international criminal court. Innovative strategies, involving cooperating and competing legal actors, but also actors from the worlds of diplomacy and military, from national governments, IGOs and NGOs, unfold in the face of uncertain outcomes before a new type of international criminal legal practice can be established.

Incentives to Intervene: The Case of the ICC

The organization of a court and the institutional rules under which it operates determine the incentives to aggressively engage in the prosecution of grave human rights violations. The ICC, for example, determined by the Rome Statute, consists of several courts or "Chambers" with a total of 18 judges, each with non-renewable 9-year terms (see Schabas 2007). Trial and appeals chambers write opinions and thus specify future international criminal law. The prosecutor is also selected for one non-renewable 9-year term through an anonymous vote of the member states. Cases can be referred to the ICC prosecutor by individual citizens of member states, by states against one of their citizens, and by the UN Security Council. The latter, for example, referred the Darfur case to the ICC. In contrast, state parties referred cases against Uganda, the Central African Republic and the Democratic Republic of Congo to the ICC.

In light of these and additional conditions, the court will not be as activist and punitive as American criminal courts have been in recent decades. First, due to the principle of complementarity the ICC can only get involved with Security Council referral or if domestic courts are unable or unwilling to prosecute. Second, given the still relatively weak institutionalization of civil society at the world level, public moral outrage is less likely to promote prosecution than at the national level. Third, prosecutors and judges, holding tenured and non-renewable positions, are less likely to be responsive to moral outrage. Yet, public mobilization may advance Security Council actions as it did in the case of Darfur. Fourth, given the strong role of nation states among the court's constituents, the ICC's legal and procedural principles will often compete with diplomatic or military actors and outcome-oriented reasoning. Finally, massive power-asymmetries between states are likely to constrain the court's agenda. The United States, for example, has entered "Bilateral Immunity Agreements" with some 100 governments that agree not to extradite American citizens to the ICC, often in exchange for international aid.

Scope of Jurisdiction: More on the ICC

The ICC's jurisdiction includes only four crimes: genocide, crimes against humanity, aggression, and war crimes, committed after April 2002 in states that have ratified the Rome Statute. Some actors would like to substantially expand jurisdiction of the ICC or other criminal courts to include violations of all rights guaranteed by the UDHR. Scholar-activists have suggested, for example, that child poverty in a wealthy country could be conceived of as an offense against the Convention for the Rights of the Child. This should, in the eyes of some, result in sanctions against responsible states and in the expansion of individual criminal liability to those whose policies had advanced these violations of international human rights standards.

Such far-reaching demands are being challenged by others, even those who in principle support expanded international criminal law (Hagan and Levi 2007). They argue that that tort law may in many cases be at least as effective as criminal law, with lower burdens of proof; charging countries and their leaders in criminal court may isolate these countries from the international community, polarizing conflict, and resulting in a loss of international influence; and that criminal law is ill-suited to address larger structural and cultural forces that contribute to broader human rights violations. Finally, human rights problems such as large-scale homelessness among children may result from national policies enacted by legitimate governments and backed by majorities of the electorate. Who then is to be charged?

Individual Versus Collective Liability

Grave human rights violations, while always involving individual agency, are typically embedded in complex organizational settings, government agencies, police and military units, and actors at different levels of hierarchy. These complexities pose challenges to prosecution. In response, criminal law has developed the concept of conspiracy, rooted in American jurisprudence. The crime of conspiracy is defined as an agreement between two or more individuals, entered into for the purpose of committing an unlawful act. It is justified as a separate crime as it is conceived of as a threat to the public in itself. Conspiracy also encourages and eases additional criminal behavior. Law seeks to deter (by increasing the cost of membership) and to destabilize (by undermining trust among co-conspirators) (see Meierhenrich 2006). “Conspiracy” makes its way into international criminal law under different names.

Article 9 of the London Charter, for example, makes membership in a “criminal organization” punishable. The American chief prosecutor, Justice Robert Jackson, argued that criminal organizations will serve as carriers of criminal plans from one generation to the next if not delegitimized by criminal law. Later, at the ICTY, crimes that must be expected in the context of a “joint criminal enterprise” constitute individual criminal liability even if charged individuals are not themselves engaged in the execution of the crime. For instance, involvement in the planning of ethnic cleansing campaigns to be conducted by armed groups should result in charges of homicide and war crimes as such campaigns must be expected to result in the killing of civilians.

And yet, “conspiracy,” “criminal organizations,” and “joint criminal enterprise,” innovative attempts at resolving the legal dilemma of individual guilt in the context of collective action, face grave challenges. They raise the specter of guilt by association, thereby endangering the *nullum crimen sine lege* principle, and they thus create doubts regarding the rigor and impartiality of international criminal law.

An alternative option of collective punishment, following the US Sentencing Guidelines’ model of Corporate Probation, would face challenges as well. Court orders are hard to enforce if a hostile regime refuses to obey, a problem long faced by UN Treaty Bodies. Further, in cases of grave atrocities, a negotiated settlement with the perpetrating collectivities such as the Iraqi Republican Guards or the Nazi SS, is often inconceivable and may undermine the legitimacy of the court. Dissolution of corporate entities and punishment of their individual members are typically called for. Collective punishment also disregards considerable shades of culpability. It may open collectivities up to the lure of oppositional cultures, if it is directed against large populations or entire countries. Post-WWI German history taught lessons, and paths chosen after the victory over Nazi Germany took these lessons seriously.

Formal Procedure Versus Political Reasoning

Given the high stakes of trials against serious violations of human rights, practical consequences of a court’s rulings are considerable. Justice Jackson had argued that failure to legally denounce the organizations involved in the Nazi crimes would open the doors to a recurrence of massive genocide in the next generation, and that humankind would not survive a repeat of such horrors. And yet, some legal philosophers argue that formal principles, a "just desert" orientation and a focus on rights of the accused must trump policy in international criminal justice. Even the IMT has come under liberal attack for its breach of formal rules, despite general agreement regarding its relative success in producing justice. Landsman (2005), for example, critiques the IMT's breach of liberal legal principles, such as the admission of hearsay evidence. Such philosophical arguments are backed by the conviction that only strictly due process-oriented international criminal justice can earn legitimacy and that only a legitimate system can prevail in the long term.

Effectiveness of Criminal Court Intervention Against Human Rights Perpetrators

The effectiveness of criminal justice intervention in preventing grave human rights violations is hotly debated. Critics challenge, for example, the rise of universal jurisdiction, the power of domestic courts to try foreign citizens, summarized in the Princeton Principles of Universal Jurisdiction and justified by the recognition that human rights violations are offenses to all humanity. Domestic courts, they argue, may have little sense of the harm their prosecutions may cause in the affected foreign country. Amnesties, truth commissions and other transitional justice programs, and thus successful transitions to peace and democracy may be at risk.

The ICC and other international courts are further targets of critique. They are said to suppress the consideration of power, necessary to assess the consequences of intervention and to balance legal accountability with political costs. Critics argue that filing charges against Serb president Milošević by the prosecutors of the ICTY, for example, made it harder for NATO to reach a deal with Serbia at the time, thereby extending war and suffering in the Balkans in the summer of 1999. In 2011, critics challenged the ICC for its decision to charge Sudanese President al-Bashir with genocide, at a time where his role in stabilizing relations with the newly independent South Sudan may be crucial. In general, the concern is that perpetrators will not be willing to negotiate and cease power if they are threatened by criminal trials (see Snyder and Vinjamuri 2003/2004).

Challengers of these skeptics include political scientist Kathryn Sikkink (2011), who offers an impressive new data set with information on domestic truth commissions and domestic, foreign and international trials for a 26-year period (1979–2004), covering 192 countries and territories. Sikkink finds that transitional justice does not typically lead to the strengthening of old forces; the severity of offences and the likelihood of trials are highly correlated (decisions for trials are thus not made lightly); and, importantly, countries with more human rights trials show greater improvements of later human rights records. Such positive outcome was further improved where trials were coupled with truth commissions. More importantly, and contrary to skeptics' point of view, not a single case in Latin America shows that holding a trial contributed to violent conflict and dislodged the transition.

If the correlation between the use of transitional justice and improved human rights records shown by Sikkink is not just due to third factors (e.g., a past rule of law regime in the affected country), but if it does indeed constitute a causal relationship, the question remains how this relationship can be explained. Theoretical efforts toward understanding the link between trials and outcomes require us to open up the "black box" between intervention and human rights. Two processes may be at work. One is a deterrence mechanism, consistent with rational choice ideas; the other is a cultural mechanism, a thorough de-legitimization of gross human rights abuses, possibly to the point where they no longer appear as options in the decision trees of rational actors.

Rational choice ideas and deterrence research have a long tradition in criminology. One common conclusion of many studies is that the certainty of punishment is most likely to have a deterrent effect, more than its severity. White-collar crime literature suggests that deterrence is most likely to work for the powerful, as they are expected to act rationally. Sikkink's (2011) explanation of the positive correlation between transitional justice mechanisms and human rights records is consistent with this mode of thought: the next generation of military officers will remember the shaming their predecessors experienced as a result of criminal sanctions or truth commission reports and be reluctant to breach human rights. But what are the conditions under which past perpetration and subsequent sanctions are being remembered? Sophisticated rational choice arguments take learning about the past seriously, as such learning will affect what costs and benefits decision makers take into consideration.

Here a new line of academic work comes to play that examines the effects of trials and other mechanisms on collective memory (Osiel 1997; Savelsberg and King 2011). This work builds on classic sociological ideas and on arguments made by politicians and jurists such as President Franklin Roosevelt and Justice Robert Jackson. Judge Samuel Rosenman, Roosevelt's confidant reports about the president: "He was determined that the question of Hitler's guilt – and the guilt of his gangsters – must not be left open to future debate. The whole nauseating matter should be spread out on a permanent record under oath by witnesses and with all the written documents" (after Landsman 2005: 6). From a Durkheimian perspective, Osiel perceives the court room drama in cases about mass atrocities as a "'theatre of ideas,' where large questions of collective memory and even national identity are engaged" (Osiel 1997: 3).

A new idea is thus added to the traditional rationales for criminal trials and sanctions such as retribution, deterrence or incapacitation: a history writing function, the construction of a collective memory of past evil that, some argue, will reduce the likelihood of future offending. Collective memory is here understood as knowledge about the past that is shared, mutually acknowledged and reinforced by a collectivity. Its shape determines how a formerly charismatic leader, after the "degradation ceremony" of a criminal trial, will be remembered.

The hope invested in criminal trials' contribution to history writing and to the formation of collective memory, however, must be cautioned. Trials follow a particular logic. Evidentiary rules differ, for example, from those used by historians. Further, trials target individuals, not the social processes and cultural patterns sociologists might focus on when constructing the past. Actions trials address are further limited by legal classification systems; producers of inflammatory rhetoric may have played central roles, but they will typically not be criminally liable. Finally, following the binary logic of criminal law, the defendant is guilty or not guilty, a gross simplification by psychological standards.

A budding literature has thus begun to critically explore the consequences of trials for collective memory. Giesen (in Alexander et al. 2004), for example, argues that German criminal trials against former Nazis served a "decoupling" function. In light of such trials, the German people could take the position of the third party, while individual guilt was assigned to a few in the court of law. As individual perpetrators were ritually expelled, the majority of Germans were offered a chance to avoid acceptance of collective guilt. Osiel (1997) applies this insight to French history. President Charles De Gaulle urged that post-World War II trials were directed against a few elite actors of the Vichy regime that had collaborated with the Nazis. Here too decoupling succeeded. By attaching guilt to some individuals through legal rituals, memory could be cleansed of the collaboration of many, and attention could be redirected from questions about their past to the reconstruction of France following war and occupation.

The French case indicates further that trials after grave human rights violations, especially after regime change and military defeat, can only be understood in their political context. Just like the legal process was affected by political will in the French case, so were the Allied trials against German perpetrators (Landsman 2005: 111ff for the IMT). The Allies decided to focus on just a few major Nazi perpetrators, as prosecuting all who had become guilty might have destabilized the country and its budding democracy. This focus on few perpetrators, however, helped shape a view of history that saw the majority of Germans as victims of a small elite, thus repressing the memory of support and collaboration many had provided (Marrus in Heberer and Matthäus 2008).

Cases against high-level perpetrators have in common that trials follow conquest and/or regime change. Trials may construct very different collective memories though when past evil is processed in the context of regime continuity (Savelsberg and King 2011). The United States trial against perpetrators of My Lai, for example, where criminal guilt was attributed only to one low-ranked military, has contributed to the exculpation, not just of broader segments of the population, but also of political and military leadership in public memory. Research shows, for example, that most American history textbooks do not mention the My Lai case. Those that do, tend to present the massacre in line with the outcome of the trial, focusing on the one convict, Lt. Calley, but silencing the role played by higher ranks and their attempted cover up of the massacre. Such processing of past atrocities may have advanced uncritical attitudes toward the institution of the military and contributed to a willingness by American military in current conflicts to offend against norms of humanitarian law. The judicial handling of recent torture and illegal imprisonment cases is likely to further advance such attitudes.

Despite such misgivings, criminal justice mechanisms may still affect collective memory in ways that decreases the likelihood of future evil. First, the de-legitimization of past perpetrators may be crucial in transitions. Second, the selectivity of legal logic must not be overrated. After all, legal trials initiate the collection of evidence. While not all may be admitted in the court of law, such evidence nevertheless may be available to future historians, or it may be directly communicated to the public through mass media. Hagan (2003) documents the diversity of extra-legal expertise of forensic scientists, victim workers, journalists and social scientists, mobilized by the ICTY to uncover forensic and interview-based empirical evidence of the atrocities committed during the Balkan wars. News of recently opened mass graves and liberated concentration camps reached a broad public through journalistic reports, independent of the success of translating these materials into evidence in the court's proceedings. Investigatory evidence may also be used in future historical documentations, independently of its legal status at the trial (Bass 2000).

In short, through deterrence and collective memory functions, criminal trials may help transitions and prevent repetitions of past evil. Optimism, however, must be tempered by insights into the selectivity and inaccuracy of trial-based memories, by the focus on "small fish" in the absence of regime transitions and by transition problems that trials may cause in some contexts.

Alternatives, Supplements and Precursors to Criminal Justice Intervention

In light of the shortcomings of criminal courts, a balance of principles may be more realistically achieved through a mix of institutions, including UN Treaty bodies, creative new models, truth commission, reparation programs, vetting proceedings, apologies, commemorations and memorials, or amnesties. A few comments on four of these must suffice.

First, while the movement toward criminal justice and other mechanisms has partly been powered by a profound sense of frustration regarding the weak enforcement of human rights through UN Treaty Bodies such as the UN Human Rights Council and traditional international courts, the latter institutions still constitute one crucial component in the mix of mechanisms through which the international community responds to human rights violations.

Second, international courts, beyond criminal courts, include the International Court of Justice (ICJ), also called the "world court," and regional courts. The ICJ addresses disputes between states, and it provides the UN with advisory opinions. Regional courts such as the European Court of Human Rights (ECHR) and the Inter-American Court of Human Rights have somewhat stronger enforcement mechanisms.

Their judgments may demand reparation and force policy changes in member countries. Judgments also become binding law to nation-level courts.

Third, innovative experimentation with traditional justice mechanisms holds some promise. Reponses to the Rwanda genocide included a new institution, called *inkiko gacaca* (*gacaca*). The *gacaca*, while inspired by traditional justice forms, is a modern system of some 10,000 community-based judicial bodies, oriented toward retributive and restorative justice, and administered by the Rwandan state. By establishing the *gacaca*, the state responded to a desperate situation in which up to 120,000 detainees waited for their trials in overcrowded cells in this extremely poor country of just six million people (see Meyerstein 2007).

Finally, truth commissions (TC) have become important supplements, predecessors, or alternatives to criminal courts. They are bodies that focus on the past, investigate long-lasting patterns of abuses, are constituted for limited periods of time and conclude their work with a report. They are officially sanctioned and authorized by the state (Hayner 2001). The name "TC," of course, is often misleading, as the truth is frequently well known, while its acknowledgement is at stake. The Argentinean TC, for example, the National Commission on the Disappeared was created in 1983 per decree by President Raúl Alfonsin after 7 years of military dictatorship, during which tens of thousands endured arrest and torture and "were disappeared." Eventually, the commission turned its files over to the prosecutor's office and thus provided critical evidence for the cases against senior members of the military Junta. Importantly, TCs may contribute to accountability in ways not available to criminal courts. Instead of attributing responsibility to particular individuals alone, they examine broader patterns of abuse, thereby encouraging institutional reforms. They may thus also challenge broad sectors of society and segments of the population that carry some degree of responsibility, from bureaucrats to torturers and profiteers all the way to by-standers who refused to speak up.

Concluding Comment

Domestic, hybrid, foreign and international courts have increasingly applied international human rights law, especially since the end of the Cold War. International courts or tribunals, once created, require resourceful actors with different types of capital, to successfully fight challengers and to actually get a functioning court off the ground. The specific institutional forms courts take have massive implications for their functioning. Establishing and operating courts involved in human rights issues, even more than other courts, faces diverse fields of tension: between a formal-procedural versus a pragmatic outcome orientation, restricted versus expanded jurisdiction, and individual versus collective accountability. Criminal courts often face considerable challenges as human rights perpetration typically occurs in complex organizational and political contexts. Yet, recent research indicates the potential of positive outcomes. Both a deterrence function and the establishment of collective memories that delegitimize past inhumane practices appear to be at work. Finally, human rights courts are supplemented by other institutions such as lustration or truth commissions, at times thought of as alternatives, but often crucial supplements of criminal justice intervention.

Recommended Reading and References

Alexander J et al (2004) Cultural trauma and collective identity. University of California Press, Berkeley

Bass GJ (2000) Stay the hand of vengeance: the politics of war crimes tribunals. Princeton University Press, Princeton

Hagan J (2003) Justice in the Balkans. University of Chicago Press, Chicago

Hagan J, Levi R (2007) Justiciability as field effect: when sociology meets human rights. Sociol Forum 22(3):372–84

Hayner PB (2001) Unspeakable truths: confronting state terror and atrocities. Routledge, New York

Heberer P, Matthäus J (eds) (2008) Atrocities on trial. University of Nebraska Press, Lincoln

Keck ME, Sikkink K (1998) Activists without borders. Cornell, Ithaca

Landsman S (2005) Crimes of the holocaust: the law confronts hard cases. University of Pennsylvania Press, Philadelphia

Meierhenrich J (2006) Conspiracy in international law. Annu Rev Law Soc Sci 2:341–57

Meyerstein A (2007) Between law and culture: Rwanda's Gacaca and postcolonial legality. Law & Social Inquiry 32(2):467–508

Minow M (1998) Between vengeance and forgiveness: facing history after genocide and mass violence. Beacon, Boston

Osiel MJ (1997) Mass atrocities, collective memory, and the law. Transaction Publishers, New Brunswick

Roht-Arriaza N (2005) The Pinochet effect: transnational justice in the age of human rights. University of Pennsylvania Press, Philadelphia

Savelsberg JJ (2010) Crime and human rights: criminology of genocide and atrocities. Sage, London

Savelsberg JJ, King RD (2011) American memories: atrocities and the law. Russell Sage, New York

Schabas WA (2007) An introduction to the international criminal court. Cambridge University Press, Cambridge

Sikkink K (2011) The justice cascade: how human rights prosecutions change world politics. Norton, New York

Snyder J, Vinjamuri L (2003/2004) Trials and errors: principle and pragmatism in strategies of international justice. Int Security 28:5–44

Human Smuggling

Joanne van der Leun[1] and Richard Staring[2]
[1]Department of Criminal Law and Criminology, Faculty of Law, Leiden University, Leiden, The Netherlands
[2]Department of Criminology, Faculty of Law, Erasmus University, Rotterdam, The Netherlands

Overview

The illegal movements of migrants across national borders have become a global issue. Human smuggling is defined as those activities that facilitate illicit crossing of national borders with consent of the individuals smuggled. It is therefore closely connected to international migration as such. The present entry discusses definitions as well as background factors and basic characteristics of the process of smuggling over time. Characteristics of human smugglers and those of smuggled migrants are presented. In addition, the ways in which states respond to the challenges of detecting and dealing with human smuggling are dealt with. The entry closes with a brief discussion of controversies and future directions including the paradox of combating human smuggling and tensions between migration control and crime policies.

Introduction: Human Smuggling and Its Definition

Human smuggling is currently often presented as one of the most profitable criminal activities worldwide. A formerly relatively minor transborder activity affecting few neighboring countries has become a global activity (Saricca 2005:8). Tragedies like Dover in 2001 and Morecambe Bay in 2004 every now and then throw a spotlight on the largely hidden human smuggling trade. This trade takes place between China and Europe, the Horn of Africa and South Africa, and Asia and Australia, only to name a few routes. Media coverage of incidents is high, and thanks to popular movies, "snakehead" has even become a familiar word in many countries.

Movement of people across borders is far from new, but it is only in the 1990s that human smuggling, human trafficking, organized crime, and illegal migration came to attract attention of policy makers and researchers alike (Salt 2000). In this first phase, scientific and policy debates were severely hampered by lack of clarity in definitions and by conceptual confusion. Over the years consensus over definitions seems to come nearer, but the field remains complex and politicized. Empirical research is almost inherently difficult and many publications are written primarily from the viewpoint of law enforcement.

Definitions

Human smuggling involves, put simply, those activities that facilitate illicit crossing of national borders with consent of the individuals smuggled, and it is therefore closely connected to migrants and international migration (Staring et al. 2005; Van Liempt and Doomernik 2006). The United Nations Protocol against the Smuggling of Migrants by Land, Sea, and Air, which supplements the Palermo Protocol, accounts for consent in its definition of human smuggling, by clarifying that the “Smuggling of migrants shall mean the procurement, in order to obtain, directly or indirectly, a financial or other material benefit, of the illegal entry of a person into a State Party of which the person is not a national or a permanent resident” (2011a:1). The protocol relates smuggling of migrants explicitly to the aim of obtaining financial benefits of the illegal entry of a person into a state (UNODC 2011a), thereby excluding helping a person to cross borders for humanitarian reasons.

Human smuggling and migration are often confused with trafficking in human beings (THB). These concepts have the following in common: they involve the movement of persons and they are all at least commonly associated with illegality, vulnerability, and exploitation. Yet, there are important differences between them. Crossing borders is not a necessary condition in the case of human trafficking (see chapter on THB by Smit and Van der Laan), where by contrast exploitation and lack of consent are crucial. Although the legal concepts are clearly different, there are also similarities and interconnections between human smuggling and human trafficking in practice, and one form can evolve into another (Aronowitz 2001). One could state that whereas human smuggling is perceived primarily in the context of international migration and organized crime, trafficking is, by and large, seen as an issue of (organized) crime as well as of human rights (with victims whose fundamental rights are violated). The idea that smuggled immigrants made a choice out of free will conflicts with notions of victimhood. Therefore, some researchers have made a case for a broader more sociological outlook on the phenomenon (Staring 2001; Van Liempt and Doomernik 2006). Migration scholars also stress that legal and illegal entry modes and flows of people are inextricably related. When facing restrictions, prospective immigrants employ an array of side doors and back doors in order to reach their destination, including the help of human smugglers (Castles and Miller 2009). In many instances, this does not have to involve criminal organizations.

Migrants who have been smuggled often become illegal immigrants once they enter the destination country. Several authors have criticized the use of the terms “illegal migrants” because of its connotations with criminality and called for replacing it by the term irregular migrants (for a discussion, see Koser 2005). Because in the end, all terms used may acquire similar connotations, the term “illegal” is used here despite its disadvantages (cf. Ngai 2003). It has the advantage that it refers to the way in which migrants relate to the construction of what is legal (Samers 2001).

Dimensions of illegal migration are entry, stay, work and, in a country like China, also exit (Van der Leun 2003). Exact regulations differ across jurisdictions. In case of human smuggling, the dimension of entry is crucial. Yet, what states consider as legitimate (legal) might not fully coincide with what individuals consider as such (Abraham and Van Schendel 2005). Transnational movements of people can be labeled illegal because they defy formal norms and authority, while in the participants’ view they can be acceptable, even licit.

Background Factors and Basic Characteristics

Global inequalities and integrated world markets shape push and pull factors for international mobility of people and, as such, also fuel processes of human smuggling. A number of factors can be identified in order to explain contemporary human smuggling, which in turn facilitates illegal migration. The globalization of communication means and transport routes are important ones: countries that have been once secluded are gradually becoming part of a global market and infrastructure. In addition, the availability of information and all forms of ties between individuals and communities brings societies closer together. Networks of immigrants and immigrant communities also play a crucial role in this respect (Massey et al. 1993). Restrictions on immigration, and sometimes on emigration, are meant to channel this mobility, but restrictions to some extent also create illegality and crime. At the same time, unregulated migration can involve social, financial, and political costs for the individuals, society, and governments alike.

Estimates of numbers of people and money involved have been made. At the end of the 1990s, IOM, for instance, estimated that four million people were smuggled across borders annually worldwide (IOM 1997). More recent UN assessments have suggested that total revenues of 10 billion dollars are involved, but these estimates are severely hampered by a lack of reliable data, conflicting perceptions and interests (Kyle and Koslowski 2011). Therefore, figures on human smuggling should be interpreted with care.

While some decades ago these forms of migration law violations and migration crime appeared to be phenomenon of the wealthier countries of the global north and west, they are now globally prevalent. Smuggling patterns have become more complex and in some cases also more dangerous.

Being smuggled is not without risks. Especially those migrants who are being smuggled by sea face high risks. Pugh (2004: 56) found that up to 3,000 people drowned in the Morocco-Spain crossings during the 1990s. More recently, in 2011, around 1,500 people are estimated to have drowned in the Mediterranean Sea only in an attempt to reach Europe (Strik 2012). Moving to the border between Mexico and the United States, it is acknowledged that due to the increased militarized security measures at the border, fewer migrants cross illegally. Rose poses that this militarization of the Mexican-American border leads smugglers and migrants to choose more dangerous and desolated routes through the desert. As a result, the number of migrants dying in these illegal crossings remains stable the last 10 years (Rose 2012). Grassroots organizations such as the Human Rights Coalition estimate the number of victims in the southern border regions of the United States among the smuggled migrants around 200 yearly (http://derechoshumanosaz.net). Apart from those who end up paying with their lives, smuggling can also be accompanied with other risks including violence.

Smuggling routes are highly diverse and constantly changing. At a global scale, Zhang (2007) has discerned four major smuggling routes to high-income regions and countries like Western Europe, Australia, and the United States: (1) from African countries through Spain and Italy to parts of Western Europe; (2) from Central Asia through countries like Kyrgyzstan, Turkmenistan, and Uzbekistan to Russia and then further to Western Europe through the Czech Republic, Slovakia, and Ukraine; (3) from the Middle East and Asia (China) to Australia and New Zealand; and finally (4) from Southern American countries or islands in the Caribbean via the Mexican-American border into the United States. The UNODC (2011a) classifies this perspective on human smuggling toward high-income countries as a rather traditional, western-centric vision, denying the dynamics and complex nature of people smuggling nowadays. Human smuggling networks do not only adapt routes and modus operandi in response to law enforcement, but smugglers have also developed routes toward West African countries, South Africa, and countries in the Middle East that also have economies that attract people from abroad (UNODC 2011a: 34–37).

Human Smugglers

Research has made it evidently clear that many illegal immigrants do not enter European countries through the help of human smuggling organizations but rather enter legally. These immigrants travel regularly and safely with tourist or business visa, guaranteed by the helping hands of relatives and friends living in the countries of destination. Only when they overstay their visa, they become illegal immigrants. Others were smuggled in by family or friends on their way back home from a holiday or visit in their home country. These transnational kinship-based networks can explain the arrival and further incorporation of illegal immigrants even in times of economic recession. This pattern coincides with what is referred to in the literature as chain migration, although in an irregular form (cf. MacDonald and MacDonald 1974). In these cases, financial gains are largely absent and a relative or friend of the illegal immigrant is often the human smuggler (Staring et al. 2005). Other prospective immigrants employ local smugglers or middleman during their travels from country to country without valid travel documents. They cross borders illegally by foot and by local means of transportation, or they pay local smugglers for difficult border crossings only. In the case of, for instance, the irregular trans-Saharan migration to North Africa and Europe, De Haas (2006) describes how local smugglers – often nomads or former immigrants – cooperate with local corrupt officials and intermediaries who have links to European employers. Finally, there are those international human smuggling networks that to a large extent dominate the media coverage and popular image as well as the political debate on combating irregular migration (Staring 2004; cf. Kleemans and van Brienen 2002): highly organized large-scale criminal organizations.

Several authors argue that the characterization of human smuggling in terms of international organized crime does not fit reality in most cases. Pastore et al. (2006) have demonstrated for Italy that the extent of organization is hugely overrated in popular belief. On the basis of a review of court cases, they argued that smuggling organizations in Italy are often little more than loose networks linking largely independent clusters of practical competencies, without a clear chain of command. Moving from Europe to the USA, Zhang (2008), a prominent expert on Chinese smuggling networks, stresses the fact that "snakeheads" should be perceived as opportunistic entrepreneurs who provide specific services in cooperation with other smugglers. They form flexible temporary business alliances rather than function in a stable organized setting. In line with these observations within Europe, Zhang's study based on fieldwork among Chinese snakeheads has smashed the popular image of snakeheads as members of well-organized mafia-like gangs. In the Netherlands, the Research and Documentation Centre of the Ministry of Justice and Safety continuously monitors organized crime by a two-yearly systematic analysis of closed investigations of criminal groups that are active in the Netherlands. In one of these monitors, the authors also concluded that human smuggling organizations show a huge diversity in terms or organizational structures (Kleemans and Van Brienen 2001). Other researchers who analyzed closed Dutch human smuggling investigations in the domain of organized crime stated that the majority of these human smuggling organizations showed more similarities with the image of small-scale informal entrepreneurship where economic motives intermingle with kinship-based loyalties. Large-scale, violent, and professionally organized human smuggling networks seem to be an exception rather than the rule (UNODC 2011a: 81).

Authorities and law enforcement agencies often claim that human smuggling is related to other forms of (organized) crime as document fraud and/or forgery, human trafficking, labor exploitation, money laundering, and terrorism. Among researchers this link between human smugglers and activities of organized crime groups is more controversial (UNODC 2011a). There is much more agreement on the existence of a shadow economy – especially in transit countries – surrounding human smuggling in which divergent facilitators operate. These include, for instance, forgers, burglars looking for identity documents, hairdressers, and beauty specialists for look-alikes; human smugglers sometimes do need local carriers and people safeguarding and taking care for the smuggled migrants in safe houses (Staring et al. 2005). Both informal economic activities and criminal activities take place within these shadow economies.

Smuggled Migrants

Migrants typically travel from relatively poor to richer areas, but the poorest people do not usually have the financial and social capital to move across borders. Those who are on the move are pushed by necessity and/or pulled by the prospect of a brighter future abroad *and* have the means to take action. As described before, many, probably most, illegal immigrants enter their country of destination in a legal manner and without the help of smugglers. Aronowitz (2001: 167) states that those "who fall prey to smugglers and traffickers are usually those most disadvantaged in their own countries: those with poor job skills or little chance of successful employment at home. They are often women and children." In other words, a selective segment of the population of illegal immigrants makes use of the services of smugglers. Whereas the majority of the illegal immigrants employ human smugglers out of a lack of other opportunities, there are also those immigrants who chose for the paid services of human smugglers in order to stay away from moral claims of relatives and friends. There is also evidence that significant proportions of asylum seekers in the United Kingdom and elsewhere in Western Europe have also been smuggled, thereby blurring the boundaries between asylum migration and illegal migration (Koser 2005).

On the whole, the literature currently available on smuggled migrants is scarce and has inherent methodological flaws. Moreover, it tends to focus on migrants from the perspective of the countries of destination and from the perspective of migration control. Qualitative sociological studies on living conditions of illegal immigrants in Europe and the USA (see, for instance, Düvell 2006; Mahler 1995; Staring 2001; Van der Leun 2003; Van Meeteren 2012) show that illegal immigrants make up a highly diverse category of people. Relatively young men between 20 and 40 years of age make up a large share, but the share of women appears to be growing. Social networks are pivotal for their opportunities. Without social resources, they risk marginalization. Deportation is a main fear and health problems including mental health problems are often found among smuggled migrants.

In literature which views smuggling of migrants primarily as a form of organized crime, migrants are often presented as "human cargo" or passive objects trapped and moved around by criminal networks. These studies also emphasize the risks involved, up to death on the way. Although this can be a realistic picture, as the figures mentioned before have illustrated, it is also biased. Studies departing from a more sociological viewpoint emphasize human agency and point to the selectivity of the other perspective. Violence and risks can indeed be involved, but human smuggling is in many cases also "just" a way of making migration possible when legal options fail. Many migrants, including many who have made use of the services of smugglers, are pursuing a project of bettering their lives. Finding a way of making a living is their top priority.

After arrival illegal migrants typically start looking for (informal) work. They are prepared to do the jobs that natives shun, many times under poor conditions, for wages usually lower than those of the legal residents (Calavita 2005: 15). Businesses in receiving countries are searching for cheap labor input, while private homeowners welcome the domestic workers that otherwise they might not afford. Certain economic sectors even only survive international completion by employing illegal workers (Van der Leun and Kloosterman 2006).

There are also instances in which exploitation takes place after arrival. Migrants who have initially consented with their smuggling have to pay high smuggling fees and later find out they have to work under exploitative conditions in order to pay the sum back. Migrants from China, for instance, appear to face a high risk of being exploited through these forms of debt bondage. These are the cases in which human smuggling can turn into human trafficking, which were briefly mentioned before.

State Policies

Tackling the smuggling of migrants requires a comprehensive approach, in which governments across the world invest in different ways, extents, and configurations. Since terrorist attacks in New York, Madrid, and many other places, the fight has become more intense. At the national level, border agencies such as customs, border police, and immigration services are responsible for the processing of people and goods at points of entry and exit as well as for the detection of human smuggling. Since the mid-1980s, EU and especially Schengen officials have been intensifying police and judicial cooperation to combat illegal or irregular migration in response to the diminishing of internal borders (Boswell 2003). This was followed by the externalization of migration control to countries outside the EU in the form of so-called pre-frontier control, or capacity building of migration management and asylum systems in transit countries, or the deployment of EU police to combat migrant trafficking in the western Balkans. In 1999 the European Council on Justice and Home Affairs has strengthened cooperation in the area of migration, asylum, and security. Five years later FRONTEX was established – the European Agency for the Management of Operational Cooperation at the External Borders of the Member States of the European Union. FRONTEX enables joint operations using member states' staff and equipment at external sea, land, and air borders.

At the international level, important steps have been taken since the late 1990s with the so-called Vienna process in response to concerns about the involvement of organized crime in international migration. This has led to the Protocol against the Smuggling of Migrants by Land, Sea, and Air, supplementing the United Nations Convention against Transnational Organized Crime (Smuggling of Migrants Protocol) which entered into force in 28 January 2004. In the whole multilevel playing field involved, a push toward a punitive approach is visible. The ultimate objective is dismantling organized crime networks that are active in migrant smuggling.

Controversies and Future Directions

Despite obvious successes, there is no convincing evidence that current policies to curb human smuggling are effective in doing what they should do (Kyle and Koslowski 2011). In fact, human smuggling is both adaptive and persistent (UNODC 2011b). Organized crime is not only feeding on and profiting by the phenomenon of illegal migration, but its activity also might be promoting it. This is partially the case because it does not influence root causes for international mobility. More effective long-term solutions to illegal migration and its facilitation would have to involve more legal opportunities for labor migration in receiving countries and diminishing economic differences between parts of the world, which have to be seen as long- and far-off-term aims. In the mean time governments still have to deal with international migration in many forms, including illicit forms.

Combating human smuggling will remain a challenging task, also in the future. Yet, law enforcement strategies and irregular migration policies are increasingly characterized by academics as not only ineffective but also inhumane because they lead to a race between law enforcement agencies, human smugglers, and the migrants. More militarization of and control at the borders will lead to new and more dangerous routes and as such will also contribute to an increasing number of victims and a lucrative market for organized crime. New technology might to some extent offer solutions, but it also will bring about new problems and adaptations. Engbersen and Broeders (2009) emphasize that the increasing digitalization of control within Europe has made the cat-and-mouse game between states and illegal migrants, pushing them further underground.

Many studies have demonstrated that rather than focussing on those behind human smuggling, many governments still, or even increasingly, see and treat smuggled individuals as criminals. In practice, people who are smuggled are often the ones to be arrested, detained, and expelled. This tendency appears to be strengthened by processes of crimmigration and securitization which puts the emphasis on "dangerous others" (Pugh 2004; Stumpf 2006). Yet, it should not be forgotten that protecting human rights is at the fundaments of the international protocols that underlie the fight against human smuggling. In line with international protocols, it is very clear that policies should focus on those behind human smuggling, rather than on those smuggled. Or as the recent UN Action Plan states: "Where migrants are simply detained and returned to countries of origin without investigating the actors involved in smuggling those migrants, the criminal processes at work continue unchallenged" (UNODC 2011b: 6).

Related Entries

- ▶ Forced Migration and Human Rights
- ▶ Human Trafficking
- ▶ Organized Crime, Types of

Recommended Reading and References

Abraham I, Van Schendel W (2005) Introduction: the making of illicitness. In: Van Schendel W, Abraham I (eds) Illicit flows and criminal things. States, borders and the other side of globalization. Indiana University Press, Bloomington, pp 1–37

Aronowitz A (2001) Smuggling and trafficking in human beings. The phenomenon, the markets that drive it on the organizations that promote it. Eur J Crim Pol Res 9:163–195

Boswell C (2003) The 'external dimension' of EU immigration and asylum policy. Int Affairs 79(3):619–638

Calavita K (2005) Immigrants at the margins. Law, race and exclusion in Southern Europe. Cambridge University Press, Cambridge

Castles S, Miller MJ (2009 [1993]) The age of migration: international population movements in the modern world. Macmillan, Basingstoke

De Haas, H. (2006). Trans-Saharan migration to North Africa and the EU: historical roots and current trends. Migration Information Source, 1 Nov 2006, available at: http://www.migrationinformation.org/feature/display.cfm?ID=484. Last accessed Aug 2012

Düvell F (2006) Illegal immigration in Europe. Beyond control? Palgrave Macmillan, London

Engbersen G, Broeders D (2009) The state versus the alien: immigration control and strategies of irregular immigrants. West Eur Polit 32(5):867–885

IOM (1997) Trafficking in migrants: policy and activities. International Organization for Migration, Geneva

Kleemans ER, Brienen MEI (2001) Van vriendendienst tot slangenkop: Een analyse van tien opsporingsonderzoeken naar mensensmokkel. Tijdschrift voor Criminologie 43(4):350–359

Koser K (2005) Irregular migration, state security and human security. Global Commission on International Migration, Geneva

Kyle D, Koslowski R (eds) (2011) Global human smuggling: comparative perspectives. John Hopkins University Press, Baltimore

MacDonald J, MacDonald L (1974) Chain migration ethnic neighborhood formation and social networks. Milbank Memorial Fund Quart 42(1):82–97

Mahler SJ (1995) American dreaming. Immigrant life on the margins. Princeton University Press, Princeton

Massey D-S, Arango J, Hugo G, Kouaouci A, Pelegrino A, Taylor J-E (1993) Theories of international migration: a review and appraisal. Popul Dev Rev 19(3):431–463

Ngai MM (2003) Impossible subjects: illegal aliens and the making of modern America. Princeton University Press, Princeton

Pastore F, Monzini P, Sciortino G (2006) Schengen's soft underbelly? Irregular migration and human smuggling across land and sea borders to Italy. Int Mig 4(4):95–119

Pugh M (2004) Drowning not waving. Boat people and humanitarianism at sea. J Refugee Stud 17(1):50–69

Rose A (2012) Death in the desert. New York Times, 21 June 2012

Salt J (2000) Trafficking and human smuggling: a European perspective. Int Mig 38(3):31–56

Samers M (2001) "Here to work": undocumented immigration in the United States and Europe'. SAIS Rev 21(1):131–145

Saricca F (2005) The smuggling of migrants. A flourishing activity of transnational organized crime. Crossroads 5(3):7–23

Staring R (2001) Reizen Onder Regie: Het Migratieproces van Illegale Turken in Nederland. Het Spinhuis, Amsterdam

Staring R (2004) Facilitating the arrival of illegal immigrants in the Netherlands. Irregular chain migration versus smuggling chains. J Int Mig Integr 5(3):273–294

Staring RHJM, Engbersen GBM, Moerland H, Lange NE de, Verburg DA, Vermeulen EH, Weltevrede A (2005) De sociale organisatie van mensensmokkel. Kerckebosch, Amsterdam

Strik T (2012) Lives lost in the Mediterranean Sea: who is responsible? Report Committee on Migration, Refugees and Displaced Persons. Strasbourg: Council of Europe

Stumpf J (2006) The crimmigration crisis: immigrants, crime and sovereign power. Am Univ Law Rev 56(2):367–419

UNODC (2011a) Smuggling of migrants a global review and annotated bibliography of recent publications. United Nations Office on Drugs and Crime, Vienna

UNODC (2011b) International framework for action to implement the smuggling of migrants protocol. United Nations Office on Drugs and Crime, Vienna

Van der Leun JP (2003) Looking for loopholes. Processes of incorporation of illegal immigrants in the Netherlands. Amsterdam University Press, Amsterdam

Van der Leun JP, Kloosterman RC (2006). Going underground. The labour market position of undocumented

immigrants in the Netherlands. Tijdschr Econ Soc Geogr 97(1):59–68
Van Liempt I, Doomernik J (2006) Migrant's agency in the smuggling process: the perspectives of smuggled migrants in the Netherlands. Int Mig 44(4):165–190
Van Meeteren M (2012) Transnational activities and aspirations of irregular migrants in Belgium and the Netherlands. Glob Netw 12(3):314–332
Zhang S (2007) Smuggling and trafficking in human beings: all roads lead to America. Praeger, Westport
Zhang S (2008) Chinese human smuggling organizations: families, social networks, and cultural imperatives. Stanford University Press, Stanford

Human Trafficking

► Trafficking in Human Beings

Hybrid Gangs

► Gangs and Social Networks

I

ICAP Theory

► Integrated Cognitive Antisocial Potential Theory

ICC

► Crime Investigations by the International Criminal Court

Identification and the Development of Forensic Science

Pierre Margot
School of Criminal Justice, Institute of Forensic Science, University of Lausanne, Lausanne, Switzerland

Overview

Forensic science as a science and profession takes its roots in the industrial revolution and the Age of Enlightenment. Measurement, classification, and databases started on habitual criminals or recidivists, with criminal anthropometry introduced in Paris by Bertillon, but this was soon followed with fingerprinting or dactyloscopy. The scientific exploitation of traces as silent witnesses of crimes or clues to their understanding extended the capabilities of forensic science throughout the twentieth century. The story of identification sees a major development in 1985, with the advent of the analysis of DNA markers which allow current individualization, through their extreme diversity. Identity has always been a fuzzy notion. It is the need to identify criminals and recidivists that led to the concept of identity and individuality that currently influences the structure of large databases such as fingerprint or DNA, but also most measurements in biometry as well as administrative identities (security documents, electronic identity, etc.). These developments arose at the end of the nineteenth century around pioneers. This entry goes through the early discussions and controversies that established forensic science as a scientific endeavor.

Definitions, History, Administrative Identity, and Biological Identity

Early man may have used identification in rituals and in relation to gods (hand impressions, finger impressions). The process of identification and individualization became essential as soon as communities grew to a size when direct knowledge and recognition of familiars was not possible anymore. There is little reference to identification of persons in the earlier texts (such as Hammurabi's code, art 18 about the identification of the owner of slaves), but the early history of man describes impostors or

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

characters impersonating others to rob, cheat, or gain undue advantages or a hiatus around the identity (in religious early manuscripts such as Esau in the Genesis) that make the essence of theatre plots and great tragedies (such as Oedipus). Persons in organized societies started to have rights (community, legal, monetary) and duties needing administrative or civil identification, and supports were soon introduced in material forms such as church registries (recording birth, death, marriages, filiations) and magistrate logs.

Within organized societies, migrations created the need for legislation, classification, and identities. The use of "identifiers" such as a written signature (in Europe) or an apposed fingerprint (in Asia) strengthened the binding of terms in documents, engagements, etc., whereas official documents "*sauf-conduits*" were given to officialize the identity of the bearer for a safe passage. This led to the first passports that, incidentally, gave little information about the identification of the bearer but carried the power of the authority given by seals such as royal or imperial heraldic seals. Passports became personal identity papers for early royal messengers of the French King Louis XI in 1464 – early postmen. Associated to the passport, a registry of passports allowed tracking every journey made by messengers and a way of controlling back the activity and the effectiveness of delivery. The development of surveillance has become a whole part of the identification process in modern days, often in contradiction with other rights (privacy, data protection, etc.).

One can distinguish two types of identities: civil or administrative (surname, name, sex, filiation, date of birth, address, profession, function, etc.) and biological. The first may be variable over time and may depend on the observer's knowledge making this identity relative, whereas the biological identity (genome, DNA, fingerprint, biometry) is usually very stable over time with notable exceptions (like the extraction of a tooth in odontology).

Any new development such as identity documents has intrinsic weaknesses that will be abused by government (spies, etc.) and criminals (forgers, counterfeiters, mafias, etc.). Groebner (2007) makes a fascinating account of the historical developments around the needs and reasons for identification in Europe from the Middle Ages to modern times and technological developments and their risks. The need for identification becomes a major issue in our electronic world with the apparent distance between a person (identified, pseudo) and her activities through the electronic medium. Identity theft is not a new crime, but it has taken dimensions never attained before the twenty-first century.

Other needs arise for the recognition and identification of deceased persons (Gremaud 2010), an act necessary for a mourning family (particularly important for victims of disasters – Disaster Victims Identification teams or DVI teams are often deployed), but also for administrative reasons (pensions, heritage, insurances, etc.) or to establish the criminal act of genocide, or stop a criminal investigation against a disappeared person.

The use of court experts and expertise is documented in Antiquity, but the needs for the identification of persons became a real issue with the comparatively recent demographic explosion (humankind reached one billion in 1860 and sees the seven billion mark will overtaken in 2013!). This was compounded by the disappearance of many forms of arbitrary justice (based on torture and various forms of godly intervention, etc.), and punishments based on mutilations (from severe cuttings of limbs or disfiguration to marking the skin with a hot stamp). The century or Age of Enlightenment in France (siècle des Lumières – Diderot (1713–1784), Montesquieu (1689–1755), and Voltaire (1694–1778)) led to major changes (American Declaration of Independence, US Bill of Rights, French Revolution, Universal Declaration of Human Rights, etc.) that affected the way justice was done. Punishment (except for the death penalty) became pecuniary or imprisonment in Europe and countries under its influence. This led to a situation where habitual criminals could not be identified as such, unless arrested by the same policeman who could identify a recidivist under whichever name (pseudo) the criminal used. The need for authentication or verification was born, and the questions of how to recognize or identify a person previously condemned for a crime, what are the causes of crimes, is there a way to identify the signs from crimes, etc. became the focus of interest of both natural and social scientists. Many theories were based on pure prejudice and have led to tragic decisions and outcomes (racial theories, facial or body stigmata, genetic syndromes, psychological traits, etc. justifying exclusions, investigation methodologies, to the extreme of genocides) that still haunt us today. Except for specific types of madness, most crimes fall under categories of moral or conventional decisions that may vary over time, and some changes come about because some broke what they considered wrong laws (see the debate about passive or active euthanasia in this beginning of the twenty-first century).

The use of measurements and statistics helped develop selectivity of observed traits that may differentiate individuals or groups within populations or to observe relationships (whether causal or correlated) between societies and crime. Quételet (1796–1874), a versatile mathematician, decided to apply the newly developed statistics and probability to measure social phenomena a discipline that he called "social physics." It was a key factor in the development of both sociology and criminology. His work on the measurements of man, to determine the "average man" (still used in medicine in the form of the body mass index), was also the seed to the development of a fundamental aspect of forensic science: criminal anthropometry or bertillonnage.

Bertillon and Criminal Anthropometry

Bertillon (1853–1914) was what we would call a failed scientist, but he had a systematic and orderly mind, almost to an obsession. Highly influenced by a family of scientists, he obtained a clerical job at the Préfecture de Police in Paris in 1879 because of strings his influential father pulled. His main duties were to copy and file criminal records, created in 1848 after indications that many crimes were made by few individual repetitive offenders (recidivists). These records were mostly vague, nonsystematic, partially inaccurate, and with photos (when photos were available) taken in different lights, angles, sizes, etc. It was obvious that there was no way this data could be used to recognize or describe known criminals with any chance of success. The only recognized recidivists were recognized because they were known to the police officers that arrested them, often under different assumed names, the real birth name remaining unknown most of times with little chance of reconstructing the criminal career. The family background of Bertillon led him to obsessively search for a combination of measurements on each individual that would be selective to the extreme so that any recorded criminal, when arrested again could be identified to a recidivist with the same set of measurements, whatever the assumed name. This would be a process of authentication or verification that the person was already known and had a criminal record. The assumed names were identified to a set of crimes and not necessarily to a specified identity in the sense that we use with biometry now days (biological identity).

The principle was based on the measurement of fixed and stable features offered mostly by bone structures in adult persons. Nine bone measures offered 19,683 possible features that could be combined to obtain a highly selective classification. These included the stature and the spanning (from the left shoulder to the tip of the major of the right hand) (Bertillon 1893). Examples are given below: Fig. 1.

Various other descriptive features such as skin color, eye/iris color, face morphology, nose, hairline, forehead, and shin as well as accidental or added features (tattoos, warts, scars) offered complementary attributes that would characterize an individual.

Bertillon also introduced the standardization of the "mug shot" (position, lighting, size of face, and profile) that is still the feature of police file photographs. Initial success led to the systematic introduction of anthropometry, despite heavy administrative demands, error prone measurements, and the limited application to the sole recognition of habitual criminals.

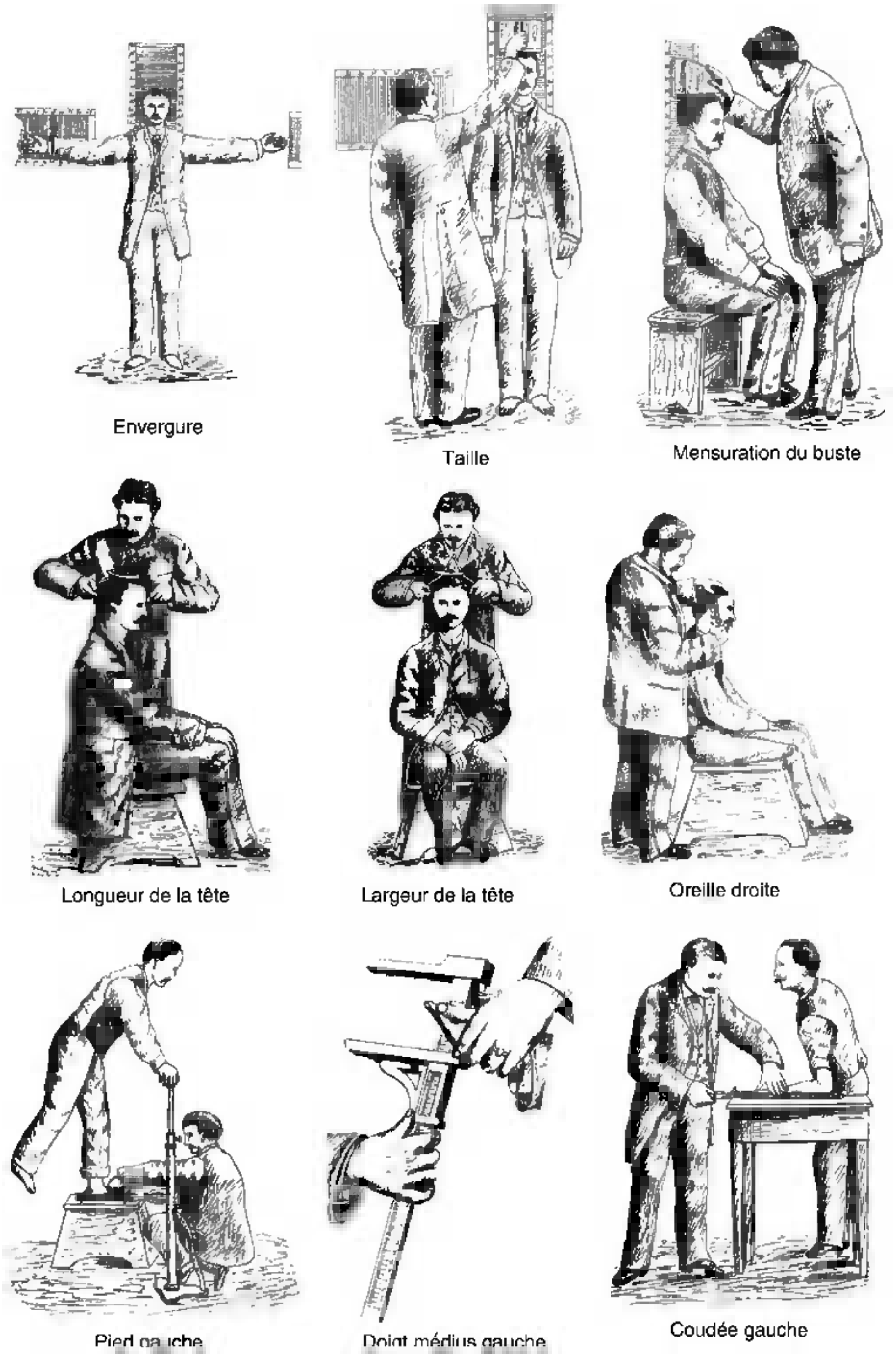

Envergure

Taille

Mensuration du buste

Longueur de la tête

Largeur de la tête

Oreille droite

Pied gauche

Doigt médius gauche

Coudée gauche

Identification and the Development of Forensic Science, Fig. 1 Examples of anthropometric measurements of the Bertillon system

Herschel, Faulds, Galton, Vucetich, and the Lyon School of Forensic Medicine (Lacassagne)

William Herschel (1833–1917) and Henry Faulds (1843–1930) fought over the discovery of fingerprints as identifiers, although both had essential merits in the development of fingerprinting. Both members of the British Empire, they observed in India (Herschel was a British civil servant) and in Japan (where Faulds was a medical doctor, part of a Scottish mission) that fingerprints were locally used as binding seals in contracts, works of arts, and judgements and decided that these minute patterns and details of papillary ridges may be used for identification.

Herschel had the merit of studying the permanence as well as the diversity of ridge details and proposed through a letter to the British Government (the so-called Hooghly letter sent in 1887) that it could be used for the identification of recidivists (Herschel 1880). The proposal was indeed similar to Bertillon's anthropometric measurements, once recorded a recidivist can be identified by his fingerprints. Faulds, on the other hand, had noticed that the contact of fingerprints with objects left an image of the pattern and details (a finger mark) that could be used to identify the source of the mark (Faulds 1880). Here the identifier is not necessarily in a file but on the fingertip of the bearer and could be used to identify the source of a contact, even if this source is not a recidivist but identified through inquiry as a potential participant to a crime. Whereas Herschel and Bertillon considered the possibility of identifying using the basic relation "person to person," Faulds proposed the identification through a "trace to person" or "person to trace" relationship, thereby being able to solve crimes from trace elements. Faulds had addressed his findings to Charles Darwin, who passed it on to a relative, Francis Galton, who mostly ignored the submission and forwarded it to the Anthropological Society of London.

The awkwardness of the anthropometric system led to question its effectiveness as a law enforcement tool. A veterinary doctor Wilhelm Eber, who had observed many times his own fingerprints in animal blood, proposed to the Preussen ministry in Berlin the use of dactyloscopy (the study of fingerprints) in 1888 rather than anthropometry, and that same year, Galton was required to discuss the identification of criminals. This led him to study Bertillon's system and reminded him of Herschel and Faulds proposals. He started to study fingerprints from an anthropological point of view and was soon convinced of its usefulness and superior identifying power in law enforcement not only in the identification of recidivists but also in crime and trace investigation. This was published in a book, now considered a classic in forensic science (Galton 1892). Bertillon resisted all discussion with Galton, and Great Britain adopted initially a mix between anthropometry and dactyloscopy. An efficient classification system (Henry's system) soon demonstrated its ability to supersede most of what anthropometry was supposed to offer (Henry 1900). The active resistance of Bertillon internationally probably was instrumental in slowing down the development of dactyloscopy.

Vucetich (1858–1925), a Croatian immigrant in Argentina, anthropologist, and police official, was in charge of setting up the Argentinian police identification system at the end of the 1880s. He came across reports of Galton and communicated with him. Convinced of the superiority of dactyloscopy over any other system, he designed a classification system (the Vucetich system) based on Galton's work and was apparently the first to use finger marks to identify the author of a crime in 1892.

To move from an administrative system designed to recognize recidivists to a system that would allow the recognition of recidivists but also the identification of criminals based on traces left at scenes of crime accepted throughout the world approximately 25 years elapsed, and the death of Bertillon in 1914 saw the abandonment of his system as a whole except for a few administrative and peculiar feature.

Ironically, beside Bertillon in France, a professor of forensic medicine in Lyon, Alexandre Lacassagne (1843–1924) was a keen observer of nature, convinced of the validity of scientific methods, and was very critical of various theories about crimes and criminals. He was very critical of theories about the criminal man propounded by Lombroso (1887) where facial appearance became an indicator of criminal propensity and critically considered the systematic use of anthropometry. Indeed, he created and opened the scientific journal he founded in 1886, "*Les archives d'anthropologie criminelle*," to conflicting views and views he opposed. He developed an active research whose influence is still felt today. Locard was one of his students, but many others carried research on aspects of trace evidence to help identify authors of crimes or reconstruct criminal activity with many fundamental theses on subjects such as pattern marks in judiciary investigations (Coutagne and Florence 1889). Other examples include research on prints in general, with detection methods for invisible or poorly visible marks (Forgeot 1891), fingerprint patterns and as traces (Frécon 1889), and blood marks and patterns (Florence 1885). Many tests and procedures to detect and analyze blood, finger marks, semen, hair, etc. survived most of the first half of last century. Locard, although his initial research was dedicated to medicolegal history, was the first, early in the twentieth century, to consider forensic science (criminalistics) as a separate branch of knowledge rather than a part of forensic medicine or police work. The research and innovations of this group were largely overshadowed by the fact that their endeavor was provincial as compared to Bertillon's central position in Paris and the importance of anthropometry developed within police circles that would view the clerical and administrative aspects as secondary support to investigation.

Hans Gross and Archibald Rodolphe Reiss

Parallel to these activities in the police, scientific, and medical world, a scholar in criminal law, and part of the intelligentsia of the Austro-Hungarian Empire, Hans Gross (1847–1915) was fascinated by many scientific and medical discoveries and had contacts with personalities that had a huge importance in developments that influenced the whole twentieth century. These included Freud (psychoanalysis), Landsteiner (blood groups), Amoëdo (odontology), and other chemists, biologists, and engineers. Gross had the insight to realize the benefit that science could bring to the criminal investigator.

Austria had a continental legal system, inquisitorial in nature, with a magistrate responsible for criminal investigation: the investigating judge/magistrate. The responsibilities for this judge was (and still is) to collect and analyze all data to try to solve crimes and when faced with a suspect to determine all information in favor or against the suspect to build a case that is forwarded to a court and the public prosecutor. Gross imagined the immense resources afforded by scientific investigation and called for an academic discipline fundamental for all magistrates to be introduced in law schools: *Kriminalistik* (German name for criminalistics or forensic science). Gross considered that forensic science was the single fundamental investigative methodology for the investigating magistrate. This was epitomized in a book (Gross 1893) that became a best seller in a second edition, translated in multiple languages. A sixth edition was produced in 1913 by Gross himself and further editions updated by various authors have been published in German until the end of the twentieth century.

Locard recognized the great influence of Gross in his own writings. Most developments in Germany, Austria, and Central and Eastern Europe recognized the key influence of Gross who saw physical traces as rich indicators for strategic, operational, and tactical decisions by the magistrate. The dimensions of intelligence, investigative or probative value of physical traces, were already perceived. The phenomenological dimension – serial crime, intelligence – and the case solving (casuistic – inclusive of investigative and evidence dimensions of traces) combined to create this brand new discipline. Gross called in 1895 for the creation of this discipline as an academic discipline in Vienna, but he finally succeeded (2 years before his death) in

1913 with the creation of an institute at the University of Graz.

A. R. Reiss (1875–1929), a chemist and photographer, immediately saw the interest of anthropometry and crime scene documentation through his interaction with Bertillon. The interest went further in the treatment of databases and international exchanges of standardized data. An international meeting in Monaco in 1914 carried the seed of what would become Interpol after the Second World War (Roux 1926) and of the current Prüm Treaty in Europe, signed in 2005, which is institutionalizing database exchanges across signatory countries (DNA, fingerprints, car registration) (Dehousse and Sifflet 2006). Of German origin, Reiss saw the merit of being part of a judiciary (legal) process and followed up scientific developments from the Lyon School. He was able to create an institute at the University of Lausanne in 1909, linked administratively to the Law faculty and to the Science and Medicine faculties for research and education. This is the first academic program in forensic science, and it is still a major part of the Faculty of Law and Criminal Sciences nowadays. His contributions (Reiss 1911) included both phenomenological and casuistic dimensions that make up the various professions of forensic science nowadays (forensic intelligence, crime analysis, scientific investigation, and evaluator in the evidence process in court). It must be emphasized that history played an important role to have Gross and Reiss become scientific "enemies" each taking opposite sides in the Balkan wars in 1915 (Collectif 2009).

From Identification to Forensic Science, and Current Organizations

The absence of a uniform view, resistance by police organizations, together with controversial and political decisions led to the construction of forensic science as a discipline regrettably based on a patchwork of miscellaneous technical and scientific applications without real fundamental or conceptual developments as a discipline and without a unifying research agenda. Kirk, in 1963, called for fundamental work and academic research (Kirk 1963), but it is still currently an identified deficiency (National Academy of Sciences 2009). This also led to various structural decisions that still haunt the discipline today.

Although many developments occurred in the detection and analysis of various trace types, including biological traces and DNA (after 1985), the early developments discussed influenced most of the current structures, and difficulties, encountered today. Some viewed the development of forensic science as a law enforcement task, almost administrative, left for identification bureaus. These identification bureaus turned into police or crime laboratories depending on policies, management decisions, and whether or not science was considered a useful component of police activities. This developed with great variations across the world and across countries. They have become large organizations often accompanied by the development of identification databases. In many cases, management by police with little scientific inclination left forensic science within secondary and supporting structures with little or no influence on investigation and intelligence for that matter.

Other countries and organizations have decided to link expertise laboratories to the judiciary and belong, often, to departments of justice. This is indicating a certain independence from police services but also creating tensions as to who is responsible for scene of crime investigations and who is in charge of databases. Some countries, especially in Eastern Europe, have had dual laboratory services, police/crime laboratories, and justice department laboratories to this day. Ironically, Locard developed his expert activities within the Lyon tribunal, but his laboratory was reclaimed by the French National Police to become the headquarters of the French forensic science laboratories in Lyon, whereas the identification bureaus remain a distinct entity within the police. The first are manned by scientists and civilian, the second by police and are considered police province, but with some overlap.

Few of the academic programs that provided expertises both for law enforcement and courts survived the two World Wars (Lausanne and Berlin), and forensic medicine absorbed part of the forensic science activities within their laboratories (especially since the advent of DNA). Newer organizations or laboratories, whether part of the justice system or law enforcement agencies, have integrated forensic medicine within forensic science laboratories but with some administrative difficulties residing in their business models. Should forensic science laboratories be financially independent industries or are they part of the overall justice systems under government control? This is an opened entry that needs fundamental and policy decisions that may, in turn, overcome some of the barriers created by history.

Related Entries

- ▶ Biometrics and Border Control Policing
- ▶ DNA Profiling
- ▶ Evidence-Based Policing
- ▶ Fingerprint Identification
- ▶ Forensic Science
- ▶ History of Criminal Investigation
- ▶ Identity Theft
- ▶ Philosophy of Forensic Identification

Recommended Reading and References

Bertillon A (1893) Identification anthropométrique; instructions signalétiques, 2nd expanded edition from the 1882 “Identification anthropométrique” 1ére edition. Imprimerie administrative, Melun

Collectif (2009) Le théâtre du crime. Presses Polytechniques et Universitaires Romandes, Lausanne

Coutagne H, Florence A (1889) Les empreintes dans les expertises judiciaires. Archives d’anthropologie criminelle et des sciences pénales 4:25–56

Dehousse F, Sifflet D (2006) Les nouvelles perspectives de la coopération de Schengen: le traité de Prüm. (New perspectives on Schengen cooperation: the treaty of Prüm.) Egmont European Affairs Publication, Bruxelles

Faulds H (1880) On the skin-furrows of the hand. Nature 22:605

Florence A (1885) Les taches de sang, leur signification en médecine judiciaire. Doctoral Thesis, University of Lyon

Forgeot R (1891) Les empreintes en général. Doctoral thesis, Univesity of Lyon

Frécon A (1889) Des empreintes en général et de leur application dans la pratique de la médecine judiciaire. Doctoral thesis, Université de Lyon

Galton F (1892) Finger prints. Macmillan, London

Gremaud J-L (2010) Processus de reconnaissance et d’identification de personnes décédées. Université de Lausanne, Lausanne

Groebner V (2007) Who are you? Identification, deception, and surveillance in early modern Europe. Zone Books, New York

Gross H (1893) Handbuch für Untersuchungsrichter als System der Kriminalistik. Leuschnen und Lubensky, Graz

Henry ER (1900) Classification and uses of finger prints, 4th edn. Georges Routledge, London

Herschel W (1880) Skin furrows of the hand. Nature 23:76

Kirk PL (1963) The ontology of criminalistics. J Crim Law Criminol Police Sci 54:235–238

Lombroso C (1887) L’homme criminel, 4ème édition italienne traduite edition. Félix Alcan, Paris

National Academy of Sciences (2009) Strengthening forensic science in the United Sates: a path forward. National Academies Press, Washington, DC

Reiss AR (1911) Manuel de police scientifique (technique). I. Vols et homicides. Payot Alcan, Lausanne

Roux JA (1926) Actes du premier congrès de police judiciaire internationale. Marchal et Billard, G. Godde succ, Paris

Identification Issues in Life Course Criminology

Robert Brame and Michael G. Turner
Department of Criminal Justice and Criminology, University of North Carolina at Charlotte, Charlotte, NC, USA

Overview

In everyday language, the word “identify” refers to the presentation of proof or convincing evidence that something is what it appears to be. This is also what identification means in the social sciences and in life-course criminology. Life-course criminologists try to understand how crime develops, changes, and evolves over the life span. Their efforts usually fall into one or more of the following two categories: (1) describe stability and change in some criminologically interesting characteristic or behavior and (2) discern whether certain factors are causes or correlates of criminal behavior. Within each of these categories, a focus on identification leads to questions about what can be measured, what is being measured, and what are the main sources of uncertainty. This entry provides a brief discussion of the key identification issues in each category of life-course research.

General Terminology

The most comprehensive contemporary works in the identification area are those of Charles Manski (1995, 2003). In general, Manski's work revisits and then extends a much older literature (Koopmans 1949; Cochran et al. 1954). A key theme throughout all of this work is that identification is a matter of degree. So it is more useful to speak of stronger and weaker identification instead of speaking of an estimate as being identified or not identified Manski (2003, p. 2).

Manski further notes that the social sciences have typically overemphasized the importance of point identification. Point estimation occurs when a researcher is able to combine the data with a set of assumptions that are strong enough to yield an answer that can be transmitted with a single estimate. Point estimates can only be obtained when a parameter to be estimated is point-identified. And point identification is generally possible in the social sciences when certain (usually quite strong) "point-identifying assumptions" are met. The main problem is that well-informed researchers often disagree about the validity of the assumptions required for point identification. These disagreements end up creating ambiguity about the accuracy of the point estimate. Manski (2003, p. 1) describes this problem as "the law of decreasing credibility: the credibility of inference decreases with the strength of the assumptions maintained." With weaker assumptions, point identification may no longer be possible. But it will often still be possible to identify a set of values (a partial identification interval) that include the correct answer. Manski (2003, p. 2) refers to the practice of combining weaker assumptions with data to yield set-identified parameters as "partial identification." An important feature of this approach is that it will often be possible for well-informed scientists and policy makers to agree on the weaker identifying assumptions, thereby increasing the likelihood of achieving consensus about the validity of the set-identified estimate.

The level of uncertainty in estimates creates a tension in social science research. On the one hand, social scientists are generally trained to think about their analyses and the presentation of their results in terms of point estimates. Such estimates are often presented along with some measure of uncertainty due to sampling error (i.e., standard errors and confidence intervals). On the other hand, a careful "partial identification" analysis typically reveals much greater levels of uncertainty than what one would normally encounter with conventional point estimators (combined with estimates of sampling error). In fact, as Manski's work (and the work of others) has shown, analyses based on weaker identifying assumptions often lead to uncertainty that dwarfs what social scientists are accustomed to seeing (see, e.g., Manski and Nagin 1998). So, a partial identification analysis will typically transmit high levels of uncertainty in exchange for weaker assumptions – assumptions which will be plausible to a wide range of observers. An emphasis on point identification often transmits small degrees of uncertainty in exchange for assumptions that are much stronger – and often not credible to many observers.

Stability and Change

Life-course criminologists are concerned with questions about what stays the same and what varies over time. Sometimes efforts to answer these questions focus on cross-sectional data. A good example of a cross-sectional analysis with life-course implications comes from Tittle and Ward (1993). In this study, the authors used survey data collected at a single time point (1972) to study the relationship between age and criminal involvement covering ages 15–94 in three states (Iowa, New Jersey, and Oregon). The purpose of the study was to examine predictors and moderating influences on the venerable "age-crime curve" (Gottfredson and Hirschi 1990). Similarly, researchers have long used cross-sectional data from the Federal Bureau of Investigation's Uniform Crime Reporting (UCR) program to study arrest rates stratified by ageacross the population (Piquero et al. 2003) – with the goal of understanding how criminal involvement covaries with age.

Unfortunately there are ambiguities with each of these efforts. The Tittle and Ward (1993, p. 11) study states that "[i]nterviews were completed successfully with respondents from 57 % of the originally selected households (77 % of the screened subjects)." Despite Tittle and Ward's (1993, p. 11) claim that the resulting sample matched up closely to census demographic statistics in the surveyed areas, the high rate of missing data does create some ambiguities. And it is well known that a small – but not insignificant – number of law enforcement agencies decline to participate in the UCR program in any given year. After all, police departments can only arrest suspects in crimes that have been reported. Since many victims choose not to report their victimizations to the police, those cases are quite unlikely to lead to an arrest. More importantly if those reporting practices change in important ways over time, the biases in the data can change as well.

Over the past 40–50 years, criminologists have increasingly turned their focus to the analysis of longitudinal micro-data or "panel" datasets. Criminologists have relied on these studies to measure key features of what has come to be known as the "criminal career" (Piquero et al. 2003). They are distinguished from cross-sectional studies by their ability to measure how behaviors remain stable or change over time for the same individual persons. There are generally two methods by which longitudinal micro-data are collected: (1) retrospective studies of administrative records that yield a sequence of dates on which particular events occurred (Tracy et al. 1990; Piquero et al. 2007) and (2) prospective studies of the same individuals over a sustained period of time to measure a range of life events and experiences (Elliott et al. 1989; Piquero et al. 2007; Thornberry et al. 1991; Huizinga et al. 1991; Loeber et al. 1991). These two types of studies are not mutually exclusive. In some instances, the retrospective studies have been supplemented with interviews, and in some instances the prospective studies have included detailed reviews of administrative data. The difference between them is generally a difference in emphasis or focus. Both types of studies yield a stream of information for each individual about particular events, experiences, and perceptions.

These studies also have some measurement obstacles that may impact identification. Administrative records may be incomplete for some individuals, particularly if they move out of the area or become incapacitated. And, as in cross-sectional studies, administrative records only measure what is recorded by an agency. Survey data can be compromised when individuals who are targeted to be in a study cannot be located or do not provide consent for participation. This can be a problem at the time of initial contact (enrollment bias), or it can become a problem as the study progresses when individuals drop out (attrition bias). Still another concern is that survey participants learn over time that they can shorten the survey if they answer questions in particular ways (a testing effect) (Thornberry and Krohn 2000; Lauritsen 1998). Regardless of the source of the measurement obstacle, data that are measured incompletely affect the degree of identification.

A recent example of the ambiguity created by these sorts of issues comes from a study of the National Longitudinal Survey of Youth (1997 version; NLSY97) conducted by Brame et al. (2012). This study's main objective was to estimate the proportion of the US population that had been arrested or taken into custody at least one time for something other than a minor traffic offense by age 23. The NLSY97 was designed to be based on a nationally representative sample of households with youth between the ages of 12 and 16 years old on December 31, 1996. Two samples were surveyed – a self-weighting random sample and a minority oversample. Brame et al. (2012) focused their efforts on the 7,335 youths who were selected for inclusion in the random sample. Each person in the sample was asked an initial question about prior arrest experiences at the first interview (1997). Then, at each subsequent interview (through 2008), the respondents were asked about arrest experiences that had occurred since the last interview.

Aside from the obvious difficulties of asking individuals to self-report their own arrest experiences, the study encountered three additional problems. First, only 6,748 (91.9 %) of the originally selected 7,335 persons in the sample actually completed a first interview in 1997. Second, among those who did complete a first interview, many persons missed at least one interview by age 23. Third, even among those who were interviewed, some declined to answer questions about their arrest experiences. By age 23, Brame et al. (2012, p. 23) were able to identify 1,858 people who had been arrested at least once and 4,299 people who had not yet been arrested. For the remaining 1,178 (16.1 %) persons, the ever-arrested status could not be determined because of incomplete data.

Criminologists who study individual-level datasets like the NLSY97 often encounter these kinds of incomplete data problems. One way to address these problems is to assume that the missing cases are a simple random sample of the cases that were originally targeted for inclusion in the study (the missing at random assumption) (Manski 1995). With this assumption, Brame et al. (2012) study would yield an estimated arrest rate of 1,858/(1,858 + 4,299) = 30.2 %. The problem is that there is no way to test the validity of this assumption. It is quite possible the arrest experiences of the missing cases look quite different from those of the observed cases. Since the missing cases are missing, however, there is no way to verify this assumption.

An important insight offered by Manski (1995, 2003) provided the authors with a path forward, however. This way forward was based on the following question: what can the data alone reveal? The answer to this question is obtained by considering two extreme cases: (1) the arrest rate assuming that all of the missing cases were arrested ([1,858 + 1,178]/[1,858 + 4,299 + 1,178] = 41.4 %) and (2) the arrest rate assuming that none of the missing cases were arrested (1,858/[1,858 + 4,299 + 1,178] = 25.3 %). So, assuming individuals who participated accurately reported their experiences, the actual arrest rate must lie somewhere between 25.3 and 41.4 %. In the special case where the arrest rate is the same for both the observed and the missing cases, the arrest rate for the entire 7,335 cases originally targeted for the survey would be 30.2 %.

Some observers might look at the [25.3 % and 41.4 %] interval and find it to be excessively wide and therefore relatively useless to draw substantive conclusions. But a careful, scientific treatment of identification leads to a different set of insights. As Manski would point out: (1) the interval before looking at the data was [0 %,100 %] so the fact that the data allow us to reduce our uncertainty to [25.3 %,41.4 %] means that significant progress has been made; (2) without additional information about the missing cases, we cannot say with any empirical basis that one value in the interval is more probable than another value; (3) the ambiguity caused by not knowing the status of the missing cases leads to much greater amounts of uncertainty than what we normally encounter when we only consider sampling error; (4) studies often do not consider this kind of uncertainty, and if they did, the conclusions of social scientists would appear to be much more ambiguous and fragile than they are typically presented as being; and (5) the extreme bounds are a starting point for further analysis, not an ending point. Researchers can follow up on the extreme bounds analysis by examining how stronger assumptions may be invoked to tighten the bounds. The advantage of this approach is that it is clear how much of the identification is based on the data alone and how much is based on fundamentally untestable assumptions about what cannot be measured.

Toward this end, Brame et al. (2012) noted that if we are willing to assume that the missing cases are at least as likely to have been arrested as the observed cases, then the bounds shrink from the extreme case of [25.3 %, 41.4 %] to [30.2 %, 41.4 %]. There are some compelling criminological reasons to make this assumption. In fact, researchers using the NLSY97 have documented higher attrition rates at later survey waves for individuals self-reporting problem behaviors in the initial wave (McCurley 2006). These data suggest that individuals with a higher propensity of problem behaviors have a lower probability to remain within the sample of survey respondents, but if some researchers or policy analysts do not accept it, the bounds revert back to the extreme case. In this sense, the bounds take on a "no-free-lunch" quality. One starts by considering what can be learned with minimal assumptions (i.e., all respondents understand the question that is being asked, they are cognitively able to answer the question, and that the information they provide is accurate) combined with the data. If one is not willing to make further assumptions, then the extreme bounds will be as tight as possible. If one is willing to make further assumptions, then the argument for those assumptions can be made, and if the audience is convinced, the uncertainty will shrink. If the audience is not convinced, the extreme bounds will form what Manski calls the "domain of consensus" about how much the data are able to reduce uncertainty.

A first order of business in any developmental study of criminal behavior is to measure both stability and change in criminal involvement. Yet, criminological datasets used to measure these quantities are nearly always hampered by incomplete or missing data. Criminologists are accustomed to adopting easy and convenient fixes for these problems. But these fixes often do not come fully to terms with the uncertainty created by missing data. The advantage of a comprehensive identification analysis is that it shows the audience exactly what the data alone can reveal. Then, different sets of researchers and policy audiences can see how much a specific conclusion depends on the data and untestable assumptions about the data.

Distinguishing Between Correlation and Cause and Effect

Of course, criminologists are often interested in moving beyond descriptive analyses in order to measure the causal effect of interventions and life experiences on criminality. Early criminologists often thought of the causes of crime in terms of events, circumstances, and experiences that occurred at about the same time as criminal involvement. But contemporary criminologists take a more nuanced position on this issue arguing that some causes of crime are likely proximate to criminal involvement while others are rooted in the formative aspects of the life course – namely infancy, toddlerhood, and early childhood (Grasmick et al. 1993).

For example, Gottfredson and Hirschi's (1990; see also Hirschi and Gottfredson 1983) theory of self-control maintains that the failure of parents to adequately socialize children contributes to poor development of self-control which, in turn, leads to a greater risk of criminal involvement. From their perspective, while "socialization continues to occur throughout life" (Gottfredson and Hirschi 1990, p. 92), differences in self-control between individuals of the same age are destined to remain about the same throughout the life span (i.e., someone who ranks low in self-control at age 8 will also rank low in self-control at age 58). Their position gives voice to an important axiom of contemporary criminology: there is a rival hypothesis for any apparent effect of a post-childhood life experience or event on criminality. The implications are profound and can be illustrated with a few prominent examples.

Among the most well-replicated findings in life-course criminology are (1) the positive correlation between past and future criminal behavior (Nagin and Paternoster 2000) and (2) the negative correlation between the time since one's last offense and the likelihood or risk that one commits new offenses (Maltz 1984; Schmidt and Witte 1988; Kurlychek et al. 2012). There are clear causal interpretations of these results. Prior offending may reduce one's opportunity to live in stable housing, develop normal healthy relationships with family and friends, obtain an education, or secure employment, and these obstacles may, in turn, lead to increased criminality. Similarly, when an individual leaves prison, there may be a sustained effort to avoid criminality, and the actual experience of avoiding criminality each day causally and gradually reduces the risk of recidivism over time. But, of course, it is also possible that higher-risk individuals are more likely to offend at all times inducing a strong positive correlation between past and future offending behavior and a strong negative correlation between time since one's last offense and the risk of future offending. Hence, an important source of ambiguity in these results is whether the correlations are the result of causal effects or whether they are simply artifacts of failure to measure and properly control for the influence of stable tendencies (like low self-control). Econometricians have long understood that persistent between-individual differences are capable of producing these exact patterns (Heckman 1978). The question is as follows: when these patterns appear, are there reliable ways to discern their source?

This is a classic identification problem. If a researcher can measure the stable between-individual differences in factors like self-control, then it would be easy to identify individuals that are comparable in terms of these differences and look at the data to see whether the ones who have offended in the past are more likely to offend in the future or whether variation in time since the last offense is able to predict the limiting risk of future offending. The problem is that researchers will typically only be able to create partial measures of the relevant between-individual differences. Econometric models can help, but these models typically require strong assumptions related to the functional form of the model, the initial conditions preceding the sequence of over-time observations, the probability distribution of the error term, and the stability of measurement error over time. The ambiguity of partial measurement and questions about how much of the analysis depends on these strong and untestable assumptions can lead to intractable debates about whether causal inference is viable.

One way to make progress is for researchers to become more specific about the factors they believe will have important effects on criminality and then design studies to rigorously test for those effects. Toward this end, Daniel Nagin and his colleagues (see Nagin and Odgers 2010 for a useful discussion) have embarked on an increasingly comprehensive program of research to investigate causal effects for residential placement of juvenile offenders, gang membership, grade retention in school, imprisonment, and adolescent employment. A common theme of this work is that an explicit effort is made to ensure that comparisons between individuals are well posed so that the basis for identification of causal effects is transparent.

The first of these efforts – Manski and Nagin (1998) – considers the very difficult problem of estimating the effect of sanctions on recidivism for juvenile offenders. This study relied on a broad-ranging sample of 13,197 juveniles who were adjudicated delinquent in Utah juvenile courts at least 2 years before their 18th birthday. Each of these juveniles was characterized as receiving a residential placement (11 %) or a community placement (89 %). All of the juveniles were then followed up over a 2-year period, and the number who reappeared in the juvenile court within that period was counted for each group. This analysis revealed that 61 % of the juveniles returned to the court within the 2 years (77 % for the residential placement group and 59 % for the community group). The key question is what can be learned about the effects of residential versus community placement from these data. It turns out that with only minimal assumptions (i.e., no measurement bias, each person's outcome is independent of each other person's treatment), the sign of the sanction effect is not identified. If one is willing to make two assumptions, then it is possible to reduce uncertainty about the sanction effect: (1) sanctioning practices vary between Utah judicial districts, and (2) the average outcome after receiving a given sanction does not vary between Utah judicial districts. This pair of assumptions is called an exclusion restriction and has identifying power because there is no obvious reason to suppose that judicial district affects recidivism. Combining these assumptions with the additional assumption that the highest risk offenders are more likely to receive residential placement (what Manski and Nagin refer to as the "skimming model") leads to the conclusion that residential placement is criminogenic. On the other hand, combining these assumptions with the assumption that each individual receives the sanction that minimizes their risk of recidivism (what Manski and Nagin refer to as the "outcome optimization model") leads to the conclusion that residential placement reduces recidivism.

The identification analysis of Manski and Nagin is useful on three different levels. First, it tells us what the data alone (with minimal assumptions) can reveal about the sign and magnitude of sanction effects. Second, it tells us what assumptions are required to make progress. Third, it tells us where the priority needs to be placed for future research. In this case, researchers learn that a better understanding of juvenile judges' priorities in assigning dispositions is needed in order to better understand the sanction effects. It is noteworthy that efforts to simply condition the comparisons on a set of observed factors in the data would not have led to these insights.

Nagin's more recent efforts have emphasized the use of semi-parametric, group-based trajectory models to understand population variation in the development of criminal behavior over time. An important by-product of these models is a means to probabilistically classify individuals as belonging to groups or "latent classes" that are comprised of individuals with relatively similar developmental trajectories (Nagin and Odgers 2010) prior to some intervention or experience of interest. These models have been somewhat controversial among criminologists because of concerns about "group reification" (for discussion see Nagin and Tremblay 2005; Sampson and Laub 2005). But these concerns do not negate the fact that the method reliably identifies individuals with relatively similar developmental trajectories of criminality. The critical insight of Nagin and colleagues is that the trajectory groups can be used as the basis for creating reservoirs of comparable individuals prior to the occurrence of a treatment or experience that could affect future criminality. In practice, Nagin and colleagues have used trajectory groups in conjunction with so-called propensity scores (Rosenbaum and Rubin 1983) to identify persons that match each other on background characteristics but experience different treatments or interventions.

The identifying power of these comparisons is based on the idea that two or more people who match each other on an extensive set of pretreatment characteristics can be sufficiently comparable to have a reasonable discussion about causal inference. The advantage of the Nagin approach is that the discussion of identification does not devolve into debates about whether the modeling assumptions are correct. Instead the discussion is based on whether the set of characteristics on which one matches are reasonable and comprehensive. Such discussions can provide the field with an understanding of what a comparison of reasonably well-matched cases looks like and they can inform the priorities for variables that should be measured in subsequent research. An added benefit is that these analyses are easy to explain to nonscientific policy audiences.

A good illustration of these issues appears in Nieuwbeerta et al. (2009). In this study, Nieuwbeerta and colleagues used data from the Netherlands Criminal Career and Life-Course Study to investigate the effects of first-time imprisonment on recidivism. The first step of the analysis was to estimate a series of conviction trajectory models for varying durations beginning at age 12 (with a maximum of age 37). Next, within each trajectory group, a propensity score was estimated for each individual in that group. The propensity score for an individual measured the estimated probability that an individual receives a first term of imprisonment conditional on an extensive set of offense-related, demographic, and life circumstances characteristics that were measured before the sentence was imposed. The propensity score estimator is based on a theorem from Rosenbaum and Rubin (1983) which demonstrates that as the sample size grows infinitely large, matching or stratifying treatment and control cases on the estimated propensity score will tend to balance or "equalize" the factors used to estimate the propensity score between the treatment and the control groups. This is a powerful result and it is increasingly being used in criminology and the other social sciences to estimate treatment effects in ways that are more convincing than the standard regression-based estimators.

The third step of the analysis involved matching each imprisoned person to 1, 2, or 3 non-imprisoned individuals. The matching process guaranteed that matched individuals are similar in terms of their prior conviction histories and in terms of the pretreatment characteristics used to estimate the propensity score. The analysis showed that the recidivism rates of imprisoned individuals were significantly higher than those of non-imprisoned individuals after matching. While this is a compelling result, the analysis had some important limitations including (1) the inability to match some of the most serious imprisoned offenders in the study with a comparable person who did not receive a confinement-based sentence, (2) a wide-ranging but still finite set of covariates and background characteristics to match on, and (3) for the sake of inferential clarity, only first-time imprisonment experiences were studied. But even with these limitations, the Nieuwbeerta study provides a useful model for criminologists to consider as they think about causal inference and identification in life-course criminology.

Conclusions

Criminologists are increasingly devoting attention to issues of identification as they conduct life-course research. This trend is bringing a greater focus and scrutiny to issues such as missing data and estimation of treatment and intervention effects among individuals that are suitably comparable to each other on developmentally appropriate factors and characteristics. It is all too easy for researchers to simply drop cases with missing information from their analysis and proceed with a study of the residual sample of nonmissing cases while hoping that the biases are not large. It is all too easy for researchers to estimate a statistical model-based treatment effect while hoping that the functional form assumptions of that statistical model are consistent with the data. A focus on identification moves away from a criminology of excessive "certitude" (Manski 2011) and toward a criminology of sensitivity analysis, robustness testing, and verification.

When cases have missing data, a focus on identification leads to what-if questions about those cases. What if the missing cases are more likely than the observed cases to be involved in criminal behavior? If that is true, how would the analysis results change? What assumptions are criminologists willing to make to achieve identification? What assumptions will criminologists agree are reasonable? What do the data reveal about criminologically interesting quantities if researchers are only willing to make minimal assumptions? These are all questions that have not been asked much in life-course criminology. But they are starting to be asked more and the methods to make those questions answerable are becoming more widely available.

Methods for causal inference in observational life-course studies are also becoming more sophisticated and thoughtful. Here, a focus on identification leads to equally useful questions. For example, here is a question that should always be asked in a causation-focused study: what is the empirical basis for comparing these cases to each other? Is this comparison an apples-to-apples comparison? How are the groups being compared different from each other in terms of characteristics that are, at least in theory, measurable before an intervention (going to prison) or life experience (getting married or securing employment) occurs?

Criminologists are accustomed to thinking in theoretical terms about the causes of criminal involvement. But a focus on identification leads to broader questions like why some people go to prison while others do not. Why is it that some people get married or get a job and others do not? Does gang involvement lead to criminality? Or does criminality lead to gang involvement?

These are questions that criminologists have generally not considered as much as the questions addressed by traditional crime theories. But a focus on identification means these questions are every bit as important as the traditional theoretical terrain.

Ultimately, identification is about researchers coming to terms with what is known and that which remains unknown. If the data are missing, it is of little use to confine our analysis to the simplistic assumption that the data are missing purely at random. It is far more fruitful to present the analysis with reasonable attention to what happens to the analysis if the assumption is wrong. If the analysis relies on a comparison of one group of people to another group of people, then criminologists should make clear and convincing arguments about why those comparisons are sensible and what would happen to the analysis if the assumptions are wrong. In sum, the more the field thinks about identification, the farther the field will advance toward a more thoughtful and rigorous science of criminal behavior.

Related Entries

- ▶ Age-Crime Curve
- ▶ Criminal Careers
- ▶ Econometrics of Crime
- ▶ Parametric Sample Selection Models
- ▶ Sample Selection Models
- ▶ Sample Selection Problems

Recommended Reading and References

Brame R, Turner MG, Paternoster R, Bushway SD (2012) Cumulative prevalence of arrest from ages 8–23 in a national sample. Pediatrics 129:21–27

Cochran WC, Mosteller F, Tukey JW (1954) Statistical problems of the Kinsey Report. J Am Stat Assoc 48:673–716

Elliott DS, Huizinga D, Menard S (1989) Multiple problem youth: delinquency, substance use, and mental health problems. Springer, New York

Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Grasmick HG, Tittle CR, Bursik RJ, Arneklev BJ (1993) Testing the core empirical implications of Gottfredson and Hirschi's general theory of crime. J Res Crime Delinq 30:5–29

Heckman JJ (1978) Simple statistical models for discrete panel data developed and applied to test the hypothesis of true state dependence against the hypothesis of spurious state dependence. Annales de l'inséé 30–31:227–269

Huizinga D, Esbensen F-A, Weiher AW (1991) Are there multiple paths to delinquency? J Crim Law Criminol 82:83–118

Koopmans TC (1949) Identification problems in economic model construction. Econometrica 17:125–144

Kurlychek MC, Bushway SD, Brame R (2012) Long-term crime desistance and recidivism patterns: evidence from the Essex County convicted felon study. Criminology 50:71–103

Lauritsen J (1998) The age-crime debate: assessing the limits of longitudinal self-report data. Soc Forces 77:127–155

Loeber R, Stouthamer-Loeber M, Van Kammen W, Farrington DP (1991) Initiation, escalation, and desistance in juvenile offending and their correlates. J Crim Law Criminol 82:36–82

Maltz MD (1984) Recidivism. Academic, Orlando

Manski CF (1995) Identification problems in social science research. Harvard University Press, Cambridge, MA

Manski CF (2003) Partial identification of probability distributions. Springer, New York

Manski CF (2011) Policy analysis with incredible certitude. Econ J 121:F261–F289

Manski CF, Nagin DS (1998) Bounding disagreements about treatment effects: a case study of sentencing and recidivism. Sociol Methodol 28:99–137

McCurley C (2006) Self-reported law-violating behavior from adolescence to early adulthood in a modern cohort. Final report to the National Institute of Justice, Washington, DC

Nagin DS, Odgers CL (2010) Group-based trajectory modeling in clinical research. Annu Rev Clin Psychol 6:109–138

Nagin DS, Paternoster R (2000) Population heterogeneity and state dependence: state of the evidence and directions for future research. J Quant Criminol 16:117–144

Nagin DS, Tremblay RE (2005) Developmental trajectory groups: fact or a useful statistical fiction. Criminology 43:873–904

Nieuwbeerta P, Nagin DS, Blokland AAJ (2009) Assessing the impact of first-time imprisonment on offenders' subsequent criminal career development: a matched samples comparison. J Quant Criminol 25:227–257

Piquero AR, Farrington DP, Blumstein A (2003) The criminal career paradigm. Crime Justice Rev Res 30:359–506

Piquero AR, Farrington DP, Blumstein A (2007) Key issues in criminal career research: new analyses of

the Cambridge Study in Delinquent Development. Cambridge University Press, New York
Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. Biometrika 70:41–55
Sampson RJ, Laub JH (2005) Seductions of method: rejoinder to Nagin and Tremblay's "Developmental Trajectory Groups: fact or fiction?". Criminology 43:905–913
Schmidt P, Witte AD (1988) Predicting recidivism using survival models. Springer, New York
Thornberry TP, Krohn MD (2000) The self-report method for measuring delinquency and crime. In: Dufee D (ed) Measurement and analysis of crime and justice: criminal justice 2000. National Institute of Justice, Washington, DC, pp 33–83
Thornberry TP, Lizotte AJ, Krohn MD, Farnworth M, Jang SJ (1991) Testing interactional theory: an examination of reciprocal causal relationships among family, school, and delinquency. J Crim Law Criminol 82:3–35
Tittle CR, Ward DA (1993) The interaction of age with the correlates and causes of crime. J Quant Criminol 9:3–53
Tracy P, Wolfgang M, Figlio RM (1990) Delinquency careers in two birth cohorts. Plenum, New York

Identification Technologies in Policing and Proof

Gary Edmond[1] and Simon A. Cole[2]
[1]Network for Interdisciplinary Studies of Law (NISL) School of Law, The University of New South Wales, Sydney, NSW, Australia
[2]Department of Criminology, Law & Society, University of California, Irvine, CA, USA

Synonyms

Admissibility; CCTV; Courts; Criminal law; DNA; Fingerprints; Forensic science; Images; Individualization; Intelligence; Reliability; Resistance

Overview

As citizens we are increasingly observed, traced, and documented in our routine and some of our not-so-routine activities. While it is common for citizens and scholarly commentators to question, and frequently criticize, invasive uses of emerging technologies, it is less common to focus attention on actual technical capacities and the use of "traces" as intelligence and evidence in criminal justice practice. This omission is curious, because there are long-standing problems with identification and proof in legal settings. This entry discusses the reliability, and therefore evidentiary (or probative) value, of techniques (and technologies) routinely used for surveillance and proof of criminal activities (including terror offenses). Recent work on surveillance and surveillance technologies has tended to focus on their potential to change or disrupt, particularly older, settlements (and values) around privacy and visibility. This entry endeavors to shift that focus by introducing some of the epistemic dimensions and institutional responses to surveillance technologies, and their role in identification, in order to understand their use as intelligence and evidence, in criminal investigations, plea bargains, and criminal prosecutions. Criminological literatures, particularly surveillance theory, have much to offer discussion of forensic science and forensic medicine, but recent challenges to the value of many technologies and derivative interpretations suggest that threats posed by "Big Brother" – that is, the state's ability to "watch" – might, in some regards, be more clumsy (and simultaneously complicated, attenuated, and remote) than is often assumed.

Thinking about the role and use of the artifacts (or products) of surveillance in investigative and prosecutorial contexts is revealing because it tends to expose, both directly and indirectly, the limits of the technologies as well as the limits of legal processes. Recent authoritative criticisms of many forensic sciences (e.g., Saks and Koehler 2005; National Research Council 2009) reinforce the importance of questioning claims made by

This research was supported by the ARC (FT0992041, LP100200142 and LP120100063). Thanks to Mehera San Roque for comments.

police, forensic scientists, lawyers, and judges. They might also lead us to doubt the value of legal institutions as appropriate regulators (or “gatekeepers”) of surveillance technologies and derivative evidence.

Substantially, our entry begins with a review of three technological assemblages ubiquitous in modern criminal investigations and prosecutions, namely, those associated with the analysis of latent fingerprints, DNA, and security and surveillance images. Reviews of these particular technologies and the transformation of traces into admissible evidence are rewarding because of their distinctive historical trajectories, divergent scientific foundations, and pervasive use across a range of intelligence, investigative, forensic, and administrative settings. Ubiquitous in forensics and surveillance literatures as well as the popular imagination, fingerprints, DNA, and images provide interesting insights and illuminating comparisons to aid our understanding of surveillance technologies and their *evidentiary* products.

Latent Fingerprints

The identification of individuals is, of course, a problem of long standing to which a variety of technological solutions have been proposed (e.g., Groebner 2007). Concerned with increasingly mobile populations, in the nineteenth century, state governments began to develop bureaucratic *systems* of identification that were attentive to the storage, deployment, and ability to retrieve information. One of these new systems, fingerprint identification, which was developed nearly simultaneously in colonial India for the control of native populations and Argentina for the control of “criminal classes” immigrating from Europe, was perhaps the first identification technology to develop, inadvertently it seems, a forensic application. While fingerprint identification was based on impressions of the “friction ridge skin” that covers the fingertips created deliberately using ink, it was noted that similar impressions are often deposited inadvertently when individuals touched objects with smooth services. Several different observers proposed that such impressions might provide evidence that a particular individual had touched an object that might, in some cases, prove useful in the investigation and prosecution of crimes. In early cases, the clerks who maintained identification bureaus were called upon to undertake the analysis of such “latent” impressions.

By the 1920s, fingerprint identification had become the world’s dominant system of identification for institutions, like prisons and police departments, charged with keeping track of “criminal” populations. These institutions began assembling large databases of inked fingerprints to prevent apprehended individuals escaping from their criminal pasts by adopting alternative identities (i.e., aliases). Latent fingerprint identification was becoming increasingly common as well. Methods for “dusting” (and photographing) objects in order to visualize latent impressions were developed, and the search for latent prints became a regular feature of police investigations, especially for more serious offenses. However, to be useful in identifying the perpetrator of a crime, latent print identification was dependent either on the selection of a suspect by other investigative means or a painstaking and resource-intensive “cold search” of a local or regional database – often a manual filing system. Nevertheless, fingerprint identification developed into the most trusted and iconic means of both forensic and administrative identification.

In court, the interpretation of latent prints was quickly deemed admissible following a modest degree of resistance by defendants. Courts based admissibility on: (1) a general assumption among informed observers that the evidence must be reliable; (2) the observation, though not systematically recorded or tested, that experts’ conclusions regarding latent prints were rarely refuted *post hoc* by external, epistemologically superior, facts; (3) an intuitive sense of the very high complexity and variability (or “selectivity,” Champod et al. 2004) of fiction ridge skin itself.

By the 1980s, automated fingerprint identification, which had been under development for decades, had become sufficiently mature to penetrate major law enforcement agencies. This development made database searches much less costly, greatly enhancing the utility of fingerprint databases and the potential value of recovered latent prints.

In the 1990s, influenced largely by the controversies over DNA typing (discussed below), many commentators began to question the legal assumption that the accuracy and discriminability of latent print association could be taken on trust. This rethinking, along with some high-profile errors, led to legal challenges to fingerprint evidence, especially in the United States (Cole 2001). For the most part, these challenges have not persuaded courts to substantially modify their practices, but they have generated renewed attention to fingerprint evidence, in the form of a number of high-profile official inquiries and reports (National Research Council 2009; Campbell 2011; Expert Working Group on Human Factors in Latent Print Analysis 2012), and arguably, they played a role in stimulating new efforts at basic research on the accuracy (Ulery et al. 2011; Tangen et al. 2011) and discriminability (e.g., Neumann et al. 2012) of latent print associations.

DNA Profiling

The development of a method for visualizing "hypervariable" regions of genotypes by Alec Jeffreys in the mid-1980s led to the proposal that genetic variations might be harnessed to provide another means of individual identification. Unlike fingerprinting, the forensic application was apparent – and, indeed, seemed the principal use of the new technique – from the outset. "DNA fingerprinting," as it was briefly called, could glean identity from genetic material, such as blood or semen. The method was applied to a serial rape-murder case in the vicinity of Jeffreys's institution, the University of Leicester, and it quickly became apparent that there were many crime scenes where biological material was present but fingerprints were absent.

"DNA profiling," as it was soon renamed, spread quickly, and the technology developed rapidly, with a series of new techniques replacing Jeffreys's original multi-locus probes. Of particular importance to forensics was the application of the polymerase chain reaction (PCR), developed by Kary Mullis. This allowed for the "amplification" of small amounts of genetic ("trace") material into volumes sufficient for analysis. Though the initial PCR techniques lacked discrimination, short tandem repeats (STRs) combined PCR with high discrimination. This technique remains the standard technology used today, although more powerful techniques are on the horizon.

As with fingerprints, state law enforcement agencies began assembling databases of genetic profiles. These databases were initially limited to serious sex crimes. Subsequently, they have expanded rapidly such that their eventual replacement of fingerprint databases seems probable. DNA databases have proved more controversial than fingerprint databases, largely based on the (in principle, questionable, Cole 2007) argument that DNA includes more "intimate" information than fingerprints (Krimsky and Simoncelli 2011).

Legal systems have differed sharply in reacting to this challenge. In the United States and the United Kingdom, the state's interest in public safety has generally trumped privacy concerns pertaining to genetic information. The European Court of Human Rights, however, recognized a fundamental privacy interest in both DNA and fingerprint information and this right, it ruled, demanded the exercise of moderation and proportionality in the state's retention of such information, a ruling which bound the United Kingdom as well (*S. and Marper v. United Kingdom* 2008). The European Court was particularly concerned by the potential practice of "familial searching," in which near correspondence in database searches would be used to generate investigative leads from the blood relatives of those in the database. American courts, in contrast, have generally approved of this practice.

In court, DNA evidence was challenged more forcefully than fingerprinting ever was, and this resulted in some adverse decisions. It is generally agreed that this small number of challenges and adverse rulings (e.g., *People v. Castro* 1989) stimulated improvement in the science. After a period of public controversy – often characterized as the "DNA wars" – the legal admissibility of routine DNA evidence became non-contentious. Issues remain, however, for cases involving complex mixtures of DNA from several individuals, more exotic techniques like mitochondrial and low copy number DNA profiling, and contamination and biased interpretation (Aronson 2007; Lynch et al. 2008; Kaye 2010).

Image Interpretation (e.g., of CCTV Recordings and Photographs)

Images have been used in and around legal proceedings and public inquiries (e.g., the Zapruder film at the Warren Commission) for more than a century – initially in the form of photographs, film, and video, and more recently in the guise of animation and "virtual" displays (Feigenson and Spiesel 2009). Photographic images, incorporated into many official documents (e.g. drivers' licenses, passports, and criminal records), provide important means of identification in administrative, commercial, and social settings. Accepting that images may have considerable potential as sources of information, here it is our intention to focus on their use in legal proceeding for purposes of identification.

In recent decades, especially with the advent of CCTV systems in public spaces, police, investigators, and prosecutors – and to a more limited extent criminal defendants – have begun to use images to piece together criminal activities and to identify the perpetrators of such acts (e.g., Goold 2004). In conjunction with the rapid expansion of security cameras in private spaces (e.g., houses, banks, bars, casinos, convenience stores) and/or quasi-public spaces (e.g., shopping malls, airports, and around ATMs), these systems have produced a very large number of images indexable to criminal activities – ranging from shoplifting to assault, armed robbery, and murder. The availability of images that, unlike DNA and fingerprint evidence, are susceptible to immediate interpretation by investigators (and others) has begun to change the way that criminal acts are investigated and presented to various publics whether as evidence in court or via various media in the public domain. By way of example, intelligence operatives, investigators, and general duty police officers – with limited, if any, specialist training – are often obliged to watch and interpret videos (and images) in order to advance investigations or assemble intelligence. Similarly, newspapers and television are used to enlist an expanded audience that might be capable of identifying individuals or providing investigative leads.

Ordinarily, images related to criminal offenses are admissible and the tribunal of fact is free to interpret the content, which may involve attempting to track events and/or identify those involved. The low quality of many crime-related images has meant that investigators and prosecutors often call upon others to provide supplementary interpretations and guidance. Courts, particularly in common law jurisdictions, have responded unevenly to the admission and use of opinion evidence to assist with identification. Approaches range from jurisdictions such as England – where police (including investigating officers) and a range of individuals, recognized as "face and body mapping" *experts* (with a background or training in intelligence, medical art, photography, information technology, anatomy or physical anthropology) or gait analysts (usually podiatrists), are permitted to express opinions about the content of images and positively identity persons of interest; to Australia – where police officers are generally prohibited from expressing opinions about identity, but "facial mappers" are allowed to describe similarities between a person of interest and the accused; to Canada – where police, prison and parole officers, possessing some familiarity with the accused, can positively identify them in incriminating images; to the United States – where investigators tend to rely upon quantitative (i.e., photogrammetric) approaches purporting to capture features such as the height or shoe size of offenders (Edmond et al. 2013).

The admission of images and incriminating interpretations of images appears to be an institutional attempt to accommodate the exponential proliferation of potentially probative evidence emerging in recent decades, in conjunction with a naive confidence in the abilities of lay people – whether lawyers, judges, or jurors – to manage image interpretation, and particularly identification, within the confines of the adversarial trial. Opinions about images are routinely admitted. Witnesses are sometimes limited to describing similarities rather than expressing positive opinions about identity. The value of such opinions is ordinarily left for the tribunal of fact – as a matter of weight.

This ready admission and reliance upon images in criminal investigations and prosecutions is revealing, particularly when contrasted to DNA profiles and latent fingerprints. For, there are no standardized or empirically validated approaches to the capture and storage, let alone interpretation, of images. Excepting some photogrammetric methods – with their origins outside of the forensic sciences – we have no indication of the validity or reliability of the variety of techniques and derivative opinions allowed in courts. To put this another way: we do not know how accurate the various "expert" witnesses are, or if they can consistently do what they claim.

Not insignificantly, experimental studies indicate that humans are error prone when asked to compare people in photographs or compare photographs and persons (i.e., typical courtroom scenarios) even when relying upon contemporaneous, high-quality, full-frontal images (Davis and Valentine 2008). Conspicuous exceptions are those with considerable familiarity (e.g., family members and close friends), especially where exposure covers significant periods of time and a variety of circumstances. On an average, when familiars are shown images and/or video and asked to express opinions about identity they tend to be reasonably accurate – even where the quality of images is low and the exposure short (Jenkins et al. 2011). Predictably, familiars are often reluctant to testify and incriminate. Such reticence helps to explain reliance on investigators, the legal recognition (or construction) of "expertise," and ready admission of opinions.

One factor, applicable to all identification technologies, though perhaps clearest in the forensic use of images, is the beguiling complexity of interpretation. The pervasiveness of photographs, films, and videos, along with the frequency or our exposure, encourages courts to admit images, frequently characterizing them as objective evidence of what transpired. Witnesses and juries are routinely asked or expected to approach images as mechanical reproductions of reality; their interpretation is implicitly straightforward – a relatively mundane or even intuitive activity. There are, however, serious complexities with image interpretation and with the kinds of intertextuality – e.g., the narration of images by prosecutors or expert witnesses during criminal proceedings – that may subtly (and sometimes unconsciously) shape or cue perception and interpretations. In contrast to electropherograms and latent fingerprints, the tribunal of fact is routinely encouraged to undertake its own interpretation of the primary data (i.e., the images), often in conjunction with "expert" opinion and the other evidence. Additional incriminating evidence may implicate the accused and be conveyed to those endeavoring to interpret images – whether analysts or juries – in ways that skew perception and attempts to convey difficulties with image interpretation. The interpretation of unfamiliar images, such as x-rays, fMRI scans, and aerial photographs, confirmed by numerous controversies in the long history of photography and film (e.g., Fenton's cannon balls and the detention of Rodney King), reinforces just how complex and controversial images and their interpretations can be (e.g., Morris 2011).

Key Issues

At this juncture, we turn to consider our three forensic technologies via a series of analytical frames and themes drawn from surveillance and other criminological and legal literatures.

Accommodation: Legal Responses Are Driven by Technological Advances and the Availability of Incriminating "Evidence"

Investigators, prosecutors, and judges tend to be rapid and uncritical adopters of identification technologies, typically embracing their outputs and opinions about their outputs as probative, and implicitly reliable, evidence. Admissibility standards and contests over the probative value of incriminating expert opinions have been ineffective – and at best inconsistent – at excluding unreliable opinions or exposing weaknesses in the interpretation and expression of opinions about fingerprints, DNA profiles, and images.

Courts in most jurisdictions – including those with reliability-based admissibility standards (e.g., in the United States and Canada following *Daubert v. Merrell Dow Pharmaceuticals, Inc.* 1993) – have been surprisingly uncritical and remarkably accommodating in their responses to opinion evidence derived from new technologies and older technologies subjected to new criticisms. Rather than require empirical support for interpretive techniques and expressions, courts express confidence in the efficacy of traditional legal processes and trial safeguards (see "The Limits of Legal Proceedings"). They have preferred to "grandfather" long-standing techniques (e.g., latent fingerprint comparison), accept witnesses with investigative experience (see "Reliability Discourses" and "Experience and the Reification (or Legal Construction) of Expertise") or broadly relevant qualifications – along with their self-serving claims – and they have relied primarily on the deconstructive abilities of poorly resourced defense lawyers to expose and convey limitations to lay jurors (and judges), rather than impose responsibility upon the state to demonstrate reliability (i.e., validity and reliability) through reference to empirical studies.

Moreover, once a technology or derivative interpretive technique is admitted in one jurisdiction, admission tends to follow in others. Conditions imposed on the initial admission are frequently elided in subsequent decisions, thereby facilitating technological creep (Risinger 2000).

Weapons of the Not So Weak: Resistance Is Not Always Useless

Following Scott (1985), a common theme in contemporary surveillance studies concerns the ability of various individuals and publics to resist or subvert surveillance, oversight, and identification. Concerned with visibility, traceability, and often privacy, a range of responses have emerged that, to varying degrees, enable individuals and groups to resist scrutiny or turn methods of surveillance and accountability back on to those who normally "watch" (e.g., sousveillance). The technologies we have selected are illuminating because they are more or less intrusive depending upon how information is captured and used, as well as their actual capabilities – both current and anticipated.

It is argued, for example, that DNA samples and profiles may be used to do more than determine a non-match or a match and a probability (or likelihood) of the accused being the source. DNA samples, which are often retained by government agencies, can theoretically be analyzed in ways that may facilitate discrimination – both positive and negative. Analysis might be used to identify a risk of a particular illness, such as susceptibility to breast cancer, and this might be used to prevent premature death or to discriminate through exclusion from insurance coverage. Many (negative) discriminatory uses are proscribed by constitutional rights or legislative enactments – always susceptible to the possibility that they might be reviewed, circumvented, or interpreted in ways that accommodate new technological capabilities, whether real or imagined.

Conversely, the current limits of interpretation mean that it is far from obvious that the availability of images will enable investigators, and even attentive publics and familiars, to definitively determine what took place and who was involved (e.g., the Vancouver and Tottenham riots). Depending on the cameras, systems, and sophistication of resistance, those capturing images, for whatever reason, may or may not be able to use the images as probative evidence. Techno-legal assemblages (see "Back to the Future: No Escape from Interpretation in

Complex Techno-legal Assemblages") have both technical and socio-legal constraints and limitations.

Anyone remotely conversant with criminal justice knows that some technologies can be resisted (or subverted) through relatively simple techniques such as wearing gloves (against fingerprint collection) or condoms (against DNA typing), or wearing bulky clothing, a hat or balaclava (for cameras) (e.g., Marx 2003). Such strategies of resistance, used by criminals and other citizens to disrupt comparison and identification, may or may not be effective. For, it is increasingly common for a variety of technologies (or artifacts) to be used collectively to assist with identification and proof. Where a DNA match is obtained, or similarity is observed between an image from a robbery and a suspect, an ATM transaction or mobile phone call may be used to place a person of interest in the vicinity of the offense. Moreover, certain forms of resistance (such as a particular disguise) might lead to propensity interpretations as criminal "signatures." Of course, where suspects offer a potentially innocent explanation or defense – such as consent in a sexual assault complaint – the value of identification technologies may rapidly depreciate along with the need for resistance or critique. Similarly, claims about what happened "off camera" might be useful to raise doubts around the cause of injury or culpability for serious injury resulting from a partially filmed (and therefore fragmentary) assault in a bar or street. This often shifts the hermeneutic exercise from the non-contentious issue of identity to questions about what actually transpired.

While there are means of resisting forensic and surveillance technologies, it is difficult to resist multiple identification technologies in court, especially when they are combined into a coherent narrative. Where a variety of different technological artifacts are aligned, especially if they are presented as independent and corroborative – it has become increasingly difficult to resist the cumulative results of skillfully integrated allegations.

Interestingly, not every criminal act is attended by resistance to the identification of its perpetrator. Many criminal acts are performed directly, some even deliberately, in front of cameras (and/or witnesses) or in ways that are seemingly indifferent to surveillance technologies and investigative possibilities.

Back to the Future: No Escape from Interpretation in Complex Techno-legal Assemblages

Drawing conspicuously on science and technology studies (STS and SSK) literatures, notably the work of Latour (1987), surveillance studies have embraced the idea of *surveillance assemblages* (Haggerty and Ericson 2000). That is, the combination of machines/equipment, techniques, procedures (all potentially *actants*), individuals, institutions, training, traditions and even cultures, as forming part of the assembly of constituent factors structuring the way in which interpretive activities are performed and understood. Although many of these features and dimensions tend to be omitted, or elided, in formal accounts and legal proceedings, their reintroduction provides a useful way of approaching the construction of surveillance and evidentiary artifacts and the *potential* for them to be opened up and pulled apart (or deconstructed, after Jasanoff 1995) in legal proceedings. The idea of the assemblage, combining human users, technologies, and traditions, reinforces the inescapability of human participation and the need for interpretation across the forensic sciences. Interpretation and risks of error occur across the spectrum of activities from the recognition and collection of traces, the analysis of results, to attempts to ascribe significance in the context of a case or administrative process. Human interpretation is perhaps most conspicuous on the margins: with low copy number DNA analysis, badly degraded DNA samples or results obtained at instrumental thresholds, partial or smudged latent fingerprints, and in responses to badly distorted or very old images.

Even sophisticated automated and semi-automated systems, such as DNA profiling machines or the automated fingerprint databases, require a human interface to interpret results as well as assess their potential significance to an investigation or prosecution. They also depend, as Lynch et al. (2008) explain, on mundane practices – chains of custody, collection and storage practices, training of personnel, accreditation of laboratories, the honesty of investigators and technicians as well as assumptions about the way traces may have been deposited. Rather than simply overcoming known problems, new and more powerful technologies (e.g., low copy number DNA profiling) may actually compound difficulties as smaller and smaller samples can be analyzed in ways that generate more mixed samples and introduce stochastic effects that require new forms of interpretation – and make the need for bias-reducing procedures, such as sequential unmasking, even more urgent.

Notably, fingerprint identification, as traditionally conceptualized, has lacked the "interpretative step," in which the analyst, having found correspondences between prints, assesses the significance of this finding by estimating the rarity of the corresponding features in some relevant population. Instead, assumptions about fingerprints lead analysts to equate a "match" with the positive identification of a particular individual to the exclusion of all others (i.e., individualization).

The interpretation of images represents an obvious example of the interpretative predicament. Whereas closed biometric systems relying upon images and/or other data such as fingerprints may perform reasonably well – especially on relatively small data sets, where the reference data and application are obtained in controlled conditions – when it comes to determining the identities of unknown individuals appearing in security and surveillance images, things are decidedly more complicated and not necessarily susceptible to algorithmic analysis – at least not without very significant levels of error (Introna and Wood 2004). In most investigations involving images, there are no relevant databases (e.g., *R v. Atkins* 2009). Rather, those recognized by courts as experts – because they are believed to be able to assist the tribunal of fact – are allowed to express their opinions about similarity and/or identity in the absence of any kind of systematic information about: the frequency of facial (or bodily) features; the independence of different facial features (e.g., lips and nose); and usually in the absence of methods capable of explaining how the analyst overcomes the distortion created by cheap cameras, poor positioning, bad lighting, image perspective, and the generally low levels of data retention.

Reliability Discourses

The epistemic value of many of the forensic sciences, particularly comparison sciences used for purposes of identification, was explicitly questioned after an inquiry by an eminent multidisciplinary committee (of the National Research Council, 2009) established by the National Academy of Science (USA). The committee's conclusions are revealing. Its report is not only a remarkably critical response to many comparison/identification techniques *in routine use*, but it also captures something of the light that scientifically predicated DNA techniques have cast on the other comparison "sciences."

Without wanting to suggest that DNA processes are infallible or even a "gold standard" to be faithfully emulated, it is useful to contrast the validation studies, sophisticated statistical work on population models and probability/likelihood ratio frameworks that have accompanied the largely extralegal negotiations surrounding the refinement of DNA evidence, with the credulous and accommodating reception of many other forensic science techniques – extending beyond latent fingerprint and image comparison to include, among others, ballistics, tool marks, handwriting, dental, hair, footprint, shoeprint, ear print, and tire mark comparison. This accommodating posture predates the introduction of DNA profiling and stubbornly survives its refinement in the 1990s. To varying degrees, non-DNA techniques have yet to produce, and courts yet to stipulate, the need for independent evidence of validation and reliability.

As things stand, and somewhat tautologically, investigators and experts tend to use prior admission, guilty pleas, and convictions as a basis for claims about the reliability of their techniques, interpretations, and evidence. In the absence of demonstrated validity and reliability studies (or independent evidence of proficiency), admission and convictions (however obtained) do not necessarily provide epistemic support for technologies and techniques. Moreover, the complex assemblages associated with "identification" are often obscured or "black boxed" and left to impecunious criminal defendants and their state-funded defense attorneys to consider challenging (or deconstructing) in the risky realm of the (adversarial) criminal trial.

Reviews of wrongful convictions suggest that techniques used to establish identity may sometimes fail to correct misleading investigative information and faulty assumptions, such as mistaken eyewitness accounts or confessions procured under duress (Garrett 2011). Where analysts are exposed to prejudicial information that is not required for their analysis (e.g., the suspect's prior criminal conduct and/or investigators' beliefs about culpability), these dangers are likely to be accentuated (Dror et al. 2006).

The Expression of Results

Considerable effort has been expended on the interpretation of DNA matches and the way results are expressed in reports and in courts. The results of DNA profiling are now routinely expressed in probabilistic terms based on widely accepted population models and statistics. Prominent, scientifically based, refinements to the reporting of DNA matches have created problems for other comparison sciences, including the interpretation of latent fingerprints and images.

Recently, in response to admissibility challenges to opinions based on latent fingerprints and images, several courts have required witnesses to qualify (i.e., read down) their results, sometimes restricting them to descriptions of apparent similarities (so-called splitting). This intervention is revealing. It reinforces the absence of validation and reliability studies able to support the historical form of expression (e.g., individualization and no errors for latent fingerprints) and exposes the willingness of these witnesses, legally recognized as experts, to express strong forms of incriminating opinion in their absence. It also evidences commitment to the admission of opinion evidence and naïve confidence in the ability of the trial to identify and convey limitations.

Curiously, it is highly unusual for forensic scientists to provide an indication of the error rate with the expression of results from most of these techno-legal assemblages. Rather, it seems to be the responsibility of those challenging "identification" evidence to retrospectively identity errors in the collection, storage, analysis, interpretation, and the expression of results.

Experience and the Reification (or Legal Construction) of Expertise

Courts have been responsible, to a considerable degree, through the legitimacy conferred by admission, along with the reliance placed on incriminating expert opinion evidence in supporting convictions (and upholding appeals), for recognizing and indeed reifying some forms of experience and/as expertise. Instead of requiring evidence of validity and reliability, courts have tended to accept weaker surrogates such as formal qualifications and experience. Many courts, including those in jurisdictions with formal reliability standards purportedly governing the admission of expert opinion evidence, have tended to place great store on formal qualifications and, particularly, experience along with earlier accommodating admissibility jurisprudence when approaching the admissibility of forensic science and medicine evidence (Edmond et al. 2013).

Rather than require evidence of ability and accuracy, in the form of published experimental studies, judges have preferred to look to practical experience and institutional affiliations (e.g., with the FBI, the FSS, and RCMP). While we accept that experience is an important dimension of expertise, we would contend that in most cases, and certainly with the kinds of technologies discussed in this entry, that experience alone (or even in conjunction with past "successes" and supplementary incriminating evidence – a strong case) cannot replace the need for demonstrable evidence of reliability. Experience is useful once the value of a technique has been demonstrated empirically.

In consequence, legal institutions have played an important if under-appreciated role in the construction and social legitimation (or "certification") of forensic science and medical evidence as "knowledge" through admission and allowing reliance to be placed upon techniques and opinions of unknown evidentiary value. Through their failure to impose credible restrictions on admission or adequately explore and convey limitations, courts – and lawyers, in particular – have facilitated admission and encouraged reliance upon techniques that were (or are) not adequately understood, nor demonstrably reliable. Somewhat iatrogenically, this accommodating posture may have actually inhibited study and testing. Disinterest in reliability has undoubtedly contaminated intelligence and investigative practice, actively contributed to wrongful convictions, and allowed serious offenders to remain undetected, free to re-offend (Garrett 2011).

The Limits of Legal Proceedings

In criminal justice systems, trial courts are conventionally characterized as the appropriate forum to contest and assess expert opinion evidence. They do not generally perform well in these capacities (Edmond and Roach 2011). Indeed, one of the reasons forensic scientists (and forensic medicine and dentistry) have been insufficiently attentive to the reliability of identification technologies and their various uses is that those managing the admission and presentation of evidence have been overconfident and accommodating because of their faith in the ability of legal processes to identify and convey limitations to lay tribunals of fact and courts of appeal. Such accommodating, and empirically indifferent, responses are difficult to reconcile with the expressed goal of "doing justice in the pursuit of truth" (Ho 2008).

Empirical evidence suggests that courts and jurisdictions vary dramatically in their ability to respond to forensic science and medicine evidence, and that individually and cumulatively trial safeguards and protections – such as prosecutorial restraint, (reliability-based) admissibility standards for expert opinion evidence, cross-examination, rebuttal experts, restrictions on the expression of results, directions and warnings to juries, and scope for appeals – are not particularly effective means of addressing problems with expert opinion (Edmond and San Roque 2012). Their limited value is compounded by under-resourced defense lawyers, the state's (near) monopoly of many types of "expert" evidence and experience (e.g., latent fingerprint analysts), excessive trust in police and investigative agencies, the complexity of many forms of statistical and probabilistic reasoning, and requiring lay persons to assess expert opinion evidence in the context of a criminal trial. There are relatively few well-resourced and technically sophisticated challenges to forensic science techniques and derivative opinions, and correspondingly few incentives for police and forensic scientists to study or improve their techniques.

Intelligence, Investigations, Prosecutions, Evidence, and Proof

In this entry, we contend that many of the technologies used routinely for surveillance, in investigations and as evidence in criminal proceedings, are not as unproblematic as routinely represented by investigative agencies, prosecutors, judges, and the media (e.g., CSI). Among commentators, considerable attention has been directed to the *frontstage*: of privacy concerns; and the ways surveillance technologies are adopted, used, and understood by institutions, operators, investigators, and various publics. Typically, less attention is dedicated to the *backstage* (or inside the "black box"): the value of the technologies and their evidentiary products, the capabilities of legal institutions and personnel, or the models of science, technology, and expertise embodied in the ways they are represented.

Following from the discussion of the epistemic value, particularly the uncertain validity and reliability of many "identification" techniques, we have to wonder about the use of terms such as "intelligence" and "evidence." Such terms are question begging: requiring further discussion about whether claims can be sustained and the kinds of models of science, technology, and expertise that might be used to support and understand their use in particular contexts and for specific uses. Without wanting to dismiss some kind of epistemic continuum, in practice, it will often be difficult to know precisely where techniques of interpretation might sit. In theory, there seems to be a spectrum from *uncertainty* (or ignorance) to *certainty*, with "intelligence," "evidence," and "proof," each respectively situated further along the spectrum toward "certainty." For the criminal justice system, there are risks, and a corresponding need for caution, if the probative value of the products of surveillance technologies is low or unknown – located toward the "uncertainty" end of the spectrum. This may require a protective attitude, formal protections, and possibly exclusion (or nonuse), particularly if technologies and derivative opinions are not demonstrably reliable.

Regardless of the nomenclature, there may be differences in the suitability of particular techniques, of varying degrees of reliability, to investigations as opposed to trials and criminal proof. Because of a range of explicit commitments and obligations – embodied in rules and procedures as well as constitutional and human rights protections – courts would seem to be constrained in the types of techniques and interpretations they might admit toward proof in criminal proceedings. In principle, criminal proceedings ought to be organized in a manner that embodies the desire for factual accuracy and fairness as well as an express commitment to avoid convicting the innocent.

Intelligence gathering and criminal investigations are not, for good reasons, constrained in the same way. Problems may emerge, however, when a range of speculative and error-prone technologies are available and relied upon uncritically by investigators and/or used as probative evidence in criminal proceedings.

Future Directions

This entry suggests that the value of surveillance technologies used routinely to observe, monitor, and interpret behaviors, and identify various publics and individuals, may be quite varied, necessarily moderating their suitability for investigations and criminal proof. There appears to be tremendous variation in what those using comparison techniques, whether based on DNA, latent fingerprints, or images (and extending to voice recordings, hair, handwriting, ballistics, tool marks, and so on), can credibly claim. Against expectations, we cannot assume that surveillance technologies, including those in routine use by investigative agencies and courts, are particularly effective. Interestingly, pervasive use and popular impressions reveal little about, and do not correlate with, evidentiary value.

This entry reveals complexity, and unevenness, in the ways institutions adopt, use, depend upon, and rationalize recourse to a range of identification technologies. It also suggests that even assemblages (including scientifically based technologies such as DNA profiling) unavoidably depend upon humans for the collection and interpretation of results and the attribution of meaning in particular investigative and prosecutorial settings. It seems likely that even with substantial advances in technological capabilities, whether through increasing sensitivity or whole genome sequencing, higher resolution cameras and greater data storage, or more accurate algorithms to enable automated comparisons, we will not eliminate – and probably not dramatically reduce – errors or the role of analysts, interpreters, and others. There may be technological advances, but there will be relatively fewer technological fixes. This seems to mean that the ability of the state (as well as the many corporate and private observers) to monitor, let alone effectively trace and identify citizens in a range of social, commercial, and criminal contexts, may be more constrained than is routinely suggested or believed. Scholars such as Lyon (2008) suggest that we need to develop more balanced responses to surveillance practices and capabilities – recognizing both the threats they may pose to traditions, privacy, and freedom as well as their ability to improve our lives and enhance safety and security. DNA profiling – with its diagnostic, discriminatory, and therapeutic possibilities, its ability to assist with investigations and identify criminals, as well as its exposure of wrongful convictions and the frailty of criminal justice processes – represents an example of the potential value of emerging empirically predicated technologies and their ability to produce valuable, if sometimes unexpected, insights.

This entry has adopted a somewhat unreflexive approach to technologies – particularly discussion of the validity and reliability of evidentiary products. We accept that the meaning of *reliability* is a social accomplishment, unavoidably extending beyond experts and domains of expertise to incorporate social, institutional, ideological, and practical dimensions (Bijker et al. 1987). Rather than adopting an essentialist or technocratic approach to the meaning of *reliability* or the effectiveness of a technology, we encourage analysts to shift focus to the particular use, the setting(s), institutional values, and traditions and even the way negotiations around the meaning and use of technologies ought to be conducted – in particular settings. It is our contention that in criminal justice settings, the state should be able to satisfy an onerous standard – guaranteeing basic trustworthiness – before interpretations derived from surveillance assemblages are admitted as incriminating opinions to assist with proof of identity and guilt. This is a response to the premium placed upon accuracy and fairness, but is also indexed to the frailty of many surveillance assemblages along with emerging evidence about the weakness of criminal trial safeguards. Legal systems have performed poorly in regulating and assessing identification technologies (NRC 2009).

Significantly, we would not necessarily suggest that the standards required of the state in relation to incriminating expert opinion evidence in criminal prosecutions should apply in other settings, such as civil justice, intellectual property, regulatory contexts, or even criminal investigations. Rather, models of science, technology, and expertise should be developed with sensitivity to goals, institutional values, and the risks associated with specific domains.

Related Entries

- ▶ Causes of Wrongful Convictions
- ▶ Cognitive Forensics: Human Cognition, Contextual Information, and Bias
- ▶ DNA Profiling
- ▶ DNA Technology and Police Investigations
- ▶ Expert Evidence and Criminal Trial Procedure
- ▶ Expert Witnesses: Role, Ethics, and Accountability
- ▶ Fingerprint Identification
- ▶ Forensic Facial Analysis
- ▶ Forensic Science
- ▶ Forensic Science and Criminal Investigation
- ▶ Forensic Science and Miscarriages of Justice
- ▶ History of Technology in Policing
- ▶ Legal Rules, Forensic Science and Wrongful Convictions
- ▶ Philosophy of Forensic Identification
- ▶ Scientific Evidence in Criminal Prosecutions
- ▶ Surveillance Technology and Policing

Recommended Reading and References

Aronson JD (2007) Genetic witness: science, law, and controversy in the making of DNA profiling. Rutgers University Press, New Brunswick

Bijker W, Hughes T, Pinch T (eds) (1987) The social construction of technological systems: new directions in the sociology and history of technology. MIT Press, Cambridge, MA

Campbell A (2011) The fingerprint inquiry report. Scotland AG, Edinburgh. The Scottish Government

Champod C, Lennard C, Margot P, Stoilovic M (2004) Fingerprints and other ridge skin impressions. CRC Press, Boca Raton

Cole SA (2001) Suspect identities: a history of fingerprinting and criminal identification. Harvard University Press, Cambridge, MA

Cole SA (2007) How much justice can technology afford? The impact of DNA technology on equal criminal justice. Sci Publ Pol 34(2):95–107

Cole S (2011) Splitting hairs? Evaluating 'split testimony' as an approach to the problem of forensic expert evidence. Syd Law Rev 33:459–485

Davis J, Valentine T (2008) CCTV on trial: matching video images with the defendant in the dock. Appl Cognitive Psych 23:482–505

Dror I et al (2006) Contextual information renders experts vulnerable to making erroneous identifications. Forensic Sci Int 156:74–78

Edmond G, Roach K (2011) A contextual approach to the admissibility of the state's forensic science and medical evidence. U Toronto Law J 61: 343–409

Edmond G, San Roque M (2012) The cool crucible: forensic science evidence and the frailty of the criminal trial. Curr Issues Crim Justice 24:51–68

Edmond G, Cole SA, Cunliffe E, Roberts A (2013) Admissibility compared. University of Denver Criminal Law Review (in press)

Expert Working Group on Human Factors in Latent Print Analysis (2012) Latent print examination and human factors: improving the practice through a systems approach. U.S. Department of Commerce, National Institute of Standards and Technology

Feigenson N, Spiesel C (2009) Law on display: the digital transformation of legal persuasion and judgment. New York University Press, New York

Garrett B (2011) Convicting the innocent: where criminal prosecutions go wrong. Harvard University Press, Cambridge, MA

Goold B (2004) CCTV and policing. Oxford University Press, Oxford

Groebner V (2007) Who are you? Identification, deception, and surveillance in early modern Europe (trans: Kyburz M, Peck J). Zone, New York

Haggerty KD, Ericson RV (2000) The surveillant assemblage. Br J Sociol 51(4):605–622

Ho HL (2008) A philosophy of evidence law: justice in the search for truth. Oxford University Press, Oxford

Introna LD, Wood D (2004) Picturing algorithmic surveillance: the politics of facial recognition systems. Surveill Soc 2:177–198

Jasanoff S (1995) Science at the bar. Harvard University Press, Cambridge, MA

Jenkins R, White D, Van Montfort X, Burton AM (2011) Variability in photos of the same face. Cognition 121(3):313–323

Kaye DH (2010) The double helix and the law of evidence. Harvard University Press, Cambridge

Krimsky S, Simoncelli T (2011) Genetic justice: DNA data banks, criminal investigations, and civil liberties. Columbia University Press, New York

Latour B (1987) Science in action. Harvard University Press, Cambridge, MA

Lynch M, Cole SA, McNally R, Jordan K (2008) Truth machine: the contentious history of DNA fingerprinting. University of Chicago Press, Chicago

Lyon D (2008) Surveillance studies: an overview. Polity, Cambridge

Marx G (2003) A tack in the shoe: neutralizing and resisting the new surveillance. J Soc Issues 59(2):369–390

Morris E (2011) Believing is seeing: observations on the mysteries of photography. Penguin, New York

National Research Council (2009) Strengthening forensic science in the United States: a path forward. The National Academies, Washington, DC, Community CoItNotFS

Neumann C, Evett IW, Skerrett J (2012) Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm. J Roy Stat Soc A 175(2):1–26

Risinger M (2000) Navigating expert reliability: are criminal standards of certainty being left on the dock? Albany Law Rev 64:99–152

Saks M, Koehler J (2005) The coming paradigm shift in forensic identification science. Science 309:892–895

Scott JC (1985) Weapons of the weak: everyday forms of peasant resistance. Yale University Press, New Haven

Tangen J, Thompson M, McCarthy D (2011) Identifying Fingerprint Expertise. Psychological Science 22:995–997

Torpey J, Caplan J (eds) (2001) Documenting individual identity: the development of state practices since the French revolution. Princeton University Press, Princeton

Ulery B, Hicklin RA, Buscaglia J, Roberts MA (2011) Accuracy and reliability of forensic latent fingerprint decisions. Proc Natl Acad Sci 108(19):7733–7773

Identity Theft

Lynne Vieraitis[1], Heith Copes[2] and Ivan Birch[1]
[1]The University of Texas at Dallas, Richardson, TX, USA
[2]University of Alabama at Birmingham, Birmingham, AL, USA

Overview

Each year millions of people worldwide have their identities stolen or misused. As a result of these illegal actions, businesses and consumers lose an estimated $16 billion to $50 billion per year. Information available from public and private agencies that collect data on identity theft suggests that the crime is becoming more common and more costly. Indeed, many have referred to identity theft as the "fastest-growing crime in America." Not surprisingly, citizens worldwide have become concerned about falling victim to identity theft and fraud. These fears have led to a watershed of protective legislation and have stimulated private investors to develop monitoring and preventative products for identity protection. Consumers spend billions of dollars a year on credit monitoring, fraud alerts, and information monitoring to protect their personal information from theft and fraud. In fact, credit monitoring services alone are approaching a billion dollars a year in sales (Lieber 2009) with Experian leading the market with nine million consumers paying $650–700 million annually. In addition, sales of personal paper shredders, the fastest-growing segment of the total office products market, have grown by double-digit revenues for the past few years (Fetterman 2009).

Likely due to the increased prevalence of the crime, identity theft has garnered the attention of the media whose coverage of cases has risen dramatically since 2000. Media regularly report on the latest scams used by identity thieves to steal personal information, the dangers of conducting routine transactions involving personal data, and the newest products and services designed to protect consumers from becoming victims of identity theft. Although much of this attention is directed toward educating consumers and marketing products, the media regularly present identity theft as an ever-increasing, ever-threatening problem. Despite rising media attention on identity theft over the past several years, academic research on the crime is limited. This entry provides an overview of what is known about identity theft. Specifically, it discusses the nature of identity theft, characteristics of offenders and victims, techniques of enacting the crime, and what can be done about this seemingly "unstoppable problem."

Definition of Identity Theft

The *Identity Theft Assumption and Deterrence Act* (ITADA), passed in 1998, states that identity theft occurs when a person "knowingly transfers, possesses or uses, without lawful authority, a means of identification of another person with the intent to commit, or to aid or abet, or in connection with, any unlawful activity that constitutes a violation of Federal law, or that constitutes a felony under any applicable State or local law." The term "means of identification" is defined as "any name or number that may be used, alone or in conjunction with any other information, to identify a specific individual." Yet, despite the federal statute, "there is no one universally accepted definition of [it] as the term describes a variety of illegal acts involving theft or misuse of personal information" (Bureau of Justice Statistics [BJS] 2006: 2). In fact, the term "identity theft" has been used to describe activities associated with a variety of offenses including checking account fraud, counterfeiting, forgery, auto theft using false documentation, trafficking in human beings, and terrorism.

One issue that has impeded development of a universally accepted definition of identity theft centers on the concept of "personal information." For example, if an offender steals a credit card, makes a purchase, and then discards the card, has the victim's identity been stolen? Does the use of a financial account identifier constitute identity theft? Or does identity theft only occur when an offender uses personally identifying data? An offender can use a credit card number (financial account identifier) to make unauthorized purchases or use a social security number (personally identifying data) to open a new credit card account and make purchases. Furthermore, some argue that since the crime of identity theft involves two separate components, theft and fraud, they should be labeled accordingly. In this case, "identity theft" occurs when an offender steals a victim's personal identifying information, such as a social security number, birth certificate, or driver's license. Whereas "identity fraud" occurs when an offender uses the stolen information to open credit card accounts, bank loans, or to deposit counterfeit checks and make withdrawals from the victim's bank account (Koops and Leenes 2006). Although identity fraud cannot occur without identity theft, identity theft is not always followed by identity fraud, and the two components may be committed by separate offenders.

Patterns of Identity Theft

Inconsistent definitions of identity theft across agencies and organizations tasked with collecting data make gauging the extent and patterning of the crime difficult. Contributing to the problem of data collection is the sheer number of federal agencies that play a role in identifying and investigating it. The manner in which an identity thief obtains and/or uses an individual's identity usually determines which federal agency has jurisdiction in the case. Within the United States, the list of agencies with potential jurisdiction in identity theft cases includes the US Postal Inspection Service of the US Postal Service, the Social Security Administration Office of the Inspector General, US Securities and Exchange Commission, US Secret Service, the Federal Bureau of Investigation, the U.S. Department of State, the US Department of Education Office of Inspector General, and the Internal Revenue Service.

Measuring the extent of identity theft is further complicated by the problem of underreporting. It is estimated that 40 % of all crime victims do not report their victimization to law enforcement, but the rate of reporting varies by type of victimization with more serious crimes showing a greater level of reporting than less serious crimes. For a variety of reasons, some "unknown" number of identity theft victims do not report their crimes to law enforcement authorities. The 2008 Consumer Sentinel Report indicated that only 35 % of victims who reported identity theft to the Federal Trade Commission (FTC) also reported to their local law enforcement agencies (FTC 2009). According to the National Crime Victimization Survey (NCVS), only 17 % of all victims of identity theft reported their victimization to law enforcement. Of those who did not report the theft to law enforcement, almost half reported it to a credit card company or bank (BJS 2010). Identity theft victims may see no reason to report their victimization if they do not suffer much financial harm, as when a credit card company quickly dismisses the unauthorized charges made on the victim's credit card. Other victims may be reluctant to report their victimization if they know the offender for fear of retaliation or getting the offender in trouble with law enforcement, especially if the offender is a family member of the victim. With these limitations in mind, the most reliable data on identity theft in the United States come from the FTC and more recently the NCVS.

Extent of Identity Theft

In their report, the FTC estimated that approximately eight million people were identity theft victims in 2005 and total losses were nearly $16 billion (Synovate 2007). The number of victims estimated by the 2008 Identity Theft Supplement to the NCVS suggests 11.7 million people, or 5 % of persons age 16 or older in the United States, were victims of at least one type of identity theft in a 2-year period (2006–2008), with losses around $17.3 billion (BJS 2010). However, recent findings from the NCVS suggest identity theft has increased since the agency first collected data from 5.5 % of households in 2005 to 7 % of households in 2010 (BJS 2011). The change appears to be due to the increase in the misuse or attempted misuse of existing credit card accounts. The differences in the number of victims reported by the FTC and NCVS may be attributed to several methodological factors and do not necessarily represent an increase in victimization.

It is also difficult to ascertain the financial costs of identity theft since estimates vary across the available data, but all sources show that it is an extremely costly crime. According to estimates from the FTC, in 2006, the average amount obtained by the offender was $1,882, and the average victim loss (i.e., out-of-pocket expenses) was $371 (Synovate 2007). According to data from the NCVS survey, the estimated loss for all types of identity theft reported by victims who had lost at least 1 dollar averaged $2,400 with a median loss of $430 (BJS 2010).

The financial costs to victims vary depending on the type of identity theft. Victims who experienced misuse of personal information reported an average loss of $2,829, while theft of existing credit card accounts resulted in the lowest average losses ($1,105). Victims of new account fraud suffered the greatest financial loss with an average direct financial loss of $8,110 (BJS 2010). These figures represent losses that may or may not have been covered by a financial institution, such as a credit card company. In most cases, the victims are protected by a variety of federal and state laws that limit consumers' liability in these situations. For example, the *Truth in Lending Act* limits consumer liability for unauthorized credit card charges to a maximum of $50 and the *Electronic Fund Transfer Act* provides limits on consumers' liability for unauthorized electronic fund transfers (provided the consumer notifies the applicable financial institution within a specified time period). State laws typically but not always protect consumers from losses associated with checking account fraud and loan fraud. However, victims may incur expenses from time spent resolving problems created by the theft, including closing existing accounts and opening new ones, disputing charges with merchants, and monitoring their credit reports. Surveys suggest that the time victims spend resolving issues associated with identity theft has declined over recent years as the banking and credit card industries as well as federal laws have made it easier for victims to recoup their losses. Results from the NCVS show that about 42 % of victims spend 1 day or less resolving financial and credit problems associated with their victimization (BJS 2010).

Regardless of the time spent resolving problems created by the identity theft, victims may still experience a great deal of emotional distress, including feelings of anger, helplessness, and mistrust; disturbed sleeping patterns; and feelings of a lack of security (Davis and Stevenson 2004). In addition, some victims report work, school, and relationship problems as a result of their victimization (BJS 2011).

Getting a clear assessment of the actual costs incurred by other victims of identity theft such as banks, businesses, and companies is also difficult. In most cases, the victim whose information is misused is not legally responsible for the costs of the fraudulent transaction by identity thieves since typically it is the credit card company or merchants who lose money. Recent reports indicate that the costs to victims has been declining in the past few years perhaps due to greater awareness of identity theft, legislation directed toward making it easier for victims to regain their financial identity, and/or protection services for potential victims. Given that most banks and credit card issuers provide zero-liability fraud protection to customers, victims are usually not liable for the fraudulent charges. Since identity theft reports indicate the majority of identity theft is credit card fraud, the "costs" of identity theft have been shifted from the individual to the credit card companies and banks. But the extent of these costs is unknown as companies and banks may be reluctant to disclose financial losses due to identity theft and fraud. Despite the differences in the number of victims and estimated dollar loss to victims and companies, the data suggest that identity theft is a serious problem affecting a large number of people.

Although this entry focuses on identity theft in the United States, it is a crime that is a problem worldwide. International sources of data on the crime suggest that prevalence of the crime varies considerably by country. A multinational telephone survey conducted by Unisys (2005) found that victimization rates varied across countries participating in the survey, with the lowest victimization rates reported by Germany (3 %). The highest victimization rates were reported by respondents in the United States (17 %) and the United Kingdom (11 %). France and Mexico reported victimization rates of 8 %, while Australia's rate was 7 %.

More recent estimates from an online survey conducted by a private company in Australia in 2011 show that nearly one in six (or 16 %) of

Australians have been a victim or know somebody who has been a victim of identity theft or misuse (DiMarzio Research 2011). Findings from a survey conducted in 2007 by the Australian Bureau of Statistics indicated that nearly half a million Australians or 3 % of persons aged 15 years or more were victimized by some form of identity fraud in the preceding 12 months. In the same year, the Office of the Privacy Commissioner undertook a national survey, which found 9 % of respondents, aged 18 or older, were victims of identity theft. A 2004 survey by a UK consumer magazine indicated one in four British adults had been victimized by identity theft or knew someone who had been victimized.

Research from government agencies also provides limited data on the costs of identity theft. The Home Office estimates there are over 100,000 victims of identity fraud in the UK at a cost of £1.2 billion per year. A report from the Dutch Parliament in 2002 claimed identity theft resulted in over 5 billion euros a year in financial damage, but prevalence rates were not reported. Estimates from Australia suggest that identity theft costs its citizens $1.1 billion in 2005 (Rollings 2008). In Canada, estimates from victim reports suggest that it costs them $16 million (Canadian) in 2006.

Identity Theft Offenders

It is difficult to paint a portrait of the "typical" identity thief based on the lack of reliable data. To date, few have sought to collect data directly from offenders (for exceptions, see Copes and Vieraitis 2012; Duffin et al. 2006; Gill 2007). Thus, what little is known about offenders comes from police reports and closed case files of various law enforcement agencies. The fact that clearance rates for identity theft are low means caution should be taken when interpreting the findings, and available evidence suggests that offenders are seldom detected and rarely apprehended. Using data from a large metropolitan police department in Florida, Allison et al. (2005) reported an average clearance rate of 11 % over a 3-year period. Similarly, law enforcement officials interviewed by Owens (2004) and Gayer (2003) estimated that only 10 % and 11 %, respectively, of identity theft cases received by their departments were solved. Several obstacles make the investigation of identity theft cases and the likelihood of arrests difficult. Specifically, identity theft cases can be highly complex, and the offender may have committed the theft in a different jurisdiction than where the victim resides, making it difficult to secure an arrest warrant. In addition, limited departmental resources may be directed toward the investigation of violent and drug-related offenses rather than identity thefts.

The results of the analyses conducted by Allison et al. (2005), Gordon et al. (2007), and Copes and Vieraitis (2012) demonstrate relative consistency among the demographic characteristics of identity thieves, with the exception of gender. Gordon et al. (2007) examined closed US Secret Service cases with an identity theft component from 2000 to 2006. They found that most offenders (42.5 %) were between the ages of 25 and 34 when the case was opened and another one-third fell within the 35–49 age group. In their study of data from a local law enforcement agency in Florida, Allison et al. (2005) found that offenders ranged in age from 28 to 49 with a mean age of 32.

Both law enforcement-based studies found similar patterns about race. Gordon et al. found that the majority of the offenders were black (54 %), with whites and Hispanics accounting for 38 % and 5 % of offenders, respectively. Allison et al. found that the distribution of offenders was 69 % black, 27 % white, and less than 1 % were Hispanic or Asian. The two studies differed in terms of the gender of offenders. Gordon et al. found that nearly two-thirds of the offenders were male, whereas Allison et al. found that 63 % of offenders were female.

Copes and Vieraitis' (2012) sample of 59 identity thieves included 23 men and 36 women which is consistent with the findings of Allison et al.; however, this may be attributed to Copes and Vieraitis' sampling strategy and the higher response rate from female participants. The racial makeup of their sample was 44 % white, 53 % black, and 3 % other. The makeup for the full list of located inmates from which the sample was drawn was 50 % white, 46 % black, and 4 % other. This is a higher percentage of white offenders than found by either Gordon et al. or Allison et al. Offenders in the sample ranged in age from 23 to 60 years with a mean age of 38 years. The majority of offenders were aged 25–34 (34 %) or 35–44 (32 %). Only 7 % were aged 18–24 years and 5 % were older than 55 years. The age distribution matches closely with the larger sampling pool and that found by Gordon et al. and Allison et al.

Both Copes and Vieraitis (2012) and Allison et al. (2005) included information on the offenders' employment status. Most of the offenders in Copes and Vieraitis' study had been employed at some point during their lifetimes. The diversity of jobs included day laborers, store clerks, nurses, and attorneys. At the time of their crimes, 52.5 % were employed, and a total of 35.5 % of the sample reported that their employment facilitated the identity thefts. The majority of those who used their jobs to carry out their crimes committed mortgage fraud. The results from Allison et al. indicated that 47 % were employed.

Little is known about the degree to which identity thieves specialize in their offenses. Prior arrest patterns indicated that a large portion of the offenders interviewed by Copes and Vieraitis (2012) had engaged in various types of offenses, including drug, property, and violent crimes. Yet, the majority of them claimed that they only committed identity thefts or comparable frauds (e.g., check fraud). In total, 63 % of the offenders reported prior arrests, and most were arrested for financial fraud or identity theft (44 %), but drug use/sales (19 %) and property crimes (22 %) were also relatively common. This finding is consistent with that of Gordon et al. (2007), who found that while the majority of defendants had no prior arrests, those who did have criminal histories tended to commit fraud and theft-related offenses.

Copes and Vieraitis' (2012) interviews with identity thieves yielded information that helps provide a richer and more detailed profile of the persons who commit this crime. Through interviews with offenders, they show that identity thieves are a heterogeneous group. Their family backgrounds, educational attainments, work histories, and criminal histories run the gamut from poverty to wealth, less than a high school education to graduate degrees, and from no prior arrests to incarcerations for everything from fraud to drugs. Some are embedded in "street life" and resemble the profile of a typical street offender, while others live lives similar to those of the conventional middle-class citizen and share characteristics in common with middle-class fraudsters or white-collar offenders. Copes and Vieraitis' data suggest that it is difficult to create a "profile" of identity thieves because the crime may be more "democratic" than most other types of crimes. That is, there is not a "typical" profile that emerges from the data because identity theft is committed by a diverse group of people from all walks of life. Despite their diversity, however, they are similar in their motivations for why they choose to commit identity theft.

Motivations

The primary motivation for identity theft is the desire for money (Copes and Vieraitis 2012). Offenders estimate that they can earn several thousands of dollars in as little as 1 hour. The amount of profit gained from identity theft varies depending on the type of personal information stolen, the method an offender uses to convert a victim's information into cash and/or goods, and the financial status of the victim. Although official statistics estimating the dollar losses of identity theft to be in the billions each year support the notion that money is a primary motivation for identity thieves, data from interviews with offenders provide more understanding about the underlying motivational factors. While it is not surprising that money is the most commonly given reason for engaging in identity theft, what offenders do with the money may not be as obvious. As previously mentioned, identity thieves are a very diverse group, and the ways in which they steal information, convert it into cash and goods, and how they spend their profits demonstrate their various lifestyles. Many live self-indulgent lifestyles that are characterized by the mentality of "desperate partying" found among persistent street thieves. For these offenders, proceeds gained through illicit activities are seldom saved for long-term plans or to pay bills. Instead, identity theft is a way of getting money to buy drugs, to party, and spend frivolously. For other offenders, their lifestyles reflect their position among the middle class. Many hailed from middle-class backgrounds and sought to support and sustain lifestyles of the affluent. They use the proceeds of identity theft to finance comfortable middle-class lives, including paying rent or mortgages, buying expensive vehicles, and splurging on the latest technological gadgets. Many face increasing debt, looming financial problems, and/or fear of losing streams of income, which led them to identity theft as a way to solve their immediate financial crises.

Victims

Virtually anyone can become a victim of identity theft, including newborns and the deceased. Establishing a pattern of those most likely to be victimized, however, is difficult with existing data. First, the pattern of victims that emerges from victimization data is affected by the operational definition of identity theft employed in the survey. For example, including existing credit card fraud as a type of identity theft increases not only the victimization rate and the costs of identity theft, but it also changes the face of victims. Specifically, women and racial minorities are at a higher risk of victimization (relative to percent of population) for existing account fraud and new credit card fraud than existing credit card fraud (Copes et al. 2010). Second, victimization patterns are also difficult to establish if cases in which the victim and offender know each other are underreported. That is, official statistics on identity theft may suffer from the same problem as other forms of crime (e.g., forcible rape) in that incidents in which the offender is a stranger to the victim are more likely reported and thus overrepresented in victimization statistics.

The results of several studies indicate that a similar percentage of men and women are victims of identity theft each year (Allison et al. 2005; Anderson 2006; BJS 2010; Kresse et al. 2007). The lowest rate of victimization is among persons age 65 or older, while the majority of victims are in the mid-twenties to mid-50s in age. Recent data from the NCVS, however, show a younger demographic – households headed by persons ages 16–24 have the greatest risk of victimization although the difference in percentages with those ages 25–34 was minimal (BJS 2011). In addition, most data suggest that those with incomes greater than $75,000 are at higher risk than households in lower income brackets. Data from the NCVS on race/ethnicity and identity theft victimization show that households headed by whites, Hispanics, and Asians experienced increased victimization from 2005 to 2010. Those reporting "two or more races," Asians, and whites had the highest rates of victimization among race and ethnic groups (BJS 2011).

Although several studies suggest that child identity theft is relatively rare, this form of identity theft in particular may be underreported. Child identity theft occurs when an offender uses the identifying information of a person under the age of 18 for personal gain. Data from the Consumer Sentinel Network indicate that 7 % of all cases reported to the FTC involved victims who were 19 years old or younger (FTC 2009), though the study by Kresse et al. (2007) reported only half of that (3.5 %) for persons under the age of 20. The perpetrator of child identity theft is typically a family member who has easy access to personal information. According to Pontell et al. (2008), over three-quarters of those who stole the identities of victims under the age of 18 were the parents. Similarly, the Identity Theft Resource Center (ITRC) survey data indicated that in child identity theft cases, 69 % of the offenders were one or both parents or stepparent, and 54 % of these cases began when the victim was younger than the age of 5 (ITRC 2007). However, strangers also target children because of their "clean" credit histories as well as the lengthy amount of time between the theft of the information and the discovery of the offense. Evidence suggests the cost to the child whose identity is obtained illegally does not take place until the child applies for a driver's license, enrolls in college, or applies for a loan and/or credit. This also makes it difficult to detect and report.

Lastly, deceased persons are not immune to identity theft. Law enforcement officials have historically thought deceased victims were the targets of choice for identity thieves, as information stolen from the deceased are often used to defraud businesses or to hide from law enforcement with relative ease. Identity thieves obtain information about deceased individuals in various ways, including watching obituaries, stealing death certificates, and even getting information from websites.

Victim-Offender Relationship

Research on the relationship between the victim and offender has produced mixed results regarding whether the offenders knew victims before stealing their information. To date, the available data suggest that the majority of victims do not know their offenders. The FTC reported that 84 % of victims were either unaware of the identity of the thief or did not personally know the thief (Synovate 2007). Similarly, Javelin Research and Strategy Group (2005) found that only 14 % of all identity theft crimes were committed by someone known to the victim. When the offender is known, he or she is typically a family member, relative, friend, acquaintance, co-worker, or customer/client. Three other studies also reported that the majority (60 %) of victim-offender relationships involved individuals who did not know each other (Allison et al. 2005; Gordon et al. 2007; ITRC 2007). Although this figure is lower than the figures provided by the FTC and Javelin, overall the data suggest the majority of victims do not know the thief who stole their identity.

Techniques of Identity Theft

To be successful at identity theft requires that the would-be offenders secure identifying information and convert it into goods or cash. Identity thieves have developed a number of techniques and strategies to do just this. Researchers and law enforcement agencies have collected information, primarily from victimization surveys and interviews with offenders, on the techniques identity thieves commonly employ. Although media reports suggest that hackers and other cyber criminals are responsible for the lion's share of identity theft, most official reports suggest that this is not the case. For instance, a survey by the Javelin Strategy & Research suggests that only 10 % of identity theft occurred online (Javelin 2006). This finding is also supported by Gordon et al. (2007) and Copes and Vieraitis (2012).

Acquiring Identifying Information

The first step in the successful commission of identity theft is to obtain personal information on the victim, a relatively easy thing for offenders to do. Offenders obtain this information from wallets, purses, homes, cars, offices, and business or institutions that maintain customer, employee, patient, or student records. Social security numbers provide instant access to a person's personal information and are widely used for identification and account numbers by insurance companies, universities, cable television companies, military identification, and banks. The thief may steal a wallet or purse, work at a job that affords him/her access to credit records, purchase the information from someone who does (e.g., employees who have access to credit reporting databases commonly available in auto dealerships, realtor's offices, banks, and other businesses that approve loans), or find victims by stealing mail, sorting through the trash, or by searching the Internet. Some offenders create elaborate schemes to dupe victims into revealing their personal information both on- and off-line.

The FTC (2009) data have also shed light on strategies of offending from the victim perspective. Based on data from the 43 % who knew how their information was stolen, the report suggests that offenders obtain information from people they knew personally (16 %), during a financial transaction (7 %), from a stolen wallet or purse (5 %), from a company that maintained their information (5 %), or through stolen mail (2 %). Of the 4.5 million victims responding to the

NCVS, nearly 30 % believed their identity was stolen during a purchase or other transaction, 20 % believed the information was lost or stolen from a wallet or checkbook, and 14 % thought the information was stolen from personnel or other files at an office (BJS 2010).

Other techniques have been identified such as organized rings in which a person is planted as an employee in a mortgage lender's office, doctor's office, or human resources department to more easily access information. Similarly, these groups will simply bribe insiders such as employees of banks, car dealerships, government, and hospitals to get the identifying information. Offenders have reported buying information from other offenders such as prostitutes, burglars, drug addicts, and other street hustlers. Others have obtained credit card numbers simply by shoulder surfing, which involves peering over someone's shoulders while they type in a credit card number.

Another method of obtaining personal identifying information is through the use of computer technology. This involves hacking into businesses that maintain information legitimately or through the use of phishing, which involves spam email campaigns that solicit information from would-be victims. Underground websites and forums operate that sell stolen information for relatively cheap prices (Holt and Lampke 2010).

Converting Information

Offenders can use information to acquire or produce additional identity-related documents, such as driver's licenses or state identification cards, in an attempt to gain cash or other goods. Offenders apply for credit cards in the victims' names (including major credit cards and department store credit cards), open new bank accounts and deposit counterfeit checks, withdraw money from existing bank accounts, apply for loans, open utility or phone accounts, and apply for public assistance programs.

According to the FTC, the most common type of identity theft in 2006 was credit card fraud (25 %) followed by "other" identity theft (24 %), phone or utilities fraud (16 %), bank fraud (16 %), employment-related fraud (14 %), government documents or benefits fraud (10 %), and loan fraud (5 %) (Synovate 2007). Although not directly comparable due to differences in methodology, units of analysis, and definition of identity theft, data from the NCVS indicate the most common type was unauthorized use or attempted use of existing credit card accounts, which accounted for more than half (64 %) of victimized households, followed by unauthorized use or misuse of existing accounts other than credit card accounts such as banking, savings, or PayPal (35 % of households) (BJS 2011). Data from the National Public Survey on White Collar Crime show that existing credit card fraud is more common than existing account fraud and new credit card fraud (Copes et al. 2010).

The most common strategy for converting stolen identities into cash is by applying for credit cards. Most offenders use the information to order new credit cards, but they also use the information to get the credit card agency to issue a duplicate card on an existing account. They use credit cards to buy merchandise for their own personal use, to resell the merchandise to friends and/or acquaintances, or to return the merchandise for cash. Offenders also use the checks that are routinely sent to credit card holders to deposit in the victim's account and then withdraw cash or open new accounts. Offenders have been known to apply for store credit cards such as department stores and home improvement stores. Other common strategies for converting information into cash and/or goods include producing counterfeit checks, which offenders cash at grocery stores, use to purchase merchandise and pay bills, open new bank accounts to deposit checks or to withdraw money from an existing account, and apply for and receive loans (Copes and Vieraitis 2012).

Conclusions and Future Research

It is clear that identity theft affects a sizable portion of the population, is costly in both time and money, and is difficult to detect and prosecute. It is even difficult to define. To understand the crime of identity theft and thus increase the likelihood that policy makers and law enforcement are effective in reducing this crime, more research is needed. The first step is to address the problem of data collection. Currently, information on identity theft is collected and housed in multiple databases including both private and government agencies. Information on one incident may be reported to local, state, and federal law enforcement agencies; credit reporting agencies; credit card companies; financial institutions; telecommunication companies; and others. This makes the collection and sharing of information among agencies difficult. It also creates significant barriers to developing reliable estimates of the extent of identity theft, patterns in victimization and offending, and the true costs associated with this crime. There is also little data on the processing of identity thieves including clearance rates, conviction rates, and sentencing.

Second, a number of laws have been passed to provide help to consumers and victims of identity theft and to assist law enforcement. However, the effectiveness of these laws has not yet been assessed. Although much of this legislation is relatively new, future research should evaluate the degree to which legislation is an effective strategy in reducing identity theft. Information is also needed on citizens' awareness of identity theft including their awareness of the laws designed to assist them in protecting their information as well as those designed to assist those who have been victimized. Recent research from Unisys (2005) suggests that people's concern with, awareness of, and willingness to take steps to secure their information vary widely across countries.

Third, there is very little research on identity thieves themselves. To date, only a few scholars have examined offenders directly (Copes and Vieraitis 2012; Duffin et al. 2006). Researchers should consider further developing this line of inquiry by expanding the sample of identity thieves to include active offenders and offenders convicted at federal, state, and local levels. These efforts would also be aided by the development of a centralized database for the collection of identity theft data.

Developments in technology and the globalization of commerce also make coordination of international efforts to stop identity theft increasingly important. Lawmakers in various countries have been relatively slow to respond to the emerging threat of identity theft. According to a survey completed by the Perpetuity Group, countries such as France, the United Kingdom, the Netherlands, and Germany had no specific laws directly addressing identity theft and identity fraud (Owen et al. 2006). The Netherlands, for example, had over 30 different laws encompassing identity theft or fraud. In the absence of an explicit law, prosecutors had to be creative to ensure convictions. The study also found that only the United Kingdom and France had dedicated units specializing in noncash payment fraud. Most countries, however, had multiple, uncoordinated divisions each of which was partially responsible for detecting identity theft and apprehending thieves. The Department of Justice (DOJ) and the Department of State have undertaken several measures to increase awareness of and provide assistance to law enforcement agencies in an effort to organize a global strategy against identity theft. These efforts include the expansion of a "24/7 network" which communicates with agencies from approximately 50 countries to provide electronic evidence in emergency cases and the training of lawyers and law enforcement agents in other countries.

Related Entries

- ▶ Corporate Crime Decision-Making
- ▶ Crimes of Globalization
- ▶ Fraud Victimization
- ▶ Money Laundering

Recommended Reading and References

Allison S, Schuck A, Lersch KM (2005) Exploring the crime of identity theft: prevalence, clearance rates, and victim/offender characteristics. J Crim Justice 33:19–29

Anderson KB (2006) Who are the victims of identity theft? The effect of demographics. J Public Policy Market 25:160–171

Bureau of Justice Statistics (2006) Identity theft, 2004. U.S. Government Printing Office, Washington, DC

Bureau of Justice Statistics (2010) Victims of identity theft, 2008. U.S. Government Printing Office, Washington, DC

Bureau of Justice Statistics (2011) Identity theft reported by households, 2005–2010. US Government Printing Office, Washington, DC

Copes H, Vieraitis L (2012) Identity thieves: motives and methods. Northeastern University Press, Boston

Copes H, Kerley K, Kane J, Huff R (2010) An exploratory study of identity theft victims using a national victimization survey. J Crim Just 38:1045–1052

Davis K, Stevenson A (2004) They've got your numbers. Kiplinger's Pers Financ 58:72–77

DiMarzio Research (2011) Survey data report on: identity theft concerns and experiences. Available at: http://www.ag.gov.au

Duffin M, Keats G, Gill M (2006) Identity theft in the UK: the offender and victim perspective. Perpetuity Research and Consultancy International, Leicester

Federal Trade Commission (2009) Consumer sentinel network data book. January–December 2008. Retrieved from http://www.ftc.gov/sentinel/reports/sentinel-annual-reports/sentinel-cy2008.pdf

Fetterman M (2009) Identity theft, new law about to send shredding on a tear. USA TODAY. Retrieved from http://www.usatoday.com/money/perfi/general/2005-01-14-shredder-cover_x.htm?loc=interstitialskip

Gayer J (2003) Policing privacy: law enforcement's response to identity theft. California Public Interest Research Group, Los Angeles

Gill M (2007) Learning from fraudsters: reinforcing the message. Perpetuity Research & Consulting, Leicester

Gordon GR, Rebovich D, Choo K, Gordon JB (2007) Identity fraud trends and patterns: building a data-based foundation for proactive enforcement. Center for Identity Management and Information Protection, Utica

Holt TJ, Lampke E (2010) Exploring stolen data markets on-line: products and market forces. Crim Justice Stud 23:33–50

Identity Theft Resource Center (2007) Identity theft: the aftermath, 2006. Retrieved from http://www.idtheftcenter.org/index.html

Javelin Strategy and Research (2005) 2005 Identity fraud survey report: consumer version. Javelin Strategy and Research Group, Pleasanton

Javelin Strategy and Research (2006) 2006 Identity fraud survey report: consumer version. Javelin Strategy and Research Group, Pleasanton

Koops B, Leenes R (2006) Identity theft, identity fraud and/or identity-related crime. Datenschutz Datensicherheit 30:553–556

Kresse W, Watland K, Lucki J (2007) Identity theft: findings and public policy recommendations. Final report to the Institute for Fraud Prevention. Retrieved from http://www.sxu.edu/Academic/Graham/Documents/identity_theft_study.pdf

Lieber R (2009) Credit monitoring services are a waste of money, critics say. The News & Observer. Retrieved from http://www.newsobserver.com/2009/11/03/172304/credit-monitoring-services-are.html

Owen K, Keats G, Gill M (2006) The fight against identity fraud: a brief study of the EU, the UK, France, Germany and the Netherlands. Perpetuity Research & Consultancy International, Leicester

Owens M (2004) Policing privacy: Michigan law enforcement officers on the challenges of tracking identity theft. Michigan Public Interest Research Group, Ann Arbor

Pontell HN, Brown GC, Tosouni A (2008) Stolen identities: a victim survey. In: McNally M, Newman G (eds) Perspectives on identity theft. Criminal Justice Press, New York, pp 57–86

Rollings K (2008) Counting the costs of crime in Australia: a 2005 update, vol 91, Research in public policy series. Australian Institute of Criminology, Canberra

Synovate (2007) Federal trade commission-2006 identity theft survey report. Retrieved from www.ftc.gov/os/2007/11/SynovateFinalReportIDTheft2006.pdf

Unisys (2005) Nearly half of consumers worldwide willing to switch banks for more security, says ground-breaking Unisys research. Retrieved from http://www1.unisys.com:8081/services/news_a_events/all__news/11118600.htm

Illicit Retail Marketplaces

► Spatial Perspectives on Illegal Drug Markets

ILP

► Intelligence-Led Policing

Images

► Identification Technologies in Policing and Proof

Impacts of Community-Oriented Policing

Antoinette Verhage and Paul Ponsaers
Faculty of Law, Research Group SVA, Ghent University, Ghent, Belgium

Introduction: The Deficit of Traditional Policing

The academic evaluative literature on police during the 1970s and 1980s concluded in an impressive consensus concerning the deficit of traditional police models (Bayley 1994; 1998). Summarized, following critiques can be considered as the most important: (1) The mere increase of the number of police officers is not an effective strategy to tackle crime or disorderly behavior. The quantitative assumption cannot resolve the necessary qualitative change of "how to do good policing" (Greene 1998). (2) The police cannot prevent crime, and more generally, cannot function without the help of the population, which means that the population is much more than "the eyes and ears" of the police (Rosenbaum 1998). (3) The classic tactics of traditional police models are too reactive, while they do not affect the circumstances that cause crime and disorder. (4) Police policy is frequently too broad and is applied to different problems in one and the same way ("one size fits all" – Skogan 1998). Observers advocated the need of "tailor-made responses." The need for linking different forms of policing to specific risks is probably the most energetic conclusion of police research during these decades.

COP as a Police Strategy

The most important attempt to the transformation and reform of policing during last decades was without any doubt the introduction of "Community (Oriented) Policing" (COP). The combination of focus on COP studies and the absence of ethnographers during the 1990s had as a consequence that the most influential books were studies on COP (Skogan and Harnett 1996), while this focus continued in the early years of this century (Skogan 2006). Without any doubt, this had a powerful and lasting effect on the image and the rhetorical capacity of the police (Manning and Yursza Warfield 2009).

Despite this evolution, Eck and Rosenbaum observe: *"There is no simple or commonly shared definition of community policing, either in theory or in practice"* (Eck and Rosenbaum 1994). Writing this, both authors suggest that COP over time became a container-notion. Bayley, who did a lot of research in different countries where COP was implemented, confirms this: "*Despite the benefits claimed for community policing, programmatic implementation of it has been very uneven. Although widely, almost universally, said to be important, it means different things to different people (. . .) Community policing on the ground often seems less a program than a set of aspirations wrapped in a slogan*" (Bayley 1988).

M. Moore states in this context (Moore 1994): *"Community policing is not a clear-cut concept, for it involves reforming decision-making processes and creating new cultures within police departments rather than being a specific tactical plan (. . .).*" He further states: *"Under the rubric of COP, American departments are opening small neighborhood substations, conducting surveys to identify local problems, organizing meetings and crime prevention seminars, publishing newsletters, helping form neighborhood watch groups, establishing advisory panels to inform police commanders, organizing youth activities, conducting drug education projects and media campaigns, patrolling on horses and bicycles, and working with municipal agencies to enforce health and safety regulations."*

Bennett arguments nevertheless that there appears to be some convergence of opinion in the recent literature that community policing is fundamentally a philosophy of policing or a policing paradigm, stating that *"It is generally agreed that these organizational structures and operational strategies do not in themselves represent community policing as they could exist* *equally well within the context of a different policing philosophy or policing paradigm. However, when they are implemented within a community policing paradigm they become community policing structures and strategies*" (Bennett 1994, see also Bennett 1990, 1998). Probably this conceptual blurring is to a large extent the consequence of the fact that COP is more a *prescriptive* model (on how police "ought to be") than a theory-based empirical statement (on how police "is").

Effects of Community-Oriented Policing

After more than twenty years now of promotion of this so-called police model (Ponsaers 2001) by governments, foundations, and leading universities, it is still not clear what effect this has had on police practice (Brodeur 1998b). The results of evaluative research seem to be unimpressive and in some cases nonexistent or immeasurable (Greene 2000; Fielding 1995). COP is stated to have little or no effect on police practice (Greene and Mastrofski 1988; Weisburd and Braga 2006), while, e.g., aging and years of service do (Mastrofski and Snipes 1995).

Impact on Public Opinion

Because COP tends to increase the contact between the police and the population, with a minimal use of compulsory measures, it is possible to improve the public satisfaction. But this coping strategy has only limited value, because those who are forced to stay in contact with the police (especially victims and offenders) seem to be precisely those who are mostly dissatisfied about the functioning of the police. This means that COP programs have a stronger impact on the improvement of the image than on the effectiveness of the police. This was also demonstrated in research; the most important effect of the implementation of COP was to be found in the improvement of the attitude of the population toward the quality of the service rendered by the police to the public (Brodeur 1998b). Moreover, it became clear that the improvement of the image of the police resulted in an intrinsic goal and was often misused to gain more (financial and personnel) facilities (Sacco 1998).

Impact on Crime

The most striking results were achieved in programs directed to intensive problem-solving strategies focused on so-called hot spots (Bayley 1994; Braga et al. 1999; Leigh et al. 1996). The realization of results nevertheless seemed almost impossible, while the police is confronted with problems they never can resolve (Brodeur 1998b).

The frequently used programs of "neighborhood watch" resulted in limited effects on crime. In the best case, the feelings of security and the communication between the public and the police are improving. As a result of that, the image of the police is reinforced and the job satisfaction of police officers is raised. But evaluative research demonstrated also that the majority of these initiatives were implemented in a defective way. Also it became clear that the involvement of citizens in these initiatives, also in England, was weak (Bennett 1998).

The difficulties to realize a more intensive collaboration seem to be more serious than most advocates expected. The empowerment of the public by means of a professional marketing strategy is certainly an interesting tool for the improvement of a more functional partnership between the police and the population. But the problems in mobilizing local inhabitants are often more structural of nature. In more deprived neighborhoods, the lack of collaboration by the public is often a result of feelings of despair and powerlessness, the fear for street gangs, and a deep embedded mistrust and conflict with the police (Rosenbaum 1998).

On the long run, COP would lead to a more or less important decrease of the number of emergency calls by the public (Brodeur 1998a). COP programs can have a regressive (instead of progressive) effect, while they are often directed toward the wrong target groups. Those groups within the population that are already organized succeed in using the police to their advantage, because the police feels themselves comfortable in this part of society. In spite of that, research evoked that COP, by means of locally initiated consultations, structures the active participation of the population in problem identification and prioritizing. It gives a channel for external accountability on police performance. Often it became clear that the initiatives were directed toward the wrong territoria and the target groups with the smallest needs (Skogan 1998).

Impact on Incivilities and Fear of Crime

Some authors come to the conclusion that COP can have some effect on the perception of crime by the population and on the appreciation of the quality of police care. Moreover, the feelings of insecurity seem to decrease, because of the increased visibility of the police in public space and the intensification of the interaction between the population and the police lead frequently to a better appreciation of the police service. COP seems to have an impact, when neighborhood problems are tackled and on the fear of crime. In any case, the results of COP are not worse than traditional policing in the control of crime, but the results in tackling incivilities and feelings of insecurity in the communities are better (Greene 1998).

Bayley, who did a lot of international comparative research on policing, concludes: "*We don't know if community policing works. Most of the time, a small effect can be detected, but sometimes also contradictory results. The best results can be observed in focused activities of problem oriented policing. It is not proven that citizens can act against insecurity in an effective way. Initiatives as "neighbourhood watch" don't have an effect on crime. Most of the time these initiatives work the best there were they are least needed and least where they are necessary. Nevertheless, most authors conclude that it is not the model that is failing, but in first instance the deficient implementation of it*" (Bayley 1994).

A Number of Positive Results

Pessimism should be avoided in this respect. Wycoff en Skogan (1994) state in this context that it *is* possible to bend granite. They report on the results of an evaluation of a successful internal reorganization of a police force, which has had a positive impact on the service of the police within a COP approach. One of the critical factors for successful intensive reform, they warn, is the creation of an instance outside and above the police, holding the police chief and his organization accountable for the realization of the new goals to achieve (Moore 1992).

Also Aronowitz (1997) points at positive consequences. He arguments that the approach has effects for the community: citizens are more involved in the identification of problems in the neighborhood and the relation with the police improves. Moreover, he stresses that the approach also increases the level of self-help of the citizens. They take a more active role in the maintenance of security and the quality of life in their own neighborhood. Another effect has a relation with the maintenance of legal order: not only are citizens more inclined to report to the police, but also the feelings of security improve.

One of the most prominent evaluative sources is the study *Preventing crime: What works, What doesn't, What's promising*. Sherman et al. conducted a systematic review, among others on COP (Sherman et al. 1997). The group of scholars introduce hypotheses on four levels concerning COP: (1) Neighborhood Watch programs are considered to be effective, while they encourage the level of surveillance by inhabitants of neighborhoods, which leads to the consequence that they have a deterrence effect on criminals. (2) The stream of information stemming from the communities is stimulated toward the police concerning suspects, offenders, and suspect circumstances, which leads to an increased probability to arrest offenders. This information exchange improves the problem-solving ability of the police. (3) The improvement of information from the police to the public empowers the population to protect oneself, certainly when it concerns recent trends in crime patterns and risks. (4) The credibility and legitimacy of the police is sustained and the population has more confidence in the police, which leads to more compliance to the law by the population.

Sherman et al. conclude that the results of tests concerning these hypotheses are ambiguous. Proof for the assumption that crime prevention is sustained by the increase of information from the population toward the police is not available. The second and third hypotheses are not supported by evidence either. The most important conclusion is nevertheless that there seems to be enough evidence for the fourth hypothesis concerning the legitimacy. There seems to be enough research and evaluation that sustains the presence of a strong correlation between COP on the one hand and the legitimacy of the police and law-abiding behavior by the population on the other hand (Sherman et al. 1997).

More recently, Sunshine and Tyler have concluded from their research on policing that the evaluation of police *legitimacy* is based on the perception of the way in which people are treated (Sunshine and Tyler 2003). Personal contacts between police and community are crucial – not the perception of the public with regard to how well the police handles crime.

In a study in Latin-American countries, Dammert and Malone (2006) indicate that the inclusion of the public in policing reduces public fear of crime. Although the authors are very careful in drawing this conclusion, they claim that this conclusion is very important in these "tough-on-crime-countries."

In this respect, procedural justice also influences the extent to which the public is *willing* to engage in crime prevention. The results of a study by Reisig (2007) show that citizens who judge police practices as fair and respectful are more open for participation in (property) crime prevention. This implies that it is not simply the assessment of effectiveness that influences willingness of the public to participate, but merely the way in which police practices are perceived. This conclusion could be made regardless of the level of property crime in the community. In this respect, the use of community policing as a police model can be seen as a crucial element in tackling crime. This finding was supported by empirical research in Australia, in which it was found that – when the police apply procedural justice – they are more likely to be judged as legitimate (Murphy et al. 2008). At the same time, social survey data showed that foot patrols – a typical practical element in community policing – meet the public demand and support "*the symbolic function of policing as a sign of social order*" (Wakefield 2007). Earlier studies had already showed that foot patrols lead to higher levels of citizen's satisfaction with police services *and* lower crime rates (except for robbery and burglary) (Trojanowicz 1982). A few years later, however, Pate showed that foot patrols did influence people's *perceptions* of safety and disorder problems, but did not influence the levels of reported crime (Reisig 2011).

Reisig (2010) concludes in his study on the effects of community- and problem-oriented policing, that in general, the results are encouraging. There is (though modest) evidence for the effects of these types of policing on levels of crime and disorder, and also for the perception of citizens with regard to their neighborhood (Reisig 2011). He also concludes that one of the important merits of the introduction of both community policing and problem-oriented policing is that it has instigated empirical research into police strategies and police practices, although of course a number of questions still remain.

A final and very recent (2012) impressive systematic review by Gill et al. (in progress) gathered both published and unpublished studies that focused on the effectiveness of community-oriented policing. This review based their final conclusions on 45 trials, published in 25 reports. Their findings show that community-oriented policing was associated with a statistically significant, but very small reduction in officially recorded crime. But, although the effect on crime figures seems to remain limited, findings for other intended effects, such as legitimacy, citizen satisfaction, fear of crime, and citizens' perceptions of local disorder, were very promising. The results showed a large increase in legitimacy and satisfaction with police, and a (more moderate) increase of odds of perceived social disorder and a decline in the fear of crime. The researchers hypothesized that short-term improvements in legitimacy may lead to longer-term effects on crime control, but emphasized the need for long-term research.

New Types of Policing Take Over

Some scholars, e.g., Manning, argument that the current attempt to consolidate and integrate research progress in community policing, problem-solving policing, hot spots policing, and crime analysis and crime mapping has collapsed into efforts of apparently preventive but actually active, aggressive- and arrest-oriented policing (Manning and Yursza Warfield 2009). This reading is supported tacitly by research (Weisburd and Braga 2006) containing little or no comment on the negative, unanticipated, or destructive impacts such types of policing have on order, sense of justice, and "community."

In addition to this, academic literature has also changed its focus and is increasingly moving away from the topic; in their review of police literature in 2007, Mazeika et al. conclude that although police strategies have remained the largest category in police literature, "community policing is no longer the most prevalent literature within this category" – for the first time in 6 years (Mazeika et al. 2010). Outcome-based research declined with over 32 %. The primary focus of research within the category of police strategies (which is, by the way, declining since 2005) is now research on target groups (Mazeika et al. 2010). A positive conclusion of their review was, however, that publications on policing have increased substantially, although it was not clear what the effect of this increase was on the distribution of research.

These developments have unfolded in the last twenty years. While many claims have been made, the cumulative progress in research based on deep and critical understanding of policing is modest in part because the research focus is far too narrow. It should therefore be emphasized that more research is needed for a good understanding of effects of police strategies and tactics, taking into account social processes that might influence the effects (Reisig 2011). It does, after all, affect our society in a fundamental way.

Related Entries

- ▶ Causes of Police Legitimacy
- ▶ Community Policing
- ▶ Comparing Police Systems Across the World
- ▶ Early Modern Police and Policing
- ▶ Evidence-Based Policing
- ▶ Intelligence-Led Policing
- ▶ Managing Innovation of Policing
- ▶ Measuring Police Performance
- ▶ Organizational Change and Police Legitimacy
- ▶ Police Culture
- ▶ Police Integrity
- ▶ Police Self-Legitimacy
- ▶ Police-Led Interventions to Enhance Police Legitimacy
- ▶ Policing in Developing Democracies
- ▶ Problem-Oriented Policing
- ▶ Role and Function of the Police
- ▶ Theories on Policing and Communities

Recommended Reading and References

Aronowitz A (1997) Progress in community policing. Eur J Crim Policy Res 5(4)

Bayley D (1988) Community policing: a report from the devil's advocate. In: Greene J, Mastrofski S (eds) Community policing: rhetoric or reality? Praeger, New York, pp 225–238

Bayley D (1994) Police for the future. Oxford University Press, New York

Bayley D (1998) What works in policing. Oxford University Press, New York

Bennett T (1990) Evaluating neighbourhood watch. Gower Publishing Company, Aldershot

Bennett T (1994) Community policing on the ground: developments in Britain. In: Rosenbaum DP (ed) The challenge of community policing: testing the promises. Sage, Thousand Oaks, pp 224–246

Bennett T (1998) Drugs and crime: the results of research on drug testing and interviewing arrestees. Home Office Research study 183. Home Office, London

Braga A et al (1999) Problem-oriented policing in violent crime places: a randomised controlled experiment. Criminology 37(3):549–555

Brodeur J-P (1998a) Tailor-made policing, a conceptual investigation. In: Brodeur J-P (ed) How to recognize good policing: problems and issues. Sage, Thousand Oaks

Brodeur J-P (ed) (1998b) How to recognize good policing: problems and issues. Sage, Thousand Oaks

Dammert L, Malone MFT (2006) Does it take a village? Policing strategies and fear of crime in Latin America. Latin Am Polit Soc 48(4):27–51

Eck J, Rosenbaum D (1994) Effectiveness, equity, and efficiency in community policing. In: Rosenbaum D (ed) The challenge of community policing – testing the promises. Sage, Thousand Oaks

Fielding N (1995) Community policing. Clarendon criminology series. Oxford University Press, Oxford, 229 pp

Gill CE, Weisburd D, Bennett T, Vitter Z, Telep CW (in progress) Community-oriented policing to reduce crime, disorder, and fear and increase legitimacy and citizen satisfaction in neighborhoods. Campbell Syst Rev

Greene J (1998) Evaluating planned change strategies in modern Law enforcement, implementing community-based policing. In: Brodeur J-P (ed) How to recognize good policing: problems and issues. Sage, Thousand Oaks

Greene J (2000) Community policing in America. In: Horney J (ed) Criminal justice, vol 3. Office of Criminal Justice Programs, Washington, DC, pp 299–370

Greene J, Mastrofski S (1988) Community policing: rhetoric or reality? Praeger, New York

Leigh A, Read T, Tilley N (1996) Problem-oriented policing. Crime detection and prevention series, Paper 75, Home Office

Manning P, Yursza Warfield G (2009) Police research in the United States: an overview and an analysis. In: Ponsaers P, Tange C, Van Outrive L (eds) Insights on police – Quarter of a century research on police in Europe and the anglo-saxon world. Brussels, Bruylant

Mastrofski S, Snipes J (1995) Law enforcement in the time of community policing. Criminology 33:539–563

Mazeika D, Bartholomew B et al (2010) Trends in police research: a cross-sectional analysis of the 2000–2007 literature. Police Pract Res: Int J 11(6):520–547

Moore M (1992) Problem-solving and community policing. In: Tonry M, Morris N (eds) Modern policing, crime and justice: a review of research. University of Chicago Press, Chicago

Moore M (1994) Research synthesis and policy implications. In: Rosenbaum D (ed) The challenge of community policing – testing the promises. Sage, Thousand Oaks

Murphy K, Hinds L et al (2008) Encouraging public cooperation and support for police. Pol Soc 18(2):136–155

Ponsaers P (2001) Reading about ‘community (oriented) policing’ and police models. Polic: Int J Police Strat Manag 24:470–496

Reisig MD (2007) Continuing the discussion on community policing, Issue 2. Procedural justice and community policing—what shapes Residents’ willingness to participate in crime prevention programs? Policing 1(3):356–369

Reisig MD (2010) Community and problem-oriented policing. Crime Just Int 39(1):1–53

Reisig M (2011) Community and problem-oriented policing. In: Tonry M (ed) The Oxford handbook of crime and criminal justice. Oxford University Press, New York, pp 538–576

Rosenbaum D (1998) The changing role of the police, assessing the current transition to community policing. In: Brodeur J-P (ed) How to recognize good policing: problems and issues. Sage, Thousand Oaks

Sacco V (1998) Evaluating satisfaction. In: Brodeur J-P (ed) How to recognize good policing: problems and issues. Sage, Thousand Oaks

Sherman LW, Gottfredson D, MacKenzie D, Eck J, Reuter P, Bushway S (1997) Preventing crime: what works, what doesn’t, what’s promising. http://www.ncjrs.org/works

Skogan W (1998) Community participation and community policing. In: Brodeur J-P (ed) How to recognize good policing: problems and issues. Sage, Thousand Oaks

Skogan W (2006) Police and community in Chicago: a tale of three cities. Oxford University Press, New York

Skogan W, Harnett S (1996) Community policing. Oxford University Press, New York

Sunshine J, Tyler TR (2003) The role of procedural justice and legitimacy in shaping public support for policing. Law Soc Rev 37:555–589

Trojanowicz R (1982) Evaluation of the neighborhood foot patrol program in flint. Michigan State University, School of Criminal Justice, Michigan

Wakefield A (2007) Continuing the discussion on community policing, issue 2. Carry on constable? Revaluing foot patrol. Policing 1(3):342–355

Weisburd D, Braga A (eds) (2006) Police innovation. Cambridge University Press, Cambridge

Wycoff M, Skogan W (1994) Community policing in Madison – an analysis of implementation and impact. In: Rosenbaum D (ed) The challenge of community policing – testing the promises. Sage, Thousand Oaks

Impartiality and Fairness in Judging

► Evolving Judicial Roles

Imprisonment

► History of Corrections

Imprisonment and Cardiac Risk

Emily A. Wang
Yale School of Medicine, New Haven, CT, USA

Synonyms

Epidemiology of cardiovascular disease; Incarceration; Racial disparities

Overview

More than 13 million individuals have a history of incarceration in the USA. In spite of their young age, current and former inmates are at increased risk for developing and dying from cardiovascular disease. In some US studies, current and former inmates have higher rates of hypertension, smoking, diabetes, and end-stage heart disease, compared to the general population and are thus more likely to die from cardiovascular disease.

Why is this the case? There are a number of plausible reasons. Current and former inmates are more likely to be ethnic minorities and poor. Between 50 % and 80 % of US inmates have a history of substance abuse of cardiotoxic agents, like alcohol, cocaine, and amphetamines. Many individuals do not exercise or eat a healthy diet in prison or on the street. And most do not have a primary care provider in the community. The focus of this entry is to delineate (1) the epidemiology of cardiovascular risk factors and disease among individuals with a history of incarceration, (2) to discuss mechanisms for this increased risk, and (3) to describe potential interventions that reduce the risk of disease in this at-risk population. Improving cardiovascular health of individuals with a history of incarceration may lower disease burden in the USA and reduce racial disparities in cardiovascular disease morbidity and mortality.

Fundamentals of Correctional Health Care

Incarceration is Common and Costly

The prison population has tripled in the past 20 years, and the USA now incarcerates more people per capita than any other nation. Worldwide, imprisonment ranges from 30 people per 100,000/year in India, 119 in China, 148 in the United Kingdom, 628 in Russia, to 750 in the USA (Wamsley 2006). Currently, 2.3 million individuals are incarcerated with an additional 4.6 million on parole or probation, such that one out of every 31 Americans is under the supervision of the criminal justice system (The Pew Center on the States 2009). An additional five million Americans are estimated to have a history of incarceration but are no longer under the criminal justice system jurisdiction (Wilper et al. 2009). In 2008, corrections was the fastest expanding segment of state budgets, with expenditures totaling $44 billion annually. Medical care is one of the principal drivers of cost, totaling $9.9 billion per year (Spaulding et al. 2011). By conservative estimates incarcerating individuals with chronic medical conditions like cardiovascular risk factors and disease by conservative estimates costs two times more than the average inmate.

Definition of Prison and Jail

Having a history of incarceration can range from a brief stay in jail (facilities that typically house those who are awaiting adjudication or serving sentences of less than 1 year) to longer sentences in prison (which house those who have been sentenced to more than 1 year). The median length of stay in jail is less than 7 days and in prison, 2 years. Despite the range of exposure to the correctional system, there are shared experiences that uniquely define the incarcerated population's health. To start, prisons and jails are one of the only places in the USA where health care is guaranteed by law. In 1976, the Supreme Court ruled in *Estelle v. Gamble* that it was "cruel and unusual punishment" not to provide basic health care in correctional facilities. Following the ruling, prisons and jails were mandated to provide acute care services. But as the prison population demographics have changed, health care services have adapted to provide chronic medical care. Between 1995 and 2010, the number of state and federal prisoners aged 55 or older nearly quadrupled (increasing 282 %), while the number of all prisoners grew by less than half (increasing 42 %). There are now 124,400 prisoners aged 55 or older. Many correctional facilities provide individuals their first access as adults to preventive and chronic medical care. An estimated 40 % of inmates are diagnosed with a chronic disease while incarcerated, and 80 % report seeing a medical provider while incarcerated. The correctional health care setting likely influences how ever-incarcerated patients understand their chronic medical condition and how to manage their disease condition.

Quality of Correctional Health Care

The quality of chronic medical care in correctional health care settings is variable, given limited state budgets or profit motives in privatized prisons. Unlike most free world health systems, which undergo annual quality measurement and public reporting of results, the quality of care within prisons and jails is not subject to government oversight or standardization. Standards of correctional health care are evolving for chronic conditions like hypertension, diabetes, and cardiovascular disease. National organizations such as the National Commission on Correctional Health Care (NCCHC) have established guidelines for accrediting prisons and jails, but it is voluntary whether prisons and jails agree to participate. Further, while the NCCHC and organizations like the American Diabetes Association have established screening, diagnosis, and treatment recommendations in prisons and jails, these have not been standardized in jails or prisons. Some prison systems, including Texas, Missouri, and California, have adapted free world measures of quality and have implemented quality improvement projects in the state prison systems (Damberg et al. 2011). For instance, California adopted 79 measures for quality of care, seven of which focus on cardiovascular risk factors and disease. Data on the performance of these prison systems are not publically available. Acute care for chronic medical conditions is typically contracted to outside community hospitals. Only one study has evaluated the quality of care for patients hospitalized with chest pain and found that on average inmate patients with heart disease stay in the hospital longer and receive treatment sooner compared to noninmate patients, indicating that inmates do not receive worse quality hospital care compared to noninmates (Winter 2011).

Transitions Between Correctional and Community Health Systems

Discharge planning does not fall under the constitutional guarantee for health care and nor does health care post-release. Among 33 states surveyed in 2004, all prisons offered some form of discharge planning to a limited number of patients and in most cases only those with HIV disease (Flanagan 2004). Patients with diabetes, hypertension, or cardiovascular disease are often released without medications or a follow-up appointment in the community (Mallik-Kane and Visher 2008). Even when provided medications upon release, many do not obtain them. Recently released inmates are less likely to have a primary care physician and disproportionately use the emergency department for health care compared to the general population. In one study, 90 % of individuals released from jail were uninsured or lacked financial resources to pay for their medical care (Conklin et al. 2000). Inmates with health problems may have a harder time returning to the community, as they are additionally confronted with the task of managing their health problems, obtaining health care, and keeping up with medications or appointments, while meeting their basic needs. Many return to the community without financial resources, housing, employment, or family support. And many individuals convicted of drug felonies are released from prison and prohibited from accessing safety net services, including food stamps, public housing, or federal grants for education.

Epidemiology of Cardiovascular Risk Factors and Disease

Cardiovascular Disease Mortality

Cardiovascular disease is one of the top causes of death of inmates in the incarcerated population. According to the Bureau of Justice Statistics, during the period of 2001–2004, 12,129 state inmates deaths were reported and 27 % of these deaths were attributed to heart disease (Mumola 2007). However, in spite of being a leading cause of death, the risk for death from cardiovascular disease is not higher than what would be expected if they were living in the free world (Spaulding et al. 2011). In contrast, recently released inmates have an increased risk for mortality compared to the general population (Binswanger et al. 2007; Rosen et al. 2008; Spaulding et al. 2011). In a retrospective study of 30,000 individuals released from Washington state prison, recently released inmates had a greater risk of all cause and cardiovascular disease mortality than the general population in the year following release (adjusted HRs 2.1, 95 % CI 1.7–2.6) (Binswanger et al. 2007). Evidence to date suggests that white inmates have an increased risk of dying compared to the general population, whereas blacks do not. A retrospective study of individuals released from North Carolina prison between 1985 and 2000 found that white ex-prisoners had 30 % greater than expected rate of death from diabetes and cardiovascular disease compared to the general population, although no difference was detected among black former inmates (Rosen et al. 2008). A study from Georgia, in which two-thirds of the prison population was black, also found no increased risk of cardiovascular death upon release (Spaulding et al. 2011). Importantly, there have been no national or prospective studies looking at the association between incarceration and cardiovascular disease mortality.

Incarceration Is Associated with Increased Cardiovascular Disease Risk Factors

In US studies, current and former inmates have higher rates of hypertension (Wang et al. 2009), smoking (Cropsey et al. 2008), and left ventricular hypertrophy (Wang et al. 2009) than the general population. Analyses comparing diabetes in patients with and without a history of incarceration have yielded conflicting results (Leddy et al. 2009). Other studies have shown higher rates of cardiovascular disease risk factors among current and former inmates, but are based on extrapolations from noninstitutionalized Americans (2002), data from single institutions (Baillargeon et al. 2000), or limited to self-reported health outcomes, (Wilper et al. 2009) which may be subject to detection bias given that health care is constitutionally guaranteed in prison. In spite of these limitations, current and former inmates appear to have increased rates of cardiovascular risk factors, including hypertension, diabetes, and smoking.

Mechanisms for Increased Cardiovascular Disease Risk

Despite the evidence supporting the association between incarceration and cardiovascular risk factors and disease, it is not known why this association might exist. There are several plausible biological, behavioral, and psychosocial factors through which incarceration may lead to increased cardiovascular disease morbidity and mortality. No study has determined which or what combination of these potential mechanisms leads to poor health outcomes (Fig. 1).

Incarceration May Be Associated with Dysregulated Stress Responses

As early as 1971, psychiatrist George Engel published a paper reconstructing the events in

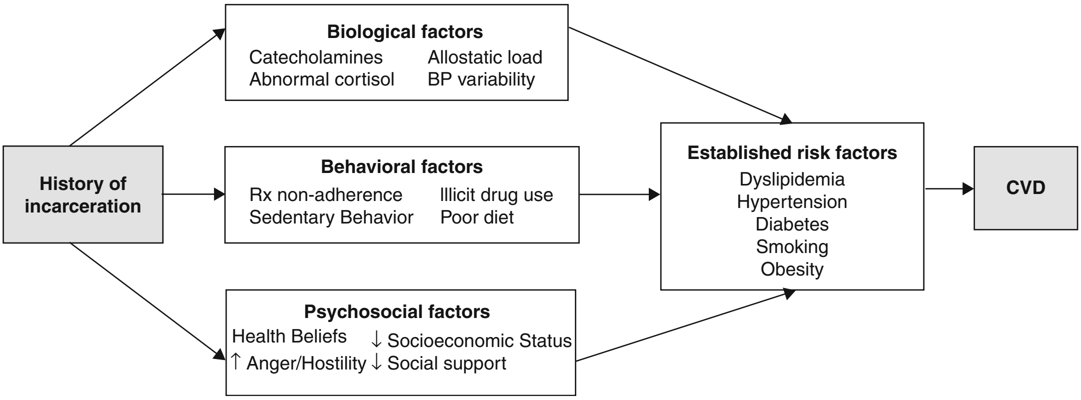
History of incarceration
Biological factors
Catecholamines
Allostatic load
Abnormal cortisol
BP variability
Behavioral factors
Rx non-adherence
Illicit drug use
Sedentary Behavior
Poor diet
Psychosocial factors
Health Beliefs
↓ Socioeconomic Status
↑ Anger/Hostility
↓ Social support
Established risk factors
Dyslipidemia
Hypertension
Diabetes
Smoking
Obesity
CVD

Imprisonment and Cardiac Risk, Fig. 1 Conceptual framework the hours before 170 people died and described the sudden death of a gentleman just released from prison as evidence that major stressful life events can lead to sudden cardiac death (Engel 1971). While researchers speculate that current and former inmates may have dysregulated stress mechanisms leading to increased risk for poor health outcomes (Massoglia 2008), no studies have examined the association between incarceration and various markers of stress – dysfunction of the hypothalamus-pituitary-adrenal axis, increased catecholamines, blood pressure reactivity, or physiological responses to chronic stresses, as measured by allostatic load. Alterations in these pathways may lead to cardiovascular disease, as elevated catecholamines have direct effects on the heart, blood vessels, and platelets. Catecholamines and cortisol have been implicated in the development of heart failure and cardiac ischemia. Blood pressure reactivity is associated with hypertension (Matthews et al. 2006). And allostatic load, a measure of cumulative stress burden, has been found to be a mediator for social disparities in cardiovascular disease (Sabbah et al. 2008). However, while life post release is stressful and may lead to dysregulated stress responses, this has not been studied.

Incarceration Is Associated with Behavioral Risk Factors That May Increase Cardiovascular Disease Risk

Higher rates of alcohol abuse and illicit drug use in patients with a history of incarceration are well documented and could contribute to increased cardiovascular risk. Injection drug use is associated with contracting HIV and hepatitis C, both of which are independently associated with developing cardiovascular disease. Whether substance abuse and its medical comorbidities contribute to the observed increased cardiovascular disease in this population is unknown. Epidemiologic studies looking at the association between incarceration and cardiovascular disease risk factors, either include measures of a single substance of abuse (Binswanger et al. 2009), do not explore it at all (Wilper et al. 2009), or find that it does not contribute (Wang et al. 2009).

Similarly understudied are the self-management strategies of current and former inmates – their physical activity, diet, pharmacologic adherence behaviors, and how the correctional system environment and its health care system might affect this. The average prison inmate spends an average of 2 years incarcerated, during which time there are limited opportunities for exercise, choice of diet, and self-management of medications. Limited studies have demonstrated that prison diet can either positively or negatively influence cardiovascular risk (Eves and Gesch 2003). Prison diets (inclusive of inconsistent timing and poor availability of sugar-free or fat-free alternatives) have been identified as a source of cardiovascular-disease-related complications in some prisons. Even fewer studies examine the association between physical activity and cardiovascular disease risk factors among incarcerated patients. These studies have found that being in prison is associated with a decrease in physical activity among women (Young et al. 2005) but an increase among men (Leddy et al. 2009). Studies examining body mass index during incarceration have found that inmates who serve shorter sentences had a nonsignificant trend for a reduction in body mass index while those who serve longer sentences have significant increases (Leddy et al. 2009). A study of older inmates reveals that being incarcerated can facilitate developing skills in managing one's cardiovascular risk factors; however, these skills may not be translatable to managing one's disease in the community (Loeb et al. 2007). In most correctional settings, patients do not manage their own medications. They do not pick up their medications at a pharmacy, dispense their own medications (for example, inject insulin or learn which medications to take at mealtime), nor develop the skills to manage complications of chronic medical conditions, including using a glucometer or sleep apnea machine. These are all skills that are essential to managing chronic conditions upon release to the community. No studies detail how best to provide chronic disease self-management support in correctional settings around hypertension, diabetes, and dyslipidemia, which is a cornerstone of cardiovascular disease risk factor management (Wagner et al. 2001).

Incarceration Is Associated with Psychosocial Factors That May Increase Cardiovascular Risk

Incarceration may augment socioeconomic disadvantage, which is independently associated with the development of cardiovascular disease. Individuals released from prison and jail often face additional barriers to obtaining employment, housing, and public entitlements and encounter various restrictions of their political rights (Travis 2005), which may further impede access to healthcare and medical treatment. Current and former inmates have been found to have higher levels of anger and hostility (Cuomo et al. 2008) which are associated with significant increases in angina, myocardial infarction, and angiographic severity of cardiovascular disease (Iribarren et al. 2000). Those with a history of incarceration may suffer from social isolation (Mallik-Kane and Visher 2008), which has also been linked to higher total mortality independent of cardiovascular risk factors. A related issue is that patients receiving psychotropic medications (e.g., olanzapine) are at increased risk of diabetes and, given the high rates of mental illness among the incarcerated, psychotropic medicine use is more common than in the general community. Another plausible mechanism that has not been studied is current and former inmates' health beliefs and attitudes about cardiovascular risk factors and disease. Research in minority and vulnerable populations has demonstrated that disparate patient adherence to health-supporting behaviors (exercising, eating a healthy diet, taking medications as directed, avoiding illicit substance abuse) is an important component of racial disparities in cardiovascular outcomes. Given the disproportionate incarceration of poor minority populations, this likely plays a role in the increased risk of cardiovascular risk factors and disease of ever-incarcerated populations.

Future Directions: Interventions to Reduce Cardiovascular Disease Morbidity and Mortality

Previous studies have demonstrated that recently released inmates have an increased risk of death from cardiovascular disease and current and former inmates are likely to have increased rates of cardiovascular risk factors. However, the reasons for this are unclear, making interventions that improve this population's health difficult to design. As the correctional population continues to age and correctional health care shifts from a focus on acute conditions and infectious diseases to noncommunicable diseases, like cardiovascular disease, important first steps are to understand why current and former inmates have increased rates of cardiovascular risk factors and disease, followed by design of prison and community-based interventions. To achieve this aim, we need more comprehensive studies that include patients with a history of incarceration and the correctional health care system.

Better Studies

Currently incarcerated individuals are not included in national surveys studying cardiovascular risk factors and disease, including the National Health and Nutrition Examination Study and the Behavioral Risk Factor Surveillance System (Wang and Wildeman 2011). Also, longitudinal cohort studies funded by the National Heart, Lung, and Blood Institute designed specifically to examine the epidemiology of cardiovascular risk factors and disease in minority populations do not follow individuals from the community into the correctional system, thus leading to loss of follow-up among minority populations, in particular black men and, thus, biased estimates of risk and disease. Why is this the case? In 1978, following decades of unethical prison research, the federal government instituted a moratorium on research in correctional settings. Recently, the Institute of Medicine revisited the issue at the behest of the Department of Health and Human Services and released a report on the ethics of conducting research on prisoners. It recommended the continuation of current restrictions but suggested updates to improve prisoners' ability to participate in limited clinical studies, particularly those with minimal risk and only interventions with demonstrated safety and efficacy. Despite these recommendations, the prohibition on prison research has not been lifted. Currently, individuals who enroll in studies while free may not be followed into prisons and jails (unless the researcher seeks OHRP approval under Subpart C) and in certain jurisdictions cannot be followed even upon release. Under the current guidelines, participants are removed from the study at the time of incarceration, unless specifications are delineated in the initial institutional review board application. Given this barrier, national surveillance systems may provide flawed estimates of cardiovascular risk factors and miss opportunities to understand how incarceration and the correctional health care system affect cardiovascular risk factors and disease. National policies should be created that allow incarcerated individuals to enroll in surveillance studies or continue their participation in a study they enrolled in prior to incarceration. Ever-incarcerated individuals, prison officials, researchers, and ethicists should be included in this dialogue so that such policies are patient-centered and acknowledge the unique logistics of conducting research in prison (or jail) without impeding good science or violating research ethics.

Understanding the Patient's Experience

Self-management is the cornerstone of treatment for patients at risk for or living with cardiovascular disease. Many studies have shown that a patient's health beliefs and experiences, medication adherence, or self-care behaviors are associated with improved blood pressure, diabetes, and cholesterol control. We know that individuals who are incarcerated are disproportionately comprised of minority racial groups and individuals from low socioeconomic environments. Additionally, given the established association between racial discrimination and hypertension, the incarcerated population has the added stigma of being incarcerated. How discrimination might impact their health and health care access has not been fleshed out. The determinants of ever-incarcerated individuals' cardiovascular disease outcomes are likely multifactorial, but their knowledge, attitudes, and health beliefs likely play a significant role and are shaped by the correctional health setting. Specific attention should be paid to studying the health knowledge, attitudes, and beliefs of ever-incarcerated individuals and how prison environmental issues, including access to healthy foods and health care in prison and upon release, affect self-efficacy for managing health in prison and upon release.

Include the Prison System

Given that around 60 % of inmates return back to prison within 3 years, understanding how the correctional health care system delivers care for individuals with cardiovascular risk factors and disease and particularly transitions of care are crucial to reducing cardiovascular disease morbidity and mortality in this population. A framework to improve the delivery of chronic care in the correctional system must be applied to study and improve care in the correctional setting. One such model, the Chronic Care Model, has already been applied in California (Wagner et al. 2001). The Chronic Care Model is a patient-centered model of care that identifies the essential elements of a health care system that encourages high-quality chronic disease care. These elements include the community, the health care system, self-management support, delivery system design, decision support, and clinical information systems. Redesigning systems of community primary care to the Chronic Care Model has led to improvements in health care outcomes for individuals with cardiovascular diseases. Adapting this model and appropriate metrics for assessing care delivery in correctional settings are underway and should incorporate corrections-specific needs (including security, transfers, and lock-downs) and the realities of the state of correctional medicine including the lack of robust informational technologies (Ha and Robinson 2011). These efforts will generate hypotheses about how prison health care, particularly self- management support and discharge planning, might impact inmates' long-term control of cardiovascular risk factors and management of disease. Moreover, current initiatives to improve quality of care in the correctional system should be mandated and overseen by state and local government entities, given that taxpayers pay for care behind bars. These initiatives may lead to lower costs and improved health for inmates with cardiovascular risk factors and disease.

Conclusion

Whether incarceration is independently associated with increased cardiovascular risk factors and disease remains unknown. However, what is clear is that prisons and jails house a disproportionate number of individuals who have or are at risk for cardiovascular disease. Acknowledging this reality and making attempts to both study and improve cardiovascular health in prison may mitigate disease among inmates.

Recommended Reading and References

Baillargeon J, Black SA, Pulvino J, Dunn K (2000) The disease profile of Texas prison inmates. Ann Epidemiol 10:74–80

Binswanger IA, Stern MF, Deyo RA, Heagerty PJ, Cheadle A, Elmore JG, Koepsell TD (2007) Release from prison – a high risk of death for former inmates. N Engl J Med 356:157–165

Binswanger IA, Krueger PM, Steiner JF (2009) Prevalence of chronic medical conditions among jail and prison inmates in the United States compared with the general population. J Epidemiol Community Health 63:912–919

Conklin TJ, Lincoln T, Tuthill RW (2000) Self-reported health and prior health behaviors of newly admitted correctional inmates. Am J Public Health 90:1939–1941

Cropsey K, Eldridge G, Weaver M, Villalobos G, Stitzer M, Best A (2008) Smoking cessation intervention for female prisoners: addressing an urgent public health need. Am J Public Health 98:1894–1901

Cuomo C, Sarchiapone M, Giannantonio MD, Mancini M, Roy A (2008) Aggression, impulsivity, personality traits, and childhood trauma of prisoners with substance abuse and addiction. Am J Drug Alcohol Abuse 34:339–345

Damberg CL, Shaw R, Teleki SS, Hiatt L, Asch SM (2011) A review of quality measures used by state and federal prisons. J Correct Health Care 17:122–137

Engel GL (1971) Sudden and rapid death during psychological stress. Folklore or folk wisdom? Ann Intern Med 74:771–782

Eves A, Gesch B (2003) Food provision and the nutritional implications of food choices made by young adult males, in a young offenders' institution. J Hum Nutr Diet 16:167–179

Flanagan NA (2004) Transitional health care for offenders being released from United States prisons. Can J Nurs Res 36:38–58

Ha BC, Robinson G (2011) Chronic care model implementation in the California State Prison System. J Correct Health Care 17:173–182

Iribarren C, Sidney S, Bild DE, Liu K, Markovitz JH, Roseman JM, Matthews K (2000) Association of hostility with coronary artery calcification in young adults: the CARDIA study. Coronary artery risk development in young adults. JAMA 283:2546–2551

Leddy MA, Schulkin J, Power ML (2009) Consequences of high incarceration rate and high obesity prevalence on the prison system. J Correct Health Care 15:318–327

Loeb SJ, Steffensmeier D, Myco PM (2007) In their own words: older male prisoners' health beliefs and concerns for the future. Geriatr Nurs 28:319–329

Mallik-Kane K, Visher CA (2008) Health and prisoner reentry: how physical, mental, and substance abuse conditions shape the process of reintegration. The Urban Institute, Washington, DC

Massoglia M (2008) Incarceration as exposure: the prison, infectious disease, and other stress-related illnesses. J Health Soc Behav 49:56–71

Matthews KA, Zhu S, Tucker DC, Whooley MA (2006) Blood pressure reactivity to psychological stress and coronary calcification in the coronary artery risk development in young adults study. Hypertension 47:391–395

Mumola CJ (2007) Medical causes of deaths in state prison, 2001–2004. Bureau of Justice Statistics. Washington, DC

National Commission on Correctional Health Care (2002) The health status of soon-to-be-released inmates: a report to congress I & II. NCCHC, Chicago

Rosen DL, Schoenbach VJ, Wohl DA (2008) All-cause and cause-specific mortality among men released from state prison, 1980–2005. Am J Public Health 98:2278–2284

Sabbah W, Watt RG, Sheiham A, Tsakos G (2008) Effects of allostatic load on the social gradient in ischaemic heart disease and periodontal disease: evidence from the third national health and nutrition examination survey. J Epidemiol Community Health 62:415–420

Spaulding AC, Seals RM, McCallum VA, Perez SD, Brzozowski AK, Steenland NK (2011) Prisoner survival inside and outside of the institution: implications for health-care planning. Am J Epidemiol 173:479–487

The Pew Center on the States (2009) One in 31: the long reach of American corrections. The Pew Charitable Trusts, Washington, DC

Travis J (2005) But they all come back: facing the challenges of prisoner reentry. Urban Institute, Washington, DC

Wagner EH, Austin BT, Davis C, Hindmarsh M, Schaefer J, Bonomi A (2001) Improving chronic illness care: translating evidence into action. Health Aff (Millwood) 20:64–78

Wamsley R (2006) World prison population list. King's College, London

Wang EA, Wildeman C (2011) Studying health disparities by including incarcerated and formerly incarcerated individuals. JAMA 305:1708–1709

Wang EA, Pletcher MJ, Lin F, Vittinghoff E, Kertesz SG, Kiefe CI, Bibbins-Domingo K (2009) Incarceration, incident hypertension, and access to healthcare: findings from the coronary artery risk development in young adults (CARDIA) study. Arch Intern Med 169:687–693

Wilper AP, Woolhandler S, Boyd JW, Lasser KE, McCormick D, Bor DH, Himmelstein DU (2009) The health and health care of US prisoners: results of a nationwide survey. Am J Public Health 99:666–672

Winter SJ (2011) A comparison of acuity and treatment measures of inmate and noninmate hospital patients with a diagnosis of either heart disease or chest pain. J Natl Med Assoc 103:109–115

Young M, Waters B, Falconer T, O'Rourke P (2005) Opportunities for health promotion in the Queensland women's prison system. Aust N Z J Public Health 29:324–327

Incapacitation

Shawn D. Bushway
School of Criminal Justice, University at Albany,
State University of New York, Albany, NY, USA

Overview

Criminologists typically consider four possible mechanisms by which incarceration can prevent crime: general deterrence, specific deterrence, rehabilitation, and incapacitation.

- General deterrence focuses on the changes in behavior created by the threat of punishment. Individuals may choose not to do a particular crime in order to avoid the possibility of incarceration.
- Specific deterrence focuses on the possible change in behavior after the experience of prison. An individual who has experienced prison may update their expectations about prison and decide, on the basis of that experience, to avoid criminal activity in the future. Of course, the experience of prison may also be less harsh than the person was expecting, leading to an increase in the rate of offending over what the person would have committed absent the prison experience.
- Rehabilitation refers to changes in the underlying "criminogenic" factors that the individual brings to prison, usually through targeted treatment programs, like drug treatment, cognitive behavioral programming, or work programs. These programs (and the resulting changes in the underlying propensity to offend) will lead to reduced crime by these individuals upon release. However, the prison experience itself may also increase criminal propensity, by reducing the person's pro-social human capital or increasing the person's criminal capital, such that the person is actually more likely to offend upon release.
- Incapacitation reduces crime by literally preventing someone from committing crime through direct control during the incarceration experience. While it is not impossible to commit a crime in prison, the possibility is greatly limited by the direct control exerted by the correctional system. People who are in prison will not be able to commit crime at the same rate as they did prior to their prison experience, and these prevented crimes should be a direct benefit to society. Unlike the first three mechanisms, incapacitation does not rely on creating changes in criminal propensity. Criminal propensity does not change at all – it simply is prevented from becoming reality. This direct, obvious connection between incarceration and crime reduction is the main attraction of incapacitation. The main drawbacks are that there are no efficiencies to scale and the effect is time limited. While one prison cell could potentially deter the criminal actions of many people or rehabilitate a person for a lifetime, one prison cell

can only incapacitate one criminal at a time for the time that the person is in prison.

The benefits of incapacitation depend directly on the offending behavior of those individuals who are incarcerated. Incapacitation benefits will be larger for policies that manage to incarcerate higher rate criminals. This is particularly true given that most criminologists believe that criminal offending is highly skewed among the offending population, with a relatively small minority of all offenders responsible for the majority of all crimes. Selective incapacitation focuses on the idea that policymakers can prospectively identify the most active offenders prior to their period of peak activity and prevent a great deal of crime through "selectively" incapacitating these high-risk individuals. However, incapacitation also implies declining marginal returns. If society starts by incapacitating the highest rate and most frequent offenders, additional incarceration will generate reduced benefits from incapacitation as society incarcerates lower rate, less frequent offenders.

Estimating Incapacitation

Historically, there are two basic approaches to the study of incapacitation that fall along basic disciplinary lines. The first, based in criminology's criminal career literature, relies on self-reported data from incarcerated individuals to generate estimates of lambda, the frequency of offending. The second, based in economics, relies on aggregate data from places to estimate the impact of prison on crime rates.

Criminologists use individual estimates to generate simulated estimates of the amount of crime averted by specific imprisonment policies. For example, inmates have been asked while incarcerated to describe their offending prior to the current term of incarceration. These responses are then used to generate estimates of the annual amount of crime committed while the person is free using the criminal career approach. The criminal career approach comes from operations research and involves the use of detailed equations which identify the specific parameters that contribute to an observed offending rate. Estimates of the underlying parameters are then used to generate an estimate of the benefits of a year of incarceration in terms of the number of crimes prevented.

Initially, research using this approach reported extremely high benefits from incapacitation, on the order of almost 200 felony crimes per prisoner year. But data on self-reported crimes is highly skewed, with a few offenders reporting a great many crimes. More recent research has reached a consensus around 15–20 felony or index I crimes per year in prison, on average. However, there is no way of knowing whether a given policy change will incapacitate the average offender or someone who commits less crime on average.

There is also no way of knowing whether the person placed in prison was simply replaced by someone else who would have otherwise not committed crime. The possibility of replacement is most plausible in the case of drug crimes, where dealers could be replaced by others. The problem of replacement is similar to the problem of displacement in the case of place-based crime prevention, which attempts to reduce crime by making a "crime-generating" place incapable of generating crime (by, e.g., increasing police presence). If criminals simply go to another place, then crime is displaced, and overall crime is not reduced by this policy, even though crime goes down in a particular place. By the same token, crime is not reduced by incapacitation if a person is incarcerated and is promptly replaced by another individual who now commits the same crimes the other person would have committed absent incarceration.

One way around the problem of replacement is to focus directly on the total amount of crime committed in a place rather than on the crime committed (or not committed) by a particular person. A change in the number of crimes committed in a certain place that can be directly tied to changes in the number of people in prison will generate an estimate of the incapacitation effect that is net of replacement. This idea of linking changes in incarceration to changes in crime for places is the approach of economists. While this approach has the twin advantages of controlling for replacement while linking policy directly to the outcome of interest, it also faces numerous empirical challenges, most notably the problem of endogeneity between prison policy and crime rates.

There is a positive correlation between aggregate crime rates and incarceration rates, not because prison causes crime but because places with higher crime rates also tend to put more people in prison. Attempts to causally identify the impact of prison on crime must therefore break this link by identifying exogenous variation in the incarceration rate. Experiments can cause this variation, although it is hard to create experimental variation in incarceration.

However, "natural" experiments can sometimes create exogenous variation. For example, Steve Levitt observed that some states in the USA were forced by the courts to reduce their incarceration levels due to charges of overcrowding (Levitt 1996). Initially, at least, states under court sanction could only meet this mandate by releasing prisoners. Likewise, the Italian government routinely releases up to 35 % of its prison population through periodic "collective pardons" (Barbarino and Mastrobuonie 2012). This creates variation in the incarceration rate over time and in different places, because the pardons release varying amounts of prisoners to each Italian region. In each case, researchers have observed an immediate increase in crime rates in the places that receive prisoners relative to places that do not, a change that that can be plausibly associated with the individuals who were just released from prison who otherwise would have been incapacitated. These papers generate estimates of between 15 and 20 more felony crimes per year in prison, a number that aligns well with the individual estimates from criminology.

It may not be completely correct to attribute all of these changes to incapacitation. In each case, the deterrent threat of incarceration has also changed as has the potential for specific deterrence and rehabilitation. For example, an individual contemplating crime in a state with a court-imposed cap on incarceration might plausibly assume that his chance of incarceration is less if he was to commit a crime. Individuals who are part of collective pardons might not be able to complete rehabilitation programs or might decide that incarceration is not as bad as they had previously thought and therefore commit more crimes than they would have without the collective pardon.

Some researchers have attempted to differentiate between incapacitative effects and deterrent effects by looking at the timing of effects on crime (Kessler and Levitt 1999). For example, three strikes laws impose long prison sentences for multiple or repeat offenders. These individuals would have been incarcerated even without the three strike provisions: the difference is that they may now have a 10- or 15-year sentence instead of a 5-year sentence. In the short term, there will be no additional incapacitative benefit to these particular laws. An immediate change in the crime rates after the implementation of the law can then plausibly considered to be a deterrent rather than an incapacitative effect. Long-term effects can be plausibly attributed to both deterrence and incapacitation.

Miles and Ludwig (2007) argued that reliable and valid estimates of incapacitation are too difficult to obtain and that time would be better spent generating estimates of the aggregate effects of prison using natural experiments. Although these aggregate measures would not enable the researcher to isolate the mechanism by which crime declined, it would give policymakers clear guidance with which to conduct cost-benefit analyses of prison. In a cost-benefit analyses, the cost of incarceration is compared with the monetized benefits of preventing a certain number of crimes. The monetary costs of crime are generated using direct accounting methods, compensatory damages from civil lawsuits, and methods that attempt to put an estimate on individuals' willingness to pay to avoid victimization. From Miles and Ludwig's perspective, generating a clean estimate of the crime-reducing benefits of a policy would then provide room to concentrate on generating good estimates of the costs and benefits of such a policy.

Recent work by economist Ben Vollaard (2012) looking at the "habitual offender" policy in the Netherlands is an exemplar of this tradition. He first estimates the reduction in crime due to the policy and then conducts a cost-benefit analysis to determine its viability as a crime prevention strategy. While Vollaard argues that the observed impact is due to incapacitation of chronic drug-addicted offenders who were no longer "deterrable" or amenable to treatment, he also acknowledges that it is still at least possible that there was also deterrent value to the statute. For example, good evaluations of "three strikes" laws in California using individual-level data have plausibly identified deterrence for individuals exposed to the risk of the third strike.

Regardless of whether or not the effects identified in the paper are entirely the result of incapacitation, the paper is important because it simultaneously highlights the attraction and limitations of incapacitation policies. Vollaard's estimates are huge, suggesting that the policy prevents over 50 reported crimes per prison year. This very large effect size is actually plausible because of the policy's laser focus on very high-rate habitual offenders in a country with a very low incarceration rate. The average offender sentenced under the law had 31 prior felony convictions, almost all for property offenses. The policy is implemented only as a last resort, after the person has shown no response to treatment. This makes sense, because in the long run it will be cheaper to simply remove the "cause" of the crime problem (the drug habit). Incapacitation only works while the person is in prison, while rehabilitation could last "forever." In cases where treatment does not work, Vollaard's paper appears to show that incapacitation could be a very cost-effective policy for reducing crime.

The problem arises when policymakers get greedy. If we can stop 50 crimes by incarcerating someone for a year, why do we not incarcerate more people? The reality is that only a few people commit crime at such a high rate. In the Netherlands, a country of 16 million, only 4,000 people are even eligible for the habitual offender label in a given year, and even fewer actually receive the penalty. Nevertheless, Vollaard shows that Dutch cities that used the habitual offender law more liberally also had smaller crime reduction per prison year. The reality of a strong positive skew is impossible to avoid – the only way to incarcerate more people is to incarcerate offenders who commit fewer crimes. This leads inevitably to diminishing marginal returns from increased incarceration. In the USA, which has seen a fourfold increase in the incarceration rate over the last 40 years, researcher after researcher has shown that the impact of incarceration on crime has declined during this period (e.g., Piehl et al. 2006). Incapacitation may work, but it is not scalable without a loss of impact.

Selective Incapacitation

The second way policymakers may get greedy would be to avoid waiting until someone reached "habitual offender" status. In the Netherlands, the habitual offender statute could only be used as an option of last resort for an offender with at least 10 felony offenses. As a result, individuals are only incapacitated after they have revealed through their own behavior that they are indeed prolific. However, it is tempting to avoid waiting until someone has committed so many offenses before identifying them as a high-rate offender. Imagine how many crimes could have been prevented in the Netherlands if these individuals could have been identified and incapacitated before they were convicted more than 10 times for felony offenses.

Prospectively identifying high-rate offenders, sometimes called selective incapacitation, holds out this tantalizing prospect of dramatic reductions in crime. It also carries with it some ethical ambiguity, since individuals are essentially being incarcerated for crimes they have not yet committed, and the inevitable cost of false positives. False positives are people identified as high-rate offenders who would have stopped without any additional intervention. Because of this concern about false positives, the debate about selective incapacitation can become quite heated, and people often make broad statements questioning whether criminologists can reliably predict high-rate offenders.

A reasonable read of the now large literature on identifying chronic offending and recidivism will necessarily conclude that criminologists can prospectively identify high-rate offenders at a rate significantly (and substantially) better than chance. It is also true that the accuracy is far from perfect. The open question is not whether criminologists can predict risk – they can – but whether the accuracy is good enough to create benefits that outweigh not only the costs of incarcerating the accurately identified high-risk offenders but also the costs generated by incarcerating the false positives.

There may also be inherent limits to the ability of researchers to prospectively identify high-risk offenders in the current environment. Empirical models require data on individuals followed over many years to validate risk prediction models. This is not a big problem in places like the Netherlands, since even the prolific offenders in the Netherlands who were subject to the habitual offender law had spent very little time in prison prior to receiving the sentence enhancement. However, in the USA, these people would have been incarcerated for substantial periods of time, drastically reducing the amount of time in which their behavior could have been observed. As a result, it would have taken them longer to accumulate the same number of offenses.

Despite the presence of these challenges, the use of risk prediction in the USA's criminal justice system, for sentencing, correctional placement, probation supervision, and parole release, has exploded and shows no sign of abating. Although not all risk prediction is used for selective incapacitation, many of the explicit goals of the risk, needs, and responsivity (RNR) model that now dominates the risk prediction field are at least consistent with selective incapacitation. For example, a central tenet of the RNR approach is the identification of low-risk offenders who will not offend even without treatment or supervision. This practice is simply selective incapacitation at the other end of the distribution – why incarcerate or otherwise restrict people who are at low risk for offending.

The Future of Research on Incapacitation

For decision makers, the most important question is the extent to which incarceration reduces crime; the path generating the reduction is irrelevant for those purposes. However, incapacitation has acquired prominence primarily because of its "obviousness." It is easy to question the power of deterrence in a world where people shoot one another over Nike sneakers. It is not easy to question that locking someone up will prevent the crimes he would have committed had he not been locked up. The debate, therefore, is not about the existence of the impact, but the relative costs and benefits of the approach. Within that framework, plausible estimates of the size of incapacitation benefits serve as useful lower bounds on the potential benefits of incarceration.

As previously discussed, there have traditionally been two approaches, one which relies on inmate surveys and the criminal career approach and the other which relies on policy changes and aggregate crime rates. The first approach, which relies on complicated calculations and expensive surveys, has not been used regularly for at least 15 years, and the second approach, taken to its logical end, reaches the conclusion that the incapacitation effect cannot be isolated from all the other mechanisms by which prison reduces crime.

A third approach has emerged in the last 5 years that combines the best parts of the two methods. The main advantage of the first approach is that it uses data from individuals and allows researchers to estimate a distribution of benefits from incapacitation. The main advantage of the second approach is that it provides a useful counterfactual, comparing the crime control effects of two different policies that can be described substantively.

These new attempts to estimate the benefits of incapacitation rely on individual-level data, like the first method, but, like the second method, attempt to create a reasonable counterfactual. In this approach, individuals are identified who are incarcerated. The researcher then creates a useful counterfactual using other observations in the data. The rate of offending generated by the non-incarceration counterfactual becomes the best estimate of the amount of crime prevented through the incarceration of the first individual.

The main attraction of this latter approach is that it does not force the researchers to assume that individuals who are not incarcerated will necessarily look like those who are. Rather, individuals are explicitly identified as comparable on the basis of their observed offending behavior and other characteristics.

For example, criminologists Gary Sweeten and Robert Apel (2007) studied the self-reported offending of a group of individuals in a contemporary USA sample. In contrast to most prior research, they do not generate estimates by relying on the reports of offending before incarceration by the same respondents. Rather, they rely on self-reported data from a matched control group (using propensity scoring) of other people who are otherwise similar but not incarcerated. This is the first time that a matched sample approach has been used to generate estimates of incapacitation.

Since it is likely that those who are incarcerated are different from those who are not in unobservable ways, it is also reasonable to assume that the Apel and Sweeten estimates also underestimate the incapacitative benefits. Their approach also has the merit of being tied to a specific change in imprisonment policy, namely, increasing the numbers who are given prison as a punishment as opposed to extending the lengths of those currently incarcerated.

Economist Emily Owens (2009) fashioned an identification strategy that should control for both unobserved and observed differences between those who are incarcerated and those who are not. It also fits nicely between the two approaches, because she uses individual-level data and a natural experiment to estimate the causal effect of incapacitation for a subsample of people affected by the policy. She takes advantage of a technical change in Maryland sentencing guidelines that had a substantial effect on the sentence for a subset of sentenced offenders.

The change involved the use of juvenile records in sentencing decisions. Until 2001, these records were included in the criminal history of all individuals up to the age of 25; after 2001, the age for which juvenile histories counted was lowered to 22. Thus, some of those aged 23–25 received shorter sentences than they would have received in the earlier years. Owens estimates that this reduced the average sentence under the Maryland guideline system by about 25 % (about 9–18 months). She also found that, during this time period when they were at liberty because of the change in rules, youths sentenced after 2001 were arrested on average 2.5 times per year. Taking account of the specific offenses for which they were arrested and the ratio of recorded arrests to recorded offenses of the same type, she estimates that they were responsible for 1.5 index crimes per year. This provides a relatively precise estimate of their recorded criminal activity during a period when they would have been incarcerated under the previous rules.

Unlike Sweeten and Apel, Owens' estimate uses information from other offenders to generate a counterfactual incarceration spell, rather than a counterfactual offending rate. Then, the individual's own offending during this time period after their initial period of incarceration is used to estimate the incapacitation effect. The estimate that Owens develops of crimes averted is smaller by an order of magnitude than the consensus estimate of 15–20 crimes previously cited in the literature. The difference comes from the fact that Owens has in all likelihood captured the lower tail of the distribution by her focus on young, relatively minor offenders.

The causal modeling in Owens' paper represents strong causal identification, but it only provides an estimate for a very narrow part of the population. This trade-off between generalizability and strong internal validity is probably typical of the research that might follow in this vein.

A final trend in this area is to extend the concept of incapacitation to include other policies besides incarceration. For example, research in economics has considered the incapacitative impact of school, which keeps youths out of the community and potentially reduces property crime, and of bad weather, which keeps people off the streets. It is also possible to talk about the incapacitative effect of police and even probation officers, who can detain and otherwise obstruct individuals from committing crime by their presence and actions. House arrest and electronic monitoring, which have become increasingly common in the USA, may serve as a deterrent but may also incapacitate people by making it more difficult to engage in criminal behavior. New monitoring policies which require individuals to check in daily for drug and alcohol tests may incapacitate offenders by requiring certain behavior (showing up for breathalyzer tests) when they would otherwise be drinking. Sex offender restrictions on housing locations are also another form of incapacitation. No attempt is being made to deter or change the underlying propensity of the individuals. Rather, policies are created which will make it harder for these sex offenders to actuate their underlying propensities.

These alternative forms of incapacitation might not be as complete as imprisonment, but they may also not carry with them the costs associated with concentrating large numbers of offenders in a prison. The costs of creating such potentially violent environments are not typically considered in the average incapacitation study, which only focuses on crimes in the community. In contrast, evaluations of alternative forms of incapacitation do consider the crimes that are committed while under supervision. For example, evaluations of electronic monitoring compare the behavior of people with the monitors to the behavior of people without monitors. No assumption is made that people with monitoring do not commit crime. A realistic appraisal of these new forms of incapacitation starts with a clear understanding of what the constraints placed on an individual can and cannot prevent. This same realistic appraisal could, and perhaps should, be applied to the constraint of prison. By broadening the definition of incapacitation, researchers may also start to question the fundamental assumptions underlying standard research on incapacitation.

Recommended Reading and References

Auerhahn K (1999) Selective incapacitation and the problem of prediction. Criminology 37:703–733

Barbarino, A, Mastrobuonie G (2012) The incapacitation effect of incarceration: evidence from several Italian collective pardons. IZA Discussion paper no. 6360

Berk R (2011a) Asymmetric loss functions for forecasting in criminal justice settings. J Quant Criminol 27:107–123

Berk R (2011b) Balancing the costs of forecasting errors in parole decisions. Albany Law Rev 74:1071–1085

Bhati A (2007) Estimating the number of crimes averted by incapacitation: an information theoretic approach. J Quant Criminol 23:355–375

Blokland A, Nieuwbeerta P (2007) Selectively incapacitating frequent offenders: costs and benefits of various penal scenarios. J Quant Criminol 23:327–353

Blumstein A (1983) Selective incapacitation as a means of crime control. Am Behav Sci 27:87–108

Blumstein A, Cohen J, Nagin D (1978) Deterrence and incapacitation: estimating the effects of criminal sanctions on crime rates. National Academy of Sciences, Washington, DC

Bushway S, Smith J (2007) Sentencing using statistical treatment rules: what we don't know can hurt us. J Quant Criminol 23:377–387

Gottfredson S, Gottfredson D (1994) Behavioral prediction and the problem of incapacitation. Criminology 32:441–474

Gottfredson S, Moriarty L (2006) Statistical risk assessment: old problems and new applications. Crime and Delinq 52:178–200

Johnson R, Raphael S (2007) How much crime reduction does the marginal prisoner buy? Working paper. Goldman School of Public Policy, University of California, Berkeley

Kessler D, Levitt S (1999) Using sentence enhancements to distinguish between deterrence and incapacitation. J Law Econ 42:343–363

Kuziemko I, Levitt S (2004) An empirical analysis of imprisoning drug offenders. J Public Econ 88:2043–2066

Levitt S (1996) The effect of prison population size on crime rates: evidence from prison overcrowding litigation. Q J Econ 111:319–351

Miles T, Ludwig J (2007) The silence of the lambdas: deterring incapacitation research. J Quant Criminol 23:287–301

Owens E (2009) More time, less crime? Estimating the incapacitative effect of sentence enhancements. J Law Econ 52:551–579

Piehl A, Liedka RV, Useem B (2006) The crime-control effect of incarceration: does scale matter? Criminol Public Policy 5:245–275

Piquero A, Blumstein A (2007) Does incapacitation reduce crime? J Quant Criminol 23:267–285

Piquero A, Farrington D, Blumstein A (2003) The criminal career paradigm: background and recent developments. Crime Justice: A Rev Res 30:359–506
Rhodes W (2011) Predicting criminal recidivism: a research note. J Exp Criminol 7:57–71
Spelman W (1994) Criminal incapacitation. Plenum Press, New York
Spelman W (2000) What recent studies do (and don't) tell us about imprisonment and crime. Crime Justice 27:419–494
Sweeten G, Apel R (2007) Incapacitation: revisiting an old question with a new method and new data. J Quant Criminol 23:303–326
Vollaard B (2012) Preventing crime through selective incapacitation. Econ J 123:262–284
Zimring F (1997) Incapacitation: penal confinement and the restraint of crime. Oxford University Press, New York

Incarcerating Terrorists

Mark Hamm
Department of Criminology and Criminal Justice, Indiana State University, Terre Haute, IN, USA

Overview

An estimated 100,000 suspected terrorists are currently locked up in prisons throughout the world (Kruglanski et al. 2010). Some of the most influential terrorists are incarcerated in Western prisons, representing a significant security challenge: These prisoners have been known to conduct extensive outreach to other inmates and free-world extremists as well, instigating fear by claiming that anti-Western terrorists, including the 9/11 hijackers, should be remembered as martyrs and heroes.

A notable example is the Blind Sheikh, Omar Abdel Rahman, who has been in US custody since 1993 for his role in the New York City Landmarks plot – a plan to bomb the United Nations Building, along with the New York office of the FBI and the Lincoln and Holland Tunnels. Years ago the Sheikh issued a fatwa, calling on al-Qaeda to attack the United States should he die in an American prison, which is most likely inevitable. Not only has the Sheikh's fatwa reached worldwide audiences via the Internet, but from federal custody he has sent similar communiques to the Egyptian terrorist organization al-Gama al-Islamiyya.

Yet when it comes to the mythopoeia of modern terrorism – an essential element in understanding how extremists from different cultures create violence in similar ways – no one is more important than Ilich Ramirez Sanchez, better known as Carlos the Jackal. The first terrorist to achieve universal media fame for organizing a series of outstanding European and Middle Eastern terrorist attacks during the 1970s, Carlos played a definitive role in turning the terrorist into an incandescent celebrity, paving the way for Osama bin Laden's enlargement as a living myth at the turn of the millennium. And today, Carlos the Jackal is serving a life sentence in France's Le Sante prison where he periodically issues press releases praising al-Qaeda.

Key Issues

The imprisonment of influential terrorists therefore raises an important question for Western nations: What is the best practice for incarcerating terrorists? There are three models to consider: the *total segregation* model used by the United States, the *partial segregation* model used by Israel, and the *dispersal* model employed by the British. The challenges of comparative criminology are well known. Cross-cultural comparisons are subject to the vagaries of language and custom, politics, law, public policy, and criminal subcultures. Yet at the end of the day, the United States, Israel, and Great Britain share a common dilemma. Each country is struggling in its own way to work out the institutional methods and conceptual frameworks for controlling the threat of radicalization brought on by the long-term incarceration of Islamic terrorists.

United States: The Total Segregation Model

A total of 362 federal prisoners were serving sentences on terrorism-related charges in the

United States at the close of 2011 (out of a combined state and federal prison population of 2.3 million). Over half of these prisoners were non-US citizens (LaFree 2009). Most were involved in international terrorism (269 inmates) with an additional 93 inmates locked up on charges of domestic terrorism (Shane 2011). Another 171 suspected terrorists were being indefinitely detained without trial by the US military at Guantanamo Bay, Cuba, including five al-Qaeda operatives charged in connection with the 9/11 attacks.

Al-Qaeda has a solid presence in American prisons. Among the 362 terrorist inmates incarcerated by the Bureau of Prisons (BOP) were about two dozen al-Qaeda operatives, most of whom were trained in bin Laden's camps. They included terrorists involved in the 1993 World Trade Center bombing, the 1998 East African embassy bombings, the 1999 millennial plot to bomb the Los Angeles International Airport, and the 2000 bombing of the USS *Cole*. New York's BOP detention centers in Manhattan and Brooklyn held another dozen al-Qaeda members awaiting trial, including one high-ranking jihadist who was once bin Laden's bodyguard. These figures may be preliminary, however, given that federal prisons are experiencing what an official 2006 report called "a constant population of new inmates arrested on terrorism-related charges" (U.S. Department of Justice 2006, p. 49).

Moreover, these inmates are considered capable of radicalizing other prisoners into the cause of violent extremism. The danger to US security, then, is not the sheer number of Muslims in prison as some experts warn (Cilluffo and Saathoff 2006); rather, it is the potential for small networks of radical Muslim inmates to instigate terrorist attacks, either by other radicalized prisoners once released or by existing support structures in the community.

The threat of prisoner radicalization first surfaced several years after 9/11 when MSNBC news correspondent Lisa Myers reported that three federal prisoners – Mohammed Salameh, Mahmud Abouhalima, and Nidal Ayyad – incarcerated at the BOP's Administrative Maximum security facility (ADMAX or Supermax) in Florence, Colorado, for the 1993 World Trade Center bombing, wrote over 90 letters to Islamic militants outside the prison between 2002 and 2004. Fourteen of these letters were sent to prisoners in Spain who had connections to the terrorist cell responsible for the 2004 Madrid train bombings. Mohammed Salameh had also managed to smuggle several letters to Arabic newspapers praising bin Laden as a hero. The government's after-action report condemned the BOP in the strongest of terms, charging that it had failed to monitor terrorists' communications, including mail, phone calls, visits with family and friends, and cellblock conversations, resulting in "little or no proactive" intelligence on the activities of terrorist inmates in custody. "Consequently," the report concluded, "the threat remains that terrorist. . .inmates can use mail and verbal communications to conduct terrorist activities while incarcerated" (U.S. Department of Justice 2006, p. ii). Thus was born the total segregation model.

As a result of the Supermax affair, between 2006 and 2008 the Justice Department transferred all but the most highly secured terrorist inmates (e.g., Unabomber Theodore Kaczynski, shoe bomber Richard Reid, Zacarias Moussaoui, the 20th hijacker of 9/11) to two newly established maximum-security Communication Management Units (CMUs) within the federal prison system – one in the former death row at the US Penitentiary in Terre Haute, Indiana, and the other at the US Penitentiary at Marion, Illinois. (Both units were secretly opened in violation of federal law requiring public notice and comment). The little that is known of these prisoners and their conditions of confinement is due to the investigative reporting of journalists from National Public Radio and the *New York Times* (Johnson and Williams 2011; Shane 2011). According to these sources, prisoners are under 24-h surveillance in the CMUs. Guards and cameras record their every move and hidden microphones pick up every word they speak – a snooping operation that costs US taxpayers more than $14 million a year. Such information – along with data gleaned from the monitoring of phone calls, mail, and visits – is routinely gathered by prison intelligence officers who share their findings with counterterrorism experts at the Justice Department (Shane 2011).

Rules within the CMUs are strict and they are strictly enforced. The CMUs prohibit group prayer beyond the authorized hour-long Jumna services on Fridays and restrict inmate visitation to lawyers and immediate family members. Visits from journalists, human rights advocates, and volunteers are strictly off limits, as are researchers, who are routinely denied access to the CMUs; hence there is no primary criminological research on the incarceration of terrorists in the United States. Inmates are required to hold all conversations in English. Most of the CMU prisoners are Arab Muslims, yet the units also hold some African-American Muslims charged with radicalizing other inmates. Several environmental and animal rights activists are in the CMUs, along with a neo-Nazi, a tax resister, and a member of the Japanese Red Army. Also locked up in the CMUs are a few inmates who have threatened officials while in custody or ordered murders using smuggled cellphones.

In addition to virtually banning the prisoners' contact with the outside world, the objective of the CMUs is to fully segregate terrorist inmates from the general prison populations to prevent them from both converting other convicts to radical Islam and plotting terrorist acts behind bars. By fully segregating terrorists, the BOP argues that it can better concentrate its resources on language translation, content analysis of letters and phone calls, and intelligence distribution. Despite repeated media requests, authorities have refused to release a full list of the CMU inmates, although reporters have compiled a partial list based on open sources. Among them are three convicts who have previously waged terrorist attacks while confined to maximum-security prisons.

Nothing is known of the prisoners' psychological status in the CMUs, the criteria by which they have been chosen for incarceration in the CMUs, their conflicts with guards and other inmates, or their current threat to public safety. Nor is anything known about their rehabilitation, their preparation for community reentry, or their rates of recidivism. Yet many of the CMU prisoners will one day finish their sentences and return to society (some 300 terrorist-related prisoners have completed their sentences and been set free since 2001). Civil rights attorneys have filed lawsuits contending that CMU inmates are denied the right to review the evidence that sent them there or to challenge that evidence. Civil liberties groups claim that prisoners' rights are being violated by the policies that monitor phone calls and severely restrict visitation privileges, sometimes including access to lawyers which leaves prisoners unprotected against mistreatment. Other evidence shows that by creating Muslim-dominated control units, the BOP has inadvertently fostered solidarity and defiance among the CMU inmates, thereby increasing the potential for radicalization. Adding to these problems, the BOP has not instituted de-radicalization programs which are common in other countries. Because of the legal complaints, combined with the atmosphere of secrecy surrounding the disproportionate placement of Muslim prisoners in the CMUs, Terre Haute and Marion have become internationally known as "Guantanamo North."

Israel: The Partial Segregation Model

The most recently available figures indicate that Israel has approximately 10,000 terrorists in custody (out of a total prison population of 25,000), representing a 1,200 % increase over the 802 terrorists held in 2000. These prisoners are divided into two categories: those in administrative detention without formal charges and those with "blood on their hands" – including convicted terrorists who have killed Israeli citizens, terrorists who were en route to a suicide bombing mission but changed their minds, suicide dispatchers, masterminds of the attacks, those responsible for making the explosives, and senior members of terrorist groups. Roughly 700 terrorist inmates are serving life sentences (like most of the world, Israel does not have the death penalty) that are often "stacked" upon one another. It is not uncommon to find prisoners serving multiple life sentences. The majority are Islamic militants from the Palestinian territories. A reported 44 % are members of Fatah (which governs the West Bank), 26 % are members of Hamas (which rules in Gaza), and 14 % belong to Islamic Jihad which operates out of Egypt. They are spread out among 32 prison facilities and three detention centers in Israel proper. About 140 of the terrorist inmates are juveniles (some as young as 14 years old) who are held at two separate facilities.

Terrorists are physically segregated from the general prison populations, yet their confinement is not secret nor are they totally segregated from other prisoners or the outside world. Journalists, documentary filmmakers, legal aid groups, the International Red Cross, politicians, and human rights activists are permitted access to the inmates, as are researchers. Between 2002 and 2009, for instance, Ariel Merari (2010) conducted a series of important psychological studies on Palestinian would-be suicide bombers and their commanders inside Israeli prisons. Another research team led by Edna Erez conducted interviews with Arab/Palestinian women locked up in Israeli prisons on terrorism offenses, exploring their motivations for attempted suicide bombings (Berko et al. 2010). Terrorist inmates have been able to cooperate with criminal inmates inside Israel's prisons, pursuing various goals, both illegal and mundane. They are also allowed to receive food from family and militant groups during Muslim holidays.

Terrorists are confined to large dormitory-style units, holding as many as 40 inmates, where they manage a common money pool and keep decorations, books, magazines, and other personal items. (There is no documented evidence of electronic surveillance on these prisoners, though one can assume such measures are taken.) The inmates wear street clothes; women wear *hijabs*, or headscarves, as part of their traditional Muslim dress. The inmates are allowed to pray together in special areas. They are housed by militant faction; a good number of them were senior leaders of these factions prior to their incarceration. Their declared purpose is to continue the Israeli/Palestinian conflict within the prison system.

According to prison rules, each political faction appoints an inmate leadership council to negotiate with the guards for prisoners' rights. Each faction is organized hierarchically in accordance with the structure of the organizations on the Palestinian street. The prisoners present themselves in a clean and orderly fashion, yet there are distinct differences between the factions. Hamas is organized like an army while the more moderate Fatah reflects the looseness of a Middle Eastern bizarre. The outstanding feature of the Israeli experiment with locking up terrorists is its emphasis on rehabilitation.

Terrorist inmates are offered a 14-point rehabilitation program, beginning with point 1: religious de-radicalization, where mainstream Muslim clerics offer Koranic lessons on reversing Islamic fundamentalism. Point 2 involves educational, vocational, and occupational training and point 3 emphasizes psychological treatment designed to "heal the anger" from years of living under Israeli occupation. Point 4 involves social and family intervention, including flexible mail and visitation rights for Palestinian families. Families from Gaza are currently forbidden from making these visits, however, due to military restrictions on border crossings into Israel. In those cases, photographs of prisoners are sent to family members twice a year in an effort to show that they are being humanely treated. Other modes of rehabilitation include the creative arts, prison coping skills, and a mandated practice of holding congenial interactions between prisoners and guards.

These rehabilitative interventions are impeded, however, by a labeling process wherein imprisonment bestows a hero status on terrorists; some have been elected to the Palestinian legislature as prisoners. Reinforcing this status is a formidable ideology of resistance. Terrorists blatantly refuse to take part in the programs afforded to them by the Prison Service. Instead they spend time educating themselves in political science, international politics, and Israeli politics – often under the auspices of the Open University. They also study terrorism tactics, demolition, and other subjects relevant to their struggle. In so doing, they have turned Israeli prisons into de facto universities of Palestinian nationalism.

Contraband smuggling is a significant problem in Israel. Though forbidden by Fatah and Hamas leadership, heroin and alcohol are problematic among the rank and file, as is the smuggling of cellphones which can sell on the prison black market for as much as $10,000 (US) per phone. Yet their primary mode of communication is the *ashgarim* – crimped notes written on thin transparent paper tightly rolled into bindles ("kites" in the USA). Weapons – in the form of metal shanks, blow darts, and all manners of improvised explosive devices – present a major problem in the hands of terrorist inmates, as does intelligence on the security arrangements of the prisons, which is occasionally furnished to inmates by visiting family members and Palestinian lawyers acting as "mailmen" for Hamas and Fatah. Inmate leaders use this intelligence to both advance militant goals and ensure internal security by interrogating fellow prisoners and punishing those who deviate from the party line.

The security situation inside Israeli prisons began to take a turn for the worse in 2002 following the 9/11 attacks and the renewal of terrorist strikes by Hezbollah in Lebanon. The trend reached a tipping point in 2006 – a year marked by Hamas' landslide victory in the Palestinian legislative election followed by Hamas' decision to call off its truce with Israel and begin attacking the South with rockets, eventuating in the June 25 kidnapping of the 25-year-old Israeli soldier Gilad Shalit (released in 2011 in exchange for 1,027 Palestinian prisoners), prompting severe military reprisals by Israel. Amidst these tumultuous events, a United Nations study declared the humanitarian situation in Gaza "intolerable," with 75 % of its 1.6 million residents dependent on food and aid.

Since then a new generation of terrorists have flooded the prisons of Israel. They are younger than those who came before, more daring, belligerent, and proficient in the use of explosives. They reject all dialogues with the Prison Service. This generational shift has led to an increase in hunger strikes, disturbances, and security breaches involving violence. Since 2006, terrorist inmates have directed terrorist activities in the streets using *ashgarims* and cellphones. They have been involved in a range of attacks on staff, including stabbings, beatings, attempted murders and kidnappings, and throwing boiling oil into the faces of guards. Suicide bombing plots have been uncovered inside prisons, some carrying the mark of extraordinary criminal ingenuity. In one case, pigeons trained by Hamas flew from the Gaza Strip to the yard of an Israeli prison carrying explosives and cellphones strapped to their talons. On balance, prisons have become the center for terrorist planning and coordination with one goal in mind: destroying the State of Israel and replacing it with an independent Palestinian state. As a result, prison guards are primarily recruited from combat units within the Israeli military.

Great Britain: The Dispersal Model

A total of 150 terrorists are currently locked up in British prisons (out of a total prison population of 95,000). Eighty-nine of them are al-Qaeda inspired extremists who justify violence through distorted interpretations of Islamic texts. These terrorists are unknown to most Americans but their deeds are not. They include Abu Doha, once al-Qaeda's main European recruiter and a central figure in the millennium plot to bomb the Los Angeles International Airport, along with Saajid Badat and Nizar ben Trabelsi, accomplices to Richard Reid in a second bombing attempt on a US-bound aircraft in 2001. Other terrorists in custody include Irish Republicans, animal rights extremists, anarchists, and a growing number of far-right activists whose crimes were motivated by neo-Nazism.

There are 140 prisons in England and Wales, designated by security classification. Category A is the top classification (with three designations: standard, high, and exceptional risk). The escape of exceptional risk inmates, according to the Prison Service, "would be highly dangerous to national security." There are eight high-security prisons deemed suitable to hold Category A prisoners. Al-Qaeda terrorists are housed in these prisons, usually in single occupancy cells. For the most part, they are model prisoners: quiet, polite, clean, and often university educated. Most are ethnic minorities. In instructive contrast to the USA and Israel, in Britain terrorists are not segregated from the general prison populations. Rather, they are "dispersed" within those populations. As a result, al-Qaeda terrorists walk the cellblocks alongside murderers, rapists, drug dealers, skinheads, and old-school IRA members.

Al-Qaeda terrorists are subject to strict security precautions which are strictly upheld. They include restricted movements, pat down searches, daily strip searches, cell searches (called "spins"), and round-the-clock camera coverage. Guards, or "personal officers," are assigned individual terrorists and are required to monitor them for intelligence which is shared with Scotland Yard detectives apportioned to the facilities. To prevent them from building power bases, terrorists are required to change cells every 28 days. They are offered a full complement of rehabilitative services, including education, current affairs, history, and politics; library, inmate tutoring, and recreation; multi-faith worship allowing for group prayer and counseling by Muslim chaplains; psychological treatment; and creative workshops as well as visitation privileges, including the opportunity to meet with selected researchers from the Home Office.

The British dispersal model has two objectives. The first is to promote a sense of inclusiveness common to civil society. One facet of this mission is finding ways to enhance mutual accommodation between Western values and respect for the peaceful traditions of Islam. The second objective is more pragmatic. As a security matter, the dispersal model is intended to prevent the clustering of terrorists in cellblocks, thereby stopping them from forming paramilitary gangs.

The model has its roots in the Irish "Troubles" which began in 1969 as Irish Republican Army prisoners became the central focus of the decades-long conflict. IRA prisoners embarked on a campaign to challenge their status as ordinary criminals, seeking instead to be classified as political prisoners, through organized resistance against British prison authorities. This prison resistance movement eventually erupted in rioting, arson, hostage taking of prison staff and volunteers, and several high-profile escapes. By 1972, IRA prisoners at Belfast's Maze Prison were running their own paramilitary training and indoctrination programs. The campaign reached a tipping point with the 1981 Maze hunger strikes. When Prime Minister Margaret Thatcher refused to give in to IRA demands concerning the hunger strikers, she became the person most responsible for the painful deaths of Bobby Sands and nine other strikers. As a result, the IRA decided to kill Margaret Thatcher, leading to the spectacular 1984 assassination attempt against her at the Grand Hotel in Brighton, England, killing five and injuring over 30 others. Meanwhile, violence between different dissident groups also spilled over beyond the prison walls, with 29 prison officers being killed on the streets of Belfast and London during the Troubles (Pantucci 2008). The goal of the dispersal model is to prevent this violence from reoccurring with al-Qaeda.

Usually, the British dispersal model is successful. Prisoners charged with al-Qaeda-related offenses have been spread out over a number of institutions, thereby undermining their concentration of power. The dispersal model has also yielded increased intelligence, low rates of recidivism, and thus far no repeat of the violence against prison officers witnessed during the Troubles. But officials are nevertheless concerned about the long-term influence Islamic terrorists are having on other inmates. Al-Qaeda inmates have been found in possession of information relating to the bombing of passenger aircraft; some have been found with smuggled laptops and cellphones used to download al-Qaeda materials on how to manufacture explosives. Al-Qaeda inmates have also found ways of recruiting followers beyond the walls.

And despite the best of intentions, al-Qaeda inmates have been able to skirt security regulations and establish themselves as de facto gang leaders who use forced conversions and other forms of coercion (theft and intimidation) to recruit vulnerable inmates into the mix. This has occurred against the backdrop of a rise in minority religious groups in British prisons and fears of Islamic radicalization in the wake of the 2005 terrorist attacks on London's transportation system (the 7/7 bombings and the 7/21 follow-on attempt). System-wide, prisoners identifying themselves as Muslim rose 141 % between 1997 and 2007 (Liebling 2012). This growth was due mainly to the appearance of a younger, more ethnically diverse generation of convicts who have formed a new gang culture in British prisons which has not only imported the values and behaviors of street life into prisons (including drug and cellphone trafficking) but also glorifies Islamic terrorists and their extremist ideologies. Accordingly, in 2009 the security minister of the Prison Service proclaimed that Britain's high-security prisons had become "incubators of extremism" (Leppard 2009).

The confinement of al-Qaeda terrorists alongside Muslim gang members has therefore presented important challenges to high-security prison administrators in Great Britain. Most dauntingly, the dispersal model has led to increased levels of radicalization and in-prison violence that, some say, now rivals the challenges once posed by the IRA. As with most wardens in the United States today, the prison governors of Britain belong to a generation trained in what Feeley and Simon (1992) termed the "new penology." Whereas traditional penology stemmed from criminal law and criminology with an emphasis on punishing and rehabilitating individual prisoners, the new penology is based on a management style in which "specialists" think about criminal subpopulations as "aggregates" that need to be "herded" through "actuarial" risk assessments of danger. For Feeley and Simon, the new penologists provide an important "waste management" function for the state.

As a result, the primary focus of the British dispersal model is intelligence gathering and risk assessment of Islamic radicals. Officials rely on a set of radicalization risk factors which manifest themselves in prison behavior and communication. Of particular concern are overt "feelings of anger, grievance and injustice." In this regard, intelligence is gathered on dialogue and actions of prisoners which reflect a preoccupation with the injustices or the corruptness of the Prison Service and/or the British Government – particularly with the British-supported wars in Afghanistan and Iraq. Although some Muslim inmates "use prison to politicize everything," efforts are made to distinguish normal prison behavior from a larger grievance against Islam. For instance, a Muslim inmate who bangs on his cell house door because he is upset about a broken toilet is one thing; complaining that the broken toilet is evidence of Western imperialism and hatred of Islam is quite another.

A host of risk factors involving dominance and control are also used in intelligence collecting (e.g., appointing oneself as a leader or group spokesman; manipulating and controlling other inmates). Risk factors associated with correspondence and telephone calls include such things as giving commands and counsel to other prisoners, especially if the prisoner is ill-qualified to do so, and receiving requests for instructions from other prisoners. Intelligence is also assembled on susceptibility to indoctrination and groupthink. Here officials are on the lookout for sudden behavioral changes among inmates after coming into contact with charismatic leaders. The primary concern is with young minority prisoners who appear to be lost in the mix or are seeking a transformative experience.

An innovative feature of this control strategy is the identification of *jail craft* or an inmate's show of personal skills necessary to survive in high-security custody. Those with immature jail craft – inmates who have not yet learned how to "play the game" – are deemed susceptible to radicalization. The concept is also used as an anchor for counseling sessions intended to emphasize incongruities between an imagined jihad and actual life in a Western prison. Al-Qaeda's training manual offers several incongruities. For example, the Qaeda document stipulates: "Inside the prison, the brother should not accept any work that may belittle or demean him or his brothers, such as the cleaning of the prison bathrooms or hallways." However, as anyone who has ever worked in prison knows, one of the most coveted jobs among prisoners is that of a janitor or custodian – it is the one assignment that affords inmates the freedom to move around the cellblocks and often to enter security offices.

Finally, for a host of applied reasons, the dispersal model has become the subject of controversy. Prison officers required to be on "red alert" for signs of Islamic radicalization run the risk of alienating ordinary Muslims. The rise in Muslim extremism has created a power imbalance leaving other inmate groups in a state of anxiety. Increased tensions between staff and prisoners, combined with growing competition for state resources and official charges that high-security prisons have become "incubators of extremism," have led some officials and private prison companies to push for a specialized high-security unit for al-Qaeda terrorists and other extremists. Essentially, there is a mounting concern in Great Britain that it may be time to cast aside the dispersal model in favor of the total segregation model employed by the Americans.

Future Directions

The read-across from America to the United Kingdom may not be as obvious as some think. America's secret prisons of the post-9/11 era are hardly a model for forward-looking counterterrorism policy. By associating with the stain of Guantanamo and by insinuating the shame of Abu Ghraib, the British could attract recruits into radical Islamic networks by making the terrorist's cause appear a just response to an unjust enemy. It could also wreak havoc closer to home. Across Europe, Guantanamo has become a symbol of what many see as America's dangerous drift away from the ideals that made it a moral beacon in the post-World War II era. The British may want to reconsider the new penology first.

An often overlooked aspect of Feeley and Simon's argument is that it is not the new penology that has failed, but administrators who have failed to go beyond the new penology "to a new way of narrating the power to punish that can help shape the public and political discourse" (Simon and Feeley 2011, p. 78). Understanding the possibilities of de-radicalization as a "soft" weapon in the fight against Islamic extremism in prison may be such a narrative. It is a way to scale responses to a level appropriate to the real threat of radicalization without doing more harm militarily. This is not easy. Even in countries like Indonesia where de-radicalization programs are routinely offered to terrorist inmates, along with a strong community support structure, recidivism is a genuine concern as is the radicalization of criminal offenders and even Muslim prison officers who watch over them. As one Indonesian terrorist told a researcher after completing his de-radicalization program, if he was released from prison today, he would bomb the US embassy tomorrow (Ungerer 2011).

Decades of criminology shows that rehabilitation lowers recidivism rates when cognitive behavioral treatments target known predictors of crime (Cullen and Gendreau 2001). Although much remains to be done, research indicates that the predictors of prisoner radicalization are gang affiliations around Salafi-inspired Islam and one-on-one proselytizing by charismatic leaders with a violent agenda (Hamm 2008; Liebling 2012). Britain already employs treatment strategies targeting these predictors. Its counseling policy aimed at minority inmates with underdeveloped jail craft is one way of targeting the predictors of radicalization. Inverting al-Qaeda's training manual by using it as a teaching tool to dramatize incongruities between the jihadist dream and the reality of everyday prison life is certainly another. In fact, it is a brilliant example of undermining al-Qaeda's appeal among young prisoners without tackling ideology head on. De-radicalizing older jihadists is more difficult. One reason for this is that Islam's sacred law, the *shariah*, is essentially silent on the subject of incarceration as a rehabilitative venue. The crucial step in de-radicalization, then, is getting to young offenders near the time of their conversions to Islam when they are not yet settled into a commitment to jihad.

Israel's elaborate rehabilitation program for Palestinian terrorists will not work until the peace process is settled. But that does not mean that

Israel's strategies will not work with at-risk inmates in other countries. Employing Muslim clerics to challenge the tenants of Islamic fundamentalism has the potential to broaden transcendent horizons among the vulnerable to the point where serious discussions can be held about the meaning of such loaded concepts as *jihad* (struggle), *takfir* (blasphemy), *shahada* (martyrdom), and the establishment of the Islamic State or *caliphate*. But perhaps most importantly, discussions can be held about al-Qaeda's long-term future in the post-bin Laden era, thereby addressing al-Qaeda's failure to offer any positive vision for building a modern society as imagined by activists of the Arab Spring. Psychological interventions intended to "heal the anger" of prior victimization have the potential to reverse rigid anti-authoritarian views held by convicts at risk for radicalization, especially when tied to meaningful educational and occupational training. Indeed, changing ideological beliefs alone is unlikely to work unless combined with rehabilitative interventions aimed at desistance. And the mandated practice of forming congenial relationships between prisoners and guards can potentially take the air out of moral panics about the transgressive other.

If these are international standards by which a new de-radicalization movement is to be judged, the United States of America has little to offer. Its practice of incarcerating terrorists is secret. Not only secret but it is apparently void of any operating penal philosophy concerning the rehabilitation of anti-Western jihadists. Consequently, most of these inmates see their imprisonment not as justice but as US revenge against *all* Muslims for the attacks of 9/11. This will not change until the emblems of American injustice and abuse – Guantanamo and now "Guantanamo North" – are abolished and replaced with prisons that uphold America's core values.

Recommended Reading and References

Ammar N, Weaver R, Saxon S (2004) Muslims in prison: a case study from Ohio prisons. Int J Offender Ther Comp Criminol 48:414–428

Atran S (2010) Talking to the enemy: faith, brotherhood, and the (un)making of terrorists. Ecco, New York

Beckford JA, Daniels J, Khosrokhavar F (2005) Muslims in prisons: challenges and change in Britain and France. Palgrave, Basingstoke

Berko A, Erez E, Globokar J (2010) Gender, crime and terrorism: the case of Arab/Palestinian women in Israel. Brit J Criminol 50:670–689

Brandon J (2009) Unlocking al-Qaeda: Islamist extremism in British prisons. Quilliam, London

Cilluffo F, Saathoff G (2006) Out of the shadows: getting ahead of prisoner radicalization. George Washington University/University of Virginia, Washington, DC

Cullen FT, Gendreau P (2001) From nothing works to what works: changing professional ideology in the 21st century. Crime Delinquency 81:313–338

Feeley MM, Simon J (1992) The new penology: notes on the emerging strategy of corrections and its implications. Criminology 30:449–474

Hamm M (2007) Terrorist recruitment in American correctional institutions: an exploratory study of non-traditional faith groups. National Institute of Justice, Washington, DC

Hamm M (2008) Prisoner radicalization: assessing the threat in U.S. Correctional institutions. NIJ J 261:14–19

Hamm M (2009) Prison Islam in the age of sacred terror. Brit J Criminol 49:667–685

Hamm MS (2013) The spectacular few: Prisoner radicalization and the evolving terrorist threat. New York: New York University Press

Irwin J (2009) Lifers: seeking redemption in prison. Routledge, New York

Johnson C, Williams M (2011) Guantanamo north: inside secretive U.S. prisons. NPR News, March 3

Knox G (2002) Melanics: a gang profile analysis. J Gang Res 9:1–76

Kruglanski A, Gelfand M, Gunaratna R (2010) Detainee de-radicalization: a challenge for psychological services. APS Observer, January

LaFree G (2009) Understanding radicalization in the U. S.: surveys and interviews with convicted terrorists. Department of Homeland Security Grant Application

Leppard D (2009) Terrorists smuggle fatwas out of secure prisons. Times (of London), November 15

Liebling A (2012) An exploration of staff-prisoner relationships at HMP whitemoor. Home Office, London

Merari A (2010) Driven to death: psychological aspects of suicide terrorism. Oxford University Press, New York

Pantucci R (2008) Britain's prison dilemma: issues and concerns in Islamic radicalization. Terrorism Monitor, March 24

Rand (2008) Radicalization or rehabilitation: understanding the challenge of extremist and radicalized prisoners. Rand, Washington, DC

Shane S (2011) Beyond Guantanamo, a web of prisons for terrorist inmates. New York Times, December 11

Simon J, Feeley MM (2011) The forms and limits of the new penology. In: Blomberg TG, Cohen S (eds) Punishment and social control. Transaction, New Brunswick, NJ

Spalek B, El-Hassan S (2007) Muslim converts in prison. Howard J Crim Just 46:99–114

U.S. Department of Justice (2006) The Federal Bureau of Prisons' monitoring of mail for high-risk inmates. Office of the Inspector General, Evaluation and Inspections Division, Washington, D.C, Report Number I-2006-009

Ungerer C (2011) Jihadists in jail: radicalization and the Indonesian prison experience. Australian Strategic Policy Institute, Sydney

Incarceration

▶ Comparative Incarceration
▶ Drug Abuse and Alcohol Dependence Among Inmates
▶ Imprisonment and Cardiac Risk
▶ Rational Choice, Deterrence, and Crime: Sociological Contributions

Incarceration Rates

▶ Comparative Incarceration

Incivilities

▶ Defining Disorder

Incivilities Thesis

▶ Broken Windows Thesis

Inclusionary Control

▶ Juvenile Diversion

Indigenous

▶ Transitional Justice

Indigent Defense

▶ Public Defenders

Individual Characteristics and General Strain Theory

Lee Ann Slocum
Department of Criminology and Criminal Justice, University of Missouri – St. Louis,
St. Louis, MO, USA

Synonyms

Continuity and change; GST; Life course

Overview

When initially proposed, General Strain Theory's (GST) primary emphasis was on explaining why some people are more likely to offend than others. In more recent years, it has been extended to address within-person patterns of offending, or more specifically, why some individuals continue to offend well into adulthood while most limit their involvement in delinquency to adolescence. Explanations of both inter- and intra-individual differences rely on the same central principle: Patterns of offending can be accounted for by differences in exposure and reaction to strain between people and over time. This entry provides a brief review of the theoretical and empirical literature on GST explanations of between- and within-person differences in offending. Emphasis is on understanding the manner in which individual characteristics, such as personality traits, social support, and social control, affect the types and quantity of strains to which a person is exposed as well as their tendency to respond to strain using illegal coping mechanisms. After assessing the state of our knowledge regarding GST explanations of between- and within-individual differences in offending, some of the most pressing questions that remain to be answered on this topic are outlined.

Introduction

In 1992, Robert Agnew took traditional strain theory in a new direction with his General Strain Theory (GST). GST conceptualizes strain as a psychosocial concept that can explain individual-level offending. The crux of this theory is that strain – resulting from exposure to noxious stimuli, the removal of positive stimuli, or blocked goals – elicits negative emotions including fear, depression, and most importantly, anger. Illicit coping mechanisms, including offending, substance abuse, and other problem behaviors, may be used in an effort to relieve these negative emotions. As originally proposed, GST set out to explain between-individual differences in offending, or why some people are more likely to offend than others. Specifically, the theory states that people will be more likely to offend if they are exposed to higher levels of certain types of strain (e.g., victimization, work in the secondary labor market, abusive peer relations, negative secondary school experiences, poor parenting). Strains are most likely to lead to crime when they have the following characteristics: (1) are seen as unjust; (2) are viewed as high in magnitude; (3) are associated with low levels of social control; and (4) create pressure for criminal coping (Agnew 2001). People are also more likely to offend if they have a tendency to respond to strain with crime. Both exposure and reaction to strain are influenced by characteristics of the individual, including personality traits, social support, and social control; therefore, individual differences are critical to understanding between-person differences in offending. Socio-demographic and group characteristics such as race, gender, and socioeconomic status are important as well, but they are reviewed in another entry ("▶ Group Characteristics and General Strain Theory," by Stacy De Coster).

In recent years, GST has been extended to explain within-person patterns of offending or more specifically, why some individuals continue to offend well into adulthood while most limit their involvement in delinquency to adolescence (Agnew 1997, 2006b). Similar to GST explanations of between-person differences in offending, Agnew argues that for most individuals offending temporarily peaks in adolescence because youth are exposed to more criminogenic strains and are more reactive to the strains they experience during this period of their lives. Likewise, the small proportion who engage in chronic offending do so because they experience stable, high levels of strain and because they possess characteristics that make them more likely to cope with strain by using crime.

This entry provides a brief review of the theoretical and empirical literature on GST explanations of between- and within-person differences in offending. Emphasis is on describing the types of individual characteristics hypothesized to affect both exposure and reaction to strain, and outlining how these individual differences are key to understanding patterns of offending. This work is not intended to be an exhaustive review; rather, I focus on select characteristics that are most central to these issues, such as negative emotionality and low constraint. I also discuss in detail work that has received less attention in the literature, including GST explanations of offending over the life course.

Background

Between-Person Differences in Offending: Explaining Differences in Exposure and Reaction to Strain

Strain is often viewed as part and parcel of modern life. Given the pervasiveness of strain, one of GST's biggest challenges has been identifying why some people react to strain with crime, while others do not. GST argues that between-person differences in offending can be accounted for, in part, by individual differences that shape the types and amounts of strain to which people are exposed. They also affect the likelihood a person will respond to strain with criminal coping and moderate the relationships between strain, negative emotions, and offending.

Individual differences that are central to GST include social support and social control, associations with criminal others, coping skills, and the personality traits of negative emotionality and low constraint. The contributions of each of these characteristics to understanding why some, but not others, engage in offending are discussed below (see Agnew 1992, 2006b for a more in-depth discussion of these issues).

Agnew argues that individuals with low levels of conventional social support are more likely to react to strain in a criminal manner. Social support, which is defined as assistance that is used to meet an individual's needs, can come from many sources (e.g., parents, teachers, friends, intimate partners, employers) and may take the form of advice, comfort and assurances, direct assistance with problems, and material goods. Individuals who can turn to others for help will be less likely to respond to strains in a criminal manner because they have a conventional support system that can provide them with resources to cope with their problems using legal means. Although not discussed by Agnew, it is also possible that people with more social support will experience fewer strains because they are better able to extricate themselves from stressful situations before these situations affect other life domains. For example, someone with high levels of social support may draw on these resources to find work before unemployment leads to financial problems.

A second factor that can explain between-person differences in offending is social control. Individuals who are low in social control generally have less direct supervision, weaker emotional bonds with others, and do not believe in conventional social norms. People with these characteristics will be more likely to react to strain with illegal behavior because they experience fewer negative consequences as a result of their actions. Specifically, they are unlikely to get caught because their behavior is not monitored; their lack of attachment to others means that they have few people to disappoint with their behavior; and they do not believe crime is wrong, so they experience little guilt. It is also possible that individuals with low social control will experience more strains. For example, youth whose behavior is unmonitored have more opportunities to become involved in situations that are likely to lead to conflict with others.

Associating with criminal others can also increase the likelihood that a person will react to strains in a criminal manner. Drawing on social learning theory, Agnew argues that interacting with offenders can cause one to develop a disposition for crime because criminal peers model illegal behavior, provide reinforcements for criminal activity, and teach beliefs that are favorable to crime. In addition, as individuals observe their peers engaging in crime without being caught or sanctioned, they may come to believe that crime is a low cost endeavor. This belief is reinforced as peers fail to sanction the illegal behavior of group members and instead laud offending. Associating with criminal peers can also enhance one's exposure to criminogenic strains as criminal peer groups are more likely to engage in conflict with one another and with those outside the group.

Finally, individuals with few personal resources to draw upon, including poor problem-solving and social skills, are more likely to engage in crime because they have limited capacity to solve problems in a legal manner. Personal resources that may be particularly important for understanding between-person differences in offending are the personality traits of negative emotionality and low constraint. In fact, Agnew and others (Agnew et al. 2002) argue that these traits may be the primary factors moderating the relationship between strain and crime. Negative emotionality captures the likelihood a person will breakdown under stress. People with high levels of this trait have a tendency "to experience events as aversive, to attribute these events to the malicious behavior of others, to experience intense emotional reactions to these events. . ., and to be disposed to respond to such events in an aggressive or antisocial manner" (Agnew et al. 2002, p. 46). In comparison, low constraint characterizes individuals who "are impulsive, are risk-taking/sensation-seeking, reject conventional social norms, and are unconcerned with the feelings or rights of others" (Agnew et al. 2002, p. 46).

Negative emotionality and low constraint are moderately stable over the life span, but relative stability (i.e., maintaining rank order compared to others) is more common than absolute stability (i.e., sustaining the level of a trait over time) (see Miller and Lynam 2001).

According to Agnew (2006b), people high in negative emotionality and low in constraint are more likely to offend for two primary reasons. First, negative emotionality affects how people perceive, and therefore react to, stimuli. In situations where the intent of the actor is ambiguous, those with high levels of negative emotionality will be more likely to attribute harmful intent to the actor and blame others for their problems (Caspi et al. 1994). These cognitive processes increase the likelihood strain will lead to anger, which is the emotion most often associated with offending and aggressive responses. Because individuals with weak constraint tend to act quickly with little thought, low constraint may enhance the effect of negative emotionality on offending by encouraging the rapid translation of negative emotions into impulsive action (Caspi et al. 1994). There is also some evidence that people with low levels of constraint perceive stressors as more threatening or negative (Taylor and Aspinwall 1996). For these reasons, negative emotionality and low constraint increase the likelihood a person will respond to strain with criminal coping.

Second, people who are high in negative emotionality and low in constraint are more likely to experience strains that are associated with offending. This is due to the fact that they have problems getting along with others and are unlikely to exert the effort necessary to excel in conventional activities. In turn, failure in conventional pursuits can result in greater exposure to strain-producing environments, like delinquent peer groups. People with these traits are also more likely to evoke negative reactions from their social environments. These responses may be stressful in-and-of themselves, or they may introduce the individual to stressful situations. Finally, the search for interpersonal interactions with others they find compatible and stimulating may lead those who are high in negative emotionality and low in constraint to environments and relationships filled with stressors.

Differences in Within-Person Patterns of Offending

Originally, GST was formulated to explain between-individual variations in offending. More recently, it has been extended to explain within-individual patterns of behavior, including why some individuals engage in persistent offending over the life span while most others engage in normative delinquency that is limited to adolescence (Agnew 1997, 2006b). In his explanation of these within-individual patterns of offending, Agnew draws on many of the same concepts he uses to explain between-individual differences including poor problem-solving skills, lack of conventional sources of social support, low social control, association with delinquent peers, and negative emotionality and low constraint. Moreover, the fundamental explanation regarding the patterning of criminal behavior remains the same. Life-course persistent offenders experience high levels of criminogenic strains throughout their lives and have time stable characteristics that make them more likely to respond to strain with crime. In comparison, adolescent limited offenders experience a relatively short-lived peak in criminogenic strains and in the characteristics that promote criminal coping during their teenage years and will generally limit their offending to adolescence.

Life-Course Persistent Offending

GST posits four mechanisms that contribute to stability in offending over the life course. All but one of these mechanisms rely on the relatively time stable traits of negative emotionality and low constraint, directly or indirectly, to explain continuity. In the first explanation, people with these traits are theorized to be more reactive to potential sources of strain. As described above, during interactions, they will be more likely to attribute harmful intent to the actions of others, more likely to respond aggressively, and more likely to believe that aggressive responses will be effective. These perceptions increase the probability they will respond to any given strain with offending at all points in their life.

Second, individuals who are high in negative emotionality and low in constraint will persist in offending because they experience more strains over their life course. They have difficulty in getting along with others and succeeding in conventional pursuits, which generates strain and increases the likelihood they will end up in aversive environments and situations. People possessing these traits also actively seek out environments or relationships that are high in strain. In turn, experiencing strain contributes to the maintenance of negative emotionality and low constraint. In these first two explanations, it is the relative stability of traits overtime, and therefore continuity in exposure and sensitivity to strain, that accounts for persistent offending.

The third explanation Agnew proposes to explain continuity involves passive selection into an aversive home environment. Children who are high in negative emotionality and low in constraint are likely to have parents who possess these traits, partially because there is a hereditary component to these characteristics and partially because parents who are high in negative emotionality and low in constraint are likely create a family environment that promotes the development of these traits in their children (Rutter et al. 1997). Parents with these characteristics are impatient and irritable and typically create home environments for their children that are high in strain. Youth rarely have recourse to change their living arrangements, and therefore, stability in offending results from consistent exposure to criminogenic strains in the family environment.

The remaining pathway is the only explanation for continuity in offending outlined by Agnew that does not rely on the traits of negative emotionality or low constraint. Instead stability is attributed to state-dependent mechanisms in which offending and its negative ramifications have a causal effect on future behavior by increasing exposure and reactivity to strain. In this process, individuals facing persistent poverty and underclass status are initially more likely to offend because they are exposed to more strain, are more sensitive to the strain they experience, and are more likely to respond to strain with offending. Offending then creates its own set of consequences, like decreased social control. It can also disrupt the normative timing of life events, such as the birth of a child, making these events more stressful. These aversive circumstances increase exposure and reactivity to strain and make offending more likely. Therefore, offending is maintained over time by its reciprocal relationship with strain.

More recently, Slocum (2010a) has extended Agnew's explanations of stability. Drawing on the sociology of stress literature and the life-course perspective, she proposes two additional ways that GST may account for persistent offending. First, she argues that when experienced early in life, particularly before age 10, trauma and severe chronic strain can have long-lasting effects on how a person reacts to later strain. The size and force of these early stressors can be great enough to overwhelm available coping mechanisms, decreasing the amount of additional stress an individual can withstand. In this manner, early stressful experiences can lead to stability in offending through their lasting effect on stress reactivity.

The second explanation for stability proposed by Slocum is based on the process of stress proliferation, which occurs when initial stressors expand into aspects of the same life domain that were previously unaffected or when they disrupt patterns of social interactions, obligations, and expectations, leading to problems in other life domains (Pearlin et al. 1997). For example, the loss of a job may lead to financial problems which, in turn can create marital strain. Importantly, stress proliferation can lead to chronic strain, which Agnew (1992) argues is likely to be dealt with through illicit coping.

Adolescent Limited Offending

GST is also able to account for the fact that most individuals engage in normative delinquency beginning in early to mid-adolescence, but then leave this behavior behind as they move into adulthood. This theory attributes the adolescent peak in delinquency to two factors: (1) Compared to other age groups, adolescents are more likely to experience criminogenic strains and (2) they are more likely to react to these strain using criminal forms of coping. I discuss each of these explanations in turn.

Youth are more likely to encounter the types of strains associated with criminal behavior; therefore, the peak in crime during adolescence partially is due to greater exposure to strain. Adolescence is a time of great change. The center of youths' world shifts from their families to their peers, they become increasingly independent and mobile, and they transition from middle school to larger high schools. As a result "adolescents. . .experience a dramatic increase in the size and complexity of their social world" (Agnew 1997, p. 114). At the same time, they are less monitored by parents and their bonds with conventional adults become weaker. Therefore, these new relations are often accompanied by adversity because they require youth to engage in unchartered patterns of interactions without the guidance of their parents or other adults. Adolescents also have more interactions with delinquent peers who are likely to treat them poorly and initiate conflict with one another as well as with parents, teachers, and other adults. Agnew argues that another source of increased strain in adolescence is the blockage of goals. The most important of these goals are autonomy and popularity among their peers, both of which can be achieved through delinquent behavior. Moreover, the desire for autonomy brings youth in conflict with those who attempt to exert social control over them, namely, parents and teachers, and can lead to frustration and anger. Importantly, adulthood is associated with diminishing exposure to these types of strains. The social world of adults is often more limited compared to that of adolescents; adults settle down with a romantic partner, move into the more socially limited world of work, and have less contact with delinquent peers. In addition, adults get the autonomy that they so desired as adolescents and the importance of peer popularity and of other goals that can be achieved through delinquency declines with age. Therefore, according to GST, the decline in crime that accompanies adulthood is explained, at least partially, by a decrease in exposure to strain.

In addition, adolescents are also more likely than other age groups to define the events they experience and their relationships as aversive (Agnew 2006b). That is, not only are they more likely to experience objective strains (i.e., events and circumstances most people would find aversive), but they are also more likely to experience subjective strains. Important cognitive changes occur during adolescence. Compared to childhood, youth are capable of increasing abstract thought that enables them to empathize with others and take on their emotional burdens. At the same time, adolescents are egocentric, and thus, they feel that their actions are always being monitored and judged by others. This self-centeredness, combined with the fact that adolescents' lives are lived in the public sphere, makes it more difficult for them to cognitively reinterpret or minimize their strains. Youth are also more likely to blame their problems on others, which increases the likelihood they will react to strain with anger, and in turn, delinquency. The tendency to interpret events as aversive declines as youth mature, becoming less self-centered and learning not to blame their problems on others. This leads to a corresponding decline in illicit behavior.

The adolescent peak in offending can also be explained by the fact that this age group is more likely to respond to strain with delinquency than other age groups. As youth become more independent, there is an increased expectation that they will handle their own problems; however, they have little experience with coping since in their younger years, they could turn to adults to help them solve their problems. As a result, they may be forced to use less effective, illegitimate forms of coping mechanisms, such as delinquency. Adolescents also lack the resources and the power to escape from their problems, so this is one less coping mechanism that is available to them. Because adolescents have few social ties to prosocial adults and have more associations with delinquent peers, the costs of engaging in delinquency are lower and this behavior may even be rewarded. Thus, for youth, illicit behavior may be a relatively attractive way to cope with strains. As youth age, they have more opportunities to engage in prosocial coping and many of the factors that enhance the likelihood an individual will respond to strain with delinquency (e.g., low social control and exposure to delinquent peers) diminish.

In summary, GST explanations of between- and within-individual differences in offending are very similar – offending is more likely when individuals experience more criminogenic strains and when they possess the characteristics that increase the probability they will cope with strain in a criminal manner. The next section briefly reviews the research on these explanations.

State of the Art

Between-Person Differences in Exposure and Reaction to Strain

A number of tests using diverse samples have examined whether the effect of strain on negative emotions and criminal offending is stronger for individuals who possess certain characteristics, such as low social control, limited social support, and poor problem-solving skills, and for those who associate with criminal others. Support for GST has been mixed at best (see Agnew 2006a, b for a review). For example, in one study that looked at 96 interaction effects, only 6 of these were consistent with GST (Aseltine et al. 2000). In a more recent comprehensive test of GST that uses a sample of Korean youth, Moon et al. (2009) also found mixed support: Good problem-solving skills and high levels of social control mitigated the effect of strain on many types of delinquent behaviors, but contrary to GST, the effect of strain on offending was weaker for individuals with more delinquent peers. Looking across studies, there are few conditioning mechanisms that have received universal support. For example, some studies (e.g., Moon et al. 2009) have found that delinquent peers condition the effect of strain on offending, while others (e.g., Tittle et al. 2008) have not. There is some evidence, however, that conditioning effects may better explain the link between strain and affective responses than strain and delinquency (see Aseltine et al. 2000).

Research examining the conditioning effects of negative emotionality, low constraint, and related traits is generally more consistent with GST. For example, Agnew and colleagues (2002) found that the relationship between self-reported strain and delinquency was stronger for youth who had high scores on a measure that combined negative emotionality and low constraint. Additional research has examined whether related traits moderate the strain-offending link. For example, some studies indicate that individuals with low self-control are more likely to react to strain with delinquency (e.g., Hay and Evans 2006; Mazerolle and Maahs 2000), but other research has found that youth with low self-control are *less* reactive to strain (e.g., Ousey and Wilcox 2007). It is possible that self-control enhances the effect of strain on some outcomes, but not others (Peter et al. 2003). Methodological limitations with many of these studies including improper causal ordering and limited measures of key concepts make it difficult to reach a definitive conclusion regarding the moderating effects of negative emotionality and low constraint. Still, this is a promising avenue for future research.

Few tests of GST have examined whether individual characteristics affect exposure to the types of strain likely to lead to offending, but there is a good deal of evidence to support this idea (see Agnew 2006b for a partial review). For example, individuals involved in delinquent peer groups are more likely to be victimized and come into conflict with others, and negative emotionality, low constraint, and related characteristics have been linked to greater exposure to stressors.

Differences in Within-Individual Patterns of Offending

Compared to other aspects of GST, there has been relatively little research that has examined the ability of this theory to account for within-individual patterns of offending across the life span, although research on development generally supports Agnew's explanations of continuity and the adolescent peak in offending (see Agnew

1997, 2006b, and Slocum 2010a for reviews). Only one study has attempted to examine all of the pathways expected to promote stability in offending simultaneously. Using data spanning approximately 30 years, Slocum (2010b) examined whether individual differences, childhood experiences, and social location account for the continuity of strain and illicit drug use from adolescence through adulthood. As hypothesized by GST, negative emotionality/low constraint had a direct effect on adolescent drug use, and strain in adulthood had a stronger effect on depression and substance use for individuals with these traits. Contrary to theoretical expectations, individuals higher in negative emotionality/low constraint did not experience more stressors than other individuals and there was no support for the passive selection hypothesis. More support was found for dynamic processes including the reciprocal relationship between strain and offending and stress proliferation. Overall, the findings suggest that traits play a role in maintaining substance use from adolescence to adulthood, but that dynamic aspects of the stress process may be more important for explaining continuity.

Studies testing GST explanations of adolescent limited offending are also rare. Hoffmann and Cerbone (1999) used hierarchical growth curve modeling to relate increases in stressful life events to the escalation of delinquency in adolescence. Consistent with GST, they found that experiencing an increasing number of life events was related to intra-individual growth in delinquent activity. Although this study did not examine whether GST could explain the later decrease in offending observed in adulthood, other researchers have examined desistance from a GST perspective. Using seven waves of data from the National Longitudinal Survey of Youth, Gunnison and Mazerolle (2007) found that youth who reported lower levels of negative relations with adults were more likely to desist from minor delinquency even after controlling for other factors theoretically predicted to account for desistance. There was no relationship, however, between decreases in other types of strain and desistance, and none of the strain measures accounted for desistance from more serious delinquency.

Another study examined both GST explanations of desistance and persistence. Eitle (2010) used five waves of data collected from a sample of middle school males who were followed through age 25. He found that decreases in both chronic strain and negative life events were associated with decreases in the number of different types of delinquency in which a person reported engaging. This is consistent with GST, which posits that desistance is partially due to a decrease in strain. Eitle's study also provided some support in favor of GST explanations of stability: The effect of chronic strain on delinquency was exacerbated for individuals with an angry disposition, a trait similar to negative emotionality. Other predictions from GST were not supported. None of the measures of personal and social resources, including social support, moderated the effect of changes in strain on changes in delinquency nor did peer delinquency. Therefore, there was no support for the idea that adolescents desist from crime because they are better able to cope with strain in a prosocial manner or because factors that promote antisocial coping diminish over time. Moreover, strain was found to decrease during adolescence, rather than increase.

Finally, a study using data from the Great Smoky Mountains Study of Youth that covered ages 9 through 16 found support for GST explanations of intra-individual patterns of offending. Youth who experienced persistent poverty were more likely to have conduct disorder throughout the study period, while those who could be classified as desisters only experienced transitory poverty. Some of the findings are also consistent with the idea that stress is produced through a dynamic process, such as stress proliferation. Specifically, changes in the family configuration were associated with residential instability, suggesting that strain in one life domain can have a causal effect on strain in a different domain.

In summary, research suggests that exposure to strain is determined, in part, by characteristics of the individual. Although there are a number of studies that examine conditioning mechanisms, the results are too inconsistent to reach a definitive conclusion on the empirical validity of GST regarding this issue. In comparison, studies using GST to explain within-individual patterns of offending have been limited in quantity and scope, but they offer some support for this theory.

Open Questions and Controversies

There are a number of unresolved issues regarding GST and individual differences in exposure and reaction to strain. Perhaps the most pressing issue is that of moderating effects. Studies that examine whether the effect of strain on offending is contingent on the characteristics of the individual have produced markedly inconsistent findings. As Agnew (2006b) suggests, this might be partially due to the difficulty of detecting interaction effects in survey research. One solution to this problem is to use research designs that are better suited for detecting interactions (Agnew 2006a). For example, an experiment could be conducted in which the treatment group participated in a program intended to reduce stress exposure and reactivity by improving problem-solving skills, increasing social control and social support, and reducing exposure to delinquent peers. GST would be supported if, compared to the control group, individuals in the experimental group experienced fewer strains and were less likely to respond to the strains they experienced with delinquency.

Another potential explanation for the mixed support for conditioning effects is that the importance of specific moderators may vary based on the characteristics of the population being studied, the types of strain that are measured, and the outcome of interest. For example, the relationship between strain and offending might be stronger for youth with delinquent peers, but not for adults for whom friends take on a less important role in offending (Tittle et al. 2008). Therefore, to fully understand the contingent relationship between strain and offending, researchers might need to explore not just two-way, but also three-way interactions. The decision regarding what variables to explore in these complex interactions should be guided primarily by theoretical expectations, but might also be informed by the empirical literature on conditioning effects. This would require a systematic review of the studies that have tested hypotheses on why some people, but not others, react to strain with crime so that researchers can better specify the conditions under which individual differences are likely to moderate the effect of strain. A meta-analysis of the GST literature would be one way to conduct such a review.

It is also necessary to consider that important moderators may exist that have not yet been identified by GST or that have been identified but not well studied. Tittle et al. (2008) suggest that criminal opportunity is a potential moderator that has not received enough attention. Other potential moderators include past experiences with strain (Slocum 2010a) and coping styles (Aseltine et al. 2000).

Given the centrality of negative emotionality and low constraint to both between- and within-individual explanations of offending, understanding the role of these traits in shaping the strain-offending link is crucial; however, there remain a number of unresolved issues. First, most studies have combined negative emotionality and low constraint into one composite measure. Other studies have used measures that capture particular aspects of these traits, but not the full range of characteristics, such as angry disposition and low self-control. Therefore, it is not known if negative emotionality, low constraint, and related traits exert independent effects or whether they interact to produce even greater reactivity to strain. Second, although research on stress has found that individuals with traits similar to negative emotionality and low constraint experience more stressors, little work has examined whether strain mediates the effect of these traits on offending. Answering this question would provide a more complete picture of the relationship between traits, strain, and criminal coping.

More generally, the etiology of strain is typically overlooked in research on GST. Studies that do consider the sources of strain tend to focus on how strain emerges from poverty, family instability, or the neighborhood environment (e.g., Foster et al. 2010). While this work is certainly important, exposure to strain is also affected by access to personal resources as well as a person's actions and interactional styles, which partially originate from individual differences. As an emerging body of research (see Foster et al. 2010; Slocum 2010b) is beginning to make evident, strain is also produced through dynamic processes, such as stress proliferation. Given the importance placed on chronic strain in both GST and the larger stress literature, additional research is needed to understand the manner in which strain can restructure the life course to produce additional strain. Studying the etiology of strain will help to provide a more complete understanding of the causes of offending.

It is apparent that more testing of GST is needed before it will be known if this theory can explain patterns of offending over the life span. Most studies examine the relationship between changes in strain and changes in offending. With a few exceptions, they do not assess the mechanisms through which these changes are expected to occur nor do they examine the factors that cause changes in exposure to strain and the reaction to strain. In addition, most of these studies have very limited measures of emotions and therefore, they cannot assess one of the core propositions of GST – that emotions mediate the link between strain and offending. Yet, it should be noted that some (Tittle et al. 2008) have suggested that strain may have a direct effect on offending that is not mediated by affective responses. Data that are necessary to simultaneously test these mechanisms is hard to come by, especially if the interest is in moving beyond adolescence. Moreover, examining some of the mechanisms expected to lead to continuity and change in offending may be methodologically challenging. For example, future work will need to take into account the reciprocal relationships between strain, illicit behavior, and traits, which will require the use of more complex nonrecursive models.

But it is not just empirical work that is lacking; additional theoretical consideration is needed as well. Agnew's theoretical explanations for patterns of offending over the life course are primarily intended to explain the existence of two groups of offenders: the relatively rare life-course persisters and the much more common adolescent limited offenders. Research is accumulating that suggests this typology might be too simplistic and that even chronic offenders may eventually desist from crime. GST implies that these individuals will desist from crime because they are exposed to fewer criminogenic strains and are less likely to respond to the strain they experience with criminal coping mechanisms; yet, because stability is driven by stable traits and the increasing consequences of delinquency and strain itself, it is not clear how or why these necessary shifts in stress exposure and reactivity will occur. In other words, GST explanations for continuity leave little room for change. Therefore, a broader theoretical account of desistance that extends beyond adolescence and early adulthood is needed.

Conclusion

GST is a broad theory that incorporates concepts from a wide range of theoretical perspectives. Therefore, in order to understand patterns of offending, both across individuals and over time, it is not enough to focus only on strain. Rather, researchers must consider the characteristics of the individual and his or her social environment as well as the complex set of interactions that occur between these two spheres. People are not passive recipients of their circumstances, but instead make decisions or engage in behaviors that shape their exposure to strain. Their decisions and actions are shaped by their social world, past history, and family background, but also by their individual characteristics. These same factors also affect reactions to strain.

These complexities have made GST a difficult theory to test empirically. Examinations of individual differences in reactivity to strain have been hampered by the inherent difficulties associated with finding interactions using survey data. Studies of within-individual patterns of strain and offending have been limited by a lack of longitudinal data that contains the full range of variables needed to examine explanations of stability and change. Yet it is critical for the advancement of the theory that more attention be paid to these issues as well as the etiology of strain. GST is no longer in its infancy and in order for this theory to remain at the forefront of criminology, it must demonstrate that it can account for observed patterns of offending. This will require systematic reviews of the best studies, additional tests of the theory using more appropriate research designs and longitudinal data, and perhaps further refinement of the theory.

Related Entries

- ▶ Age-Crime Curve
- ▶ Desistance from Crime
- ▶ General Strain Theory
- ▶ Group Characteristics and General Strain Theory

Recommended Reading and References

Agnew R (1992) Foundation for a General Strain Theory of crime and delinquency. Criminology 30:47–87

Agnew R (1997) Stability and change in crime over the life course: a strain theory explanation. In: Thornberry T (ed) Developmental theories of crime and delinquency. Transaction, New Brunswick

Agnew R (2001) Building on the foundation of General Strain Theory: specifying the types of strain most likely to lead to crime and delinquency. J Res Crime Delinq 38:319–361

Agnew R (2006a) General Strain Theory: current status and directions for future research. In: Cullen FT, Wright JP, Blevins KR (eds) Taking stock: the status of criminological theory. Transaction, New Brunswick

Agnew R (2006b) Pressured into crime: an overview of General Strain Theory. Roxbury Publishing Company, Los Angeles

Agnew R, Brezina T, Wright JP, Cullen FT (2002) Strain, personality traits, and delinquency: extending General Strain Theory. Criminology 40:43–72

Aseltine RH, Gore S, Gordon J (2000) Life stress, anger and anxiety, and delinquency: an empirical test of General Strain Theory. J Health Soc Behav 41:256–275

Caspi A, Moffitt TM, Silva PA, Stouthamer-Loeber M, Krueger RF, Schmutte P (1994) Are some people crime-prone? Replications of the personality-crime relationship across countries, genders, races, and methods. Criminology 32:163–195

Eitle D (2010) General Strain Theory, persistence, and desistence among young adult males. J Crim Justice 38:1113–1121

Foster H, Nagin DS, Hagan J, Angold A, Jane EJ (2010) Specifying criminogenic strains: stress dynamics and conduct disorder trajectories. Deviant Behav 31:440–475

Gunnison E, Mazerolle P (2007) Desistance from serious and not so serious crime: a comparison of psychosocial risk factors. Crim Justice Stud 20:231–253

Hay C, Evans MM (2006) Violent victimization and involvement in delinquency: examining predictions from General Strain Theory. J Crim Justice 34:261–274

Hoffmann JP, Cerbone FG (1999) Stressful life events and delinquency escalation in early adolescence. Criminology 37:343–373

Mazerolle P, Maahs J (2000) General Strain and delinquency: an alternative examination of conditioning influences. Justice Q 17:753–778

Miller JD, Lynam D (2001) Structural models of personality and their relation to antisocial behavior: a meta-analytic review. Criminology 39:765–795

Moon B, Morash M, McCluskey CP, Won Hwang H (2009) A comprehensive test of General Strain Theory: key strains, situational- and trait-based negative emotions, conditioning factors, and delinquency. J Res Crime Delinq 46:182–212

Ousey GC, Wilcox P (2007) The interaction of antisocial propensity and life-course varying predictors of delinquent behavior: differences by methods of estimation and implications for theory. Criminology 45:313–335

Pearlin LI, Aneshensel CS, LeBlanc AJ (1997) The forms and mechanisms of stress proliferation: the case of AIDS caregivers. J Health Soc Behav 38:223–236

Peter T, LaGrange TC, Silverman RA (2003) Investigating the interdependence of strain and self-control. Can J Criminol Crim Justice 45:431–464

Rutter M, Dunn J, Plomin R, Simonoff E, Pickels A, Maughan B, Ormel J, Meyer J, Eaves L (1997) Integrating nature and nurture: implications of person-environment correlations and interactions for developmental psychology. Dev Psychopathol 9:335–364

Slocum LA (2010a) General Strain Theory and continuity in offending over time: assessing and extending GST explanations of persistence. J Contemp Crim Justice 26:204–223

Slocum LA (2010b) General Strain Theory and the development of stressors and substance use over time: an empirical examination. J Crim Justice 38:1100–1112

Slocum LA, Simpson SS, Smith DA (2005) Strained lives and crime: examining within-individual variation in strain and offending in a sample of incarcerated women. Criminology 43:1067–1109

Taylor SE, Aspinwall LG (1996) Mediating and moderating processes in psychosocial stress: appraisal, coping, resistance, and vulnerability. In: Kaplan HB (ed) Psychosocial stress: perspectives on structure, theory, life-course, and methods. Academic, San Diego

Tittle CR, Broidy LM, Gertz MG (2008) Strain, crime and contingencies. Justice Q 25:283–312

Individual-Based Modelling

► Agent-Based Assessments of Criminological Theory

Individual-Based Simulation Modeling

► Agent-Based Modeling for Understanding Patterns of Crime

Individualization

► Identification Technologies in Policing and Proof

Inferential Crime Mapping

Michael Townsley
School of Criminology and Criminal Justice,
Griffith University, Brisbane, QLD, Australia

Synonyms

Geographical analysis; Geography of crime; Spatial approaches to crime

Overview

Crime mapping has become extremely popular as an analytic technique in the last 20 years and is arguably the default mode of analysis for law enforcement officers. Spencer Chainey (this volume) explores the driving forces behind this rise as well as what is current best practice. A more detailed treatment can be found at Chainey and Ratcliffe (2005).

The purpose of this entry is to outline a number of ways that the practice of crime mapping could be improved. Many analysts receive incomplete training in spatial analysis, and most in-house training consists of how to interrogate the various crime databases with very little or no time spent on theories of crime and criminality nor *how* to conduct analysis (generating and testing hypotheses, drawing inferences). There is currently a paucity of guidance on how to construct the analysis endeavor with notable exceptions being Ekblom (1988), Weisel (2003), and Hirschfield (2005). Two recent examples of hypothesis testing in crime analysis have been provided by Townsley, Mann, and Garrett (2011) and Chainey (2012).

A lack of analytical training has grave consequences for the quality of intelligence products. Without guiding principles, it is very easy to "uncover" patterns and relationships in data that are spurious. For instance, Diana Duyser, of Florida, found an image of the Virgin Mary in her grilled cheese sandwich (BBC 2004). While it is possible that this and other images are demonstrations of divinity, another explanation is *pareidolia*, when vague and random stimuli are perceived as clear and important. Sagan (1995) speculated that pareidolia may be a side effect of evolution as facial recognition is instinctive in humans.

A more germane example to law enforcement is the birthday problem: how many people would have to be in a room before you are confident that two people present share the same birthday? The answer is 23 for a 50 % chance and 57 to be 99 % confident (Saperstein 1972). Most people find this a surprisingly small number for a small initial probability. Yet intelligence analysts regularly construct social network graphs of individuals of interest trying to linking these to other individuals of interest. But the birthday problem demonstrates it is reasonably easy to establish connections once sufficient individuals are in the group. What no analyst can say, or even thinks to ask, is what they would expect to find for someone who has no connection to the criminal underworld.

Operational law enforcement relies on analysts that interpret the criminal environment, produce intelligence products that influence decision-makers, who then take action to impact the criminal environment (Ratcliffe 2008). The realities of operational law enforcement are that decisions need to be made with incomplete data, in imperfect conditions, and under significant time pressure. The intention of this entry is to offer broad guidelines to analysts on how to better construct hypothesis testing and draw inferences. Without a systematic approach to assessing evidence, there seems a high chance of misinterpretation. This entry contains five fundamental principles of analysis, influenced heavily by statistical reasoning and inspired by Dawid (2005). Statistical reasoning is the focus here because it is science of making decisions in the face of uncertainty.

Statistical Principle 1: Frequencies Versus Rates

A basic tenet of scientific enquiry is analysis using *rates* instead of frequencies. A rate is a frequency adjusted for the underlying population at risk. They are considered superior to count data because frequencies do not reflect the different exposure "at-risk" populations face. For instance, more crime is committed in the USA compared to the UK, but when the population of each country is controlled for, the USA has much a lower crime rate than the UK for nearly all crime classifications except murder (Langan and Farrington 1998).

In the context of crime, there are two underlying problems with the application of rates for measurement. First, the frequency of crime has tangible meaning for practitioners. For instance, while the property crime rate for the City of Omaha, Nebraska, is roughly the same as the City of Philadelphia, Pennsylvania (3730.7 and 3708.2 property crimes per hundred thousand population, respectively), their very different populations (464,628 versus 1,558,378, respectively) mean that the City of Philadelphia hosts roughly 3.3 the number of property crimes of the City of Omaha (FBI 2011). No one would suggest that both jurisdictions require the same level of resources for operational purposes, yet the crime levels, once the at-risk population is accounted for, are practically identical.

Second, locating valid denominators for crime rates is challenging because the population at risk is far from clear for many crime types. Criminal justice researchers are most familiar with crime incidence, the number of crimes per hundred thousand population, but residential population is not appropriate for certain crime types. Central business districts, night-time entertainment precincts, sporting arenas, and tourist locations have very low residential populations but have the potential to host large volumes of crime.

A related issue is the temporal relevance of the denominator. Vehicle crime has almost no connection with residential population; Rengert (1997), for instance, shows that the distribution of vehicle crime in Philadelphia can be explained by the time of day and not by residential population. At a larger scale, seasonal variation in crime rates (McDowall et al. 2011; Baumer and Wright 1996) suggests different limitations with conventional indicators of the population at risk.

A review by Harries (1981) found three main ways of reporting crime: (1) frequencies, (2) population-based rates, and (3) risk-related opportunity. As to be expected with any list of alternative measures, each has advantages over the others in particular circumstances. Frequencies make sense in an operational context or for highly focused, local research efforts. Population-based rates, where the residential population is used as the denominator, are suitable for interpersonal crimes at areal units around city size (Pyle and Hanten 1974). Units of analysis smaller than this do not typically have residential populations estimated with sufficient precision.

Opportunity-based rates use the population at risk as the denominator in rate calculation. For instance, to compute a commercial burglary rate the number of retail outlets could be used as the denominator. While there may be a strong correlation between residential population and shops, the scope of a crime problem is heavily influenced by the available targets in an area. As noted by Phillips (1973, p. 224, emphasis added) [] "[o]ne thing is certain, true risk-related crime rates reveal a far different pattern than had been previously described and *are a necessary starting point in any geographic research concerning crime* in urban areas."

Opportunity-based rates are difficult to calculate due to the wide variety of alternate denominators that could be chosen for any given crime type. Moreover, there are considerable logistical difficulties in compiling these indicators. To estimate vehicle crime rates, Boggs (1965) used the amount of parking space as a proxy for the number untended cars, the latter an impractical measure for routine crime analysis.

Another issue with opportunity-based crime rates is they largely represent the number of available targets, and R. V. Clarke (1984) argues convincingly that targets are not created equal and a simple count of targets does not correspond to opportunity. A number of crucial factors may operate to alter a given target's suitability, such as ease of exploitation, attractiveness, and defended perceptions. Moreover, available targets express different level of vulnerability over the course of a day (see J. H. Ratcliffe (2006) for a theoretic and empirical justification).

One of the methodological problems facing most police organizations is that information systems record crime incidents as point level information, with precision measured in meters when accurate property parcel databases are used. Rate calculation requires some level of aggregation to cruder levels of resolution (e.g., streets, beats, communities). For example, the following figure is taken from the Safe as Houses project in Enfield, UK (Enfield 2011), a finalist in the Herman Goldstein Problem-Oriented Policing Awards in 2011. Problem analysis showed concentrations in particular sections of the borough, one of which is depicted in Fig. 1. It shows frequencies of burglaries for electoral wards around the A10 motorway. The analysts inferred that because the crime count is much higher around the A10 that these properties face an increased risk of victimization, consistent with the empirical literature that more accessible streets host more crime because they are exposed to offenders at a greater rate than less accessible streets (Beavon et al. 1994; Johnson and Bowers 2010).

There are two issues with this figure and the inference drawn. First, it is impossible to gage the underlying population at risk. The difference in crime counts may simply be a reflection that there are many more residential dwellings in the electoral wards around the A10. Second, the Modifiable Area Unit Problem (MAUP) (Openshaw 1983) may be operating. MAUP is a pernicious problem in spatial analysis, occurring when arbitrary spatial units (such as administrative boundaries) are used to aggregate data. Different boundaries, especially when they have no connection with the phenomenon of interest, produce contrasting spatial patterns. Gerrymandering is a powerful illustration of the impact of the MAUP.

This project was selected because it is unusually thorough and is clearly the product of tremendous effort. The fact that it was a finalist is evidence that this represents an atypically nuanced and empirical treatment of a crime problem. The point is that if basic problems can be identified in an exemplar case study, then how prevalent are they in less rigorous analytic products?

There are a number of recent examples of opportunity-based rates in crime mapping. Chainey and Desyllas (2008) compute the pedestrian population in various London streets to generate assault rates per hour per street segment. This revealed that some areas that have a notorious reputation were actually some of the safest seats in London. A similar type of analysis, although used for a completely different purpose, was conducted by Kurland and Kautt (2011). They used street population data to quantify the level of crime around football stadiums in the UK. Ambient pedestrian population measures from LandScan (http:/www.ornl.govscilandscan) were used to estimate population at risk on certain days and times.

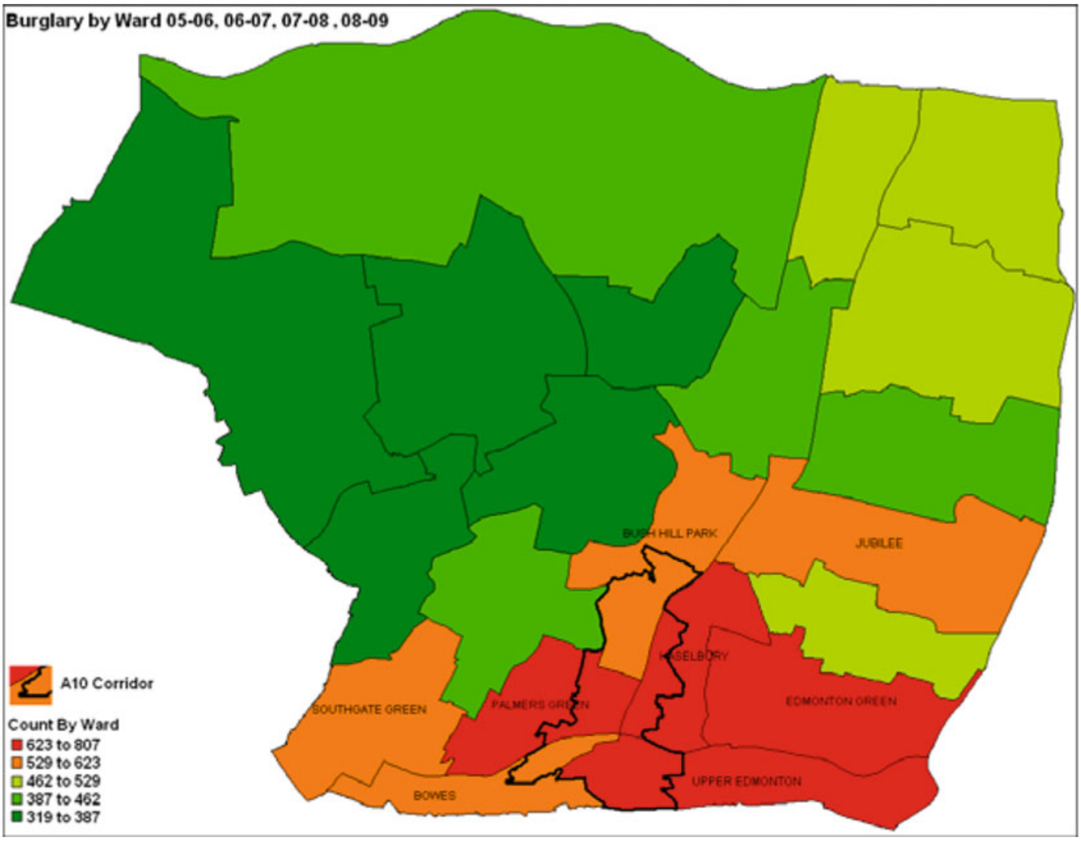
Burglary by Ward 05-06, 06-07, 07-08 , 08-09
A10 Corridor
Count By Ward
🟥 623 to 807
🟧 529 to 623
🟨 462 to 529
🟩 387 to 462
🟩 319 to 387
BUSH HILL PARK
JUBILEE
HASELBURY
EDMONTON GREEN
SOUTHGATE GREEN
PALMERS GREEN
UPPER EDMONTON
BOWES

Inferential Crime Mapping, Fig. 1 Counts of Burglary by electoral ward (Source: Safe as houses project)

Statistical Principle 2: Making Comparisons

The next foundational principle of statistical reasoning is *making comparisons*. Comparative information features prominently in statistical reasoning, from baseline measurements, to the use of control groups, even to inferential testing (the null hypothesis provides an expectation and this is compared with reality; test statistics quantify the extent of differences in this comparison).

Crime figures, used in isolation, are meaningless without reference to a comparison area or some baseline crime level because they cannot be put into perspective. The newspaper headline declaring “Over 40 robberies this year” implies this is a high volume of incidents, but without reference to other statistics it is difficult to interpret the claim with confidence.

A common comparison choice is to use an area as its own “control”; that is, the area’s history serves as a baseline indicator to assess current performance. This is analogous to a *repeated measures design*, a type of analysis where a cohort is measured on more than one occasion. The advantages of this design are beyond the scope of this entry, but it has a number of highly desirable inferential features. The main point is that each observation in the sample is compared with itself to detect a change.

Even if a baseline period was established, a confounding issue is *regression to the mean*, a tendency for an area with an extreme score at one point in time to have a score much closer to the average on the next measurement period. This is a frequent problem in crime prevention where areas selected for attention because they contain a hot spot. When regression to the mean occurs, it is natural for analysts, decision-makers, and researchers to credit an intervention or tactical responses with curbing or suppressing a rising crime trend, after all a reduction was expected as a result of the tactics employed.

Another popular method of conducting comparisons is to use a *comparison group*. Pawson and Tilley (1997) provide a persuasive note of caution about their value. The argument is that interventions trigger mechanisms that fire in particular contexts. An intervention is designed explicitly for the particular area in question, yet comparison groups are conventionally identified through matching demographic and socioeconomic indicators, not the context. Two areas may be strikingly similar with respect to aggregate census indicators yet have neighborhood dynamics which would render some interventions infeasible.

The importance of making comparisons is reasonably straightforward, and spatial researchers could easily argue that crime maps actually possess an inherent comparison; crime maps contain simultaneous displays of hot and cold spots. However, this does not entirely capture the statistical principle of making comparisons. A map of crime density displays the *where* of crime, serving simply as a form of analytical triage indicating where attention should be placed. What next? What should law enforcement attempt to do to the criminal environment to change to reduce the level of crime?

The overarching goal of analysis is to make some causal claim or inference. For example, poor place management leads to more assaults is a causal statement that links the presence or absence of place management with varying degree of assaults. Statisticians would formally label poor place management as an *independent variable* (sometimes explanatory variable) and assaults as the *dependent variable*. The volume of assaults is said to "depend" on the amount of place management.

In order to *make comparisons*, analysts need to demonstrate that the spatial pattern of some independent variable is closely aligned with the observed pattern of crime incidence (the dependent variable). Explanatory factors are rarely a feature of crime mapping products (Eck (1997) provides a rare example of crime mapping that fully incorporates theory).

Statistical Principle 3: Retrospective Versus Prospective

Many risk factors are expressed in terms of a *retrospective proportion*. This is where the prevalence of risk factors among an interest group (victims or offenders, say) is computed. For instance, a study claims that 80 % of sexual assaults take place on the public transport system. The inference drawn is that there is a high chance of victimization if you use public transport. Notice the shift in emphasis between the data and the inference:

- If victimized, there is a high chance of using public transport (data).
- If using public transport, there is a high chance of being victimized (inference).

It is clear there is a difference between these two statements, yet the vagaries of language make it easy to slip between these two distinct meanings. In statistical parlance, each of these statements is a *conditional probability* of the form, $P(A \mid B)$ where the probability of outcome A is conditioned on (or constrained by) the necessary occurrence of B. The conditioning event is always located to the right the "|" symbol.

To clarify, a conditional probability is the probability of event A occurring if B also occurs. A simple form of inference is to compare conditional probabilities with simple probabilities, probabilities of a single event – $P(A)$. The extent of difference between simple and conditional probabilities implies an association between A and B. Of course, this just establishes an association, a type of correlation. This is not the same as claiming that A *causes* B. In order to do this, the analyst needs to demonstrate (1) the temporal sequencing such that a change in B corresponds to a change in A and (2) dismissing competing explanations. The statements can be expressed in formal statistical terms:

- $P(PT \mid V)$ the probability of public transport use among victims (data)

Inferential Crime Mapping, Table 1 Frequencies of victimization status versus public transport use

	Victimized	Not victimized	Total
Public transport	80	10,000	10,080
Not public transport	20	11,000	11,020
Total	100	21,000	21,100

- $P(V \mid PT)$ the probability of victimization among public transport users (inference)

These are two entirely different quantities as demonstrated in the following table.

It is reasonably straightforward to see that of those people who have been sexually assaulted, 80 % (=80,100) were on public transport. It is equally easy to see that for those people who are not sexually assaulted, 48 % (1,000,021,000) used public transport. These suggest an association between victimization and public transport and are consistent with the first two statistical principles – these quantities are rates and comparisons are being made. Importantly, these quantities relate to the likelihood of public transport patronage for victims and non-victims.

Unfortunately there is a further problem with retrospective proportions as they give an inflated sense of the association. The statistic of immediate interest is the likelihood of victimization among public transport users, the *prospective proportion*. Referring to Table 1, this prospective proportion is .8 % (=8,010,080). The corresponding percentage for nonpublic transport users is .2 % (2,011,000).

Two points should be obvious. First, prospective proportions maintain the same order as their retrospective counterparts. The association between public transport and victimization is preserved such that the presence of public transport does correspond to a higher victimization rate regardless of which type of proportion is used.

Second, the prospective proportions are considerably smaller in magnitude than the retrospective proportions, yet this is of more relevance. The problem is that retrospective proportions create the impression that there is a high chance of crime, yet they quantify the likelihood of observing the risk factor (public transport), not the outcome measure (crime). The prospective proportion is about the likelihood of crime.

Analysts need to be aware that due to the nature of crime, they will invariably be calculating retrospective proportions about crime risk factors. Following individuals prospectively when the event of interest is rare or has low frequencies is time and cost prohibitive. Instead, it is far simpler to take a sample of individuals that have observed the event of interest and examine their antecedent attributes. These are useful but are limited in their inferential insight. Retrospective proportions will point to associations, but the strength of those relationships is not immediately apparent. As seen in the example here, quoting the retrospective proportion implies that victimization is a near certain occurrence if one was foolish enough to use public transport, yet the prospective proportion seems far, far less alarming.

Statistical Principle 4: Selection Bias

This principle reflects the observation that forces influencing the data collection process shape the results of analysis performed on the data. Selection bias occurs when the process of data collection is non-probabilistic. Celebrated examples of selection bias occur in election polling (Squire 1988; Jowell et al. 1993), leading to inaccurate estimates. Increased use of mobile telephones has further compromised telephone polling, with some households having no landline telephone (Keeter 2006).

Selection bias occurs naturally whenever *secondary data analysis* is performed. This is a well-recognized problem that accompanies *all* crime analysis. Maguire (2012) writes in the most recent volume of the Oxford Handbook of Criminology of the differences between victimization surveys and official statistics. Commonplace problems – which offences are reported to police, which get recorded by police – act as filters on the information contained in police information system. The cumulative impacts of these and other recording practices are to skew the crime analyst's depiction of the criminal environment.

With respect to spatial analysis, there are a number of additional pressures on crime data which may further impact selection bias. J. H. Ratcliffe (2001) writes that the process of *geocoding*, where spatial reference information is assigned to crime event data, could be faulty due to a number of reasons:

- *Out of date cadastral map*. Cadastral maps are a register of property parcels in a region. They contain validated, government information about spaces. Cadastral information is very useful for geocoding because street addresses can be linked to known land survey information. However, if this information is out of date, such as buildings have been demolished or property parcels have been combined to form larger spatial units, then it is likely that geocoding will be inaccurate if applied to new spatial units that do not appear in the cadastral map.
- *Abbreviations and misspellings*. Reported crime may be misspelled, mispronounced, or incomplete in some way, such as an absent street number. Under these conditions, geocoding is unlikely to operate correctly, with addresses either failing to find a match or being mismatched to a separate address.
- *Local name variations*. Certain places develop a nickname or local business name which may not appear in an address database. Geocoding works on known street addresses, which victims may be oblivious of if they are not an employee or local community member. Tourists are often unaware of street addresses of locations they visit, for instance.
- *Address duplication*. Certain street names are more common than others and, without ancillary suburb or community level information, can easily be placed in another area.
- *Nonexistent addresses*. Mistakes in data entry can easily modify the reported address so it is positioned out of range of a street segment. For example, 18 Baker Street can be transcribed as 80 Baker Street.
- *Non-addresses*. Public space is a notoriously difficult to record and/or capture reliably, and there are a number of locations that may not exist in a cadastral map. For instance, public parks may be defined as a polygon of the geographic extent of the park, yet it is difficult to see how any geocoding engine could distinguish where any incident would have occurred. In most cases, anything occurring in the park is "recorded" at a single location for the entire park such as the centroid of the polygon.

Bichler and Balchak (2007) show that address matching procedures in the major mapping software applications are prone to distinctive systematic biases in geocoding errors. J. H. Ratcliffe (2001) demonstrates geocoding inaccuracy of between 5 % and 7 % in a dataset where crimes recorded in an incorrect *census tract*. This level of inaccuracy seriously undermines any spatial analysis performed at all but the crudest of spatial resolutions.

Statistical Principle 5: Simpson's Paradox

This principle operates when patterns of rates (or proportions) calculated for an entire sample are not consistent for patterns for subgroups of the data. The paradox is a reflection of changing denominators in crime rates and is a result of only relying on proportions or rates as indicators of activity. The contradiction is that Simpson's Paradox directly contradicts the first principle listed here. Of course, it really underlines the importance that counts *and* rates need to be incorporated in analysis.

To illustrate, consider you were responsible for two policing jurisdictions and the crime rates for three priority crimes were as listed in the following table. Being a diligent analyst you compute rates for each of these based on appropriate denominators. It appears that Area 2 is safer than Area 1, in aggregate (Table 2).

Inferential Crime Mapping, Table 2 Aggregate crime rates for Areas 1 and 2

	Area 1	Area 2
Total	7.75	7.20

Inferential Crime Mapping, Table 3 Subgroup crime rates for Areas 1 and 2

Crime type	Area 1 freq	Area 1 denominator	Area 2 freq	Area 2 denominator	Area 1 crime rate	Area 2 crime rate
Assault	256	41,250	430	54,000	6.21	7.96
Commercial burglary	178	2,800	30	350	6.36	8.57
Car theft	69	20,850	66	18,750	3.31	3.53
Total	503	64,900	526	73,100	7.75	7.20

Given that Area 2 has the lower crime rate, it would be worth considering what examples of best practice might be transferred to Area 1. Exploring differences across different crime types will aid in identifying areas worth targeting. However, examining each crime type shows that Area 2 has a *higher* crime rate for all crimes (Table 3).

The reason this apparently contradictory result occurs is due to differences in the denominator that are used to compute the rates as well as a reliance on only rates to assess relative risk. Examining either the "at-risk" populations or the volume of crime would indicate that the two areas have certain fundamental differences which may need to be incorporated in addition to considering crime rates.

This principle is a reflection that variation is unequally distributed among locations and facilities. Aggregate summaries of crime may not be representative at finer levels of resolution. For instance, Brantingham, Dyreson, and Brantingham (1976) provide a nice depiction of the changing nature of crime patterns viewed at different levels of resolution. National patterns differ greatly from regional patterns which are different still from local patterns.

Even the use of sophisticated modeling techniques does not protect research findings from Simpson's Paradox. Roncek and Maier (1991) found a positive association between the presence of licensed venues and the volume of crime. They use the technique which estimated the average impact of a bar on an average block. However, the frequency of crimes for individual bars shows that the effect of an "average" bar is atypical and that only a minority of bars account for the majority of alcohol-related crime and instabilities (Eck et al. 2007).

Spatial Principles

The five principles outlined pertain to any quantitative analysis. Spatial data and analyses in particular have a number of unique attributes that need to be controlled for, or at least accounted, in order to provide valid findings. These are mentioned briefly here:

- *Modifiable area unit problem* (Openshaw 1983). This is an issue that occurs when point level information is aggregated to areal spatial units that are defined by arbitrary administrative boundaries. If the boundaries used do not have a clear alignment with the social phenomena in question, it is possible to generate a depiction of crime that is a reflection of the properties of the boundaries. For instance, a hot spot located close to the border between two police beats will appear diluted if the hot spot is bisected by the border of the police beats. Using a different set of spatial units will probably "undercover" this hot spot.
- *Spatial autocorrelation*. Tobler's first law of geography states observations that located close tend to be more similar to each other than far observations (Tobler 1987). This is a problem because it suggests that observations that are proximate are not statistically independent. This is a fundamental assumption of many inferential techniques in hypothesis testing. Michael Townsley (2009) provides a list of techniques useful for incorporating spatial autocorrelation in analysis or investigating whether is present.

Conclusion

The purpose of this entry has been to outline fundamental principles of statistical reasoning with a view to informing operational crime analysts about better ways of approaching pattern detection. The target audience for this entry may very well ask exactly how should analysis be conducted in order to observe these principles. There are three main strategies:

1. *Be more scientific*. The discipline of crime science has emerged in the last decade and is focused on ways to make crime harder to commit and how to catch offenders quicker. It emphasizes that practitioners should think more scientifically, generate hypotheses, test them, and collect high-quality data. One of the missing gaps in this literature is an acknowledgement of how to conduct analysis in a scientific way. Recent studies by M. Townsley, Mann, and Garrett (2011) and Chainey (2012) put forward methods and examples of how to carry this out.
2. *Employ more sophisticated methods*. One of the advantages of more sophisticated statistical techniques is that many of the principles of statistical reasoning are "baked in," especially regression models. There needs to be much more investment in increasing the training of analysts so that they can exploit well-developed quantitative methodologies. If crime science is analogous to medical science, at present there is no analogue to epidemiology in the study of crime.
3. *Be more focused and use crime theories*. Analysts need to define their problem precisely, restricting their attention to a finite set of behaviors to analyze. For instance, analysts should not study assaults by criminal classification (aggravated assaults, grievous bodily harm, assault with a weapon); rather, they should focus on the relationship between the parties involved. These will range typically from spouses or partners (domestic violence), road rage incidents, school pupils, bar patrons and door staff, and strangers. Each of these assault types has a unique opportunity structure and probably different space-time signatures. Failure to partition each of these discrete problems means that aspects of each are consolidated into an aggregate uninformative blob.

Related Entries

- ▶ Agent-Based Models to Predict Crime at Places
- ▶ Applied Geographical Profiling
- ▶ Biological Geographical Profiling
- ▶ COMPSTAT
- ▶ Crime Location Choice
- ▶ Crime Science
- ▶ Criminal Investigative Analysis
- ▶ Early Chicago School Theory
- ▶ Effectiveness of Situational Crime Prevention
- ▶ History of Geographic Criminology Part I: Nineteenth Century
- ▶ Hot Spots and Place-Based Policing
- ▶ How to Make Crime Mapping More Inferential
- ▶ Longitudinal Crime Trends at Places
- ▶ Longitudinal Studies in Criminology
- ▶ Near Repeats and Crime Forecasting
- ▶ Outdoor Serious Violence: The Role of Place
- ▶ Prediction and Crime Clusters
- ▶ Predictive Policing
- ▶ Problem-Oriented Policing
- ▶ Rational Choice Theory
- ▶ Routine Activities Approach
- ▶ Spatial Models and Network Analysis
- ▶ Spatial Perspectives on Illegal Drug Markets

Recommended Reading and References

Baumer E, Wright R (1996) Crime seasonality and serious scholarship: a comment on Farrell and Pease. Brit J Criminol 36:579–581

BBC (2004) Woman 'blessed by the holy toast'. http://news.bbc.co.uk/2/hi/americas/4019295.stm

Beavon DJK, Brantingham PL, Brantingham PJ (1994) The influence of street networks on the patterning of property offenses. In: Ronald V, Clarke G (eds) Crime prevention studies, vol 2. Criminal Justice Press, Monsey

Bichler G, Balchak S (2007) Address matching bias: ignorance is not bliss. Pol Int J Police Strat Manag 30:32–60

Boggs SL (1965) Urban crime patterns. Am Sociol Rev 30:899–908

Brantingham PJ, Dyreson DA, Brantingham PL (1976) Crime seen through a cone of resolution. Am Behav Sci 20:261–273

Chainey S (2012) Improving the explanatory content of analysis products using hypothesis testing. Policing 6 (2):108-121

Chainey S, Desyllas J (2008) Modelling pedestrian movement to measure on-street crime risk. In: Liu L, Eck JE (eds) Artificial crime analysis systems: using computer simulations and geographic information systems. Idea Group, Hershey, pp 71–91

Chainey S, Ratcliffe JH (2005) GIS and crime mapping. Wiley, Chichester

Clarke RV (1984) Opportunity-based crime rates: the difficulties of further refinement. Brit J Criminol 24:74–83

Dawid PA (2005) Probability and proof. In: Anderson TJ, Schum DA, Twining WL (eds) Analysis of evidence. Cambridge University Press, Cambridge

Eck JE (1997) What do those dots mean? Mapping theories with data. In: Weisburd DL, McEwen T (eds) Crime mapping and crime prevention, vol 8. Criminal Justice Press, Monsey, pp 377–406

Eck JE, Ronald V, Clarke G, Guerette RT (2007) Risky facilities: crime concentration in homogeneous sets of establishments and facilities. In: Farrell G, Bowers KJ, Johnson SD, Townsley M (eds) Imagination for crime prevention: essays in honour of Ken Pease, vol 21. Criminal Justice Press, Monsey, pp 225–264

Ekblom P (1988) Getting the best out of crime analysis. Crime prevention unit. Home Office, London

FBI (2011) Crime in the United States by metropolitan statistical area, 2010 (Table 6)."

Harries KD (1981) Alternative denominators in conventional crime rates. In: Brantingham PJ, Brantingham PL (eds) Environmental criminology. Sage, Beverly Hills, pp 147–165

Hirschfield A (2005) Analysis for intervention. In: Tilley N (ed) Handbook of crime prevention and community safety. Willan, Cullompton, pp 629–673

Johnson SD, Bowers KJ (2010) Permeability and crime risk: are cul-de-sacs safer? J Quant Criminol 26:89–111

Jowell R, Hedges B, Lynn P, Farrant G, Heath A (1993) Review: the 1992 British election: the failure of the polls. Public Opin Q 57:238–263

Keeter S (2006) The impact of cell phone noncoverage bias on polling in the 2004 presidential election. Public Opin Q 70:88–98

Kurland J, Kautt P (2011) The event effect: demonstrating the impact of denominator selection on 'floor' and 'ceiling' crime rate estimates in the context of public events. In: Morina AD (ed) Crime rates, types and hot-spots. Nova, Hauppauge, pp 115–144

Langan PA, Farrington DP (1998) Crime and justice in the United States and in England and Wales, 1981–96. US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics

London Borough of Enfield (2011) Safe as houses: reducing domestic burglary project. Finalist, Goldstein POP Awards, http://www.popcenter.org/library/awards/goldstein/2011/11-09(F).pdf

Maguire M (2012) Crime data and statistics. In: Maguire M, Morgan R, Reiner R (eds) The Oxford handbook of criminology. Oxford University Press, Oxford, pp 241–301

McDowall D, Loftin C, Pate M (2011) Seasonal cycles in crime, and their variability. J Quant Criminol 28(3):389–410

Openshaw S (1983) The modifiable areal unit problem. Geo Books, Norwich

Pawson R, Tilley N (1997) Realistic evaluation. Sage, New York

Phillips PD (1973) Risk-related crime rates and crime patterns. Proc Assoc Am Geog 5:221–224

Pyle GF, Hanten EW (1974) The spatial dynamics of crime. Research Paper Series. Department of Geography, University of Chicago, Chicago

Ratcliffe JH (2001) On the accuracy of tiger-type geocoded address data in relation to cadastral and census areal units. Int J Geog Inform Sci 15:473–485

Ratcliffe JH (2006) A temporal constraint theory to explain opportunity-based spatial offending patterns. J Res Crime Del 43:261

Ratcliffe JH (2008) Intelligence-led policing. Willan, Cullompton

Rengert GF (1997) Auto theft in Central Philadelphia. In: Homel R (ed) Policing for prevention: reducing crime, public intoxication and injury. Criminal Justice Press, Monsey, pp 199–220

Roncek DW, Maier PA (1991) Bars, blocks, and crimes revisited: linking the theory of routine activities to the empiricism of hot spots. Criminology 29:725–753

Sagan C (1995) The demon-haunted world: science as a candle in the dark. Random House, New York

Saperstein B (1972) The generalized birthday problem. J Am Stat Assoc 67(338):425–428

Squire P (1988) Why the 1936 literary digest poll failed. Public Opin Q 52:125–133

Tobler WR (1987) Experiments in migration mapping by computer. Cartogr Geog Inform Sci 14:155–163

Townsley M (2009) Spatial autocorrelation and impacts on criminology. Geog Anal 41:450–459

Townsley M, Mann M, Garrett K (2011) The missing link of crime analysis: a systematic approach to testing competing hypotheses. Policing 5:158–171

Weisel DL (2003) The sequence of analysis in solving problems. In: Knutsson J (ed) Problem oriented policing: from innovation to mainstream, vol 15. Criminal Justice Press, Monsey, pp 115–146

Infiltration

► Law of Undercover Policing

Informal Guardianship

Danielle M. Reynald
School of Criminology and Criminal Justice,
Griffith University, Griffith, Australia

Overview

Guardianship is the process by which citizens function as effective informal crime prevention and control agents. The concept was first introduced in Cohen and Felson's (1979) routine activity approach and was originally defined as "any spatio-temporally specific supervision of people or property by other people which may prevent criminal violations from occurring" (Felson and Cohen 1980). In spite of the capable guardian's central role in determining criminal victimization, research focusing exclusively on guardianship and the mechanisms that facilitate and inhibit it is limited, particularly compared to that of offending and victimization. Recent research has begun to address this gap in the criminological literature to reveal that guardianship against crime is effective when citizens are available and is boosted when they engage in monitoring or supervision of their surroundings and intervene when necessary. Empirical evidence demonstrates that guardianship intensity in residential environments is a place-based characteristic, as it is critically determined by the opportunities available for supervision generated by spatio-physical and sociodemographic contextual factors.

Fundamentals

The Concept of Guardianship

Within criminology the concept of guardianship first surfaced as a critical component of Cohen and Felson's (1979) routine activity approach. This theory put forward the perspective that in order for a crime event to occur, it requires the convergence of three minimal elements: a likely offender, a suitable target (defined as a person, object, or place), and the absence of a capable guardian. In doing so, they draw attention to the capable guardian as the ultimate protector and defender of people and property against criminal violations. Felson and Cohen (1980) define guardianship as "any spatio-temporally specific supervision of people or property by other people which may prevent criminal violations from occurring." In spite of the fact that the capable guardian is one of the only actors within the crime event model who has the power to prevent criminal violations from occurring, it has been argued that "the issue of guardianship lags in theoretical development and empirical testing" (Tewksbury and Mustaine 2003, p. 203).

The Capable Guardian

In their original formulation of routine activity theory, Cohen and Felson (1979) explained crime as the product of the convergence of three minimal elements in time and space: (1) likely offenders, (2) suitable targets, and (3) the absence of capable guardians. Thus, the routine activity explanation of crime rate trends identified the capable guardian as one of the only actors within the crime event who has the power to prevent a crime from being executed. The theory defines a capable guardian as any person or thing that discourages crime from taking place (Cohen and Felson 1979). Cohen and Felson (1979) and Felson (1995) emphasize that guardians "keep an eye on potential crime targets" and serve "by simple presence to prevent crime, and by absence to make crime more likely" (Felson 1995). Felson (1995, pp. 53–54) goes on to further clarify the concept by providing examples of who capable guardians are and how they may serve to prevent crime: "a retired person at home might well discourage daytime burglary of his or her own home or even the home next door. Conversely, someone working away from home during the day contributes to a greater risk of burglary by their absence. Two persons walking down the street might serve as effective guardians for one another against a mugging or other attack, while each individual serves as a guardian for his or her own immediate property."

With these examples, Felson reveals that guardians may be household residents (who can act as guardians over their own as well as nearby properties) and/or passersby (who may act as guardians over others in their presence). Who fulfills the role of a guardian, therefore, is dependent in the first instance on who is present within the context of the crime opportunity. Felson (1995) argues that his conceptualization of the guardian within the crime event model is not a formal policeman or security guard but an informal person or ordinary citizen whose presence, during the course of their daily routines, provides security to potential targets/victims of crime. It is therefore informal guardians like residents and passersby who are more likely to be present to discourage crime than formal guardians like police officers and security guards.

Guardians and Other Crime Controllers

Recent developments in routine activity theory highlight the various roles assumed by the actors who are responsible for crime control, of which guardians represent just one. These roles have been expanded to include handlers who supervise potential offenders (Felson 1995), and managers who supervise places (Eck 1994), to accompany the original group of guardians who supervise potential crime targets (people or property) (Cohen and Felson 1979). Felson developed the concept of the "handler" by linking the routine activity approach and Hirschi's control theory to explain how parents and others in a community who have knowledge of and close proximity to a likely offender can serve as "intimate handlers" who can exert control over them if they are aware of their rule-breaking. Eck (1994) identified another group of controllers – place managers – who have the power to discourage crime at places through the effective supervision of those places. Place managers can be restaurant managers, apartment building managers, doormen, or concierges – all have the ability to keep an eye on places due to their close proximity to or frequent presence at places, as well as the responsibility they have been assigned to manage those places. These developments in the routine activity approach point to the convergent mechanisms that may come into play for those who exercise control over likely offenders, defined (micro) places, and individual people and/or property targets.

Crime Prevention and Control Through Guardianship

Guardianship and Situational Crime Prevention

Clarke (1997) highlights the importance of increasing guardianship as one of the core situational crime prevention techniques. Based on the premise that specific situational characteristics help generate opportunities for offenders to commit crime, Clarke (1997) explains that crime can be prevented by systematically manipulating situational elements in order to block these opportunities using a range of situational crime prevention strategies and techniques. Situational crime prevention is designed to affect the offender's decision-making process by manipulating the perceived costs and benefits associated with a particular crime opportunity. The aim therefore is to manipulate situations so that the costs of committing crime outweigh the benefits, thereby discouraging the offender from committing the criminal act. Thus, Clarke (1997) explains that the ultimate goal is to create unfavorable circumstances for the offender through "discrete managerial and environmental change" (1997, p. 2) in order to reduce opportunities for crime and prevent its occurrence.

With this objective in mind, one of the central strategies of situational crime prevention is to *increase the risks* for offenders to get caught. The key techniques associated with this strategy include extending or increasing guardianship, using place managers, and providing assistance for guardians or tools to facilitate natural and other forms of surveillance including improved street lighting and use of CCTV. Clarke explains that the implementation of these guardianship (and related) techniques is fundamental to crime prevention as it heightens the offender's perception that he or she will be detected and subsequently puts him or her off from offending in that situation. In short, guardians serve to discourage crime and protect targets through their presence and through supervision.

Support for the role of guardianship as an effective crime prevention technique comes from offender self-reports on factors affecting target selection. In Bennett and Wright (1984), for example, qualitative interviews with burglars reveal that factors related to guardianship in terms of household occupancy and the risk of being seen or detected by others played a significant role in the targets selected for burglary. In selecting their targets, burglars explained that they would look for signs or cues that a dwelling was occupied, in which case the dwelling would be deemed "unsuitable for burglary" (Bennett and Wright 1984, p. 65). Similarly, offenders were also concerned with the proximity of neighboring houses and the presence of neighbors who might act as guardians (Bennett and Wright 1984, p. 63). Offenders were discouraged not only by the presence and proximity of guardians in the form of neighbors but also by their capability as guardians, that is, whether offenders would be noticed by neighbors and whether neighbors would take action upon noticing their presence. They admitted being concerned about "being seen by someone familiar with the property who might intervene if something suspicious occurred" (Bennett and Wright 1984, p. 64).

Guardianship and Informal Social Control

The three groups of controllers or supervisors identified in the routine activity approach – guardians, place managers, and handlers – draw attention to the fact that crime can be prevented by controlling potential targets, crime sites, or offenders themselves through supervision. In this way, the concept of guardianship is subsumed under the umbrella of informal social control, where supervision by the guardian over the target, the handler over the offender, and the place manager over the place is enhanced by social bonds and informed by social norms, which leads ultimately to crime control. Felson and Cohen's (1980) definition of guardianship as spatio-temporally specific supervision of potential crime targets emphasizes the function of supervision – that is, looking out for, keeping an eye on, monitoring, and natural surveillance – of people or property that is likely to prevent victimization. This definition also explains that guardianship is place and time specific, highlighting that guardianship as a crime prevention technique is situational; it needs to be targeted at specific places and times in order to maximize its effectiveness.

Availability of Guardians

Citizens' capability to guard against crime is defined, in the first instance, by their routine activities as this dictates a person's availability to act as a guardian. Guardianship for crime prevention thus rests on the notion that any individual can serve as a guardian against crime by being present at the scene of a potential crime event (Felson 1995). Local residents, for example, can guard their residential communities against crime through frequent presence in their residential locale. Routine activities that involve a higher proportion of time spent at or around the home rather than away from home help increase the likelihood that guardians will be available if necessary. The mere presence of a guardian has the potential to discourage offenders who are in search of low-risk opportunities to commit crime. Availability therefore can be viewed as the primary dimension of guardianship.

This is corroborated by empirical evidence which shows that household occupancy or frequency of time people spend at home is a negative correlate of burglary victimization. Garofalo and Clarke (1992) used a survey measure of residential guardianship asking household members how often they were at home, as well as indicators such as the presence of a dog, an alarm system, a timer for lights or radio, informal arrangements with neighbors, and other security measures. Results showed that the combined measures of guardianship contribute to reducing the likelihood of residential burglaries, with the strongest effect when they are used together. Similarly, Miethe et al. (1990) operationalized guardianship in terms of household occupancy, using the number of persons in a household over the age of 12 as its measure. Results showed that individuals who had a decreased level of guardianship over time showed an increased risk of victimization. Time spent away from home not only decreases property guardianship by leaving properties unattended but also increases individual exposure to the risk of personal victimization. In some situations, more important than being present is the *evidence* of availability. "Traces of household occupancy" in the form of visual or auditory cues inform potential offenders that guardians are present and therefore are likely to play a vital role in discouraging offenders.

As such, the majority of empirical research on guardianship utilizes proxy measures for home occupancy to measure guardianship. Indicators such as the number of owner-occupied households, percentage of labor-force participation, and number of household members have traditionally been used to approximate the level of guardianship at places in lieu of direct measures. The British Crime Survey (BCS) and Garofalo and Clark (1992) provide notable exceptions, as they use self-report survey items to measure the amount of time people spend at home as an approximation of the level of guardianship at the property level. Availability, however, is just one dimension of guardianship and conveys little about actual supervision by the guardian who may be present or whether they would take action to stop a crime in progress. It has been argued that this reflects the tangible problem of the inadequacy of some of the proxy measures of guardianship that have traditionally been used in tests of the concept (see Miethe and Meier 1994, for discussion). While availability is a necessary condition for guardianship, the visible presence of guardians is merely the starting point of the spectrum of capability that citizens can draw upon to guard against crime.

Guardianship in Action: Including Supervision and Intervention

While presence/availability is arguably its most fundamental dimension, Reynald (2009, 2010, 2011a) demonstrates that capable guardianship is intensified through monitoring and intervention when necessary (*see* Cohen and Felson 1979). Given availability, effective guardianship can be enhanced by actual supervision or monitoring of residential space and property. Felson (2006) explains that the absence of actual supervision determines when and where crime occurs since offenders primarily seek to escape supervision by finding targets and places that are not well attended to. The power to guard against crime can be further intensified when guardians take action to intervene upon observing untoward activities in their surroundings. Intervention can be defined as any action taken to disrupt or prevent a crime event and can encompass a range of actions including shouting out to perpetrators, calling the police, or physically intervening to stop a crime. Intervention as a form of guarding against crime is supported by Brown and Bentley (1993) who explain that the risk of being seen is not always sufficient to deter a burglar since "burglars also need to know whether potential onlookers would care about their presence. . .Burglars also may decide that some neighbours will intervene and others will not" (Brown and Bentley 1993, p. 52).

Reynald (2009, 2011) elaborated on these action dimensions of capable guardianship, arguing that these components of supervision and intervention are directly observable and provide a more ecologically valid measure of the core dimensions of guardianship than traditionally used proxy measures. Measuring these action components of guardianship through observation revealed that guardianship intensity is strongest and is significantly associated with lower property crime rates when guardians are (1) available, (2) engage in supervision/monitoring/surveillance over targets in their surroundings, and (3) take action by intervening when necessary in order to prevent or disrupt crime or related violations. These findings were replicated by Hollis-Peel and Welsh (forthcoming). Together, availability, monitoring, and intervention have been established as a reliable and valid integrated measure of guardianship and can therefore be viewed as the dimensions of guardianship in action (GIA).

Decision Making by Guardians: Willingness to Supervise and Intervene

In support of the premise that guardianship can extend beyond the availability of guardians, an interview study with available residential guardians revealed that neither supervision nor intervention when necessary is an inevitable consequence of availability (Reynald 2010). In fact, of a sample of 255 available guardians, 15 % admitted to choosing not to supervise or pay attention to occurrences in their surroundings. Moreover, supervision was shown to enhance guardianship capabilities, as residents explained that monitoring enhanced their contextual awareness by helping them to recognize people and situations that were suspicious and out of the ordinary in their particular residential context. These findings support the results from the GIA studies showing that monitoring or supervision boosts the intensity of guardianship at places.

Consistent with the social psychology literature on bystander intervention, only 16 % of the 217 residents that reported witnessing crime or related events in their surroundings in the past also reported being willing to intervene directly, irrespective of the type of crime event they witnessed. Of equal interest was the fact that 20 % of this subsample admitted they would not take any action to intervene, either directly or indirectly, upon observing suspicious, crime-related activity. The majority of the sample reported indirect intervention by calling the police for help as their chosen response to a crime event, with many residents explaining that their choice of direct or indirect intervention would depend on the severity of the crime event in progress. Through their discussion of these varying responses to crime, some of the core stages involved in guardians' willingness to take action to disrupt crime events – that is, to supervise and/or intervene – were able to be elucidated.

Results also revealed that once a crime event is observed by an available guardian, his/her responsive action is decided through an assessment of a range of individual and situational factors that can be broadly categorized as (1) perceived risk to personal safety, (2) perceived capability, and (3) perceived responsibility. When the perceived risk to safety is low and the perceived capability and responsibility is high, available guardians expressed a greater willingness to take some form of direct or indirect action to intervene. Clarke (1997) suggests that guarding against crime within the situational crime prevention context is about encouraging guardians and place managers to take responsibility for monitoring and guarding the places they inhabit. For residents it means taking responsibility, not just for private residences, but also for public areas around residences, while for place managers like shop assistants, parking lot attendants, or train conductors, it means taking responsibility for monitoring their workplaces (Clarke 1997). Felson (1995) builds on this premise by identifying varying levels of responsibility that guardians take for guarding against crime which are determined by physical distance to potential targets/victims, as well as social distance in terms of familiarity and attachment (see Reynald 2011b).

Interviews with available guardians reveal that there are distinct decision-making stages that determine the critical action dimensions of guardianship – supervision and intervention. In particular, they suggest that once a crime event is observed, the responsive action taken by an available guardian is the product of a series of rational choices and the evaluation of the costs and benefits involved. By determining the intensity of guardianship available within particular contexts, these decisions ultimately have an indirect effect on the crime levels within defined contexts. Whether or not guardians take the necessary action to supervise and, in particular, to intervene is affected by individual, contextual, and situational factors which are often interrelated (see Reynald 2010). In order to develop this line of research, these guardianship decision-making stages need to be further extrapolated by developing an understanding of the mechanisms that facilitate intervention by guardians, as well as those that hinder it. A systematic construction of how these mechanisms interrelate to produce the decision to intervene or not is a critical first step in realizing how effective forms of intervention by guardians can be encouraged.

Guardianship and Environmental Support

The aforementioned results point to the importance of contextual factors in generating opportunities for effective guardianship. Wilcox et al. (2007) argue that guardianship is a multidimensional concept which is affected by contextual factors not appropriately considered by most of the previous studies in this area. Alongside the behaviors required to guard against crime as described in the GIA model, the local environment or the characteristics of a place play a critical role in supporting efforts by local citizens to guard against crime. Physical, spatial, social, and demographic factors all make a significant contribution to what is perceived holistically as the residential environment. In turn, each of these groups of factors affects the (visible) availability of residents, their monitoring capabilities, and whether or not they are willing to intervene to disrupt or prevent crime events. Citizens can therefore be empowered to guard against crime through opportunities that are generated by their environment.

The following section will explain which environmental factors have emerged as most influential when it comes to guarding against crime, as well as the mechanisms through which the environment and guardianship likely interrelate. Relevant theoretical perspectives will be highlighted along with empirical evidence to explain the relationship between the environment and the dimensions of guardianship depicted in the GIA model. This comparison will reveal how the same environmental cues vital for crime prevention may be conversely interpreted by the offender in search of crime opportunities.

Guardianship Opportunities and the Spatio-physical Context

Natural Surveillance

Both the visible availability of guardians and their monitoring capabilities are critically mediated by the opportunities for surveillance that are provided by the design and layout of the built environment. Newman's (1972) defensible space theory explains how the spatio-physical environment affects residents' capability to act as guardians by either facilitating or blocking opportunities for natural surveillance. According to Newman's theory, residents can maximize their ability to monitor their surroundings if windows and doors of properties are oriented towards public space and designed to face each other along a street, and by ensuring that the lines of sight between private residences and public space are clear and unobstructed. This makes it easier for residents to oversee their property, neighboring properties, and public space immediately surrounding them. With the support of these features of environmental design and layout, residents can supervise their surroundings purposely or inadvertently while carrying out regular household activities (Newman 1972). Felson (2006) reinforces this argument, suggesting that supervision is maximized when guardians can see as much of their residential space as possible. These natural surveillance opportunities function as a tool for crime prevention as they increase the risk that potential offenders will be seen by increasing the observability of residential areas as well as the visibility of residents who are at home.

Image and Maintenance

The image of a residential environment reflects the extent to which it is controlled, maintained, and well cared for by its residents (Newman 1972). Image provides information about the management of a residential place and whether or not its residents are effective place managers. The extent to which residents take responsibility for keeping a place clean and well maintained determines the image of residential places. Residents can help guard against crime through the regular upkeep of properties and the wider residential environment. Broken windows theory corroborates this perspective as it suggests that cues reflecting poor management and maintenance, such as broken windows and graffiti, promote an image of vulnerability and encourage crime. Empirical evidence also shows how physical incivilities like graffiti, broken windows, litter, and general disrepair encourage social incivilities and lead to increased crime, fear of crime, and withdrawal from the community by residents (e.g., Skogan 1990). This evidence reveals the ways in which the image of a place can determine resident participation as it can affect the frequency with which residents are present, as well as their willingness to monitor and intervene when necessary.

Accessibility, Land Use, and the Routine Activities of Place

While residents can contribute to the maintenance of their physical environment, image is also generated by the routine activities of that place. The social organization of behavior at a particular place determines its routine activities and therefore its risk of being targeted for victimization. The routine activities of a place can be viewed as a function of the level of activity and the type of activity encouraged therein. The accessibility of a place and its land-use patterns in turn contribute to the level and type of activity that is routine at a place. Accessibility and land use affect residents' ability to guard against crime through their capacity to either generate or inhibit opportunities for effective monitoring and direct intervention when necessary. Taylor et al. (1995, p. 122) explain that the presence of nonresidential land use on a block essentially "alters the ratio of outsiders to regulars on a block, lessening the familiarity of faces residents see around them."

Similarly for areas that are highly accessible, a high proportion of users are likely to be outsiders and this diminishes residents' monitoring capabilities by reducing the likelihood that they will be able to detect suspicious activity or individuals who are there for illegitimate purposes. Empirical studies have shown that crime tends to be higher in more accessible, highly utilized areas of the street network. High accessibility and activity level at a place also generates increased opportunities for crime by encouraging opportunities for physical and social incivilities (Taylor et al. 1995). The image of the place that is generated as a result of the number of external users that are attracted to it can have a negative effect on resident's place attachment, participation, and their territorial behavior. Taylor et al. (1995) explain that the lack of familiar faces and the physical deterioration that is encouraged by the high usage cause residents to withdraw from public spaces and become much less engaged in place management. This consequently serves only to reinforce the negative place image. Exacerbating the issue further still is that nonresidential land uses on a residential block create "holes" in the residential fabric that residents do not take responsibility for (Taylor et al. 1995, p. 122).

There is, however, some debate over the effect of accessibility and mixed land use on residents' guardianship capabilities and on crime in general. Built on the work of Jane Jacobs (1961), New Urbanism promotes residential developments comprised mixed land use that includes public transportation stops and other amenities. Jacobs (1961) argues that mixed land usage and the flow of people it attracts at various times of the day are factors that encourage, rather than detract from, opportunities to guard against crime since more people means more available guardians.

The Sociodemographic Context and Guardianship

The Density of Acquaintanceship and the Social Context

Citizens' willingness to monitor and their willingness to intervene are both influenced by their sociodemographic context. Freudenberg (1986) argues that while monitoring by residents in an area discourages offenders and helps control crime levels, this is likely to be mediated by the social context and the degree to which residents know each other. Thus, he argues that the *density of acquaintanceship* or the degree to which residents in a community know each other is the critical factor as it determines the degree of "watchfulness" by residents in areas with stable population (Freudenburg 1986). Likewise, Sampson et al. (1997) found evidence to suggest that residents' willingness to intervene is linked to social cohesion among neighbors and is negatively associated with neighborhood variations in violent victimization. These studies emphasize the importance of the social environment in encouraging residents' capability to guard against crime.

Residential Mobility, Ethnic Heterogeneity, Income, and Social Ties

Residents' immediate social environment is fundamentally shaped by demographic characteristics of the broader neighborhood context. Sampson and Groves' social disorganization model predicts that neighborhood demographic characteristics, including ethnic heterogeneity, low income, and residential mobility, contribute to high crime and delinquency rates by weakening informal social control and disrupting social organizations and networks within neighborhoods. This has been corroborated by micro-level empirical studies. Xie and McDowall's (2008) study showed that residential turnover independently increases the risk of property victimization due to changes in the opportunity structure for crime that result from constrained capabilities to guard against crime.

According to Xie and McDowall (2008), high residential turnover on streets increases the vulnerability of households to crime through the confluence of several mechanisms that highlights the relationship between the sociodemographic environment and the guardianship dimensions of monitoring and intervention: "Residential moves sever local ties and weaken social control, so newcomers tend to have fewer friends in the neighbourhood who might act as guardians of their homes. . .Long-term neighbours are more likely to recognize unusual patterns and unfamiliar people that may be signs of criminal activity. They will also be more likely to intervene, ask questions, and notify the police of clear indications of unusual activities" (Xie and McDowall 2008, p. 543). High rates of resident turnover mean high numbers of new residents whose lack of knowledge of their residential area has a negative effect on their monitoring capabilities and perhaps even on their confidence to intervene. Xie and McDowall (2008, p. 543) suggest that "potential offenders may perceive the vulnerabilities of newcomers," including their diminished capacity to distinguish outsiders from their own neighbors.

Alternative Applications and Extensions of the Guardianship Concept

Mechanical Surveillance

Cohen and Felson (1979) explain that while capable guardians are usually people, the function can also be served to some extent by mechanical objects. As such, CCTV cameras have been used increasingly as technological/mechanical replacements for the supervision and control provided by human guardians. When guardians are not able or available to provide supervision over targets, people, and places, it has been proposed that CCTV cameras are an effective form of surveillance. Moreover, whether manned or not, CCTV cameras are said to create the perception that someone is watching. It is therefore meant to have a similar effect as a guardian who is watching by discouraging offenders in search of crime opportunities because of the elevated risk of detection (see Clarke and Eck 2005).

Guardianship and Physical Control

Target Hardening

Another dimension of the guardianship concept to be introduced is that of *physical guardianship* or the protection of targets through physical control. Physical guardianship has been defined as "routine activities that provide physical protection of homes" (Wilcox et al. 2007, p. 774) or other property targets. According to Wilcox et al. (2007), these include target hardening measures that overlap with guardianship and are consistent with situational crime prevention, such as the use of alarm systems and locks. One of the principles of SCP is that target hardening can be used effectively to block opportunities for crime by increasing the effort required to commit crime. Several studies have distinguished these types of target hardening measures as a specific form of guardianship (e.g., Miethe and Meier 1994; Garofalo and Clark 1992). It is again important to note, however, that this is distinct from the routine activity concept of guardianship, and even within the situational crime prevention literature, target hardening and guardianship are considered separate techniques for crime prevention. Increasing guardianship is categorized as a technique that helps prevent crime by increasing the risk for offenders that they will be identified or caught. On the other hand, increasing target hardening is a separate technique designed to increase the effort required by offenders to illegally acquire or access a target (Clarke 1997). Though related, these are viewed as distinct techniques that operate successfully individually and in combination to prevent crime through separate strategies.

Self-Protection Measures

The fusion of the routine activities approach with the lifestyle-exposure perspective of Hindelang et al. (1978) inspired the birth of another branch in the conceptualization of guardianship, one that remains distinct from the original form it takes in the routine activity approach. Moving away from the perspective of guardianship as an expression of social control, studies by Mustaine and Tewksbury (1998) and Tewksbury and Mustaine (2003) conceptualize guardianship in terms of specific behaviors geared towards self-protection and target protection. Starting with the perspective that individual lifestyle or routine activities determine "differential exposure to dangerous places, times, and others" (Miethe and Meier 1994, p. 32), Mustaine and Tewksbury (1998, p. 834) define guardianship as "the degree of protection afforded to property or persons." In keeping with this definition, guardianship is operationalized by the use of self-protective behaviors by individuals, including weapon possession (e.g., carrying a knife or gun) and possession of body alarms or mace (Tewksbury and Mustaine 2003). There remains, however, a great deal of debate about the extent to which personal protection measures like these are synonymous with the original concept of guardianship and reduced risk of victimization (*see* Wilcox et al. 2007; Reynald 2009).

Conclusions and Future Directions

"Private citizen's involvement in crime prevention is being given a high priority as a public policy issue" (Geis and Huston 1983, p. 398). This has fuelled debates about whether or not private citizens are capable of guarding against crime, whether or not they have the expertise to do so, and the dangers inherent in involving citizens at this level. These concerns are especially interesting in light of the fact that private citizen involvement in crime control is not a novel phenomenon and actually is an historical antecedent of formal policing in westernized states (Zedner 2006). In fact, crime prevention through guardianship in the form of informal surveillance and shared responsibility has been put forward as an alternative to the ideals of the present-day formal state police (Zedner 2006). This informal guardianship by citizens preceded formal policing as it now exists within the modern criminal justice state. Indeed, the technique of guardianship as a form of situational crime prevention would thus fit in well with what Zedner (2006, p. 85) describes as the eighteenth-century policing experts who "regarded prevention as more important than the retrospective functions of arrest, detention and prosecution that later came to characterize modern policing." Critical for the development of research on guardianship is understanding the ways in which informal guardians can enhance the functioning of formal guardianship processes and, as a consequence, have a positive impact on crime control (see Reynald 2011b, for further discussion).

While a recent empirical focus on understanding the mechanisms that facilitate and inhibit effective informal guardianship has revealed the importance of environmental/contextual factors in determining guardianship intensity, further microlevel research is needed to elucidate ways in which situational and individual factors interact to affect the intensity of guardianship at places. In his article on communities, crime, and neighborhood organization, Skogan (1989, p. 437) asserts the claim that "the ability of individuals to act in defense of their community is shaped in important ways by the opportunities for action that are available to them." This highlights the importance of mobilizing citizens to take action through guardianship by raising their awareness of the active role they can potentially play in crime prevention. Aside from continuing to develop our understanding of the function of informal guardians as crime control agents in their residential environments, much needs to be explored about guardianship in other domains. Research on guardianship is dominated by a focus on the way it functions in the residential context, but very little is known about how citizens function as capable guardians in domains away from home.

Related Entries

- ▶ Routine Activities Approach
- ▶ Situational Crime Prevention

Recommended Reading and References

Bennett T, Wright R (1984) Burglars on burglary: prevention and the offender. Gower, Aldershot

Brown BB, Bentley DL (1993) Residential burglars judge risk: the role of territoriality. J Environ Psychol 13:51–61

Clarke RV (1997) Situational crime prevention: successful case studies, 2nd edn. Harrow & Heston, Albany

Clarke RV, Eck J (2005) Become a problem-solving crime analyst: in 55 small steps. Jill Dando Institute of Crime Science, University College London, London

Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44(4):588–608

Eck JE (1994) Drug markets and drug places: a case–control study of the spatial structure of Illicit drug dealing. Unpublished Ph.D. dissertation, University of Maryland, College Park

Felson M (1995) Those who discourage crime. In: Eck JE, Weisburd D (eds) Crime and place: crime prevention studies, vol 4. Criminal Justice Press, Monsey

Felson M (2006) Crime and nature. Sage, Thousand Oaks

Felson M, Cohen LE (1980) Human ecology and crime: a routine activity approach. Hum Ecol 8(4):389–406

Freudenburg WR (1986) The density of acquaintanceship: an overlooked variable in community research? Am J Sociol 92(1):27–63

Garofalo J, Clark D (1992) Guardianship and residential burglary. Justice Q 9(3):443–463

Geis G, Huston TL (1983) Bystander intervention into crime: public policy considerations. Policy Stud J 11:398–408

Hindelang MJ, Gottfredson MR, James G (1978) Victims of personal crime: an empirical foundation for a theory of personal victimization. Ballinger, Cambridge, MA

Hollis-Peel ME, Welsh BC (forthcoming). What makes a guardian capable? A test of guardianship in action. Secur J. Advance online publication, August 13, 2012; doi:10.1057/sj.2012.32

Jacobs J (1961) The death and life of great American cities. Random House, New York

Miethe TD, Meier RF (1994) Crime and its social context: toward an integrated theory of offenders, victims and situations. State University of New York Press, Albany

Miethe TD, Stafford MC, Sloane D (1990) Lifestyle changes and risks of criminal victimization. J Quant Criminol 6(4):357–376

Mustaine EE, Tewksbury R (1998) Predicting risks of larceny theft victimization: a routine activity analysis using refined lifestyle measures. Criminol: Interdiscip J 36:829–858

Newman O (1972) Defensible space: crime prevention through urban design. Macmillan, New York

Reynald DM (2009) Guardianship in action: developing a new tool for measurement. Crime Prev Commun Saf: Int J 11(1):1–20

Reynald DM (2010) Guardians on guardianship: factors affecting the willingness to monitor, the ability to detect potential offenders & the willingness to intervene. J Res Crime Delinq 47(3):358–390

Reynald DM (2011a) Factors associated with the guardianship of places: assessing the relative importance of the spatio-physical and socio-demographic contexts in generating opportunities for capable guardianship. J Res Crime Delinq 48(1):110–142

Reynald DM (2011b) Guarding against crime: measuring guardianship within routine activity theory. Ashgate, Surrey

Sampson RJ, Raudenbush SW, Earls F (1997) Neighbourhoods and violent crime: a multilevel study of collective efficacy. Science 277(5328):918–924

Skogan WG (1989) Communities, crime, and neighbourhood organization. Crime Delinq 35(3):437–457

Skogan WG (1990) Disorder and decline. Free Press, New York

Taylor RB, Koons B, Kurtz E, Greene J, Perkins D (1995) Street blocks with more nonresidential land use have more physical deterioration: evidence from Baltimore and Philadelphia. Urban Aff Rev 30:120–136

Tewksbury RA, Mustaine EE (2003) College students' lifestyles and self-protective behaviors: further consideration of the guardianship concept in routine activity theory. Crim Justice Behav 30(3):302–327

Wilcox P, Madensen TD, Tillyer MS (2007) Guardianship in context: implications for burglary victimization risk & prevention. Criminology 45(4):771–803

Xie M, McDowall D (2008) The effects of residential turnover on household victimization. Criminology 46(3):539–575

Zedner L (2006) Before and after the police: the historical antecedents of contemporary crime control. Br J Criminol 46:78–96

Informal Social Control

María B. Vélez
Department of Sociology, University of
New Mexico, Albuquerque, NM, USA

Overview

Research in sociology and criminology consistently indicates that crime clusters in places characterized by adverse conditions such as neighborhood disadvantage and residential mobility. Scholars draw on social disorganization theory and its idea of community social organization – the capacity of local communities to solve commonly experienced problems and realize collective goals – to help explain this clustering of crime. In neighborhoods characterized by high levels of informal social control, social disorganization is comparatively low, suggesting the importance of informal social control as a strategy to achieve community social organization. Informal social control reflects the ability of local neighborhoods to *supervise* the behavior of their residents and the capacity of neighborhoods to *socialize* their residents conventionally (Bursik 1988; Bursik and Grasmik 1993; Sampson and Groves 1989). In addition to promoting social organization, a predominance of evidence shows the crime-controlling effect of informal social control. Not only do these neighborhoods have lower crime rates, such contexts also translate into lower levels of offending and victimization risk among individuals. Put simply, the impact of informal social control in shaping neighborhood crime is "widely accepted" (Kubrin and Weitzer 2003, p. 376).

Fundamentals

Crime is unevenly distributed across neighborhoods, and research suggests that it tends to cluster in those communities with adverse conditions such as poverty, racial/ethnic heterogeneity, and residential mobility (Bursik and Grasmik 1993; Kubrin and Weitzer 2003; Sampson et al. 1997; Sampson and Groves 1989; Shaw and McKay 1942; Sampson 2012). Scholars draw on social disorganization theory and its idea of community social organization to help explain this clustering of crime. Community social organization refers to the ability of local communities to solve commonly experienced problems such as crime and realize common goals important to their residents such as achieving a crime-free environment (Bursik 1988, p. 521).

A major strategy to achieve community social organization is for neighborhood residents to exert informal social control. Informal social control is thought to have two central elements: the ability of local neighborhoods to *supervise* the behavior of their residents and the capacity of neighborhoods to *socialize* their residents conventionally (Bursik 1988; Bursik and Grasmik 1993; Sampson and Groves 1989). *Supervision* occurs through a variety of activities including resident's surveillance of public spaces, a shared sense of responsibility for taking care of the neighborhood and its residents such as calling parents when kids are misbehaving or neighbors taking care of each other's property especially when out of town, providing information to fellow residents about which parts of the neighborhood are unsafe and should be avoided, and directly intervening such as breaking up street fights, questioning strangers in the neighborhood, or admonishing fellow residents for behavior defined as unacceptable (Bursik 1988; Bursik and Grasmik 1993). *Socialization* occurs in the presence of strong conventional institutions such as two-parent families or well-funded and viable schools, which help to provide a setting where support for conventional processes of socialization can take hold. Neighborhoods characterized by extensive supervision and effective socialization are well situated to ensure norm compliance. As such, they enjoy little crime.

The goal for this encyclopedia entry is to provide a preliminary introduction for readers to the concept of informal social control. This entry is by no means exhaustive but is designed to outline the major contours of the concept and how it fits within the larger literature on communities and crime. This entry begins with a brief sketch of social disorganization theory, which serves as the theoretical footing for the role of informal social control in reducing community-level crime rates. Then, current conceptualizations and operationalizations of informal social control are discussed. The entry provides the current state of knowledge on its distribution and impact on local levels of crime and victimization as well as individual-level offending and victimization risk. The entry concludes with directions for future research.

Theoretical Background. Social disorganization theory serves as the primary theoretical backdrop of our understanding of the role of informal social control in shaping community crime rates. It is a long-standing theory in criminology originally pioneered by Shaw and McKay (1942), researchers grounded in the Chicago school, who sought to understand the neighborhood ecology of social problems such as juvenile delinquency. Utilizing a variety of secondary data sources including Cook County court information, they found that high delinquency rates associated positively with a variety of adverse neighborhood conditions including poverty, racial/ethnic heterogeneity, and residential mobility. They also found that high delinquency rates persisted in certain Chicago neighborhoods for several decades despite change in racial and ethnic composition. They drew on life-history data and argued that the uneven pattern of crime across urban neighborhoods and its persistence over time were rooted in adverse neighborhood conditions that impeded community social organization. In particular, they proposed that neighborhoods with economic deprivation tended to have high rates of population turnover (or mobility) and racial/ethnic heterogeneity, which in turn socially disorganized a community, and increased delinquency rates as well as other social problems.

Shaw and McKay (1942) argued that two key processes accompany the structural mechanisms that characterize disorganized areas and reinforce neighborhood disorganization. First, disorganization creates a context in which there is a wide diversity of values and attitudes with respect to conformity to the law. Because delinquency as a way of life competes with conventional ways, disorganized areas are characterized by divergent systems of values. This aspect of social disorganization theory aligns with cultural deviance theories. Second, because of weakened conventional institutions, socially disorganized communities are unable to control the behavior of residents and facilitate socialization processes so that they comply with conventional norms, especially those relevant for law-abiding behavior.

Shaw and McKay's emphasis on the need for neighborhood structures to regularize norm compliance is consistent with notions of social control, one of the most widely discussed concepts in sociology (Meier 1982). Social control is considered to be a collection of mechanisms (a social process) to induce compliance to norms (Meier 1982). Controls can be invoked internally with feelings of guilt or shame (products of effective conventional socialization) or commitment in conformity. The logic is that people feel good when they adhere to conventional norms, allowing them to continue a conventional life. When they do not adhere to these universal beliefs, they feel bad, self-critical, and guilty, deterring such norm violations in the future. Controls can also be induced externally by others in direct ways such as with scrutiny or surveillance engaged in by formal agents like the police or informal agents like neighbors. Others like family, school officials, or the community can exert control via the withdrawal of sentiment. The logic is that people conform to norms or rules because they are rewarded with things like status, prestige, money, and freedom when they do adhere to them and are punished with the loss of them when they do not. Taking this logic to the macro level, social units by way of their social relationships can regularize norm compliance by providing rewards for compliance and prompting high costs for deviance.

Social disorganization theorists took the above logic to heart when they reframed the theory in the 1970s and 1980s. They contended that neighborhoods vary in their ability to exert social control; they vary in their ability to provide awards for compliance and punishments for deviance. As such, contemporary articulations of this theory are basically a "group-level analog of control theory and is grounded in very similar processes of internal and external sources of control" (Bursik 1988, p. 521). The current entry focuses on contemporary discussions of social disorganization (see Bursik 1988; Bursik and Grasmik 1993; and Kornhauser 1978 for historical accounts of social disorganization as well as greater elaboration on the re-articulation as a macro-control theory).

This encyclopedia entry centers on the *informal* ways neighborhoods exert social control in urban communities; the informal level also is where the bulk of theorizing and research has taken place (Bursik 1988; Bursik and Grasmik 1993; Carr 2003; Kubrin and Weitzer 2003; see also Hunter 1985; Bursik 1988; Bursik and Grasmik 1993 for discussion of two other levels of networks and social control: private and public). How do neighborhoods informally provide rewards for enforcing norms and engender high costs for engaging in deviance so that they ensure a crime-free environment? The logic is that social disorganization arises in contexts of adverse neighborhood conditions such as – residential mobility, racial/ethnic heterogeneity, family disruption, and neighborhood disadvantage. These adverse neighborhood conditions translate into less informal neighboring such as fewer people getting together to chat and neighbors not knowing the names of their fellow residents. When residents have limited social interaction, it makes it difficult for a community to develop mutual trust and shared expectation among residents, impinging on the community's ability to provide the rewards for adhering to norms and the punishments for deviant behavior. This means that residents are unlikely to intervene during suspicious activity, admonish residents and local youth for deviant behavior, or watch out for one another's property. In turn, public spaces are not supervised effectively and crime rates increase. Empirically this logic also means that the intervening mechanisms of informal networks and social control will absorb much of the effects of neighborhood adverse conditions on crime and victimization. Discussed below is how researchers study informal social control and its impact on levels of crime.

Conceptualizations of Informal Social Control. Informal social control is hypothesized to emerge and facilitate reductions in crime by way of three neighborhood-level processes: social interaction, collective efficacy, and community organizations. First, the dominant approach termed the systemic model views communities as a "complex system of friendship and kinship networks and formal and informal associational ties rooted in family life and ongoing socialization processes" (Kasarda and Janowitz 1974, p. 329). As such, informal social control is conceptualized as the regulatory capacity of a neighborhood, presumed to arise from the community's affiliation, interactional, and communication networks among residents (Bursik 1988, 1999; Kasarda and Janowitz 1974). From this vein, larger, interconnected, and active associational networks provide the rewards for adhering to norms and the punishments for deviant behavior. The systemic model argues that poverty, residential instability and racial heterogeneity are expected to hamper the establishment of relational network structures that serve to facilitate mechanisms of informal social control. In turn, crime is expected to increase (Bursik 1999, p. 87). Importantly the systemic model assumes that places with strong ties among neighbors and to local institutions will bring about informal social control.

Empirical work suggests that strong informal ties among neighborhood residents do not always translate into low crime rates. For instance, ethnographic work on disadvantaged and minority communities indicates that although strong bonds exist among neighbors, they do not translate into reductions in crime (Pattillo 1998; Wilson 1996). Wilson (1996) finds that despite high levels of neighboring, inner-city residents in Chicago reported feeling they had little informal social control over their children. Wilson (1996, p. 64) states that "A primary reason is the absence of a strong organizational capacity or an institutional resource base that would provide an extra layer of social organization in their neighborhoods." In her research about a predominantly

African-American middle-class neighborhood in Chicago, Pattillo (1998) found a socially cohesive neighborhood characterized with abundant ties among residents. However, these ties did not always translate into an effective exercise of informal social control; rather, they often inhibited the ability of residents to admonish the behavior of other residents such as local youth hanging out in the street corner. Pattillo (1998, p. 763) argued that "… the more people are familiar with one another; the more illicit networks are absorbed into mainstream connections and thereby normalized." She found this reluctance especially pronounced when it came to residents calling the police. As one resident said about a young man's gang-affiliated behavior to Pattillo (1998, p. 763), "I didn't wanna give this young man's name to the police because his mama is such a sweet lady." Bellair (1997) in a study of Seattle neighborhoods finds that social interaction among neighbors that takes place at least once a year or more has consistent and negative effects on rates of burglary, motor vehicle theft, and robbery across Seattle neighborhoods. He, however, finds that if social interaction is frequent such as once a week, it does not have consistent effects on crime rates. Bellair (1997) suggest that in terms of the regulatory and sanctioning capacity of local networks, it may matter very little whether neighbors are close friends as long as they interact with one another enough to provide the mutual support needed for "doing something" on behalf of the community.

Such findings set the stage for another conceptualization of informal social control. The focus here shifts to collective efficacy, which highlights the *nature* of social ties rather their presence or absence (Warner and Rountree 1997; Wilson 1996; Pattillo 1998). It may be that networks originating in churches, after school clubs, or schools may not lead to the social control of crime because these networks are not strictly focused on reducing crime. Whereas networks that result from being involved in a neighborhood watch group where neighbors meet once a month to talk about crime issues do lead to the social control of crime because the content or nature of the network is focused on crime. This logic is consistent with notions of social capital – networks are used in part to transmit group norms and expectations pertaining to appropriate behavior for neighborhood residents and to impose a wide variety of informal negative sanctions if those expectations are violated. So in the case of crime, ties have to be geared towards collective action to secure a crime-free environment. The concept offered to address this issue is collective efficacy, a form of informal social control, measured as mutual trust, social cohesion, and a willingness to intervene (Sampson et al. 1997). Collective efficacy thus only counts informal ties among neighbors when they are coupled with high levels of informal social control; otherwise, they are not assumed to reduce crime from this perspective.

A third strategy for conceptualizing informal social control is to account for the presence of or resident participation in community organizations. Community organizations, especially those that have indigenous roots, act as a central vehicle for neighborhoods to bring about informal social control (Bursik and Grasmik 1993; Sampson and Groves 1989; Skogan 1989, 1990). For instance, community organizations provide spaces and events for the fostering of informal networks, the planning out of collective action to address crime and other neighborhood problems, and the supervision and socialization of neighborhood youth with programming. Moreover, community organizations help to minimize isolation by linking primary (i.e., family and friends) and secondary institutions (i.e., neighborhood watch groups and local businesses). Thus, community organizations help to organize neighborhoods so that they can "protect and defend their interests and values" (Kornhauser 1978, p. 79). It is important to note that the above literature is not clear on what types of organizations are most consequential for the control of crime.

Operationalization of Informal Social Control. To study informal social control, researchers have focused primarily on secondary relational networks (relationships among neighbors) and the exercise of informal (parochial) control. The bulk of studies on informal social control is quantitative and uses questions from community surveys that sample residents across neighborhoods in a particular city. In most cases, researchers compile a multiple-item index to capture informal types of social interaction among residents that take place in the neighborhood. Responses are usually aggregated to the community level (i.e., census tracts, police beats, community areas, and neighborhood clusters).

Informal Neighboring. Some researchers measure social interactions among neighbors that are presumed to yield informal social control. Social interaction among neighbors is usually captured by a variety of aspects related to informal neighboring such as how often respondents get together with their neighbors either in their neighbor's or their own home, how many people respondents know by sight or by name in the neighborhood, do respondents ask neighbors for favors such as borrowing tools or looking out for each other's property, and how many neighborhood kids are known by name (Bellair 1997; Elliott et al. 1996; Simcha-Fagan and Schwartz 1986, Warner and Rountree 1997). Further, Bursik (1988) points to the importance of residents patronizing local business establishments as a strategy for integrating residents into the social life of a community. As such, Bursik (1999) measured informal networks with a six-item factor that tapped into how often (range from nearly always to never) respondents do their shopping for groceries, medical goods, and clothing; do car repairs; went out to eat; and did their banking in the neighborhood.

Direct Intervention. Other papers attempt to capture neighborhood levels of direct intervention by residents with the extent to which residents "do something" or that they can count on their neighbors to "do something" about crime. Specifically, respondents are asked how likely it is they can count on their neighbors to intervene to solve problems in the neighborhood like someone breaking into a house in plain sight, someone is trying to sell drugs to a child in plain sight, a fight in front of the house and someone is getting beaten up, and their kids are getting into trouble. As an early example, Maccoby et al. (1958) asked respondents hypothetical questions about whether they would step in and try to do something if they observed children using abusive language, engaging in property damage, fighting, or drunkenness. Bellair (2000) captures informal social control with a two-item factor scale that combines neighborhood percentages on whether respondents watch their neighbor's property when the neighbors are out of town and if neighbors watch out for the respondent's property when the respondent is out of town. Silver and Miller (2004) use a four-item Likert scale that asks respondents the likelihood they can count on their neighbors to take action if children were skipping school or hanging out on a street corner, children were spray painting on a local building, children were showing disrespect to an adult, or if there was a fight in front of the house and someone was being beaten up. In another study, Sampson et al. (1999) utilize a similar measure as Silver and Miller (2004) to capture child-centered control by asking respondents the likelihood that neighbors would take action if children were skipping school or hanging out on a street corner, children were spray painting on a local building, and children were showing disrespect to an adult. It is important to note that most measures capturing the extent that residents do "doing something" or that they can count on their neighborhoods to "do something" are based on hypothetical situations.

Collective Efficacy. To operationalize collective efficacy, scholars often replicate Sampson et al. (1997) who combine two factors that were closely associated across neighborhoods: (1) social cohesion and mutual trust are part of this construct because residents are unlikely to intervene when the rules are unclear or people mistrust or fear one another and (2) the willingness to intervene (informal social control). Sampson et al. (1997) state that the two measures tap aspects of the same latent construct, which they term collective efficacy (see also Sampson 2012). To measure the willingness to intervene on behalf of the neighborhood, Sampson et al. (1997) used a five-item Likert-type scale that asks residents about the likelihood (ranging from very likely to very unlikely) that their neighbors could be counted on to intervene in various ways if children were skipping school and hanging out on a street corner, children were spray-painting graffiti on a local building, children were showing disrespect to an adult, a fight broke out in front of their house, and the fire station closest to their home was threatened with budget cuts. To capture "social cohesion and trust," respondents were asked how strongly they agreed that "people around here are willing to help their neighbors," "this is a close-knit neighborhood," "people in this neighborhood can be trusted," "people in this neighborhood generally don't get along with each other," and "people in this neighborhood do not share the same values" (the last two statements are reverse coded). Note that had Sampson et al. (1997) only used their willingness to intervene measure, it would have been consistent with above work on direct intervention. But to capture the notion of collective efficacy, they combine direct intervention with social cohesion and trust into one measure.

Community Organizations. Researchers measure the role of community organizations in one of two ways. First, using survey data, some researchers examine the extent of resident reported participation in community organizations. For example, Sampson and Groves (1989) computed the percentage of residents in a neighborhood that had attended a committee or club meeting in the week prior to the survey. As another example, Morenoff et al. (2001) measure the extent residents report participation in voluntary associations such as local religious organizations, neighborhood watch programs, block group associations, community councils, business or civic groups or ethnic/nationality clubs, or local political organizations. Second, scholars attempt to capture the presence of organizations in a community by either counting their presence with a directory or by asking respondents in community surveys about their knowledge of organizations present in their neighborhoods. For instance, Morenoff et al. (2001) put together an index of the number of resident reported organizations and programs in the neighborhood such as a community newspaper, block group, crime prevention program, alcohol/drug treatment program, mental health center, or family health service. Skogan (1989, p. 445) created an index that summed resident responses to three questions concerning respondent's awareness of "volunteer citizens patrolling residential areas," "groups that encourage citizen to undertake crime prevention efforts," and "groups that work to improve police-community relations" within two to three blocks of their houses.

The State of Evidence Regarding Informal Social Control

The Distribution of Informal Social Control. Extant work from the systemic or collective efficacy approach shows that informal social control is harder to achieve in neighborhoods characterized by adverse neighborhood conditions including concentrated disadvantage, immigrant concentration, racial/ethnic heterogeneity, and residential mobility. In the now classic assessment of the systemic model, Sampson and Groves (1989) using data from the British Crime Survey find strong support for the idea that adverse neighborhood conditions like low economic conditions dampen informal strategies to control crime across neighborhoods in Great Britain. In particular, they find that neighborhoods characterized by low socioeconomic status, ethnic heterogeneity, family disruption, and urbanization have relatively high levels of unsupervised peer groups, their measure of a community's diminished capacity to supervise youth. They also find that poor neighborhoods report less organizational participation than their more affluent counterparts. Similar patterns emerge from studies grounded in collective efficacy. In the first study on collective efficacy, Sampson et al. (1997) find that concentrated disadvantage and immigrant concentration decreases levels of neighborhood collective efficacy across Chicago neighborhoods while residential stability increases its presence. Morenoff et al. (2001) find that collective efficacy is significantly lower in Chicago neighborhoods characterized by concentrated disadvantage, immigrant concentration, and with prior high rates of homicide, whereas affluence, resident participation in voluntary associations, the presence of organizations, and strong kin/friendship ties increase collective efficacy levels across Chicago neighborhoods.

Although the above work establishes support for the idea that adverse structural contexts do not foster informal social control as successfully as more advantaged contexts, theoretical and empirical work has yet to fully articulate the source of this gap. In research directly addressing this question, Silver and Miller (2004) argue that if residents feel that local institutions have not met their needs or aspirations, which is often the case for disadvantaged neighborhoods, they will feel more like they are "passing through" and thus be less willing to intervene on behalf of the community. Moreover, when resident satisfaction with the police is low, typical in disadvantaged neighborhoods, residents will be reluctant to take personal risks to safeguard the community and so will be less willing to exert informal social control. Silver and Miller (2004) find support for this logic in that relatively higher levels of neighborhood attachment and satisfaction with police contributed significantly to increases in neighborhood levels of informal social control. Moreover, they find that neighborhood attachment and police satisfaction help to mediate a substantial portion of the effects of concentrated disadvantage and immigrant concentration on levels of informal social control.

Evidence for the Impact of Informal Social Control on Crime. Though more theoretical and empirical work needs to be done to account for the uneven distribution of informal social controls across neighborhood contexts, the fact that it is more limited in disadvantaged communities serves to compound the crime problem in these communities. At the same time, research also suggests that informal social controls can act as a protective factor, buffering the criminogenic influence of disadvantage. Evidence suggests that weak informal social control mechanisms increase neighborhood levels of crime and victimization. Early work by Maccoby et al. (1958) compared two Cambridge, Massachusetts neighborhoods that were similar in terms of socioeconomic status but with different levels of juvenile delinquency. One neighborhood had relatively high levels of juvenile delinquency, while the other had relatively low rates of juvenile delinquency. They find that area differences in informal social control at least partly explain differences in juvenile delinquency across the two neighborhoods. Specifically, they find significant area differences in informal social control, measured as the percentage of residents who "would do something about it" in hypothetical instances of fighting and drinking. They also found a higher percentage of interventions among non-victims in low versus high delinquency areas. At the same time, they did not find significant neighborhood differences in informal social control when it pertained to children involved in property damage or being openly rude such as using insulting language.

Recall that during the 1970s and 1980s, social disorganization was reframed as a systemic model, which emphasized the role of informal social networks as precursors of informal social control. In the first contemporary assessment of the systemic model, Sampson and Groves (1989) examined the impact of unsupervised peer groups, conceptualized as a neighborhood's inability to regulate youth in public spaces, and local participation in formal and voluntary organizations (a measure of informal networks among neighbors) on 12 different victimization and offending rates across a large number of local British areas. They found strong direct effects of unsupervised peer groups in increasing crime and victimization rates, while participation in local organizations lowered crime and victimization rates. They also found strong support for the mediation arguments of social disorganization theory in that the intervening measures of informal social control absorbed much of the effects of poor socioeconomic conditions, ethnic heterogeneity, residential mobility, family disruption, and urbanization on crime and victimization rates.

Sampson and Groves (1989) strong support for the social disorganization framework paved the way for subsequent tests within the systemic and collective efficacy approaches to informal social control (Bellair 1997, 2000; Elliott et al. 1996; Simcha-Fagan and Schwartz 1986; Warner and Rountree 1997). Elliott et al. (1996) in an analysis of Chicago and Denver neighborhoods found neighborhood levels of informal social control significantly lowered juvenile delinquency rates across Chicago but not Denver neighborhoods. Supporting expectations of social disorganization theory, they also find that informal social control mediated the effect of neighborhood disadvantage on adolescent reports of delinquency in both Chicago and Denver. Two additional papers demonstrate the impact of collective efficacy on neighborhood levels of crime. In the first paper to advance the concept, Sampson et al. (1997) finds that higher levels of collective efficacy yield lower levels of homicide across Chicago neighborhoods. In a second paper, Morenoff et al. (2001) find that collective efficacy decreases both local levels of homicide victimization and homicide offending across Chicago neighborhoods.

Research has also examined the role of community organizations in shaping neighborhood levels of crime. Simcha-Fagan and Schwartz (1986) in a study of 12 neighborhoods in New York City found that neighborhoods with relatively high levels of organizational involvement had relatively lower levels of rates of delinquency. Similarly, Sampson and Groves (1989) investigate the impact of resident participation in community organizations on various types of crime. They find that neighborhoods with relatively high levels of organizational participation have lower levels of burglary, auto theft, mugging, street robbery, stranger violence, and total victimization across British neighborhoods. Other research suggests that the reason organizational participation reduces crime is because it fosters collective efficacy (Morenoff et al. 2001). This research indicates that beyond participation, the number of organizations available to community members matters. In fact, Morenoff et al. (2001) find the presence of community organizations is important for producing collective efficacy which, in turn, brings about decreases in homicide in Chicago neighborhoods. Regardless of how community organizations are measured, research suggests that they play a role in bolstering informal social control and thereby help to reduce crime.

Scholars also investigate how informal social control at the neighborhood level shapes individual-level offending or victimization risk using multilevel analytical strategies. For example, Elliott et al. (1996) find that, under some circumstances, informal social control decreases the likelihood of youth engaging in problem behavior. Though they find this to be the case among youth in Chicago, this pattern did not hold for youth in Denver. Simcha-Fagan and Schwartz (1986) find that community organizational participation, a composite of factors meant to capture informal social control, at the neighborhood level decreased the likelihood of self-reported delinquency for youth across 12 neighborhoods in New York City. Sampson et al. (1997) find that increases in neighborhood levels of collective efficacy translated into decreases in individual perceptions of neighborhood violence and individual-level victimization risk.

In the end, a preponderance of empirical evidence shows that residents in adverse neighborhood conditions struggle to engage in the kinds of informal social control necessary to reduce crime such as organizing neighborhood watches, intervening when suspicious activity occurs, or making the police more accountable to resident needs. Concentrated disadvantage, racial segregation, and residential mobility are powerful forces that dampen the ability of neighborhoods to provide (in informal ways) rewards for conformity and punishments for deviance. In turn, neighborhoods with relatively low levels of informal social control, measured in a variety of ways as outlined above, are the kinds of places that experience relatively high levels of crime. Such contexts can translate into higher levels of offending and victimization risk among individuals.

Future Directions for Informal Social Control

As established above, social disorganization research has focused almost exclusively on what residents can do inside the neighborhood to address crime (Bellair 1997; Bursik and Grasmik 1993; Sampson and Groves 1989; Sampson et al. 1997; Warner and Rountree 1997). As such, high rates of neighborhood crime are viewed as largely a result of the inability of residents within a neighborhood to organize collectively against crime. But a growing body of criminology shows that the ability of a community to control crime is shaped profoundly by its relationship with local political and economic actors. Favorable ties to political and economic elites are thought to translate into the allocation of lucrative resources to improve local public spaces, increase police presence in a community, and speed-up city approval for resident-driven initiates related to crime control. This process is often termed public social control. Recall that Silver and Miller (2004) found that neighborhoods with strong ties to the police and local public officials have high levels of informal social control. As another example, Carr (2003) documents how informal social control efforts are facilitated when cities enact policies that reinforce these efforts. He describes that when Chicago implemented a community policing program, it provided the resources and support for residents to coordinate two crime control programs: (1) the Beltway Night Patrol, the neighborhood problem-solving group that orchestrated the closing down of a troublesome tavern, and (2) the Court Advocacy Program that organized residents to attend hearings and pressure local judges to hand down harsher penalties for residents charged with criminal/delinquent acts. This growing body of work offers evidence that external actors from outside the neighborhood can facilitate informal social control. Criminologists should further investigate the interplay between informal dynamics of social control with local political and economic entities. Doing so should help to make more central the idea that broader political and economic forces are consequential for the ability of residents to control crime.

The bulk of our understanding of the role of informal social control on neighborhood crime is cross-sectional and thus does not adequately take into account that the relationship between informal social control and neighborhood crime is likely multifaceted, non-recursive, and dynamic. Two issues are particularly important. First, cross-sectional research designs do not handle very well the possibility of reverse causality: crime possibly influences the ability of a community to exert informal social control. For instance, research finds that disorder and incivilities often go hand-in-hand with high-crime areas, and residents often view these conditions as troublesome and threatening (Skogan 1990; Taylor 1995). In particular, violence, disorder, and incivility correspond to a variety of negative emotions felt by residents – fear of victimization, mistrust of others, desire to move, and general dissatisfaction with the neighborhood – that fuel an unwillingness to intervene on behalf of the common good. In a study of neighborhoods across five cities, Skogan (1990) finds that disorder decreased neighborhood satisfaction, the percentage of residents who planned to stay in the neighborhood, and the percentage of residents reporting that neighbors help each other. As another example, Hipp (2009) finds that households that perceived more social and physical disorder reported lower levels of satisfaction with their neighborhood, especially in communities where residents perceived crime as a problem. Clearly, crime itself can be a disorganizing force that can impair informal social control. A handful of studies have found support for this claim with cross-sectional designs (Bellair 2000; Markowitz et al. 2006; Sampson et al. 1999). But without longitudinal data, these findings must be taken cautiously. Future research should utilize longitudinal data that allow for a more careful examination of the informal social control – crime relationship. Specifically, longitudinal data will allow for the assessment of the sequencing of the relationship between informal social control and crime. By better incorporating strategies for feedback processes among informal social control and crime, a more definitive assessment of this causal relationship can be established.

Second, cross-sectional data are not well suited for providing information on how informal social control evolves over time. But there is reason to believe that informal social control is a dynamic process that changes within a neighborhood over time as events take place, policies are enacted, new populations move in or leave the neighborhood, and crime fluctuates up or down. Work by Carr (2003) and Small (2002) is instructive here. Carr (2003) spent 5 years conducting field work in a working-class Chicago neighborhood (Beltway). He found that although traditional forms of informal social control were low such as the direct supervision of youth, a hybrid version called new parochialism had evolved partly because of the introduction of a community policing program. He argues that new parochialism is a set of strategies of ongoing informal social control that are enabled and stimulated by the actions of city officials. Successful implementation of informal social control, Carr finds, only happens when there is intervention by institutions and groups from outside the neighborhood. Carr suggests that such a complex process would not have been discovered if he had used a shorter time frame. Similarly, Small (2002) attempts to understand why community participation declined in Villa Victoria, a poor Puerto Rican housing project in Boston. He finds that the decline in local participation in the housing project was driven by changes in cohorts (different generations of residents) and not changes in structural conditions and that the most important of these changes was the transformation in the cultural categories through which the cohorts framed their neighborhood. In particular, Small (2002, p. 25) finds that the first generation of residents framed the neighborhood positively "as a success, something beautiful to cherish and preserve" and so were willingness to participate in local community activities that enhanced informal social control. In contrast, the younger generation of residents framed their neighborhood poorly as "physically dirty and deteriorating . . .above all, a ghetto" and were less compelled to participate in local community activities, hindering most efforts at informal social control (Small 2002, p. 25). Had Small (2002) taken a single snap shot of one time period, he would have concluded that participation was either high or low depending on the cohort under observation. But by taking the long view, he is able to map out the process by which a disadvantaged community goes from being an efficacious one to a marginalized and socially isolated one. Longitudinal designs are a critical strategy for incorporating the dynamic and evolving nature of informal social control and assessing how this dynamism is consequential for neighborhood crime or victimization risk.

In sum, crime rates tend to cluster in neighborhoods characterized by adverse conditions such as disadvantage, residential segregation, and mobility. To explain this pattern, the bulk of social disorganization research has focused on the roles of informal neighbor networks and their attendant social controls. Informal social control reflects the ability of local neighborhoods to *supervise* the behavior of their residents and the capacity of neighborhoods to *socialize* their residents conventionally. As such, informal social control is a process by which norm compliance is achieved. Since Shaw and McKay's (1942) early work, a majority of evidence shows the crime-controlling effect of informal social control. Neighborhoods with relatively high levels of informal social control are the kinds of places that enjoy relatively little crime and lower levels of offending among individuals. Put simply, the impact of informal social control in shaping neighborhood crime is "widely accepted" (Kubrin and Weitzer 2003, p. 376).

Related Entries

- ▶ Broken Windows Thesis
- ▶ Communities and the Police
- ▶ Community Policing
- ▶ Crime and the Racial Composition of Communities
- ▶ Defining Disorder
- ▶ Differential Social Organization
- ▶ Disadvantage, Disorganization and Crime
- ▶ Early Chicago School Theory
- ▶ Impacts of Community-Oriented Policing

- ▶ Law of Community Policing and Public Order Policing
- ▶ Longitudinal Crime Trends at Places
- ▶ Longitudinal Studies in Criminology
- ▶ Neighborhood Effects and Social Networks
- ▶ Public Housing and Crime Patterns
- ▶ Race and Ethnicity in Social Disorganization Theory
- ▶ Rational Choice, Deterrence, and Crime: Sociological Contributions
- ▶ Social Capital and Collective Efficacy
- ▶ Social Control
- ▶ Social Control and Self-Control Through the Life Course
- ▶ Social Disorder and Physical Disorder at Places
- ▶ Theories on Policing and Communities

Recommended Reading and References

Bellair PE (1997) Social interaction and community crime: examining the importance of neighbor networks. Criminology 35:677–701

Bellair PE (2000) Informal surveillance and street crime: a complex relationship. Criminology 38:137–169

Bursik RJ Jr (1988) Social disorganization and theories of crime and delinquency: problems and prospects. Criminology 26:519–552

Bursik RJ Jr (1999) The informal control of crime through neighborhood networks. Sociol Focus 32:85–97

Bursik RJ Jr, Grasmik HG (1993) Neighborhoods and crime. Lexington Books, New York

Carr PJ (2003) The new parochialism: the implications of the beltway case for arguments concerning informal social control. Am J Sociol 108:1249–1291

Elliott DS, Wilson WJ, Huizinga D, Sampson RJ, Elliot A, Ranking B (1996) The effects of neighborhood disadvantage on adolescent development. J Res Crime Delinq 33:389–426

Hipp JR (2009) Specifying the determinants of neighborhood satisfaction: a robust assessment in 24 metropolitan areas. Soc Forces 88:395–424

Hunter AJ (1985) Private, parochial and public school orders: the problem of crime and incivility in urban communities. In: Suttles GD, Zald MN (eds) The challenge of social control: citizenship and institution building in modern society. Alex, Norwood

Kasarda JD, Janowitz M (1974) Community attachment in mass society. Am Sociol Rev 39:328–339

Kornhauser RR (1978) Social sources of delinquency: an appraisal of analytic models. University of Chicago Press, Chicago

Kubrin CE, Weitzer R (2003) New directions in social disorganization theory. J Res Crime Delinq 40:374–402

Maccoby EE, Johnson JP, Church RM (1958) Community integration and the social control of juvenile delinquency. J Social Issues 14:38–51

Markowitz FE, Bellair PE, Liska AE, Liu J (2006) Extending social disorganization theory: modeling the relationships between cohesion, disorder, and fear*. Criminology 39(2):293–319

Meier RF (1982) Perspectives on the concept of social control. Annu Rev Sociol 8:35–55

Morenoff JD, Sampson RJ, Raudenbush SW (2001) Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. Criminology 39:517–559.

Pattillo ME (1998) Sweetmothers and gangbangers: managing crime in a black middle-class neighborhood. Soc Forces 76:747–774

Sampson RJ (2012) Great American city: Chicago and the enduring neighborhood effect. University of Chicago Press, Chicago

Sampson RJ, Groves WB (1989) Community structure and crime: testing social-disorganization theory. Am J Sociol 94:774–802

Sampson RJ, Raudenbush SW, Earls F (1997) Neighborhoods and violent crime: a multilevel study of collective efficacy. Science 277:918–924

Sampson RJ, Morenoff JD, Earls F (1999) Beyond social capital: spatial dynamics of collective efficacy for children. Amer Sociol Rev 633–660

Shaw CR, McKay HD (1942) Juvenile delinquency and urban areas. The University of Chicago Press, Chicago

Silver E, Miller LL (2004) Sources of informal social control in Chicago neighborhoods. Criminology 42(3):551–584

Simcha-Fagan OM, Schwartz JE (1986) Neighborhood and delinquency: an assessment of contextual effects. Criminology 24(4):667–699

Skogan WG (1989) Communities, crime and neighborhood organization. Crime Delinq 35:437–457

Skogan WG (1990) Disorder and decline: crime and spiral of decay in American neighborhoods. University of California Press, Berkeley

Small ML (2002) Culture, cohorts, and social organization theory: understanding local participation in a Latino Housing Project. Am J Sociol 108:1–54

Taylor RB (1995) The impact of crime on communities. Ann Am Acad Pol Soc Sci 539:28–45

Vélez MB (2001) The role of public social control in urban neighborhoods: a multilevel analysis of victimization risk. Criminology 39:837–864

Warner BD, Rountree PW (1997) Local social ties in a community and crime model: questioning the systemic nature of informal social control. Soc Probl 44:520–536

Wilson WJ (1996) When work disappears: the world of the new urban poor. Random House, New York

Information Technology and Police Work

Peter K. Manning
College of Criminal Justice, Northeastern
University, Boston, MA, USA

Synonyms

COMPSTAT; Crime mapping; Evidence-led policing; History of technology in policing; How to make crime mapping more inferential; Intelligence-led policing

Overview

This entry addresses the role of information technology in shaping policing organizations and police practices. It is comprised of five parts. The first part is a brief overview of police and technology. The second part outlines types of technology in police organizations. The third part discusses current police innovations in information technology. The fourth sets out the problematic collocation of data in police organizations. The fifth part discusses the focus on the incident, the core of operational policing, and the ways in which it controls and shapes *how* whatever technology is available is used (or not). The entry concludes with some propositions that might guide future research. Overall, while there are abundant capacities based on information technology available, they are selectively used and patterned by traditional notions of the job and the occupational culture of the patrol division.

Introduction

Literature on technology and organizational change in management science, organizational theory, and social theory more generally reflects a long history of what has been called the "iron cage" thesis of Weber (Weber 1958: 181–182). In compact form, this states that rationality in its various forms drives and shapes all modern bureaux, and that this rationalization process, linked to capitalism, is inevitable, universal, and unidirectional. Implicit in this statement is the notion that organizations in their dance with rationality, forward and back, are much the same in their adaptive and environmentally sensitive maneuvers. This thesis remains a proposition, something of an assumption, such that research short circuits the general proposition and then proceeds to examine the extent to which it is true. Assuming the rational nonproblematic nature of the organization permits investigation of putative innovations in policing, corporate life, schools, and the military as if the rationality of these innovations and their success is merely a question of goodness of fit with the current practices within the organization. At the heart of this powerful assumptive assemblage is the role of technology, a sponge concept whose definition remains as elusive as any other of the treasured social science concepts on which we trade.

Inevitably, the introduction of technology into discussions of innovation always raises questions because technology is not a thing, but an imagined process and a product of practices. The term "imagined process" refers to the fact that technology is a *social object* that is created by actions directed toward it and responses that are made to these actions by those involved in the installation, maintenance, and use of such a social object. It is a collectively created and sustained matter with many facets. Granted that "technology" has a variety of connotations and is always embedded in an organizational structure, it can be succinctly summarized as the ways in which work is accomplished including both material and symbolic matters, the understanding of which is required to get this work done. "Technology" has a visible, material aspect (weight, shape, and size), a symbolic or representational aspect (it stands for something as shown in the actions of people who use it), a logic, as well as a range of connotative meanings – implied attitudes toward it and talk about its functions (Roberts and Grabowski 1996). What is rarely discussed (although see

Manning 2008) is the cognitive and imaginative work required to understand, fix, maintain, and use the technology routinely. The imaginative and cognitive aspects are situated and situational in their unfolding, and the workings of technologies, including information technologies, are best seen or revealed by looking at how people use them in key or *thematic situations* that are directly and indirectly relevant to their work tasks. *Technologies are situationally and contingently relevant*. Their use while in some sense is routinized is always touched off by a work-situation and in that sense is *occasioned*. This is true because information technologies, as noted above, are invisible in their interior workings, complex in their effects, not all of which are understood, appreciated, or anticipated, and viewed with some degree of mystification. Thus, it is important to keep an eye on what is out of sight – both within the machinery as well as in the minds and bodies of users. In addition, since whatever "technology" is introduced will often be in conflict with current practices, dialectic and associated technological dramas are inevitable (Manning 1992). Information technology (IT) in particular produces response and counter-responses which are attempts by those in power to stabilize the organization. The time frame of introduction and response to new technology will shape the types of responses the organization makes (Manning 1992, 2008). Technology is a plastic or sponge concept, one that soaks up meaning and reflects the meanings attributed to it. It could be said also that technology is a blank screen onto which ideas, hopes, and dreams can be projected. In this way, the ambiguity of technology and its far-reaching consequences make defining a "technological innovation" and evaluating its effects difficult (Byrne and Marx 2011). Finally, the technology of an organization speaks to the socially constructed environment, the *external realities*, in which the organization operates. The organization officially only knows what is recorded and officially sanctioned; bureaucracies are webs of secrets, and what is known is a function of the channels by which the information of interest is sought processed, encoded, stored, and retrieved. External realities, globalization, and political trends shape the framing of technology, how and when the technology is used. Technology also speaks to, and is shaped by, the *internal realities*, or the pattern of power in the organization.

The position adopted in this entry is that while technology may alter specific organizational processes, the null hypothesis should be that organizations shape and alter the impacts of technologies. Placing technology in organizational context as such requires also interrogating the matter of what is being "produced" and the related decision-making. The process reproduces the patterned and conventionally sanctioned deciding that takes place and into which the technology is slotted. Policing is in fact a process. Policing is a people-processing system, above all and in that sense alone cannot be judged by the criteria or standards of economic organizations with a product, a market, and conventional standards for judging "profit" and "loss." Certainly, "people work" is complex and daunting because of the roles and statuses of the people who are dealt with. Ostensively, their social class, gender, ethnicity, and occupational status are relevant to the incident at hand, regardless of the basis for the negotiation; the standards employed are flexible and situational as people morally are defined as ends in themselves, rather than merely tools of the trade. The assumption of technological rationality, that technology has exclusively positive, systematic, predictable, and endearing consequences, is untenable. The craft of policing, archaic as it may be, is predicated on managing the vagaries of the human condition.

Technologies have powerful and seductive appeal to police departments. Large departments in particular routinely acquire new technologies, the most central of which are variations on computer-based activities and analysis. Thus, *prima facie*, the topic of police technology or technologies used by the police is a topic of growing interest to scholars (Nunn 2001; Chan 2003; Chan et al. 2001; Manning 2008; Byrne and Marx 2011; Mastrofski and Willis 2010; Willis and Mastrofski 2011). Some reviews of technology in police organizations are benignly optimistic and naively hopeful, but conclude with rather disappointed summaries (Willis and Mastrofski 2011; Mastrofski and Willis 2010), while others are dystopian (Nunn 2001; Byrne and Marx 2011). But because these reviews are descriptive, accept generalizations about reported changes, for example, from LEMAS data of dubious credibility, and lack any theoretical framework, it is unclear what explains the absence of "innovation." The tendency in practitioner-law enforcement journals is to be benignly positive about all new technologies, absent any systematic evaluation. Chan (2003) critically and systematically reviewed technology in police organizations using a framework adopted from Bourdieu. She concluded in a balanced overview that not only are there negatives associated with increased technology, but that ironically, increased technology may exacerbate the problems of factual overload, noise, stored data of modest validity, and resistance to change among patrol officers. Manning's study of crime mapping and crime analysis in three cities concludes that little has changed in the routines and practices of police as a result (See also Willis and Mastrofski 2011).

Consider now, prior to review of the topic of police and technology, two fundamental ideas that should shape any research on policing as an organization. Egon Bittner, in many respects the theorist of record in regard policing, states at the beginning of his well-regarded classic, *The Functions of the Police in Modern Society* (1970): "The [function] of the police can't be understood by working down from broadly conceived programmatic idealizations." (p. 4), He continues, arguing that the task of his book is ". . . to elucidate the role of the police in modern American society by reviewing the exigencies located in the practical reality which give rise to police responses *and* [italics in the original – ed.] by attempting to relate the actual routines of responses to the moral aspirations of a democratic polity." (p. 5). By "programmatic idealizations," Bittner is referring to social theory of the Weberian sort, disconnected from the exigencies and routines of organizational life. This means in turn that technologies are situationally contingent and relevant, rather than concretely obvious in their functions, meanings, and uses. Insofar as it is possible, then this framework requires an examination of the relationships between technology and the police organization to be ethnographically rich and textured, and somewhat reticent to accept evidence that does not acknowledge this ground in the contingencies and routines of police work. This postulate is explored below. It would appear that technology has replaced craftwork as the domain means by which work is accomplished (Bittner 1983). An ideal technology would be one in which the worker was not required to have any knowledge or interest in the task, only to follow instructions (Bittner 1983: 254). This creates an everyday contradiction with the practice of policing. Police work, in contrast, is above all a craft, a craft rooted in the intuitive grasp of contingencies, demonstrating collective trust using shared routines to manage the endlessly problematic, and connecting the ethics of the work with craft-like standards. These standards are disconnected with the technological efficiency. Policing in many respects is based on a "reading" of local knowledge, subjective knowledge of persons and places, and nonstandardized ways of knowing and doing.

This entry has several parts. The first part is a brief overview of police and technology and outlines types of technology in police organizations. The second part outlines innovations in police technology. The third part discusses current police innovations in information technology. The fourth sets out the problematic collocation of data in police organizations. The fifth part discusses the cynosure of the incident, the core of operational policing, and the ways in which it controls and shapes *how* whatever technology is available is used (or not).

Police and Technology: An Overview

The notion that scientific or technological innovations in police organizations would fundamentally alter the nature of policing and, in turn, social security and safety is almost 100 years old. It was set out as a principle of policing by August Vollmer and has long been associated with "professional policing." The steady march of more efficient communicational devices, the telegraph, call box, radio and two-way radio, telephone and cell phone (with the numbers for reaching police as 911, 999, and 311 in addition to routine local numbers), computer, and mobile digital terminals, have had impact. All have reduced the "pass through or processing time" of requests for service, escalated demand for policing, and stimulated a variety of ways in which the police screen, set aside, delay, and otherwise manage demand that is sometimes, albeit rarely, overwhelming. On the other hand, the long history of innovations suggests the following: active and passive officer and civilian resistance to innovations; abundant short-term and abandoned nonevaluated projects; on the ground cynicism about the latest-soon-to-be-abandoned "new thing;" a devotion to the craft-like nature of the work; **and** a slow creeping intrusion of some technological innovations that build on already processed facts-as-knowledge. These changes are overlays or superficial elaborations upon the truculently guarded craft skills.

Nevertheless, let us step back and recognize that many technologies are in operation in policing, and some of them are quite new. In general, and without question, *the most obvious and salient "technology," or means of accomplishing police work tasks, is talk* – face to face verbal communication. This is the baseline "information-processing system". The efficacy of what might be called "tactical talk" has not been researched by social scientists. While most social science attention has been paid to information technology and its role in shaping the police organization and police practices, and that is the focus here, there are four other types of technology at work. It could be argued that *weapons technology* – both lethal and nonlethal – is understudied by social scientists. *Transportation technology* is also assumed to be gradually improving in some sense, but emphasis has been placed on speed, presence, and availability, rather than economy or environmental prudence. *Training technologies* are used in the academy. The core remains symbolic (shaming, harassing, conditioning, and rapid response to orders) and secondarily, lengthy lectures, "academic learning" about the law, diversity, interpersonal relations, and problem-solving, combined with rote learning and exams. Police are also adopting *sensory processing and extending technologies*, transformative devices, means to enhance the senses: sight, hearing, smell, taste, and touch. They include radar guns, listening devices, night sights, crime scene and accident scene technologies, DNA, CCTV, fingerprinting, and ballistics, as well as the breathalyzer (See Nunn 2001). Increasingly, police are adopting these sense-extending devices to monitor events and track ex-offenders and the general population (Nunn 2001; Byrne and Marx 2011). While these last named technologies decrease time spent on tasks, and increase reliability, they are dependent on the quality of the input and interpretation, usually scientific, to convert facts or data, placing it in a context, thus creating information. The primary interests of social scientists are in the *information technologies*. The central operating information technologies have expanded rapidly in the last 40 years since the introduction of 911 systems and computer-assisted dispatch (CAD). The primary systems are: records management system (internal matters such as employee records – payroll, attendance, leaves, and illness); jail booking system; criminal histories and offender identification; CAD data; mobile digital terminal (MDT) records; GIS; crime analysis and crime mapping; UCR/NIBRS, AFIS, NICS, data and links to the Internet, search engines, national databases such as NCIC, and the FBI's new (2012) N-Dex, a national data exchange system and website [**http://www.fbi.gov/about-us/cjis/n-dex**]. Also national are UCR/NIBRS, AFIS, NICS, and the FBI's various crime lists. Not all states participate in submitting data to these databases, the quality of data varies widely and is unaudited, and the lists are frequently redundant and contain useless or dated information (Geller and Morris 1992). The maxim is that the more the better, regardless of the quality of the input. It must be said that the use of cell phones, police-issued, multiple private phones, "personal assistants," "smart phones," and other private communication – computer devices, including handheld cameras, scanners, and in-car fax-machines and mapping capacities, for example – have not been researched and their impact on practices not well understood.

Recent Police Innovations

Technologically driven management-based innovations are sprinkled across the many departments in the USA and elsewhere but their effect is not known. Research with known results has been carried out briefly, in only a few cities, and has not been widely and consistently evaluated and diffused (Willis and Mastrofski 2011). Police departments are:

- Developing and publishing on accessible websites and in hard copies strategic plans, annual reports, mission statements, and the like. Organizations may submit and revise annually a policing plan with goals, objectives, and a broad mission statement. These are overwhelmingly focused on "crime" and based upon unaudited and unreviewed data generated by the departments or Constabularies themselves. In recent years, especially in the UK, these may be related to performance criteria for ranks and linked in turn to bonuses or other perks.
- Offering varied communication channels to and from the public. Websites now display descriptive materials, some data on calls for service or crime patterns, and hyperlinks to other websites. Often headlined is a current crime, wanted offender, or seasonal concern, for example, traffic and school closures for holidays. Some are appropriately mundane such as warnings about drinking and driving, congestion as a result of sporting events, community activities, and local holiday celebrations. Chan et al. suggest that these channels, including crime reporting, are growing in use in the UK, but they are limited for the same reason that reporting in general is sociologically constrained.
- Assembling more integrated databases and systems. This process may include linking software and geocoded databases and computer-assisted dispatching (CAD) data that facilitate crime analysis and crime mapping (Manning 2008).
- Managing tactical deployment including short-term efforts directed to manage known problems of crime or disorder or longer programs based on crime analysis. The rise of an experimental criminology devoted to controlled, randomized experiments focused on crime control has resulted in a number of such research projects in Boston, Jersey City, Lowell, Massachusetts, and St. Louis.
- Adopting tools for career-assessment of young officers while also monitoring complaints to identify high-risk officers and guide supervisors in their evaluation tasks.
- Supporting a modest research capacity. In part as a result of Community Oriented Policing Services agency (COPS) funding, growing political pressure to demonstrate and be seen as "efficient," and pressure to employ systematic and unrelenting "smart management," large departments now have a "research and development office." Many departments receive and manage state and federal grants. For the most part, the research units are in fact merely data entry and tabulation centers.
- Establishing public relations units with explicit policies, specialized, nonsworn personnel, equipment, and offices (Mawby 2002). These units and their spokes persons are increasingly sophisticated and well-funded and are designed not only to respond to queries but also to propagate "good news." They communicate with media and the public routinely, in a noncrime-focused as well as in a crisis modality.
- Arranging institutionalized communicational channels by which public can contact police, such as providing business cards, e-mail addresses, and voice mail for uniformed officers as well as sites by which police can contact villains, victims, or witnesses on a case by case basis. Some departments have introduced programs to target groups, neighborhood associations, corporations, and other governmental agencies (Mawby 2002). There has been a move toward more communication from the public to the police. These include

incompatible *channels of communication* that connect the public to the police and between units within the police department. Police use websites, e-mail, cell and land-based phones, "snail" mail, personal visits to stations, face to face encounters, network communication via fiber optic cables, paper documents, and "blogs." None of these data is assembled or checked for overlap, inconsistency, validity, or utility. Included and increasingly important are the social media. In particular, Facebook, YouTube, MySpace, Twitter, and variations on these are used by the Police Service of Northern Ireland, Greater Manchester Police (UK), and New York City Police Department. These can be used, of course proactively to solicit information as well as reactively to be in receipt of useful information.

- Organizing internal management-oriented meetings of a compstat-like character that involve presenting crime maps, recent trends in crime, quality of life issues, current problems of police interest such as gang activity, changes in drug use, or guns fired. These have become wide-spread but their functions and the depth of analysis presented varies. The relationship between these meetings and the drop in crime in New York City in the 1990s is now seen as problematic (Eterno and Silverman 2012).
- Presenting data routinely to public meetings, in some places required by law. As a part of political changes brought on by the Patten Report, PSNI officers are required to report in local districts patterns of crime and disorder. Similar meetings are also held regularly in Chicago.
- Linking with city governments and private corporations to monitor public areas via CCTV (Goold 2004). These may include parks, train stations, bus stations, airports, and harbors, as well as areas of the city deemed trouble spots. A few police departments have added the aural and visual monitoring equipment designed to detect shots fired in high crime areas (Mares and Blackburn 2012).
- Developing some information-based operational infrastructures. This includes transforming management information systems into networks linking electronically created and sustained databases; making these networks accessible wirelessly; as well as seeking integration of various nonlinked (typically) paper databases such as criminal records, juvenile records, CAD data, and social facts about neighborhoods (Dunworth et al. 2000).
- Increasing the number of nonsworn employees. Demands for service as well as management by objective approaches, crime analysis, and some augmented research capacity, has led to an increasing civilianization. The present composition of police departments shows about 27–30 % of their employees are civilians (LEMAS 2000 survey). This figure is some 5–7 % higher in the UK.
- Innovating in regard to thinking about crimes that are planned, politically imagined, and organized, and belief-based such as human trafficking, anti-terrorist, and homeland security programs mounted by the NYPD and other large departments. These are funded by massive infusions of money through federal and state-based grants.
- Introducing more systematic database access and storage for investigative work. These systems also include the possibility of exchange of information concerning crimes that have area-wide significance. Formalized systems are used in the UK, such as HOLMES and HOLMES 2; they provide a format for entering data on serious crimes and enable sharing of key facts and evidence (See website http://www.holmes2.com/holmes2/index.php).
- Engaging in networks of data internationally while developing dedicated systems for specialized crime and investigative work. For example, in the UK, the Serious Organised Crime Agency (SOCA), established in 2006 the Home Office, which operates within the UK and collaborates (through its network of international offices) with many foreign law enforcement and intelligence agencies (www.soca.gov.uk/). It is the result of the merger of the National Crime Squad, the National

Criminal Intelligence Service, the National Hi-Tech Crime Unit, the investigative and intelligence sections of HM Revenue and Customs that investigate serious drug trafficking, and the Immigration Service, and the Assets Recovery Agency.

- Linking data and analysis through the Internet. The FBI has a massive network of data including phone calls, e-mails, addresses, and other information that is available for police agencies, and is working on linking the many state agencies including police, corrections, and social services in this database. (N-Dex cited above). There are also planned databases in the UK centered at GCHQ to monitor all cell phone, e-mail, and web-use, while the UK national database of DNA data includes 3.4 samples, and the fingerprint database includes 5.5 million prints. There are large datasets based on number plate recognition in the UK, and these are used for enforcement, fines, and other penalties assessment. It is estimated that there are four million CCTV sites in the UK although research suggests they have limited value (Welsh and Farrington 2009). Some of these are monitored directly in police stations, for example, in Belfast, while others are either privately owned or owned by local authorities.

Each of these examples is built upon a core of information processing. The new information channels operate embedded in a structure, strategies, and values largely unchanged since the 1920s. As Chan et al. (2001: 113) have argued correctly, these new channels produce more work, more paperwork, and more data within organizations already awash with uncollated facts. The fantasy of efficiency of course has not been validated. There have been no structural changes, setting aside 911 and computer-assisted dispatch (itself a means), since the mid-nineteenth century when detective work was added. The rest are add-ons to the basic model developed by Sir Robert Peel prior to 1829 – the model featuring reactive, uniformed, visible, full time patrol. The pastiche that has resulted has not been systematically evaluated; that is, an evaluation that assessed the extent of integrated functioning with respect to clear and established criteria and valid and reliable data in one or more police organizations over time.

The Problematic Collocation of Police Data

While the term system is used to describe the information modalities within a police organization it is better seen as an *occasioned assemblage* based on "need to know" and patterned by use and power. While many police departments have acquired new information technologies, and most departments typically possess many, they remain fragmented and scattered throughout the organization even if electronic, scarcely used beyond daily needs, and marginal to the core of the job as seen by working officers. They serve many functions and are assumed to perform in connection with certain assumed rationalities. These rationalities, connecting ends and means, are incompatible in theory and are mediated and ignored in practice. It is not surprising that any given system or facet of the information system is so intertwined with others that it cannot be distinguished for research purposes. This makes evaluation very imprecise and narrow, while the impact of information technology is amplified throughout the organization and its practices.

Consider these features of current modern information systems in policing. As Bittner wrote (1970: 64), "... police departments accommodate a colossally complicated network of secret sharing combined with systematic information denial." A reflection of this, among other structural aspects of the craft, he also avers (1970: 65), correctly "... even the most advanced among our police departments are not anywhere near the objective of developing adequate information and storage retrieval systems." His point is that the majority of information known is seen as personal knowledge, not shared, may not even be written down, a product of a combination of individualism and defensive occupational solidarity (p. 65). If this position is definitive, and it is consistent with my fieldwork, whatever technology shapes and is used to enter, store and retrieve information, is a modest part of what is known.

The stated *capacity* to gather and process data quickly, to store them in an accessible and orderly fashion, and to develop vast fact-based files of fingerprints, criminal records, lab reports, arrest documents, and cases, most of which are still in written and printed form (or both), is considerable. There is vast storage capacity and it is increasing, but storage capacity and use are not calibrated. The growth in storage capacity absent access and use reveals the tendency within policing and perhaps other public agencies to acquire systems without clear standards or stated purposes, and without considering the complexity of creating usable, and simple modes of interface, collocation, and analysis. What is useful is not fact, but fact in a context or information.

While the systematic use of databases such as the Internet is growing and linked to trends in policing new forms of crime such as human trafficking, white collar and Internet-based crime, these are not connected within and between police departments, or done so unsuccessfully, for example, HOLMES. Individual cases crossing police boundaries are rarely connected and investigated systematically.

Given these databases and their related platforms, it is not surprising to find that numerous *software* systems exist (e.g., ArcView, for geocoded material; Pop TRAK, for monitoring problem-solving; specialized programs for workload management; many spread sheets for accounting and noncriminal records maintenance). Changes in software are seldom done well "in house" and when out-sourced, lead to confusion because training is not provided, new systems have new "bells and whistles," and the oral culture for actually using them well may not exist. A parallel matter is that when federal or state funding expires for a given software, it is abandoned in time because those who knew the routines were transferred or retired, and no one knows how to run it (Nesbary 2001).

Many inconsistent user and backside *technology interfaces* are in place. A large police department will house: several servers, diverse and uneven mainframe access and work-place terminals as well as Internet connections via fiber optic cables. These work-place terminals have varied and varying memory and memory capacity. Perhaps as a result of the ad hoc accretion of these via technologies, uneven and unplanned purchasing, the influence of grants, vendors, trends, and fads, and now abandoned failed innovations, police possess disparate information technology clusters. These are not additive or cumulative in their effects. Unlike technology-sensitive organizations, such as universities and corporations, they do not systematically and routinely, that is, every 2–3 years on a rotated basis, replace software, laptop, and desktop computers. The work stations are often acquired ad hoc and used by several officers over the course of a 24 h period of three or more shifts. Large departments have diverse work stations, running software of various generations (several versions of Microsoft) and a sprinkling of MACS and IBM clones that do not "speak" to each other. Research suggests that they are rarely and poorly utilized (Dunworth et al. 2000; Manning 2008); and data transfer, especially of large data sets, is awkward and flawed. A variety of limited and many nonlinked access points exist.

The Incident Cynosure and Police Beliefs

The ground of policing is the work of the patrol officer. Ideas about the what, why, and how of the work is surrounded by and protected by beliefs about it. *These beliefs are the most profound impediments to information-based technologically driven change*. These beliefs vary within and across the segments of the organization, among civilians and sworn officers, and probably vary by years of service and gender, but they underlie the powerful world view of the uniform patrol segment. They are in many respects the primary interface – where the organization's boundaries are maintained, expanded, or contracted over time – between the police and the public.

Consider the incident itself: a situated face to face encounter. This matter of some interest to the police may arise from an on-view matter; from a call to that unit or another; from a cell phone, or a decision to make a stop. These incidents, especially those of most public importance, are seen as perceived as distributed in time and space, as semi-autonomous and bounded, and have little if any connection except as they are clustered within an area. These are here and now moments, and the future alone can tell whether some level of crime or disorder will arise and have to be dealt with, but it is believed that without some action, things will get worse. Officers believe one cannot predict the occurrence of any given event as regards time, even given local knowledge of an area and its residents or habitués. A general policing proposition is: social life is odd; "things happen." The patrol culture and cops' stories, a vague set of configured resources, are loosely fitted to guide response to a given incident confronting an officer. The somewhat heroic yet quixotic world of the officer is always at risk from society's oddities. One might say, "society is what it is, and changes little." Correcting such lasting and troubling things as homelessness, mental illness, poverty, crime, and inequality are beyond the necessarily limited scope of policing, and of a given officer, and to some degree their causes and consequences are beyond the understanding of the average police officer. This is not a comment on intelligence or education, but seeing the "big picture" is regarded as a hindrance, and ignoring it a logical necessity, given the pressing, repeated, anomalous demands of police work. Policing is about practical, decisive action that minimizes the consequences of the visible and present here and now incident. Thus, ratiocination, reflection, and paperwork are secondary to "real police work." Details, abstractions, reflection, and cogitation may in fact impede accomplishing this work, extend the job, and add unnecessary complexity. This concern, the **defined incident**, is itself a flexible and elastic matter. Many events can be overlooked, ignored, or made irrelevant, but those in which the police intervene are truly multidimensional, could be handled in a number of ways, and as argued above, are governed by forces largely out of control of the officer. Once defined, the incident shapes the needed information: that which is accessible and useful in the incident at hand, and arrives "just in time" for use by an officer. The extent to which the incident extends to other matters is problematic and an information search, if undertaken, is guided by past success with the source, whether personal or technological in nature. Inquiries from patrol officers or investigators tend to be restricted to referring to or querying a handful of trusted (external) data sources: vehicle registration and driving licenses, outstanding warrants and/or criminal records, and other field stops. It follows, then, that the defined incident and its contours drive the nature, number, and types of records kept. It may not be written up, or may be dismissed by a number of routine modes of closing an open incident. Officers emphasize: "keeping up the numbers," "output," or showing what counts: processing calls, moving on quickly to the next, making the odd traffic stop and citation, and showing one is part of a team. This means, by extension, that some information will be passed on often verbally, about dogs, guns, dangerous alley ways, and "cock-ups" that are heard as cautionary tales. Always, the technologies and related information sources required or used are matters for the officer at the scene to best determine. The cynosure of the incident is combined with the notion that "you had to be there" and the reluctance of officers to judge others officers' decisions.

While there are technologies that actively circumscribe action – the mobile digital terminal in the car and the cell phone if used – the technologies used by officers are not used for the intrinsic merits ascribed by computer scientists, sociologists, or other experts; they are occasioned or used when an incident is fraught with rich and troubling potentialities. They are used as and when, by the needs of the moment, interpersonal trust, and a subtle assessment of the audiences to which the information might be presented. The imagined audience for the report, for example, a sergeant or a prosecutor, or internal affairs officer, shapes the written report. If the record associated with a technology is to serve the immediate situation, one's partner, citizens, and on-lookers, it may have a lower horizon than something written for management. If it is seen as having broader, media-based attention, it has a longer and more uncertain horizon of possibilities. How problems are defined, approached, managed, and disposed of is situational, governed by context, and what is seen as a resource shifts as contexts change.

The incident cynosure reduces the likelihood of open channels of communication between the public and the police. Consider these beliefs that are intimately related to patrol work. The focus on "doing something" shapes interactions with the public, and is based on a distrust of the public. Police are often lied to, misled, and subject to provocative dramas. They learn to keep some distance from what is said, make quick judgments, trust their "gut feelings" and learn from experience, and to honor the oral traditions of the job. The cloud of the media, their preoccupations and biases are misleading and public fears, insecurities, and concerns are generally inconsistent with police interpretations of these problems. They are obligated to act, as Bittner (1970: 40) writes, in *spite* of the citizens' natural attitude (Citizens' requests, definitions of solutions to the situation) about a possible solution. There is parsimony about the job and its craft. It is assumed that what is written is just enough and just in time to accomplish what the officer anticipates accomplishing, no more. This often limits radically additional comments, asides, observations, and other matters that might, if included, link this incident to others. Although it is most entrenched and honored in the patrol segment, the ideology of the power of the present and the incident is not restricted to the patrol segment. It is revealed in the focus of detective work on the case and its clearance, and the ideology of crisis management at the command level.

The Fundaments of Police Information Technology: Some Propositions

What are the *grounding fundamentals* of the police organization and of policing as a practice? One might argue that police organizations shape the uses of technology, that is, simple assumptions about the "what" and the "how" of policing – the basis for the craft, what is seen as "the job." In this sense, the rationality of the worker, poised between work control and rate-busting, persists and determines production, for example, the speed and quality of responses to calls for service, clearance of crimes, visits to schools and neighborhood meetings, clearance rates, and cold cases. In the same sense, technology in use, whether in the communications center, patrol, or investigation, produces pressures to speed up deciding. Operators are monitored by several means to coerce them to rapid processing of calls for service and for categorical "by the numbers" dispatching. Patrol officers are encouraged to "keep up their numbers," for example, running number plates, making traffic stops, and returning to service promptly (Meehan 1998). Detectives are pressured to "produce" clearances; they quickly focus on cases that can be cleared. Technology compresses deciding time and simplifies complexity. There is a "group think" aspect to deciding because the greater the time constraints on action, given ambiguity in technological human interaction, the greater the tendency for coping based on collective cues. Nevertheless, work fills the time available to do it. One aspect of this (Goffman 1959: 110–111) is "make work," activities that maintain the appearance of working, whatever the actual level of activity, as well as the appearance accuracy, involvement, pace, and personal interest. In every police department, detectives claim they are overloaded regardless of the workload; patrol officers complain of the same. As MDTs are installed in patrol cars, officers with patrols increase their inquiries to databases (Meehan 1998; Meehan and Ponder 2002), especially if they have a partner in the car. Police technologies stimulate and shape work *routines* which in turn shape roles (Goffman 1959: 16). The segments of the police organization, for example, top management, middle management, and the lower participants, tend to be loosely connected through occasioned routines. Management dicta flow down, but little information flows upward. Given this segmentation of the organization communication networks, technology has different meanings and functions when the police organization is in a crisis mode. The influx of information about a riot, a disaster, or a flood of calls for service may require changes in command authority (it moves up in emergency), resources available ("assets"), and tactical deployment. In an emergency mode, the organization communicates internally and externally quite differently, more quickly and with less reflection and anticipation of outcomes. Technology amplifies police communications in crises and emergencies and increases uncertainty (matters that cannot be answered factually and are contingent and contextual). Police organizations have more images, facts, and records than ever before. As the police organization increases its capacity to make short-term surveillances and interventions, it increases the use of "high tech"-based activities, for example, heavily armed and elaborately uniformed SWAT teams for relatively benign incidents. These encounters increase the chances for a mistake in judgment, a bad shooting, a false entry, a response to an exaggerated risk. As has been emphasized above, information technologies are variously rewarding to the officer. Those with features that decrease effort in respect to valued routines – checking data from traffic stops, running field stops, credit checks by detectives, ad hoc retrievals – are used and praised; those that are associated with unwanted efforts or disvalued routines, "paperwork," whether electronic, typed, or handwritten, are ignored, sabotaged, or seldom used. Uneven performance of the IT is frustrating and leads to reduced use of the technology. As Chan has shown (2001), skill and knowledge of the workings of information technologies is not randomly distributed within police departments. It is used for different purposes by the command, patrol, and investigative units; varies by age and gender; and increasingly is a source of social distance between young patrol officers and senior officers. When a form of IT violates the zone of indifference at the bottom, that is, when monitoring and surveillance are seen as "big brother" activities, the relevant IT will be sabotaged, damaged, turned off, or not used. It is resisted. This is seen in connection with audio and video monitoring in patrol cars, transponders, and GPS devices used to track the movements of unit (Meehan 1998). While technology strives for unity in the face of complexity, it separates content or "propriety" from efficiency (Bittner 1983: 258). As Bittner points out (1970: 55), policing is a means to deal with complexity and with exceptions or contradictions. Rules for organizing are bureaucratic and formalized, but cannot define when and how to intervene. Quality of response is less valued than pragmatic resolution of incidents.

The illusion of technological advance is made murky by those who use the technology, and define the utility of its apparent capacities. Bittner (1970: 64) states this succinctly "...police departments accommodate a colossally complicated network of secret sharing combined with systematic information denial." And later, in a phrase that no evidence has yet contradicted since he observed this pattern in the late 1960s (1970: 65): "... even the most advanced police departments are not anywhere near the objective of developing adequate information storage and retrieval systems. Even is present efforts to make modern electronic technology would succeed in coding existing information, this would not encompass the knowledge that is currently neither shared nor recorded." His point is that much knowledge is "local knowledge" gathered by officers and occasionally shared within the oral culture; that much knowledge is not recorded or written down in any form, and that this body of knowledge is more useful in practice than that which is seen as centralized, administrative, stored, and formatted (Brown and Brudney 2003). The reticence is in part ignorance and lack of training, in part due to the here and now incident focus; and in part due to a resistance to sharing information that might advance the career of a given officer. The working rules that govern policing include a set of contradictory themes in action – information is personal property of a kind to be guarded and protected; the most useful technologies are those that have visible, useful, local, immediate, and consequential properties: checking on driving licenses, insurance; ownership and registration of the vehicle, databases concerning previous stops, arrests, and criminal record of driver and passenger; watch your back and avoid close surveillance (Chan et al. 2001). These join other reliable and trustworthy technologies such as vehicles, weapons, and everyday tools of the job (logbooks, second gun, accident scene equipment, cell phones, and the radio-computer when help is needed). Those technologies seen as functioning to carry out surveillance, track and monitor (and perhaps to punish) workers such as in-car videos, sound recording capacities, and traffic stop records, are untrustworthy and might be sabotaged, turned off or on as need be in the situation, or used as self-protective devices in problematic circumstances.

Conclusion

The focus and implicit purpose of police technology is somehow to increase efficiency, or the level of resources, effort, and time required to accomplish a given task or set of tasks. Policing, however, is ill-suited to considerations of efficiency minimally, because it acts in a turbulent rapidly changing environment, responds to complex, problematic, human situations, and is "people work." It requires slack resources in anticipation of a complex, lasting, and engaging emergency, natural or man-made, such as a hurricane, earthquake, riot, terrorist attack, or power loss. Patrol officers, the front line and most powerful segment politically in the organization, spend more than 50 % of their time doing "nothing," driving around randomly (Famega 2005) and this is valued and lauded as the essence of police work, and much admired by the public. The emphasis in technological innovation has been upon time saving, yet the most fundamental waste of time is random, unfocused, and aimless patrol. The work, on the other hand, is viewed to some considerable degree as a sacred, representative entity rather than as an industry. It produces at best trust. It is not surprising, furthermore, that innovations in policing vary widely in their purpose, focus, cost, and utility, and that they are rarely evaluated internally. This ensemble of technological add-ons can only at best be an augmentation of the fundamental technology, interpersonal skills, verbal and nonverbal, that are the tools of policing everywhere. The ensemble also reflects the *competing rationalities* within a police organization: the focus on various ends such as crime, on crime prevention, public trust and respect – the mandate itself – or on means of policing, more patrol officers, more arrests, more partnerships, and better and faster equipment.

This review has set out types of technology and focused on the information technology-based innovations that have been most visible in the last 30 years or so that are typically called "technology." They are numerous. The framework presented here, a concern for practices that constitute the social objects that are considered technology, has identified how the data and information systems are loosely coupled within police organizations, rarely integrated, inconsistent in detail and format, varying as to content and form, and widely scattered in many locations. Thus, claims for capacity are misleading as they do not consider actual usage and outcomes of such utilization. The incident or case and how to handle it is the focal concern of policing; technologies that are seen as reducing paperwork, surveillance, time and energy, and supervisory intervention in entrepreneurship on the ground are valued.

On balance, the research on the impact of technology on police practices is thin and there is very little systematic evaluation of force-level impact available except the work of Chan in Australia. It is very likely, however, that the primary impacts are subtle, not directly affecting the work on the street or anything resembling "crime," but on the following: the speed and availability of response, given a call for service from a land line or cell phone; the availability of records for jail occupancy, arrests, field stops, CAD, and MDT inquiries for the previous year or so (others tend to be destroyed or stored); the skill of younger officers with various forms of e-communications, and the quality and quantity of internal communications, for example, e-mails.

The resistance to innovative thinking, for example, problem-solving (Brown and Brudney 2003) and the power of the patrol segment to define the nature of the job remain. The ways in which the lower participants work around, with, and innovate in their usage of, technology remain unexplored (see however, Meehan 1998; Meehan and Ponder 2002).

Related Entries

▶ Conceptualizing of Police
▶ History of Technology in Policing
▶ Managing Innovation of Policing
▶ Role and Function of the Police
▶ Surveillance Technology and Policing

Recommended Reading and References

Bittner E (1970) The functions of the police in modern society: a review of background factors, current practices, and possible role models. United States Government Printing Office, Washington, DC

Bittner E (1983) Technique and the conduct of life. Soc Probl 30:249–261

Bittner E (1990) Aspects of police work. Northeastern University Press, Boston

Brown MM, Brudney JL (2003) Technology: learning organizations in the public sector? A study of police agencies employing information and technology to advance knowledge. Public Admin Rev 63:30–43

Byrne J, Marx G (2011) Technological innovations in crime prevention and policing. A review of the research on implementation and impact. J Police Stud 3:17–40

Chan JBL (2001) Technological game: how information technology is transforming police practice. Crim Justice Int J Policy Pract 1:139–159

Chan JBL (2003) Police and new technologies. In: Newburn T (ed) Handbook of policing. Willan, Cullompton

Chan J (2010) Internet technologies and criminal justice. In: Jewkes Y, Yar M (eds) Handbook of Internet crime. Willan, Cullompton

Chan JBL et al (2001) E-policing: the impact of information technology on police practices. Criminal Justice Commission, Brisbane

Dunworth T, Cordner G, Greene J, Bynum T, Decker S, Rich T, Ward S, Webb V, M. A. Abt Associates Inc. Cambridge (2000) *Information systems technology enhancement project (ISTEP)*. United States

Eterno J, Silverman EB (2012) The crime numbers game: management by manipulation. CRC Press, Boca Raton

Famega CN (2005) Variation in officer downtime: a review of the research. Policing Int J Police Strateg Manag 28:388–414

Geller WA, Morris N (1992) Relations between federal and local police. In: Tonry MNM (ed) Crime and justice: a review of research, vol 15. University of Chicago, Chicago, p 231

Goffman EM (1959) The presentation of self in everyday life. Doubleday Anchor, New York

Goold BJ (2004) CCTV and policing: public area surveillance and police practices in Britain. Oxford University Press, Oxford/New York

Manning PK (1992) Information technology and the police. In: Tonry M, Morris N (eds) Modern policing. The University of Chicago Press, Chicago

Manning PK (2008) The technology of policing: crime mapping, information technology, and the rationality of crime control. New York University Press, New York

Mares D, Blackburn E (2012) Evaluating the effectiveness of an acoustic gunshot location system in St. Louis, MO. Policing 6:26–42

Mastrofski SD, Willis JJ (2010) Police organization continuity and change: into the twenty-first century. Crime Justice 39:55–144

Mawby RC (2002) Policing images: policing, communication and legitimacy. Willan, Devon

Meehan AJ (1998) The impact of Mobile Data Terminal (MDT) information technology on communication and recordkeeping in patrol work. Qual Sociol 21:225–254

Meehan AJ, Ponder MC (2002) Race and place: the ecology of racial profiling African American motorists. Justice Q 19:399–430

Nesbary D (2001) Reports and communications: the acquisition of computer-aided dispatch systems: administrative and political considerations. Soc Sci Comput Rev 19:348–356

Nunn S (2001) Police information technology: assessing the effects of computerization on urban police functions. Public Admin Rev 61:221–234

Roberts KH, Grabowski M (1996) Organizations, technology and structuring. In: Clegg S, Hardy C, Nord WR (eds) Handbook of organization studies. Sage, London/Thousand Oaks

Weber M (1958) The Protestant ethic and the spirit of capitalism. Scribner, New York

Welsh B, Farrington DP (2009) Making public places safer: surveillance and crime prevention. Oxford University Press, Oxford/New York

Willis JJ, Mastrofski SD (2011) Innovations in policing: meanings, structures, and processes. Ann Rev Law Soc Sci 7:309–334

Information-Led Policing

▶ Intelligence-Led Policing

Innocence Commissions

Kent Roach
Faculty of Law, University of Toronto, Toronto, ON, Canada

Synonyms

Criminal Cases Review Commission

Overview

Increased attention paid to wrongful convictions over the last quarter of the century has led to the formation by the state of permanent commissions that have powers to refer convictions back to the courts after ordinary appeals have been exhausted. The best known of these commissions is the Criminal Cases Review Commission (CCRC) of England and Wales. It was created by statute in 1995 and has been operating since 1997. Before its creation, it was necessary to convince an elected official and a member of the executive to refer a conviction back to the courts once appeals had been exhausted. Indeed, such petitions to the executive remain the norm in Europe and North America.

Commissions with powers to refer convictions back to the courts after appeals have been exhausted are sometimes called innocence commissions (Roach 2010a, b) because of their concern about the wrongful conviction of innocent people. At the same time, the criteria that various commissions apply in referring convictions back to courts depend on the precise form of their enabling statute. The CCRC, for example, applies criteria relating to the safety of the conviction that are used by appeal courts in England and Wales, and such criteria may not directly address the question of innocence (Naughton 2010).

The phrase "innocence commission" has been used to describe a wide variety of very different institutions including (1) self-appointed study commissions with an interest in systemic reform of the criminal justice system (Gould 2007), (2) volunteer civil society groups that investigate or advocate on behalf of the wrongfully convicted (Roberts and Weathered 2009), (3) temporary or permanent state-appointed inquiries into specific cases and/or systemic causes of wrongful convictions (Roach 2010a), and (4) permanent state-appointed commissions with a mandate to investigate claims of miscarriages of justices and to refer convictions in individual cases back to the court after ordinary appeals have been exhausted. This entry will focus only on the fourth category, namely, state-based commissions with a power to refer convictions back to the courts after the exhaustion of ordinary appeals.

This entry will examine the background of the formation and composition of various commissions. It will then examine the state of the art in the scholarship and identify areas of controversy within that scholarship. Finally, the entry will conclude with an examination of issues for future research.

Background

Most jurisdictions do not have independent commissions with powers to refer convictions back to the courts after appeals have been exhausted. The background of the creation and composition of the few commissions that operate at present will be reviewed in this section.

The Criminal Cases Review Commission

The pioneer of commissions is the Criminal Cases Review Commission (CCRC) created for England, Wales, and Northern Ireland in the Criminal Appeal Act of 1995 c. 35, s. 8. The CCRC was created following recommendations by two Royal Commissions and after successive applications had been made to the Home Secretary as an elected official to refer several suspected wrongful convictions in terrorist cases back to the court. In 1992, Lord May in his second report concerning one of many wrongful convictions in terrorism cases discovered in the late 1980s and early 1990s recommended that new independent machinery be established by statute that would have the power and resources to investigate cases of suspected miscarriages and refer them to the Court of Appeal (May 1992, 92ff). The next year, the Runciman Royal Commission on Criminal Justice made more elaborate recommendations for the creation of what was to become the CCRC. It recommended that the power to refer cases of suspected miscarriages of justice to the courts should be removed from the Home Secretary and be given to a body that was independent from the executive and the courts. The body would be composed of both lawyers and lay people and have the power to appoint police and others individuals such as forensic scientists to investigate cases of suspected miscarriages of justice. It contemplated that decisions made by the body would not be subject to appeal or judicial review, but that applicants whose cases were not referred to the courts could reapply to the commission (Runciman 1993).

Although there was much support for the creation of an independent commission to replace the role of the elected Home Secretary in referring convictions back to the Court of Appeal and this proposal was eventually put into effect by Parliament in the Criminal Appeal Act, 1995, c. 35, some reservations were expressed about the commission from the start. One reservation was a concern that the commission would be too closely tied to the Court of Appeal because the Court of Appeal would have the ultimate power to overturn the conviction. This concern was heightened by the fact that the Court of Appeal had expressed some reluctance to overturn convictions in the terrorism wrongful conviction cases that led to the creation of the CCRC in the first place. Concerns were also expressed that applicants would not have a hearing or even full disclosure before the CCRC and that the CCRC might have to rely on the police for investigations even though police misconduct had been a contributing cause of the wrongful convictions that had led to the creation of the CCRC in the first place. There were also concerns that the new body would not act as an advocate on behalf of the wrongfully convicted (Pattenden 1996: 407–408, 421; Thornton 1993). These concerns have all resurfaced in later scholarship critical of the CCRC (Naughton 2010).

In the Criminal Appeal Act of 1995, the CCRC was created as an independent body appointed by the Queen on the advice of the Prime Minister. It consists of at least 11 commissioners, one third of whom must be lawyers with 10 years of experience and two thirds of whom must have knowledge or experience of the justice system (Criminal Appeal Act 1995, c. 35, s. 8). The CCRC has subpoena powers, but only with respect to public bodies such as the police (Criminal Appeal Act 1995, c. 35, s. 17). It can appoint police officers when necessary to conduct investigations (Criminal Appeal Act 1995, c. 35, s. 19), although this is rarely done. More commonly, the CCRC relies on its own caseworkers as well as its ability to commission reports from experts (Criminal Appeal Act 1995, c. 35, s. 21).

The CCRC has the power to refer a conviction to the Court of Appeal where it considers that there is a "real possibility" that the conviction will not be upheld (Criminal Appeal Act 1995, c. 35, s. 13). The "real possibility" standard has been defined by the courts "as more than an outside chance or a bare possibility but which may be less than a probability or a likelihood or a racing certainty. The Commission must judge that there is at least a reasonable prospect of a conviction, if referred, not being upheld. The threshold test is carefully chosen: if the Commission were almost automatically to refer all but the most obviously threadbare cases, its function would be mechanical rather than judgmental and the Court of Appeal would be burdened with a mass of hopeless appeals; if, on the other hand, the Commission were not to refer any case unless it judged the applicant's prospect of success on appeal to be assured, the cases of some deserving applicants would not be referred to the Court and the beneficial object which the Commission was established to achieve would be to that extent defeated. The Commission is entrusted with the power and the duty to judge which cases cross the threshold and which do not" (R. *v*. *Criminal Cases Review Commission ex parte Pearson*, [1999] EWHC 452 at para 17 (Admin.) (Eng.)).

One commissioner can make a decision to reject an application, but three commissioners must agree to a referral to the Court of Appeal. The CCRC rejects over 95 % of the approximately 1,000 applications it receives each year (Walker and McCartney 2008: 198). The CCRC does not publicly release its decisions on whether or not to refer a case to the Court of Appeal, and it is an offense for the CCRC to disclose information it has collected except in relation to its own statutory functions or for purposes of criminal, civil, or disciplinary proceedings (Criminal Appeal Act 1995, c. 35, §§ 23–24).

The basis for a conclusion by the CCRC that a conviction should be referred to the Court of Appeal will generally be "an argument, or evidence," not raised in the proceedings, but a reference can be made in "exceptional circumstances" in other cases (Criminal Appeal Act 1995, c. 35, § 13). Originally, the accused was able to raise any ground of appeal once a case had been referred back to the courts by the CCRC, but this was amended in 2003 to only allow appeals on grounds certified by the CCRC or on other grounds where the Court of Appeal has granted leave (Criminal Appeal Act 1995, c. 35, § 14). The CCRC's reasons for making a referral are not released publicly, and it has a discretion not to refer a case even if the statutory requirement of a "real possibility" that the conviction will be not be sustained on appeal is satisfied. In one well-publicized case, the courts upheld the CCRC's decision not to refer a case to the Court of Appeal because the applicant had already received a posthumous pardon and compensation (*Westlake v. Criminal Cases Review Comm'n*, 2004 EWHC (Q.B.) 2779 (Admin)).

Despite having some lay representation, the CCRC ultimately makes its decisions on legal criteria relating to the hearings of appeals. The CCRC does not directly consider factual innocence and has referred cases back to the Court of Appeal on technical legal grounds relating to changes in the law and procedural irregularities (Elks 2008, Chap. 1; Naughton 2010).

As of May 31, 2012, the Court of Appeal had heard 461 referrals from the CCRC, quashing 324 convictions but upholding 137 cases (CCRC 2012). Although the CCRC is not a party in cases that it refers to the Court of Appeal, it has a 70 % success rate in the cases it refers to the Court of Appeal. The courts have generally been receptive to the role of the CCRC, but in some cases have expressed reservations about the CCRC's decisions to refer historical cases or cases where there was no new evidence to consider (Elks 2008).

Since its inception in 1997, the CCRC has completed a review of over 14,000 applications, but has referred only 503 cases (CCRC 2012). In other words, the CCRC has rejected just over 96 % of the applications it has received. The difference in cases referred by the CCRC and those decided by the Court of Appeal is explained by the delays in preparing and hearing the new appeals ordered by the CCRC. The cases that the CCRC has referred to the Court of Appeal and the convictions that have been quashed cover a wide range of cases, but only two cases in the first 10 years of the commission have been classified as DNA exonerations (Elks 2008: 83) of the type that often characterize exonerations in the United States, which are much more dominated by murder and sexual assault cases (Garrett 2011; Elks 2008).

The Scottish Criminal Cases Review Commission

There is a separate criminal case review commission for Scotland, which has a different legal system than the rest of the United Kingdom. The Scottish commission has nine members. Like the CCRC, they include both lawyers and lay representatives. The Scottish commission may refer a case back to the courts on the basis that a miscarriage of justice may have occurred and that it is in the interests of justice that the case be referred (Crime and Punishment (Scotland) Act 1997, c. 48, § 25). A miscarriage of justice is the sole ground of appeal in Scottish law. Thus, the Scottish commission, like the CCRC, essentially applies the test for a successful appeal when deciding whether to refer a case back to the courts.

The Scottish commission has broader investigative powers than the CCRC, which include the ability to obtain a judicial order to examine any person with relevant information under oath as well as to require production of relevant material from any person, not just public bodies as is the case for the CCRC (Elks 2008). As of July 2012, the commission had referred 111 of 1,489 completed applications since the start of its work in April 1999 (Scottish Criminal Cases Review Commission 2012). This referral rate of 8 % is considerably higher than the CCRC's. The High Court has allowed a successful appeal by the accused against either a conviction or sentence in 59 of these cases.

The Norwegian Criminal Cases Review Commission

In 2004, Norway established its own Criminal Cases Review Commission by an amendment of Chapter 27 of its Criminal Procedure Act. It can refer cases back to the court if new evidence or new circumstances may lead to an acquittal or a more lenient sentence or if an international court has determined that a conviction violates international law. It can also refer cases if a participant in the case such as a judge, prosecutor, juror, or witness has been convicted of a criminal offense in such a way that might affect an applicant's conviction or sentence. Unlike in the CCRC and the Scottish commission, the Norwegian commission refers cases for new trials as opposed to new appeals.

The commission has five permanent members, three of which must have law degrees. They are appointed by the King in Council and are assisted by investigating officers with legal and police backgrounds. The Norwegian commission may summon witnesses to appear before it and subpoena evidence. It also can, and frequently does, appoint expert witnesses. It can also hold formal hearings, though this has only been done in one case. Although the commission does the investigation, it can and does appoint lawyers to represent the applicant especially in complex cases or in cases where the applicant may not be guilty because of mental disorder. More recently, the commission also has the power to appoint lawyers to represent victims of sexual offenses who may be interviewed as a result of an application to the commission. The most common ground for reopening cases is evidence of mental disorder and that occurs in about 40 % of case where the commission sends the case back for a new trial (Stridbeck and Magnussen 2012). The cases reopened include not only crimes of violence including sexual offenses but also financial, drug, and minor offenses (Stridbeck and Magnussen 2012).

Except in its first years of operation when there were higher numbers, the commission receives about 150 applications each year. Since 2004, the commission has reviewed 1,045 cases and has reopened 120 or 11 % of the applications. The conviction was quashed in 80 % of reopened cases, and the sentence was reduced in another 14 % of cases, suggesting a higher success rate for the Norwegian applicant than is found in the other commissions.

The North Carolina Innocence Inquiry Commission

The first, and at the time of writing only, jurisdiction in North America to create a CCRC-type body was the state of North Carolina. In August 2006, legislation was enacted creating the North Carolina Innocence Inquiry Commission (NCIIC) to determine claims of factual innocence from living persons (N.C. Gen. Stat. §§ 15A–1460–75 (Supp 2006)). This legislation grew out of recommendations made by another body, the North Carolina Actual Innocence Commission, which was a self-appointed body chaired by Chief Justice Lake of the North Carolina Supreme Court. The legislation creating the NCIIC was originally subject to a 4-year sunset provision imposed by the legislature, but this has now been extended indefinitely.

The North Carolina Innocence Inquiry Commission (NCIIC) is composed of eight voting members appointed by the Chief Justices of the North Carolina Supreme Court and Court of Appeals. The membership of the NCIIC must include a superior court judge, a prosecuting attorney, a victim advocate, a defense attorney, a sheriff, a person who is not an attorney or employed by the judiciary, and two others (N.C. Gen. Stat. § 15A–1463).

Unlike the British, Scottish, and Norwegian commissions, the NCIIC is limited to claims of factual innocence, which are defined as "a claim on behalf of a living person convicted of a felony in the General Court of Justice of the State of North Carolina, asserting the complete innocence of any criminal responsibility for the felony for which the person was convicted and for any other reduced level of criminal responsibility relating to the crime, and for which there is some credible, verifiable evidence of innocence that has not previously been presented at trial or considered at a hearing granted through postconviction relief". (N.C. GEN. STAT. § 15A–1460.) Thus, unlike the other commissions examined above, the NCIIC is also precluded from considering applications on the behalf of deceased persons (N.C. GEN. STAT. § 14–1467(a)).

Like the commissions examined above, the NCIIC screens out most applications made to it, but also has extensive investigative powers that include the power to issue subpoenas and compel the attendance of witnesses (N.C. GEN. STAT. §§ 15A 1467(d)–(f)). The NCIIC can refer a case to the courts for review by a majority vote on the basis that there is "sufficient evidence of factual innocence to merit judicial review"; a unanimous decision is required if the convicted person pled guilty (N.C. GEN. STAT. § 15A–1468). This is a narrower standard than that used by the CCRC or the Scottish commission, which as discussed above can refer cases on the basis that they will be overturned on any existing legal standard for appeals, including miscarriages of justice that are not related to factual innocence.

The NCIIC has discretion whether to hold public hearings when it decides whether to refer a conviction back to the courts; however, if it does hold a hearing, the transcript of these hearings must be released in cases that are referred back to the courts. If a case is referred back to the courts, the case will be heard by a three-judge panel with no previous involvement in the case. The judicial panel can dismiss all charges on the basis of a unanimous decision that there is a clear and convincing case of innocence (N.C. GEN. STAT. § 15A–1469(h)). Both the decisions of the NCIIC and the three-judge panel are deemed to be final and not subject to appeal or judicial review (N.C. GEN. STAT. § 15A–1470(a)).

As of January 2012, the NCIIC had closed 953 applications of 1,102 applications since it started operating in 2007, effectively denying relief in all but four of these cases. The commission has held four formal hearings and referred three of those cases to the special three-judge panel. Of those cases, the special three-judge judicial panel concluded that factual innocence had been established in two of the three cases with one of the two successful cases resulting in the exoneration of two previously convicted persons. Thus, the North Carolina process has so far resulted in the judicial exoneration of three previously convicted persons on factual innocence grounds (North Carolina Innocence Inquiry Commission 2012).

State of the Art and Current Controversies in the Literature

There is a growing amount of scholarship on the commissions, and a number of themes and controversies can be identified.

Commissions are often seen as "inquisitorial" institutions that allow state-financed but independent persons to investigate claims of wrongful convictions. The exact extent to which a commission engages in proactive investigation or relies on the result of investigations by others is not known, but research conducted on the role of legal representatives for applicants before both the CCRC and the Scottish commission engages with some of these issues. A study has found that while about one third of applicants to the CCRC have legal representation, almost two thirds of the small number of applicants who are successful in the sense that the CCRC refers their case to the Court of Appeal have legal representation (Hodgson and Horne 2009). This finding reveals only a correlation between legal representation and success in the form of a referral and might be inconclusive given the fact that applicants with stronger cases might be more likely to attract legal representation. Nevertheless, the qualitative part of this research revealed a number of cases where legal representatives for the applicant were able to change initial decisions made by the CCRC to dismiss the application. This suggests that lawyers can have a determinative effect in some cases, especially those in which the CCRC reverses its initial decision not to refer a case (Hodgson and Horne 2009).

A study of applications made to the Scottish commission from 2007 to 2009 found that the applicant was represented in 155 applications and unrepresented in 99 applications and that applicants with representation fared better at each stage of the proceedings before the commission (Scottish Criminal Cases Review Commission 2010: 19). The problem of whether this measures a simple correlation between strong cases of miscarriages of justice and representation is not resolved in this research (Scottish Criminal Cases Review Commission 2010). The Norwegian commission also encourages representation of applicants in more serious cases, again complicating the idea that commissions are purely inquisitorial institutions (Stridbeck and Magnussen 2012). More research is needed into how the commissions operate and the role that representatives play in determining the strength of applications made to the commissions.

Much of the literature on the CCRC has from the start focused on its relation with the Court of Appeal. From the creation of the CCRC, commentators have raised concerns that it will be too dependent on the relevant tests used by the Court of Appeal to accept new evidence and to quash convictions (Nobles and Schiff 2001, 2005). More recent literature in this regard has argued that the CCRC is not truly independent because of its reliance on the Court of Appeal and that it is not concerned enough about factual innocence (Naughton 2010). These arguments have not gone unchallenged in the literature, and others have argued that the Court of Appeal's rejection of a significant number of referrals by the CCRC and its occasional criticisms of the CCRC referrals demonstrates that the CCRC is not captive of the Court of Appeal (Zellick 2006). As discussed above, the Court of Appeal quashes convictions in about 70 % of cases referred by the commission. The controversy in the literature over whether commissions are or should be independent from the legal system is likely to continue. Michael Zander, however, had made the important point that "some of these criticisms of the CCRC... could perhaps with more justice be directed at the Court of Appeal" (Zander 2010: xviii). Future research into commissions should not be conducted in isolation but rather should study how commissions interact with other parts of the justice system.

A common feature of all four commissions reviewed above is that they reject the vast majority of applications that they receive. Concerns have been expressed with respect to the CCRC that "many failed applicants are now voicing their dissatisfaction with the paucity of reasons given for the Commission's refusal to refer their case" (Walker and McCartney 2008: 198). There is a need for more research about why applications are rejected and the response of the rejected applicants. Such research will contribute to better understanding of whether commissions enhance the legitimacy of criminal justice systems by admitting of the possibility of error or whether they may actually increase dissatisfaction with the justice systems among those convicted of offenses and other affected persons including crime victims.

The Runciman commission contemplated that the CCRC would be well placed to make recommendations on the basis of its cases about how to prevent wrongful convictions in the future (Runciman 1993: 185). Other commentators have made similar predictions (Findlay 2009). Some research suggests, however, that the annual reports of the CCRC have failed to address issues such as the regulation of forensic experts and the use of anonymous witnesses that may increase the risk of wrongful convictions and have instead focused on matters that affect the case review process including feedback it has received from the Court of Appeal and budgetary matters (Roach 2010a, b). This research also suggests that there may be a tension between a commission's quasi-judicial role in deciding whether a conviction merits referral back to the court and less impartial advocacy for systemic reforms. This research found that the CCRC intervened in matters of legislation before Parliament only once and more from the perspective of how a proposed narrowing of the grounds of appeal would affect the commission's own work and has stayed silent with respect to other legislative initiatives that arguably have increased the risk of wrongful convictions.

Other research has also found a decrease in press coverage of wrongful convictions since the creation of the commission (Mason 2010). This suggests that the CCRC may have bureaucratized and regularized miscarriages of justice. Before the creation of the CCRC, extensive media campaigns had often proven necessary to convince the Home Secretary as an elected official to refer convictions back to the court, whereas this power is now exercised by the independent CCRC. The media campaigns previously designed to persuade the Home Secretary created a sense of crisis that some commentators observed undermined the criminal justice system's legitimacy (Nobles and Schiff 2001).

Much can be gained by comparing the profiles of cases referred by a commission to the courts to the profile of wrongful convictions addressed without a commission. Britain has opted for a state-financed error correction model which has enjoyed considerable success both in processing large numbers of applications and in referring close to 500 cases back to the Court of Appeal. A striking feature of these cases is that two thirds are non-homicide cases and only two cases in the first 10 years apparently involved DNA (Elks 2008: 184). This profile of cases can be contrasted with those found in the United States where DNA evidence plays a much larger role and most exonerations have involved homicide/sexual assault cases. Much work remains to be done in comparing the profile of reversed convictions generated by publicly subsidized commissions with those produced by the volunteer work of innocence projects and lawyers, who may quite understandably focus on the most serious cases, and those in which it is easiest to find new evidence to overturn a conviction. Such research should examine the interactions and frequent tensions between volunteer innocence projects and commissions that are not limited to work for the factually innocent (Roberts and Weathered 2009; Naughton 2010). Such research also poses questions about the larger purposes of the criminal justice system and the influence that increased concerns about the conviction of the factually innocent have on the traditional law and politics of criminal justice systems normally focused on crime control and due process (Findlay 2009).

Future Directions for Research

The issue of the relationship between commissions and the courts to which they refer cases is likely to continue to command scholarly attention. The present state of the literature has focused on competing viewpoints about whether the CCRC is sufficiently independent of the Court of Appeal. Perhaps a more productive approach would be to identify particular areas where Court of Appeal decisions have presented obstacles to possible referrals. This could expand scholarship beyond the commissions themselves to the courts and their receptivity to the idea that convictions should be quashed after ordinary appeals have been exhausted.

Commissions in all jurisdictions reject the vast majority of applications made to them, and there is a dearth of research on these rejected applications. Relevant questions are whether the applications are frivolous and should be screened out or whether they provide an important safety valve in the system. Qualitative and quantitative research about the attitudes of unsuccessful and successful applicants to the process would be promising. Another question is whether the jurisdiction of the CCRC and the Scottish commission over sentences as well as convictions should be retained or whether the focus should be more squarely placed on wrongful convictions.

There is a continued need to conduct research on the proper definition and approach to wrongful convictions. Arguments have recently been made that the CCRC is not concerned enough with factual innocence (Naughton 2010). The existence of the North Carolina commission that is limited to factual innocence claims provides a fascinating natural experiment; some early comparisons suggest that the North Carolina process may overturn fewer convictions than under the broader approaches used in the England and Scotland (Sangha et al. 2010: 351–356). Although all the commissions decide not to refer the vast majority of applications that they receive back to the courts, the North Carolina commission refers an even smaller percentage of its applicants back to the courts, and this may be related to the difficulties of establishing factual innocence on clear and convincing evidence (Roach 2012). Research comparing North Carolina's factual innocence commission with other commissions will also engage with broader questions about the purposes of the criminal justice system.

Comparative research between jurisdictions that have commissions and between jurisdictions that do not have commissions is promising. So far, much of this research has compared England and the United States; it demonstrates how most American states (except North Carolina) have been reluctant to create commissions and instead have committed resources to extensive but frequently futile habeas corpus review (Griffin 2001, 2009). Comparative research should not just focus on commissions but take into consideration other aspects of the criminal justice system that may make it easier or more difficult to reverse convictions (Griffin 2009). There is a need to expand the universe of countries that are the subject of comparative review especially to include countries with both adversary common law backgrounds and those from more inquisitorial systems (Roach 2010b).

Other research based on a natural experiment approach could compare the profile of convictions that are overturned as a result of a commission process as opposed to the profile of cases that are overturned in jurisdictions that do not have commissions such as Australia, Canada, most of Europe, and most of the United States. One hypothesis that finds some support in the case profiles of convictions overturned after referral by the CCRC (see Elks 2008) is that commissions are more likely to overturn convictions in non-murder and sexual assault cases than those that occur in jurisdictions without a commission. Jurisdictions that rely on volunteer lawyers and innocence projects and that do not have publicly funded commissions may understandably prioritize murder and DNA cases over less serious cases where the accused is less likely to be subject to long-term imprisonment and where it may be more difficult to find new evidence to undermine the conviction. More research also needs to be done on the subsets of cases referred by commissions. Almost a quarter of the CCRC's referrals are in sexual cases (Griffin 2009: 123; Elks 2008, Chap. 9), and this high referral rate may in part be related to the CCRC's powers and practice of reviewing victim compensation and other claims for post-conviction statements that can impeach the complainant's credibility (Griffin 2009: 124). This raises the question of whether different standards are applied in different types of cases.

The different commissions that now exist should be compared on a number of grounds relating to issues such as the transparency of their processes. The CCRC in particular may be less open than the North Carolina commission. The investigative powers and resources of each commission can be compared with an eye to determining the optimal model for commissions that may be created in new jurisdictions. Similarly, the composition of each commission can be compared as can their engagement with systemic reform that attempts to prevent wrongful convictions in the future.

Another area of future research that should be pursued is whether the existence of a commission bureaucratizes miscarriages of justice. The routine work of those who process and reject the vast amount of applications sent to the commissions needs to be examined. The relation between commissions and civil society groups that advocate on behalf of the factually innocent and also with the media needs to be examined. At the same time, the ability of commissions to act proactively, for example, where the practices of a particular police force or expert witness have come into question, needs to be carefully examined. In such cases, commissions may not necessarily wait for applications but could conduct proactive audits (Roach 2010b).

Conclusion

Commissions with the power to refer convictions back to the courts after appeals have been exhausted have emerged as interesting and innovative new criminal justice institutions in a number of jurisdictions. Much of the first generation of scholarship on these commissions has undertaken the necessary steps of describing the basic composition and operation of each commission. The next generation of scholarship should use the work of commissions as a window into wrongful convictions and into the larger purposes of the criminal justice system.

One research priority should be to compare the profiles of convictions reversed as a part of a state-financed commission process and those that are reversed by the efforts of innocence projects and other volunteers. These cases can be compared with each other and with the systems of petitioning the executive to reopen convictions that are still used in most American and European states. In general, the next generation of research and scholarship on commissions should focus on situating commissions in the larger context of the criminal justice system.

Related Entries

- ▶ Causes of Wrongful Convictions
- ▶ Forensic Science and Miscarriages of Justice

Recommended Reading and References

Criminal Case Review Commission (2012) Statistics. http://www.justice.gov.uk/about/criminal-case-review-commission

Elks L (2008) Righting miscarriages of justice: ten years of the Criminal Case Review Commission. Justice, London

Findlay K (2009) Towards a new paradigm of criminal justice. Texas Tech Law Rev 41:1

Garrett B (2011) Convicting the innocent, Harvard University Press, Cambridge

Gould J (2007) The Innocence Commission. New York University Press, New York

Griffin L (2001) The correction of wrongful convictions: a comparative perspective. Am Univ Law Rev 16:1241

Griffin L (2009) Correcting injustice: studying how the United Kingdom and the United States reviews claims of innocence. Univ Toledo Law Rev 41:107

Hodgson J, Horne J (2009) The extent and impact of legal representation on applications to the Criminal Case Review Commission. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1483721

Kyle D (2004) Correcting miscarriages of justice: the role of the Criminal Cases Review Commission. Drake Law Rev 52:657

Maiatico JM (2007) All eyes on us: a comparative critique of the North Carolina Innocence Inquiry Commission. Duke Law J 56:1345

Mason P (2010) Press and release: UK News coverage of the CCRC since 1996. In: Naughton M (ed) The Criminal Cases Review Commission Palgrave, New York

May J (1992) Return to an address of the Honourable House of Commons dated 3 December 1992 for a report of the inquiry into circumstances surrounding the convictions arising out of the bomb attacks in Guildford and Woolwich in 1974, H.C. 296, § 12.25, at 93–94

Naughton M (ed) (2010) The Criminal Cases Review Commission: hope for the innocent? Palgrave, New York

Nobles R, Schiff D (2000) Convicting the Innocent. Oxford University Press

Nobles R, Schiff D (2001) The Criminal Cases Review Commission: reporting success? Mod Law Rev 64:280

Nobles R, Schiff D (2005) The Criminal Cases Review Commission: establishing a workable relationship with the court of appeal. Crim Law Rev 173

North Carolina Innocence Inquiry Commission (2012) Case statistics as of October 2011. http://www.innocencecommission-nc.gov/stats.html

Pattenden R (1996) Ashworth A (ed) English criminal Appeals 1844–1994, 407–08, 421

Roach K (2010a) The role of Innocence Commissions: error discovery, systemic reform or both? Chicago Kent Law Rev 85:89

Roach K (2010b) Wrongful convictions: adversarial and inquisitorial themes. North Carol J Int Law Comer Regul 35:387

Roach K (2012) Independent Review Commissions: lessons from North Carolina. Crim Law Q 58:283–302

Roberts S, Weathered L (2009) Assisting the factually innocent: the contradictions and compatibility of innocence projects and the Criminal Cases Review Commission. Oxford J Legal Stud 29:43

Runciman V (1993) Royal Commission on Criminal Justice. Her majesty's Stationary Office: London.

Sangha B, Roach K, Moles R (2010) Forensic investigations and miscarriages of justice. Irwin Law, Toronto

Scottish Criminal Cases Review Commission (2010) The impact of legal representation on applications to the

Scottish Criminal Cases Review Commission. http://www.sccrc.org.uk/ViewFile.aspx?id=476

Scottish Criminal Cases Review Commission (2012) Case statistics. http://www.sccrc.org.uk/casestatistics.aspx

Stridbeck U, Magnussen (2012) Opening potentially wrongful convictions – look to Norway. Crim Law Q 58:267–282

The Royal Commission on Criminal Justice (1993) The Royal Commission on Criminal Justice, report. HMSO, London

Thornton P (1993) Miscarriages of justice: a lost opportunity. Crim Law Rev 103:926–935

Walker C, McCartney C (2008) Criminal justice and miscarriages of justice in England and Wales. In: Ronald Huff C, Killias M (eds) Wrongful conviction: international perspectives on miscarriages of justice., pp 183–198

Wolitz D (2010) Innocence commissions and the future of post-conviction review. Ariz Law Rev 52:1027

Zander M (2010) Foreword. In: Naughton M (ed) The Criminal Cases Review Commission: hope for the innocent? Palgrave Macmillan, London

Zellick G (2006) Comment, facing up to miscarriages of justice. Manit Law J 31:555

Innovation and Crime Prevention

Paul Ekblom[1] and Ken Pease[2,3]
[1]Design Against Crime Research Centre, Central Saint Martins College of Arts and Design, University of the Arts, London, London, UK
[2]Jill Dando Institute, University College London, London, UK
[3]Loughborough University, Loughborough, UK

Synonyms

Adaptation; Arms races; Coevolution; Creativity; Design; Learning

Overview

The UK Cox Report on creativity in business (HM Treasury 2005: 2) identifies three key interlinked terms:

"Creativity" is the generation of new ideas – either new ways of looking at existing problems or of seeing new opportunities, perhaps by exploiting emerging technologies or changes in markets.

"Innovation" is the successful exploitation of new ideas. It is the process that carries them through to new products, new services, new ways of running the business, or even new ways of doing business.

"Design" is what links creativity and innovation. It shapes ideas to become practical and attractive propositions for users or customers. Design may be described as creativity deployed to a specific end.

Innovation, creativity, and design of course occur not just in the scientific and technological domains but also in the social, institutional, economic, environmental, commercial, and legal domains. And the business in question could be a criminal enterprise, whether run by a lone, casual thief or the Camorra. Design is covered in Entry "▶ Designing Products Against Crime" of this encyclopedia; this entry focuses on innovation and creativity.

Why should criminologists and crime scientists be interested in these concepts? Simply put, there is a dynamic such that innovation regulates both the quantity and the quality of crime opportunities and crime reduction solutions. Examples are given of obsolescent crimes and the process whereby they become obsolete. The metaphor of co-evolution or arms race between offender and preventer is found helpful in understanding the dynamic. It is noted that the consequence of this insight is the relative unhelpfulness for understanding crime trends of statistics of recorded crime, since the same crime label is attached to crimes whose nature and method has changed with the profile of presenting opportunities. The crime drop experienced throughout the Western world in the last decade of the twentieth century is interpreted in terms of a temporary advantage enjoyed by preventers, notably via enhanced security of motor vehicles. The integration is advocated of this understanding into the crime reduction enterprise and into the strategy and tactics that follow.

The scope of this entry encompasses three strands of thought. First, innovation may be taken to refer to deliberate changes in economic, technological, and social arrangements in society. These in turn change the supply of opportunities and wider opportunity structures

(Clarke and Newman 2006) for crime and/or influence the prevalence of precipitating factors such as prompts and provocations (Wortley 2008 and [this encyclopedia]). Second, innovation may be taken to refer to changes in offenders' repertoire and application of resources (Ekblom and Tilley 2000), organizational and business models, tactics and strategies for committing crime, and actively *generating* opportunities. Third, it may be taken to refer to counterpart changes in crime reduction. Clearly, the three strands are yoked together, with the first opening avenues in the second and requiring responsive changes in the third, which will in turn reverberate back on the others.

Fundamentals

Obsolescence

Given that criminals have a finite amount of time available to offend, new methods and targets of crime must drive out some of the old. Dermot Walsh, in an early attempt to describe the "obsolescence of crime forms" (Walsh 1994), distinguished a number of reasons for crime obsolescence (the classification given here is slightly reworked from the original). They were:

1. Legal abolition: legislative changes in permissible sexual behavior represent the most obvious examples here, with the decriminalization of suicide, eavesdropping, and challenging to fight affording more arcane examples.
2. Court action: Walsh is properly skeptical that claimed effects of court action were real, with the closest to an example he found persuasive being the decline of garrotting (strangling with cord or wire) following the passage of the Garrotters Act (1863) which enabled exemplary sentences. This instance is of course vulnerable to the criticism that legislators, acting when things are at their worst, are condemned to success by the operation of regression to the mean (Campbell 1969; Ekblom and Pease 1995).
3. Changes in Police and Forensic Science: "Forensic science has been most notably successful in thwarting the crime of poisoning. This has been accomplished through the techniques of small sample identification, and through co-ordinated knowledge, which meant that there ceased to be such a thing as an unknown poison" (p. 152).
4. Economic change: Walsh identifies the shift in transport from horses to cars as making crimes concerned with the market in horses obsolescent. Because of his focus, he omits to mention the corresponding rise in crimes linked to motor vehicles. Safebreaking was deemed a casualty of control of explosives and increased surveillability. The diminished need for companies to keep cash on the premises to pay wages could also have been mentioned.
5. Social change: A less censorious attitude to sexual proclivities and adventures was taken to reduce the scope for offences like blackmail, as well as to the decriminalization of some sexual acts, as mentioned above.
6. Population density: More contentiously, Walsh claims that increased population density "unfavourably affects the prospects for the commission of some crimes, for example bribery of voters or nomad-related crimes" (p. 154).

While one may not agree with all Walsh's conclusions, his work is a useful first review of the association between innovation, social change, and crime.

Innovation and Crime Statistics

The dynamic of innovation and desuetude has, as a little-acknowledged consequence, the relative meaninglessness of some aspects of aggregated crime statistics. This is because the practical actions which constitute offences with the same legal label have changed. At the simplest level, the targets of "theft" have changed from, say, pocket watches in the nineteenth century to mobile phones in the twentieth century. More subtly, for example, at one point in time, the offence of "unlawful abstraction of electricity," originally associated with consumers getting free energy by bypassing the point at which the electricity supply was metered, had to be used

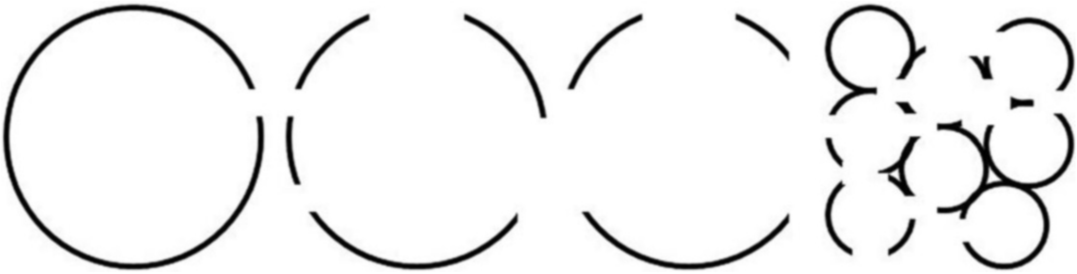

Innovation and Crime Prevention, Fig. 1 Schematic representation of opportunities in terms of obviousness and effort

to prosecute some early instances of computer hacking. This thereby illustrates innovation not only by criminals but also by the crime control authorities, faced as they were with agile offenders on the one hand and arthritic legislative responsiveness on the other. The recorded crime landscape likewise changed with the advent of one-to-very-many digital communication, where the techniques of fraud and pedophilic grooming are adjusted almost beyond recognition. With these changes have come huge increases in the number of potential victims (which of us has not received online invitations to profit in exchange for helping to release a bequest held in a foreign bank?) and the scope for joint offending, with illegal encounters being arranged from the comfort of one's home.

Innovation and Criminal Opportunity

While "criminal opportunity" is normally considered solely as a property of the *situation*, in reality (Ekblom 2011), it is an ecological concept, co-generated by the resources of the *offender* to exploit the vulnerabilities, gain the rewards, and handle crime preventers. An open window three floors up is only an opportunity to an offender equipped with courage, agility, and a ladder. One important kind of resource for innovation is J.J. Gibson's concept of *affordance* (Gibson 1979), the ability of some offender (say) to perceive opportunity and utility as they go about their environment (in a terrorist context, see Taylor and Currie 2012).

To take things further, Clarke and Harris (1992) distinguish between criminals who are opportunity *takers* and those who are opportunity *makers*. This concept is best described graphically in Fig. 1, which schematically shows a range of opportunities. At the left end is the perceptually obvious and logistically simple – an easy chance to steal an unattended laptop from a library desk, say. Moving towards the right, the shape of the opportunity becomes progressively less distinct, and progressively more causal preconditions for success have to be actively created by the offender (copy the key, time the guard patrols, prepare an excuse for being in the factory at the weekend. . .). Here, ingenuity is required to thread the crime technique between tricky multiple constraints and manufacture/exploit numerous simultaneous or sequential opportunity factors. At the far right, it is not even clear *which* if anything is the opportunity in question – what possible desired criminal event can be made to happen from what manipulable causes. Quite easy opportunities could be hidden in this tangle of possible conjunctions if only they could be discerned. One can think of talented entrepreneurs spotting a gap in the market that nobody else has. And the difference can also be noted between "regular earners" – criminal techniques that once invented continue to deliver for some time – and spectacular "silver bullets" that can only ever be used once, as with the 9/11 terror attack, where retrospective preventive action ensures opportunities are never the same again.

One presumes that, as in honest walks of life, there are more creative and less creative criminals. Studies of the former are rare although the issue is dealt with (Cropley et al. 2010) (Gamman and Raein 2010), who draw interesting psychological parallels between criminals, designers, and artists (in terms, e.g., of dyslexia). Whatever the proportion of creatives, the dissemination tools of the modern age enable the less well endowed to pick their brains (and indeed their locks – a quick check on Google reveals many lock-picking sites and even a lock-pickers' T-shirt!).

In the Industrial Revolution, the advent of machine tools was a bellwether of the "evolution of evolvability." A possible machine tool for the next generation of skilled offenders is the 3D printer which can manufacture realistic and effective scanning ports to be attached to the card slots of cash machines or reproduce high security handcuff keys. Another is the "script kiddie" – computer virus-making kits readily available to empower Internet criminals lacking their own expert knowledge to write the source code themselves. In today's open-source, open innovation (Chesbrough 2003) society, turning the clock back to an era of tools restricted to expert use or confined to professional guilds cannot be envisaged. Preventive interventions have to be put together which have a reasonably long effective life despite criminals' knowledge of their limitations and vulnerabilities and despite attempts to create countermeasures. The key strategic question then becomes one of how it might be possible to reduce the innovative capacity of offenders while building it up with preventers and/or designers.

A consideration of what is truly innovative serves as a counterpoint to the above comments on the flux-concealing nature of crime statistics. The UK government Foresight project "Cyber Trust and Crime Prevention" (Collins and Mansell 2004) wrestled with the issue of whether there were ever truly novel kinds of crimes – introduced in that case due to the arrival of the cyber world – or merely familiar techniques in new guises. Was online "phishing," for example, something entirely novel or merely the latest variant of a familiar "social engineering" con trick designed to extract personal details from an unsuspecting victim? In most cases the novelty is at the level of technical detail, but at a certain level of abstraction, the "crime script" (Cornish 1994) – the organized sequence of actions necessary to prepare, undertake the crime, cover tracks, and handle the proceeds – is similar. Following on from this, it may be helpful to systematically document changes in the nature and volume of modus operandi or perpetrator techniques over time to complement the numbers falling into given legal offence categories.

Offender Versus Preventer: Co-evolution and Arms Races

Change always presents new crime opportunities, which in turn call forth new preventive responses. There are usually three stages linking new products, and new ways of doing things, with crime. They are:

1. Innovation without considering the crime consequences
2. Reaping the crime harvest
3. Retrofitting a solution, usually partial

This sequence has been common. The Penny Black postage stamp was introduced in 1840 and withdrawn in 1841. This is because it was canceled by red franking ink, which in 1840 was water soluble, leading people to wash the franking ink off and reuse the stamp. The Penny Black was replaced by the Penny Red, canceled by black ink, which could not be washed off. Similar sequences are universally evident, most recently seen in mobile phones and laptop computers, where the mass manufacture of small, anonymous, valuable items, and the complexities surrounding numbers used for their unique identification, meant that they became a target for robbers and in thefts from cars. The magnetic stripe on credit cards was secure until thieves discovered techniques for "skimming," where data from the stripe on one card is copied without the cardholder's knowledge and placed on another card (to be combined with the cardholder's PIN number, obtained, e.g., by "shoulder surfing" in the cash machine queue). When silver mining opened up in California, silver was cast into conveniently sized ingots for transport to the East. These were stolen en route, despite protection by guards. The thefts ceased when the metal was cast in the largest available cannon ball mold. This was conveyed East on unprotected wagons without further problem (Lingenfelter 1986). Pocket watches were fastened to one's waistcoat by a rigid circular eye around the winder. This was a boon for cutpurses, who could simply snap the rigid eye and make off with the watch. The design response to this involved making the circular eye swivel, rendering it extremely difficult to steal and providing an interesting example of "target softening." Daguerreotype photographs can apparently be dated by this change.

The coin clipping example is particularly interesting because of its historical associations, namely, that Sir Isaac Newton was Warden of the Royal Mint when recoinage occurred. So the sequence was as above, first with smooth-edged (hammered) coins susceptible to clipping, the inevitable clipping problems, and replacement by milled edge coins whose well-defined perimeter apparently solved the initial problem. On his first morning at the Mint, Newton had to take an oath that:

> You shall swear that you will not reveal or discover to any person or persons whatsoever the new invention of rounding the money and making the edges of them with letters or graining or either of them directly or indirectly, so help you God. (White 1998 [229])

Unfortunately, the replacement coins were melted down and sold abroad as bullion because the metal commanded a higher price than the value of the coin set by the Treasury. Thieves were also inducing people to give up their old coins at unfavorable rates, then exchanging the old coins for new. Newton instituted official exchange mechanisms, and the recoinage eventually succeeded, but it illustrates that solutions to one crime are themselves untested innovations and can reap crime harvests of their own if not well thought through. It also illustrates the iterative nature of innovation, whether by criminals or crime preventers, because human ingenuity and social complexity (Chapman 2004) hinder getting things right first time and usually engender unforeseen consequences in what may be a "complex adaptive system" of diverse agents, each pursuing their own agendas in response to, or in anticipation of, the actions of the others.

A recent example of the neglect-harvest-retrofit sequence is the story of the Europrofile lock. This form of lock is almost ubiquitous in more recently built homes (or older homes with replacement doors) in Western Europe. In time, burglars realized that it could be easily overcome by the use of a wrench or a hammer and screwdriver. Once the protruding elliptical face of the lock is removed, the door will open. This form of attack is depicted through You Tube (another innovation making for communication with the effect of reducing the period between design and crime harvest). The increasing use of the technique in burglary led to the design of hardware devices to make disabling Europrofile locks more difficult.

Arguably the best metaphor for the dynamic linking innovation of crime is that of an arms race or the process of coevolution (Ekblom 1997, 1999). Offenders, over shorter or longer timescales, adapt their methods of attack to circumvent current preventive measures. They do so in several ways:

- Making tactical countermoves in situ – spraying quick-setting foam in car alarm speakers to deaden the sound.
- Turning crime prevention devices to their own advantage – antishoplifting mirrors work both ways; communal CCTV in blocks of flats has been misused by residents to spot which neighbors are going out, prior to burgling their flat.
- Turning designer themselves and developing tools – perhaps doing sophisticated reverse engineering of locks to understand and defeat the mechanisms of protection.

Preventers in their turn readjust by creating new devices or employing new methods of defense. The offenders then make further countermoves, and the process spirals on indefinitely. Safebreaking was one of the examples used by Dermot Walsh to illustrate obsolescent crime. The evolutionary struggle between safe manufacturer and safebreaker was described in more detail by Neil Shover (Shover 1996). The safebreaker, erstwhile skilled artisan of the crime business, is now an endangered species if not extinct.

Shover's story of the safe design and breaking "arms race" can be taken up in the mid-nineteenth century. By the turn of that century, safe cracker skills had required safe designers to turn to manganese steel in safe manufacture by dint of its resistance to drilling and fire. The spread of the oxyacetylene torch left even safes so constructed vulnerable to attack. The next countermove involved the development of laminated safes with alternate layers of manganese steel and copper, the better to conduct heat away from the point of attack. With safecracking by torches effectively rendered obsolete, offender attention switched to the locks, developing techniques to exploit vulnerabilities (sometimes analyzed by "reverse engineering" – careful dismantling of legitimately purchased mechanisms to see how they worked and how to overcome them). In some cases specialist tools to extract an entire lock became favored. After World War II, carbide and diamond-tipped drill bits briefly rendered the safe walls vulnerable again, forcing manufacturers to develop new laminates, on which most drill bits shatter. The production of ever-more sophisticated intrusion alarm systems placed time pressure on the safebreaker. Thieves who were unable to acquire expensive electronic equipment and specialist expertise to defeat the alarms had to give up safecracking. In a parallel development, electronic banking meant that many companies kept little in safes, so there were usually slim pickings for the endangered safebreaker species who had recently invested heavily in specialist equipment (overspecialization to best exploit what turns out to be a temporary niche or opportunity is a classic route to biological extinction).

But any such "victory" by one or other side remains forever provisional and precarious. Some new technology, for example, could appear which again tips the scales the other way. Like the frog which has evolved an enzyme which currently manages to protect its skin against fungal attack, society's currently functioning security measures may be just one or two "mutations" of criminal ideas away from susceptibility. And there are some grounds for pessimism here. To continue with a biological parallel, coevolution between pathogen and host, or predator and prey, has been likened to the Red Queen's Game, from *Alice through the Looking Glass* – you have to keep running just to stay on the same spot (Ekblom 2005; Schneier 2012). Worse, Schneier maintains that criminals have a natural advantage because they can usually come up with innovations faster than society can invent and deploy theirs; he also refers (p. 229) to von Clausewitz' nineteenth century military concept of the "position of the interior," wherein defenders have to protect against every possible attack, while the attackers just have to find one vulnerability in the defenses.

The rate of obsolescence will depend on the kinds of offenders involved and their resources and the kinds of social and technological changes that occur. And the rate of change is constantly accelerating. This means that the "breathing space," which society gets from a new preventive method before it is bypassed, is tending to diminish. In the face of this, Cohen, Vila, and Machalek wrote in 1995 that "contemporary crime control policies are hopelessly static" (p216).

Implications for Crime Control

If the arms race/coevolution characterization of the relationship between innovation and crime is accepted, there are a number of implications for how would-be crime preventers should act. The first thing is to address the pessimist position just stated. One could take the arms race to mean that coming up with new preventive measures is ultimately futile because, in time, criminals will inevitably find a way to defeat them. At the *tactical* level, many crime prevention approaches are undoubtedly a wasting asset (although this applies over a rather longer timescale than conventional displacement). Nevertheless, they cannot be abandoned. The long term can only ever be influenced via a succession of actions in the short term provided they are intelligently concerted. Led by responses to the rapidity of changes in criminal targets and techniques within the information and communications technology

(ICT) world, it is fair to say that the "hopelessly static" critique of 1995 no longer fully applies, though there is still far to go.

Viewed *strategically*, it must be accepted that crime reduction is *not* about individual innovations and their sometimes limited shelf life; it *is* about maintaining a dynamic imbalance between creativity and innovation by preventers and that by offenders.

In an earlier paper (Ekblom 1997), it was argued that those responsible for developing crime prevention policy and practice have to recognize the time-limited nature of security products and procedures and learn to cope with innovative, adaptive offenders, changing opportunities for crime and arms races by *gearing up* to become innovative and adaptive ourselves. (The concept of developing and building innovative capacity among preventive designers and crime prevention practitioners more generally is further considered in Entry "▶ Designing Products Against Crime" in this encyclopedia.) Strategies for gearing up include:

- Exploiting new technology for prevention (Ekblom 2005).
- Avoiding rigidity – where crime changes but the installed security system cannot.
- Encouraging variety (in contrast to the "Europrofile" lock – crack one, crack all).
- Basing the generation of *plausible* variety (i.e., that which has a good chance of working) on some combination of tested theoretical principles, research into what works at the level of detailed practical methods, and "think thief" imagination; here, integrated theoretical models such as the Conjunction of Criminal Opportunity Framework (Ekblom 2010, 2011) and design frameworks such as the Security Function Framework (Ekblom 2012) can help, where simple slogan-type frameworks hinder the creativity and innovation needed to customize specific interventions to crime problem, evidence of what works, and context.
- More generally getting crime prevention practitioners of all backgrounds and responsibilities to "draw on design" in generating good quality interventions with a strong chance of being cost-effective, practically implementable and acceptable, durable, and devoid of undesirable side effects; interventions which address the messy and complex nature of the constraints and requirements in crime prevention and the not unrelated prevalence of "implementation failure" in rolling out mass crime prevention programs.
- Designing products, places, procedures, and systems to performance standards rather than fixed technical specifications – enhancing "design freedom."
- Studying offender resources (Ekblom and Tilley 2000), current and future, to aid anticipation of what the product and its owner or guardian is up against (resources may be tools, as with a bolt cutter for defeating a bike chain; skills, such as deft movements for pickpocketing a music player, or more cognitive and social ones, such as intimidation or distraction techniques; more systematically professional preventers can consult or prepare "technology roadmaps" to get an idea of upcoming tools for attack).
- Making "attack testing" the default position in relation to the security of new products.
- Undertaking future proofing – whether of products (Armitage 2012) or even of laws (as with new drugs designed to circumvent too-specific proscriptions in law, by alteration of some nonfunctional part of the molecule – the answer here being to describe such drugs generically or even by their mechanism).
- Creating "pipelines" of new security arrangements to deploy as soon as current ones fail (as with satellite TV revenue-protection encoders, which have the rare advantage of immediate, "broadcastable" transmission of security upgrades to all users; with the increasing tendency to pervasive computing and networking, or "chips and wifi with everything," this strategy may become progressively more available to protect all kinds of property, though one can expect offenders to find ways to misuse the facility).

The above strategies center on crime preventers outracing offenders by innovation of their own or heading them off by anticipation. But, like Dick Dastardly in the *Wacky Races* cartoons covertly tying an anchor behind competitors' vehicles, it may be possible to thwart offenders by actively disrupting their innovative capacity. In effect, this could be done by turning all that is known about supporting legitimate enterprise and innovation into reverse or sowing uncertainty and distrust to reduce the level of networking that enables open innovation-type sharing of knowledge and other resources among offenders. Needless to say, this can only be done by subtle design and continuous response to feedback, because it is simultaneously necessary to avoid disrupting legitimate business.

Innovation and the Crime Drop: A Key Issue

The image of the crime/crime prevention struggle as a process of coevolution is consistent with one account of the crime drop which has characterized the Western world in the last decade of the twentieth and the first decade of the twenty-first centuries. Because the trend is common across nations, it cannot be the result of (divergent) national policies. The account with which it is consistent is the security hypothesis, still modestly characterized as an hypothesis despite its greater plausibility than its rival accounts. The hypothesis (Clarke and Newman 2006) is that change in the quantity and quality of security was a key driver of the crime drop:

> the one thing in common amongst all these countries, including the United States, is that they have all made a huge investment in security during the past 25 years, affecting almost every aspect of everyday life (p. 220).

Jan van Dijk and colleagues (van Dijk et al. 2007) independently reached the same conclusion:

> Perhaps a more significant factor inhibiting crime across the Western world is the universal growth in the possession and use of private security measures by households and companies over the past few decades. ICVS [International Crime Victimization Survey] -based trend data on the use of precautionary measures confirm that in all Western countries, without exception, the use of measures to prevent property crimes such as car thefts and household burglaries has risen drastically over the past 15 years (p. 23).

From evidence relating to vehicle theft in two countries, it was concluded that electronic immobilizers and central locking were particularly effective. It is suggested that reduced car theft may have induced drops in other crime including violence. Rival accounts, including demographic or economic change, better policing, gun legislation, the death penalty, lead levels in blood, and crack cocaine use, have all been subject to doubt after empirical analysis (Farrell et al. 2008; 2011). These authors point out that previous analyses focused on crimes of violence. They thereby failed the "phone theft test." They cannot therefore parsimoniously explain why many crime types fell in the 1990s while others including phone theft and e-crimes increased. They note the (inverse) coincidence of vehicle theft and security measures. Other trends, for example, in temporary relative to permanent theft, age of vehicle taken, and method of entry, yielded a "signature" consistent with the security hypothesis. Graham Farrell and his colleagues plausibly extend their vehicle crime analysis to other crime types, seeing car theft as a debut crime for young offenders and vehicle availability being the facilitator for many other crime types.

Related Entries

▶ Crime Science
▶ Designing Products Against Crime
▶ Desistance from Crime
▶ Green Criminology

Recommended Reading and References

Armitage R (2012) Making a brave transition from research to reality. In: Ekblom P (ed) Design against crime: crime proofing everyday objects. Crime prevention studies 27. Lynne Rienner, Boulder

Campbell DT (1969) Reforms as experiments. Am Psychol 24:409–429

Chapman J (2004) System failure: why governments must learn to think differently. Demos, London
Chesbrough H (2003) Open innovation: the new imperative for creating and profiting from technology. Harvard Business School Press, Boston
Clarke R, Harris, P (1992) Auto theft and its prevention. In: Tonry M (ed) Crime and justice: a review of research 16. University of Chicago Press, Chicago
Clarke R, Newman G (2006) Outsmarting the terrorists. Praeger, Westport
Cohen L, Vila B, Machalek R (1995) Expropriative crime and crime policy: an evolutionary ecological analysis. Stud Crime Crime Prev 4:197–219
Collins B, Mansell R (2004) Cyber trust and crime prevention: a synthesis of the state-of-the-art science reviews. Department for Business Innovation and Skills, London
Cornish D (1994) The procedural analysis of offending and its relevance for situational prevention. In: Clarke R (ed) Crime prevention studies 3. Criminal Justice Press, Monsey
Cropley D, Cropley A, Kaufman J, Runco M (2010) The dark side of creativity. Cambridge University Press, Cambridge
Ekblom P (1997) Gearing up against crime: a dynamic framework to help designers keep up with the adaptive criminal in a changing world. Int J Risk Secur Crime Prev 2:249–265
Ekblom P (1999) Can we make crime prevention adaptive by learning from other evolutionary struggles? Stud Crime Crime Prev 8(1):27–51
Ekblom P (2005) Designing products against crime. In: Tilley N (ed) Handbook of crime prevention and community safety. Cullompton, Willan
Ekblom P (2010) The conjunction of criminal opportunity theory. Sage Encycl Victimol Crime Prev 1:139–146
Ekblom P (2011) Crime prevention, security and community safety using the 5Is framework. Palgrave Macmillan, Basingstoke
Ekblom P (2012) The security function framework"; and "conclusion". In: Ekblom P (ed) Design against crime: crime proofing everyday objects. Lynne Rienner, Boulder Col
Ekblom P, Pease K (1995) Evaluating crime prevention. In: Tonry M, Farrington D (eds) Building a safer society: strategic approaches to crime prevention, crime and justice. Chicago University Press, Chicago
Ekblom P, Tilley N (2000) Going equipped: criminology, situational crime prevention and the resourceful offender. Brit J Criminol 40:376–398
Farrell G, Tilley N, Tseloni A, Mailley J (2008) The crime drop and the security hypothesis. Brit Soc Criminol Newslett 62:17–21
Farrell G, Tseloni A, Mailley J, Tilley N (2011) The crime drop and the security hypothesis. J Res Crime Delinquen 48:147–175
Gamman L, Raein M (2010) Reviewing the art of crime: what, if anything, do criminals and artists/designers have in common? In: Cropley D, Cropley A, Kaufman J, Runco M (eds) The dark side of creativity. Cambridge University Press, Cambridge
Gibson J (1979) The ecological approach to visual perception. Houghton Mifflin, Boston
Lingenfelter R (1986) Death Valley and the Amargosa: a land of illusion. University of California Press, Berkeley
Schneier B (2012) Liars and outliers: enabling the trust that society needs to thrive. John Wiley & Sons, Hoboken NJ
Shover N (1996) Great pretenders: pursuits and careers of persistent thieves. Harper Collins, London
Taylor M, Currie M (2012) Terrorism and affordance. Bloomsbury, London
van Dijk J, Manchin R, Nevala S, Hideg G (2007) The burden of crime in the EU. EU, ICS
Walsh D (1994) The obsolescence of crime forms. In: Clarke RV (ed) Crime prevention studies, vol 2. Willow Tree Press, Monsey, pp 149–163
White M (1998) Isaac Newton: the last sorcerer. Fourth Estate, London
Wortley R (2008) Situational precipitators of crime. In: Wortley R, Mazerolle L (eds) Environmental criminology and crime analysis. Willan, Cullompton

Institutional Anomie Theory

Beth Bjerregaard
Department of Criminal Justice and
Criminology, University of North Carolina at
Charlotte, Charlotte, NC, USA

Synonyms

American Dream; Anomie; Cross-national crime rates; Economic dominance; Institutional imbalance; Macro-social criminology

Overview

What began as a monograph aimed at accounting for the excessively high rate of violent crime in the United States (Messner and Rosenfeld 1994), Messner and Rosenfeld's *institutional anomie theory* (IAT), an identity provided by Chamlin and Cochran (1995), is perhaps the quintessential sociological theory of crime among the more recently minted theoretical approaches over the past two or three decades. The theory is a purely macro-social theory of crime and is in fact a cross-national theory of crime stressing the interplay of social structural and cultural feature of modern nation-states as the ultimate cause of crime. The entry below traces the origins of the theory, describes the theory in detail, reviews the empirical work testing key propositions of the theory, and discusses current controversies surrounding the development of the theory.

In 1994, Messner and Rosenfeld published their book entitled Crime and the American Dream where they first introduced their conceptualization of anomie.

Building on the work of Durkheim, Merton, and others (or building upon the Mertonian paradigm), Messner and Rosenfeld proposed the institutional anomie theory (hereafter, IAT) arguing that anomie is derived from the interplay between the culture and social organization of a society. Their theory shifts the focus away from Merton's development of individualized strain theory and toward the expansion and reformulation/clarification of the macro-level relationships articulated by Merton throughout his career. Thus far, empirical support for IAT has been mixed. The purpose of this entry is to describe Messner and Rosenfeld's institutional anomie theory and its theoretical underpinnings, review the empirical support for this theory, and discuss some of the controversies in the literature surrounding this theory as well as the policy implications derived from this theoretical perspective.

Origins of the Theory

Messner and Rosenfeld draw heavily upon the works of Emile Durkheim and Robert K. Merton to lay the foundation for their theory. Durkheim first introduced his concept of anomie in the late 1800s when he was describing the division of labor in societies as they transformed from the simple mechanical societies to the more complex organic societies of today. Durkheim noted that as societies become more complex and increase in size and density, the division of labor becomes more specialized and the capacity of societies to regulate behaviors weakens. Durkheim was not particularly concerned with explaining crime rates, but rather sought to explain conditions of deregulation or normlessness, which he termed "anomy" that he felt led to various forms of deviant behavior.

Durkheim believed that humans had insatiable desires and it was the responsibility of society to provide the appropriate restraints. Durkheim thought, "human activity naturally aspires beyond assignable limits and sets itself unattainable limits" (Durkheim 1951: 247). Durkheim proposed that in mechanical societies, the society exercised a relatively high degree of control over individuals. In societies undergoing rapid change, such as those evolving toward organic societies, this degree of regulation or control was weakened. In these situations, society loses the ability to regulate man's aspirations, and they can escalate beyond their abilities to fulfill them. If man's desires exceed his ability to satisfy them, then he will be in a state of anomie (Durkheim 1951: 246). Therefore, Durkheim was proposing that both the type and occurrence of deviance (including crime) is tied to changes in social conditions. Durkheim also described other situations where a society might lose its ability to regulate individual goals and therefore might lead to increases in anomie, including those of both economic prosperity and economic distress.

Merton, setting out to explain the relatively high crime rates present in the United States, attributes his initial ideas to the writings of Durkheim. While Merton borrowed the concept of anomie from Durkheim, he greatly expanded its meaning and proposed a slightly different interpretation. Merton's concept of anomie differed from Durkheim's in several important ways. First, Merton believed that it was society, not man's innate desires or human nature, that ascribed man's aspirations. Merton and later theorists recognized that American society was uniquely different from other societies. He believed that the distinctive feature of American society was its emphasis on what has been termed the "American Dream." He contended that American society placed an overemphasis/exaggerated emphasis on monetary success. As a result, all citizens are taught that success is measured by the achievement of monetary or material goals. Therefore, it is the society, not a biological drive, that provides man's aspirations. Society is also responsible for prescribing the acceptable means of achieving these cultural goals. Merton (1938) refers to these methods as the institutional norms. The institutional norms emphasized in America include means that are primarily based on what is referred to as the "Protestant work ethic." These include education, hard work, perseverance, etc. A balanced society, Merton contended, is one in which there is an equal emphasis or a balance between the cultural goals and the institutionalized means of achieving these goals (1957: 134). In such a society, utilizing these means should in and of themselves provide citizens with success or a sense of achievement. Further, in such a society, there would be a variety of values through which citizens could achieve/obtain culturally validated virtue or worth (Kornhauser 1978: 162).

In American, however, there is an overemphasis on the goals to the detriment of the means, setting up an environment where the achievement of the goals (winning) is more important than the process (playing the game) (Merton 1938). It is this disjunction between the goals and the means that Merton referred to as anomie. American society is, therefore, an imbalanced culture (Kornhauser 1978; Merton 1938). Kornhauser (1978: 163) points out that our culture is imbalanced because it does not provide a "sufficient variety of values whose achievement carries public recognition of moral worth." In American society it is the accumulation of wealth that is correlated with the moral values of stability, hard work, thrift, discipline, ambition, diligence, etc. (Kornhauser 1978: 165). In other words, achievement of the monetary goal affirms the existence of these other virtues. A balanced society would provide other ways to demonstrate these virtues. As Kornhauser points out in our society, there are no rewards for "hero of production" or "heroine mother of the year" (Kornhauser 1978: 163). In addition, Merton maintained that citizens are also taught to believe that this universal cultural goal is available to all and prescribes the acceptable means of achieving these goals. In reality, this is not the case. He believed that the availability of the institutional means varies by social class and is not equally available to all. In particular, the institutionalized means are restricted to those in the lower class structures. As Kornhauser (1978: 165) describes when items such as a "diploma, job title and bank account provide the only public insignia of moral worth, large numbers of people are shut out of the moral community defined by these achievements."

Although Merton proposes a variety of methods in which individuals can adapt to this strain or the inability to achieve the culturally prescribed goals through the institutionalized means, his theory is most useful in explaining utilitarian, profit-motivated crimes. Individuals who engage in these types of criminal behaviors, termed innovators, are simply utilizing the most efficient means available to them to achieve the universally accepted cultural goals. Individuals who engage in this type of criminal behavior accept and believe in the culturally induced goals. In fact, it is society that has pressured them into engaging in this type of conduct, by limiting their access to the institutionalized means of achieving these goals. Merton's proposed theory was, therefore, a theory of societal anomie. Bernard (1987: 268) simplifies Merton's cultural argument to the following propositions:

1. "The value attached to the goal of monetary success varies in different cultures and
2. At a cross-cultural level, variations in the value attached to monetary success are positively related to rate of utilitarian, profit-oriented criminal activity."

Institutional Anomie Theory

It is upon this foundation that Messner and Rosenfeld developed their version of anomie theory. Much like Merton, they developed their theory to explain the disproportionately high crime rates in America compared to other countries. Also similar to Merton, Messner and Rosenfeld believe that America places a high value on monetary success. However, they expand upon this idea by proposing that American society is unique in its emphasis on the cultural values embraced in the "American Dream." According to Messner and Rosenfeld, the cultural ethos of the American Dream refers not only to the goal of material and monetary success but also to the ideal that this goal can be pursued by everyone under conditions of open and individual competition (Rosenfeld and Messener 1995). This high level of crime coupled with a strong belief in the American Dream is termed "American exceptionalism" by Messner & Rosenfeld (2005).

The American Dream as envisioned by Messner and Rosenfeld embodies four fundamental value orientations: (1) individualism, (2) universalism, (3) achievement, and (4) materialism. Individualism is a highly prized value that stresses not only individual rights and freedoms but also the belief that individuals should be self-reliant and that success or achievement is individually accomplished.

Universalism is the belief that all individuals, regardless of their social class and background belief, are encouraged to aspire to the goal of monetary success. These two values combined underscore the belief that with hard work, anyone has the ability to succeed. Achievement stresses the importance of this success. One is ultimately judged by the achievements he or she has accumulated. Success in America is defined by the accumulation of monetary rewards. This belief that money is the ultimate reward in our society is the value of materialism. These four values together promote what has been referred to as the Horatio Alger ideal of rags to riches.

Embracement of the American Dream has had a number of benefits for the United States leading to an expanding economy, technological innovation, and social mobility (Messner and Rosenfeld 2007: 8). However, this value orientation, along with the capitalist economy present in the United States, also operates to influence citizens toward the unrestrained pursuit of materialistic and monetary goals. It is both the exceptionally high crime rates, along with the strong emphasis on the American Dream that sets the United States apart from other nations.

However, these cultural values in and of themselves are not what lead to exceptionally high rates of serious crime. IAT theory postulates that it is the interaction between culture and the social structure that leads to crime. The social institutions highlighted by Messner and Rosenfeld include the economy, polity, family, and educational system. These institutions assist societies in (1) adapting to the environment, (2) mobilizing and deploying resources for the achievement of collective goals, and (3) socializing members to accept society's fundamental normative patterns (Messner and Rosenfeld 2007: 72).

Each of these social institutions assists with one or more of these goals. For example, the economy, in its responsibility for the production and distribution of goods, helps in the adaptation to the environment. The family assists in propagating society and in caring and socializing the children. The political system is responsible for protecting members of society as well as mobilizing the resources to assist in attaining the normative goals. The education system not only helps to prepare members of society for assuming adult roles but also shares responsibility with the family to socialize society's youth. The relationship between these institutions is both dependent and antagonistic. For a harmonious society to exist, these institutions must cooperate with one another. For example, the economy is reliant on both the family and the educational system to produce quality citizens who will make efficient and effective workers. On the other hand, they are in competition with each other for resources and value orientations.

This interplay between these social institutions is crucial to IAT. One of the central assumptions of IAT is that a dominant economy coupled with weakened controls from the other social institutions in society produces a state of anomie conducive to serious crime. Economic dominance is expressed in a number of ways. First, the role of noneconomic social institutions is devalued in comparison to the economy.

For example, the role of the educational system in imparting knowledge is devalued, as the system is more valued for its role in training future workers. The second way this is expressed is by accommodation of economic norms by other institutions. This is exemplified by the family sacrificing family time for the sake of work schedules. Finally, economic values penetrate these other domains. For example, Messner and Rosenfeld point to terms such as "accountability" and "consumer-driven classroom" as evidence of the penetration of economic values into the educational system.

When the institutional balance of power is skewed, these noneconomic institutions are weakened in terms of their ability to regulate social behavior. When the economy is dominant, this can lead to crime by both stimulating anomie and by simultaneously weakening the organizational controls associated with the noneconomic institutions. IAT not only predicts anomic offenses when the institutional imbalance of power is in favor of the economy but also predicts criminal behavior in situations of imbalance. For example, corruption is expected to result in situations where the polity is the dominant institution. Likewise, "crimes in defense of the moral order" result in situations where family or religion dominates (Messner and Rosenfeld 2001). In sum, anomie results from "a strong cultural emphasis on ends over means" that is not guided by a strong moral foundation (Rosenfeld and Messner 1997: 214). In this type of situation, the type of means selected, including criminal behavior, is dependent on a cost-benefit or utilitarian analysis. Consequently, a society where the economy is dominant and the culture is anomic is expected to also have high crime rates, particularly serious and utilitarian or instrumental offenses.

Messner and Rosenfeld expanded upon Durkheim and Merton's concept of anomie by both refocusing on the macro-level causes of anomie and by examining the connections between the cultural ethos of the American Dream and the interrelationships among social institutions (Rosenfeld and Messener 1995). In more recent iterations of their theory, they have discussed how IAT explains the racial and gender distribution of crime. For instance, they suggest that females are less involved in criminal behavior because of their greater investment in noneconomic institutions such as the family as well as gender differences in value orientations (Messner and Rosenfeld 2007). In contrast, Messner and Rosenfeld argue that African-American investments are different with struggling attachments and alienation from the major institutions in our society.

Policy Implications

The policy implications derived from this theory are typically outside of the control of the criminal justice system. For instance, one way that societies can mitigate the influence of a dominant economy is by reducing its citizen's reliance on the market economy for their well-being. This is typically accomplished by welfare policies such as unemployment, social security, and health insurance. These policies help to restore the institutional balance of power as well as protecting individuals from market fluctuations (Batton and Jensen 2002).

IAT theory also suggests more complex initiatives that would be difficult and that many might be unwilling to undertake. Because the source of high crime rates lies within the tenets of the American Dream, reducing crime significantly would necessitate rethinking the basic tenets of the dream – many of which are responsible for the innovation and success that is often considered uniquely American. Likewise, crime would be reduced if we strengthened the noneconomic institutions in our society. Messner and Rosenfeld, for example, suggest strengthening the family by promoting initiatives such as paid family leave, flexible work schedules, and employer-provided child care (2007: 112). The primary criminal justice reform associated with IAT theory is to reduce our reliance on mass incarceration and to promote effective prisoner reentry programs (for a thorough discussion of this issue, see Rosenfeld and Messner 2010).

State of the Art (Empirical Tests of IAT)

While IAT theory is relatively new to the field, there have been a number of empirical tests conducted to determine the efficacy of the theory. These studies have primarily focused on examining the impact of the institutional balance of power between the social institutions and its effect on crime rates. Comprehensive and frequent investigations are most likely hampered by the methodological difficulties involving both the measurement of and collection of empirical data on the key concepts in the theory.

One of the first tests of IAT's main propositions was conducted by Chamlin and Cochran (1995). They examined 50 states in the USA to determine the impact of institutional imbalance on property crimes. They utilized the percentage of families below poverty level as their measure of absolute economic deprivation. In addition, they included measure for the polity, family, and religion. They examined the institutional balance by testing interactions between their poverty measure and measures for the church, family, and polity. They found all of these interactions to affect property crime rates in the directions predicted by the theory. They found that in the presence of strong noneconomic institutions, the influence of poverty on property crime was lowest. With one exception (family disruption), they also found that these results held up when they utilized alternate measure of income inequality in place of poverty rate.

Piquero and Piquero (1998) also examined property and violent crime rates using cross-sectional data from the United States. They also measured the strength of the economy by the percentage of persons below the poverty level. They likewise measured the strength of the polity (percent receiving public aid and percent voting in presidential election), education system (proportion of population enrolled full time in college, percentage of high school dropouts, teacher's salary relative to the average pay), and the family (percent of single parent). They explored whether the strength of these institutions moderated the effect of the economy on crime while varying their operationalization of the key concepts. They found mixed support for the theory. The percentage of persons in college (education) and the percentage receiving welfare (polity) decreased both property and violent crime rates. Also consistent with the theory, the percentage of the population below the poverty level (economy) and the percentage of single-family homes (family) increased both types of crimes. When examining the interactions, they found the cross-product term of economy and education to be significantly related to crime rates. Specifically, when more persons were enrolled full time in college, the economy had its weakest influence on property crimes and violent offenses. Similarly, the interaction between the economy and the polity significantly influenced the violent crime rate. However, when they utilized alternative measures for these concepts, the results were not replicated. They concluded that such tests of the theory are sensitive to the operationalizations of the key concepts.

In 2003, Maume and Lee predicted homicide rates in US counties with populations of 100,000 or greater. They examined the influence income inequality (economy), voter turnout for presidential elections (polity), the average educational expenditure per person (education), rates of divorce (family), and rates of adherence to civically engaged religious dominations (religion). One of their primary focuses was on determining the exact nature of the relationship between the economy and non-economy institutions on crime rates. Specifically, they tested whether noneconomic institutions mediated or moderated the influence of the economy on homicide rates. They also disaggregated homicide rates to predict both instrumental and expressive homicides. They found the economy to be a strong predictor of both types of homicides. In their analysis they found more support for the notion that noneconomic institutions mediated the relationship between income inequality and homicide. They found only limited support for the notion that they moderated this relationship as only the level of welfare expenditures interacted with income inequality.

Kim and Pridemore (2005b) advanced the literature by examining the utility of IAT theory in explaining homicide rates in 78 regions in transitional Russia. They measured the strength of the economy using a ratio of the top 20 % of income to the bottom 20 %. They also examined the impact of socioeconomic change on homicide. They only found limited support for direct effects on crime and no indication that the noneconomic institutions of the polity, education, or family interacted with change to affect crime. When they substituted their socioeconomic measure for the change measure and then looked at absolute deprivation (population below poverty line), their findings remained the same. Therefore, in this setting, social institutions did not help ameliorate the effects of poverty and social change on violence.

Schoepfer and Piquero (2006) expanded the discussion to examine the efficacy of IAT theory in predicting embezzlement rates between states in the USA. They found their measures of the economy (percent unemployed), policy (percent of registered voters who voted in general state and local elections), and education (percent that did not graduate from high school) to all be directly related to embezzlement rates. They also tested several cross-product terms and found the economy polity to be associated with embezzlement such that high levels of voter participation reduced the impact of the economy on embezzlement rates.

In 2008b, Bjerregaard and Cochran as well as Frerich, Munch, and Sanders assessed IAT utilizing cross-national data sets. Frerichs et al. (2008) predicted robbery and homicide rates among twenty developed, post-welfarist nations. They evaluated the institutional balance of power between several measures of the economy (Gini coefficient, P90/P10 ratio, union density rates, public social expenditures, unemployment, and mass imprisonment). Contrary to the expectations of IAT theory, they found that low female employment, high levels of union density, and low unemployment sometimes increased crimes. Likewise, they found that decommodification also increased crime. In contrast, low divorces and high public social expenditures and high imprisonment rates reduce crime. Bjerregaard and Cochran (2008b) also investigated homicide and theft rates across 49 nations. They found that the family and education institutions moderated the relationship between the economy and homicide. Likewise, education and the polity tempered the relationship between the economy and theft rates. Similar to Piquero and Piquero (1998), they found these results to be sensitive to the various operationalizations of the key concepts.

Collectively the results of these studies provide moderate support for IAT theory. The majority of the studies support the notion that noneconomic institutions moderate the relationship between the economy and crime (for an exception, see Maume and Lee 2003). Utilizing indirect measures of the key theoretical concepts support for the various propositions exists. The most common finding is that various measures of the economy including income inequality, poverty, and unemployment are related to crime rates. The findings regarding the influence of the noneconomic institutions are more mixed with various studies confirming these relationships and several refuting them. In addition, these results appear to be sensitive to the measurements utilized in the studies.

Another set of studies has focused on the idea that decommodification or welfare policies ameliorate the impacts of a dominant economy. This was first addressed by Rosenfeld and Messner themselves in 1997 when they studied the influence of Esping-Andersen’s measure of welfare expenditures (decommodification) on homicide rates in 45 nations. They found decommodification to be negatively related to homicide controlling for a variety of demographic and socioeconomic factors.

Hannon and DeFronzo (1998), integrating IAT with social support theory, looked at whether welfare assistance moderated the relationship between income inequality and crime in the United States. Confirming IAT theory, they found that high levels of welfare assistance did reduce the impact of economic deprivation on crime. Pratt and Godsey (2003) assess this relationship cross-nationally by measuring both social support (GDP spent on health care) and economic inequality (ratio of median incomes of the richest to poorest 20 % of citizens). They found economic inequality to have a strong, positive effect on homicides and social support to have a negative effect. Looking at the interaction between these two concepts, they found the effect of economic inequality on homicides to be lessened in situations of high levels of social support.

Savolainen (2000) went a step further and looked at sex-specific homicide rates using two cross-national data sets including the one previously used by Rosenfeld and Messner (1997). Income inequality (Gini coefficient) interacted with decommodification to influence homicide. Specifically, in situations with high levels of decommodification, income inequality reduced homicide rates for both men and women. These findings were supported in both samples. However, they also found the influence of income inequality on homicide to be restricted to nations with low levels of decommodification. Taking a slightly different approach, Stucky (2003) found the effects of economic deprivation on crime to be weakest in those metropolitan areas with strong local political structures.

Messner et al. (2002) studied this issue longitudinally investigating 150 countries over a 20-year time period. Confirming previous findings, they noted that income inequality had a positive impact on homicide rates. They also supported the notion that well-developed social welfare institutions mitigate the effect of income inequality on homicide. Batton and Jensen (2002) likewise researched this issue longitudinally across the United States, but they found that high levels of decommodification did not reduce homicide rates. They also examined these relationships across various time periods and found a negative relationship between unemployment and homicide for recent years. They attribute this finding to increasing levels of juvenile violence, potentially rising nonemployment rates and more females joining the workforce. Again, the findings from these studies are mixed, although the majority shows support for the proposition that decommodification buffers the influence of market forces on crime rates.

The above-discussed studies have been criticized in the literature for failing to test the fundamental assumptions underlying IAT. The studies discussed below have all focused on a different approach to researching IAT theory by attempting to confirm the notion of American exceptionalism. Specifically, they sought to determine both whether the values associated with the American Dream are universally supported and whether or not the United States is unique in its value system.

Value Orientations

Jensen (2002) was one of the first to decry the lack of empirical data on the degree to which societies embraced the goals and values described by Messner and Rosenfeld. Utilizing data from the World Values Survey, he examined value orientations across 38 nations. He found that the USA did not rank as high on materialistic values or normlessness compared with other nations. In fact, the United States scored toward the top in terms of rating the family, religion, and leisure as important. The USA also ranked sixth in agreement with the statement that “less emphasis on money and material possessions would be a good thing” (p. 59). In addition to ranking value orientations, Jensen also analyzed the extent to which decommodification influenced crime rates by strengthening the family. He discovered that the higher the investment in decommodification, the less likely citizens were to have children and the more likely they were to divorce (p. 62). Decommodification was also positively related to the importance of leisure and negatively related to the importance of religion, and not significantly related to the importance of family. Ultimately decommodification was found to significantly increase burglary rates. The findings from this study are clearly contrary to the expectations of institutional anomie theory.

Chamlin and Cochran (2007) followed suit and utilized the World Values Survey to measure cultural attitudes and ultimately determine how they correlated with homicide and robbery rates.

They also do not find evidence that the United States is particularly materialistic. In fact, they find that the "fledgling, market economies that emerged from the dissolution of the Soviet Union most enthusiastically embrace the so-called 'American Dream'" (p. 53). There was also no significant correlation between the importance of work and homicide and robbery rates. Interestingly, they found that if they limited their analysis to the economically developed nations, this correlation became strong and significant, implying that IAT may have salience only among developed nations.

In order to determine cultural values, Muftic (2006) took a unique approach and compared American to foreign-born college students. She asked the students several questions designed to tap into core value orientations such as individualism, achievement, and fetish for money and then correlated those values with cheating behaviors. Consistent with IAT theory, she found that students born in the USA indicated higher dedication to universalism and monetary fetishism and higher levels of cheating behaviors. These value orientations were also found to be related to cheating behaviors (except for the value of achievement). She also investigated attachment to and importance of family, polity, education, and the economy. In a multivariate analysis, the family and polity were found to be significant and negatively related to cheating. However, students who were employed (involvement in the economy) were less likely to cheat. Finally, attachment to or involvement in noneconomic institutions were not found to moderate the influence of values on cheating. Cullen et al. (2004) applied IAT to explain unethical business practices. Their findings were also mixed. The findings with regard to the values of universalism and pecuniary materialism support IAT, while the results for individualism and achievement do not.

Baumer and Gustafson (2007) also looked at geographic areas in the United States. They used individual-level data from the General Social Survey to measure commitment to pursing monetary success and the legitimate means of achieving that success and macro-level measures of the polity, education system, family, and religion. They found broad support for IAT. Their interactive models showed that when there are strong beliefs in the cultural goals of monetary success coupled with low attachment to legitimate means, higher rates of instrumental crimes result. When they considered the impact of social institutions only, the family and welfare assistance reduced this effect.

Controversies in the Literature

While these studies advance certain tenets of IAT, the findings overall are mixed with several propositions finding support and others not. All in all there is more support for the notion that social institutions moderate rather than mediate the influence of the economy on crimes. It is important to note that several studies find these results to vary depending on the operationalizations utilized.

Despite the advances made in empirically testing the theory in the past decade, a number of issues remain unresolved. For instance, the exact nature of the causal relationship between social institutions and the economy in affecting crime rates is somewhat uncertain. While the majority of studies have supported the notion that social institutions moderate the relationship between economy and crime, a few have refuted this perspective instead finding that these institutions play a mediating role (e.g., Maume and Lee 2003).

Another controversy that currently exists is the extent to which the core values of the American Dream central to IAT are universal. Findings surrounding this question have been mixed, with several researchers refuting this claim (e.g., Jensen 2002). Further, some research appears to suggest that the values held by Americans are not as unique as the theory suggests failing to provide support for the notion of American exceptionalism (Chamlin and Cochran 2007; Jensen 2002).

One of the most challenging issues facing the advancement of this theoretical perspective is the fact that IAT is not readily amenable to falsification. In general, the studies reviewed imply partial support for IAT. This may be in large part due to the fact that researchers are only indirectly testing main propositions of the theory due to the difficulty in both operationalizing the key concepts (e.g., economic dominance) and data challenges inherent in cross-national research. In the future, we need to find adequate sources of measuring crime rates cross-nationally. It would be especially helpful if we could disaggregate homicide offenses so that researchers could focus on the violent acts committed for instrumental or utilitarian purposes. Better data sources would allow future researchers to expand the scope of the dependent variables utilized.

Open Questions

There are still several open questions that exist in determining the efficacy of IAT. One of the primary issues involves the extent to which IAT is compatible with other theoretical perspectives. Rosenfeld and Messner (1997: 218) acknowledge that they are "heavily indebted to insights associated with traditional bonding and social-disorganization theories." They also recognize that "Mertonian insights about the social structuring of criminal incentives and opportunities have yet to be incorporated systematically and formally into the framework of IAT" (Messner and Rosenfeld 2006: 144).

Relatedly, Jensen (2002) has pointed out that other theoretical perspectives could likewise explain several of the findings from the literature testing IAT. For example, he points out that the mediating structure relating economic dominance to crime is consistent with social control and social disorganization perspectives. Further, he criticizes previous empirical tests of IAT for failing to incorporate variables central to other criminological theories into the analysis (p. 56). Future research needs to more fully incorporate key concepts from other theoretical perspectives in order to disaggregate the impact of IAT.

Another issue is the degree to which IAT can be integrated with micro-level explanations of crime. Both Durkheim and Merton advance individual-level explanations along with their structural focus. In fact, Messner et al. (2008: 177) point out that in integrating these two levels of analysis in research on IAT, it will be necessary to include indicators from other theoretical perspectives, and they encourage further efforts along these lines.

Finally, the applicability of the theory to explaining offenses other than serious crimes such as robberies and homicides is not well developed. Although these offenses were clearly the focus of Messner and Rosenfeld's presentation of the theory, it is unclear the extent to which the theory would apply to other types of crimes. Researchers have started to explore this issue predicting unethical business practices (Cullen et al. 2004), cheating behaviors among college students (Muftic 2006), and embezzlement rates (Schoepfer and Piquero 2006).

Messner and Rosenfeld offer us a unique reformulation and expansion of Merton's original anomie theory. In doing so they focus on the role of cultural values, social institutions, and the ability of these institutions to regulate behavior. Their theory helps describe the role of anomie in explaining cross-national variations in crime in general and high rates of serious crimes in the United States in particular. Finally, they refocus policy initiatives toward the wider cultural and social structure issues. The challenge will be to operationalize the key concepts in the theory to better advance our empirical tests of this theory. Messner and Rosenfeld have introduced an intriguing explanation for America's serious crime rates and presented a challenge for future researchers to develop the means to empirically assess its validity.

Related Entries

- ▶ Anomie and Crime
- ▶ Monetary Strain and Individual Offending
- ▶ Poverty, Inequality, and Area Differences in Crime

Recommended Reading and References

Batton C, Jensen G (2002) Decommodification and homicide rates in the 20th-century United States. Homicide Stud 6:6–38

Baumer EP, Gustafson R (2007) Social organization and instrumental crime: assessing the empirical validity of classical and contemporary anomie theories. Criminology 45:617–663

Bernard TJ (1987) Testing structural strain theories. J Res Crime Delinquency 24:262–280

Bernburg JG (2002) Anomie, social change and crime: a theoretical examination of institutional-anomie theory. Brit J Criminol 42:729–742

Bjerregaard B, Cochran JK (2008a) Want amid plenty: developing and testing a cross-national measure of anomie. Int J Confl Violence 2:182–193

Bjerregaard B, Cochran JK (2008b) A cross-national test of institutional anomie theory: do the strength of other social institutions mediate or moderate the effects of the economy on the rate of crime? West Criminol Rev 9:31–48

Cao L (2004) Is American society more anomic? A test of Merton's theory with cross-national data. Int J Comp Appl Crim Justice 28:17–31

Chamlin MB, Cochran JK (1995) Assessing Messner and Rosenfeld's institutional anomie theory: a partial test. Criminology 33:411–429

Chamlin MB, Cochran JK (2007) An evaluation of the assumptions that underlie institutional anomie theory. Theor Crim 11:39–61

Cochran JK, Bjerregaard B (2011) Structural anomie and crime: a cross-national test. Int J Offender Ther Comp Criminol 55:1–15

Cullen JB, Praveen Parboteeah K, Hoegl M (2004) Cross-national differences in managers' willingness to justify ethically suspect behaviors: a test of institutional anomie theory. Acad Manage J 47:411–421

Durkheim E (1951) Suicide. The Free Press, New York

Frerichs S, Munich R, Sander M (2008) Anomic crime in post-welfarist societies: cult of the individual, integration patterns and delinquency. Int J Conflict Violence 2:194–214

Hannon L, DeFronzo J (1998) The truly disadvantaged, public assistance, and crime. Soc Probl 45:383–392

Jensen G (2002) Institutional anomie and societal variations in crime: a critical appraisal. Int J Sociol Soc Policy 22:45–74

Kim SW, Pridemore WA (2005a) Social change, institutional anomie and serious property crime in transitional Russia. Brit J Criminol 45:81–97

Kim SW, Pridemore WA (2005b) Poverty, socioeconomic change, institutional anomie, and homicide. Soc Sci Quart 86:1377–1398

Kornhauser R (1978) Social sources of delinquency. University of Chicago Press, Chicago

Maume MO, Lee MR (2003) Social institutions and violence: a sub-national test of institutional anomie theory. Criminology 41:1137–1172

Merton RK (1938) Social structure and anomie. Am Sociol Rev 3:672–682

Merton RK (1957) Social theory and social structure revised. Free Press, Glencoe

Messner SF (2003) An institutional-anomie theory of crime: continuities and elaborations in the study of social structure and anomie. Cologne J Sociol Soc Psychol 43:93–109

Messner SF, Rosenfeld R (1994) Crime and the American dream, 1st edn. Wadsworth, Belmont

Messner SF, Rosenfeld R (1996) An institutional-anomie theory of the social distribution of crime. In: Siegel L, Cordella P (eds) Contemporary criminological theory. Northeastern University Press, Boston

Messner SF, Rosenfeld R (1997) Markets, morality, and an institutional – anomie theory of crime. In: Passas N, Agnew R (eds) The future of anomie theory. Northeastern University Press, Boston

Messner SF, Rosenfeld R (2001) An institutional-anomie theory of crime. In: Paternoster R, Bachman R (eds) Explaining criminal and crime: essay in contemporary criminological theory. Roxbury, Los Angeles

Messner SF, Rosenfeld R (2004) Institutionalizing criminological theory. In: McCord J (ed) Institutions and intentions in the study of crime: beyond empiricism, advances in criminological theory, vol 13. Transaction, New Brunswick

Messner SF, Rosenfeld R (2005) The present and future of institutional anomie theory. In: Cullen FT, Wright JP, Blevins KR (eds) Taking stock: the status of criminological theory advances in criminological theory. Transactions, New Brunswick, pp 127–148

Messner SF, Rosenfeld R (2006) The present and future of institutional-anomie theory. In: Cullen FT, Wright JP, Blevins KR (eds) Taking stock: the status of criminological theory, advances in criminological theory, vol 15. Transaction, New Brunswick

Messner SF, Rosenfeld R (2007) Crime and the American dream, 4th edn. Thompson, Belmont

Messner SF, Rosenfeld R (2009) Institutional-anomie theory: a macro-sociological explanation of crime. In: Krohn MD, Lizotte AJ, Hall GP (eds) Handbook on crime and deviance. Springer, New York

Messner SF, Raffalovich LE, Shrock P (2002) Reassessing the cross-national relationship between income inequality and homicide rates: implications of data quality control in the measurement of income distribution. J Quant Criminol 18:377–395

Messner SF, Thome H, Rosenfeld R (2008) Institutions, anomie, and violent crime: clarifying and elaborating institutional-anomie theory. Int J Conflict Viol 2:163–181

Muftic LR (2006) Advancing institutional anomie theory: a microlevel examination connecting culture, institutions, and deviance. Int J Offender Ther Comp Criminol 50:630–653

Piquero A, Piquero NL (1998) On testing institutional anomie theory with varying specifications. Stud Crime Crime Prev 7:61–84

Pratt TC, Godsey TW (2003) Social support, inequality, and homicide: a cross-national test of an integrated theoretical model. Criminology 41:611–644

Rosenfeld R, Messener SF (1995) Crime and the American dream: an institutional analysis. In: Adler F, Laufer WS (eds) The legacy of anomie theory. Transaction, New Brunswick

Rosenfeld R, Messner SF (1997) Markets, morality, and an institutional anomie theory of crime. In: Passas N, Agnew R (eds) The future of anomie theory. Northeastern University Press, Boston

Rosenfeld R, Messner SF (2003) Crime and the American dream. In: Cullen FT, Agnew R (eds) Criminological theory past to present: essential readings, 2nd edn. Los Angeles, Roxbury

Rosenfeld R, Messner SF (2006) The origins, nature, and prospects on institutional-anomie theory. In: Henry S, Lanier M (eds) The essential criminology reader. Westview, Boulder

Rosenfeld R, Messner SF (2010) The normal crime rate, the economy, and mass incarceration: an institutional-anomie perspective on crime-control policy. In: Decker S, Barlow H (eds) Criminology and public policy: putting theory to work. Temple University Press, Philadelphia

Rosenfeld R, Messner SF (2011) The intellectual origins of institutional-anomie theory. In: Cullen FT, Jonson CL, Myer AJ, Adler F (eds) The origins of American criminology, advances in criminological theory, vol 16. Transaction, New Brunswick

Savolainen J (2000) Inequality, welfare state, and homicide: further support for the institutional anomie theory. Criminology 38:1021–1042

Schoepfer A, Piquero NL (2006) Exploring white-collar crime and the American dream: a partial test of institutional anomie theory. J Crim Justice 34:227–235

Stucky TD (2003) Local politics and violent crime in U.S. cities. Criminology 41:1101–1135

Institutional Environment

► Organizational Change and Police Legitimacy

Institutional Imbalance

► Institutional Anomie Theory

Institutional Legitimacy

► Organizational Change and Police Legitimacy

Institutional Theories of Punishment

Joachim J. Savelsberg
Department of Sociology, University of
Minnesota, Minneapolis, MN, USA

Synonyms

Comparative Punishment; Criminal Punishment; Institutions (political, legal, knowledge); Rates of Punishment (spatial and temporal variation)

Overview

Long- and midterm changes in punitive attitudes and penal practices can only be understood by taking complex interactions between multiple social forces into account. Statistical analyses show that many factors are at work (Jacobs/Jackson 2010), while theoretically informed studies provide an in-depth understanding of the role played by specific conditions such as everyday life experiences (Garland 2001), including economic conditions and racial/ethnic conflict (Wacquant 2009), elite strategies (Beckett 1997; Simon 2007), and long-term cultural patterns (Melossi 2001; Savelsberg 2004; Whitman 2003). Recent literature, following earlier suggestions (Savelsberg 1994, 1999), has drawn special attention to the institutional arrangements under which ideas regarding crime and punishment are generated and penal decisions are made (Barker 2009; Campbell 2011; Garland 2010; Lacey 2008; Miller 2008; Page 2011; Savelsberg et al. 2004; Savelsberg and Flood 2011; Schoenfeld 2010; Sutton 2000; Whitman 2003). This entry begins with an illustration of what is to be explained: massive divergences in midterm trends in punitive attitudes and penal practices, illustrated for Europe and the United States. The main body of this text introduces competing (and mutually complementing) factors and approaches, each supported by distinct literatures, that seek to make sense of penal variation across time and space. While each of these approaches proves to be insufficient by itself, a synthesis is more promising, especially if it takes institutional conditions seriously.

What is to be Explained: Cross-National Patterns and Trends

Differences between crime concerns, penal attitudes, and intensities of punishment in Europe vis-à-vis the United States are substantial. The United States of the last third of the twentieth and the early twenty-first centuries is home to considerable fear of crime. Such fear and related cognitions inform not only adaptations in everyday behavior but also political attitudes and behavior. They are a precondition for the extraordinarily high rate of public support for capital punishment that has been hovering between 70 % and 80 % for most years since the late 1970s, for America's continuing use of the death penalty (the only Western country), albeit not in all of its states, and for rates of imprisonment that are unparalleled in American history and in any Western democracy.

A closer look shows that punitive attitudes and tough punishment practices are not a constant feature of the United States. Populations of other countries, for that matter, are not consistently mild-minded. A comparison of longer-term trends in support of capital punishment in Germany and the United States, for example, shows that both populations strongly supported the death penalty in the immediate post-World War II era. Support among Germans was even higher, despite (or because) of the most recent experience of the murderous Nazi regime. The 1960s witness a considerable weakening of support in both countries. Analyses for Germany show that this decline is primarily the result of cohort effects, only secondarily of period effects. It reaches new lows in the 1980s and 1990s. The United States, on the other hand, experience rising support of capital punishment, beginning in the mid 1970s, reaching new heights within a few years, and remaining at a very high level ever since.

Penal practice shows some parallels to, but also marked differences from, attitudes. Rates of executions in the United States, for example, are closely associated with trends in public opinion. They declined through the 1960s until a Supreme Court-mandated moratorium brought executions to a standstill from 1968 through 1976, but they increased steadily from there through the late 1990s to almost 100 executions per year in 1999, followed by a drop in most recent years. In European countries, by contrast, the death penalty is now abolished. In Germany, for example, it has been unconstitutional since the foundation of the Federal Republic in 1949, a time when public support for the death penalty was still high. Here, public opinion followed political decisions and constitutional norms, rather than vice versa.

The trend line for incarceration rates in the United States shows a similar development to that of executions. It decreased through the early 1970s and then began its dramatic increase that culminated in levels never seen before by the turn of the millennium (from just above 100 per 100,000 population to a rate of above 600). The new increase beginning in the 1970s is not just associated with the renewed support for the death penalty but also with increases of other indicators of punitiveness. For instance, a growing number of people supported the proposition that courts in their region are too lenient in their practices.

In Europe, on the other hand, long-term imprisonment rates were relatively stable and less in line with changes in public opinion. Consider again Germany. Even though the last decade of the twentieth century shows a marked 50 % increase in imprisonment rates, this trend built on a historical low of around 50 inmates per 100,000 population in 1992. Recent increases in punitive attitudes among Germans have been diagnosed, but they are relatively moderate (Reuband 2003).

How can we explain the syndrome of fear, punitive attitudes, and increasingly severe penal practice in the United States? What about the long-term trends toward leniency and recent moderate shifts in the opposite direction in Europe? Can we develop a theory that is congruent with these patterns and general enough to explain distinct trends in different countries alike? Several explanatory approaches have been offered in recent years, each with its own promises and limits.

Explanatory Controversies: Toward Multiple Factors

Crime as a Basis for Punitiveness?

Looking at the extraordinary rates of punitive attitudes and practices in the United States, some point at high American crime rates as the likely cause. Yet, American crime rates for most offenses are quite comparable with rates in other Anglo-Saxon countries and not dramatically higher than those in most European countries. Important exceptions are particular types of violent crime such as robbery and homicide. These are concentrated, however, in poor inner-city ghettos, where dealing with illegal drugs constitutes one remaining opportunity to earn a living after the cities began to deindustrialize in the 1950s. When conflicts develop in the world of illegal markets, the resolution of resulting grievances is not left to the state's monopoly of force, but is carried out through self-help, thus contributing to high rates of violence in these areas.

Punitive attitudes, on the other hand, are highest not among those who are most exposed to the risks of violent crime but rather among residents of America's suburbia. Also, the increase in punitive attitudes in the United States is far from synchronized with crime trends. Similarly in Germany, while crime rates increased for the first half of the post-World War II era, support for capital punishment declined during this period. Patterns and trends of crime and victimization risk thus neither offer a simple and direct explanation for punitive attitudes in the United States nor for international differences. In short, crime rates are not direct determinants of punishment trends, especially not in Europe. It is still conceivable though that they play an indirect role once translated into fear through different ways of cultural processing and political instrumentalization (Garland 2001). Such processing is thus to be considered below.

National Culture?

Others point to cultural particularities of different countries as a basis for their patterns of punishment. For the United States, it has been suggested that a profoundly religious culture with clear distinctions between "good" and "evil" may dramatize evil and thus produce fear and punitive attitudes. Research finds some supporting evidence. For example, judges who are members of fundamentalist congregations tend to sentence more harshly than other judges; also members of fundamentalist Protestant congregations are more likely to support imprisonment, corporal punishment, and capital punishment (e.g., Grasmick et al. 1993). Not just religion, also other aspects of culture may encourage punitive attitudes. Some suggest that American individualism advances an individual as opposed to collective attribution of guilt, thus encouraging punitive rather than social reform responses.

A closer look shows, however, that reference to enduring national cultures by itself does not suffice. Historically, the USA is the country in which the Quakers promoted the replacement of corporal punishment for the sake of a rehabilitation-oriented prison system already in the early nineteenth century. Also in the nineteenth century, Gustave de Beaumont and Alexis de Tocqueville traveled to America to learn about this system and its potential benefits for Europe. Further, the American imprisonment rate varied considerably in the course of the twentieth century, and Americans' support for capital punishment tended to fall below that of Germans for much of the 1950s and 1960s. A simple reference to American culture is thus equally misleading as the reference to American crime rates when we seek to explain fear and punitive attitudes and practices.

Reference to a long-term culture as a straightforward explanation of punitiveness elsewhere is just as simply cast into doubt. Consider again Germany as a convenient example. The post-World War II era of the Federal Republic experienced considerable variation in support of capital punishment. A look beyond this Republic's history reveals the most dramatic shifts in penal attitudes and practices. While we do not have good measures for public attitudes during the Nazi era or for the German Democratic Republic (the Communist part of the divided Germany), we do know that penal and police practice in the former were unspeakably harsher and in the latter considerably more severe than in the Weimar Republic or in the Federal Republic. Again, enduring national culture surely is unable to provide a straightforward explanation of penal practices and probably not for penal attitudes either, neither for Germany nor for the United States or any other country.

And yet, the role of culture should not be discarded altogether. Instead, the concept of culture as "tool kit" provides help, a notion developed by sociologist Ann Swidler. A tool kit is a collection of tools from which actors can draw in a situation of need. It is then decisive what tools are available in a given national context. Whitman (2003) suggested that aristocratic sensibilities played a part in Europe's historically less shaming and ferocious punishment regimes. Others highlight dominant religious traditions in different countries that provide distinct penal tools (Savelsberg 2004). More broadly, Philip Smith (2008), in his Punishment and Culture, develops a neo-Durkheimian argument that alters us to notions of the sacred, of rituals, and of myths that feed penal practices. The question then arises: What types of myths, sensitivities, religious doctrines, and other tools are available in different societies?

Culture as a tool kit thus means that certain tools are available and can be drawn on depending on a given situation. Cultural tools are necessary but not sufficient conditions of the use of penal strategies. Further, tools also need to be legitimate. It is noticeable that the use of force by government appears to have been delegitimized in countries such as Italy and Germany after the brutal misuse by the Fascist and Nazi regimes (albeit with a time lag) but that it remained intact in liberator and victor countries (Melossi 2001).

In short, long-term cultural traditions affect punitive attitudes and practices in complex interaction with the shorter-term legitimization of the state's use of force. They further interact with actions by concrete actors, such as elite campaigns and support by the populace that is associated with concrete living conditions.

Concrete Actors 1: Conservative Elites

The social science literature has shown that a focus on concrete actors is highly promising. Katherine Beckett's (1997) work, for example, successfully links increases of punitive attitudes and practices to activities of members of conservative elites, beginning in the 1960s. Her historical account quotes from political campaigns in which conservative Southerners equated support for the civil rights movement (as expressed in tactics of passive resistance and limited law violations) with the promotion of lawlessness generally to thus explain the dramatic increase in crime rates during this period. Charles Whittaker, a former US Supreme Court Justice, for example, argued that the 1960s crime wave was "fostered and inflamed by the preachments of self-appointed leaders of minority groups . . . [who told their followers] . . . to obey the good laws and to violate the bad ones. . . This simply advocates the violation of the laws they do not like. . . and the taking of the law into their own hands" (as quoted in Beckett 1997:31). More recent work has thus rightfully highlighted racial dimensions of the punitive turn in the United States (Wacquant 2009), especially in the context of the "war on drugs" (Tonry 1995).

Conservatives applied the same logic when they fought President Johnson's expansion of the welfare state through his Great Society programs. A quote from a 1964 speech by the Republican presidential candidate Barry Goldwater may serve as an illustration: "If it is entirely proper for the government to take away from some to give to others, then won't some be led to believe that they can rightfully take from anyone who has more than they? No wonder law and order has broken down, mob violence has engulfed great American cities, and our wives feel unsafe in the streets" (as quoted in Beckett 1997:35).

Insights gained from such historical narratives are supported by evidence from systematic data collection and hypothesis testing, including multivariate tests of the relationship between political initiatives, media coverage, and public perception. Examining when respondents to public opinion surveys consider crime or drug use the most important problem the nation faces, Beckett's (1997):14–27 work yields four insights: First, the reporting of drug and crime issues by the media and their ranking in public opinion surveys are highly uneven across the years, even within 2–4 year spans. Second, neither media reporting nor public opinion is clearly associated with the actual seriousness of the problem as measured by the crime rate and the incidence of drug use. Third, the fluctuations of media reporting are closely associated with political campaigns that highlight crime or drug issues. Fourth, public opinion tends to be motivated by media reporting. In short, the causal path leads from elite strategies through media reporting to public opinion. Americans tend to think of crime or drug use as the most important problems when politicians raise these topics in their campaigns and when media report about the issues raised by politicians. Yet, an important factor is missing from this account which is now to be addressed.

Concrete Actors 2: Everyday People and Their Fears

New critics within the constructivist tradition of social problems research, with which authors such as Beckett are affiliated, have noted that the public is not always receptive to the campaigns of political elites. Such campaigns may thus be necessary, or at least important, conditions of social problem careers, but not sufficient ones. Elite campaigns are effective only if they touch a sensitive nerve of the public. This sensitivity is a central theme in David Garland's (2001) book *The Culture of Control*. Garland argues here that people, under conditions of late modernity, are exposed to extraordinary challenges that produce severe anxieties. Old certainties about stable jobs, streamlined work careers, and social security have been lost. Arrangements of everyday life become more complicated as people are confronted with longer commutes; more single parent or dual income situations; longer periods of study, training, and retraining with less secure career expectations; and a weakening of the welfare state's safety net. In addition, and in line with routine activity arguments, public space is ever less secured by informal social control, which helps explain actual increases in crime rates in most Western societies in recent decades, and has thus contributed to further insecurities among the public. In short, Garland argues that we can understand the public's receptivity for punitive elite campaigns only against this background of general insecurity under conditions of late modernity.

Such arguments are supported by empirical investigation. Social psychologist Tom Tyler and collaborators examine public support for the "three strikes" law in California (Tyler and Boeckman 1997). This well-known law, passed through a public referendum, demands that offenders convicted of a third felony offense be given a life prison term with no possibility of parole. Tyler's research indicates that California's "three strikes" law was most strongly supported not by crime victims but by those who feel that the basic foundation of their social life has become undermined as the stability of their families is threatened or as immigrants move into their neighborhoods.

Conditions of everyday life and resulting public sensitivities must thus be taken seriously when we seek to explain punitive attitudes generally and people's receptivity to punitive campaigns by conservative elites specifically. Yet, this argument too faces an important limitation. The United States is not the only country whose residents experience conditions of late modernity. Consider the small and seemingly peaceful alpine country of Austria where, according to new survey results, the majority of young people no longer believe that their retirement benefits are secured. In some of its states, almost half of all children are born out of wedlock. Or consider

Germany where family life has abandoned its traditional and standardized patterns. Statistical trends show an almost double divorce rate by the end of the twentieth century compared to two decades earlier. In the same time span, the number of non-married cohabitating couples quadrupled, accompanied by a massive increase in the rate of out of wedlock birth. Finally, the participation of women in institutions of higher learning and in occupational life has increased dramatically in most European countries, albeit not to the same degree as in the United States.

In short, life has become less certain and life courses less predictable in the United States, in other Anglo-Saxon countries, and in continental European countries such as Austria or Germany. Late modernity has arrived in all of these places. Yet, most countries do not show the same intense increases in punitive attitudes and practices as do the United States (for common law countries, see Sutton 2000).

Institutional Particularities 1: The Weight of Knowledge Markets in the United States

The puzzle of country-specific responses to the challenges of late modernity can only be solved if we consider nation-specific institutions that process the insecurities people experience, that produce knowledge and ideas on how to respond to these insecurities, that promote or inhibit attitudes, and that channel those processes in the political and judicial worlds where punitive decisions are made (Savelsberg 1994). Such institutions differ considerably even across Western style democracies. Consider mass media as one institution through which information and knowledge are generated and diffused. Mass media in the United States are largely steered through market mechanisms, while most European countries, at least until a few years ago, were served by parallel public radio and television stations. Mass communication research has shown that market-driven media tend more toward dramatized and sensationalized presentation of hot news items, including crime, thereby enhancing fear of crime and cultivating punitive attitudes.

Further, public opinion surveys have been institutionalized in the United States for many decades. Media regularly report survey results back into the public, thus contributing to a positive feedback slope that further intensifies public sentiments. The regular publication of public opinion survey results in mass media has become more common in European countries as well; yet, it was the exception for much of the post-WW II era.

Institutional arrangements even in the world of scholarship as it contributes to knowledge about crime and punishment open up this field to market pressures much more in the United States than in European countries. The resource distribution to universities and academic departments depends on how well they do in the competition for student enrollment and research grants. Signs of competitiveness are annual rankings of department members according to their achievement in research, teaching, and service and the rankings of academic programs and universities alike that are widely known across the country. It is not surprising then that the past decades of growing public concern about crime have also witnessed a rapid growth of specialized programs of criminology and criminal justice studies. This specialization has been partly driven by funding from Justice Departments and by the creation of demand for higher education among law enforcement personnel.

This institutional adaptation of academia to concerns of its environment has also resulted in the incorporation of punitive ideas. Content analyses of large sets of scholarly journal articles on issues of crime and crime control in leading sociology and criminology journals, combined with in-depth interviews with scholars, show that research in specialized programs is more closely oriented toward themes and theoretical perspectives that dominate in the increasingly conservative political sector. This effect is further enforced where research is funded by political institutions (e.g., Savelsberg et al. 2004; Savelsberg and Flood 2011). These findings may partly be due to the facts that funding has shifted away from (1) departments dealing with health, housing, or human services to the Justice Department and (2) from researcher initiated to strategic funding programs. While comparable studies do not exist for other countries, it is likely that less competition-oriented European academic institutions will not adapt as readily to their political environment as their American counterparts do.

Institutional Particularities 2: The Weight of Political Markets in the United States

Not only institutions of knowledge production but also institutions of political and legal decision-making play crucial roles in the unfolding of penal practice. In a recent series of publications, Nicola Lacey (e.g., Lacey 2008) directs our attention toward the weight of political decentralization, the dynamics of different electoral systems and cycles. Her work leads her to challenge the notion that globalization leads to a simple convergence of criminal justice policy across nation states. Instead, she stresses that several nation-level institutional and cultural particularities filter global processes, dislodging them before they prove effective at the national level.

Applying these insights to a European-US comparison informs us that political institutions too are more market driven in the United States (Savelsberg 1994). One decisive factor is the relatively weak position of political parties that are not guaranteed a place in the American constitution, in contrast to some European countries. Candidates for elected office, for example, who are selected by the public in primary elections in the USA, are chosen through party committees and assemblies in Europe. American political candidates who seek to advance their political careers thus have to orient their positions at least as much toward opinions and sentiments of the electorate as toward enduring positions of their party. Not only legislators, but also heads of the executive branch, president, and governors alike, are elected through direct popular elections, not through a vote within the legislature as is the case in parliamentary democracies. Again, public opinion is more likely to translate into political decision-making in the United States. Further, high-ranking civil servants in the USA are much more frequently political appointees, selected by the president. They will return to their prior positions in the business world, academia, foundations, or law firms once the administration changes. Here too public sensitivities make their way into public administration more easily than in most of Europe. Finally, in the judicial branch of government, the public is represented in court through the American jury system, and most prosecutors and judges are also selected through popular elections. They depend on reelection if they wish to keep their position for a subsequent term. It is likely that such judges and prosecutors will orient their decisions more toward concerns of the public than their life tenured civil servant colleagues in most continental European countries.

It is decisive that all of these institutional spheres are more open to adapting to public opinion in the United States than in Europe. As a result, anxieties and punitive public attitudes are more likely to be reflected in the production and diffusion of knowledge and in political and legal decision-making. And they become communicated back to the public through a positive feedback slope to thus reinforce the initial expression of public opinion. Sutton's sophisticated and theoretically guided statistical analyses provide empirical support for this type of institutional argument in his comparative study of common law countries (Sutton 2000). Such institutional arguments are not meant to answer the "chicken or egg" question as to what comes first, public opinion about crime and punishment or punitive programs. It is decisive instead that those relatively "open" institutions of knowledge production and decision-making characteristic for the United States tend to lead to a mutual reinforcement of both poles, while Europe's relatively "closed" institutions are more likely to support stability.

Two modifications to this typological cross-national argument are needed. First, the American War on Crime has contributed to a transformation of America's democratic institutions, as Jonathan Simon (2007) has aptly documented. Prominent among them is a shift toward the use of executive power. Further, Joshua Page (2011) has powerfully demonstrated for the case of California that the imprisonment boom has strengthened an important group in the penal field, the corrections officers union. In turn, this union constituted an effective lobbying group on behalf of the passage of penal laws.

Second, the above argument, focusing on nation-level characteristics, has been challenged in recent years as states below the nation-level show remarkably different trends in their penal practice. David Garland (2010), in his recent book on capital punishment, highlights the fact that the United States, while retaining the death penalty longer than other Western democracies, is constituted by states, some of which abolished this penalty when no European country even considered abolition. It is exactly the decentralization of political power and localized decision-making that account for this country simultaneously being an early abolitionist and a late retainer of the death penalty.

Rightfully then, others have taken the next step by studying punishment at the state level. A comparison of penal practice in the states of California, New York, and Washington shows, for example, that state-level differences in political institutions and participatory practices help explain the distinct penal patterns in these three states (Barker 2009). Specifically, Barker argues that "polarized populism" in California contributed to a retributive penal regime. This pattern starkly contrasted with Washington State's "deliberative democracy" that fostered de-escalation and New York's "elite pragmatism" that advanced a managerial practice of punishment. Others have explored institutional particularities of individual states to explain their specific patterns of punishment (e.g., Lynch 2010 on Arizona; Schoenfeld 2010 on Florida). Campbell (2011) finds for Texas that a period of unsettled and contested change generated a highly punitive penal code and created an additional layer of prison facilities to manage lower-level offenders. His analysis suggests that the patterns he identifies reflect institutionalized conflict between state and local governments. Campbell further alerts the reader to interactions between such institutional arrangements and contextual factors that are deeply embedded in the history of the state. Finally, examining differences in the institutionalization of civil society between the local and the national levels, Miller (2008) shows that anti-punitive interest groups are generated at the local level, where residents are in fact most affected by crime but where political pressure has little influence on penal policy making. State- and nation-level organization of interests, on the other hand, representing those least affected by crime, tends to be punitive and effective.

Clearly thus, institutional arrangements, characteristic of particular levels of government, and even of particular states must be taken seriously. This does not mean, of course, that higher-level factors become obsolete. The next step in this institutional line of scholarship should thus aim at theorizing – and examining empirically – the interaction between global-, national-, state-, and local-level institutions as they jointly filter social structural and cultural forces and thus shape penal practices.

In short, a growing body of literature has shown that institutional arrangements are a crucial factor in the equation that explains practices of punishment. Open institutions, at least in some contexts, tend to communicate public concerns and punitive demands into legal and political decision-making. Yet, the question remains unanswered why people under conditions of insecurity, provoked by the strains of late modernity, resort to punitive demands. It is not at all certain, after all, that punitive strategies will actually create security and improve living conditions. Much to the opposite, evidence indicates considerable destruction of human, social, and cultural capital as a consequence of massive imprisonment. Alternative responses to insecurity such as increased unionization or demand for new welfare protections may be much more effective in fighting insecurity. Here, institutional arguments need to link back to the notions of tool boxes and interested actors discussed above. Open institutions are especially receptive to punitive turns where popular pressure suggests the use of punitive tools and where political leaders and members of the judiciary have an interest in using these tools.

Institutional Change and Transnational Convergence?

Given the impressive differences between Europe and the United States in terms of cultural repertoires, institutional arrangements, and social structure, should we still see a convergence as many globalization theorists would predict? There are some reasons to follow their argument and even stronger challenges that caution us.

First, institutions in Europe have been undergoing considerable change. The safety net of the welfare state is weakening and some functions are being passed on to the market, even if changes in Europe are much slower and meet more resistance than they do in the United States. Also, the weight of public radio and television stations has been declining over the past two decades as private channels, operating under the pressure of markets, have become more prominent. Simultaneously, the number of programs that address crime, especially violent crime, in movies and reports has multiplied. Consumers of such programs tend to overestimate crime as a problem, perceiving serious increases even for types of crime that have been on the decline as documented by German criminologist Christian Pfeiffer. Finally, leading politicians are increasingly exposed to public pressure due to a growing media presence in combination with increased public opinion polling. They are thus likely to be more oriented toward public opinion.

Finally, institutional shifts in Europe are supplemented by one change of a different sort. New cohorts of judicial decision-makers are socialized under conditions of late modernity and new institutional arrangements. Future cohort studies may measure effects of this turnover in judicial personnel on changing penal practices in Europe.

Despite such changes, however, many institutions in Europe maintain their distinct character and continue to differ markedly from those in the United States. The role of public media is still more prominent in Europe. Platform parties remain strong. Candidates for political office continue to be selected by parties, rather than by the public in primary elections. Heads of the executive branch continue to be elected by the members of parliaments rather than in direct popular votes in most countries. Prosecutors and judges still do not get elected by the general public; the proposal of such selection mechanism would in fact continue to appear strange to most Europeans. Due to these institutional specifics, I expect that increases in punitiveness in Europe will not reach American dimensions. Empirical opinion data seem to support this contention as recent studies by German sociologist Karl-Heinz Reuband (2003) have shown. The percent of residents of the (old) West German states who demand prison terms in response to burglary convictions, for example, increased from 13 % to 19 % between 1989 and 2002. Fifty-nine percent demand community service (down 1 %), 7 % suggest fines (down 2), 10 % probationary terms (down 2), and 5 % “other” penalties (up 3) (others “don’t know”). The increase in demand for prison terms is surprisingly modest in light of the advent of late modernity, a high rate of unemployment for much of the 1980s and 1990s, the burden of the 1990 German unification with its cultural clashes and dramatic demographic and economic transformations, and the institutional changes sketched above. Each of these factors by itself might, according to substantial bodies of social science literature, result in massive increases in punitiveness, was it not for the continuing features of Europe’s institutional outfit.

In line with these observations about public opinion, the 50 % increase in German imprisonment since 1992, beginning from a historic low, and similar prison growth in other European countries, has been significant but modest compared to the American experience of recent decades, where the 1970s and the 1980s each experienced almost 100 % increases in imprisonment rates; the 1990s, despite steep declines in serious crime, saw another increase of around 50 %. Again, distinct cultural repertoires and institutions and important institutional differences in the legislative, executive, and judicial branches of government are likely to account for this contrast between the USA on the one hand and Germany (and most European countries) on the other.

Concluding Thoughts: Toward Synthesis

Punitive attitudes and practices, especially pronounced in the United States, have been increasing in Europe as well. Changes in crime are only a very limited and indirect explanation for this trend. Crime needs to be culturally processed before it can cause anxiety and punitive attitudes. Conservative elites may play a major role in this process as the American case indicates. Their effect, however, is partly contingent on the growing insecurity of many segments of the population under conditions of late modernity. Both effects of elite strategies and of conditions of late modernity on punitive attitudes and practices have been shown by researchers in the United States.

Yet, punitive responses to insecurity are not a natural outcome. They are more likely where respective cultural repertoires are available and legitimized, which (during the post-World War II period) seems to apply more to the United States than to Europe. One additional factor, too frequently overlooked in earlier scholarship, is the profoundly different institutional arrangements in the United States vis-à-vis most continental European countries. This applies to institutions for the production and dissemination of knowledge as much as for institutions of political and legal decision-making. The more open political and legal institutions are for public sentiments, the more will public concerns affect political and legal decision-making. Anxieties will often not be calmed by punitive decisions, but instead may even be reinforced as the American case seems to indicate. This does not mean, of course, that democracy accounts for punitiveness but that some forms of democratic institutions may have that effect. Autocratic regimes, in fact, are characterized by particularly severe and arbitrary practices. Fortunately, the earlier neglect of institutional forces has been corrected by a wave of recent research at the national, state, and local levels. We now face the challenge of understanding the interaction between global forces and national, state, and local institutional conditions.

In sum, an appropriate theory of punitive attitudes and practices must be conceived as a multifactorial, historically, and institutionally grounded theory, as Max Weber would have suggested. Such a theory must take seriously crime, changing living conditions, elite strategies, cultural repertoires and experiences with the state's use of force, and finally institutional patterns through which knowledge is produced and diffused and through which decisions on punishment are made.

Recommended Reading and References

Barker V (2009) The politics of imprisonment: how the democratic process shapes the way America punishes offenders. Oxford University Press, Oxford

Beckett K (1997) Making crime pay: law and order in contemporary American politics. Oxford University Press, New York

Campbell MC (2011) Ornery alligators and soap on a rope: Texas prosecutors and punishment reform in the Lone Star State, Theor Criminol 15:289–311

Garland D (2001) The culture of control: crime and social order in contemporary society. The University of Chicago Press, Chicago

Garland D (2010) Peculiar institution: the death penalty in the United States. Harvard University Press, Cambridge, MA

Grasmick H, Cochran JK, Bursick RJ Jr, Kimpel ML (1993) Religion, punitive justice, and support for the death penalty. Justice Q 10:289–314

Jacobs D, Jackson AL (2010) On the politics of imprisonment: a review of systematic findings. Annu Rev Law Soc Sci 6:129–149

Lacey N (2008) The prisoners' dilemma: political economy and punishment in contemporary democracies. Cambridge University Press, Cambridge

Lynch M (2010) Sunbelt justice: Arizona and the transformation of American punishment. Stanford Law Books, Stanford

Melossi D (2001) The cultural embeddedness of social control: reflections on the comparison of Italian and North American cultures concerning punishment. Theor Criminol 5:403–424

Miller LL (2008) The perils of federalism: rave poverty and the politics of crime control. Oxford University Press, Oxford

Page J (2011) The toughest beat: politics, punishment, and the prison officers union in California. Oxford University Press, Oxford

Reuband KH (2003) Steigende Repressionsneigung im Zeitalter der 'Postmoderne?' Das Sanktionsverlangen der Bundesbueger 1989 und 2002 im Vergleich. N Kriminalpolit 15:100–115

Savelsberg JJ (1994) Knowledge, domination, and criminal punishment. Am J Sociol 99:911–943

Savelsberg JJ (1999) Knowledge, domination, and criminal punishment revisited: Incorporating state socialism. Punishment and Society 1:45–70

Savelsberg JJ (2004) Religion, historical contingencies, and criminal law: the German case and beyond. Law Soc Inq 29:373–401

Savelsberg JJ, Cleveland LL, King RD (2004) Institutional environments and scholarly work: American criminology: 1951–1993. Social Forces 82:1275–1302

Savelsberg JJ, Flood SM (2011) Collins meets criminology: global theory of intellectual change and a policy-oriented field. Sociol Forum 26:21–44

Schoenfeld H (2010) Mass incarceration and the paradox of prison conditions litigation. Law Soc Rev 44:731–768

Simon J (2007) Governing through crime: how the war on crime transformed American democracy and created a culture of fear. Oxford University Press, Oxford

Smith P (2008) Punishment and culture. University of Chicago Press, Chicago

Sutton JR (2000) Imprisonment and social classification in five common-law democracies, 1955–1985. Am J Sociol 106:350–385

Tonry MH (1995) Malign neglect. Oxford University Press, New York

Tyler TR, Boeckmann R (1997) Three strikes and you are out, but why?: The psychology of public support for punishing rule breakers. Law Soc Rev 31:237–265

Wacquant L (2009) Punishing the poor: the neoliberal government of social insecurity. Duke University Press, Durham

Whitman JQ (2003) Harsh justice: criminal punishment and the widening divide between America and Europe. Oxford University Press, Oxford

Institutional Theory

► Organizational Change and Police Legitimacy

Institutions (Political, Legal, Knowledge)

► Institutional Theories of Punishment

Integrated Cognitive Antisocial Potential Theory

David P. Farrington
Institute of Criminology, University of Cambridge, Cambridge, UK

Synonyms

Antisocial potential; Developmental criminology; ICAP theory; Life events; Life-course criminology; Risk and protective factors

Overview

Developmental and life-course criminology (DLC) focuses on the development of offending and antisocial behavior, risk and protective factors, and the effects of life events. The ICAP theory specifies the risk and protective factors that influence the development of long-term antisocial potential (varying between individuals) and the situational factors that influence short-term antisocial potential (varying within individuals over time). It also specifies the cognitive processes that influence whether the antisocial potential becomes the actuality of offending in any situation and the consequences of offending that have feedback influences on antisocial potential. Tests of the ICAP theory and its research and policy implications are described.

Developmental and Life-Course Theories

Developmental and life-course criminology (DLC) is concerned mainly with three topics: (a) the development of offending and antisocial behavior from the womb to the tomb; (b) the influence of risk and protective factors at different ages; and (c) the effects of life events on the course of development (Farrington 2003). Whereas traditional criminological theories aimed to explain between-individual differences in offending, such as why lower-class males commit more offenses than upper-class males, DLC theories aim to explain within-individual changes in offending over time, such as why males commit more offenses when they are unemployed than when they are employed.

DLC theories aim to explain offending by individuals (e.g., as opposed to crime rates of areas). "Offending" refers to the most common crimes of theft, burglary, robbery, violence, vandalism, minor fraud, and drug use and to behavior that in principle might lead to a conviction in Western industrialized societies such as the United States and the United Kingdom. These theories aim to explain results on offending obtained with both official records and self-reports. Generally, DLC findings and theories particularly apply to offending by lower-class urban males in developed countries in the last 80 years or so. To what extent they apply to other types of persons (e.g., middle-class rural females) or offenses (e.g., white-collar crimes or sex offenses against children) are important empirical questions that cannot be addressed here.

In conducting research on development, risk and protective factors, life events, and DLC theories, it is essential to carry out prospective longitudinal surveys. I have directed the Cambridge Study in Delinquent Development, which is a prospective longitudinal survey of over 400 London males from age 8 to age 56 (Farrington et al. 2013). The main reason why developmental and life-course criminology became important during the 1990s was because of the enormous volume and significance of longitudinal research on offending that was published during this decade. Particularly influential were the three "Causes and Correlates" studies originally mounted by the US Office of Juvenile Justice and Delinquency Prevention in Denver, Pittsburgh, and Rochester. Other important longitudinal projects that came to prominence in the 1990s were the Seattle Social Development Project, the Dunedin study in New Zealand, the Montreal Longitudinal-Experimental study, and the further analyses by Laub and Sampson of the classic study by the Gluecks. (For reviews of all these studies, see Farrington 2013.)

What Do We Know?

There are ten widely accepted conclusions about the development of offending that any DLC theory must be able to explain. (For references, see Farrington and Loeber 2013.) First, the prevalence of offending peaks in the late teenage years – between ages 15 and 19. Second, the peak age of onset of offending is between 8 and 14, and the peak age of desistance from offending is between 20 and 29. Third, an early age of onset predicts a relatively long criminal career duration and the commission of relatively many offenses.

Fourth, there is marked continuity in offending and antisocial behavior from childhood to the teenage years and to adulthood. What this means is that there is relative stability of the ordering of people on some measure of antisocial behavior over time and that people who commit relatively many offenses during one age range have a high probability of also committing relatively many offenses during another age range. However, neither of these statements is incompatible with the assertion that the prevalence of offending varies with age or that many antisocial children become conforming adults. Between-individual stability in antisocial ordering is perfectly compatible with within-individual change in behavior over time. For example, people may graduate from cruelty to animals at age 6 to shoplifting at age 10, burglary at age 15, robbery at age 20, and eventually to spouse assault and child abuse later in life. There is also continuity in offending from one generation to the next.

Fifth, a small fraction of the population (the "chronic" offenders) commit a large fraction of all crimes. In general, these chronic offenders have an early onset, a high individual offending frequency, and a long criminal career. Sixth, offending is versatile rather than specialized (although specialization increases with age). For example, violent offenders are indistinguishable from frequent offenders in childhood, adolescent, and adult risk factors. Seventh, the types of acts defined as offenses are elements of a larger syndrome of antisocial behavior, including heavy drinking, reckless driving, sexual promiscuity, bullying, and truancy. Offenders tend to be versatile not only in committing several types of crimes but also in committing several types of antisocial behavior.

Eighth, most offenses up to the late teenage years are committed with others, whereas most offenses from age 20 onward are committed alone. This aggregate change is not caused by dropping out processes or group offenders desisting earlier than lone offenders. Instead, there is change within individuals; people change from group offending to lone offending as they get older. Ninth, the reasons given for offending up to the late teenage years are quite variable, including utilitarian ones (e.g., to obtain material goods or for revenge), for excitement or enjoyment (or to relieve boredom), or because people get angry (in the case of violent crimes). In contrast, from age 20 onward, utilitarian motives become increasingly dominant. Tenth, different types of offenses tend to be first committed at distinctively different ages. For example, shoplifting is typically committed before burglary, which in turn is typically committed before robbery. In general, there is increasing diversification of offending up to age 20; as each new type of crime is added, previously committed crimes continue to be committed. Conversely, after age 20, diversification decreases and specialization increases.

The main risk factors for the early onset of offending before age 20 are well known (Farrington and Welsh 2007):

- *Individual factors*: low intelligence, low school achievement, hyperactivity-impulsiveness and risk taking, and antisocial child behavior including aggression and bullying
- *Family factors*: poor parental supervision, harsh discipline and child physical abuse, inconsistent discipline, a cold parental attitude and child neglect, low involvement of parents with children, parental conflict, broken families, criminal parents, and delinquent siblings
- *Socioeconomic factors*: low family income and large family size
- *Peer factors*: delinquent peers, peer rejection, and low popularity
- *School factors*: a high delinquency rate school
- *Neighborhood factors*: a high crime neighborhood

By definition, a risk factor predicts a high probability of offending, whereas a protective factor predicts a low probability of offending in a group who are at risk. Less is known about protective factors. (For reviews of risk and protective factors, see Farrington et al. 2012.)

The main life events that encourage desistance after age 20 are getting married, getting a satisfying job, moving to a better area, and joining the military. The distinction between risk factors and life events is not clear-cut, since some life events may be continuing experiences whose duration is important (e.g., a marriage or a job), while some risk factors may occur at a particular time (e.g., loss of a parent). Other life events (e.g., converting to religion) may be important but have been studied less.

While the focus in DLC is on the development of offenders, it is important not to lose sight of factors that influence the commission of offenses. It is plausible to assume that offenses arise out of an interaction between the person (with a certain degree of criminal potential) and the environment (providing opportunities and victims). Existing evidence suggests that people faced with criminal opportunities take account of the subjectively perceived benefits and costs of offending (compared with other possible activities) in deciding whether or not to offend. DLC theories should explain the commission of offenses as well as the development of offenders.

Contentious Issues

I now turn to some more contentious issues. (For reviews of knowledge about these more or less contentious issues, see Piquero et al. 2003.) First, while it is clear that the prevalence of offending peaks in the late teenage years, it is much less clear how the individual offending frequency (i.e., the frequency of offending by those who offend) varies with age. Some studies suggest that the individual offending frequency accelerates to a peak in the late teenage years and then decelerates in the twenties, whereas others suggest that the individual offending frequency does not change with age. Second, it is not clear whether the seriousness of offending escalates up to a certain age and then de-escalates or whether it does not change with age.

Third, while it is clear that an early age of onset of offending predicts a long criminal career duration and many offenses, it is much less clear whether an early age of onset predicts a high individual offending frequency or a high average seriousness of offending. Nor is it clear whether early onset offenders differ in degree or in kind from later onset offenders or how much there are distinctly different behavioral trajectories. Fourth, while chronic offenders commit more offenses than others, it is not clear whether their offenses are more serious on average. Nor is it clear whether chronic offenders differ in degree or in kind from nonchronic offenders.

Fifth, it is clear that certain types of offenses occur on average before other types and hence that onset sequences can be identified. However, it is not clear whether these onset sequences are merely age-appropriate behavioral manifestations of some underlying theoretical construct (e.g., antisocial potential) or whether the onset of one type of behavior facilitates or acts as a stepping-stone toward the onset of another. In other words, onset sequences could reflect persistent heterogeneity or state dependence (Nagin and Farrington 1992). Similarly, little is known about onset sequences in which childhood antisocial behavior has some kind of influence on later offending, which might suggest opportunities for early prevention.

Sixth, while the main risk factors for the early onset of offending are well known, the extent to which these risk factors have causal effects on offending is not clear. Seventh, many risk factors could either be causes of offending or indicators of the same underlying construct, or even both. For example, heavy drinking could reflect the same underlying construct as offending (e.g., antisocial potential) in comparisons between individuals but could be a cause of offending in comparisons within individuals (e.g., if people committed more offenses while drinking than while not drinking). In other words, heavy drinking could be a factor that influenced short-term within-individual variations in offending: why people commit offenses in some times and places but not in others.

There are many other unresolved issues concerning risk factors for offending. While a great deal is known about family risk factors (especially) and individual risk factors, much less is known about biological or neighborhood risk factors. Little is known about risk factors for continuation of offending after onset, for later onsets after age 20, or for persistence or desistance of offending after age 20. Little is known about risk factors for the duration of criminal careers. Little is known about the causal processes that intervene between risk factors and offending. And little is known about protective factors, whether defined as factors that are opposite to risk factors (e.g., high school achievement compared with low school achievement) or as factors that interact with and counteract the effects of risk factors (Farrington and Ttofi 2012). While the main life events that encourage desistance are well known, much less is known about life events that influence onset or continuation after onset. Also, the effect of the criminal justice system (police, courts, prison, and probation) on desistance is highly controversial. The labeling effects of convictions, in increasing the probability, frequency, variety, or seriousness of subsequent offending, are also controversial. Few DLC theories include specific postulates about the effects of interventions.

DLC Theories

Several DLC theories have been proposed to explain findings in developmental and life-course criminology, of which two of the most famous were proposed by Moffitt and by Sampson and Laub (see Farrington 2005a). My “ICAP” theory is described here.

For many years, I did not attempt to formulate a wide-ranging DLC theory of offending. In line with Bernard and Snipes (1996), who advocated a risk factor approach, I focused on identifying independently predictive risk factors, testing specific hypotheses (e.g., about the effects of unemployment on offending), and on investigating possible causal mechanisms linking risk factors and offending (e.g., mediators between criminal parents and delinquent sons). However, I was criticized for being "atheoretical" (and sometimes described as a "dust-bowl empiricist," whatever that means) for focusing on empirical variables rather than underlying theoretical constructs. Therefore, encouraged by Joan McCord, I proposed a tentative theory in 1992 (Farrington 1992).

My theory tries to overcome some of the perceived deficiencies of classic criminological theories. First, they focused primarily on offending during the teenage years, on factors that correlated with offending, and did not focus on development through life. They emphasized immediate influences on offending, as opposed to longer-term influences from infancy and early childhood. Second, they neglected the continuity between offending and antisocial behavior, and the continuity in antisocial behavior from childhood to adulthood. Third, they focused on a limited number of influences on offending (especially those connected with socioeconomic status) and did not include biological or psychological factors. Fourth, they focused on differences between individuals rather than on changes over time within individuals.

The ICAP Theory

My "Integrated Cognitive Antisocial Potential" (ICAP) theory was primarily designed to explain offending by lower-class males, and it was influenced by results obtained in the Cambridge Study (Farrington 2005b; Zara 2010). It integrates ideas from many other theories, including strain, control, learning, labeling, and rational choice approaches; its key construct is antisocial potential (AP); and it assumes that the translation from antisocial potential to antisocial behavior depends on cognitive (thinking and decision-making) processes that take account of opportunities and victims. Figure 1 is deliberately simplified in order to show the key elements of the ICAP theory on one page; for example, it does not show how the processes operate differently for onset compared with desistance or at different ages.

The key construct underlying offending is antisocial potential (AP), which refers to the potential to commit antisocial acts. Long-term persisting between-individual differences in AP are distinguished from short-term within-individual variations in AP. Long-term AP depends on impulsiveness; on strain, modeling, and socialization processes; and on life events, while short-term variations in AP depend on motivating and situational factors.

Regarding long-term AP, people can be ordered on a continuum from low to high. The distribution of AP in the population at any age is highly skewed; relatively few people have relatively high levels of AP. People with high AP are more likely to commit many different types of antisocial acts including different types of offenses. Therefore, offending and antisocial behavior are versatile rather than specialized. The relative ordering of people on AP (long-term between-individual variation) tends to be consistent over time, but absolute levels of AP vary with age, peaking in the teenage years, because of changes within individuals in the factors that influence long-term AP (e.g., from childhood to adolescence, the increasing importance of peers and decreasing importance of parents).

A key issue is whether the model should be the same for all types of crimes or whether different models are needed for different types of crimes. Because of their focus on the development of offenders, DLC researchers have concluded that, because offenders are versatile rather than specialized, it is not necessary to have different models for different types of crimes. For example, it is believed that the risk factors for violence are essentially the same as those for property crime or substance abuse. However,

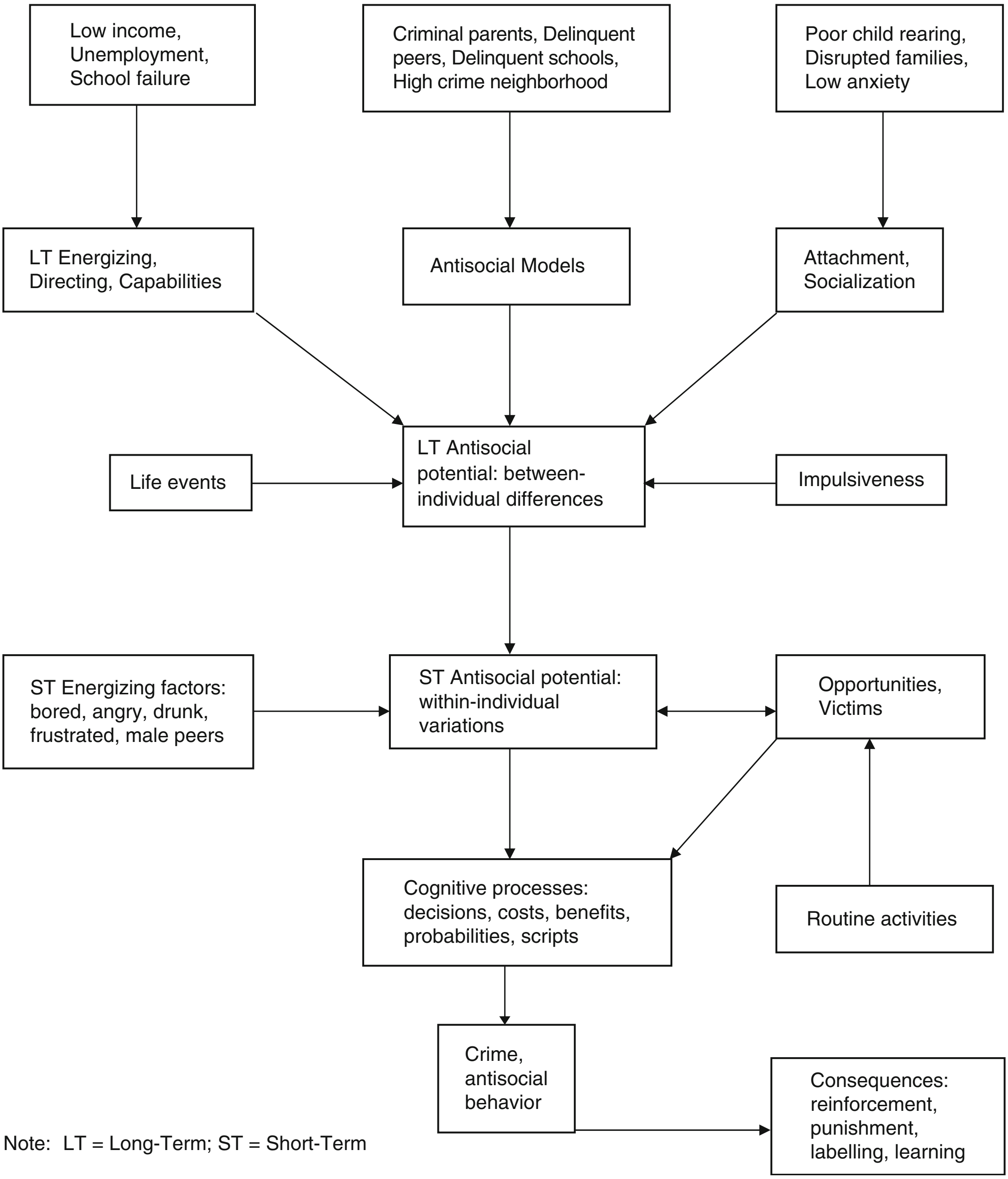
Low income, Unemployment, School failure
Criminal parents, Delinquent peers, Delinquent schools, High crime neighborhood
Poor child rearing, Disrupted families, Low anxiety
LT Energizing, Directing, Capabilities
Antisocial Models
Attachment, Socialization
Life events
LT Antisocial potential: between-individual differences
Impulsiveness
ST Energizing factors: bored, angry, drunk, frustrated, male peers
ST Antisocial potential: within-individual variations
Opportunities, Victims
Cognitive processes: decisions, costs, benefits, probabilities, scripts
Routine activities
Crime, antisocial behavior
Consequences: reinforcement, punishment, labelling, learning

Note: LT = Long-Term; ST = Short-Term

Integrated Cognitive Antisocial Potential Theory, Fig. 1 The integrated cognitive antisocial potential (ICAP) theory

researchers who have focused on situational influences have argued that different models are needed for different types of crimes. It is suggested that situational influences on burglary may be very different from situational influences on violence.

One possible way to resolve these differing viewpoints would be to assume that long-term potential was very general (e.g., a long-term potential for antisocial behavior), whereas short-term potential was more specific (e.g., a short-term potential for violence).

The top half of the model in Fig. 1 could be the same for all types of crimes, whereas the bottom half could be different (with different situational influences) for different types of crimes.

In the interests of simplification, Fig. 1 makes the ICAP theory appear static rather than dynamic. For example, it does not explain changes in offending at different ages. Since it might be expected that different factors would be important at different ages or life stages, it seems likely that different models would be needed at different ages. Perhaps parents are more important in influencing children, peers are more important in influencing adolescents, and spouses and partners are more important in influencing adults.

Long-Term Risk Factors

A great deal is known about risk factors that predict long-term persisting between-individual differences in antisocial potential. For example, in the Cambridge Study, the most important childhood risk factors for later offending were hyperactivity-impulsivity-attention deficit, low intelligence or low school attainment, family criminality, family poverty, large family size, poor child-rearing, and disrupted families. Figure 1 shows how risk factors are hypothesized to influence long-term AP. This figure could be expanded to specify protective factors and to include influences on onset, persistence, escalation, de-escalation, and desistance.

I have not included measures of antisocial behavior (e.g., aggressiveness or dishonesty) as risk factors in the ICAP theory because of my concern with explanation, prevention, and treatment. These measures do not cause offending; they predict offending because of the underlying continuity over time in AP. Measures of antisocial behavior are useful in identifying risk groups but less useful in identifying causal factors to be targeted by interventions. Similarly, I have not included variables that cannot be changed, such as gender or ethnicity. I assume that their relationships with offending are mediated by changeable risk factors (Farrington et al. 2003).

A major problem is to decide which risk factors are causes and which are merely markers or correlated with causes (Murray et al. 2009). Ideally, interventions should be targeted on risk factors that are causes. Interventions targeted on risk factors that are merely markers will not necessarily lead to any decrease in offending. Unfortunately, when risk factors are highly intercorrelated (as is usual), it is very difficult to establish which are causes in between-individual research. For example, the particular factors that appear to be independently important as predictors in any analysis may be greatly affected by measurement error and by essentially random variations between samples.

It is also important to establish how risk factors or causes have sequential or interactive effects on offending. Figure 1 shows how risk factors are hypothesized to influence long-term AP. Following strain theory, the main energizing factors that potentially lead to high long-term AP are desires for material goods, status among intimates, excitement, and sexual satisfaction. However, these motivations only lead to high AP if antisocial methods of satisfying them are habitually chosen. Antisocial methods tend to be chosen by people who find it difficult to satisfy their needs legitimately, such as people with low income, unemployed people, and those who fail at school. However, the methods chosen also depend on physical capabilities and behavioral skills; for example, a 5-year-old would have difficulty in stealing a car. For simplicity, energizing and directing processes and capabilities are shown in one box in Fig. 1.

Long-term AP also depends on attachment and socialization processes. AP will be low if parents consistently and contingently reward good behavior and punish bad behavior. (Withdrawal of love may be a more effective method of socialization than hitting children.) Children with low anxiety will be less well socialized, because they care less about parental punishment. AP will be high if children are not attached to (prosocial) parents, for example, if parents are cold and rejecting. Disrupted families (broken homes) may impair both attachment and socialization processes.

Long-term AP will also be high if people are exposed to and influenced by antisocial models, such as criminal parents, delinquent siblings, and delinquent peers, for example, in high crime schools and neighborhoods. Long-term AP will also be high for impulsive people, because they tend to act without thinking about the consequences. Also, life events affect AP; it decreases (at least for males) after people get married or move out of high crime areas, and it increases after separation from a partner (Theobald and Farrington 2009).

There may also be interaction effects between the influences on long-term AP. For example, people who experience strain or poor socialization may be disproportionally antisocial if they are also exposed to antisocial models. In the interests of simplicity, Fig. 1 does not attempt to show such interactions.

Figure 1 shows some of the processes by which risk factors are hypothesized to have effects on AP. It does not show biological factors, but these could be incorporated in the theory at various points. For example, the children of criminal parents could have high AP partly because of genetic transmission, excitement-seeking could be driven by low cortical arousal, school failure could depend partly on low intelligence, and high impulsiveness and low anxiety could both reflect biological processes.

Many researchers have measured only one risk factor (e.g., impulsiveness) and have shown that it predicts or correlates with offending after controlling for a few other "confounding factors," often including social class. The message of Fig. 1 is don't forget the big picture. The particular causal linkages shown in Fig. 1 may not be correct, but it is important to measure, analyze, and explain all important risk factors in trying to draw conclusions about the causes of offending or the development of offenders.

Ideally, protective factors should be included in the ICAP theory as well as risk factors. The risk factor prevention paradigm (Farrington 2000) suggests that risk factors should be reduced while protective factors are enhanced. Often, programs focusing on protective factors (e.g., building on strengths, promoting healthy development) are more attractive to communities and consequently easier to implement than programs focusing on risk factors (which seem to emphasize undesirable features of communities and hence imply blame). Of course, a key question is whether the risk and protective factors for offending (or for the onset of offending) are the same as or different from the risk and protective factors for reoffending (or for the persistence of offending). Perhaps different models are needed in Fig. 1 for onset, persistence, escalation, de-escalation, and desistance.

Explaining the Commission of Crimes

According to the ICAP theory, the commission of offenses and other types of antisocial acts depends on the interaction between the individual (with his immediate level of AP) and the social environment (especially criminal opportunities and victims). Short-term AP varies within individuals according to short-term energizing factors such as being bored, angry, drunk, or frustrated or being encouraged by male peers. Criminal opportunities and the availability of victims depend on routine activities. Encountering a tempting opportunity or victim may cause a short-term increase in AP, just as a short-term increase in AP may motivate a person to seek out criminal opportunities and victims.

Whether a person with a certain level of AP commits a crime in a given situation depends on cognitive processes, including considering the subjective benefits, costs and probabilities of the different outcomes, and stored behavioral repertoires or scripts. The subjective benefits and costs include immediate situational factors such as the material goods that can be stolen and the likelihood and consequences of being caught by the police. They also include social factors such as likely disapproval by parents or female partners and encouragement or reinforcement from peers. In general, people tend to make decisions that seem rational to them, but those with low levels of AP will not commit offenses even when (on the basis of subjective expected utilities) it appears rational to do so. Equally, high short-term levels of AP (e.g., caused by anger or drunkenness) may induce people to commit offenses when it is not rational for them to do so.

The consequences of offending may, as a result of a learning process, lead to changes in long-term AP and in future cognitive decision-making processes. This is especially likely if the consequences are reinforcing (e.g., gaining material goods or peer approval) or punishing (e.g., receiving legal sanctions or parental disapproval). Also, if the consequences involve labeling or stigmatizing the offender, this may make it more difficult for him to achieve his aims legally and hence may lead to an increase in AP. (It is difficult to show these feedback effects in Fig. 1 without making it very complex.)

A further issue that needs to be addressed is to what extent types of offenders might be distinguished. Perhaps some people commit crimes primarily because of their high long-term AP (e.g., the life-course-persistent offenders of Moffitt 1993), others primarily because of situational influences (e.g., getting drunk frequently) and high short-term AP, while still others offend primarily because of the way they think and make decisions when faced with criminal opportunities. From the viewpoint of both explanation and prevention, research is needed to classify types of people according to their most influential risk factors and most important reasons for committing crimes.

Predictions About Contentious Issues

The ICAP theory makes the following predictions about the contentious issues listed earlier. First, it predicts that the individual offending frequency and the seriousness of offending should increase to a peak in the teenage years and then decline, because both reflect AP. Second, the theory predicts that an early age of onset of offending predicts a high individual offending frequency and a high average seriousness of offending, again because both reflect AP. An early age of onset indicates a relatively high level of AP, and the ordering of people on AP tends to stay relatively consistent over time. Similarly, chronic offenders should have both a high frequency and a high seriousness of offending. Third, the ICAP theory predicts that onset sequences generally are age-appropriate behavioral manifestations of AP and therefore that they reflect persistent heterogeneity. However, the theory also incorporates learning and labeling processes following offending, so this would predict some state dependence effects.

Fourth, the ICAP theory predicts that within-individual changes in the major risk factors should be followed by within-individual changes in AP and hence in offending. Fifth, the theory predicts that short-term energizing factors such as heavy drinking influence short-term within-individual variations in offending: why people commit offenses in some times and places but not others. Sixth, the ICAP theory predicts that different risk factors would influence early onset before age 20, continuation of offending after onset, later onset after age 20, and persistence or desistance of offending after age 20. This is because it assumes that risk factors have different effects on AP at different ages. Seventh, the theory predicts that the criminal justice system should have deterrent or labeling effects on future offending. Future research should aim to establish to what extent these predictions are correct or incorrect.

Testing the ICAP Theory

The first independent test of the ICAP theory was carried out by Van Der Laan et al. (2009) in the Netherlands. Nearly 1,500 youth ages 10–17 completed a survey that enquired about long-term and short-term (situational) risk factors for delinquency. Nearly 300 of these answered questions about the circumstances of their last offense. In agreement with the ICAP theory, Van Der Laan and his colleagues found that long-term individual, family, and school factors correlated with serious delinquency, and the probability of serious delinquency increased with the number of factors. However, after controlling for long-term factors, short-term situational factors such as the absence of tangible guardians and using alcohol or drugs prior to the offense were still important. While the results of this test are very encouraging, it would be useful in the future to test the theory in a longitudinal survey.

The ICAP theory could be tested in the Cambridge Study by using antisocial attitude scores at different ages as measures of long-term antisocial potential. These could be related to individual, family, peer, and school risk factors to investigate the relative strength of relationships at different ages. They could also be related to self-reported and official offending at different ages. Different causal models could be tested at different ages. For example, in the Pittsburgh Youth Study, Defoe et al. (2013) concluded that hyperactivity and low socioeconomic status led to low school achievement, which led to delinquency, which in turn led to depression. In the Cambridge Study, the importance of long-term risk factors and short-term situational influences should be compared.

Conclusions

The ICAP theory, like other DLC theories, explains the widely accepted findings that I have listed above. However, more efforts should be made to compare and contrast the ICAP theory point by point with other DLC theories in regard to their predictions and their agreement with empirical results (Farrington 2006). Detailed computer models of the ICAP theory should be developed that allow users to click on the boxes of Fig. 1 to see more underlying detail or to see different theories for onset, continuation, and desistance. It would be better to make quantitative rather than qualitative predictions; criminological theories typically predict that X is related to Y or that X is greater than Y, but not the magnitude of relationships. These could build on the mathematical models of MacLeod et al. (2012), which hypothesize three types of offenders with different rates of offending and different probabilities of desistance. It is particularly important to specify crucial tests, where predictions from the ICAP theory differ from predictions from another DLC theory.

The ICAP theory could be improved in several ways. First, different models could be specified at different ages, based on the relative importance of various long-term risk factors (strain, modeling, socialization, impulsiveness, and life events). It might be expected that parental and peer influences would be differentially important at different ages, and the same is true of life events such as leaving school, getting a job, leaving home, and getting married. Second, risk factors that are not currently in the theory could be incorporated (e.g., biological factors or empathy). Third, the theory could be modified to explain differences in offending by age, gender, race and nationality, or country (Farrington and Loeber 1999). Fourth, different models could be proposed to explain the prevalence and frequency of offending, persistence versus desistance, and the length of criminal careers. Fifth, different models of short-term situational influences could be proposed to explain the commission of different types of crimes. Potential offenders could be asked about their likelihood of committing crimes in different situations to investigate their cognitive processes and motivations (Farrington 1993), and efforts could be made to study the learning processes involved in the development of criminal careers.

In order to advance knowledge about DLC theories and test them, new prospective longitudinal studies are needed with repeated self-report and official record measures of offending. Future longitudinal studies should follow people up to later ages and focus on desistance processes. Past studies have generally focused on onset and on ages up to 30. Future studies should compare risk factors for early onset, continuation after onset (compared with early desistance), frequency, seriousness, later onset, and later persistence versus desistance. DLC theories should make explicit empirical predictions about all these topics. Also, future studies should make more effort to investigate protective factors, and biological and neighborhood risk factors, since most is known about individual and family risk factors. And future research should compare development, risk factors, and life events for males versus females and for different ethnic and racial groups in different countries.

Because most previous analyses of risk factors for offending involve between-individual comparisons, more within-individual analyses of offending are needed in longitudinal studies. These should investigate to what extent within-individual changes in risk and protective factors are followed by within-individual changes in offending and other life outcomes (Farrington et al. 2002). These analyses could provide compelling evidence about causal mechanisms. More information is needed about the predictability of future criminal careers at different ages and stages, in order to know when and how it is best to intervene.

DLC researchers should carry out more research on explaining why crimes are committed, and especially on situational influences. Research should also be conducted on interactions between individual and situational factors in explaining the commission of crimes. Studies are also needed to classify types of people according to their most influential risk factors and most important reasons for committing crimes. Questions about situational influences on offending should be included in longitudinal studies of the development of offenders.

DLC theories have many policy implications for the reduction of crime. First, it is clear that children at risk can be identified with reasonable accuracy at an early age. The worst offenders tend to start early and have long criminal careers. Often, offending is preceded by earlier types of antisocial behavior in a developmental sequence, including cruelty to animals, bullying, truancy, and disruptive school behavior. It is desirable to intervene early to prevent the later escalation into chronic or life-course-persistent offending and to develop risk-needs assessment instruments to identify children at risk of becoming chronic offenders. These instruments could be implemented soon after school entry, at ages 6–8.

It would be desirable to derive implications for intervention from DLC theories and to test these in randomized experiments. In principle, conclusions about causes can be drawn more convincingly in experimental research than in nonexperimental longitudinal studies. The results summarized here have clear implications for intervention. The main idea of risk-focused prevention is to identify key risk factors for antisocial behavior and then implement prevention methods designed to counteract them (Farrington 2007). In addition, attempts should be made to enhance key protective factors.

Because of their emphasis on development through life, DLC theories suggest that it is "never too early, never too late" (Loeber and Farrington 1998) to intervene successfully to reduce offending. In other words, it is highly desirable to focus not only on early intervention to prevent the adolescent onset of offending but also on later programs to prevent adult onset and to intervene with offenders to prevent continuation and encourage early desistance.

The fact that offenders tend to be antisocial in many aspects of their lives means that any measure that succeeds in reducing offending will probably have wide-ranging benefits in reducing, for example, accommodation problems, relationship problems, employment problems, alcohol and drug problems, and aggressive behavior. Consequently, it is very likely that the financial benefits of successful programs will greatly outweigh their financial costs. The time is ripe to mount a new program of research to compare, contrast, and test predictions from the ICAP and other DLC theories, in the interests of developing more valid theories and more effective crime prevention policies.

Related Entries

- ▶ Age-Crime Curve
- ▶ Career Criminals and Criminological Theory
- ▶ Cognitive/Information Processing Theories of Aggression and Crime
- ▶ Desistance from Crime
- ▶ Establishing Causes of Offending in Longitudinal and Experimental Studies
- ▶ Labeling Theory
- ▶ Longitudinal Studies in Criminology

- ▶ Moffitt's Developmental Taxonomy of Antisocial Behavior
- ▶ Offender Decision Making and Behavioral Economics
- ▶ Social Learning Theory

Recommended Reading and References

Bernard TJ, Snipes JB (1996) Theoretical integration in criminology. In: Tonry M (ed) Crime and justice, vol 20. University of Chicago Press, Chicago, pp 301–348

Defoe IN, Farrington DP, Loeber R (2013) Disentangling the relationship between delinquency and hyperactivity, low achievement, depression, and low socio-economic status: analysis of repeated longitudinal data. J Crim Justice 41:100–107

Farrington DP (1992) Explaining the beginning, progress, and ending of antisocial behavior from birth to adulthood. In: McCord J (ed) Facts, frameworks and forecasts. Transaction, New Brunswick, pp 253–286

Farrington DP (1993) Motivations for conduct disorder and delinquency. Dev Psychopathol 5:225–241

Farrington DP (2000) Explaining and preventing crime: the globalization of knowledge – The American Society of Criminology 1999 presidential address. Criminology 38:1–24

Farrington DP (2003) Developmental and life-course criminology: key theoretical and empirical issues – The 2002 Sutherland Award address. Criminology 42:211–255

Farrington DP (ed) (2005a) Integrated developmental and life-course theories of offending. Transaction, New Brunswick

Farrington DP (2005b) The integrated cognitive antisocial potential (ICAP) theory. In: Farrington DP (ed) Integrated developmental and life-course theories of offending. Transaction, New Brunswick, pp 73–92

Farrington DP (2006) Building developmental and life-course theories of offending. In: Cullen FT, Wright JP, Blevins KR (eds) Taking stock: the status of criminological theory. Transaction, New Brunswick, pp 335–364

Farrington DP (2007) Childhood risk factors and risk-focused prevention. In: Maguire M, Morgan R, Reiner R (eds) The Oxford handbook of criminology, 4th edn. Oxford University Press, Oxford, UK, pp 602–640

Farrington DP (2013) Longitudinal and experimental research in criminology. In: Tonry M (ed) Crime and justice, 1975–2025. University of Chicago Press, Chicago

Farrington DP, Loeber R (1999) Transatlantic replicability of risk factors in the development of delinquency. In: Cohen P, Slomkowski C, Robins LN (eds) Historical and geographical influences on psychopathology. Lawrence Erlbaum, Mahwah, pp 299–329

Farrington DP, Loeber R (2013) Two approaches to developmental/life-course theorizing. In: Cullen FT, Wilcox P (eds) The Oxford handbook of criminological theory. Oxford University Press, Oxford, UK, pp 226–252

Farrington DP, Ttofi MM (2012) Protective and promotive factors in the development of offending. In: Bliesener T, Beelman A, Stemmler M (eds) Antisocial behavior and crime: contributions of developmental and evaluation research to prevention and intervention. Hogrefe, Cambridge, MA, pp 71–88

Farrington DP, Welsh BC (2007) Saving children from a life of crime: early risk factors and effective interventions. Oxford University Press, Oxford, UK

Farrington DP, Loeber R, Yin Y, Anderson SJ (2002) Are within-individual causes of delinquency the same as between-individual causes? Crim Behav Ment Health 12:53–68

Farrington DP, Loeber R, Stouthamer-Loeber M (2003) How can the relationship between race and violence be explained? In: Hawkins DF (ed) Violent crime: assessing race and ethnic differences. Cambridge University Press, Cambridge, UK, pp 213–237

Farrington DP, Loeber R, Ttofi MM (2012) Risk and protective factors for offending. In: Welsh BC, Farrington DP (eds) The Oxford handbook of crime prevention. Oxford University Press, Oxford, UK, pp 46–69

Farrington DP, Piquero AR, Jennings WG (2013) Offending from childhood to late middle age: recent results from the Cambridge Study in delinquent development. Springer, New York

Loeber R, Farrington DP (1998) Never too early, never too late: risk factors and successful interventions for serious and violent juvenile offenders. Stud Crime Crime Prev 7:7–30

MacLeod JF, Grove PG, Farrington DP (2012) Explaining criminal careers: implications for justice policy. Oxford University Press, Oxford, UK

Moffitt TE (1993) Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. Psychol Rev 100:674–701

Murray J, Farrington DP, Eisner MP (2009) Drawing conclusions about causes from systematic reviews of risk factors: the Cambridge quality checklists. J Exp Criminol 5:1–23

Nagin DS, Farrington DP (1992) The onset and persistence of offending. Criminology 30:501–523

Piquero AR, Farrington DP, Blumstein A (2003) The criminal career paradigm. In: Tonry M (ed) Crime and justice, vol 30. University of Chicago Press, Chicago, pp 359–506

Theobald D, Farrington DP (2009) Effects of getting married on offending: results from a prospective longitudinal survey of males. Eur J Criminol 6:496–516

Van Der Laan A, Blom M, Kleemans ER (2009) Exploring long-term and short-term risk factors for serious delinquency. Eur J Criminol 6:419–438

Zara G (2010) David P. Farrington: the integrated cognitive antisocial potential theory. In: Cullen FT, Wilcox P (eds) Encyclopedia of criminological theory, vol 1. Sage, Thousand Oaks, pp 313–322

Integrating Rational Choice and Other Theories

Stephen G. Tibbetts
Department of Criminal Justice, California State University, San Bernardino, San Bernardino, CA, USA

Synonyms

Choice theory; Decision making theory; Rational choice theory; Theoretical integration

Overview

This entry will discuss the integration of rational choice theory with other theoretical perspectives in developing a better understanding of why individuals engage in various criminal activities. Both the strengths and weaknesses of such integration will be examined.

Fundamentals/Historical Context

Rational Choice theory, also known as reasonable action theory or choice theory, is a perspective of human behavior that is based on the assumption that individuals will naturally make decisions of whether or not to act in a specific situation in order to maximize their pleasure or benefits, while minimizing potential pain or punishment. This perspective readily lends itself to integration with other theories of crime, given that it can be applied to virtually every type of activity (including all conventional activities as well). The fact that rational choice theory was originally developed by economists, and then adopted by criminologists to explain criminal behavior, is a strong indication of the versatility and general validity of this perspective. Thus, rational choice theory is inherently an integrated theory, and furthermore is quite accommodating to integration with other theories of crime.

Historical Context

Rational Choice theory was first presented as a framework for explaining and understanding of why individuals choose to engage (or not to engage) in certain behaviors in a variety of economic or social behaviors. Emerging primarily from the economic field, the most notable proponent of this perspective was Gary Becker (1968), who later won a Nobel Prize in 1992 in the area of Economic Sciences. In the last couple of decades, the rational choice framework has become perhaps the dominant perspective in the area of microeconomics, which emphasizes the reasons for why people make the decisions they make regarding many aspects of their life.

This perspective assumes that the primary reasoning of individuals in virtually all decisions they make is based on perceived expectations of the potential benefits versus potential costs that are likely to result in such decisions. Obviously, this relates to economic decisions regarding personal finances, but also applies directly to decisions made regarding whether or not to engage in a variety of criminal acts. After all, illegal activities virtually always involve potential risk of punishment or costs, and also typically involve some kind of potential benefit, whether it be financial profit, peer-respect, etc. Thus, it is not surprising that criminologists readily embraced the rational choice perspective and promoted it as a robust framework in developing models of explaining why individuals choose to commit (or not commit) criminal acts. Regardless of the activity being examined, such as financial, social, criminal, etc., the bottom line of rational choice theory is that when individuals make a decision, they are assumed to balance their perceptions of likely benefits versus likely costs of their decision.

Context in Criminological Literature

The development of rational choice theory in the criminological literature can not be understood without acknowledging the rebirth of classical deterrence theory in the field. In the late 1960s, studies began to highlight and examine the importance of the rates of being apprehended by police (often measured as ratios of arrests per crimes reported to police for given offenses), as well as the severity of being caught (typically measured by the length of jail/prison sentences given if convicted for certain types of crimes). Comparisons were made between jurisdictions based on these certainty ratios and length of sentences, with most studies concluding that the certainty ratios appeared to be consistently and significantly related to lower crime rates, but the length of sentence not having as much influence on crime rates, but still related to lower crime rates in some of the studies.

However, researchers soon realized that such aggregate measures are likely not to be considered by individuals who make decisions to commit criminal offenses. Thus, researchers advanced their models to examine individual perceptions of potential risks, such as the likelihood of being caught and the severity that the formal punishments would be if convicted of such crimes. The acknowledgment of this weakness led to cross-sectional studies (i.e., taken at one point in time) that examined individuals' perceptions of certainty and severity of punishment of being caught and then compared them to their actual offending. Studies showed that such perceptions of higher certainty of punishment were consistently linked to lower reported criminal activity, whereas perceptions of the higher severity of being punished were more inconsistent, but many studies showed that such perceptions were also linked to lower reported criminal offending. But once again, researchers realized that their methodology was flawed in the sense that it was likely that individuals' experience in committing crimes was altering their perceptions of getting caught (i.e., experiences affect perceptions of risk, labeled an "experiential effect"), as opposed to the proposed model of perceptions of risk affecting their decisions to commit crimes (which is the classic deterrence model).

To address this weakness in their methodology, a new wave of research emerged that focused on longitudinal or panel studies that measured individuals' perceptions of potential risks, as well as prior criminal offending, in the first wave. Then, at a later date (such as a year later), the subsequent wave(s) measured their reported offending and analyzed this while taking into account their perceptions and past offending measures collected at the first wave. This approach was largely initiated by Raymond Paternoster and his colleagues in a series of studies published in the early- to mid-1980s (e.g., Paternoster et al. 1983). These seminal studies revealed that previous suspicions were indeed true; specifically, an "experiential effect" existed and had a highly significant effect on persons' perceptions of risk. This was a key finding that highly influenced future research in both deterrence and emerging rational choice models for years to come.

Despite the acknowledgement and accounting of an experiential effect in determining individuals' perceptions of risk, the studies that accounted for this effect still typically showed significant effects for levels of perceived formal sanctions, such as the risk of being caught and the severity of being caught/convicted. Still, these models were limited to a more traditional deterrence model, which emphasized only formal sanctions and past offending. This set the stage for a new theoretical framework that introduced other important factors that go into most individuals' decisions to commit crime, namely the rational choice perspective.

Early Rational Choice Models of Criminal Behavior

According to surveys of empirical research (see Tibbetts and Gibson 2002), prior to 1986 there were approximately a dozen of studies on applying the rational choice framework to criminal behavior. Most of these studies were based largely on more traditional deterrence factors, such as perceptions of the certainty of being caught and the perceived severity of punishment that would likely result from being caught. Virtually all of these pre-1986 studies tend to lack an examination of other factors, such as the benefits or profits from committing crimes, as well as the influence of a variety of other factors, such as moral beliefs, emotions, perceived informal factors, etc., nor did they typically attempt to integrate rational choice theory with other theoretical perspectives in the criminological field. Although they did not necessarily create much attention or inspire a flurry of research on rational choice theory, they certainly built the foundation for what was soon to come.

In 1986, the first notable burst of research regarding rational choice theory in explaining criminal behavior emerged by Derek Cornish and Ronald Clarke, who published their seminal book, *The Reasoning Criminal: Rational Choice Perspectives on Offending*. This anthology included many chapters that explored applying rational choice to a variety of offenses, as well as promoting the expansion of the perspective to include or integrate more factors that would likely be important in an individuals' decision to engage in criminal activity. This book was well-received by scholars and the public alike, and was likely almost entirely responsible for the later surge of rational choice studies that were published in the early- to mid-1990s. The several years gap between their 1986 publication makes sense given a normal time-gap between the publication of a notable work and the research that results from it. This gap is expected due to reading the new work (in this case Cornish & Clarke's book), developing new research agendas, creating new survey instruments and measures, administering such surveys or acquiring new data, analyzing the results, writing up the manuscripts, and then going through the process of getting published in peer-reviewed journals. Regardless, the large attention given to rational choice theory a handful of years later showed that the call for more research was quickly answered.

Integration of Informal Sanctions and Personal Sanctions (Self-Conscious Emotions)

In fact, in only a few years after the publication of *The Reasoning Criminal*, a series of studies were published by Harold Grasmick, Robert Bursik, and their colleagues that examined the rational choice framework on a variety of crimes, such as littering, theft, and drunk driving (e.g., Grasmick and Bursik 1990). Beyond the more traditional deterrence factors of official risks of being caught and severity of such punishments if caught, some of these studies also included more informal factors, such as how the individual would expect their family, friends, employers, etc., to react to them being caught for engaging in such behaviors. These informal sanctions that were integrated into estimated rational choice models were largely derived from social disorganization theory (or Chicago School of criminology), which places a focus on the influence that one's social environment – such as family, peers and neighborhood – has on their likelihood of committing crime and delinquency. These studies consistently found such informal factors to be consistently significant influences on predicting the likelihood of people to decide to commit such offenses, and were often far more important than any formal, legal sanctions that were expected. Specifically, the higher the levels of reported expectations that friends, family, employer, etc., would affect the individual negatively for committing such crimes, the less likely the individual was to engage in such behaviors. This was not too surprising given that it is likely that most people care more about their standing with people they interact with every day than with the official sanctions of going to jail for a few months. The informal sanctions are likely to impact their long-term future, as opposed to the typically short-term consequences of official, legal sanctions. Although this may not seem to be a major breakthrough, it was at the time for models on deterrence and rational choice. Furthermore, although not recognized or stated at that time, the incorporation of such factors as informal sanctions and emotions into models of rational models were a form of integration.

Also notably in these studies, the effects of new factors were introduced and estimated, such as the levels of expected embarrassment or shame a person would feel for being caught for engaging in such activities. These emotional factors, especially shame, not only showed consistently significant inhibitory effects on decisions to commit such illegal acts, but also inspired many subsequent studies on the importance of such self-conscious emotions in rational choice models of offending. As with the influence of informal sanctions described above, the integration of these emotional components added an entirely new dimension in terms of factors that should be included in subsequent rational choice models of offending, if they are to be fully specified. Thus, the published research in early 1990s not only added support to what would become the largest wave of studies on rational choice theory in the criminological literature, but also set the stage for integrating other factors into such models.

Key Issues in Integration of Rational Choice with Other Theories

Peak of Research on Rational Choice and Integration with Social Control, Social Learning, and Social Disorganization Theories

Empirical studies have shown that the first peak of attention given to the rational choice perspective was clearly between the years of 1992–1997, according to the database of Criminal Justice Abstracts (see discussion in Tibbetts and Gibson 2002). This 6-year period experienced a significant increase in studies published in the criminological literature regarding applying rational choice theory to various types of criminal offending. Specifically, in these 6 years the total amount of rational choice studies more than doubled the number that had been done in all years prior to 1992. Many of the studies in this period appropriately took the earlier versions of rational choice theory and made them more fully specified, and thus more complete, models of criminal offending on given types of crimes.

Much of this expansion of the rational choice perspective involved the elaboration of the rational choice framework to include various factors or concepts that had not been traditionally considered deterrence variables, as discussed above and below. Other studies actually merged rational choice theory with other established theoretical perspectives, such as social control/bonding theory, social disorganization, social learning, and others. Therefore, a number of the most cited studies in this period integrated the rational choice perspective with various other types of theories, or at least key concepts taken from other established theories of crime.

The key studies in the period of the early- to mid-1990s attempted to integrate rational choice theory with such concepts from other established theories such as social control/bonding, social learning and social disorganization, by incorporating the concepts of attachments to significant others, commitment/involvement in the local community, and perceptions of the reactions of others in their local neighborhood or community would react to their offending as important factors in decisions to commit crimes. For example, the importance of perceptions of how their family, friends, employers, and others in their community were not only examined, but seemed to have more importance in most cases than the perceptions of formal sanctions (i.e., police, courts, corrections) that had been emphasized in the previous rational choice studies that were based almost exclusively on expectations of official punishments. Specifically, the findings of these studies were consistent in showing that the perceived responses of such informal factors, appeared to have just as much, if not more, of an influence on individuals' decisions to engage in a variety of criminal acts.

Such findings make sense because most people in society desire to please those who are closest to them and to the others they interact with on a daily basis. The formal sanctions, on the other hand, are typically considered somewhat removed from their everyday life, in the sense that the daily lives of most individuals are not typically directly affected by the opinion or reaction of police officers, judges, etc. Thus, it is not surprising that the key studies in the early 1990s that examined subjects' perceptions of informal sanctions were quite influential in their decision-making when it comes to engaging in various crimes.

The introduction of these informal sanctions were largely drawn from other established theories such as social learning, which emphasizes the influence of learned attitudes and behaviors of significant others regarding a specific criminal activity, as well as the social rewards and punishments that come from deciding to engage (or not engage) in certain illegal acts. According to social learning theory, the internalization of such attitudes and behaviors of others by an individual regarding a given act increase the likelihood that individuals will either engage or not engage in such activities. In cases in which a person's significant others actually promote the commission of crime, then the individual will view engaging in the act as a benefit in terms of respect from the friends or others that will respect her or him for doing the act, which are seen as a form of reward. On the other hand, if most of the person's significant others look unfavorably on the commission of such an act, the individual is far less likely to engage in such behavior, because this is seen as a form of punishment. One of the benefits of the rational choice perspective is that any factor that creates a potential benefit or cost to their daily life can easily fit in this framework.

Integration of the Benefits or Thrills of Offending

The early 1990s wave of studies also, for the first time, integrated and placed a large emphasis on individuals' perceptions of benefits, or pay-off, from engaging in various criminal activities. Much of this attention was due to the recent publication of the book in 1988 by Jack Katz, *Seductions of Crime*, in which he emphasized the importance of the perceived benefits of committing crime or delinquency. Specifically, Katz had pointed out that there were many positive reasons for why certain individuals, especially young people, who engage in crime for various benefits, such as getting "sneaky thrills" out of offending. Youth tend to get a thrill from committing some offenses that have no actual pay-off in terms of utility, especially among certain groups of youths (such as graffiti, vandalism, unprovoked assaults, being a member of a street gang, etc.).

A good example is vandalism, such as destroying mail-boxes, breaking windows, or other similar forms. Katz specified that much of these behaviors that seemed to have no actual pay-off really did have a large pay-off other terms, such as peer respect, confidence, and relief of built-up anger/stress. Recent studies have shown that such activity actually increases dopamine and testosterone levels, which adds a physiological aspect to the benefits of crime that were not directly examined by most rational choice studies in the early 1990s period, but would surface later. Ultimately, the integration of Katz' model of benefits or pay-off, either from a financial or social or physiological aspect, was one of the key contributions of this wave of studies in the early 1990s. This element fit well into the rational choice perspective as an element of benefit from engaging in criminal acts.

Integration of Moral Beliefs (Social Bonding Theory)

Another notable integration in this large wave of key studies on rational choice theory in the early 1990s was the emphasis on personal moral beliefs, which was largely derived from the social control/bonding theory of crime (Hirschi 1969). Although social bonding theory specified four elements, some of these key elements – attachments to others, commitment/involvement to society, and conventional activities – had already been indirectly integrated into rational choice studies by prior models via informal factors (see above). But it was the last element of moral beliefs that had previously been unexplored, but became an emphasis on influencing the other variables in estimated models of rational decision-making regarding criminal behavior. Specifically, the integration of the concept of moral beliefs, previously seen as a social bonding concept, was presented in rational choice models of crime as a conditioning factor on the effects of individuals' perceptions of the certainty or severity of getting caught when considering engaging in certain types of crimes.

Some of the first key studies to integrate the concept of moral beliefs as a conditional variable was Ronet Bachman, Raymond Paternoster and their colleagues (e.g., Bachman et al. 1992; Paternoster and Simpson 1996), which revealed that the likelihood of decisions to engage in a variety of crimes such as sexual assault and corporate crimes were significantly influenced by their personal beliefs regarding the morality (or immorality) of committing such acts. Furthermore, these studies also consistently showed that the effects of perceptions of likelihood of getting caught officially or severity of subsequent punishment tend to not as be as important when individuals' have high levels of moral beliefs regarding the inherent immorality to engage given specific criminal acts. The findings of these studies further confirmed that there was much more to personal decisions to engage in a given criminal act than simply balancing the perceived risk of formal and/or informal costs of the act versus the perceived likelihood of benefits or potential pay-off. Ultimately, the incorporation of moral beliefs represented a radical departure from the simple framework presented by the early rational choice theorists, which was limited to expected costs and/or benefits. These studies suggested that long-term, time-stable personal beliefs and perhaps many other personality traits or dispositions should be considered as potential factors that may alter the effects of perceived costs and benefits of committing certain crimes.

Integration of Self-Control Theory

The rational choice studies of the early 1990s were also highly influenced by the publication of the book (and prior studies) by Michael Gottfredson and Travis Hirschi, *A General Theory of Crime* (1990). This book claimed that virtually all criminal offending and deviant behaviors are caused by a single factor – low self-control. The theory proposed by the authors quickly became not only perhaps the most researched theory in the field, but also became in the 1990s the perspective ranked by criminologists as the second most important theory in terms of validity of what explains most criminal behavior (Ellis and Walsh 1999), ranked just after social control/bonding theory which was also integrated with rational choice models around the same time period (as discussed above).

Many of the studies on rational choice integrated the concept of self-control into their models of offending, primarily as a conditional or interaction variable, in terms of determining how individuals perceive the risks of being caught and the severity of punishments of being caught. Specifically, studies in the early 1990s by Daniel Nagin and Raymond Paternoster showed that many of the primary factors of the traditional classical/deterrence framework were conditioned by individuals' levels of self-control (Nagin and Paternoster 1993). These findings were further supported by subsequent studies that showed that not only were such traditional deterrence factors conditioned by self-control, but that other factors – specifically anticipated shame and anticipated benefits – also played a significant conditioning role in terms of decisions to engage in various criminal activities (Piquero and Tibbetts 1996). The findings of these studies created an entirely new area of research that explored the conditional effects of the more traditional deterrence/rational choice factors of perceived certainty and severity in light of many other factors that had not been included in traditional models of rational choice.

The amount of research given to rational choice theory in criminological research experienced a relative decrease in the late 1990s, as well as the early 2000 years (Tibbetts and Gibson 2002). In light of this notable decrease in attention given to rational choice theory, some criminologists specifically advised that in order to rational choice theory to make an attempt to elaborate on the rational choice framework and integrate various concepts from more established theories of crime. Subsequently, in the first decade of the new century, the attention and number of studies began to pick up again in the mid-2000s period; this is evidenced by the large number of studies that are currently referenced in Criminal Justice Abstracts between the years of 2005–2010, which averaged 28 studies for each year between 2005 and 2009 (with a high of 34 studies in 2007), which is far higher than any other 5-year period in the history of rational choice studies in criminological literature. Much of this attention was inspired by the studies that were done before in the early- to mid-1990s, that showed that numerous other factors can and do play a large part in individuals' decision-making beyond the traditional deterrence factors regarding perceived costs and benefits of committing crimes. Such studies began to apply rational choice theory to areas that had not been directly emphasized before, such as white-collar crime (for a review, see Simpson et al. 2002). This led to a burst of research that emphasized the integration of rational choice with other established theories, which was largely based on elaborating the rational choice framework to include other key concepts from such traditional, relatively more established theories that will be discussed below.

Recent Research Integrating Rational Choice with Other Theoretical Models and Concepts

It is notable that rational choice theory did not even make the list of the top 23 "most valid" theories in the 1999 study above that examined responses from over a hundred leading criminologists regarding which theories they thought had the most validity regarding serious crime (Ellis and Walsh 1999). This is one of the reasons that some researchers soon after claimed that in order for the rational choice framework to thrive in the criminological literature, researchers must begin integrating other theoretical concepts and propositions into the rational choice model of criminal decision making. Specifically, after a review of the extant research, Tibbetts and Gibson claimed in 2002, "the survival of rational choice theory largely depends on its integration with frameworks that emphasize individual propensities, which has been shown to complement both perspectives" (p. 18). Fortunately, it appears that this call for research has been taken by many researchers in the field. As mentioned above, the first decade of the new century saw a resurgence in the attention and number of studies on rational choice theory, especially those that integrated various concepts or propositions of other established theories into such estimated models. Notably, most of these studies were indeed related to elaborating the rational choice framework to account for individual propensities or dispositions toward committing criminal activity.

Enduring Individual Differences and Rational Choice Theory

The call for more rational choice studies to account for individual propensities in the early 2000 years appears to have been answered, because the most current research in rational choice theory are studies that integrate various other established theories or concepts from such theories, especially those accounting for individual dispositions toward committing crime. These personal propensities that are typically emphasized in most recent studies tend to relate to personality traits or other time-stable factors that are apparent from an individual's early antisocial behavior to later manifestations of criminality. After all, one of the key findings from empirical studies show that people tend to behave quite differently given the same situation, which is likely due to such personal dispositions and personality traits.

Related to the integration of other theories to rational choice, a recent review by Alex Piquero, Raymond Paternoster, and their colleagues recently explored the evidence related to this "differential deterrability" of that exists across different individuals with respect to the perceived sanctions (i.e., certainty and severity of punishment) of traditional rational choice models of crime (Piquero et al. 2011). This review showed that various traditional theories of crime provided much insight toward why certain individuals are more (or less) likely to engage in criminal activity. Specifically, the theories of social bonding, moral inhibition, self-control, impulsivity, emotional arousal, and pharmacological arousal were all shown to be significant additions to understanding the decision making of individuals in whether they would commit or not commit a given criminal act.

Regarding the factors that had not been discussed above in this entry, emotional and pharmacological arousal was one of the most important aspects that this review presented.

Specifically, the Piquero et al. review noted the pioneering studies of the 1990s and recent years that integrated contemporaneous emotional states into rational choice models of offending by such scholars as George Loewenstein, Daniel Nagin, Raymond Paternoster, and their colleagues. Such studies compared sexually aroused males (such as exposing some subjects to nude photographs) to non-aroused males on responses of their likelihood of committing date rape. Not only did the aroused males score higher in their intentions to commit such acts, but these studies also showed that their decisions to commit date rape were not as influenced by traditional rational choice/deterrence factors (e.g., perceived certainty/severity of punishment). Other studies have shown that the perceived benefits of such a date rape sexual act were only partially mediated by such traditional rational choice/deterrence factors.

Additional studies have explored the pharmacological influence of alcohol on individuals' intentions to commit crimes. Specifically, Assaad and Exum (2002) reviewed the extant literature on the influence of alcohol on individuals' likelihood of engaging in criminal activity, as well as how alcohol influences their perceptions of risks of being caught and punished, as well as their perceptions of potential benefits of committing such acts. This review concluded that (p. 77) "under the influence of alcohol, individuals are more likely to attend to salient, provocative cues in the environment. . .are less likely to recognize the costs associated with their own aggressive behavior. . .and may experience increases in arousal that in turn interferes with the cognitive processes that normally inhibit aggression." Furthermore, there are many other pharmacological influences that involve many other substances (e.g., cocaine, methamphetamine, etc.) that were not examined in this review, but are quite likely to influence to influence individuals' perceptions of risks versus benefits of engaging in illegal behaviors. Far more research is needed in this area of how other illegal substance usage influences the various perceptions of costs and benefits of a given act, and thus creating a better understanding of human behavior according to rational choice theory.

Ultimately, the most important point of this entry is the conclusion that the evolution of rational choice theory has recently become far more integrated with other concepts and propositions of more established theories of crime, as well as factors that have not been included in such traditional theories (e.g., emotional arousal, etc.). Such elaboration of the rational choice framework to incorporate these concepts and propositions has created a more robust theory that explains criminal behavior much better than the earliest forms of the theory presented in the 1970s and 1980s. However, despite this success, such integration of rational choice theory with other theories has seen its fair share of criticism.

Criticisms of the Integration of Other Theories with the Rational Choice Perspective

Criticisms of Rational Choice Theory and Integration with Other Theories

Some researchers would likely call such studies/theories that use rational choice theory as a framework, and then integrate other key concepts or propositions from other more established theories, as more of an elaboration of rational choice as opposed to integration of two or more separate theories. However, theoretical elaboration is generally considered a form of theoretical integration (Messner et al. 1989), so such elaborations or expanded models of the rational choice framework that include such key concepts or propositions from other established theories are generally accepted as theoretical integration. Ultimately, the goal of both elaboration or integration is to create a theory that represents the best explanation or understanding of the phenomenon being examined. Thus, whether it be theoretical elaboration or a more complete form of theoretical integration is somewhat moot, because both seek to develop the best model for predicting and understanding the offense being considered. This is what the many of the key rational choice studies of from the early 1990s to contemporary times has sought to do. Apparently the elaboration or integration of rational choice with concepts and/or propositions of other established theories has been successful and enlightening. This can be observed in the dramatic increase in the number of studies in the last decade on rational choice, as well as the significant increase in the explained variation of offending decisions in estimated statistical models of these expanded and integrated rational choice models of illegal behavior.

Other criticisms of the integration of rational choice with other theories of crime have claimed that the proponents of rational choice are in some ways stealing from the other established theories (such as differential reinforcement/association theory) because many of their concepts can easily fit into either a cost (e.g., peer disapproval) or benefit (e.g., peer approval) in a rational choice framework. However, it is true that one of the most beneficial aspects of rational choice models of offending is that it does indeed allow for the introduction of any aspects or factors that may influence individuals' perceptions of anticipated costs or benefits given a particular opportunity to commit a criminal act. Given that the rational choice framework is more of a general theory of behavior, it *should and must include* all relevant factors in terms of specifying models for engaging in illegal acts, especially factors that significantly influence a person's reasoning when deciding whether or not to engage in such activity. Therefore this criticism of being allowed to include many factors that influence individual decision making regarding various behaviors is also somewhat moot, because the inclusion of such factors that affect a person's perception of expected costs versus benefits is what rational choice models of offending are inherently about.

Ultimately, the major criticisms against the integration or elaboration of rational choice models to include factors from more traditionally established models of crime have been shown to be weak at best. After all, why should any theory have "property claim rights" over any concept, factor, or proposition regarding human behavior. Thus, the introduction of the various concepts to the rational choice theoretical framework over the last decades should be embraced, especially in light of how much better the recent estimated rational choice models of crime have shown success in creating a far better understanding of individuals' perceptions in choosing whether or not to engage in a given criminal act.

Related Entries

- ▶ Deterrence
- ▶ Deterrence: Actual Versus Perceived Risk of Punishment

Recommended Reading and References

Assaad J, Exum ML (2002) Understanding intoxicated violence from a rational choice perspective. In: Piquero A, Tibbetts S (eds) Rational choice and criminal behavior. Routledge, New York, pp 65–84

Bachman R, Paternoster R, Ward S (1992) The rationality of sexual offending: testing a deterrence/rational choice conception of sexual assault. Law Soc Rev 26:343–372

Cornish D, Clarke R (1986) The reasoning criminal: rational choice perspectives on offending. Springer, New York

Ellis L, Walsh A (1999) Criminologists' opinions about causes and theories of crime and delinquency. The Criminologist 24:1–4

Gottfredson M, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Grasmick H, Bursik R (1990) Conscience, significant others, and rational choice: extending the deterrence model. Law Soc Rev 24:837–861

Hirschi T (1969) Causes of delinquency. University of California Press, Berkeley

Katz J (1988) Seductions of crime. Basic Books, New York

Lowenstein G, Nagin D, Paternoster R (1997) The effect of sexual arousal on expectations of sexual forcefulness. J Res Crime Delinq 34:443–473

Messner S, Krohn M, Liska A (1989) Theoretical integration in the study of deviance and crime: problems and prospects. State University of New York Press, Albany

Nagin D, Paternoster R (1993) Enduring individual differences and rational choice theories of crime. Law Soc Rev 27:467–496

Paternoster R, Simpson S (1996) Sanction threats and appeals to morality: testing a rational choice model of corporate crime. Law Soc Rev 30:378–399

Paternoster R, Saltzman L, Waldo G, Chiricos T (1983) Perceived risk and social control: do sanctions really deter? Law Soc Rev 17:457–480

Piquero A, Tibbetts S (1996) Specifying the direct and indirect effects of low self-control and situational factors in offenders' decision making: toward a more complete model of rational offending. Justice Quart 13:481–510

Piquero A, Paternoster R, Pogarsky G, Loughran T (2011) Elaborating the individual difference component in deterrence theory. Annu Rev Law Soc Sci 7:335–360

Simpson S, Piquero N, Paternoster R (2002) Rationality and corporate offending decisions. In: Piquero A, Tibbetts S (eds) Rational choice and criminal behavior. Routledge, New York, pp 25–39

Tibbetts S, Gibson C (2002) Individual propensities and rational decision-making: recent findings and promising approaches. In: Piquero A, Tibbetts S (eds) Rational choice and criminal behavior. Routledge, New York, pp 3–24

Intelligence

► Identification Technologies in Policing and Proof

Intelligence-Led Policing

Jerry H. Ratcliffe
Temple University, Philadephia, PA, USA

Synonyms

ILP; Information-led policing

Overview

Intelligence-led policing is a policing business model that incorporates data analysis and criminal intelligence into a strategy that coordinates strategic risk management of threats with a focus on serious, recidivist offenders. It originated in the UK but has been increasingly adopted around the world.

Definition

Intelligence-led policing (ILP) is a term now embedded in the lexicon of policing around the world. From its origins in the new public management ethos of 1990s Britain to hurried US recommendations for ILP adoption in the wake of the 9/11 terrorist attacks, law enforcement agencies have looked to ILP as the contemporary model (Carter 2009) to integrate the intelligence function into the operations of the police service. The National Criminal Intelligence Sharing Plan (GIWG 2003), developed for the US Department of Justice by the Global Justice Information Sharing Initiative (Global) Intelligence Working Group (GIWG), advocated heavily for the need for ILP, yet they neither defined it nor explained how it should be implemented (Carter and Carter 2009). Wardlaw and Boughton (writing in 2006) point out that “the concept of intelligence-led policing is now widely espoused by police services as a fundamental part of the way they do business. But for such a widely talked about concept, there is remarkably little clarity about its definition and fundamental concepts” (2006: 134). So while the term has joined the jargon of police departments seeking to establish their credentials as innovative and progressive, the absence of a definition from which to direct and measure implementation hampered initial development (Ratcliffe 2008a).

Various definitions of ILP now exist to fill the void. One stresses the philosophical role of the intelligence function within policing, describing it as “an underlying philosophy of how intelligence fits into the operations of a law enforcement organization. Rather than being simply an information clearinghouse that has been appended to the organization, ILP provides strategic integration of intelligence into the overall mission of the organization” (Carter 2004: 4). The Bureau of Justice Assistance has three definitions of ILP within the same document (BJA 2009): first describing it as “a business process for systematically collecting, organizing, analyzing, and utilizing intelligence to guide law enforcement operational and tactical decisions” (p. 3) and then as “a collaborative law enforcement approach combining problem-solving policing, information sharing, and police accountability, with enhanced intelligence operations” (p. 4). The same page finally settles on, “executive implementation of the intelligence cycle to support proactive decision making for resource allocation and crime prevention. In order to successfully implement this business process, police executives must have clearly defined priorities as part of their policing strategy" (p. 4). Thus, within the space of two pages, ILP is a business model, an operational tactic, and a strategic leadership function.

One definition has been increasingly adopted by a variety of writers (e.g., Burch and Geraci 2011; Gottschalk 2010; Hess and Orthmann 2011; Smith 2009) as the de facto definition:

> Intelligence-led policing is a business model and managerial philosophy where data analysis and crime intelligence are pivotal to an objective, decision-making framework that facilitates crime and problem reduction, disruption and prevention through both strategic management and effective enforcement strategies that target prolific and serious offenders. (Ratcliffe 2008a: 89)

This definition was arrived at after consultation and wordsmithing from a number of influential analysts and intelligence managers involved in the implementation of ILP within Britain's National Intelligence Model and beyond. The definition integrates analysis and intelligence into the decision-making structure of a police department and combines a traditional enforcement approach for serious and prolific offenders with the problem-oriented strategic management of crimes and other policing problems. With the inclusion of problems in the definition, there exists the possibility for the business model component to be applied to a variety of (noncrime) policing issues, such as traffic management and crash prevention (as is the case in New Zealand). As the original architects of the UK National Intelligence Model have pointed out, "we knew that the intelligence model was in fact an ideal model for the whole business of policing that would enable police commanders to understand and anticipate risks and threats across the public safety domain" (Flood and Gaspar 2009: 54).

Origins

Rising crime levels in many countries during the 1980s and 1990s drove some police departments to seek out more cost-effective strategies and greater efficiencies from increasingly limited police budgets (Ratcliffe 2008a). The "demand gap" of greater expectations placed on police not being met with equivalent increases in resources placed a strain across the policing sector and drove some to seek out more strategic approaches to the business of policing (Flood and Gaspar 2009). While the USA was the originator for problem-oriented policing, the UK's different policing structure and "more sophisticated legacy" with criminal intelligence (Carter and Carter 2009) provided other opportunities. Emerging from various criminal intelligence reforms during the 1970s (Flood and Gaspar 2009), two influential British government reports, *Helping With Enquiries: Tackling Crime Effectively* (Audit Commission 1993) and *Policing with Intelligence* (HMIC 1997), laid out the government's interpretation of the problems; existing police roles and accountability lacked integration, police were not making the best use of resources, and targeting criminals would be more effective than focusing on crime. The reports also pointed to a potential solution, with *Policing with Intelligence* stating in the first paragraph:

> Good quality intelligence is the life blood of the modern police service. It allows for a clear understanding of crime and criminality, identifies which criminals are active, which crimes are linked and where problems are likely to occur. It enables valuable resources to be targeted effectively against current challenges and emerging trends ensuring the best opportunities for positive intervention and maximum value for money. (HMIC 1997: 1)

Unencumbered by the damaging domestic intelligence failures and bad practices that plagued the development of criminal intelligence in the USA (Carter 2004; Ratcliffe 2008a), British policing leaders looked to reform the police service with a greater focus on the strategic management of crime risks through a better understanding of the criminal environment.

This period of sustained reform in British policing also saw the development of the UK National Intelligence Model (NIM). Enthusiastically adopted by a number of police leaders, some saw it as an occasion to push back against the prevailing neighborhood reassurance/community policing model popular at the time. In particular Sir David Phillips (then Chief Constable of Kent Constabulary) saw the NIM as an opportunity to restore the status of detectives and the criminal investigation department. However, the NIM, actively promoted by the National Criminal Intelligence Service and the Association of Chief Police Officers, established a business model that situated criminal intelligence and analysis – rather than criminal investigation – at the center of police decision-making and crime fighting (Harfield and Harfield 2008; NCIS 2000).

In the USA, the move toward intelligence-led policing has been hampered by a number of factors, both historical and organizational. The Church Report, which investigated the activities of the FBI's counterintelligence program that ran through the 1950s and 1960s, identified numerous shortcomings of the management of criminal intelligence programs. Furthermore, many police departments in the USA were sued under civil rights legislation for maintaining files on people with no criminal involvement (Carter and Carter 2009; Ratcliffe 2008a). There still exists to this day a distrust of police involvement in any activities labeled as "intelligence" (irrespective of how this is defined) both from some members of the community and, as a result, from within the justice system. For many years, the National Institute of Justice and other federal agencies have preferred to refer to ILP with the more vague and inaccurate term "information-led policing" (intelligence is information that has gone through a process of analysis to determine relevance and context to a crime problem).

Organizationally, the fragmented and disorganized way that law enforcement has evolved in the USA has influenced the development of many policing innovations. If the USA had the same population/police department ratio as the UK, it would have fewer than 300 police departments and not in excess of 18,000. Many of these departments consist of fewer than ten sworn officers, and as such their lack of size hampers the development of an intelligence capacity to support ILP. Furthermore, as Carter and Carter (2009) point out, in the USA there is no legacy of widespread intelligence use, the intelligence function has until recently been poorly defined, and there exists to this day a lack of training and expertise to develop the necessary function.

The plethora of small departments with no intelligence capability and neither a tradition nor expertise in said function has steered American development of ILP to emphasize intelligence sharing rather than strategic crime control. A central report that propelled ILP acceptance across the USA was called the National Criminal Intelligence Sharing Plan. The executive summary did not emphasize the needs of efficiency in the face of rising crime, as had the British reports a decade previously. Rather, it alluded to the perceived intelligence failures that led to the terrorist attacks of 9/11 and how that day "initiated a concerted effort by American law enforcement agencies to correct the inadequacies and barriers that impede *information and intelligence sharing*" (GIWG 2003, emphasis added). Tapping into the Zeitgeist may have enhanced readership, but it led many in US law enforcement to incorrectly assume that intelligence-led policing was predominantly a counterterrorism strategy, one that had little to do with many US policing environments.

A positive development of the concerns about lack of intelligence sharing has been the development of the fusion center concept. Locally in the USA, the initial response to 9/11 was overwhelmingly to invest in infrastructure support for first responders and to provide a protective function at locations perceived to be sensitive: there was no major investment in an offender-focused capacity. But with time, the development of intelligence capabilities has continued, both within individual police departments and regionally. The regional capacity has been enhanced with the development of fusion centers, defined as "a collaborative effort of two or more agencies that provide resources, expertise, and/or information to the center with the goal of maximizing the ability to detect, prevent, apprehend, and respond to criminal and terrorist activity" (DoJ 2005: 3). Within the homeland security context, due to the localized and decentralized nature of American policing, local government is more suited to protective functions of places rather than building a capacity to identify and target specific offenders. However, with at least 72 fusion centers now across the USA, fusion centers have (at least in theory) the capacity to develop intelligence packages focused on serious and recidivist offenders across the region, a capacity often beyond the abilities of local police departments. One of their guidelines is to "provide a multi-tiered awareness and educational program to implement intelligence-led policing and the development and sharing of information" (DoJ 2005: 7).

Distinctive Features

ILP arose as a response to the failure of the standard model of policing (Ratcliffe 2008a). The standard model emphasized random patrol, rapid uniformed response, deployment of officers to crime investigations, and reliance on enforcement and the legal system to reduce crime. The failure of the standard model in the 1960s and 1970s paved the way for a variety of responses, ultimately including ILP. In particular, ILP emphasizes the need to focus on serious and repeat offenders rather than a broad-brush reaction to crime events after they have occurred. In other words, targeting the right offenders will help police proactively address much of the crime these offenders commit. Empirical support for this assertion can be found in the classic studies of criminal careers. The Cambridge Study in Delinquent Development found that frequent offenders had not only a high offense rate but also a relatively low likelihood of desisting their criminal careers after each conviction. From the original group of nearly 10,000 boys studied by Wolfgang et al. (1972) in Philadelphia, only 627 (6.3 %) were identified as chronic offenders, but this group were responsible for 52 % of all offenses recorded against the entire cohort. The largest study, following various UK cohorts, found that by the time British men (born in 1953) reached their 31st birthday, 7 % would have at least six convictions against them, and this group accounted for 65 % of all convictions for their birth year.

A second component of ILP is the need for a business model for policing in order to influence the key resource allocation choices: "Intelligence process in law enforcement cannot be merely a voyage of discovery but demands predictability in the delivery and content of intelligence products for managers' decision-making" (Flood and Gaspar 2009: 60). This then requires knowledge of a further distinctive feature, a cross-jurisdictional strategic focus, as evidenced by the enthusiasm for ILP among many officers deployed to fusion centers.

With regard to Compstat, there are some organizational similarities between ILP and Compstat. For example, in both models leadership at the highest echelons, commitment to crime prevention, and accountability at mid-level command are important. However, Carter and Carter (2009) identify a number of distinguishing features, including ILP being multijurisdictional and threat driven (i.e., proactive and focused on anticipating criminality) whereas Compstat is intra-jurisdictional and crime driven, and ILP being strategic and more focused on criminal enterprises while Compstat is time sensitive and focused on street crime.

ILP is quite distinctive from community policing, and there is some discussion about where complimentary components exist (Carter 2009; Ratcliffe 2008a). Community policing places a high value on, and allocates considerable resources to, developing collaborative relationships with members of the community. Recent developments of ILP in the UK recognize the benefits of these collaborative relationships for holistic crime prevention, but as yet do not necessarily incorporate the development and sustenance of those collaborative relationships into ILP programs. There does appear to be agreement that in the area of counterterrorism, the relationships built up over years of community policing work are invaluable to information gathering. Yet where ILP seems a more centralized and objective method of resource allocation and priority triage, community policing emphasized local initiatives across a wide spectrum of policing areas and other (sometimes non-policing) concerns. ILP and problem-oriented policing share a number of objectives, and the overlap between these strategies is considerable, perhaps with the exclusion of the focus on serious and repeat offenders. It is a key component of ILP but one of an optional suite of solutions from the perspective of problem-oriented policing.

ILP is supported by a number of models, both conceptual and functional. The three-i model is a conceptual model that places the intelligence cycle into the business of policing. The three-i model of *interpret*, *influence*, and *impact* requires analysts and intelligence staff not only to interpret the criminal environment but also to influence the thinking of key decision-makers with their analysis. The final stage involves the decision-maker assigning tasks and allocating resources with the aim of having an impact on the criminal environment (for more information, see Ratcliffe 2008a). This idealized conceptual model is operationalized by the UK National Intelligence Model (NIM) (NCPE 2005). The NIM also has multiple levels to recognize the complexity of the modern criminal world. Level 1 deals with local criminality, level 2 addresses cross-border crime between agencies, and level 3 relates to serious and organized crime at national and international levels. The NIM has particular intelligence products for analysts to complete as well as roles for knowledge, system, source, and people assets. All of this is designed to address key ILP targets of individual criminals and criminal groups, crime and disorder hotspots, identified crime series, and the application of preventative measures (Flood and Gaspar 2009).

Critiques and Responses

As with any initiative, ILP is not without its critics. Writing during the early years of ILP development, James lamented the lack of research on which to base the hopes of improved police efficiency through the new (at least in 2003) intelligence focus and the National Intelligence Model (James 2003). He also raised concerns about the increased use of police informants and civil liberties issues with the potential for ILP to increase intrusion into the lives of citizens, a critique that has been raised by the American Civil Liberties Union (ACLU) (Guidetti and Martinelli 2009). Few doubt that work is always required to improve the management, handling, and operationalization of informants (Maguire and John 1995) and that effective mechanisms to manage intelligence files are essential (Carter 2009), but equally there is considerable sensitivity to this issue within local and state law enforcement in the USA, and a return to the COINTELPRO days of unfettered police surveillance seems unlikely (COINTELPRO is an acronym for the now discredited FBI political surveillance and disruption Counter Intelligence Program of the 1950s and 1960s).

From the perspective of 2003 and given journal publication times, one can estimate that James was writing at a time when the NIM had been unveiled no more than a couple of years before. As such, his measured concerns are largely an apt warning for more oversight of the police development of intelligence-led policing in the future. Indeed he notes at the time that “Forces favour strategies that appear to be rational, intelligence-led, based on empirical analyses and responsive to public need” (James 2003: 55). He continues to question the empirical basis for ILP, as did Ratcliffe a year earlier writing in regard to “the absence of a conceptual framework for intelligence-driven policing that is supported by an evidence-based theory. The dearth of any empirical and contextual research into the effectiveness of intelligence-led policing in either the academic or professional literature leaves the police with little option other than to experiment and do the best they can” (Ratcliffe 2002: 64).

Research on the police intelligence function and on police operations has made considerable progress in the last decade. As a result, the theoretical foundations of ILP have also developed over the last decade. An increased police willingness to allow researchers to evaluate their tactics and operations has dramatically improved knowledge on what works in policing, and numerous examples are relevant. For example, an evaluation of Operation Anchorage in Canberra, Australia, found that when police focused on burglary hotspots and used surveillance teams and investigative resources to target known repeat burglars, they were able to reduce burglaries in the city by 53 per week (from an average of around 150). Moreover, this reduction lasted for 45 weeks after the operation concluded, most likely due to the incarceration of a number of high-volume offenders (Makkai et al. 2004). At least 25 chronic offenders involved with an inner-city drug market in the Brightwood section of Indianapolis were targeted with a covert police interdiction and subsequent arrest swoop. The impact of removing these long-time repeat offenders was substantial, with a statistically significant reduction in calls for service for burglary, guns, violence, and robbery across the neighborhood for 2 years (Nunn et al. 2006).

A development of direct offender targeting is the pulling levers strategy. This approach focuses the attention of the criminal justice system on a few key chronic offenders in an attempt to encourage their deterrence from criminal activity. Evaluations have been generally positive, and this initiative is spreading across the USA and a number of other countries (Kennedy 2006, 2009). Pulling levers is described in depth elsewhere in this encyclopedia.

An additional component of ILP is the identification and targeting of police attention to crime hot spots. This aspect of ILP in enshrined as a key component of the UK National Intelligence Model. The targeting of resources to crime hot spots has a growing evidence base, in a subfield termed hot spot policing (Taylor et al. 2010; Weisburd and Braga 2006). The weight of evidence has some even calling in the USA for a national hot spot policing strategy (Mastrofski et al. 2010). The experimental basis for this claim continues to grow. For example, the Philadelphia Foot Patrol Experiment, a summer-long strategic management program to place foot patrol officers at violent crime hotspots, reduced violence by 23 % compared to equivalent control areas (Ratcliffe et al. 2011). And Operation Nine Connect, an extensive surveillance operation that resulted in over 100 arrests, effectively broke up a gang identified by intelligence analysts as an emerging threat and the most violent in New Jersey (Ratcliffe and Guidetti 2008).

So while at least initially in the development of ILP, some research suggested that police attempts to identify key offenders to target were fairly crude (Ratcliffe 2005; Townsley and Pease 2002), more recent evidence appears more hopeful that more “strategic, future oriented and targeted” (Maguire 2000: 315) tactics are emerging. A fruitful path would appear to blend all of these elements, with police identifying crime hot spots and then paying particular attention to the chronic offenders who commit offenses within that area.

Some analysts have critiqued ILP by comparison to the military arena, lamenting that “intelligence” within the policing environment is not well defined (a problem previously identified some years earlier in Ratcliffe 2008a) and that police intelligence analysts would be overstepping their role by making crime prevention recommendations to police managers. As noted by Mark Evans (Evans 2009: 194), bringing both his military and police experience to bear, “in general such thinking mainly reflects weak analytical product, a relatively unsophisticated decision-making model and a lack of understanding about the purpose and value of recommendations.” The military comparison is not relevant because while intelligence has long been a staple of the military and national security areas, the use of intelligence is a “contemporary phenomenon” in modern law enforcement Christopher and Cope (2009: 236). As a result, law enforcement commanders are less likely to be familiar with the use of intelligence resources to guide decision-making. Moreover, it is highly unlikely that they have received any training in this area, given the near total absence of courses available to police managers on problem-oriented policing, intelligence-led policing, or general crime prevention and reduction (Ratcliffe 2004). There is a distinction between an analyst making recommendations and a police officer making decisions (Cope 2004).

Sklansky (2011) worries that ILP (and predictive policing) represents a move toward a technologically driven new version of the "profession model" of policing and away from the philosophies and tenets of community policing. Like James (2003), the concerns are more abstract and serve as an abject warning to policing not to lose the community legitimacy gains of recent years in a wave of technological wizardry and policing by objectives. The challenge may be getting police commanders to embrace "advanced community policing" (Sklansky 2011: 12) anew, when it has not been shown to reduce crime measurably and when police chiefs are facing budgetary constraints and calls for greater efficiency and demonstrated productivity. The issue may rather be the mechanism by which community policing and ILP are integrated. As one commentator has written, "community policing has developed skills in many law enforcement officers that directly support new ILP responsibilities" (Carter 2009: 86).

Future Directions

The future of ILP rests with police leadership and their enthusiasm to adopt more objective, evidence-driven decision paths rather than value-based decision-making (Guidetti and Martinelli 2009). With an explosion of technology-based platforms to manage ever-greater volumes of information, there is a threat that fusion centers, real time crime centers, and intelligence analysts general fall prey to a "naïve empiricism" where information volume trumps quality resulting in an inability to distinguish between information and intelligence (Gill 2000: 211). In this future a balance has to be struck between new data sources ("new knowledge" such as crime mapping and statistical analysis) that tell us a great deal about *what* is going on in our world and policing's "old knowledge" (such as confidential informants and community intelligence) that can tell us *why* the patterns we see are emerging (Ratcliffe 2008b). In this role of answering *why*, it is increasingly clear to scholars of policing that criminal investigators have a role beyond that of case-by-case delivery of justice. Rather than simply investigating individual cases and arresting suspects, detectives have insight into broader patterns and emerging crime trends that can have wider strategic significance.

At present, while strategic assessments are key to the NIM and ILP, many analysts are not involved in strategic analysis tasks, a problem that exists not just in the UK but across most of the developed world. The same can be said for intelligence officers at fusion centers across the USA. Fusion centers are an indication of a more nodal form of governance of the security problem (Wood and Shearing 2007). Where before the police maintained a hegemonic status, now myriad agencies, both government and nongovernment, public and private, have a role in domestic crime control and security. A risk-based approach to the crime problem, such as intelligence-led policing, has a key role in this future. "The problem is that American law enforcement is neither structurally nor substantively ready to support the ILP infrastructure. Just like a building, the foundation must first be in place" (Carter 2009: 82). There is still therefore a pressing need for greater education of intelligence analysts and staff, and even more so of the police leaders that command them, before ILP can demonstrate its full potential.

Related Entries

- ▶ Information Technology and Police Work
- ▶ Predictive Policing

Recommended Reading and References

Audit Commission (1993) Helping with enquiries: tackling crime effectively (Management handbook). HMSO, London

BJA (2009) Navigating your agency's path to intelligence-led policing. Bureau of Justice Assistance, US Department of Justice, Washington, DC

Burch JH, Geraci MN (2011) Data-driven approaches to crime and traffic safety. The Police Chief

Carter DL (2004) Law enforcement intelligence: a guide for state, local, and tribal enforcement agencies (Report). Office of Community Oriented Policing Services, Washington, DC

Carter DL (2009) Law enforcement intelligence: a guide for state, local, and tribal enforcement agencies (Report), 2nd edn. Office of Community Oriented Policing Services, Washington, DC

Carter DL, Carter JG (2009) Intelligence-led policing: conceptual and functional considerations for public policy. Criminal Justice Policy Review 20(3):310–325

Christopher S, Cope N (2009) A practitioner's perspective of UK strategic intelligence. In: Ratcliffe JH (ed) Strategic thinking in criminal intelligence, 2nd edn. Federation Press, Sydney, pp 235–247

Cope N (2004) Intelligence led policing or policing led intelligence?: Integrating volume crime analysis into policing. Brit J Criminol 44(2):188–203

DoJ (2005) Fusion center guidelines (Version 1). Department of Justice, Washington, DC

Evans RM (2009) Influencing decision-makers with intelligence and analytical products. In: Ratcliffe JH (ed) Strategic thinking in criminal intelligence, 2nd edn. Federation Press, Sydney

Flood B, Gaspar R (2009) Strategic aspects of UK national intelligence model. In: Ratcliffe JH (ed) Strategic thinking in criminal intelligence, 2nd edn. Federation Press, Sydney

Gill P (2000) Rounding up the usual suspects? Developments in contemporary law enforcement intelligence. Ashgate, Aldershot

GIWG (2003) The national criminal intelligence sharing plan. Department of Justice (Global Intelligence Working Group), Washington, DC

Gottschalk P (2010) Policing organized crime: intelligence strategy implementation. CRC Press, Boca Raton, FL

Guidetti R, Martinelli TJ (2009) Intelligence-led policing—a strategic framework. Police Chief 76(10)

Harfield C, Harfield K (2008) Intelligence: investigation, community, and partnership. Oxford University Press, Oxford

Hess KM, Orthmann CH (2011) Introduction to law enforcement and criminal justice. Delmar Cengage Learning

HMIC (1997) Policing with intelligence (Thematic inspection on good practice). Her Majesty's Inspectorate of Constabulary, London

James A (2003) The advance of intelligence-led policing strategies: the emperor's new clothes? The Police Journal 76(1):45–59

Kennedy DM (2006) Old wine in new bottles: policing and the lessons of pulling levers. In: Weisburd D, Braga AA (eds) Police innovation: contrasting perspectives. Cambridge University Press, New York, pp 155–170

Kennedy DM (2009) Deterrence and crime prevention: reconsidering the prospect of sanction. Routledge, New York

Maguire M (2000) Policing by risks and targets: some dimensions and implications of intelligence-led crime control. Policing and Society 9(4):315–336

Maguire M, John T (1995) Intelligence, surveillance and informants: integrated approaches. Police Research Group: Crime Detection and Prevention Series, Paper 64:58

Makkai T, Ratcliffe JH, Veraar K, Collins L (2004) ACT recidivist offenders. Research and Public Policy Series 54:83

Mastrofski SD, Weisburd D, Braga AA (2010) Rethinking policing: the policy implications of hot spots of crime. In: Frost NA, Freilich JD, Clear TR (eds) Contemporary issues in criminal justice policy. Wadsworth, Cengage Learning, Belmont, CA, pp 251–264

NCIS (2000) The national intelligence model. National Criminal Intelligence Service, London

NCPE (2005) Guidance on the national intelligence model. National Centre for Policing Excellence on behalf of ACPO, Wyboston, UK

Nunn S, Quinet K, Rowe K, Christ D (2006) Interdiction day: covert surveillance operations, drugs, and serious crime in an inner-city neighborhood. Police Quarterly 9(1):73–99

Ratcliffe JH (2002) Intelligence-led policing and the problems of turning rhetoric into practice. Policing and Society 12(1):53–66

Ratcliffe JH (2004) Crime mapping and the training needs of law enforcement. Eur J Crim Pol Res 10(1):65–83

Ratcliffe JH (2005) The effectiveness of police intelligence management: a New Zealand case study. Police Practice and Research 6(5):435–451

Ratcliffe JH (2008a) Intelligence-led policing. Willan Publishing, Cullompton

Ratcliffe JH (2008b) Knowledge management challenges in the development of intelligence-led policing. In: Williamson T (ed) The handbook of knowledge-based policing: current conceptions and future directions. Wiley, Chichester, pp 205–220

Ratcliffe JH, Guidetti RA (2008) State police investigative structure and the adoption of intelligence-led policing. Policing: An International Journal of Police Strategies & Management 31(1):109–128

Ratcliffe JH, Taniguchi T, Groff EE, Wood J (2011) The Philadelphia foot patrol experiment: a randomized controlled trial of police patrol effectiveness in violent crime hotspots. Criminology 49(3):795–831

Sklansky DA (2011) The persistent pull of police professionalism (No. New Perspectives in Policing). Harvard Kennedy School, Cambridge, MA

Smith R (2009) A call for the integration of 'Biographical Intelligence' into the national intelligence model. Policing 3(2):191–199

Taylor B, Koper CS, Woods DJ (2010) A randomized controlled trial of different policing strategies at hot spots of violent crime. J Exp Criminol. doi:10.1007/s11292-010-9120-6

Townsley M, Pease K (2002) How efficiently can we target prolific offenders? Int J Police Sci Manag 4(4):323–331

Wardlaw G, Boughton J (2006) Intelligence-led policing: the AFP approach. In: Fleming J, Wood J (eds) Fighting crime together: the challenges of policing and security networks. University of New South Wales Press, Sydney, pp 133–149

Weisburd D, Braga AA (2006) Hot spots policing as a model for police innovation. In: Weisburd D, Braga AA (eds) Police innovation: contrasting perspectives. Cambridge University Press, New York, pp 225–244

Wolfgang ME, Figlio RM, Sellin T (1972) Delinquency in a birth cohort. University of Chicago Press, Chicago

Wood J, Shearing C (2007) Imagining security. Willan Publishing, Cullompton

Intensive Probation and Parole

Charlotte E. Gill
Center for Evidence-Based Crime Policy, George Mason University, Fairfax, VA, USA

Overview

Intensive supervision probation or parole (ISP) is a supervision approach that has emerged to deal with serious offenders in the community. ISP involves increasing the frequency of contact between probation officers and clients, placing offenders in small caseloads so that probation officers have more time to spend with their clients, and more recently a focus on identifying individual risk and need levels and directing probationers to appropriate services and treatment. ISP in the United States, and to a lesser extent in the United Kingdom, became popular as an intermediate sanction to provide a noncustodial alternative to dealing with high-risk offenders and has therefore traditionally focused on surveillance and control of offenders rather than treatment and service delivery. A substantial body of rigorous research studies has suggested that this approach may be ineffective and even harmful if used with lower-risk offenders. However, programs that emphasize a combination of surveillance and carefully-matched treatment, focused deterrence, and behavioral management approaches have shown more promise. ISP may also have suffered because it is difficult to devote sufficient attention to high risk and need offenders in understaffed and under-resourced probation agencies. Thus, ISP as an area of study and practice is currently in flux, but a number of promising future directions may prevent the approach from being discredited as a failed program. Note that in this entry, the terms “probation” and “parole” interchangeably mean a period of supervision or license in the community imposed post-release from prison or as a community-based sentence in its own right. In practice, there is often little difference in supervision practices for these two populations.

Fundamentals and Historical Development of Intensive Probation

In its most basic form, intensive supervision probation or parole (hereafter ISP) involves “tightening up” the supervision of convicted offenders on probation or supervised release from prison by reducing caseloads. This is usually achieved in one or both of two ways: reducing the size of probation officers’ caseloads so that they can spend more time focused on each client and/or enhancing reporting requirements such as the frequency and duration of contacts or increased drug testing requirements. Electronic monitoring, in the form of transmitting devices that notify authorities when an offender breaks a curfew or leaves a specified area or voice-assisted check-in devices, is also frequently added to ISP schemes. The belief behind most intensive probation programs is that if the most serious offenders are to be released back into the community, the more supervision and control they are subjected to, the better.

ISP programs developed over the past 50 years reflect theoretical notions of specific deterrence, surveillance and formal social control, and to a lesser extent informal social control. Offenders are offered the opportunity to remain in the community on the understanding that they are being constantly monitored, and the consequence of failure is the loss of liberty. While the formal social control and deterrence elements of ISP are most often discussed, ISP – as an alternative to incarceration – can also promote informal social control by allowing offenders to remain in the community. This can minimize the disruption of positive social networks and opportunities, such as access to prosocial role models within an offender's family and community, and conventional activities and treatment.

Indeed, the earliest ISP-like programs, which emerged in the United States in the 1960s, focused heavily on the idea of treatment and rehabilitation. They were guided by the belief that smaller caseloads allowed probation officers more time to help their clients improve their lives (Petersilia and Turner 1990). However, a series of field studies of these programs showed that these initiatives made very little impact on recidivism and in some cases even increased probation failure and technical violation rates. A potential explanation for the lack of effectiveness of these initiatives was that no attention was given in their development to how probation supervision activities best served the rehabilitative goal. Although probation officers had extra time to spend with clients, they did not know how to use that time to improve their clients' outcomes. Experimental testing of the programs focused almost exclusively on identifying the optimal client to probation officer ratio rather than supervision content (Clear and Hardyman 1990).

Although ISP does have some roots in traditional criminological theories, in its best-known form, it was primarily intended as a practical solution to a crisis in US corrections and a punitive alternative to imprisonment, rather than a program developed in accordance with theories of offending and behavior or a desire to better understand effective supervision practices. The lack of success of early ISP programs meant that they had fallen out of favor by the 1970s. At the same time, the rehabilitative ideal gave way to a retributive philosophy across the entire criminal justice arena in the United States and beyond, driven by a sharp rise in crime and the popular belief that "nothing worked" to reduce recidivism. In the United States, this convergence of factors triggered an exponential increase in the incarcerated population and the cost of corrections in general. This put substantial pressure on probation services. The probation population was growing in its own right, with an increasing number of serious and high-need offenders being granted parole or sentenced to probation outright. Probation officers' caseloads were becoming unmanageably large (Petersilia and Turner 1993).

Thus, by the 1980s, ISP had begun to make a comeback as a key strategy among other intermediate sanctions (alternatives to imprisonment) that sought to alleviate prison overcrowding and save money. At the same time, intermediate sanctions needed to maintain the appearance of being tough on offenders who would otherwise have been incarcerated and deter future offending through swift, certain punishment effected by close supervision (Petersilia and Turner 1990). Probation supervision in general has long suffered from image problems, particularly a public perception that it is a "soft" approach to dealing with serious offenders who are highly likely to recidivate. In the United States, for example, it is estimated that over 40 % of probationers and more than half of parolees do not successfully complete a term of supervision and that parole violators account for nearly 35 % of admissions to state prisons (Solomon et al. 2008). The new style of ISP was thus much more geared toward reducing strain on the criminal justice system and reassuring the public that serious offenders were not simply "getting away with it" by being released on probation (Petersilia and Turner 1993).

The "Georgia Model" and Experimental Research on ISP

The new generation of ISP programs that emerged in the 1980s thus focused on surveillance and control of offenders through small caseloads, frequent contacts, increased drug testing, and mandatory employment. Georgia was the first US state to implement a program that followed this model, with positive results. Participants in the ISP program had very low recidivism rates, maintained their employment, and paid probation fees that helped offset the cost of their supervision. As a result, the "Georgia model" was adopted elsewhere, with more mixed results. The interest in developing intermediate sanctions, coupled with uncertainty about the effectiveness of ISP, led the Bureau of Justice Assistance (BJA) of the United States Department of Justice to fund a multisite randomized controlled trial, conducted by the RAND Corporation (Petersilia and Turner 1993). The experiment remains one of the largest randomized controlled trials in criminal justice.

The BJA/RAND experiment was implemented in 14 locations. In 12 sites, ISP was compared to routine probation or parole supervision, which usually involves a brief meeting between the probation officers and client once a month, or existing intensive probation caseloads that were not as intense as the Georgia model. In an additional two sites, ISP was compared to incarceration. The exact nature of the program depended on the study site – each jurisdiction chose which components of the Georgia model to implement. However, there were a number of common features across all the evaluations. All included reducing probation officers' caseloads to around 25–30 clients per officer, usually compared to 100 or more in regular supervision caseloads. The frequency of contact was increased from a general standard of about once per month to at least once a week at first, with the frequency gradually decreasing in phases. All the programs also included drug testing and mandated employment.

Participants in the ISP experiments had to be adults. They were generally high-risk offenders, since the primary goal of the Georgia model ISP was to offer a punitive alternative to incarceration for offenders deemed too serious to be supervised in regular caseloads. However, offenders convicted of homicide, robbery, or sexual offenses were excluded from the experiment. Most participants were males in their late twenties and early thirties with extensive criminal histories, and a substantial proportion were drug dependent. Beyond these initial requirements, the study sites set their own additional eligibility criteria for their programs.

The methodology of the experiment involved random assignment to treatment and control conditions. Several waves of data collection followed, including a baseline assessment of demographic characteristics and criminal history, recording of supervision details and services received at 6–12 months, recording of recidivism at 12 months, and monthly drug test data. The researchers also collected data on cost-effectiveness and time at risk. Recidivism was based on official records of technical violations, arrests, convictions, and incarceration.

As with the ISP programs of the 1960s, the results of the BJA/RAND evaluations were disappointing. In general ISP did nothing to reduce recidivism, and because of the increased scrutiny to which probation clients were subjected, the rate of technical violations increased compared to standard practice. A program that had been intended to reduce the strain on the prison system actually resulted in more incarceration, as the greater level of surveillance and drug testing increased the likelihood of probation failure (Petersilia and Turner 1993). The rigorous research design used by the RAND team may have revealed limitations that were not uncovered by earlier evaluations: Petersilia and Turner (1993) note that comparison group participants in some of the more successful program evaluations of the early and mid 1980s were not well matched to probationers who experienced ISP.

Beyond Georgia: Intensive Probation in the 1990s and 2000s

Around the same time as the research evidence against ISP was growing, the "what works" movement was also gaining momentum. Researchers and practitioners began to move away from the idea that nothing worked and to acknowledge the limitations of existing evidence and program design. This led to a renewed interest in the factors that influence successful programming and outcomes. An early turning point in the "what works" movement was the development of the "principles of effective intervention" or PEI (Andrews et al. 1990) – the idea that correctional programs and treatment should be designed to be responsive to offenders' risk and need levels (also known as “risk-need-responsivity” or RNR). The “risk principle” element of the PEI, which has been empirically validated (Lowenkamp et al. 2006), suggests that the most treatment should be targeted only at the highest-risk individuals.

In probation research, the PEI have helped to refine supervision practices and better understand the factors that influence successful outcomes. Although participants in the BJA/RAND experiments were generally deemed to be high risk according to a range of characteristics including age at first conviction, offending history, current offense, and drug treatment needs, there was some variability across evaluation sites in the average risk level of participants, since sites generally set their own eligibility criteria (Petersilia and Turner 1993). Risk/needs assessment and treatment provision were not priorities of the experimental programs. Despite the increased contact with and control of their clients, probation officers did not in general focus on using that time to establish their clients’ risk and need levels using validated instruments and direct them to appropriate treatment (Latessa et al. 1998). However, the RAND evaluators noted a positive interaction between treatment provision, where it did occur, intensive supervision, and reduced recidivism and called for more research into effective treatment as part of probation supervision (Petersilia and Turner 1993).

Subsequently, and as a result of the general interest in “what works,” a new generation of ISP programs emerged in the 1990s. The programs still involved surveillance and control but increased the emphasis on making use of the extra contact time to establish probationers’ risk and need levels and direct them into appropriate treatment. Recently, several reviews and meta-analyses have supported the hypothesis that more modern ISP programs that adhere to the PEI and balance surveillance and control with a treatment component are more promising for reducing recidivism than purely surveillance-based programs (Aos et al. 2006; Drake 2011; Mackenzie 2006). Additional refinements include “community probation” efforts that draw upon additional law enforcement resources to supervise offenders. In these programs, probation officers work in tandem with police officers who are assigned to check in with ISP clients in their beat areas or develop closer relationships with smaller groups of probationers. Although ISP programs have been moving further away from pure surveillance and control, the purpose of most of these multiagency approaches has been surveillance-based, in particular, increasing contact between probation and law enforcement agencies to help police officers learn who are the most serious offenders in their jurisdictions and where they reside (Giblin 2002; Piquero 2003).

International Perspective

Intensive probation has most often been studied and discussed in the United States context. This is perhaps due to its recent history as a cheaper but still punitive alternative to incarceration during a time of significant increase in the prison population, which has occurred to a far greater extent in the United States than anywhere else in the world. Furthermore, post-release probation supervision and aftercare in general are less common in many other countries, especially in Europe (van Kalmthout and Durnescu 2008). However, several other countries have implemented intensive probation and monitoring schemes along similar lines to those seen in the United States. For example, the United Kingdom, whose prison population is the highest in Western Europe and has grown at a record pace, has a highly developed ISP system.

The earliest experimental ISP program in the United Kingdom was the IMPACT (Intensive Matched Probation and After-Care Treatment) scheme in the early 1970s. Surprisingly, the program recognized many of the factors that modern ISP programs in the United States focus on today: that high-risk offenders on probation have complex needs that contribute to their failure rates and probation officers spend little time dealing with them. Offenders were placed in small caseloads and met more frequently with their probation officers than control group participants. Supervision strategies were to be tailored to the needs and risk factors of individual offenders, combine treatment and support with control, and be delivered in the probation office and the client's own environment. The program showed some success for male participants, but not females (Folkard et al. 1976).

Intensive supervision reemerged in the last decade as a provision for serious juvenile offenders. The Intensive Supervision and Surveillance Programme (ISSP) was introduced in 2001 by the Youth Justice Board specifically to deal with prolific and persistent young offenders. It was not originally intended as an alternative to youth detention but was introduced to target a small proportion of serious juvenile offenders thought to be responsible for a substantial proportion of youth crime and is billed as the most rigorous noncustodial approach in the justice toolkit of England and Wales (a similar scheme, Intensive Monitoring and Supervision, exists in Scotland). The program was described as evidence based and involved a variety of approaches to supervision, including at least two surveillance checks per day, electronic tagging, tracking of movements by probation and police officers, and an intensive program of structured activities including education, training, and correctional treatment. However, the evaluation of the pilot programs showed that although the frequency of new offenses declined in the ISSP samples, it also declined for the comparison groups, and the failure rate for ISSP participants was very high. While it is to be expected that high-risk probationers will remain at high risk of reoffending, the increased intensity of supervision and programming did not appear to affect recidivism relative to standard practice.

Nonetheless, ISSP was rolled out nationally and spurred the development of additional intensive probation and community programs that combine frequent monitoring by probation and police officers with treatment and structured programming, including a similar strategy targeted at young adults, "Intensive Supervision and Monitoring" aimed at persistent offenders of all ages who commit volume property crime, and prolific and other priority offender schemes (Gray et al. 2005). In 2008, ISSP was incorporated into statute through the Criminal Justice and Immigration Act. It can now be attached to a community-based sentence for juvenile offenders called the Youth Rehabilitation Order or imposed as a condition of bail or post-release supervision for juveniles. Intensive supervision conditions can be implemented for 6 or 12 months, with the intensity gradually decreasing over time. Like the earlier version of the scheme, it is intended to offer a mixture of punishment and "positive opportunities," incorporates a set number of hours of supervision per week, and must include elements of education, training, or employment; restorative justice; programming to address offending behavior and increase interpersonal skills; and family support (Youth Justice Board 2009).

New Zealand also provides for a sentence of intensive supervision, which can be imposed along with a fine or community service scheme, under its Sentencing Act 2002. As in other locations, the sentence is intended for high-risk offenders and requires frequent contacts between the client and probation officer for an average of 12 months (up to 24 months). However, the New Zealand program differs from others in its emphasis on the rehabilitative nature of the sentence. The goal is to reintegrate the offender back into society, on the understanding that the level of need is sufficiently great that this could not be achieved through regular supervision or other programming. Probation officers are expected to direct their clients into appropriate treatment and assist them with addressing specific needs related to reintegration, such as employment and housing issues.

Key Issues and Controversies

Finding an ISP Model That "Works"

The key controversy surrounding ISP programs is whether or not they are effective. Both of the main waves of interest in ISP programs have been deemed a failure, especially given that ISP is one of the few aspects of criminal justice that has been subjected to multiple rigorous experimental tests. ISP was categorized as a program that "doesn't work" in the influential 1997 University of Maryland report to the United States Congress, *Preventing Crime: What*

Works, What Doesn't, What's Promising (Sherman et al. 1997). By the mid-1990s, interest in the surveillance/control model of probation supervision was largely dead.

However, as discussed earlier, the emergence of the PEI as the dominant paradigm in correctional treatment also drove the development of new ISP programs that focused more on risk, need, and appropriate services. Some of these more recent programs that attempt to balance surveillance with treatment have shown promise. MacKenzie (2006), in a detailed update to the University of Maryland report, lists intensive probation supervision with a treatment component as a "promising" correctional strategy, meaning that further rigorous research is needed but several studies have produced encouraging results. Most recently, a systematic review and meta-analysis of ISP programs confirmed a lack of positive evidence for surveillance-based ISP (Gill 2010). The meta-analysis did not reveal a clear positive effect of treatment-based programs but noted that the programs studied were highly variable and often did not include enough common components to draw meaningful conclusions about the effects of different program features.

There are a number of challenges in drawing conclusions from the overall body of evaluation research on ISP, several of which have already been noted. ISP programs were designed on the basic assumption that supervision is good for offenders in the community, so more of it must be better. However, there have been very few attempts to develop a theoretical basis for probation supervision or identify the qualities of effective supervision strategies (the latter point is discussed in more detail below). Thus, as several scholars have pointed out, there has been very little guidance on *how* to make supervision more intensive or which strategies work best. This may limit the utility of what the field can learn from ISP research.

It is also clear that much of the ISP research suffered from considerable implementation problems. Although surveillance-based programs were intended to be directed at high-risk offenders, definitions of "high" and "low" risk were not standardized across studies, risk levels varied, and the most serious offenders were often excluded from experiments for safety reasons even though it was not unusual for homicide, robbery, and sexual offenders to appear in probation and parole caseloads. This means that many of the programs studied did not necessarily comply with the PEI. Furthermore, the research studies often involved implementing innovative strategies in agencies that were already using standard probation procedures that had been in place for years and were struggling with limited resources (even when research funding was available). Thus, measurements of effect were sometimes based on an experimental supervision strategy that was not as intensive as intended. The planned ratios of intensity between treatment and control groups in the less successful ISP experiments were often similar because agencies found it difficult to meet their targets for caseload size or contact frequency in both the intensive and regular supervision caseloads. This problem also raises the question of whether ISP is a sustainable strategy in practice.

The uncertainty about the effectiveness of ISP has indicated a clear need for further development of programs to understand the complexities of the relationship between surveillance and treatment, probation officer and client, and the question of which supervision strategies should be employed during more frequent or intensive supervision sessions. In addition to the promising evidence for treatment-based and PEI/RNR-based ISP programs, two recent studies of modern ISP programs that have indicated relatively successful effects on recidivism are the Maryland Proactive Community Supervision (PCS) program (Taxman et al. 2006) and Hawaii's Opportunity Probation with Enforcement (HOPE) approach (Hawken and Kleiman 2009). Although the two programs follow different approaches, they reveal some promising strategies for further refining future programs.

The PCS approach focuses on service brokerage and individual case planning by probation officers in a proactive and "seamless" model of supervision and treatment. While PCS retains some of the characteristic surveillance and control elements of the classic ISP model, it emphasizes the “nature and intent” of the contacts between probation officers and their clients to allow for greater rapport building. This relationship is the foundation for leveraging services and treatments that will address the offender’s criminogenic needs and build prosocial networks and skills, ultimately facilitating behavioral change. A hallmark of the PCS approach is a “behavioral contract” or supervision plan that is responsive to the offender’s specific needs. This approach follows the principles of behavioral management, which include incentive and sanction schemes; a focus on criminogenic factors, which leads to tailored, but not mandated, treatment and services; and offender accountability. The probation officer acts as a behavioral manager, working with the offender to understand his or her criminogenic risk and needs and also encourage the offender to do so, develop the supervision plan, arrange services, and emphasize, incentivize, and encourage desistance from crime and achievement of goals.

HOPE, in contrast, strongly emphasizes enforcement, surveillance, and deterrence. However, while the potential deterrent effect of traditional ISP programs lay in the general threat of a return to prison if further offending or violations were detected through more frequent and intense surveillance, HOPE is highly specific and targeted. The main focus of HOPE is on increased drug testing. Participants in a randomized controlled trial of HOPE (Hawken and Kleiman 2009) were notified on a daily basis whether or not they had been selected for a random drug test. Test failures were met with swift, certain adjudication and shock incarceration. However, violators were not simply removed from probation and re-incarcerated – after spending a brief period (usually a weekend) in jail, they continued their supervision with probation officers trained in therapeutic techniques, and repeat violators were directed to specialized treatment as well as receiving traditional punishment in the form of jail stays of increasing duration. In addition, prior to being sent to jail, violators appeared before a judge who reminded them of their probation responsibilities. This hearing emphasized the desire of the criminal justice system to see offenders ultimately succeed in their probation term rather than fail.

The HOPE approach closely aligns with the emerging idea of “focused deterrence,” rather than the broader, classical specific deterrent qualities of traditional ISP. PCS embodies some of the same principles. Focused deterrence strategies have most commonly been used in policing. The framework involves implementing a menu of highly specific, targeted sanctions aimed at preventing offending through a multiagency approach that leverages community, police, and probation resources to directly communicate the reasons for sanctions to offenders. In the policing setting, this approach shows considerable promise (Braga and Weisburd 2012). Focused deterrence, incentives and sanctions, and the combination of treatment and accountability have also shown promise in other corrections settings, such as drug courts (Mitchell et al. 2012), and are consistent with the PEI and other bodies of research suggesting that interventions should be implemented at a high level of focus – whether at small places or with the small proportion of individuals who are high risk – and incorporate specific risk factors. These approaches also contrast with traditional surveillance-based ISP programs that leverage increased scrutiny and a higher probability of detecting a violation as threats to deter future offending but do not focus on the content of supervision, the role of the probation officer, or the relationship between officers and clients.

Another important difference between the HOPE and PCS approaches and traditional ISP programs is that the newer programs do not focus on specific alterations to probation intensity, such as finding the optimal frequency of contact or caseload size. As the earlier research suggested, a reduction in caseload size is not automatically accompanied by a guarantee that probation officers will actually be able to spend more time with their clients or offer more intensive treatment. As noted earlier, the realities of implementation often made it difficult for probation officers to sustain more frequent contact with ISP clients in the RAND research, and resource limitations meant that caseload sizes often crept back up to pre-evaluation levels. Studies that have focused on improving the content of supervision as well as the levels of surveillance, for offenders who are at greatest need of supervision, appear to show the most promise.

Understanding the Content of Supervision

The question of what probation officers actually do when required to supervise offenders more closely is largely unanswered by current research. It is clear from the previous discussions that a lack of focus on the content of supervision has contributed to the failure of both the earliest treatment-based ISP programs and the traditional surveillance-based programs (Clear and Hardyman 1990). On the other hand, more recent programs that have set out specific guidance for how probation officers should spend their time with clients have shown more promising results.

One of the few qualitative studies that have attempted to look inside this "black box" of supervision was by Bonta and colleagues (2008). They examined recordings of over 60 interviews between officers and probationers. They discovered that the probation officers in their sample spent too much time focusing on enforcement and insufficient time on service delivery. The officers did not appear to account for the PEI or offenders' criminogenic needs and were not equipped with the necessary tools to help clients change their behavior – the elements that the recent research on ISP programs suggests are most crucial to successful outcomes. While this study did not specifically examine ISP programs, the development of ISP research to date has indicated that ISP is not successful when it focuses on enforcement through increased surveillance and control and is most successful when it adheres to the PEI and incorporates behavioral management principles.

Shifting the focus of ISP away from surveillance and enforcement toward directing offenders into appropriate services may also help to reduce the "surveillance artifact" problem. Research on ISP shows that participants tend to be more likely to violate their probation. However, this does not necessarily mean that they are less likely to comply. The increased level of surveillance and drug testing to which they are subjected means that probation officers have more opportunities to detect a violation. This is the "surveillance artifact." It presents a challenge for evaluations of ISP because one of the goals of the program has been to keep offenders out of prison, yet the higher probability of failure does not necessarily mean that ISP makes offenders' behavior worse. Pearson (1988) suggested that in some surveillance-based programs in the 1980s, probation officers would "go out in the field actively looking for violations." This is in contrast to PCS and HOPE, which focus more on the content of supervision and responses to violations. It is clear that the future of efforts to improve intensive probation lie in focusing on how well the programs adhere to the PEI and help to manage clients' behavior, and moving away from the idea that there is an optimal caseload size or ratio of clients to probation officers that allows for more intensive intervention without clarifying what that intervention should involve.

Low-Intensity Supervision

The principles of effective intervention suggest that intensive supervision, as well as treatment, should be reserved for the highest-risk offenders. More recent attempts at developing ISP programs have taken this into account and have included a greater emphasis on treatment and service brokerage by probation officers in addition to smaller caseloads and more frequent contact. However, even when ISP programs are successful, it can be difficult to implement them well in practice. A key challenge is ensuring that probationers actually receive *more intensive* supervision. In reality, many probation agencies are understaffed and lack financial resources to hire more officers, and intensive probation clients often end up receiving no more supervision than their traditional supervision counterparts, especially when trial programs (and the associated money and research support) come to an end or are rolled out across the whole agency as standard practice. Low-intensity supervision – allowing the lowest-risk offenders to receive minimal supervision in large caseloads – is an emerging method for dealing with resource issues and allowing existing staff to be reallocated to concentrate on higher-risk clients who pose a greater public safety risk. However, given the prevailing attitude in ISP research that large caseloads have been portrayed as detrimental to crime prevention, this approach raises the question of whether it is "safe" to reduce the level of supervision given to certain offenders in terms of controlling recidivism.

Probation agencies have recently begun experimenting with reducing the intensity of supervision for those at the lowest risk of recidivism. New York City's probation department initiated these efforts in the mid-2000s by pilot testing an electronic kiosk reporting system for a substantial proportion of its lowest-risk caseload. Probationers checked in regularly using an ATM-style device and could be compelled to see a probation officer if adverse circumstances arose. Although no formal evaluation of the kiosk approach was conducted, an assessment of two-year rearrest rates for probationers assigned to the kiosks showed a slight decline in crime, suggesting that reducing supervision did not affect their outcomes. At the same time, high-risk probationers received more frequent supervision and their rearrest rates also declined (Wilson et al. 2007). Similarly, a study in Oregon found that reducing supervision for low-risk probationers in order to free up resources for intensive supervision of high-risk offenders showed an overall decline in crime for the local probation population compared to the previous "one size fits all" approach.

The most rigorous test of low-intensity probation to date is a randomized controlled trial implemented in the Philadelphia Adult Probation and Parole Department, in which a risk prediction model was used to designate 60 % of the agency's caseload as low risk, with the ultimate goal of increasing resources and supervision for 10 % of the caseload predicted to be highest risk. Low-risk offenders were randomly allocated to low-intensity supervision, involving a face-to-face or telephone check-in every 3 months in a caseload of 400 per officer, or a control group receiving regular supervision, involving one face-to-face meeting with their probation officer per month in a caseload of 150–200. The experiment revealed no significant differences between the treatment and control groups after 1 or 2 years, suggesting that low-intensity probation is a safe strategy that does not lead to increased recidivism for low-risk offenders (Barnes et al. 2010; Gill 2010).

Further development of the low-intensity supervision concept will likely be crucial to the future of intensive probation. Not only does it allow resources to be allocated more efficiently, potentially giving probation officers working with the highest-risk clients more time to focus on risk and need assessment and service delivery, but it could also play a key role in enhancing the effectiveness of ISP by moving it further into line with the PEI. As discussed above, many of the ISP strategies tested in the 1980s were not always targeted at probationers at highest risk of recidivism and in greatest need of treatment. Furthermore, some of those evaluations found that ISP was especially ineffective – and in some cases even backfired – for the lowest-risk offenders (Hanley 2006). This is consistent with broader research on the PEI that suggests recidivism among low-risk offenders tends to increase when they are subjected to more intensive programming and that they can safely be assigned to the least restrictive settings (Lowenkamp and Latessa 2004). Research also suggests that low-level delinquents are more susceptible to the influence of more delinquent peers, suggesting that they may be better served by spending more time away from the probation office where they will mix with other offenders and focusing on building positive social relationships and opportunities away from criminal networks (Barnes et al. 2010; Lowenkamp et al. 2006; Lowenkamp and Latessa 2004). In theory, by reducing the time probation officers spend dealing with low-risk clients who are unlikely to recidivate again regardless of the level of intervention, ISP for the highest-risk offenders can be properly resourced and subsequently approved.

Conclusions and Future Directions

Understanding effective practices in probation supervision is important not just to alleviate concerns about probation as a "soft option" and to efficiently use limited resources. It is crucial to understand what works in probation supervision in its own right, not just as a background to leveraging treatment and other programs. Without a strong supervisory foundation, probation officers may not have the opportunity to learn how best to target programming most effectively. Further, when agencies lack funding and caseloads are large (often 150–200), brief supervision meetings may be the only interaction that occurs between the client and the agency. Outside the ISP context, supervision levels can vary widely, from weekly or twice-weekly meetings for high-risk or delinquent probationers to telephone reporting for those near the end of their sentences. In especially busy agencies, clients sometimes simply mail in a card periodically to confirm their contact details (Petersilia and Turner 1993).

The controversies surrounding ISP and the research on treatment effectiveness strongly indicate that probation is most effective when related to the client's risk and need, rather than the agency's operational capabilities. Of course, operational capabilities cannot be disregarded, but finding a balance between the two is crucial. Programs like HOPE appear to be having some success in that regard by minimizing the time probation officers spend on surveillance and control unless the client's behavior or need warrants it. The probation population in the United States has only recently stabilized after growing by more than half a million between 2000 and 2008 alone. Currently 1 in 48 adults is under probation or parole supervision (Glaze and Bonczar 2011). With nearly five million individuals under community supervision, probation agencies must identify efficient and effective ways to restructure supervision and focus proven strategies where they are needed the most.

The current state of research, knowledge, and practice concerning intensive probation highlights a number of future directions for researchers and practitioners seeking to study and improve ISP programming. There is still a need for increased focus on the content of supervision to understand which supervision strategies are most effective for high-risk offenders. It is clear that when the development of ISP programs has focused on what has been termed the "search for the magic number" – trying to discover the optimal caseload size or ratio of clients to probation officers – outcomes for offenders have not been favorable. While intermediate sanctions like ISP are an important strategy for reducing the financial burdens in other areas of the criminal justice system, the ultimate goal should be to protect public safety and ensure that offenders receive a system response that balances punishment with effective intervention, treatment, and assistance to prevent future recidivism. Thus, the content and goals of supervision strategies should be central to the future development of intensive probation. A number of researchers and practitioners have already recognized this and have developed pilot programs centered around focused deterrence and behavioral management approaches to supervision and treatment that are showing promise. These approaches should be further refined and tested using rigorous evaluation methods including randomized trials (a multisite replication of the HOPE experiment is already under way, funded by the United States National Institute of Justice and Bureau of Justice Assistance).

The discussion of low-intensity supervision suggests that intensive probation programs could be most successful when they are implemented as part of a whole-agency approach to restructuring supervision and resource allocation around offender risk and need, rather than as an add-on to regular probation. As probation agencies continue to grapple with increasing caseloads and dwindling budgets, innovative strategies will be needed to maintain effective supervision. Most probation agencies typically follow a "one size fits all" approach to probation supervision and do not vary resources according to risk, largely because typical caseloads are so large that probation officers cannot devote extra time to the supervision of higher-risk or higher-need offenders.

However, this might be achievable when agencies adhere to the principles of effective intervention and classify their entire caseload according to risk. Research and practice in this context should therefore continue to focus on developing and refining reliable risk prediction instruments at both ends of the spectrum.

Finally, for criminologists interested in intensive probation, there remains a need to clarify the theoretical basis for ISP programs. While a number of theories have been suggested for how ISP should work, programs have typically been developed in purely practical settings. As the principles of effective intervention show, returning to theoretical principles and causal mechanisms can help to identify common features of effective programs and inform the development of future programs, especially when there is uncertainty about the effectiveness of an approach. Future research in this area should focus on identifying and testing the mechanisms that underlie successful ISP programs and differentiate them from those that have been less effective and seek to clarify the role and need for post-offense or post-release intensive supervision. Does ISP serve a purpose in itself over and above regular supervision, or is it simply a foundation for directing offenders to appropriate treatment?

Related Entries

▶ Actualizing Risk-Need-Responsivity
▶ Behavioral Management in Probation
▶ Effective Supervision Principles for Probation and Parole
▶ Electronic Monitoring
▶ Focused Deterrence and “Pulling Levers”
▶ History of Probation and Parole in the United States
▶ Parole and Prisoner Reentry in the United States
▶ Probation and Parole Practices
▶ Probation Officer Decision-Making
▶ Risk Assessment, Classification, and Prediction
▶ Sex Offenders and Criminal Policy

Recommended Reading and References

Andrews DA, Bonta J, Hoge RD (1990) Classification for effective rehabilitation: rediscovering psychology. Crim Justice Behav 17(1):19–52. doi:10.1177/0093854890017001004

Aos S, Miller M, Drake E (2006) Evidence-based public policy options to reduce future prison construction, criminal justice costs, and crime rates (No. 06-10-1201). Washington State Institute for Public Policy, Olympia, Retrieved from http://www.wsipp.wa.gov/rptfiles/06-10-1201.pdf

Barnes GC, Ahlman LC, Gill C, Sherman LW, Kurtz E, Malvestuto R (2010) Low-intensity community supervision for low-risk offenders: a randomized, controlled trial. J Exp Criminol 6(2):159–189. doi:10.1007/s11292-010-9094-4

Bonta J, Rugge T, Scott T-L, Bourgon G, Yessine AK (2008) Exploring the black box of community supervision. J Offender Rehabil 47(3):248–270. doi:10.1080/10509670802134085

Braga AA, Weisburd DL (2012) The effects of focused deterrence strategies on crime: a systematic review and meta-analysis of the empirical evidence. J Res Crime Delinq 49(3):323–358. doi:10.1177/0022427811419368

Clear TR, Hardyman PL (1990) The new intensive supervision movement. Crime Delinq 36(1):42–60. doi:10.1177/0011128790036001004

Drake EK (2011) “What works” in community supervision: interim report (No. 11-12-1201). Washington state institute for public policy. Washington State Institute for Public Policy, Olympia, Retrieved from http://www.wsipp.wa.gov/rptfiles/11-12-1201.pdf

Folkard MS, Smith DE, Smith DD (1976) IMPACT intensive matched probation and after-care treatment. Volume II: The results of the experiment (HORS 36). Home Office Research, Development and Statistics Directorate, London, Retrieved from http://www.homeoffice.gov.uk/rds/pdfs05/hors36.pdf

Giblin MJ (2002) Using police officers to enhance the supervision of juvenile probationers: an evaluation of the anchorage CAN program. Crime Delinq 48(1):116–137. doi:10.1177/0011128702048001005

Gill C (2010) The effects of sanction intensity on criminal conduct: A randomized low-intensity probation experiment. Doctoral dissertation. Department of Criminology, University of Pennsylvania, Philadelphia, Retrieved from http://repository.upenn.edu/edissertations/121/

Glaze LE, Bonczar TP (2011) Probation and parole in the united states, 2010 (NCJ 236019). United States Department of Justice, Bureau of Justice Statistics, Washington, DC, Retrieved from http://bjs.ojp.usdoj.gov/content/pub/pdf/ppus10.pdf

Gray E, Taylor E, Roberts C, Merrington S, Fernandez R, Moore R (2005) Intensive supervision and surveillance

programme: the final report. Youth Justice Board for England and Wales, Oxford, UK
Hanley D (2006) Appropriate services: examining the case classification principle. J Offender Rehabil 42(4):1–22. doi:10.1300/J076v42n04_01
Hawken A, Kleiman M (2009) Managing drug involved probationers with swift and certain sanctions: evaluating Hawaii's HOPE (NCJ 229023). U.S. Department of Justice, National Institute of Justice, Washington, DC, Retrieved from http://www.ncjrs.gov/pdffiles1/nij/grants/229023.pdf
Latessa E, Travis L, Fulton B, Stichman A (1998) Evaluating the prototypical ISP: final report. University of Cincinnati, Division of Criminal Justice and American Probation and Parole Association, Cincinnati, Retrieved from http://www.uc.edu/ccjr/Reports/ProjectReports/ISP.pdf
Lowenkamp CT, Latessa EJ (2004) Understanding the risk principle: how and why correctional interventions can harm low-risk offenders. Topics in community corrections. National Institute of Corrections, Washington, DC, Retrieved from http://www.uc.edu/content/dam/uc/ccjr/docs/articles/ticc04_final_complete.pdf
Lowenkamp CT, Latessa EJ, Holsinger AM (2006) The risk principle in action: what have we learned from 13,676 offenders and 97 correctional programs? Crime Delinq 52(1):77–93. doi:10.1177/0011128705281747
Mackenzie DL (2006) Reducing the criminal activities of known offenders and delinquents: crime prevention in the courts and corrections. In: Sherman LW, Farrington DP, Welsh BC, Mackenzie DL (eds) Evidence-based crime prevention (revised edition. pp. 330–404). Routledge, New York
Mitchell O, Wilson DB, Eggers A, Mackenzie DL (2012) Drug courts' effects on criminal offending for juveniles and adults. Campbell Systematic Reviews, 2012 (4). Retrieved from http://campbellcollaboration.org/lib/project/74/
Pearson FS (1988) Evaluation of New Jersey's intensive supervision program. Crime Delinq 34(4):437–448. doi:10.1177/0011128788034004005
Petersilia J, Turner S (1990) Intensive supervision for high-risk probationers: findings from three California experiments (No. R-3936-NIJ/BJA). RAND corporation. RAND Corporation, Santa Monica, Retrieved from http://www.rand.org/pubs/reports/2007/R3936.pdf
Petersilia J, Turner S (1993) Intensive probation and parole. Crime Justice 17:281–335
Piquero NL (2003) A recidivism analysis of Maryland's community probation program. J Crim Justice 31(4):295–307. doi:10.1016/S0047-2352(03)00024-2
Sherman LW, Gottfredson D, Mackenzie D, Eck J, Reuter P, Bushway S (1997) Preventing crime: what works, what doesn't, what's promising. Ncjrs.gov. United States Department of Justice, National Institute of Justice, Washington, DC
Solomon AL, Jannetta J, Elderbroom B, Winterfield L, Osborne JWL, Burke P, Stroker RP et al (2008) Putting public safety first: 13 strategies for successful supervision and reentry (public safety policy brief No. 7). The Pew Charitable Trusts, Washington, DC, Retrieved from http://www.urban.org/url.cfm?ID=411800
Taxman FS, Yancey C, Bilanin JE (2006) Proactive community supervision in Maryland: changing offender outcomes. Maryland division of parole and probation. Maryland Department of Public Safety and Correctional Services, Towson, Retrieved from http://www.dpscs.state.md.us/publicinfo/publications/pdfs/PCS_Evaluation_Feb06.pdf
van Kalmthout AM, Durnescu I (2008) European probation service systems: a comparative overview. In: Probation in Europe 2008. European Council on Probation. Retrieved from http://www.cepprobation.org/uploaded_files/1_Chapter_1_Comparative_overview.pdf
Wilson JA, Naro W, Austin JF (2007) Innovations in probation: assessing New York City's automated reporting system. The JFA Institute, Washington, DC, Retrieved from http://www.nyc.gov/html/prob/downloads/pdf/kiosk_report_2007.pdf
Youth Justice Board (2009) Youth rehabilitation order with intensive supervision and surveillance (ISS). Youth Justice Board for England and Wales, London, Retrieved from http://www.justice.gov.uk/downloads/youth-justice/reducing-re-offending/YROwithISSOperationalGuidance.pdf/

Interactional Theory of Delinquency

Terence P. Thornberry
University of Maryland, College Park, MD, USA

Synonyms

Transactional theory

Overview

Interactional theory offers a broad explanation for the causes and consequences of involvement in antisocial behavior. When first proposed by Thornberry in 1987, it primarily focused on delinquency and drug use during adolescence and early adulthood. The theory proposed that delinquent behavior was caused by weak social bonds and involvement in delinquent networks but that delinquency also had feedback effects to further weaken prosocial bonds and further embed the individual in deviant networks and belief systems. Prolonged and serious involvement in antisocial behavior gradually evolved over the life course as a function of these reciprocal processes. The theory was subsequently expanded in two major directions. First, Thornberry and Krohn (2001, 2005) added a fuller developmental, life course perspective to account for involvement in antisocial behavior across the life span. The theory offered explanations both for the onset of offending, from childhood through adulthood, as well as for continuity and change in offending as a person ages and develops. Second, Thornberry (2005, 2009) extended the logic of interactional theory to the explanation of intergenerational continuity in antisocial behavior. One of the central premises of this theoretical orientation is that involvement in antisocial behavior has consequences for a person's later development. Those consequences, such as difficulty in the transition from adolescence to adulthood, interfere with patterns of family formation, and with the quality of parenting behaviors, helping to account for why parental involvement in delinquency is a significant risk factor for their children's involvement in delinquency.

The Original Statement

Interactional theory starts with the basic assumptions and premises of a social control theory of delinquency. Youth who are strongly bonded to conventional society are unlikely to have the behavioral freedom to engage in delinquency, drug use, and other problem behaviors. In particular, youth who are strongly attached to their parents and family, who are committed to conventional institutions like school, and who have strong beliefs in conventional values are thereby constrained from engaging in antisocial behaviors. These elements of the bond to society are likely to prevent any prolonged and serious involvement in antisocial behavior. In contrast, youth who have weak social bonds have more behavioral freedom and a higher probability of engaging in antisocial behavior.

Beginning with this basic perspective, Thornberry elaborated upon it to develop a more comprehensive and dynamic explanation for delinquent behavior during adolescence. Theoretical elaboration (Thornberry 1989) is an approach to theory construction that starts with a core theoretical orientation and then elaborates upon it to develop a broader theoretical explanation. There is no specific intent to integrate different theoretical models into a single theory, but insights are taken from other theories and from empirical investigations to develop as comprehensive an explanation as possible. One defining aspect of the theoretical elaboration that is evident in interactional theory is that the causal processes that generate delinquency vary developmentally. For example, the factors that cause offending for a 12-year-old are likely to be different from those that cause offending for an 18-year-old or a 25-year-old. Underlying processes – such as the strength of bonds to conventional society – may be the same, but the way they are manifested and the specific social institutions that are involved vary. Accordingly, Thornberry (1987) presented several developmentally specific causal diagrams of interactional theory. To summarize the logic of the theory, Fig. 1 presents the model for middle adolescence. It should be noted that the general concepts presented in the figure are proxies for more specific causal processes within each of these developmental domains. For example, "attachment to parents includes the affective relationship between parent and child, communication patterns, parenting skills such as monitoring and discipline, parent-child conflict, and the like" (Thornberry 1987, p. 866). Similar extensions can be drawn for the other concepts. In the interest of space, however, we will only discuss the more general concepts here.

Two core aspects of interactional theory are immediately evident in Fig. 1. The first concerns the inadequacy of a basic social control theory to account for delinquent behavior. Weak bonds to conventional society are viewed as creating increased behavioral freedom; that is, they allow individuals to deviate from conventional activities, but they do not necessarily lead

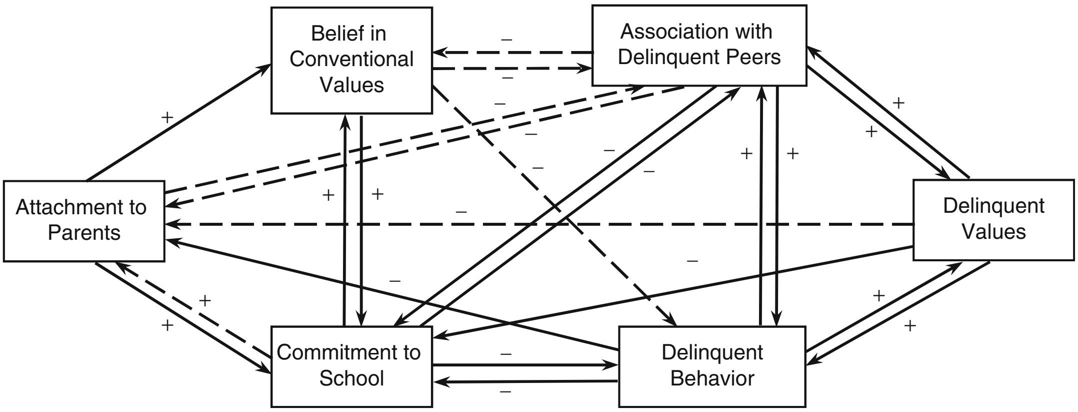

Interactional Theory of Delinquency, Fig. 1 A reciprocal model of delinquent involvement at middle adolescence. Note: Solid lines indicate stronger effects; dashed lines indicate weaker effects

directly to involvement in antisocial behavior. For that to occur, a learning environment is required that channels the behavioral freedom specifically into delinquent behavior, as opposed to other maladaptive responses such as internalizing problems, retreatist behavior, alienation, and withdrawal. This learning environment is represented by two key concepts – association with delinquent peers and delinquent values. Associating with others who engage in delinquent behavior provides role models for imitation and social reinforcements for engaging in delinquent behavior. It also increases the likelihood of spending unstructured and unsupervised time with peers which is also related to delinquent behavior (Osgood and Anderson 2004). In addition, beliefs that it is acceptable to engage in delinquent behavior in certain circumstances, or at least that it is not morally wrong to engage in those behaviors, also increase the likelihood of delinquent behavior.

The second core aspect of interactional theory that is evident in Fig. 1 is its reciprocal and dynamic nature. First, these central causes of delinquency are likely to become reciprocally related over time. For example, youth who are poorly attached to their parents are likely to be less engaged and successful in school, and in turn, youth who fail in school are likely to grow more distant from their parents. Second, delinquent behavior is not viewed as a mere outcome of earlier factors, such as weak commitment to school or associations with delinquent peers, as it is in many traditional theories. In contrast, delinquency is viewed as embedded in a set of mutually reinforcing causal relationships that develop over time and that create trajectories toward or away from prolonged involvement in delinquency. For example, associating with delinquent peers, via learning processes and changes in routine activities, increases the chances of delinquent behavior. But, the more the individual engages in delinquent behavior, the more likely they are to associate with fellow delinquents in part because of social selection processes and in part because delinquent youth are often rejected from prosocial peer groups. Similarly, weak social bonds such as attachment to parents increase the likelihood of delinquent behavior as youth are less constrained to follow the wishes and values of their parents, but involvement in delinquency is likely to further erode the bond to parents since it is typically antithetical to parental wishes and desires.

Thus, delinquent behavior is not merely the outcome of earlier risk factors and causal processes. It is part and parcel of its own development. Weak bonds to conventional society allow for the emergence of delinquent behavior, values, and associations. But as individuals become enmeshed in these delinquent influences, it is likely that their bonds to the conventional world will become further eroded. As a result, alienation from parents and family is likely to increase as does school disengagement. These reciprocal relationships are reflected in the bidirectional arrows presented in Fig. 1. Some are stronger during middle adolescence than others, as reflected in the solid lines, but the orientation of the theory as a state-dependence model (Nagin and Paternoster 1991) is evident. Past delinquent behavior is correlated with future delinquent behavior because of the consequences – the erosion of bonds to conventional society and embeddedness in deviant networks – that it generates.

These reciprocal relationships create behavioral trajectories that lead to different patterns of offending over the life course. At one extreme are youth who are strongly bonded to conventional institutions of family and school and are not embedded in delinquent networks. For them prolonged involvement in delinquency is highly unlikely. At the other extreme are youth from dysfunctional families who have weak bonds with conventional institutions and abundant sources of delinquent peer networks. For these youth chronic, serious offending is much more likely. And, of course, there can be a variety of offending trajectories between these extremes. Moreover, the starting values that give rise to these different trajectories are systematically related to structural variables. Two of the most central are social class and neighborhood composition. Children from impoverished families, especially those living in areas of concentrated disadvantage, suffer from a number of stressors that increase the likelihood of weak bonds and delinquent embeddedness as compared to children from families living in more advantaged neighborhoods with more resources and better schools.

Life Course Extensions

Thornberry and Krohn (2001, 2005) extended interactional theory to account for offending over longer portions of the life course. Antisocial behavior can emerge at virtually any age, from toddlerhood to adulthood. Childhood onset of offending is accounted for by three broad factors – individual characteristics, ineffective parenting, and social disadvantage. Children who exhibit early involvement in antisocial behavior are likely to suffer from a variety of negative developmental traits and neuropsychological deficits, such as impulsivity, risk-taking, and negative emotionality. They are more likely to be born to families experiencing severe structural adversity such as chronic poverty, unemployment, and residence in areas of concentrated poverty. They are also more likely to experience less effective parenting styles such as low affective ties with parents, ineffective monitoring and discipline, and, at the extreme, child maltreatment. More importantly, very early onset of antisocial behavior is likely to be brought about by the intense coupling and interaction of these influences. For example, bidirectional influences between the child's temperament and the parent's child management style have been observed as early as toddlerhood (Shaw and Bell 1993). Also, social disadvantage contributes to both parenting deficits and to negative temperamental qualities in the child.

Antisocial behavior that emerges during late childhood and early adolescence is more likely to be caused by environmental influences. Those influences, which were largely described above and in the original statement of interactional theory (Thornberry 1987), include the weakening of social bonds, associations with delinquent peers, the formation of delinquent beliefs, and the reciprocal influences that develop among them over time.

Thornberry and Krohn (2005) also account for unusually late onset – late bloomers or offenders whose criminal careers do not begin until late adolescence/early adulthood, after the normative peak of onset as described by the age-crime curve. Interactional theory hypothesizes that late bloomers have a number of individual deficits (e.g., internalizing problems like anxiety and depression) as well as reduced human capital (e.g., lower intelligence and lower academic competence). At earlier ages, however, these risk factors for delinquency are offset by compensating factors such as a supportive family, residence in more organized communities, and strong school environments. During emerging adulthood the individual leaves the protective environment of the family and school, begins to seek his or her own identity, and enters adult roles such as employee, partner, and parent. As the prosocial cocoon of adolescence is lost, the individual deficits begin to emerge, creating stress which is likely to be linked to late onset of antisocial behavior such as drug use and criminality.

Continuity and Change

In addition to issues related to onset, life course theories are also interested in the twin topics of continuity and change. Continuity in offending is somewhat more likely for those who start early than for those who start later. Unlike taxonomic theories (e.g., Moffitt 1993), however, interactional theory assumes that continuity is not tightly linked to age of onset. Although there is a positive correlation between onset and persistence, the magnitude of the correlation is moderate (Krohn et al. 2001). Regardless of when one's delinquent career begins, it is possible for that career to continue. Continuity is accounted for by two general processes. The first, which is much more important for those with an early onset of offending, is the stability of factors associated with delinquency. Negative temperamental qualities, ineffective parenting styles, and extreme social disadvantage are all relatively stable over time. Therefore, these factors are likely to continue causing involvement in delinquency at later ages offering a partial explanation for persistence. The second general process, which is important for all offenders regardless of their age of onset, stems from the bidirectional relationships within which delinquency is embedded. As we saw earlier, delinquency has feedback effects that further intensify the factors associated with its causality. Delinquency, especially serious delinquency, erodes social bonds with family, school, and prosocial peers and embeds the individual in deviant social networks and belief systems. As those reciprocal relations intensify, persistent involvement in delinquency becomes more likely.

Although some offenders do indeed persist in antisocial behavior throughout long portions of their lives, most offenders stop or desist from involvement in crime. Involvement in delinquent and criminal behavior peaks during adolescence, and desistance from crime typically unfolds during the transition from adolescence to adulthood. This transition affords new opportunities for social bonds that can reverse previous patterns of behavior. In particular, the major developmental challenges of this developmental stage are to complete the development of autonomy and independence from the family of origin, to establish one's own family of procreation, and to establish stable patterns of work. Successfully doing so is likely to lead to reductions in involvement in antisocial behavior. First, all of these transitions increase bonds to conventional society – especially the roles of partner, parent, and worker. Second, as the individual becomes enmeshed in the conventional roles, they are less likely to associate with deviant peers and to engage in risky time with friends; friendships gradually change to prosocial networks organized around family and work. All of these changes are likely to reduce involvement in antisocial behavior, and at this point, the reciprocal influences begin to work in reverse – increasing social bonds enhances prosocial behavior patterns, and those behavior patterns feed back to increase prosocial bonds. Of course, smooth and successful transitions to adulthood are not available to all. For example, the more extensive an individual's delinquent career, the harder it is to move from antisocial to prosocial behavior patterns. Nevertheless, most youth come through adolescence with enough human and social capital to give them entrée to successful transitions to adulthood. It may be through a "good marriage" or a set of skills that opens employment opportunities, but once those transitions are initiated, the movement away from antisocial behavior is enhanced.

Intergenerational Extensions

Linked lives, a basic premise of the life course perspective, argue that an individual's life course does not unfold independently of that of others. Quite the contrary, the life courses of intimate relations – husbands and wives, best friends, and parents and children – become interwoven over time and mutually influence one another. Consistent with this, there is evidence of intergenerational continuity in delinquency, drug use, and other antisocial behaviors. That is to say, children born to parents who have a history of involvement in delinquent behavior are themselves at significantly increased risk of also becoming involved in delinquent behavior (Farrington 2011; Thornberry 2005, 2009). There are a number of possible explanations for this level of intergenerational continuity ranging from purely genetic influences to shared environmental influences. Interactional theory's explanation for how parental behavior can influence child behavior in this regard is firmly rooted in the life course perspective and is summarized in Fig. 2.

One of the basic premises of interactional theory is that antisocial behavior has important causal influences on subsequent aspects of the person's development. Involvement in delinquency, especially if it is prolonged and serious, is not cost-free – it has negative consequences for the individual and, eventually, for the individual's children. Some of the important consequences that bear on the next generation are represented in Fig. 2. First, there is considerable evidence that delinquent careers are associated with disorderly transitions from adolescence to adulthood. Delinquency interferes with the successful completion of the major developmental tasks of adolescence, and delinquents are significantly more likely to drop out of high school, to become teen parents, to cohabit and exhibit unstable patterns of family formation, and to have higher rates of unemployment or underemployment. Both a history of delinquency and experiencing these disorderly transitions increase the likelihood of experiencing structural adversity as indicated by such factors as poverty, receipt of welfare, and residence in areas of concentrated disadvantage. All of these influences increase stressors such as depression, financial stress, aversive life events, and partner conflict. In other words, delinquent careers can initiate a cascade of negative influences that interfere with adolescent development, disrupt an orderly transition to adult roles, and create

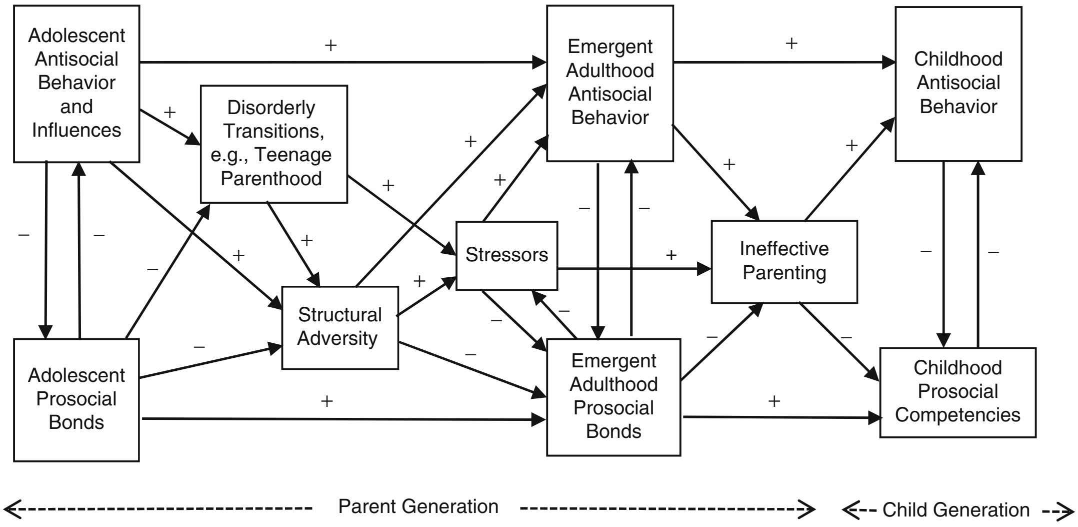
Adolescent Antisocial Behavior and Influences
Disorderly Transitions, e.g., Teenage Parenthood
Emergent Adulthood Antisocial Behavior
Childhood Antisocial Behavior
Stressors
Ineffective Parenting
Structural Adversity
Adolescent Prosocial Bonds
Emergent Adulthood Prosocial Bonds
Childhood Prosocial Competencies
Parent Generation
Child Generation
+
−

Interactional Theory of Delinquency, Fig. 2 A general conceptual model for understanding the intergenerational transmission of antisocial behavior stress and disadvantage during the period of early or emerging adulthood. These outcomes are not preordained by delinquent careers, but they are significantly more likely to occur for serious delinquents as compared to nondelinquents. Young adults experiencing this constellation of factors are poorly prepared to enter the role of parent and to provide a safe, stable, and nurturing environment for their children. They are more likely to remain involved in antisocial behaviors including substance use and to remain embedded in deviant social networks. In addition, they are less likely to develop strong prosocial bonds to conventional institutions, in particular in the arenas of family and work.

These problematic developmental processes have important intergenerational implications. The most potent pathway by which they influence the child's development and likelihood of delinquency is via their effect on family processes such as family conflict, hostility, and especially the quality of parenting behaviors. Less effective parenting styles include low affective ties and reduced involvement with the child, inconsistent monitoring and standard-setting, explosive and inconsistent disciplinary styles, and, at the extreme, child maltreatment. A robust literature describes the impact of ineffective parenting on the onset and maintenance of delinquent careers (Patterson et al. 1992). Children exposed to this style of parenting are more apt to be involved in delinquency, drug use, and other antisocial behaviors. They are also unfortunately less likely to develop prosocial competencies that may buffer them from these parental risk factors.

In general, the developmental processes depicted in Fig. 2 describe a set of pathways that lead from adolescent antisocial behavior by the parent to an elevated risk for involvement in antisocial behavior by their children. Parental delinquency creates disorder in their life course that enhances the probability of structural disadvantage, stressors, and antisocial behavior during the early-adult years, all of which interferes with the development of competent and effective parenting styles for rearing their children. Those inappropriate parenting styles are the most proximal influences on the child's behavior and the most powerful mediators between parental and child antisocial behavior. These processes certainly do not account for all of the intergenerational continuity in antisocial behavior that has been observed, but they are proposed to account for a substantial share of the variance.

And, at this point, we have come full circle. The children initiate involvement in delinquent behavior in part because they are linked to their parent's history and prior behaviors and to the consequences of those behaviors. But once initiated, their own delinquency begins to have the reciprocal effects described in the original version of interactional theory. Delinquency potentially weakens their bonds to society and enhances their embeddedness in deviant networks thereby increasing the likelihood of an escalating level of involvement in delinquent careers.

Empirical Support

Over the years a considerable amount of empirical support for interactional theory has been demonstrated. The core contention of interactional theory is that delinquency is embedded in a set of mutually reinforcing reciprocal relationships and that delinquency has consequences for the person's later development. In 1996 Thornberry reviewed 17 studies of delinquency and crime that examined bidirectional relationships. Overwhelmingly, they provided empirical data consistent with this core contention. For example, weak social bonds increase delinquent behavior, and involvement in delinquency further attenuates the strength of social bonds. Similarly, association with delinquent peers and delinquent beliefs increase the likelihood of delinquent behavior, and delinquency further cements those associations and beliefs.

In the context of the Rochester Youth Development Study, Thornberry and colleagues conducted a number of empirical assessments of interactional theory (see Thornberry et al. 2003b, for a summary). Jang and Smith (1997) found that delinquency and parental supervision are involved in mutually reinforcing relationships over time. In contrast, the relationship between attachment to parents and delinquency is unidirectional, but interestingly, delinquent behavior was found to attenuate attachment, while attachment was not significantly related to delinquency. Thornberry et al. (1991) and Thornberry and Henry (2009) demonstrated that engagement in school and delinquency are also reciprocally interrelated. Finally, relationships among beliefs, association with deviant peers, and antisocial behavior were investigated in several studies. Thornberry et al. (1994) focused on delinquent behavior, while Krohn et al. (1996) focused on drug use. In both cases support for interactional theory's hypotheses was evident. Reciprocal influences between associations and behavior, between beliefs and behavior, as well as between associations and beliefs were generally evident. Overall, there is little evidence to suggest that the major causal influence is simply from social bonds to delinquency or from peer associations to delinquency. In contrast, there is abundant evidence that these relationships are bidirectional and that delinquent careers develop over time both because of these influences and because of the feedback effects of delinquency on those influences.

Another central contention of interactional theory is that involvement in delinquent behavior has long-term consequences disrupting the normal course of adolescent development and transitions from adolescence to adulthood. Consistent with expectations, involvement in delinquency and drug use was found to be significantly related to early pregnancy, teenage parenthood, high school dropout, and living independently from one's parents during the teenage years (Krohn et al. 1997; Smith et al. 2000). In line with these findings from the Rochester Youth Development Study, many other longitudinal studies also find that involvement in adolescent antisocial behavior leads to later disruption in life course development (see, e.g., Huizinga et al. 2003; Newcomb and Bentler 1988).

Core aspects of the intergenerational theory presented in Fig. 2 have been examined in the context of the Rochester Intergenerational Study. First, there is clear evidence of intergenerational continuity both for general antisocial behavior and for drug use. Parental involvement in these behaviors significantly increases the likelihood of child involvement in these behaviors (Thornberry 2005, 2009; Thornberry et al. 2003a, 2006, 2009). Interestingly, the level of intergenerational continuity is moderated by the level of ongoing parental contact with the child in a manner that is quite consistent with interactional theory's emphasis on environmental influences. In particular, the impact of parental criminality on child criminality is statistically significant for mothers, virtually all of whom are primary caretakers of their children, as it is for supervisory fathers, who have ongoing contact with their children. In contrast, for nonsupervisory fathers – those who have little or no contact with their children – the relationship is not statistically significant and is quite close to zero (Thornberry 2005, 2009; Thornberry et al. 2003a, 2006, 2009). A similar pattern also appears to emerge between the drug-use patterns of grandparents and grandchildren (Thornberry et al. 2006). Overall, the pattern of findings in the Rochester study suggests that ongoing contact between the parent and child is essential for the intergenerational transfer of risk for antisocial behavior to occur (see also Jaffee et al. 2003).

Second, these intergenerational analyses from the Rochester study also offer strong support for the mediational model of intergenerational continuity described above (see Fig. 2). These studies (Thornberry 2005, 2009; Thornberry et al. 2003a, 2009) examined different aspects of parental antisocial behavior; different disorderly transitions; varying types of stressors such as depression, financial stress, exposure to negative life events, and parenting stress; different aspects of parenting styles sometimes based on interview data and sometimes based on direct observational data; and different aspects of the child's antisocial behavior measured by different reporters. Despite these methodological and measurement variations, the results are uniformly supportive of interactional theory's predictions. Adolescent antisocial behavior on the part of the parent leads to later life course disorder that in turn leads to ineffective parenting styles that increase the chances of the child's antisocial behavior.

The findings presented by Thornberry et al. (2009) illustrate this general pattern. For mothers, they found that parental drug use and delinquency during adolescence, as well as being a teenage mother, significantly increased their early-adult depressive symptoms. In turn, depressive symptoms significantly reduced their attachment to the child and that significantly impacted the child's externalizing behavior problems. In addition, maternal depression had a direct impact on the child's behavior. For the supervisory fathers, adolescent involvement in drugs and delinquency increased their early-adult depressive symptoms; depressive symptoms reduced the father's level of attachment to the child and that, in turn, was related to the child's externalizing behavior problems.

A number of other intergenerational studies also report results that confirm interactional theory's basic explanation for intergenerational continuity in antisocial behavior (e.g., Capaldi et al. 2003; Conger et al. 2003). First, these studies find that a history of parental delinquency increases the likelihood of child delinquency. More importantly, they also find that this intergenerational impact is largely indirect, mediated by the types of development processes depicted in Fig. 2. Of particular importance are disorderly transitions to adult roles, continued involvement in antisocial behavior, and, especially, parenting behaviors. As hypothesized by interactional theory, an ineffective parenting style – low attachment, poor monitoring, harsh discipline, etc. – is the strongest and most proximal mediator of the impact of a parent's history of delinquency on the likelihood of delinquent involvement by their offspring.

Conclusion

Interactional theory, first proposed in 1987, offers a broad explanation for involvement in antisocial behavior across the life course as well as for intergenerational continuity in antisocial behavior. It attempts to account both for the major causal processes that are associated with the onset of delinquent careers at different stages of the life course as well as for the consequences of involvement in delinquent behavior. This theoretical approach views delinquency as part of a dynamic developmental process in which factors like individual characteristics, family processes, and peer networks bring about delinquency, but involvement in delinquency also has causal impacts on those factors and, more generally, the person's later development. Those consequences help to account for levels of change and continuity in offending patterns as well as the manner in which the lives of parents and their children are linked with respect to antisocial behavior. There is also a body of empirical support for the basic tenets of the theory and its explanation of offending across the life course and across generations.

Related Entries

- ▶ Career Criminals and Criminological Theory
- ▶ Control Theory
- ▶ Desistance from Crime
- ▶ Pathways to Delinquency
- ▶ Social Control Theory of Sexual Homicide Offending

Recommended Reading and References

Capaldi DM et al (2003) Continuity of parenting practices across generations in an at-risk sample: a prospective comparison of direct and mediated associations. J Abnorm Child Psychol 31:127–142

Conger RD et al (2003) Angry and aggressive behavior across three generations: a prospective, longitudinal study of parents and children. J Abnorm Child Psychol 31:143–160

Farrington DP (2011) Families and crime. In: Wilson JQ, Petersilia J (eds) Crime and public policy. Oxford University Press, Oxford, pp 130–157

Huizinga D et al (2003) Delinquency and crime: some highlights from the Denver Youth Survey. In: Thornberry TP, Krohn MD (eds) Taking stock of delinquency: an overview of findings from

contemporary longitudinal studies. Kluwer-Academic/Plenum, New York, pp 47–92
Jaffee SR et al (2003) Life with (or without) father: the benefits of living with two biological parents depend on the father's antisocial behavior. Child Dev 74:109–126
Jang SJ, Smith CA (1997) A test of reciprocal causal relationships among parental supervision, affective ties, and delinquency. J Res Crime Delinq 34:307–336
Krohn MD et al (1996) Reciprocal causal relationships among drug use, peers, and beliefs: a five-wave panel model. J Drug Issues 26:405–428
Krohn MD et al (1997) The interrelationship between substance use and precocious transitions to adult statuses. J Health Soc Behav 38:87–103
Krohn MD et al (2001) Later delinquency careers. In: Loeber R, Farrington DP (eds) Child delinquents: development, intervention, and service needs. Sage, Thousand Oaks, pp 67–93
Moffitt TE (1993) Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. Psychol Rev 100:674–701
Nagin DS, Paternoster R (1991) On the relationship of past and future participation in delinquency. Criminology 29:163–189
Newcomb MD, Bentler PM (1988) Consequences of adolescent drug use: impact on the lives of young adults. Sage, Newbury Park
Osgood DW, Anderson AL (2004) Unstructured socializing and rates of delinquency. Criminology 42:519–547
Patterson GR et al (1992) Antisocial boys. Castalia, Eugene
Shaw DS, Bell RQ (1993) Developmental theories of parental contributors to antisocial behavior. J Abnorm Child Psychol 21:35–49
Smith CA et al (2000) The effect of early delinquency and substance use on precocious transitions to adulthood among adolescent males. In: Fox GL, Benson ML (eds) Families, crime and criminal justice. JAI Press, Amsterdam, pp 233–253
Thornberry TP (1987) Toward an interactional theory of delinquency. Criminology 25:863–891
Thornberry TP (1989) Reflections on the advantages and disadvantages of theoretical integration. In: Messner SF, Krohn MD, Liska AE (eds) Theoretical integration in the study of deviance and crime. State University of New York Press, Albany, pp 51–60
Thornberry TP (1996) Empirical support for interactional theory: a review of the literature. In: Hawkins JD (ed) Delinquency and crime: current theories. Cambridge University Press, New York, pp 198–235
Thornberry TP (2005) Explaining multiple patterns of offending across the life course and across generations. Ann Am Acad Pol Soc Sci 602:156–195
Thornberry TP (2009) The apple doesn't fall far from the tree (or does it?): intergenerational patterns of antisocial behavior. Criminology 47:297–325
Thornberry TP, Henry KL (2009) The interplay of school disengagement and drug use: an interactional perspective. Monatsschrift für Kriminologie und Strafrechtsreform (Criminol Crim Law Reform) 92:240–252
Thornberry TP, Krohn MD (2001) The development of delinquency: an interactional perspective. In: White SO (ed) Handbook of youth and justice. Plenum, New York, pp 289–305
Thornberry TP, Krohn MD (2005) Applying interactional theory to the explanation of continuity and change in antisocial behavior. In: Farrington DP (ed) Integrated developmental and life course theories of offending, vol 14, Advances in Criminological Theory. Transaction, Piscataway, pp 183–210
Thornberry TP et al (1991) Testing interactional theory: an examination of reciprocal causal relationships among family, school, and delinquency. J Crim Law Criminol 82:3–35
Thornberry TP et al (1994) Delinquent peers, beliefs, and delinquent behavior: a longitudinal test of interactional theory. Criminology 32:47–83
Thornberry TP et al (2003a) Linked lives: the intergenerational transmission of antisocial behavior. J Abnorm Child Psychol 31:171–184
Thornberry TP et al (2003b) Causes and consequences of delinquency: findings from the Rochester Youth Development Study. In: Thornberry TP, Krohn MD (eds) Taking stock of delinquency: an overview of findings from contemporary longitudinal studies. Kluwer Academic/Plenum, New York, pp 11–46
Thornberry TP et al (2006) Intergenerational roots of early onset substance use. J Drug Issues 36:1–28
Thornberry TP et al (2009) The impact of parental stressors on the intergenerational transmission of antisocial behavior. J Youth Adolesc 38:312–322

Intermediate Sanctions

▶ History of Boot Camps

Internal Security Strategies

▶ Strategies of Policing Terrorism

International

▶ International Policy Transfer

International Crime Victimization Survey

Pat Mayhew[1] and Jan Van Dijk[2]
[1]Crime and Justice Research Centre, Victoria University of Wellingon, New Zealand
[2]Human Security and Safety International Victimology Institute, Tilburg University, Tilburg, The Netherlands

Overview

Surveys of people's experience of crime are now well established. They go under various titles – "crime," "victims," or "victimization" surveys. This entry places the International Crime Victimization Survey (ICVS) in the context of the development of other victimization surveys. It discusses the ICVS from the point of view of its main rationale for making international comparisons. It also looks at what has been done on the survey front in terms of international comparisons outside the context of the ICVS and what might be in store for the future.

Victimization surveys have focused on various types of potential victims, although most take a nationally representative sample of people living in private households, as is the case with the ICVS. Those sampled are asked whether they have recently been a victim of crime, regardless of whether the victimization was reported to the police. By including non-reported crimes, victimization surveys provide a different (and larger) count of crime to that from police records – which encompass in large part only crimes reported by victims and which the police choose to record. Several studies have shown that survey-based estimates of commonly committed crimes can give more reliable evidence as to their volume and trends than police figures (see, e.g., Lynch 2006). Victimization surveys also collect social and demographic information to show which groups most often fall victim to crime. Police information offers relatively little on this front.

The Development and Types of Victimization Surveys

The first round of the ICVS took place in 1989 after about a decade and a half of victimization survey development. The most important early national victimization survey took place in the United States in the 1960s. This was followed in 1972 by the first round of what is now called the National Crime Victimization Survey (NCVS), which has been conducted continuously since. Early victimization surveys were also carried out in the late 1970s and early 1980s in Finland, the Netherlands, Switzerland, Canada, Australia, and the UK. These countries still conduct their stand-alone "bespoke" national surveys. By now, so too do many other countries, for example, New Zealand, Belgium, Denmark, Italy, Sweden, and South Africa. The surveys are done on a regular basis, although not necessarily every year. In some countries, victimization questions have been added to other generic household surveys (France, Ireland, Italy, and Romania are examples). In other countries, the ICVS has been adopted as the national survey, using augmented samples (Argentina, Estonia, Poland, and Japan are examples).

In many countries, household victimization surveys have been complemented by surveys focusing on special groups or particular forms of victimization. Violence against women (VAW) surveys are particularly popular, since "general purpose" household surveys are not considered to give good measures of sexual or assaultive victimization, which often involves offenders whom the victim knows. VAW surveys are notable for their differences in approach and have thus produced widely different estimates of harm to women (see, e.g., Percy and Mayhew 1997). Of increasing interest are ethnic minorities as victims of crime, the key issue being whether their different experiences are due to racial prejudice or to broader structural factors which negatively influence victimization risk. Business (or commercial) victimization surveys are also becoming popular. The types of business sectors looked at vary, although retailers and manufacturers have been a common focus. Recently, there has been some activity in terms of measuring cybercrime (or "e-crime") among businesses.

The ICVS

The ICVS has been by far the most serious attempt to obtain a survey-based measure of victimization in a wide range of different countries. It was set up at the end of the 1980s by a small group of academic criminologists (the two authors among them) who were keen to apply survey methodology to explore international comparisons of crime. All the early ICVS academics (along with many others) were clear that using police statistics for international comparisons was fraught because of differences in legal definitions, police recording practices, and the readiness of victims to report to the police. They were also aware that results from bespoke victimization surveys were extremely difficult to compare on account of design differences – in particular, with regard to the questions on victimization and procedures for counting it. The thrust of the ICVS, therefore, was to standardize design and use the same questionnaire and analysis methods to produce equivalent across-country results.

The first ICVS took place in 1989 in 13 industrialized countries (see Van Dijk et al. 1990). It was then repeated in 1992, 1996, 2000, and (with different management procedures) in 2004/2005. Another set of surveys, which have been called ICVS-II, were carried out in 2010 in six countries, and these are returned to. The ICVS was also repeated in 2010 in Estonia, Georgia, and Switzerland.

Some of the initial 1989 survey countries took part again in subsequent rounds, although not all of them. Other new countries joined. To help consistency, much of the data collection in each of the five main rounds of the ICVS was supervised by one polling company. In the first four rounds, survey coordinators were appointed in each country to liaise with the central team to minimize deviations from the central ICVS model (or "template"). From 2000 onwards, efforts were made by the United Nations Interregional Crime and Justice Research Institute (UNICRI) in Italy to execute surveys (usually at city level) in countries in transition and developing countries (see Gruszczyńska (2004) and Alvazzi del Frate (1998) for results). Western industrialized countries have largely paid their own way in the ICVS, whereas others have had financial support. The cost of financing the ICVS has been a major factor in adopting rather modest sample sizes. The template was that national samples should cover at least 2,000 randomly chosen respondents (aged 16 or more), although some countries increased sample size to aid better local measurement. The issue of sample sizes in the ICVS is returned to.

The fifth round of the ICVS in 2004/2005 was organized rather differently. The first was the *European Survey on Crime and Safety* (*EU ICS*) in which all the 15 current member states of the European Union (EU) took part. This was organized by a consortium led by Gallup Europe and financed by the European Commission. Results are in Van Dijk et al. (2007). The second set of surveys was done in countries outside the EU, coordinated by the UN. Van Dijk et al. (2008) report on the results for all countries, in and outside the EU. All told, 30 countries were covered at national level, with another 33 surveys in main or capital cities.

Across all rounds of the ICVS over a period of two decades, more than 140 surveys have been done in over 80 different countries (with national-level surveys in 37). Over 320,000 respondents have been interviewed with a questionnaire that has been translated into 30 or more languages. The full dataset can be consulted at http://rechten.uvt.nl/icvs/ICVS2005_3full.zip (last accessed 11.07.2010).

Other Comparative Surveys

Before the ICVS began, there were a few attempts at cross-national comparisons using victimization surveys. Some studies took results from stand-alone national surveys, but they largely ran into the sand because the victimization count was so heavily influenced by survey design and counting protocols (cf. Lynch 2006). A few studies mounted standardized surveys in a limited number of countries, for instance, in Scandinavia. These exercises have largely sunk into obscurity.

Apart from these early studies and the ICVS, there are some other recent comparative surveys worth documenting.

- In 1996, Eurostat (the statistical arm of the European Commission) piloted a small *Eurobarometer of Crime* (covering the populations of the EU member states). It included questions about victimization experience, with small samples of 1,000 in most countries (Van Dijk and Toornvliet 1996).
- As part of the ICVS program, a comparative International Commercial Crime Survey (ICCS) was mounted in eight countries in 1994, although problems of different sector coverage and small sample sizes meant that little became of it. A similar questionnaire was later used in six other countries. It focused on experiences of business victimization, safety around the business area, pollution issues, and the extent and cost of security. The ICCS questionnaire was modified again by UNICRI in the late 1990s to include more items on fraud, corruption, extortion, and intimidation. Surveys took place in 2000 with small samples of 500 managers in eight capital cities in Central and Eastern European countries (Alvazzi del Frate 2004). There has also been a global, comparative survey of economic crimes against businesses, including cybercrime (PricewaterhouseCoopers 2005; Bussman and Werle 2006). This survey was repeated in 2011 for the sixth time among respondents from 78 countries.
- In terms of violence against women, a World Health Organization survey collected data between 2000 and 2003 from women in 12 countries, although there appears to be some differences in samples and survey administration (Garcia-Moreno et al. 2005). A tighter comparative exercise was the International Violence against Women Survey, which used a fully standardized questionnaire and analysis methods. Between 2004 and the end of 2005, surveys were conducted with women in eleven countries (see Johnson et al. 2008).
- The European Union's Fundamental Rights Agency (FRA) conducted pilots in six EU member states in the late 2006 and early 2007 to test different probability sampling approaches to identify selected ethnic minority and immigrant groups in countries where population data on minorities is often limited. The pilots were the forerunner to the *European Union Minorities and Discrimination Survey* (EU-MIDIS) conducted in 2008 across the EU member states. The victimization questions were taken mainly from the ICVS questionnaire. Risks for ethnic minority residents were considerably higher than for others (EU-MIDIS 2009).

General Methodological Approaches

As said, the ICVS aimed for standardized methodology and questionnaire content. It built upon (and mirrors) some fairly generic features of victimization surveys. These are summarized below, with the ICVS put in context.

- **Samples** – Household victimization surveys generally adopt a stratified random sampling approach to achieve a representative sample in terms of age, gender, and geographical area. (Post-survey weighting is often done to improve representativeness.) Virtually all national surveys are cross-sectional (taking a different sample in each round). The ICVS has not deviated in these respects. Usually one person in each household is interviewed – and this was the case in the ICVS. This is generally someone aged 16 years or over (ICVS) or 18 years or over.
- **Mode of interview** – Interview modes have changed over time. Mail (or postal) surveys were a cheap option early on. These are now largely discounted mainly because of low (and probably unrepresentative) response – although Germany has recently conducted a postal survey in a pilot test for an EU-wide victimization survey that is discussed later. Face-to-face interviewing is still seen as the

"gold standard" – because of higher response rates and the belief that they build more rapport with respondents. Face-to-face interviews now generally use computer-assisted personal interviewing (CAPI) in which the questionnaire is programmed into a laptop which the interviewer uses to enter responses. Telephone interviewing has increased substantially as telephone coverage has grown. It was also boosted by methodological work that showed respondents willing to answer questions of a sensitive nature over the telephone (see Smith (1989) for an early test). Telephone interviews are now nearly always done through computer-assisted telephone interviewing (CATI), with samples typically drawn using some form of random digit dialling. The ICVS design "template" recommends CATI on cost grounds, although in countries where there is insufficient national telephone penetration, face-to-face interviews were conducted, generally with samples of 1,000–1,500 respondents in the main cities.

- **The screening process** – Respondents are screened for experiences of victimization over a given "recall period" (see below). The ICVS screener questions cover ten types of crime. The screener questions use definitions and concepts based on colloquial language rather than the law. In many surveys (the ICVS included), it is common for the selected respondent to answer questions about possible victimizations that can be seen as affecting the household as a whole (in the ICVS case: theft of a car, theft from a car, theft of a motorcycle or moped, theft of a bicycle, burglary, and attempted burglary). The respondent answers only about his/her experience in relation to theft of personal property, robbery, sexual offences and assault, and threats. (This is because it is felt that personal victimizations are not necessarily shared with other household members).

  Screener questions usually aim to elicit a simple "yes" or "no" answer in the first instance, with further questions about the nature of what happened coming later. (This is because a respondent faced with a long list of questions immediately after saying "yes" may soon learn to say "no" to a subsequent screener.) The count of victimization sometimes comes from affirmative answers to screener questions (the ICVS procedure). Sometimes surveyors post hoc count and define the crimes using more detailed information about what happened.

  While the focus of household victimization surveys is on "volume" crime, some surveys have also taken up issues of emerging interest. In later rounds of the ICVS, for instance, additional questions have been added on experiences with street-level corruption, consumer fraud (including Internet-based fraud), credit card theft, drug-related problems, and "hate crime". Counts of these are usually not added to the conventional volume crime count.
- **The recall period** – Victimization surveys aim to estimate victimization over a limited time period. This needs to be short enough that experiences are reliably remembered, but long enough to generate an adequate number of victimization incidents to report upon. Most surveys have a "recall period" of 1 year. Some surveys (the ICVS included) initially ask people about possible victimization over the last 5 years and then focus down – for the main measurement – on the last 12 months, or (as in the ICVS) the last calendar year.
- **Details of victimization** – Many surveys follow up affirmative answers to the victimization "screeners" with detailed questions about what happened. In the case of the ICVS, details of victimization incidents are collected about the "last" incident of a particular type (the most recent incident of assault, for instance). This approach reduces interviewing time, although it may risk bias insofar as respondents choose a "last" incident which is most salient to them or about which they have more to say.
- **Respondent information** – All victimization surveys collect sociodemographic information to assess how risks for different groups vary. This includes age, gender, household income, and personal education level. Some surveys also try to measure "lifestyle" or "routine activities," which the victimological literature

has shown to be important in explaining why some people seem to be in the wrong place at the wrong time or behave in ways that "attract" victimization (e.g., Hindelang et al. 1978). The ICVS has made some small incursions here.

- **Other crime-related information** – Making contact with respondents is the biggest survey cost, so surveyors typically take the opportunity to ask them about crime-related issues, as well as victimization. The questions asked in different surveys have differed but popular topics which the ICVS has also covered include the use of crime prevention measures, attitudes to the police, fear of crime, and (to victims) the police response and the need for victim support.

General Limitations

There have been extensive discussions of the limitations of victimization surveys (e.g., Mayhew 2007). The brief review here applies to the ICVS and virtually all other surveys.

- Sample surveys do not represent the entire population as most sampling frames exclude those in communal establishments, for instance, and the homeless. This makes little difference to national victimization estimates, but it obviously precludes a full picture of victimization patterns.
- Findings are subject to sampling error. Margins of error are obviously greater for surveys with smaller samples. Estimates are most imprecise for crimes that happen relatively infrequently and for answers to questions answered only by part of the sample. The ICVS template, as said, was that national samples should cover at least 2,000 randomly chosen respondents (aged 16 or more). These are sufficient for "top-level" comparisons across different countries for broad categories of prevalent crimes. For very accurate estimates, and estimates for subpopulations, the ICVS struggles.
- Household victimization surveys omit "victimless" crimes (such as drug possession) and crimes against businesses and society at large. There are omissions even for crimes against private individuals. The victimization of children is largely ignored, and it is difficult to cover fraud well, since people will not always know they have been victimized.
- By no means all potential respondents can be contacted, and some who are contacted refuse to take part. Bias from nonresponse needs to be acknowledged, although its extent is somewhat contested. One view is that non-contacts and refusers may have "more to say" in victimization terms. The other view is that people who are available and willing to be interviewed are those who have "something to say." This point is returned to in the context of the ICVS.
- There are other types of response bias too. Serious victimization incidents may be overcounted because of "forward telescoping," whereas more minor incidents may well be undercounted. Incidents of a sexual nature and/or those perpetrated by someone well known to the victim are also likely to be undercounted – because they may not be perceived as "criminal," or because respondents will be reluctant to talk about them. (As said, dedicated surveys are often seen as more suitable here).
- A further challenge is what is known as repeat or "series" victimization (such as domestic violence). It can be hard for respondents to remember series victimization as discrete events located accurately in time. Series victimization also poses a problem for "incidence rates" (the number of victimizations spread across a given number of respondents). Victims of series incidents often cannot count them reliably, and very high values, taken at face value, can distort victimization rate estimates for some groups. Reporting simple "prevalence rates" (the number of respondents victimized once or more in the previous year) is another option and one which the ICVS has generally taken (with the rates published with their margins of error at the 90 % confidence level).

Particular Challenges for the ICVS

Discounting the general limitations of household victimization surveys discussed above, a central question is whether the ICVS has achieved its purpose in terms of standardization and providing culturally relevant measures of crime.

The approach of the ICVS was that all countries should use the same questionnaire, adopt the same fieldwork period, and use telephone interviews. While much has been achieved here, the ICVS has fallen short of full standardization. This was especially so in the surveys in developing countries, where face-to-face interviews were mainly used, and sometimes less experienced interviewers. Marginalized groups living in informal housing were also hard to reach, as were the more affluent living in gated communities living in South America, for instance (Kury et al. 2002). While comparability may not have been seriously undermined, small differences in design and people's response to the survey may have influenced ICVS results to some extent. Some of the main issues are set out below.

- **The questionnaire**. Country coordinators were responsible for ensuring that the initial English language questionnaire was well translated. In the first ICVS, all translations were back-translated and checked by the coordinating group, but back-translation did not always occur subsequently. Particularly demanding was translating the concepts and terms of the screener questions. Other questions and terms were also problematic. For instance, one question asked victims how "serious" they felt their crime had been, but the term "serious" proved difficult to translate. Other problems arose, for instance, with "stranger" (which in some countries is nearer to immigrant) and "bribery" (too serious a term in some countries for the type of low-level bribery the ICVS was trying to capture).

  On other fronts, too, it was inevitable that some countries baulked at the standardized questionnaire, feeling that "they could do better" or that some questions were inappropriate. Some countries were also keen to add additional questions or restructure the ordering of questions, thereby introducing possible "context effects." Some made minor change to the ordering of response categories – causing huge complications in analysis.
- **Response rates** – These have been variable across countries and sometimes rather low. This was so especially in the first 1989 ICVS, when privacy regulations hindered sufficient recontacting of nonrespondents in some countries. Response rates generally improved subsequently, but then fell again in the 2004/2005 ICVS, reflecting a general trend in survey research. It may be that variable and low response affects ICVS results, although the extent of nonresponse bias is contested, as has been said. Analysis of results from the fifth ICVS also showed no statistical relationship in developed countries between the number of attempts needed to reach a respondent and overall victimization rates (Van Dijk et al. 2007). This suggests that reluctance to take part in telephone interviews may not have a serious impact on cross-national comparability. Whether this holds in developing countries is more arguable. The difficulties of contacting marginalized groups, or privileged groups, may mean that victimization rates in developing countries are underestimated.
- **Mode of interview** – An obvious challenge is whether ICVS face-to-face interviews produce similar results to those in CATI mode. Analysis of ICVS results so far has not indicated any systematic bias such that one mode produces higher victimization rates than the other (e.g., Scherpenzeel 2001). The burden of the methodological evidence is that survey productivity is most affected by quality control (e.g., the selection, training and supervision of interviewers).
- **Cultural sensitivity** – It is difficult to know whether people in different countries answer questions about victimization with the same degree of truthfulness. Different cultural thresholds for defining certain behaviors as crime may also apply, especially in relation to violent and sexual victimization. Rates of minor sexual victimization have typically

been higher in Western countries where the social position of women is most advanced, suggesting that subjective thresholds are lower (Kangaspunta 2000). In pilot tests for the proposed *European Security Survey* (discussed later), there were high refusal rates in relation to sexual and intimate partner offences in some Eastern European countries specifically. Thus, there is still a question mark over how well a standardized survey can cope with cultural sensitivity in this domain.

- A rather different issue is whether any one survey can capture experience of and reactions to "crime across the globe." One might suppose that the globalization of markets and mass media information has brought some attitudinal consistency as regards most conventional crime, especially in urban settings. Counter to this, though, is that there are inevitable limitations to the ICVS as a fine-tuned measure of the impact of crime. Despite much common ground in terms of people's usual experience of victimization, some countries may have particular concerns which bespoke surveys could cover better. Questions on bribery and corruption, for instance, are very sensitive to different interpretations as to what they mean and how seriously they are assessed. In some former Soviet countries, street-level extortion in the form of neighbors asking to borrow money is a major concern.

What Has the ICVS Shown?

Space precludes documenting the full extent of ICVS results here. Just four sets of results from the ICVS are singled out. The first is what it has shown as to the level crime in different countries compared to the picture from police figures. The second concerns what ICVS measures of *trends* in crime show relative to trends according to police figures. The third set of results relates to reporting behavior and assessments of police performance and the fourth to victims' assessments of the seriousness of crime across different countries.

ICVS and Police Measures of Crime

Van Dijk (2009) looked at 39 countries across the world in relation to "total ICVS crime" (as measured mainly in 2000) and "total police-recorded crime" around the same period (drawing mainly on the 2000 UN Crime Survey, which collects a range of criminal justice statistics). Figure 1 shows the results. "Total ICVS crime" is generally more restricted in scope than "total recorded crime," with the result that in some countries (Sweden, the UK, and Finland, for instance), total recorded crime per capita is higher.

Figure 1 shows no correlation between the actual level of victimization and the rates of crime recorded by the police (r = 0.212; n = 39; n.s.). Some countries with exceptionally high recorded crime also show a comparatively high victimization rate (South Africa), but many others do not (e.g., Finland, Canada, and Switzerland). Comparisons between country rankings according to the ICVS and police figures were repeated for different types of crime. The results showed positive correlations for robbery (r = 0.663, n = 37), and theft of cars (r = 0.353; n = 34), but much weaker (and statistically insignificant) correlations for other types of crime.

Trends in Crime

Of the countries which have taken part in the ICVS since 1989, there are 15 developed countries about which information is available from at least four different rounds, enabling an analysis of trends in crime over the last 10–15 years. The average victimization rates for these countries showed them to have peaked halfway the 1990s but declined since. The pattern for individual countries was nearly always the same. The drops are most pronounced in vehicle-related crime and burglary.

In looking at trends in crime, Van Dijk and Tseloni (2012) used results from countries that had taken part in four or five rounds of the ICVS between 1989 and 2004/2005. Rates of overall victimization in North America (Canada and the USA), Australia, and the nine European countries for which ICVS-based trend data are available show distinct downward trends. The drop in

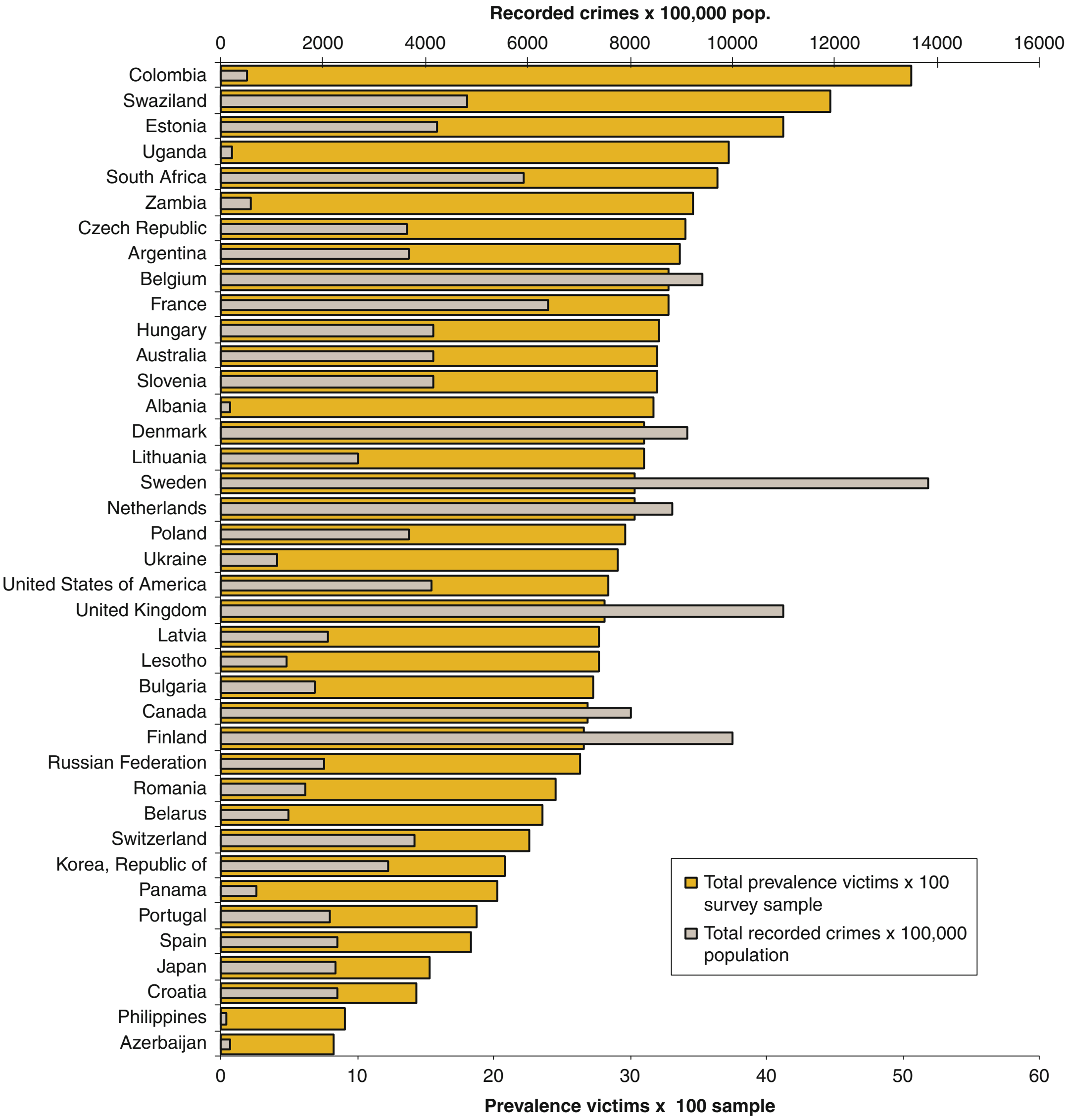

International Crime Victimization Survey, Fig. 1 Total crime by countries (Source: ICVS 2000 and UN Crime Survey 2002 or latest data available (Van Dijk 2009))

crime in the USA was already evident between the 1989 and 1992 ICVS rounds. The turning point came somewhat later in Canada, most European countries, and Australia. Trend data were available from only two middle-income cities in the developing world, but these also point at a downturn in overall victimization since 1996 (Buenos Aires) or 2000 (Johannesburg). In a separate analysis, Van Dijk and Tseloni centered on ten European countries, of which four had data until 2010 from the ICVS-II (see below). In all the countries except Belgium, rates of victimization in 2005 and/or 2010 were significantly lower than those of about 10 years before.

The purpose of Van Dijk and Tseloni's (2012) examination of trends in crime was not simply to focus on the ICVS. They also looked at survey-measured trends from a number of stand-alone national surveys, as well as the police figures in the countries that ran those surveys. Nonetheless, there are points worth noting from an ICVS perspective. First, Van Dijk and Tseloni note the fortuity of the ICVS being initiated in time to register the rise in volume crime, its general "peaking" around the mid-1990s and its subsequent decline since. Second, the ICVS covered a wider range of countries than alternative standalone survey counts did. Third, across this broader range of countries, the ICVS offered a picture of crime trends to compare with police figures. The fact that police figures could have their own momentum in registering changes in crime levels – due to changes in recording procedures or the scale and effectiveness of policing activities – bought the ICVS into play. It would not be fair to say that the ICVS holds credit for the criminological "bombshell" of a fall in crime in many countries in recent years. But in helping to document this, it has greatly added to the case that parochial (national) popular explanations of the fall in crime – such as policing improvements, greater imprisonment, or economic conditions – are not adequate.

Reporting and Police Performance

The ICVS measures levels of reporting to the police to see the extent of differences across countries and whether these help explain variations in police figures at country level. Without going into detail, suffice it to say that fairly marked differences in reporting levels emerge. There are generally higher rates in more affluent countries but still differences nonetheless. Referring back to the data on which Fig. 1 is based, the concordance between police-recorded crime and a measure of ICVS crime *reported to the police* was rather closer than was the case for police-recorded crime and *all* ICVS crime. This is important testimony to the fact that victims' reporting habits are one factor influencing the officially recorded output of police forces.

Although the decision to report a victimization has been shown to be mainly driven by a simple cost-benefit assessment, it also serves to some degree as a measure of confidence in the police. Using data from the ICVS 2000, Goudriaan (2004) found that country-level variations in reporting property crimes were related to confidence in local policing (among victims and non-victims). Van Dijk et al. (2008) also showed an interrelationship between levels of reporting, confidence in local policing and – a third factor – the degree to which victims who did report felt satisfied with the police response. A "police performance" indicator based on the three measures showed that the police performance was perceived to be best in Northwest Europe, Australia, and New Zealand. Scores on the index were lowest for Brazil, Mexico, Argentina, Estonia, Turkey, Cambodia, Greece, and Poland.

The Seriousness of Crime

As said, a challenge for the ICV is how far results might be affected by cultural differences as regards the meaning of victimization and responses to it. One result at least that undermines such concerns relates to victims' assessment of the seriousness of what happened to them – which they were asked to rate on a three-point scale. Van Dijk (1999) looked at results based on the 1992 and 1996 ICVS in relation to 14 types and subtypes of crime in countries grouped into five regions. Because of possible translation differences in the connotation of "serious," mean scores were translated into rank numbers per region. There was a striking similarity in the ranking of the crime types in terms of seriousness. Regression analysis also showed that seriousness ratings were determined much more heavily by crime type than any social characteristic of the victim. That said, there were a few differences. Motorcycle theft, car theft, and joyriding ranked highest in seriousness in developing countries but rather lower elsewhere – no doubt reflecting different economic values of cars and motorcycles. In the Nordic countries where levels of violence are comparatively low, simple assaults and threats are considered as more serious than elsewhere. In terms of victim characteristics, after controlling for crime type, gender was the strongest predictor of how seriously crimes are judged – with women considering crimes somewhat more serious than men.

Conclusions

Victimization surveys are now firmly established in criminological research. Surveys comparing the level and nature of crime internationally are a rather newer enterprise than stand-alone national surveys, although the ICVS started over two decades ago. It came into its own because of growing recognition of the difficulties of comparing police figures. This was something of a kiss of death for traditional comparative criminology, although it is fair to acknowledge improvement of late in exposing the problems of noncomparability. For instance, the *European Sourcebook of Crime and Criminal Justice Statistics* now itemizes the difference in legal definitions, counting protocols, etc. that affect police statistics (Aebi et al. 2006).

While the ICVS has not provided impeccable or comprehensive data by any means, it has arguably improved on what can be gained from police records as regards comparative levels and patterns of victimization in different parts of the world. It has also offered an important addition to the picture from police records on trends in crime.

While the ICVS has shown that it is feasible to conduct standardized surveys in a large number of countries, it nonetheless carries several practical lessons. Conducting surveys in a standardized way, more or less on time, and with sound adherence to the ICVS template posed a considerable logistical challenge. Tight oversight from the central team was needed to ensure that country coordinators engaged with tedious technical detail to maintain consistency. Financial matters were time consuming and challenging, especially with regard to developing countries. Moreover, there were occasional issues as regards ownership of results. When and how these were released was also a minefield – as crime "league tables" are politically sensitive. For countries taking part that had their own national surveys, country coordinators also sometimes needed to be able to explain why the ICVS could give different results as to victimization risks.

Future Developments

Policy interest in international comparisons of crime using victimization surveys has been enhanced by the results to date. Thus, countries are likely to be keen to enter into future comparative survey exercises – especially if fieldwork costs are met. But even where self-funding is necessary, countries that have not taken part in a standardized survey to date may want to sign up, particularly when they have insufficient resources for a bespoke national survey. This makes it attractive for them to take a ride on the back of another survey vehicle, especially when comparisons with other countries are on offer. A striking example of this are the repeats of nationwide surveys based on the ICVS in Georgia in 2010, 2011, and 2012 which have demonstrated a very significant drop in crime (Van Dijk and Chanturia 2011). An abridged version of the ICVS has also been used in 2012 in Moldova and Azerbaijan and is being pilot tested in Tamil Nadu, India. This said, the conduct of victimization surveys in developing countries seems to have stayed somewhat behind. This is unfortunate since in many of these countries, crime remains a major concern and statistics of recorded crimes are of poor quality, if they exist at all.

The logistical challenges of mounting comparative victimization surveys remain. So too do the methodological challenges of improving crime measurement internationally through survey techniques that are hard to standardize in diverse environments. "Survey saturation" in Westernized countries may also become an issue, whatever mode of interviewing is adopted. Over the lifetime of the ICVS, the development of CATI has reduced fieldwork costs and helped consistent questionnaire administration. Now, though, many people (particularly more heavily victimized younger ones) rely only on mobile phones. A key issue is whether viable representative samples can be obtained with the growing diversity of telephone provision – although survey companies will be under strong commercial pressure to achieve a solution. The inclusion of mobile phone users in sampling designs is an obvious priority.

As Internet use grows, computer-assisted web interviewing (CAWI) offers a way forward for surveys in the future, with cost advantages. Pilot testing with CAWI for the national victimization surveys in the Netherlands and Finland has begun, with fairly encouraging results. However, the methodological challenges of CAWI include possible bias due to differential Internet access, a degree of respondent self-selection, and low response rates. In the medium to long term, ways round these problems may be found, particularly by using incentives and representative panels that polling companies are increasingly likely to offer.

What of initiatives currently in place for further international comparative surveys? At present, there are no elaborate plans to repeat the ICVS globally in either of the two forms it has taken in the past – i.e., with a largely academic input or (as in the fifth round) with both an academic and EU input. However, there have been other developments afoot – albeit with a less global focus – which have taken or are taking forward comparative survey-based measures of victimization in different countries.

ICVS-II

The first initiative – the so-called ICVS-II – replicated the ICVS with a reduced questionnaire and across a broader range of interviewing modes. An initial plot was conducted in 2008 by a Dutch agency, NICIS, at the request of the International Government Research Directors (IGRD). NICIS commissioned pilots in Canada, Germany, Sweden, and England/Wales. NICIS subsequently mounted a round of surveys in 2010 with co-funding from the European Commission. Six countries took part: the four who participated in the first pilot, together with Denmark and the Netherlands. The questionnaire used in the first pilot was amended somewhat. Each country was to provide a net sample of 4,000, half of which was to be achieved using CATI and half through CAWI (this split between a register of personal data and addresses' and a "panel"). Results are available on the NICIS website, differentiating between CATI- and CAWI-based results (http://62.50.10.34/icvs/).

European Union Safety Survey (EU-SASU)

A second initiative, led by Eurostat, is a standardized household victimization survey – the *EU Safety Survey*. At the time of writing, this is planned to take place in 2013 or 2014 in all 27 member states. The survey can be seen as an extension of the EU's sponsorship of the 2004/2005 ICVS. It is intended to meet the perceived need for a full EU-wide comparative survey tailored to the current legal and social realities of the EU and its particular policy interests. The results will complement the numbers of crimes recorded by the EU member states annually published by Eurostat.

In 2009, the Universities of Tilburg and Lausanne were contracted by Eurostat to evaluate pilot tests in 17 member states using a draft questionnaire prepared by the European Institute for Crime Prevention and Control, affiliated with the United Nations (HEUNI). The pilot was to assess the questionnaire using different modes of interview, and recommendations were to be made in the light of results as to the best way forward. In the event, the piloted questionnaire proved problematic in the field on several fronts. Moreover, a subsequently revised version (trying to maintain some comparability with a set of core questions from the ICVS) did not entirely meet the approval of the survey agencies in all countries either – testimony to the difficulties of gaining consensus when jurisdictional sensitivities and interests differ. One difficulty has been the degree of stringency that can be imposed as regards interview mode. Although standardization is highly advisable (with CATI currently still the best option), dictating a common mode is unrealistic given different survey capabilities in different countries. There is also pressure from some countries to bypass the *EU Safety Survey* and provide "adjusted" results from their own large-scale national surveys instead – a procedure that may prove tricky to say the least.

European Union Crime Against Businesses Survey

The EU is also testing whether to take forward standardized, comparative surveys to assess the crimes against businesses in the 27 EU member states and the now three candidate countries. Pilot work started at the end of 2011, led by Gallup Europe in collaboration with Transcrime (in the Università Cattolica del Sacro Cuore, Milan). Large pilot surveys are being conducted in 20 of the EU member states, encompassing interviews with 19,000 businesses. A two-stage approach is being tested to minimize nonresponse. The first CATI stage gather information about locality and business activity, and it will screen for the prevalence and incidence of victimization of selected types of crime. In the second stage, CAWI will mostly be used to contact businesses that had reported victimization to collect more details of what happened.

European Union Agency for Fundamental Rights (FRA) Violence Against Women Survey

In 2012, the FRA conducted an EU-wide survey of violence against women. Led by HEUNI and UNICRI, this encompassed interviews with 40,000 women across the 27 EU member states and Croatia (a candidate country). The survey used standardized face-to-face interviews with about 1,500 randomly selected women in each country. (A pretest took place in six EU member states to test the survey questions). The interviews covered women's so-called "everyday" experiences of violence (including physical, sexual, and psychological violence, and harassment and stalking) by current partners, former partners, and non-partners in the past 12 months and since the age of 15. The survey also covered experiences of violence before the age of 15, to give a "lifetime" perspective. The aim is to compare the frequency and severity of violence against women in the different EU countries, including their access to and experience of police, healthcare, and victim support services. First results are expected in 2013.

Related Entries

► Bullying
► Cross-National Performance in Policing
► Cultural Criminology
► National Victimization Surveys
► Poverty, Inequality, and Area Differences in Crime
► Repeat Victimization
► Surveys on Violence Against Women

Recommended Reading and References

Aebi M, Aromaa K, Aubusson de Cavarlay B, Barclay G, Gruszczynska B, von Hofer H, Hysi V, Jehle J-M, Killias M, Smit P, Tavares C (2006) European Sourcebook of Crime and Criminal Justice Statistics 2006. http://www.europeansourcebook.org

Alvazzi del Frate A (1998) Victims of crime in the developing world. UNICRI, Rome, Publication 57

Alvazzi del Frate A (2004) The international business survey: findings from nine central-eastern European cities. Eur J Crim Policy Res 10:137–161

Bussman K-D, Werle M (2006) First findings from a global survey of economic crime. Br J Criminol 46:1128–1144

EU-MIDIS (2009) European union minorities and discrimination survey: main results report. European Union Agency for Fundamental Rights (fra.europa.eu/eu-midis)

Garcia-Moreno C, Jansen H, Ellsberg M, Heise L, Watts C (2005) World Health Organisation multi-country study on women's health and domestic violence against women: initial results on prevalence, health outcomes and women's responses. World Health Organisation, Geneva

Goudriaan H (2004) Reporting to the police in western countries: a theoretical analysis of the effects of social context. Justice Q 21(4):933–969

Gruszczyńska B (2004) Crime in central and eastern European countries in the enlarged Europe. Eur J Crim Policy Res 10:123–136

Gruszczynska B, Gruszczynski M (2005) Crime in enlarged Europe: comparison of crime rates and victimization risks. Transit Stud Rev 12:1337–1345

Hindelang MJ, Gottfreson MR, Garofalo J (1978) Victims of personal crime: an empirical foundation for a theory of personal victimization. Ballinger, Cambridge, MA

Kangaspunta K (2000) Secondary analysis of integrated sources of data. In: Alvazzi del Frate A, Hatalak O, Zvekic U (eds) Surveying crime: a global perspective. ISTAT/UNICRI, Rome

Johnson H, Ollus N, Nevala S (2008) Violence against women. Springer, New York

Kury H, Obergfell-Fuchs JM, Würger M (2002) Methodological problems in victim surveys: the example of the ICVS. Int J Comp Criminol 2(1):38–56

Lynch J (2006) Problems and promise of victimization surveys for cross-national research. In: Tonry M, Farrington D (eds) Crime and justice: an annual review. University of Chicago Press, Chicago, pp 229–287, Volume 34

Mayhew P (2007) Researching the state of crime: national and international and local surveys. In: King R,

Wincup E (eds) Doing research on crime and justice, 2nd edn. Oxford University Press, Oxford
Percy A, Mayhew P (1997) Estimating sexual victimisation in a national crime survey: a new approach. Stud Crime Prev 6:125–150
PricewaterhouseCoopers (2005) Global economic crime survey 2005. In collaboration with Martin Luther University
Scherpenzeel A (2001) Mode effects in panel surveys: a comparison of CAPI and CATI. Bundesamt fűr Statistik, Neuenberg
Smith MD (1989) Women abuse: the case for surveys by telephone. J Int Violence 4:308–324
Van Dijk JJM (2009) Approximating the truth about crime. www.crimprev.eu
Van Dijk JJM (2008) The world of crime; breaking the silence on problems of crime, justice and development across the world. Sage, Thousand Oaks
Van Dijk JJM (1999) The experience of crime and justice. In: Newman G (ed) Global report on crime and justice. Oxford University Press, New York, pp 25–41
Van Dijk JJM, Chanturia T (2011) The remarkable case of Georgia. Ministry of Justice of Georgia, Tbilisi
Van Dijk JJM, Tseloni M (2012) International trends in victimization and recorded crime. In: van Dijk JJM, Farrell G and Tseloni M (eds) The international crime drop: new directions in research. Palgrave Macmillan, Basingstoke
Van Dijk JJM, Manchin R, Van Kesteren J, Hideg G (2007) The burden of crime in the EU: a comparative analysis of the European crime and safety survey (EU ICS) 2005. Gallup Europe, Brussels, http://www.europeansafetyobservatory.eu/downloads/EUICS_The%20Burden%20of%20Crime%20in%20the%20EU.pdf. Last accessed 11 July 2010
Van Dijk JJM, Mayhew P, Killias M (1990) Experiences of crime across the world: key findings of the 1989 international crime survey. Kluwer, Daventer
Van Dijk JJM, Toornvliet LG (1996) Towards a Eurobarometer of public safety. Paper presented at the seminar on the prevention of urban delinquency linked to drug dependence. European Commission, Brussels, November
Van Dijk JJM, Van Kesteren JN, Smit P (2008) Criminal victimisation in international perspective: key findings from the 2004-2005 ICVS and EU ICS. Boom Legal Publishers, The Hague, http://rechten.uvt.nl/icvs/pdffiles/ICVS2004_05.pdf. Last accessed 11 July 2011

International Crimes

▸ Restorative Justice and State Crime

International Criminal Court

▸ Crime Investigations by the International Criminal Court
▸ Human Rights Violations in Criminal Court

International Evidence Exchange

▸ Transnational Exchange of Forensic Evidence

International Human Rights Standards and Community Sanctions

Christine Morgenstern[1] and Dirk van Zyl Smit[2]
[1]Ernst Moritz Arndt Universität Greifswald, Greifswald, Germany
[2]University of Nottingham, Nottingham, UK

Overview

In criminal justice, the question of the place of human rights often focuses on issues relating to imprisonment. However, imposing and enforcing community sanctions, such as probation, community work, or electronic monitoring, may also pose severe threats to the basic human rights of the offenders and their families. The rights potentially affected differ according to the sanction imposed and the stage of the criminal process – human dignity, privacy, and the presumption of innocence are examples of the rights that may be endangered. These rights are generally protected by binding international human rights law as expressed in the International Covenant on Civil and Political Rights or the European Convention on Human Rights. More specific international standards are set by recommendations and other instruments by international bodies. Questions arise, however, about the extent to which these standards are implemented and their impact monitored. These questions are becoming increasingly important as international cooperation, also with respect to community sanctions, increases and sanctions imposed in one country are implemented in another.

Fundamentals

It is indisputable that general legal standards exist that all sanctions must meet. These standards relate to the substance of the sanctions as such as well as to their appropriate imposition and implementation. They can be found in international human rights law and in national legal systems. The key standard is the absolute prohibition of cruel and unusual or inhumane punishment, although historically it has more often been applied to the death penalty and to the implementation of sentences of imprisonment than to sanctions that are implemented in the community.

It is often thought that human rights violations are less likely in the case of sanctions that are not to be implemented in a prison. Moreover, because community sanctions may serve as alternative to prison sentences, such sanctions are seen as benefits or privileges for offenders – as "doing them good." Therefore, the potential risks that they pose to the rights of offenders are routinely underestimated (Von Hirsch 1990; McNeill 2013) and need to be investigated. Social theorists as diverse as Foucault (1975) and Cohen (1985) have warned that in modern societies, the discipline and control that characterizes prison life may be reproduced within the community when sanctions are enforced there. This is particularly true in countries where a rise in punitiveness (von Hirsch 1990; van Kalmthout and Durnescu 2008; McNeill 2013) has led to many sanctions other than imprisonment becoming harsher, thus transforming measures designed to facilitate the rehabilitation of the offender into measures aimed primarily at supervision and control (van Zyl 1993). In this regard it is noteworthy that more recently scholars have recognized not only the pains of imprisonment but also the "pains of probation" (Durnescu 2011) as posing a potential threat to human rights.

The focus in this entry is on community sanctions and, to a lesser extent, on related community measures. According to the definition of "community sanctions and measures" adopted by the Council of Europe (1992), they refer to penalties that maintain the offender in the community, that is, outside prison, and involve some restrictions of his liberty through the imposition of conditions and obligations which are implemented by designated bodies. The composite term therefore also applies to diversion, compensation, reparation, and even mediation, as long as these forms of intervention are subject to some form of control by law-enforcement agencies in the widest sense. It does not, however, include noncustodial sanctions such as fines, which do not involve a specific "community" element. In most countries, community sanctions in the sense that it is used here far outnumber sanctions or measures that are enforced in prisons.

Safeguarding Human Rights in the Context of Community Sanctions

Rights at Risk

As a general proposition, it must be recognized that the human rights of offenders may be infringed by the nature of any sanction which is incompatible with their human dignity: The near universal prohibition on cruel, unusual, or degrading punishment is a specific application of this human dignity standard. A characteristic of the human dignity test is that it recognizes the ability of all people, including offenders, to reflect, to make choices, and to exercise social responsibilities. The right to human dignity may overlap with other rights. An example is a sanction that forces the offenders to express publicly sentiments that they do not hold. That may be inherently humiliating but it may also infringe a related human right to freedom of expression. Other fundamental rights may be infringed if a court were to order that an offender has to marry the mother of his child (the right to marriage and family life) or has to attend church services regularly (freedom of religion) or has to take birth control measures (the right to privacy). All these are examples of “community sanctions” imposed in various jurisdictions, which have been held to infringe human rights (for details and reported examples, see van Zyl Smit 1993).

Human rights violations also occur, arguably more frequently, when a generally acceptable community sanction is imposed for a disproportionately long period or implemented inappropriately. Particularly if the legal prerequisites for such a sanction are not entirely clear and too much discretion is left to judges and enforcement agencies, the suffering may well go beyond what is acceptable from the human rights point of view, either relative to the offence committed or in absolute terms. Here again, human dignity may be at risk, for example, when a community service order is enforced in a way that exposes the offender to dangerous or unhealthy conditions or otherwise unacceptable risks. Another example is electronic monitoring if it exercises such close control that the offender becomes a mere object of this control. Here, of course, the right to privacy is affected as well, and this also refers to the privacy rights of third parties such as the family of the offender. Further human rights issues that may arise from the ways community sanctions or measures are enforced stem from procedural issues, in particular with regard to due process and complaints procedures.

International Standards

Sources

Against the background of prison population inflation and prison overcrowding, as well as the increasing use of noncustodial alternatives to imprisonment that was emerging without much system, international bodies in the middle of the 1980s started to draft comprehensive instruments that would both encourage the use of community sanctions and other alternatives to imprisonment and protect the human rights of those subject to them. All these instruments build on existing human rights codifications, namely, the Universal Declaration of Human Rights (UDHR), the International Covenant on Civil and Political Rights (ICCPR) on the international level, and, at the regional level, particularly the European Convention on Human Rights (ECHR) but also the American Convention on Human Rights (ACHR) and the African Charter on Human and People’s Rights (ACHPR).

The first international standard minimum rules that dealt with community sanctions were drafted by experts of the United Nations Asia and Far East Institute for the Prevention of Crime and the Treatment of Offenders. Their development was heavily influenced by a set of rules formulated by an international NGO, the International Penal and Penitentiary Foundation (IPPF 1988), which aimed at creating an analogous and complementary human rights instrument to the pioneering 1955 United Nations Standard Minimum Rules for the Treatment of Prisoners, which were widely accepted throughout the world. The final version of this UN instrument, the Tokyo Rules, as they came to be called after their place of birth, or, more formally, the United Nations Standard Minimum Rules for Non-custodial Measures, was adopted by the United Nations’ General Assembly by consensus in December 1990 (United Nations 1990). The development of the Tokyo Rules (TR) was the result of a “careful process of international consensus building” (van Zyl Smit 1993). It is noteworthy that they focused both on encouraging the use of community sanctions and other alternatives to imprisonment and also, albeit to a lesser extent, to the creation of human rights protections against their abuse (United Nations Office on Drugs and Crime 2007).

European standards in this area were also prepared by governmental and independent experts, nongovernmental organizations such as the European Conference on Probation and Aftercare, and the secretariat of the Council of Europe. However, they were focused more clearly on the setting of standards in order to prevent abuses and to a lesser extent on propagating the use of community sanctions. The key standards were adopted in 1992 by the Council of Europe’s Committee of Ministers in the form of the European Rules on Community Sanctions and Measures (ERCSM) (Council of Europe 1992), which remain the primary regional instrument in this area to this day. They were supplemented in 1999 by a recommendation "concerning prison overcrowding and prison population inflation" (Council of Europe 1999). This latter recommendation, when read with the ECRSM, covers the same area as the Tokyo Rules but gives more details on the measures to reduce prison overcrowding as well as offering additional protection for those serving their sentences in the community. Following a critical evaluation of the ERCSM and their implementation, a further supplementary recommendation was adopted at the end of 2000 (Council of Europe 2000). It largely confirmed the existing ERCSM and concentrated on promoting their dissemination.

More was to follow. A further step was taken in 2003 with the adoption of a recommendation dealing with conditional release/parole, which sought to develop safeguards for persons who serve their sentences in prisons but complete them in the community (Council of Europe 2003). The European standard-setting process continues and most recently led to the adoption of the Council of Europe Probation Rules (CEPR) (Council of Europe 2010). Their rules are meant to be a practice-oriented addition to the existing ERCSM and rightly focus on probation agencies as their practices are crucial for the enforcement of community sanctions in conformity with human rights principles.

In contrast to the abovementioned UDHR, which is legally binding as customary international law (Meron 1989; Alston and Simma 1989), or the ICCPR, which is legally binding on all states that have ratified it, the more specific instruments are "soft law" and take the form of nonbinding recommendations. Nevertheless, they derive a degree of authority and legitimacy from the fact that they passed through a long drafting process that involved many governmental and nongovernmental parties and that they were adopted unanimously by the representatives of the states concerned (van Zyl Smit 1993; Morgenstern 2002).

Overriding Principles

Absolute limits for the imposition and enforcement of community sanctions and measures can be found in fundamental principles of international human rights law. As mentioned above, the respect for human dignity (Art. 1 UDHR) as well as the prohibition of cruel, unusual, inhuman, and degrading treatment and punishment (Art. 5 EDHR; Art. 7 ICCPR; Art. 3 ECHR; Art. 5s. 2 ACHR; Art. 5 ACHPR) plays a central role in this regard. Other widely recognized human rights, namely, the prohibition of discrimination (e.g., Art. 2 and 3 ICCPR) or general habeas corpus and due process requirements (e.g., Art. 9, 14 ICCPR), equally have to be respected. Consequently, both the Tokyo Rules (TR) and the ERCSM contain general "hold-all" provisions that refer to "norms accepted by the international community concerning human rights and fundamental freedoms" (ERCSM 21 and 22; TR 4). The precise meaning of these norms in the context of community sanctions needs to be specified, however, a task that the TR and the ERCSM undertake.

Both the TR and ERCSM stipulate the need for legal safeguards for every sanction or measure that has a penal content, while emphasizing that the idea of alternative or community sanctions being soft options or even privileges for the offender in comparison with a custodial sentence is wrong. They therefore call for clear and explicit legal provisions on the introduction, definition, and application of community sanctions and measures (that means sentencing and enforcement, especially regarding conditions and obligations and consequences of noncompliance, ERCSM 3, 4, 7, 9; TR 3.1). One of the fundamental principles spelled out clearly by the ERCSM is the concept of proportionality. Thus ERCSM 6 provides:

> The nature and the duration of community sanctions and measures shall both be in proportion to the seriousness of the offence for which an offender has been sentenced or of which a person is accused and take into account his personal circumstances.

Similar provisions can also be found in the TR (TR 2.6).

To safeguard offenders' rights and thus to implement the "legal citizenship" of the offender, both instruments contain detailed provisions on complaints procedures both against the imposition and the enforcement of community sanctions

(ERCSM 13 and 15; TR 3.5 and 3.6; Morgenstern 2002; van Zyl Smit 2012). Furthermore, particularly the ERCSM highlight that the cooperation of the offender is crucial for the success of the measure. Both instruments also stress the right to privacy of the offender and his family, which is of utmost importance, on the one hand with regard to the use and dissemination of personal data and on the other hand with regard to visits and other personal contacts.

The Pretrial Stage

If a measure is taken before or instead of a criminal prosecution, it is vital that appropriate steps are taken to recognize the human right of being presumed innocent until proven guilty (Art. 11(1) UDHR; Art. 14(2) ICCPR; Art. 6 ECHR; Art. 8(2) ACHR; Art. 7(1b) ACHPR). Although the scope of this principle may differ between jurisdictions, it is widely accepted that when it comes to the decision on whether or not to detain a suspect, it means that during the pretrial phase, generally, the offender should be left at liberty wherever possible, that is, granted unconditional bail. If unrestricted liberty would threaten the proper functioning of the criminal justice process in some way, there is a presumption in favor of conditional bail or its equivalents, that is, keeping the offender in the community but subject to restrictions. It can be observed, however, that this seemingly undisputed principle is particularly at risk when such alternatives to remand custody are set. Bail conditions may serve, whether intentionally or not, coercive, therapeutic, or even punitive purposes, and what should be (pretrial) community measures easily becomes informal sanctions. Moreover, even when they are not punitive, bail conditions are often introduced merely to allow the alleged offender to make a smooth transition from unconvicted to convicted status (Hucklesby 2011), which neglects the crucial principle that the person affected is still presumed to be innocent. It is essential therefore that whenever pretrial measures serve purposes other than to secure the functioning of criminal process – for example, if they are designed as rehabilitative measures or therapy – then the alleged offender must consent to them being implemented at that stage. The presumption of innocence and its consequences are incorporated in the TR and the ERCSM. With regard to the pretrial stage, the TR only mention the need for pretrial dispositions that include the possibility of discharge for minor offences (TR 5.1) and the need to avoid pretrial detention as far as possible (TR 6, United Nations Office on Drugs and Crime 2007). The ERCSM deal extensively with the question of cooperation and consent as one of the core problems of enforcement (ERCSM 31–36) and require it, as do the TR, particularly where any unconvicted person is involved (ERCSM 35, also CEPR 7; TR 3.4).

Sentencing and Enforcement

According to internationally accepted human rights standards, the key guiding principle to be used in sentencing is that of parsimony, particularly if the use of imprisonment is to be reduced. That implies that imprisonment may only be imposed as ultima ratio, both less often and for shorter periods (TR 2.6). The principle of minimum intervention also applies to the nature and duration of community sanctions, in particular with regard to the number and quality of potential conditions and obligations.

According to the preamble of the ERCSM, the application of community sanctions has to balance the need to protect society and the needs of offenders having regard to their social adjustment. The victims' interests are only incorporated indirectly in this preamble in the form that CSM must be implemented in a way that allows for "reparation for the harm caused to victims." It can thus be said that the ERCSM draw mainly on the concept of rehabilitation of offenders: All enforcement activities should gear towards it. The TR contain a similar commitment: They state that the sentencing decision should take into consideration the "rehabilitative needs of the offender, the protection of society, and the interests of the victim." They add, however, that the victim "should be consulted, whenever appropriate" (TR 8.2).

All these sentencing principles derive from respect for human dignity (TR 3.9). A sanction may infringe human dignity, when its penal content is unacceptable. This relates not only to corporal punishment or to humiliating public confessions, discussed above, but also to excessive sentences whose impact on the offender is unpredictable. Where, for example, the law provides for some form of unpaid work as a noncustodial punishment, it should also require the court to determine the total hours to be worked and in which period this has to be done.

A specific question in this regard is whether a sentence of community service requires the offender's formal consent before it can be imposed. Although this requirement is not universal, it exists in several European countries (Morgenstern 2010) and enjoys some recognition as an international standard (United Nations Office on Drugs and Crime 2007). This is partly due to national (constitutional) requirements which outlaw forced labor in general terms. The position in European human rights law is uncertain as Art. 4(3) of the ECHR creates an exception in the penal sphere only for work "done in the course of detention . . . or during conditional release from such detention." On the face of it, it does not extend to community service imposed directly as sanction in its own right (and not a substitute for prison), which should therefore be regarded as forced labor if no prior consent is gained. However, some scholars would extend the exception in the ECHR by analogy and argue that such consensual community service is not forced labor either (Emmerson et al. 2007). Whether the consent requirement may be circumvented when community service is labelled as exclusively rehabilitative is an open question (van Zyl Smit 1993; Morgenstern 2010). The International Labour Organization, rather than entering the debate directly, stresses that the nature of the work must be adequate for the rehabilitative purpose and may not serve a commercial function if it is to avoid being regarded as forced labor (Eradication of forced labour – General Survey concerning the Forced Labour Convention 1930 (No. 29); Abolition of Forced Labour Convention 1957 (No. 105); Morgenstern 2010).

Similar problems arise with regard to the necessity of consent to medical and psychological treatment as part of a community sanction: Does compulsory psychological therapy violate human dignity or other human rights? The instruments explicitly only prohibit "medical and psychological experiments" and "undue risk of physical and mental injury" (TR 3.4, a similar provision can be found in ERCSM 26). In any case, consent has to be informed and voluntary. The argument that the offender always acts under the threat of an otherwise unconditional prison sentence and therefore is never "free" cannot be dismissed immediately. In the end, however, it is not convincing because it leaves the offender at least the choice between two options, and prison is not always the greater evil (Durnescu 2011; Morgenstern 2010).

Both the ERCSM and the TR recognize that the rights of third parties, namely, the family of the offender but also under certain circumstances the victim are at stake. According to ERCSM 23, "the nature, content and methods of implementation of community sanctions and measures shall not jeopardize the privacy or the dignity of the offenders or their families, nor lead to their harassment. Nor shall self-respect, family relationships, links with the community and ability to function in society be jeopardized." Particularly intrusive sanctions or forms of enforcement, such as electronically monitored house arrest or even intrusive home visits and interviews by probation officers, may infringe third persons' rights to privacy and therefore require their informed consent and voluntary cooperation. The same principle applies to victims of crime when they are involved in the enforcement of a sanction as is the case with victim-offender mediation.

Attention must also be paid to the provisions that are made for consequences of an offender's failure to fulfill the conditions of the community sanction. One aspect that has to be considered in the sentencing process is the burden placed on the offender by conditions attached to the sanction. Even if on the face of it, the burden is not excessive, it may become disproportionate and infringe the offender's rights if it is enforced too strictly. In the enforcement process, there should be some flexibility in assessing whether there has been substantive compliance with the terms of the sentence. Most importantly, imprisonment should not be the automatic default sentence for failure to fulfill the requirements of the non-custodial sentence (TR 14.1; ERCSM 86). In deciding what further action is to be taken against the offender, other non-custodial possibilities must be considered, and partial fulfillment must be seen as a proportionately positive aspect. In the breach procedure, all due process requirements must be met (broadly spelled out in ERCSM 76-86).

Post-Sentencing Stage

Most countries have mechanisms that allow prisoners to be released before they have completed their full prison terms. Conditional release (parole) both reduces the burden on the prison system and takes into account the principle of minimum intervention by effectively converting (part of) a term of imprisonment into a community sanction. Early release can have a significant rehabilitative effect when it is accompanied by resettlement measures (Padfield et al. 2010). Nevertheless, early release from prison may raise human rights issues: Particular care must be taken to ensure that power to grant early release is not abused to put prisoners under pressure during the prison sentence. Where early release is conditional on good behavior in prison, it is important that the presence or absence of such behavior is determined fairly. The European Court of Human Rights has recognized that a penalty of loss of remission for a disciplinary infringement may be regarded as the equivalent of an additional sentence of imprisonment (*Ezeh and Connors v United Kingdom* appl. nos. 39665/98 and 40086/98, 19 October 2003). The consideration of early release in general must follow the same principles as stated above: It must be based on clear and explicit legal provisions that allow sufficient discretion to make the system flexible but not unpredictable. Conditions imposed should relate either to assisting the reintegration of prisoners into society or to exercising a measure of control on them while they are subject to such conditions, but must not constitute an additional punishment. Like other post-sentencing dispositions, conditional release and requirements attached to it must be subject to review if the offender so chooses (TR 9.3).

Enforcement and Monitoring of International Standards for Community Sanctions

The impact of international standards depends not only on their formal legal status but on the way in which they are enforced. The TR and the ERCSM are to a great extent the non-custodial equivalents of the respective sets of prison rules developed by the United Nations and the Council of Europe. However, they are less well known, and they do not have the same back up as the prison standards, which at the European level and to a lesser extent at the international level are supported by the monitoring activities of bodies designed to prevent the torture and ill treatment of persons in detention. Even if the ERCSM and the TR are based, in the same way as the prison rules, on the principles of the ECHR and the ICCPR, respectively, almost no jurisprudence by the European Court of Human Rights or of the United Nations Human Rights Committee can be found that deals with the imposition or implementation of community sanctions and measures. This is in stark contrast to the prison rules that are regularly referred to by these bodies when they deal with prison conditions. This lack of wider reference to the rules dealing with community sanctions is regrettable because both sets of rules adopt a highly principled approach towards the human rights-based use of community sanctions and represent what human rights lawyers call “evolving standards of decency” in this field (van Zyl Smit 2006).

It can be argued that the problems of enforcing and monitoring the Tokyo Rules also stem from the fact that they are drafted relatively loosely. For example, TR 1.3 allows implementation of community sanction in a way that “take[s] into account the political, economic, social and cultural conditions of each country and the aims and objectives of its criminal justice system.” This wording reflects the difficulties that are often experienced in arriving at universal understanding and acceptance of one set of human rights (Alston 1994).

More often than not, the inclusion of such very general provisions is the price for unanimity in adopting human rights instruments (Morgenstern 2002) and it is certainly a weakness. On the other hand, the official justification for such a provision in the TR – that they are not meant as one single model because they aim at being considered in very different criminal systems – cannot totally be dismissed if one takes into account that the UN had more than 180 member states at the time of their adoption. The situation, of course, is different at a regional or at least at the European level: Notwithstanding divergent interpretations of specific articles of the ECHR, the question of a universal understanding of human rights is not a source of major controversy within the Council of Europe.

Mechanisms to monitor the implementation of the instruments are limited but vary significantly between the two bodies. The United Nations used to collect data and reports on the implementation of their standards and norms including the TR. However, this material had little value because the information provided by the member states was incomplete. As a result, the systematic collection of such information has largely lapsed. The standards therefore are more often used as tool for technical assistance. The TR, for example, served as basis for the "Handbook of basic principles and promising practices on Alternatives to Imprisonment" (United Nations Office on Drugs and Crime 2007) which is used for training courses or seminars in countries who wish to reform this aspect of their criminal justice systems.

For the ERCSM, too, no formal monitoring mechanism is in place, but the Council of Europe tries to provide a follow-up on their dissemination and implementation by way of expert reports and attempts to collect data on community sanctions through its SPACE II program. Increasingly, also the European Union at least indirectly helps to implement criminal justice standards formulated by the Council of Europe by relying on them when developing its own criminal justice policies and projects (see below and Morgenstern 2009, 2011).

Future Developments

Recently, debates about community sanctions and their human rights implications have returned to the penal policy agenda. The rehabilitative ideal may have had been seriously challenged but in many jurisdictions is (again) acknowledged as the key principle governing community sanctions (Robinson 2008; McNeill 2013). It is also being recognized increasingly that the legitimacy of community sanctions is based on perceptions of how fair and humane they are and that human rights standards provide precisely the necessary and generally available criteria for evaluating them (Sparks and McNeill 2010).

How such standards may impact on criminal policy in the field of community sanctions is revealed by recent developments within the European Union: Historically, the European Union showed little interest in criminal justice and human rights-related questions, which were regarded as matters for the Council of Europe that includes all European states except Belarus. When the European Union became more active in crime and criminal justice-related fields in recent years, the engagement was usually associated with more effective law-enforcement instruments, as these were seen as closely related to its economic and political objectives. One more recently adopted EU instrument, however, aims at facilitating the cross-border enforcement of community sanctions within the European Union. The Framework Decision on Probation (Framework Decision 2008/947/JHA of 27 November 2008) will enable all member states of the EU to enforce a foreign probation sanction or measure according to their national practice; the enforcing state generally has to accept the judgment. Its objectives are "facilitating the social rehabilitation of sentenced persons, improving the protection of victims and the general public, and facilitating the application of suitable probation measures and alternative sanctions in case of offenders who do not live in the State of conviction." Given the very different sanction systems, enforcement modalities, and infrastructures of the 27 member states of the European Union, this is not an easy task, as member states are very reluctant to accept foreign judgments. This particularly refers to an area such as community sanctions where the details of how the punishments should be implemented are still largely shaped by nationally bound cultural norms. A lot of confidence-building measures are necessary, and steps have begun to be taken for EU member states to exchange detailed information about their community sanction systems (Flore et al. 2012). Here the established human rights standards of the Council of Europe come into play because they are already available, have been adopted by an even larger European institution (the Council of Europe to which all EU member states belong), and are based on fundamental common human rights values. They can thus be used as yardsticks or criteria against which national practices can be assessed (Morgenstern 2009, 2011).

Even if we consider that "human rights lawyers are notoriously wishful thinkers" (Humphrey 1973), this can be regarded as a positive development for the rights of those affected who may get to serve their community sentences in their countries of origin. It may be a harbinger of a wider international movement to recognize human rights in the enforcement of community sanctions.

Related Entries

- ▶ Alternatives to Pre-trial Detention
- ▶ Community Service in Europe
- ▶ Early Release from Prison
- ▶ Electronic Monitoring

Recommended Reading and References

Alston P (1994) Human rights in 1993: how far has the United Nations come and where should it go from here? In: Nowak M (ed) World conference on human rights. United Nations, Vienna. The Contribution of NGO's. Reports and Documents. Vienna, June 1993

Alston P, Simma B (1989) The sources of human rights law: custom, jus cogens, and general principles. Aust Yearb Int Law 25:82–108

Cohen S (1985) Visions of social control: crime, punishment, and classification. Polity Press, Cambridge

Council of Europe (1992) Recommendation No. R (92) 16 of the committee of ministers to member states on the European rules on community sanctions and measures adopted by the committee of ministers on 19 Oct 1992 at the 482nd meeting of the ministers' deputies

Council of Europe (1999) Recommendation No. R (99) 22 of the committee of ministers to member states concerning prison overcrowding and prison population inflation adopted by the committee of ministers on 30 Sep 1999 at the 681st meeting of the ministers' deputies

Council of Europe (2000) Recommendation Rec (2000) 22 of the committee of ministers to member states on improving the implementation of the European rules on community sanctions and measures adopted by the committee of ministers on 29 Nov 2000 at the 731st meeting of the ministers' deputies

Council of Europe (2003) Recommendation Rec (2003) 22 of the committee of ministers to member states on conditional release (Parole) adopted by the committee of ministers on 24 Sep 2003 at the 853rd meeting of the ministers' deputies

Council of Europe (2010) Recommendation CM/Rec (2010) 1 of the committee of ministers to member states on the council of Europe probation rules adopted by the committee of ministers on 20 Jan 2010

Durnescu I (2011) Pains of probation: effective practice and human rights. Int J Offender Ther Comp Criminol 55:530–545

Emmerson B, Ashworth A, Macdonald A (2007) Human rights and criminal justice. Sweet & Maxwell, London

Flore D, Bosly S, Honhon A, Maggio J (eds) (2012) Probation measures and alternative sanctions in the European Union. Intersentia, Cambridge

Foucault M (1975) Surveillir et punir. La naissance de la prison. Gallimard, Paris (English version: discipline and punish: the birth of the prison, 1977. Pantheon, Harmondsworth)

von Hirsch A (1990) The ethics of community-based sanctions. Crime Delinq 36:162–173

Hucklesby A (2011) Bail support schemes for adults. The Policy Press, Bristol

Humphrey J (1973) Foreword. In: Lillich R (ed) Humanitarian intervention and the United Nations. University of Virginia Press, Charlottesville

International Penal and Penitentiary Foundation (1988) The standards minimum rules for the implementation of non-custodial measures involving the restriction of liberty (the Groningen rules). IPPF, Groningen

van Kalmthout A, Durnescu I (2008) European probation service systems: a comparative overview. In: van Kalmthout A, Durnescu I (eds) Probation in Europe. Wolf Legal, Nijmegen

Meron T (1989) Human rights and humanitarian norms as customary law. Oxford University Press, Oxford

McNeill F (2013) Community sanctions and European penology. In: Daems T, Snacken S, van Zyl Smit D (eds) European penology? Hart Publishing, Oxford

Morgenstern C (2002) Internationale mindeststandards für ambulante strafen und maßnahmen. Forum Verlag Godesberg, Mönchengladbach

Morgenstern C (2009) European initiatives for harmonisation and minimum standards in the field of community sanctions and measures. Eur J Probat 1:128–141
Morgenstern C (2010) The requirement of the offender's consent to community service. In: Groenhuijsen M, Kooijmans T, de Roos T (eds) Fervet opus. Liber amicorum Anton van Kalmthout. MAKLU, Apeldoorn, pp 151–160
Morgenstern C (2011) Estándares europeos sobre penas y medidas comunitarias. In: Larrauri E, Blay E (eds) Penas comunitarias en Europa. Trotta, Madrid
Padfield N, van Zyl Smit D, Dünkel F (eds) (2010) Release from prison. European policy and practice. Willan Publishing, Cullompton
Robinson G (2008) Late-modern rehabilitation. Punishment Soc 10:429–445
Sparks R, McNeill F (2010) Incarceration, social control and human rights. In: The International Council of Human Rights Policy (ed) Social control and human rights. ICHRP, Geneva. Research paper
United Nations (1990) United Nations standard minimum rules for non-custodial measures (The Tokyo rules) adopted by general assembly resolution 45/110 of 14 Dec 1990
United Nations Office on Drugs and Crime (2007) Handbook of basic principles and promising practices on alternatives to imprisonment. United Nations, New York
van Zyl Smit D (1993) Legal standards and the limits of community sanctions. Eur J Crime, Crim Law Crim Justice 1:309–331
van Zyl Smit D (2006) Humanising imprisonment: a European project? Eur J Crim Policy Res 12:107–120
van Zyl Smit D (2012) Non-custodial sanctions and European human rights law. In: Roberts J, Zedner L (eds) Principles and values in criminal law and criminal justice: essays in honour of Andrew Ashworth. Oxford University Press, Oxford, pp 191–208

International Policy Transfer

Rob Canton
Community and Criminal Justice, Health and Life Sciences, De Montfort University, Leicester, UK

Synonyms

Exchange; International; Policy exchange; Policy transfer

Overview

There is a long history of the international exchange of ideas, research findings, policies and practices in criminal justice. In recent years, there has been a marked increase in more deliberate and strategic activity to "learn lessons" from the achievements (and sometimes the failures) of other nations and to "import" policies and practices. This has become known as *policy transfer*, which has been defined as "the process by which knowledge of ideas, institutions, policies and programmes in one setting is fed into the policy-making arena in the development and change of policies and programmes in another setting" (Dolowitz et al. 2000, 9). Transfer can take different forms, including copying, emulation, or inspiration. A nexus of factors – economic, political, social, cultural, and organizational – will influence the character and outcome of transfer attempts, and the complexity of these interactions makes the consequences of transfer inherently unpredictable. Different people may have different and changing motivations to transfer and, since neither the intended nor the actual consequences are always easy to identify, evaluation is far from straightforward. Consideration of policy transfer foregrounds the importance of agency and choice in penal policy, redressing theoretical tendencies to explain penal development solely in structural terms. Experiences of policy transfer often expose taken-for-granted features of criminal justice systems and illuminate influences that shape their character and development.

Background

Countries have long exchanged ideas, research findings, and practices in all aspects of criminal justice. As Radzinowicz put it (1999, 357), "Imitations, mutations and cross-fertilization of all kinds of institutions and measures in response to crime can be traced in abundance throughout the world. . .." These have included the study of crime and criminals; the collection, collation, and interpretation of criminal statistics; law; sentencing practice, policing, prisons, probation, working with victims, and crime prevention. Taking place through academic conferences and professional congresses, international publication, correspondence, and study visits, these exchanges have sometimes led to deliberate attempts to "import" institutions or practices from another country. Well-known examples may be found in historical accounts of developments in imprisonment and probation. The character and regime of Pentonville penitentiary in North London, the opening of which was an iconic event in prison history marked by historians as a radical new approach to imprisonment in England, were strongly influenced by American ideas: study visits had been undertaken to explore the relative merits of the silent associated system at Auburn and the solitary confinement regime in Philadelphia and to decide on the better model (Ignatieff 1978). In its turn, Pentonville was influential in shaping prison design in many other parts of Europe.

John Augustus, the Boston shoemaker, whose practice of standing bail for "for an array of convicted adult drunkards, delinquent children and young prostitutes" (Nellis 2007, 28) led many to think of him as the father of probation, inspired the work of probation officers in Massachusetts. These initiatives were later brought to the attention of the British Home Secretary by the Howard Association (*ibid.*) and contributed to the development of probation in England. Anglo-American models of probation were to become widely influential in many other countries. It is to be noted, however, as with the lessons taken into Pentonville, that it seems more accurate to speak of influence and adaptation rather than simple copying.

The case of Mettray, the boys' reformatory, near Tours in central France, affords an instructive early example of both the possibilities and complications of transfer. In the mid-nineteenth century, Mettray was heralded as a model reformatory: many visitors were welcomed from abroad and penal institutions or colonies modelled on Mettray were established in a number of countries. Sydney Turner and Thomas Paynter were sent out from England to find out if the principles of the institution could be replicated in their own country. While they found much to admire, they drew particular attention to differences between French and English culture and tradition so that "... there can be no mere transplanting of it [Mettray] to this country" (Smith 2008, 76). They also noted that the character of Mettray was powerfully influenced by the personal qualities of its charismatic founder, Frédéric-Auguste Demetz, raising questions about the longer-term future of the reformatory as well as precluding any simply imitative "transfer."

The reflections of Turner and Paynter, recognizing that culture, context, setting, and individuals are quite as important as particular techniques or methods in determining the effectiveness of penal practice, should stay in the minds of all involved in modern policy transfer endeavors. Many initiatives are developed regionally or locally and may not be assumed to be readily transferable even to other parts of the same country (Crawford 1998; Edwards and Hughes 2005). In the case of programs to reduce reoffending, for example, even within a single country, initiatives and programs that have had encouraging outcomes in a pilot stage, once "rolled out" nationally, often fail to achieve such impressive results. Plausible explanations include the likelihood that the enthusiasm of the staff who were involved in the first programs, their commitment to and understanding of a new program, may not be replicated when the program is introduced somewhere else (Raynor 2004). Other contextual factors – staff training, committed leadership, targeting (making sure the right people attend), resourcing, and staff supervision – can make a decisive difference. When these programs are taken to another country, with different practices, institutions, and traditions, such challenges increase exponentially.

Thinking About Transfer

David Dolowitz (*ibid.*) framed a series of questions that constitute a framework within which to explore and interpret what can take place when a jurisdiction considers adopting policies, practices, or institutions from elsewhere. These include:

- *Why transfer?* This opens an inquiry into the motivations of those involved. It includes consideration of the reasons why it was felt that policy needed to change at all and why reference to another country was found to be the way to innovate (rather than have recourse to local “indigenous” solutions). It has been suggested that specific transfer initiatives can be placed on a spectrum from voluntary to “coerced” (Newburn and Sparks 2004). Some governments are actively keen to introduce policies from other countries – perhaps because they are believed to be effective and/ or they are thought to be politically advantageous. At the other extreme, conformity with international requirements (e.g., human rights conventions) can require countries to introduce policies even with a degree of reluctance. For example, since no country has ever been admitted to the European Union without first being a member of the Council of Europe (which requires a formal commitment to the European Human Rights Convention), economic ambitions and the aspiration to eventual membership of the EU have been a prominent incentive for some former Soviet countries to enhance human rights by reforming their penal practices and to be seen to be modernizing (Canton 2006). At the same time, many people in these countries hoped to repudiate the Soviet legacy, to find less repressive (and less expensive) responses to crime and to develop criminal justice practices that were consistent with their standing and self-awareness as transitional democracies. It is not clear in such circumstances whether adopting penal policies from other countries should be described as voluntary or coerced. No less significantly, motivations may *change* in the course of the transfer process: as the full implications of transfer become clearer, what was initially sought may now be much less welcome or, on the other hand, what was resisted may now be embraced. Again, a government may adopt a policy voluntarily, while its administration and judiciary may be rather less enthusiastic but find that they have little choice. Many agents are involved, then, and their respective motivations may be different and variable. This leads to our second question.

- *Who is involved?* In the receiving country, very many people may be involved in interpreting new policy or practice and in putting it into effect. Many more may be affected by the transfer and may seek to mediate, mold, or even block the innovation. Transfers typically also involve “exporters” (advisers or consultants), who will support the receiving country in implementation, but who may well have their own interests in the way in which the transfer takes place and develops. A reminder of the many individuals involved, the relationships of power among them, and their various and shifting motivations helps to counter purely structural and potentially deterministic accounts of transfer – for example, the idea of *globalization* – and emphasizes the place of choice and agency in penal policy (Newburn and Sparks 2004).
- *From where?* Commentators have often noted the influence of the USA on penal policy in other countries, as indeed on so many other areas of social policy and on culture: globalization often seems close to Americanization (Newburn and Sparks 2004). Many of the resonant and (too) familiar crime and punishment slogans – *zero tolerance*, *naming and shaming*, *three strikes and you’re out* – are of American provenance, although they have now become part of the discourse of criminal justice debate in any number of countries. In Europe and in many Anglophone countries, the United Kingdom has often been influential – sometimes acting as a conduit for Canadian and American ideas, but also in its own right Dolowitz et al. (1999). Much of the research into *what works* with offenders was led by scholars and researchers in the USA and in Canada; adapted, developed, and applied in the UK; and then spread to other European countries. At the same time, other ideas and practices have a quite different origin: some of the principles of restorative justice and of reintegrative shaming have developed out of studies of premodern, indigenous communities and become well known through academic and practice (rather than political) networks (Braithwaite 1989).

- *What is transferred?* Countries may seek to import institutions, technologies, methods, and techniques. But transfer will also introduce new concepts and meanings, ideas, ways of thinking and talking about crime and punishment, either as part of a new discourse and/or in association with new arrangements and practices. Newburn and Jones (2007) distinguish between the "hard" transfer of particular policies and the "soft" transfer of ideas, principles, and symbols: in their discussion of the spread of the idea of *zero tolerance*, they conclude that there is much clearer evidence for the transfer of the idea and the stance of zero tolerance than of any associated concrete policies and practices. More generally, in the early 1990s, the United Kingdom took note of the way in which the politics of crime and punishment was presented and deployed for electoral advantage in the USA, while other countries too have increasingly recognized the political gains achievable through demonstrably robust responses to matters of crime and punishment. Especially at times of insecurity and enhanced sensitivity to risk, the symbolic weight of being seen to be *tough on crime* has been extensively exploited in this way (Jones and Newburn 2007). This enhanced sensitivity to risk is itself an international phenomenon and has had a wide influence, leading to what has been called a "preventive turn" in the policies of many countries (Crawford 2009). Other examples include ideas about the proper place of "the market" in security or corrections and, reciprocally, the role of the state, which may also be raised when countries look for models from elsewhere.
- *To what degree?* Dolowitz and colleagues (2000) distinguish different degrees of transfer – copying, emulation, a mixture of these, and inspiration (emulation differs from copying in that it implies transfer of the ideas behind, but not the details of, the policy or program). While the conceptual distinctions can be useful, in practice these differences can become conflated, and neither the intentions nor the consequences of transfer can always be confidently assigned to one "degree" rather than another.

Influences on Transfer

Countries differ from one another in many ways and these have their effects upon transfer activities. They include history and culture; the size, geography, and demography of the country; different legal traditions and their relative strength and maturity; the administration of criminal justice and the political management of its institutions; and the training and management of staff. Again, since there will always be differences between "exporting" and "receiving" countries, questions arise about the "tolerance" of these factors – how favorable or antagonistic must they be to support or to spoil transfer? At the least, a provisional identification of the factors which constitute the parameters of transfer, which shape its implementation and its prospects of success, would have to include:

- The criminal law. It is obvious, but easily overlooked, that transferred practices have to be consistent with national law. With political support, the law may be changed, but these processes can be slow and uncertain. Many countries have criminal or penal codes that may have to be supplemented with procedural codes if they are to be put into effect. In one country in the author's experience, the criminal code empowered the court to impose community service orders which were then to be put into effect by the designated competent authority. But no authorities had been so designated, and in the absence of such supporting procedural legislation, no community service orders could be made. Countries also differ markedly in the discretion, both in statute and in case law, that decision makers are accorded. The development of community sanctions, for example, may be thwarted by statutory requirements to impose prison sentences in particular cases.
- Criminal justice institutions and practices. Any innovation will need to find (or, more likely, will need to make) a place for itself within existing arrangements and will inevitably impinge upon the interests of other institutions and professional groups; their response will mediate or even resist implementation.

For example, the introduction of presentence/probation reports for the court makes (is intended to make) a difference to the sentencing process. This disturbs familiar practices and role relationships among the principal agencies, notably in jurisdictions where the prosecutor proposes the sentence to the court. The prosecution and judiciary will react variably to what may be seen as a challenge to their authority.

- Cost. Innovations have to be affordable and criminal justice is always in competition for scarce resources. One example is imprisonment. It is commonly said that prison is more expensive than probation, but where no probation service exists, new money must be found to pay for staff, training, and infrastructure. In comparison, the marginal costs of adding other prisoners to a swollen prison population are relatively small. Meanwhile, to the (very debatable) extent that establishing community sanctions and measures will lead to a reduction in the numbers of people in prison, these gains will only be realized gradually and perhaps no more than notionally.
- Research. Countries vary in the extent to which research is regarded as a foundation for policy or practice. Sometimes the claim is made that policy is *evidence led*, although such claims may be disputed, not least (though not only) because the implications of research findings are typically contested in this area. Broader political considerations and ideology, moreover, inevitably shape the interpretation of research findings so that penal policy is rarely *determined* by research. Even so, research may lend support to (or discourage) policy initiatives. The claim that a practice that has been shown to be effective in one country is now to be adopted can also be politically advantageous. Examples include the introduction in the UK of American strategies – *zero tolerance* or the *broken windows* approach to reducing crime (Jones and Newburn 2007) – and the "modernization" of Eastern European systems by seeking models of effective probation practice from the west of the continent (Canton 2006). On the other hand, most research findings have been made in specific national (or, as we have seen, regional or local) contexts and may not simply be assumed to have the same significance in another country. For instance, actuarial instruments to predict rates of reconviction depend upon statistical correlations established in particular countries that may not be obtained elsewhere. Bauwens and Snacken (2010) have also drawn attention to what may be called *second-hand transfer* where an instrument already modified (and as it may be compromised or even distorted) has been exported again in that revised form. For example, an assessment instrument may be taken from country A, adapted by country B, and then exported in this modified form to country C. The extent to which practices – assessment instruments, offending behavior program – can be modified in this way without prejudice to their integrity is insufficiently understood and may not be taken for granted.
- Political economy. There is a strong association between penal policy and political economy (Cavadino and Dignan 2006), and penal projects will be variably accomplished in societies with different socioeconomic structures. Even within broadly similar socioeconomic systems, there will be political variations that will mediate the transfer and can cause it to take an unexpected direction. Crime and punishment are strongly influenced by broader social policies and trends. Even if politically supported in its own terms, penal change may fail to thrive if it takes a markedly different trajectory from other policies. Desistance research is increasingly drawing attention to the importance of social capital – for example, access to employment and accommodation – in providing viable pathways out of crime; political economy influences the availability of social resources as well as their accessibility by offenders and ex-offenders. Relatedly, there are likely to be differences in the extent to which people regard crime reduction as something that the state should (or even could) achieve. The idea, for example, prominent in many western democracies, that the

police are in the service of the community is not the legacy of many countries where they have been viewed with suspicion as an (often repressive) arm of government.

- Technology and commerce. Technology is increasingly a feature of modern penality in all kinds of ways, from the processing and exchange of information to the electronic monitoring of offenders. The manner in which information is collected and processed can reciprocally influence the way in which the penal subject is understood (Aas 2004). Technological capacity and infrastructure constrain the potential of an innovation and shape its development. This consideration is linked too to commercial markets, as private enterprise becomes increasingly, though variably, involved in penal practices that have traditionally been the prerogative of the state. This has direct implications for transfer. Liberal criminologists, for example, may well be wary about net widening (Cohen 1985) – the idea that penal programs can draw in groups of offenders who had been dealt with before in less intrusive or more informal ways, but commerce in the nature of the case seeks to expand its markets and may give priority to the interests of shareholders over a wider public interest.
- Pressure groups, networks, public opinion. Innovations need their champions and the support or opposition of a pressure group can make a critical difference to the acceptance of a new idea, as can its representation in the mass media. Policy transfer has to engage with these challenges, not least in transitional democracies, where penal developments ought to progress in step with public opinion and with proper regard for their legitimacy. At the same time, where politicians have tried to seek political gain through "talking tough" about crime and punishment, mass media may well oppose and misrepresent any innovation that appears to be liberal.
- The ethical environment. This elusive but significant concept refers to "the surrounding climate of ideas . . . [that] determines what we find acceptable or unacceptable . . . our conception of when things are going well and when they are going badly . . . our conception of what is due to us, and what is due from us, as we relate to others" (Blackburn 2001, 1). The language of human rights, perhaps, as an accepted international ethical discourse, is often the way in which transferred innovations are discussed and interpreted. Even so, there are likely to be decisive differences in conceptions of the rights and duties of offenders and of victims and of the obligations of society towards them. In discussing crime prevention, Crawford comments on variation among countries in the ways in which "crime prevention strategies (seek to) embody a dynamic of social integration or exclusion" (1998, 243). He connects this with wider problems of social and economic cohesion and exclusion. Just the same observations are relevant to penal policies where societies take a variety of positions on whether punishment should be inclusive or exclusive, when (if ever) an offender should be considered an ex-offender, and what obligations (if any) the community has in this regard, as well as putative rights against them (in the name of public protection).

The list is not exhaustive and no doubt other influences could be identified. However that may be, these different factors, acting in different ways and with varying weight, make the development and outcome of transfer activity inherently unpredictable. Bluntly, what works here may not work there.

Culture and Transfer

Culture is increasingly recognized as central to an understanding of crime and punishment (Smith 2008) and is accordingly a critical variable in policy transfer and implementation. The term, however, is contested and used in a number of different ways. Garland (2006) warns that, in appropriating the term from other academic disciplines, criminology risks reproducing some of the confusions and ambiguities that prevail elsewhere. He distinguishes usefully between (1) culture as a collective entity (i.e., *this* culture as opposed to *that*, e.g., African culture contrasted with European) and (2) culture as a distinct influence in shaping penal development – that is, "“cultural” forces (or ideas, or symbols, or values, or meanings, or sentiments . . . {as opposed to} other kinds of entity (such as social, political, economic or criminological factors) . . ." (2006, 422). Culture is important in both senses in considering policy transfer, but the distinction is analytically valuable. A further complication is the use of the word *culture* – as in *legal culture* or *court culture*–where the term seems to mean not much more than tradition, habit, or the familiar way of doing things.

The claim that transfer must be culturally sensitive or culturally appropriate risks losing these distinctions. Transfer, to be sure, must take account of customary ways of doing things, but the purpose of transfer is often precisely to make changes in these respects. Transfer enterprises must assuredly take account of "the cultural" – the ideas, values, and meanings of penality. It could be said penal policies thrive or fail as much or more in the *meanings* they manage to convey than in their effectiveness in reducing crime (Smith 2008), and accordingly, disregard for the meanings of current practice and of the innovations that are intended to vary these meanings can vitiate transfer.

In its other sense – "this culture as opposed to that" (Garland 2006) – culture matters, but should not be seen as either insulated or static. Indeed, transfer typically occurs when a country is already exposed to other kinds of outside social, economic, and cultural influences, and its culture is consequently evolving, often rapidly. Even so, lack of sensitivity to cultural variation could constitute a form of neocolonialism or imperialism (Worrall 2000) – with prosperous countries telling relatively poorer ones how they should respond to crime – and is in any case likely to lead to failure in transfer. In this connection, the attempt to introduce community service/unpaid work into Zimbabwe as an alternative to custody is instructive. The idea that this new penalty of the court might *make amends* was greeted as an affirmation of restorative African traditions that had been suppressed in the first place by a "procedure dominated and prison-happy philosophy" introduced under colonialism (Stern 1999. The quotation, on page 25, is by Elufemi Odekunle). This seems an example of transfer reviving or reaffirming indigenous traditions that had been suppressed in the first place by "transfer" through colonization.

This discussion shows that while the insistence that transfer must respect culture is salutary, it is a claim that can be understood in a number of different ways. Cultural differences can (perhaps too easily) be invoked to account for some of the obstacles to transfer or to reject an initiative that some may find threatening. At one level, it is morally and practically imperative that innovation be culturally sensitive; on the other hand, the whole point of transfer is to bring about change.

Obstacles to Transfer

Among the ways in which transfer may fail is through uninformed or "unknowledgeable" transfer (Dolowitz et al. 2000). This arises from misunderstandings about the context in which innovation is to be made and/or the nature of the challenge that the transfer is intended to meet. Among the most immediate (and often persisting) problems confronting a consultant from another country, then, is trying to understand the criminal justice practices in the country accepting transfer. Difficulties here can be formidable. Most consultants may be presumed to be very knowledgeable about their own national systems, but similarities in institutions and practices can turn out to be superficial and to conceal significant differences. The temptation is to assimilate unfamiliar practices to ones with which the consultant is already familiar. Yet equally there may be "functional equivalence" (an expression, however, to be treated with care) between practices that at first sight seem very different. Consultants must therefore rely heavily upon local experts, although as Nelken warns (2000, 6), "those relied on for descriptions of the aims or results of legal reforms are themselves part of the context they are describing, in the sense of favouring one position rather than another."

Secondly, transfer may be incomplete–although this partly depends on whether copying, emulation, or inspiration is the objective (see above) and remembering too that different actors may have different aspirations here. Thirdly, inappropriate transfers can fail because "insufficient attention may be paid to the differences between the economic, social, political and ideological contexts in the transferring and borrowing systems" (Dolowitz et al. 2000, 34). Transfer seems most likely to be successful where there is social and political affinity and/or ideological proximity between the originating and borrowing systems (Newburn 2002).

There is also the consideration that in many (perhaps most) societies, criminal justice and penal practices are not so much created as *evolve* under the influence of an indeterminate number of interrelated and sometimes antagonistic factors. There should be no assumptions about the internal coherence of the resulting arrangements. As Garland puts it, "Having developed as a means of managing tensions, arbitrating between conflicting forces, and getting certain necessary things done, social institutions typically contain within themselves traces of the contradictions and pluralities of interest which they seek to regulate" (Garland 1990, 282). The introduction of a new institution or practice may disturb the equilibrium. This is part of the reason, no doubt, why penal policy so frequently results in unexpected and indeed perverse outcomes: for example, attempts to provide alternatives to custody can often be associated with *increases* in its use. Contradictory influences, then, can distort transfer and make outcomes inherently unpredictable.

Nor should the challenges of linguistic translation be underestimated. How, for example, is a specialized (and barely even English) term like *responsivity* be presented to an interpreter? Again, the differences among, for example, *deferred*, *conditional*, or *suspended* sentences, are significant, but linguistically subtle. Importantly, too, there is the problem of what translators and interpreters refer to as *faux amis* – false friends – words in different languages whose apparent similarity conceals substantial differences of connotation, resonance, and even sometimes denotation. Crawford (2009) gives the example of the word *community*, whose false friend is *communauté*. When *community* is rendered literally in French (*la communauté*) or German (*Gemeinschaft* is one among a number of possible translations), other and different nuances are imported that could lead to a quite different conception of *community safety*, *community crime prevention*, *community policing*, or *punishment in the community* (see also here the seminal work of Lacey and Zedner 1995). Crawford goes on to emphasize "... the need to excavate and understand the culturally specific essence of key terms and locate them within the context of wider horizons of interpretation" (Crawford 2009, 17).

Evaluating Transfer

If the success of a transfer is to be appraised in terms of the achievement of the intended goals, the first problem in evaluating transfer is in determining the intentions of those involved, which, as we have seen, are often complex, variable and changing over time. Different actors, as we have seen, are likely to have different intentions and consequently may not agree about what would count as success.

With that reservation in mind, one way of investigating the difficulties of evaluating transfer would be to consider a specific initiative and to ask what would count as success here, what as failure. For example, suppose the attempt was being made to introduce community service/unpaid work into a jurisdiction as part of a strategy to provide alternative sanctions to imprisonment (Canton 2006). The attempt would count as unsuccessful, presumably, if no community service orders were made by the courts. But another possibility is that orders are made, but instead of functioning as alternatives to custody (diverting people from imprisonment), they merely widen the net by drawing in offenders who would in the past have been dealt with through other and perhaps less burdensome and intrusive sanctions (this is an altogether likely consequence, as Cohen's (1985) work shows).

Even the size of the prison population would not resolve the question: a decrease could be attributable to other initiatives (reforming administrations, after all, rarely proceed on a single front), while an increase would not necessarily be a sign of failure (perhaps the increase would have been even steeper, but for the introduction of community service). Another possibility is that such community sanctions might be made in respect only of defendants who were powerful or privileged, with custody remaining for the poor. Again, there are other corruptions of transfer that can be envisaged: for instance, a scheme where unpaid work brought personal profit to individuals or commercial interests rather than providing a wider community benefit.

A further challenge is the timescale for evaluation. Some initiatives – the introduction of a new institution or system, of new legislation or changes to the powers of courts – could (at least in principle) be assessed quite quickly. But their longer-term significance, the extent to which they become "owned" and embedded in the criminal justice system, calls for an extended period of inquiry and for sophisticated methods of evaluation. In particular, the "soft transfer" (Newburn and Jones 2007) of new ideas, new ways of thinking about crime and punishment, is likely to have their effects much later on in time.

In summary, the effects of transfer should be evaluated and arguably this has not usually taken place with sufficient rigor. At the same time, evaluation poses formidable challenges–not only of determining effects, which may change over time, but of deciding on the criteria of evaluation in the first place.

Conclusions

Policy transfer is a topic of considerable practical importance. Whether celebrated as the dissemination of good practice and experience or denounced as the imposition of the preferences of (typically) more powerful and affluent countries upon others, transfer activity is common and only likely to increase. A principal stimulus here, which is likely to become still more compelling, is the increasing influence of supranational influences. Respect for international law and a recognition that some level of accountability to the international community is a bulwark against human rights abuses have prompted the promulgation of several treaties and conventions. These have given a mandate for supranational entities – for example, the Council of Europe – to concern themselves with the penal practices of their member states. Penal policy, in short, is no longer the sole concern of the nation state. Such developments are likely to lead to further transfer initiatives as countries wishing and/or required to change look for acceptable and effective models to copy or emulate. Another driver for transfer, as we have seen, is the claim that research can give new insights into effective practice which accordingly should be adopted everywhere – although some of the limitations and caveats to be observed have already been noted here.

Penal policy is also of great theoretical interest. Transfer projects soon expose taken–for-granted characteristics of penal practice in one's own country. Such activities also contribute to contemporary debates in comparative criminology about international convergence and divergence. It is not that convergence should always be seen as an indication of the "success" of transfer activities: it could as well be a function of countries having to confront similar challenges and predicaments (Jones and Newburn 2007). Nor should persistent difference or divergence be taken as a sign of failed transfer, especially if the objective of transfer is taken to be the development of systems and practices that truly belong to the receiving nation and not just the copying of a practice from another place (Canton 2009). But consideration of the processes of transfer and the people involved highlights, as we have seen, the place of choice and agency in penal development to counter explanations that depend on purely structural (and sometimes excessively abstract and deterministic) accounts that invoke concepts like globalization. As Cavadino and Dignan well put it in their discussion of the influences on the character of criminal justice (2006, 452): "... however many factors we incorporate into our theory, it will still not give us the whole story. Individual nations, and their cultures, histories and politics, can be just as quirky and esoteric as individual human beings."

Policy transfer could even be seen as an exercise in *applied comparative criminology*. The determinants of the character of penality and the drivers of change have usually been explored historically (notably Garland 1990). But transfer endeavors put these hypotheses to test and may therefore contribute to fuller and richer understandings of the complex network of influences that shape the institutions and practices of punishment.

Related Entries

- ▶ Comparative Legal Cultures
- ▶ Comparative Political Economy of Punishment
- ▶ Comparing Police Systems Across the World
- ▶ Cross-National Performance in Policing
- ▶ Effective Community Supervision Programming

Recommended Reading and References

Aas KF (2004) From narrative to database: technological change and penal culture. Punishment Soc 6(4):379–393

Bauwens A, Snacken S (2010) Modèles de guidance judiciaire. Paulus-Akademie, Zürich 1, 93–107

Blackburn S (2001) Being good: a short introduction to ethics. Oxford University Press, Oxford

Braithwaite J (1989) Crime, shame and reintegration. Cambridge University Press, Cambridge

Canton R (2006) Penal policy transfer: a case study from Ukraine. Howard J 45(5):502–520

Canton R (2009) Taking probation abroad. Eur Prob J 1(1):66–78

Cavadino M, Dignan J (2006) Penal policy and political economy. Criminol Criminal Justice 6(4):435–456

Cohen S (1985) Visions of social control. Polity Press, Cambridge

Crawford A (1998) Crime prevention and community safety: politics, policies and practices. Longman, London

Crawford A (ed) (2009) Crime prevention policies in comparative perspective. Willan, Cullompton

Dolowitz D, Greenwold S, Marsh D (1999) Policy transfer: something old, something new, something borrowed, but why red white and blue. Parliam Aff 52:719–730

Dolowitz D, Hulme R, Nellis M, O'Neill F (eds) (2000) Policy transfer and British social policy. Open University Press, Buckingham

Edwards A, Hughes G (2005) Editorial to special issue. Theor Criminol 9(3):259–263

Garland D (1990) Punishment and modern society: a study in social theory. Oxford University Press, Oxford

Garland D (2006) Concepts of culture in the sociology of punishment. Theor Criminol 10(4):419–447

Ignatieff M (1978) A just measure of pain: the penitentiary in the industrial revolution 1750–1850. Random House, New York

Jones T, Newburn T (2007) Policy transfer and criminal justice: exploring US influence over British crime control policy. Open University Press, Maidenhead

Lacey N, Zedner L (1995) Discourses of community in criminal justice. J Law Soc 22(1):301–320

Nelken D (ed) (2000) Contrasting criminal justice: getting from here to there. Ashgate, Aldershot

Nellis M (2007) Humanising justice: the English probation service up to 1972. In: Gelsthorpe L, Morgan R (eds) Handbook of probation. Willan, Cullompton

Newburn T (2002) Atlantic crossings: 'policy transfer' and crime control in the USA and Britain. Punishment Soc 4(2):169–194

Newburn T, Jones T (2007) Symbolizing crime control: reflections on zero tolerance. Theor Criminol 7(2):221–243

Newburn T, Sparks R (eds) (2004) Criminal justice and political cultures: national and international dimensions of crime control. Willan, Cullompton

Radzinowicz L (1999) Adventures in criminology. Routledge, London

Raynor P (2004) Rehabilitative and reintegrative approaches. In: Anthony B, Sue R, Gwen R (eds) Alternatives to prison: options for an insecure society. Willan, Cullompton

Smith P (2008) Punishment and culture. University of Chicago Press, Chicago

Stern V (1999) Alternatives to prison in developing countries. International Centre for Prison Studies, London

Worrall A (2000) What works at one arm point? A study in the transportation of a penal concept. Probat J 47(4):243–249

International Responses to Victims in Criminal Justice

Matthew Hall and Joanna Shapland
School of Law, University of Sheffield,
Sheffield, UK

Synonyms

Critical criminology; Restorative justice; Social harms; State crime; Victimology

Overview

In recent years, victims of crime have moved from the peripheries of both academic and policy discussions concerning criminal justice to become at least rhetorically "central" in modern criminal justice policies in many developed states (Hall 2010). This entry examines some of the newest territory into which victims and victimology are now moving, and, in so doing, it is argued that traditional criticisms espoused by victimologists, namely, that official recognition of victims often fails to account for victims' own perspectives, remain extremely pertinent to this debate. In addition, this entry examines how the scope of "victimization" is expanding, with many commentators now adopting a "social harm" approach to the question of victimization rather than limiting themselves to restrictive, legalistic categories. Nevertheless, it is argued that the classification of such individuals and groups as victims of "social harm" (as opposed to "criminal harm") raises the concern that criminal justice systems may shirk their responsibilities for such victims. It can be argued therefore that notions of social harm represent a challenge to justice processes to better incorporate these "new" victims through the continued development of victims' rights and restorative justice.

Discussion

Expanding Notions of Victimhood

"Criminal Harm" or "Social Harm"?

Over time, the ambit of both *academically* and *politically* recognized "victimhood" has tended to expand. That said, much of victimology (and criminology) continues to be centered on notions of victimization espoused by official sources. Indeed, some go further to argue that society's narrow conception of victimization is brought about by selective definitions of crime, construed for political purposes (Elias 1986). Such arguments have led to the development of "critical victimology" and its expanded notions of victimhood beyond straightforward, criminal classifications to include categories of "real, complex, contradictory and often politically inconvenient victims" (Kearon and Godey 2007, p. 31). This in turn has prompted some authors to approach "victimization" from a cultural perceptive and to focus on "social harms" as experienced by victims and their communities, as opposed to harms necessarily caused by breaches of the criminal law (Hillyard and Toombs 2003).

The conceptualization of victims as those who have suffered harm (as opposed to more technical, legal, or prescriptive definitions) has two key implications. Firstly, it gives scope for a wide cross section of individuals or organizations to be included within the ambit of victimhood. Criminal law, as argued by Hulsman (1986), "fails to capture the more damaging and pervasive forms of harm" (Hillyard and Toombs 2003, p. 12). Focusing on harm thus has the potential to include the often legally ambiguous activities carried out by big corporations or the state, and also activities which the state or its agencies choose not to regulate. This approach also allows for the consideration of "mass harms," whereas traditional criminology has struggled to fully embrace mass victimization, largely remaining focused on the individual. Of course, "harm" similarly is a fuzzy concept, with strong cultural roots, and definitions of it from a state's, a victim's, or a community's point of view are likely to differ. This notwithstanding, a social harm approach also potentially allows victims to be self-defined, something which at present is usually reserved for a small subset of victims (e.g., victims of hate crimes) – though victimization surveys take a victim's eye view. This implies that prosecutors should seek the view of individuals and groups who feel they have been harmed.

The adoption of "harm" as the central element of victimization has been witnessed in many jurisdictions (Hall 2010). In this respect, many jurisdictions have followed the definition set down by the 1985 UN Declaration of Basic Principles of Justice for Victims of Crime and Abuse of Power (hereafter the 1985 UN Declaration):

> 'Victims' means persons who, individually or collectively, have suffered harm, including physical or mental injury, emotional suffering, economic loss or substantial impairment of their fundamental rights, through acts or omissions that are in violation of criminal laws operative within Member States, including those laws proscribing criminal abuse of power. (para.1)

The 2001 EU Framework Decision of the Council of the European Union on the Standing of Victims in Criminal Proceedings [(2001/220/JHA, at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32001F0220:EN:NOT) follows a similar model (Art.1(a)] as does the South African Victims' Charter and its Dutch equivalent. The US Justice for All Act of 2004 defines a victim as "a person directly and proximately harmed as a result of the commission of a federal offense" (18-USC-3771). This includes victims of white collar and financial crimes.

There are numerous examples of this broadening of victimhood among policy makers and government actors (see Hall 2010). In this entry, two pertinent examples are considered, victims of state actions and victims of environmental crime, to illustrate the complexities inherent in this development and, in particular, the difficulties when such developments occur without direct consultation with victims.

Victims of State Crime and the International Criminal Court

While much of the policy and academic focus on victims has rested on traditional breaches of the criminal law, there has also been growing awareness of victimization at the hands of the *state*. "State crime," itself a markedly underdeveloped concept, is generally used to describe criminal actions carried out by state employees, such as the police, as well as by governments. Academic focus on *victimization* by the state or state employees has been similarly limited, although it was included under the 1985 UN Declaration in the guise of victims of "abuse of power." The inclusion of *omissions* under the definition of "abuse of power" used by the Declaration clarifies that a state is responsible for failure to regulate activities within its borders which lead to transboundary harm. This is particularly significant given the proliferation of cross-border crimes such as human trafficking and international terrorism.

Kauzlarich et al. (2001) characterize victimization by the state as a manifestation of attempts to achieve organizational, bureaucratic, or institutional goals of state agencies. They note that victims of state crime tend to be among the least socially powerful, are often blamed for their suffering, and rely on the victimizer, an associated institution, or civil social movements for redress. Such victims also tend to be easy targets for repeat victimization. From the victimizer's perspective, the nature, extent, and harmfulness of institutional policies tend to be ignored or misunderstood. If suffering and harm are acknowledged, the authors contend that it is often neutralized through a sense of "entitlement," which the authors illustrate by drawing on what they see as "the long history of U.S. abuses in Latin and Central America" (p. 185).

In relation to victims of *crime*, the 1985 UN Declaration has prompted significant criminal justice reform in multiple jurisdictions (Hall 2010). Notwithstanding this, Bottigliero (2004) argues that the victims' movement (at activist, academic, and public policy levels) has focused on what she calls "ordinary" crimes like assault and theft, which can be dealt with in the domestic context, while ignoring the victims of more "severe" crimes such as war crimes, human rights violations, or genocide. At the international level, one significant development is the creation in 2002 of the International Criminal Court (ICC). Constituted by the Rome Statute of 1998, the purpose of the ICC is to try war crimes, genocide, and crimes against humanity which, by their nature, are often perpetrated by those in positions of authority within governments. In most cases such activities will also constitute breaches of domestic criminal laws, but when such crimes are committed within the ambit of wider atrocities by representatives of a state, they are rarely prosecuted. This means that the ICC rarely deals with "abuses of power" as defined by the 1985 UN

Declaration, which only includes activities *not* officially prescribed as crimes, although Wemmers (2009) characterizes the inclusion of "abuse of power" within the Declaration as "a source of inspiration" (p. 124) for the ICC.

Significantly, the Rome Statute legislates for the participation of victims within the ICC process, and the court has adopted a judicial code of ethics which includes judicial obligations to have regard to the interests of victims. The court has also ruled that the definition of "victims" within the Statute is a wide one and "cannot be interpreted as limited to certain categories of individuals" (ICC-01/04-01/06-1119, para.27).

From a victimological perspective, the ICC is a rare example of a discrete judicial system developed *from the outset* with an appreciation for the arguments of the victims' movement. The Court's Rules of Procedure and Evidence contain many safeguards to what are called the rights of victims: including the right to participate in proceedings as a party with legal representation. The rules often bring together an "interests of justice" test with an "interests of victims" test. Perhaps most significantly, the rules provide the general principle that "a Chamber in making any direction or order, and other organs of the Court in performing their functions…shall take into account the needs of all victims and witnesses" (Rule 86). There is however a limitation inherent in the Rome Statute in that the court is not *required* to establish what may be "in the interests of victims" through consultation with the victims themselves but rather, under Rule 93, *may* seek the view of victims or their legal representatives. This leaves the procedure open to the criticism that judges are only taking into account what *they think* the interests of victims may be.

Bottigliero (2004) argues that the infrastructure set up by the ICC begins the process of addressing the needs of victims of state crime, particularly their need for restitution. Of course, the true impact of the ICC on victims will only become clear as the number of cases it deals with increases. On this point, Wemmers (2009) conducted qualitative interviews with court personnel to determine how they felt its engagement if victims would operate. The results indicated serious engagement with the "victim issue" as well as a recognition that the success of the court as a whole to a large extent relied on victims' perceptions, especially at the community level of it as a credible and legitimate institution. A view generally held was that the true test of the ICC was whether it can "effect real changes in the communities, such as restoring the safety of victims and preventing further violence by arresting and punishing offenders" (p. 124). One theme which did not emerge in many of the interviews, but which Wemmers sees as crucial, is victims' relationship with their legal representation at the ICC given the reality that the vast majority will never attend the court.

A further issue raised by Wemmers (2009) and examined in more detail by Bottigliero (2004) concerns the selection of cases for trial by prosecutors working at the ICC. As Wemmers notes:

> The selection of cases by the Prosecutor is a central point in the decision-making process. It is vital that the Court be seen as neutral and fair and that its choice of cases is both a-political and representative of the atrocities committed. (p. 225)

Clearly this reflects a concern that prosecutions at the ICC may be seen as politically motivated. Bottigliero (2004) however has shown that major selection by the prosecutor as to which cases should proceed to the court seems to be based on which cases are most likely to succeed rather than the needs of victims. Prosecutors at the ICC are bound to consider the position of victims, and Article 53(c) of the Rome Statute notes that a prosecutor may decline to pursue a prosecution if it is in "the interests of victims" to do so. The question remains however as to whether prosecutors will approach victims directly for their views or work on their own assumptions as to their needs and the perceived seriousness of the crime.

Environmental Victims

In recent years, progress made in our understanding of the causes of environmental degradations has posed a number of challenges for lawyers, as questions are increasingly raised concerning the responsibility of corporations and states for so-called environmental harms. Indeed, following the previous section, the increasing prevalence of "green" debates among criminologists has rekindled notions of "mass victimizations" and state crime. Lynch et al. (2010), for example, have exposed complex interactions between state and corporate liability following what they call the "politicization of global warming." White (2008) has discussed state crime in relation to the environmental effects of the use of depleted uranium munitions in the Gulf wars. Of course, much environmental crime is committed by corporations *against* governmental regulations or legislation. In such cases the question becomes one of enforcement, and whether the prosecuting agents of the state approach such transgressions in the same manner as they do other crimes. A large body of literature now suggests that the enforcement of regulatory or criminal instruments against corporate offenders is highly variable (Hutter 1997), especially in relation to environmental degradation (Bell and McGillivray 2008).

"Environmental harms" can include direct impacts on human health and economic impacts on a range of industries (fishing, agriculture, tourism, and significant social and cultural damage to traditional ways of life). The implications of environmental harm for issues such as energy security and food security are equally significant. As a consequence, criminal justice systems (at the national and international levels) are increasingly called upon to incorporate "environmental victims" into their procedures. This is a challenging proposition, given the vested economic and political interests inherent to the industrial practices which can foster environmental damage (although it should not be assumed that all industrial or commercially viable practices lead to environmental degradation). As such, environmental victimization encapsulates many of the problematic issues addressed by the critical school and covered by notions of "social harm."

Unfortunately, specific empirical research even on the *human* victims of environmental degradation is scarce, although the first call for the development of "environmental victimology" came as early as 1996. For Williams (1996), victimology represents a solution to what he perceives as an environmental justice movement dominated by activists and relying too heavily on subjective understandings of victimhood. Certainly "environmental justice" is usually acknowledged as a wide concept which emphasizes the involvement of people and communities in decisions (apparently meaning state decisions) which impact upon their environment. "Environment" tends to be defined broadly to include cultural norms, values, rules, regulations, and behaviors (see also Bryant 1995, p. 6). "Intergenerational justice" is often cited as a core component of the environmental justice model, highlighting responsibility to future descendents.

Examples of how criminal justice systems are now being obliged to adapt to meet the challenges of environmental victimization can be drawn from a number of sources. At a national level, for example, the US Crime Victims' Rights Act 2004 has been applied to victims of environmental harm in *Re: Parker; U.S. v U.S. District Court and W.R. Grace & Co.* Nos. 09-70529, 09-70533 (9th Cir). In Europe the EU has adopted Directive 2008/99/EC on the protection of the environment through the criminal law. Though the enforcement of environmental legislation by means of the criminal law is by no means novel, it has often played a secondary role to administrative sanctions and civil penalties. Indeed, Bell and McGillivray (2008) have drawn on the "enforcement pyramid" posited by Ayres and Braithwaite (1992) to describe enforcement tactics of many jurisdictions in relation to environmental "crime." Essentially this pyramid puts "persuasion" at its apex (and, interestingly, does not have criminal recourses there), principally because revocation of a company's license to operate is in fact far more damaging than a relatively small fine. The 2008 Directive is thus indicative of a hardening of attitudes over breaches of environmental law, albeit the response of the EU so far to environmental crime has neglected the impact of such crime on victims. Potential for greater involvement of such victims can be identified within the earlier 1998 Council of Europe Convention on the Protection of Environment through Criminal Law. Though this treaty has not yet entered into force, and indeed has secured little support, Article 11 provides for the participation of environmental groups in relevant criminal proceedings. This has the potential to be a ground-breaking article, raising the possibility of opening up environmental criminal proceedings to wider participation.

At the international level, the most progressive source of legally binding environmental obligations is found in the 1998 UNECE ("Aarhus") Convention on Access to Information, Public Participation in Decision-Making, and Access to Justice in Environmental Matters. The convention requires governments to bring individuals who may be affected into the decision-making process when environmental issues are at stake. What is significant about the convention is that, almost uniquely within international law, members of the public are able to refer possible breaches of their rights under the Convention to its Compliance Committee. Of course, while the Aarhus convention offers something of a "way in" to the international legal order for the individual in practice, this compliance mechanism can be subject to criticisms. There is a lack of real compulsive power on behalf of the Compliance Committee to really address victims' complaints and ensure restitution/apologies from perpetrator states are forthcoming. That said, there is presently a dearth of empirical research on what victims of environmental crime might actually want from a criminal justice (or other) process (Williams 1996). Consequently, moves to increase the attention paid to such victims in many jurisdictions is once again progressing without reference to the victims they are purporting to help.

Addressing the Needs of Victims

To summarize the preceding section, the scope of officially recognized victimhood is expanding. However, it appears to be doing so without asking victims what their perceptions and needs might be. The second section of this entry examines the development of victims' rights and the place of victims in restorative justice processes, both of which provide cogent examples of a more evidence-based process and, as such, shed light on how the above developments might be adapted.

Victims and Criminal Justice: The Perennial Question of Rights or Services

Should victims have legislated rights which they can use to demand better treatment in relation to criminal justice – or should criminal justice agencies be mandated to provide services to improve the victim experience? The question of "rights or services" has been a perennial one since the first empirical research into victims and criminal justice, in both the USA and the UK, showed how they were essential to the system's operation but ignored in terms of their needs and expectations. Subsequent research has continued to reinforce the same needs and expectations of victims, but has also documented the reluctance of criminal justice personnel and policy makers to provide for those needs, partly through a continued focus on professionalism and reducing delay, partly through a misapprehension that meeting victims' expectations disadvantages offenders. In fact, the research on offenders' views shows they would also like to be treated with respect, given information on what is happening and generally feel they were being treated fairly – just like victims.

The slow pace of change has spurred international bodies to promote action to produce a more helpful environment for victims. Following a review of the "patchy" implementation of the 2001 Council Framework Decision, in May 2011 (see Press Release, "Victims" rights: Frequently asked questions,' at http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/11/310&format=HTML&aged=0&language=), the European Commission is introducing a Directive on minimum standards for victims, the EU's Justice Commissioner, Vice President Reding saying, "Victims of crime need respect, support, protection and to see that justice is served. That is why I am putting victims at the heart of criminal justice in the EU by making sure they can rely on minimum rights and support" (Press release, "European Commission ensures better protection of crime victims," 18 May 2011, IP/11/585, at http://europa.eu/rapid/pressReleasesAction.do?reference=IP/11/585). The EU sums up the needs of victims as being to be recognized, to be treated with respect and dignity, to receive protection and support, and to have access to justice, compensation, and restoration.

It is suggested that the answers to the question of "rights or services?" depend on ideology and on evaluated effectiveness. Criminal justice and legal scholars, seeing criminal justice as primarily a matter for the nation-state and the offender, are loathe to grant any more than service rights to victims, primarily focusing the burden of providing such rights on personnel at the entry to criminal justice: the police, welfare agencies, and volunteers, such as Victim Support. Given that, for victims, most contact with criminal justice is with the police (and Victim Support) – because offenders are rarely apprehended – this is indeed the way to supply most victims with most of what they need. It is underlined by social activists who wish to widen the idea of victim to embrace other harms, including traffic accidents and disasters, as discussed above. For the police or support agencies to help victims, what is key is political (rather than justice) prioritization to victims rather than the multitude of other tasks required of the police and money. Both, however, are currently rather in short supply in this time of recession and fears over public disorder and terrorism.

It is argued that procedural rights (see Edwards (2004) for distinctions between service rights and procedural rights) for victims become more essential further into the criminal justice system. Criminal procedure in all legal systems is shaped and operated through legislated rules. If victims are to receive consideration during prosecution and at court (and in relation to compensation), then the rules need to provide them with those means, and key decision makers within criminal justice must be appointed to ensure they are operationalized. It is notable that the relatively recent provisions for vulnerable witnesses in courts in England and Wales have come into force and been effective for victims in a short time frame (Hamlyn et al. 2004) – with few qualms (other than the technical ones of shortage of court space and learning the technology). The provisions include familiarization visits and giving evidence via video link or from behind screens. They were enacted through legislation (Youth Justice and Criminal Evidence Act 1999), and decisions are made by the presiding judge.

In contrast to the speed of action over vulnerable witnesses – which may be linked to them being for *witnesses*, rather than victims perse – provisions which depend upon individual police officers, prosecutors, or politicians taking action seem to be more tardy and patchy. Victim personal statements (statements made by victims about the effects of the offense on them, also called victim impact statements (in the USA, Australia, etc.) or victim effect statements), for example, though mandated in case law and by government (Shapland and Hall 2010), depend on the police officer giving them to the victim at the right time, and in England and Wales, less than half of victims recalled being offered that opportunity (Roberts and Manikis 2011). Nonetheless, of those offered the opportunity, the majority took it, victims of more serious offenses being particularly likely to participate. Research in Canada, the USA, Australia, and the UK indicates that those who made a statement were generally satisfied (see summary in Roberts and Manikis 2011). Victims, in the studies carried out in Australia, Canada, and the UK, seem to make a statement primarily for expressive reasons (desiring a voice at court), though instrumental reasons, such as affecting sentence, are also important (Roberts and Manikis 2011). Though the introduction of victim personal statements for nonfatal cases seems to have been welcomed by the judiciary (as providing more information), it does not often seem to have affected sentence to a significant extent. However, this would be expected if judges are sentencing primarily on a deserts basis to an "average" range of sentence (only more unusual cases can be affected). In death penalty cases, its use has been far more controversial. Paternoster and Deise (2011), in a randomized controlled simulation experiment with potential jurors in the USA, found that jurors who viewed a victim impact statement were more likely to impose the death penalty, because they sympathized with the victim's family. This has been the fear, but of course most jurisdictions do not have sentence pronounced by juries – and it is highly unlikely professional judges would do the same (Shapland and Hall 2010).

There has always been a dysfunctional tension between criminal justice system personnel's reluctance to provide services (or rights) to victims and their dependence on victims, both for the immediate case and for political support. The public's confidence in the system is mediated through victims' views, with a key determinant of comparative confidence between countries being victims' treatment (van Dijk et al. 2007). This has recently been tested empirically. Bradford (2011), analyzing the national victimization survey for England and Wales, the British Crime Survey, found that victims who did have contact with Victim Support had more favorable views of the fairness of the criminal justice system and higher levels of confidence in its effectiveness. The link between contact with Victim Support and confidence in the criminal justice system was mediated by trust in the fairness of the system; it appears that if victims have someone who listened to their concerns (and may take action in relation to them), they trust the system more and have greater confidence in it. This continues to support Tyler's theory of procedural justice: that those members of the public who have contact with the system view the system through the lens of their perceptions of how fairly it takes decisions and how it treats them (with respect, sensitivity, and dignity scoring highly).

It is interesting that the European Commission, introducing the proposed Directive, mentions the need for both services and rights, though it is clearly moving towards procedural rights. It specifies its reasons for needing a Directive as improving provisions on an equal and nondiscriminatory basis, enhancing trust in the justice system, improving the quality of justice, that minimum standard must apply in all Member States, and that "fundamental rights must be respected" (Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions, COM(2011) 274, at http://ec.europa.eu/justice/policies/criminal/victims/docs/com_2011_274_en.pdf), in which it links victims' rights to the EU Charter and the European Convention on Human Rights, in terms of victims' human dignity, private and family life, and property.

Victims and Restorative Justice

The European Commission refers to restorative justice, in its Communication (COM(2011)274 final) in its section on compensation and restoration:

> Restorative justice, which is a relatively new concept in criminal proceedings, goes beyond purely financial compensation to focus on the recovery of the victim. As an alternative to, or in combination with formal justice, it aims to restore victims to the position they had before the crime by giving them, if they so wish, an opportunity to confront their offenders face to face and for the offenders to take responsibility for their acts. (p. 6)

Reparation has certainly been one key element behind the growth of restorative justice in many countries, particularly in mainland Europe and Scandinavia. However, restoration of the victim, in the fuller sense of acknowledging the victim has been damaged by the offense, showing how the victim and others have been affected, allowing apology by the offender and determining what best to do next, has been one of the key purposes behind mediation and conferencing as opportunities for interaction between victim and offender.

From a recent survey of key knowledgeable informants, it seems that mediation (victim-offender mediation: in which a mediator meets with victim and offender) is now available in at least 39 countries worldwide, including almost all of Europe (Zinsstag et al. 2011). Conferencing (in which the victim, offender, and mediator are joined by victim and offender supporters, and sometimes community representatives) is now available in at least 48 countries, more frequently outside Europe. Though more commonly used (and legislated for) in relation to young offenders, both mediation and conferencing are growing in relation to adult offenders as well. Mediation is starting to be used widely by the police, more informally, as part of neighborhood or community policing (Gunther-Moor et al. 2009), where police officers either act as mediators themselves or refer cases to community mediators or specialized mediators employed by the police. Mediation can thus be available at many stages of the criminal process (or even outside it altogether): as diversion from prosecution, presentence, post-sentence, or while the offender is in custody. Though widespread, at least in theory, and increasingly backed by both European and national authorities (see the Council of Europe's Recommendation No. R(99)19 on Mediation in penal matters), its implementation has, however, like many initiatives for victims, been patchy, constrained by money and a lack of trained mediators/facilitators.

Restorative justice needs to be voluntary, both for the victim and the offender. What proportion of victims, then, when offered it, decides to take part? Mostly, victims will be offered one "take it or leave it" choice between participating in restorative justice and not participating. It is uncommon for victims to be offered several formats (e.g., conferencing, a direct mediation meeting or "shuttle" mediation, in which the mediator passes questions and answers between victim and offender), though increasingly, schemes are offering several possibilities, depending on the type of offense and the circumstances of the parties. In England and Wales, a recent large-scale evaluation of three schemes found that the percentage of victims offered restorative justice where they agreed to participate was between 30 % and 89 % (Shapland et al. 2011). Victims were more likely to agree where the case involved a young offender (all these sites had agreement rates of 75–89 %), but there was no difference between victims offered mediation and victims offered conferencing. Victims offered indirect mediation were more likely to choose it over a direct meeting with the offender, but the overall rate for conferencing was the same as for indirect and direct mediation combined. Even very serious offenses with adult offenders offered conferencing presentence in the higher courts found a majority of victims prepared to participate. Restorative justice is not, though, suitable for offenses in which there are considerable power imbalances between offender and victim (domestic abuse, some sexual assault cases).

The evaluation results for victim reactions to restorative justice are almost uniformly positive – providing that the offender acknowledges responsibility for the offense and the victim is invited to attend. Processes and the extent of victim participation vary considerably between schemes and countries. In early New Zealand family group conferencing, victims were not always informed or invited at helpful times, but in Northern Ireland youth conferencing, based on the New Zealand model, victims have attended in around 70 % of cases (Shapland et al. 2011). In the Thames Valley diversionary police-run conferencing, victims also tended not to be invited (Hoyle et al. 2002). In family group conferencing, victims may be asked to leave after they have talked about the effects of the offense and before the family and young offender decide with the mediator what should happen thereafter. In contrast, in conferencing based on the Australian RISE model, victims (and their supporters) are active participants in all stages of the conference, discussing what happened, what effects it had, and what should then be done (Shapland et al. 2011). In this model, the conference is intended to end, where possible, with an outcome agreement signed by all present (and 95 % of conferences in the evaluation did so). Despite all these differences between models, victims in all these evaluations were positive about their experiences. However, it is highly likely that these positive results depend upon the process being suitable for the seriousness of the offense. If restorative justice were to be used to replace criminal justice as a diversion measure (as opposed to being offered in addition, as is normal), then particularly for serious offenses, victims would be highly likely to feel offended.

There has been a significant amount of evaluation of restorative justice, most of it concentrating upon whether it reduces re-offending, but including victims' views and the effects on victims. In terms of satisfaction ratings, Latimer et al. (2005) did a meta-analysis of 22 studies which had a control group of 35 schemes and found there was significantly more satisfaction among those who had experienced restorative justice. Sherman and Strang (2007), in a further meta-analysis, concluded that restorative justice with adult offenders reduced crime victims' posttraumatic stress symptoms and related costs, provided both victims and offenders with more satisfaction with justice than did criminal justice alone, and reduced crime victims' desires for violent revenge against their offenders. It is difficult to measure whether restorative justice (or any other initiative taken in conjunction with criminal justice) reduces the harmful effects of crime on victims, because many psychological, social, and financial perceived effects of crime on victims reduce with time (Shapland and Hall 2007), so one would be measuring the "extra" effect of restorative justice beyond that of time and other justice processes. Victim perceptions, however, are that it is helpful in reducing negative effects of the offense as well as enabling victims to have their questions answered about the offense and to communicate with the offender. In the major recent evaluation of conferencing in England and Wales, where conferencing occurred in parallel to the normal criminal justice process and cases were randomly allocated, victims participating in restorative justice said they felt significantly more secure (Shapland et al. 2011). In addition, over half the victims said that it had helped them have closure, with another 20 % saying it had done so to some extent. Victims of more serious offenses were as likely to feel closure and were significantly more likely to feel satisfied with the criminal justice system as a whole.

The extent of satisfaction and the generally positive effects of well-run restorative justice, particularly conferencing, are now so prevalent in evaluations that, in the evaluation authors' view, it would be difficult ethically to justify running further random controlled trials between the presence or absence of restorative justice. It would be difficult to put victims into a control group and so deny them the benefits of restorative justice. That evaluation also found that, at whatever stage of criminal justice the restorative justice was offered, participants were fine about it being done then. However, it is not known exactly which forms of restorative justice are best to offer for which groups. Having a "future-oriented" stage to the restorative justice where participants can discuss what should happen next seems to be important for both serious offenses and neighborhood disputes (Shapland et al. 2011). Supporters bring extra voices and resources to the debate and can help to mitigate any power imbalances (Strang 2002; Shapland et al. 2011). Indirect mediation seems to lead to less satisfaction than a direct meeting, because communication, prized by victims, is circumscribed (Shapland et al. 2011) – but if this is the only kind of contact victims want, should it be withheld?

Conclusion

The development of scholarly interest in victims of crime, as well as policy interest among the world's governments, looks set to continue well into the new century. In the above paragraphs, a number of key areas of development have been set out which will no doubt raise a whole host of questions and challenges for victimologists and criminologists in the coming years. A key concern however is that at least some of these developments are occurring in the absence of detailed research or engagement with victims themselves. Indeed, it is truly fascinating how in both state crime and environmental crime fields, the definition of victims by victims themselves is largely being ignored. In these two fields, detailed victimization studies are absent possibly because they would mean confronting what victims themselves say and think, rather than "using" them as exemplars of damaged people. Of course, in this sense, victims are threats not only to lawyers and the state but also to activist groups wishing to attack states if, for example, it turns out that they desire only respectful treatment, information, understanding, and an apology as opposed to more retributive outcomes. The research into what kind of services or rights victims actually want, as well as their response to restorative justice processes, seems to bear out this hypothesis – although actual answers can only follow detailed empirical work. In sum, while the expansion of victimhood and the development of concepts like "social harm" may benefit those victims still ignored by governments and criminal justice actors, it also places obligations on those actors, and the academy, to find out what these people actually need – and only then consider how best it may be achieved and the extent to which it is valued by victims themselves.

Related Entries

- ▶ Bullying
- ▶ Crimes of Globalization
- ▶ History of Restorative Justice
- ▶ Monitoring and Evaluation of Restorative Justice
- ▶ Multiple Victims and Super Targets
- ▶ Resistance to State Crime
- ▶ Restorative Justice and Practice
- ▶ Restorative Justice and State Crime
- ▶ State Crime
- ▶ State-Corporate Crime
- ▶ Victim Input at Sentencing
- ▶ Victimization, Gender, and the Criminal Justice System
- ▶ Victims and Restorative Justice
- ▶ Victims and the International Criminal Court
- ▶ Victims of Corporate Crime
- ▶ Victims of State Crime

Recommended Reading and References

Ayres I, Braithwaite J (1992) Responsive regulation: transcending the deregulation debate. Oxford University Press, New York

Bell S, McGillivray D (2008) Environmental law, 7th edn. Oxford University Press, Oxford

Bottigliero I (2004) Redress for victims of crimes under international law. Martinus Nijhoff, Leiden

Bradford B (2011) Voice, neutrality and respect: use of victim support services, procedural fairness and confidence in the criminal justice system. Criminol Crim Justice 11(4):345–366

Bryant B (1995) Environmental justice. Island Press, Boston

Edwards I (2004) An ambiguous participant: the crime victim and criminal justice decision-making. Brit J Criminol 44:967–982

Elias R (1986) The politics of victimization: victims, victimology and human rights. Oxford University Press, New York

Gunther-Moor G, Peters T, Ponsaers P, Shapland J, van Stokkom B (eds) (2009) Restorative policing. Journal of Police Studies 2/Cahier Politiestudies Cahier no. 11. Maklu, Antwerp

Hall M (2010) Victims and policy making: a comparative approach. Willan, Cullompton

Hamlyn B, Phelps A, Turtle J, Sattar G (2004) Are special measures working? Evidence from a survey of vulnerable and intimidated witnesses, Home office research study 283. Home Office, London

Hillyard P, Toombs D (2003) Introduction. In: Hillyard P, Toomb D, Pantazis C (eds) Beyond criminology: taking harm seriously. Pluto Press, London, pp 1–9

Hoyle C, Young R, Hill R (2002) Proceed with caution: an evaluation of the Thames valley police initiative in restorative cautioning. Joseph Rowntree Trust, York

Hulsman L (1986) Critical criminology and the concept of crime. In: Bianachi H, van Swaaningen R (eds) Abolitionism. Towards a non-repressive approach to crime. Free University Press, Amsterdam

Hutter B (1997) Compliance: regulation & environment. Clarendon, Oxford

International Criminal Court (2002) Rules of procedure and evidence. International Criminal Court, The Hague

Kauzlarich D, Matthews R, Miller W (2001) Toward a victimology of the state. Crit Criminol 10:173–194

Kearon T, Godey B (2007) Setting the scene: a question of history. In: Walklate S (ed) Handbook of victims and victimology. Willan, Cullompton, pp 17–36

Latimer J, Dowden C, Muise D (2005) The effectiveness of restorative justice practices: a meta-analysis. Prison J 85:127–144

Lynch M, Burns R, Stretesky P (2010) Global warming and state-corporate crime: the politicalization of global warming under the Bush administration. Crime Law Soc Change 54:213–239

Paternoster R, Deise J (2011) A heavy thumb on the scale: the effect of victim impact evidence on capital decision making. Criminology 49(1):129–161

Roberts J, Manikis M (2011) Victim personal statements: a review of empirical research. Report for the Commissioner for Victims and Witnesses in England and Wales. Ministry of Justice, London. Available at http://www.justice.gov.uk/downloads/news/press-releases/victims-com/vps-research.pdf. Accessed 12 Jan 2012

Shapland J, Hall M (2007) What do we know about the effects of crime on victims? Int Rev Victimol 13:175–218

Shapland J, Hall M (2010) Victims at court: necessary accessories or principal players at centre stage? In: Bottoms AE, Roberts J (eds) Hearing the victim: adversarial justice, crime victims and the state. Willan, Cullompton, pp 169–188

Shapland J, Robinson G, Sorsby A (2011) Restorative justice in practice. Routledge, London

Sherman L, Strang H (2007) Restorative justice: the evidence. Smith Institute, London

Strang H (2002) Repair or revenge: victims and restorative justice. Clarendon Press, Oxford

Van Dijk J, van Kesteren J, Smit P (2007) Criminal victimization in international perspective. WODC report 127. Ministerie van Justitie, The Hague

Wemmers J-A (2009) Victim and the International Criminal Court (ICC): evaluating the success of the ICC with respect to victims. Int Rev Victimol 16:211–252

White R (2008) Crimes against nature: environmental criminology and ecological justice. Willan, Cullompton

Williams C (1996) An environmental victimology. Soc Sci 23:16–40, reprinted in R White (ed) *Environmental Crime: A Reader*, Cullompton: Willan Publishing at 200-22

Zinsstag E, Teunkens M, Pali B (2011) Conferencing: a way forward for restorative justice in Europe. Report to the European Union. European Forum for Restorative Justice, Leuven

International Sentencing

Barbora Hola
Department of Criminology, VU University
Amsterdam, Amsterdam, The Netherlands

Introduction

International crimes, such as genocide, crimes against humanity, and war crimes, are manifestations of large-scale and serious violations of human rights that have been defined as the most serious crimes of international concern. As noted by *Mettraux* it would be hard to identify crimes more difficult to sentence than international crimes (Mettraux 2005). This is not only due to the atrocious nature of international crimes but also due to their collective, systematic character, often involving state authorities and a huge number of perpetrators. In the last decade a number of international institutions have been established to prosecute and sentence perpetrators of international crimes. Compared to national criminal law systems, there is no "international criminal code" to govern the functioning of these courts. Each tribunal has its own statute outlining in general terms its jurisdiction, and the various tribunals exercise their mandates rather independently with no formal and institutional links to each other. With respect to sentencing, provisions of the statutes, and rules of procedure and evidence, the written law establishing the tribunals and governing their functioning, is usually very sketchy providing judges with only very loosely defined limits to their sentencing discretion. The positive law does not pronounce any sentencing rationales or general principles of sentence determination. It does not contain any penalty structure in the sense of minimum or maximum penalties for individual offences. Usually, no list of relevant aggravating and mitigating factors is provided. As opposed to many domestic jurisdictions, legislation thus plays only a very marginal role at the international level, and one has to examine judicial decisions in order to disentangle the phenomenon of international sentencing.

This contribution will provide an overview of the current practice of sentencing of international crimes. First, it will briefly discuss sentencing practices at various international courts and tribunals that have been set up to prosecute international crimes. Then, the most controversial issues, such as sentencing goals at the international level, hierarchy of different categories of international crimes, and sentencing consistency, will be discussed. The final section will provide suggestions for further research. It will be demonstrated that in the past few decades, sentencing of international crimes at international courts and tribunals has developed rapidly and inspired scholarly discussions and controversies. This attention, however, is still rather marginal. Compared to other aspects of international criminal law, sentencing has been relatively neglected by academics and commentators. There are, therefore, many unexplored areas of sentencing of international crimes that call for further research and inquiry.

Fundamentals of International Sentencing

Plurality of International Criminal Courts and Tribunals

The scholars date the history of international criminal law centuries back, however, the first time an international tribunal held an individual responsible for international crimes came only after World War II. In 1945 the International Military Tribunal in Nuremberg (IMT) and International Military Tribunal for Far East in Tokyo (IMTFE) were established to punish the major Axis war criminals. Due to the stalemate at the international arena during the period of Cold War, there were no similar institutions created. There were, however, some ongoing efforts to set up an international criminal court. After the fall of the Berlin wall ending the period of Cold War, the picture changed significantly. Media coverage of stories of concentration camps in Bosnia, showing pictures which evoked memories of the Holocaust, and of the wholesale slaughter committed in Rwanda in the presence of the UN peacekeepers in the early 1990s caused a public outcry and led to demands that something should be done. The UN Security Council reacted by the establishment of the International Criminal Tribunal for the Former Yugoslavia ("ICTY") and International Criminal Tribunal for Rwanda ("ICTR") – the first ad hoc international criminal tribunals. Shortly thereafter, the permanent International Criminal Court ("ICC") was set up at the Rome Conference in 1998. The increased interest in achieving justice and thus preventing international crimes in the future led to the establishment of many other so-called internationalized courts and tribunals such as for Sierra Leone, Kosovo, East Timor, Cambodia, or Lebanon. As opposed to the ICTY, ICTR, and ICC, all the internationalized courts have a much stronger "national element" (hybrid jurisdiction over international and domestic crimes; mixed composition with international and domestic judges, prosecutors, and defense attorneys; applicability of international and domestic law). In this respect, the ICTY, ICTR, and ICC can be perceived of as the only "purely international criminal tribunals." Furthermore, sentencing of international crimes is exercised not only at the international level by international courts and tribunals but also by many domestic courts in nation states.

International Sentencing Law and Practice

IMT and IMTFE

Traditionally, statutes of international criminal tribunals have paid only very limited attention to sentencing. The Charters of the postwar military tribunals included only one provision regulating sentencing basically authorizing "death or such other punishment as shall be determined [. . .] to be just" (Art 27 IMT Charter; Art 16 IMTFE Charter). Also, the case law of these tribunals and successor trials held by various national courts in the aftermath of World War II left only very few sentencing guidelines (Schabas 2006). The IMT's judges recognized a wide range of mitigating factors for sentencing of international crimes, including superior orders, age, position in the military hierarchy, efforts to reduce victims' suffering, and duress (Schabas 2010). Otherwise, however, sentencing argumentation of the post-World War II judges was not very developed. The IMT in Nuremberg rendered 19 sentences (three defendants were acquitted) while the Tokyo tribunals convicted 25 individuals. Twelve Nuremberg defendants and seven Tokyo defendants were sentenced to death. The IMT convicted three other defendants to life imprisonment while the IMTFE sentenced 16 individuals to life in prison. Otherwise, determinate sentences ranged from 7/10 (Tokyo/Nuremberg) to 20 years imprisonment. More than 100 defendants were sentenced during the 12 succession trials held in Germany by occupying powers; among them, 12 defendants were executed (Heller 2011).

ICTY and ICTR

International sentencing law and practice have become more sophisticated after the Cold War with the jurisprudence of the ad hoc international criminal tribunals. The regulatory framework for sentencing is almost identical for both tribunals. Applicable penalties are limited to imprisonment, and when determining the terms of imprisonment, judges shall have recourse to the local courts' practices regarding prison sentences (Yugoslavian or Rwandese). Articles 24/23 ICTY/ICTR Statutes contain very general instructions as to what factors should be taken into account in imposing sentences: the gravity of the offence and the individual circumstances of the convicted person. What is actually meant by the "gravity of crime" or which "individual circumstances" are relevant is unclear. The judges are left to evaluate the gravity of a crime and relevance of individual circumstances on a case-by-case basis.

The ICTY and ICTR judges consider the gravity of a crime to be the primary consideration in sentencing and the litmus test for the appropriate sentence. In this sense, the ICTY and ICTR judges endorsed the classic "just desert" principle of proportionality and argued that punishment should be proportional to the gravity of the particular offence(s) committed by a defendant. It should reflect the magnitude of harm caused by the defendant and his role and degree of participation in the offence(s) (i.e., the so-called offence gravity proportionality) (D'Ascoli 2011). Given the fact, however, that all international crimes tend to be extremely serious entailing extreme harm, it is unclear how much guidance this metric offers (Sloane 2007a). In many domestic jurisdictions, criminal codes provide for sentencing range for individual offences and thus set limits to judges' discretion. This is not the case at the international tribunals, and judges are free to evaluate proportionality of sentences on a case-by-case basis.

The proportionality assessment in the case of international crimes, however, is complicated not only by the lack of positive law guidance and extreme seriousness of crimes under the tribunals' jurisdiction but also by the fact that the tribunals deal with crimes committed by multiple perpetrators often acting in hierarchically organized groups. In this respect, the so-called defendant-relative proportionality of sentences becomes relevant. The defendant-relative proportionality dictates that a punishment should be proportional relative to the other defendants who committed international crimes within the same conflict, i.e., high-ranking individuals who organized, planned, and ordered the crimes should be punished more severely than low-ranking followers who "just" executed criminal orders. According to Ohlin, the relatively low ICTY sentences are the consequence of the fact that "the Tribunal was implicitly prioritizing defendant-relative proportionality over offence-gravity proportionality" and sentenced many defendants to relatively low sentences "because it was reserving space at the top of the scale for Milosevic, Mladic and Karadzic" (Ohlin 2011). The judges at the ICTY and ICTR implicitly endorsed the defendant-relative proportionality by applying the so-called principle of gradation. The principle of gradation is closely related to the position a defendant occupied and the role he/she played in the overall conflict situation. It stipulates that sentences should be gradated along with increasing authority of a defendant in the state structure and significance of his/her role in crimes. The relative position of the accused, however, has to always be balanced against the seriousness of the committed crimes. It is not the steadfast rule that all low-ranking defendants automatically receive low sentences and all authorities are automatically punished the most. Accordingly, offenders receiving the most severe sentences tend to be senior authority figures, such as ministers or governmental officials. However, very severe sentences can also be imposed on those at lower levels who zealously orchestrated or participated in crimes (Ewald 2010).

The provisions of the Statutes are supplemented by the Rules of Procedure and

Evidence. Only one rule, Rule 101, of the 165/154 ICTY/ICTR rules governing the proceedings before the tribunals is dedicated to factors relevant to sentencing. Rule 101 clarifies the regulation of the sentencing process only to a very limited extent. It limits the range of applicable sentences – the maximum sentence available to the judiciary is life imprisonment. It also instructs judges to take into account any aggravating and/or mitigating circumstances when determining the sentence. However, no list of aggravating and mitigating factors is provided. Only two potential mitigating factors are explicitly mentioned: "superior orders" and "substantial cooperation with the Prosecutor." Effectively, judges are left to determine on a case-by-case basis what factors justify an increase or reduction in sentence length. According to case law, mitigating factors need to be established upon the balance of probabilities and need not directly relate to the charged offences. The standards applicable to aggravating factors are more stringent. Aggravating factors must be proven beyond any reasonable doubt, and only those circumstances directly related to the commission of the offence charged and to the offender himself when he committed the offence may be considered in aggravation. A wide range of factors have been accepted by the tribunals in aggravation/mitigation of a sentence. "Abuse of superior position/authority/influence" is the aggravating factor most frequently used, while "family circumstances of a defendant" or "his/her assistance of victims" is among the mitigating factors cited the most by the ICTY and ICTR judges (Bagaric and Morss 2006).

As of July 2012, the tribunals have convicted more than 120 individuals for their involvement in genocide, crimes against humanity, and war crimes during the conflicts in the former Yugoslavia and Rwanda. The sentences range from 2 years imprisonment to the maximum: life imprisonment. Four ICTY and 21 ICTR defendants have been convicted to life imprisonment. The average length of determinate sentences is 15.9 years at the ICTY and 22.6 years at the ICTR. There are differences in sentence length between the ICTY and ICTR – the ICTR sentences are generally longer compared to the ICTY (Drumbl 2007). This difference is mainly connected to the different case composition at the tribunals: in contrast to the ICTY, the majority of the ICTR defendants are convicted of genocide and also many key figures (members of government and other high-ranking figures) and organizers of violence stood trial in Arusha; furthermore, all ICTR defendants are convicted for killing and/or serious violence against victims (Hola et al. 2011). Other possible reasons for the difference in sentence severity between ICTY and ICTR include the reference to national practice (incorporation of national law and sentencing practice) (Beresford 2001) or the sheer gravity of atrocity in Rwanda (Drumbl 2007).

ICC

The Rome Statute of the ICC establishes a very similar, though a bit more elaborate, sentencing framework compared to the Statutes of the ad hoc tribunals. Article 77 provides two types of custodial sentence: "imprisonment for a specified number of years, which may not exceed a maximum of 30 years" and "life imprisonment when justified by the extreme gravity of the crime and individual circumstances of the convicted person." The possibility of life imprisonment was included in the Rome Statute on the assumption that its imposition would be rare (Schabas 2010). In addition to imprisonment, the court may order a fine or a forfeiture of proceeds and property. Article 78 establishes general sentencing criteria in the same way as the Statutes of the ICTY and ICTR: the court shall take into account the gravity of the crime and the individual circumstances of the convicted person. The ICC Rules of Procedure and Evidence complement this general framework and provide judges with a more detailed sentencing guidance. Rule 145 states that the totality of any sentence must reflect the culpability of a defendant; and judges are asked to balance all the relevant factors, including any mitigating and aggravating factors and consider the circumstance of both: the convicted person and the crime. The Rule 145 further contains a demonstrative list of sentencing factors, including aggravating and mitigating circumstances. The lists resemble factors accepted in mitigation or aggravation in the case law of the ICTY and ICTR. Consequently, similar to the ICTY and ICTR, the ICC judges are vested with broad discretionary powers when it comes to sentencing. How this discretion will materialize in practice and whether the ICC judges will draw upon the rich sentencing jurisprudence of their predecessors at the ICTY and ICTR remain to be seen. Currently, the first ICC defendant Thomas Dyilo Lubanga has been convicted for the war crime of “conscripting and enlisting children under the age of 15 years and using them to participate actively in hostilities” and sentenced to 14 years of imprisonment.

Internationalized Courts and Tribunals

The Special Court for Sierra Leone (SCSL), Special Panels for Serious Crimes in East Timor (SPSC), and Extraordinary Chambers in the Courts of Cambodia (ECCC) constitute examples of the so-called internationalized tribunals combining international elements with much stronger links to domestic systems (in the sense of the applicable law, punishable offences or composition of tribunals). The sentencing practices and case law of these courts have been neglected in the scholarship, and most attention of commentators focused on the ICTY and ICTR. This lack of attention might be the result of the fact that case law and case composition at these tribunals have not been as extensive and/or varied compared to the ad hoc tribunals (SCSL rendered nine sentences; SPSC rendered 84 sentences focusing exclusively on low-ranking defendants; ECCC has so far convicted only one perpetrator). Similar to the other international courts, judges of the SCSL, SPSC, and ECCC have been vested with large discretionary powers when it comes to sentencing, and no detailed sentencing guidelines are provided by the positive law. Especially at the SCSL and the ECCC, judges seem to find inspiration in the case law of the ICTY and ICTR and often refer to the principles endorsed by the ad hoc tribunals’ judges. The sentencing argumentation of the SPSC judges, on the other hand, is very rudimentary and limited to description of factual findings and relevant aggravating and mitigating circumstances. Similar to the ICTY and ICTR, the sentence ranges at these courts are rather broad: (i) SCSL, 15–52 years; (ii) ECCC, handed down just one sentence of life imprisonment; and (iii) SPSC, 11 months–33 years and 4 months imprisonment. The high average sentence at the SCSL – 38.6 years – might be related to the fact that the tribunal is not allowed to render life sentences and instead it has handed out rather long sentences (40–50 years) to a couple of defendants. In contrast, the average sentence at the SPSC is comparably low – 8.6 years. This might be influenced by the fact that all defendants at the SPSC were lower-ranking crime executioners convicted usually of single instances of crime. More research is needed to describe and analyze sentencing practices at the internationalized tribunals and compare them to the sentencing regimes of the ICTY/ICTR/ICC but also domestic sentencing of international crimes.

Current Issues and Controversies

Goals of International Sentencing

The positive law outlined above does not contain any general provision explaining objectives to be pursued in sentencing. The Preamble of the ICC Statute seems to incorporate the purposes of retribution and general deterrence. The ICTY and ICTR Statutes are virtually silent in this respect. Accordingly, ICC, ICTY, and ICTR judges are generally free to switch from one self-chosen rationale to another as they see fit (Bagaric and Morss 2006). Some principles specific to sentencing have emerged in the ICTY and ICTR case law. In this respect, judges clearly found inspiration in classic “domestic” penal theories (Drumbl 2007). Over the years, the following purposes have been listed by international judges as relevant for international sentencing: retribution, justice, deterrence (general and specific), rehabilitation, expressivism, reprobation, stigmatization, affirmative prevention, incapacitation, protection of society, social defense, and finally restoration/maintenance of peace and reconciliation. Different combinations of some of these principles are usually listed at the beginning of the sentencing part of a judgment. There is no uniform approach in the tribunals' case law regarding the objectives of international punishment. In general, deterrence and retribution are emphasized in the majority of the cases. Next to retribution and deterrence, the third most frequently cited sentencing objective by the ICTY and ICTR judges is rehabilitation. Nonetheless, in the majority of cases, it is emphasized that rehabilitation should not be given undue weight given the seriousness of committed crimes. The exact way how the various purposes guide sentence determinations at the tribunals is, however, unclear (Henham 2005). The influence of retribution is apparent from the fact that gravity is consistently emphasized by the judges as the primary consideration in sentence determination and that proportionality between gravity of crime and sentence severity is often underscored. Decisions of judges regarding relevant aggravating and mitigating factors could be linked to retribution, rehabilitation, and reconciliation. While not being decisive during sentence determinations, rehabilitation is conversely emphasized at the sentence enforcement stage, and demonstration of rehabilitation seems to be one of the decisive factors of decisions granting early release to imprisoned convicts.

The relevance of these various sentencing objectives to international crimes has been discussed at length in legal doctrine. The legal scholars, however, also do not seem to agree on the objectives of international sentencing. Bagaric and Morss noted that goals of international sentencing in the form of reconciliation, retribution, and rehabilitation are either highly speculative or misguided. The only justification for the practice is according to these scholars general deterrence. Its relevance is, however, significantly undermined by the selective and infrequent prosecution of international crimes (Bagaric and Morss 2006). In contrast, Ohlin argues that deterrence together with rehabilitation and expressivism are inapposite in case of sentencing international crimes, and international sentencing ought to be based on retributive considerations (Ohlin 2009). Beresford, on the other hand, observes that retributive sentiments should play only a slight role in the development of international sentencing since they may prove counterproductive and disruptive to the restoration and maintenance of peace in the afflicted communities. He also cautions against according undue prominence to deterrence in meting out of sentences, especially taking into account the length of the proceedings at the international level. Rehabilitation should also be approached with caution since the judges cannot predict what rehabilitative programs are available in states enforcing the sentences (Beresford 2001). In contrast, Harmon and Gaynor advocate (general) deterrence as the sentencing objective in the case of international crimes and state that "there is some evidence that the international justice system is forcing western military commanders to listen more closely to their legal advisers when selecting targets and weapons for bombing missions" (Harmon and Gaynor 2007, p. 695). Noting the relative leniency of international sentences compared to domestic jurisdiction, Harmon and Gaynor consider the application of the concept of retribution in cases of large-scale crimes and mass atrocities unsatisfactory. Applying rehabilitation in the context of sentencing military and political leaders is according to these authors also inapposite since usually they are individuals with high social skills and do not clearly pose a criminal threat to society (Harmon and Gaynor 2007) Nemitz argues that retribution should not play the dominant role in the international criminal justice system especially because the heinous nature of the international crimes makes it impossible to state that the perpetrator of various crimes against humanity deserves a sentence of so and so many years. The sentencing purpose of deterrence also has, according to Nemitz, shortcomings especially (i) given the fact that many perpetrators of international crimes actually do not act rationally, do not weigh pros and cons of committing crimes, but act in the heat of a combat situation and (ii) given the slight chances that a perpetrator will indeed be apprehended and tried at one of the tribunals. He however acknowledges that deterrent effect might be greater for high-ranking perpetrators.

Nemitz argues that affirmative general prevention in a sense of educating public that violations of ICL are always prohibited should be the primary aim of sentencing (Nemitz 2001) Drumbl is also very skeptical regarding the ability of international trials to attain retributive, deterrent, and expressive aims due to the selectivity of international prosecutions, pervasive discretion of international judges, and political contingency of the process of international criminal law (Drumbl 2010). Similarly, Sloane argues that the efficacy of justifications for punishment derived from domestic systems is compromised at the international level. International criminal law differs from domestic criminal law in several respects that are relevant to the social institution of punishment, including the nature of (i) the community that authorizes international criminal law, (ii) the crimes addressed by it, and (iii) the perpetrators judged by it. He then argues that the principal value of international punishment lies in its expressive dimensions, which more accurately capture the nature of international sentencing (Sloane 2007b). Consequently, a rather vivid academic debate regarding the purposes of international punishment has developed in international criminal law scholarship. To determine goals for international sentencing that would reflect a specific character of international crimes being a form of collective criminality and of international criminal justice system based on multiplicity of international criminal courts and institutions with limited mandates is a challenging task that would for sure spark off academic discussions also in the future.

Hierarchy of International Crimes

As noted above the statutes of the international tribunals do not contain any sentencing tariff. Similarly, judges in the current case law emphasize that there is no hierarchy among individual categories of international crimes and that all the crimes are very serious violations of international (humanitarian) law. The question of hierarchy among genocide, crimes against humanity, and war crimes, however, was not so clear-cut in the past. In the early ICTY and ICTR case law, judges endorsed the idea of hierarchy among genocide, crimes against humanity, and war crimes. Genocide was labeled as the crime of crimes and considered the most serious category of international crimes (Prosecutor vs. Kambanda, ICTR-97-23, Judgment, 4 September 1998, para. 16). In the first ICTY cases, judges indicated that, all else being equal, an offence committed as a crime against humanity, that is, with awareness of a widespread or systematic attack on a civilian population, is more serious than an ordinary war crime (Prosecutor v. Tadic, IT-94-1-T, Sentencing Judgment, 14 July 1997, para. 73). Later, both tribunals, however, seem to adopt the stance that there is no preestablished hierarchy between individual categories of crimes emphasizing that *all* crimes under their jurisdiction are very serious violations of international humanitarian law.

In contrast, several legal scholars have argued that there should be a hierarchy of individual categories of international crimes based on the assessment of their objective gravity. According to these scholars, when the same act, e.g., murder, is committed as either genocide or a crime against humanity or a war crime, there should be a gradation in punishment reflecting different contextual requirements of individual categories. Each category of crimes requires proof of different *chapeau elements*, i.e., genocide requires special intent to destroy in whole or in part a specific group of people; crimes against humanity must be committed as part of a widespread or systematic attack against a civilian population with knowledge of such an attack; whereas war crimes could be isolated incidents committed within an armed conflict. Due to these inherent differences, it should matter for the purposes of sentencing if an act is classified e.g., as genocide or as a war crime. Scholars have advocated a hierarchy of individual categories of international crimes based on their objective severity (Carcano 2002; Danner 2001; Olusanya 2005). According to these scholars, a genocidal murder should be considered as the most serious entailing the severest punishment, followed by a murder as a crime against humanity and the least severe punishment should be ascribed to a murder as a war crime. Numerous empirical studies have also shown that despite the proclamation of international judges to the contrary, there appears to be differences in the actual sentence length between individual categories of crimes indicating an empirical ordering among genocide, crimes against humanity, and war crimes for the purposes of sentencing (Hola et al. 2011; Meernik 2011).

Consistency of International Sentencing

The requirement of consistency of sentencing could be perceived as one of the fundamental principles of any sentencing practice. It stems from the modern internationally recognized human rights principles and has been explicitly endorsed by the international sentencing case law (Prosecutor v. Mucic et al., IT-96-21-A, Judgment, 20 February 2001, para. 756). Many legal scholars have criticized international sentencing practice, in particular that of the ICTY and ICTR, and labelled it as "irrational" (Henham 2003), "akin to a lottery system" (Olusanya 2005), or "confusing, disparate, inconsistent, and erratic" (Drumbl 2007). These criticisms, however, have never been founded on a comprehensive empirical analysis. Lately, more and more empirical legal studies have emerged and addressed the alleged inconsistencies of international sentencing from an empirical perspective. The majority of these studies focused on the ICTY and ICTR since these international tribunals produced the most extensive sentencing practice. Without any exceptions these empirical studies have identified consistent patterns in the tribunals' sentencing practice. The first to introduce empirical analysis into the research on international sentencing was *Meernik*. Together with his colleagues, he has published several studies analyzing primarily the early sentencing practice of the ICTY and ICTR. In all these studies, they concluded that sentencing at the tribunals seems not to be influenced by political factors and that "there is a fair degree of consistency in the sentences conferred on the guilty and sentences are premised on those critical factors that the judges are asked to apply by the Statutes and Rules of Procedure and Evidence" (Meernik and King 2003; Meernik et al. 2005). Subsequent empirical studies confirmed these findings and identified consistent and predictable patterns in the international sentencing practice (Ewald 2010) Hola in her PhD dissertation identified the following sentencing patterns in the ICTY and ICTR sentencing practice: (i) an empirical ordering/hierarchy among different categories of international crimes with genocide being considered the most serious category of crime, followed by crimes against humanity and war crimes; (ii) a culpability level of a perpetrator is assessed by the ICTY and ICTR judges taking into account primarily a defendant's position in the overall state hierarchy, his role, and the scope of his/her crime with high-ranking organizers convicted on multiple crimes being subjected to the most severe sentences; and (iii) relevance of individualization of sentences with aggravating and mitigating factors cited by the ICTY and ICTR judges accounting for an increase/reduction in sentence length (Hola 2012). All these studies have demonstrated that international sentencing can be statistically predicted to a considerable extent – 70 % of sentence variation can be predicted by legally relevant factors (Hola 2012). It seems that at the international level, despite the lack of sentencing guidelines and large discretionary powers of judges, the sentences appear to be as statistically predictable as sentences in domestic legal systems with more detailed legal regulation of sentence determination such as the USA (Hofer et al. 1999, p. 243). The majority of the empirical studies, however, focused on the two ad hoc international criminal tribunals and examined sentencing consistency within one tribunal at a time. In order to evaluate consistency of international sentencing practice, it is also necessary to not only examine sentences across different international and internationalized tribunals (Meernik 2011) but also evaluate sentencing of international crimes by domestic courts. Some authors have noted undue leniency of international sentences compared to sentences pronounced by domestic courts (Bagaric and Morss 2006; Drumbl 2007).

Specifics and Further Criticism of International Sentencing

Next to the criticism stemming from the alleged inconsistencies of international sentencing, the practices of international courts have been criticized for being disconnected from the reality of victimized communities – as not reflecting the perception of "justice" in states where the crimes were committed. According to some authors, this failure to find a way of relating the outcomes of international sentencing to local perceptions of justice continues to be critical for the future credibility of international criminal justice. There is therefore a need to close this gap between global and local in order to enhance legitimacy of international courts and tribunals (Henham 2007). Drumbl calls for international criminal law interventions to engage with local practices that actually reflect the customs, procedures, and mores of those individuals affected by violence (as perpetrators and victims) (Drumbl 2005). In a similar vein Henham argues that the ultimate challenge for international trial justice is to ensure that the "real" experiences of victims and communities in post-conflict states really do engage with the sentencing outcomes produced by the international courts and tribunals, so that they symbolize legitimacy and justice, rather than rhetoric and partiality (Henham 2005). Damaska warns that realization of this ideal would entail fragmentation of international criminal law and that there should be a uniform legal regime for all international criminals. International judges should, however, always strive to explain the reasons for any deviations in their practice from local norms or practices (Damaska 2008).

International sentencing and focus on individualized guilt and punishment have also been called into question by authors who see fundamental problems with such individualized responses to collective and systematic crimes. International crimes are usually offences entailing a large number of victims, committed as part of a large-scale campaign, and implicating many individuals with very different tasks. These crimes are often state sanctioned and characterized by mass involvement of both military functionaries and civilians. Criminologists have labeled international crimes as "crimes of obedience" and argued that international crimes are a different type of criminality than ordinary crimes (Smeulers 2008). They are crimes of conformism rather than crimes of deviance. In times of collective violence, many otherwise law-abiding citizens get involved and commit extremely cruel acts simply by conforming to their environment. Some authors have argued that this specific character of international crimes calls for a specific reaction. Drumbl maintains that "adequately redressing collective violence requires a discursive shift to other accountability mechanisms, including *collective forms of responsibility*" (Drumbl 2010).

Conclusion and Future Research

International criminal law and international sentencing have come a long way since the World War II when the first international criminal tribunals held individuals legally accountable for international crimes. After the end of the Cold War, there has been a proliferation of various international(ized) courts and tribunals that are called to hold perpetrators of international crimes to account. These developments brought a heightened scholarly attention to various issues of international criminal justice. Relative to the amount of commentary on other aspects of international criminal law, international sentencing still remains to be a rather neglected phenomenon. There are therefore many areas of international sentencing that need to be further explored by scholars.

Above all, there is a paucity of empirical studies on international sentencing. The highest priority for future research is evaluation of effects of international sentencing. Given the panoply of goals international sentencing institutions have ascribed to themselves, it is necessary to assess to what extent these goals are attainable and have been attained by the existing tribunals. Given the specific context within which international criminal justice institutions function and specific character of international crimes, the achievement of these various goals is not self-evident. The effects of international sentencing on perpetrators, victims, and global public (the purported audiences of international criminal justice) shall be operationalized and evaluated. In this respect, further theoretical research should further problematize various goals of international criminal justice and international sentencing and develop sui generis theories that would reflect specifics of international punishment and international crimes.

Given the multiplicity of institutions delivering international sentences, it is also important to examine sentencing practices of other international and internationalized tribunals than ICTY and ICTR in order to assess consistency/fragmentation of "international sentencing system." What are the relationships and influences, in terms of sentencing for international crimes, among all the international criminal courts and tribunals that have emerged in the past decade? In view of the principle of complementarity, one of the governing principles at the ICC, interactions between domestic courts dealing with international crimes and their international counterparts also need to be explored. Researchers should also devote more attention to the context of international sentencing and transitional justice in general. It is necessary to examine how international criminal trials affect and complement various local endeavors to deal with a violent past – not only domestic criminal trials but also other transitional justice mechanisms such as truth and reconciliation commissions or various other local modalities of justice.

Theoretical inquiries should address the questions of linking international prosecutions and sentencing to the etiology of international crimes. The current practice of the international tribunals has been criticized as not sufficiently addressing the peculiarities of international crimes and simply transposing concepts from domestic criminal law. More criminological research is needed in this respect to further promote the understanding of international crimes as collective criminality and its peculiarities with respect to punishment and sentencing individual perpetrators.

Recommended Reading and References

Bagaric M, Morss J (2006) International sentencing law: in search of a justification and coherent framework. Int Crim Law Rev 6:191–225

Beresford S (2001) Unshackling the paper tiger – the sentencing practices of the ad hoc international criminal tribunals for the former Yugoslavia and Rwanda. Int Crim Law Rev 1:33–90

Carcano A (2002) Sentencing and the gravity of the offence in international criminal law. Int Comp Law Q 51:583–609

D'Ascoli S (2011) Sentencing in international criminal law, the un ad hoc tribunals and future perspectives for the ICC. Hart Publishing, Oxford

Damaska M (2008) What is the point of international criminal justice? Chi-Kent Law Rev 83:329–365

Danner AM (2001) Constructing a hierarchy of crimes in international criminal law sentencing. Va Law Rev 87:415–501

de Roca EM (2008) Sentencing and incarceration in the ad hoc tribunals. Stanf J Int Law 44:1–62

Drumbl MA (2005) Collective violence and individual punishment: The criminality of Mass Atrocity, 99 Northwestern University Law Review: 539–610.

Drumbl MA (2007) Atrocity, punishment and international law. Cambridge University Press, New York

Drumbl MA (2010) Collective responsibility and post-conflict justice. Washington & Lee Public legal studies research paper series, working paper available at http://ssrn.com/abstract=1601506

Ewald U (2010) Predictably irrational – international sentencing and its discourse against the backdrop of preliminary empirical findings on ICTY sentencing practices. Int Crim Law Rev 10:365–402

Harmon MB, Gaynor F (2007) Ordinary sentences for extraordinary crimes. J Int Crim Justice 5:683–712

Haveman R, Olusanya OA (eds) (2006) Sentencing and sanctioning in supranational criminal law. Intersentia, Antwerp

Heller KJ (2011) The Nuremberg military tribunals and the origins of international criminal law. Oxford University Press, Oxford

Henham R (2003) Some issues for sentencing in the international criminal court. Int Comp Law Q 52:81–114

Henham R (2005) Punishment and process in international criminal trials. Ashgate, Aldershot

Henham R (2007) Developing contextualized rationales for sentencing in international criminal trials, a plea for empirical research. J Int Crim Justice 5:757–778

Hofer PJ, Blackwell KR, Ruback RB (1999) The effect of the federal sentencing guidelines on inter-judge sentencing disparity. J Crim Law Criminol 90:239–306

Hola B (2012) International sentencing, a game of Russian roulette or consistent practice. Uitgeverij BOX Press, Oisterwijk

Hola B, Smeulers A, Bijleveld C (2011) International sentencing facts and figures: the sentencing practice at the ICTY and ICTR. J Int Crim Justice 9:411–439

Meernik J (2011) Sentencing rationales and judicial decision making at the international tribunals. Soc Sci Q 92:588–608

Meernik J, King KL (2003) The sentencing determinants of the international criminal tribunal for the former Yugoslavia: an empirical and doctrinal analysis. Leiden J Int Law 16:717–750

Meernik J et al (2005) Judicial decision making and international tribunals: assessing the impact of individual, national, and international factors. Soc Sci Q 86:683–703

Mettraux G (2005) International crimes and the ad hoc tribunals. Oxford University Press, New York

Nemitz JC (2001) The law of sentencing in international criminal law: the purposes of sentencing and the applicable method for the determination of the sentence. Yearb Int Humanit Law 4:87–127

Ohlin JD (2009) Towards a unique theory of international sentencing in international criminal procedure: towards a coherent body of law. In: Sluiter G, Vasiliev S (eds) International criminal procedure, towards a coherent body of law. Cameron May, London, pp 373–404

Ohlin JD (2011) Proportional sentences at the ICTY. In: Swart B, Sluiter G, Zahar A (eds) The legacy of the international criminal tribunal for the former Yugoslavia. Oxford University Press, New York, pp 322–341

Olusanya OA (2005) Sentencing war crimes and crimes against humanity under the international criminal tribunal for the former Yugoslavia. Europa Law Publishing, Groningen

Schabas WA (2006) Sentencing by international tribunals: a human rights approach. Duke J Comp Int Law 7: 461–517

Schabas WA (2010) The international criminal court, a commentary on the Rome statute. Oxford University Press, New York

Sloane RD (2007a) Sentencing for the ‘crime of crimes’: the evolving ‘common law’ of sentencing of the international criminal tribunal for Rwanda. J Int Crim Justice 5:713–734

Sloane RD (2007b) The expressive capacity of international punishment: the limits of the national law analogy and the potential of international criminal law. Stanf J Int Law 43:39–94

Smeulers A (2008) Punishing the enemies of all mankind. Leiden J Int Law 21:971–993

International Tribunals

▶ Human Rights Violations in Criminal Court

Internet Pornography

▶ Pornography

Interrogation

▶ Interview and Interrogation Methods Effects on Confession Accuracy

Interrupted Time Series Analysis

▶ Interrupted Time Series Models

Interrupted Time Series Models

David McDowall[1] and Richard McCleary[2]
[1]School of Criminal Justice, University at Albany, State University of New York, Albany, NY, USA
[2]School of Social Ecology, University of California, Irvine, Irvine, CA, USA

Synonyms

Interrupted time series analysis; Time series intervention analysis; Time series quasi-experiment

Overview

Interrupted time series models compare the levels of a time series before and after the introduction of a discrete intervention. The time series design is often a good match to the questions posed in criminology and criminal justice studies, and many examples of its use appear in the research literature.

The popularity of interrupted time series designs is largely attributable to the strength of the causal inferences that they allow. Shadish, Cook, and Campbell (2002) evaluate a variety of common research designs in terms of their ability to guard against threats to four types of validity: statistical conclusion validity, internal validity, construct validity, and external validity. Within each threat category, they consider a more detailed list of potential problems. Interrupted time series models are vulnerable to comparatively few of the issues that Shadish et al. identify. Several threats are nevertheless plausible in typical applications, and these require attention before making causal assertions.

Researchers can reduce threats to statistical conclusion validity by using long time series and analytical methods that are suited to temporal data. Threats to internal validity, the most important challenges to causal inference, require a combination of design variations and modifications to the statistical analysis. Methods for addressing construct validity threats also rely on statistical adjustments, especially dynamic modeling of possible impact patterns. Ruling out threats to external validity, in contrast, depends on replicating the quasi-experiment in diverse settings.

Fundamentals of Interrupted Time Series Models

Interrupted time series models use repeated measurements on an outcome variable to estimate the causal impact of an intervention. The interrupted time series design is among the strongest types of quasi-experiments, and it has many applications in criminology and criminal justice research. Interrupted time series models are especially useful in evaluating the outcomes of legal interventions, because they are well suited to the aggregate archival data that these studies often employ.

The time series design is intuitive in its structure, and variations on it have a long history of use in the natural and social sciences. Careful assessments of the design by Donald T. Campbell and his collaborators (Campbell and Stanley 1963; Cook and Campbell 1979; Shadish et al. 2002) nevertheless did much to increase its popularity among social researchers. The discussions by Campbell and associates are largely responsible for current understandings of the interrupted time series design's structure and also of its strengths and limitations.

In skeletal form, an interrupted time series model breaks a long series of observations into pre-intervention and post-intervention segments, and a data analysis then compares the segment means. A rejection of the null hypothesis of equality between the means supports a conclusion that the intervention produced a change in the outcome variable.

Figure 1 presents a simple example of a time series quasi-experiment. The data are 108 months of gun homicide counts from Detroit, Michigan (Loftin et al. 1983). A law that imposed a mandatory sentence enhancement for persons convicted of firearm crimes became effective

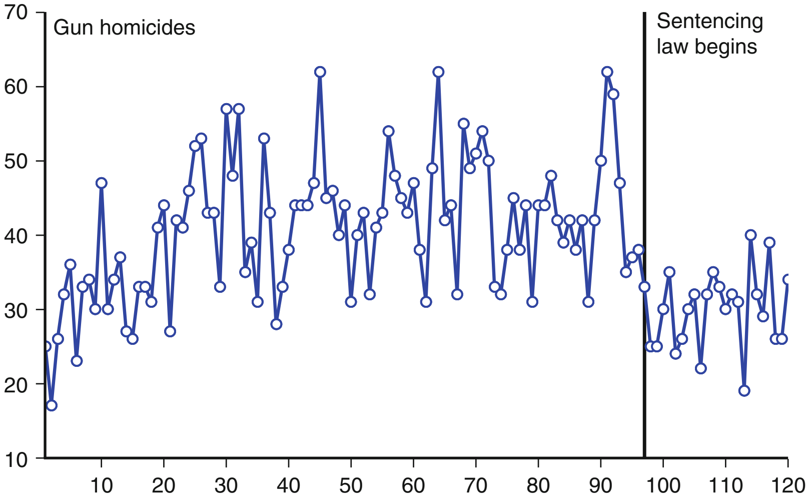

Interrupted Time Series Models, Fig. 1 Monthly Detroit homicide counts, 1971–1979 (Source: Loftin et al. 1983) during the 97th month of the study. A plot of the time series appears to show that gun homicides fell after the law began. Consistent with this appearance, the post-intervention mean represents a 29 % drop from the pre-intervention baseline. Solely from these results, one might conclude that the law reduced gun homicides.

Such a conclusion would be premature, however, because the analysis ignores several major threats to valid inference. The most obvious of these involve statistical questions about whether the series decreased by more than a chance amount during the post-intervention period. Visual inspections of time series are notoriously prone to error, making formal hypothesis tests necessary to establish the existence of a post-intervention change. Yet standard tests assume independent observations, and this is a requirement that time series data usually do not meet. Later observations in a series are almost always predictable from earlier ones, and serial dependence will bias the test results. A satisfactory data analysis therefore requires attention to statistical issues.

Time series quasi-experiments are also subject to design-based challenges to inferences about causal effects. Campbell and his associates systematically examined potential threats to validity that are inherent in a variety of research designs. They concluded that interrupted time series studies guard against most of these threats relatively well, but that several threats are plausible in most applications. The following sections further discuss both analytical and design issues in time series studies, considering threats to validity and ways of reducing their influence.

Four Threats to Valid Causal Inference

Campbell and Stanley (1963; Campbell 1963) divided validity threats into two broad categories, internal validity and external validity. Threats to internal validity considered whether the intervention was the cause of an observed change in the outcome variable. Threats to external validity considered whether any causal effect of the intervention held outside the original research setting.

Interrupted Time Series Models, Table 1 Common threats to validity in interrupted time series quasi-experiments

Type of validity	Threat to validity
Statistical conclusion validity	Low statistical power
	Violated assumptions of the test
Internal validity	History
	Maturation
	Regression artifacts
	Instrumentation
Construct validity	Novelty and disruption
External validity	Interaction over treatment variations
	Interaction with settings

Cook and Campbell (1979) expanded the Campbell and Stanley classification, adding two new threat categories. Threats to statistical conclusion validity arose from problems that Campbell and Stanley had previously considered under the heading of internal validity. Threats to construct validity involved sources of confounding that Campbell and Stanley had regarded as components of external validity. In a further revision of the classification, Shadish et al. (2002) retained Cook and Campbell's four-category scheme but substantially added to the set of specific threats.

Table 1 lists the threats to each type of validity that apply to the interrupted time series design. Shadish et al. (2002) include more threats within each category than appear in the table, and any of these might in theory operate in an interrupted time series analysis. Many of the threats are usually implausible, however, and the table includes only the most common and realistic possibilities.

Threats to Statistical Conclusion Validity

Threats to statistical conclusion validity come from situations in which analysts err in concluding that a correlation exists between a pair of variables. Shadish et al. (2002: 45) identify nine specific ways in which such threats can arise, but all of these fall into two more general categories that involve either Type I or Type II error rates. A Type I error, or α-error, involves rejecting the null hypothesis when it is true. In a time series quasi-experiment, this would occur if the data analysis yielded false evidence that the intervention influenced the outcome variable. A Type II error, or β-error, involves failing to reject the null hypothesis when it is false. Here the intervention did affect the outcome variable, but the statistical analysis was not successful in detecting the impact. Conventionally, analysts fix the Type I error rate at $\alpha \leq .05$ and the Type II error rate at $\beta \leq .2$ (see, e.g., Cohen 1988).

Threats to statistical conclusion validity involve distortions to the values of α and β, and the usual solution is to control the error rates through an appropriate statistical model. Although other reasonable classes of models exist, the most common choice in interrupted time series studies is the *Auto*R*egressive* I*ntegrated* M*oving* A*verage* (ARIMA) models of Box and Jenkins (1970).

Under the null hypothesis of no intervention effect, one can write an ARIMA model as

$$\varphi(B)Z_t = \theta(B)a_t \qquad Zt = Y_t - \mu_Y$$

Here Z_t and a_t are observations from a stationary time series and an *iid* $N(0,\sigma^2)$ error term, and $\varphi(B)$ and $\theta(B)$ are polynomial lag operators. Given appropriate constraints on $\varphi(B)$ and $\theta(B)$, one can solve the ARIMA model for a_t:

$$\theta(B)^{-1}\varphi(B)Z_t = a_t$$

To carry out the null hypothesis test, one first codes the intervention as a binary variable,

$$X_t = 0 \text{ for } t \leq N_{pre};\, X_t = 1 \text{ thereafter}$$

One then adds to the ARIMA model a function of X_t – conventionally called the "transfer function" – to represent the following causal impact:

$$\theta(B)^{-1}\varphi(B)Z_t = a_t + \omega X_t$$

In this model, ω estimates the effect of X_t on Z_t. Since a_t is *iid* $N(0,\sigma^2)$, a test of the null hypothesis that $\omega = 0$ now follows standard procedures.

Interrupted Time Series Models, Table 2 Parameter estimates for dynamic mandatory sentencing intervention model

Response model: $\varphi(B)$ (homicide) – $\theta_0 = a_t + \omega(\text{law}_t)$ $\text{law}_t = 0$ for observations 1–84; = 1 for observations 85–108	
$\theta_0 = 42.25$	$t(\theta_0) = 39.16$
$\omega = -12.04$	$t(\omega) = -3.99$
$\varphi = 0.28$	$t(\varphi) = 3.00$

Methods for building ARIMA models appear in Glass, Willson, and Gottman (1975), McCleary and Hay (1980), and McDowall et al. (1980). The $\varphi(B)$ and $\theta(B)$ operators are themselves of no interest for the hypothesis test, and they are present only to control irrelevant variation. The transfer function measures the impact of the intervention on the series, and estimating it accurately is the major goal of the analysis.

Table 2 presents an ARIMA impact analysis of the Detroit firearm homicide series from Fig. 1. The intercept, θ_0, shows that on average about 42 gun homicides occurred each month before the mandatory sentencing law began. Following the law, the mean number of firearm homicides dropped by 12 per month, a 28 % reduction. The *t*-value for the intervention coefficient allows a rejection of the null hypothesis of no change at $p < .0001$. The φ term in the model controls for serial dependence between the observations that would otherwise inflate the Type I error rate. In the example, the serial dependence is statistically significant, but small enough that only a slight difference exists between the ARIMA impact estimate and the original naive model. Obviously, this type of outcome often does not occur.

Although a satisfactory ARIMA model will control for Type I error threats, it does not fully address Type II errors. Failing to reject the null hypothesis does not mean that it is true, and accepting the null requires showing that the research design had sufficient power to detect an effect if one existed.

Of the factors that determine the power of a time series study, the underlying effect size is the most important consideration. Small effects are difficult to separate from random variation, and they result in tests with low power. Besides the effect size, the power of a time series quasi-experiment also depends on the length of the series, and on the balance between the pre-intervention and post-intervention segments. Authorities often recommend 50 observations as the minimum series length necessary if an analysis is to detect an effect (e.g., McCleary and Hay 1980; McDowall et al. 1980). Statistical power will also increase as the design more closely approximates perfect balance, with equally long pre-intervention and post-intervention segments.

Threats to Internal Validity

Shadish et al. (2002: 55) consider nine threats to internal validity, and any one of these might apply to time series quasi-experiments. Under realistic circumstances, however, only four threats are serious concerns. *History* and *maturation*, which are inherent in the use of time series data, may occur in any application of the design. *Instrumentation* and *regression* are also sometimes problems, but they arise only in designs that evaluate planned interventions. Many interrupted time series studies instead involve *natural experiments* (Campbell 1963), where an intervention occurs without advance planning. In these settings, instrumentation and regression threats are usually implausible.

The most serious threats to the internal validity of a time series quasi-experiment come from the operation of history. Historical threats are changes in a time series that occur along with the intervention but that are due to other causes. Some critics argue that such threats are ubiquitous, and that they will always undermine conclusions from an interrupted time series design (e.g., Piquero 2005; Harcourt and Ludwig 2006). These claims rest on a too simplistic understanding of the time series quasi-experiment, but they are helpful in emphasizing the need to take history seriously. Although interrupted time series studies can never completely rule out lurking historical threats, they can greatly reduce their plausibility through appropriate design adaptations.

A standard design-based approach to historical threats is to analyze one or more comparison series. A nonexperimental comparison will never be fully equivalent to the outcome under study, and this approach will therefore always be open to question. Comparisons can nevertheless take numerous forms, and a careful selection of them can help rule out possible historical artifacts.

An analysis might consider other treatment series that the intervention should have affected, for example, or control series that it should not have affected. The analysis might also examine the intervention's impact within different subgroups, or consider switching replications in which the intervention shifted from one series to another. A consistent pattern of results across different variables or time periods undermines historical explanations and provides support for an intervention effect. In general, historical explanations become more difficult to justify as the pattern of comparisons becomes increasingly dense (Morgan and Winship 2007).

Hennigan et al. (1982), who examined changes in crime rates following the introduction of commercial television, exemplify the use of comparisons to address historical artifacts. In 34 "early" television cities that Hennnigan et al. studied, TV broadcasting began in 1950. In 34 "late" TV cities, television was not available until 1954. Figure 2 presents annual time series of property crimes for the "early" and "late" city groups.

Historical events are plausible threats to inferences about the impact of television in either the "early" cities alone or the "late" cities alone. Many conditions varied during the period in which each group adopted television, and some of these could explain a change in crime. The Korean War was underway in 1951, for example, and the economy had slid into a recession during 1955. More generally, alternative explanations for a change in either set of cities are easy to construct.

Yet when considering the two time series together, the changes in crime rates are more difficult to dismiss simply as the product of historical events. Plausible historical explanations would have to account for increases that occurred

Interrupted Time Series Models, Fig. 2 Annual property crimes for cities with commercial TV broadcast service in 1951and for cities with TV in 1955 (Source: Hennigan et al. 1982)

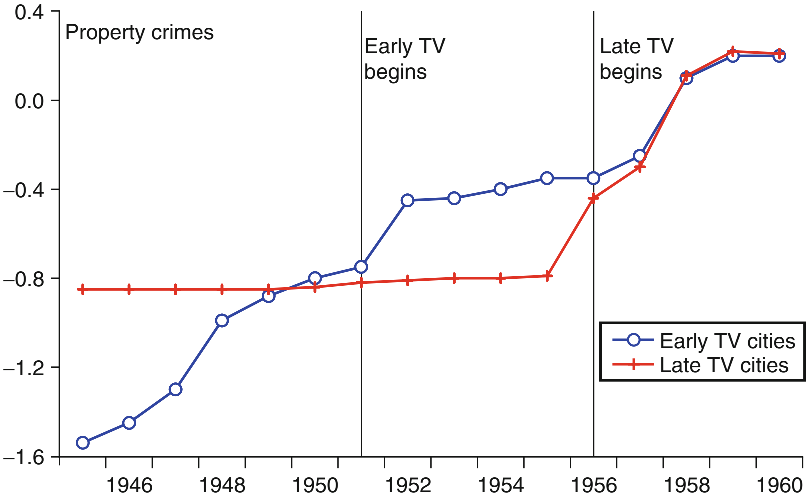

at different time points, but at each point affected only one group of cities. A critic might conceivably create such an explanation, of course, but persuasively justifying it could be another matter.

A one-time change in a single time series will usually allow multiple historical explanations. As the number of pertinent comparisons increases, however, historical threats become easier to dismiss. Although a time series quasi-experiment can never completely eliminate history, it can rule out all but the most complex (and therefore unlikely) possibilities.

In contrast to history, methods for addressing the other three threats to internal validity do not rely on extensions of the design. Maturation threats, in particular, require the use of statistical modeling. Maturation is the result of a developmental process that operates independently of the intervention, and it will generate a secular trend in the outcome series. Social science data sometimes display such trends, and trending patterns are especially common in the long series that are most desirable for analysis. Maturational trends can therefore take the appearance of an intervention effect, and this makes them a problem in any application of the design.

An illustration of a maturation threat appears in Fig. 3, which shows monthly homicide counts from New York City between January 1993 and December 2003. Pridemore, Chamlin, and Trahan (2008) used similar (although not identical) data to study whether the 2001 World Trade Center attack affected the frequency of homicides in the city. The series is generally flat for several years before and after September 2001, inconsistent with the possibility that the terrorist attack was a catalyst to broader changes in violence. Yet the downward trend in the early years dominates Fig. 3, and this suggests that a maturational artifact may account for the apparent lack of impact. Pridemore et al. recognized this threat, and in an analysis that controlled for it also found no evidence that the World Trade Center attack influenced homicides.

Controlling for maturational threats requires modifications to the quasi-experiment's statistical analysis. The usual ARIMA model for an interrupted time series design,

$$\theta(B)^{-1}\varphi(B)Z_t = a_t + \omega X_t$$

assumes that $\varphi(B)$ and $\theta(B)$ fall within the bounds of stationarity and invertibility (Box and Jenkins 1970: 49–51). If this is not so, the series under analysis will be nonstationary, and the model will not have a proper solution.

Nonstationarity can generate a secular trend, but it also manifests itself in other patterns that lack constant means. The New York homicide series does not exhibit a consistent trend over its entire length, for example, and instead displays

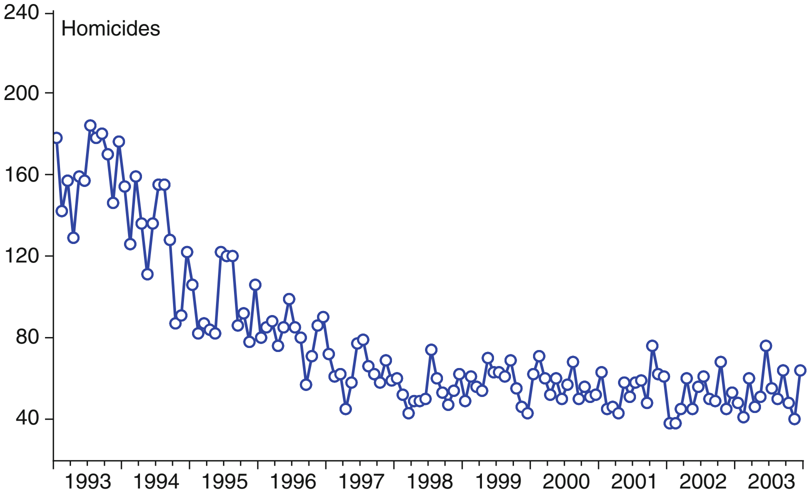

Interrupted Time Series Models, Fig. 3 Monthly New York City homicide counts, 1993–2003 (Source: U.S. Department of Health & Human Services (various years))

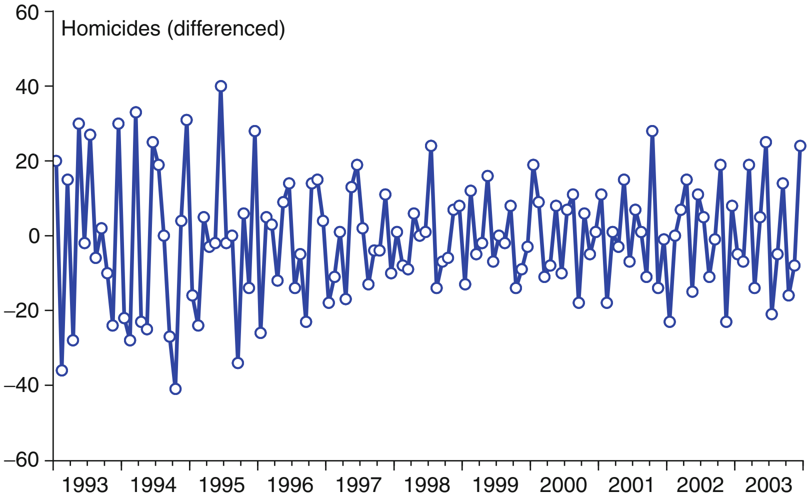

Interrupted Time Series Models, Fig. 4 New York City homicides from Fig. 3, differenced

the "drifting," up-and-down, appearance that characterizes a trendless random walk.

Although nonstationarity is a common occurrence in social science applications, converting a series to first-differences will almost always resolve the problem. Using "∇" to represent the differencing operation,

$$\nabla Y_t = Y_t - Y_{t-1}$$

a nonstationary series can be modeled as

$$\nabla Y_t = \varphi(B)^{-1}\theta(B)\, a_t$$

$$\text{or} \quad Y_t = \nabla^{-1}\varphi(B)^{-1}\theta(B)\, a_t$$

Figure 4 plots the first-differences of the New York homicide series. The differences fluctuate around a constant and unvarying level, and the differencing operation has removed maturation as a realistic threat. If the mean of the differenced series is zero, the series follows a directionless random walk; otherwise, the mean difference is an estimate of its secular trend.

Besides resolving maturation threats, differencing a series also removes the influence of any and all time-constant causes of the outcome variable. To take only one example, homicide rates are generally higher in large cities like New York than they are in rural areas or suburbs. Suppose that Y_t is New York's homicide rate, and that W represents population size and all other variables that make New York's rate consistently different from those elsewhere. After differencing the series,

$$Y_t - Y_{t-1} = f(a_t) - f(a_{t-1}) + W - W$$

$$\nabla Y_t = f(a_t) - f(a_{t-1})$$

and the influence of W and other time-constant causes disappears from the model. This property establishes a link between the interrupted time series quasi-experiment and the fixed-effects designs that are popular for analyzing panel data (Allison 2009).

As with maturation, regression threats to internal validity distort conclusions from a statistical analysis. Yet, unlike maturation threats, regression is plausible only in a limited range of circumstances. Regression arises mainly when an intervention occurs after a time series reaches a level that observers find so alarming that they take action to change it. In these reactive situations, regression to the mean can produce an increase or decrease that is independent of any intervention effect. When an intervention occurs naturally, without advance planning, threats due to this type of mean reversion are much less plausible.

Reactive interventions are common in criminal justice settings, because new legal initiatives are often the result of unusually high levels of some undesirable phenomenon. In one of the best-known early applications of the time series quasi-experiment, Campbell and Ross (1968) studied the impact of a speeding crackdown on Connecticut highway fatalities. Traffic deaths dropped significantly after the crackdown began, but this was largely due to a regression artifact. Fatalities had soared immediately prior to the intervention, which occurred as a response intended to reduce them. A predictable drop in traffic deaths followed the crackdowns as accidents moved down toward their historically average levels.

Regression becomes a less serious threat as the time series becomes longer. Introducing the intervention when the series is at an unusually high or low point will bias estimates of the pre- and post-intervention means. The bias decreases as the pre- and post-intervention segments grow longer, however, and it will eventually shrink to zero. Use of at least 50 observations for a time series quasi-experiment will generally be sufficient to make regression artifacts implausible, but longer series will obviously be more desirable than this minimum.

Like regression, instrumentation can be a realistic threat to validity in situations with planned interventions. These interventions often come with a package of other changes, including new ways of measuring the outcome variable. The new measurement methods are then confounded with the intervention, and they will provide an alternative explanation for any post-intervention difference in the series.

Figure 5 is from McCleary, Nienstedt, and Erven (1982), and shows a time series of monthly burglary counts from Tucson, Arizona. Before 1978, uniformed patrol officers investigated all burglary reports within the city. Detectives then assumed the investigative responsibility for a 2-year period between 1978 and 1980, after which investigations passed back to the uniformed officers. Burglary counts dropped when detectives investigated the cases, consistent with the hypothesis that they were more skilled in crime prevention.

The switching interventions help eliminate historical explanations for the lower burglary counts, but instrumentation threats remain plausible. Additional analysis by McCleary et al. showed that in fact the detectives and uniformed officers used different methods of keeping records. This instrumentation artifact explained the lower counts during the period that detectives handled the cases, and after controlling for it, burglaries did not vary with the type of investigating officer.

Unlike cases with planned interventions, the risk of instrumentation threats is typically low when time series quasi-experiments involve natural interventions. Measurement methods could

Interrupted Time Series Models, Fig. 5 Monthly burglaries for Tucson, 1975–1981. During a 24-month period, burglaries are assigned to detectives for investigation (Source: McCleary et al. 1982)

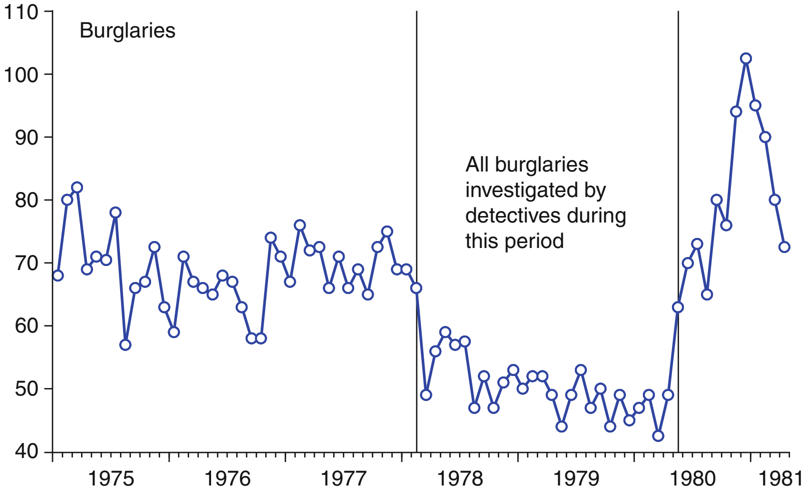

in some cases change if an intervention occurred naturally, but this would require an unusual set of circumstances. Natural experiments are therefore ordinarily not plausible settings for instrumentation threats to exist.

Threats to Construct Validity

Threats to construct validity involve the adequacy of "the match between study operations and the constructs used to describe those operations" (Shadish et al. 2002: 73). Of the 14 specific threats to construct validity that Shadish et al. list, only one, *novelty and disruption* is relevant to the interrupted time series design. Because interventions seek to alter established patterns of conduct and action, the outcome variable may react to features of an intervention that are independent of its intended causal impact.

As with maturation threats, the data analysis can allow for threats due to novelty and disruption. To accomplish this, however, a researcher must have a general idea about how the construct validity threats should operate.

Figure 6 displays four impact patterns that an intervention might produce, including three patterns that incorporate the effects of novelty and disruption. The models in the first row of the figure illustrate permanent intervention responses. In these cases a series may change instantaneously or gradually after the intervention occurs but, in either case, it changes permanently. The *abrupt, permanent* response model requires the least elaborate theory, and it is the pattern that researchers most often entertain. This model assumes that any influence of the intervention is immediate, and that the impact holds constant throughout the post-intervention period. If these assumptions are true, the analysis faces no threats due to novelty and disruption.

In contrast, the *gradual*, *permanent* response model assumes that the full influence of an intervention is distributed over time. Here the novelty and disruption associated with the intervention act to slow its full impact, and the response follows a damped curve. This pattern is sometimes a reasonable model for the influence of new laws on crime rates. In these cases, organizational inertia and the diffusion of information among potential offenders may work to spread the law's effects over multiple time points. The novelty and disruption threats then help mask the total impact of the law, and an analysis that does not allow for them will produce downwardly biased estimates.

The models in the second row of Fig. 6 depict temporary responses to an intervention. Both the *abrupt*, *temporary* and *gradual*, *temporary* patterns involve cases in which novelty and disruption threats entirely account for an observed change. Here the outcome variable responds to the intervention initially, but this influence

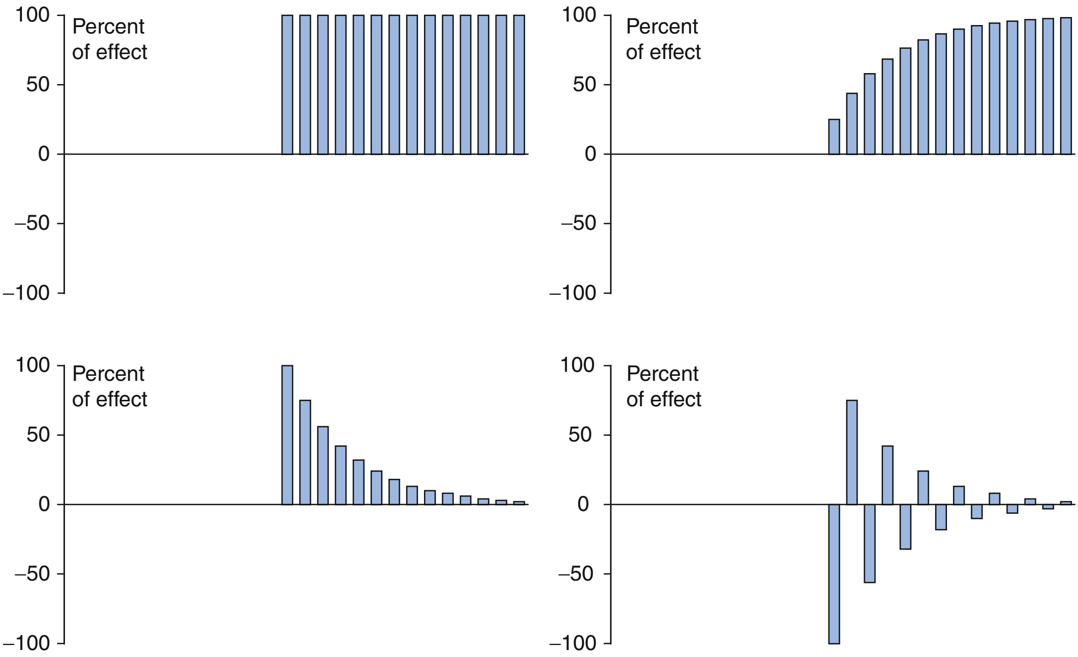

Interrupted Time Series Models, Fig. 6 Model responses to an intervention

eventually wears off. Under ideal conditions, with a long post-intervention time series, an analysis may ignore the threats to validity that these temporary impacts pose. Under less-than-ideal conditions, the researcher must accommodate the effects within the structure of the analytical model.

The structure of the usual linear ARIMA model does not allow the types of dynamic responses that novelty and disruption typically produce. To incorporate these more elaborate patterns, Box and Jenkins (1970; Box and Tiao 1975) proposed a lagged polynomial transfer function,

$$\theta(B)^{-1}\varphi(B)Z_t = a_t + \delta(B)\omega X_t$$

where $\delta(B)$ is a polynomial lag operator and ω is the usual impact estimate of the effect of X_t on Z_t.

The polynomial lags make the ARIMA model nonlinear, and this complicates estimation and interpretations of the results. These inconveniences aside, the lag model directly addresses novelty and disruption, and it is an effective approach to controlling them in practice.

Besides the individual effects, permanent and temporary responses to an intervention may also exist together. Figure 7 shows a time series of divorce rates before and after adoption of the Australian Family Law Act of 1975. The Act allowed no-fault divorce, and its opponents argued that divorce rates would increase as a result. An evaluation conducted by the Australian government concluded that divorces did initially rise, but that over the next 3 years they dropped back to their initial levels. The government evaluation attributed the temporarily higher post-1975 divorce rates to the operation of a trend.

Reanalyzing these data, McCleary and Riggs (1982) modeled the impact of the Act as partly temporary and partly permanent. A temporary increase occurred immediately, but this response then rapidly decayed over the next 3 years. McCleary and Riggs found that divorce rates never fully returned to their earlier values, and that they instead eventually stabilized at a new and higher level.

Whether or not they have permanent effects, legal interventions can sometimes display

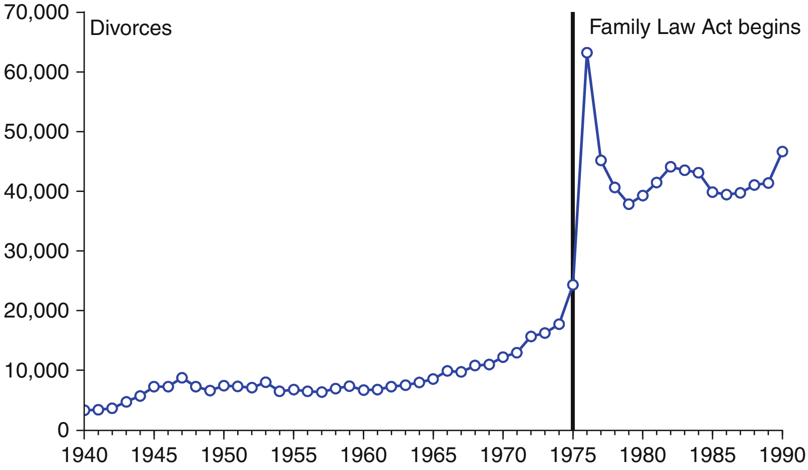

Interrupted Time Series Models, Fig. 7 Australian divorces before and after the 1975 Family Law Act (Source: McCleary and Riggs 1982)

temporary impacts that amount to decaying spikes. Failing to allow for such transitory changes can lead to invalid inferences, and this gives researchers an incentive to develop models that closely mimic observed data patterns. Yet a data-dependent modeling strategy raises another serious threat to statistical conclusion validity: *fishing*. Fishing refers to tailoring a statistical model to match the peculiarities of a specific set of data, and it invalidates standard approaches to statistical inference. To avoid fishing threats, researchers must specify the form of intervention models before undertaking the analysis, and this substantially limits the situations in which complex temporary response models are possible.

Threats to External Validity

A time series quasi-experiment usually considers an intervention's influence on only a single series, and this restriction threatens its external validity. External validity considers whether a finding generalizes beyond a particular sample of data to other settings, treatments, and outcome variables (Shadish et al. 2002: 83). If a new law influences crime rates in one area, for example, will the same law – or a slightly modified version of it – also influence crime rates elsewhere?

Threats to external validity are limitations on the generality of an impact, and they come from interactions between the intervention and other components of variation. Two sources of interactions are variations in treatments and variations in settings. In the first case, seemingly irrelevant differences in the intervention, such as small changes in the wording or penalty structure of a law, will alter its impact. In the second case, the effect of the intervention will not be constant across areas, perhaps due to unmeasured differences in environmental conditions. A new law may affect crime rates in some states, for example, but not in others.

Threats to external validity are impossible to evaluate when an analysis considers only one time series. Ruling out these threats necessarily requires replicating the quasi-experiment under a diverse set of circumstances. Concerns about external validity will decrease as the range of variation in the interventions and settings grows.

A complication in using replications is that they can produce results in support of very different conclusions about an intervention's impact. Evaluations of the preventive effects of laws regulating firearms, for example, have yielded an array of positive, negative, and null estimates (Hahn et al. 2005). In these situations, the variance of the impact distributions can be more important than are point estimates of the average effect size.

Given the availability of replications of a time series quasi-experiment, a researcher can evaluate external validity using one of two approaches.

First, the analysis can pool the data from the set of individual time series to form a single overall model. It can then use the diversity in the combined data to estimate both the average effect size and its expected variability (e.g., McGaw and McCleary 1985). Second, the analysis can pool only the impact estimates from each of the time series models, decomposing them into components associated with various types of validity threats.

The first approach uses more of the information contained in the replications than does the second, and this ideally makes it the better choice. The first approach also requires a full set of observations from each of the original series, however, while the second approach only requires information about the impact estimates. The second approach is therefore often more practical to use, and it is also generally easier to apply. Statistical models for the second approach come from meta-analysis, and divide the effect variance into parts due to the setting, intervention, and other potential confounders. McDowall, Loftin, and Wiersema (1992) use a meta-analysis model to estimate the overall impact and variability of mandatory sentencing laws for firearm crimes.

By training and custom, social scientists usually concentrate on the results of individual studies, and this can sometimes make replication strategies controversial. Tonry (1995), and Durlauf and Nagin (2011), for example, both criticize the McDowall et al. study for its use of pooled estimates. Much of their concern apparently reflects a strong ideological opposition to mandatory sentencing laws, but some of it is also likely due to misunderstandings about what replications can accomplish. Although external validity threats can threaten inferences from interrupted time series studies, current practice often gives less weight to these threats than to those from the other three validity types.

Conclusions

Interrupted time series models are important to all of the social science disciplines, including criminology and criminal justice. Of the various ways in which criminologists use time series data to make causal inferences, interrupted time series analysis is among the most popular approaches. The interrupted time series design has applications in a wide range of situations, and it allows stronger causal statements than do most nonexperimental alternatives.

Time series models are of course also vulnerable to multiple validity threats, and one would normally not use them if randomized controlled trials were possible. Still, given the practical constraints imposed by field research, a time series study is in many settings the strongest feasible approach. In criminal justice and criminology, especially, situations often arise in which researchers want to estimate the impact of a well-defined policy change on an aggregate outcome variable. Interrupted time series studies provide a natural fit to the data and research questions in these situations, and this can make them the quasi-experimental design of choice.

Related Entries

► Propensity Score Matching
► Randomized Experiments in Criminology and Criminal Justice

Recommended Reading and References

Allison PD (2009) Fixed effects regression models. Sage, Thousand Oaks, CA

Box GEP, Jenkins GM (1970) Time series analysis: forecasting and control. Holden-Day, San Francisco

Box GEP, Tiao GC (1975) Intervention analysis with applications to economic and environmental problems. J Am Statistical Assoc 70:70–79

Campbell DT (1963) From description to experimentation: interpreting trends as quasi-experiments. In: Harris CW (ed) Problems in measuring change. University of Wisconsin Press, Madison

Campbell DT, Ross HL (1968) The Connecticut crackdown on speeding: time series data in quasi-experimental analysis. Law Soc Rev 3:33–53

Campbell DT, Stanley JC (1963) Experimental and quasi-experimental designs for research. Rand-McNally, Chicago

Cohen J (1988) Statistical power analysis for the behavioral sciences, 2nd edn. Lawrence Erlbaum, Englewood Cliffs, NJ
Cook TD, Campbell DT (1979) Quasi-experimentation: design and analysis issues for field settings. Houghton-Mifflin, Boston
Durlauf SN, Nagin DS (2011) Imprisonment and crime: can both be reduced? Criminol Public Policy 10:13–54
Glass GV, Willson VL, Gottman JM (1975) Design and analysis of time series experiments. Colorado Associated University Press, Boulder, CO
Hahn RA, Bilukha O, Crosby A, Fullilove MT, Liberman A, Moscicki EK, Snyder S, Tuma F, Briss PA (2005) Firearm laws and the reduction of violence: a systematic review. Am J Prev Med 28(2S1):40-65
Harcourt BE, Ludwig J (2006) Broken windows: new evidence from New York City and a five-city social experiment. Univ Chicago Law Rev 73:271–320
Hennigan KM, Del Rosario ML, Heath L, Cook TD, Wharton JD, Calder BJ (1982) Impact of the introduction of television on crime in the United States: empirical findings and theoretical implications. J Personal Social Psychol 42:461–477
Loftin C, Heuman M, McDowall D (1983) Mandatory sentencing and firearms violence: evaluating an alternative to gun control. Law Soc Rev 17:287–318
McCleary R, Hay RA Jr (1980) Applied time series analysis for the social sciences. Sage, Beverly Hills, CA
McCleary R, Nienstedt BC, Erven JM (1982) Uniform Crime Reports as organizational outcomes: Three time series quasi-experiments. Social Problems 29:361–72
McCleary R, Riggs JE (1982) The 1975 Australian Family Law Act: a model for assessing legal impacts. New Directions Program Evaluation 16:7–18
McDowall D, McCleary R, Hay RA, Meidinger EE (1980) Interrupted time series analysis. Sage, Beverly Hills, CA
McDowall D, Loftin C, Wiersema B (1992) A comparative study of the preventive effects of mandatory sentencing laws for gun crimes. J Criminal Law Criminol 83:378–394
McGaw DB, McCleary R (1985) PAC spending, electioneering, and lobbying: a vector ARIMA time series analysis. Polity 17:574–585
Morgan SL, Winship C (2007) Counterfactuals and causal inference: methods and principles for social research. Cambridge University Press, New York
Piquero A (2005) Reliable information and rational policy decisions: does gun research fit the bill? Criminol Public Policy 4:779–798
Pridemore WA, Chamlin MB, Trahan A (2008) A test of competing hypotheses about homicide rates following terrorist attacks: an interrupted time series analysis of September 11 and Oklahoma City. J Quant Crimnol 24:381–396
Shadish WR, Cook TD, Campbell DT (2002) Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin, New York
Tonry M (1995) Malign neglect: race, crime, and punishment in America. Oxford University Press, New York
U.S. Department of Health & Human Services. (various years). Public use data tape documentation: multiple cause of death for ICD-9 1989 Data. [Data file]. Author, Hyattsville, MD

Intervention and Prevention in China

Bill Hebenton[1] and Susyan Jou[2]
[1]Centre for Criminology and Criminal Justice, School of Law, University of Manchester, Manchester, UK
[2]Graduate School of Criminology, National Taipei University, Taipei, Taiwan

Overview

The key elements of the overall approach to prevention and intervention on crime in contemporary China are outlined in Fig. 1. The approach is best understood as state-initiated comprehensive management or governance of social order. Developed in the new era of economic liberalization and reform, the policy of comprehensive governance was first promulgated in 1981 as a key organizational tool for crime control but only legislated for in 1991 (Situ and Liu 1996a). The approach emphasizes prevention, education, and punishment of offenders through coordination and cooperation between political, legal, educational, cultural, economic, and social bodies yet sits in an integrated manner alongside "strike-hard" campaigns by police against more serious and organized crime (Jiang and Dai 1990; Situ and Liu 1996a; Zhong 2008). While the development, nature, and dynamic interplay between these elements can only be fully explicated through detailed sociohistorical analysis, this entry describes and contextualizes the relevant organizational and thematic issues contained within the comprehensive approach to the prevention and intervention on crime, including the complex relationship between public and private agencies in prevention and intervention on crime in China.

Intervention and Prevention in China, Fig. 1 The crime prevention framework in China

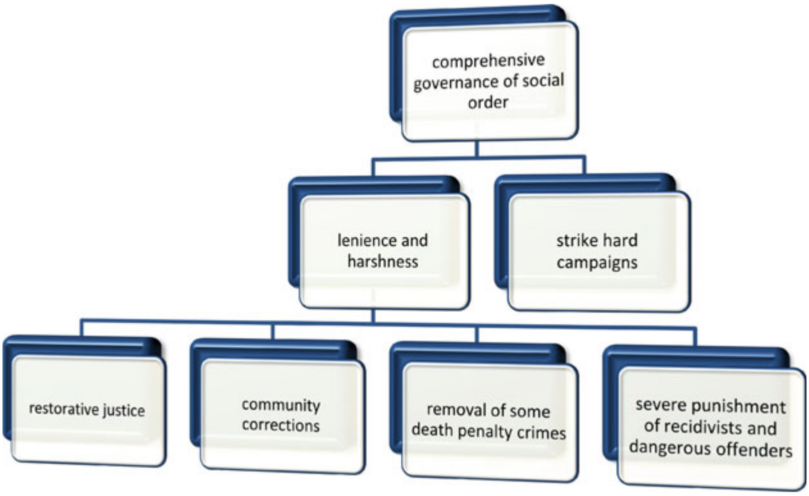
comprehensive governance of social order
lenience and harshness
strike hard campaigns
restorative justice
community corrections
removal of some death penalty crimes
severe punishment of recidivists and dangerous offenders

Introduction

In 1962, Leifeng, a Chinese People's Liberation Army soldier, died; he was decorated by the state for his exemplary kind deeds. His helpful actions were universally acknowledged, and the month of March was designated as the month of "Learning from Leifeng," when everyone was encouraged to follow the example of Leifeng and undertake helpful and kind actions (Bakken 2000). Fifty years on, modern socialist China is both a very different yet similar place. At the societal level, a depoliticized place; yet in terms of the language of governance, including management of crime, familiar (Dutton 2005: 252–255). In the social control mechanisms of contemporary China, "exhortation through example" and mass participation are still regarded as key tenets by the ruling Chinese Communist Party (Bakken 2000; Chen 2004). Social order in China is supposed to be maintained by the capacity of society's members to understand one another and to act in concert in achieving common goals through common rules of behavior. The most important distinction is the efforts of the Chinese state to control both the behavior and the minds of the people (Wilson et al. 1977). Social conformity in the Chinese vocabulary is not limited to behavioral conformity with the rule of law but always moralistically identifies with the officially endorsed beliefs of social standards and behavioral norms. The Chinese tradition, both before the creation of the People's Republic (in 1949) and since, is that of "greatest unity" of society (Ren 1997). Thus, the current crime prevention framework known as "comprehensive governance (or management) of social order" takes crime as caused by a complex set of psychological, social, and economic factors and to be prevented and "cured" by a similarly rich interplay of state mechanisms.

The policy of comprehensive governance of social order adopted since 1991 is, in essence, traditional strategies with new components: thus, it continues the Maoist ideology of the state as "moral entrepreneur" (Wilson et al. 1977), combined with formal and informal mechanisms; the policy continues with the perceived need to develop citizens' knowledge about laws in the Republic; it revives the Maoist "mass line" – listening to the ideas of the people and the ruling party working with them (Dutton 2005); finally, it continues to promote lenience and recognizes the importance of rehabilitating offenders through correction while adding a bifurcatory "harsh" aspect – namely, that of severely punishing recidivists and organized criminals through long prison sentences and via the death penalty if required (Liang 2008). These key themes are enlarged upon in what follows.

Key Conceptual and Organizational Themes

Along with modernization, crime rates have risen significantly in China's cities over recent decades (Liu 2005). While there has been no national-level victimization survey in China, there is city-level data from the key city of Tianjin, where estimates on risk of major crime types are available to scholars (Zhang et al. 2007a). Surveys of this type in China are not without research challenges and concerns, for example, around gaining access, designing and deriving samples, protecting privacy of respondents, and ensuring data integrity and quality (Zhang et al. 2007b). Crime is now, to use David Garland's term, a "normal social fact" in urban China (Garland 2001). A fellow traveler alongside economic development has been insecurity, built into the fabric of everyday life. Crime has increasingly become a daily consideration for anyone who owns property or uses public spaces in urban China. People have felt the need to establish control over risks and uncertainties, and the desire to assuage feelings of insecurity has become ever more urgent. For the Chinese Communist Party, the increasing crime rate has posed a threat to the legitimacy and stability of its rule, and social order has become a major government concern (Dutton 2005; Muhlhahn 2009). As originally developed in the 1980s, one key organizational theme within the policy of comprehensive governance is to engage in so-called "strike-hard" anticrime campaigns. This prong of the policy focused on attacking crime directly, as opposed to the longer-term goal of prevention and social reintegration of offenders. "Strike hard" has been described as an extreme form of the educative and deterrence function of punishment used as a method of crime prevention and social control. It is essentially a policy of employing severity and swiftness to deal with criminal cases considered to be a major threat to public or social order (Trevaskes 2007a, b). "Strike-hard" anticrime campaigns take two forms. One is the large-scale campaign that runs for up to 3 years, focusing on a variety of crime targets chosen by the state. The other is the "specialized" smaller-scale campaign targeted at one category of crime. Party authorities initiate both the national campaigns and the specialized anticrime drives, which are run under the leadership of Party committees and criminal justice personnel at provincial, municipal, and local levels. Campaigns against organized crime and drug crime are well described in recent literature (Trevaskes 2007a, b, 2008). Police, prosecutors, and courts work together to ensure swift arrests, swift prosecution, and swift and severe punishment for the targeted types of crime that the authorities identify. "Severity" limits judges' discretionary power and requires them to punish relatively harshly. "Speedy" investigation, prosecution, and adjudication of cases to deal efficiently with identified criminals, often within several days of their arrest, distinguish these campaigns from the usual criminal justice process (Muhlhahn 2009). "The People's War on Drugs" is a major specialized campaign that operated as the drug crime counterpart to an existing campaign against organized crime (Trevaskes 2008). It was launched in mid-2005 and exemplifies comprehensive management of social order as an approach to crime fighting through interagency cooperation between the criminal justice agencies. It is the latest in a long succession of antidrug campaigns since the late 1980s and is China's key countermeasure against an alarming rise in heroin use. The core strategy is to block the sources of drugs and stop their circulation. The national modus operandi of the campaign in its first year was to separate tasks across individual agencies rather than to use a multiagency method. Individual agencies included customs, railways, water police, airline security, forestry police, and the postal services.

Criticisms of "strike-hard" campaigns abound in the Chinese literature (see Trevaskes 2007a, b for details). Among senior policy makers and senior administrators, anticrime campaigns are overwhelmingly supported, but among practitioners, there is greater ambivalence and criticism. Problems identified in the Chinese-language literature include the difficulty of control and accountability where criminal justice agencies overstep the mark in campaigns, that like cases are not treated alike in terms of severity across place and time, and the problems of confusion on the ground as different agencies compete to "deliver swift justice." Many Chinese scholars question the overall effectiveness of the strike-hard policy (see Trevaskes 2007a, b). Apart from no significant reduction of crime rates following the campaigns, there are criticisms of the lack of long-term prevention effects (Liang 2005). However, some research argues that the anticrime campaigns have significantly impacted on lowering serious crime (Chen 2012). In this regard, it is interesting to note that the recent fourth "strike-hard" campaign in 2010 targeting two types of criminals – violent criminals (esp. with gun and organized criminals) and so-called "mass-hatred" criminals (including telephone fraudsters, children and women traffickers) – has been remarked upon more favorably by commentators. Unlike previous campaign waves, the police this time were focusing more on prevention – for example, increased patrol, problem-oriented policing, and mediating community conflicts rather than arresting and delivering severe and swift punishment (Yang 2010).

As an element of the comprehensive governance of social order, "strike-hard" campaigns draw from deep cultural wellsprings, and any adequate understanding of why such a practice is continued, despite what appears to be very limited success, needs to recognize how the practice partly reflects the remnants of classic Maoist "mass-line" thinking: the categorization and targeting of specific criminal types and defining social harm according to the socialist social relations violated an important aspect of Mao's early campaigns against "political enemies," the flexible nature of administration of law in practice during campaigns, and finally the educative role performed by state agencies (police, prosecution, and courts). In essence, the "strike-hard" campaign performs a number of noninstrumental, expressive political functions for the ruling Party (see Liang 2005). Thus, short-term severe strike campaigns became an indispensable complement to decreasing traditional mass-line neighborhood-level preventive control. This leads to a second key theme within the comprehensive governance of social order, namely, mobilization of broad preventive forces, particularly at neighborhood (community) level.

The origin of mobilization lies in the working style of the early Chinese Communist guerilla movement of the 1930s, linking popular participation with the guerilla struggle (see Selden 1971). Mao framed it in the following terms, "...take the ideas of the masses and concentrate them, then go to the masses and propagate and explain these ideas until the masses embrace them as their own..." (Mao 1967, vol 3, 119). In the Mao period, the Chinese peasant was mobilized out of any passive role into willing, positive, active participation based not on broad notions but on a single commitment to the total social and economic transformation of society. Campaigns were highly organized mobilizations of mass commitment to push over barriers to proletarian dictatorship and socialism, and mass-line thinking has continued to echo throughout the post-Mao period. Wang (1989) sets out the theory of comprehensive governance and basic rationale for the revival of somewhat moribund mass-line security organs and renewal of community activism under Party committee leadership. In this context, commentators point to the pioneering role of the Fengqiao district in Zhejiang province, where in the early 1960s local mass democracy was put into effect, illuminating how long-term effective crime prevention role relies upon taking wider administrative measures (Dutton 2005).

Under the 1991 law on comprehensive governance, all government organizations at different levels regard taking comprehensive measures to secure prevention as one of the most important duties, with an appropriate responsibility system for management; it is also accepted that it is imperative to mobilize and rely on the broad mass of ordinary people in the community. Within urban environments, the "subdistrict" level typifies how these arrangements are structured. For example, in Shanghai, the subdistrict would have an office of comprehensive management and a community structure would contain members of the residents' committee, community police, property owners, housing management, private security personnel overseen by the police together with social workers responsible for meeting the needs of local offenders, and finally local volunteers who are usually retired and receive small financial subsidies. Financial support and human resources are underpinned by government spending. Emphasizing nonlegal methods to solve neighbor disputes is another key characteristic of the comprehensive approach. People's mediation is a process whereby the parties concerned to a dispute seek to reach a mediation agreement on the basis of equal negotiation and free will and thus solve the dispute between them, and the mediation agreement reached upon mediation is binding to all parties concerned, and the parties concerned are asked to fulfill it as agreed. There is a mediation committee as part of every residents' committee (Hong 2011). Since the 1990s, people's mediation has faced some unprecedented challenges. Along with rapid social changes, China's dispute resolution system has gone through significant transformations. Courts have increasingly been favored by the public as a forum to solve disputes. In 1990, the people's courts handled about 1.8 million civil lawsuits. This number jumped to 3.4 million in 2000, a nearly twofold increase. Meanwhile, the numbers of mediation committees and disputes solved through mediation have declined steadily. In 1990, there were approximately 1 million mediation committees that solved 7.4 million disputes. In 2000, these numbers dropped to 960 thousand mediation committees that solved 5 million disputes (Hong 2011). In the 1980s, people's mediation committees disposed of an average of almost 11 times more disputes than trial courts, but this ratio had decreased to 1:1 in 2001. In response, and as part of the comprehensive approach, the Chinese government has made significant policy reforms to revitalize mediation, during the Fourth National People's Mediation Work Conference held in Beijing in 1999; the Ministry of Justice added an important phrase, "using multiple methods to solve problems," to the original people's guideline of "combining mediation and prevention, with prevention as the main goal" for mediation. The Ministry called for people's mediation committees across the country to gain cooperation and resources from administrative and judicial organizations to expand the range of disputes that can be solved by mediation (Hong 2011).

Details of the practice of the comprehensive approach to prevention can be gleaned from research studies on Guangzhou and Shenzhen (Situ and Liu 1996b; Zhong 2008). In Guangzhou, government aimed for "the establishment of 1,000 security neighborhoods in which criminal events must be substantially reduced, the transient population must be adequately managed, and the public environment must be made safe and clean" (Situ and Liu 1996b: 95). Organizationally this involved establishing the coordinating Office of Comprehensive Management which ensured coordination of crime prevention, monitored the level of transient population, and developed environmental schemes; it is a well-developed contractual responsibility system – contracts not only clearly define responsibility but also make security a responsibility that is legally binding on all people involved. Failure to meet the conditions of the contract can result in administrative penalties, such as resignation or dismissal, or financial punishments, such as a fine or elimination of bonus. Those who are responsible for the occurrence of serious criminal incidents are also subject to criminal charges. By linking the security duties to individual interests, the system of responsibility effectively promotes the motivation of contractees and improves the quality of their performance in carrying out preventive duties. Patrol-based neighborhood teams were also established, with volunteers receiving some training and augmenting police patrol. Linked to this was police-initiated "anticrime" campaigns against specific crimes (as described earlier in this entry). Finally, situational measures were implemented, including "gated" entry and the securing of ground floor apartments with "bars-to-windows".

In Shenzhen, five neighborhoods were studied to examine the practice of comprehensive governance (Zhong 2008). Shenzhen local government sought to develop "little safe and civilized communities" (LSCC). In terms of safety, residents should feel safe, with few serious crimes and general public security cases kept under control. In terms of civilization, active politics, economy, culture, education, hygiene, and so on were required. Thus, in an LSCC in Shenzhen, there must be a strong conformity to law and regulation, good public order, high degree of civilization, upright social morale, beautiful environment, and harmonious neighborhood relationship. Piloting of schemes began in 1992 and took place over a number of years, with full development in 1998. For a little community to be eligible for LSCC, it has to meet these criteria and go through a complicated rating process. The rating categorizes the communities into "exemplary," advanced, pass, or failure, on an annual basis. If a community fails to meet the standards in the second year, it is stripped of the original level. It can also strive to upgrade to a higher level in subsequent years. The important measures in implementing the comprehensive approach to building such communities have four elements: organizational features (ideological underpinning, responsibility system, mass prevention and mass management, and fund raising), safety measures (police, private security, situational measures, and management of the transient population), civilization measures (moral education, promoting harmony, community development, and environment improvement), and of course the rating system itself.

A final significant theme in understanding the changing conceptual basis of contemporary crime prevention and intervention in China is that of the recent promotion of the idea of "balancing leniency and harshness"; this idea has been brought to bear in the areas of alternatives to normal punishment (restorative justice) and the death penalty in particular (Trevaskes 2010). Within the wider policy of comprehensive governance, this development has resulted in a redefinition and use of what are considered "serious crime" and "minor crime." After 2006, a series of criminal justice reforms began to be publicized after President Hu Jintao espoused "building a socialist Harmonious Society" as the touchstone of the Sixteenth Party Congress governance objectives for the coming decade. These reforms were aimed at addressing the crucial political issue of encouraging social harmony by preventing further social conflicts during a period of accelerated economic and social transformation. In 2006/2007, the Chinese Communist Party used this policy as a rationale for momentous changes to the death penalty, thus the change in 2006, requiring the Supreme People's Court to ratify death sentences, in contrast to the previous practice of allowing provincial courts this power. More recently still and in keeping with this shift, recent amendments to the Chinese Criminal Law in 2011 have removed the death sentence from many nonviolent economic crimes. In 2007, criminal justice authorities announced changes at the other end of the spectrum – moves to decriminalize a wide range of minor offenses and to encourage courts to use a greater variety of alternatives to custodial sentencing, such as restorative justice and involving community-based corrections. Restorative justice techniques are intended to provide assistance, guidance, and direction to residents who have committed minor offenses, have been caught by police agencies, or have been released from correctional institutions. Traditionally, these have been organized in the form of help groups that normally consist of the offender's parents or relatives, a member of the neighborhood committee, and an officer from the neighborhood police station. Currently, social workers and volunteers are also likely to participate in such groups. The primary goal is to intervene in the lives of residents, especially young residents, who are at risk of becoming offenders and to reintegrate the ex-offenders back into the community. Nevertheless, restorative justice involving victims is not without its critics in China, where certain scholars have argued that it is immature and confused, incompatible with Chinese constitutional law (Zhao 2005). Application of community corrections, involving management of sentences outside of prison, is itself still "experimental" and is permitted in some 18 Chinese cities. A recent review of the scientific literature on community corrections in China highlighted a number of issues: the small number of recidivism studies following release from prison, a lack of reliable information surrounding recidivism rates in general, and very few studies comparing China's particular circumstances with those of the West (Liang and Wilson 2008).

Conclusions

Research from the USA, Europe, and the UK suggests that community crime prevention transcends both situational and social prevention and indeed transcends the borders between developmental and situational prevention, with its focus on social relationships within and beyond individual communities (Ekblom and Pease 1995). In the West, criminological commentary on the community approach to crime prevention has journeyed backwards and forwards, at times stressing the significance of informal social control on crime and disorder inside individual communities, while at other times orienting around the opportunity reduction model (or situational crime prevention) and emphasizing partnership/interagency working including community and problem-oriented policing. This entry has analyzed the policy and practice of crime prevention and intervention in China, a process conforming to recognizable forms of community crime prevention in the West. However, its "Chinese characteristics," resting upon unique forms of social capital, are also evident (Zhuo 2012). In addition, the state, through all the levels of the Chinese Communist Party, plays a determining role and it could hardly be otherwise. The state-led comprehensive approach both structures and mandates community organizations working with the police and other government agencies, in concerted efforts to achieve reductions in crime. As mentioned in the introduction, traditional strategies with new components underpin the approach: its ideological underpinnings and mass-line tradition reinvented by a responsibility system and incentivized by financial contract arrangements and government investment at all levels of governance make it distinctive from community crime prevention approaches described by Western commentators (Welsh and Hoshi 2002). Neighborhood committees still function as an important part of the grassroots organization but are under financial pressures and under challenge from newly emerging market influences and organizations. The public security bureaus (police) still uphold the mass line and place great emphasis on good relationships with the masses – yet actively try to adapt to the new environment of greater "police professionalization" and mixed markets in security provision. Dutton notes the irony of the revival of Maoist rhetoric and mass-line forms, since the "revival" actually signaled the demise of the significance of transformational politics; instead it was part of the slow process of political and ideological drainage that would change the nature of mass mobilization in China. The new dynamic fueling these developments has been monetary incentives rather than "revolutionary" volition (Dutton 2005: 260–262). The new money-incentivized forms are, in this sense, actually the commodification of mass-line security (Dutton 2005: 298–299).

Contracted services in private security vary appreciably across urban neighborhoods in China. For some newly established, affluent neighborhoods, private security companies provide a variety of services ranging from neighborhood cleaning to guarding services (Trevaskes 2007c). They engage in night patrol and manage security devices (CCTV cameras). Some of the services overlap with the activities traditionally performed by neighborhood police stations and residents' committees. Furthermore, some neighborhoods, especially older ones, do not have any contracted services. Very little research has been conducted on the interconnections between the traditional organizational forms of prevention in neighborhoods (the comprehensive governance model) and contracted services.

Returning to the bifurcatory nature of China's comprehensive approach, one should note that it is not without precedent. It has been argued that in the last three decades, the practice of crime prevention and intervention in both the UK and the USA exhibited two distinct patterns of action (Garland 2001): an adaptive strategy stressing prevention and partnership and a strategy aimed at identifying "dangerous" offenders and expressive punishment. The roots of such bifurcation lie in complex political rationalities – and China is not an exception to the playing out of such governmental processes (Liang 2005: 395). In the end, the nature and dynamics of change in China's comprehensive crime prevention strategy can be understood as an anthem to how financial incentivization came to appear both normal and even natural, despite the fact of it being so alien to the planned economy of the past.

Finally, for those unfamiliar with the literature on policy and practice of crime prevention in the People's Republic of China, this entry has perhaps had a somewhat "strange" quality principally because there has been little comment on policy evaluation. Bowles et al. (2005) point to generic obstacles to evaluation of policy in developing countries, namely:

- A lack of empirical information about the functioning of institutions
- A lack of articulation of policy options and policy objectives
- A lack of articulation of the inputs, of the outputs and outcomes associated with interventions, and thus of the raw material for estimating the costs and benefits associated with interventions.

The lack of research-led evaluation in China merely confirms this litany of obstacles (Liang and Lu 2006).

Related Entries

▶ Shaming in Asian Societies
▶ Triads and Tongs

Recommended Reading and References

Bakken B (2000) The exemplary society: human improvement, social control, and the dangers of modernity in china. Oxford University Press, Oxford

Bowles R, Akpokodje J, Tigere E (2005) Evidence-based approaches to crime prevention in developing countries. Eur J Crim Pol Res 11:347–377

Chen X (2004) Social and legal control in China: a comparative perspective. Int J Offender Ther Comp Criminol 48:523–536

Chen Y-L (2012) A political and economic analysis of the strike hard policy. Develop Law Soc 104:87–95 (in Chinese)

Dutton M (2005) Policing Chinese politics. Duke University Press, London

Ekblom P, Pease K (1995) Evaluating crime prevention. In: Tonry M, Farrington D (eds) Building a safer society: strategic approaches to crime prevention. University of Chicago Press, Chicago

Garland D (2001) The culture of control: crime and social order in contemporary society. Oxford University Press, Oxford

Hong D (2011) Research on changes to the mediation system in contemporary China. Shanghai People's Press, Shanghai (in Chinese)

Jiang B, Dai Y (1990) Mobilizing all possible social forces to strengthen public security – a must for crime control. Police Stud 13:1–9

Liang B (2005) Severe strike campaign in transitional China. J Crim Just 33:387–399

Liang B (2008) The changing Chinese legal system, 1978-present. Routledge, New York/London.

Liang B, Lu H (2006) Conducting fieldwork in China: observations on collecting primary data regarding crime, law, and the criminal justice system. J Cont Crim Just 22:157–172

Liang B, Wilson C (2008) A critical review of past studies on China's corrections and recidivism. Crim Law Soc Change 50:245–262

Liu J (2005) Crime patterns during the market transition in China. Brit J Criminol 45:613–633

Mao Z (1967/2010) Collected writings of Chairman Mao, volume 3 – on policy, practice and contradiction. El Paso Norte Press, San Francisco

Muhlhahn K (2009) Criminal justice in China: a history. Harvard University Press, Cambridge, MA

Ren X (1997) Tradition of the law and law of the tradition. Greenwood, Westport

Selden M (1971) The Yenan way in revolutionary China. Cambridge University Press, Cambridge

Situ Y, Liu W (1996a) Comprehensive treatment to social order: a Chinese approach against crime. Int J Comp Appl Crim Just 20:95–115

Situ Y, Liu W (1996b) Restoring the neighborhood, fighting against crime: a case study in Guangzhou city, People's republic of China. Int Crim Just Rev 6:89–102

Trevaskes S (2007a) Swift and severe justice in China. Brit J Criminol 47:23–41

Trevaskes S (2007b) Courts and criminal justice in China. Lexington Press, Lanham

Trevaskes S (2007c) The public/private security nexus in China. Soc Justice 34:38–55

Trevaskes S (2008) The war against organised crime and drug crime in china. Regional Outlook Paper No.15. Griffith Asia Institute, Brisbane, Australia

Trevaskes S (2010) The shifting sands of punishment in the era of ‘harmonious society’. Law Policy 32:332–361
Wang ZF (1989) (ed) Theory and practice of comprehensive management of public order in China. Masses Press, Beijing (in Chinese)
Welsh B, Hoshi A (2002) Communities and crime prevention. In: Sherman L, Farrington DP, Welsh BC, MacKenzie DL (eds) Evidence-based crime prevention. Routledge, London
Wilson AM, Greenblatt SL, Wilson RW (eds) (1977) Deviance and social control in Chinese society. Praeger, New York
Yang YM (2010) Analyzing the fourth wave of the strike hard policy. Gov Leg Syst 36:22–23 (in Chinese)
Zhang L, Messner SF, Liu J (2007a) A multilevel analysis of the risk of household burglary in the city of Tianjin, China. Brit J Criminol 47:918–937
Zhang L, Messner S, Liu J (2007b) Criminological research in contemporary China: challenges and lessons learned from a large-scale criminal victimization survey. Int J Offender Ther Comp Criminol 51:110–121
Zhao J (2005) Conflicts between advantages and disadvantages in restorative justice. Law 5:113–115 (in Chinese)
Zhao R, Liu J (2011) A system’s approach to crime prevention: the case of Macao. Asian J Criminol 6:207–227
Zhong L (2008) Communities, crime and social capital in contemporary China. Willan, Cullompton
Zhuo Y (2012) Social capital and satisfaction with crime control in urban China. Asian J Criminol 7:121–136

Intervention Research

► Bullying Prevention: Assessing Existing Meta-Evaluations

Interview and Interrogation Methods Effects on Confession Accuracy

Christopher E. Kelly[1], Allison D. Redlich[1], Jacqueline R. Evans[2] and Christian A. Meissner[3]
[1]University at Albany, State University of New York, Albany, NY, USA
[2]The University of Texas at Tyler, Tyler, TX, USA
[3]The University of Texas at El Paso, El Paso, TX, USA

Synonyms

Confessions; Criminal interviewing; Interrogation; Meta-analysis

Overview

The interviewing and interrogation of suspects is important to securing convictions against the guilty and freeing the wrongly accused. There are two general methods of questioning suspects: information gathering and accusatorial. The information-gathering approach, used in the United Kingdom, New Zealand, Australia, and elsewhere, is characterized by rapport building, truth seeking, and listening. The accusatorial approach, used primarily in the United States and Canada, is characterized by accusation, confrontation, psychological manipulation, and the disallowing of denials. Academics and practitioners hotly debate which method is more effective, particularly in light of increased awareness of the problems with false confessions. Two separate but related meta-analyses were conducted to address this question. The first relied upon data from five observational field studies and the second from 12 experimental, laboratory-based studies. The primary outcome measures were true and false confessions. Results revealed that both information-gathering and accusatory methods were associated with the production of confessions in field studies, though the experimental data indicated that the information-gathering method was more diagnostic in that it increased the likelihood of true confessions while reducing the likelihood of false confessions. Although continued research is needed to better understand the methods that produce true and false confessions, the results of these meta-analyses suggest that the information-gathering approach may be more effective in comparison to the accusatorial approach.

Introduction

The elicitation of false confessions is an international problem that has been documented in almost every continent (Kassin et al. 2010; Lassiter and Meissner 2010). Two general factors have been linked to the incidence of false confessions: personal (psychological) vulnerabilities of the individual and the use of accusatorial (psychologically based) interrogative methods. Accusatorial methods are common practice in the United States, Canada, and many Asian nations, and the use of these methods has been linked to false confessions (Kassin et al. 2010; Lassiter and Meissner 2010). In response to the increased awareness of the prevalence of false confession, the United Kingdom, Norway, New Zealand, Australia, and several other countries have amended their interrogation practices to employ information-gathering methods of interrogation (Bull and Soukara 2010). This entry summarizes the research evidence on the relative diagnosticity of the accusatorial interrogation and information-gathering interview approaches. Diagnosticity is the ratio of true to false confessions, with an effective interview/interrogation method maximizing true confessions while minimizing false confessions.

This entry is based on the systematic review conducted for and published by the Campbell Collaboration (http://www.campbellcollaboration.org) (Meissner et al. 2011). As will be shown below, the existing research evidence suggests that although both accusatorial and information-gathering approaches increase the likelihood of obtaining a confession in the field context, experimental laboratory studies demonstrated that the information-gathering approach yielded more diagnostic confessions (by increasing true confessions and reducing false confessions) when compared with the accusatorial method.

Key Issues and Controversies

Generally speaking, two distinct methods of suspect interview and interrogation approaches have emerged: information gathering and accusatorial. The information-gathering approach to interviewing and interrogation is based on establishing rapport between the interviewer and suspect and seeks to elicit information rather than obtain confessions. The interviewer uses positive confrontation and an open-ended questioning approach to elicit information from the suspect; psychological manipulation is anathema to the goals of the interview. This approach is exemplified and codified in Great Britain’s adoption of the Police and Criminal Evidence (PACE) Act of 1984 and later modified as the PEACE model in 1993. The PACE Act explicitly prohibited the use of psychologically manipulative techniques and required the recording of all custodial interrogations. Furthermore, the PEACE model focuses on developing rapport, explaining the allegation and the seriousness of the offense, emphasizing the importance of honesty and truth gathering, and requesting the suspect’s version of events. Suspects are permitted to explain the situation without interruption, and questioners are encouraged to actively listen. Finally, the information-gathering approach relies on cognitive-based strategies to detect deception.

In contrast, the accusatorial method is typified by the United States model and is best illustrated by the “Reid Technique” (Inbau et al. 2001). In this method, the interviewer establishes control over and confronts the suspect, presumes guilt, and ultimately aims to obtain a confession. As distinct from the information-gathering approach, accusatorial methods use closed-ended and confirmatory questions and employ psychological manipulation which generally involve three components: (a) *custody and isolation*, in which the suspect is detained in a small room and left to experience the anxiety, insecurity, and uncertainty associated with police interrogation; (b) *confrontation*, in which the suspect is presumed guilty and told (sometimes falsely) about the evidence against him/her, is warned of the consequences associated with his/her guilt, and is prevented from denying his/her involvement in the crime; and finally (c) *minimization*, in which a now sympathetic interrogator attempts to gain the suspect’s trust, offers the suspect face-saving excuses or justifications for the crime, and implies more lenient consequences should the suspect provide a confession. Lastly, this approach relies on anxiety-based cues (e.g., verbal, nonverbal, and paralinguistic)

Interview and Interrogation Methods Effects on Confession Accuracy, Table 1 Five distinguishing dimensions of information-gathering and accusatorial interrogation approaches

Dimension	Information gathering	Accusatorial
Interpersonal dynamic	Rapport building	Control oriented
Approach	Direct, positive confrontation	Psychological manipulation
Question type	Open-ended, exploratory	Closed-ended confirmatory
Primary goal	Elicit information	Obtain confession
Clues to deception	Cognitive	Anxiety

in order to detect when suspects are being deceptive. Table 1 summarizes the differences between the information-gathering and accusatorial approaches along four dimensions.

Meta-Analysis Methods

Researchers have relied on two primary methods of examining the influence of questioning technique on confessions: field studies of actual interrogations and experimental laboratory studies with mock crimes and transgressions. Whereas the field studies offer greater external validity, it is not possible to determine the "ground truth" or the veracity of confessions obtained in this context. In contrast, experimental studies allow for the manipulation of guilt and innocence but are limited in their ability to precisely model the context of a criminal interrogation. As such, two systematic reviews and meta-analyses of the accusatorial and information-gathering literature were conducted, one examining field studies and the other examining experimental studies.

The first meta-analysis focused on observational field studies; the second, on experimental laboratory-based studies. Although briefly summarized here, readers interested in more details regarding the methods are referred to Meissner et al. (2011). To identify studies that might be included in the analyses, more than a dozen reference databases were searched (using over 20 search terms), the reference sections of several comprehensive sources on interviewing and interrogation (books, articles, and reports) were reviewed, and researchers in the United States and overseas were contacted for any unpublished research. To be eligible for inclusion in the meta-analyses, field and laboratory studies must have met the following criteria:

- *Intervention*: Eligible field studies must have coded or quantified the use of at least one interview or interrogation technique, which was then categorized into information-gathering, accusatorial, or general interrogation approaches. Eligible laboratory studies must have involved the experimental manipulation of information-gathering and/or accusatorial methods with one another or with a control interview method (e.g., a simple request for compliance).
- *Outcomes*: Eligible field studies must have reported the analysis of confessions and have had sufficient quantitative data to calculate effect sizes, specifically including the relationship between the use of the certain interview/interrogation methods and elicitation of a confession. Laboratory studies must have reported the number of true and/or false confessions from "guilty" or "innocent" participants, respectively, in a manner that would permit calculation of an effect size.
- *Population/samples*: The population of interest for both the field and experimental studies was criminal (mock) suspects of any age, nationality, or status who were accused of committing a criminal act, a transgression, or withholding important information.

The initial search identified over 2,000 studies, though the vast majority was screened out. After this preliminary screening, 33 field studies and 22 laboratory studies were coded to determine final eligibility, of which five field studies and 12 laboratory studies were ultimately included in the meta-analyses. All included and excluded field and experimental studies (as well as the reasons for exclusion) can be found in Meissner et al. (2011).

Results of the Meta-analyses

Field Studies

The five field studies were identified through the systematic search, three of which were conducted in the United Kingdom, one in Canada, and one in the United States. These five studies represented data from 608 interrogation sessions. Of these sessions, 39.4 % were coded from audio recordings, 36 % from video recordings, 20 % were live interrogation sessions, and the remaining 4.4 % were from transcriptions of the sessions. The alleged crimes for which the suspects were being interrogated varied from social security benefit fraud to serious violent crime such as murder and robbery. Between 39 % and 64 % of the suspects in the interrogation sessions confessed or offered some admission of guilt. The studies classified the interrogations as accusatory, information-gathering, or combined interrogative method, resulting in eight possible pair-wise comparisons across the five study samples.

The effect size was Hedges' *g* based on dichotomous outcomes (i.e., confess vs. not confess) using the Cox computation method (see Lipsey and Wilson 2001; Sánchez-Meca et al. 2003). The relationship between the use of certain interrogative methods (accusatorial, information-gathering, or general interrogative methods) and elicitation of a confession was the focus of the analysis. All analyses were based on a random-effects model, and the results are presented in Table 2.

Three field studies assessed the relationship between use of accusatorial methods and elicitation of a confession in a real-world context and showed that such a method was associated with a large and significant increase in confession rates. Furthermore, there was no significant degree of variability across the studies suggesting consistency in findings across studies.

The relationship between information-gathering methods and elicitation of a confession was assessed through two field studies. These studies also found that the use of such a method was associated with a large and significant increase in confession rates. The results varied meaningfully between these two studies, however, with a substantially smaller effect in one study (.38 vs. 1.04).

Interview and Interrogation Methods Effects on Confession Accuracy, Table 2 Mean weighted effect sizes for field studies

Interrogative method	*k*	*N*	*g*	95 % CI	*Q*
Accusatorial	3	306	0.090***	(0.38, 1.41)	4.89
Information gathering	2	222	0.86*	(0.04, 1.69)	5.54*
General interrogative methods	3	422	0.19	(−.069, 1.06)	25.35***

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

A number of tactics observed in these studies could reasonably be coded as a part of accusatorial and information-gathering approaches. Three field studies examined the influence of these combined methods in eliciting confessions in a real-world context as opposed to those methods that might be exclusively linked to either accusatorial or information-gathering approaches, and showed no significant relationship between the use of these general methods and confession statements provided by suspects. However, these results were highly variable across studies, ranging from a low of − .40 to a high of .79.

Based on these field studies, it appears that the use of accusatorial and information-gathering methods of interrogation was significantly associated with the elicitation of confession evidence in a real-world context in contrast to those characterized as not accusatorial or information gathering, or even those techniques characterized as both accusatorial and information gathering. Although these results suggest that such methods are effective tools for elicitation of confession evidence, it is important to note that these findings fail to distinguish the diagnostic value of the interrogative evidence – field studies offer little or no information to distinguish between innocent and guilty suspects, and ground truth in such contexts is nearly impossible to determine. As such, researchers have assessed the diagnostic value of interrogative methods

Interview and Interrogation Methods Effects on Confession Accuracy, Table 3 Mean weighted effect sizes for experimental studies

Interrogative contrast	Outcome	*K*	*N*	*g*	95 % CI	*Q*
Accusatorial versus control	True confession	6	272	0.46*	(0.06, 0.86)	7.52
	False confession	14	892	0.74***	(0.35, 1.12)	32.99**
Information gathering versus control	True confession	2	110	0.67*	(0.02, 1.32)	1.41
	False confession	2	110	−0.23	(−0.98, 0.52)	0.11
Accusatorial versus information gathering	True confession	3	215	0.64*	(0.01, 1.28)	3.62
	False confession	3	215	−0.77*	(−1.46, −0.08)	4.43

$*p < .05$; $**p < .01$; $***p < .001$

by modeling the interrogative process in an experimental, laboratory context.

Experimental Studies

Twelve experimental studies were identified through the systematic search, and these studies provided 30 contrasts between the different interrogation methods. All but one of the studies was conducted in the United States. Nine were published in peer-reviewed journals and three are currently unpublished.

Eleven of the 12 studies used variations of the Kassin and Kiechel (1996) or the Russano et al. (2005) paradigm. The Kassin and Kiechel paradigm is one in which all participants are "innocent" of the mock crime of crashing the computer. The Russano et al. paradigm includes participants randomly assigned to an innocent or guilty condition of a known, intentional act (i.e., cheating in an academic context). Eleven of the studies used undergraduate students as participants, with two studies including participants from other age groups. The results are presented in Table 3.

Accusatorial versus *Control*. The contrast between an accusatorial interrogative method and a control interview condition demonstrated that accusatorial methods yielded a moderate and significant increase in the frequency of true confessions and large and significant increase in false confessions. Typically, the control technique was the absence of a specific technique; for example, if the experimental manipulation was the presentation of false evidence, the control condition did not present false evidence. The findings for false confessions were highly variably across studies.

Information Gathering versus *Control*. Two studies examined the influence of information-gathering interrogative methods versus a control condition in eliciting true confessions and false confessions. These studies demonstrated that information-gathering methods yielded a greater frequency of true confessions, but did not significantly influence the likelihood of eliciting false confessions.

Accusatorial versus *Information Gathering*. Three studies assessed the direct contrast between accusatorial and information-gathering interrogative methods in eliciting true confessions and false confessions. The results demonstrated that information-gathering methods produced a significantly greater frequency of true confessions, while significantly reducing the frequency of false confessions, when compared with accusatorial methods. Neither analysis produced significant variability, suggesting consistency in findings across studies (*Q*s < 4.43, *ns*).

This meta-analysis of the eligible experimental literature demonstrated several key findings that may have implications for policy and practice. First, while accusatorial methods significantly increased the likelihood of obtaining a true confession (when compared with a no-tactic control condition), these methods also significantly increased the likelihood of obtaining a false confession – a finding consistent with many cases of wrongful conviction in the United States (see Kassin et al. 2010; Lassiter and Meissner 2010). In contrast to this, information-gathering approaches significantly increased true confession rates, but showed no significant increase in the rate of false confessions when compared with a no-tactic control condition. In fact, information-gathering approaches appeared to show a numerical decrease in the rate of false confessions obtained. When compared directly against accusatorial methods, information-gathering approaches showed superior diagnosticity by significantly increasing the elicitation of true confessions and significantly reducing the incidence of false confessions.

Future Directions

A number of variables were considered for inclusion in a moderator analysis of the influence of accusatorial methods in eliciting false confessions. Unfortunately, studies varied little in several key factors. For example, only two of the 14 independent samples involved children or adolescents, while the remainder involved college students. In addition, none of the studies manipulated race or ethnicity in participant recruitment or analyses of the data. Similarly, only one study was conducted outside of the United States.

It is important here to note that field studies fail to offer us important information regarding the relative diagnostic value of the confession that is elicited. That is, such studies lack "ground truth" that would enable us to factually determine the veracity of the statement provided by a suspect and thereby preclude the ability to assess the diagnostic value of the information elicited and the effectiveness of such techniques when employed in the field. One method often used to assess veracity in field studies has been to evaluate the "strength" of available evidence against the defendant; however, none of the studies took this approach to evaluating the likely credibility of the confession evidence obtained as a moderator of interrogative efficacy.

Additionally, each of the studies included in the field study meta-analysis examined the bivariate relationship between certain interrogative methods and elicitation of a confession. As indicated in the review, a number of control variables could reasonably be included in such analyses (e.g., factors related to interrogator experience, crime type, interrogator/suspect ethnic backgrounds, geographic characteristics), and more complex modeling approaches (such as multilevel modeling or path analysis) could have been pursued, albeit many (if not all) of these studies may not have had a large enough sample size to consider multiple factors simultaneously. Researchers are strongly encouraged to initiate more systematic, multilevel analyses of the influence of interrogative methods. Further, there is a great need for the use of quasi-experimental methods in the field context as the effects of certain interrogative methods are currently unknown. Quasi-experimental methods could include the examination of the influence of certain factors in real-world interviews and interrogations, such as the use of the cognitive interview, whether suspects are told they are being recorded, and many of the variables under consideration here. Such quasi-experimental methods are effective tools for assessing the policy implications of alternative approaches to police interviewing and interrogation and should be considered in the years ahead.

Conclusions

The meta-analyses summarized above suggest that the information-gathering approach introduced by Great Britain can be effective in eliciting confession evidence but also has the advantage of eliciting more diagnostic information. In the experimental meta-analysis, when the information-gathering and accusatorial approaches were contrasted, the information-gathering approach clearly produced more advantageous outcomes (although caution is warranted given the small number of eligible studies). Specifically, the information-gathering approach produced significantly more true confessions, whereas the accusatorial approach produced significantly more false confessions. As such, the results suggest that law enforcement, military, and intelligence agencies should consider the use of information-gathering approaches to interrogation. In addition, research should be conducted to further refine and solidify the understanding of the effects of various interrogative methods in eliciting true and false confession evidence, therein providing a stronger foundation for evidence-based practice and policy recommendations.

Related Entries

- ▶ Comparing Police Systems Across the World
- ▶ Cross-National Performance in Policing
- ▶ False Confessions and Police Interrogation
- ▶ Law of Police Interrogation
- ▶ Methodological Issues in Evaluating Police Performance
- ▶ Randomized Experiments in Criminology and Criminal Justice
- ▶ Video Technology and Police Interrogation

Recommended Reading and References

Bull R, Soukara S (2010) What really happens in police interviews. In: Lassiter GD, Meissner CA (eds) Police interrogations and false confessions: current research, practice, and policy recommendations. American Psychological Association, Washington, DC

Cox DR (1970) Analysis of binary data. Chapman & Hall/ CRC Press, New York

Educing information. Interrogation: Science and Art. Intelligence Science Board, Phase 1 Report (2006). National Defense Intelligence College, Washington, DC

Gudjonsson GH (2003) The psychology of interrogations and confessions. Wiley, Chichester

Inbau FE, Reid JE, Buckley JP, Jayne BC (2001) Criminal interrogation and confessions, 4th edn. Aspen, Gaithersberg

Justice BP, Bhatt S, Brandon SE, Kleinman SM (2009) Army field manual 2-22.3 interrogation methods: A science-based review. Unpublished manuscript

Kassin SM, Drizin S, Grisso T, Gudjonsson G, Leo RA, Redlich AD (2010) APLS-approved white paper, police-induced confessions: risk factors and recommendations. Law Hum Behav. doi:10.1007/ s10979-009-9188-6

Kassin SM, Gudjonsson GH (2004) The psychology of confessions: a review of the literature and issues. Psychol Sci Public Interest 5:33–67

Kassin SM, Kiechel KL (1996) The social psychology of false confessions: compliance, internalization, and confabulation. Psychol Sci 7:125–128

Lassiter GD, Meissner CA (2010) Police interrogations and false confessions: current research, practice, and policy recommendations. American Psychological Association, Washington, DC

Lipsey MW, Wilson DB (2001) Practical meta-analysis. Sage, Thousand Oaks

Meissner CA, Redlich AD, Bhatt S, Brandon S (2011) Interview and interrogation methods and their effects on true and false confessions. Final report to the campbell collaboration, crime and justice group. http://www.campbellcollaboration.org/

Russano MB, Meissner CA, Narchet FM, Kassin SM (2005) Investigating true and false confessions within a novel experimental paradigm. Psychol Sci 16:481–486

Sánchez-Meca J, Marín-Martínez F, Chacón-Moscoso S (2003) Effect-size indices for dichotomized outcomes in meta-analysis. Psychol Methods 8(4):448–467

Schollum M (2005) Investigative interviewing: the literature. New Zealand Police Department. Retrieved 15 Jan 2006. http://www.police.govt.nz/resources/2005/investigative-interviewing/investigative-interviewing.pdf

Williamson T (2006) Investigative interviewing: rights, research, and regulation. Willan Publishing, Devon

Interviewing Eyewitnesses

Julie Gawrylowicz and Amina Memon
Department of Psychology, Royal Holloway,
University of London, Egham, UK

Overview

Eyewitness researchers have attempted to take what we know about memory and social influence and wrap it into a set of procedures for interviewing eyewitnesses. This has resulted in a number of empirically based investigative interviewing tools. Some of these tools have been successfully implemented into current police practice. The most developed and researched procedural package for gathering detailed reports from cooperative eyewitnesses without compromising overall accuracy is the Cognitive Interview (CI). Since its development 25 years ago, novel and alternative interviewing strategies have been developed, including the Self-Administered Interview© (SAI), a pen-and-paper version of the CI. The SAI can be administered at the crime scene and to multiple witnesses simultaneously, thereby reducing police time and resources.

Although a lot has been achieved during the past two decades with regard to improving and standardizing investigative interviewing of witnesses, victims, and suspects, especially in the UK, there remain controversial policy issues, such as how to best train investigative interviewers and which methods to use for recording interviews. Under researched areas include how to best interview vulnerable suspects and procedures and practices for interviewing traumatized victims and witnesses. It is important to promote the collaboration between practitioners and academics to ensure that investigative interviewing practice continues to evolve and is driven by contemporary theories and research findings.

Fundamentals

Investigative interviewing is an essential information-gathering tool in any context where the interviewer needs to find out what if anything has happened and where the goal is to achieve as accurate recall as possible. The need to obtain credible information in an investigative context has become even more important given the increase in terrorist offences around the world. Köhnken (1995), in a review of the information-processing approach to interviewing, highlights the numerous uses of an interview ranging from an expert witness interviewing children and adults during competency assessments and custody disputes to obtaining information about present moods, attitudes, opinions and to establishing rapport with the interviewee. This entry focuses primarily on the forensic applications of interviewing where the goal is to obtain accurate and complete information from a victim or witness about some action or behavior they experienced in the past.

In order to be able to interview eyewitnesses effectively, police officers and other members of the Criminal Justice System need to understand how memory works. It is crucial to consider which factors may impact memory and how to obtain reliable and accurate information about past events. In a nutshell, the process of remembering information can be subdivided into three different components: encoding, storage, and retrieval (Tulving 1995). Encoding refers to the initial intake of information by our sensory systems. Thus, for an eyewitness to be able to describe the face of a perpetrator during a police interview, he or she would need to have paid attention to the perpetrator's face when the crime occurred. Assuming the information is encoded, it will be stored in our memory system until it needs to be accessed again. Information that has been successfully encoded and stored can then be retrieved.

Why Does Our Memory Fail?

People may fail to remember information due to problems with encoding, storage, or retrieval. For instance, information might not have been encoded in the first place. This might happen due to external factors such as poor visibility or objects obstructing the view or internal factors such as the witness not paying attention to specific details. Several cognitive psychology studies have investigated the change-blindness phenomenon, the failure to observe large changes in a scene which would normally not go unrecognized (Simons and Rensink 2005). A well-known psychology study of change blindness is the gorilla experiment by Simons and Chabris (1999) in which participants were asked to watch a video clip showing a basketball match and to count how many passes the players were making. In the middle of the clip, a gorilla walked through the scene. It was found that more than half of the participants (66 %) failed to notice the gorilla. The change-blindness phenomenon has been replicated numerous times with different paradigms and different stimuli (see Simons and Rensink (2005) for an overview), and the findings have strong practical implications particularly in forensic settings. During the occurrence of a crime, the failure to detect changes in the scene can have dramatic consequences. Investigative officers need to be aware that problems with encoding are not uncommon in humans and might influence what the witness subsequently recalls and how accurate this information is.

Some information we encode may enter our memory but will be forgotten quickly afterwards. Not all information is stored efficiently, and most information is forgotten soon after the incident has happened (Ebbinghaus 1964/1885). Therefore, it is not unusual for an eyewitness to say, "I don't remember what the perpetrator looked like."

During the interview process itself when the witness engages in retrieval, several factors may impact how much will be recalled and how accurate this information is. The way a question is phrased and characteristics of the interviewer can influence how much a witness discloses and/or remembers and how accurate this information is likely to be (Milne and Bull 1999). Generally, open-ended questions and free-recall instructions, such as "Tell me everything," do elicit accurate responses from witnesses, but often the provided information is sparse (Davies et al. 1989). The interviewer might feel the need to ask more specific questions to generate a detailed account or to elaborate on specific details; however, the more specific the questioning style becomes, the less accurate the responses may be. Leading and suggestive questions should be avoided, since they imply the desired answer to the interviewee.

The misinformation studies by Elizabeth Loftus provide a potent illustration of this problem. During a well-known and often-replicated experiment (Loftus et al. 1978), participants were presented with slides depicting an automobile-pedestrian accident at a junction with either a yield or a stop sign. Subsequently, half of the participants were asked whether the car stopped at a stop sign and the other half whether it stopped at a yield sign. Thus, half of the participants received misleading post-event information in the form of a suggestive question. After a delay, participants engaged in a forced-choice recognition test. The two presented choices were slides depicting the car accident with either the yield sign or the stop sign. A significant difference was obtained between the control and the misled group, with the misled group making considerably more errors and reporting considerably more often to have seen the misleading post-event information in the original slide.

Misleading post-event information can manifest itself in many different ways, such as in written format or provided by an individual, for example, a co-witness. The feedback a police officer may give to a witness, for example, during or after an identification task, can be also regarded as post-event information and may influence how confident the witness is about the task at hand, as well as how he or she remembers the original event. Wells and Bradfield (1998) conducted a study during which participants had to identify suspects from photo spread arrays. After identifications had been made, participants received either confirming feedback that they picked the suspect or disconfirming feedback that they selected the wrong person. Feedback manipulation did not only affect participants' reports about how confident they were when they made the identification but also the accounts about the original event, such as memories about the viewing conditions present at the original event. These studies highlight how important it is for an investigative interviewer to be aware of the fragility of memory and how differences in questioning style and content can influence the completeness and accuracy of a witness account.

Current Investigative Interviewing Techniques in the UK

Given the importance of accurate recollection during investigative interviews, researchers have developed techniques to increase the quantity and quality of witness accounts, which are based on theoretical knowledge about how our memory works (Fisher et al. 2011). Some of these techniques are developed for specific witness populations, such as children, whereas others are developed to be used in much broader settings.

The National Institute of Child Health and Human Development (NICHD) developed a structured interview protocol for interviewing child witnesses, specifically focusing on child abuse victims (Lamb et al. 2007). The protocol can be divided into two phases: the pre-substantive and substantive interview phase. During the pre-substantive phase, the interviewer introduces himself or herself and explains the purpose and the ground rules of the interview. The rapport phase creates a comfortable and supportive atmosphere and enhances the interpersonal relationship between interviewer and interviewee. The child will be asked to report an experienced event unrelated to the incident in question, to familiarize himself or herself with the open-ended questioning approach, and to demonstrate the level of detail which is expected during the substantive interview phase. During the transition from pre-substantive to substantive phase, a couple of open-ended prompts are used to identify the target event under investigation. If the child makes an allegation, the interviewer starts the substantive phase with an open invitation ("Tell me what happened") followed by other free-recall prompts ("Tell me more about it" or "What happened then?"). After the initial free narrative, the interviewer prompts the child with cued questions ("When did it happen?" or "Who was there?") to obtain a more complete and detailed account. If necessary, option-posing questions can be used in the end to obtain crucial details not previously disclosed by the child. Interviewers are advised to not use any suggestive questions at any time during the interview (Lamb et al. 2007). The NICHD has been tested empirically in several field studies in different countries, among others, in the UK, France, and Israel. A recurrent finding is that interviewers using the NICHD protocol asked in general more open-ended questions and fewer focused, option-posing, and suggestive questions than interviewers who employed a standard interview. In addition, interviewers using the NICHD technique imposed focused, option-posing, and suggestive questions much later during the progression of the interview. Regardless of age, children interviewed with the NICHD protocol provided more information during the open-ended prompts, and significantly fewer details were revealed with direct, option-posing, and suggestive questions. Importantly, no age differences were found, suggesting that the NICHD protocol can be regarded as a suitable interview technique for children as young as 3 years of age (Cyr and Lamb 2009).

An interview technique focusing more on the general population is the Cognitive Interview (CI). It was developed by Ed Geiselman (University of California, Los Angeles) and Ron Fisher (Florida International University) in the 1980s (Fisher and Geiselman 1992) and since then has been tested and applied in various interviewing settings and with different witness populations, including adults with and without intellectual disabilities, children, and the elderly (see the meta-analysis by Memon et al. (2010) for a comprehensive overview). In its structure, the CI is very similar to the NICHD; however, it is unique in that it comprises several memory-retrieval techniques also known as mnemonics, which are based on fundamental theoretical principles regarding memory organization, storage, and retrieval. One of these principles is Tulving and Thomson's encoding specificity hypothesis (1973), which states that successful retrieval of information is most likely when the context and the cues present at retrieval match those present at the initial encoding. Reinstatement of the initial encoding context should therefore lead to an increase in accessibility of the stored information. Godden and Baddeley (1975) tested this assumption empirically by asking participants to learn word lists either underwater or on land. Subsequently half of the participants who learned the words underwater engaged in retrieval underwater, whereas the other half retrieved the words on land. The same retrieval procedure was applied to participants who encoded the words on land. Intriguingly, participants who encoded the words underwater were better at recalling them underwater compared to on land, whereas participants who learned the word lists on land recalled more items on land than underwater.

In an eyewitness situation, it is unlikely that cues which were present during encoding will be physically present at retrieval, since the investigative interview will probably take place in a police station. However, witnesses can be encouraged to reinstate the context of the original event mentally, thereby inducing cues which may facilitate their recall. The second memory principle is Bower's multiple-component memory trace theory (1967), which proposes that our memory is a network of associations rather than an accumulation of single, unconnected incidents. Consequently, there are multiple ways to access or cue one specific memory.

Initially, the CI consisted of four general retrieval techniques: mental context reinstatement, report-everything instruction, change-perspective, and reverse-order technique. During the mental context reinstatement, the interviewee is instructed to form a mental image of the physical and personal context present at the witnessed event. The interviewer can help the interviewee during this task by encouraging them to imagine what they could see, hear, and smell when the incident occurred and to think about any experienced feelings. The second technique is to ask the interviewee to report everything they have seen, including small, unimportant, and partial information. This technique may encourage witnesses to recall the event from a more detailed memory level. Research by Milne and Bull (2002) suggests that the mental context reinstatement together with the report-everything instruction is the most effective component of the CI. The change-perspective and reverse-order techniques are based on the assumption that information not accessible via one route might be accessible via another route. The change-perspective technique encourages the interviewee to recall the event from different perspectives, for example, from the perspective of another witness or from another angle. The reverse-order technique asks the witness to recall the event from different starting points or from a different chronological order. Both techniques encourage the interviewee to use different retrieval routes to access the information in question.

The CI was revised several times to increase its effectiveness and to make it more suitable to the needs of different witness populations. The revisions focused mostly on the enhancement of social dynamics and communication between interviewer and interviewee (Fisher et al. 2011). The following techniques are part of the revised CI: rapport building, focused retrieval, and interviewee compatible questions. Rapport building refers to the attempt to get to know the witness better and to make him/her feel more at ease. An important part of the rapport building is the transfer of control, during which the interviewer makes clear that it is the witness who has the knowledge of what happened and the interviewer does not know anything. Therefore, the interviewee is the active one doing all the talking, while the interviewer, who is the listener, is passive. During the questioning phase, where the interviewer takes the lead in asking questions based on what the witness has said in free recall, some additional techniques are used. Focused retrieval encourages the interviewee to form mental images of specific details. Interviewee compatible questions are guided by the interviewee's pattern of recall and prevent disruptions during retrieval. Empirical tests as well as field studies have revealed that the CI elicits significantly more correct information than control interviews (25–50 % more), without a significant decrease in accuracy rates (see Fisher et al. (2011) for a review and for a meta-analysis; Memon et al. 2010).

Although the beneficial effects of the CI are well established through laboratory and field studies, research has revealed that police officers often do not follow the whole procedure and/or only use aspects of the CI they consider as useful (Oxburgh and Dando 2011). Moreover, officers tend to find the CI too time consuming, difficult to administer, and sometimes not appropriate for specific cases they are dealing with. In particular, police officers report that they have difficulty applying the mental context reinstatement component of the CI (Dando et al. 2009). On the basis of these practical limitations, Dando and colleagues developed a modified mental context reinstatement technique called Sketch Mental Context Reinstatement (MCR). The sketch procedure appears to be more time efficient and easier to administer especially for frontline police officers.

In a recent laboratory study (Dando et al. 2009), participants received either Sketch MCR, traditional MCR, or no MCR during a structured interview protocol after having watched a criminal event. Participants in the Sketch MCR condition were provided with a pen and paper and were instructed to draw a detailed plan or sketch of the event they had witnessed. Interview protocols including the original MCR and the new Sketch MCR elicited significantly more correct details than the no MCR condition. The accuracy rate was significantly higher for Sketch MCR interviews than for original and no MCR interview protocols. These findings imply that a CI interview protocol including the Sketch MCR technique can be regarded as effective as an original CI. Beyond that, CI protocols comprising the Sketch MCR were significantly shorter in duration than the original MCR CI protocols, which can be regarded as an enormous benefit considering the limited resources available to frontline police officers. In addition to the pragmatic benefits, the Sketch MCR technique may be also particularly suitable for vulnerable witness populations, such as the elderly, since it places fewer cognitive demands on the witness.

Recently, a pen-and-paper version of the CI was developed by Gabbert et al. (2009) to protect witness memory from the negative effects of delay and misleading post-event information. The novel recall tool is called Self-Administered Interview© (SAI) and can be administered to witnesses after the incident has happened. Empirical research has shown that it is crucial to interview witnesses as soon as possible after an incident has occurred to be able to obtain accurate and reliable accounts (Tuckey and Brewer 2003); however, this is often not possible due to demands on police resources and time. In addition, the police frequently have to deal with incidents where there are multiple witnesses at the scene, such as during natural disasters, mass accidents, or terrorist attacks. It is not always possible for the police to interview all witnesses simultaneously, and often witnesses are not interviewed for several days, weeks, or even months after the incident has happened. During these delays, forgetting may occur and memory is also more susceptible to misleading information. The SAI facilitates police practice, since it can be administered promptly at the scene of an incident. It is totally self-explanatory and can be completed by witnesses in their own time (Gabbert et al. 2009).

The most recent version of SAI comes in the form of a booklet comprised of seven sections with information, instructions, and cues that aim to facilitate witness recall and are essentially based on the underlying memory principles of the CI (Hope et al. 2011). Section A includes mental context reinstatement instructions which encourage the witness to place themselves back to the point in time when the incident has taken place and to think about the event in as much detail as possible before writing down the memories. Thereafter, witnesses are asked to “report everything” that they remember. The next section focuses on eliciting a detailed person description by asking the witnesses several questions about the perpetrators appearance (e.g., hair, build, distinguishing features). Section C encourages witnesses to draw a sketch of the crime scene to generate important spatial information, which might be especially important during road traffic incidents, and to get additional information about possible escape routes of the perpetrators. In Section D, witnesses are instructed to describe any other people who were present at the crime scene, also when they were not directly involved to identify other witnesses. Section E requests information about any present vehicles and Section F about general conditions which may have impacted the visibility of the event, such as weather, obstructions, and time of day. The final Section G allows witnesses to report any additional information not previously mentioned or covered in any of the prior sections (Hope et al. 2011).

Early laboratory studies revealed that participants provided considerably more accurate information about a to-be-remembered event when interviewed with the SAI compared to free-recall instructions. The amount of correct information recalled was similar to that elicited with a C-I. Furthermore, participants who received the SAI directly after they witnessed the event remembered more information about it during a 1-week later free-recall test than individuals who had not completed the SAI (Gabbert et al. 2009).

In 2009, field trials of the SAI started in England and Wales initiated by the Association of Chief Police Officers (ACPO) Investigative Interview Committee. The main aims of the field trials were to examine the type of information provided by witnesses and the nature of event, to get end-user evaluations, and to identify situations in which the SAI is most useful (Hope et al. 2011). Trials have also focused on police officers' opinions about the efficiency of the SAI. It was found that the majority of officers experience the SAI as user-friendly and time saving. However, some officers raised concerns such as that it is only suitable for cooperative witnesses, that vital information may be missed, and that officers may use it incorrectly. The Greater Manchester Police in the UK is now providing officers with a SAI training package to inform them about its usage and theoretical underpinnings (Hope et al. 2011).

Policy Issues

Given the extensive research on investigative interviewing, predominantly on the CI technique, and the light it has shed on faulty interviewing practices, the question can be asked whether police interviews have improved in the 20 years since the CI was first introduced. One of the main issues that arises from the in-depth analysis of the research literature on investigative interviewing is that, despite numerous psychological studies, police officers are reluctant to use the complete enhanced version of the CI (Kebbell et al. 1999). In a recent analysis of police interview techniques, Fisher and Schreiber (2007) requested 23 detectives in Miami experienced in investigations of robbery, sexual assault, homicide, and internal affairs to tape-record their witness interviews. Analysis of these interviews revealed interviewing behaviors that were similar to those identified 20 years earlier. This was particularly disappointing in view of the scientific progress made in the field and the efforts to disseminate findings to practitioners and implement training programs. The evidence from the UK suggests that many police officers are inadequately trained in investigative interviewing techniques. In an evaluation of 75 police interviews, for example, Clarke and Milne (2001) found no evidence of the CI procedure having been used at all in the vast majority (83 %) of interviews. As a result of this, there have been changes in the way training is delivered in the UK. Most police officers are subject to a tiered approach with novice police officers receiving basic training in interview skills (the Tier 1 procedure), and as they progress through their career, they can acquire additional interview tiers. Advanced courses build on skills acquired in earlier courses, and there are courses directed at specialist interviewers such as police working in child protection. The tiered model encourages continuous learning about and acquisition of advanced knowledge on the best interviewing practices and recent discoveries in cognitive psychology. Furthermore, it ensures that police officers will only learn about interviewing techniques which are most relevant to their current role and responsibilities in their career. It is essential to continuously monitor and assess police interviewing practice and training to ensure best practice.

Another controversial issue with respect to investigative interviewing is whether interviews should be electronically recorded and be made available to the courts and experts. In some parts of the UK (England and Wales), interviews with witnesses are electronically recorded, whereas in other countries (Scotland), interviews with witnesses are based on hand-written notes taken during the interview. This may be problematic as we know from research that interviewers' recollections regarding the content and structure of interviews are far from being infallible. Lamb et al. (2000) compared 20 audiotaped forensic interviews of sexual abuse cases with the verbatim notes made by interviewers during the course of the interview. It was revealed that interviewers often failed to report in their notes many of the details provided by the children and many of the utterances they used to elicit them. Twenty-five percent of forensically relevant information was not provided in the notes. In addition, interviewers often failed to accurately recall the question format utilized and stated that they used more often open-ended questions than they actually did. The failure to accurately recall when and which questions are asked during an interview can have severe consequences especially in forensic settings, such as undermining the credibility of a child witness during a sexual abuse case. It is crucial to use objective and accurate methods of recording, such as video recordings, which do not affect adversely the credibility of the evidence.

Future Research Directions

It should be noted that the NICHD, the CI, and the SAI all focus on interviewing cooperative witnesses. Few studies have investigated the effectiveness of these protocols with uncooperative witnesses, such as suspects. Suspect interviewing plays an important part during criminal investigations, particularly to gather further information or evidence about additional perpetrators and other related criminal activities. In the USA, suspect interviewing is very confrontational, and the most frequently taught and used interview technique is the so-called Reid technique, which focuses on identifying truth tellers and liars and eliciting a confession (Kassin et al. 2010). In contrast, suspect interviewing in the UK changed dramatically in the early 1990s from "obtaining a confession" to "information gathering" (Bull and Milne 2004). Since then, recommendations on suspect interviewing were published and a new interviewer training program was developed which standardized for the first-time interview practices in England and Wales (Bull and Milne 2004). The aim of the new guidelines and the training program was to make suspect interviewing more transparent and fair. The training constitutes of a five-part model called PEACE, which stands for planning/preparation, engaging and explaining to the suspect, account from the suspect, closure, and evaluation (Bull and Milne 2004). Although the PEACE model has been successfully implemented in current UK police practice, there remain unanswered questions. O'Mahony, Milne, and Grant (2012) raise the issue that there is no guidance available on how to best interview vulnerable suspects, such as suspects with an intellectual disability. This is an important topic, given the high prevalence rate of people with intellectual disabilities in the UK (Emerson et al. 2001) and their increased risk of getting involved in criminal activities, due to living in underprivileged neighborhoods with high crime rates and/or being more likely exploited by criminals to assist in illegal activities because of their low understanding of their actual involvement in a crime and their heightened need to be accepted by other people (Davis 2005). The majority of research in the area of offenders with intellectual disabilities has looked at assessment and treatment of sexual offenders (see the book by Craig et al. (2010)). Empirical research on investigative interviewing has focused merely on interviewing vulnerable victims and witnesses (Milne 1999) and has neglected somewhat the research question of how to best interview vulnerable suspects.

Another area which needs to be researched in more depth is the impact of trauma on witness memory and how the different interview protocols work best with traumatized witnesses. For example, should the SAI be administered at the scene of a crime to a traumatized witness, or should witnesses and victims first be assessed and allowed to recover before they are subjected to an interview? Would a face-to-face CI meet the special needs of these types of witnesses better than the SAI? Until now psychological research has focused predominantly on the impact of emotional events on people's memory, without considering the impact of the applied interview technique (Reisberg and Hertel 2004). Some evidence suggests that highly emotional or distinctive events are remembered well and very vividly. Researchers call these memories flashbulb memories, because of their intense and detailed nature (Brown and Kulik 1977). Many people report, for example, to have remembered the death of Princess Diana (1997) or the 9/11 attacks on the World Trade Center (2001) as if they had happened yesterday. People are also often very confident in the accuracy of such flashbulb memories. However, a detailed and confidently held memory does not necessarily need to be accurate as demonstrated by Neisser and Harsch (1992). They interviewed people after the 1986 Challenger space-shuttle disaster about their whereabouts when the incident had happened and when and where they first heard about it. Three years later, the same people were reinterviewed. People still reported detailed and highly confident memories; however, these memories were often prone to errors and sometimes were even completely different from their first recollections. Findings from laboratory studies suggest that arousal can aid as well as impair memory for an experienced event and that it strongly depends on the type of detail one is trying to recall. Central details, such as details associated with the gist of the event or which are spatially central, are usually remembered very accurately, whereas peripheral details, which are plot irrelevant, are often remembered poorly (Reisberg and Hertel 2004). Thus, witnesses of and victims to traumatic events can provide accurate accounts; however, they may be less complete. The empirical and practical question of which interview technique is best for interviewing traumatized witnesses and victims and the most appropriate time at which to interview them remains as yet unclear.

Conclusions

The knowledge acquired through psychological research has helped to develop interviewing techniques which are based on well-established psychological theories regarding memory, communication, and social interaction. The UK, by basing its national investigative interviewing strategy on psychological theory and research, provides a good model for other jurisdictions to follow; however, this does not mean that enough has been achieved. Enhancing the efficacy of investigative interviews is a continual process, and there is always room for improvement, especially at a time where police forces face major cuts in their resources and need to consider alternative more time-efficient methods to interview eyewitnesses. The SAI provides one way in which police can save time and resources and the available research suggests it might be a practical alternative to the face-to-face interview. Of course, there will be situations where the police will need to assess the suitability of a witness for an investigative interview such as the CI particularly if the witness is traumatized or intellectually challenged. This is an area where further research is needed and specialist interview teams need to consider how best to deal with vulnerable witnesses or those with special needs. Another area where interviewing practices need further consideration is in dealing with offenders, particularly those with intellectual disabilities. It may be that existing protocols can be modified to meet the needs of specific groups. Practitioners working in the Criminal Justice System and academics need to work together to ensure that investigative interviewing follows best practices and that miscarriages of justice based on faulty interviewing practices are minimized.

Related Entries

- ▶ Estimator Variables and Eyewitness Identification
- ▶ Evaluating Truthfulness: Interviewing and Credibility Assessment
- ▶ Eyewitness Identification Evidence
- ▶ Eyewitness Research
- ▶ False Confessions and Police Interrogation
- ▶ Interview and Interrogation Methods Effects on Confession Accuracy
- ▶ Investigative Psychology
- ▶ Jury Decision Making and Eyewitness Testimony
- ▶ Law of Police Interrogation
- ▶ Victims and the Police

Recommended Reading and References

Bower G (1967) A multicomponent theory of the memory trace. Psychology of Learning and Motivation 1:229–325

Brown R, Kulik J (1977) Flashbulb memories. Cognition 5(1):73–99

Bull R, Milne B (2004) Attempts to improve the police interviewing of suspects. In: Lassiter GD (ed) Interrogations, confessions, and entrapment (pp. 181–196). New York: Kluwer Academic

Clarke C, Milne R (2001) A national evaluation of the PEACE Investigative Interviewing Course. London: Home office

Craig L, Lindsay WR, Browne KD (2010) Assessment and treatment of sexual offenders with intellectual disabilities: a handbook: Wiley-Blackwell, Chichester

Cyr M, Lamb ME (2009) Assessing the effectiveness of the NICHD investigative interview protocol when interviewing French-speaking alleged victims of child sexual abuse in Quebec. Child Abuse & Neglect 33(5):257–268

Dando C, Wilcock R, Milne R (2009) The cognitive interview: the efficacy of a modified mental reinstatement of context procedure for frontline police investigators. Appl Cogn Psychol 23(1):138–147

Davies G, Tarrant A, Flin R (1989) Close encounters of the witness kind: children's memory for a simulated health inspection. Br J Psychol 80(4):415–429

Davis LA (2005) People with intellectual disabilities in the criminal justice system: victims & suspects. Arc Q & A

Ebbinghaus H (1964/1885) Memory: a contribution to experimental psychology. Dover, New York

Emerson E, Hatton C, Felce D, Murphy G (2001) Learning disabilities: the fundamental facts. The Foundation for People with Learning Disabilities, London

Fisher RP, Geiselman RE (1992) Memory-enhancing techniques for investigative interviewing: the cognitive interview. Charles C Thomas, Springfield, Illinois, US

Fisher RP, Schreiber N (2007) Interview protocols for improving eyewitness memory. In: Toglia MP, Read JD, Ross DF, Lindsay, RCL (eds) The handbook of eyewitness psychology. Memory for events, vol I, pp 53–80, Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers

Fisher RP, Milne R, Bull R (2011) Interviewing cooperative witnesses. Curr Dir Psychol Sci 20(1):16

Gabbert F, Hope L, Fisher RP (2009) Protecting eyewitness evidence: examining the efficacy of a self-administered interview tool. Law Hum Behav 33(4):298–307

Godden DR, Baddeley AD (1975) Context-dependent memory in two natural environments: on land and underwater. Br J Psychol 66(3):325–331

Hope L, Gabbert F, Fisher RP (2011) From laboratory to the street: capturing witness memory using the self-administered interview. Leg Criminol Psychol 16:211–226

Kassin SM, Appleby SC, Perillo JT (2010) Interviewing suspects: practice, science, and future directions. Leg Criminol Psychol 15(1):39–55

Kebbell MR, Milne R, Wagstaff GF (1999) The cognitive interview: a survey of its forensic effectiveness. Psychol Crime Law 5(1–2):101–115

Köhnken G (1995) Interviewing adults. In: Bull R, Carson D (eds) Handbook of psychology in legal contexts, (pp 216–233), Chichester, UK, Wiley

Lamb ME, Orbach Y, Sterberg KJ, Hershkowitz I, Horowitz D (2000) Accuracy of investigators' verbatim notes of their forensic interviews with alleged child abuse victims. Law and Human Behavior 24(6):699–708

Lamb ME, Orbach Y, Hershkowitz I, Esplin PW, Horowitz D (2007) A structured forensic interview protocol improves the quality and informativeness of investigative interviews with children: a review of research using the NICHD Investigative Interview Protocol. Child Abuse Negl 31(11–12):1201–1231

Loftus EF, Miller DG, Burns HJ (1978) Semantic integration of verbal information into a visual memory. J Exp Psychol Hum Learn Mem 4(1):19

Memon A, Meissner CA, Fraser J (2010) The cognitive interview: a meta-analytic review and study space analysis of the past 25 years. Psychol Publ Policy Law 16(4):340–372

Milne R (1999) Interviewing children with learning disabilities. In: Memon A, Bull R (eds) Handbook of the psychology of interviewing, pp 165–180, Chichester, UK, Wiley

Milne B, Bull R (eds) (1999) Investigative interviewing psychology and practice. Wiley, Chichester

Milne R, Bull R (2002) Back to basics: a componential analysis of the original cognitive interview mnemonics with three age groups. Appl Cogn Psychol 16(7):743–753

Neisser U, Harsch N (1992) Phantom flashbulbs: false recollections of hearing the news about challenger. In: Winograd E, Neisser U (eds) Affect and accuracy in recall: studies of "flashbulb" memories, vol 4, pp 9–13, 9–31, London, Cambridge University Press

O'Mahony BM, Milne B, Grant T (2012). To challenge, or not to challenge? Best practice when interviewing

Oxburgh GE, Dando CJ (2011) Psychology and interviewing: what direction now in our quest for reliable information? Br J Forensic Pract 13(2):135–144

Reisberg D, Hertel P (2004) Memory and emotion. Oxford University Press, New York

Simons DJ, Chabris CF (1999) Gorillas in our midst: sustained inattentional blindness for dynamic events. Percep Lond 28:1059–1074

Simons DJ, Rensink RA (2005) Change blindness: past, present, and future. Trends in Cogn Sci 9(1):16–20

Tuckey MR, Brewer N (2003) The influence of schemas, stimulus ambiguity, and interview schedule on eyewitness memory over time. J Exp Psychol Appl 9(2):101

Tulving E (1995) Organization of memory: quo vadis. In: Gazzaniga MS (ed.), The cognitive neurosciences, (pp 839–850), Cambridge, MA: Bradford, MIT Press

Tulving E, Thomson DM (1973) Encoding specificity and retrieval processes in episodic memory. Psychol Rev 80(5):352

Wells GL, Bradfield AL (1998) Good, you identified the suspect: feedback to eyewitnesses distorts their reports of the witnessing experience. J Appl Psychol 83(3):360

Investigative Interviewing

▶ Evaluating Truthfulness: Interviewing and Credibility Assessment

Investigative Psychology

Donna Youngs and Elizabeth Spruin
International Research Centre for Investigative Psychology (IRCIP), University of Huddersfield, Huddersfield, UK

Synonyms

Criminal investigation; IRCIP; Offender profiling

Overview

Investigative Psychology (IP) is a subdiscipline of psychology developed by David Canter for the integration of a diverse range of aspects of psychology into all areas of criminal and civil investigation and legal processing (Canter 1995b, 2011; Canter and Youngs 2009). It considers all forms of offending action from arson, stalking, and robbery to murder, rape, or terrorism, exploring the psychological processes involved in the detailed actions and the influences on and characteristics of perpetrators. The starting point for such studies is the operational challenges that arise during investigation or in court, but the models and theories that have emerged inform the general psychological understanding of offending action and offenders.

The discipline may be understood in terms of the three key strands of academic activity which it encapsulates: (1) the modelling of criminal action and perpetrator inferences (the focus here is on the differentiation of patterns of criminal action and the determination of the psychological processes underlying these differences that have their origins in individual differences and experiential or lifestyle characteristics of the offender (see Canter and Youngs 2009)), (2) the investigative processes (this includes the study of investigative interviewing techniques as well as the decision-making processes of investigators and juries; although less developed, it extends to considerations of effective legal questioning techniques in court), and (3) the assessment and improvement of investigative information/legal evidence (this includes the study of eyewitness testimony, as well as addressing questions of veracity whether through the study of false allegations, false appeals, false confessions, or forensic psycholinguistic studies of authorship attribution).

Although the discipline is based upon a distinct philosophical approach to human beings and a particular meta-methodological approach to studying them, the discipline may also be readily understood in terms of the specific and concrete operational questions it seeks to answer (See Table 1: Key Concepts and Terms).

Moving Beyond Offender Profiling to Investigative Psychology

A major emergence of IP was the drawing of outline pictures of offenders, which subsequently became known as “offender profiles.” This activity of giving informed advice to criminal investigations can be traced back to biblical times; this advice grew and developed over the centuries, as illustrated by early description of the characteristics of witches and subsequent proposals of the classification of people on the basis of their physical characteristics and the inferring of criminality from certain subsets of those characteristics. Even though these early attempts at identifying the indicators of deviance were fundamentally flawed, they did point in the direction from which an Investigative Psychology could eventually emerge (see Canter and Youngs 2009). The contributions that IP can make to police investigations are most widely known and typically understood in terms of “offender profiles.” Offender profiling is held to be the process by which individuals, drawing on their clinical or

Investigative Psychology, Table 1 Key concepts and terms

Investigators' questions	Informed by psychologists' understanding of:
What type of crime is this?	The differentiation of criminal action
What are the likely characteristics of the sort of individual who might commit a crime such as this?	Investigative inferences
Which of the possible suspects is most likely to have committed the crime?	Investigative inferences
Where is the offender most likely to live in relation to the crime?	Geographical offending profiling
Which other crimes are likely to have been committed by the same perpetrator(s)?	The consistency of criminal action
Are the decisions made during the investigation free from distortion and bias?	Decision-making under stress/cognitive load
Have the lines of enquiry pursued been determined systematically?	Decision support tolls
Can we get the witness to remember more?	How to interview a witness
Can we tell if a suspect is lying?	The detection of deception
Can we tell who wrote something?	Forensic psycholinguistics
Did this crime really happen?	False allegations
Did this person really do the crime they say they did?	False confessions

other professional experience, make judgments about the personality traits or psychodynamics of the perpetrators of crimes. Recently, Canter has pointed out that the idea of a distinct "profiling" process is a fallacy.

Certainly, from the perspective of scientific psychology, any such "process" would be flawed in its reliance on clinical judgment rather than actuarial assessment. The clinically derived theories upon which much "offender profiling" has relied are equally questioned by research psychologists (Canter 1995, 2011). Canter modelled the relationship between the offence (A)ctions and offender (C)haracteristics, as canonical equations between particular sets of A and C variables.

The lack of scientific rigor evident in the profiling process has, for two decades, driven proponents of IP research to map out the scientific discipline that could underpin and systematize contributions to investigations. Interestingly, this more academically grounded approach is opening up the potential applications of psychology beyond those areas in which "profiling" first saw the light of day, rather than moving away from operational concerns. Early "profilers" insisted that their skills were only relevant to bizarre crimes in which some form of psychopathology was evident, notably serial killing and serial rape, but investigative psychologists now study and contribute to investigations across the full spectrum of illegal activities.

Nevertheless, as outlined, through the development of a scientific psychological approach to the study of criminal actions and detection processes, a much more challenging set of questions emerge than is apparent in fictional accounts of "profilers." These require consideration of what is actually consistent for an offender from one situation to another. There is also the challenge of dealing with how a criminal's actions may change over time due to many different processes of maturation and learning, as well as the variations introduced by any particular situation.

Fundamentals of Investigative Psychology

The Modelling of Criminal Action and Perpetrator Inferences

The potential links between the actions associated with a crime and the characteristics of the person who committed that offence are informed by a number of questions that are implicit within the concept of "profiling" and associated activities (Canter 1993; Canter and Youngs 2009).

Salience

In order to generate hypotheses and investigative possibilities and to select from them, detectives and other investigators must draw on some understanding of the actions of the offenders involved in the offence they are investigating. They must have some idea or template of typical ways in which offenders behave that will enable them to make sense of the information obtained. A central research question, then, is to identify the behaviorally important facets of offences – those facets that are of most use in revealing the salient psychological processes inherent in any given crime.

So many different aspects of a crime can be considered when attempting to formulate views about that crime that there is the challenge, before any scientific arguments can be developed, of determining which of those features are behaviorally important, in the sense of carrying information on which reliable empirical findings can be built. Accordingly, determining salience turns out to be a much more complex issue than is often realized with parallels in many areas of information retrieval. It requires the determination of base rates and co-occurrences of behaviors as well as an understanding of the patterns of actions that are typical for any given type of crime.

Consistency

One aspect of these salient features that also needs to be determined as part of scientific development is whether they are consistent enough from one context, or crime, to another, such that they can form the basis for considering and interpreting those crimes and accompany them with other offences. The establishment of consistency is not straightforward. There are weaknesses in the sources of data, particularly "patchiness," whereby in some instances certain features – or their absence – will be recorded and in others not. To deal with these problems with the data, there is an emerging range of methodologies that focus on working with the sort of information that is available from police and other crime-related sources.

In general, there are two rather different aspects of criminal consistency; one is the consistency from the actions in a crime to other aspects of the offender's life. This may thus allow some aspects of a one-off crime, such as the expertise involved, to be used to form views about an offender's previous training. So, the question here is which features of an offence are consistent with which, if any, aspects of an individual's background and personal characteristics.

A quite distinctly different form of consistency is that between one crime and another committed by the same person. This is often, rather loosely, referred to as an offender's modus operandi, or M.O. The question here then will be which features of an individual's offending behavior are consistent across situations, separate offences, and even across different types of offences. There are many aspects of this latter form of consistency that need to be clarified; for the moment we just need to be aware of its relevance at the core of IP.

A more conceptual challenge to determining consistency, as in all human activity, is that some variation and change is a natural aspect of human processes. There therefore will be criminals who are consistently variable or whose behavioral trajectories demonstrate some form of career development, as well as those whose criminal behavior will remain relatively stable over time. Research around all these possibilities of consistency is therefore central to any development of a scientific basis for offender profiling.

Development and Change

Any consideration of consistency will be against a backdrop of likely change in a serial offender's activities. Some variation is a natural aspect of human processes. Accordingly, there will be criminals who are consistently variable or whose behavioral trajectories demonstrate some form of career development, as well as those whose criminal behavior will remain relatively stable over time. The question of what is characteristic of an offender is one aspect of the operational problem of linking crimes to a common offender in circumstances where there is no forensic evidence to link the case. To date there has been remarkably little research on linking crimes. The challenge to police investigations can be seen as the reverse of the personality question. Psychologists usually have a person and want to know what will be consistent about that person from one situation to another. Police investigators have a variety of situations, criminal events, and need to know what consistencies can be drawn out of those events to point to a common offender.

However, if the basis of these changes can be understood then they can be used to enhance the inference process. The following five relevant forms of change are identified and need to be considered when establishing actions to crime scenes and their relationship to offender characteristics.

Responsiveness. A criminal's actions may not be the same on two different occasions because of the different circumstances he/she faces. An understanding of these circumstances and how the offender has responded to them may allow some inferences about his/her interpersonal style or situational responsiveness to be made that have implications for the conduct of the investigation.

Maturation. Maturation is an essentially biological process of change in a person's physiology with age. Knowledge of what is typical of people at certain ages, such as sexual activity or physical agility, can thus be used to estimate the maturity of the person committing the crimes and to explain the possible basis for longer term variations in an individual's criminal activity.

Development. The unfolding psychological mechanisms that come with age provide a basis for change in cognitive and emotional processes. One reflection of this is an increase in expertise in doing a particular task. Evidence of such expertise in a crime can thus be used to help make inferences about the stages in a criminal's development that he/she has reached and indeed to indicate the way their crimes might change in the future.

Learning. Most offenders will learn from their earlier offending in the same way that learning theorists have shown that behavior generally is shaped by experience. So, for example, an offender who struggled to control his first victim may be expected to implement some very definite restraining measures during subsequent offences. Indeed, for offenders, the particularly salient, potentially negative consequences of their actions (e.g., prison) may make this a powerful process for change in the criminal context. An inferential implication of this is that it may be possible to link crimes to a common offender by understanding the logic of how behavior has changed from one offence to the next.

Careers. The most general form of change that may be expected from criminals is one that may be seen as having an analogy to a legitimate career. This would imply stages such as apprenticeship, middle management, leadership, and retirement. Unfortunately, the criminology literature often uses the term "criminal career" simply to mean the sequence of crimes a person has committed. It is also sometimes confused with the idea of a "career criminal," someone who makes a living entirely out of crime. As a consequence, much less is understood about the utility of the career analogy for criminals than might be expected. There are some indications that the more serious crimes are committed by people who have a history of less serious crimes and that as a consequence, the more serious a crime, the older an offender is likely to be. But a commonly held assumption, such that serious sexual offences are presaged by less serious ones, does not have a lot of empirical evidence in its support.

Cultural Shifts. There are also important changes that come about from the developments in society. Increased security precautions will lead offenders to change their behavior.

Differentiation

Although an offender's consistency is one of the starting points for empirically based models of investigative inference, in order to use these models operationally, it is also necessary to have some indication of how offenders can be distinguished from each other (Canter 2000). In part, this reflects a debate within criminology about whether offenders are typically specialist or versatile in their patterns of offending. Research suggests that offenders may share many aspects of their criminal styles with most other criminals, but there will be other aspects that are more characteristic of them. It is these rarer, discriminating features that may provide a productive basis for distinguishing between offences and offenders. Canter also posited a conceptual structure known as the Radex (see Canter 2000; Canter and Youngs 2009), as the basis for modelling these patterns of variation and similarity in offending action. As Canter and Youngs (2009) note then, the Radex of Criminal Differentiation hypothesizes that the patterns within criminal action will be organized according to a combination of quantitative and qualitative themes. Numerous empirical studies have supported the Radex hypothesis, leading to recent explorations of the theoretical and conceptual import of this structure.

The rarer features combine to create that which may be regarded as a "style" or "theme" of offending. Therefore, the production of a framework for classifying these different styles is a cornerstone of IP. The classification frameworks are ways of organizing the themes that distinguish crimes which can then be drawn on in inference models so that, rather than being concerned with particular, individual clues as would be typical of detective fiction, the IP approach recognizes that any one criminal action may be unreliably recorded or may not happen because of situational factors. But a group of actions that together indicate some dominant aspects of the offender's style may be strongly related to some important characteristic.

Recent work has identified a single, generic theoretical framework for understanding the different styles of offending that can be identified across the range of offending forms from burglary or robbery and sexual assault to serial killing or stalking. This is the Narrative Action System model (Canter and Youngs 2009) with its corollary Victim Role Assignments model (Canter 1994; Canter and Youngs 2012). The Narrative Action System identifies four generic styles of offending that reflect the focus and origins of the action (termed an action system mode; see Canter and Youngs 2009) combined with the underpinning criminal narrative : the Hero's Expressive Quest, the Professional's Adaptive Adventure, the Revenger's Conservative Tragedy, and the Victim's Integrative Irony. The Victim Role model emphasizes the interpersonal components of these generic offending styles, distinguishing the offender's interpersonal approach to the victim (or "role assignment") on the basis of different forms of empathy-deficit and control strategies (Canter 1994; Canter and Youngs 2012).

Inferences

As outlined, the heart of these questions is what has become known as the "profiling equations" (Canter 1993, 2011). These are equations that provide the scientific bases for inferring associations between the actions that occur during the offence – including when and where they happen and to whom – and the characteristics of the offender, including the offender's criminal history, background, base location, and relationships to others. They are also referred to as the "A → C equations," where A are the actions related to the crime, and C are the characteristics of typical offenders for such crimes, and → is the argument and evidence for inferring one from the other.

Although this relationship is not an "equation" in strict mathematical sense, it is helpful to keep the looser meaning implied by the simple formulation. Thinking of the relationships as a family of equations simplifies the discussion of what the central questions of IP are. It also helps to clarify the many challenges and complexities that arise in trying to establish empirically any relationships implied by these equations. It is important to emphasize, though, that specifying the problem as a set of equations does not mean that these are a purely mathematical or statistical way of carrying out profiling. Rather, by studying what would be involved in solving these equations, it becomes clear that we have to use more conceptual, theory-driven approaches if we are to have any hope of producing a scientific basis for psychological contributions to investigations.

A starting point for understanding the challenges inherent in the A → C relationships is to recognize that the A → C mapping will rarely take the form of simple one to one relationships, rather there are a range of complexities in the ways that As relate to Cs. These complexities can be thought of as "canonical equations," which are "the relationships between two sets of variables." There is always a mixture of criminal actions that are being related to a mixture of characteristics or other criminal actions. The relationships are between combinations of action variables and combinations of characteristic variables. Therefore, the whole concept of a "canonical equation" shows that small changes in any one variable can influence the overall outcome.

Investigative Psychologists have conducted – and continue to conduct – a wide range of empirical studies of different types of crime, with the purpose of establishing solutions to these equations, in the hope of providing objective bases for investigative inferences. The complexity of the A-C relationship does however require the development of theoretical models to explain the relationships that may be expected for different combinations of variables, in different contexts.

Psychological Processes Underlying Investigative Inferences

Once there is some information to work with, there is the significant challenge of determining what conclusions can be drawn from this information. These conclusions are inferences that claim that certain features are linked. The inferences that detectives make in an investigation about the perpetrator's likely characteristics will be valid to the extent that they are based on appropriate ideas about the processes by which the actions in a crime correlate with the characteristics. Investigative Psychologists have advanced five possible processes that may be drawn upon to develop these inferences: (a) Personality theory, (b) Psychodynamic theory, (c) the career route perspective from criminological theory, (d) social processes, and (e) Interpersonal Narrative theory. Any or all of these theories could provide a valid basis for investigative inferences if the differences in individuals that they possess correspond to real variations in criminal behavior.

One general hypothesis is that offenders will show some consistency between the nature of their crimes and characteristics they exhibit in other situations. This is rather different from psychological models that attempt to explain criminality as a displacement or compensatory activity, resulting from psychological deficiencies. The evidence so far is consistent with this general consistency model, suggesting that processes relating to both the offender's characteristic interpersonal style and his/her routine activities may be particularly useful in linking actions and characteristics. Valid inferences also depend upon an understanding of the way in which a process operates. Conceptually there are a number of different models that can be drawn on to link an offender's actions with his/her characteristics. One is to explain how it is that the offender's characteristics are the cause of the particular criminal actions. For example, if a man is known to be violent when frustrated, this knowledge provides a basis for inferring his characteristics from his actions. A different theoretical perspective would be to look for variables that were characteristic of the offender and that would influence the particular offending actions. A highly intelligent person, for instance, may be expected to commit a fraud rather differently from someone with educational difficulties. The intelligence may be reflected in the style of action even if not in the actual cause of the action. A third possibility is that actions give rise to some consequences from which characteristics may be inferred. An example of this would be when particular types of goods are stolen that imply that the thief must have contact with other offenders who would buy or distribute those goods.

Other approaches to inference are less based on assumptions about the enduring characteristics of the offender and more on their interpersonal and social context. In terms of "criminal career," this is the idea that because criminals develop and learn as they commit more and more crimes, their most recent crimes may tell us more about the crimes they have committed in the past than about their enduring characteristics. The need to take on board consistencies in offenders' characteristics as well as the developments and changes of their lives has been articulated in personal narrative models. This approach also gives more emphasis on the sense the offender makes of his/her life and their agency in acting out their life story as they see it unfolding. This perspective has had particular power in assisting the identification of themes within criminal activities.

From an applied perspective, it is also important that the variables on which the inference models draw are limited to those of utility to police investigations. This implies that the A variables are restricted to those known prior to any suspect being identified, typically crime scene information and/or victim and witness statements. The C variables are limited to those on which the police can act, such as information about where the person might be living, his/her criminal history, age, or domestic circumstances.

These inference models operate at the thematic level, rather than being concerned with particular, individual clues as would be typical of detective fiction. This approach recognizes that any one criminal action may be unreliably recorded or may not happen because of situational factors. But a group of actions that together indicate some dominant aspect of the offender's style may be strongly related to some important characteristic of the offender.

Investigative Processes: Investigation as Decision-Making

The decision-making phase of the investigative cycle is to select from the various options that are revealed through the earlier stages and act on those options, either to seek further information or to prepare some action that will lead to an arrest or proceedings in court. More specifically, the decision-making tasks that constitute the investigation process can be derived from consideration of the sequence of activities that detectives follow, starting from the point at which a crime is committed through to the bringing of a case to court. As they progress through this sequence of activities, detectives reach choice points, at which they must identify the possibilities for action on the basis of the information they can obtain. For example, when a burglary is committed, they may seek to match fingerprints found at the crime scene with known suspects. This is a relatively straightforward process of making inferences about the likely culprit from the information drawn from the fingerprints. The action of arresting and questioning the suspect follows from this inference.

However, in many cases, the investigative process is not so straightforward. Detectives may not have such clear-cut information, for example, if it is suspected that the style of the burglary is typical of one of a number of people they have arrested in the past. Or, in an even more complex example, they may infer from the disorder at a murder crime scene that the offender was a burglar disturbed in the act. These inferences will either lead them to seek other information or to select from a possible range of actions, including arresting and charging a likely suspect.

Investigative decision-making thus involves the identification and selection of options, such as possible suspects or lines of enquiry that will lead to the eventual narrowing down of the search process. Throughout this process, detectives must gather the appropriate evidence to identify the perpetrator and prove their case in court. A clear understanding of the investigation process as a series of decision-making tasks allows the challenges implicit in this process to be readily and appropriately identified. The main challenge to investigators is to make important decisions under considerable pressure and in circumstances that are often ambiguous. The events surrounding the decisions are likely to carry a great emotional charge, and there may be other political and organizational stresses that also make objective judgements very difficult. A lot of information, much of which may be of unknown reliability, needs to be amassed and

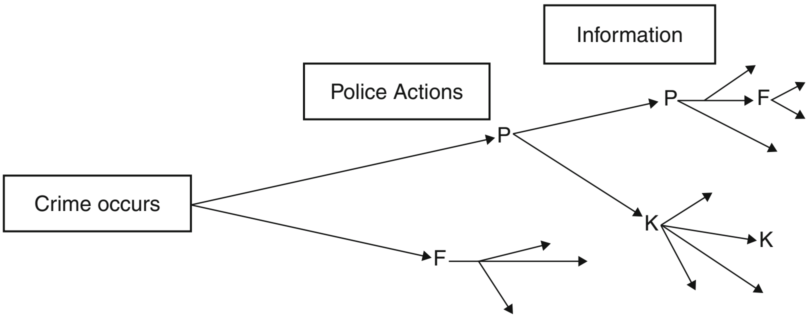

Investigative Psychology, Fig. 1 Schematic indication of the rapidly developing information in a major police investigation

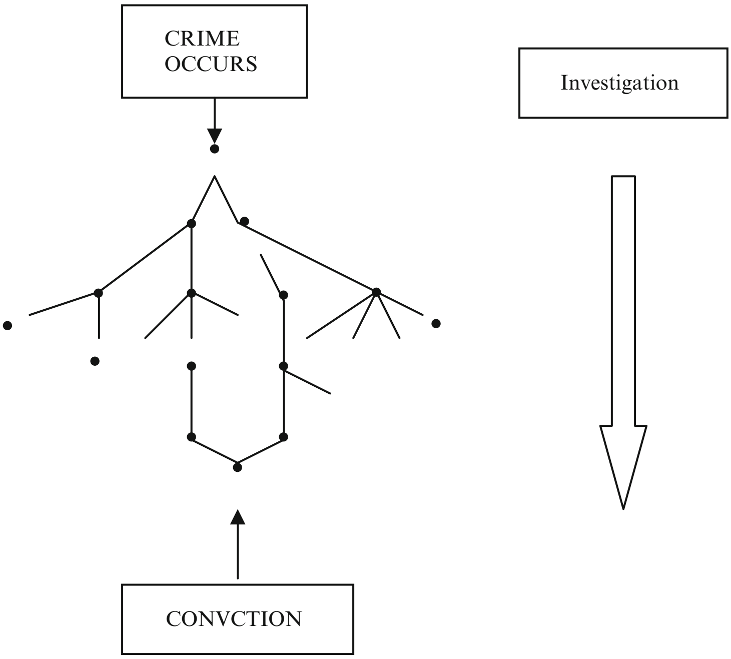

Investigative Psychology, Fig. 2 Police investigation indicating the convergence on a suspect and conviction

digested. In decision-making terms, the investigative process can be represented as in Fig. 1.

In this diagram the lines represent investigative actions by the police, while the nodes are the results of those actions, i.e., new pieces of information or facts. Immediately after a crime occurs, detectives often have few leads to follow up. However, as they begin to investigate, information comes to light opening up lines of enquiry. These produce more information, suggesting further directions for investigative action. The rapid buildup of information in these first few days will often give rise to exponential increases in the cognitive load on detectives, reaching a maximum weight after some short period of time. At this point, investigators will often be under considerable stress.

As the investigation progresses, detectives will eventually be able to start to narrow down their lines of enquiry by establishing facts that close off all but one of them, substantially reducing the general level of demand. The general diamond shape in Fig. 2 represents the typical progression of an investigation. The diagram depicts the initial buildup and then the subsequent narrowing down of the information (the nodes) under consideration and investigative steps related to these (the lines) as the investigation becomes increasingly focused to the point where an arrest occurs.

The Assessment and Improvement of Investigative Information/Legal Evidence

Decisions clearly rely upon information, which is often carved out of real-world events. In a typical police investigation, a mass of information accumulates. The sheer volume of information can reduce the effectiveness of an investigation, as happened in the Yorkshire Ripper enquiry in the UK or more recently in the inquiry into the murder of the English television journalist Jill Dando.

On top of the demands of the amount of information that the police have to deal with is the great variety of material. Each different type of material has to be managed and examined in different ways, requiring different skills and different assessment processes. This information comes from a variety of different sources. There may be photographs or other recordings of the crime scene. There may be records of other transactions such as bills or telephone calls. Increasingly there are records available within computer systems used by witnesses, victims, or suspects. Often there will be witnesses to the crime or results of the crime will be available for examination. Further, there will be information in police and other records that may be drawn upon to provide indications for action. Once suspects are identified, there is further potential information about them, either directly from interviews with them or indirectly through reports from others. In addition, there may be information from various experts that has to be understood and that may lead to specific actions. Apart from being potentially overwhelming in terms of its sheer quantity, the information takes many different forms. It is slanted in many different ways, having been either originally generated for other purposes, or distorted by human perception and memory processes, or, indeed, particular motives. Moreover, while the information available does have certain strengths (such as the fact that it may have been given under oath), it has not been collected with the careful controls of laboratory research. The information on which police rely on in an investigation is therefore often incomplete, ambiguous, and unreliable.

Furthermore, because the information is carved out of real-world events, it must be treated with great respect and considerable caution within IP. A continuum of trustworthiness can be regarded as running from the most trustworthy evidence (the hard evidence at the crime scene) through to the least (the statement of a culprit who denies guilt). The hurdles in the way of trustworthy evidence are present at every stage of an investigation, from initial witness statements, through comments from suspects, to how evidence is presented in court. The challenges this poses for investigations have a different emphasis from those that are posed for research. The biggest problems for research are the biases in the sampling and the gaps and lack of detail in the data. So the researcher always has to work from a record made of the events by some third party.

It is by conceptualizing and treating this information as “data,” and the ways in which it is obtained as research processes, that psychologists can make a further broad class of contribution to investigative activity. Understanding it in this way allows us to use psychological principles and knowledge to evaluate and improve the information detectives need to progress an investigation or to back up a case in court. Two aspects of scientific data assessment are particularly relevant to investigative information: its usefulness and detail and its accuracy and validity.

There are a variety of contributions psychologists can make to the improvement of the accuracy and validity of the information available in an investigation. A number of formal validity assessment techniques have been developed to assess the truthfulness of witness accounts when no objective means of doing this are available. Most of these techniques are based on the assumption that honest accounts have identifiable characteristics that are different from fabricated accounts.

There is much more evidence to indicate that for many people, there are psychophysiological responses that may be indicators of false statements. The procedure for examining these responses is often referred to as a polygraph or "lie detector." In essence, this procedure records changes in the autonomic arousal system, i.e., emotional response. Such responses occur whenever a person perceives an emotionally significant stimulus. The most well-established indicator is when the respondent is asked to consider information that only the perpetrator would be aware of, known as the "guilty knowledge" test. A more controversial procedure is to ask "control questions" that many people would find emotionally significant, in order to determine if they elicit responses that can be distinguished from those questions relating directly to the crime. However, in both these applications of psychophysiological measures, the most important element is the very careful interview procedure before measurements are made and during the process.

In general "lie detection" is more productive in supporting a claim of innocence than in providing proof of guilt. For this reason, many jurisdictions do not allow "lie detector" results to be presented as evidence in court. Its value in eliminating possible suspects is used in a variety of jurisdictions around the world. In some investigations, the suspect may deny that he/she made a statement that has been attributed to him/her. In such cases, techniques based on the quantitative examination of the language used may be drawn upon to evaluate the claim. So, for instance, a forensic linguist may try to establish whether the use of particular nouns is typical of the suspect or not. Interestingly, however, indications are that the psychological components of written or verbal accounts, i.e., what is meant and how it is expressed, may be more useful in attributing authorship than the linguistic features, such as counts of particular words.

Sometimes the concern will be not with the veracity of the suspect's account but with that of an alleged victim. This can be an issue in cases of sexual or other abuse. In such cases, the complainant is not a suspect, and the more intrusive processes of lie detection are rarely used. However, investigators can draw on studies of the circumstances under which such false allegations are made and use those as guidelines for more intensive examination of the circumstances. Whether or not this is a valid way of identifying false allegations is a topic awaiting further research.

Investigative Psychology in Action

There has been a long journey from trying to profile the characteristics of a witch through the attempts to fathom the criminal mind of the late nineteenth and twentieth centuries. The experienced detectives, notably at the FBI training academy in Quantico, who drew attention to the possibility of working directly with what occurred in a crime to draw systematic inferences about offenders, pointed the way to a more empirical, scientific basis for psychological and related behavioral science contributions to police investigations. The blossoming of applied psychology into many areas of professional activity laid the foundations for an Investigative Psychology that goes far beyond the initial speculations about serial killers of 50 years ago.

It has been demonstrated throughout this entry that IP is relevant to the full range of offence activity. Psychology and related disciplines need no longer be assigned to the "hit-and-run" offering of an "offender profile," which is only relevant to bizarre, extreme crimes. It is now taking its place as of central importance to many aspects of police training and beyond. The groundbreaking work of IP has continuously emphasized the value of systematizing and presenting visualizations of investigative information and the development of decision support systems than can be an integrated aspect of police activities. Two areas in particular are proving fruitful for such decision support: geographical offender profiling and the linking of cases to a common offender. Both of these will benefit from an increasing amalgamation of spatiotemporal and behavioral information as has been attempted in recent investigative decision support systems.

Related Entries

- ▶ Applied Geographical Profiling
- ▶ Behavioral Investigative Advice
- ▶ Criminal Profiling
- ▶ FBI Influence on State and Local Police
- ▶ Interview and Interrogation Methods Effects on Confession Accuracy
- ▶ Modus Operandi of Sex Offenders
- ▶ Profiling Arson
- ▶ So What Criminology?

Recommended Reading and References

Canter D (1993) The wholistic organic researcher: Central issues in clinical research methodology, In: Powell G, Young R & Frosh S (eds) Curriculum in clinical psychology, Leicester: The wholistic organic researcher: Central issues in clinical research methodology. BPS pp 40–56

Canter D (1994) Criminal Shadows. London: HarperCollins

Canter DV (1995a) Criminal shadows. HarperCollins, London

Canter D (1995b) Psychology of offender profiling. In: Bull R, Carson D (eds) Handbook of psychology in legal contests. Wiley, Chichester, pp 343–335, Chapter 4.5

Canter D (2000) Offender profiling and criminal differentiation. Leg Criminol Psychol 5:23–46

Canter DV (2003) Mapping murder: the secrets of geographical profiling. Virgin, London

Canter DV (2011) Resolving the offender "profiling equations" and the emergence of investigative psychology. Curr Dir Psychol Sci 20:5–10

Canter D (2012) Forensic psychology for dummies. Wiley, New York

Canter DV, Alison LJ (eds) (1999a) Profiling in policy and practice, networks. Offender profiling series. Aldershot, Dartmouth, pp 157–188, Vol II

Canter DV, Alison LJ (eds) (1999b) The social psychology of crime: teams, groups, networks. Offender profiling series. Aldershot, Dartmouth, Vol III

Canter D, Fritzon K (1998) Differentiating arsonists: a model of firesetting actions and characteristics legal and criminal psychology, 3, 73–96

Canter D, Youngs D (2008a) Principles of geographical offender profiling. Ashgate, Aldershot

Canter D, Youngs D (2008b) Applications of geographical offender profiling. Ashgate, Aldershot

Canter D, Youngs D (2009) Investigative psychology: offender profiling and the analysis of criminal action. London: John Wiley & Sons

Canter D, Youngs D (in press) Sexual and violent offenders' role assignments: a generic model of offending style. J Forensic Psychiatr Psychol

Canter D, Bennell C, Alison L, Reddy S (2003) Differentiating sex offences: a behaviourally based thematic classification of stranger rapes. Behav Sci Law 21:157–174

Salfati G, Canter D (1999) Differentiating stranger murders: profiling offender characteristics from behavioural styles. Behav Sci Law 17:391–406

Youngs D (2004) Personality correlates of offence style. J Investig Psychol Offender Profiling 1:99–119

Youngs D The behavioural analysis of crime: studies in David Canter's Investigative Psychology (in press). Ashgate, Aldershot

IRCIP

▶ Investigative Psychology

Italian Mafia

Giulia Berlusconi
Università Cattolica del Sacro Cuore and
Transcrime, Milan, Italy

Synonyms

Italian organized crime

Overview

Different ways of conceptualizing the mafia have been proposed over the time, emphasizing either the cultural or the organizational factors. The former describes the mafia as a subculture typical of certain areas of Southern Italy, whereas the latter focuses on its internal configuration. The mafia has also been conceptualized as a business enterprise, although its political dimension has also been emphasized by delineating it as an industry of private protection. Indeed, Italian mafias differ from other criminal organizations for their will to political dominion in the areas of their traditional presence. Another peculiarity of mafia groups in Italy is their internal configuration and the resort to symbols and rituals that promote the creation of strong ties among the affiliates.

Italian mafia organizations include Cosa Nostra (Our Thing) in Sicily, the 'Ndrangheta (Society of the Men of Honor) in Calabria, and the Camorra in Campania. Although these criminal organizations share common elements, they differ in terms of history, internal organization, illegal market activities, and infiltration in the legitimate and informal economy, and cannot thus be considered as a single phenomenon. The historical and cultural background, the internal organization, and the expertise provide an explanation for the differences that emerge with regard to the presence of mafia groups in illegal markets and the way they conduct their criminal activities. Extortion is instead common to all Italian mafia organizations, and it constitutes the core business of Italian mafias, since it allows them to gain control over a specific territory and thus exercise political power in that area.

The large profits coming from illegal activities are invested both in the illegal markets and in the legitimate economy. Investments in the legitimate economy allow mafia organizations to extend their political dominion by infiltrating into and controlling legitimate economic sectors. Among the economic sectors, mafia groups seem to favor the public construction industry, as well as other "protected" sectors where the public administration regulates the participation by requiring a specific license or permission and thus reduces the competition.

To this day, most research on Italian mafias focused on areas of their traditional presence and explored the reasons they emerged in those territories, whereas little is known about the colonization and transplantation processes. In the future, the adoption of more complex interpretations of the mobility of criminal groups might help identify the factors that foster transplantation and that make it successful.

Mafia and Organized Crime

For a long time, the term "mafia" has been used as a synonym for organized crime (Blok 2008), and Italian mafia has been considered a paradigm of organized crime (Paoli 2003). Nonetheless, the term "mafia" refers to a specific form of organized crime (Finckenauer 2005), and the illegal organizations that have been presented as its archetype include large-scale criminal groups other than Italian mafias, such as the Chinese Triads and the Japanese Yakuza (Paoli 2002).

The main trait that distinguishes mafia groups from other criminal organizations is their political dimension. Although actively involved in illegal market activities, mafia groups cannot be reduced to these activities (Paoli 2002) nor is their internal configuration the consequence of their participation in illegal businesses (van Duyne 1997). Besides the provision of illegal goods and services and the maximization of profits, mafia groups aim at exercising political dominion over the areas where their members live and where most of their illegal activities are conducted. The control over the territory is guaranteed by a generalized system of extortion, which also allows mafia organizations to increase their income (Paoli 2003).

Another peculiarity of mafia groups is their internal configuration. Despite variations, mafia organizations are characterized by the presence of government bodies that coordinate the activities and regulate the relationships among their member. Symbols and rituals, including a ceremony of affiliation, promote the creation of ties of brotherhood among the members of the organization, who become "men of honor" and are subject to the duty of silence (Paoli 2003).

Italian Mafias

The first comprehensive study of the mafia was published in 1876 by Leopoldo Franchetti (1974), who described it as an extralegal social system, namely, a specific and socially accepted way of exercising violence to solve conflicts. After Franchetti's work, different ways of conceptualizing the mafia have been proposed over the time, emphasizing either the cultural or the organizational factors (Sciarrone 2009). The cultural approach describes the mafia as an attitude, a subculture typical of certain areas of Southern Italy, rather than a formal organization (Hess 1973), although organizational factors have also been considered by analyzing the network of family, friendship, and business relations of mafia members (Blok 1988; Schneider and Schneider 1976). The organizational approach instead emphasizes the internal configuration of mafia groups and describes them as formal organizations with articulated structures and a set of symbols and rituals. The depiction of mafia organizations as criminal associations can also be found in the Italian penal code, where Article 416bis describes the mafia as a specific type of criminal organization and lists the main features of mafia groups.

In the mid-1980s, the focus shifted from cultural and organizational aspects to the illegal activities carried out by criminal organizations, and the mafia was conceptualized as an enterprise. According to Pino Arlacchi (1988), between the 1950s and the 1960s, the Italian mafia had experienced an entrepreneurial transformation and had focused on the accumulation of capital. Other scholars refused the identification of mafia groups with business enterprises and the sharp distinction between a modern and an old mafia. They instead emphasized either the importance of the threat or use of violence to regulate economic transactions (Catanzaro 1992) or the political dimension, describing the mafia as "a specific economic enterprise, an industry which produces, promotes and sells private protection" (Gambetta 1993, 1). More recently, the complexity of the phenomenon has been recognized (Santino 2006), and Italian mafias have been described as "multifunctional organizations" involved in a plurality of illegal activities and aiming at pursuing a plurality of goals, including the exercise of political power over a specific territory (Paoli 2004, p. 274). Mafia groups cannot thus be reduced to their economic dimension, both because they originated before the illegal markets they operate in and because their activities include noneconomic ones, such as the exercise of political dominion (Paoli 2002).

Most of these conceptualizations of the mafia have been proposed with particular reference to the Sicilian mafia, also known as "Cosa Nostra." However, mafia organizations which originated from and mainly operate in Italy include Cosa Nostra (Our Thing) in Sicily, the 'Ndrangheta (Society of the Men of Honor) in Calabria, and the Camorra in Campania.

Although these criminal organizations share common elements, the history of the mafia "cannot be reduced to a single scheme or interpretation, valid in every situation and time" (Lupo 1996, p. 26). Italian mafias differ in terms of history, internal organization, illegal market activities, and infiltration in the legitimate and informal economy (Sciarrone 2009), and they thus need to be discussed separately.

Cosa Nostra

The term "mafia" first appeared in 1865 in a report by the then-prefect of Palermo Filippo Gualterio, who used it to refer to Sicilian criminal groups. Only 20 years later, the same term was adopted in the United States to refer to a criminal organization of Italian origins. The same organization was also referred to as "La Cosa Nostra," as Joseph Valachi testified in 1963 before the Permanent Subcommittee on Investigations of the US Senate Committee on Government Operations. The same term has also been adopted by Sicilian mafiosi, together with other names, such as "Onorata Società" (Honored Society) (Lupo 1996).

Evidence of antecedents of Cosa Nostra in Sicily can be found starting from the second half of the nineteenth century, although only in the 1920s, the Italian government ordered a systematic repression of Sicilian mafia groups. After World War II, Cosa Nostra extended its power, mainly by expanding from the countryside into the largest cities and infiltrating into the public construction industry. In the mid-1970s, the Sicilian mafia entered the illicit drug market further increasing its power and profits, until the beginning of the 1990s. In 1992–1993, the killing of the state prosecutors Giovanni Falcone and Paolo Borsellino was followed by an intensification of law enforcement repression.

As a consequence, Cosa Nostra adopted the so-called "strategia dell'inabissamento" (sinking strategy), aimed at reducing its visibility and thus the attention from law enforcement agencies by minimizing violence. However, this strategy also led mafia groups to concentrate their illegal activities in Sicily, mainly focusing on extortion and infiltration in the legitimate economy (Lupo 1996; Savona 2012).

The internal structure of Cosa Nostra has not changed since the 1950s. It is composed of a number of *cosche* or families. Each *cosca* controls a specific territory, and a boss, who is elected by the affiliates, is in charge of its management. His counselors are also elected by the affiliates and assist the boss in his decisions. The hierarchical structure of the *cosche*, the direct election of the high-ranking positions, and the frequent turnover guarantee the cohesion of the group and help strengthen the ties of brotherhood among mafia members. In 1957 a Commission was also created to regulate the relationships among the different families. The meetings were attended by representatives of the *mandamenti*, namely, groups of three or four families. The Commission had the aim of settling disputes, sanctioning violations and taking any decision related to the use of violence, whereas each *cosca* was still autonomous in doing business (Paoli 2003).

The intensification of the law enforcement repression since the early 1990s and the arrest of Bernardo Provenzano, the latest chief of Cosa Nostra, in 2006 have considerably weakened the Sicilian mafia, which has experienced a strong reorganization process. However, despite the undergoing crisis, the organization can still rely on a widespread network of entrepreneurs and politicians that ensure Cosa Nostra the ability to commit extortion and infiltrate into the Sicilian legitimate economy (DIA 2010).

The 'Ndrangheta

Evidence of the presence of the 'Ndrangheta in Calabria dates back to the end of the nineteenth century, although the term "'ndrangheta" has been consistently used only since the 1960s. Different names instead appear in the first judicial documents, such as "mafia" or "camorra" – which were used to refer to Sicilian and Neapolitan mafia organizations – or "onorata società" (honored society) and "picciotteria" (Paoli 2003). The limited violence and the fact that it was considered outdated in comparison to Cosa Nostra and the Camorra made law enforcement agencies and scholars not to pay too much attention to the 'Ndrangheta (Ciconte 2011). Today the Calabrian mafia is instead considered the most powerful among Italian criminal organizations and the most involved in drug trafficking, and in June 2008, it was included into the US Treasury Department Office of Foreign Asset Control (OFAC) Foreign Drug Kingpin list (U.S. Department of State 2011).

The 'Ndrangheta is organized in *cosche* or *'ndrine* which are usually formed by the members of the same family and exercise their influence over a specific territory according to the arrangements made with the other families. The fact that members of the *'ndrine* are bonded by kin ties guarantees secrecy and trust (Paoli 2003). Within each *'ndrina*, specific tasks are also assigned to the affiliates, according to their formal rank in the organization. A two-level hierarchical structure is present; low-level affiliates are referred to as members of the *società minore* (lower society), whereas bosses are part of the *società maggiore* (higher society). Within both the *società*, a long series of different roles and tasks can then be identified.

The familistic structure of the Calabrian mafia has been considered for a long time a sign of backwardness, compared with the Sicilian mafia and the Neapolitan Camorra. Rather than being old fashioned, this type of structure guarantees a high degree of cohesion among its members and minimizes the risk of defections. The number of defectors has indeed always been lower than that of other Italian mafia groups (CPA 2008; Ciconte 2011). The internal cohesion is also guaranteed by a set of formal rules, an extensive use of rituals and symbols, and strong barriers to entry. As in Cosa Nostra, an affiliation ceremony or *battesimo* is indeed required for a new affiliate to be appointed and become a "man of honor" (Paoli 2004).

For a long time, the ‘Ndrangheta was conceived as a confederation of families without a coordinating body which establishes common rules in order to avoid conflicts among the *cosche*. This idea of the ‘Ndrangheta as an horizontal organization started to be reexamined in the 1990s and has been abandoned when the existence of a collective body (*Crimine* or *Provincia*) entitled to establish common rules and settle disputes emerged from recent criminal investigations (TrMi 2010). However, despite the presence of a collective body, each family enjoys a monopoly on its territory and the autonomous running of illegal activities, even though a collaboration among different *cosche* is frequent, especially for the trafficking of drugs (Ciconte 2011).

The power of the Calabrian mafia also comes from its ramifications, both in other Italian regions and abroad. In the last decades, the ‘Ndrangheta has indeed created outposts both in Central and Northern Italy and in foreign countries, where it has developed an extensive network of contacts to facilitate its involvement in illegal market activities. Kin ties maintain their importance in the internal organization of the *‘ndrine* also outside Calabria, since they allow for a quick replacement of affiliates in case of arrests by law enforcement agencies. However, in addition to mafia associates, ‘Ndrangheta groups resort to other individuals outside the organization and can rely on an extended network of contacts with both common criminals and legal entrepreneurs and politicians (CPA 2008; Ciconte 2011).

The Camorra

The Camorra refers to a variety of independent criminal groups and gangs which conduct their illegal activities in Campania. These criminal groups – unlike Sicilian and Calabrian ones – vary considerably from long-lasting family businesses to gangs developed around charismatic leaders, often in competition among each other (Paoli 2004).

First mentions of the Camorra as a criminal organization can be found in police records which date back to the early nineteenth century, when written statutes describing symbols and rituals, as well as signs of the presence of stable coordination mechanisms, were found (Behan 1996). Although most symbols and rituals have been maintained, the Campanian underworld has changed, and the contemporary Camorra has turned into a different criminal organization. In the 1960s, the Camorra groups entered the tobacco and later the drug market, making large profits which were then invested both in the same illegal markets and in the legitimate economy, dealing mainly with public sector contracts (Paoli 2004). With the development of the contraband industry, competition among rival groups increased, as well as their dynamism and the territorial control (Behan 1996).

The presence of a stable mechanism that coordinates the activities of the different groups and settles disputes among them no longer exists, making wars against rival gangs more likely to occur than in Sicily and Calabria (Paoli 2004; DIA 2010). The Camorra is indeed horizontally organized, with criminal groups forming temporary alliances and maintaining their autonomy (Barbagallo 2010). As for Cosa Nostra and the ‘Ndrangheta, the direct physical control over a territory is fundamental for Camorra groups. “Once this control is won, criminal gangs can not only extend their protection rackets, illegal gambling, usury, cigarette, and drug trading; they can begin to become power brokers, mediating between ordinary people and ruling politicians” (Behan 1996, p. 110).

The heterogeneity of criminal groups and gangs which populate the Campanian underworld is reflected upon the illegal activities they conduct. Smaller groups are mostly involved in the supply of illegal goods and services, such as extortion, loan sharking, drug trafficking, counterfeiting, exploitation of prostitution, and illegal gambling; the most powerful and long-lasting groups combine these illegal activities with infiltration into the legitimate economy and the public administration sector (Paoli 2004; DIA 2010).

Illegal Activities

Extortion and protection constitute the core business of Italian mafias. Although extortion is not as profitable as other illegal activities, such as usury, drug trafficking, or infiltration in the public sector contracts, it does not require any initial investment, and it can be carried out without high risks or high managing costs. Besides being a source of profit, it allows mafia groups to gain control over a specific territory and thus exercise political power in that area, as well as it enables them to establish contacts with legal entrepreneurs and penetrate into the legitimate economy. The sum that legal entrepreneurs are forced to pay is usually calculated on the basis of their legal profits, so to encourage their compliance and minimize the risk of reporting to law enforcement agencies. In some cases, however, extortion is also exploited by mafia groups to take possession of profitable businesses by forcing entrepreneurs to undersell their company when they are no longer able to handle mafia organization's request of money (Paoli 2003; Savona 2012).

Extortion can take shape in different ways. The most common one involves the request of money in exchange for protection. However, mafia groups also resort to "indirect extortion" (Sciarrone 2009, p. 76). In this case, the mafiosi establish special relationships with a selected number of legal entrepreneurs, and they ask for materials or services normally provided by the entrepreneurs' businesses in exchange for protection. Another form of indirect extortion involves the request to hire affiliates or other people related to mafia members. By giving suggestions for hiring, mafia groups create new jobs for those who accept their presence and collaborate with them, and thus increase the collusive behavior within their communities (Sciarrone 2009).

Whereas extortion is common to all Italian mafia organizations, some differences emerge with regard to their presence in illegal markets and the way they conduct their criminal activities. These differences depend on the history and culture of the three mafia groups, as well as their internal organization and their expertise. These three factors – historical and cultural background, organization, and expertise – provide an explanation for the involvement of the 'Ndrangheta and the Camorra in drug trafficking, an illegal business formerly run by Cosa Nostra. The same factors also explain the involvement of the Camorra in counterfeiting and in the illegal traffic in waste. The Neapolitan mafia has indeed exploited the need for new places for waste disposal and the inefficiency of the public administration and corrupt politicians to enter the market, whereas the lack of the same combination of criminal opportunities in Sicily and Calabria have prevented Cosa Nostra and the 'Ndrangheta to do the same (Savona 2012).

Together with extortion, drug trafficking is among the most profitable illegal activities conducted by mafia organizations, and it is carried out in coexistence with other criminal groups. Indeed, in the 1980s and 1990s, mafia groups tried do gain the monopoly of the drug market both in their traditional communities and in some Italian cities, enforcing retailers to buy solely from them, but they never succeeded in controlling the whole exchange of drugs in an entire city or even in one neighborhood, except for a few mafia strongholds in southern Italy (Paoli 2004; DIA 2010).

The 'Ndrangheta is particularly active in the illegal drug market since 1980s, when the Calabrian mafia has succeeded in entering this profitable market partly at the expense of Cosa Nostra, whose members were no longer considered trustworthy interlocutors, especially by South American drug producers. Among drug trafficking activities, the 'Ndrangheta is mainly involved in the importation and wholesale distribution of drugs, especially cocaine and heroin. The position of strength of 'Ndrangheta families in the illegal drug market is guaranteed by their extensive network of contacts in Center and Northern Italy, both with their affiliates and with external drug dealers. In the last decades, the Calabrian mafia has also created outposts in foreign countries which serve as transit countries for the smuggling of drugs (CPA 2008; DIA 2010; Ciconte 2011).

The illegal activities mafias are involved in are not limited to the trafficking of drugs. Other illegal goods are provided by the Italian criminal organizations, such as firearms and counterfeited products. Mafia groups are also involved in the illicit traffic in waste, illicit gambling, loan sharking, frauds, as well as robberies and homicides (DIA 2010). The large profits are then invested both in the illegal markets and in the legitimate economy.

Infiltration of the Legitimate Economy

The infiltration of mafia organizations in the legitimate economy "has to do with the very nature of Italian mafia groups and their claim to exercise a political dominion within their communities" (Paoli 2004, p. 284). Extortion allows mafia groups to achieve an indirect control of companies, by forcing them to pay for protection services, hire specific people, or supply materials or services. Investments in the legitimate economy allow mafia organizations to extend their political dominion by infiltrating into and controlling legitimate economic sectors. The management of legal companies is indeed a way mafias can create new job opportunities and extend their power within their communities, as well as establish relationships with legal entrepreneurs and politicians. Also, the large illegal profits generated from illegal activities cannot be completely reinvested in the same illegal markets. The management of legal companies thus enables the mafiosi to launder part of their money and make even more profits from investments in the legal markets (Arlacchi 1988; Ruggiero 1996; Fantò 1999).

Among the economic sectors, mafia groups seem to favor the public construction industry, as well as other "protected" sectors, i.e., those where the public administration regulates the participation by requiring a specific license or permission and thus reduces the competition. These economic sectors are generally characterized by a low productivity, but they do not require specific skills nor technological innovations (Sciarrone 2009). Also, they allow mafia groups to capitalize on their competitive advantages, namely, the constant availability of floating assets and the possibility to discourage the competitors by resorting to violence (Arlacchi 1988).

Future Directions

Most research on Italian mafias focused on areas of their traditional presence, namely, Western Sicily (Palermo, Agrigento, and Trapani provinces), Southern Calabria (Reggio Calabria province), and Naples and its surroundings. Only recently the interest in the mobility of criminal groups has generated a small but significant number of studies on mafia transplantation into new territories (Sciarrone 2009; Varese 2011; Campana 2011). Different interpretations have been proposed on whether, why, and how mafia groups expand to new and distant territories. Peter Reuter (1985) and Diego Gambetta (1993) have argued that mafia organizations tend to be local in scope and are not likely to migrate. However, examples of successful mafia transplantation do exist (Varese 2011). The possibility that generalized migration of the population or of convicted mafiosi relocated outside their areas of origin might foster the mobility of criminal groups has also been discussed, as well as the hypothesis that involves specific mafia expansion strategies (Sciarrone 2009). To this day, little is known about the colonization and transplantation processes, whereas a significant body of research has explored the reasons Italian mafias emerged in their territories of origin. The adoption of more complex interpretations of the mobility of criminal groups might help identify the factors that foster transplantation and that make it successful.

Related Entries

- ▶ Co-offending
- ▶ Drug Trafficking
- ▶ Money Laundering
- ▶ Organized Crime, Types of
- ▶ Racketeering

Recommended Reading and References

Arlacchi P (1988) Mafia business. The Mafia ethic and the spirit of capitalism. Oxford University Press, Oxford (Original edition: 'La mafia imprenditrice. L'etica mafiosa e lo spirito del capitalismo'. Il Mulino, Bologna, 1983)

Barbagallo F (2010) Storia Della Camorra. Laterza, Roma-Bari

Behan T (1996) The camorra. Routledge, London

Blok A (1988) The Mafia of a Sicilian Village, 1860–1960. A study of violent peasant entrepreneurs. Polity Press, New York

Blok A (2008) Reflections on the Sicilian Mafia: peripheries and their impact on the centres. In: Siegel D, Nelen H (eds) Organized crime: culture, markets and policies. Springer, Dordrecht, pp 7–13

Campana P (2011) Assessing the movement of criminal groups: some analytical remarks. Global Crime 12(3):207–217

Catanzaro R (1992) Men of respect. A social history of the Sicilian Mafia. Free Press, New York

Ciconte E (2011) ’Ndrangheta. Edizione aggiornata. Soveria Mannelli, Rubbettino

CPA (Commissione Parlamentare Antimafia) (2008) Relazione annuale sulla ’Ndrangheta. Doc. XXIII, N. 5, XV Legislatura. Commissione parlamentare di inchiesta sul fenomeno della criminalità organizzata mafiosa o similare, Roma

DIA (Direzione Investigativa Antimafia) (2010) Relazione Del Ministro dell’Interno Al Parlamento Sull’attività Svolta e Sui Risultati Conseguiti Dalla Direzione Investigativa Antimafia. 2° semestre 2010. Direzione Investigativa Antimafia, Roma

van Duyne PC (1997) Organized crime, corruption and power. Crime, Law Soc Change 26:201–238

Fantò E (1999) L’impresa a Partecipazione Mafiosa. Economia Legale Ed Economia Criminale. Edizioni Dedalo, Bari

Finckenauer JO (2005) Problems of definition: what is organized crime? Trends Organ Crime 8(3):63–83

Franchetti L (1974) Condizioni Politiche Ed Amministrative Della Sicilia. In: Franchetti L, Sonnino S (eds) Inchiesta in Sicilia. Vallecchi, Firenze

Gambetta D (1993) The Sicilian Mafia: the business of private protection. Harvard University Press, Cambridge

Hess H (1973) Mafia and Mafiosi: The structure of power. Saxon House, Farnborough (Original edition: ‘Mafia. Zentrale Herrschaft und lokale Gegenmacht’, Mohr, Tubingen, 1970)

Lupo S (1996) Storia Della Mafia: Dalle Origini Ai Giorni Nostri, 2nd edn. Donzelli Editore, Roma

Paoli L (2002) The paradoxes of organized crime. Crime, Law Soc Change 37(1):51–97

Paoli L (2003) Mafia brotherhoods: organized crime, Italian style. Oxford University Press, New York (Original edition: ‘Fratelli di Mafia: Cosa Nostra e ‘Ndrangheta’. Il Mulino, Bologna, 2000)

Paoli L (2004) Organised crime in Italy: Mafia and illegal markets – exception and normality. In: Fijnaut C, Paoli L (eds) Organized crime in Europe: concepts, patterns and policies in the European Union and beyond. Springer, Dordrecht, pp 263–302

Reuter P (1985) The organization of illegal markets: an economic analysis. National Institute of Justice, Washington, DC

Ruggiero V (1996) Economie Sporche: L’impresa Criminale in Europa. Bollati Boringhieri, Torino

Santino U (2006) Dalla Mafia Alle Mafie: Scienze Sociali e Crimine Organizzato. Rubbettino, Soveria Mannelli

Savona EU (2012) Italian Mafias’ asymmetries. In: Siegel D, van de Bunt H (eds) Traditional organized crime in the modern world: responses to socioeconomic change. Springer, New York, pp 3–25

Schneider J, Schneider P (1976) Culture and political economy in Western Sicily. Academic, New York

Sciarrone R (2009) Mafie Vecchie, Mafie Nuove: Radicamento Ed Espansione. Nuova edizione riveduta e ampliata. Donzelli Editore, Roma

TrMi (Tribunale di Milano) (2010) Ordinanza Di Applicazione Di Misura Coercitiva Nei Confronti Di Agostino + 155. Operazione ‘Il Crimine’. Procedimento n.43733/06 RGNR, n.8265/06 R.GIP

U.S. Department of State (2011) International Narcotics Control Strategy Report (I.N.C.S.R.). U.S. Department of State, Bureau for International Narcotics and Law Enforcement Affairs, Washington, DC

Varese F (2011) Mafias on the move: how organized crime conquers new territories. Princeton University Press, Princeton

Italian Organized Crime

► Italian Mafia

J

Jail

▶ Drug Abuse and Alcohol Dependence Among Inmates

Job of a Judge

▶ Evolving Judicial Roles

Joint/Group Offending

▶ Co-offending and Offender Attributes

Journey to Crime

▶ Applied Geographical Profiling

Judicial Leadership and Performance

Ingo Keilitz
CourtMetrics, Williamsburg, VA, USA

Synonyms

Governance; Organizational performance measurement; Performance management

Overview

Judicial leadership is what leaders of court systems do to translate values, visions, and goals into exceptional organizational performance. It means mobilizing and inspiring employees and other stakeholders to get extraordinary things done for their organizations. This entry explores the relationship between leadership – widely considered the most important element of effective justice system administration – and self-governed, well-managed, and operationally efficient courts. It focuses on organizational performance measurement, a precondition of effective self-governance of courts.

The theme of this entry is that the relationship between leadership and effective organizational self-governance of courts, especially the role of performance measurement, is uniquely important for courts. Buttressed by the doctrine of separation of powers doctrine and the principle of judicial independence, courts enjoy, at least theoretically, a greater level of freedom from scrutiny and interference than other components of the executive and legislative branches of government. This freedom and institutional independence cannot maintain legitimacy in practice without effective self-governance; and effective self-governance is dependent on judicial leadership. In practical terms, effective self-governance of courts is the manifestation of the separation of powers doctrine and the principle of judicial independence.

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

To achieve a truly independent (and, practically speaking, a self-governing) judiciary requires judicial system leadership to overcome resistance from various entrenched interests that view courts as not deserving or not capable of self-governance, or both. To be effective, self-governance should drive a judicial system leadership committed to transparency and accountability and to rigorous performance measurement and performance management that ensures adherence to the highest standards of organizational performance.

The organization of this entry is as follows. It begins with a brief definition of court performance measurement and performance management. The next section makes the case for the unique importance of transparency and accountability to the self-governance of courts. Stated simply, without transparency and accountability for performance, it is unlikely that judicial systems can be institutionally independent. Judicial leaders must ensure that a transparent and accountable self-governed framework exists not just in the "law and the books" but also in the "law in practice." Described in the next section is the imperative of value-oriented performance measurement and how this imperative folds into a broad vision of self-governed, well-managed, effective, and operationally efficient courts.

The difficulties of thinking in terms of organizational performance measurement and its relationship to effective self-governance contribute to the theme of the next section of this entry under the heading "How Are We Performing?" It describes two related impediments that have stood in the way of the advance of performance measurement, transparency, and accountability, and the self-governance of courts. The first is the reliance of courts and justice systems throughout the world on third-party monitoring and evaluations, which are quite different from performance measurement on various dimensions, to assess their performance. The second is a related tendency to think in terms of a research or program evaluation paradigm when addressing the question of *How Are We Performing*? – a paradigm ill suited to answer the question except in a very narrow sense. The theoretical foundations and methodology of the disciplines of performance measurement and research overlap, of course, but they are very different in important aspects – purposes, sponsorship, organization, audience, functions, timing, and data interpretation rules – that are critical for self-governance of courts.

The final section of this entry concludes with the proposition that judicial system leaders should focus on ensuring an effective self-governance of courts that is transparent and accountable and to committing to the primacy of performance measurement to drive success.

Court Performance Measurement Defined

Court performance measurement is the process of monitoring, analyzing, and using performance data on a regular and continuous basis for the purposes of transparency and accountability and for improvements in efficiency and effectiveness. This definition encompasses both performance measurement per se and the use of performance data in management (sometimes referred to as performance management). Measures of court performance include, for example, those promulgated by the National Center for State Courts (National Center for State Courts 2005, 2009):

- Court user/citizen assessment and satisfaction with court services including access to the courts, fairness and integrity, timeliness and expedition, and their general trust and confidence in the court
- Clearance rates, that is, the number of outgoing cases as a percentage of the number of incoming cases
- Case finalization time (also referred to "time to disposition," "on-time case processing," and other terms)
- Time in custody prior to trial (median number of days in pretrial detention)
- Reliability and accuracy of court files
- Backlog, that is, percentage of cases in the system longer ("older") than established timeframes
- Trial date certainty expressed the proportion of all trials that are held when first scheduled

- Court employee engagement (a proxy for court success)
- Recovery of criminal and civil court fees as a proportion of fees imposed
- Money expenditure per case (net costs per finalization)

These measures used by United States courts and, on a more limited basis, by courts in other parts of the world are linked to success factors and core values; they form a set of balanced measures of a court's or court systems' performance; they focus on outcomes, that is, how well the courts are making things better as a result of their efforts instead of how much effort has been expended or by how many resources are used; they represent strategic planning in action; and, finally, they provide clarity and strategic focus to guide judicial leadership.

With the rapid advance of data analytics and business intelligence into judicial systems around the world, the capacity of courts to measure and manage their own performance has grown exponentially (Keilitz 2010). Court leaders and managers today can access and analyze performance data in seconds, what in the traditional world of courts would take weeks or months. It is not yet clear, however, to what extent courts have actually used performance data on a regular basis to guide planning, budgeting, strategy, and management decisions, a state of affairs that is probably true of governments in general, at least in the United States (Hatry 2010).

Judicial Independence, Transparency, and Accountability: The Law in Practice

In the United States, an independent judiciary is taken for granted as a matter of principle. Central to United States constitutional law is the doctrine of the separation of powers and the dynamic of checks and balances of three separate and independent functions of what we today see as the legislative, executive, and judicial branches of government. Article III of the United States Constitution vests judicial power in the third branch of government. The constitutions of 39 other States also *expressly* require that governmental power is divisible among three coequal separate branches. The function of checks and balances renders the judiciary dependent on the other two branches for resources. The purpose of this separation of powers is to prevent the concentration of governmental power and to provide for checks and balances.

Unfortunately, this constitutional framework and related legislative and policy mandates are not self-executing in practice, not in the United States or elsewhere throughout the world. There are few constitutions and laws, if any, that require judicial systems to be self-governing or specifically authorize the means to ensure that they are. Institutionally independent does not necessarily mean self-governing. As argued by some scholars, the difference between the United States, where self-governance is arguably more in evidence, and other judiciaries has less to do with prevailing theories of how popular sovereignty relates to jurisprudence and political theory but with the institutional capacities of courts to act independently in practice.

Three critical elements of modern democracies – and, arguably, the preconditions of a market-based economy – are a capable state, the state's subordination to the rule of law, and government accountability to its citizen. There can be little doubt that the ideas and principles underlying these elements that are enshrined in law are fundamental causes of why some states fail and others prosper. Yet, the "law on the books" is not sufficient (Dakolias 2005). It is, of course, important to know what the law established as formal practices, but this will not be enough to understand the reality on the ground. Creating remedies for case processing delay and court congestion, for example, requires examining causes beyond the law to organizational structure and practice including the size of demand for dispute resolution and the points in the processing of cases where perverse incentives might exist that are contrary to efficiency.

Rule of law reform, which has been underway throughout the world for quite some time, began with a focus on drafting and enacting laws, codes, and rules. These initiatives, however, soon gave way to attention paid to the "law in practice" as opposed to the "law on the books," as it became clear that laws are quite literally of little use if no efforts are devoted to establishing the means of their implementation and enforcement.

In a 2002 study of 75 countries, Lars Feld and Stefan Voigt looked into the effects of an independent judiciary (Feld and Voigt 2003). They distinguished between de jure independence (that expressed in formal law) and de facto independence (that which is implemented and enforced). Their de jure indicator was based on 12 variables made up of 23 characteristics concerning the legal basis for judicial independence (including institutional arrangements, appointments, tenure, salaries, and transparency). The de facto indicator based on eight variables measured over time in 75 countries shows the degree of judicial independence in practice (effective average term length, changes in number of judges, income, and court budgets). Not surprisingly, the main finding of the study is that judicial independence "on the books" is insufficient. They found that de jure independence was not correlated with economic growth, whereas de facto independence was conducive to economic growth. Indeed, de jure and de facto independence appeared to be almost completely unrelated, as there was no overlap whatsoever between rankings of the top ten countries examined.

Because courts traditionally have a greater degree of freedom from scrutiny compared to other public institutions, courts need to balance this independence with transparency and accountability for performance. As noted by the former president of the International Association for Court Management, Marcus Zimmer, judicial leaders often "fail to fathom that attaining such independence is conditioned upon their willingness to accept the commensurate levels of individual and institutional accountability." Institutional independence, he writes, means "having in place an accountable self-governance framework built on effective internal control and enforcement mechanisms to ensure the confidence of the other branches and the trust of the people in its capacity to independently administer justice" (Zimmer 2011, p. 138). Regular and continuous performance measurement addresses this need. Performance measurement promotes an independent judiciary by providing transparency and accountability.

The Performance Measurement Imperative

An imperative of value-oriented performance measurement and performance management folds well into a broad vision of judicial leadership of self-governed, well-managed, effective, and operationally efficient courts. This imperative rests on five basic assumptions that speak to the relationship between judicial leadership and performance measurement.

Performance matters. Successful leaders show a strong preference for outcome measurement that gauges the desired results of program of services instead of measures of inputs (such as the number of staff, costs, or hours worked by judges and staff). Nothing else really matters as much as results defined in terms of quality, that is, the achievement of good results as efficiently as possible. The quality of courts' success should not be measured by how many hearings are held or even by the number of cases that are resolved or by the number of programs and processes they call for. What matters are outcomes that matter to the people served by courts. Self-governance of courts should be organized for performance.

Effective performance measures focus on ends, not the means to achieve them. They emphasize the condition or status of the recipients of services or the participants in court programs (outcomes) rather than the internal aspects of processes, programs, and activities (inputs and outputs). They focus on results rather than quantification of resources or level of effort. Traditionally, court managers have relied on measures of volume or frequency in three categories: (a) amount of work demand (such as the number of cases filed), (b) number of products or services delivered (such as the number of cases filed), and (c) the number of people served. Increasingly, there is the recognition that while such measures show demand and how much effort has been expended to meet that demand, they reveal nothing about whether the effort has made any difference – that is, whether anyone is better off as a result.

Second, performance is not about the numbers. Performance measurement uses numbers, but it is not about the numbers. It is about the perception, the understanding, and the insight required of effective leadership. Ultimately, it is not the measure itself that is important, but rather the questions that it compels judicial leaders to confront (Spitzer 2007) questions such as:

- How well is the court, court system, or justice system performing?
- Where is the court now (performance level, baseline)? What is the current performance level compared to established upper and lower "controls" (e.g., performance targets, objectives, benchmarks, or tolerance levels)?
- How well is the court performing over time? Is performance better, worse, or flat? How much variability is there? (trend analysis)
- Why is this particular performance happening (analysis and problem diagnosis)? What happened to make performance decline, improve, or stay the same. What are some credible explanations?
- What is the court doing to improve or maintain performance levels? (planning future outcomes)
- What actions should be started, continued, or stopped altogether as a result of what the measure reveals? What should be done to improve poor performance, reverse a declining trend, or recognize good performance? (strategy formulation)
- What performance targets and goals should we set for future performance (goals)?

Third, courts must count what counts and measure what matters. Figuratively and literally, performance does not count unless it is related to the things that really matter and are critical to the success of a court. Key success factors have been referred to in the literature of organizational performance measurement as major performance areas, standards of success, perspectives, domains, performance criteria, key results factors, and key outcomes. Whatever they are called, they form the framework of a court's accountability and transparency to the public and other stakeholders.

Alignment of performance measurement with purpose and fundamental responsibilities is vital. Well-conceived performance measures serve to align an organization's efforts with the achievement of its mission. The requirement of linkage of performance measures to a court system's mission and strategic goals is vital.

Fourth, performance measurement is a powerful antidote for too much information. Information overload is one of the biggest irritations of modern life. And it seems to be getting worse. It can make justice system executives and managers feel anxious and powerless, reduce their creativity, and render them less productive. A profusion of phrases describe the anxiety and anomie caused by too much information: data asphyxiation, data smog, information fatigue syndrome, and cognitive overload. Surveys have found that most managers believe that the data deluge has made their jobs less satisfying or hurt their relationships. Some think that it has damaged their health. Many managers think most of the information they receive is useless. The explosion of information hitting the courts is going off within systems too fragmented and disorganized to absorb it. Performance measurement, supported by business intelligence tools like performance dashboards, is an antidote and a remedy to data fog. It focuses on what counts and filters out the rest.

Finally, measuring and managing court performance is an essential survival skill for court leaders and managers. The right performance measures effectively delivered are clear, unambiguous, and actionable. Focus and clarity are factors of effective leadership. Above all, leaders need to be clear. Performance measures such as clearance rates or court user/citizen satisfaction focus on a limited number of success factors like access, fairness, and timeliness of case processing. They count only what counts and measure only what matters.

The discipline of performance measurement provides a conceptual shortcut to a host of organizational competencies like strategic planning, resource management, and communication with stakeholders. The benefits of an effective court performance measurement and management system are the same, for example, as those of strategic planning – that is, accountability, consensus building, focus, coordination, control, learning, communication, hope, and inspiration.

To identify the right performance measures, a court system must address the same fundamental questions about guiding ideals, values, mission, goals, and broad strategies as it must address in strategic planning. When it identifies a core performance measure such as court user/citizen satisfaction with court services, for example, it communicates a clear, simple, and penetrating theory of its "business" – its ideals and purpose – that informs decisions and actions.

How Are We Performing?

The question *How Are We Performing*? lies at the heart of self-governance and effective leadership of courts. The capacity and political will to address this self-directed question regularly and continuously using the tools of performance measurement is the hallmark of a successful court organization. More and more court leaders and managers are turning to performance measurement to drive success. However, the trend is on a slow march impeded by two entrenched ways the judicial sector has tended to measure its success: (1) reliance on third-party monitoring and evaluations of court performance and (2) an adherence to a research paradigm for assessing court performance.

Self-governance, transparency, and accountability of courts will depend on the degree to which these two impediments are attenuated. Court leaders will need to champion and support the self-assessment of performance by courts, instead of relying on third-party evaluations. They will need to seek the replacement of the methodologies of the disciplines of *research* or *program evaluation* (or evaluation research), which have dominated justice sector assessment in the past, with that of *performance measurement and management*. These suggestions are consistent with the values, principles, and tools of the National Center for State Courts' High Performing Courts Framework (Ostrom and Hanson 2010), which is oriented toward United States courts, and the International Framework for Court Excellence (International Consortium for Court Excellence 2010).

Self-Assessment of Court Performance

As with most changes in life, self-directed change is the most meaningful and long lasting. Self-assessment, and not third-party monitoring and program evaluation (or evaluation research), is the hallmark of successful courts. It is integral to effective self-governance. A successful court has the capacity and the political will for self-directed rigorous performance measurement and management that addresses the question *How Are We Doing*? The "we" in the question suggests the critical differences between court performance measurement and third-party evaluations of courts.

Monitoring and Evaluation

Performance measurement is not yet the norm in the United States and around the world, though it has a strong foothold in the United States and large parts of the developed and developing world. Most assessments of programs, processes, and reform initiatives in courts are accomplished instead by monitoring and evaluations instigated and conducted by third parties, including funding agencies, donors, aid providers, and their agents (researchers, analysts, and consultants). The abiding concern of these third parties is return on their investments, and this concern does not necessarily align with the expressed purposes and fundamental responsibilities of courts, at least as these might be conceived by court leaders and managers, and in seminal authorities like the *Trial Court Performance Standards* (Keilitz 2000). For the most part, the focus of these third-party assessments is a specific initiative, program, or process (e.g., juvenile drug courts, small claims mediation, and summary jury trials).

Performance data produced by monitoring and evaluation efforts are used primarily in service of decisions to increase, decrease, or redirect funding or other support. While some performance data may be shared with courts and justice system, as a practical matter, most are collected, analyzed, interpreted, and used, first and foremost, by third parties, that is, funders, donors, aid providers, and their agents. The measures of success are defined and results interpreted by these third parties with little or no shared responsibility or accountability (ownership) by the courts or justice systems implementing the programs, processes, and reform initiatives.

Because of their time requirements and costs, these evaluation efforts are, at best, limited and generally unsatisfactory responses to the question *How Are We doing*? The research paradigm within which these third-party monitoring and evaluation operate makes the results less than responsive to the question. Large-scale studies can take 2–3 years before data reports become available. Therefore, the result of regular and continuous performance measurement by courts themselves is likely to be the major source of information for addressing the question.

Indexing Performance

Another type of third-party monitoring and evaluation of performance of courts takes the form of indexing of justice sector performance. Indexes are useful tools. The idea that complex things can be pinned down and quantified on a simple scale seems universally appealing. Indexes that reduce justice sector performance to a single number for purposes of comparing, ranking, and rating countries are being used throughout the world to understand everything from governance, corruption, economic vitality, health, and education to the quality of life. Governments and reformers are taking such indexes seriously. They are closely watched. “There is nothing,” wrote the *economist* in an October 9, 2010, report of the results of the Mo Ibrahim Foundation’s Index of African Governance, “like a bit of naming, shaming, and praising.”

The crowded field of performance indexes includes comprehensive indicators and indexes that encompass entire countries but include aspects of justice, like the Mo Ibrahim Foundation’s Index, and others more narrowly focused on rule of law like the World Justice Project’s *WJP Rule of Law Index*™ and the American Bar Association’s Judicial Reform Index. The World Justice Project’s *WJP Rule of Law Index*™ incorporates ten elements of the rule of law – such as limited government powers, fundamental rights, and clear, publicized laws – and 49 associated sub-factors that make general reference to various justice “systems” but reference to courts is conspicuous by its absence.

While the indexes are successful in getting people’s attention by “naming, shaming, and praising” the jurisdictions rated and ranked, buy-in of the leaders and managers of courts and court systems may be limited. Well-known indexes like the World Justice Project’s *WJP Rule of Law Index*™ and the American Bar Association’s (ABA) Rule of Law Initiative’s Judicial Reform Index might be seen to reflect the ethos of the sponsoring organizations, and not necessarily the values, purposes, and fundamental responsibilities of courts. Both of these well-known indexes rely heavily on polls of commissioned experts who assess the factors the third parties deem important to judicial reform.

Many court leaders and managers and most elected officials are interested in comparing courts. However, they are not enthusiastic about having such comparisons reported externally unless they stack up well against others (Keilitz 2005). Perhaps stemming in part by a broad interpretation of judicial independence that does not embrace transparency and accountability, many judges have a viscerally negative reaction to public reporting on the quality of judicial services, and some may be appalled that their judicial systems are ranked numerically by outside “experts” on the basis of what they perceive as misinformation.

Third-party assessment of performance, whether it takes the form of indexing or monitoring and evaluation, is the antithesis of self-governance. Because it is not self-directed by courts, it is not as likely to be embraced by its leaders and managers and lead to reform. As might be expected, when performance assessment is, or perceived to be, wholly initiated and executed by external organizations, more energy and resources may be expended by the leaders of courts and court systems on refuting poor evaluation results or low rankings than on developing strategies for sustained reform.

Performance Measurement

Both the disciplines of performance measurement and research adhere to the scientific method and use statistical thinking to draw conclusions. But, as already suggested above, performance measurement and research, or evaluation research, are vastly different in their purposes, functions, sponsorship, uses, and the way they are funded and structured. Sorting out the differences between them, and choosing to employ one over the other, is not just academic hairsplitting.

First, the *purposes* of the two disciplines are quite different. At a very fundamental level, the purpose of performance measurement is to answer the question *How Are We Doing*? in response to the demands for transparency and accountability from stakeholders and to provide a basis for improvement. Performance measurement can give clues to why outcomes are as bad or good as they are, but it cannot go the full distance of determining why things are as they are. Performance measurement can help to identify variations in performance and to isolate where and when those variations occur (e.g., an upward trend in the public's rating of the courts is largely due to attorneys' increased satisfaction with case processing timeliness after the court initiated electronic filing) so that decisions and actions can target improvements. It does not determine causes. Causal inference is the domain of research that helps us to understand why something has occurred. (Of course, an important value of performance data is to trigger in-depth evaluation research.) As noted above, the purpose of evaluation research in the justice sector, for the most part, is to answer the question "What has worked and what has not, and why?" in order to justify donors' or funders' investments various initiatives, programs, and processes.

Second, performance measurement and research differ in their *sponsorship and audience* – that is, who is doing it and for (or "to") whom. Performance measurement is done by the courts, for the courts. Results are made known to court leaders and managers. Distribution of performance data to the public and other stakeholders is done at the discretion and direction of the court, preferable in real (or near real time) in a wholly transparent manner. Evaluation research, on the other hand, is more often than not sponsored or instigated by third parties (e.g., administrative offices of the courts or funding outside agencies). At the extreme circumstances, courts and other justice sector institutions are mere "subjects" of the research. Those conducting the research have no compunction to share the research results with court leaders or managers except as a courtesy or as a quid pro quo for the courts' participation in the research.

Third, the *functions* of performance measurement are specific and targeted, that is, establishing a baseline for current performance, setting organizational goals and assessing whether performance is within determined boundaries or tolerances (controls), identifying and diagnosing problems, determining trends, and planning. Performance measurement is done for the utilitarian and practical purposes of making improvements in court programs, services, and policies. Court leaders and managers use performance information to make improvements in programs and services. Specifically, they might use performance measurement for a number of management purposes such as translating vision, mission, and broad goals into clear performance targets; communicating progress and success succinctly to the public and other stakeholders; responding to legislative and executive branch representatives' and the public's demand for accountability; formulating and justifying budget requests; responding quickly to performance downturns (corrections) and upturns (celebrations) with day-to-day resource allocation; motivating employees to make improvements in programs and services; setting future performance expectations based on past and current performance levels; and insulating the court from inappropriate performance audits and appraisals imposed by external agencies or groups. Evaluation research, on the other hand, seeks truth about the worth and merit of an initiative, program, or processes and is intended to add to our general knowledge and understanding, especially regarding future investments in those initiatives, programs, or processes.

Fourth, performance measurement and evaluation research adhere to different *design and data interpretation* protocols. Consistent with self-governance, performance measurement is focused on the performance of individual courts with the aim of individual accountability. Research, on the other hand, is interested in the generalizability of findings to all courts.

Of course, both performance measurement and evaluation research must adhere to the requirements of the scientific method. Both use quantitative and qualitative methods including surveys and questionnaires, interviews, direct observation, recording, descriptive methods, tests and assessments, and statistical analysis. But these requirements and methods are applied differently in performance measurement and evaluation research. For example, sample sizes may be smaller and levels of confidence lower in performance measurement primarily because replication of results is done on a regular and continuous basis as a critical matter of design. Evaluation research, on the other hand, is episodic. It is done when time and funds permit.

The matter of replication of results highlights a critical design difference between performance measurement and evaluation research. Basically, replication means repeating the performance measurement or evaluation research to corroborate the results and to safeguard against overgeneralizations and other false claims. Repeated measurements – that is, replication – on a regular and continuous basis are part of the required methodology of performance measurement. Analyzing trends beyond initial baseline measurement requires replication of the same data collection and analysis on a monthly, weekly, daily, or, in the case of automated systems, on a near real-time basis. In contrast, replication in research is a methodological safeguard that is universally lauded by scientists, but seldom done in practice.

Conclusion

Judicial leadership is the essential ingredient of the effective and efficient administration of courts. Effective judicial leaders fold the imperative of value-oriented performance measurement into a broad vision of self-governed, well-managed, effective, and operationally efficient courts. They are committed to transparency and accountability. They recognize that institutional independence and self-governance requires courts to be open and accountable. Effective judicial leaders focus on outcomes and use performance measurement as a tool to mobilize and inspire employees and other stakeholders and to drive improvements.

Consider how this relationship between leadership and self-governed, well-managed, and operationally efficient courts plays out in the minds of two successful judicial leaders, Christine M. Durham and Daniel Becker, the chief justice and the state court administrator of Utah, respectively:

> We in the courts should know exactly how productive we are, how well we are serving public need, and what parts of our system and services need attention and improvement. This includes measuring the accessibility and fairness of justice provided by the courts as measured by litigants' perceptions and other performance indices. And we should make that knowledge a matter of public record. (Durham and Becker 2011)

There is no doubt that to achieve a truly independent and self-governing judiciary requires vigorous judicial leadership to overcome resistance from various entrenched interests that view courts as not deserving or capable of self-governance. Most states in the United States today have some form of legislation requiring performance measurement of government institutions and agencies (see Lu et al. 2009). Who does it, how it is done, and whether it will lead to successful self-governance of courts will be determined by its leadership.

Recommended Reading and References

Dakolias M (1999) Court performance around the world: a comparative perspective Yale Hum Rts & Dev L J. Available online at: http://siteresources.worldbank.org/BRAZILINPOREXTN/Resources/3817166-1185895645304/4044168-1186404259243/14pub_br176.pdf

Dakolias M (2005) Methods for monitoring and evaluating the rule of law. Paper presented at the Center for International Legal Cooperation's 20th Anniversary Conference. Applying the "Sectoral Approach" to the Legal and Judicial Domain, The Hague, Netherlands, 22 Nov 2005

Durham CM, Becker D (2011) A case for court governance principles. Perspectives on state court leadership. Harvard University Executive Session for State Court leaders in the 21st century

Feld LP, Voigt S (2003) Economic growth and judicial independence: cross country evidence using a new set of indicators. Eur J Polit Econ 19(3):497–527

Hatry HP (2010) Looking into the crystal ball: performance management over the next decade. Public Administration Review, Special Issue, December 2010, S208–S211

International Consortium for Court Excellence (2010) International framework for court excellence. Available online at http://www.courtexcellence.com/pdf/IFCE-Framework-v12.pdf

Keilitz I (2000) Standards and measures of court performance. In criminal justice 2000, vol 4. Measurement and analysis of crime and justice. U.S. Department of Justice, Office of Justice Programs, National Institute of Justice, Washington, DC, July 2000, pp 559–593. Available online at http://www.ncjrs.org/criminal_justice2000/vol_4/04k.pdf

Keilitz I (2005) How do we stack up against other courts? The challenges of comparative performance measurement. Court Manag 19(4):29–34, Winter 2004–2005

Keilitz I (2010) Smart courts: performance dashboards and business intelligence. In: Future trends in state courts 2010. National Center for State Courts, Williamsburg. Available online at http://vis-res.com/pdf/Trends2010.pdf

Lu Y, Willoughby K, Arnett S (2009) Legislative results: examining the legal foundations of PBB systems in the states. Publ Perform Manag Rev 33(4):671–676

National Center for State Courts (2005) CourTools: trial court performance measures. National Center for State Court, Williamsburg, Available online at http://www.ncsconline.org/D_Research/CourTools/index.html

National Center for State Courts (2009) CourTools: appellate court performance measures. National Center for State Courts, Williamsburg, Available online at http://www.ncsconline.org/D_Research/CourTools/index.html

Ostrom B, Hanson R (2010) Achieving high performance: a framework for courts. Working paper series. National Center for State Courts, Williamsburg. Available online at http://www.ncsc.org/information-and-resources/high-performance-courts.aspx

Spitzer DR (2007) Transforming performance measurement: rethinking the way we measure and drive organizational success. AMACOM, New York

Zimmer MB (2011) Judicial institutional frameworks: an overview of the interplay between self-governance and independence. Utah Law Rev 2011(1):121–139

Judicial Temperament

▶ Evolving Judicial Roles

Junk Science

▶ Legal Rules, Forensic Science and Wrongful Convictions

Juries, Lay Judges, and Trials

Toby S. Goldbach and Valerie P. Hans
Cornell University Law School, Ithaca, NY, USA

Overview

"Juries, Lay Judges, and Trials" describes the widespread practice of including ordinary citizens as legal decision makers in the criminal trial. In some countries, lay persons serve as jurors and determine the guilt and occasionally the punishment of the accused. In others, citizens decide cases together with professional judges in mixed decision-making bodies. What is more, a number of countries have introduced or reintroduced systems employing juries or lay judges, often as part of comprehensive reform in emerging democracies. Becoming familiar with the job of the juror or lay citizen in a criminal trial is thus essential for understanding contemporary criminal justice systems in many countries. This entry reviews procedures for selecting jurors and lay judges and outlines lay participation in fact finding and in sentencing phases of the criminal trial. It also assesses the promises and challenges of lay participation in law. Reviewing and evaluating the effects of the different approaches that countries have taken to incorporating lay citizens, it reflects on whether the goals of democratic deliberation are being met in both jury and lay judge systems. It concludes with suggestions for future directions for research.

Doing the Job of Democracy

The custom of employing ordinary citizens as legal decision makers in criminal trials is widespread. Originating in early legal systems, this practice has continued in modern times as a vehicle for including a democratic element in law. In approximately 50 countries, ranging from Australia to Kazakhstan to Spain and Sri Lanka, citizens serve as jurors, deciding cases in independent bodies separate from professional judges. In many other countries, such as Italy, Poland, and Japan, citizens participate as lay judges (alternately called lay assessors), deciding cases together with professional judges in mixed decision-making bodies. Whether citizens serve as jurors or lay judges often depends on whether the legal system derives from a common or a civil law tradition. Juries are more likely to act as independent decision makers in common law adversarial trials, whereas lay judges appear more often in civil law inquisitorial trials and tribunals (Thaman 2011).

Delegating the task of adjudicating a criminal trial to ordinary citizens promotes democracy in both legal and political institutions. Laypersons bring the voice of the people to the law. They draw on their own life experiences, allowing trial decisions to more accurately reflect the society that the legal system serves. This is a form of representative democracy in legal institutions (Malsch 2009). When the trial is finished, lay citizens return to the community, sharing the lessons they learned about the law, ensuring an accountable and transparent legal system.

Employing juries and lay judges supports and enhances democratic political institutions. Laypersons act as a check on authority. They balance the power imbued in state officials such as prosecutors or professional judges. And by contributing to decisions that directly affect the people, juries and lay judges engage in an act of self-governance, reestablishing the people's sovereignty. This act of self-governance in the legal realm affects political participation. As individuals perform the duties of a juror or lay judge, they become more connected to the world around them. The French political thinker Alexis de Tocqueville believed that jurors internalized the duties they owed to society and the role they ought to play in government. Writing in admiration of the US jury system, he observed that the jury communicates "the spirit of the judges to the minds of all the citizens" and that "this spirit, with the habits which attend it, is the soundest preparation for free institutions" (Tocqueville 1835/1945, p. 289).

Jurors and lay judges also act as conflict resolvers. They attend to the conflict between the state (as representative of the victim and the community) and the defendant by reaching a verdict in cases that may feature difficult and often complex competing narratives. And, depending on the jurisdiction or type of crime, lay judges and juries may participate in punishment decisions for convicted defendants.

Despite the important role that they play in the criminal trial, juries and lay judges face challenges on several fronts. First, critics argue that the evidence and legal arguments now being introduced in contemporary trials are too complex for untrained eyes. These critics question the lay participant's ability to understand new technologies and scientific methods such as DNA or mtDNA and accuse juries in particular of relying on expert credentials rather than evaluating the substance of expert testimony. A broad trend toward professionalization in the nineteenth and twentieth centuries deepened a mistrust of lay participation in the law and other expert domains. Professionals were thought to be superior to ordinary citizens because of their theoretically rich education, instruction in the practical skills of decision making, and greater adherence to rational scientific principles. Now, in a contemporary twist, modern technology invites new concerns about the soundness of lay decision making, with some worrying about jurors' use of cell phones or Internet to obtain information about cases and parties that would otherwise not be available in the courtroom proceedings (Waters and Hannaford-Agor, in press).

Moving from courtroom proceedings to posttrial coverage, a second concern about lay participation is evident in the way the public scrutinizes criminal trial verdicts. Media and members of the public often disparage jury decision making following controversial judgments, especially in cases where a notorious defendant is found not guilty. The popular press in the United States conveyed public outrage following verdicts of not guilty for O.J. Simpson and Casey B. Anthony, trials that were decades apart. Similarly, support for Spain's recently introduced jury system dropped precipitously following a jury's acquittal of a defendant who was on trial for killing two police officers (Hans 2008).

Remarkably, one of the greatest challenges to lay participation in legal decision making is the sharp decline in criminal trials. Despite a constitutional right to trial by jury, in the United States, plea bargaining and other settlements resolve almost all criminal cases. As a result, the number of jury trials has declined in both state and federal courts. Today, US juries decide just 5–10 % of all criminal cases. In the United Kingdom, the number is even lower, at 1 % of cases. In a surprising contrast, a number of countries outside the United States and the United Kingdom have recently introduced new lay participation systems, often as part of broad efforts to support emerging democratic systems (Hans 2008; Kovalev 2010; Marder 2011).

Historically, European countries transplanted all or part of their legal systems to their colonies. Thus, juries were found in legal systems in the United States and in British colonies in Africa, parts of India, Australia, New Zealand, and some Caribbean and South American countries (Hans 2008; Vidmar 2000). Several countries dismantled the institution postindependence, particularly where native populations saw jury verdicts as furthering the interests of colonizers or where the right to a jury trial was not extended to indigenous peoples. But now, citizen participation is seeing a rebirth, particularly in countries that are democratizing following periods of authoritarian rule (Lempert 2007). Notable examples include Spain, which began conducting jury trials in the mid-1990s, as well as Russia (Thaman 2011). In fact several post-Soviet countries, including Armenia, Kazakhstan, Ukraine, and Azerbaijan, enshrined a right to trial by jury in their constitutions (Kovalev 2010). The newly independent country of Georgia modeled its jury system on US juries and held its first criminal jury trial in the fall of 2011. In the last decade, East Asian countries enacted court reforms which created new and innovative methods of lay participation: Japan adopted a mixed court with lay judges, whereas South Korea introduced a jury system. Other countries, while not adopting the independent jury model, have altered their trial processes away from dossier-based inquisitorial approaches toward more adversarial oral evidence presentation to facilitate eventual citizen participation.

Becoming familiar with the job expected of jury and lay citizens is therefore essential for understanding contemporary criminal justice systems in many countries. This entry first describes the independent fact finding of jury systems and then turns to a consideration of mixed courts of lay and professional judges. It summarizes research evidence about how well lay decision makers function and reflects on whether the goals of democratic deliberation are being met. The entry concludes by considering contemporary challenges and future directions for research.

Juries at Trial

The trial by a jury of one's peers dates back to thirteenth-century England. These early juries were composed of land-owning white men, who testified about their personal knowledge of local disputes. Over time the institution changed so that jurors functioned less like witnesses and more like fact finders. In England, distinct roles for judge and jury developed in the 1700s. A similar evolution took place in the United States toward the end of the nineteenth century (Vidmar and Hans 2007). Formally, the judge determines the law and the jury applies that law to the facts as it finds them. However, in practice, the distinction between applying the law and finding facts is not always so clear.

A sketch of the jury trial reveals a handful of basic steps, starting with selecting individuals from a pool of persons summoned to serve on the jury (National Center for State Courts Jury

Topic Page). Once the trial is underway, the jury listens to the evidence presented and will alone make the decision about whether there is enough evidence to support a finding of guilt. If the defendant is found guilty, in some jurisdictions and some types of trials, the jury may be involved in sentencing the defendant.

Jury Selection

> A large government-funded organization is seeking 12 thoughtful people for group decision-making. Applicants must be willing to put their regular lives on hold for a year or more in return for low pay and zero benefits. Out-of-pocket expenses will not be covered (Butler 2007 p. B1).

This imaginary job advertisement for jury duty captures the challenges that confront judges and other tenured members of the court in their search for citizens to serve as jurors, particularly in lengthy trials. Selecting members of the jury, however, may be the most important stage of the trial because the jury's ability to inject a democratic perspective into the legal system depends on its composition. A jury drawn from a representative cross-section of the population is in a better position to express the full range of the community's views. Yet, for much of its history, juries have reflected the more privileged and elite segments of their communities. Men, whites, and those with more education, higher occupational status, and higher incomes were overrepresented on juries compared to their numbers in the population.

In the last several decades, jury commissioners have taken advantage of technological advances to reform the summoning process in order to achieve more representative groups of prospective jurors. Commissioners combine lists of potentially eligible residents and send out summonses for jury duty to residents living in geographically diverse areas. Some jurisdictions even use targeted replacement mailings to ensure representation from parts of the community that tend to have lower return rates. Governments have raised the compensation for jury service, especially for long trials when jurors may be away from their employment for a substantial period of time. While in many countries these methods still fall short of producing fully representative groups, they are a decided improvement over the past when jurors served based on personal recommendations or were handpicked by jury commissioners.

The process of summoning a large group of eligible jurors begins the trial. While the specifics vary across jurisdictions, typically both the judge and the attorneys are able to remove individuals before finalizing the jury that will ultimately decide the case. In the United States, the judge has the ability to decide "for cause" challenges to eliminate clearly biased jurors, whereas attorneys have a limited number of "peremptory" challenges that they may exercise without providing reasons. Other countries such as England and Wales do not permit peremptory challenges. The Korean system allows each side five peremptory challenges and an unlimited number of challenges for cause (Park 2010). In Canada, subgroups of jurors determine whether other jurors are impartial or should be removed for cause.

Providing a mechanism for removing apparently biased individuals from the jury helps to ensure the integrity and quality of the jury's decision. Nevertheless, opponents criticize the jury selection process, with some pointing to evidence that race and other impermissible factors infect it, and others doubting the ability of judges and lawyers to identify and remove biased jurors. In the United States, a robust field of jury consulting has developed to assist lawyers in selecting jurors; it is employed predominantly in high-profile criminal trials and with wealthy defendants (Tanovich et al. 1997; Vidmar and Hans 2007).

Juries as Fact Finders

During the trial, the jury acts as a fact finder, working through witness testimony, exhibits and the lawyers' arguments in order to uncover the facts of the case. Researchers interested in how juries process truncated and conflicting sketches of events use several theories to explain the jury's fact finding role. The most common theory is the "story model," in which jurors arrange the evidence into a narrative account. Other theoretical explanations focus on the psychological processes of anchoring and adjustment. Jurors anchor on their initial perception but adjust their opinion about the probability of the event having occurred as new evidence is presented at trial. A final theory is that jurors integrate information by weighing relevant information and combining discrete pieces of evidence. Ideally, jurors would evaluate the strength and weakness of evidence using relevant and impartial criteria. However, research indicates that factors unrelated to the strength of the evidence influence some juries. In the aggregate, jurors are more likely to attribute accuracy to eyewitness identification when a witness displays confidence, even though this is not an indication of reliability. As well, juries tend to give more weight to identifications made by police officers.

At its best, jury decision making embraces many of the features of deliberative democracy. Members of the jury hold a diverse set of viewpoints which can be brought forward and evaluated through reason-based discussion. Open discussion helps ensure that trial evidence is thoroughly evaluated, that rival explanations are examined, and that mistaken recollections are corrected. Through this process of deliberation and discussion, the jury reaches binding conclusions on the defendant's guilt or innocence. Empirical studies on group decision making confirm some but not all of the predictions of deliberative democracy theory. Studies show that groups outperform individuals in recalling facts, in correcting errors, and in pooling information. However, studies also show that during the deliberative process, once jurors are made aware of the majority view, they will tend to move in that direction, regardless of whether the view is to convict or acquit. Other studies indicate a leniency bias, where jurors in a numerical minority arguing for acquittal have more impact than a minority arguing for conviction. Either way, as the majority increases, so does the pressure on the minority to conform to the majority view.

Several studies reveal that judges and juries agree on the verdict in a substantial majority of cases, and that when judges and juries disagree, it is for reasons other than the difficulty of the evidence. A recent study of the first 3 years since the introduction of jury trials in South Korea confirms these findings (Kim et al. 2013). While in most cases judges and juries agree on the verdict – 91.4 % of the time in South Korea – when they do disagree, juries tend to be more lenient than judges, perhaps because they have a more generous interpretation of reasonable doubt, because they sympathize with the defendant, or because they disagree with the law. Judges' professional training and prior experience deciding criminal cases are also factors. The sizeable overlap between professional judges and juries should assuage concerns about jury incompetence. On the other hand, the research lends some support to beliefs that juries are more generous to criminal defendants than judges sitting alone.

Juries instantiate civic duty and their work strengthens legal and political institutions. Jurors report higher regard for the institution following their trial experiences (Diamond 1993). In a national survey of over 8,000 former jurors in the United States, 63 % reported that they were more favorable about jury duty after serving. Other studies similarly show enhanced regard for the courts and for judges after jury service. Surveys of jurors in South Korea's new advisory jury system show positive views of the experience (Park 2010).

Moreover, there may be a direct link between jury decision making and interest in politics (The Jury and Democracy Project website). A multidisciplinary team conducted research on the salutary effects of jury service on civic participation (Gastil et al. 2010). An initial study of Thurston County, Washington, residents found that jurors who took part in deciding the verdict at trial voted more frequently in subsequent elections than people who were called for jury duty but were dismissed, were alternates, or were on hung juries that could not reach a verdict. A follow-up study conducted nationwide included a sample of more than 13,000 jurors. This latter study found that jurors who had been infrequent voters were more likely to vote after serving jury duty. In this case, it didn't matter whether the jury had reached a verdict or not.

The job of serving as a juror in a criminal trial does, however, come with a cost. Jurors experience stress, especially following lengthy trials or in trials where the offense and potential punishment is severe (Anand and Manweiller 2005). In a 1998 study, jurors identified several sources of stress, including deciding on the verdict, jury deliberations, disruption to daily routine and dealing with lurid evidence (National Center for State Courts, 1998). A 2001 Canadian study revealed similar sources of stress, including reaching a verdict and the deliberation process. Canadian juries present a particular challenge for researchers, as they are prohibited by law from disclosing any information about jury proceedings. The closed nature of jury proceedings in Canada acts as an additional source of stress for these jurors, who are not able to process potentially emotional or difficult information with others (Chopra 2002). The responsibility placed on juries is particularly acute in murder trials. Jurors play a critical role in bringing legitimacy to punishments that take life or deprive liberty for life. However, murder trials can be lengthy and the evidence presented can be disturbing. For example, the multiple-victim murder trial of Robert "Willie" Pickton, who was suspected in the deaths of 26 women in British Columbia, Canada, lasted just over 9 months (Butler 2007). In order to address these issues, some courts provide psychological, psychiatric, or social work services to jurors following trial (Anand and Manweiller 2005; Chopra 2002).

Juries at the Sentencing Phase

Even if a jury decides the verdict, it is not necessarily involved in sentencing. For most types of crimes in most US jurisdictions, professional judges have the sole responsibility for sentencing criminal offenders. In six US states, jury sentencing is an option in felony trials and defendants may choose either judge or jury sentencing. However, in capital cases, most US states give the decision on whether or not the defendant should be sentenced to death to the jury. In Canada, juries do not participate in sentencing offenders. Murder convictions carry automatic life sentences, and for other convictions, judges determine the sentence bearing in mind sentencing ranges in the Criminal Code, the codified principles of sentencing, and previous similar cases. However, when a defendant is convicted of first- or second-degree murder charges, the jury can make a nonbinding recommendation of how long the defendant should remain ineligible for parole. In South Korea, juries decide on a sentencing recommendation and submit it to the judges. Similar to Canadian parole recommendations, Korean jury recommendations on sentencing are not binding (Park 2010).

Judges and juries will experience the task of sentencing offenders differently in the United States. In many jurisdictions, judges consider statutory ranges, sentencing guidelines, and typical sentences in similar cases as they determine an individual defendant's sentence. In contrast, states often limit the information jurors receive about the offense, the offender, and sentencing guidelines. In Virginia, juries but not judges must impose the statutory minimum sentence. Ironically, jury sentencing may reduce the use of juries at trial, as prosecutors use the prospect of unfettered jury sentencing to encourage defendants to enter a plea before trial (King and Noble 2004).

Lay Judges and Mixed Courts

In countries that follow the civil law tradition, lay judges (alternatively, lay assessors) sit in mixed courts and tribunals with professional legally trained judges. Lay judges serve jointly with professional judges in many European countries, including Germany, Austria, Denmark, France, Finland, Hungary, Italy, Norway, Poland, and Sweden, and in post-Soviet countries including the former Czechoslovakia and in countries of the former Yugoslavia (Hans 2008; Kutnjak Ivković 2007). Several new systems of lay participation use the mixed court model: for example, Japan's mixed court of lay citizens and professional judges, *Saiban-in Seido*, and Argentina's mixed court in Córdoba (Corey and Hans 2010).

Selection of Lay Judges

Similar to the way that jurors are selected, lay judges are ordinary citizens selected from a list of potential candidates who fulfill age, citizenship, literacy, or education requirements. For example, in Italy, lay judges must be Italian citizens of "good moral conduct," between the ages of 30 and 65 years, with at least a high school diploma (Catellani and Milesi 2006). Citizens with specialized skills may be sought for certain cases, such as offenses involving juvenile offenders. In those cases, lay judges may be expected to have parenting experience or a degree in education, psychiatry, social work, or sociology (Kutnjak Ivković 2007; Catellani and Milesi 2006). Some countries explicitly prohibit selection from among occupations with legal education or experience, such as professional judges, prosecutors, attorneys, or police officers; whereas in South Korea, soldiers, police officers, and firefighters are exempted from serving as lay judges because of the essential nature of their jobs.

Typically, a presiding judge or a commission of the court reviews lists of randomly selected candidates. Candidates who meet the legal requirements are appointed or elected and will serve for a period of time (Malsch 2009). The process of selecting lay judges may be more or less democratic and transparent. Citizens may elect lay judges to serve and sit on trials intermittently throughout a period of several years (Kutnjak Ivković 2007). On the other hand, in some countries, important members of the community such as mayors or municipal commissioners appoint potential candidates (Kaplan et al. 2006). To address transparency concerns, the president of the court in Italy posts lists of prospective lay judges in public places for members of the public to review.

In reality, the selection process can be political and can preference certain segments of the population. In Norway, a nomination committee picks candidates from among registered political party members. Some procedures or qualifications will favor middle-class citizens, leading to a disproportionate number of middle-class lay judges. And in countries with a large immigrant population, citizenship requirements may create tribunals that do not represent the population.

Lay Judges at Trial

The numbers of lay and professional judges who sit at trial differs across the various jurisdictions and often corresponds to the severity of the case, with larger tribunals sitting for more serious cases. In Japan, six lay judges and three professional judges decide guilt and sentencing in serious felony cases (Fukurai 2011), whereas in Poland, only three panel members try less serious criminal cases. Other countries have different combinations of lay and professional judges depending on whether the crime is a lesser crime or a more serious felony offense. For example, in Germany, two lay assessors and one professional judge sit for most criminal cases, whereas two lay and three professional judges try the more serious crimes. In South Korea, five to nine lay judges sit in mixed courts, with nine lay judges participating in cases where the defendant could receive the death sentence or life imprisonment. And in Italy, six lay judges and two professional judges hear cases where the defendant could be incarcerated for at least 24 years or life or where the crime is against the state (Catellani and Milesi 2006), whereas a law graduate appointed as an honorary judge for a term of 4 years can decide misdemeanor criminal cases. As in some of the panels noted above, lay judges may outnumber professional judges.

In many mixed court systems, lay judges decide all of the factual and legal issues in conjunction with the professional judges. This is not the case, however, in Japan. There, only the professional judge has the authority to determine questions of law and procedure. It is typical for a professional judge to control the trial. For example, a presiding professional judge may determine the trial date and summons the defendant and witnesses. In certain courts, the presiding judge will examine the witness before other members are allowed to speak. Lay judges can usually take an active role in the trial if they are so inclined. They may examine the evidence and question the witnesses including the defendant, either directly or indirectly through the professional judge.

The mixed court trial will work best when lay judges actively participate. Research indicates that professional judges, the presiding judge in particular, have a role to play in facilitating active lay participation. Lay judges are more likely to ask questions and participate in deliberations when encouraged to do so by professional judges. Lay judges are also more likely to be active during the trial and deliberations when they have special expert or technical knowledge that relates to the issues at trial. These were the findings of observers of mixed courts in Croatia, who noted that while expert lay judges were active at trial, regular lay judges asked questions only infrequently (Kutnjak Ivković 2007). This research also found that the legal professionals, the lawyers and judges, had more respect for the expert lay judges.

Access to the case file or dossier affects the extent to which lay judges are able to do the work of assessing the evidence at trial. Germany prohibits lay judges from reviewing the case dossier and in France, only the presiding judge has access to the dossier (Hans and Germain 2011). Studies of mixed courts in Poland found that even though lay judges were allowed access to the dossier, most did not read the file. Access to the case dossier, laudable if one seeks equality of information between lay and professional judges, is nonetheless controversial in that it is prepared by the prosecution and often contains potentially biasing information such as a defendant's criminal record. Moving from a dossier-based trial to oral presentations, as was done in Japan, tends to equalize the information available to lay and professional judges.

Several jurisdictions have formal rules about the deliberation and decision-making processes to ensure full and fair contributions by citizen participants. In Japan, the judgment must be agreed upon by a majority of the panel, with at least one citizen and one professional in the majority. Norwegian lay judges must consent to the decision written by the professional judge (Malsch 2009). Other jurisdictions may provide for lay judges to vote on guilt or innocence before the professional judges give their verdicts. Research has shown, though, that even with formal requirements, lay judges often agree with professional judges, and rarely use their larger numbers to outvote them. A study in Sweden found that lay judges outvoted the professional judges in only 1–3 % of all criminal cases. In cases where lay judges disagree with professional judges, it is the lay judges who are likely to modify their opinion to resolve the disagreement (Kutnjak Ivković 2007).

Nevertheless, lay participants tend to be positive about their experiences, just as jurors usually are. More than 85 % of Croatian lay judges had positive opinions about their participation. And in a study of Japanese citizen participation, 94 % of the lay judge respondents reported having a positive experience. Lay judges believe they have a substantial and beneficial impact on verdicts. A study of German lay judges indicates that most respondents felt that the court would have decided differently "in a few cases" and only 20 % responded that the court's decision would have been the same without the participation of lay judges.

Even though lay judges serve at the trial level, in most countries, appeals are decided by panels of professional judges only. Some exceptions include Sweden, which uses lay judges at the appellate level, but with a greater number of professional judges sitting on the panel, and France and Italy, which allow lay judges to serve in appeal courts. Lay judges can also sit on certain post-conviction reviews in Germany.

Sentencing Circles

Another way that laypersons participate in the criminal trial is through sentencing circles (Goldbach 2010). In Canada, sentencing circles modeled on Aboriginal healing circles are being used at the sentencing phase of a criminal trial for adult offenders and in juvenile diversion programs, as a way to encourage Aboriginal participation in the criminal justice system. Sentencing circles allow for a more restorative justice approach to sentencing, one that focuses on moral growth, constructive resolution of differences, and empowerment of individuals and the community to take responsibility for harm done.

The Canadian Department of Justice funds approximately 275 community-based justice committees which provide all levels of criminal justice support. In these communities, the justice committee works with the court to decide who will participate in sentencing circles and to identify a local community leader to act as keeper of the circle. In most cases, criminal justice participants, including the judge, the Crown prosecutor, defense counsel, the court reporter, the offender, the victim, and their respective families, form an inner circle. This inner circle can also include probation officers, court workers, youth workers, or police officers. Surrounding that circle is an outer circle of friends, relatives, and interested members of the community. Participants develop recommendations for sentencing and present those to the judge. The procedure can be lengthy and demands substantial commitment from victims, defendants, and the community.

Aboriginal communities participate in sentencing circles to construct sentences for such crimes as aggravated assault, assault causing bodily harm, robbery with violence, criminal harassment, breaking and entering, and arson. Judges are less likely to allow a request for a sentencing circle following conviction of murder or manslaughter. The use of sentencing circles is controversial, particularly in cases of domestic or sexual assault (Dickson-Gilmore and La Prairie 2005). However, this turn to community sentencing reflects a trend in Canada and in other jurisdictions encouraging greater victim and community involvement in all parts of the criminal justice process, as well as a move away from incarceration and retributive justice toward more restorative sentencing approaches. For example, in Canada, the form that the punishment takes following a sentencing circle may involve a conditional sentence, where the offender serves the sentence in the community and undertakes to fulfill certain conditions such as doing community service or enrolling in drug or alcohol treatment programs.

Controversies

Sentencing circles, mixed courts, and juries share similar problems and are confronted by overlapping critiques, which are in large part due to apprehension over lay persons participating in criminal trials. Concerns about lay competency to grasp complicated legal and factual issues are common to both jury and mixed trial systems. Similarly, both juries and lay judges face obstacles in accessing the full array of information that is available to professional judges. Lay judges in mixed courts confront the additional challenge of having to overcome professional judges' control of the hearing and deliberation processes. On the other hand, independent juries are under pressure to make decisions about guilt or innocence on their own, and both jurors and sentencing circle participants must determine punishment without the benefit of the germane experiences or expert knowledge that professional judges possess.

Research offers some reassurance. Studies indicate that lay participants are able to understand complex issues and often match judges in the accuracy of their decisions. In addition, studies show that reforms of the trial system, such as allowing note taking, asking questions, mid-trial deliberations, and the use of notebooks to organize the evidence, can increase the quality of lay citizen fact finding. Other controversies, however, are less easily resolved. Even when procedures are enacted to ensure active participation, lay judges arguably play a minor role in mixed courts. And research confirms that juries are influenced by pretrial publicity about a case, by defendant characteristics such as criminal record, and by preexisting biases about particular types of crimes. It should be noted that professional judges are not immune to some of the same influences. However, judges are generally required to produce written reasons which are thought to protect against biased or arbitrary decisions. Following this line of thinking, in 2009, the European Court of Human Rights (ECtHR) overturned a jury conviction in Belgium, finding that the jury's failure to provide reasons violated the defendant's right to a fair trial under the European Convention of Human Rights (ECHR). To meet human rights requirements, jury verdicts must be justifiable in some fashion (Thaman 2011). Although the Grand Chamber of the European Court observed that the decision was not a general indictment of jury systems, merely an objection to the jury trial procedure in the specific case, it remains to be seen whether this will trigger a widespread call for juries to produce written reasons in Europe.

Two additional controversies that have not yet been discussed are worth reviewing. The first is the problem of wrongful or erroneous convictions that become apparent when exculpatory DNA or other evidence comes to light. Juries in particular are at risk of being portrayed as error-prone and overly eager to convict when it comes to attributing blame for the conviction of an innocent defendant (Vidmar and Hans 2007). Thaman (2011) suggests that requiring juries to give reasons for their decisions would protect against convicting innocent defendants. Yet research shows that the most frequent cause of wrongful conviction, found in roughly 75 % of the cases, is mistaken eyewitness identification. A second important cause is false confessions. Thus, in the overwhelming majority of cases where a defendant is cleared by DNA evidence, the contributing factors relate to the evidence presented, not the jury's decision-making ability. Research through the Innocence Project in Toronto, Canada, illustrates systemic pressures at the trial and investigation stage, either because a particular case has a high profile or because of other institutional pressures (Martin 2002). This research suggests that a pressure to convict creates a bias in favor of building a case as opposed to solving the crime, which in turn shapes how police gather evidence and may even lead officials to disregard or suppress exculpatory evidence.

A second troubling controversy relates to capital punishment cases. Yet here too the problems are largely located in trial and court procedures as opposed to the fact of lay participation. In the early 1970s, the US Supreme Court struck down all state death penalty statutes because of evidence of arbitrariness, inconsistency, and racial bias in capital punishment decisions. Unfortunately, at that time, juries were given little guidance as to how to arrive at the decision to order a death sentence. Presently, state statutes require that juries consider or find certain aggravating circumstances before a death sentence can be ordered, reducing the impact of discretionary decision making.

Jury selection in death penalty cases, however, continues to be a problem for fair and equal decision making. Where the prosecution intends to seek the death penalty, only those jurors who are willing to impose a death sentence are considered to be fair and impartial. Yet research indicates that these individuals are more prone to convict. So-called "death-qualified" juries are more likely to believe the prosecution and have a general crime-control orientation which shapes their evaluation of the evidence. Analysis of capital case outcomes in the USA reveals that the sentence is related to the race of the victim, with black defendants who kill white victims more likely to be sentenced to death than other race of defendant-race of victim combinations. Nonetheless, the political, legal, and ethical justifications for including citizen input in a decision to sentence a defendant to death continue to be compelling. The US Supreme Court confirmed the importance of jury determinations in *Ring v Arizona* (2002). That decision held that in a capital jury trial, at a minimum, the jury and not the judge must decide the elements that make a case eligible for the death penalty, whether they are components of the criminal offense or additional aggravating factors.

Conclusion: Future Research

In the early decades of the twenty-first century, the role of citizen decision making in legal systems worldwide is at an interesting juncture. In many countries like the United Kingdom, the United States, and Canada, where juries have been a prime fact finder for centuries, the jury lives on as a potent symbol of democracy even as the number of jury trials declines. Jurors are largely positive about their experiences and studies indicate that participation on juries increases civic activity. Similarly, lay judges in mixed courts report being positive about their contributions, even though they tend to play secondary roles compared to the professional judges. Systematic research, from surveys of judges and juries in the United States in the 1950s to data collected on juries in South Korea between 2008 and 2010, supports the basic soundness of lay decision making in that it corresponds considerably to professional decision making. On the other hand, research also reveals that sources of bias continue to influence both lay and professional decision making in law.

More recent experiments with juries and mixed courts of professional and lay judges present new opportunities for democratic decision making in a host of countries. These new systems also offer potential for greater theoretical understanding and systematic scientific study. Research already underway in these countries will tell us much about what difference it makes to include laypersons in legal decision making. Because lay participation systems are introduced at particular points in time, researchers may be able to pinpoint their effects more precisely than has been possible in countries with long-standing jury and mixed court systems.

Although there is now research on how jurors and lay judges react to their experiences, how the presence of lay citizens affects criminal defendants is largely unexplored. Advocates of sentencing circles propose the use of restorative justice approaches based on the assumption that the community's active participation in sentencing assists in rehabilitating the criminal defendant. This reasoned connection between democratic, accessible legal institutions and defendants' experiences is theoretically robust and suggests several research questions. Are defendants more willing to accept the verdict or punishment when recommendations or decisions are made by lay members of the community rather than by professional judges? Does the inclusion of lay perspectives increase or otherwise affect perceptions of procedural justice? These questions connect the conduct of trials with justifications for criminal punishment and therefore deserve further investigation.

Related Entries

▶ Jury Impartiality in the Modern Era

Recommended Reading and References

Anand S, Manweiller H (2005) Stress and the Canadian criminal trial jury: a critical review of the literature and the options for dealing with juror stress. Crim Law Q 50:403–440

Butler D (2007) Trial by jury: the hours are long, the benefits are lousy and the work is highly stressful. The Ottawa citizen B1

Catellani P, Milesi P (2006) Juries in Italy: legal and extra-legal norms in sentencing. In: Kaplan MF, Martín AM (eds) Understanding world jury systems through social psychological research. Psychology Press, New York

Chopra SR (2002) Juror stress: sources, severity, and solutions. Ph.D. dissertation, Simon Fraser University

Corey Z, Hans VP (2010) Japan's new lay judge system: deliberative democracy in action? Asian Pacific Law Policy J 12:72–94

Diamond SS (1993) What jurors think: expectations and reactions of citizens who serve as jurors. In: Litan RE (ed) Verdict: assessing the civil jury system. Brookings Institution, Washington, DC

Dickson-Gilmore EJ, La Prairie C (2005) Will the circle be unbroken? Aboriginal communities, restorative justice, and the challenges of conflict and change. University of Toronto Press, Toronto

Fukurai H (2011) Japan's quasi-jury and grand jury systems as deliberative agents of social change: de-colonial strategies and deliberative participatory democracy. Chicago Kent Law Rev 86:789–829

Gastil J, Deess EP, Weiser PJ, Simmons C (2010) The jury and democracy: how jury deliberation promotes civic engagement and political participation. Oxford University Press, Oxford

Goldbach TS (2010) Sentencing circles, clashing worldviews, and the case of Christopher Pauchay. Illumine 9(2):53–76

Hans VP (2008) Jury systems around the world. Annu Rev Law Soc 4:275–297
Hans VP, Germain CM (2011) The French jury at a crossroads. Chicago Kent Law Rev 86:737–768
Kaplan MF, Martín AM, Hertel J (2006) Issues and prospects in European juries: an overview. In: Kaplan MF, Martín AM (eds) Understanding world jury systems through social psychological research. Psychology Press, New York
Kim S, Park J, Park K, Eom J (2013) Judge-jury agreement in criminal cases: the first three years of the Korean jury system. J Empir Leg Stud 10:35–53
King NJ, Noble RL (2004) Felony jury sentencing in practice: a three-state study. Vanderbilt Law Rev 57:885–964
Kovalev N (2010) Criminal justice reform in Russia, Ukraine, and the former republics of the Soviet Union. Edwin Mellen Press, Lewiston
Kutnjak Ivković S (2007) Exploring lay participation in legal decision-making: lessons from mixed tribunals. Cornell Int Law J 40(2):429–454
Lempert R (2007) The internationalization of lay legal decision-making: jury resurgence and jury research. Cornell Int Law J 40:477–488
Malsch M (2009) Democracy in the courts: lay participation in European criminal justice systems. Ashgate, Farnham
Marder NS (ed) (2011) Symposium on comparative jury systems. Chicago-Kent Law Rev 86(2):449–853
Martin D (2002) Lessons about justice from the "laboratory" of wrongful convictions: tunnel vision, the construction of guilt and informer evidence. UMKC Law Rev 70:847–864
National Center for State Courts (1998) Through the eyes of the juror: a manual for addressing juror stress. Willamsburg, NCSC Publication No. R-209
National Center for State Courts Jury Topic Page, available at http://www.ncsc.org/Topics/Jury.aspx
Park R (2010) The globalizing jury trial: lessons and insights from Korea. Am J Comp Law 58:525–582
Tanovich DM, Paciocco DM, Skurka S (1997) Jury selection in criminal trials: skills, science, and the law. Irwin Law, Concord
Thaman SC (2011) Should criminal juries give reasons for their verdicts?: The Spanish experience and the implications of the European court of human rights decision in Taxquet v Belgium. Chicago Kent Law Rev 86:613–668
The Jury and Democracy Project website, available at http://www.la1.psu.edu/cas/jurydem/
Tocqueville AD (1945) Democracy in America (trans: Phillips B, Reeve H, Bowen, F). A. A. Knopf, New York. French edition: Tocqueville AD (1835) De la démocratie en Amérique. C. Gosselin, Paris
Vidmar N (2000) World jury systems. Oxford University Press, Oxford
Vidmar N, Hans VP (2007) American juries: the verdict. Prometheus, Amherst

Jury Decision Making and Eyewitness Testimony

Amanda S. Nicholson, Angela M. Yarbrough and Steven D. Penrod
Department of Psychology, John Jay College of Criminal Justice, The City University of New York, New York, NY, USA

Overview

Jurors are assigned the arduous task of examining and processing copious amounts of evidence – in light of their legal instructions – to determine an appropriate verdict. While jurors do a relatively good job at sorting through the evidence and the law, the complex nature of evidence may lie outside jurors' "common knowledge," and additional education may aid jurors as they process such information. This is especially true in the domain of eyewitness testimony. Eyewitness testimony is extremely influential despite its potential to be unreliable. Eyewitnesses may appear very confident in their identification of the perpetrator, yet be completely mistaken. Indeed, over 75 % of wrongful convictions overturned due to DNA testing have been linked to faulty eyewitness identifications. Unfortunately, traditional safeguards, such as cross-examination of eyewitnesses, result in little improvement in jurors' ability to discriminate between accurate and inaccurate eyewitness identifications. This has led the courts to establish additional safeguards against wrongful convictions based on faulty eyewitness identifications including the use of expert testimony and detailed instructions on how to evaluate eyewitness testimony.

Jury Decision-Making Research

Understanding how jurors make decisions has been a topic that has interested social scientists for the better part of a century. Legal decision making has been studied by examining both group-level deliberations and individual juror decisions. At the group level, deliberation styles have been characterized as either evidence or verdict driven. Evidence-driven deliberations typically entail reconstructing the chain of events and focusing on the evidence presented before determining a verdict. On the other hand, deliberations that are verdict driven tend to be shorter and involve an immediate vote followed by discussion that is centered on how the evidence supports the verdict options. Unfortunately, deliberations in general do little to remedy mistakes or improve evidence and instruction comprehension. However, 12 member juries (as opposed to six) appear to result in more diverse jury composition, more accurate recall of trial testimony, and increased satisfaction among jury members (Saks and Marti 1997).

Because of the cost and time required to conduct research involving juries, the bulk of legal decision-making research focuses on individual jurors. One of the most popular cognitive theories is the Story Model, which takes into account individual jurors' attitudes and beliefs. Jurors proceed through three stages during the course of a trial. First, jurors develop one or more stories based on evidence presented during the trial that are supplemented by personal knowledge of similar events and individual beliefs and attitudes. Second, jurors are provided with judge's instructions that explain the verdict options available to them and the required elements for a finding of each verdict. Finally, a categorical decision is made based on the best match between a chosen story and verdict options available. Story construction is the focal point of this theory and is what will determine one's verdict choice.

Broader cognitive theories have also been applied to juror decision making. For example, dual process theories propose that people process information more or less thoroughly depending upon an individual's motivation and ability to understand such information. Jurors may be overwhelmed by large amounts of information. For example, expert testimony often entails complex information that is difficult to understand. When jurors defer to heuristic cues when evaluating the testimony. Heuristic cues, such as source attractiveness or length of testimony, do not directly contribute to the overall quality of the argument. Jurors should be relying on the clarity and strength of the experts' argument, but if they are unable to assess these features, they may rely on more superficial factors. An eyewitness expert, for example, may present scientific research to the jury outlining what factors may increase or decrease the reliability of eyewitness identifications. If this research is too difficult for jurors to understand, or so lengthy that they lose motivation to process the information, they may not absorb the message the expert is trying to present. However, if the expert presents information in a way that is understandable and interesting to the jurors, they will be more likely to systematically process the information and use it to assist with their evaluation of the eyewitness testimony presented at trial.

Eyewitness Research

Eyewitness testimony can be powerfully persuasive evidence. To the detriment of innocent defendants, such testimony is often damning. Extensive research has determined a variety of factors that may explain how misidentifications can occur.

Research on factors influencing the accuracy of eyewitness identifications has examined two types of variables: *estimator variables* and *system variables*. *Estimator variables* are situational variables associated with the crime, eyewitnesses, and perpetrators that can only be considered after the fact to evaluate the quality of the identification. *System variables* are variables related to police procedures dealing with lineup administration. Both system and estimator variables can impact the reliability and quality of eyewitness identifications and their accuracy.

Estimator variables are outside the control of the criminal justice system. For example, the viewing conditions experienced by the witness when they saw the perpetrator can influence the ability of the witness to later make an accurate identification. Conditions include lighting at the time of the witness' viewing, the witness' distance from the culprit, and the length of time the witness had to view the culprit.

A number of witness characteristics can impact the reliability of eyewitness testimony aside from any situational factors. For example, young children and the elderly appear more likely to make a mistaken identification compared to young adults. In addition, cross-race identifications are less accurate than identifications of culprits who are members of the same race as the witness (Meissner and Brigham 2001). People who witness crimes under high stress are less likely to make accurate identifications than witnesses who are not under stress (Deffenbacher et al. 2004). Furthermore, memory for an event may be limited when a weapon is present as attention is focused on the weapon and shifted away from the perpetrator (Steblay 1992). Lastly, the amount of time between witnessing the event and identifying the individuals is another important consideration in the accuracy of eyewitness testimony as memory deteriorates quickly for details of the event and perpetrator.

During the identification procedure, there are several important factors that can act as indicators of accuracy. For example, the speed at which a witness makes an identification is associated with witness reliability. Quicker identifications tend to be more accurate (Wells and Olson 2003). Perhaps the most important research regarding estimator variables is on eyewitness confidence. Although people commonly use the confidence level of an eyewitness to evaluate the reliability of eyewitness testimony, confidence is not highly correlated with accuracy (Sporer et al. 1995). In some situations the confidence of a witness can be a better predictor of accuracy, as when the confidence rating is taken directly after an identification. However, the most important manner in which the confidence-accuracy relationship is of concern is in relation to how confidence can be impacted by confirmatory feedback (Bradfield et al. 2002) or by a biased lineup (Charman et al. 2011). Confidence is malleable and susceptible to suggestive procedures used by lineup administrators. For instance, witnesses express higher levels of confidence when they receive feedback that they have correctly identified the suspect than when they do not receive this feedback (Bradfield et al. 2002). This occurs regardless of whether the witnesses' identification was accurate or not. Thus, if an identification is inaccurate and the witness receives positive feedback, they will appear more confident to a jury, but actually be mistaken about the identification.

System variables are typically under the control of the criminal justice system. Most often, these variables refer to procedures employed by the police when administering a lineup. There are a variety of biases related to lineup administration, specifically lineup composition, instructions, and presentation method in addition to the behavioral influence of the administrator that has the potential to cause suggestive lineups.

Foil bias occurs when the "foils" or the members of the lineup that are not the suspect (also known as "fillers") do not match the original description provided by the eyewitness (Charman et al. 2011). A distinction must be made between choosing foils that are similar to the suspect and those that are similar to the description provided by the eyewitness. In order to create a fair lineup, the police will often choose foils that look similar to the suspect. While this seems like a reasonable way to make sure that witnesses do not disproportionately identify suspects because they "stand out" in an array, there are several reasons why this matching strategy may be a problem. If foils are chosen based on their similarity to the suspect, the question arises about how similar the foils should be. If a witness is presented with a lineup consisting of almost identical individuals, the lineup task will be too hard, and there is the potential for a large number of false rejections or errors. When a witness provides the police with a description, they are not providing the police with an exact picture of the suspect. The witness may mention the suspects' hair color, eye color, height, weight, or distinctive tattoos or scars. If foils are chosen based on this short list of characteristics, the members of the lineup will differ from each other on a variety of characteristics the witness did not describe to the police. This raises the chances that the eyewitness can correctly identify the perpetrator – but not as a result of suggestiveness created by the police. According to a survey of experts in eyewitness research, 71 % felt that research on foil bias was consistent enough to present in court (Kassin et al. 2001).

Instruction bias is another type of bias that can occur with lineup administration. When using *biased instructions*, the lineup administrator does not explicitly instruct the witness that the suspect may or may not be in the lineup. Failing to provide this information implies that the suspect is present in the lineup, which has the potential to put pressure on the witness to make a choice from the lineup. *Unbiased instructions*, on the other hand, inform the witness that the suspect may or may not be present in the lineup. They also inform the witness explicitly that if they are uncertain about the identity of the suspect, they have the ability to reject the lineup. These instructions have been demonstrated to result in more accurate identifications (and particularly reduce mistaken identifications) as compared to biased instructions (Steblay 1997).

Presentation bias in lineup administration deals with how the various photographs in a lineup are displayed to the witness. In *simultaneous lineups*, all the photographs are displayed to the witness at the same time. The witness may be inclined to compare the photographs and choose the person who most closely resembles their memory of the culprit. However, this "relative judgment" procedure increases the likelihood of a false identification when the perpetrator is not in the lineup. *Sequential lineups* involve showing the witness one photograph at a time. This forces the witness to compare each face to their memory and use an "absolute judgment" when deciding each time whether the photograph is the culprit they remember or not. The sequential lineup presentation procedure results in fewer false identifications with only a modest reduction in the number of suspect identifications – and those lost identifications are likely the result of reduced guessing by witnesses (Wells et al. 2011).

Double-blind administration of lineups is another important method for reducing inaccurate eyewitness identifications. A number of researchers have recommended this method for improving the accuracy and eliminating biases associated with lineup administration (Greathouse and Kovera 2009). *Double-blind administration* of a lineup means that the person administering the lineup to the witness does not know who the suspect is. If the administrator is aware of the identity of the suspect, they may communicate this information unconsciously to the witness through subtle means (i.e., hesitations, pauses, smiles). They also may provide biased feedback to the witness that can affect confidence judgments of that witness going forward. If they are aware of the suspects' identity, they may inform the witness they were correct in their identification, which has the potential to increase future confidence assessments of that witness. Double-blind procedures can eliminate these biases and prevent the possibility that the lineup administrator intentionally or unintentionally communicated their belief in the suspects' guilt during the identification procedure.

How familiar are jurors with the factors that influence eyewitness performance? In order to answer questions about how much lay jurors know and understand about factors that affect the reliability of eyewitness identifications, Desmarais and Read (2011) performed a meta-analysis of various surveys concerning lay knowledge of eyewitness issues. Although laypeople understood more about factors affecting the accuracy of identifications than expected, the authors still concluded that "75 % of the eyewitness factors reviewed herein are 'beyond the ken' of potential jurors" (Desmarais and Read 2011, p. 209). They then concluded that jurors would be assisted in their decision making by expert testimony on these factors. These findings suggest that one way to improve the evaluation of eyewitness testimony by jurors should be to educate them about factors affecting the reliability of eyewitness testimony.

The Efficacy of Traditional Safeguards

Given the various issues with eyewitness identifications, there are obvious concerns about the impact eyewitness testimony will have on jurors in court. To help jurors make sound verdict decisions when eyewitness testimony is involved, the courts have devised a number of methods for sensitizing jurors to factors affecting eyewitness identification accuracy. There has been some debate about the respective efficacy of these various safeguards, but overall it appears these safeguards are not as effective as the courts intended.

Some potential safeguards that occur before a trial begins are *jury selection* and *voir dire*. Before a trial, both the defense and the prosecution have the opportunity to remove jurors from the courtroom that may be hostile to their case or biased on some issue relevant to the trial. This procedure typically involves the judge, prosecutor, and defense attorney interviewing potential jurors to determine whether they are biased. If a potential juror is biased on an issue relevant to the trial, the defense or prosecutor may bring this to the attention of the judge. It is then up to the judge to determine whether they will remove the juror from participating in the jury by exercising a *for-cause challenge* or whether they will attempt to *rehabilitate* that juror and remove their biases. The judge in the case may then ask the juror if they can make a decision about the outcome of the case putting aside their particular biases. The judge will then determine, based on that juror's response, whether they believe that a juror can be rehabilitated. If the judge thinks that juror cannot put aside their biases and decide the case in an unbiased manner, the juror will be dismissed. If the judge does not dismiss a juror, the defense or prosecution may excuse the juror using one of their limited number of *peremptory challenges*.

Another potential safeguard against erroneous convictions is cross-examination of witnesses. Cross-examination is considered the traditional method to prevent unreliable evidence from impacting jurors. It is expected that whatever weaknesses exist in the prosecutor's case will be exposed by the defense attorney through cross-examination (and vice versa). Eyewitnesses are like other witnesses in that they can be questioned by the opposing counsel. In order for cross-examination to be effective, both attorneys and jurors need to possess some degree of knowledge about eyewitness issues. An attorney must know what factors impact the reliability of eyewitness identifications in order to identify such weaknesses in the witness' testimony. Yet describing eyewitness problems for the jury is fruitless if jurors are not sensitive to these issues. Jurors may have difficulty believing information that is contrary to beliefs they have already developed about eyewitnesses, especially if that information is provided to them by a potentially biased defense attorney.

The adversarial nature of cross-examination renders it ineffective as an educational tool for jurors, thus creating a need for some other way to disseminate information about eyewitness issues to jurors. The final two safeguards – *judicial instruction* and *expert testimony* – involve expert advice. *Judicial instruction* is typically used to inform jurors about what evidence they should and should not use when making a decision. Often, a judge will issue an instruction informing jurors to ignore any media coverage of the trial and to avoid using information from media sources when making a decision. Judicial instruction may also be used to instruct jurors to ignore any inadmissible evidence from entering verdict decisions. Judicial instructions are largely ineffective in persuading jurors to disregard inadmissible evidence (Steblay et al. 2006). However, one study found that jurors were more likely to disregard inadmissible evidence if told that the evidence was inadmissible because it was unreliable, rather than if its inadmissibility was due to the fact that it was obtained illegally (Kassin and Sommers 1997). The results of this study suggest that judicial instructions informing jurors about reliability issues with eyewitness identifications could cause them to question the reliability of identifications.

Although standard pattern judicial instructions are regularly used in all cases, recently courts have been supporting the use of case-specific instructions on eyewitness testimony. These instructions would be in place to educate jurors on factors affecting the reliability of eyewitness identifications in a specific case. General eyewitness instructions have existed for

40 years – *US v. Telfaire* (1979). The *Telfaire* instruction was created by the courts to address the problems with eyewitness identifications, but this instruction has not been particularly effective in sensitizing jurors to factors affecting the reliability of eyewitness identifications. Revised instructions may be more effective in educating jurors. Whether such instructions will work is an open question – most research has found judicial instruction to have little to no effect on jurors' verdict responses. Furthermore, it is reasonable to predict judicial instructions may, for the most part, have a skepticism effect on potential jurors. If jurors are informed by the judge about the unreliability of eyewitness evidence, they may assume that the judge believes the identification to be unreliable. Despite these caveats, some courts (e.g., *New Jersey v. Henderson* 2011) are advocating for case-specific instructions that will focus on reliability issues at play in a given identification. It is possible that the proposed judicial instructions will assist jurors in evaluating the reliability of eyewitness evidence, but as yet, none have demonstrated success in sensitizing jurors to the differences between reliable and unreliable identifications.

The most strongly supported safeguard against eyewitness misidentifications affecting jury verdicts is expert testimony. Expert testimony can increase jurors' sensitivity to eyewitness issues. Experts on eyewitness research and human memory are most often hired by the defense to speak to the court and jury about the factors that impact the accuracy of eyewitness identifications. Expert testimony typically focuses on specific system and estimator variables relevant to the case at hand. Support for juror sensitivity has been mixed. However, given the low level of lay knowledge about eyewitness issues, as outlined in Desmarais and Read (2011), expert testimony may serve as a useful educational tool to help jurors understand eyewitness issues. Those authors recommend expert testimony as the most effective safeguard, highlighting that the educational component may make up for the relative lack of knowledge jurors have about eyewitness issues.

Current Controversies and Issues with Eyewitness Research

Research on juror decision making, particularly experimental jury simulations, has been criticized for a variety of reasons. These reasons include the low stakes involved in mock jurors' decisions, the lack of ecological validity, use of college samples, and the lack of deliberating groups. Of course, it is simply unethical and impractical to manipulate variables in actual court cases that could impact the outcome for research purposes. However, the quality and verisimilitude of trial materials used when conducting stimulation studies has continued to improve. Additionally, in a review of the existing literature, Bornstein (1999) found few differences between college student and community samples serving as mock jurors. Overall, the results of posttrial surveys of actual jurors tend to corroborate the findings of stimulation studies, thereby increasing its validity. Still, the courts have been slow to embrace the research despite much consensus among experts regarding the findings.

While most courts have been slow to rely on psychological findings, the New Jersey Supreme Court in *New Jersey v. Henderson* goes as far as any court ever has in embracing psychological research regarding eyewitness identification. The court in *Henderson* appointed a Special Master to hear testimony from experts on human memory and eyewitness identification. Drawing on the Special Master Report, the court exhaustively reviewed eyewitness identification research and the general acceptance of this research among experts.

The *Henderson* court decided to modify the existing criteria used to determine eyewitness identification admissibility in pretrial hearings. These criteria were based on *Manson v. Brathwaite* (1977) which resulted in a two-pronged test. First, courts determine whether suggestive police procedures were used to secure an identification. If suggestive action was taken, the reliability of the identification must be assessed by reviewing the opportunity of the witness to view the criminal at the time of the crime, the witness' degree of attention, the accuracy of his prior description of the criminal, the witness' certainty, and the time between the crime and the identification – criteria initially outlined in *Neil v. Biggers* (1972). *Manson* made it clear that identifications can still be admissible despite the presence of suggestive procedures when the identification is deemed reliable according to the five criteria. The defendant has the burden of demonstrating the identification was made under suggestive circumstances that the police improperly created *and* is unreliable based on the totality of the circumstances.

Extensive research conducted since 1977 has demonstrated that the criteria are incomplete and inadequate measures of reliability. In addition, three of the five reliability criteria in *Manson* are, themselves, influenced by the presence of suggestion (i.e., the witness' certainty, witness reports about opportunity to view, and witness reports about degree of attention). Recognizing these deficiencies, the New Jersey Supreme Court ruled that when an identification is challenged by a defendant, all relevant system and estimator variables should be reviewed, including suggestive factors that may not involve police action, such as lightning and distance. If the defendant is unable to demonstrate a strong likelihood of misidentification, the jury will be charged with the task of determining whether the identification was reliable based on the evidence presented in court. Because juries tend to overestimate the accuracy of eyewitnesses and are unfamiliar with the literature establishing the influence of multiple factors on identification accuracy, the court directed a committee to develop new jury instructions addressing case-specific eyewitness identification factors. Unfortunately, as noted above, the effectiveness of these instructions has not been empirically demonstrated.

Relative to New Jersey, the US Supreme Court in its review of *Perry v. New Hampshire* (2011) was not as receptive to eyewitness research. This was the first case involving eyewitness identification before the US Supreme Court since *Manson v. Brathwaite* in 1977. The Supreme Court affirmed the state court's decision that due process protection against unreliable identification evidence applies only to identifications made under suggestive circumstances that are caused by the police. In this instance, the suggestive circumstances were happenstance, involving a woman looking out of her window into a dark parking lot and identifying a single black male among a group of police officers from 100 ft away. Despite research that demonstrates jurors are unable to determine the reliability of eyewitness testimony, the Court ruled that eyewitness evidence is like any other evidence and is therefore at the discretion of the jury to determine its reliability.

Future Directions

As demonstrated by the large number of people who have been wrongfully convicted, the traditional safeguards used by the courts to prevent wrongful convictions based on eyewitness misidentifications have not fully sensitized jurors to factors affecting eyewitness identification accuracy. The recommendation of the court in *Henderson* to develop case-specific instructions regarding eyewitness factors is a step in the right direction, but it may not be enough.

The Innocence Project has made recommendations focused on improving system variables, including double-blind administration, improving lineup composition, recording confidence statements immediately after an identification, videotaping the entire lineup and identification procedure, and using sequential lineups. Law enforcement agencies across the countries are heeding these recommendations. In addition, the *Henderson* Special Master made several recommendations not accepted by the New Jersey Supreme Court, but worthy of future attention by the courts. First, the Special Master recommended mandatory pretrial hearings for all cases involving eyewitness identifications. Second, it was recommended that the initial burden be shifted from the defendant, who now needs to establish that suggestive procedures were present during an identification, to the prosecution – the party seeking to introduce identifications into evidence. The prosecution would then have the initial burden to prove an identification was reliable. This in essence treats eyewitness evidence like physical trace evidence.

Focusing the majority of recommendations on improving system as opposed to improving the decision-making ability of jurors leaves open the possibility that a confident eyewitness may outweigh any efforts to understand the circumstances that led to the identification. Yet, some jury reforms have been proposed. For example, researchers have called for a return to 12 member juries and requiring unanimity of verdicts to ensure more thorough decision making. In addition, expert testimony may be the best option in court for sensitizing jurors to factors affecting eyewitness identification accuracy. Future research should focus on how these experts can improve the decision-making ability of jurors.

Related Entries

- ▶ Behavioral Science Evidence in Criminal Trials
- ▶ Causes of Wrongful Convictions
- ▶ Estimator Variables and Eyewitness Identification
- ▶ Expert Witnesses: Role, Ethics, and Accountability
- ▶ Eyewitness Evidence: Historical Background
- ▶ Eyewitness Identification Evidence
- ▶ Eyewitness Research
- ▶ Interviewing Eyewitnesses
- ▶ Juries, Lay Judges, and Trials
- ▶ Jury Impartiality in the Modern Era
- ▶ Measuring Wrongful Convictions
- ▶ Overturning Wrongful Convictions and Compensating Exonerees
- ▶ Prosecution and Wrongful Convictions
- ▶ Scientific Evidence in Criminal Adjudication
- ▶ Scientific Evidence in Criminal Prosecutions
- ▶ Wrongful Convictions in the Inquisitorial System

Recommended Reading and References

Bornstein B (1999) The ecological validity of jury simulations: is the jury still out? Law Hum Behav 23:75–91

Bradfield A, Wells G, Olson E (2002) The damaging effect of confirming feedback on the relation between eyewitness certainty and identification accuracy. J Appl Psychol 87:112–120

Charman S, Wells G, Joy S (2011) The dud effect: adding highly dissimilar fillers increases confidence in lineup identifications. Law Hum Behav 35:479–500

Cutler B, Dexter H, Penrod S (1990) Nonadversarial methods for sensitizing jurors to eyewitness evidence. J Appl Soc Psychol 20:1197–1207

Deffenbacher K, Bornstein B, Penrod S, Kiernan McGorty E (2004) A meta-analytic review of the effects of high stress on eyewitness memory. Law Hum Behav 28:687–706

Desmarais S, Read D (2011) After 30 years, what do we know about what jurors know? A meta-analytic review of lay knowledge regarding eyewitness factors. Law Hum Behav 35:200–210

Devenport J, Cutler B (2004) Impact of defense-only and opposing eyewitness experts on juror judgments. Law Hum Behav 28:569–576

Greathouse S, Kovera M (2009) Instruction bias and lineup presentation moderate the effects of administrator knowledge on eyewitness identification. Law Hum Behav 33:70–82

Kassin S, Sommers S (1997) Inadmissible testimony, instructions to disregard, and the jury: substantive versus procedural considerations. Personal Soc Psychol Bull 23:1046–1054

Kassin S, Anne Tubb V, Hosch H, Memon A (2001) On the "general acceptance" of eyewitness testimony research: a new survey of the experts. Am Psychol 56:405–416

Luus CAE, Wells G (1991) Eyewitness identification and the selection of distracters for lineups. Law Hum Behav 15:43–57

Manson v. Brathwaite (1977) 432 U.S. 98

Meissner C, Brigham J (2001) Thirty years of investigating the own-race bias in memory for faces: a meta-analytic review. Psychol Publ Policy Law 7:3–35

Neil v. Biggers (1972) 409 U.S. 188

New Jersey v. Henderson (2011) 208 N.J. 208

Perry v. New Hampshire (2011) 131 S. Ct. 2932

Saks M, Marti M (1997) A meta-analysis of the effects of jury size. Law Hum Behav 21:451–467

Sporer S, Penrod S, Read D, Cutler B (1995) Choosing, confidence, and accuracy: a meta-analysis of the confidence-accuracy relation in eyewitness identification studies. Psychol Bull 118:315–327

Steblay N (1992) A meta-analytic review of the weapon focus effect. Law Hum Behav 16:413–424

Steblay N (1997) Social influence in eyewitness recall: a meta-analytic review of lineup instruction effects. Law Hum Behav 21:283–297

Steblay N, Hosch H, Culhane S, McWethy A (2006) The impact on juror verdicts of judicial instruction to disregard inadmissible evidence: a meta-analysis. Law Hum Behav 30:469–492

Steblay N, Dysart J, Wells G (2011) Seventy-two tests of the sequential lineup superiority effect: a meta-analysis and policy discussion. Psychol Public Policy Law 17:99–139

United States v. Telfaire, 469 F.2d 552 (D.C. Cir. 1979)

Wells G, Olson E (2003) Eyewitness testimony. Annu Rev Psychol 54:277–295

Wells G, Steblay N, Dysart J (2011) A test of the simultaneous vs. sequential lineup methods: an initial report of the AJS national eyewitness identification field studies. American Judicature Society, Des Moines

Jury Impartiality in the Modern Era

Nicole L. Waters[1] and Paula Hannaford-Agor[2]
[1]National Center for State Courts, Williamsburg, VA, USA
[2]Center for Jury Studies, National Center for State Courts, Williamsburg, VA, USA

Overview

The strength of the jury system in an adversarial system of justice depends on the impartiality of the jurors. Yet the rapid evolution of Internet-based communication technologies poses serious challenges to the traditional concept of juror impartiality. It is now possible for jurors to access virtually any piece of published information about pending cases in minutes, and the volume of potentially case-relevant information is growing exponentially. Many jurors have become accustomed to using these technologies to conduct research and communicate with friends and family. For some jurors, reliance on these technologies has become so ingrained that it would require conscious effort to refrain from doing so for the duration of a trial.

This entry discusses the notion of what it means for a juror to be and to remain impartial in the digital age. First, it focuses on the impact of new media on how jurors acquire and process information. Then it discusses the impact of new media on public perceptions of the justice system, especially in high-profile trials. Finally, it examines the viability of traditional judicial responses to the newly wired and media-saturated jury pool. It concludes with some sobering reflections about the ability of the justice system to keep up with these technological changes.

The Traditional Concept of Juror Impartiality

The strength of the jury in an adversarial system of justice is the impartiality of the jurors. Impartial jurors are those who are willing and able to consider the evidence presented at trial without preconceived opinions about the defendant's guilt or innocence, to apply the governing law as instructed by the trial judge, and to deliberate in good faith to render a legally and factually justifiable verdict. Traditionally, the process of identifying impartial jurors focused on voir dire during which judges and lawyers questioned jurors about their knowledge of the facts of the case, opinions about issues that might arise during trial, or life experiences that might affect how jurors perceive the evidence they would hear during the course of the trial. Once the judge and lawyers had removed biased jurors, the trial would begin. Except under very unusual circumstances, there was little risk that the selected jurors might lose their impartiality during the remainder of the trial.

The rapid evolution of various types of new media over the past two decades poses serious challenges to this concept of juror impartiality. Internet-based technologies now make it possible for jurors to access virtually any piece of published information about pending cases in minutes, regardless of when or where published. Many jurors have become accustomed to using these technologies to conduct research and to communicate with friends and family. For some jurors, reliance on these technologies for everyday tasks has become so ingrained that it would require conscious effort to refrain from doing so for the duration of a trial. As a result, judges and lawyers can no longer be confident either that a sufficient number of prospective jurors on any given panel will meet the traditional definition of impartiality or that the jurors selected for trial will remain so for the entire trial.

The volume of potentially case-relevant information that might jeopardize juror impartially is also growing exponentially with the proliferation of various types of news media including traditional media outlets as well as cable news organizations, online print media, and specialty blogs. The traditional news cycle involved at most daily updates, but some trials now receive continual, minute-by-minute, 24/7 news coverage as well as ongoing commentary and background information based on interviews with trial attorneys, litigants, witnesses, and even less central players such as coworkers, neighbors, and childhood friends. This level of media saturation exposes a larger number of prospective jurors to potentially prejudicial information about more upcoming trials than ever before in history, making it more difficult to select impartial jurors for trial and to maintain their impartiality throughout the trial.

Traditional approaches to minimizing these effects seem to be losing their effectiveness given the volume and intensity of trial information available to prospective jurors. Some of the most pressing concerns in contemporary jury system management are the impact of these technologies on juror decision-making and on public perceptions of the justice system.

The Impact of New Media on Juror Decision-Making

In a provocative essay published in the July/August 2008 issue of *Atlantic Monthly*, Nicholas Carr described how his use of the Internet seemed to be changing how his brain operates, especially his memory and his capacity for sustained concentration. The article provided an overview of how the development of various technologies – written language, the mechanical clock, the printing press, radio and television – all affected the brain's neural circuitry. For the most part, those changes had a positive impact on civilization, both increasing the scope of human knowledge and distributing it more widely. He then posited that the Internet, which he describes as an immeasurably powerful computing system, might be affecting a similar shift in human cognitive ability. Preliminary neurological studies suggest that Carr's insights may be quite accurate. If so, that prospect will have a profound impact on juror decision-making, especially how trial jurors receive and interpret information during the course of a trial.

Much of the existing scientific literature on juror decision-making is grounded in theories derived from cognitive psychology that individuals engage in schematic processing to interpret their environment efficiently and effectively. "Schemas" act as cognitive filters through which individuals identify people and situations quickly, according to familiar paradigms. For jurors, these schemas take the form of preconceptions and knowledge of the world that they use to construct narratives or stories from trial evidence and fill in missing details to increase the story's internal consistency and convergence with their world knowledge. This cognitive processing helps jurors assess the trial evidence for credibility, consistency, and relative importance. Contemporary researchers refer to this theory as the "story model" of juror decision-making. In more colloquial terms, jurors bring their common sense and community values to inform judgments about a criminal defendant's guilt or a civil defendant's liability for damages. During deliberations, jurors compare these individual narratives and, except in very rare exceptions, arrive at a consensus about the "correct" interpretation of the evidence and application of the governing law to produce a legally valid decision.

Historically, this process took place during trial as the lawyers presented each new piece of evidence to the jury through direct and cross-examination of witnesses. The question-and-answer format through which attorneys elicit oral testimony to support their respective theory of the case was originally intended to provide jurors with an unvarnished and neutral presentation of trial evidence. The format is exceedingly archaic and is almost never employed in other settings in which information is communicated to a lay audience. The jurors' task can be tremendously complicated insofar that it involves taking the individual bits of trial evidence and piecing them together into a coherent picture. In most trials, however, the relatively slow and methodical nature of the trial process, often interrupted by trial recesses and sidebar conferences between the judge and trial attorneys, provided ample time for jurors to reflect carefully on the evidence and make sense of the disparate pieces.

There are two significant implications of the changes wrought in neurological processing by increased use of new media. First, contemporary jurors are increasingly accustomed to the fast-paced and constant mode of transmission that one expects from handheld devices and Internet surfing in which readers jump from hyperlink to hyperlink, skimming materials for key nuggets of critical information without stopping to digest the entire webpage. In addition to reflexively seeking out information online with which to better understand the world, contemporary jurors are also accustomed to receiving constant updates in the form of e-mail and text messages, tweets, and notices from social networking sites that do not require active intent to acquire new information. They just arrive, unsolicited, on one's computer screen or smart phone with information formatted in the highly abbreviated style of headlines, sound bites, and bullet points. Communication in the Internet Age must conform to the "140 characters or less" requirement or risk losing the intended audience in the confusion or boredom of excessive detail and nuance. The traditional style of trial procedure is more and more likely to perplex and antagonize jurors who will have greater difficulty making sense of how its organizational framework presents disparate and detailed pieces of trial evidence.

Second, there is the possibility that contemporary jurors are cognitively either less reliant on or less confident in their collective common sense and community values and thus find it necessary to verify initial impressions about the evidence or to supplement it with external sources found online. Their cognitive schemas are no longer purely internal psychological constructions, but rather exist as an externalized collective schema in "the cloud" where they can be accessed with the click of a mouse. As the urge to use these technologies becomes stronger and the ability to do so becomes easier, judges and trial lawyers will find it increasingly difficult to block juror access to these potentially prejudicial sources of extraneous information so that jurors might maintain some semblance of impartiality until the evidentiary portion of the trial is complete.

At this time, it is uncertain how often jurors already access the Internet on a routine basis. In a preliminary study of the frequency of juror and jury use of new media, the National Center for State Courts (NCSC) found that sizeable portions of trial jurors reported interest in using new media to conduct research on case-related topics and to communicate with friends and family about their jury service experience. Although the vast majority of jurors in that study had daily, if not immediate, access to new media, none of them admitted to acting on that desire. That study involved a very small sample of trials; however, it is clear from court opinions and news stories discussing the problem of the "Googling juror" that the risk is not purely hypothetical. In a review of court opinions published between 1998 and 2010, Thompson-Reuters reported that at least 90 verdicts were challenged based on claims of Internet-related juror misconduct. One-half of those challenges occurred between 2008 and 2010. In 28 cases, civil and criminal, new trials were granted or verdicts overturned. Even where judges declined to declare a mistrial, in three-quarters of the cases, the courts held that Internet-related misconduct had occurred. Indeed, it is likely that the frequency of juror use of new media is much greater than written court opinions reflect as many such instances would not result in a written opinion. In fact, most would likely go undetected.

The crux of the dilemma for the justice system is the impending collision between the traditional view of juror impartiality and contemporary jurors' increasing reliance on new media to inform their decision-making. The traditional strength of the jury system rests on the assumption that the jury considers only evidence properly admitted at trial. Jurors take an oath at the beginning of the trial to render a "true verdict ... according to the evidence, without fear, favor, or affection, and ... governed by the instructions of the court." Intentionally seeking extraneous information about case-related topics is a clear violation of the juror's oath and can result in a mistrial or overturned verdict. As individuals increasingly rely on the Internet to access information to help navigate their environment and interpret the world, it will likely become ever more difficult to prevent them from doing so when serving as trial jurors. After all, jurors understand that jury service is a serious task that requires the greatest degree of attention and competence. It will become increasingly counterintuitive to jurors that they would violate a solemn oath by using the very tools on which they normally rely to inform their judgments in serious matters.

To a certain extent, trial courts have already accepted, and even embraced, a seismic shift in jurors' role in the trial process by adopting trial techniques (e.g., juror note taking, juror submission of written questions to witnesses, juror discussion of evidence before final deliberations) that facilitate active learning styles. Traditionally, it was assumed that juror passivity helped to maintain their impartiality. Contemporary empirical research confirms that enforced passivity does not significantly enhance impartiality and, in fact, can seriously undermine juror performance and satisfaction. The transition to the "active juror" model mirrors many trends in contemporary life in which individuals are encouraged to assume a more active role. In health care, financial management, and continuing education, for example, responsible and competent behavior is defined by a person's willingness and ability to undertake an active partnership with professionals to accomplish both personal and collective tasks. The key question in the context of contemporary jury service is whether Internet use is a legitimate tool to aid juror decision-making (as appears to be the case for increasing numbers of jurors) or a serious breach of juror impartiality that threatens the legitimacy of the jury's verdict.

The latter viewpoint predominates among contemporary trial judges and lawyers. Responses to this type of "juror misconduct" run the gamut from education and outreach to persuade prospective jurors not to engage in Internet use during trial and deliberations, to preventive measures intended to block juror access to the Internet, to punitive measures imposed on jurors who disobey direct orders to forego the Internet for the duration of the trial. Increasingly, informational booklets and orientation programs for prospective jurors emphasize the importance of not accessing the Internet during trial. Many judges and lawyers now question prospective jurors about their Internet use during voir dire to screen out jurors who indicate reluctance to adhering to prohibitions on juror use of new media. Jury instructions have become increasingly specific about prohibitions on various types of the online activities. Judges repeat these instructions more frequently throughout the trial to remind jurors of these prohibitions and to emphasize their importance. Hoping that jurors who understand the underlying rationale for the prohibition on Internet use will be more likely to comply, some judges also explain that extraneous information encountered online is not evidence and deserves no weight in the jurors' deliberations. Finally, some courts ban all forms of electronic devices from the courthouse or confiscate such devices from jurors during trial and deliberations. Ironically, some courts have proposed using technology to combat problems associated with juror use of technology including blocking electronic transmissions in courtrooms and jury deliberation rooms to prevent juror misconduct. When prevention is insufficient, judges are also becoming more willing to punish jurors for violating the admonition and to consider posttrial challenges to verdicts based on juror misconduct.

Impact of New Media on Public Perceptions of the Jury System

Because jurors are drawn from the community at large, they reflect the general social outlook and values of their communities. Indeed, one of the primary roles of the jury is to inject community values into judicial decision-making. Although, as designed, the voir dire process identifies and removes jurors who hold such strong opinions about case-specific issues that they could not serve fairly and impartially, those opinions will still be present in the public at large. This is particularly the case in high-profile trials that generate considerable media attention. One of the great ironies of contemporary society is the apparent disconnect between the jurors' trial and the public trial. Trial jurors will largely be isolated from ongoing media coverage of the trial. All the while, public sentiment may become even more inflamed over the course of the trial in reaction to evidence admitted at trial as well as media commentary on that evidence and non-trial information disclosed by the litigants, lawyers, and witnesses. In essence, a high-profile trial is actually two very different trials – one that the sworn trial jurors experience and one that the public observes as quasi-jurors, which can sometimes lead to very different conclusions about the appropriate verdict. Recent examples include O.J. Simpson's acquittal of murder charges, Michael Jackson's acquittal of child molestation charges, life sentences rather than the death penalty for Terry Nichols (coconspirator in the Oklahoma City bombing) and Zaccarias Moussoui (the alleged 20th hijacker in the September 11th terrorist attacks), and the acquittal of Casey Anthony on charges of murdering her 2-year-old daughter.

Different degrees of information presented to the public also result in a blurring of the line between news reporting, education, and entertainment, between fiction and reality. Much of what the public knows about what actually occurs in the courtroom is what the ever-merging news and entertainment outlets portray. While reports on trial events provide a glimpse into how the justice system works or does not work, most people do not have a realistic sense of what it is like to serve as a juror in an actual trial. People routinely report that their primary source of courtroom knowledge comes from television trial shows such as the *People's Court* and crime dramas such as *Law and Order*.

The various iterations of entertainment shows and news outlets affect the public's expectation about the justice system and jury verdicts. The number of law enforcement and forensic-based crime dramas on television (e.g., *CSI* and its numerous iterations) outnumbers the number of actual trial drama shows. Yet an underlying theme across all of them portrays a fast-paced trial that is resolved in an hour or less with justice unequivocally done. Cable television's 24-hours news coverage and shows such as *Court TV* provide what appears to the public as the inside story with all of the facts revealed including commentaries by so-called experts. Inherent in most jury trials, however, is the reality that trial evidence is often ambiguous, conflicting, and incomplete; the law articulated in jury instructions sometimes borders on incoherence; jury deliberations can be quite contentious; and jurors may nevertheless harbor some doubts (albeit not reasonable doubts) about a defendant's guilt even after returning a guilty verdict.

The public forum for hearing jury trials, as guaranteed by the Sixth Amendment, was seen by the founders as a safeguard for the defendant against abuses by the government. Freedom of the press under the First Amendment was intended to protect the people, to be the voice of the community. The tension between the two, spurred along by new media in the Information Age, contributes to a decline in the public's trust and confidence in the jury system. The courts depend on the jurors as representatives of their respective communities to provide legitimacy to the justice system. As such, central to the mission of the courts is a way to maintain the public's trust and confidence in trial by jury as an effective way to resolve disputes. When there is public outrage over a perceived injustice, especially in a notorious trial, the courts must work quickly and effectively to counter the public's doubt.

Courts now use many contemporary communications tools and techniques to make the justice system appear more accessible and more transparent. Some provide online access to court documents including filings and decisions. Others offer real-time video feeds of court proceedings. Many courts, especially in urban areas, now employ public information officers who are specifically tasked with communicating information about the court's mission and role in contemporary society to the public. Inherent in that task, however, is the paradoxical challenge that the culture of the court is fundamentally at odds with the societal culture that has developed with and in response to these new communication technologies. A recent study entitled "New Media and the Courts: Current Status and a Look at the Future," undertaken by the Conference of Court Public Information Officers, observed that courts rely almost exclusively on textual communication – written opinions and court orders – to speak publicly. This mode of public communication, detailing the proven facts and logic on which court decisions were made, underscored the message that the court's legitimacy rested firmly on the rule of law. Moreover, court communication is primarily hierarchical and unidirectional; opinions and court orders are intended as the final word to be obeyed. They are not intended as an invitation for further discussion except within the highly stylized procedures of a legal challenge to those orders.

In contrast, the Internet is a multimedia environment offering visual and audio formats to communicate in addition to traditional text. These technologies also are intended to be interactive and to encourage collective decision-making on the largest scale possible. While many court public information officers have made tremendous progress in incorporating some new media tools and strategies, at least to communicate non-case-specific information to the public, it is not clear that they will ever fully harmonize these two incongruous cultures without a radical reconceptualization of many of the fundamental principles of judicial independence and legitimacy.

The Continued Viability of Judicial Responses to Counter the Effects of Pretrial and Trial Publicity

High-profile trials cause the most difficulty by far for judges and lawyers in terms of how to mitigate the impact of pretrial publicity on prospective jurors. Trials can become the focus of intense media attention for a variety of reasons. Sometimes the litigants, witnesses, or victims are celebrities, such as in the O.J. Simpson and Michael Jackson trials. Sometimes the case involves particularly violent or heinous crimes that shock the community, including the Oklahoma City bombing trials, the Unabomber trial, and the Moussaoui terrorism trial. Sometimes the case raises controversial social or political issues, including the California Proposition 8 trial involving the constitutionality of same-sex marriage or the prosecution of financial fraud charges against key executives at Enron and WorldCom. The media themselves sometimes highlight particular cases, such as when Headline News (HLN) anchor Nancy Grace took on the Casey Anthony trial as a personal cause célèbre to see justice done for a murdered child. And sometimes there is no apparent reason other than a slow news day for a case to suddenly attract great media attention.

The key issue for courts concerning both the scope and tone of media treatment of pending cases is the impact that it will have on jurors' judgments of defendant guilt, the conditions under which those effects will most likely occur, and the remedial efforts, if any, that are most likely to minimize those effects. Over the past 40 years, numerous empirical studies have attempted to examine these questions, sometimes with inconclusive or even contradictory results, using a variety of methodological and analytical approaches. A meta-analysis (Steblay et al.) of 23 such studies published between 1966 and 1997 offers some well-documented findings on this question. First and foremost, it is clear from the studies that jurors exposed to negative pretrial publicity are more likely to judge defendants guilty compared to jurors exposed to less pretrial publicity or at least more neutral pretrial publicity. The effect was documented most acutely in those studies that employed jury-eligible citizens as study participants compared to those that employed students. The amount of detail communicated in media accounts (e.g., crime details, arrest information, confessions, prior criminal record, and other incriminating evidence) as well as accounts that employed both video and print media produced greater effects than studies that focused on just one type of pretrial publicity. Crimes involving violence, especially homicide and sexual abuse, also produced greater effects on juror judgments of defendant guilt than other types of crimes. Even general publicity, not specifically related to the case at hand, which included a discussion of similar legal concepts (e.g., eyewitness identification) or case facts (e.g., acquaintance rape) had an indelible impact on juror decision-making.

Of critical importance, these studies collectively confirm that the impact of pretrial publicity on individual juror judgments about defendant culpability carries through to the collective verdicts rendered by juries. Remedial efforts employed by courts (e.g., brief trial continuances, expanded voir dire, judicial instruction, trial evidence, and jury deliberation) do not effectively counter the biasing effects of pretrial publicity. Instead, pretrial publicity exerts a disproportionate imprint on juror memory compared to the evidence actually presented at trial. Most of these studies predate the advent of the Internet Age, or at least its apparent effects on human neural circuitry, so it remains to be seen whether these effects are accentuated or attenuated by frequent Internet use.

Traditional mechanisms for mitigating the impact of pretrial publicity include trial delay, a change of venue or impaneling an out-of-county jury, and extensive voir dire including the use of written juror questionnaires. Unfortunately, the approaches to addressing publicity concerns may not be as viable as they once were given the geographic reach and intensity of 24-hours news coverage. Multiple news media, along with the varied reliability of the information source, simply reach more people, often delivered as an unfiltered or even politicized message. Finally, court use of anonymous juries to prevent communication affecting jury decision-making, such as that of external jury tampering or intimidation, is also explored as a valid response.

Change of Venue

One response to pretrial publicity is to delay a trial date as news often migrates “off the front page.” Ironically, that phrase was derived from the traditional print newspaper of the past. In our technology-saturated culture, such a topic is no longer the most popular tweet or the post no longer appears in the current news feed section. As a result, people forget the details they read initially. While Internet postings are virtually eternal, they migrate to the less-visible archive sections. There is no guarantee, however, that an interested juror would be unable to access the old information quickly, if he or she desired. This is in stark contrast to the era of newspapers in which the juror would be required to spend significantly more time to be able to uncover the details of a past news event and would most likely need to physically leave the courtroom to accomplish this task. The loss of control over the flow of information into and out of the courtroom has indeed left some courts unprepared.

Although a change of venue is an option for courts, numerous notorious trials have been successfully tried in the original venue (e.g., the Nanny trial and a series of well-known Mafia cases), and legitimate concerns about the logistical and financial burdens arise when moving the trial. The proposed venue must also resemble the original community in terms of both demographic and attitudinal characteristics due to the historical importance of public access. The US Supreme Court in *Murphy v. Florida*, 421 US 794 (1975), addressed the level of pretrial publicity that deems one incapable of being impartial. The decision did not require jurors to be completely unaware of publicity, but to be able to set aside the information and judge the defendant solely on the information provided in the courtroom. Courts have developed a fairly consistent analytical framework for determining whether the extent and tone of pretrial publicity have so "poisoned" the local jurisdiction that a change of venue is necessary. In *Irvin v. Dowd*, 366 US 723 (1961), the US Supreme Court wrote that if "an appellant can demonstrate that prejudicial, inflammatory publicity about his case so saturated the community from which his jury was drawn as to render it virtually impossible to obtain an impartial jury, then proof of such poisonous publicity raises a presumption that appellant's jury was prejudiced, relieving him of the obligation to establish actual prejudice by a juror in his case." The presumption is rebuttable, however; if the government demonstrates that an impartial jury was actually impaneled in the appellant's case, the conviction will stand despite appellant's showing of adverse pretrial publicity.

The question for contemporary courts trying the most high-profile cases is whether *any* venue can satisfy these requirements. Timothy McVeigh's trial moved from Oklahoma City to Denver, but that was only possible because the trial was held in federal court. State court, by definition, would be required to maintain jurisdiction and hold the trial within the same state. The Washington DC Sniper trial moved from Fairfax, Virginia (a suburb of Washington, DC), to the southern tidewater area of Chesapeake/Virginia Beach. Complicating the matter, some news accounts indicated the sniper had been in the tidewater area of Virginia, possibly seeking additional victims, before capture, raising the possibility that prospective jurors in that venue would be similarly biased. The saturation of national news undermines the widespread effectiveness of granting a change of venue to overcome the challenge of impaneling a fair and impartial jury for it will be increasingly difficult to locate an alternative venue not equally affected.

The Supreme Court recently revisited this framework in *US v. Jeffrey Skilling*, in which the Enron CEO was convicted of multiple counts of securities and wire fraud involving the collapse of the Enron Corporation. The court ultimately held that the amount and tone of pretrial publicity about the Enron collapse was insufficient to establish a presumption that the jury pool was prejudiced, and in any case, the fact that Skilling only challenged one juror for cause and the jury returned acquittals on nine counts of insider trading demonstrated that the impaneled jury was, in fact, impartial.

Intensive Voir Dire

Another traditional mechanism for addressing pretrial publicity is to conduct intensive voir dire. A written juror questionnaire is often given to potential jurors to identify potential bias. Although case law varies in state courts concerning the scope of permissible questions, the strength of using such a questionnaire is its ability to elicit truthful information from the potential jurors about bias, including attitudes about the specific case as well as underlying attitudes about relevant case issues (e.g., personal experience with substance abuse). Conducting voir dire through written questionnaires in addition to oral questioning in the courtroom is more likely to uncover sensitive or personal information that may affect the juror's ability to be fair and impartial. Courts have identified a number of factors relevant to whether the voir dire in a case involving extensive and prejudicial pretrial publicity would be adequate to impanel an impartial jury including (1) the percentage of the entire pool of veniremembers who evidenced bias; (2) whether the court questioned the veniremembers individually; (3) whether the court questioned the veniremembers thoroughly concerning their knowledge of the circumstances surrounding the alleged crime; (4) whether the court asked each veniremember specifically about the nature and extent of any preconceived notions*;* (5) whether the court asked each veniremember about his or her capability to render an impartial verdict; (6) the length of time the process took; (7) whether the court examined the veniremembers outside the presence of other veniremembers; (8) whether the attorneys had the opportunity to recommend further inquiries, and (9) whether the judge . . . inquired into the prospective jurors' exposure to publicity and ability to render a fair and impartial verdict.

If these remedies are no longer viable, at least in the most notorious trials, what else can be done? Or do courts simply acknowledge that the traditional view of juror impartiality cannot be achieved under these circumstances? If impartiality is an elusive goal, the courts would have to accept a compromised version of justice, mitigating to the greatest extent possible the problems while hoping for the best outcome.

Anonymous Juries and Jury Sequestration

In recent times, judges have become more likely to use anonymity measures (protecting the juror's identities). In the trial of Illinois governor Rod Blagojevich on public corruption charges, Judge James B. Zagel ordered that jurors' names not be disclosed publicly until after the trial ended. He noted that he had personally received dozens of letters, telephone calls, and e-mails from the public concerning the trial and was concerned that jurors would also be targeted for harassment or intimidation if their names became public. He noted that prohibiting jurors from using e-mail or other Internet technologies to communicate with friends and family about non-trial matters for the duration of the 4-month trial would impose an extraordinary burden on them, and other means of screening jurors' personal correspondence, e-mail, and telephone calls would be similarly intrusive. Jury sequestration, which can also be used to insulate jurors from outside influence, is expensive to the court and onerous on the jurors, their families, and the courts. Consequently, it is rarely employed except in the most extreme circumstances. Local and national media outlets challenged the order on First Amendment right to access government proceedings in an interlocutory appeal to the federal Seventh Circuit of Appeal. Historically, anonymous juries were rarely permitted in the federal courts unless the trial judge made specific factual findings concerning the immediate risk of jury tampering or intimidation. The fact that Judge Zabel's order ultimately prevailed illustrates the point that many courts have come to appreciate that contemporary communications technologies pose as great a risk or more to juror impartiality as traditional in-person approaches, in part due to the substantially larger pool of Internet-savvy people who might be inclined to contact jurors in high-profile cases.

Conclusions

The introduction and evolution of Internet technologies that has taken place over the past two decades has introduced a number of challenges to the concept of juror impartiality. In many instances, these challenges are simply extensions of the types of challenges that courts have faced in the past – and for which highly effective strategies exist. A thorough voir dire can identify jurors who cannot serve fairly and impartially due to personal knowledge about the case, exposure to pretrial publicity, or preconceived opinions or bias about case-related factors. Effective pretrial instructions about independent research and communications with family and friends can inform jurors about necessary restrictions on their activities during trial to prevent them from being exposed to potentially prejudicial information. Very high-profile trials may require additional measures including anonymous or sequestered juries to prevent jury tampering or intimidation by electronic means.

Although these are all tried and true techniques that require only a little tweaking to be equally useful for maintaining juror impartiality in the Internet Age, some aspects of modern telecommunications technologies appear to affect jurors and jury trials in a qualitatively different way. Perhaps the most troublesome is the apparent impact that frequent juror use of Internet technologies is having on cognitive behavior, especially the ability to retain and interpret information. Future jurors may not be as effective as decision-makers unless they are permitted to access the Internet to supplement and interpret the evidence they are given at trial. That access, however, is currently prohibited on grounds that doing so would undermine juror impartiality. It is hard to imagine a more direct confrontation of traditional trial procedure and modern technological innovation. It is not clear how new media will ultimately change how we think about courts, about jurors and their role in the justice system, and about how jurors should fulfill that role, but it is clear that some change will ultimately occur.

Similarly, the overwhelming volume of news, its iterative manifestations, and its expanding geographical distribution to all parts of the globe pose an immense challenge for courts. Even if the tone of media coverage of an upcoming trial is relatively neutral, the sheer level of detail may so saturate the potential jury pool that it becomes increasingly difficult to impanel a fair and impartial jury in that jurisdiction. But it may be equally difficult to locate an alternative jurisdiction where the level of pretrial publicity and community impact is sufficiently less. Even after a fair and impartial jury is impaneled and the trial has commenced, the court faces the ongoing possibility that the public audience watching both the trial proceedings, often in real time, and corresponding news and commentary will draw very different conclusions about the appropriate outcome of the trial than the impaneled jury would based solely on the evidence and law.

Courts can no longer blandly assume that the public understands their mission and the underlying rationale for trial procedures and that trial outcomes will be inevitably accepted as valid judgments in the court of public opinion. Ultimately, courts will have to devise more persuasive arguments, and more effective strategies to promulgate those arguments, of the continued importance and validity of its core function in the justice system. Courts are institutionally reactive organizations that have been slow to adapt to the implications of new media on court operations generally and in the context of jury trials specifically. But just as new media is affecting changes in human cognitive processing, it is similarly affecting – in a dynamic and interactive way spurred by the use of new media – both public perceptions about the courts and courts' own perceptions about themselves and their role in contemporary society.

As a final note, it is important to recognize that contemporary technologies are changing very rapidly. Courts have been taken entirely by surprise by many of the communications technologies in widespread use in contemporary society. They are even less aware of and prepared for newer technologies that likely have already been developed and deployed; they have not begun to imagine the future implications that these technologies will have on court operations. There is a distinct possibility that the issues discussed in this entry will already have become moot by the time this encyclopedia is published by the introduction of newer technologies. At the very least, however, this entry will provide a historical glimpse of the issues and problems that courts once confronted. Future researchers will have to assess whether their reactions and adaptations were ultimately adequate and satisfactory, or insufficient, in the long run.

Recommended Reading and References

Amey G (2011) Social media and the legal system: analyzing various responses to using technology from the jury box. J Leg Prof 35:111–130

Artigliere R (2011) Sequestration for the Twenty-First Century: disconnecting jurors from the internet. Drake Law Rev 59:621–645

Blackman J, Brickman E (2011) Let's talk: addressing the challenges of Internet-Era jurors. Jury Expert 23(2):35–45

Carr N (2010) The shallows: how the internet is changing our brains. New York, W. W Norton

Chesterman M (1997) O.J. and the dingo: how media publicity relating to criminal cases tried by jury is dealt with in Australia and America. Am J Comp Law 45:109–147

Dann BM (1993) "Learning lessons" and "speaking rights": creating educated and democratic juries. Indiana Law J 68:1229–1279

Farrell A, Givelber D (2010) Liberation reconsidered: understanding why judges and juries disagree about guilt. J Crim Law Criminol 100:1549–1586

Fiske ST, Taylor SE (eds) (1991) Social cognition, 2nd edn. McGraw-Hill, New York

Greene E (1990) Media effects on jurors. Law Hum Behav 14:439–450

Hannaford-Agor PL (2008) When all eyes are watching: trial characteristics and practices in notorious trials. Judicature 91:197–201

Hannaford-Agor P, Hans VP, Mott NL, Munsterman GT (2002) Are hung juries a problem? National Center for State Courts, Williamsburg

Hannaford-Agor P, Rottman D, Waters NL (2012) Jurors & new media: an experiment. Perspectives on state court leadership. National Center for State Courts, Williamsburg

Hans V (2007) Deliberation and dissent: 12 Angry Men versus the empirical reality of juries. Chicago-Kent Law Rev 82:579–589

Hans V, Vidmar N (2007) American juries: the verdict. Prometheus Books, Amhearst

Hans VP, Hannaford-Agor PL, Mott NL, Munsterman GT (2003) The hung jury: The American jury's insights and contemporary understanding. Crim Law Bull 39(1):33–50

Hastie R, Penrod SD, Pennington N (1983) Inside the jury. Harvard University Press, Cambridge

Irvin v. Dowd, 366 U.S. 717 (1961)

Marder N (2001a) Juries and technology: equipping jurors for the Twenty-First Century. Brooklyn Law Rev 66:1257–1299

Marder N (2001b) Cyberjuries: a new role as online mock juries. Univ Toledo Law Rev 38:239–269

McGee A (2010) Juror misconduct in the Twenty-First Century: the prevalence of the Internet and its effect on American courtrooms. Loyola Los Angel Entertain Law Rev 30:301–326

Meringolo JC (2010–11) The media, the jury, and the high-profile defendant: a defense perspective on the media circus. NY Law Sch Law Rev 55:981–1012

Munsterman GT, Hannaford-Agor PL, Whitehead GM (2006) Jury trial innovations, 2nd edn. National Center for State Courts, Williamsburg

Murphy v. Florida, 421 US 794 (1975)

Murphy TR, Hannaford PL, Loveland GK, Munsterman GT (1998) Managing notorious trials, 2nd ed. National Center for State Courts, Williamsburg

Salaz K, Hodson T, Davey C (2010) New media and the courts: the current status and a look at the future. A report of the New Media Committee of the Conference of Court Public Information Officers (available at www.ccpio.org)

Steblay NM, Besirevic J, Fulero SM, Jimenez-Lorente B (1999) The effects of pretrial publicity on juror verdicts: a meta-analytic review. Law Hum Behav 23:219–235

United States v. Jeffrey Skilling, 130 S. Ct. 2896 (2010)

United States v. Rod Blagojevich and Robert Blagojevich, 612 F. 3d 558 (7th Cir. 2010); 614 F. 3d 287 (7th Cir. 2010); 743 F. Supp. 2d 794 (N.D. Ill. 2010)

Vidmar N (2002) Case studies of pre- and midtrial prejudice in criminal and civil litigation. Law Hum Behav 26:73–105

Vidmar N (2006) When jurors talk about their verdict. In: Jury ethics: juror conduct and jury dynamics. Paradigm Publishers, Boulder

Waters NL, Hans VP (2009) A jury of one: opinion formation, conformity, and dissent on juries. J Empir Leg Stud 6:513–540

Justice Reinvestment

Rob Allen
Justice and Prisons

Overview

While the idea of justice reinvestment (JR) has its origins in the USA, politicians, policy makers, and academics in the UK have shown a growing interest in its applicability over the last 8 years. The level of public spending on prisons in the UK does not yet approach that of the USA whose 2.5 million prisoners represent an imprisonment rate four times higher than in Britain. Yet the UK prison population, particularly in England and Wales, which is by far the largest jurisdiction, has more than doubled in the 20 years since 1991 to reach more than 87,500 in the autumn of 2011. Concerns about the cost and effectiveness of this vastly increased use of imprisonment have led to a widespread desire to find ways of curbing the spending involved or at least diverting it into more productive community-based measures.

The formation of a new coalition government in May 2010 led to a marked change of criminal justice policy with the new Conservative and Liberal Democrat administration aiming to halt the seemingly inevitable rise in prison numbers. The coalition's program for government promised to introduce a "'rehabilitation revolution' that will pay independent providers to reduce re-offending, paid for by the savings this new approach will generate within the criminal justice system" (Cabinet Office 2011). The government has described this as "a radical and decentralizing reform which will deliver a fundamental shift in the way rehabilitation is delivered. It will make the concept of justice reinvestment real by allowing providers to invest money in the activity that will prevent offending rather than spending money on dealing with the consequences" (Ministry of Justice 2010).

It is too early to say what impact this form of justice reinvestment might have. Before it was under way, serious disturbances on the streets of English cities in the summer of 2011 led to demands for a tougher approach to law and order in some quarters but the need to cut public-spending acts as a restraint on those who would reverse policy and return to a the more expansionist prison-building program that the Conservative party favored in advance of the 2010 general election.

A new Justice Secretary appointed in 2012 has announced me acceleration of the introduction of "Payment by returns" with plans to open most probation services to competition. Savings will be used to expand community supervision of offenders released from short prison sentences. But reduced spending on prisons will be sought not by cutting prison numbers but through reductions in the unit costs of prison management.

Some Important Differences between the USA and the UK

At the outset, it is important to recognize some important differences between the USA and the UK which make the idea of justice reinvestment more compelling to Americans than in Britain. The still vastly differing scale of imprisonment in the USA has created a large pool of prisoners who, on any reasonable assessment, could safely be released from prison. As a result of the comparative harshness of sentencing and sentence execution policies, there is inevitably considerable scope for prison-reduction measures. For example, mandatory and truth in sentencing policies have led to the long-term incarceration of large numbers of nonviolent prisoners. Reducing prison capacity potentially frees up sums for reinvestment (or indeed to translate into tax cuts) which are thus substantially higher across the Atlantic than in the UK. There are unlikely to be many "million dollar blocks" in England and Wales – residential neighborhoods where that amount is spent on imprisonment each year – although analysis in the northeast of England found that in 2005 magistrates in Gateshead incurred over half a million pounds worth of costs in sending just over a hundred individuals to prison, on average for a few weeks. Crown court decisions generated substantially greater costs (Allen et al. 2007). A more recent study found that it cost the taxpayer £2.5 million in 2009/2010 to send nonviolent and nonsexual offenders from the London borough of Lewisham to prison for periods of less than a year. The authors concluded that funds that could be made available to local agencies to prevent reoffending through a process of justice reinvestment are therefore considerable (Lanning et al. 2011). But they are substantially lower than in the USA.

It is true too that America's lack of welfare structures and the highly concentrated zones of urban deprivation mean that resources to strengthen housing, employment, substance misuse, and other services are likely to yield quicker and more visible positive impact on the resettlement of released prisoners than might be the case in the UK.

Nonetheless, there is considerable scope for enhancing measures to rehabilitate offenders in the UK. In England and Wales, prisoners sentenced to less than 12 months do not receive any supervision or support from the probation service on release. There are gaps in services even for those offenders who are being supervised in the community. The National Audit Office found in 2008 that "some community order requirements, for example alcohol treatment... are not available or rarely used" – this despite strong links between alcohol and offending. They also found long waiting lists for some order requirements, in particular group programmes on domestic violence" (NAO 2008).

It is also the case that the aspects of the structure of American governance lend themselves to JR approaches. The shared responsibilities exercised by counties, states, and the federal government in running and financing imprisonment provide opportunities to introduce financial incentives to control or, as well as (as has been the case), to expand prison numbers. A variety of fiscal arrangements have been introduced which reward counties which develop measures that reduce demand for custodial places at state level. A virtuous circle is created in which state savings on incarceration are reinvested in local alternatives which in turn further reduce demand for expensive state placements. This process has been seen, for example, in New York State where four closed facilities for juveniles have been closed with resources invested in family therapy and other alternative programs in New York City (Solomon and Allen 2009), but the approach has a history going back to the California probation subsidy scheme in the 1970s. Federal grants have also been used to encourage reductions in custody particularly in respect to disproportionate minority confinement.

In England and Wales, the costs of custody are met almost entirely by central government through the National Offender Management Service, an agency of the Ministry of Justice. There is a modest exception in the case of juvenile custody. Local authorities are currently required to pay a third of the costs when a court orders a child under the age of 17 to be placed in secure care during a remand period in criminal proceedings. This local liability applies in a relatively small number of cases – most juveniles who are remanded into custody are not placed in secure care institutions but in prison establishments which are paid for centrally. However, as part of their program of criminal justice reform, the government is seeking to make local authorities financially responsible for the costs of all secure remands of young people under the age of 18 wherever they are placed. Although responsibility for commissioning and placements will be retained by central government, the aim is to "simplify current arrangements and provide a powerful incentive for local authorities to invest in alternative strategies for this group of young people" (Ministry of Justice 2011). The legislation to introduce this change is currently being debated in Parliament.

The system of American governance also gives the executive a somewhat greater say over the size of custodial populations than in the UK where the courts play a more decisive role. In juvenile justice, it is the executive which decides whether custodial placements are made in some states. In others such placements are in large part made following recommendations by probation officers. This provides a straightforward lever to change rates of committal to custody. In the adult system, executives have often retained the right to adjust the proportion to time served and of course can determine the level of flexibility applied in cases of parole violations. Reintroducing or making more generous schemes for good time credits and introducing a little more discretion in responses to parole violations can have substantial impacts. Legislative change is not always needed to reduce the actual lengths of sentences served or the rate of recalls to prison. In England and Wales, these are the two key factors which lie behind the sharp rise in prison numbers in England and Wales since the early 1990s.

In short, in the USA, the extraordinary rise in mass incarceration has made JR a highly attractive proposition for cash-strapped states and provides numerous opportunities to scale back the use and cost of prison. In the UK, because the rate of imprisonment is lower, the opportunities are somewhat more modest.

Recent Policy Developments in England and Wales

Although the overall rate of imprisonment in the UK is much lower than in the USA – 155 per 100,000 population compared to 743, between 1992 and 2011 – the prison population in England and Wales more than doubled in size. About three quarters of the rise is accounted for by increase in the numbers sentenced to immediate imprisonment and about 16 % by an increase in those recalled to prison for breaching the terms of their release on parole or license. A recent Ministry of Justice analysis suggests that the reasons lie in tougher sentencing and enforcement outcomes (which is not in question) and a more serious mix of offenders coming before the courts (for which the evidence is more contested) (Ministry of Justice 2009). Rising prison numbers represent the costliest tip of a criminalization iceberg which has seen a large extension of the reach of the criminal justice system, during a period of falling crime. The creation of 3,000 new criminal offenses, the development of hybrid forms of social control such as Anti-Social Behavior Orders, and the introduction of indeterminate sentences form a pattern in which social problems are increasingly treated by way of punishment and control. Overseas observers express surprise that this trend should have occurred under a Labor government. Indeed as recently as 2002, in the White Paper "Justice for All," the Labor government listed the record prison population, with its costs and poor outcomes as something that "is not working" (Home Office 2002).

Since then, record prison numbers have like numbers of police officers become something to be trumpeted. Until 2010, the then Labor government relied on the analysis contained in a report they had commissioned from Lord Carter called "Securing the Future" (Carter 2007). In large measure, this offered a "predict and provide" approach to the supply of prison places. The report did contain one or two proposals to reduce or stabilize demand for imprisonment, but its scope was narrowly confined and was dismissed as "deeply unimpressive" by the House of Commons Justice Committee (Parliamentary Select Committee on Justice 2008). It ignored entirely the case for a comprehensive and wide-ranging strategy to reduce the resort to imprisonment and to develop alternative ways of dealing with offenders. Carter's report stood in sharp contrast to the work of the Scottish Prisons Commission which was established in 2007. Its report "Scotland's Choice" set out a vision of reducing Scotland's prison population over time from 8,000 to 5,000 (McLeish 2008). Among its recommendations, the Scottish report called for progress "in developing services that are available nationwide to address the social and health related needs of many offenders" and called on the devolved government "to promote recognition across all departments, all public services, all sectors and all communities of a duty to reintegrate both those who have paid back in the community and those who have served their time in prison" (McLeish 2008).

Lord Carter's report about England and Wales had by contrast nothing to say about social, educational, and health-care policies which are needed to prevent crime, rehabilitate offenders, and reintegrate those leaving prison. There is evidence that despite substantial investment in public services in recent years, major shortfalls in the availability of social provision remain particularly in the most deprived areas and among the most excluded populations in England and Wales. A separate government review of people with mental health problems and learning disabilities in the criminal justice system, for example, concluded in 2009 that for diversion into the health-care system to be introduced effectively, there needed to be sufficient capacity in mainstream services, as well as confidence in those services for those making decisions about offenders (Bradley 2009). While the review did not make a detailed assessment of the adequacy of services, it tellingly noted that in 2006, courts made use of their powers to add a mental health treatment condition to a community sentence on only 725 occasions. Such conditions represented less than 0.4 % of the 203,323 requirements imposed during the course of the year. Similarly, despite considerable investment in drug treatment, the number of residential rehabilitation places is still low compared to other countries. There are about 2,500 beds in England with about 16,000 individuals accessing residential services for substance misuse each year. Given that half of male prisoners and two thirds of women have used class A drugs in the 6 months prior to imprisonment, there is a strong prima facie case for increasing capacity significantly. Despite some progress in establishing community-based services for women offenders following a major review of women in prison, women are still receiving short prison sentences for want of constructive alternatives.

It is arguable too that the distorted priority given to increasing imprisonment has been encouraged by a costly centrally driven system of offender management, the child of an earlier review of correctional services undertaken by Lord Carter. The creation of the National Offender Management Service (NOMS), a central government agency responsible for prison and probation services (and to a lesser extent the Youth Justice Board) has produced a situation in which local health, education, employment, and social services can slough off their responsibilities for people in the criminal justice system, safe in the knowledge that their needs will be addressed by a central government agency. This is most starkly illustrated in the juvenile system where local authorities can shunt the costs of meeting the needs of demanding teenagers onto central government, but it is more generally the case that local mainstream agencies have little incentive to address and absorb crime and delinquency problems in the way that they might.

The formation of the coalition government in May 2010 brought an important change of emphasis in criminal justice policy. The Programme for Government which set out the key policy commitments included a commitment to more effective sentencing policies, an overhaul of rehabilitation, and an exploration of alternative forms of secure treatment-based accommodation for mentally ill and drugs offenders (Cabinet Office 2010).

Specific proposals for criminal justice reform were published 6 months later in the consultation paper "Breaking the Cycle." This proposed a series of measures which aimed to result in "fewer crimes being committed overall, stemming the unsustainable rise in the prison population and ultimately achieving a reduction in the amount of money spent on the criminal justice system. There will be more payback to victims and communities and the public will be better served" (Ministry of Justice 2010).

During the period of consultation, some of the measures proved controversial – particularly a proposed increase in the sentencing discount offered to offenders who plead guilty at an early stage in proceedings. When it became clear that increased discounts would apply to serious crimes such as rape, there was something of a backlash. As a result of pressure from the Labor opposition and from harder line members of the Conservative party, policy on criminal justice assumed a markedly harder line in the weeks leading up to the publication in June 2011 of the government's response to the Green Paper consultation and of the Legal Aid Sentencing and Punishment of Offenders Bill which gave legislative expression to the policy.

The post-consultation policy shows signs that the government has watered down its prison reform proposals on both the supply and the demand side. The response to the consultation makes it clear that the government is "not aiming to cut the prison population – and we are clear that there must always be places for those that judges sentence – but we will manage a stable, effective system rather than undermining it by endlessly and irresponsibly inflating prison numbers for their own sake" (Ministry of Justice 2011). This disappointed many reformers looking for reductions in the use of prison which was but not surprising given that one of the main mechanisms for reducing prison numbers – the increased sentence discounts for early guilty pleas – was to all intents and purposes abandoned.

Disappointing too for reformers was the decision that the government would not push for community sentences to be used instead of prison. Instead, the government has decided to try to "transform community orders into more credible punishments that stop offenders getting to the point where custody is the only option. Non-custodial sentences need to be tough and demanding. For too long, they have fallen short of what is required" (Ministry of Justice 2011). Simultaneously making community sentences more demanding while moving them lower down the tariff of sentencing options no longer looks like a strategy for prison reduction. It rather spells the danger of escalating offenders towards custody by increasing the risk of failure to comply rather than diverting them from it.

The government's commitment towards rehabilitation and a more measured use of prison has also been put at risk by the public disturbances in English cities in August 2011.

Although the government has sought to use the fact that large numbers of rioters appear to be heavily convicted to remake the case for better rehabilitation, courts have adopted a harsh line on the remanding and sentencing of riot-related offenders. The prison population reached record levels in five consecutive weeks in September and October 2011.

2012 saw a new Justice Secretary appointed who published radical proposals to reform the delivery of offender services (Ministry of Justice 2013). The aim is clearly reduce spending by cutting the unit costs both of prison and probation and not by reducing the number of prisoners.

Justice Reinvestment in the UK Developments and Proposals

Policy

Notwithstanding the important differences between the USA and UK, a range of specific policy proposals drawing on JR have been produced in Britain in recent years. Attention was first drawn to the possible application of the concept in the UK in 2004 (Allen 2004), and since then, there has been a range of ideas put forward by academics, think tanks, and charities which look to apply various of the key elements of justice reinvestment. These include suggestions for:

(a) A much more locally driven system of so-called primary justice which would involve shifting control of prisons and key supporting services away from Whitehall to a local level and the creation of a local "safety and justice" budget to fund local prisons and neighborhood policing. This pot of money would be used to commission local services, either from existing providers or by setting up new local services. A local budget could include approximately 35 % of the prison budget, the administration budget for magistrates' courts, local policing, and probation (LGIU 2009).

(b) An extension to the remit of locally based Youth Offending Teams (multiagency professional teams which deal with juvenile offenders) to the adult age group (LGA 2005; Allen and Stern 2007).

(c) More locally based management of the prison system by abolishing the National Offender Management Service (NOMS) which is the central government agency responsible for prison and probation services and replacing it with a network of Community Prison and Rehabilitation Trusts (CSJ 2009).

(d) Using funds earmarked by the government for prison expansion to strengthen measures in the community such as restorative justice which could reduce demand for imprisonment (RCP 2008).

The most thorough-going proposals have come in the House of Commons Justice Committee's substantial report "Cutting Crime – The Case for Justice Reinvestment" (Justice Committee 2010). The report focused attention on not only how much is being spent on imprisonment but what alternative which uses the public money consumed by prison could be put if demand for prison places could be reduced. It also drew attention to how people going to and returning from prison are disproportionately drawn from the poorest neighborhoods and how targeted investment in these areas could help develop more initiatives both to prevent crime and improve reintegration of ex-prisoners. The report argued that the prison population in England and Wales should be capped with an objective of reducing it to two thirds of its then level. It coined the term "pre-habilitation" to argue that resources should be invested in the most deprived communities which produced the greatest numbers of candidates for custody and suggested that mainstream local agencies responsible for providing health, education, social care, and employment should be incentivized to do more to assist offenders subject to community sentences and returning from prison.

Building on the interest in places as well as cases, in 2011, a more detailed strategy plan was proposed by the influential think tank the Institute for Public Policy Research (IPPR) for the funds spent on imprisoning adult offenders for periods of less than 12 months to be devolved from central government to local councils so that they can put in place measures to deal with the causes of offending behavior. Councils would be charged back by the prison service every time someone from their local area was sent to prison for a period of less than 12 months. The IPPR also recommended that probation services should be decentralized and fully integrated into crime-reduction work locally, by placing it under local authority control (Lanning et al. 2011).

These proposals have so far had limited specific impact on policy makers, although the general thrust of the thinking that lies behind them – reduced use of prison and greater localism – resonates strongly with aspects of the rehabilitation revolution. There is however one specific current that underpins government policy on the provision of a whole range of public services which shares some features with JR – that is, payment by results or PBR. A number of government departments are moving towards an outcome-based system of payment for services. Funding for private or not-for-profit organizations involved in getting the long-term unemployed into work, for example, is paid exclusively or heavily based on delivering results – not just placing people in work but keeping them there, with payment made after results have been delivered. Similar initiatives are being piloted in respect to drug and alcohol treatment and have been proposed as a way of funding early intervention programs with children at risk.

In the justice area, the government has pledged to "pioneer a world first – a system where we only pay for results, delivered by a diverse range of providers from all sectors. This principle will underpin all our work on reoffending. This is a radical shift" (Ministry of Justice 2011). One aspect if this relates to diversification of the market – the government says they will no longer provide rehabilitation services directly without testing where the private, voluntary, or community sectors can provide them more effectively and efficiently. An Offender Services Competition Strategy was published in July 2011 which signals a series of competitions for contracts to run nine prisons and a range of community-based programs (Ministry of Justice 2011b).

A second aspect of this relates to the way contracts for the provision of services are framed. A key element of the approach to competition is to "continue to develop and expand a range of models where providers are paid based on the results they achieve. This will allow providers both to innovate and to ensure that value for money is achieved. Providers will be encouraged to put forward innovative models to focus more effectively on rehabilitation..." (Ministry of Justice 2011b). For example, a recently agreed contract to run Doncaster prison (which in fact predates the competition strategy) includes a 4-year payment-by-results pilot starting in October 2011. This contract, apparently the first of its type in the world, will mean that private provider Serco are held to account both for running a safe, decent, and secure prison and the reoffending of their offenders on release. For each year of the pilot, 10 % of the contract value will be placed at risk and will only be retained by Serco if they reduce reoffending by 5 percentage points. If they fail to achieve this target, then money will be paid back to the government. If they deliver more, then they will be paid more than the original contract price (Ministry of Justice 2011). Serco claims that the Doncaster prison scheme could see more than 15,000 fewer offenses a year if it hits its targets.

Practice

Alongside these developments and proposals for policy change, a number of practical pilot initiatives have drawn inspiration from aspects of JR. One of the first which claimed to be based on JR was the Diamond District initiative which operated in six London boroughs in 2009–2010. Multiagency teams including police and probation officers and local authority workers were located in neighborhoods with high numbers of people returning from short spells in prison. (The term diamond was coined to communicate the disproportionate value which would attach to reducing reoffending in these areas.) The teams offered enhanced resettlement support to prisoners returning to the neighborhoods. Although there was no obvious mechanism by which any savings to the prison system would be recouped into the communities concerned, the question has not been pursued since the £11million initiative did not find that the experimental group was reconvicted any less than a carefully matched control group – although interestingly the rates of reoffending for both experimental and control groups were well below the national rate. The evaluation of the project argued that "whilst the headline finding of the report is that there was no evidence of reduced reoffending as a result of the Diamond Initiative within the original cohort-this does not demonstrate that Diamond has failed as an approach: such a conclusion would be erroneous" (LCJP 2011).

Thus, despite the disappointing results, interest in justice-reinvestment-inspired schemes has not waned. Three pilot programs involving JR are under way at the time of writing each of which aims to reduce the demand for custody.

The first relates to juvenile offenders, for whom a Youth Justice Reinvestment Pathfinder Initiative was set up in early 2011. The Green Paper Breaking the Cycle proposed that local authorities should share both the financial risk of young people entering custody and the financial rewards if fewer young people require a custodial sentence. The pilots aim to test how this might be achieved.

Under the scheme, central government will be investing a proportion of the budget used to fund custodial places into grants for consortia of local authorities to commission alternatives to custody in four pathfinder areas. The aim is that as a result the use of incarceration for both sentenced and remanded young people in the pathfinder areas will be reduced by an agreed amount over a 2-year period. If the authorities fail to reduce their use of custody by the agreed amount, they will have to pay back the funding – or a proportion of it – to the Ministry of Justice. Pathfinders will be free to commission and deliver schemes in a way that best addresses their local needs and circumstances.

The second initiative known as the Financial Incentive Model has also been developed by the Ministry of Justice to test the extent to which local partners can work together more effectively to reduce crime and reoffending. This approach is being tested in six local areas: Greater Manchester and five separate London boroughs (Croydon, Hackney, Lewisham, Lambeth, and Southwark). The project covers both adult and youth aspects of the criminal justice system. The basic idea is that the Ministry of Justice would pay local authorities and service providers if they were successful in reducing court convictions and court disposals for adult and youth offenders in the local areas. Sharing the savings that accrued to the Ministry of Justice from the reductions would enable the local authority to reinvest further in crime prevention activity at the local level in line with their priorities.

The rationale for the project is based on the hypothesis that giving local partners more financial accountability for the demand placed on criminal justice services, such as courts, prison places, and probation services, will provide them with an incentive to work more effectively to reduce crime and reoffending locally and thereby to reduce that demand. Under the model, the local partners are free to implement plans to reduce crime and reoffending, targeting their resources on specific groups of offenders in line with the local priorities and crime patterns. Demand on criminal justice services will be measured using a series of metrics focused on offending that results in short custodial sentences and less severe sentences. Demand will be measured across two consecutive 1-year periods, with the measurement period starting on 1 July 2011 and closing on 30 June 2013.

The third initiative, which has attracted significant attention in the UK and internationally, is a novel form of financing services to resettle prisoners leaving Peterborough prison after a short custodial sentence – the so-called Social Impact Bond ("SIB").

A Social Impact Bond is a contract with the public sector in which investment is raised from socially motivated investors and used to pay for a range of interventions to improve social outcomes. As improved social outcomes result in savings to government and wider benefits to society, the government agrees to pay a proportion of any savings back to the investors if the interventions succeed. The payments from government cover the initial investment plus a financial return which depends on the extent to which outcomes are improved. If the outcomes do not improve, investors lose their investment. The term "bond" is a misnomer because all of the capital is theoretically at risk should the project fail to make any impact.

In the Peterborough project, which was officially launched in September 2010, £5 million has been raised from 17 investors – mainly charitable foundations interested in a "mission-aligned" investment opportunity – to fund the provision of services to 3,000 prisoners returning to the community having served a sentence of less than 12 months. Investors stand to make a return if the three cohorts of offenders eligible for the service are reconvicted 7.5 % less frequently than a matched control group of prisoners from other prisons. Nationally about 60 % of short-term prisoners are reconvicted within a year of release, and few of these prisoners have access to support which might help them find a stable home, a job, and assistance with other problems which left unaddressed can rapidly lead back to crime. The funds are being used to commission not-for-profit organizations to provide intensive help before and after release – so-called "through-the-gate" services.

The attractiveness of the SIB to a range of stakeholders is clear. For government a SIB removes the financial risk that they pay for services that prove to be ineffective at addressing social needs and improving outcomes. For investors a SIB offers a potential return on investment – if reconviction events are reduced by 10 % across all cohorts, investors are expecting an annual internal rate of return of around 7.5 % up to a maximum of around 13 %, depending on the scale of the reduction in reconviction events (Disley et al. 2011). The SIB also offers the chance to use capital in a socially useful way. For service providers a SIB provides up-front funding for the delivery of services (so they do not carry the risk of not being paid). For the public and service users, funding raised through a SIB may pay for services that fill a gap in existing provision.

The first evaluation report on lessons learned from the planning and early implementation of the SIB at Peterborough did not find it possible to draw conclusions about or comment on any outcomes from the scheme. It raised a number of questions about its complexity of and the costs of establishing it – the intermediary who arranged the SIB, Social Finance, estimated that it has invested approximately 2.5 person – years of its resources, and more than 300 h of legal advice (provided pro bono), as well as specialist tax advice, in developing the SIB. The report also pointed out that the scheme *is* too small to deliver substantial "cashable" savings – "The ability of the SIB model to lead to identifiable savings for government is yet to be tested, if the SIB model is implemented on a larger scale" (Disley et al. 2011).

If this model were rolled out nationally, the report suggests that there could be incentives to "cherry pick" by prison or area – that is, for funders to choose sites for investment or indeed particular types of beneficiaries for the service who are likeliest to achieve improved outcomes – by implication leaving or "parking" as it is sometimes called more hard to reach people or places. More technical difficulties were also identified – for example, that future SIBs may face the challenge of sharing outcome payments across central and local government departments or other agencies who may stand to benefit from savings brought about by the intervention. Also SIBs in areas of work that currently receive statutory provision will need to devise outcome metrics that enable the government to isolate the effect of SIB-funded interventions from existing statutory services. Short-term prisoners do not usually receive any statutory help so the issue does not arise at Peterborough. The evaluation report also points out that using a Social Impact Bond to fund a new service in a field where no existing interventions are in place also "avoids the many difficult issues raised by the possible need to decommission statutory services in such circumstances – another possible challenge for future SIBs" (Disley et al. 2011).

Supporters of SIBs point to the fact that they provide a vehicle which could attract new money into underfunded activities like resettlement and community support for offenders. Over half the investment (by value) and half the investors (by number) in the Peterborough SIB are either using their endowment or investing in the UK criminal justice area for the first time.

Others are less convinced. The head of the Howard League, a leading penal reform charity reportedly described the system of payment by results for justice services as "nonsense." "Using the example of a prolific burglar who has recently been released from prison, Frances Crook was quoted as saying: "You come out of jail, you're sofa-surfing and you're on drugs, but you clean yourself up, get a job and get a girlfriend. Who should be able to make money out of you? Public service is something which does great good and which we should value. If you introduce profit into public service it changes everything" (Civil Society Finance 2011).

Others have attacked the use of investment bonds more generally as "fairy dust, disguising short-sighted cuts that will do much social damage in future years. We get the social services we are willing pay for with our taxes: this is fool's gold" (Toynbee 2011).

It will of course be some time before data is available to make an empirical judgment about the impact in criminal justice of the payment-by-results approach in general and the Social Impact Bond in particular. That has not prevented the government committing to applying the principles of payment by results across all services which reduce reoffending by 2015.

Conclusion

Justice reinvestment has come to have a range of meanings in England and Wales, and it is possible to identify three types of policy and programs. First, there are relatively pure examples of initiatives that are true to what are generally taken to be the key tenets of the concept, aiming to transfer resources way from imprisonment and into community-based crime-reduction measures of one sort or another (Allen 2011). The Youth Justice Reinvestment Pathfinders and Financial Incentive Model fall into this category. Each is based at the local government level (albeit through consortia of local authorities in the case of the youth scheme), but it is not yet clear how much of a focus they will have in the particular neighborhoods whose residents are disproportionately imprisoned. These programs are all publicly funded with local agencies given incentives to reduce the use of custody through the opportunity to spend resulting savings but facing a concomitant risk of meeting additional custodial costs which might be incurred.

There is a second form of JR which involves the investment of funds from outside government into activities designed to produce positive social outcomes such as reduced reoffending. The Social Impact Bond is an example of a model which offers incentives and risks for investors; if the program succeeds, they will receive outcome payments from the government's savings in prison spending produced by reduced reoffending. If the program fails, government pays for the prison places but not the unsuccessful attempt to reduce reoffending. The scale of likely investment in financial instruments such as this is not yet known. Nor is it clear whether Social Impact Bonds could practicably be used to replace existing activities rather than fund new pieces of work.

Some uncertainty also surrounds the third type of JR which is the approach to contracting based on payment by results. The hope is that attaching financial rewards to successful rehabilitation will promote innovation and that reduced reoffending will bring down demand for prison. But unlike the first type of JR, there is no commitment to reinvest any savings in communities. Unlike the second type, there is no outside funding brought into the system. Given the substantial budget reductions being faced by the Ministry of Justice, it is all too likely that PBR could prove a form of justice disinvestment rather than reinvestment.

Despite the range of JR-related initiatives in place in 2011, there are still possibilities for more extensive and deeper application of the approach. This is certainly the conclusion reached by the parliamentary committee which championed JR in 2010. In an inquiry into the role of the probation service the following year, the Justice Committee concluded that "the model of justice reinvestment explored by our predecessor Committee – of greater investment in a package of reforms to reduce the use of custody, including increased spending on probation services, allowing significant resources to be freed by halting the prison-building programme and enabling current inefficient prisons to close – has not been fully exploited; this will undoubtedly impede the pace at which capital can be transferred from prisons to community-based interventions and will therefore continue to leave probation services and local communities deprived of desperately needed resource" (Parliamentary Select Committee on Justice 2011).

Related Entries

- ▶ Comparative Courts and Sentencing
- ▶ Conditional Release
- ▶ Crime Mapping
- ▶ International Sentencing
- ▶ Juvenile Diversion
- ▶ Law, Diversion and Community Sanctions in Juvenile Justice
- ▶ Sex Offenders and Criminal Policy
- ▶ Social Network Analysis and the Measurement of Neighborhoods

Recommended Reading and References

Allen R (2004) Correctional creativity guardian 7th July 2004

Allen R (2008) The cost of custody: whose responsibility? In: Blyth M et al (eds) Children and young people in custody managing the risk. Policy Press, Bristol

Allen R (2011) Justice reinvestment and the use of imprisonment. Criminol Public Policy 10:617–627

Allen R, Stern V (2007) Justice reinvestment a new approach to crime and justice. King's College, London

Allen R, Jallab K, Snaith E (2007) Justice reinvestment in gateshead the story so far. In: Allen R, Stern V (eds) Justice reinvestment a new approach to crime and justice. King's College, London

Bradley L (2009) The Bradley report. Lord Bradley's review of people with mental health problems or learning disabilities in the criminal justice system. Department of Health, London

Cabinet Office (2011) Programme for government. The Stationery Office, London

Carter P (2003) Managing offenders, reducing crime. London Prime Minister's Strategy Unit, London

Carter P (2007) Securing the future — proposals for the efficient and sustainable use of custody in England and Wales. Ministry of Justice, London

Centre for Social Justice (CSJ) (2009) Locked up potential: a strategy for reforming prisons and rehabilitating prisoners. Centre for Social Justice, London

Civil Society Finance (2011) http://www.civilsociety.co.uk/finance/news/content/10555/charities_that_run_jails_should_lose_charitable_status_says_howard_league_chief

Disley E, Rubin J, Scraggs E, Burrowes N, Culley D (2011) Lessons learned from the planning and early implementation of the Social Impact Bond at HMP Peterborough, vol 5/11, Ministry of Justice Research Series. Ministry of Justice, London

Fox C, Albertson K, Warburton F (2011) Justice reinvestment: can it deliver more for less? How J Crim Justice 50(2):119–136 (18)

Home Office (2002) Justice for all. Home Office, London

Lanning T, Loader I, Muir R (2011) Redesigning justice reducing crime through justice reinvestment. IPPR, London

Local Government Association (LGA) (2005) Going straight reducing re-offending in Local Communities. LGA, London

Local Government Information Unit (LGIU) (2009) Primary justice an inquiry into justice in communities. LGIU, London

London Criminal Justice Partnership (LCJP) (2011) An evaluation of the diamond initiative: year two findings. London Criminal Justice Partnership, London

McLeish H (2008) Scotland's choice report of the Scottish Prison Commission. Scottish Prison Commission, Edinburgh

Ministry of Justice (2009) Story of the prison population 1995–2009. Ministry of Justice, London

Ministry of Justice (2010) Breaking the cycle effective punishment, rehabilitation and sentencing of offenders. The Stationery Office, London

Ministry of Justice (2011a) Breaking the cycle: government response. The Stationery Office, London

Ministry of Justice (2011b) Offender services competition strategy. The Stationery Office, London

Ministry of Justice (2013) Transforming rehabilitation. The Stationary Office, London.

National Audit Office (NAO) (2008) The National Probation Service: The Supervision of Community orders in England and Wales

Parliamentary Select Committee on Justice (2008) Towards effective sentencing HC 184–I fifth report of session 2007–08. The Stationery Office, London

Parliamentary Select Committee on Justice (2010) Cutting crime: the case for justice reinvestment HC 94-I, first report of session 2009–10, vol I. The Stationery Office, London

Parliamentary Select Committee on Justice (2011) The role of the probation service HC 519-I eighth report of session 2010–12. The Stationery Office, London

Rethinking Crime and Punishment (RCP) (2008) The Manifesto Esmee Fairbairn Foundation. London

Solomon E, Allen R (2009) Reducing child imprisonment in England and Wales – lessons from abroad. Prison Reform Trust, London

Toynbee P (2011) The guardian July 4th. http://www.guardian.co.uk/commentisfree/2011/jul/04/sub-prime-behaviour-bonds-fools-gold

Juvenile Awareness Programs

► Scared Straight Programs

Juvenile Crime

► Juvenile Violence

Juvenile Diversion

Simon I. Singer
School of Criminology and Criminal Justice,
Northeastern University, Boston, MA, USA

Synonyms

Inclusionary control; Rehabilitation, treatment; Rhe

Overview

Juvenile diversion refers to dispositional options that attempt to avoid the stigma of an official delinquent or criminal label (Lemert 1981). It is an overly general term because it can apply to virtually any programmed response that avoids the formality of the justice system. Historically, the exact shape of juvenile diversion has depended on the availability of various treatment and punishment options in a particular time and place. In earlier times, juvenile diversion meant the *reformatory* as opposed to the *penitentiary* and then later the *juvenile* court as opposed to the *criminal* court (Schlossman 1977). More recently, juvenile diversion refers to the avoidance of juvenile court; it has become a court from which youths are to be diverted *from* rather than a court to which youths are to be diverted *to* (Miller 1979). Moreover, the types of youths eligible for juvenile diversion tend to be those who are considered most likely to succeed through its various dispositional options (Cicourel 1968; Cohen 1985). These options include a range of programmatic and non-programmatic responses to common acts of delinquency by adolescents (Mears 2012). The effectiveness of these various diversionary options is an important area of policy-relevant juvenile and criminal justice research (Shelden 2008; Patrick 2005; Cocozza et al. 2005; Chapin and Griffin 2005).

This entry begins with a brief historical review. It then proceeds in section "Definitions of Juvenile Diversion" with a definition of juvenile diversion. Section "Who is Likely to be Diverted" discusses the most likely candidates for juvenile diversion. Sections "Evaluating Successful Forms of Diversion" and "An Example of Diversion" provide an example of juvenile diversion and figures on diversion based on police arrest and juvenile court intake statistics. The concluding section presents future directions for the evaluation of juvenile diversion.

In the Beginning

The idea of diverting juveniles from the justice system appears to be as old as juvenile justice itself. Children were always considered as a category of their own. Early nineteenth-century industrialization and urbanization extended childhood into an older-age category that would be known as adolescence (Empey and Stafford 1999). This shift necessitated the need for juvenile justice. The beginnings of juvenile justice can also be considered the basis for diversionary justice. Under the common law doctrine of *parens patriae*, the state could act as a parent in deciding the best interests of the child even if the child turned out to be an older adolescent (Schlossman 1977). Constitutional requirements for a fair hearing were avoided merely through a judge ordering a juvenile to be placed in the care of the state. As the age of compulsory school education began to extend into older-age categories, officials required a backup diversionary system of governmental control, for example, youths who refused to attend school.

During the 1820s, houses of refuge thus appeared in large cities like New York City as a way of enforcing various status-offending categories, such as not going to school or hanging out aimlessly on the city streets. The house of refuge was the place to divert such youths not only for status-offending acts but also for common acts of theft (Pickett 1969: 20). It was framed as a residential school for the treatment of unruly youths. Most importantly, however, houses of refuge exemplified the beginnings of juvenile diversion because they were considered a more reasonable alternative to the city jails and emerging state penitentiaries (Schlossman 1977). As such, they represented a profound shift in the way that the public and its officials thought about punishment; it was no longer to be based on exclusionary punishment but inclusionary control. Juvenile diversion from criminal justice was to foster the normal return of an adolescent once he had experienced the redemptive and formative assistance of a residential school in the shape of a house of refuge, reformatory, or training school.

But by the end of the nineteenth century, diversion in the shape of out-of-home placement was no longer viewed as desirable. Instead, reformatories became known as ordinary places of imprisonment that could turn an adolescent into a serious adult criminal offender. Reports surfaced about how the house of refuge moved from just being a residential school to a brutal place of punishment (Pickett 1969). A new face to juvenile diversion was needed; this would take the form of a juvenile court. Zimring (2002) in relating the early juvenile court to juvenile diversion notes Judge Tuthill's 1904 explanation of why the juvenile court could save juveniles from a life of crime. Tuthill explained that prior to the juvenile court, delinquents were "educated in crime and when discharged were well fitted to become the expert criminals and outlaws who have crowded our penitentiaries and jails. The state had educated innocent children in crime, and the harvest was great" (145). The juvenile court could divert juveniles from this life of crime by providing a more inclusionary form of control.

The preferred means for avoiding a life of crime was through a system of inclusionary control that avoided the isolation, criminalization, and stigma of incarceration. This kind of diversion from punishment could be organized through a juvenile court. The main method for doing so was in the shape of probation. The history of the juvenile court as a diversionary court is the history of probation as a community-based supervisory form of diversion. The advantage was not only to treat the adolescent in the community but also to save the offending juvenile from a criminal identity; that identity was conceived as ongoing, to be shaped through the malleability of childhood as extended into adolescence. The best place to create a respectable, law-abiding adolescent was to be maintained through the treatment-oriented juvenile court and its emphasis on prevention and control within the community and family.

By the beginning of the twentieth century, officials had made the case for a more all-encompassing system of juvenile diversion from criminal justice. The juvenile court would have a multitude of options as its disposal. In large cities, there would be a clinic in which to diagnose troubled youths. There would also be a separate probation department. The emerging social science of adolescence provided the justification for confirming various techniques for treating troubled youths. The stories of juvenile court reformers told of how best to divert adolescents from a life of crime (Levine and Levine 1992).

As a diversionary court of jurisdiction, the main technique of treatment emphasized inclusion over exclusion; it emphasized probation as a first resort for a wide range of offense. Through a system of probation, parents could be guided to be better parents. But the power of probation would extend far beyond supervision. The probation officer became a key official, critical to both intake and outtake processes. As an intake official, probation officers could appear at various stages in the juvenile justice process. Police officers were often assigned the task of acting as intake probation officers (Wolcott 2001). They could easily dismiss cases, bring a child to the child's home and warn the child and the parent that the offending act could lead to an official arrest. As outtake officials, probation officers had the authority to incarcerate adolescents who repeated their offensive behavior.

The diversionary techniques of the juvenile court at various stages could proceed because it was an insulated, nonpublic legal setting (Miller 1979). The rationale for confidentiality was to protect the identity of adolescents from a public criminal label. This veil of confidentiality worked not only to keep the name of delinquents away from newspaper reporters but also to protect officials from public scrutiny over their decisions to treat and to punish juveniles. The informal nature of the traditional juvenile court became less tolerable as the court appeared to ignore the justice component of juvenile justice. A new wave of reforms would shift the purpose of the juvenile court and the shape of juvenile diversion.

By the 1960s, when critics of the juvenile court were writing about its informal procedures, officials were ready for a new set of diversionary reforms. The US Supreme Court, in a series of decisions, would decide that the juvenile court was much too informal in the way that it adjudicated adolescents (Feld 2000). Constitutional rights needed to be applied to juveniles in a juvenile court, including the right to legal representation. The juvenile court could no longer proceed informally as it had in the past. Several years after the 1964 Gault decision, the President's Crime Commission in 1968 explicitly recommended the diversion of status offenders from a less informal diversionary juvenile court.

In one of the first evaluations to appear on juvenile diversion, Cressey and McDermott (1973) predicted the contemporary shape of juvenile diversion by observing that "there will be a polarization of attitudes and programs: Lawbreaking juveniles are likely to be processed along the lines of the adult model and hence will receive more due process and less humanistic consideration—after all, are they not merely small criminals? Juveniles who have been called 'predelinquents,' because they can't get along at home or in school, will be diverted (61)." Cressey and McDermott's prediction would become true at both ends of the justice system. In less than a decade, waiver legislation expanded the eligible population of juveniles who could be brought into criminal court. Singer (1996) has referred to this hard-end of the juvenile and criminal justice system as *recriminalization*. A segment of juveniles viewed as ineligible for the treatment-oriented mission of the juvenile court are now seen as criminally responsible and deserving of criminal sentencing in criminal court. But this kind of diversion would be at one extreme end of the juvenile justice system, while at another softer end, inclusion would persist through the proliferation of programs that avoided the more formal juvenile court. The informal system of juvenile justice expanded beyond the juvenile court's ability to adjudicate delinquent status (Mears 2012).

Definitions of Juvenile Diversion

Most definitions of juvenile diversion focus on its inclusionary technique of invoking control. As noted, the term diversion has over time encompassed a variety of agencies, institutions, and programs. Diverting juveniles from the juvenile court to the criminal court through waiver legislation is an aberration of terms since it avoids the inclusionary, non-labeling function of juvenile justice (Singer 1996). For juveniles that are deemed criminally responsible for their offenses based on the seriousness of the crime and prior history, diversion is the juvenile court (Feld 2000).

A definition of diversion suggests the following possibilities. First, there is *true* diversion. Edwin Schur (1973) defined true diversion as a form of "radical non-intervention." But any offense that invokes a response may be viewed along a scale of first and last resorts of dispositional options (Emerson 1969). The most frequent source of juvenile diversion is the police (Wolcott 2001). For common acts of delinquency, the police are the first to respond, and they frequently respond by warning adolescents not to repeat the offense. This kind of diversion may be considered non-programmatic. It avoids mandating an intervention that might disrupt the life of the adolescent. Non-programmatic diversion operates on the theory that a warning is sufficient to avert the repetition of an offense.

Juvenile diversion may also be referred to as programmatic. The police, intake officials, and even prosecutors as well as judges may refer juveniles to a variety of treatment options. They may be directly related to the operation of the juvenile court but without any formal adjudication or official penetration into the juvenile justice system. They include a range of options – including drug, youth, and family relation courts (Chapin and Griffin 2005). These newly established centers for diversion provide an alternative way of confronting adolescent offending while allowing the juvenile court to focus on serious acts of delinquency.

Thus, juvenile diversion as typically defined is inclusionary and cost-effective. Juvenile courts when first created as an informal diversionary court were cheaper to operate than the criminal court. Probation is less expensive than incarceration. Today, the informal system of juvenile justice is less expensive than the post-Gault formality of the juvenile court. Juvenile diversion is also defined by its more inclusionary form of control. The house of refuge was more inclusionary than the city penitentiary. Similarly, probation is more inclusionary, less stigmatizing than out-of-home placement. Last but not least, juvenile diversions into drug rehab, stop-shoplifting programs, or any number of treatment-oriented nonformal juvenile justice proceedings are less stigmatizing than adjudication in juvenile court. The most inclusionary form of juvenile diversion is the type that is referred to as non-programmatic in producing a warning rather than a programmed intervention.

Who Is Likely to Be Diverted

Advocates of juvenile diversion would indicate through their mission statements that their programs are for those adolescents who deserve a particular form of treatment. Criminologists who have evaluated diversion see more than just the objective need of the adolescent for treatment. Juvenile diversion tends to be defined by a set of local administrative rules and procedures as well as by the availability of various treatment options (Hasenfeld and Cheung 1985). Criminologists also recognize that these options are not evenly distributed. They depend on the resources of a community and the ability of parents to afford various diversionary options. More affluent communities are in a better position to provide their youths with a range of diversionary options. Similarly, affluent families have at their disposal a range of options in a private system of treatment for their troubled youths. The police may be attuned to the possibilities of true, non-programmatic diversion because they see affluent parents as competent to control their youths. Petty acts of shoplifting and property destruction may be settled civilly, because affluent parents are in a better position to pay the cost of their child's offending.

Closely associated with class indicators of diversion is race. Disproportionate minority confinement in juvenile institutions may be viewed as a function of a lack of suitable alternative diversionary programs. In a report sponsored by the US government's Office of Juvenile Justice and Delinquency Prevention on racial disparities within New York State, the authors concluded that the differences in the race of youths locked in juvenile facilities can be largely attributed to "the greater availability of diversion programs in suburban areas, coupled with the greater affluence of suburban families and their ability to access services for their children and compensate victims, leads to over-representation [of minorities] in the juvenile justice system" (Nelson and Lansing 1997). A variation of the above quote can be found in nearly all federally mandated, state studies of the disproportionate confinement of minority juveniles and numerous scholarly articles and books (Feld 2000, Leiber and Stairs 1999; Ward 2013).

In Stanley Cohen's 1985 book, *Visions of Social Control*, he relates a theory of diversion on who is most likely to receive the more inclusionary justice system's controls. He illustrates soft-end diversionary justice as targeting those who are believed to most likely to succeed. He refers to the acronym YARVIS (Young, Attractive, Rich, Verbal, Intelligent, and Successful) to describe the most likely candidates for diversion. *Younger* adolescents are more likely to be diverted because they are viewed as too naïve to be held fully responsible for their behavior. The *Attractive* may be viewed as those who look more like the children of officials, who would reflect majority class and race characteristics. This distinction is invoked by Jerome Miller (1979) when he suggests that there is a basic split between how some youths are perceived as more or less deserving of treatment. Juvenile diversion reproduces this basic bifurcation in decision-making. Similarly, the *Rich* as noted have options that are simply unavailable to the impoverished. Therapeutic schools can easily exceed the cost of the most expensive private colleges, and only the wealthy can afford this kind of juvenile diversion. Moreover, professional therapy is generally a luxury for those who have insurance to cover its cost or who can pay for the out-of-pocket expense of a therapist. The *Intelligent* have always maintained an advantage in that they can articulate their troubles and the reasons for their delinquency. Their prognosis for success is generally considered better than those who suffer from a range of intellectual and learning disabilities.

Evaluating Successful Forms of Diversion

Determining the rate of success of juvenile diversion is a difficult task. The criteria for selecting youths can confound any evaluation of the true effectiveness of a treatment program. The ability to assign youths to programmatic and non-programmatic treatments requires an experimental design that cannot easily be implemented. Programmatic and non-programmatic efforts at diversion may be viewed as desirable independent of any proof of recidivism. Judgments about who is eligible for a diversion are often based on judgments about the probability of success; creating selection effects can easily confound experimental and quasi-experimental research designs. Moreover, there may be long-term effects to juvenile diversion that cannot easily be detected through short-term studies.

The success of programmatic efforts at juvenile diversion most often rests on the sense that officials have of a particular policy. A diversionary program may make sense because it is cost efficient when compared to alternatives and when it reflects the desire of officials and the public to see a minimum sanction applied. A litany of sensemaking approaches to understanding organizational structure (Weick 1995) and how juvenile diversion makes sense in contrast to a range of alternatives is the reason for programmatic and non-programmatic responses. So houses of refuge made sense as a residential school for troubled youths in contrast to the city penitentiary. Similarly, probation made sense as supervised care within the community in contrast to the reformatory. Last but not least, the informal system of juvenile justice today makes sense when contrasted to the more formal procedures of the juvenile court. The litany of juvenile diversionary responses is generally considered appropriate for more deserving youths as subjectively defined by officials in their case processing decisions. Juvenile diversion will continue to proliferate in a multitude of forms independent of any proof of success in terms of recidivism. Instead, success is defined in relation to a range of possible alternatives.

An Example of Diversion

As stated in the previously cited government report, juvenile diversion is more likely to take place among youths in affluent communities and affluent families. The following unedited letter from an affluent youth in an affluent suburb illustrates the point of diversion:

> Dear Chief Probation Officer Riley:
>
> I would like to point out how thankful I am for being granted the opportunity to make right of what I have done wrong. The pre-trial diversion program gave me a chance to realize how serious and real breaking the law is. Along with being granted the pre-trial program I was obligated to pay all fees to the court, finish the JDP [Juvenile Diversion Program] program, visit a personal counselor, finish twenty hours community service, and take random drug tests. Because my record was so important to a successful future for me, I completed all the tasks I was allotted to accomplish within six months. In doing so I learned not to get in trouble. I do not plan on ever having to show up to court for anything again. I think every kid with a first offense deserves the pre-trial diversion program. Because ones record is so important, a clean one is important for their future. I am very grateful for the opportunity I was given.
>
> Sincerely, an arrested affluent youth in an affluent suburb

Success is deemed possible because this particular adolescent's family was able to hire a private attorney, and in this affluent suburban town, a juvenile diversion program created the opportunity to avoid adjudication as a delinquent. Community service, drug testing, and therapy enabled this particular youth to meet the conditions of his pretrial diversion. By meeting those conditions, the youth escaped a possible conviction not only in juvenile court but also in criminal court. He could have been charged with a felony and faced prison time for a large quantity of marijuana that was found in his possession along with other drug-related paraphernalia. The fact that he was able to avoid formal adjudication as a delinquent is related not only to his youth but to his affluence, intelligence, and attractiveness as a good candidate for diversion. His diversion might not have been possible if his parents were unable to afford the cost of a private attorney, the time spent in court, and the insurance money for private counseling and professional therapy. The empathetic response of officials in this affluent suburb recognized the youth's adolescence and identified with his parents' desires.

Data on Juvenile Diversion

A large part of the business of understanding diversion involves evaluating various programmatic designs. Various diversionary programs are considered panaceas when first implemented. The reliability and validity of these evaluations are beyond the scope of this entry and are discussed elsewhere. For now, it is important to bear in mind that a range of diversionary programs exists, independent of any proof of effectiveness. They exist because they make sense to officials and the public who support them as alternatives to the expense and stigma of incarceration. Rather than a non-programmatic form of diversion (true diversion), programs like youth and drug courts, restorative justice, first offender programs, and a wide range of other programs are thriving under the banner of diversion.

It is important to bear in mind that the traditional role of the police has been to act as a gateway to juvenile justice. This point has been emphasized by David Wolcott (2001: 363) who concluded that "the police made the juvenile justice system run." This conclusion is based on the filtering function of the police to decide which adolescents would be brought into juvenile court. He examined statistic for Chicago, Detroit, and Los Angeles during three distinct periods of time. The story that he tells is one of juvenile diversion that reflected the organizational practices of the times. It is not an even story given that the availability of options created the possibilities for juvenile justice to take on a variety of diversionary tracks.

In one affluent suburban town which is similar to where the previously mentioned diverted youth resided, data on the disposition of police arrests reveal a familiar story. First, it is important to bear in mind that the police can act as the first source of diversion. They can decide if an offending youth should just be warned or referred to the juvenile court. Figure 1 shows the arrests of juveniles younger than 16 in an affluent suburb. Of the 137 recorded official arrests for acts of delinquency (non-status offenses), only 36 % were referred to the juvenile court (New York's

Juvenile Diversion, Fig. 1 Officially recorded police recorded dispositions of juvenile arrests for 2001, Amherst, New York (Source: Singer 2014)

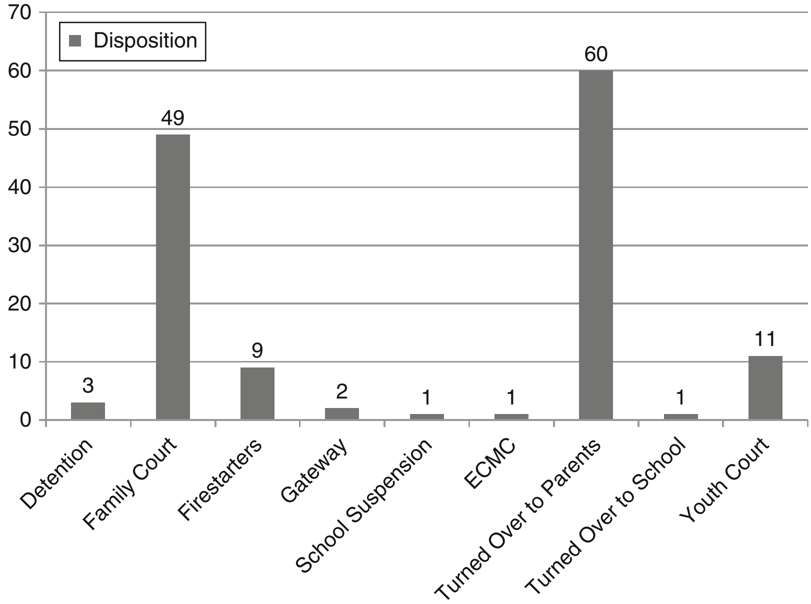

Family Court). The remaining arrests were diverted either to their parents or a variety of treatment-oriented programs. The most common of these was the town's youth court. This type of diversion avoids any kind of juvenile court record, and its most severe sanction is of the community service type.

But what happens if a juvenile is referred to the juvenile court? Once juveniles are referred to juvenile court, they may receive diversion into any number of programs through intake independent of the formal hearing. Figure 2 presented national court statistics for case processing decisions in 2008 based on nationwide juvenile court statistics as compiled by the National Center for Juvenile Justice. Of those who were not formally processed in the juvenile court, most were given a range of sanctions including probation. About 42 % were dismissed or released. This non-programmatic form of diversion as exemplified by true diversion is after the fact that in some jurisdictions the police have already acted as filtering agents in only bringing a small proportion of youths with whom they make a recorded contact into the juvenile court.

Figure 2 is for the entire USA; it does not include the wide array of programs to which adolescents can be placed as a condition of their diversion in affluent suburbs such as the one represented by the graph of arrests in Fig. 1. Informal dispositions should be higher in more affluent counties and towns. In this particular affluent suburb, youths may be eligible for its Drop-Out and Early School Intervention, First Offender Youth Board program, Serving at Risk Adolescent Homes program, Responsible Adolescent Parenting program, and the nationally recognized Tough Love program. Many of these programs are family based and require family therapy as a way of confronting an adolescent's offensive behavior.

The impact of these programs is difficult to assess except in its most general form. There are advantages and disadvantages to non-programmatic forms of juvenile diversion. For Lemert and other theorists of labeling, these programmatic forms of diversion had the risk of isolating youths with other youths who were equally delinquent.

Future Directions

Programmatic efforts to deal with offending youth exist to respond to trouble. Thus, qualitative research by Robert Emerson (1969) has defined trouble along a continuum of first resort and last

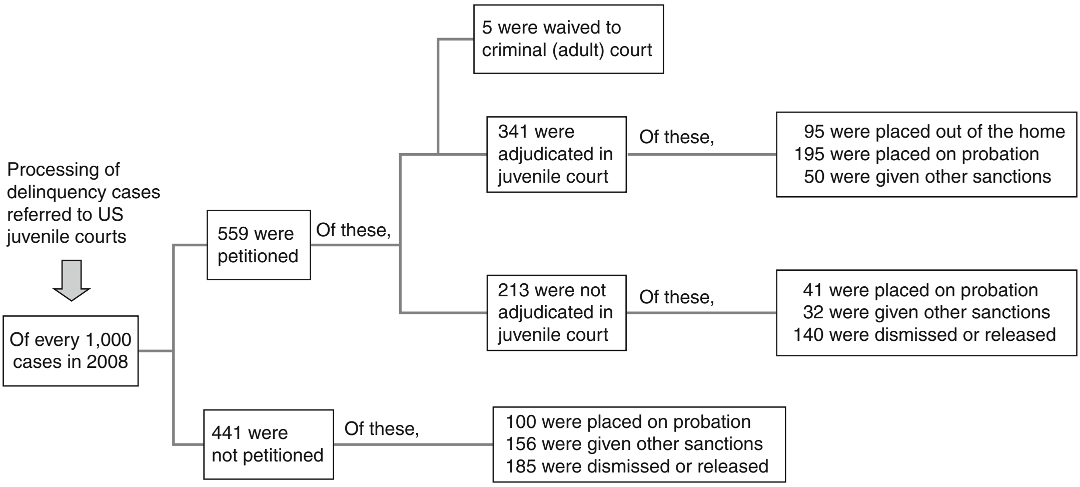
Processing of delinquency cases referred to US juvenile courts
Of every 1,000 cases in 2008
559 were petitioned
Of these,
5 were waived to criminal (adult) court
341 were adjudicated in juvenile court
Of these,
95 were placed out of the home
195 were placed on probation
50 were given other sanctions
213 were not adjudicated in juvenile court
Of these,
41 were placed on probation
32 were given other sanctions
140 were dismissed or released
441 were not petitioned
Of these,
100 were placed on probation
156 were given other sanctions
185 were dismissed or released

Juvenile Diversion, Fig. 2 Juvenile court case processing decisions (2008) (Source: Puzzanchera et al. 2011)

responses. These first and last responses are historically place specific. The house of refuge, as noted, was initially conceived as a first resort, an alternative to the city penitentiary. It became a last resort as new ways of seeing how to respond to adolescents evolved in the shape of probation and its juvenile court. The emerging social science of adolescence has created new opportunities for considering how best to respond to the reasons for adolescent offending. For those jurisdictions that could afford the luxury of a mental health clinic and a professional staff of probation and social workers, there was little need to divert from the juvenile court.

However, a more formal juvenile court brought a more expensive juvenile court – one where lawyers would have to be paid and rules followed similar to criminal court. This was to reduce the risk of injustice that might be caused by sentencing a juvenile as delinquent for a crime that they did not commit. The Gault decision was to safeguard the risks of a too informally driven juvenile court. But this informality would be extended in an expanding front-end of juvenile justice where most adolescents would receive some sort of sanction based on the assumption that diversion was more just for its treatment protocol and less stigmatizing than formal handling in juvenile court.

This assumption of the justice of juvenile diversion needs to be examined in its organizational and legal sense of justice. The decentering of the contemporary juvenile court became so blatant that Jerome Miller would refer to the juvenile court as a setting from which juveniles were to be diverted. This was not always the case as previously noted. But the reasons for diversion need to be examined based on (1) the sense of justice that juvenile diversion produces in an informal system of juvenile justice, (2) the risks that the treatment will be just as stigmatizing as formal processing, and (3) the meaning of the treatment for officials and the public, that is, how success and effectiveness are to be defined.

Edwin Lemert's (1972) theory of secondary and primary deviance is often cited in textbooks as the reason for diversion. But this can be disputed based on Lemert's writings that avoided the simplistic notion of a labeling theory. Trouble was identifiable in the shape of an offense. The method of reacting to that offense defined the type of control that would be deemed as most effective. The courts can make a bad situation worse, leading to secondary deviance by separating the adolescent from the mainstream of society. This process of social attenuation could occur through an adolescent's repeated identification with delinquents. It is them against us as exemplified by the concept of social attenuation that Lemert warned against. The problem for many programmatic efforts to divert juveniles is they risk further isolating youths by not dealing with their troubling behavior. Moreover, the best place for treatment may not be in the home when the source of trouble is the home, such as in instances of child abuse. How to maintain the social status of troubled youths may require an individualistic and not programmatic response, such as individual therapy.

The social organization of juvenile justice in each community also needs to be revisited to assess the extent to which juvenile diversion matters in the lives of adolescents. This entry has referred to research on programs in one affluent community and a government report that points to disparities in diversion based on the availability of services. Affluence matters both on community and familial levels. Cicourel (1968) made this point in his study of the social organization of juvenile justice when he compared two cities that varied in their policing and official responses to delinquency. The less affluent city was less professional, and officials were more inclined to make arrests rather than to negotiate diversion. The status of parents was considered equally important as well as the possibilities that parents were willing to negotiate a disposition.

To look deep inside of juvenile justice systems is to examine the reasons for a veil of confidentiality that is there not so much to protect the status of adolescents but to protect officials from public scrutiny of their decisions. Juvenile justice in general may be considered a black box of decision-making, and this makes decisions regarding juvenile diversion difficult to trace. Officials must be convinced about the importance of transparency in their decision-making along with the rights of adolescents to have their identities protected.

Criminologists can provide that independent evaluation. But it requires the hard work of gathering the intimate details of administrative decision-making. Criminologists should also look at the extent to which diversion is supported independent of concerns about offender recidivism.

Conclusion

Juvenile diversion is an integral aspect of our justice system. It was the basis for houses of refuge, reformatories, juvenile courts, drug courts, and youth courts. Juvenile diversion will remain with us as an integral part of our juvenile justice systems. The principle of juvenile justice is rooted in the concept of diversion. But it is a principle that revolves around doing nothing (radical nonintervention) or doing something (programmed treatment). This difference too often neglects the fact that doing nothing often involves at a minimum a warning and announcement that a rule has been violated. Similarly, doing something often turns into doing nothing if the juvenile is inappropriately placed in a program that has little effect on their troubling behavior.

Diversion is not only to be found everywhere in a complex system of juvenile justice, but its basic characteristics are closely identified with the notion that adolescents are deserving of another chance. Yet eligibility for diversion is not evenly distributed based on the resources of the community where an adolescent lives and the status of their parents to afford a number of possible options.

The most simplistic approach is to consider diversion as a first resort in a system of justice where individuals are given a multitude of chances. Yet not only are these chances dictated by a prior history of offending but also by the seriousness of the offense. Generally, the more serious the offense, the less likely the offending adolescent will be eligible for diversion.

Future research needs to examine the mix between place, characteristics of adolescents, and the availability of diversionary programs if they are to be able to fully assess the meaning of juvenile diversion in a complex modern-day system of juvenile justice.

Related Entries

▶ History of Juvenile Justice
▶ Labeling: History and Concept
▶ Law, Diversion and Community Sanctions in Juvenile Justice

Recommended Reading and References

Chapin DA, Griffin PA (2005) Juvenile diversion. In: Heilbrun K, Goldstein NES, Redding RE (eds) Juvenile delinquency: prevention, assessment, and intervention. Oxford University Press, New York, pp 161–178

Cicourel A (1968) The social organization of juvenile justice. Wiley, New York

Cocozza J, Veysey BM, Chapin DA, Dembo R, Walters W, Farina S (2005) Diversion from the juvenile justice system: the Miami-Dade Juvenile Assessment Center Post-Arrest Diversion Program*. Subst Use Misuse 40:935–951

Cohen S (1985) Visions of social control. Polity Press, Oxford, UK

Cressey D, McDermott RA (1973) National assessment of juvenile corrections. In: National assessment of juvenile corrections. University of Michigan, Ann Arbor

Dannefer D, Schutt RK (1982) Race and juvenile justice processing in court and police agencies. Am J Sociol 87:1113–1132

Emerson R (1969) Judging delinquents: context and process in juvenile court. Aldine, Chicago

Empey LT, Stafford MC (1999) American delinquency: its meaning & construction. A Division of Wadsworth, Belmont

Feld BC (2000) Legislative exclusion of offenses from juvenile court jurisdiction: a history and critique. In: Fagan JE, Zimring FE (eds) The changing borders of juvenile justice: transfer of adolescents to the criminal court. University of Chicago Press, Chicago, pp 83–144

Gelsthorpe L, Nellis M, Bruins J, Van Vliet A (1995) Diversion in English and Dutch juvenile justice. In: Fennell PE, Harding C, Jorg N, Swart B (eds) Criminal justice in Europe: a comparative study. Clarendon, Oxford, pp 199–226

Hasenfeld Y, Cheung PPL (1985) The juvenile court as a people-processing organization: a political economy perspective. Am J Sociol 90:801–824

Klein MW (1979) Deinstitutionalization and diversion of juvenile offenders: a litany of impediments. Crime Justice 1:145–201

Leiber MJ, Stairs JM (1999) Race, contexts, and the use of intake diversion. J Res Crime Delinq 36:56–86

Lemert EM (1972) Human deviance, social problems, and social control (edited by Lemert EM). Prentice-Hall, Englewood Cliffs

Lemert EM (1981) Diversion in juvenile justice: what hath been wrought. J Res Crime Delinq 18:34

Levine M, Levine A (1992) Helping children: a social history. Oxford University Press, New York

Mears DP (2012) The front end of the juvenile court: intake and informal versus formal processing. In: Feld BC, Bishop DM (eds) The Oxford handbook of juvenile crime and juvenile justice. Oxford University Press, Oxford/New York, pp 573–605

Miller JG (1979) The revolution in juvenile justice: from rhetoric to rhetoric. In: Empey ELT (ed) The future of childhood and juvenile justice. University Press of Virginia, Charlottesville, pp 66–111

Nelson JF, Lansing SE (1997) Disproportionate minority confinement in New York State. New York State Division of Criminal Justice Services, Albany

Patrick S, Marsh R (2005) Juvenile diversion: results of a 3-year experimental study. Crim Justice Policy Rev 16:59–73

Pickett RS (1969) House of refuge; origins of juvenile reform in New York State, 1815–1857. Syracuse University Press, Syracuse

Puzzanchera C, Adams B, Sickmund M (2011) Juvenile court statistics 2008. National Center for Juvenile Justice, Pittsburgh

Rojek D, Erickson M (1981) Reforming the juvenile justice system: the diversion of status offenders. Law Soc Rev 16:241

Schlossman SL (1977) Love & the American delinquent: the theory and practice of “progressive” juvenile justice, 1825–1920. University of Chicago Press, Chicago

Schur EM (1973) Radical nonintervention: rethinking the delinquency problem. Prentice-Hall, Englewood Cliffs

Shelden RG, Macallair D (2008) Juvenile justice in America: problems and prospects. Waveland Press, Long Grove

Singer SI (1996) Recriminalizing delinquency: violent juvenile crime and juvenile justice reform. Cambridge University Press, Cambridge/New York

Singer SI (2014) Trouble in suburbia: delinquency and modernity in America’s Safest City. New York University Press, New York

Ward GK (2013) The black child-savers: racial democracy and juvenile justice. University of Chicago Press, Chicago

Weick KE (1995) Sensemaking in organizations. Sage, Thousand Oaks

Wolcott D (2001) “The Cop Will Get You”: the police and discretionary juvenile justice, 1890–1940. J Soc Hist 35:349–371

Zimring FE (2002) The common thread: diversion in the jurisprudence of juvenile courts. In: Rosenheim MK, Zimring FE, Tanenhaus DS, Dohrn B (eds) A century of juvenile justice. The University of Chicago Press, Chicago, pp 142–157

Juvenile Justice

► Juvenile Justice System Responses to Minority Youth

Juvenile Justice in the Get Tough Era

Donna M. Bishop[1] and Barry C. Feld[2]
[1]School of Criminology and Criminal Justice,
Northeastern University, Boston, MA, USA
[2]University of Minnesota Law School,
Minneapolis, MN, USA

Overview

During the last quarter of the twentieth century, a conservative reform movement took hold that quickly came to dominate the national debate over juvenile justice policy. Legislators in virtually every state enacted punitive legislation that challenged foundational principles of the juvenile justice system and, for a time, appeared to threaten its continuation. This entry considers this "Get Tough Era" in juvenile justice – the factors that brought it about, the changes it introduced, and its legacy for the future.

In order to appreciate just how radical a departure the contemporary Get Tough movement represents, it is helpful to recount a bit of history.

Juvenile Court and the Progressive Era: A separate justice system for juveniles took shape over the course of the nineteenth century and came to fruition in 1899 with the founding of the juvenile court in Cook County (Chicago), Illinois. Beginning in about 1820, industrialization, urbanization, and immigration transformed what was then a largely rural America: Factories sprang up from the Eastern seaboard to the Midwest, drawing internal migrants from rural areas and massive waves of immigrants from Western and Eastern Europe. In a single generation, small towns like New York and Chicago grew into densely inhabited and chaotic cities characterized by a host of social problems: slums, infectious disease, alcoholism, mental illness, homelessness, civil unrest, and burgeoning levels of crime and delinquency. In impoverished urban neighborhoods, children living in crowded ghetto tenements took to the streets to play, to scavenge, and to work – as newsboys, shoe shiners, peddlers, beggars, and thieves. The presence of destitute children in the streets, unsupervised and exposed to the perceived evils of urban life, disturbed more affluent residents. By the 1880s, groups of educated, middle-class women (the "Child Savers") came together to address the plight of these endangered (and potentially dangerous) children. They advocated successfully for many reforms, including child labor laws, compulsory school attendance laws, mothers' pensions (the precursor to Aid to Families of Dependent Children), and a specialized court to handle delinquents and children at risk for delinquency (Feld 1999).

Up to this time, children who violated the law were prosecuted in criminal courts and received the same punishments as applied to adults. In urging the removal of children from the purview of the criminal justice system, the Child Savers argued that there are important differences between youth and adults that ought to inform societal responses to youth misbehavior. Drawing from the emerging discipline of child psychology, they portrayed juvenile offenders as immature, dependent, and vulnerable children who require special protection. In contrast to adults, they argued, children are immature not just physically, but mentally: They lack adult capacities for reasoning, moral understanding, and judgment. Because these capacities are the grounds on which attributions of criminal culpability rest, it is unjust to subject youth to adult criminal punishments. Moreover, they contended, whereas adults are relatively fixed as to personality and disposition, children are malleable and especially vulnerable to external influences. Consequently, they ought to be shielded from exposure to unwholesome elements in their neighborhood and home environments and – should they commit crime – from the corrupting influence of adult offenders in jails and prisons. At the same time, young people's malleability makes them especially good candidates for intervention to prevent them from becoming adult offenders. Finally, unlike adults, children are dependent on others to meet their basic needs and to provide guidance and supervision. They argued that when the natural parents are not equal to this task, the state has an obligation to step in and assume this responsibility.

On these grounds, the Progressives urged the creation of a juvenile court that would replace the punitive model of criminal justice with a civil-therapeutic one: The new court would function as a benign, nonpunitive institution with a mission to act "in the best interest of children" to safeguard their welfare and promote their positive development. Based on this optimistic vision, the first juvenile court was established, supported by the common law doctrine of *parens patriae* – the state as parent and arbiter of child rearing. Because the court's mission was benevolent, its framers did not hesitate to draw its jurisdiction broadly to include not only children accused of crime but also those merely "at risk" for delinquency (Tanenhaus 2004). Further, its jurisdiction over children brought under its care extended to the age of majority: A child found within the purview of the juvenile court law might remain under state care and supervision for a brief period or an extremely lengthy one, depending on the nature and extent of his problems and needs.

To protect children, the court closed its proceedings to the public and created euphemisms for the stigmatizing terminology used in criminal courts. ("Petition," "adjudication," and "disposition" replaced "charge," "conviction," and "sentence.") In court, a judge met with the child and his parents, urged the child to admit wrongdoing, gathered information – less about what the child had done than about why he had done it – and developed a treatment plan to address the youth's problems and needs. The treatment most often involved probation or commitment to a correctional "school" for the purpose not of punishment but rehabilitation. Although the court's unregulated procedures had enormous potential for abuse, supporters trusted that the judge would "put himself in the position of a wise, affectionate, and careful parent" (Cardozo 1925) and exercise only benevolent despotism.

Setting the Stage for the Get Tough Era: For about 70 years, the juvenile justice system functioned in relative obscurity without challenge to the lawlessness of its court proceedings or the punitive realities of its correctional institutions. This came to an end in the late 1960s and early 1970s when, in a series of decisions, the Supreme Court compared the juvenile court's lofty rhetoric with the reality of the juvenile justice system in practice and, drawing parallels between the juvenile court and the Star Chamber, concluded that youth were entitled to constitutional procedural safeguards to protect them from erroneous findings of delinquency. Most importantly, the Court afforded juveniles the right to counsel, the right to confront and present witnesses, and the privilege against compelled self-incrimination, and required that offenses be proven beyond a reasonable doubt.

The Court did not endorse punishment by the juvenile courts, nor did it abandon the idea of rehabilitation. Indeed, it encouraged states to continue to pursue the "noble experiment" with which it began. However, by providing youth with procedural protections, the Court shifted the focus of juvenile proceedings from identifying and addressing the causes of a youth's misbehavior to proof that she or he had committed a crime. In so doing, the Court made explicit a relationship between law violations and juvenile court dispositions that had previously been implicit and unacknowledged (Bishop and Feld 2012). Notwithstanding the Court's continued endorsement of the juvenile court's therapeutic mission, its decisions had the effect of creating a hospitable climate for the punitive reforms that would soon follow. Immediately, for example, its decisions made it imperative that prosecutors participate in juvenile court proceedings for the first time. Prosecutors established juvenile divisions, staffed them with traditionally trained attorneys, and infused the court with an adversarial criminal law orientation.

An additional impetus to the Get Tough Era was the publication in the 1970s of a series of widely publicized reviews of research on the (in) effectiveness of treatment programs. Especially influential was the Martinson Report (Martinson 1974), which drew the bold conclusion that "nothing works." Critics responded that the negative results could be explained by research flaws and poor program implementation, rather than by the absence of positive treatment effects. But the critics' cautions – and even the subsequent retraction of the report's conclusion by its authors – received little notice. Instead, the rehabilitative mission of the juvenile court was increasingly viewed with skepticism.

Rising juvenile crime rates further contributed to punitive shifts in juvenile justice policy. Juvenile arrests increased substantially from the mid-1960s until 1980, reinforcing the view that the juvenile justice system was ineffective. Although juvenile crime then declined briefly in the early 1980s, another much more dramatic upswing followed. This one turned out to be a watershed.

Moral Panic and the Get Tough Era: From the mid-1980s to 1994, rates of juvenile gun homicide escalated sharply, especially among minority youth in impoverished areas of the nation's inner cities. The violence received an extraordinary amount of media attention. Images of young black youths wielding guns tapped into racial stereotypes, ignited public fears, and fueled what has been described as a "moral panic" (Chiricos 1994). In a moral panic, the media, politicians, and the public reinforce each other in an escalating alarmist response that exaggerates the magnitude of the threat and produces urgent calls to "do something." In the midst of a panic, legislators rush to contain the threat, frequently by adopting drastic measures that are ill-conceived and overreaching.

The media responded to the urban, predominantly black youth gun violence with heavy and sensationalized coverage. Media portrayals of juvenile offenders shifted from the traditionally rather benign images of puerile and corrigible delinquents to menacing portraits of savvy, ruthless, and morally impoverished "superpredators," whose numbers were forecast to swell rapidly. Political scientist John Dilulio (1995 p. 23) famously claimed that "Americans are sitting atop a demographic crime bomb," while others warned of a coming "bloodbath" of youth violence (Fox 1996).

Politicians fueled public fears for electoral advantage. They pushed violent crime to the top of the political agenda and, wary of being labeled "soft on crime," vied to outdo their opponents in the rush to take a hard line on juvenile offending. They sharply challenged the assumptions about youth that had animated the earlier juvenile court movement. By embracing the imagery of shrewd, adult-like predators, they gave traction to a new, retributive sound bite – "adult crime, adult time." Further, and without any evidence of its efficacy, they claimed that severe punishment of young offenders would protect the public safety.

Get Tough Measures: Legislatures in nearly every state responded by taking steps to facilitate the transfer of greater numbers of juvenile defendants to criminal court for prosecution and punishment as adults. Historically, juvenile court judges had the power to waive youths to criminal court, but it was a power they exercised rarely. A 1966 Supreme Court decision (*Kent v. United States*, 383 U.S. 541) made it more difficult for them to do so. They first had to conduct a "full investigation" and hold an adversarial hearing, then set forth in writing facts sufficient to support a determination that the child was either too dangerous to remain in the juvenile system or was not amenable to treatment. Believing that juvenile court judges were loath to waive even serious offenders, legislators created alternative, expedited transfer methods that altogether circumvented the juvenile court or that sharply constrained juvenile court judges' discretion to retain youths in the juvenile system (Bishop and Frazier 2000).

In many states, lawmakers shifted authority to choose the forum in which a youth would be tried from the juvenile court judge to the prosecutor, who was now permitted to "direct file" a case in either juvenile or criminal court. Most often, prosecutorial direct file laws fail to specify any standards or considerations for decision-making. Direct file is carried out without a hearing, without explanation, and with no opportunity for judicial review. In the 1970s, only two states permitted prosecutorial direct file; by 2000, the number had risen to 15 (Griffin et al. 2011). In addition, most states made transfer "automatic" for certain crimes or offense/prior record combinations, often without regard to the offender's age. Before1970, only eight states had automatic transfer laws, and these most often applied only to murder and other then-capital crimes; by 2000, 38 states had enacted these laws and had expanded their scope to include lengthy lists of offenses. The adoption of "once an adult, always an adult" provisions amplified the effects of these other changes by mandating criminal court processing of all post-transfer offenses. Finally, two states (New Hampshire and Wisconsin) took the most sweeping steps to get tough on juveniles: They lowered the maximum age of juvenile court jurisdiction from the 18th to the 17th birthday, which transformed all 17-year-old juvenile offenders into adult criminals, no matter how minor their offenses. Analysts estimate that, as a result of these reforms, somewhere between 200,000 and 250,000 youth under age 18 were being tried annually in American criminal courts by the turn of the century. Many were neither particularly serious nor particularly chronic offenders, and some had not yet reached their teens (Feld and Bishop 2012).

In the criminal courts, judges did not treat youthfulness as a mitigating factor in sentencing transferred youths. Indeed, they imposed harsher sentences than those imposed on comparable adult defendants. Even extreme punishments were not deemed inconsistent with youth. Until the Supreme Court's decision in 2005 (*Roper v. Simmons*, 543 U.S. 551) to ban capital punishment of offenders under age 18, judges in several states sentenced teens to death. By the turn of the century, nearly ten thousand inmates were serving life sentences for crimes they committed before they were 18. Moreover, more than two thousand were serving sentences of life without the possibility of parole (LWOP) in contravention of the United Nations Convention on the Rights of the Child. The harshness of youth crime policy in the USA clashed with policy and practice in the rest of the world. Juveniles received LWOP sentences in only three other countries, and in these three, a total of about a dozen offenders were serving these sentences (Human Rights Watch/Amnesty International 2005).

Punitive policy changes were not limited to the removal of juveniles to the criminal courts. Legislatures in nearly every state also amended their juvenile code purpose clauses to endorse "punishment," "holding youth accountable," and "protecting the public safety" as new objectives of the juvenile court. Although the traditional goal of rehabilitation remained, it was eclipsed by these other objectives. Legislatures in nearly half the states adopted offense-driven sentencing in their juvenile courts. Some adopted sentencing guidelines to impose presumptive, determinate, and proportional sentences based on age, offense seriousness, and prior record. Others enacted mandatory minimum sentencing provisions that prescribed either a minimum period of confinement or a minimum level of secure placement commensurate with the seriousness of the offense. Finally, some created a "blended sentencing" option that permitted or required juvenile court judges to impose criminal sentences (40 years, at the extreme) on youth adjudicated delinquent for certain crimes. Each of these sentencing methodologies applied offense-based principles of proportionality to rationalize sentencing decisions, increase the penal bite of juvenile sanctions, and demonstrate legislators' toughness (Feld 1998).

Other changes chipped away at special protections that youths in the juvenile system had traditionally enjoyed. Juvenile court proceedings had historically been closed to the public, juvenile records were confidential and subject to later expungement, and at least nonserious juvenile offenders were exempt from photographing and fingerprinting after arrest. During the Get Tough Era, the vast majority of states opened juvenile proceedings to the public, allowed juveniles to be photographed and fingerprinted, and required them to give DNA samples (Szymanski 2009). States increasingly used delinquency adjudications obtained without a jury trial to enhance adult criminal sentences and to subject youths to longer terms (Feld 2003). Especially controversial were decisions to make juvenile sex offenders – even those adjudicated in juvenile court – subject to Megan's Law, which required them to register with police and, in some instances, provided for community notification of their identities, whereabouts, and offenses (Zimring 2004). By the turn of the century, most states permitted or required juveniles convicted of sex crimes to register as sex offenders. In about half of the states, juvenile sex offenders faced a possible lifetime of registration.

The frenzy to "get tough" on juveniles reflected a broader climate of fear of young people – especially other people's children – and a simplistic view of "punishment as panacea" that soon spread to other arenas. In response to escalating youth violence, Congress passed the Gun Free Schools Act of 1994 (20 U.S.C. § 7,151 [2003]), which required that states receiving federal funding mandate expulsion for at least 1 year of any student found in possession of a firearm on school grounds. States and schools responded by adopting zero tolerance policies, often expanding the prohibition beyond possession of firearms to include all manner of weapons or drugs, as well as violation of school rules. In some jurisdictions, school administrators applied these draconian policies broadly and senselessly to children who made innocent mistakes, such as bringing scissors or over-the-counter medications to school. Zero tolerance policies have been linked to increased dropout rates and referrals to juvenile court for behaviors that school officials have traditionally handled informally (American Psychological Association Zero Tolerance Task Force 2008). Further, because these policies have more often been implemented in urban schools, they have had a disproportionate impact on minority youth (Fenning and Rose 2007).

In the 1980s, school departments also began hiring police as School Resource Officers (SROs) to combat drugs, enforce school rules, and teach antidrug and antigang curricula and, in the 1990s, to provide heightened security following some high-profile school shootings (Redding and Shalf 2001). Expanded use of metal detectors and drug sniffing dogs accompanied the heightened police presence. Increasingly, police arrested young violators not for drug and weapons offenses but for disorderly conduct, disturbing the peace, and disruption of school assembly – based on normative adolescent acting-out behaviors, like cursing, clowning, and participating in cafeteria food fights, that used to be handled informally by school officials through school detention or intervention with the youth's parents. The combination of zero tolerance policies and SRO programs resulted in the removal of thousands of youths from mainstream educational environments and produced a dramatic escalation in arrests and referrals to juvenile courts – a phenomenon that has come to be known as the School-to-Prison Pipeline.

In sum, as the twentieth century came to a close, the United States embraced policies that increased the punitive powers of juvenile courts, expanded the reach and bite of transfer laws, and criminalized much ordinary adolescent misbehavior. The harshness with which the United States responded to youthful misconduct was unparalleled, and the punitive fervor showed few signs of abating.

Legacy of the Get Tough Era: At the start of the twenty-first century, the future of juvenile justice looked bleak. But a number of developments in the first decade of the twenty-first century hint at a counterbalance to the repressive policies of the recent past. Beginning in 1994, youth violence plummeted and stabilized in the 2000s at levels not seen since the 1970s (Snyder 2012). This precipitous decline seems to have alleviated the moral panic that sparked the unprecedented spate of punitive legislation, allowing for consideration of more moderate, humane, and sensible policies.

Important new discoveries in developmental psychology and neuroscience have affirmed the Progressives' belief that adolescents are immature, vulnerable, and prone to poor and impulsive decision-making. In a signal development, the Supreme Court gave its imprimatur to notions of adolescent immaturity, impetuosity, and vulnerability and to their legal corollary – lesser culpability – in two decisions that limited the punishments to which even the most serious young offenders can be subject. In 2005, the Court in *Roper v. Simmons* abolished the death penalty for persons convicted of crimes committed before they were 18 years of age. In *Graham v. Florida* (2010), the Court banned sentences of life without possibility of parole for youths convicted of nonhomicide crimes. The Court in *Graham* emphasized youths' immature judgment, reduced self-control, susceptibility to negative peer influences, and ongoing character development and cited recent developments in developmental psychology and neuroscience as the linchpin of its decision. *Roper* and *Graham* are groundbreaking decisions that shatter the superpredator theory of juvenile offending and reaffirm bedrock principles on which the juvenile court was founded.

Signs of disenchantment with harsh punishment as a response to juvenile crime are evident in a number of arenas. Expansion of transfer policies has slowed almost to a halt; in some instances, legislatures have retrenched and modified some of the harshest laws. Most dramatically, the Connecticut legislature recently raised the lower bound of the criminal court's jurisdiction, from 16 to 18, while the Illinois legislature shifted the lower bound of criminal court jurisdiction from 17 to 18 for youths charged with misdemeanors. In 2008, the voters of California defeated Proposition 6, which would have created a presumption that 14-year-olds charged with gang-related felonies are unfit for juvenile court. Some states have recently restricted the offense or offender criteria for transfer or mitigated the criminal sentences to which transferred youths may be subject. Others have taken steps to rein in prosecutorial discretion either by severely restricting the scope of offenses subject to direct file or by requiring prosecutors to provide criteria and assessment procedures for choosing the forum in which youths will be tried (Bishop and Feld 2012).

Recent developments in juvenile corrections also suggest a softening of the punishment agenda. The 1980s saw the establishment of boot camps, first in the adult system (1983), then in the juvenile system (1985). The federal Office of Juvenile Justice and Delinquency Prevention (OJJDP) supported the boot camp initiative and funded an evaluation of three programs in three different states, which showed that boot camp participants had higher recidivism rates than controls. Despite the negative results and allegations of staff abuse, boot camps "caught on" in the same way that other deterrence-based programs have. However, the tide is beginning to turn. Several states have eliminated juvenile boot camps, and a comparison of OJJDP's Juvenile Residential Facility Census for 2006 with that for 2002 shows that the number of boot camps declined by nearly 50 % over that period.

In addition, several states have closed (or been forced by the courts to close) training schools following the disclosure of terrible conditions of confinement (California, New York, Mississippi, Texas, Louisiana). Under court order, California has shut down all but two of its training schools and is slated to close the remainder in the very near future. Texas also shut down four maximum security juvenile facilities and reduced its juvenile institutional population by more than half. Florida not only shut down all of its boot camps but also downsized its training schools and youth development centers and placed a cap of 165 beds on all residential placements. The downsizing that is taking place across the nation is motivated in part by external pressures to close large institutions where conditions of confinement are most problematic, in part by budgetary constraints, and in part by research evidence that smaller institutions and community-based programs are more likely to realize rehabilitative aims.

Recently, the results of a number of public opinion polls have received considerable publicity. They show that, although the public supports punishment of serious juvenile offenders as a means of protecting the public safety, there is much broader public support for rehabilitation than was previously believed (Nagin et al. 2006). The vast majority of survey respondents support treatment as a response to all but the most violent juvenile offenders. Consistent with this public support, there are signs of a revitalized treatment orientation in juvenile justice practice. This is apparent in the rapid proliferation throughout the nation of juvenile drug courts, which harken back in many respects to juvenile courts of a century ago. In drug courts, judges are expected to establish strong supportive relationships with juvenile offenders, to monitor closely youths' progress in treatment, and to apply modest sanctions for noncompliance and strong positive reinforcements for youths' active participation in the court's treatment plan (Butts et al. 2012).

While research has shown that punitive sanctions are largely ineffective in reducing the risk of recidivism, there is considerable evidence that some correctional treatment strategies are effective when implemented well. In the last 15 years, a number of private, state, and federal initiatives have produced reviews of treatment program evaluation research and provided authoritative lists of "what works" (Center for the Study and Prevention of Violence n.d.; Campbell Collaboration 2009; Lipsey 2009) These, in turn, have supported a movement toward "evidence-based practice." In an increasing number of jurisdictions, governmental funding of programs is contingent upon the adoption of evidence-based best practices. The combination of solid public support for rehabilitation coupled with social science evidence of effective treatment methodologies bode well for a return to more traditional, treatment-oriented juvenile justice systems.

Signs of a counterbalance to the repressive policies of the recent past are evident in the Supreme Court's endorsement of views of youth much more in line with those of the Progressives and in the softening of criminal punishments, the modest retrenchment of transfer laws, the introduction of specialized treatment-oriented courts, and the renewed focus on evidence-based treatment in juvenile corrections. This is not to say that the Get Tough approaches of the 1980s and 1990s will soon disappear. Punitive approaches institutionalized through legislative reforms are stubbornly resistant to reversal and are unlikely to be quickly undone. The legacy of the Get Tough Era will surely continue for some time to come.

Related Entries

- ▶ Fear and Punishment
- ▶ History of Juvenile Justice
- ▶ Juvenile Violence
- ▶ Youth Homicide in the United States

Recommended Reading and References

American Psychological Association Zero Tolerance Task Force (2008) Are zero tolerance policies effective in schools? An evidentiary review and recommendations. Am Psychol 63:852–862

Bishop D, Feld B (2012) Trends in juvenile justice policy and practice. In: Feld B, Bishop D (eds) The Oxford handbook of juvenile crime and juvenile justice. Oxford University Press, New York, pp 898–926

Bishop D, Frazier C (2000) The consequences of transfer. In: Fagan J, Zimring F (eds) The changing borders of juvenile justice: transfer of adolescents to the criminal court. University of Chicago Press, Chicago, pp 227–276

Butts J, Mears D (2001) Reviving juvenile justice in a get-tough era. Youth Soc 33(2):169–198

Butts J, Roman J, Lynn-Whaley J (2012) Varieties of juvenile court: nonspecialized courts, teen courts, drug court, and mental health courts. In: Feld B, Bishop D (eds) The Oxford handbook of juvenile crime and juvenile justice. Oxford University Press, New York, pp 606–635

Campbell Collaboration (2009) www.cmpbellcollaboration.org

Cardozo JB (1925) In *Finlay v. Finlay*, 148 N.E. 624,626

Center for the study and prevention of violence. (n.d.) Blueprints for violence prevention. (n.d.) Boulder, CO: Institute of Behavioral Science, University of Colorado. http://www.colorado.edu/cspv/blueprints/index.html

Chiricos T (1994) The media, moral panic & the politics of crime control. In: Cole G, Gertz M, Bunger A (eds) Criminal justice: law & politics, 9th edn. Wadsworth, Belmont

Dilulio J (1995) The coming of the super-predators. Wkly Stand 1(11):23–29

Fagan J, Zimring F (eds) (2000) The changing borders of juvenile justice: transfer of adolescents to the criminal court. University of Chicago Press, Chicago

Feld B (1998) Juvenile and criminal justice systems' responses to youth violence. Crime Justice 24:189–261

Feld BC (1999) Bad kids: race and the transformation of the juvenile court. Oxford University Press, New York

Feld B (2003) The constitutional tension between *Apprendi* and *McKeiver*: sentence enhancements based on delinquency convictions and the quality of justice in juvenile courts. Wake Forest Law Rev 38:1111–1224

Feld B, Bishop D (2012) Transfer of juveniles to criminal court. In: Feld B, Bishop D (eds) The Oxford handbook of juvenile crime and juvenile justice. Oxford University Press, New York, pp 801–842

Fenning P, Rose J (2007) Overrepresentation of African American students in exclusionary discipline: the role of school policy. Urban Educ 42:536–559

Fox J (1996) Trends in juvenile violence: a report to the United States attorney general on current and future rates of juvenile offending. U.S. Department of Justice, Washington, DC
Griffin P, Addie S, Adams B, Firestone K (2011) Trying juveniles as adults: an analysis of state transfer laws and reporting. Juvenile offenders and victims: national report series bulletin (September). U.S. Department of Justice, Washington, DC
Human Rights Watch/Amnesty International (2005) The rest of their lives: life without parole for child offenders in the United States. Amnesty International, New York
Lipsey M (2009) The primary factors that characterize effective interventions with juvenile offenders: a meta-analytic overview. Vict Offenders 4:124–147
Martinson R (1974) What works? Questions and answers about prison reform. Public Interest 35:22–54
Nagin DS, Piquero AR, Scott E, Steinberg L (2006) Public preference for rehabilitation versus incarceration of juvenile offenders: evidence from a contingent valuation study. Criminol Public Policy 5:627–652
Redding R, Shalf S (2001) The legal context of school violence: the effectiveness of federal, state, and local law enforcement efforts to reduce gun violence in schools. Law Policy 23:297–343
Scott ES, Steinberg L (2008) Rethinking juvenile justice. Harvard University Press, Cambridge
Snyder H (2012) Juvenile delinquents and juvenile justice clientele: trends and patterns in crime and justice system responses. In: Feld B, Bishop D (eds) The Oxford handbook of juvenile crime and juvenile justice. Oxford University Press, New York, pp 3–30
Szymanski L (2009) DNA registration of juvenile offenders (2008 update). *NCCJ Snapshot* 14 (1) (January). Pittsburgh: National Center for Juvenile Justice
Zimring FG (2004) An American travesty: legal responses to adolescent sexual offending. University of Chicago Press, Chicago

Juvenile Justice System Responses to Minority Youth

Michael J. Leiber and Jennifer H. Peck
Department of Criminology, University of South Florida, Tampa, FL, USA

Synonyms

Differential offending; Disproportionate minority contact; Juvenile justice; Race; Selection Bias

Overview

The issue of racial and ethnic disparities has been a controversial topic within the juvenile justice system, as the overrepresentation of minority youth is a continued concern. In 2008 alone, minority youth represented 13.3 % of the United States population, yet represented 36.9 % of all cases handled by the juvenile courts. Differential offending (i.e., minorities commit more and more serious delinquent acts compared to Whites) and racial bias among juvenile justice decision-makers are the two main explanations for understanding the relationship between race and juvenile justice system proceedings. The following chapter will demonstrate how the juvenile justice system responds to minority youth through the use of theory, empirical research, and policy. From a review of the literature throughout these three areas, possible reasons for minority overrepresentation will be explained, including specific policy recommendations to lead to the equitable treatment of all youth, regardless of race and ethnicity.

Conceptualization of Racial Bias

Compared to Whites, minorities are disproportionately involved within all stages of the juvenile justice system. They are more likely to be arrested, petitioned, held in secure detention, adjudicated (found guilty), and receive harsher sentences at disposition. One explanation focuses on differential offending, while the other argues that racial bias by juvenile justice decision-makers (police, prosecutors, officers, judges, etc.) influences the way minority youth are treated throughout each stage. Researchers generally acknowledge that both of these explanations may account for the overrepresentation of minority youth in the juvenile justice system. For the purpose of this chapter, the focus will be on racial and ethnic bias as an explanation to explain race/ethnic differences in the juvenile justice system.

Racial/Ethnic Bias

Underlying most empirical research examining race and the adult and juvenile justice systems centralizes on the concept of "racial/ethnic bias." Up until the 1980s, people perceived that blatant, overt, or intentional racial bias and racism occurred against minorities. However, this was not necessarily the case. Zatz (1987) reviewed over 50 years of sentencing research and concluded that from the time of the civil rights movement, there has been a transition from overt to covert racial bias. While overt discrimination was present throughout the 1930s to mid-1960s (race was directly related to more harsh sentencing outcomes), sentencing decisions from the late 1960s to the early 1980s confirmed the effects of race were more implicit, subtle, or subconscious. Recent research has found the presence of unconscious negative race stereotypes that result in Black youth being viewed as more blameworthy, resulting in harsher outcomes in juvenile justice proceedings, while overt or conscious race bias had little impact (Graham and Lowery 2004). However, subtle bias is no less harmful than overt bias.

Examinations of the existence of implicit bias have found that that not including African Americans, every racial group unconsciously associated "African American" with the terms crime or danger. Specifically within the juvenile justice system, probation officers have been found to inhibit a form of implicit racial bias, by describing Black offenders as actively making poor and destructive choices to maintain a criminal lifestyle (Steen et al. 2005). However, this thought process was not extended to White youth. Additional research has also suggested that juvenile justice decision-makers perceive Hispanic youth as a social threat, therefore are treated more severe within court proceedings.

From this overt to hidden transition, racism or racial bias was evident through the effects of race that operate indirectly or in interaction with other conditions (e.g., age, employment status, prior record, family situations). For example, age and employment status may seem race neutral but can increase racial disparities at numerous decision-making stages (Leiber and Johnson 2008). Within the juvenile justice system, age is considered a mitigating factor due to the belief that younger youth lack *mens rea* (intent) due to immaturity, inexperience, and inability to resist peer pressure. From this, older youth are seen as more responsible and handed more formally than younger youth, who receive a "youth discount." However, Leiber and Johnson (2008) found that White adolescents received a "youth discount" at intake, yet similarly situated Black youth were referred to further proceedings. Therefore, the "youth discount" did not extend to African American youth, regardless of the age criterion supported by the juvenile court.

Another example of race effects that are masked by other conditions is the research by Bishop and Frazier (1996) who found by juvenile justice officials perceived single-parent minority families as more broken and dysfunctional than single-parent White homes. In sum, additional research has found that both African Americans and Hispanics are treated more severely than similarly situated Whites, even though other legal and extra legal factors (including the gender of the decision-maker) were taken into consideration (Bishop et al. 2010).

Contexts of Juvenile Justice Decision-Making

Traditional Theories

Prior research examining race and decision-making in the juvenile justice system have traditionally explained minority overrepresentation based on consensus or conflict approaches. Established by Emile Durkheim, the consensus approach is based on the idea that laws, punishment, and treatment are formed from a broad consensus of societal norms and values. From this, racial bias in the criminal justice system is seen as a random occurrence for the reason that laws bound the discretion of decision-makers to follow strict legal criteria. In other words, race differences in offending are due to minorities' differential involvement in crime.

The conflict approach assumes the opposite of the consensus approach, which argues that there is a lack of consensus within society. According to the conflict tradition, majority groups utilize their power to exert control over minority groups. By controlling the "powerless" groups, majority groups are able to protect their own interests and power by controlling the political, economic, and social realms of society, including the law, law enforcement, and the court system. Majority groups exercise social control over the powerless group, which results in minorities becoming labeled as outsiders, deviants, delinquent, and criminal, and therefore subject to harsher sanctions.

The labeling process is central to one theory within the conflict tradition, labeling theory. The labeling perspective stresses that minorities are stereotyped based on negative labels and the inability to acquire necessary resources, which results in differential treatment compared to Whites. Juvenile justice decision-makers may take into consideration negative images and stereotypes of minorities (e.g., dangerous, drug users) to assess culpability, blameworthiness, or treatment possibilities (Graham and Lowery 2004).

A criticism related to both conflict theory (deemphasizes the role of stereotypes) and labeling theory (deemphasizes the role of power) is the inability of both theories to account for research that has shown equitable treatment for both minorities and Whites or instances where Whites are treated more harsh. Due to the economic orientation of conflict theories, the majority and minority groups were based on class, ignoring the specific effect of race. In other words, race effects were masked by economic status, and the reason for differential treatment was due to Blacks being disproportionately poor and powerless (Blalock 1967). Consensus theory also been critiqued due to its inability to conceptualize the complexities between structural and interactional characteristics within race and decision-making (Sampson and Lauritsen 1997). From these criticisms, recent research has begun to include the role of contextual and structural factors into the understanding of decision-making outcomes within the juvenile justice system (Rodriguez 2010; Sampson and Laub 1993). The types of contextually situated theories expand beyond consensus, conflicts, and labeling perspectives to take into account structural characteristics that may also influence decision-making.

Macrolevel Theories

As discussed earlier, the emergence of contextual approaches to understanding minorities and differential treatment compared to Whites was due to shortcomings from consensus and conflict perspectives. In the sections that follow, both macro- and microlevel theories are examined within the context of juvenile justice decision-making with possible reasons for minority overrepresentation. First, macrolevel theories are discussed and can be identified by a focus on structural and contextual characteristics of neighborhoods and communities. Crime rates, impoverishment, minority representation, community urban-rural composition, and political orientation are some of the factors that research has found to influence decision-making within juvenile courts. Second, microlevel theories are examined utilizing evidence focused at the individual level for racial disparities within the juvenile justice system, including race stereotypes, decision-makers use of discretion, and attribution bias.

Urbanism and Formal Rationality

The theoretical foundation for much of the earlier research that examined community characteristics and criminal justice decision-making was the work by Max Weber (1969). His research called attention to the interrelationships between urbanization and bureaucratization and its influence on social control. Weber wanted to understand why and how people orient themselves to the legal system based on their own values and rules. He argued that the two institutions that regulate and guide people within a society are the economy and legal system. The link between the two systems is formal rationality, which refers to universally applied laws, standards and rules. Weber believed that in a capitalist society, all decision-making within the formal rational legal system is grounded the rules themselves, therefore the law is applied to similarly situated individuals in an equal manner based on legal criteria. From this, discretion on the part of the decision-maker is limited. Once differences in crime severity are taken into account, few differences should exist between minorities and Whites concerning their contact and involvement with the criminal justice system.

In contrast to formal rationality, there is also informal or "substantive" rationality, which refers to a plethora of external values including beliefs about justice and respect. In this situation, decision-making depends more on informal criteria and social tradition, instead of an equal manner based on legal criteria (formal rationality). Within this type of legal system, discretion is enhanced among decision-makers, concluding that similarly situated individuals may be treated differently. For example, in a more rural environment, minorities should experience higher rates of involvement in the criminal justice system compared to Whites, due to decision-makers reliance on racism, stereotypes, and other subjective discretionary factors.

However, other theorists have argued that urban courts will demonstrate greater racial discrimination than rural courts. According to this perspective, decision-makers who adhere to a formal rationalized legal system may not question the procedure or substance of the law but will accept them as binding and therefore rely on factors that may be biased (e.g., prior record, family situations). Also, because urban courts are bombarded with a large number of cases, decision-makers might classify offenders into routine types. In doing this, the types are based on stereotypes about minorities, the poor, moral character, motivation, and behavior (Gaarder et al. 2004). Therefore, the influence of race and other extralegal variables (e.g., stereotypes) on more severe outcomes will be greater in urban than rural settings. Overall, research is mixed within these varying perspectives. Some research has revealed that rural courts are less formal and more discriminatory against minorities, while other research has shown that urban courts are more likely to be punitive and influenced by racial stereotypes.

Minority Group Threat/Racial Threat Theory

Another theoretical perspective important for understanding the influence of structural characteristics, race, and social control is Blalock's (1967) minority group power threat thesis. This perspective is in contrast to the traditional interpretation of conflict theory, where the focus is on the perceived or actual threat that minority groups pose on majority/advantaged groups. From this, the interchangeable terms "racial treat" or "minority group threat" are contained within a general "group threat" orientation of the traditional conflict theory. Within this group threat orientation are two interrelated factors: the size of the minority population and the economic situation of minorities compared to the majority/advantaged group.

Blalock (1967) argues that the larger the proportion of the minority group population, the greater the competition for economic resources (jobs, money, property, prestige), which results in perceived challengers to the majority group's status. Also argued by Blalock is that increases in income and wealth by the minority group compared to Whites should make the majority group (Whites) feel more threatened. Prejudicial and discriminatory attitudes and practices will be employed by the majority/dominant group in order to diffuse the minority group threat.

However, research examining the power threat thesis within the context of the juvenile justice system has shown mixed results. Some research has found that police within communities with a higher proportion of minorities have responded in a prejudicial manner towards minority youth, yet the bias was corrected in later stages of juvenile justice proceedings (Bishop et al. 2010). Other research has found that racial composition of a community did not have effects on the level of minority youth confinement.

In 1992, Frazier and colleagues tested Hawkins version of the power threat thesis (a revised conflict approach) who believed that the historical contexts of race and punishment should be included within Blalock's (1967) power threat thesis. However, Frazier and associates argued that Hawkins' thesis was directly opposite of numerous traditional conflict theories. In traditional conflict theories, a lower proportion of minorities allows the majority group to exert social control, yet Hawkins argued that a higher proportion of minorities would lead to more political and economic threat towards the majority group, resulting in the advantaged group to exert social control. Results from Frazier et al. (1992) supported the traditional conflict perspective rather than the power threat thesis, indicating White majority populations exert social control over minorities.

Theory of Inequality and Social Control

An integrated perspective including structural contexts and racial stereotyping based on Tittle and Curran's (1988) symbolic threat thesis, the war on drugs, and increased sanctioning of juvenile offenders was developed by Sampson and Laub (1993). They argue that the poor, underclass, and minorities will be perceived by juvenile justice decision-makers as threatening and in need of social control from residing in communities with high levels of economic and racial inequality. In line with Tittle and Curran (1988), Sampson and Laub (1993) emphasized that decision-makers view minority youth as aggressive, sexual, and lacking discipline. In conjunction with certain social psychological emotions of juvenile court officers and how these emotions interplay between the perceived characteristics of minority youth, feelings of fear and jealousy from juvenile court officers are manifested into beliefs that minority youth pose symbolic threats to public safety and middle-class standards (Sampson and Laub 1993, pp. 289–290).

The interplay between the war on drugs, the use of stereotyping, and disadvantaged structural characteristics further refined the symbolic threat thesis at the macrolevel. More specific, Sampson and Laub (1993) emphasized the evolving stereotype of minority males as drug users and dealers. Due to two trends within the 1980s (the dramatic increase in incarceration of Black males and drug offenders), this prominent stereotype leads to the overall effect that poor, Black men (especially those who were involved with drugs) would be subject to increased social control by the juvenile justice system.

Support for this perspective was found in Sampson and Laub's (1993) examination of over 200 counties across the United States. Results indicate that measures of underclass poverty and racial inequality were significantly related to increase sanctioning of juvenile offenders especially within the stages of predisposition detention, and out-of-home placement. These results were also more pronounced for African Americans and to some degree, those involved in drug offenses. Additional support for Sampson and Laub's (1993) perspective was found in the research by Leiber (2003), including suggestions for the refinement of the theory to encompass aspects of historical organization, sociopsychological processes, and the role of stereotyping involving the family and respect for the court. Additional research within the context of Sampson and Laub's (1993) perspective is the work by Rodriguez (2010), who found that racial and ethnic biases exist throughout numerous stages of the juvenile justice system. Black and Hispanic youth were more likely to be detained compared to White youth. Also important, youth who lived in structurally disadvantaged areas were more likely to be detained, and indirectly were treated more severely in later stages of juvenile justice proceedings. These results show the interrelated nature of race, structural disadvantage, and the juvenile justice system. Other research has shown race to have an indirect effect on case outcomes through detention that further results in a cumulative disadvantage as detained youth move on from intake to adjudication to disposition and receive more severe case outcomes than non-detained youth (Holman and Ziedenberg 2006, p. 4). In other words, decisions made at earlier stages (such as detention) affect outcomes at later stages and in particular judicial disposition. This means that being detained strongly predicts more severe treatment at judicial disposition. From this, if racial bias occurs at earlier stages, it may reappear indirectly at later stages, masked by youth who are previously detained.

Microlevel Theories

In contrast to macrolevel perspectives, microlevel theories focus on the individual level of social psychological characteristics of juvenile justice decision-makers, including the roles that stereotypes and discretion play in this process. The following microlevel perspectives will be discussed in relation to minority overrepresentation: symbolic threat thesis, racial stereotyping within the realm of attribution theory, the liberation hypothesis, and the focal concerns perspective.

Symbolic Threat Thesis

As briefly mentioned earlier in the chapter, Tittle and Curran's (1988) symbolic threat thesis focuses on the interplay between the characteristics of minority youth, and the emotions of juvenile justice decision-makers. The perspective argues that juvenile justice decision-makers may view minority youth as aggressive, sexual, lacking discipline, and other fear or resentment – provoking qualities. These emotions make juvenile court officers feel uneasy, uncomfortable, and unable to identify with the youth, leading to greater social control. These stereotypes of minority youth lead juvenile court officers to base decisions on perceived thoughts about juvenile offenders, not the specific behavior of the youth. Symbolic threats are connected to structural characteristics, minority threat, inequality, and public perceptions of gang involvement (Fagan 2010). In turn, while this threat is more symbolic (threat to middle-class standards and public safety) than real, it still may have an impact on how minority youth are treated within juvenile justice proceedings.

Differential treatment within the juvenile justice system could be based on decision-makers' perceptions and emotional reactions to what minority youth represent. Tittle and Curran (1988) found support for this perspective, indicating that a larger minority and younger population resulted in different case outcomes of minority you who were charged with drug and/or sexual offenses. Research has also found that minority youth who are involved with drug offenses lead to differential treatment compared to other youth (Leiber and Johnson 2008). For example, Bishop, Leiber, and Johnson (2010) found that minorities charged with drug crimes were more likely to continue into further juvenile justice proceedings than similarly situated White drug offenders.

Attribution Theory and Racial Stereotyping

The reliance of racial stereotypes by juvenile justice decision-makers has been utilized as an explanation on how subjective thought processes influence case outcomes. For example, Bridges and Steen (1998) integrated theory and prior research concerning racial stereotyping and attribution theory to explain differential treatment of African Americans compared to Whites. According to attribution theory, attributions provide individuals with explanations for previous experiences and criteria for evaluating one's own behavior, as well as others behavior, actions, and attitudes. Prior research has indicated that individuals who are more punitive or retributive in treatment philosophies have been more likely to blame the person. Bridges and Steen (1998) incorporated attribution theory, prior research on racial discrimination, and arguments that decision-makers develop and rely on past experiences and prior cases to influence case outcomes, and found support for their argument.

More specific, results indicated that probation officers used different types of attributions to assess delinquent behavior between Black and White youth. Involvement in delinquent acts committed by African American youths' were seen by decision-makers as related to internal or dispositional attributions (e.g., defects of character, lack of responsibility), while White youths' involvement in delinquency was related to external causes (e.g., disadvantaged neighborhoods, delinquent peers, poor school performance). Internal attributions compared to external attributions resulted in perceptions by decision-makers that youth were more at risk to re-offend, less amenable to treatment, and more dangerous, which resulted in longer sentences for Black compared to White youth. Overall, Bridges and Steen (1998) concluded that certain values and beliefs of juvenile probation officers lead to a stereotyped image of juvenile offenders that resulted in differential treatment throughout the decision-making process.

Research has also found that differential treatment with respect to ethnicity has also been found within the decision-making proceedings. Gaarder et al. (2004) found that attributions related to delinquency and victimization assigned to Hispanic females by juvenile justice decision-makers were linked to racial and gender stereotypes. In another study, the juvenile justice system responded to Hispanic females more severe than their White counterparts, finding that Hispanics from economically advantaged and disadvantaged communities are treated harsher than Whites (Rodriguez 2007).

The Liberation Hypothesis

Certain macro- and microlevel theories that seek to explain racial and ethnic bias and disparities have an underlying theme that incorporates what types of outcomes occur depending on the amount of discretion by decision-makers. The basic premise of the liberation hypothesis is that it identifies certain conditions that result in decision-makers ability to "liberate" to and from certain legal criteria and uniformity. When involving adult offenders, jurors were found to be more likely to exercise discretion when evidence against the defendant was weak and a less serious crime occurred. In other studies, judges were less likely to exercise discretion involving cases of murder, robbery, or rape, but in less serious cases, both legal and extralegal variables influenced judges' decision-making. Blacks who were convicted of less serious crimes were treated harsher at sentencing, while Blacks who were convicted of more serious crimes were no more likely than Whites to receive a harsher sentence.

The liberation hypothesis has also been examined within the juvenile justice system, suggesting that for more serious cases, court outcomes will be influenced by legal factors, while for less serious cases, court outcomes will be influenced by extralegal factors, such as race. Research suggests some support for the liberation hypothesis at the juvenile level; however this depends on which stages are examined (Guevara et al. 2011). In one study, minorities were more likely to be adjudicated compared to Whites (Freiburger and Burke 2010); yet another study did not find support for the hypothesis at the stage of detention. Overall, while some support has been found for the liberation hypothesis, the issue of *why* decision-makers liberate from strict legal criteria when evaluating less serious crimes is missing from the theory, especially the notion *why* increased discretion by juvenile court officers results more severe treatment for minority offenders compared to Whites.

A "Focal Concerns" Perspective

Originally, Steffensmeier and associates developed and tested a "focal concerns" perspective to explain discretion and increased social control within adult court processing decisions. This perspective has also been applied within the juvenile justice system. The underlying premise of this perspective is that judges have a limited amount of time and information about defendants and may rely on three focal concerns or attributions involving race, gender, and class stereotypes when making decisions. The three focal concerns include (1) the defendants' blameworthiness and culpability based on the seriousness of the offense, (2) society's concern to protect the community, and (3) organizational considerations involving available correctional resources (Steffensmeier et al. 1998). Even though the three focal concerns are interrelated, Steffensmeier and colleagues argue that judges do not have complete information about each. From this, judges exercise their own discretion and rely on attributions such as age, race, and social class while making decisions. In other words, judges develop a "perceptual shorthand," where they rely on legal factors (crime severity, prior record) and racial stereotypes to determine case outcomes.

In order for this perspective to be applicable to the juvenile justice system, the focal concerns approach requires some modification. Compared to the adult system, the juvenile justice system has a dual focus of social control and social welfare of all youth, directing attention towards treatment and rehabilitation of juvenile offenders. For example, the social welfare orientation of the system influences decision-makers to focus on extralegal factors such as school, family, and peers, when assessing possible treatment outcomes. However, just as adult court officers may rely on stereotypes when making outcome decisions, juvenile court officers may also rely on stereotypes concerning juveniles' family, peers, and school situations. Utilizing a previous example, juvenile court officers who believe that single-parent family are unable to adequately provide for youth has been shown to influence decision-making (Bishop and Frazier 1996).

More recent, Bishop and colleagues (2010) integrated the focal concerns perspective with the "loose coupling" perspective. They argued that race effects may vary by each stage in juvenile justice proceedings due to numerous decision-makers involved and associated responsibilities and concerns. On the one hand, Bishop et al. (2010) contend that the influence of race will be more likely to exist at the loosely coupled stages of intake and judicial disposition; stages where multiple actors have a say, multiple goals of decision-making exist, are often tied to racial stereotypes, and discretion is prominent. On the other hand, at the tightly coupled stages of petition and adjudication where there are fewer decision-makers involved, legal factors and in particular the severity of the crime drive decision-making. In other words, discretion is constrained, and race effects are predicted to have a diminished role in decision-making. Consistent with the focal concerns and loose coupling perspective, results from Bishop et al. (2010) confirmed these expectations.

In addition, support for the focal concerns perspective has been found throughout numerous studies (Steffensmeier et al. 1998). Steffensmeier and colleagues (1998) found that young, Black males received more severe sentenced compared to any other age, race, or gender subgroup. This result confirmed the stereotype that young, Black, males are perceived as more dangerous, threatening, and unsuitable for release.

However, some researchers have questioned that the focal concerns concept is not a true theory but rather a perspective. On the one hand, there are no testable propositions, and the focal concerns concepts are underdeveloped and in need of refinement. On the other hand, empirical tests of the perspective have lead to supportive conclusions, and overall may only need some additional modifications.

Summary

In sum, traditional consensus, conflict, and labeling theories paved the way for the incorporation of more recent contextual and structural perspectives to better understand race and juvenile justice decision-making. At the macrolevel, the social control of minority youth has been fostered by structural and contextual characteristics, and stereotypes enforced by juvenile justice decision-makers. At the microlevel, juvenile court officers may be dependent on racial stereotypes, biases, symbolic threats, and internal attributions that have resulted in differential treatment and harsher sanctions of minority youth. In the final section, a specific policy implication called the disproportionate minority confinement/contact mandate (DMC) will be described in detail, in order to explain a present-day response to racial disparities in the juvenile justice system.

Policy Implications

It has already been established that minority youth are overrepresented throughout all stages of the juvenile justice system. However, it was not until the 1980s that this occurrence was officially acknowledged as a "problem." A policy implication and potential solution to this "problem" was the implementation of the DMC mandate. The following paragraphs will describe the history of DMC, specific stages within the mandate, results of present assessment studies, the overall effectiveness of DMC, and future research and efforts.

The Disproportionate Minority Confinement/ Contact (DMC) Mandate

History

In 1989, the disproportionate minority confinement mandate (DMC) was passed as part of the reauthorization of the Juvenile Justice and Delinquency Prevention Act (JJCPA) of 1974. In 2002, the JJDP Act was modified, shifting the emphasis from "disproportionate minority confinement" to "disproportionate minority contact" requiring the examination of possible minority youth overrepresentation throughout all decision points in the juvenile justice system. This examination is continuously conducted throughout five interconnected phases of the DMC mandate. Throughout DMC's history and continuing today, the underlying goal of the mandate is the equitable treatment of all youth within the juvenile justice system (Department of Justice 2009).

Stages

As stated earlier, the DMC mandate has five interrelated and ongoing phases: identification of DMC, assessment into possible causes, intervention, evaluation, and monitoring. The identification stage includes identifying and monitoring if and at which stages of the juvenile justice system DMC exists. More specific, the identification stage provides a count of the occurrence of DMC, as well as the extent of minority overrepresentation within different stages and jurisdictions. This phase sets the stage for the assessment studies of DMC by presenting descriptive data about the occurrence and location of DMC. At this stage, research has found that minority youth are overrepresented most often during arrest, detention, and waiver decisions. Compared to Whites, minority youth are also more likely to be petitioned and receive a disposition of placement outside of the home. From this, research aimed at identifying DMC has found that minority overrepresentation occurs at the beginning stages of juvenile justice system, which results in a cumulative effect of increased disparities throughout later case proceedings as well.

The second interrelated phase is the assessment phase which includes the examination of possible causes or factors that contribute to minority youth overrepresentation. The assessment stage examines in greater detail the specific decision points that DMC has been shown to exist from the identification phase. In addition, the assessment stage identifies certain mechanisms that are occurring in different jurisdictions that contribute to DMC. Research at this stage has found numerous contributing characteristics of DMC throughout many states and jurisdictions, which is described later in this section.

The intervention stage is based on taking information that is gather from the previous two phases and develop a comprehensive list of intervention strategies that have that ability to reduce minority youth overrepresentation. Several initiatives have been created and implemented across numerous jurisdictions and stages that center on direct services of youth, training of key personnel within the juvenile justice system, and to some extent, encourage change in the system itself. Jurisdictions are encouraged to rely on the results from the identification and assessment stages to guide decision-making concerning what specific initiatives and interventions should be implemented. For example, direct services attempt to address and provide at-risk youth with skill development, educational attainment, and positive relationships with family and peers. Prevention programs may include strategies that focus on familial relationships and educational deficiencies, while intervention programs seek to reduce antisocial behavior. Other types of direct services include diversion programs/services, advocacy, and alternatives to secure detention. Cultural diversity training is another strategy used to expose and educate individuals about DMC and racial bias in the juvenile justice system. Some examples of this type of strategy include training of law enforcement and juvenile court officers, participation in national and regional conferences, and day-long workshops.

Unfortunately, significant system change within juvenile court proceedings has been a slow occurring process. However, changes that have been implemented include legislative reform, administrative changes, and structural and procedural changes that impact decision-making (US Department of Justice 2009). One of the most popular changes involving the DMC mandate has focused on changing detention procedures with an emphasis on alternatives to secure detention, which previous research has found that minority youth are more likely to be detained compared to White youth (Rodriguez 2010).

The final stages of the DMC mandate include the evaluation and monitoring of the DMC mandate. It is important to evaluate the effectiveness of interventions that have been implemented on a continued basis. Overall, there are some interventions that have been evaluated, and results indicate that there has been some success at reducing minority youth overrepresentation (US Department of Justice 2009). However, relatively few known evaluations of strategies have been conducted, including a lack of systematic evaluation of certain interventions across multiple jurisdictions. Overall, it is the evaluation and monitoring stages that have seen a lack of implementation, which will later be discussed involving a discussion of the failure or success of the DMC mandate.

Results of Assessment Studies

Recall that the second phase of the DMC mandate is the assessment into possible causes or factors that contribute to minority overrepresentation. Once again, the assessment process examines at which decisions points DMC occurs from the identification phase. From this, information will be obtained to help policy makers choose appropriate interventions and strategies to reduce DMC. Presently, results from assessment studies found that explanations for the occurrence of DMC center on the juvenile justice system itself and differential involvement in offending. It is also important to note that the results from assessment studies parallel and mirror empirical research in general of how the juvenile justice system responds to minority youth.

Within the juvenile justice system, four specific factors contribute to DMC. First, racial stereotyping and cultural insensitivity on the part of police, juvenile court officers, prosecutors, and judges have been found to further move minority youth into the system. Second, there has also been evidence from assessment studies concerning unintended consequences that impact minority youth through the implementation of certain laws and policies. Examples of these include zero-tolerance and transfer/waiver adult certification of youths. Third, research has found that due to a lack of alternatives to diversionary programs, minority youth receive more severe outcomes and further movement into the juvenile justice system, ultimately resulting in placement in secure corrections. In other words, minority youth are not provided the same opportunities to participate in diversion programs in general, and specifically as an alternative to secure detention. Last, assessment studies have found that a lack of bicultural and bilingual staff and the use of English – only information materials lead to an inability to effectively communicate with non-English-speaking youth and families, resulting in a hindered ability to successfully navigate the juvenile justice system.

Assessment studies have also found that some of the strongest explanatory factors involving DMC are legal and extralegal criteria. Differential involvement in crime has been said to lead to a cumulative disadvantage that continues to hinder the life chances of minorities, further increasing the chance of continued criminal involvement in both juvenile and adult systems. Once again, four factors were found to be related to the continued occurrence in minority overrepresentation as offenders. First, educational deficits in the form of poor school attendance, lack of an emphasis on the value of education, disciplinary problems, and learning and cognitive abilities have been related to minorities committing more delinquency. Second, certain types of family dysfunctions have also shown to predict more harsh case outcomes, especially minorities who live in single-parent homes, reside in high-poverty neighborhoods, experience parental abuse, and have insufficient supervision. Third, assessment studies have found that many youth referred to juvenile court suffer from mental and emotional disabilities, often lacking the necessary services to address these issues due to family and financial problems. Finally, assessment studies have also considered certain structural characteristics of juvenile courts that may lead to minority overrepresentation. Results have found that juvenile courts in urban areas that are characterized by racial inequality, underclass poverty, and high crime rates are more likely to formally process, detain, and impose harsher sentences on minority youth.

Overall Effectiveness of the DMC Mandate

Over the last 20 years since the implementation of the DMC mandate, key improvements and advancements have emerged. For example, research has expanded its focus to include numerous minority groups (Hispanics, Native Americans, etc.) in addition to African Americans. In addition, factors that decision-makers take into consideration when deciding case outcomes for youth (school problems, assessments of the family, and mental health issues) have been more closely examined than in previous years. Assessment studies have also begun to examine how certain factors such as age, detention, and technical violations by the youth may work to the disadvantage of some minority groups relative to Whites. Lastly, recent assessment studies have adopted a more macro-sociological perspective, inquiring whether racial and ethnic disparities may be explained or linked to the larger social or structural context of where individual courts are located. Stated differently, research on DMC has expanded its inquiry to attempt to understand if differential offending and racial disparities in juvenile justice processing may be mutually affected by underlying social, structural, and cultural context in which they occur.

While key improvements and advancements have emerged in the last 20 years, the overall effectiveness of the DMC mandate has lead to three different opinions and positions concerning its success (Leiber and Rodriguez 2011). The first position argues that the issue of DMC is misguided and instead should focus on the prevention of delinquency of minority youth. In other words, this position believes that resources and strategies should divert its efforts towards differential offending by minority youth, instead of attempting its previous efforts to understand racial and ethnic bias.

The second position believes that overall, the DMC mandate has failed to bring about change to reduce racial and ethnic bias, and therefore more aggressive measures need to be taken into consideration. Some researchers have specifically argued that a lack of benchmarks and the failure to change procedures in the juvenile justice system reveals reluctance from jurisdictions to change procedures and policies that deal with overall system issues. In some cases, evidence was found that the DMC legislation was unsuccessful at changing decision-makers view of relying less on race and more on legal and extralegal considerations. From this, OJJDP has been criticized about not doing enough to address the issue of DMC. Overall, procedural changes within the juvenile system have been too slow to occur.

A third position focuses on broadening the inquiry of DMC to include both the causal factors associated with juvenile involvement in offending and increased social control, as well as the understanding the perceptions of racial stereotypes that further disadvantages minority youth in the juvenile justice system. In other words, supporters of this position believe that efforts need to focus on both the conditions that contribute to involvement in delinquency, including factors that foster the reproduction of stereotypes and biases of minority offenders through the eyes of juvenile justice decision-makers.

Future Research and Efforts

There is still a need for states and jurisdictions to continue to conduct assessment studies. In other words, future research needs to continue and extend analyses to gain a better understanding of the complex intricacies involved in the relationship between race, crime, and juvenile justice system involvement. For example, research needs to begin to assess the relationship between race and gender (e.g., being a Black male compared to a White female in the juvenile justice system). Some research has suggested a community commitment to reduce DMC, through addressing issues including a truancy program as well as collaborations with school systems, law enforcement officials, juvenile courts, and political offices.

Future efforts should also continue to sensitize individuals working within the juvenile justice system that the laws in general, legal and extra legal criteria, and system procedures may disadvantage minority youth compared to their White counterparts. More generally speaking, structural, neighborhood, and community changes need to be made to decrease the amount of minority overrepresentation but acknowledge that this specific recommendation will be a difficult and slow process to conquer. Lastly, additional research is needed to examine *why* minority youth continue to be overrepresented within the juvenile justice system. Even though it has been 20 years since the implementation of DMC, there are still numerous conflicting views of its success, as well as a need to continue future research and efforts.

Conclusion

In sum, while the overrepresentation of minority youth within the juvenile justice system has been a concern for many years, recent policy implications like the DMC mandate have attempted to identify, assess, intervene, evaluate, and monitor the occurrence and location of racial disparities within juvenile court processing. Through the use of traditional, macrolevel, and microlevel theories, researchers have attempted to understand how and why minority youth are treated more severe. Even though there have been gaps in the implementation and monitoring of certain DMC initiatives, progress has occurred. This proves that the DMC mandate is an ongoing and interrelated progress that in the end will hopefully lead to the equitable treatment of all youth, regardless of race and ethnicity.

Recommended Reading and References

Bishop D, Frazier C (1996) Race effects in juvenile justice decision-making: findings of a statewide analysis. J Crim L Criminol 86:392–414

Bishop D, Leiber M, Johnson J (2010) Contexts of decision making in the juvenile justice system: an organizational approach to understanding minority overrepresentation. J Youth Violence Juvenile Justice 8(3):213–233

Blalock HM (1967) Toward a theory of minority group relations. Wiley, New York

Bridges G, Steen S (1998) Racial disparities in official assessments of juvenile offenders: attributional stereotypes as mediating mechanisms. Am Sociol Rev 63:554–570

Fagan J (2010) The contradictions of juvenile crime & punishment. Dædalus 139:43–61

Frazier CE, Bishop DM, Henretta JC (1992) The social context of racial differentials in juvenile justice dispositions. Sociol Q 33(3):447–458

Freiburger TL, Burke AS (2010) Adjudication decisions for black, white, Hispanic, and Native American youth in juvenile court. J Ethn Crim Justice 8(4):231–247

Gaarder E, Rodriguez N, Zatz M (2004) Criers, liars, manipulators: probation officers' views of girls. Justice Q 21(3):547–578

Graham S, Lowery BS (2004) Priming unconscious racial stereotypes about adolescent offenders. Law Hum Behav 28(5):483–504

Guevara L, Boyd LM, Taylor AP, Brown RA (2011) Racial disparities in juvenile court outcomes: a test of the liberation hypothesis. J Ethn Crim Justice 9(3):200–217

Holman B, Zeidenberg J (2006) The dangers of detention. Justice Policy Institute, Washington, DC

Leiber M (2003) The contexts of juvenile justice decision making: when race matters. State University of New York Press, Albany

Leiber M, Johnson J (2008) Being young and Black: what are their effects on juvenile justice decision making? Crime Delinq 54:560–581

Leiber M, Rodriguez N (2011) The implementation of the disproportionate minority contact confinement/contact (DMC) mandate: a failure or success? Race Justice 1:103–124. doi: 10.1177/2153368710377614

Rodriguez N (2007) Juvenile court context and detention decisions: reconsidering the role of race, ethnicity, and community characteristics in juvenile court processes. Justice Q 24:629–656

Rodriguez N (2010) The cumulative effect of race and ethnicity in juvenile court outcomes and why preadjudication detention matters. J Res Crime Delinq 47(3):391–413

Sampson RJ, Laub J (1993) Crime in the making. Harvard University Press, Cambridge

Sampson RJ, Lauritsen JL (1997) Racial and ethnic disparities in crime and criminal justice in the United States. Crime Justice 21:311–374

Steen S, Bond C, Bridges G, Kubrin C (2005) Explaining assessments of future risk: race and attributions of juvenile offenders in presentence projects. In: Hawkins D, Kempf-Leonard K (eds) Our children, their children: confronting racial and ethnic differences in American juvenile justice. The John D. and Catherine T. MacArthur foundation, research network on adolescent development and juvenile justice. The University of Chicago Press, Chicago, pp 23–82

Steffensmeier D, Ulmer J, Kramer J (1998) The interaction of race, gender, and age in criminal sentencing: the punishment cost of being young, black, and male. Criminology 36:763–797

Tittle CR, Curran DA (1988) Contingencies for dispositional disparities in juvenile justice. Soc Forces 67(1):23–58

US Department of Justice (2009) Disproportionate minority contact technical assistance manual (4th ed.). Rockville, U.S. Dept of Justice, Office of Juvenile Justice and Delinquency Prevention

Weber M (1969) The nature of the city. In: Sennett R (ed) Classic essays on the culture of cities. Meredith Corporation, New York

Zatz MS (1987) The changing forms of racial/ethnic biases in sentencing. J Res Crime Delinq 24(1):69–92

Juvenile Sex Offenders

► Risk Factors for Adolescent Sexual Offending

Juvenile Victimization

Christopher J. Schreck
Department of Criminal Justice, Rochester
Institute of Technology, Rochester, NY, USA

Overview

Juvenile victimization can take many forms, and research has documented that victimization during childhood is connected with negative long-term patterns in criminal activity, health, and prospects for the future. Research on victimization patterns began in earnest during the 1970s, with the advent of the National Crime Victimization Survey (NCVS). Although the victimization of youth is a topic of considerable importance and interest to the scientific community, less is known about young victims of crime either because large-scale data collections, such as the NCVS, refrain from collecting data from this group or else information is only available from official sources. Nevertheless, a sizable scholarly literature has developed in order to better understand the causes, extent, and consequences of juvenile victimization, whether in the family or outside the family.

In recent years, research has begun to offer a much clearer picture about juvenile victimization, finding that many forms of victimization among youth are concentrated among African-Americans and the poor. As a result, researchers have begun to build theories to account for patterns of juvenile victimization as well as to evaluate programs designed to prevent youth from becoming victims.

Extent of Juvenile Victimization

The main source of our knowledge about victimization patterns comes from the National Crime Victimization Survey (NCVS), which collects data from a representative sample of Americans age 12 or older. Although the NCVS is the largest source of data on victimization, there are important limits with respect to understanding juvenile victimization. First, the focus of the NCVS is on street crimes (such as assaults, thefts), and so information on child maltreatment is therefore lacking. Second, although street crime victimization data is available for teenagers, younger children are excluded. Because of this omission, relevant information about those crimes and groups must come from alternative and more limited data sources. This entry will turn to these sources of information later.

With these limitations in mind, the National Crime Victimization Survey generally shows that violent victimization is most highly concentrated among those aged 24 or younger. This pattern has held true since the advent of the NCVS in 1973. From 1973 well into the 1980s, the violent victimization of youth was at a fairly stable and relatively high plateau, with rates of victimization ranging from 75 to as much as 100 incidents per 1,000 people each year. Starting in the late 1980s, however, violent victimization began to trend significantly upward. In 1994, the annual rate peaked at 125 per 1,000 individuals and then began a precipitous decline that has continued through the most recent year for which there is data. In 2009, the rate of violent victimization for youths aged 12–19 had fallen to less than 37 per 1,000. Juvenile violent victimization is thus presently at its lowest recorded levels. Of the violent crimes, simple assaults were the most likely to occur among juveniles. Property crime victimization, however, was substantially more frequent among juveniles than violence. Juveniles (age 12–19) who were classified as "heads of households," reported a property victimization rate of 268 per 1,000 in 2008, which was the highest rate of any age-group by a considerable margin. Most of these victimizations were thefts. In short, while juveniles have the highest risk of victimization for any age-group, most youths will not become victims and of those that do the most likely victimization will take the form of a theft.

The NCVS does not collect data on murder victims, but juveniles are sometimes the victims of homicides. The Supplemental Homicide Report (SHR) is the key source of data, and the SHR records the age of those who die in a criminal attack no matter how young. The data show that juveniles have rates of homicide victimization that are relatively low compared to other age-groups. Since 1975, children who were 13 or younger have consistently had the lowest murder rate of any age-group (an annual rate of approximately 1–2 per 100,000 between 1975 and 2005), even when contrasted with rates for the elderly. Youths aged between 14 and 17 had somewhat higher rates (less than five incidents per 100,000 people in 2005), but these rates were generally comparable to middle-aged adults. These older juveniles, however, have shown more fluctuation in their homicide victimization pattern over the years. Starting in the late 1980s and through the mid-1990s, their homicide victimization rates jumped to 12 per 100,000. This bulge was only mirrored in the 18–24 age-group (the group with consistently the highest rate of murder victimization); all the remaining age-groups generally reported stable levels of homicide victimization during this period.

Although the NCVS shows that violent victimizations are rare, even among the young, youth may also be exposed to child maltreatment. Child maltreatment, as generally understood, encompasses the sexual and physical abuse of minors as well as neglect. Beyond this, however, there is no consensus about what these terms exactly mean, and states have accordingly established differing criteria as to what constitutes maltreatment. As a result, precise estimates of how many children are maltreated, and developing a profile of who they are, are somewhat problematic tasks. The two major sources of child abuse and neglect data are the National Child Abuse and Neglect Data System (NCANDS) and the National Incidence Study (NIS). Unlike the NCVS, NCANDS does not collect relevant data directly from the youth but rather from state Child Protective Services (CPS) records. Because of this, unreported maltreatment or those that in fact occur and that CPS cannot substantiate are not reflected in the data. According to 2009 data, approximately 64 % of allegations made to CPS were found to be unsubstantiated. The NIS goes beyond CPS and surveys community professionals who have contact with children, but this is not an annual survey.

The NCANDS data, since 2005, shows that child victimization rates are low and fairly stable over time. In 2005, the victimization rate for children age 17 or younger was 12 confirmed incidents of maltreatment per 1,000 juveniles, or only 1.2 %. The prevalence rate that year, which is the number of children victimized at least once, was 10.9 per 1,000. Each year since 2005, the child victimization rate has declined slightly. By 2009, the victimization rate was ten per 1,000, with the prevalence rate being less than 1 %.

The children most likely to be reported as victims of child maltreatment tended to be less than 3 years old and most of these were less than a year old. Juveniles least likely to be substantiated as victims of maltreatment typically are older (for instance, 16–17 years old). These substantiated cases were almost equally divided between boys (48.2 %) and girls (51.1 %; the remainder were "unknown"). White children comprised 44 % of victims of child maltreatment, followed by African-Americans (22 %) and Hispanic (21 %). African-Americans, however, had the highest rate of victimization. Neglect figured in more than three-quarters of all cases reported in the NCANDS data, with physical abuse (17.8 %) being the next most common form of child victimization. These patterns have remained generally stable since 2005.

Although substantiated allegations child abuse and neglect, according to these estimates, is statistically rare, a large amount of research has examined the predictors of maltreatment. Research has located consistent risk factors for maltreatment since research on the topic began in earnest in the 1960s. These risk factors appear to be correlated with one another; that is to say, the presence of one risk factor is likely to be accompanied by one or more additional predictors. Research has shown that families where the parents are isolated from stable social support networks tend to have higher incidences of abuse victimization. Likewise, abuse is more prevalent among those families living in poverty. Theorists have speculated that impoverished families are less able to properly allocate financial resources, which create stress within the family. Abuse appears to be more concentrated among racial/ethnic minorities, particularly African-American households. This pattern likely arises from the fact that racial minorities are often ranked lower on the economic ladder, and they tend to live in severely disadvantaged areas where social support structures for families are very weak. Parents who themselves have problems with alcohol and drug addiction, or mental health issues, are more likely to mistreat their children.

Consequences of Victimization for Juveniles

Apart from estimating the extent of victimization among youth, scholars have also invested some attention on the effects of victimization. The life course approach develops the idea that quality of life over the years, including antisocial behavior, is largely (although not completely) shaped during childhood. Children acquire, during these years, the social capital needed to build their lives: education, experience with relationships, work, and so forth. The basic idea is that childhood victimization represents a life course event that can have long-term adverse effects on child development and adult outcomes. To the extent that victimization impairs the acquisition of social and personal capital, then it represents an important domain for research.

There is a considerable body of literature exploring the long-term effects of familial maltreatment on children, and the findings consistently show that children whose parents victimize them are more likely to experience adverse effects going forward. Abused children tend to:

- Experience delays in physical development
- Show higher levels of antisocial behavior
- Have lower self-esteem
- Have lower levels of academic achievement
- Experience symptoms of depression, posttraumatic stress disorder (PTSD), and anxiety

Beginning in the 1960s, scholars discussing the long-term effects of physical victimization of children spoke of the "cycle of violence." The cycle of violence can refer to two phenomena. First, abused children may, when they become parents, turn toward abusing their own children. Second, abused children may have offense repertoires that are characterized by higher levels of violence. Considerable research shows that abused children have a greater probability of engaging in violent and sexual offending in adulthood. Although the cycle of violence as an idea is well ensconced in the child abuse field and among policy-makers and advocates, the research suggests some caution. Such experts as Cathy Spatz Widom found that most physically maltreated children did not get arrested, whether as juveniles or adults. Children who were maltreated did tend to have a greater probability of arrest than other children, but arrests were for a range of crimes – sometimes violent but often not. Or, put another way, physical abuse increased the risk of arrest for not only violence but also nonviolent crime.

The literature is less large with respect to victimization's effects on children more generally; however, there do appear to be similar patterns to those reported in the child abuse literature. For instance, child victims of nonfamilial violence and sexual assault are more likely to experience PTSD and feelings of distress and depression. These feelings, as with abuse victimization, appear to persist into adulthood. Other long-term effects include stunted educational and socioeconomic consequences. Adolescents who were victimized were less apt to have a record of quality educational achievement. This, in turn, undermined work-related accomplishments and occupational status. Victimized children also have a greater probability of offending as well. In short, research indicates that many of the problems experienced by abused children are evident among those children who are victimized by those outside the family. In both cases, maltreatment and victimization among juveniles were more likely to result in a degraded quality of life as adults.

Victimization in School

Youth spend a considerable amount of their waking hours at school, and it therefore follows that school may be a context in which victimization frequently takes place. Victimization at school can take many forms – street crimes but also harassment and bullying. For different reasons, the topic of school crime has drawn great interest from parents, policy-makers, and students. Students who are victims of crime often have a more difficult time meeting the academic and social expectations of school life, such students can drop out of school, which can have adverse long-term effects on higher education and employment prospects. In turn, struggling students could endanger the existence of the school itself, as such programs as No Child Left Behind can mandate the closure of schools that repeatedly fail to make progress toward meeting educational benchmarks for reading and math.

Since 1992, the National Crime Victimization Survey has reported victimization incidents that occurred on school grounds. That year, the rate of theft victimization per 1,000 students was 95; the corresponding rate for violent victimizations was 48 ("serious violence" had a rate of only ten per 1,000 students). The level of victimization for all types of crime has decreased dramatically since the early to mid-1990s, particularly for theft victimizations. As of 2008, both theft and violent victimization rates were 24 per 1,000 individuals (serious violence at only four per 1,000). A comparison of the overall rates of victimization across the years suggests that school is an unsafe place for juveniles. In 1992, for example, the combined rate of victimization was 138 per 1,000 (compared to 144 per 1,000 at school). By 2008, schools still had a higher rate of victimization for juveniles than elsewhere (47 per 1,000 at school, versus 38 per 1,000 away from school). These numbers are somewhat misleading, at least by the metric of safety. Most of these victimizations are thefts. Rates of victimization in 2008 for "at school" and "away from school" are also tiny, especially when compared to what they were 15 years previous. Most importantly, however, serious violent victimization is more likely to occur away from school (rate of eight per 1,000) than at school (four per 1,000). This pattern is reflected in other data sources. For instance, National Incident-Based Reporting System (NIBRS) data shows that during the school week violent offending committed by juveniles (and, presumably, violent victimization) peaks at approximately 3 pm, which is around the time when students are dismissed from school and are unsupervised. Although a considerable amount of victimization takes place at school, its levels are small and the most serious forms of violent victimization are most likely to transpire, not at school, but rather on nearby playgrounds and en route home from school.

Information on bullying is considerably scantier. Some research indicates that bullying affects a relatively small portion of students (less than 15% are either perpetrators or victims of bullying). Estimating the extent of bullying is somewhat difficult. Aggressive behavior calculated to intimidate or harass someone, such as by threatening them or calling them names, is commonly included as part of the definition of bullying. Some scholars, however, include passive-aggressive behaviors that the bullying victim might not even be aware of – where bullies ostracize disliked peers or spread rumors and gossip. Many of the long-term consequences of bullying victimization resemble those for victimization and child maltreatment, such as degraded educational achievements and depression.

Of particular concern, however, are school shootings. For instance, in a 2001 Gallup survey administered at a time when the Columbine and Jonesboro school massacres were a fresh memory in the minds of the public, 70 % of parents felt that a shooting was “likely” at their school. While school shootings garner much media attention, during 2008–2009 academic year there were only 15 homicides that took place on a United States elementary or secondary school campus. This is contrasted with the approximately 1,700 homicides of youths age 5–18 over that same period. Homicide is itself a very rare event; however, when the murder of children does take place, they are extremely unlikely to occur at school. The rate of homicide at school amounts to approximately one youth each year per 2.5 million students. Since 1992, the number of school homicide victimizations has held fairly steady between 14 and 34 per year, with recent years tending toward the lower boundary of that range.

What sort of student is most likely to fall victim at school? The findings are clear that victimization at school, although unlikely to befall any given student during a single year, is not a random occurrence; if it was, then everyone would share equal risk. Researchers have examined the correlates for school victimization, and findings have begun to emerge implicating a combination of factors, such as particular aspects of the community where the school resides, the school context, and the students themselves. Given that much victimization occurs off school grounds after children are dismissed, the quality of the neighborhood plays a significant role in promoting the risk of victimization. The most important school-related risk factor for victimization appears to be the concentration of offenders at the school, such as children who bring weapons or else gang activity. The most consistently salient contributor toward youthful victimization, however, seems to be factors connected with the individual student. Students who bring weapons to school, who resent school and feel isolated, and who have friends who are delinquent are more likely to experience victimization.

Research on the effectiveness of school security strategies – guards, metal detectors, corporal punishment – have generally produced unimpressive results, especially after controlling for the other known correlates of youthful victimization. This is likely because many school preventative strategies are irrelevant to when and where victimization actually occurs. The presence of a school resource officer and locked school doors has little to do with theft, which is the most common type of victimization at school. Rather, these strategies are most clearly targeted toward preventing violence; however, research has yet to demonstrate a significant beneficial effect even for reducing the occasional and usually low-level scuffling that occurs during the school day. Since serious forms of victimization are most likely to happen after final dismissal from school, then the usefulness of guards and metal detectors is neutralized. Research has even shown that students where there are invasive forms of security do not feel any safer and are somewhat more likely to be afraid. Nevertheless, numerous other factors are much more strongly connected with higher levels of fear – for instance, experiencing victimization and observing criminal activity while at school.

Explanations of Nonfamilial Juvenile Victimization

Given these patterns of juvenile victimization, it is clear that victims possess distinct characteristics that somehow increase their risk. Victimization researchers have, over the past several decades, developed several important theories in an effort to explain why some people fall victim while others do not. The most long-standing theoretical tradition in victimization work is the routine activities/lifestyles approach. Lawrence Cohen and Marcus Felson, Michael Hindelang, Michael Gottfredson, and James Garofalo were the creators of this approach. Over the following years, other theorists would offer additional perspectives on victimization that included social-psychological or cultural risk factors. Recently, building on studies that show that offenders and victims share demographic characteristics and profiles, researchers have even applied criminological theories to account for victimization risk. Note that these theories are organized to explain nonfamilial forms of victimization; the risk factors for child maltreatment were presented earlier.

Lifestyles-Routine Activities Theory. The lifestyles/routine activities (lifestyles/RAT) approach is a long-standing theoretical tradition in victimization research and was the first victimization explanation to be widely accepted and researched. The approach argues that for victimization to occur, a given situation or context must have a combination of (1) a likely offender, (2) a worthwhile target, and (3) ineffective guardianship. Circumstances where these three factors are present will have the highest incidence of victimization. Other, relatively more recent, scholarship has identified the qualities of routine activities or lifestyles that ought to be predictive of victimization: that those who engage in unsupervised and unstructured social activity with peers will have a higher risk. Such activities would include hanging out with friends and going to parties, activities that are often characteristic of youth.

In the late 1970s through the mid-1980s, research was primarily concerned with the demographic correlates of victimization such as those mention earlier. Starting in the 1980s and continuing into the 1990s, research turned toward specific situations and activities that enhanced the probability of victimization. For instance, researchers noted that self-reported offending was highly correlated with victimization among not only adults but youth as well. In other words, offenders tend to also be victims. The lifestyles/RAT approach accounted for this by claiming that those who engaged in a deviant lifestyle were more likely to spend time with those who commit crime, the very people who would not scruple at victimizing their own peers (thus implying proximity to motivated offenders). Not all juvenile victims are also deviants, but studies show that there is considerable overlap between offending and victimization. Note that the child maltreatment and child victimization literatures point to offending as a byproduct of victimization, thus offering their own somewhat differing accounts of the victim-offender overlaps.

Self-Control Theory. Michael Gottfredson and Travis Hirschi's self-control theory suggests that a person's ability to focus on long-term consequences of behavior, as opposed to short-term solutions to problems, explains conformity. People with low self-control, in contrast, make self-centered decisions that address short-term problems that fail to consider the long-term consequences of those decisions. One can understand how much self-control a juvenile has by identifying actions that consistently speak to ignoring long-term risk, such as acting angrily, impulsively, in flagrant disregard for risk, and selfishly. Such actions also elevate the risk of victimization as well. Children who are aggressive toward others may be attacked and beaten. Impulsive behavior can result in greater vulnerability. Selfish behavior can ruin friendships that have the potential to become strong, and thereby weaken guardianship potential. In this respect, victimization is not something that those low in self-control voluntarily seek out; however, their decisions place them at greater risk than would otherwise be the case. Research so far shows that those with low self-control are more likely to become victims of all manner of crime.

The theory asserts that through childhood differences in self-control tend to be unstable; younger children, in particular, are going to have great difficulty in being able to act consistently with self-restraint. In this respect, low self-control may not offer much of an account for early childhood maltreatment because low self-control is assumed to be universal in that population. Indeed, Gottfredson and Hirschi attributed childhood maltreatment to parents who were poorly equipped to teach their children self-control. Once children can be reliably rank-ordered in terms of their self-control, however, then the theory can offer a plausible account of why some individuals become victims.

Social Interationist Theory. Richard Felson introduced the "social interactionist" (or SI) theory to explain violent victimization. The main argument of the theory is that victimization occurs as a result of interplay between two or more individuals involved in social exchange. The basic process begins with a person who has experienced a negative event and who is consequently feeling emotionally down. The emotional response of the stressed individuals is usually to become short-tempered, irritated, or angry. These behaviors elicit backlash from those around the individual who, rather than take pity on the stressed person, become annoyed or angry themselves. As a consequence, the other person in the interaction will have a grievance, which may provoke physical attack if the distressed person is unable or unwilling to make satisfaction (e.g., apologize). This social interaction produces a cycle of retaliation and counter-retaliation that is hard to stop. A few major studies have tested the tenets of the theory to explain victimization. Richard Felson examined the role of negative events as predictors of victimization, finding that they indeed made victimization more likely to occur. Later work has connected negative affect or emotions with significantly elevated victimization risk among youth.

This particular framework is of particular relevance to those interested in juvenile victimization insofar as early maltreatment and victimization appear to produce the emotional distress that is implicated in this theory. That is, childhood victimization (familial or nonfamilial) results in emotional disorders and worsened performance at a variety of tasks. These circumstances, according to SI, would create the grievances that might provoke attack.

Subcultures of Violence Theories. For many years, researchers have studied how subcultural norms of violence are connected with victimization. These subcultures of violence place value on acting tough and retaliating when disrespected. Inherently, this creates a cycle where victims retaliate and become offenders against the initial offenders (who now become the victims). Therefore, the subculture is characterized by individuals who alternate roles as offenders and victims. In fact, this literature has consistently found that violent offenders living in areas where norms of violence predominate experienced high rates of victimization.

More recently, scientific research has begun to directly test the influence of subcultural norms. Elijah Anderson, for instance, introduced an influential theory of offending in his book *Code of the Street*, an ethnographic study of African-Americans in disadvantaged neighborhoods in Philadelphia. The street code is a system of rules governing the use of violence to informally address grievances. These codes are adopted by many individuals in disadvantaged communities, and even if they are not ascribed to by all, they are known by almost everyone living in that area. Here, the individual has a system of beliefs that allow him or her to become physically violent toward others in situations where social control is perceived as necessary. These beliefs are seen as a way to obtain respect among peers and in the community as a whole. The reputation gained through adhering to these codes provides protection against violence. These codes also allow for functional recourse against violations against the individual. Unlike their counterparts outside disadvantaged communities, who can use the police and the courts as allies, individuals in communities marked by street codes rely on violence. Put another way, having a reputation for violence enhances "respect," which in turn deters would-be offenders (who would thus risk their lives by trying a person who has "respect").

Contrary to Anderson's assertion that the code of the street protects individuals from harm, research has shown that African-American youth who reported having attitudes favorable to the use of violence, in fact, had a greater risk of experiencing violent victimization. A system of beliefs that espoused violence and retaliation, rather than being protective, was correlated with more violent victimization. Similar conclusions have been drawn from studies on youth gangs. Although many youth join together to form delinquent gangs for protection, research has shown that gangs, in fact, increase the likelihood that a person will experience violent victimization. In short, victims of violence often have profiles similar to offenders when it comes to attitudes toward violence; both tend to show support for violence as a mechanism for addressing disrespect and physical injury.

Conclusion

Juvenile victimization is a topic of concern, but it is nevertheless perplexing. On the one hand, there are the occasional horrific accounts of youthful victimization reported in the media, which in turn creates a demand for a response from government authorities. This has resulted in considerable investment in school security, for instance. Research consistently shows that the effects of victimization upon youth appear connected with unfortunate decisions and outcomes later in life. Children victimized in school and maltreated at home are more likely to quit school, have a lower income than everyone else, and suffer employment and legal problems throughout their lives. From this perspective, any effort to protect children, no matter how expensive or invasive, would appear to be warranted.

On the other hand, victimization is an exceedingly rare occurrence even for children. Youth who become victims are most likely to have been targets of theft. The percentage of children who have been maltreated at home, who have been violently attacked elsewhere, and who have directly experienced bullying at school is very small. When violence happens, the incident itself tends to be not very serious in terms of immediate effects; however, research indicates that there are many adverse long-term effects that degrade a youthful victim's quality of life as an adult. Levels of victimization are slightly higher at school than outside of school, but this paints a somewhat misleading picture. Much of the victimization is in the form of theft and minor assaults. Serious violence is most likely to occur off of school grounds, where school security cannot offer any protection.

The research also indicates that many efforts intended to curtail the onslaught of youthful victimization are disconnected from the reality of where and when the victimization occurs and to whom. Increasing the safety of students is of course a laudable goal, but misallocating resources toward ineffective programs – and even policies that elevate student fear – undermine the efforts of society to produce children who can function independently as adults.

Research is still attempting to understand the reasons some youths fall victim and not others. There is a clear profile of the families that are most likely to mistreat their children. These theories have focused on the situations that juveniles often are in (and how they fuel victimization risk), their levels of self-control, degrees of emotional distress, and their attitudes toward violence. It is interesting to note that many theories of victimization highlight variables that are consequences of childhood maltreatment, which in turn are correlates of victimization in adolescence and adulthood.

Recommended Reading and References

Anderson E (1999) Code of the street. Norton, New York

Bureau of Justice Statistics (2008) Criminal victimization, 2007. US Department of Justice, Washington, DC

Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44:588–608

Felson RB (1992) Kick 'em when they're down: explanations of the relationship between stress and interpersonal aggression and violence. Sociol Q 33:1–16

Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Hindelang MJ, Gottfredson MR, Garofalo J (1978) Victims of personal crime: an empirical foundation for a theory of personal victimization. Ballinger, Cambridge

Lauritsen JL, Laub JH (2007) Understanding the link between victimization and offending: new reflections on an old idea. Crime Prev Stud 22:55–75

Lauritsen JL, Sampson RJ, Laub JH (1991) The link between offending and victimization among adolescents. Criminology 29:265–291

Macmillan R (2001) Violence and the life course: the consequences of victimization for personal and social development. Annu Rev Sociol 27:1–22

Melde C, Taylor TJ, Esbensen FA (2009) I Got your back: an examination of the protective function of gang membership in adolescence. Criminology 47:565–594

Olweus D (1978) Aggression in the schools: bullies and whipping boys. Halsted, New York

Robers S, Zhang J, Truman J (2010) Indicators of school crime and safety. National Center for Education Statistics, Washington, DC

Schreck CJ (1999) Criminal victimization and low self-control: an extension and test of a general theory of crime. Justice Q 16:633–654

Schreck CJ, Miller JM (2003) Sources of fear of crime at school: what is the relative contribution of disorder, individual characteristics, and school security? J Sch Violence 2:57–79

Schreck CJ, EA Stewart (2011) The victim-offender overlap and its implications for juvenile justice. In Oxford handbook of juvenile crime and juvenile justice. New York: Oxford University Press

Schreck CJ, Miller JM, Gibson C (2003) Trouble in the school yard: a study of the risk factors of victimization at school. Crime Delinq 49:460–484

Schreck CJ, Stewart EA, Osgood DW (2008) A reappraisal of the overlap of violent offenders and victims. Criminology 46:871–906

Singer S (1981) Homogeneous victim-offender populations: a review and some research implications. J Criminal Law Criminol 72:779–788

Stewart EA, Schreck CJ, Simons RL (2006) I ain't gonna let no one disrespect me: does the code of the street reduce or increase violent victimization among African American adolescents? J Res Crime Delinq 43:427–458

US Department of Health and Human Services, Administration for Children and Families, Administration on Children, Youth and Families, Children's Bureau (2010) Child maltreatment 2009. Available from http://www.acf.hhs.gov/programs/cb/stats_research/index.htm#can. Washington, DC

Widom CS (1989) The cycle of violence. Science 244:160–166

Wolfgang M, Ferracutti F (1967) The subculture of violence: towards an integrated theory in criminology. Tavistock, London

Juvenile Violence

James Alan Fox[1], Michael Rocque[2] and Chad Posick[1]
[1]School of Criminology and Criminal Justice, Northeastern University, Boston, MA, USA
[2]Department of Sociology, University of Maine, Orono, ME, USA

Synonyms

Criminological theory; Juvenile crime; Trends

Overview

In the mid- to late 1990s, America was seemingly faced with a frightening epidemic: senseless acts of large-scale violence taking place in ostensibly safe, middle-class schools. Multiple instances of rampage-style shootings occurred in suburban and rural locales across the nation, the most deadly of which was the April 20, 1999 massacre at Columbine High School. Schools no longer *felt* safe. However, despite several isolated cases of school shootings involving multiple (oftentimes random) victims, schools *were* safe. In fact, by the late 1990s, school-related homicides were on the decline – a trend that has continued to the present. Thus, notwithstanding the seriousness of the rampage school shootings of the late 1990s, defining the problem as an epidemic was a hyperbolic overreach. Figure 1 displays school-related homicides since the early 1990s. As shown, the actual count of homicide victims has remained relatively low over time.

Despite this disjuncture between impression and reality with respect to school danger, juvenile violence – both inside and outside of school – remains a pressing concern for the public, criminologists, and criminal justice professionals alike. As such, it is important to have timely and accurate data on which to base policy. Without such data, criminal justice policies are more likely to be based on media-driven panics, as in

Juvenile Violence, Fig. 1 School-related homicides, 1992–2010

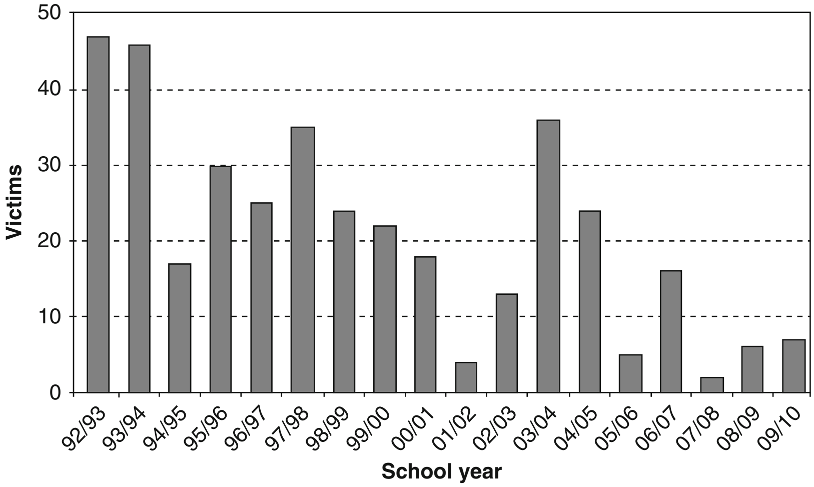

the last 10–15 years, when ineffective and unnecessary measures, including zero-tolerance and fortress-like access control, were implemented within schools.

The purpose of this entry is to review patterns and trends in juvenile violence, using the most up-to-date data available. The first section of the entry provides definitions and describes the categories of violence used by law enforcement officials and researchers. The second section provides some historical context for the state of juvenile violence in the United States today. This is followed by a review of trends in juvenile violence, examining how different types of offending have shifted or remained stable over the years. This section also briefly highlights several explanations for juvenile violence. The entry concludes with a discussion of policy implications emerging from the juvenile violence data.

Definitions of Juvenile Violence

Juvenile violence is defined as the threat, attempt, or actual use of force by one or more persons that results in physical or nonphysical harm to one or more persons, committed by a person under the age of 18 (Barkan 2009). Generally, behavior is only considered "violence" if the perpetrator caused harm (either physical or emotional) to a victim. Juvenile violence can be classified in a number of different ways, including the Federal Bureau of Investigation's designation of Part I (criminal homicide, forcible rape, robbery, and aggravated assault) and Part II (less serious acts such as carrying a weapon and simple assault) violent crimes.

A Brief History

A comprehensive history of juvenile violence is beyond the scope of this entry. However, because the majority of the information presented here refers to the modern period (i.e., mid-1970s to the present), it is important to provide some context. Juvenile violence is hardly a new phenomenon. While data and research on such behavior have increased in sophistication over time, juveniles have presumably engaged in illegal acts of violence throughout history. Records indicate that the first juvenile to be sentenced to death in Colonial America was Thomas Graunger, for the offense of bestiality committed in 1642 (Twersky-Glasner 2007). Reasonably valid statistics are not available until the nineteenth century. Before that time, juveniles accused of crimes were often treated as, and imprisoned with, adults by the criminal justice system. William Blackstone's *Commentary on the Laws of England*, from which American Law in the eighteenth Century was largely derived, argued that the "malice supplies the age" – that is, the seriousness of the offense determined how the offender was treated (ABA 2007). With the establishment of houses of refuge in 1820, juvenile offenders began to be increasingly separated from adult offenders. Records of offenders incarcerated in reformatories from 1890 show that juveniles were imprisoned for such crimes as homicide, rape, and assault, though these cases were rare (Walker 2007).

In 1899, the juvenile justice system with which most people are familiar today was ushered in with the formation of the first juvenile court in Cook County, Illinois. The model was initiated to provide guidance for wayward youth, rather than punishment. However, the middle part of the twentieth century saw many changes to juvenile justice, leading to a system that is increasingly similar to the adult system. While there remains a separation between juvenile and adult criminal justice systems, the distinction is blurring, with increasing stipulations allowing juveniles to be tried as adults (Singer 1996).

Juvenile Violence: Correlates, Trends, and Theory

There are two main sources of data that researchers use to study juvenile violence. First, is the Uniform Crime Report (UCR) which is released yearly by the FBI. The UCR is based on information gathered by local police departments who submit their data to the FBI. Thus, the UCR is comprised of "crimes known to the police." The FBI also assembles and disseminates arrest data, which are more detailed – because of the inclusion of offender information – but underestimate the number of actual crimes that occur, and so may be somewhat biased (relatively few crimes, e.g., result in an arrest). UCR data for juveniles show that the number and rate of arrests for violent crimes is inversely related to the seriousness of the act – in other words, the likelihood of a juvenile committing and being arrested for a violent act declines as the seriousness of the crime increases.

In addition to the long-standing UCR program, the National Incident-Based Reporting System (NIBRS) is a relatively recent development in which the police report a wide array of details on crimes, offenses, arrests, and victims, providing researchers with extensive information with which to examine the correlates and causes of crime. However, most police agencies (especially larger ones) do not as yet participate in the NIBRS initiative, thereby limiting its value for tracking patterns and trends in criminal behavior. Finally, the Supplementary Homicide Reports (SHR), available in its current form since 1976, provides detailed information on over 90 % of the homicides in the United States. Because of the level of detail and high degree of coverage, SHR data are used here extensively for examining the nature and extent of juvenile violence.

With respect to "unofficial" sources of data on juvenile violence, the most common method used by researchers involves surveys. When conducting self-report surveys, researchers gather a sample of youth and ask them about whether they have committed various criminal acts. Examples of such studies are the National Longitudinal Study of Youth, the National Youth Study, and the Behavioral Risk Factor Surveillance System. Research has indicated that despite some shortcomings, self-report surveys of violence tend to be fairly reliable and valid sources of information (Hindelang et al. 1981; Mosher et al. 2002).

One reason that academics and practitioners are concerned with juvenile delinquency is the disproportionate amount of crime committed by adolescents. Another reason for concern is the consistent finding that individuals who are involved in offending are also more likely to be victimized (Posick 2013). In other words, offenders are more likely to end up as victims compared to those who avoid criminal behavior. Taken together, juveniles are disproportionately involved in violence as both offenders and victims. Trends in juvenile delinquency allow researchers to highlight potential patterns and causes of crime and delinquency that can be addressed through prevention and intervention programs.

The Rise and Fall of Juvenile Violence

The late 1980s and early 1990s saw a significant and disturbing surge in youth violence, including homicide. While the overall violent crime rate also rose as a whole during this period, this was largely due to increases in violence among juveniles and young adults. Criminologists disagree about exactly why the youth crime rate rose so rapidly during this period. Some have pointed to the emerging crack markets and the violence that surrounded drug selling in the 1980s, while others implicated the decline of intact families and family values. Regardless of the explanation of this increase, the rise in juvenile violence was experienced nearly ubiquitously across the nation, although it was far more pronounced in urban areas.

Following the sharp rise in juvenile violence that peaked in the early 1990s, the USA experienced a major decline in youth violence that in large measure continues to this day. Some of this decline was expected as rates were already so high that they essentially had almost nowhere else to go but down. However, research has indicated that this decline was steeper than expected and not only due to a return to normal levels. For example, by some measures, crime fell so much during the mid- and late 1990s that it reached levels not seen since the 1960s. Just as criminologists were interested in explaining the rise in crime rates from the 1980s until the early 1990s, they were equally interested in explaining the decline. Criminologists also disagree about why there was a drop in violent crime in particular, especially among juveniles. Researchers have attributed the drop to several factors, including:

- Shifting demographics (e.g., aging of the population)
- The decline of crack markets
- Effective policing strategies
- Legalization of abortion (reducing the supply of youths likely to commit crimes)
- Enhanced enforcement of gun laws
- Increased use of incarceration and longer sentences

Incarceration may or may not have been a major contributor to declining, rates of juvenile crime rates in the 1990s; however, with the "get tough" approach that began to characterize justice systems during that time period, juveniles were increasingly charged and punished as adults. Thus, incarceration of juveniles may have played a role in the juvenile crime drop. Despite the disagreement on the cause of the crime decline, there is little debate over the extent of this decline.

More recent trends in juvenile violence show that while rates declined from the early 1990s to the early twenty-first century, they now appear to be plateauing. Research suggests, however, that some forms of juvenile violence, such as gang activity, are rebounding. Predicting where the trend will move over the next decade is rather difficult; only time will tell.

Trends within the United States show peaks and valleys, but the rank order of the country in comparison to others around the globe has remained steady. There is little controversy over the violent nature of the United States in comparison to other developed countries. The USA remains consistently ranked near the top in violent crime, particularly murder. Per capita, the USA is one of the most lethal developed nations in the world. These rates are largely driven by violence by teens and young adults. Self-report surveys also confirm this pattern. The International Self-Report of Delinquency-II (ISRD-II) study has shown that youth in the "Anglo-Saxon" country cluster (which is comprised of Ireland, Canada, and the United States) has the highest rates of violence (Junger-Tas et al. 2012).

Juvenile crime rates have risen and fallen but there are certain elements of juvenile offending that have remained fairly consistent across time. For example, males are more likely than females to be involved with violent crime, and this holds across all ages and demographic groups. This fact is illustrated in Fig. 2.

As can be seen in Fig. 2, which displays UCR arrest data from 2000 to 2010 for males and females under the age of 18, males have a consistently higher arrest rate than females. The figure shows all Part I violent crimes combined (homicide, rape, robbery, and aggravated assault). Interestingly, the trend for that time

Juvenile Violence, Fig. 2 Juvenile arrest rates by sex, 2000–2010

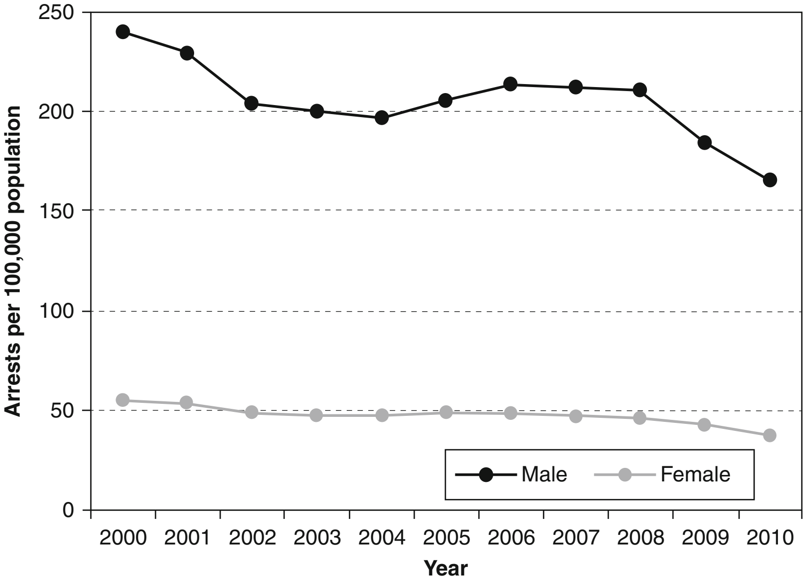

period was very similar for both males and females, dipping a bit around 2002, peaking around 2008 and then falling again in 2010. While males have higher arrest rates than females for all violent crimes, the disparity is most stark for homicide, robbery, and especially rape. The violent crime in which male and female rates are the closest is aggravated assault (where male rates are around three times higher than female rates).

Youth Homicide

Homicide, or lethal violence, is the most feared crime and also the one for which the most detailed information (provided by the Supplementary Homicide Reports) is available. Data indicate that lethal violence is not distributed evenly across the population. In other words, there are certain clear and consistent demographic, geographical, and temporal patterns that data have revealed with respect to juvenile homicide. The following figures explore how lethal violence is distributed demographically and geographically, and how these patterns have changed over time.

As is shown in Fig. 3, homicide is an offense committed disproportionately by juveniles and young adults. People over 25 years of age have a much lower rate of homicide. Young adults ages 18–24 are the most likely homicide perpetrators followed by youth ages 14–17. In terms of trends, homicide rates for all age groups increased in the 1970s and then began to decline in the early 1980s. Prior to the 1980s, the homicide rates for youths aged 14–17 and those aged 25 were virtually indistinguishable. However, as noted above, at that point, there was a noticeable spike in homicide by young offenders. By the early 1990s, the homicide rates for both the 14–17 and 18–24 age categories had nearly doubled their 1980 levels. Interestingly, the homicide rate for the over 25 age group remained relatively unchanged during the 1980s, compared to the other two age groups. During this time, some scholars worried about a violence epidemic gripping the nation’s youth and argued that the existence of “super-predators” may account for some of this increase in violence and indeed would continue to keep homicide rates high (Bennett et al. 1996).

However, as discussed earlier, in the United States as well as other western, industrial nations, beginning in the early 1990s, crime began to fall. Homicide rates dipped to the lowest they had been in decades. As is shown in Fig. 3, homicide rates for juveniles and young adults stabilized

Juvenile Violence, Fig. 3 Homicide offending rate by age group, 1976–2010

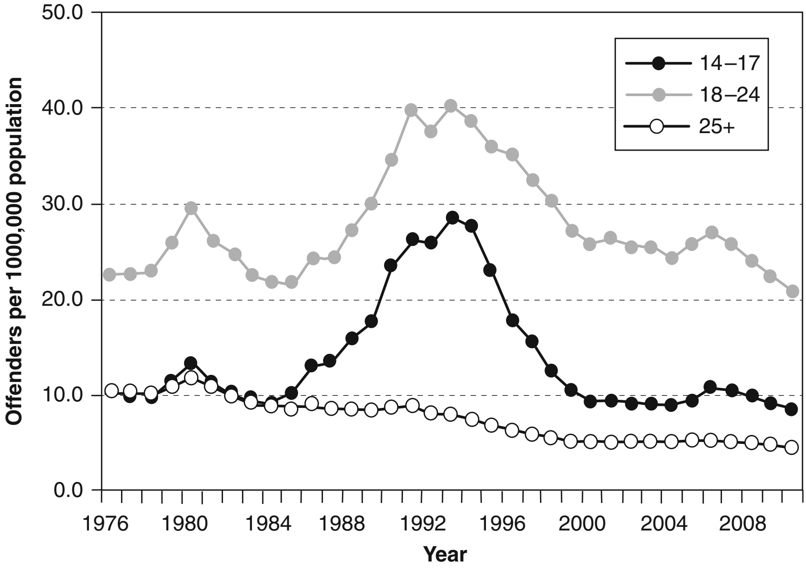

Juvenile Violence, Fig. 4 Percent gun homicide by age group, 1976–2010

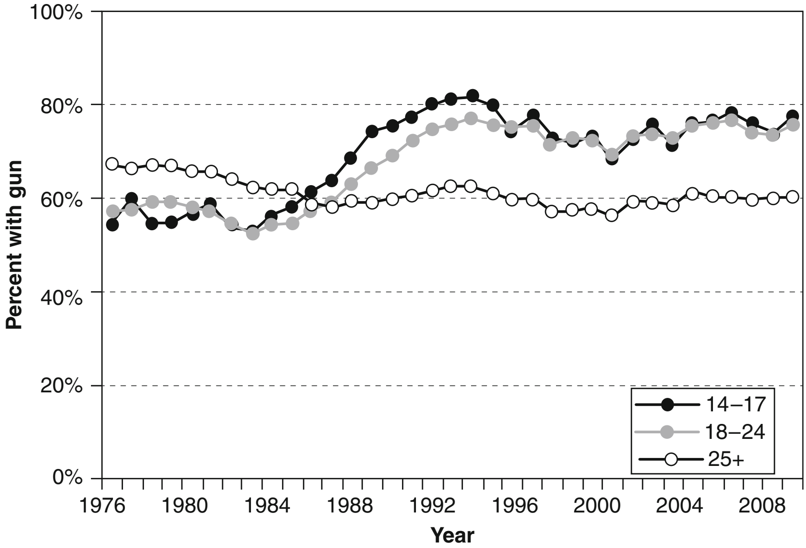

around the year 2000, but since then have continued to dip. This “great American crime decline” remains largely unexplained, although the reductions in the crack cocaine market and use of handguns by juveniles likely are contributing factors (Blumstein and Wallman 2006). As Fox et al. (2012) note, the rise of the crack market in inner cities during the late 1980s and early 1990s resulted in a dangerous and often very violent “arms race” in which profits had to be protected by the youths directly involved. It appears that when this market declined, the juvenile homicide rate followed suit.

The emerging crack market required that drug couriers defend themselves and their territories with deadly weapons, namely, guns (Blumstein 1995). As shown in Fig. 4, homicides involving guns increased in the mid-1990s and then declined for both juveniles and young adults. Much like the trend shown in Fig. 3, the rate of

Juvenile Violence, Fig. 5 Homicide offending rate by race, 1976–2010

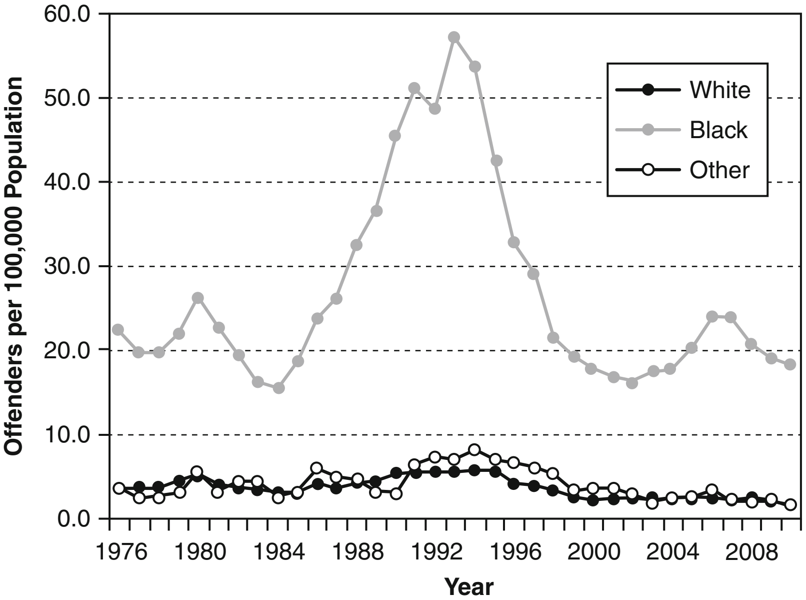

homicides using a gun did not follow the same pattern for those aged 25 and older. Thus, it would seem that much of the decline in violent crime for juveniles and young adults was due to the changing crack market and the use of guns accompanying it. While crack use was still fairly prevalent, the marketplace for the drug was not nearly as volatile, competitive, and/or violent.

As can be seen in Fig. 4, in the past four decades, a solid majority of all homicides were committed with a gun. This finding is consistent across all age groups; in fact, the 14–17 and 18–24 age groups are almost identical in their pattern throughout the past several decades. During the peak of the violent crime spike, over 80 % of the homicides committed by the youngest group were committed with guns. Many researchers suggest that it may be the widespread availability of guns that contributes to high homicide rates in the USA and that this may explain some of the peaks and valleys in the patterns of lethal violence.

When one considers the saga of the crime decline in the 1990s following the "near-epidemic" levels in the late 1980s, perhaps the starkest trend can be found with respect to race. As shown in Fig. 5, the peak in homicides was much more drastic for blacks than for other racial groups. As is well known, the crack market was concentrated among minority groups, particularly blacks. This lends more credence to the thesis that the crack market was a large contributing factor to the rise and fall of juvenile violence since the mid-1980s.

In comparison to other crimes, homicide is unevenly distributed across racial groups. Regardless of the trend, blacks have a much higher homicide rate than any other racial group. Whites and other races have had almost identical rates throughout the last few decades, which range from about four times lower (in the early 1980s) to about six times lower (in the mid- to late 1980s) than the rate for blacks.

The crack market, as previously noted, was concentrated in the minority-dominated, inner cities. As shown in Fig. 6, the rise and fall of juvenile homicides was most pronounced in large cities. This again supports the thesis that the crack market and associated gang activity played a large role in juvenile homicide patterns since the early 1990s. Nonetheless, juvenile homicide rose and fell in cities of all sizes, suggesting that crack and gun use were not the only factors affecting violent crime rates.

In general, homicide is not evenly distributed geographically and this is true for juvenile

Juvenile Violence, Fig. 6 Juvenile homicide offenders by location, 1976–2010

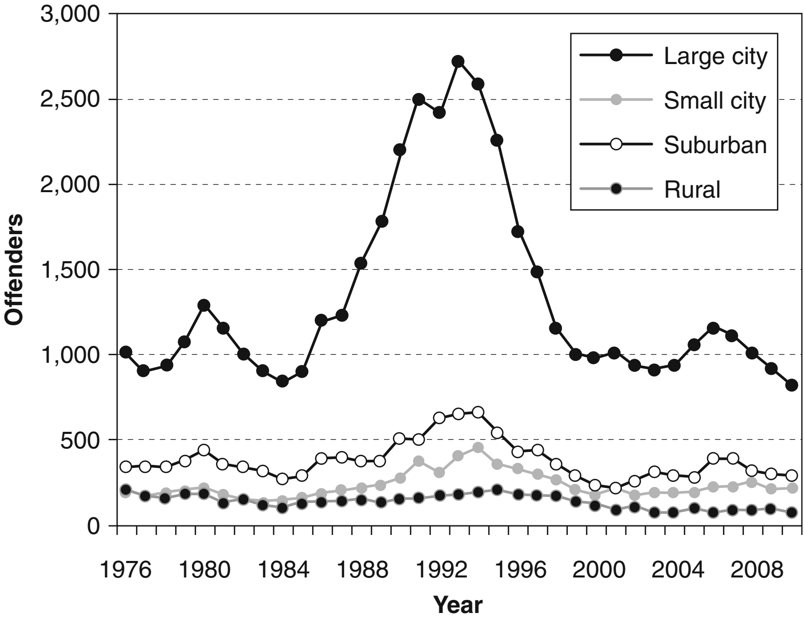

homicide as well. As Fig. 6 indicates, the highest juvenile homicide rates are found in large cities. These are followed by suburban, small cities, and rural areas. The disparity between large and smaller cities with respect to juvenile homicide is similar to that between blacks and other races, as discussed previously. For example, at the point of the homicide peak, large cities had a homicide rate of more than four times that of other city sizes. That this disparity (despite some fluctuations) is relatively consistent across time suggests that factors such as poverty, relative deprivation and "subcultures of violence" may contribute to the variation in homicide rate by city size. A counterargument to the "more guns, more violence" hypothesis is that guns are also prevalent in rural areas where hunting and gun culture is prominent. Perhaps it is not guns alone, but guns in conjunction with a culture of violence that influences violent crime rates.

Figure 7 illustrates trends by age group for homicides committed by multiple offenders. There is a near-linear increase over time for each age group, suggesting that more homicides are being committed in partnerships. While homicide is generally committed by a lone offender, younger individuals are more likely than older individuals to commit homicide with other offenders. In other words, violent offending is more of a group activity among youths. Younger individuals are more likely to associate with and be influenced by delinquent peers, while older individuals are more likely to be involved with family or work. This is concordant with research that shows juveniles are more affected by peer influence and more likely to offend in groups early in life as compared to later (Farrington 2003). About 20–25 % of homicide offenders ages 14–17 commit the crime with the help of accomplices, compared to 15–20 % for 18–24-year-old assailants and only about 10–15 % for offenders over age 25. When older individuals do commit lethal violence, it is more likely to be done alone – perhaps related to domestic or intimate partner violence. In general, the literature suggests that adults tend to commit crime alone, although lethal violence in particular is among the least likely to be committed with co-offenders.

The timing of juvenile violence (e.g., time of day, day of week, and season of year) is another key piece of information with clear policy implications. Logic would suggest that youth crime should peak in the early afternoon, directly after school lets out), and data support this notion. In Fig. 8, NIBRS data are used to examine patterns of violent

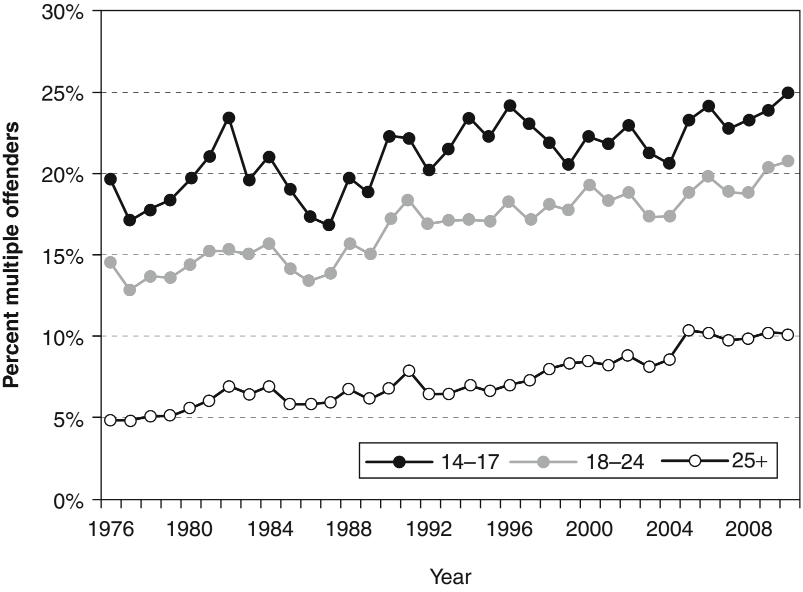

Juvenile Violence, Fig. 7 Percentage multiple offenders by age group, 1976–2010

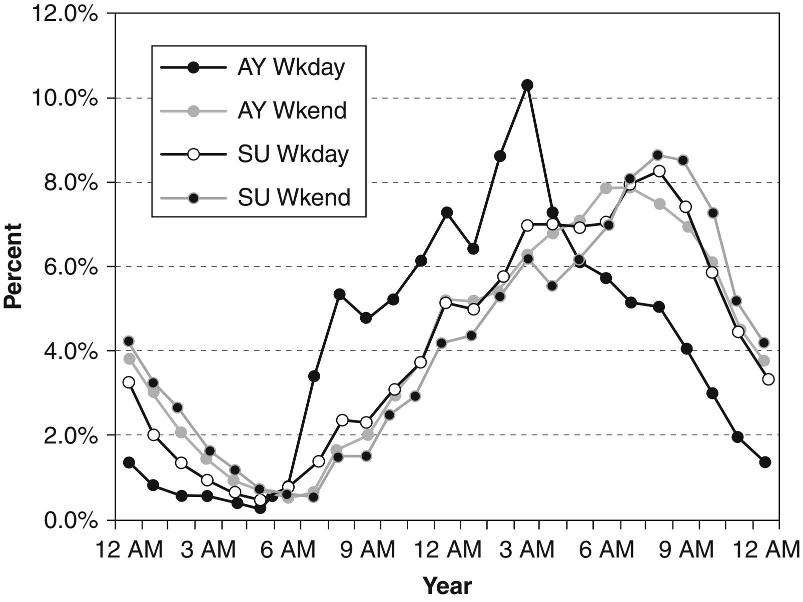

Juvenile Violence, Fig. 8 Time of day patterns for juvenile offending by time of the week, 2009

crime committed by juveniles by time of day for the academic year (AY) and summer months (SU), for weekdays and weekends separately. While the overall patterns for offending are fairly similar, there are two deviations that are worthy of attention. First, a spike is apparent in offending around 3 pm on weekdays during the school year. This uptick is related to the routine school-day activities of likely offenders and victims in the after-school hours. Students leave school around 2–3 pm, bringing together many youth who, now out of school but not yet at home, tend to be less supervised by teachers and parents. This situation is prime for both offending and victimization.

A second deviation from the general offending pattern is seen in offending around 9 pm in the

Juvenile Violence, Table 1 Patterns in juvenile homicide by time period

		Year group		
Item	Category	1976–1985	1986–1995	1996–2010
Victim age	Under 14	5.00 %	4.20 %	4.20 %
	14–17	17.00 %	20.60 %	19.50 %
	18–24	27.20 %	31.70 %	33.30 %
	25–34	16.10 %	20.50 %	17.40 %
	35–49	16.30 %	15.10 %	16.30 %
	50–64	9.50 %	4.60 %	6.00 %
	65+	8.90 %	3.30 %	3.30 %
Relationship	Intimate	3.30 %	2.40 %	2.20 %
	Family	13.40 %	8.50 %	9.20 %
	Acquaintance	47.10 %	52.70 %	52.10 %
	Stranger	36.10 %	36.30 %	36.50 %
Weapon	Gun	56.00 %	76.00 %	74.50 %
	Knife	22.90 %	10.90 %	9.90 %
	Other	21.10 %	13.10 %	15.60 %
Circumstances	Felony	35.50 %	33.70 %	32.60 %
	Argument	36.30 %	34.70 %	33.70 %
	Other	28.50 %	31.60 %	33.70 %

summer and on the weekends. Again, this is likely due to the routine activity patterns by young people (those aged 18 and under) during these hours. After work and during the weekends, adolescents often enjoy partying and carousing. This brings together young people who may be under the influence of alcohol or drugs in environments with little formal supervision. During the academic year, when youth are in school, the percentage of total offending during these later hours is smaller because students are more likely to be involved with schoolwork or asleep during this time of day.

Figure 8 suggests that crime prevention efforts should be focused on peak hours of offending (which covary with victimization) to prevent violence. Some research in criminology has examined the extent to which after-school programs can reduce offending. This research has shown that the peak in offending after school tends to be starker for official records than for self-reports. In addition, after-school programs do, in fact, reduce juvenile offending, in part because of an increase in supervision, but more importantly because of effective socialization that reduces antisocial tendencies (Gottfredson et al. 2001, 2004).

The discussion of trends and correlates of juvenile violence closes with a look at the characteristics of juvenile offenders, including their victims across several decades. These data are shown in Table 1. The first panel of Table 1 illustrates patterns with respect to victim age. Since the first time period (1976–1985), victims of teenagers have tended to be younger. For example, in the period between 1976 and 1985, 17 % of victims were aged 14–17 compared to around 20 % in later years. In addition, about 18 % of victims were over the age of 50 in the first time period, compared to around 8–9 % in later years. Overall, mirroring the offender ages for homicide, the largest age category for victims in all years is 18–24. This supports the argument that offenders and victims of lethal violence tend to share demographic characteristics.

In terms of the relationship between victim and offender for juvenile homicide, most are acquaintances. The relative rank ordering of these relationships has remained consistent over time. However, there has been an increase in the percentage of victims who are acquaintances, and a decrease in family homicides since the 1976–1985 time period. The percentage of homicide victims who are intimates or strangers has remained relatively consistent over time.

Juvenile homicides also predominantly involve guns. This is shown in the third panel of Table 1. Interestingly the percentage of juvenile homicides using a gun increased dramatically in the mid-1980s from 56 % to 76 %. This follows the trend shown in Fig. 4.

The circumstances leading to juvenile homicide have not changed much over time. As shown in the bottom panel of Table 1, the three classifications of circumstances (felony-related, argument-related, or other) are relatively evenly distributed, with argument-related episodes slightly more commonplace. Adolescents tend to act impulsively and have lower self-control. As a result, most juvenile homicides appear likely spontaneous or episodic, rather than being planned and premeditated.

Conclusion

Juvenile violence remains a significant concern for policy-makers, criminologists, and the public alike. Too often, however, misconceptions fueled by sensational media reports lead to policies or programs that either do not work or actually serve to make things worse. In the 1980s, two particular programs (Scared Straight and D.A.R.E.) were popular and supported handsomely by government funding. Unfortunately, evaluations of these initiatives have shown that they did not work; in fact, in some situations, they actually increased the risk of offending (see Mackenzie 2006).

Because juvenile crime – especially violence – remains a significant concern, reducing its occurrence continues to be a priority. The best interventions and prevention strategies are based upon a solid understanding of the patterns of juvenile violence, including prevalence, trends, and correlates. As discussed here, juvenile violence has followed some pronounced trends over the last 30 years, including a major spike in the 1980s (especially for inner-city, minority youths using handguns), followed by a large decline that in large measure continues to this day.

Juvenile violence also seems to cluster in terms of day of the week and time of day. As we have shown, crime tends to increase on the weekends and during after-school hours. This information represents a powerful tool for policy-makers who wish to design effective interventions that will decrease juvenile violence. However, it is important to bear in mind that it is not solely supervision or surveillance that seems to be what matters with respect to reducing violence, but impacting attitudes and other risk factors related to crime. Those wishing to reduce or prevent juvenile violence would do well to remember that it is not a simple phenomenon, but one that is complex and influenced by a variety of factors, including peer influences and lack of self-control (see Farrington 2007). Understanding what juvenile violence is (and is not) as well as its trends and correlates is an important first step in achieving this goal.

Related Entries

- ▶ Co-offending
- ▶ Development of the UCR and the NCVS
- ▶ History of Criminological Theories: Causes of Crime
- ▶ History of Juvenile Justice
- ▶ Juvenile Justice in the Get Tough Era
- ▶ Juvenile Victimization
- ▶ Law, Diversion and Community Sanctions in Juvenile Justice
- ▶ School Crime Statistics
- ▶ School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis
- ▶ Theories of Co-offending
- ▶ Youth Homicide in the United States

Recommended Reading and References

ABA (2007) Dialogue on youth justice. American Bar Association, Washington, DC

Barkan SE (2009) Criminology: a sociological understanding. Pearson, Upper Saddle River

Bennett WJ, DiIulio JJ, Walters JP (1996) Body count: moral poverty and how to win America's war against crime and drugs. Simon and Schuster, New York

Blumstein A (1995) Youth violence, guns, and the illicit drug industry. J Crim Law Criminol 86(1):10–36

Blumstein A, Wallman J (2006) The crime drop in America. Cambridge University Press, New York

Farrington DP (2003) Developmental and life-course criminology: key theoretical and empirical issues—the 2002 Sutherland award address. Criminology 41(2):221–256

Farrington DP (2007) Developmental criminology and risk-focused prevention. In Maguire M, Morgan R & Reiner R (eds) The Oxford handbook of criminology (Vol. 4, pp. 657–701). Oxford, UK: Oxford University Press

Fox JA, Levin J, Quinet K (2012) The will to kill. Pearson, Upper Saddle River

Gottfredson DC, Gottfredson GD, Weisman SA (2001) The timing of delinquent behavior and its implications for after-school programs. Criminol Public Policy 1(1):61–86

Gottfredson DC, Gerstenblith SA, Soulé DA, Womer SC, Lu S (2004) Do after school programs reduce delinquency? Prev Sci 5(4):253–266

Hindelang MJ, Hirschi T, Weis JG (1981) Measuring delinquency. Sage, Beverly Hills

Junger-Tas J, Marshall IH, Enzmann D, Killias M, Steketee M, Gruszczynska B (2012) The many faces of youth crime: contrasting theoretical perspectives on juvenile delinquency across countries and cultures. Springer, New York

Mackenzie DL (2006) What works in corrections: Reducing the criminal activities of offenders and delinquents. New York: Cambridge University Press

Mosher CJ, Hart TC, Miethe TD (2002) The mismeasure of crime. Sage, Beverly Hills

Posick C (2013) The overlap between offending and victimization among adolescents Results from the second international self-report delinquency study. Journal of Contemporary Criminal Justice. doi: 10.1177/1043986212471250

Singer SI (1996) Recriminalizing delinquency: violent juvenile crime and juvenile justice reform. Cambridge University Press, New York

Twersky-Glasner A (2007) Juvenile violence, 1600–1800 (Colonial Era). In: Finley LL (ed) The encyclopedia of juvenile violence. Greenwood Press, Westport

Walker L (2007) Juvenile violence, 1861–1885 (Civil War Era). In: Finley LL (ed) Encyclopedia of juvenile violence. Greenwood Press, Westport

Juveniles, Young Adults

► Youth Homicide in the United States

K

Key Skills Involved in Judging

► Evolving Judicial Roles

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

L

Labeling and Deviance

▶ Cultural Criminology

Labeling Theory

Erich Goode
Sociology Professor Emeritus, Stony Brook
University, Stony Brook, NY, USA

Overview

Deviance is a label attached to certain acts and individuals; those acts need not be harmful to be designated as bad, and those individuals may not have even committed the acts in question, but it is the label – the designation, the *judgment* – that renders the acts and the individuals as deviant. Other acts may be more harmful but considered less heinous; other individuals may be more dastardly without attracting an equal measure of stigma. Labeling theory argues that, from a sociological perspective, what *counts* is this designation. Other theories of deviance attempt to explain the incidence or prevalence of concretely real acts with concretely real consequences – robbery, adultery, murder, drug use, rape, and the like. Labeling theorists do not say that such designated behavior is not real, or that its consequences are not real, until they are labeled; they say that the behavior is not specifically *deviant* until the label (self-labeling included) is applied. One man may kill another, but it is not *murder* – a deviant category of killing – until audiences (a judge, a jury, a community, oneself) label the act *as* murder. It is still a killing: Someone has died, and forensics may be able to determine whether the first man caused the second's death. What is deviant is determined socially and culturally. Labeling theory is a constructionist perspective *par excellence*.

Deviance may be viewed *horizontally* or *vertically*. At the micro level, deviance is manifested by how one or more persons or members of one small collectivity react to another; labeling entails judgments made by individuals or social circles. When Person A evaluates and reacts to Person B's behavior, Person A acts as an "audience." Multiplied dozens of times, this represents the labeling process at the horizontal level. In contrast, *vertically*, labeling can be viewed at the cultural, institutional, and social-structural level. When labeling is institutionalized, it is coagulated into preexisting potential judgments, analogous to socioeconomic status or occupational prestige; anyone who is a member of the relevant society is likely to be subject to them. And institutionally, organizations possess the capacity to label and deal with individuals as *deviants*. A prison has the power to label, and treat an inmate as, a *convict*; a school, college, and university possess the power to judge student a *failure* – a loser, an educational flop; the psychiatric profession, both individually and

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

collectively, dispense diagnoses of schizophrenia, autism, and bipolar disorder that can have consequences for the labelee as momentous as the conditions themselves. Hence, labeling tends to be *asymmetrical*: Although audiences with equal power and influence can label one another, vertically, the most consequential deviance-labeling processes flow from institutions to individuals who are so labeled.

The Emergence of Constructionism

Hints of the central concepts of deviance-labeling can be found in the works of Émile Durkheim, in *Suicide* and *The Elementary Forms of the Religions Life*; nonetheless, most scholars locate the roots of the labeling theory of deviance in the pragmatic school of George Herbert Mead (1863–1931). Pragmatism denies the significance of mysticism and essentialism, emphasizing instead that *the meaning is in the response* – what defines something is not its abstract, indwelling "isness," material or spiritual, but the concrete, identifiable consequences it has, the results it affects.

The connection between pragmatism and the perspective that defines deviance as a consequence of labeling is obvious. Rather than regarding deviance as a type of action with objective features, the labeling perspective views the violation of rules as an *infraction*, that is, what *makes* certain actions (and beliefs and conditions) infractions. Why are rules constructed and enforced? Who makes and forces the rules? Why are *certain kinds* of rules made? Why are certain *persons* or *types of persons* apprehended and punished? What *consequences* do rule-making and rule-enforcement have? This approach turns the focus of attention around. Now the spotlight is not on the rule-violator or the conditions that make for rule violation but on the society and the groups in the society that *make and enforce the rules*. And rules take on meaning only insofar as they are applied; deviant behavior is what tends to attract punishment, condemnation, stigma, censure; a deviant person is someone whose identity is saturated with public scorn.

Labeling or Interactionist Theory

In 1938, Tannenbaum published the volume *Crime and the Community*, which argued that in a slum area, nearly all boys engage in a wide range of mischievous, sometimes illegal behavior – getting into fights, skipping school, stealing apples, throwing rocks at windows. These actions, taken for granted by the boys themselves, are often regarded as deviant, even criminal, by the authorities – by the teachers, the police, and the courts. The police may admonish, informally punish, or arrest these boys. If these boys persist in this behavior, they may be sent to reform school. However, punishment could have the unintended, undesired, and ironic effect of escalating the seriousness of the deeds that these boys commit. Arrest and incarceration often result in the community regarding a boy as incorrigible, rebellious, unmanageable – a deviant. By being treated as a delinquent and forced to associate with slightly older and more experienced young criminals in reform schools, the troublemaker may come to adopt the identity of the delinquent. In all likelihood, this will escalate his deviant career – increasing the chance that he will go on to a life of crime. Labeling creates major crimes out of minor sins. Tannenbaum also emphasized that social characteristics – specifically social class – play a role in deviance labeling. Though he added labeling to the roster of the causes of deviance, Tannenbaum did not deny that other factors caused delinquency. *Crime in the Community* did not address the social construction of deviance, a central tenet of labeling theory; hence, Frank Tannenbaum is a *precursor* of labeling theory.

In 1951, Edwin Lemert published a textbook with the anachronistic title, *Social Pathology* (1951); Tannenbaum does not appear among its author index. Lemert distinguished between *primary* and *secondary* deviation. Primary deviation is simply the enactment of deviant behavior itself – any form of it. Lemert argued that primary deviation is *polygenetic* (1951, pp. 75–76) – caused by a wide range of factors. For instance, someone may drink heavily for a variety of reasons – the death of a loved one, a business failure, belonging to a group whose members call for heavy drinking, growing up in a subculture or a society in which heavy drinking is taken for granted. In fact, Lemert asserted, the original causes of a particular form of deviance is not especially important, and pursuing them, not especially fruitful. What counts is the social reaction *to* the behavior. Lemert's treatise represented a major departure from the mainstream of essentialistic/positivistic, or scientific cause-and-effect, reasoning about deviance.

Secondary deviation occurs when the individual who enacts deviant behavior deals with the problems created by social reactions *to* his or her primary deviations (1951, p. 76). "The secondary deviant, as opposed to his [or her] actions, is a person whose life and identity are organized around the facts of deviance" (1972, p. 63). When someone is isolated, singled out, stigmatized, or condemned for engaging in deviant behavior, it becomes necessary to *deal with* and *manage* this social reaction in certain ways. One comes to see oneself in a certain way, define oneself in different terms, adopt different roles, associate with different individuals. Being stigmatized forces one to become a deviant – to engage in secondary deviation. The concept of "secondary deviation" does not *necessarily* imply that labeling causes more frequent enactment of the behavior that generated the label, but it is more likely than in the absence of labeling.

During the late 1950s, Howard Becker circulated a manuscript, along with four previously published papers, that constituted the core of *Outsiders*. He mentions that he wrote the first draft of the volume without having read or cited Lemert's book (Debro 1970, p. 165), and so his reference to *Social Pathology* in the published version of this book (p. 9) was tacked on after the fact – in effect, a *bogus genealogy*. In the 1960s, Becker, along with a small group of like-minded researchers, produced a small body of work that exerted an enormous influence on the sociological approach to deviance; it came to be looked upon as a more or less unified perspective that is widely referred to as *labeling theory*. Labeling theory grew out of a more general perspective in sociology called *symbolic interactionism*. The interactionist approach is based on "three simple premises." First, people act on the basis of the *meaning* that things have for them. Second, this meaning grows out of *interaction* with others, especially intimate others. And third, meaning is continually modified by *interpretation* (Blumer 1969, p. 2). These three principles – meaning, interaction, and interpretation – form the core of symbolic interactionism and likewise of labeling theory as well. People are not simple "products" of their upbringing or socialization or of their environment, but they are active and creative in how they see and act on things in the world and arrive at what they think, how they feel, and what they do through a dynamic, creative process. All behavior, deviance included, is an interactional product; its properties and impact cannot be known until we understand how it is defined, conceptualized, interpreted, apprehended, and evaluated – in short, what it *means* to participants and relevant observers alike. Labeling theory is not a separate theory but a direct application of symbolic interactionism to phenomena than are designated as deviant.

According to Becker (1973, pp. 177–208) and Kitsuse (1972), labeling theory is not so much an explanation of why certain individuals engage in deviant behavior as it is a perspective whose main insight tells us that the labeling process is crucial and cannot be ignored. And labeling "theory" is not so much a theory as it is an orientation perspective, a set of useful concepts with no clear-cut propositions. In fact, most "labeling theorists" preferred the term, "the interactionist approach to deviance" to the easy-to-remember tag their perspective was stuck with – "labeling theory" (Becker 1973) – here we have an irony that even labeling theory came to be labeled. The labeling approach shifts attention away from the traditional question of "Why do they do it?" to a focus on how and why judgments of deviance come to be made and what their consequences are. What consequences does labeling have for stigmatization? What is the difference between enacting rule-breaking behavior which does not result in getting caught and enacting that same behavior and being publicly denounced for it?

These are some of the major issues labeling theorists have concerned themselves with.

In many ways, labeling theory is a model or quintessential example of the constructionist approach. It addresses issues such as the creation of deviant categories, the social construction of moral meanings and definitions, social and cultural relativity, the how and why of social control, the politics of deviance, the criminalization of behavior, and the role of contingency in the labeling process. Together, these issues constitute the foundation stone of constructionism. Several issues or concepts represent the hallmark of the labeling theory's central concerns: *audiences, labeling and stigma*, *reflexivity*, and *the "stickiness" of labels* and *the self-fulfilling prophecy*.

Audiences. An audience is an individual or any number of individuals who observe and evaluate an act, a condition, or an individual. An audience could be one's friends, relatives, or neighbors, coworkers, the police, teachers, a psychiatrist, bystanders or observers – even oneself, for one can be an observer and an evaluator of one's own behavior or condition (Becker 1963, p. 31). *Audiences* determine whether something or someone is deviant: no audience, no labeling, therefore, no deviance. However, an audience need not *directly* view an act, condition, or person; audiences can witness behavior or conditions "indirectly," that is, they can hear or be told about someone's behavior or condition, or they can simply have a negative or condemnatory attitude toward a class or category of behavior: "The critical variable in the study of deviance . . . is the social audience rather than the individual actor, since it is the audience which eventually determines whether or not any episode of behavior *or any class of episodes* is labeled deviant" (Erikson 1964, p. 11; my emphasis). In other words, audiences can evaluate *categories* of deviance and stand ready to condemn them, even before they have actually witnessed specific, concrete cases of these categories. Audiences are absolutely central in the sociological definition of deviance. Whether an act, a belief, or a trait is deviant or not depends on the audience who does or would evaluate and react to the actor, the believer, or the possessor accordingly. Without specifying real-life audiences, the question of an act's, a belief's, or a trait's deviance is meaningless. Audiences include the society at large, agents of formal social control, and the significant others of the actor, such as intimates – or any minority collectivity within a given society. Different audiences may or may not evaluate or react to the actor's behavior in the same way.

Labeling and stigma. The key elements in "becoming" deviant are labeling and stigma. The processes of labeling and stigmatizing are done by a relevant audience. The audience is relevant according to the circumstances or context. Gang members can label the behavior of a fellow member of the gang, for instance, the refusal to fight, as deviant *within the gang context*; the police can label an action of a gang member, for instance, fighting, as deviant, wrong, or illegal, by arresting or harassing him. Each and every audience or person can label each and every action, belief, or condition of each and every person as deviant – but the weight or *consequences* of that labeling process vary according to the context, as we saw in considering the *vertical* conception of deviance labeling.

The labeling or stigmatization process entails two steps. First, an audience labels an *activity* (or belief or condition) deviant, and second, it labels a specific *individual* as *a* deviant. In these two labeling processes, if no audience labels or would label a given person or someone deviant, strictly speaking, *no deviance exists*. An act, belief, condition, or person cannot be deviant *in the abstract*, that is, without reference to how an audience does or would label it. Something or someone must be defined as such by the members of a society or a group *as* deviant – it must be *labeled*, concretely or potentially, as reprehensible or wrong. An act, belief, or condition need not be *actually* or *concretely* labeled to be regarded as deviant, however, but it is deviant if it belongs to a category of similar actions, beliefs, or conditions. In other words, *we already know* that the public regards shooting the proprietor of a store and taking the contents of that store's cash register as deviant. Even if the robber gets away with the crime, *if known about*, the act is likely to be regarded as deviant in the society at large – that is, it is an instance of "societal deviance." At the same time, the role of *contingencies* – who, what, where, why, and so on – may influence the outcome of the labeling process.

Labeling involves attaching a *stigmatizing* definition to an activity, a belief, or a condition. Stigma is a stain, a sign of reproach or social undesirability, an indication to the world that one has been singled out as a shameful, morally discredited human being. Someone who has been stigmatized is a "marked" person; he or she has a "spoiled identity." Once someone has been thereby discredited, relations with conventional, respectable others become difficult, strained, problematic. In other words, "being caught and branded as a deviant has important consequences for one's further participation and self-image. . . . Committing the improper act and being publicly caught at it places [the individual] in a new status. He [or she] has been revealed as a different kind of person from the kind he [or she] was supposed to be. He [or she] is labeled a 'fairy,' 'dope fiend,' 'nut,' or 'lunatic,' and treated accordingly" (Becker 1963, pp. 31, 32).

So crucial is this labeling process that, in some respects, it does not necessarily matter whether or not someone who has been stigmatized has actually engaged in the behavior of which he or she is accused. According to the logic of labeling theory, *falsely accused* deviants – if the accusation sticks – are still deviants (Becker 1963, p. 20). In many important respects, they resemble individuals who really *do* commit acts that violate the rules. For example, women and men burned at the stake for the crime of witchcraft in the fifteenth and sixteenth centuries were deviants in the eyes of the authorities and the community, even though they clearly did not engage in a pact with the devil. Two individuals, one who engaged in a deviant act and the second of whom is falsely accused, will share important experiences and characteristics in common, *by virtue of the labeling process alone*, even though they are poles apart with respect to having committed the behavior of which they were accused. While their lives are unlikely to be *identical* simply because both are seen by the community as deviants, the similarities they share are likely to be revealing.

Reflexivity. Reflexivity means looking at oneself in part through the eyes of others. It is what is widely referred to, although too mechanistically, as the "looking glass self." Labeling theory is based on a seemingly simple but fundamental observation: As Mead and his pragmatic peers said, we tend to see ourselves through the eyes of others, and when others see us in a certain way, we tend to begin seeing ourselves that way, too. In other words, deviance labeling by others, if those others are sufficiently influential and the labeling persists long enough, may become *internalized*.

Both direct *and* indirect, or concrete *and* symbolic, labeling operate in the world of deviance. "Indirect" or "symbolic" labeling is the awareness by a deviance enactor that his or her behavior is saturated with public scorn, his or her identity is potentially discreditable, that he or she *would be* stigmatized if discovered. People who violate norms have to deal with the probable and potential, as well as the actual and concrete, reactions of the respectable, conventional, law-abiding majority. All violators of major norms must at least ask themselves how others react to them and their behavior. If the answer is that others will condemn and humiliate them, then the rule breaker must try to avoid detection, remain within deviant or minority circles, or be prepared to be punished and stigmatized.

The "stickiness" of labels and *the self-fulfilling prophecy*. Labeling theorists argue that stigmatizing someone as a socially and morally undesirable character has important consequences for that person's further rule-breaking. Under certain circumstances, being labeled may intensify one's commitment to a deviant identity and contribute to further deviant behavior. Some conventional, law-abiding citizens believe, "once a deviant, always a deviant." Someone who has been stigmatized and labeled "is ushered into the deviant position by a decisive and often dramatic ceremony, yet is retired from it with hardly a word of public notice." As a result, the deviant is given "no proper license to resume a normal life in the community. Nothing has happened to cancel out the stigma imposed upon him" or her. The original judgment "is still in effect." The conforming members of a society tend to be "reluctant to accept the returning deviant on an entirely equal footing" (Erikson 1964, pp. 16, 17).

Deviant labels tend to be "sticky." The community tends to stereotype someone as, above all and most importantly, a deviant. When someone is identified as a deviant, the community asks, "What kind of person would break such an important rule?" The answer that is given is "one who is different from the rest of us, who cannot or will not act as a moral being and therefore might break other important rules" (Becker 1963, p. 34). Deviant labeling is widely regarded as a quality that is "in" the person, an attribute that is carried wherever he or she goes; therefore, it is permanent or at least long lasting. Deviant behavior is said to be caused by an indwelling, essentialistic trait; it is not seen as accidental or trivial, but a fixture of the individual. Once a deviant label has been attached, it is difficult to shake. Ex-convicts find it difficult to find legitimate employment upon their release from prison; once psychiatrists make a diagnosis of mental illness, hardly any amount of contrary evidence can dislodge their faith in it; ex-mental patients are carefully scrutinized for odd, eccentric, or bizarre behavior.

Such stigmatizing tends to deny to deviants "the ordinary means of carrying on the routines of everyday life open to most people. Because of this denial, the deviant must of necessity develop illegitimate routines" (Becker 1963, p. 35). As a consequence, the labeling process may actually increase the deviant's further commitment to deviant behavior. It may limit conventional options and opportunities, strengthen a deviant identity, and maximize participation in a deviant group. Labeling someone, thus, may become "a self-fulfilling prophecy" (Becker 1963, p. 34) in that *someone becomes what he or she is accused of being* – even though that original accusation may have been false (Merton 1948).

Labeling Theory Today

Labeling theory left two legacies to the contemporary study of deviance. The first, its "major" mode, which Plummer refers to as its "broader" version (1979, p. 88, 2011, pp. 84–85), was its constructionist vision. Other, earlier approaches were careful to point out that deviance and crime were a matter of violating rules, norms, and laws, which are socially constructed and vary somewhat historically and culturally. But labeling theory stressed and highlighted this point more forcefully. Indeed, it went further and emphasized that definitions of wrongdoing vary not only from society to society but from one *category* or *social context* to another. This remains a basic and crucial assumption in all sociological work on deviance. The second legacy of labeling theory, its "minor" mode (which, unfortunately, critics stress as its main point) – which Plummer refers to as its "narrow" version (2011, pp. 83–84) – is its argument about the causal mechanism of deviance: Being labeled as a wrongdoing inevitably or usually leads to a strengthening of a deviant identity and hence, an escalation in the seriousness and frequency of deviant behavior. This argument is as often wrong as it is right; its lack of empirical verification should not negate the perspective's "major" mode or constructionist legacy.

The foundational works of labeling theory were published in the 1960s; the number of citations or "hits" in Harzing's "Publish or Perish: Google Scholar" to "labeling theory" increased from 85 per year in the 1965–1969 period to over 825 in 1970–1974, then peaked in the 1990s (5,720 per year in 1990–1994 and 5,660 in 1995–1999), and declined into the 2000s. In the 1960s, the field's younger scholars and researchers yearned for a fresh, unconventional, and radically different way of looking at deviance. At the same time, the labeling perspective was immediately and subsequently widely and vigorously attacked, and many of these criticisms stuck; common wisdom has it that, today, both labeling theory and the sociology of deviance are dead or dying (Sumner 1994; Best 2004). Eventually, researchers generated competing perspectives, and, in reaction to critiques, the perspective's supporters modified or adapted labeling theory's insights. Currently, no single approach or paradigm dominates the study of deviance in the way that labeling theory did circa 1970–1974. What we see now is diversity, fragmentation, and theoretical dissensus. In spite of the criticisms, however, the labeling school left a legacy to the field that even its critics incorporate into their work, albeit, for the most part, implicitly.

By the 2000s, the central insights of labeling theory have become so taken for granted, so densely interwoven into the conventional wisdom of criminology and the sociology of deviance – a kind of "quiet orthodoxy" that appears in "different guises" (Plummer 2001, pp. 193–194) – that its place in these fields offers a case of "obliteration by incorporation" (Merton 1979). In other words, "the central strands of the perspective live on in cognate areas of inquiry" (Grattet 2011a, p. 186). Ongoing research has demonstrated that the consequences of negative labeling tend to be long lasting and often dire. Grattet's summary of this literature is most revealing (2011a, b). Matsueda's study of troublesome boys reveals that parental definitions ("informal social control") often results in self-conceptions that increase the likelihood of further delinquencies (1992). The research of Bruce Link and his associates on mental disorder likewise demonstrates the baleful impact of stigma and deviance labeling. Working with a "modified labeling theory," Link uncovered how mental illness processing agencies reinforce "perceptions of patient dangerousness" and put social distance between themselves and the patient, making their conditions more serious (Link et al. 1989). Sampson and Laub test the hypothesis of "cumulative disadvantage" (Grattet 2011b, p. 124) – that criminal justice sanctioning commonly results in offenders repeating and increasing the seriousness of their involvement in offending over the life course. Sampson and Laub's contention is that there is only "one theoretical position in criminology that is inherently developmental in nature – labeling theory" (1997, p. 3). Cumulative disadvantage represents a kind of "snowballing effect" which increasingly "mortgages" the offender's future, especially when negative evaluations in the realms of school and employment further reduce their life chances. For instance, convicted felons face increasingly difficult conditions for reintegrating into civil society, disenfranchising them and making the choice of further criminal activity increasingly attractive; negative labeling by work settings, marriage, and family dynamics all make desistance from crime and wrongdoing increasingly difficult (Petersilia 2003; Uggen and Manza 2006; Western 2006). A criminal record has a powerful chilling effect on employment outcomes. Pager introduces the concept of "negative credentials" to stress this process; these are the "official markers that restrict access and opportunity rather than enabling them" (2007, p. 32). As Grattet argues (2011a, b), recent research powerfully argues for the ongoing influence of the labeling/interactionist tradition in the study of crime and deviance.

Related Entries

- ▶ History of Criminological Theories: Causes of Crime
- ▶ Labeling and Deviance
- ▶ Social Control Theory of Sexual Homicide Offending

Recommended Reading and References

Becker HS (1963) Outsiders: studies in the sociology of deviance. Free Press, New York

Becker HS (ed) (1964) The other side: perspectives on deviance. Free Press, New York

Becker HS (1973) Labeling theory reconsidered. In: Becker HS (ed) Outsiders: studies in the sociology of deviance, (2nd ed). Free Press, New York, pp 177–212

Best J (2004) Deviance: career of a concept. Wadsworth, Belmont, CA

Blumer H (1969) Symbolic interactionism. Prentice-Hall, Englewood Cliffs, NJ

Debro J (1970) Dialogue with Howard S. Becker. Issues Criminol 5(Summer):159–179

Erikson KT (1964) Notes on the sociology of deviance. In: Becker HS (ed) The other side: perspectives on deviance. Free Press, New York, pp 307–314

Grattet R (2011a) Societal reactions to deviance. Annual Rev Sociol 37:185–204
Grattet R (2011b) Labeling theory. In: Bryant CD (ed) The Routledge handbook of deviant behavior. Routledge, London/New York, pp 121–128
Kitsuse JI (1972) Deviance, deviant behavior, and deviants: some conceptual problems. In: Filstead WJ (ed) An introduction to deviance: readings in the process of making deviants. Markham, Chicago, pp 233–243
Lemert EM (1951) Social pathology: a systematic approach to the theory of sociopathic behavior. McGraw-Hill, New York
Lemert EM (1972) Human deviance, social problems, and social control (2/e). Prentice-Hall, Englewood Cliffs, NJ
Link BG, Cullen FT, Struening E, Shrout P, Dohrenwend BP (1989) A modified labeling theory approach to mental disorders: an empirical assessment. Am Sociol Rev 54(June):400–423
Matsueda RL (1992) Reflected appraisals, parental labeling, and delinquency: specifying a symbolic interactionist theory. Am J Sociol 97(May):1577–1611
Merton RK (1948) The self-fulfilling prophecy. Antioch Rev 7(Summer):193–210
Merton RK (1979) Foreword. In: Eugene Garfield (ed) Citation indexing: its theory and application in science, technology and humanities. Wiley, New York, pp. ix–x
Pager D (2007) Marked: race, crime, and finding work in an era of mass incarceration. University of Chicago Press, Chicago
Petersilia J (2003) When Prisoners Come Home: Parole and Prisoner Reentry. Oxford University Press, New York
Plummer K (1979) Misunderstanding labelling perspectives. In: Downes D, Rock P (eds) Deviant interpretations. Martin Robinson, Oxford, UK, pp 85–121
Plummer K (2001) Labeling theory. In: Bryant CD (ed) Encyclopedia of criminology and deviant behavior, vol I. Brunner-Routledge/Taylor & Francis, New York, pp 191–194
Plummer K (2011) Labelling theory revisited: forty years on. In: Peters H, Dellwing M (eds) Langweiliges Verbrechen. VS Verlag, Wiesbaden, Germany, pp 83–103
Sampson RJ, Laub JH (1997) A life-course theory of cumulative disadvantage and the stability of delinquency. In: Thornberry TP (ed) Developmental theories of crime and delinquency. Transaction, New Brunswick, NJ, pp 133–161
Sumner C (1994) The sociology of deviance: an obituary. Open University Press, London
Tannenbaum F (1938) Crime and the community. Ginn, New York
Uggen C, Manza J (2006) Locked out: Felon disenfranchisement and American democracy. Oxford University Press, New York
Western B (2006) Punishment and inequality in America. Russell Sage Foundation Press, New York

Labeling: History and Concept

► Cultural Criminology

Labelling Perspective

► Labelling, Deviance, and Media

Labelling Theory

► Labelling, Deviance, and Media

Labelling, Deviance, and Media

Chris Greer[1] and Robert Reiner[2]
[1]Department of Sociology, City University London, London, UK
[2]Department of Law, London School of Economics and Political Science, London, UK

Synonyms

Labelling perspective; Labelling theory

Overview

The labelling perspective emerged as a distinctive approach to criminology during the 1960s and was a major seedbed of the radical and critical perspectives that became prominent in the 1970s. It represented the high point of an epistemological shift within the social sciences away from positivism – which had dominated criminological enquiry since the late 1800s – and toward an altogether more relativistic stance on the categories and concepts of crime and control. It inspired a huge amount of work throughout the 1960s and 1970s and still resonates powerfully today. This short entry maps out some of the ways in which labelling, deviance, media, and justice interact at the levels of definition and process. It presents an overview and analysis of key mediatized labelling processes, such as the highly influential concept of moral panics. It discusses how the interconnections between labelling, crime, and criminal justice are changing in a context of technological development, cultural change, and media proliferation. The conclusion offers an assessment and evaluation of labelling theory's long-term impact on criminology.

The Roots of the Labelling Perspective

The labelling perspective emerged at a time of radical intellectual change in the 1960s. The intellectual problem, as labelling theorists saw it, was that the study of crime had narrowed into two key questions: (1) Why do they do it and (2) how do we stop them from doing it? Government bodies and funding agencies reinforced the notion that "they" were different from "us," and that "crime" was entirely distinct from "criminal justice." Such thinking had laid foundations for the resurgence of a separate academic discipline of "positivist criminology," which functioned as policy science of crime, the criminal and crime control. There was an assumed consensus over what constituted crime, and the operations of criminal justice were seen as of interest only in terms of making them more effective in controlling crime. Yet nobody appeared to be asking why some behaviors were deemed criminal in certain contexts, while others were not. And why some people were deemed deviant and in need of correction or punishment, while others – who engaged in similar behaviors – were not.

Though Durkheim had discussed the problematic definition of crime in the late nineteenth century (Durkheim 1895/1964, pp. 69–72), and scholars such as Frank Tannenbaum, George Herbert Mead, and W. I. Thomas had been writing more or less directly about labelling processes since the 1930s, it is Howard Becker's (1963) *Outsiders: Studies in the Sociology of Deviance* that has become the best known and most influential statement of labelling theory. Building on the theoretical foundations of social constructionism and symbolic interactionism, Becker sought to problematize systematically precisely those questions that mainstream "correctionalist" criminology left unexplored. He argued (1963, p. 14), "The same behaviour may be an infraction of the rules at one time and not at another; may be an infraction when committed by one person, but not when committed by another; some rules are broken with impunity, others are not. In short, whether a given activity is deviant or not depends in part on the nature of the act (that is whether or not it violates some rule) and in part on what other people do about it."

Understanding the highly selective nature of labelling, the equally selective nature of the social reaction to which it may or may not give rise, and the consequences for those who are labelled required analyzing complex and contested power relations as both micro-interactional and macro-social processes. For Edwin Schur (1979, p. 160, italics in original), the labelling perspective is thus concerned with both "*definition* and *process* at all the levels that are involved in the production of deviant situations and outcomes. Thus, the perspective is concerned not only with what happens to specific individuals when they are branded as deviant ('labeling' in the narrow sense) but also with the wider domains and processes of social definitions and collective rule-making that frequently lie behind such concrete applications of negative labels."

Key questions for labelling theorists therefore include the following: How are labels created or socially constructed? How are labels imposed? How and why do particular behaviors become defined as "normal" or "deviant"? What enables labellers to impose their particular definitions upon behaviors, actions, and situations? How does the labelling process work and with what consequences?

Key Issues/Controversies

Labelling, Media, and Crime

The mass media play at least a subordinate role in all the major theoretical perspectives attempting to understand crime and criminal justice. To illustrate this, the predominant theories of crime can be assembled in a simple model. For a crime to occur, there are five logically necessary preconditions, which can be identified as labelling, motive, means, opportunity, and the absence of controls (Reiner 2007, pp. 80–90). The media potentially play a part in each of these elements and thus can affect levels of crime in a variety of ways (Greer and Reiner 2012, pp. 256–61).

Labelling

For an act to be "criminal" (as distinct from harmful, immoral, antisocial, etc.), it has to be labelled as such. This involves the creation of a legal category. It also requires the perception of the act as criminal by citizens and/or law enforcement officers if it is to be recorded as a crime. The media are an important factor in both processes, helping to shape the conceptual boundaries and recorded volume of crime.

The role of the media in helping to develop new (and erode old) categories of crime has been emphasized in most of the classic studies of shifting boundaries of criminal law within the "labelling" tradition. Becker's (1963) seminal book *Outsiders* analyzed the emergence of the Marijuana Tax Act in the USA in 1937, emphasizing the use of the media as a tool of the Federal Bureau of Narcotics and its moral entrepreneurship in creating the new statute. Since this pioneering work many studies have illustrated the crucial role of the media in shaping the boundaries of deviance and criminality, by creating new categories of offense or changing perceptions and sensitivities, leading to fluctuations in apparent crime (Young 1971; Cohen 1972; Hall et al. 1978). For example, Roger Graef's celebrated 1982 fly-on-the-wall documentary about the Thames Valley Police was a key impetus to reform of police treatment of rape victims (Greer and Reiner 2012, p. 256). This also contributed, however, to a rise in the proportion of victims reporting rape and thus an increase in the recorded rate. Many other studies document media-amplified "crime waves" and "moral panics" about law and order (Goode and Ben-Yehuda 2009).

What all these studies illustrate is the significant contribution of the media to determining the apparent level of crime. Increases and (perhaps more rarely) decreases in recorded crime levels are often due in part to the deviance construction and amplifying activities of the media.

Motive

A crime will not occur unless there is someone who is tempted, driven, or otherwise motivated to carry out the "labelled" act. The media feature in many of the most commonly offered social and psychological theories of the formation of criminal dispositions. Probably the most influential sociological theory of how criminal motives are formed is Merton's version of *anomie* theory (Merton 1938/1957), echoes of which are found in more recent work (see Special Issue of *Theoretical* Criminology 11/1 2007; Reiner 2007, pp. 14–5, 84–5). The media play a key role in these accounts of the formation of anomic strain generating pressures to offend. The media are pivotal in presenting for universal emulation images of affluent lifestyles, which accentuate relative deprivation and generate pressures to acquire ever higher levels of material success regardless of the legitimacy of the means used.

Psychological theories of the formation of motives to commit offenses also often feature media effects as part of the process (Greer and Reiner 2012, pp. 247–62). It has been claimed that the images of crime and violence presented by the media are a form of social learning and may encourage crime by imitation or arousal effects. Others have argued that the media tend to erode internalized controls by disinhibition or desensitization through witnessing repeated representations of deviance ().

Means

It has often been alleged that the media act as an open university of crime, spreading knowledge of criminal techniques. This is often claimed in relation to particular *causes célèbres* or horrific crimes, for example, during the 1950s campaign against crime and horror comics. A notorious case was the allegation that the murderers of Jamie Bulger had been influenced by the video *Child's Play 3* in the manner in which they killed the unfortunate toddler. A related line of argument is the "copycat" theory of crime and rioting. Despite a plethora of research and discussion, the evidence that this is a major source of crime remains weak.

Opportunity

The media may increase opportunities to commit offenses by contributing to the development of a consumerist ethos, in which the availability of tempting targets of theft proliferates. The domestic hardware and software of mass media use – TVs, videos, radios, CDs, personal computers, and mobile phones – are the common currency of routine property crime, and their proliferation has been an important aspect of the spread of criminal opportunities.

Absence of Controls

Motivated potential offenders, with the means and opportunities to commit offenses, may still not carry out these crimes if effective social controls are in place. These might be *external* – the deterrent threat of sanctions represented in the first place by media made criminality the police – or *internal*, the still, small voice of conscience, what Eysenck has called the "inner policeman."

A regularly recurring theme of respectable anxieties about the criminogenic consequences of media images of crime is that they erode the efficacy of both external and internal controls. They may undermine external controls by derogatory representations of criminal justice, for example, ridiculing its agents, a key complaint at least since the days of Dogbery, resuscitated in this century by the popularity of comic images of the police, from the Keystone Cops onward. Serious representations of criminal justice might undermine its legitimacy by becoming more critical questioning, for example, the integrity and fairness or the efficiency and effectiveness of the police. Negative representations of criminal justice could lessen public cooperation with the system, or potential offenders' perception of the probability of sanctions, with the consequence of increasing crime.

Probably the most frequently suggested line of causation between media representations and criminal behavior is the allegation that the media undermine internalized controls, by regularly presenting sympathetic or glamorous images of offending. In academic form this is found in the psychological theories about disinhibition and desensitization, which were referred to in the section above on the formation of motives. In sum, there are several possible links between media representations of crime and criminal behavior which are theoretically possible and frequently suggested in criminological literature and political debate. In the next section I will review some of the research evidence examining whether such a link can be demonstrated empirically.

Labelling, Media, and Moral Panics

The successful labelling of a particular situation or set of conditions as deviant and in need of amelioration can, in the extreme, result in "moral panic." The term was first used by Young (1971) in his study of subcultures and drugtaking. Cohen (1972) developed and extended the concept in his analysis of the sensationalistic, heavy-handed, and ultimately "disproportionate" reaction to the mods and rockers disturbances in an English seaside resort in 1964. Though the damage was in financial terms minor, Cohen traces the spiralling social reaction through initial intolerance, media stereotyping, moral outrage, increased surveillance, labelling and marginalization, and deviancy amplification leading to further disturbances that seemed to justify the initial concerns. The flamboyant misbehavior of youth subcultures, independent and sexually and economically liberated, affronted the postwar values of hard work, sobriety, and deferred gratification. At a time of rapid social change, they were a visible index of a world that was slipping away – "folk devils" who provided a crystallizing focus for social anxiety and "respectable fears." Cohen used the building blocks of labelling theory – social constructionism, symbolic interactionism, deviancy amplification, and social psychology – but also incorporated the lesser known academic literature on "disaster research" to describe the various phases of a moral panic, warning, impact, inventory, and reaction and chart its progression.

Hall et al. (1978) politicized the concept by locating it within a broader political economy perspective in their analysis of a "mugging" moral panic which, they argued, was constructed to address an escalating crisis in state hegemony. Drawing from an eclectic mix of influences, their account connects "new deviancy theory, news media studies and research on urban race relations with political economy, state theory and notions of ideological consent" (McLaughlin 2008, p. 146). For some critical criminologists, it represents the high point of Marxist theorizing about crime, law and order, and the state. While fully acknowledging the sophistication of this work, Cohen (2011) has nonetheless noted a wider tendency to over-politicize the concept at the expense of its sociological meaning and application. Hall (2007) has suggested in response that politicization was a necessary developmental stage and that the full explanatory potential of the moral panic concept was, in fact, only realized through its construction as ideology.

Goode and Ben-Yehuda (2009) developed Cohen's discussion of moral panic by paying particular attention to the criteria that should be in place before it can be suggested that a "moral panic" is occurring. They identify five key features of the phenomenon: (a) *concern* (a reported condition or event generates anxiety), (b) *hostility* (the condition or event is condemned and, where there are clearly identifiable individuals who can be blamed, these are labelled as "folk devils"), (c) *consensus* (the negative social reaction is widespread and collective), (d) *disproportionality* (the extent of the problem and the threat it poses are exaggerated), and (e) *volatility* (media attention and the associated panic emerge suddenly and with intensity, but can dissipate quickly too). Media are central to all of these.

"Moral panic" is one of the most widely used terms in the sociological analysis of crime and justice and has transcended academic discourses to become commonplace in political rhetoric and popular conversation (Altheide 2009). Given its prolific usage, it is surprising that few commentators have subjected the concept to sustained and rigorous critical investigation. With the split in the criminological left in the late 1970s, the concept was dismissed by left realists as "left idealism" and accused of obfuscating the painful "realities" of criminal victimization by propagating the view that "the crime problem" is socially constructed (Young 1979). In exploring the anatomy of the concept, critics have queried the notions of "disproportionality" and "volatility": first, since this assumes a superior knowledge of the objective reality of the issue against which the reaction is measured and a corresponding assumption of what a "proportionate" reaction would look like. Second, because in a contemporary multimedia world characterized by ontological insecurity and state of a permanent free-floating anxieties, the notion of discreet, self-contained, and volatile moral panics may need some rethinking (McRobbie and Thornton 1995). Cohen has responded to all of these criticisms. But such critical interventions, both from within and outside of criminology, have barely interrupted the general tendency to arbitrarily apply the concept to explain everything from global warning to "swine flu." The broadly uncritical application of the moral panic concept has led Garland (2008) to reassert two elements of the original analysis, which are absent from many contemporary studies: (a) the *moral dimension* of the social reaction – most issues can be moralized, but many are not in and of themselves "moral" and cannot automatically be analyzed as such – and (b) the idea that the deviant conduct in question is somehow *symptomatic* of a wider problem, a threat to established values, or a particular way of life. Struggles over the power to label and to label effectively via media discourses, of course, remain fundamental to the moralization of particularly "social problems," the identification of folk devils, the persuasive representation of threats to particular forms of social existence, and the prescription of ameliorative action.

Labelling and Trial by Media

Another way in which the news media are directly involved in labelling is the phenomenon of "trial by media" (Greer and McLaughlin 2011, 2012a). "Trial by media" can be defined as a dynamic, impact-driven, news media-led process by which individuals – who may or may not be publicly known – are tried and sentenced in the "court of public opinion." The targets and processes of "trial by media" can be diverse and may range from prejudging the outcome of formal criminal proceedings against "unknowns" to the relentless pursuit of high-profile celebrity personalities and public figures deemed to have offended in some way against an assumed common morality.

Despite their clear diversity, such "trials" share certain core characteristics. In each case, the news media behave as a proxy for "public opinion" and seek to exercise parallel functions of "justice" to fulfill a role perceived to lie beyond the interests or capabilities of formal institutional authority. Due process and journalistic objectivity can give way to sensationalist, moralizing speculation about the actions and motives of those who stand accused in the news media spotlight. Judicial scrutiny of "hard evidence" yields ground to "real-time" dissemination of disclosures from paid informants and hearsay and conjecture from "well-placed sources." Since the news media substitute for the prosecution, judge, and jury, the target may find themselves rendered defenseless. The default position is "guilty until proven innocent." Those found "guilty" will be subjected to righteous "naming and shaming" followed by carnivalesque condemnation and ridicule. The public appeal of "trial by media" is evidenced by increased circulation and web traffic. And by no means is it restricted to the British press.

Thus, in recent years police commissioners, senior politicians, banking executives, and, in the UK, the entire political establishment, as well as countless members of the public who are suspected of, but not yet charged with, any range of alleged criminal activities, have been subjected to mediatized scandal and trial by media (Greer and Mclaughlin 2010, 2011, 2012a, 2012b). The results of such high-profile labelling and public shaming, depending on the target, can range from deep and lasting reputational damage, public apologies, high-level resignations, radical political reform, or criminal proceedings.

Labelling, Media, and Criminal Justice

Powerful organizations and institutions tend to hold a distinct advantage in defining the nature of reality as represented via news media. Despite considerable variation in theoretical and methodological approach, decades of research has confirmed that the institutionally powerful enjoys privileged positions as "primary definers" at the top of the "hierarchy of credibility" and that a pro-establishment perspective is structurally and culturally advantaged, if not necessarily guaranteed in news media representations (Ericson et al. 1991; Hall et al. 1978). Historically, then, the power to label has tended to rest more or less firmly in the hands of those who command institutional authority. There is good evidence to suggest, however, that with the proliferation and diversification of media in recent decades, the power of institutional authority to effectively "define how things are" and set the terms of public debate is becoming increasingly contested and unstable.

In a digital multimedia age, a proliferation of news platforms, sites, and formats has been paralleled by a rapidly expanding array of news sources and producers of content, leading to the creation of an unprecedented amount of potentially newsworthy information, and a remarkable number of "news spaces" in which to broadcast/publish it. In the process, increasingly sophisticated, interactive news audiences are reconstituted as consumers – once content to be told what the news is, now increasingly interested in being part of the production process. Armed with cellphones, Blackberries or iPhones, all citizens are potential "citizen journalists." A photo can be taken on a mobile phone, tweeted on Twitter.com, picked up by other users, and disseminated like a virus online. Internet monitoring by mainstream news media outlets means that dramatic amateur photographic, audio or video content can become headline news. Citizen journalism has been instrumental not only in providing newsworthy images but also in defining the news itself – in shaping representations of key global events. From the police brutality against Rodney King filmed by camcorder in Los Angeles in 1991, to the 7/7 London bombings of 2005, to Hurricane Katrina, street protests in Tehran, and the Haiti earthquake, many of the defining images that now form a key part of the "official record" of events were taken with handheld recording devices and posted on social media sites. The emergence of the citizen journalist carries significant implications for official institutions that would seek to control the representing of crime and justice in the news. This phenomenon has been seen as a significant modification of existing power relations, offering what has been called "synopticon" (Mathiesen 1997), providing the mass of the population with some potential to record elite deviance.

The police can no longer simply "deny out of existence" incidents of police violence in public protest situations, since these are even more likely to be captured on camera and broadcast to the world (Greer and Mclaughlin 2010, 2012b). The same can be said of governments that would engage in larger scale abuses of their people and seek to conceal this from international scrutiny. And politicians or public officials, who may previously have fiddled expenses, taken bribes, engaged in illicit affairs when they should have been attending to the affairs of state, have all become more "visible" and, thus, more vulnerable to public exposure, labelling, trial by media, and mediatized justice.

The democratization of public involvement with the news production process, and the "new visibility" (Thompson 2005) to which institutional and state authority are continually subjected, is altering the dynamics of "communication power" that shape our constructed realities. Of course, citizen journalists are neither automatically nor naturally imbued with cultural or official authority: They are not "authorized knowers," who command access to mainstream news media "as of right." Their position in the "hierarchy of credibility" is entirely precarious and contingent. Media access is not granted because of who citizen journalists are, but rather because of where they are and what they have. Their credibility and authenticity as news sources derives from their capacity to provide "factual" visual evidence of "live events" which, in a multiplatform news media market, constitutes an important and cost-effective resource for "making news." Nevertheless, citizens are becoming increasingly involved in the processes of public labelling and social construction that determine what, and who, is defined as honest or corrupt, competent or incompetent, legitimate and illegitimate, or compliant or deviant. As such, they are centrally involved in the reconfiguration of notions of "crime" and "justice" in multimedia worlds.

Labelling Theory: Evaluation and Critique

The labelling perspective has transformed criminological theory and practice since the 1960s. It has made many permanently valuable contributions, above all the recognition of criminal law and justice as problematic research areas, that shape at least as much as they control crime. Criminology conferences and textbooks today devote as much attention to research on and analysis of criminal justice, from a non-correctionalist standpoint, as they do to the study of offending, a legacy (albeit often unrecognized) of labelling theory. The two subfields that the authors of this entry have spent most of their careers researching (policing and media representations of crime/criminal justice) were almost entirely absent from criminologists' agendas until the 1960s, and the questions raised then by labelling theorists. The problematic character of crime statistics, now universally recognized, is another contribution of labelling theorists. These impacts reflect the labelling perspective, but its influence is largely unacknowledged, and the developments have come to be taken for granted and domesticated within mainstream criminology.

What is more questionable is the imperialistic version of labelling theory that was trumpeted in its heyday and made large claims about itself as offering a total theory of crime. This grandiose version of labelling theory originated and flourished as the criminology of the 1960s counterculture and could only be plausible as a general theory in that context.

The imperialistic version claimed that concepts of crime were entirely relative and dependent on perceptions and labelling. It further suggested that labelling and social reaction were the principal explanations of crime and deviance. These claims are epitomized by two frequently cited quotes from key architects of the perspective. The relativity assertion is captured by Becker's statement that "deviance is not a quality of the act. . . but of the application. . . of rules and sanctions" (Becker 1963). The explanatory power of labelling is asserted most explicitly by Lemert: "Older sociology tended to rest heavily upon the idea that deviance leads to social control. . . (T)he reverse idea i.e. that social control leads to deviance, is equally tenable and the potentially richer premise for studying deviance in modern society" (Lemert 1967).

Both claims have some validity, but the exaggerated imperialistic versions, postulated by Becker, Lemert, and others, were neither new nor true without considerable qualification. Criminology before labelling theory (and indeed even nowadays) often took the concept of crime for granted. But its problematic character had already been emphasized by Durkheim and others. Seeing the making and enforcement of criminal law as a part of criminology was indeed acknowledged by some criminologists long before the labelling revolution. Moreover, it was assumed by criminal lawyers, both in textbooks and judicial decisions (*Proprietary Articles Trade Assn*. v. *Alt. Gen. for Canada* [1931] AC at 32, per Lord Atkin). Legal scholars had studied the emergence and change of criminal laws long before the advent of labelling theory (e.g., Hall 1935/1952). Recognizing the historical and social diversity of what precisely is criminalized at different times and places (Reiner 2007, Chap. 2; Lacey and Zedner 2012) does not entail complete relativity. As Hart suggested persuasively, there seems to be a "minimum content of natural law," activities that are regulated in all societies because they are conditions of viable social existence, even though the precise content and manner of proscription and sanctioning is variable (Hart 1961, Chapter IX, Part II).

The recognition of labelling as a cause of crime was also not entirely new and had been anticipated even by some criminologists in the positivist tradition (most explicitly Wilkins 1964, whose concept of deviance amplification in turn influenced labelling theorists). While it is the case, as Lemert claims, that often "social control leads to deviance," it is disputable whether it is the "richer premise for studying deviance." Lemert's claim rests on the assumption that "secondary deviance," which follows labelling, is more pervasive and problematic than "primary deviance," which precedes it. But this is an empirical question that is likely to vary in different times and places, and with regard to different kinds of deviance and social reaction, not a "premise."

Any plausibility the imperialistic claims of labelling theory had derived from the limited nature in practice of their empirical research. These tended to concentrate on marginal or exotic forms of deviance, which lend themselves to being seen as harmful or problematic not intrinsically but primarily if not solely because of labelling: marijuana use, the bohemian subculture of jazz musicians (Becker 1963); "hustlers, beats and others" (Polsky 1967); and "crimes without victims" (Schur 1965). An early critique castigated this pithily as the "sociology of nuts, sluts and "preverts" (sic)" (Liazos 1972).

The labelling theory pioneers' focus on the dramatic and colorful made it much easier to ignore the harms done by some primary deviance. They concentrated on the creation of crime by the labelling activities of low-level control agents, reversing the moral assessments of criminal law and justice – as explicitly advocated by Becker in his call for criminologists to ask "Whose Side Are We On?" (1967). This not only neglected the harms done by some crime but bracketed out its structural causes and the structural determinants of control activity – law, culture, political economy, wider social patterns, and institutions (as Gouldner argued in his 1968 repost to Becker "The Sociologist As Partisan"). This critique stimulated the morphing of labelling theory into more politically radical forms of "new criminology" and "deviance theory" in the 1970s (the core classics were Cohen 1971; Taylor et al. 1973; as well as the seminal studies discussed extensively in this paper, Young 1971; Cohen 1972; Hall et al. 1978).

Labelling theory has had a huge impact, fundamentally shifting the criminological paradigm away from a taken for granted correctionalist stance and stimulating a variety of forms of critical perspective. Much of its influence is now hidden, domesticated in the proliferating analyses of policing, media, and criminal justice. Although the sweeping claims of its originators are hard to sustain, its legacy lives on explicitly in contemporary cultural criminology and other qualitative and critical approaches.

Related Entries

- ▶ Labeling Theory
- ▶ New Media and Crime Images
- ▶ Use of Social Media in Policing

Recommended Reading and References

Altheide D (2009) Moral panic: from sociological concept to public discourse. Crime Media Cult Int J 5(1):79–99

Becker HS (1963) Outsiders: studies in the sociology of deviance. Free Press of Glencoe, London

Becker H (1967) Whose side are we on? Soc Probl 14(1):239–247

Cohen S (ed) (1971) Images of deviance. Penguin, London

Cohen S (1972) Folk devils and moral panics: the creation of the Mods and Rockers. MacGibbon and Kee, London

Cohen S (2011) Whose side were we on? The undeclared politics of moral panic theory. Crime Media Cult Int J 7(3):28–38

Durkheim E (1895/1964) The rules of sociological method. Free Press, Glencoe

Ericson RV, Baranek PM, Chan JBL (1991) Representing order: crime, law, and justice in the news media. University of Toronto Press, Toronto

Garland D (2008) On the concept of moral panic. Crime Media Cult Int J 4(1):9–30

Goode E, Ben-Yehuda N (2009) Moral panics: the social construction of deviance, 2nd edn. Wiley-Blackwell, Chichester/Malden

Gouldner A (1968) The sociologist as partisan: sociology and the welfare state. Am Sociol 3(2):103–116

Greer C, Mclaughlin E (2010) We predict a riot? public order policing, new media environments and the rise of the citizen journalist. Br J Criminol 50(6):1041–1059

Greer C, Mclaughlin E (2011) Trial by media: policing, the news 24–7 news mediasphere, and the politics of outrage. Theor Criminol 15(1):23–46

Greer C, Mclaughlin E (2012a) Media justice: Madeleine McCann, intermediatisation and' trial by media', in the British Press. Theor Criminol 16(4):395–416

Greer C, Mclaughlin E (2012b) 'This is not justice': Ian Tomlinson, Institutional Failure and the Press politics of outrage. Br J Criminol 52(2):274–293

Greer C, Reiner R (2012) Mediated mayhem: media, crime, criminal justice. In: Maguire M, Morgan R, Reiner R (eds) The Oxford handbook of criminology, 5th edn. Oxford University Press, Oxford

Hall J (1935/1952) Theft, law and society. Bobbs Merrill, Indianopolis

Hall S (2007) 'Moral panics: then and now' *moral panics: then and now*. Presentation at the British Academy, London, 9th March 2007 – available at: http://www.britac.ac.uk/audio.cfm/assetfileid/9513

Hall S, Critcher C, Jefferson T, Clarke J, Roberts B (1978) Policing the crisis: mugging, the state, and law and order. Macmillan, London

Hart H (1961) The concept of law. Oxford University Press, Oxford

Lacey N, Zedner L (2012) Legal constructions of crime. In: Maguire M, Morgan R, Reiner R (eds) The Oxford handbook of criminology, 5th edn. Oxford University Press, Oxford

Lemert E (1967) Human deviance. Prentice Hall, Englewood Cliffs

Liazos A (1972) The poverty of the sociology of deviance: nuts, sluts, and preverts. Soc Probl 20(1):103–120

Mathiesen T (1997) The viewer society: Michel Foucault's "panopticon" revisited. Theor Criminol ½:215–234

McLaughlin E (2008) Hitting the panic button: policing/ 'mugging'/media/crisis. Crime Media Cult Int J 4(1):145–154

McRobbie A, Thornton S (1995) Rethinking "moral panic" for multi-mediated social worlds. Br J Sociol 46(4):559–574

Merton R (1938/1957) Social structure and anomie. Am Sociol Rev 3:672–682

Polsky N (1967) Hustlers, beats and others. Penguin, London

Reiner R (2007) Law and order: an honest citizen's guide to crime and control. Polity, Cambridge

Schur E (1965) Crimes without victims. Prentice Hall, Englewood Cliffs

Schur E (1979) Interpreting deviance: a sociological introduction. Harper and Row, New York

Taylor I, Walton P, Young J (1973) The new criminology. Routledge, London

Thompson JB (2005) The new visibility. Theory Cult Soc 22(6):31–51

Young J (1971) The drugtakers: the social meaning of drug use. Paladin, London

Young J (1979) Left idealism, reformism and beyond: from the new criminology to Marxism. In: Bob F, Kinsey R, Lea J, Picciotto S, Young J (eds) Capitalism and the rule of law: from deviancy to Marxism. Hutchinson, London

Latent Class Growth Modeling

▶ Group-Based Trajectory Models

Latent Growth Curves

▶ Longitudinal Crime Trends at Places

Latent Trajectory Modeling

▶ Group-Based Trajectory Models

Law Enforcement

▶ Drug Enforcement

Law of Community Policing and Public Order Policing

Tracey L. Meares
Yale Law School, Yale University, New Haven, CT, USA

Overview

It is likely impossible today to find a jurisdiction in the United States wherein a policing agency does not claim to practice community policing. Certainly the term is ubiquitous among law enforcement practitioners and scholars. The discussion around the practice for the last decade or so has not been whether to implement the approach but, rather, how exactly to implement it and the extent to which the practice is effective.

Less discussed is the relationship between law and community policing. By its very nature, community policing is focused upon patrol officers engaging in proactive conduct through exercise of discretion. Police discretion is, of course, framed and constrained by law – very typically constitutional law. However, unlike the constitutional rules that structure searches, seizures, and interrogations, the law pertaining to community policing and public order maintenance often is much more fluid. This entry will sketch out the fundamentals of that body of law as well as document current debates.

Fundamentals

There is a great deal of variation among the definitions of community and public order policing offered by scholars, but most agree that police engagement, collaboration, and partnership with private citizens are central features. An additional key feature of community policing emphasizes willingness on the part of individual police officers to proactively engage citizens to work on neighborhood problems and to promote community safety. While internal agency policy (essentially administrative rules) may regulate street-level policing in communities and neighborhoods, very few statutory rules govern these kinds of police practices. Constitutional law also seems mostly inapposite. Of course, the Fourth Amendment to the United States Constitution protects citizens from unreasonable searches and seizures, and the Fifth Amendment regulates police interrogation of citizens. Together with the Sixth Amendment and the Due Process Clause of the Fourteenth Amendment,courts have developed a muscular, detailed, and complex body of law to govern important police practices. This jurisprudence, however, is focused primarily upon police investigation of crimes. State and federal courts have recognized, though, that a large portion of police responsibilities is not related to criminal investigation (Walker 1992) and that traditional constitutional requirements may be inappropriate for regulating non-investigatory police activity.

Consider police searches as an example. Typically, to satisfy the Fourth Amendment, police must obtain a warrant from a magistrate certifying that they have probable cause to believe that the places, persons, and things to be searched are involved or related to a criminal act. Moreover, it is typically said that warrant requirement is extinguished only for certain categorical exceptions. But what happens when police engage in activities, such as responding to a leak in a private home that may damage property (*United States v. Boyd*, 407F. Supp. 693 (1976)), responding to a serious noise nuisance (*United States v. Rohrig*, 98F.3d 1506, 1522 (6th Cir. 1996)), responding to a missing person's claim (*State v. Brideswell*, 759 P.2d 1054 (Or. 1988); *Commonwealth v. Bates*, 548N.E. 2d 889 (1989)), retrieving a police-issued firearm from a private automobile (*Cady v. Dombrowski*, 413 U.S. 433 (1973)), investigating a potential burglary in progress (*United States v. Johnson*, 9F.3d 506 (6th Cir. 1993)), and assisting injured or endangered persons (*Wright v. State*, 7S.W.3d 148 (Tex. App. Ct. 1999)). In these and related situations, courts have not required police to first obtain a warrant from a judicial magistrate before entering homes and searching people. When discussing these activities, courts often point to the "Community Caretaking doctrine" (CC), in order to grant police relief from some of the traditional Fourth Amendment procedures. Courts typically apply CC where the police invade a traditional area of Fourth Amendment privacy but do so for purposes other than criminal investigation. Given the range of activities to which CC has been applied, there is a very real sense in which it can be considered the constitutional law of community policing.

Federal judge and legal scholar Debra Livingston has articulated two primary justifications for decreased procedural protections in CC cases (Livingston 1998). First, CC intrusions are thought not to impose stigma upon the victim because the police typically are not motivated by criminal suspicion when engaging in the intrusion. Second, many courts have concluded that community caretaking activities are less intrusive than investigative searches. During investigative searches, the police search all locations specified in the warrant in which the evidence may reasonably be found. The search continues until the police locate the relevant criminal evidence or determine it is not present. By contrast, when engaging in intrusions justified by CC, the police usually have no suspicion of criminal activity and thus have no reason to proactively search for criminal evidence.

Specifics of the Community Caretaking Doctrine

The courts are divided on the requirements imposed upon police officers when the CC doctrine applies. There are two primary camps.

One group of courts has interpreted CC as independent of traditional Fourth Amendment jurisprudence. On this view, CC intrusions are not governed by the usual probable cause and warrant requirements. Instead, a warrantless intrusion justified by the CC doctrine is permitted if the officer's actions were reasonable given the totality of the circumstances. For example, in *State v. Pinkard*, 319 Wis. 2d 234 (2010), a private citizen informed the police that two persons were sleeping in a home next to cocaine and other contraband and that the door to the home was open. The police entered the home to "make sure that the occupants . . . were not the victims of any type of crime[,] that they weren't injured . . . and to safeguard any life or property in the residence." Upon entering the home, the police seized drugs and other contraband in plain view. The Supreme Court of Wisconsin held that "the officers were engaged in a bona fide community caretaker function and that the community caretaker function was reasonably exercised under the totality of the circumstances." Similarly, in *Wright v. State*, 7S.W.3d 148, 151 (Tex. App. Ct. 1999), the court held that a "police officer may stop [a car to] assist an individual whom a reasonable person – given the totality of the circumstances – would believe is in need of help." This post hoc reasonableness test contrasts sharply with the traditional categorical approach under the Fourth Amendment. It is not, however, a complete anomaly. The Supreme Court has, for example, applied reasonableness tests to a range of police functions that fall under the special needs exception, such as searches of student property (*New Jersey v. T.L.O.*, 469 U.S. 325 (1984)) and highway roadblocks (*Delaware v. Prouse*, 440 U.S. 648 (1979)).

A second group of courts, rather than interpreting CC as an independent doctrine with its own set of rules, have applied CC as an extension of the traditional exigent circumstances exception to the warrant requirement. The exigency exception extinguishes the warrant requirement when one or more categorically defined exigencies are satisfied (i.e., hot pursuit, destruction of evidence); and it is not practical for police to obtain a warrant. Importantly, under this line of reasoning, the police must still demonstrate probable cause. They are relieved only from the requirement of obtaining a warrant. Thus, some courts have held that the CC doctrine does not permit warrantless Fourth Amendment intrusions that lack probable cause (or reasonable suspicion when appropriate). Nor do such courts permit warrantless intrusions when police had sufficient time to obtain a warrant but did not. Courts have applied this interpretation of the CC doctrine in a number of circumstances including missing person cases (*State v. Brideswell*, 759 P.2d 1054 (Or. 1988); *Commonwealth v. Bates*, 548N.E. 2d 889 (Mass. 1990)) and responses to home burglary calls (*United States v. Erickson*, 991F.2d 529 (9th Cir. 1993)).

Judge Livingston has argued that the first interpretation of the CC doctrine is superior. She writes, CC "does not fit within the central assumptions [of traditional Fourth Amendment theory]. The 'reasonableness theory' … can better and more sensitively accommodate those cases in which police officers have intruded on private places principally to serve legitimate community caretaking ends. Though the Court's 'special needs' cases have been subject to legitimate criticism, these cases in fact respond to the plausible intuition that some intrusions on privacy implicate a different set of social practices than traditional law enforcement and are sufficiently unlike law enforcement intrusions so as to justify a distinct Fourth Amendment approach. This same intuition, however, applies to police intrusions to protect life and property or to serve other important community caretaking purposes" (Livingston 1998 at 261, 265).

In addition to these two primary interpretations of the CC doctrine, some courts have conceptualized CC in terms of a third doctrine, the "emergency doctrine." Recently, the Supreme Court defined the standard for the emergency doctrine as follows: "police may enter a home without a warrant when they have an objectively reasonable basis for believing that an occupant is seriously injured or imminently threatened with such injury"(*Brigham City v. Stuart*, 547 U.S. 398, 403 (2006)). John Decker discusses CC in relation to the emergency doctrine extensively (Decker 1999).

When Does the CC Doctrine Apply?

There are three key legal questions for courts to answer when determining whether the CC doctrine applies in a particular case. First, what kind of public interests and police objectives lay within the scope of CC? Second, what happens when CC objectives overlap with clear criminal investigative objectives? Third, does CC apply to private residences?

The Courts have provided some guidance on the first question and, by implication, the second. The United States Supreme Court in *Cady v. Dombrowski*, 413 U.S. 433, 441 (1973), concluded that the CC doctrine applies, at the very least, to cases in which the police objective was "totally divorced from the detection, investigation, or acquisition of evidence relating to the violation of a criminal statute." Following this logic, lower courts have applied CC to a range of police activities, such as responding to a leak in a private home that may damage personal property (*United States v. Boyd*, 407F.Supp. 693 (1976)), responding to a serious noise nuisance (*United States v. Rohrig*, 98F.3d 1506, 1522 (6th Cir. 1996)), responding to a missing person's claim (*State v. Brideswell*, 759 P2d 1054 (Or. 1988)), retrieving a police-issued firearm from a private automobile (*Cady v. Dombrowski*, 413 U.S. 433 (1973)), and assisting injured or endangered persons (*Wright v. State*, 7S.W.3d 148 (Tex. App. Ct. 1999)). Collectively, case law suggests that the CC doctrine applies to three main categories of police functions: protection of public or private property, protection of individual or community safety, and elimination of public nuisance. Each of these categories, of course, is an area ripe of potential for community policing activity.

Note that the *Dombrowski* Court equivocated on whether the CC doctrine applies when criminal investigative objectives also are present. The reality is that even when such objectives are not present at the outset of an engagement between citizens and police when police are acting in their community caretaking capacity, it is likely that at some point the encounter may take on a criminal investigative character. What to do in such situations? The answer requires resolving two sub-issues.

First, may an officer legally conduct a search under the CC doctrine when his true subjective intent is to investigate criminal activity? At least one court has held that the CC doctrine can apply even if the subjective purpose of the officer is investigative in nature stating that “a court may consider an officer’s subjective intent in evaluating whether the officer was acting as a bona fide community caretaker; however, if the court concludes that the officer has articulated an objectively reasonable basis under the totality of the circumstances for the community caretaker function, he has met the standard of acting as a bona fide community caretaker, whose community caretaker function is totally divorced from law enforcement functions” (*State v. Kramer*, 759N. W.2d 598 (Wisc. 2009)). In contrast, John Decker insists that an “officer’s actions must be motivated by an intent to aid or protect, rather than to solve a crime” in order for the CC doctrine to apply (Decker 1999). Decker then makes an argument similar to the Supreme Court’s reasoning in *Indianapolis v. Edmond*, 531 U.S. 32 (2000). There the Court invalidated a randomized checkpoint utilizing dog sniffs of cars to detect drugs because each sniff of a car was not motivated by individualized suspicion and therefore lacked probable cause.

The second sub-issue is this: given the almost inevitable presence of some criminal/investigative purpose in a CC case, how much is too much? Judge Livingston proposes that for the CC doctrine to apply, “a legitimate community caretaking purpose [must have] clearly predominated over any law enforcement purpose that was present.” Moreover, the CC purpose must “constitute an independent and substantial justification for the intrusion” (Livingston 1998).

Finally, a live question among the courts is whether the CC doctrine applies to private residences. The Supreme Court has never provided specific guidance on this issue, although in a case concerning police use of without a warrant of a heat sensor outside of a home in an attempt to detect increased thermal energy resulting from growing marijuana plants indoors, the Court seemingly has indicated that warrantless searches of homes presumptively are invalid. The lower federal courts are divided on the issue. Three circuits permit the application of the CC doctrine to homes, and four circuits (3rd, 7th, 9th, 10th) do not (Marinos 2012). Commentator Mary Naumann concludes that many state courts are more lenient than the federal courts, and several allow police to argue, when a citizen objects to an intrusion, that their actions as public servants were reasonable in the context of a balancing test wherein an individual’s interest in privacy or autonomy is balanced against the state’s interest in having the police act as public servants (Naumann 1999).

While the police can perform valuable and useful service as community caretakers, the potential danger of sanctioning this activity without limits should be obvious. To the extent that one believes that the Fourth Amendment’s individualized suspicion required provides broad and sturdy protection against invasions of individual rights, then one should be skeptical of a doctrine that not only relaxes the warrant requirement but does not necessarily require even probable cause alone to justify a police action (but see Harcourt and Meares 2011). Thus, there is a risk that the doctrine could be used as pretext for criminal investigation, providing police with a loophole for getting around Fourth Amendment strictures.

Conclusion

In a world in which people increasingly believe that police can and should play a large role in producing public safety, there is an increasing likelihood that police will utilize their discretion to play a proactive rule with respect to the public as opposed to a reactive one. Community policing as a philosophy presupposes this kind of activity. However, the constitutional law that provides a framework for policing has largely assumed that police act in a reactive investigatory fashion activated by citizen complaints regarding crime. The Supreme Court's community caretaking jurisprudence illustrates the dilemmas that both law enforcers and private citizens face as they attempt to navigate the realities of public safety production in their neighborhoods against the backdrop of a constitutional law that has been for the last few decades very skeptical of police discretion. Continued interplay among the judicial, legislative, and executive branches and public activism at the local level will be necessary to work out the rules of engagement that are both protective of individual rights while allowing police enough flexibility to do the jobs as public agents that their principals expect them to do on their behalf.

Related Entries

- ▶ Causes of Police Legitimacy
- ▶ Communities and the Police
- ▶ Community Policing
- ▶ Defining Disorder
- ▶ Law of Police Seizures and the Exercise of Discretion
- ▶ Legal Control of the Police
- ▶ Order Maintenance Policing
- ▶ Police Discretion and Its Control
- ▶ Police Legitimacy and Police Encounters
- ▶ Problem-Oriented Policing
- ▶ Procedural Justice, Legitimacy, and Policing
- ▶ Role and Function of the Police
- ▶ Social Disorder and Physical Disorder at Places
- ▶ Theories on Policing and Communities

Recommended Reading and References

Decker J (1999) Emergency circumstances, police responses, and fourth amendment restrictions. J Crim Law Criminol 89:433

Harcourt B, Meares T (2011) Randomization and the fourth amendment. Univ Chic Law Rev 78:809

Livingston D (1998) Police, community caretaking, and the fourth amendment. University of Chicago Legal Forum, p 261

Marinos MP (2012) Breaking and entering or community caretaking? A solution to the overbroad expansion to the inventory search. Civ Rights Law J 22:249, 264, 270

Naumann ME (1999) The community caretaker doctrine: yet another fourth amendment exception. Am J Crim Law 26:325

Walker S (1992) The police in America: an introduction, 2nd edn. McGraw-Hill, NewYork

Law of Confessions

- ▶ Law of Police Interrogation

Law of Crime Concentrations at Places

David Weisburd[1,2] and Cody W. Telep[3]
[1]Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA
[2]Faculty of Law, The Hebrew University Mt. Scopus, Jerusalem, Israel
[3]School of Criminology and Criminal Justice, Arizona State University, Phoenix, AZ, USA

Overview

Crime is not distributed randomly across jurisdictions but instead clusters geographically. As Eck and Weisburd (1995, 12) note in their chapter on theories of crime and place, “Crime events are not uniformly distributed, a fact known for over a century. At every level of aggregation, some geographic areas have less crime than others.” This has been demonstrated historically at multiple levels of geography (Weisburd et al. 2009a), and it is a fact that is typically well recognized by even private citizens, who may characterize some locations or neighborhoods as “good” and others as “bad.” What has received less empirical attention until recently, however, is the strong concentration of crime at particular small places across cities. Places in this “micro” context are specific locations within the larger social environments of communities and neighborhoods (Eck and Weisburd 1995). Recent studies point to the potential theoretical and practical benefits of focusing research on crime places. In particular there has been a consistent finding that crime is tightly concentrated at just a small number of micro places in a city. These places are typically referred to as crime hot spots (Sherman and Weisburd 1995). They represent small geographic areas with high levels of criminal activity (typically measured by crime incidents or emergency calls for service) relative to other places in the city. This microgeographic conception of place approaches the distribution of crime with much smaller units of analysis than the neighborhoods or communities that have traditionally been of interest to criminologists studying crime and place (Weisburd et al. 2009a).

In one of the pioneering studies in this area, for example, Sherman et al. (1989) found that only 3.5 % of the addresses in Minneapolis, Minnesota, produced 50 % of all calls to the police in a single year. Fifteen years later, in a retrospective longitudinal study in Seattle, Washington, Weisburd et al. (2004) reported that between 4 % and 5 % of street segments in the city accounted for 50 % of crime incidents for each year over 14 years. These findings suggest that a focus on “good” and “bad” neighborhoods misses an important part of the story (see Groff et al. 2010). The unit of interest in understanding the distribution of crime should be much smaller in scope. A very small proportion of places in a jurisdiction are typically responsible for a substantial percentage of citywide crime.

The findings of remarkable concentrations of crime at place raise a more general question about the phenomenon of crime in cities. Is there some general law that applies across cities that dictates the general concentration of crime? This is the question raised in a recent book by Weisburd et al. (2012) entitled *The Criminology of Place: Street Segments and Our Understanding of the Crime Problem*. Studying crime at street segments in Seattle, they found a remarkable stability of crime concentrations each year over a 16-year period. They argue that these data, as well as prior studies showing similar concentrations of crime for specific years in other cities (e.g., see Pierce et al. 1988; Sherman et al. 1989), suggest that crime concentrations at micro places are relatively constant with about 5 % of places producing about 50 % of crime in a city each year.

The idea of a “law” of crime rates is not a new one. Emile Durkheim raised this possibility more than a century ago. Durkheim suggested that crime was not indicative of pathology or illness in society but at certain levels was simply evidence of the normal functioning of communities (Durkheim 1893 [1964], 1895 [1964]). For Durkheim, the idea of a normal level of crime reinforced his theoretical position that crime helped to define and solidify norms in society. While Durkheim’s proposition regarding a normal level of crime in society does not seem to fit recent experience and is seldom discussed by criminologists today, Weisburd et al. (2012) argue that there is indeed a “normal level of crime” in cities, but one that relates to the concentration of crime at place and not to the overall rate of crime. While the absolute levels of crime in cities vary year to year, the extent of crime concentrations remains similar. The idea of a “law of crime concentrations” will be discussed in more detail after first reviewing more thoroughly the empirical research suggesting that crime is highly concentrated at micro units of geography.

Empirical Examples of the Concentration of Crime

A number of studies over the past 20–30 years have found that a relatively small number of micro places are responsible for a significant proportion of total crime in a city. One of the most important early studies in this area was Sherman et al.'s (1989) analysis of emergency calls for service to addresses over a single year (December 1985–December 1986). Sherman et al. (1989) found that only 3.3 % of the addresses in Minneapolis produced just over 50 % of all calls to the police. If crime were randomly distributed across the 115,000 addresses in Minneapolis, one would not expect any places to have 15 or more calls in a single year. Instead, Sherman and colleagues (1989) found 3,841 such addresses, indicating that crime is far more concentrated than would be expected by chance. Pierce and colleagues (1988) found almost identical results when examining crime call concentrations in Boston, Massachusetts.

Weisburd et al. (1992) examined the distribution of crime in Minneapolis when aggregating address level data for a subsequent year (June 1987–June 1988) to crime hot spots of about one street block in length. Examining blocks with at least 20 hard crime (i.e., serious violent and property crime) calls for service, they found 365 hot spots in the city, which represented about 2.5 % of Minneapolis' street segments. These streets accounted for 27.3 % of all hard crime calls for service in the city. These streets also accounted for 27.8 % of soft crime (i.e., disorder) calls for service across the 1-year period. Overall then, whether examining individual addresses or clusters of high crime addresses, crime was highly concentrated at a small number of places.

More recently, Weisburd, Telep, and Lawton (in press) found high rates of concentration when examining crime incidents in New York City in 2009 and 2010. About 77 % of total incidents were geocoded to street segments. Approximately 52 % of the incidents on street segments were found in the top 5 % of street segments both years. Just under a quarter of crime incidents were found at 1 % of the street segments. Similar concentrations were found when examining incidents geocoded to intersections. About half of incidents at intersections were found at the top 5 % of the intersections, while slightly more than 20 % were found at the top 1 % of intersections. Between 55 % and 60 % of the street segments and intersections each year had no crime incidents.

Weisburd and Amran (forthcoming) found similarly high rates of the concentration of crime at place when examining crime incidents at street segments in 2010 in Tel Aviv-Jaffa, Israel. One quarter of all incidents were found at just 0.9 % of the street segments in the city, and 50 % of incidents were located at about 4.5 % of total street segments. About 60 % of street segments in the city did not experience any crime incidents in 2010. These findings suggest the broad application of the "law of concentrations" for crime at place. Even in a jurisdiction thousands of miles from the United States, similar levels of crime concentration were found.

Spelman (1995) offers some words of caution about the use of cross-sectional data in analyzing the concentration of crime at place (see Eck and Weisburd 1995). He examined calls for service at schools, public housing projects, subway stations, and parks and playgrounds longitudinally in Boston. He found evidence of a high degree of stability of crime at the "worst" of these places over a 3-year period. While the top 10 % of public places in terms of crime were responsible for about 30 % of calls for service (see also Eck et al. 2000), he notes that "Much of the concentration of crime among locations is due to random and temporary fluctuations that are beyond the power of the police and the public to control reliably" (Spelman 1995, 142). Thus, even though many of the highest crime places remained high crime, some of the change over time reflected random variation, suggesting the importance of longitudinal approaches to the study of crime at place. Despite these cautions, Spelman (1995) still advocated for efforts such as community problem-solving to address underlying problems at persistent high crime locations.

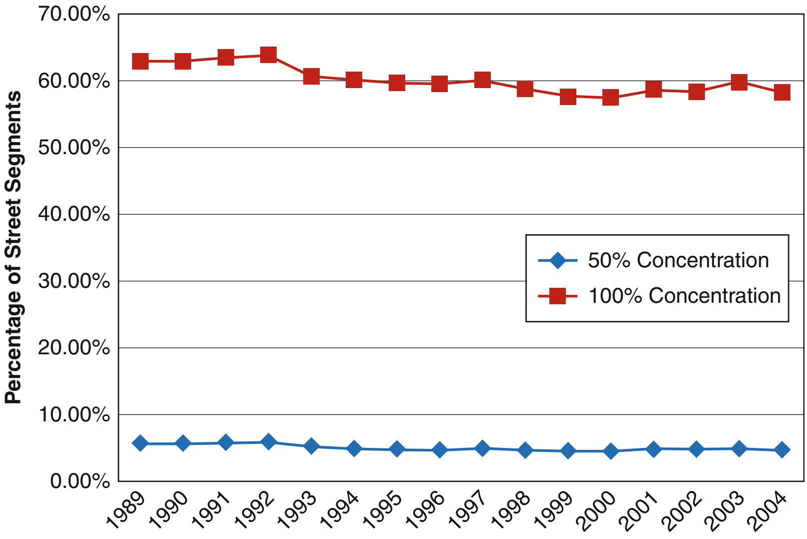

Law of Crime Concentrations at Places, Fig. 1 Crime incident concentrations in Seattle, Washington, 1989–2004

More recently, two studies in Seattle have examined the distribution of crime at street segments using a much longer time series. Weisburd et al. (2004) not only confirmed the concentration of crime but also the stability of such concentrations across a long time span. Weisburd et al. (2004) examined street segments in the city of Seattle from 1989 through 2002. They found that 50 % of crime incidents over the 14-year period occurred at between 4 % and 5 % of the street segments each year. This concentration was also very stable year to year.

Weisburd et al. (2012) updated and expanded these analyses, examining crime incidents at street segments in Seattle from 1989 to 2004. They again found that crime was highly concentrated and these concentrations remained fairly stable across the time period under study. They found that each year 50 % of crime incidents occurred on between 4.7 % and 6.1 % of street segments (see Fig. 1). All crime was found on around 60 % of street segments each year, suggesting that about 40 % of street segments recorded no crime each year in Seattle. The 247 highest crime street segments, about 1 % of the total in Seattle, were responsible for over 23 % of crime incidents across the 16-year period. Crime was even more heavily concentrated when examining a smaller proportion of citywide crime. For example, in the year 2000 in Seattle, just 11 street segments out of over 24,000 in total were responsible for 5 % of crime in the city, and only 31 streets produced over 10 % of the crime incidents.

These strong crime concentrations remain when focusing on specific types of offenders or crimes. For example, Weisburd et al. (2009b) examined the concentration of crime incidents in which a juvenile was arrested in Seattle. They found even stronger levels of concentration than for crime incidents more generally. Less than 1 % of street segments were responsible for 50 % of juvenile arrest incidents each year from 1989 to 2002. Only 86 street segments accounted for one third of all official juvenile arrest incidents over the 14-year period. As another example, Weisburd and Mazerolle (2000) examined the number of calls for service in 56 drug hot spots in Jersey City, New Jersey, that made up about 4.4 % of the street segments and intersection in the city. About half (47 %) of both narcotics arrests and calls for service were found in these hot spots, suggesting the strong concentration of drug activity in Jersey City.

In Boston, Braga et al. (2010) found that just 4.8 % of the street segments were the site of 73.9 % of all gun assaults over a 29-year period. Braga et al. (2011) also examined the concentration of robbery incidents at street segments and intersections in Boston from 1980 to 2008.

Each year, about 2 % of street units accounted for 50 % of the robbery incidents in the city. The top 52 street units, which represented just 0.18 % of total units in the city, were responsible for 10,886 of the 135,276 robberies (8.0 %) over the 29-year period. Commercial robberies are even more highly concentrated than street robberies. From 1980 to 2008, just 1.3 % of street segments were responsible for about half of commercial robberies. In Minneapolis, Sherman et al. (1989) found that the 54 addresses (about 0.047 % of the total addresses in the city) that had at least ten calls for service for public predatory crime (a combination of robbery, auto theft, and rape) were responsible for over 7.5 % of all such calls.

These trends for specific crime types are once again consistent when looking at data from outside the United States. Andreson and Malleson (2011) examined the distribution of calls for service for seven offense types in Vancouver, Canada, in 1991, 1996, and 2001. They found that for each crime type, 50 % of calls were found on between about 1 % and 8 % of street segments. Burglary showed the lowest level of concentration, although in 2001, for example, 50 % of burglary calls were on 7.61 % of street segments and only 39.43 % of streets had any burglary calls. Robbery was incredibly concentrated in Vancouver. In 2001, 50 % of robbery calls were found on just 0.84 % of street segments in the city, and only 5.32 % of streets had any robbery calls.

Sherman (1995) argues that such clustering of crime at places is even greater than the concentration of crime among individuals. Weisburd et al. (2004) found that 5 % of streets produced 50 % of crime throughout the 14-year study period. In Wolfgang et al.'s (1972) groundbreaking Philadelphia, Philadelphia birth cohort study, they found that 6 % of the cohort produced 50 % of the offenses committed by the cohort. While these statistics seem similar, 6 % of a city's population is likely to yield thousands of targets, while 5 % of street segments in a city like Seattle is only around 1,500 places. And those places are not "moving" targets for police since they stay in the same place. This suggests that focusing on places may be more efficient than focusing on people (Weisburd and Telep 2010). As Sherman (1995, 37), "Why aren't we thinking more about wheredunit, rather than just whodunit?"

Importantly, Weisburd and colleagues (2012) also found that the micro place concentrations in Seattle were not just proxies for neighborhood-level effects (see also Groff et al. 2010). While they did find some evidence of larger area effects in their analyses and some clustering of high crime streets, particularly in the Central Business District, their overall conclusion was that there is a great deal of street-by-street heterogeneity in crime patterns. In other words, crime hot spots can be found throughout the city, and so there is a great deal of variability street-to-street within both "good" and "bad" neighborhoods.

A Law of Crime Concentrations

These findings about the concentration of crime are particularly interesting in light of Emile Durkheim's classic proposition that the level of crime is stable in society or rather that there was a "normal level" of crime in society. For Durkheim, this meant that crime was not necessarily an indication of an illness or pathology in society but rather that healthy societies would inevitably have some normal level of crime. Crime waves and crime drops in this context can be seen as the result of some "abnormality" in a society that results from crisis or dramatic social change.

Underlying Durkheim's proposition was his understanding of crime as a product of social definition. Kai Erickson (1966) was to build upon this idea in his classic study *Wayward Puritans*, where he sought to show that the definition of crime had a social function. By defining others as deviant, society can help draw the boundaries between acceptable and unacceptable conduct (see also Adler and Adler 2009). Defining people as criminal in this sense serves a function in defining the moral boundaries of society. One can know the boundaries of acceptable behavior by observing "deviants" who are sanctioned for violating societal norms.

Crime rates over the last few decades would seem to strongly contradict Durkheim's conception of normal levels of crime in society. Between 1973 and 1990, violent crime doubled, and then in the 1990s, the USA experienced a well-documented "crime drop" (Blumstein and Wallman 2000). In the 1970s, Blumstein and Cohen (1973) hypothesized that Durkheim's proposition could be applied to punishment in America, where imprisonment rates had remained static for a long period of time. But recent dramatic increases in the US incarceration rate in the 1980s and 1990s would seem inconsistent with the normal crime or "normal punishment" (Blumstein and Cohen 1973) hypothesis.

Weisburd et al. (2012), however, take a different approach to Durkheim's theory and instead argue that there is indeed a "normal" level of crime in cities, but one that relates to the concentration of crime at place and not to the overall rate of crime. They claim that a different proposition from Durkheim's can be raised at this juncture and should be examined in future studies. There appears to be a "law of concentrations" of crime at place. The consistency of crime concentrations at micro places over time and across geographic locations as diverse as Minneapolis, Minnesota, and Tel Aviv-Jaffa, Israel, suggests that there is some underlying social process pushing crime to certain levels of concentration in modern cities.

What might lead to this stability? It is important to note at the outset that the concentration of crime follows patterns of concentration in many other areas of scientific inquiry (e.g., see Eck et al. 2007; Koch 1999; Sherman 2007). Joseph Moses Juran (1951) first noted this concentration in looking at economic activities, coining the phrase "the vital few and the trivial many." Juran sought to emphasize to managers that they should focus on the small number of events or cases that produce the majority of relevant business activities, for example, the small number of defects that cause most complaints about products or the small number of clients that are responsible for a majority of revenue. Juran termed this phenomenon the "Pareto Principal" after Vilfredo Pareto (1909), who first brought attention to what is sometimes referred to as the 80-20 rule (see Koch 1999). Pareto observed that a number of distributions seem to follow this specific pattern of concentration. For example, in studying land ownership in Italy, he found that 80 % of the land was controlled by just 20 % of the population. He also observed that 20 % of the pea pods in his garden produced 80 % of the peas. The 80-20 rule is generally seen only as an approximation, but it applies fairly well to Weisburd et al.'s (2012) data in Seattle. Eighty percent of the crime incidents each year were found on between 19 % and 23 % of the street segments. Of course, the question remains, what leads to the tremendous concentration and stability of crime at microgeographic places over time?

In their discussion of the concentration of crime at facilities, Eck and colleagues (2007) discuss how crime follows the J-curve pattern found in a number of other phenomena. As they describe, "To reveal a J-curve, the number of crimes in a given time period at each facility needs to be known, and then the facilities ranked from those with the most crimes to those with the fewest. If a bar chart of the crime frequency is drawn, a few facilities at the left end of this distribution will have many crimes, but as one moves to the right there will be a steep drop-off in crimes that flattens out at a very few or no crimes for the majority of the facilities. The resulting graph resembles a reclining J" (Eck et al. 2007, 228). The words "micro place" can easily be substituted for "facilities."

They point to five potential reasons why crime might be concentrated at particular facilities, which are relevant to consider when understanding the concentration of crime at micro places. The first was noted above; some portion of the concentration can be explained by random variation. That is, some hot spots of crime would likely cool off without any intervention. Random variation, however, is not the whole story, as Weisburd et al. (2012) demonstrated by showing that the hottest street segments in Seattle remained high crime throughout the 16-year study period. A second explanation is changes in reporting processes. If the police changed their crime incident reporting protocols, for example, this could have some impact on the level and concentration of crime, although again this does not explain the stability of concentration in Seattle, which experienced no major change in reporting patterns from 1989 to 2004.

The last three potential explanations focus on the characteristics of particular facilities (or places). Eck and colleagues (2007) point to offenders, targets, and place management as potential reasons why some facilities may be riskier than others. Some places may be crime attractors if they tempt potential offenders with rampant opportunities for criminal activity. Places can also be rich in targets and be crime generators. Finally, poor place management (i.e., a lack of supervision and guardianship) can be a crime enabler. Examining the characteristics of places was the focus of Weisburd et al.'s (2012) research in Seattle, and they demonstrated that crime opportunity factors play a key role in explaining the concentration of crime in hot spots. These factors are important for understanding why some places become hot spots, but they do not necessarily explain why such a small number of hot spots are responsible for such a high percentage of crime citywide.

Can Durkheim's initial insights also be used to consider possible reasons for this law of concentrations of crime at place? Following Durkheim and other theorists that built on his work, one would look to the role of crime at place in defining normative boundaries in society. In this case, it could be argued that a certain number of places in the city with severe crime problems serve as lessons for the city more generally. This would fit well with the finding by Weisburd et al. (2012) that crime hot spots are found throughout the city. Accordingly, everyone would have direct visceral experiences with the "bad places" in the city, and perhaps that serves to define the "moral boundaries" of place for individuals. The normal level of crime concentrations in this context would relate to the proportion of problem places that are needed to bring the lessons of moral boundaries to the city's residents.

Another possible explanation for a law of concentrations comes from the concentration of other characteristics of places in the city. For example, Weisburd et al. (2012) note that the concentration of bus stops or number of public facilities, like crime, stays relatively stable over long periods. Perhaps the law of concentrations of crime is related to the overall distribution of social and environmental characteristics of places in cities. Does the stability of patterns of business and employment in a city, for example, reflect more general patterns of concentration that are related to the growth and development of urban areas? Cities regulate such concentrations, by defining commercial, business, and industrial use of property. Perhaps the normal concentrations of crime are simply a reflection of the normal concentrations of other social activities in the city. The law of concentrations of crime at place may simply be a reflection of a more general law of the stability of concentrations of specific aspects of social and economic life in the city, as discussed above with Juran's (1951) notion of the Pareto Principle.

But this brings the discussion back to Durkheim, because crime is a social phenomenon and its tolerance is a social construct. Is society willing to tolerate crime at only a certain proportion of the landscape of a city? Is the law of concentrations a result of the boundaries of crime at place that citizens are willing to tolerate? Will people become worried and call for action when crime hot spots increase beyond a specific proportion of places in the city, and will they become more lax when the concentrations are below that level? Certainly, this law of concentrations needs to be studied across other metropolitan centers around the world to see how widely it applies. More generally, it is time for scholars to explore more directly the explanation for the law of concentrations of crime trends at micro places. In the spatial context, scholars should explore both social and environmental characteristics of street segments that are important to understanding the concentration of crime at place.

Related Entries

- ▶ Criminology of Place
- ▶ Geography of Crime and Disorder
- ▶ History of Geographic Criminology Part I: Nineteenth Century

- ▶ History of Geographic Criminology Part II: Twentieth Century
- ▶ Hot Spots and Place-Based Policing
- ▶ Longitudinal Crime Trends at Places

Recommended Reading and References

Adler PA, Adler P (2009) Constructions of deviance: social power, context, and interaction, 6th edn. Wadsworth Publishing, Belmont

Andreson MA, Malleson N (2011) Testing the stability of crime patterns: implications for theory and practice. J Res Crime Delinq 48:58–82

Blumstein A, Cohen J (1973) A theory of the stability of punishment. J Crim Law Criminol 64:198–207

Blumstein A, Wallman J (eds) (2000) The crime drop in America. Cambridge University Press, Cambridge

Braga AA, Hureau DM, Papachristos AV (2011) The relevance of micro places to citywide robbery trends: a longitudinal analysis of robbery incidents at street corners and block faces in Boston. J Res Crime Delinq 48:7–32

Braga AA, Papachristos AV, Hureau DM (2010) The concentration and stability of gun violence at micro places in Boston, 1980–2008. J Quant Criminol 26:33–53

Durkheim E (1893 [1964]) The division of labour in society. Free Press, New York

Durkheim E (1895 [1964]) The rules of sociological method. In: Catlin GEG (ed) Translated by Solovay SA, Mueller JH. Free Press, New York

Eck JE, Clarke RV, Guerette RT (2007) Risky facilities: crime concentration in homogenous sets of establishments and facilities. In: Farrell G, Bowers KJ, Johnson SD, Townsley M (eds) Imagination in crime prevention. Crime prevention studies, vol 21. Criminal Justice Press, Monsey, pp 225–264

Eck JE, Gersh JS, Taylor C (2000) Finding crime hot spots through repeat address mapping. In: Goldsmith V, McGuire PG, Mollenkopf JH, Ross TA (eds) Analyzing crime patterns: frontiers of practice. Sage, Thousand Oaks, pp 49–64

Eck JE, Weisburd D (1995) Crime places in crime theory. In: Eck JE, Weisburd D (eds) Crime and place, crime prevention studies, vol 4. Willow Tree Press, Monsey, pp 1–33

Erikson KT (1966) Wayward puritans: a study in the sociology of deviance. Wiley, New York

Groff ER, Weisburd D, Yang S-M (2010) Is it important to examine crime trends at a local "micro" level?: a longitudinal analysis of street to street variability in crime trajectories. J Quant Criminol 26:7–32

Juran JM (1951) Quality control handbook. McGraw-Hill, New York

Koch R (1999) The 80/20 principle: the secret of achieving more with less. Nicholas Brealey Publishing, Naperville

Pareto V (1909) Manuel d'économie politique. Vol. 7 of oeuvres complètes. Droz, Geneva

Pierce G, Spaar S, Briggs LR (1988) The character of police work: strategic and tactical implications. Center for Applied Social Research, Northeastern University, Boston

Sherman LW, Gartin P, Buerger ME (1989) Hot spots of predatory crime: routine activities and the criminology of place. Criminology 27:27–55

Sherman LW (1995) Hot spots of crime and criminal careers of places. In: Eck J, Weisburd D (eds) Crime and place. Crime prevention studies, vol 4. Willow Tree Press, Monsey, pp 35–52

Sherman LW (2007) The power few: experimental criminology and the reduction of harm. J Exp Criminol 3:299–321

Sherman LW, Weisburd D (1995) General deterrent effects of police patrol in crime 'hot spots': a randomized, controlled trial. Justice Q 12:625–648

Spelman W (1995) Criminal careers of public places. In: Eck JE, Weisburd D (eds) Crime and place, crime prevention studies, vol 4. Willow Tree Press, Monsey, pp 115–143

Weisburd D, Amram S (forthcoming) The law of crime concentrations at places: the case of Tel Aviv. Police Pract Res Int J

Weisburd D, Bruinsma GJN, Bernasco W (2009a) Units of analysis in geographic criminology: historical development, critical issues, and open questions. In: Weisburd D, Bernasco W, Bruinsma GJN (eds) Putting crime in its place: units of analysis in geographic criminology. Springer, New York, pp 3–31

Weisburd D, Bushway S, Lum C, Yang S-M (2004) Trajectories of crime at places: a longitudinal study of street segments in the city of Seattle. Criminology 42:283–321

Weisburd D, Groff ER, Yang S-M (2012) The criminology of place: street segments and our understanding of the crime problem. Oxford University Press, New York

Weisburd D, Maher L, Sherman L (1992) Contrasting crime general and crime specific theory: the case of hot spots of crime. In: Adler F, Laufer W (eds) New directions in criminological theory, advances in criminological theory, vol 4. Transaction Press, New Brunswick, pp 45–69

Weisburd D, Morris NA, Groff ER (2009b) Hot spots of juvenile crime: a longitudinal study of street segments in Seattle, Washington. J Quant Criminol 25:443–467

Weisburd D, Mazerolle LG (2000) Crime and disorder in drug hot spots: implications for theory and practice in policing. Police Q 3:331–349

Weisburd D, Telep CW, Lawton BA (in press) Could innovations in policing have contributed to the New York City crime drop even in a period of declining police strength?: the case of stop, question and frisk as a hot spots policing strategy. Justice Quarterly

Weisburd D, Telep CW (2010) The efficiency of place-based policing. J Police Stud 17:247–262

Wolfgang ME, Figlio RM, Sellin T (1972) Delinquency in a birth cohort. University of Chicago Press, Chicago

Law of Police Interrogation

George C. Thomas III[1] and Richard A. Leo[2]
[1]Rutgers University School of Law, Newark, New Jersey, NJ, USA
[2]University of San Francisco School of Law, San Francisco, CA, USA

Synonyms

Law of confessions; Police interrogation

Overview

The law of interrogation in the United States is largely a product of United States Supreme Court cases interpreting the United States Constitution. Three parts of the Constitution regulate interrogation practices. The most obvious regulation is found in the Fifth Amendment right not to be compelled to incriminate oneself. The United States Supreme Court also found in the Due Process Clause of the Fifth and Fourteenth Amendments a right not to be coerced into confessing. The Court read the Sixth Amendment right to the assistance of counsel to offer protection against interrogation in some circumstances. Each of these rights provides a slightly different protection, but together they seek to ensure that a suspect makes an uncoerced decision about whether to confess and has the assistance of a lawyer when desired.

Due Process Requirement that Confessions Be Voluntary

The Court's due process analysis initially drew heavily on the English common law. As far back as the late seventeenth century, English judges were skeptical of confessions that were influenced by threats or promises of favor. By 1788, the leading technical treatise of English law reported that the law of confessions refused to admit statements obtained "by the flattery of hope, or by the impressions of fear, however slightly the emotions may be implanted . . . for the law will not suffer a prisoner to be made the deluded instrument of his own conviction" (Hawkins 1788). Confessions taken under those conditions were later referred to as "involuntary."

The core notion of "involuntary" follows the English common law, with "flattery of hope" understood as promises of benefit and "impressions of fear" understood to be threats. In the infamous case of *Brown v. Mississippi*, 1936, for example, the United States Supreme Court held that the Fourteenth Amendment Due Process Clause forbids the use of confessions that are produced by brutal beatings and threats that the torture would continue. The Court held in several cases that threats made the confessions involuntary even where there was no torture. In *Payne v. Arkansas*, 1958, the chief of police said that he would probably be able to protect the suspect from the mob outside if he told the truth about what happened. In *Lynum v. Illinois*, 1963, the police told the suspect that she would lose custody of her children if she did not "cooperate" with them. In *Arizona v. Fulminante*, 1991, a government informant promised to protect the suspect from violent reprisals from other prisoners but only if the suspect told the truth about raping and murdering his stepdaughter.

The Supreme Court also applied the term "involuntary" to confessions that did not include classic examples of threats or promises of benefits. One way to describe the Court's involuntariness doctrine as it evolved is that it serves "a complex of values . . .; which, by way of convenient shorthand, this Court terms involuntary, and the role played by each in any situation varies according to the particular circumstances of the case" *Blackburn v. Alabama*, 1961. In its more than 40 involuntary confessions cases, the Court has expressed concern about unreliable confessions, about protecting the will of the suspect, and about morally offensive police conduct.

For example, *Ashcraft v. Tennessee*, 1944, held that 36 h of nonstop questioning made the defendant's confession involuntary, even though police made no threats or promises, because the questioning deprived him of the "mental freedom" to decide whether to answer. *Townsend v. Sain*, 1963, reached the same result under a mental freedom theory when the police gave the suspect a drug that functioned as a form of truth serum. In *Spano v. New York*, 1959, a combination of factors rendered the confession involuntary. First, the suspect was "emotionally unstable and maladjusted." Police also repeatedly refused his requests for counsel and told him, falsely, that if he did not confess, his police officer friend might lose his job. The Court held that this combination of factors made the suspect's confession involuntary, in part because the police conduct was itself outside the law and thus morally offensive. In *Malinski v. New York*, 1945, the police strip-searched the suspect, gave him a blanket, and interrogated him while he was naked in an effort, the prosecutor said, "to humiliate him." The Court held the confession involuntary even though the interrogation was relatively brief.

Cases involving threats and promises of favor have traditionally raised issues about the reliability of the confession. Brutal torture, such as was used in *Brown v. Mississippi*, 1936, is as likely to produce a false confession as a true one. But in 1986, the Court held that the unreliability of a confession, standing alone, is not relevant to whether the Due Process Clause has been violated. *Colorado v. Connelly*, 1986, held that as long as the confession is not coerced or compelled by state actors, its reliability is completely a matter of state law. DNA evidence has recently made plain that police methods well short of torture can produce false confessions that are used to convict innocent defendants. These DNA exonerations have caused courts to focus on police deception and trickery – for example, lies about evidence – used to obtain confessions. The underlying notion is that if the police make the case against the innocent suspect look hopeless, he might falsely confess to some lesser involvement to escape the most severe punishment.

Police have long relied on deception during interrogation; in the famous case of *Miranda v. Arizona*, 1966, the Court noted a police strategy in which someone pretending to be an eyewitness would pick the suspect out of a lineup. Other deceptive techniques include fake lie detector results, fake tests that "prove" the suspect had recently fired a gun, lying about the suspect's fingerprints being at the scene of the crime, and lying about the existence of eyewitnesses. The issue that courts must address in deception cases is whether police techniques, either alone or in conjunction with other high-pressure interrogation strategies, render a subsequent confession involuntary. This might be true because the technique overbore the suspect's will, was a morally offensive police technique, or created too high a risk that an innocent suspect might confess falsely. American courts have generally permitted the use of confessions obtained by deception, while sometimes condemning the police conduct, on the ground that an innocent suspect would know that the evidence was false and could resist the temptation to confess.

But there are limits beyond which police may not go. A New Jersey Court held that it violated due process to play an audiotape of a fake eyewitness who claimed to have seen the shooting and identified the suspect as the shooter (*State v. Patton*, N.J. Superior Court, Appellate Division 2003). In *Cayward v. Florida*, Florida District Court of Appeal, 1989, the police showed the suspect two official reports that "proved" that his semen was found on the rape/murder victim. "One false report was prepared on stationery of the Florida Department of Criminal Law Enforcement; another was prepared on stationery of Life Codes, Inc., a testing organization." The Florida Court held that the use of manufactured documents crossed the line between permissible and impermissible deception.

As the cases discussed to this point suggest, the voluntariness test was difficult to apply. There are many factors to consider. More fundamentally, the inquiry itself is hopelessly indeterminate. Absent the kind of torture in *Brown* or overt threats of severe harm, it is not obvious whether a suspect answers police questions because it is his will to answer or because the police have overcome his volition. These difficulties led the

Supreme Court, and all the state courts of appeal, to review many state court decisions, each with unique facts.

When 1966 dawned, the Court realized that it needed a confession rule that made the volition inquiry easier and thus could be applied more mechanically than the voluntariness test.

Miranda v. Arizona's Requirement of Free Choice

In *Miranda*, 1966, the Court turned to the Fifth Amendment right not to be compelled to be a witness against oneself, often called the privilege against compelled self-incrimination. While freedom from compulsion seems roughly the same as freedom from forces that produce involuntary confessions, the five-justice majority in *Miranda* read the privilege to guarantee that suspects who answer police questions have to make a "free choice." Noting that little is known about what actually goes on in police interrogation rooms, the Court drew from textbook examples of police interrogation strategies, such as lying about the evidence against the suspect and discouraging him from remaining silent or asking for a lawyer. The Court sketched the history of the privilege against compelled self-incrimination, emphasizing its role in protecting the mental freedom to decide whether to cooperate with the State when it is seeking evidence of crime. From all of this, the Court concluded that the atmosphere of custodial police interrogation was inherently compelling. Thus, the Court held that *all statements* made in response to custodial police interrogation are compelled unless steps are taken to ameliorate the compulsion of the interrogation. Otherwise, the suspect could not make a "free choice" whether to answer police questions.

Though the Court left the door open for the states and Congress to come up with another way to ameliorate the inherent compulsion of police interrogation, it held that in the absence of an equally effective remedy, the Constitution required the famous *Miranda* warnings. Suspects must be told that that they have a right to remain silent, that anything they say will be used against them in court, that they have a right to counsel, and that if they cannot afford counsel, a lawyer will be appointed at no cost to them. If a suspect answers questions after being told that he has these rights, answering might be his free choice. If so, the voluntariness question would, in a sense, answer itself in a *Miranda* world.

Most observers thought that *Miranda* would reduce the incidence of incriminating statements, perhaps drastically. While innocent suspects have an incentive to answer questions and explain why the police have arrested the wrong person, a guilty suspect's most rational course of action is to exercise the right to silence and to counsel. In the immediate wake of *Miranda*, some police officials, including the police chief of Los Angeles, accused the Court of trying to "end the use of confessions in convicting criminals" (Baker 1983).

Though the precise effect of *Miranda* on the confession rate is yet unknown, there is widespread agreement that at least 80 % of suspects waive their *Miranda* rights and consent to answer police questions, very often incriminating themselves. Although the *Miranda* Court said that the State had a "heavy burden" to prove the voluntariness of a *Miranda* waiver, later cases held that these waivers may be implied (from the actions or the words of the suspect) rather than explicit (*North Carolina v. Butler* 1979). Moreover, the Court recently held that "a suspect who has received and understood the *Miranda* warnings, and has not invoked his *Miranda* rights, waives the right to remain silent by making an uncoerced statement to the police" (*Berghuis v. Thomkins* 2010). When suspects waive *Miranda* (whether implicitly or explicitly), its protections typically play no role in mitigating the process or outcome of the subsequent interrogation. In almost all cases, the only protection available after *Miranda* has been waived is the protection against involuntary confessions, the very protection that *Miranda* was designed to supplement. Once waived, *Miranda* neither regulates nor restricts psychologically manipulative or deceptive interrogation techniques, hostile or overbearing questioning styles, lengthy interrogation and confinement, or any of the many stressful conditions of modern accusatorial police interrogation. The leading police interrogation manuals continue to encourage some of the very same deceptive practices that were condemned in the *Miranda* opinion.

It seems likely that suspects who waive *Miranda* are responding to police pressure but that they also, at some level, may *want* to talk to police, to tell their side of the story, and to persuade the police that they should be released from custody. David Simon observed the Baltimore Homicide Department for a year. He concluded that every suspect who faces interrogation imagines a "small open window" that is "the escape hatch, the Out" (Simon 1991). Every suspect who "opens his mouth during an interrogation . . . envisions himself parrying questions with the right combination of alibi and excuse" and then "crawling out the window to go home and sleep in his own bed." Thus, it seems that suspects waive their *Miranda* rights at least in part because they often *want to talk to police*. In most cases, of course, talking to police is a supremely irrational act. As David Simon imagines a detective's reflections: "[E]ven after all these years working homicides, there is still a small part of him that finds it completely mystifying that anyone utters a single word in a police interrogation." After all, "what do homicide detectives do for a living? Yeah, you got it, bunk. And what did you do tonight? You murdered someone."

However irrational it is for guilty suspects to talk to police, waiver is surprisingly easy to obtain. Police can also avoid *Miranda* protections because they do not apply to a substantial amount of police questioning. The warnings are required only when police engage in custodial interrogation, and police are free to engage in persistent, marathon questioning – for hours on end with no warnings – as long as they do not take the suspect into custody. Police often tell the suspect that he is not under arrest, to create a noncustodial situation, and then suggest that it is in his best interests to cooperate with them. Any reluctance to answer questions can be met with reassurances that he is not under arrest and that police just want to get his side of the story. Technically, the suspect is not in custody in this situation, and police interrogation without warnings is completely legal. In 1977, Justice Marshall urged the Court to expand *Miranda* to include coercive interrogations where technical custody was absent, arguing that while *Miranda* was initially limited to custodial interrogation, the "rationale of *Miranda* . . . is not so easily cabined" (*Oregon v. Mathiason* 1977, Marshall, J., dissenting). But by 1977, the Court had settled on a course intended to keep *Miranda* from interfering with what it viewed as legitimate police interrogation.

In addition to the ease of proving waiver and the limitation of *Miranda* to custodial interrogation, four doctrines now permit the State to benefit from evidence learned from a violation of *Miranda*. The Court first created a "public safety" exception that permits police to question without warnings if the officer reasonably believes it is necessary to protect the public safety (*New York v. Quarles* 1984). There, police arrested a suspect in a supermarket; they expected him to be armed but his shoulder holster was empty. One officer asked where the gun was, and the Court allowed the State to use his answer on the ground that the officer had an obligation to find the gun and protect the public. A later doctrine that ignores police failure to comply with *Miranda* is that physical evidence found as a result of a *Miranda* violation can be introduced into evidence (*United States v. Patane* 2004). Thus, in the supermarket case, the State could have used the gun itself as evidence even without a public safety exception.

Evidence is also admissible despite a violation of *Miranda* when police succeed in remedying the violation. In most cases, police are permitted to remedy a failure to warn by providing warnings later even if the suspect has already made an incriminating statement. Thus, if the suspect makes an incriminating admission in his parents' home without warnings and it turns out that this is a violation of *Miranda*, giving warnings later at the police station will permit a second confession to be admissible (*Oregon v. Elstad* 1985). *Elstad* would not apply, however, if the police had intentionally withheld warnings in the hope that the suspect would confess and then repeat the confession after the warnings are given (*Missouri v. Seibert* 2004).

A fourth area where statements taken in violation of *Miranda* can still be used is in impeaching the credibility of a defendant who testifies (*Harris v. New York* 1971). For example, if a defendant testifies that he has never sold narcotics, his statement to police admitting a sale can be introduced to suggest that he is not testifying truthfully, even though the statement was taken in violation of *Miranda*. The jury will be instructed not to consider the statement as evidence of guilt but only as to the defendant's credibility. Conventional wisdom is that juries would have a difficult time ignoring an admission of criminal behavior on the issue of guilt. At least a few police departments have advised officers to be willing to violate *Miranda* because the statements will always be available to impeach the defendant's testimony, a threat that seems likely to deter defendants from testifying.

In sum, the kind of robust guarantee of a "free choice" to decide whether to answer police questions probably intended by *Miranda* does not seem to exist. The most frequent loss of *Miranda* protections is through waiver, which leaves suspects facing police interrogators with nothing to protect them from zealous interrogation beyond the vague due process prohibition of involuntary statements. Moreover, in a *Miranda* world, the suspect's agreement to talk to police in effect creates a kind of soft presumption that statements made later in the interrogation must be voluntary. The choice to talk to police was, superficially at least, a free choice. And the suspect was told that he did not have to answer questions. Thus, answering questions an hour, or 2, or 30 h, later seems to be clothed in voluntariness. As Justice Souter noted in 2004: "[G]iving the warnings and getting a waiver has generally produced a virtual ticket of admissibility; maintaining that a statement is involuntary even though given after warnings and voluntary waiver of rights requires unusual stamina, and litigation over voluntariness tends to end with the finding of a valid waiver" (*Missouri v. Seibert* 2004 (plurality)).

Moreover, there is no right to warnings when the intense questioning takes place in a noncustodial situation. And, finally, even if *Miranda* is violated, four doctrines permit the State to use statements and physical evidence in any event.

Right to Counsel

Two years prior to *Miranda*, the Court held in *Massiah v. United States* (1964) that *indicted* defendants have a Sixth Amendment right to counsel when government actors are seeking to elicit incriminating statements. Thus, the government violated Massiah's Sixth Amendment rights when it used an informant to elicit statements from him after he was indicted. Massiah did not know that he was speaking to a government informant and thus there was no issue of waiver of the right to counsel. Waiver first arose in *Brewer v. Williams* (1978). The Court acknowledged that indicted defendants could waive their *Massiah* right to counsel but, on the facts of *Brewer*, the Court held that the state had failed to prove waiver.

The Court did not address how to prove waiver of the *Massiah* right to counsel until 1988. In *Patterson v. Illinois* (1988), the Court held that when an indicted defendant waived his *Miranda* right to counsel, he had simultaneously waived his *Massiah* right to counsel. This followed, the Court held, even though the two rights have different textual homes – *Massiah* the Sixth Amendment and *Miranda* the Fifth Amendment self-incrimination clause – and even though indictment is the beginning of the adversary criminal process. Like *Miranda*, the vote was 5-4. The dissent argued that the Sixth Amendment right to counsel carries with it duties and responsibilities that go beyond advice about answering questions, but the majority said that the issue was the usefulness of a lawyer in the particular proceeding. In the context of interrogation, *Miranda* provides the measure of the usefulness of a lawyer.

One issue left undecided by *Patterson* was whether it matters if a judge has appointed a lawyer during a pretrial, and post-indictment, proceeding before police seek a *Miranda* waiver. Would a waiver of *Miranda* also waive *Massiah* when the indicted defendant is actually represented by counsel? Continuing a string of 5-4 decisions, the Court in *Montejo v. Louisiana* (2009) held that *Miranda* continues to provide the measure of a usefulness of a lawyer even if the defendant is formally represented by counsel. Thus, a waiver of *Miranda* is a waiver of *Massiah* whether or not the indicted defendant is represented by counsel.

In sum, *Massiah* provides indicted defendants with a right not to have incriminating statements elicited surreptitiously – the actual facts of the *Massiah* case – but it is not clear that the *Massiah* right to counsel adds anything to the *Miranda* right to counsel when police seek to interrogate indicted defendants.

Conclusion

The law of custodial police interrogation in 2012 begins with *Miranda*. Suspects who have been indicted also have a *Massiah* Sixth Amendment right to counsel. But as most suspects waive *Miranda*, and thus also waive *Massiah*, the real story continues to be the voluntariness test that has been evolving in Anglo-American law since the late seventeenth century. Nonetheless, *Miranda* continues to exert an enormous influence on the law of interrogation. Unlike the vague, subjective, and indeterminate due process test, *Miranda* provides police interrogators an easily administered procedure that is clear and whose outcomes are predictable. Its clarity and predictability are of great benefit to prosecutors and judges. In addition, *Miranda* has contributed to a civilizing of police behavior generally and to the professionalism of the interrogation process. It also serves symbolic and educational functions. It has increased popular awareness of the right not to answer police questions and of the right to counsel, so much so that the Rehnquist Court in 2000 concluded that *Miranda* has "become embedded in routine police practice to the point where the warnings have become part of our national culture" (*Dickerson v. United States* 2000).

Related Entries

- ▶ Criminal Defense Profession
- ▶ False Confessions and Police Interrogation
- ▶ FBI Influence on State and Local Police
- ▶ History of Criminal Investigation
- ▶ Legal Control of the Police
- ▶ Police Lying and Deception
- ▶ Sex Offenders and Criminal Policy

Recommended Reading and References

Arizona v. Fulminante, 499 U.S. 279 (1991)

Ashcraft v. Tennessee, 322 U.S. 143 (1944)

Baker L (1983) Miranda: crime, law and politics. Atheneum, New York

Berghuis v. Thompkins, 130 S.Ct. 2250 (2010)

Blackburn v. Alabama, 361 U.S. 199 (1961)

Brewer v. Williams, 430 U.S. 387 (1978)

Brown v. Mississippi, 297 U.S. 278 (1936)

Colorado v. Connelly, 479 U.S. 157 (1986)

Dickerson v. United States, 530 U.S. 428 (2000)

Florida v. Cayward, 552 So.2d 971(Fla. Dist. App. Ct. 1989)

Harris v. New York, 401 U.S. 422 (1971)

Hawkins W (1788) A treatise of the pleas of the crown; or, a system of principal matters relating to that subject, digested under proper heads, 6th edn. Eliz. Lynch, Dublin: London

Lynum v. Illinois, 372 U.S. 528 (1963)

Malinski v. New York, 324 U.S. 401 (1945)

Massiah v. New York, 377 U.S. 201 (1964)

Miranda v. Arizona, 384 U.S. 436 (1966)

Missouri v. Seibert, 542 U.S. 600 (2004)

Montejo v. Louisiana, 129 S. Ct. 20179 (2009)

New York v. Quarles, 467 U.S. 649 (1984)

North Carolina v. Butler, 441 U.S. 369 (1979)

Oregon v. Elstad, 470 U.S. 298 (1985)

Oregon v. Mathiason, 429 U.S. 492 (1977)

Patterson v. Illinois, 487 U.S. 285 (1988)

Payne v. Arkansas, 356 U.S. 560 (1958)

Simon D (1991) Homicide: a year on the killing streets. Houghton Mifflin, Boston

Spano v. New York, 360 U.S. 315 (1959)

State v. Patton, 826 A.2d 783, N.J. Super. A. D. (2003)

Townsend v. Sain, 372 U.S. 293 (1963)

United States v. Patane, 542 U.S. 630 (2004)

Law of Police Searches

Fabio Arcila Jr.
Touro Law Center, Central Islip, NY, USA

Overview

In the United States, all governmental law enforcement agents, whether federal, state, or local, must comply with the Fourth Amendment's search and seizure requirements. Unfortunately, those requirements are often unclear, and there is no consensus about the analytical method for determining them. One option is historical analysis, and the United States Supreme Court has increasingly employed it. However, there are significant differences between the law of police searches today and what existed at the nation's founding, and even the Court itself has resisted the consequences of more fully returning to a historical model. This resistance results from one of the great challenges in Fourth Amendment law, which is how to balance conflicting and likely irreconcilable imperatives, such as retaining a historical model in light of social, cultural, and technological change, or limiting governmental discretion or power in a time of increased interest in preventative policing and searching.

This entry covers only that part of Fourth Amendment law that applies to governmental agents engaging in search and seizure activity for traditional criminal law enforcement purposes. For ease of reference, such agents will be referred to as "law enforcement" or simply as "police."

Main Text

Fourth Amendment law as we know it today did not exist for the first century of our nation's existence, despite that search and seizure issues have been at the forefront of public concerns since colonial times. It was not until *Ex Parte Jackson*, 96 U.S. 727 (1877), that the Supreme Court considered the Fourth Amendment in a substantive (albeit brief) way, and the Court did not significantly discuss it until *Boyd v. United States*, 116 U.S. 616 (1886). During the preceding period, search and seizure issues were litigated through common law trespass claims, and partly for this reason, the Fourth Amendment came to be viewed through a property law lens.

As for the scope of its protection, for most of our nation's history, the Fourth Amendment bound only the federal government. This is because the Bill of Rights, of which the Fourth Amendment is a part, was originally understood to limit only federal power. During this period, the Fourth Amendment was technically inapplicable to non-federal actors, such as state or local police, whose actions were constrained only by state and local law. The reason why the Fourth Amendment applies to state and local police today is that the Supreme Court, in the mid-1900s, ruled that the Fourteenth Amendment's Due Process clause made the Fourth Amendment binding against the states (*Mapp v. Ohio*, 367 U.S. 643, 1961; *Wolf v. Colorado*, 338 U.S. 25, 1949).

The Fourth Amendment did not become an active subject of Supreme Court litigation until Prohibition, which had resulted in a large expansion in both federal law enforcement and its search and seizure efforts. Even then, the property-based focus persisted, despite the pressure that technological change was exerting on the jurisprudence. In *Olmstead v. United States*, 277 U.S. 438 (1928), for example, the Court rejected a Fourth Amendment challenge to the wiretapping of telephone conversations, emphasizing that the wiretapping was accomplished "without trespass upon any property of the defendants."

As time and technology advanced, however, concerns about the government's investigatory and surveillance powers continued to increase, to the point where the Supreme Court finally expressed dissatisfaction with the property model and introduced a privacy model in *Katz v. United States*, 389 U.S. 347 (1967), which held that electronic eavesdropping from the exterior of a telephone booth constituted a search subject to Fourth Amendment protections. Since then, *Katz* and its privacy model have served as the touchstones for all Fourth Amendment law. This dominance, however, may be ending.

Recently, in *United States v. Jones*, 132 S. Ct. 945 (2012), the Court clarified that both a privacy model and a property-centric trespass model are available bases upon which to claim Fourth Amendment rights.

In addition to the privacy/property dichotomy, the other debate that has strongly influenced Fourth Amendment jurisprudence concerns the relationship between its two clauses, the Reasonableness and Warrant Clauses. Through the 1960s and into the 1970s, the Court often engaged in Warrant Clause-primacy, in which reasonableness under the opening clause was determined in light of whether a valid warrant had been obtained. This approach underlies the ubiquitous formulation that searches and seizures are per se unreasonable when unsupported by a valid warrant, except in certain carefully limited exceptions. Starting in the 1980s, this approach has tended to give way to Reasonableness Clause-primacy, in which the ultimate touchstone of Fourth Amendment constitutionality is whether the governmental conduct at issue was reasonable. Under this view, the ultimate question of reasonableness need not, and indeed often is not, determined in light of whether law enforcement had obtained a valid warrant, and the Warrant Clause's function is simply to set forth the constitutional minima for obtaining a warrant, while saying nothing about when warrants are actually required. Increasingly, the Supreme Court has settled into the pattern of generally judging the constitutionality of criminal searches under the Warrant Clause, and that of civil searches by reference to the Reasonableness Clause.

Fourth Amendment jurisprudence in general, and that part which governs police searches specifically, is often criticized for lacking coherence. A great deal of that criticism stems from disagreement among Supreme Court justices about whether to apply a privacy or property model, whether to proceed from Reasonableness Clause primacy or Warrant Clause primacy, and from disputes about how to do so once those choices are made. One evident dynamic is that Warrant Clause-primacy has often resulted in increasing Fourth Amendment protections from governmental searches – and concomitantly greater restrictions on police search power – both on a substantive basis and because it results in clearer standards against which to gain suppression of evidence. A challenge to moving toward a Reasonableness Clause approach is assuring a sufficiently high threshold of protections against governmental searches and seizures and assuring sufficient coherence and predictability in the jurisprudence so that it amounts to more than merely the subjective reaction of the judicial officer reviewing a given Fourth Amendment issue.

Fundamentals

The Fourth Amendment's text explicitly mentions both searches and seizures. But the Fourth Amendment does not cover all searches and seizures, only those performed by governmental actors – such as all police searches – or by private actors but in mere acquiescence to governmental authority. Thus, for example, a spouse or landlord acting in a voluntary, private capacity is free to engage in search and seizure behavior that would violate the Fourth Amendment had it been engaged in by police and then share the results with police. That person's conduct might violate some other body of law, such as tort law, but it will not amount to a Fourth Amendment violation because it is private, not governmental, conduct.

The next issue in any Fourth Amendment challenge to police conduct is whether it amounts to a "search" or "seizure" covered by the Fourth Amendment. There are two search tests. *Katz* is ubiquitously understood as having established a two-factor privacy test for a search: a Fourth Amendment search occurs only if the target had (1) an actual, subjective privacy expectation and (2) an objective privacy expectation that society accepts as reasonable. Recently, *Jones* engaged in a historical analysis to clarify that an alternate, property-based search test continues to exist as well: a Fourth Amendment search occurs when (1) government physically trespasses (2) for the purpose of collecting information (3) upon an enumerated Fourth Amendment item (i.e., "persons, houses, papers, and effects"). Either *Katz*'s privacy-based search test or *Jones*'s property-based one or both may be invoked when seeking

Fourth Amendment protection. (When property is involved, the seizure standard is whether government has meaningfully interfered with a possessory interest (*Soldal v. Cook County*, 506 U.S. 56, 1992). Numerous seizure standards exist with regard to the detention of persons, e.g., *United States v. Drayton*, 536 U.S. 194 (2002); *United States v. Mendenhall*, 446 U.S. 544 (1980); *Terry v. Ohio*, 392 U.S. 1 (1968). Because this entry is devoted to police searches, it will not consider seizures in any detail.)

Once it has been established that a Fourth Amendment "search" has occurred, the next question to be determined is what conditions the Fourth Amendment imposes for the search to be constitutional. It is at this stage that a great deal of complexity and ambiguity exists in Fourth Amendment jurisprudence.

Many police searches are assessed under the Warrant Clause. Consequently, often, the next inquiry is whether a valid search warrant was obtained and, if not, whether a valid exception to a warrant requirement existed. The Fourth Amendment's text imposes only three requirements for a valid warrant: (1) probable cause, (2) particularity, and (3) oath or affirmation. Of these, most commonly disputed is whether probable cause existed, though such challenges have been less likely to succeed since *Illinois v. Gates*, 462 U.S. 213 (1983), because it imposed a practical, commonsense, nontechnical totality-of-the-circumstances test. To a lesser extent, particularity is also sometimes disputed. Though not explicitly specified in the Fourth Amendment, the Supreme Court has also imposed a judicial preclearance requirement for warrants. It mandates that police, when applying for a search warrant, explain the factual basis upon which probable cause is claimed to exist such that a neutral and detached magistrate can make an independent determination of whether it, as well as the particularity requirement, is satisfied prior to issuing the warrant (*Lo-Ji Sales, Inc. v. New York*, 442 U.S. 319, 1979; *Giordenello v. United States*, 357 U.S. 480, 1958; *Nathanson v. United States*, 290 U.S. 41, 1933; *Byars v. United States*, 273 U.S. 28, 1927). Search warrants can also be subject to other, nonconstitutional requirements, like those in Federal Rule of Criminal Procedure 41, or in relevant statutes such as those that apply to certain forms of electronic surveillance.

Though it sounds odd, Warrant Clause primacy need not result in imposition of a warrant requirement, strict application of which is incompatible with the realities of quotidian police duties. Often, police are called upon to respond to quickly developing situations, and they could not perform their jobs and adequately protect public safety if required to obtain a warrant prior to intervening. Traffic stops, for example, could not possibly comply with a warrant requirement. As a result, the presumptive warrant requirement quickly began developing exceptions. Though these are claimed to be "carefully delineated," many commentators and at least one Supreme Court justice believe that the list of exceptions is now so broad as to have swallowed the rule (*California v. Acevedo*, 500 U.S. 565, 1991; Scalia, J., concurring).

Of such exceptions, an important one is that seizures can be made based upon only probable cause, with the scope of the seizure (e.g., its duration) being limited by the nature of the probable cause. Traffic stops provide an example. This probable cause standard is considered to be consistent with Warrant Clause primacy given that "probable cause" appears in that clause. Because no warrant is required, this probable cause standard is not subject to judicial preclearance. A probable cause challenge can be raised after the search, however.

Police searches can alternately be analyzed under the Reasonableness Clause. A prime example is police stop-and-frisk activity, which the Supreme Court allowed in *Terry* based upon reasonable suspicion, which is a lower suspicion threshold than probable cause. The move away from the Warrant Clause in favor of the Reasonableness Clause resulted in greater flexibility, allowing the Court to evade an otherwise applicable probable cause requirement. Stops and frisks are comprised of an initial seizure and search encounter (the stop) and may be followed by a second, more intrusive search (the frisk). The stop requires reasonable suspicion that a crime may be afoot and justifies police in detaining (seizing) an individual and engaging in investigatory questioning (a limited search). Frisk authority, because it is more intrusive, does not automatically follow from the stop, but instead requires reasonable suspicion not just of a crime but also of danger. The greater discretion that stop-and-frisk authority provides police has often caused it to be highly controversial and has contributed to claims of abused police authority and to distrust of police, particularly in low-income or minority communities.

There are many circumstances in which police may search based only upon probable cause or reasonable suspicion, and sometimes, no suspicion at all. Many of these circumstances can be grouped under the umbrella category of exigency, and they include hot pursuit, the automobile exception (which allows automobiles to be searched based upon only probable cause), and the search-incident-to-arrest doctrine (which, after an arrest, allows an immediate search of the arrestee as well as the "grabbing area" around him).

Two doctrines that allow police to engage in suspicionless searches, but do not fall under the exigency rubric, are plain view searches and consent-based searches. Plain view searches comply with the Fourth Amendment when police are lawfully in the place from which they observe an item whose incriminating character is immediately apparent. Thus, due to the plain view doctrine, a valid traffic stop can result in a much more serious criminal prosecution if police see illegal drugs in open view inside the vehicle. Consent-based searches are discussed in greater detail below.

Key Issues and Controversies

Role of History: Limited vs. Big Government, Urbanism, and Prohibition

Though the Supreme Court has increasingly called for giving greater weight to history in Fourth Amendment analysis, the utility of history as an analytical tool is actually quite limited because the modern world bears little resemblance to the world that existed at the nation's founding, and consequently, many policy imperatives have changed over time, sometimes completely reversing course. For example, the historical record is quite clear that, at nationhood, the primary Fourth Amendment goal was to constrain governmental search discretion. By contrast, Fourth Amendment jurisprudence today is designed to extend tremendous discretion to governmental searchers. Another reason why history is of limited utility is that, although widespread agreement exists about many of the relevant historical data points, radically different opinions exist about the implications of that history.

When considering history, Anglo-American history naturally dominates. At the nation's founding, constraining official discretion was an understandable goal given colonial history, which included complaints and, especially, extravagant rhetorical warnings that royal agents were infringing or would infringe upon colonial rights through an overbearing search power. This search power was exercised in the customs context to enforce tax laws. An attempt to extend it to excise taxes resulted in the 1765 Stamp Act riots. The Fourth Amendment is generally thought to include the Warrant Clause to constrain search discretion by establishing demanding standards under which search warrants can issue, and thus constitutionalize a ban against "writs of assistance" (official documents that were used to enforce customs laws) and general warrants, both of which provided unconstrained discretion to the search official.

In the years after nationhood, the federal government enacted customs laws that essentially parroted the British ones that had applied to the colonies – and which colonists had protested – including with regard to search and seizure provisions. The British statutes had constrained discretion by extending immunity to the searchers if, after the search, it was determined that probable cause had existed. Federal revenue laws used probable cause in the same way.

Those aggrieved by searches had recourse through a common law trespass action or, if federal revenue laws were at issue, possibly also through a civil law forfeiture action. However, if a federal revenue law was at issue, the probable cause determination was usually taken away from a jury and reserved for a federal judge. Searches that had occurred for purposes other than enforcing federal revenue laws, such as those that we think of today as classic police searches – searches for stolen goods, for example – were also subject to a legal regime that sought to constrain discretion through a combination of warrant procedures and liability for unsuccessful searches. This system, however, bears little resemblance to Fourth Amendment jurisprudence today because in our early history there was no professionalized police force. Thus, investigatory and search authority was left to private parties, who could seek a search warrant and assistance from a constable or sheriff in executing the warrant. Such governmental officials who acted under authority of an ostensibly valid search warrant were immunized from damages. The private party who swore out the warrant, however, would be liable if the search was unsuccessful, which acted to constrain search discretion.

A good example of the disjunction between modern and historical approaches to Fourth Amendment law concerns probable cause. Today, probable cause is often an important mechanism for protecting the public from improper police searches. This is certainly true as to warrant-based searches and is often true with regard to warrantless searches, which often are authorized merely upon probable cause. (One important difference in how probable cause is regulated in these two situations is that, in the former, probable cause can be assessed twice, once through judicial preclearance when the warrant is applied for and again after the search during a suppression or damages proceeding, but in the latter, it is assessed only after the fact.)

But much evidence indicates that during the nation's early history probable cause may have been a much less protective device. Though explicitly mentioned in the Fourth Amendment, when it was included there it was an immature concept that legal treatises indicated could be satisfied under conditions that would never be acceptable today, such as being a nightwalker or living an idle, vagrant, and disorderly life. Moreover, evidence suggests that probable cause played an especially important role only during federal revenue enforcement actions, where statutes used it as an immunity factor. Probable cause may have been a much less protective concept under the common law, for several reasons. It is accepted today that judges have a well-established duty to independently assess the adequacy of probable cause before issuing search warrants to police. But considerable historical evidence indicates that in our nation's early history the validity of probable cause supporting a warrant often would have been subject only to an after-the-fact jury determination through a common law trespass action. Judicial officers who issued search warrants may often have perceived their role with regard to probable cause as being merely ministerial, such that they may have treated an assertion of probable cause as a mere pleading requirement, the validity of which a jury could later decide. Juries, however, could so decide only if a judicial challenge was brought – which was rare then, as now – and even then could have their opinion nullified by a federal judge if a federal revenue law was at issue.

Using history as a guide to current Fourth Amendment law is a challenge because so much has changed over time. The conception and importance of probable cause, and the role of the jury, are just several examples. As time has passed, the original guiding imperative of limiting search discretion has given way to a desire to greatly expand search discretion. The reasons for this switch are numerous and include the move from rural to urban populations, which increased anonymity and hence crime and resulted in the creation of modern professionalized police forces, as well as persistent advances in technology that have vastly magnified both the harm one individual can inflict and the interest in, and capabilities of, surveillance, which increase the pressure to engage in preventative searches. For example, recent research persuasively argues that the key event in transforming the Fourth Amendment law of police searches to what we have today was the advent of the temperance movement of the late 1800s, which culminated in Prohibition and a significant increase in federal law enforcement (Oliver 2011). Indeed, several Prohibition cases, such as *Carroll v. United*

States, 267 U.S. 132 (1925), which involved automobile searches, and *Olmstead*, which involved wiretapping, resulted in important Fourth Amendment decisions (though later *Katz* famously overruled *Olmstead*). These changing circumstances have resulted in a desire to assure adequate governmental power to protect us from third parties, and thus the law of police searches has moved dramatically in the direction of expanding governmental discretion.

Consent, Police Intent, and Discrimination

The law of police searches with regard to consent and police intent operates to greatly expand police discretion. The premises underlying consent searches are that police should not operate at any disability as compared with a private actor and thus may seek voluntary assistance and compliance as could any private actor, and that the target has full knowledge about whether the encounter is truly voluntary or is subject to governmental compulsion (*Kentucky v. King*, 131 S. Ct. 1849, 2011). The latter presumption often operates to make consent-based searches controversial. For example, often there is no way for a person stopped by police on a sidewalk or in an airport or bus terminal to know whether police are acting under probable cause or reasonable suspicion, which would authorize a compelled detention. Similarly, every traffic stop at some point exhausts the probable cause justification for the seizure, but that point is often inscrutable to the driver. Thus, when a person is approached by police and asked, “Do you mind if I ask you a few questions?” or a driver is asked “By the way, do you mind if I look around?” it is often unclear whether a truly voluntary choice exists, particularly given that many jurisdictions make it a criminal offense to cut short a justified police encounter, as in the case of traffic stops. The controversy exists in large measure because the Supreme Court has ruled that consent need only be voluntary, not knowing, and thus police need not inform the target that he or she has a right to refuse consent (*Schneckloth v. Bustamonte*, 412 U.S. 218, 1973) or a driver that he or she is free to go instead of granting consent (*Ohio v. Robinette*, 519 U.S. 33, 1996).

Adding further controversy to the law of police searches is that, generally, only objective police intent is relevant to Fourth Amendment constitutionality, while subjective police intent is immaterial. Consequently, Fourth Amendment law makes it very difficult to prevail on claims of either invidious discrimination in police searches or on pretext claims such as that a minor traffic stop was really a subterfuge for forcing a police interaction (*Whren v. United States*, 517 U.S. 806, 1996). The Supreme Court justifies the focus upon objective intent based on ease of judicial application, a desire to treat similar police actions similarly regardless of whether different subjective motivations animated them, and the difficulty of establishing subjective intent.

Fourth Amendment consent and intent doctrines are often the source of great tension between police and poor or minority communities because these doctrines can effectively immunize police behavior motivated by actual invidious discrimination. For example, they can make it very difficult to challenge racial profiling, despite that it has been documented in traffic stops. Whether the Fourth Amendment law of police searches is properly formulated on these points is a significant issue because, by far, the greatest amount of face-to-face interaction between police and the public is during traffic stops. Thus, these Fourth Amendment doctrines are often highly criticized.

Arrest Authority and Strip Searches

Yet an additional example of the large amounts of deference extended to police under modern Fourth Amendment law is seen in the low levels of protection extended against dubious arrests and their ensuing consequences such as strip searches at detention centers. The Supreme Court has rejected a mother’s claim that the Fourth Amendment protected her from being arrested for the minor criminal offense of failing to secure herself and her two children in her pickup truck, despite acknowledging that “the physical incidents of arrest were merely gratuitous humiliations imposed by a police officer who was (at best) exercising extremely poor judgment" (*Atwater v. City of Lago Vista*, 532 U. S. 318, 2001). A very persuasive case has been made that *Atwater* is completely at odds with a historical view of the Fourth Amendment (Davies 2002). The Court has also refused to provide Fourth Amendment protections from arrests that violate state law, reasoning that Fourth Amendment and state standards are distinct and that the Fourth Amendment is satisfied so long as probable cause supported the arrest (*Virginia v. Moore*, 553 U.S. 164, 2008).

Arrestees are taken to detention centers, and recently, the Supreme Court has ruled that, once there, all adult arrestees may be subjected to suspicionless strip searches, including visual body cavity searches, and that the Fourth Amendment provides no protection even if the arrest was wrongful, such as because it was based upon the erroneous belief that an arrest warrant existed (*Florence v. Board of Chosen Freeholders*, 132 S. Ct. 1510, 2012). The degree of deference that the Supreme Court extended to detention centers was striking because the Court could easily have endorsed a reasonable suspicion requirement for such strip searches given that a prior case, *Bell v. Wolfish*, 441 U.S. 520 (1979), had widely been interpreted as imposing such a suspicion threshold before arrestees – who, after all, have yet to be adjudicated guilty of any offense – could be strip searched. Substantial uncertainly exists as to whether similar institutional deference will be extended in the context of suspicionless strip searches of juvenile arrestees given that the Court has required reasonable suspicion to justify strip searches in high schools, as well as proportionality between a search's intrusiveness and the suspected harm being investigated (*Safford Unified School Dist. No. 1 v. Redding*, 557 U.S. 364, 2009).

Surveillance and Technology

Fourth Amendment law has consistently struggled with whether, and how, to adapt to technological change. This dynamic became evident in the early 1900s as technology started rapidly advancing, which, for example, made wiretapping possible. Today, this is a consistently pressing issue given the accelerating pace of technological change, which has contributed to an increased interest in preventative searches as a result of heightened security concerns in a post-9/11 world.

The Supreme Court usually has taken a careful approach to technological development, often refusing to recognize Fourth Amendment privacy barriers to its use, though infrequently the Court has intervened even at the risk of dramatically changing Fourth Amendment law. One prominent distinction the Court has applied in Fourth Amendment technology cases is whether the technology merely involves sense enhancement or augmentation, on the one hand, or sense replacement or the use of what the Court believes is a disruptive technology, on the other.

Two of the Supreme Court's most important Fourth Amendment rulings fell into the latter category, which shows what an important role technology has had in developing the jurisprudence. One was *Katz*, in which the Court adopted a privacy model, and rejected a property model, in holding that the public enjoyed Fourth Amendment protections against electronic eavesdropping. The second was *Jones*, a recent decision in which the Supreme Court unanimously ruled that police installation and monitoring of a GPS device upon a suspect's vehicle was subject to Fourth Amendment oversight. *Jones* is significant not just for its reaffirmation that a property-based trespass model exists as a basis to seek Fourth Amendment protections (in addition to *Katz*'s privacy model) but also for what the concurring opinions suggest about the future direction of Fourth Amendment law in general and the law of police searches in particular.

Justice Alito wrote a concurring opinion, explaining that he would have resolved the case through a *Katz*-ian privacy analysis that would have focused upon objective privacy expectations. In the context of GPS technology, he would have applied two new, revolutionary factors to the objective prong: a temporal factor and an offense-specific factor. With regard to the temporal factor, he would have ruled that no objective privacy expectation exists in the context of "short-term" GPS tracking but that one does usually exist in the context of "longer term" GPS tracking. He provided no guidance as to what separates brief from prolonged GPS tracking other than to say that the GPS tracking at issue, which spanned 28 days, clearly qualified as prolonged. As for the offense-specific factor, he would have ruled that prolonged GPS tracking would violate an objective privacy expectation in the case of "most offenses" but suggested that this might not be true for "extraordinary offenses." His discussion of the offense-specific exception indicates that it is not defined in terms of distinct offense categories. Rather, it is really a resource-intensity measure, which the government can successfully invoke if it can show that it would have approximated similar surveillance results through traditional means had it not used GPS tracking, even if that meant an unusual commitment of resources. Because three other justices joined Justice Alito's concurrence, there are at least four votes on the current Supreme Court for incorporating these two revolutionary factors into *Katz*'s objective privacy expectation analysis.

There also appears to be a fifth vote available because Justice Sotomayor, though she provided a fifth vote for resolving *Jones* on property/trespass grounds, wrote a separate concurrence indicating that she was prepared to support Justice Alito's privacy approach.

Significantly, she also specified that she was willing to go even further than Justice Alito in extending Fourth Amendment protections against technological surveillance, even if it meant fundamentally reforming Fourth Amendment privacy jurisprudence. Police searches often take advantage of three interrelated Fourth Amendment doctrines, all of which are closely tied to *Katz*-ian privacy. One is the third-party doctrine, under which exposure of information to third parties, even in instances where there is little alternative, can obviate any privacy interest and thus eliminate Fourth Amendment protections (*Smith v. Maryland*, 442 U.S. 735, 1979; *United States v. Miller*, 425 U.S. 435, 1976). Another is the assumption of risk doctrine, under which privacy interests in information can be lost when voluntarily shared with a third party because one assumes the risk that the third party will disclose the information (*United States v. Jacobsen*, 466 U.S. 109, 1984; *Frazier v. Cupp*, 394 U.S. 731, 1969). The assumption of risk doctrine is particularly useful to police in the context of informants and co-occupants (*United States v. Matlock*, 415 U.S. 164, 1974; *United States v. White*, 401 U.S. 745, 1971). Finally, some Fourth Amendment law suggests a private-versus-public space distinction, in which no privacy expectation exists as to activities conducted in public, such as movements through public spaces or roads (*United States v. Knotts*, 460 U.S. 276, 1983). *Jones* is important in part because Justice Alito's and Justice Sotomayor's concurring opinions could have transformative implications for these three Fourth Amendment doctrines.

Additionally, some judicial opinions assessing GPS tracking had embraced mosaic theory as a means for the Fourth Amendment to keep up with technology, and Justice Alito and the three justices who joined his opinion seem to endorse it in *Jones*, and Justice Sotomayor clearly embraces it. In a search and seizure context, mosaic theory posits that privacy interests should be protected in a manner that guards against collections of small bits of information that individually may not be particularly revealing but that when aggregated can reveal a great deal. Justice Alito's temporal factor seems to implicitly endorse mosaic theory because a prime differentiating factor between brief and prolonged GPS tracking is the amount of data revealed to police. Justice Sotomayor was more direct in endorsing mosaic theory, so she clearly seems to be one available vote for importing it into Fourth Amendment jurisprudence. If these five justices are indeed willing to make this outcome law, it would significantly alter the law of police searches, at least in the context of technological surveillance.

Legislation, Technology, and National Security

In part because the Supreme Court often has been cautious in extending Fourth Amendment protections in technological contexts, the political branches have played an important role in defining the law of police searches. For example, when the Supreme Court created the third-party doctrine by ruling in *Smith* that there is no Fourth

Amendment privacy interest in telephone numbers dialed from one's home (because the telephone company keeps records of those numbers) and in *Miller* by holding that no privacy interest exists in one's banking records (because the bank holds them), the political branches responded by creating statutory protections through passage of the Pen Register Act, 18 U.S.C. §§ 3121–27 and the Right to Federal Privacy Act, 12 U.S.C. §§ 3400–22. Privacy standards that the political branches have created and defined have also strongly influenced the law of police searches through the Federal Wiretap Act, 18 U.S.C. §§ 2510–22 and the Stored Communications Act, 18 U.S.C. §§ 2701–12. When such statutory protections exist, it is important to understand that they may not necessarily comply with Fourth Amendment standards, especially in contexts where the political branches have taken the lead. For example, currently there are numerous controversies concerning the terms under which the Stored Communications Act allows the police to access cellular and GPS location data held by third-party service providers and whether those statutory terms meet Fourth Amendment standards.

Some of the uncertainty exists because the law of police searches has been significantly impacted by national security concerns. After the September 11, 2001 terrorism attacks, for example, Congress statutorily authorized the Attorney General, through one portion of the USA PATRIOT ACT, 50 U.S.C. § 1843, to use pen registers without prior court approval in emergency circumstances relating to international terrorism and foreign intelligence. There is evidence that governmental surveillance powers that were expanded post-9/11 for national security purposes have been used in the domestic law enforcement arena.

Remedies: The Exclusionary Rule, the Right to Counsel, and Qualified Immunity

An important issue in Fourth Amendment law is what remedies, if any, exist for a violation of the rights conferred against governmental searches and seizures. The primary question concerns the legitimacy of the exclusionary rule: whether exclusion of wrongfully obtained evidence should be an available remedy. Textually, the Fourth Amendment identifies rights but is silent about remedies. Early on, the Supreme Court in *Boyd* at least presumed that exclusion was proper, and then formally adopted exclusion as a remedy for Fourth Amendment search and seizure violations by the federal government in *Weeks v. United States*, 232 U.S. 383 (1914). Thirty-five years later, in *Wolf*, the Court declined to impose exclusion as a Fourth Amendment remedy against violations by state actors, on the theory that the device by which the Fourth Amendment was incorporated against the states – the Fourteenth Amendment – did not require it, but the Court switched course 12 years later in *Mapp*, in part on the ground that experience had shown that exclusion was the only effective remedy.

The exclusionary rule is controversial because, as famously phrased, "[t]he criminal is to go free because the constable has blundered," *People v. Defore*, 150 N.E. 585 (N.Y. 1926), and as a result, bare majorities of the Supreme Court have been on an active campaign for decades to minimize or eliminate it, as in *United States v. Leon*, 468 U.S. 897 (1984) (adopting good-faith exception for reasonable police reliance on search warrant subsequently held invalid), and this campaign has accelerated in the last decade. As originally conceived, the exclusionary rule was justified on the twin bases of judicial integrity and deterrence value. The campaign against it has rejected the judicial integrity rationale and has used the only remaining justification – deterrence – to require a cost-benefit analysis before wrongfully obtained evidence can be suppressed, while both emphasizing its costs and minimizing its benefits (*Herring v. United States*, 555 U.S. 135, 2009; *Hudson v. Michigan*, 547 U.S. 586, 2006).

A major debate surrounding the exclusionary rule is whether, as *Mapp* conceived, it is the only effective remedy for Fourth Amendment violations. Opponents of the exclusionary rule identify numerous intervening reforms since *Mapp* which, they assert, mean that the rule is no longer needed. They point to the increased professionalism of police forces and to federal civil rights statutes that authorize civil suits and the availability of attorney's fees for prevailing plaintiffs, as well as to increases in the number of public-interest law firms and civil rights lawyers (*Hudson*). Supporters of the exclusionary rule point to the increased professionalism and improved training of police forces as evidence of both the success and necessity of exclusion as a remedy, *Herring* (Ginsburg, J., dissenting), and assert that alternate remedies are not sufficiently available and successful.

In this debate, insufficient attention has been paid to the role of assigned counsel and to the impact of the qualified immunity defense. The Sixth Amendment provides a right to counsel in criminal cases involving a felony or before a sentence of incarceration may be imposed. This provision of counsel has played a singular role in the development of the law of police searches because it has guaranteed the presence of a legal professional to assess each client's situation and to vigorously and aggressively litigate potential Fourth Amendment claims. If, however, the exclusionary rule is eliminated or so emasculated that there is no incentive to litigate such claims during a criminal trial, the development of Fourth Amendment law will greatly suffer. This is because alternative remedies to the exclusionary rule often will be more theoretical than real for several reasons. The number of public-interest law firms and civil rights attorneys available to bring such claims pales by comparison to the number of assigned counsel in criminal cases. An aggrieved Fourth Amendment claimant who cannot enlist a public-interest law firm is left with no other choice but to hire a private attorney, but few such claimants can afford such legal representation. This representation deficit is further exacerbated by the lack of meaningful damages that would be available in most civil cases – particularly where the client was held guilty in the criminal proceeding – and by the difficulty in collecting statutory attorney fees in civil rights cases, a difficulty that has been magnified through rulings spearheaded by many of the same Supreme Court justices who are hostile to the exclusionary rule and point to such attorney fees as an alternative.

The qualified immunity defense also has a significant role in limiting the effectiveness of remedies other than the exclusionary rule. The qualified immunity defense protects police from civil suits for damages when a constitutional right has been violated but was not clearly established at the time of the violation. The law of qualified immunity is often extremely generous to police – again often through efforts by the same group of Supreme Court justices who oppose the exclusionary rule – which often makes it difficult to win on claims that a constitutional right was clearly established. This can be particularly ironic because, though the Court has increasingly insisted that historical analysis should play a leading role in Fourth Amendment law, the Court has at the same time rejected the relevance of historical Fourth Amendment analysis when it would operate to defeat a qualified immunity defense (*Anderson v. Creighton*, 483 U.S. 635, 1987). This rejection is significant because the common law of search and seizure was often quick to impose liability upon a government official who acted improperly.

Conclusion

Fourth Amendment law in general, and as it applies to police searches in particular, is at a crossroads. Though recently the Supreme Court has given more emphasis to historical analysis, in reality the Court's record in aligning modern Fourth Amendment law with the historical law of search and seizure is poor.

This outcome was foreordained because circumstances have simply changed too dramatically for a historical model to dominate. A good example of why can be seen in the advance of technology, which forces a Fourth Amendment reckoning absent any historical analogue. Moreover, technology also puts pressure on the core Fourth Amendment protective concept upon which we have relied – prior suspicion. Prior suspicion has come to occupy pride of place in Fourth Amendment law because of history. Historically, suspicion was the primary concept used to restrict search discretion, which was the overarching goal for which the Fourth Amendment was adopted. For more than a century, suspicion fulfilled this function well. At the nation's founding, the world was simpler and in that context suspicion was a particularly useful mechanism for limiting governmental power and discretion. As time advanced, however, circumstances changed and the world became increasingly complex and interconnected – for example, life in the United States moved from an agrarian to an urban existence, and technological advances accelerated – and consequently suspicion became less protective. This dynamic can be seen in the evolution of Fourth Amendment law, which started from a presumptive warrant requirement, then began departing from that requirement and instead relied more often only upon probable cause, and then in *Terry* created an exception to even probable cause through the creation of a less demanding reasonable suspicion model. Technological advances continue to put pressure upon suspicion because they increase the ability to engage, and interest in engaging, in preventative searches, such as through various forms of electronic surveillance.

Therefore, a significant challenge that currently exists in Fourth Amendment law, including with regard to police searches, is how to assure adequate search and seizure protections when suspicion can no longer play the protective role that it once did. This crisis comes with a benefit – the opportunity to expand and reassess our thinking about the Fourth Amendment – but it is one that must be embraced and pursued. Doing so requires the recognition that suspicion is not alone sufficient to assure a search's constitutionality because the Fourth Amendment protects a host of values, and suspicion protects only some of them. A search can be unconstitutional under the Fourth Amendment for many reasons. It might have been overly invasive or intrusive, the governmental interest being pursued might not have been sufficiently compelling, the search target's privacy interests could have been so substantial as to render the search illegitimate, or the search may not have been sufficiently efficacious or had sufficient deterrence value to justify the measures employed. Such interests are ones that the Fourth Amendment protects, but that suspicion does not. Thus, one avenue that might beneficially be pursued is to reform the Fourth Amendment law of police searches to more precisely and distinctly identify and consider the varied and multiple interests that the Fourth Amendment protects in various circumstances and ask whether the police conduct adequately respected them given the circumstances.

Related Entries

- ▶ British Police
- ▶ Civil Remedies
- ▶ Communities and the Police
- ▶ Comparing Police Systems Across the World
- ▶ Control of Police Misconduct
- ▶ Democratic Policing
- ▶ Early Modern Police and Policing
- ▶ History of Criminal Investigation
- ▶ History of Technology in Policing
- ▶ History of the Police Profession
- ▶ Hot Spots and Place-Based Policing
- ▶ Information Technology and Police Work
- ▶ Intelligence-Led Policing
- ▶ Law of Community Policing and Public Order Policing
- ▶ Law of Police Seizures and the Exercise of Discretion
- ▶ Law of Undercover Policing
- ▶ Legal Frameworks for Third-Party Policing
- ▶ National and Local Policing
- ▶ Police Culture
- ▶ Police Discretion and Its Control
- ▶ Police Discretion in Law Enforcement
- ▶ Police Integrity
- ▶ Police Legitimacy and Police Encounters
- ▶ Police Lying and Deception
- ▶ Police Officers in Schools
- ▶ Policing Terrorism and Legitimacy
- ▶ Policing the Police
- ▶ Preventing of Terror at Shopping Malls
- ▶ Role and Function of the Police
- ▶ School Security Practices and Crime
- ▶ Strategies of Policing Terrorism
- ▶ Surveillance Technology and Policing
- ▶ Theories on Policing and Communities
- ▶ Traffic Police

Recommended Reading and References

Amar AR (1994) Fourth amendment first principles. Harv Law Rev 107:757–819

Arcila, Jr. F (2010) The death of suspicion. William & Mary Law Rev 51:1275–1341

Clancy TK (1995) The role of individualized suspicion in assessing the reasonableness of searches and seizures. Univ Memphis Law Rev 25:483–635

Cuddihy WJ (2009) The fourth amendment: origins and original meaning 602–1791. Oxford University Press, New York

Davies TY (1999) Recovering the original fourth amendment. Mich Law Rev 98:547–750

Davies TY (2002) The fictional character of law-and-order originalism: a case study of the distortions and evasions of framing-era arrest doctrine in *Atwater v. Lago Vista*. Wake Forest Law Rev 37:239–437

Dickerson OM (1939) Writs of assistance as a cause of the revolution. In: Morris RB (ed) The era of the American revolution: studies inscribed to Evarts Boutell Greene. Columbia University Press, New York, pp 40–75

Lasson NB (1937) The history and development of the fourth amendment to the United States Constitution. Johns Hopkins Press, Baltimore

Maclin T (1997) The complexity of the fourth amendment: a historical review. Boston Univ Law Rev 77:925–974

Oliver WMN (2011) The modern history of probable cause. Tennessee Law Rev 78:377–429

Sklansky DA (2000) The fourth amendment and common law. Columbia Law Rev 100:1739–1814

Taslitz AE (2006) Reconstructing the fourth amendment: a history of search and seizure, 1789–1869. New York University Press, New York

Thomas GC III (2005) Time travel, hovercrafts, and the framers: James Madison sees the future and rewrites the fourth amendment. Notre Dame Law Rev 80:1451–1518

Law of Police Seizures and the Exercise of Discretion

I. Bennett Capers[1] and Alafair S. Burke[2]
[1]Brooklyn Law School, Brooklyn, NY, USA
[2]Maurice A. Deane School of Law,
Hofstra University, Hempstead, NY, USA

Overview

Our society grants the police extraordinary powers. This is not to suggest those powers are limitless. While many police departments have internal rules and guidelines that limit their behavior with respect to the seizure of persons or property, the most important limitations originate in the Fourth Amendment or more specifically in Supreme Court cases explicating the Fourth Amendment. The Fourth Amendment provides in its entirety, "The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized." From these 54 words, a rich body of case law has developed governing searches and seizures. In addition, the Supreme Court has developed a rule of exclusion that generally requires the exclusion of evidence obtained in violation of the Fourth Amendment. The law of searches and the exclusionary rule are both beyond the scope of this entry, which focuses on the law of seizures and the exercise of discretion.

By its terms, the Fourth Amendment governs two types of seizures: the seizure of things (or more precisely, tangible property) and the seizure of persons. The law of each is different, and each is addressed in turn.

Fundamentals and Key Issues

The Seizure of Tangible Property

The first task in understanding constitutional limitations with respect to the seizure of tangible property is to understand what constitutes a seizure. The Supreme Court has stated that a seizure occurs "when there is some meaningful interference with an individual's possessory interests in that property." In fact, this is a rather common sense approach to defining seizures. As such, a seizure, within the meaning of the Fourth Amendment, occurs when a police officer takes possession of someone's property, for example, a purse or watch. Likewise, a seizure occurs when the police exercise control over property such that it is no longer available to a person.

For example, when a police officer conducts a traffic stop for a traffic violation, a seizure of the vehicle has occurred during the period that the driver or other occupants are prohibited from using the vehicle. The key is that there must be more than only minor interference over the tangible property. Accordingly, moving a sheet of paper a few inches because it is obscuring the serial number of a television does not constitute a seizure of the paper. Similarly, photographing the exterior of a house from a public sidewalk would not constitute a seizure of the house.

However, caution must be exercised in distinguishing between the seizure and search implications of police action. In the examples described above, the movement of the sheet of paper does not so meaningfully interfere with possessory rights to form a seizure of the paper; however, the conduct would constitute a search because it exposes the otherwise obscured serial number to public view. Similarly, photographing the exterior of a house is not a seizure of the house, but "photographing" the interior of the home using heat-detecting thermal technology would constitute a search within the meaning of the Fourth Amendment.

If police conduct amounts to a seizure of property, the second task is to determine whether the seizure is permissible. The Supreme Court has held that police may initiate brief, temporary seizures of property as long as they are reasonable. For example, if the police suspect – but do not yet have probable cause to show – that a suitcase contains narcotics, they can interfere with the owner's possessory rights for a very short period, only so long as necessary to bring a drug-sniffing dog to the scene.

In general, however, the Supreme Court has required that all police seizures of property be supported by probable cause to believe that the item seized will aid in a particular apprehension or conviction. Although the preference is for the determination of probable cause to be made by a neutral and detached magistrate through the issuance of a warrant, in recent years the Supreme Court has relied on the Fourth Amendment's prohibition against "unreasonable" searches and seizures to excuse the warrant requirement when a search or seizure without a warrant is reasonable. In general, when it is practical to secure a warrant, a warrant will be required. When it is not practical, the warrant requirement may be excused. A brief example illustrates this distinction. If the police have probable cause to believe that evidence of an identity theft operation will be found in someone's home, a warrant will normally be required to search and seize the evidence. By contrast, if the police have probable cause to believe that evidence of an identity theft operation will be found in someone's home *and* they have reason to believe that the evidence may be destroyed or relocated before they can secure a warrant, a search and seizure will be permitted without a warrant under the exigency exception. Because of the mobility of automobiles, this exception will often permit a search and seizure of items in an automobile without a warrant, so long as there is probable cause to believe the automobile contains contraband. Similarly, the police may seize, without a warrant, any item that is in "plain view" so long as the officer observes the item from a lawful vantage point, has physical access to the item, and has probable cause to believe the item is contraband or evidence of crime. Consider an example in which the police are summoned to respond to a domestic violence call. If, while responding to the domestic violence call, the police observe illegal drugs in plain view, they may seize it.

Lastly, it should be noted that the warrant requirement, and even the probable cause requirement, may be excused when the government has special needs, other than the detection of crime, justifying a seizure. For example, automobile checkpoints designed to check drivers for evidence of intoxication have been justified under the special needs exception, when the primary goal is to insure public safety on the roads, rather than to make arrests. A similar rationale allows temporary detentions at international borders as well as temporary searches and seizures at airports. Since 9/11, the use of special needs searches and seizures, of both persons and property, has grown exponentially.

The Seizure of Persons: Arrests

The seizure of persons is more complicated. Clearly, an arrest of a person constitutes a seizure. Such a seizure must be supported by probable cause to believe that a crime was committed and that the person to be arrested committed the crime. Beyond the probable cause requirement, the law regarding arrests is fairly straightforward. Absent exigent circumstances or consent, the police must secure an arrest warrant before making an arrest in the arrestee's home. If the police want to enter a third party's home to locate the arrestee, they must go still further and obtain a search warrant for the home. By contrast, the police may arrest a person in a public space with or without a warrant, so long as the probable cause requirement is satisfied. To protect against erroneous arrests, a defendant is entitled to a "prompt" postarrest assessment of probable cause by a magistrate, often referred to as a "probable cause hearing." In *County of Riverside v. McLaughlin*, 500 U.S. 44 (1991), the Supreme Court held that a jurisdiction "that provides judicial determinations of probable cause within 48 h of arrest will, as a general matter, comply with the promptness requirement."

The Seizure of Persons: Terry Stops

A seizure can also occur as a result of activity that falls far short of an arrest. The seminal case permitting seizures short of arrest is *Terry v. Ohio*, 392 U.S. 1 (1968).

In *Terry v. Ohio*, the Supreme Court considered for the first time whether a person could be detained in the absence of probable cause to believe that he had committed a crime, the usual prerequisite for an arrest. On its face, such a seizure would seem to violate the "probable cause" language of the Fourth Amendment. However, weighing the Fourth Amendment in the context of rising crime rates, and placing newfound interest in the Fourth Amendment's reasonableness clause, the Supreme Court interpreted the Fourth Amendment as permitting a limited detention and questioning of a person as long as an officer has specific and articulable facts, that is, reasonable suspicion, to believe that "criminal activity may be afoot." Expressing concern for the safety of officers, the Court went a step further. If the officer also has reasonable suspicion that a person is armed and dangerous, the officer can couple the limited detention and questioning with a pat down for weapons: in common parlance, a stop and frisk.

It should be noted that *Terry v. Ohio* recognized that permitting stops and frisks based on reasonable suspicion would invest officers with a fair amount of discretion. The Court also noted that stop and frisk practices, which the police had already been engaging in for years, were not race neutral and would continue to disproportionately burden minorities. Nonetheless, the Court accepted these disadvantages and interpreted the Fourth Amendment as permitting the practice of forcibly stopping individuals based on "reasonable suspicion."

Several external factors might explain the Court's decision to allow reasonable suspicion as a compromise between barring all stops absent probable cause and ceding complete discretion to the police to engage in stops without judicial oversight. Just 4 months after oral argument, and 2 months before issuing its decision, riots broke out in cities across the nation, including Washington, DC, suggesting that what was needed was more state police power, not more individual rights. In addition, the Court and Chief Justice Warren in particular had been criticized during the 1964 presidential campaign for promoting individual rights at the expense of law enforcement and were expected to be targeted again in the 1968 campaign.

While it is impossible to say with any certainty what role these external events played in *Terry*'s outcome, what is certain is that, by settling for the compromise of reasonable suspicion, *Terry* had the effect of ushering in a shift in direction that, while providing a basis for judicial oversight, would eventually invest officers with almost unfettered discretion to initiate brief, investigatory seizures of individuals.

Because arrests must be supported by probable cause, and stops merely by reasonable suspicion, disputes often arise regarding whether a person's seizure amounts to an arrest or a stop.

Arrests include not only formal arrests but also de facto arrests in which what may have started as a stop takes on the indicia of a custodial arrest. Courts evaluating the level of seizure consider the totality of circumstances surrounding the situation, including the length of the detention, whether the person was moved during the detention, and whether the police seized the person's possessions.

Controversies

What *Terry* has meant in practice, and what *Terry* means to many, is the apprehension of criminals. *Terry* is even preventative. It allows the police to stop and question not only those individuals they suspect of having committed a crime but also those individuals who they suspect may be about to commit a crime.

But the discretion that is inherent in *Terry* stop and frisks remains controversial. The vast majority of individuals stopped and questioned by the police are not engaged in criminal activity nor are they carrying weapons or contraband. That most stops and frisks affect innocent people, however, does not render them unlawful. Put differently, *Terry* authorizes stops and frisks based on mere reasonable suspicion (less than probable cause), regardless of whether those suspicions prove accurate or not.

Furthermore, it is difficult to know what proportion of stops and frisks that affect innocent people are in fact based on suspicions that were, at the time, reasonable. For the vast majority of the law-abiding citizens stopped, questioned, and even frisked by the police before being let go, there is no record of an arrest, no charge number, and no prosecution. Accordingly, those stops and frisks for the most part remain beyond public or judicial scrutiny. By definition, there is no exclusion of evidence, the remedy for a Fourth Amendment violation, because there is nothing to exclude. And because there is nothing to exclude, the affected citizens are unlikely to initiate judicial review of the police conduct, thereby leaving most stops and frisks outside of public scrutiny.

Of equal concern is the malleability of reasonable suspicion. As Justice Thurgood Marshall noted some years after *Terry* was decided, reasonable suspicion is often little more than a "chameleon-like way of adapting to any particular set of observations." This is especially true when it comes to profiling. In *United States v. Sokolow*, 490 U.S. 1 (1989), Justice Marshall offered a string cite of cases in which a suspect matched one of the DEA's profiles to show that almost any behavior can be deemed suspicious and thus satisfy *Terry*'s reasonable suspicion requirement:

> Compare, e.g., United States v. Moore, 675 F.2d 802, 803 (C.A.6 1982) (suspect was first to deplane), with United States v. Mendenhall, 446 U.S. 544, 564 (1980) (last to deplane), with United States v. Buenaventura-Ariza, 615 F.2d 29, 31 (C.A.2 1980) (deplaned from middle); United States v. Sullivan, 625 F.3d 9, 12 (C.A.4 1980) (one-way tickets), with United States v. Craemer, 555 F.2d 594, 595 (C.A.6 1977) (round-trip tickets), with United States v. McCaleb, 552 F.2d 717, 720 (C.A.6 1977) (non-stop flight), with United States v. Sokolow, 808 F.2d 1366, 1370 (C.A.9 1987), vacated, 831 F.2d 1413 (C.A.9 1987) (changed planes); Creamer, supra, at 595 (no luggage), with United States v. Sanford, 658 F.2d 342, 343 (C.A.5 1981) (gym bag), with Sullivan, supra, at 12 (new suitcase); United States v. Smith, 574 F.2d 882, 883 (C.A.6 1978) (traveling alone), with United States v. Fry, 622 F.2d 1218, 1219 (C.A.5 1980) (traveling with companion); United States v. Andrews, 600 F.3d 563, 566 (C.A.6 1979) (acted nervously), with United States v. Himmelwright, 551 F.2d 991, 992 (C.A.5 1977) (acted too calmly).

Moreover, the concern raised in *Terry v. Ohio* that the costs of allowing stops and frisks based on reasonable suspicion would be disproportionately borne by racial minorities appears to have ample merit. Consider recent numbers from New York City, one of the few jurisdictions that require officers to make a record of certain stops and frisks. According to recent data analyzing 867,617 stops over a 2-year period, blacks and Hispanics constituted over 80 % of the individuals stopped, a percentage far greater than their representation in the population. Moreover, of the blacks stopped, 95 % were *not* found to be engaged in activity warranting arrest. When considered as a percentage of the population, the numbers are even more jarring. Stops of whites, if spread across the population of New York City, would amount to stops of approximately 2.6 % of the white population during the period. By contrast, stops of blacks, if spread across the population, would amount to stops of approximately 21.1 % of the population.

The Seizure of Persons: Pretext Stops

Another seminal case in any discussion of the law of police seizures and the exercise of discretion is *Whren v. United States*, 517 U.S. 806 (1996), in which the Court rejected a Fourth Amendment challenge to a pretextual car stop used to search for drugs and other contraband. Justice Scalia, writing for the Court, rejected this argument and concluded that so long as the stop itself was based on an actual traffic violation, the subjective motivation of an officer in singling out a particular motorist, even if race-based, is irrelevant under the Fourth Amendment. Many scholars argue that *Whren* essentially green-lights the police practice of singling out individuals for pretextual traffic stops in the hope of discovering contraband. Given that most drivers routinely violate traffic laws, that is, by slightly exceeding the speed limit, *Whren* in fact gives officers almost unlimited discretion in deciding whom to target for a traffic stop.

Controversies

Like *Terry* before it, *Whren* has had racial consequences. Indeed, the terms "driving while black" and "driving while brown" have become part of the common lexicon to describe the belief that many officers consider race and ethnicity in determining whom to target for a traffic stop. Empirical evidence supports the belief that traffic enforcement disproportionately falls on racial minorities. For example, a report compiled by the Maryland State Police revealed that, during the period examined, African-Americans comprised 72.9 % of all of the drivers that were stopped and searched along a stretch of Interstate 95, even though they comprised only 17.5 % of the drivers violating traffic laws on the road.

Again, numbers like these are only part of the story. The other part is how these numbers impact law-abiding minority citizens. For example, in the Maryland study, even though blacks were disproportionately the subjects of searches, the hit rate for blacks, that is, the rate at which contraband was found, was statistically identical to the hit rates for whites. What this means in numbers is that the vast majority of the individuals stopped and searched were law-abiding minorities not in possession of contraband. A more recent analysis of over 810,000 "field data reports" collected by the Los Angeles Police Department found that the stop rate was 3,400 stops higher per 10,000 residents for blacks than for whites and 350 stops higher for Hispanics than for whites. There is also the issue of the discretionary practices that accompany such stops. The same study found police were 127 % more likely to search stopped blacks than to search stopped whites and 43 % more likely to search stopped Hispanics than stopped whites. Notwithstanding the fact that these groups were searched more often, blacks in fact were 37 % less likely to be found with weapons than searched whites and 24 % less likely to be found with drugs than searched whites. Similar numbers were found for searched Hispanics: Hispanics were 33 % less likely to be found with weapons than searched whites and 34 % less likely to be found with drugs than searched whites.

In addition to race, there is evidence that officers consider economic status in determining whom to target for a stop. For example, there is empirical evidence that police officers are significantly less likely to search vehicles driven by those with above-average incomes than vehicles driven by those with below-average incomes.

The Seizure of Persons: Non-seizures

Finally, any discussion of the law of police seizures would be incomplete without a discussion of a line of Supreme Court cases that categorize many police-citizen encounters as essentially "consensual," mere "encounters" that are not seizures and are thus outside the purview of the Fourth Amendment. In *United States v. Mendenhall*, 446 U.S. 544 (1980), and *Florida v. Bostick*, 501 U.S. 429 (1991), the Supreme Court held that absent a show of force or other circumstances that would lead a reasonable person to believe he or she was not free to leave, a "stop" is not always a "stop" within the meaning of the Fourth Amendment. An officer need not have even reasonable suspicion to initiate an encounter. An officer surprising an individual, asking that individual to accompany the officer, and asking that individual a series of questions – *Where are you going? Do you live nearby? Are you visiting someone here?* – would likely be categorized as a consensual encounter, not a stop within the meaning of the Fourth Amendment. As the Supreme Court has interpreted the Fourth Amendment, "Even when law enforcement officers have no basis for suspecting a particular individual, they may pose questions, ask for identification, and request consent to search luggage – provided they do not induce cooperation by coercive means." The facts in *Mendenhall* are illustrative. In *Mendenhall*, two federal agents approached an African-American woman in an airport, asked to see her identification and airline ticket, and asked her to accompany them to another location for further questioning, where they asked her if she would consent to a strip search. Justice Stewart, proposing the "free to leave" test for the first time, rejected her claim that she was ever "seized" within the meaning of the Fourth Amendment, reasoning that she was always free to leave, and as such the officers needed neither probable cause nor reasonable suspicion to engage her in what could be considered a "consensual encounter." The case of *Immigration and Naturalization Service v. Delgado*, 466 U.S. 210 (1984), is also illustrative. *Delgado* involved a factory sweep to determine the presence of undocumented immigrants. Several INS agents positioned themselves near the buildings' exits, while others dispersed throughout the factory to question most of the workers. The agents "displayed badges, walkie-talkies, and were armed [as they] approached employees and, after identifying themselves, asked from one to three questions relating to their citizenship." Writing for the majority, Justice Rehnquist rejected any claim that the workers were in any way detained and categorized the encounters as entirely consensual and thus outside the purview of the Fourth Amendment.

As with *Terry* and *Whren*, the *Mendenhall* line of cases likely has race and class implications. Notably, all of the "free to leave" cases from the Supreme Court involved racial minorities. Unfortunately, there is little data about the pervasiveness of consensual encounters as a law enforcement tool, precisely because they are outside the purview of the Fourth Amendment. Turning again to New York's stop and frisk data, what this means is that the 867,617 stop and frisk reports tell us virtually nothing of the number of additional individuals who were questioned during supposedly "consensual encounters."

Related Entries

- ▶ Biased Policing
- ▶ Law of Police Searches
- ▶ Police Discretion and Its Control
- ▶ Police Legitimacy and Police Encounters

Recommended Reading and References

American Civil Liberties Union (2007) Analysis of new NYPD stop-and-frisk data. Available on-line at http://www.aclu.org/racialjustice/racialprofiling/33095prs20071126.html. 26 Nov 2007

Ayres I (2008) Racial profiling and the LAPD: a study of racially disparate outcomes in the Los Angeles Police Department. Available at http://www.aclu-sc.org/contents/view/3

Capers IB (2009) Policing, race, and place. Harv Civ Rights-Civ Liberties Law Rev 44:43

Capers IB (2011) Rethinking the fourth amendment: race, citizenship, and the equality principle. Harv Civ Rights-Civ Liberties Law Rev 46:1

Dressler J, Michaels AC (2006) Understanding criminal procedure, 4th edn. LexisNexis, New York, NY

Harris D (1997) 'Driving While Black' and all other traffic offenses: the Supreme Court and pretextual stops. J Crim Law Criminol 87:544

Maclin T (1990) The decline of the right of locomotion: the fourth amendment on the streets. Cornell Law Rev 75:128

Maclin T (1998) Race and the fourth amendment. Vanderbilt Law Rev 51:333

Law of Police Use of Force

Erik Luna
Washington and Lee University School of Law,
Washington and Lee University, Lexington,
VA, USA

Overview

Perhaps no topic in policing has generated more debate than the use of force. In news coverage, as well as in movies and television dramas, police officers seem to confront violent offenders in lethal melee on a constant basis. Sometimes the stories concern allegations of excessive force or misconduct: the beating of Rodney King, the shooting of Amadou Diallo, the sexual assault of Abner Louima, the deadly FBI raid in Waco, the sniper fire in Ruby Ridge, and, most recently, the alleged rough treatment of Occupy Wall Street protestors. From these examples, people may extrapolate not only the incidence of official violence but also the law governing the use of force. This entry provides some background on the police use of force and a critical summary of the relevant law. The question remains as to whether the jurisprudence is sufficiently clear, consistent, and coherent in order to guide law enforcement, protect individual rights, and inform the citizenry of the legal boundaries.

Force, Police, and the Law

Definitions, Contexts, and Types of Force

For present purposes, the *use of force* may be defined as “any physical strike or instrumental contact with a person; any intentional attempted physical strike or instrumental contact that does not take effect; or any significant physical contact that restricts the movement of a person” (IACP, 116). Although some works on the topic provide a transnational perspective (Chevigny, 145–248), this entry focuses solely on the law of police use of force in the United States. The term *police* refers to any government agency empowered to enforce the criminal law of a given jurisdiction, including police departments, sheriff’s offices, highway patrol, and the Federal Bureau of Investigation. The *law* of police use of force comprises the legal justifications and limitations for the use of force, as well as the remedies for the use of excessive force. The resulting framework addresses the issues of when the police may use force and what types of force may be used in terms of the amount of force and the method employed.

In general, the police use of force may be justified to prevent the commission of a crime, to implement the criminal justice process, or to preserve public order. Many jurisdictions and the American Law Institute’s Model Penal Code specifically provide for the use of force in law enforcement (MPC, § 3.07). In certain contexts delineated by statute, police officers are said to possess a privilege to use force. As a constitutional matter, the police use of force is typically analyzed as a seizure of persons subject to the constraints of the Fourth Amendment. Over the years, the US Supreme Court has defined the term *seizure* in various ways. As used here, a seizure occurs when official force is intentionally applied to terminate or restrain an individual’s freedom of movement (*Brendlin v. California*, 551 U.S. 249, 254 (2007)).

The rubric of crime prevention comprises the police use of force to protect persons and property. For instance, the US Supreme Court has held that a police officer may temporarily detain an individual based on a reasonable suspicion that “criminal activity may be afoot” (*Terry v. Ohio*, 392 U.S. 1, 50 (1968)). If the officer also reasonably suspects that the individual is armed and dangerous, he may conduct a frisk for weapons to ensure his own safety and the security of others. Recently, the Supreme Court upheld the use of force to detain occupants during a home search due to “the risk of harm to both officers and occupants” (*Muehler v. Mena*, 544 U.S. 93, 100 (2005)).

The use of force to effect an arrest or otherwise seize a suspected criminal is the most obvious example of force employed to implement the legal process. Law enforcement may conduct an arrest – that is, taking an individual into custody to answer for a criminal charge – based on "probable cause," which is defined as sufficient facts and circumstances to warrant an officer of reasonable caution to believe that a crime has been committed and the person to be arrested committed it (*Brinegar v. United States*, 338 U.S. 160, 175–76 (1949)). Likewise, the police may use force to prevent the escape of a suspected criminal or to prevent the destruction of evidence. As for preserving public order, the paradigmatic example is the police use of force to suppress a riot after the rioters have been ordered to disperse and warned that force will be used if they do not obey.

Traditionally, the police use of force has been divided into two categories: *deadly force* (or lethal force) and *non-deadly force* (or nonlethal or less-lethal force). Deadly force is commonly defined as force that is likely to cause death or serious physical injury. Most prominently, this category encompasses discharging firearms at other individuals. But it also includes, for instance, the use of a police cruiser against another occupied vehicle, as when law enforcement performs a "Precision Intervention Technique" (PIT) maneuver – hitting another vehicle with the goal of causing it to spin to a stop. Non-deadly force comprises the use of all other types of force, including physical contact (e.g., tackling and punching), restraining devices (e.g., handcuffs), impact weapons (e.g., batons and flashlights), chemical weapons (e.g., pepper spray and mace), electronic weapons (e.g., stun guns), kinetic energy weapons (e.g., beanbag guns), explosive devices (e.g., flash-bang grenades), and police dogs.

Constitutional Limitations on Police Use of Force

At common law, government agents and private citizens were allowed to use both lethal and nonlethal force to apprehend a felon or prevent the commission of a felony. Nonlethal force could be used to prevent the commission of a misdemeanor amounting to a breach of the peace, but force could not be used to prevent the commission of any other misdemeanor. These standards were carried over to the American colonies and subsequently adopted in judicial opinions or codified by statute in the post-revolutionary states. Certainly, the common-law rules made some sense in an era when all felonies were punishable by death and private citizens often served as the community's law enforcers, dutifully answering the "hue and cry" to help apprehend criminals (Perkins & Boyce, § 10.1).

In modern times, however, this justification was undermined by the near monopoly of force exercised by the police, the proliferation of felony offenses in penal codes, and the reservation of capital punishment for the crime of murder. Nonetheless, some states maintained the so-called "fleeing felon" rule of the common law, which allowed the police to use whatever force was necessary, including deadly force, to prevent the escape of a felon. In its 1985 decision in *Tennessee v. Garner*, the US Supreme Court found that applications of this rule could violate the Fourth Amendment's ban on unreasonable seizures. "It is not better that all felony suspects die than that they escape," the Court opined, declaring that a "police officer may not seize an unarmed, nondangerous suspect by shooting him dead" (*Tennessee v. Garner*, 471 U.S. 1, 11 (1985)). The *Garner* Court then announced a constitutional standard for the use of deadly force:

> "Where the officer has probable cause to believe that the suspect poses a threat of serious physical harm, either to the officer or to others, it is not constitutionally unreasonable to prevent escape by using deadly force. Thus, if the suspect threatens the officer with a weapon or there is probable cause to believe that he has committed a crime involving the infliction or threatened infliction of serious physical harm, deadly force may be used if necessary to prevent escape, and if, where feasible, some warning has been given." (*ibid.*, 11–12)

Subsequent decisions have both expanded upon *Garner* and limited its categorical prohibition on deadly force to the basic fact pattern presented by the case. In 1989, the Supreme Court in *Graham v. Connor* made clear that excessive force claims arising from arrests, investigatory stops, and other seizures are properly characterized as invocations of the Fourth Amendment. Other constitutional provisions apply to individuals already in lawful custody, with pretrial detainees and convicted prisoners protected from excessive force by the Due Process Clause and the Eighth Amendment, respectively. Most importantly, the *Graham* Court held that all Fourth Amendment claims of excessive force, both deadly and non-deadly, should be evaluated for "reasonableness." This standard balances the nature and quality of the police intrusion against the governmental interests at stake, "including the severity of the crime at issue, whether the suspect poses an immediate threat to the safety of the officers or others, and whether he is actively resisting arrest or attempting to evade arrest by flight" (*Graham v. Connor*, 490 U.S. 386, 396 (1989)).

The *Graham* Court emphasized that the evaluation must not employ the "20/20 vision of hindsight." Instead, the proper perspective is that of a reasonable officer who is on the scene and may be "forced to make split-second judgments – in circumstances that are tense, uncertain, and rapidly evolving – about the amount of force that is necessary in a particular situation" (*ibid.*, 396–97). Moreover, the inquiry is one of "objective reasonableness," that is, whether the use of force was reasonable in light of the facts and circumstances confronting the police officer but without regard to his underlying intent or motive. An officer's corrupt intentions, even those that are "malicious and sadistic," do not transform an objectively reasonable use of force into a Fourth Amendment violation, just as an officer's good intentions do not render constitutional an objectively unreasonable use of force.

Most recently, the Supreme Court's decision in *Scott v. Harris* considered a specific fact pattern involving the use of police vehicles to seize a fleeing motorist. Although the use of a police cruiser to ram the motorist's vehicle did amount to deadly force, the Court rejected the idea that its earlier decision in *Garner* established "a magical on/off switch" (*Scott v. Harris*, 550 U.S. 372, 382 (2007)). The ultimate test is not whether the police used deadly force but whether the use of such force was objectively reasonable under the circumstances. In terms of the danger to others, the present case was quite unlike the escape on foot of an unarmed felony suspect in *Garner*. Pointing to a videotape of the chase, the *Scott* Court weighed the risks of the PIT maneuver versus the fleeing motorist's endangerment of pedestrians, other civilian motorists, and officers involved in the chase. According to the Court, the test of reasonableness can take into consideration the number of people imperiled by the fleeing motorist, as well as the relative culpability of that motorist versus those he placed in harm's way.

Means of Evaluating the Police Use of Force

In theory, the police use of force might be assessed through a number of legal paths. For instance, statutory schemes and departmental regulations on the use of force might be challenged on their face as constitutionally unreasonable. However, this option is largely foreclosed by legal and practical barriers to bringing facial challenges under the Fourth Amendment (*Sibron v. New York*, 392 U.S. 40, 59–60 (1968)), which might explain the persistence of statutes of dubious constitutionality in light of *Garner* (McCauley and Claus, 3). Moreover, the internal rules of an executive agency are often non-justiciable (*United States v. Caceres*, 440 U.S. 741 (1979)).

Another option would be to evaluate the police use of force in criminal proceedings. Defendants claiming to be victims of excessive force might seek, inter alia, the suppression of incriminating evidence at trial. But the courts have refused to apply the exclusionary rule absent a causal connection between the excessive force alleged by the defendant and the evidence he seeks to suppress (*United States v. Watson*, 558 F.3d 701 (7th Cir. 2009)). State prosecutors can bring charges against police officers whose use of force constitutes a criminal offense, and federal prosecutors can indict officers for uses of force that amount to deprivations of constitutional rights (18 U.S.C. § 242). Although beyond the scope of this entry, various legal requirements and practical impediments limit the viability of criminal prosecutions as a method to assess the use of force and check police abuses (Harris, 55–64).

Instead, civil suits have proven to be the most feasible means to contest the use of force. Purported victims of excessive force can bring state tort law actions against offending police officers, although claims based on simple negligence may be barred by sovereign immunity for the discretionary acts of government officials made in the course of duty. Alternatively, a 1994 federal law empowers the US Department of Justice to investigate and bring suit against police departments that have a "pattern or practice" of conduct violating the civil rights of people within their jurisdictions (42 U.S.C. § 14141). Since its enactment, the law has been used to investigate dozens of police departments, including those exhibiting patterns of excessive force. Typically, the cases result in a settlement where the department in question agrees to implement a slate of reforms. However, the law is not designed to investigate a specific case or to institute a civil action for a particular use of force.

Today, most use of force litigation is brought in federal court under a provision of the Civil Rights Act of 1871 (42 U.S.C. § 1983) which imposes liability on persons who, while acting under color of law, deprive others of their civil rights. Section 1983 itself is not a source of rights but instead provides a cause of action to vindicate established federal rights. As discussed, the relevant right in excessive force cases is the Fourth Amendment prohibition against unreasonable seizures. Although § 1983 only applies to state officials, the Supreme Court has ruled that an implied cause of action exists against federal officials for civil rights violations (*Bivens v. Six Unknown Federal Narcotics Agents*, 403 U.S. 388 (1971)).

Before trial, § 1983 cases can be stymied, or at least complicated, by the threshold issue of qualified immunity. This doctrine shields government officials from liability so long as his or her conduct does not violate clearly established rights that would have been known to a reasonable officer. The US Supreme Court has held that the first step in qualified immunity analysis is determining whether the police officer's conduct violated a constitutional right. At times, "a court might find it necessary to set forth principles which will become the basis for a holding that a right is clearly established," which, in turn, "serves to advance understanding of the law and to allow officers to avoid the burden of trial if qualified immunity is applicable" (*Saucier v. Katz*, 533 U.S. 194, 201 (2001)).

Over the past half century, thousands of federal lawsuits have raised claims of excessive force under § 1983, thereby providing the principal means for shaping the law of police use of force (Harmon, 1126). Indeed, *Garner*, *Graham*, and *Scott* were all cases brought under this provision. A number of reasons help explain the dominance § 1983 litigation, including the potential for successful plaintiffs to obtain attorney fees and costs (which, somewhat ironically, can be greater than the damages awarded to plaintiffs), as well as the fact that federal cases tend to come to trial much faster than state cases.

Controversies

The Supreme Court opinions on the use of force are often "cast at a high level of generality" and "[t]he test of reasonableness under the Fourth Amendment is not capable of precise definition or mechanical application" (*Brosseau v. Haugen*, 543 U.S. 194, 199 (2004)). Or as one federal judge put it, "There is no such thing as a per se violation of the Fourth Amendment" (*Cox v. Treadway*, 75 F.3d 230, 241 (6th Cir. 1996) (Ryan, J., concurring in part)). However, the idea that there are no per se rules in this area is belied by the fact that *Garner* did, in fact, adopt a bright-line rule regarding the use of lethal force. The same can be said of *Scott*, despite claims to the contrary in the majority opinion. Specifically, the *Scott* Court delivered the following conclusion: "A police officer's attempt to terminate a dangerous high-speed car chase that threatens the lives of innocent bystanders does not violate the Fourth Amendment, even when it places the fleeing motorist at risk of serious injury or death" (*Scott*, 550 U.S. at 386). This rule may be "sensible," as the *Scott* Court believed, but it certainly appears to be "absolute" (*ibid.*, 389 (Breyer, J., concurring)).

Objections to a per se approach might be tempered by reading *Scott* as creating a special rule regarding deadly force, as some lower court opinions have done with *Garner* (*Quintanilla v. City of Downey*, 84 F.3d 353, 357 (9th Cir.1996)). Yet the distinction between deadly and non-deadly force is not always clear. For instance, striking an individual in the head with a baton or other hard object may properly be viewed as deadly force, and some police departments classify the use of chokeholds and the firing of warning shots as deadly force. Moreover, the use of ostensibly non-deadly instruments can still produce lethal consequences, such as when a flash-bang grenade – an explosive device that creates a disorienting light and sound – ignites flammable accelerants contained in a building. Rather than relying solely on the Supreme Court's rough dichotomy between deadly and non-deadly force, most law enforcement agencies employ far more discerning methodologies to guide their officers. For instance, use-of-force continuums incorporate multiple levels of force, which typically begin with officer presence and verbal communication and end with the use of deadly force.

The totality of the circumstances approach espoused by *Graham* presents its own difficulties. As then-Judge (now-Stanford Law Professor) Michael McConnell opined: "[B]ecause excessive force jurisprudence requires an all-things-considered inquiry with 'careful attention to the facts and circumstances of each particular case,' there will almost never be a previously published opinion involving exactly the same circumstances" (*Casey v. City of Federal Heights*, 509 F.3d 1278, 1284 (10th Cir. 2007) (quoting *Graham*)). To reiterate, the Supreme Court in *Graham* listed three non-exclusive factors that may be relevant when considering the objective reasonableness of police use of force: the severity of the crime, whether the suspect poses an immediate threat to the safety of the officers or others, and whether the suspect is actively resisting arrest or attempting to evade arrest by flight. In *Scott*, the Court referenced the number of people endangered by the suspect and the relative culpability of that suspect versus those he placed in harm's way. The lower courts have cited additional considerations in assessing reasonableness, including:

- the type and amount of force used by the police
- the extent of injuries inflicted upon the suspect
- the amount of time and any changing circumstances during which the officer had to determine the type and amount of force that appeared necessary
- the relative physical characteristics of the officer and the suspect (e.g., height, weight, age, gender, and strength)
- the nature of the arrest charges
- the availability of alternative methods of capturing or subduing the suspect
- whether a warrant was used
- whether the suspect resisted or was armed
- whether more than one suspect or officer was involved
- whether a warning was given before the use of force
- whether the suspect was sober
- whether it was apparent that the suspect was emotionally disturbed
- whether the officer harbored ill will toward the suspect

Any number of other circumstances might be added to this list, either explicitly or by use of a catch-all phrase, such as "whether other dangerous or exigent circumstances existed at the time of the arrest" (*Chew v. Gates*, 27 F.3d 1432, 1440 n.5 (9th Cir. 1994)). Without constraints on the relevant circumstances or at least some guidance as to the hierarchy or relative weight of factors, the reasonableness standard can have the semblance of a Rorschach test. Along these lines, Professor Rachel Harmon has argued that "the reasoning in these cases is ad hoc, often inconsistent, and sometimes ill-considered," meaning that "the outcomes of future cases are largely unpredictable, even by the Supreme Court's own measures" (Harmon, 1123).

As noted above, some cases have taken into consideration the type and amount of force used and the availability of alternative methods of capturing or subduing a suspect. Other cases, however, have held that police officers need not use the least intrusive means of responding to an exigent situation (*Scott v. Henrich*, 39 F.3d 912 (9th Cir. 1994)). Likewise, officers are not required to use, and their departments are not required to provide, particular types of weapons (*Plakas v. Drinski*, 19 F.3d 1143 (7th Cir. 1994)). And as a matter of law, officers might not have a duty to retreat before using deadly force (*Penley v. Eslinger*, 605 F.3d 843, 855 (11th Cir. 2010)). Although several cases emphasized the communication of warnings before using force, other cases have found it objectively reasonable for an officer to forego a warning (*Colston v. Barnhart*, 102 F.3d 85 (3d Cir. 1997)). More generally, the police are not required to make any announcement when carrying out an arrest in a public place (*Catlin v. City of Wheaton*, 574 F.3d 361 (7th Cir. 2009)).

Another subject of debate in excessive force cases relates to the level of force used and the kind of injury inflicted. In passing, the *Graham* Court stated that "[n]ot every push or shove, even if it may later seem unnecessary in the peace of a judge's chambers, violates the Fourth Amendment" (*Graham v. Connor*, 490 U.S. 386, 397 (1989)). Along these lines, some courts have held that a Fourth Amendment violation only occurs when the police use more than a de minimis amount of force (*Cook v. City of Bella Villa*, 582 F.3d 840 (5th Cir. 2009)) and that the suspect must suffer an actual or non-de minimis injury (*Fisher v. City of Las Cruces*, 584 F.3d 888 (10th Cir. 2009)). But other courts have held that excessive force claims do not require more than a de minimis injury ((*Chambers v. Pennycook*, 641 F.3d 898 (8th Cir. 2011)), which is consistent with the Supreme Court's jurisprudence on inmate claims of excessive force under the Eighth Amendment (*Wilkins v. Gaddy*, 130 S. Ct. 1175 (2010)).

Additional areas of tension in the case law include the significance of verbal or physical threats by the police, particularly in the absence of physical injury; the pertinence of an officer's ill will toward a suspect in a regime premised on objective reasonableness; and the relevance of department policies given that a violation of such policies is not a basis for an excessive force claim. But perhaps the most glaring inconsistency in the jurisprudence, and the clearest split among the federal circuit courts, concerns the appropriate time frame for analyzing the reasonableness of police use of force (Avery, 267–89). Some decisions have focused on the Supreme Court's language of "immediate flight," "on the scene," "reasonableness at the moment," and "split-second judgments," all of which connote short time spans and temporal proximity (*Graham v. Connor*, 490 U.S. 386, 396 (1989)).

With these phrases in mind, the Second, Fourth, Fifth, Eighth, and Eleventh Circuits have limited the assessment to those circumstances existing at the time of the seizure. For instance, the Second Circuit concluded that an officer's

> "actions leading up to the shooting are irrelevant to the objective reasonableness of his conduct at the moment he decided to employ deadly force. The reasonableness inquiry depends only upon the officer's knowledge of circumstances immediately prior to and at the moment that he made the split-second decision to employ deadly force." (*Salim v. Proulx*, 93 F.3d 86, 92 (2d Cir. 1996))

By contrast, the First, Third, Ninth, and Tenth Circuits have held that proper analysis of excessive force claims must take into consideration the actions leading up to a seizure. In rejecting decisions espousing a narrow time frame, the Third Circuit could

> "not see how these cases can reconcile the Supreme Court's rule requiring examination of the 'totality of the circumstances' with a rigid rule that excludes all context and causes prior to the moment the seizure is finally accomplished. 'Totality' is an encompassing word. It implies that reasonableness should be sensitive to all of the factors bearing on the officer's use of force." (*Abraham v. Raso*, 183 F.3d 279, 291 (3d Cir. 1999))

Still another approach has been employed by the Sixth and Seventh Circuits, which divide use of force incidents into separate, sequential parts for analysis. As explained by the Seventh Circuit, "we carve up the incident into segments and judge each on its own terms to see if the officer was reasonable at each stage," but without returning to prior segments to reconsider the event "in light of hindsight" (*Plakas v. Drinski*, 19 F.3d 1143, 1150 (7th Cir. 1994)).

Conclusion

Given the above, it is not surprising that scholars have criticized the current law of police use of force. Professor Harmon has described the Supreme Court's jurisprudence as "profoundly impoverished," with the "paucity of reasoned analysis in this area" undermining "the evolution of a principled case law defining clear requirements for the legitimate use of police force" (Harmon, 1119, 1183). Another commentator claims that "[a]s brutality remains a problem and law enforcement challenges continue to grow, there has never been a better time to add flesh to the skeletal understandings of force found in the text of the Fourth Amendment" (Note, 2009, 1721–22). Arguably, the problem is not just the quantity of law on the use of force or the appropriate choice (or mix) of bright-line rules and flexible standards. Rather, there appears to be a disconnect between the case law on the one hand and people's perceptions on the other.

Consider, for example, the "added wrinkle" in the Supreme Court's decision in *Scott v. Harris*: "existence in the record of a videotape capturing the events in question" (*Scott v. Harris*, 550 U.S. 372, 378 (2007)). Based on its review of the tape, the *Scott* Court concluded that the fleeing motorist's "version of events is so utterly discredited by the record that no reasonable jury could have believed him," deriding the lower courts for having "relied on such visible fiction" when "it should have viewed the facts in the light depicted by the videotape" (*ibid.*, 380–81). In dissent, Justice John Paul Stevens believed that "the tape actually confirms, rather than contradicts, the lower courts' appraisal of the factual questions at issue" (*ibid.*, 389–90 (Stevens, J., dissenting)). Along the way, Justice Stevens panned "eight of the jurors on this Court" for reaching a different verdict than the lower court judges, "who are surely more familiar with the hazards of driving on Georgia roads than we are" (*ibid.*, 390). Nonetheless, the majority was "happy to allow the videotape to speak for itself" (*ibid.*, 378 n.5 (majority opinion)). Since *Scott*, the lower courts appear to have taken the cue, relying on videotapes in ruling on summary judgment motions in use of force cases (*Dunn v. Matatall*, 549 F.3d 348 (6th Cir. 2008)).

As it turns out, however, "what [a videotape] says depends on to whom it is speaking," as demonstrated by an empirical study conducted by Professors Dan Kahan, David Hoffman, and Donald Braman. The video in *Scott* was shown to a diverse sample of more than 1,300 Americans, who were then asked to describe what they saw and to provide their opinions on key issues identified by the Supreme Court. The authors found that a substantial majority of study participants had viewpoints consistent with those expressed by the eight-member majority of the Supreme Court. This perspective was hardly uniform, however. "Whites and African Americans, high-wage earners and low-wage earners, Northeasterners and Southerners and Westerners, liberals and conservatives, Republicans and Democrats – all varied significantly in their perceptions of the risk that [the fleeing motorist] posed, of the risk the police created by deciding to pursue him, and of the need to use deadly force against [the motorist] in the interest of reducing public risk" (Kahan et al., 903). As such, the vice of *Scott* was not necessarily the outcome but the way in which the Court reached it, namely, by asserting that no reasonable jury could have come to a different conclusion regarding the reasonableness of the police use of force.

The authors suggested several alternative rationales – for instance, the need for uniform, predictable rules – by which the *Scott* Court could have rendered the same judgment without dismissing other coherent viewpoints as unreasonable. To these suggestions, one might add the larger project identified by Professor Harmon: the need for "an accessible and transparent framework that the public may use to analyze highly publicized uses of police force" (Harmon, 1183). It may well be true that only a small percentage of police-citizen contacts involve the use of physical force (Adams et al., 3). But as mentioned at the beginning of this entry, the use of force tends to be understood through the lens of mass-media portrayals, which can distort popular perceptions or at least focus attention on extreme circumstances, at times fostering distrust of the police and disrespect for the legal system (Luna, 1112–19). Finding the means to educate the citizenry about the jurisprudence of force and to inform legal decision-makers about perceptions of police-citizen interactions might provide the first step toward an understanding of the police use of force that goes beyond the "Hollywood-style car chase of the most frightening sort" (*Scott v. Harris*, 550 U.S. 372, 379 (2007)).

Related Entries

- ► Law of Police Seizures and the Exercise of Discretion
- ► Legal Control of the Police
- ► Police and the Excessive Use of Force
- ► Police Use of Firearms

Recommended Reading and References

American Law Institute (1962) Model Penal Code: official draft and explanatory notes (MPC). American Law Institute, Philadelphia

Adams K et al (1999) Use of force by police: overview of national and local data. U.S. Dept. of Justice, Washington, DC

Avery M (2003) Unreasonable seizures of unreasonable people: defining the totality of circumstances relevant to assessing the police use of force against emotionally disturbed people. Columbia Hum Rights Law Rev 34(2):261

Avery M, Rudovsky D, Blum KM (1996) Police misconduct: law and litigation. Clark Boardman Callaghan, New York

Chevigny P (1995) Edge of the knife: police violence in the Americas. The New Press, New York

Civil Rights Division, U.S. Department of Justice, Conduct of law enforcement agencies. http://www.usdoj.gov/crt/split/police.php

Harmon RA (2008) When is police violence justified? Nw UL Rev 102(3):1119–1187

Harris D (2012) The interaction and relationship between prosecutors and police officers in the United States, and how this affects police reform efforts. In: Luna E, Wade M (eds) The prosecutor in transnational perspective. Oxford University Press, New York

International Association of Chiefs of Police (IACP) (2006) Protecting civil rights: a leadership guide for state, local, and tribal law enforcement. U.S. Dept. of Justice, Washington, DC

Luna E (2000) Transparent policing. Iowa Law Rev 85(4):1107–1194

Kahan DM, Hoffman DA, Braman D (2009) Whose eyes are you going to believe? *Scott v. Harris* and the perils of cognitive illiberalism. Harv Law Rev 122(3):837–906

McCauley RP, Claus LN (2007) Police use of force statutes: states inconsistent with established constitutional law. Crim L Bull 43(5):726–738

Note (2009) Retreat: the Supreme Court and the new police. Harv Law Rev 122(6):1706–1727

Law of Undercover Policing

Jacqueline E. Ross
University of Illinois College of Law,
Champaign, IL, USA

Synonyms

Covert policing; Infiltration; Sting operations

Overview

Undercover policing is an investigative technique that is largely unregulated by constitutional criminal procedure in the United States, though the European Court of Human Rights has proven more willing to regulate the impact of undercover tactics on fundamental rights. American criminal law regulates undercover policing primarily through the entrapment defense, of which there are different variants. Current tests rest on empirical assumptions about the ways in which police investigations influence their target environment. Current versions of the entrapment defense do not take account of the varieties of undercover investigations and the different types of influences undercover agents may exert on their targets.

Key Issues and Controversies Surrounding the Choice of Regulatory Framework

Undercover policing is proactive an investigative tactic that allows police officers and informants to infiltrate criminal organizations or to test the criminal resolve of suspected offenders by disguising their true identities and orchestrating criminal opportunities. Undercover tactics have been examined through a rich variety of regulatory prisms, each of which captures a different feature that makes such operations vulnerable to abuse. Criminal law scholars worry about entrapment, which is an affirmative defense for targets of undercover stings who can show that government agents convinced them to commit a crime when they were not predisposed to engage in such criminal activity. Criminal law scholars also consider the potential criminal liability of undercover agents and informants who participate in the crimes they investigate. Criminal procedure scholars examine the impact of undercover investigations on the constitutional rights of criminal defendants, including their legitimate expectations of privacy under the Fourth Amendment and the ways in which undercover operations can compromise the privilege against compelled self-incrimination under the Fifth Amendment and the Sixth Amendment right to counsel.

Many civil libertarians look beyond the rights of criminal defendants to the First and Fourth Amendment interests of ordinary citizens. These concerns center on covert operations that pursue intelligence rather than evidence, since such investigations are likely to affect not only criminals but anyone who belongs to religious and political organizations that the government seeks to infiltrate. Because such investigations may not be predicated on concrete suspicion of wrongdoing, they may cast a wider net, yielding information whose validity may never be tested in court; covert scrutiny of this sort may burden the exercise of freedom of speech, assembly, and religion and may compromise the privacy of confidential communications between members of targeted organizations. Finally, state bar associations and courts that interpret state ethics codes have from time to time become concerned that prosecutors who supervise undercover operations might infringe disciplinary rules that prohibit attorneys from sanctioning acts of deception, or from sponsoring direct or indirect contact with represented parties, outside the presence of their lawyers.

Among this multiplicity of regulatory options, the entrapment defense remains the most significant, if only because most of the other types of challenges have proven either legally ineffectual or limited in scope. Appeals to civil liberties, in particular, have been notably less successful in the United States than in the European Union, where the European Court of Human Rights (ECtHR) has interpreted the Convention on Human Rights (ECHR) to protect the fair trial rights of criminal defendants whose convictions rest on evidence obtained undercover. Within the European Union, many national legislatures have understood the Convention's protections for privacy (set forth in Article Eight) as requiring statutory regulation of undercover operations and the enactment of a warrant procedure that ensures prior approval and continuing oversight by judicial officers. The Convention's privacy protections have been interpreted to limit the use of undercover tactics primarily to the pursuit of serious crimes, after showing that other, less intrusive investigative tactics seem unlikely to yield evidence against highly secretive offender groups. Accordingly, European legal systems authorize undercover operations primarily for the investigation of organized crime. Statutory constraints limit both the types of crimes undercover agents can investigate and the tactics they can use, as undercover agents in many European countries risk criminal sanctions for facilitating or taking part in crimes for which undercover tactics are not authorized by law – even when such assistance or participation serves exclusively investigative purposes. The European Court of Human Rights has also sustained challenges to undercover operations on grounds of entrapment, or because of the use of secret evidence, without adequate opportunity to question undercover agents and informants at trial. Both types of challenges invoked the right to a fair trial, as guaranteed by Article Six of the ECHR.

By contrast, undercover operations in the United States are not authorized by statute and may be initiated without obtaining a warrant or establishing probable cause or even reasonable suspicion that a crime is being committed. In *Hoffa v. United States*, 385 U.S. 293 (1966), the Supreme Court held that the use of undercover agents and informants does not tantamount to a search, within the meaning of Fourth Amendment, as targets waive their expectations of privacy in the information they voluntarily impart to others; in essence, suspects assume the risk of betrayal by their associates, or the risk that those with whom they commit crimes may turn out to be undercover agents. While listening in on a target's telephone conversation is a search that infringes on a protected expectation of privacy, eliciting such a target's confidences through undercover agents or informants does not count as a search within the meaning of the Fourth Amendment. Accordingly, resort to undercover tactics, unlike the use of electronic surveillance, requires no showing that less intrusive investigative methods have been tried or are likely to fail and no showing that the crimes being investigated are sufficiently serious to warrant the use of the undercover technique. In *Illinois v. Perkins*, 496 U.S. 292 (1990), the Supreme Court also rejected claims that undercover questioning of a custodial defendant violates his Fifth Amendment right to silence, even if he has already indicated his unwillingness to speak to the police, because the Fifth Amendment protects defendants only from compelled self-incrimination; if the defendant does not realize he is speaking to a representative of the police, the Court has reasoned, he cannot experience the conversation as an exertion of pressure by the government. Defendants have had more success challenging undercover questioning of defendants who have invoked their right to counsel. In *Massiah v. United States*, 377 U.S. 201 (1964), the Supreme Court held that once formal charges are in place, undercover questioning can violate a criminal defendant's Sixth Amendment right to counsel – but only if the questioning concerns the crimes with which he is charged and only if he has already invoked his right to counsel. Finally, sufficiently outrageous sting operations can violate defendants' substantive Due Process rights under the Fifth and Fourteenth Amendments. But this remains largely a theoretical possibility, as very few convictions have been vacated on that basis.

In recent years, US courts have also rebuffed claims that undercover policing invades their First Amendment rights of freedom of speech and assembly, though similar claims made in the 1970s and 1980s resulted in consent decrees under the terms of which police departments for cities like New York and Chicago agreed to refrain from surveillance of political and religious organizations absent any concrete evidence of ongoing or incipient criminal conduct. These consent decrees were quietly abandoned in the wake of the September 11 attacks, and the FBI, too, has modified the internal guidelines that restricted domestic intelligence agencies absent evidence of a criminal threat.

If American courts have rarely sustained legal challenges that are framed in the language of rights, and if American legislatures have never enacted a statutory warrant requirement, they have also avoided the regulatory path taken by most European legal systems before undercover tactics came to be regulated by a statutory warrant requirement: that of restraining undercover tactics through the threat of criminal sanctions for undercover agents or informants who take part in the crimes they investigate. Courts have taken the position that "[c]riminal prohibitions do not generally apply to reasonable enforcement actions by officers of the law" (*Brogan v. United States*, 522 U.S. 398 (1998)). And state legislatures have enacted broad immunities for undercover agents, through the so-called "public authority" defense, which protects law enforcement officers generally from criminal liability for enforcement actions that were duly authorized by their superiors. Unlike French and Italian immunities, which spell out the undercover tactics in which undercover agents and informants may lawfully engage, the American defense makes no effort to enumerate the enforcement actions – undercover or otherwise – for which the defense is designed.

The Dominant Framework: Entrapment

Accordingly, the entrapment defense remains the primary regulatory constraint on the criteria by which targets of undercover operations are selected and on the ways in which undercover investigations are planned and carried out. But there is an ambiguity at the heart of the entrapment defense, which accounts for the divergent ways in which it has been codified and discussed. Does the rationale for the defense rest on the view that targets do not deserve to be punished if they would have been unlikely to commit the charged offense without the criminal opportunity they were offered by government? Or does the entrapment defense exist primarily to reorient investigators away from targets of opportunity to "real" criminals, to whom the government merely tenders a convenient occasion to commit acts in which they would otherwise engage undetected?

The so-called objective test accords with the latter rationale, as it concerns itself with the nature and strength of the inducement employed by the government instead of the predisposition of offenders. While a particular defendant may well have been predisposed to commit the crime with which he is charged, an excessively appealing inducement may nonetheless amount to entrapment if it has a tendency to overcome the resistance of the average law-abiding citizen. Only a minority of jurisdictions, such as California and Oregon, have embraced this version of the entrapment defense. By contrast, the subjective variant of the entrapment defense accords with the former rationale, as it makes the defense available only to those offenders who were not predisposed to commit the crime they were eventually encouraged or persuaded to commit. A third, hybrid variant makes the defense available only if the defendant can establish both that he was not predisposed to commit the crime *and* that the government's tactics were unfair, meaning that the pressures and inducements it used were excessive. This is the most restrictive variant of the entrapment defense, as a purely subjective test would sustain a defense of entrapment even when the government offered a target otherwise reasonable inducements, so long as the targeted offender could establish a lack of predisposition.

In *Sherman v. United States*, 356 U.S. 369, 372 (1958), the Supreme Court embraced this hybrid version of the test by asking whether the inducement was objectively excessive and whether the government's tactics in fact implanted the idea for the crime in the mind of an "unwary innocent," reasoning that the legislature that defined the criminal offense could not have intended the criminal prohibition to apply to those who would not have committed such a crime without encouragement by undercover agents or informants. Critics of the subjective entrapment defense and its hybrid variant have pointed out that any claim that targets who lack predisposition do not deserve criminal punishment must explain the unavailability of any similar defense to those "unwary innocents" who were led astray by their friends rather than government agents or informants.

Critics of all three variants of the entrapment defense point out that all current versions of the test make the assumption that one can meaningfully distinguish "true criminals" from "unwary innocents." Commentators argue that almost anyone can be induced to commit a crime if the "criminal offer" is sufficiently tempting, so that the distinction between deserving and undeserving targets is at best a fluid one. The true question, for many reformers, is what level of inducement is considered reasonable, with some commentators using a market framework to argue that so-called above-market offers should be prohibited, because they may ensnare at least some targets who would be unlikely to commit crimes under "normal market conditions," meaning that they would be unlikely to take advantage of the ordinary of criminal opportunities they are likely to encounter in their normal surroundings.

Legal scholars who would prohibit the government from offering targets more than the "market price" as inducements to commit a crime have often been vague about what constitutes an excessive inducement, as they have generally not examined or systematized the variety of government sting operations that might give rise to a defense of entrapment. In the realm of sociology, however, Gary Marx's path breaking study, *Undercover: Police Surveillance in America* (1988), has identified a wide range of ways in which infiltration can alter a target's conduct or environment, along with a large number of factors that can distort the "naturalism" of an undercover operation. Marx points out that undercover operations can alter not only the opportunity structure for criminal conduct but also the motives, rewards, markets, or resources that shape targets' decisions about whether to offend and how. And among types of undercover operations, those in which an operative investigates past crimes are less likely to shape criminal conduct than anticipatory investigations that seek to prevent or facilitate future offenses. Investigations that offer criminal opportunities randomly will pose different risks of abuse than sting operations undertaken in response to prior intelligence about specific targets. Even if efforts to avoid concerns about entrapment lead investigators to emulate the natural criminal environment in their design of a criminal opportunity, too much realism may also overbear a target's autonomy, since "[i]n genuine criminal encounters, one party may coerce or threaten another party into participating" or may offer sex or drugs as an inducement.

Applicability of Entrapment Defense to the Varieties of Undercover Policing

The distortions that can be created by undercover operations may be compared to those inherent in either cognitive science or anthropology. In cognitive science, an experimental design may not correlate well with the real-life setting it seeks to emulate, if the experimental scenario is overly artificial. In the same way, an undercover agent may, perhaps unwittingly, offer a target a criminal opportunity that may be a poor substitute, or proxy, for the types of offenses that a target commits on his own, independently of the government. This may be a particularly salient risk for transactional undercover investigations, in which the undercover agent may agree to buy a much larger quantity of contraband from a suspected dealer than the target ordinarily sells to others. (American courts call this phenomenon "sentencing entrapment," if the investigator's aim is to increase the level of offending for the purpose of triggering a mandatory minimum prison term, or some other sentence enhancement.)

But some undercover investigations may more profitably be compared to ethnography, anthropology, or undercover sociology rather than the experimental designs of cognitive science. If an undercover agent infiltrates a criminal organization, learning about kinship patterns and power hierarchies, his presence as a facilitator or coconspirator can reshape some of the internal dynamics of the organization or help the organization to branch out into new territories or to take advantage of emerging criminal opportunities. In this, agents may resemble anthropologists, ethnographers, or sociologists, who become part of a community they are studying, and in the process may influence and alter the social environment, for example, by bringing with them weapons and tools that lead natives to abandon their own technologies or by helping native communities patent their knowledge of the therapeutic effects of local plants and herbs. If undercover agents participate in the organization's crimes, instead of proffering criminal opportunities that the government can control, the agents may not only help reshuffle established hierarchies but may shape at least some of the organization's criminal activities. Agents may also become entangled with and perhaps complicitous in the commission of crimes against innocent third parties or in retaliatory violence against unwitting informants (who may vouch for undercover agents without knowing their true identity.)

Most American undercover investigations can be grouped along a continuum, one end of which resembles the artificial, experimental scenarios through which cognitive scientists seek to reproduce natural occurrences under controlled conditions, while the other end of the spectrum features true infiltration of a natural environment which may, however, be altered by the presence of an outside observer. If undercover narcotics buys are closest to controlled experiments, other analogues to cognitive science experiments include the whole range of "honeypot" operations in which the police offer a criminal opportunity to targets who self-select by taking agents up on their criminal offers. Such investigations include random integrity testing of bank tellers, the establishment of storefront fencing operations, the positioning of bait cars filled with tempting merchandise, or the deployment of decoy officers posing as prostitutes on the street or as underage girls on the Internet. At the other end of the spectrum, undercover agents, sociologists, and journalists have infiltrated mental hospitals, supermarket chains, and extremist political parties, much as long-term moles have infiltrated the Cosa Nostra, the Hell's Angels, and the Klu Klux Klan.

To be sure, all undercover operations allow the government secretly to influence the crimes it investigates. But the concerns such influence might raise will be different when the government orchestrates an offense as a provable proxy for other, secret criminal activity, than when it allows its agents to take part in and perhaps even steer offenses that are orchestrated by others, in settings and with consequences that the government may either not be able to control or may be able to steer only at the cost of making the organization's true ambitions and capabilities difficult to disentangle from the government's own contribution.

Neither objective nor subjective tests for entrapment are currently able to distinguish among these types of influence in assessing targets' criminal responsibility for crimes to which government operatives contribute in an undercover capacity. That the entrapment doctrine functions as an affirmative defense makes it difficult to treat government influences as a matter of degree. By contrast, Italy and Germany treat entrapment as a scalar concept by making the doctrine available as a mitigating factor at sentencing. (The American doctrine of sentencing entrapment is rarely successful in reducing punishment to reflect government influence on the severity of a target's offense, as it is invoked only in truly exceptional cases.) This allows targets' penalties to be adjusted for the degree of government influence on targets' criminal activities, so that punishment may correspond to the nature and severity of the offenses that targets would have committed on their own. Compared with many of the member states of the European Union, the United States remains unusual in resisting both statutory regulation and warrant requirements, which would compel advance scrutiny of undercover operations, while making much less use of doctrines such as sentencing entrapment to adjust for distorting government influences in their aftermath.

Related Entries

- ▶ Comparative Legal Cultures
- ▶ Comparing Police Systems Across the World
- ▶ Drug Enforcement
- ▶ Drug Trafficking
- ▶ High Policing
- ▶ Police Lying and Deception
- ▶ Risk Management in Policing
- ▶ Sex Offenders and Criminal Policy
- ▶ Victims and the Police

Recommended Reading and References

Deis JG (2001) Economics, causation, and the entrapment defense. Ill L Rev 1207

Dworkin G (1985) The serpent beguiled me and I did eat: entrapment and the creation of crime. Law Phil 4:17

Hay B (2005) Sting operations, undercover agents, and entrapment. Mo L Rev 70:387

Joh EE (2009) Breaking the law to enforce it: undercover police participation in crime. Stan L Rev 62:155

Marx GT (1982) Who really gets stung? Some issues raised by the new police undercover work. Crime Delinq 28:165–169

Marx GT (1988) Undercover: police surveillance in America

Marx GT, Fijnaut C (eds) (1995) Undercover: police surveillance in comparative perspective

McAdams R (2005) The political economy of entrapment. J Crim L Crim 96:107

Moore CD (1995) The elusive foundation of the entrapment defense. Nw U L Rev 89:1151

Park R (1976) The entrapment controversy. Minn L Rev 69:163

Roiphe R (2003) The serpent beguiled me: a history of the entrapment defense. Seton Hall L Rev 33:257

Ross JE (2004a) Impediments to transnational cooperation in undercover policing: a comparative study of the United States and Italy. Am J Comp Law 52:569

Ross JE (2004b) Valuing inside knowledge: police infiltration as a problem for the law of evidence. Chi-Kent L Rev 79:1111–1115

Ross JE (2007) The place of covert surveillance in democratic societies: a comparative study of the United States and Germany. Am J Comp L 55:493–512

Seidman LM (1981) The Supreme Court, entrapment, and our criminal justice dilemm, 1981. Sup Ct Rev 111

Sherman J (2009) "A person otherwise innocent": policing entrapment in preventative, undercover counterterrorism investigations. U PaJ Const L 11:1475

Stevenson D (2004) Entrapment and the problem of deterring police misconduct. Conn L Rev 37:67–73

Yaffe G (2005) "The government beguiled me": the entrapment defense and the problem of private entrapment. J Ethics Soc Phil 1:2

Law, Diversion and Community Sanctions in Juvenile Justice

Ineke Pruin
Department of Criminology, University of Greifswald, Greifswald, Germany

Overview

Most criminal justice systems today have introduced alternatives to imprisonment. The general idea behind diversion and alternative sanctions is to keep the offender within society and to save him from the socially detrimental outcomes of imprisonment. This entry aims to give an overview about diversion strategies and alternative sanctions in juvenile justice systems which can in many countries be seen as pioneers for law reforms in the adult criminal justice systems.

Fundamentals of Juvenile Justice Systems

Generally speaking, the common idea of juvenile justice systems is that minors or juveniles should be dealt with differently than adults. According to criminological research results worldwide, juvenile delinquency and crime are episodic and regularly disappear in early adulthood. Consequently – and in line with Art. 40 (4) of the United Nations' Convention on the Rights of the Child (CRC) and many subsequent international instruments – justice systems have introduced special regulations for juvenile offenders which provide educational measures and sanctions instead of imprisonment as responses to youth criminality. The intention is to avoid compromising the developmental process of young persons in the transitional stage from youthfulness to adulthood. This development and the many international recommendations and conventions in the field of juvenile justice can be seen as the major achievement in modern juvenile criminal policy worldwide.

Justice systems differ in their approaches to youth offending. Parts of the juvenile justice systems are more justice oriented, which can mean that they sentence young offenders based on notions of punishment and accountability but also on the idea of proportionality, thus limiting interventions in order to avoid disproportionate sanctions. The idea of education and rehabilitation which is inherent also to justice-oriented juvenile justice systems is implemented by giving priority to diversion and community sanctions ("subsidiarity" of punishment or "education instead of punishment"). Therefore, in justice systems like in *Canada* or some *European states*, we find justice-oriented criminal procedural laws for juvenile offenders, oftentimes with extensive modifications compared to those for adults but by nature based on criminal law. Those are regularly focused on the offender rather than the offence, and the aim of rehabilitation plays a special role. Other juvenile justice systems are more oriented at family or youth welfare law. This includes that the offence is rather seen as a sign for the "need of help" of the juvenile and therefore leads to welfare or family law measures (e.g., in *Poland* or *Bulgaria*), and in some countries, family conferences play an important role in the practice of juvenile justice (e.g., *New Zealand*, *Northern Ireland*, or *Canada*). As more or less pure welfare-oriented systems do not provide for criminal sentences they allow for sometimes far-reaching transfers of juvenile offenders to adult criminal courts as in the *United States*.

Nowadays, as a result of manifold developments, we can hardly find a pure welfare or justice approach in one country. Systems that combine welfare as well as justice elements and that have introduced even more and different approaches to responding to young offenders are more common (Winterdyk 2002; Doob and Tonry 2004; Cavadino and Dignan 2006; Goldson and Muncie 2006; Junger-Tas and Decker 2006; Hartjen 2008; UNICEF 2008; Dünkel et al. 2011a). They differ in their scope concerning the age groups or the behavior they encompass (e.g., in many systems status, offences can open the door to juvenile justice, in others only "criminal" behavior can lead to reactions from the youth court; see for an overview Doob and Tonry 2004; Pruin 2011). However, due to the requirements of the CRC and their subsequent international instruments, many juvenile justice systems have managed to introduce procedural safeguards to make sure that the state reactions remain proportional to the seriousness of the offence.

Diversion

Definition and Functions of Diversion

One state reaction which is usually used more extensively in juvenile justice compared to adult criminal justice systems is the possibility of diverting young offenders from trial.

In the context of criminal justice, diversion is seen as a headword for decisions, measures, and strategies which aim to avoid formal penal prosecution, trial, and sentences (Koffmann and Dingwall 2007). A more concrete definition of diversion can be as follows: Diversion is the dismissal of the case when the offence is of minor gravity and if formal proceedings do not seem to be appropriate. Diversion follows the procedural principle of "expediency" (in contrast to a strict principle of "legality" which obligatorily requires formal proceedings and court decisions in any case) and its main aim is to avoid stigmatization through formal court proceedings (Dünkel et al. 2011b, p. 1651). In the juvenile justice systems, different forms of diversion are to be found: Diversion can be unconditional or conditional; furthermore, it often means the referral of juvenile offenders to health or social services or to mediation schemes instead of judging them in criminal court proceedings.

Diversion is regularly a decision of the prosecutor at a pre-court level. In many countries diversion can also be adjudicated by the judge, if after an accusation the case seems to be appropriate for a dismissal (e.g., because of reparation efforts by the offender that have been performed in the meantime). In some countries even the police are competent to divert juvenile offenders and therefore to avoid criminal proceedings more or less completely (e.g., police cautionings or warnings in Anglo-Saxon countries).

Particularly in the field of juvenile justice, since the mid-1980s many international recommendations have emphasized that diversion should be given priority as an appropriate and effective strategy of juvenile crime policy (e.g., Convention on the Rights of the Child of 1989, Article 40 (3) b), emphasized by Comment No. ten on Children's Rights in Juvenile Justice from 25 April 2007; the United Nations Standard Minimum Rules for the Administration of Juvenile Justice [Beijing Rules] of 1985, Rules No. 11.1–11.4; the United Nations Guidelines for the Prevention of Juvenile Delinquency [Riyadh Guidelines] of 1990, Rules No. 5 and 6; the United Nations Standard Minimum Rules for Non-custodial Measures [Tokyo Rules], Rule No. 5 and on the European level; the Council of Europe's Recommendation on "New ways of dealing with juvenile delinquency and the role of juvenile justice" of 2003 [Rec 2003 (20)], Rules 7, 8, and 10; and the European Rules for Juvenile Offenders Subject to Sanctions or Measures [ERJOSSM, Rec (2008) 11], Rules 5, 10, and 12).

The concept of nonintervention (or better avoiding formal prosecution) was developed in combination with decriminalization (particularly of so-called status offences) and deinstitutionalization (from youth custody and residential homes). Since the 1960s particularly in North America and across Europe, tendencies in juvenile criminal policy have emerged that are based on the notions of the principles of "subsidiarity" and "proportionality" of state interventions against juvenile offenders (Dünkel 2009). More specifically, these developments also involve the expansion of procedural safeguards on the one hand and the limitation or reduction of the intensity of interventions in the field of sentencing on the other. One major element of this philosophy was the idea of diversion, i.e., to avoid possibly stigmatizing state interventions in favor of a more lenient and – with regard to future social integration – more appropriate approach.

In spite of heavy criticism in the early 1980s, blaming "net-widening" effects and informal social control that would even surpass formal social control of the youth courts (Austin and Krisberg 1982), diversion has continued its "triumphant" expansion due to national and international developments in juvenile crime policy in the 1980s. Nowadays, apart from the aim to avoid (unnecessary) stigmatization, diversion in juvenile justice is based on the idea that education should be prioritized over punishment. From the perspective of sociology of law, the advantage of nonintervention or less severe punishment (e.g., probation instead of imprisonment) lies in the increased expectations of future norm conformity, which are expressed by the competent punishing authority to the offenders in question. The violator of the norm is under the pressure of a special (informal) obligation as he has been given a "social credit" which contributes to better compliance with the norm (see Dünkel et al. 2011b, p. 1628 with further references.).

Furthermore, the use of diversion is oftentimes related to the pragmatic consideration of reducing or limiting the courts' caseload (see in general Jehle and Wade 2006). It can be shown that an increase of cases in the criminal justice system needs to be compensated by diversionary or other bureaucratic strategies that make the "input" manageable.

Types of Diversion

Types of diversion can be distinguished by the different authorities who are competent to decide on the dismissal of the case (diversion). Diversion can be initiated by the police, the prosecutor, or the judge.

A young offender's first contact with the justice system is mostly the police. This is why police diversion is generally recommended by the international instruments (see, e.g., No. 11.2 of the Beijing Rules or No. 5.1 of the Tokyo Rules). The advantage of diversion at this early stage of the procedure is that the police can react promptly, so that there is immediacy between committing the offence and the justice system's response. On the one hand, a swift reaction is seen positively from a pedagogic point of view (if the police act cautiously and with respect). On the other hand, the amount of time during which the offender could be stigmatized is reduced. The police have a lot of discretional power in such a system. The resulting dangers could be prevented through the condition that the police have to undergo specific training on contact with young offenders.

Consequently, many juvenile justice systems, mostly in countries where the role of the public prosecution service is not very strong, introduced the possibility of "police diversion." "Police diversion" can mean that the police, to whose attention an offence committed by juveniles has come, is competent to simply take no further action because the behavior in question is viewed as being very minor, petty, and unimportant (e.g., *Cyprus, England/Wales*). In other systems the police can issue informal or formal warnings or can refer the juvenile to special diversion schemes in close cooperation with the Social Services (e.g., *Northern Ireland, Ireland, the Netherlands*).

Other countries strictly follow the procedural principle of "legality." Originally the principle of legality means that a human conduct must be declared as a crime by a specific statute or law before it can be considered as a criminal act. The strict consequence from this principle is that a special prosecution service must bring a charge against any criminal offender. If the laws in these countries provide for any form of pre-court diversion, usually the prosecution services, consisting of lawyers, are competent to decide about dismissing the case or to refer it to a special diversion scheme. Usually the decisions of the prosecution services are conditional. This means that the prosecution service offers to close the criminal proceedings when the offender has fulfilled special obligations like repairing the damage or participating at anti-aggression trainings.

In some countries, diversionary decisions can also be made by a judge at the court level. Such court diversion makes sense particularly in cases when the juvenile has paid reparation or has otherwise resolved the conflict with the victim after the prosecutor has submitted the indictment to the court, which gives the judge the impression that further prosecution would not be appropriate or necessary. Especially the countries following the traditional welfare model have facilitated diversionary strategies on a court level because of the wide discretionary power of the juvenile judge.

In consideration of the different competent authorities, juvenile justice systems generally introduced a way to divert a case and combine the conditional dismissal of the case with educational measures or interventions. The systems for this kind of "interventionalist diversion" vary.

In many countries, minor offences can be dismissed after educational measures have taken place (e.g., mediation, victim-offender reconciliation, reparation, apology to the victim). The idea is that if the conflict is already solved within the society, there is no need to further stigmatize the offender or to spend much money on cost-intensive criminal procedures. So, for example, if in *Germany* at the time of the beginning of the prosecution the offender has already apologized and paid for the damage he produced, the public prosecutor can dismiss the case without any further conditions (the German Juvenile Justice Act even encourages the prosecutor to regularly do so; see Sect. 45 al. 2 JJA).

A similar approach is followed in countries where diversion in combination with (minimum) educational interventions is seen as a possible option. One alternative is that the prosecutor or the judge can suspend the case for a certain period of time. The case will be dismissed after the offender has fulfilled special obligations, like community work or reparation of the damage or participation at certain "training courses." A comparable approach is quite common in Eastern European countries with the so-called release from criminal liability, which can be combined with educational measures.

In some countries, diversion can be combined with a referral to the Social Services (*Sweden*) or special administrative authorities/bodies, like the "Children's Hearings System" (*Scotland*) or the "Juvenile Commissions" in *Bulgaria*, *Estonia*, or *Russia*. These bodies can partly issue and partly negotiate the fulfillment of special educational obligation. The transfer of responsibility and sanctioning power to administrative bodies makes it important to guarantee that juvenile justice standards are respected on this level as well. According to the international recommendations, these kinds of administrative authorities likewise have to avoid each form of deprivation of liberty as a reaction to criminal behavior. Questionable is if "Juvenile Commissions" as they can be found in Eastern and Central Europe constitute an appropriate way of diversion. Not always the commission's decisions are subject to judicial review, even if they may include the placement of a juvenile in a closed institution (UNICEF 2008, p. 25). An interesting approach in this sense is the idea of introducing conferences to divert juvenile offenders from criminal proceedings which is apparently most developed in *New Zealand*: The police have to refer all juvenile offenders who have not been arrested and charged to a family group conference. If the conference can resolve the matter, there will be no further public reaction. The youth court is required to refer all cases coming to it for a family group conference as well (Maxwell and Morris 2006, the European approaches for the introduction of family conferences are described by Doak and O'Mahony 2011).

Efficiency of Diversion

Measuring the efficiency of sanctions is one of the most complicated issues in criminological research. Usually different sanctions are applied for different kinds of offences and offenders. This leads in practice regularly to a selection bias which makes it almost impossible to ascribe recidivism (or legal behavior) to the inefficiency (or efficiency) or the sanction (validity). Reducing a desired behavior to a special sanction is only possible with the help of studies using random assignment or quasi-experimental design (standards for evaluations were developed by Sherman et al. 1998; see also the chapters of MacKenzie and Shapland in this volume). Most study designs do not follow these standards, as it is in a criminal justice system difficult to allocate one or another sanction randomly.

With regard to diversion, there are at an international level study results which indeed used random assignment or quasi-experimental design which predominantly arrived at the conclusion that recidivism rates after diversion are lower, or at least not higher, than after formal court procedures and convictions. According to German studies, the strategy of expanding informal sanctions has insofar proved to be an effective means, not only to limit the juvenile court's workload, but also with respect to special prevention (see Heinz 2005; Dünkel 2011).

According to German data, reconviction rates of offenders who were "diverted" instead of being formally sanctioned are significantly lower (reoffending rates after a risk period of 3 years were 27 % vs. 36 %, see Dünkel 2011). Even for repeat offenders the reoffending rates after informal sanctions were not higher than after formal sanctions (see Heinz 2005, p. 306). Another study demonstrates that the increase in the use of diversion in Germany during the 1980s and 1990s does not correspond to an increase of juvenile delinquency rates. On the contrary, the recidivism rates of comparable delinquents (for different typical juvenile delinquent acts) were significantly lower when diverted as compared to those formally sanctioned by the youth court (see Dünkel et al. 2011b, p. 1639 with further references). The evident methodological problems of comparing different sanctions (concerning the seriousness of different crimes, previous convictions, etc.) were addressed by British empirical research which strictly controlled the different "sanction groups" for key variables such as age, sex, and previous criminal history. The research results demonstrate that conditionally discharged offenders had lower reconviction rates (39 %) than those sentenced to fines (43 %), probation (55 %), or community service (48 %, see Moxon 1998, p. 91). The meta-analysis of *Whitehead* and *Lab* supports the findings that diversion has the best prognosis with regard to recidivism. As far as can be seen, only the study of Morton and West (1983) could not find any reduction in recidivism from the use of youth diversion as opposed to youth courts (in *Canada*).

Looking at the costs and the impact of different sentences and interventions, it is evident that informal warnings and cautions are the least expensive measures. Further research on "what works, with whom under which circumstances" including serious control of the selection bias needs to be continued in this context.

Community Sanctions

Definition and Functions of Community Sanctions

If a case is not diverted but reaches the level of the court, many juvenile justice systems provide a lot of different dispositions as well. The so-called community sanctions define all sanctions and measures which do not lead to deprivation of liberty in any form (see Beijing Rules, No. 18). They are important as a basis to comply with the principle that no child shall be deprived of liberty except as a "last resort," as required by Article 37(b) of the Convention on the Rights of the Child. The reform movement to widen the scope of community or alternative sanctions started in the 1970s in the *United States* and *England* and later developed in the 1980s in many other juvenile (and adult criminal) justice systems. The base for the introduction of community sanctions is seen as twofold: On the one side a shift in people's attitude towards punishment emerged which tried to make criminal justice more humane (Junger-Tas 1994, p. 1). On the other hand, there was the problem of overcrowded prisons in many countries which forced to find and to use alternatives (e.g., Bala and Roberts 2006, S. 37). If we look at today's different juvenile justice systems, we do find many differences regarding the extent and types of community sanctions.

Types of Community Sanctions

As a general rule, the applicable sanctions and measures follow a certain hierarchy that is based on the order in which priority shall be given to the most educational, most appropriate sanction. This regularly opens up the possibility to combine several educational measures or sanctions with each other. We can find the following levels of sanctioning, ordered from the least to the most intrusive:

1. Warnings, reprimands, conviction without sentence, educational "directives"
2. (Day) fines, community service, reparation orders, mediation
3. Social training courses and other more intensive educational or supervision sanctions
4. Mixed sentences, combination orders (which can be characterized as a more "repressive" (intrusive) way of dealing with juvenile offenders)
5. Suspended sentences without supervision by the Probation Service
6. Probation
7. Suspended sentences with supervision by the Probation Service, electronic monitoring
8. Educational residential care, youth imprisonment, and similar forms of deprivation of liberty

The least invasive sanctions are warnings or reprimands (verbal sanctions) and followed by a wide range of alternative sanctions that exert more or less influence on the life of the offender. Many sanction systems provide educational measures (such as educational "directives" in *Austria* and *Germany*) either as independent sanctions or as complementary elements of other sanctions like probation or suspended prison sentences (e.g., *Denmark*). The aim of such educational directives is always to improve the educational impact on the one hand and to reduce the impact of risk factors in the juvenile's daily life on the other. The laws should confer a certain degree of discretionary power on the judge to enable him or her to find the most appropriate directive. In between we find the possibility to impose a day fine on juvenile offenders which is theoretically possible in many, albeit not all juvenile justice systems. Indeed, one may question if fines could be seen as educational sanctions, facing the fact that juveniles will often be unable to pay for fines with their own money.

Restorative justice elements like victim-offender mediation for juveniles do often play a special role in juvenile justice. Yet, in some (European) countries, mediation is never or only seldom practiced due to a lack of organizational infrastructure at the local level, as reported by the *Czech Republic*, *Poland*, *Romania*, and *Serbia* (Dünkel et al. 2011a). Such kinds of "restorative justice orders" can be manifold and creative in aiming at compensating the victim.

In many countries, community service combines slight "punishment" with reparative and rehabilitative elements. The offender shall offer "a 'payback' to the community via unpaid work" (Goldson 2008, p. 78). Some countries have special age limits for the imposition of community service: For example, in *England* and *Wales*, *Ireland*, and *Northern Ireland*, community service can only be imposed on juveniles aged 16 or older. Huge differences can be observed with respect to the maximum number of hours: The limit lies between 30 h in *Belgium* and 250 h in *Canada*, 300 h in Denmark, or even 400 h in New Zealand (see Dünkel and Lappi-Seppälä in this volume). The differences in legislation are partly due to different approaches and settings for community service orders: For example, in *Finland* a high number of hours (regularly only for young adults aged 18–20) will replace a sentence of up to 8 months of unconditional imprisonment (Dünkel et al. 2011a, p. 1647).

Different from community service is the sentence of "corrective labor," which can be found primarily in Central and Eastern Europe. Corrective labor may be imposed on a juvenile offender at the place of his/her regular employment for a particular length of time. In the course of corrective labor, deductions from the offender's earnings shall be made in favor of the state in the amount specified in the court ruling within a certain limit (UNICEF 2008, p. 30).

Many countries have successfully implemented creative and constructive measures such as social training courses (*Germany*) or so-called labor and learning sanctions or projects (the *Netherlands*), where the juveniles can learn to deal with their aggressive potential or where they can be trained according to their personal skills.

Some countries have introduced high intensive supervision or educational orders, e.g., special "centers" to which juvenile offenders can be sent for a few hours a day. In *England* and *Wales*, the attendance center order requires a young person to be present at a (usually) police-run institution on Saturday afternoons, where juveniles engage in physical education and other activities designed to inculcate a sense of discipline or social skills, for sessions up to a maximum of 36 h. In *Kosovo*, the court can commit a minor to a disciplinary center for a maximum of 1 month (for up to 4 h per day) or for a maximum of 4 days of a school or public holiday (for up to 8 h per day). In *France*, the law of 5 March 2007 created a new educational measure, *activities during the day* (*mesure d'activités de jour*), in which the juvenile is involved in vocational or school insertion activities at a public or qualified private institution or agency. In *Italy*, the magistrate can order the minor to carry out study or work activities in special working groups. Looking at the strict definition above it may be questionable if these measures fall into the scope of "community sanctions." However, because they are used in practice to avoid imprisonment or comparable forms of deprivation of liberty in predominant closed institutions, they deserve to be listed within this category. This is not the case for short-detention centers, where juveniles can be sent to for some days or weeks with the aim to deter the offender via a "short sharp shock." Even if some jurisdictions try to avoid "real" imprisonment with the introduction of this short-term incarceration, this however falls within the definition of deprivation of liberty as it takes the juvenile out of his social surroundings and endangers him with the negative aspects of stigmatization and other negative outcomes of imprisonment.

In many countries, supervision or surveillance orders play a special role as alternatives to imprisonment. In most countries, the Social Service or the Probation Service is responsible for the execution of these measures, but in some countries, juvenile offenders are usually supervised by the legal representative, normally the parents (e.g., "house arrest" in *Italy*). The aim is to avoid isolating the minor from his/her familiar and social surroundings in order to prevent disturbances to his/her personal development.

In *Finland* the Juvenile Punishment Order consists of work programs, supervision, and activity programs that aim to promote social adjustment, the person's sense of responsibility, and his/her social relations. There is a strict requirement that this sentence can only be issued in high-risk cases. This requirement may prevent net-widening effects as the Juvenile Punishment Order is definitively only applied in cases of repeat offenders who have already been sentenced to conditional imprisonment.

Contrarily, in *England and Wales* the introduction of the so-called referral order could well be having a net-widening effect, since it is more invasive and rigorous than the conditional discharge that it has essentially replaced in practice. "Action plans" or "referral orders" in contrast to the *Finnish* "Juvenile Punishment Order" follow a more punitive approach (Dünkel et al. 2011a, p. 1651 with further references).

In some juvenile justice systems, it is possible to confiscate a person's driver's license or to issue a prohibition from driving a vehicle as independent sanction or measure. In these cases the courts have to consider a certain susceptibility to unequal treatment, because there are special groups of juveniles or young adults who are more dependent on driving a car than others (due to work obligations, poor local infrastructure, etc.). There are serious reservations against the temporary withdrawal of a driver's license as a standalone sanction, especially if it is used for other than only traffic-related offences. The future integration of juveniles is often more difficult when their mobility is hampered. Therefore, educational efforts should be made to allow juveniles to participate in traffic in a responsible manner.

Social traffic training courses seem to be the appropriate answer, rather than excluding juveniles from mobility – particularly when they live in rural areas (Dünkel et al. 2011a).

Many juvenile justice systems provide suspended juvenile prison sentences that frequently go hand in hand with supervision by the Probation Service or a similar service with a social work approach. The "Continental European Model" of suspended sentences implies the imposition of a youth prison sentence, the execution of which is not immediate. Should an offender fail to meet the conditions of probation, suspension (as a last resort) can be revoked and the juvenile serves the term of imprisonment set at the first trial (e.g., in *Austria*, *Bulgaria, Germany*, or *Spain*). Other criminal systems like in the *United States* or the states of the *United Kingdom* introduced probation as a special sanction. This sanction is – as its name indicates – always connected with support from and control by the Probation Service. Contrary to the "Continental European Model," in these countries no term of detention is fixed. Therefore, where an offender fails to comply with his or her probationary requirements, the term of imprisonment is determined in a second sentencing trial. Here one finds another explicit example for how the same terms can mean different things in international comparative analyses: Many countries use the term "probation" to describe the "Continental European" approach of "suspended sentences with supervision."

Apparently no juvenile justice system has managed to totally avoid imprisonment or detention for juveniles. Many different forms of deprivation of liberty with corresponding institutions can be found worldwide, like youth prisons, detention centers, closed educational care, or schools "for juveniles with special needs" (see Dünkel and Stańdo-Kawecka 2011).

Especially researchers and practitioners from Central and Eastern European countries claim that oftentimes the laws in their countries do provide a lot of alternative sanctions, but the judges or prosecutors do not apply them. The reason mostly lies in the lack of infrastructure. The law might allow for victim-offender mediation – but if no organization or no service is available to offer mediation, this promising new approach will gain no importance in practice. Another problem related to the implementation of alternative sanctions is oftentimes that the question of funding is not responded or clear.

Efficiency of Alternative Sanctions

Research results on the efficiency of alternative sanctions are not as clear as for diversion (see above) and share the same problems with respect to the selection bias. However, there is some evidence that alternative sentences "work" better than liberty-depriving sentences with regard to recidivism, but the results request for differentiations: For example, Latimer et al. (2001) or Sherman and Strang 2007 showed that restorative justice programs can reduce recidivism, but there is a wide variation in their effects (see also the chapters of MacKenzie and Shapland in this volume). In general, international results show that positive effects rather are to be expected from programs comprising behavior therapy, oriented towards social learning, and tailored to the needs of the offender than from punitive sentencing and/or imprisonment (e.g., Murphy et al. 2010 with further references; MacKenzie in this volume). In contrast, programs focusing on discipline or deterrence through fear of consequences showed negative or minimal positive effects (Lipsey and Howell 2012, p. 517 with further references). Not surprisingly alternative sanctions have proven to be cost-effective compared to the immediate and belated costs of imprisonment (Aos 2006; UNODC 2007).

The Relevance of Diversion and Community Sanctions in Practice

Oftentimes the country's "law in the book" does not correspond to the "law in practice." To investigate whether the countries use their manifold possibilities in sentencing juvenile offenders is not always an easy task: Apart from the general problems that arise when working with statistical data about crime and reactions to crime in a large number of countries (e.g., offences are in some countries registered in relation to the offence, in other countries in relation to the offender), the collection of data about sentencing in a juvenile justice system varies a lot. Sometimes an absence of reliable statistical records can be observed (see Goldson and Muncie 2006, p. 2), and even where statistical records do exist, practice of how crimes (and clear-ups) are recorded varies greatly (see Cavadino and Dignan 2006, p. 4). Thus, comparability can only be achieved through lots of interpretation. This makes an international comparison of statistical data difficult enough. The above-described wide variety of alternative sanctions in the different juvenile justice systems complicates matters further, and the different age groups that are covered by the different juvenile justice systems all over Europe additionally hinder comparability. With regard to the use of alternative sanctions in practice, it is possible to present some structural tendencies (for Europe see Dünkel et al. 2011b). Diversion has experienced a triumphant expansion in many European countries such as *Austria*, *Germany*, *Ireland*, the *Netherlands*, *Northern Ireland*, *Romania*, *Slovenia*, *Spain*, and *Sweden*, where more than 50 % and up to 70 % (Germany) or even about 80 % (*Northern Ireland*) of cases involving juvenile offenders are diverted. Other countries like *New Zealand*, *Scotland*, or *Sweden* regularly refer juvenile offenders to special institutions, conferences, or social services.

Many of those countries who do not make extensive use of diversion, primarily apply court-based community sanctions, e.g., in the *Czech Republic*, *Spain* (Catalonia), and *Switzerland*. *Slovenia* is an extreme case as in 98 % of court decisions educational measures are applied. The same is true for *Serbia* (95 %). In many Central and Eastern European countries, the suspended prison sentence is still the predominant community sanction, often because of a lack of infrastructure for other, more educational alternatives (e.g., in the *Czech Republic*, *Hungary*, *Latvia*, *Russia*, and *Slovakia*). The *United States* are infamous to sentence young offenders rather based on notions of punishment and accountability than rehabilitation, but alternative sanctions are still seen on the rise (Bishop and Decker 2006, p. 29).

Still there are countries where custodial sentencing is obviously seen as a promising answer to juvenile offending and of considerable importance in practice. *Bulgaria*, *Lithuania*, *Romania*, *Russia*, and *Spain* (particularly Catalonia) could be classified as belonging to this group. In *Bulgaria* traditionally 80–90 % of court disposals had been sentences to imprisonment, but after the law reform of 1999, the proportion of prison sentences for juvenile offenders dropped to “only” 47 % (2005). In *Romania* and *Russia*, however, compared to the Soviet time, decreasing proportions of custodial sanctions are evident. In *Spain* increasing numbers of custodial sanctions have been imposed only recently. As indicated above, some countries, mainly from Central and Eastern Europe, report that they have introduced a wide variety of alternative sanctions which are not in use to a greater extent in practice due to the missing infrastructure and unclear funding.

Conclusion

Juvenile justice systems reveal numerous possibilities to divert juvenile offenders from the criminal justice system or to offer them alternative sanctions. In many countries lawmakers and practitioners seem to be convinced that juveniles shall be saved from the socially detrimental outcomes of imprisonment, and the international recommendations which request the use of custodial sanctions only as a last resort demand the criminal justice systems to seek for effective alternatives.

As available research results allow for the gently conclusion that diversion “works” in means of the reduction of recidivism, data records on the efficiency of alternative sanctions do not allow for entirely clear positive conclusions. However, is it the correct way to ask the alternatives to imprisonment to proof their efficiency when we definitely know that imprisonment or comparable ways of custodial sanctions/measures are not at all adequate to reduce recidivism? Alternative sanctions violate human rights less than imprisonment does and are to be preferred due to the minimum intervention principle (UNODC 2007; see, however, to human rights issues in this area the chapter of Morgenstern and van Zyl Smit in this volume). And even if research shows that the cost-effective sanctions are not always superior in preventing reoffending, it is true that it is attractive for criminal justice systems to make a wide use of alternative sanctions from an economic point of view as well. This is why diversion and alternative sanctions should be extended or, where available, used more frequently in practice. Further research with random assignment or quasi-experimental design would be desirable to support the triumphal procession of alternatives to imprisonment in juvenile justice systems.

Related Entries

- ▶ Community Service in Europe
- ▶ Examining the Effectiveness of Correctional Interventions
- ▶ History of Juvenile Justice
- ▶ International Human Rights Standards and Community Sanctions
- ▶ Juvenile Diversion
- ▶ Juvenile Justice in the Get Tough Era
- ▶ Probation and Community Sanctions

Recommended Reading and References

Aos S (2006) Evidence-based public policy options to reduce future prison construction, criminal justice costs, and crime rates. http://www.wsipp.wa.gov/rptfiles/06-10-1201.pdf

Austin J, Krisberg B (1982) The unmet promise of alternatives to incarceration. Crime Delinq 28:374–409

Bala N, Roberts JV (2006) Canada's juvenile justice system: promoting community-based responses to youth crime. In: Junger-Tas J, Decker S (eds) International handbook of juvenile justice. Springer, Berlin/New York, pp 37–64

Bishop DM, Decker SH (2006) Punishment and control: juvenile justice reform in the USA. In: Junger-Tas J, Decker S (eds) International handbook of juvenile justice. Springer, Berlin/New York, pp 3–35

Cavadino M, Dignan J (2006) Penal systems. A comparative approach. Sage, London/Thousand Oaks/New Delhi

Doak J, O'Mahony D (2011) Developing mediation and restorative justice for young offenders across Europe. In: Dünkel F, Grzywa J, Horsfield P, Pruin I (eds) Juvenile justice systems in Europe – current situation and reform developments, 2nd edn. Forum-Verlag, Bonn-Bad Godesberg, pp 1717–1746

Doob AN, Tonry M (2004) Varieties of youth justice. In: Tonry M, Doob AN (eds) Youth crime and justice. University of Chicago Press, Chicago/London, pp 1–20, Crime and Justice Vol. 31

Dünkel F (2009) Diversion – a meaningful and successful alternative to punishment in European juvenile justice systems. In: Junger-Tas J, Dünkel F (eds) Reforming juvenile justice. Springer, Berlin/New York, pp 147–163

Dünkel F (2011) Germany. In: Dünkel F, Grzywa J, Horsfield P, Pruin I (eds) Juvenile justice systems in Europe – current situation and reform developments, 2nd edn. Forum-Verlag, Bonn-Bad Godesberg, pp 547–622

Dünkel F, Stańdo-Kawecka B (2011) Juvenile imprisonment and placement in institutions for deprivation of liberty – comparative aspects. In: Dünkel F, Grzywa J, Horsfield P, Pruin I (eds) Juvenile justice systems in Europe – current situation and reform developments, 2nd edn. Mönchengladbach, Forum-Verlag, pp 1789–1838

Dünkel F, Grzywa J, Horsfield P, Pruin I (eds) (2011a) Juvenile justice systems in Europe – current situation and reform developments, 2nd edn. Forum-Verlag, Mönchengladbach

Dünkel F, Pruin I, Grzywa J (2011b) Sanctions systems and trends in the development of sentencing practices. In: Dünkel F, Grzywa J, Horsfield P, Pruin I (eds) Juvenile justice systems in Europe – current situation and reform developments, 2nd edn. Forum-Verlag, Mönchengladbach, pp 1649–1716

Goldson B (2008) Dictionary of youth justice. Willan Publishing, Cullompton/Devon.

Goldson B, Muncie J (2006) (eds) Comparative youth justice. Sage, London, Thousand Oaks, New Delhi

Hartjen CA (2008) Youth, crime and justice: a global inquiry. Rutgers University Press, New Brunswick

Heinz W (2005) Zahlt sich Milde aus? Diversion und ihre Bedeutung für die Sanktionspraxis. Teil 1 und 2: Zeitschrift für Jugendkriminalrecht und Jugendhilfe. 16, p 166–178, 302–312

Jehle J-M, Wade M (2006) Coping with overloaded criminal justice systems. The rise of prosecutorial power across Europe. Springer, Berlin

Junger-Tas J (1994) Alternatives to prison sentences. Experiences and developments. Kugler Publications, Amsterdam

Junger-Tas J, Decker S (eds) (2006) International handbook of juvenile justice. Springer, Berlin

Junger-Tas J, Dünkel F (2009) Reforming juvenile justice: European perspectives. In: Junger-Tas J, Dünkel F (eds) Reforming criminal justice. Springer, Berlin, pp 215–233

Koffmann L, Dingwall G (2007) The diversion of young offenders: a proportionate response? In: Web journal of current legal issues. http://webjcli.ncl.ac.uk/2007/issue2/koffman2.html

Latimer J, Dowden C, Muise D (2001) The effectiveness of restorative justice processes: a meta-analysis. Research and Statistics Division, Department of Justice, Ottawa

Lipsey MW, Howell JC (2012) A broader view of evidence-based programs reveals more options for state juvenile justice systems. Criminol Public Policy 11:515–523

Maxwell G, Morris A (2006) Youth justice in New Zealand: restorative justice in practice? J Soc Issues 62(2):239–258

Morton M, West G (1983) An experiment in diversion by a citizen committee. In: Corrado R, LeBlanc M, Trepanier J (eds) Current issues in juvenile justice. Butterworths, Toronto

Moxon D (1998) The role of sentencing policy. In: Goldblatt P, Lewis C (eds) Reducing offending: an assessment of research evidence on ways of dealing with offending behaviour. Home Office Research Study 187, London, pp 85–100

Murphy P, McGinness A, McDermott T (2010) Review of effective practice in juvenile justice. http://www.djj.nsw.gov.au/pdf_htm/publications/general/Juvenile Justice Effective Practice Review FINAL.pdf

Pruin I (2011) The scope of juvenile justice systems in Europe. In: Dünkel F, Grzywa J, Horsfield P, Pruin I (eds) Juvenile justice systems in Europe – current situation and reform developments, 2nd edn. Forum-Verlag, Mönchengladbach, pp 1539–1582

Sherman LW, Strang H (2007) Restorative justice: the evidence. Smith Institute, London, http://www.smith-institute.org.uk/pdfs/RJ_full_report.pdf

Sherman LW, Gottfredson D, MacKenzie D, Eck J, Reuter P, Bushway S (1998) preventing crime: what works, what doesn't, what's promising. http://www.ncjrs.org/works/download.htm

UNICEF (2008) Lost in the justice system. Children in conflict with the law in Eastern Europe and Central Asia. http://unicef.hu/documents/10552/30013/Lost_in_the_Justice.pdf

United Nations Office on Drugs and Crime [UNODC] (2007) Handbook of basic principles and promising practices on alternatives to imprisonment. http://www.unodc.org/pdf/criminal_justice/07-80478_ebook.pdf

Whitehead JT, Lab SP (1989) Meta-analysis of juvenile correctional treatment. J Res Crime Delinq 26(3):276–295

Winterdyk JA (2002) Juvenile justice systems: international perspectives, 2nd edn. Canadian Scholars' Press, Toronto

Lawful Killings

Anna Alvazzi del Frate[1], Giulia Mugellini[2], Irene Pavesi[1] and Takhmina Karimova[3]
[1]Small Arms Survey, Geneva, Switzerland
[2]Institute of Criminology, University of Zurich, Zurich, Switzerland
[3]Geneva Academy of International Humanitarian Law and Human Rights, Geneva, Switzerland

Synonyms

Dowry deaths; Executions; Honor killings; Legal interventions; Mob killings; Witchcraft

Overview

Homicide is frequently used as an indicator of violence. Its statistical superiority originates from the seriousness of this crime and from a broad availability of relevant data from law enforcement and criminal justice sources. Definitions of homicide for statistical purposes tend to focus on the intentionality and premeditation of death inflicted on a person by another person and exclude unintentional acts (non-culpable – manslaughter). Nevertheless, definitions of homicide across countries show marked differences, depending on penal codes and what is considered unlawful. Indeed, different cultures tend to criminalize different acts and behaviors.

A cross-country comparison of data based on "homicide," therefore, should take into account that the indicator includes the elements which are criminalized in each country and may exclude others, which are not.

Behaviors defined in this entry as "lawful killings" are acts with lethal consequences which are tolerated by some cultures and therefore do not appear in some homicide statistics. This is the case with "honor" killings, which are not criminalized in several countries.

The term "lawful killings" is used throughout the entry to indicate many different types of killings which in some contexts are considered "justified," or less serious from a legal point of view than plain killings. This may happen because of cultural, ethical, political, or legal reasons. Depending on countries and cultures, this could be the case with any of the following: killing in self-defense; killing in revenge or retaliation; killing as a result of provocation; killing to eliminate the pains of incurable patients; killing to defend the "honor" of a person or a family. In some cases, States take in their hands the authority to kill to defend the public order (legal interventions) or to punish someone for having killed someone else (death penalty).

This entry focuses on some specific forms of killings, namely, "honor" killings, dowry deaths, killing of persons accused of witchcraft, and mob killings, and how they relate with homicide concepts and categories. The current debate highlights two sides of the problem: on the one hand, in some countries/contexts these acts still fail to be recognized as forms of violence, sometimes even as crimes, depending on specific traditions and beliefs. Perpetrators of these killings frequently get away with committing these crimes, or enjoy more lenient investigations and sentences. On the other hand, in the countries where these killings are part of the homicide category, the seriousness of these acts may be insufficiently emphasized, and some advocate for perpetrators to be subject to sanctions harsher than those inflicted for homicides.

The entry also considers some examples of data collection or specific studies aimed at capturing relevant data to measure the extent of these phenomena.

What Is in a Label: "Homicide," Criminalization, Culture, and Human Rights

Criminalization

The United Nations Survey of Crime Trends and the Operations of Criminal Justice Systems (UN-CTS), through which UN Member States exchange crime and criminal justice statistical information, defines homicide as "unlawful death inflicted on a person by another person," which could be intentional or not.

What is considered "unlawful" depends on the definition of homicide in domestic criminal law and procedure, as well as the specific legal classification of homicide in each country. The degree of "unlawfulness" and the type of sanction foreseen for specific forms of killing vary across societies "to the extent to which different countries deem that a killing could be classified as such" (United Nations Office on Drugs and Crime 2011, p. 10).

According to Waters (2007, pp. xiii–xiv), "a murder is a social act that involves not only killing, but also judgment and evaluation by society at large" but, still, "[e]ach society has a process through which killings are described as legal justifiable homicide, illegal unjustified homicide (i.e. murder), or just plain killings. The process varies not only with a technical or legal capacity to assign blame but also with the capacity of a particular society to respond in a manner it defines as decisive and appropriate."

The process of defining a particular killing as "unlawful" involves the social context where the act is committed as well as the traditional and cultural codes of the victim and the perpetrator. Some killings may be considered as "lawful," mitigating circumstances invoked and, in some cases, impunity granted to the perpetrators, because the society at large does not consider they have committed any illegal acts. It can therefore happen that killings which are criminalized in some countries may not be in others. Boundaries for such considerations are represented by the cultural and sociological perception of specific issues such as honor, status distinctions, culpability, innocence and blame, as well as the state's approach to legitimate use of force.

In some societies, perpetrators of killings in the name of "honor" or "morality" still get reduced penalties and/or may be exculpated. This happens because the killing is "publicly justified" according "to a social order claimed to require the preservation of a concept of honor vested in male (family and/or conjugal) control over woman" (Welchman and Hossain 2006, p. 4). As sustained by Wolfgang and Ferracuti's "subculture of violence theory," within some specific ethnic groups and when the values of "honor" and "morality" are strongly perceived and shared in the society, one's capacity for violence may not only be admired but also perceived as an indispensable trait of personality.

Culture

The ecological approach to human development has been applied to criminology to explain the roots of violent crimes across different societies. It highlights the influence of community contexts in which social relationships are embedded and of the larger societal factors, such as cultural norms and attitudes, in explaining the level and dynamics of some specific forms of lethal violence.

Following this approach, homicide can be seen as a social and cultural construct in which the level of "seriousness" and "unlawfulness" can vary not only across different countries but also across different communities within the same country. Because of the way countries are composed and as a result of migration processes, most countries include a variety of ethnic and religious groups, even where a majority of the population shares the same culture.

The influence of culture is crucial to determine which behaviors are criminalized and which are not, reflecting majoritarian cultural values at the time when legislation is passed. Changes in culture and social acceptance of specific values could influence the consideration of specific behaviors as unlawful or not.

Analysis of the World Values Survey (WVS) results (a worldwide investigation of sociocultural and political change, conducted by a network of social scientists at leading universities on national samples of at least 1,000 people in 97 societies in all 6 continents) reveals that many of the basic cultural values closely relate with each other and can be grouped along two major cross-cultural dimensions, namely, one going from "Traditional" to "Secular-rational" values and one from "Survival" to "Self-expression" values.

Moving from "Traditional" to "Secular-rational" values reflects the shift from traditional religious values to secular-rational attitudes. Moving from "Survival" values to "Self-expression" reflects the shift of individual focus from personal and economic security to personal self-expression and quality of life.

Self-expression values, which focus on subjective well-being, freedom to express oneself, and quality of life, may give high priority to environmental protection, promotion of gender equality, tolerance of different characteristics and behaviors (such as migrants, foreigners, gays, and lesbians), and rising demands for participation in decision-making in economic and political life. Moreover, societies that rank high on self-expression values also tend to rank high on interpersonal trust. These types of society may therefore be much less tolerant with regard to any form of deprivation of life and crimes against the individual. Traditional and survival values stress the importance of parent–child ties and traditional family values and reject divorce, abortion, euthanasia, and suicide.

Many of these factors could have a strong influence on some types of lethal violence perpetrated within families. Societies and communities based on strong traditional values could also show higher levels of tolerance with regard to specific types of killings committed in the name of the family's respect and honor (Luopa 2010).

Nevertheless, as a result of globalization, there may be no clear-cut attribution of specific cultural values to specific countries. Therefore, different communities based on very different cultural values and norms may coexist within the same country.

Some types of killings could be "justified," or even considered "necessary" under specific circumstances, by some types of cultural values. More in-depth knowledge of causes and reasons for such considerations may allow preventing and contrasting these forms of lethal violence by focusing on their cultural roots.

Citizens of Western countries still tend to believe that "honor crimes" are a prerogative of Asian and African cultures, while this phenomenon is visibly increasing also in Europe (particularly in France, Sweden, the Netherlands, Germany, UK, and Turkey) and in the US (Council of Europe Parliamentary Assembly 2009, p. 6;

Chesler 2010, p. 3). For example, a recent study found a correlation between higher homicide rates observed in the US South and "culture of honor" as a heritage of herding carried out by early settlers, indicating that "the Southern propensity for homicide stems largely from the cultural background of the early Southern settlers alone and, in any case, is fully accounted for by the cultural background of early settlers together with current economic conditions and demographic characteristics" (Grosjean 2010, p. 20).

Indeed "honor" crimes disappeared from criminal codes in Europe (especially Western Europe) and in the US a long time ago and events of similar nature may be currently registered under the label of "domestic violence." However, their characteristics, motivations, and evolution are completely different. For example, "unlikely domestic violence, honor killings often involve multiple family members as perpetrators" (Chesler 2010, p. 3).

A proper assessment of the extent and distribution of "lawful" killings across different countries is necessary in order to properly prevent, contrast, and sanction them. In this respect, proper recognition of different cultural settings and understanding of how such data would be contributing in terms of policy relevance are necessary. For example, the proposal for an international classification of crime for statistical purposes currently discussed at the UN level considers the categories of *infanticide* and *euthanasia* separately as specific types of intentional homicide (UNODC/UNECE Task Force on Crime Classification 2011, p. 25).

Human Rights Perspective

It is important to place the present subject within the human rights law discourse. Human rights are universal and belong to every human being in every society. Therefore, human rights are generally regarded as neutral.

The issue of homicide has found its corollary in the discussion on "right to life." For human rights law, it is taken for granted that States are not only required to refrain from directly violating the right to life, but are also obliged to take positive measures to protect the individual from abuses committed by other individuals, that is to prevent such crimes and prosecute and punish the perpetrators. If initially, the right to life was envisaged as a protection from direct violations by the state, it has developed to impose on states obligations of prevention in relation to acts perpetrated by private persons.

Under human rights law, impunity for a homicide may constitute a violation of the right to life. Within the international human rights law system, only the regional legal framework, namely, the European Convention on Human Rights, specifies in a detailed way limitations on the right to life. In particular, it allows limitations of a right to life under certain conditions resulting from the use of force among others in self-defense from unlawful violence, "which is no more than absolutely necessary" (Article 2). Thus, domestic legal order may allow self-defense under conditions specified by the relevant human rights treaty. In reality, however, the debate on self-defense is much more complex.

Human rights consequences of the types of killings discussed here do not concern exclusively the right to life. Indeed, many of them involve different forms of discrimination, particularly on the basis of gender and age. A concrete case in point is "honor" killing, where discrimination can be dissected in the fact that laws applicable to this crime envisage an unequal treatment of men and women. Other potential human rights issues at stake may include nondiscrimination and equality as well as violations related to violence against women.

In practice a number of measures have been undertaken to address these issues. The human rights Treaty Bodies and Special Procedures of the United Nation's Human Rights Council have accumulated a considerable practice generally on the right to life but also as a result of the work of the treaty bodies in the context of protection of women and children against violence. General Assembly Resolution 59/165 of 10 February 2005 recalls the obligation of all States to use legislation to prevent and combat crimes in the name of honor with a view to their elimination. It also reminds the duty of states to investigate, prosecute effectively, and document cases of crimes against women and girls committed in the name of honor and punish perpetrators. Similar recommendations have been put forward in relation to the phenomenon of witchcraft. It is clear that human rights consequences of the "lawful killings" are expansive. International legal framework tends to prohibit all the practices falling under this classification by virtue of guarantees of the right to life and nondiscrimination both at universal and regional levels. States are therefore saddled with a duty to criminalize homicide and homicide-related practices.

Measuring the Extent of "Lawful Killings"

Measuring "lawful killings" requires broadening the perspective beyond criminalization to include the social and cultural aspects related to crime definition and reporting/recording behaviors and procedures. Available homicide statistics may be an adequate tool to measure the incidence of "lawful killings," depending on two levels of challenges to be considered. The first level is related to the general difficulty in comparing crime statistics across countries. The quality of statistics depends indeed on the efficiency of national reporting systems, so that discrepancies in homicide rates across countries may originate from different degrees of efficiency and/or from national-level underreporting or nonreporting issues rather than real differences in violence incidence. Moreover, homicide cases are recorded according to counting rules for statistical classification by using different units (persons, cases, victims) and time of registration (at the time of reporting/discovery of a case, at different point in time during investigation, etc.) so that direct cross-national comparison should be pondered over methodological limitations to avoid misleading interpretations of data.

The second level of challenges depends on operative definitions of crime for data collection and how they affect the classification of homicide data for statistical purposes. The general heading "homicide" may cover substantial differences in the patterns of violence. In general, definitions for unlawful killings vary depending on the involvement of legal concepts related to the perpetrator, such as motivation, involvement, responsibility, and planning. A recent study on European homicide research (Liem and Pridemore 2012) has identified two main elements in national definitions of homicide: intent and premeditation. The role of both intent and premeditation (and/or other aggravating circumstances) determines the criteria for statistical classification of unlawful killings. Differences in homicide statistics across countries, thus, depend also on whether premeditation, intent, and aggravating circumstances are defined by autonomous provisions (Liem and Pridemore 2012, p. 10).

Even when the characteristics of homicide appear clearly defined by criminal law, there are cases that may fit the definition but "there is less consensus on whether or not they should be comprised under the label homicide" (Liem and Pridemore 2012, p. 12). This refers in particular to the categories of attempted homicide, assault leading to death, euthanasia, and infanticide, which in some countries are not considered elements of homicide. Analysis carried out by the European Sourcebook of Crime and Criminal Justice Statistics about the compliance of European countries to the operative definition of homicide ("intentional killing of a person; where possible it should include assault leading to death, euthanasia, infanticide, attempts; but exclude assistance in suicide") based on inclusion/exclusion of such elements in national definitions of homicide demonstrates that only 13 countries out of 36 fully comply with the definition (Aebi et al. 2010, pp. 343–344). Liem and Pridemore (2012, p. 16) have reviewed elements included in national statistics on homicide in 34 countries also in relation to causing death by dangerous driving (which is included by 13), justified killing (10 countries), and nonintentional killings (17 countries). Altogether these findings help in understanding the impact of cross-national differences in defining homicide on the comparison of homicide rates across countries.

Consensus on what is to be considered homicide and what may be excluded, thus "excusable," depends on a wide range of cultural influences, which may go back to ancient customs incorporated in many countries. This includes a shared acceptance of which circumstances can be considered as mitigating the offense, and the concept of provocation.

While there may be a large consensus in considering mitigating circumstances for killings that occur on duty or to defend life, some cultures may consider legitimate or justifiable the killing of an unfaithful wife or of a child suspected of witchcraft. This generates several types of "lawful killings," for example, mob killings, witchcraft/ritual killings, "honor" killings, and dowry killings, for which some patchy statistics are available.

Table 1 provides some examples of available data and related sources on selected types of "lawful killings", without the ambition to provide a complete picture of their incidence and distribution. Data presented in the table are commented in the relevant sections below.

Mob Killings

The UN Special Rapporteur on extrajudicial killings and summary or arbitrary executions has described mob killings as "those undertaken by individuals or groups who take the law into their own hands. They are killings carried out in violation of the law by private individuals with the purported aim of crime control, or the control of perceived deviant or immoral behavior. Specific incidents of vigilante killings can most usefully be categorized along various axes – such as spontaneity, organization, and level of State involvement – and can be considered in relation to various characteristics – including the precise motivation for the killing, the identity of the victim and the identity of perpetrators" (United Nations 2009, para 51).

These forms of killing have been reported from all around the world to the extent that they do not represent an issue limited to any particular region but they should be considered a potential concern for all States (the report refers to cases recorded in Australia, Brazil, Benin, Burundi, Cambodia, the Central African Republic, Democratic Republic of Congo, Ghana, Guatemala, Guinea, Haiti, Hungary, Indonesia, Jamaica, Kenya, Liberia, Mexico, Nepal, Nigeria, Papua New Guinea, the Philippines, South Africa, Tanzania, Uganda). Victims are most frequently young males, who are often suspected of having committed theft, robbery, or murder, or being accused of witchcraft. The modalities of these killings vary from stoning to burning to lynching by the use of canes, machetes, or any type of weapon (Ng'walal and Kitinya 2006). For example, the Uganda Police provides statistics for *death by mob action*. Data refer to the number of investigations. Since each episode could involve more than a victim, the actual number of persons killed is probably largely underestimated (see Table 2). The increase observed in 2011 was commented by the police as "attributed to thefts, robbery, suspected witchcraft, and dissatisfaction with delayed/omission of justice" (Uganda Police Force 2011, p. 9).

Literature suggests that mob killings are likely to occur where the judicial system is weak and affected by corruption (United Nations 2009; HRW 2010). The summary executions of suspects by angry crowds would then be a response to the limited presence of the State which is perceived as lacking or delaying proceeding against criminals, causing public distrust towards justice and law enforcement institutions. Others point to inefficacy of formal and informal dispute resolution systems and the conflict between the cultural and the legal frameworks. Witchcraft is the typical example of a behavior frequently not recognized by courts, but strongly felt by the traditional culture of society. In these cases, the lynching of alleged witches is perceived by communities as an acceptable alternative form of administration of justice.

Killing of Witches

Another form of "lawful killings" related to belief systems is related to killing of persons accused of witchcraft. This practice is found mainly among tribal communities in Africa, Asia, and Pacific Islands and appears to target mainly women, elders, and children (United Nations 2012). Victims of witchcraft accusations are likely to belong to the most vulnerable

Lawful Killings, Table 1 Examples of data reported in literature on “lawful” killings (mob killings, witchcraft/ritual killings, “honor” killings, dowry killings)

Region	Country	Mob killings	Witchcraft/ ritual killings	“Honor” killings	Dowry killings	Source
Africa	Burundi	75 cases (2009)				Human Rights Watch 2010
	South Africa		132 plus unreported cases (2000)			Petrus 2009
			455 cases (1990–1995)			Carstens 2009
	Tanzania		50 cases (2007–2009)			Dave-Odigie 2010
		1,249 victims (2000–2004)				Ng’walal and Kitinya 2006
	Uganda	1,241 cases (2007–2010)	68 killings (2008–2010)			Uganda Police 2010
	Zimbabwe		42 cases			United Nations 2012
Asia	Bangladesh	161 cases (2011)		2,303 killings and 167 suicides (2000–2011)		Odhikar 2012
					437 deaths (2007–2009)	United Nations 2012
	India		2,028 killings (2000–2011)		89,964 deaths (2000–2011)	NCRB 2000–2011
			240 cases in 2011		8,186 deaths in 2011	
		10 killings in Vaishali (2007)				Digital Journal 2007
	Pakistan			791 murders; 719 suicides; 414 attempted suicides (2011)		HRCP 2010
Caribbean	Jamaica	14 cases (2011)				Jamaica Constabulary Force 2012
Europe	Albania			1 killing (2005)		Chesler 2009
	Denmark			1 killing (2006)		Chesler 2009
	France			1 killing (2002)		Chesler 2009
	Germany			78 killings from (1996–2005)		Oberwittler and Kasselt 2011
				3 killings (2006–2008)		Chesler 2009
	Italy			127 killings[a] (2010)		Karadole and Pramstrahler 2011
	Netherlands (The)			2 killings (2003–2004)		Chesler 2009
	United Kingdom			12 reported cases and 109 cases reinvestigated (2000–2006)		Smartt 2006
				117 cases re-investigated		Khan 2007
				14 killings (1989–2007)		Chesler 2009

(*continued*)

Lawful Killings, Table 1 (continued)

Region	Country	Mob killings	Witchcraft/ ritual killings	"Honor" killings	Dowry killings	Source
North America	Canada			4 killings (1999; 2003; 2007)		Chesler 2009
	United States	4 killings (2005–2006)				United Nations 2009
				10 killings (1989–2007)		Chesler 2009
Oceania	Papua New Guinea	500 cases				United Nations 2012

[a] 127 cases of "honor killing" can be identified out of 159 cases of death, which also include suicides of offenders who committed such killings. Cases refer to women killed by men, often with the involvement of acquaintances, relatives, children, etc. (Karadole and Pramstrahler 2011, p. 35)

Lawful Killings, Table 2 Number of investigations of death by mob actions in Uganda. Period 2007–2010

Year	Investigations
2011	383
2010	357
2009	332
2008	368
2007	184

Source: Uganda Police Force (2008–2011)

positions in society: elders and people with some form of disability or unique characteristics, such as albinos, for which they are discriminated (Dave-Odigie 2010). These killings often take place in post-conflict and post-disaster settings, or regions burdened by public-health crises. These phenomena are still frequent and rooted in lack of education and opportunities for protection of the weakest elements of the society. Such killings are captured by official statistics in India; according to the NCRB, 240 cases of witchcraft killings were reported in India in 2011. Reports of witchcraft killings have almost doubled over the last decade, with sharp increases between 2004 and 2005 and again between 2010 and 2011 (see Fig. 1). Figure 1 also shows an increase in reports of dowry deaths, the characteristics of which are described below.

Available statistics on ritual killings published by the Uganda Police Force account for 29, 14, and 8 ritual killings for 2009, 2010, and 2011, respectively (Uganda Police Force 2010, 2011). According to the Uganda Police, the remarkable decline in ritual killings was achieved against some serious challenges, including that many people still believe in witchcraft and practice rituals and some quack healers demand human body parts for their rituals (Uganda Police Force 2011, p. 10). Given the above-mentioned impact of underreporting and deficiencies in the recording systems, it may be estimated that the incidence of such killings is actually higher.

The most recent Report of the Special Rapporteur on violence against women notes the occurrence of 42 cases of killings of women due to accusations of witchcraft in Zimbabwe and other 500 in Papua New Guinea (United Nations 2012, p. 11).

"Honor" Killings

"Honor killings can be understood as an extreme result of the combination of patriarchal dominance over women and their sexuality, rigid behavioral norms, and the importance of honor for social relations in economically and socially backward, agrarian societies" (Oberwittler and Kasselt 2011). "Honor" killings exist as a mitigating circumstance in many countries/territories. For example, in 6 Mexican States out of 32 (Michoacán, Baja California Sur, Chiapas, Jalisco, Yucatán, and Zacatecas) "femicide for honor reasons" is sanctioned with sentences between 3 days and 5 years imprisonment (Chouza 2012). The perpetrators of such killings are convinced they act to safeguard the "honor" of men and families to remedy inappropriate behaviors or shameful incidents, most frequently committed by women and girls. The reasons behind killings include adultery, the choice of

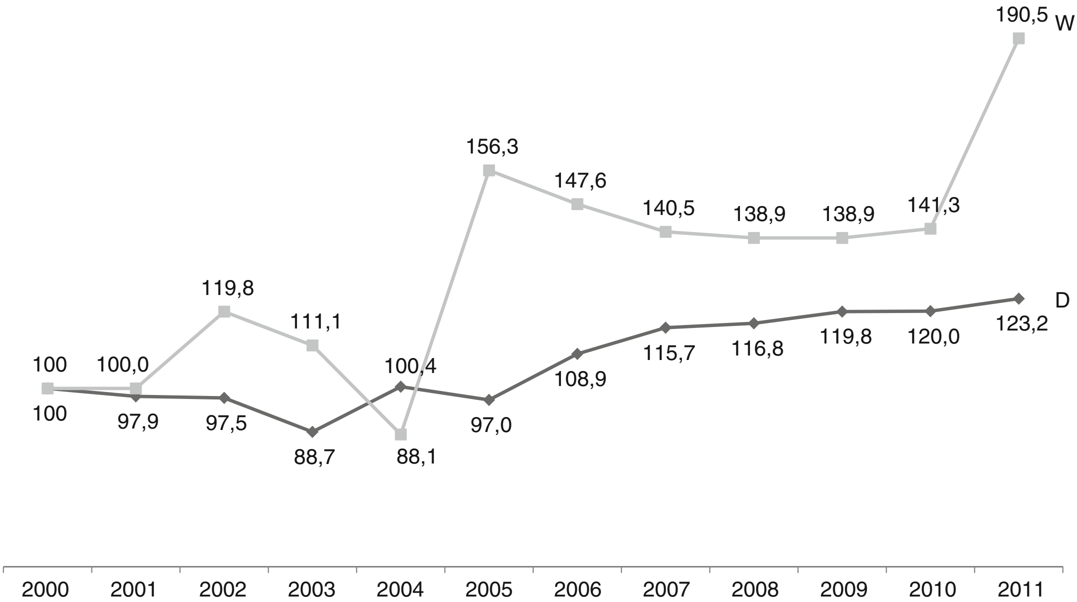

Lawful Killings, Fig. 1 Trends of selected forms of "lawful" killings in India: dowry and witchcraft-related killings, 2000–2011 (index: 2000 = 100) (Source: NCRB 2000–2011)

a partner, seeking divorce, and also homosexuality, bisexuality, and transsexuality. In some cultures, when a woman is the victim of rape, this is considered to bring shame to the family and victims may be killed in the belief that this will restore the lost "honor."

The clash between different cultural values within the same country or the adaption of some members of one specific community to the "new" cultural context could be one reason for "honor" killings. These issues have been analyzed by Chesler (2010, pp. 4–5) in her study on 230 victims of honor killings in North America, Europe, and 14 countries in the Muslim world (Afghanistan, Bangladesh, Egypt, Gaza Strip, India, Iran, Iraq, Israel, Jordan, Pakistan, Russia, Saudi Arabia, Syria, Turkey) between 1989 and 2009. The results of this study demonstrated that "worldwide 58 % of the victims were murdered for behaving 'too Western' and/or for resisting or disobeying cultural and religious expectations." Indeed, since the process of adaptation to different cultures can be perceived as dishonorable by the victim's family, immigrant communities seem to be vulnerable to the problem of "honor" killings to the extent that the risk of being killed in the name of "honor" is higher among immigrant communities in the "Western world" than it is in the countries those immigrants come from (Luopa 2010, p. 7).

However, there is no specific cultural or local connotation for "honor" killings. For example, a study on 78 cases in Germany showed that "in 80 % of the cases an unwanted love affair by a woman, outside or after marriage, was the main reason for the homicide, whereas a desire to live an autonomous 'Western' lifestyle was the only central factor in very few cases." This study also demonstrated that "honor killings frequently occur in the context of 'arranged marriages', either when young women violate the norm that their partner will be chosen by the family or when married women want to escape from an unbearable relationship which is the consequence of an arranged marriage." Moreover, two-thirds of these homicides were committed in families of Turkish or Kurds origins, by first-generation immigrants, without Germany citizenship (Oberwittler and Kasselt 2011, pp. 2–3).

A recent study shows that in Italy, 127 women were killed by an intimate partner or former partner during 2010. Of these 70 % of victims and

76 % of offenders were Italians and the motives of killings included conflict in the relationship, men's unemployment, and "honor" (United Nations 2012, p. 8, Karadole and Pramstrahler 2011). The occurrence of "honor" killings in countries where "honor" crimes are not supposed to exist implies that these killings are likely to be recorded as domestic violence, thus limiting the knowledge on the real nature and extent of the phenomenon (see Table 1). The Council of Europe has expressed concern about the increase in the frequency of "honor" crimes and the insufficiency of adequate recording of their occurrence (Council of Europe 2003), which lead to progressive raising of awareness of the scale and the dynamics of the problem and reviewing cases (for example, a national review of nearly 120 cases of murder in the UK lead to the identification of at least 13 cases of suspected "honor" killings).

Reliable data are lacking. In 2003 the UN Population Fund estimated that approximately 5,000 women and girls are killed in "honor killings" every year in the world. According to the last report by the UN Special Rapporteur on violence against women, "honor killings take many forms, including direct murder; stoning; women and young girls being forced to commit suicide after public denunciations of their behavior; and women being disfigured by acid burns, leading to death" (United Nations 2012, p. 12;).

Dowry Deaths

Dowry-related killing is a practice related to religious and cultural traditions of South Asian countries; antidowry laws have been enacted in India (1961), Pakistan (1976), Bangladesh (1980), and Nepal (2009) (United Nations 2012). Indian criminal law describes dowry deaths as follows: "where the death of a woman is caused by any burns or bodily injury or occurs otherwise than under normal circumstances within 7 years of her marriage and it is shown that soon before her death she was subjected to cruelty or harassment by her husband or any relative of her husband for, or in connection with, any demand for dowry, such death shall be called 'dowry death', and such husband or relative shall be deemed to have caused her death" (304b IPC). The payment of dowry for marriages was abolished in 1961 with the Dowry Prohibition Act; however, the incidence of killings related to dowry is still problematic (see Table 2 and Fig. 1). According to the statistics provided by the NCRB, dowry-related killings have increased by 23 % over the last decade (2000–2011) and dowry represents the fifth motive for murder in India (see Fig. 1 and Table 2). According to official statistics, dowry is the fifth cause of murder. The distribution of these types of killings within India appears patchy as no cases have been reported in 10 regions, while dowry is the main motive for killings in Orissa and West Bengal, accounting for, respectively, 48 and 45 % of identified cases for which a cause was identified in 2011. Despite attempts towards criminalization of "honor" and dowry-related crimes, and removing the right to pledge for mitigating circumstances, these killings are still very frequently considered as "lawful." States responses to these forms of violence often fail in eradicating traditional customs, which frequently also involve police officers and judges who might use a discriminatory and gender-biased approach in applying the law (Luopa 2010).

Conclusions

In conclusion, "lawful" killings represent a hidden form of lethal violence, supported by some consensus and even by law in some cultures, which mostly affects women and other vulnerable groups in the society. A large portion of this violence goes undocumented, a part is recorded but fails to prominently feature in official statistics, which struggle in describing the extent of such phenomena. Beyond violence caused by the hand of husbands to wives, parents to daughters, relatives and community members to persons who are perceived as "different" or having infringed some (frequently unwritten) community laws, there are also many cases of suicides of women who are unable to find a way out from the psychological and physical abuses suffered.

Capturing the extent of these forms of violence faces both technical and cultural obstacles, as regards general limitations of comparative crime statistics and different understanding on what is to be considered "unlawful" death. In order to enhance the understanding of lethal violence patterns, more efforts should be paid in the recording and classification of the circumstances related to killings, such as, for example, a detailed description of the act/event, the alleged motives for the act, the characteristics of the perpetrator and the victim. The accuracy in the collection of this information is crucial in terms of analysis and policy making at both national and international levels.

Indeed, as stated by the UNODC/UNECE Task Force on Crime Classification (2011, p. 12), "a crime classification at international level, under the principle of exhaustiveness, should cover all possible acts or events that could carry criminal responsibility and sanctions anywhere in the world." Therefore, it would be challenging to understand how to provide an unique interpretation of "honor" killings, dowry deaths, or killing of witches, which in some parts of the world do not imply criminal responsibility and do not carry sanctions.

Related Entries

- ▶ Cultural Criminology
- ▶ Domestic Violence
- ▶ Human Rights Violations in Criminal Court
- ▶ Moral Crimes

Recommended Reading and References

Aebi M et al (2010) European sourcebook of crime and criminal justice statistics – 2010 (4th edn). Boom Juridische uitgevers, Den Haag. Onderzoek en beleid series, no. 241, Ministry of Justice, Research and Documentation Centre (WODC). Accessed on 27 July 2012

Carstens Pieter A (2009) The cultural defence in criminal law: South African perspectives. In: Foublets MC, Renteln AD (eds) Multicultural jurisprudence: comparative perspectives in the cultural defence. Hart, Oxford

Chesler P (2009) Are honor killings simply domestic violence? Middle East Quart 16(2):61–69

Chesler P (2010) Worldwide trends in honour killings. Middle East Quart Spring 2010;17(2):3–11

Chouza P (2012) Feminicidio "por honor". El Pais. http://sociedad.elpais.com/sociedad/2012/03/05/actualidad/1330981386_402961.html

Council of Europe (2003) So-called "honour crimes". Report, Committee on Equal Opportunities for Women and Men. http://assembly.coe.int/Documents/WorkingDocs/doc03/edoc9720.htm

Council of Europe Parliamentary Assembly (2009) Doc. 1193 The urgent need to combat so-called "honour crimes", 8 June 2009

Dave-Odigie CP (2010) Albino killings in Tanzania: implications for security. Peace Stud J 3(1):69–75

Digital Journal (2007) Another mob killing reported from India's Bihar state. http://www.digitaljournal.com/article/231650#ixzz1qENu9NNS

Grosjean P (2010) A history of violence: testing the "culture of honor" in the US South. Mimeo, University of San Francisco. http://www.feem.it/userfiles/attach/201011815184441_Grosjean_paper.pdf

Human Rights Commission of Pakistan (HRCP) (2010) Annual Report. State of Human rights in 2010. http://www.hrcp-web.org/Publications/AR2010.pdf

Human Rights Watch (HRW) (2010) Mob justice in Burundi. Official Complicity and Impunity, United States

Jamaica Constabulary Force (JCF) (2012) Annual major crime statistics review

Karadole C, Pramstrahler A.ed (2011) Femicidio. Dati e riflessioni intorno ai delitti per violenza di genere. Regione Emilia Romagna – Assessorato Promozione Politiche Sociali. http://www.casadonne.it/cms/images/pdf/pubblicazioni/pubblicazioni/femicidio_pdf.pdf

Khan R (2007) Honour-related violence (HRV). In: Scotland: a cross- and multi-agency intervention involvement survey, Internet Journal of Criminology

Liem MCA, Pridemore WA (2012) Handbook of European homicide research. Patterns, explanations and case studies. Springer, New York

Luopa K (2010) State responses to honour killings. Åbo Akademi University. http://www.abo.fi/instut/imr/norfa/Katja%20Luopa%20killings.pdf

NCRB (National Crime Records Bureau) (2000–2011) Crime in India. Ministry of Home Affairs, Government of India, New Delhi

Ng'walal PM, Kitinya JN (2006) Mob justice in Tanzania: a medico-social problem. Afr Health Sci 6(1):36–38

Oberwittler D, Kasselt J (2011) Ehrenmorde in Deutschland. Eine Untersuchung auf der Basis von Prozessakten [Honour Killings in Germany. A study based on prosecution files] (Polizei+Forschung, Bd. 42, hrsg. vom Bundeskriminalamt). Wolters Kluwer Deutschland, Köln

Odhikar (2012) Human rights report. Odhikar Report on Bangladesh, January

Petrus TS (2009) An anthropological study of witchcraft-related crime in the Eastern Cape and its implications for law enforcement policy and practice. PhD dissertation, Nelson Mandela Metropolitan University, Port Elizabeth, South Africa

Smartt U (2006) Honour killings. Justice Peace 170(7):4–7

Uganda Police Force (2008–2011) Annual crime and traffic/road safety report. http://www.upf.go.ug/Reports/Annual%20Report%202008.pdf

United Nations (2009) Report of the special rapporteur on extrajudicial, summary or arbitrary executions, A/64/187. http://www.extrajudicialexecutions.org/application/media/64%20GA%20SR%20Report%20(A_64_187).pdf

United Nations (2012) Report of the Special Rapporteur on violence against women, its causes and consequences, Rashida Manjoo. A/HRC/20/16 http://www.ohchr.org/Documents/Issues/Women/A.HRC.20.16_En.pdf

United Nations Office on Drugs and Crime (2011) Global study on homicide, Vienna. http://www.unodc.org/documents/data-and-analysis/statistics/Homicide/Globa_study_on_homicide_2011_web.pdf

UNODC/UNECE Task Force on Crime Classification (2011) Principles and framework for an international classification of crime for statistical purposes. Report to the Conference of European Statisticians

Waters T (2007) When killing is a crime. Lyenn Rienner, Boulder

Welchman L, Hossain S (2006) Honour – crimes, paradigms and violence against women. Zed Books, London

Leadership

▶ Prison Warden Stress and Job Performance

Learning

▶ Innovation and Crime Prevention

Learning Theory

▶ Social Learning Theory

Left Realism

Walter S. DeKeseredy[1] and Martin D. Schwartz[2]
[1]Institute of Technology, University of Ontario, ON, Canada
[2]George Washington University, Washington, DC, USA

Overview

Left realism began in the 1980s in Great Britain partially as a reaction to those on the left who felt that talk about street crime was just a racist-fueled media scare. It was an attempt to take back the crime issue from conservatives with progressive socialist analyses and short-term solutions. Left realists felt that lower or working class members were not only the primary victimization targets of street criminals, but were being attacked from above by white collar, corporate and state crime. Rather than automatically rejecting all mainstream analysis, left realists have adopted elements of strain theory and subcultural theories to make their point that absolute poverty is not a cause of crime. Rather, relative deprivation that leads to dissatisfaction with the social order is the criminogenic factor. When such dissatisfied people lack the legitimate means to attain socially approved goals, they may join together in patriarchal, pro-violent subcultures that support their members in criminal activities and in actions to buttress their masculinities (e.g., respect, status) that have been battered by the economic policies of modern governments. The most recent left realist theories, then, which have spread from Great Britain to the United States, Canada and Australia, have added feminist and male peer support elements to concerns for progressive short term solutions.

Introduction

Left Realism can be located on a political map in comparison with so-called "right realism" and "left idealism." Right realism is a neo-conservative doctrine that attacks mainstream criminological theory, blames the working class for crime, and calls for extensive harsh measures to conduct a war on crime. It has been extremely influential in providing for massive increases in imprisonment at the same time that crime has gone down, and in convincing many police departments to engage in zero tolerance policing that targets minor annoying behaviors on the theory that this will reduce serious crime in its wake. Left idealism was a reaction to such policies on the political left that protested the implicit racism in discussions of street crime and crime statistics. Complaints about working class crime, they argued, as a media and conservative politician-induced scare tactic designed to create a moral panic that would justify oppressive actions against the poor. Rather, these theorists turned their attention on the horrific and harmful practices of corporations, the wealthy and the state.

Left realism, while actually agreeing with these varied claims of the idealists, felt that the primary focus on crimes of the powerful meant rather explicitly that crimes of the powerless or less powerful were being ignored. While idealists often argued that a concern with burglary, robbery and rape was a thinly disguised form of racism, the left realists instead argued that the predatory crimes of the underclass and the predatory crimes of the upper class squeezed the working class in the middle. As primary targets both from above and below, the working class suffers enormous victimization.

Background Description

Left realism had its beginnings in the United Kingdom in the 1980s, among theorists with extensive experience in decrying the power and actions of the law and the state, but partially as a reaction to various theorists on the political left. Jock Young, John Lea and Roger Matthews were upset that the debate over interpersonal and street crime only had two sides. The larger voice consisted of conservatives had no problem playing off racial fears to convince voters to fund a massive war on crime and an extraordinary increase in the use of imprisonment. The only other major voice came from the left, where many were arguing that any concern for street crime was a racist reaction against inner city and minority group members. After all, these people claimed (quite correctly) most of the money stolen in the world was obtained by white collar and corporate criminals, and much of the injury from illegal activities also tended to be caused by these same people through illegal code violations, lack of safety practices in manufacturing and agriculture, and substantial health threatening industrial pollution.

However, limiting the debate to only these two sides, left realists felt, meant that there remained a major problem. City people were hiding behind locked doors, giving up sitting on park benches or their front stoops to meet with their neighbors, not because they were scared that General Electric or Halliburton would rig bids to steal millions in public construction projects. Older people were dying in heat waves after nailing their windows shut despite not having air conditioning, because of a major fear, but certain not because they were afraid that a company like Union Carbide might release poison gas into the atmosphere, as they did in India in 1984. Rather, the crime problems that people in the working class are most worried about are a variety of acts committed by people at the lowest end of the socioeconomic ladder. They were afraid of rape, robbery, break-ins and burglary, assault and murder. In opposition to the left idealist notion that crime fear was a social construction based on racist and anti-minority thought, left realism took the position that working class people were being attacked both from above and below. While at the mercy of white collar and corporate criminals for a wide variety of crimes, they were also the primary victims of street criminals such as rapists, car thieves, purse snatchers and armed robbers. Thus, one of the most important early arguments of the left realists was that complaints that the justice system was racially biased (which it was), and that the legal system was put into effect to serve the interests of the upper classes (which it was), still masked the fact that the primary victims of both violent and economic crimes were lower class and working class people.

Worse, people in the general public had little respect for theorists of the left who belittled the problem of rape or robbery, the object of great fears on their part. Over 20 years ago, American left realist Elliott Currie laid out a crucial concern that completely ignoring intra-class street crime serves to "help perpetuate an image of progressives as being both fuzzy-minded and, much worse, unconcerned about the realities of life for those ordinary Americans who are understandably frightened and enraged" (1992: 91). Ignoring or belittling a concern for street crime and victimization meant handing over the entire issue of street crime to conservative politicians, who had no problem using it to their benefit first to elect their candidates, and then to implement a series of progressively harsher criminal justice policies, such as longer sentences, mandatory sentences, increased imprisonment, reduced defendant rights, and reduced services for rehabilitation or treatment.

These were not the only places where left realism diverged from common positions taken by other theorists of the left. In the first place, left realists made it clear that they did not feel any necessity to be constrained by the dictates of Marxist theory (Hayward 2010). This was not unique on the left, it was still controversial in some quarters when it became an essential element in left realist theory. While explicitly remaining on the left politically, they borrowed any ideas from mainstream criminology that they believed had merit or explanatory power that would advance or help their explanations of criminal behavior. It was much more common then (and to some degree today also) for left-leaning criminologists to reject wholesale any theories that originated in mainstream criminology, for no other reason than the place of their origin. Thus, it was a major divergence from the norm for left realism to adopt some mainstream ideas just because they seemed valuable or had explanatory power. For example, Lea and Young (1984) borrowed extensively from concepts brought into criminology by Robert K. Merton (strain theory) and Albert K. Cohen (subcultural theory). Lea and Young argued that although criminologists have commonly claimed that crime was caused by poverty, one cannot attribute or explain crime solely through absolute deprivation or extreme poverty. There are many cultures where people were much poorer than the poor of the United State or Great Britain on some objective scale (income, standard of living, purchasing power), but these poor people in other cultures as a group committed much less of what is usually termed street crime (e.g., burglary, robbery, rape, assault). Thus, just the state of having a small amount of money – poverty – was not a sufficient explanation for criminal activity. Rather, the left realists said, crime has its roots in "relative deprivation," or the extreme differences between rich and poor that leads to discontent with the political structure. This was particularly true in a context where such people do not see any hope of a political solution. Such people who are frustrated and disempowered by relative deprivation tend to band together – to come into contact with other disenfranchised people, and eventually form subcultures that in many cases encourage or legitimate criminal behavior.

State of the Art

There have been a number of more current left realist perspectives. For example, DeKeseredy and Schwartz (2010) have more recently argued that criminological subcultural development has been strongly influenced in recent years by the destructive consequences of conservative Chicago School economic policies, and by marginalized men's attempts to live up to the principles of hegemonic masculinity in an environment that offers few constructive outlets for men concerned with masculinity to act out this gendered behavior. Without jobs, social structures, organized sports or educational support, marginalized men sometimes find that crime provides one of the few bases of support available to them for gendered behavior. As newer conservative economic models predominate, jobs continue to be eliminated or exported and salaries for those that remain often cut drastically. Deindustrialization and the drastic decline in family owned farms provide more challenges to young men's masculine identity than ever before, and the addition of the effects of institutionalized racism no doubt contributes to the mushrooming presence of criminal gangs in the United States.

Although many or even most men who are economically marginalized engage in male to male violence as a form of compensatory masculinity, DeKeseredy and Schwartz (2010) argue that others, including those influenced by a patriarchal culture, engage in a variety of forms of male-to-female victimization as an effective means of repairing damaged masculinity (Messerschmidt 1993). Women abuse has become in many violent subcultures a legitimate way of maintaining patriarchal authority and control.

Other left realist perspectives have attempted to explain other problematics. For example, Gibbs (2010) argues that left realist theory is useful in attempts to explain terrorist acts. Economically disenfranchised men not only are a pool from which street criminals are drawn, but they also under certain circumstances join terrorist-supporting subcultures. Thus, she argues, current political "get tough" policies are likely to be just as unsuccessful in fighting terrorist acts as they have been in fighting street crime.

If there is one thing that separates left realist theory from other theoretical perspectives, and particularly mainstream criminology, but also a substantial portion of critical criminology, it is the importance to any theoretical discussion of short-term, anti-crime policies and practices. British left realism has been centrally concerned with criminal justice reform. In the early formulations, it received extensive attention for proposals dealing with the democratic control of policing. The goal was to implement a pattern of minimal policing in areas where local communities were opposed to a heavy police presence, and to oppose such policies as zero tolerance policing. The goal was to locate and implement police reforms that people in a district actually want.

Further, left realists understood that limiting policing suggestions to issues related to the criminal justice system would doom them to failure. To some degree, the criminal justice system is asked to clean up society's messes, caused by problems in low employment, bad housing, poor schools, and cuts in other city services. Left realists have argued that many of these issues are related to crime, and included in their short-term proposals for solutions things that might deal directly with these problems, such as an increase in the minimum wage, and affordable day care. To curb crimes of the powerful, left realists have called for such policies as the democratization of corporations, representative citizen patrols to study complaints of corporate crime, and enforcement of regulatory procedures and mechanisms. Still, this area has not been a major strength for left realists.

Controversies

One of the most potent critiques of left realism, especially in Great Britain, has been the attack that it does not offer a coherent theory of the state (Coleman et al. 2009). For the most part, this has been an attack on the 1980s concept of the Square of Crime, which focused on four interacting elements that can be visualized at the corners of the square: the victim, the offender, state agencies such as the police, and the public. Jock Young explained the relationship (1992: 27):

> It is the relationship between the police and the public which determines the efficacy of policing, the relationship between the victim and the offender which determines the impact of the crime, the relationship between the state and the offender which is a major factor in recidivism.

This conceptualization, it was often argued, was most relevant to inner-city street crime, and less relevant to a broader understanding of various societies and crimes being studied today. Further, the development of the state, while inherent in the square of crime, was not fleshed out in significant detail.

One recent attempt to deal with this problem came from Roger Matthews (2009) who published a refashioned left realist theory that prioritizes the state and views it as one of the "fundamental organizing concepts that provide the conceptual frameworks through which we make sense of the social world" (2009: 346). Matthews developed an extensive call for linking theory, method and intervention in his re-energized left realist theory. However, this is a single author, and certainly much more extensive work on this subject needs to be done.

As mentioned, one of the central features of left realism is that the main architects did not feel constrained by traditional leftist theory, and in fact took their influences from any source they felt advanced explanatory power or knowledge. Unsurprisingly, then, there has been considerable controversy on the left among those people who are in fact bound to traditional Marxist or other leftist structures, and find that any incorporation, integration or use of theories developed by mainstream theorists to be a sign of right-wing tendencies. Thus, to give just one example, Jock Young's use of such mainstream concepts as strain and subculture to attempt to harness a left realist explanation earned him what Yar and Penna (2004) no doubt viewed as their ultimate insult: he was labeled a "positivist" for using a politically incorrect theory. Others on the left have similarly attacked left realism for being politically incorrect, rejecting the claim of the left realists that they are engaged in socialist analysis. Similar attacks were made on left realists for their use of social science data collection such as victimization surveys to develop information to inform their theories.

Another strong controversy in England was the series of attacks from the left from people who believed that short-term solutions that fix problems with the police only plays into the hands of the state. After all, critical changes that make people happy with the police are proposed and implemented, then they will turn their attention away from other issues that might fuel their discontent. This improvement of such things as the police ultimately will only strengthen the existing power structure (e.g., Jamieson and Yates 2009). An alternative formulation of this notion is that the state, the legal system and the criminal justice system are all part of an overarching system that makes any efforts at reform end up as mere tinkering with the inside structures and ultimately meaningless. Left realists, on the other hand, have always believed that ignoring street crime and policing problems while waiting for the outcome of the eventual revolution means ignoring and condemning the women and men who are the victims of both street criminals and also the overzealous activities of some police forces. Any delay in making reforms just means that more lives will be destroyed by criminal or police action. Thus, one of the major contributions of left realists was to argue that it was possible to be a complete realist while at the same time pushing for left-oriented progressive solutions to crime. They argued that it is a legitimate goal to "chip away" at the capitalist patriarchal order, rather than holding off to await the success of some policy to overthrow all of society's structures. Of course, there are many who see this lack of support for the revolution as a serious flaw in left realism.

There have been sharp attacks on left realism for ignoring crimes of the capitalist order. Of course, those criminologists that were called left idealists here were heavily focused on corporate and state crime, and they obviously felt that such a focus was important. However, as also made plain in this essay, from the beginning one of the central tenets of left realist thought was that members of the working class were victimized from above (white collar, state and corporate crime) as well as below (street crime). Further, there were a series of important works those arose from left realist theory on corporate crime in the 1980s and 1990s. Yet, there is no question that the primary focus of left realists was to correct the imbalance of looking only at state and corporate crime, and ignoring the victimization of people by street criminals.

The other area of major attack on left realism has been the argument that some of the theorists have been gender blind. There is some truth to the argument made by some feminist critics that the earlier left realists, and even some of the more recent formulations, are not particularly sensitive to the unique problems of women, although certainly much left realist analysis affected men and women equally. It is not unfair to call them gender blind, as they were indeed more directed toward issues that were of major import to men, such as bar fights and police harassment of young men. Still, left realists were some of the first to recognize the importance of gendered issues. One of their most famous tools, the local victimization survey such as the Islington Crime Survey, was developed specifically to look at measures often seen as gendered, but ignored by most victimologists, such as fear of crime, avoidance behaviors on the streets, sexual harassment, and injury from rape or domestic violence. Still, early left realism was not strong on issues related to how patriarchal structures affected not only women's victimization but women's criminal patterns. More recently, left realists have begun to look at some of these issues, such as violence against women in a global perspective (Currie 2008), and a gendered subcultural theory that deals with violence against women (DeKeseredy and Schwartz 2010).

Although there are numerous places where left realist theories were used to study violence against women, one fertile area has been the study of male peer support as a form of subcultural behavior. In particular, men who use violence against women as a method of repairing damaged masculinities, or in an attempt to maintain and uphold patriarchal structures, often have subcultural male peer support for their actions. A variety of studies have found a relationship between the extent of male violence against women, and the amount of time that such men spend with their male friends, and indeed the amount of commitment these men have toward maintaining these male peer support structures. In particular, drinking alcohol with such friends seems to be related to such male violence. In a national representative sample of men, Schwartz, et al. (2001) found that men who went out drinking two or more times a week, had male friends who offered them support for the emotional abuse of women, and offered support for the physical abuse of women, were ten times as likely to sexually abuse women as men who did not have such friends or drinking patterns.

Another recent left realist approach was developed by Dragiewicz (2010), who attempted a left realist explanation for the existence of the anti-feminist fathers' rights group activism. Such groups have been active in arguing that women are as violent as men, and are politically active in fighting any legislation proposed to reduce violence against women, and in fighting attempts to make men responsible for child support payments. Dragiewicz argues that which such group members are not socioeconomically marginalized by the usual standards, but that they experience divorce and child support as socially and economically marginalizing. She argues that "These marginalized men seek out like-minded peers in person and online, drawing upon and adapting mainstream discourses around families, violence, and gender to reassert patriarchal masculinity in the face of challenges" (Dragiewicz 2010: 205).

Conclusion

Criminal justice policy and government policy in general have been trending further and further to the right over the past 30 years. Even in the United States when Democrats have been elected to the presidency during that time, they have engaged in policies that incorporated a great deal of the current conservative thinking, in what Elliott Currie has termed "progressive retreatism." In general, the entire fight against crime has been ceded to the right, and many observers on the left have limited themselves over the past 30 years to sitting on the sidelines and carping.

DeKeseredy and Schwartz (2012) have issued a call for left realists to do even more to publicize solutions, form alliances with progressive community agencies, and in general to translate left realist ideas into action steps. Currie (2008: 117) suggested that: "the choice is stark and simple: We can either let the process continue and fortify outselves against it, with more gated communities and more prisons, or we can decide that it is not tolerable and work to change it. What we cannot do is pretend we don't know it's happening."

Related Entries

- ▶ Anomie and Crime
- ▶ Cultural Criminology
- ▶ Feminist Criminological Theory
- ▶ Marxist Criminology
- ▶ Police Legitimacy and Police Encounters

Recommended Reading and References

Coleman R, Sim J, Tombs S, Whyte D (2009) Introduction: state, power, crime. In: Coleman R, Sim J, Tombs S, Whyte D (eds) State, power, crime. Sage, London

Currie E (1992) Retreatism, minimalism, realism: three styles of reasoning on crime and drugs in the United States. In: Lowman J, MacLean BD (eds) Realist criminology: crime control and policing in the 1990s. University of Toronto Press, Toronto

Currie E (2008) The roots of danger: violent crime in global perspective. Prentice Hall, Upper Saddle River

Currie E (2010) Plain left realism: an appreciation and some thoughts for the future. Crime, Law Social Change 54:111–124

DeKeseredy WS, Schwartz MD (2010) Friedman economic policies, social exclusion, and crime: toward a gendered left realist subcultural theory. Crime Law Social Change 54:159–170

DeKeseredy WS, Schwartz MD (2012) Left Realism. In: DeKeseredy WS, Dragiewicz M (eds) Handbook of critical criminology. Routledge, London

Dragiewicz M (2010) A left realist approach to antifeminist fathers' rights groups. Crime Law Social Change 54:197–212

Gibbs JC (2010) Looking at terrorism through left realist lenses. Crime Law Social Change 54:171–185

Hayward KJ (2010) Jock Young (1942–). In: Hayward KJ, Maruna S, Mooney J (eds) Fifty key thinkers in criminology. Routledge, London

Jamieson J, Yates J (2009) Young people, youth justice and the state. In: Coleman R, Sim J, Tombs S, Whyte D (eds) State, power, crime. Sage, London

Kinsey R, Lea J, Young J (1986) Losing the fight against crime. Blackwell, Oxford

Lea J (2010) Left realism, community and state building. Crime Law Social Change 54:141–158

Lea J, Young J (1984) What is to be done about law and order? Penguin, New York

Matthews R (2009) Beyond "so what?" criminology. Theor Criminol 13:341–362

Messerschmidt JW (1993) Masculinities and crime: critique and reconceptualization. Roman & Littlefield, Lanham

Pearce F, Tombs S (1992) Realism and corporate crime. In: Matthews R, Young J (eds) Issues in realist criminology. Sage, London

Schwartz MS, DeKeseredy WS (1997) Sexual assault on the college campus: the role of male peer support. Sage, Thousand Oaks

Schwartz MD, DeKeseredy WS, Tait D, Alvi S (2001) Male peer support and routine activities theory: understanding sexual assault on the college campus. Justice Quart 18:701–727

Yar M, Penna S (2004) Between positivism and post-modernity?: Critical reflections on Jock Young's the exclusive society. Brit J Criminol 44:533–549

Young J (1992) Ten points of realism. In: Young J, Matthews R (eds) Rethinking criminology: the realist debate. Sage, London

Legal Control of the Police

Rachel Harmon
University of Virginia School of Law,
Charlottesville, VA, USA

Overview

Police officers are granted immense authority by the state to impose harm. They walk into houses and take property. They stop and detain individuals on the street. They arrest. Sometimes they kill. The problem of policing the police is how to regulate police officers and departments to protect individual liberty and minimize the social costs the police impose while allowing them to do what is necessary to achieve the ends of policing: reducing fear, promoting civil order, and pursuing criminal justice. Constitutional law as interpreted by the United States Supreme Court provides the most well-known check on police conduct. In addition, many other federal, state, and local statutes, constitutional provisions, court decisions, and administrative regulations also govern the police. Since federal constitutional law cannot alone ensure that the benefits of policing are worth the harms it imposes, this participation by other government actors is essential to ensure adequate regulation of the police. However, the laws that presently govern the police are not tailored to balance the individual and societal interests at stake when police officers act, they lack coordination, and responsibility for regulating the police is haphazardly allocated. As a result, the present array of laws that polices American policing does not promote law enforcement that is maximally effective and protective of civil rights.

Policing the Police Through the Courts and Constitution

Constitutional law has long been a central component of legal efforts to police the police. In the early 1960s, under the leadership of Chief Justice Earl Warren, the United States Supreme Court authorized broad new remedies for violations of federal constitutional rights by local police. First, in 1961, in *Monroe v. Pape*, the Court interpreted 42 U.S.C. § 1983, a long-standing civil rights statute, to permit civil liability for police officers who violated federal rights even if those officials also violated state law. Later the same year, in *Mapp v. Ohio*, the Court imposed the exclusionary rule on states, mandating that state courts – like federal courts – exclude from state criminal trials evidence obtained in violation of the US Constitution's Fourth Amendment ban on unreasonable searches and seizures. These decisions gave victims the incentive and means to challenge police conduct, the courts the opportunity to refine constitutional doctrine, and the police new reasons to comply with constitutional norms.

A few years after the Court broadened remedies for constitutional violations by the police, it expanded the constitutional standards that set limits on police conduct. In *Katz v. United States* and opinions following it, the Supreme Court eliminated technical requirements that previously limited the scope of the Fourth Amendment, reframed Fourth Amendment analysis to bring a broader array of police practices within the Amendment's ambit, and emphasized the importance of warrants issued by neutral magistrates as a means of ensuring the constitutionality of police activity. In *Miranda v. Arizona* and its progeny, the Court applied the Fifth Amendment privilege against self-incrimination to police interrogations and imposed prophylactic rules and an exclusionary remedy to protect suspects during those interrogations. Like *Katz*, *Miranda* subjected additional police activity to judicial review, and, like *Mapp*, the case increased incentives for litigation and police compliance. Through these cases, the Court firmly established constitutional law as an important mechanism for regulating the police.

While the Supreme Court's doctrines have changed over time, its enterprise has not. Since 1968, the Court has considerably loosened the constraints on the investigation and detection of crime imposed by the Fourth and Fifth Amendments, substantially narrowed the scope of the exclusionary rule, and, after expanding § 1983 liability significantly in the 1970s, contracted liability in more recent decades. Even so, the paradigm arising from the Warren Court doctrines remains largely intact. Courts continue to apply the Fourth Amendment and the *Miranda* doctrine to impose detailed regulations governing police searches of homes, cars, and people and defining the procedural protections defendants are entitled to during custodial interrogations. Moreover, courts continue to require exclusion of evidence and to permit civil liability to remedy civil rights violations by the police. Thus, courts continue to delimit and protect constitutional rights through criminal cases and civil suits.

Although the judicial enterprise of defining and enforcing constitutional rights is and has been for many decades an important means to regulate the police, the judiciary cannot alone effectively prevent police misconduct, and the policy problem policing presents is not limited to constitutional rights. Courts suffer systematic limitations that inhibit them from undertaking the complex analysis of policing that effective regulation of the police demands, and constitutional rights are structurally unable to balance fully the harm police conduct imposes against its beneficial effects. Consequently, while courts and the Constitution play an important role in regulating the police, the judiciary and the Constitution can never successfully address the problem of policing without assistance.

Fourth Amendment doctrine makes apparent some of the limitations of courts. When courts evaluate whether a police search or seizure is "unreasonable" and therefore unconstitutional, courts frequently must appraise the nature of the intrusion on the individual, the strength of the government's interest in the intrusion, and the consequences for law enforcement of various possible rules and then balance these interests against each other. Assessing these considerations requires courts to draw factual conclusions about matters beyond the circumstances of the particular case, such as whether law enforcement can achieve its ends by alternate means. Similarly, when courts seek to interpret the exclusionary rule and civil remedies to prevent future constitutional violations by police officers, they do so based entirely on the likely effects of doctrines on police behavior. Thus, in order to tailor remedies to encourage lawful police behavior, courts must make correct judgments about what police officers value and how particular legal decisions are likely translate into incentives for them.

Although courts can define and vindicate constitutional rights only if they are capable of these kinds of empirical assessments and predictive analyses, they are notoriously ill suited to these tasks. For one thing, courts act with grossly inadequate data. Most Fourth Amendment questions are contested in state criminal cases in which neither party is likely to have adequate resources or incentives to effectively litigate significant empirical questions, and even a civil plaintiff hoping for compensation after a violent arrest cannot cost-effectively litigate many matters. As a result, courts deciding constitutional criminal procedure matters have no effective mechanism to obtain evidence about policing and incorporate it into their normative judgments. Even when courts are able to engage in effective empirical analysis, they have little opportunity or ability to adjust a doctrine as the facts and social science underlying it evolve, since they are bound by precedent and limited to deciding questions presented by the cases before them. As a result, courts have systematic difficulty formulating effective rules for the police and structuring remedies to prevent constitutional violations.

Even if courts could overcome these barriers, a major objective of police regulation would still remain beyond their reach. Ideal regulation of the police would take into account considerations such as how harmful any police action is to individuals and communities, as well as how it compares to other means of producing law and order in terms of cost, harm, effectiveness, and officer safety, in order to specify the conditions under which the police should harm individual interests for the greater good. If courts regulate the police, then the legal problem of policing is limited to constitutional violations. Constitutional rights are, however, ill suited to balance societal interests in law enforcement and individual freedom.

While constitutional rights accommodate both individual and societal interests, the well-known process by which constitutional rights are articulated and enforced dictates that rights provide only a limited tool for shaping police conduct. First, rights establish only minimum standards for law enforcement. Because individuals assert rights against the police rather than the other way around, constitutional criminal procedure rights are always framed as a ceiling on government action. They thus cannot reflect a full analysis of how to balance competing interests when the police enforce the law and individuals are harmed. Constitutional criminal procedure rights are therefore commands about what the police *cannot* do, not standards for what they *should* do. Second, because constitutional criminal procedure rights set unbreakable rules for police officers in advance, the rights themselves must be defined to permit law enforcement flexibility in pursuing societal aims even if this produces societally undesirable results in individual cases. That is, the "ceiling" set by constitutional rights must be higher – more generous to law enforcement – than would result from a full balancing of the interests at stake. Third, because rights are held and enforced by individuals, usually with respect to specific actions, they do a poor job of measuring aggregate costs and benefits of law enforcement activity. Intrusions by the police, such as stopping and frisking pedestrians to investigate crimes, may be constitutionally justified, and yet when multiplied thousands or hundreds of thousands of times, those intrusions may impose total costs that substantially undermine the quality of life in a community in a manner the Constitution cannot check.

Because of these characteristics, constitutional rights are structurally incapable of encouraging law enforcement to impose only necessary, fair, and efficient harms on legitimate interests in privacy, equality, autonomy, and the like. Instead, the Constitution provides only a rough measure of whether police conduct is justified. Adequately protecting individual and communal interests therefore requires nonconstitutional regulation of the police.

The Law of the Police

Government actors other than the courts, using legal tools other than the Fourth and Fifth Amendment, already create, empower, influence, and constrain the police and those that supervise them. There exists a vast web of law regulating the police, which can be divided into four categories: (1) law that authorizes or restricts the conduct in which police may engage; (2) law that remedies, punishes, or disincentivizes violations of the first category of law; (3) law that governs the hiring, management, and organization of police officers and departments; and (4) law that governs the availability of information about police activities.

Laws Authorizing and Restricting Conduct

Laws that regulate police conduct come from sources ranging from the Vienna Conventions on Consular and Diplomatic Relations, which limit police power to engage in searches, seizures of property, and arrests involving diplomats, to a San Francisco ordinance that prohibits police officers from questioning people about immigration status. Federal law, for example, contains more than a dozen statutes that regulate police searches, electronic surveillance, and access to private information. These statutes range from Title III, which governs federal wiretaps, to the Health Insurance Portability and Accountability Act, which restricts law enforcement access to and use of medical records. And federal law concerning the police is not limited to statutes governing searches and seizures. For instance, the Illegal Immigration Reform and Immigrant Responsibility Act provides for local police enforcement of federal immigration law, and the Law Enforcement Officers Safety Act allows qualified active and retired law enforcement officers to carry a concealed firearm anywhere in the United States, even if forbidden by state law. In addition to federal statutes, federal constitutional doctrines outside of the Fourth Amendment and *Miranda* also affect the police. Most notable among these, the First Amendment limits the conditions under which police officers may make arrests for breach of the peace, disorderly conduct, and resisting arrest. As these examples suggest, federal regulation of police conduct beyond constitutional criminal procedure doctrine is considerable.

State constitutions, statutes, and regulations regulate police conduct even more extensively. Local police officers are created by state law, which both grants power to police officers and restricts its exercise. Thus, state statutes permit police officers to engage in community caretaking and criminal law enforcement, allow officers to use force, require police to aid citizens in limited circumstances, mandate that officers arrest suspects in domestic violence cases, and forbid the police from asking questions unrelated to the subject of a traffic stop, for example. State constitutional law frequently mirrors federal law, regulating searches, seizures, and interrogations, but it is often interpreted more expansively to control police behavior that is beyond federal constitutional protection. Local ordinances further restrict police conduct, limiting the use of race in police actions, for example. Finally, departmental administrative rules – often known as general orders – provide the most important and extensive guidance to police officers about what they must, may, and may not do.

Laws to Remedy and Punish Violations

The second category, laws that provide remedies for violations of rules governing police conduct, includes federal constitutional decisions mandating that evidence illegally obtained by the police be excluded from federal criminal trials, as well as federal statutes that authorize criminal prosecution and private civil suits against local officers, departments, and municipalities. States provide similar remedies that supplement federal law, including statutes authorizing criminal prosecution, evidentiary exclusion, civil suits for damages and other relief, and sometimes, structural reform of departments. States, municipalities, and departments themselves often also provide other remedies, without analogs in federal law, for police conduct that is unconstitutional, illegal, or merely against administrative regulations. For example, most states authorize revocation of police officer certification for some kinds of misconduct, which prevents officers from reentering law enforcement in the same state. Municipalities frequently provide for civilian review of citizen complaints concerning police misconduct. And internal administrative rules within police departments establish procedures for taking, investigating, and resolving complaints and impose punishments for misconduct, often through an internal affairs unit. These internal administrative processes provide the most commonly used remedy for misconduct and, in many jurisdictions, are also subject by local ordinance, charter amendment, or public referendum to external review by an auditor or civilian oversight agency.

Laws Governing Hiring, Management, and Organizational Requirements

Some of the most important rules and laws governing the police are laws that set standards for hiring, training, and managing police officers. Policing is usually organized as a function of municipal or county government. Municipal ordinances, city charters, and other local laws dictate matters such as who hires and fires the police chief and thus often who ultimately controls policy in the police department. Although local governments largely oversee policing, federal and state laws check local political control of the police.

For example, every state has a peace officer standards and training commission (“POST”) that establishes minimum qualifications and training requirements for police officers as well as a process for licensing them. These commissions control how old and how educated police officers must be and what kind of criminal record they can have, factors that may affect whether officers are likely to engage in misconduct. They regulate the hiring process for officers, including whether officers must pass psychological or medical screening. Finally, they are the primary determinant of what kind and how much training police officers receive.

Other statutes affecting the police may impose less direct standards of police qualifications. The Lautenberg Amendment to the Gun Control Act of 1968, for example, changed federal gun laws to prohibit individuals convicted of misdemeanor domestic violence crimes from possessing a firearm. The Gun Control Act also applies the prohibition to those subject to a restraining order or dishonorably discharged from the military. As a result of this law, individuals in these categories cannot serve as sworn police officers in most jurisdictions, even if state law would otherwise allow them to serve. These federal requirements also affect the size of the pool of potential officers as well as who becomes an officer and therefore constitute one more piece of the web of laws that governs American policing.

In addition to laws mandating qualifications and training for police officers, federal and state laws constrain the organization of police departments and management of police officers. Federal law imposes constraints on how police officers are hired and fired through statutes and constitutional doctrines that prohibit employment discrimination. In addition, federal laws influence how officers are managed: The Fair Labor Standards Act, for example, influences when police work, including how shifts are structured to address overtime thresholds, a major financial and administrative issue for police departments. And the Fifth Amendment privilege against self-incrimination has a distinctive application to government employees that frequently restricts the use of statements compelled in administrative investigations of police officers against them in criminal prosecutions. State law is even more significant. State laws and regulations do everything from authorizing the existence of police departments to setting qualifications for the police chief. And state employment and labor laws, including civil service law, collective bargaining law, employment discrimination law, and law enforcement officer bills of rights, constrain most employment decisions by police departments and local governments.

Laws Governing Information About the Police

Finally, a variety of state and federal laws govern the production and distribution of information about the police. Most of the information that exists about policing is collected by police officers themselves pursuant to internal administrative policies demanding that officers complete form reports about arrests, uses of force, responses to citizen calls, and other interactions with the public. Other information is produced by departmental practices, such as installing video cameras in police cars or recording interrogations that occur at the station house. While these local practices are driven in significant part by departmental concerns, external legal constraints on departments influence them substantially. Thus, a department may collect data about the race of those targeted in traffic stops to fulfill the terms of a lawsuit settlement over racial profiling or in response to a local ordinance or state law. Of course, internal policies and state and local law can inhibit as well as facilitate the production of information about policing. Some departments collect little information on daily police activities, a practice that makes scrutinizing those activities difficult. And some states have applied laws governing recorded communications to prohibit private citizens from videotaping or audiotaping their interactions with the police.

Access to information about police conduct is often similarly shaped by a combination of departmental practice, state law, and civil settlements. State statutes and rules of civil and criminal procedure mandate government disclosure of some kinds of information about police conduct and departmental policies to civil plaintiffs and criminal defendants. Open records laws permit the broader public to obtain some information about police departments and their management, and some states expressly require departments to collect and disclose data about policing. But many states restrict public access to data about police misconduct, either through generally applicable statutory exemptions, such as exemptions for personnel records or for criminal investigations, or through specific exceptions for law enforcement. As a result, for example, internal disciplinary records and citizen complaints against an officer can be unavailable to the public, limiting political accountability for police practices.

Weaknesses of the Law Governing the Police

As this description suggests, local, state, and federal actors use a range of legal mechanisms to influence policing. Although this participation by institutions other than courts and regulation outside constitutional law is essential to effective governance of the police, the existing web of laws governing the police has arisen organically, the product of institutional arrangements and historical contingencies that have little to do with policing. As a result, these laws are often not tailored to serve the end of making law enforcement worth its costs, they are inadequately coordinated to promote efficient police practices, and responsibility for generating and enforcing them is haphazardly allocated among governmental institutions. As result, American policing does not fully ensure that police effectively control crime, fear, and disorder without imposing unjustifiable and avoidable costs on individuals and communities, despite extensive legal regulation.

First, existing law is not well designed to promote harm-efficient policing and cannot easily be made to do so. Legal regulation of the police should promote harm-efficient policing – that is, policing that imposes harms only when, all things considered, the benefits for law, order, fear reduction, and officer safety outweigh the costs of those harms. Presently, harm efficiency is largely left to the local political process, and police departments and local and state governments already take it into account in governing policing, at least to some degree. Some departments adopt, for example, internal regulations forbidding consent searches without reasonable suspicion of criminal activity, even though constitutional law demands no individualized suspicion before requesting consent to search. Nevertheless, public debates about police practices often focus on whether their conduct is constitutional or effective, not whether it is harm efficient, distracting political actors from this central question. Moreover, in order for regulatory actors to promote harm-efficient means of policing, they must have a basis for doing so, including an account of what the relevant harms and benefits of policing practices are and empirical work measuring and comparing harms and policing efficacy. Presently, however, insufficient data is collected about policing to provide the basis for harm analysis, and no substantial academic literature exists that can be used to compare policing techniques with respect to both effectiveness and harm. Until regulators mandate the necessary data production and collection, and scholars lay the conceptual and empirical groundwork for understanding harm efficiency, institutional actors will be stymied in their efforts to regulate the police toward this end.

Second, existing laws regulating the police interact in ways that may undermine the goal of ensuring that policing practices adequately protect individuals and communities. For example, courts have tailored civil remedies for unconstitutional policing to encourage departmental reforms likely to reduce misconduct, such as careful hiring practices and adequate training and supervision of police officers, but these judicial efforts cannot achieve their aim because state collective bargaining and civil service laws create countervailing incentives discouraging departments from implementing the same reforms. In a majority of states, civil service laws heavily regulate the treatment of public employees, including police officers. These laws allow costly legal battles when police departments demote, transfer, or fire an officer and thus impose significant additional costs on police departments trying to manage officers who commit misconduct. They therefore disincentivize precisely the same conduct that civil remedies are intended to encourage. Collective bargaining rights similarly deter department-wide changes intended to prevent constitutional violations. In states that mandate collective bargaining before a department changes its disciplinary process or promotion standards, departments face significant additional costs for increasing internal accountability for police officers, a key means of preventing police misconduct.

As these examples suggest, departments enter a legal minefield whenever they take employee action or make new policies. This is not to say that we should eliminate civil service laws or collective bargaining: These laws have complicated goals and effects that go beyond facilitating harm-efficient policing. Nor does it mean that civil remedies should not be used to incentivize reform. But until courts and legislatures consider the interactions among the myriad laws governing the police, legal efforts to regulate the police may not promote effective reform.

Third, although police departments, local governments, states, and the federal government all influence police conduct, as the example of the courts suggests, government institutions are not equally well suited for the tasks governing the police demands. Courts cannot, for example, determine the consequences of the real-world trade-offs between effective policing and individual freedoms that policing puts at stake. Regulating policing effectively therefore requires allocating institutional responsibility for regulating the police and choosing the best legal mechanisms for influencing police conduct. The existing allocation of responsibility and existing choice of mechanisms are unlikely to serve the goal of effectively regulating the police because they reflect historical and political contingencies rather than a considered choice about the best institutional approaches to regulation of the police. Unfortunately, reassigning responsibility for regulation to institutional actors with adequate capacity and incentive to ensure that the communal benefits of policing are worth its costs to individuals and communities cannot be done easily.

Police departments have enormous influence over the conduct of police officers, but they are nevertheless imperfect regulators of policing. Departments have difficulty engaging in broader causal analysis about how effective and harmful alternative law enforcement practices are. Moreover, police chiefs and local political actors lack sufficient incentive to ensure that individual interests are adequately vindicated in policing, because they are usually better rewarded for maintaining order and reducing crime than for protecting civil rights of the minority of residents targeted by police activities. Thus, police departments and local governments cannot be counted on to produce consistently harm-efficient policing.

Although states are critical in shaping police conduct, much of state regulation governing the police is aimed solely at ensuring law enforcement effectiveness rather than balancing effectiveness and civil rights. In addition, state law provides numerous mechanisms for remedying misconduct, but they seem uniformly weak. This state of affairs suggests that state actors have inadequate incentives to promote harm-efficient policing.

Thus, while local and state actors already promote civil rights to some degree and they could do so more, they likely cannot be expected to take adequate account of individual constitutional rights and constitutional interests that extend beyond them: They cannot and do not have sufficient reason to reach the appropriate trade-offs between effective policing and individual freedoms.

By contrast, federal actors can engage in thorough analysis about how to reduce harms to constitutional interests while engaging in effective law enforcement and may be able to encourage local and state action toward these ends. Congress already plays an active role in regulating the police. It has long regulated some police uses of private information deferentially regulated by the Court. It funds the Department of Justice's significant civil rights efforts, including criminal prosecution of abuses by local law enforcement, funding for nonprofits that promote civil rights in law enforcement, and technical assistance to police departments. And it is responsive to change: After court decisions restricted federal civil suits for injunctive relief against police departments, Congress passed 42 U.S.C. § 14141 authorizing the Department of Justice to bring suits for equitable remedies against police departments that engage in a pattern or practice of unconstitutional misconduct.

Though federal actors are capable of regulating the police, congressional attention to civil rights is piecemeal and irregular, and federal intervention in policing has long been politically controversial. Congress has not yet required even mandatory data reporting for local police departments, though the need for such data as a foundation for the regulation of the police is obvious, and though this could easily be carried out by existing Department of Justice components. More comprehensive regulation of the police, including the administrative analysis of the consequences of alternative law enforcement practices for individual and societal interests, is therefore unlikely and in any case may not justify its costs. As this analysis suggests, reaching a combination of government institutions with both the ability and the motive to regulate police officers and police departments effectively represents an ongoing challenge to efforts to ensure that police practices are carried out to promote law and order while minimizing harm.

Conclusion

The police have always represented both hope and harm. They contribute to social order but also threaten it. Though courts can judge the moral and historical imperatives that underlie constitutional rights, they cannot assess conditions on the ground or predict the consequences of legal rulings on civil rights and law enforcement. The project of defining and protecting constitutional rights inevitably requires input from other institutional actors. Although other government institutions have long participated in regulating the police, these efforts cannot approach ideal regulation of the police. First, political actors lack sufficient incentive and an adequate basis for determining when law enforcement should harm individual interests for societal ends, given the risks to human dignity and the costs and benefits of law enforcement activity. Second, because existing efforts to regulate the police operate without coordination, the multiple efforts to govern policing sometimes conflict and undermine rather than reinforce the policy goals of balancing the costs and benefits of law enforcement. Finally, various political institutions face structural limitations and incentives that may be incompatible with effective regulation of the police. Future efforts to improve law governing the police must address these substantial challenges.

Related Entries

- ▶ Civil Remedies
- ▶ Control of Police Misconduct
- ▶ Law of Police Interrogation
- ▶ Law of Police Searches
- ▶ Law of Police Seizures and the Exercise of Discretion
- ▶ Law of Police Use of Force
- ▶ Law of Undercover Policing
- ▶ History of Police Unions

Recommended Reading and References

Amsterdam AG (1974) Perspectives on the fourth amendment. Minnesota Law Rev 58:349

Armacost B (2004) Organizational culture and police misconduct. George Washington Law Rev 72:453

Avery M, Rudovsky D, Blum K (2012) Police misconduct: law and litigation. 3d ed

Bradley, CM (1993) The failure of the criminal procedure revolution

Harmon RA (2012) The problem of policing. Michigan Law Rev 110:761

Katz v United States 389 U.S. 367 (1967)

Mapp v Ohio 367 U.S. 643 (1961)

Miller ML, Wright RF (2005) Criminal procedures: cases, statutes, and executive materials. 2d ed

Miranda v Arizona 384 U.S. 436 (1966)

Monroe v Pape 365 U.S. 167 (1961)

Skogan WG, Frydl K (eds) (2003) National research council. Fairness and effectiveness in policing: the evidence, Washington, D.C

Stuntz WJ (2001) The pathological politics of criminal law. Michigan Law Rev 100:505

Walker S (2005) The new world of police accountability

Legal Frameworks for Third-Party Policing

Janet Ransley
School of Criminology and Criminal Justice,
Griffith University, Mt. Gravatt, QLD, Australia

Synonyms

Plural Policing; Regulatory Crime Control; Third Party Policing

Overview

Public police are now one part of a network of law enforcement that includes private police and security, regulatory agencies, businesses, schools, and private individuals. Police have always formed partnerships to help them in responding to, and preventing crime, but increasingly these partnerships are moving beyond joint approaches to solving community problems. Third-party policing is an approach that rests on the use of legal levers available to non-offending third parties under regulatory and private law frameworks. There has been a blurring of these frameworks in recent times, so that many noncriminal levers are available for police and their partners to use for crime reduction purposes. The types of levers commonly used in third-party policing are discussed, along with the types of crime and disorder problems which they have been used to address. Challenges to the broader use of regulatory frameworks are overviewed, and future directions and especially the need for further research, are suggested.

Introduction

Third-party policing (TPP) occurs when police form crime control or prevention partnerships or networks with non-offending third parties (Mazerolle and Ransley 2005; Buerger and Mazerolle 1998). Police partner with a range of state and non-state agencies, private organizations, and individuals to focus on crime reduction or prevention goals. Typical partners include housing agencies, property owners, rental agents, health and building inspectors, private businesses and professionals, parents, schools, and private security providers. Typical crime problems at which TPP has been directed include illicit drugs, alcohol-related violence and disorder, young offenders, and property offenses.

The distinctive feature of TPP, compared with other innovative and proactive policing strategies such as problem-oriented policing (POP), is its placement within a legal and regulatory framework. This framework enables police to use the authority and powers of their voluntary, co-opted, or coerced partners to help control crime problems. In particular, police gain access through their partners to a range of criminal and noncriminal legal mechanisms or levers. These levers include civil regulation and remedies, along with contractual and tortious rights, which can greatly extend the reach of police (Mazerolle and Ransley 2005, 2006). Unlike POP, TPP goes beyond merely "harnessing the crime control capacities of third parties" (Cherney 2008: 631) in identifying and developing responses to crime problems. It draws on developments in regulatory theory and practice that allow police to use new techniques, institutions, and mentalities of control. Police become part of the trend to regulatory justice (Crawford 2003, 2009) that can extend and improve upon traditional criminal justice responses. In its focus on problem places, people, and activities, TPP also brings together opportunity and place-based theories of crime and crime prevention with innovative policing strategies, and gives police a new role as regulators of crime (Eck and Eck 2012; Mazerolle and Ransley 2012).

One appeal of TPP for governments and law enforcement is that non-offending third parties assume, or are made to assume, some of the responsibility and costs of controlling and preventing crime. Especially in times of fiscal restraint this has economic advantages (Eck and Eck 2012), as well as giving police access to new levers and remedies, such as civil orders and injunctions, property closures, parental responsibility contracts, and asset forfeiture schemes. These levers can be directed at the underlying causes and supports of crime, and can have a preventive and intelligence-based focus rather than the purely reactive responses of most traditional policing practices. A second appeal of TPP is that by relying on noncriminal measures, some of the restrictions and protections of the criminal justice system such as high evidentiary standards and burdens of proof, can be avoided (Mazerolle and Ransley 2005; Crawford 2009).

The discussion below canvasses these issues, beginning with an analysis of the regulatory and legal context of TPP. It compares TPP with traditional approaches to crime control and prevention, and then discusses in detail the legal frameworks most used in TPP. The discussion moves on to some of the key issues and debates arising from the topic, including the potential impact of TPP on increased criminalization and net-widening, and for inequitable and discriminatory policing. The final comments address future directions for TPP and the need for more evidence-based research on what levers work best and when, and why this is so.

Fundamentals

Context: Regulatory Approaches to Criminal Justice

Third-party policing approaches are part of a broader shift in regulation and governance which has occurred in many societies over the past 40 years, and which has led to new responses to crime along with many other social and economic problems (Feeley and Simon 1992; Ericson and Haggerty 1997; Garland 2001; Ericson 2007). The new approaches to governance have involved a move away from state sovereignty and control to networks of power (Braithwaite 2000), a diffusion of responsibility (Garland 2001), and an expectation that non-state actors and individuals must contribute to responses and solutions. Regulatory theorists see this as a shift from state-centered command and control to new modes of pluralistic, networked, or nodal regulation (Burris et al. 2005).

Central to this shift has been the rise of risk as an organizing concept, so that the new forms of regulation rely on notions of risk assessment, identification, and analysis, rather than on general rules applied equally to all regulated individuals (Sparrow 2012). Feeley and Simon (1992) applied this concept to criminal justice, arguing that it has seen a shift from a focus on individual offenders, the assignment of blame and responsibility, and goals of treatment and rehabilitation. In their "new penology," the focus instead is on actuarial classifications of risk and dangerousness and reliance on selective surveillance and incapacitation to reduce crime opportunities (Willis and Mastrofski 2012). The shift from old to new penology is also a shift from seeing crime as deviance to crime as normalized, and from ideas of treatment and rehabilitation to the management of risk (Garland 2001).

At the same time, the development of the new regulatory state (Braithwaite 2000) has led to a "proliferation of new administrative agencies as well as an enormous expansion in private policing. Regulatory and private policing entities are part of a new division of policing labour with the public police" (Ericson 2007: 387). Regulatory styles have also shifted, from a focus on command and control and detailed prescriptions and inspectorate models, to new forms of responsive and smart regulation tailored to particular industries and contexts, and in which risk identification and management again plays a central role (Mazerolle and Ransley 2005).

As a result of these trends to risk, regulation, and pluralization, public police have now been joined by regulatory agencies with many police-like powers, and private policing organizations, in overlapping networks of security. Policing is now provided by "a growing plethora of governmental and non-governmental providers" (Stenning 2009: 22; Jones and Newburn 2006). Contemporary policing is marked by diversity, complexity, and a "messy reality" where rather than smart and responsive regulation playing a less coercive role than traditional law enforcement, tools of intervention have become increasingly interchangeable and overlapping (Crawford 2006: 469). In this environment the role of public police has changed, so that policing services are increasingly provided by networks of public, private, and welfare agencies, with public police as one node of the network (Ransley and Mazerolle 2009), albeit a coordinating or anchoring node (Ericson and Haggerty 1997; Crawford 2006).

In addition to these new regulatory approaches to criminal justice and policing, new theories have been developed about crime control and prevention. The rise of crime opportunity theory and its focus on places, patterns, routine activities, and rational choice has led to recognition that regulation and regulatory measures or instruments have a central role to play in criminal justice and crime control (Eck and Eck 2012; Ayling et al. 2008; Mazerolle and Ransley 2012). TPP provides an organizing framework for how this occurs, and highlights the role of police in developing and implementing new regulatory approaches to crime control and prevention.

As discussed earlier, TPP occurs when police persuade or coerce non-offending partners into assuming crime reduction or prevention responsibilities. As such, police form formal or informal partnerships directed at specific crime problems. The development of police partnerships is not new – as Smith and Alpert (2011) point out, Sir Robert Peel's eighteenth-century "new police" rested on a community partnership model. Contemporary versions of community policing also rely heavily on policing as a series of collaborative partnerships (Willis and Mastrofski 2011; Skogan 2011) focused on identifying and solving community problems. The "joint policing" model involves "collaborative, proactive, preventive actions to reduce or remove crime and to protect and secure spaces and places in lieu of traditional reactive models" (Smith and Alpert 2011: 136). Similarly, problem-oriented policing approaches often focus on police working with community members to identify problems and their solutions. What sets third-party policing apart from these other strategies is the way in which it involves police harnessing not just partners but their formal and informal legal powers over offenders or potential offenders. These partnerships occur within a context of changing legal frameworks.

Changing Legal Frameworks

The messy reality of policing now draws on a range of legal frameworks, in addition to traditional criminal justice models. The major distinction between these frameworks has been in their goals and targets – traditionally criminal law has focused on detecting and prosecuting those who commit morally wrong offenses, while regulatory and private law measures have focused on governing legal business and private activities (Mazerolle and Ransley 2012), mainly using negotiation, inspection, and administrative and civil sanctions, as shown in Table 1. As an organizing and analytical mechanism, the different categories of law are described by reference to their goals, attributes, targets, methods, outcomes, actors, and styles (see Mazerolle and Ransley 2005: 67–69; Cheh 1998).

In the context described above, where trends in governance, risk, and pluralization have broken down traditional categories, there has been a blurring of legal frameworks. As shown in Table 2, criminal law now draws on regulatory and civil law mechanisms, and vice versa. Mazerolle and Ransley (2005: 67) refer to this as a convergence of legal frameworks. In this environment, criminal law makes use of civil processes and remedies, while in both regulatory and private law serious misbehaviors are criminalized. The major implication of this convergence in legal frameworks has been to facilitate the use in criminal law enforcement of a large range of sanctions and processes outside the traditional criminal justice system, such as civil remedies and administrative sanctions. The principal facilitator of this convergence has been the concept of risk. Proactive policing strategies have focused on risky places (i.e., hotspots of crime) and people (i.e., repeat offenders, particular groups), while regulators have adopted self-regulation approaches based on risk assessment (Sparrow 2012). The main effect of this convergence can be seen in Table 2 in that there has been a blurring of targets, methods, outcomes, actors, and styles across the different categories of law, in contrast to the relatively clear-cut divisions shown in Table 1.

The rise of TPP is a reflection of this convergence, in that it enables police to draw on a much wider variety of legal levers than was previously available to them. The next section discusses some of those levers.

Legal Levers for Third-Party Policing

Despite this convergence of legal frameworks, it is important to understand the source of law that underpins each legal lever used in TPP, because this source affects the nature, extent, and limitations of the lever. Table 3 summarizes the levers commonly used in TPP and their main features. For each type of lever, there is a description of the source of authority (whether legislation, regulation, contract, or tort), the extent to which the lever applies (i.e., generally or targeted), and the types of legal outcomes that can arise. The final column indicates the range of non-offending third parties who are utilized by each lever to achieve crime control or reduction benefits, through the co-option of their legal authority.

Table 3 is not a comprehensive list of legal levers available, but it does illustrate the range and breadth of measures, and of potential third parties, that can be used in TPP. Mazerolle and Ransley (2005) give comprehensive examples of the ways in which each of these levers has been used across different jurisdictions, in the two broad areas of controlling and prevent drug problems, and other forms of crime. For drug problems, laws and regulations focused on premises (e.g., crack houses, methamphetamine laboratories, derelict buildings), and accompanying nuisance abatement civil suits, have been used in many places in the United States and Australia. These strategies involve police partnering with city and state inspectors and sometimes neighbors, to require property owners to become more responsible for controlling activities on their premises that lead to drug problems. A range of civil sanctions can be used (fines, injunctions, rectification orders, loss of license or permit, forfeiture) to persuade or coerce these third parties to assume crime control responsibilities. This not only shares responsibility but shifts some of the costs associated with crime reduction to third parties. The use of civil levers removes the need

Legal Frameworks for Third-Party Policing, Table 1 Traditional categories of law

Focus	Criminal	Regulatory	Private
Goal	Detection, prosecution, punishment, rehabilitation of offenders	Orderly conduct of economically desirable activities	Dispute resolution in private financial and family activities
Activities	Morally repugnant offenses	Business activities	Commercial and domestic interactions
Targets	Individual offenders	Businesses	Private individuals and groups in dispute
Methods	Investigation, arrest, criminal trial	Conciliation, negotiation, self-regulation, inspection, limited prosecution	Private court action to enforce rights under contracts and tort
Outcomes	Fines, incarceration	Penalties, rectification orders, license revocation	Damages, injunctions, specific performance
Actors	Police, prosecutors, courts, correctional staff	Regulatory agencies, local governments, industry associations	Individuals in private relationships, civil courts
Style	Guilt and punishment for individual offenses	Creation of business obligations, consequences enforced by sanctions	Creation of private obligations, consequences enforced by sanctions

Source: based on Mazerolle and Ransley (2005, 68)

Legal Frameworks for Third-Party Policing, Table 2 Converging categories of law

Focus	Criminal	Regulatory	Private
Goal	Identification and prevention of potential crime risks, detection and prosecution of offenders	Identification and prevention of potential regulatory risks	Prevention and resolution of private disputes
Activities	Morally repugnant offenses, antisocial behavior	Business activities	Commercial and domestic interactions
Targets	Groups and places with high risk of crime, offenders, nuisances	Businesses and industries at risk of noncompliance	Individuals and groups at risk of or in dispute
Methods	Investigation, arrest, trial, surveillance, statistics, profiling, incapacitation, responsibilization	Codes of practice, reporting, statistics, audit, enforcement pyramids, responsive regulation, criminalization	Private court action to enforce contract and tort rights, statutory remedies, criminalization
Outcomes	Fines, incarceration, forfeiture, injunctions, preventive detention, behavior contracts	Penalties, rectification orders, license revocation, fines, incarceration, damages, injunctions	Damages, injunctions, specific performance, fines, incarceration
Actors	Police, prosecutors, courts, correctional staff, local governments, health and housing agencies, community groups, forensic specialists	Regulatory agencies, local governments, industry associations, consumer groups, international regulators	Individuals in private relationships, government and community agencies, police, prosecutors, courts, tribunals, welfare agencies
Style	Guilt and punishment, consequences enforced by sanctions, therapeutic, restorative	Guilt and punishment, consequences enforced by sanctions	Guilt and punishment, consequences enforced by sanctions, therapeutic

Source: based on Mazerolle and Ransley (2005, 68)

for police to gather evidence to a criminal standard of proof, to prove individual responsibility for breaches, and often also involves a reversal of the burden of proof, so that property owners must show they were not aware of criminal activities to avoid responsibility. Other drug problem-related forms of TPP include monitoring and reporting schemes, such as when doctors, pharmacists

Legal Frameworks for Third-Party Policing, Table 3 Legal frameworks for third-party policing

Extent of application	Legal outcomes	Types of TPP levers and interventions	Third parties targeted by the intervention
Statutes			
General or a specified population, e.g., liquor retailers, parents, gangs	Criminal or civil action leading to criminal or civil penalties, or administrative measures	*Orders to control behavior*, e.g., antisocial behavior orders; truancy and gang orders, child welfare and domestic violence orders, mental health and vagrancy laws	Parents, schools, non-offending gang members, institutions and shelters, housing authorities
		Movement and association limits, e.g., curfews; traffic, move-on, and crowd control powers; probation and community corrections conditions limiting criminal contacts	Parents, local councils, probations/corrections staff
		Conduct licensing, e.g., alcohol, firearms and prostitution licensing, pawn/second-hand shops and dealers	Retailers, service providers
		Formalized surveillance, e.g., offender notification, probation and community corrections reporting conditions	Probations/corrections staff
		Property controls, e.g., drug house and rave site limits	Property owners, rental agents local councils
		Mandatory reporting, e.g., methadone prescribing, cash transactions, chemical sales, child and spousal abuse	Doctors, pharmacists, health funds, banks, retailers, schools
		Civil forfeiture, e.g., assets gained through crime, "hoon" cars	Banks, vehicle licensers
Regulation/subordinate legislation			
General, within a specified area, e.g., local council or housing authority area	Criminal or civil action leading to criminal or civil penalties, refusal of consent, eviction, rectification orders, property forfeiture or confiscation	*Orders under regulatory codes*, e.g., building, fire, health, safety, noise, animals, liquor licensing, environmental, public housing, parking, venue by-laws and codes requiring closure, rectification, removal of hazards, or restrictions on use	Property and business owners, local councils, housing authorities, regulatory agencies
		Product and service standards, e.g., requiring vehicle immobilizers	Manufacturers and service providers, standards associations
		Controlled zones, e.g., begging, busking, drug and alcohol free areas, limits on street prostitution	Local councils, retailers and business owners, prostitutes
Contract			
Specific – parties to a contract	Civil action for specific performance, damages or injunction, non-renewal of contract, eviction, loss of bond	*Enforcement of conditions*, e.g., property use and maintenance in private housing and commercial leases; crime reduction measures in service provision contracts, e.g., lighting and security	Property and business owners and associations, housing authorities

(*continued*)

Legal Frameworks for Third-Party Policing, Table 3 (continued)

Extent of application	Legal outcomes	Types of TPP levers and interventions	Third parties targeted by the intervention
		standards in commercial car parks	
		Incentives, e.g., insurance bonuses and rebates for crime prevention measures	Landlords, insurers
Tort			
Specific – duty of care owed	Civil action for damages or injunction	*Actions for nuisance, trespass, or negligence*, e.g., against noise, physical disorder, pets, physical access, breach of duty of care, misfeasance	Landlords, local councils, housing authorities, liquor retailers, licensed venues

or chemical manufacturers, wholesalers and retailers are required to record and report transactions involving drug precursors. Such schemes co-opt these third parties into systems of surveillance and sources of intelligence for proactive policing.

For other nondrug crimes, the most widespread use of TPP has been in schemes designed to reduce or prevent antisocial behavior and disorder. The most comprehensive of these schemes is the antisocial behavior agenda introduced in Britain from 1998 onward, which featured a range of civil orders and contracts imposed to control offending and pre-offending activities. Crawford (2009: 817) identifies 18 new tools used in this scheme, including civil orders (e.g., antisocial behavior, drug intervention, parenting, individual support, child curfew, dispersal, drink banning orders) and contracts (e.g., acceptable behavior, parenting, tenancy contracts). All of the measures are civil, involving a mix of regulatory and contractual techniques, but breach of most leads to a criminal sanction. The scheme has been much criticized, and while the new government elected in 2010 announced its intention of withdrawing the system, its most recent proposals suggest that many of the new tools will be rebranded and retained (see The Independent, 23/5/2012). While the British antisocial behavior scheme is comprehensive, many of its individual components are seen in other jurisdictions. In particular, measures directed at parental responsibility for juveniles, such as truancy and curfew orders, are common in Australia and some parts of the United States (see Mazerolle and Ransley 2005). Most jurisdictions have anti-vagrancy and begging regulations, and several Australian states have introduced civil measures directed at alcohol-related violence (see Mazerolle et al. 2012). These types of levers induce third parties, whether parents, schools, club-owners, or city administrators, to accept responsibility for reducing crime and disorder using civil or private law measures available to them.

As discussed above, increasingly, this type of approach is coupled with crime opportunity approaches that seek to identify and reduce facilitating environments for crime. As Eck and Eck (2012: 308) note, “rather than viewing crime as simply a matter between offenders and police, a place focus requires consideration of the morality of crime facilitation by third parties.” They suggest that such approaches can not only reduce crime, but also reduce the economic demands on police and policing, by sharing responsibility and costs. However, the extension of responsibility in this way, and the increased use of TPP and regulatory strategies, is not without problems and controversies, and the next section examines some of these.

Key Issues and Controversies

A threshold problem for TPP approaches is that, typically, police agencies are different to regulators and other third parties. Police do not view themselves as regulators, have different systems of training and organization, and a different culture. Sparrow (2012: 347) refers to this as the "divide" between police and regulators and attributes it to police unfamiliarity with the role of regulations, the distinctive police culture and domain, and the special risks they face. These differences can impede police from recognizing opportunities for TPP, and identifying regulators and legal levers that can be used to respond to particular crime problems. However, Sparrow (2012) also sees similarities between the shift in regulatory approach to risk-based strategies and the rise of POP and other innovative policing strategies which are similarly based on risk. He argues that POP has become single-dimensional, focused almost entirely on place-based interventions, and that increased use of a toolbox of regulatory strategies can expand the reach of innovative policing. But Sparrow (2012: 354) concedes a significant barrier to this occurring – how can police and regulators agree on who does what to deal with particular problems? Tilley (2012) goes further to ask who should pay for what and who should bear the costs of crime prevention efforts? This is particularly important when crime and its prevention is not the core business or concern of most third parties identified in the discussion above.

Tilley (2012) and Eck and Eck (2012) identify other potential problems with regulatory approaches to crime control and prevention. Spreading responsibility to non-offending third parties can create costs for the blameless, as well as those who facilitate crime, for example, when all bar owners are required to upgrade security in response to violence that occurs only at a few sites. Some strategies can be seen as too interventionist, such as the requirement for bars to use plastic glasses or serve low-alcohol drinks (Mazerolle et al. 2012). Such measures can add to the cost of doing business and favor those traders who cut corners rather than those doing the right thing.

Mazerolle and Ransley (2005) grouped the disadvantages of TPP as unintended side effects: the disproportionate allocation of policing and regulatory resources creating either over- or under-policing; the displacement of crime problems; the intensification of crime and disorder in already disadvantaged areas; the co-option of regulatory resources and agendas away from their core concerns, such as public health and safety; and the dilution of criminal law protections and restraints. Additionally, there is a danger of regulatory creep that extends the acceptable limits of government intervention in the lives of citizens and businesses. Crawford (2009: 818–819), discussing the British antisocial behavior agenda, notes similar concerns including: the use of civil measures to evade criminal justice process; the introduction of new forms of hybrid liability that potentially lead to increased criminalization and net-widening; the reliance on subjective perceptions of risk in place of objective assessments of past conduct; the privileging of public anxieties, fears, and concerns as triggers for formal action (as in dispersal orders); the trends to preemption and individualized rather than generally applied controls; and the substitution or supplementation of judicial decision making by police and regulatory discretion.

Many of these concerns and issues remain unexamined in any systematic way, with the possible exception of the British antisocial behavior system. How TPP affects the communities in which it is practiced, whether it does so equitably and with accountability are empirical questions that remain largely unanswered. Also insufficiently examined is the effectiveness of TPP – does it work, and if so, when and in what forms? Mazerolle and Ransley (2005) conducted an extensive review of the evaluation literature, identifying 80 studies of TPP strategies. They found that business owners are the third parties most often targeted to share crime control or reduction responsibility, in dealing with problem bars, property crime in unsecured car parks, drunks and disorderly behavior in public places, and street prostitution. Other third parties included parents, schools, liquor licensing authorities, car manufacturers, local councils and public housing authorities. Overall, the use of these third parties and their legal levers was found to be an effective tactic for crime control.

Future Directions/Conclusion

The use of TPP is part of a shift to regulatory justice that positions police at the center of crime control and prevention networks. Most TPP currently occurs in an ad hoc and variable way, apart from the British antisocial behavior agenda. In times of financial constraint, TPP will become increasingly attractive to governments wanting to share costs with business and individuals, and with police agencies wanting to draw on the resources and expertise of other regulatory bodies.

While there is some evidence that TPP is effective (see Mazerolle and Ransley 2005; Eck and Eck 2012), there is a clear need for more research that encompasses at least the following:

- The full range of legal levers for TPP that already exist and the contexts in which they are available for use in TPP strategies
- The effectiveness and limits of various levers in various situations – what works, when, and why
- The full costs of financial burdens shifted to others through TPP, and the impact on their private or regulatory functions and concerns that results
- The extent to which TPP is used differentially in different communities – are disadvantaged neighborhoods over-targeted or neglected?
- The extent to which TPP is used differentially against different social and demographic groups – to their advantage or disadvantage
- The extent to which TPP has resulted in expanded criminalization, leading to new offenses and criminalizing conduct previously seen as disorder or nuisance
- The extent to which TPP has resulted in net-widening, particularly by criminalizing noncompliance with regulatory and civil sanctions
- The extent to which TPP has diluted criminal justice processes, by criminalizing those against whom noncriminal justice processes have been used
- The impact of TPP on regulators and the extent to which their agenda has become co-opted by crime reduction concerns

There seems little doubt that the trend to regulatory justice, and the accompanying rise of TPP strategies, is occurring. The challenge for police is to manage this trend in a thoughtful and useful way. The challenge for researchers is to study the features, utility, and costs of the trend and to suggest ways of better harnessing its advantages while minimizing the potential costs and disadvantages.

Related Entries

- ▶ Cost-Benefit Analysis
- ▶ Civil Remedies
- ▶ Pulling Levers Policing
- ▶ Third Party Policing and School Truancy
- ▶ Third-Party Policing

Recommended Reading and References

Ayling J, Grabosky P, Shearing C (2008) Lengthening the arm of the law: enhancing police resources in the twenty-first century. Cambridge University Press, New York

Braithwaite J (2000) The new regulatory state and the transformation of criminology. Brit J Criminol 40(2):228–238

Buerger M, Mazerolle L (1998) Third party policing: a theoretical analysis of an emerging trend. Justice Quart 15(2):301–308

Burris S, Drahos P, Shearing C (2005) Nodal governance. Aus J Legal Philos 30:30–58

Cheh M (1998) Civil remedies to control crime: legal issues and constitutional challenges. In: Mazerolle L, Roehl J (eds) Civil remedies and crime prevention. Crime Prevention Studies, Monsey, New York: Criminal Justice Press, Vol 9, pp 45–66

Cherney A (2008) Harnessing the crime control capacity of third parties. Policing: Inter J Police Strat Manage 31(4):631–647

Crawford A (2003) Contractual governance of deviant behaviour. J Law Soc 30:479–905

Crawford A (2006) Networked governance and the post-regulatory state? Theoretical Criminol 10:449479

Crawford A (2009) Governing through anti-social behaviour: regulatory challenges to criminal justice. Brit J Criminol 49:810–831

Eck J, Eck E (2012) Crime, place and pollution: expanding crime reduction options through a regulatory approach. Criminol Public Policy 11(2):281–316

Ericson RV (2007) Rules in policing. Theoretical Criminol 11(3):367–401

Ericson RV, Haggerty KD (1997) Policing in the risk society. Toronto University of Toronto Press, Toronto

Feeley M, Simon J (1992) The new penology: notes on the emerging strategy of corrections and its implications. Criminology 30(4):449–470

Garland D (2001) The culture of control: crime and social order in contemporary society. Oxford University Press, Oxford

Jones T, Newburn T (2006) Plural policing: a comparative perspective. Routledge, London/New York

Mazerolle L, Ransley J (2012) Crime, place and pollution: expanding crime reduction options through a regulatory approach. Criminol Public Policy 11(2):335–344

Mazerolle L, Ransley J (2005) Third party policing. Cambridge University Press, New York

Mazerolle L, Ransley J (2006) The case for third party policing. In: Weisburd D, Braga A (eds) Police innovation: contrasting perspectives. Cambridge University Press, New York, pp 191–206

Mazerolle L, White G, Ransley J, Ferguson P (2012) Violence in and around entertainment districts: a longitudinal analysis of the impact of late-night lockout legislation. Law Policy 34(1):55–79

Ransley J, Mazerolle L (2009) Policing in an era of uncertainty. Police Practice Res: Inter J 10(4):365–381

Skogan W (2011) Community-based partnerships and crime prevention. Criminol Public Policy 10(4):987–990

Smith HP, Alpert GP (2011) Joint policing: third parties and the use of force. Police Practice Res 12(2):136–147

Sparrow M (2012) Joining the regulatory field. Criminol Public Policy 11(2):345–359

Stenning P (2009) Governance and accountability in a plural policing environment – the story so far. Policing 3(1):22–33

The Independent (23rd May 2012) 'Crimbos' replace Asbos – but will they rush children into custody? Accessed at http://www.independent.co.uk/news/uk/crime/crimbos-replace-asbos–but-will-they-rush-children-into-custody-7778907.html

Tilley N (2012) Crime reduction: responsibility, regulation and research. Criminol Public Policy 11(2):361–378

Willis JJ, Mastrofski SD (2011) Innovations in policing: meanings, structures and processes. Annual Rev Law Soc 7:309–334

Willis JJ, Mastrofski SD (2012) Compstat and the new penology: a paradigm shift in policing? Brit J Criminol 52(1):73–92

Legal Interventions

► Lawful Killings

Legal Rules, Forensic Science and Wrongful Convictions

Carole McCartney
School of Law, University of Leeds, Leeds, UK

Synonyms

Junk science; Miscarriages of justice

Overview

Law and science have long had a strained relationship although their tendency to "clash" may have been exaggerated over the years. The source of their disharmony has most often been seen as the result of the two domains having both dissimilar methods and goals. Regardless of the extent to which this is true, there remain apposite and acute concerns regarding the status of science utilized by the law, in particular forensic science and its interplay with criminal law. Without delving into the philosophy of science or law, or the psychology of the courtroom, it is considered here whether legal rules have, and can, prevent flawed scientific evidence from entering the courtroom by empowering judges to rule questioned scientific evidence inadmissible. It concludes that where evidential hurdles for forensic science exist, decisions to permit forensic science into evidence may still appear arbitrary, as well as inconsistent between and within jurisdictions. In addition, such hurdles may come too late in the criminal process to prevent all injustices. Current attempts to regulate forensic science are welcome but do not yet go far enough to ensure that wrongful convictions will not continue to occur, with forensic science and scientists playing a significant role in these miscarriages of justice and judges unable to execute their gatekeeper role effectively and consistently.

Fundamentals

The prefix "forensic" simply means the use of science in legal fora: applying science to answer a legal question. Efforts have been made to circumscribe the field of "forensic science," with the UK's (now defunct) Forensic Science Service stating that the task of forensic science is "to serve the interests of justice by providing scientifically based evidence relating to criminal activity," and the definition given in the US Congress Forensic Science and Standards Bill of 2012 3(a): "'forensic science' means the basic and applied scientific research applicable to the collection, evaluation and analysis of physical evidence, including digital evidence, for use in investigations and legal proceedings, including all tests, methods, measurements and procedures."

Most often, forensic science is a subdiscipline of a larger scientific body, such as pathology, toxicology, archaeology, or genetics. There are also a variety of "experts" who may be called upon to testify on matters in issue at trial (e.g., forged documents, firearms), each being a "forensic specialist" for the purposes of a criminal trial. There is potentially no limit to the scope of physical evidence (or "trace" evidence) that can be utilized, from the smallest fragment of glass, the most obscured fingerprint, or the tiniest swab of human tissue to the entire scene of a major disaster or the realms of cyberspace. There are, therefore, potentially no limits on the number of disciplines that may be utilized in a forensic investigation. Herein lies one of the pivotal difficulties with forensic science: the multiplicity of techniques, technologies, and disciplines, which can sit under the "forensic science" umbrella, risks over-inclusion. It is this over-inclusion that lies at the crux of some of the frustrations with forensic science. While some forensic technologies are born of rigorous scientific testing and experimentation (the oft-used example being DNA profiling), many others have no, or a minimal, scientific basis or grounding in experimentation and testing.

This variety of disciplines and techniques and their great variances in reliability and validity create an issue for users of forensic science: how to discern between those reliable and valid techniques that can assist, and those that may be insufficiently reliable and valid, and can mislead. So, while on first blush it would appear contradictory to discuss forensic science and judicial oversight alongside wrongful convictions, this question occupies the minds of legal professionals, researchers, practitioners, and forensic experts alike: how can courts ensure that valid and reliable scientific evidence, and *only* valid and reliable scientific evidence, is adduced during trials, thereby using forensic science to *prevent* wrongful convictions?

Ruling potentially unreliable or invalid forensic evidence inadmissible at trial, while clearly having an impact on that specific trial, may also have important wider ramifications. If a particular expert, expertise, or evidence type is regularly deemed inadmissible, then police or legal professionals will lose confidence in them and stakeholders may insist upon greater controls or regulation. It may be that, as we have seen in the USA, Canada, and the UK in recent years, political demands are made for changes to the forensic science community. Public confidence in the operation of the criminal justice system is an essential prerequisite of any legitimate legal system, and if forensic science brings that system into disrepute or diminishes public confidence, then the forensic science community should not be surprised if their reputation en masse is tarnished and stricter regulation is insisted upon.

Key Issues/Controversies

The forensic scientist (or rough equivalent) has been popularly portrayed in both factual and fictional media as a heroic figure in the fight against crime. In the twenty-first century, that portrayal has had to become more nuanced as forensic science has been pinpointed as playing a role in causing wrongful convictions while failing to prevent others. Forensic scientists have been accused of inter alia: negligence, corruption, and perhaps most critically, of not being "scientists" at all.

The growing international "innocence movement" has paradoxically utilized modern forensic techniques and technologies to free innocent prisoners, simultaneously demonstrating the flaws of many forensic techniques that were used to convict those prisoners and the failure of the courts to prevent misrepresented or misinterpreted forensic evidence from being relied upon at trial. Yet wrongful convictions most often result not from a single error but a composite of failures that can include, but are clearly not limited to, the gathering, interpretation, and communication of forensic evidence. It is most often human fallibility that lies at the heart of wrongful convictions. Forensic science can then overcome the failings of many other, unreliable, and often highly questionable evidence types, regularly relied upon at court. For this reason, forensic science cannot fairly be portrayed as the cause of wrongful convictions, as it plays probably a more significant role in prevention and cure; these three roles are considered, before turning to the place of forensic science in the courtroom.

During police investigations, the gathering of scientific evidence cannot prove any one theory of events, or conclusively prove anyone guilty of a criminal offense in isolation of other evidence, but it can be incredibly valuable in explaining events and implicating individuals involved in criminality. Most often, its value can lie in supporting, or refuting, a version of events provided by those deemed to be involved (or by the police, prosecution, or defense). Indeed, forensic science is now considered integral to many criminal investigations, particularly serious offenses, but also "volume" (mostly property) crimes. Forensic science as neutral, objective, evidence is perhaps of most value at the outset of police investigations, where erroneous judgements or false assumptions can send investigators down a blind alley. A piece of scientific evidence gained early in an investigation can ensure that the wrong suspects do not continue to be the center of investigators attentions, thus preventing a wrongful conviction.

Some traditional police "practices," hopefully now becoming obsolete, have not always leant themselves to ensuring that only the guilty are ever convicted. Many suspects in the past have been "persuaded" to confess, and not only the guilty succumbed to such pressure. Informants were also heavily relied upon, and written police statements were not always entirely accurate. However, police practices were not always to blame. Police relied – and still do to a greater extent – on witnesses. These may be eyewitnesses, which are considered highly valuable in securing a conviction. However, as psychologists have demonstrated for many years, people can be highly unreliable at the same time as highly persuasive, especially when they truly believe themselves to be telling the truth. There is a wealth of research on how poor human memory is, how bad we are at identifying people (particularly people we don't know or who are different from us ethnically or in terms of age), how easily we can be persuaded of something or change our recollections to match them with expectations, etc. In short, relying on humans to assist in the detection and prosecution of offenses can be highly problematic, even when those humans are not deliberately lying.

In light of such unreliability, scientific evidence would seem to avoid these "human" problems. There is no need to rely on a person testifying that they saw the defendant at the scene of the crime, if we also have their fingerprints found at the scene. There have been numerous cases where eyewitnesses – victims even – have identified with 100 % certainty their attacker, and yet later DNA evidence has proven that they are mistaken. It would seem far safer to rely upon a DNA match of a rapist than the testimony of the victim (certainly if they don't know their attacker before the crime). Furthermore, the popular (and populist) perception of an almost continuous increase in crime, coupled with a perceived inability on the part of the police to detect criminals or convict them, raises the hope that scientists will be able to tip the justice scales back in favor of the police and prosecution. The growth in forensic science is then based upon the assumption of science as capable of being a neutral, objective, "arbiter of truth" (Dreyfuss and Nelkin 1992: 339). However, uncritical faith can be misplaced, and as scientists themselves warn, "the general public often has an unrealistic expectation of what forensic science can achieve" (Ross 1998: 41).

Yet, forensic science has come to the aid of many wrongfully convicted, most clearly in DNA exonerations, of which there have been over 300 in the USA alone since 1989, with undoubtedly many more to come. The discovery and proliferation of DNA profiling has had a dramatic impact because it has come to be trusted as almost indisputable evidence. The strength of DNA evidence has meant many convictions have been overturned that would have been impossible to overturn otherwise. In some of these cases, DNA has subsequently been used to locate the real perpetrators, sometimes many years after the crime. In just one example, Sean Hodgson spent 27 years in an English prison for a murder that he had confessed to but had not committed. Blood type matching had been used at trial to support Hodgson's confession, but after DNA testing of the evidence 27 years later, it was proven that he was not the killer and freed. There have now been numerous cases around the world where innocent people have called for evidence to be tested or retested in their cases, in the hope that DNA will finally see justice in their case.

It is not hard when reading of famous miscarriages of justice to find instances of flawed scientific evidence being used to convict the innocent, from terrorist bombers found to have nitroglycerine on their hands, which it later transpired could have been a number of harmless chemicals from many household items, to mothers convicted of killing their babies when they had died of unknown, but entirely natural causes. Such cases exemplify the risks of relying upon techniques or "theories" that are erroneous. While experts in these cases were often castigated after the conviction is overturned, many had continued working and testifying before their mistakes were brought to light. The experts were also singled out as isolated "mavericks" in many instances, providing a scapegoat for the wider scientific community. Many other cases however, have relied upon valid scientific principles, but the expert has provided misleading testimony, often overstating the probative value of their evidence or its reliability.

The Innocence Project in New York has of February 2013 exonerated over 302 people wrongly convicted using DNA testing. In 50 % of those cases, the DNA testing has identified the true perpetrator. An early analysis of the first 86 of those exonerations found that faulty forensic evidence had played a role in two-thirds of the convictions. Latest analysis demonstrates that unvalidated or improper forensic science played a role in approximately 50 % of wrongful convictions later overturned (Innocence Project 2009). In such cases, if available evidence had been subjected to DNA testing, then the suspect would have been excluded and most often, another suspect indicated.

The US National Registry of Exoneration, launched in May 2012 by the University of Michigan Law School and the Center for Wrongful Convictions at Northwestern University, is an online database containing a list of exonerations in the United States since 1989. The registry lists more than 1066 wrongfully convicted individuals and is growing. In an analysis of 873 of these cases, 24 % featured false or misleading forensic evidence (Gross and Shaffer 2012). These figures are slightly lower than found by Neufeld and Garrett (2009) in their analysis of 137 trial transcripts of convictions later overturned by DNA. They found experts in 60 % of the cases to have given invalid testimony that overstated the evidence, was unscientific, or contrary to empirical data. They gave instances of erroneous or unsupported testimony about the accuracy and results of forensic techniques including hair comparison, bite-mark comparison, serology, fingerprint comparison, and even DNA testing.

While the work of a forensic expert can take place in any number of environments, particularly when examining crime scenes, for instance, the laboratory is most often the typical workplace of a forensic scientist. However, for forensic evidence to be of utility, it must be able to make the transition from crime scene, via the laboratory, to the courtroom, where it is ultimately used. While much forensic testing will not produce any positive results, or may simply be retained as "intelligence" for possible future use, if test results are not able to be admitted at court as relevant evidence, then the resources used to obtain that evidence will have been wasted. This transition to admissible evidence at trial is then essential if forensic evidence is to be able to play a role in criminal justice.

Experts have been permitted to give evidence in courts for centuries, and their word and expertise have been rarely questioned. Medical men of the middle ages, for example, were often called upon to testify as to their opinion on questions such as cause of death or the sanity of the defendant. Fingerprint experts have, for the last century, been testifying that the prints found at the scene of a crime, for example, “matched” the defendant to the exclusion of all others. However, it has been the massive expansion in forensic science in the last couple of decades, and the increasing incidence of DNA exonerations, that has led to the current paradox being brought into sharp relief. While it has compelling value in preventing and overturning wrongful convictions, forensic science can still also lie at the heart of flawed investigations and trials. In fact, forensic science is introduced to reduce uncertainty and bring objectivity to legal disputes, yet it is increasingly under fire for obfuscation, the introduction of partiality and partisanship, and creating new sites of dispute, that is, for *increasing* uncertainty rather than alleviating it.

One of the perennial predicaments is that while a technique may have a valid scientific basis and prove reliable in its application, how can it be ascertained that it was correctly applied in this case? What are the error rates associated with the technique that may mean that the testing could be wrong in this particular instance? Furthermore, are there any pertinent factors that could jeopardize the reliability or validity (or both) of the testing in the particular instance? For example, the operator (scientist) is incompetent or unqualified, the technique is novel and untested, contamination has occurred, the provenance of the exhibits or results cannot be attested to, and the results have been misinterpreted. As one can quickly see, the use of forensic science can be a complex and complicating factor in any legal dispute.

As the omnipotent umpire, the judge has the task of ensuring the legality and fairness of a trial, and for as long as experts have been allowed to give evidence, there have been rules regarding the expert’s remit and the special allowances afforded them. In light of the powerful influence of science at trial, efforts have been made to bolster the role of the judge in ensuring that science is only used at trial when relevant and reliable and is represented impartially and accurately for the assistance of the fact finders. This has led to the common representation of judges as “gatekeepers,” guarding the doors of the courts to ensure that invalid or unreliable science cannot enter. This could be argued to be the most obvious way to keep “junk” or bad science from admission at trial, but there are complicating factors. These are not limited to the fact that judges themselves are fallible, not always as impartial as one might hope, and rarely have a scientific grounding upon which to rely when ruling. The ability of judges to interpret and apply “correctly” (assuming there is a “correct” decision) complex exclusionary rules is further complicated by the difficulty of the task itself and the clarity, or otherwise, of the rules.

In a survey of 400 judges on their understanding and acceptance of their role as “gatekeeper,” Gatowski et al. (2001: 443) found that 91 % believed the role was appropriate. Those doubting believed that judges had insufficient scientific training, making the role “‘difficult, untenable, or inappropriate’” (Gatowski et al. 2001: 444). Indeed, the researchers found judges equally divided on whether their education adequately prepared them to deal with scientific evidence, many stating that the extent to which judges could properly apply criteria to judge scientific evidence was “questionable at best” (Gatowski et al. 2001: 451–452). The authors concluded that the lack of scientific literacy among trial judiciary, and the increasingly complex nature of the science coming before courts, demonstrates a need for more science-based judicial education (Gatowski et al. 2001: 455). While a laudable aim, the exclusionary rules themselves have been widely criticized (there is a substantial literature but, e.g., see Edmond (2000); Risinger (2000–2001); Danaher (2011)). These differ between jurisdictions with the USA, Canada, Australia, and the English courts all making attempts to improve the exclusionary rules applying to expert evidence, some meeting with more success than others. Some of the important cases are discussed below.

International Perspectives

The USA, perhaps unduly, has earned a reputation as both a highly litigious nation and one that has most utilized "junk" science to ensure a favorable legal outcome for clients, in particular in multimillion dollar "toxic tort" civil cases. Indeed, in the decades leading up to one particular toxic tort case, that of *Daubert v Merrell Dow Pharmaceuticals* 509 U.S. 579 (1993), there was a great deal of controversy over the perceived "flood of 'junk' science that, according to some popular critics, threatened to inundate the courts" (Beecher-Monas 1998: 58). The so-called *Daubert* case was the progenitor of two further cases which have come to be known as the "*Daubert* trilogy" (the other two being *General Electric Co. v Joiner, 522 U.S. 136, 118 S Ct. 512* (*1997*) and *Kumho Tire Co. Lts v Carmichael, 119 S Ct. 1167* (*1999*)). Prior to these, the standard for evaluating expert testimony was the "Frye" standard, from *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923). At the core of Frye was the contention that judges should refer to scientists, admitting evidence when the method utilized was "sufficiently established to have gained general acceptance in the particular field in which it belongs"(1014). This came to be known as the "general acceptance" rule and was reinforced by the Federal Rules of Evidence adopted in 1976, of which Rule 702 stated that "if scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise."

Many states continued to utilize the Frye standard, but the "general acceptance" rule gave rise to problems and two primary concerns: (1) it required the establishment of an "orthodoxy," but such an orthodoxy may be indulgent, in that it is based upon a body of work that has no scientific validity or has any checks or objective standards, and (2) waiting for "general acceptance" may stifle innovation or deprive courts of novel techniques or scientific breakthroughs. "Acceptance" within the scientific community can be misleading, in that some methods or theories are "accepted" almost by faith, and are not proven flawed for many years, perhaps refuted by people on the periphery who lack "acceptance" by the wider scientific community. The Frye standard and Federal Rules were then proving inadequate. A more stringent standard was therefore outlined in *Daubert*, where general acceptance was just one of five criteria by which to "test" scientific evidence:

1. Whether the theory or technique in question can be (and has been) tested
2. Whether it has been subjected to peer review and publication
3. Its known or potential error rate
4. The existence and maintenance of standards controlling its operation
5. Whether it has attracted widespread acceptance within a relevant scientific community

Most states have adopted the *Daubert* standard, some still work with the *Frye* test, and a few have their own tests they adopt on a case-by-case basis. *Daubert* explicitly placed judges in the role of gatekeeper who must evaluate the scientific validity and reliability of scientific evidence. It also switched the focus from whether there was a scientific consensus upon which to base the evidence given, to whether the techniques and methodology used were valid. The *Kumho Tire* case clarified that the *Daubert* analysis applies to scientific, technical, and otherwise specialized knowledge, and not exclusively to scientific evidence (so applying to engineers, etc.) (at 1175). *Kumho* also strengthened trial judge's discretion in noting that the judge had "considerable leeway" and "broad latitude" in flexibly applying the criteria set out in *Daubert* (at 1176).

Daubert has provided a useful checklist but is by no means uncontroversial. In his judgement in *General Electric Co. v Joiner*, 522 U.S. 136 (1997), Justice Stephen Breyer alluded to the difficulties that judges faced: "scientific evidence often requires judges to make subtle but sophisticated determinations," (p. 1) and "those duties often must be exercised with special care" (p. 2). Scholars have produced many theoretical, quantitative, and qualitative analyses of the operationalization of *Daubert*, debating the relative merits of each of the criteria; considering how judges are deciding on admission, the extent to which judges are able to apply the criteria; as well as assessing various forensic methods using the *Daubert* criteria. Many have found the *Daubert* criteria wanting, as well as doubting the ability, or willingness of judges to apply them strictly. Moriarty and Saks (2005: 29) concluded that "some forensic sciences have been with us for so long, and judges have developed such faith in them, that they are admitted even if they fail to meet minimum standards under *Daubert*. Faith, not science, has informed this gatekeeping." In their major report "Strengthening Forensic Science: The Path Forward," the National Academy of Sciences was pessimistic about the contribution that could be made by judges and their gatekeeper role in preventing "junk science," concluding that "*Daubert* has done little to improve the use of forensic science evidence in criminal cases" (NRC 2009: 106). Their research found that US appellate courts were too deferential to admissibility decisions made by trial judges and were simultaneously being too generous in admitting prosecution expert evidence while generally excluding expert evidence for the defense (NRC 2009: 96).

Both Canadian and Australian State and federal legal jurisdictions have seen expert evidence and forensic science playing a role in wrongful convictions. In both countries, public inquiries and Royal Commissions following exonerations have driven forensic science policy and reform. In Canada, the case of *Mohan* [*1994*] *2 SCR 9* established a four-part test for expert evidence, requiring that it be (1) relevant, (2) necessary in assisting the trier of fact, (3) not otherwise subject to an exclusionary rule, and (4) given by a properly qualified expert. The subsequent Canadian Supreme Court cases of *R v J-LJ* [2000] 2 SCR 600 & *Trochym* [2007] 1 SCR 239 affirmed that there are enhanced tests of reliability for expert evidence and courts must particularly scrutinize novel science or methodologies, similar to the approach in *Daubert*.

The 2008 Inquiry into Pediatric Forensic Pathology in Ontario (the Goudge Report) inquired into a series of wrongful convictions relating to the pathologist Charles Smith. The report affirmed that judges play "an important role in protecting the legal system from the effects of flawed scientific evidence" and that "judges bear the heavy burden of being the ultimate gatekeeper in protecting the system from unreliable expert evidence" (Goudge Report 2012: 470). Goudge asserts that "trial judges should be vigilant in exercising their gatekeeping role" and that a test of reliability is embedded within the *Mohan* test. Evidence must thus be excluded by judges if not satisfying standards of threshold reliability, whether or not the science is novel (Recommendation No. 130). The report stressed that judges needed to pay close attention to the methodological and reliability issues identified in *Daubert* (pp. 483–484). Enhanced judicial education was also recommended to enable judges to undertake their gatekeeping role competently.

Similarly, Australia has suffered wrongful convictions that have led to Royal Commission reports highly critical of scientific evidence. In 1987, Judge Morling released a report into the Azaria Chamberlain conviction that dismissed the majority of the prosecution's scientific evidence and saw the conviction overturned. Previously, in 1984, the Shannon Royal Commission into the conviction of Edward Charles Splatt had implicated flawed scientific evidence which led to his release and pardoning. Such Royal Commissions have seen improvements in forensic science provision across Australia. However, legal reforms have been more fitful and uneven across the continent.

Australian courts follow a variety of state, federal, and common law, with many adhering to uniform evidence legislation which admits opinion evidence under s. 79, where the opinion must be (1) relevant; (2) from a person who has specialized knowledge; (3) the specialized knowledge is based upon the person's training, study, or experience; and (4) the opinion is wholly or substantially based on that specialized knowledge. However, the legislation does not define "specialized knowledge" or require certain criteria for the "field of expertise" to be met. This has given rise to debate about the role of reliability. In *HG v The Queen* (1999) 197 CLR 414, the common law is cited as requiring an expert's knowledge or experience to be in an area "sufficiently organized or recognized to be accepted as a reliable body of knowledge or experience" (at 58). However, the New South Wales Court of Criminal Appeal has said that evidentiary reliability is not a consideration under s79 (*R v Tang* (2006) 161 A Crim R 377; [2006] NSWCCA 167). The focus of attention must be on "specialized knowledge," not on the introduction of "an extraneous idea such as reliability" (at 137). For a country that has seen wrongful convictions based upon unreliable expert evidence, this is a worrying judgement indeed.

Presently the English and Wales legal system operates a "case-by-case" assessment of experts (their evidence tested via cross-examination), which has on occasion proved flawed. Expert opinion evidence is admissible under *R v Turner* [1975] QB 834, where "an expert's opinion is admissible to furnish the court with scientific information which is likely to be outside the experience and knowledge of a judge or jury...." In recent years, judges have approved and adopted the South Australian case of *Bonython* [1984] 38 SASR 45, which requires consideration of the subject matter of the expert's opinion, considering whether it "forms part of a body of knowledge or experience which is sufficiently organized or recognized to be accepted as a reliable body of knowledge or experience"(at 47). This still stops short of an explicit or stringent gatekeeping role for English and Welsh judges, an omission that the Law Commission of England and Wales in their 2011 report on expert evidence is hoping to fill. Their Draft Bill seeks to introduce a statutory reliability test for expert evidence, requiring the party wishing to rely on the evidence to demonstrate that it is sufficiently reliable to be admitted. It is intended that this will be an enhanced test of admissibility, with a suggested list of criteria for judges to consider.

Future Directions

Given the difficulties highlighted with the gatekeeping role of the judge and the vagaries of expert evidence admissibility at trial, it may be more prudent to ensure the reliability and validity of scientific evidence *prior* to admission at trial. Indeed, given that most wrongful convictions have at their core defective investigative decision making, it is vital that forensic evidence utilized by investigators is reliable to ensure decisions made at this stage are based upon sound evidence. If forensic evidence is flawed at the investigative stage, it is often too late, and the damage irreversible at trial. Given then that it is essential that all forensic evidence is reliable and valid, whether used at trial, during an investigation, or held as "intelligence" by law enforcement agencies, there must be systems in place to ensure the quality of forensic evidence from the outset of the criminal process. This requires regulation and oversight of forensic science from the crime scene to the courtroom and quality assurance standards for the education, training, and operation of forensic scientists and the quality assurance and accreditation of their working environments and practices.

During the massive expansion of forensic science provision in the late twentieth and early twenty-first centuries, there have been a series of reports commenting upon forensic services. In England and Wales, the Royal Commission on Criminal Justice of 1993 (The Runciman Report) made 13 recommendations specific to forensic science. Of these, the establishment of an oversight body was deemed a priority. A subsequent report into serious contamination at a military forensic explosives laboratory by Professor Caddy in 1996 recommended the creation of an "Inspectorate of Forensic Sciences" and advocated the registration of individuals as forensic practitioners, a call repeated by the House of Commons Science and Technology Committee (2005), when it proposed greater regulation of forensic science. In 1999 the establishment of the Council for the Registration of Forensic Practitioners (CRFP) sought to register "competent" forensic practitioners. However, the CRFP stopped far short of bringing rigorous scrutiny to bear upon forensic science and in 2009 was closed in the light of financial difficulties, lack of stakeholder support, and the newly created Forensic Regulator role.

The role of the UK Forensic Regulator was created in 2007 and was tasked with establishing and monitoring quality standards and oversees accreditation via the UK Accreditation Service (UKAS) using the international laboratory testing ISO17025 standard, necessitating UKAS establish supplementary standards and modifications to tailor the standard to forensic science. However, there remain questions over whether regulation reforms are being applied equally to all aspects of forensic science. On the one hand, the introduction of the regulator was presented as creating a generic standard for forensic science providers in the UK and “a light touch” in steering service providers. However, there remain concerns about a perceived lack of teeth and gaps in regulation, with a fear that accreditation may prove to be superficial. The regulator also faces serious resource restrictions.

There is increased recognition in the USA for the need for proper regulation and oversight of forensic science. The New York State Commission on Forensic Science which accredits and monitors forensic laboratories was established in 1994 and in Texas in 2005; the legislature established an independent oversight body for forensic laboratories to identify and oversee rectification of problems that have blighted forensic science in that state. Despite such innovations, in February 2009, the National Academy of Sciences report, *Strengthening Forensic Science in the United States: A Path Forward*, is almost unremittingly condemnatory. This critical tone was not unexpected, given that the committee had been established in the wake of high-profile failings and against a backdrop of the ongoing exoneration of innocent people, convicted often with the aid of flawed or misrepresented forensic science. In addition, the USA was still struggling to make inroads into significant backlogs in forensic laboratories despite increases in federal funding. The report unsurprisingly found serious issues facing forensic science and concluded that any remedy would have to be national in scope and demands both leadership and funding.

In July 2012 legislation was introduced to the US Congress to help prevent wrongful convictions by bringing reliable, science-based standards to forensic evidence. The Forensic Science and Standards Act of 2012 seeks to strengthen forensic science and standards: “yielding evidence that judges, prosecutors, defendants, and juries can fully trust.” The Forensic Science and Standards Act of 2012 would require the National Institute of Standards and Technology (NIST) to develop forensic science standards while a Forensic Science Advisory Committee ensure the implementation of standards. Further, a National Forensic Science Coordinating Office, housed at the National Science Foundation (NSF), would also develop a strategy to support a forensic science grant program to promote research.

With international forensic data exchange increasingly common, quality standards for forensic science have become ever more important. The requirement that forensic science providers have demonstrable quality standards and accreditation has now been mandated with EU Council Framework Decision 15905/09. Accreditation of laboratories to the ISO17025 standard was viewed as providing “mutual trust in the validity of the basic analytic methods used,” although it does not mandate particular methods to be used, only that the method be suitable for its purpose. This decision only covers laboratories, omitting any quality oversight of the retrieval of forensic information, whether at crime scenes or in police stations, for example, its scope restricted to the results of laboratory activities.

It is unlikely that any standard can regulate every aspect of a forensic practitioner’s work. The lack of oversight of crime scene examination and evidence retrieval will be very difficult to overcome, particularly where police personnel are working in their domains without external supervision or oversight. Well over half of forensic science services (measured by cash value) in England and Wales are delivered within police forces’ own scenes of crime operations and scientific support services, with this set to increase. These services are not yet subject to the same quality standards regimes as apply to commercial providers. Yet differential standards operate against the public interest, increasing the risk of flawed results being relied upon, or challenged in the courts. In many countries, all forensic science services come under the auspices of the police, with their accreditation status and quality assurance regimes unknown. The use of personnel directly employed by the police has been roundly criticized by all reports looking into forensic science. Indeed, high-profile wrongful convictions in England and Wales were tainted by the suspicion that scientists had been too easily influenced by the police when undertaking testing and reporting results. The US National Research Council (2009) report specifically recommends that all forensic laboratories be removed from law enforcement premises and/or their administrative authority.

Despite the overwhelming support for a stronger and more regulated range of forensic services, difficulties persist. There remains the ever-present problem of funding: to oversee all forensic science provision to a standard which some would still view as inadequate has required considerable investment. It is also dangerous to implement quality standards and operating procedures that are not underpinned by rigorous scientific research, yet it is this lack of underlying "science" that has been most strongly criticized. Significant investment in research is still required prior to regulation. Further, the proper extent of the reforms in the forensic area, as opposed to their direction, remains largely unconsidered. For example, it might be asked why fingerprints are currently collected and analyzed by the police without reference in the vast majority of cases to scientists or independent laboratories. The answer seems in the main to be historical; fingerprinting was developed by the police themselves two or three decades before the establishment of formal laboratories. Thus, forensic science should be defined in wide terms and regulators given a wide remit if the quality of justice is to be improved.

Another problem concerns how forensic evidence is handled and the particular danger that tests adverse to the contentions of the prosecution will be disregarded and suppressed. This leads to the observation that acceptable forensic detection ultimately depends not only on the imposition by society of training, rules, sanctions, and supervision but also on the internalization by scientists of the ethics behind that training. As has been found to be indispensable in relation to police interrogation practices, it is necessary to impact upon the whole culture with which the police approach investigations. A corresponding approach to forensic evidence should spark further debate as to qualifications and training.

Despite almost comprehensive favorable media representation, forensic science has not eradicated the potential for factual errors in legal investigations and criminal trials resulting in wrongful convictions. While the realization of the extent of human unreliability has grown at the same time as scientific and technological power and knowledge have increased, undoubtedly preventing many wrongful convictions, they may still flourish in a culture which fails to properly scrutinize and question forensic evidence. Research into the causes of wrongful convictions clearly demonstrates that while the utilization of forensic evidence can assist the pursuit of justice, it can also seriously hamper fact finders and triers of fact in criminal cases.

One attempt to prevent wrongful convictions has been in the development of evidentiary rules to prevent flawed scientific evidence from entering the courtroom, empowering judges to rule questioned scientific evidence inadmissible. However, such admissibility standards have also been criticized, with commentators concluding that there remains a "conspicuous need for further refinement and greater vigilance to make these standards effective" (Edmond and Roach 2011).

While forensic scientists have joined calls over the years for better regulation, to provide assurance to the courts that experts before them will be qualified, and their evidence valid and reliable, there remain serious limitations to current regulation that means that the gatekeeping role of judges in courtrooms remains as vital as ever. The task cannot be left to judges alone however. Without "good" forensic science, authorities run the risk not only of wrongful convictions but also a loss of public confidence in the criminal justice system. It is critical that attention be paid to the delivery of forensic services: how scientists are trained and standards are set, monitored, and maintained across the forensic science sector. If this were to be done to a high standard, then the burden upon judges in their role as gatekeeper and the opportunities for error would be lessened.

Related Entries

- ▶ Behavioral Science Evidence in Criminal Trials
- ▶ Causes of Wrongful Convictions
- ▶ Comparative Legal Cultures
- ▶ Critical Report on Forensic Science
- ▶ Evolving Judicial Roles
- ▶ Expert Evidence and Criminal Trial Procedure
- ▶ Expert Witnesses: Role, Ethics, and Accountability
- ▶ Eyewitness Identification Evidence
- ▶ False Confessions and Police Interrogation
- ▶ Fingerprint Identification
- ▶ Forensic Science and Criminal Investigation
- ▶ Forensic Science and Miscarriages of Justice
- ▶ Forensic Science and the Paradigm of Quality
- ▶ Innocence Commissions
- ▶ Overturning Wrongful Convictions and Compensating Exonerees
- ▶ Philosophical Basis of the Forensic Process
- ▶ Philosophy of Forensic Identification
- ▶ Probability and Inference in Forensic Science
- ▶ Prosecution and Wrongful Convictions
- ▶ Scientific Evidence Before International Criminal Tribunals
- ▶ Scientific Evidence in Criminal Adjudication
- ▶ Scientific Evidence in Criminal Prosecutions
- ▶ Transnational Exchange of Forensic Evidence
- ▶ Wrongful Convictions in the Inquisitorial System

Recommended Reading and References

Beecher-Monas E (1998) Blinded by science: how judges avoid the science in scientific evidence. Temple Law Rev 71:55–102

Danaher J (2011) Blind expertise and the problem of scientific evidence. Int J Evid Proof 15:207–231

Dreyfuss R, Nelkin D (1992) The jurisprudence of genetics. Vanderbilt Law Review 45(2):313–348

Dwyer D (2008) The judicial assessment of expert evidence. Cambridge University Press, Cambridge

Edmond G (2000) Judicial representations of scientific evidence. Modern Law Rev 63:216

Edmond G, Roach K (2011) A contextual approach to the admissibility of the state's forensic science and medical evidence. University of Toronto Law Journal 61:343

Gatowski SO, Dobbin SA, Richardson JT, Ginsburg GP, Merlino ML, Dahir V (2001) Asking the gatekeepers: a national survey of judges on judging expert evidence in a post-Daubert world. Law Hum Behav 25(5):433–458

Goudge Report (2012) Inquiry into pediatric forensic pathology in Ontario. Available at: www.goudgeinquiry.ca. Accessed 10 August 2012

Gross S, Shaffer M (2012) Exonerations in the United States, 1989–2012. First report by the National Registry of Exonerations. http://globalwrong.files.wordpress.com/2012/05/exonerations_us_1989_2012_full_report.pdf

House of Commons, Science and Technology Committee (2005) Forensic science on trial. HMSO, London, HC 96-1

Innocence Project (2009) Wrongful convictions involving unvalidated or improper forensic science that were later overturned through DNA testing. http://www.innocenceproject.org/docs/DNA_Exonerations_Forensic_Science.pdf

Law Commission (2011) Report on expert evidence in criminal proceedings (Law Comm. No. 325)

Moriarty JC, Saks MJ (2005) Forensic science: grand goals, tragic flaws, and judicial gatekeeping. Judges J 44:16–33

National Research Council (2009) Strengthening forensic science in the United States: a path forward, committee on identifying the needs of the forensic, sciences community. http://www.nap.edu/catalog/12589.html

Neufeld P, Garrett B (2009) Invalid forensic science testimony and wrongful convictions. VA Law Rev 29(1):1–97

O'Brien WE (2003) Court scrutiny of expert evidence: recent decisions highlight the tensions. Int J Evid Proof 7:172–184

Odgers SJ, Richardson JT (1995) Keeping Bad science out of the courtroom – changes in American and Australian expert evidence law. UNSW Law J 18:108

Redmayne M (2001) Expert evidence and criminal justice. Oxford University Press, Oxford

Risinger M (2000–2001) Navigating expert reliability: are criminal standards of certainty being left on the dock? Albany Law Rev 64:99–152

Roach K (2009) Forensic science and wrongful convictions: some lessons from comparative experience. Jurimetrics J 50:67–92

Ross A (1998) Controversy corner: the quest for truth. Aust J Forensic Sci 30(1):41–44

Legal Status of Abortion

Gilbert Geis
School of Social Ecology, University of
California, Irvine, Irvine, CA, USA

Overview

The moral issues in regard to the proper legal status of abortion do not involve a dispute over fundamental demonstrable facts but rather concern a myriad of considerations that lie beyond empirical verification or dismissal. It is this ambiguity that allows the dispute to be so contentious and continuous. There also is considerable argument but few facts concerning the consequences of the triumph of either the pro-choice position favoring legal abortions or the pro-life position that advocates that there be no government-accorded right to an abortion. Some pro-lifers, although by no means all, qualify this last view by granting the right to an abortion when a conception is the result of rape or incest or if the abortion seems necessary to save the life of the pregnant woman. When the "health" of the women becomes a proabortion criterion – especially her mental health – pro-life forces tend to object on the ground that pro-choice proponents define "health" so loosely that it would allow abortions almost at will. The incest exception also can be controversial. Most incest likely occurs between teenage and somewhat older brothers and sisters: should abortions be allowed in these instances if they are otherwise banned?

Key Issues

Some people note that permissible abortions detract from the population growth, although there is no agreement whether this is a good or an undesirable consequence. There also are those who insist that if legal abortions were not available, nearly as many women who now have them would choose illegal ways to get rid of the embryo growing inside them. Disputes can become highly emotional with the pro-choice advocates publishing frightening scenarios depicting poor and desperate women resorting to back-alley quacks or to dangerous folk remedies to deal with what they regard as an intolerable situation. Botched amateur abortion procedures can prove fatal; the pro-choice forces point out that the death rate from legal abortions, which increases with the length of the pregnancy, is about 0.7 per 100,000 operations (Bartlett et al. 2004). Depending on the source you rely upon, the death rate for legal abortions may or may not be higher than that in which pregnancies are carried to term. The pro-choice forces, for their part, often employ grisly pictures of aborted matter, particularly in regard to late-stage abortions when the images bear obvious humanlike features.

There is among the pro-life advocates a belief that permissible abortions are evidence of the trend toward more degeneracy in the United States, part of a burgeoning spirit that favors self-interest over social and moral imperatives. Others believe that permitting legal abortions demonstrates the freedom that should characterize a democratic society.

In terms of empirical evidence, some scholars have reported that the number of what they label "marginal children" is decreased by the availability of legal abortion. They maintain that cohorts of children born after legalized abortion have experienced a significant reduction in the number of adverse outcomes compared to the average child and that the "marginal child [who was aborted] would have been 40–60 % more likely to live in a single-parent family, to live in poverty, to receive welfare, and to die as an infant" (Gruber et al. 1999, p. 269). Others maintain that abortion has contributed to a decline in the rate of juvenile delinquency and crime (Donahuse and Levitt 2001). That finding, if accurate – it is disputed by Joyce (2003) – would have to be considered with a research conclusion that having a baby significantly decreases the likelihood of delinquency among adolescent girls (Hope et al. 2003). Pro-lifers generally find such studies unpersuasive, declaring that alternative explanations are just as plausible for the reported results, and, besides, the outcomes, even if accurate, do not make what they consider to be fetal homicide either justifiable or excusable.

For the pro-life group, a fundamental moral consideration is that for them at the moment of conception the embryo becomes a human being and has the right to have his or her life sustained. Some regard this as a sacred right and rely on biblical sources to make their case; others see the issue as one of secular morality. Carried to its logical extreme, this position regards abortion as murder, though only extremely rarely do pro-life people maintain that women who have abortions ought to be tried in criminal courts for homicide.

For the pro-choice camp, the decision whether to carry the fertilized embryo to term should be the woman's because her body is deemed to be her personal domain. They maintain that the state has no right to demand that a woman give birth to what is now in her womb. Those adopting this position have to grant that the state often intrudes to dictate what they must do with their body, such as put it inside a classroom in their early years.

Some in both the pro-life and pro-choice camps believe that the father of the child should have a voice in the decision about whether or not an abortion ought to be elected. If the couple is not married, nor a pair, advocates of this position typically agree that the responsibility to raise the child should be totally on the man who objects to an abortion. The subtitle of a law review article illustrates how some women feel about any involvement of a putative father-to-be: "A Women's Womb is not a Man's Castle" the subtitle reads; the text of the article notes that in law the unwed father has no rights and to afford him any would make a woman's right to an abortion "virtually meaningless" (Preshiran 1990, p. 1365).

Roe v. Wade

The clash between opposing views about abortion began most dramatically after a Supreme Court opinion – *Roe v. Wade*, handed down in 1973 – decreed that under the umbrella of the judicially decreed doctrine of privacy abortions must be legally available to women in the first trimester of pregnancy. During the decades, since the *Roe* decision, pro-choice advocates have whittled away at the Supreme Court opinion, hedging it with requirements that make access to abortions more burdensome.

A major pro-life group goal is to have the Supreme Court, if it is reluctant to overturn *Roe v. Wade*, to alter its ruling by declaring that abortion is not a federal matter but that individual states should be allowed to adopt whatever rules they desire in regard to abortion. Justice Antonin Scalia, who strongly supports this view, has declared that the Court should repudiate *Roe*, which he sees as the Court's "self-awarded sovereignty over a field in which it has little proper business" (*Webster v. Reproductive Health Services* 1989, p. 532). A result of state control of abortions obviously would be that persons living in a state such as Utah who desired an abortion would have to travel to a state such as California to undergo the procedure. Obviously, the less affluent would be most affected if the states were to determine their position on abortion.

The pro-choice advocates, since they continue to retain the right to a legal abortion, have largely been relegated to a defensive position, trying to erect barricades against the incremental advance of the agenda of the pro-life forces who have benefited by having a Supreme Court with a conservative majority that recently had the very unlikely arrangement of six Roman Catholic justices (all but one a conservative) and three Jews (all liberals) sitting on the nine-person bench.

South Dakota and Abortion

In 2011, as illustrative of the whittling away process, the South Dakota legislature, with a three-to-one majority of Republicans, enacted the first statute in the nation that requires women seeking abortions to undergo a consultation at a "pregnancy help center" where they are to be informed what assistance is available "to help the mother keep and care for her child." The statute also established the nation's longest waiting period – 3 days – after an initial visit with an abortion provider before the procedure can be done. About half of the states in America have a 24-h waiting period – none, until South Dakota's action, any longer than that. It is presumed that the waiting period, the extra trip to the sites for counseling, and the cost of overnight stays will deter some South Dakota women from seeking abortions. Since no doctors in South Dakota will perform abortions, women in the state who desire abortions are dependent upon physicians who once a week fly into South Dakota from Minnesota.

When signing the legislation into law, South Dakota's governor wrote that he hoped that women considering abortion will use the 3-day waiting period to become "as fully aware as possible of the implications of the grave decision to terminate the most sacred gift of life." The governor praised the law as promoting "voluntary, uncoerced, and informed" decision-making in regard to abortion. He presumably had in mind husbands, family, friends, or others who might encourage an abortion when he employed the word "uncoerced," but opponents found the usage ironic in view of the fact that the required counseling would be by individuals without accreditation who are personally committed to the pro-choice ideology. One of the counseling sites, the Alpha Center in Sioux City, for example, had declared that abortions increase the risk of breast cancer, infertility, and depression, a conclusion decried by the most reputable scientific studies (see, e.g., Varmos 2003).

A Democrat member of the South Dakota legislature expressed the basis for her opposition to the new law, stating that "South Dakota women should not need any in-person lecture from an unqualified, uncertified faith-based counselor or volunteer at an anti-choice clinic" (Sulzberger 2011).

In 2011, the eighth circuit of the federal appellate court upheld the constitutionality of the major elements of the South Dakota law, including the fact that the person seeking an abortion must be told that what she was contemplating would "terminate the life of a whole. separate, unique human being." At the same time, the court ruled against the requirement that the woman should be informed that her behavior increased the risk of "suicide ideation" and suicide itself. The court maintained that there was no satisfactory evidence to support this conclusion. Critics would argue that the same was true of the "unique human being" message. In 2012, the legislature mandated that abortion counselors must be certified but rejected a proposal that the doctor and the proposed patient could conduct their business on the telephone, thereby avoiding the possible expenses of an extra trip to the abortion site.

The Personhood Movement

The boldest drive by the pro-life forces has involved an effort by a group called Personhood USA to sponsor referendums that seek to alter state constitutions so that a fetus at the moment of fertilization is deemed to be a human being and its destruction therefore becomes a grave criminal offense. The proposed amendment makes no exception for fertilizations that were the consequence of rape or incest or that were found to be likely to produce seriously deformed children. An egg fertilized in a test tube was also to be regarded as a person.

A Personhood initiative in November 2011 appeared on the ballot in Mississippi, arguably the most conservative state in the nation. Proponents thought its passage would be a slam dunk and were stung when it went down to defeat by a 58–42 % vote. Many persons who opposed legal abortion voted against the measure because they believed it was too broad and overly ambiguous. Some pro-lifers also thought that the measure was certain to be struck down by the courts and thus would set their movement back in the minds of the public. There also was voter concern that the measure would have implications for the legality of some birth control procedures such as the use of intrauterine devices that allow egg fertilization but prevent the embryo from attaching to the uterine wall. Supporters of the initiative maintained that many of the views of opponents were scare tactics and not realistic assessments of the proposed measure (Markoe 2001, December 13).

The Mississippi legislature responded to the Personhood proposal's defeat by placing particularly cumbersome restrictions on which doctors could do abortions and where they must be granted permission to do so. A temporary injunction against the law was granted in mid-2012 (*Jackson Women's Health Organization v. Currier* 2012), but the case is likely to be under legal review for a considerable time. About 2,200 abortions were performed in 2010 in Jackson, the site of the only clinic in Mississippi offering the service.

Religion and Abortion

Catholicism

The hierarchy and communicants of the Roman Catholic Church constitute major supporters of the pro-life movement. Catholics, the country's largest religious group, make up 24 % of the population of the United States. Clearly, not all Catholics follow every church doctrine literally, and a large number of individuals with other religious affiliations and many with no such allegiances are part of the pro-life camp.

The historical record shows that in earlier times the core issue for Roman Catholic theologians regarding abortion was the question of "ensoulment," that is, when the soul unites with the body to produce an actual human being. For some time, Catholic authorities held that abortion during the initial stages of pregnancy did not constitute a religious sin because the soul had not yet entered the fetus. St. Augustine in his fourth-century writings declared that the amalgamation of body and soul did not take place until the time of fetal quickening, which usually occurs during the fifth month of pregnancy with the development of the spinal cord. Pope Innocent III, the church leader from 1198 to 1216, set the dividing line for when abortion became a mortal sin at 40 days after conception for a male fetus and 80 days for a female fetus, numbers based on the time when it became possible to determine genital development in spontaneously aborted fetuses. In practice, since the gender of the gestating fetus was unknowable at the time, 80 days became the sanctioned time for a legal abortion (Asma 1994). The problem the church faced was that in regard to ensoulment, the task of sustaining its legitimacy was made significantly more difficult by virtue of the fact that religious conceptions of the soul are often hybrids of fundamentally inconsistent notions.

The current position of the Catholic Church regarding abortion was forcefully enunciated by Pope Benedict XVI in 2010:

> From the moment of its conception life must be guarded with the greatest care. With regard to the embryo in the mother's womb science itself highlights its autonomy, its capacity for interaction with the mother, the coordination of biological processes, the continuity of development, the growing complexity of the organism.
>
> It is not an accumulation of biological material but rather of a new living being, dynamic and marvelously ordered, a new individual of the human species. This is what Jesus was in Mary's womb; this is what we all were in our mother's womb. We may say with Tertullian, an ancient Christian writer, "the one who will be a man is one already"[Bindley 1890, ch. 9], there is no reason not to consider him a person from conception. (Benedict 2010)

The Pope's statement represents a departure from the Church's earliest emphasis on allowing abortions until ensoulment. That criterion, obviously lying beyond any possibility of demonstration, was open to being scorned by nonbelievers as no more than folklore. To declare that humanity begins at birth is a much less vulnerable position. The best that skeptics can say is "Maybe so, maybe not." Pro-choice advocates might add that they resent having the church's arguable interpretation determine their personal behavior in regard to whether or not to undergo an abortion.

Abortion in Brazil

The United States has the fourth largest number of Roman Catholics of any country in the world. As the largest religious group in the United States, the Church is able to influence to some extent the alignment between its doctrines and secular law. In Brazil, which has the largest number of Catholics of any country, the Church, home to 75 % of the country's population, inevitably exerts a great deal more power over daily affairs than it does in the United States.

Brazilian law permits abortion in cases of rape and in circumstances in which giving birth could cause the death of the pregnant women. To perform an abortion on oneself or another person carries the possibility of a 3-year prison term.

The status of abortion was vaulted into the limelight in Brazil in 2009 when it became known that a 9-year-old girl had undergone a medical abortion in the city of Recife after she was made pregnant with twins by her 23-year-old stepfather. The doctors claimed that the girl's 80-lb body was not mature enough to give birth safely. Church authorities retorted that the doctors could have resorted to a Caesarian section to bring the twins into the world.

The rift between church and state came into play when the Church excommunicated the girl's mother (but declared there was no theological doctrine to allow it to do so in regard to the stepfather) as well as all members of the medical team that had participated in the abortion. The President of Brazil deplored this action. The Archbishop replied that the President ought to study up on his theology. "We know people have other ideas," the Archbishop declared, "but if they do they are not Catholics. We want people who adhere to God's word." The Archbishop later added: "Abortion is much more serous than killing an adult. An adult may or may not be innocent, but an unborn child is most definitely innocent. Taking that life cannot be ignored" (Downie 2009).

The statistics concerning aspects of the abortion phenomenon in Brazil portray a situation difficult to interpret. There were 3,093 legal abortions performed in the country from January to November of 2008. The Ministry of Health indicates that 200,000 women, almost all of them Catholics, appear at hospitals or medical clinics each year for treatment for errant illegal abortions. It is estimated that there are about 14 million clandestine abortions performed in Brazil annually. Yet, 67 % of the population responding in a national poll indicated that their preference is to leave the law as it is (Downie 2009).

Abortion and Other Religions

Protestants

Protestant religions run a gamut in regard to abortion, although the most prominent tendency is to echo the Catholic Church's position but less categorically, lest the ministers offend too many communicants. A policy statement by the Evangelical Lutheran Church in America, for instance, reads as follows:

> …the number of induced abortions is of deep concern to this church. We mourn the loss of life that God has created. Abortion ought to be an option only of last resort. Therefore, as a church we seek to reduce the need to turn to abortion as the answer to unintended pregnancies. (Baker and Ehlke 2011, p. 122)

The "ought" in statement is somewhat equivocal, and the words "we seek" do not suggest as vigorous a condemnation of abortion as that promulgated by Catholic theologians.

Mormons

The Church of Latter-Day Saints (Mormons) allows abortions under limited circumstances (rape, for instance) but advises that those contemplating the procedure first consult with the local presiding church authority and then go ahead only after receiving divine conformation of their decision through prayer (Hunter 1990).

Judaism

We can examine the theological position of Jews as an interesting example of an approach that, depending on which segment of the religion a communicant adheres to, comes close to or differs notably from Catholic doctrine.

In terms of theology, Jews divide into three major subgroups: Orthodox, Conservative, and Reform. The issue of Jews and abortion is more complex than with Catholics because there is no one person, such as the Pope, who can put forward a position that would represent the official stance for all the observant.

Orthodox Jews oppose abortion except under very limited circumstances, while some conservative and reform rabbis take the position that a fetus is not a person until some segment of the birthing baby moves outside the mother's body. Theology aside, Jewish authorities find themselves faced with a survival problem that tilts them away from endorsing abortion. In the United States, about half of the Jews marry outside their religion, and their children may or may not be raised in the Jewish faith. Besides, the Jewish birth rate is not sufficiently high to reach replacement levels.

Orthodox Jewish theology takes a position opposite to that of the Catholic Church on the priority of mother and fetus in instances when a decision has to be made which of the two is to be sacrificed to save the other. A one scholar has enunciated Jewish doctrine: "If a woman suffered hard labor in travail, the child must be cut up in her womb and brought out piecemeal, for her life takes precedence over its life; if the greatest part already has come forth, it must not be touched, for the claim of one life can not supersede that of another" (Zoloth 2003, pp. 40–41).

Attitudes Toward Abortion

A comprehensive analysis of the views regarding abortion among white and black men and women from the 1970s date demonstrated shifts over four decades. Factors said to have had an impact on these swings include the increase in female participation in the labor force, the alterations in the extent of nonmarital sexual activity (a dramatic upward move followed by a decline after the appearance of the AIDS epidemic), the further secularization of the society, the increase in educational attainments, the decreasing vitality of the feminist movement, and the growing vigor of the antiabortion movement (Carter et al. 2009, p. 3).

The trend analysis is based on the results of the General Social Survey that questions a representative sample of noninstitutionalized English-speaking adults in the United States. Abortion attitudes were determined by a composite score based on answers to seven questions, asking the respondent if he or she favored allowing abortion under each of the following conditions:

1. If there is a strong chance of a defect in the baby?
2. If she is married and does not want any more children?
3. If the woman's own health is seriously jeopardized by the pregnancy?
4. If the family has a very low income and cannot afford any more children?
5. If she becomes pregnant as a result of rape?
6. If she is not married and does not want to marry the man?
7. The woman wants it for any reason?

The was a slight decrease in support for abortion during the 1980s compared to the period between 1973, when *Roe v. Wade* was decided, and the following decade. That decline has been tied to the influence of President Reagan's antiabortion views but, of course, could have been a cause for his election. White males have been and remain the most liberal of the demographic groups studied when it comes to abortion attitudes. White females are the second most liberal group followed closely by black females and black males. The striking decline in support for abortion among black females from the 1980s to 2008 brings their score close to the point at which it stood in the 1970s before a dramatic rise from the 1980s to the 1990s. It is black males who show the greatest increase in support for abortion between the 1970s and current times (Carter et al. 2009).

Abortion and the Law

Frederick J. Tausig, a leading authority on abortion, observed of the pre-*Roe* days that he knew of "no other instances in which there had been such frank and universal disregard for a criminal law" (Taussig 1936, p. 422). There are greatly varying estimates of the number of illegal abortions in the United States that set the annual figure for the years right before the *Roe* decision at about 200,000–1.2 million. If all women who underwent illegal abortions then were included in the crime statistics, the female crime rate would have far exceeded that for males. In practice, sanctions were not applied to women who had abortions but to those who performed them. Most often, this happened when the aftereffects of the procedure brought the woman to a hospital. There she sometimes offered or was persuaded to testify against the abortionist (Regan 1997).

As with many highly controversial issues in the United States, such as racial desegregation, it was the judiciary that took the lead in endorsing the dramatic national shift in abortion practices. Part of the reason for this is that federal judges are appointed for life and need not be unduly concerned about public disfavor with their decisions. Presidents, governors, and legislators, if they want to remain in office, cannot afford to alienate too significant a portion of their constituency and often decide to stay on the sidelines in regard to actions that might split the public.

Fifteen states had already relaxed restrictions on abortion by 1973. In tandem with *Roe v. Wade*, the 7–2 Supreme Court decision that allowed abortions to be performed legally throughout the nation during the first trimester of pregnancy, the *Doe v. Bolton* (1973) opinion delivered the same day declared unconstitutional a Georgia law requiring that a doctor's decision regarding abortion had to be confirmed by another doctor or a committee and had to be performed in a hospital. Among the forces that appeared to underlie the *Roe* decision was the power of the feminist movement, particularly in terms of the considerable number of women moving into the work force. Birth control pills (replacing more cumbersome and less employed means of contraception) and the growing use of condoms called into question the inviolability of the birth process, since it could readily have been prevented before it began. Condoms became available in college campus restrooms, and birth control pills often sit on the kitchen counter, waiting to be swallowed with orange juice at breakfast.

Parental Notification

The matter of parental notification became the first major abortion battleground. In *Planned Parenthood of Kansas City, Missouri v. Ashcroft* (1983), the US Supreme Court ruled that it was constitutionally acceptable to demand that minors inform their parents of the intention to have an abortion; in some instances, telling a judge could substitute for parental notification. By 2012, in 22 states a parent had to consent to the procedure and in three of these, both parents. In 11 states, a parent only had to be notified and in one of these, both parents. In seven states, the parental notification had to be notarized. In the remaining states, an abortion could be had by a minor without parental involvement.

The irony in the success of the campaign to incorporate parental notification by underage women in the requirement for an abortion surfaced in a later study that indicated that the ruling appeared to have produced only a scant effect on teenage abortions. In Tennessee, for instance, the abortion rate went down after a federal court put a parental notification law on hold but increased when the requirement was reinstated. The rate of abortions fell in Texas after parental notification was mandatory but not nearly as much as it had dropped before then. All told, the study found "no evidence that the laws had a significant impact on the number of minors who got pregnant, and, once pregnant, the number who had abortions" (Lehren and Leland 2006, p. A1).

The *Webster* Case

An important victory for pro-lifers came 16 years after *Roe* when the Supreme Court in *Webster v. Reproductive Health Services* (1989) upheld the constitutionality of a Missouri law that prohibited the use of public facilities, such as municipally owned hospitals, to perform abortions, except when necessary to save a woman's life. The Missouri law also disallowed public funding for programs that included proabortion counseling. The court further agreed that a doctor could be required to inform a pregnant woman seeking an abortion whether her fetus was viable and might possibly survive if delivered prematurely.

The *Casey* Case

The Supreme Court's ruling in *Planned Parenthood v. Casey* (1992) found the justices deciding by a five to four vote to uphold a Pennsylvania law that required a woman seeking an abortion to listen to a lecture or watch a film about fetal development and then wait a day before undergoing the procedure. The doctor was obligated to provide information about alternatives to abortion, its medical risks, and the probable gestational age of the fetus. The doctor also had to inform the patient regarding medical benefits for prenatal care, childbirth, and neonatal assistance, among other matters. The Court, however, refused to endorse the segment of the Pennsylvania law that required the woman, unless she had compelling reasons for not doing so, to sign a statement that she had informed her husband that she was going to have an abortion.

The *Casey* decision was something of a disappointment for the pro-life camp which had thought there was a good chance that the Court would totally jettison *Roe v. Wade*. It led the antiabortion forces to decide to take a lesson from the civil rights movement and to incrementally pick away at elements of *Roe* until it had been so weakened that it would virtually fall of its own accord. They took particular heart from the clause in *Casey* that stated that the Court would be hospitable to abortion arrangements so long as they did inflict an "undue burden" upon women seeking an abortion. Pro-lifers anticipated that "undue burden" was so vague a criterion that it left a great deal of room for them to chip away at the guidelines for abortions.

Late-Term Abortions

The controversy between supporters of legal abortion and their opponents, like many campaigns for the minds of the public, has relied upon the manner in which disputed matters are worded. In the overall denotation of the two sides, those opposing abortion enjoy the verbal high ground. Pro-life is a more appealing slogan than pro-choice. Who can declare that they are not in favor of life, especially when considering its alternative? Pro-choice, on the other hand, sounds self-indulgent. Choices are always limited and where would its adherents draw the line? The unequal appeal of the way the sides have designated themselves has been noted elsewhere: "Perhaps 'pro-choice' once was good enough shorthand for liberty, human dignity, individualism, pluralism, self-government and woman's equality," writes Nancy Cohen (2010, p. A20), who then adds: "But anyone who thinks it still is sufficient as we enter our fifth decades of the cultural wars [over abortion], hasn't been paying attention." By the last, she means attention to the growing strength of the pro-life movement.

The same kind of verbal warfare was prominent in regard to what here is being called "late-term abortions." Pro-life forces defined the matter as "partial-birth abortion" and described the procedure in repellent terms. Pro-choice advocates tended to rely on the terms "dilation and extraction" (d & x) or "dilation and evacuation" (d & e), the medical designations for the procedure. These were abortions generally performed during the fifth to the sixth month of pregnancy. Estimates placed their number at somewhere between 2,000 and 5,000 a year.

The procedure was banned by the federal Congress in 2003. Signing the measure into law, President Bush noted that "for years a terrible form of violence has been directed against children who are inches from birth." The constitutionality of the measure was upheld by a 5-4 vote in the US Supreme Court (*Gonzalez v. Carhart* 2007). The law decreed a fine of not more than $250,000 and/or 2 years imprisonment for "any physician who, in or affecting interstate commerce, knowingly performs a partial birth abortion and thereby kills a human fetus." The Court's opinion noted, rather oddly, that "while we find no reliable data on the phenomenon it seems unexceptional to conclude some women come to regret their decision to abort the infant they once created and sustained" (*Gonzales v. Carhart* 2007, p. 159). This piece of obiter dicta (off-the-cuff musings not directly relevant to the case) is undoubtedly accurate, but the same can be said, perhaps more tellingly, about births – or, for that matter, about people who marry and Supreme Court judges who render decisions they come to regret.

Pro-choice groups had little leverage against the groundswell of public support in opposition to late-term abortions. The view of the liberal wing of the medical profession was expressed in an article in the prestigious *New England Journal of Medicine*:

> This is the first time the Court has ever held that physicians can be prohibited from using a medical procedure deemed necessary by the physician to benefit the patient's health. (Annas 2007, p. 2201)

Subsequent Developments

For a time thereafter, the abortion issue was confined to the periphery of social concerns in the face of America's involvement in wars in Iraq and Afghanistan and the drastic economic meltdown in late 2008. Then abortion recaptured in headlines in March 2011 when the Republicans in Congress held up to almost the very last moment passage of a budget bill that would keep the government functioning. It was not monetary concerns that were being fought over but the Republican's demand to end all funding for Planned Parenthood, a national agency that devotes a small portion of its work to abortion. The Republicans also wanted to stop funding for abortions in the District of Columbia, which is ruled by Congress, and to end the distribution of funds to any overseas government that reallocates some part of the money to agencies encouraging or performing abortions. The dispute over these matters finally was postponed at the last moment in order to keep the government running. But the fact that abortion again has become a political issue is indicated by the fact 944 bills were introduced intro state legislatures during the first 3 months of 2012 seeking to rein in abortions (about 3 % of these bill make it into law).

The Demographics of Abortion

The most reliable statistics on abortion in the United States are those gathered by the Alan Guttmacher Institute, a nonprofit organization that conducts research on human reproduction and performs policy analyses and sponsors public education programs.

The Institute notes that there have been about 50 million abortions since the procedure was authorized by the *Roe* decision in 1973 and 2008 (2008 was the latest year for which US figures were available as late as mid-2012.4 That figure represents a decrease from the highest levels of 1.6 million in the 1980s, despite a growing population. Part of the decrease is due to the appearance of the so-called morning-after abortion pills (RU-486 or Mifepristone) that were sold over the counter to women over 18 and by prescription to those younger.

The legal abortion rates by 4-year periods are presented in Table 1.

Nearly half of the pregnancies in the United States are reported to be unintended, and 40 % of these pregnancies are terminated by abortion. Eight percent of the abortions are performed on teenagers. The figures for other age groups are 20–24 years old (33 %) and 25–29 (24 %). Women who have never married and are not currently cohabiting accounted for 45 % of all abortions, while 61 % are performed on women who already have one or more children. Twenty-eight percent of the women who obtain abortions report that they are Catholics (Jones et al. 2010).

Four of the 13 possibilities were the median number offered by the 1,160 women questioned about the reason they chose to undergo an abortion. Three-fourths cited concern for or responsibility to other individuals; three-fourths said that they could not afford a child; three-fourths said that having a baby would interfere with work, school, or the ability to care for dependents; and half indicated that they did not want to be a single parent or were having problems with their husband or partner (Finer et al. 2005).

For and Against Legal Abortion

Most of the positions taken by pro-choice advocates and those favoring the pro-life positions are summarized below.

The Pro-life Claims

1. It is argued by pro-lifers that to allow legal abortion represents a wedge into more drastic reinterpretations of life and its value. They maintain that since the elderly tend to be a drain on the economy, especially in terms

Legal Status of Abortion, Table 1 Abortion rate per 100,000 women, 1973–2008

Year	Rate
1973	16.3
1977	26.4
1981	29.9
1985	28.0
1989	26.8
1993	25.0
1997	21.9
2001	21.1
2005	19.4
2008	19.48

Source: Alan Guttmacher Institute

of their pensions and medical costs, attitudes favoring abortion could readily be expanded to form a basis for a program of euthanasia directed against old people.

2. Legal abortion is said to make fetal experimentation and human cloning more acceptable, actions most pro-lifers deplore.
3. Pro-lifers maintain that many women who undergo abortions later come to deeply regret that irreversible action; therefore, all reasonable means ought to be allowed to keep them from doing something they subsequently would wish they had not done.
4. Relatively few persons, pro-lifers point out, no matter how wretched their situation may be, choose to end their lives. Nor do many people wish they had never been born. The argument is offered that the fetus being aborted has no way of registering his or her view on the matter and that pro-lifers must defend the fetus.
5. Pro-lifers believe that abortion encourages immoral behavior, such as premarital or extramarital sexual intercourse.
6. The term "genocide" is sometimes used by pro-lifers to describe what they view as the slaughter of millions of unborn human lives.
7. Pro-lifers often insist that the Supreme Court put its nose (and votes) into business that is far removed from its judicial concerns and competence.
8. In regard to *Doe v. Bolton*, pro-lifers maintain that the definition of a pregnant woman's health was overbroad and ill defined since it includes emotional, familial, and other conditions that can be employed as excuses for a self-indulgent act.
9. With ultrasound techniques now able to determine the gender of the fetus, pro-lifers note that abortion will be used to discriminate against female fetuses, skewing the population ratio and in time having far-reaching detrimental consequences for social life.
10. Pro-lifers argue that the loss of manpower and womanpower because of legal abortions creates a necessity for the United States to rely upon foreigners to maintain an adequate workforce. This necessity, they argue, has led to a considerable influx of illegal aliens who are claimed to be a drain on the economy.
11. Pro-lifers say that as a nation that largely derives its moral principles from biblical writings, we should be obligated to follow Christian theological doctrines that preach divine objections to abortion.

The Pro-choice Claims

1. Pro-choice advocates emphasize that making abortion illegal has racist implications. They point out that when abortion was against the law, fatalities from outlawed procedures were found primarily among poor minority group women.
2. Pro-choicers believe that if women bear children they do not want, they are very likely to severely undermine the quality of life in store for those children and, as a result, impose social costs upon others.
3. Pro-choicers dispute the claim that there exists biblical justification for the crusade against legal abortions and that, even if there are such doctrines, they should not be used to interfere with options available to those who do not accept them as guidelines for their own conduct.
4. While granting that it was a Supreme Court of males that enunciated the *Roe v. Wade* doctrine, pro-choicers point out that it has been only males who have allowed inroads against

the unfettered application of the original ruling. None of the women who are or have been on the court have ever voted in favor of an abortion restriction.

5. The pro-choice forces find the ideologies of pro-lifers illogical and contradictory if not hypocritical. They point out that pro-lifers tend to argue that they want the government to get off their backs, yet at the same time, they advocate that the government intrude into people's freedom by forbidding them to undergo abortions.

Conclusion

The fight (battle may be the better term) over abortion shows few signs of abating. It is difficult to think of a resolution that would appease and subdue both sides. Perhaps the best shorthand summary of the situation is that of Laurence Tribe, an eminent constitutional law professor: "What we have here," Tribe (1991, p. 6) has observed, "is a clash of absolutes."

Related Entries

▶ Gendering Traditional Theories of Crime
▶ Moral Crimes
▶ Sex Offenders and Criminal Policy

Recommended Reading and References

Annas GJ (2007) The supreme court and abortions. N Engl J Med 350:2200–2207

Asma ST (1994) Abortion and the embarrassing saint. Humanist 54(May-June):30–33

Baker RC, Ehlke R (2011) Natural law: a Lutheran reappraisal. Concordia, St. Louis

Bartlett LA, Berg CJ, Schulman HB, Zane SB, Green CA, Whitehead S, Atrash HK (2004) Risk factors for legal induced abortion-related mortality in the United States. Obstet Gynecol 103:729–737

Benedict XVI (2010) Celebration of first vespers of the First Sunday of Advent for unborn life. Libera Editice Vaticana, Vatican City

Bindley TH (1890) The apology of Tertullian for the Christians. Parkler, Oxford, UK

Bolzendahl CI, Myers DL (2004) Feminist attitudes and support for gender equality: opinion change in women and men. Soc Forces 83:759–789

Cahill LS, Farley MA (1995) Embodiment, morality, and medicine. Kluwer, Boston

Carter JS, Carter S, Dodge J (2009) Trends in abortion attitudes by race and gender: a re-assessment over a four-decade period. J Sociol Res 1:1–17

Cohen NL (2010) Framing today's abortion debate. Los Angeles Times, A10, 10 Apr 2010

Davis JJ (1984) Abortion and the law: what every Christian should know. Presbyterian and Reformer, Phillipsburg

Doe v. Bolton (1973) 410 U.S. 179

Donahuse JJ III, Levitt SR (2001) The impact of legalized abortion on crime. Q J Econ 116:399–420

Downie A (2009) Nine year olds abortion outrages Brazil's Catholic Church. Time World, 21–22, 6 Mar 2009

Finer LB, Frohwirth LF, Lindsay A, Dauphenee SS, Moore AM (2005) Reasons U.S. women have abortions: quantitative and qualitative perspectives. Perspect Sex Reprod Health 37:110–118

Gonzalez v. Carhart (2007) 550 U.S. 154

Gruber J, Levine P, Staiger D (1999) Abortion legislation and child living conditions: who is the 'Marginal Child'? Q J Econ 114:269–291

Haugern DM, Musser S, Lovelace K (eds) (2010) Abortion: opposing viewpoints. Greenhaven, Detroit

Hess JA, Rueb JD (2005) Attitudes toward abortion, religion, and party affiliations among college students. Curr Psychol 24:23–42

Hope TA, Wilder EI, Watt TT (2003) The relationships among adolescent pregnancy, pregnancy resolution and Juvenile delinquency. Sociol Q 44:555–576

Hull NE, Hoffer W, Hoffer PC (eds) (2004) The abortion rights controversy in America. University of North Carolina Press, Chapel Hill

Jackson Women's Health Organization v. Currier (2012) 12–00436 (U.S. District Court, Southern District of Mississippi (Jackson)

Jones RK, Finer LB, Singh S (2010) Characteristics of abortion patients, 2008. Guttmacher Institute, New York

Joyce T (2003) Did legalized abortion lower crime? J Hum Resour 39:1–28

Lehren A, Leland J (2006) Parental notification has scant effect in teen abortion. New York Times, A1, A19, 6 Mar 2006

Markoe L (2010) After Mississippi defeat, what next for 'Personhood'? Christ Century 128:18, December 10

Planned Parenthood of Kansas City, Missouri v. Ashcroft (1988) 462 U.S. 476

Planned Parenthood of Southeastern Pennsylvania v. Casey (1992) 505 U.S. 802

Presshirran AM (1990) Potential fathers and abortion: a woman's womb is not a man's castle. Brooklyn Law Rev 1359–1404

Regan LJ (1997) When abortion was a crime: women, medicine and the law in the United States, 1867–1973. University of California Press, Berkeley
Roe v. Wade (1973) 410 U.S. 413
Sulzberger AG (2011) Women seeking abortions to get anti-abortion advice. New York Times, A16, A18, 23 Mar 2011
Tribe LH (1991) Abortion: clash of opposites. Norton, New York
Varmos H (2003) Summary report: early reproductive events and breast cancer workshop. National Cancer Institute, Washington, DC
Webster v. Reproductive Health Services (1989) 492 U.S. 490

Lie Detection

▶ Detecting Deception with fMRI

Lie/Truth Detection

▶ Evaluating truthfulness: Interviewing and Credibility Assessment

Life Course

▶ Individual Characteristics and General Strain Theory

Life Events

▶ Integrated Cognitive Antisocial Potential Theory

Life-Course Criminology

▶ Integrated Cognitive Antisocial Potential Theory

Life-Course-Persistent Offending

▶ Onset of Offending

Lifestyle Theory

Glenn D. Walters
Kutztown University of Pennsylvania,
Kutztown, PA, USA

Synonyms

Criminal thinking model; Lifestyle theory of crime

Overview

Lifestyle theory holds that crime is a developmental process guided by an ongoing interaction between three variables (incentive, opportunity, and choice). During each phase of the criminal lifestyle (initiation, transition, maintenance, burnout/maturity), incentive, opportunity, and choice take on different values and meanings. Existential fear serves as the incentive for the initiation phase of a criminal lifestyle. Once initiated, the incentive for continued lifestyle involvement becomes a fear of losing out on the benefits of crime. By the time the individual enters the third (maintenance) phase of a criminal lifestyle, incentive has changed once again, this time to a fear of change. With the advent of the burnout/maturity phase of the criminal lifestyle, incentive has changed yet again, this time to a fear of death, disability, or incarceration. Comparable transformations take place in opportunity and choice. Current controversies include (1) generalizing results obtained from research on male prison inmates to that on female inmates and nonincarcerated offenders, (2) documenting the clinical predictive efficacy of assessment procedures designed to measure key lifestyle constructs, and (3) examining the possibility of an allegiance effect for much of the research on lifestyle theory. Open questions requiring further study include (1) investigating the hierarchical model of criminal thinking proposed by lifestyle theory, (2) exploring the role of criminal thinking and other quasi-time-stable cognitive factors in mediating and clarifying important crime relationships, and (3) ascertaining whether the lifestyle approach to intervention qualifies as evidence based.

Introduction

The lifestyle theory of crime has its roots in Bandura's (1986) social cognitive model, Sykes and Matza's (1957) techniques of neutralization, and Yochelson and Samenow's (1976) work on the criminal personality. As such, it is designed to explain habitual criminal activity by focusing on the cognitive factors that support antisocial behavior. The cognitive-behavioral proclivities of lifestyle theory are obvious given its emphasis on cognition, behavior, and the cognition-behavior relationship. Less obvious, perhaps, is how lifestyle theory integrates cognition and behavior. This entry, in addition to describing the lifestyle theory of crime, is designed to provide the reader with an understanding of exactly how cognition and behavior interact to increase or decrease a person's risk for future criminal involvement.

Background

For much of its history, criminology has restricted itself to a small portion of the relevant data, resulting in the simplistic single-variable theories that now dominate the field. Biology and psychology have been largely ignored by criminologists and while biology and psychology are no more capable of providing a complete explanation of crime than criminology, a complete explanation necessitates their inclusion. The lifestyle theory of crime attempts to highlight psychological variables that may be helpful in explaining certain well-known crime relationships. In this vein, lifestyle theory seeks to integrate constructs from divergent conceptual models rather than perpetuate the artificial dichotomies that seem to have limited theory development in the field of criminology (i.e., classicism vs. positivism, propensity vs. development, continuity vs. change). How lifestyle theory integrates these artificial dichotomies is discussed next.

Early criminological theory emphasized the classical perspective that humans are rationale decision makers who engage in behaviors they believe will provide them with the greatest amount of pleasure and the least amount of pain. The criminal justice system still relies heavily on the classical notion that people choose to commit crime. Theoretical criminology, however, has largely rejected the notion of choice in favor of a deterministic model of criminal behavior in which crime is seen as a function of sociological and environmental factors over which the actor has no control. Lifestyle theory integrates these opposing points of view into a single perspective in which incentive (pushes from within), opportunity (pulls from without), and choice (the decision-making apparatus) are equally important in the development of criminal behavior.

Some criminological theories view offenders as *exhibiting a* propensity for crime; other criminological theories postulate that crime follows a developmental sequence or pattern. The first approach underlies the career criminal paradigm and the second approach lays the foundation for the criminal career paradigm. Lifestyle theory asserts that crime is both a propensity and developmental process and that the career criminal and criminal career paradigms, rather than being diametrically opposed, are actually complementary. Integration of the career criminal and criminal career paradigms gives rise to four overlapping but sequential phases of lifestyle development: initiation, transition, maintenance, and burnout/maturity. Certain propensities lead some individuals to drop out of the sequence during an early phase (initiation, transition) and others to avoid the sequence altogether. Alternate propensities lead some individuals to remain in the sequence or in a particular phase of the sequence longer than most people.

Continuity versus change is another dichotomy that has preoccupied the field of criminology. Some theorists conceptualize crime as a time-stable characteristic that is largely impervious to change, whereas other theorists conceptualize crime as an unstable developmental pattern that is subject to regular and dramatic periods of change. Crime continuity, whereby past offending serves as one of the best predictors of future offending across multiple studies, is often used to support the stability argument. The age-crime relationship, in which crime peaks during mid-adolescence and then drops off sharply in late adolescence regardless of whether crime is measured with official, self-report, or victimization data, is often used to support the instability argument. Lifestyle theory agrees with both positions and integrates stability and instability into its framework by proposing the existence of quasi-time-stable cognitive and behavioral variables that give the lifestyle the appearance of being both stable and changeable.

Another dichotomy that has helped shape the lifestyle theory of crime is whether individual differences in criminality are categorical (difference in kind) or dimensional (difference in degree). This time, however, one side of the controversy (i.e., dimensional latent structure) rather than an integration of the two sides receives the bulk of empirical research support. In a recent review of the taxometric literature on antisocial personality, psychopathy, and criminal lifestyle, Haslam (2011) concludes that these crime-related constructs are dimensional rather than categorical in nature and that individual differences in these crime-related psychological constructs are quantitative (people can be ordered along one or more dimensions) rather than qualitative (people can be grouped into types or categories). Based on taxometric and confirmatory factor analytic research (Walters 2009), lifestyle theory proposes that a criminal lifestyle is composed of two overlapping dimensions: proactive criminality and reactive criminality.

State of the Art

The lifestyle theory of crime, as described in Walters (1990), has undergone several revisions and elaborations. This section on the state of the art of lifestyle theory provides a summary of the most recent version or iteration of lifestyle theory as applied to habitual criminal conduct (Walters 2012a).

Precursors to a Criminal Lifestyle

Prior to entering the initial phase of a criminal lifestyle, certain conditions are already in place. Two commonly observed precursors of a criminal lifestyle are templates and trial runs. A template consists of cultural and subcultural factors that provide a context for subsequent development of a criminal lifestyle. Sundry environmental and familial factors help shape an individual's attitudes and thinking toward antisociality and crime, which, in turn, increase or decrease the person's susceptibility to future criminogenic influences. The child's interactions with the interpersonal environment consequently form a template that makes it more or less likely that he or she will pursue criminal opportunities present in that environment. Trial runs are the individual's initial attempts to employ antisocial solutions to solve interpersonal problems (e.g., acquiring a toy they want, avoiding punishment for stealing a treat from the proverbial cookie jar). The reaction the child receives from his or her interpersonal environment will go a long way toward shaping his or her thinking with respect to future criminal opportunities.

Phase I: Initiation

The first phase of a criminal lifestyle is referred to as initiation. This phase begins with commission of the first arrestable crime and ends when the individual either adopts a conventional lifestyle or moves into the next phase of a criminal lifestyle. Each phase of a criminal lifestyle is a function of specific incentives, opportunities, and choices, around which the current discussion is organized.

Incentive: Existential Fear

Existential fear is a fear of nonbeing combined with a sense of separation and alienation from the environment. Although it is a fear shared by all humans, it is shaped and molded by a person's experiences in the survival-relevant areas of affiliation, control/predictability, and status.

Those whose fears have been shaped by affiliative concerns might experience existential fear as a fear of fitting in or being rejected. Those whose fears have been shaped by control concerns might experience existential fear as a fear of losing control. Those whose fears have been shaped by status concerns may perceive existential fear as a fear of being anonymous or unsuccessful. Existential fear serves as an incentive for behavior in that it motivates, pushes, or encourages the individual to engage in behaviors designed to reduce or alleviate the fear. The degree to which a criminal lifestyle promises the individual relief from existential fear is the degree to which the individual is motivated to enter a criminal lifestyle.

Initially, existential fear is tied to one's physical survival. It is, in effect, a manifestation of the person's survival instinct. This can become distorted over time, however, to where the individual paradoxically favors psychological survival of the lifestyle over physical survival of the organism. This is particularly true of a criminal lifestyle. Research indicates that criminality is often associated with low anxiety or fearlessness (Newman and Schmitt 1998). Lifestyle theory conceptualizes fearlessness as a relatively weak bond between existential fear and physical survival, keeping in mind that lifestyle theory is a dimensional model and that bond strength is a matter of degree rather than an all-or-nothing proposition. Because existential fear needs to be attached to something, the individual with low fearlessness will often attach their existential fear to a lifestyle. Under the proper conditions, this lifestyle could turn out to be a criminal lifestyle.

Opportunity: Early Risk Factors

Opportunity factors increase or decrease a person's risk of entering a lifestyle. In line with the criminal lifestyle's dimensional structure, no single risk factor determines a person's position on the proactive and reactive dimensions of the lifestyle; rather, it is the total number of relevant risk factors that is important (additive etiology). Some risk factors have a greater impact than other risk factors, certain risk factors interact with one another, and some risk effects are moderated by a third variable. Childhood temperament is considered a particularly salient risk factor for the purpose of initiating a criminal lifestyle. Novelty seeking, negative emotionality, and physical activity are three temperament dimensions likely to be elevated in someone at risk for future antisociality. Because childhood temperament is a function of both genetics and early environment, it demonstrates the complex interaction that exists between biological and developmental factors in the formation of a criminal lifestyle. Other risk (opportunity) factors vital during the initiation phase of a criminal lifestyle include stress, weak socialization to conventional groups, strong socialization to deviant groups, and the availability of criminal opportunities in the current environment.

Choice

Lifestyle theory rejects the hard determinism of the positivistic tradition. Even though incentive drives behavior and opportunity shapes it further, the individual still makes choices. The active decision-making that gives rise to choice is a two-stage process: generation and evaluation. The goal of the generation stage of the decision-making process is to come up with as many alternative solutions to a problem as possible. The goal of the evaluation stage of the decision-making process is to systematically and effectively evaluate the pros and cons of each alternative option. For a variety of reasons, from intelligence to experience to poor integration of decisional economics (rational choice) and emotion (empathy), the crime-prone individual often has trouble generating alternatives, properly evaluating alternatives, or both. This is a problem during all four phases of a criminal lifestyle, but during the initiation phase the issue often is resisting the powerful effects of actual and anticipated positive reinforcement for criminality (excitement, curiosity, peer acceptance). Lifestyle theory rejects traditional rational choice and deterrence theory in favor of a model that encompasses both rational and irrational choice.

Phase II: Transition

The transitional phase of lifestyle development is characterized by increased involvement in, commitment to, and identification with the criminal lifestyle. In contrast to the experimentation of the initiation phase, the transitional phase is marked by a growing sense of comfort with the lifestyle accompanied by the belief that certain basic needs will be satisfied by the lifestyle.

Incentive: Fear of Lost Benefits

It is during the initiation phase of a criminal lifestyle that the individual comes to realize the material and psychological benefits that can be derived from being involved in a regular pattern of criminality. Existential fear facilitates the transition to a higher level of lifestyle involvement by stimulating a person's fear of losing the material (money, excitement, power) and psychological (affiliation, control, status) benefits of a criminal lifestyle. This transformation of existential fear into a fear of losing the material and psychological profits of a criminal lifestyle is instrumental in transitioning the individual to the next phase of the lifestyle whereby commitment to the lifestyle and certain corollaries (initial incarceration, labeling, rejection of conventional values) take precedence.

Opportunity: Schematic Subnetworks

With increased internalization of the criminal lifestyle, opportunity transitions from behavioral to cognitive. This is another way of saying that a person starts acting like a criminal before he or she starts thinking like one. Once the individual starts thinking like a criminal, his or her opportunities are shaped primarily by six cognitive factors known as schematic subnetworks. Lifestyle theory proposes that criminal cognition is hierarchically organized; with belief systems (self-view, worldview, past view, present view, future view) at the top of the hierarchy, individual criminal thoughts are the bottom of the hierarchy, and several layers of schematic subnetwork in between (see Fig. 1). The six primary schematic subnetworks in the lifestyle model are thinking styles, attributions, outcome expectancies, efficacy expectancies, goals, and values (see Table 1 for more details).

Choice

Decision-making tends to narrow as the individual's commitment to a criminal lifestyle grows. Many people who enter the transitional phase of a criminal lifestyle never possessed good problem-solving skills to begin with. Of those who enter this phase with adequate problem-solving skills, it is the generation stage of the problem-solving process that is most adversely affected by the individual's growing commitment to a criminal lifestyle. As the person begins narrowing the focus of his or her problem-solving deliberations to criminal options, he or she starts a process, facilitated, in part, by criminal thought patterns and other schematic subnetworks (which can be considered purveyors of irrational choice), of discarding and denigrating the noncriminal options available to him or her. Near the end

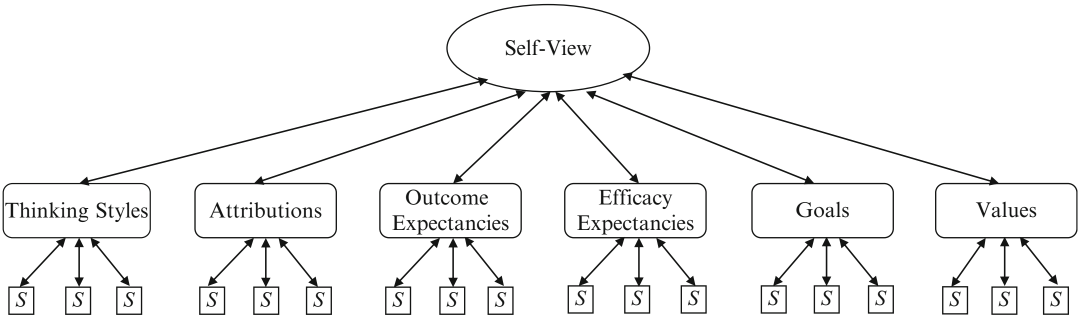

Lifestyle Theory, Fig. 1 The hierarchical organization of criminal cognition (Note. S = individual criminal thoughts or schemes)

Lifestyle Theory, Table 1 The six quasi-time-stable cognitive variables in lifestyle theory

Cognitive factor	Description
Criminal thinking	Thinking styles that support criminal behavior. The criminal thinking hierarchy is organized, from top to bottom, into a superordinate factor (general criminal thinking), two higher-order factors (proactive and reactive criminal thinking), and eight individual thinking styles (mollification, entitlement, power orientation, and superoptimism under proactive; cutoff, cognitive indolence, and discontinuity under reactive; sentimentality by itself)
Attributions	Beliefs about the causes of one's own or others' action; self-labeling and hostile attribution biases, whereby the individual perceives an unintentional act on the part of another person as a deliberate and hostile act, are two ways attributions help maintain a criminal lifestyle
Outcome expectancies	Beliefs about the anticipated consequences of crime and other behaviors; the criminal lifestyle is characterized by strong positive outcome expectancies for crime (money, power, acceptance) and weak negative outcome expectancies for crime (death, injury, incarceration)
Efficacy expectancies	Beliefs about one's chances of successfully engaging in a behavior or completing a task; the criminal lifestyle is associated with strong self-efficacy for crime and weak self-efficacy for conventional behavior
Goals	Objectives one pursues; the goals that direct a criminal lifestyle tend to be of short (reactive) to intermediate (proactive) range rather than long term
Values	Priorities that govern one's actions; a criminal lifestyle tends to be driven by physical (reactive) and mental (proactive) hedonistic values

of the transition phase, it is not uncommon to find the proactive (scheming, cold-blooded, e.g., "what's in it for me?") and reactive (impulsive, hot-blooded, e.g., "I'll get you for that!") dimensions of decision-making exerting independent, simultaneous, and poorly modulated effects on behavior.

Phase III: Maintenance

Whereas the majority of those who enter the initiation phase of a criminal lifestyle never progress to the transitional phase and a sizeable minority of persons who move into the transitional phase drop out before entering the maintenance phase, there is virtually no attrition from the maintenance phase. This is because the maintenance phase of a criminal lifestyle is a period of maximum lifestyle involvement, commitment, and identification.

Incentive: Fear of Change

As scary as a criminal lifestyle and its consequences (injury, death, and incarceration) may be, they are often not as frightening as the prospect of change. Fear of change consequently becomes the primary incentive for remaining in a criminal lifestyle long after it has stopped being fun. The individual feels compelled to remain in the lifestyle even though the benefits no longer seem to outweigh the costs, and it is a fear of change or better yet, a fear of the unknown, that is behind the individual's inactivity and apparent immobility. Things may not be as the person would like them to be but there is comfort to be found in the familiar, even when the familiar is no longer as comfortable as it once was. Giving up crime may be interpreted as symbolic death by someone in the maintenance phase or at least tacit acceptance that the years spent in a lifestyle were a waste of time and a poor decision on his or her part.

Opportunity: Psychological Inertia

Psychological inertia is based on Newton's first law of motion, which states that a body at rest will remain at rest and a body in motion will remain in motion unless acted upon by an outside force. Once a criminal lifestyle is in motion, it will maintain itself until acted upon by an outside force for change. The progenitors of psychological inertia are the six schematic subnetworks mentioned in the previous section on transition. Each of these quasi-time-stable cognitive factors gives rise to continuity in criminal behavior. Criminal thinking, for instance, provides offenders with a way of understanding the world that is consistent from one situation to the next and helps rationale their ongoing criminal behavior. Attributions like self-labeling, outcome expectancies of unlimited power and control, high self-efficacy for crime and low self-efficacy for conventional behavior, short-term goals, and hedonistic values all keep the individual locked in chronic pattern of offending by way of psychological inertia.

Choice

During the maintenance phase of a criminal lifestyle, the individual may truly believe that he or she has no choice other than to remain in the lifestyle. Given that lifestyle commitment is maximal during this phase, it is easy to see why many late phase offenders feel "stuck" in the lifestyle or view continued criminal involvement as their "fate." Hence, attitudes expressed by probation officers, correctional staff, or even counselors suggesting that offenders never change (i.e., "once a criminal, always a criminal") may serve to inhibit the natural changes process that leads to change in even the most recidivistic of offenders. The ability to generate and evaluate alternatives is almost universally weak in those who reach the maintenance phase of a criminal lifestyle. Fear of change and psychological inertia only serve to reinforce the fatalistic belief that their situation will never change. There is hope, however, and this hope arrives in the form of the fourth and final phase of a criminal lifestyle, burnout and maturity.

Phase IV: Burnout and Maturity

The combined effect of the negative long-term consequences of a life of crime and the aging process leads to the fourth and final phase of a criminal lifestyle, burnout and maturity. Whereas burnout is a decrease in physical energy and stamina that makes crime more difficult and less pleasurable, maturity is a psychological process involving a genuine change in interests, goals, values, and activities. Physical burnout is inevitable, psychological maturity is not. An individual who is physically burned out but has not yet achieved psychological maturity may switch to a less physically taxing criminal activity, like dealing in stolen property; nevertheless, he or she will remain on the outskirts of the lifestyle. Whereas the transition from the maintenance phase of a criminal lifestyle to burnout and maturity can be abrupt, it is normally a gradual and uneven process.

Incentive: Fear of Death, Disability, and Incarceration

With respect to fear, the offender has come full circle once he or she enters the burnout/maturity phase of a criminal lifestyle. One of the factors associated with initiation of the lifestyle is a weakened bond between existential fear and physical survival and the creation of a robust bond between existential fear and some activity, in this case, the criminal lifestyle. During burnout/maturity the bond between existential fear and physical survival strengthens, while the existential fear-lifestyle bond weakens. This is achieved by way of a growing fear of incarceration, a fear of dying in prison, and fears associated with the negative consequences of a criminal lifestyle, namely, death, injury, and disability.

Opportunity: Approach and Avoidance

Several factors support desistence from crime by increasing opportunities for prosocial activity and decreasing opportunities for antisocial activity. This can be accomplished by approaching goals, options, and outcomes incompatible with crime, such as marriage, parenthood, and conventional employment, or avoiding goals, options, and outcomes compatible with crime, such as drug use, criminal associates, and settings where one has committed crime in the past. As approach and avoidance opportunity factors interact, burnout and maturity increase and the risk of future criminal involvement drops dramatically.

Choice

Choice plays a vital role in crime initiation, maintenance, and desistance. During burnout and maturity, its role is to focus the individual on the rapidly accumulating negative consequences of criminal behavior, from incarceration to death. Many individuals, as they get older, become better problem solvers. Age seems to have a positive effect on a person's ability to both generate and evaluate alternatives and the negative consequences that accompany a criminal lifestyle are more difficult to accept and tolerate as the person ages. More people exit a criminal lifestyle during the initiation phase, but the greatest proportion of people exit the lifestyle during burnout and maturity. Fear of death, disability, and incarceration; approaching prosocial situations and avoiding antisocial ones; and placing greater emphasis on the negative consequences of crime than on the perceived benefits of crime are at the heart of the burnout and maturity phase.

Current Controversies

Nearly all of the research on lifestyle theory has been conducted on male inmates serving time in US federal prisons. A handful of studies have been conducted on US state prisoners and forensic patients and several studies have been done on European samples but only one study used female participants (Walters et al. 1998), and no studies have tested lifestyle theory in community corrections clients. The research base for lifestyle theory must consequently be expanded. Not only is there a need for more research on female offenders and community samples, but research on the invariance of lifestyle principles across ethnic groups and crime categories is also required. Research on juvenile samples is also needed. Whereas application of lifestyle theory to children and adolescents has been covered recently (i.e., Walters 2012a), there have been no research studies on lifestyle theory in which juveniles have served as subjects.

Assessment procedures have been developed to measure key concepts in lifestyle theory but the clinical utility of these measures remains largely untested. Thus far, the Lifestyle Criminality Screening Form (LCSF: Walters et al. 1991) has been developed to assess the behavioral dimensions of a criminal lifestyle (irresponsibility, self-indulgence, interpersonal intrusiveness, and social rule breaking), the Psychological Inventory of Criminal Thinking Styles (PICTS: Walters 1995) can be used to assess criminal thinking, and the Outcome Expectancies for Crime scales (OEC: Walters 2003) are available for assessing positive and negative outcome expectancies for crime. In the name of clinical utility, these measures should be capable of predicting important crime outcomes like recidivism with at least modest to moderate effectiveness. In addition, they should also possess incremental validity relative to easily obtained measures like age and criminal history.

Meta-analyses have been conducted on the LCSF ($k = 11$) and PICTS ($k = 7$) as predictors of recidivism, with r serving as the effect size measure and studies being combined using the random effects model. The results reveal a weighted effect size of .23 (95 % CI = .15–31) for the LCSF and .20 (95 % CI = .15–24) for the PICTS General Criminal Thinking (GCT) score. Viewing these results relative to Cohen's (1988) guidelines for small (.10), moderate (.24), and large (.37) effect sizes, we can see that the LCSF and PICTS both achieved effect sizes in the small to moderate range. Because the LCSF is a measure of criminal history, it is not possible to evaluate its ability to predict recidivism beyond the effects of age and criminal history. However, when the PICTS GCT score was entered into a regression equation behind age and criminal history, it continued to predict recidivism above and beyond the contributions of age and criminal history (Walters 2012b). A meta-analysis or incremental validity analysis could not be performed on the OEC because of a lack of recidivism data.

Theoretical articles and research reports on lifestyle theory have been published largely by one individual, the author of lifestyle theory. An allegiance effect can arise any time an author is evaluating his or her own theory and is most commonly observed in studies where an assessment device or therapeutic modality is being evaluated. Eight of the eleven LCSF studies and five of the seven PICTS studies included in the previously mentioned meta-analyses were performed by the author of lifestyle theory (Walters). A small difference was observed when the 13 effect sizes obtained in studies by Walters ($r = 0.23$, 95 % CI = 0.17–0.28) were compared with the five effect sizes obtained in studies conducted by outside researchers ($r = 0.18$, 95 % CI = 0.09–0.26); however, the difference disappeared when a single outlying study was removed from the outside researcher group ($r = 0.22$, 95 % CI = 0.14–0.29). Only two empirical studies have evaluated the lifestyle approach to intervention and both were published by Walters (1999; 2005). Until outside researchers conduct more studies on lifestyle theory, the possibility of an allegiance effect for research on lifestyle theory remains an open question.

Open Questions

Besides investigating whether an allegiance effect accounts for some of the positive results obtained in research on lifestyle theory, there are three other open questions that demand attention. First, there is a need for more research on the hierarchical structure of criminal thinking. This hierarchy fits into the criminal thinking schematic subnetwork box found in Fig. 1, with general criminal thinking at the top, proactive and reactive criminal thinking in the middle, and the eight individual criminal thinking styles at the bottom. Applying item response theory (IRT) principles and confirmatory factor analysis to a sample of nearly 3,000 incarcerated male offenders, Walters, Hagman, and Cohn (2011) determined that the sentimentality scale did not load onto either of the two higher-order factors (proactive, reactive) or the superordinate general criminal thinking factor. Based on these results, it has been recommended that instead of calculating the GCT score by combining the raw scores of the eight thinking style scales, the seven thinking style scales other than sentimentality be used to calculate the GCT. Further research is nonetheless required to cross-validate these results in a noninstitutionalized sample.

A second open question is whether the cognitive factors in lifestyle theory mediate important crime relationships. In the first of several studies, Walters (2011) discovered that the PICTS GCT score partially mediated the relationship between a history of serious mental health problems and subsequent institutional violence in a group of federal prison inmates. A second study found that the GCT score partially mediated the relationship between race and recidivism (Walters in press b) and a third study revealed that the PICTS Reactive Criminal Thinking score partially mediated the relationship between prior substance abuse and subsequent recidivism (Walters 2012c). In a fourth study, Walters (in press a) determined that the GCT score and weak self-efficacy to avoid future police contact both mediated the relationship between past and future criminal conduct. It would appear that at least some of the quasi-time-stable cognitive factors in lifestyle theory are capable of mediating crime-relevant relationships, although further research is required to ascertain the extent to which the effect applies to all six factors.

A third open question is whether lifestyle intervention can be considered evidence based. Given that there have been only two empirical studies on lifestyle intervention to date (Walters 1999; 2005), there is insufficient evidence at this time to conclude that lifestyle intervention is evidence based. Nevertheless, the approach is manualized and has been adapted for use in an outpatient substance abuse program in Denmark (Thylstrup and Morten Hesse in press). One of the founding principles of lifestyle intervention is that while behavior proceeds cognition in the development of a lifestyle (i.e., a person starts acting like a criminal before he or she starts thinking like one), cognition proceeds behavior in lifestyle change (i.e., a person stops thinking like a criminal before he or she stops acting like one). Although cognition and behavior cannot be meaningfully separated, the early stages of lifestyle intervention focus primarily on challenging criminal thinking patterns and other cognitive mediators of criminal behavior, with the behavioral interventions becoming more prominent at latter stages of the treatment process. Research is required, however, to determine whether the progression proposed by lifestyle theory (i.e., start by focusing on cognition and then move into behavior) is justified.

Conclusion

The lifestyle theory of crime is presented for the purpose of illustrating how psychological factors are capable of furthering our understanding of criminal behavior. Lifestyle theory seeks to reconcile popular dichotomies in the field of criminology (classicism vs. positivism, propensity vs. development, continuity vs. change) by incorporating features of criminality that have been largely ignored by traditional criminological theories. After reviewing the developmental progression vital in initiating and maintaining a criminal lifestyle, controversial topics and open questions concerning the theory are discussed. The future of lifestyle theory depends on its ability to attract the attention of outside researchers and clinicians so that the model's potential can be tested and its limitations delineated.

Related Entries

- ▶ Career Criminals and Criminological Theory
- ▶ Cognitive/Information Processing Theories of Aggression and Crime
- ▶ Criminal Careers
- ▶ Desistance from Crime
- ▶ Interactional Theory of Delinquency
- ▶ Offender Change in Treatment
- ▶ Offender Decision Making and Behavioral Economics
- ▶ Psychopathy
- ▶ Psychopathy and Offending
- ▶ Social Control Theory of Sexual Homicide Offending

Recommended Reading and References

Bandura A (1986) Social foundations of thought and action: a social cognitive theory. Prentice-Hall, Englewood Cliffs

Cohen J (1988) Statistical power analysis for the behavioral sciences, 2nd edn. Lawrence Erlbaum, Hillsdale

Haslam N (2011) The latent structure of personality and psychopathology: a review of trends in taxometric research. Sci Rev Mental Health Prac 8:17–29

Newman JP, Schmitt WA (1998) Passive avoidance in psychopathic offenders: a replication and extension. J Abnorm Psychol 107:527–532

Sykes GM, Matza D (1957) Techniques of neutralization: a theory of delinquency. Am Sociol Rev 22:664–670

Thylstrup B, Hesse M (in press) The impulsive lifestyle counseling program for antisocial behavior in outpatient substance abuse treatment. Int J Offender Therapy Comp Criminol

Walters GD (1990) The criminal lifestyle: patterns of serious criminal conduct. Sage, Newbury Park

Walters GD (1995) The psychological inventory of criminal thinking styles: part I. Reliability and preliminary validity. Crim Justice Behav 22:307–325

Walters GD (1999) Short-term outcome of inmates participating in the lifestyle change program. Crim J Behav 26:322–337

Walters GD (2003) Changes in outcome expectancies and criminal thinking following a brief course of psychoeducation. Personal Individ Differ 35:691–701

Walters GD (2005) Recidivism in released lifestyle change program participants. Crim Justice Behav 32:50–68

Walters GD (2009) Latent structure of a two-dimensional model of antisocial personality disorder: construct validation and taxometric analysis. J Personal Disord 23:647–660

Walters GD (2011) Criminal thinking as a mediator of the mental illness-prison violence relationship: a path analytic study and causal mediation analysis. Psychol Serv 8:189–199

Walters GD (2012a) Crime in a psychological context: from career criminals to criminal careers. Sage, Thousand Oaks

Walters GD (2012b) Criminal thinking and recidivism: meta-analytic evidence on the predictive and incremental validity of the psychological inventory of criminal thinking styles (PICTS). Aggress Violent Behav 17:272–278

Walters GD (2012c) Substance abuse and criminal thinking: testing the countervailing, mediation, and specificity hypotheses. Law Human Behav 36:506–512

Walters GD (in press a) Cognitive mediation of crime continuity: a causal mediation analysis of the past crime-future crime relationship. Crime Delinq

Walters GD (in press b) Relationships between race, education, criminal thinking, and recidivism: moderator and mediator effects. Assessment

Walters GD, Elliott WN, Miscoll D (1998) Use of the psychological inventory of criminal thinking styles in a group of female offenders. Crim Justice Behav 25:125–134

Walters GD, Hagman BT, Cohn AM (2011) Toward a hierarchical model of criminal thinking: evidence from item response theory and confirmatory factor analysis. Psychol Assess 23:925–936

Walters GD, White TW, Denney D (1991) The lifestyle criminality screening form: preliminary data. Crim Justice Behav 18:406–418

Yochelson S, Samenow SE (1976) The criminal personality: vol. 1. A profile for change. Aronson, New York

Lifestyle Theory of Crime

► Lifestyle Theory

Linkage Analysis for Crime

Craig Bennell[1], Jessica Woodhams[2] and
Rebecca Mugford[1]
[1]Department of Psychology, Carleton University,
Ottawa, ON, Canada
[2]School of Psychology, University of
Birmingham, Birmingham, UK

Synonyms

Behavioral linkage analysis; Comparative case analysis; Crime linkage analysis

Overview

Police investigators must often determine whether multiple crimes have been committed by the same offender. In ideal situations, this decision is based on an analysis of physical evidence left at crime scenes, such as DNA, fabric fibers, and/or fingerprints. However, despite what is portrayed in the popular media, such evidence is not always available to be processed (Davies 1991). Given this, the police have had to establish alternative methods for linking serial crimes. One of the most commonly used approaches is behavioral linkage analysis.

When using this form of analysis, an attempt is made to link crimes based on the behaviors that offenders engage in while committing their offenses. Specifically, the goal is to identify patterns of behavior across an offender's crimes that meet two criteria: behavioral stability and behavioral distinctiveness (Canter 1995). Behavioral stability exists when offenders behave in the same or similar way across their crime series (i.e., high levels of within-series similarity). Behavioral distinctiveness exists when the behaviors exhibited by one serial offender are different from those exhibited by other offenders committing similar types of crimes (i.e., low levels of between-series similarity). When offenders behave in a relatively stable and distinct fashion, it may be possible to link them to their crimes and to differentiate between crimes committed by different offenders (Bennell et al. 2009).

In the investigative setting, linkage analysis is most often carried out by crime analysts or police officers who have specialized training. Generally speaking, there are two ways that the linking task is approached (Woodhams et al. 2007). Proactive linking involves attempts to determine whether a new crime series can be identified by examining the similarities and differences that exist between unsolved crimes archived in a large database. In contrast, reactive linking typically involves attempts to determine whether unsolved crimes can be linked to a particular offense of interest to the police (it may also involve attempts to determine whether a specified set of crimes are the work of a common offender). In some cases, the perpetrator of an index crime may already be known to the police, and the task is to determine whether any unsolved crimes are the responsibility of that particular offender.

Although there is no one method for conducting behavioral linkage analysis, many approaches consist of the following steps: (1) searching for crimes that share similar behavioral features; (2) isolating all the similarities and differences between the crimes that are identified; (3) evaluating the importance of those similarities and differences by, for example, considering the base rates of the behaviors within larger samples of offenses (given that frequently occurring behaviors are unlikely to be useful for distinguishing between crimes committed by different offenders); (4) determining the likelihood of actual crime linkages; and (5) reporting the results of the analysis to investigators (Woodhams et al. 2007).

Historical Developments

The idea that an offender's behavior can be useful for linking crimes is not new. In Gross's (1906) classic book, *Criminal Investigation*, he highlighted the potential value in using an offender's modus operandi, or MO, to link his crimes. Even before this, police agencies in England and Wales had begun developing sophisticated systems for categorizing MOs for the specific purpose of linking crimes (Fosdick 1915). For example, around the turn of the century, Chief Constable William Atcherley of the West Riding Yorkshire Constabulary developed a coding scheme for burglary MOs that increased the ease with which these crimes could be compared to one another. This system was further refined in North America to allow for more detailed comparisons (Vollmer 1919).

Such coding systems, which were used throughout the early 1900s, assumed that offenders exhibit behavior in a highly stable fashion across their crimes. However, thinking around this issue gradually changed. Investigators began to realize that an offender's MO can vary across his crimes for a number of reasons, including learning, maturation, and situational factors (Douglas and Munn 1992). This led to a search for crime features that would remain more stable over time. Law enforcement professionals began to distinguish between three related but distinct constructs: MO, ritual, and signatures (Hazelwood and Warren 2003). MO refers to functional behaviors that are required to successfully commit a crime. Ritual refers to fantasy-based behaviors, which are symbolic in nature and reflect the psychological needs of an offender. Finally, signatures refer to combinations of behaviors (MO or ritual) that are assumed to be relatively stable and unique to each offender. Despite the lack of empirical research to support their use, by the 1990s, linking crimes based on behavioral signatures became a popular approach (e.g., Keppel 1995).

In order to systematize the analysis of an offender's crimes, including their MO and ritual behaviors, police professionals have historically used charts that allow them to compare common and distinctive features exhibited across a set of crimes. While such charts are still commonly used today (Burrell and Bull 2011), sophisticated computer databases are also sometimes relied upon to track and analyze the behaviors of offenders (Collins et al. 1998). One of the first systems designed for this purpose was the Violent Crime Apprehension Program (ViCAP), which was developed by the Federal Bureau of Investigation for the purpose of enabling cross-jurisdiction crime linkages (Howlett et al. 1986). With the cooperation of police investigators from across the country who provide ViCAP with detailed information about the crimes they are investigating, this system allows the ViCAP team to organize, search, and analyze offense-related data in an attempt to identify patterns across crimes that may indicate the presence of serial offenders. Since the development of ViCAP, other systems have also been developed. Most notable among these is the Violent Crime Linkage Analysis System (ViCLAS). Constructed by the Royal Canadian Mounted Police in the mid-1990s (Collins et al. 1998), ViCLAS is currently being used by police forces around the world in an attempt to link serial crimes and is generally considered the gold standard for linkage systems.

Current Research

With the advent of standardized data collection protocols came the possibility of conducting empirical research to examine the degree to which behavioral information can be used to reliably link crimes to the same offender. While this field of research is still relatively new, the number of studies in this area has grown rapidly over the last decade. Existing studies examine a wide range of issues.

Tests of the Linking Assumptions

The majority of linking research conducted to date has attempted to determine whether the assumptions of behavioral stability and distinctiveness (and thus linking) are empirically supported. Many studies in this area use the same basic procedure: (1) the researcher categorizes crime scene behaviors into domains based on their particular function (e.g., wearing a mask is assigned to a "planning" domain in cases of robbery); (2) behavioral similarity scores are calculated for each domain for pairs of crimes that have either been committed by the same offender (linked pairs) or different offenders (unlinked pairs); (3) a statistical procedure, such as logistic regression analysis, is then used to determine the degree to which these similarity scores discriminate between the two types of crime pairs; and (4) linking accuracy is quantified using a measure known as the area under the curve (*AUC*), derived from receiver operating characteristic analysis, which typically varies from 0.50 (chance accuracy) to 1.00 (perfect accuracy). In general, research of this type provides support for the assumptions underlying linkage analysis in crimes ranging from serial burglary (e.g., Bennell and Canter 2002) to serial homicide (e.g., Melnyk et al. 2011), with *AUC* values in the range of .60–.90 being frequently reported. Evidence is even emerging that it is possible to use certain types of behavioral information (e.g., spatial and temporal information) to accurately link offenses that fall into different crime categories (e.g., a residential burglary and a sexual assault committed by the same offender; Tonkin et al. 2011).

Using Behavioral Themes to Link Crimes

Rather than assigning behaviors to domains based on their particular function, other researchers have examined the linking assumptions by using clustering procedures (e.g., multidimensional scaling) to group crime scene behaviors into psychological themes (e.g., hostility). These themes then form the basis for further analysis. For example, Santtila et al. (2008) used Mokken scaling, which is similar to factor analysis, to derive behavioral themes (or dimensions) from the crime scene behaviors of Italian murderers. Using scores derived from seven dimensions as independent variables, and the series an offense belonged to as the dependent variable, they were able to use discriminant function analysis to correctly assign 62.9 % of crimes to the correct series. While lower accuracy rates have been reported for other types of crimes (e.g., serial arson; Santtila et al. 2004a), even these studies suggest that linkage accuracy exceeds levels that would be expected by chance.

Identifying Factors That Influence Linkage Accuracy

Many studies have also focused on identifying factors that might influence the accuracy with which crime linkages can be identified. While a range of factors have been explored (e.g., the type of similarity coefficient used to measure across-crime similarity; Melnyk et al. 2011), most research has examined whether certain crime scene behaviors (or themes) are more useful for linking purposes. Recent findings suggest that this is the case. For example, crime scene behaviors which are “offender-driven” appear to be exhibited in a more stable and distinct fashion by offenders (and are thus better predictors of whether crimes are linked) than behaviors that are more “situation-driven.” In cases of serial burglary, for instance, the distance between crime site locations (which is obviously determined by where an offender decides to commit their crimes) can be a very effective linking cue, whereas the type of property stolen by an offender (which depends on what is available to be stolen) tends to lack high levels of predictive accuracy (Bennell and Canter 2002). Similar conclusions can be reached from studies of linkage analysis that have used the thematic approach. In Grubin et al. (2001) study of serial sexual assaults, for instance, behaviors related to a “control” theme (e.g., having a weapon, which would appear to be planned in advance of the crime) were exhibited more consistently than behaviors related to a “style” theme (e.g., taking money from the victim, which depends on situational factors).

Understanding the Complexities of Linking Decisions

Another body of research has investigated some of the complexities involved in linkage analysis by examining how decision-makers perform on laboratory-based linking tasks. For example, Santtila et al. (2004b) examined the performance of experienced car crime investigators, experienced general investigators, novice general investigators, and naive participants on a mock linking task using car crime data. They found that, while investigators were significantly more accurate than naive participants, there were no differences between the different types of investigators in terms of their linking accuracy, with each group identifying about half of all possible links (although experienced car crime investigators did rely on significantly less information than every other group to make their decisions). The reasons why participants in these types of studies experience problems with identifying crime linkages are still not well understood, but they likely have to do with difficulties in selecting appropriate linking cues and in adequately processing this information (Bennell et al. 2010). It may also be the case that participants who have been examined in these types of studies lacked the necessary training or experience in linkage analysis to perform well on the tasks or that the tasks themselves were inappropriately designed.

The Development of Computerized Linking Algorithms

In contrast to the research just described, studies have also explored whether more "mechanical" approaches can be used to identify crime linkages, which is potentially important given the problems that people seem to encounter when faced with this task. In addition to various statistical approaches that have been examined (e.g., Bennell and Canter 2002), a variety of computer algorithms have also been tested (e.g., Yokota and Watanabe 2002). Studies indicate that these algorithms have the potential to automatically link crimes, sometimes with a high degree of accuracy, and that these procedures are associated with a number of advantages over other, more subjective, procedures. For example, some of the tested algorithms can weight behavioral similarity scores between crimes, such that information which is more prone to errors is given less weight (Brown and Hagen 2002). While it is still too early to make recommendations for how these computer algorithms (or similar tools) should be used in investigative settings, they may be able to effectively support the decisions currently being made by linkage analysts.

Examining Design Decisions from Previous Linking Studies

Finally, much of the research being conducted recently examines the impact of potentially problematic design decisions that have been made in previous studies. For example, most studies conducted on linkage analysis have relied on samples of solved serial crimes because, in order to determine linking accuracy, researchers need to know which crimes in their sample are actually linked. While this makes sense, designing studies in this way can bias results if the sampled crimes were originally linked and solved because they were characterized by highly similar or distinct MOs (Bennell and Canter 2002). If that were the case, levels of accuracy reported in these studies may overestimate what is actually possible when linkage analysis is applied to unsolved crimes. Fortunately, recent research that has examined linking accuracy for crimes first linked by MO versus DNA has indicated that similar levels of accuracy are observed, suggesting that there may be no need to worry about previous results (Woodhams and Labuschagne 2012). Similar attempts have been made to explore the impact of other design decisions, such as the common but questionable practice in most linking studies of excluding non-serial offenses from samples.

Remaining Challenges

Despite the growth in studies that has occurred over the last decade, challenges still remain, both for researchers working in this area and for practitioners who conduct linkage analysis.

Dealing with Problematic Data

Some of the most difficult challenges faced by both researchers and practitioners relate to issues around data quality (Burrell and Bull 2011; Woodhams et al. 2007). In both research and practice, linking decisions tend to be based on certain sources of data – typically victim statements or crime reports – that can be highly problematic. For example, with respect to victim statements, victims can forget what occurred during a crime, they can be reluctant to talk about certain events, and reports of what they say can be distorted by the police (Alison et al. 2001). Crime reports, on the other hand, are often plagued with missing data, and certain pieces of information contained within them can be highly unreliable (e.g., the time that a burglary occurred). Any research results that emerge from such data sources, or any linkage decisions that are based on them, must obviously be viewed with caution, though improvements to data coding and recording protocols can improve the situation.

This type of data can also be very limited with respect to what it reveals about offenders and the crimes they commit, which can prevent, or at least hinder, useful lines of research or inquiry. For example, in other research contexts where high-quality data can be collected, it has been discovered that individual differences in noncriminal behavior are more stable across "psychologically similar" situations (e.g., Shoda et al. 1994). This is clearly an important finding that might have relevance to the linking context. Unfortunately, despite the efforts by some (e.g., Woodhams et al. 2008), it is difficult to examine these issues within the investigative context. It is typically not possible to observe crimes taking place, and data sources like victim statements do not tend to include detailed information about situational factors. This prevents researchers from examining potentially interesting (and useful) issues. For the linkage analyst, such problems limit their ability to understand how situations influence offenders and to interpret the meaning or significance of specific behaviors, which is often necessary to accurately link crimes.

Conducting (and Getting Access to) Valid Research

A related challenge for researchers in this area is to conduct studies which produce results that generalize to real investigations (i.e., externally valid studies). The challenge for practitioners is to make sound linking decisions in the absence of externally valid linking research. The primary issue for researchers is that attempts to improve the external validity of their studies can sometimes make the research more challenging to conduct. For example, in order to increase the validity of studies concerned with linking accuracy, researchers might want to include serial and non-serial crimes in their samples, rely on serial crimes that have been linked initially by DNA, and include every crime from each offender's series. The problem is that it is difficult, though not impossible (e.g., Woodhams and Labuschagne 2012), to get access to such data. Unfortunately, the further studies stray away from these sorts of samples, the more the generalizability of results will become a serious concern.

Given these issues, practitioners often face challenges in applying existing research to linking tasks. While the sorts of sampling issues described above are certainly part of the problem, there are additional (potentially more serious) issues that need to be addressed. Indeed, some practitioners argue that much of what is studied by researchers in this area is largely irrelevant for decisions that analysts routinely have to make, leading to calls for research that is more pragmatic in nature (Alison and Rainbow 2011). For example, most research to date has examined proactive linking, where large samples are searched, crime linkages are established, and accuracy is determined. However, analysts often encounter reactive tasks where they are asked to determine whether a set of offenses, thought to be connected by an investigator, are likely to be linked (Rainbow in press). At best, research on proactive linking tasks is relevant here and researchers have simply not made this clear. At worst, a large body of research that has been conducted is of little use in assisting analysts on common tasks that they encounter.

Withstanding Legal Scrutiny

Increasingly, linkage analysis is finding its way into court when questions are raised about whether a defendant is responsible for multiple crimes (Labuschagne 2006). This presents challenges for both researchers and practitioners. While practitioners will likely be the ones presenting at trial, for their testimony to be admitted in many jurisdictions good linking studies will be required. Indeed, although standards of admissibility differ across jurisdictions, at least some rely on criteria that relate directly to research. For example, beyond having to establish expertise and demonstrate that testimony is relevant and useful (i.e., goes beyond the common understanding of the judge or juror), if one were to apply the *Daubert* criteria from the USA to any sort of linkage analysis, issues related to its reliability would also come up (Woodhams et al. 2007). Questions about the testability of the linking assumptions would likely be asked, in addition to other research-related questions pertinent to this issue (e.g., have the assumptions been tested, have the results of those tests been adequately published). It is doubtful that all the *Daubert* criteria would be met at the moment (Woodhams et al. 2007). Whether they are in the future will depend, to some extent, on the type of research that is conducted and published. In this way, researchers have a role to play in developing research that can allow linkage analysis to withstand legal scrutiny.

For practitioners, there are two issues they could potentially speak to at trial – behavioral similarities that exist across the crimes in question and/or how peculiar (distinct) the similar behaviors actually are (Ormerod 1999). As just discussed, the primary challenge for practitioners will be getting their evidence admitted. Usefulness might prove to be a controversial issue. Despite evidence to the contrary in some jurisdictions (e.g., Labuschagne 2006), it is debatable whether testimony related to similarities would be commonly heard by the courts given that jurors and judges can arguably perceive similarity for themselves (Ormerod 1999). However, evidence related to distinctiveness might be viewed as more useful. Issues of reliability could also be contentious. The practitioner will have to clearly demonstrate that distinctiveness exists and convince the court that their testimony is sound. Currently, there appears to be no standard (i.e., commonly accepted) protocol for measuring distinctiveness, and there is insufficient published research on this particular topic, especially as it relates to reactive linking tasks. While the increased use of crime linkage systems, such as ViCLAS, will assist in establishing base rates of behavior, the reliability of the data contained within these systems will have to be confirmed.

Related Entries

- ▶ Behavioral Investigative Advice
- ▶ Criminal Investigative Analysis
- ▶ Investigative Psychology

Recommended Reading and References

Alison LJ, Rainbow L (eds) (2011) Professionalizing offender profiling: forensic and investigative psychology in practice. Routledge, London

Alison LJ, Snook B, Stein KL (2001) Unobtrusive measures: using police information for forensic research. Qual Res 1:241–254

Bennell C, Bloomfield S, Snook B, Taylor PJ, Barnes C (2010) Linkage analysis in cases of serial burglary: comparing the performance of university students, police professionals, and a logistic regression model. Psychol Crime Law 18:507–524

Bennell C, Canter DV (2002) Linking commercial burglaries by modus operandi: tests using regression and ROC analysis. Sci Justice 42:153–164

Bennell C, Jones NJ, Melnyk T (2009) Addressing problems with traditional crime linking methods using receiver operating characteristic analysis. Leg Criminol Psychol 14:293–310

Brown DE, Hagen S (2002) Data association methods with application to law enforcement. Decis Support Syst 34:369–378

Burrell A, Bull R (2011) A preliminary examination of crime analysts' views and experiences of comparative case analysis. J Gov Inf 13:2–15

Canter DV (1995) Psychology of offender profiling. In: Bull R, Carson D (eds) Handbook of psychology in legal contexts. Wiley, Chichester, pp 345–355

Collins PI, Johnson GF, Choy A, Davidson K, MacKay RE (1998) Advances in violent crime analysis and law enforcement: Canadian violent crime linkage analysis system. J Gov Information 25:277–284

Davies A (1991) The use of DNA profiling and behavioral science in the investigation of sexual offences. Medic Sci Law 31:95–101

Douglas JE, Munn C (1992) Violent crime scene analysis: modus operandi, signature, and staging. FBI Law Enforce Bull 61:1–10

Fosdick RB (1915) The modus operandi system in the detection of criminals. J Am Inst Crim Law Criminol 6:560–570

Gross H (1906) Criminal investigation. G. Ramasawmy Chetty, Masdras

Grubin D, Kelly P, Brunsdon C (2001) Linking serious sexual assaults through behaviour. Home Office, London

Hazelwood RR, Warren JI (2003) Linkage analysis: modus operandi, ritual, and signature in serial sexual crime. Aggress Violent Behav 8:587–598

Howlett JB, Hanfland KA, Ressler RK (1986) The violent criminal apprehension program: a progress report. FBI Law Enforce Bull 55:14–22

Keppel RD (1995) Signature murders: a report of several related cases. J Forensic Sci 40:670–674

Labuschagne G (2006) The use of linkage analysis as evidence in the conviction of the Newcastle serial murderer, South Africa. J Investig Psychol Offender Profiling 3:183–191

Melnyk T, Bennell C, Gauthier D, Gauthier D (2011) Another look at across-crime similarity coefficients for use in behavioural linkage analysis: an attempt to replicate Woodhams, Grant, and Price (2007). Psychol Crime Law 17:359–380

Ormerod D (1999) Criminal profiling: trial by judge and jury, not criminal psychologist. In: Canter DV, Alison LJ (eds) Profiling in policy and practice. Ashgate, Aldershot, pp 207–261

Rainbow L (in press) A practitioner's perspective: theory, practice and research. In: Woodhams J, Bennell C (eds) Crime linkage: theory research and practice. CRC Press, Boca Raton

Santtila P, Fritzon K, Tamelander AL (2004a) Linking arson incidents on the basis of crime scene behavior. J Police Crim Psychol 19:1–16

Santtila P, Korpela S, Häkkänen H (2004b) Expertise and decision-making in the linking of car crime series. Psychol Crime Law 10:97–112

Santtila P, Pakkanen T, Zappalá A, Bosco D, Valkama M, Mokros A (2008) Behavioural crime linking in serial homicide. Psychology, Crime Law 14:245–265

Shoda Y, Mischel W, Wright JC (1994) Intra-individual stability in the organization and patterning of behavior: incorporating psychological situations into the idiographic analysis of personality. J Pers Soc Psychol 67:674–687

Tonkin M, Woodhams J, Bull R, Bond JW, Palmer EJ (2011) Linking different types of crime using geographical and temporal proximity. Crim Justice Behav 38:1069–1088

Vollmer A (1919) Revision of the Atcherley modus operandi system. J Am Inst Crim Law Criminol 10:229–274

Woodhams J, Bull R, Hollin CR (2007) Case linkage: identifying crimes committed by the same offender. In: Kocsis RN (ed) Criminal profiling: international theory, research, and practice. The Humana Press, Totowa, pp 117–133

Woodhams J, Hollin CR, Bull R (2008) Incorporating context in linking crimes: an exploratory study of situational similarity and if-then contingencies. J Invest Psychol Offender Profiling 5:1–23

Woodhams J, Labuschagne G (2012) A test of case linkage principles with solved and unsolved serial rapes. J Police Crim Psychol 27:85–98

Yokota K, Watanabe S (2002) Computer-based retrieval of suspects using similarity of *modus operandi*. Int J Police Sci Manag 4:5–15

Longitudinal

► Establishing Causes of Offending in Longitudinal and Experimental Studies

Longitudinal Crime Trends at Places

Elizabeth Griffiths[1] and Jorge M. Chavez[2]
[1]School of Criminal Justice, Rutgers University, Newark, NJ, USA
[2]Department of Sociology, Bowling Green State University, Bowling Green, OH, USA

Synonyms

Latent growth curves; Neighborhood crime trends; Time series; Trajectories

Overview

The role of space and place has a long tradition in American criminology largely germinating from the ground-breaking research of Shaw and McKay (1942). Yet by the 1960s and 1970s, criminological attention had turned almost wholly to individual-level causes of crime. Over the past three decades, however, researchers have rediscovered the central role of communities in the causation and control of crime. Like people, communities have a criminal history or “criminal career”; they experience relatively more or less criminal activity than both other places in the city and compared to their own levels at earlier periods in time. To the extent that the natural history of a community affects its current crime rates, its reputation for violence, and its projected levels and patterns of crime and violence in the future, the idea that a community may follow a specific type of trajectory, or “career” may be more theoretically useful to criminologists than are cross-sectional or static patterns across space. This entry describes the theoretical foundations of the literature on longitudinal crime trends at places, provides an abbreviated overview of various methodological approaches to modeling community crime trajectories, elaborates on some key findings, considers the conceptual implications of the unit-of-analysis (or what is meant by “place”), and ends with a discussion of both thorny issues and future directions for research.

Introduction

If it is true that the whole is greater than the sum of its parts, then a key to the success of any social group is its ability to build and sustain a sense of community – to generate a capacity for realizing the shared interests, actions, and fortunes of its members (Sampson and Groves 1989). The criminological literature on space and place primarily is concerned with investigating how one’s residential environment either creates vulnerabilities to risk of victimization or provides conditions conducive to offending among the already predisposed. Much scholarly work has identified the risk factors associated with crime in space, including: (1) the social and economic characteristics related to social disorganization such as poverty, family structure, residential mobility, ethnic and racial heterogeneity, and urbanization (Shaw and McKay 1942; Sampson and Groves 1989); (2) private, parochial, and public social controls (Bursik and Grasmick 1993; Carr 2003); (3) risky terrains rife with concentrated public housing, bus stops, liquor stores, leisure establishments, bars, and other entertainment venues (Caplan et al. 2011; Felson 1998; Griffiths and Tita 2009); and (4) gang territory or drug market locales (Tita and Ridgeway 2007), among others. These features affect the likelihood of criminal events occurring within neighborhoods.

Until recently, however, the neighborhood-effects literature treated local characteristics as if they were relatively static or fixed. That is, the presence, absence or level of various population and land use features were used to predict the distribution of crime across urban space. Despite having much success in explaining cross-sectional aggregate crime patterns, this approach fundamentally ignores one of the basic lessons of the early disorganization scholars: communities change over time, either as a consequence or independently of the outcome under study. For this reason, it is important to consider the joint spatial and temporal aspects of communities or, said differently, longitudinal crime trends at places.

In 1986, Reiss and Tonry edited a volume of *Crime and Justice: A Review of Research* entitled *Communities and Crime*. This volume emerged prior to what would become a plethora of studies on social disorganization and neighborhood effects over the succeeding decades. In anticipation of this scholarly interest, the authors of *Communities and Crime* prepared thoughtful and important essays on the dynamics of crime across urban space, focusing on issues of stability, change, and community careers in delinquency, crime, and violence (Reiss and Tonry 1986). Unfortunately, their call for attention to the dynamics of community crime trends went largely unheralded. While the neighborhood-effects literature mushroomed (Sampson et al. 2002), it remained predominantly targeted at difference in crime across space rather than changes in crime at places over time. Criminologists are certainly much better prepared to discuss the role of communities in crime causation as a consequence of the neighborhood-effects literature, but our understanding of how crime changes over time across place remains in its infancy.

The remainder of this entry is organized as follows. The first section describes the central theoretical arguments for explaining the dynamics of crime in space. In the second section, three common methodologies that have been employed to capture crime trends in empirical models are briefly outlined. The third section provides a short review of the literature on community trends in delinquency, crime, and violence. This is followed by a section that problematizes the plethora of “places” operationalized by researchers. Finally, the entry ends with a consideration of a number of data problems confronting scholars of neighborhoods and crime and touches upon anticipated directions of future research.

Theoretical Fundamentals

It was at the Chicago School that Clifford Shaw and Henry D. McKay first studied the distribution of delinquency across the city and pioneered one of the most important ideas in criminology of the twentieth century. In brief, Shaw and McKay (1942) observed that neighborhoods in the transitional zone – or in neighborhoods contiguous to but immediately outside of the central city – repeatedly represented the highest rates of juvenile delinquency over more than three decades (1900–1933), irrespective of the particular ethnic group residing in these places at any given period. In effect, then, rates of delinquency must be explained as a consequence of the conditions of the community rather than the personal characteristics of the persons who reside there. This finding, which undermines the premise that specific individuals or groups have an inherent predisposition toward antisocial behavior, was possible precisely because Shaw and McKay (1942) found stability in delinquency rates over a period in which extensive population change took place. It was the longitudinal nature of their research, in particular, that set the stage for this key theoretical breakthrough.

Despite its import, the heyday of social disorganization gave way as scholars turned toward individual-level explanations for offending behavior by the late 1950s and certainly the 1960s. A quarter of a century later, however, interest in places and crime trends reemerged. For example, Scheuerman and Kobrin (1986) explored the concept of stability and change in the crime rates of dangerous communities in Los Angeles and, in doing so, introduced the idea that communities have "careers" in crime in much the same way that individuals do. Specifically, they found that high-crime neighborhoods followed three distinct trends over two decades: emerging, transitional, and enduring. Moreover, the relationship between neighborhood deterioration and crime is a reciprocal one, with deterioration preceding initial spikes in crime but prompting additional deterioration once crime rates have reached high and stable levels. Scheuerman and Kobrin were focused on changes in social, economic, and family population characteristics of communities whereas both Wilson and Kelling (1982) and Skogan (1990) considered the potential reciprocity between crime and signs of neighborhood incivilities including social (panhandling, public intoxication, etc.) and physical (graffiti, broken windows, etc.) forms of disorder. According to this logic, places that appear to be uncared for by neighborhood residents and places that are void of private, parochial, and public controls are vulnerable to criminal invasion. The "spiral of decline" that ensues is inherently a temporal process (Skogan 1990).

While not focused specifically on the communities and crime question at its inception, it is noteworthy that Cohen and Felson's (1979) Routine Activity theory also centers on how large-scale social changes influence the dynamics of crime over time. The routine activities of everyday life, including the work and home activities that shape the tempo, timing, and guardianship capabilities of communities, help to explain aggregate crime trends. At its core, then, place-based crime trends are a reflection of broad social and economic transformations that can be tracked and modeled over time. Like social disorganization theory, the central premise of routine activities theory is rooted in an examination of crime rate trends in places, rather than static or cross-sectional levels.

Methodological Diversity

A series of methodological innovations make examining crime trends at places possible. Some of the first studies focused on stability and volatility in community crime relied on cross-lagged correlations or residual change scores to decompose the degree of change between *t* and *t*-1 that could not have been predicted on the basis of levels at time *t*-1 (Bursik 1986). For example, Bursik (1986) compares change score models to cross-sectional models predicting community delinquency rates in Chicago at each decade between 1930 and 1970. These comparisons lead him to conclude that, "without testing the assumption of stability, ecological models will not only have a limited degree of theoretical power, but they will have an interpretive framework that is generally unrelated to the dynamics of modern urban areas" (Bursik 1986, p. 59). That is, static models ignore both expected and unexpected ecological changes that are entirely consistent with dominant theoretical explanations for community crime rates. A key limitation of these early methodological approaches is that they are "relatively cumbersome when multiple waves of data are analyzed and, for this reason, applications of such techniques have generally focused on a series of two-wave trends, providing a truncated sense of neighborhood change" (Kubrin and Herting 2003, p. 332).

More recent approaches to modeling neighborhood change and associated trends in crime rates utilize growth-curve modeling and/or semi-parametric group-based trajectory procedures. Growth-curve regression models take advantage of information about the individual slopes of each geographical unit-of-analysis under study to predict how various independent variables affect changes or trends in the dependent variable (Kubrin and Herting 2003; Braga et al. 2010, 2011). The advantage of such an approach is related to the fact that no trend information is lost in the estimation of the models. Alternatively, some scholars have exploited the capacity of Nagin's (1999) semi-parametric group-based trajectory procedure to reduce the temporal homicide rates of *N* census tracts, street segments, or other places into a smaller number of latent classes or groups exhibiting similar violence levels and following like-trends over some period of time (Griffiths and Chavez 2004; Groff et al. 2010; Weisburd et al. 2004). The resulting number and shapes of place-based homicide trajectories ostensibly capture the various types of place-based "criminal careers."

Key Findings

This section provides a brief (and, of necessity, incomplete) overview of a few of the key findings in the longitudinal crime trends at places literature. Much of the research focused on crime trends across place has modeled longitudinal homicide rates – sometimes disaggregating by victim-offender relationship, motive, or weapon type – in part due to data availability and quality issues (Griffiths and Chavez 2004; Jennings and Piquero 2008; Kubrin and Herting 2003; Stults 2010). For example, in using growth-curve models to capture-distinctive trends in general altercation, felony, and domestic homicide trends in St. Louis neighborhoods, Kubrin and Herting (2003) find that disaggregation of homicide type is imperative as neighborhoods vary in temporal trends by homicide type. In particular, various aspects of neighborhood structure are related to both the amount but also the nature of neighborhood homicide. Griffiths and Chavez (2004) likewise demonstrate that Chicago neighborhoods exhibiting a high and increasing gun homicide trend between 1980 and 1995 have marked unobserved heterogeneity in their criminal careers (i.e., in their total homicide trajectories). Some of these neighborhoods are characterized by a high and unstable homicide trajectory driving citywide trends over the period, as one might expect. Yet four times as many neighborhoods following a high and increasing gun homicide trend are characterized by a moderate and relatively stable total homicide career over the same period, relative to other neighborhoods in the city. Cumulatively, these studies provide substantial evidence that place-based patterns in homicide type reveal considerable variability and instability both over time and across space.

Research at the county-level also illustrates the importance of place in understanding temporal homicide patterns. Jennings and Piquero (2008) show, for example, that rural counties are significantly more likely to exhibit non-declining intimate-partner homicide trends between 1980 and 1999, compared to their urban counterparts. And McCall et al. (2011) have recently extended this literature to examine temporal changes in the homicide rates of US cities over 30 years. Not only are cities distinguished by multiple and varied homicide trajectories, but disadvantage and other characteristics associated with social disorganization more generally (McCall et al. 2011; see also Land et al. 1990) predict membership in the consistently more violent city groupings compared to those cities that are members of the lower and more stable latent classes.

The literature on longitudinal crime trends at places is not limited to homicide research. For example, Weisburd et al. (2004) find that a very small group of street segments in Seattle are responsible for generating volatility in the city-wide pattern of total criminal incident reports to police, as the vast majority of street segments follow relatively stable trends between 1989 and 2002. Studies employing trajectory or growth-curve models for crime trends at places have also measured juvenile arrests (Weisburd et al. 2009), robbery rates (Braga et al. 2011), and disorder (Yang 2010), among others. The bulk of the literature on trends in all types of criminal events at most aggregate micro (e.g., street segment, face block, intersection, etc.) and macro (e.g., census tracts, cities, counties, etc.) units-of-analysis has repeatedly shown that very few places are responsible for high rates of crime and violence, that these places tend toward volatility rather than stability in trends over time, and that the nature of the events under study (e.g., gun violence, intimate versus non-partner victim-offender relationship, age of offender, etc.) affect levels, trends, and the distribution of incidents both across places and over time.

A Plethora of "Places"

The neighborhood-effects literature is rightfully concerned with what geographers call the Modifiable Areal Unit Problem – or the extent to which the unit-of-analysis used by the researcher as an operationalization of neighborhoods influences the findings of the study (Openshaw and Taylor 1979). Scholars interested in understanding neighborhood levels and changes in violence have tended to use census tracts as an approximation of neighborhoods (Griffiths and Chavez 2004; Hipp 2007; see Fagan 2008) while those focused on the emergence and stability of hot spots in micro-places have elected to use even smaller aggregates, such as street segments or block faces (Braga et al. 2010; Weisburd et al. 2004, 2010). Any unit-of-analysis is associated with both advantages and disadvantages, some of which are related to data availability (particularly over time), the spatial units at which policing agencies release information on criminal incidents within their jurisdictions, census units required to access structural characteristics of place, and size of the residential population or daily population flows into and out of the physical space.

One consideration in selecting the geographical unit of analysis is the extent to which the findings at one level are either robust across levels of analyses or almost wholly dependent upon the researchers' definition of place (Hipp 2007). The problem is this: while findings that can be replicated at multiple aggregate units are arguably more robust than findings which vary across aggregations, robustness across units that bear little relation to one another or to what we mean by "neighborhood" suggests that the concept of neighborhood is relatively unimportant. It is not the extent to which the same findings can be replicated at different operationalizations that should matter, so much as the fit between the operationalization of neighborhoods in empirical research and neighborhoods on-the-ground, as identified and understood by residents.

The longitudinal crime trends at places literature is among the most open to the use of extremely varied aggregate units. This is, in part, because not all research in this area is concerned with neighborhood crime trends. Specifically, Braga and Weisburd (2010, p. 2) note that "micro-places" are "specific locations within the larger environments of communities and neighborhoods…[] sometimes defined as buildings or addresses, sometimes as block faces or street segments, and sometimes as clusters of addresses, block faces, or street segments." It is in these micro-units that crime "hot spots" emerge. St. Jean (2007) makes the same point in his ethnographic study of a high-crime neighborhood in Chicago. He finds that very few streets within this community are the sites of crime and violence. The same result is found by Groff and colleagues (2010) in their quantitative study of street segments in Seattle, Washington.

In this case, Groff et al. (2010) establish that there is considerable variability in the 16-year crime trajectories of street segments that are adjacent to one another; consequently, they argue that this heterogeneity in the criminal careers of very small geographic units which are situated in close spatial proximity reflects "the importance of looking at the micro level" (Groff et al. 2010, p. 26).

The openness of scholars in the longitudinal crime trends at places literature to using various different spatial aggregation of the unit-of-analysis generates a great deal of information about how crime is situated both in space and over time. Both the hot spots and the neighborhood crime literatures have benefitted from consideration of the size and shape of "places.' At the same time, the opportunity to cherry-pick the size of the unit under study may necessitate careful theoretical unpacking. For example, how many street segments comprise a neighborhood? How is living beside a street segment following a high-crime trajectory different from living far away from one or any? Are any of the census classifications (e.g., tracts, blocks, etc.) reasonable proxies for neighborhoods? Does the concept of neighborhood really matter? What do we mean by "place" and how do places of varying sizes, shapes, and histories change the empirical picture?

Methodological Considerations and Future Directions

The longitudinal-crime-trends-at-places literature is an exciting field of study that remains largely in its infancy. As a consequence, scholars who study crime trends at places face difficulties related not only to the theoretical issues associated with defining places but also to methodological complications. A continual challenge is the availability of appropriate spatial and longitudinal data. For example, to predict changes in crime trends at places, scholars require data capturing changes in the various independent variables that are associated with aggregate levels of crime. The most widely accessible and most commonly used data on the social and economic conditions of place are compiled by the United States Census, which are available only on a decennial basis for most areas of the country. Land use and other characteristics of place not drawn from the Census can be coded to more fully specify quantitative models predicting crime trends. The problem these scholars face, however, is in explaining *changes* in crime as a consequence of risk factors in the setting (e.g., bars, leisure establishments, public housing) that remain relatively static over 5, 7, or even 10 years. In essence, researchers need to distinguish the "vulnerability" to crime created by these relatively stable criminogenic environmental conditions from "exposure" to crime, captured in unstable local crime trends (Caplan and Kennedy 2011).

In order for researchers to better model and specify how vulnerabilities in the form of local criminogenic conditions can result in different rates of exposure (or changes in crime over time), the gap between theory and data must be addressed. In particular, future researchers need to develop and collect more dynamic indicators of place-based crime and associated behaviors, social and demographic characteristics, land use, and other temporal aspects of place. Researchers need to consider innovative sources of data and more refined measures of key characteristics of places.

Greater attention should also be focused on measuring the intervening mechanisms which link risk and place, and which underlie the process of change in crime at places. Recent research on crime and place has identified a number of potential avenues to consider, such as social capital, collective efficacy, formal and informal social control, culture, and social isolation, among others. A renewed focus on intervening mechanisms requires that researchers integrate spatial and temporal data on these dynamics as a function of place. While developing innovative ways of substantively measuring intervening mechanisms is important, future research must also better specify the proper time lags for capturing change over time, as well as the reciprocal effects of patterns of crime, place-based characteristics, and intervening mechanisms.

Finally, future research needs to consider how the unique spatial characteristics of place create important considerations for change over time. While much of the focus on the dynamics of longitudinal trends in crime at places has centered on temporal change, we must not forget about space. A focus on space means that researchers must aim for a substantive understanding of the environment which surrounds a place, both socially and physically. Researchers need to consider how external factors may influence trajectories of crime and other place-based characteristics. The characteristics of proximate places may directly and indirectly influence levels and changes in crime. Future research should assess the interrelationship between internal rates and changes in both crime and place-based characteristics as well as external (environmental) factors. External influences include, but are not limited to, the structural and demographic characteristics of proximate places, as well as larger economic and political factors. Places may be influenced by their immediate neighbors, local law enforcement, political and economic policies, and decisions which occur well beyond the boundaries of a particular place. Focused attention toward these theoretical and methodological issues will further develop our collective understanding of the meaning of space, place, and time in the criminological literature.

Related Entries

- ▶ Crime and the Racial Composition of Communities
- ▶ Crime Mapping
- ▶ Disadvantage, Disorganization and Crime
- ▶ Group-Based Trajectory Models
- ▶ History of Geographic Criminology Part I: Nineteenth Century
- ▶ Hot Spots and Place-Based Policing
- ▶ Neighborhood Effects and Social Networks
- ▶ Police Legitimacy and Police Encounters
- ▶ Social Network Analysis and the Measurement of Neighborhoods
- ▶ Spatial Perspectives on Illegal Drug Markets

Recommended Reading and References

Braga AA, Weisburd DL (2010) Editor's introduction: empirical evidence on the relevance of place in criminology. J Quant Criminol 26(1):1–6

Braga AA, Papachristos AV, Hureau DM (2010) The concentration and stability of gun violence at micro places in Boston, 1980–2008. J Quant Criminol 26(1):33–53

Braga AA, Hureau DM, Papachristos AV (2011) The relevance of micro places to citywide robbery trends: a longitudinal analysis of robbery incidents at street corners and block faces in Boston. J Res Crime Delinq 48(1):7–32

Bursik RJ Jr (1986) Ecological stability and the dynamics of delinquency. In: Reiss AJ, Tonry M (eds) Communities and crime. Crime and justice: a review of research, vol 8. University of Chicago Press, Chicago

Bursik RJ Jr, Grasmick HG (1993) Neighborhoods and crime: the dimensions of effective community control. Lexington Books, New York

Caplan JM, Kennedy LW (eds) (2011) Risk terrain modeling compendium. Rutgers Center on Public Security, Newark

Caplan JM, Kennedy LW, Miller J (2011) Risk terrain modeling: brokering criminological theory and GIS methods for crime forecasting. Justice Q 28(2):360–381

Carr PJ (2003) The new parochialism: the implications of the Beltway case for arguments concerning informal social control. Am J Sociol 108(6):1249–1291

Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44(4):588–608

Fagan J (2008) Crime and neighborhood change. In: National Research Council Committee on Understanding Crime Trends, Committee on Law and Justice, Division of Behavioral and Social Sciences and Education (ed) Understanding crime trends: workshop report. The National Academies Press, Washington, DC

Felson M (1998) Crime and everyday life, 2nd edn. Pine Forge Press, Thousand Oaks

Griffiths E, Chavez JM (2004) Communities, street guns and homicide trajectories in Chicago, 1980–1995: merging methods for examining homicide trends across space and time. Criminology 42(4):941–978

Griffiths E, Tita G (2009) Homicide in and around public housing: is public housing a hotbed, a magnet, or a generator of violence for the surrounding community? Soc Probl 56(3):474–493

Groff ER, Weisburd D, Yang S-M (2010) Is it important to examine crime trends at a local "micro" level?: a longitudinal analysis of street to street variability in crime trajectories. J Quant Criminol 26(1):7–32

Hipp JR (2007) Block, tract, and levels of aggregation: neighborhood structure and crime and disorder as a case in point. Am Sociol Rev 72(5):659–680

Jennings WG, Piquero AR (2008) Trajectories of non-intimate and intimate partner homicides, 1980–1999:

the importance of rurality. J Crim Justice 36(5): 435–443

Kubrin CE, Herting JR (2003) Neighborhood correlates of homicide trends: an analysis using growth-curve modeling. Sociol Q 44(3):329–350

Land KC, McCall PL, Cohen LE (1990) Structural covariates of homicide rates: are there any invariances across time and social space? Am J Sociol 95(4):922–963

McCall PL, Land KC, Parker KF (2011) Heterogeneity in the rise and decline of city-level homicide rates, 1976–2005: a latent trajectory analysis. Soc Sci Res 40(1):363–378

Nagin DS (1999) Analyzing developmental trajectories: a semi-parametric, group-based approach. Psychol Methods 4(2):139–157

Openshaw S, Taylor PJ (1979) a million or so correlation coefficients: three experiments on the modifiable areal unit problem. In: Wrigley N (ed) Statistical applications in the spatial sciences. Pion, London

Reiss AJ Jr, Tonry M (1986) Communities and crime. Crime and justice: a review of research, vol 8. The University of Chicago Press, Chicago

Sampson RJ, Groves WB (1989) Community structure and crime: testing social disorganization theory. Am J Sociol 94(4):774–802

Sampson RJ, Morenoff JD, Gannon-Rowley T (2002) Assessing 'neighborhood effects': social processes and new directions in research. Annu Rev Sociol 28:443–478

Scheuerman L, Kobrin S (1986) Community careers in crime. In: Reiss AJ, Tonry M (eds) Communities and crime. Crime and justice: a review of research, vol 8. University of Chicago Press, Chicago

Shaw CR, McKay HD (1942) Juvenile delinquency and urban areas. University of Chicago Press, Chicago

Skogan WG (1990) Disorder and decline: crime and the spiral of decay in American neighborhoods. University of California Press, Berkeley

St. Jean PKB (2007) Pockets of crime: broken windows, collective efficacy and the criminal point of view. The University of Chicago Press, Chicago

Stults BJ (2010) Determinants of Chicago neighborhood homicide trajectories: 1965–1995. Homicide Stud 14(3):244–267

Tita G, Ridgeway G (2007) The impact of gang formation on local patterns of crime. J Res Crime Delinq 44(2):208–237

Weisburd D, Bushway S, Lum C, Yang S-M (2004) Trajectories of crime at places: a longitudinal study of street segments in the city of Seattle. Criminology 42(2):283–321

Weisburd D, Morris NA, Groff ER (2009) Hot spots of juvenile crime: a longitudinal study of arrest incidents at street segments in Seattle, Washington. J Quant Criminol 25(4):443–467

Weisburd D, Groff ER, Yang S-M (2010) Understanding developmental crime trajectories at places: social disorganization and opportunity perspectives at micro units of geography. Final report to the National Institute of Justice, 2005-IJ-CX-0006, March 2010

Wilson JQ, Kelling G (1982) Broken windows. *The Atlantic Monthly*, March, 29–38

Yang S-M (2010) Assessing the spatial-temporal relationship between disorder and violence. J Quant Criminol 26(1):139–163

Longitudinal Studies in Criminology

Holly Nguyen and Thomas A. Loughran
Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA

Overview

The purpose of this entry is to "take stock" of the important contributions of longitudinal studies in the field of criminology. First, classic longitudinal studies in criminology are reviewed. Second, the authors review key findings from the major contemporary longitudinal studies that focus on the development of crime and delinquency. Third, there are numerous longitudinal studies that have a broader focus but have important provided important insight into crime and delinquency; thus, some of the contributions of these studies are examined. To conclude, potential avenues for future research are suggested.

Introduction

Core issues that capture the imagination of criminologists, such as identifying the antecedents of criminal behavior or why some individuals desist from crime while others continue criminal lifestyles, are issues which require longitudinal data to disentangle. Sheldon and Eleanor Glueck conducted a series of seminal studies (1950) that demonstrated that long-term follow-ups studies were both feasible and valuable to the study of crime and delinquency. Since then, there have been specific calls for more longitudinal studies (e.g., Farrington et al. 1986) and recently, there has been considerable growth in both number and quality of longitudinal studies of crime and delinquency. These studies are integral in shaping how the study of crime has been advanced.

Longitudinal, or panel, data consists of repeated observations on the same units of analysis over time, offering many advantages over cross-sectional data. Some of these advantages include having a better precision of the sequence of events timing and measurement; studying intra-individual change; and disentangling direct and indirect causal chains of analyses (Thornberry and Krohn 2003). Despite the vast potential, many early researchers who had access to longitudinal data did not take advantage of this potential and analyzed their data as though it were cross sectional (Farrington et al. 1986). The growth in the number of studies that use longitudinal data in recent years can be attributed to a number of factors including methodological and analytical advances especially developed for panel data. Further, many of the contemporary studies in criminology were initiated in the 1980s with cohorts born in the 1970s. Thus, these data are just now coming to into fruition. Undoubtedly, the use of longitudinal data has advanced the field of criminology both in terms of theory and policy (this, however, is not without controversy).

The purpose of this entry is to "take stock" of the important contributions of longitudinal studies in the field of criminology. Given the abundance of studies, however, this is not a straightforward task. Past reviews of longitudinal research have broadly consisted of two approaches. Some reviews of the literature have discussed key findings of each study (e.g., Thornberry and Krohn 2003), while other reviews choose key topics and summarize the findings across studies (e.g., Farrington 2003). The authors approach the topic by using both approaches. First, the classic longitudinal studies in criminology are reviewed. These studies are important in that they provided a springboard both methodologically and theoretically in which contemporary researchers could be inspired. Second, the key findings from the major contemporary longitudinal studies (that focus on the development of crime and delinquency) are examined. Some of these studies are still ongoing and scholars are just beginning to witness the returns to these data. Third, there are numerous longitudinal studies that have a broader focus, such as the *National Longitudinal Survey of Youth*, but include crime and delinquency measures. The contributions of these studies are reviewed. Finally, this entry concludes by suggesting potential avenues for future research.

Classic Longitudinal Studies

Classic longitudinal studies are characterized by the following of one cohort over time. Information was often collected through official records on a number of important factors such as family life, school, employment, police contacts, and court dispositions. Records would be collected over a number of times. By following one cohort however, potential problems such age, period, and cohort effects can arise (see Farrington et al. 1986). Another criticism is that early researchers did not pay attention to the exact timing of specific events so that causal mechanisms can be established (see Farrington 1988). Although single cohort designs are criticized, it is important to spend some time looking at some of the seminal works in the study of the development of crime and delinquency.

Sheldon and Eleanor Glueck are considered the pioneers of longitudinal research in crime and delinquency. The Gluecks' study consisted of 500 delinquent boys (recruited at ages 7–11) living in the Boston area. Each delinquent was matched with a nondelinquent by age, family background, general intelligence, ethnicity, and residence in an under-privileged neighborhood (Glueck and Glueck 1950). The Gluecks' inquiry fell into five main categories: family and personal background, body types, health, intelligence, and temperament and character. The major contribution of the Gluecks' was their discovery that factors relating to temperament and family characteristics played an important role in later delinquency, and emphasized early detection and prevention.

Another reason why the Gluecks' study is one of the most important longitudinal inquiries is because of the highly influential work of Robert Sampson and John Laub (1993). Sampson and Laub reanalyzed the Glueck data using modern analytic techniques and formulated their age-graded theory of informal social control, which posits that informal social bonds can be an important factor in continuity in and desistance from crime. In 2003, Laub and Sampson conducted a follow-up study of the original participants of the study. In addition to following up with official records of all the original participants, they conducted in-depth interviews with 50 of the men who were in their 70s at the time of follow-up. This makes their study the longest follow-up study in criminology. A number of important contributions have stemmed from Sampson and Laub's work including insight into the mechanisms of desistance and turning points, which include importance of marriage, work, and military service.

The Cambridge Study in Delinquency Development is a prospective study composed of a sample of 411 boys from South London who were aged 8 or 9 between 1961 and 1962. The goal of the original Cambridge study was to test several hypotheses about delinquency: first was to describe development of delinquent and criminal behavior; to assess the efficacy of prediction; and to explain why or why not it continued to adulthood. The investigators examined socioeconomic conditions, schooling, friendship, parent–child relationships, extracurricular activities, school records, and criminal records. They also performed psychological tests to determine the causes of crime and delinquency. The importance of this study is that there have since been follow-up studies of the men up to the age of 48. Few studies have followed as many subjects for as long as the Cambridge study. Researchers are still analyzing data from the Cambridge Study to assess important criminological puzzles.

Another seminal longitudinal study is the Philadelphia Birth Cohort, collected by Wolfgang et al. (1972), which has been described as "one of the turning points in criminological research in the United States" (Laub 2004). The study is composed of two birth cohorts, 1945 and 1958. The 1945 cohort consisted of males who resided in Philadelphia from 10 years (most from birth) to at least 18 years old, and contains almost 10,000 cases. Information contained in school records, police records, and dispositions from juvenile and criminal court records was used to provide important descriptive information and also to allow the researchers to predict offending behavior. Since 1968, Wolfgang et al. were able to follow up and conduct interviews with 567 of the original sample from 1945. The follow-up consisted of questions pertaining to offenses in which the participants were not arrested for. The 1958 cohort study was a replication of the 1945 study to assess whether there were any cohort effects. The 1958 cohort was much larger, over 28,000 subjects and resulted in substantively comparable findings.

The most important contribution of Wolfgang et al.'s study is their identification of the "chronic offender," which they identified as boys with four or more offenses. They found that the chronic offenders consisted of 18 % of the delinquents but were responsible for over half of all the offenses. Further, the Wolfgang et al. (1972) study found that the probability of committing an offense increased with each successive offense committed. That is, the probability of committing a second offense after the first was about 50 %, the third after the second was about 79 %, and about 80 % thereafter. In fact, Wolfgang et al. (1972) have been credited as being the seed of the criminal career paradigm (Piquero et al. 2007). The notion of the chronic offender and methods to prospectively identify such individuals still captures the interest of scholars today.

Other classic studies have also provided a foundation for the study of criminal careers. For example, Blumstein et al. (1985) were able to prospectively identify chronic offenders based on low IQ, poor parent–child rearing, convicted parents, conduct, and other behavioral problems among a London cohort. Other scholars have also argued that factors such as parent–child rearing and early signs of aggression can be predictive of later delinquency. Robins (1978) showed that there is continuity in antisocial behavior and offending. These studies support the famous "Robin's Paradox," which argues that "antisocial behavior virtually requires childhood antisocial behavior; yet most antisocial children do not become antisocial adults" (p. 611). Horney and Marshall (1991) focused on the reliability of life event calendars and emphasized the importance of short-term within-individual variability. Other issues such as work and crime have been made possible through longitudinal studies such as the National Supported Demonstration Work Project. Lyle Shannon also added to criminal career work by following three cohorts from Racine, Wisconsin.

The authors have spent time on some detail reviewing several classic longitudinal studies in the hopes of illustrating the innovative and creative nature of early longitudinal research. These studies provide a foundation in which contemporary researchers are able to refine research designs and move toward a better understanding of the development of crime and delinquency. Indeed, scholars today are still using classic longitudinal studies to provide commentary on new questions (e.g., the next section turns to contemporary longitudinal studies. In reviewing these studies, the authors take a slightly different approach and use the key findings from the past 20 years of longitudinal research as a guide.

Contemporary Longitudinal Studies

The contemporary longitudinal studies reviewed in this entry are more eclectic than the classic studies. They are, however, often larger-scale studies with various methodological improvements over classic studies. For example, rather than one cohort, many of them employ "accelerated cohort" designs, which tracks several cohorts over a shorter period of time while others follow subjects who are closely clustered in age. Many are also conducted at multiple sites. Krohn and Thornberry (2008) identify three main purposes of contemporary longitudinal studies. The primary purpose is to describe delinquent and criminal careers. The second purpose is to identify causal mechanisms of delinquent behavior. The third purpose is to identify the consequences of delinquent/criminal behavior. To be sure, different longitudinal studies have different emphases. For example, some may focus on structural characteristics while others focus on biological factors. There are dozens of studies that can be considered contemporary longitudinal studies, with many still ongoing. These studies have been conducted across various samples and throughout numerous countries. The burgeoning of information makes the task of reviewing contemporary longitudinal studies challenging. Thus, as an organizing framework, the authors identified five key findings gleaned from longitudinal studies in criminology over the past several decades. The five findings include: problem behavior begins early; parents, school, peers, and social structure remain consistent risk factors; there is often co-occurrence of problem behaviors and delinquency; there is a substantial amount of continuity in crime and delinquency; and there is also a considerable amount of change. Each will be reviewed in succession.

1. *Problem behavior begins early*

Since Wolfgang et al.'s finding of the chronic offender, there has been great interest in identifying individuals who are at high risk of becoming chronic or life-course persistent individuals. Several longitudinal studies focus on identifying early signs of risk factors. Risk factors can be identified prenatally or when children are very young (before the age of 6). Scholars have found that prenatal complications, such as exposure to substances like cigarettes, alcohol, and other drugs, interact with psychosocial and environmental factors, such as maternal rejection, poor parenting, neighborhood disadvantage, and low-income, to produce a heightened risk of later antisocial behavior.

Early childhood signs include disruptive behavior like aggression and negative emotionality (Tremblay et al. 2003). Researchers have found that there is a moderately stable relationship between early disruptive behavior and other similar behaviors (heterotypic). Important findings such as those from the Pitt-Mother and Child Project show that very few children have an onset of conduct behavior before the age of 5 or 6. Further studies suggest that the earlier a child is identified and treated, the higher is the likelihood of prevention. Several longitudinal studies have been central to our understanding of early childhood risk factors on later delinquency. This group of studies would include the Montreal Longitudinal and Experimental Study (Tremblay et al. 2003), which tracks subjects from kindergarten to high school and has a special focus on parent–child relationships. The Dunedin Study is a multidisciplinary study that follows 1,000 individuals for almost 40 years and the Longitudinal Study of Biosocial Factors Related to Crime and Delinquency which studies the prenatal and child-related health of a high-risk sample of boys.

2. *Important factors: parents, schools, peers, and social structure*

The previous section discussed how early childhood risk factors can set a child on to a path to later delinquency. Similarly, there are factors in later childhood and adolescence that are also indicative of persistent delinquent behavior. In 1986, the Office of Juvenile Justice and Delinquency Prevention funded a large-scale project entitled the Program of Research on the Causes and Correlates of Delinquency. These projects consisted of: the Denver Youth Study, the Pittsburgh Youth Study, and the Rochester Youth Development study. In a review of the findings of these and four other large-scale longitudinal studies, Thornberry and Krohn (2003) identified parenting, schools, peers, and social structure as consistent factors that can affect children's lives. These predictors can be taken as part of a risk factor approach or guided by criminological theory. Nevertheless, examining these factors provides a useful approach to reviewing longitudinal studies.

Various indicators of poor parenting have been consistently found to be positively related to delinquent or problem behavior. Using the Rochester Youth Development Study, childhood and adolescent maltreatment is linked with later antisocial behavior. Sampson and Laub (1993) found that parent–child attachment mediated the relationship between poverty and delinquency. Similarly, findings from the Montreal Longitudinal Study suggest that healthy child-rearing practices and supervision are crucial protective factors against future antisocial behavior (Tremblay et al. 2003). Using data from the Cambridge Study, scholars investigated childhood neglect and juvenile delinquency and found that the odds of juvenile conviction were over four times higher for subjects who were exposed to childhood neglect than subjects who were not.

Children spend vast amounts of time in school, and thus, experiences in school are an important factor to consider in the development of delinquency. Low levels of intellectual ability and academic achievement are a consistent predictor of delinquency and other problem behaviors (Thornberry and Krohn 2003). For example, Raine et al. (2002) found that life-course persistent offenders had lower visuomotor functioning than other subjects. A birth cohort from Kuai was followed and results suggest that individuals who had below average intellectual skills had a greater likelihood of antisocial behavior.

The relationships between delinquent peers and one's own delinquency have been a part of many longitudinal studies. There is strong support for the causal role that deviant peers play. For example, research suggests that deviant peers provide support and opportunity for delinquent behavior which in turn has an effect on the patterns of associations an individual has. Moreover, in mid-adolescence, selection and socialization are important whereas in late adolescence, socialization is more important. New longitudinal methods have been developed to assess the impact of delinquent peers. Longitudinal social network data gathered by the Netherlands Institute for the Study of Crime and Law Enforcement's School Study, a study that focuses on social networks and the role of peers in delinquency. Scholars have used Simulation Investigation for Empirical Network Analyses, a method that can estimate the effects on individual changes in network ties as well as effects on changes in individual behavior, and found that the average delinquency of an individual's friends has a significant causal effect on the individual's behavior.

The idea of delinquent peers has extended to inquiry into the causes and consequences of gang membership. Krohn and Thornberry (2008) commented that longitudinal studies are an “ideal design for investigating the impact of gang membership on life-course development” (p. 130). For example, the Seattle Social Development Project is an example of a longitudinal study that looks at the development of both positive and problem behaviors among adolescents and young adults. Researchers using data from the study found that youth who are gang members more often commit property and violent offenses.

Since Shaw’s work on juvenile delinquency prevention in the 1930s in Chicago and the founding of the Chicago Area Project, structural factors have been an important part in investigating the development of delinquent behavior. Shaw and McKay’s (1942) seminal work on social disorganization pointed to community factors such as poverty, population mobility, and population heterogeneity as correlates of delinquency. Several contemporary studies have also incorporated structural factors in their design. For example, the Project on Human Development in Chicago Neighborhoods (PHDCN) is an interdisciplinary study which assesses families, schools, and neighborhoods affect child and adolescent development. The PHDCN has two components: a neighborhood level component and an individual level component. Researchers using the PHDCN have provided important insight into the dynamics of the mediating effects of social control on neighborhood disadvantage (Sampson and Morenoff 2006). The Denver Youth Study also integrated high-risk neighborhoods as an integral part of the study.

3. *Co-occurrence of problem behaviors and delinquency*

There have been calls by prominent criminologists for the discipline to consider a broader set of behaviors beyond crime (Sampson and Laub 1993), with the rationale that there may be an underlying dispositional trait between crime and other risky behaviors. Important questions such as whether all problem behaviors, such as substance abuse, delinquency, school problems, aggression, promiscuity, and violence, represent one underlying syndrome and to what extent risk factors are the same for all problem behaviors are of central concern for developmental researchers. Results to date suggest that this is the case. Huizinga et al. (2000) found that individuals with serious delinquency also are more likely to have mental health problems. Similarly, data from the Denver Youth Study show that there is overlap between violent behavior and mental health problems. Using data from the Cambridge study, Piquero and colleagues (2011) found that chronic offenders were more often hospitalized and have a registered disability. Loeber et al. (1998) looked at eight problem behaviors, including substance use, conduct problems, physical aggression, and depression, and found that they were interrelated. Thus, how extensive the overlap between offending behavior and other problem behavior is unclear but commentary on this topic has been consistently positive.

4. *There is a significant amount of continuity*

According to Farrington (2005), one of the most widely accepted conclusions about the development of offending is that there is a marked continuity in antisocial behavior. That is, there is a relative stable ordering of individuals on some level of antisocial tendency. This important finding has motivated scholars to investigate what the mechanisms behind continuity are. Two divergent explanations have dominated criminology: state dependence and population heterogeneity. State dependence argues that the effects of earlier offending have an effect on later offending whereas population heterogeneity offers persistent individual differences as an explanation to continuity. Blokland and Nieuwbeerta (2010) use data from the Criminal Career and Life-course Study, which follows 4,684 Dutch offenders for the better part of their lives, and found that both population heterogeneity and state dependence partially explain continuity in offending. The more an individual offends, however, the effect of state dependence weakens. Violent and aggressive behavior also appears to have continuity. Farrington (1993) suggests that there is continuity in bullying behavior – both within and between generations. Individuals who self-report bullying at 14 were more likely to report bullying at age 32. Their children were also more likely to be bullies.

5. *There is also change*

Thus far, the authors have not portrayed an optimistic view of the development of offending. While there is a marked continuity in offending and other problem behaviors, there is also a great deal of change. Investigation into change was sparked by Sampson and Laub's (1993) study. The scholars found that there are multiple pathways to desistance and identified four major factors associated with change: marriage, the military, reform school, and neighborhood change (Laub and Sampson 2003). The study of desistance remains a challenge however, because measuring desistance is not straightforward. For example, there is debate regarding the definition of what constitutes cessation from crime and the data are ultimately right censored. Moreover, the methodological advancement developed has allowed scholars to study population heterogeneity in developmental offending trajectories by identifying groups of distinct trajectories. Piquero (2008) reports over 60 studies that have employed such group-based trajectory modeling to study change over time.

The idea of change provides researchers with incentive to explore the factors and mechanisms that contribute to desistance. For example, Hawkins et al. (1992) advocate for a risk factor preventive approach whereby risk factors are identified and then addressed to promote desistance. The idea of change was the impetus behind The Pathways to Desistance project. The Pathways study is a large collaborative, multidisciplinary project that followed 1,354 serious juvenile offenders aged 14–18 for 7 years after their adjudication. The main purpose of the study is to identify factors related to continued desistance. General findings suggest that almost all offenders greatly reduce their offending and longer stays in institutions do not reduce recidivism.

Other Important Longitudinal Studies

It is important to note that the authors have based our review primarily on panel studies that focus on the development of crime and delinquency. There are, however, many longitudinal studies that have a broader focus but also include delinquency measures. These studies are often large in scale and are of a representative population sample of adolescents. These studies are rich sources of data that provides insight into many important criminological questions that only longitudinal studies can shed light on.

A couple of the national longitudinal surveys commonly used in criminological research include: the National Longitudinal Survey of Youth (NLSY) and the National Youth Survey (NYS). The National Longitudinal Survey of Youth is a study of 9,000 individuals who were born in 1980–1984 and has been used to study issues such as education, work, and crime (Lochner 2004). The National Youth Survey began in 1976 and is ongoing. Parents and youth were interviewed about both conventional and deviant types of behavior by youths. Important studies on drug use deterrence, labeling, and peer influence have used the NYS.

Monitoring the Future is an annual nationally representative survey of American high school students. Annual follow-up questionnaires are mailed to individuals in each graduating class for a few years after their initial participation. Data from Monitoring the Future has provided insight into patterns of drug use and work and delinquency. The National Longitudinal Study of Adolescent-Health (Add-Health) is also a nationally representative sample of adolescents in grades 7–12. In addition to collecting information on neighborhood, family, social, psychological, health, and economic well-being, the Add-Health collects information on the social networks of adolescents. This has generated a number of important studies that provide commentary on the nature of delinquent peers.

Future Directions

The discipline of criminology owes much of its knowledge of the development of offending to past and current longitudinal studies. In this entry, the authors reviewed classic studies that are the foundation of longitudinal studies in criminology, key findings gleaned from contemporary panel studies, and some of the important nationally representative longitudinal studies on adolescent well-being. However, it is realized that there are other important longitudinal studies that were not included in this entry. To this end, the authors urge readers to explore previous syntheses of longitudinal studies (Blumstein et al. 1986; Delisi and Piquero 2011; Farrington 1988b; Liberman 2008; Thornberry and Krohn 2003).

Despite considerable advances made in the study of the development in crime and delinquent behavior over the last few decades, there are avenues that would benefit from greater scholarly investment. One such avenue is longitudinal experimental designs. Despite highlighting the value of longitudinal experimental studies over 20 years ago, Farrington (2006) was able to only identify four longitudinal experiments that have incorporated significant pre- and post-developmental measures, suggesting that there is a dearth of research that assesses the long-term effects of interventions.

As ongoing longitudinal studies extend into their subjects' early adulthood, it is important to investigate intergenerational study of crime and delinquency. Elder (1985) noted that one of central tenets of studying the life course is the notion of linked lives, which refers to the interaction between individuals and their social world. He argues "Each generation is bound to fateful decisions and events in the other's life course" (p. 40). Issues of central concern such as continuity and change can be extended to look at intergenerational patterns in crime and problem behaviors. Ongoing studies such as the Rochester Developmental Youth Survey and the Denver Youth Study are gathering information on the children of the participants. Preliminary results from these studies and past studies like the Ohio Life Course Study suggest that there is a clear relationship between the antisocial behaviors of parents and their children.

Finally, an important yet often neglected line of inquiry is investigation into the processes of co-offending. Co-offending has important implications for the development of criminal careers. For example, co-offenders tend to be more prolific offenders, are more likely to engage in violence, and have longer criminal careers (McGloin and Nguyen 2012). Indeed, greater insight into co-offending can be important in terms of theory and policy. In conclusion, longitudinal studies in criminology have made significant strides since the Glueck's Unraveling Delinquency study. Continued improvements in methodological designs, advancements in statistical applications, and theoretical refinements will provide scholars of crime with a number of interesting avenues for research.

Recommended Reading and References

Blokland AAJ, Nieuwbeerta P (2010) Considering criminal continuity: testing for heterogeneity and state dependence in the association of past to future offending. Aust N Z J Criminol 43(3):526–556

Blumstein A, Farrington DP, Moitra S (1985) Delinquency careers: innocents, desisters and persisters. In: Tonry M, Morris N (eds) Crime and justice: an annual review of research, vol 6. University of Chicago Press, Chicago, pp 187–219

Blumstein A, Cohen J, Roth JA, Visher CA (eds) (1986) Criminal careers and "career criminals". National Academy Press, Washington, DC

Delisi M, Piquero AR (2011) New frontiers in criminal careers research, 2000–2011: a state-of-the-art review. J Crim Justice 39:289–301

Elder GH (ed)(1985) Life course dynamics: trajectories and transitions, 1968–1980. Project of SSRC Committee on the Life Course. Cornell University Press, Ithaca

Farrington DP (1988a) Advancing knowledge about delinquency and crime: the need for a coordinated programme of longitudinal research. Behav Sci Law 6:307–331

Farrington DP (1988b) Studying changes within individuals: the causes of offending. Studies of psychosocial risk: the power of longitudinal data. In: Rutter M (ed) Studies of psychosocial risk: the power of longitudinal data. Cambridge University Press, Cambridge, pp 158–183

Farrington DP (1993) Understanding and preventing bullying. In: Tonny M, Morris N (eds) Crime and justice, vol 17. University of Chicago Press, Chicago

Farrington DP (2003) Key results from the first 40 years of the Cambridge study in delinquent development. In: Thornberry TP, Krohn MD (eds) Taking stock of delinquency: an overview of findings from contemporary longitudinal studies. Kluwer, New York, pp 137–174

Farrington DP (2006) Key longitudinal-experimental studies in criminology. J Exp Criminol 2:121–141

Farrington DP, Ohlin LE, Wilson JQ (1986) Understanding and controlling crime: toward a new research strategy. Springer, New York

Glueck S, Glueck E (1950) Unraveling juvenile delinquency. Commonwealth Fund, New York

Hawkins J, Catalano R, Miller J (1992) Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: implications for substance abuse prevention. Psychol Bull 112:64–105

Huizinga D, Loeber R, Thornberry TP, Cothern L (2000) Co-occurrence of delinquency and other problem behaviors. Bulletin. U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention, Washington, DC

Krohn MD, Thornberry TP (2008) Longitudinal perspectives on adolescent street gangs. In: Lieberman A (ed) The long view of crime: a synthesis of longitudinal research. Springer, New York, pp 128–160

Laub JH (2004) The life course of criminology in the United States: the American Society of Criminology 2003 presidential address. Criminology 42(1):1–26

Laub JH, Sampson RJ (2003) Shared beginnings, divergent lives: delinquent boys to age 70. Harvard University Press, Cambridge

Liberman AM (2008) The long view of crime: a synthesis of longitudinal research. Springer, New York

Lochner L (2004) Education, work, and crime: a human capital approach. Int Econ Rev 45(3):811–843

Loeber R, Farrington DP, Stouthamer-Loeber M, Van Kammen WB (1998) Antisocial behavior and mental health problems: explanatory factors in childhood and adolescence. Lawrence Erlbaum, Mahwah

McGloin JM, Nguyen H (2012) “It was my idea”: considering the instigation of co-offending. Criminology 50(2):463–494

Piquero A (2008) Taking stock of developmental trajectories of criminal activity over the life-course. In: Lieberman A (ed) The long view of crime: a synthesis of longitudinal research. Springer, New York, pp 23–78

Piquero AR, Sheperd I, Sheperd JP, Farrington DP (2011) Impact of offending trajectories on health: disability, hospitalisation and death in middle-aged men in the Cambridge study in delinquent development. Crim Behav Mental Health 21(3):189–201

Raine A, Yarlian PS, Reynolds C, Venables PH, Mednick SA (2002) Spatial but not verbal cognitive deficits at age 3 years in persistently antisocial individuals. Develop Psychopathol 14(1):25–44

Robins LN, Radcliffe KS (1980) The long-term outcomes of truancy. In: Hersov L, Berg I (eds) Out of school. Wiley, New York

Sampson RJ, Laub JH (1993) Crime in the making: pathways and turning points through life. Harvard University Press, Cambridge

Sampson RJ, Morenoff JD (2006) Durable inequality: spatial dynamics, social processes and the persistence of poverty in Chicago neighborhoods. In: Bowles S, Durlauf S, Hoff K (eds) Poverty traps. Princeton University Press, Princeton, pp 176–203

Shaw CR, McKay HD (1942) Juvenile delinquency in urban areas. University of Chicago Press, Chicago

Thornberry TP, Krohn MD (2003) The development of panel studies of delinquency. In: Thornberry TP, Krohn MD (eds) Taking stock of delinquency: an overview of findings from contemporary longitudinal studies. Kluwer, New York, pp 1–6

Tremblay RE, Vitaro F, Nagin D, Pagani L, Seguin JR (2003) The Montreal longitudinal and experimental study rediscovering the power of descriptions. In: Thornberry TP, Krohn MD (eds) Taking stock of delinquency: an overview of findings from contemporary longitudinal studies. Kluwer, New York, pp 205–254

Wolfgang ME, Figlio RM, Sellin T (1972) Delinquency in a birth cohort. University of Chicago Press, Chicago

Low Self-Control

► General Theory of Crime

Low-Level Data Forensics

► Computer Forensics

M

MABS

► Simulation as a Tool for Police Planning

Macro-social Criminology

► Institutional Anomie Theory

MACS

► Simulation as a Tool for Police Planning

Malls

► Preventing of Terror at Shopping Malls

Managerial Court Culture

Brian J. Ostrom and Roger A. Hanson
National Center for State Courts, Williamsburg,
VA, USA

Overview

Court culture as a concept follows from the modern management maxim that there is more than one way to get things done and done well in the workplace. Formulating an effective strategy for a particular workplace requires a good understanding of not only the formal structure and lines of authority but also the unwritten rules, unofficial networks, and underlying norms and behaviors that shape how work is accomplished. As a result, knowledge of an organization's cultural web is a crucial factor when searching for ways to improve operational effectiveness.

This entry outlines the development and current state of court managerial culture by reviewing five pertinent inquiries. All of the studies focus primarily on American felony criminal state trial courts and share a comparative approach, examining courts in several states. The first pair provides initial impetus by popularizing the concept of local legal culture as it relates to why case resolution in some courts is timelier than in others. Timeliness is an undeniable mark of performance due to the US Constitution's enshrinement of a "speedy trial" in the provisions of the Seventh Amendment. The third study extends the notion of culture by elaborating three distinct work orientations made up of different constellations of norms. The fourth study provides evidence that culture affects different behavior patterns among courts as well as in different attitudes on the part of prosecutors and defense attorneys. The fifth project draws on a broad range of research traditions, including organizational effectiveness in both the private and public sectors and

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

modern methodologies. The inquiry develops a conceptual framework, applies it, and shows how results are solid grounds for changing behavior. The conceptual framework has the dual advantage of creating a typology of culture based on a set of values appropriate to the American context and simultaneously amenable to using alternative values as might exist in other legal systems. Hence, the entry concludes with suggestions for future research, including work in other countries.

Introduction

A perennial theme in contemporary criminal court administration is that better management enables higher organizational performance benefitting all participants in the legal process. Superior administration of the legal process comes about supposedly by the actions of enlightened leaders who cultivate collegial agreement on administrative goals to resolve cases expeditiously and fairly treat litigants and others, and then put objectives into place. Yet, the nature of court organizational structure introduces notable challenges to such aspirations.

Minimal hierarchy of authority exists around which to fashion purposeful management. Professional attorneys, who by virtue of public election or legislative or executive appointment, occupy the position of state constitutional officers. Judges do not select their colleagues and thereby have limited means to control the behavior of their fellow members of the bench. As a result, a call for reaching collective decisions in a court requires an appreciation of the coequal status of judges. Even a judge in a leadership position, such as the presiding or chief judge, possesses limited formal authority to set administrative policy for the court, an equal among firsts. Compare this type of setting to an executive agency led by a chief cabinet officer that appoints assistant secretaries. If the judges are willing to share, collaborate, and follow the lead of a presiding judge, supported by a court administrator, establishing court-wide ways of doing business is a viable option. On the other hand, if individual judges seek to maintain autonomy in administrative matters, reaching agreement on standard work practices is much more difficult. As a result, voluntary commitment to collective decision making inhibits the adoption of court-wide policies and procedures – even those deemed to be best practices.

Courts are not the only institutions challenged by calls for synchronized management that seem quite natural in the bottom-line motivated private sector. They fit within a category of structures called loosely coupled systems. “In loosely coupled systems, the forces for integration—for worrying about the whole, its identity, its integrity and its future—are often weak compared to forces for specialization” (Hirschorn 1994). Coordination is difficult because the members do not think about system-wide goals, but, as in the case of courts, focus instead on their own subunit, assigned caseload, or courtroom. Recognizing the decentralized character of how decision making occurs in each court is another way of saying court culture matter.

Early Efforts

The first term to describe court culture was local legal culture. The idea of local legal culture arose in two complementary studies of delay reduction in criminal courts during the 1970s. In 1976, Nimmer (1976) observed that the “local discretionary system” is a major obstacle to criminal court reform efforts and later went on to claim that lengthy case processing times are “most directly associated with prevailing informal norms of the judicial process and with the personal motivations of participating attorneys and judges” (1978). Following this pioneering work, what caught the imagination of the field of court administration as well as law professors, judges, and social scientists was a subsequent study of both civil and criminal trial courts by Church et al. (1978): They formulated a working hypothesis that challenged the conventional wisdom that resources determined success in resolving cases expeditiously. They instead postulated

> The speed of disposition of civil and criminal litigation in a court cannot be ascribed in any simple sense to the length of its backlog, any more than court size, caseload, or trial rate can explain it. Rather, both quantitative and qualitative data generated in this research strongly suggest both speed *and* backlog are determined in large part by established expectations, practices, and informal rules of behavior of judges and attorneys. For want of a better term, we have called this cluster of related factors the "local legal culture."

Similar to Nimmer, Church et al. defined local legal culture as the "established expectations, practices and informal rules of behavior of judges and attorneys." Their interpretation of the wide variation in the timeliness of different courts attributed the speed of resolution more to the views of judges and attorneys than to structural, resource, or procedural distinctions among courts. Both sets of studies consider expectations stable implying that efforts to reduce court delay encounter strong resistance unless the expectations themselves are the subject of planned change. The concept of local legal culture enjoyed acceptance of the proposition that timely court performance is primarily governed by shared beliefs, expectations, and attitudes, although the substantive nature of the determining views lacked specificity and confirming evidence. As a result, local legal culture does not distinguish different types of cultures and does not offer a basis for putting courts into different categories.

Culture as a Working Hypothesis

Local legal culture is the subject of considerable attention and conceptual enrichment in a third inquiry by Eisenstein, Fleming, and Nardulli (Eisenstein et al. 1987). These scholars studied three criminal courts in each of three states, Illinois, Michigan, and Pennsylvania. They claim norms define how the courts should operate and that differences in norms contribute greatly to the observed variety in how courts conduct business. They extend the previous research by focusing more on what constitute the norms. By calling particular sets of norms "work orientations," they signal that norms are complex and have multiple consequences.

Work orientations "are rationalizing principles" that court leaders "use to explain why particular tasks or functions" are structured the way they are in their respective courts (Nardulli et al. 1988). These powerful forces are not the same from court to court, but they also are not unique to each court community. Indeed, the scholars see three types of distinctive work orientations in the courts they studied: (1) structural or formal, (2) efficiency, and (3) pragmatic orientations.

The first orientation emphasizes the compliance with professional norms including close adherence to rules and the rejection of expediency. The second orientation places a premium on the efficient use of resources and promotes the expeditious handling of cases, even if the achievement of smooth handling occasionally calls for deviation from rules. Finally, a pragmatic orientation is the most flexible approach because the routine is whatever the consensus is at a specific point in time. With no long-term commitment made to a particular manner of conducting business, the prevailing norm is deemed satisfactory until a problem emerges calling the existing paradigm into serious question. Then, new agreements on ways of resolving cases are necessary and appropriate and the emerging consensus guides the court until it too, ultimately proves problematic. This intriguing configuration of cultures advanced the concept of local legal culture but emerged as a working hypothesis at the conclusion of the inquiry. None of courts under study actually were placed in one of the orientation categories and no tests were conducted on the relationship between orientations and expected consequences.

Judges Nudge Attorneys to an Expeditious Culture

A fourth inquiry by B. Ostrom and Hanson (1999) provides empirical evidence pertinent to a critical aspect of managerial culture. Their study proposed that timeliness and quality are compatible. This proposition challenged the popular view that efforts to resolve cases more quickly compromise quality because delay reduction activities cut corners. This research focused on nine midsized communities: Albuquerque, Austin, Birmingham, Cincinnati, Grand Rapids, Hackensack, Oakland, Portland, and Sacramento. An initial part of the research involved dividing the courts according to their observed degrees of timeliness and placing them into one of three subgroups: very expeditious, moderately expeditious, and least expeditious. To test the study's hypothesis, frontline attorneys completed questionnaires that tapped into their views on the workings of the state trial court where they practiced and how they assessed the opposing side.

Findings from the investigation revealed that when courts act decisively to manage the litigation, they achieve both timeliness and enable attorneys for both the state and the defendant to prepare and present their arguments effectively. Attorneys have distinctive views toward the leadership role played by a court, a court's ability to communicate its expectations clearly, the degree to which the opposing side is well trained and prepared for trial, and the extent to which the opposing side operates in an adversarial manner. On the one hand, if attorneys see a court exercising firm leadership, a court stating its policies clearly, and the opposing side is equipped for trial, the court is among the most expeditious subgroup. Conversely, if the attorneys see a court as a hazy communicator, a source of weak leadership, and the opposing side as ill prepared, the court is of lesser expeditiousness. In all three subgroups, attorneys view the opposing side as a strong adversary, suggesting the timely resolution of cases does not require counsel to abandon the goals of protecting society (prosecutor) or to defend a criminal defendant's constitutional rights (defense attorney). These relationships offer strong evidence that the degree to which the Seventh Amendment is consistent with reality depends on effective management actions taken by judges and administrators and embraced by legal practitioners. However, what is missing concerns the rationale prompting judges to exercise control over the legal process in some courts to a greater degree than in other courts.

Taken together, this previous body of research does converge on the central point that court culture, norms, and beliefs held by judges, attorneys, and court administrators shape how courts operate. Moreover, these investigations document that legal practitioners are aware of court policies to the extent that judges and administrators clearly articulate when, how, and why court proceedings are going to occur. These four studies provide solid justification for further inquiry into the specific norms and values associated with different culture types. A fundamental component missing from the initial research was a methodology for measuring culture and an investigation into the consequences of possible cultural variation.

Constructing Managerial Court Culture

The earlier lines of research primarily affirm what court administrators and judges know: culture plays an important role in how courts function. Some management cultures inhibit modernization, reform, and performance. Others are more conducive to the development and adoption of better ways of doing things. In essence, the notion of local legal culture emerged as a shorthand phrase to refer to a host of norms and resulting behaviors not otherwise easily explained. Unfortunately, simply naming a phenomenon is not the same as measuring it and using it to both explain and improve court performance. Without a vocabulary and set of tools to distinguish fundamental types of cultures, courts will continue to struggle in building a management culture that supports and expects to achieve high-quality case resolution.

Building on and refining previous studies, a fifth approach by Ostrom et al. (2007) provides a more comprehensive framework, along with a set of steps and tools to assess and measure a court's current and preferred culture. Culture assessment furthers observation and measurement of the abstract concept of court culture, thereby making it an explicit part of court management and reform efforts. Because assessing court culture can yield systematic information compatible with and useful to understanding court performance and the allocation of court resources, an expository summary of this fifth link in a long line of research indicates the current state of knowledge.

B. Ostrom et al. follow the lead of managerial experts who contended culture is the glue that operates at many different levels in an organization. Schein (1999) argued that to comprehend what matters in culture, one must strive to understand the espoused values (i.e., the values that shape why an organization acts in a particular way) and basic assumptions (i.e., jointly learned values, beliefs, and assumptions that become shared and taken for granted in an organization) that shape the way work gets done in the organization. For Schein, culture is the mental representation of the work environment that members of the organization carry in their heads.

Schein's advice, applied by B. Ostrom et al. to the study of courts focused on shared mental models that judges, administrators, and staff hold and take for granted. A court's management culture manifests itself in what is valued, the norms and expectations, the leadership style, the communication patterns, the procedures and routines, and the definition of success that makes the court unique. More simply: "The way things are done around here."

B. Ostrom et al. accepted the proposition that culture evidences itself in the "accumulated shared learning of a given group" (Schein 2004). For courts, this shared learning comes from parallel experiences in meeting the aims of the judicial branch. The live and unanswered question then became for B. Ostrom et al. as, if a court's culture is the result of accumulated learning, how does one describe and catalogue the content of that learning?

Types of Managerial Court Culture. The conceptual framework developed by B. Ostrom et al. organized managerial culture into four distinct types of cultures, communal, networked, autonomous, and hierarchical, defined as follows:

- ***Communal***: Judges and managers emphasize the importance of getting along and acting collectively. Communal courts emphasize group involvement that produces mutually arrived at agreements. Procedures are open to interpretation and creativity is encouraged when it seems important to "do the right thing."
- ***Networked***: Judges and managers emphasize inclusion and coordination to establish a collaborative work environment and effective court-wide communication. Efforts to build consensus on court policies and practices extend to involving other justice system partners, groups in the community, and ideas emerging in society. Judicial expectations concerning the timing of key procedural events are developed and implemented through policy *guidelines* built on the deliberate involvement and consensus of the entire bench. Court leaders speak of courts being accountable for their performance and the outcomes they achieve.
- ***Autonomous***: Judges and managers emphasize the importance of allowing each judge wide *discretion* to conduct business. Many judges in this type of court are most comfortable with the traditional adversary model of dispute resolution. Under this traditional approach, the judge is a relatively passive party who essentially referees investigations carried out by attorneys. Individual judges exercise latitude on key procedures and policies. Limited discussion and agreement exist on court-wide performance criteria and goals.
- ***Hierarchical***: Judges and managers emphasize the importance of established *rules* and *procedures* to meet clearly stated court-wide objectives. These courts seek to achieve the advantages of order and efficiency in a world of limited resources and calls for increased accountability. Effective leaders are

good coordinators and organizers. Recognized routines and timely information reduce uncertainty, confusion, and conflict in how judges and court staff make decisions.

Empirical Findings. B. Ostrom et al. measured the extent that courts in the real world fit into one of these four types. Moreover, they assessed both the current culture and the type of culture that judges and administrators preferred to see in place in the future. Additionally, both prosecutors and public defenders provided evaluations of the performance of the court where they practiced on a daily basis. As a result, conclusions surfaced on what, if any, consequences flowed from the judicial embrace of a particular culture. The following synopsis highlights their essential observations.

Courts with hierarchical cultures achieve objective standards of timeliness promoted by the American Bar Association (ABA) and other groups, more closely than courts with other dominant cultures do. Interestingly, all of the courts under study exhibited awareness of this connection. Every court, according to the preferences of judges and administrators, seeks to increase its hierarchical culture in the area of case management. This common desire likely arises because none of the courts currently meet the ABA-prescribed time frames. Courts are cognizant of this situation and seek to do better by moving toward a culture more conducive to expedition and timeliness.

Like culture, court performance has multiple dimensions. Moving beyond timeliness, practicing attorneys evaluate courts on how well they achieve other important goals, such as access, fairness, and managerial effectiveness. With these criteria, the picture becomes more complex. The courts with more hierarchical cultures rated less satisfactorily on access and fairness, for example, than courts with other cultures. Prosecutors and public defenders do not see only virtue in a court's emphasis on timeliness and a move toward a more hierarchical culture to achieve that goal. Both sets of attorneys also can see greater benefits to themselves from the courtroom workgroup relations fostered by an autonomous culture's minimal emphasis on solidarity. Vested interests in established work patterns perhaps lead attorneys to view courts with autonomous cultures as also doing better when it comes to access, fairness, and managerial effectiveness. Further complexity arises because prosecutors and public defenders disagreed over the relative merits and limitations of other cultures. The former see the greatest advantage in networked cultures because of the emphasis by the court on guidelines consistent with a prosecutor's desire for some firmness in guilty plea policies. Public defenders have an opposite orientation and like a communal culture's emphasis on flexibility that they view as enhancing their role in gaining the best resolution for their clients. Managing the competing interests of prosecutors and public defenders, while maintaining an effective internal work environment for judges and court staff members, highlights the management challenges facing court leaders.

Methodology. The development of this fourfold typology follows from an analysis of how expert practitioners believe core values affect and relate to carrying out work. Sixteen values culled from the literature on court administration include such distinct values as collegiality, continuity with the past, discretion, standard operating procedures, flexibility, rule-oriented, innovation, judicial consensus, and self-managing. Using a tightly structured questionnaire, 53 seasoned practitioners, including judges, administrators, prosecutors, and defense attorneys, compared and contrasted the values. This exercise asked the practitioners to indicate how closely each of the 16 values relates to each of the other 15 values. The results of paired comparisons, using the technique of multidimensional scaling, showed four clusters of four values each.

The clusters illustrate the core values of different types of cultures and fit along two dimensions called solidarity and sociability. Solidarity refers to the degree to which a court has clearly understood shared goals, mutual interests, and common tasks and sociability refers to the degree to which people work together and cooperate in a cordial fashion. Each of the four

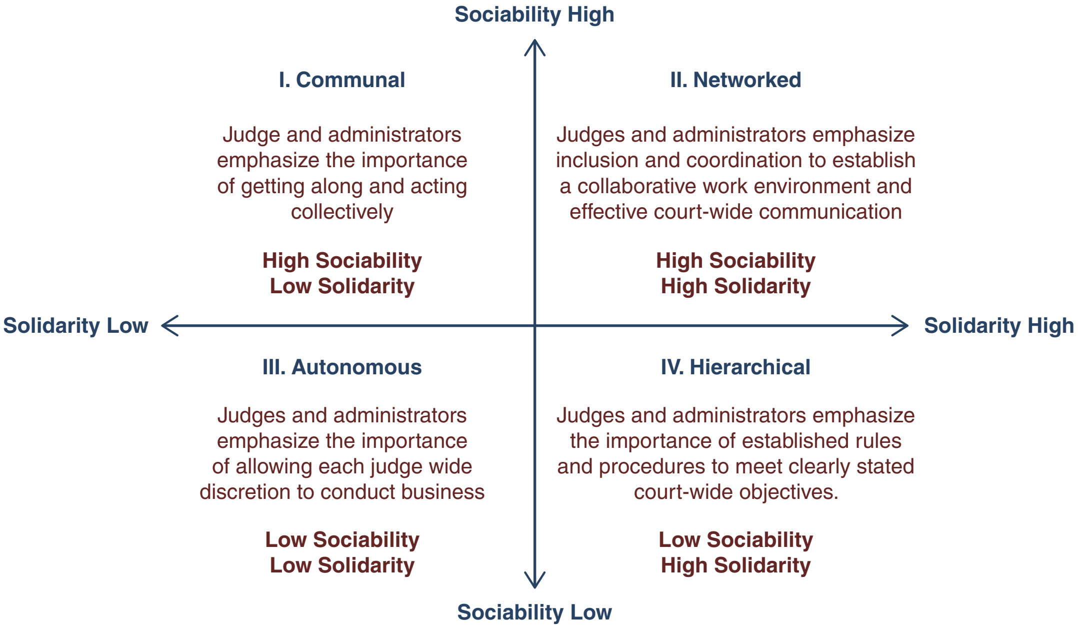
Sociability High
I. Communal
Judge and administrators emphasize the importance of getting along and acting collectively
High Sociability
Low Solidarity
II. Networked
Judges and administrators emphasize inclusion and coordination to establish a collaborative work environment and effective court-wide communication
High Sociability
High Solidarity
Solidarity Low
Solidarity High
III. Autonomous
Judges and administrators emphasize the importance of allowing each judge wide discretion to conduct business
Low Sociability
Low Solidarity
IV. Hierarchical
Judges and administrators emphasize the importance of established rules and procedures to meet clearly stated court-wide objectives.
Low Sociability
High Solidarity
Sociability Low

Managerial Court Culture, Fig. 1 Court culture classifications

cultures is a particular combination of solidarity and sociability, as shown in Fig. 1.

Communal culture is low on solidarity and high on sociability. Its distinctive values are flexibility, egalitarianism, negotiation, and trust. A network culture seeks both sociability and solidarity. Its values include judicial consensus, innovation, visionary thinking, and human development. An autonomous culture emphasizes neither sociability nor solidarity. Its values are self-managing, continuity, independence, and personal loyalty. A hierarchical culture stresses solidarity but not sociability. Its values are rules, modern administration, standard operating procedures, and merit-based staff promotions.

B. Ostrom et al.'s proposed concept of culture is manifested in familiar and recognizable activities called "work areas," such as the handling of cases, the responsiveness of courts to the concerns of the community, the division of labor and allocation of authority between judges and court staff members, and the manner in which court leadership is exercised. Each particular culture's way of doing things exists in four work areas as seen in the Value Matrix (Fig. 2).

B. Ostrom et al. developed a survey instrument to determine what individual judges and administrators believed about the execution of work in key areas. Because each culture was presumed to manifest itself differently, the instrument asks individuals to indicate how closely each of four ways of getting work done corresponds to what happens in their court (current culture) and what they would like to see as the work style in the future (preferred culture). An application of the framework to courts in California, Colorado, Florida, Maryland, Minnesota, Ohio, Oregon, Utah, and Washington and the Tax Court of Canada found examples of each of the four cultures, although the autonomous culture is the most frequent. This balanced distribution suggested courts are not monochromatic in their work orientations. On the other hand, regardless of the current culture, the vast majority of courts under study indicated a similar mosaic-like preference for the future. Specifically, they tend to desire hierarchical orientations to dominate in the work areas of case management and change management, networked orientations to dominate judge–staff

	Communal	*Networked*	*Autonomous*	*Hierarchical*
Case Management Style	Flexibility— Judges follow accepted principles for the timing of key procedural events, but are comfortable fashioning their own approach to "do the right thing."	Judicial Consensus— Judicial expectations concerning the timing of key procedural events are developed and implemented through policy guidelines built on the deliberate involvement and consensus of the entire bench.	Self-managing— Individual judges are relatively free to make their own determinations on when and how key procedural events are to be completed.	Rule oriented— Judges are committed to the uniform use of standard case flow management techniques (e.g., early case control, case coordination, and firm trial dates) with the support of administrative and courtroom staff. Written court rules and procedures govern what judges do.
Judge and Court Staff Relations	Egalitarian— Characterized by teamwork, cooperation, and participation. Judges, court managers, and staff work things out flexibly as they go along. Judges agree all individual staff members should obtain satisfaction from work, but no set training program applies to all staff uniformly.	People Development— Characterized by commitment to innovation, diversity of ideas, and widespread managerial and courtroom staff development. Attention is paid to developing effective court-wide communication. Regular systematic performance evaluations are encouraged.	Personal Loyalty— Characterized by personal loyalty to individual judges. Non-standardized procedures are the norm as judges have wide discretion in how they recruit, manage and reward their courtroom support staff.	Merit— Characterized by formal rules and policies, with people following clear guidelines and written instructions about work. Reasons for rewards and demerits are clear. Poor performance is dealt with quickly. Maintaining a smooth running organization is important.
Change Management	Negotiation— The change process tends to occur incrementally through negotiation and agreement. Procedures are seldom rigid so that the actual application of policy changes may reflect revision and compromise among work teams of individual judges and corresponding court managers and staff.	Innovation— The change process tends to be proactive in order to achieve desired goals. Judges and court managers are open to new challenges and acquiring new resources to support innovation. Monitoring and reacting to broad court performance targets are encouraged.	Continuity— The change process tends to occur sporadically as the court is generally content to preserve established ways of doing business. Centralized change initiatives are a challenge because each judge exercises a wide scope of latitude in the choice of practices and procedures.	Modern Administration— The change process tends to emphasize improved efficiency and using new techniques to measure the way work is done. Judges and court managers seek and use court performance information, data, and technologies to help make better business decisions.
Courthouse Leadership	Trust— Leadership in the court is generally considered to exemplify building personal relationships and confidence among all judges and court employees; and seeking to reconcile differences through informal channels.	Visionary— Leadership in the court is generally considered to exemplify innovation, inclusion, and coordination by the presiding judge and/or court management team to establish a collaborative work environment.	Independence— Leadership in the court is generally considered to exemplify preserving individual judicial discretion, allowing judges to use their own criteria in defining success, and not necessarily relying on the same indicators of achievement.	Standard Operating Procedures— Leadership in the court is generally considered to exemplify centralized control and organization to achieve administrative efficiency. A presiding judge and/or court management team typically has authority to establish a clear division of labor and set court wide expectations.

Managerial Court Culture, Fig. 2 Value matrix member relations, and a communal culture to dominate the area of courthouse leadership.

Origins and Controversy. The effort to study managerial court culture has roots in initial court research and from private sector studies and selected public organizational experts on culture. Concerning private sector management and performance, B. Ostrom et al. looked to the work of business school scholar Robert Quinn and his colleagues. They had embarked on nearly two decades of sequential studies and made clear the relevance of culture as they summarize the traits of successful private sector businesses:

> The major distinguishing feature in these companies, their most important competitive advantage, and the most powerful factor they all highlight, as a key ingredient in their success, is their organizational culture. The sustained success of these firms has had less to do with market forces than company values; less to do with competitive positioning than personal beliefs; less to do with resource advantages than vision. In fact, it is difficult to name even a single highly successful company...that does not have a distinctive, readily identifiable, organizational culture. (Cameron and Quinn 1999)

Growing awareness that culture matters to performance and long-term success in the world of business triggered the emergence of culture analysis as a definable area in the field of management and organizational studies largely beginning in the 1980s.

B. Ostrom et al. then found support in applying Quinn's ideas to courts in the views of public sector scholars, such as Wilson and Dilulio. Wilson (1989) had made the case for including organizational culture in the study of public sector organizations. Wilson opened the door for integrating work on private sector organizational culture into the study of the public sector with the following claim: "Organizations matter, even in government agencies" (1989). He argued culture is a core topic for public sector organizational analysis because organizations have a culture just as an individual has a personality. Organizational culture is a relevant and important facet of all bureaucracies – public and private.

However, when shifting the focus to public institutions, Wilson anticipated some difficulties in diagnosing a public organization's culture. Specifically, public organizations do not necessarily have a single culture. In response, B. Ostrom et al. expected courts might manifest a mosaic of complementary and competing cultures. Identifying and understanding the resulting mosaic and the relationship to performance becomes vital to public sector management.

Additional insights come from DiIulio (1993) who had suggested that the measurement of organizational culture is an important part of any public administration improvement strategy. He had proposed the following agenda for doing so:

1. To observe how members at every level of an organization "really behave"
2. To relate systematically these observations to the formal character of the organization in order to see what (if any) connections exist
3. To search systematically for the connections (if any) between organizational activities and real-world outcomes

The key research focus is to discover the relationship between "the way things are" in an organization and the ability of the organization to reach its stated ends.

Following the direction of Wilson and DiIulio's arguments, B. Ostrom et al. found three reasons to suggest private sector studies inform the specific topic of court culture. First, the concept of court culture and the parallel private sector investigations both revolve around the idea that "the way things get done" (i.e., work orientations) defines the character of the organizations. To the extent that judges and administrators in one court have particular views on how cases should be resolved, they will organize themselves differently than individuals with dissimilar views in another court. Second, work orientations shape the degree of a court's effectiveness. This linkage in the private sector made the search for a parallel relationship in public institutions, including courts, worthwhile. Third, business school scholars have shown private organizational culture is susceptible to measurement and the results are interpretable within a typology of cultures.

Of course, some scholars argue that knowledge about private organizations cannot be a basis for understanding public bodies. Wallace Sayre expressed this viewpoint in his well-known assertion that public and private organizations are "fundamentally alike in all unimportant respects." More specific assumptions underlying this argument are public agencies that, in contrast to private bodies, lack a clear bottom line (e.g., profit and market performance), have a diverse set of goals and performance criteria, and are more "open" and with greater exposure to public scrutiny, and managers of public organizations have shorter time horizons. Yet, when looking at the temperaments, skills, and techniques of judges, court managers, and court employees, the differences between public sector and private sector organizations and management are minimal. This point of view is expressed in Lynn's observation that

> [t]he two sectors are constituted to serve different kinds of societal interests, and distinctive kinds of skills and values are appropriate to serving these different interests. The distinctions may be blurred or absent, however, when analyzing particular managerial responsibilities, functions, and tasks in a particular organizations. The implication of this argument is that lesson drawing and knowledge transfer across sectors is likely to be useful and should never be rejected on ideological grounds. (2003, 3)

The ability to measure culture by borrowing the tools available in the broader organizational culture literature highlights the relevance of the methodologies to the field of court administration.

Unsettled Issues

An unresolved issue concerns the scope of the managerial court culture approach. Even in the strictly American context, the question is whether a culture framework applies to different types of disputes, such as civil, family, juvenile, or probate. For B. Ostrom et al., broadening the scope of study beyond criminal cases is possible because despite the unique set of substantive laws governing other types of disputes, they all share an essential characteristic. In every court, cases move from one key procedural event to another. To execute this common process, the activities represented by the four areas of the court culture framework occur in every type of case. As a result, case management style happens in civil cases just as in criminal cases. Judge–staff relations exist in probate cases just as in criminal cases. Courthouse leadership governs juvenile cases just as in criminal cases. Furthermore, how the work is to be done in each area reflects cultural values such as independence, discretion, inclusiveness, and efficiency more than by differences in evidentiary standards, rights of the parties, and severity and type of possible sanctions imposed on a losing party among courts.

Legal criteria specify desirable goals more than how to achieve them. In fact, appellate courts are likely to be amenable to cultural examination because the key procedural events are more definable and similar across appellate courts than trial courts (Chapper and Hanson 1990). Experts and practitioners are likely to have multiple hunches and notions to explain anticipated cultural variation. What judges and court administrators think are appropriate work orientations will likely vary by case type and differ from what criminal courts demonstrated. For example, in the area of case management style, judges with a civil docket may prefer to give greater deference to the bar. In the area of judge–staff relations, probate judges may want to grant greater independence to legally trained staff. In the area of courthouse leadership, judges handling juvenile dependency cases may seek to develop strong connections with social welfare agencies. For all these reasons, B. Ostrom et al. believe that despite variation in work orientations across areas of law, the managerial court culture framework opens up a valuable vista to future court studies.

Ongoing efforts to understand and diagnose court organization culture in American trial courts highlight the relevance of the approach to legal systems outside the borders of the United States. Without claiming that the use of those tools answers all questions, court scholars in other countries should benefit by adapting what has been set in motion. Applicability of the methods likely hinges on three conditions: (1) the opportunity for collegial discussion among judges and staff members exists, (2) judges and staff members have some say over the design and implementation of court administration policies and practices, and (3) the judge with official leadership responsibilities has an interest in organizational improvement. The more vibrant the collegial discussion, the more aware judges are that their administrative decisions have independent consequences, and the greater the interest of administrative leaders, the greater the utility of investigating managerial court culture. The use of the same methods that are free and open to inspection augurs well for truly comparative research across boundaries.

Conclusion

Culture focuses attention on variables exercising a strong, independent influence on the administration of the legal process. Reforms need to bond with cultural values to stand a chance of influencing court performance. For this reason, the effort to understand cultural values as indicators of the current state of affairs and future possibilities is as deserving of attention as any other aspect of court management, such as structure, process, or resources. Implementation of new practices becomes less difficult under a supportive culture.

Recommended Reading and References

Cameron KS, Quinn RE (1999) Diagnosing and changing organizational culture: based upon the competing values framework. Addison-Wesley, Reading, MA

Chapper JA, Hanson RA (1990) Intermediate appellate courts: improving case processing. National Center for State Courts, Williamsburg, VA

Church TW Jr (1982) The ‘old’ and the ‘new’ conventional wisdom. Justice Syst J 8:712

Church TW Jr, Carlson A, Lee J-LQ, Tan T (1978) Justice delayed: the pace of litigation in urban trial courts. National Center for State Courts, Williamsburg, VA

DiIulio JJ Jr (1993) Measuring performance when there is no bottom line. In: Performance measures for the criminal justice system. U.S. Department of Justice, Bureau of Justice Statistics, Washington, DC,pp 143–156

Eisenstein J, Jacob H (1977) Felony justice: an organizational analysis of criminal courts. Little Brown, Boston

Eisenstein J, Fleming R, Nardulli P (1987) The contours of justice: communities and their courts. Little and Brown, Boston

Feeley M (1983) Court reform on trial: why simple solutions fail. Basic Books, New York

Fleming R, Nardulli P, Eisenstein J (1992) The craft of justice. University of Pennsylvania Press, Philadelphia, PA

Friesen EC, Gallas EC, Gallas NM (1971) Managing the courts. Bobbs-Merrill, Indianapolis

Goffee R, Jones G (1998) The character of a corporation. Harper Business, New York

Hirschorn L (1994) Leading and planning in loosely coupled systems. CFAR, Cambridge, MA

Lipsky M (1976) Street level bureaucracy. Russell Sage, New York

Lynn L Jr (2003) Public Management. In: Handbook of Public Administration. Sage, London

Nardulli PF, Eisenstein J, Fleming RB (1988) The tenor of justice: criminal courts and the guilty plea process. University of Illinois Press, Urbana, IL

Nimmer R (1976) A slightly moveable object: a case study in judicial reform in the criminal justice process: the omnibus hearing. Denver Law J 48:206

Nimmer R (1978) The nature of system change: reform impact in criminal courts. American Bar Foundation, Chicago, IL

Ostrom BJ, Hanson RA (1999) Efficiency, timeliness, and quality. National Center for State Courts, Williamsburg, VA

Ostrom BJ, Hanson R (2010) Understanding and diagnosing court culture. Court Rev 45:104–109

Ostrom BJ, Hall DJ, Schauffler RY, Kauder N (2005) CourTools. National Center for State Courts, Williamsburg, VA

Ostrom BJ, Ostrom CW, Hanson RA Jr, Kleiman M (2007) Trial courts as organizations. Temple University Press, Philadelphia, PA

Quinn RE (1988) Beyond rational management. Jossey-Bass, San Francisco, CA

Quinn RE, Rohrbaugh J (1983) A spatial model of effectiveness criteria: towards a competing values approach to organizational analysis. Management Sci 29(March):363–377

Schein E (1999) The corporate culture survival guide. Jossey-Bass, San Francisco, CA

Schein E (2004) Organizational culture and leadership, 3rd edn. Jossey-Bass, San Francisco

Trial Court Performance Standards (1997) Trial court performance standards measurement system implementation manual. Bureau of Justice Assistance. U.S. Department of Justice

Wilson JQ (1989) Bureaucracy. Basic Books, New York

Managing Innovation of Policing

Marleen Easton
Faculty of Economics and Business
Administration, Ghent University, Ghent,
Belgium

Overview

This entry seeks to address some of the main issues in relation to the management of innovation in public policing without the pretense of being exhaustive. This entry will first contextualize innovation in public policing by paying attention to evolutions in our contemporary society. Second, the entry zooms in on the content of innovations in public policing and aspects of the process of diffusion. Third, attention is focused on challenges related to the process of managing these innovations. Finally, the entry concludes by suggesting some ideas for future research, which are needed to move forward in the field of managing innovation in public policing.

Introduction

> All innovations represent change but not all changes are innovative (Etter 1995, p. 278).

Managing the innovation of public policing is a challenge for police leaders around the world. These innovations are driven by evolutions in our contemporary society, and chance events are often related to the upsurge of reforms in this field. There is significant diversity in the content of innovations taking place. Furthermore, these innovations are conceptualized, studied, and measured in different ways. It is unclear what the effects of most innovations are on public police performance. Nevertheless, they are disseminated and mimicked all over the world. Police leaders have no choice but to get up to date. Knowledge on the content of the innovations is required to get ready for the process of adoption and diffusion. Unfortunately, concepts and ideas of innovation often lack a translation to practices. There is little coherence in broader understandings of what works, in which specific context, based on which implementation strategies, or how it was supposed to work but actually failed. This hampers fundamental changes in the functioning of the police and requires more research in the field of innovation in public policing. In this entry, we address each of these aspects more in detail.

By addressing the management of innovation in public policing, we enter the crossroad of literature on innovation and change management in public organizations, on the one hand, and literature on public police studies and innovations, on the other hand.

A first observation to be made is that the definition of the word “innovation” has subtly changed since the late 1960s. It is no longer enough to introduce a new idea, to be creative, to talk about innovation. The idea has to be brought into use, be successful, before innovation can be said to have taken place. Mulgan and Albury (2003, p. 3) define innovation as “new ideas that work” and to be more precise *successful innovation is the creation and implementation of new processes*, *products*, *services and methods of delivery which result in significant improvements in outcomes efficiency*, *effectiveness or quality*. The change in the content of this concept reflects the current societal pressure on public and private organizations all over the world to be successful in dealing with societal transformations.

Since the eighties, this evolution has been known as New Public Management in the public sector, which has featured the transfer of principles from the private sector to the management of public organizations. Principles such as external orientation, service delivery, partnership, debureaucratization, performance management, accountability, efficiency, and effectiveness are at the core of the movement and were used in (inter)national public sector reforms (Pollitt and Bouckaert 2011). These principles can also be found in the philosophy of community-oriented policing, which frequently has been operationalised as external orientation, accountability, empowerment, partnership and problem-solving.

When looking at high-performing innovative organizations in the public sector, leadership is one of the key characteristics.

> These leaders, these senior managers and lead professionals are hungry, thirsty, exploring all the time the external environment. These are people who are both externally and frontline oriented in their approach to management. And they maintain a 'split screen narrative'; they are interested in innovation but not for its own sake, rather they are concerned about how to continue to improve their day-to-day operations and services and products while at the same time building innovative capacity to address present and future challenges. (Albury 2011, p. 230)

Furthermore, there is a difference being made between management and leadership, which is crucial to the management of innovation. In essence, leadership is all about inspiring, motivating, coaching, and mobilizing individuals and groups. Leaders have a vision by which they give directions for the future. Although managers can be leaders, the management process as such is more about bringing consistency within the organization by planning, organizing, budgeting, controlling, and the like (Bennis 1993). From this perspective, innovation is mainly related to the leading capacities of managers in terms of being creative and entrepreneurial and which has been described as "transformational leadership" (Bass and Avolio 1990).

Although innovation might be considered to be the new buzzword in policy documents on public organizations, it is worthwhile to transcend fads and fashions in addressing the management of innovation in public policing (Abrahamson 1991). Therefore, it is essential to use a historical approach to this subject and to contextualize it in the dynamics of our contemporary society/societies. The role of the environment of any public organization in stimulating the demand for change is inevitably and should be taken into account. It is a lesson learned based on the contingency approach and (new) institutional perspective in the change management literature (Powell and DiMaggio 1991; Scott 2001; Maguire and Uchida 2000).

Contextualizing Police Innovation

It cannot be denied that the times we are living in today are characterized by constant change. Concepts such as "*postmodern* society," "*risk* society" or "*Information* Age," and the "*Network* society" are being used to refer to the major transformations in our global world. Changes in relation to our economy, politics, ecology, technology, demography, culture, religion, values, and social interactions are at the heart of it.

The combination of economic forms of globalization (interconnectedness of world financial markets, extensions of free trades, technological standardization, and internationalization), cultural globalization (through movies, fashions, and sporting events), and social changes in the pattern of life (increased life expectancy, changes in the patterns of family life, rise in the level of unemployment and migration flows, etc.) makes the nature and pace of change today unique (Pollitt and Bouckaert 2011).

Apart from the new opportunities and chances, these driving forces influence the problems arising in our societies. The nature of threats, risks, crime, criminals, and criminal organizations is changing, and the "localized citizens' demands" for more security and protection are rising all over the world. To answer these "demands," the public police has up until now gone through a period of significant change and innovation (Mastrofski and Willis 2010). Processes, services, and methods of delivery have been revised to improve the functioning of the public police. To get accustomed to new social media such as Twitter and Facebook, mobile communication, and the need to closely work together with other (security) actors in the field are just some of the challenges the public police face today.

The current European and North American (economic) crisis is adding greater complexity to it all. Castell and his network of academics talk about "the Aftermath" referring to an era before and after the economic crisis of 2008 (www.aftermathproject.com). It is a social crisis which is bringing about a fundamental transformation of societies at large. Citizens are being challenged to become creative in dealing with this new context and so are policy makers and (leading) managers in the public sector. The time has come to produce significantly better outcomes for significantly lower costs (Albury 2011). Austerity measures generate immense pressures on all public services, including the police. Instead of doing more with less, in some countries, the police might have to find a way to do less with more (involving more people and delivering more impact). This means that a smaller (less public money is available), smarter (better use of informational and human assets), and sharper (interventions targeted to where they can make a difference) style of policing is considered to be a possible way to deal with this challenge (Innes 2011, p. 80). The way in which this context may be fertile ground for innovations is still to be seen but worthwhile to follow up as it might depend on policy choices made by politics.

It is important to note that these driving forces may provide pressure, but they do not supply the ideas for innovation, which is to be generated by (leading) managers of public policing. The question arises: "what is yet to come and what will be the content of the innovations for policing in the future?" Groups of academics and/or practitioners working on a vision on policing for the future address these questions, and most of them rely on the societal transformations mentioned above to start thinking creatively about possible future challenges for the functioning of the police. Interesting in this respect is the RAND-publication on "Moving Toward the Future of Policing" and the activities of the "Pearls of Policing" network, which bring together police leaders from around the world to think about a vision on policing the future. These networks raise questions on what the main challenges will be for police work in the next decades. In their writings and conferences ideas, knowledge and experiences are shared which can be inspiring for law enforcement leaders around the world.

Finally, chance events such as scandals, disasters, accidents, and unpredictable tragedies are quite often related (as a catalyst) to the upsurge of reforms in the public sector (Pollitt and Bouckaert 2011). For example, the Marc Dutroux scandal (serial killer and child molester) and more specifically the lack of information exchange between the existing police agencies at the time was one of the reasons for the substantial reorganization of Belgium's law enforcement sector in 1998. With the law on the integrated police force (07/12/1998) structured on a local (local police) and a federal (federal police) level, the State Police, Municipal Police, and Criminal Investigation Department, who used to be separate agencies, were integrated. During the reorganization community policing was introduced by law. This illustrates the dynamic relation between the public police and its societal environment in terms of change and innovation.

In what follows, the content of public police innovations and some aspects of the process of diffusion are considered.

The Content and Diffusion of Innovations in Public Policing

Although there might be some overlap (Scott 2009), community policing, broken windows policing, problem-oriented policing, pulling levers policing, third-party policing, hot spots policing, CompStat, evidence-based policing, intelligence-led policing, telemetric policing, restorative policing, and zero-tolerance policing are all considered to be major innovations in (American) public policing developed during the last three decades (Weisburd and Braga 2006; Scott 2009). The classification as "major" refers to the influence of these innovations on the broad array of public police tasks and on the practices and strategies that broadly affected the public policing of American communities. In the wake of the September 11, 2001, terrorist attacks on New York and Washington *terrorist-oriented public policing* can be added to this list as one of the broad and quite complicated innovations (Mastrofski and Willis 2010, p. 117).

Nevertheless, there is some discussion on whether the innovations above are cultural and/or structural; incremental or radical; strategic, administrative, technical, or programmatic; and focusing on a single service, a department, or the total organization of the police (Damanpour 1991; King 2000). In relation to innovations, the question often persists whether it is rhetoric or reality. This discussion creates a division between “believers” and “nonbelievers” on each of these innovations between and within the community of scholars and practitioners. It is reflected in the lively debate in the literature of which Weisburd and Braga (2006) present a representative sample. Currently, however, it is not clear at all where the state of the art is in research on these innovations in public policing.

The notion of a policing model is found to be useful in making the complexity of innovations comprehensible. Ponsaers (2001, p. 473) roughly discerns four police models: the military bureaucratic model, the lawful policing model, the community-oriented policing model, and the public-private divide policing model. Each of these models is considered to have an internal-logical vision which has a critical function and involves basic assumptions in itself about the role and the place of the police in society. This includes a position in relation to key questions about police discretion, the role of the law, responsibility, relationships with the population, professionalization, legitimacy, prevention, and pro-/reactive police force policy. These are the key issues being influenced by the broader societal context mentioned above. From this perspective, these police models can be considered to be important innovations in policing.

The importance of these police models for the topic of managing police innovation is threefold. First, it shows that police innovation is related to the timeframe in which it takes place. The military bureaucratic model, now considered to be the traditional model, was innovative at the time in dealing with the problem of corruption, politicization, and the lack of professionalism. The postmodern “Public-Private Divide Model” fits into the current challenges for the police in an era of economic crisis and an evolving security field in which there is an increasing degree of interaction between public and private bodies. A second reason is that this approach makes it possible to distinguish these police models from other kind of concepts and to clear up the conceptual “mess.” For example, discerning four police models makes it possible to emphasize that the broken windows theory is not a model but a theory. In this line of reasoning, problem-solving/oriented policing and reassurance policing are a variant of the community policing model, and intelligence-led policing is simply the result of an evolution in technology (Ponsaers 2001).

The third reason why police models are important to address the issue of innovation is that police models are not to be considered as sequences in time, they are not chronological episodes but logic deductions. In other words, reform processes inhibit a programmatic choice and the danger of a return to more conservative police models is always present. Importantly, a police organization can be considered as a combination of police models. The combination will vary in relation to the political consensus and integration within society. In that respect, the responsibility of police chiefs and politicians is huge (Ponsaers 2001, p. 492).

Despite the discussions, it is important to note that most of the innovations mentioned above have to varying degrees been “exported” throughout the (Western) world. Overall, there is optimism among scholars as Weisburd and Braga (2006, p. 12) observe *police agencies have become open to the idea of innovation, and that new programs and practices have been experimented with and adopted at a rapid pace over the last few decades*. Efforts are being made to “translate” some of these concepts to different contexts (see, e.g., Ponsaers and Easton 2008) and to implement the innovations, often conceptualized in the Anglo-Saxon world, to other continents where public police units, departments, and agencies have their own structure and culture. At the same time, efforts are being made to integrate some of these innovations to avoid too many new notions in the world of public policing and to try to capture the overall picture (Ponsaers and Easton 2008; Scheider et al. 2009).

Around the world, this process of “translation” and “integration” is being done by organizing seminars in which attention is being paid to new concepts and evolutions in the field of policing and which are being discussed by academics and practitioners. Networks such as the Centre for Excellence in Policing and Security in Australia, the Asian Association of Police Studies, the African Security Sector Network, the UK Police Foundation, the Scottish Institute for Policing Research, The Dutch Society, Security, and Police Foundation, and the Center for Police Studies in Belgium are explicitly investing in these exchanges. An example are the efforts to explain the concept of "reassurance policing" and the link with "community policing" in order to inform police chiefs on these new evolutions. It was done during a seminar on "Reconsidering Reassurance Policing" in May 2009 organized by the Belgian Flemish Center for Police Studies in collaboration with the Dutch Society, Security, and Police Foundation and the UK Police Foundation.

The main reason why the "export" of such innovations remains quite remarkable is that there is no clear evidence on the effect the innovations have on police performance (Weisburd and Braga 2006, p. 339). "Does it work?" often remains the question which cannot always be answered on the basis of current empirical research in the field of policing. The "contrasting perspectives" on police innovations are fueled by the finding that the success of these innovations is seldom revealed due to the complex nature of police work (complex relation between outcome and effect of police work). Studies rather indicate that the standard model of policing continues to dominate the work of most public police agencies (Skogan and Frydl 2004). In other words, the public police are more likely to follow traditional police practices and models than to choose innovative approaches. Nevertheless, not everybody is that pessimistic about the public police. According to Mastrofski and Willis (2010, p. 116), generalization about change in relation to public policing is hazardous. They refer to the existing tension in academic research between the temptation to overstate the capacity of the public police to resist change and the temptation to overstate their susceptibility to change.

There is an urgent need for more (comparative) research work in this domain. Measures of innovations and change in public policing are far less developed than that available in innovation studies of other organization types. *Innovativeness studies* which quantify the adopted innovations and aim at explaining why some organizations adopt more innovations than others are scarce. It has limited understanding of why some departments are more innovative than others. There is in fact some disagreement about the impact of environmental factors, on the one hand, and internal, departmental factors, on the other hand (King 2000, p. 306). Furthermore, there is certainly a gap in the *diffusion studies* on innovations in public policing which describe and predict the spread of an innovation or innovations across a group of adopters, over time (King 2000, pp. 303–305). In relation to public policing, these studies conclude that innovations are slowly adopted by a few departments at first, and over time, the rate of adoption increases. Larger departments tend to adopt before smaller departments (Weiss 1997).

To sum up, there is significant diversity in the content of innovations taking place. Despite this diversity and the fact that it is unclear what the effects of most innovations are on police performance, these innovations are disseminated and mimicked all over the world. Without presuming that all innovations are transferable, the question arises as to what is known about the process of managing these innovations as management is a crucial aspect of any type of dissemination. In what follows, we further elaborate on this.

The Process of Managing Innovations in Public Policing

When Pettigrew (1992) stresses the interaction between the context, content, and process of changes, (leading) police managers are at the heart of these interactions. To learn more about the process of managing police innovation, it is interesting to have a look at *process studies* being done in the field of public policing in relation to some of the innovations mentioned above. *Process studies* explain the way in which an innovation is discovered, adopted, adapted, and employed by an organization. The study of the adoption of innovation has only recently become a subject of interest for police scholars (Skolnick and Bayley 1986; Weiss 1997).

Nevertheless, there have been few systematic studies of these processes, and scholars were generally not concerned about the emergence of innovation as a research problem when these innovations were being developed (Weisburd and Braga 2006, p. 11). As a consequence, there is remarkable little research on the innovation process or how it can be facilitated or implemented (Maguire 2009). There is likewise little systematic, cross-agency research on the extent and effectiveness of organizational change strategies in policing or on the role of police leadership in relation to these strategies (Skogan and Frydl 2004, p. 9).

There are, nonetheless, some challenges related to the management of innovation in public policing that need to be addressed. The complicated nature of this management process is related to the fact that police managers do not operate in a vacuum but find themselves in an "arena" of forces. Key players are the internal police **p**ersonnel and organization, the **p**ublic or the communities being policed, the **p**oliticians and the civil servants, diverse external **p**artners of the police, and the **p**ress or media. Numerous sources of resistance and challenges are related to this "5**p**" complexity (Skogan 2008).

A big challenge is related to the competing demands and high expectations of the public towards the police. Because time and money is limited, the police are often unable to meet the "list" which different communities generate. Police managers need to respond to priorities with the available means and effectively communicate with the public to achieve the necessary legitimacy and support. This task is further complicated by the competing nature of the demands coming from different communities. However, it is important to note that societies are quite different in this respect. The public police are not the first point of reference for all citizens around the globe. For example, South African citizens are contacting their neighbors or private security agencies first when confronted with security problems. The public police over there are not at all the first to be called in emergencies.

Furthermore, managers in police innovation need to build public and political support. Police innovation might not be isolated to the police organization alone. The innovation must be supported by the (local, federal, regional) network of (community) agencies that influence the functioning of the police. The failure of (interagency) cooperation can be a substantial source of resistance in relation to the numerous external (state and non-state) partners the police are now expected to work with. These partners have their own view on security and well-being which might not fit the assumptions of the police innovation (Reiner 1992; Skogan 2008). However, external support can be crucial as it might help the change manager to respond to internal opposition to the innovation that has to be implemented. Taking this into account, it might be wise to invest time in the transition of innovation to successors in leadership, in the police organization, and within the governmental and community environment (Skogan 2008, p. 33).

In addition to these external contextual elements, managing innovations in public policing implies addressing the traditional internal resistance to change. Changes to established workplace cultures and practices in police organizations are difficult to generate. This is related to the nature of public police work and the structures and cultures that this has generated. This is a culture that has been called traditional or conservative as it holds on to established practices and symbolic representations of "discipline" (Marks and Fleming 2004). This hinders attempts at developing more participatory management techniques which are crucial to some of the innovations mentioned above. A culture based on "top-down command and control" is, for example, resistant to take into account critical reflections of police personnel on internal or external work-related procedures.

Furthermore, managers have to take into account that within this culture, police unions play a role in resisting innovations which include performance measurement and augmented accountability mechanisms (Reiner 1992). This may be understood from their attempt to preserve the professional discretion of the rank-and-file in the police organization. The fact that the core of the innovations cannot easily be translated into actual practices is fueling this kind of resistance. This may also be related to the fact that it is difficult to measure and reward all aspects of police work. It is, for example, easy to measure interventions at traffic incidents on the basis of police reports, but it is much more difficult to measure and reward the efforts of police to strengthen their relationships with local communities. This complicates the implementation of innovations such as community policing that include important activities which cannot always be measured (Skogan 2008).

A significant internal source of resistance, which must not be underestimated, is the rank-and-file officers. They are often asked to leave their traditional roles and meet new and high expectations for which they do not always have the necessary skills and training. These officers are quite often suspicious about new ideas that they easily call "fads" and put aside to deal with "real" police work. Furthermore, middle managers are not always enthusiastic. Empowerment of the lower ranks, for example, which is a key element in community-oriented policing, is challenging police power which is traditionally built on command and control and doing everything by the book. Managers are asked to have trust in the professionalism of their police officers and encourage the exercise of police discretion. For established managers, this is a radical change that is not always welcome (Skogan 2008, pp. 24–25).

Being aware of this kind of resistance to change, managers need to fine-tune their strategies to support their personnel depending on their stage in the adoption of the innovation. On a Gauss curve, the stages of each employee may vary from being part of the small group of "innovators" and "early adapters," being part of the "early majority" or "late majority" which is representing most of the group to being a member of a small group of "laggards" (Rogers 1962). Reeducation might work in relation to innovators and early adopters, but most of all, persuasion strategies will be needed to get the rest going in the desired direction. In this regard, the police is far from close to the many innovators and early adapters in the world of crime and criminal organizations they are fighting against (Treverton et al. 2011).

The ultimate challenge is to incorporate new management practices into police organizations in such a way that these changes are experienced "as unremarkable as the air they breathe" (Shearing 1992 in Marks and Fleming 2004, p. 786). According to Toch (2008), reform can get considerable credibility if officers are enlisted as change agents, encouraging them to get involved in the design and implementation of change. In this line of reasoning, police unions can be considered as change agents also (Marks and Sklansky 2011). It might help if the utility function is taken into account. Police personnel, as cultural subjects, only subscribe to a certain kind of policing and policy if they also derive an individual or social added value from this (Easton and Van Ryckeghem 2010). Financial or promoting rewards for extra investments in relation to innovations can make a difference in the motivation for change. However, it is important to not allocate the innovation to a new unit within the organization or to ask people to allocate their extra overtime to the implementation of new practices (Skogan 2008, p. 27). An innovation such as community policing should be an overall strategy for the organization as a whole.

This complexity of issues and factors requires a flexible leadership style in which attention for continuity and innovation is combined. Purely instrumental (transactional) leadership will remain important to develop a working environment in which one can work effectively and efficiently. Transformational elements will be essential as well they empower people by formulating a vision and mission for the future of the police organization.

At times managers might find themselves confronted with even contradictory requirements on the organizational, strategic, and operational level of the police organization. The community policing model asks him/her to accept decentralized decision-taking, generalization as principle for labor division, a preventive orientation, a social strategy, and an open, proactive interaction with the environment. This may conflict with the demands of the military, bureaucratic model that stresses centralization, specialization, repression, and a closed system that is reactive in the interaction with the environment. Furthermore, a high degree of fragmentation in the decision-taking, specialization, a risk calculation, and contracting with consumers may be essential to implement the public-private divide model. In relation to various task environments, different police models coexist within a single police organization (Ponsaers 2001, p. 492). Police managers might find themselves "juggling" with inherently conflicting demands. This might be considered to be one of the main challenges in managing innovation in public policing.

To sum up, the challenge for police managers and leaders in the twenty-first century is to adapt their police organization to a demanding context and translate societal changes into the functioning of their organization by implementing innovations as mentioned above. The fact that different authors conceptualize, study, and measure police innovation in different ways generates different findings and makes the field of police innovation a "morass" wherein (leading) managers try to "survive." Although managing innovations in public policing inhibits challenges that are well known in the field, there is an urgent need for more empirical research on this topic.

Conclusion and Future Research

This overview makes clear that insights on the management of public police innovations might be "lost in translation." Firstly, innovators in policing have lent insights from public sector reform and were inspired by New Public Management. The application of the concept in the world of policing might need some further attention due to the specific nature of police work reflected in its culture and structure. Secondly, innovations tend to be captured in concepts and ideas for too long and lack a translation to practices within the police organization to be able to get implemented and radically change the functioning of the police. Thirdly, different authors conceptualize study and measure police innovation in different ways which generates different findings. It makes the empirical field of police innovation a morass and fuels a constant debate on how to manage police innovation. Fourthly, it is a challenge for academics and practitioners to develop and promote future innovations in a more integrated fashion to police. It is clear that more research is needed on the management of innovations in public policing.

Most innovation studies have not concentrated sufficiently on the dynamics behind police leadership, managerial, or tactical programs to provide understanding and/or guidance as to how these changes and innovations came about, how they were designed, what processes facilitated or hindered the finalization of project or program goals and objectives, and what outcomes were produced. The context in which such outcomes produce or fail to produce is often "black-boxed." There is a general absence of a body of literature that can be used to both identify the processes employed to design or implement police strategies or tactics which are part of major innovations, and the outcomes of those efforts. Simply put, policing has an oral tradition, where learning from the successes and failures of others has not been well routinized. Rather, police leaders emerge, have success and failures, and then often fall back into obscurity, without passing their learning and experience on to future generations of police leaders. This is an important conclusion for three reasons.

First, in the absence of collective understanding of environmental and institutional dynamics in policy choices and their implementation, police leaders are left to a hit-and-miss process, ill informed by current experience and knowledge. As such police chiefs often operate with policy and tactical blind spots, often mimicking what are thought to be national trends, whether or not those trends suit the particular police agency, or community in question. Replicating the practices of larger police agencies may actually do harm to smaller agencies, where there are perhaps greater ties between the police and the community. Moreover, appearing to look innovative in mimicking larger agency practices can actually lessen police effectiveness and the legitimacy the public accords the police in smaller communities. In short, mimicking bigger is not always better. Structures and practices of small agencies remain under examined (Mastrofski and Willis 2010, p. 57).

Second, given the large gaps between academic research and police policy and decision making, police chiefs may be uninformed about "what works" and therefore must rely on institutional and programmatic myths. It may be argued that the current police research literature itself suffers from questions of validity and scientific credibility, a criticism leveled by what is known as the evidence-based policing movement. Further, it can be asserted that what are called "best practices" are often absent in police discussion and in the preparation of a new generation of police leaders. Consequently, old practices, ideologies, and routines persist. This may account for the near uniform persistence in patrol and investigative practices that have little evidence of actually working, such as general patrol and general investigations.

Third, at the tactical level, there is no body of literature that actually assesses how certain tactics were called for, designed, implemented, and which impact they generate. There are assessments of parts of these questions scattered throughout the research and program evaluation literature, but there is little coherence in broader understandings of what works, in which specific context, based on which implementation strategies (how it was made to work), or how it was supposed to work but actually failed. The exception to this statement, perhaps, is in the area of problem-solving, where there has been an attempt to provide a systematic review (see, e.g., www.popcenter.org).

This makes clear that to support the management of police innovation in the future, two kinds of investments are essential. The first is the development of "evidence-based police practices" that support leaders in managing police innovations in the implementation process and confrontation with numerous constraints related to the functioning of the public police. The second is investment in future police leaders to have the necessary attributes, skills, and knowledge to manage it all. These investments can help future leaders to play a role in managing police innovation and translate broader societal transformations (sociological, economic, political, technological, etc.) into the functioning of the police organization of tomorrow taking into account its own character.

Related Entries

- ▶ Broken Windows Thesis
- ▶ Community Policing
- ▶ Compstat
- ▶ Evidence-Based Policing
- ▶ Hot Spots and Place-Based Policing
- ▶ Impacts of Community-Oriented Policing
- ▶ Intelligence-Led Policing
- ▶ Police Leadership Styles
- ▶ Problem-Oriented Policing
- ▶ Pulling Levers Policing
- ▶ Reassurance Policing and Signal Crimes
- ▶ Telemetric Policing
- ▶ Third-Party Policing

Recommended Reading and References

Abrahamson E (1991) Managerial fads and fashions: the diffusion and rejection of innovations. Acad Manage Rev 16(3):586–612

Albury D (2011) Creating the conditions for radical public service innovation. Aust J Publ Adm 70(3):227–235

Bass B, Avolio B (1990) Developing transformational leadership: 1992 and beyond. J Eur Ind Train 14(5):21–27

Bennis WG (1993) Managing the dream: leadership in the 21th century. In: Rosenbach WE, Taylor RL (eds) Contemporary issues in leadership. Westview Press, Boulder

Damanpour F (1991) Organizational innovation: a meta-analysis of effects of determinants and moderators. Acad Manage J 34(3):555–590

Easton M, Van Ryckeghem D (2010) Conceptualising the role of police culture in change strategies. In: Cools M, De Ruyver B, Easton M, Pauwels L, Ponsaers P, Vande Walle G, Vander Beken T, Vander Laenen F, Vermeulen G, Vynckier G (eds) New empirical data, theories and analyses on safety, societal problems and citizens perceptions, Governance of security research

papers series III. Maklu, Antwerpen-Apeldoorn, pp 135–153
Etter B (1995) Mastering innovation and change in police agencies. In: Etter B, Palmer M (eds) Police leadership in Australasia. Federation Press, Sydney, pp 278–309
Innes M (2011) Doing less with more. The ‘new’ politics of policing. Public Policy Res. 73–80
King WR (2000) Measuring police innovation: issues and measurement. Int J Police Strateg Manag 23(3):303–317
Maguire ER (2009) Taking implementation seriously: response to Mastrofski, Weisburd, and Braga. In: Frost NA, Freilich JD, Clear TR (eds) Contemporary issues in criminal justice policy. Wadsworth, Belmont, pp 265–268, Papers from the American society of criminology conference
Maguire ER, Uchida CD (2000) Measurement and explanation in the comparative study of American police organizations. In: Duffee D (ed) Criminal justice 2000, volume 4: measurement and analysis of crime and justice. National Institute of Justice, Washington, DC, pp 491–557
Marks M, Fleming J (2004) As unremarkable as the air they breathe? Reforming police management in South Africa. Curr Sociol 52(5):784–808
Marks M, Sklansky D (eds) (2011) Police reform from the bottom up: officers and their unions as agents of change. Police Practice and Research, Routledge, London
Mastrofski SD, Willis JJ (2010) Police organization continuity and change: into the twenty-first century. J Crime Justice 39(1):55–144
Mulgan G, Albury D (2003) Innovation in the public sector. Paper strategy unit, cabinet office, London. www.cabinetoffice.gov.uk, version 1.9 October 2003 consulted 15 Feb 2012
Pettigrew A, Ferlie E, McKee L (1992) Shaping strategic change: making change in large organizations, the case of the NHS. Sage, London
Pollitt C, Bouckaert G (2011) Public management reform: a comparative analysis: new public management, governance, and the neo-Weberian state. Oxford University Press, Oxford
Ponsaers P (2001) Reading about “community (oriented) policing” and police models. Int J Police Strateg Manag 24(4):470–496
Ponsaers P, Easton M (2008) Community (oriented) policing reassured: significance within a Flemish context. In: Easton M, Gunther Moor L, Hoogenboom B, Ponsaers P, Van Stokkom B (eds) Reflections on reassurance policing in the low countries. Boom: Eleven International Publishing, Reeks Het groene gras, Den Hague, pp 31–51
Powell WW, DiMaggio PJ (eds) (1991) The new institutionalism in organizational analysis. University of Chicago Press, Chicago
Reiner R (1992) The politics of the police. Harvester Wheatsheaf, Hemel Hempstead
Rogers EM (1962) Diffusion of innovations. McMillan, New York
Scheider M, Chapman R, Shapiro A (2009) Towards the unification of policing innovations under community policing. Policing 32(4):694–718
Scott WR (2001) Institutions and organizations, 2nd edn. Sage, Thousand Oaks
Scott MS (2009) Progress in American policing? Reviewing the national reviews. Law Soc Inq 34(1):171–185
Skogan WG (2008) Why reforms fail. Polic Soc 18(1):23–34
Skogan W, Frydl K (eds) (2004) Fairness and effectiveness in policing: the evidence committee to review research on police policy and practices. www.nap.edu., National Research Council, consulted 15 Feb 2012
Skolnick JH, Bayley DH (1986) The new blue line: police innovation in six American cities. Free Press, New York
Toch H (2008) Police officers as change agents in police reform. Polic Soc 18(1):60–71
Treverton GF, Wollman M, Wilke E, Lai D (2011) Moving toward the future of policing. RAND Cooperation, National Security Research Division
Weisburd D, Braga A (2006) Police innovation: contrasting perspectives. Cambridge University Press, Cambridge
Weiss A (1997) The communication of innovation in American policing. Int J Police Strateg Manag 20(2):292–310

Manual

► Automated and Manual Forensic Examinations

Marxist Criminology

Rick A. Matthews[1] and William J. Chambliss[2]
[1]Carthage College, Kenosha, WI, USA
[2]George Washington University, Washington, DC, USA

Synonyms

Critical criminology; Radical criminology

Overview

Despite the fact that Marx did not address crime in a systematic way, criminologists have used Marxist theory to analyze laws, crime, and the criminal justice system. Over the past 40 years, Marxist criminology has become a core component of what has been broadly referred to as critical criminology. Many critical criminologists have contributed to the development of Marxist criminology, and its position within the field of critical criminology is well established. At the same time, the insights provided by Marxist criminological theory have received little interest among mainstream criminologists, despite the fact that some share with Marxist criminologists a macrolevel or structural approach to studying crime. While recent developments in Marxist criminological theory have slowed, Marxist criminologists are still able to provide a viable framework to understand law, crime, and state responses to crime.

Marx and Crime

The works of Karl Marx, Friedrich Engels, Antonio Gramsci, Emma Goldman, Willem Bonger, and a host of others who have stood on their shoulders provide what the philosopher of science T.S. Kuhn aptly named a "paradigm." A paradigm is a general theoretical perspective composed of concepts, hypotheses, and perspectives from which political, economic, and social relations can be analyzed, explained, and understood. The Marxian paradigm, which serves as a foundation for a corpus of knowledge in the social sciences generally, and criminology in particular, is thus a very general perspective from which specific (although sometimes conflicting) theories are developed. With such a large fountain of ideas, it is not surprising that no one "Marxist theory" can claim an exclusive interpretation.

Despite the fact that there are a wide variety of "Marxist theories," there are two fundamental characteristics of Marxist theories that tie them all together and differentiate Marxist theory from other paradigms. First is the belief that the material conditions of a particular time and place – not norms or values of culture – are the starting point for an analysis of why political, economic, and social relations have the shape and texture they do. Second is the assumption that social theory generally, and criminological theory in particular, is best served by approaching the subject from a *structural perspective*.

The structural perspective is of course not unique to Marxist theory. Emile Durkheim argued, in his classic text *Suicide*, that individual explanations of suicide were impossible to come by, but explanations of why suicide rates varied according to social structural characteristics were not only possible but also scientifically defensible. Durkheim, however, did *not* see the importance of material conditions in determining suicide rates (social class is never mentioned in the entire book). On the contrary, Durkheim based his analysis on the importance of norms, specifically normative integration, a paradigm that has dominated criminological theory ever since.

The Marxian paradigm stressing material conditions has influenced many theoretical paradigms in criminology, albeit often rather tentatively. Edwin Sutherland's essentially psychological theory of the cause of crime explained individual criminality by the degree of association an individual had with "attitudes favorable or unfavorable to the commission of crime." The implicit assumption was that people from the lower classes (thus material conditions) were more likely to have associations with attitudes favorable to the commission of crime. The theory was, however, basically normative in that it was attitudes and norms that were the deciding factors in the etiology of criminal behavior.

Robert Merton's strain theory is likewise a combination of Marxist materialism and Durkheimian emphasis on norms. The "strain" in Merton's theory derives from the location of groups or individuals in the social structure relative to their access to legitimate means for achieving culturally prescribed goals. It is not material conditions per se; it is material conditions as they relate to cultural norms (goals and means) that provide the impetus for crime.

Sutherland and Merton exemplify the tendency of criminological theory to emphasize the importance of norms and to give passing reference to the fact that material conditions (in particular social class) may contribute to the learning of social norms (Sutherland) or the experience of social strain (Merton).

Herein lies the most fundamental difference, then, between the Marxist paradigm and that of other criminological theories. The Marxist paradigm, with few exceptions, does not seek to explain why individuals commit crime (for exceptions, see Bonger, 1916; Colvin and Pauly, 1983) but seeks to explain why different types of crime are situated where they are in the social structure and why different economic systems (e.g., capitalism and socialism) produce different types and rates of crime. The starting point for Marxist theory is not the norms, values, and culture of the historical period. For Marxist theory the starting point for explaining any social phenomenon lies in the material conditions and the social structure of a particular historical period.

Social Order and Crime

Since Marxist criminologists look to social order when explaining crime and societal responses to it, they begin with an analysis of the existing social arrangements. For Marxist criminologists, the laws that define behaviors as crimes emerge from the existing social order. In this view, social order comes before the laws that define behaviors as crimes (Michalowski 1985). This is very different from mainstream criminological theory that assumes crime creates the need for laws, which then in turn produce social order (Michalowski 1985).

Marxist criminologists have argued that political economy is dynamic and develops in stages over time. Thus, the particular form of capitalism may change depending on historical social and political developments. For example, the emergence of the welfare state changed many capitalist economies in that politicians decided to add (or strengthen) social safety nets, such as unemployment benefits and welfare for the poor. Marxists have argued that the development of the welfare state was necessary to stem the tide of class conflict that arose from growing class inequality and labor unrest. Marxists have argued that because capitalist economies contain inherent contradictions, they must constantly undergo change to address the problems created by them. In many cases, the state intervenes to dampen the effects of particular social problems created by these contradictions. However, these attempts to dampen the effects of structural contradictions create new dilemmas which then must be resolved, thus continuing the process of reproducing capitalism Chambliss (1993).

Marxist criminologists have been careful to note that some capitalist systems are more likely to be criminogenic than others. For instance, some capitalist societies have stronger social safety nets that reduce inequality and provide basic necessities like health care to its members (e.g., Western European capitalism as opposed to the capitalism found in the United States). Capitalist countries that provide stronger social safety nets also produce different cultural values that are less likely to promote the welfare of the individual over that of the community.

Beginning the analysis of crime from a structural, material paradigm leads to different theories than does beginning from an individualistic, normative paradigm. The structural-materialist paradigm also generates very different *questions.* For example, criminological theory generally assumes what the Marxist paradigm seeks to explain: Why are acts defined as criminal? Why are some people more likely to be labeled and punished than others? What effect does the classifying of some acts as crimes have on economic, political, and social relations? How does the classification of acts and people as criminals serve to perpetuate and maintain the class structure of the society (be in capitalist, socialist, or feudal)? What interests are served by the particular criminal law system in effect at different times and places?

The Development of Marxist Criminology

Neither Marx nor his writing partner, Friedrich Engels, developed a systematic theory of crime. Rather, Marx chose to focus his attention on analyzing capitalism as one stage in the larger history of economic development. While Marx was very critical of the negative aspects of early industrial capitalism, such as the exploitation of the working class by industrialists, he also viewed capitalism as a necessary stage of development that would create the technology and industrial capacity to lay the foundation for a more just society based on mutual benefit, rather than the exploitation of one class (workers) by another (owners). Crime, for Marx, was a social problem like any other in capitalist societies.

Marx believed that capitalism is an inherently unstable system that is prone to crises. For example, capitalism needs competition between business owners, since without it, monopolies emerge that control markets and prices. However, as one company becomes powerful, it can eliminate competition by buying up competitors, or simply crushing them with vast resources (e.g., a large retailer, because of enormous purchasing power, is often able to sell its products at a much lower price than small retailers). Eventually, the state has to step in to prevent this type of monopoly capitalism, passing antitrust and other laws. Thus, Marx viewed each crisis in capitalism as the result of some underlying contradiction in the system. As contradictions give rise to crises, someone (usually the state) must intervene to save the capitalist system from its own destructive tendencies.

Subsequent Marxist theorists have argued when public policies and social institutions are created to resolve one dilemma arising from an inherent contradiction in the system, another one emerges. Because capitalist economies are crisis prone, they undergo stages of growth and decline, leading to boom cycles characterized by economic growth and prosperity, as well as bust cycles characterized by a lack of economic growth, high unemployment, and increased poverty. These cycles influence both crime and societal responses to it. As changes are made to the system to reduce the problems created by one dilemma, new problems emerge that must be dealt with.

One of the first scholars to apply Marx's theories to crime was the German criminologist Willem Bonger. Bonger (1916) argued in his book *Criminality and Economic Conditions* that crime in capitalist societies arose from egoism (self-interest). Bonger (1916) viewed capitalism as an economic system that tended to bring out the most egoistic, and least altruistic, tendencies in people. Bonger (1916) was careful to argue that egoism does not, in itself, cause people to become criminal. Rather, he argued that certain environments make people more capable of crime. Related to the egoistic tendencies of capitalism, Bonger (1916) also claimed that the capitalistic economic system of exchange, that encourages people to act primarily out of self-interest, weakens social bonds. In such systems, the interests of the individual are not balanced with the interests of the collective and create an environment of intense competition and individualism. Much crime, then, occurs because the social bonds that tie the interests of people together are weakened in favor of an ethos of individualism and self-interest.

Another important and early development in Marxist criminology was *Punishment and Social Structure*, by George Rusche and Otto Kirchheimer (1939). Rusche and Kirchheimer (1939) continued advancing a Marxist analysis of crime in their study of punishment. They argued that capitalist systems create forms of punishment that correspond to their productive systems and that large pools of surplus labor in the market are likely to lead to more severe penal policies (i.e., large pools of unemployed people are more likely to result in longer prison sentences).

The 1970s saw the emergence of a sustained interest in Marxist theory and crime by critical criminologists. One of the most significant attempts to develop a Marxian framework for studying crime was *The New Criminology*, by Taylor et al. (1973). Giving primacy to a structural approach to understanding crime, Taylor et al. (1973) challenged criminologists to avoid reductionist explanations of crime that focused on individual causes of crime while providing a strong case for developing a theory of crime that located its causes within the existing political economy.

Structuralist and Instrumentalist Marxists

Early attempts to develop Marxist criminology saw debates emerge between scholars who had differing views on the role of the state in making laws and resolving crises within the capitalist system. On one hand, those such as Quinney (1974) argued that the purpose of laws and the criminal justice system was to maintain the class interests of the powerful. In this view, the state has little autonomy from the ruling class, and laws are instruments used to advance their interests. Others, such as Greenberg (1981) and Chambliss (1993), argued that the primary purpose of the state is to preserve the system itself. Greenberg (1981) made a compelling case that the ruling class is not as cohesive as Quinney and other instrumentalists had argued, and Chambliss (1993) advanced a comprehensive analysis of contradictions, showing that at times, the state passes laws that do not directly favor the ruling class, but are nonetheless intended to stem the negative consequences of a particular crisis in the system.

While the instrumentalist perspective provided a useful starting point to analyze the role of the ruling class in law making, today most Marxist criminologists reject this view. While it is true that the interests of the ruling class are often promoted in the social policies of the state, most Marxists view the overarching goal of the law, and the law making process, as a means to preserve the capitalist system.

Who Are the Criminals?

One of the important contributions of Marxist criminology has been the ways in which it has expanded the study of crimes to include crimes of the powerful. Many early attempts to develop Marxist criminological theory focused on crimes committed by the lower classes (e.g., Bonger 1916; Rusche and Kirchheimer 1939). While the early Marxists studied street crimes within the broader context of political economy, other Marxist criminologists later turned their attention to the study of corporate, state, and state-corporate crimes (e.g., Michalowski and Kramer 2006).

Since Marxist criminologists believe that the origin of all crime can be found in the political economy, it may be tempting to view crime and other social problems arising from the inherent contradictions of capitalism as necessary and positive features that will eventually cause capitalism to collapse. The danger in this view is that street crime can become romanticized, either as a symptom of system crisis necessary for collapse or as some type of rebellion against the system. In this latter case, criminals are viewed as forces resisting capitalist domination.

As left realists and other Marxist criminologists have warned, this view of crime is counterproductive. If criminologists have learned anything over the past 100 years, it is that most victims of street crime are very similar, demographically, to those engaging in crime. Thus, poor minorities living in urban areas are likely to be victimized by other poor minorities living in their neighborhoods. Many left realists have pointed out that the most harmful effects of street crime are not realized by the ruling class – which in many cases can insulate itself from street crime – but by other members of society that are exploited by the system. In the view of left realists and many Marxist criminologists, there is nothing romantic or revolutionary about street crime.

The fact that many Marxist criminologists have examined the crimes of the powerful, combined with the fact that they do not view street crime as revolutionary, does not mean that they have ignored it (see Carlson and Michalowski 1997). In fact, Marxist criminologists have studied all different kinds of crime, from crimes on the street to crimes in the suites.

Crime in the Streets

When explaining traditional forms of street crime, Marxists look to the ways in which economic conditions foster crime. For example, some areas have much higher levels of crime than others, and a Marxist explanation for those differences would begin with an analysis of how factors like poverty, persistent unemployment, and other “structures of disadvantage” play important roles in creating a criminogenic environment in these areas (Taylor 1999).

Since capitalist systems are dynamic, the amount and type of crime may vary from one period to the next and from one country to the next. For example, some capitalist economies have relatively high unemployment rates but lower levels of serious poverty because of public assistance programs (e.g., many countries in Western Europe). Other countries, such as the United States, have persistent unemployment and poverty, with fewer public dollars spent on social welfare. In addition, these structural differences give rise to significant differences in cultural values. For example, the egoism that Bonger (1916) identified differs from one culture to the next. In the United States, serious "structures of disadvantage" are coupled with an ethos of individualism, egoism, and extreme competition for scarce resources. In many urban areas in the United States, it is common to see highly concentrated poverty and unemployment, with little or no opportunity for success in the formal economy. Therefore, it is not uncommon to find underground economies in these areas, where drugs and other illegal goods and services are exchanged.

Crime in the Suites

While the social costs of street crime are significant, the costs of corporate crime are even higher. From a Marxist perspective, the crimes of the powerful are just as important as street crimes and have the same genesis in the political economy.

If capitalism reduces the altruistic tendencies of people and increases their competitiveness, the same is even more so for corporations, since their very survival depends on competition and making as much profit as possible. For corporations, making a profit trumps all other considerations. One of the clearest ways in which this can be seen is in the harm that is caused when corporations circumvent regulatory laws in order to maximize profits. Michalowski and Kramer's (2006) anthology documents many cases where corporate crimes were committed because regulatory laws were either circumvented or simply not enforced by governmental agencies.

Profiting from Crime

The social responses to crime in the United States have resulted in the highest per capita imprisonment rate (750 per 100,000) of any advanced industrial nation (Cowling 2008). With only 5 % of the world's population, the United States imprisons 25 % of the world's prisoners (Herivel and Wright 2007).

Carlson et al. (2010) have argued that the roots of this problem can be found in the economic crisis of the 1980s. During the 1980s, many areas in the United States underwent deindustrialization, where manufacturers moved their plants overseas, in large part to save money on labor and regulation. Deindustrialization caused unemployment to rise dramatically in many industrial cities and devastated local communities. In many areas, drug markets emerged, followed by violent crime. Politicians of all stripes endorsed "get tough on crime" policies, resulting in mandatory sentencing, longer sentences, and exploding prison populations.

Furthermore, Carlson et al. (2010) argued that the net effect of the "get tough on crime" policies that emerged from the economic crisis of the 1980s resulted in increasing numbers of people from the surplus population being incarcerated. In addition, they noted that more money has been spent on the criminal justice system itself, increasing amounts of which have been channeled to private corporations that either run prisons or provide key services for them.

Politics and Crime

Marxist criminologists have long argued that politicians have profited from crime in using "get tough on crime" platforms to win elections. Beginning with the "war on crime" that began in the 1960s, there has been a successive line of presidents and politicians who have advanced policies leading to increased incarceration rates as well as increased spending on all aspects of the criminal justice system.

For example, Chambliss (2001) noted that increased expenditures for the criminal justice system compete directly with other important expenditures such as education, programs for the poor, and other social services. While not Marxist in orientation, Jonathan Simon's (2007) *Governing Through Crime* extends Chambliss' (2001) earlier work by arguing that the focus on crime has become so normative that hardly anyone questions the ways in which crime control policies have radically changed American institutions. While states have been quick to cut funding for education during tough times, they have been slow to cut expenditures on criminal justice – even when crime rates are falling.

Mass Incarceration

Criminologists have noted that one of the most obvious results of the various "wars on crime" that have been waged over the past 50 years is the enormous increase in the prison population in the United States. Poor, young, urban, minority men have disproportionately been incarcerated, leading to startling inequalities in incarceration rates between whites and other minority groups. As Western (2006) has shown, young African-American males are six to eight times more likely to be incarcerated than their white counterparts. This alarming trend has been termed "mass incarceration."

Young minority males have been systematically excluded from economic opportunities, experiencing unemployment rates and poverty rates far higher than whites. Confined to the poorest urban neighborhoods, many minority males have not had access to well-paying manufacturing jobs that disappeared during the wave of deindustrialization during the 1980s. Excluded from the legitimate economy, many of these young men have turned to the underground economy.

Rather than addressing the underlying economic realities of urban areas and their structural problems, politicians have decided to become more punitive, passing a series of "get tough on crime" bills. Aside from incarcerating large numbers of poor, minority males, these policies have resulted in strained budgets and cuts in other essential programs like education and public assistance to the poor.

From a Marxist perspective, mass incarceration is not only a matter of crime and politics but also indicative of larger economic trends. Carlson et al. (2010) have argued that the current trend of imprisonment stems from repressive policies, not necessarily linked to increased crime. For example, mass incarceration serves to reduce the surplus labor force while simultaneously increasing spending in the criminal justice system, employing more people, and increasing the need for goods and services (Carlson et al. 2010). As Carlson et al. (2010) noted, this "imprisonment binge" has "served capital accumulation indirectly by restraining large segments of the surplus population from posing public problems in the form of crime, and directly by serving as a vehicle for private profit realization" (p. 256).

Social Structures of Accumulation and the Neoliberal Project

From a Marxist perspective, the periods of growth and decay in capitalism repeat themselves because of the inherent contradictions within the system itself. Some stages of capitalism have been characterized by economic growth and relative prosperity (i.e., the years after WWII). Other periods, like the Great Depression during the 1930s, were characterized by severe problems such as mass unemployment and deep poverty. Because capitalism is crisis prone, no period of growth lasts forever. Eventually, a crisis emerges that forces a new stage of capitalist development. Each stage of capitalist development, however, relies on institutional structures to support it. In some cases, this means that institutional structures undergo significant change to support capital accumulation. In other instances, social institutions are intentionally weakened to promote capital accumulation (e.g., deregulating the financial sector that leads to the most recent economic crisis). Social Structures of Accumulation (SSA) theory, which has its roots in Marxist theory, views important social institutions – such as the criminal justice system – as playing vital roles in undergirding any sociohistoric stage of capitalist development.

For over 30 years, neoliberal politicians have sought to weaken the welfare state and the social institutions and structures that have been created to support it. While the philosophical roots of this movement have been around for a long time, the period of economic stagnation of the 1970s created a political opportunity to advance neoliberal economic theory. In the early years, champions of neoliberalism such as Ronald Reagan and

Margaret Thatcher sought to wage a war on government itself, characterizing it as the enemy of economic prosperity and personal freedom. Other important components of the neoliberal project have included efforts to reduce taxes (particularly for the rich), the deregulation of business and industry, transferring collective risk to individuals, the decline in labor rights and the power of unions, the promotion of privatization, and perpetual budget crises caused by reducing taxes while increasing spending (see Harvey 2010). The most radical version of neoliberal politics can be seen in the tea party movement in the United States, with its extreme forms of privatization, antitax, and antigovernment rhetoric.

The neoliberal movement has brought with it repressive policies of increased social control that have filtered through many social institutions, most notably the criminal justice system. Carlson et al. (2010) have argued that the criminal justice system, while not a core social institution, plays a significant role in the accumulation of capital by controlling crime and civil disorder and thereby creating a good investment climate. In addition, the growth of mass incarceration has kept significant numbers of the "surplus population" of young men behind bars (Carlson et al. 2010).

From a Marxist perspective, neoliberalism has been a significant force that has given rise to the increasing privatization of prisons, resulting in "prisons for profit." Private sector prisons, like other businesses, are not run for the collective good. Rather, the goal for privately owned prisons is to make a profit. Critics of the neoliberal project have noted that private prisons transfer public money to private stakeholders, shifting the (ostensible) purpose of prisons from protecting the public to making money (Herival and Wright 2007).

On the other end of the class spectrum, some have argued that neoliberalism has resulted in the deregulation and the weakening of regulatory agencies that were created to prevent various forms of corporate crime. In this sense, the neoliberal project has created many opportunities for corporations to engage in criminal activity, such as the financial fraud that lead to the meltdown of the housing market.

The overarching ideology of the neoliberal movement has been that invisible market forces are better at controlling the behavior of people and institutions than laws or regulations. As an extreme form of Bonger's (1916) egoism, the ideology of market forces has entered nearly every facet of human relations, resulting in the radical change of traditional forms of social control such as community norms and other social institutions, like the family and education. As Currie (2004) noted, this trend has resulted in an extreme version of social Darwinism that can be seen throughout contemporary social institutions, ranging from "sink or swim" families where adolescents are left to fend for themselves to the draconian zero tolerance policies found in schools.

The net result of neoliberalism has been an enormous transfer of wealth from the bottom of the class structure to a very small group at the top, against the backdrop of various "wars" (i.e., crime, drugs, terrorism) that have, in many significant ways, undermined democratic institutions and processes. To paraphrase Reiman and Leighton (2010), during the neoliberal era, the rich have gotten much richer, while the poor have gotten prison.

In a very paradoxical way, however, the current politics of crime do not fit with the neoliberal trend of shifting collective risk from the collective to the individual. When it comes to crime, neoliberals are very supportive of strong government intervention to provide collective security. Yet, as Simon (2007) has noted, very few politicians have questioned the fact that "neither contemporary liberal nor conservative principles extol the kind of penal state and gated civil society we are building by governing through crime." From a Marxist perspective, such paradoxical views can only be understood as ideology, the purpose of which is not to promote some principle like individualism but is rather an attempt to promote the particular interests of the ruling class (preserving a system of capital accumulation) by extolling the virtues of collective risk in one situation (crime) while ignoring them in another (e.g., attempts to privatize social programs like social security).

The Future of Marxist Criminology

The most recent and comprehensive overview of Marxist criminology is Cowling's (2008) *Marxism and Criminological Theory*: *A Critique and a Toolkit*. Cowling (2008) argued that Marxist criminology can be advanced by further refining its analysis of the ways in which capitalism reproduces itself, including both the roles that criminal justice institutions and laws play in this process. Given the recent crisis of the global capitalist system following the housing market collapse, Marxist scholars find themselves with ample opportunity to study the ways in which states attempt to respond to the inherent contradictions within capitalism and in doing so provide insight to significant social problems like class inequality and crime.

The current crisis of capitalism provides several potential areas of inquiry for Marxist criminologists. For example, Marxist criminologists might examine the enormous impact of the financialization of the US economy over the past 30 years and the ways in which the exponential growth in the financial sector has led to the current economic crisis. There is little doubt that the neoliberal policies that emerged in the 1980s eventually lead to the deregulation of the banking and finance industry, creating the opportunity for banking and finance to take tremendous risks in order to increase profits. The rewards of these risks were privatized, but as the subsequent collapse showed, the costs were socialized – in the form of a $700 billion taxpayer bailout. Given the tremendous cost and harm created by the banking industry, why is it that there has not yet been a single prosecution? In addition, a Marxist analysis could provide insight to the fact that very little legislation has been passed to curb the practices of the banking and finance industries.

Marxist criminologists can also benefit from the good analysis that is currently being provided by both general Marxist theorists and other critical criminologists. For example, few Marxist criminologists have carefully examined the ways in which current Marxist theorizing can be applied to crime (c.f., Carlson et al. 2010; Carlson and Michalowski 1997). These theoretical insights could easily be applied to the work that has been done by mainstream criminologists. Take, for example, mass incarceration that has continued, unabated, for over 10 years. The great paradox of mass incarceration is that incarceration rates have steadily increased, while crime rates have decreased. Clearly, there is amble room for Marxist criminologists to contribute to the understanding of mass incarceration by providing a structural explanation that is grounded in political economy.

Also, Marxist criminologist would do well to address the criticisms leveled at Marxist criminology by other critical criminologists who have been critical of the fact that Marxist criminologists have been slow to adequately address important issues like gender and race.

Finally, while there have been several varieties of Marxist criminology over the years, and the amount of scholarship in the area has waned in recent years, Marxist criminology still provides a useful, macrolevel approach to understanding the problem of crime. As long as capitalism and the social problems like crime that it generates exist, Marxist criminology will continue to be relevant.

Related Entries

- ▶ Anomie and Crime
- ▶ Corporate Crimes and the Business Cycle
- ▶ Crimes of Globalization
- ▶ Left Realism
- ▶ Political Crime
- ▶ Postmodern Criminology
- ▶ Poverty, Inequality, and Area Differences in Crime
- ▶ Punishment and the State
- ▶ State Crime
- ▶ State-Corporate Crime

Recommended Reading and References

Bonger W (1916) Criminality and economic conditions. Methuen and Company, London

Carlson SM, Gillespie MD, Michalowski RJ (2010) Social structures of accumulation and the criminal justice system. In: McDonough T, Reich M, Kotz DM (eds)

Contemporary capitalism and its crises: social structure of accumulation theory for the 21st century. Cambridge University Press, New York, pp 239–263

Carlson SM, Michalowski RJ (1997) Crime, unemployment, and social structures of accumulation: an inquiry into historical contingency. Justice Q 14(2): 209–241

Chambliss WJ (1993) On lawmaking. In: Chambliss WJ, Zatz M (eds) Making law: the state, the law and structural contradictions. Indiana University Press, Bloomington, pp 3–35

Chambliss WJ (2001) Power, politics, and crime. Westview Press, Boulder

Colvin M, Pauly J (1983) A critique of criminology: toward an integrated structural-Marxist theory of delinquency production. Am J Soc 89(3): 513–551

Cowling M (2008) Marxism and criminological theory: a critique and a toolkit. Palgrave Macmillan, London

Currie E (2004) The road to whatever: middle class culture and the crisis of adolescence. Metropolitan Books, New York

Greenberg D (1981) Crime and capitalism. Mayfield Publishing Company, Palo Alto

Harvey D (2010) The enigma of capital and the crises of capitalism. Oxford University Press, Cambridge

Herivel T, Wright P (eds) (2007) Prison Profiteers: who makes money from mass incarceration. The New Press, New York

Michalowski RJ (1985) Order, law, crime: an introduction to criminology. McGraw-Hill, New York

Michalowski RJ, Kramer RC (eds) (2006) State-corporate crime: wrongdoing at the intersection of business and government. Rutgers University Press, Piscataway

Quinney R (1974) Critique of legal order. Little, Brown and Company, Boston

Reiman J, Leighton P (2010) The rich get richer and the poor get prison: ideology, class and criminal justice. Allyn and Bacon, Boston

Rusche G, Kirchheimer O (1939) Punishment and social structure. Columbia University Press, New York

Simon J (2007) Governing through crime: how the war on crime transformed American democracy and created a culture of fear. Oxford University Press, Cambridge

Taylor I (1999) Crime in context: a critical criminology of market societies. Polity Press, Cambridge

Taylor I, Walton P, Young J (1973) The new criminology: for a social theory of deviance. Harper and Row, New York

Western B (2006) Punishment and inequality in America. Russell Sage, New York

Mass Transit

► Crime on Public Transport

MAT

► Drug Abuse and Alcohol Dependence Among Inmates

Measurement of Self-Control

Chris L. Gibson
Department of Sociology and Criminology & Law, University of Florida, Gainesville, FL, USA

Overview

Gottfredson and Hirschi's concept of self-control is one criminological construct that has sparked a large amount of theoretical and empirical debate among criminologists. The interest in measuring self-control and understanding the psychometric properties of self-control measures is perhaps due to the explanatory power that Gottfredson and Hirschi (1990) attribute to self-control and the theoretical propositions they make regarding this construct. It is therefore surprising that little effort has been made to systematically examine and compile the state of criminological research on the measurement of self-control. This entry is designed to fill this void by providing a review of several issues related to the measurement of self-control and empirical studies on self-control measurement.

This review will accomplish several tasks. First, conceptual and operational definitions of self-control will be discussed. In doing so, this entry highlights the disagreements on whether the concept should be operationalized in terms of attitudes and behavior and whether it should be treated as unidimensional or multidimensional. Second, although different measures of self-control have been used by criminologists, this entry focuses extensively on Grasmick et al.'s 24-item, attitudinal scale, which is arguably the first and most commonly used measure of self-control. Finally, directions for future research are offered that cover a broad range of topics related to the conceptualization and measurement of self-control.

Self-Control Definitions

Gottfredson and Hirschi (1990, p. 89) provide a meticulous account of the six "elements of self-control." Those lacking self-control will have a "concrete here and now orientation"; "lack diligence, tenacity, or persistence in a course of action"; are "adventuresome, active, and physical, are indifferent, or insensitive to the suffering and needs of others"; and "tend to have minimal tolerance for frustration and little ability to respond to conflict through verbal rather than physical means" (Gottfredson and Hirschi 1990, pp. 89–90). They (Gottfredson and Hirschi 1990, pp. 89–90) link each element of self-control to the criminal act. First, the "here and now" orientation reflects the immediate gratification provided by crime, and those with low self-control have an inclination to respond to tangible stimuli in the immediate environment. Second, lacking diligence, tenacity, or persistence reflects the easy and simple gratification provided by crime, and those with low self-control tend to want immediate rewards without much effort. Third, being adventuresome, active, and physical are reflective of the excitement, risk, and thrill attached to the criminal act. Fourth, being insensitive or indifferent reflects the lack of relevance of the discomfort or pain the victims of criminal acts may experience. Fifth, possessing a marginal tolerance for frustration reflects relief from temporary irritation. These individuals will also possess a volatile temper indicative of their low tolerance for frustration. They note that "there is a considerable tendency for these traits to come together in the same people" (Gottfredson and Hirschi 1990, pp. 90–91).

Some criminologists have interpreted Gottfredson and Hirschi's definition of self-control as a unidimensional construct because its dimensions should coalesce in a person. According to Trochim (2001, pp. 136), unidimensionality can be thought of as a number line, meaning that one line should be used to reflect higher or lower levels of a trait. For self-control, this would mean that all elements specified by Gottfredson and Hirschi are one and the same and therefore do not represent different attributes as they can all be captured on one number line to indicate more or less self-control.

In one of the first empirical tests of Gottfredson and Hirschi's theory, Grasmick et al. (1993, pp. 9) explicitly interpreted the concept of self-control as a unidimensional construct:

> A factor analysis of valid and reliable indicators of the six components is expected to fit a one factor model, justifying the creation of a single scale called low self-control. In effect, this is a very crucial premise in Gottfredson and Hirschi's theory. A single, unidimensional personality trait is expected to predict involvement in all varieties of crime as well as academic performance, labor force outcomes, success in marriage, various "imprudent" behaviors such as smoking and drinking, and even the likelihood of being involved in accidents. Evidence that such a trait exists is the most elementary step in a research agenda to test the wealth of hypotheses Gottfredson and Hirschi have presented

Others have also pursued measurement of self-control under a conceptual framework of unidimensionality. For example, Nagin and Paternoster (1993, p. 478) note that "the construct was intended to be unidimensional." Further, Piquero and Rosay's (1998, p. 157) interpretation shows consistency with a unidimensional definition when they stated, "evidence for a solution that has more than one factor would not be consistent with Gottfredson and Hirschi's claim."

Criminologists have also interpreted Gottfredson and Hirschi's original conceptualization of self-control as a multidimensional construct, likely due to the elements that embody the concept of self-control. Unlike unidimensionality, it can be problematic to measure a multidimensional construct on a single number line (Trochim 2001, p. 135). For instance, intelligence consists of multiple dimensions such a math and verbal ability. A person may have strong verbal ability and weak math ability. Some argue the same is true for self-control. For example, some individuals may have a high level of impulsivity but also have lower levels of frustration. Various elements of self-control may indicate different, yet correlated, constructs. As such, it would be incorrect to depict a person's level of self-control using one number line.

While some criminologists suggest that evidence of multidimensionality would be damaging to Gottfredson and Hirschi's original conceptualization (Longshore et al. 1996), others interpret such evidence as support for their claim (Arneklev et al. 1999; Vazsonyi et al. 2001). Arneklev and colleagues (1999) argued that Gottfredson and Hirschi specify six dimensions of self-control, so the characteristics cannot be anything but multidimensional. What is questionable, according to Arneklev and his colleagues (1999), is whether or not these six elements account for a final, higher-order construct. Vazsonyi et al. (2001) also argued that Gottfredson and Hirschi conclusively outline self-control as a multidimensional trait. However, they also suggest that this is not in contrast to a unidimensional interpretation when they stated that "a multidimensional measure of self-control still can and does imply that these elements together form a single latent trait of self-control" (Vazsonyi et al. 2001, p. 98).

The conceptual confusion resulting from interpretations of Gottfredson and Hirschi's description of self-control complicates efforts to validate scales designed to specifically test self-control. This is largely because there is no general consensus on how to interpret their construct.

An appropriate operational definition is the second source of controversy regarding self-control as a construct (Akers 1991; Hirschi and Gottfredson 1993, 1994). The controversy surrounds two different operationalizations: attitudinal and behavioral. For Hirschi and Gottfredson (1993), behavior-based operationalizations are preferred. They (Hirschi and Gottfredson 1993, p. 49) explicitly stated, "behavioral measures of self-control seem preferable to self-reports" and "multiple measures [items] are desirable." In contrast, Akers (1991) has warned against such operationalizations due to a tautology issue of not having indicators of self-control that are independent of outcomes that it should predict. Nevertheless, Hirschi and Gottfredson argue that behavioral indicators can be identified independent of crime. They propose the following:

> With respect to crime, we would propose such items as whining, pushing and shoving (as a child); smoking and drinking and excessive television watching and accident frequency (as a teenager); difficulties in interpersonal relations, employment instability, automobile accidents, drinking, and smoking (as an adult). (Hirschi and Gottfredson 1994, p. 9)

Furthermore, Hirschi and Gottfredson (1993) prefer direct behavioral observations when operationalizing self-control, and they put less faith in self-reports. They argue that "the level of self-control itself affects survey responses. . .self-report measures, whether of dependent or independent variables, appear to be less valid the greater the delinquency of those whom they are applied" (Hirschi and Gottfredson 1993, p. 48). While they do not argue for abandoning operational definitions that employ self-report methods to test self-control theory, they do suggest that differences among respondents should be considered in measurement when testing their theory (Hirschi and Gottfredson 1993, p. 48).

Few studies have used direct behavioral observations to measure self-control. For example, Keane and colleagues (1993) used direct observations (i.e., failure to wear a seatbelt) as well as self-report behavioral items (i.e., drinking, perceived risk of being stopped by police) to predict driving under the influence. Most behavioral operationalizations of self-control, however, rely on self-reports. In operationalizing self-control, Zager (1994, p. 75) used an index consisting of "six self-report delinquency items, including alcohol use, marijuana use, making obscene phone calls, avoiding payment, strong arming students, and joyriding." Similarly, Evans and his colleagues (1997) used an operational definition of low self-control that consisted of self-reports of violating the speed limit, drunk driving, illegal gambling, and using drugs. Many of these self-report behaviors are deviant and criminal acts that Gottfredson and Hirschi (1990) would argue are predicted by low self-control, but they are used in some studies to measure low self-control as well. This can be a problem for researchers when attempting to disentangle cause and effect. Another problem with behavioral measures is the degree of measurement error they contain. For instance, Gibbs and Giever (1995) state that "crime and analogous behaviors as measures of self-control can be expected to contain substantial error because they reflect several underlying variables or constructs" (Gibbs and Giever 1995, p. 249).

The other, probably more favored, operational definition of self-control among criminologists has been attitudinal and/or personality-based self-report items (Grasmick et al. 1993; Gibbs and Giever 1995). Some argue that this method is a way to overcome the tautology problem (Stylianou 2002), while others argue that this approach implies psychological positivism and is incongruent with the self-control construct (Hirschi and Gottfredson 1993).

Gibbs and his colleagues (1998, p. 95) suggest that a variable used to explain behavior "can be most clearly grasped and tested when it is defined as something broader or different than behavior." An advantage to such an approach is that it "leaves no space for tautology" (Stylianou 2002, p. 538). Additionally, Gibbs and Giever (1995) argue that self-inventory, personality-based operational definitions have advantages over behavioral ones for two reasons: (1.) they are more useful in tapping more cognitive aspects of self-control and (2.) allow for a more comprehensive coverage of domains of self-control because items can be developed to capture typical modes of behavior that relate to everyday life (Gibbs and Giever 1995, p. 249).

Few self-report attitudinal and/or personality-based operational definitions have been developed specifically to test Gottfredson and Hirschi's construct of self-control (Gibbs and Giever 1995; Grasmick et al. 1993). While Gibbs and Giever (1995) created such a measure, its creation was intended to be relevant only to college students and has not received much empirical attention beyond their own exploratory scrutiny. Grasmick et al. (1993) created a 24-item attitudinal/personality scale based on their interpretation of Gottfredson and Hirschi's conceptual definition of self-control, and it has been used widely in tests of Gottfredson and Hirschi's (1990) theory. For instance, Pratt and Cullen (2000) show that at least 12 studies have used Grasmick et al.'s (1993) scale in examining self-control theory, and many others have used it since then (Ward et al. 2010; Gibson et al. 2010). Since Grasmick et al.'s attitudinal scale has been one of the most utilized measures of self-control, the next section provides an extensive review of studies that have been conducted to examine its reliability and validity.

In the Beginning: Creation of the Grasmick et al.'s Self-Control Scale

Grasmick and colleagues paid close attention to how Gottfredson and Hirschi (1990) conceptually defined the elements of self-control and derived an operational definition to reflect its conceptual properties. Grasmick and colleagues (1993) suggested that the six components of self-control can be interpreted as a "personality trait" that should conform to a unidimensional trait.

Grasmick and his colleagues used many items in pre-testing college students to identify the final 24 items. This resulted in four items for each of the six elements of self-control. Please see below for lists of the original items. Items were scored on a four-point Likert scale ranging from (1) strongly disagree, (2) disagree somewhat, (3) agree somewhat, and (4) strongly agree.

Grasmick et al.'s (1993) self-control items

Item
Impulsivity
I often act on the spur of the moment without stopping to think
I don't devote much thought and effort to preparing for the future (reverse coded)
I often do whatever brings me pleasure here and now, even at the cost of some distant goal
I'm more concerned with what happens to me in the short run than in the long run
Simple Tasks
I frequently try to avoid projects that I know will be difficult
When things get complicated, I tend to quit or withdraw
The things in life that are easiest to do bring me the most pleasure

(*continued*)

Item
I dislike really hard tasks that stretch my abilities to the limit
Risk Seeking
I like to test myself every now and then by doing something a little risky
Sometimes I will take a risk just for the fun of it
I sometimes find it exciting to do things for which I might get in trouble
Excitement and adventure are more important to me than security
Physical Activities
If I had a choice, I would almost always rather do something physical than something mental
I almost always feel better when I am on the move than when I am sitting and thinking
I like to get out and do things more than I like to read and contemplate ideas
I seem to have more energy and a greater need for activity than most other people my age
Self-Centeredness
I try to look out for myself first, even if it means making things difficult for other people
I'm not very sympathetic to other people when they are having problems
If things I do upset people, it's their problem not mine
I will try to get the things I want even when I know it's causing problems for other people
Temper
I lose my temper pretty easily
Often, when I'm angry at people I feel more like hurting them than talking to them about why I am angry
When I'm really angry, other people better stay out of my way
When I have a serious disagreement with someone, it's usually hard for me to talk calmly about it without getting upset

While Grasmick and colleagues were quite methodical when determining the face and content validity of the items they used to measure self-control, they neglected other important measurement-related issues. For example, Grasmick and his colleagues (1993) did not give advice on which populations the instrument is appropriate (e.g., college students, juveniles, and incarcerated populations), if directions were explicitly given to respondents, and which scoring procedures should be used for scale construction.

Consequently, it is not clear whether Grasmick et al.'s self-control measure should be equally applied to different samples to discriminate between levels of self-control (or a lack thereof). Perhaps, their scale items are more suitable for low-risk samples, such as college students, rather than high-risk samples, such as serious criminal offenders. The scale items may be too easy or too agreeable for a sample of respondents who, on average, are likely to have lower self-control. This could result in the inability of Grasmick et al.'s scale to accurately measure levels of self-control in some populations.

Although Grasmick and his colleagues (1993) advise readers not to accept their work as a definitive measure of self-control, their scale remains the measuring instrument of choice among researchers attempting to quantify self-control (see Delisi et al. 2003). To support its continued use, however, there must be evidence showing that the scale is empirically reliable and valid and that it is applicable to different samples.

Reliability of Grasmick et al.'s Self-Control Scale

A handful of studies have now explicitly examined the psychometric properties of Grasmick et al.'s self-control scale (Arneklev et al. 1999; Delisi et al. 2003; Gibson et al. 2010; Grasmick et al. 1993; Longshore et al. 1996; Piquero et al. 2000; Piquero and Rosay 1998; Vazsonyi et al. 2001). Most studies employing Grasmick et al.'s scale typically estimate Cronbach's alpha as a measure of scale reliability. Assuming that the self-control construct is unidimensional, the alpha for this scale should be quite high, or at least modest, ranging from 0.7 to 0.9. Lower internal consistency scores may be a function of the scale's multidimensionality; however, this would remain an empirical question that reliability estimates alone cannot answer. Although many studies report alpha for the Grasmick et al.'s scale when testing relational propositions of self-control theory, this section specifically summarizes studies that have focused on the psychometric properties of this scale.

Using data from a simple random sample of 395 adult respondents who completed the

Oklahoma City Survey, Grasmick and his colleagues (1993) conducted the first reliability analysis of their scale. They concluded that by dropping one item (the last item under the Physical Activities component) from the scale, they could increase reliability from 0.80 to 0.81.

Two studies using the same data set emerged in 1996 and 1998, which came from a multisite evaluation of Treatment Alternatives to Street Crime (TASC) programs to identify drug-using adult and juvenile offenders in the criminal justice system to gauge their treatment needs. The sample consisted of 623 offenders. Most sample respondents had lengthy criminal histories, and the sample varied in sex, race, and age. It should be noted that the version of the scale in these studies diverges slightly from its original form in a few ways. First, Longshore et al. (1996) modified the original response scale and added an additional category to make it a five-point Likert scale: never (0), rarely (1), sometimes (2), often (3), and almost always (4). Second, item wording was changed and often reversed to detect any bias from yes-saying. Longshore et al. (1996) reported a Cronbach's alpha of 0.80 for the Grasmick et al.'s scale, whereas Piquero and Rosay (1998) reported a Cronbach's alpha of 0.71. Piquero and Rosay (1998) reported gender-specific alpha's, 0.72 for males and 0.68 for females.

Both Longshore et al. (1996) and Piquero and Rosay (1998) reported reliability estimates for each self-control dimension, which were low compared to acceptable standards. Longshore et al.'s (1996) estimates were 0.65 for Impulsivity/Self-Centeredness, 0.48 for Simple Tasks, 0.58 for Risk Seeking, 0.35 for Physical Activities, and 0.71 for Temper. Piquero and Rosay's (1998) estimates were 0.45 for Impulsivity (0.46 for males and 0.43 for females), 0.44 for Simple Tasks (0.47 for males and 0.28 for females), 0.58 for Risk Seeking (0.58 for males and 0.56 for females), 0.37 for Physical Activities (0.40 for males and 0.31 for females), 0.68 for Temper (0.71 for males and 0.59 for females), and 0.57 for Self-Centeredness (0.59 for males and 0.49 for females). The difference was that Longshore et al. (1996) reported alpha's on only five of the six components; they combined Impulsivity and Self-Centeredness items due to the results of their factor analysis that is discussed in the next section.

Delisi et al. (2003) used data collected from 208 male parolees residing in work release facilities in a midwestern state that had been previously released from prison. They reported that Cronbach's alpha for the total scale was 0.91. They also computed reliability estimates for each component showing coefficients of 0.79 for Impulsivity, 0.81 for Simple Tasks, 0.79 for Risk Seeking, 0.72 for Physical Tasks, 0.81 for Self-Centeredness, and 0.86 for Temper.

In the largest study undertaken to investigate the psychometric properties of Grasmick et al.'s scale, Vazsonyi et al. (2001) gathered data on over 8,000 adolescents from four different countries including schools in Hungary ($n = 871$), Netherlands ($n = 1{,}315$), Switzerland ($n = 4{,}018$), and the United States (2,213). While total scale reliability estimates are not reported, they do report them for self-control subscales for the total and country-specific samples. Cronbach's alpha was 0.50 for Impulsivity ranging from 0.45 to 0.62; 0.68 for Simple Tasks ranging from 0.61 to 0.73; 0.79 for Risk Seeking ranging from 0.69 to 0.84; 0.63 for Physical Activities ranging from 0.55 to 0.74; 0.60 for Self-Centeredness ranging from 0.45 to 0.68; and 0.76 for Temper ranging from 0.68 to 0.76.

Gibson and colleagues (2010) analyzed a sample of 333 college students attending a university in the southern United States. They reported alpha coefficients for the full 24-item self-control measure and each of the six components for males and females separately. They found that the 24-item self-control measure had an alpha coefficient of 0.84 for males and 0.85 for females. Their assessment of the reliability of each dimension found an alpha coefficient for Impulsivity of 0.65 and 0.69 for males and females, respectively; for Simple Tasks was 0.75 and 0.77 for males and females, respectively; Risk Taking was 0.78 and 0.75 for males and females, respectively; Physical Activities was 0.81 and 0.80 for males and females, respectively; Self-Centeredness was 0.77 and 0.65 for males and females, respectively; and for Temper the estimates were 0.79 and 0.76 for males and females, respectively.

Dimensionality of Grasmick et al.'s Scale

Grasmick et al. (1993) were the first to assess the dimensionality of their scale. First, they performed a principal components exploratory factor analysis (EFA) with one-, five-, and six-factor solutions. They could not "find strong evidence that combinations of items into subgroups produces readily interpretable multidimensionality" (Grasmick et al. 1993, p. 17). Their analysis led them to conclude that "the strongest case can be made for a one-factor unidimensional model" (Grasmick et al. 1993, p. 17). Several studies have since shown similar results using the same method across different samples (Arneklev et al. 1999; Nagin and Paternoster 1993; Piquero and Tibbetts 1996; Piquero et al. 2000; Delisi et al. 2003), concluding that all items load moderately on the first factor which is interpreted as evidence of unidimensionality. However, these exploratory analyses also provided evidence for six factors. As such, the methods used in these studies were not ideal for concluding whether Grasmick et al.'s scale is unidimensional or multidimensional.

Depending on researchers' perceptions of the original conceptualization of self-control, findings from these studies have been interpreted as having multiple factors or one factor. Furthermore, EFA, e.g., principal components analysis, reduce multiple variables (or items) without an imposed theoretical structure, and they attempt to extract the most variance possible from the initial factor. EFA leaves the task of defining the number of factors up to the factor analysis program, therefore being inadequate for construct validity purposes (DeVellis 1991).

Some criminologists have used confirmatory measurement models that are more appropriate for construct validation. Results from these models have often led to divergent conclusions compared to Grasmick et al.'s (1993) original analysis. For instance, Longshore et al. (1996) and Piquero and Rosay (1998) used the same data from a sample of drug-using offenders and found that the Grasmick et al.'s scale items fit two different models, a unidimensional solution and a multidimensional solution with slight modifications to the scale items, e.g., dropping items from the analysis.

Longshore et al. (1996) did not find initial support for a single underlying factor, and the scale did not appear to function equally well across subgroups defined by race, sex, and age. They modified the scale by dropping two items and allowing several error terms to correlate in a confirmatory measurement model, still concluding that a one-factor solution did not adequately fit the data. Next, they assessed a five-factor solution by combining two of the components, i.e., Impulsivity and Self-Centeredness. This solution also provided a poor fit to the data until they allowed four error terms to correlate and one item to load on two different factors. Their modified five-factor solution provided a better fit to the data especially for juveniles (CFI = 0.89), males (CFI = 0.92), Caucasians (CFI = 93), African-Americans (CFI = 0.92), and adults (CFI = 0.91). This solution, however, provided a poor fit for women (CFI = 0.80). Most importantly, their results questioned the unidimensionality of the scale for a criminal population.

Piquero and Rosay (1998) reanalyzed the data used by Longshore and colleagues. They argued that the scale could conform to a one-factor solution, could be equally reliable and valid across gender, and could produce a good fit to the data without allowing error terms to correlate. Their confirmatory measurement model showed that a unidimensional model fit the data for both males and females, but they did drop several items from the scale reducing it to 19 items. For example, the Physical Activities component was reduced to two items, and the Impulsivity, Simple Tasks, and Self-Centeredness components were all reduced to three items.

Although Piquero and Rosay (1998) are confident that their results garnered supported for unidimensionality, others disagree and conclude that their results are analogous to a second-order factor analysis where one overarching factor accounts for the relationships among lower-level factors (Longshore et al. 1996). This criticism is based on Piquero and Rosay's (1998) averaging of the scores within each component or dimension and their use of the final six composite scores as indicators in a one-factor measurement model. In sum, the results produced by these two studies do not provide a clear, unambiguous understanding of the internal validity of self-control. In both cases, modifications were made to the scale so that the results from the analyses could conform to either a multidimensional or unidimensional measure.

Arneklev et al. (1999) employed a second-order, confirmatory measurement model to test the dimensionality of Grasmick et al.'s scale. In doing so, they argued that a valid measure of self-control should have six distinct dimensions that load on a higher-order self-control factor. Using a random sample of adults and a convenience sample of college students, they concluded that a second-order model fits the data well. For the adult sample, the results showed that the coefficients between both the indicators and the six dimensions of self-control and the six dimensions and the self-control latent construct were sufficiently large. Although each of the six dimensions was significantly related to the second-order self-control factor, they found that Impulsivity had the highest factor loading. Additionally, the Physical Activities dimension loaded less strongly on self-control than any other dimension. Overall, Arneklev and his colleagues (1999) concluded that the second-order factor model for the adult sample provided support for Gottfredson and Hirschi's (1993) theory; the goodness of fit statistic (GFI = 0.89) had an acceptable magnitude, indicating that the proposed theoretical model fit the data well. Arneklev et al. (1999) showed similar results for their college sample. Arneklev et al. (1999) concluded that evidence of six distinct dimensions exists, but evidence also indicated that all six dimensions loaded on a higher-order construct that they called self-control.

Vazsonyi and his colleagues (2001) used both exploratory and confirmatory factor analysis in their study of Grasmick et al.'s self-control scale for adolescents across four countries. In an a priori fashion, they interpreted Gottfredson and Hirschi's (1990) conceptualization of self-control to be multidimensional, consisting of six separate dimensions. They first estimated exploratory factor analysis models for the total sample as well as for groups by sex, age, and country. Vazsonyi et al. (2001) argued that their preliminary results indicated the existence of six factors and that the scale is not unidimensional. Second, they use all 24 items to conduct a series of more rigorous confirmatory models including a one-factor and six-factor model to confirm their exploratory efforts. Using several fit statistics (e.g., CFI = 0.65, GFI = 0.82, and RMSEA = 0.09 for the total sample), they concluded that a one-factor model was not a good fit to their data. Each item, however, did show a statistically significant loading. In contrast, they showed that a six-factor solution fits the data much better (e.g., CFI = 0.91, GFI = 0.95, and RMSEA = 0.05 for the total sample), even across groups by age, sex, and country. In a final attempt to improve the six-order factor model, they allowed for two correlated error terms and dropped two items from the scale to achieve a consistent, overall, improved fit (e.g., CFI = 0.93, GFI = 0.96, and RMSEA = 0.04 for the total sample) which did not vary much across groups. Unlike others (Arneklev et al. 1999; Delisi et al. 2003; Piquero et al. 2000), Vazsonyi and his colleagues did not test a second-order factor model.

Delisi et al. (2003) tested the dimensionality of Grasmick et al.'s scale using a sample of male offenders residing in work release facilities in a midwestern state. They employed confirmatory measurement models to test one-factor, six-factor, and second-order factor models. While all loadings were statistically significant in all models, they concluded that all models fit the data poorly. These conclusions were drawn using numerous fit statistics. Most troubling for the unidimensionality hypothesis was their finding showing that the one-factor solution had the worst fit among the confirmatory factor models (GFI = 0.65, AGFI = 0.59, RMR = 0.11, and x^2/df = 4.27). From Delisi et al.'s (2003) confirmatory analyses, a model building effort was undertaken. They found that a modified six-factor model was able to fit their data well.

In sum, the evidence for the dimensionality of the most commonly used measure of self-control turns out to be inconclusive. While researchers continue to use it as if it is measuring one construct, evidence suggests that this may not be the case. This can have implications for the impact of self-control on outcomes hypothesized by Gottfredson and Hirschi. For instance, if Grasmick et al.'s self-control measure is actually multidimensional, then not treating it as such may mask the importance of each dimension's influence on criminal behavior, substance use, and other antisocial behaviors. Specifically, some dimensions may be relatively more important than others in predicting particular behavioral outcomes, or some dimensions may even interact with one another in important ways. By treating this measure as unidimensional, these possibilities will not be realized. Some criminologists have actually found that not every dimension of the scale is important in explaining criminal outcomes (Delisi et al. 2003). Others have found particular dimensions of Grasmick et al.'s scale, e.g., Impulsivity and Risk Seeking, to be more important than the overall self-control scale when predicting criminal outcomes (Longshore et al. 1996).

The Rasch Measurement Model and Self-Control

Piquero and his colleagues (2000) were the first to apply a Rasch measurement model to investigate the internal validity of Grasmick et al.'s self-control scale. The Rasch model is a confirmatory measurement model that tests for scale unidimensionality, but it diverges from traditional confirmatory measurement models discussed thus far in several important ways. First, the Rasch model produces distribution-free estimates in that the values do not depend on the distribution of the trait, i.e., self-control, across samples as do conventional exploratory and confirmatory factor models. This is important because results from Rasch models can be compared across samples when the same scale is employed, while results from factor analysis models are questionable for comparative purposes (Piquero et al. 2000; Bond and Fox 2001).

Second, a Rasch model separates person ability and item difficulty estimates and then places them both on the same logit ruler for comparative purposes. This function allows for comparisons of item difficulty in relation to people's level of ability, i.e., self-control, on the same interval-level scale. By taking into account the interaction between persons and scale items, the Rasch model overcomes the test-based approach of conventional confirmatory factor analysis. As such, the ability estimates do not depend on the difficulty of items in the scale. A Rasch model allows a researcher to detect the difficulty of endorsing items in relation to the distribution of self-control in the sample.

Third, the Rasch model is mathematically defined to assess unidimensionality. In doing so, each self-control scale item is examined to assess its fit to the model. Finally, Rasch models create interval-level measures from ordinal items, whereas factor analysis methods in several statistical software packages mistake ordinal responses for continuous responses violating statistical assumptions (see Piquero et al. 2000; Wright and Masters 1982).

Piquero et al. (2000) were interested in whether respondents used item response categories as they were supposed to. The use of response categories was orderly; those with low levels of self-control had higher probabilities of agreeing to low self-control items (selecting response categories that reflect low self-control) than those with high self-control. Second, they were interested in how well scale items fit the unidimensional Rasch model. They found that many items had poor fit to the unidimensional expectation of the model. Particularly, they found that 11 of the 24 items showed statistically significant misfit when all items were considered as a unidimensional measure. Third, Piquero et al. (2000) investigated a person/item logit ruler generated by Rasch modeling software to determine that several items were too difficult for their college sample to endorse. A logit ruler, or person/item map, allows researchers to assess the distributions of ability and item difficulty on the same metric (i.e., logit scale) to determine if items are too difficult to endorse relative to the distribution of the sample's self-control. Most of their sample had very low ability indicating high levels of self-control. The Grasmick et al.'s items did not discriminate well among a college sample with disproportionately high levels of self-control. Finally, Piquero et al. (2000) conducted a differential item function (DIF) analysis to assess item responses across high and low self-control groups. Low-self-control individuals were found in some instances to respond to items differently than those having high self-control. Particularly, the low self-control group was more (or less) willing to agree with some items than would have been expected by the Rasch model.

In a more recent study, Gibson et al. (2010) used a Rasch Model to assess how Grasmick et al.'s self-control items may perform differently for a sample of male and female college students. Gibson and colleagues' investigation found that self-control items performed differently for men and women. In fact, 33 % of the items showed differential item function. In some cases, women were less likely to agree to items compared to men even when males and females had similar levels of the trait. This means that several of the items were biased for men and women and that this could inflate or deflate aggregate self-control differences between them. When the biased items were excluded from the scale, a gender difference in aggregate self-control was still observed, but the difference was reduced.

In sum, the few studies that have used a Rasch model to investigate the internal validity of Grasmick et al.'s self-control measure have found that the items do not conform to a unidimensional trait and potential biases exist for using it to measure self-control for different groups. While this is a relatively new approach to investigating the validity of self-control measures, researchers should continue to use it in the future for several reasons. First, the differential item functioning capabilities of the Rasch model will be helpful for understanding if self-control measures can or should be applied to different groups including gender, race, community samples, adolescents, and prison samples to name only a few. Second, and from a developmental perspective, this approach will allow researchers to determine if measures of self-control such as the Grasmick et al.'s scale items will have similar or different meanings as people age. As such, this will allow a researcher to determine if true change in this construct is occurring within individuals over time or if differences observed within individuals over time are due to the items having different meanings for individuals as they age. Third, the Rasch model has the valuable property of being able to determine if self-control items are able to discriminate well between individuals with low to high levels of self-control.

Conclusions and Future Research

This entry has provided a review of key issues related to the measurement of self-control. Of these, several noteworthy themes emerged. First, much less criminological research has focused on the measurement of self-control compared to the large body of work that has investigated the influence of self-control on a variety of behavioral outcomes. This is unfortunate because the research on self-control measurement has produced little agreement among criminologists as to what measure should be used to capture this important construct.

Second, further complicating the disagreement among criminologists is the conceptual and operational definitions provided by Gottfredson and Hirschi. Although they were creative in linking elements of self-control directly to the criminal act, they leave readers questioning whether self-control should be interpreted as a unidimensional or multidimensional construct and the merits of behavioral measures. These issues have implications for how criminologists should proceed in testing the influence of self-control on behavioral outcomes; that is, should researchers examine the effect of each of self-control's dimensions on outcomes or the combined effect of self-control or both? Moreover, recent studies that attempt to expand self-control theory have used quite different and diverse measures of self-control. It is likely that some measures being used are more valid and reliable than others, but this must be acknowledged and empirically addressed before determining the true effects of self-control on behavioral outcomes or the casual underpinnings of self-control.

Third, it was noted that one of the most common measures used by criminologists to assess differences in self-control among individuals has been Grasmick et al.'s scale or a reduced version of it. The empirical studies conducted to date on the psychometric properties of this measure are mixed. Studies estimating the reliability of Grasmick et al.'s scale have employed diverse samples ranging from convicted offenders, adult community members, adolescents residing in different countries, and college students. Only one of these studies reports a Cronbach's alpha above 0.9 for the total scale (Delisi et al. 2003); however, other studies do indicate that the scale has modest to high reliability. This is necessary, but not sufficient, for demonstrating scale validity. The scale items appear to cohere more closely in some samples than others, and some studies show low reliability for subscales. In addition, Grasmick et al. (1993) explicitly stated that a factor analysis of valid and reliable indicators of self-control should produce a unidimensional measure and they found support for their hypothesis, but they used an exploratory analytic method to make such an inference. Since then, several rigorous examinations of the scale have refuted their claims. In contrast, the scale has often been shown to be multidimensional reflecting either six-factor or a second-order factor structure where dimensions load on a latent self-control factor with some dimensions loading more strongly than others. As such, more studies in the future should investigate closely how each of these dimensions differentially influence outcomes that Gottfredson and Hirschi would predict are affected by self-control.

Fourth, conflicting results have emerged as to the Grasmick et al.'s scale's dimensionality and validity across demographic groups using traditional confirmatory measurement models and the Rasch model. For example, conflicting results have been found for whether the scale is equally reliable and valid for blacks, Hispanics, whites, males, and females. Gottfredson and Hirschi (1990) do not clearly specify the factor structure that a valid measure of self-control should posses, let alone how such a structure would hold up across different races. The Rasch model holds promise for understanding if the Grasmick et al.'s scale and other measures of self-control are unbiased across different groups.

Since the 1990s, few studies have been published in criminology that primarily center on the measurement of self-control. With some exceptions, most studies that have done so focus on the dimensionality debate of Grasmick et al.'s self-control scale using a variety of samples. As such, research on the measurement of self-control is open to new ideas and can hold much promise for criminology. A set of ideas for those pursuing research on the measurement of self-control is discussed below.

First, compiling a systematic list of all self-control measures used in past criminological research would be beneficial for researchers interested in exploring self-control theory. As such, criminologists would be able to assess the positive and negative attributes of each and make decisions on which measures or measure best fits their purpose.

Second, different conceptual and operational definitions should be pursued by criminologists, especially given Gottfredson and Hirschi's disapproval of the Grasmick et al.'s scale and Hirschi's more recent re-conceptualization of the self-control concept which suggests that self-control and social control are the same (Hirschi 2004). A recent effort has been made by Marcus (2004). Opposed to the idea that six elements define the construct of self-control, Marcus (2004) argued that Gottfredson and Hirschi's (1994, p. 3) construct is simply "the tendency to avoid acts whose long-term costs exceed their monetary advantages." Marcus acknowledges that Grasmick et al.'s 24 items reflect six traits and these traits overlap with a large body of literature on the structure of personality. He argued that elements such as risk seeking and preference for easy tasks measured by this instrument reflect a motivational basis of behavioral choice and do not reflect behavioral constraint. Given this conceptual flaw, he argued that Grasmick et al.'s measure is incompatible with the main premise of self-control theory. In fact, he argued that the six elements measured by Grasmick et al.'s scale are linked to the five-factor model (FFM) of personality. So, if Grasmick et al.'s scale is a unique measure of Gottfredson and Hirschi's construct of self-control and not merely a measure that is tapping into several personality traits, then it should show divergent validity when compared to a widely used and psychometrically sound instrument of the FFM. Future research should explore this idea by empirically investigating the correlations between Grasmick et al.'s subscales and the most well-researched instrument used to measure the FFM of personality, the NEO-personality inventory.

Third, given Gottfredson and Hirschi's preference for behavioral self-control measures, an important line of future research on the measurement of self-control will be to create behavioral measures and empirically compare them to Grasmick et al.'s measure to assess divergent and convergent validity. Although such investigations have received limited empirical attention (Tittle et al. 2003), very few attempts have been made to create a behavioral measure that is consistent with Gottfredson and Hirschi's conceptual definition of self-control. Marcus (2004) has created a self-control measure that consists of 67 strictly behavioral statements that are designed to assess the frequency of prior conduct across developmental stage that have long-term negative consequences. Future research should not only investigate how Marcus's scale correlates with Grasmick et al.'s measure but should assess the predictive validity of his measure and compare it to the predictive validity of Grasmick et al.'s measure (see Ward et al. 2010).

Fourth, researchers should further explore the validity of self-control measures for different types of people, especially given the possibility that one's own self-control can affect his/her survey responses. Responses may be less valid for individuals with higher criminality and/or lower self-control (Hirschi and Gottfredson 1993; Piquero et al. 2000). With regard to the measurement of self-control, this is a topic that has received minimal empirical attention. Hirschi and Gottfredson (1993) have argued that a solution would be to collect behavioral indicators of self-control that are measured independent of respondents; for instance, having parents or teachers report on the behaviors of children and adolescents (see also Gibson et al. 2010). Acknowledging that such a solution could be a financial burden to researchers that do not have funds to collect direct observational data, an alternative to Hirschi and Gottfredson's solution would be to employ methods that would enhance the accuracy of survey responses from those having lower levels of self-control. In doing so, researchers could consider the nature of self-control and how it may guide item creation and scale formatting.

The research ideas discussed here should be seen as ways for improving self-control measures used by criminologists, but they should not be the only measurement endeavors pursued. If the concept of self-control is as important for understanding criminal and antisocial outcomes as criminologists suggest, it will be even more important to refine measures of self-control in ways that most accurately reflect the conceptual and operational definition of the construct. At the same time it is important for criminologists to move toward the creation of a psychometrically reliable and valid measure that will be used to capture differences in self-control for populations of interest to criminologists.

Recommended Reading and References

Akers RJ (1991) Self-control as a general theory of crime. J Quant Criminol 7:201–211

Arneklev BJ, Grasmick HG, Bursik RJ Jr, (1999) Evaluating the dimensionality and invariance of "low self- control." J Quant Criminol 15:307–331

Bentler PM (1990) Comparative fit indexes in structural equation models. Psychol Bull 107:238–246

Bond TG, Fox CM (2001) Applying the Rasch model: fundamental measurement in the human sciences. Lawrence Erlbaum, Mahwah

Delisi M, Hochstetler A, Murphy DS (2003) Self-control behind bars: a validation study of the Grasmick et al. scale. J Quarterly 20:241–263

DeVellis RF (1991) Scale development. Sage, Newbury Park

Evans TD, Cullen FT, Burton VE Jr, Dunaway RG, Benson ML (1997) The social consequences of self-control theory. Criminology 35(3):475–501

Gibbs JJ, Giever D (1995) Self-control and its manifestations among university students: an empirical test of Gottfredson and Hirschi's general theory. JQ 12:231–255

Gibbs JJ, Giever D, Martin JS (1998) Parental management and self-control: an empirical test of Gottfredson and Hirschi's general theory. J Res Crime Delinq 35:40–70

Gibson CL, Sullivan C, Jones S, Piquero A (2010a) Does it take a assessing neighborhood effects on self-control. J Res Crime Delinq 47:31–62

Gibson CL, Ward JT, Wright JP, Beaver KM, DeLisi M (2010b) Where does gender fit in the measurement of self-control. Criminal Justice Behav 36:883–903

Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Grasmick HG, Tittle CR, Bursik RJ, Arneklev BJ (1993) Testing the core empirical implications of Gottfredson and Hirschi's general theory of crime. J Res Crime Delinq 30:5–29

Hambleton RK, Swaminathan H, Jane Rogers H (1991) Fundamentals of item response theory. Sage, Newbury Park

Hirschi T (2004) Self-control and crime. In: Baumiester RF, Vohs KD (eds) Handbook on self-regulation: research, theory, and application. Guilford, New York, pp 53–552

Hirschi T, Gottfredson MR (1993) Commentary: testing the general theory of crime. J Res Crime Delinq 30:47–54

Hirschi T, Gottfredson MR (1994) The generality of deviance. In: Hirschi T, Gottfredson MR (eds) The generality of deviance. Transaction Publishers, New Brunswick

Keane C, Maxim PS, Teevan JJ (1993) Drinking and driving, self-control, and gender: testing the general theory of crime. J Res Crime and Delinq 30:30–46

Longshore D, Turner S, Stein JA (1996) Self-control in a criminal sample: an examination of construct validity. Criminology 34:209–228

Marcus B (2004) Self-control in the general theory of crime: theoretical implications of a measurement problem. Theor Criminol 8:33–35

Nagin DS, Paternoster R (1993) Enduring individual differences and rational choice theories of crime. Law and Soc Rev 27:467–496

Nunnally JC, Bernstein IH (1994) Psychometric theory, 3rd edn. McGraw Hill, New York

Paternoster R, Brame R (1998) The structural similarity of processes generating criminal and analogous behavior. Criminology 36:633–670

Piquero AR, Tibbetts S (1996) Specifying the direct and indirect effects of low self–control and situational factors in offender's decision making. Justice Quart 13:481–510

Piquero AR, Rosay AB (1998) The reliability and validity of Grasmick et al.'s self-control scale: a comment on Longshore et al. Criminology 36:157–173

Piquero AR, MacIntosh R, Hickman M (2000) Does self-control affect survey response? Applying exploratory, confirmatory, and item response theory analysis to Grasmick et al.'s self-control scale. Criminology 38:897–929

Pratt TC, Cullen FT (2000) The empirical status of Gottfredson and Hirschi's general theory of crime: A meta–analysis. Criminology 38:931–964

Stylianou S (2002) Relationship between element and manifestations of low self-control in a general theory of crime: two comments and a test. Deviant Behav 23:531–557

Tittle CR, Ward DA, Grasmick HG (2003) Self-control and crime/deviance: cognitive vs. behavioral measures. J Quantitative Criminol 19:333–365

Trochim WMK (2001) The research methods knowledge base. Atomic Dog Publishing, Cincinnati

Vazsonyi AT, Pickering LE, Junger M, Hessing D (2001) An empirical test of a general theory of crime: a four-nation comparative study of self-control and the prediction of deviance. J Res Crime Delinq 38:91–131

Ward JT, Gibson CL, Boman J, Leite WL (2010) Is the general theory of crime really stronger than the evidence suggests? Assessing the validity of the retrospective behavioral self-control scale. Criminal Justice Behav 37:336–357

Wright BD, Masters GN (1982) Rating scale analysis. MESA Press, Chicago

Zager MA (1994) Gender and crime. In: Hirschi T, Gottfredson MR (eds) The generality of deviance. Transaction Publishers, New Brunswick

Measuring Crime Specializations and Concentrations

Martin A. Andresen
School of Criminology, Institute for Canadian
Urban Research Studies, Simon Fraser
University, Burnaby, Canada

Overview

Research on crime at places is a rapidly growing literature. Though Eck and Weisburd (1995) coined the term "crime at places," the starting point of this literature was Sherman et al. (1989) in a microspatial analysis of predatory crime in Minneapolis. This research area finds that calls for police service are highly concentrated in very few places: 50 % of calls for police service are generated from 5 % of street segments (or less), particularly when one considers detailed crime classifications (Sherman et al. 1989; Andresen and Malleson 2011). Moreover, research has shown that spatial crime patterns at microspatial units of analysis vary significantly across crime classifications (Andresen and Malleson 2011). Because of this consistency, Weisburd et al. (2012) put forth the Law of Crime Concentrations.

One implication from this research is that opportunity surfaces for different crimes, though they may overlap, are different – the opportunity surface for a crime is the distribution of targets across the landscape. From the theoretical perspective of environmental criminology, this statement is rather trivial: routine activity theory (Cohen and Felson 1979) posits that crime occurs when a motivated offender and suitable target converge in time in space without the presence of a guardian – this convergence occurs at a place. In the geometric theory of crime (Brantingham and Brantingham 1981), crimes occur at discrete locations (places) within the boundaries of our activity spaces. And in rational choice theory (Clarke and Cornish 1985), crimes are the result of context-specific choices that consider the immediate environment (places) within which a crime occurs. Even beyond these various theories, one only needs to consider where targets are located for different crime classifications: commercial burglary and robbery only occur in commercial districts, and residential burglary only occurs in residential districts, for example. Needless to say, whether one considers theory or the most straightforward aspects of our (built) environment, crime specializes at places.

Crime specialization is different from crime concentration. We have long known that crime concentrates in places, and those places vary by crime classification (Quetelet 1842). Specialization does not necessarily occur in the same places as concentration. Crime concentration is typically measured using crime rates (a crime count scaled by a population at risk); by the nature of this calculation, places with high crime rates (high degrees of concentration) have greater levels of crime per capita than places with low crime rates. Crime specialization, however, may occur in places with low levels of crime per capita – specialization may be defined relative to the entire study area. As such, the form of measurement necessary must consider the crime mix in some form. Location quotients provide an excellent example of this form of crime measurement.

Location Quotients: The State of the Art

The location quotient is a geographic measure of specialization. One of the benefits of the location quotient is that it only relies on data for the phenomenon in question; if a researcher is analyzing crime only, crime data are needed. This proves useful because measures of concentration (crime rates) have a number of methodological issues that emerge because they require other data (Andresen 2006). Of course, the location quotient still suffers from all the problems associated with crime reporting issues and the dark figure of crime, a concern in quantitative-based studies of crime dating back to at least the early nineteenth century (Bulwer 1836). Regardless, the location quotient is calculated as follows:

$$LQ_{in} = \frac{C_{in}/C_{tn}}{\sum_{n=1}^{N} C_{in} \Big/ \sum_{n=1}^{N} C_{tn}}, \qquad (1)$$

where C_{in} is the count of crime i in subregion n, C_{tn} is the count of all crimes in subregion n, and N is the total number of subregions. In this context, the location quotient is a ratio of the percentage of a particular type of crime in a subregion relative to the percentage of all crime in that same region. If the location quotient is equal to one, the subregion has a proportional share of a particular crime; if the location quotient is greater than one, the subregion has a disproportionately larger share of a particular crime; and if the location quotient is less than one, the subregion has a disproportionately smaller share of a particular crime. Specifically, if a subregion has a location quotient of 1.20, that subregion has 20 % more of that crime than expected given the percentage of that crime in the region as a whole – that subregion then "specializes" in that crime classification. Miller et al. (1991) provide the following classifications that are useful for interpreting the location quotient: very underrepresented areas, $0 \leq LQ \leq 0.70$; moderately underrepresented areas, $0.70 < LQ \leq 0.90$; average represented areas, $0.90 < LQ \leq 1.10$; moderately overrepresented areas, $1.10 < LQ \leq 1.30$; and very overrepresented areas, $LQ > 1.30$.

Brantingham and Brantingham (1993, 1995, 1998) introduced the location quotient into criminological research in the early 1990s, but its adoption as a standard criminological measurement has been slow. With a primary concern for violent crime, Brantingham and Brantingham (1993, 1995, 1998) use the location quotient to measure the crime mix and specialization within a municipality; they compare the location quotient results to those for crime counts and crime rates. The larger municipalities in their study, not surprisingly, had the highest counts. However, after controlling for population size, the large municipalities no longer ranked at the top of the list, but now towards the bottom of the list. Smaller hinterland municipalities topped the list when using crime rates. Yet another ranking emerged with the location quotient: some municipalities with high crime rates also have high location quotients, but there are also a number of municipalities that have low crime rates with high location quotients. Consequently, there are municipalities that have low risk of crime in general (low crime rates), but specialize in violent crimes.

Though this phenomenon of specializing in a particular crime when the risk of crime is rather low is interesting in itself (and the focus of this entry), this property may allow one to investigate a curious result that emerges when considering crime rates. In the context of illegal drugs in the United States, Rengert (1996) expects the north central region of the United States to have the greatest proportions of marijuana crimes (of all drug-related crimes) because of a lack of a coastline and its agricultural base – heroin and cocaine are expected to be greatest in coastal regions because of a greater need to transport the drugs internationally. However, Rengert (1996) finds that the spatial pattern of marijuana crimes effectively follows the same spatial pattern as heroin and cocaine when using calculations based on the crime rate: the north central region of the United States is ranked last. However, when Rengert (1996) uses the location quotient to consider crime specialization, the pattern reversed. Therefore, the other regions of the United States had greater volumes and corresponding rates of marijuana crimes than the north central region, but if one were to commit a drug crime in the north central region, it is most likely to be related to marijuana.

Curiously, aside from one study that uses the location quotient as a component in a composite index, its use in criminology is absent for 10 years. In this more recent research, McCord and Ratcliffe (2007) use the location quotient to measure crime intensity across neighborhoods and find that drug markets tend to cluster close to pawnshops, drinking establishments, and mass transit stations. Andresen (2007) performs an inferential analysis (spatial regression) using the location quotient as a dependent variable and finds that if independent variables are interpreted as attractors of a particular crime, it performs well in a predictive manner. Ratcliffe and Rengert (2008) use the location quotient to identify areas that have greater intensities of shootings, relative to the city as a whole. Andresen (2009) investigates the phenomenon of crime rates in Canadian provinces increasing as one moves east to west – crime rates in the territories are even higher than in the western provinces. Andresen (2009) uses the location quotient to show that simply because all crime rates are greatest in the west does not mean that western provinces specialize in all crimes. In fact, crime specialization is present in all provinces for at least two crime classifications in each province. This analysis vividly shows that crime concentration (crime rates) does not necessarily imply crime specialization. Most recently, Block et al. (2012) use the location quotient in the context of automotive theft in the United States. They find that states and counties that contain or are near heavily trafficked borders and ports specialize in automotive theft; Block et al. (2012) speculate that this is an indication for the presence of a "theft for export" problem in these areas. Moreover, these are not always areas that have high rates of automotive theft.

Clearly, the location quotient's popularity in spatial crime analysis has increased in recent years. However, its general lack of adoption in criminological research is somewhat surprising because it allows for the measurement of crime specialization. Regardless, there are a number of other measures of specialization that may prove to be useful to criminologists. Some of these measures are already being used within the criminological literature in other (nonspatial) contexts, whereas others are used in geography and economics that prove insightful in a criminological context. To illustrate these, calls for service data are used below to show their utility.

An Empirical Example of Measuring Specialization

In order to perform the calculations below (crime rates, location quotients, and other measures of specialization), crime data are necessary. All crime data are for 2001 in the City of Vancouver, British Columbia, Canada. The areal unit of analysis is the dissemination area (DA). These units are defined by Statistics Canada and are similar to the census block groups in the United States Census. There are 990 DAs in Vancouver in 2001. The calculations below consider both property crime and violent crime classifications: assault, burglary, robbery, sexual assault, theft, theft of vehicle, and theft from vehicle – homicide is excluded because there are so few cases, twelve homicides for the entire year. Property crime counts and percentages dominate violent crime counts and percentages, as expected; property crimes account for just over 83 % of the 60 427 incidents included in the calculations below.

All crime data used below come from the Vancouver Police Department's Calls for Service Database (VPD-CFS Database) generated by its computer-aided dispatch system. Though the VPD-CFS Database is actually a proxy for actual crime data because not all calls for service represent actual crimes, the primary advantage of the VPD-CFS Database is this raw form – these data are not dependent on a criminal charge, though few calls for service are subsequently unfounded by the VPD. The geocoding success rate for these data was 93 %; though computer algorithms for geocoding have limitations that may lead to concerns over accuracy (Ratcliffe 2001), the current success rate is well above the minimum acceptable success rate of 85 % (Ratcliffe 2004).

Mapping the Location Quotient

Though the discussion regarding the location quotient above may be instructive to show the utility of its calculation, an example using burglary highlights the distinction between concentration and specialization. The burglary rate per 1,000 households in Vancouver (2001) is shown in Fig. 1a. It should be clear to the reader that burglary has its greatest concentrations in the eastern section of the central business district, the downtown eastside, as well as in a number of other areas that correspond to commercial land use, mixed land use, and arterial roadways; burglary is also in greater concentrations within the eastern side of the city – the eastern side of Vancouver has lower income levels than the western side of Vancouver, on average.

The spatial pattern of the location quotient for burglary (Fig. 1b), however, has a very different spatial pattern. Though there are areas throughout Vancouver that exhibit specialization, by far specialization occurs within the western side of Vancouver. This region of the city contains some of the wealthiest neighborhoods of Vancouver that all have relatively low levels of concentration in burglary. If one were to analyze other crime classifications in Vancouver, the results are not as stark as with burglary, but this example shows that where crimes are concentrated are not necessarily the same places in which specialization occurs.

Alternative Measures of Specialization

The first alternative measure considered here is a normalized version of the location quotient, normalized such that its values range from zero to unity. Though the algebra provides the same value, this normalized location quotient is based on the Balassa (1965) Index and is commonly used in the international trade literature (De Benedictis et al. 2009):

$$BI = \frac{C_{in} \Big/ \sum_{n=1}^{N} C_{in}}{C_{tn} \Big/ \sum_{n=1}^{N} C_{tn}}. \qquad (2)$$

The calculation of the normalized location quotient, referred to as the relative Gini in the literature (De Benedictis et al. 2009), is calculated as follows:

$$\text{Relative Gini (RG)} = \frac{\sum_{i=1}^{N-1} (p_i - q_i)}{\sum_{i=1}^{N-1} p_i}, \qquad (3)$$

where p_i is the share of crime i within the study area as a whole (Vancouver in the current

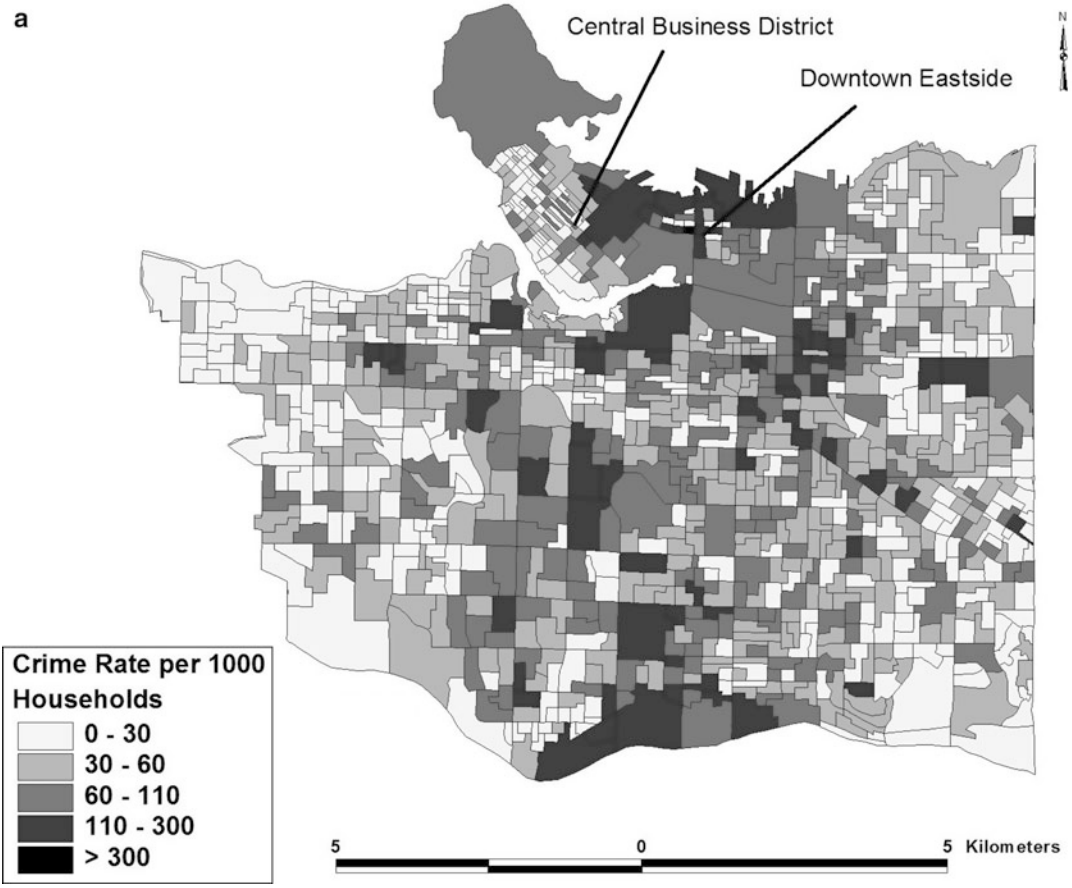

Measuring Crime Specializations and Concentrations, Fig. 1 (continued)

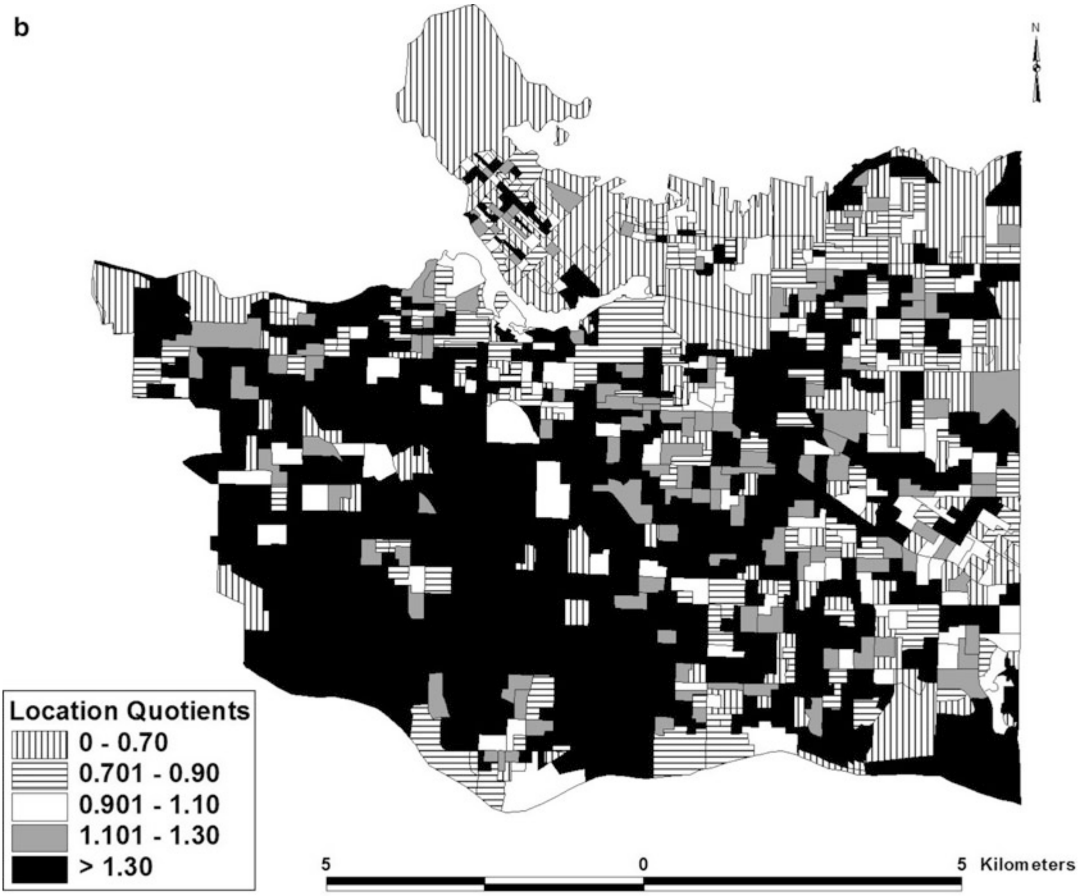

Measuring Crime Specializations and Concentrations, Fig. 1 (**a**) Crime rates per 1,000 households, burglary, Vancouver 2001. (**b**) Location quotients, burglary, Vancouver 2001

context) and q_i is the share of crime i within the subregion (dissemination area in the current context). In order to perform this calculation, all q_i are ranked according to their values with the top value being excluded. All values of q_i are ranked for the purposes of identifying the top value so it can be excluded from the calculation. This is done in order for the index to have a range from zero to unity. If q_N is included and there is total specialization (that must occur in that classification, by definition), the index takes on a value of zero. Clearly this makes the index of little value if this situation is a possibility. In the case of total specialization in a subregion, $q_N = 1$ and the index value is equal to zero. As such, if the research wishes to have total specialization have a value of unity, the calculation may easily be modified to be 1 – RG.

Though the location quotient has now been normalized, this measure is not without limitations. First, the natural interpretation of the location quotient has been lost: a value of 1.20 indicates that crime i specializes in a subregion by 20 %, relative to the study area as a whole. Such statements can no longer be made. Second, the normalized location quotient is not specific to a crime classification. Rather, it is a measure of specialization, generally speaking.

Another index that uses the location quotient in its calculation is one of the many variants of the

Theil (1967) Index. The Theil Index is a measure of relative specialization:

$$T_j = \sum_{k=1}^{n} \frac{x_{jk}}{x_j} \ln\left(LQ_{jk}\right), \tag{4}$$

where x_{jk} is the count of crime k in subregion j, x_j is the count of all crime in subregion j, ln() is the natural logarithm operator, and LQ_{jk} is the location quotient for crime k in subregion j, as calculated above in Eq. 1. When $T_j = 0$ there is no specialization and T_j has an upper bound of ln(z), where z is the number of basic units (crime classifications); in the current context, the upper bound of $T_j = \ln(7) = 1.94$.

The Theil Index, though instructive, also has its limitations. As with the normalized location quotient, though the Theil Index is calculated for each subregion, it is a general measure of specialization and does not provide information regarding which type of crime classification specialization occurs in a particular subregion. Also, the Theil Index is undefined if the share of crime k in a subregion is zero because the location quotient is then zero and the natural logarithm of zero is not defined. Of course, one may use a small number in place of zero, but this is dependent on the concerns of the study. If knowing that none of a particular crime occurred in a subregion is critical information, a different measurement of specialization would be more important. This increasingly becomes a problem in the context of crime as the spatial unit of analysis becomes smaller. As shown in the crime at places literature, many places consistently have no crime (Sherman et al. 1989; Andresen and Malleson 2011).

The Diversity Index is the next specialization measure to consider here. The Diversity Index is calculated as follows:

$$DI_i = 1 - \left(\sum_{n=1}^{N} {p_n}^2\right), \tag{5}$$

where p_n is the percentage of offenses in category n and i refers to the subregions. This index has been used in the criminological literature as a measure of offender specialization (Johnson et al. 2009) – this index may only be calculated using the latter summation term, depending on whether or not the researcher wants the value of zero or unity to indicate specialization. Though criminologists often point to other criminological or sociological research as its source, this index is a variant of the Herfindahl-Hirschman Index in economics that measures industrial concentration – see Herfindahl (1950) and Hirschman (1945).

In its original formulation,

$$H = \sum_{i=1}^{N} {S_i}^2. \tag{6}$$

H ranges from $1/N$ to unity, with N being the number of firms in an industry and S_i being the market share of firm i; in the current context, the number of crime classifications is under consideration. In order to have the index range from zero to unity, the following normalized H (H*) has been proposed:

$$H^* = \frac{H - \frac{1}{n}}{1 - \frac{1}{n}}. \tag{7}$$

As with the previous two measures of specialization, the Diversity Index is a general measure of specialization and does not provide information regarding what type of crime classification specialization occurs in a particular subregion.

The following measure of specialization has not been given a name in the economics literature, so it will simply be referred to as the Specialization Index in this discussion. The Specialization Index is calculated as follows:

$$S_j = \frac{\sum_{i=1}^{n} \left|h_{ij} - h_j\right|}{2}, \tag{8}$$

where h_{ij} is the share of crime j at subregion i and h_j is the share of crime j in the study area as a whole. The Specialization Index ranges from zero to unity with unity representing total specialization. The Specialization Index does not provide information regarding which type of crime classification specialization occurs in a particular subregion.

The last measure of specialization considered here is the Entropy Index – this index has been used in the public health literature to measure mixed land use (Frank et al. 2004). The Entropy Index for area j is calculated as follows:

$$E_j = 1 - \left(\frac{-\sum_{i=1}^{n} p_i \ln p_i}{\ln n} \right), \tag{9}$$

where p_i represents the share of crime classification i and n is the total number of crime classifications. The Entropy Index ranges from zero (evenly distributed crime classifications) to unity (perfect specialization). As with all the alternative measures of specialization, the Entropy Index does not provide information regarding what type of crime classification specialization occurs in a particular subregion. Also, as with the Theil Index, the Entropy Index has difficulties dealing with zero values for p_i because ln(0) is undefined. However, in this case ln(0) may be set to zero without loss of meaningful interpretations.

There are, of course, many other measures of specialization, primarily stemming from the economics, geography, and regional sciences literatures. Though they vary in the details, they generally convey the same type of information as those shown above. The primary difference between the measures shown above and other measures of specialization found in the literature is the complexity of the calculation. However, at this stage of introducing alternative measures of specialization that may be used in the context of places, the more straightforward measures were chosen to show their utility.

Mapping the Alternative Measures of Specialization

Before the discussion turns to the maps of the alternative measures of specialization, the similarity between the various alternative measures (and the location quotient for burglary) is of interest. The bivariate correlations (Pearson and Spearman) are shown in Table 1: Pearson correlations are shown in the upper right triangle of the table, whereas Spearman correlations are shown in the lower right triangle of the table. Though the location quotient for burglary is not highly related to the other measures of specialization (the correlations are positive and statistically significant), the alternative measures of specialization are, generally speaking, highly related with each other. Such relationships are not particularly surprising because the location quotient for burglary only considers burglary and the alternative measures of specialization are all quantifying the general nature of specialization, not being concerned with any particular crime. In fact, most of the (Pearson) correlations between the alternative measures of specialization are approximately 0.80, indicating that there is significant

Measuring Crime Specializations and Concentrations, Table 1 Specialization measures, Pearson and Spearman bivariate correlations, Vancouver, 2001

	Location quotient, burglary	Theil	Herfindahl-Hirschman	Herfindahl-Hirschman (Normalized)	Specialization	Entropy	Mean
Location quotient, burglary	1.00	0.44	0.59	0.56	0.31	0.53	0.49
Theil	0.41	1.00	0.79	0.81	0.80	0.72	0.71
Herfindahl-Hirschman	0.52	0.73	1.00	0.98	0.55	0.82	0.72
Herfindahl-Hirschman (Normalized)	0.52	0.73	1.00	1.00	0.67	0.83	0.75
Specialization	0.30	0.82	0.63	0.63	1.00	0.55	0.59
Entropy	0.50	0.70	0.94	0.94	0.57	1.00	0.73

Notes. Pearson correlations shown in upper triangle; Spearman correlations shown in lower triangle. All correlations significant at the 1 % level

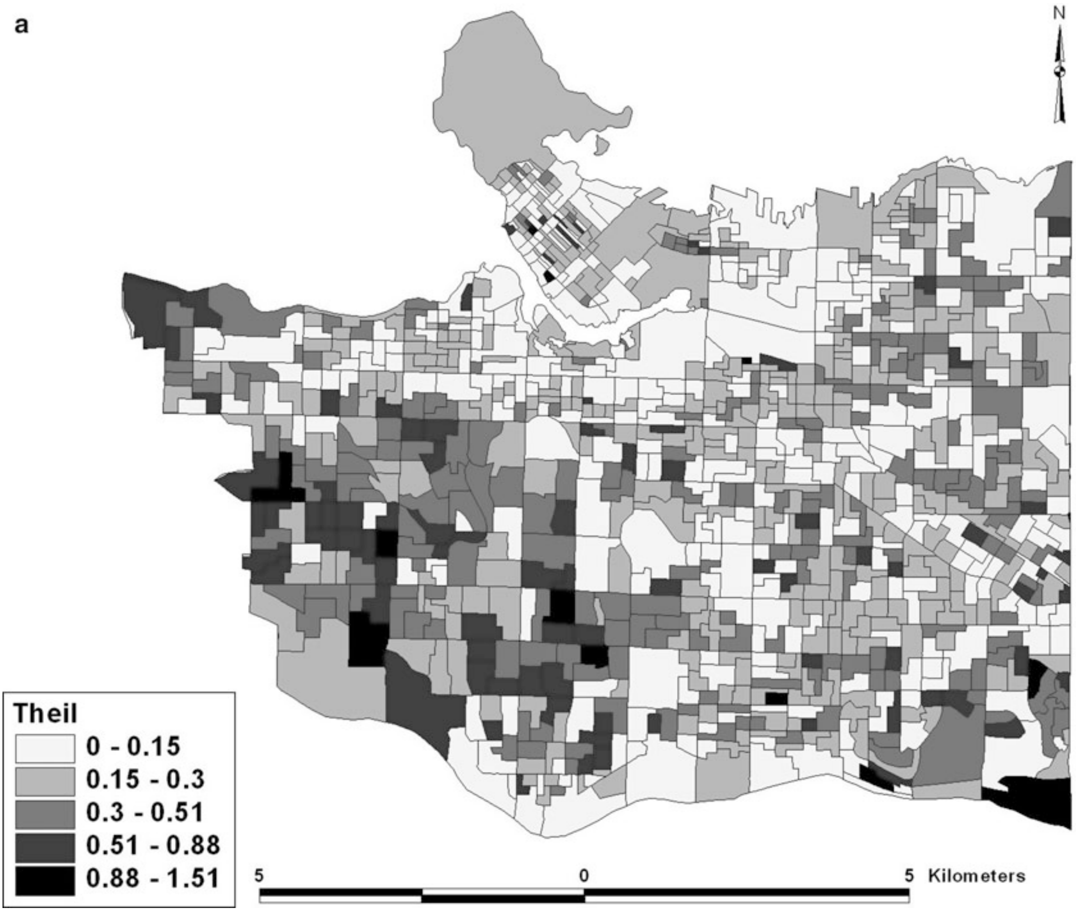

Measuring Crime Specializations and Concentrations, Fig. 2 (continued)

interchangeability among these alternative measures of specialization. The maps of these alternative measures indicate a similar conclusion.

Of all the alternative measures of specialization discussed above, only four are mapped for subsequent discussion: Theil Index (Fig. 2a), Herfindahl-Hirschman (Normalized) Index (Fig. 2b), Specialization Index (Fig. 2c), and Entropy Index (Fig. 2d) – the normalized location quotient is not mapped because the results are so similar to the location quotient shown in Fig. 1b, and the Herfindahl-Hirschman Index is not mapped because of its similarity to the Herfindahl-Hirschman (Normalized) Index. Because there are no classifications for these four alternative measures, unlike the location quotient, they are shown using natural breaks and five groups.

Immediately obvious with the Theil Index (Fig. 2a) is that the spatial pattern is similar to that of the location quotient for burglary (Fig. 1b). Though the spatial pattern is not as strong as with the location quotient, crime appears to specialize within the western side of Vancouver. This pattern is the opposite of expected when considering crime concentration because the west side of Vancouver is the wealthier section of the city – it contains some of the wealthiest neighborhoods. The DAs that have the greatest concentration is a smaller area than the similar classification for the location quotient for burglary, but is present nonetheless. The difficulty with this mapped output, as mentioned above, is that the Theil Index is a general measure of specialization and the research is forced to ask: specializes in what?

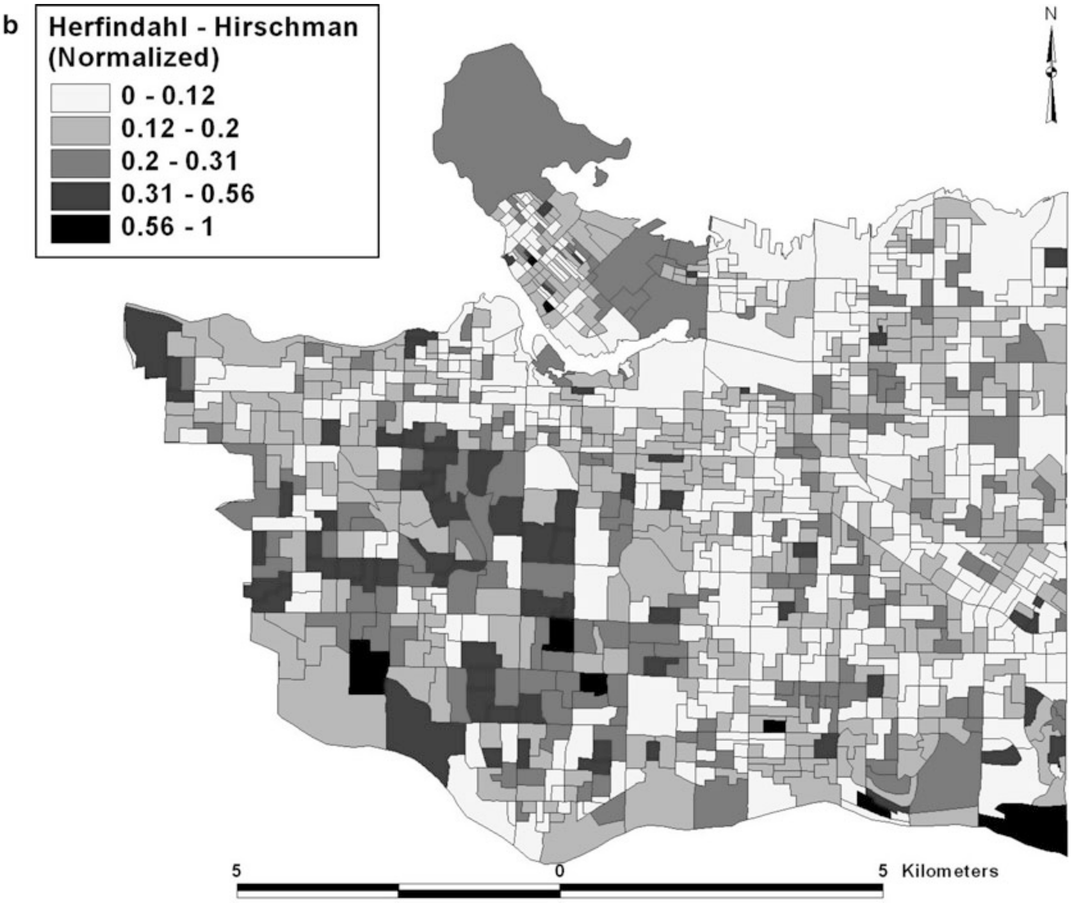

Measuring Crime Specializations and Concentrations, Fig. 2 (continued)

The Herfindahl-Hirschman (Normalized) Index (Fig. 2b) shows a similar pattern to that of the Theil Index: crime concentration in the western port of Vancouver, but not nearly as much as shown using the location quotient for burglary. Also noteworthy are the legend categories for the Herfindahl-Hirschman (Normalized) Index. Inspection of the legend and the mapped output reveals that there are not many DAs that have index values greater than 0.50. This is not problematic but shows that the Herfindahl-Hirschman (Normalized) Index is a more conservative measure of specialization than the Theil Index.

The Specialization Index (Fig. 2c) has the same general pattern as shown in Figs. 2a and 2b but is a blend of these two specialization indices: the mapped output for Specialization Index is more similar to the Theil Index, but the legend categories are more similar to the Herfindahl-Hirschman (Normalized) Index. The primary difference that emerges with the Specialization Index is that there are more DAs that exhibit specialization than in the previous two alternative measures of specialization.

Lastly, the Entropy Index (Fig. 2d) again exhibits a similar spatial pattern to the previous alternative measures of specialization. However, the number of DAs that show a high degree of specialization using the Entropy Index is rather low. Overall, and not surprising given the high magnitude correlations between the different measures of specialization (Table 1), the spatial patterns of the different alternative measures of specialization are quite similar.

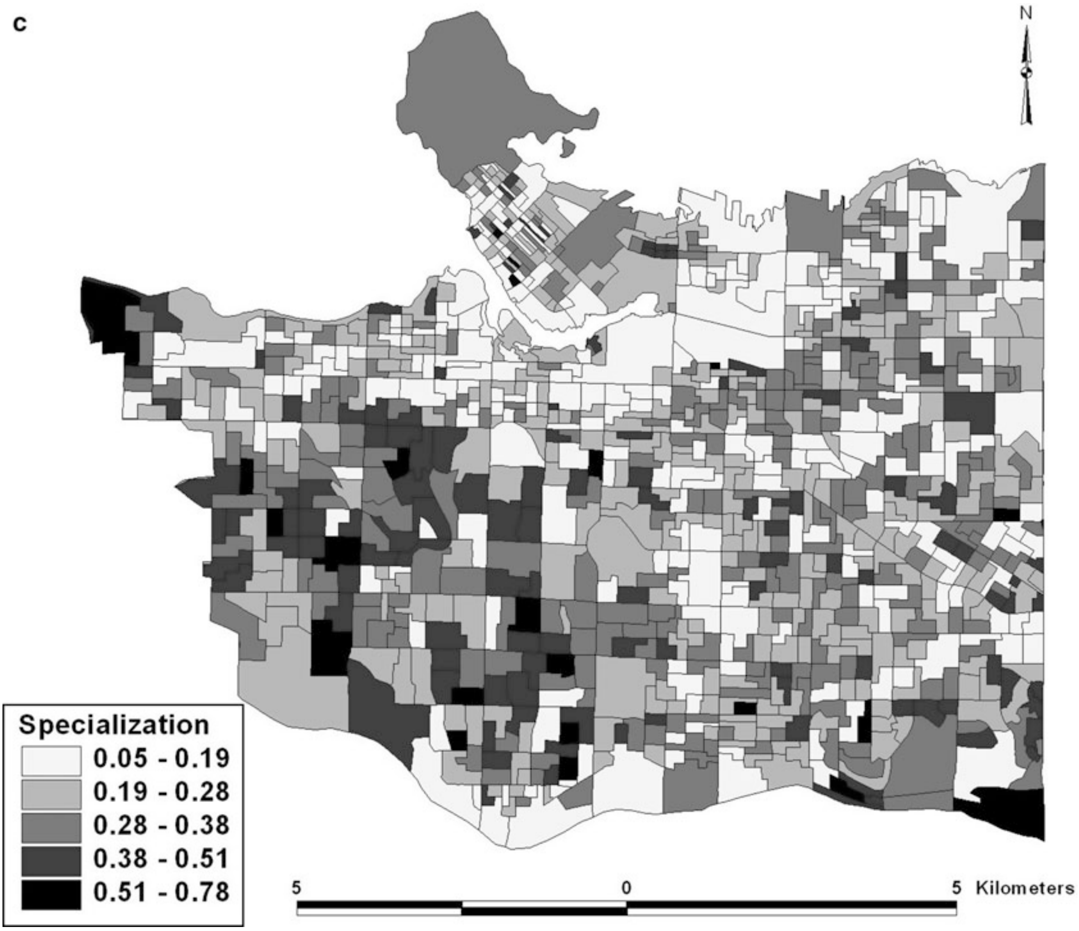

Measuring Crime Specializations and Concentrations, Fig. 2 (continued)

Future Directions for Crime Specialization

This entry has discussed the need to separately consider crime concentration and crime specialization, presented the location quotient and reviewed the criminological literature that has employed the location quotient, and presented and mapped some alternative measures of specialization. It has also opened the debate on which method might prove the most useful in measuring crime specialization. Though the different measures of specialization vary in their calculations, the output of these different measures of specialization is remarkably similar. This indicates that the choice of which measure of specialization the research decides to use has little impact on the overall spatial pattern – all of the alternative measures of specialization are general measures in that they do not show which crime classification an area is specializing in.

Because of this, unless there are very few zero values for crime in an analysis, the Theil Index and the Entropy Index are not ideal in the context of crime in smaller areal units of analysis (micro-places). This leaves the Herfindahl-Hirschman (Diversity) Index and the Specialization Index. Because of its current use in the criminological literature and its ease of calculation, the Herfindahl-Hirschman (Diversity) Index is probably the best of all the alternatives.

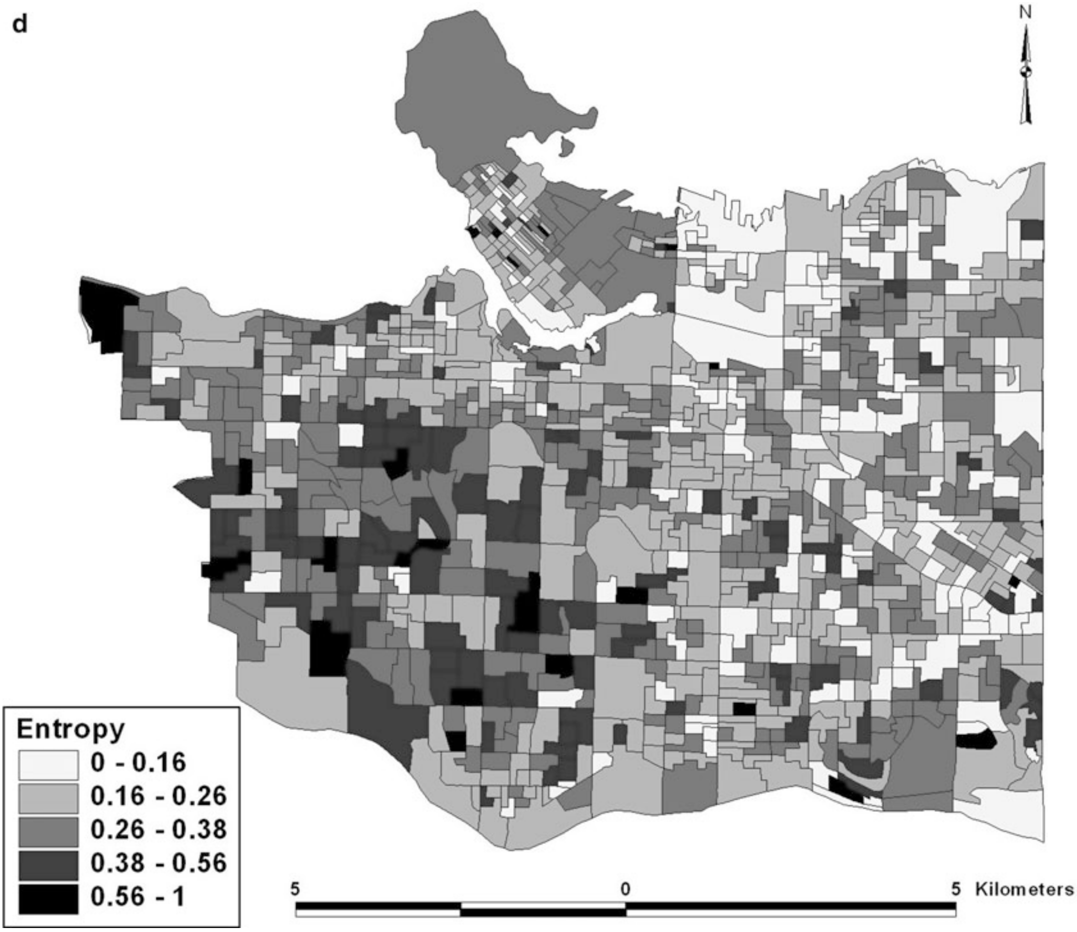

Measuring Crime Specializations and Concentrations, Fig. 2 (**a**) Theil Index, Vancouver 2001, (**b**) Herfindahl-Hirschman Index, Vancouver 2001, (**c**) Specialization Index, Vancouver 2001, (**d**) Entropy Index, Vancouver 2001

Again, the researcher is still left to ask: this place specializes in which crime? It is recommended that an initial measure of specialization, generally speaking, be calculated to investigate whether or not *any* specialization exists. If specialization does occur, then the time may be taken to calculate the various location quotients for each crime classification under analysis. In the case of Vancouver, only burglary specializes in the spatial patterns evident in Fig. 2a-d. As such, the alternative measures of specialization all indicate that specialization does indeed occur in Vancouver and the location quotient can then be used to identify what type of specialization is taking place. This information may then be used to identify the opportunity surface relative to each crime classification to understand the particulars of crime specialization.

With the respective opportunity surfaces identified, an obvious extension is to find what drives the specialization of different crime classifications. In other words, why is crime specialization particularly conspicuous in certain places? There are a number of data sources that may prove instructive for this endeavor. The census is an obvious data set that is available to most researchers. However, because of the time lag between census years (as long as 10 years), these data may not be the most reliable. Another data source that has proven useful in identifying crime generators and crime attractors is land use data (Kinney et al. 2008); Kinney et al. (2008) found that crime generators and crime attractors are best identified using highly specific land use classifications. It is expected to be similar for crime specialization.

Lastly, with information regarding the presence of crime specialization, the location of crime specialization, and why it is present in certain places, this may be used to inform policy decisions. Specifically, such information may be of interest to police in order to address specific crime issues. For example, the police may not only be interested in burglary in the places where it is most frequent (based on counts) but also in the places where it is the most problematic even if its incidence level is not particularly high. Similarly, this information may be of use for crime prevention initiatives, more generally, not just policing.

Related Entries

- ▶ Crime Mapping
- ▶ History of Geographic Criminology Part I: Nineteenth Century
- ▶ Network Analysis in Criminology
- ▶ Spatial Models and Network Analysis

Recommended Reading and References

Andresen MA (2006) Crime measures and the spatial analysis of criminal activity. Br J Criminol 46:258–285

Andresen MA (2007) Location quotients, ambient populations, and the spatial analysis of crime in Vancouver, Canada. Environ Plann A 39:2423–2444

Andresen MA (2009) Crime specialization across the Canadian provinces. Can J Criminol Crim Justice 51:31–53

Andresen MA, Malleson N (2011) Testing the stability of crime patterns: implications for theory and policy. J Res Crime Delinq 48:58–82

Balassa B (1965) Trade liberalisation and “revealed” comparative advantage. Manchester Sch Econ Soc Stud 33:99–123

Block S, Clarke RV, Maxfield MG, Petrossian G (2012) Estimating the number of U.S. vehicles stolen for export using crime location quotients. In: Andresen MA, Kinney JB (eds) Patterns, prevention, and geometry of crime. Routledge, New York, pp 54–68

Brantingham PL, Brantingham PJ (1981) Notes of the geometry of crime. In: Brantingham PJ, Brantingham PL (eds) Environmental criminology. Waveland Press, Prospect Heights, pp 27–54

Brantingham PL, Brantingham PJ (1993) Location quotients and crime hot spots in the city. In: Block CR, Dabdoub M (eds) Workshop on crime analysis through computer mapping, proceedings. Criminal Justice Information Authority, Chicago. pp 175–197

Brantingham PL, Brantingham PJ (1995) Location quotients and crime hot spots in the city. In: Block CR, Dabdoub M, Fregly S (eds) Crime analysis through computer mapping. Police Executive Research Forum, Washington, DC, pp 129–149

Brantingham PL, Brantingham PJ (1998) Mapping crime for analytic purposes: location quotients, counts and rates. In: Weisburd D, McEwen T (eds) Crime mapping and crime prevention. Criminal Justice Press, Monsey, pp 263–288

Bulwer HL (1836) France, social, literary, political, volume I, book I: crime. Richard Bentley, London

Clarke RVG, Cornish DB (1985) Modeling offenders’ decisions: a framework for research and policy. Crime Justice Annu Rev Res 6:147–185

Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44:588–608

De Benedictis L, Gallegati M, Tamberi M (2009) Overall trade specialization and economic development: countries diversify. Rev World Econ 145: 37–55

Eck JE, Weisburd D (eds) (1995) Crime prevention studies, volume 4, crime and place. Criminal Justice Press, Monsey

Frank LD, Andresen MA, Schmid TL (2004) Obesity relationships with community design, physical activity, and time spent in cars. Am J Prev Med 27:87–96

Herfindahl OC (1950) Concentration in the U.S. steel industry. Unpublished doctoral dissertation, Columbia University

Hirschman AO (1945) National power and the structure of foreign trade. University of California Press, Berkeley

Johnson SD, Summers L, Pease K (2009) Offender as forager? A direct test of the boost account of victimisation. J Quant Criminol 25:181–200

Kinney JB, Brantingham PL, Wuschke K, Kirk MG, Brantingham PJ (2008) Crime attractors, generators and detractors: land use and urban crime opportunities. Built Environ 34:62–74

McCord ES, Ratcliffe JH (2007) A micro-spatial analysis of the demographic and criminogenic environment of drug markets in Philadelphia. Aust N Z J Criminol 40:43–63

Miller MM, Gibson LJ, Wright NG (1991) Location quotient: a basic tool for economic development studies. Econ Dev Rev 9:65–68

Quetelet LAJ (1842) A treatise on man and the development of his faculties. W and R. Chambers, Edinburgh

Ratcliffe JH (2001) On the accuracy of TIGER type geocoded address data in relation to cadastral and census areal units. Int J Geogr Inf Sci 15:473–485
Ratcliffe JH (2004) Geocoding crime and a first estimate of a minimum acceptable hit rate. Int J Geogr Inf Sci 18:61–72
Ratcliffe JH, Rengert GF (2008) Near repeat patterns in Philadelphia shootings. Secur J 21:58–76
Rengert GF (1996) The geography of illegal drugs. Westview Press, Boulder
Sherman LW, Gartin PR, Buerger ME (1989) Hot spots of predatory crime: routine activities and the criminology of place. Criminology 27:27–56
Theil H (1967) Economics and information theory. Rand McNally, Chicago
Weisburd D, Groff ER, Yang S (2012) The criminology of place: street segments and our understanding of the crime problem. Oxford University Press, New York

Measuring Police Performance

Rick Jacobs[1], Don Zettlemoyer[2] and
Lawrence Houston[3]
[1]Department of Psychology, The Pennsylvania State University and EB Jacobs, University Park, PA, USA
[2]Penn State Justice and Safety Institute, The Pennsylvania State University, University Park, PA, USA
[3]The Pennsylvania State University, University Park, PA, USA

Overview

Successful organizations have one critical thing in common: they get the right people on the bus (Collins 2001). Selecting top quality individuals is a critical first step in creating a highly effective organization. In general, this is true but nowhere is it more the case than in a police department. Consider the fact that in almost every other type of organization, deficiencies in selection can at least be partially corrected by successfully introducing supervisors and other leaders from outside the organization. In police departments, it is rare, in fact highly unlikely, that supervisors and other leaders come from anywhere else except those who joined the force as new officers. The pool of applicants for promotions and ultimately departmental leadership is almost exclusively made up of individuals who have come up through the ranks. Ineffective selection programs and the inevitable hiring mistakes that occur even with the best selection processes result in limitations in not only current officers' performance but also in the quality of the candidate pool for promotions.

While getting the right people on the bus or in the present case, into the patrol car, is vital to organizational success, it is equally important that organizations possess well-developed performance appraisal systems to motivate and improve performance of those on the force. Such systems are essential to providing developmental feedback to officers, helping organizations increase the readiness of officers for their current assignments as well as future positions, and understanding where the department's strengths and developmental opportunities exist. Unfortunately, it is the exception rather than the rule that a department has an effective system for appraising and managing officer performance. This missed opportunity for improving the effectiveness of the department can be reversed.

The following entry addresses the role that performance appraisal research can play in the evaluation of police officer performance. First, key areas of research on performance appraisal in organizations will be summarized, pointing out traditional issues found in theory, research, and practice. Next, information is offered on those system elements that lead to effective performance appraisal programs. The process of system implementation is also presented, including preparing officers for performance evaluations, steps departments can take to create a positive environment for launching the process, and proper training of supervisors who will be the evaluators. The entry also addresses issues related to providing performance assessment feedback to officers. Highlighted also is the critical role of trust in the feedback process and the creation of an organizational climate conducive to performance feedback. Finally, the entry concludes with a discussion on the future of performance appraisal and management broadly and specifically within law enforcement organizations.

Key Research Questions in Performance Appraisal

How do we define performance? – Industrial psychologists have debated the definition of performance for decades. Austin and Villanova (1992) referred to this issue as the “criterion-problem,” suggesting that traditional measures of job performance fail to conceptualize and measure performance constructs that are complete, accurate, and multidimensional in nature. Specifically, such measures often suffer from criterion deficiency and contamination. While the former refers to aspects of the performance domain not measured by the appraisal instrument (leaving out of the evaluation critical performance factors), the latter refers to specific measurement error created by including in the assessment measures that are either not part of the job requirements or things that are out of the control of the officer being evaluated (i.e., unrelated elements that are unintentionally measured as a part of the performance domain). As such, work psychologists have underscored the importance of conducting a thorough job analysis to fully define the performance domain, develop a holistic understanding of key job-related behaviors, and thereby develop a valid appraisal system.

Before conducting a job analysis, however, it is important to have an understanding of what performance means in a given job. Campbell (1999) defined performance as “behavior or action that is relevant for the organization’s goals and that can be scaled (measured) in terms of the level of proficiency (or contribution to goals) that is represented by a particular action or set of actions.” (pp. 402–403). Motowidlo (2003) defined performance as “the total expected value to the organization of the discrete behavioral episodes that an individual carries out over a standard period of time” (p. 51). In short, performance is what employers pay employees to do (Campbell 1999). In the context of police performance, it includes departmental expectations, as well as what the public demands (viz., behaviors associated with preserving life and property).

Additionally, job performance can be described in terms of two broad arenas, task performance and contextual performance. Task performance is related to the core elements of the job while contextual performance supports the organizational, social, and psychological environment in which the task performance occurs.

For example, while task performance for police officers might include monthly documentation of traffic stops, creating accurate reports, testifying in court, and listening to citizens describing events, contextual performance includes behaviors related to boosting squad morale, providing counsel to a fellow officer, or being available for additional assignments. Both factors contribute independently to overall performance ratings.

What have we done to better understand performance assessment? – Much of the research on performance appraisal (PA) prior to 1980 was dominated by measurement issues and the search for “the right” rating format. This research focused on comparing various rating formats, including graphic ratings scales, behaviorally anchored ratings scales, traditional Likert scales, behavioral checklists, and other formats (Jacobs et al. 1980; Landy and Farr 1980). However, after thirty years of research on the topic, the field became less interested in issues of rating formats, concluding that although raters may have a preference for one format over another, differences in ratings due to format are minimal and overall inconsequential when it comes to the overall performance information provided to the individual and the organization. They did, however, note that behavioral anchors are better than simple numerical scales or ones that rely on common adjectives, and that anchoring should be based on methods that are grounded in firm psychometric theory. In fact, right now any organization, police or other, can find a variety of tools for measuring performance and can be ready to launch a program without the burden of creating a site-specific rating form. With respect to police officer performance, the reader of this entry can contact the senior author and receive a set of scales to be used for officers, supervisors, or command personnel.

Moving beyond issues of rating format, researchers began to consider the role of rater cognition in the performance appraisal process, leading to a flurry of research on the topic during the 1980s. During this time, increased attention was paid toward understanding *how* raters reach their judgments of employee performance. Such research targeted various cognitive processes of raters including observation and information acquisition, encoding and categorization, storage, retrieval, integration, and evaluation (Murphy and Cleveland 1995). Moreover, one of several themes to emerge from this literature was the effects of raters' implicit theories on ratings, including the tendency for raters to commit various rating errors including halo, leniency, and systematic distortion based on assumed rather than actual ratee behavior (Ostroff and Ilgen 1992).

The 1990s brought with it an emphasis on performance appraisal as embedded within the larger social-organizational context (Ilgen et al. 1993). Specifically, Ilgen and colleagues (1993) suggested that the appraisal process as entrenched within a rating environment or "social milieu" that raters and ratees occupy during that process. Performance appraisal became an even more complex process, given this newer social-psychological approach emphasizing a variety of contextual influences on performance ratings. Specifically, performance appraisal began to be viewed as a function of several different categories of environmental factors, including distal variables (i.e., organizational climate, culture, economic conditions, etc.), process proximal variables (rater accountability, feedback environment, etc.), structural proximal variables (appraisal goals and purpose, appraisal training, etc.) and rater and ratee behavior (i.e., rater/ratee attitudinal reactions, cognitive reactions, perceptions of justice, etc.). While cognitive research during the 1980s served to shed light on the limits of raters to make accurate ratings, the social-organizational approach expanded our view of the appraisal process by incorporating issues of rating context into the discussion.

More recent research has tended to focus on motivation and reactions to appraisal processes (Levy and Williams 2004). In terms of motivation, it appears that participation in the appraisal process is critical to motivating employees, who must perceive the appraisal system as fair and ethical. With regard to ratee reactions, research has tended to focus on system and session satisfaction on the part of the person being evaluated, perceived utility and accuracy of the performance information, and various forms of justice including procedural, distributive, and interactional, as well as due process. One important finding, in particular, by Folger, Konovasky, and Cropanzano (1992), suggests that perceptions of fairness are achieved when adequate notice, fair hearing, and judgment based on evidence are present in the appraisal system. In general, it appears that fair appraisal systems lead to a wide array of positive rater and ratee outcomes, including less emotional exhaustion, increased acceptance of feedback, more positive reactions to one's supervisor and the organization, and more satisfaction with the appraisal system and the job by the rater and ratee.

Moreover, ratees' reactions seem to be strongly tied to the opportunity to participate in the appraisal process and the amount of information provided about the process (Cawley et al. 1998). Other research suggests that aspects of the leader-member relationship may impact ratee evaluations. In particular, trust (a topic that will be returned to later), and specifically trust in one's supervisor, seems to be particularly important in predicting both incumbents' and supervisors' acceptance of their appraisal and satisfaction with the appraisal system (Hedge and Teachout, 2000). Further, the presence of a feedback culture, often enhanced by trust and strong supervisor-subordinate relationships, has been empirically related to satisfaction with feedback, motivation to use feedback, and feedback seeking (Steelman et al. 2004).

Finally, a considerable amount of work has addressed the role of rater training in the appraisal process and that work supports the importance of properly training those responsible for conducting performance evaluations. While a variety of methods are available to train raters, the general consensus is the most effective rater training is achieved using what is known as frame-of-reference ("FOR") training (Sulsky and Day 1992).

The performance appraisal literature has come a long way in the past sixty years. What began with research primarily dedicated to measurement issues and the search for the perfect rating format was followed by two shifts in the research literature, the cognitive approach of the 1980s, and the social-organizational perspective of the 1990s. This rich research history has provided us with a strong empirical framework for designing effective performance appraisal systems. Clearly, performance assessment is more than just assigning numerical values to a few categories that loosely reflect what an officer has done over the past year. Given this historical backdrop, the next section highlights system elements that lead to effective programs and evidence for performance improvement as a result of such systems.

Steps in Successfully Implementing a Performance Appraisal Programs

The Preamble: Setting the stage – Defining which competencies are required to effectively perform as a police officer is the first step in providing a framework for a useful performance appraisal and management system. However, once these competencies are in place, the question of how to best create a system that accurately measures these competencies still remains. There are a variety of aspects to consider when implementing a performance appraisal system. From training the raters to using the resulting data to make administrative decisions, there are many pieces of the performance appraisal puzzle that have to be in place in order to ensure success.

Often overlooked, some of the most important parts of the performance appraisal process take place even before the system is put in motion. For example, departments need to first ensure that officers are aware of the competencies being evaluated, how they are linked to the job, and why they are important. As referenced previously, this method of performance appraisal implementation is called "due process." Due process refers to a lifecycle approach to performance appraisal – one in which officers are informed from the start which competencies are important in defining job performance and which behaviors are viewed as effective or ineffective. This would provide officers with a clear understanding of how performance ratings will be derived.

Ensuring that officers are "on board" with the way in which performance is defined and measured, as well as the key features of the system, will result in an environment of trust in the appraisal process, as opposed to suspicion. One common finding from both research and practice is that employees do not like being evaluated. Thus, creating clear guidelines for performance from the start allows organizations to avoid the potential for surprise during the evaluation process, and enhances the overall perceptions of system fairness. Employees who perceive political motives underlying the performance appraisal process are likely to have lower job satisfaction and affective commitment, which were both found to be negatively related to job performance.

Ensuring that employees perceive the performance appraisal process as valid and worthwhile is important for maintaining a productive and satisfied workforce. Table 1 reflects a list of areas of responsibility for police officers and the basics upon which a system can be built for evaluating the performance of police officers. As can be seen, the areas of performance being rated are specific and clearly a part of a police officers' job. Specific performance assessment scales further enhance the performance evaluation process for each of the areas. Table 2 reflects a similar list that represents police supervisor/sergeant. Scales based on this type of approach are available by contacting the senior author of this entry.

In addition to presenting employees with clear and relevant criteria, it should also be made salient that it is possible for employees to voice their opinion if a part of the system appears to be unfair or unclear. The ability to participate in the appraisal process by respectfully questioning the system is also a part of due process. In line with this idea, Cawley et al. (1998) empirically demonstrated the effects of employee participation in the appraisal process. The authors found that employees had more favorable reactions to

Measuring Police Performance, Table 1 Performance evaluation dimensions for police officer

Dimension	Description
Patrol Preparation and Relief	Activities performed at the beginning and end of a shift during shift changes. This includes safeguarding and accounting for all agency property and equipment.
Patrol/Guard Duties and Responsibilities	Activities involving observing assigned area to detect unusual activities or violations of the law and preserving the safety of individuals and security of property in assigned area.
Vehicle Enforcement and Control	Activities involving the safe and legal operation of vehicles. This includes enforcing laws, handling accidents, assisting accident victims, and eliminating public safety hazards.
Responding to Crimes and/or Disturbances	Activities performed while responding to all types of offenses including crimes in progress, civil matters and domestic disputes, property crimes, serious crimes, etc. This includes the initial investigation of crimes and/or disturbances.
Apprehension and Control of Suspects, Prisoners/Inmates	Activities involving pursuit, isolation, containment, apprehension, and control of suspects or prisoners/inmates. This also includes activities performed in the process of searching to locate suspects, prisoners/inmates, or missing persons.
Search	Activities involving searches of suspects, prisoners/inmates, or crime scene/facility to locate physical evidence or contraband.
Arrest and Detention	Activities involving arrest processing procedures, transporting, detention, and lodging of prisoners/inmates or juveniles.
Evidence/Property	Activities involving the safeguarding, storing, and accounting for evidence and non-agency property.
Emergency Situations	Activities involving responding to emergency situations (e.g., natural disasters, riots, hostage situations, fires, escapes, floods/high water, etc.), maintaining order and notifying appropriate agencies to secure assistance.
Care of Victim, Prisoner/Inmate Welfare	Activities involving attending to medical and emotional needs of victims, agency personnel, other citizens, or prisoners/inmates.
Inmate Supervision	Activities involving supervising or controlling inmates during daily activities to maintain safety and security of inmates and staff. This includes communicating with inmates, overseeing the visitation of inmates and maintaining a clean and safe environment during work, meals, showers, etc.
Notifications/ Communications/ Administration	Activities involving distribution of information regarding incidents and conditions. This also includes communications with superiors, other personnel, and individuals from other agencies to coordinate activities.
Reports, Forms, and Memo Books	Activities involving the preparation of written forms, reports, photo/video records, or memo books. Forms and reports may be of a variety of types, such as: captioned (fill-in-the-blank), written narrative, or a combination of these types.
Court Activities	Activities involving providing assistance with court procedures including appearing in court and presenting testimony.
Public Interaction	Activities involving interactions with members of the public. This includes answering questions about community conditions and problems.
Agency Policies, Procedures, Rules, and Laws	Activities involving keeping current with agency policies and rules, and City, State, and Federal laws that govern the activities of officers and applying/enforcing them correctly.
Professional Development	Activities performed to improve one's skills or to improve the agency. This includes participating in specific training or education, participating in professional organizations, or conferences.

a performance appraisal system in which they could voice their opinion about the accuracy of their performance ratings and the system in general, regardless of whether or not the feedback would have an impact on organizational practice in the future. Thus, it is clear that employees are more comfortable with appraisal when they are well informed and feel that they have the opportunity to provide input about the system.

Rater Training: An Absolute Requirement – Once the performance appraisal system has been rolled out, it is important to identify those

Measuring Police Performance, Table 2 Performance evaluation dimensions for police sergeant

Supervision of Field/Case Work	
Incident Supervision	This duty involves assuming responsibility for incidents or special operations, to ensure the efficient and safe handling of the incident/operation and the preservation of life and property.
Case Management	This duty involves assuming responsibility for investigations coming under the sergeant's supervision to ensure the productive and appropriate handling of all investigative matters and to preserve the integrity of these investigations.
Personnel Management	
Personnel Assignment and Coordination	This duty involves assigning or allocating personnel so as to ensure sufficient personnel resources are available to handle workload. This area also involves planning and preparing for special circumstances such as sporting events and emergencies.
Personnel Evaluation	This duty involves observing subordinate performance to identify strengths and areas of needed improvement. It also includes conducting formal performance evaluation sessions and informal counseling sessions with subordinates to recognize performance strengths and discuss and resolve performance problems.
Policy Implementation and Enforcement	This duty involves interpreting, enforcing, and explaining the rules and regulations that govern the activities of Department personnel.
Training	This duty involves ensuring that personnel are properly trained to carry out their assigned duties. This includes planning, developing, conducting, monitoring, and evaluating formal and informal training programs. The training context can be the Academy or the field. The trainees may be Department personnel at any level, personnel from other agencies, or civilians.
Administrative Activities	
Record and Report Management	This duty involves reviewing, preparing, and/or maintaining logs, records, forms, memos, reports, and other field and administrative documents and correspondence used in the course of performing the job. This includes reviewing documents prepared by subordinate or other personnel for completeness and accuracy, reviewing and authorizing personnel requests and integrating information from multiple sources into summary documents.
Internal/External Communication and Coordination	This duty involves communicating or coordinating activities, with Department personnel and individuals from other agencies, to accomplish work objectives and discuss issues of mutual concern.
General Administration	This duty involves participating in the development, implementation, and evaluation of Division/Department programs, policies, procedures, and objectives.
Hands-on Field/Case Work	
Traffic Observation, Enforcement, and Control	This duty involves all activities conducted to ensure the safe and legal operation of vehicles on the road. This includes performing or supervising others in performing tasks such as assisting motorists, enforcing traffic laws, handling accidents, assisting accident victims, and identifying and eliminating public safety hazards.
Responding to Crimes and Disturbances	This duty involves the activities performed while responding to all types of offenses including crimes in progress, civil matters and domestic disputes, property crimes, serious crimes, etc.
Investigation	This duty involves all activities performed as part of internal or criminal investigations. This includes performing or supervising others in performing tasks such as obtaining and analyzing information, collecting physical evidence, and participating in the judicial/administrative process.
Arrest-Related Activities	This duty involves activities performed for the purpose of apprehending, restraining, arresting, transporting, and detaining suspects to be taken into custody according to the law and Department guidelines. It also includes activities performed or supervised in the process of searching vehicles, persons and/or premises for weapons, fruits of a crime, or contraband in order to effect arrest, protect self and the public, and/or to obtain evidence.

(*continued*)

Measuring Police Performance, Table 2 (continued)

Public and Community Relations	
Public and Community Relations	This duty involves activities which have an impact on the Department's image in the community by virtue of the subordinate's interactions with individual members of the public, community organizations, and the media.
Professional Development	
Professional Development	This duty involves participating in activities to keep apprised of job-related developments in laws, technology, policies, and procedures as well as to enhance job-related knowledge, skills, and abilities. This can include reading internal memos and bulletins; reading external publications; participating in training or certification drills and classes; and/or attending outside conferences, seminars, and courses.

individuals who will be responsible for providing ratings and to properly train them to use the system. As discussed previously, raters may fall prey to a variety of rating biases, such as halo (employees are either all good or all bad) or leniency (every employee is great) without proper training. Further, raters may decide to pursue personal goals when providing ratings for their officers (e.g., providing low ratings to an exemplary officer as a motivator for even better performance or providing high ratings to a poor performing officer to get them out of a particular unit and under someone else's command). There is also evidence that when raters have other goals in mind such as promoting harmony and a sense of equality across ratees, their ratings tended to be less discriminating, no longer sorting strong performers from weak ones and resulting in higher ratings in general. This provides support for the notion that when raters have a goal in mind, they can paint a picture of performance that is in alignment with that goal but not necessarily consistent with what they observe.

In the end, a system is only as good as the people who use it. It is completely possible for an organization to have a fantastic performance appraisal system in place and still achieve mediocre or even disastrous results. Thus, it is helpful to view the performance appraisal system as a vehicle – all of the parts may be in place but, without a properly trained driver and periodic servicing, the car does not function properly and thus at risk of malfunctioning. That being said, providing rater training is clearly a key component of a successful performance appraisal process.

As mentioned earlier, one type of rater training that has been touted as particularly effective in driving an accurate performance system is frame-of-reference (FOR) consisting of a variety of components which all help to make certain that raters will be able to accurately assess officers' behavior. FOR training helps in creating a unified definition of performance across raters, so that officers will receive similar ratings regardless of their rater. As such, FOR training is an extremely valuable tool for increasing the validity and predictive capabilities of performance ratings. FOR training also provides raters the opportunity to practice assessing behavior and to compare ratings with other raters to help provide a consistent and clear basis upon which raters should be formulating performance ratings. For example, raters who can effectively evaluate officers will gain additional and reinforcing insights from the training, while officers who are not accomplished performance evaluators will learn important skills and gain confidence in their ability to properly evaluate others. Systems that fail to provide rater training (i.e., FOR) will be far less effective.

Of equal concern for performance management efforts is feedback. Anxiety and lower levels of motivation in ratees often arise from misused and improperly delivered feedback. In the next section, proper feedback delivery and environment will be discussed to ensure that organizations are making the best possible use of performance information. As should be obvious, any rater training process must include information on the proper conduct of the feedback process.

Feedback in Performance Appraisal

Reference different fields of study and you will find various ways in which the term “feedback” has been conceptualized and defined. Industrial psychologists generally refer to feedback as information provided to an individual that clarifies expectations of their job and alerts them to how well their current performance meets such expectations (Ashford and Cummings 1983; Spector 2006). Although it was long assumed that employee feedback resulted in positive gains in performance, a meta-analysis by Kluger and DeNisi (1996) exposed the reality that feedback not only has the potential to produce performance gains, but also performance declines. As Taylor, Fisher, and Ilgen (1984) note, “Feedback may have no impact on the recipient at all, it may cause the individual to lash out angrily, or it may result in a response quite different from that desired by the source” (p. 82). Ilgen, Fisher and Taylor (1979) suggest that variability in response to feedback might be due to a series of psychological or cognitive processes elicited by the feedback process.

Many studies have sought to determine what characteristics of the feedback process contribute to its effectiveness or lack thereof but before considering such characteristics, it is important to understand that the effectiveness of any feedback system hinges upon recipients’ acceptance of such feedback (Ilgen et al., 1979). Ilgen and colleagues (1979) defined feedback acceptance as “the recipient’s belief that the feedback is an accurate portrayal of his or her performance” (p. 356). One way to think about this is that feedback acceptance acts as a gatekeeper or moderator between feedback delivery and effectiveness. Using self-consistency theory, Korman (1970) argued that, “individuals will be motivated to perform on a task or job in a manner which is consistent with the self-image with which they approach the task or job situation” (p. 32). Thus, one’s response to performance feedback is expected to be congruent with his or her acceptance of such feedback. As such, positive or negative feedback, which is perceived as accurate and therefore accepted by the recipient, is more likely to motivate one to respond than feedback which is perceived as inaccurate.

Given the crucial role feedback acceptance appears to play in the feedback delivery-performance relationship, what factors might predispose recipients to perceive feedback as inaccurate and unacceptable, and what can organizations do to foster acceptance? In the following sections, the effects of feedback content, specificity, frequency, delivery, and source on individuals’ willingness to accept and utilize performance feedback will be discussed. Moreover, the role of trust in the feedback process will be explored, as well as how organizations can create environments conducive to feedback.

Feedback Content – Given the sensitive and potentially anxiety-inducing nature of feedback in organizations, it is no surprise that the content of feedback may have a substantial impact on the likelihood of feedback acceptance. In general, individuals are more accepting of positive feedback, which likely stems from the motivation to preserve one’s self-esteem from the perceived failure resulting from negative feedback. Individuals are more likely to reject feedback that reflects discrepancies between one’s actual performance, relayed through feedback, and one’s perceived performance.

This does not necessarily indicate that individuals wholly reject negative feedback. It does indicate, however, that when performance discrepancies exist, individuals prefer positive feedback signs that portray them in a positive light and confirm their own self-perceptions. Moreover, it suggests that feedback providers ought to balance their negative feedback with positive feedback in order to reduce perceived performance overestimations as much as possible. By narrowing the perceived gap between one’s actual performance and perceived performance, feedback providers may be more likely to gain feedback acceptance, which in turn may lead to performance improvements on the part of feedback recipient. Additionally, Smither and Walker (2004) showed that providing a small amount of negative feedback enhances performance, but that a large amount of negative feedback impairs future performance. Atwater and Brett (2006) further suggest that feedback sessions should begin with positive feedback in order to spur feedback acceptance. Thus, feedback providers should be aware of the amount of negative feedback they provide, making sure to buffer it with positive feedback at the beginning of the feedback session and ensuring that subordinates are not overwhelmed by too much negative feedback during the process.

Other research indicates that feedback is perceived as valuable when it helps employees reduce uncertainty, provides recipients information regarding their goal progress, and indicates how one's performance is being evaluated by others. Additionally, feedback is more likely to be accepted and used when it is perceived as valid, accurate, and reliable. Thus, feedback providers should always err on the side of providing higher quality feedback during feedback sessions.

Kluger and DeNisi (1996), however, note that such feedback ought to be focused more on the performed task than the person receiving the feedback. In their influential meta-analysis, they found that feedback intervention (FI) effectiveness decreased as feedback became more directed toward the feedback recipient than the task. This is consistent with other research that supports the idea that individuals are more likely to accept feedback that compares them to a neutral standard, rather than to their peers. It is believed that such comparisons can hurt recipients' self-concept and reduce their likelihood of accepting feedback. This suggests that recipients may be less likely to accept and ultimately use feedback that they perceive as a personal attack rather than an honest assessment of performance on the task. Thus, feedback providers should be careful that their feedback is not directed toward personal characteristics or compares recipients to their peers, but rather should focus on specific performance aspects related to the task.

Feedback Specificity – To date, feedback specificity, or the level of information presented in feedback messages, has received quite a bit of attention in the performance appraisal literature. Generally speaking, it is believed that effective feedback includes specific information about behaviors that were performed incorrectly and how to correct them, rather than statements that are generalized or ambiguous. Such feedback provides a rich source of information about explicit behaviors that are appropriate or inappropriate for effective performance, thereby allowing recipients to learn from and correct their behavior. In doing so, feedback specificity is thought to decrease information-processing activities such as error diagnosis, encoding, and retrieval, which subsequently reduce the cognitive load required to make links between actions and outcomes. Moreover, specific feedback provides information on how individuals are progressing toward their goals and informs them about how their job performance is being evaluated.

In fact, empirical evidence suggests that specific, objective feedback that is consistent with actual performance leads to higher performance than less specific, more subjective feedback. Moreover, feedback interventions that are supplemented with information regarding tasks, strategies, and appropriate behaviors have been found to result in increased short-term performance during practice and training. Interestingly, Steelman and Rutkowski (2004) demonstrated that employees are motivated to use negative feedback when it is perceived to be of high quality and directed toward improving their performance than similar feedback of lower quality.

Despite such findings, overly specific narrative feedback may also introduce excessive cognitive load on recipients, hindering their understanding of the feedback. The positive impact of feedback can clearly be enhanced when the information being conveyed is easy to consume and in a form that is readily understood by the recipient. This point argues for a graphic presentation of information as perhaps a takeaway from the feedback session. It may also include the use of summary statistics that make it easy to compare current performance to expectations or even to the performance of peer groups.

In general, research has suggested that specific feedback is more effective than vague, generalized feedback. Employees must learn which specific behaviors are indicative of good and poor performance – not only so that they can correct such behaviors but also in order to protect against perceptions of arbitrariness in the feedback process.

Feedback Frequency – Feedback frequency has also been explored as a potential factor influencing feedback acceptance and perceived utility. In general, it is thought that frequently delivered feedback is more effective in promoting performance improvements than less frequent feedback. The logic that drives this conclusion is that the close temporal pairing of feedback with employee performance behavior has an overall positive effect on future performance. Frequently delivered feedback reduces the chances that interference will occur in the periods between behavior and the feedback, which have the potential to distort feedback effectiveness. This stands in stark contrast to most performance evaluation processes, which call for assessment and feedback once per year. In the course of performing any job and especially a police officer's, much can happen over that 12-month period.

In fact, research suggests that frequency of feedback is positively related to recipients' perceived accuracy and fairness of feedback. Because feedback tends to not be delivered as frequently as it should in today's organizations, recipients may view feedback as a scarce, valuable resource that decreases role ambiguity by continually informing employees about where they stand with regard to their performance levels. Thus, increasing the frequency of feedback may facilitate more positive employee reactions and outcomes during the feedback process. While once per year seems too little, other factors must be taken into consideration when deciding whether to provide more frequent feedback. For example, instituting more frequent feedback sessions may send implicit messages to employees that supervisors do not trust them as being competent and capable of carrying out their tasks without increased micro-management. In fact, it has been shown that while positive feedback provided immediately promotes a strengthening of previous behavioral responses, negative feedback is more effective after a time lag, which allows for decomposition in the strength of the previous incorrect behavioral response. As such, it stands to reason that while providing feedback on how to do a job may be perceived positively when done once or twice, such feedback may be reacted to more negatively when delivered in excess.

Supervisors should seek to provide regular feedback to employees, but they should also recognize that more is not always better. Specifically, because recipients' feedback perceptions are crucial to its acceptance, supervisors should be aware that the intent behind providing increased amounts of feedback may not be perceived accurately by recipients. What may be intended as supportive may in actuality be perceived as controlling. One simple recommendation is that supervisors find out how often their employees desire feedback and use that in conjunction with the requirements of the organization to create a dialogue about performance.

Feedback Delivery – It is also important that feedback providers take active steps to preserve and respect recipients' self-esteem by being sensitive and understanding in order to obtain feedback acceptance, rather than simply viewing their role as a provider of information. The manner in which feedback is delivered may significantly impact employees' acceptance and perceived utility of the feedback. For example, it has been found that a supportive, constructive attitude on the part of feedback providers contributes to increased recipient satisfaction, perceptions of fairness, and motivation to improve job performance. Furthermore, feedback consisting of greater interpersonal fairness has been found to be related to more favorable attributions toward the feedback provider, greater acceptance of feedback provider, and more favorable reactions to the organization. Officers can be seen as motivated to improve their job performance following negative feedback when feedback is delivered in a considerate manner and with information that helps them better understand what was done improperly as well as what might be seen as more effective.

Feedback Source – Not surprisingly, much research has also investigated whether feedback acceptance depends on who provides such feedback. This research has often investigated the feedback source's credibility, which refers to one's expertise and trustworthiness in providing feedback. Expertise on the part of the person conducting the evaluation includes knowledge of the job requirements, knowledge of actual job performance, and the ability to evaluate performance in an accurate manner, while trustworthiness represents whether or not the individual being evaluated has a belief that the feedback source can and will provide accurate performance information.

Credible feedback providers are seen as possessing expertise relative to the specific tasks being evaluated and are trusted to provide feedback that is objective and free from bias (e.g., political considerations, feedback source's mood at the time of feedback, etc.). In police settings, the former is generally the case since virtually every police officer providing evaluations came through the position they are evaluating (sergeants were police officers, lieutenants were sergeants, etc.). However, trust is a much more elusive characteristic, and one that must be developed over time between the evaluator and the officer being evaluated. This suggests that organizations should make a strong effort to ensure that feedback is delivered by highly qualified individuals with experience on the task being evaluated and professional status that is greater than the feedback recipient's. In doing so, recipients may be more likely to both accept and use feedback in order to improve their future performance.

The Role of Trust in Feedback – Further examination of the feedback process highlights the importance of recognizing that feedback does not exist within a vacuum, but rather is embedded within the larger social-organizational context or what is often referred to as the "social milieu." In fact, research suggests that context plays a potentially pivotal role in shaping aspects of the appraisal process, including feedback, and the ways in which employees react to such processes (Farr and Jacobs 2006; Murphy and Cleveland 1995). Specifically, Folger, Konovsky, and Cropanzano (1992) discussed performance appraisal as a process in which individuals at all organizational levels (including subordinates, supervisors, and upper management) have a stake, leading to potentially conflicting interests regarding the results of a given performance appraisal or feedback session. Feedback providers, in such contexts, are believed to hold power over subordinates, who inevitably make themselves vulnerable during such sessions. Therefore, in order for feedback sessions, and performance appraisal processes in general to be effective, those engaged in the process must trust in it. For feedback recipients, this means trusting their supervisors and the system used to evaluate them. For supervisors, this means having faith in the quality of their feedback and appraisal system that informs it.

An officer's trust in their sergeant influences the communication link between the two. As trust fades, information fails to make a substantive impact during communication. Thus, it stands to reason that when trust decreases so does the impact of the sergeant's feedback on officer's behavior. Providing empirical support for this idea, Herold and Greller (1977) found that supervisors who were more "psychologically close" with their subordinates had a greater impact in terms of feedback than those who were more psychologically distant. In addition, Earley (1986) found that workers' trust in the feedback source partially mediated the relationship between feedback and workers' response to and value attached to praise and criticism. Jablin (1979) provides a nice overview of the supervisor-subordinate relationship and its importance in this context.

In sum, these findings suggest that trust plays a crucial role in individuals' willingness to accept and respond to feedback. Thus, supervisors should actively try to promote trust among their subordinates by educating subordinates about the performance appraisal and feedback process, showing how the feedback process may be used to promote employee development, demonstrating consideration for employees' work and well-being, exhibiting dependability on a regular basis, and communicating a sense of honesty and forthrightness in daily activities, among other things. In doing so, it may be possible to promote a climate of trust surrounding the feedback process, which will facilitate feedback acceptance and ultimately improved job performance.

Summary of What We Know About Police Officer Performance Management

To this point, the primary focus has being on giving the reader a lot of information regarding research and practice as it pertains to performance assessment and management. Much has been done in this area and over the decades of theory building, research, and practical application, quite a bit have been learned. While the importance of being very specific about what has been done in the past is acknowledged, it is often important to step back and simply summarize what is known and what it means in terms of creating a successful performance management program. Here is the authors' perspective.

Defining Performance – Without a doubt, nothing works unless time is spent and effort is expended to accurately define the key responsibilities and expected behaviors of those who are being evaluated. Traditionally, this starts with a job analysis or a very comprehensive job description. This step helps raters and ratees when it comes to understanding performance expectations and accepting the accuracy of performance evaluations. Further, with respect to fairness and meeting potential legal challenges, a system of performance management will never be seen as reasonable unless the agency has done its homework and clearly documented what is required by the job. These descriptions should be multidimensional covering critical areas of job performance and behaviors specifying the types of actions that represent poor, average, and strong job performance.

Setting the Stage – Performance management is not just an organizational requirement. It is an organizational event. In police agencies and other organizations, the process of evaluating performance is often approached with dread. Evaluators do not enjoy conducting evaluations, those being evaluated feel like they are being unnecessarily put under the microscope, and those responsible for the process within the HR organization often feel they have to harass and cajole individuals just to do what needs to be done. It is not a happy time of the year for many members of the organization. With this backdrop, it is important that a more positive environment is created. This can be accomplished by focusing on the process and potential positive outcomes that will result from effective performance management. Creating a positive environment requires a campaign initiated by the agency and supported from the top down. Without this kind of effort, the process will wallow in apathy, and the benefits will never emerge.

Training, Training, and More Training – Organizations consistently make a faulty assumption when it comes to performance appraisal and management. They too often assume that because someone has become a supervisor, they CAN and WILL be competent in the evaluation of others. Many supervisors and command personnel are good at reviewing the work of others and may enjoy the process, but not all. The role of training evaluators on the how to's and why's of performance evaluation is critical. Even for those who are positively disposed to the activity, training helps. An essential part of any performance management system is the training of raters on all aspects of the system.

It is also critical that those being evaluated are informed regarding the process (i.e., setting the stage), which is often a neglected area. While it may be an annual agency requirement that performance reviews will occur, it is not necessarily clear to individuals being evaluated exactly how the evaluations will be conducted, what is being measured, and how the data will be used. All these topics should be explained in advance of system implementation.

Keeping Track/Monitoring the System – The most effective processes do not stop at the collection of data and the checking off of yet another organizational event required by the agency. The best systems review results, hold raters accountable for doing the evaluations in a timely fashion, analyze data for potential problems (bias, halo, and missing information as examples), and use this information to enhance the future system implementations.

Using the Data – Traditionally data from performance evaluations are used to understand individual performance, to provide feedback to the person being evaluated, and/or to trigger personnel actions such as administering rewards or reprimands. While these are clearly important uses for the data, organizations are missing opportunities unless they look across individuals to understand where the agency may have repeated performance decrements indicating needed training programs, or where the department is truly excelling and how that information can be leveraged for future motivational programs. Suffice it to say, the best performance management programs continue beyond the individual level of analysis and look at groups, departments, and agency wide data to better understand performance.

Feedback – "Don't collect information that you just put in a folder:" Over the years, quite a bit has been learned about how organizations can move from performance appraisal to using that information to drive future performance. It begins with a well-developed feedback process that brings supervisors and officers together to talk about the past and to plan for the future. Feedback does not happen automatically. Many involved in the process of delivering and receiving feedback need guidance to effectively turn observations of performance into performance improvements.

The Future of Performance Appraisal and What it Means for Police Departments

Thus far, this entry has outlined best practices for performance appraisal in organizations – from traditional issues regarding instrumentation to how to deliver performance feedback and the importance of trust between an evaluating supervisor and the officer being evaluated. However, performance appraisal methodology has never been stagnant. Changes have been seen in research emphasis and practice over the past five decades. The future will bring improvements and updates to best practice as it stands today. Although it is impossible to predict exactly what the future will hold, there are a few areas that have already begun to show promise, which may allow police departments to enhance the effectiveness of their performance assessment processes.

The first of these areas lies in the realm of online assessment. Online assessment is quick, easy, and fits in well with today's demand for technology-driven solutions. For example, E-learning has become very popular in organizational training because it is easy to use and can be completed by employees at any time. The same is true for online performance appraisal systems. Supervisors can complete surveys about officer behavior at a convenient time and in a quick, easy fashion. As the computerization of police work marches forward, more and more systems will be available for documenting and evaluating police officer performance. The use of online performance appraisal also facilitates self-evaluation and lends itself to automated comparisons between subordinate and supervisor, which can form the basis of performance discussions and goal setting. Online assessments are also extremely useful tools for organizations because they save time, energy, and allow for the automated collection and storage of performance assessment data. Making performance appraisal systems user-friendlier and more efficient is undoubtedly in the best interest of the organization. And, online appraisals certainly help to propel these goals forward.

Second, there has been a recent upswing in the use of self-assessment for performance appraisal. In this case, employees are able to provide ratings regarding their own performance. However, the literature suggests that self-appraisals are ridden with bias. For example, Atkins and Wood (2002) demonstrated that self-ratings were negatively and nonlinearly related to assessment center performance. The worst performers may report that they are doing the best as a way of overcompensating for poor performance. Atkins and Wood's (2002) findings further suggest that high performers may be hard on themselves, creating a performance profile much lower than that which exists in reality. It has also been demonstrated that self-ratings tend to be less variable than supervisor, peer, or subordinate ratings, suggesting possible halo or leniency errors (Scullen et al. 2003). Thus, while self-evaluations may provide interesting information, this information is not suggested for use in making administrative decisions. Organizations should be aware of employees' intent to distort self-ratings and should proceed with caution when determining how to make use of self-report performance data.

Finally, it is time for police organizations to move away from the perspective that they conduct appraisals because their policies and procedures manual indicates they must do so to viewing the performance appraisal system as a vehicle for enhancing communication between supervisor and subordinate, facilitating change, and setting goals for the future performance. It is tempting to consider performance appraisal in police departments a truly lost opportunity but that seems excessively negative. Rather the tone of this entry should convey the tremendous opportunity for organizational success that is embodied in well thought-out and well-executed programs in performance management and how they can truly transform a department. In this time of shrinking budgets and disappearing resources, developing and implementing a state-of-the-art performance appraisal system can be a very low cost, high impact process for any law enforcement agency. This entry, hopefully, has helped convince you and has provided you with information that can move you forward in putting a performance management program into practice in your department.

Recommended Reading and References

Ashford SJ, Cummings LL (1983) Feedback as an individual resource: personal strategies of creating information. Organizational Behavior Human Performance 32:370–398

Atkins PWB, Wood RE (2002) Self-versus others' ratings as predictors of assessment center ratings: validation evidence for 360-degree feedback programs. Personnel Psychol 55:871–904

Atwater L, Brett J (2006) Feedback format: does it influence manager's reactions to feedback? J Occupational Psychol 79:517–532

Austin JT, Villanova P (1992) The criterion problem: 1917–1992. J Appl Psychol 77:836–874

Campbell JP (1999) The definition and measurement of performance in the new age. In: Ilgen DR, Pulakos E (eds) The changing nature of performance. San Francisco, Jossey-Bass, pp 399–430

Cawley BD, Keeping LM, Levy PE (1998) Participation in the performance appraisal process and employee reactions: a meta-analytic review of field investigations. J Appl Psychol 83:615–633

Collins J (2001) Good to great. Harper Collins, New York

Earley PC (1986) Trust, perceived importance of praise and criticism, and work performance. An examination of feedback in the United States and England. J Manage 12:457–473

Farr JL, Jacobs R (2006) Unifying perspectives: the criterion problem today and into the 21st century. In: Bennett W, Lance C, Woehr D (eds) Performance measurement: current perspectives and future challenges. LEA, Mahwah, NJ, pp 321–337

Folger R, Konovsky MA, Cropanzano R (1992) A due process metaphor for performance appraisal. Res Organizational Behav 14:129–177

Hedge JW, Teachout MS (2000) Exploring the concept of acceptability as a criterion for evaluating performance measures. Group Organization Manage 25:22–44

Herold DM, Greller MM (1977) Feedback: the definition of a construct. Acad Manage J 20:142–147

Ilgen DR, Fisher CD, Taylor SM (1979) Consequences of individual feedback on behavior in organizations. J Appl Psychol 64:349–371

Ilgen DR, Barnes-Farrell JL, McKellin DB (1993) Performance appraisal process research in the 1980s: what has it contributed to appraisals in use? Organizational Behav Human Decision Processes 54:321–368

Jablin FM (1979) Superior-subordinate communication: the state of the art. Psychol Bull 86:1201–1222

Jacobs RR, Kafry D, Zedeck S (1980) Expectations of behavioral rating scales. Personnel Psychol 33:595–640

Kluger AN, DeNisi A (1996) Effects of feedback intervention on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychol Bull 119:254–284

Korman AK (1970) The prediction of managerial performance: a preview. Studies Personnel Psychol 2:4–26

Landy FJ, Farr JL (1980) Performance rating. Psychol Bull 87:72–107

Levy PE, Williams JR (2004) The social context of performance appraisal: a review and framework for the future. J Manage 30:881–905

Motowidlo SJ (2003) Job performance. In: Borman WC, Ilgen DR, Klimoski RJ (eds) Handbook of psychology: industrial and organizational psychology. Wiley, Hoboken, NJ, pp 39–53

Murphy KR, Cleveland JN (1995) Understanding performance appraisal: social, organizational, and goal-based perspectives. Sage, Thousand Oaks, CA

Ostroff C, Ilgen DR (1992) Cognitive categories of raters and rating accuracy. J Business Psychol 7:3–26
Scullen SE, Mount MK, Judge TA (2003) Evidence of the construct validity of developmental ratings of managerial performance. J Appl Psychol 88:50–66
Smither JW, Walker AG (2004) Are the characteristics of narrative comments related to improvement in multirater feedback ratings over time? J Appl Psychol 89:575–581
Spector PE (2006) Industrial and organizational psychology: research and practice. Wiley, Hoboken, NJ
Steelman LA, Rutkowski KA (2004) Moderators of employee reactions to negative feedback. J Manag Psychol 19:6–18
Steelman LA, Levy PE, Snell AF (2004) The feedback environment scale: construct definition, measurement, and validation. Educational Psychol Measure 64:165–184
Sulsky LM, Day DV (1992) Frame-of-reference training and cognitive categorization: an empirical investigation of rater memory issues. J Appl Psychol 77:501–510
Taylor MS, Fisher C, Ilgen D (1984) Individuals reactions to performance feedback in organizations: control theory perspective. In: Rowland K, Ferris G (eds) Research in personnel and human resource management. JAI, Greenwich, CT, pp 81–124

Measuring Police Unit Performance

Gary Cordner
Department of Criminal Justice, Kutztown University, Kutztown, PA, USA

Overview

Although police performance measurement has gotten increased attention in recent decades (Shane 2007; Fleming and Scott 2008; Davis 2012), it has been a longstanding concern of scholars and policy makers (Lind and Lipsky 1971; Parks 1971). The ability to measure how well the police are performing is central to notions of police effectiveness, accountability, reform, and the current emphasis on evidence-based policing. Without accurate performance measurement, aspirations toward implementation of strategic policing, scientific policing, and a new approach to police professionalism (Weisburd and Neyroud 2011; Stone and Travis 2011) are mostly rhetoric, not reality.

In the United States, where most police agencies are extremely small (one-half have ten or fewer police officers; Reaves 2011), the majority of the focus has been on the overall performance of the entire police organization, rather than on police unit performance (Langworthy 1999). Even in the United States, however, many police organizations are comprised of multiple subunits; this is the normal characteristic of police organizations in most of the rest of the world. These subunits are important to consider, because the overall performance of a medium-sized or large police agency is, at least to a degree, the sum of the performance of its parts. Top-level police executives have the challenge of measuring the performance of their organizations' subunits in order to guide resource allocation, hold lower-level managers accountable, fix problems, and coordinate interunit activity. Lower-level police managers, especially if they are responsible for specific subunits, need information about how well their own units are performing in order to carry out their responsibilities of planning, directing, and controlling in order to maximize their units' performance.

Fundamental Considerations

Conceptualizing and operationalizing "performance" is not as easy as it sounds. Commonly, a distinction is made between effort, output, and outcome. For example, the performance of a manufacturing unit within a larger company might be defined in three different ways:

1. How hard did the members of the unit work?
2. How many widgets did the unit produce?
3. How much profit was derived from the unit's performance?

Each of these measures of performance is legitimate and might be relevant. The company's top management probably wants to be assured that the members of the manufacturing unit are working hard. If not, the unit might be overstaffed and therefore costing more than necessary. Second, top management certainly wants to know the level of production (output) of the unit, particularly in the case of a manufacturing

Measuring Police Unit Performance, Table 1 Unit measures for a police district or precinct

Dimension	District/Precinct measures
Reduce crime	Reported violent crime per 100,000 population Reported property crime per 100,000 population
Hold offenders accountable	Clearance rate for crimes investigated at the district/precinct level Conviction rate for crimes investigated at the district/precinct level
Reduce fear	Personal fear of crime measured with community surveys Perceived likelihood of being a crime victim measured with community surveys
Ensure civility	Number of injury and fatal traffic crashes Perceived level of social disorder measured with community surveys
Services/ satisfaction	Public confidence in the police measured with community surveys Customer satisfaction with specific services measured with follow-up surveys
Use of force & authority	Percent of custodial arrests that involve police use of force Disproportionality index for vehicle and pedestrian stops (reflecting minority overrepresentation)
Use of financial resources	Percent of annual district/precinct budget expended Overtime expenditures per district/ precinct employee

Measuring Police Unit Performance, Table 2 Measures for a police repeat offender unit

Dimension	Repeat offender unit measures
Reduce crime	Estimated number of offenses not committed by arrested offenders (offending rate times years of incarceration)
Hold offenders accountable	Proportion of targeted offenders arrested, prosecuted, and convicted Average sentence length of targeted offenders
Use of force & authority	Percent of custodial arrests that involve police use of force Percent of cases declined for prosecution due to evidentiary concerns Percent of cases not resulting in conviction due to entrapment findings
Use of financial resources	Percent of arrests not qualifying as repeat offenders Average number of previous convictions of unit arrestees (seriousness factor) Average number of arrests per unit member Average cost per arrest, per conviction, and per crime not committed (personnel, informant payments, buy money, etc.)

unit, since that is its fundamental purpose (produce widgets). Ultimately (third), top management wants to know the unit's contribution to the company's profit. That is arguably the most important dimension of the unit's performance, since it directly coincides with the entire company's bottom line (its profit). After all, if the company is not profitable, it will go out of business, and then there will not be a manufacturing unit or any jobs for the members of the unit.

Several significant complications arise even from this simple example. With regard to effort, members of a unit might work hard, but make many mistakes. They might work hard, but be doing the wrong things. They might work hard, but not have the proper equipment needed to complete their work correctly. Also, if the unit's role is more analytical, intellectual, or creative than this manufacturing example implies, then "working hard" might not be the primary ingredient of the effort associated with good performance. Rather, wise decision making, innovative problem solving, imaginative thinking, or just "working smart" might be better ways than "working hard" of conceptualizing the effort dimension of performance as it relates to some organizational units.

With regard to output, an organizational unit might produce a satisfactory number of widgets but the quality of the widgets might not be acceptable. Similarly, the number of widgets produced might be adequate but the amount of resources consumed in producing them could be excessive – i.e., there might be wastage of raw materials,

Measuring Police Unit Performance, Table 3 Unit measures for a Police Communications Center

Dimension	Communication Center Measures
Services/ satisfaction	Caller satisfaction with services measured with follow-up surveys Percent of 911 (emergency line) calls answered within three rings Quality control assessment of call-answering performance (review of random sample of calls) Police officer satisfaction with dispatcher services measured with surveys
Use of financial resources	Calls handled per unit employee Percent of annual unit budget expended

driving up the cost of producing the widgets. As these two scenarios suggest, the output of a unit cannot be assessed in a meaningful way without incorporating some standard of quality as well as a measure of cost.

In principle, a unit's outcomes are even more important than either its effort or its output, since the outcomes are the end results of the unit's performance. This is particularly true when the outcomes are closely associated with the entire organization's "bottom line," such as its profit. Two related complications arise at this level, however. First, simply determining the end results of a unit's performance can often be difficult. A company's training unit, for example, might have as its desired end results increased employee knowledge, increased employee skills, and improved employee performance. The first two could be measured immediately after the conclusion of training programs. This would give some indication of training outcomes, but subsequent measurements would have to be made to determine whether the employees' knowledge and skills were maintained over time. It would be even more involved (and costly) to measure employee performance back on the job, in order to see whether there was any post-training improvement in actual performance. The greatest challenge would be in "proving" that the training unit's efforts either did, or did not, cause any changes in employee performance. Changes in performance could be caused by other factors – pay, supervision, new technology, etc. Or, measurement of employee performance might indicate no improvement following training – but that might be caused by workgroup pressures (e.g., output limitation norms), failure of supervisors to reinforce training, or many other factors, not necessarily some shortcoming of the training. In other words, validly determining whether a unit's efforts did or did not cause some specified end results (outcomes) sometimes requires a genuine research project, not simply collecting some information.

These challenges are multiplied when one seeks to determine the contribution of a unit's performance to the entire organization's achievement of its specified outcomes. Again, this is what matters the most, but it is difficult to determine. The manufacturing unit may have produced perfectly good widgets at a satisfactory cost, but if the marketing unit failed to market them effectively, or if the sales unit failed to sell them, or if the packaging unit failed to design and produce the right packaging for them, or if the public's taste in widgets changed, the widgets might still be in the warehouse nine months later, not contributing to the company's profit. Would it be fair then to assess the manufacturing unit's performance negatively? This scenario illustrates the fact that one unit's contribution to the overall organization's performance may be heavily dependent on the performance of other units.

A second major challenge arises when a unit's contribution to overall organizational outcomes is indirect or tangential, as in the example of the training unit. In principle, the ultimate value of the training unit is tied to its contribution to the organization's overall performance and effectiveness. If there was no connection, the training unit should not exist. However, most of what the training unit does is provide knowledge and skills to employees. Whether the employees actually use that knowledge and skill, whether they use it correctly, whether they are allowed to use it by co-workers and supervisors, whether they have the equipment they need, and many other factors are far beyond the control of the training unit. Equally important, the employees' performance might improve due to training, but the impact of this improvement on overall organizational performance may not be detectable amidst all the other dynamic conditions, both inside and outside the organization, which are constantly in flux. For organizational units, such as training, that are somewhat distant from core operating functions, assessment based on contribution to the organization's overall bottom line may be impractical or impossible.

In addition to all these complications, it should be noted that there is a tendency toward suboptimization in organizations. This refers to the tendency of organizational subunits to emphasize accomplishment of their units' objectives (outputs or outcomes) even when those conflict with overall organizational goals and effectiveness. If the manufacturing unit is judged on the number of widgets produced, it may maximize that production, regardless of whether they are selling. Or, if it is judged on the percentage of widgets produced that meet a high standard of quality, it may slow production in order to take more time in making each widget, regardless of whether the sales unit then loses customers because the rate of delivery is too slow. This happens, of course, in part because members of a unit come to identify with it, often more closely than they identify with the entire organization. Also, if a unit is judged on a specific criterion of effort, output, or outcome, its members and its manager will often focus their entire energy on meeting that criterion, to the exclusion of other criteria that may be important, and without regard to how it may affect other units or the organization as a whole. This is a reflection of the common refrain, "what get's measured gets done." Sometimes, what does not get measured, does not get done.

Measuring Police Performance

All of the general considerations noted above apply to police organizations. Unit performance is sometimes measured based on effort, such as vehicles stopped, miles driven, average response time to reported incidents, properties checked, number of vehicle crashes reconstructed, and number of K-9 unit responses. It is particularly common to measure police unit performance in terms of outputs – number of reports taken, quantity of drugs seized, number of citations issued, number of arrests, proportion of cases cleared, etc. It is substantially less common to assess police organizational units based on outcomes. This is not surprising since the same is generally true regarding overall police organizational performance – police agencies have traditionally resisted being held accountable for the level of crime in their jurisdiction, and few if any other outcome measures were even considered. Until fairly recently, overall police organizational performance has generally been measured using response time, arrests, citations, and case clearances – measures that mainly address effort and outputs, not outcomes.

A few key characteristics of the police mission make performance measurement particularly complex. One is that a police organization's "bottom line" is multidimensional. In the simple example given above, the widget company's bottom line was *uni*dimensional – profit. The company could easily and accurately tell how well it was doing, and, within the constraints discussed above, determine how well or poorly the manufacturing unit was contributing to the bottom line. Compare that to a police department that has a hypothetical three-part bottom line: reduce crime, maintain order, and protect constitutional rights. Right away, it is more difficult for the police department to know how well it is doing, because it must monitor three separate indicators, each of which may be quite hard to measure. In addition, working toward accomplishing any one of the three criteria may lead to some neglect of the others, since time and resources are always limited. Moreover, pushing hard on one may actually conflict with accomplishment of another. What is created is a tricky balancing act built on a foundation of uncertainty, since data on each of these conditions are liable to be incomplete at best.

Another important factor is that "how the police behave" (process) is of equal importance to "what the police accomplish" (outcome).

For the most part, this is not true of the fictional widget company and its manufacturing unit. The company CEO could say to the head of the manufacturing unit, "I want you to produce 1,000 widgets a day and I do not care how you do it." Certainly, there would be cost considerations, but the method by which the unit produced the widgets could be left entirely to the discretion of the manufacturing unit head. The same cannot be said in police organizations (Sunshine and Tyler 2003; Cordner 1978). A police chief might say to a police tactical unit commander, "I want you to reduce crime by 10 %," but the unspoken remainder of the message would be that it would have to be done within the budget, within the law, without interfering with the accomplishment of other important outcomes, and without upsetting the community. And it would have to look good in the next day's newspaper (or that night's blogs). The point is not simply that police organizations operate under constraints, but also that police organizations are judged as much on how they behave (their actions, efforts, and outputs) as they are on outcomes (what they accomplish). To put it another way, even though, in principle, we should want to assess organizations (and their units) as much as possible in relation to outcomes, in the case of police organizations (and their units), there are valid reasons to measure what they do in addition to what they accomplish.

Because of the nature of the police mission, it may also be important to measure performance outcomes that are not immediately obvious. To give an exaggerated example, a desired outcome might be reduction of traffic accidents and fatalities. A police organization might achieve this outcome by reducing all speed limits to 5 miles per hour and writing citations to every driver who went any faster. Merely on the basis of the specified outcome, it would seem that the organization's performance could be judged exemplary. However, there would probably be other outcomes of this brand of police performance, including greatly inhibited traffic flow, commercial trucking losses, irate drivers, and reduced police performance in relation to other outcomes, since officers would be totally consumed with traffic enforcement. The point is that measurement of police performance may sometimes need to take account of the side effects, negative effects, and unintended consequences of performance. Otherwise, false conclusions might be drawn about the effectiveness of the performance.

A useful framework for police performance measurement that incorporates a multidimensional bottom line and that also includes process and outcome dimensions was developed by Moore and Braga (2003). The framework identifies seven categories of performance that can be used to assess a police organization's overall effectiveness, as noted below. A nice feature is that the seven categories also cover some of the most likely side effects and unintended consequences of police activity, such as abuse of authority, decreased legitimacy in the eyes of the public, and excessive cost. The first four categories reflect mainly outcomes, the fifth combines outcome and output, and the last two refer to processes. These same categories can be used in assessing the performance of individual police units, particularly to determine whether they contribute to overall organizational effectiveness, recognizing that all seven dimensions may not be directly applicable to every separate unit within the organization.

- Reduce crime and victimization
- Call offenders to account
- Reduce fear and enhance personal security
- Ensure civility in public spaces (ordered liberty)
- Quality services/customer satisfaction
- Use force and authority fairly, efficiently, and effectively
- Use financial resources fairly, efficiently, and effectively

Police Unit Performance

Based on the preceding discussions, it makes the most sense to measure police unit performance in two ways: (1) measures of effort, output, and outcome directly tied to the unit's specific responsibilities, and (2) measures of the unit's performance in relation to the seven performance categories noted just above. The first set of measures is likely to be the most specific and tangible, while the second set is ultimately the most important. In other words, the number of citations issued might be a clear indication of how hard the traffic unit is (or is not) working, but what ultimately matters the most is the number of traffic crashes and fatalities (as long as the means of reducing crashes is not too draconian, as per the earlier example), an element of “ensure civility in public spaces.”

Measurement of police performance in three different types of police units is discussed below: a police district or precinct, a specialized operational unit, and a support unit. These three types of units represent the vast majority of units likely to be found in police organizations.

Police District or Precinct

Large police organizations are usually divided into districts or precincts, which are geographically based units responsible for providing a variety of police services. Districts or precincts almost always provide patrol services – preventive patrol, response to calls, community policing, and problem solving. Whether they are also responsible for other operational functions such as investigations and traffic enforcement varies. Some police organizations decentralize one or more specialized functions to the district/precinct level, while others centralize everything other than patrol at the headquarters level.

Compared to other police units, the responsibilities of a district or precinct are broad, and therefore, a variety of measures are likely to be needed to measure performance. In fact, all seven of Moore and Braga’s dimensions of police performance might be applicable. In order to assess the performance of a police district/precinct, a top police executive might want data for the following measures:

A positive feature of using this measurement approach is that it employs a breadth of measures to match the breadth of functions performed by the district/precinct unit. Also, the measures are mainly associated with outcomes, which arguably is both preferable and reasonable when focusing on a core operational unit – the district/precinct has the opportunity to reduce crime, solve crimes, make the public feel safer, and so forth, in contrast to support units that can only hope to accomplish these kinds of outcomes indirectly. Of course, the two specific measures listed for each performance dimension can certainly be debated and better measures might be identified.

One disadvantage of this measurement scheme is complexity. Fourteen measures are quite a few, and it could be argued that the tendency would be to divide the district/precinct’s attention across too many indicators, resulting in scattered rather than focused performance. This is similar to the observation that too many objectives are almost worse than none. In response, though, it must be argued that multiple indicators are necessary when functions are so broad. It then becomes the manager’s responsibility to set priorities and identify areas of performance in need of greater emphasis, based on the performance data.

This multi-measure approach contrasts sharply with Compstat, which has become the most popular method for measuring the performance of police districts/precincts, and for using performance measurement to hold commanders accountable (Henry 2002). Compstat systems typically rely on just one measure, reported crime, perhaps supplemented with information about arrests and crimes solved. Pressure is put on district/precinct commanders to reduce reported crime and to show arrests in response to any crime increases that occur. While this system has the virtue of focusing everyone’s attention on the one measure that probably does deserve the highest priority, it typically fails miserably in taking into account the full range of the police mission in a geographic area, and thus, its one measure grossly misrepresents the full performance picture. Also, the extreme emphasis on one measure creates temptations on the part of unit managers to manipulate the data in order to look better on that measure (Eterno and Silverman 2012).

Another feature of the seven-dimension, fourteen-measure framework presented above is that it does not incorporate much information on effort or outputs. It does not measure miles driven, number of arrests, or response time (or at least it does not present them as measures of unit performance). A three-part explanation can be offered. First, these measures of effort and output may simply not matter in and of themselves, but only in relation to the accomplishment of larger outcomes. A police chief probably does not care whether the public in a district is reassured by the sight of many police cars driving about, or by the sight of police officers walking about, or by the sight of police officers in the mall when they go shopping, or by their knowledge that their local officers are systematically working to solve neighborhood problems – the police chief only wants to know the degree to which the district's public is reassured.

Second, by not emphasizing some of the most traditional measures of effort and output, the framework avoids sending a mixed message about whether it is necessary to perform those kinds of activities at some specified level, in order to avoid getting in trouble. The intention is to encourage the district/precinct unit and its commanders to emphasize the achievement of important outcomes, and to give them reasonable flexibility in determining how best to do it. The measures do not discourage the use of motorized patrol, quick response, or making arrests; they just do not send a message to the district/precinct that some level of those activities is required. It will inevitably be recognized by district/precinct commanders that some level of motorized patrol is probably necessary for public reassurance, that slow response to emergencies will result in low customer satisfaction, and that too few arrests might allow crime and disorder to increase. Thus, including these kinds of measures in a district/precinct performance measurement framework is probably not necessary – they will happen regardless.

Third, the truth is that someone, somewhere in the police organization will be measuring miles driven, response times, and arrests anyway. If customer satisfaction in a district/precinct declines, the unit commander as well as his or her bosses at headquarters will want to know why, and response time data may (or may not) provide some of the explanation. Lots of effort and output data are routinely generated and stored within police organizations, and can be used to diagnose and analyze unit-level performance-related problems. The framework above simply suggests avoiding most of those types of measures for the ongoing review of unit performance, especially core operational units, such as a district or precinct, whose performance is directly tied to the accomplishment of most of the overall organization's primary missions.

Repeat Offender Unit

One type of specialized police operational unit is a repeat offender unit (Martin and Sherman 1986; Abrahamse et al. 1991). Typically, this type of unit targets repeat offenders, also called prolific offenders. The underlying rationale is that some offenders commit a disproportionate number of offenses. The earlier in such offenders' careers that police can identify, arrest, and successfully prosecute them, the quicker they come under correctional supervision (including incarceration), and therefore the fewer offenses they will be able to commit. Similarly, if offenders are released from jail or prison and resume a prolific career, the sooner police rearrest them, the fewer offenses they will be able to commit.

Police repeat offender units usually target robbers, burglars, and thieves, as these are high-volume crimes, although other types of offenders may also be targeted, and it is not uncommon for targeted offenders to commit a variety of types of crimes, to be drug offenders as well, and/or to be gang members. It is not unusual for such a unit to conduct surveillance on a suspected active robber and catch him in the act of a theft or drug purchase. Methods for targeting often include plainclothes operations, surveillance, sting operations, and the use of informants.

One other method used by repeat offender units is to conduct enhanced investigations following the arrest of offenders who qualify under "three strikes" laws as career criminals or prolific offenders (criteria vary from jurisdiction to jurisdiction), whether these offenders are arrested by the unit or by other officers in the police department. These enhanced investigations are intended to build stronger cases on the instant offense, as well as assemble the necessary documentation to prove to the court that the individual meets the eligibility criteria for an additional sentence under the three strikes law.

As described, the responsibilities of a repeat offender unit are fairly narrow, but they are directly tied to some of the overall police organization's important dimensions of performance. A framework like the following one might provide a well-rounded picture of the unit's performance.

This suggested framework includes ten measures, nine of which are within three of the seven dimensions of police performance. It does not seem reasonable to utilize measures related to reducing fear, ensuring civility, and satisfying the public when assessing the performance of a repeat offender unit, simply because the unit's functions are not tied closely enough to these outcomes. It *is* anticipated that the unit's activities might contribute to crime reduction, but care needs to be taken since such a unit is usually comprised of just a few officers/detectives – expecting the impact of their work to be detectable on the overall crime rate in the jurisdiction is somewhat unrealistic. The suggested measure focuses on the crimes not committed by the offenders who are arrested by the unit. Of course, the measure is nothing more than an estimate. But if a robber who has been committing ten robberies a year is incarcerated for 3 years, it is not too unreasonable to give the unit credit for preventing thirty robberies from occurring.

The other nine measures try to get to the heart of the repeat offender function. Reflecting the basic issues of productivity and effectiveness, they focus on whether the unit is successful in arresting, prosecuting, and convicting the repeat offenders that are targeted. Recognizing one important negative consequence of this type of strategy, they focus on whether the unit is able to operate within the law. Similarly, they focus on whether the unit is able to maintain its focus on repeat offenders, rather than broadening its criteria and making arrests of less significant offenders – experience has shown that this kind of "mission creep" often occurs, especially if the unit feels pressure to produce larger quantities of arrests in order to justify its existence. Lastly, the measures incorporate some cost criteria. These are particularly important because, again, experience has shown that repeat offender units tend to be costly on a per-arrest basis, because their tactics are personnel-intensive. However, on a per-conviction basis, or in relation to the seriousness of the offenders who are arrested, the cost of the unit's performance may be rated more favorably.

A question might be raised as to why arrests figure so prominently in the repeat offender unit's measures, when they were not suggested among the district/precinct measures. The response would be that the repeat offender unit's narrower function is directly focused on identifying and apprehending particular offenders, whereas the district/precinct's function is much broader – so much so that the number of arrests made by the district's or precinct's officers might or might not contribute positively to the achievement of its desired outcomes.

Communications Center

One type of support unit within a police organization is the communications center. This unit normally answers telephone calls to the police agency from the public, including emergency calls, and also uses the police radio system to dispatch police officers to incidents and to respond to inquiries from police officers in the field. The communications center may also be the hub for inquiries to databases on wanted persons, vehicle registrations, driver licenses, and so forth, although in many police agencies today, officers can initiate these kinds of inquiries themselves via hand-held computers or computers in their patrol cars, without requiring any assistance from dispatchers.

A police communications center does interact directly with the public, mainly via the telephone, but otherwise its performance is entirely in support of other units of the organization. Because the communication center's performance is only indirectly tied to the accomplishment of a majority of the overall organization's desired outcomes, it is not reasonable to use most of those seven categories when measuring the performance of the unit. A framework like the following, emphasizing two of the seven categories, might provide an adequate indication of the unit's performance.

This framework includes six measures, four associated with providing quality services that satisfy customers, and two related to efficient use of resources. The main emphasis is on measuring the quality of services provided to the communications center's primary clientele – members of the public who call to request assistance or seek information, and police officers who interact with dispatchers over the police radio system. The quality control assessment mentioned in the measures reflects a common practice in the communications field. Telephone conversations are routinely recorded, and a sample is later reviewed, either by supervisors or quality assurance specialists, using a standard rubric.

Comparisons, Benchmarks, and Standards

This entry has focused mainly on measuring police unit performance, but a brief discussion of analysis and interpretation of the measures is in order. A few of the measures identified above are meaningful on their face, such as the percent of the unit budget expended, which is expected to be 100 % or less. If it exceeds 100 %, the unit manager would likely be required to explain why the budget was exceeded. Any subsequent action would probably depend on the explanation.

Most of the measures of police unit performance become meaningful through comparison. As an example, 75 % of the recipients of police service in a police district/precinct might report that they were satisfied with the services they received, or 75 % might report that the officer treated them with respect (the specific survey items used to measure customer satisfaction are likely to vary from one organization to another). The question then arises, is 75 % a good score? Does it indicate that police performance in the district/precinct is excellent, satisfactory, or in need of improvement? Without any additional context, we would likely conclude that performance is fairly good, but not fantastic. But interpretation could become much more meaningful by making two types of comparisons: (1) comparison to all the other districts/precincts in the same police organization, and (2) comparison to previous reporting periods (years, in most cases). Such comparisons would reveal whether the public's satisfaction with police services received in the district/precinct is higher, about the same, or lower than in other districts, and also whether it is improving, staying the same, or declining over time. These kinds of simple comparisons typically add a great deal of meaning to most of the unit performance measures described above.

A third type of comparison is external. Comparing performance in one police department to others is usually done at the organizational level, such as comparing crime rates and clearance rates. Particularly when scores on performance measures are published, they can serve as benchmarks to which a police organization can easily compare itself. This becomes somewhat more challenging at the police unit level, in part because such data are rarely published, and in part because similarly named units in two different police departments may not actually perform the same duties or have comparable resources devoted to the functions. Nevertheless, a police communications center, for example, could identify other police agencies that also measure the proportion of incoming emergency calls answered within three rings, or even participate in a consortium of agencies that agree to collect data on the same unit performance measures. Establishing a system that enables such external comparisons of performance would be very valuable for the police communications center manager, allowing she or he to gauge much more accurately how well the unit is performing, and which aspects of performance might be most in need of improvement. Naturally, this kind of performance information would also be extremely useful for higher-level executives who, otherwise, would have a difficult time determining whether their department's communications center is performing as well as it should.

An additional form of comparison is sometimes available when recognized standards of performance have been established. Within a police organization, for example, a standard of three minutes or less might be established for average response time to high-priority calls. This might be an organization-wide standard monitored monthly or annually, but it could also be applied to districts/precincts. When a performance measure includes such a standard, a message is sent to unit managers that they should allocate their resources and direct their personnel in such a way that the standard is met. If a unit does not meet such a standard, its manager knows that a convincing explanation will be expected, and in fact, the manager should be able to point to efforts made to meet the standard, and should have notified his or her superiors in advance that the unit was at risk of not meeting the standard.

External standards of police performance are potentially the most powerful, but they are rarely available. A standard-setting body does exist, the Commission on Accreditation for Law Enforcement Agencies (CALEA). However, the commission's standards primarily address administrative processes rather than true performance standards. For example, a CALEA standard requires periodic audits and inventories of evidence and property under the control of the police agency. The standard has some specific requirements, such as a minimum of an annual audit and a required audit whenever a new person is assigned to the position of property custodian. But the standard does not specify that the audit must be able to account for 100 %, or 99 %, or 95 % of the evidence and property that is supposed to be in storage. Similarly, a standard specifies that "if the agency participates in a tactical team, the agency requires that all personnel assigned to the team engage in training and readiness exercises." CALEA standards do not specify how much training, though, nor do they set any kind of standard for the tactical team's actual performance.

Despite the general lack of external performance standards for police units, it is practical and feasible for a police organization to establish a robust system for measuring the performance of its subunits, as described in this entry. Certainly there are challenges to both measurement and interpretation, but establishment of such a system is essential for effective police administration, accountability, reform, and ultimately, for the delivery of quality services and adequate protection to the public.

Related Entries

- ▶ Evidence-Based Policing
- ▶ Measuring Police Performance
- ▶ Methodological Issues in Evaluating Police Performance
- ▶ Police Performance Measurement

Recommended Reading and References

Abrahamse AF, Ebener PA, Greenwood PW, Kosin TE (1991) An experimental evaluation of the Phoenix Repeat Offender Program. Justice Quart 8:141–168

CALEA (2012) Standards for Law Enforcement Agencies: The Standards Manual of the CALEA Law Enforcement Accreditation Program, Fifth Edition as revised. Commission on Accreditation for Law Enforcement Agencies, Gainesville, VA

Cordner G (1978) Open and closed models of police organizations: traditions, dilemmas, and practical considerations. J Police Sci Administr 6(1):22–34

Davis RC (2012) Selected international best practices in police performance measurement. RAND Corporation, Santa Monica, CA

Eterno JA, Silverman EB (2012) The crime numbers game: management by manipulation. CRC Press, Boca Raton, FL

Fleming J, Scott A (2008) Performance measurement in Australian police organizations. Policing: J Policy Practice 2:322–330

Henry VE (2002) The Compstat paradigm: management accountability in policing, business and the public sector. Looseleaf Law Publications, Flushing, NY

Hoover LT (ed) (1998) Police program evaluation. Police Executive Research Forum, Washington, DC

Langworthy RH (1999) Measuring what matters: proceedings from the Policing Research Institute meetings. National Institute of Justice, Washington, DC

Lind RC, Lipsky JP (1971) The measurement of police output: conceptual issues and alternative approaches. Law Contemporary Prob 36:566–588

Martin SE, Sherman LW (1986) Selective apprehension: a police strategy for repeat offenders. Criminology 24:155–175

Moore MH, Braga A (2003) The bottom line of policing: what citizens should value (and measure) in police performance. Police Executive Research Forum, Washington, DC

Parks RB (1971) Measurement of performance in the public sector: a case study of the Indianapolis Police Department. Indiana University, Bloomington, IN

Reaves BA (2011) Census of state and local law enforcement agencies, 2008. Bureau of Justice Statistics, Washington, DC

Shane JM (2007) What every chief executive should know: using data to measure police performance. Looseleaf Law Publications, Flushing, NY

Stone C, Travis J (2011) Toward a new professionalism in policing. New perspectives in policing. National Institute of Justice, Washington, DC

Sunshine J, Tyler TR (2003) The role of procedural justice and legitimacy in shaping public support for policing. Law Soc Rev 37:513–548

Weisburd D, Neyroud P (2011) Police science: toward a new paradigm. New perspectives in policing. National Institute of Justice, Washington, DC

Measuring Wrongful Convictions

Marvin Zalman
Department of Criminal Justice, Wayne State University, Detroit, MI, USA

Overview

Measuring wrongful convictions includes counting exonerations, estimating the incidence of wrongful convictions, and measuring the correlates of wrongful convictions in known exoneration cases. Measuring wrongful convictions is an important task because it can be used to evaluate the inaccuracy and unfairness of the criminal justice system.

A wrongful conviction is a conviction that is factually inaccurate, that is, the wrong person was convicted or a person was convicted for a crime that did not occur. An actual innocence exoneration is an official act that erases the legal status of guilt based on a judgment of factual innocence.

Wrongful convictions are inherently difficult to establish because they are "invisible" when they occur, the process of reconstructing equivocal evidence that led to a conviction in the first place is difficult, and prosecutors and courts are resistant to opening old convictions.

Lists of exonerations are not kept by government agencies but by private organizations like the Innocence Project and the National Registry of Exonerations.

Quantitative studies of wrongful death penalty convictions estimate that about three in 100 defendants sentenced to die are factually innocent. Qualitative estimates of wrongful convictions among all felonies, based on descriptions of the quality of criminal justice and other factors, plausibly conclude that at least 1 % of all felony convictions are factually inaccurate.

Studies of exoneration cases show that such factors as eyewitness misidentification, police tunnel vision, false confessions, informants' deception, erroneous forensic evidence, and misconduct and incompetence by prosecutors and defense attorneys occur in wrongful conviction cases in significant quantities.

Introduction

This entry discusses the measurement of wrongful convictions occurring in the United States. Two types of measurement are discussed: (1) the number and rates of factually wrongful convictions and (2) measures of the factors (correlates) discovered in such cases. Measuring the number of wrongful convictions is divided into counting wrongful convictions and estimating their rates.

Inaccurate convictions are indicative of an improperly functioning criminal justice system. Convicting factually innocent people means that the process is not effective or efficient.

More important, convicting the innocent is an injustice that every legal system wishes to avoid. Knowing something about such cases, including their overall numbers and correlates, is necessary to decide whether the problem requires a policy response and, if so, how to respond.

It should be noted that "false acquittals," when juries acquit defendants who in fact committed criminal acts with the requisite criminal intent, are also injustices. As both kinds of errors of justice may stem from operational flaws and informational deficiencies in the criminal justice process, research on measuring and correcting wrongful convictions should help reduce wrongful acquittals. Although a large number of factually guilty people are never convicted of crimes, such "errors of impunity" mostly result from crimes never being discovered or reported in the first place and other deficiencies of police agencies (Forst 2004). Such injustice is not usually related to the kinds of justice system processing errors that lead to wrongful convictions or acquittals.

The terms prevalence and incidence, used in medicine and epidemiology, which are sometimes confused, have been used to describe rates of wrongful convictions. Prevalence is an epidemiological measure of the existence of a disease or condition in a population and is calculated by dividing the number of persons with a disease or condition at a particular time by the number of individuals examined. Incidence measures the rate of occurrence of new cases and is calculated as the number of new cases of a disease or condition in a specified time period (usually a year) divided by the size of the population under consideration who are initially disease free. Both measures can be expressed as rates (Crichton 2000). It makes some sense to measure the prevalence of wrongfully convicted *prisoners*, but the more useful measure of wrongful conviction rates for the purpose of assessing the criminal justice process is the incidence that occurs in a year.

Defining and Measuring Wrongful Convictions

To measure a social category, it must be defined. Wrongful conviction can mean different things. A conviction can be wrong and reversed on appeal because it was obtained on the basis of a serious violation of the defendant's procedural or constitutional rights. Such a conviction is wrongful in the procedural or legal sense. A conviction can also be wrong because a jury was mistaken about a defendant's culpability, as, for example, regarding a self-defense claim. In this entry, a wrongful conviction means actual or factual innocence. This occurs when a crime was committed and the person convicted had nothing to do with it (i.e., the "wrong person" scenario, where a jury erroneously discounted an alibi), or no crime was committed, but a court imputed criminal liability (Risinger 2007). The latter situation has occurred in cases involving fire-related deaths and the deaths of infants, where guilt was wrongfully based, respectively, on erroneous technical or medical evidence about arson or shaken baby syndrome. The term false conviction is a synonym for convictions that are wrongful in the factual sense. Scholars agree that false convictions should be included in measurements of factual innocence. There is some uncertainty about whether cases that are wrongful in the culpability sense should be included in the measurement of wrongful convictions.

A measurable entity "must be identifiable by some means and identified by us. And we must know something about its properties and behavior, so that we can relate these to our measures of it" (Gorard 2010:390). This insight raises the question of what exactly is meant by a "conviction," as a wrongful conviction presupposes a conviction. A conviction is an official decision that is formally entered in the record of a politically legitimate entity following a process of investigation, prosecution, and adjudication. The process leading to a conviction involves many actors and subprocesses and can be brief and routinized or very complex.

A conviction, therefore, is a socially constructed artifact. As a result, a wrongful conviction simply does not "exist" but is itself a socially constructed artifact that comes at the end of a process of deconstructing a prior conviction. The process of deconstruction can be done by a private entity (e.g., journalist, legal scholar, social scientist, or research group), resulting in an unofficial judgment of a false conviction. The process can also be accomplished by an official body, resulting in an official exoneration. Understanding the process by which a conviction (and therefore a wrongful conviction) is produced allows researchers to not only count them but also to measure the factors present that may logically have led to the erroneous verdict.

The definition of a false conviction raises several difficulties. The inherent limitations of proof in the absence of incontrovertible evidence have led to the standard of conviction "beyond a reasonable doubt." This means that a conviction might be false and that, concomitantly, an acquittal might also be false. Lingering uncertainty about the relationship between legal and factual guilt or innocence provides an initial basis for exploring whether some convictions are false, especially when non-DNA exonerations are compared to DNA exonerations. To avoid this difficulty, some enumerations of wrongful convictions or exonerations either apply or claim to apply "objective" standards. However, in some cases (e.g., an acquittal in a retrial of a case after a successful appeal), the "objective" *legal* acquittal can raise the same lingering uncertainty as an initial acquittal regarding factual innocence. This suggests that some level of judgment about case facts (similar to the *judgment* leading to a conviction, but not similar to the *process* of jury deliberation) is necessary to generate a conclusion that a conviction was false and that the exoneree is factually innocent.

Another issue is that some person or institution must define whether a conviction was wrongful. The standard applied by the particular writer or research group, or by a governmental body, must be examined to determine how it defines a wrongful conviction; the goals and organizational imperatives of the decision-making person or body should also be considered. This will be considered below in the section on "Counting Wrongful Convictions."

A clear distinction must be drawn between exonerations and wrongful convictions. In common usage, the terms are sometimes employed interchangeably. An exoneration is an official act by a judge or governmental officer that erases the legal status of guilt previously imposed on a person. Researchers listed four kinds of exonerations: pardons issued by governors or other executive officers based on evidence of innocence; cases dismissed by courts after new evidence of innocence, like DNA, is presented; acquittals in retrials, but only if the basis is evidence that the defendant had no role in the crime; and a few cases where states posthumously acknowledged the innocence of defendants who had died in prison (Gross et al. 2005). This definition of exonerations includes both objective and subjective or judgmental elements. Exonerations may also be issued by official bodies with jurisdiction to determine the amounts of compensation for wrongful convictions or to declare a status of innocence (discussed below).

Official exonerations are a subset of a larger number of cases identified as wrongful convictions by authors, research scholars, and private organizations, often as a result of exhaustive research. Unless an official body declares that a conviction was wrongful, and as a matter of law erases the conviction, the wrongful conviction cannot properly or legally be called an exoneration. This is so even if the judgment that a false conviction occurred is based on ample and powerful supporting evidence. Indeed, prosecutors in some cases in which a prisoner won an appeal and is facing retrial will offer to dismiss the case if the prisoner, facing the risk of reconviction, pleads guilty to a lesser crime and is released for time served. This allows prosecutors to claim that the prisoner was actually guilty in the face of contrary evidence, thereby saving face and avoiding a lawsuit against police and prosecutors. Such action depresses and distorts the number of justifiable exonerations.

A more controversial point, addressed in the section on "Estimating Wrongful Convictions," is that known exonerations and wrongful convictions are a minor subset of all wrongful convictions. This is plausible, considering that social measurements, including the United States census and official crime data, tend to undercount the phenomenon being measured (Loftin and McDowall 2010).

The Inherent Difficulty of Measuring Wrongful Conviction

Whether data about wrongful convictions are collected by researchers or potentially by government agencies, the task is inherently difficult. In most jurisdictions, there is no routine legal mechanism to review prior convictions for the possibility that they were factually erroneous, although this is beginning to change (discussed below). Moreover, a wrongful conviction is invisible when it occurs. Miscarriages of justice are typically uncovered years after they occur and only after arduous investigations on behalf of prisoners by friends and relatives or by innocence projects. They are often discovered by chance, as when a forensic analyst happens to save biological evidence samples (when to do so is not standard procedure), which later provide exonerating DNA profiles. As a result, although lists of wrongful convictions have been collected by several organizations, it is not possible to relate them to the context of similar cases that occurred at the same time. This makes it close to impossible to directly measure the incidence of wrongful convictions (Gross and O'Brien 2008).

Another difficulty in establishing that a conviction was false is that many cases have equivocal evidence, and determining whether the conviction was accurate in the first place requires an element of judgment. As a result, convictions that are deemed false by many will be disputed by others. Additionally, the limited resources of innocence projects and their high standards in accepting cases, other attendant costs of seeking exonerations, and the general reluctance or opposition of government agencies to open closed cases inevitably mean that many wrongful convictions will never be uncovered.

Innocence scholars and critics generally agree that it is impossible to precisely measure the incidence of false convictions (Schehr 2005). This understanding has led some to opine that the number of such miscarriages of justice is minuscule (Marquis 2006; Hoffman 2007). No writer now suggests that wrongful convictions never occur, acknowledging at least the accuracy of the Innocence Project's list of DNA exonerations (Innocence Project n.d.). Analysts who closely study wrongful convictions, however, assert with confidence that the small number of known DNA and non-DNA exonerations are indicative of much larger numbers of false convictions (Gross et al. 2005; Garrett 2011).

Government Institutions That Establish Actual Innocence

The traditional methods of establishing official exonerations – pardon, judicial dismissal, and acquittal after a reversal – are ambiguous. These methods may exonerate former and current prisoners for legal or humanitarian reasons, rather than on account of factual innocence. Even if raw counts of such decisions were kept, the records that underlie the decisions would have to be inspected to determine if the ground was actual innocence. In addition to these modalities, courts and agencies given the authority to grant state compensation for wrongful convictions have the potential to keep and publish records of false convictions. The federal government has had a compensation law since 1938, and awards are made by the United States Court of Federal Claims. In recent years, the number of states that authorize compensation for wrongful conviction has grown to twenty-seven (Norris 2011).

Nevertheless, some recent compensation statutes set very high standards of proof and include numerous disqualifications that may deny compensation and, by implication, a finding of innocence to claimants with justifiable wrongful conviction cases (Norris 2011). For example, James Richardson had been exonerated by Florida for murder in 1989 in a special proceeding in which Janet Reno, later Attorney General of the United States, was appointed a special prosecutor. She moved to dismiss his prosecution on a nolle prosse in a memorandum which concluded that he was "probably wrongfully accused." When Florida enacted the Victims of Wrongful Incarceration Compensation Act in 2009, Richardson petitioned for compensation, but was denied because the nolle prosse did not establish by clear and convincing evidence that he was actually innocent. The anomalous situation of one state institution exonerating a person and another institution of the same state not exonerating the person for the same crime can play havoc with clarity in determining the existence of exonerations and wrongful convictions (Zalman 2012).

Several states have, by statute, authorized courts or other agents to declare convicted persons actually innocent. In 2001, Virginia authorized its Supreme Court to issue writs of actual innocence based on biological evidence and in 2004 authorized the issuance of such writs on the basis of nonbiological evidence. North Carolina established an Innocence Inquiry Commission in 2006 to investigate petitions by prisoners and where it finds sufficient evidence of actual innocence to refer such cases to a superior court for a three-judge trial to determine actual innocence. In 2008, Illinois authorized circuit courts to issue certificates of innocence to petitioners who can prove their innocence on a preponderance of the evidence. In that year, Utah also established a procedure allowing district courts to find a person factually innocent. Several states, including Illinois and Texas, now have provisions for governors or pardon boards to issue pardons based on innocence.

Courts typically do not keep accessible records that are useful to the public or the research community, and records of findings of actual innocence will be no different. Innocence issues, however, are now visible to such organizations as the National Registry of Exonerations (n.d.) and Justice Denied (n.d.), making it more likely that exonerations announced under new state compensation laws and declaration-of-innocence procedures by court and pardon boards will be tracked and listed.

Counting Wrongful Convictions

No government agency or court maintains an accessible list of exonerations, although several have the authority to issue exonerations. Most information on wrongful convictions, including counts, has been produced by journalists, writers, researchers, and organizations.

Big Picture Books. Beginning with Borchard's 1932 book, *Convicting the Innocent*, a number of "big picture books" (Leo 2005) detailed cases of exonerations, describing the factors that appeared to be sources of the miscarriages of justice. Borchard listed 65 cases, most of which had occurred within a decade or two of the book's publication. Although he was a law professor, Borchard wrote a journalistic book designed to arouse popular concern. His goal was to firmly establish that wrongful convictions can and do occur. He used the book to successfully lobby for the passage of a federal compensation law for exonerees. His list of 65 exonerations (most were American, a few were English) was the first such compilation. The number grew as other similar books have been published since the 1950s (Frank and Frank 1957; Yant 1991). In addition, scores of books, typically written for a popular audience, have been written about single wrongful conviction cases. They provide a wealth of detailed information about wrongful convictions but typically do not relate the information to a growing body of wrongful conviction scholarship (Leo 2005).

Scheck et al. (2000), a popular and influential book describing DNA exonerations obtained by the Innocence Project in the 1990s, focused on factors related to wrongful convictions. An appendix provides data on 74 false convictions, including jurisdiction, twelve causal factors, length of maximum sentence, time served before exoneration, the number of mistaken witnesses per case, and the race of the crime victims and the exonerated defendants.

Scholarly Lists and Evaluations. Rattner (1988) analyzed 205 "old and new legal cases" of wrongful conviction occurring in the United States after 1900 "derived from books, documents, and newspaper clippings." Most of the cases had been previously described in big picture books, supplemented with more recent news reports. Most resulted in legal exonerations (three were paroled, and data was missing for 15). In 40.5 % of the cases, the real culprit was apprehended. The violent crimes of murder, robbery, and rape accounted for 87 % of the convictions. Rattner described the distribution of sentences (45.8 % of the known sentences were death or life in prison) and the types of error associated with the wrongful convictions (52 % of the cases with known or listed errors included eyewitness identification error).

Gross (1987) assembled 136 cases of misidentification that led to criminal prosecution in the twentieth century. The cases were divided between those occurring before and after the Supreme Court's lineup decision (U.S. v. Wade 1967). The pre-Wade cases were accounts of convictions derived from the big picture books, while the post-Wade cases were derived from news accounts. Thirty-nine cases, mostly from the post-Wade set, did not result in convictions. Gross used the cases to describe the processes of misidentification and exoneration and to describe historical trends in case processing.

Bedau and Radelet (1987) enumerated 350 capital or potentially capital murder or rape convictions imposed in the twentieth century believed to have been factually erroneous. The list was based on evaluations of the case facts in which the scholars exercised their judgment to conclude that either no crime occurred or the defendant was legally and physically uninvolved in the crime. The article listed the jurisdiction, date, crime of conviction and type of sentence, the primary evidence for the judgment that the conviction was false, and the factors associated with the erroneous convictions.

Gross et al. (2005) scoured many sources to identify 340 official exonerations occurring in the DNA era – for the years 1989–2003. More than half, 57.6 %, were non-DNA exonerations. The vast majority were convicted by trial rather than by plea, and all but 14 cases were convictions for murder or rape. The authors suggested that the fortuitous nature of exoneration must have excluded many robbery cases based on faulty eyewitness identification. They excluded two and possibly three "mass exoneration" cases in which corrupt police planted evidence on innocent people, followed, after discovery, by the exoneration of masses of such people. They also excluded from the exoneration count pending cases, pleas of guilty or no contest where evidence of innocence was strong, inexplicable failures to exonerate people who were patently innocent, and people falsely convicted in the child care sex abuse and satanic ritual cases that swept through America in the 1980s and 1990s.

Organizational Lists. The Death Penalty Information Center (n.d.) (DPIC) maintains an on-line list of exonerated death row inmates since the reinstatement of capital punishment in 1976. The number was 142 as of March 2013. The list includes name, state of conviction, race, year convicted and exonerated, number of years served, the reason for exoneration, and whether it was a DNA exoneration. The DPIC applies an objective standard of legal innocence, which means that a few of the exonerees may have been factually guilty, although wrongful conviction scholarship has established the actual innocence of many.

The Innocence Project (n.d.), affiliated with the Benjamin Cardozo Law School of Yeshiva University, is a self-described "national litigation and public policy organization dedicated to exonerating wrongfully convicted individuals through DNA testing and reforming the criminal justice system to prevent future injustice." It lists the number of DNA exonerations that have been obtained in the United States since 1989 on its website home page. As of March 2013, the number stood at 303. An accessible list includes each exoneree, the state and years of conviction and release, and a link to a page of more detailed information, including whether the real perpetrator was found, contributing causes, and whether compensation was awarded. This list is frequently cited and is so prominent that some conflate the number of DNA exonerations with the total number of exonerations. This may result from the fact that until recently there has been no other prominent and easily referenced source of total exonerations. Furthermore, DNA exonerations are viewed as virtually incontrovertible, while non-DNA exonerations can be subject to some levels of doubt.

The National Registry of Exonerations (n.d.) (NRE), launched in the spring of 2012, is a joint product of the University of Michigan Law School and the Center on Wrongful Convictions (CWC) at the Bluhm Legal Clinic of Northwestern University School of Law. Its exoneration list, numbering more than 1,060 in early 2013, was drawn from raw lists kept by CWC, the data collected by Gross et al. (2005), the Innocence Project (n.d.), and other sources. It relies on the methodology pioneered by Gross et al. (2005). Each case is reviewed by staff to meet the criteria for exonerations based on actual innocence. For each exoneree, data include county and state of conviction; federal jurisdictions; most serious crime; additional conviction; year of crime; conviction and exoneration; exoneree's sentence, race, sex, and age at conviction; whether DNA exoneration or not; and contributing factors: mistaken eyewitness identification, false confession, perjury or false accusation, false or misleading forensic evidence, official misconduct, and inadequate legal defense. Clicking on the exoneree's name displays a few pages listing relevant data items and providing a narrative account of the case. Some narratives are by named authors.

Forejustice.org (n.d.) maintains a searchable database of the wrongly convicted that as of early 2012 included 3,310 exonerated people from 98 countries. The list can be searched by last name of the party or by location. Searching by location allows a search by American state and federal cases. Each name links to a page that lists years imprisoned, the charge, sentence, year convicted and cleared, location of trial, result, and case summary. Each page provides such sources of information as news articles and legal cases with citations and/or links and books. The listed cases are legal exonerations that include miscarriages of justice that cannot always be classified as cases of factual innocence. Forejustice.org is closely related to Justice Denied (n.d.), a volunteer group that publishes *Justice Denied* magazine in paper and on-line, which "publicizes cases of wrongful conviction, and exposes how and why they occur."

These lists provide a wealth of raw information about miscarriages of justice. The Innocence Project list is widely cited because it provides a "gold standard" of DNA cases that establish wrongful convictions with little doubt of error. The DPIC list focuses exclusively on death sentence exonerations and is often consulted given the prominence of the death penalty issue. The NRE has the potential of becoming the authoritative source for known DNA and non-DNA exonerations in the United States. The Justice Denied list is accessible and useful for gaining a quick overview of the scope of the numbers of wrongful convictions in the United States and other countries. None of the on-line lists are in formats that allow for quantitative analysis by researchers, although they are useful starting points for scholars wishing to explore the cases more systematically.

Estimating Wrongful Convictions

The fortuitous nature of obtaining exonerations means that counting and cataloging them can never amount to a "census" of all exoneration, let alone of all wrongful convictions that occur in a year. As a result, the incidence of wrongful convictions can only be estimated. In the alternative to estimating, many writers simply assume that the number of wrongful convictions is large enough to constitute a policy issue. Two kinds of estimates have been generated: precise quantitative estimates and qualitative estimates.

Precise Estimates. Poveda (2001) estimated a wrongful conviction rate of 1.4 % for released New York inmates convicted of murder based on (1) an official 1989 New York state study establishing that 21.2 % of inmates committed for murder were granted new trials, (2) the fact that 21.2 % of murder inmates whose cases were discharged by new trials in 1995 were discharged for substantial error, and (3) the assumption that these discharges, constituting 1.4 % of all murder commitments to New York prisons in 1992, were actually innocent.

Risinger (2007) established a wrongful conviction rate for rape-murder death sentences based on eleven DNA exonerations for rape-murder convictions occurring between 1982 and 1989. The numerator was set at 10.5 to produce a conservative estimate. The denominator of 319 was determined by finding (and listing) those individuals sentenced to death for rape-murders in the same time frame when the exonerees were convicted, discounted by the rate of similar cases where no usable DNA was available. This resulted in a wrongful conviction rate of 3.3 % (or 2 % before discounting the denominator for cases in the same period for which no usable DNA was available). Risinger also speculated that in some rape-murder cases the facts of guilt were so clear that DNA testing was not requested, which would further deflate the denominator of rape-murder death sentences consistent with the 11 (or 10½) exonerees, possibly yielding a maximum exoneration rate for this crime of 5 %. Risinger asserted the highly plausible conjecture that rates of wrongful convictions are unevenly distributed.

Gross and O'Brien (2008) examined 2,394 death sentences pronounced in US courts from 1973 to 1984. By 2004, the process of identifying exonerations for these cases had run their course. Fifty-four had been exonerated – a rate of 2.3 %. They then calculated the exoneration rate as of 2004 for those sentenced to death through 1989, again finding a rate of 2.3 % (86/3792). They then assessed the relation between capital exoneration and wrongful conviction rates among those sentences to death. Some capital defendants who were innocent may not have been exonerated because they did not participate in the arduous exoneration process or because they were executed, died from other causes, or had their sentences commuted to life terms (which reduces the pressure on innocence projects or others to seek their exoneration). It is possible that some among the exonerated participated in the murders for which they were convicted. Nevertheless, Gross and O'Brien estimate that the probability of innocence is very high for capital exonerees and the number of misclassifications low enough for the capital exoneration figure (2.3 %) to stand as a proxy for false death sentences of the actually innocent.

Roman et al. (2012) examined an unbiased cohort of 634 Virginia post-conviction sexual assault and/or homicide cases with DNA evidence resulting in 715 convictions, dating from 1973 to 1987. In 5.3 % of the cases, the offender was eliminated as a source of DNA, and the DNA exclusion was accompanied with probative evidence that appeared to support exoneration. In another 2.5 % of the cases, the offender was eliminated as a source of DNA, but the exclusion was not accompanied with probative exoneration evidence. This is the strongest empirical evidence to date that establishes the existence of a substantial proportion of wrongful convictions in a sample of *general* cases, rather than in group of cases that were targeted because they appeared to be miscarriages of justice.

General Estimates. Death sentence data tracked by the government and private groups are unusual in providing precise numbers of capital sentences, exonerations, and other case disposition information. This allows precise estimates of exonerations and wrongful convictions for this tiny fraction of the approximately one million felony convictions handed down by courts annually. The lack of precise or easily accessible data from court records makes it impossible to estimate wrongful felony conviction levels in the way that has been accomplished with capital sentences. In the absence of any general estimate, many wrongful conviction writings, especially by legal scholars, simply assume that the number is sufficiently large to constitute a problem needing to be addressed.

Huff et al. (1986) surveyed the opinions of state attorneys general and Ohio justice officials and defense lawyers about the annual incidence of wrongful conviction. Their estimates in the mid-1980s ranged from never (5.6 % of respondents) to less than 1 % (71.8 %), to 1–5 % (20.3 %), and to 6–10 % (2.3 %). The authors suggested that if an overall estimate of 0.5 % was accurate, it would mean more than 5,000 wrongful convictions occur each year. Observers have noted that these estimates are essentially guesses. Huff et al., however, did not assert that these estimates are a precise measure of wrongful convictions, but rather that the broader sampling of expert opinion provided some boundaries for the authors' *own* estimate of wrongful conviction incidence.

Ramsey and Frank (2007) updated and replicated Huff et al. (1986) with a sample of Ohio justice officials and defense lawyers reporting their beliefs about the frequency of wrongful convictions. Their modal response, taken as a whole, was that such error occurred in less than 0.5 % of cases in their own jurisdiction and in 1–3 % of cases in the United States. Defense attorneys reported higher estimates and police and prosecutors lower estimates, with the estimates of judges lying between. Zalman et al. (2008) replicated Ramsey and Frank's study and reported very similar opinions for Michigan officials. Neither survey claimed that officials' estimates were precise measures of wrongful convictions. Zalman et al. (2011) conducted a general residents' survey in Michigan at the time of their survey of experts and found that among the general population in 2005, the modal category of believed wrongful conviction incidence was 4–5 % and that 61 % of respondents believed that wrongful convictions occurred in between 4 % and 10 % of all felony convictions.

Zalman (2012) provided his own qualitative estimate of wrongful convictions for all felony convictions generally in the United States at a *minimum* of 0.5–1 %. The qualitative estimate was based primarily on a review of descriptive literature showing widespread defects in the quality of criminal case processing throughout the United States, supplemented by information provided by the precise death penalty wrongful conviction estimates (about 3 %), the validity of qualitative estimation techniques, and the approximations of officials, which are lower than estimates provided by residents in a statewide opinion survey (Zalman et al. 2011). Zalman (2012) concluded that a lower estimate than 0.5 % was highly implausible and that a higher general wrongful conviction rate of 2–3 % was possible, but less plausible than the lower rate. A 1 % wrongful conviction incidence rate would mean that about 10,000 wrongful convictions occur each year in the United States and that about 4,000 of those cases would result in prison sentences.

Estimates of the incidence of wrongful conviction have been contested by critical scholars and criminal justice system personnel who argue that in the absence of firm quantitative data, it can be assumed that decisions made by criminal justice system actors are virtually flawless. Controversy about wrongful conviction rates is not unique to criminal justice. As soon as a social problem becomes a political issue, legislators begin asking for numbers. It then becomes important to distinguish between scientific numbers and political numbers (Jencks 1994). Scholars who study child sexual abuse, for instance, report that there are probably more incentives to hide facts than to reveal them (Tucker and Cheit 2010). Some critical scholars and officials may compare qualitative wrongful conviction estimates to the seeming precision of crime statistics, without being aware of the many difficulties that accompany the development and interpretation of official crime data (Loftin and McDowall 2010). The nature of wrongful convictions and the inherent difficulty of finding them, however, make any attempt to generate a precise quantitative national estimate impossible.

Measuring the Correlates of Wrongful Convictions

What is known about wrongful convictions is derived from the cases that were discovered by various authors, researchers, and organizations, which then analyzed and catalogued them. Virtually every author or organization identified in the section on "Counting Wrongful Convictions," and many others, has described the factors associated with false convictions and typically has tallied the frequency with which the correlates of wrongful conviction have occurred.

It is as or more important to measure the correlates of false convictions by deconstructing the original conviction than simply to tally up a case and place it on a wrongful conviction or exoneration list. Examining the facts of a case to determine whether it led to the conviction of an innocent person proceeds simultaneously with determining what factors – errors in an investigation, prosecution, or adjudication – might have been responsible for the miscarriage of justice. Most wrongful conviction studies focus on these correlates. The most prominent sources of wrongful conviction, according to many sources, include eyewitness misidentification combined with suggestive lineup procedures, police tunnel vision that leads investigators to overlook exculpatory evidence, false confessions, reliance on informants and other witnesses who lie, erroneous forensic evidence (as a result of substandard laboratories, use of highly unreliable techniques or judgment errors, and deliberate falsehoods by forensic examiners), prosecutorial error and misconduct (especially the failure to divulge exculpatory evidence to the defense), ineffective assistance of defendant counsel, and such trial errors as misleading forensic science testimony and erroneous or biased judicial evidentiary rulings.

Although the correlates are often called *causes* of wrongful convictions, they cannot be deemed causal factors in a scientific sense (Gould and Leo 2010). Many errors such as misidentifications or false confessions occur in cases that do not result in convictions (Gross 1987; Drizin and Leo 2004). A handful of studies have compared samples of wrongful convictions to other convictions and have found some of the usual factors to be statistically associated with wrongful convictions but not others (Harmon 2001; Harmon and Lofquist 2005; Garrett 2008; Gross and O'Brien 2008). Because of the difficulty of matching wrongful conviction cases, which are discovered episodically and often by chance, with a cohort of similar, purportedly accurate convictions, this kind of research is arduous. Nevertheless, such measurements of wrongful convictions are research paths that ought to be pursued by criminologists (Leo 2005).

Another serious limit of false conviction-correlate accounts is that they overgeneralize their findings. For example, many writings claim that eyewitness identification error is the most common wrongful conviction correlated factor. The factors that are discerned in false conviction cases are dependent on the fortuitous discovery of exonerations and wrongful convictions and their unsystematic inclusion in a database. Studies identifying the frequency of correlates should be limited to the particular source of cases and not be extended to wrongful convictions generally. This is recognized by researchers who are aware of the unusual circumstances by which particular groups of cases come into being. Both Garrett (2011), analyzing the first 250 DNA exonerations listed by the Innocence Project, and Gross et al. (2005), analyzing 340 exonerations they discovered, have noted that the cases were an extremely unrepresentative sample of convictions. Almost all resulted from jury trials, whereas more than 95 % of convictions are obtained by guilty pleas. Their samples were almost entirely murder and rape cases, which constitute less than 1 % of all convictions. Thus, Garrett (2011) reports that flawed eyewitness evidence is found in 88 % of his sample, while the 2005 sample of Gross et al. reports such error in 50 % the murder cases and 88 % of the rape cases. Gross (2011) compared the correlates of his expanded exoneration list to the smaller percentage of murder and rape cases found in Gross et al. (2005) or Garrett (2011) and indicated that the proportionate list of correlates has changed. Covey (2011) studied the correlates in two sets of "mass exoneration cases" (which were acknowledged by Gross et al. (2005) and Gross (2011) but were not included in those exoneration lists). Covey discovered that the correlates of error in the wrongful conviction cases among the mass exonerations, virtually all of which were drug or gun possession convictions, were perjury and false confessions, with misidentification playing no role in those wrongful convictions.

Conclusion

The measurement of wrongful convictions is a question asked by all who study the subject (Gross 2008). Providing an answer to a general "how many" question is not easy. The inherent difficulties in discovering wrongful convictions, and governmental resistance and/or inability to identify and list them, hobble a more complete understanding. Nevertheless, precise estimates of wrongful death sentences, combined with a sound overview of the problems within the criminal justice process, lead to a conclusion that the numbers are more than minuscule.

Studies of individual wrongful convictions and lists of miscarriages of justice provide a wealth of information about how they occur, how they are discovered and exonerated, and what factors are correlated with them. A large number of studies by psychologists, legal scholars, and forensic experts are exploring ways to correct criminal justice processing errors, whether or not they directly cause wrongful convictions. Viewing the correlates of wrongful convictions as likely sources of these miscarriages of justice, however, is a sound conjecture. Criminologists can contribute to this research by applying the tools of social science to advance the study of the correlates and causes of wrongful convictions and the relationships among correlates and to more systematically study wrongful convictions in order to clarify our understanding of why errors of justice occur and how to correct them.

Recommended Reading and References

Bedau HA, Radelet ML (1987) Miscarriages of justice in potentially capital cases. Stanford Law Rev 40:21–120

Borchard EM (1932) Convicting the innocent: errors of criminal justice. Yale University Press, New Haven

Covey R (2011) Mass exoneration data and the causes of wrongful convictions (manuscript). Available at SSRN: http://ssrn.com/abstract=1881767 or doi: 10.2139/ssrn.1881767

Crichton N (2000) Information point: prevalence and incidence. J Clin Nurs 9:188 [Following Brenda Roe and Helen Doll (2000). Prevalence of urinary incontinence and its relationship with health status. J Clin Nurs 9:178–188]

Death Penalty Information Center (n.d.). http://www.deathpenaltyinfo.org/home Last retrieved 1 Jan 2012

Drizin SA, Leo RA (2004) The problem of false confessions in the post-DNA world. North Carol Law Rev 82:891–1007

Forejustice.org (n.d.). Forejustice.org

Forst B (2004) Errors of justice: nature, sources, and remedies. Cambridge University Press, Cambridge

Frank J, Frank B (1957) Not guilty. Doubleday, Garden City

Garrett BL (2008) Judging innocence. Columbia Law Rev 108:55–141

Garrett BL (2011) Convicting the innocent: where criminal prosecutions go wrong. Harvard University Press, Cambridge

Gorard S (2010) Measuring is more than assigning numbers. In: Walford G, Tucker E, Viswanathan M (eds) The Sage handbook of measurement. Sage, Los Angeles, pp 389–407

Gould JB, Leo RA (2010) One-hundred years later: wrongful convictions after a century of research. J Crim Law Criminol 100(3):825–868

Gross S (2008) Convicting the innocent. Annu Rev Law Soc Sci 4:173–192

Gross SR (1987) Loss of innocence: Eyewitness identification and proof of guilt. J Leg Stud 16:395–453.

Gross SR (2011) New data on exonerations in the United States. Paper presented at the Sixty-third annual meeting of the American society of criminology, Washington, DC

Gross SR, O'Brien B (2008) Frequency and predictors of false conviction: why we know so little, and new data on capital cases. J Empir Leg Stud 5(4):927–962

Gross SR, Jacoby K, Matheson DJ, Montgomery N, Patil S (2005) Exonerations in the United States, 1989 through 2003. J Crim Law Criminol 95(2):523–560

Harmon TR (2001) Predictors of miscarriages of justice in capital cases. Justice Q 18(4):949–968

Harmon TR, Lofquist WS (2005) Too late for luck: a comparison of post-Furman exonerations and executions of the innocent. Crime Delinq 51:498–520

Hoffman MB (2007) The myth of factual innocence. Chicago-Kent Law Rev 82(2):663–690

Huff CR, Rattner A, Sagarin E (1986) Guilty until proved innocent. Crime Delinq 32:518–544

Innocence Project (n.d.). http://www.innocenceproject.org/. Last retrieved 11 March 2013

Jencks C (1994) The homeless. Harvard University Press, Cambridge

Justice Denied (n.d.). http://justicedenied.org/. Last retrieved 1 Jan 2012

Leo RA (2005) Rethinking the study of miscarriages of justice: Developing a criminology of wrongful conviction. J Contemp Crim Just 21(3):201–223.
Loftin C, McDowall D (2010) The use of official records to measure crime and delinquency. J Quant Criminol 26(4):527–532
Marquis J (2006, January 26) Op-ed, the innocent and the shammed. N Y Times, p A23
National Registry of Exonerations (n.d.). http://www.law.umich.edu/special/exoneration/Pages/about.aspx. Last retrieved 11 March 2013
Norris RJ (2011) Assessing compensation statutes for the wrongly convicted. Crim Just Policy Rev 23 (3):352–374.
Poveda TG (2001) Research note: estimating wrongful convictions. Justice Q 18(3):689–708
Ramsey RJ, Frank J (2007) Wrongful conviction: perception of criminal justice professionals regarding the frequency of wrongful conviction and the extent of system errors. Crime Delinq 53(3):436–470
Rattner A (1988) Convicted but innocent: wrongful conviction and the criminal justice system. Law Hum Behav 12(3):283–294
Risinger DM (2007) Innocents convicted: an empirically justified factual wrongful conviction rate. J Crim Law Criminol 97(3):761–806
Roman J, Walsh K, Lachman P, Yahner J (2012) Postconviction DNA testing and wrongful conviction. Urban Institute, Washington, DC
Scheck B, Neufeld P, Dwyer J (2000) Actual innocence: five days to execution, and other dispatches from the wrongly convicted. Doubleday, New York
Schehr RC (2005) The criminal cases review commission as a state strategic selection mechanism. Am Crim Law Rev 42:1289–1302
Tucker W, Cheit R (2010) Sensitive issues and the difficulty to measure: the case of measuring child sexual abuse. In: Walford J, Tucker E, Viswanathan M (eds) The Sage handbook of measurement. Sage, Los Angeles, pp 429–454
Wade v. United States, 388 U.S. 218 (1967).
Yant M (1991) Presumed guilty: when innocent people are wrongly convicted. Prometheus Books, Buffalo
Zalman M (2012) Qualitatively estimating the incidence of wrongful convictions. Crim Law Bull 48(2):221–279
Zalman M, Smith B, Kiger A (2008) Officials' estimates of the incidence of 'actual innocence' convictions. Justice Q 25(1):72–100
Zalman M, Larson MJ, Smith B (2011) Citizens' attitudes toward wrongful convictions. Crim Justice Rev 37(1):51–69

Medication Assisted Therapy

▶ Drug Abuse and Alcohol Dependence Among Inmates

Men

▶ Victimization, Gender, and the Criminal Justice System

Mental Disorder

▶ Mentally Disordered Offenders Under Community Supervision

Mental Health Courts

Chrysanthi Leon and Kevin Ralston
Sociology and Criminal Justice, University of Delaware, Newark, DE, USA

Synonyms

Behavioral health courts

Overview

Mental health disorders are overrepresented within United States prison systems. Available national statistics show that nearly one quarter of state prisoners had received a clinical diagnosis or treatment within a 12-month period; when prisoners with symptoms of a mental disorder were also included, 56 % of state prisoners were defined as having a mental health problem (James and Glaze 2006 4). This alarming overrepresentation has been linked to the deinstitutionalization of mental care facilities and the lack of creation of outpatient services to fill the void created by their dismantling. Some of these mentally ill individuals that have fallen through the cracks of the health-care system have instead been swept up by the criminal justice system, in many cases for offenses related to the lack of treatment for their mental illness. Once they become a part of the criminal justice system, these mentally ill individuals often receive inadequate care for their mental illness and begin a cycle of reoffending and institutionalization in which their mental health needs are never appropriately addressed. James and Glaze found that almost one in four of state prisoners or jail inmates who had a mental health problem had been incarcerated three or more times (compared to one-fifth of those without, *Id.* at 1). Their analysis of national data also documents more homelessness, substance abuse, and prior traumatic abuse for mentally ill inmates compared to other offenders, as well as less utilization of treatment for these populations within the criminal justice system. Criminal justice agencies from police to corrections are increasingly concerned about the costs associated with using the justice system as the treatment facility of last resort, a function it is rarely equipped to adequately perform.

Mental health courts (MHCs) fall under the auspice of problem-solving courts as they seek not only to reduce the number of mentally ill persons cycling through the criminal justice system but also to improve the lives of their clients by helping them cope and better manage their mental illnesses. MHCs attempt to do this by restructuring the courtroom into a cooperative work group, where defense lawyers, prosecutors, probation officers, and mental health case workers work together providing monitoring, treatment, and linkage to external mental health facilities for the court's clients. During these proceedings the judge abandons his/her traditional role as a neutral arbitrator and interpreter of the law and becomes a risk manager who relies on the expertise of the courtroom staff but has the final decision about the potential risk the court's clients pose to themselves and the public, as well as implementing treatment plans designed specifically for each client and monitoring clients (Talesh 2007).

The Nature of Mental Health Courts

While there are many ways to address the insufficient approach to mentally ill offenders, one of the most notable and popular approaches is through the adaptation of problem-solving courts, best known through the longest-running variation, the drug court. Problem-solving courts use the concept of therapeutic jurisprudence. According to Berman and Feinblatt (2001) problem-solving courts share five elements: (1) They seek tangible outcomes for victims, offenders, and society such as reduced time in hospitalization for mental illness, (2) they seek to restructure how the government responds to social problems such as drug addictions and mental illness, (3) the use of judicial authority to change behavior of litigants and solve problems, (4) the collaboration between the courts and social services, and (5) nontraditional roles for judges, attorneys, and prosecutors.

Nationally, mental health courts are one of the fastest-growing types of problem-solving courts. Examples such as the Nathaniel Project in New York and the Behavioral Health Court in San Francisco are models for innovations across the country. The first MHCs were developed in the 1990s in the United States and Canada out of grassroots movements combined with government interest in diversion (Scheider et al. 2007, 34). The growth in the United States is partially due to the availability of seed money from the federal government, including a mental health court grant program established by the America's Law Enforcement and Mental Health Project Act of 2000 (Scheider et al. 2007, 167). However, matching funds are often required and the federal start-up funds are short term, so jurisdictions must be committed to providing internal funding. The spread of MHCs is also assisted through the information sharing made possible by organizations like the Criminal Justice/Mental Health Consensus Project, coordinated by the Council of State Governments Justice Center, which releases reports and maintains a database of best practices and model programs (http://consensusproject.org/, accessed 4/12/13). The well-established Center for Court Innovation, a public/private partnership between the New York State Unified Court System and the Fund for the City of New York, includes a focus on mental illness and also provides research, expert assistance, and demonstration projects such as the Brooklyn Mental Health Court (http://www.courtinnovation.org/, accessed 5/19/12).

The Process

Although MHCs fall under the rubric of problem-solving courts, their rapid proliferation, the multiple forms they assume, and the way they operate make defining what exactly a MHC is and how it operates problematic. For example, Griffin et al. (2002) describe how MHCs use at least three different methods of disposition of criminal charges to get their clients to adhere to their treatment regime. While people may enter a MHC at different stages of the justice process (at arrest in a true diversion program or after conviction or sentencing), referral is typically initiated by the public defender's office, a judge, or a probation officer. Most programs restrict participation based on an offender's criminal record, current offense, and mental health diagnosis. Eligibility is often determined by the prosecutors, and then selection is finalized by the mental health court team itself. Although the selection and admission details vary depending on the stage and level of the criminal justice process at which the court functions (e.g., misdemeanor diversion vs. felony violation of probation), generally, if an offender with mental illness agrees to participate, he or she is read a mental health court explanation that explains the trade-offs that must be accepted in order to participate in court, including the right to contest the charges if the MHC is a pretrial diversion. Often, the judge estimates a time frame for program enrollment and provides a sense of the graduated sanctions that may be employed. Common conditions include regular contact with a probation officer, regular therapy or counseling sessions, drug and alcohol treatment, and abstinence from criminal behavior. Time frames range from a few months to a few years.

While the client is in the mental health court program, they attend regular status hearings. Before each status hearing there is a pre-court meeting of the MHC team members: the judge, the prosecutor and defense attorney, the probation officers, the treatment staff, and some support personnel. At these pre-court meetings the team members review each client's behavior, discuss the client's situation, and give recommendations to the judge regarding the disposition of each client. The clients are not present at these pre-court meetings.

The in-court status hearings resemble traditional court proceedings in some ways; they take place in a courtroom open to the public, with the judge, the prosecuting attorney, and the defense in their traditional places within the courtroom. One of the team members typically gives an account of a client's progress since the previous status hearing, the individual and the judge talk briefly, and the judge rewards, sanctions, and advises the client accordingly. The performative nature of these sessions is even more striking than in traditional court proceedings; for example, the judge often gets his/her script from the other team members prior to court ("What should I say to Ms. Smith?"), and the order of client appearances may be calculated to send a message to other participants. For example, graduations may be conducted first, to dramatize what the other participants are working towards.

Though the exact definition of what a mental health court is and how it operates is elusive, Steadman et al. (2001) put forth four basic criteria which MHCs follow: (1) All persons with mental illness identified for referral to community-based services on initial booking are handled on a single court docket, (2) a courtroom team's approach is used to arrive at recommended treatment and supervision plans with a person specifically designated as a "boundary spanner" to ensure actual linkage, (3) assurance of existing appropriate treatment slots is necessary before the judge rules, and (4) appropriate monitoring occurs under the court aegis with possible criminal sanctions for noncompliance, such as reinstituting continued charges or sentences. Redlich et al. (2005) describe a second generation of MHCs which are more likely to accept felony cases, use jail as a sanction, and offer different forms of court supervision. In 2006, Redlich et al.

conducted a comprehensive survey of all the existing MHCs, and their results supported Redlich et al. (2005) in concluding that MHCs are continuing to evolve, as well as describing additional characteristics that are common to MHCs. In a 2009 review of the literature and based upon a consultation with leading researchers and practitioners, Almquist and Dodd also found that MHCs increasingly include violent offenders and recognize the co-occurring substance abuse needs that accompany the serious mental illnesses of their clients (2009).

Empirical Results

Increasing Treatment

While the earliest work on MHCs was descriptive in nature, studies of the effectiveness of MHCs have recently begun to emerge. These studies have generally relied on outcome comparisons between a certain MHC and a similarly situated court that handles the same type of offenders but that does not follow a MHC model. The results indicate that MHCs produce favorable outcomes in both providing their clients with increased mental health care and services, as well as reducing recidivism.

Boothroyd et al. (2003), in their evaluation of the Broward misdemeanor MHC, found that the use of behavioral health services by MHC participants increased from 36 % to 53 % after enrollment, and clients were 50 % more likely to continue receiving treatment after their court appearance than defendants in the comparison group. They also found that the amount of behavior health services that MHC participants received increased by 61.6 %, while the amount of behavior health services received by those in the comparison court decreased by 18.3 %. Trupin and Richards (2003), in their evaluation of two Seattle MHCs, found that those who opted into the MHC received more treatment and access to mental health care than mentally ill offenders that opted not to participate in the MHC. Using a 12-month pre-post comparison design, Herinckx et al. (2005) found that offenders who participated in the Clark County misdemeanor MHC in Washington "received more hours of case and medication management and more days of outpatient service after enrollment" in the 12 months after enrollment as compared to the 12 months before enrollment into the program, as well as spending less time in inpatient treatment facilities (855). Keator et al. (2013) also found that participants in MHCs accessed community treatment more frequently and at a quicker pace following discharge than those who did not participate in a MHC. The preceding studies indicate that MHCs have increased the amount of treatment their clients receive, which is an important step in breaking the cycle of continually re-incarcerating mentally ill offenders. Not only do MHC participants seek out community treatment at a greater rate than those who do not participate in a MHC, but participants have greater perceptions of procedural justice than those who do not participate in a MHC (Canada and Watson 2013).

Reducing Recidivism

A recent meta-analysis performed by Sarteschi et al. (2011), which quantitatively examines a large number of MHC studies to determine a pattern of significant results, found two distinct patterns: (1) that MHCs are consistently found to reduce recidivism among participants and (2) that the majority of participants in MHCs in the United States are white males in their mid-30s. The idea that MHCs reduce recidivism for participants is important because it is one of the main goals of a MHC; this illustrates the potential success of MHCs for the misdemeanor offenders who are chosen to participate. A multisite study conducted by Steadman et al. which compared 407 MHC participants with 600 matched offenders who had no contact with MHC, similarly found that in the 18 months following entry into mental health court, the MHC group had a lower rearrest rate, fewer post-18-month arrests, and fewer post-18-month incarceration days than the comparison group (Steadman et al. 2010).

Sarteschi et al. (2011) also indicate that MHCs appear to reduce recidivism among participants. MHCs are moderately effective in curtailing subsequent criminal activity, and as expected, the reduction in recidivism is most pronounced for individuals who participate and complete the guidelines outlined by the MHC. Examining the results of the meta-analysis, it appears that the reason for the reduction in recidivism for participants who complete a MHC has to do with their GAF (Global Assessment of Functioning) scores. The successful completion of a MHC has been found to increase participants' GAF. The GAF test measures how well an individual deals with problems they encounter in their daily lives. The higher the GAF score, which ranges from 0 to 100, the lesser extent to which mental illness symptoms are present in the individual. These increased GAF scores indicate that individuals who complete a MHC are able to function to a greater degree in their daily lives and are less likely to be bothered by an unexpected problem. This could lead to a reduction in psychiatric emergency room visits for successful MHC participants, thus reducing individual and state costs, a finding supported by the meta-analysis performed by Sarteschi et al. (2011).

Specific case studies can allow for a more in-depth analysis of how MHCs can reduce recidivism. McNiel and Binder compared 170 participants in the San Francisco Behavioral Health Court to 8,067 other adults with mental disorders who were booked into an urban county jail after arrest during the same interval (2007, 1396). Based on an intent-to-treat sample (i.e., all of those who enrolled in MHC regardless of completion), both the participants and the graduates were less likely to have any new charges than the comparison groups. After 18 months, only 34 % of MHC graduates had received any new charges, compared to 56 % of the treatment comparison (*Id.*). Most strikingly, the MHC graduates evidence half the rate of violent charges when compared to the other group (just 6 % of MHC graduates were charged with violent offenses during the follow-up, compared to 13 % of the others) (*Id.*).

Examining the 12 months after enrollment in a MHC, crime rates for individuals with a severe and persistent mental illness who completed the program were four times less than individuals who did not complete a MHC program (Herinckx et al. 2005). When examining just those participants who completed the program, after 1 year approximately 55 % had no arrests and probation violations were reduced by approximately 60 % (Herinckx et al. 2005). Overall, successful completion of a MHC reduced the likelihood of rearrest by approximately 3.7 times compared to participants who do not successfully complete a MHC program (Herinckx et al. 2005). MHCs also reduced the likelihood of new charges for violent crimes, and when new charges did occur, participants in MHCs had a greater period of time between charges than those who do not participate (Herinckx et al. 2005). The examination of violence and MHCs also indicates the reduction in new charges is maintained once the MHC graduate is no longer under court supervision. In general, studies indicate that MHCs reduce recidivism for individuals who successfully complete the court-mandated guidelines by showing the participant how to deal with problems that can arise in their daily lives, which reduces psychiatric emergency room visits and costs for both the individual and the state. However, it should be noted that the reduction in recidivism might only be a result of structured follow-ups with the MHC. According to Burns et al. (2013), once the structured period of follow-up with the MHC concludes, participants may see recidivism more in-line with the average of those who did not participate in a MHC. Their conclusions are based on a MHC in northern Georgia and a comparison of completers and non-completers 2 years after involvement in the MHC. While Burns, Hiday, and Ray's results may not be representative to all MHCs, they are worth mentioning as an issue facing graduates of MHCs.

An additional cautionary note is needed here: the empirical support for mental health courts is promising, but it suffers from selection biases. No large random assignment study has been performed to date (but see Cosden et al. 2003). As McNiel and Binder explain in their study, "Although propensity weighting helps control for nonrandom assignment, it can only adjust for observed covariates, and it may not have adequately adjusted for unobserved variables that may have influenced selection into mental health court (e.g., as participants who entered mental health court voluntarily agreed to have their cases handled in mental health court rather than traditional court, it is possible that they differed in unobserved variables such as treatment motivation) (2007, 1402)." Thus, whether MHCs are helping the offenders who would likely have performed better without this intervention remains unknown.

Key Issues and Controversies

As discussed in the previous section, MHCs appear to be successful based on the empirical analyses completed to date, including the ability to successfully link participants with mental health services, reduce recidivism, reduce costs, and aid the decriminalization of individuals with less severe mental illnesses. However, this does not mean there are no controversies surrounding MHCs. Critics contend that they are coercive, ineffective when the proper mental health services are not provided, and that selection bias is present along gender and racial lines. This section discusses some of these controversies more closely.

MHCs as Reactive

The first major controversy surrounding MHCs is that they are reactive and fail to address the larger systemic and structural factors that account for the overrepresentation of the mentally ill in the criminal justice system (Leon 2007). Individuals with a mental illness who participate in these programs only receive treatment after they commit a criminal offense. While the specialty courts provide a link to treatment for those who need it, scholars echoing this criticism feel that mental health services need to be more adequately available for all individuals who need them. Instead of waiting for an individual to commit a criminal offense, mental health services must be more fully funded by all levels of government and available to individuals with mental illnesses who are untreated or undertreated. A more proactive response could include advertisement to inform individuals about available services, along with outreach to the community. Allocating resources for preventative care instead of providing a reactive "drop in the bucket" will entail providing services to all relevant persons within a community, instead of only those committing criminal acts. This would reduce any privileging of one group over another.

MHCs as Coercive

Another major criticism of MHCs is the troubling concern that individuals may not have a full understanding of the potential consequences when they choose to participate. While Canada's approach demonstrates an insistence on reminding participants that they can withdraw at any stage, this is not typically the case in courts within the United States. Many of the studies find that although participants agree to participate, many claim they were not told that the program is voluntary (Sarteschi et al. 2011). Participants may not be told that the increased supervision they will undergo through the specialized program by its nature extends the opportunity for their failures to be detected and penalized. Another important concern to come out of this line of research is that many of the participants in MHCs do not understand facts about court procedures (Sarteschi et al. 2011). This is troubling because of the legal requirements of MHCs. Participants who are involved in the programs usually must plead guilty to their charge (Sarteschi et al. 2011). Once the participant successfully completes the program, their record may be expunged and the guilty plea should be erased. However, that does not always happen. Once the individual pleads guilty, a criminal record is established. If the individual does not complete the program, the criminal record stays with them and can subsequently haunt them. When an individual successfully completes the program, the charges and plea should be automatically expunged or dismissed and the criminal record should also disappear. Unfortunately, studies have found that this is not always the case. For whatever reason, some successful graduates of MHCs continue to have a criminal record even though the charge should have been expunged or dismissed.

Selection Bias

The meta-analysis performed by Sarteschi et al. (2011) indicated that the majority of participants in MHCs are white males in their 30s, even though African Americans, especially young African American males, are disproportionately represented in the criminal justice system. This indicates selection bias in MHCs, another major criticism of the specialty courts. This is not uncommon for jail diversion programs in general, as a disproportionate number of individuals being placed into jail diversion programs are older white females. Not only are there racial differences in admittance into MHCs, but there are also gender differences. While the majority of participants in MHCs are white males, females, especially white females with a severe mental illness, are more likely to be diverted into a MHC than men. This is even true when controlling for type of charge, crime, and race.

Furthermore, the other major finding of the Sarteschi et al. meta-analysis is troubling: the majority of participants in MHCs in the United States are white males in their mid-30s. Considering that a disproportionate number of African Americans are involved in the criminal justice system and that mental illness is found in a similar proportion of blacks and whites, the number of blacks and whites participating in MHCs should be almost equal. Having a majority of white males as MHC participants could indicate selection bias by the courts.

Critics of the selection criteria of MHCs argue that it is hard to examine whether a selection bias exists because there is often great leeway in selecting participants for the specialty court. Instead, it is up to the MHC to decide when someone fits the criteria to participate in the program; this comes up, for example, when deciding whether the details of past offenses count as "violent." This allows for the MHC to choose participants that they think most likely to succeed and ignore a possibly more deserving individual who they think will not succeed: "cherry-picking." A selection bias could indicate that MHCs are ignoring a certain population who needs treatment but might be seen as too difficult to treat. These views about who will be successful could ultimately lead to skewed research findings indicating MHC efficacy when they have not been tested across the eligible population.

Future Directions

Given the institutional support for mental health courts, as well as the promising data regarding their efficacy, their continuation and expansion seem likely. Researchers, practitioners, and advocates must remain attuned, however, to concerns about coercion and biases in the selection and treatment of participants, as well as to the broader efforts needed to prevent contact with the criminal justice system through comprehensive mental health care for all.

Recommended Reading and References

Almquist L, Dodd E (2009) Mental health courts: a guide to research-informed policy and practice. Council of State Governments Justice Center: New York, NY

Berman G, Feinblatt J (2001) Problem-solving courts: a brief primer. Law Policy 23:125–140

Boothroyd RA, Poythress NG, McGaha A, Petrila J (2003) The Broward mental health court: process, outcomes and service utilization. Int J Law Psychiatry 26:55–71

Burns PJ, Hiday VA, Ray B (2013) Effectiveness 2 years postexit of a recently established mental health court. Am Behav Sci 57:189–208

Canada KE, Watson AC (2013) "Cause everybody likes to be treated good": perceptions of procedural justice among mental health court participants. Am Behav Sci 57:209–230

Cosden M, Ellens J, Schnell JL, Yamini-Diouf Y, Wolfe MM (2003) Evaluation of a mental health treatment court with assertive community treatment. Behav Sci Law 21:415–427

Griffin PA, Steadman HJ, Petrila JD (2002) The use of criminal charges and sanctions in mental health courts. Psychiatr Serv 53(10):1285–1289

Herinckx HA, Swart SC, Ama SM, Dolezal CD, King S (2005) Rearrest and linkage to mental health services among clients of the Clark county mental health court program. Psychiatr Serv 56:853–857

James DJ, Lauren EG (2006) Mental health problems of prison and jail inmates. U.S. Department of Justice, Bureau of Justice Statistics. September 2006, NCJ 213600

Keator KJ, Callahan L, Steadman HJ, Vesselinov R (2013) The impact of treatment on the public safety outcomes of mental health court participants. Am Behav Sci 57:231–243

Leon CS (2007) Should courts solve problems? Connecting theory and practice. Crim Law Bull 43:879–899

McNeil DE, Binder RL (2007) Effectiveness of mental health court in reducing criminal recidivism and violence. Am J Psychiatry 164:1395–1403

Redlich AD, Steadman HJ, Monahan J, Petrila J, Griffin PA (2005) The second generation of mental health courts. Psychol Publ Policy Law 11:527–538

Redlich AD, Steadman HJ, Monahan J, Robbins PC, Petrilla J (2006) Patterns of practice in mental health courts: a national survey. Law Hum Behav 30:347–362

Sarteschi CM, Vaughn MG, Kim K (2011) Assessing the effectiveness of mental health courts: a quantitative review. J Crim Justice 39:12–20

Schneider RD, Bloom HY, Heerema M (2007) Mental health courts. 1st ed. Toronto, Ontario: Irwin Law

Steadman Henry J, Redlich A, Callahan L, Robbins P, Vesselinov R (2011) Effect of mental health courts on arrests and jail days a multisite study effect of mental health courts on arrests and jail days. Arch Gen Psychiatry 68:167–172

Steadman HJ, Davidson S, Brown C (2001) Mental health courts: their promise and unanswered questions. Law Psychiatry 52:457–458

Talesh S (2007) Mental health court judges as dynamic risk managers: a new conceptualization of the role of judges. DePaul Law Rev 57:92–132

Trupin E, Richards H (2003) Seattle’s mental health courts: early indicators of effectiveness. Int J Law Psychiatry 26:33–53

Wolff N, Fabrikant N, Belenko S (2011) Mental health courts and their selection processes: modeling variation for consistency. Law Hum Behav 35: 402–412

Mental Illness

▶ Drug Abuse and Alcohol Dependence Among Inmates

▶ Mentally Disordered Offenders Under Community Supervision

Mentally Disordered Offenders Under Community Supervision

Sarah M. Manchak[1], Jennifer Eno Louden[2] and Jennifer L. Skeem[3]
[1]School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA
[2]The University of Texas at El Paso, El Paso, TX, USA
[3]University of California, Irvine, Irvine, CA, USA

Synonyms

Mental disorder; Mental illness; Parole; Probation

Overview

Offenders with serious mental illness (schizophrenia, bipolar, major depression) are overrepresented in criminal justice settings. Most offenders – both with and without mental illness – are not incarcerated but instead are supervised in the community on probation and parole. Offenders with serious mental illness are more likely to fail on community supervision than their relatively healthy counterparts. With a focus on community supervision in North America, this entry distills research on (1) explanations for why offenders with mental illness are at increased risk for supervision failure and (2) promising- and evidence-based practices for reducing this risk.

Statement of the Problem

Overrepresentation of Offenders with Mental Illness

Of the 7.2 million people under correctional supervision (Bureau of Justice Statistics 2011), nearly one-sixth have a serious mental illness (SMI; Fazel and Danesh 2002). Disorders that fall under this classification include schizophrenia (i.e., symptoms such as delusions and hallucinations, disorganized thought processes, and even catatonia), bipolar disorder (i.e., fluctuations in mood and activity levels that range from severely depressive to manic), and major depressive disorder (i.e., intense feelings of sadness and anhedonia). All three disorders can severely affect daily functioning. The current rates of SMI in correctional populations are approximately three times higher for men and twice as high for women than that found in the general population (see Teplin 1990; Kessler et al. 2006).

Community Supervision Failure

The majority (~70 %) of offenders both with and without SMI are supervised in the community on probation or parole (Bureau of Justice Statistics 2011). While on community supervision, offenders with SMI are at greater risk of failure (i.e., breach supervision and return to custody in jail or prison) than offenders without SMI. Based on a cohort of 45,000 California State parolees (20 % of whom were pre-identified as having a mental illness), Eno Louden and Skeem (2011) found that those with mental disorder were more than three times more likely to have parole revoked (i.e., terminated) than those without mental disorder. Moreover, those with acute (i.e., severe, sudden increase/onset of symptoms) mental illness had even higher rates of return to custody than parolees with more stable (i.e., ongoing, managed symptoms) mental illness. Similarly, in an unpublished dissertation that compared 115 probationers with SMI to 518 randomly selected non-disordered probationers, Dauphinot (1996) found that probationers with SMI were more likely to be arrested (54 %) and revoked (37 %) than those without SMI (30 %, 24 %, respectively) over a 3-year follow-up.

Technical violations appear to be an especially problematic concern for offenders with SMI. Relative to non-disordered offenders, offenders with SMI have disproportionately high rates of technical violations and are more likely to fail community supervision failure as a result of a technical violation. This finding has been demonstrated in a small sample of Canadian federal offenders (37 SMI vs. 37 non-disordered; Porporino and Motiuk 1995) and parolees in New York (147 SMI vs. 400 non-disordered; Feder 1991) and California (Eno Louden and Skeem 2011). All three studies indicate significant proportional differences between offenders with SMI and non-disordered offenders in their failure due to technical offenses. One study, in particular, showed that the relative risk of supervision failure due to technical violations (vs. person, property, drug, and minor offenses) is almost twice as great for offenders with SMI (52.9 %) than non-disordered (29.7 %) offenders (Eno Louden and Skeem 2011).

The Link Between Mental Illness and Recidivism

Several explanations for community supervision failure among offenders with SMI have been proposed. Skeem and colleagues (2006, 2011) provide a framework for understanding three important paths to failure for offenders with SMI. The first two paths, direct and indirect proximal effects, keep the role of mental illness central to the equation (see Skeem and Eno Louden 2006). The third path by which offenders with SMI can fail community supervision de-emphasizes mental illness as a proximal or immediate cause of supervision failure. Instead, this path views mental illness as a distal factor that is related to – or mediated by – other, stronger predictors of supervision failure (see Skeem et al. 2011). Although these pathways can overlap, evidence for each of these three routes is distilled separately below.

The Proximal Direct Route

The direct route between mental illness and community supervision failure suggests that symptoms of mental illness lead directly to criminal behavior. For example, someone may experience hallucinations in the form of voices compelling them to engage in an illegal act (e.g., theft). In this particular case, the criminal act is directly attributable to the experience of psychotic symptoms.

Although lay people may endorse beliefs that support a link between mental illness and crime or violence, the empirical evidence suggests such beliefs may be unfounded. In fact, among offender samples, mental illness is either unrelated or negatively correlated with future criminal behavior. In a meta-analysis conducted by Bonta et al. (1998), p. 35 predictors of general recidivism and 27 predictors of violent recidivism were examined across 64 unique samples. A clinical domain, comprised of clinical variables such as psychosis, treatment history, mood disorder, and current mental disorder, showed significant negative effects for both general ($r = -.02$) and violent recidivism ($r = -.03$; for a recent update, see Quinsey et al. 2006).

Even though mental illness does not predict recidivism for the majority of offenders with SMI, the symptoms of mental illness may directly lead to criminal behavior for a small proportion of offenders. In one study, 113 individuals with mental illness in a jail diversion program were interviewed about the cause of their index offense. Three independent raters used a 100-point scale (definitely not [0] to definitely [100]) with 5 point increments to rate the likelihood that the offense was attributable to either "the specific influence of concurrent delusions or hallucinations on the criminal offense identified in the police arrest report" or "any other symptom-based influence, such as confusion, depression, thought disorder, or irritability" (Junginger et al. 2006, p. 880). Results indicated that the index offense for only about 8 % of the sample was attributable to the symptoms of mental illness.

In a similar study, Peterson et al. (2010) matched 111 parolees with SMI to 109 parolees without SMI and examined their pattern of offending using official parole records. Parolees were classified into one of five groups based on their offense patterns by four independent raters: (1) psychotic (violent offenses driven by delusions or hallucinations), (2) disadvantaged (survival crimes such as shoplifting), (3) reactive (crimes driven by hostility or impulsivity), (4) instrumental (goal-driven crimes that involve manipulation or deceit), or (5) gang or drug related. Here, the offense pattern was attributable to psychosis in only 5 % of the parolees with SMI. Although replication is needed, the two studies taken together emphasize the importance of the link between mental illness and crime for a small proportion of offenders – less than 10 %. Meanwhile, the relationship between mental illness and crime for the majority of offenders with SMI may be better understood through other avenues, which are reviewed next.

The Proximal Indirect Route

This second avenue by which offenders with SMI can fail community supervision also emphasizes a proximal link between psychiatric symptoms and failure, but this route is indirect. In this path, offenders with SMI are believed to fail supervision not because their psychiatric symptoms lead to criminal behavior but because these symptoms lead to another outcome such as declined functioning or substance use. This third variable then leads to the termination of supervision.

For instance, it could be that the symptoms of mental illness prevent offenders on community supervision from doing what is required of them. If a supervisee is severely depressed and physically unable to get out of bed to attend probation or parole appointments, he could incur a technical violation or even be revoked for noncompliance. Alternatively, supervision can be revoked due to the inability to stay clean and sober – a standard condition of community supervision. Indeed, having SMI raises the likelihood of a co-occurring substance disorder; almost 75 % of offenders with SMI also have a substance use problem (James and Glaze 2006; Abram and Teplin 1991). In both scenarios, mental illness leads to the failure of meeting conditions of community supervision, which can then lead to revocation.

The Third Variable Route

The final pathway of the mental illness-recidivism link suggests that a third variable – not mental illness – is actually responsible for community supervision failure. In this route, something that is separate from – but related to – mental illness can explain supervision failure. The mental health status of the individual remains important, but it is viewed as a distal variable that exposes offenders to more proximal factors that contribute to supervision failure.

The research examining this "third variable" route has identified two very distinct but important factors that contribute to the community supervision failure of offenders with SMI. The first factor, the role of supervision practices among probation and parole officers, demonstrates a truly spurious relationship of the mental illness-recidivism link. The research on this topic suggests the manner in which an officer supervises an offender can directly affect the offender's criminal justice outcomes. This process can occur with offenders with and without SMI, but there are some concerns that officers' stigmatizing attitudes about mental illness can influence officers' supervision style with this population in particular. The second factor, an offender's criminogenic risk, demonstrates a mediated relationship between mental illness and recidivism. Here, mental illness exposes individuals to risk factors that strongly relate to criminal behavior (e.g., poverty, antisocial associates).

Supervising Officer Effects. Officer orientation (i.e., guiding viewpoint) toward rehabilitation versus public safety can influence offenders' outcomes, whether they have mental illness or not. In a seminal ethnographic study of 7,000 probationers and parolees and their supervising officers, Klockars (1972) found that officer supervision orientation characterized by extreme emphasis on law enforcement/control or social casework/rehabilitation, as opposed to a blend of these approaches, can negatively affect supervisees' outcomes. Decades later, Paparozzi and Gendreau (2005) provided further evidence to support this theory when they examined similar supervision styles (punishment, social casework, and balanced) in a matched sample comparing 142 parolees on intensive supervision to 142 offenders on standard parole. Offenders who had "balanced" officers were least likely to be revoked for any offense or due to a new conviction than offenders who had punishment-oriented or strictly social casework-oriented officers.

Results from these studies of general offenders likely also apply to offenders with SMI, and recent evidence indicates that officers may be particularly likely to adopt these supervision strategies with offenders with SMI. Doing so can adversely affect their outcomes. In other words, officers may be inclined to supervise offenders with SMI in ways that increase their likelihood of returning to custody.

In an experimental study, Eno Louden and Skeem (2013) presented 234 probation officers with a vignette portraying a probationer with mental disorder, substance abuse, both, or neither. Officers were asked to provide risk assessment ratings for the probationers then predict how they would supervise that probationer if he were placed on their caseload. Officers judged probationers with mental disorder, particularly schizophrenia, as being at high risk for both new offense and violence, compared to probationers with no disorder or only substance abuse. Specifically, officers estimated that a probationer with schizophrenia had a 68.6 % chance of committing a new offense and a 55.2 % chance of committing violence while on probation, compared to 49.3 % and 39.0 %, respectively, for a probationer with no mental disorder. Further, they also sought to monitor these probationers closely via frequent meetings – wanting to meet with a probationer with schizophrenia an average of 6.1 times per month compared to 3.3 times per month for a probationer with no disorder (Eno Louden and Skeem 2013). Other research has shown that closer supervision can increase detection of noncompliance, particularly for technical violations (Petersilia and Turner 1993; Paparozzi and Gendreau 2005). Although much work remains to be done on why officers may hold certain beliefs about the risk of offenders with mental illness and the reasons for supervising them more closely, the strong influence officers have on supervisees' outcomes is apparent.

Shared Criminogenic Risk. In addition to officer effects, another "third variable" that has a strong influence on community supervision failure of offenders with SMI is criminogenic risk (e.g., risk factors related to criminal behavior). As established through several comprehensive meta-analytic studies, the strongest risk factors for criminal behavior – dubbed the "central eight" – are history of antisocial behavior; antisocial personality pattern; antisocial cognition; antisocial associates; family and marital instability; deficits in school or work aptitude, productivity, or motivation; lack of pro-social leisure and recreational pursuits; and substance use (see Andrews et al. 2004, 2006; Andrews and Dowden 2006). Research suggests that the criminogenic risk factors most strongly related to recidivism in non-disordered offenders also predict recidivism for offenders with SMI (Bonta et al. 1998).

In fact, mental illness may disproportionately expose offenders with SMI to specific criminogenic risk factors such as poverty and antisocial associates that lead to community supervision failure. Several studies have indicated that not only do offenders with and without SMI share the same risk factors for recidivism but also that those with SMI actually are higher on many general criminogenic risk factors. In one study of 627 offenders with and without mental health problems (defined as the presence of depression, psychosis, prior suicide threats or attempts, and other emotional distress), Girard and Wormith (2004) found higher risk/needs scores for offenders with mental illness than those without mental illness on an Ontario version of the Level of Service Inventory, a variation of the popular, well-validated risk assessment. Moreover, the relationship between offenders' specific risk/needs profile and general recidivism was stronger for those with mental health problems than those without them. In another study, Skeem et al. (2008) examined 112 parolees with SMI and 109 parolees without SMI and found those with SMI had higher general (total scores) and specific risk/needs profiles on the Level of Service/Case Management Inventory (LS/CMI; Andrews et al. 2004) than non-disordered parolees.

Routes of Supervision Failure: A Summary

In summary, the relationship between mental illness and community supervision failure for offenders with SMI is complex. There are various routes by which offenders with SMI fail community supervision. For a small proportion of offenders with SMI, community supervision failure is directly attributable to the symptoms of mental illness. However, the symptoms of mental illness can also influence a trajectory of noncompliant behavior that leads to technical violations, an especially problematic concern for this group. For the majority of offenders with SMI, however, supervision failure is likely attributable to criminogenic risk factors that are shared with non-disordered offenders. Finally, and perhaps to a lesser extent, the specific supervision styles and strategies employed by community corrections officers can also play an important role in community supervision failure. Equipped with an understanding of how offenders with SMI can fail community supervision, this entry now turns to an overview of the current approaches used to supervise offenders with SMI in the community, highlights their evidence base, and discusses the strengths and weaknesses of the current policy model in light of this empirical evidence.

Strategies for Supervising Offenders with SMI in the Community: Specialty Mental Health Caseloads

Perhaps the most common current approach employed in the community supervision of offenders with SMI is specialty mental health caseloads. The last decade has seen increasing implementation of specialty mental health caseloads to manage offenders with SMI on community supervision. In 2002, the Council of State Governments recommended the implementation of specialty caseloads in response to the problem of supervision failure for this population and the growing concerns about the consequences of offenders with SMI becoming more deeply entrenched in the criminal justice system. At the time of this recommendation, there was unfortunately no data to support the effectiveness of implementing this type of caseload, nor did the recommendation include directives for establishing specialty caseloads. Since that time, Skeem et al. (2006) conducted a national survey of specialty mental health probation. This study provided some basic insight into what these caseloads look like in practice.

Although there is heterogeneity in how specialty mental health caseloads are implemented across jurisdictions, they do share five distinct features: (1) reduced caseload sizes, relative to non-specialty caseloads, (2) caseloads comprised solely of offenders with SMI, (3) officer interest or training in mental health, (4) coordination of internal (i.e., probation) and external (i.e., treatment) resources, and (5) emphasis on problem-solving over sanctioning (Skeem et al. 2006). Additionally, specialty mental health community supervision places a very strong emphasis on mental health treatment (psychiatric medications and therapy). In doing so, as argued elsewhere (see Skeem et al. 2011), these programs make an implicit assumption that symptom reduction can lead to improved criminal justice outcomes.

Perhaps the most rigorous examination of specialty probation performed to date is that recently completed by Skeem and colleagues. In this study, 183 offenders with SMI on specialty mental health probation were matched to 176 probationers with SMI on traditional probation on the following criteria: age, gender, race, time on probation, and index offense. Probationers and their supervising probation officers were followed over the course of 1 year. During this time, probationers completed three semi-structured interviews – one approximately 2 months after beginning probation and again 6 and 12 months later. Officers completed questionnaires on the same schedule. Participants' official criminal justice outcomes (FBI arrests rap sheets and violation and termination reports filed with the court) were collected 2 years post baseline.

Two important preliminary findings have emerged from this study. First, what officers do in specialty supervision differs from what they do in traditional supervision. Specifically, specialty probation officers used more problem-solving strategies than threats and sanctions, had better skills in navigating and working with outside social service agencies (i.e., "boundary spanning"), and established better relationships with their supervisees than officers in traditional probation supervision (Manchak et al. 2007). Additionally, offenders with SMI on specialty probation received more evidence-based integrated dual diagnosis treatment (i.e., treatment that targets mental health and substance use symptoms simultaneously) than those on traditional probation (Manchak et al. 2010). Second, specialty probation reduced recidivism but not for reasons that were expected. Effectiveness of specialty probation was not mediated by probationers' symptom reduction but by high-quality officer-probationer relationships characterized by firmness, caring, and trust (Skeem et al. 2010). These findings would suggest that the implicit model guiding practices like specialty supervision may be flawed; effective officer practices – not treatment – appear to be essential for affecting outcomes for offenders with SMI.

Moving Research into Practice: Prescribing Evidence-Based Practices

Practice

The available research on specialty caseloads underscores the importance of what is already being emphasized in correctional intervention with general offenders: "Core Correctional Practices" (CCP). First discussed by Andrews and Kiessling (1980), these practices place strong emphasis on *how* correctional interventions are implemented. Appropriate use of authority, modeling, behavioral reinforcement, and problem-solving are viewed as essential CCPs. Additionally, CCPs entail linking offenders to needed community resources and services. Perhaps most important, however, is the establishment of strong relationships between supervisees and probation and parole officers; these relationships are most effective when they are characterized by warmth, flexibility, empathy, openness, mutual respect, and liking and use "directive, solution-focused, structured, non-blaming, or contingency-based communication" (Dowden and Andrews 2004, p. 208). A meta-analysis of 273 effect sizes shows that the relationship component correlates strongly with average effect sizes in recidivism reduction ($r = .25$), and the additional CCP components enumerated above can increase this correlation ($r = .41–.47$; Dowden and Andrews 2004).

Other work that has come to dominate the landscape of general correctional intervention, and which is slowly beginning to take hold in policy recommendations with offenders with SMI (see Prins & Draper, 2009), is the application of the principles of effective intervention: risk, need, and responsivity (RNR). These principles suggest that (1) intensity of supervision should be matched to the offender's individual level of risk (risk principle), (2) supervision should focus on the primary criminogenic risk factors known to relate to recidivism (need principle), and (3) the manner in which supervision is delivered should be consistent with the offender's learning style and abilities (responsivity principle). Like CCP, adherence to the RNR model can substantially reduce recidivism risk among general offenders. One meta-analysis, for example, found an average reduction of 20–38 % (depending on how many principles were applied) among high-risk offenders (Andrews and Dowden 2006). Additionally, the more time officers spend discussing criminogenic needs in supervision, the greater the reduction in recidivism (Bonta et al. 2008).

Although little is known about how well officers use RNR practices when supervising offenders with SMI, there is some evidence that suggests they may intuitively apply these principles without being specifically trained to do so. In an investigation of a specialty mental health probation agency, Eno Louden et al. (2012) examined audiotaped meetings between 83 probation officer-probationer pairs and coded them for adherence to these principles. Although the most common topic discussed in these meetings was issues related to probationers' general mental health, officers also discussed probationers' criminogenic needs. One-third of meetings (33.7 %) included discussion of criminogenic needs that are most predictive of recidivism (antisocial attitudes, antisocial personality, criminogenic peers), and three-quarters of these meetings (75.9 %) also included discussion of criminogenic needs moderately predictive of recidivism (financial problems, employment, family problems, and substance abuse). Thus, supervision practices in specialty mental health caseloads may already adhere in some respects to important components of RNR.

Implementation

Even though officers may unintentionally employ the evidence-based practices of RNR and interact with supervisees in ways consistent with core correctional practices, greater recidivism reduction for offenders with SMI is more likely achieved through officer training and effective implementation practices. There are several important steps that must be taken to do so effectively. First, researchers must be able to communicate their findings to practitioners in a manner that is not only accessible but also conveys the value of adopting new evidence-based practices. Directors and frontline staff must recognize the short- and long-term benefits of these practices and be committed to ensuring their implementation. Second, agencies need to be provided with guidance in the application of these practices. Without proper training, certain evidence-based practices may be misinterpreted or misapplied, therefore undermining their intended goal of recidivism reduction. Finally, safeguards must be in place to ensure the sustainability of the evidence-based practices. For example, procedures should be in place for training new hires and maintaining quality control in the long term.

Several scholars in the USA and Canada are already making strides in bridging this research-practice gap. Researchers at George Mason University and the University of Cincinnati have been actively involved in training correctional officers in the application and use of RNR and CCP using innovative implementation strategies and rigorous training protocols. Agencies such as Policy Research Associates and the National GAINS Center have also made significant contributions to both research and practice by disseminating their findings on correctional programs for offenders with SMI, in particular. Such programs are essential in translating the research evidence into real-world practice in an accurate, ethical, and efficient manner.

Policy

The research supporting the various routes to supervision failure among offenders with SMI has informed the current policy model recommended by the Council of State Governments for managing this population on probation or parole. This model provides a classification system for offenders with SMI where offenders are classified and managed based on their individual level of criminogenic risk, issues with substance abuse/dependence, and the severity of the individual mental health symptoms and functioning (D'Amora et al. 2012). In accordance with the risk principle, the model recommends that services in each domain be contingent upon severity of problems. For example, high criminogenic risk is matched by more intensive supervision that adheres to the RNR principles. Severe substance problems are met with intensive substance abuse treatment and, if coupled with mental illness, integrated dual diagnosis treatment. Finally, for those with severe mental health symptoms and poor psychosocial functioning, more intensive psychiatric treatment is proscribed. For offenders who are "high" in mental health, substance use, and criminogenic risk, integration of intensive services and supervision is recommended.

Extensive evidence now indicates that there is little utility in emphasizing the role of mental illness in recidivism, except for a small minority of offenders with SMI. As policy and practice shifts away from mental illness as the master status around which correctional interventions and programming are designed, future research should seek to identify new approaches to reducing recidivism for offenders with SMI. Even more, efforts should be made to identify the mechanisms by which (i.e., mediators) and the conditions under which (moderators) current evidence-based practices operate. Doing so can maximize recidivism reduction for this population. Current and forthcoming research examining the effectiveness of practices like specialty caseloads and RNR will continue to help refine the policy model guiding interventions and improve outcomes for offenders with SMI.

Related Entries

- ▶ Effective Supervision Principles for Probation and Parole
- ▶ History of Probation and Parole in the United States
- ▶ Intensive Probation and Parole
- ▶ Probation and Parole Practices
- ▶ Probation Officer Decision-Making

Recommended Reading and References

Abram K, Teplin L (1991) Co-occurring disorders among mentally ill jail detainees: implications for public policy. Am Psychol 46:1036–1045

Andrews D, Dowden C (2006) Risk principle of case classification in correctional treatment: a meta-analytic investigation. Int J Offender Ther Comp Criminol 50:88–100

Andrews D, Kiessling JJ (1980) Program structure and effective correctional practices: a summary of the CaVIC research. In: Ross RR, Gendreau P (eds) Effective correctional treatment. Butterworth, Toronto

Andrews D, Bonta J, Wormith S (2004) The Level of Service/Case Management Inventory (LS/CMI). Multi-Health Systems, Toronto

Andrews D, Bonta J, Wormith S (2006) The recent past and near future of risk and/or need assessment. Crime Delinq 52:7–27

Bonta J, Law M, Hanson C (1998) The prediction of criminal and violent recidivism among mentally disordered offenders: a meta-analysis. Psychol Bull 123:123–142

Bonta J, Rugge T, Scott T, Bourgon G, Yessine A (2008) Exploring the black box of community supervision. J Offender Rehabil 47(3):248–270

Bureau of Justice Statistics (2011) Adult correctional populations, 1980–2009. Retrieved on 17 Feb 2009 from http://www.ojp.usdoj.gov/bjs/glance/corr2.htm

D'Amora D, Jarrett N, Osher F (2012) Adults with behavioral needs under correctional supervision: a shared framework for reducing recidivism and promoting recovery. Council of State Governments Justice Center, New York. Retrieved from http://consensusproject.org/jc_publications/adults-with-behavioral-health-needs/Behavioral_Framework.pdf

Dauphinot L (1996) The efficacy of community correctional supervision for offenders with severe mental illness. Unpublished doctoral dissertation, Department of Psychology, University of Texas at Austin

Dowden C, Andrews D (2004) Importance of staff practice in delivering effective correctional treatment: a meta-analytic review of core correctional practice. Int J Offender Ther Comp Criminol 48:203–213

Eno Louden J, Skeem J (2011) Parolees with mental disorder: toward evidence- based practice. Bulletin of the center for evidence-based corrections, 7(1). Center for Evidence-Based Corrections, Irvine

Eno Louden J, Skeem J (2013) How do officers assess and manage risk for probationers with mental disorder? An experimental investigation. Law and Human Behavior, 37:22034

Eno Louden J, Skeem J, Camp J, Vidal S, Peterson J (2012) Supervision practices in specialty mental health

probation: what happens in officer-probationer meetings? Law and Human Behavior 36:109–119
Fazel S, Danesh J (2002) Serious mental disorder in 23,000 prisoners: a systematic review of 62 surveys. Lancet 359:545–550
Feder L (1991) A comparison of community adjustment of mentally ill offenders with those from the general prison population: an 18-minth follow up. Law Hum Behav 15:477–493
Girard L, Wormith S (2004) The predictive validity of the Level of Service Inventory-Ontario Revision on general and violent recidivism among various offender groups. Criminal Justice & Behavior 31:150–181
James D, Glaze L (2006) Mental health problems of prison and jail inmates. U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics, Washington, DC
Junginger J, Claypoole K, Laygo R, Cristiani A (2006) Effects of serious mental illness and substance abuse on criminal offenses. Psychiatr Serv 57:879–882
Kessler R, Chiu W, Colpe L, Demler O, Merikangas K, Walters E, Wang P (2006) The prevalence and correlates of serious mental illness (SMI) in the National Comorbidity Survey Replication (NSC-R). In: Manderschied R, Berry J (eds) Mental health, United States, 2004. Substance Abuse and Mental Health Service Administration, Rockville, pp 134–148
Klockars C (1972) A theory of probation supervision. J Crim Law Criminol, Police Sci 64:549–557
Manchak S, Skeem J, Vidal S (2007) Care, control, and mental disorder: comparing practices and outcomes in prototypic specialty vs. traditional probation. Poster presented at the first annual North American Correctional and Criminal Justice Psychology Conference, Ottawa, Canada
Manchak S, Skeem J, Vidal S (2010) Boundary spanning: assessing skills in bridging systems for offenders with mental disorder. Paper presented at the American Psychology-Law Society (AP-LS) annual conference, Vancouver, British Columbia, Canada
Paparozzi M, Gendreau P (2005) An intensive supervision program that worked: service delivery, professional orientation, and organizational supportiveness. Prison J 85:445–466
Petersilia J, Turner S (1993) Intensive probation and parole. In: Tonry M (ed) Crime and justice: an annual review. University of Chicago Press, Chicago, pp 281–335
Peterson J, Skeem J, Hart E, Vidal S, Keith F (2010) Analyzing offense patterns as a function of mental illness to test the criminalization hypothesis. Psychiatr Serv 61:1217–1222
Porporino F, Motiuk L (1995) The prison careers of mentally disordered offenders. Int J Law Psychiatry 18:29–44
Prins S, Draper L (2009) Improving outcomes for people with mental illness under community corrections supervision: a guide to research-informed policy and practice. Council of State Governments Justice Center, New York
Quinsey V, Harris G, Rice M, Cormier C (2006) Violent offenders: appraising and managing risk, 2nd edn. American Psychological Association, Washington, DC
Skeem J, Eno Louden J (2006) Toward evidence-based practice e for probationers and parolees mandated to mental health treatment. Psychiatr Serv 57:333–342
Skeem J, Emke-Francis P, Eno Louden J (2006) Probation, mental health, and mandated treatment: a national survey. Crim Justice Behav 33:158–184
Skeem J, Eno Louden J, Polaschek D, Camp J (2007) Assessing relationship quality in mandated treatment: blending care with control. Psychol Assess 19:397–410
Skeem J, Nicholson E, Kregg C (2008) Understanding barriers to re-entry for parolees with mental disorder. In: Kroner D (Chair) Mentally disordered offenders: a special population requiring special attention. Symposium conducted at the meeting of the American Psychology-Law Society, Jacksonville, FL
Skeem J, Kennealy P, Manchak S (2010) Do “firm but fair” relationships mediate the effect of specialty mental health supervision on recidivism? Paper presented at the American Psychology and Law Society (AP-LS) annual conference, Vancouver, British Columbia, Canada
Skeem J, Manchak S, Peterson J (2011) Correctional policy for offenders with mental illness: creating a new paradigm for recidivism reduction. Law Hum Behav 35:110–126
Teplin L (1990) The prevalence of severe mental disorder among urban male jail detainees: comparison with the epidemiologic catchment area program. Am J Public Health 80:663–669

Meta-analysis

► Bullying Prevention: Assessing Existing Meta-Evaluations

► Interview and Interrogation Methods Effects on Confession Accuracy

Methadone

► Drug Abuse and Alcohol Dependence Among Inmates

Methamphetamine

► Drug Abuse and Alcohol Dependence Among Inmates

Methodological Issues in Evaluating Police Performance

Cody W. Telep
School of Criminology and Criminal Justice,
Arizona State University, Phoenix, AZ, USA

Overview

Measuring and evaluating police performance is not easy. Both technical issues, such as designing the best measures, and normative concerns, such as deciding what tasks the police should be performing, are important to consider (Moore and Braga 2004). The focus here is on methodological issues, not normative ones, but one cannot assess the methodological quality of police performance measures without some consideration of what "good" performance entails.

Police performance traditionally has been evaluated with data on crime-related outcomes (e.g., calls for service, incidents, and arrests). "Measurement of the police's impact on crime has long been the *sine qua non* of those endeavors that might be loosely dubbed 'police performance measurement'" (Wycoff and Manning 1983: 15). This does not mean that crime measures are the only relevant performance measures, but crime is logically one important indicator of police performance. The focus below is on the flaws of using aggregate crime data to assess departments and the prospects for using crime data from evaluations as one means to assess performance.

Measures of arrests, clearance rates, and response time are all crime-related outcomes, but they shift the focus to the activities the police are engaged in. None of these measures, however, gives a true representation of how the police are spending their time. Better data on actual activities and resource allocation are needed. Advances in technology offer possibilities for improving departmental knowledge of officer activities, and methods such as systematic social observation (Mastrofski et al. 2010) allow for a rigorous assessment of policing activities.

One component missing from traditional measures is an assessment of citizen views of the police. This seems particularly important in efforts to evaluate the dual police goals of fairness and effectiveness noted by the National Research Council (2004). What can the police do to assess the fairness of their actions? Interviews and surveys of citizens offer police important feedback, although these methods are not without flaws. Police cannot truly evaluate their performance without understanding the demands and viewpoints of their "customers," the citizenry at large, and particularly the citizens they have contact with.

Multicomponent measurement systems help generate as valid a picture of police performance as possible. Quantifying success remains a challenge however, and so proxy measures are often used. Such measures are necessary but problematic (Swindell and Kelly 2000). Measuring police outputs may not necessarily provide reliable data on outcomes. These proxy measures that police can use for assessing performance are discussed below. Methodological issues in using official data on crime are addressed first, followed by issues in understanding what the police are spending their time on, and then a discussion of measuring citizen perceptions. Finally, concluding remarks on the prospects for evaluating police performance are included.

Official Data on Crime

If measures of crime are one piece of the puzzle to evaluating police performance, how can official crime data best be used? One major concern in most current assessments of police performance is the focus on citywide crime data and trends. In some ways, this focus makes sense, because such data are easily available. Over 18,000 police agencies provide data to the Federal Bureau of Investigation (FBI), which then presents annual crime totals by agency for the Uniform Crime Report (UCR).

The problems with the UCR are numerous (see Mosher et al. 2002), but a few important points are worth reviewing. First, the UCR only covers crimes reported to the police, and for crimes other than homicide, a large portion of crime goes unreported. Second, the FBI has specified definitions for each offense type that do not necessarily match up with each jurisdiction's statutory definitions. As an example, the 2010 Annual Report of the Washington, DC Metropolitan Police Department (MPD) presents FBI versus DC Code definitions for a number of offenses to demonstrate that UCR data and DC data do not always match. Such differences are not always minor. The MPD reported 3,538 aggravated assaults to the FBI in 2010, but using the DC code definition for the close to equivalent offense of assault with a deadly weapon, the city had 2,615 offenses. Third, even when focusing on just the most serious Part I crimes, jurisdiction-wide crime rates are dominated by the least serious of these crimes, larceny thefts. These issues are raised here to emphasize the difficulties of using UCR data to assess performance. If police performance measurement is inherently difficult, starting with a flawed data source makes things even more challenging.

While the FBI strongly discourages using the data from the UCR for ranking cities, this does not prevent annual rankings of the "most dangerous" and "safest" cities from being published by media outlets. As the FBI warns on its website, these rankings do not consider the many variables that affect crime in particular jurisdictions and thus can be misleading and incomplete. Ranking cities based on UCR data is like running a regression model on the correlates of crime and only including a single predictor variable. There is far too much variance unexplained in an approach that looks exclusively at crime at the jurisdiction level and ignores other jurisdiction-level characteristics.

Despite the warning from the FBI, agencies are interested in knowing how they compare to their peers. Evaluating performance is difficult without benchmarks, and so comparisons to "similar" departments would ideally be an important part of judging how typical the results in one agency are. Colleges and universities keep a close eye on rankings from *US News and World Report*, and many evaluate their performance in relation to peer institutions. Should police departments be expected to do the same? Not everyone believes ranking cities is inherently problematic. Sage Publications publishes an annual book ranking cities by crime rate. The latest version is *Crime City Rankings 2011–2012: Crime in Metropolitan America*. The International City/County Management Association (ICMA) provides performance measures across 18 areas including policing for over 200 cities through their Center for Performance Management. Cities report information to ICMA, which then compiles a variety of crime and service-related measures for agencies. The organization actively encourages comparisons between peer organizations. What ICMA offers to cities that the FBI does not is an extensive effort to ensure collected data are valid. While the issues in comparing across jurisdictions still exist, one improvement here is the review of the data by an outside organization. Similar arguments have been made by Sherman (1998), who points to the benefits of police agencies using auditors just like private companies to ensure that crime data are reported validly. Alpert and Moore (1993: 122) also argue that "existing measures could be improved to live up to the challenge of professionalism. This would include audited clearance and arrest rates."

In contrast, in the United Kingdom the government is explicitly involved in assessing and ranking police forces. The UK is not entirely comparable to the USA because statutes and crime definitions are consistent across the UK, and the government has national oversight of the police. Still, the work of Her Majesty's Inspectorate of Constabulary is far more comprehensive than anything in the USA at the local or national level. The public has access to comparable data for each of the 43 forces on crime, quality of service, and expenditures. Importantly, the data are checked for quality, and each agency has been assigned peer agencies, which are as similar as possible in population and demographics.

While comparing agencies is one challenge, the major issue with the use of crime data to assess performance is with the unit of analysis. As the unit increases in geographic size and scope, it becomes increasingly difficult to estimate to what extent police practices contribute to changes in observed crime. Linking police performance to crime at the jurisdiction level is most difficult of all. In recent years, scholars have increasingly recognized that the police can have some beneficial impact on crime (see National Research Council [NRC] 2004). This is a change from conventional wisdom even as recently as the early 1990s. Still, even the strongest advocates for the police role in reducing crime do not believe that police have complete control over crime rates. A multitude of factors combine to influence crime, and disentangling how each component influences crime is an impossible task. That is what makes evaluating police performance at the city level so challenging. There is no validated way of estimating how much the police contributed to the crime rate in a city for a particular year versus all the other factors that influence crime.

Such efforts to link police performance to crime rates are not hopeless, however. Importantly, when the scope of the geographic analysis is narrowed, it becomes possible to link police activities to crime declines, and this becomes especially true when smaller geographic units are used in concert with rigorous research designs to test the impact of particular police practices. In many ways, this makes more sense than a citywide assessment. While police are responsible for crime across the entire city, in reality their resources tend to be concentrated in a small number of places, and it is in those places where one would expect police performance to make the most difference.

This point is made by Bayley (1996: 49), who argues that police efforts to prevent crime should be evaluated only when they are targeted on specific sorts of crime in specific locations, “To demonstrate their effectiveness, the police have to target their shots.” Similarly, Wycoff and Manning (1983) note that efforts to improve the measurement of police performance should focus on either improving the methodology of measurement or better conceptualizing what is to be measured. Aggregate official crime data should not be used in comparative studies, because the data across cities are not comparable. They point to a model of evaluating performance that relies on rigorous evaluation and close attention to differences across cities: “The most reliable analyses will be experimental studies conducted on a site-by-site basis...The next most reliable method will be longitudinal studies, again done site-by-site, with particular attention to contextual and organizational factors” (Wycoff and Manning 1983: 19).

Wilson (1993) also argues persuasively about this point, recognizing that a problem in police performance measurement is that there are no adequate performance measures, because it is impossible to precisely estimate what portion of any measure of public safety in a city is attributable to police actions. If the police are unable to adequately assess their performance with citywide crime rates, they often turn to proxy measures, such as response time or clearance rates, which have their own flaws, and are not necessarily related to the crime rate. Focusing attention on specific police actions in specific geographic areas is Wilson’s (1993) solution. He suggests the usefulness of police targeting particular problem neighborhoods (however, the agency defines neighborhood) and then using “microlevel” measures (e.g., changes in calls for service in that neighborhood) to assess the effectiveness of police efforts to address neighborhood problems. As Wilson (1993: 162) argues, “These micro-measures are likely to be among the few valid measures of police performance. They may well lead to conclusions quite at variance with city-wide, aggregate data.”

Randomized experiments are particularly useful in combination with these micro-measures because such designs are assumed to have the highest internal validity and allow for the strongest causal statements about the effects of interventions (see Cook and Campbell 1979). This is the case because subjects are randomized into treatment and control conditions, and thus, their inclusion in one group or another is simply determined by chance. If an agency uses a randomized design to assess an intervention in a specified geographic area, then a crime decline observed in this area (when comparing the treatment area to a randomly allocated control site) can reasonably be attributed to the efforts of police. Randomized experiments of geographic police interventions offer the opportunity for police to persuasively and validly evaluate how their performance influences crime and disorder. Randomized experiments do not solve the problems with official data and the UCR, such as the vast underreporting of most crimes, but experiments offer a way to more rigorously assess how police performance affects crime.

Hot spots policing is a particularly good example of a policing intervention that takes place in a small unit of geography and which has been rigorously evaluated with a series of randomized trials (NRC 2004). A large body of research suggests that police can have a significant beneficial impact on crime and disorder when they narrow their focus to microgeographic units with high rates of crime (see Weisburd and Eck 2004). Importantly, work examining hot spots of crime in Seattle also suggests that citywide crime trends tend to mask the trends at particular street blocks in the city (Weisburd et al. 2012). While Seattle enjoyed a citywide crime decline in the 1990s, this decline was isolated to a small percentage of blocks, while the vast majority of blocks showed little change and some even showed substantial crime increases. This reinforces the problem of using citywide crime data to assess police performance.

Compstat provides another good example of how crime data can be used as a means of evaluating police performance. Compstat forces police commanders to focus in on crime problems occurring in short-time periods and assess how police actions can address these problems quickly and efficiently. Compstat was an effective approach for evaluating and altering performance in the New York Police Department. Moore and Braga (2003: 446) note that in Compstat, “the measures are simple, objective, reliably measured, and continuous so that changes in performance can be observed over time within an operational unit, and across units that are roughly similar.” While Compstat does not make use of randomized experiments in its assessment of performance, it does make important strides in generating measures that are consistently measured over time, allowing for comparisons within and across units and making it possible to more closely link police actions in a specified time period to changes in crime.

What Are the Police Spending Their Time on?

The police rely heavily on crime measures to assess their performance, but these are not the only measures traditionally used. Police also are interested in understanding what officers are doing with their time, and hence, they often track measures of police activity such as response time to 911 calls, arrests made (and citations written), and clearance rates. These measures, however, provide only limited data on what officers are doing with their time, and, if crime control remains a primary measure of police performance, research suggests that rapid response and general increases in arrests are not effective ways to address crime (see Weisburd and Eck 2004). Additionally, these measures give insufficient attention to police resource allocation. Systematic social observation research (see below) suggests that only a very small portion of officer time is spent arresting offenders and a fairly small proportion of time is spent responding to calls for service. Departments are not adequately measuring performance by focusing on these two measures alone.

Clearance rates can also be a problematic measure. The FBI reporting rules allow for situations in which data are presented in a way that is technically correct but misleading. For example, the MPD widely touted a 94 % homicide clearance rate in 2011. One would likely assume that DC police made an arrest in 94 % of 2011 murder cases. The FBI, however, counts homicide closures that occurred in any year in the current year's clearance rate. DC homicide detectives closed 62 of the 108 homicides in 2011 and also closed about 40 homicides that occurred in earlier years. While the 94 % homicide clearance rate is accurate under FBI reporting rules, it would seem to make more sense to report on the percentage of homicide cases closed within 12 months, rather than report a percentage that is not entirely related to current year data. Additionally, while clearance rates have been on the decline in recent decades, the factors associated with these declines may be outside police control. Ousey and Lee (2010) find that changes in police personnel and workload are not associated with changes in the clearance rate, but that changes in immigration can decrease clearance rates. Like crime rates, certain factors that police have no control over may contribute to decreased clearance rates.

Number of arrests and police response time are both measures, to some extent, of police productivity. These measures, however, give police little information about what officers are spending their time on day to day. How are departmental resources being used, and is resource allocation effective and efficient? These are also important questions for police and present challenges for measurement. Unlike profits in the private sector, there is no single bottom line in policing, and so assessing whether resources are being properly used is not easy.

Ostrom (1983) makes this clear in her discussion of equity in the delivery of police services. What should be the standard for equity and how should this standard be measured? Should services be distributed based on taxes paid, or based on need, or to all citizens equally, or to ensure that all citizens get equal results? These are normative questions, but the effort to assess equity presents a host of methodological problems. The service that police are providing is ambiguous, making it difficult to measure whether service is being distributed equitably. Expenditure data can be used for assessing level of service, although budgeted resources do not necessarily correspond to services delivered to particular communities. Ostrom (1983) points to four flawed ways researchers typically assess equity in policing. Data on the distribution of police across jurisdictions typically shows that higher crime areas receive more police (i.e., the needier areas get more), but this says little about what police are actually doing in these areas. As Ostrom (1983: 110) notes, "The difference between an occupying army and a service-oriented police force cannot be determined solely by examining personnel allocation." Similarly, response time data and crime data do not provide much information on the actual delivery of police services in these areas. A final less commonly used method is citizen surveys to assess fear of crime, which are potentially useful, but survey data do not necessarily provide insights into why citizens feel the way they do (see below).

A recent study in Dallas (Weisburd et al. 2013) was the first randomized experiment to make use of data from automatic vehicle locators (AVL). AVL technology tracks the exact location of a patrol car repeatedly. This has obvious benefits for officer safety in terms of monitoring patrol cars, but it also has benefits for assessing where officers are going during their shift. In hot spots policing studies, saturation patrol is often an important component of the intervention, and so AVL technology could allow the department to assess how often officers are visiting hot spots and whether treatment is being delivered as intended. While police unions are reluctant to allow departments to closely monitor AVL data, there is potential here to better identify officers' spatial location during their shift.

To fully evaluate performance, police agencies must know not only where officers are but what they are doing. While agencies have increasingly moved towards a greater embrace of problem solving and working with the community, there is little or no official departmental data on whether officers are actually engaging with the community or using proactive problem solving (Alpert and Moore 1993). One option is to better survey officers about their activities while on duty. Most police agencies currently keep track of how long officers spend on particular calls for service, but have much less data on what officers are doing in their uncommitted time. Dadds and Scheide (2000) discuss the use of activity measurement in Australia as a way to better monitor what officers are doing in the field and how these activities may be linked to desired outcomes. Officers spend 1 or 2 weeks annually filling out a survey on specifically how they are spending their time on each shift. This method relies on the veracity of officer responses, and it can be difficult to place officer activities into specific categories. Still, the survey provides an opportunity to estimate how officers' time is divided between traffic services, community police services, crime management, emergency response management and coordination, and criminal justice support.

Similarly, Mastrofski (1983) calls for greater police attention to noncrime services provided to citizens. While incident reports and arrests are well recorded, police devote a substantial portion of their time to providing information or assistance, dealing with nuisances, or breaking up disputes, and unless a report is written, none of this noncrime service is well recorded. Providing officers with a set of response activity codes that could be used to report what actions were taken in a particular citizen encounter would be useful and would provide more data on how police were allocating their noncrime services.

Mastrofski (1996) provides five different methods for evaluating what officers are doing in interactions with citizens. The first, as described above, is officer self-reports of behavior. While inexpensive, this method suffers from potential problems in reporting validity. The reports of citizen who have interactions with the police are another possibility and are discussed more below. A third possibility is using evaluations by other human service professionals who regularly interact with the police. While such evaluations would be less subjective than self-report data, such professionals would not be present during a large proportion of citizen interactions. The last two methods offer the most objective means of evaluating police performance: the use of indirect third-party observation (e.g., video recording) and the use of direct third-party observation. Direct third-party observation is discussed below as a means for systematic social observation. The use of police video recordings has been fairly uncommon to date, although as more departments record police-citizen interactions increases and as technology improves, this would seem to be a promising and inexpensive route for future research on police performance.

Systematic social observation (SSO) offers an even better way to assess what activities officers are engaged in and provides a more objective measure of police-citizen interactions. SSO makes use of protocols in field observations of the police to ensure that data collection is standardized within and across observers. As Mastrofski et al. (2010: 243–244) note, "it offers enhanced prospects of validity, and in many situations it provides for increased confidence in reliability, because of the researcher's direct access to the phenomenon of interest and greater control and transparency of data encoding. Further, it affords greater precision in capturing details of the phenomenon and its context." SSO is not without potential threats to validity. A primary one is the possibility of observer error in recalling events. While multiple observers of the same scene would be ideal to avoid this, it becomes difficult both in terms of cost and logistics to use more than one observer per officer. Another concern is reactivity effects. Are officers responding differently when observers are watching their actions? Research on this topic is limited, but in a long-term project where officers are being observed repeatedly, it seems less likely that reactivity would be a major problem. Systematic social observation has been an important component of a small but growing number of research studies on police behavior. Famega et al. (2005) used systematic social observation in Baltimore to assess both how officer time was allocated and the reason for officer actions over the course of 163 shifts. Their finding that 75 % of officer time is unassigned suggests that simply examining calls for service data is insufficient for understanding what officers are doing during their shifts.

An important question with data on police activities and interactions with citizens is how these data can be used to evaluate police performance. Even with the problems with crime data, it seems clear what the desired outcome is – overall crime reduction. When evaluating what officers are spending their time on and how they interact with citizens, the outcomes become fuzzier. In terms of what officers spend their time on, if the goal is to relate these activities to crime control effectiveness, then it is easier to assess whether or not officers are performing in ways that are evidence based. But what about evaluating how an officer handles a particular domestic violence call, or how officers carry out a traffic stop? One part of evaluating performance here would be assessing levels of citizen satisfaction and beliefs about the fairness of the interaction, an issue discussed in the next section.

But police, like other professionals, do not want their performance assessment based only on the views of outsiders. Mastrofski (1996) suggests that police master craftsman, as chosen by surveys of fellow officers, could work to develop performance criteria for difficult or common police problems. Then, observed officer actions could be compared to these criteria. The development of such criteria would not be easy and would take much deliberation among highly regarded officers, but could be quite useful for helping to evaluate what good performance should look like. Recent research by Mastrofski et al. (2011) suggests both that officers can identify master craftsman (i.e., officers whose performance and opinion they respect) and that officers can be interviewed about evaluating police performance in particular situations. While officer views are not always consistent on what constitutes good performance, this is a promising area for future research on how police can use observational data to evaluate police performance based on agreed-upon departmental standards.

Measuring Citizen Perceptions

While innovations such as community policing have stressed the need for police to partner and work closely with community members to effectively address crime and fear of crime, the police have not been quick to add measures that will help gauge the community's view of the department (Alpert and Moore 1993; Moore et al. 2002). As Mastrofski (1996: 210) notes in arguing for more attention to be placed on police-citizen encounters, "Crime rates, fear levels and citizen satisfaction – though keenly felt at the personal level – are hard to observe directly. . .It is not at all clear that the public holds the police accountable for the crime rate, but they do hold them accountable for their actions." This makes assessing how citizens view officer actions an important aspect of evaluating police performance. While departments have traditionally collected data on citizen complaints, such data are not sufficient for evaluating performance (Moore et al. 2002). While complaint data are an important way for departments to learn about and remedy improper police behavior, complaint data do not give a full picture of police-citizen encounters, and importantly for efforts to evaluate "good" performance, complaints provide no data on positive police-citizen interactions.

Data on citizen perceptions can be gathered in multiple ways. Some police departments, regularly survey 911 callers in satisfaction surveys. A number of researchers have used telephone surveys to gather information on recent citizen encounters with the police and how citizens viewed such encounters, or on citizens' perceptions of their neighborhood and/or recent victimization experiences (see Skogan 1999). Rosenbuam et al. (2008) were successful in assessing citizen views of the police through an Internet survey in Chicago. In addition to large-scale surveys, qualitative interviews can also be an important way to gather detailed information on police-citizen encounters.

Brown and Benedict (2002) point to some methodological issues that can arise in research on citizen perceptions of the police. Broad-based citizen surveys may not capture a sufficient number of witnesses, victims, and suspects and thus may oversample individuals who have not had direct contact with the police. Even when researchers have access to lists of individuals who have had recent police contact, it is not always easy to track these people down and to acquire their consent to participate in research. Such individuals, particularly suspects, are often skeptical of any research related to the police and are reluctant to believe that confidentiality measures are in place. Victims or witnesses may not feel comfortable discussing a sensitive matter, particularly for violent crimes. It may be especially difficult to survey young minority males, a group of particular interest in assessing police performance, because of their typically high levels of contact with the police and low levels of perceived police legitimacy. Wolfer and Baker (2000) suggest the additional problem of "halo effects" in survey research, especially in small and rural agencies. Citizens in these areas may rate the police as satisfactory based on personal relationships and not on actual good performance. In their study of problem-oriented policing, citizens tended to rate the police highly in areas where other objective measures suggested the police were not doing a particularly good job.

Brown and Benedict (2002: 562) also stress the benefits of efforts to "develop objective, independent measures of police activity to better determine the impact of officers' behaviors on attitudes towards the police." Efforts to combine SSO and citizen perception data could be particularly beneficial here. While this raises concerns about citizen privacy, an ideal research design would be to use SSO data to objectively assess police-citizen encounters and then conduct follow-up interviews with these citizens to better understand their views of the encounter and the police. Such a design would help better link citizen attitudes and officer performance.

While large-scale survey research has the benefit of generating larger sample sizes, Shilston (2008) stresses the importance of using qualitative data to assess police performance. General satisfaction surveys will likely be unable to provide detailed depictions of citizen views of and attitudes towards the police. Such surveys can be useful as a quantitative summary of perceptions, but they fail to provide insight into the individual experiences of those who interacted with the police. Interviewing those who had direct contact with the police is a better method for gaining in-depth information on citizen perceptions.

Fielding and Innes (2006: 135) also call for a more qualitative approach to assess public perceptions of the police and argue that "A better approach than using randomly selected respondents to measure change may be to employ judgements made by key informants: individuals who have detailed knowledge of communal life in a locale and are in a position to provide a meaningful assessment of how policing there has improved or worsened." This purposive sampling technique may provide a richer data source for police. They also point to the benefits of qualitative assessments of the narratives from complaint data to get a more descriptive view of negative police-citizen encounters. This could provide greater insight for police departments than a simple count of the number of complaints.

Conclusions: What Should Police and Researchers Be Measuring?

In efforts to evaluate police performance, it is important for police and researchers to use multiple methods of data collection to create a fuller portrait of police performance. Cordner (1996: 200) sums this up well by noting "The bottom line is that policing must be judged by *bottom lines*." Police should continue to collect data on the traditional measures above, but should be cognizant of the problems with these measures. They should also expand the list of measures to include variables that reflect the importance of police activities and interactions with citizens and the views of citizens, particularly in terms of the fairness of police actions. For example, official data on drug calls for service and drug arrests are problematic, because they do not always match up with drug market boundaries. Official data should be supplemented with resident surveys to assess the severity and visibility of problems, as well as data sources from other agencies, such as the location of 911 calls for drug overdoses. Just as multiple data sources are a better way to identify drug markets, they would also be important to assess whether police are having an impact on drug markets. Arrest patterns reflect more about police enforcement priorities than whether police reduced harm from particular drug markets. But the triangulation of official data, health data, and resident data could help indicate whether police actions were having impacts on citizen safety and well-being. Skogan (1999) also points to the benefits of triangulation in his review of methodological challenges in using both survey and official crime data. When victimization data, official crime data, and citizen perception data are all pointing in the same direction, then evaluating police performance can be done with greater confidence.

Moore et al. (2002) provides a series of goals for police and measures that would be useful to include in what he describes as a "model annual report." These include not only traditional measures, in efforts to reduce criminal victimization and call offenders to account, but also measures explicitly focused on the views of citizens in efforts to reduce fear of crime and satisfy customer demands and achieve legitimacy with those policed. Such efforts to view citizens as the customers of police fit in well with the private sector notion of total quality management (TQM). Hoover (1996) describes how TQM focuses on a number of aspects relevant for evaluating police performance, including using routine and constant measurement to enhance performance, focusing on the concerns of customers, and being explicitly concerned with counting and quantifying interactions with customers and the outcomes of these interactions. Management strategies from the private sector do not transfer perfectly to policing. "Good" policing has no clear definition, and so success is not as easy as raising profits. Hoover (1996: 19) points out that "Not all qualitative elements of police work can be quantified. There are too many exigencies, contingencies and intangibles." Nonetheless, efforts to better measure various aspects of police performance are worthwhile and important to more fully evaluating police fairness and effectiveness.

One final issue is whether police performance measurement should be focused on outputs or outcomes. Stephens (1996: 121) points to the need to supplement standard measurements of police performance with measures of outcomes and notes, "The issue of measuring effectiveness is not entirely one of discarding the old and replacing it with new measures. It is one of putting the traditional measures in perspective (using them when appropriate) and placing greater emphasis on outcomes than process or 'proxy measures.'" While it seems reasonable to focus on the outcomes of policing, Moore et al. (2002) argues that while outcomes are often seen as more important, one cannot ignore the outputs (and inputs) of policing. Outcome measures can be challenging to collect, and in policing, the concern is not just about outcomes but also how police are doing their job. While reducing crime and disorder is a desired outcome that should be assessed as well as possible, it is equally important that the police are being fair in their use of force, their allocation of resources, and their interactions with citizens.

Starbuck (2004: 337) summarizes the problems and prospects of evaluating performance in any agency: "Performance measures are everywhere, but they are filled with errors, and these errors are likely to cause faulty inferences. We should distrust performance measures, but we cannot ignore them because they are powerful motivators that can produce dramatic improvements in human and organizational performance." While evaluating police performance entails a number of methodological challenges, measuring performance is an important way to both bring about desired changes in police organizations (e.g., Compstat) and assess whether police performance fits in with notions of what "good" policing should entail.

Related Entries

- ▶ Cross-National Performance in Policing
- ▶ Measuring Police Performance
- ▶ Measuring Police Unit Performance
- ▶ Police Performance Measurement

Recommended Reading and References

Alpert G, Moore M (1993) Measuring police performance in the new paradigm of policing. In: DiIulio JJ et al (eds) Performance measures for the criminal justice system. Bureau of Justice Statistics, U.S. Department of Justice, Washington, DC, pp 109–142

Bayley DH (1996) Measuring overall effectiveness. In: Hoover LT (ed) Quantifying quality in policing. Police Executive Research Forum, Washington, DC, pp 37–54

Brown B, Benedict W (2002) Perceptions of the police: past findings, methodological issues, conceptual issues and policy implications. Policing: Int J Police Strategies Manag 25:543–580

Cook TD, Campbell D (1979) Quasi-experimentation: design and analysis issues for field settings. Rand McNally, Chicago

Cordner GW (1996) Evaluating tactical patrol. In: Hoover LT (ed) Quantifying quality in policing. Police Executive Research Forum, Washington, DC, pp 185–206

Dadds V, Scheide T (2000) Police performance and activity measurement, vol 180, Trends and issues in crime and criminal justice. Australia Institute of Criminology, Canberra

Famega CN, Frank J, Mazerolle LG (2005) Managing police patrol time: the role of supervisors. Justice Quart 22:540–559

Fielding N, Innes M (2006) Reassurance policing, community policing and measuring police performance. Policing Soc 16:127–145

Hoover LT (1996) Translating total quality management from the private sector to policing. In: Hoover LT (ed) Quantifying quality in policing. Police Executive Research Forum, Washington, DC, pp 1–22

Mastrofski SD (1983) The police and noncrime services. In: Whitaker GP, Phillips CD (eds) Evaluating performance of criminal justice agencies. Sage, Beverly Hills, CA, pp 33–61

Mastrofski SD (1996) Measuring police performance in public encounters. In: Hoover LT (ed) Quantifying quality in policing. Police Executive Research Forum, Washington, DC, pp 207–241

Mastrofski SD, Parks RB, McCluskey JD (2010) Systematic social observation in criminology. In: Piquero AR, Weisburd D (eds) Handbook of quantitative criminology. Springer, New York, pp 225–247

Mastrofski SD, Willis JJ, Revier L (2011) How police distinguish quality in police work. Paper presented at the Annual Meeting of the American Society of Criminology, Washington, DC

Moore MH, Braga AA (2003) Measuring and improving police performance: the lessons of Compstat and its progeny. Policing: Int J Police Strategies Manag 26:439–453

Moore MH, Braga AA (2004) Police performance measurement: a normative framework. Crim Just Ethics 23:3–19

Moore M, Thacher D, Dodge A, Moore T (2002) Recognizing value in policing: the challenge of measuring police performance. Police Executive Research Forum, Washington, DC

Mosher CJ, Miethe TD, Phillips DM (2002) The mismeasure of crime. Sage, Thousand Oaks

National Research Council (2004) Fairness and effectiveness in policing: the evidence. Committee to review research on police policy and practices. In: Skogan W, Frydl K (eds) Committee on law and justice, division of behavioral and social sciences and education. National Academies Press, Washington, DC

Ostrom E (1983) Equity in police services. In: Whitaker GP, Phillips CD (eds) Evaluating performance of criminal justice agencies. Sage, Beverly Hills, pp 99–125

Ousey GC, Lee MR (2010) To know the unknown: the decline in homicide clearance rates 1980–2010. Crim Just Rev 35:141–158

Rosenbuam D, Schuck A, Graziano L, Stephens C (2008) Measuring police and community performance using web-based surveys: findings from the Chicago internet project. National Institute of Justice, U.S. Department of Justice, Washington, DC

Sherman LW (1998) Evidence-based policing. Ideas in American policing. Police Foundation, Washington, DC

Shilston TG (2008) One, two, three, what are we still counting for? Police performance regimes, public perceptions of service delivery and the failure of quantitative measurement. Policing 2:359–366

Skogan WG (1999) Measuring what matters: crime, disorder, and fear. In: Langworthy RH (ed) Measuring what matters: Proceedings from the police research institute meetings, pp 37–54. Washington, DC: National Institute of Justice and Office of Community Oriented Policing Services, U.S. Department of Justice

Starbuck WH (2004) Methodological challenges posed by measures of police performance. J Manag Gov 8:337–343

Stephens D (1996) Community problem-oriented policing: measuring impacts. In: Hoover LT (ed) Quantifying quality in policing. Police Executive Research Forum, Washington, DC, pp 95–129

Swindell D, Kelly JM (2000) Linking citizen satisfaction data to performance measures:

a preliminary examination. Public Perform Manag Rev 24:30–52
Weisburd D, Eck JE (2004) What can the police do to reduce crime, disorder, and fear? Ann Am Acad Pol Soc Sci 593:42–65
Weisburd D, Groff ER, Yang SM (2012) The criminology of place: street segments and our understanding of the crime problem. Oxford University Press, New York
Weisburd D, Groff E, Jones G, Amendola KL, Cave B (2013) The Dallas AVL experiment: evaluating the use of automated vehicle locator technologies in policing. Police Foundation, Washington, DC
Wilson JQ (1993) The problem of defining agency success. In: DiIulio JJ et al (eds) Performance measures for the criminal justice system. Bureau of Justice Statistics, U.S. Department of Justice, Washington, DC, pp 156–164
Wolfer L, Baker TE (2000) Evaluating small town policing: methodological issues. J Police Crim Psychol 15:52–63
Wycoff MA, Manning PK (1983) The police and crime control. In: Whitaker GP, Phillips CD (eds) Evaluating performance of criminal justice agencies. Sage, Beverly Hills, pp 15–32

Microsimulation Models of the Criminal Justice System

Anna Stewart
School of Criminology and Criminal Justice,
Griffith University, Brisbane, Australia

Overview

This entry provides an introduction to the use of microsimulation models for informing decision making by criminal justice system professionals. Microsimulation models are placed within a framework for understanding computer simulation models. A brief history of the use of simulation models and microsimulation models in the criminal justice system is provided. The processes involved in the development of the Queensland Juvenile Justice Simulation model are described to illustrate the steps required to develop, validate, and use a microsimulation model. Finally, some of the challenges and advantages to using microsimulation models in the criminal justice system are explored.

Overview of Social Science Microsimulation Modelling

Within the social sciences three broad approaches to computer simulation models have been identified: macrosimulation approach, microsimulation approach, and agent-based modelling approach (Macy and Willer 2002). The development of macrosimulation models (also known as system dynamics models, world models, stock and flow models, dynamic systems simulation modelling) coincided with the first use of computers in university research in the 1960s. These models have their roots in systems of difference and differential equations and are used for understanding the dynamic nature of complex nonlinear systems. The second modelling approach, microsimulation (also known as discrete event simulation modelling), simulates a “macro” system by modelling and simulating the actions and interactions of the systems “micro” units (usually individuals) (Spielauer 2011). The results of the effect of system changes on each unit can then be analyzed at the microlevel or aggregated up to show the overall “macro” effect of change. This type of modelling has its roots in economics, engineering, and operations research. The final approach to simulation modelling, agent-based modelling, was developed during the 1980s coinciding with the advent of the personal computer (Macy and Willer 2002). This type of modelling has its roots in artificial intelligence and computer science. Unlike macro- and microsimulation the focus of agent-based modelling is on theoretical development and explanation rather than applied research (Gilbert and Troitzsch 2005).

Social science microsimulation enables the experimentation with a virtual society of thousands of individuals created by a computer. These models can either be static (examine the short-term effect of a policy change) or dynamic (follow individuals over the life course). In static models time has no effect on the individuals modelled. However, dynamic simulation models explicitly include time. For dynamic microsimulation models characteristics associated with the individual units change in response to accumulated experience or the passage of time.

Consequently these models enable us to simulate future conditions and project the outcomes of a proposed policy or operational change. By understanding the likely consequences of a proposed change, it is possible to evaluate if the policy change would be beneficial, contributing to the process of policy analysis and implementation. Dynamic microsimulation models allow virtual individuals to be followed over their entire life course adding a whole new dimension to policy analysis.

There are two steps to social science simulation: first, building the computer model and the second is to "play" or "experiment" with the model. To build a microsimulation model, detailed data are required on the individuals in the system. Individuals are then modelled with a number of key attributes (e.g., sex, age, race) and a number of transition probabilities that reflect changes over time. Once the models are built, they are run (multiple times) to establish a baseline scenario. These baseline scenarios reflect expectations of how current conditions will impact future populations, assuming the policies and the system remain constant and the only changes are the underlying demographics.

Simulation scenarios ask the "what if" questions. They are like mini experiments that identify the downstream impact on the system of a proposed change if everything else was held constant. Of course, systems are extremely dynamic and models do not and cannot predict the future. Rather, models provide predictions on the basis of past trends and take into account what is known about a particular system. As such, policy simulation modelling provides decision makers with additional information that would assist them in making rational decisions on the optimal use of scarce resources and improve the accountability of the criminal justice system.

Microsimulation provides a tool for examining the impact of the behavior of individuals on the whole system. The social science literature is rich with statistical micro data. However, most research stops after the estimation of individual processes. As Spielauer (2011) stated, "With a microsimulation model, we go one step further: microsimulation adds synthesis to analysis" (p. 11). Spielauer (2011) identifies three circumstances under which microsimulation is the appropriate simulation approach, when population heterogeneity is important, when behaviors are complex at the macro level but better understood at the microlevel, and when individual histories matter. These circumstances make microsimulation modelling an appropriate modelling technique for policy analysis in the criminal justice system.

History of Simulation Modelling of the Criminal Justice System

Microsimulation is widely used in the social sciences for policy analysis. Policy-orientated microsimulation modelling is commonly used in areas such as pensions, tax policy, income and poverty issues, and health policy. Within the criminal justice system microsimulation is not currently widely used; however, this modelling approach was quite prominent in the criminal justice system in the late 1960s (Auerhahn 2008a).

The pioneer of criminal justice system simulation modelling is Alfred Blumstein. Over 40 years ago, he, in his role as Director of President Johnson's Commission on Law Enforcement and Administration of Justice, introduced techniques from operations research, quantitative modelling, and system perspective and planning to criminal justice (Blumstein 2007). This work formed the basis for the development of the seminal computer-based criminal justice simulation model – the Justice System Interactive Model (JUSSIM) (Belkin et al. 1972). JUSSIM made it possible for the first time to examine how actions in one part of the criminal justice system affected other parts of the system pioneering the use of policy analysis tools in the formulation of criminal justice policy (Nagin 2012). This model represented the criminal justice system as a sequence of stages, starting with crime, leading to arrest and leading through the stages of the criminal justice system, ending with parole and release from parole. A major limitation of the model was the inability to incorporate feedback (recidivism) into the model. This limitation was addressed in the second version of the model JUSSIM2 (Belkin et al. 1973). This model and its variants are still in use in the USA.

Following on from this work simulation models were developed in the UK (Morgan 1985a, b) and in Australia (Crettenden et al. 1983). These early representations of the operation of the criminal justice system did so in great detail allowing for a greater range of policy simulations to be conducted. However, this increase in detail resulted in a more complex model with more assumptions included and more parameters to be estimated. This made these models difficult, time consuming, and expensive to update and maintain, and mostly these models have disappeared (see review Stewart et al. 2004).

One of the most significant findings of the early work on understanding and modelling the criminal justice system was that people in the criminal justice system are likely to end up back in the criminal justice system – the problem of recidivism. This initiated a productive line of research by Blumstein and colleagues examining the criminal careers of offenders. Criminal careers are the characterization of longitudinal patterns of offending and the distribution of these over population subgroups (Blumstein 2002). This research is still very active today (Piquero et al. 2003), and the findings have considerable policy implications for the control of crime including prevention, career modification (deterrence), and incapacitation. Furthermore, this work is the forerunner to developmental and life-course criminology (Farrington 2003), currently the most influential theoretical paradigm in criminology (Cullen 2011).

In recent years there has been a resurgence in criminal justice simulation models for policy analysis. Much of this work is carried out in house and is not published in the academic literature. Examples of work that has been published included Zarkin and colleagues (Zarkin et al. 2005, 2012) who utilized microsimulation to demonstrate the lifetime costs and benefits to society of improving prison-based post release substance abuse programs. Auerhahn (2002, 2007, 2008b) used system dynamic modelling to examine sentencing reforms, the increase of the length of stay, and the impact of aging prison populations on future prison populations. The New South Wales Bureau of Crime Statistics built a stock and flow model of the criminal justice system (Clark and Lind 2003; Lind et al. 2001) and used this model to simulate flows through the New South Wales criminal justice system.

The construction and use of the Queensland Juvenile Justice System Model is described in the next section. This description will provide the reader with a brief introduction to steps involved in building a model. There is very little publicly available documentation on the building of criminal justice microsimulation models. Following this description two published examples of the use of this model for policy analyses are briefly described. The difficulties in maintaining and using the model will then be briefly outlined.

Queensland Juvenile Justice System Model

An example of a criminal justice system microsimulation model is the Queensland Juvenile Justice System (QJJSM). The development of this model is fully explained in a technical report (Stewart et al. 2004). This model was developed in collaboration with the Office of Economic and Statistical Research, Queensland Treasury, to provide a tool for policy makers and legislators to estimate the *relative* impact of potential changes to the system in the medium term. The model was designed to assess "what if" questions, with the underlying assumption that, apart from the system change being modelled, the only changes to the youth justice system relate to demographic changes. Therefore, the focus of the model was on a comparison over time between the baseline situation (i.e., the current system) and the proposed change to the system.

This model is a parsimonious model that simulates the initiation of new offenders, the commission of specific offences, the decision of the youth court, and reoffending behavior (Fig. 1). The model is a microsimulation model, that is,

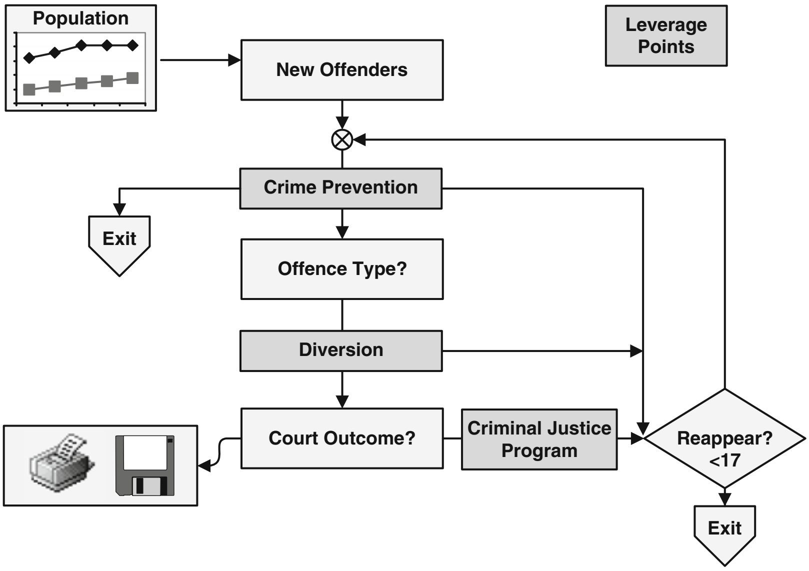

Microsimulation Models of the Criminal Justice System, Fig. 1 Diagrammatic description of the Queensland Juvenile Justice Simulation Model including leverage points

individual offenders are simulated. This allows for the examination of both the way individuals move though the system and the number of young people at various points in the system (e.g., detention). The outputs of this model are individual-based data that can be summarized and analyzed using standard statistical techniques to examine and explore the medium-term impacts of the intervention programs on the number of young people flowing through the system.

There are a number of leverage points included in the model. These leverage points represent components of the youth justice model where the implementation of a program, policy, or legislative change may result in the reduction of offending. These leverage points were designed to allow the user to add crime prevention, diversion, and post-court intervention programs to the base system. This enabled the comparison of the impact of these interventions with the baseline system.

Keeping the model simple resulted in three benefits. First, the model captured the critical components of both the system (court outcomes and supervision) and individual offender behavior (initiation, desistance, and reoffending) required. Second, the model's simple structure made easy to explain to policy makers thereby avoiding alienating the model's user base through over complexity. Third, the structure of the final schema ensued that the data required to parameterize the mode were readily available in the administrative datasets maintained by Queensland government.

The model was developed in the proprietary microsimulation package Extend. The logic for choosing a proprietary package and this particular package is provided in the technical report (Stewart et al. 2004). Prior to building the model, the administrative data collected by Queensland government to manage the operation of the youth justice system were statistically examined to develop offending rates, offence probabilities, sentencing models, and models of reoffending and desistence. Separate analyses were conducted for each stage of the system and included a series of logistic regressions for the court decisions and survival analyses for reoffending. The details of these analyses are presented in Stewart et al. (2004).

In the model new offenders enter the system from the general population. Based on the analyses of the administrative data, these offenders are probabilistically assigned demographic characteristics (age, sex, Indigenous status, and geographic region). The offence types and court outcomes were reduced to categories that meaningfully reflected statistical differences in behavior and outcomes among the demographic groups. Eight offence types were modelled. These included offences against the person, break enter and burglary, theft and related offences, drug offences, traffic offences, public order offences, property damage, and other offences. Analyses indicated that gender and Indigenous status determined the offending profiles for young offenders and there was no evidence of offence specialization. That is, regardless of the number of previous offences, the probability of a particular offences category remains constant. Consequently, the model assigns probability parameter to each demographic group, corresponding to the offending profiles identified in the data.

To determine the court outcomes (divert from formal order, non-supervised order, community supervised order, detention order suspended (immediate release order), and detention order served), a series of logistic regressions were performed to examine a range of predictive variables (number of prior appearances, Indigenous status, gender, offence type, previous detention, number of previous detention orders, number of offences finalized at appearance, total number of prior finalized offences). These analyses indicated that two variables accounted for the majority of the variation in sentencing (offence type and the number of prior finalized appearances). Probability tables based on these two variables (and gender, due to the simplicity of including it and its significance in some, but not all, of the logistic models) were created and used in the model to assign court outcomes to offenders.

The final, and arguably the most critical parameter, to be modelled was the reappearance of offenders. Young offenders re-appeared in the model if they reoffended before they turned 17 years. The analyses revealed that 65 % of court appearances were by young offender reappearing in the system. Consequently, the reappearance parameters were critical to ensure the model replicated the system. The number of times an individual reappears in the model also determines the number of prior appearances assigned to that individual. This variable was identified as critical in predicting sentencing outcomes. Survival analyses indicated that there were significant differences in the time to reappear between the four demographic groups (gender by Indigenous status). These differences were apparent in the time from first to second appearance, second to third appearance, and third to fourth appearance. After the fourth appearance the survival curves for the four groups were converging (and the numbers in each group were small). Two population Weibull distributions (with a desistance term) were fitted to the 16 categories (gender, indigenous status, and appearance number (1,2,3,4+)) to provide reappearance parameters for the model.

The model was validated by comparing the simulation results for court outcomes for the financial year 2002/2003 with the actual court outcomes for that year. These results were aggregated by court outcome and Indigenous status of the young offender. Only the detention outcomes for both Indigenous and non-Indigenous offenders fell outside the 95 % upper and lower bounds of the simulated values. The simulation model underestimated the number of young people sentenced to detention orders. These numbers are small, show high variability, and therefore are difficult to estimate. Consequently, these underestimations do not invalidate the model. However, the detention numbers produced by the model need to be treated with caution.

Policy Analysis Using the QJJSM

The primary purpose of the QJJSM is the analysis of proposed changes to policy in the youth justice system. The model cannot predict exactly what is going to happen in the future, as there are too many influential factors that cannot be modelled accurately. Instead, the model provides a baseline set of data assuming the currently system behavior will remain stable over the time period modelled, with only the underlying demographics changing (Baseline Scenario). These demographics are obtained from the projected population estimates, which indicate differential population growth between Indigenous and non-Indigenous young people. This baseline model provides a set of standard output that can be used for comparison with the proposed system changes (Policy Scenario).

Once the baseline results have been recorded, the user can include one or more prospective programs at the leverage points. The model is then rerun with the proposed programs included and the relative reduction in the flow of young people through the youth justice system can be examined. Due to the nature of microsimulation models, the resulting data are individual based and consequently can be analyzed using conventional statistical techniques to examine the impact of the proposed intervention both on the system (e.g., the reduction in court appearances or detention orders) or individuals (e.g., the reduction in Indigenous offenders). Furthermore, the model allows some simple cost analysis by incorporating the costs associated with court appearances, supervision of community-based orders, and detention.

Two examples of policy analysis using the Queensland Juvenile Justice Simulation Model have been published (Livingston et al. 2006; Stewart et al. 2008). Each of these will be briefly described.

Modelling an Early Intervention Program

Livingston et al. (2006) presented the results of the introduction of a family-based counselling program aimed at 5- to 10-year-olds. This simulated program targets young people who have yet to come into contact with the youth justice system. The results of Farrington's (1994) meta-analysis were used to provide estimates of the efficacy to parameterize this scenario. This meta-analysis indicated that such programs demonstrate a 12 % reduction in the initiation of offending by young people. Consequently, the model assumed that this program will prevent 12 % of generated new offenders from commencing a criminal career. Furthermore this simulated program was geographically targeted into areas with high populations of young Indigenous Australians.

The results of simulation indicated that immediately after introduction the program there was little impact of the program on the number of young people initiating offending. This was primarily due to the young age of the program participants. However, after a period of 10 years (the life of the simulation), the simulation estimated a reduction of 6.5 % in the number of court appearances by Indigenous young people and a reduction of 1.7 % in court appearances by non-Indigenous young people. This example highlights the benefits resulting from programs that target preadolescents can take a number of years to realise but over the longer term such programs can have a substantial impact on the number of young people appearing in the youth justice system.

Modelling a Youth Justice Conferencing Program

In Australia the use of restorative justice conferencing for youth offenders to divert young people from formal processing in the court has been introduced in all state jurisdictions. The QJJSM model was used to examine the impact of the introduction of youth justice conferencing in Queensland on the overrepresentation of Indigenous young people in the youth justice system (Stewart et al. 2008). Wider utilization of youth justice conferencing was proposed as a key strategy for reducing the high levels of overrepresentation of Indigenous Australians in the criminal justice system. Young offenders who have pleaded guilty are referred to a youth justice conference by police as a diversion from court or by the courts either in lieu of sentencing or prior to sentencing. Limited research has examined the impact of youth justice conferencing on reappearance in the criminal justice system. Work conducted in New South Wales (Luke and Lind 2002) compared the rate of rearrest for young offenders whose matter was finalized in the youth court with those who attended a conference. They estimated the rate of rearrest for conferences was 15–20 % lower than that for offenders who attended the youth court. This figure was used in the model as a measure of the efficacy of youth justice conferencing.

In 2003 when this simulation was carried out, a pilot of youth justice conference had been completed and the program was about to be rolled out across Queensland. In the policy scenario youth justice conference was equally available to all young people in Queensland regardless of gender or Indigenous status. This scenario was consistent with legislation and anticipated practices. The results of this simulation indicated the introduction of youth justice conferencing had an immediate effect of the number of finalized court appearances for both Indigenous and non-Indigenous young people. By 2011, there was a drop of 12.5 % in court appearances; non-Indigenous young people had a 13.7 % drop and Indigenous had a 10.5 % drop. However, these results need to be interpreted cautiously as over half the estimated reduction results from diversion for the court system, not a reduction in offending. The differential impact of youth justice conferencing is a result of the different offending profiles of Indigenous and non-Indigenous young people. Indigenous young people are younger at their first appearance, offend more frequently, and are less likely to commit property offences that their non-Indigenous counterparts (Stewart et al. 2004). Consequently a diversion program available principally for first time offenders and property offenders will have a greater impact on the number of court appearances and offending behavior of non-Indigenous young people.

The results of this simulation indicate that youth justice conferencing is unlikely to contribute to a reduction in the overrepresentation of Indigenous young people in the criminal justice system. In fact it has the potential to increase the disparity in the ratio of Indigenous to non-Indigenous offenders. These results contribute to the increasing debate about the need to focus outside the criminal justice system to reduce the rates of offending in Indigenous communities. Rather than focusing on criminal justice interventions (such as youth justice conferencing), more progress might be made if the focus was shifted to the underlying causes of aboriginal crime (e.g., substance abuse, family violence, poor school performance, and unemployment).

Current Status of the QJJSM

Unfortunately, the QJJSM has followed the footsteps of many previous models and is no longer functioning. Not only are microsimulation models extremely resource intensive to build; they also require ongoing resources to maintain and operate them. This model needed to be re-parameterized annually with new data to ensure that the model reflected current practices in the system. The model also needed to be updated to reflect changes occurring within the system such as new legislation and new sentencing options. Additionally, computer technology and microsimulation software is constantly changing and this model needed to be updated to take advantage of these new developments. Finally, it was difficult to convince criminal justice policy analysts about the value of simulation modelling. Within Queensland Treasury the model had “champions” who assisted with the initially building of the model. However, across the criminal justice system, policy analysts struggled to understand how they could use the model. The developers identified and tested the different policy scenarios. However, the ultimate aim was to have these policy scenarios generated from within government. Despite best efforts to educate the policy analysts, this did not happen. Consequently government did not continue to resource this model. However, the work in building the QJJSM has resulted in two important outcomes.

First, to build the model it was necessary to gain access to and develop a good understanding of the administrative data held by the various agencies of the Queensland criminal justice system. A critical parameter in building the model was the reappearance of young people in the criminal justice system. To analyze reappearance a longitudinal database was built by linking together administrative data for individuals born in 1983 or 1984, who offended and had contact with the youth justice system (police cautioning and youth court appearances). This work is ongoing and administrative data for these birth cohorts was last linked in 2011 when the individuals were 25/26 years old. The current longitudinal database (1983/84 Queensland Longitudinal Database) includes administrative data on contacts with the child protection system, the youth justice system, the adult court system and administrative data concerning their contacts with the mental health system are currently being negotiated. These data provide the basis of a productive ongoing research program examining the life course of offending, transition and turning points, cost of offending, evaluation of interventions, and geographic distribution of chronic and serious young offenders.

The second major benefit of the modelling program has been the development of strong relationships with influential public servants across the various agencies of the criminal justice system. These relationships have facilitated the research policy nexus, enabling researchers to influence the development of policies and legislation within the Queensland criminal justice system.

The Future of Microsimulation Modelling of the Criminal Justice System

In a special issue of *Social Science Computer Review*, Spielauer (2011) presented a very optimistic future for microsimulation modelling in the social sciences:

> Microsimulation is an approach whose time has come. More than half a century after the introduction of microsimulation into the social sciences, the main obstacles to this approach have almost disappeared. Computer power has increased exponentially, the collection individual data has become routine, social scientists are trained in micro data analysis and longitudinal research and the research itself has moved from a macro to a micro approach and is on the way toward a multilevel integration. The life course perspective has become a dominant paradigm and many of the pressing problems we faces are of a nature which makes dynamic microsimulation the most suitable study approach. (p. 10)

Despite this optimistic approach and the acceptance of policy-orientated microsimulation modelling in areas such as pensions, tax policy, income and poverty issues, and health policy, microsimulation modelling is still rare in the criminal justice system (Anderson and Hicks 2011). As Blumstein (2002) concluded, *Management of the agencies of the criminal justice system is still far from the model of efficiency one might like, and is still slowly moving into the information technology era* (Blumstein 2002, p. 22).

There are two key obstacles to simulation modelling practices becoming an inherent part of criminal justice planning, program, and policy development: the lack of skills among criminal justice policy makers and the politicized nature of the criminal justice system. The first obstacle to the use of simulation models is that criminal justice policy makers and practitioners do not understand simulation technology. These professionals struggle with assumptions required to build a model of a highly complex system, do not understand the various sources of data and analyses required to parameterize the model, and do not have the skills to develop and quantify policy scenarios. These professionals consider simulation models to be hard-to-operate-and-understand black boxes (Spielauer 2011). The traditional training of the criminal justice policy analyst and professional decision makers provides little or no exposure to simulation modelling. While graduates from criminal justice programs need not have the expertise to develop the models, they should be taught to appreciate the benefits of modelling and to incorporate the outputs of models into their professional practice. Without such training it is difficult for these professionals to develop a professional culture conducive to the use of models and simulations in criminal justice decision making. Without an understanding of the benefits of modelling, it is difficult for governments to justify the financial investments necessary to build, maintain, and operate the models.

The second obstacle to simulation modelling becoming normative practice in the criminal justice system is the highly politicized nature of the system. Arguably, crime and justice issues are the most politicized of public agendas. In recent years, with an increasing public concern about crime and crime rates, we have seen increasingly unsophisticated responses by our politicians to the management and prevention of offending. In these circumstances it is very difficult for researchers to encourage the use of "evidence-based" policy development let alone the use of simulation models. Furthermore, the reactive nature of current policy responses result in fluctuations within the criminal justice system that are extremely difficult to model. Auerhahn (2008a) argues that given the politicized nature of criminal justice policy, simulation modelling enables the consideration of more “radical” policy ideas at very little political or operational cost. Unfortunately, there is limited evidence that these policy ideas are implemented into actual practice in the criminal justice system.

Despite these challenges to the use of microsimulation modelling, three ways in which simulation modelling has the potential to aid decision making within the Australian criminal justice system have been identified. First, and perhaps most importantly, modelling has the potential to lead to a more informed debate about the allocation of scarce public resources. The Australian criminal justice system cost over $13 billion in the 2010–2011 financial year. These costs are increasing dramatically. However, crime victimization data indicates only slight increases in actual crime. In essence, the additional costs to the system can be largely attributed to changes in policing practices, public policy, and legislation, which are often the result of politically driven decisions in response to public concern about crime. This is reflected in criminal justice policies summed up in political catch phrases such as “get tough on crime,” “three strikes and you are in,” “zero-tolerance policing,” and “truth in sentencing” which have resulted in increases in police numbers, prison beds, and more punitive legislation. Indeed, as a result of such policies, the male incarceration rate in Australia has grown from 281 per 100,000 in 2002 to 324 in 2011. However, little attention has been paid to both the initial and cumulative costs of adopting such policies and services over a period of time (Austin et al. 1992).

A second way that policy simulation modelling may aid decision making is that such models can provide information that could assist decisions aimed at providing a more just and equitable criminal justice system. As demonstrated by the use of the QJJSM, microsimulation modelling enables an examination of the differential impact of indigenous status, gender, and geographic locality on young people entering and reappearing in the youth justice system. Currently, Indigenous young people are grossly overrepresented across Australian criminal justice systems. The modelling process facilitates the identification of points in the system where Indigenous young people (and other population groups) are differentially treated. Consequently, the use of modelling has the potential to lead to a more effective, accountable, and equitable criminal justice system.

A third benefit of policy simulation modelling is that it may provide information to decision makers crucial for the effective short-term and long-term planning of the system and the implementation of a whole-of-government approach to criminal justice in Queensland. There is an increasing recognition of the need for a more coordinated criminal justice system. Each section of the criminal justice system has its own policies, data collection strategies, and legislation. However, the complex interdependent nature of the criminal justice system means that any change in one section of the system could have substantial flow on effects to other parts of the system. Statistical models are required in order to determine the consequences of changes in policy and legislation not only for the host agency but also for other agencies in the criminal justice system.

The criminal justice system has much to gain from the use of dynamic microsimulation models. The process of modelling requires a sophisticated understanding and analysis of both the system and the data in the system. The modellers are required to understand what data are collected, why they are collected, what the limitations of these data are, and what gaps there are in the data collection. Simulation modelling transforms these data, which are routinely collected at great expense by the criminal justice agencies, into sophisticated decision support systems. This information has the potential to facilitate policy analysis, planning, and decision making leading to a more equitable and just criminal justice system.

Related Entries

- ▶ Agent-Based Assessments of Criminological Theory
- ▶ Agent-Based Modeling for Understanding Patterns of Crime
- ▶ Agent-Based Models to Predict Crime at Places
- ▶ Dynamical Simulation in Criminology
- ▶ Simulation as a Tool for Police Planning

Recommended Reading and References

Allard R (2010) Understanding and preventing indigenous offending. Indigenous justice clearing house brief No 9. Australian Institute of Criminology, Canberra

Anderson RE, Hicks C (2011) Highlights of contemporary microsimulation. Soc Sci Comput Rev 29:3–8

Auerhahn K (2002) Selective incapacitation, three strikes and the problem of aging prison populations: using simulation modelling see the future. Crimin Publ Policy 1:353–388

Auerhahn K (2007) Do you know who your probationers are? Using simulation modelling to estimate the composition of California's felony probation population 1980–2000. Justice Q 24:28–47

Auerhahn K (2008a) Dynamic systems simulation analysis: a planning tool for the new century. J Crim Justice 36:239–300

Auerhahn K (2008b) Using simulation modelling to evaluation sentencing reform in California: choosing the future. J Exp Criminol 4:241

Austin J, Cuvelier S, McVey A (1992) Projecting the future of corrections: the state of the art. Crime Delinq 38:385–408

Australian Bureau of Statistics (2003) Population projections, Australian, 2002–2101. Australian Government, Canberra

Australian Bureau of Statistics (2011) Prisoners in Australia, catalogue 4517.0. Commonwealth Government of Australia, Canberra

Australian Institute of Criminology (2012) Australian crime: facts and figures: 2011. Australian Institute of Criminology, Canberra

Belkin J, Blumstein A, Glass W, Lettre M (1972) JUSSIM, an interactive computer program and its uses in criminal justice planning. In: Cooper G (ed) Proceedings of the international symposium on criminal justice information and statistics systems. Project SEARCH, Sacramento, pp. 467–477

Belkin J, Blumstein A, Glass W (1973) Recidivism as a feedback process: an analytical model and empirical validation. J Crim Justice 1:7–26

Blumstein A (2002) Crime modeling. Oper Res 50: 16–24

Blumstein A (2007) An OR missionary's visits to the criminal justice system. Oper Res 55:14–23

Cassidy RG (1985) Modelling a criminal justice system. In: Farrington DP, Tarling R (eds) Prediction in criminology. State University of New York Press, Albany, pp. 193–207

Clark J, Lind B (2003) The New South Wales criminal justice system simulation model: further developments. Crime Justice Bull No 76. New South Wales Bureau of Crime Statistics and Research. http://www.lawlink.nsw.gov.au/bocsar1.nsf/pages/index

Crettenden I, Parker J, Macalpine S (1983) A computer simulation model of the district criminal court of New South Wales. New South Wales Bureau of Crime Statistics and Research, Sydney

Cullen FT (2011) Beyond adolescence-limited criminology: choosing our future – the American society of criminology 2010 Sutherland address. Criminology 49:287–330

Farrington D (1994) Early developmental prevention of juvenile delinquency. Crim Behav Ment Health 4:209–226

Farrington DP (2003) Developmental and life-course criminology: key theoretical and empirical issues – the 2002 Sutherland award address. Criminology 411:221–256

Gilbert N, Troitzsch KG (2005) Simulation for the social scientist, 2nd edn. Open University Press, Berkshire

Gordon MB (2009) A random walk in the literature on criminality: a partial and critical view on some statistical analysis and modelling techniques. Eur J Appl Math 21:283–305

Lind B, Chilvers M, Weatherburn D (2001) Simulating the New South Wales criminal justice system: a stock and flow approach. New South Wales Bureau of Crime Statistics and Research. http://www.lawlink.nsw.gov.au/bocsar1.nsf/files/r50.pdf/$file/r50.pdf

Livingston M, Stewart A, Palk G (2006) A microsimulation model of the juvenile justice system in Queensland. Trends and issues in crime and criminal justice No. 307. Australian Institute of Criminology, Canberra

Luke G, Lind B (2002) Reducing juvenile crime: conferencing versus court. Crime Justice Bull No. 69, New South Wales Bureau of Crime Statistics

Macy MW, Willer R (2002) From factors to actors: computational sociology and agent-based modelling. Annu Rev Sociol 28:143–166

Maltz MD (1996) From Poisson to the present: applying operations research to the problems of crime and justice. J Quant Criminol 12:3–61

McEwen TJ (1992) CJSSIM: Criminal justice system simulation model user manual. Institute for Law and Justice, Alexandria

Mollona E (2008) Computer simulation in the social sciences. J Manag Gov 12:205–211
Morgan PM (1985a) Modelling the criminal justice system. In: Moxon D (ed) Managing criminal justice. HMSO, London, pp 29–45
Morgan PM (1985b) Modelling the criminal justice system. Home office research and planning unit paper 35. Home Office, London
Nagin D (2012) Alfred Blumstein. In: Assad A (ed) The founders of operations research. Springer, New York, pp 707–719
Piquero AR, Farrington DP, Blumstein A (2003) The criminal career paradigm background and recent developments. Crime Justice Rev Res 30: 359–506
Productivity Commission (2012) Report on government services 2011. Australian Government, Canberra
Spielauer M (2011) What is social science microsimulation? Soc Sci Comput Rev 29:9–20
Stewart A, Spencer N, O'Connor I, Palk G, Livingston M, Allard T (2004) Juvenile justice simulation model: a report on the Australian Research Council Strategic Partnerships with Industry Research and Training Grant No C00106983. http://www.griffith.edu.au/__data/assets/pdf_file/0009/54846/jmagmodel.pdf
Stewart A, Hayes H, Livingston M, Palk G (2008) Youth justice conferencing and indigenous over-representation in the Queensland juvenile justice system: a micro-simulation case study. J Exp Criminol 4:357–380
Weatherburn D, Fitzgerald K, Hau J (2003) Reducing aboriginal over-representation in prison. Aust J Public Adm 62:65–73
Zarkin GA, Dunlap LK, Hicks KA, Mamo D (2005) Benefits and costs of methadone treatment: results from a lifetime simulation model. Health Econ 14: 1133–1150
Zarkin GA, Cowell AJ, Hicks KA, Mills MJ, Belenko S, Dunlap LJ, Houser KA, Keyes V (2012) Benefits and costs of substance abuse treatment programs for state prison inmates: results from a lifetime simulation model. Health Econ 21:633–652

Migration Policing

► Biometrics and Border Control Policing

Militarization

► Police and the Military Nexus

Minorities Within the Police Workforce

Simon Holdaway
School of Law, Centre for Criminological
Research, The University of Sheffield,
Sheffield, UK

Overview

It is important for police forces to serve all communities. Questions about the just use of police powers have fostered criminological research, not least in relation to their racially discriminatory use. Notions like racial profiling, for example, have signalled police unfairness and injustice. It has become clear that some communities are policed in ways that are unfair and unjust. Research, however, has also turned attention to justice and fairness within police forces. Questions about black and other minority ethnic officers' experiences of racial prejudice and discrimination as an aspect of relationships with colleagues; of whether and, if so, why women are underrepresented in the police ranks; about unequal promotion opportunities for majority and minority ethnic and other minority officers within the workforce; and many other related questions have been explored by social scientists.

Social inequalities within the police workforce might follow those found in wider society, but there might also be a distinct experience of prejudice and discrimination among minority ethnic and women officers. Gay officers' experiences might be different from those of other minorities. Race and gender are not fixed characteristics of individuals: they are social constructs and, if policies to alleviate discrimination and prejudice from within police forces are to be developed, it is necessary to understand how they are articulated and sustained.

Introduction

Many societies retain deeply embedded laws, policies, customs, norms, and related ideas that sustain social inequalities. Equality between men and women, between the members of majority and minority ethnic groups, between people of different sexual orientations, and between able bodied people and those with physical disabilities has therefore become a major theme of politics, law and policy across the world. Societies are at different points of development when their attempts to reform and create a more equal society are evaluated. Although often faltering, major reforms of law and policy, as well as many other aspects of social change, have nevertheless contributed to attempts to form societies within which social inequalities diminish and, ideally, become wholly insignificant. Indeed, reforms to create social equality have stimulated phenomena like the civil rights movement in the USA and womens' movements that reach across continents. What has come to be called "identity politics," where participants engage in debates and action about inequalities on the basis of one criterion, their personal identification as a woman, black, or gay, for example, is recognized as a keynote of contemporary politics (Phillips 2007; Young 1990).

Police forces are not immune from these wider social trends. Discriminatory policing of ethnic minorities has become a subject of significance in many societies and of criminological research (Lamberth 2003). The failure of police officers adequately to protect women who are victims of domestic violence has gained public and criminological attention (Council of Europe 2002). A glance at programs for all the major criminology conferences will reveal sessions on ethnicity and gender as standard subjects, and scholarly societies like the American Society of Criminology have established women's and Afro-American sections. Criminology has not been immune from identity politics.

When a person is recruited into a constabulary, they join an institution that predates their membership. This means that police recruits are socialized into a structure of existing relationships and ideas, some of which relate to minorities and, therefore, to prejudice and discrimination. Individuals respond to these relationships and ideas in different ways but always with reference to what have become dominant, powerful, taken-for-granted ideas, and relationships that officers accept as commonsense (Holdaway 1983). In this sense, the police is an institution with a structure that frames individual officers' ideas and actions, including ideas about and actions related to women officers and ethnic minority officers that include views about their appropriate career paths, the ways in which they understand policing, their place within a wider social hierarchy of equality and status, and many other subjects. These ideas and related actions, appropriately called aspects of the police occupational culture, can be prejudicial and discriminatory to minorities serving in the police workforce. Criminological research has therefore to some extent focussed upon the police workforce and the experiences of women, ethnic minorities, and gay officers within it.

Two aspects of prejudice and discrimination within the police workforce are of particular importance. The first is the mediation and expression of wider social inequalities that lead to prejudice and/or discrimination within constabularies. Police officers, police policies, and related actions are not immune from widely held assumptions about women, for example. There is a strong view, for instance, that policing is a man's job; that masculinity, so called, is essential to an officer's fulfilment of his duties (Brown 2007).

The second point is that, during the last two decades, an increasing number of applicants for recruitment to the police have been women, members of ethnic minorities, and gay people. They are more than familiar with being categorized by others in prejudicial and discriminatory ways. Despite expectations to the contrary conveyed in recruitment literature and policy statements, the experience of being a member of a constabulary becomes different for women officers and for those from majority and minority ethnic groups. Indeed, the experience can lead to a view that prejudice and discrimination pervade police culture (Holdaway 1991).

This is not to suggest that all police officers are individuals who are racists, women haters, or homophobic. Neither is it to suggest that all officers from minorities perceive and respond to prejudice and discrimination in the same way. Police officers are not clones. Rather, the research described and analyzed in this entry charts contours of relationships, beliefs, ideas, and related actions that form a framework for police employment and, therefore, the experience of officers from minority groups in the workforce. Particular importance is given to the occupational culture of the police, which is precisely the context within which categorizations and understandings of group membership vie with and yield to each other. Categorizations, remember, are external to and therefore imposed upon women, ethnic minorities, and gay people. Ascriptions of group membership, however, are subjective responses to categorizations that may lead to particular actions. The institution of the police and its occupational culture are not set in stone. Rather, they should be conceptualized as processes of interaction between categorizations and understandings of group membership (Brubaker et al. 2004).

Police forces are obviously not the only work institutions that exhibit prejudice and discrimination within employment. They are however particularly important because they are a central state institution, policing on behalf of the whole population. Inequalities within police forces that demonstrate a failure to address prejudice and discrimination portray a central institution of the state with a workforce that is not representative of the general population and one that perpetuates rather than ameliorates prejudicial and discriminatory experiences of policing for minorities. This is why the experiences of minorities within the police workforce are important and have captured the attention of criminologists.

Researching Minorities Within Police Forces

Questions about whether or not a police workforce is representative of the population it serves are usually researched by statistical comparison of the percentages of a particular minority within the general population and their number within the police workforce. Controls for the age profile of the minority population ensure that only people within the recruitment age range, requisite educational qualifications, and other variables are included in the analysis. The research task here, however, is rather uncomplicated.

Other studies, mostly about women officers, have used questionnaires to ask a sample about aspects of their work. They have then often compared findings with a sample of male officers asked the same questions. Subjects of relevance, for example, have been the balance between police work and family commitments, attitudes towards police work and particular aspects of it, and relationships with male officers (Brown and Savage 1992).

“Race” and “gender,” however, are not facts that can be measured straightforwardly. Their meanings differ across societies, within different contexts and they change. The category “Asian police officer,” for example, means a person from the Indian subcontinent in England, but in the USA a person with origins in Japan, China, Indonesia, or other close geographical areas. As far as membership of an ethnic group is concerned, some British Asian police officers regard themselves as “black” and some do not. Some Asian officers prefer to signify their ethnicity by identifying with their affiliation to Islam – they are Muslim police officers – others do not and call themselves “Asian” or “black.” Similarly, characteristics attributed to women officers by male officers have been found to be highly partial: their natural ability to deal with children, for example.

The main point here is that “race,” “ethnicity,” and “gender” are social constructs. We attribute particular meanings to them within particular contexts, often assuming they are shared generally and constitute commonsense. From this perspective, interesting research questions arise about the meanings attributed to women and to ethnic minority officers. Analysis of the interaction between categorizations and ascriptions of ethnicity and of gender within the particular context of the police workforce reveals social processes through which race, ethnicity, and gender are constructed, sustained, and, maybe, changed (Holdaway 1997a). The systematic analysis of qualitative data from interviews with minority and majority ethnic and with women police officers, of policy documents, and employment law informs research of this type.

Although gender and ethnicity are distinct subjects for research, they are combined in this entry. The aim is to make clear a theoretical perspective based on the view that they are both social constructions, understood through an analysis of similar social processes that constitute and sustain them within constabularies.

Beginnings: The Power of Ideas About Subordination

The general recruitment of women into the police is a relatively recent phenomenon, realized within a historical legacy of resistance to the idea that policing can be an occupation for females (Brown 2000; Heidensohn 1992). Comparing the history of women police officers in many European societies, Jennifer Brown and her colleagues concluded that, "The idea of women officers appears to have been met with both incredulity and hostility wherever the suggestions were first made" (Brown et al. 1999).

Initial advances in the recruitment of women are due more to pressures from circumstances external to constabularies than the fostering of the idea within the police or by governments that, indeed, policing is as much a job for women as for men. In England, for example, women could only fill the void left by the diminution of men available for recruitment following national conscription in 1914 (Brown et al. 1999). Expediency rather than views about equality was in the ascendancy. Elsewhere and to some extent in England, the advancement of women into constabularies developed by a perceived need for the policing of moral concerns suited to women's attentions. Bathers who lacked adequate clothing, the spread of venereal disease in postwar Cologne, the care of lost and wayward children, fears about the sexual promiscuity of women at a time when men were at war, and a need to police prostitution formed in different societies a rationale for the recruitment of women into forces (Brown et al. 1999). The suitability of women to care for children, to prescribe and police women's morals, and, by default, not to engage in the round of routine policing that includes arresting criminals, investigating crime, and patrolling the streets sustained ideas about women's roles.

The origins of these ideas lay beyond constabularies. Given expression within the particular context of policing, however, characterizations of women as "naturally" morally sensitive, as the carers of children, in summary, as different and supporters of rather than independent from or equal to male officers were sustained. Women recruited into this context and who did not wholly share these and related dominant ideas had to develop strategies and tactics to accommodate their own understandings and related styles of policing.

A historical context of subordination to white, often colonial, or immigrant ruling minorities is a key point for understanding the early recruitment of ethnic minorities into the police. The legacy of slavery, itself a policing system, forms the background and to some extent continuing context of the position of Afro-American officers in the USA (Dulaney 1996). In Africa and in India, colonial rule framed recruitment from indigenous populations. When first recruited into constabularies, indigenous Africans and Indians entered the rank and file. Colonial rule established a continuing, natural "officer class" for what were usually white immigrants into Africa, India, and the USA. Whereas a notion of "femininity" framed women officers, different, lasting ideas about the innate intellectual and moral inferiority of indigenous people structured commonsense views about the recruitment of ethnic minorities into the police.

Following the abolition of slavery and the movement of Afro-American, former slaves from southern to northern states, the policing of exclusive, black areas of cities had to be considered (Dulaney 1996). In most northern cities, black officers were recruited to police black communities. They were prohibited from arresting white suspects and not promoted above the rank of sergeant. In many places black officers were not allowed to wear uniforms similar to those worn by white officers. Slavery ended but confining the work of black officers in terms of what was suited to the innate characteristics of "blacks" perpetuated what would eventually be regarded by Afro-American officers as systemic racism.

Understanding Origins and the Perpetuation of Structures and Ideas

The history of ethnic and gender relations within particular societies is of importance when the contemporary situation of police recruitment is analyzed. The point is not that slavery and colonial rule remain, perhaps in a covert form. It is not that the notion of women as traditionally feminine wholly perpetuates police recruitment. Ideas and hierarchical relationships that flowed from them have nevertheless been influential and resistant to change. They have formed a structure, sometimes tightly woven, sometimes much less so, with a legacy that for many years and now places constraints on women and ethnic minority officers. Highly partial categorizations related to the long-standing subordination of ethnic minorities and of women officers remain and to some extent have become taken for granted as "commonsense."

Women and ethnic minority officers work within this structure of constraint, but their experiences of policing have not been determined wholly by it. Police officers, as said earlier, are not clones (Holdaway 1997b). Some will have accepted and maybe taken for granted constraints of gender or ethnicity. Others, experiencing the symptoms of gender and racialized inequalities within policing, have responded in ways that challenge apparent commonsense. Human beings are creative and in the next section constraints that have led to distinct responses to reform are discussed.

The Beginnings of Change

Social change is complex and many different circumstances should be taken into account when understanding why the recruitment of women and of ethnic minorities into the police are subjects of public debate and policy. Although every society exhibits unique points, key features that research needs to consider apply to all of them.

The first point is that basic changes in the governance of a significant group previously denied police employment, sometimes realized through political settlement, sometimes through protest and violent conflict, provide opportunities for women and minorities to enter new sectors of the labor market. The abolition of slavery in the USA created opportunities for recruitment to the police; the end of colonial rule in parts of Africa and in India opened up similar prospects. In England, although not central to police recruitment, a women's political protest movement, the suffragettes, lobbied successfully for the extension of the political franchise to women, acting as an agent of general change to ideas about suitable employment for women (Brown et al. 1999).

No matter how dramatic events that create change might be, the past nevertheless continues to constrain the present. It was some time before black officers patrolled white areas of the USA (Dulaney 1996). A key influence was the entry of Afro-Americans into city politics and the lobbying for the election of black mayors, who then appointed black police officer, including some chiefs. The civil rights movement would eventually pave the way for greater change. Similarly, it was many years before women officers in England undertook routine rather than specialist duties. In India, the caste system prohibited members of particular social groups from police employment. Apart from this constraint, after the end of colonial rule, it was some time before indigenous Indians could hold police rank. The same situation applies to the context of independence in many African societies.

Legal change, which can follow wider political changes of the type described, has also been very important for minorities' and women's recruitment to the police (Jones 1986; Heidensohn 1992). In the USA, the introduction of positive discrimination laws and related quotas for ethnic minority officers allowed the mounting of important legal challenges in cities like New York, where a deficit of Afro-American officers was apparent (Leinen 1984). With limited success, federal law in Ontario, Canada, has required constabularies to justify why they have not realized targets for the recruitment of women and for ethnic minorities (Todd and Todd 1992). In England, the Equality Act, 2010, is the latest legislation that builds on the Sex Discrimination Act 1975 and the Race Relations Act, 1976. Public bodies now have a statutory duty to demonstrate that their policies do not discriminate against legally defined minorities within their workforce.

Once equalities legislation has been passed, chief officers can introduce policies based on positive or affirmative action and, sometimes, positive discrimination (Edwards 1987). They can and in many countries have been required to publish equalities policy statements. Combined with legislation, these policies have a number of important functions. Law and policy publicly states actions to which a police force should aspire and against which they are monitored. They make clear what should be occurring in police forces and therefore offer a tangible reference point for redress when minorities within the police workforce are underrepresented or are the subjects of prejudice and/or discrimination. They therefore offer a protection, providing a measure of confidence among potential recruits and serving officers.

Cases brought by officers who allege discrimination in employment indicate a further function of equalities legislation, which is to create cases that become public landmarks for minorities in police forces (Leinen 1984). Such cases symbolize police forces' and chief officers' lack of commitment to legislation and that chief officers can be brought to account. Reporting of cases raises the public profile of inequality within the police; litigants are named and precedents are set.

Legislation also plays a part in placing a subject like equal opportunities in the police on the wider public policy agenda, which leads to further action within and out with constabularies. Governments begin routinely to collect statistics about the number of people from different groups – ethnic minorities and women, for example – entering police employment. In many cases the under-recruitment from these groups is revealed and further questions are asked about why this is so. Regulatory bodies concerned with monitoring and in some instances the prosecution of police forces (among other public bodies) become the justification for these organizations' existence. Regulatory bodies rarely work to abolish themselves.

These conditions relevant to the beginnings of change also have a wider relevance than, for example, individual acts of legislation being passed, discrete cases being brought, and unconnected policies being developed and regulated. They do not constitute a necessary, fixed point of development but are important features of a social process leading to changing the position of women and ethnic minorities within constabularies. The personal experience of, for example, a woman who is hesitant about police recruitment or of a black British person concerned about the possible experience of police employment is not viewed as events in the lives of untypical people. They are recognized as a part of a shared experience within an identified institution and therefore amenable to public comment. As the sociologist C. Wright Mills put it, "personal issues" become "public problems" (Mills 1959). Once realized, the status of a public issue changes. It begins to be secured as a recognized subject of public debate and confirms a reality of police employment for women and for ethnic minorities. Its dismissal as a personal grudge of a few individuals or a partial concern of minorities is challenged strongly.

Experience After Joining

Most research about women police officers has been concerned with documenting dimensions of prejudice and discrimination. One of the much cited, early studies is Susan Martin's participant observer research with Washington's police.

She summarized women officers' experiences in the following way:

> (Women) are excluded from the information exchange network and informal social life... Police women's behaviour is circumscribed by the stereotypical roles in which they are cast... which reminds women that as females they are sex objects, vulnerable to harassment yet held responsible for the outcomes of the interaction. (Martin 1980)

Attempts to integrate women into the police workforce have been challenged by what some have called a "cult of masculinity" articulated through the occupational culture (Heidensohn 2005). Among others, beliefs about the need for physical strength, bravado, and reliability to support colleagues in trouble have framed women officers' experiences of police work. Jokes and stories about women officers have sustained these views. Women have been excluded from socializing and working with male colleagues. Before considering them as colleagues who can be trusted, male officers have wanted to determine where their loyalties lie. Reviewing research studies undertaken in a good number of societies, Jennifer Brown found policemen exhibit hostile attitudes about women as patrol partners in the USA, Sweden, England, Scotland, Northern Ireland, and India (Brown 1997).

A number of important consequences have arisen from these circumstances. One is that women have not been promoted in adequate numbers to the supervisory ranks (Silvestri 2003). The other is that women officers are underrepresented in some specialist areas of policing traditionally regarded as most suitable for men. Brown found this in firearms squads in English constabularies (Brown and Sargent 1994). Westmarland found during her study of a provincial force that women were working disproportionately in specialisms related to child victims (Heidensohn 2005). More recent work has found that women working in the London Metropolitan Police child abuse units have been afforded a low status (Heidensohn 2005, p. 569).

Ethnic minority officers have a similar experience of discrimination and prejudice. Studies conducted in England and Wales documented the widespread use of racist language by white officers, racist jokes and stories, exclusion from socializing with white colleagues, and a diminished status within the workforce (Holdaway and Barron 1997). Research in American police forces has documented similar findings (Bolton and Feagin 2004).

The occupational culture has been analyzed to understand more precisely how race and ethnicity are articulated through it. A routine and necessary use of typifications which lend themselves to stereotypes of race (and gender), the demands of perceived teamwork as fundamental to policing, and the telling of stories and jokes that can articulate race prejudice sustaining discrimination have been found to be important (Holdaway 1997a). Officers' racialized identities were dominated by their ethnic majority colleagues' categorization of them as "black" rather than of police officers who happen to be black. White officers place the racialized identity of colleagues in the ascendancy.

Further dimensions of discrimination have been charted, identifying a lack of promotion prospects, premature resignation from constabularies, and unwillingness by senior officers to understand the context within which their minority ethnic officers work (Holdaway and Barron 1997).

Responses

One of the most interesting changes in policing has been the response of ethnic minority and of women officers to experiences of prejudice and discrimination. Ethnic minority and women police associations have worked within a context of legal change, where sex and race discrimination laws have become prominent and public expectations about the composition and management of the police workforce have been supportive.

Police forces are hierarchical organizations. Policy is developed by senior officers, with more junior ranks expected to implement it. An important feature of change in the position of ethnic minority and women officers, however, has been the reversal of this policy process (Holdaway 2009). Women and ethnic minority officers have formed police associations with memberships based solely on gender or ethnic identification.

The associations have in many respects become social movements within police forces. Forms of identity politics have entered debates and action to change police policy. Virtually all of the 43 constabularies in England and Wales have a Black Police Association and representative membership of the British Association of Police Women. American police forces have hundreds and probably thousands of similar associations. Discrimination and prejudice is challenged by all of these associations, as it is by larger groupings, the National Black Police Association and the European Network of Women Police, for example.

Criminologists can continue to describe and to some extent analyze the outcomes of associations' actions. The experience of officers can be documented. Previous studies can be repeated to assess the extent and character of change. This type of work will provide interesting information but not address important theoretical and related, empirical questions about the ways in which "race" and "gender" are constructed and sustained within the police.

It was argued in an early part of this entry that race and gender are social constructions. From this theoretical perspective it becomes possible to research changing meanings of such phenomena and, therefore, the ways in which race and gender are articulated within police forces, processes that lead to the changes, specific contexts within which prejudice and discrimination are experienced, and to compare and contrast what might be similar processes that marginalize minorities within the police workforce.

Race and gender are now matters for police policy. That has not always been the case. Race prejudice and discrimination were once phenomena of relevance to individuals who sometimes sought legal remedies – individual cases – to deal with the difficulties faced. The same could be said of gender and the situation of women officers. Once defined as phenomena to be addressed by policy, the meaning and status of race and of gender change. They are phenomena representing a collective experience requiring the attention of senior officers. What was an individual experience of a black officer or a women officer is now an organizational problem. In England what was once defined as race prejudice and/or discrimination and sometimes as "equal opportunities" became "institutional racism," called "systemic racism" in North America and Canada. The history of women's and ethnic minority police associations can be understood as aspects of a process within which race and gender are refined as they are related to a distinct, shared experience of police employment (Holdaway and O'Neil 2006). Differences between male and female officers and between black and white officers are emphasized within this process and presented as the widely shared experience of police employment. The notions of women officers and of black officers, for example, are accepted as representing particular experiences of police employment related to distinct realms of policy. While apparently secure, such experiences require the continuing vigilance of women's and black police associations. Their work sustains particular constructions of the relevance of gender, race, and ethnicity to police work and its management by senior officers.

One feature of the development of black police associations in the USA and in the UK has been an awareness of a distinct experience of police employment among minority ethnic officers who, despite having a dark skin color, do not identify as "black." These officers cannot unite with black officers, who define the experience of race prejudice and discrimination as common to all colleagues from minority ethnic groups (Holdaway and O'Neill 2004). The consequence of this situation has been the development of many police associations representing different ethnic groups – Asian, Greek, Italian, and Turkish police associations, for example. Other associations have been formed by the criteria of religious affiliation and in the UK, for example, there are Muslim, Jewish, Hindu, and Sikh police associations. The difference between these associations and those for black colleagues is a focus upon culture rather than race. Their claim is that unless particular cultural practices are recognized by a police force, they experience discrimination and prejudice. Halal food, access to prayer rooms in police stations, and freedom from work on the Sabbath have become subjects of policy.

Each ethnic group of officers, with its distinct culture, has been recognized, leading to conflicting understandings of minorities within the police workforce. Black officers argue that there is one experience of race discrimination, irrespective of membership of an ethnic minority group. To argue otherwise is to cause division, to weaken the claims of those who experience discrimination, and to privilege partial, cultural interests above those shared by all ethnic groups. Multicultural policies recognize differences between many groups; race policies address a fundamental, shared reality of discrimination.

Police forces throughout the world have responded to the claims of women officers and, though to a lesser extent, minority ethnic officers. Work is not yet finished, but policies to address many of the problems encountered by the members of these groups have been developed. Despite this situation, research in the UK suggests that while overt prejudice and discrimination has all but disappeared, their covert forms frame the reality of police employment. When asked for examples of covert discrimination and/or prejudice, as part of a study of black police associations, black officers presented examples that relied on their ability as a black person to know the motive lying behind a white person's action. They could sense it in a way a white person could not identify prejudice or discrimination (Holdaway and O'Neill 2005). It would be interesting to research women police experiences to understand whether they identify with similar forms of prejudice or discrimination.

The difficulty with these explanations of covert racism, as it is called, is that they rest on an assumption that black police officers are *essentially* different from white officers. They have a privileged access to the motives of white officers because, and only because, they are black. Their claim is an essential one, which is in conflict with ideas about race equality because claims to race equality are dependent on the assumption that all human beings are essentially the same. Black officers who define covert racism in these terms are thus arguing that they are the same as the ethnic majority but also essentially different.

An implicit argument here is that future research about minorities in the police workforce is that both similarities and differences between women and minority ethnic officers characterize experiences of discrimination and prejudice. Research has so far focussed upon differences between majority and minority groups in the police, which are important and the continuing subject of policy. Comparative research within and between police forces could however analyze in greater detail whether or not similar processes and forms of prejudice and discrimination lead to what are defined as essentially different experiences. We know, for example, that racist and sexist jokes, that exclusion from informal networks, and that stereotyping are processes that marginalize officers from both groups. Closer attention to this matter could clarify the essential and phenomenal forms of marginalization and exclusion within police forces and, maybe, in other work organizations.

Related Entries

▶ Biased Policing
▶ Police Culture
▶ Women in Policing

Recommended Reading and References

Bolton KJ, Feagin JR (2004) Black in Blue: African American police officers and racism. Routledge, New York

Brown J (1997) European policewomen; a comparative research perspective. Int J Sociol Law 25:1–19

Brown J (2000) Gender and policing. Palgrave/Macmillan, Basingstoke

Brown J (2007) From cult of masculinity to smart macho culture: gender perspectives on police occupational culture. In: O'Neill M. Marks, M and Singh, A (eds)

Police occupational culture: new debates and new directions. Elsevier, Oxford
Brown J, Sargent S (1994) Policewomen and firearms in the British police service. Police Stud 18:1–16
Brown J, Savage S (1992) Policewomen's career aspirations: some reflections on the role and capabilities of women in policing in Britain. Police Stud 15:13–19
Brown J, Hazenberg A, Ormiston C (1999) Policewomen: an international comparison. In: Mawby RI (ed) Policing across the world: issues for the twenty-first century. UCL Press, London, Policewomen, Police, Gender
Brubaker R, Loveman M, Stamov P (2004) Ethnicity as cognition. Theor Soc 33:31–64
Council of Europe (2002) Recommendation of the Committee of Ministers to member states on the protection of women against violence. Adopted on 30 Apr 2002; and Explanatory Memorandum, Strasbourg Council of Europe
Dulaney WM (1996) Black police in America. Indiana University Press, Bloomington
Edwards J (1987) Positive discrimination, social justice and social policy. Tavistock, London
Heidensohn F (1992) Women in control?: the role of women in law enforcement. Clarendon, Oxford
Heidensohn F (2005) Gender and policing. In: Newburn T (ed) Handbook of policing. Willan, Collumptom, pp 556–577
Holdaway S (1983) Inside the British police: a force at work. Blackwell, Oxford
Holdaway S (1991) Recruiting a multi-racial police force. HMSO, London
Holdaway S (1997a) Constructing and sustaining 'race' within the police workforce. Br J Sociol 48:19–34
Holdaway S (1997b) Responding to racialised divisions within the workforce – the experience of black and Asian police officers in England. Ethn Racial Stud 20:69–90
Holdaway S (2009) Black police associations: race and ethnicity within constabularies. Oxford University Press, Oxford
Holdaway S, Barron A-M (1997) Resigners? The experience of black and Asian police officers. Macmillan, Basingstoke
Holdaway S, O'Neil M (2006) Ethnicity and culture: thinking about police ethnicity. Br J Sociol 57:483–502
Holdaway S, O'Neill M (2004) The development of black police associations in the UK. Britsh Journal of Criminology 44:854–865
Holdaway S, O'Neill M (2005) Institutional racism after Macpherson: an analysis of police views. Policing 16:349–369
Jones S (1986) Policewomen and equality. Macmillan, Basingstoke
Lamberth J (2003) Racial profiling data analysis study: final report for the San Antonio police departments. Lamberth Consulting, Pennsylvania
Leinen S (1984) Black police: White society. New York University Press, New York
Martin S (1980) Breaking and entering: policewomen on patrol. University of California Press, Berkeley
Mills CW (1959) The sociological imagination. Oxford University Press, New York
Phillips A (2007) Multiculturalism without culture. Princeton University Press, Princeton
Silvestri M (2003) Women in charge: policing, gender and leadership. Willan, Collompton
Todd R, Todd F (1992) Ethnic relations and community policing: a Canadian model for Europe. Br J Can Stud 7:277–289
Young M (1990) Justice and the politics of difference. Princeton University Press, Princeton

Miscarriages of Justice

► Legal Rules, Forensic Science and Wrongful Convictions
► Prosecution and Wrongful Convictions

Mitigation and Aggravation at Sentencing

Julian V. Roberts
Faculty of Law, Centre for Criminology,
University of Oxford, Oxford, UK

Overview

The determination of factors which aggravate or mitigate punishment is an under-researched yet vital subject in the field of sentencing. Sentencing factors affect the severity of sentences imposed. Indeed, the determination of sentence may be regarded as a judicial weighing of all relevant mitigating and aggravating circumstances. The important word here is "relevant." Before a court takes a particular factor – X – into account, it needs to be satisfied that this factor is relevant to the sentencing decision. Factors unrelated to sentencing – such as the offender's income, gender, race/ethnicity, or social class – should be ignored. As will be seen, some sentencing guidelines explicitly direct courts to ignore irrelevant factors such as race and employment status.

After some introductory comments this chapter notes the sources of guidance for sentencers with respect to mitigation and aggravation. This is followed by a discussion of the problems associated with this guidance. At present, guidance regarding sentencing factors tends to be modest: guideline schemes around the world generally adopt a relaxed approach to the use of sentencing factors. Even in the US-based systems where some factors are excluded from consideration, guidance regarding the application of other factors is quite limited. This state of affairs may reflect the view that consideration of mitigation and aggravation is more properly left to the exercise of judicial discretion, with only minimal direction from the legislature, the guidelines authority, or the appellate courts (for further discussion on aggravation and mitigation, see Walker 1999; Ashworth 2010; Roberts 2011b).

Establishing the Relevance of Factors

How do we establish that a factor is relevant? One definition is that a factor is relevant if it relates to the objectives of sentencing. For example, an offender's expressed intention to take a drug treatment program is relevant to the sentencing purpose of rehabilitation. Accordingly, a court may reasonably take this fact into account and mitigate the sentence imposed, because the offender is taking steps towards rehabilitation. Now consider an aggravating factor such as hate motivation. Imagine that an offender is convicted of racially abusing his victim prior to and during an assault. This factor makes the crime more serious: the victim's social identify and social group have been threatened. Since one of the purposes of sentencing is to impose a sentence which reflects the harm caused, the offender can legitimately be punished more severely; hate motivation is therefore a relevant aggravating circumstance.

In most jurisdictions, the objectives of sentencing are enshrined in statute. For example, the Criminal Justice Act 2003 identifies the sentencing purposes in England and Wales, while Section 718 of the *Criminal Code* of Canada lists the sentencing objectives for courts in that country. Across the USA, guidelines manuals such as those in Minnesota or at the federal level provide this information for courts. Consulting these purposes will provide courts with insight into the relevance of any particular factor at sentencing. Since most countries pursue multiple sentencing objectives, including deterrence, rehabilitation, incapacitation, and retribution, the list of potential mitigating and aggravating factors is vast, hence the need for guidance from either the legislature, courts, or a sentencing council or commission.

Factors Unrelated to Sentencing Purposes

Finally, there are some factors considered by courts – usually mitigating in nature – which are clearly unrelated to any specific sentencing purpose but which are nevertheless recognized as legitimately affecting sentencing. The best known of these is probably the guilty plea discount. Pleading guilty does not make the crime less serious or reduce the offender's level of culpability for the offence. Nor can it be said that giving reduced sentences to offenders who plead guilty is relevant to deterrence. If anything, significant reductions in severity for offenders who plead guilty may undermine the deterrent power of sentencing: offenders may be encouraged to offend – or at least not deterred from offending – by knowing that they can get a lighter sentence by pleading guilty.

Yet almost all countries offer offenders who plead guilty a significantly discounted sentence. Why? It is done in the broader interests of justice. When an offender pleads guilty, the state saves the cost of a trial, and victims and witnesses are saved the trouble of testifying. So in order to promote more efficient administration of justice and to spare victims and witnesses, we offer an incentive to offenders in the form of a reduced sentence. The same argument is used to justify sentence reductions in cases where the offender provides assistance to the state in prosecuting other offenders. These two factors – a guilty plea and providing assistance to the state – are unrelated to crime seriousness or offender culpability but are nevertheless taken into account to promote broader goals of the criminal justice system.

The concept of crime seriousness is primordial across most jurisdictions, whether common or civil law. Gravity is a function of harm and culpability. Crimes cannot be precisely calibrated by legislatures – although gross differentiations can be made and are reflected in the relative severity of statutory maximum penalties. Accordingly, courts must weigh a wide range of offence – and offender-related factors in order to determine the seriousness of the offence and hence the severity of the sentence. Some factors enhance sentence by increasing or lowering the seriousness of the crime; others influence sentence by enhancing or diminishing the culpability of the offender. Mitigating and aggravating factors are often taken for granted by members of the public – we all have intuitions about whether factor X should result in a more lenient or a harsher sentence. But many problematic circumstances exist. Factors may be considered relevant by some, irrelevant by others.

Consider intoxication, a circumstance explored by Padfield (2011). Should the fact that the offender was drunk at the time of the offence mitigate the sentence on the grounds that the offence was uncharacteristic ("I have never seen him act like that – it must have been the booze")? On this line of reasoning, intoxication sustains a more general claim that the offence was "out of character." People who normally lead law-abiding lifestyles sometimes commit criminal acts under the disinhibiting effect of alcohol. But are they reasonably adjudged less blameworthy as a result? There is a strong counterargument in cases in which the offender has a history of heavy drinking: he or she may be seen as being more culpable for failing to exercise some restraint regarding a known problem. Might intoxication actually aggravate the sentence – as in the case of an offender who drinks heavily knowing that this often results in criminal assaults in pubs? As with a number of other factors then, intoxication may aggravate or mitigate, depending upon the particular circumstances of the offender. Intuition, in short, is often a poor guide to principled sentencing.

Conceptual Frameworks

Despite the importance of sentencing factors, as Ashworth notes, the concepts of aggravation and mitigation have attracted little close examination or theoretical discussion (2010, see also Walker 1999; Roberts 2011b). Categorical breakdowns of sentencing factors exist but surely more than this is needed. More recently, Jacobson and Hough propose a categorical framework for factors relating to personal mitigation (Jacobson and Hough 2007a; Roberts and Hough 2011). They propose classifying factors into categories relating to the crime, the offender's response to the offence and prosecution, and the offender's past (including previous convictions) and future. Classifying factors is one way of bringing some order to the vast array of factors potentially relevant to the determination of sentence. As will be seen, there is considerable variation across the common law world with respect to the degree and nature of guidance regarding sentencing factors.

Sources of Guidance: Legislature and Appellate Courts

The Legislature: Statutory Sentencing Factors

Legislatures affect sentencing factors through codification. Almost all common law jurisdictions have placed certain mitigating and aggravating factors on a statutory footing – albeit with some exceptions: in Tasmania, for example, the sentencing legislation does not identify any factors. According to some statutes judges are directed to consider these statutory factors when determining the seriousness of the offence. For example, Sects. 2 and 3 of Chap. 29 of the

Swedish Criminal Code 1988 specify factors that enhance or diminish the penal value of the crime. In addition, Sect. 5 of the chapter specifies eight additional mitigating circumstances that a court shall consider "to a reasonable extent." These include factors such as whether the punishment would have a disproportionate effect on the offender due to advanced age or ill health. Finland adopts the same approach.

The number of codified factors is generally small: in Finland a sentencing statute specifies only five factors which should increase the severity of the punishment, while the Canadian *Criminal Code* specifies only a handful of aggravating factors and no mitigating factors. This reflects a trend found elsewhere, including England and Wales, to provide more guidance with respect to aggravation than mitigation. The explanation for this asymmetry is unclear; it may reflect a view that aggravating factors are relatively straightforward and consistent in application. This means that the legislature can specify a factor as being relevant, confident in the knowledge that this factor is easily and consistently applied by courts. For example, if one agrees that hate motivation is an aggravator, it surely operates across all offences and offenders. There may be less consensus about the relevance and weight of different mitigators, in which case it is hard to be more prescriptive.

In England and Wales, only a few sentencing factors have been placed on a statutory footing. The *Criminal Justice Act 2003* provides no indication regarding the relative weight of these factors, their potential to affect the decision of whether a term of custody is imposed, or whether they are more important than other sentencing factors emerging from the case law. The statute thus provides only limited guidance to sentencers regarding the sources of aggravation and mitigation. This reflects the existence of the sentencing guidelines issued by the Sentencing Council and which are designed to supplement and interpret the statutory framework. Other jurisdictions provide more comprehensive lists of statutory factors. Thus, the Crimes (Sentencing Procedure) Act 1999 in New South Wales identifies 21 aggravating and 16 mitigating factors. Section 9 of the *Sentencing Act 2002* in New Zealand notes ten aggravating, seven mitigating, and one statutorily excluded sentencing factor (see Young and King 2011).

Guidance from the Appeal Courts

Until the advent of sentencing councils and commissions to disseminate guidelines, the appellate courts have been the traditional source of guidance at sentencing. An evaluation of the adequacy of appellate guidance regarding mitigating and aggravating circumstances is beyond the scope of this chapter. However, numerous academic commentators and commissions have identified the limitations on appellate guidance in general and the area of mitigation and aggravation in particular. Most sentence appeals address a specific point of law or provide a test for whether the sentence imposed was manifestly unfit. Guideline judgments in which the court sets out the mitigating and aggravating factors relevant to the offence are rare.

Sentencing Guidelines and Mitigating/Aggravating Factors

United States

Under most American sentencing guideline schemes, the role of mitigating and aggravating factors is generally to justify the imposition of a sentence that is outside the range prescribed by the guidelines or to move the case from the normal range up to an aggravated or down to a mitigated range. In the landmark decision in *United States vs. Booker*, the US Supreme Court held that the federal sentencing guidelines are merely advisory rather than bindingly presumptive in nature. The mitigating circumstances therefore may be reasonably described as mitigating factors rather than "departures" in the stricter sense of the pre-*Booker* era (see also discussion in Berry 2011).

The Utah guidelines manual notes that "There are occasionally circumstances that compel deviation from the guidelines" (Utah Sentencing Commission 2007, p. 12). Directions are provided to courts with respect to the consideration of the sources of mitigation and aggravation. A non-exhaustive list of factors is provided, and courts are directed that in considering all aggravating and mitigating factors in a particular case, the number of each should not merely be added up or otherwise mechanically applied in the balancing process. Rather, the totality of the mitigating factors should be compared against the totality of the aggravating factors. The sentencing guidelines in North Carolina provide sentencers with three sentence ranges: presumptive, aggravated, and mitigated (see Lubitz 1996; North Carolina Sentencing and Policy Advisory Commission 2007).

England and Wales

Sentencing guidelines exist for most common offences and are issued by the Sentencing Council for England and Wales or its predecessor, the Sentencing Guidelines Council (see discussion in Ashworth 2010). Rather than forming part of a grid encompassing all offences, the English guidelines are offence specific. They provide sentence ranges for categories of seriousness – most offences are divided into three levels. Each definitive guideline also contains a list of relevant mitigating and aggravating factors of particular relevance to the specific offence which is the subject of the guideline. For example, for the crime of domestic burglary, committing the offence at night, when a child is present in the residence, and creating damage to property are all cited in the guideline as aggravating factors (Sentencing Council 2011). In addition, there is a guideline for the determination of offence seriousness which is applicable to all offences (Sentencing Guidelines Council 2004).

In its definitive guideline regarding the determination of offence seriousness, the Council provides two lists containing 31 aggravating factors that arise from a higher level of culpability or a greater degree of harm (or sometimes both conditions). These are described as being the most important aggravating circumstances with application across many offence categories. The lists are non-exhaustive and provide no hierarchy of importance. The lists include both statutory and non-statutory factors but make no commentary on the relative importance of the two. They are then followed by four factors that indicate when an offender's culpability is unusually low or that the harm is less than usually serious. There is an imbalance in the lists therefore: 31 factors are identified that enhance sentence severity, while only four factors are noted which may reduce the level of culpability or harm (Sentencing Guidelines Council 2004).

Guilty Plea Discount

With respect to one important sentencing factor – a guilty plea – the Sentencing Guidelines Council (now the Sentencing Council) has provided clear and detailed advice. The Council's definitive guideline makes it clear that the reduction for a guilty plea "derives from the need for the effective administration of justice and not as an aspect of mitigation" (Sentencing Guidelines Council 2007, p. 4). Nevertheless, in other jurisdictions a guilty plea is interpreted as a mitigating factor and so the issue is worth comment in the present context. The definitive English guideline published in July 2007 specifies the range of the reduction in sentence as well as the circumstances that justify different levels of reduction (Sentencing Guidelines 2007). This level of guidance reflects the importance of this factor in terms of sentencing and the administration of justice (see earlier discussion). With respect to other aggravating and mitigating factors, however, guidance is less comprehensive or precise. In all the guideline schemes, a number of elements are missing or require greater elucidation.

General Issues Arising from Mitigating and Aggravating Factors

Rationale for Mitigating or Aggravating Sentence

It would be useful for a guidelines authority to articulate the *rationale* for considering specific mitigating and aggravating factors. This is not simply an academic exercise; awareness of the rationale for mitigation will focus sentencers' attention on the question of relevance. The English Sentencing Guidelines Council's

Seriousness guideline provides only a limited rationale for the invocation of any specific factor. Sentencing factors are particularly relevant when they indicate a higher or lower level of culpability or harm. Some sentencing factors however are unrelated to these concepts and are applied because they are relevant to the statutory objectives of sentencing identified in s.142 of the *Criminal Justice Act 2003*. Sentencers should be encouraged to use this section to guide their application of the factors. When contemplating a potential mitigating or aggravating circumstance, it is important to consider whether it is relevant to one of the sentencing objectives or principles.

Some factors may be relevant independent of any link to the statutory objectives, but in general, the coherence of the sentencing process is impaired if many factors unrelated to the sentencing purposes are taken into account. A clear statement of the rationale for any specific factor would also help members of the community and crime victims to understand why the factor was taken into account at sentencing. A good example of an ambiguous or controversial factor is positive actions by the offender which speak to his or her good character but which are unrelated to the offence.

Criminal History

Perhaps the most important sentencing factor operates in mitigation and aggravation, potentially increasing the severity of sentence in some cases and reducing it in others. Generally speaking, the absence of any prior convictions is a powerful mitigating factor: the offender claims a "first offender" sentencing discount. The argument is advanced that the offence was out of character since he has never been convicted before. Moreover, first offenders receive a discount because the state recognizes human frailty – everyone can fall into crime at some point; this is known as the "lapse" theory (see von Hirsch 2010). First offenders are also less likely to re-offend (compared to repeat offenders), and so the absence of prior convictions justifies more lenient treatment under a rehabilitation or deterrent sentencing rationale.

On the other hand, as the offender re-offends, accumulating more convictions and becomes a repeat offender, sentences generally get harsher. The presence of prior convictions is in fact the most important aggravating factor under the US guideline schemes. Some US systems (including the federal guidelines) use a two-dimensional grid where one dimension or axis is crime seriousness, while the second is criminal history.

Several justifications exist for imposing harsher sentences on repeat offenders. First, they may be regarded as more blameworthy for having committed a crime after repeatedly being sentenced for other offences (Roberts 2008). Second, repeat offenders are in general more likely to re-offend and therefore should receive harsher sentences according to a deterrent or incapacitation perspective. It is important that guideline schemes articulate the justification for the use of prior convictions and also the weight that previous convictions should carry at sentencing. Otherwise, courts will apply their own criteria in considering this common factor. The US guidelines provide this direction, but no other jurisdictions have followed this example.

Worthy Conduct Unrelated to the Offence of Conviction

A recurrent mitigating circumstance in many jurisdictions arises when prior to or after conviction the offender has performed some commendable action for the community. This is sometimes recognized in sentencing law. For example, s.6 of the Sentencing Act 1991 in the Australian state of Victoria states that "any significant contributions made to the community by the offender" is a relevant factor to be considered at sentencing. The classic case involves the offender who shows some act of exceptional bravery shortly before or after conviction. It is unclear why this kind of action should justify leniency, yet sentencers sometimes mitigate the sentence on this basis. Indeed, there is a strong intuitive appeal to a policy of mitigating sentences for offenders who have a history of very creditable social behavior.

Most jurisdictions permit courts to recognize a mitigating factor that is unrelated to the principal theoretical orientation of sentencing. The Wisconsin Sentencing Commission provides a list of "additional factors" that do not relate to the primary sentencing rationale. The list includes items as diverse as the contents of the victim impact statement and whether the offender suffered public humiliation or loss of employment as a result of having been convicted (Wisconsin Sentencing 2007, p. 8). However, these factors should not carry the same weight as more fundamental sentencing factors – those relating to crime seriousness or culpability. Guidance would be particularly useful with respect to the weight that should be attached to collateral consequences of the offence. Here, too, distinctions will need to be made between consequences that the offender should have foreseen – such as loss of driving privileges and subsequent employment for taxi drivers – and other, exceptional and unusual consequences such as the death of a relative or friend as a result of a fatal accident arising from a driving offence. It is surely more reasonable to consider the latter than the former as justifying a mitigated punishment.

Power of Sentencing Factors to Influence Sentence

Guidance is seldom provided with respect to the power or influence of a specific sentencing factor. For example, how much should a sentence be reduced to reflect the fact that the offender is remorseful and has apologized to the victim? How much harsher should a sentence be because the offender spent some considerable time planning the crime? One explanation for this absence of guidance is that it is hard to specify in advance; the importance of premeditation or remorse will depend on the individual circumstances of the case appearing for sentencing. In some cases remorse may be much more important than others – and it is therefore impossible to specify this in advance. Legislatures for this reason leave courts free to make the decision as to how much weight should be ascribed to particular factors.

Some guidance may be useful regarding the relative weight to be assigned different aggravating and mitigating factors – they are clearly not all equally important. The circumstances identified by the English Council in its Seriousness guideline represent a wide range, some important, others less so. For example, by virtue of its statutory foundation and the number of cases to which it is relevant, the existence of previous related convictions should carry far more weight than factors such as "causing gratuitous damage to property over and above what is needed to carry out the offence." Sentencers should be provided with some sense of the relative power of these factors to influence the quantum of punishment. This is not to say that precise weights should be assigned to factors, rather that if a factor is particularly important, sentencers should be aware of this fact – otherwise they are free to develop their own hierarchies of impact.

This level of guidance is provided in some guideline schemes. For example, in Sweden the existence of mitigating factors permits the court to impose a sentence below the statutory minimum sentence. Allowing a factor this much influence is controversial, but at least sentencers are given an indication of the extent to which a mitigating factor may reduce the custodial term. In a similar fashion the sentencing statute in the state of Alaska identifies two gradations of mitigation. If the presumptive term is 4 years or less, mitigating factors may reduce the sentence by up to this amount. If the presumptive term is longer than 4 years, mitigating factors may reduce the sentence by up to 50 %. Placing limits on the impact of sentencing factors provides sentencers with an idea of the extent to which these factors can affect the sentence. Moreover, it would surely be useful to indicate to courts whether sentencing factors taken together, or on an individual basis, can make the difference between custody and a community sentence. Some guidance should be provided with regard to the factors which have the potential to move the offender above or below the custodial threshold.

Categorizing Sentencing Factors

The sentencing factors provided by statutes or guideline schemes include a mixture of different factors, some relating to the offender, some to the offence. A more coherent approach would involve providing categories of factors. This approach has been recommended in a report by the Australian Law Reform Commission (ALRC 2006). The ALRC proposed grouping sentencing factors into eight categories with the intention of setting them out "in a logical and clear manner, such as within groups that share a common theme" (p. 168). If sentencers worked through each category systematically, this might result in a more thorough and consistent review of the circumstances of each case.

Excluding Factors and Controversial or Asymmetrical Factors

The English Sentencing Guidelines do not specify factors that should *not* be considered in mitigation or aggravation. It would be both unrealistic and undesirable to construct a sentencing guideline which would specify all potential mitigating and aggravating circumstances. However, greater consistency of application would be promoted by identifying problematic factors that should not mitigate or aggravate or by specifying the way that factors should be considered. Most sentencing guideline systems identify a number of factors that should be excluded from consideration. For example, the detailed series of proposals made by the Australian Law Reform Commission includes a number of grounds which should not aggravate sentence.

Similarly, many US guidelines identify specific grounds which should not be considered by a sentencing court. For example, in Minnesota, the guidelines exclude a number of factors including race. Establishing a list of proscribed as well as permitted factors also compels the guidelines authority to confront the underlying principles and consequences of different mitigating factors. Consider the offender's employment status or employment history. This circumstance is often cited by sentencers as justifying the imposition of a mitigated sentence, in order to preserve the offender's source of income or possibly to prevent the infliction of collateral harm upon his or her dependants. Research has repeatedly demonstrated that employment status is a significant predictor of whether a custodial sentence is imposed (e.g., Moxon 1988). However, the application of this factor obviously disadvantages unemployed defendants or offenders with poor employment histories. Visible or immigrant minority offenders may be particularly affected, which is why a number of commentators have advocated the prohibition of this factor and why some sentencing guidelines (such as those in Minnesota) explicitly exclude it.

Personal Mitigation

Finally, sentencing guidelines around the world say little about "personal mitigation" – factors relating to the offender rather than the offence and which typically serve to mitigate punishment. In his sentencing text Thomas argued that allowance for mitigation should not be considered to be an entitlement of the offender (1979, see p. 174). Times may have changed since he expressed this opinion. Today, personal mitigation is generally accepted as a fundamental element of the sentencing process – as noted by Lovegrove (2010). A distinction has sometimes been made between personal mitigation and mitigating factors relating to offence seriousness – to which an offender is entitled. The distinction is spurious and invites disparity of application. If a particular factor reduces crime seriousness or the offender's level of culpability – whether as a result of personal factors or not – the sentence should be mitigated accordingly.

The English Council's Seriousness guideline notes only the statutory provision in the Criminal Justice Act 2003 which authorizes courts to take account of any matters that "in the opinion of the court, are relevant in mitigation of sentence" and then states that courts should "consider any offender mitigation. The issue of remorse should be taken into account at this point along with other mitigating features such as admissions to the police in interview" (Sentencing Guidelines Council 2004). The Magistrates' Court Sentencing Guidelines simply direct magistrates to consider "offender mitigation" and offer three or four examples.

Personal mitigation surely requires more attention from scholars. Research conducted in 1980 found that personal mitigating factors appeared more than any other category in speeches for mitigation (Shapland 1981). More recent research with sentencers in England and Wales has demonstrated that personal mitigation still plays the largest part in tipping the balance away from the imposition of a term of custody (see Hough et al. 2003, at p. 37). The most recent English study found that in approximately one-third of the cases observed, judicial recognition of personal mitigation changed what would have been a custodial sentence to a community-based sanction (Jacobson and Hough 2007; Roberts and Hough 2011). This research study also revealed that many different personal mitigating factors were taken into account by sentencers and that while agreement was expressed with respect to some factors, participants disagreed on many others. As the authors note, the Council needs to provide better guidance in this area, beginning with the principles on which personal mitigation is based.

Role of Public Opinion

Ideally, mitigating and aggravating factors should be recognized by the community as being relevant to sentencing. If courts take a factor into account deemed irrelevant by the community, confidence in sentencing will decline. There is a considerable literature on public attitudes to mitigating and aggravating factors. This research reveals that the public recognize a wide range of factors as being relevant to sentencing (e.g., Lovegrove 2010; Roberts and Hough 2011). In this sense they are like judges, who, as noted, take a wide range of factors into account at sentencing. On the other hand, there are some factors where the community and courts diverge. In practice, the absence of prior convictions is an important mitigating factor; first offenders usually receive a significantly reduced sentence compared to repeat offenders convicted of the same offence. The public, however, are less enthusiastic about offering this discount to first offenders, particularly for more serious crimes.

In order to promote a greater degree of "fit" between courts and communities regarding sentencing factors, guidelines authorities often undertake research into public attitudes to mitigation and aggravation, to see which factors attract greatest support (e.g., Roberts et al. 2008, 2009). The findings from this research then help guidelines authorities shape the mitigating and aggravating factors found in the guidelines (Roberts and Hough 2011). Of course there are limits to this kind of exercise; an offender should not receive a harsher or more lenient sentence just because his case contains a factor the public perceive to be relevant to sentencing; there must be some legal relevance to the factor.

Conclusion

The number of potentially relevant mitigating and aggravating factors is, as noted, huge. For this reason, it is important to establish which are relevant and to provide courts with some guidance regarding the application of sentencing factors. Over a generation ago, scholars called for more judicial guidance regarding the criteria for mitigation and aggravation. Thus, Ashworth lamented the absence of "any thorough consideration of the claims of the various personal factors" and argued that "this is a sphere in which discretion has led largely to anarchy" (1987, pp. 30–31). So long as it does not unduly impair a court's ability to impose an appropriate disposition, any attempt to structure judicial discretion is welcome. Courts around the world would benefit from more detailed and structured guidance regarding the use of mitigating and aggravating factors at sentencing. The search for relevant sentencing factors should not involve a fishing expedition but rather an inquiry into principled adjustments to the sentence.

Related Entries

► Sentencing as a Cultural Practice

Recommended Reading and References

Ashworth A (1987) Criminal justice rights and sentencing: a review of sentencing policy and problems. Unpublished manuscript. Worcester College, Oxford

Ashworth A (2010) Sentencing and criminal justice, 5th edn. Cambridge University Press, Cambridge

Australian Law Reform Commission (2006) Same crime same time. Sentencing of federal offenders. Australian Law Reform Commission, Sydney

Berry W (2011) Mitigation in federal sentencing in the United States. In: Roberts JV (ed) Mitigation and aggravation at sentencing. Cambridge University Press, Cambridge

Cox E (1877) The principles of punishment as applied in the administration of criminal law, by judges and magistrates. Garland, London

Hough M, Jacobson J, Millie A (2003) The decision to imprison: sentencing and the prison population. Prison Reform Trust, London

Jacobson J, Hough M (2007a) Mitigation: the role of personal factors in sentencing. Prison Reform Trust, London

Jacobson J, Hough M (2007b) Personal mitigation: an empirical analysis in England and Wales. In: Roberts JV (ed) Mitigation and aggravation at sentencing. Cambridge University Press, Cambridge

Lovegrove A (2010) The sentencing council the public's sense of justice and personal mitigation. Crim Law Rev July:769–781

Lubitz R (1996) Offender characteristics and departures under North Carolina's sentencing guidelines. Fed Sentencing Rep 9:132–135

Moxon D (1988) Sentencing practice in the crown court. Home Office, London

North Carolina Sentencing and Policy Advisory Commission (2007) Structured sentencing. Training and reference manual. NCSPAC, Raleigh

Padfield N (2011) Intoxication as a sentencing factor: mitigation or aggravation? In: Roberts JV (ed) Mitigation and aggravation at sentencing. Cambridge University Press, Cambridge

Roberts JV (2008) Punishing persistent offenders. Oxford University Press, Oxford

Roberts JV (2011a) Community views of sentencing: intuitive and principled responses to offending. In: Tonry M (ed) Punishment futures. Oxford University Press, Oxford

Roberts JV (ed) (2011b) Mitigation and aggravation at sentencing. Cambridge studies in law and society. Cambridge University Press, Cambridge

Roberts JV, Hough M (2011) Custody or community? Exploring the boundaries of public punitiveness in England and Wales. Criminol Crim Justice 11:185–202

Roberts JV, Hough M, Jacobson J, Bredee A, Moon N (2008) Public attitudes to sentencing offenders convicted of offences involving death by driving. Crim Law Rev July:525–540

Roberts JV, Hough M, Jacobson J, Moon N (2009) Public attitudes to sentencing purposes and sentencing factors: an empirical analysis. Crim Law Rev April:771–782

Sentencing Council of England and Wales (2011) Burglary offences. Definitive guideline. Sentencing Council of England and Wales, London

Sentencing Guidelines Council (2004) Overarching principles: seriousness. www.sentencingcouncil.org.uk

Sentencing Guidelines Council (2007) Reduction in sentence for a guilty plea. Definitive guideline. www.sentencingcouncil.org.uk

Shapland J (1981) Between conviction and sentence. The process of mitigation. Routledge & Kegan Paul, London

Thomas D (1979) Principles of sentencing, 2nd edn. Heineman, London

Utah Sentencing Commission (2007) Adult sentencing and release guidelines. Utah Sentencing Commission, Salt Lake City

von Hirsch A (2010) Proportionality and the progressive loss of mitigation: some further reflections. In: Roberts JV, Von Hirsch A (eds) Previous convictions and sentencing. Hart Publishing, Oxford

Walker N (1985) Sentencing. Theory, law and practice. Butterworths, London

Walker N (1999) Aggravation, mitigation and mercy in English criminal justice. Blackstone Press, London

Wisconsin Sentencing Commision (2007) Sentencing guidelines. http://wsc.wi.gov

Young W, King A (2011) Addressing problematic sentencing factors in the development of guidelines. In: Roberts JV (ed) Mitigation and aggravation at sentencing. Cambridge University Press, Cambridge

ML

► Money Laundering

Mob Killings

► Lawful Killings

Modern Marxist and Radical Theory

► Cultural Criminology

Modus Operandi of Sex Offenders

Eric Beauregard
School of Criminology, Simon Fraser University,
Burnaby, BC, Canada

Overview

Traditionally, the research on sex offenders has focused on the etiology of these crimes in order to inform treatment modalities as well as risk prediction practices. Needless to say that the classification of the individuals committing sexual crimes has attracted a lot of attention, first from clinicians and then from academics. The first typologies have mainly focused on the offender and his characteristics. But subsequent typologies have started to take into account the specific characteristics of the offense as well. These behavioral features were easier to observe, thus facilitating the classification of offenders based on these features. However, when one carefully examines the behavior of sex offenders, several problems emerge with the existing typologies. To begin with, typologies of the offending process focus on the aggressor, ignoring the entire criminal event – that is, the offender, the victim, and the context (Meier et al. 2001). Second, these typologies presuppose that the offending process, more specifically the modus operandi, does not fluctuate with the given sexual aggressor; they assume, that is, that such offenders always act in the same way, "specializing" in a single type of victim and remaining insensitive to situational factors and opportunities. Finally, these typologies were unable to account for the situational factors, the majority of them suggesting that individual characteristics (e.g., personality, sexual preferences, thoughts) are the principal factors influencing the offending process. In order to understand the modus operandi used by the offender – that is, the pattern of actions and behaviors prior to, during, and following the illicit act in order to perpetrate the offense successfully (Sutherland 1947) – and the different resulting outcomes at the crime scene, a different approach is needed.

From the Offender to the Criminal Event

In criminology, two complementary theoretical approaches are useful to examine and understand the modus operandi of sex offenders. The criminal event perspective (CEP) (Sacco and Kennedy 2002) is the term that has been used to describe a technique of organizing ideas and data (Meier et al. 2001). CEP is not itself a theory of criminality or criminal behavior but rather a tool that can be used to design explanatory models of crime that account for the importance of interactions (Anderson and Meier 2004). Thus, in order to understand the process and structure of crime, it is crucial to identify and understand the social context in which offenders live and interact with victims (Anderson and Meier 2004). According to Sacco and Kennedy (2002), the behavior of any one participant (offender and victim) in the criminal event intersects with and influences the behavior of other participants, shaping the course of the event and determining the stages through which it progresses and the extent to which it will be judged a serious one. CEP, then, can be used as an exploratory tool which organizes information about multiple elements of a criminal case or criminal event and which ultimately could lead to inductive theory building. Criminal events are different from criminal acts (Sacco and Kennedy 2002). Acts are instances of behavior, while events involve the context of the behavior. The major advantage of CEP is that it conforms to the way the world works; like all forms of social events, criminal events have a beginning and an end and occur over time in a sequential fashion (Sacco and Kennedy 2002).

Similarly, according to the *social interactionist* perspective, in the course of any personal crime, the behavior of one actor is shaped by the behavior of the other (Goffman 1967). Within the context of a sexual crime, this could mean that the victim's behavior depends – in whole or in part – on the offender's level of violence and coercion. Similarly, the offender may change his modus operandi depending upon the victim's perceived willingness or resistance. Luckenbill (1977) applied this concept of a social interaction between victims and perpetrators of homicide, referring to this exchange as a "collective transaction." Essentially, Luckenbill suggests that each participant develops a role within the criminal interchange; this role is shaped by the other actor and ultimately plays its own part in the resulting fatality. Similarly, Block (1981) examines the effect of the interaction between victim and offender on the outcome of violent crimes. However, rather than viewing the event as a "working agreement" between victim and offender, Block (1981) simply states the importance of the victim's role and actions on those of the offender. What takes place within the confines of the microenvironment surrounding the crime is most often a result of the actions of the victim and how those actions intersect with the strategies and modus operandi of the offender. Tedeschi and Felson (1994) further developed these ideas through their decision-making theory of coercive action, wherein harmdoing is goal-oriented behavior arising from social interactionist processes. The essence of goal-oriented behavior assumes a certain level of rationality within the decisions made during the course of a coercive encounter. Borrowing from the rational choice perspective (Cornish et al. 1986), choices regarding one's actions are made based on the perceived value of rewards (positive outcomes), the perceived value of costs (negative outcomes), and the estimated probabilities or likelihoods of the positive and negative outcomes achieving fruition. Self-preservation is intrinsic to human nature, and thus, all behavioral choices are made based on what is perceived as the best possible option in a particular scenario that is most likely to yield valued rewards.

Modus Operandi and Rational Choice

In line with the previous theoretical perspectives described above, the *modus operandi* (MO) is conceived as a dynamic process which needs to be adapted to situational circumstances (Cornish 1993). In criminology, the rational choice perspective is a theoretical framework specifically designed to address offender decision-making in the course of a specific crime (Cornish et al. 1986). The rational choice perspective assumes that criminals offend because crime provides the most effective means of achieving desired benefits (e.g., sexual gratification). As a corollary, the choice of methods for carrying out these crimes is best regarded as instrumental behaviors to achieve these goals. Offenders decide whether or not to commit a crime by weighing the effort, rewards, and costs involved in alternative courses of action. The making of decisions and choices, however rudimentary this process might sometimes be, exhibits a measure of rationality, albeit constrained by limits of time, ability, and the availability of relevant information (Johnson and Payne 1986). In fact, an offender's cost-benefit analysis may not take into account every possibility, both because such an analysis is time-consuming and because the relevant information is only partially available to him. In addition, an offender's cognitive deficits may restrict his information-processing concerning the crime (e.g., an offender intoxicated prior to the crime). Consequently, offenders usually rely on heuristics partially based on the success and failure of previous criminal activities, including previously used detection-avoidance strategies which did not lead to apprehension. The course of action selected by an offender is usually the first minimally satisfactory one identified, rather than the optimal one (Tversky and Kahneman 1981).

In departure from previous studies which classified sex offenders' behaviors into typologies, recent studies on the modus operandi of sex offenders have shown the complexity of the criminal event – more so when the context, the victim, and the dynamic aspects of the crimes are taken into account. Some recent studies have examined different outcomes related to sex offenders' modus operandi using different approaches.

Offenders' Reaction to Victim Resistance

The most influential action a victim can take within the criminal event is one of resistance. Varying levels and types of resistance can have a range of effects on the offender's modus operandi, especially with respect to how violent and coercive the offender becomes. Interactions between victim and offender become exceptionally important within this particular context, as the resistance itself is a reaction to the modus operandi of the offender. In fact, the resistance level employed by the victim has been found to closely match the offender's strategy to commit the assault (Ullman 2007). Thus, if an offender attempts to verbally coerce or threaten his victim, the victim will likely utilize screaming and pleading in an attempt to stop the sexual assault. However, if the offender physically attacks the victim, the victim is more likely to use a similar level of physical forcefulness to avoid the assault.

Victim resistance has been examined in its effects on two outcomes: victim injury and/or sexual assault completion. Although these are pertinent factors to investigate, two recent studies have shifted the focus from these more terminal outcomes to the more direct relationship between the victim's resistance and the offender's immediate reaction to that resistance. There could very well be a multitude of factors that come into play between when the victim resists and her later injuries (e.g., victim frailty) or whether the crime results in its completion (e.g., interference or not of a bystander). However, the dependent variable used for these two studies questions the *offender* (rather than the victim) as to his *direct response* to that resistance.

In the first study, Balemba et al. (2012) found that specific variables and interactions among variables increase the risk of sexual assault victims experiencing a violent reaction to their resistance. Overall, the most influential situational factor is the strategy used by the offender to commit the crime. If the offender uses a violent persuasive strategy, the most likely scenario is one of violence; even after interaction with other variables, violence remains the consistently predicted offender response. This relationship is logical, since it is reasonable to suppose that an offender who begins his crime in a violent manner is more likely to react violently to resistance from his victim. However, it is also plausible that situational factors influence violence and increase the offender's anger or modulate other factors that increase aggression. Tark and Kleck (2004) have found that when victims who resisted were hurt, it was almost always injury that came first, suggesting that the offender had decided to use violence before any interaction with the victim.

Moreover, the results found the type of victim resistance to be significant and to interact with offender strategy. The importance of this variable, and specifically its interaction with offender strategy, is consistent with previous findings indicating that resistance strategies of the victim tend to match offender strategies (Ullman 2007). Essentially, if the offender begins the assault using physical aggression, the victim is more likely to react with physical resistance. Finally, it was found that use of a weapon by the offender would lead to a violent reaction (Balemba et al. 2012).

The second study used a similar approach but this time distinguished the offender's reaction to victim resistance between adult and child victims. In this study, Balemba and Beauregard (2012) concluded that characteristics of the victim – often beyond their control – may increase their likelihood of victimization, the type of offender they will attract, and the amount of violence that may occur during the assault itself. Additionally, the victim's actions during the commission of the crime affect the offender's modus operandi, just as the offender's actions will affect the victim's (see for instance Block 1981). Thus, substantial differences in situational factors and interactions were found when comparing offenses against adult victims to those against child victims. In general, the findings show that adult victims are more likely to encounter violence from the offender than child victims, as has been previously shown in the literature (e.g., Scott and Beaman 2004; Weaver et al. 2004). Moreover, the findings suggest a different offense planning and stratagem, dependent upon the type of victim. In essence, it appears that when an offender chooses a child victim, there is a higher degree of preparation that occurs before the crime takes place. In contrast, an offense against an adult victim appears to be somewhat more impulsive and a reaction to the situation.

These results suggest that, with an adult victim, an offender is more likely to resort to violence according to the resistance level of the victim. This is a very important finding, especially because previous studies have found the resistance of the victim to have no effect on the physical forcefulness of the offender (Ullman 2007). However, previous studies have consistently suffered from an inability to distinguish a temporal sequence in the event. If, for example, the offender begins with a violent offense strategy and the victim reacts physically, this is a much different scenario than if an offender reacts violently to a victim's physical resistance. Such a distinction has been made in these studies, as the dependent variable was specifically designed to represent, as the name states, the offender's reaction to victim resistance.

There appears to be a completely different social environment surrounding a sexual assault of a child. According to the findings, such assaults seem to be premeditated to a greater degree in an attempt to circumvent the use of expressive violence to complete the assault. Offenders against children are prepared, with regard to the surrounding setup and physical environment, and are aware of the amount of force they are likely and willing to use prior to the commencement of the assault (thus, the importance of the strategy to commit the crime). These interpretations are congruent with findings showing that planning is an important step, either implicitly or explicitly, in the offense process of child molesters (Ward et al. 1995). This planning often involves an assessment of victim vulnerability and chance of apprehension (Leclerc and Tremblay 2007). Furthermore, the notion of offenders "grooming" children as a way of normalizing or legitimizing sexual contact also corroborates the assertion that the element of premeditation is an important factor within the offending strategy of sexual abusers of children.

The social interactionist perspective emphasizes the true importance of behaviors of the various players within a coercive interchange (Luckenbill 1977; Tedeschi and Felson 1994). As evidenced by the findings within the two studies previously discussed, offender modus operandi has the potential to drastically alter victim behavior, especially the level of victim resistance. However, the current analyses have shown the reverse to be true as well: victim behavior also considerably affects offender modus operandi. Behavior appears to be cyclical in the sense that actors continually affect one another: offender assaultive behavior leads to victim resistance, which, depending upon the presence or degree of certain factors, may then lead to offender coercion in response to that resistance.

Victim's Death

Sometimes during the interaction between the offender and the victim in a sexual assault event, the victim does not survive. Research has shown that the likelihood of a sexual assault becoming a homicide differs mainly based on victim characteristics and circumstantial factors. For instance, Felson and Messner (1996) found that the likelihood of the victim suffering serious injury or death was higher when sexual offenders used a weapon. Mieczkowski and Beauregard (2010) identified particular combinations of characteristics differentially associated with sexual assaults that end in the homicide of the victim. Using the method of conjunctive analysis (Miethe et al. 2008), they highlight those attributes in the victim, the situation, and the modus operandi which are associated with a lethal outcome. In brief, what was found in this analysis is that the most telling domain for a lethal outcome is contained within the modus operandi. For instance, the most lethal combination is when the offender uses a weapon and spends more than 30 min with the victim but does not commit intrusive sexual acts on the victim nor force the victim to commit sexual acts on him (ratio approaching 25 to 1). However, the weakest differentiation is seen in the *situational characteristics*. The domain of *victim characteristics* occupies an intermediate level. It has a relatively high number of combinations – five – which attain a ratio value greater than 1. Of these five combinations, the most lethal attains a value of 4.53, which is characterized by a victim aged under 14, stranger to the offender, and not from a criminogenic environment. These findings, although still preliminary, demonstrate the importance of looking at combination of factors instead of individual factors and the "practical" usefulness of the CEP.

Sex Offenders' Criminal Mobility and Choices of Crime Location

A different outcome which can be influenced by the modus operandi of sex offenders is the geography of the crime. More specifically, recent studies have looked at sex offenders' criminal mobility as well as the choice of crime location for these same offenders. Beyond the basic measure of the journey to crime, some authors have attempted to further our understanding of the criminal mobility of sex offenders by examining two types of factors influencing their journey to crime: the offender and the modus operandi. Here we are particularly interested in the latter factor, the modus operandi. Lebeau (1987) focused on how the journey to rape varies as a function of the offender's approach method. Results revealed that offenders traveled the shortest mean distance to assault their victims when they illegally entered the victim's residence. In a different study, Canter and Gregory (1994) found that rapists who offend during the weekend travel farther than those who commit rape during a weekday and that rapists who attack outdoors traveled approximately 2.7 times farther to offend as those who raped indoors (e.g., in a house). Davies and Dale (1995) suggested also that rapists who target victims from a particular area (e.g., prostitutes from a red-light district), who commit sophisticated property offenses during a sexual assault, who spend large amounts of time roaming and using public transportation, and who are familiar with numerous neighborhoods (previous habitation, locations of significant people, current or past workplace locations) travel longer distances to commit their crimes. This was echoed by Warren et al. (1998) who found similar results in relation to the sophistication of the crime. Their results showed that rapists who used forced entry and who burglarized the victim during the assault tended to travel farther. Although interesting, it appears that all the studies have taken for granted that journey to crime was the only and probably the most adequate measure of criminal mobility. Whether or not this is true, none of these studies have questioned the use of the journey to crime to measure their criminal mobility.

In order to go beyond the journey-to-crime measure, criminal mobility has been conceptualized as the use of multiple locations for the purpose of repetitive sexual contact with the same victim. In their study, Leclerc et al. (2010) set out to examine whether offending differences existed between perpetrators who used multiple locations for sexual contact and those who used a single location for the entire crime-commission process. Overall, the results demonstrate that mobile offenders are more likely to isolate their victims, use violence, involve the victim in several sexual episodes, abuse the victim for over a 1-year period, and make the victim participate and perform sexual behaviors on them during sexual episodes. The authors concluded that by examining mobility of pedophiles from a location angle rather than measures of distance and direction provides a different perspective on the crime-commission process of these offenders (Leclerc et al. 2010).

Following Leclerc et al. (2010) study, Beauregard and Busina (2012) used a similar approach with serial sex offenders. They proposed that criminal mobility can be defined as the number of changes of location during the criminal event. As discussed previously, rape events present different stages – that is, encounter, attack, crime, and victim release – that may be associated with different locations. Although some sex offenders decide to commit all their action at the same location (i.e., stable offender with zero change of location), other mobile offenders may change location up to three times during the same event. As criminal mobility can be interpreted as a purposive action necessary to successfully commit a crime, the aim of their study was to predict the criminal mobility patterns exhibited in serial rape events from situational and modus operandi characteristics. The situational characteristics of the rape events and the modus operandi used in serial sex crimes might explain why some offenders need to be mobile and change location during the criminal event while others do not. Using negative binomial regression, the authors found that events which involve child or adolescent victims, committed during daytime, when the offender did not use pornography prior to crime, and where victim resistance is observed should display more criminal mobility. Moreover, when the victim is selected, the victim is alone when approached by the offender, and the crime is characterized by sexual penetration and a lack of premeditation is exhibiting more criminal mobility. These results point toward the fact that criminal mobility is a goal-oriented action taken by serial sex offenders in order to complete successfully their crime and to avoid detection and apprehension (Beauregard and Busina 2012).

As mentioned previously, much research in the field of spatial decision-making has concentrated on distance to crime, but less focus has been on those factors that affect the types of environmental locations where crimes are likely to take place. As rapes are complex crimes largely influenced by contextual factors, environmental variables, and offender decision-making (Beauregard et al. 2010), it is necessary to examine these factors when determining offenders' choice of crime site locations. The study by Hewitt, Beauregard, and Davies (Hewitt et al. 2012) analyzes temporal factors, offender hunting behavior, and modus operandi strategies in a sample of 361 sexual crimes committed by 72 serial rapists, to determine their utility in predicting the type of location where victim encounter and release sites are likely to be, as these two sites are the most likely to be known by the police during the investigation (Rossmo 2000). Findings showed that temporal factors, hunting behavior, and modus operandi variables are all important to the victim encounter and release sites, but the significance of these factors varies depending on whether or not the location is in a residential land-use area, private site, inside location, and familiar to the offender. For instance, the study shows that in terms of offender modus operandi strategies, victims are more likely to be encountered in residential land-use areas, private sites, and inside locations when the crime is structurally premeditated. These findings suggest that offenders purposefully hunt for victims in these locations because they know that they will find victims who satisfy their criteria. As noted in the *home-intrusion rape track* proposed by Beauregard and colleagues (2007), offenders following this script enter a private, inside location that is unfamiliar to them (e.g., the victim's residence). Thus, crimes taking place in private or inside locations require more premeditation on the offenders' behalf as they are entering a space that is removed from the public domain. Because the offender does not belong in these locations, it requires more effort and planning on his part to enter these spaces unlawfully and be able to successfully commit the crime in areas where he is considered to be an outsider. Moreover, offenders who structurally premeditate their crimes tend to release their victims in an inside location. As rapists who premeditate their crimes are more common than those who do not (Rossmo 2000), many offenders select their victims by window peeping or following women on their way home, which suggests some degree of planning on behalf of the perpetrator. As such, the victim attack, crime, and release locations tend to be at the victim's residence, which in many cases is classified as an inside location. In addition, victims of sexual assault where the rapist uses a vehicle to commit the crime are less likely to be released in a private site or inside location than those victims where the offender does not use a vehicle. As Beauregard and Leclerc (2007) note, sex offenders do not usually use vehicles throughout the commission of their crimes for several reasons, including lack of a driver's license or car, or the rape takes place in an inside location, thereby not necessitating the use of a vehicle. However, in crimes where rapists do utilize a vehicle, they may do so because it helps them to search for victims and provides them with a private site to rape the victim and/or a means to move the victim from one location to another (e.g., moving her from the rape site to the release site; Beauregard and Leclerc 2007). By using a vehicle as the rape site and/or to move the victim from one geographic location to another, the rapist may release the victim in a variety of land-use or public spaces.

These studies illustrate not only the complex dynamics of sexual assault events but also how the modus operandi strategies used by the offenders may impact the different geographical outcome of the crime, that is, the decision to move during the event or the decision as to where to encounter and release the victim.

Forensic Awareness

One of the stages of the criminal event which has been understudied is the aftermath – that is, what are the actions and behaviors of the offenders after committing the crime. One of these modus operandi strategies which is of great interest for the criminal investigation is the use of particular precautions by sex offenders in order to avoid getting caught. When committing their crimes, offenders take actions that inherently have the potential to leave evidence at the crime scene (e.g., breaking and entering into a residence may leave fingerprints, ejaculating during a rape may leave DNA). Despite all these potential opportunities, unavailability of many forms of forensic evidence has been observed at crime scenes. In fact, research has shown that physical evidence is being collected in less than 10 % of cases investigated by the police (Horvath and Meesig 1996). Although little empirical evidence has been found for the "*CSI* effect" – by which juries supposedly will no longer convict without forensic evidence (Podlas 2006) – some authors have contended that perhaps a more serious aspect of the availability of all these TV shows is the knowledge gained by potential criminals (Durnal 2010). For example, it has been noted that offenders are increasingly wearing gloves to avoid leaving fingerprints, using tape to seal envelops to avoid leaving DNA, and even using bleach (which destroys DNA) to clean up blood (Stevens 2008).

Some offenders will adapt their modus operandi or take precautions before, during, or post crime commission to decrease their risk of apprehension. Arguably, this adaptation of crime strategy may be deemed an indication of evolving criminal sophistication on the part of the offender. Offenders who adapt their modus operandi to thwart police investigative efforts may be said to be exhibiting investigative awareness (i.e., a knowledge or understanding of police investigative practice). A related concept coined in criminological research as "forensic awareness" (Davies 1992) refers to an offender's knowledge or understanding of the importance of forensic evidence (e.g., DNA, fingerprints, dental impressions) to police investigation.

There is a dearth of research on the extent of forensic awareness among offenders and the impact that this awareness has on the crime process and investigation. A small number of studies, specific to sexual crime, have addressed the issue of forensic awareness and provide some insight on how certain offender behaviors may be indicative of forensic and/or investigative awareness. Davies and Dale (1995) studied stranger rapists and suggested that traveling longer distances to commit an offense may be indicative of forensic awareness. Beauregard and Field (2008) identify a precaution (i.e., moving a victim's body post murder) that may be undertaken to delay apprehension but may also be considered an indicator of both investigative and forensic awareness. Moving the body complicates the investigation by decontextualizing the crime, potentially decreasing the likelihood that the victim will be found or, if found, identified. This action may also indicate forensic awareness as the offender is removing a significant source of forensic evidence from the scene of the homicide. In their examination of 222 stranger sexual assaults, Beauregard and Bouchard (2010) found that certain modus operandi strategies (i.e., nonrandom selection of a victim, selecting a victim who is alone, and hunting for a victim in certain locations, such as a prostitution stroll), arguably indicators of investigative awareness, were related to forensic awareness, which was evident at the crime scene. Furthermore, offenders who break and enter the homes of their victim and exhibit certain sexual behaviors (i.e., penetrate the victim and ejaculate) are more likely to exhibit forensic awareness. A negative relationship was found between forensic awareness and the use of drugs and/or alcohol prior to commission of the offense (Beauregard and Bouchard 2010).

Conclusion

Although typologies based on offenders' behavior have been useful to classify offenders for the purpose of suggesting treatment modalities and assessing the risk, recent studies on the modus operandi of these sex offenders have shown the complex dynamics in the criminal event. The findings reveal that the modus operandi decided by an offender may change during the criminal event, depending on the context, more specifically on the victim's behavior. But in order to uncover these complex dynamics, it is important to look at the modus operandi of sex offenders through a criminal event perspective, where the offender, the victim, and the context of the crime are taken into account. Such an approach requires not only a theoretical shift in the study of sex offenders' modus operandi but also a shift in the analytical methods used. Methods which can take into account the multiple interactions between the variables under study (e.g., conjunctive analysis, regression tree) should be prioritized over classification methods (e.g., cluster analysis). These new approaches help to gain a better understanding of the sex offender's decision-making during the criminal event as well as better predict different outcomes within the criminal event. The modus operandi is a purposive set of actions taken by an offender which will inevitably lead to an outcome (e.g., death of the victim of not). But to understand why certain offenders are associated to one outcome in particular, it is important to be able to look at the dynamics involved during the criminal event and examine which factors can influence the offender's modus operandi.

Related Entries

- ► Crime Specialization, Progression, and Sequencing
- ► Integrating Rational Choice and Other Theories
- ► Offender Decision Making and Behavioral Economics
- ► Rational Choice Theory
- ► Specialization and Sexual Offending

Recommended Reading and References

Anderson AL, Meier RF (2004) Interactions and the criminal event perspective. J Contemp Crim Justice 20:416–440

Balemba S, Beauregard E (2012) Reactions to resistance: the role of contextual factors in sex offending. Violence Vict 27:148–165

Balemba S, Beauregard E, Mieczkowski T (2012) To resist or not to resist? The effect of context and crime characteristics on sex offenders' reaction to victim resistance. Crime Delinq 58:588–611, OnlineFirst

Beauregard E, Bouchard M (2010) Cleaning up your act: forensic awareness as a detection avoidance strategy. J Crim Justice 38:1160–1166

Beauregard E, Busina I (2013) Journey "during" crime: Predicting criminal mobility patterns in rape. J Interpers Violence. OnlineFirst.

Beauregard E, Field J (2008) Body disposal patterns of sexual murders: implications for offender profiling. J Police Crim Psychol 23:81–89

Beauregard E, Leclerc B (2007) An application of the rational choice approach to the offending process of sex offenders: a closer look at the decision-making. Sex Abuse 19:115–133

Beauregard E, Proulx J, Rossmo K, Leclerc B, Allaire J-F (2007) Script analysis of the hunting process of serial sex offenders. Crim Justice Behav 34:1069–1084

Beauregard E, Rebocho MF, Rossmo K (2010) Target selection patterns in rape. J Investig Psychol Offender Profiling 7:137–152

Beaver KM (2010) The promises and pitfalls of forensic evidence in unsolved crimes. Criminol Public Policy 9:405–410

Block R (1981) Victim-offender dynamics in violent crime. J Crim Law Criminol 72(2):743–761

Canter D, Gregory A (1994) Identifying the residential location of rapists. J Forensic Sci Soc 34:169–175

Cornish DB (1993) Theories of action in criminology: learning theory and rational choice approaches. In: Clarke RV, Felson M (eds) Routine activity and rational choice. Transaction, New Brunswick, pp 351–382

Cornish DB, Clarke RV et al (1986) Introduction. In: Cornish DB, Clarke RV (eds) The reasoning criminal: rational choice perspectives on offending. Springer, New York

Davies A (1992) Rapists' behaviour: a three aspect model as a basis for analysis and the identification of serial crime. Forensic Sci Int 55:173–194

Davies A, Dale A (1995) Locating the stranger rapist, Special interest series: paper 3. Police Research Group, Home Office Police Department, London

Durnal EW (2010) Crime scene investigation (as seen on TV). Forensic Sci Int 199:1–5

Felson RB, Messner SF (1996) To kill or not to kill? Lethal outcomes in injurious attacks. Criminology 34:519–545

Goffman E (1967) Interaction ritual. Pantheon, New York

Hewitt A, Beauregard E, Davies G (2012) "Catch and release": predicting encounter and victim release location choice in serial rape. Polic Int J Police Strateg Manag 35:835–856, OnlineFirst

Horvath F, Meesig R (1996) The criminal investigation process and the role of forensic evidence: a review of empirical findings. J Forensic Sci 41:963–969

Johnson E, Payne J (1986) The decision to commit a crime: an information processing analysis. In: Cornish DB, Clarke RV (eds) The reasoning criminal: rational choice perspectives on offending. Springer, New York, pp 170–185

Lebeau JL (1987) The journey to rape: geographic distance and the rapist's method of approaching the victim. J Police Sci Adm 15:129–136

Leclerc B, Tremblay P (2007) Strategic behavior in adolescent sexual offenses against children: linking modus operandi to sexual behaviors. Sex Abuse 19:23–41

Leclerc B, Wortley R, Smallbone S (2010) Investigating mobility patterns for repetitive sexual contact in adult child sex offending. J Crim Justice 38:648–656

Luckenbill DF (1977) Criminal homicide as a situated transaction. Soc Probl 25(2):176–186

Meier RF, Kennedy LW, Sacco VF (2001) Crime and the criminal events perspective. In: Meier RF, Kennedy LW, Sacco VF (eds) The process and structure of crime: criminal events and crime analysis. Transaction, New Brunswick, pp 1–28

Mieczkowski T, Beauregard E (2010) Lethal outcome in sexual assault events: a conjunctive analysis. Justice Q 27:332–361

Miethe TD, Hart T, Regoeczi W (2008) The conjunctive analysis of case configurations: an exploratory method for discrete multivariate analysis of crime data. J Quant Criminol 24:227–241

Podlas K (2006) The "CSI effect" and other forensic fictions. Loyola Los Angel Entertain Law Rev 27:87–125

Rossmo DK (2000) Geographic profiling. CRC Press, Boca Raton

Sacco VF, Kennedy LW (2002) The criminal event: perspectives in space and time, 2nd edn. Wadsworth, Belmont

Scott HS, Beaman R (2004) Demographic and situational factors affecting injury, resistance, completion, and charges brought in sexual assault cases: what is best for arrest? Violence Vict 19(4):479–494

Stevens DJ (2008) Forensic science, wrongful convictions, and American prosecutor discretion. Howard J 47:31–51

Strom KJ, Hickman MJ (2010) Unanalyzed evidence in law-enforcement agencies: a national examination of forensic processing in police departments. Criminol Public Policy 9:381–404

Sutherland EH (1947) Principles of criminology. J. B. Lippincott, New York

Tark J, Kleck G (2004) Resisting crime: the effects of victim action on the outcomes of crime. Criminology 42:861–909

Tedeschi JT, Felson RB (1994) Violence, aggression, and coercive actions. American Psychological Association, Washington, DC

Tversky A, Kahneman D (1981) The framing of decisions and the psychology of choice. Science 211:453–458

Ullman SE (2007) A 10-year update of "review and critique of empirical studies of rape avoidance". Crim Justice Behav 34(3):411–429

Ward T, Louden K, Hudson SM, Marshall WL (1995) A descriptive model of the offense chain for child molesters. J Interpers Violence 10(4): 452–472

Warren J, Reboussin R, Hazelwood RR, Cummings A, Gibbs N, Trumbetta S (1998) Crime scene and distance correlates of serial rape. J Quant Criminol 14:35–59

Weaver GS, Wittekind JE, Huff-Corzine L, Corzine J, Petee TA, Jarvis JP (2004) Violent encounters: A criminal event analysis of lethal and nonlethal outcomes. J Contemp Crim Just 20(4):348–368

Moffitt's Developmental Taxonomy of Antisocial Behavior

Alex R. Piquero[1] and Terrie E. Moffitt[2]
[1]Department of Criminology, The University of Texas at Dallas, Richardson, TX, USA
[2]Departments of Psychology and Neuroscience, Psychiatry, and Behavioral Sciences, Institute for Genome Sciences and Policy, Duke University, Durham, NC, USA

Overview

This entry provides an overview of Moffitt's developmental taxonomy, highlights key findings from subsequent research, identifies critical challenges to the taxonomy, and outlines several important directions for future research. The entry concludes by noting that some of the taxonomy's key hypotheses have been supported by empirical research, while other findings have presented challenges to the taxonomy.

Introduction

The age/crime relationship is one of the brute facts of criminology, which any theory of crime must be able to explain. The curve evidences a slow rise during early adolescence, peaks in mid to late adolescence, and then decreases rapidly in early adulthood. The fact of the curve is unassailable; however, what is especially contentious is the reason(s) underlying this consistently documented correlate of crime. One theory in particular, Terrie Moffitt's (1993) developmental taxonomy, seeks to develop a better understanding of the age/crime relationship by focusing on the mixture of two distinct groups of offenders, whose offender styles combine to represent the overall aggregate relationship – yet with unique age/crime curves. This entry provides a background description of Moffitt's theory, provides a very broad review of key findings, identifies controversies that exist with respect to the theory and empirical research, highlights open questions, and provides an overall conclusion.

Background Description

In her original statement of the developmental taxonomy, Moffitt suggested that the age/crime relationship could be summarized by the existence of two distinct types of offenders, each of whom evince unique age/crime patterns and who offend for distinct reasons. The first group, adolescence-limited (AL), is comprised by a large segment of the offending population who constrict their antisocial behavior to the adolescent time period. For ALs, delinquency is a function of two features of adolescence: the "maturity gap" and the peer social context. The maturity gap recognizes that adolescents biologically resemble adults but are not socially or legally recognized as adults and are consequently denied most adult privileges and responsibilities. Young people's recognition of this mismatch between their social maturity and their biological maturity leads to a feeling of strain that is also characteristic of their similarly situated peers. Because peers are an exceedingly influential socializing agent during adolescence, and because all ALs are negotiating the adolescent period and experiencing the same sets of strain, they encounter potential role models in other older-aged peers who have alleviated their sense of strain by becoming involved in delinquent activities that bring them forbidden adult pleasures. Life-course-persistent age peers, as will be seen below, also constitute delinquent role models. Thus, ALs engage in delinquent activities that resemble adult status such as smoking, drinking, sexual activity, and theft (so as to obtain economic resources and goods) in order to alleviate their strain and acquire a sense of being recognized and treated like adults. As they exit out of adolescence and enter emerging and early adulthood, and because they are now legally permitted all of the things they once coveted, most but not all AL offenders cease their antisocial behavior. It is expected that most ALs will successfully transition into adult roles and responsibilities and leave their antisocial ways behind. However, a subset of ALs may encounter snares as a result of their antisocial behavior, such as drug addiction, school dropout, or a criminal record, that may entrap them in a delinquent lifestyle and delay desistance.

Unlike their AL counterparts, LCP offenders comprise a very small group of individuals (<10 % of males), who exhibit antisocial tendencies and behaviors very early in the life course (i.e., early childhood). The antisocial repertoire of LCPs is of a varied nature, chronic, persistent, and serious and spans distinct phases of the life course (childhood – adolescence – adulthood) and continues largely unabated via different manifestations and across various life domains well into adulthood. Unlike the more social origins of antisocial behavior among ALs, LCP-style antisocial behavior originates from compromised neuropsychological development that is exacerbated by disadvantaged familial and economic environments. Specifically, LCPs tend to be born with neurodevelopmental problems of weak cognitive abilities and low self-control that exert negative effects on their immediate caretakers. If caretakers are unskilled at parenting, these early deficiencies are not addressed nor corrected. As a result, LCPs become involved from early life in a series of antisocial behaviors that eventually produce negative outcomes in virtually all aspects of their lives, including failures in education, interpersonal relationships, and employment. Thus, their antisocial behavior gradually becomes part of their overall personality. This developmental process of cumulative continuity results in adverse reactions by prosocial agents and produces barriers to, or knives off, conventional opportunities for success. The neglected injurious childhoods, poor neurocognitive abilities, and resultant antisocial lifestyles of LCPs limit their prospects for change.

Although Moffitt's taxonomy focused on two distinct groups of offenders, it also describes a third group of young people who stand in stark contrast to the more normative, adaptational ALs: a group of teens who abstain from delinquency. Moffitt views these "abstainers" as somewhat extraordinary because they eschew delinquency during a period of the life course when delinquency is the norm. Abstainers are viewed as encountering barriers that preclude their foray into delinquency, either through structural barriers that prevent them from learning about delinquency, no maturity gap because of early access to adult roles, or personal characteristics unappealing to other teens that cause them to be excluded from teen social group activities.

Any serious contender for a theory of delinquency is obliged to address two of the most salient facts about delinquency: sex variation and ethnic-minority variation. The original statement of Moffitt's taxonomy asserted that the theory describes the behavior of females as well as it describes the behavior of males. Thus, it is important to discuss Moffitt's (1994, pp. 39–40) specific expectations of the role of gender in the taxonomy:

> The crime rate for females is lower than for males. In this developmental taxonomy, much of the gender difference in crime is attributed to sex differences in the risk factors for life-course persistent antisocial behavior. Little girls are less likely than little boys to encounter all of the putative initial links in the causal chain for life-course persistent antisocial development. Research has shown that girls have lower rates than boys of symptoms of nervous system dysfunction, difficult temperament, late verbal and motor milestones, hyperactivity, learning disabilities, reading failure, and childhood conduct problems... Most girls lack the personal diathesis elements of the evocative, reactive, and proactive person/environment interactions that initiate and maintain life-course persistent antisocial behavior.
>
> Adolescence-limited delinquency, on the other hand, is open to girls as well as to boys. According to the theory advanced here, girls, like boys, should begin delinquency soon after puberty, to the extent that they (1) have access to antisocial models, and (2) perceive the consequences of delinquency as reinforcing... However, exclusion from gender-segregated male antisocial groups may cut off opportunities for girls to learn delinquent behaviors... Girls are physically more vulnerable than boys to risk of personal victimization (e.g., pregnancy or injury from dating violence) if they affiliate with life-course persistent antisocial males. Thus, lack of access to antisocial models and perceptions of serious personal risk may dampen the vigor of girls' delinquent involvement somewhat. Nonetheless, girls should engage in adolescence-limited delinquency in significant numbers....

The original theory thus proposed that (a) fewer females than males would become delinquent (and conduct disordered) overall and that (b) within delinquents the percentage who are LCP would be larger among males than females. Following from this, (c) the majority of delinquent females will be of the AL type, and further, (d) their delinquency will have the same causes as AL males' delinquency (Moffitt 2006).

Moffitt's (1994, p. 39) taxonomic writings also anticipated that the taxonomy would apply to socially disadvantaged ethnic-minority populations as well as to whites:

"In the United States, the crime rate for black Americans is higher than the crime rate for whites. The race difference may be accounted for by a relatively higher prevalence of both life-course persistent and adolescence-limited subtypes among contemporary African-Americans. Life-course persistent antisocials might be anticipated at elevated rates among black Americans because the putative root causes of this type are elevated by institutionalized prejudice and by poverty. Among poor black families, prenatal care is less available, infant nutrition is poorer, and the incidence of exposure to toxic and infectious agents is greater, placing infants at risk for the nervous system problems that research has shown to interfere with prosocial child development. To the extent that family bonds have been loosened and poor black parents are under stress,. . .and to the extent that poor black children attend disadvantaged schools. . ., for poor black children the snowball of cumulative continuity may begin rolling earlier, and it may roll faster downhill. In addition, adolescence-limited crime is probably elevated among black youths as compared to white youths in contemporary America. If racially-segregated communities provide greater exposure to life-course- persistent role models, then circumstances are ripe for black teens with no prior behavior problems to mimic delinquent ways in a search for status and respect. Moreover, black young people spend more years in the maturity gap, on average, than whites because ascendancy to valued adult roles and privileges comes later, if at all. Legitimate desirable jobs are closed to many young black men; they do not often shift from having "little to lose" to having a "stake in conformity" overnight by leaving schooling and entering a good job. Indeed, the biological maturity gap is perhaps best seen as an instigator of adolescent-onset delinquency for black youths, with an economic maturity gap maintaining offending into adulthood."

Thus, the taxonomy expected that both LCP and AL processes should work the same way within African-American and white American groups, but any excess of offending among poor African-American youth could be attributed to an excess of the risk factors for both delinquent subtypes (Moffitt 2006).

The original theoretical taxonomy asserted that two main trajectories of offenders (LCP and AL) account for much of the age/crime relationship and thus warrant the lion's share of attention by theory and research. As a result of subsequent empirical studies, a small but additional group of offenders have been identified that does not fit either of the other two offender groups. First identified in trajectory analyses of a British cohort (Nagin et al. 1995), this third group of offenders has been labeled "low-level chronics" because they have been found to offend persistently but at a low rate from childhood to adolescence or from adolescence to adulthood.

State of the Art

There has been a very active set of studies seeking to explore various aspects of Moffitt's developmental taxonomy and the results from this research have been effectively summarized in several publications (see Moffitt 2006; Piquero and Moffitt 2005). Here, a very brief overview of the findings regarding key hypotheses from the taxonomy is presented.

Since the publication of the taxonomy in 1993, it can be stated with a good degree of certainty that the hypothesized LCP antisocial individual exists, at least during the first three decades of life. Near universal agreement about this group has emerged from all studies that have applied trajectory-detection analyses to a representative cohort sample having longitudinal repeated measures of antisocial behavior (Piquero 2008). In fact, it remains the case that no research that has looked for a persistent antisocial group has failed to find it. A number of other postulates from the taxonomy have received empirical support, ranging from robust to modest, but overall they appear to be supported and are listed here.

First, there is good evidence that LCP antisocial behavior emerges from early neurodevelopmental and family-adversity risk factors, but AL delinquency does not. Second, many studies show genetic etiological processes contribute more to LCP than AL antisocial development. Third, longitudinal follow-ups reveal that childhood-limited aggressive children, if followed to adulthood, become low-level chronic criminal offenders with personality disorders. Fourth, there is some evidence that abstainers from delinquency are rare individuals, who become unpopular with teen peers. Fifth, LCP and AL delinquents have been shown to develop different personality structures by adulthood. Sixth, LCP development is differentially associated with serious offending and violence in adulthood in all studies. Seventh, epidemiological studies confirm that LCP antisocial development is almost exclusively male, whereas most female antisocial behavior is of the AL type.

Some findings have received beginning support, but more research is needed. For example, AL antisocial behavior is influenced by the maturity gap between childhood and adulthood and by social mimicry of antisocial role models. Also, childhood-onset antisocial behavior persists at least into middle adulthood, whereas adolescent-onset antisocial behavior desists in young adulthood. Recent studies report that LCP antisocial individuals are at high risk in midlife for poor physical health, cardiovascular disease, and early disease morbidity and mortality (see Odgers et al. 2007; Piquero et al. 2010, 2011).

Some predictions from the taxonomy have received only minimal empirical attention. For example, according to the theory, AL offenders must rely on peer support for crime, but LCP offenders should be willing to offend alone (although in adolescence they serve as magnets for less expert offenders). Additionally, "snares" (such as a criminal record, incarceration, addiction, or truncated education without credentials) should explain variation in the age at desistence from crime during the adult age period, particularly among AL offenders. Lastly, the two groups should react differently to turning point opportunities: ALs should get good partners and jobs that help them to desist from crime, whereas LCPs should selectively get undesirable partners and jobs and in turn expand their repertoire into domestic abuse and workplace crime. These predictions need testing.

Possible Controversies in the Literature

As is true of many theories of antisocial behavior, results from empirical research provide some challenges and controversies for the original statement of the theory. Here, the single most important challenge to the taxonomy is highlighted. Two life-course criminology theorists and researchers, Robert Sampson and John Laub, have called into the question not only whether there are indeed LCP and AL offenders and whether such groups could be prospectively identified in childhood, but also whether the two groups follow their hypothesized crime trajectories.

Sampson and Laub (2003) conducted a follow-up study of a sample of 500 adolescent male inmates from the 1940s, covering the period from age 7 years to the end of each offender's life up to age 70 years. Sampson and Laub reported two findings from their analysis that challenged the taxonomy. First, they found that almost all of the men desisted from criminal offending sooner or later, and second, they found heterogeneity in adulthood crime career patterns within the sample of adolescent inmates, and they found that this heterogeneity was not explained by measures of childhood risk.

Moffitt (2006) has observed that Sampson and Laub's findings are potentially premature and perhaps overstated for several reasons. First, virtually all of the men in the sample might have been regarded as candidates for the LCP subtype (i.e., they had been incarcerated as young adolescents as inmates in reform schools; they had backgrounds of marked family adversity, social disadvantage, and childhood antisocial conduct, low IQ, and an early first arrest). Second, Moffitt argued that Sampson and Laub misrepresented the taxonomy's prediction because they set up a "straw" prediction to test, i.e., that LCP offenders would continue committing crimes at the same high rate from adolescence through old age, and to their deaths. Moffitt's taxonomy had not implied this expectation, largely because of the well-known population-wide process of aging out of crime in midlife. The taxonomy claims that LCP-type delinquents would continue offending well beyond the age when most young men in their cohort would ordinarily have stopped offending and ought to retain detectable features of an antisocial personality into late life (such as lying, irresponsibility, or callous attitudes), but perhaps not criminal offending itself.

With respect to Sampson and Laub's second finding of heterogeneity in adulthood crime career patterns, this again is viewed by Moffitt as a "straw man." The alternative would be that males who spent their youth and early adulthood on the LCP pathway can show no variation in subsequent offending during midlife and aging, over a span of many years. Such uniformity is implausible, and no such prediction was made in the taxonomy. As Sampson and Laub argue, child and family characteristics were unable to distinguish between the different offender trajectories they identified. But Moffitt contends that this failure of discrimination is not unexpected in the sample they studied, because the childhood backgrounds of the males were almost uniformly high risk. At the same time, this particular result suggests that to the extent that different crime careers emerge during midlife within a group of LCP men, concurrent life experiences must account for the divergence. This would constitute an interesting extension to the taxonomy: heterogeneity within LCPs in the ways they age out of crime. In any case, the main point Sampson and Laub made was that it is a risky business to predict that a child is doomed to a life of crime, and thus early identification advocates may have gone too far. Moffitt's work provides consensus on this important point about potential misuse of the taxonomy.

Open Questions

Although there has been extensive research into Moffitt's taxonomy, open questions remain. Here, a few of these questions that are important for subsequent inquiry are provided.

First, there have been few empirical investigations regarding the extent to which the taxonomy explains the association of key demographics of crime, namely, that of race and gender. Although Moffitt anticipated that the theory would explain the antisocial behavior of males and females as well as whites and socially disadvantaged nonwhites (especially African-Americans), the lack of data and requisite measures of key predictors has prohibited extensive analyses on these associations.

Second, while Moffitt clearly advanced some sort of neurocognitive deficit among LCP offenders, she also noted the relevance of a sociobiological disjuncture to predict patterns of offending among AL offenders: the maturity gap. Thus, the role of biology and genetics, especially in concert with social environments, is an important aspect of Moffitt's taxonomy. Yet, due to data constraints, there has been little investigation of the taxonomy's implications about joint effects of biological and social factors and the extent to which such factors help explain variation across offenders and in offending.

Third, there has been virtually no research exploring the decision-making process across the two offender types. How do LCPs weigh the risks and rewards of offending at the point of decision-making? Do AL offenders consider the costs more than their LCP counterparts? Do AL offenders alter their decision-making criteria as they enter adulthood? Different personality styles and different cognitive abilities characterize the two groups, which imply that they should differ on their typical decision-making strategies too. Integration of research from deterrence/rational choice would provide a good extension of Moffitt's theory to consider situational factors.

Conclusions

Moffitt's developmental taxonomy was originally advanced as a framework to help explain and understand the age/crime relationship. Since its publication in 1993, it has generated much interest and research. Some findings have been consistent with some of the key, original hypotheses, some findings have pointed to important revisions needed to improve the fit between the taxonomy and empirical reality, and some findings have brought critical challenges to various aspects of the taxonomy. All three kinds of findings are important as the empirical knowledge base accumulates with respect to the taxonomy and its predictions. Continued research – especially with respect to some of the understudied and unresolved questions – will permit a more concrete statement about how well the taxonomy explains criminology's most basic facts.

Related Entries

▶ Criminal Careers
▶ Desistance from Crime

Recommended Reading and References

Nagin DS, Farrington DP, Moffitt TE (1995) Life-course trajectories of different types of offenders. Criminology 33:111–139

Moffitt TE (1993) Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. Psychol Rev 100:674–701

Moffitt TE (1994) Natural histories of delinquency. In: Cross-national longitudinal research on human development and criminal behavior. Kluwer, Dordrecht, pp 3–61

Moffitt TE (2006) Life-course-persistent versus adolescence-limited antisocial behavior. In: Cicchetti D, Cohen DJ (eds) Developmental psychopathology, vol 3: risk, disorder, and adaptation, 2nd edn. Wiley, Hoboken, pp 570–598

Moffitt TE, Caspi A, Rutter M, Silva PA (2001) Sex differences in antisocial behaviour. Cambridge University Press, Cambridge

Odgers CL, Caspi A, Broadbent JM, Dickson N, Hancox RJ, Harrington HL, Poulton R, Sears MR, Thomson WM, Moffitt TE (2007) Prediction of differential adult health burden by conduct problem subtypes in males. Arch Gen Psychiatry 64: 476–484

Piquero AR (2008) Taking stock of developmental trajectories of criminal activity over the life course. In: Lieberman AM (ed) The long view of crime: a synthesis of longitudinal research. Springer, New York, pp 23–78

Piquero AR, Moffitt TE (2005) Explaining the facts of crime: how the developmental taxonomy replies to Farrington's invitation. In: Farrington DP (ed) Integrated developmental & life-course theories of offending: advances in criminological theory. Transaction, New Brunswick, pp 51–72

Piquero AR, Farrington DP, Nagin DS, Moffitt TE (2010) Trajectories of offending and their relation to life failure in late middle age: findings from the Cambridge study in delinquent development. J Res Crime Delinq 47:151–173

Piquero AR, Shepherd I, Shepherd J, Farrington DP (2011) Impact of offender trajectories on health: disability, hospitalization, and death by middle age in the Cambridge study in delinquent development. Criminal Behav Ment Health 21: 189–201

Sampson RJ, Laub JH (2003) Life-course desisters? Trajectories of crime among delinquent boys followed to age 70. Criminology 41:555–592

Monetary Strain and Individual Offending

Stephen W. Baron
Department of Sociology, Queen's University, Kingston, ON, Canada

Overview

Monetary strain theories, including anomie and general strain perspectives, outline that individuals who are unable to reach monetary goals utilizing legal channels, or who are experiencing negative economic circumstances, will be more likely to engage in a range of illegal activities. The probability of an illegal response is strengthened when individuals perceive their opportunities to achieve the monetary goal to be blocked, attribute the cause of their failure to external sources, and feel deprived relative to the others with whom they compare themselves. The failure to reach the monetary goal, or the negative economic experience, in these situations leads to frustration, dissatisfaction, and anger. These negative emotions supply the incentive for individuals to utilize numerous illegal avenues that focus on financial gain to facilitate the realization of monetary goals or address negative economic circumstances. Individuals experiencing monetary strain may also strike out at others in response to their feelings of failure and frustration or they may use drugs and alcohol in an attempt cope with the negative emotions generated by the strain. Illegal reactions in response to monetary strain are also more likely when individuals associate with peers who also experience monetary strain and utilize deviant means to cope with the strain. Further, individuals who have adopted values that support the use of illegal means to cope with strain and who lack conventional social support and social control are more likely to respond to monetary strain with crime. Monetary strain is also more likely to lead to crime when individuals lack self-efficacy and self-esteem, view the rewards of crime as greater than the potential costs of punishment, and have developed personality traits that leave them more likely to react to strain with negative emotionality and low constraint.

Fundamentals

The link between monetary strain and individual offending has a long tradition in criminology. Beginning with the work of Merton (1968) through to more recent arguments (see Agnew 2006; Baumer 2007), scholars have attempted to theoretically and empirically show that the failure to reach monetary goals, or the experience of adverse economic circumstances, can lead individuals to engage in crime to address these failures, improve their circumstances, and reduce the frustration that these experiences create. Scholars have provided a complex and nuanced map of the way that monetary strain is linked to individual offending and have presented a range of "research puzzles" that researchers have tackled to lend evidence to support our understanding of the causal process.

Scholars working within the strain perspective of criminology argue that everyone within North American society, regardless of class position, is encouraged and expected to strive to become monetarily successful. The goal of monetary success is emphasized and promoted in the media, family, and education system, making it difficult to avoid its influence. The result of this process is that a "substantial" or "significant" portion of individuals across all social classes adopt strong aspirations for financial success (Merton 1968). When an individual places a great deal of importance on the goal of financial success and prioritizes it over other goals, the failure to reach financial objectives creates dissatisfaction. When individuals experience monetary dissatisfaction, they begin to consider utilizing alternative means to satisfy their monetary goals including illegal means.

Beyond the failure to reach the goal of monetary success, negative economic circumstances including poverty, unemployment, and homelessness can also create monetary strain. These types of negative economic situations are viewed as undesirable in the broader society and for the individual immersed in these circumstances, the experience can be extremely negative. These types of situations serve to threaten an individual's goals and identity and influence their ability to carry out activities and satisfy basic needs. The adverse economic experience in the context of cultural failure creates an additional possible starting point for utilizing crime to cope with financial difficulties. Important here is not only that these are objectively negative financial circumstances, but also that an individual subjectively perceives them to be negative and is dissatisfied with these conditions. If an individual does not interpret their economic circumstances to be negative, then strain will not emerge despite the adverse economic circumstances being experienced.

While the cultural emphasis on monetary success is advanced for all in the population, large portions of the populace are inhibited from achieving financial success due to barriers in legitimate channels. Monetary goals are more likely to lead to crime when individuals experience hurdles to the achievement of their financial goals. The major channel emphasized to achieve financial success by strain theorists is the acquisition of education and the subsequent securing of high quality employment. These avenues, however, are more available to some individuals than others. These channels can be undermined by prior familial and community socialization that fail to promote qualities that facilitate educational and employment success, including the development of work ethics, deferred gratification, self-discipline, and thorough planning skills. Certain familial and community situations may also be unable or neglect to provide the knowledge, connections, and social and financial support necessary to facilitate an individual's success at school and in the job market. Finally, individuals from certain social and economic backgrounds may experience discrimination in educational and economic institutions that restrict their ability to reach their financial goals. Beyond inhibiting educational success and the reaching of financial goals, community and family dynamics can lead directly to negative economic outcomes including unemployment, poverty, and homelessness that are not only experienced as strain but can also serve as barriers in and of themselves to monetary goal achievement (Hagan and McCarthy 1997). Individuals with strong monetary goals are more likely to respond with crime when they encounter obstructions that undermine their ability to reach their goals through legal avenues. Further, those experiencing adverse economic circumstances are also more likely to engage in crime when they experience blockages to legal methods for improving financial circumstances.

The emergence of monetary strain does not take place in social isolation. Beyond the failure to achieve financial goals, monetary dissatisfaction, or the experience of negative economic circumstances, the development of monetary strain is also influenced by an individual's perceptions of the efforts, financial achievements, and life situations of other people. Scholars argue that individuals will be more likely to experience monetary strain when they conclude that they are financially worse off than the people with whom they compare themselves (see Agnew et al. 1996; Burton and Cullen 1992; Passas 1997). This negative evaluation of one's financial situation or goal achievement in comparison to others is referred to as relative deprivation. What is important here is the reference group that the individual adopts to make this comparison of achievements or life situations. Monetary strain is less likely to occur when reference groups selected share efforts and achievements similar to the individual's own efforts and achievements. The choice to adopt a reference group with achievements much greater when compared to the individual will lead to greater feelings of monetary strain. The selection of a reference group can be influenced by one's peer group who can guide one's choice to appropriate comparisons. Further, in some societies, there is a significant weight placed on individuals to choose nonmembership reference groups. For example, the egalitarian ideology of North America and the "American Dream," often lead an individual to evaluate themselves with reference to those higher in the stratification system (see Passas 1997). While this allows for individuals at all economic levels of achievement, except those at the very top, to experience relative deprivation, it is those lower in the stratification system that are at greater risk of feeling relatively deprived. This is because the selection of possible reference groups for those lower in the stratification system is broader and more likely to differ economically (see Passas 1997). It may also be that relative deprivation has a greater impact on the behavior of those in lower socioeconomic situations since the "subjective experienced gravity of these discrepancies" will be greater (Passas 1997, p. 65). The negative evaluation of one's financial situation in comparison to others generates feelings of resentment and hostility. Perceptions of relative deprivation, therefore, can lead directly to crime. It is argued, however, that it is those with strong monetary goals, or who are experiencing negative economic situations, that are more likely to engage in crime when they perceive themselves to be relatively deprived.

Monetary strain is also more likely to emerge and lead to crime when individuals place the blame for their failure on sources outside of themselves (Cloward and Ohlin 1960; Hoffman and Ireland 1995). Individuals who externalize blame and attribute their failure to the labor market, education system, economy, or governmental activities are more likely to become alienated and withdraw their support for the legitimacy of conventional social norms. In contrast, individuals who make internal attributions and internalize blame for their economic failure or circumstances are likely to continue to support conventional avenues for goal achievement and refrain from illegal activities. Further, the impact of these external attributions is greater when individuals believe that the causes of their goal failure or adverse economic circumstances are controllable, intended, or preventable by others. In addition, perceptions that the external causes of one's failure and financial difficulties are unlikely to change in the future can exacerbate the role that external attributions play in increasing the risk that monetary strain will lead to crime (Baron and Hartnagel 1997). Like the selection of reference groups, the directions and content of attributions can be influenced by one's peer group. Situations where blame is placed externally, perceived as intentionally caused or controllable and unlikely to change, generate feelings of injustice and perceptions of unfairness that facilitate the potential link between monetary strain and crime. Those who fail to reach their monetary goals and are monetarily dissatisfied, and/or who are experiencing adverse economic circumstances, are more likely to engage in crime when they make external attributions for these situations.

The experience of failing to reach monetary goals, monetary dissatisfaction, and adverse economic circumstances, along with blocked opportunities, relative deprivation, and external attributions for failure, generate a negative emotional reaction. Possible reactions can include depression, frustration, and anger. Anger, in particular, is a key factor that intervenes between negative economic circumstances, monetary goal failure, and crime. Anger promotes feelings of power and stimulates desires for revenge, retaliation, and the righting of perceived unfairness and injustices which can lead to criminal reactions (Agnew 2006). How angry one becomes in response to monetary strain may be influenced by people's beliefs about the level and degree of emotional reaction appropriate for these types of negative experiences. Once again, one's peer group may also influence interpretations of monetary strain and the type and level of emotional response that evolves.

Recent theoretical work also outlines that the chronic failure to reach monetary goals and long-term experiences with negative economic circumstances can leave an individual at risk to develop trait anger, a personality characteristic where one is persistently angry (Agnew 2006). In these situations, individuals will often be in a state of angry arousal leaving them vulnerable to become easily angered when faced with additional negative economic experiences and outcomes. Those who respond to negative economic experiences and outcomes with anger are more likely to use criminal means to reach their goals, reduce their negative economic circumstances, and cope with the negative affect. This may involve active means, such as violence, vandalism, and offenses involving instrumental gain, or through more passive means, such as illegal substance use, to soothe the negative emotions.

There are a number of other factors that can strengthen the link between monetary strain and crime. First, all strain theorists outline that monetary strain is more likely to generate a criminal response when the commitment to, and belief in, institutional norms is weak. The failure to reach monetary goals, the negative economic circumstances, the monetary dissatisfaction, and the perceived injustice generated by relative deprivation and external attributions can lead individuals to "withdraw legitimacy from conventional social norms" (Cloward and Ohlin 1960; Hoffman and Ireland 1995). It may also be the case that individuals from certain backgrounds may never acquire strong conventional values that emphasize the use of legal avenues for reaching monetary goals (Merton 1968). When commitment to the conventional norms is weak, or destroyed, individuals may create, embrace, and implement illegal methods to achieve success and overcome negative economic situations. Further, they will be more likely to retaliate against perceived sources of their monetary strain, or to act on, or cope with, the negative emotions generated by monetary strain more easily than if they felt these criminal actions were morally wrong. The weakening of conventional beliefs allows for the development and adoption of nonconventional beliefs that promote various forms of law breaking as a method to cope with monetary strain. Therefore, deviant values increase the likelihood that negative economic situations and monetary dissatisfaction will lead to crime. Further, monetary goals themselves will be more likely to lead to crime in the presence of deviant values, particularly, when legitimate opportunities are blocked, people are economically dissatisfied, and there is peer approval (Baumer 2007).

Strain theorists emphasize that criminal peers can have a strong influence on whether individuals experiencing monetary strain turn to crime. Individuals who fail to reach their monetary goals, who experience negative financial situations, and who are monetarily dissatisfied often seek or develop collective solutions with other people who are also experiencing monetary strain. These similarly situated peers can work together in attempts to overcome financial frustrations and negative economic circumstances. Peers can provide strained individuals with encouragement, support, and rewards for engaging in crime to deal with monetary strain. If peers are utilizing criminal means to cope with monetary strain, they may educate the individual in this avenue as a means to deal with strain and provide training, opportunities, and assistance to successfully engage in these behaviors. Criminal peers can assist in creating an alternative value system and may further break down allegiance to conventional norms and promote beliefs that encourage breaking the law increasing the potential causal link between monetary strain and crime. Alternatively, they may fill the gap left by the breakdown of conventional norms caused by strain providing a value system that favors illegal means to cope with strain. Peer groups can also affect perceptions of what is considered to be monetary strain, influence causal attributions of monetary strain, impact the choice of reference groups when making assessments of relative deprivation, and provide parameters for the appropriate emotional reaction to negative economic outcomes and goal failure (Cloward and Ohlin 1960; Hoffman and Ireland 1995).

Strain theorists also suggest that an individual's personal resources can influence reactions to monetary strain (Agnew 2006). Levels of self-esteem will influence an individual's sensitivity to monetary strain and sway the ability to engage in legal coping strategies in response to strain. High self-esteem will buffer an individual from monetary strain, leaving the person less likely to utilize criminal coping strategies. Those with low self-esteem who are overwhelmed by monetary strain will be less able or likely to adopt conventional coping strategies. Self-efficacy will also influence an individual's sensitivity and reaction to monetary strain. Individuals high in self-efficacy will be more likely to feel that they can cope with their monetary strain in a legal manner, leaving them less likely to utilize criminal coping. Lacking a sense that they can deal with monetary strain in a legal manner effectively will lead those low in self-efficacy to utilize criminal means.

Strain theorists also recognize that levels of social control can influence whether monetary strain leads to crime. Individuals who have few emotional attachments to conventional others and who have a low commitment to conventional institutions are more likely to respond to monetary strain with crime. People in these circumstances have little to risk or damage in terms of relationships, opportunities, and reputations if caught engaging in crime. In contrast, bonds with conventional others, a continued commitment to conventional activities, and a belief in the conventional value system may buffer the effect of monetary strain since the person risks damaging relationships, opportunities, and reputations by engaging in crime. Further, those with high social control have greater opportunities to learn skills and access resources to cope with monetary strain in a legal manner. Many negative economic situations, however, including unemployment and homelessness are associated with low social control. These negative economic circumstances often destroy or even evolve out of low social control. These conditions are often related to reduced emotional attachments to conventional others, a decrease in investments in conventional institutions, including employment, and a distancing of values supportive of conventional behavior (Agnew 2006; Baron 2008).

Related to this is the degree of social support that an individual has at their disposal to cope with monetary strain in a legal manner (Agnew 2006; Cullen and Wright 1997). Social support refers to the transfer of resources. Social support can come in two broad forms. First, it can be instrumental. This includes material or financial assistance, networks for positive advancement in conventional society, or the provision of information, guidance, and advice. Alternatively, social support can be expressive. This includes one person offering another understanding, allowing the display and sharing of emotions, as well as providing avenues that enable the expression of dignity, self-worth, and positive regard. When individuals experience monetary strain, those with social support can draw on this multifaceted resource to mitigate the potentially criminogenic impact associated with strain. The availability of supportive relations guards individuals against the impact of monetary strain by providing legitimate coping skills that inhibit the development of negative emotions and prevents the withdrawal of support for, and commitment to, conventional institutions and values. Social support can also reduce opportunities for crime by providing alternative settings and isolating at-risk individuals from criminally supportive peers. This is important because criminal social support from strained peers can lead to increased rather than decreased criminal reactions to strain. Therefore, the link between monetary strain and crime will be greater for those who lack conventional instrumental and expressive social support (Cullen and Wright 1997).

Individuals who exhibit personality traits such as negative emotionality and low constraint are also expected to be more likely to react to their monetary strain with crime (Agnew 2006). Individuals with negative emotionality and low constraint are easily upset and angered, tend to be impulsive risk-takers, and lack empathy for other people. These characteristics are likely to increase the relationship between monetary strain and crime because individuals with low constraint and negative emotionality are more likely to be sensitive to their goal failure and adverse economic situations, interpret these situations as strain, and view these situations as unjust. Individuals with these characteristics care little about the reactions of others and give little thought to the consequences of their actions, leaving them more likely to adopt criminal coping strategies. Further, individuals with these traits tend to have an aggressive interaction style and proneness for risk that increase the likelihood of an illegal reaction to monetary strain (Agnew 2006).

Finally, the perceived likelihood of formal punishment for crime also influences whether an individual will choose to cope with their monetary strain using criminal means. Criminal coping is much more likely when individuals experiencing monetary strain find themselves in situations where the costs of engaging in crime are low and the benefits are high. Thus, those who fail to reach their monetary goals or who are unemployed, homeless, or living in poverty will view criminal behavior as an attractive alternative when they believe that the odds of apprehension and punishment are minimal and the material, social, and emotional rewards great (Agnew 2006). The conditioning impact of potential punishment may be even more important when the strong monetary goals or negative economic circumstances are accompanied by minimal moral reservations against crime (Baumer 2007).

State of the Art

There have been a number of recent research projects that examine the complex relationship between monetary strain and crime. This recent research has tended to show support for the argument that monetary strain can lead to criminal behavior. Some of this work shows support for the basic relationships between monetary strain and crime, as outlined by the theorists. For example, research shows that those who express monetary dissatisfaction are more likely to engage in property offending, violent offenses, drug distribution, and drug use (Agnew et al. 1996; Baron 2004, 2006, 2008). There is also evidence that economic problems (Agnew et al. 2008) and adverse economic circumstances like homelessness and unemployment can lead to a range of offenses (Baron 2004; Hagan and McCarthy 1997). Work also tends to show that relative deprivation is linked to property and violent offending (Baron 2004, 2006, 2008; Burton and Dunaway 1994).

While encouraging, the more complex arguments outlined by theorists stress the need to examine combinations of factors that link monetary strain to crime. The cursory work conducted so far has provided initial support for a number of the components of the multifaceted reasoning contained in these claims. Turning first to monetary goals, evidence suggests that such goals are not in and of themselves criminogenic. Instead monetary goals are more likely to lead to property and violent offending when individuals do not *expect* to reach those goals (Baron 2006). Further, monetary goals also appear more likely to lead to crime for people with lower levels of monetary satisfaction, and lesser occupational and educational achievement (Cernkovich et al. 2000) or if they are experiencing adverse economic circumstances (Baron 2006). Thus, it is in situations where reaching monetary goals are not viewed as attainable, are difficult to reach because of economic position, or are not satisfactorily attained, that individuals with strong monetary goals are more likely to engage in crime.

At an even more complex level, some work has revealed that the availability of legitimate opportunities reduces the criminogenic impact of monetary goals, even in the presence of reduced support for conventional values. Similarly, being economically satisfied reduces the criminogenic relationship between monetary goals and instrumental crime even when one is supportive of crime as a method to reach these goals. Alternatively, holding values that support the use of illegal means to reach monetary goals when an individual places strong importance on those goals is extremely criminogenic, particularly when opportunities to reach the goals through legitimate means are blocked or one is monetarily dissatisfied (Baron 2011).

Research has also shown that monetary dissatisfaction is more likely to lead to crime when accompanied by negative economic circumstances, including homelessness and unemployment (Baron 2006). Further, scholarship reveals that the criminogenic effect of relative deprivation is greater when accompanied by negative economic circumstances like homelessness. This suggests that while negative interpretations of economic situations can lead to crime, the likelihood of these negative interpretations leading to illegal behavior is stronger when individuals are also experiencing adverse economic situations. The research also shows that although adverse economic circumstances, like unemployment and homelessness, can lead directly to crime, they are more likely to result in offending when individuals experiencing these conditions associate with deviant peers and hold values that support breaking the law. Similarly, negative interpretations of adverse economic circumstances, like monetary dissatisfaction and relative deprivation, are more likely to lead to various forms of crime when individuals have adopted values that favor law breaking and also associate with peers who break the law. Therefore, the criminogenic effect of negative economic circumstances and negative interpretations of these circumstances is influenced by the types of people one associates with (specifically, people who support using illegal means to attain monetary goals), as well as the types of values one adopts (that is, where there is a withdrawal of support for the legitimacy of conventional routes to success) (Agnew et al. 1996; Baron 2004, 2006; Baron and Hartnagel 2002).

It has also been uncovered that negative economic situations like unemployment are more likely to lead to crime when people make external attributions for their circumstances (Baron and Hartnagel 1997). In addition, when negative economic outcomes are viewed as unfair, they are more likely to lead to offending behavior when people place blame externally (Baron and Hartnagel 2002). Further, negative interpretations of economic circumstances, like monetary dissatisfaction, are also more likely to lead to crime when one attributes blame to external sources (Baron 2008). Thus, adverse economic situations and negative interpretations of economic circumstances are more likely to lead to a criminal response when blame for the negative circumstance is placed on outside sources like industry, government, and the economy.

Research has also provided support for the argument that monetary strain leads to the development of negative emotions. For example, research shows that negative interpretations of economic circumstances, like relative deprivation and monetary dissatisfaction, lead to the development of anger over economic circumstances (Baron 2004, 2008). Work has also provided insights to suggest that negative economic circumstances are more likely to lead to anger when individuals attribute the cause of their situation to external sources (Baron 2008). Further, peer influence, together with external attributions for the cause of negative economic circumstances, also contributes to the development of anger over economic circumstances (Baron 2008). Finally, there is support for the argument that the angry emotions that develop out of negative economic circumstances are predictive of some forms of criminal offending, including violence and drug dealing (Baron 2008).

In sum, recent work has provided a great deal of initial support for various aspects of the argument that monetary strain leads to individual offending. This research, however, also clearly demonstrates that this linkage is anything but straightforward and that evidence is still required to test some of the proposed linkages outlined by theoretical scholars.

Possible Controversies in the Literature

Despite what has been called a common sense explanation to crime, the link between monetary strain and individual offending has been attacked because of the perceived lack of empirical support for the perspective. Early work examining the perspective tended to offer minimal or marginal support for the approach (see Burton and Cullen 1992). Many prominent scholars went so far as to declare that searches for links between monetary strain and crime should be abandoned. However, as can be seen from the summary of recent work reviewed, there appears to be a significant connection between monetary strain and crime. The controversy appears to be a product of researchers oversimplifying and frequently misinterpreting the way monetary strain is linked to crime (see Agnew and Passas 1997). For example, empirical tests have often reduced the monetary strain/crime explanation to the simple assertion that when individuals are unable to gain monetary success through legal channels, or experience negative economic circumstances, they will engage in crime. Further, these tests have often ignored theoretical developments that expand the explanation of the role monetary strain plays in the generation of individual offending (Agnew 2006; Agnew et al. 1996; Baumer 2007; Cullen and Wright 1997; Hoffman and Ireland 1995; Passas 1997). The result is that much of the research does not do justice to the complex process outlined by theorists. The failure to take into account reference groups, attributions, perceptions of monetary dissatisfaction, negative emotions, deviant peers, deviant values, social support, social control, deterrence, and personality characteristics, such as self-esteem, self-efficacy, and negative emotionality/low constraint, has been a serious oversight. The tendency to ignore the way that these factors condition the development and impact of monetary strain means that many of the hypothesized linkages have never actually been explored. As can be seen, preliminary explorations into these more complex relationships have provided evidence that monetary strain leads to crime. Through the efforts of a number of scholars, the argument linking monetary strain to individual offending has survived (see Agnew and Passas 1997).

In addition to the poor theoretical application, much of the past empirical research examining the monetary strain/crime link has been beset by methodological issues. First, the operationalization of strain has proved problematic. For a considerable period of time, researchers relied on a financial strain measure focusing on the disjunction between educational/occupational aspirations and educational/occupational expectations. Criticism has emerged outlining that this type of operationalization is dubious since it does not tap into the notion of the importance of monetary goals so core to the explanation, nor does it capture the sense of dissatisfaction associated with the failure to reach those goals or the subjective unhappiness with one's monetary situation (Agnew et al. 1996; Baron 2006).

While advances have been made regarding this issue, there remains some debate regarding how to actually conceptualize and measure monetary strain (Agnew 2006; Agnew et al. 1996; Baumer 2007; Burton and Cullen 1992; Passas 1997). Although there is some agreement among researchers about the core concepts that need to be taken into account when exploring the theory, the way that researchers have chosen to interpret where in the causal process these factors should be placed and how they might possibly come together has led to an array of causal model types to be tested. These alternative interpretations will certainly provide research puzzles for future researchers to resolve. Over time researchers will hopefully clarify the causal linkage between monetary strain and crime.

A third issue that has plagued the research has been the tendency for scholars to utilize school populations to explore if monetary strain is linked to very minor forms of crime (Baron 2006; Burton and Cullen 1992). This trend has led observers to suggest that the monetary strain/crime research has been misdirected. Critics argue that the monetary strain/offending link should be strongest among those not in school, for whom money is of a greater concern, and among the urban poor who experience larger barriers to achieving their goals. Further, the explanation, they argue, is designed to explain more serious forms of offending. Again recent work tends to support the approach advocated by critics, thereby rescuing monetary strain from being discarded as a potential cause of individual offending.

Open Questions

It is becoming clear that the link between monetary strain and individual offending is quite complex. Researchers continue to debate the precise manner in which monetary strain is linked to crime, regularly providing alternative interpretations, extensions, and operationalizations. This process has resulted in a myriad of ways to both conceptualize and measure monetary strain. As the overview to the current state of affairs illustrates these extensions, alternative readings and methodological advances have all provided valuable insights into the manner in which monetary strain can be linked to crime. Yet, these extensions also create further questions that will need to be resolved. For example, more work is required outlining the precise direct and indirect paths by which various forms of monetary strain can lead to crime. Here, for instance, the role of emotions has been overlooked in most research exploring monetary strain. While there is some supportive research, more work will need to be carried out to determine how anger emerges from the array of variables that can be characterized as strain, and whether anger mediates their impact on crime, or whether their role is direct. Also, to date, there has been little research exploring the conditioning roles of social support, social control, deterrence, and personality characteristics such as self-esteem, self-efficacy, and negative emotionality/low constraint and certainly, more work on reference groups, attributions, perceptions of monetary dissatisfaction, deviant peers, and deviant values is required.

Another issue is that the array of factors that can moderate the effect of monetary strain on crime is quite extensive, leading to debates over how these factors should be combined and which ones are the most important both in terms of creating anger and increasing the likelihood of offending directly. Related to this is the fact that it is still unclear how many of the factors should or need to be combined to provide an accurate understanding of the link between monetary strain and crime. Work has begun exploring how combinations of two factors, for example, monetary goals and low expectations, work together to create crime. However, recent theoretical reinterpretations stress the need to examine how a larger number of factors, for example, three factors, work together to increase offending likelihood (Baumer 2007). The potential combinations between theoretically central variables provide a number of avenues for exploration and suggest that the link between monetary strain and crime may be more complex and nuanced than what has been discovered so far.

There is also still some uncertainty over the application of the perspective to various populations. For example, while there is a theoretically strong argument linking monetary strain and crime for disadvantaged racial and ethnic groups, little work has explored this possibility and the minimal empirical work is not supportive to date (see Cernkovich et al. 2000). Further, the role that gender plays in the criminogenic process of monetary strain has been relatively underexplored (Baron 2007). This is important since perceptions of monetary strain may vary across gender, emotional reactions might be different, and criminal responses might be dissimilar depending on gender (Agnew 2006). Sparse work suggests that perceptions of monetary strain and emotional reactions to monetary strain may not vary across gender. The way males and females react to negative economic circumstances, and interpretations of those circumstances with crime, however, may be more gender specific and differ depending on deviant peers, deviant values, and attributions (Baron 2007). It is also the case that while the monetary strain explanation is certainly pertinent to understanding the crimes of economically marginal populations, it can also be applied to understanding crime in other socioeconomic locations. Again, there is sparse work that suggests that the ideas here can be utilized to explain white-collar offending as well as street-oriented offenses (see Passas 1997).

There is also some disagreement about the link between strain and different types of offenses. Classic scholars, for example, suggest that goal failure leads to giving up monetary goals which in turn leads to using illegal substances. In contrast, contemporary work focuses on how having strong monetary goals can lead to substance use under the more complex process outlined. These seemingly contradictory arguments need to be further explored. It may be that the types of emotions that can be generated by strain, for example, depression, need to be examined, and linked to specific offenses. Further, internal attributions, or self-blame, that lead to more passive deviant responses to monetary strain may be a fruitful avenue to explore. It is clear that research examining a diverse range of populations and offenses, attempting to apply a more complex model using appropriate operationalizations, is certainly needed to provide broader support and explanatory refinement to the way monetary strain leads to crime.

Conclusions

In sum, monetary strain emerges when an individual is unable to reach monetary goals, experiences negative economic circumstances, or is dissatisfied with their monetary situation. This experience is exacerbated by blocked opportunities, feelings of relative deprivation, and external attributions of blame. This strain creates a negative emotional reaction that motivates individuals to engage in crime to cope with the monetary strain. The likelihood that monetary strain will lead to crime is increased when the individual experiencing the strain has few moral reservations about using illegal means, associates with criminal peers, has low social control and little conventional social support, has lowered personal coping resources like self-esteem and self-efficacy, exhibits negative emotionality/low constraint, and views the chances of legal apprehension and punishment to be minimal. The instrumental illegal actions taken can be seen as an attempt to satisfy monetary goals and overcome negative economic circumstances, while violence and substance use can be viewed as efforts to reduce the negative emotional affect created by monetary strain. Recent empirical work has provided early support for this explanation of crime. Yet, much work is still to be done to explore the full complexity of the theory and its application to an array of populations to explain a range of offending behaviors. These research challenges encourage current and future scholars to make strides to more precisely understand the important link between monetary strain and individual offending.

Related Entries

- ▶ Anomie and Crime
- ▶ General Strain Theory
- ▶ Group Characteristics and General Strain Theory

Recommended Reading and References

Adler F, Laufer WS (1995) Advances in criminological theory, vol 6, The legacy of anomie theory. Transaction, New Brunswick

Agnew R (2006) Pressured into crime: an overview of general strain theory. Roxbury, Los Angeles

Agnew R, Passas N (1997) Introduction. In: Passas N, Agnew R (eds) The future of anomie theory. Northeastern University Press, Boston

Agnew R, Cullen FT, Burton VS Jr, Evans TD, Dunaway RG (1996) A new test of classic strain theory. Justice Q 13:681–704
Agnew R, Keith Mathews S, Buchner J, Welcher AN, Keyes C (2008) Socioeconomic status, economic problems, and delinquency. Youth Soc 40:159–181
Baron SW (2004) General strain, street youth and crime: a test of Agnew's revised theory. Criminology 42:457–483
Baron SW (2006) Street youth, strain theory, and crime. J Crim Justice 34:209–223
Baron SW (2007) Street youth, gender, financial strain and crime: exploring Broidy and Agnew's extension to general strain theory. Deviant Behav 28:273–302
Baron SW (2008) Street youth unemployment and crime: is it that simple? Using general strain theory to untangle the relationship. Can J Criminol Crim Justice 50:399–434
Baron SW (2011) Street youths and the proximate and contingent causes of instrumental crime: untangling anomie theory. Justice Q 28:413–436
Baron SW, Hartnagel TF (1997) Attributions, affect and crime: street youths' reactions to unemployment. Criminology 35:409–434
Baron SW, Hartnagel TF (2002) Street youth and labor market strain. J Crim Justice 30:519–533
Baumer EP (2007) Untangling research puzzles in Merton's multilevel anomie theory. Theor Criminol 11:63–93
Burton VS, Cullen FT (1992) The empirical status of strain theory. J Crime Justice 15:1–30
Burton VS, Dunaway RG (1994) Strain, relative deprivation, and middle-class delinquency. In: Barak G (ed) Varieties of criminology: readings from a dynamic discipline. Praeger, Westport
Cernkovich SA, Giordano PC, Rudolph J (2000) Race, crime and the American dream. J Res Crime Delinq 37:131–170
Cloward RA, Ohlin L (1960) Delinquency and opportunity: a theory of delinquent gangs. Free Press, New York
Cullen FT, Wright JP (1997) Liberating the anomie-strain paradigm: implications from social support theory. In: Passas N, Agnew R (eds) The future of anomie theory. Northeastern University Press, Boston
Hagan J, McCarthy B (1997) Anomie, social capital and street criminology. In: Passas N, Agnew R (eds) The future of anomie theory. Northeastern University Press, Boston
Hoffman JP, Ireland T (1995) Cloward and Ohlin's strain theory reexamined: an elaborated theoretical model. In: Adler F, Laufer WS (eds) The legacy of anomie theory, vol 6, Advances in criminological theory. Transaction, New Brunswick
Merton RK (1968) Social theory and social structure, enlarged edn. Free Press, New York
Passas N (1997) Anomie, reference groups, and relative deprivation. In: Passas N, Agnew R (eds) The future of anomie theory. Northeastern University Press, Boston
Passas N, Agnew R (1997) The future of anomie theory. Northeastern University Press, Boston

Money Laundering

Brigitte Unger
Utrecht University School of Economics,
Utrecht, The Netherlands

Synonyms

AML; ML

Overview

Laundering money entails disguising the illicit origins of proceeds of financial manipulation, drug trade, fraud, corrupt enrichment, or other crimes, bringing back money into the financial circuit of the legal economy. Some underlying misbehavior, the so-called predicate offense generates illicit gains which are to be turned legal. Hiding or disguising the source of these proceeds will not amount to money laundering *unless* they were obtained from a criminal activity (Busuioc 2007).

Predicate offenses for laundering can be "hard" crime involving drugs or violence, but may also be white collar crimes, such as illicit real estate dealings, fraud and embezzlement, and finally tax fraud and tax evasion. Laundering the resulting (typically large amounts of) money can be achieved through fake invoicing for exports and imports or unusually priced real estate transactions. Disguising is further attempted through pumping money around foreign banks (typically in tax haven countries), through gambling or buying expensive consumer goods which can be moved around without effective controls (e.g., diamonds, highly priced art, or yachts). Each of these innovative ways of spending money (which can be recovered in legal markets) may require a different set of expertise of the investigation agencies: It might be the criminal police, customs, or forensic auditing firms. In many cases, only a combination of these might be able to trace these complex crimes.

With the growth of financial markets and instruments, money-laundering constructions have become ever more sophisticated. Swaps and other derivatives can be involved; multiple housing transactions and complicated insurance instruments can blur the origins of the illegal proceeds. Take, for example, a real estate agent collecting say 10 million dollars from a tax evader in cash money for selling a business complex. The listed price of the object is 5 million dollars; the true price is 15 million dollars. The real estate agent takes the cash money to his or her bank and transfers this money to his abroad company X, who in turn gives a fake loan of 10 million dollars to a company Y. Company Y pays back the fake loan to company X. The real estate agent now has an official receipt of income of his company X of 10 million dollars. Tax evading now results in an object bought worth 15 million dollars of which only 5 million in official money has been paid, the rest with laundered money. When the object is sold again, there are official sales receipts for 15 million dollars. Finally, the tax evader has a non-traceable supposedly legal revenue of 10 million dollars which can be spent on investments in business, hotels, or for luxury consumption of cars, jewelry, and real estate.

Money laundering, however, owes its name to a much simpler technique dating back to Chicago gangster Al Capone. He used the cash-intense business of launderettes to hide his illegal alcohol proceeds in times of prohibition in the 1930s. Slipping money from alcohol proceeds into the cash register of his launderettes, he pretended this was turnover of the regular business. By 2009, money laundering is estimated to amount between 1.5 and 3 trillion US dollars that circulate across the globe as illegal money transfers (see Walker 1995; Unger 2007; Walker and Unger 2009; Review of Law and Economics 2009). With the exception of the group around Truman and Reuter (2004) who think that the estimates of laundering are exaggerated, money laundering is considered a huge global phenomenon.

Laundering transfers can take place using the banking sector. They can use transfers by regular invoice or payment including new electronic methods like digital cash, e-gold, or prepaid phone cards (see http://www.fatf-gafi.org). Pressure from national regulators has led to the introduction of more electronic control systems. Financial institutions set barriers for the transfer of larger amounts (such as 10,000 dollars in cash or unusual transfer) which are to be checked for their source or purpose. Gambling casinos are obliged to control the identity of customers who invest unusual amounts of cash. Internet gambling, however, can easily avoid national borders and controls.

Such controls are devised to detect the predicate crime. The United States has developed a list of more than 130 predicate offenses suspect for money laundering. More predicate offenses were added into the money laundering definition over the years. Upon political requests, also terrorist financing and lately tax evasion have been included (see: FATF 2010). As a consequence, the amount of laundering has increased by definition.

Setting Money Laundering on the International Agenda

Money laundering is a relatively young topic and only recently an issue on the international agenda (see Unger 2007 for a survey of the money-laundering literature). Sure, already in ancient Chinese trade, there were other forms of hiding proceeds from authorities and there were punishments for dealing in stolen goods (“fencing”). However, money laundering in its current meaning dates back to the 1980s when it was first criminalized in the United States.

Fighting drug dealing has for a long time been of central concern in combating money laundering. After several decennia of the unsuccessful US “War on Drugs,” the Clinton Administration chose a hard crime fighting approach: If drug dealers and other criminals could not got caught directly, at least they should be discouraged by not being able to reap the monetary benefits of their dealings (see Unger 2011). Thus, in 1986 (Title 18, US Code Sec. 1956), money laundering got punished in the USA with up to 20 years in prison and $500,000 in fines. Further legal arrangements permitted seizing, freezing, and confiscation of assets by the authorities.

Since money laundering is a global crime without borders, the USA made large efforts to convince the international community of the importance of the topic. On a global level, anti-money-laundering policy started in the late 1980s with the UN Convention on Drug and Narcotics of 1988. In 1989, the Financial Action Task Force (FATF), an intergovernmental body to combat money laundering, was established by the G-7 countries. Since then the global movement against money laundering steadily gained support. Today, the FATF comprises of 34 member jurisdictions and two regional organizations, representing most major financial centers in all parts of the globe. The combat accelerated after the terrorist attacks of 9/11, when getting hold of terrorists by combating the financing of terrorism became part of anti-money-laundering policy.

However, one principal objection to FATF's work is its nondemocratic character. Groups of experts are nominated by their governments; they are not accountable to any parliament, and nobody can dismiss them. In such an "expertocracy," some countries send delegates to FATF meetings who may not be fully aware which enormous consequences these meetings can have on their national legal system and institutional setting.

Hindrances for Compliance: Victimless Crime and Unproven Effects

Money laundering itself does not seem to have direct victims. No individual suffers from the Colombian drug mafia sending 10 billion dollars from one country to another; on the contrary, many might rightfully profit from the resulting business. There are no victims who would complain, no liability claims that would be raised, and no tort law that could be applied. Legal literature therefore debates whether crime without victims should be punished at all and whether punishment is a deterrent. Would society be damaged by this maleficent behavior and fall into decadence (Mill, Chapter 1, p. 21)? The damage incurs on the level of the common goods of society. "*Public order crime*" would be the preferred term in place of "victimless" crime.

The harm of money laundering is indirect. Unger (2007) lists 25 negative and positive effects that money laundering can have. It can deter honest business activities: If criminals invest their money in particular sectors like the transport industry, restaurants, housing, it can lead to changes in relative prices, savings, output, employment, and growth or it can affect the liquidity, reputation, integrity, and stability of the financial sector. The public sector can suffer from unpaid taxes and criminals might buy up public enterprises during privatization efforts. There are also social and political effects, such as increased corruption and bribery. Criminals might undermine political institutions: As laundering needs facilitators, professionals such as lawyers, notaries, real estate agents, and accountants can become contaminated. None of these effects, however, can be observed directly which renders it difficult to prove. Reaching compliance for anti-money-laundering policy needs the backing of civil society.

Development of International Anti-laundering (AML) Standards That Need Compliance

For larger amounts which are dealt within international finance, the FATF sets international standards, which, in order to comply to, member countries have to arrange for enforcement and execution. Lawyers in ministries, policemen, public prosecutors, and judges may be involved in antiterrorism and money-laundering combat. Special organizations, called Financial Intelligence Units (FIU), have to be established in each country. Countries have to introduce plans to implement ongoing customer due diligence (CDD), identify nondomestic politically exposed persons (PEPs), and ascertain identification of beneficial ownership of offshore accounts. Banks, real estate agents, notaries public, traders in large values are obliged to screen their clients and to identify persons, activities, or transactions suspicious of money laundering or terrorism financing among their clients. The FIU collects their suspicious transaction reports. Supervisory authorities are involved to control the compliance of banks and other sectors with the anti-money-laundering regime (AML). Some countries, such as the USA, have high sanctions for not reporting, with fines as great as $250,000 or 5-year imprisonment (IRS 2011).

However, there has also been criticism of too harsh penalties. Takats (see Masciandaro et al 2007) warned stricter sanctioning might lead to a "crying wolf problem," referring to the boy who always cried wolf but once the wolf really came nobody paid attention. The private sector might, from fear of being punished, drown the FIU by reporting all sorts of transactions as suspicious without indicating what might a predicate offense.

Multiple Actors Involved in Money Laundering

The money-laundering process may be divided into three phases. In the placement phase, the criminal, after he has committed the predicate crime, such as selling drugs, committing fraud, working illegally, or producing fake brand products, wants to get rid of the criminal cash. He or she tries to hand the proceeds of crime over to a person from the "upper world." This can be putting it into the cash register of a launderette as Al Capone did and bringing it to the bank as legal proceeds from washing people's laundry, bringing illegal cash to a bank employee who accepts a suitcase of cash money, to a company who issues fake invoices, or to a transport entrepreneur who transports cash as a courier to another country where it is placed at a bank. Other actors involved in this first phase may be bank employees, business owners, or employees from the transport sector, the restaurant business, etc. After, the second phase called "layering" starts, which tries to make it difficult to detect the illicit proceeds. Transfers may be wired; offshore banks may be involved; fake shell companies may pump the money around the globe. Here the international financial sector would be involved. In the last phase, the integration phase, the money is parked in either financial investment and business. Actors involved may be real estate investors or notaries who are involved in buying and selling houses and accountants in companies, tax advisors, etc. Smaller amounts which could possibly be carried in cash might be laundered by luxury consumption or gambling. Actors involved might be dealers in diamonds or cars, custom officials, etc.

All private sector actors listed, such as bank employees, accountants, tax advisors, dealers, notaries, and lawyers, can get consciously or unconsciously involved in money-laundering transactions. Complicity is an offense wherever laundering is prosecuted. It is the task of all professional services involved to stay alert to money-laundering transactions and to report suspicious transactions to public authorities. Their commitment can be either voluntary, as to help fighting crime, but increasingly statutes try to enforce it by reporting duties and threat of punishment.

Compliance of a Great Variety of Actors

Combating money laundering is as strong as its weakest link in the chain. Coordination and compliance of many policy actors is crucial. Diverse actors for whom the policy is not necessarily given priority or not even in their own interest must be brought into one line. Unlike food or tobacco regulation, combating money laundering does not have a private sector lobby. It is governments and with them the media engaging in a fight against laundering. Not *all* governments are however as active in engaging in the combat. For some countries, it can be beneficial to attract criminal money. Unger and Rawlings (2008) speak of the "Seychelles effect," referring to the island state which in the 1990s deliberately advertised that if investors invested more than ten million dollars, they would not check where the money came from. As they are hardly confronted with the underlying crime, offshore islands and small countries may see the incentive of attracting criminal money. Large countries on the other hand often attract a bigger volume of laundering due to their financial centers and expertise (see Gnutzmann, McCarthy and Unger 2009).

By membership of the FATF or belonging to a FATF-Regional Style Body, states are obliged to implement international standards of anti-money-laundering policy, even though there are divergent national interests – some are tax havens and profit, others suffer from the experience of crime. Lack of a common interest among countries and the divergence of who bears the costs and who the benefits of the combat against money laundering makes it difficult to effectively implement anti-money-laundering policy. Seeing the variety of actors, different ways of circumvention and peculiar forms of compliance can be expected.

Public Actors Involved in Combating Anti-money-laundering

Governments must implement the international standards of FATF into national law. For some countries, this can pose a legal problem. Export-intensive countries such as Austria, Germany, and the Nordic countries feared double punishment for laundering money as well as the underlying crime and therefore originally excluded self-laundering from the money-laundering definition. In the FATF Third Mutual Evaluations of Austria and Germany, authorities explained that the predicate crime includes the concurring act, the laundering, and that the sanction set for the predicate crime is deemed to cover the entire unlawfulness of the criminal's act (FATF 2010). Most of these countries had to revise their money-laundering definition due to pressure of the FATF. In strongly legalistic countries, this change in the principles of law might lead to resistance when practicing the law, so that law in the books and law in practice diverge. Compliance would here be only in the books but not in practice.

Combating money laundering goes from designing the national anti-money-laundering law(s) to the establishment of an FIU. It can be located at the Ministry of Finance as financial intelligence unit, at the Ministry of Justice, the police, or at the Central Bank. It receives reports of suspicious transactions from the private actors, especially from banks and the financial sector, and has to screen these reports. A finance person may concentrate on different reports than a police officer used to hunting drug dealers. The FIU located in the Ministry of Finance might be more inclined to identify tax fraud, while the police might be more after corruption and drug cases. Thus, data access and the interpretation of data vary widely among the different FIUs.

The further chain of judicial procedures can be frustrating if police investigations are judged to deliver insufficient proof. Reports found sufficient by the investigators are sent to the office of public prosecutors who may decide whether to indict and proceed with an accusation or to drop the case, either because of lack proof or because of lack of legal relevance. In principle, the prosecution has to prove suspects guilty, but in some countries, the judge may allow for reversal of the burden of proof about where a big amount of money (relative to the suspect's means) came from.

An effective indictment presupposes all public actors, the legislative, the executive, and the judiciary are willing and able to recognize anti-money-laundering policy as an important goal to pursue and that they are willing and able to cooperate. They should have a similar understanding of the goals of anti-money-laundering policy and how it can be practiced in each country. Compliance policy will be counteracted from within, when public actors are themselves involved in money laundering: such as may be the case when government officials or customs and police officers are part of the laundering gang.

Reporting and Supervision

Anti-money-laundering policy requires penal codes which clearly define offenses and punishment as well as an active enforcement administration. If private actors (such as banks) are obliged to report suspicious transactions to the FIU, noncompliance may risk high fines (as in the USA) or even a jail sentence (as in Luxembourg).

Reporting obligations bring into life reporting systems. Banks and financial institutions may have to establish a software system to recognize suspicious transactions including compliance officers who can be held responsible. They have to check on lists of internationally blacklisted persons and companies, to fill in forms and forward them to the FIU, sometimes for privacy reasons handing them over personally. In short, the reporting system is costly and cumbersome for the private sector. As the private sector has no particular interest in fighting money laundering (there is no private sector lobbying for this policy issue) and in so far as compliance poses extra costs, it might be difficult to be reached.

Supervision of reporting entities is important. Many supervisory authorities are engaged in supervising reporting. In many countries, the Banking Supervision Authority also supervises anti-money-laundering matters. But do they find money laundering as important as, for example, the Financial Intelligence Unit might find it? Lawyers are usually supervised by their Bar Association which might put more emphasis on legal privilege and privacy of customer relations. In some countries, the FIU itself supervises. It remains doubtful whether the FIU knows enough about diverse sectors in order to supervise banks, financial service sectors, accountants, car dealers, and diamond dealers. Countries differ with respect to the competent actor for supervision. Some agencies might be centrally organized, others at a state or provincial level.

Hindrances to Compliance

Anti-money-laundering policy is mainly a public sector concern; hence, the private sector needs to be persuaded to collaborate. This means both good relations between the public and private sector as well as strict laws, punishment, and supervision should be in place. However, some countries benefit from the underlying crime while others bear the costs it. Disinterested countries may need to risk international sanctions or receive side payments in order to combat money laundering. As long as some countries hunt drug dealers but stick to bank secrecy while others fight tax evaders, compliance lacks a common policy goal to which one should comply.

Private actors have different priorities and often a double relationship: A launderer is also a client of banks and lawyers. Reporting a client with a close relationship to a bank (particularly when cash business is involved) means losing money, while not reporting may entail reputation damage to the bank. Similarly, there is the legal privilege of lawyers and notaries, which guarantees privacy of the lawyer-client relationship. Reporting a client therefore becomes a delicate balance between reporting duty and right of legal privilege, which might be very difficult to supervise and actors involved are definitely not interested in cooperation.

Seeing the many reasons why the private sector might not comply with anti-money-laundering policy, and seeing the many ways in which anti-money-laundering policy might be circumvented, the FATF has opted for blacklisting as a drastic step of enforcement.

Blacklisting Countries for Noncompliance

Many institutions, among them all central banks of the OECD countries (18 members) as well as the European Union, are maintaining white lists of countries: countries which are cleared for controlling their banks and financial institutions as well as multinational corporations for money-laundering transfers. Transfers from countries which are cleared may be accepted on the basis of simplified identification research of their financial institutions.

Similar lists, only with a reversed definition as *black lists* of countries with insufficient controls, are issued by the FATF (23 members). It has issued forty anti-money-laundering recommendations for controlling the financial sector and nine recommendations for combating terrorist financing. Countries face regular mutual evaluations and in case of noncompliance with the recommendations, they can get blacklisted as noncooperative countries (Unger and Ferwerda 2009). This can be economically very harmful, since countries risk that no American bank will do business with them. It is doubtful, however, whether blackmailing dogs really want to bite. In a first evaluation of the International Monetary Fund (IMF) in 1999 when a definition of financial havens was used, 69 noncooperative countries were listed, among them 19 European countries such as the financial centers of the UK, the Netherlands, Switzerland, and Luxembourg. This cross-border transfer of capital leads to extraction of capital in countries where the crime is committed and leads to an inflow of capital in countries where the integration phase takes place. The IMF (with 188 members including developed as well as developing countries) stopped their rankings because they felt that they could not take into account both ends of the chain. Consequently, they criticize the FATF (with 18 predominantly richer members) which simply checks the financial control infrastructure by intergovernmental assessments (Table 1).

Money Laundering, Table 1 Blacklists for noncooperative countries and territories

Blacklist	Number of countries worldwide	IMF: financial havens FATF: anti-money laundering and countering the financing of terrorism
IMF list 1999	**69**	**Nineteen** European countries: Austria, Hungary, Ireland, Luxemburg, Netherlands, Russia, Switzerland, Ukraine, UK, Liechtenstein, Luxembourg, and 8 smaller European jurisdictions
FATF 2000	**15 high risk**	**Two** European countries: Liechtenstein, Russia
FATF 2001	**19 high risk**	**Three** European countries: Hungary, Russia, and Ukraine
FATF 2002	**10 high risk**	**One** European country : Russia
FATF 2005	**3 high risk**	**no** European country
		Only Nigeria, Myanmar, and Nauru left on the list
FATF 2012	**2 high risk**	**no** European country
		Iran, North Korea

IMF, FATF: only high risks
In addition to the high-risk label, the FATF since 2011 indicates a gray list of deficiencies every year. In 2012, it included 15 countries: Cuba, Bolivia, Ethiopia, Ghana, Indonesia, Kenya, Myanmar, Nigeria, Pakistan, São Tomé and Príncipe, Sri Lanka, Syria, Tanzania, Thailand, and Turkey

And indeed, on the FATF list of noncooperative countries, none of the financial centers are listed. The first official FATF rating in 2000 lists 23 countries, later blacklists included only 15 noncooperative countries of which two were European: Liechtenstein and Russia. Eventually the list got shorter and shorter. In 2006, Myanmar was the only noncooperative country left, and was finally also removed from the list in October 2006. The regular mutual evaluations of the FATF look for "risks of anti-money laundering and countering the financing of terrorism." As soon as the Financial Control Agency cooperates (such as the financial center countries), the country does not appear on the blacklists.

The FATF blacklist was heavily criticized for its lack of transparency and legitimacy, as it did not rate all countries and was biased against small countries (Unger and Ferwerda 2009). Since 2011, there is a new "gray list" which distinguishes in four categories in how far they are fulfilling the FATF recommendations (either fully, making efforts in the right direction or not having made any effort for change). Countries make great efforts to be removed from this list, but this might show only "shallow compliance" in the books instead of effective action. They might strategically produce statistics to satisfy the FATF (such as by just reclassifying cases or counting a case with several transactions by each transaction separately).

Related Entries

- ▶ Deterrence of Tax Evasion
- ▶ Naming and Shaming of Corporate Offenders
- ▶ Risk Assessment, Classification, and Prediction
- ▶ Role and Function of the Police

Recommended Reading and References

Baker R (2005) Capitalism's Archilles heel: dirty money and how to renew the free-market system. Hoboken, NJ: John Wiley

Busuioc EM (2007) Defining money laundering. In: Unger B (ed) The scale and impacts of money laundering. Edward Elgar, Cheltenham/Northampton, pp 15–280

Deleanu I, van den Broek M, Ferwerda J, Unger B (2011) Performing in the books, assessing countries' anti-money-laundering policy in the light of strategic reporting. Paper presented at the ECPR Conference 2011 in Reykjavik, Iceland. http://www.ecprnet.eu/MyECPR/proposals/reykjavik/uploads/papers/2734.pdf

FATF (2010) Financial action task force report on tax crime, Third round mutual evaluation report on Germany, Paris

Financial Action Task Force homepage. http://www.fatf-gafi.org

Gnutzmann H, McCarthy KJ, Unger B (2010) Dancing with the devil: country size and the incentive to tolerate money laundering. Int Rev Law Econ 30(3): 244–252

IRS (2011) Suspicious activity reports. Retrieved from http://www.irs.gov

Masciandaro D, Takats E, Unger B (2007) Black finance: the economics of money laundering. Edward Elgar, Cheltenham, UK and Northampton, MA

Sabitha A (ed) (2009) Combating money laundering. Transnational perspectives. Amicus Books: The Icfai University Press

Tackling money laundering, Special issue of Rev Law Econ 5(2):807–985. December 2009

Truman E, Reuter P (2004) Chasing dirty money: the fight against money laundering. Institute for International Economics, Washington, DC

Unger B (2007) The scale and impact of money laundering. Edward Elgar, Cheltenham, UK and Northampton, MA

Unger B (2011) Regulating money laundering: from Al Capone to Al Qaeda. In: Levy D (ed) Handbook of regulation. Edward Elgar, Cheltenham, UK and Northampton, MA

Unger B, den Hertog J (2012) Water always finds its ways. Crime Law Social Change 57:287–304

Unger B, Ferwerda J (2011) Money laundering in the real estate sector. Edward Elgar, Cheltenham. 2011 Review of Law and Economics, Special Issue on Money Laundering, Dec 2009. http://www.toezichtencompliance.nl/nieuws?cID=914&PHPSESSID=jdeaul6hts3u2cv6kdeubdhce6

Unger B, Rawlings G (2008) Competing for criminal money. Global Bus Econ Rev 10(3):331–352

United Nations Office for Drugs and Crime (UNODC). www.unodc.org

Walker J (1995) Estimates of the extent of money laundering in and through Australia. Paper prepared for the Australian Transaction Reports and Analysis Centre, Sep 1995. John Walker Consulting Services, Queanbeyan

Walker J, Unger B (2009) Estimating money laundering: the walker gravity model. Rev Law Econ December

Monitoring and Evaluation of Restorative Justice

Joanna Shapland
School of Law, University of Sheffield,
Sheffield, UK

Overview

As restorative justice programs spread throughout the world, it has become more and more important to have adequate evaluations of such programs upon which policymakers, practitioners, and academics can rely. This entry distinguishes between *monitoring*, which all programs and practitioners should do themselves, and *evaluation*, which requires more specialized knowledge and skills. Both monitoring and evaluation need to be appropriate to the aims and context of the program itself and should meet the needs of and be accountable to the stakeholders of the program, including funders, those referring cases to the program, participants, and staff. Evaluation is thus not just "measuring satisfaction" or answering the question "does it make offenders reoffend less?" but must consider all the outcomes desired for the program, whether they have been achieved and, preferably, whether the program is cost-effective, or value for money. The entry provides some guidance on how to set up and conduct evaluations of restorative justice programs, but each program needs to consider its own requirements and aims.

What Is Evaluation?

Evaluation itself is a contested topic and the subject of many academic papers in its own right. Within criminology, evaluation has been the subject of lively exchanges, for example, on whether outcome evaluation is the key or whether the contexts of programs and the theoretical mechanisms through which the outcomes might be expected to occur are equally important (Farrington 1997, 1998; Pawson and Tilley 1998).

Historically, the forms of restorative justice being developed have been highly attuned to their cultural environment. As a result, there are many different types of program or scheme, even when considering only those dealing with crimes and which remain within the sphere of criminal justice (this entry is restricted to this sector). There is victim-offender mediation, conferencing, indirect mediation, sentencing circles, community panels, and many others. There is even more diversity if one moves away from crime and considers the use of restorative justice in civil disputes, in schools and organizational settings, or in peace-making. It is also important to reflect as to whether the restorative justice program is concerned only with offences which have direct victims (people who are themselves injured or lose property) or whether it is also concerned with indirect victims (family members who are affected, neighbors, the community). The entry will use the term "scheme" to denote a program of work delivered by an organization within a defined geographical area in relation to criminal cases, such that the work is done within an agreed framework or process by one or more mediators or facilitators. It is possible for a scheme to have a national remit, as does the Youth Conference Service of Northern Ireland, which has statutory authority to deal with all cases referred by prosecutors or judges in youth justice cases (Campbell et al. 2006). More typically, however, restorative justice schemes have grown up locally, to serve one town, or one region, often specializing in particular types of referrals.

Within restorative justice schemes, there is considerable local diversity – and this may need to be so if participants (victims, offenders, those supporting each party) are to feel comfortable and safe in the restorative setting and so able to communicate with each other. However, it makes it much more difficult to say which forms of restorative justice are most suitable for whom or for what kind of dispute. We need to develop methods of evaluation which are sympathetic to the local aims and procedures of different restorative justice procedures but which are still rigorous enough to allow us to show whether there are differences in outcomes and to compare schemes.

Evaluation itself is a word which means different things to different people – which is usually the prerequisite for mutual misunderstanding, particularly in a comparative field. First, there is the need to distinguish what will be called "monitoring" from "evaluation."

Monitoring: A Basic Duty for Each Scheme and Practitioner

Monitoring is an activity which needs to be undertaken by all schemes and by all practitioners. It is the essence of the reflective practitioner: essentially considering at regular intervals what one has done and reflecting on what has worked and what might be changed.

The basic level of monitoring implies record keeping of what cases have been dealt with, from whom referrals have come, who the participants were (including basic demographic details such as age and gender), what happened in the case and the dates of contacts, and when the case was closed. Having such systems is the basis of accountability and a requirement in codes of practice for restorative justice practitioners (e.g., Restorative Justice Council 2011, in the UK). They are important in answering queries from participants (such as "I've lost my copy of the outcome agreement: could I have another one?"), from those referring cases and from funders. Without good data systems and monitoring, further evaluation is impossible. Nor is it possible for governments and funders to see whether funds are being used correctly.

Yet new schemes do find it very difficult to remember to set up such systems, which provide not just a case management tool but also allow reports to be derived easily to provide management information (Shapland et al. 2011). There is no "off-the-peg" software currently available for running restorative justice schemes, and it is strongly suggested that new schemes contact others (even others in different countries) to find out their experience in practice. Restorative justice is much more complicated than criminal justice processes, because it is voluntary for both victims and offenders – so there is a need to ensure all the participants, venue, etc., are in place before there is a direct meeting.

Monitoring implies that data will be available from each restorative justice scheme in terms of its activity and some agreed outcome measures. So it will be possible to know how many cases have been referred to the scheme from whom, how many facilitators/mediators dealt with them, when they were referred and when they were closed, and what the process outcome was (one party declined the service, indirect mediation, a meeting between parties, etc.). Some of the latent antipathy towards evaluation by some restorative justice workers seems to have stemmed from the feeling that restorative justice must never be bureaucratized as a product, to be achieved identically each time, and to be able to be counted like tins of beans. This is a valid concern: restorative justice must be flexible and able to meet each group of participants' needs. The very definition of restorative justice indicates this continuous democratic creation: "a process whereby parties with a stake in a specific offence collectively resolve how to deal with the aftermath of the offence and its implications for the future" (Marshall 1999). If the parties are collectively resolving, then necessarily the content of their discussions and any outcome agreement will be slightly different each time. With monitoring, however, it becomes possible to consider how restorative justice is faring across schemes or across countries (see Shapland et al. 2011, for a review of evaluation results; also Sherman and Strang 2007).

Understanding the Monitoring Figures: Providing a Description of the Scheme

Given the intrinsic plurality of restorative justice, bare statistics tend to be meaningless. What does it mean to say that one scheme had 30 referrals from prosecutors while another had 150? What does it mean for the participants that 80 % of cases were closed within four months? Monitoring is vital, but collection of figures needs to be amplified by a detailed description of the process of restorative justice being run by that scheme, including:

- What kinds of cases does the scheme take? What kinds of cases are referred? What processes are undertaken to see whether cases are suitable?
- Who is contacted by whom when a case starts? What is offered? Typically, how does the case proceed?
- Do participants meet face-to-face? Just victim and offender with a mediator (or two mediators?) Or with others? How many meetings?
- What are typical outcomes? Is there a written outcome agreement (agreed by all parties)?
- What happens after any meeting? Is there feedback to those referring cases? Is there feedback to participants about any criminal justice outcomes?

One of the most detailed but most helpful descriptions of what was done by a scheme was Justice Research Consortium Thames Valley (2004) "From soup to nuts" folder. The "from soup to nuts" title describes the typical pattern of an English dinner party – the idea of the folder was that new staff could find out from it how conferences were organized from first receiving the referral or data on the case (soup) all the way through the possible outcomes and processes, to the final feedback contacts with the parties (nuts).

Financial Cost Data

Alongside statistics of cases and this detailed description, it is also important that schemes keep details of their own scheme's financial operations. It will never be possible to see whether the scheme is value for money unless the financial cost of running the scheme over a certain period of time and undertaking a certain number of cases is known. Value for money, on the British definition of value for money in public services, is achieved if the cost of running the scheme is exceeded by the financial benefits of its work in terms of reoffending prevented or well-being created (if well-being can be measured in financial terms, as it is in health economics and, for example, by the English National Institute for Clinical Excellence). Details of the kinds of financial cost figures that need to be kept are given in Shapland et al. (2008), but essentially they include salaries or fees to personnel, buildings and premises costs, travel costs reimbursed to participants, publicity, insurance, training, and so forth. The costs need to be able to summed over a particular time period, which is the same as that used for case management summary reports, so that it is possible to calculate a cost per case (even though such an idea of an "average" case is difficult for restorative justice, where cases may stop at different stages) or a cost per facilitator.

Some might comment why bother undertaking the burden of keeping such cost and time data, when the scarce time of personnel could be used in undertaking more restorative justice work? The answer is that cost and time data are not just for the benefit of economists, hard-edged governments wanting value for money, or criminal justice practitioners who hold budgets but also for schemes themselves. They can help to answer questions such as the following: Does direct mediation take longer? Can one split it from the effort needed to be put in to undertake indirect mediation? What are the costs of recruiting, training, and managing volunteers compared to paying a fee per case or employing paid facilitators? All would agree that proper preparation of participants is vital, but over what length of contact should it be encouraged? How much more costly is work in prison (given the need to have security personnel present, get all participants to the prison, etc.) compared to work in the community? These are key questions for the setting up and development of any scheme – and for those funding restorative justice. It is important that new developments or new areas where referrals are taken are coded as such in the data. It can be very difficult to work back through financial data to undertake these costs analyses retrospectively. It is much easier to work out a simple system of codes to allocate to payments as they are made.

Evaluating What the Scheme Is Trying to Achieve: Specifying Aims and Outcomes

Some practitioners and academics tend to be referring to monitoring when they talk about evaluation. Evaluation, however, is a rather more long-term and fundamental exercise – looking at whether the scheme has achieved its desired outcomes. Monitoring is not sufficient. Any scheme also needs to know whether it is achieving its desired outcomes or aims. It also needs to know how its services are being perceived by participants and where it may need to make changes.

Specifying Aims: The First Key Step

In order to start on evaluation, the scheme needs to know what its aims are in sufficiently clear and precise terms so that it is possible to measure whether they have been achieved. Unless the scheme has spent time in working this out and in agreeing what the precise aims are, it is impossible to set out what its outcomes are supposed to be and so evaluation is close to impossible.

This would seem to be obvious. However, the history of crime reduction and criminal justice evaluation is a rather sorry one in this respect – and it is important restorative justice evaluation does not follow the same path. In England and Wales, for example, considerable amounts of money have been invested by government, police, and local authorities on crime reduction initiatives. Yet some of these initiatives have not been implemented as intended, and others have been almost impossible to evaluate (Hamilton-Smith 2004; Hope 2004). One of the reasons has been a lack of precision in the intended aims of the initiatives or a change in those aims during the initiative. The problem is that initiatives designed to reduce burglary may be intended just (a) to reduce recorded offences of burglary; and/or (b) to reduce other linked crime, such as criminal damage; and/or (c) to reduce fear of burglary; and/or (d) to reduce fear of crime more generally; and/or (e) to make residents feel the police or other local authorities are doing their job in the local area better. All these aims are possible; more than one may be desired – but each creates a different outcome measure. If the program does not specify exactly which aims are its priority, the evaluation has real problems.

Restorative justice has potentially as many aims as crime reduction programs. Is a particular restorative justice scheme aiming to:

- Achieve particular aims for participants (e.g., increase offenders' empathy, restore victims' hurt, provide a sense of closure for victims or offenders, reduce offenders' reoffending)
- Provide a process whereby participants may find their own solutions to situations/conflict (e.g., by allowing each to state the effects of the offence on themselves and ask questions, allowing each to have a fair say in resolution, encouraging a joint problem-solving agreement, providing a safe space for discussion)
- Provide a more civilized or cheaper or more participatory alternative to traditional criminal justice.
- Or more than one of these?

Each of these aims has been advocated as an aim of restorative justice. In our evaluation of three large restorative justice schemes in England and Wales, the schemes differed as to how the managers perceived their relation to criminal justice, for example, and so whether measuring reoffending was a relevant outcome measure (Shapland et al. 2006a). Facilitators within the same scheme put emphasis on different aims, because typically each scheme had multiple aims.

It is important to realize that each aim implies that specified people should judge whether it has been achieved. The first aim above, for example, is primarily about participants' perceptions, rather than schemes' or referrers' perceptions. It needs to be evaluated using data from participants or relating to participants. The second relies on participants as well but could also be measured by observers at restorative justice meetings and through questions to facilitators. The third could be judged by all of these, but it can be argued that if it is a publicly funded service, then criminal justice system professionals and policymakers should also be judging its worth against other alternative processes in criminal justice.

Exactly which aims are seen by the scheme as paramount will depend upon its history but also upon its ideology and the theoretical mechanisms through which it believes restorative justice is impinging on the participants and others. Aims are inseparable from theory. Many theoretical strands have combined to make up the rich tapestry of modern restorative justice, and they differ according to the key purposes they see for restorative justice (Shapland et al. 2006b). Zehr (1990), for example, stresses its healing and conflict-reduction potential and this is a dominant strand in victim-offender mediation, for example, in Nordic countries. Australasian restorative justice has stressed the democratic, participatory, emotional, and procedural potential of restorative justice conferencing (Braithwaite 1999; Harris et al. 2004). Recent schemes in England and Wales, following the dominant crime reductive imperative of national and local criminal justice agencies and government, have sought its potential in reducing offending, though our own evaluation has always stressed that both reducing offending and responding to victim needs must be equal aims (Shapland et al. 2011).

For restorative justice, specifying theoretical mechanisms is even more important than for crime reduction. This is because restorative justice is an intrinsically process-based initiative. It cannot be seen in evaluation terms as a black box model, where participants are put in one end, a mediator added, one shakes them all together, and outcomes come out at the other end. For restorative justice workers, it is precisely what happens in the middle, during the process, which is important. How exactly is it expected that victims will be reinstated or restored? What will be a sign of this? Are there mediations or conferences where this occurs more often or where conditions seem more fruitful? How can we make those occur more often? Is it something the mediators say? Or the surroundings? Or what? Similar questions can be asked in relation to offenders.

Setting Out Realistic Outcome Measures: The Difference Between Outcomes and Outputs

Evaluation is primarily about measuring outcomes – representing the final aims of the scheme. However, commonly, measuring outcomes takes some considerable time. If one aim of a scheme is to reduce reoffending, this is typically done by looking at reconviction rates or rearrest rates (in the USA). It is necessary to leave a period of time for reoffending to occur (1 year minimum currently in the UK and in the USA) and for official systems to record it (i.e., cases to reach conviction and sentence). It is important that reoffending be checked to see if it is being reduced in relevant ways (is it most relevant to measure likelihood of reoffending, or frequency, or seriousness, or all three?), but this tends to be a complex and costly business, because of the ways in which official conviction or arrest data are held. Would it then be sensible to measure intermediate goals called "outputs" more often, restricting evaluations looking at the final reconviction outcome to periodic sweeps every 5 years or so? Are there relevant outputs that could be measured? For reoffending or official conviction, it may be that successfully completing a release program from prison or a community program may be a proper output measure (though there is always the danger of contamination of the results if the criminal justice program workers know which offenders have been through restorative justice and are either favorable or very unhappy about it).

Restorative justice has been accused more recently of just relying on "satisfaction" measures (i.e., sending small questionnaires to participants immediately after closing the case asking if they were "satisfied" with the process and the result of the restorative justice). There is nothing wrong with sending out satisfaction questionnaires – they provide an opportunity for participants to feedback their experiences and some idea for supervisors as to whether any particular type of case or individual facilitator is providing problematic. They are a typical, simple output measure of participant perceptions. What is important, though, is that periodically it is checked that when an independent researcher does a more in-depth interview with participants at different times after the restorative justice is finished, it is considered whether there are different views about preparation, processes, outcomes, and feedback by the scheme about the result of the case and whether "satisfaction" immediately after the case ends continues for some months afterwards. Few restorative justice evaluations have been able to look in detail at all these elements, but Shapland et al.'s (2011) evaluation of victims' and offenders' views found that views on different aspects of the process and outcomes tended to be highly correlated and that there was not a big dip in positive views with time. They also found, though, that schemes might be regarded highly in terms of facilitator performance and outcome agreements but still fall down, for example, on feedback to participants about whether offenders had completed agreement items or what had happened to cases later in criminal justice.

Minimum Evaluation Is Not Optional

Some evaluation periodically is essential for all schemes. Without minimal evaluation which goes beyond monitoring, even a very major and potentially very successful program will leave nothing behind. It may have benefited the participants and criminal justice practitioners who have been involved, but their memories will fade in time and no one will be able to build on those useful foundations to develop or start new programs in the future. The minimum evaluation which is needed for all schemes and for all work funded through the public purse therefore seems to me to be:

- Monitoring.
- Process evaluation which provides a clear description of the scheme and its processes: a how to do it manual.
- Identification of clear outcome measures: what the scheme aimed to do and how it expected these outcomes to occur (its theoretical view).
- Evaluation using outcome measures periodically, supplemented as necessary by output measures (such as participants' feedback and satisfaction) regularly. Ideally, this evaluation will be conducted by independent researchers or data gathered by the scheme analyzed by researchers independent of the scheme and its funders.

Further Important Questions for Evaluation

Funders and developers of restorative justice, as well as schemes themselves, however, will have many further questions. Answering these will often, however, require more specialized help from different academic disciplines. They may be matters which each country or scheme will want to ask at intervals, in a planned program of developing restorative justice through evaluation. They are certainly elements on which countries trying to develop restorative justice will wish to try to acquire comparative data, so that it is possible to think about different forms of restorative justice and different procedures which provide the most suitable platforms for participants' particular needs.

Consideration of Human Rights and Legal Issues

The first specialist area relates to the fact that restorative justice done in connection with criminal justice will raise many legal and human rights issues. This is so whenever the action at the heart of the justice process involves behavior which is against the criminal law, though the precise issues will differ when the restorative justice is never dealt with formally by criminal justice (community restorative justice), when it is diversionary (i.e., the case no longer will be dealt with by criminal justice personnel), or when it feeds into criminal justice processes (e.g., presentence or prerelease restorative justice).

Some key issues are:

1. Examining the scheme's processes to ensure participants have adequate information and explanation about the process and its outcomes (including whether outcome agreements can be enforced and what happens if they are not completed) to be able to give informed consent on participation. This includes any likely effect on criminal justice proceedings (see Campbell et al. 2006; Daly 2003 for analyses of such effects in relation to young participants).
2. Considering power imbalances between participants not only in terms of the consequences of the offence but in criminal justice terms. This can be very complicated in terms of victims and offenders (Shapland et al. 2006b), even when the offence does not itself present unusual power issues (participants know each other well, live near each other, the offence involves interpersonal power).
3. Thinking how participants may be able to maximize communication within the restorative justice process (translators, dealing with disabilities, etc.). Communication is one of the key benefits both victims and offenders experience from restorative justice, compared to criminal justice (Shapland et al. 2011). Similarly, one of the reasons given for dissatisfaction in conferences, in the few cases in which this occurred in the same research, was when communication was difficult to achieve. If the aim of restorative justice is to facilitate communication and participation, then personal or situational difficulties need to be minimized.
4. If the case involves a serious offence or is still within criminal justice, then it may need to be ensured that offenders (sometimes others as well) can access legal representation in order to advise them on whether to participate in restorative justice.
5. The question of how open the restorative justice event itself is a key one. Normally, events (direct mediation meetings, conferences, circles) are private to the participants and facilitators. This has correlated with the dominant use of restorative justice allied to criminal justice for young offenders, for diverted cases, and for minor criminal cases where it is seen as similar to civil dispute resolution. The typical paradigm in these instances would have been a closed process if the case had been dealt with through normal criminal justice processes. However, often, these events are also open to observation by those training as facilitators and sometimes to other practitioners – though permission should be sought from the participants for this. Normally, they are not videoed or recorded, though a report is often prepared on the case to be sent to those referring it to the scheme. There are major questions about accountability in relation to

restorative justice, which has not often been considered (Dignan et al. 2007). If, for example, there were to be a disturbance during the event, or a complaint afterwards about the mediator/facilitator, what mechanisms exist to answer those questions? Is there a need to have some form of recording to ensure accountability to participants? Moreover, if restorative justice does increasingly deal with more serious offending, is there a need for recording to provide accountability to judges, referrers, those hearing appeals, etc.? Recording itself does not necessarily breach privacy, as long as that recording remains within the group of those properly concerned with that case.

Evaluation to Develop a Better Appreciation of What Restorative Justice Is Achieving

In order to discern whether restorative justice schemes are delivering their aims, there will need to be periodic investment in more in-depth and specialized evaluation. Just one research design may not suit all aims. For example, the current most stringent research design to measure decreases in reoffending is a randomized controlled trial (RCT) (see Sherman et al. 1997). In the evaluation of three large schemes in England and Wales, one of the schemes evaluated, Justice Research Consortium, used this design, and this enabled the evaluators to provide results on reoffending devoid of selection effects in participants volunteering to participate in restorative justice (Shapland et al. 2011). However, the use of RCT is rare in criminal justice research.

The use of an RCT is valuable in considering reoffending, but it does not by itself establish what it is about restorative justice which was found particularly beneficial by those victims (and offenders) who experienced it. An RCT can only look at whether victims who experienced restorative justice were more satisfied, or more helped, than victims who only experienced criminal justice, not the mechanisms by which this was occurring. Considering what elements of restorative justice were most helpful needed to be assessed by detailed interviewing of victims (and offenders) at different points after the process ended and comparison of their answers with observation of the conferences they attended (where this was possible) (Shapland et al. 2011).

Moreover, an RCT is unhelpful in determining cost value parameters, unless it is accompanied by accurate time measurements by schemes of the time taken on experimental group cases, compared to that on control group cases – and this was not always the case (Shapland et al. 2008). This is because the time taken on control group cases randomized out of the experiment after consent by victim and offender is obviously far less than that on experimental group cases – and one needs to know both in order to work out whether the financial cost of any decreases in reoffending attained by the experimental group over those achieved by the control group was greater than the cost of the scheme.

The choice of evaluation methods, therefore, depends very much on exactly which aim is being evaluated. Evaluations so far have concentrated upon participants' views, whether the restorative justice was completed and agreements made, whether outcome agreements were fulfilled, and reoffending (see Shapland et al. 2011 for a review of the findings in relation to youth and adult restorative justice). A concise summary of the results would be that, overall, participants tend to say they are highly satisfied with processes and usually with outcomes. Where outcome agreements were made, they were fulfilled at least to the same extent that similar criminal justice outcomes (e.g., financial penalties, community work) were occurring in that country at that time. There is some limited, but positive evidence on reoffending, particularly for direct meetings.

However, there are key remaining areas where evaluation in different countries is still lacking. They include:

- Why do some potential participants decline to take part? Most evaluations so far have been confined to those who do take part.
- What are participants' expectations of restorative justice – and were they fulfilled? Answering this question suggests interviewing participants before and after participation, which has rarely been done.

- What happens to those who do not agree to participation in restorative justice? Do they still obtain victim/offender support services?
- Which kind of procedure is most suitable for which kinds of cases and participants? There is some limited evidence that direct meetings (mediation or conferencing) produce greater satisfaction and greater drops in reoffending than indirect (shuttle) mediation (Shapland et al. 2011), and that conferencing is more effective in reducing power imbalances between participants, because the presence of victim and offender supporters is a mitigating one (Strang 2002; Shapland et al. 2011), but there have been very few direct comparisons between procedures.
- What would happen if there were no scheme – so to what extent is this scheme and its way of working better than none, or alternative methods available in that country?

In many ways, we are only at the beginning of thinking through the most effective ways of evaluating restorative justice and of exploring its potential for all its participants and the communities in which they live. As we accumulate more results of monitoring and more specialist evaluation results, hopefully that future may unfold. Continuing comparative ways of working, evaluating and discussing evaluation results will continue to play a crucial role.

Related Entries

- ▶ Examining the Effectiveness of Correctional Interventions
- ▶ History of Restorative Justice
- ▶ Restorative Justice and Practice
- ▶ Victims and Restorative Justice

Recommended Reading and References

Braithwaite J (1999) Restorative justice: assessing optimistic and pessimistic accounts. Crime Justice: Rev Res 25:1–127

Campbell C, Devlin R, O’Mahony D, Doak J, Jackson J, Corrigan T, McEvoy K (2006) Evaluation of the Northern Ireland youth conferencing service. NIO research and statistical series report no. 12. NIO, Belfast, at http://www.nio.gov.uk/evaluation_of_the_northern_ireland_youth_conference_service.pdf

Daly K (2003) Mind the gap: restorative justice in theory and practice. In: von Hirsch A, Roberts J, Bottoms AE, Roach K, Schiff M (eds) Restorative justice and criminal justice: competing or reconcilable paradigms? Hart Publishing, Oxford

Dignan J, Atkinson A, Atkinson H, Howes M, Johnstone J, Robinson G, Shapland J, Sorsby A (2007) Staging restorative justice encounters against a criminal justice backdrop: a dramaturgical analysis. Criminol Criminal Justice 7:5–32

Farrington D (1997) Evaluating a community crime prevention program. Evaluation 3(2):157–173

Farrington D (1998) Evaluating “communities that care”: realistic scientific considerations. Evaluation 4(2):204–210

Hamilton-Smith N (2004) The Reducing Burglary Initiative: design, development and delivery. Home Office Research Sudy 287. Home Office, London

Harris N, Walgrave L, Braithwaite J (2004) Emotional dynamics in restorative conferences. TheorCriminol 8:191–210

Hope T (2004) Pretend it works: evidence and governance in the evaluation of the reducing burglary initiative. Criminal Justice 4(3):287–308

Justice Research Consortium Thames Valley (2004) From soup to nuts: a manual for restorative conferencing. Thames Valley, Bicester

Marshall T (1999) Restorative justice: an overview. Home Office Occasional Paper, London, Home Office

Pawson R, Tilley N (1998) Caring communities, paradigm polemics, design debates. Evaluation 4(1):73–90

Restorative Justice Council (2011) Best practice guidance for restorative practice. Restorative Justice Council and Ministry of Justice, London

Shapland J, Atkinson A, Atkinson H, Chapman B, Colledge E, Dignan J, Howes M, Johnstone J, Robinson G, Sorsby A (2006a) Restorative justice in practice: the second report from the evaluation of three schemes. The University of Sheffield Centre for criminological research occasional paper 2. Faculty of Law, Sheffield, at http://www.sheffield.ac.uk/law/research/clusters/ccr/occasionalpapers

Shapland J, Atkinson A, Atkinson H, Chapman B, Colledge E, Dignan J, Howes M, Johnstone J, Robinson G, Sorsby A (2006b) Situating restorative justice within criminal justice. Theor Criminol 10:505–532

Shapland J, Atkinson A, Atkinson H. Chapman B, Dignan J, Howes M, Johnstone J, Robinson G, Sorsby A (2007) Restorative justice: the views of victims and offenders. Ministry of Justice Research Series 3/07. Ministry of Justice, London

Shapland J, Atkinson A, Atkinson H, Dignan J, Edwards L, Hibbert J, Howes M, Johnstone J, Robinson G, Sorsby A (2008) Does restorative justice affect reconviction? The fourth report from the evaluation of three schemes. Ministry of Justice Research Series 10/08. Ministry of Justice, London

Shapland J, Robinson G, Sorsby A (2011) Restorative justice in practice. Routledge, London

Sherman L, Strang H (2007) Restorative justice: the evidence. Smith Institute, London

Sherman LW, Gottfredson D, MacKenzie D, Eck J, Reuter P, Bushway S (1997) Preventing crime: what works, what doesn't, what's promising. US Department of Justice, Washington, DC

Strang H (2002) Repair or revenge: victims and restorative justice. Clarendon, Oxford

Zehr H (1990) Changing lenses: a new focus for crime and justice. Herald Press, Scottdale

Moral Crimes

Robert Meier
School of Criminology and Criminal Justice,
University of Nebraska at Omaha, Omaha,
NE, USA

Synonyms

Crimes without victims; Public order crimes

Overview

The study of moral crimes requires awareness of differences in the degree of moral condemnation of conduct. While a number of crimes seem to generate a consensus regarding their immoral character, there are some that do not. These crimes set the stage for conflict and to understand their nature one must be sensitive to the role of power in putting such acts within the criminal law.

What Are Moral Crimes?

In April 2009, the Supreme Court of the State of Iowa ruled unanimously to permit same-sex marriage. In November 2010, three of the judges, who were on the ballot for their job retention, were defeated and removed from office, largely because of their vote on same-sex marriage. One poll found that 41 % of Iowans said they would vote for a ban, while 40 % said they would not; the rest said they would not vote or were undecided (Clayworth and Beaumont 2009). And 92 % said the legalization of gay marriage had brought no real change to their lives. Clearly, the political drive to target the judges in November 2010 was well orchestrated and political in nature.

Chief Iowa Supreme Court Justice Mark Cady defended the 2009 decision based solely on the law and consistent with the Iowa Constitution; he said: "... the public should understand the important voice they have and be informed of the consequences of using that voice to remove judges for simply doing their jobs" (Nation 2011).

A similar scenario played out in New York when legislators there legalized same-sex marriage in 2011. There, four Republican legislators who backed gay marriage faced an uncertain future from their constituents in the next election, constituents who were very vocal in claiming their opposition to the votes of the legislators (Hakim et al. 2011).

What happened in Iowa and New York has been happening for a long time in this country. It was not long after the founding of this country that early settlers discovered that the law could be a resource and that law was not some independent, value-free force in social life. Groups learned that the law could be used to protect their interests and values. It unmistakably began in colonial times when the content of many statutes were lifted wholly from the Christian Bible (Erikson 1966). Even later statutes contained biblical phrases or intents.

It is not surprising that law should reflect moral values. Few people, including those who are not especially religious, would consider such crimes as murder, rape, and the theft of other people's property as immoral; rather, there are strong moral pronouncements against such behavior and a strong consensus that such acts violate both law and moral sentiments. Murder, rape, and robbery are condemned everywhere, even if there are exceptions to the degree of these acts' wrongfulness and therefore the offender's legal responsibility. One traditional argument about the role of law is the reinforcement and protection of the mores, those strongly held beliefs of a people. And, by extension one could ask: “If behavior isn’t immoral, why would there be a law against it?”

But there are other crimes about which there is no consensus as to their morality. There are disagreements, some strong, others weaker, about the immorality of the use of certain drugs, consensual sexual acts among adults such as prostitution, and pornography. Many disagreements also exist with respect to abortion. To some abortion is clearly immoral and condemnable; others disagree. In these so-called moral crimes, the law has come to be used to side with those who are offended by such acts. Some observers have claimed that such conflicts are the result of shifts of “moral time” (Black 2011), while others express the dynamic in terms of shifting power arrangements (Quinney 1970). Regardless of how it is characterized, shifts in opinions, disputes, claims, or differences in moral evaluations mark the border of moral crimes.

To say that there is general agreement about the immorality of many common crimes should not suggest that every single person shares that evaluation. Nor should it suggest that conceptions of immorality alone will prevent such crimes. In spite of strong condemnation of murder, there are murders every year. So too are there robberies, burglaries, and assaults. Yet, there is strong consensus that such acts ought not to happen.

If prostitutes and other sex workers are “moral criminals,” sex offenders are not, regardless of what they actually did. The term “sex offender” conjures horrific images of adults physically harming children. But many of the people on sex offender registries have not engaged in that behavior. Not all registrants are child molesters (Lancaster 2011a). Statutory rapists, who commit an offense involving neither coercion nor violence, are registrants in some states. Some states require “Peeping Toms” and exhibitionists to be on sex offender registries. Louisiana forced some prostitutes to register, and two-thirds of those on the North Carolina registry had been convicted of the nonviolent crime of “indecent liberties with a minor,” a crime which does not necessarily involve physical contact (Lancaster 2011b). In some states, indecent exposure can land an offender on a sex offender registry. There are also other issues with sex offender registries that warrant extensive discussion but are beyond this entry (Sample and Evans, 2009).

There are, to be sure, offenders who do physical and emotional harm to children, as alleged in the case at Penn State University. Such offenders should be dealt with harshly but we should not be guided in our reaction even to these serious offenders by stereotypes that all such offenders are likely to be recidivists. One federal study found that only 5 % of released sex offenders (those who committed forcible rape or sexual assault) were rearrested for a sex crime 3 years after their release from prison (Durose et al. 2003). This recidivism rate is below that of many other offenders, such as those convicted of burglary and robbery.

The Narrowing of Morality

Gay and lesbian behavior is not illegal in the United States but this does not mean that there are no legal issues surrounding the lesbian, gay, bisexual, and transgendered (LGBT) community. These issues include the status of gay marriage, the adoption of children by LGBT couples, the extent to which legal restrictions offend equal protection standards, and the presence of gays in the military, among others. Lesbian and gay rights continue to be problematic in the United States because of different evaluations of these people, their conduct, and perceptions of the origins of homosexuality.

That said, it seems clear that gay men and lesbians have made much progress, even if the gay people do not still enjoy the same privileges and rights as straight people. In 1986, the US Supreme Court upheld a sodomy law in Georgia (Bowers v. Hardwick), but an opposing decision made in 2003, which also dealt with a sodomy law (Lawrence v. Texas), left little doubt that courts were acting more positively toward gay issues. In the Lawrence decision, the justices believed that such laws “demeaned the very existence of gay people” (Carpenter 2011).

In 1998, four Harris (Houston) County Sheriff’s deputies, responding to a false report of someone waving a gun, entered an apartment to find John Lawrence and Tyron Garner allegedly (Lawrence and Garner denied they were having sex) violating the Texas law against sodomy. Lawrence was said to be belligerent, which was apparently a major factor in the arrest of the two men.

Lawrence and Garner were unlikely people to have figured too strongly in this major US Supreme Court case (see Carpenter 2012). Both had criminal records and with Lawrence, being white, and Garner, a black man, the relationship had a racial dimension to the case. Texas had a long history of making sodomy criminal. Legislation was refined over time to reflect different examples of possible homosexual behavior, and in 1943 oral sex was added to those acts made illegal. But the law stipulated that oral and anal sex would be prohibited only with another individual of the same gender. One observer said: “Put bluntly, it was now legal in Texas to have sex with a farm animal but not with someone of the same gender” (Oshinsky 2012: 10).

In other developments, the prestigious Washington, D. C., law firm of King and Spalding informed the United States Justice Department in 2011 that it would no longer assist Congress in defending the law that forbids the federal recognition of recognizing gay marriage (Carpenter 2011). The law, called the Defense of Marriage Act, was passed in 1996.

As of early 2012, seven states had legislation to give same-sex couples the right to marry. Those states – Connecticut, Iowa, Massachusetts, New Hampshire, New York, Vermont, and Washington State (and the District of Columbia) – are likely to be joined shortly by other states. The New Jersey Senate voted in February 2012 to legalize same-sex marriage, a shift from 2 years previously when a similar measure failed.

But while progress has been made, the history of homosexual rights has been a rocky one and progress has sometimes been slowed, perhaps because the cause of gay rights has not had a national figure to serve as champion. Black civil rights had Martin Luther King and feminism had Gloria Steinem, both of whom championed their causes very successfully. But gay rights activists have been more local figures, such as Harvey Milk in San Francisco, and it is difficult to think of gay activists who have captured a national stage. As a result, most progress toward equal protection of gays has been slow and sporadic.

The victories have been small but not meaningless. The Bureau of Labor Statistics, for example, examined the extent to which domestic partner benefits were enjoyed by all workers. Thirty-three percent of state and local government employees had access to health benefits for same-sex couples, slightly higher than the 29 % of employees in the private sector (Tavernise 2011a). In addition, California became the first state to require public schools to teach gay and lesbian history in 2012. The legislation followed a strongly publicized series of suicides among gay teenagers (Lovett 2011).

The result of local battles suggests that the stigma of being gay is still strong but weakening as people are becoming more accepting. For example, San Francisco is still considered a stronghold for gays but there are other places that are ranked higher than San Francisco in same-sex couples per 1,000 population (Tavernise 2011b).

Perhaps more compellingly, the repeal of “don’t ask, don’t tell” in the United States military set an official tone of increased tolerance. Signed into law in December 2010, the act repealed the discriminatory policy that forced American military men and women to serve with anxiety and isolation. The repeal of don’t ask, don’t tell was an affirmation that even in the military, sexual preference is irrelevant.

The Politics of Moral Crimes

Early forms of criminal law came about when the state replaced the tribal unit as the dispenser of justice. In so doing, the state (usually in the form of a monarchy) became the unit that administered procedures and sanctions; the blood feud was replaced by the king's punishment. Since collective responsibility of the tribe or clan no longer existed, the individual offender was punished and restitution for a large number of offenses disappeared (Pennington 1993). Offenses for which there was no compensation were considered crimes, while tort actions were those for which compensation was available. Over time, the number of crimes increased.

Criminal law was used to combat social problems associated with increased industrialization and urbanization both in Europe and the United States. With the manufacture of inexpensive alcoholic beverages, drinking – and drunkenness – was democratized to the lower classes. About the same time, increased industrialization witnessed the increased need for labor to run the factories. The Gin Act of 1736 and the Vagrancy Act of 1744 were designed to address these specific problems in England.

Clearly, some law is the result of processes other than a collective move to reinforce or amplify group mores. The creation of "Sunday laws" which make illegal some activities on Sunday, such as the sale of alcohol and the prohibition of certain kinds of labor, reflects the interests of religious groups who felt the need to protect that day as special from other days.

If there is, by definition, differences of moral opinion regarding these crimes, how has legislation been possible in defining these acts as illegal? Most explanations for offenses where we are all agreed on the wrongfulness of the act suggest that these laws simply grow out of our common conceptions of what is right and wrong. As noted earlier, the content of laws in colonial America was undeniably religious in nature. An analysis of these laws and their enforcement found that crimes brought before the Essex County (Massachusetts) court from 1636 to 1682 were, in order of their frequency, disturbing the congregation, absence from church, contempt of clergy, criticism of the government, and "delivering first child within too short a period of time after wedding" (Erikson 1966: 171).

But an explanation for the existence of laws where there is moral disagreement is lacking. In these instances, theories of law that recognize such conflict is an explicit characteristic of the legislative process are needed. Richard Quinney's (1970/2001) interest group theory of law serves as one example. Quinney begins with the observation that criminality is not determined by the nature of the behavior; rather, criminality is created and imposed by legitimate agents of social control. Crime is created politically. It is a judgment about the behavior of others who are perceived to be threatening to powerful, entrenched interests.

Conflict is an essential part of the law-making process. Law-making is unmistakably political – it was and is made by political entities (legislatures, Congress) and the judiciary, many of whom are elected officials. What is important is who has influence over those political entities and on what basis they are acting. The central concepts for Quinney are pluralism, power, and conflict.

Modern industrialized nations are pluralistic; they are composed of a multitude of groups who are defined by their interest structure. Some groups are defined solely by their interests and their pursuit of those interests in legislation. Consider the following interests: population control, labor issues, and gun rights. Each of these interests has groups organized with a particular position on the issue and, often, these positions are oppositional. Population control, such as the use of condoms and abortion, elicits strong and opposing views. The interest of the Roman Catholic Church and Planned Parenthood are discernibly different with respect to population control. Similarly, the interests of labor (e.g., unions) and management are often different, as are the interests of such groups as the National Rifle Association and the Brady Campaign to Prevent Gun Violence which seek different goals.

Regardless of the nature of group interests, another unmistakable fact is that there an unequal distribution of power among these groups. Not all groups have access to the same resource base. Some have more members than others; some have more money than others. As a result, some groups have more influence over the legislative process than do others.

As a result of differences in interests, implying conflict among groups, and power, it is inevitable that some interests will be protected in the legal system in the form of the content of laws and interpretation of public policy based on those laws. It is also inevitable, of course, that the content of law will change as some groups increase, and other groups decrease, the amount of power they can bring to bear on their interests.

Conclusion

The study of moral crimes requires awareness of differences in the degree of moral condemnation of conduct. While a number of crimes seem to generate a consensus regarding their immoral character, there are some that do not. These crimes set the stage for conflict and to understand their nature one must be sensitive to the role of power in putting such acts within the criminal law.

Recommended Reading and References

Black D (2011) Moral time. Oxford University Press, New York
Carpenter D (2011) How the law accepted gays. New York Times, April 29, p A25
Carpenter D (2012) Flagrant conduct: the story of Lawrence v. Texas: how a bedroom arrest decriminalized gay Americans. Norton, New York
Clayworth J, Beaumont T (2009) Iowa poll: Iowans evenly divided on gay marriage ban. Des Moines Register, September 20: Accessed online at: http://www.desmoinesregister.com/article/20090921/NEWS10/909210321/1001/NEWS
Durose MR, Langan PA, Schmitt EL (2003) Recidivism of sex offenders released from prison in 1994. Bureau of Justice Statistics, Washington D.C, Accessed online at: http://bjs.ojp.usdoj.gov/index.cfm?ty=pbdetail&iid=1136
Erikson K (1966) Wayward puritans: a study in the sociology of deviance. Wiley, New York
Hakim D, Kaplan T, Barbaro M (2011) After backing gay marriage, 4 in G.O.P. face voters' verdict. New York Times, July 4, pp A1 and A3
Lancaster RN (2011a) Sex panic and the punitive state. University of California Press, Berkeley
Lancaster RN (2011b) Sex offenders: the last pariahs. New York Times, August 21, pp 6–7
Lovett I (2011) California to require gay history in schools. New York Times, July 15, p A16
Nation C (2011) Iowa justice urges duty of citizens: be informed. Omaha World-Herald, April 17, 5B
Oshinsky D (2012) Flagrant conduct. The New York Times Book Review. March 18, pp 1, 10–11
Pennington K (1993) The prince and the law, 1200–1600: sovereignty and rights in the western legal tradition. University of California Press, Berkeley
Quinney R (1970) The social reality of crime. Little Brown, Boston
Sample L, Evans M (2009) Sex offender registration and community notification in Wright R, Sex offender laws: failed policies, new directions. Springer, New York, pp 211–242
Tavernise S (2011a) Access to health care benefits for gay partners is gauged. New York Times, July 27, p A12
Tavernise S (2011b) New numbers, and geography, for gay couples. New York Times, August 25, A1 and A4

Moral Norms

► Situational Action Theory

Morality

► Situational Action Theory

Mortality After Release from Prison

Stuart A. Kinner[1,2] and Ingrid A. Binswanger[3,4]
[1]Melbourne School of Population and Global Health, The University of Melbourne and School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia
[2]School of Medicine, The University of Queensland, Brisbane, Australia
[3]Division of General Internal Medicine and Division of Substance Dependence, University of Colorado School of Medicine, Aurora, CO, USA
[4]Denver Health Medical Center, Denver, CO, USA

Overview

Studies from Europe, Australia, the United States and Taiwan have shown that the rate of mortality in ex-prisoners is elevated in the weeks immediately following release, compared with subsequent weeks. These studies have also shown that mortality rates in ex-prisoners remain elevated, compared with the non-institutionalized population, for many years post-release. Several mechanisms have been hypothesized to explain the elevated risk and a key debate relates to the relative contribution of incarceration itself and the high level of underlying risk among former inmates. Preventive efforts need further development and evaluation, and methodological limitations hinder synthesis of the literature.

Introduction

Correctional services have a duty of care to prisoners in their charge, and in most jurisdictions, deaths in custody are closely monitored (Curnow and Larsen 2009; Noonan 2010). In Australia for example, deaths in police custody and in prison are routinely monitored and reported annually through the National Deaths in Custody Program (NDICP). Similarly in the United States, deaths in custody are reported to the Bureau of Justice Statistics through the Deaths in Custody Reporting Program. Almost all prisoners return to the community eventually, most after a relatively short period of time in custody; however, it has only been relatively recently that researchers have turned their attention to deaths after release from custody. One of the first such studies examined mortality outcomes among persons released from a 300-bed prison in Geneva, Switzerland, from 1982 to 1986; the researchers identified 102 unnatural deaths in the cohort with a disproportionate number occurring in the first year following release, often due to opiate poisoning (Harding-Pink and Frye 1988). In a subsequent study, Seaman and colleagues used record linkage to examine mortality outcomes for a cohort of 238 HIV-infected injecting drug users after release from prison in Edinburgh, Scotland. The researchers found that the risk of death (the "crude mortality rate," or CMR) from overdose was 7.7 times higher in the 2 weeks after release, than in the subsequent 10 weeks, highlighting an acute period of vulnerability after release from prison (Seaman et al. 1998). Subsequent studies in Europe, North America, and Australia confirmed that this phenomenon was widespread, with a meta-analysis finding that the risk of drug-related death was between 3 and 8 times higher in the first 2 weeks following release from custody, than in the subsequent 10 weeks (Merrall et al. 2010). Some studies have also shown that the risk of death in recently released prisoners is higher than that of current prisoners (Binswanger et al. 2007; Kariminia et al. 2007; Kinner et al. 2011), and higher than that of age- and sex-matched members of the general community (the "standardized mortality ratio," or SMR) (Binswanger et al. 2007; Kariminia et al. 2007; Rosen et al. 2008; Stewart et al. 2004).

Although drug-related causes account for a large proportion of deaths in former prisoners, most studies report that the majority of deaths are not drug related. Another key cause of preventable death in this population is suicide, which may reflect the very high rates of mental illness in prisoners, particularly among women (Fazel and Danesh 2002). For example, in a study of almost 250,000 ex-prisoners in England and Wales, Pratt and colleagues found that the rate of suicide in the first year after release from prison was 8.3 times higher for males and 35.8 times higher for females, than for age- and sex-matched members of the community. Reflecting the concentration of risk in the immediate post-release period, 21 % of these deaths happened in the first 28 days post-release (Pratt et al. 2006). Similarly, in a cohort of male ex-prisoners in New South Wales, Australia, Kariminia and colleagues found that the rate of suicide was 3.9 times higher in the first 2 weeks post-release, than after 6 months post-release (Karaminia et al. 2007). Although the risk of suicide is greatest in the immediate post-release period, there is some evidence that it remains elevated for much longer. In a study of 168,001 ex-prisoners in North Carolina followed for a median of 10.3 years, Rosen and colleagues found that the rate of death from suicide was significantly elevated for white ex-prisoners (compared with age- and sex-matched white people in the community), although not for black ex-prisoners (compared with age- and sex-matched black people in the community) (Rosen et al. 2008).

Drug overdose and suicide are the leading causes of death in ex-prisoners; however, other key causes of unnatural death in this population include motor vehicle accidents (MVAs) and drowning and, in some countries, homicide. One study of 9,381 ex-prisoners in Western Australia found that of 326 observed deaths, 29 % were drug related, 20 % were due to suicide, and 12 % due to motor vehicle accidents (Stewart et al. 2004). Similarly, in the Australian study by Kariminia and colleagues, 33 % of deaths were drug related, 16 % were suicides, 8 % were transport related, and 4 % were due to homicide (Kariminia et al. 2007). By contrast, in the Rosen study in North Carolina, only 12 % of deaths were drug related, 5 % were due to suicide, 11 % due to homicide, and 9 % due to motor vehicle accidents (Rosen et al. 2008). Understanding patterns in mortality in ex-prisoners is complicated by the fact that most deaths have multiple contributing causes, and coronial processes may differ between jurisdictions and over time. For example, suicides involving drugs may be classified as a suicide or a drug-related death; accidental deaths such as MVAs may be related to intoxication from alcohol or other drugs; homicides may be related to substance use, involvement in the drug trade, and criminal subculture or, particularly for women, the return to violent relationships.

Although the majority of deaths in ex-prisoners are due to unnatural causes, ex-prisoners are also at increased risk of death from natural causes such as cancer; liver, cardiovascular, and respiratory disease; and infectious diseases such as HIV and hepatitis C (Kariminia et al. 2007; Rosen et al. 2008), reflecting the high prevalence of health risk behaviors such as injecting drug use, unprotected sex, and tobacco use in this population (Binswanger et al. 2007; WHO 2007). Furthermore, although the risk of death is greatest in the period immediately following release from custody, this risk remains elevated relative to the sex- and age-adjusted non-institutionalized population for at least a decade (Karaminia et al. 2007; Rosen et al. 2008). To date, no studies have examined whether the key causes of death or risk factors for death in ex-prisoners vary over this time.

Not all ex-prisoners are at equal risk of death following release from custody. Using survival analysis, some studies have charted the probability of "surviving" (i.e., remaining alive) over time, from the point of release from prison, for different subgroups. These studies provide a graphic illustration of the precipitous drop in survival in the immediate post-release period, and show that different subgroups of ex-prisoners (for example, defined by sex, age and race) have a different probability of "survival" after release from prison. Record linkage studies have also been used to identify risk factors for death in ex-prisoners. Although findings have not always have been consistent, most studies find that each additional episode of imprisonment (and subsequent release) increases the risk of death; not surprisingly, a history of substance abuse and mental illness is also associated with increased risk of death (Singleton et al. 2003; WHO 2010). Findings regarding age have been mixed, with some studies finding that younger people are at greater risk of death, while others find the converse (Binswanger et al. 2007; Chen et al. 2010; Larney 2010). One study of drug-using ex-prisoners in Taiwan found that HIV infection was associated with increased risk of death while retention in methadone maintenance treatment (MMT) post-release was protective (Chen et al. 2010). In another study of drug-related death among ex-prisoners in the United Kingdom, a history of relationship breakdown and limited social support post-release were associated with death (Singleton et al. 2003), highlighting the importance of post-release support mechanisms. Many studies find that ethnic and racial minority groups are at greater absolute risk of death (Stewart et al. 2004), although when compared with race-matched peers in the community, this difference is usually attenuated (Rosen et al. 2008; Stewart et al. 2004). Conversely, although most studies find that the *absolute* risk of death is lower for women than men, the *elevation* in risk of death for women, compared with age- and sex-matched peers in the community, is consistently greater for women than for men (Larney 2010; Stewart et al. 2004). This apparent paradox is explained by the fact that women in the community have a lower age-adjusted rate of death than men, but this difference is smaller among ex-prisoners. This may be due to the greater underlying risk among women who get incarcerated than men (differential selection).

Hypothesized Mechanisms

Several mechanisms have been proposed to explain the elevated risk of death among former inmates, depending on the cause of death and the time since release. Early studies focused on an acute elevation in risk of drug-related death (Seaman et al. 1998). The presumed mechanism was diminished physiologic tolerance, with opioid (usually heroin) dependent individuals ceasing or reducing their drug use while incarcerated, and upon release relapsing to substance use and experiencing an accidental poisoning (overdose). Accordingly, the main response to prevent such deaths was to educate soon-to-be-released prisoners about drug tolerance, and counsel them to avoid drug use or at least consume smaller quantities. This approach remains central to many correctional authorities' response to mortality prevention in ex-prisoners. In some jurisdictions, opioid tolerance is maintained throughout the period of incarceration via opioid substitution therapy, such as methadone or buprenorphine. Although it is clear that a substantial proportion of preventable deaths in ex-prisoners is due to drug-related causes, it remains unclear what proportion of these deaths is attributable to reduced drug tolerance. Furthermore, while former inmates may have knowledge gaps about tolerance to drugs and its potential ramifications (Adams et al. 2011), it is not clear whether education alone is sufficient to reduce drug-related mortality.

In addition to reduced drug tolerance, researchers have identified a range of other factors that may help explain the acute elevation in risk of drug-related death among recently released prisoners. Key among these is poly-substance use, and in particular the combined use of multiple central nervous system depressants such as opioids, benzodiazepines, and alcohol. Poly-substance use is a contributing factor in many drug-related deaths among former prisoners (WHO 2010); however, in order to explain the acute elevation in risk of drug-related death in the immediate post-release period, poly-substance use would need to be more common or more intense during this time. There is some indirect evidence for this, with some researchers describing an acute period of injecting risk behavior in the immediate post-release period (Seal et al. 2001). Another potential risk factor is single-room occupancy: Many recently released prisoners experience difficulty finding accommodation and may resort to living alone in single-room dwellings; drug overdoses in such settings are more likely to be fatal because the victim is unlikely to receive assistance from other users or passersby. Another potential contributing factor relates to the transition from the predictability of the prison environment to an ever-changing and unfamiliar outside world. Due to the process of classical conditioning, it appears that drug effects are attenuated in familiar environments and potentiated in novel environments, such that drug use in a novel environment may produce greater intoxication (Siegel 2001). Finally, some authors have argued that risk of fatal overdose may be increased by systemic disease, in particular lung and hepatic dysfunction (Warner-Smith et al. 2001). Consistent with this, the prevalence of both tobacco smoking and hepatitis C infection is very high among prisoners (WHO 2007).

Although the majority of deaths in former prisoners are not drug related, there has been less consideration of the mechanisms underpinning other causes of death in this population. One contributing factor in suicide deaths is likely to be mental illness, which is highly prevalent in prisoners (Fazel and Danesh 2002). Given that many prisoners experience mental illness and only a small minority commit suicide post-release, mental illness may be better conceived of as a predisposing factor than a direct cause. Given the concentration of suicide deaths in the immediate post-release period, at least part of the explanation for these deaths must be related to the transition from prison to community. For vulnerable prisoners, particularly those with a mental illness, the transition to the community is likely to be a highly stressful time characterized by uncertainty, frustration, and disappointment (Binswanger et al. 2011b). Many ex-prisoners experience homelessness, poverty, and discrimination, and may have difficulty reestablishing relationships with family and friends. These challenges may be compounded by a sudden reduction in access to healthcare and other support services, which are more readily accessible in prison environments (Adams et al. 2011; Levy 2005). For those experiencing mental illness, one consequence of this may be reduced access to and adherence to medication, and a subsequent relapse to decompensated mental illness. In this context, intentional overdose and suicide have been perceived by some former inmates as a "way out" for intolerable conditions in the community (Binswanger et al. 2011b).

Poor access to health care may also be an important factor in other causes of death in ex-prisoners, particularly given the high prevalence of chronic and infectious disease in this population. For example, there is evidence that among HIV-infected injecting drug users, release from prison is associated with reduced adherence to antiretroviral medications and subsequent increase in viral load (Palepu et al. 2004). Even if appropriate health services are available in the community, unless recently released prisoners are connected with these services before release, they may "fall through the cracks" after release, failing to access appropriate health care until an acute health issue arises. Increased support during the transition from prison to community, and improved access to healthcare for recently released prisoners, may reduce the risk of death from both unnatural and natural causes; however, the links between transitional health services and mortality outcomes are poorly understood (Kinner 2010).

Prison as a Cause or Marker of Underlying Risk

Although there is clearly a strong association between release from prison and death, this does not necessarily imply that the release "caused" the death. As discussed above, the reasons for death after release from prison are likely to be complex and variable over time. One view is that prison is a "marker" for underlying risk and that death may even be delayed by imprisonment. That is, populations who move through prison systems are already at increased risk of death, because the factors predisposing to death, such as substance abuse, mental illness, impulsivity, and poverty, also predispose to crime and incarceration. Epidemiological studies consistently reveal a high prevalence of disadvantage, poor health, and health risk behaviors such as injecting drug use among prison samples (Fazel and Danesh 2002; WHO 2007). Also consistent with the view that imprisonment is a "marker" for mortality risk is evidence that ex-prisoners remain at elevated risk of death for many years after release from custody (Kariminia et al. 2007; Rosen et al. 2008). Although it is possible that the experience of imprisonment and release has a long-term impact on mortality risk (for example, through promoting marginalization and disadvantage, initiation into risky behaviors such as injecting drug use, and/or exposure to chronic infections such as hepatitis C or HIV), it is clear that many people entering prisons for the first time are already characterized by a number of risk factors for mortality.

Despite this some authors, pointing to the acute elevation in risk of death immediately after release from custody, have implicated correctional authorities, implying a causal relationship between incarceration and death. In this view, correctional authorities are responsible for the deaths of recently released prisoners. Not surprisingly, most correctional authorities reject this view, arguing that deaths occurring in the community are, by definition, a community responsibility. While this may be true in some cases, the situation is less clear for those ex-prisoners subject to some sort of post-release supervision. At least one study has examined mortality outcomes according to post-release supervision and found no relationship, suggesting that involvement with correctional services after release may not be protective against mortality (Binswanger 2011a). Whether or not correctional authorities are in any way "responsible" for the deaths of recently released prisoners, the well-established association between release from prison and death suggests some sort of causal relationship between release from custody and death.

Given the acute elevation in risk of death in the first few weeks after release from custody, and evidence of ongoing elevation in risk of death for years after release, prison may be both a risk marker and a cause of mortality in ex-prisoners. That is, release from prison precipitates a period of particularly elevated risk of death, but given the preexisting risk factors that characterize many ex-prisoners, this risk never returns to that of non-incarcerated community peers. From a public health perspective, assigning blame for the deaths of ex-prisoners is less important than preventing these deaths. Although the majority of deaths in ex-prisoners do not occur in the weeks immediately following release, a disproportionate number of deaths do occur in this period and in the context of limited resources, it is probably prudent to target preventive interventions in the immediate post-release period.

Prevention

A growing body of literature has described patterns of mortality in ex-prisoners, but to date, there have been few attempts to evaluate interventions designed to prevent these deaths. Based on a review of largely descriptive evidence, the World Health Organization (WHO) proposed strategies to prevent drug-related death in recently released prisoners at the policy and program level (WHO 2010). Although oriented to the prevention of acute drug-related deaths, many of these recommendations are also relevant to the prevention of death due to other causes. Among the key policy responses recommended by WHO is the adoption of an evidence-based drug treatment framework that diverts individuals with a substance use disorder into treatment rather than prison wherever possible, and the integration of prison and community services so that gaps in service and accountability are minimized. Given persuasive evidence that the criminalization of drug use and dependence has driven rapid increases in prison populations internationally (Pettit and Western 2004), the recommendation to minimize the incarceration of drug users seems well grounded. Efforts to improve service continuity and integration for ex-prisoners, and to reduce financing silos, may also prove beneficial, although evaluation findings to date have been mixed.

At the program level, WHO advocates for evidence-based practices including opiate substitution treatment (OST) in custody and post-release. Recent evidence suggests that in-prison OST is associated with reduced drug-related mortality in ex-prisoners (Larney 2010), although it remains unclear whether this is a direct effect due to the maintenance of opioid tolerance in prison, or an indirect effect whereby those on OST in prison are more likely to transition to OST post-release, with the latter providing the protective effect. WHO also endorses education for prisoners, prison staff and support workers to make all stakeholders aware of the increased risk of death post-release, strategies for the prevention of drug use, the risks posed by reduced drug tolerance and poly-substance use, and strategies for overdose prevention. Although there is little evidence that education alone can prevent deaths, education is appropriate as part of a broader prevention strategy, supported by pharmacotherapy in prison and post-release for high-risk individuals.

Consistent with a growing body of evidence, WHO (2010) joins a chorus of researchers specifically advocating for training the families and peers of at-risk prisoners in first aid, including the administration of naloxone, an overdose reversal drug used routinely by paramedics. There is growing evidence that with appropriate training, naloxone can be safely administered by peers and is effective in reversing overdoses, almost certainly saving lives (Lenton et al. 2009). In Scotland, naloxone kits and training are currently being provided to friends and family of soon-to-be-released prisoners, on a trial basis. If the evaluation of this initiative is favorable, it is likely that the same approach will be adopted elsewhere; however, naloxone is not a panacea. Naloxone is only effective if the overdose victim used opioids, and if the medication is administered by a bystander early and with at least basic knowledge of its use.

Given the complexity of the post-release experience, the broad range of causes of death, and the high degree of marginalization and stigma experienced by former inmates, effective prevention will likely require a combination of behavioral, environmental, service-based, and pharmacologic interventions. Financing mechanisms that support continuity and integration of care during the transition from prison to the community may also help ameliorate transitional risk, although few such initiatives have been rigorously evaluated. Similarly, diversion initiatives (e.g., drug and mental health courts) designed to reduce the imprisonment (and subsequent release) of drug-dependent and mentally ill individuals, by providing appropriate community-based treatment and supervision, may reduce mortality risk. Under-investment in these programs has limited their impact and rigorous evaluations have been scarce.

Despite the dearth of evidence regarding "what works" to improve health outcomes and reduce mortality in ex-prisoners, correctional authorities in many countries are embracing a "through-care" model, implementing policies and programs designed to facilitate a successful reentry. Although this is rarely the intention and has yet to be demonstrated empirically, such programs may also reduce mortality. Whereas deaths in prison are routinely monitored in some countries, the same cannot be said for deaths after release from prison. A recent Australian study estimated that the annual number of deaths among adult ex-prisoners in Australia, within a year of release from custody, is about ten times the number of deaths in custody over the same period (Kinner et al. 2011). Highlighting this largely forgotten group, the authors advocated for routine monitoring of ex-prisoner mortality through record linkage, which is both feasible and cost-effective in many developed countries. In the Australian context, monitoring of deaths in recently released prisoners will feature in a "national minimum dataset for prisoner health" (AIHW 2011) in the near future, allowing policy makers and researchers to examine annual trends in mortality in this group.

Many questions remain about the most effective way to prevent deaths in ex-prisoners, and there remains a need for further research (Kinner 2010; WHO 2010): First, to identify who among ex-prisoners is most at risk of death, when and why, so that preventive interventions can be appropriately targeted and tailored. And second, to rigorously evaluate both targeted and broad-based interventions designed to improve health outcomes for ex-prisoners, to reduce untimely deaths in this profoundly marginalized population.

Methodological Issues

Evidence-based prevention requires sound evidence; however, methodological limitations of existing studies have made interpretation of the evidence complex and quantitative synthesis of the evidence difficult. Methodological limitations of the existing literature can be grouped into four categories: sampling, data quality, analysis and reporting. Early studies typically included a single prison and/or selected a subgroup of prisoners thought to be at high risk: for example, HIV-infected injecting drug users (Seaman et al. 1998). Although these studies were the first to identify a period of acute risk immediately post-release, the incidence of mortality observed in these cohorts would be a poor estimate of population-wide mortality rates, and the risk factors for mortality may also differ at the (prisoner) population level. A related problem is that imprisonment may be defined differently in different countries. In the United States, jails are differentiated from prisons on the basis of several factors including conviction, sentencing status, and length of stay; it is likely that patterns of mortality will differ for those released from jails and prisons. The definition of a release can also be complex. For instance, at least one study has excluded release for extraordinary medical placement (medical parole) from the definition of release (Binswanger et al. 2007), on the premise that inclusion of persons released on the basis of imminent death will inflate the numerator. Work release, transfer to low-security facilities or to locked half-way houses all pose additional methodological challenges. Furthermore, in countries with a number of discrete, State-based correctional systems, such as the United States and Australia, it is unclear whether the findings from one State can be generalized to other States. More broadly, despite the fact that the majority of the world's prisoners are in low- and middle-income countries, almost all studies of mortality in ex-prisoners have been conducted in wealthy, western countries.

Studies of mortality in ex-prisoners rely on linkage of prisoner data with death data. Limitations of the quality and scope of source data, and of the linkage process, can compromise the findings and thus the accuracy of the conclusions drawn. Studies that involve linkage to a national death register will potentially identify all deaths in the cohort, except for those who are deported or otherwise leave the country before death. However, some studies have relied on State-based registers, missing those who move interstate and die. Others have relied on coronial databases to identify deaths, thereby excluding all deaths that have not been brought to the attention or a coroner. The linkage process itself is imperfect, as it relies on probabilistic matching of prisoner and death data on the basis of identifying information such as sex, date of birth, and name, often including multiple aliases. This introduces uncertainty into the linkage process and in the United States, there is likely to be greater uncertainty for individuals of Latino ethnicity and African-Americans (Boyle and Decoufle 1990), who are more highly represented among former inmates. As a consequence, there may be selective underestimation of mortality in these groups.

Key analysis issues include numerator bias, correction for reincarceration, and selection of reference population. Examination of the association between release from prison and death is complicated by the fact that many ex-prisoners return to custody and are subsequently released again, requiring the researcher to decide which release to consider for the purposes of analysis. For example, if a prisoner is released from custody and is reimprisoned 6 months later, spends a further 3 months in custody, is released again and dies 1 month later, what is the time between release and death: 10 months or 1 month? When examining the link between release from prison and death, the latter approach biases toward releases that are followed by death, introducing "numerator bias." Studies adopting this approach overestimate the link between release and death, and were excluded in a recent meta-analysis of this literature (Merrall et al. 2010). A related analysis issue is that in calculating crude mortality rates, some studies fail to exclude subsequent time in custody, despite the fact that the risk of death in custody is considerably lower than the risk in the community (Binswanger et al. 2007; Kariminia et al. 2007). Where this occurs, the rate of mortality may be substantially underestimated.

A final analysis issue relates to the calculation of standardized mortality ratios (SMRs), which compare the observed number of deaths in the study population with the expected number of deaths based on a reference population (usually the general population), adjusting for key demographic differences. SMRs show how much the rate of death is elevated in the study population compared with the general population: For example, in a cohort of Australian ex-prisoners followed for a median of 7.7 years, Kariminia and colleagues (2007) found that the all-cause SMR was 3.7 for men and 7.8 for women, indicating that male ex-prisoners were 3.7 times more likely to die than age-matched males in the general population, while female ex-prisoners were 7.8 times more likely to die than age-matched females in the general population. In another study, Rosen (Rosen et al. 2008) followed a cohort of male ex-prisoners in North Carolina for an average of 10.3 years and calculated an

SMR of 2.08 for white ex-prisoners (vs. white males in the community, adjusted for age) and 1.03 for black ex-prisoners (vs. black males in the community, adjusted for age). Although these findings indicate that the rate of mortality was elevated to a greater degree for white than black ex-prisoners, compared with age- and race-matched members of the community, they mask the fact that the mortality rate in the community is higher for blacks than for whites. Another Australian study (Stewart et al. 2004) has illustrated this in a cohort of 9,381 ex-prisoners followed for a median of 3.3 years: Among Aboriginal males, the SMR was 9.4 when the reference population was all age-matched males in the community, but fell to 2.9 when the reference population was restricted to age-matched Aboriginal males. The reason for this dramatic change is that in the Australian community, the rate of mortality among Aboriginal males is considerably higher than among non-Aboriginal males. Whether or not adjustment for race is appropriate is a matter of debate.

Interpretation and synthesis of the literature is also made challenging by heterogeneity in reporting of study findings. For example, while some studies report crude mortality rates (CMRs) and/or standardized mortality ratios (SMRs), others compare the rate of mortality across two different time periods (e.g., the first 2 weeks post-release vs. the subsequent 10 weeks). Different studies report CMRs and SMRs across different time periods (e.g., 1 year post-release, 5 years post-release), and some do not indicate the time period, making their findings difficult to interpret. While some studies report findings disaggregated by key demographic variables (e.g., sex, race), others present whole-population findings; only a minority have reported both. A number of studies have examined only drug-related mortality, while others have considered all-cause mortality, suicide deaths, or some other grouping of death codes. This heterogeneity makes both narrative and quantitative synthesis of the literature challenging.

In order to inform evidence-based policy and practice, a more sophisticated understanding of the incidence and risk factors for mortality in ex-prisoners is required. Most studies to date have used record linkage methods which, by definition, rely exclusively on routinely collected data. Such data are usually not collected for research purposes and as such, they tend to provide only limited information about potential risk factors for death, or circumstances surrounding death (Kinner 2010). Until electronic records completely supersede paper-based correctional and coronial records, novel research designs will be required to complement the knowledge already gleaned from record linkage studies, and advance the evidence base to inform preventive efforts. Such methods may also allow examination of mortality outcomes for ex-prisoners in resource-poor settings.

Conclusions and Recommendations

Prior research has suggested an elevated risk of death among former inmates relative to incarcerated populations and non-institutionalized, general populations. Additionally, it has demonstrated significant time trends, suggesting an early peak in mortality after release from custody. We recommend the inclusion of diverse countries, including resource-poor countries, in future research efforts. In terms of surveillance, we recommend systematic monitoring of mortality in this population, particularly given increasing rates of incarceration that have been observed worldwide and in certain countries (e.g., the United States) over the past 20 years. To enhance efforts at data synthesis, we suggest improvements to the design, analysis, and reporting of research findings. Furthermore, more intervention research across a broad range of outcomes and populations is needed. Intervention efforts, which may include policy changes, should be supported by rigorous data. Collaborative efforts between public health, corrections and advocates are most likely to be successful. Policy efforts should address structural problems such as financing silos. Special attention to physiologic changes, medical comorbidity, poly-substance abuse, environmental factors and health service barriers should be routinely included in transitional interventions, as purely educational or individual behavioral interventions are unlikely to be effective in isolation, given the complexity of the physiologic, structural and policy issues.

Related Entries

- ▶ Drug Abuse and Alcohol Dependence Among Inmates
- ▶ Early Release from Prison
- ▶ Health Issues in Prison Reentry Models
- ▶ Imprisonment and Cardiac Risk
- ▶ Parole and Prisoner Reentry in the United States
- ▶ Re-entry
- ▶ Suicide and Prisons
- ▶ Unintended Effects of Imprisonment

Recommended Reading and References

Adams J, Nowels C, Corsi K, Long J, Steiner JF, Binswanger IA (2011) HIV risk after release from prison: a qualitative study of former inmates. J Acquir Immune Defic Syndr 57(5):429–434

AIHW (2011) The health of Australia's prisoners 2010. Australian Institute of Health and Welfare, Canberra

Binswanger IA, Stern MF, Deyo RA, Heagerty PJ, Cheadle A, Elmore JG et al (2007) Release from prison – a high risk of death for former inmates. N Engl J Med 356(2):157–165

Binswanger IA, Blatchford PJ, Lindsay RG, Stern MF (2011a) Risk factors for all-cause, overdose and early deaths after release from prison in Washington state. Drug Alcohol Depend 117(1):1–6

Binswanger IA, Nowels C, Corsi KF, Long J, Booth RE, Kutner J et al (2011b) "From the prison door right to the sidewalk, everything went downhill," A qualitative study of the health experiences of recently released inmates. Int J Law Psychiatry 34(4):249–255

Boyle CA, Decoufle P (1990) National sources of vital status information: extent of coverage and possible selectivity in reporting. Am J Epidemiol 131(1):160–168

Chen CY, Wu PN, Su LW, Chou YJ, Lin KM (2010) Three year mortality and predictors after release: a longitudinal study of the first-time drug offenders in Taiwan. Soc Stud Addict 105:920–927

Curnow J, Larsen JJ (2009) Deaths in custody in Australia: national deaths in custody program 2007. AIC Monitoring Reports No. 03. Australian Institute of Criminology, Canberra

Fazel S, Danesh J (2002) Serious mental disorder in 23 000 prisoners: a systematic review of 62 surveys. Lancet 359:545–550

Harding-Pink D, Frye O (1988) Risk of death after release from prison: a duty to warn. British Medical Journal 297:596

Karaminia A, Law MG, Butler TG, Levy MH, Corber SP, Kaldor JM et al (2007) Suicide risk among recently released prisoners in New South Wales, Australia. Med J Aust 187(7):387–390

Kariminia A, Butler TG, Corben SP, Levy MH, Grant L, Kaldor JM et al (2007) Extreme cause specific mortality in a cohort of adult prisoners – 1998 to 2002: a data-linkage study. Int J Epidemiol 36(2):310–318

Kinner SA (2010) Understanding mortality and health outcomes for ex-prisoners: first steps on a long road [Commentary]. Addiction 105(9):1555–1556

Kinner SA, Preen D, Kariminia A, Butler T, Andrews J, Stoové M et al (2011) Counting the cost: estimating of the number of deaths among recently released prisoners in Australia. Med J Aust 195(2):64–68

Kinner SA, Forsyth S, Williams GM (2013). Systematic review of record linkage studies of mortality in ex-prisoners: why (good) methods matter. Addiction 108 (1):38–49

Larney S (2010) Opioid substitution treatment in prison and post-release: effects on criminal recidivism and mortality. University of New South Wales, Sydney

Lenton SR, Dietze PM, Degenhardt L, Darke S, Butler TG (2009) Naloxone for administration by peers in cases of heroin overdose. Med J Aust 191(8):469

Levy M (2005) Prisoner health care provision: reflections from Australia. Int J Prison Health 1(1):65–73

Lidberg L, Wiklund N, Jakobsson SW (1989) Mortality among criminals with suspected mental disturbance. Scand J Soc Med 17:59–65

Merrall ELC, Kariminia A, Binswanger IA, Hobbs MS, Farrell M, Marsden J et al (2010) Meta-analysis of drug-related deaths soon after release from prison. Addiction 105(9):1545–1554

Noonan M (2010) Mortality in local jails 2000–07. US Bureau of Justice Statistics, Washington, DC

Palepu A, Tyndall MW, Chan K, Wood E, Montaner JSG, Hogg RS (2004) Initiating highly active antiretroviral therapy and continuity of HIV care: the impact of incarceration and prison release on adherence and HIV treatment outcomes. Antivir Ther 9:713–719

Pettit B, Western B (2004) Mass imprisonment and the life course: race and class inequality in U.S. incarceration. Am Sociol Rev 69:151–169

Pratt D, Piper M, Appleby L, Webb R, Shaw J (2006) Suicide in recently released prisoners: a population-based cohort study. Lancet 368(9530):119–123

Rosen DL, Schoenbach VJ, Wohl DA (2008) All-cause and cause-specific mortality among men released from state prison, 1980–2005. Am J Public Health 98(12):2278–2284

Seal KH, Kral AH, Gee L, Moore LD, Bluthenthal RN, Lorvick J et al (2001) Predictors and prevention of nonfatal overdose among street-recruited injection heroin users in the San Francisco Bay Area, 1998–1999. Am J Public Health 91(11):1842–1846

Seaman SR, Brettle RP, Gore SM (1998) Mortality from overdose among injecting drug users recently released from prison: database linkage study. Br Med J 316:426–428

Siegel S (2001) Pavlovian conditioning and drug overdose: when tolerance fails. Addict Res Theory 9(5):503–513

Singleton N, Pendry E, Taylor C, Farrell M, Marsden J (2003) Drug-related mortality among newly released offenders. UK Home Office, London

Stewart L, Henderson C, Hobbs MST, Ridout SC, Knuiman MW (2004) Risk of death in prisoners after release from jail. Aust N Z J Public Health 28(1):32–36

van Dooren K, Forsyth S, Kinner SA (in press). Young people have heightened risk of death in the year following release from adult prison. Australian and New Zealand Journal of Public Health

Warner-Smith M, Darke S, Lynskey M, Hall W (2001) Heroin overdose: causes and consequences. Addiction 96(8):1113–1125

WHO (2007) Health in prisons: a WHO guide to the essentials in prison health. World Health Organisation, Geneva

WHO (2010) Prevention of acute drug-related mortality in prison populations during the immediate post-release period. World Health Organisation, Geneva

Motivation

► Offender Change in Treatment

Motor Vehicle Theft

Marissa P. Levy
School of Social and Behavioral Sciences,
Richard Stockton College of New Jersey,
Galloway, NJ, USA

Synonyms

Auto theft; Car theft; Stolen auto; Stolen vehicle

Overview

According to the National Crime Victimization Survey (NCVS), motor vehicle theft has been on a decline in the United States during the last decade. Since 2001, the number of motor vehicle theft victimizations has decreased from just over one million to 606,990 per year. This is nearly a 40 % change in the last 10 years, with an average annual decrease of about 3.5 % (Truman 2011). The Uniform Crime Report (US Department of Justice 2011a) indicates a similar decline, with the motor vehicle theft rate decreasing from 430.5 per 100,000 people in 2001 to 258.8 per 100,000 people in 2009. Europe joins the United States in a reprieve from the inconvenience associated with higher rates of motor vehicle theft. Europe has also seen a 5.7 % decrease in motor vehicle theft between 2000 and 2007 (European Commission 2010).

Although these rates are representative of both regions as a whole, certain areas of the United States are still experiencing a disproportionate amount of stolen vehicles. In the United States, large cities are experiencing higher rates of motor vehicle theft (286 per 100,000 people), while smaller cities and rural areas experience much lower rates (142.5 and 105.3 per 100,000 people, respectively). The western and southern coasts of the United States experience disproportionately higher rates of motor vehicle theft. Cities in California close to the United States-Mexican border report that motor vehicle thefts account for 44 % of the total crime index. Cities experiencing such a high level of motor vehicle theft victimization are certainly inconvenienced by this type of crime.

Fundamentals of Motor Vehicle Theft

Definition of Motor Vehicle Theft

The Bureau of Justice Statistics (US Department of Justice 2011b) defines motor vehicle theft as the stealing or unauthorized taking of a motor vehicle. *Completed thefts* are those in which the vehicle is successfully taken by an unauthorized person. *Attempted thefts* are those in which the vehicle is unsuccessfully taken by an unauthorized person.

Sometimes motor vehicle theft and theft from motor vehicles are misunderstood. *Motor vehicle thefts* take place when a motor vehicle is removed from the location where it was left by the owner or a person who was authorized by the owner to drive the vehicle. Motor vehicle thefts are sometimes reported by the owner of the motor vehicle or a representative of the owner. Sometimes the reports are truly accurate in that the vehicle has been stolen by a person both *unauthorized* and *unknown* to the owner. Other times vehicles are reported stolen, but it is later determined that the owner's child or relative was the thief. In this case, the thief was *unauthorized* but not unknown.

Thefts from motor vehicles are those thefts which occur when items are taken from on or within the vehicle. Items stolen from within the vehicle can include items that are stolen from inside a locked or unlocked vehicle. These items can have monetary value, personal value, or no value at all. Items that may be stolen off of the vehicle include things such as hubcaps, pendants, magnetic signs, or other memorabilia on the vehicle for either functional or decorative purposes.

The number of both *motor vehicle thefts* and *thefts from motor vehicles* varies greatly dependent upon the type of vehicle and the location of the vehicle when it is parked. In the United States, foreign motor vehicles suffer more thefts than those made domestically. Location of the vehicle is also an important consideration. Location variables include aspects of physical space (garage, lot, driveway, curb) and city structure (urban, suburban, rural) to become part of the risk calculation.

Thefts of and from motor vehicles typically occur as a result of two types of thieves: opportunistic or purposive. *Opportunistic thieves* are more likely to strike when the vehicle is unlocked or able to be unlocked easily (e.g., through an open window or sunroof). *Purposive thieves* are more likely to steal from a vehicle or the vehicle itself, whether locked or unlocked, if they want a specific vehicle or something specific from the vehicle. These thieves are not deterred by locked windows or doors.

History of Motor Vehicle Theft

In 1919, the National Motor Vehicle Theft Act, or Dyer Act (18 USCS § 2311 et seq.), became federal law. The purpose of the Dyer Act was to obstruct and control vehicles stolen by organized thieves. According to the Dyer Act, the following must be established beyond a reasonable doubt: the vehicle must have been stolen, the stolen vehicle operator must be a person who knows the vehicle was stolen, and the stolen vehicle must have been transported interstate. The punishment for committing these acts is a fine, up to 10 years imprisonment, or both. According to the Dyer Act, a motor vehicle is defined as including "an automobile, automobile truck, automobile wagon, motor cycle, or any other self-propelled vehicle not designated for running on rails."

In 1931 the US Supreme Court debated the definition of a *motor vehicle* as described by the Dyer Act. The question before the court was whether or not the definition outlined above applies to aircraft. In McBoyle v. United States, 283 U.S. 25, the Court found that the language should be more specific if it wanted to include aircraft. Since the word "aircraft" was not mentioned in the language of the law, aircrafts were not to be considered motor vehicles. More recently the term *motor vehicle* is thought to include automobiles (vans, cars, trucks) as well motorcycles, scooters, and mopeds.

In 1984 Congress enacted the Motor Vehicle Theft Law Enforcement Act in response to the growing problem of motor vehicle theft in the 1980s. This Act led to the Federal Motor Vehicle Theft Prevention Standard, which mandated manufacturers to etch Vehicle Identification Numbers (VIN) into various parts of the motor vehicle. Despite the recommendations in 1984, motor vehicle theft was still on the rise in the 1990s. In response, Congress enacted the Anti-Car Theft Act of 1992. The 1992 Act added federal penalties to motor vehicle theft, provided grant money to law enforcement to combat motor vehicle theft, and improved vehicle titles, registration, and salvage requirements.

Since the mid-1990s, motor vehicle theft has been decreasing nationwide. This translated into a significant shift of the top 10 cities for motor vehicle theft in the last decade. In 2000, the National Insurance Crime Bureau (NICB) indicated that cities with major port facilities or those near international borders recorded the greatest number of thefts: Los Angeles, New York, and Philadelphia occupied the top three slots (National Insurance Crime Bureau 2000). In 2007, Las Vegas was ranked #1 (NICB 2008). In 2010, however, eight of the worst 10 cities for motor vehicle theft were in California with Fresno, Modesto, and Bakersfield in the top three (NICB 2011). This indicates a recent shift in motor vehicle thefts from coastal regions to inland areas and also demonstrates that California has been hit hardest by motor vehicle thieves.

Clearance Rates and the Impact of Technology

Historically, clearance rates have shown that it is difficult for police to track and arrest offenders. Modern cities provide outsiders with easier access to target vehicles via better road networks, more anonymity, and increased awareness of city routes and pathways during everyday travel. Until about 10 years ago, the battle between the owner and thief was that of one-upmanship. An advance in technology would provide the owner with the option of buying anti-theft devices, and the thief would find a way around such technology. One example of such a product is the steering wheel lock. The vehicle owner could purchase a steering wheel lock that would prevent the steering wheel from being turned until the lock was removed. Since the steering wheel locks were made of durable metal, they were difficult to cut. The thief could, however, cut the steering wheel so that the device would slide off. Once cut, the steering wheel still functions, at least well enough to drive the vehicle.

In the past 10 years, we have seen an increase in technology that has been more difficult for offenders to circumvent. Recent advances in technology, including ignition kill devices, smart keys, and GPS, make vehicles more difficult to steal. Ignition kill devices work to prevent motor vehicle theft because a switch is inserted into the ignition circuit. The offender would need to know where the switch is located in order to flip it and steal the vehicle. Smart keys contain a coded computer chip or radio frequency. If the smart key is not located inside the vehicle, the vehicle will not start. Both ignition kill devices and smart keys make the vehicle hard to steal because it is difficult or impossible to start the vehicle without them working properly.

Global Positioning Systems (GPS) locator devices make the vehicle easier to find once it has been stolen. The deterrent effect with GPS is that once the police or GPS service providers are notified that the vehicle was stolen, the vehicle can be easily tracked and recovered. This may be one reason why, overall, the number of motor vehicle thefts is on the decline.

Research on Motor Vehicle Thefts

Current research suggests that motor vehicle thieves find certain types of locations more suitable than others. In general thieves are concerned with not getting caught, so all things being equal, they will target areas that provide adequate cover (including poor lighting, low visibility, and obstructed view), a selection of targets, and adequate time needed for the theft to take place. Parking facilities, transportation hubs, hotel and motel parking lots, shopping centers, and motor vehicle repair and parking lots provide unique opportunities for motor vehicle thieves.

Parking Facilities

The physical structure of a parking facility can greatly impact the ability of residents and store owners to provide guardianship and informal surveillance over the vehicles. Several criminologists (Jacobs 1961; Newman 1972; Clarke and Harris 1992; Tilley 1993; Poyner 1997) acknowledge that the way the environment is built impacts perceptions of safety and can send cues to the criminal that one area is more suitable for the commission of a crime than another. Parking lots that are properly built can maintain the positive aspects of natural surveillance and encourage thieves to look elsewhere. Parking lots that are enclosed or located too far from residences or businesses reduce levels of natural surveillance and provide thieves with targets that are ripe for theft. Poor surveillance combined with a condensed number of vehicles in close proximity greatly increases the chance of motor vehicle theft. Research findings support the idea that parking lots should be created with increased surveillance both inside and outside the lots.

Current research has found that the layout of parking lots, level of lighting, presence of closed-circuit television (CCTV), and controlled vehicular access are also important factors utilized to enhance surveillance and on-site security. Lots with on-site security and personnel to monitor exits have better surveillance and report fewer thefts. Webb et al. (1992) found that parking lots, referred to as "surface car parks" (p. 15), have better attendant visibility; attendants can be seen from the entire lot and, also, can see all of the vehicles in the lot. Lots open in the evening and those with "pay and display" methods of payment are more likely to be victimized. Facilities that park the vehicles in the block method and keep the keys for owners tended to have fewer motor vehicle thefts. The block method prevents any vehicle from being stolen without moving at least one other vehicle. Parking in such a manner is one technique used to reduce motor vehicle theft in parking lots, decks, and garages.

Transportation Hubs

Transportation hubs provide a suitable environment for motor vehicle theft. This is related to both the concentration of people and the criminal's expectation that the owner will be away from the vehicle for a long period of time while watching a movie, shopping, working, or taking public transportation to another destination.

Transportation hubs impact the risk for motor vehicle theft in three ways: by increasing the number of offenders who can easily access an area, by decreasing the number of capable guardians who can provide surveillance over these areas, and by increasing the number of suitable targets. The transportation hubs, themselves, provide easy access for *offenders* to travel to the location from which the motor vehicle will be stolen. For an offender to steal a motor vehicle, he first needs to arrive on the scene. If he takes his own vehicle, it will be left at the scene while he is committing the crime. Therefore, parking facilities that are within walking distance from transportation hubs are attractive to motor vehicle thieves. They provide transportation to the crime scene.

Parking facilities near transportation hubs such as airports, train stations, and bus terminals may have fewer opportunities for *guardians* to provide informal surveillance since these large parking areas are not usually in city centers or residential areas where citizens would recognize the vehicles and their owners. Those who may be traveling through the parking facilities do not know the vehicle owner and could mistake a thief for the owner. Pedestrians in these areas may not observe with the same critical eye as they would if the vehicle were parked on a street in their own neighborhoods.

The number of *targets* in large parking facilities provides increased opportunities for motor vehicle thieves. A large quantity of vehicles provides greater selection and an increased chance that some vehicles will be unlocked or easy to access. The close proximity of vehicles in the parking facility, such as a deck or garage, makes seeing between vehicles more difficult. This obstructed view provides increased cover for offenders and provides thieves with more time to open the door and access the vehicle without suspicion. Therefore, more vehicles provide more opportunities, and these opportunities provide a greater chance of success.

Overall, residential and commercial locations near transit stops and hubs experience an increased amount of crime as compared to similar establishments not located near transportation stops. Huang et al. (1998) found that hotels with easy access to public transportation tend to have a greater amount of motor vehicle thefts than hotels without transportation stops. Brantingham and Brantingham (1999) and Levine and Wachs (1986) identified transportation hubs as types of environmental gatherings that present a greater risk for victimization. In addition to crime within the transportation hub, the vehicle parked in a lot, without residents or interested consumers to play the role of capable guardian, is at great risk.

Accommodations

Motor vehicles create an increasing amount of opportunity. Most people who travel long distances in the United States travel by vehicle. Travelers going long distances stop at hotels or motels for evening accommodations. Offenders anticipate that vehicle owners will not return to their vehicles until at least the early morning hours, providing offenders with hours of uninterrupted time to steal vehicles or items from them.

Hotels and motels create an increased risk for all types of crime, especially motor vehicle theft, since they provide a great number of opportunities for motor vehicle thieves. Guests who are staying at hotels or motels are often transient individuals, if not all of the time, certainly at the time in which they are staying at the hotel or motel. Often these individuals drive out of state vehicles. When out of state vehicles are stolen, local police have a more difficult time tracking down the owners and maintaining contact. Hotel guests and other travelers also tend to have more items in their possession since they are going to be away from home for longer periods (Huang et al. 1998). Travelers do not see the harm in leaving items in the vehicle instead of carrying them in and out of the hotel room. When high-priced items such as cameras, video equipment, computers, tablets, or other portable devices are visible, they increase the chance that the vehicle will be targeted for both motor vehicle theft and theft from the motor vehicle.

Shopping Centers

Similar to transportation hubs, shopping centers provide thieves with a large number of targets in a condensed area. Thieves also have easy access to the vehicles and are fairly certain that individuals will be away from their vehicles for a significant amount of time once they park and enter the shopping center. Engstad (1975) believed that parking lots provide opportunities for theft from vehicles and vehicle theft because massive parking lots with thousands of spots present increased opportunities for motor vehicle victimization. Like transportation hub parking facilities, the legitimate users do not know each other and are unable to spot thieves approaching a vehicle.

Supermarket parking lots also provide additional incentives for motor vehicle thieves. Supermarkets are often located off major roadways or adjacent to highways. They can be considered ideal targets since they are open late hours and have access to roadways which can make escape fast and detection impossible. By the time the motor vehicle is reported stolen, the thief can be in another county or state.

Motor Vehicle Repair Shops and Auto Parts

Year after year, vehicles that are reported stolen most often are older and foreign. In 2010, according to the National Crime Insurance Bureau, the top five stolen vehicles were the 1994 Honda Accord, 1995 Honda Civic, 1991 Toyota Camry, 1999 Chevrolet Pickup (full size), and the 1997 Ford F150 Series/Pickup. The top three stolen vehicles are older Hondas and Toyotas, which have made the top of the list for a decade. These motor vehicles are stolen for their parts; vehicle parts for many models are worth much more than the vehicle's street value.

Research in this area has supported these popular beliefs. Gant and Grabosky (2002) interviewed a snowball sample of motor vehicle thieves. Thieves in this sample indicated that professionals in search of parts to repair vehicles actually steal vehicles (especially older vehicles) and dismantle the vehicle for parts to sell them to motor vehicle repair businesses. Thieves steal for a number of reasons such as profit, transportation, or recreation. Those who steal vehicles to profit are thought to be more “professional offenders” and do not steal for reasons such as quick transportation and/or recreation. These professional offenders are often driven by the demand for certain makes and models of motor vehicles and their parts. Professional offenders may spend more time searching than offenders who are simply looking to joyride or find transportation home. La Vigne, Fleury and Szakas (2000) studied motor vehicle thieves and their travel patterns to chop shop locations. According to research by La Vigne et al., police or researchers can estimate the offender's journey to crime by considering the average distance from the location of the theft to the location of the chop shop. Lu (2003) also found that travel distances between the location of the stolen vehicle and vehicle dump sites were short. Research by both La Vigne et al. (2000) and Lu (2003) indicate that the search for thieves can be narrowed if the locations of the theft and dump site are known.

Current Issues and Controversies

The current debate over motor vehicle theft presents three key environmental factors that influence the likelihood of a motor vehicle to be stolen: location, lighting, and security. Surely motor vehicles that are parked in locations that have poor environmental characteristics, poor lighting, and poor security are more likely to be victimized, but current research has found that each characteristic varies in its level of contribution to victimization.

Locations

The location of the potential motor vehicle target is an important factor that offenders consider before they offend. The characteristics of the motor vehicle theft location help the offenders determine if it is a "good" target to select. Offenders will use landscape and design features to determine if their crime will be concealed and if the layout of the area will impede their escape after the vehicle is stolen (Keister 2007). Jeffery (1971) noted that there is a decision model that offenders use during the actual commission of a criminal act. "The decision to commit a crime involves the past experience of the subject, the immediate opportunity for a crime, plus the chances of apprehension or injury" (Jeffery 1971, p. 251). The last part of the decision, the likelihood of apprehension, is discussed with regard to the layout of the location to be targeted. All other things being equal, offenders will choose locations that provide an easy get away and enhance the chance of escaping apprehension. Nichols (1980) suggests that after offenders decide to commit a crime they engage in a search pattern of potential sites. This is conducted by using a "mental map" (Nichols 1980, p. 156) of known locations to help offenders evaluate their options. Lu (2006) also addressed the importance of escape routes for motor vehicle thieves. Major roads and those connected to major roads had disproportionately more motor vehicle thefts than less traveled roadways.

The safest place to park a vehicle is a residential garage. The best places to park, in order from most safe to least safe, are residential garage, residential driveway, residential street parking, then parking lots/garages/decks. Clarke and Mayhew found that "parking in a domestic garage at night is safer by a factor of 20 than parking in a driveway or other private place, and safer by a factor of 50 than parking in a street near home" (1994, p. 91). This indicates an increased risk for victimization when a person parks in the street instead of in a personal garage located at the residence. Street parking provides reduced surveillance and, in high density locations, a similarly plentiful group of targets as do parking garages.

Many cities prefer garage parking since it takes up less ground space and is more cost-effective for those who own the facilities. However, due to decreased surveillance opportunities, parking garages are less safe than parking lots. Parking garages that are either partially or fully enclosed and elevated above ground offer less natural surveillance than parking lots. This problem is enhanced by sloping ramps and multiple floors which further cut surveillance from other individuals inside the garages.

Lighting

Poor lighting can provide cover for offenders. Just as offenders use hedges and walls to prevent their detection when breaking into a home or business, dark alleys and dimly lit streets hinder residents and other watchers from their important involvement in informal surveillance. It is critical that lighting is both provided in adequate levels and is aimed in the proper direction. Historically, lighting has been a problem with regard to many types of crime and public safety issues. Government organizations (such as police) and private organizations (such as residents' associations) must work together to increase lighting.

Lighting is most important in parking lots and garages (Smith 1996; Webb et al. 1992) when garages are either partially or fully enclosed, limiting natural light from entering and prohibiting natural surveillance from the street (Tseng et al. 2004; Poyner 1997). Illuminance (the intensity of light) and vertical illuminance are important traits that must be considered when installing lighting. In addition to the intensity of light, placing white stain on the ceiling can reflect light and increase uniformity in the parking garages (La Vigne 1997; Painter 1994). Garages that can minimize ramps and utilize flat surfaces for parking areas will maximize the positive effects of light and reflected light and enhance natural surveillance from outside of the garage.

The presence and quality of lighting is an important site-level factor considered before the commission of a motor vehicle theft (Levy 2009; Levy and Tartaro 2010). Both the amount of lighting that is properly working and the intensity of that lighting are important in the overall lighting of streets at night. Street lighting is just one part of the lighting requirement of specific sites. Most businesses and residential areas obtain evening lighting from the streetlights and lights located on the store or residential property. Both the number of lights on the street and the property and the intensity of the light in these locations, are needed to determine if the lighting is adequate.

Farrington and Welsh (2002) conducted a meta-analysis of studies where street lighting was measured with regard to its impact on crime. Studies eligible for inclusion in the meta-analysis had to meet the following criteria: the study utilized improved street lighting as the intervention, the outcome measure was crime, the methodology was sound with both pre- and post-measures of lighting in experimental and control groups, and the experimental and control had at least 20 crimes in each group prior to the intervention. Farrington and Welsh (2002, 2004) concluded that improving street lighting decreased crime by roughly 20 %. This and other recent literature indicate that increasing lighting is an economical and effective way to reduce crime.

Security and Technology

Almost every kind of location employs some method of security in order to protect goods. For locations such as parking lots, controlling entrances and exits is an important security feature (Hunter and Jeffery, 1992). Webb et al. (1992) noted that the layout of the parking structure is a factor considered in motor vehicle theft. Parking structures with manned exits have the lowest risk for both motor vehicle theft and theft from motor vehicles. Manned exits not only provide increased surveillance, but bars or mechanical arms on the exits also provide a physical barrier to prevent vehicles from leaving the structures.

Owners of motor vehicles have their own set of precautions and security devices. Alarms are frequently installed by the manufacturer or are installed after market if the vehicle was not sold with an alarm. Research on motor vehicle alarms is mixed. Light et al. (1993) found that 35 % of motor vehicle thieves interviewed indicated that they would be deterred by alarms, 18 % said it would depend on the make of the alarm, while 9 % said it depended on the model of the alarm (p. 50). Conversely, Fleming, Brantingham, and Brantingham (1994) found that nearly 75 % of offenders would be deterred by an alarm. Many thieves also indicated that audio equipment "pull outs," radio faces that are designed to be detached and taken with the driver, are usually left in the vehicle and therefore are not a major deterrent to thieves. Research on steering column locks is similarly mixed with steering column locks working better on some makes and models of motor vehicles and not as well on others.

More recently, LoJack and OnStar have shown to be very effective tools used to locate vehicles after they are stolen. Arrest rates for vehicles with LoJack and similar GPS locator systems are three times greater than for vehicles without the devices. The LoJack company states that 90 % of vehicles with LoJack are recovered (LoJack 2011). GPS devices have been the most effective tool available to date, but often do not prevent vehicles from being stolen; such devices only aid in motor vehicle recovery.

The presence of security hardware, especially in the case of motor vehicle theft, is only effective if the hardware is installed and maintained properly. Motor vehicle owners who leave doors unlocked or leave the radio face in place are not going to deter a thief from stealing the vehicle or even the radio. Similarly, parking lots with gated access and egress are most effective in preventing motor vehicle theft, but only if these gates are working properly all of the time. Technology can play a major role in decreasing motor vehicle theft, but it can only work if used properly. And, sometimes, despite proper use, crime prevention techniques may not thwart the skilled offender.

Repercussions of Motor Vehicle Theft on Society

When compared to criminal homicide, forcible rape, and robbery, some of the other Part I offenses recorded in the *Uniform Crime Report*, motor vehicle theft typically has less of a direct personal impact on the victim. Despite this, victims of motor vehicle theft report crippling effects associated with the loss of a motor vehicle. Some of these effects include problems with insurance coverage or response times, the emotional impact of losing one's possessions, and the increase in policy premiums for all motorists living in areas that have high motor vehicle theft rates.

When a motor vehicle theft occurs, the owner is left to report the loss to the police department and to file an insurance claim. Many insurance carriers do not provide replacement vehicles while the claim is being settled, so – even if temporary – the owner does not have a motor vehicle. If the vehicle is not recovered, the owner may be compensated for the current value of the motor vehicle. In many cases this does not cover the full amount owed on the vehicle nor allow the owner to purchase a replacement vehicle of similar style or value. If the vehicle is recovered, it may have sustained serious damage during the time that it was missing. What is returned may not be the same as what was stolen. This can create more complicated interactions between the owner and the insurance company, often resulting in little or no compensation for the loss.

In addition to the loss of the vehicle, owners may suffer emotional distress in a variety of ways. The owners may have lost items of personal significance or sentiment which were in the vehicle when it was stolen. These items may have been irreplaceable but not of great financial value. Other items that were in the vehicle may not be replaced unless receipts and proof of purchase are provided. Owners may also suffer a fear of future victimization. They may fear that the replacement vehicle will also be stolen and that they will have suffer greater insurance premiums if future motor vehicles are stolen.

Finally, motorists living in areas with high rates of motor vehicle theft will have similarly high motor vehicle insurance premiums. As thefts continue to increase, so too will the premiums. Motor vehicle thefts impact the entire city because insurance companies base premiums on the location of the policy holder's home, the distance the policy holder travels, and the locations most frequented by the policy holder (work or school locations), in addition to the rate of motor vehicle theft in those areas. Motor vehicle insurance premiums are calculated based on all of these factors and are higher in areas with higher motor vehicle theft rates.

The Complicated Nature of Repeat Victimization

Motor vehicle theft presents researchers with a challenge when defining the role of "victim" in the motor vehicle theft event. Technically, there are three ways in which a victim of motor vehicle theft can be defined: address, vehicle, and owner.

An address can be a repeat victim. Motor vehicles parked in the same lot, same driveway, or same street segment can suffer multiple motor vehicle victimizations. In this example the address or location has been victimized. Criminologists can study the characteristics of these addresses or locations to better understand why the area may be suffering repeat victimization. Some potential causes of repeat victimization in these settings are reduced visibility by residents or business owners, little or no lighting in the evening, and poor security devices. Security devices that can be used to protect addresses or locations are gates, fences, security alarms on residential or commercial garages, and surveillance equipment.

A vehicle can also be a repeat victim. The same motor vehicle can be stolen from the same owner, or the same vehicle can be stolen from a different owner or renter. In this situation, the vehicle is likely to have certain features that make it more likely to be stolen. These features can be parts that were installed by the manufacturer such as engines and catalytic converters. Vehicles can also be repeatedly stolen or have parts repeatedly stolen, which were added after market. Items added after market that have high theft rates include portable GPS units, entertainment systems, rims, and tires. Security devices that can be used by vehicles in order to prevent repeat victimization are alarms, ignition devices, smart keys, and vehicle locator devices such as LoJack and OnStar.

Finally, an owner can be a repeat victim. The same owner can have many vehicles stolen at one or many points in time. Owners who find that they have many stolen motor vehicles likely have behaviors which put them at increased risk of victimization. Owners may continually purchase foreign-made vehicles which are more likely to be stolen. They may live in or routinely travel to an area that has a high motor vehicle theft rate. Finally, they may fail to lock vehicle doors, set alarms, or leave valuable items in plain sight. Any of these behaviors may put vehicle owners at an increased risk for repeat motor vehicle victimization.

Looking Forward

It appears that motor vehicle theft victimization is following a downward trend. However, that does not mean that motor vehicle theft is no longer a problem worthy of study. Over 737,000 motor vehicles were stolen in the United States in 2010. With each motor vehicle that is stolen, the costs to the owner, insurance company, and all motorists increase. With that said, it seems that advances in technology have created a significant obstacle for motor vehicle thieves to overcome. It may be that the risk has now met or exceeded the reward.

Related Entries

- ▶ Bicycle Theft
- ▶ Disorder: Observational and Perceptual Measures
- ▶ Fencing/Receiving Stolen Goods
- ▶ Situational Approaches to Terrorism

Recommended Reading and References

Brantingham PL, Brantingham PJ (1999) A theoretical model of crime hot spot generation. Stud Crime Crime Prev 8(1):7–26

Clarke RV, Harris PM (1992) Auto theft and its prevention. Crime and Justice 16:1–54

Clarke RV, Mayhew P (1994) Parking patterns and car theft risks: policy relevant findings from the British Crime Survey. In: Clarke RV (ed) Crime prevention studies, vol 4. Criminal Justice Press, Monsey

Engstad PA (1975) Environmental opportunities and the ecology of crime. In: Silverman RA, Teevan JJ (eds) Crime in Canadian Society. Butterworths, Toronto

European Commission (2010) Crime statistics. Retrieved from Eurostat: http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Crime_statistics

Farrington DP, Welsh BC (2002). Effects of improved street lighting on crime: a systematic review. Home Office Research Study 251: Home Office Research, Development and Statistics Directorate

Farrington DP, Welsh BC (2004) Measuring the effects of improved street lighting on crime. Br J Criminol 44:448–467

Fleming Z, Brantingham PL, Brantingham PJ (1994) Exploring auto theft in British Columbia. In: Clarke RV (ed) Crime prevention studies, vol 3. Criminal Justice Press, Monsey, pp 47–90

Gant F, Grabosky P (2002) The stolen vehicle parks market. In Crime facts info, No. 16. Australian Institute of Criminology

Huang WS, Kwag M, Streib G (1998) Exploring the relationships between hotel characteristics and crime. FIU Hosp Rev 16(1):81–93

Hunter RD, Jeffery CR (1992) Preventing convenience store robbery through environmental design. In: Clarke RV (ed) Situational crime prevention: successful case studies. Harrow and Heston, Albany

Jacobs J (1961) The death and life of great American cities: the failure of town planning. Peregrine Books, London

Jeffery CR (1971) Crime prevention through environmental design. Sage, Beverly Hills

Keister T (2007) Thefts of and from cars on residential streets and driveways, vol 46, Problem-oriented guides for police, problem-specific guides series. US Department of Justice, Office of Community Oriented Policing Services, Washington, DC

LaVigne NG (1997) Visibility and vigilance: Metro's situational approach to preventing subway crime. U.S. Department of Justice, Office of Justice Programs, National Institute of Justice, Research in Brief, Washington, DC

LaVigne NG, Fleury JK, Szakas J (2000) Auto theft and detecting chop shop locations. In: Turnbull LS, Hendrix EH, Dent BD (eds) Atlas of crime: mapping the criminal landscape. Oryx Press, Phoenix

Levine N, Wachs M (1986) Bus crime in Los Angeles: I–measuring the incidence. Transp Res A 20(4):273–284

Levy MP (2009) Opportunity, environmental characteristics and crime: an analysis of auto theft patterns. LFB Scholarly, El Paso

Levy M, Tartaro C (2010) Repeat victimization: a study of auto theft in Atlantic City using the WALLS variables to measure environmental indicators. Crim Justice Policy Rev 21:296–318

Light R, Nee C, Ingham H (1993) Car theft: the offender's perspective. HMSO Books, London

LoJack (2011) Theft facts. Retrieved from http://www.lojack.com/why/pages/theft-facts.aspx

Lu Y (2003) Getting away with the stolen vehicle: an investigation of journey-after-crime. Prof Geogr 55(4):422–433

Lu Y (2006) Spatial choice of auto thefts in an urban environment. Secur J 10:143–166

Maxfield MG, Clarke RV (eds) (2004) Understanding and preventing car theft. Crime prevention studies. Vol. 17. Lynn Rienner, Boulder

National Insurance Crime Bureau (2000) Vehicle theft booming in port and border communities; people should be more vigilant than ever, Warns NICB. Retrieved from http://www.nicb.org/services

National Insurance Crime Bureau (2008) Vehicle theft brochure. Retrieved from https://www.nicb.org

National Insurance Crime Bureau (2011) Hot wheels, 2010. Retrieved from https://www.nicb.org/newsroom/news-releases/hot-wheels-2010

Newman O (1972) Defensible space. Macmillan, New York

Nichols WW (1980) Mental maps, social characteristics, and criminal mobility. In: Georges-Abeyie DE, Harries K (eds) Crime: a spatial perspective. Columbia University Press, New York

Painter KA (1994) The impact of street lighting on crime, fear, and pedestrian street use. Secur J 5:116–124

Poyner B (1997) Situational crime prevention in two parking facilities. In: Clarke RV (ed) Situational crime prevention: successful case studies. Willow Tree Press, Monsey

Smith MS (1996) Crime prevention through environmental design in parking facilities. US Department of Justice, Office of Justice Programs, national Institute of Justice. Research in Brief, Washington, DC

Tilley N (1993) The prevention of crime against small businesses: the safer cities experience, vol 45, Police research group: crime prevention unit series. Home Office Police Department, London

Truman JL (2011) National crime victimization survey: criminal victimization, 2010. (Report No. NCJ 235508). Retrieved from Bureau of Justice Statistics http://bjs.gov/content/pub/pdf/cv10.pdf

Tseng A, Duane J, Hadipriono F (2004) Performance of campus parking garages in preventing crime. J Perform Constr Facil 18(1):21–28

US Department of Justice (2010) Crime in the United States, 2009. Retrieved from the Federal Bureau of Investigation http://www2.fbi.gov/ucr/cius2009/data/table_01.html

US Department of Justice (2011a) Crime in the United States, 2010. Retrieved from the Federal Bureau of Investigation http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2010/crime-in-the-u.s.-2010/property-crime

US Department of Justice (2011b) Property crime. Retrieved from the Bureau of Justice Statistics http://bjs.ojp.usdoj.gov/index.cfm?ty=tp&tid=32

Webb B, Brown B, Bennett K (1992) Preventing car crime in car parks, vol 34, Crime prevention unit series. Home Office Police Department, London

Motorcycle Clubs

► Motorcycle Clubs or Gangs?

Motorcycle Clubs or Gangs?

Tom Barker
Department of Criminal Justice, College of Justice and Safety, Eastern Kentucky University, Richmond, KY, USA

Synonyms

1%ers; Biker clubs/gangs; Motorcycle clubs; Motorcycle gangs; One percent clubs/gangs; Outlaw motorcycle clubs/gangs

Overview

Motorcycle clubs, that is, voluntary social organizations based on a common love of motorcycles and riding together, formed in the early 1900s. However, by the 1930s groups of displaced young men riding motorcycle grouped together in what were nascent gangs. These nascent gangs evolved into outlaw bikers and then into a subculture of bikers known as One Percenters aka Outlaw Motorcycle Gangs. The One Percenters evolved into motorcycle gangs more into crime for profit than a common love of motorcycles and riding together. Today, these Outlaw Motorcycle Gangs are national and international criminal organizations.

Are motorcycle clubs really clubs, that is, voluntary social organization built around the love of motorcycles, or criminal gangs whose members happen to ride motorcycles? There is no easy answer: some are and some are not. Some well-known criminal gangs, such as the Hells Angels Motorcycle Club (HAMC) and other Outlaw Motorcycle Gangs (OMGs) aka one percent clubs (1 %ers), vehemently argue that they are clubs and not gangs. Their behavior and the criminal history of their members do not support their argument. Therefore, we must examine the differences between conventional motorcycle clubs and criminal motorcycle clubs and how the 1 % outlaw motorcycle clubs/gangs fit in the motorcycle club grouping.

Motorcycle Clubs: Conventional and Criminal

Since the first American "motor-driven cycle," the Indian, was built in 1901 Springfield, Massachusetts, motorcycle riders have formed clubs based on their common interest in riding the iron machines (Hayes 2010). The first clubs were composed of foolhardy men racing these dangerous and unwieldy machines, then; in 1903 the fledgling American motorcycle scene changed dramatically. In 1903, the Harley-Davidson Company in Milwaukee, Wisconsin began building a motorized bicycle that changed the machines, the way the machines were used, and brought about a change in the riders and how they grouped together. Harleys began dominating sport motorcycle racing and then in the 1920s Harley's became more a means of transportation and pleasure rather than racing (Davidson 2002). At the time, individual travelers had a limited range of transportation options; they could ride a horse or a horse drawn buggy, ride a bicycle, or walk. The adventurous of that day who desired to travel between cities or farther had to take the train. There were few highways and coast-to-coast was a challenge. Motorcycles made it possible to ride within cities, between cities, and across the country for transportation or pleasure. The sidecar also allowed for the entire family to share the motorcycle experience.

Motorcycle riders changed as women and other family members began riding motorcycles. The clubs they formed also changed. In 1924, the American Motorcycle Association (AMA) was formed and provided the main motorcycling racing sanctioning body and soon began promoting what was known as Gypsy Tours that were "statewide, good-clean-fun, jamboree-picnic soirees, which proved to be the ground-seed for events like the Sturgis Rally (Hayes 2010, 43)." The AMA would play a huge role in the definition of outlaw motorcycle clubs. Following the formation of the AMA, groups of motorcycle riders organized clubs that are broadly classified as either conventional or outlaw clubs (Barker 2007). By the 1930s, the AMA had chartered 300 conventional motorcycle clubs (Fuglsang 1997). These clubs represented the "responsible" motorcycle riders and often had strict dress codes. Clubs that obtained a charter from the AMA were considered legal clubs; the others were classified as "outlaws" (Wolf 1999). The "outlaw" label in this case meant non-AMA, not criminal. However, the deviant label had consequences: outlaw clubs were not allowed to participate in AMA-sanctioned events.

The conventional clubs began a tradition of sponsoring mixers, charity events, and hill-climbing contests and promoting responsible motorcycling as a family activity that still exists. Historically, the members of conventional clubs came from different social strata and dressed and acted differently from deviant club members.

At about the same time, groups of motorcyclists began to appear in the motorcycle-friendly weather of Southern California in the early 1930s (Barker 2007). Depression stricken "Okies" looking for work and riding motorcycles because they were cheap transportation, appeared on the scene. Although most were loners who rarely stayed in one town for long, some banded together in the squalid Southern California industrial districts and formed loose-knit "outlaw" motorcycle clubs. Riding together in groups, these outsiders worked menial jobs and lived a deviant lifestyle of drinking and rowdy behavior, with some dealing in stolen motorcycle parts and other criminal pursuits (Yates 1999). Harley-Davidson motorcycles were the most popular and abundant bikes at that time, so Harleys became the motorcycles of choice for these groups – for both possession and theft. These outlaw clubs were more gangs than clubs and different in dress and behavior from the AMA clubs. Their appearance and behavior foreshadowed the development of Outlaw Motorcycle Gangs and the one percent biker.

At first, outlaw clubs were not much of a threat to conventional motorcycle clubs or the image of motorcycle riders in general, but this changed in the late 1940s. Following World War II, outlaw clubs began disrupting AMA-sanctioned events and the label "outlaw" motorcycle clubs took on a new meaning and changed the public image of motorcycle riders. "Wino Willie" Forkner has come to be representative of the change and the men who brought it about.

A hard-drinking young motorcyclist "Wino Willie" Forkner joined the 13 Rebels Motorcycle Club, an AMA conventional club, and rode with them for 2 years before the attack on Pearl Harbor. After Pearl Harbor, the dipsomaniac Forkner joined the Army Air Corps and was soon manning a 0.50 caliber machine gun on a B-24 Liberator for the next 30 months (Reynolds 2000). In the summer of 1945, the hell-raising Wino Willie returned to California and rejoined the 13 Rebels. It was not long before Wino Willie's hell-raising behavior, possibly intensified by his wartime experiences, and the conventional behavior of the 13 Rebels clashed.

Forkner, according to his widow, Teri, was like many of the returning veterans "back from the war and letting off steam (Bill Hayes Video Interview: The Original Wild Ones, 2003)." "Letting off steam" for these veterans meant riding their bikes and drinking together. In the summer of 1946, Forkner and his drinking buddy and fellow 13 Rebels' member, Blackie, were "letting off steam" in their favorite watering hole, the All-American Bar, when they decided to show up at 13 Rebels sponsored quarter-mile race in San Diego after an all-night drinking session (Reynolds 2000, 34–36). Thoroughly plastered and frustrated, Wino Willie decided to liven up a boring race. Forkner drove his bike through the wooden gate leading from the parking lot onto the track. The drunken Forkner burst onto the track in a shower of shattered wood and loud applause from the audience who shared his view of the race. Forkner roared down the straightaway and made four laps around the track before losing control and turning his bike over. The subdued whirling dervish was promptly arrested and hauled off to jail. After a weekend in jail, Forkner pleaded guilty to trespassing and being drunk and disorderly. He hitchhiked back to Los Angeles and faced an enraged group of 13 Rebels members. The club members demanded his 13 Rebels sweater and threw him out of the club. The unrepentant Forkner reportedly returned the sweater after defecating on it (Reynolds 2000, 36). The evolution to one percent biker clubs/gangs had begun.

The angry Wino Willie decided to put the 13 Rebels and the raceway fiasco behind him by drinking and sulking at the All-American Bar (Reynolds 2000, 36–38; Barker 2007). Forkner began drinking with three former servicemen. The four inebriated bikers decided to form a new motorcycle club but could not come up with a name. Overhearing the animated discussion, an All-American regular named Walt Porter lifted his head off the bar and called out "Call it the Boozefighters." The four inebriated bikers thought that was a perfect name. During the following weeks, Wino Willie and his fellow Boozefighters recruited 16 other members. The Boozefighters even applied for AMA membership but were turned down by the AMA president, who allegedly said: “No goddamn way I am giving a name like that a charter (Forkner 1987).” By 1947, three charters of the Boozefighters MC had formed in Los Angeles, San Pedro, and San Francisco.

The Boozefighters MC was not the only outlaw club formed by ex-servicemen “letting off steam.” World War II veterans joined outlaw clubs and called themselves the Galloping Gooses, the Pissed Off Bastards of Bloomington, Satan’s Sinners, and the Market Street Commandoes. Conventional clubs used simple and innocent-sounding names like the Road Runners, the Glendale Stokers, and the Side Winders. On occasion, these conventional clubs engaged in reckless street races, but their behavior was harmless compared to the sometimes violence-prone behavior of the new outlaw clubs (Yates 1999). To further distinguish themselves from AMA clubs members who wore ties and racing caps, the outlaw veterans were unkempt in appearance and rowdy in behavior. The stage was set for the creation of a new label – Outlaw Motorcycle Gangs. One event is considered to be the incident that changed the social definition of outlaw clubs: The Hollister Motorcycle Incident/Riot in 1947.

The events, real, imagined, and contrived of the Hollister Incident “riot”, that took place in the small California town of Hollister during the 1947 July 4th weekend led to a staged photograph, which inspired a short story that led to a movie, which influenced a genre of biker movies and media publicity that ultimately constructed a new social definition of outlaw bikers. The West Coast town of Hollister, California was chosen as the site for a gypsy tour and would draw in bikers from across the country. Hollister was [and still is] a big motorcycle town in the 1940s, with 27 bars and 21 gas stations. However, the town had only six police officers. Hollister held its first gypsy tour in 1936 and held regular motorcycle races and hill climbs. The residents were not expecting any trouble at the 1947 rally.

Members of all three Boozefighters MC chapters made plans to attend the motorcycle rally. The Los Angeles members, including Wino Willie, met on Thursday evening at the All-American to prepare for the ride to Hollister. Preparation included drinking at the All-American, riding to Santa Barbara for more drinks, and then riding to San Luis Obispo, where they became too drunk to go on. The Los Angeles Boozefighters “slept it off” in a bus terminal for 3 h and roared off to King City, where they stopped at a liquor store to finish “preparing” for the ride into Hollister. By the time they arrived in Hollister, several thousand motorcyclists were already there and thoroughly “prepared.” It wasn’t long before intoxicated bikers turned the blocked-off main thoroughfares into drag racing strips and stunt-riding exhibitions.

Conventional club members were at the nearby Bolado Racetrack, which was filled to capacity for the scheduled races. The hard core bikers of Wino Willie’s new club, the Boozefighters, and the other hell-raising outlaw clubs were driving the town’s six police officer’s crazy. Soon the jail was filled to capacity with drunken motorcyclists, including Wino Willie. While the commotion was going on, a photographer for the *San Francisco Chronicle* decided to take a picture of a drunken motorcyclist. For some reason, he staged a photograph even though there were plenty of drunken motorcyclists available. The photographer sat a Harley-Davidson motorcycle on top of a pile of beer bottles and persuaded an intoxicated biker in a leather jacket to sit astride the bike. The picture never appeared in the *San Francisco Chronicle*, but it would become one of the most famous pictures in motorcycle history after the Associated Press picked it up and printed it in *Life* magazine.

On Sunday night the Hollister police officers called for help from the California Highway Patrol (CHP). Forty CHP officers arrived at dusk. The CHP lieutenant in charge began moving the compliant bikers toward the end of town. The lieutenant spotted a group of musicians unloading their instruments for a dance and ordered them to set up in the street on a flatbed truck and play for the crowd. The crowd started dancing and continued dancing into the night under the watchful eyes of the CHP. While the drunken bikers danced, the local Chief of Police went to all the bars and had them closed at midnight, 2 h early. The "riot" was over, but the exaggerated and distorted publicity of the incident was just beginning. The media blitz that followed the alleged Hollister "riot" would create a new moral panic and "folk devil" – the outlaw biker. The next motorcycle "riot" would forever solidify outlaw bikers as folk devils.

A 1948 Memorial Day weekend gypsy tour rally in Riverside, California, attended by more than 1,000 drunken motorcyclists recreated the scene that had occurred at Hollister – this time with violent consequences. While they were drag racing and partying, an Air Force officer and his wife, trying to get through, honked at the pack of drunks blocking the street. The mob smashed the car windows, punched the driver, and manhandled his frightened wife (Reynolds 2000, 60). While this was occurring, one of the drunken riders wiped out on his bike, killing his girlfriend, who was riding with him. Boozefighters were in the area but not involved in the incident. However, local newspapers reporting the incident carried pictures of two Boozefighter members drinking beer and sitting on their bikes. The caption under the picture read that they were members of the same group who "started the Hollister Riot the previous year (Reynolds 2000, 60)." The national story coming on the heels of the Hollister "riot" led to a motorcycle short story that created a genre of low-budget exploitation biker films that would cement outlaw bikers as folk devils and a part of modern pop culture.

In 1951, *Harper's Magazine* published a short story by Frank Rooney entitled "Cyclists Raid." The story an exaggerated account of the Hollister incident would have been forgotten except for the movie it inspired – *The Wild One*. The short story's plot is about an unnamed motorcycle gang taking over and terrorizing a small town somewhere on the West Coast. Filmmaker Stanley Kramer read "Cyclist Raid" and decided to make a movie based on the story. He, writer Ben Maddow and film star Marlon Brando spent 3 weeks interviewing ex-Boozefighters before filming started (Stidworthy 2003). The resulting screen adaptation ushered in a new genre of biker movies. The 1954 movie *The Wild One*, starring Marlon Brando became a cult movie and solidified the image of outlaw bikers as folk devils and spread their cultural lifestyle and dress throughout the world. After the film's release, hundreds of American bikers would emulate Brando and costar Lee Marvin in dress and behavior. Biker expert Arthur Veno asserts that motorcyclists across the Western world "saw the Hollywood version of an outlaw motorcycle rebel" and copied their attitudes, clothes, disrespect for society, and the way they treated women (Veno 2002, 30). According to Veno, motorcycle clubs in England, Australia, Canada, South Africa, New Zealand, Germany, Denmark, and Italy mimicked the screen outlaw bikers. The next evolution of Outlaw Motorcycle Gangs comes with the Hells Angels Motorcycle Club.

Hells Angels Motorcycle Club and the One Percenters

After the Hollister incident, the World War II veterans outlaw motorcycle club, the Pissed Off Bastards of Bloomington, had disbanded, partly because of the bad publicity spin-off from the sensationalized Hollister "riot." Reynolds (2000) contends that Arvid Olsen, a former squadron leader with the WWII "Flying Tigers," of China suggested the name "Hell's Angels" for a new motorcycle club that he, Otto Friedl, and several other former Pissed Off Bastards of Bloomington decided to form in 1948 in San Bernardino, California. The new members chose as a logo a grinning skull wearing a pilot's helmet with attached wings. A decade later the name and logo would be adopted by a group of motorcycle-riding young toughs and thugs in Oakland, completing the one percent bikers aka Outlaw Motorcycle Gangs' progression.

During the next decade, Hells Angels Motorcycle Club chapters formed throughout California as nomadic members moved from one city to the next. The 1957 Sacramento Hells Angels chapter was formed out of the Hell Bent for Glory Motorcycle Club started by two teenage toughs, James "Mother" Miles and his younger brother Pat, breaking the connection between World War II veterans and the establishment of outlaw motorcycle clubs. These Hells Angels chapters were more like separate clubs operating autonomously and independent of each other, often not knowing of the others. One man, Ralph "Sonny" Barger is credited with bringing the chapters together and creating the largest outlaw motorcycle club in the world.

Barger grew up in the blue-collar jungle of Oakland, California, and joined the Army in 1955 with an altered birth certificate (Barger & Zimmerman 2000). Less than a year later, the Army discovered his actual age and discharged him. When he returned to Oakland, there were numerous motorcycle clubs, like the Oakland Motorcycle Clubs, that were conventional clubs, and there were also "freewheeling clubs" like the Oakland Panthers. Barger joined the Panthers but soon left because they did not provide enough action.

Barger found a new wild bunch of motorcyclists with whom to ride. The wild group had no name but was searching for a common identifier. One of Barger's riding buddies "wore a modified Air Force-like patch he'd found in Sacramento, a small skull wearing an aviator cap inside a set of wings (Barger & Zimmerman 2000, 30)." The young "outlaws" thought the patch was cool and decided to name their club Nomad Hells Angels after the patch. In April 1957, the club had patches made based on the design (which later became the easily recognizable copyrighted HAMC death's head). Sometime later Barger met another biker wearing the very same Hells Angel patch. This Angel filled Barger in on the history of the club formed in San Bernardino in 1948, including the other chapters and the rules, regulations, and procedures for becoming a Hells Angel chapter. Angels from the SoCal (Southern California) chapter visited the quasi-Hells Angels in Oakland. A series of meeting later and the Oakland Hells Angels Chapter came into being. In 1958, Sonny Barger became president of the Oakland Chapter; he then became President of the National Hells Angels Motorcycle Club and changed the HAMC forever. Although the basic organization was in place when Barger took over as Oakland president, under his leadership and guidance new rules were added pertaining to new members, club officials, and induction of new charters/chapters. With Barger at the helm, the HAMC would expand into the largest and most well-known outlaw motorcycle club in history. They would also become a criminal gang rather than a motorcycle club and spread its criminal tendrils all over the world.

Early outlaw clubs like the Boozefighters were formed by veterans, later outlaw clubs, particularly the one percent clubs, would not be veteran centered and would be in the tradition of the Sacramento Hells Angels chapter formed by young toughs and criminals or those with criminal tendencies.

George Wethern, the Oakland Hells Angels Vice President, describes a historic meeting that was "sort of like the Yalta conference" in 1960, held at the home of Frank Sadilek, the president of the San Francisco Hells Angels chapter (Wethern and Colnett 1978). At the meeting were Hells Angels leaders from across the state and former warring California biker club leaders – of the Gypsy Jokers, Road Rats, Galloping Gooses, Satan's Slaves, the Presidents, and the Mofos. The purpose of the meeting was to discuss police harassment, but discussion soon turned to recent hostile comments by the AMA.

"To draw a distinction between its members and us renegades, the AMA had characterized 99 % of the country's motorcyclists as clean-living folks enjoying pure sport. But it condemned the other one percent as antisocial barbarians who'd be scum riding horses or surfboards too (Wethern and Colnett 1978, 54)." The clubs decided to accept the one percent label (1 %) as a tribute and not as insult and adopt the "1 %" patch to identify themselves as righteous outlaws. Wethern and Sonny Barger were the first to get the "1 %" tattoos. From this day forward Outlaw Motorcycle Gangs would also be known as one percent (1 %) clubs/gangs.

Outlaw Motorcycle Gangs: The One Percenters (1 %ers)

In the 1960s, the formerly hell-raising and nonconforming outlaw motorcycle clubs such as the Hells Angels moved into criminal activity, particularly drug dealing, and became more like gangs (organized for crime as profit) than clubs of motorcycle enthusiasts (Barker 2007). There is no accurate count of the number of 1 % clubs/gangs. However, Outlaw Motorcycle Gangs (1 %er clubs) can be categorized into the Big Five Outlaw Motorcycle Gangs (OMGs), Puppet/Support OMGs, Independent OMGs, and Black/Interracial Gangs.

Big Five Outlaw Motorcycle Gangs

The designation of Big Five Outlaw Motorcycle Gangs is, in many ways, an arbitrary judgment, first made by law enforcement authorities in the form of the Big Four (Hells Angels, Bandidos, Outlaws and Pagans) to designate the criminal potential of the gangs. The Sons of Silence were added later becoming the Big Five Outlaw Motorcycle Gangs – the Hells Angels, Bandidos, Outlaws, Pagans and the Sons of Silence.

Hells Angels – Barger, in his autobiography, admits that he sold drugs and got into a "lot of shit" in what he describes as the "gangster era" of the Hells Angels (Barger & Zimmerman 2000). It appears that the Hells Angels are still in their "gangster era." The US Department of Justice (DOJ 2011) estimates that the HAMC has between 2,000 and 2,500 members in 27 US states and 26 foreign countries. Internationally, the HAMC is a criminal threat on six different continents. The DOJ reports that the OMG is involved in the production, transportation, and distribution of marijuana and methamphetamine. The international criminal gang is also involved in the transportation and distribution of cocaine, hashish, heroin, LSD, ecstasy PCP, and diverted pharmaceuticals.

Bandidos MC – As a testament to the criminal nature of the Bandidos Motorcycle Gang, starting with the founder, Donald Chambers, every national president of the Bandidos MC has been convicted of a felony, usually drug trafficking, and sentenced to prison. Since its founding in 1966, the Bandidos have expanded throughout the USA and internationally. The gang's first European expansion was in 1989 when the Bandidos "patched over" [assimilated] the Club de Clichy Marseille, France, igniting a war with the Hells Angels. The DOJ (2011) estimates that there are 2,000–2,500 Bandido members in the US and 13 other countries. According to law enforcement authorities, the Bandido MC is involved in transporting and distribution of cocaine and marijuana and the production, transportation, and distribution of methamphetamine.

Outlaws MC – According to their national website, http://www.outlawsmc.com/, the Outlaws MC is the oldest Outlaw Motorcycle Club, having been established in 1935 as the McCook Outlaws Motorcycle Club "out of Matilda's Bar on old Route 66 in McCook, Illinois just outside of Chicago." This motorcycle club was quite different from the Outlaws Motorcycle Gang of today. The DOJ (2011) estimates that this OMG has an estimated 1,700 members in 20 US states and 12 foreign countries. As is common with all OMGs, the Outlaws are involved in the production, transportation, and distribution of methamphetamine. They are also involved in the transportation and distribution of cocaine, marijuana, and to a lesser extent, ecstasy. The HAMC and the Outlaws MC are bitter enemies and have a long history of violence toward each other.

Pagans MC – The Pagans MC, a fierce 1 % biker gang with ties to organized crime groups in Philadelphia, Pittsburg, and New York, did not start out as a 1 % club. Lou Dobkins, a biochemist at the National Institute of Health, established the club in 1959 in Prince George's County, Maryland. Tradition says that the original 13 members wore white denim jackets with Surt the pagan Fire Giant carrying a flaming sword, logo on the back. The original members rode Triumph motorcycles. Their benign beginning did not foreshadow what the club would evolve into. During the 1970s, the fierce fighting club became an extremely violent criminal organization under the leadership of John Vernon “Satan” Marron. The Department of Justice reports that the Pagans are involved in the distribution of cocaine, methamphetamine, and PCP (DOJ 2011). The Pagans MC are the only Big 5 Outlaw Motorcycle Gang that does not have any chapters outside the United States, although there are rumors that they are attempting to establish a presence in Canada.

Sons of Silence MC – The Sons of Silence MC (SOSMC) also did not start out as a 1 % biker gang. The club was founded by Bruce, “The Dude Ruchardson” in Niwot, Colorado, in 1966. The club appears to have been formed as a “drinking” social club. As the club evolved into a gang, Richardson left the club. Today, the DOJ (2011) reports that the SOSMC is an Outlaw Motorcycle Gang with 30 US chapters and 5 chapters in Germany. Club members, according to the DOJ, have been involved in a wide range of criminal activities, including drug and weapons trafficking.

Puppet/Support 1 % Clubs

Puppet/Support Clubs are Outlaw Motorcycle Gangs affiliated with a dominant OMG. The puppet/support clubs do the biding of dominant clubs, serve as potential recruiting sources, and provide cannon fodder in the club wars. The OMG puppet/support clubs give a portion of their illegal gains to the dominant OMG. The larger OMGs will handle the wholesale distribution of the drugs and the puppet clubs will handle the more dangerous retail sale, thereby insulating the dominant club members from prosecution. On occasion if the subordinate club member is caught committing a crime and demonstrates that he “has class” [keeps his mouth shut and takes the fall], he will be rewarded by becoming a “prospect” for the dominant club.

The Red Devils MC is a well-known puppet/support club for the HAMC, and the Black Pistons MC and Forsaken Few are puppet clubs for the Outlaws MC. The Outlaws MC and the Bandidos MC list their puppet/support clubs on their national websites.

Independent Outlaw Motorcycle Gangs

Independent Outlaw Motorcycle Gangs are considered to be those who are not listed in the Big 5 and are not a puppet/support club of a Big 5 club. Because of the jealously of their territory by Big 5 clubs, it is not easy to be an independent club. Independent clubs operate in areas where there are no Big 5 Clubs or in the same area with permission of the dominant club. They are on friendly or tolerated terms with the larger club but remain independent of them.

Other independent clubs, such as the Mongols MC (national and international), operate in the same areas due to their ferocity and willingness to use violence. The Mongols MC and the HAMC fought a 17-year war over the wearing of the California bottom rocker. Other major independent clubs include the Vagos MC (national and international), Warlocks MC (national and international), and the Iron Horsemen MC (regional and national) and the smaller independents such as the Avengers MC, the Breed MC, and the Renegades MC who operate in limited geographical areas.

Black or Interracial One Percent OMGs

There are numerous one percent bikers that vehemently insist that there are no black 1 % bikers, insisting that like women, blacks are not welcome in the one percent subculture. It is true that the one percent subculture is basically sexist (women cannot be patch-holders in any 1 % club) and racist, with some clubs/gangs being more racist than others; however, it is not true that there are no black 1 % bikers or clubs. The Oakland, California, Eastbay Dragons MC is an all black 1 % club (Levington & Zimmerman 2003). The Wheels of Soul MC’s website says that they are “an Outlaw Motorcycle Club, founded in 1976 and is the only one of its kind. . ..we have a make-up of black white, Latino and Asian members (www.wosmc-trenton.com/history.html).” The Ching-a-Lings MC in Queens, New York City, are also interracial. Winterhalder reports there is a black, nationwide one percent club, the Hell's Lovers (Winterhalder 2005).

Related Entries

- ▶ Drug Enforcement
- ▶ Drug Trafficking
- ▶ Gangs and Social Networks
- ▶ History of Organized Crime in the United States
- ▶ Labeling Theory
- ▶ Labelling, Deviance, and Media
- ▶ Organized Crime, Types of

Recommended Reading and References

Barger R, Zimmerman K (2000) Hell's Angels: the life and times of Sonny Barger and the Hell's Angels Motorcycle Club. William Morrow, New York

Barker T (2007) Biker gangs and organized crime. LexisNexis/Anderson Publishing, Cincinnati

Davidson WG (2002) 100 years of Harley-Davidson. Bulfinch Press, Boston, p 2

DOJ (2011) Motorcycle gangs. U.S. Department of Justice. http://www.justice.Gov,criminal/oggs/gangs/motorcycle

Forkner W (1987) Easyriders: video interview. Cited in Reynolds, 2000

Fuglsang RS (1997) Motorcycle menace: media genres and the construction of a deviant culture. Unpublished Ph.D. Dissertation, University of Iowa

Hayes B (2010) American biker: the history, the clubs, the lifestyle, the truth. Flash Productions, Birmingham

Hayes B (2011) The One Percenter Encyclopedia: the world of outlaw motorcycle clubs from abyss ghosts to zombies elite. Motorbooks, Minneapolis

Levingston TG, Zimmerman KK (2003) Soul on bikes: the East Bay Dragons MC and the black biker set. MBI Publishing Company, St. Paul

Reynolds T (2000) Wild ride: how on the hogs: a biker filmography. McFarland & Company, Jefferson

Stidworthy D (2003) High on the hogs: a biker filmology. McFarland & Sons, Jefferson, NC

Veno A (2002) The brotherhoods: inside the outlaw motorcycle clubs. Allen & Unwin, Crows Nest

Wethern G, Colnett V (1978) A wayward angel. Richard Marek, New York

Winterhalder E (2005) Out in bad standing: inside the Bandidos Motorcycle Club – the making of a worldwide dynasty. Blockhead City Press, Owasso

Winterhalder E, De Clerco W (2008) The assimilation: rock machine become bandidos – bikers against the Hells Angels. ECW Press, Toronto

Wolf DR (1999) The rebels: a brotherhood of outlaw bikers. University of Toronto Press, Toronto

Yates B (1999) Outlaw machine: Harley-Davidson and the search for the American soul. Broadway Books, New York

Motorcycle Gangs

▶ Motorcycle Clubs or Gangs?

Multicenter Trials

▶ Multisite Trials in Criminal Justice Settings

Multiple Victimization

▶ Repeat Victimization

Multiple Victims and Super Targets

Ken Pease[1,2] and Graham Farrell[3]

[1]Jill Dando Institute, University College London, London, UK

[2]Loughborough University, Loughborough, UK

[3]Institute for Canadian Urban Research Studies and School of Criminology, Simon Fraser University, Burnaby, BC, Canada

Synonyms

Chronic victims; Multi-victims; Serial victims

Overview

The plight of multiple victims of crime is identified, with the observation that their needs are often ill-served by police and other agencies. Reasons for this are identified. These include the often inadequate information provided to officers attending the homes of such victims, together with their failure to recognize the cumulative effects on victims of crimes which taken individually may not be of great seriousness. The typical failure of victimization surveys to reflect the extent of chronic victimization is bewailed and remedies suggested.

Fundamentals

Let us start with an appeal to your personal experience. Please bring to mind a fellow pupil at school, a fellow student at college, a colleague at work, or someone in your neighborhood who is or was subjected to repeated verbal or physical attack, practical jokes, or theft or damage to their property. Everyone whom the first author has approached with this modest task has been prompt in their nomination of such a person and often very voluble and detailed in their description of what the person nominated has suffered. The case which most embarrassed British police in recent years was that of Fiona Pilkington, who killed herself and her disabled daughter rather than suffer continued victimization at the hands of young people living near her, about which she had called police 33 times without effective action being taken. The Independent Police Complaints Commission concluded:

> Police missed several opportunities to take robust action, inadequately investigated criminal allegations on some occasions and failed to record information on their own intelligence system.(http://www.bbc.co.uk/news/uk-england-leicestershire-13504618 accessed August 16th 2012)

The Chief Constable of the force concerned, Leicestershire, apologized unreservedly for police shortcomings in the case, mention of which ever since has sent a frisson of anxiety through senior police officers who are all too aware that chance rather than professionalism has saved them from similar embarrassment. When Fiona Pilkington died, the Home Secretary of the day said that there were “some hard lessons to learn about past failures, which will be the subject of further investigations.” The shadow Home Secretary concurred, “This case has horrified the nation” (http://www.guardian.co.uk/uk/2009/sep/28/fiona-pilkington-suicide-mother-police accessed August 16th 2012).

The Pilkington case was far from the only occasion when senior police officers have been called on to apologize for cases of chronic victimization which failed to “get above the radar” and trigger official action. On recent case was that of a pedophile ring which organized sexual abuse of children (http://menmedia.co.uk/manchestereveningnews/news/s/1493200_police-council-and-cps-apologise-for-failures-that-allowed-rochdale-sex-grooming-gang-to-abuse-girls-for-2-years-after-crime-was-first-reported accessed August 16th 2012), where one of the victims “was being driven to different houses or flats where there would be other men waiting. The abuse continued for months, until one evening she was arrested by police for smashing the counter at a takeaway where some of the men met the girls.” “Towards the end,” she asserted, “it could be up to five men in a day, sometimes every day, at least four or five times a week” (http://www.bbc.co.uk/news/uk-england-manchester-17914138 accessed August 16th 2012).

Mark Dyche murdered Tanya Moore. The pair had met at a Young Farmers’ ball and was soon engaged. But in February 2003, Miss Moore, tired of Dyche’s jealous and threatening behavior, ended the relationship. For a year Dyche waged a hate campaign against Moore, which included repeated threats to kill her. In June 2003, he even paid three men armed with baseball bats to rob and beat her at her family’s farmhouse home in Alkmonton, near Ashbourne, Derbyshire. Nottingham Crown Court heard that Dyche, who had a history of terrorizing women, “wanted her hurting, wanted her legs breaking, wanted her eyes gouging out, wanted to be in control.” He offered criminal associates £50,000 to kill her, but when no one came forward, he did it himself, lying in wait on a country road in March 2004 and blasting her in the face with a shotgun. A few days before she was murdered, Miss Moore presented officers with a bundle of threatening text messages from Dyche. Derbyshire Police Federation said that the incidents leading up to Tanya Moore's death were not properly examined due to systemic failures and overwork (http://news.bbc.co.uk/1/hi/england/derbyshire/6109108.stm accessed August 16th 2012).

Sadly the whole of this entry could consist of cases of this kind, but we must move on to some of the relevant research on the topic which satisfied that chronic victimization, serious in itself, can have ultimately fatal consequences.

Why Super Targets?

The use of aggression to establish dominance hierarchies is a commonplace among nonhuman primates and has long been a standard explanation for child behavior (Strayer and Strayer 1976), with sexual aggression among normally heterosexual prisoners also being interpreted in dominance hierarchical terms (Fagan et al. 1996). The stability of dominance hierarchies over time is consistent with such analyses (see Perry et al. 2001). These authors note that aggressive children selectively target their peers, with individual victimization rates becoming stable by the later elementary school years. The consequence is a set of serious adjustment problems being manifested, including depression, anxiety, social withdrawal, low self-esteem, suicidal tendency, avoidance by mainstream peers, and poor school performance. The problems were evident into adulthood (Olweus 1978).

In the classic Milgram studies of obedience, the willingness to inflict pain diminished as the distance from the "victim" increased (Milgram 1974). It is therefore no surprise to find that Internet "trolling" and cyberbullying now feature prominently in the suffering of the multiply victimized (McQuade et al. 2009). Googling "cyberbullying" at the time of writing in August 2012 yields over 53 million hits. As with the multiple victimization of Fiona Pilkington, some cases result in the suicide of the victim (http://www.truecrimereport.com/2010/01/phoebe_prince_15_commits_suici.php accessed August 18th 2012).

As for multiple victimization by offenses against property, a number of studies have reported offender accounts of why they return (Ashton et al. 1998; Shaw and Pease 2000). Most such reasons are self-evident, along the lines of "it was easy," with some laconic explanations as "Big house, small van" and some a little more subtle, as below:

> He went into a house and got a good haul. Two months later, walking by, he saw new stuff so did it again . . . took the same stuff (the replacements). 'If I do a place and its got good stuff in it – I think to myself "I'll come back 2 months later". Guaranteed – y'know what I mean, If it's good I'll come back. I did this geezer once three times before he got a [security] system up there . . . alarm bell. After I saw the alarm I thought . . . ah I'll leave it now.

While the most tragic climaxes to chronic victimization, such as the suicide of Fiona Pilkington and Phoebe Prince, are thankfully rare, the evidence of the long-term effects of school bullying suggests possible widespread sub-fatal effects of chronic victimization by crime.

The central practical issue besetting a police response proportional to harm suffered concerns the dilemma which an attending police office faces in gearing her response to the presenting situation. If the response is geared to the seriousness of the most recent presenting event, it will not be appropriate to the cumulative harm suffered as a consequence of a series of similar events. Unless the officer is apprised of the event history, the response will be too meager. Stated another way, the perennial problem is that sequences of events, none of which individually gets over the threshold of seriousness which demands police action, is often cumulative in its impact on victims. Mandy Shaw (2001) draws a parallel with the classic five stages of bereavement, denial, anger, bargaining, depression, and acceptance. In Shaw's view, chronic victims of crime cannot work through the stages to acceptance because new victimizations overtake them and they return to the initial stage. They never complete the recovery sequence because the offense (or a variant of it) is repeated. Mandy Shaw and Sylvia Chenery (2007) note the particular distress of men whose families are repeatedly targeted since it challenges their self-perceived role as family protector.

Certain crime types almost always render their victims chronically disadvantaged in their everyday life. Such offenses include intrafamilial violence, incest, stalking, and hate crimes, especially when the prevalence of hating is high and the prevalence of victims with a particular attribute (e.g., a particular ethnicity or sexual orientation) is locally low. Lest the tone of this entry seems unremittingly grim, it must be noted that there has been an astonishing change for the better in the perception of such offenses in the last 40 years or so, though we should not delude ourselves that the situation is as yet satisfactory. On BBC Radio's archive channel, warnings about possible offense being caused by dated "humor" are not infrequent. Two examples can be taken from the work of Eric Morecambe and Ernie Wise, British comedians whose material was at the time (1970s–1980s) deemed entirely uncontroversial and would still be judged as such by those who were around at the time. It would still be judged so, that is, *until it was heard again* with current sensibilities in place. When heard 35 years later in 2012, the impression is shocking.

In the radio program *The Morecambe and Wise Show* aired on December 4, 1977, a sketch depicts a couple seeking marriage guidance. One exchange goes as follows:

> *Wife: He put the boot in on me last week.*
> *Husband: A man needs a hobby.*

And the studio audience laughed.

A song whose lyrics now sound like a stalker's anthem passed without comment when first featured by Morecambe and Wise. They are reproduced in full here as a reflection of how far we have progressed since 1970 when the song was written.

> *Anywhere, I will find you,*
> *Don't care where, look behind you,*
> *I'll be there, following you around.*
> *Rain or shine, you won't shake me,*
> *I don't mind, where you take me,*
> *Spend my time following you around.*
> *Now listen, don't you know,*
> *Hiding from me does no good,*
> *Where ever you may go, I'll be in the neighbourhood.*
> *If you fly, I will follow,*
> *I don't care about tomorrow,*
> *As long as I'm, following you around.*
> *If you run, I'll run faster,*
> *Gonna stick, like a piece of plaster,*
> *Get my kicks, following you around.*
> *Get my kicks, following you around.*

Attempts at theoretical explanation of super targets within conventional criminology have been sparse and provisional (Farrell et al. 2005; Tseloni et al. 2010) and will not be dealt with here. Rather, the more important issue is addressed of how super-targeted people fail to come to official notice.

Key Issues/Controversies

Until the last third of the twentieth century, the sole measure of crime was a count of crime known to and/or recorded by the police. This remains, for the purposes of effective police intervention, the key index. A police officer cannot be expected to respond to something of which she remains unaware. Many, sometimes most, crimes are not reported to the police. Let us assume that there is a 50 % chance (0.5 probability) that any household burglary is reported to and recorded by the police. For a household which has been burgled twice, there is thus only a 25 % chance that both burglaries will be reported to the police (because $0.5 \times 0.5 = 0.25$). If three burglaries have been suffered, the probability that all three have been recorded is 0.125, and so on. (This makes the unrealistic assumption that the decision of any householder to report to the police is 50 % for every burglary suffered. This assumption is clearly unrealistic, but alternative assumptions, while they make the arithmetic more complicated, leave the central point untouched). The attrition is even greater if repeat crimes are less likely to be recorded by the police, as found by Mukherjee and Carcach (1998). So multiple victimizations will be diminished in police awareness by dint of non-report or non-recording, if for no other reason. We will never know how many times Fiona Pilkington suffered "low-level" criminality which she did not report to the police. All we know is that on 33 occasions she did report events, and the police recorded the fact. There are other reasons why a police officer responding to a call for service from a multiple victim fails to recognize the fact and extent of prior victimization. Historically, an inability to trace the same people or addresses in police data systems has been a further way in which multiple victimizations has been underrepresented in police data, although this may be declining with improvements in police information technologies. However, we should contrast the experience of those calling the police with the experience of calling large commercial concerns. Over the last 30 years, citizens calling commercial concerns have come to expect the person responding to the call to be able quickly to bring up on screen the caller's address, history as a client, and possibly the brand of toothpaste they use! During this same period of 30 years, however, it seems that the experience of someone calling the police has changed far less.

Police awareness of victimization history remains the key practical information. In a wider sense, knowledge of how chronic victims contribute to the extent and shape of the "crime problem" is equally important as providing a motivation to act. For this purpose, the victimization survey has come to take center stage. It provides a measure gleaned from asking representative samples of the adult population what they have experienced by way of crime over a specified recall period (usually 6 months or 1 year). A key aim of such crime victimization surveys has been to reveal the extent and nature of unrecorded crime. It would reasonably be concluded that victimization surveys would provide a good idea of the extent and nature of levels of chronic victimization. Analyzed in one way, they do. Analyzed in the conventional way, they do not.

In the writers' view, the central obstacle to putting in place routinely a mode of analysis of victimization surveys that adequately represents levels and distribution of multiple victimization has been the central concern of state funders of (expensive) victimization surveys, to see how and where statistics of crime known to the police coincide and where they diverge from statistics of crime which citizens say they have suffered. To do this, certain conventions are adopted which, taken together, have the effect of air-brushing chronic victims from the picture. The one described below is probably the most important and is described fully elsewhere (Farrell and Pease 2007). The example is taken from the British Crime Survey (BCS), but the practice is standard and accounts of the situation in the USA (together with a recent recognition of how the conventions distort the resulting picture of victimization) can be found elsewhere. It should be explained that series crimes are repeated crimes of the same type, committed in the same circumstances, and probably by the same person. It will be obvious to the reader that crimes of violence will be more solidly attributed to the same perpetrator. The BCS Training Notes (Budd and Mattinson 2000) comprise the principal reference point for present purposes. Pages 59–65 of the Training Notes contain the exact SPSS syntax (the software commands) used. The syntax contains the line:

$$\text{If (nseries ge 5) number} = 5.$$

which, translated, means: If there are five or more incidents in a series of crimes reported on a victim form, only five of them are included in the calculation of the crime rates. Simply put, nobody in the UK is allowed to experience more than five crimes in any one series. A woman beaten daily by her partner is counted as having suffered five assaults. Why is this apparently bizarre de-emphasis of multiple victimizations chosen? The BCS Training Guide explains capping as follows:

> For 'series' incidents the number of incidents is capped at 5. Therefore if someone reports 10 incidents in a 'series' only 5 are counted. The limit is to avoid extreme cases distorting the rates. (Budd and Mattinson 2000, p. 32)

Let us consider the interesting use of the word "distortion" here. Does it mean that chronic victims are presumed to be lying? If so, the belief

Multiple Victims and Super Targets, Fig. 1 'Capped' vs actual England and Wales crime (British Crime Survey 2005–2006)

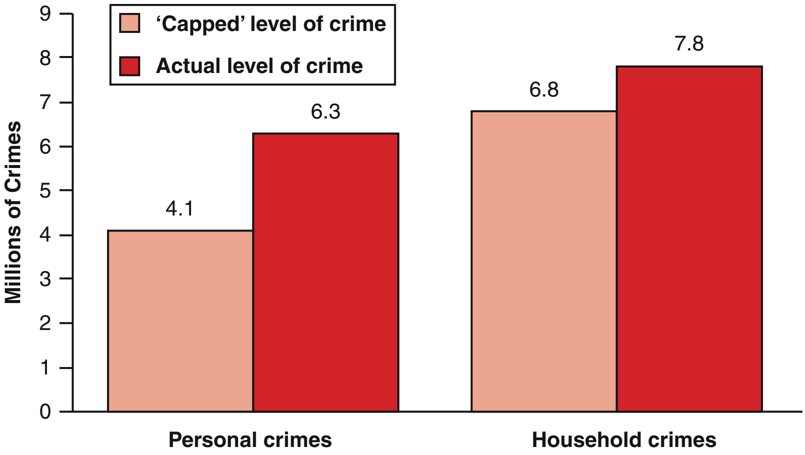

remains unstated although hinted at. When frequency distributions are examined for particular types of crime, the frequency of crimes said to have been suffered clusters around easy decimal system reference points: 10, 20, 30, 40 incidents, and so on, a phenomenon found elsewhere (Rand and Rennison 2005). There is also some indication of clustering around monthly reference points: 12, 24, and 36. These patterns are to be expected, and almost certainly reflect a tendency of respondents and interviewers to round, or approximate, to easy reference points (Rand and Rennison 2005). If interviewers suggest they were victimized "about" once per month, then an interviewer who is encouraged to report precise numbers would record 12 incidents. Analytically, such clustering is not a problem as discrepancies would be expected to be normally distributed around the cluster values. Any suggestion that this use of easy reference points compromises the veracity of victims is offensive. If you have been assaulted 47 times, rounding to 50 does not make you a bad person. So what is the "distortion" alleged to result from taking crime victims at their word? What it really means is that the overall incidence of crime would be more variable if it allowed chronic victims to be represented properly. Chronic victims being (thankfully) relatively uncommon, the crime incidence yielded by the survey would depend substantially on how many chronic victims were sampled. This would make the comparison with police statistics more problematic.

In short, the capping of series crimes at five events is a flagrant example of statistical convenience taking precedence over an accurate understanding of the distribution of human distress. As noted above, there are alternative approaches, for example the use of the Gini coefficient, which measures inequality and has been applied to crime victimization (Tseloni and Pease 2005). There has been a recent modest attempt to move toward the position advocated above by the US Bureau of Justice Statistics (BJS) (Lauritsen et al. 2012). It states the Bureau's intention as follows.

> Given the findings from this research, BJS will enumerate series victimizations using the victim's estimates of the number of times the victimizations occurred over the past 6 months, capping the number of victimizations within each series at a maximum of 10. This strategy for counting series victimizations balances the desire to estimate national rates and account for the experiences of persons with repeated victimizations while noting that some estimation errors exist. (p. 3)

While applauding the direction of movement, the objection in principle remains, with the stability of national estimation being given equal status with the facts about people whose lives are blighted by suffering multiple crimes, thereby becoming "super targets." The extent of the resulting difference in estimates of national crime incidence, particularly for crimes against the person, is evident from Fig. 1 below.

The "fat" tail of the victimization distribution arguably contains the bulk of the problem of crime and of the suffering inflicted upon its victims. There is an unanswerable case for crime counts and policing responses to reflect this. In fact, and to our collective shame, statistical and methodological convention has hitherto served to deflect attention from this fundamental truth.

Related Entries

- ▶ BCS
- ▶ National Victimization Surveys
- ▶ Repeat Victimization
- ▶ Understanding Victimization Frequency

Recommended Reading and References

Ashton J, Brown I, Senior B, Pease K (1998) Repeat victimisation: offender accounts. Int J Risk Secur Crime Prev 3:269–279

Budd T, Mattinson J (2000) British crime survey training notes. Home Office, London

Fagan TJ, Wennerstrom D, Miller J (1996) Sexual assault of male inmates: prevention, identification, and intervention. J Correct Health Care 3:49–65

Farrell G, Pease K (2007) The sting in the tail of the British Crime Survey: multiple victimisations. In: Hough M, Maxfield M (eds) Surveying crime in the 21st century. Willan, Cullompton

Farrell G, Clark K, Ellingworth D, Pease K (2005) Of targets and super-targets: a routine activity theory of high crime rates. Internet J Criminol

Lauritsen J, Gatewood Owens J, Planty M, Rand MR, Truman JL (2012) Methods for counting high-frequency repeat victimizations in the National Crime Victimization Survey. Bureau of Justice Statistics, Washington, DC

McQuade SC, Colt JP, Meyer NBB (2009) Cyber = Bullying. Praeger, London

Milgram S (1974) Obedience to authority: an experimental view. Harper & Row, New York

Mukherjee S, Carcach C (1998) Repeat victimisation in Australia. Australian Institute of Criminology, Canberra

Olweus D (1978) Aggression in the schools: bullies and whipping boys. Hemisphere, Oxford

Perry DG, Hodges EVF, Egan SK (2001) Determinants of chronic victimisation by peers: a review and new model of family influence. In: Juvonen J, Graham S (eds) Peer harassment in school: the plight of the vulnerable and victimised. Guilford, New York

Rand M, Rennison CM (2005) Bigger is not necessarily better: an analysis of violence against women. Estimates from the National Crime Victimisation Survey. J Quant Criminol 21:267–291

Shaw M (2001) Time heals all wounds? In: Farrell G, Pease K (eds) Repeat victimization. Crime prevention studies. Criminal Justice Press, Monsey

Shaw M, Chenery S (2007) Kings and castles, cavemen and caves: the impact of crime on male victims. In: Farrell G, Bowers KJ, Johnson SD, Townsley M (eds) Imagination in crime prevention. Crime prevention studies, vol 21. Willan, Cullompton

Shaw M, Pease K (2000) Research on repeat victimisation in Scotland. Scottish Executive, Edinburgh

Strayer FF, Strayer J (1976) An ethological analysis of social agonism and dominance relations among preschool children. Child Dev 47:980–989

Tseloni A, Pease K (2005) Population inequality: the case of repeat victimisation. Int Rev Victim 12:75–90

Tseloni A, Ntzoufras I, Nicolaou A, Pease K (2010) Concentration of personal and household crimes in England and Wales. Eur J Appl Math 21:326–348

Multisite Trials in Criminal Justice Settings

Faye S. Taxman[1] and Anne Giurianna Rhodes[2]
[1]Criminology, Law and Society, George Mason University, Fairfax, VA, USA
[2]Department of Criminology, Law and Society, George Mason University, Richmond, VA, USA

Synonyms

Clinical trials; Field trials; Multicenter trials

Overview

This entry describes the issues associated with conducting multisite trials in criminal justice settings, including how to determine the interventions to be tested, the selection of control groups, and the potential threats to internal validity of the trial, and gives two real-world examples of trials conducted and how they were managed and analyzed.

Introduction

Researchers conducting an experiment must decide on the type of design and the number of sites to include. The latter decision is one that receives little attention. As previously argued by

Weisburd and Taxman (2000), multicenter trials have many advantages, including the opportunity to test a new protocol or innovation within various settings. Single-site trials provide a starting point to test out the feasibility of a new innovation, but multisite trials have a clear advantage in testing the innovation under varied conditions. The multisite trial may offer the potential to accumulate knowledge more quickly by using each site as a laboratory for assessing the characteristics or factors that affect the question of "value-added" or improved outcomes. It creates the opportunity to simultaneously test the innovation as well as learn from the sites about the parameters that affect outcomes. Multisite trials provide insights about areas where "drift" may occur, whether it is in the innovation, the management of the study, external factors that affect the innovation, or the inability of the environment to align to the innovation.

This entry provides insight into the complexity of a multisite trial and how different design, site preparation, and research management characteristics can affect the integrity of the experiment. Case studies of two recently completed multisite trials that involve complex innovations or innovations that involve more than one operating agency are used. In these case studies, the various types of threats to internal validity of the trial are used as a conceptual framework to illustrate the decisions that researchers must make in a study.

The Multisite Trial

While randomized control trials are the "gold standard" of research, they are often challenged by design and implementation issues. Adding diverse sites complicates matters, particularly when the goal is to develop a standardized protocol across sites and the study sites have a wide range of expertise and resources. These issues can be even more pronounced in criminal justice settings, where design considerations are often impacted by the structure and administration of local, state, and federal agencies. While there has been a push for more randomized trials in the criminal justice field (Weisburd 2003), the number of randomized studies remains relatively low to the field of medicine, substance abuse, education, and other similar disciplines.

The first step of a multisite trial is to develop a study procedure detailing the intervention for the study sites and research team. The protocol should have (1) the rationale for the study, (2) a description of the intervention, (3) a list of the procedures to implement the study, (4) a copy of the instruments and key measures, and (5) human subject procedures including consent forms and certificates of confidentiality. This protocol is needed to provide the guidance in terms of implementation of the study in real-world settings. The protocol should consider the unique characteristics of the sites involved, particularly if these characteristics are likely to impact findings or influence the fidelity of the intervention. Up-front analyses of these issues, through pipeline or feasibility analyses or piloting of procedures, will increase the likelihood that the trial is uniformly conducted across study sites. The goal is to mitigate differences, which reduces the need to control for these issues during the analytic stage. The issues to be addressed in the protocol are:

Rationale for the Study. Every trial has a purpose that should be specified in study aims and hypotheses. The rationale provides a guidebook as to why the trial is necessary and what will be learned from the trial.

Theoretical Foundation. Trials are about innovations, either the introduction of a new idea or a change in existing procedure or process. The innovation should be guided by theory which explains why the new procedure or process is likely to result in improvements in the desired outcomes. It is important to ensure that each component of the procedure is theory-based to provide an explanation of why the component is needed as part of the innovation.

Defining the Study Population. Based on the study aims, the criteria for inclusion and exclusion should be specified to ensure that the innovation reaches the appropriate target population. Generally inclusion criteria relate to traits of the individual or setting. But in real-world criminal justice settings, other selection forces may be present that are a function of the process or environment. Researchers may not have access to the full base population, which can potentially bias results. Wardens or correctional officers may decide that offenders under certain security restrictions are ineligible for study participation. Selection forces can vary between study sites because of factors that are often not evident. Also, sites may operationalize selection criteria differently. A recent article on multisite trials in health services research recommended that the lead center develops selection criteria but allows each site to operationally define these criteria given the local environment (Weinberger et al. 2001).

Defining the Intervention: The innovation that is being tested should, by definition, be different than traditional practice. As discussed above, the intervention needs to be outlined in detail, including theoretical rationale, dosage and duration, procedures essential to implementation, and core components that will be tracked using fidelity measures (Bond et al. 2000). Most interventions consist of different components that are combined to produce the innovation. It is important to specify who will implement the intervention and how they will become proficient in the skills needed to deliver the intervention. Clear delineation of how the innovation will be integrated into the existing infrastructure of an organization and what potential issues are expected to occur and how these will be handled is also needed.

Defining the Control Group(s) or Practice as Usual: The main purpose of the control group is to serve as a contrast to the innovation. The control group represents practice as usual. This group is expected to have the same distribution of other factors that could potentially affect outcomes as the treatment group, which avoids the possibility of confounding. Randomization gives each eligible subject an equal chance of being assigned to either group in a two-armed design. Blinding, in conjunction with randomization, ensures equal treatment of study groups. In general, the services received by the control group should follow the principle of equipoise, meaning that there should be uncertainty about whether the control or the experimental condition is a better option (Djulbegovic et al. 2003). It is important for the hypotheses to be concerned about how alternative strategies (including treatment as usual) impact outcomes.

Measurement and Observations: Another important component of a study is how effects what instruments that is, will be used to measure outcomes and how and when these instruments will be used. Dennis and colleagues (2000) outline four types of variables to be collected: design variables (sites, study condition, weights, etc.), other covariates (demographics, diagnoses, etc.), pre- and post-randomization intervention exposure, and pre- and post-dependent variables. Most experiments have multiple waves of data collection. Outcome measurements are subject to a number of potential biases, including recall bias (self-report), sensitivity and specificity issues (biomedical tests), and rater error (file reviews). The types of data to be collected should be based on the hypotheses to be tested and the availability and cost of data. Multiple sources of measurement are recommended for the main outcomes, such as self-report and urine testing for drug outcomes.

Quantitative Analysis: The analysis of multisite trial data is dependent on the types of measures collected and the format of the variables. While the trial is ongoing, interim analyses may be conducted, mainly for data monitoring purposes. Some trials in the justice system require oversight from a data and safety monitoring board (DSMB). A DSMB reports on a regular basis to determine if subject flow and follow-up are adequate and if treatment effects are evident in the data. These interim data runs often utilize t-tests, ANOVAs, or effect sizes

(Cohen 1988). For more formal analyses, multivariate models are used to account for the repeated measures and multisite designs. With repeated measures data, hierarchical models can be used, which nest observations within person over time (Raudenbush 2001). Different types of these models are available including latent growth models (LGM), which capture individual differences in change trajectories over time (Meredith and Tisak 1990).

A multisite trial allows for the ability to pool the overall and specific impacts of treatment within separate sites in the context of a single statistical model. The sites in a multicenter trial can be seen in this context as building blocks that can be combined in different ways by the researchers. The randomization procedures allow each center to be analyzed as a separate experiment as well as combined into an overall experimental evaluation where the researcher is able to identify direct and interaction effects in a statistically powerful experimental context.

A multisite study is not a multicenter study where there is a random sample of sites themselves. Without a random sample of sites, a number of restrictions must be placed on the generalizations that can be made and the type of statistical models that can be estimated. Site must be defined as a fixed factor in the statistical models employed, because of the possible variations in the context of the sites. It is important to consider that the generalizability of these effects is limited and cannot be applied to the larger population.

Weisburd and Taxman (2000) recommend a statistical process for using mixed-effects models in analyzing multisite studies. A model for this example is presented in Eq. 1 below.

$$Y = \mu + S_i + T_j + ST_{ij} + O_{k(i,j)} + e_{k(i,j)} \qquad (1)$$

Tj is the treatment effects averaged across sites to obtain larger population of centers for the mixed-effects model.

S_i is the average impact of a site (or center) on outcomes, averaged over treatments.

ST_{ij} is the interaction effect between treatment and site, once the average impacts of site variation and treatment variation have been taken into account.

$O_{k(i,j)}$ is nested within a site by treatment combination.

Tools to Monitor Implementation: CONSORT Flow Charts in CJ Settings

Understanding the flow of subjects through the different phases of an experiment is important to manage the quality and integrity of the experiment. In multisite trials, subject flow may be influenced by facets of the particular criminal justice system, the behavior of study subjects, and/or factors that are difficult to identify. In 2001, the Consolidated Standards of Reporting Trials (CONSORT) statement was issued as a way to improve the reporting of information from experiments in healthcare (Moher et al. 2001), and it has now become a standard that experimenters strive for in social policy areas like criminal justice. The CONSORT flow chart illustrates how to report subjects through each trial phase; it was amended in 2008 to specifically address issues related to non-pharmacologic treatments (such as different types of therapy or different modes of sanctions). A copy of the amended chart is in Fig. 1 (Boutron et al. 2008). The CONSORT chart provides documentation for each phase of the experiment and provides sufficient information to assess the quality of the implementation of the experiment. The CONSORT chart is a standardized reporting format that clarifies the steps in the study and the attention to recruitment and retention issues.

This section describes the CONSORT flow chart for non-pharmacologic treatments (Fig. 1) as it relates to implementing trials in criminal justice settings. The unique challenges presented by these settings not only include the target population as potentially vulnerable to coercion and exploitation but also reflect the difficulties of work in settings that are, by their nature, authoritarian and often not amenable to the rigors of randomized trials.

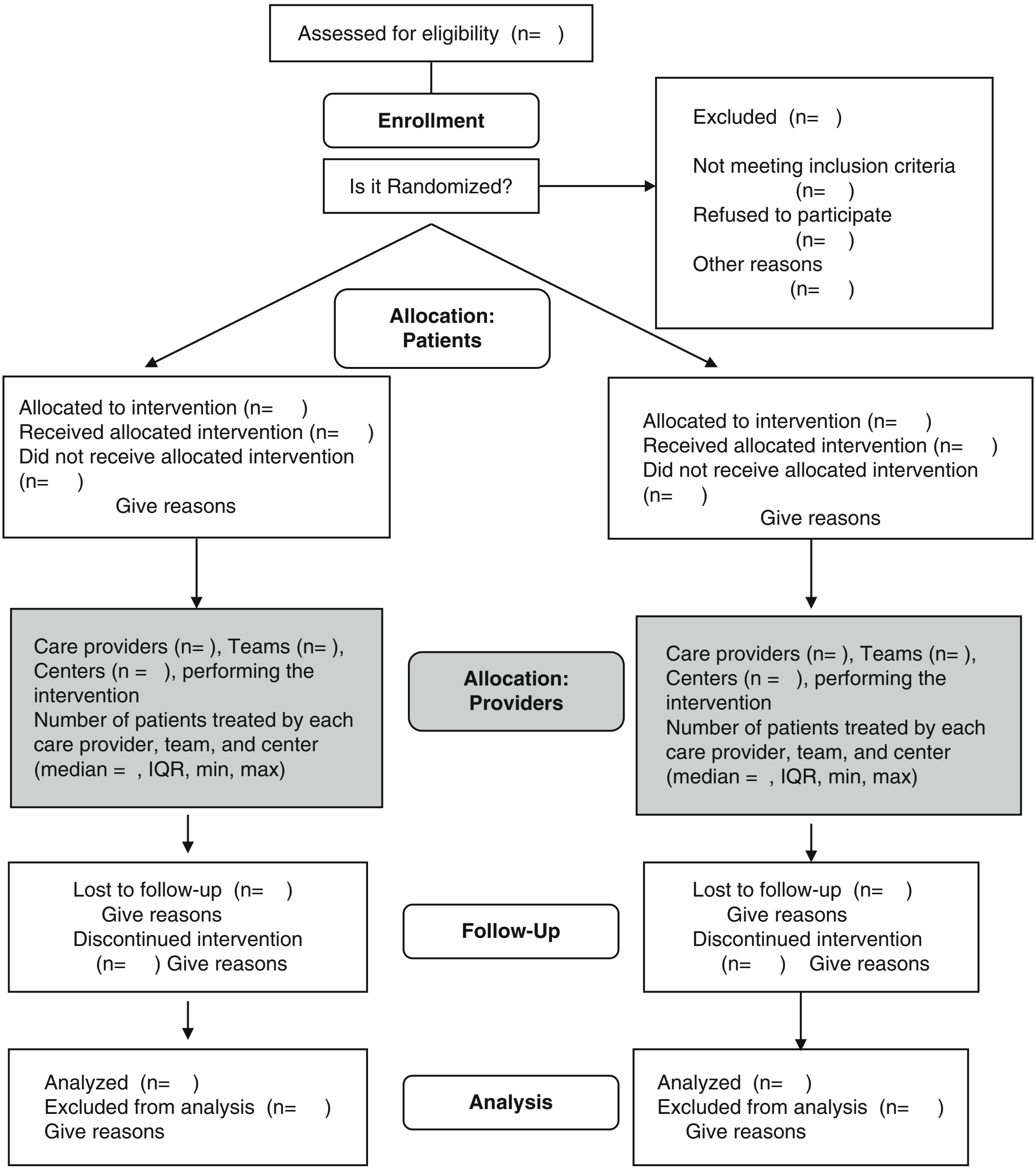
Assessed for eligibility (n=)
Enrollment
Is it Randomized?
Excluded (n=)
Not meeting inclusion criteria (n=)
Refused to participate (n=)
Other reasons (n=)
Allocation: Patients
Allocated to intervention (n=)
Received allocated intervention (n=)
Did not receive allocated intervention (n=)
Give reasons
Allocated to intervention (n=)
Received allocated intervention (n=)
Did not receive allocated intervention (n=)
Give reasons
Care providers (n=), Teams (n=), Centers (n =), performing the intervention
Number of patients treated by each care provider, team, and center (median = , IQR, min, max)
Allocation: Providers
Care providers (n=), Teams (n=), Centers (n =), performing the intervention
Number of patients treated by each care provider, team, and center (median = , IQR, min, max)
Lost to follow-up (n=)
Give reasons
Discontinued intervention (n=) Give reasons
Follow-Up
Lost to follow-up (n=)
Give reasons
Discontinued intervention (n=) Give reasons
Analyzed (n=)
Excluded from analysis (n=)
Give reasons
Analysis
Analyzed (n=)
Excluded from analysis (n=)
Give reasons

Multisite Trials in Criminal Justice Settings, Fig. 1 Sample CONSORT flowchart for non-pharmacologic treatments

The first section concerns eligibility. Most trials involve decisions about specific inclusion and exclusion criteria that should be in place for the trial so that the results can be generalizable to the appropriate population. The identification of people that fit the desired criteria may involve a multistep process. For example, a specific protocol may require the exclusion of sex offenders. This will require the research project to define the term "sex offenders" and then develop specific criteria to meet this term. In practice, researchers will generally need to conduct a prior chart review of offenders to make a determination of which offenders might meet the eligibility criteria. This screening often occurs before the consent process and baseline research interview. Other types of eligibility criteria are offense or location specifics, age or gender, language, or other characteristics. In general, screening criteria should be geared toward obtaining a study population that would be appropriate for the intervention. The questions for the CONSORT chart are how many subjects meet the criteria and what process is used to identify the target population. This defines the intent to treat group.

Prior to randomization, subjects may be dropped from the study because they fail to meet screening criteria or refuse to participate in the study. Once they are randomized, they may either receive their allocated intervention or they may not receive it for a variety of reasons including not meeting the eligibility criteria when they are interviewed, dropping out of the study, transferring to another CJ facility, or having a change in criminal justice status. All of these are examples of post-randomization factors that may affect participation in a study. The randomization process is such that once a subject is assigned to a condition, they are eligible for follow-up even if they only receive a partial dosage of the intervention or they did not complete the period of time devoted to the intervention.

The provider allocation section of the CONSORT chart was added in 2008 and characterizes the settings where the treatment(s) takes place. The researcher needs to identify all of the potential settings that might be involved in the experiment. These could be treatment facilities, jails, prisons, probation offices, courts, or other areas. Included are the number of providers, teams, and/or centers and the number of subjects in each area. Each organization will have different characteristics, and these require documentation.

A key issue is the follow-up rates for subjects in each arm of the study. The documentation should include the number that has been followed up and the reasons for missed follow-ups. Reasons for attrition from a study are critical to ensure sufficient power for the analyses. It is important to document follow-up rates at each wave, along with reasons for exclusions of any cases.

The last component of the CONSORT documentation is the results from the final outcome variables that relate to the study aims and the timing of all variables. The researcher is also required to document any assumptions that were made about the data.

Some criticisms have been made of the lack of detail present in the CONSORT chart, specifically the lack of information on implementation details. As our case studies demonstrate, the CONSORT chart does not capture a number of issues that can present threats to the internal validity of a multisite trial. Other scales and checklists have been developed to measure the quality of randomized trials, including the Balas scale for health services research (Balas et al. 1995) and the Downs scale for public health research (Downs and Black 1998). A review of these scales found little widespread use of them and also found that most had not been tested for construct validity or internal consistency (Olivo et al. 2008).

Case Studies of RCTs in Criminal Justice Settings

The following section present case studies of two RCTs that were implemented in settings covered in the Criminal Justice Drug Abuse Treatment Studies (CJDATS), a ten-center research cooperative sponsored by the National Institute on Drug Abuse from 2002 to 2008. Part of the objective of CJDATS was to use rigorous scientific methods to conduct studies in criminal justice settings. The purpose of the studies was to conduct trials at the point where offenders transition from incarceration to the community, referred to as the structured release or reentry process. The studies involved numerous agencies including prisons, probation/parole departments, prison or community treatment providers, and nonprofit organizations.

The case studies illustrate the issues regarding study design and management.

The conceptual framework for this analysis is the threats to internal validity that are specified in many textbooks on conducting social experiments (Campbell and Stanley 1966). By examining these issues, as well as looking at how the investigators managed the study, researchers can learn about the issues inherent in implementing multisite trials.

Threats to Internal Validity

- *Confounding*: The degree to which a third variable, which is related to the manipulated variable, is detected. The presence of spurious relations might result in a rival hypothesis to the original causal hypothesis that the researcher may develop.
- *Selection (bias)*: Both researchers and study participants bring to the experiment many different characteristics. Some of these characteristics that may not have been included in the eligibility criteria but may influence who participates in an experiment. These factors may influence the observed outcomes. The subjects in both groups may not be alike in regard to independent variables.
- *History*: Study participants may be influenced by events outside of the experiment either at the onset of the study or between measurement periods. These events may influence attitudes and behaviors in such a way that it might be detectable.
- *Maturation*: During the experiment, subjects may change and this change may result in how the study participant's reacts to the dependent variable.
- *Repeated testing*: Conducting multiple measures over period of time may result in bias on the part of the study participants. These impacts may occur from the failure to recall correct answers, or the prospects of being tested.
- *Instrument change*: An instrument used during the experimental period could be altered.
- *Mortality/differential attrition*: Study dropouts may be the result of the person, the experiment, or the situation. An understanding of reasons for the attrition and the differential rates of attrition can address the issue of whether the attrition is natural or due to some other factors.
- *Diffusion*: The experimental category should be different than traditional practice and movement from the experimental to control group or vice versa may pollute the assigned conditions.
- *Compensatory rivalry/resentful demoralization*: The control group behavior could be tainted by the result of being in the study.
- *Experimenter bias*: The researcher or research team may inadvertently affect the experiment by their actions.

Step'N Out: An RCT at Six Parole Offices in Five States

Step'N Out tested an integrated system of establishing target behaviors for offenders on parole, known as collaborative behavioral management (CBM). CBM requires the parole officers and treatment counselors to work together with offenders over a 12-week period, using a system of graduated sanctions and incentives. A more detailed description of the study can be found elsewhere (Friedmann et al. 2008).

Theoretical Intervention. CBM has three major components. First, it articulates parole and treatment staff roles, as well as the offenders and also makes explicit the expectations of each party. The clarification of expectations is covered in role induction theory where the emphasis is on clients understanding the roles and responsibility of different functions of the process. Second, a behavioral contract defines the consequences if offenders fail to meet those expectations. The behavioral contract specifies concrete target behaviors in which the offender is expected to engage in on a weekly basis; these target behaviors include requirements of supervision and formal addiction treatment and involvement in targeted behaviors that compete with drug use (e.g., getting a job, enhancing nondrug social network). Third, it regularly monitors adherence to the behavioral contract and employs both reinforcers and sanctions to shape behavior. The motto is "Catching People Doing Things Right," which is to say, the intervention creates the conditions to notice and reward offenders for achieving incremental pro-social steps as part of normal supervision (covered in the research literature on contingency management).

The CBM contract specifies expectations in terms of concrete target behaviors that the offender must meet before the next weekly session. These target behaviors are managed using a computer program which prints out a contract with copies for all three parties to sign and keep for their records. The CBM contract is monitored weekly to expedite identification and reinforcement of compliance and sanction of noncompliance, and then the contract is renegotiated and printed for the following week. Compliance with the contract earns points and, when preestablished milestones are reached, material and social rewards.

Control Condition. Offenders were supervised by the assigned parole officer in the normal manner that was used by that jurisdiction ("treatment as usual"). As shown in Table 1, each site had different standards for parole visits and drug testing.

Study Sites. Five CJDATS centers participated in Step'N Out – the University of Delaware (Wilmington parole office), Brown University/ Lifespan Hospital (Providence parole office), UCLA (Portland, Oregon, parole office), Connecticut (Bridgeport and Hartford parole offices), and George Mason University (Richmond, Virginia, parole office). All sites except Connecticut had one parole officer and one treatment counselor assigned to the CBM condition at their sites. In Connecticut, two or three officers were trained in CBM at each site, with one counselor also trained at each office.

Site Preparation. The initial two-and-half-day training for the Step'N Out teams occurred in December 2004. This training brought together parole officer and addiction counselor teams and their supervisors. After an overview of the intervention, most of the training focused on having each team practice skills in case-based role-plays with reinforcement and corrective feedback. A checklist of the key elements for fidelity to the protocol guided the role-plays and feedback. The teams were encouraged to negotiate roles with regards to initiating the role induction discussion, establishing goals, and setting target behaviors, but the protocol recommends that the parole officer (PO) takes primary responsibility for rewards and sanctioning and the counselor oversees problem-solving.

Additional on-site trainings were also done due to the lag time between the initial training and the time that sites began recruitment, the addition of new sites, and staff turnover. A 2-day booster training session in September 2006 focused on enhancing both the fidelity and finesse with which teams delivered the intervention.

POs were also asked to tape selected sessions. Feedback on the sessions was to be given on a monthly basis to build the skills of the POs in using the contingency management procedure.

Study Implementation. Table 2 outlines the issues encountered in implementing the study across the five sites. One of the first challenges was the point of randomization. Because those who were randomized to the CBM condition had to be assigned to a specific officer (the one who had received the CBM training), randomization had to occur at a point where PO assignment had not yet happened. This point varied among sites, with some assigning POs prior to an offender's release from prison, while others assigned the officer once the offender appeared at the parole office. While the ideal situation was to randomize at the parole office where the intervention was occurring, this was not possible in Delaware and Rhode Island. At those sites, the agency practice was to assign the officer prior to an offender's release, so screening and randomization were done while the offender was still incarcerated; this allowed the offender who was randomized to CBM to be assigned to the CBM officer. In Oregon, randomization began at the parole office, but recruitment was very slow in the initial months of the study and it was determined that the research staff had a better working relationship at the prisons; thus, recruitment was moved to the prison after

Multisite Trials in Criminal Justice Settings, Table 1 RCT implementation issues for Step'N Out

Issue	BP/HT	DE	OR	RI	VA
Place of randomization	Parole office	Halfway house	Parole office, switched to prison	Prison	Parole office
Selection bias	BP – 2 persons randomized but did not make initial parole session (CBM = 1, control = 1); HT – 2 persons (CBM = 1, control = 1)	85 offenders were not released to parole (CBM = 42, control = 43)	3 offenders not released (all control)	1 offender not released in CBM	Gang members were excluded
History	An event occurred (a parolee committed homicide) that resulted in a crackdown on parolees which increased violations			Recruitment was suspended in 2005. All subjects recruited were dropped from the study. Recruitment began again in July 2006 with offenders placed on electronic monitoring	
Maturation – time to touch-base session	6.1 days (0–27 days)	68.3 days (0–470 days)	51.6 days (0–175 days)	17.1 days (5–69 days)	10.7 days (0–78 days)
	CBM = 5.9 vs. control = 6.2 for Bridgeport	CBM = 47 vs. control = 92.8	CBM = 51.2 vs. control = 52.1	CBM = 9.8 vs. control = 20.4	CBM = 10.4 vs. control = 10.9
	7.7 days (2–25) for Hartford				
	CBM = 7.4 vs. control = 12				
Mortality/ differential attrition – 3-month FU	3M FU BP:	3M FU:	3M FU:	3M FU	3M FU
	CBM = 100 %	CBM = 98.9 %	CBM = 79.4 %	CBM = 100 %	CBM = 79.5 %
	Control = 100 %	Control = 97.9	Control = 87.5 %	Control = 85.7 %	Control = 91.3 %
	3MFU HT:				
	CBM = 100 %				
	Control = 95.7 %				
Mortality/ differential attrition – 9 months	9M FU BP:	9M FU BP:	9M FU:	9M FU:	9M FU:
	CBM = 97.2 %	CBM = 97.2 %	CBM = 73.5 %	CBM = 88.2 %	CBM = 64.1 %
	Control = 97.2 %	Control = 97.2 %	Control = 90.6 %	Control = 78.6 %	Control = 78.3 %
	9M FU HT:	9M FU HT:			
	CBM = 92.3 %	CBM = 92.3 %			
	Control = 82.6 %	Control = 82.6 %			
Selection-maturation interaction (average)	Control group saw parole Officers once a month	Control group saw parole Officers once a week	Control group saw parole Officers two times a month	Control group saw parole Officers once a week	Control group saw parole Officers once a month
Diffusion		Possibly in group SA sessions since treatment sessions for control and CBM at parole office			Possibly in group SA sessions as treatment for both control and CBM was at parole office

(*continued*)

Multisite Trials in Criminal Justice Settings, Table 1 (continued)

Issue	BP/HT	DE	OR	RI	VA
Compensatory rivalry		Anecdotal stories of resentment by control subjects of material and social rewards given to CBM			
Experimenter bias	Had a number of POs implementing CBM at each site, may have been differential implementation by PO, POs also had none CBM clients, may have diluted effects	Changed CBM POs partway through the study, may have had different implementation styles	Changed CBM counselor partway through process, may have different implementation styles		CBM PO felt isolated, also had substantial non-CBM clients, may have diluted effects

Multisite Trials in Criminal Justice Settings, Table 2 RCT implementation issues in HIV/Hep C Study

Issue	DE	KY	VA
Place of randomization	Halfway house	Prisons (2 male, 1 female)	Local jail
Start of follow-up	Randomization	Release date	Release date
Selection bias	Unknown selection forces for initial groups – had 406 randomized but only 391 released (96.3 %)	Unknown selection forces for initial groups – had 184 randomized but only 125 released (67.9 %)	Unknown selection forces for initial groups – had 113 randomized but only 67 released (59.2 %)
History	A confidentiality breach in DE on a different project shut down interviewing and recruiting until new processes were put in place	KY was unable to recruit in the men's prisons	Jail staffing issues and changes in security procedures affected recruitment; suspended for 2 months
Maturation		Longer time between randomization and follow-up than DE because of release dates	Longer time between randomization and follow-up than DE because of release dates
Repeated testing			
Mortality/differential attrition	30-day FU: Std = 98 %; NIDA = 100 %; DVD = 99 %; 90-day FU: Std = 98 %; NIDA = 98 %; DVD = 96 %	30-day FU: Std = 100 %; NIDA = 76 %; DVD = 92 %; 90-day FU: Std = 100 %; NIDA = 85 %; DVD = 92 %	30-day FU: Std = 90 %; NIDA = 73 %; DVD = 94 %; 90-day FU: Std = 75 %; NIDA = 60 %; DVD = 88 %
Selection-maturation interaction			
Diffusion	All arms of the study were all housed in the same facility for a period of time. Those in the intervention arms could share information with those in the control arms	All arms of the study were all housed in the same facility for a period of time. Those in the intervention arms could share information with those in the control arms	All arms of the study were all housed in the same facility for a period of time. Those in the intervention arms could share information with those in the control arms
Experimenter bias	Changed DVD interventionist partway through the study, also changed NIDA standard interventionist. Occasionally had same person do both arms		

6 months. In the end, half of the sites were randomizing in prison and half were randomizing at the parole office.

A consequence of the place of assignment is that not all randomly assigned participants were exposed to the intervention. A substantial portion of those who were randomized in prison never made it to the parole office. In Rhode Island, this was due to the fact that many offenders "flattened" in prison, or served out their parole sentences in prison, because of the unavailability of space in halfway houses where offenders were assigned upon release. In Delaware, many offenders committed violations in the halfway house that resulted in them being returned to prison before they could begin their parole sentence. This group of "post-randomization dropouts" was problematic, because they were technically part of the study population to be followed, yet they were not exposed to the intervention or any study activities. Many had no time in the community post-randomization. Rhode Island dropped their initial cases and changed their recruiting practices so that all of their cases were to be placed on electronic monitoring after release. Delaware also changed their recruiting to make their screening closer to the initial parole session. These measures drastically reduced post-randomization dropouts. The final CONSORT chart for Step'N Out is given in Fig. 2 (note that when the study was done, the revised CONSORT for non-pharmacologic interventions had not yet been published). There were a total of 93 post-randomization dropouts, with the majority (n = 85) from the Delaware site.

The "place of randomization" turned out to be a significant confounding variable in statistical analyses. Since the studies had different places of randomization (i.e., prior to release from prison, at the parole office, etc.), we were able to disentangle the impact of place of randomization on the study outcomes. Also, as is shown in Table 1, other selection forces also occurred during the study at different sites. In Virginia, gang members were never sent to the research interviewer for screening because a special program for gang members was being conducted with an assigned PO.

Another important site difference in implementation was the time it took for a person to begin the intervention after randomization. In some sites, those in the CBM group were able to begin parole services sooner than those in the control group, perhaps due to the availability of a dedicated PO for the CBM intervention who did not have the same large caseloads as other officers. In both Delaware and Rhode Island, it took about half the time for those in CBM to begin parole services as compared to those in the control group. This affects the outcomes which may be a function of the timing rather than the superiority of the CBM intervention attributes (MacKenzie et al. 1999).

The follow-up clock for Step'N Out began with the date of the first parole session (the "touch-base" session). Because the time to the touch-base session was longer for some subjects than for others, the time to the follow-up interviews also varied. In some cases, the 3-month follow-up occurred about 3 months after randomization (when randomization occurred close to the first parole session), but in other cases, there was over a year of time between randomization and the first parole session resulting in over 15 months between the baseline and the 3-month follow-up. While the follow-up was capturing the first 3 months on parole, there was a wide variation in how much time had elapsed since randomization for each subject. It was also complicated since some sites (Delaware) tended to have different timing between follow-up periods.

An attrition-related issue involved missed follow-ups. The Delaware center, which had the largest number of clients in Step'N Out, had staffing issues during the project and missed the 3-month follow-up window for a number of subjects. Because the main outcomes relied on data from the 3-month interview, provisions were made to obtain these data at a later date. At the 9-month interview, a full timeline follow back was completed for these subjects which included all time back to the initial parole session. The chances for historical and maturation issues were increased by this procedure, especially

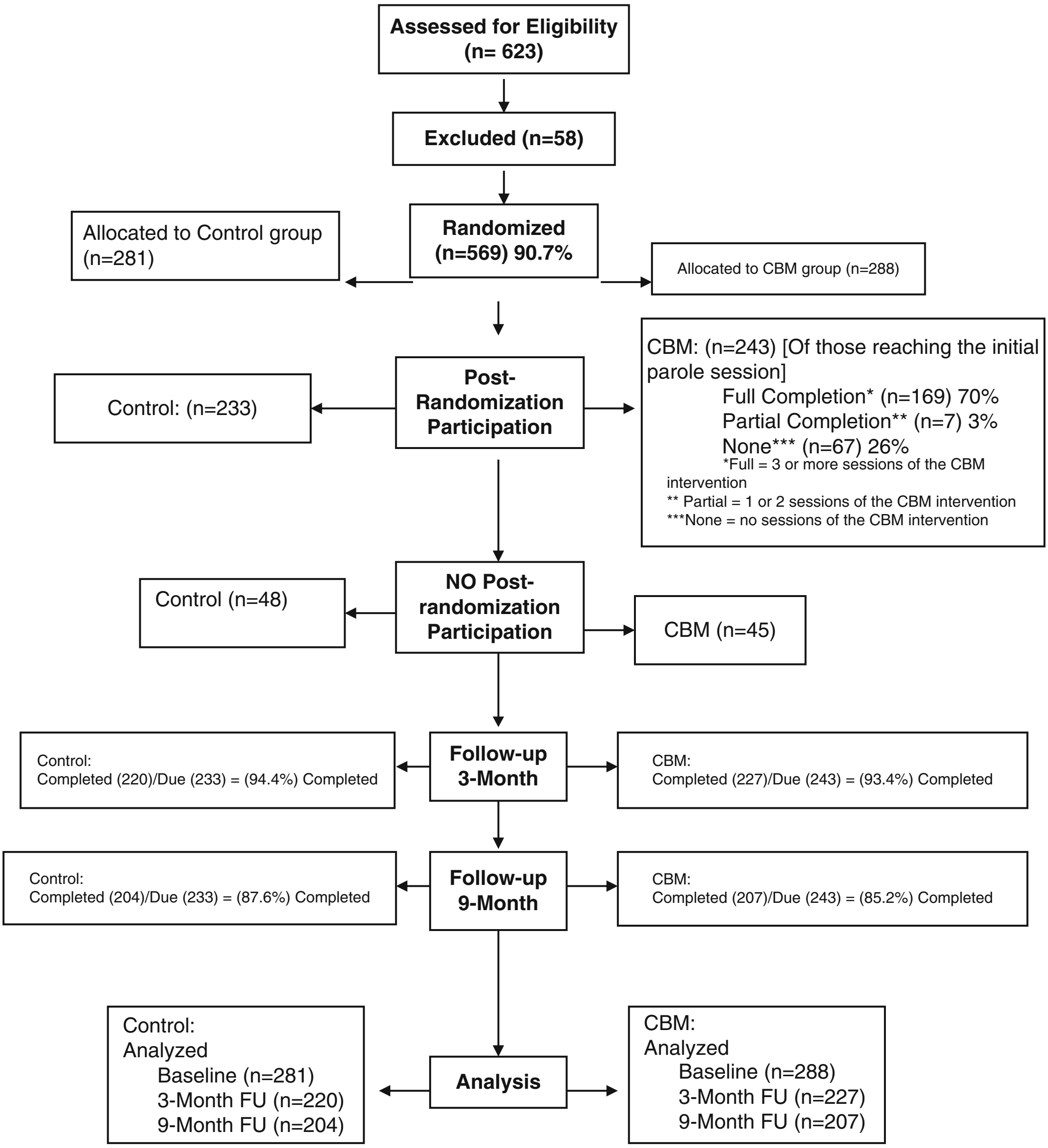
Assessed for Eligibility (n= 623)
Excluded (n=58)
Randomized (n=569) 90.7%
Allocated to Control group (n=281)
Allocated to CBM group (n=288)
Post-Randomization Participation
Control: (n=233)
CBM: (n=243) [Of those reaching the initial parole session]
Full Completion* (n=169) 70%
Partial Completion** (n=7) 3%
None*** (n=67) 26%
*Full = 3 or more sessions of the CBM intervention
** Partial = 1 or 2 sessions of the CBM intervention
***None = no sessions of the CBM intervention
NO Post-randomization Participation
Control (n=48)
CBM (n=45)
Follow-up 3-Month
Control: Completed (220)/Due (233) = (94.4%) Completed
CBM: Completed (227)/Due (243) = (93.4%) Completed
Follow-up 9-Month
Control: Completed (204)/Due (233) = (87.6%) Completed
CBM: Completed (207)/Due (243) = (85.2%) Completed
Analysis
Control: Analyzed
Baseline (n=281)
3-Month FU (n=220)
9-Month FU (n=204)
CBM: Analyzed
Baseline (n=288)
3-Month FU (n=227)
9-Month FU (n=207)

Multisite Trials in Criminal Justice Settings, Fig. 2 CONSORT flowchart of participants in Step'N Out study

with the different time frames for data collection across the various sites. This strategy allowed for the study to obtain needed outcome data, but it also created potential recall issues, where subjects were asked to self-report drug use, arrest, crime, and living situation data on a daily basis up to 18 months after it occurred. The varying lengths of the follow-up period may also have affected the integrity of the data.

An issue that was inherent to the specific criminal justice systems in each state was the status of the control group. Because this group received parole and treatment services "as usual" in each state, the services received by the control group were not uniform across all sites, as shown Table 1. In some sites, parole officers met with subjects once a month; in others, once a week. These were due to standard conditions of parole, but other variations could occur because individual offenders had different conditions which required different schedules. The same issues applied to substance abuse treatment. While the screening instrument used to assess eligibility ensured that those in the study had substance abuse dependency issues, it was not required that members of the control group would receive treatment since this decision depended upon the local parole systems. Not all parole officers recommended their clients for treatment. Thus, variations occurred in the conditions under which the control group was supervised both within a site and across sites.

The HIV/Hepatitis C (Hep C) Study: A Three-Armed RCT at Three Centers

This RCT was designed to test the efficacy of a criminal justice-based brief HIV/HCV intervention administered during the reentry period. Rates of both HIV and hepatitis C are disproportionately high for those involved in the criminal justice system, and the reentry period has been found to be a particularly risky time for offenders. This study was a three-group randomized design, with participants assigned to the following: (1) current practice, a group of transitional releases who saw an HIV awareness video shown as part of group sessions in the facilities in the prerelease process; (2) the NIDA Standard Version 3 HIV/Hep C prevention intervention delivered by a health educator via cue cards in a one-on-one didactic setting; and (3) the CJDATS Targeted, a near-peer facilitated intervention that used an interactive DVD format with gender/race congruent testimonials, also in one-on-one setting. Participants in all three groups were offered HIV and HCV testing. The full study design has been described elsewhere (Inciardi et al. 2007). The study was conducted by three CJDATS centers, the University of Delaware (lead center), the University of Kentucky, and George Mason University. The main purpose of the study was to test the interactive DVD method to determine if those who were exposed to a gender and culturally specific message had greater reductions in risky behaviors than those that receive the NIDA standard and current practice arms of the study. Follow-ups for this study were done at 30 and 90 days after either randomization or the date of release from prison.

Theoretically Driven Intervention. Based on focus groups with offenders, the study team developed a DVD that addressed issues related to reducing risky behaviors that involve the transmission of HIV/Hep C. The messages were then developed to be consistent with the gender and race (i.e., Caucasian, African American) of the offender. The theory was that the messages delivered in a gender- and culturally specific manner would result in greater compliance. The DVD arm of the study was administered by a near-peer, usually a former addict and/or someone with HIV or HCV. There were four separate DVDs, based on race/gender combinations (black male, black female, white male, white female), and they included testimonials from former offenders and persons infected with HIV and HCV discussing their situations.

Control Conditions. In this study, several treatment as usual conditions were used: (1) traditional practice where a group of offenders were shown an a HIV awareness video as part of group sessions in the facilities in the prerelease process and (2) the NIDA Standard Version 3 HIV/Hep C prevention intervention delivered by a health educator via cue cards in a one-on-one didactic setting (National Institute on Drug Abuse 2000). These represented two different conditions that reflected the "variety of practices" available in the field.

Study Recruitment. Potential participants were shown a video on HIV prevention in a group setting while still incarcerated (see section on setting by site) and were given information about the study and asked if they would like to participate. If they said yes, they were then screened individually to determine eligibility, and if eligible, they went through the informed consent process and were randomized into one of the three arms of the study. Testing for HIV and HCV was done at baseline and was voluntary.

Implementation. Table 2 presents some of the implementation issues of this study. Most of the challenges resulted from the varied settings where the intervention took place. For Delaware, the study was implemented in a halfway house facility, a place which ensured an adequate flow of, and continuous access to, participants. This site selection was convenient but in many ways altered the original intent of the study, as the offenders at this facility were already at risk in the community since they reside in the halfway house but still can spend time in the community. In Kentucky, the study was implemented at a men's prison and a women's prison, while in Virginia it was implemented at a local jail. Because the point of the follow-up period was to obtain information on risky behaviors in the community, the start of the follow-up clock was different according to site. In Delaware, because subjects were already at risk at the time of randomization, the follow-up clock started at the randomization date. In both Kentucky and Virginia, the follow-up clock did not start until the subject was released from the institution. This situation created a number of problems, including the fact that the type of site (prison, jail, halfway house) was completely nested within state and that the time at risk was not comparable among all subjects. Those in Delaware may have had some time at risk prior to being enrolled in the study. While baseline behavior data were gathered for the period prior to the last incarceration, the follow-up data collected may not reflect the same period of initial risk for those in Delaware as for those just released from incarceration in Kentucky and Virginia.

Another issue that may have affected implementation was the initial selection of the group of offenders to watch the video at each site. Generally, these groups were gathered by personnel at the facilities who were given the criteria for the study (offenders had to be within 60 days of release, speak English, and not have cognitive impairments). Researchers were not often privy to the decisions about the type of offenders that participated in the sessions. It may be that different types of groups were selected for inclusion at the different sites.

Since follow-up was dependent on release in both Kentucky and Virginia, a substantial percentage of those randomized were ineligible for follow-up because they were not released during the time frame of the study. While those randomized were initially scheduled for release within 60 days of recruitment, the release date often changed because of pending charges or other issues. In Kentucky, 33 % of the randomized population was never released, and in Virginia, 41 % of the randomized population was not released during the time frame of the study. This issue had the potential to create selection bias and differential attrition. The final CONSORT chart for this study is given in Fig. 3. A substantial number of those assessed were lost both prior to completing the baseline and also post-randomization, because of prison release issues.

Finally, in one site, the same interventionist, a peer counselor, administered the NIDA arm and DVD arm for a period of time. Bias should have been minimized since the interventionists could also do the testing for those in the control arm. There is a potential that the interventionists provided information to that group that was specific to their arm of the study. The interventionists were not blind to the study allocation of the subjects.

Discussion

Multisite trials are important tools to accomplish the scientific goals of (1) testing the efficacy of an innovation; (2) simultaneously "replicating" the study in multiple sites to determine the robustness of the innovation; and (3) allowing the unique characteristics of agencies, communities, and/or target populations to be concurrently examined in the test of an innovation to specify acceptable parameters. The trials represent the science of conducting field experiments but the process also reveals that there is an art in conducting randomized trials. It is the art – the crafting of the experiment, the mechanisms to conduct the experiment, and the daily decisions regarding the handling of "real-world" situations – that determines whether the experiment achieved the goal of testing the theoretically based innovation.

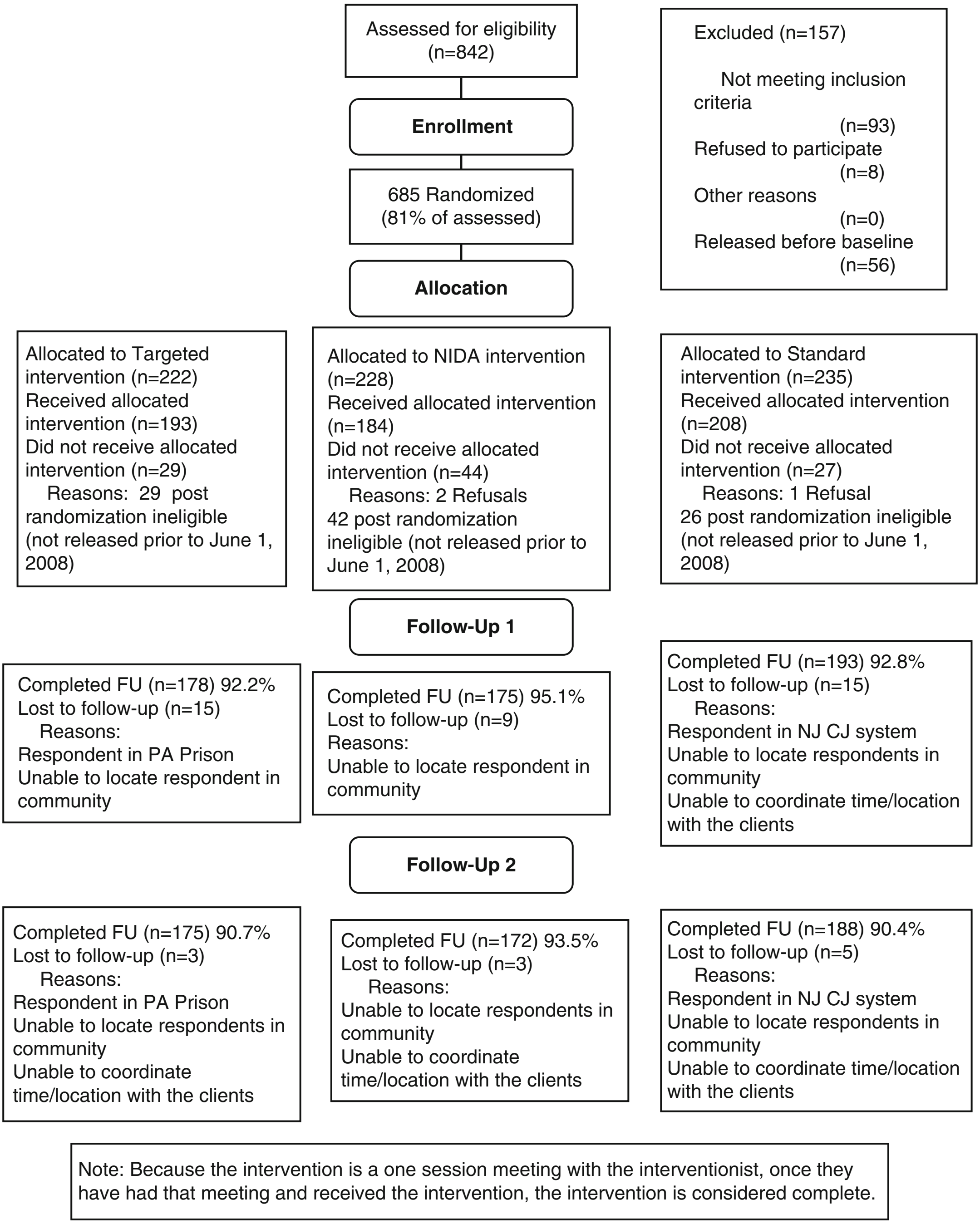
Assessed for eligibility (n=842)
Enrollment
685 Randomized (81% of assessed)
Allocation
Excluded (n=157)
Not meeting inclusion criteria (n=93)
Refused to participate (n=8)
Other reasons (n=0)
Released before baseline (n=56)
Allocated to Targeted intervention (n=222)
Received allocated intervention (n=193)
Did not receive allocated intervention (n=29)
Reasons: 29 post randomization ineligible (not released prior to June 1, 2008)
Allocated to NIDA intervention (n=228)
Received allocated intervention (n=184)
Did not receive allocated intervention (n=44)
Reasons: 2 Refusals
42 post randomization ineligible (not released prior to June 1, 2008)
Allocated to Standard intervention (n=235)
Received allocated intervention (n=208)
Did not receive allocated intervention (n=27)
Reasons: 1 Refusal
26 post randomization ineligible (not released prior to June 1, 2008)
Follow-Up 1
Completed FU (n=178) 92.2%
Lost to follow-up (n=15)
Reasons:
Respondent in PA Prison
Unable to locate respondent in community
Completed FU (n=175) 95.1%
Lost to follow-up (n=9)
Reasons:
Unable to locate respondent in community
Completed FU (n=193) 92.8%
Lost to follow-up (n=15)
Reasons:
Respondent in NJ CJ system
Unable to locate respondents in community
Unable to coordinate time/location with the clients
Follow-Up 2
Completed FU (n=175) 90.7%
Lost to follow-up (n=3)
Reasons:
Respondent in PA Prison
Unable to locate respondents in community
Unable to coordinate time/location with the clients
Completed FU (n=172) 93.5%
Lost to follow-up (n=3)
Reasons:
Unable to locate respondents in community
Unable to coordinate time/location with the clients
Completed FU (n=188) 90.4%
Lost to follow-up (n=5)
Reasons:
Respondent in NJ CJ system
Unable to locate respondents in community
Unable to coordinate time/location with the clients
Note: Because the intervention is a one session meeting with the interventionist, once they have had that meeting and received the intervention, the intervention is considered complete.

Multisite Trials in Criminal Justice Settings, Fig. 3 CONSORT flowchart of participants in HIV/Hep C study

The analysis in this entry presented the art in terms of issues that have been identified as threats to internal validity (Dennis et al. 2000). We used these tests of internal validity to demonstrate how experimental processes and research designs affect the integrity of, and findings from, an experiment. These two experiments allow us to draw lessons about study management to reduce the degree to which the threats affect the dependent variables:

1. *Confounding*. An important step in reducing the influence of third variables is to conduct three pipeline flows: (1) the research process in each participating site, (2) the core components of the innovation should be outlined to determine what other factors affect the independent variable, and (3) the potential mediators or moderators that may affect the independent variables. All three of these preexperimental processes are important to minimize extraneous variables.
2. *Selection Bias*. The inclusion and exclusion criteria of an experiment are critical in assessing potential sources of bias. But more important than simply analyzing these a priori decisions is to map these criteria to the study population to ensure that the research processes are not affecting selection bias. For example, in the Step'N Out experiment, the researchers in one study site were unaware that the agency did not release offenders without the use of extra controls (e.g., electronic monitoring). Yet, in the early stages of the experiment, when this was discovered, it resulted in consideration of new eligibility criteria. Maps of the system would assist in this process.

   The two multisite trials analyzed here had other selection biases that crept into the study such as the use of different types of recruitment sites in the HIV/Hep C study. While this variation was useful in assessing the elasticity of the innovation across different correctional settings (prison, work release, jail), it introduced new problems, including a large population of post-randomization ineligibles who were not released from the prison/jail within the study period of 12 months. This created a potential bias because it affects the intent to treat group, and the differences in releasees versus detainees must be assessed to determine the degree to which the study protocol affected study outcomes. The use of a historical control group can mitigate issues with selection bias for this part of the population and can also reduce other biases, including diffusion, compensatory rivalry, and experimenter bias. The researcher must usually rely on records to obtain data for the historical group, but this strategy is useful, especially when the experimental condition appears to be clearly superior to the treatment as usual and denying it to a portion of the current population would be unethical. Another related issue is to have a historical control group that could define the base rates for the dependent variables which can assist with understanding whether there was a Hawthorne effect with the study or whether the inclusion and exclusion criteria affect the dependent variable(s).
3. *History*. The impact of events outside of the experiment, either at the onset of the study or between measurement periods, may influence measured outcomes or independent variables. In Step'N Out, a crackdown on parolees in Connecticut occurred during the follow-up period, resulting in a higher number of parole violations than would have otherwise occurred. While it was felt that this event did not differentially affect the study groups, it was presumed that the overall number of parole violations was higher in the study than it normally would have been. This information is important to know and to report with all analyses of the study, especially as parole violations were considered a main outcome and ended up being much lower for the CBM group than the control group.
4. *Maturation*. During the experiment, subjects may change, and this change may result in how the study participant reacts to the dependent variable. In both studies, some sites had longer times between randomization and the follow-ups because of timing and release issues and some subjects also had longer times until their follow-ups were completed. In these cases, it is difficult to determine if the intervention caused positive changes or if more time in the community and exposure to other factors could be a possible cause. Future studies could focus on ensuring that randomization and follow-up schedules are implemented so that there is little room for

variance and that research interviewers are trained to understand the importance of completing follow-ups in the given time windows. For example, studies that focus on the release period could delay randomization until an offender is released to avoid the uncertainty associated with release dates.

5. *Mortality/Differential Attrition*. Study dropouts may be the result of the person, the experiment, or the situation. In some Step'N Out sites, the CBM group had lower follow-up rates at both 3 and 9 months, indicating that offenders may have found it more difficult to complete the CBM intervention and to stay connected with the parole office. The implementation of the intervention at these sites should be explored to determine how it differed from other sites. In Oregon, for example, those participating in CBM had to attend parole sessions at a different office which may have been inconvenient for them. In another site, the CBM officer was required to also maintain a caseload of regular parole clients, and she often had scheduling conflicts with her CBM clients. These issues provide insight into how implementation can affect participation. One suggestion for future iterations of Step'N Out was to have it be an office-based, rather than officer-based intervention, where the entire parole office practiced CBM.
6. *Diffusion*. The experimental category should be different than traditional practice, and movement from the experimental to control group or vice versa may pollute the assigned conditions. While there was no known crossover from one study condition to another, there was potential in both studies for offenders from different conditions to mix with each other and share information. In Step'N Out, one site had a control group that was not dissimilar to the CBM group, as the study site had specialized parole services where POs met with their clients weekly. These POs often used novel methods of sanctions and rewards. In these cases, it is difficult to determine the explicit effects of the intervention under study, as (1) the control group may have been influenced by the intervention and (2) the control group may not reflect current practice accurately. For multisite trials, this variation among current practice at sites is an issue that is often raised. An initial feasibility study at each site should identify potential issues and provide insights on how each site works. Each study protocol should have some flexibility so that it can be adapted to multiple sites but also must specify minimum conditions for the control group that must be met in order for the site to participate.
7. *Compensatory Rivalry/Resentful Demoralization*. The control group behavior could be tainted by being in the study. This is more of a possibility in longer lasting interventions where the treatment under study appears as a "better" alternative to those who are targeted for treatment. In Step'N Out, there was some evidence that persons wanted to be in the CBM condition (often asking for this at randomization) and felt those in CBM were getting an "easy ride." This could lead to resentment from the control group, especially if they were aware that the CBM group would receive material rewards for good behavior and they would not. This could create a disincentive for good behavior for those in the control group, and study results could be biased upward, with results better than they should be under normal circumstances. Researchers should also be careful in using language to describe the study conditions to ensure that the experimental condition is not seen as highly preferable to the current practice and should also work with staff at the agencies to train them not to present the intervention as an improvement or something better, as staff attitudes are usually quickly conveyed to offenders.
8. *Experimenter Bias*. The researcher or research team may inadvertently affect the experiment by their actions. Both of the studies we examined here had interventionists on-site at justice agencies who were not blinded to the study condition assignment of the offenders. Blinding cannot be done for the interventionist in these types of studies as they must know

what condition a person is in so that they can deliver the correct intervention. While a number of measures were in place to ensure fidelity (tapes in Step'N Out, checklists in HIV), there were concerns that POs may have implemented CBM differently at different sites and even at the same site, based on the person involved. The same is true for the health educator and peer interventionist in HIV, especially the peer counselor who had less of a script to work with and used more of their own experience in working with their clients. While the tapes were a good idea in Step'N Out, they were not used effectively as a feedback tool, as there was a long delay in coding and providing feedback to the POs. Because delivery of the intervention is such a key component of the study, researchers should make provisions for ongoing and timely fidelity monitoring of the intervention, with a clear plan on how fidelity will be tracked, how feedback will be provided, and how corrective actions will be taken (Carroll et al. 2007).

As a researcher considers validity threats to a multisite experiment, they should ask themselves the following five questions and care fully consider how the decisions that are made affect the experiment:

1. How does the issue affect the flow of subjects on the CONSORT chart?
2. How does the issue affect inclusion and exclusion (selection forces)?
3. Does the issue disrupt the theory underlying the intervention? Or does it disrupt the treatment as usual?
4. How might the issue affect the dependent variable(s)?
5. How might the issue affect possible confounders?

Conclusion

Multisite field experiments can test an innovation as well as the environment where an innovation can be offered. In addition, the multisite field experiment requires the researcher and/or research team to be even more sensitive to the impact of design and implementation decisions on the integrity of the experiment, as well as the generalizability to the wider target population. This entry has been designed to help researchers think about the issues related to internal validity threats and how to make decisions about the target population, eligibility processes, intervention design and processes, and instruments. Managing one study site presents sufficient challenges, while managing more than one study site requires the researchers and investigator(s) to be cognizant of how each decision at each site might affect whether the goals of the study can be obtained. In both case studies, we have examples where practical realities crept into the experiment. More work will need to be done to determine the impact on the outcome variables. In the end, it is apparent that scientists must also be good research administrators to ensure the study sites uphold the experimental design and that each decision is weighed against the concern about the integrity of the design.

Acknowledgements This entry was funded in part under a cooperative agreement from the US Department of Health and Human Services, Public Health Service, National Institutes of Health, and National Institute on Drug Abuse (NIH/NIDA) to George Mason University (Grant U01 DA016213-01, Action Research to Advance Drug Treatment in the CJS). The funding for this cooperative agreement was supplemented by the Center for Substance Abuse Treatment, Bureau of Justice Assistance, Centers for Disease Control and Prevention, and National Institute on Alcohol Abuse and Alcoholism. The authors acknowledge the collaborative contributions by federal staff from NIDA and the other nine research center grantees of the NIH/NIDA CJDATS Cooperative. The contents are solely the responsibility of the authors and do not necessarily represent the official views of NIH/NIDA or other participants in CJDATS.

Related Entries

▶ Differences-in-Differences in Approaches
▶ Place-Based Randomized Trials
▶ Quantile Regression Models to Analyze Experimental Data
▶ Randomized Block Designs

- ▶ Randomized Experiments in Criminology and Criminal Justice
- ▶ Sample Selection Models
- ▶ Sample Selection Problems

Recommended Reading and References

Balas EA, Austin SM, Ewigman BG, Brown GD, Mitchell JA (1995) Methods of randomized controlled clinical trials in health services research. Med Care 33(7):687–699

Bond G, Evans L, Saylers M, Williams J, Kim H (2000) Measurement of fidelity in psychiatric rehabilitation. Ment Heal Serv Res 2(2):75–87

Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P (2008) Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. Ann Intern Med 148(4):295–309

Campbell DT, Stanley JC (1966) Experimental and quasi-experimental designs for research. Houghton Mifflin, Boston

Carroll C, Patterson M, Wood S, Booth A, Rick J, Balain S (2007) A conceptual framework for implementation fidelity. Implement Sci 2:40–48

Cohen J (1988) Statistical power analysis for the social sciences, 2nd edn. Lawrence Erlbaum Associates, New York

Dennis ML, Perl HI, Huebner RB, McLellan AT (2000) Twenty-five strategies for improving the design, implementation and analysis of health services research related to alcohol and other drug abuse treatment. Addiction 95:S281–S308

Djulbegovic B, Cantor A, Clarke M (2003) The importance of the preservation of the ethical principle of equipoise in the design of clinical trials: relative impact of the methodological quality domains on the treatment effect in randomized controlled trials. Account Res 10(4):301–315

Downs S, Black N (1998) The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Commun Health 52(6):377–384

Friedmann PD, Katz EC, Rhodes AG, Taxman FS, O'Connell DJ, Frisman LK, Burdon WM, Fletcher BW, Litt MO, Clarke J (2008) Collaborative behavioral management for drug-involved parolees: rationale and design of the step'n out study. J Offender Rehabil 47(3):290–318

Inciardi JA, Surratt HL, Martin SS (2007) Developing a multimedia HIV and hepatitis intervention for drug-involved offenders reentering the community. Prison J 87(1):111–142

MacKenzie D, Browning K, SKroban S, Smith D (1999) The impact of probation on the criminal activities of offenders. J Res Crime Delinq 36(4):423–453

Meredith W, Tisak J (1990) Latent curve analysis. Psychometrika 47:47–67

Moher D, Schulz KF, Altman DG (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 357(9263):1191

National Institute on Drug Abuse (2000) The NIDA community-based outreach model: a manual to reduce the risk of HIV and other blood-borne infections in drug users. National Institute on Drug Abuse, Rockville, Report nr (NIH Pub. No. 00-4812)

Olivo SA, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ (2008) Scales to assess the quality of randomized controlled trials: a systematic review. Phys Ther 88(2):156–175

Raudenbush SW (2001) Comparing personal trajectories and drawing causal inferences from longitudinal data. Annual Reviews, Palo Alto

Weinberger M, Oddone EZ, Henderson WG, Smith DM, Huey J, GiobbieHurder A, Feussner JR (2001) Multisite randomized controlled trials in health services research: scientific challenges and operational issues. Med Care 39(6):627–634

Weisburd D (2003) Ethical practice and evaluation of interventions in crime and justice: the moral imperative for randomized trials. Eval Rev 27(3):336–354

Weisburd D, Taxman FS (2000) Developing a multicenter randomized trial in criminology: the case of HIDTA. J Quant Criminol 16(3):315–340

Multi-victims

- ▶ Multiple Victims and Super Targets

Naltrexone

▶ Drug Abuse and Alcohol Dependence Among Inmates

Naming and Shaming of Corporate Offenders

Judith van Erp
Erasmus School of Law, Criminology
Department, Rotterdam, The Netherlands

Synonyms

Disclosure; Publicity sanctions; Reputation sanctions

Overview

Pressure from the social environment of business is a major explanatory factor for corporate compliance. The threat of negative publicity, reputation damage, and social stigma may prevent business offenses or socially irresponsible or unethical business behavior. Therefore, both public and private regulatory actors attempt to activate social control by *naming and shaming* firms that offend legal or social norms. Exposing corporate misbehavior to the public may trigger stakeholders to impose "reputation sanctions" that largely exceed the impact of legal sanctions, such as fines. Public enforcement authorities therefore increasingly make use of publicity in their enforcement strategies, and disclose names of offenders in an attempt to increase the impact of regulatory sanctions. Similarly, NGOs actively "name and shame" firms involved in socially or environmentally harmful behavior to invoke extralegal pressure in the absence of strong formal legal enforcement. This entry addresses naming and shaming as a reaction to corporate crime. It discusses its form, the impact on various types of corporate offenders, and its potential and capacity to prevent corporate crime.

Naming and Shaming: Definition and Goals

A distinction should be made between naming offenders and subsequent shaming. Naming offenders can be a neutral activity when public enforcement agencies disclose names of offenders in the context of transparency about their enforcement results. In his standard work "*Crime, shame and reintegration*," the Australian criminologist John Braithwaite (1989, p. 100) defines shaming as follows:

> Shaming means all social processes of expressing disapproval which have the intention or effect of invoking remorse in the person being shamed and/or condemnation by others who become aware of the shaming.

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

Formal punishment can have a shaming effect, but shaming can also occur as an extralegal punishment: We often speak of shaming sanctions or publicity sanctions. The act of shaming can be performed both by the "naming" authority itself, or as a reaction to more neutral public disclosure of offenders by third parties such as media, consumer organizations, interest groups, or other NGOs. Because no clear demarcation line exists between disclosure of offenders' names and shaming them, it is relevant to conceptualize naming and shaming as continuing scale with more neutral disclosure intended to increase market transparency on the one end and shaming, intended to invoke a negative reaction, on the other end.

A few recent examples:

- The European Directive on public access to environmental information obliges member states to publish names of polluting companies in the European Pollutant Emission Register.
- In many countries, "scores on the doors" of restaurants show the extent to which the restaurant complies with hygiene standards.
- The British Department for Business, Innovation and Skills publicizes names of employers found in breach of minimum wage regulations.
- Financial market authorities, such as the US Security Exchange Commission, the British Financial Services Authority, and the Dutch Authority for Financial Markets, publish sanctions against offending financial market firms.
- Greenpeace campaigns for "greener electronics" by rating consumer electronics producers on an index of their performance regarding responsible for disposal of electronic waste as well as by shaming firms who fail to adequately prevent waste dumping.
- Amnesty International targets Shell for its oil spills in the Niger Delta.

Since naming and shaming comes in many different forms, a variety of goals can be identified. More or less neutral naming of offenders is often associated with the goal of transparency and accountability of regulatory agencies. Many public enforcement agencies systematically disclose offenders' names systematically in a repository of inspection reports, in a public sanction register, or in generic offenders' indices, with the aim of informing the public about their output. Enforcement decisions are visible results of market regulation, and their publication is one of the most direct ways of informing the public of the results of supervision. Transparency can also increase the democratic legitimacy and acceptance of supervision; public trust in market supervision may increase when the public and business become aware that bad practices are being addressed. Publication of sanctions may thus also contribute to the legitimacy of supervision through the creation of an image of "tough" enforcement. The purpose may also be to avoid blame in case of a disaster – publication confirms that the enforcement authority was aware of the problem and did not hide information about risks from the public.

In addition to the general purpose of government transparency, naming offenders may serve a more instrumental goal when it serves to repair information asymmetries in markets. Disclosure aims at contributing to the functioning of markets by providing consumers or other stakeholders with information that may be relevant to their choices. In the context of regulatory offenses, the goal is to enable consumers to include risks or unwanted side effects that result from regulatory noncompliance in their decisions.

Shaming adds a normative element to the publication of offenders' names. It serves to express and invoke public condemnation for the offender and the offense. Therefore, it not only involves the publication of offenders' names, but adds a negative evaluation, such as a press release announcing a regulatory fine or explaining an offense, or placing offenders on a blacklist. Shaming may also consist of more informal and incidental negative comments by public officials. President Obama's remark that BP CEO Tony Hayward "wouldn't be working for me" in the course of the Deepwater oil spill is a recent example. Shaming by NGOs may involve a whole array of advocacy strategies, such as negatively branding a product ("blood diamonds"); targeting a company (Greenpeace's attacks on Nestlé for producing Kitkat with palm oil from endangered tropical forests) to addressing the (ir)responsibility of individual CEOs.

How Does Shaming Work?

Again, a distinction should be made between consumer-oriented disclosure of names aiming at increased market transparency and shaming. In the case of disclosure in the context of market regulation, the audience may react to disclosure by avoiding the product or firm ("exit") or by voicing their concerns and exerting pressure on the company to change ("voice") (Hirschman 1970). Such a consumer-oriented disclosure policy explicitly or implicitly departs from the rational economic assumption that information allows consumers to protect themselves, instead of being protected by authorities. Moreover, the aggregate market pressure coming from consumers' individual actions can also result in third-party enforcement when they "punish" offending companies and "reward" firms with a good compliance status. It is because of this last mechanism that disclosure is considered such a powerful regulatory instrument. Fung et al. (2007, p. 6, 51) describe the process that is triggered by publication of offenses as a chain reaction:

- Information users perceive and understand newly disclosed information.
- They therefore choose safer, healthier, or better-quality goods and services.
- Firms perceive and understand users' changed choices.
- They therefore improve practices or products.
- That in turn reduces risks or improves services.

Theories about shaming generally identify two working mechanisms of shaming: deterrence and moral education (Braithwaite 1989; Ayres and Braithwaite 1992). Naming and shaming may deter firms from offending, because firms fear reputation damage in addition to the costs of the formal legal sanction. A good reputation pays out in terms of confidence of business partners and investors, consumer trust, and goodwill and support of the community and the political and bureaucratic environment. All these assets are necessary for the continuity and growth of a company (Gunningham et al. 2004). Publication of a firm's status as an offender may give rise to negative publicity that may result in reputation damage in the form of loss of clients, employees, business partners, or investors, that add up to the formal costs of sanctions. General deterrence theory has therefore expanded to include the cost of reputation damage in the cost-benefit equation (Williams and Hawkins 1986; Thornton et al. 2005).

Naming offenders may not only increase the cost of sanctions by adding the financial costs of stakeholder reactions, but also by adding nonfinancial reputation damage, such as loss of face, and being shamed in the eyes of colleagues, peers, and relatives. Convicted business people often feel the loss of their status and prestige, and the loss of trust in their authority as corporate executives, as their biggest punishment (Benson 1990). Although impossible to quantify, the threat of being publicly shamed probably deters more than financial reputation damage.

The "moral education" working mechanism in shaming is most prominently outlined in the theory of responsive regulation (Ayres and Braithwaite 1992), the paradigmatic current criminological theory of regulation of corporate compliance. Responsive regulation regards corporate compliance not primarily as a result of fear of legal sanctions, but as a result of a combination of an intrinsic motivation to behave responsibly, and external social pressures, such as fear of negative publicity, reputation damage, and, eventually, fear of legal sanctions. Responsive regulation aims to improve business compliance by connecting with firms' perceptions of appropriate and responsible behavior (Ayres and Braithwaite 1992). Firms are initially addressed with a cooperative, persuasive strategy, and only when firms do not respond, a regulator may escalate with a variety of interventions. A responsive regulatory approach is therefore visualized as a pyramid. The majority of regulatory action takes place at the bottom of the pyramid, where regulators use a variety of cooperative strategies, and enable third-party action. Pressure for compliance does not primarily come from legal sanctions, but from advocacy groups as watchdogs and informal regulators.

Adverse publicity is important as a catalyst for third-party action. But in addition to its function as a deterrent, shaming is essential in responsive regulation theory to communicate the harmfulness of corporate offenses and to strengthen communities' shared expectations about appropriate business behavior. Public exposure of the offender underlines the unacceptability of the conduct and is expected to evoke the normative disapproval of the general public or significant peers (Braithwaite 1989; Parker 2006). Adverse publicity may thus contribute to public awareness of the harmfulness of corporate crime and foster normative attitudes against corporate crimes both in the general public and business communities (Braithwaite 1989). Shaming is also less likely to result in neutralizing calculating behavior, and more likely to lead to a discussion about moral or normative aspects of business behavior. A higher awareness of the societal and environmental impact of business behavior may also increase firms' willingness to invest in "beyond-compliance" behavior.

Whereas formal punishment communicates that legal norms have been breached, shaming communicates the message that social or moral norms have been offended. It is an expressive instrument that works through emotions such as the wish to be respected and the fear of being humiliated (Murphy and Harris 2007), and addresses consciences by expressing to the offender that he did not take his responsibilities toward the community.

Shaming, however, can easily result in humiliation and stigmatization of the offender. Reintegrative shaming theory predicts that stigmatization will make crime worse, whereas reintegrative shaming will prevent it (Braithwaite 1989). Reintegrative shaming means that public disapproval of the offense is combined with communicating trust in the offender's willingness to improve (Makkai and Braithwaite 1994). The offender is addressed as a good person who committed a bad act. Shaming directed against the offender, instead of inviting him to show his better self, will be experienced as stigmatizing. Stigmatizing will result in defiance and resistance of the offender and often also of the business community in general (Braithwaite 1989). The offender will refuse to acknowledge shame he feels inside or attempt to reason away his part in the offense or to neutralize the offense (Harris 2001; Braithwaite and Drahos 2002; Braithwaite et al. 2006; Murphy and Harris 2007).

Empirical research in responsive regulation has shown that offenders will only perceive shaming as reintegrative when performed by direct and respected peers. Shaming by officials is often experienced as stigmatizing, even when officials do not intend this (Makkai and Braithwaite 1994; Ahmed et al. 2001). Effective shaming therefore works best in a normative community of conscience between regulators, third parties, and business executives. In the case of white collar crime, however, consensus about the harmfulness of offenses, or a community of conscience formed by regulators, interest groups, and businesses will frequently be lacking (Parker 2006). Therefore, governments or NGOs should engage in formal, but respectful shaming, encouraging business actors to become reengaged in the substantial goals of the law. Although these formal shaming statements will unavoidably be experienced as stigmatizing to some extent, chances are that they will be translated in reintegrative shaming by peers – friends, family members, or colleagues or, internal compliance departments who have the capacity for reintegrative shaming.

In sum, shaming may work as a market incentive, as a deterrent "big stick" and as a form of moral education. Empirical research has shown that formal legal sanctions such as fines often fail to fulfill these functions. Naming and shaming therefore appears to compensate for some of the shortcomings of "traditional" legal financial sanctions. It is therefore often expected that naming and shaming is more effective in promoting compliance. What empirical evidence is available to support this claim?

Is Shaming Effective?

Several factors complicate the answer to the question whether shaming is effective in combating corporate crime. The first is the interdisciplinary character of research. The question of shaming addresses the interplay between business, NGOs, media, regulatory and enforcement authorities, and consumers, shareholders, and other stakeholders. An understanding of this interplay requires contributions from a variety of disciplines. Research into the effects of shaming is conducted within management science, economic science, law and society scholarship, empirical legal studies, social psychology, corporate governance research, business law, regulatory governance research, criminology, and media studies. While all these approaches provide valuable insights about effects of shaming, they also each depart from a particular theoretical perspective and study different types of shaming, complicating the aggregation and integration of research results.

A second complication in effect studies is the measurement of effects and causality. Concepts such as "reputation" and "shame" are intangible values which are difficult to measure. Several studies have attempted to solve this problem by measuring quantifiable consumer or shareholder reactions to publicized sanctions, such as fluctuations in sales, stock rates, or revenue after publication of offenses. Also, corporate reputation indices or media monitors may generate sophisticated indicators of corporate reputations following scandals. But although these measures may generate insight into the height of reputation sanctions, they do not answer the key causal question how the threat of reputation damage prevents corporate crime. The preventive, deterrent, or educative effects of naming and shaming depend on perceptions. Qualitative case-studies or interviews may give more in-depth explanations of the relation between naming and shaming and compliance, but the findings of these studies are context-dependent.

Given these limitations, a selection of important results from different disciplines may be highlighted. They shed some light on the effects and effectiveness of shaming.

A first line of research addresses the assumption that publication of offenses creates a market incentive by generating consumer reactions. Behavioral economic research and cognitive psychology have shown that markets where consumers act as rational information-processing "econs" are rare. Consumers struggle with an overload of information and make decisions under influence of cognitive biases. In particular, they value the status quo, meaning that they are loyal to their existing choices even if negative information about a firm becomes available. In line with the changed perspective on consumer decision making, several empirical studies of regulatory disclosure initiatives show that most of them did not evoke strong or consistent consumer reactions (Fung et al. 2007; Weil et al. 2006; Pawson 2002; Van Erp 2010; Kraft et al. 2011). Information about offending firms or unsafe products plays a less prominent role in consumer decision making than a consumer-oriented disclosure logic expects. Consumers do not actively gather information from regulatory agencies' websites; fail to correctly understand the information; and, most importantly, do not act upon this information. They often remain loyal to firms despite their negative performance. In addition, firm's legal compliance or even its socially responsible behavior often does not play a prominent role in consumer choices. This is not to say that naming offenders has no effect on business behavior, but the incentive generally does not come from consumer reactions to disclosed inspection results directly.

Two types of circumstances are suggested that alter this equation. The first is the type of stakeholder. It is often expected that shareholders are more active in seeking and processing information and act more rationally than consumers. Disclosed offenses of companies traded on the stock exchange are therefore expected to generate severe reputation sanctions in the shape of stock rate devaluation. Stock reactions to disclosed corporate offenses are also one of the most concrete indicators for actual reputation damage. Several research projects have therefore measured stock rate reactions to (for an overview, Van Erp 2011a). Research by Karpoff, Lee, and Martin (2008) shows that the disclosure of financial misrepresentation can lead to reputation sanctions that multiply the size of legal sanctions. Although part of this loss can be explained as a correction for the misrepresentation of the value of the company, or anticipation of costs of legal procedures and fines, the majority of this loss is reputation loss. Comparable studies in other regulatory arenas, however, find much lower stock rate reactions than those for financial misrepresentation. The publication of environmental offenses also leads to a drop in stock value, but this loss can be entirely accounted for by the costs of legal fines and restoration of the damage (Karpoff et al. 2005). There is no additional reputation sanction, because stockholders do not value the company less for its failure to comply with environmental regulations. Karpoff et al. conclude that investors will sooner impose a reputation sanction for an offense that damaged the interests of a firms' contract partner, than for an offense that damages a third party or an intangible value such as the environment.

Second, active negative publicity, generated by public advocacy campaigns by NGOs, may trigger larger reputation sanctions than public disclosure of sanctions alone. Kagan, Gunningham, and Thornton (2011) summarize the combination of pressures on companies by describing how companies recognize three "licenses to operate": the economic license (the obligation to meet the expectations of financial stakeholders); the legal license (compliance to law and permit conditions); and the social license: expectations from neighbors, employees, community groups, news media, and advocacy groups. Their research on pulp mills showed that firm's environmental performance did not vary in relation to differences in their legal or economic license, but in relation to social license pressure. Naming and shaming corporate irresponsible actors may create large pressures on corporations to put an end to pollution and to improve their environmental policies. Indeed, many examples exist about successful campaigns that have forced corporations to withdraw from profitable, but damaging activities or even beyond-compliance behavior where companies voluntary exceed legal norms.

Adverse publicity may also have an implicit preventative effect in the sense that media exposure increases the visibility and public awareness of the damage of harmful corporate activities; thus increasing the legitimacy of public enforcement and creating support for new legislation. It may contribute to the development of self-regulation such as certification schemes and codes of conduct. In the absence of public enforcement, voluntary compliance of businesses to "soft law" can decrease damage to the environment and violations of human rights of laborers and local communities. This may provide an opportunity particularly for issues that exceed jurisdictions of individual countries, such as environmental harm, deforestation, and other extinctions of natural resources, or globalized production chains where Western producers produce goods in countries where public enforcement against businesses and legal protection of employees are weak. But is this more than a drop in the bucket? In an analysis of corporate social responsibility initiatives for working conditions in developing countries, for the environment, and for human rights; David Vogel (2005) concludes that the "market for virtue" is small. There is no business case for corporate social responsibility: Simply because consumers are not willing to pay higher prices and only a small proportion of investors are interested in CSR. For example, few industries have been as heavily shunned in the media as the tobacco industry; yet, investment in this industry has flourished and the stock rates of tobacco firms have outperformed other stock. Vogel therefore concludes that informal naming and shaming by NGOs cannot take the place of legally binding public law and strong enforcement.

It can be added that only a proportion of corporate offenses generate negative publicity because not all form attractive news stories. Likewise, a market reaction to negative publicity may not always occur. Firms invisible to the public and not known to end-consumers are unlikely to experience reputation sanctions, if they do not comply (Thornton et al. 2009). Likewise, companies facing bankruptcy, and monopoly actors do not depend on a good reputation. Finally, even when publicity causes significant reputational damage, this may not translate into better compliance because of large necessary investments in production technology or simply a lack of responsiveness of businesses (Kraft et al. 2011).

A last and entirely different line of research discussed here does not directly address firm's reactions, but regulatory authorities' shaming strategies and corporate reactions. Researchers in responsive regulation ask the question to what extent enforcement authorities are capable of powerful, yet reintegrative naming and shaming, as is expected in responsive regulation. Sometimes, regulators that lack powerful enforcement tools attempt to actively invoke negative publicity and publicly shame corporate officials by portraying them as irresponsible, greedy, or criminal. Case-studies have demonstrated that an active shaming strategy requires broad political and societal consensus about the harmfulness of the condemned behavior, and unquestioned authority of the regulatory agency as the messenger. Conversely, when regulators lack support and authority, their messages will invoke resistance and undermine business' intrinsic motivations for compliance in the long run. They will also undermine the political support for the regulatory agency itself and result in attacks on its political legitimacy by the business community (Parker 2006).

It should be added that corporate offenders do not passively undergo adverse publicity, but actively attempt to manage publicity about their involvement in offenses. The ambiguous nature of corporate crime presents an opportunity for corporate officials to deny its harm by presenting their side of the story and to neutralize the severity of the offense or their responsibility (Braithwaite and Drahos 2002; Levi 2006). Defendants may even try to present themselves as victims of abusive regulators or prosecutors. For the media, initial reports of corporate crime may be salient news facts, but the newsworthiness of disclosed inspection results may decline over the course of the years, as happened in the 25-year existence of the American Environmental Protection Agency's toxic release inventory (Kraft et al. 2011).

This illustrates the crucial role of the media in presenting the shaming message that regulatory actors attempt to broadcast. Media are no neutral intermediaries, reproducing the news as it is presented by legal authorities. By selecting, interpreting, and framing the news, they may reinforce the moral message that a punishing authority sends, but may also detach it of its shaming component. Thus, publicity can raise doubts about the rightfulness of enforcement and either remove the blame from, or even create sympathy for, business. Also, authorities' accusations of businesses may lead to further questioning of the supervision practices that enabled business offenses to occur in the first place, resulting in the shaming to backfire to the enforcement agency (Almond 2009).

Last, enforcement authorities fear that adverse publicity will lead to disproportional reputation damage for firms that may damage the interests of employees and other firms or even the economy or financial system stability in the case of corporations "too big to fail." Not surprisingly, then, many public regulators wish to avoid conflict, and revert to "naming without shaming": formalistic, technical, and generic offender's indexes or sanction registers using legal language rather than explicitly condemning corporate malpractice (Van Erp 2011b). As a result, shaming by enforcement agencies often does not contain the moral message, nor results in the negative publicity and reputation sanctions that would be required to trigger a broader debate about the harmfulness of the underlying behavior.

Future Developments and Directions for Research

The broad phrase "naming and shaming" may refer to a wide array of activities, aims and potential effects, and target groups addressed. Any naming and shaming policy or research should start by specifying the audience addressed and effects that are expected. Shaming may work as a market incentive, as a deterrent "big stick" and as a form of moral education. Often, these effects interact and reinforce each other. The impact of naming and shaming therefore often largely exceeds that of formal legal sanctions, and can be more effective in promoting compliance than "regular" enforcement. However, the effects of shaming cannot easily be predicted. The type of business and market; the existing reputation of the offender; the type of offense; the authority and legitimacy of the shaming actor; and the media all influence the amount and intensity of adverse publicity. It is clear that the process in which naming offenders and attempts to shame, translate to financial or immaterial reputation damage, and subsequently improves compliance of the offender and other firms, is a very complex process that is difficult to control. Case-studies of successful and failed shaming campaigns of both formal and informal regulators; media analyses of the nature of the media representation of corporate crime; and quantitative studies of the impact of disclosed offenses on business performance or compliance rates are necessary research strategies to contribute to theoretical understanding of the relations between naming, shaming, and corporate crime.

A few research topics in particular are relevant to mention here. First, a large part of enforcement against corporate offenses is performed by administrative regulatory agencies rather than criminal prosecutors. Disclosure of inspection results or sanctions in sanction registers, offenders indices or performance inventories, is more and more common for these administrative agencies in their quest for transparency and accountability. This may change the character of administrative enforcement, which is generally thought to generate less media attention and therefore to result in less stigma, less reputation damage, and less moral condemnation of white collar offenders. The increased powers of disclosure by administrative authorities may shift the balance between criminal and administrative enforcement regimes, because a publicized fine may result in more negative publicity than a criminal verdict, which is often anonymous. One of the few studies in this area reveals unexpected differences in press coverage of cartels in Germany. In Germany, cartel offenses of firms are regulated by federal administrative law, whereas cartel offenses of individuals are enforced under criminal law. The administrative cartel authority issues press releases, naming offending companies. Public prosecutors however do not systematically issue press releases and do not reveal the names of perpetrators. Accordingly, press coverage about bid-rigging prosecutions and convictions is much sparser than press coverage of administrative fines in Germany. “This is the opposite of the effect one would usually expect – that the increased saliency of criminal cases would lead to wider reporting” (Wagner 2011).

The effects of the rise of new and social media on corporate reputations are another important topic for future research. The impact of shaming is already highly dependent on crowd dynamics, but the new media may increase the unpredictability of the reactions to public disclosure of offenders’ names even more. Publication of offenses on the internet results in long-lasting blame for small offenders, whereas more professional or strategic firms may replace negative news by a proactive public relations strategy.

Intensified disclosure also increases the risk that firms manipulate performance scores as these become increasingly important. In other words, disclosure intended to prevent fraud, at the same time, creates new opportunities for fraud. It should also be taken into account that media are corporate actors themselves, who may be connected to corporations involved in illegal activities. Business interests may influence the way in which corporate crime is being reported. Corporate involvement in crimes could be less likely to generate negative publicity when the corporations it concerns are active advertisers on the news networks that report the crimes.

Last, the topic of shaming illustrates that state control of corporate and white collar crime does not operate in isolation from social control. In order to understand how external pressures are translated in firm’s actions, a multidisciplinary approach is required which combines criminology with governance studies and management science.

Related Entries

▶ Compliance and Corporate Crime Control
▶ Corporate Crime Decision-Making
▶ Crimes of the Powerful
▶ Shame in Criminological Theory
▶ Victims of Corporate Crime

Recommended Reading and References

Ahmed E, Harris N, Braithwaite J, Braithwaite V (2001) Shame management through reintegration. Cambridge University Press, Cambridge

Almond P (2009) The dangers of hanging baskets: 'regulatory myths' and media representations of health and safety regulation. J Law Soc 36:352–375

Ayres I, Braithwaite J (1992) Responsive regulation: transcending the deregulation debate. Oxford University Press, Oxford

Benson ML (1990) Emotions and adjudication: status degradation among white-collar criminals. Just Q 7:515–528

Braithwaite J (1989) Crime, shame and reintegration. Cambridge University Press, New York

Braithwaite J, Drahos P (2002) Zero tolerance, naming and shaming: is there a case for it with crimes of the powerful? Aust NZ J Criminol 35:269–288

Braithwaite J, Ahmed E, Braithwaite V (2006) Shame, restorative justice, and crime. In: Cullen F, Wright J, Blevins K (eds) Taking stock: the status of criminological theory – advances in criminological theory. Transaction Publishers, New Brunswick, pp 397–417

Fisse B, Braithwaite J (1983) The impact of publicity on corporate offenders. State University of New York Press, Albany

Fung A, Graham M, Weil D (2007) Full disclosure: the perils and promise of transparency. Cambridge University Press, Cambridge

Gunningham N, Kagan R, Thornton D (2004) Social license and environmental protection: why businesses go beyond compliance. Law Soc Inq 29:307–341

Harris N (2001) Shaming and shame: regulating drink-driving. In: Ahmed E, Harris N, Braithwaite J, Braithwaite V (eds) Shame management through reintegration. Cambridge University Press, Cambridge, pp 73–210

Hirschman A (1970) Exit, voice, and loyalty: responses to decline in firms, organizations and states. Harvard University Press, Cambridge

Kagan R, Gunningham N, Thornton D (2011) Fear, duty, and regulatory compliance: lessons from three research projects. In: Parker C, Lehman Nielsen V (eds) Explaining compliance. Business responses to regulation. Edward Elgar, Cheltenham, pp 37–58

Karpoff J, Lott J, Wehrly E (2005) The reputation penalties for environmental violations: empirical evidence. J Law Econ 48:653–675

Karpoff J, Lee D, Martin G (2008) The cost to firms of cooking the books. J Financ Quant Anal 43:581–612

Kraft ME, Stephan M, Abel T (2011) Coming clean, information disclosure and environmental performance. MIT Press, Cambridge

Levi M (2006) The media construction of financial white-collar crimes. Br J Criminol 46:1037–1057

Makkai T, Braithwaite J (1994) Reintegrative shaming and compliance with regulatory standards. Criminology 32:361–385

Murphy K, Harris N (2007) Shaming, shame, and recidivism: a test of reintegrative shaming theory in the white-collar crime context. Br J Criminol 47:900–917

Parker C (2006) The compliance trap: the moral message in responsive regulatory enforcement. Law Soc Rev 40:591–622

Pawson R (2002) Evidence and policy in naming and shaming. Policy Stud 23:211–230

Thornton D, Gunningham N, Kagan R (2005) General deterrence and corporate environmental behavior. Law & Policy 27:262–288

Thornton D, Kagan R, Gunningham N (2009) When social norms and pressures are not enough: environmental performance in the trucking industry. Law Soc Rev 43:405–436

van Erp J (2010) Regulatory disclosure of offending companies in the Dutch financial market: consumer protection or enforcement publicity? Law & Policy 32:407–433

van Erp J (2011a) Naming and shaming in regulatory enforcement. In: Parker C, Lehman Nielsen V (eds) Explaining compliance. Business responses to regulation. Edward Elgar, Cheltenham, pp 322–343

van Erp J (2011b) Naming without shaming: the publication of sanctions in the Dutch financial market. Regul Gov 5(3):287–308

Vogel D (2005) The market for virtue: the potential and limits of corporate social responsibility. Brookings Institution Press, Washington

Wagner-Von Papp F (2011) What if all bid riggers went to prison and nobody noticed? In: Beaton Wells C, Ezrachi A (eds) Criminalising cartels, critical studies of an international regulatory movement. Hart, Oxford, pp 157–182

Weil D, Fung A, Graham M, Fagotto E (2006) The effectiveness of regulatory disclosure policies. J Policy Anal Manage 25:155–181

Williams K, Hawkins R (1986) Perceptional research on general deterrence: a critical review. Law Soc Rev 20(4):545–572

National and Local Policing

Barry Loveday
Institute of Criminal Justice Studies, University of Portsmouth, Portsmouth, UK

Overview

This entry considers the differing structures of policing particularly those pertaining to the Anglo-American and European policing tradition. It argues that both national and local experience of policing should be viewed as reflecting the differing purposes of the police within which protecting the state has preeminence within national systems and where protection of the community takes precedence within local policing arrangements. It looks in detail at the experience of national police in France and contrasts this with the policing traditions that have characterized the development of local policing in both America and the United Kingdom.

Key Controversies

There has been an ongoing debate among both practitioner and academic commentators as to the advantages and drawbacks of two competing models of police organization. It is fair to say that until recently, the commitment to local policing was one that very much reflected the Anglo-American approach to devolved police structures that were closely linked to systems of local government (Marshall 1965). On the other hand, the national and more centralized police model was linked to mainstream European models of policing where a highly centralized state exercised clear hegemony over national policing (Mawby 1990: see Mawby this volume). Indeed, historically, the division between the two very different types of policing models may have also reflected a much more refined definition of what constituted the primary role of the police in different states (Stead quoted in Loveday 1999).

Thus within many states in Europe, traditionally the central role of the police was – and remains – the protection of the state. Certainly within France which has remained committed to national policing over many centuries, the police system as constituted in the late seventeenth century exhibited a "state protection responsibility" which over the years has remained a crucial function for the national police and the gendarmerie. The Ministry of the Interior and the Minister of Defense continue to exercise immediate responsibility for all policing in urban and rural areas within France (Mouhanna 2011). Historically a very similar situation pertained in Spain where the public order function of the police gave it a paramilitary profile with state and public protection providing the overriding raison d'être for the paramilitary Guardia Civil (Mawby 1990).

In contrast, the Anglo-American approach was to place much more emphasis on local delivery, linked closely to local government (Mawby 1999). This system was to demonstrate its most extreme form in the United States where local policing was to prove to be, in its earliest manifestation, overwhelmingly local in terms of structure, recruitment, and direction (Conley 1995). But early nineteenth-century developments in policing in America only reflected to a greater degree the creation of the local policing system in the United Kingdom.

Following the Municipal Corporation Act of 1835 and the County and Borough Police Act of 1856, local authorities were encouraged [but not at this time required] to establish local police forces. The Municipal Corporations Act of 1835 led to the creation of municipal police forces based on local authority boundaries which were to survive until well into the mid-twentieth century. Locally financed and locally led, the municipal police were in effect under the direction and control of local Watch Committees made up primarily of locally elected councillors (Marshall 1965). In effect, outside London, the police system reflected a patchwork of local forces exercising duties that would be closely monitored by locally elected representatives [albeit based on a very limited franchise]. The primary point remains, however, that the "protection of the state" and centralized control of the police never appeared as a primary function or organizational characteristic of policing in either the UK or the US. Even in London where the Metropolitan Police was made accountable to the Home Secretary, there was to be little interference in policing which was made the immediate responsibility of the Metropolitan Police Commissioner (Marshall 1965).

In America the relationship between the local government and the police force was to prove to be even stronger both from the inception of the police in the early nineteenth century and later. Arguably this very close relationship was to provide the basis for later claims that the police were heavily politicized, a criticism that continues to be raised against local policing models today. It is fair to say that given the strength of local political engagement in America, it was always unlikely that any local public service could expect to evade the tentacles of local politics. Yet, as will be argued later, the development and professionalization of policing in both the US and UK was to create situations where police forces – particularly the bigger urban forces – exhibited a degree of professional autonomy that may have generated a significant democratic deficit (Marshall 1978).

Key Issues: National and Local Police Systems

While the historical background to the division of police models remains of interest, it is also the case that policing structures remain a matter of continuing concern to both public policy makers and chief police officials. This may be linked to either party political processes or the perception that structural change can be expected to improve police efficiency and effectiveness. Some current, and very contemporary, examples of this development can be identified within two distinct police jurisdictions.

In Scotland, for example, the decision has been made within the Scottish National Party that there are significant advantages to be gained from the amalgamation of the existing eight police forces to create one national police force for Scotland. Plans are now well advanced to initiate this major structural reform of policing which could provide a significant indicator of what benefits accrue from a "national police force" within mainland UK where traditionally local structures have been seen as a defense of policing. Interestingly one justification used by the Scottish assembly in defense of nationalization of the police has proved to be that other European states have embarked on a very similar pathway (Fyfe 2012). In Norway, Denmark, the Netherlands, and Finland, the respective governments have recently decided to move from local to national police systems.

Indeed within the Netherlands, a country which in the past has challenged the European stereotype of centralized policing (Jones 1995; Mawby 1990), the political process has resulted in a decision to effectively scrap the local policing system which differentiated it from mainstream Europe. It has instead opted for the introduction of a national policing system. Based on a perception that a national police model would prove both more efficient and less costly, the myriad local police forces that characterized the Netherlands will form part of a national Netherlands police service. Interestingly, in both cases the primary consequence of the reforms planned for the police service will be a change in police governance away from local government to national government within which the single chief officer will be made either directly or indirectly accountable and answerable to a national assembly which could prove to be highly political.

Developments in both Scotland and the Netherlands provide a remarkable contrast on the other hand to developments in England and Wales. Here the Coalition government within the 2011 Police Reform and Social Responsibility Act intends to move to strengthen local police systems by radically reforming local police governance to provide much greater local accountability of the police to their local communities. The 2011 Police Reform Act represents the most significant recommitment on the part of national government to local policing and local police governance (Loveday 2012). To the extent that the current Coalition government in England and Wales remains committed to entrenching the local policing dimension, this policy development reflects a common law Anglo-American approach to policing which highlights local police delivery along with a corresponding commitment to "low" rather than "high" policing.

Indeed the division of policing into "high" and "low" policing styles has accurately reflected the differing approaches to policing responsibilities.

Traditionally "high" policing was [and remains] closely associated with European systems where immediate police responsibilities centered upon the protection of the state. Within the Anglo-American approach far less importance is attached to state protection and far more is accorded to protecting the community. This may well reflect the common law basis upon which policing in England was established and upon which the "new" policing model of the early nineteenth century was based. Within the 1829 Act creating the London Metropolitan Police, the Home Secretary Robert Peel made much of the police being "citizens in uniform" and where it was an offense for a citizen not to assist a police officer in his duties if called upon to do so. While therefore the high policing model could remove the police from any immediate responsibility to the community, in contrast "low" policing aimed to cement the police system within the community.

International Perspectives: Typologies of Policing

Typologies of policing have been expanded to include a range of policing systems which provide for a better comparative evaluation of the structure and effectiveness of those systems (Mawby 1990; see Mawby this volume). While national and local police structures provide a description of police systems, there are a range of models that can be applied to current police jurisdictions and which go beyond this. One of the most useful typologies was developed by Stead in his influential study of the French police in the 1950s, which continues to have a ready application (Loveday 1999).

Stead was to develop three typologies of police systems which he classified as "fragmented," "combined," and "national centralized" police systems. In developing these typologies, Stead also provided in effect a heuristic device in approaching the analysis of policing systems. Thus, it might be argued that using these typologies, it is relatively easy to identify examples of each of the typologies developed by Stead. Traditionally the police systems of the USA and Britain would have been classified respectively as examples of fragmented and combined police systems: in the USA a large number of police forces operated in relative autonomy, whereas in the UK local and central government were made jointly responsible for the government of the police. Similarly France has traditionally been identified as a clear example of a national centralized police system fully established with the creation of the Police Nationale.

One evident feature of government of the police remains the degree to which a police force can be made accountable for what is does [or doesn't do]. This is not an esoteric matter as accountability can be closely linked, as will be argued later, to issues surrounding police effectiveness and police efficiency. However, it might be noted that each typology identified by Stead demonstrates a different form of accountability. Thus, national systems are theoretically made accountable at national political level usually by way of ministerial responsibility to a national parliament of assembly. In France, for example, the Minister of the Interior has traditionally been identified as the senior minister responsible for the National Police. Similarly the Minister of Defence has immediate responsibility for the gendarmerie, responsible for all policing outside of the larger urban areas.

Within fragmented police systems, the accountability mechanism is in contrast much more local and diffuse. Here as in the USA police forces will be viewed as being immediately accountable to the local electorate. As if to reinforce this within some jurisdictions, both the police chief and the local mayor may stand for election (Loveday and Reid 2003). Here, as with many police forces in the USA, the police will be viewed as being immediately accountable to the local electorate, as the sheriff or chief officer of police may well be an elected official or may hold the office of chief officer which is the gift of the locally elected mayor (Loveday and Reid 2003). In between the national centralized and fragmented models of policing lies the "combined" system where a mix of local and central government is designed to maximize the accountability and responsiveness of the police to the public interest.

Yet it is also the case that each model policing and police system may not provide the degree of accountability which the public may see as being appropriate. National police systems may have been the responsibility of ministers who as in France have shown until recently a strong desire to protect police interests from effective scrutiny by the national assembly (Loveday 1999). On the other hand, local fragmented police systems as demonstrated in America can be subject to local partisan interests, a feature which as Conley notes can be seen as being a reflection of the early development of police forces in that country. Fragmented systems may also be susceptible to central-federal government interference particularly where local units of policing are resource poor and may require assistance from external police forces or federal agencies.

Within "combined" systems, the very division of responsibility between local and central government may over time have allowed for the exercise of power without responsibility by central government. Given the interesting lacuna which may surround each typology of policing, it may be of use to explore examples of each typology and following on from that attempt to assess the overall accountability and effectiveness of each model of policing.

A National Police System

The recent decision on the part of a number of European states which had earlier supported local policing models to move to integrated national police systems suggests that some evaluation of existing national police structures would be of some value. In relation to this the best example of a state-led and national centralized police system would be the National Police of France.

Certainly the French policing system has traditionally been cited as the most extreme form of centralized police activity in Europe, where the collective interests of the state have often appeared to have overshadowed the rights of citizens. Over the years observers have often concluded that given its origin and raison d'être, the national police has been more concerned with protecting the interest of the state than with the rights of the individual (Loveday 1999). However, the role of the national police may provide some evidence of the potential gains and hazards which may be associated with a national police system.

While allowance must be made for differing domestic political cultures, it is evident that the history of the national police is not one that appears to place high value on either local accountability or responsiveness to the community. In the late 1950s some evaluation of the role of the national police was to be made by one commentator who argued in relation to the National Police of France that it was "not too sweeping a judgement to say that the French police system could be more easily fitted than that of any other Western countries into the pattern of totalitarian government. The political activities and the building up of dossiers of persons unconnected with crime which form so large a part of the work of the police of totalitarian countries has been among the activities of the French police since the days of the first Napoleon" (Coatman 1959:77–8 quoted in Loveday 1999).

Moreover, along with the highly dirigiste nature of policing in France has gone a willing acceptance of a very wide discretion accorded to the police in the exercise of their powers. As has been noted elsewhere, for example, "The French police are allowed a latitude – license would not be too strong a word, in their dealings not only with suspected and arrested persons which is probably unique in Western democratic countries" (Coatman 1959:75 quoted in Loveday 1999).

Despite police reforms, it remains the case that the national police continue to have, for example, uninterrupted access to suspects for the first 48 h of their arrest and when suspects may be at their most vulnerable. Under the system of "garde à vue," suspects for this period are denied access to legal representation. Within local fragmented policing systems where the potential danger of police abuse of their authority may arise, similar police freedom has been drastically curtailed by way of either statute or rulings handed down from the higher courts.

The lack of effective accountability at either local or central government level has been evidenced over the years and not the least of these proved to be the policing of an Algerian protest march through Paris in October 1961 which was to lead to numerous fatalities among the many Algerian protestors (Rudorff 1970:55). There was never to be any kind of investigation as to how those fatalities at the hands of the police had arisen. This proved to be largely the result of a decision made by the Minister of the Interior at that time who also intervened to ensure that there was a blanket ban on any media coverage of the police attacks during the course of the protest march (Rudorff 1970).

Some of the worst examples of police abuse of their authority arose while under the leadership of Maurice Papon, chief of Paris police, at this time. One commentator provides an interesting insight into the potential dangers of overt politicization of the police and the threat this can present to the rule of law, which can arise in relation to national police systems. Reflecting on the police violence exhibited at this time and supported by the French government, Rudorff notes that "No matter what crimes were committed by the police for whose behaviour Papon was responsible, he was untouchable and even after the end of the Algerian war he remained in office always protected by the government, never disavowed and never out of favour" (Rudorff 1970:55).

Yet while the national policing system of France is often identified as a model of national policing, it has never been a unified national police system. This is because of the dual system of policing within which two central departments exercise separate responsibilities for the national police and the gendarmerie. Interestingly, as is also to be found within the Italian national police system, the division of responsibilities between the two police forces has led to continuing conflicts between them over time (Gleizal quoted in Loveday 1999). There may be some suspicion that institutionalized conflict has been a characteristic which may have been encouraged by political elites in both Italy and France, to undermine the autonomy of national police forces and limit their potential influence.

The level of institutional dysfunction exhibited within the French national police system was to be best evidenced in the Jobic affair in 1987 when officers of the gendarmerie arrested a very senior officer of the National Police in Paris (Guyomarch 1991). This arrest served only to exacerbate an already deep antipathy between the two national police forces which was to be reflected in a continuing lack of cooperation between them. Indeed, it is evident that rather than seeking to allay such tension, national politicians have at times exploited it to encourage greater police responsiveness. In 1987, for example, the Minister of Defence was to allow the gendarmerie to operate in plain clothes. In the early 1980s the national police was to be publicly humiliated by the decision of the then President Mitterand to transfer responsibility for his personal protection to the gendarmerie (Guyomarch 1991:325).

While the institutionalized tension between the two forces has retarded the development of any unified professional policing strategy, it may, ironically, have encouraged greater political accountability of both services. This in itself might be seen as no small achievement where in the recent past the police had become "too powerful to criticize and where national politicians feared their hostility and whose support had to be courted" (Rudorff 1970:47). While it may exhibit characteristics that are specific to that country, it is however evident that there are potential dangers which may arise from national police systems.

These would include the danger of overt politicization at a national level as the police are seen as an extension of government rather than being independent of it along with the danger of operational dysfunctionality as the police role and mission is potentially subsumed within the realm of national politics. In effect the primary function of the police is reinforced within this political environment and becomes "public protection and public order," a role which in

France has been clearly evidenced in the policing style of the Compagnies Républicaines de Sécurité [CRS]. The public order role which has traditionally fallen to national police systems as a primary function is indeed best demonstrated in France by the CRS.

As has been noted by one commentator, the mobile units of the CRS were created and placed under the explicit command of the Minister of the Interior. While these units had to deal with natural disasters [floods, forest fires], their primary function was to provide "a weapon against strikers, rioters and political demonstrations which got out of hand or were judged to be dangerous by governments which made increasing use of this force in political affairs" (Rudorff 1970:49). A highly confrontational policing style developed by both the CRS and elements of the police national came at the expense of alienating the public and also of developing any form of policing by consent. This is a traditional policing advantage pursued by local police forces, not least because of the real and potential benefits to increasing police effectiveness which this approach can encourage.

The potential problems generated in France by way of national policing have been recently addressed by Body-Gendrot. In her analysis of policing Paris, she notes that rather than giving citizens a sense of safety with community policing, for the CRS maintaining order during street demonstrations and pursuing organized crime have long been seen as the national police mission (Body-Gendrot 2012:84). The same commentator also finds that while the national police have over recent years pursued a strategy of "proximity policing," police officers involved in this strategy are not "policing by consent or pursuing a public reassurance function" either. Rather they are merely required to be "visible," and in problematic suburbs, there is often a glaring difference between "proximity policing" during daylight hours and more brutal methods used by police forces at night or during street demonstrations (Body-Gendrot 2012:85). Significantly, some commentators, such as Mouhanna (2011), have concluded that the apparently more militaristic gendarmerie in fact has traditionally had a stronger community orientation than the national police.

Moreover, as with any established professional policing model, most officers within the national police will not be available for public patrol work as they can be expected to be engaged in "intelligence or investigation work" analyzing "cyber crimes, underworld and organised crime" (Body-Gendrot 2012:84). This commitment to specialization which may be encouraged within national police systems only provides further evidence that the larger the unit of policing, the more difficult it can become for the police organization to fulfill its most basic functions which engage it with the public and community (Loveday 1999). This is because the level and degree of specialization which is encouraged by the very number of police officers available siphons off police resources from the public domain to police professional priorities that may not be shared by the public. It is also the case that within the French National Police, a lack of leadership, the inertia of bureaucracy, and the limited accountability exhibited by the police together have served to weaken police efficiency.

It is interesting to note that in the absence of effective accountability, the national police remain committed, within the capital, to a more basic and highly political function. As Body-Gendrot succinctly argues, as the core of Paris remains a defensible space with numerous official buildings, public and private mansions, palaces, and banks, the police are required to protect them as well as provide for the safety of state VIPs. She adds that "Whenever it is anticipated that disorders could occur a strong paramilitary police deployment is placed in charge and made responsible for law and order" (Body-Gendrot 2012:87).

Nor is the paramilitary response to disorder by the police national confined to Paris. Most recently in August 2012, in response to disturbances in Amiens in Northern France, the police were to use rubber bullets and teargas "which some residents considered excessive" (Chrisafis 2012). It appeared that the rioting was sparked by resentment over "spot" checks conducted on residents by the police and also on a highly negative relationship between the police and some sections of the local community.

In response, President Hollande was to state that his government would mobilize all of its resources to combat the violence and that his priority was "security." It meant that the next budget would include "additional resources for both the gendarmerie and the police" (Chrisafis 2012). It is evident that in the situation that confronts the police in many French banlieues, increases in police resources alone are unlikely to resolve the underlying alienation experienced by many residents as a consequence of aggressive public order policing which remains the primary response and central characteristic of the national police.

Fundamentals: Local Policing

In contrast to national policing experience within the Anglo-American, police system has continued to emphasize the importance of local service and integration with the community. This is best exampled within America where there is a continuing commitment to a wide range and variety of police agencies which operate within it. Most recently, the total number of police agencies in America has been estimated to stand at just below 20,000. However, the very number of local agencies suggests that America continues to provide the best example of a fragmented police system. While the long-term goal of many police professionals may have been to consolidate smaller police agencies, the evidence from larger police departments suggests that these continue to experience greatest difficulty in sustaining public police patrol. It has also proved difficult to persuade local residents of the value of relinquishing their own local police department in the interest of any perceived future gain in police efficiency.

As was to be argued over 40 years ago in what proved to be the most revealing study and analysis of policing in America, the President's Commission on Crime found that the community service function of the police could not be underestimated (The Challenge of Crime 1967). This was because analysis of the role of the police highlighted how dependent they remained on information given to them by the public. In its evaluation of the investigation function, the evidence of such dependence was stark. Thus, in its analysis of crimes reported in Los Angeles in 1966, it was to discover that "where the police were furnished a suspect's name then the case was very likely to be resolved." It noted however that in "1,375 crimes where no suspect was named only 181 cases were cleared by the police," and that in the LAPD one of the largest police departments in America, of all reported serious crimes against property recorded by the department, "78 % were never solved" (The Challenge of Crime 1967:97).

All the evidence indicated that a primary indicator of police effectiveness was the degree to which the police enjoyed the support and confidence of the public. This was because police dependence on the public for information was the primary driver of most crime case clear-up. Nor was it evident that increased expenditure on police specialization could necessarily be expected to impact on the crime rate or the clearance rate. The Commission discovered that in "practically every department the caseloads carried by detectives were too heavy to allow them to follow up thoroughly more than a small percentage of cases assigned to them" (The Challenge of Crime 1967:96).

In this situation increased further investment in the detective specialism, however large, would be unlikely to significantly increase the clear-up rate. It concluded that in the great majority of cases, personal identification by a victim or witness remained the only clue to the identity of the criminal (The Challenge of Crime 1967:97). Thus, the critical issue became one of encouraging victims and members of the public to come forward to help the police resolve cases.

The landmark 1967 report remains the bedrock for any evaluation of police effectiveness, and its central discovery of police dependence on the public it serves continues to have an immediate relevance to any contemporary debate over the organizational structure of policing and policing systems. This is because the evidence suggests that where the police are removed from regular contact with the public and cease to be integrated into the community the police ostensibly serve, then the decline in police efficiency and effectiveness can be expected to be quite dramatic. This may be measured in fact by the level of non-reporting of crime by the public who fail to report incidents because they believe the police cannot or will not do anything about it (The Challenge of Crime 1967:96).

This may indeed have been the overall experience in the UK as a consequence of the major police amalgamation program embarked upon in the 1960s and early 1970s. Bigger police departments ended in the creation of bigger police bureaucracies and much greater specialization all of which came at the expense of contact with the public by way of visible police patrol (Loveday 1999). Experience of the growth of police departments in the UK provided clear support for the inverse relationship identified by Monkennon in the 1980s that existed between police size and effectiveness. He noted how as police departments grew in size police patrol activity could thereafter be expected to decline (Monkennon cited in Loveday 1999).

Devolving Police Services in Centralized Systems

It is interesting to note that within France, one of the most centralized police systems in Europe, there has been growing recognition of the need for local police service delivery and evident failure of the national police model. Ironically while representatives of Anglo-American policing have highlighted its benefits, the French have implemented change at a local level which has significantly strengthened local poling in many towns and cities in France.

This can be traced back to the early 1980s with the interesting discovery that under the Law of 1884 passed by the national assembly, the mayor was given responsibility for public order in his commune (Journes 1993 cited in Loveday 1999). Under the 1884 Law, the mayor was given responsibility for the appointment of all municipal employees including police constables and inspectors, although this responsibility is balanced by the authority of the prefect who has a greater role in the recruitment of police in towns with over 40,000 residents. Thus, from the early 1980s in response to the perceived failure of the national police to provide effective policing, particularly in relation to responding to crime and delinquency and the decline in public police patrol, there has been a steady increase in the number of municipal police forces in many French towns and cities. In 1986, it was estimated that 530 municipal police forces were in existence, and that of these, 107 competed with the national police in terms of duties and responsibilities (Guyomarch 1991:331). In 1993 it was computed that there were some 10,000 municipal police officers who worked in 2,860 localities.

This interesting development remains a continuing phenomenon in response primarily to the failure of the national police organization to deliver effective police services. One consequence of this has been the decision of an increasing number of local mayors to establish their own police forces where the police are made directly accountable to the mayor. In effect the vacuum created by the withdrawal of the national police from visible policing has led to the creation of a policing system that is closely allied to the local government structure.

It is, however, evident that in France the future expansion of municipal policing may be influenced by the political position adopted by police unions on what they may understandably view as a threat to their role and status. In relation to this it is interesting to note that as identified by Body-Gendrot, one of the factors explaining the current malaise of policing in France is "a general lack of leadership to confront police unions' resistance to necessary change" (Body-Gendrot 2012:85). Yet in relation to this, it might be relevant to note that in terms of devolved policing, the experience of post-Franco Spain has demonstrated a similar movement to devolved policing with the creation of municipal police forces which in most towns and cities now act as an integral element to local municipal administration. With this has gone the part demilitarization of the Guardia Civil and closer engagement by it with local communities.

The realignment of policing with local government has also been demonstrated in England and Wales. Following the introduction of local partnership arrangements under the Crime and Disorder Act 1998, police forces now have a shared responsibility with the local authority to reduce crime within their areas. Formal arrangements under the 1998 Act mean that the local authority and other local partners are required to develop on an annual basis a strategy to reduce crime in their area. The strategy document is preceded by a local audit of crime and criminal damage and since the introduction of the 1998 Act has served to provide a far more comprehensive picture of the extent and nature of criminal activity than existed before.

The same legislation requires the partnership to consult with the local community about what should be the local strategic priorities before determining what they will be. Interestingly in this context has been the discovery that one of the highest priorities identified by the public is not in fact crime but antisocial behavior [ASB]. This has proved to have been [and remains] a significant factor in sustaining high rates of "fear of crime" in England and Wales despite the near-record fall in recorded crime. It is also evident that although ASB has long been a matter of concern to many residents, it was not one that, unfortunately, appeared to be shared by the police (Loveday 2012). Over the period of the partnership, one primary objective has been to realign police priorities with community priorities where in the past a commitment to a professional policing approach has often created a lacuna that has put local communities at risk.

Experience in England and Wales suggests that there is much to be gained from establishing close links between the police and the local authority. This is because while the police may be able to identify problems that may generate crime and disorder, unlike the local authority, they do not have the resources to respond to them. In an era of significant reductions in public spending, it is of interest to note in this context that in England and Wales, the Coalition government remains committed to increasing local accountability of the police by way of directly elected police and Crime Commissioners [PCCs].

While there may some argument over the mechanism surrounding the planned PCCs, created to achieve this, there can be little dispute about the principle that drives this reform which is to ensure that communities are in future better able to influence what kind of police service is locally delivered. To encourage greater public engagement, all police forces are now required to provide communities with local crime maps to enable them to make a more-informed judgement about how effective the local police force is proving to be in relation to the incidence of crime and disorder in their area.

Moreover, to assist local police forces in achieving greater responsiveness to local demand, the Coalition government is also in the process of establishing a National Crime Agency that can be expected to relieve local police forces from responding to serious, organized crime and terrorism as this new body is given a much higher status and role than has been accorded before to "national" agencies in the past.

Future Directions

The evidence from experience suggests that "local" policing only becomes meaningful when it is linked closely to local government structures. Moreover, even where local policing is in operation, there will be a continuing requirement for a national arm of policing to respond to the challenges of international crime and the terrorist threat. This is perhaps best evidenced in America where despite a plethora of local police departments, there are numerous federal agencies that exercise considerable power domestically. This federal responsibility cannot be ignored and has been substantially increased with the Patriot Act and the creation of the Department for Homeland Security [DHS] which now employs around 200,000 personnel and has a budget of $98.8 billion.

As in England and Wales the commitment to local policing is underpinned by a significant role accorded to national or federal agencies and reflects a reality that in the contemporary world there is a role for both local and national police capabilities and that neither can be expected to operate in isolation from the other. As the national police structures roll out in the Netherlands, Scandinavia, and Scotland, it will be of interest to discover how successful reliance on a single national force proves to be. Already in the Netherlands, local mayors are challenging the surrender of their responsibilities for the police and their removal to a national forum at The Hague.

In Scotland the argument for the introduction of a national force has been in part predicated on the need to protect existing police establishment [17,234 officers] by saving on overheads generated by the current eight forces (Fyfe 2012). Unfortunately no activity analysis of policing has been undertaken to assess whether current police establishment can be justified. In its absence the highly political nature of the Scottish governments' reform program becomes more apparent.

While local police commands will be aligned with Scotland's 32 local councils, it is noticeable that there will be limited influence accorded to the local council in the determination of policing priorities identified by the local police commander. Nor will the "partnership structure" developed in England and Wales be replicated in Scotland. It is also of some interest that in a public consultation on options for police reform, undertaken in 2011, less than 10 % of respondents supported the national police option, with most respondents preferring, in fact, a regional structure (Fyfe 2012).

In this analysis of types of policing structures, it is evident that national and local policing cannot be addressed in terms of simple alternatives. Within most advanced states, there is a perceived need for a mix of police structures which can respond to differing demands and challenges at differing levels. The degree of interdependence might be best evidenced in America where a continuing and strong commitment to local policing and community engagement is balanced by significant responsibilities given to an increasingly important federal arm.

It is also clear that national political cultures are significant in determining both domestic police structure and public expectation of policing styles. This of course is reinforced by the significance of the political context within which policing is delivered. The highly dirigiste approach to policing exhibited in France is replicated to a high degree in the Russian Federation where there may also be lower public concern about potential abuse by police of their authority and a greater indifference to the importance of legal procedures. These may, in fact, be seen by the public as impeding the work of the police (Kelly 2005:xxix).

It is also evident that as with many other areas of public service delivery there is very little evaluation of how police structures might impact on performance and general service delivery. Nor has any effective analysis has undertaken to demonstrate with any clarity whether in moving from local to national police systems the long-term cost savings which are claimed in defense of it are ever actually achieved. Additionally such savings may come at the expense of reducing public satisfaction with the police by encouraging specialization and reducing visible policing. As a result, it might then be concluded that any immediate financial savings could be more than outweighed by significant costs to police service delivery to the public.

Recommended Reading and References

Body-Gendrot S (2012) Globalisation, fear and insecurity: the challenges for cities north and south. Palgrave Macmillan, London

Chrisafis A (2012) 16 officers hurt and buildings set alight in French riots. The Guardian, London
Conley JA (1995) Police history. In: The encyclopedia of police science, 2nd edn. Garland Publishing, New York
Fyfe N (2012) In search of sustainable policing? Creating a national police force in Scotland. In: Fyfe N, Terpestra J, Tops P (eds) Restructuring policing: Western European perspectives on centralism, localism and the politics of police reform. Boom Publications, The Hague, Forthcoming 2012
Guyomarch A (1991) Problems of law and order in France in the 1980s: politics and professionalism policing and Society vol 1, 1991
Jones T (1995) Policing and democracy in the Netherlands. PSI, London
Kelly C (2005) Comrade Pavlik: the rise and fall of a Soviet boy hero. Granta Books, London
Loveday B (1999) Government and accountability of the police. In: Mawby RI (ed) Policing across the world. UCL Press, London
Loveday B (2012) Police reform in England and Wales: a new dimension in accountability and service delivery in the 21st century. In: Fyfe N, Terpestra J, Tops P (eds) Restructuring policing: Western European perspectives on centralism, localism and the politics of police reform. Boom Publications, The Hague, Forthcoming 2012
Loveday B, Reid A (2003) Going local. Who should run Britain's police? Policy Exchange, London
Marshall G (1965) Police and government. Methuen and co, London
Marshall G (1978) Police accountability revisited. In: Butler D, Halsey AH (eds) Policy and politics. Macmillan, London
Mawby RI (1990) Comparative policing issues: the British and American experience in international perspective. Routledge, London
Mawby RI (ed) (1999) Policing across the world. UCL Press, London
Mouhanna C (2011) Rural policing in France: the end of genuine community policing. In: Mawby RI, Yarwood R (eds) Rural policing and policing the rural, a constable countryside. Ashgate, Aldershot
Rudorff R (1970) The myth of France. Hamish Hamilton, London
The Challenge of Crime in a Free Society (1967) Report by the Presidents Commission on Law enforcement and Administration of Justice. US Government Printing Office, Washington, DC

National Crime Survey

► US National Crime Victimization Survey

National Victimization Surveys

Marcelo F. Aebi[1] and Antonia Linde[1,2]
[1]School of Criminal Sciences, Institute of Criminology and Criminal Law, University of Lausanne, Lausanne, Switzerland
[2]Open University of Catalonia, Barcelona, Spain

Overview

The aim of this entry is to trace the development and the expansion of general victimization surveys across the world. The description focuses on national general victimization surveys, although references to international surveys cannot be avoided because many countries use the International Crime Victims Survey (ICVS) as their national measure of victimization. Specific surveys, such as those focused on businesses, domestic violence, or minority victimizations, are excluded. The entry begins with a historical note, continues with a description of the main surveys already carried out – and those currently conducted periodically – across the five continents, and concludes with a summary of the main findings and a brief mention of known methodological issues in comparative cross-national research based on national victimization surveys.

Historical Development of Crime Victim Surveys

In 1730, following a series of citizens' complaints about an increase of property offences, the City Council of Aarhus, Denmark, sent six persons throughout the households of the city to ask whether they had been victims of domestic burglary (Sparks 1981). This study has often been considered as the first victimization survey.

Two hundred years later, on 20 October 1935, George Gallup published his first survey of public opinion (Newport 2010). In 1945, Gallup established an agency in Finland and conducted a general survey that included a question that can be translated from Sweden – one of the Finnish national languages – as "Do you have this year personally suffered from a crime?" In case of an affirmative answer, respondents were asked to specify whether it was theft, burglary, robbery, assault, trespassing, fraud, or something else. A search in Gallup's database (http://brain.gallup.com) shows that similar questions were not asked in the United States until 1972, when the survey included the following question: "During the last 12 months, have any of these happened to you?- A) My home was broken into or an attempt was made to do so; B) I was mugged or assaulted; C) Money or property was stolen from me or some other member of my household; D) My home, or car, or other personal property was vandalized; or E) My car, or a car owned by a member of my household was stolen?" According to our search, before 1972, Gallup polls included some questions on crime – for example, in 1954, the public considered that comic books, TV, and radio crime programs were partly to blame for teenage violence; in 1957, it considered that parents were mainly to blame for that; and, since 1965, polls included a question on fear of crime while walking alone in their own neighborhood at night – but not phrased as a typical victimization question. Thus, the 1945 Finish Gallup Poll seems to be the first major public survey to have asked a question on victimization. The Nordic experience with these questions is reflected in the fact that it is precisely a Finish criminologist, Inkeri Anttila, who is often quoted as one of the main promoters of victimization surveys. However, the publication that is usually quoted refers to self-reported delinquency survey and includes only this sentence about victim surveys: "By Gallup investigations it is also possible to establish how many of the general public have been made objects of a certain crime, and how many of these have reported the crime to the authorities" (Anttila 1964: 413).

The rising public worries about crime in the United States by the mid-1960s – sometimes qualified as a "moral panic" but at least partially supported by a 1965 Gallup Poll (Chambliss 1994) – combined with the increased use of surveys and the inclusion of questions such as the ones mentioned below attracted the interest of criminologists. Thus, the President's Commission on Law Enforcement and Administration of Justice (1967) sponsored three surveys in the United States, conducted by Biderman et al. (1967), Reiss (1967), and Ennis (1967). Between 1969 and 1972, several surveys were conducted in cities across the United States (Sparks 1981, with references) and in the Nordic countries, starting in Finland in 1970 (Aromaa 1971) and then in Denmark, Norway, and Sweden in 1971 and 1972. Also in 1972, the United States launched the National Crime Surveys (NCS) program which led to the first national survey being conducted in 1973 and annually since then (Rand 2006). Australia carried out a survey in 1974 (Congalton and Najman 1974). In the same year, the Dutch Ministry of Justice commissioned Gallup/NIPO to carry out a pilot survey following the format of the Gallup US survey of 1972 (Buikhuisen 1975). Building on this experience, the Ministry of Justice of the Netherlands launched in 1975 its series of national surveys which has been repeated annually – first by the Ministry of Justice and, from 1980 on, by the Central Bureau of Statistics – ever since (Van Dijk and Steinmetz 1979). Following a series of local surveys that started in the city of London in the early 1970s, the United Kingdom launched the British Crime Survey (BCS) in 1982, which has been carried out periodically since then. In the 1980s, Finland and Norway also introduced a periodical survey, as did the city of Barcelona/Spain in 1983 and Australia in 1975. Around the same time, Clinard (1977) noted that victimization surveys could be especially useful in developing countries, in which police statistics of crime were sometimes unavailable. However, the breaking point for the development of these surveys across the world is undoubtedly the first International Crime Victim Survey (ICVS), conducted in 1989 and published in 1990 (Van Dijk et al. 1990), largely building on the experiences of the Dutch, British, and Swiss national surveys. Thus, in the 1990s, roughly 70 countries participated in at least one of the waves of the

ICVS or, at least, used its questionnaire for conducting victimization surveys. Also in the 1990s, regional Latin Americans victimization surveys were developed and countries such as Italy, France, and the Autonomous Region of Catalonia introduced periodical victim surveys. Since the beginning of the twenty-first century, some Latin American (Argentina, Chile, and Mexico) and European countries (Denmark and Sweden) have also introduced national periodical victimization surveys.

National Victimization Surveys by Continent

The following sections present the development of national general victimization surveys by continent. The information available does not always allow a differentiation between the year in which the survey was conducted, the year of reference of the survey, and the year of publication. For example, in most European Union countries, the EU ICS was conducted in 2005, using 2004 as the year of reference, but the results were published in 2007 and 2008. In principle, the dates indicated in this entry refer to the date of publication of the results. This review draws on the ones conducted for Latin America by Dammert et al. (2010), for Asia by Chan (2008) and for Europe by Aebi and Linde (2010).

America

In the American continent, there are three regional multipurpose surveys that include victimization modules: the *Latinobarómetro*, conducted by a nonprofit NGO (Latinobarómetro Corporation); the *Ecosocial* (Encuesta de Cohesión Social en América Latina), conducted by another nonprofit NGO (Corporación de Estudios para Latinoamérica – CIEPLAN); and the *AmericasBarometer* of the Latin American Public Opinion Project (LAPOP) of the Vanderbilt University. The questions included in these surveys are sometimes too broad – for example, the Latinobarómetro asks: Have you or a relative been robbed, aggressed, or victim of a crime during the last 12 months? (Lagos and Dammert 2012) – and, as a consequence, they only provide a rough estimate of the extent of victimization.

The *Latinobarómetro* is conducted annually since 1995 in Latin America. Eight countries participated in 1995, 17 from 1996 to 2003, and 18 since 2004. Currently, Cuba is the only Latin American country that does not participate in the survey (see Table 1). The *Ecosocial* survey was conducted in 2007 in seven Latin American countries (see Table 1). The *AmericasBarometer* is conducted every second year since 2004 in the whole American continent. Eleven nations participated in the first survey in 2004, 20 in 2006, 23 in 2008, and 26 in 2010 (see Table 1). Thus, Cuba is the only country not covered by the survey.

In the following country-by-country analysis, it can be seen that victimization surveys became a standard measure of crime in the United States in the 1970s, in Canada in the 1980s, and in most Latin American countries in the 1990s. The latter introduced such dedicated surveys mainly through their participation in the ICVS with city samples. In particular, in 1992, Argentina, Brazil, and Costa Rica took part in the second wave of the ICVS. The same countries participated in the third wave, in 1996/1997, in which Bolivia, Colombia, and Paraguay also took part, and the number of participating countries continued to increase in the following years. In the 2000s, the UNDP sponsored some victimization surveys in collaboration with governments of countries of the Americas, such as the one conducted in 2004 in Costa Rica and El Salvador and the IDHAC survey, conducted in 2009–2010 in seven Central American countries (Belize, Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua, and Panama).

In **Argentina**, the first victimization survey was conducted in the capital, Buenos Aires, in 1989, with the support of a private foundation. Always with samples from the capital, the country participated in the ICVS in 1992, 1996, and 2004. From 1997 to 2003, the Ministry of Justice used a slightly adapted version of the ICVS to conduct an annual victimization survey in the city of Buenos Aires and its surroundings, including also in some of its waves the cities of Rosario,

National Victimization Surveys, Table 1 Regional and national victimization surveys in American countries

	Latinobarómetro (wave 2010)	Ecosocial (2007)	AmericasBarometer (wave 2010)	National survey
Argentina	✓	✓	✓	✓
Belize			✓	
Bolivia	✓		✓	✓
Brazil	✓	✓	✓	✓
Canada			✓	✓
Chile	✓	✓	✓	✓
Colombia	✓	✓	✓	✓
Costa Rica	✓		✓	✓
Dominican Republic	✓		✓	
Ecuador	✓		✓	✓
El Salvador	✓		✓	✓
Guatemala	✓	✓	✓	✓
Guyana			✓	
Haiti			✓	
Jamaica			✓	
Honduras	✓		✓	
Mexico	✓	✓	✓	✓
Nicaragua	✓		✓	✓
Panama	✓		✓	✓
Paraguay	✓		✓	✓
Peru	✓	✓	✓	✓
Suriname			✓	
Trinidad and Tobago			✓	
United States			✓	✓
Uruguay	✓		✓	✓
Venezuela	✓		✓	✓

Mendoza, and Cordoba. Since 2006, some local governments (city of Buenos Aires, province of Santa Fe) started conducting their own surveys with the scientific support of universities. Also, since 2006, the Di Tella University conducts a monthly victimization survey in the main 40 urban centers of the country.

In **Barbados**, the Victimization Survey of Barbados was conducted in 2002 using the ICVS questionnaire and in 2009–2010 including also a business victimization section. In **Bolivia**, three victimization surveys were conducted in 2006. One of these surveys was conducted by IPSOS and covered the cities of La Paz, El Alto, Cochabamba, and Santa Cruz. Another one was sponsored by the UNDP and covered cities with more than 2,000 inhabitants. The third one was conducted by the Observatory of Democracy and Security and covered the cities of La Paz and El Alto. Bolivia conducted the ICVS in 1996 with a sample of the capital, La Paz.

Brazil conducted four national victimization surveys. The ones of 1988 (PNAD) and 2002 (PESB) were multipurpose surveys that included a module on victimization and were conducted by the Brazilian Institute of Geography and Statistics (IBGE) and the Universidade Federal Fluminense, respectively. The IBGE conducted also in 2009 a national victimization survey, while, in 2010, the UNDP sponsored a survey which used an adapted version of the ICVS questionnaire. Indeed, Brazil participated in the ICVS in 2002 and 2005 with samples of the cities of Rio de Janeiro, Sao Paulo, Recife, and Vitoria. Apart from that, many local surveys were conducted throughout the country. That was the case in Rio de Janeiro at the level of the city (1992, 1996, 2002, 2005, and 2006), some of its neighborhoods (2001), the metropolitan region (1996, 2000, and 2006), and one of the municipalities of that region (2000). In Sao Paulo, victimization surveys were conducted at the level of the city (1997 and 2003), the metropolitan region (1999), and the state (1998). In Belo Horizonte, surveys were conducted at the level of the city (2002 and 2003) and the metropolitan region (2006). Finally, other victimization surveys were conducted in the cities of Marilia (2003) and Alvorada (2004).

In **Canada**, a national victimization survey is conducted roughly every 5 years as part of the General Social Survey program. The survey took place in 1988, 1993, 1999, 2004, and 2009. Before that, a national victimization survey had been conducted in 1981. Canada participated also in the five waves of the ICVS (1989, 1992, 1996, 2000, and 2004) with national samples and in the ICVS-2 pilots of 2009 and 2010.

In **Chile**, the National Urban Survey on Citizens Security (ENUSC) was conducted in 2003 by the Ministry of Interior, and it became annual since 2005. Apart from that, since 1999, an ONG (Fundación Paz Ciudadana en Chile) conducts surveys every 6 months in different municipalities of the main cities of the country.

In **Colombia**, the first victimization survey was conducted in the city of Cali in 1978 by a researcher of the University of Wales (United Kingdom) with the support of the Colombian Statistical Institute (DANE). Between 1978 and 1980, the National Association of Financial Institutions (ANIF) carried out short victimization surveys in several cities, but no scientific publications were produced out of them. The Cisalva Institute of the University del Valle conducted also a victimization survey in six cities in 1996 and in Pasto in 2008. Since 2006, the Foundation *Seguridad y Democracia* carried out some victimization survey in the six main urban centers of Colombia. In the capital city of Bogota, the Chamber of Commerce has created a Security Observatory that, since 1996, conducts periodically a victimization survey. Bogota provided also the sample for the participation of Colombia in the ICVS in 1997 and 2000. In 2003, a victimization survey sponsored by the World Bank and the DANE took place in the cities of Cali, Medellin, and Bogota, while in 2009, a pilot survey took place in Soacha. At a national level, a module on victimization was included in the National Households Survey in 1985, 1991, and 1995, and the first national victimization survey – sponsored by the Inter-American Development Bank (IDB) – should take place in 2013.

In **Costa Rica**, the United Nations Latin American Institute for the Prevention of Crime and the Treatment of Offenders (ILANUD) carried out a victimization survey in the capital city of San José in 1982. The same city provided the main sample for the participation of Costa Rica in the ICVS in 1992 and 1996, and the design of both surveys included also a small rural sample. National victimization surveys were conducted in 2006, 2008, and 2010. These three national surveys were sponsored by the UNDP, and the ones of 2008 and 2010 were cosponsored by the National Institute of Statistics. The **Dominican Republic** conducted in 2005, through its National Institute of Statistics, a national multipurpose survey (ENHOGAR) that included a module on victimization. In **Ecuador**, after some regional surveys conducted by NGOs, the government conducted the National Survey on Victimization and Perceptions of Insecurity in 2008 and 2010. In principle, this survey should be conducted every 2 years. In **El Salvador**, the University Institute of Public Opinion (IUDOP) conducted national victimization surveys in 1998, 2004, and 2009 and in the cities of Sonsonate and Cojutepeque in 1999. In **Guatemala**, since 2004, the UNDP conducts surveys on victimization and insecurity perception in Guatemala City every 6 months.

In **Mexico**, the first victimization survey was conducted in the city of Jalapa in 1976 by the Veracruzana University, using a questionnaire developed by the Texas Department of Public Safety (United States). A similar questionnaire was used for a survey carried out in 1983 in the Federal District and its surroundings, sponsored by the National Institute of Criminal Sciences (INACIPE). The Mexican National Institute of Statistics (INEGI) conducted victimization surveys in the Federal District and the State of Mexico in 1988, 1990, 1992, and 1994 and in Monterrey, Oaxaca, Veracruz, Cuernavaca, Ciudad Juárez, and the metropolitan area of the city of México in 1992, 1993, and 1997. In the Federal District and the State of Mexico, the Center for Research and Teaching in Economics (CIDE) has been carrying out every 6 months since 2005 a Survey on Victimization and Institutional Efficacy (ENVEI). At a national level, in 2002, the Citizen's Institute of Studies on Insecurity (ICESI) conducted for the first time the National Survey on Insecurity (ENSI), with an instrument inspired by the ICVS questionnaire and the one developed by the Autonomous University of Mexico (UNAM) for a national survey conducted in 2000. The ENSI was conducted again in 2003, 2005, 2006, and annually from 2008 to 2010, by the ICESI, sometimes in collaboration with the INEGI, and with some modifications of the questionnaire. Since 2011 the ENSI became the National Survey on Victimization and Perception of Public Security (ENVIPE) and was placed under the responsibility of the INEGI, which continues to carry it out annually with a revised questionnaire. INEGI also conducts since 2009 the Continuous Survey on the Perception of Public Security (ECOSEP) that generates a monthly index of public security perception. Finally, Mexico participated in the ICVS in 2004 with a national sample.

In **Panama**, within the framework of a project organized by the University of Santa Barbara (California), the University of Panama conducted in 1973, with a sample of the capital, the first Latin American victimization survey. **Paraguay** conducted a national victimization survey in 2009 through its Ministry of Interior using the ICVS questionnaire. The country also conducted the ICVS in 1996 with a sample of the capital, Asuncion. **Peru** participated in the 2005 ICVS with a sample of the city of Lima. The same year, the Ministry of Interior also conducted through a private enterprise a victimization survey in the main urban centers. The ONG *Ciudad Nuestra* conducted the First National Urban Victimization Survey in 2011. An annual survey on citizens' security conducted by an ONG (Grupo de Opinión Púbica) since 2004 includes some questions on victimization. In Lima, a local victimization survey was conducted in 2007.

As mentioned in the historical introduction, in the **United States of America**, three pilot victimization studies were sponsored by the President's Commission on Law Enforcement and Administration of Justice in 1967. They were followed by several city surveys, by the 1972 survey conducted by Gallup/USA (Gallup Poll Number 861) and, since 1973, by the annual *National Crime Victimization Survey* (NCVS, conducted under the name of National Crime Survey until 1990).

The Ministry of Interior of **Uruguay** conducted victimization surveys in the cities of Montevideo and Canelones in 1999, 2000, 2001, 2002, and 2004. A national survey was sponsored by the same ministry in 2007. In **Venezuela**, local victimization surveys were conducted in the city of Mérida (in 1980, 1983, and 1985), Maracaibo (1987, in the context of a comparison with the United States of America), and Caracas (2005). In 2001, 2006, and 2009, national victimization surveys were conducted for the Ministry of Interior and Justice by a private enterprise (2001) and by the Institute of Statistics (2006 and 2009). In 2007, another national survey was conducted by a consortium of institutes and universities.

Belize and **Nicaragua** do not have their own national victimization surveys, but victimization data on them for 2009–2010 can be found in the IDHAC survey and in the *AmericasBarometer* for 2008. Nicaragua also participated in the *AmericasBarometers* of 2004 and 2006.

Europe

In Europe, there are no genuine regional comparative *general* victimization surveys yet. The *Eurobarometer* 44.3 (1996), 54.1 (Autumn 2000), and 58.1 (Autumn 2002) included some questions on fear of crime and victimization, and the same is true for the *European Social Survey* of

2002, 2004, and 2006, but they cannot be compared to a victimization survey and not all European countries participated in them. Eurostat is currently preparing the *European Security Survey*, a full victimization survey which should take place in 2013 or 2014 (van Dijk et al. 2010).

Austria participated in the ICVS in 1996 and 2005 (EU ICS) with national representative samples and conducted the pilot study on the EU victimization survey module in 2009. Since 2006, the Austrian Safety Board conducts annually the Security Barometer (*Sicherheitsbarometer*).

Belgium participated in four sweeps of the ICVS – in 1989, 1992, 2000, and 2005 (EU ICS) – with national representative samples. Since 1997, the country conducts regularly a national victimization survey called Security Monitor (*Moniteur de Sécurité*). This survey has been conducted in 1997, 1998, 2000, 2002, 2004, 2006, and 2008–2009. Apart from that, since 1996, the Flanders Authority conducts annually the APS-SCV survey (*Administratie Planning en Statistiek – Sociaal Culturele Veranderingen*) among the Flemish-speaking population.

In **Bulgaria**, victimization questions were included in the monthly surveys led within the framework of the UNDP's Early Warning project in 1996–1997. The ICVS questionnaire was used, with a sample from the capital, Sofia, in 1997, and with national representative samples in 2002 and 2004 (EU ICS). Also in 2004, another national victimization survey was conducted with an ad hoc questionnaire. Since then, the ICVS questionnaire is being used for periodical national victimization surveys conducted in 2005, 2007, 2008, and 2009. **Cyprus** only conducted the pilot study on the EU victimization survey module in 2009 with a sample of two urban areas.

The **Czech Republic** participated, as part of Czechoslovakia, in the 1992 ICVS with a national representative sample. After the separation from Slovakia, the ICVS was conducted with a national representative sample in 1996 and with a sample from the capital, Prague, in 2000. Apart from that, between 2000 and 2003, four victimization surveys with national representative samples were conducted with an ad hoc questionnaire. The country also conducted the study Victimization of Citizens of the Czech Republic by Some Types of Criminality in the Year 2004. In 2006, an adapted version of the ICVS questionnaire was used for the survey Experiences of Czech Republic Citizens with Some Offences. The country also conducted the pilot study on the EU victimization survey module in 2008/2009.

Denmark is considered as the birthplace of victim surveys as, in 1730, the council of the city of Aarhus reacted to the complaints of the citizens by asking six persons to go through all the households of the town asking their inhabitants if they had been victims of burglary during the last 3 or 4 years. In September 1971, Denmark was the second European country to conduct a modern victimization survey, which was indeed a replication of the one conducted in Finland 1 year before. The country participated in the ICVS in 2000 and 2005 (EU ICS) with national representative samples and conducted both the pilot study on the EU victimization survey module in 2009 and the second ICVS-2 pilot study in 2010. Since 2005, Denmark has an annual victimization survey which is not based on the ICVS questionnaire but includes some comparable questions.

Estonia participated in four sweeps of the ICVS, in 1993, 1995, 2000, and 2004 (EU ICS) with national representative samples. The ICVS became then a sort of regular national victimization survey, and, as such, it was conducted again in 2009.

Finland conducted in 1970 the first contemporary European victimization survey. With the cooperation of the Scandinavian Research Council for Criminology, that survey was subsequently replicated in Denmark, Norway, and Sweden in 1971/1972. The Finnish government also financed a replication of that survey in Finland in 1973. Finland carries out a periodical victimization survey called the Finnish National Safety Survey. That survey has been conducted in 1980, 1988, 1993, 1997, 2003, 2006, and 2009. The country also participated in the five sweeps of the ICVS, in 1989, 1992, 1996, 2000, and 2005 (EU ICS) with national representative samples.

In **France**, the first nationwide victimization survey, covering the years 1984–1985, was conducted in 1986. The country participated in four sweeps of the ICVS, in 1989, 1996, 2000, and 2005 (EU ICS), with national representative samples. Between 1996 and 2006, France conducted 11 annual surveys on the Living Conditions of Households that contained a module on victimization loosely based on the ICVS questionnaire. Since 2005, the country launched an annual victimization survey called Framework of Life and Security, whose questionnaire is inspired by the one used in the British Crime Survey. Local victimization surveys were also conducted in Épinay and Toulouse in 1989, in Amiens in 1999, and in the Île-de-France (Paris) region in 2001, 2003, 2005, 2007, and 2009, as well as in five cities that are members of the French Forum for Urban Safety in 2005.

In **Germany**, the first victimization survey was conducted in Stuttgart, in 1973. Reflecting the federal structure of the country, most of the German victimization surveys took place at the local level, at an average rhythm of four surveys per decade until 1990 and up to five per year since then. Some of these surveys used a longitudinal design, such as the Victims of Crime in Bochum study, conducted in 1975, 1986, and 1998 in the city of Bochum. In 1989, the ICVS was conducted in Western Germany with a sample representative of the population of the Federal Republic of Germany. After the reunification of the country, victimization surveys with national representative samples were conducted by different researchers in 1991, 1992, 1993, 1995, and 1997–1998. In 2005, Germany took part in the EU ICS. The country also participated in the pilot of the EU victimization survey module in 2009 with a sample of four states and in the ICVS-2 pilots of 2009 and 2010.

In **Greece**, in 1991, a pilot study of the ICVS was conducted in the city of Athens. In the same city, surveys on fear of crime were conducted in 1998 and 2004. At the national level, only one victimization survey was conducted, in 2001.

Hungary participated in the ICVS in 1996 and 2005 (EU ICS). In 2003, the country also conducted the survey Victims and Opinions with a national representative sample and a questionnaire that was not based on the ICVS but included some comparable questions.

In **Ireland**, the first victimization survey was conducted in 1982–1983 with a national representative sample. In 1994, two victimization surveys were conducted. One used a national representative sample and the other a sample covering only Dublin. In 1996, the survey Victims of Recorded Crime in Ireland was conducted with a sample of victims drew from the records of the Ireland's National Police. Apart from that, Ireland participated in the 2005 EU ICS with a national representative sample. Finally, the Quarterly National Household Surveys (QNHS), introduced in 1997, include regularly a module on victimization and, as a consequence, can be considered as a sort of regular national victimization survey. The module on victimization was included in 1998, 2003, 2006, and 2009.

Italy participated in the ICVS in 1992 and 2005 (EU ICS) with national representative samples. In 1991 and 1994, national victimization surveys were conducted by UNICRI (with the support of the Ministry of the Interior) and the Cattaneo Institute, respectively. From 1993 to 2003, the national multipurpose survey called Everyday Life Aspects (*Aspetti della vita quotidiana*) included some questions on victimization, but, currently, only fear of crime in the area of residence is measured through this survey. At the same time, the Italian Citizens' Safety Survey (*Sicurezza dei cittadini*) is a multipurpose national study conducted every 5 years (1997/1998, 2002, and 2008/2009) that includes questions on victimization. This survey used oversampling for the Emilia-Romagna region in 1997/1998, for that region and another four in 2002, and for five provinces of the South in 2008/2009. Finally, in Bologna, a local victimization survey was conducted in 1994.

Luxembourg participated in the 2005 EU ICS with a national representative sample. Apart from that, the police services of Luxembourg conducted in 2007 a survey on feelings of insecurity with a national representative sample.

National annual victimization surveys started in the **Netherlands** in 1974 under the responsibility of the Ministry of Justice, initially out-contracted to Gallup/NIPO. These surveys were later adopted by Statistics Netherlands (CBS) that has been carrying out national victimization surveys since 1980. The Permanent Survey on Living Conditions (POLS – *Permanent Onderzoek Leefsituatie*) including the Justice and Security Module was conducted from 1980 to 2005 (from 1980 to 1985 every year and from 1986 to 1992 every 2 years). At the same time, from 1980 to 1992, the Crime Victim Survey (ESM) was held first annually and, from 1984 to 1992, biannually. From 1992 onwards, the ESM was succeeded by the Justice and Security Survey (ERV – *Enquête Rechtsbescherming en Veiligheid*). Since 2005, the ERV was replaced by the National Security Monitor, which integrated elements from the former POLS Justice Module and from the Police Monitor (PMB). The latter is conducted in every police region since 1993, and it was conducted every second year until 2001 and annually since then. The National Security Monitor was stopped in 2008 and replaced by the annual Integral Security Monitor. Apart from that, the Netherlands participated in the five sweeps of the ICVS in 1989, 1992, 1996, 2000, and 2005 (EU ICS) with national representative samples and in the second ICVS-2 pilot study in 2010.

In **Norway**, by the end of 1971, the Scandinavian Research Council for Criminology supported the replication of the victimization survey conducted in Finland in 1970. Currently, the Survey of Living Conditions includes periodically a module on victimization. That module was used in 1983, 1987, 1991, 1995, 1997, 2001, 2004, and 2007. Norway participated in the ICVS in 1989 and 2004 (EU ICS).

Poland participated in the five sweeps of the ICVS and in the pilot study on the EU victimization survey module in 2009. In 1989, the survey was conducted in the city of Warsaw only, and in 1992, 1996, 2000, and 2004 (EU ICS), it was based on national representative samples. In 2000 and 2004, the country used an adapted form of the ICVS questionnaire. The ICVS questionnaire was also used for a survey conducted in Warsaw in 2005.

Portugal participated in the ICVS in 2000 and 2004 (EU ICS) with national representative samples and conducted the pilot study on the EU victimization survey module in 2009. Portugal also conducted national victimization surveys with an ad hoc questionnaire in 1991, 1992, and 1994.

Romania participated in the ICVS in 1996 and 2000. In 1996 the country combined an urban sample from the capital city, Bucharest, and a small rural sample. In 2000, the country used a sample from Bucharest only. From 2001 to 2006, the National Institute of Statistics (Institutul Naţional de Statistică, INS) conducted annually the multipurpose Living Conditions Survey (*Condiţiile de viaţă ale populaţiei din România*, ACOVI), which included questions on victimization, with national representative samples.

Russia participated in the ICVS in 1992, 1996, and 2000 with a sample of the capital, Moscow. In 2001, a victimization survey was conducted in the cities of Smolensk, Omsk, and Volgograd. The **Slovak Republic** participated, as part of Czechoslovakia, in the 1992 ICVS with a national representative sample. After the separation from the Czech Republic, the ICVS was conducted in Slovakia in 1997. The country also conducted the pilot study on the EU victimization survey module in 2008/2009. **Slovenia** conducted the ICVS in 1992, 1996, and 2001 and participated in the pilot study on the EU victimization survey module in 2009. In 1992, the sample was restricted to the capital, Ljubljana; in 1996, it combined an urban (Ljubljana) and a rural subsample; and, in 2000, it was representative at the national level.

Spain participated in the ICVS in 1989 and 2005 (EU ICS) with national representative samples and conducted the pilot study on the EU victimization survey module in 2008/2009. The Centre for Sociological Research (CIS – *Centro de Investigaciones Sociológicas*) conducted six victimization surveys with national representative samples (in 1978, 1980, 1991, and 1995), one in Madrid in 1980, and one with a sample of several cities in 1982. The CIS also conducted in 1987 a study on terrorism and citizens' security. Finally, the barometer conducted periodically by the CIS includes a few questions on citizens' concerns, among which delinquency is often included. The Andalusian Institute of Criminology has also carried out a series of surveys in different cities of the Autonomous Community of Andalusia in 1994, 2005, 2006, and 2007, as well as a national survey in 2009, using the successive versions of the ICVS questionnaire, adapted to the Spanish case. The Autonomous Community of Catalonia participated in the ICVS in 1996 and 2000 and conducted the pilot study on the EU victimization survey module in 2009. From 1983 to 2001, the Survey on Public Security in Barcelona was conducted annually in the city of Barcelona. This victimization survey inspired the Survey on Public Security in Catalonia, which is conducted annually with a sample representative of the population of Catalonia since 1999.

Sweden conducted its first victimization survey in 1972 as part of a project sponsored by the Scandinavian Research Council for Criminology in Nordic countries. Since 1978, the country conducts annually the multipurpose survey called Living Conditions Survey (ULF), which includes a module on victimization. Since 2006, the country also carries out a national annual victimization survey called the Swedish Crime Survey (NTU). Sweden participated also in four sweeps of the ICVS in 1992, 1996, 2000, and 2005 (EU ICS) with national representative samples. The country also conducted the pilot study on the EU victimization survey module in 2009 and the first and second pilot studies of the ICVS-2 in 2008/2009 and 2010, respectively.

Switzerland conducted seven national victimization surveys, of which four were placed under the umbrella of the ICVS. These surveys took place in 1984/1987, 1989 (ICVS), 1996 (ICVS), 1998, 2000 (ICVS), 2005 (ICVS), and 2011. Since 1989, the survey is always based on the ICVS questionnaire.

In the **United Kingdom**, the General Household Survey was the first survey that covered victimization issues. This multipurpose survey, conducted during the 1970s, included some questions about domestic burglary victimizations. Since 1982, the United Kingdom has carried out periodically the British Crime Survey (BCS) which, as explained below, took different forms in England and Wales, Scotland, and Northern Ireland. In **England and Wales**, the first victimization survey was the London survey, conducted in the early 1970s. Other local surveys took place in Sheffield (1975), London (1983 and 2002), the county of Nottingham (1985), Merseyside (1984), and in the London Borough of Islington (1986). In the 1980s, the BCS was conducted three times in England and Wales (in 1982, 1984, and 1988). From 1992 to 2000, it was conducted every other year, and, since 2001, it became a continuous survey with monthly interviews. From 2012, the survey's name was changed from BCS into the Crime Survey for England and Wales. England and Wales participated also in the five sweeps of the ICVS, in 1989, 1992, 1996, 2000, and 2005 (EU ICS) and in the ICVS-2 pilots of 2009 and 2010. **Northern Ireland** participated four times in the ICVS, in 1989, 1996, 2000, and 2005 (EU ICS) with national representative samples. The country also conducts periodically the Northern Ireland National Crime Survey (NICS). This survey was conducted in 1994/1995, 1998, 2001, and 2003/2004, and, since 2005, it became a continuous survey with monthly interviews that follows closely the model of the BCS. **Scotland** conducts periodically a national victimization survey since 1982. In 1982 and 1988, the country used the BCS. The first independent Scottish Crime Survey was conducted in 1993 and repeated in 1996, 2000, and 2003. That survey was replaced by the Scottish Crime and Victimization Survey (SCVS), which took place in 2004 and 2006. In 2008, the SCVS was replaced by the Scottish Crime and Justice Survey (SCJS) which is a continuous survey with monthly interviews, inspired by the methodology of the BCS, but with a slightly different questionnaire. Scotland also participated in four sweeps of the ICVS, in 1989, 1996, 2000, and 2005 (EU ICS).

The following European countries do not have a national victimization survey but have conducted the ICVS, usually with samples of their capital city complemented in some cases with a small rural sample (the year and the name of the city from where the main sample was extracted are indicated between brackets): **Albania** (Tirana, 1996 and 2000), **Belarus** (Minsk, 1996 and 2000), **Croatia** (Zagreb, 1996 and 2000), the **Former Yugoslav Republic of Macedonia** (Skopje, 1996), **Latvia** (1995 and 1998, with multiple-city samples and 2000 with a national representative sample), **Lithuania** (1996/2007, with a multiple-city sample, 2000 with a sample from Vilnius, and 2005 with a national representative sample), **Malta** (1997, unpublished, except in ICVS reports), **Serbia** (Belgrade, 1996), and **Ukraine** (Kiev, 1996 and 2000).

Oceania

In **Australia**, the Australian Bureau of Statistics (ABS) conducted national victimization surveys in 1975 (except the Northern Territory), 1983, 1993, 1998, 2002, 2005, and, annually, since 2008–2009 (as part of the Multipurpose Household Survey – MPHS). Questions on fear of crime and perception of security were also covered in other surveys, such as the General Social Survey (conducted in 2002, 2006, and 2010, which includes also some questions on crime victimization) and the Personal Safety Survey (2005) and in the Southern Australia Health Omnibus Survey, conducted regularly in that region of the country. Apart from that, the ABS conducted regional surveys in Adelaida (1985); Victoria and Queensland (1987); Tasmania (1989); New South Wales (1990); Queensland, Southern Australia, New South Wales, and Western Australia (1991); New South Wales (1992); and Victoria and New South Wales (1995). Australia also participated in the ICVS in 1989 with a sample of New South Wales and Victoria and, with national samples, in 1992, 2000, and 2004. In **New Zealand**, the Ministry of Justice conducted national surveys in 1996 and 2001 (New Zealand National Survey of Crime Victims) as well as in 2006 and 2009 (New Zealand Crime and Safety Survey). The country also participated in the ICVS in 1992 and 2004 with national samples. In **Papua New Guinea**, the ICVS was conducted in 1992. The Safer Cities Programme of UN-HABITAT, with the financial support of the UNDP, launched a project in 2002 that allowed carrying out a victimization survey in the capital, Port Moresby. In 2004 and 2005, victimization surveys were conducted in the cities of Arawa and Buka.

Africa

Egypt participated in the ICVS in 1992 with a sample of the capital, Cairo. In 2006, a national victimization survey was carried out by the National Center for Social and Criminological Research (NCSCR) with a random sample of 6 governorates including Al Fayoum, Assiut, Behera, Cairo, Dakahlia, and Port Said. **Nigeria** conducted the ICVS in 1998 with a sample of the capital, Lagos. In 2010, the MacArthur Foundation supported a national victimization survey with a representative sample drawn from the 36 states of Nigeria and the Federal Capital Territory. **South Africa** conducted the ICVS in 1993, 1996, and 2000 with samples of the capital, Johannesburg. Since 1998, the country has a periodical national victimization survey. It was conducted in 1998, 2003, and 2007 under the name of South Africa National Victims of Crime Survey (NCVS), and in 2011, it became the Victims of Crime Survey (VOCS). In 2005, South Africa conducted a youth victimization survey, with a sample of youths aged 12–22. The Safer Cities Programme of UN-HABITAT, with the financial support of the Dutch government sponsored since 1997 victimization surveys in several cities, including Johannesburg, Durban, Pretoria, Port Elizabeth, and Cape Town. **Tanzania** conducted the ICVS in 1992 with a sample of the main urban center, Dar es Salaam. The Safer Cities Programme of UN-HABITAT, with the financial support of the Dutch and the Swedish government, sponsored victimization surveys in 2000 (Dar es Salaam), 2004 (Dar es Salaam, Arusha and Mtwara), and, in collaboration with the UNODC, 2008 (Dodoma, Mbeya, Moshi, Mwanza, and Tanga). **Uganda** conducted the ICVS in 1992, 1996, and 2000 with samples of the capital, Kampala. The UNODC sponsored a national victimization survey in 2008.

The following African countries do not have a national victimization survey but have conducted the ICVS, usually with samples of their capital city complemented in some cases with a small rural sample (the year and the city from where the main sample was extracted are indicated between brackets): **Botswana** (1997 and 2000, Gaborone), **Lesotho** (2000, Maseru), **Mozambique** (2002, with a sample of the city of Maputo and the regions of Nampula, Sofala and Zambézia), **Namibia** (2000, Windhoek), **Swaziland** (2000, Mbabane), **Tunisia** (1992, Tunis), **Zambia** (2000, Lusaka), and **Zimbabwe** (1996, Harare).

Different bodies of United Nations, namely, the UN-HABITAT through its Safer Cities Programme, the United Nations Office on Drugs and Crime (UNODC), and United Nations Development Programme (UNDP), have sponsored victimization surveys – sometimes with the support of the Dutch, the Belgium, or the Swedish government – in the following African countries, which do not have a national victimization survey (year and city from where the main sample was extracted are indicated between brackets): **Cameroon** (2002, Yaoundé; 2003, Douala), **Ghana** (2009, in the metropolitan areas of Sekondi-Takoradi, Accra, Kumasi and Tamale), **Ivory Coast** (1999, Abidjan), **Kenya** (2003, Nairobi; 2010, national representative sample), and **Rwanda** (2008, national representative sample).

Asia

Cambodia participated in the ICVS in 2000 with a sample of the capital, Phnom Phen, and a small rural sample. In 2011, the Survey of Cambodian Public Opinion, conducted by the International Republican Institute with a national representative sample, included a module on victimization. **Georgia** participated in the ICVS in 1992 and 1996 with national samples. The country also participated in the 2000 ICVS with a sample of the capital, Tbilisi. In 2010, 2011, and 2012, the Gorbi Institute carried out, with the financial support of the European Union, national victimization surveys using the questionnaire of the ICVS with some adjustments from the future European Union Security Survey. In **Hong Kong**, ICVS-based victimization surveys were conducted in 1990, 1995, and 1999. In **India**, the ICVS was conducted in 1992 and 1996 with samples of the capital, Mumbai. In 1996 a small rural sample was also included. In 2003, a victimization survey was conducted in four cities (Madurai, Coimbatore, Trichy, and Chennai) of the State of Tamil Nadu. The survey was funded by the University Grants Commission, New Delhi. In **Indonesia**, national victimization surveys were conducted in 1982, 1991, and 1994, as a part of its National Economic and Social Survey. The ICVS was conducted in 1989 with a sample of the city of Surabaya. In 1992 and 1996, the ICVS was repeated with a multiple-city sample accompanied by a small rural sample. **Japan** participated in the ICVS in 1989, 1992, 2000, and 2004 with national samples, and, since then, it has been using the ICVS questionnaire for national surveys every 4 years. In 2006, an independent victimization survey was conducted by Kyoto's Ryukoku University with a national random sample and a questionnaire based on the ICVS and British Crime Survey questionnaire. Surveys with samples of crimes victims were conducted in 1991–1992, 2000, and 2003. In South **Korea**, the Korean Institute of Criminology participated in the ICVS in 2000 and carries out a national victimization survey every 3 years since 1994. In **Taiwan**, victimization surveys were conducted in 2000 by the Prevention Institute of the Central Police University and in 2005 by the National Taiwan University. A survey with a sample of crime victims was conducted in 2005. In **Thailand**, a victimization survey was conducted in 1983 with a sample of the city of Bangkok.

The following Asian countries do not have a national victimization survey but have conducted the ICVS, usually with samples of their capital city complemented in some cases with a small rural sample (the year and the city from where the main sample was extracted are indicated between brackets): **Azerbaijan** (2000, Baku), **China** (1992, Beijing), **Kyrgyzstan** (1996, with a multiple-city sample), and **the Philippines** (1992 and 1996, Manila).

Conclusion

The first victimization question in a major public opinion survey seems to be the one included in a Gallup survey carried out in Finland in 1945. By the end of the 1960s, national victimization surveys were developed in the USA and the Nordic countries. The first ICVS in 1989 gave a major impulse to these surveys across the world. Currently, they are carried out regularly in the five continents, and many countries have their own national survey, conducted regularly. In particular, the regions of the world where national victimization surveys have become a systematic measure of delinquency are North America, the Nordic countries, Oceania, as well as some countries of Latin America and Western Europe.

Cross-national comparisons of independent national surveys can be performed – see, for example, the ones conducted by Langan and Farrington (1998) on the basis of the NCVS and the BCS and the studies compiled by Farrington et al. (2004) and Tonry and Farrington (2005) – but are usually problematic as the methodology of the surveys vary from country to country. The main methodological problems have been pointed out by Lynch (2006). In particular, differences in the labelling of the questions cannot be easily overcome. In this context, the increased use of the ICVS questionnaire – which was inspired by the victimization surveys conducted in the United States (namely, the one conducted by Gallup in 1972), the Netherlands, and the United Kingdom – has introduced a relative homogeneity that simplifies to some extent the task. However, the use of different methods of interviewing introduces a supplementary bias. Currently, participants in a survey can be contacted personally, by letter, by telephone, or by e-mail, and then give their answers to the questionnaire orally (by telephone or during a face-to-face interview) or in a written form (by filling a paper or an electronic questionnaire). Across the world, the technique most commonly used is the face-to-face interview, which can lead to PAPI (paper and pencil interview that can be filled by the respondent or the interviewer), CASI (computer-assisted self-interview), or, most frequently, CAPI (computer-assisted personal interview, in which the interviewer introduces the answers directly in a computer). Face-to-face interviewing has always been the preferred methodology in nonindustrialized countries, usually because their telephone and computer penetration rate are relatively low. On the contrary, in industrialized countries, the 1990s provided apparently the perfect window of opportunity for surveys conducted through CATI (computer-assisted telephone interview), as most of the households disposed of a fixed telephone line and the costs of such surveys were relatively low. Thus, during the first five waves of the ICVS (1989, 1992, 1996, 2000, and 2004/2005), most industrialized countries used this methodology.

The increased use of mobile phones instead of landlines and the growth of unwanted telephone marketing led to a fall of the response rates of victimization surveys conducted with the CATI methodology in industrialized countries, especially since the 2000s. For example, for the five countries that participated in the five waves of the ICVS (Canada, England and Wales, Finland, the Netherlands, USA), the rates increased from 52 % in 1989 to 64 % in 1996 but decreased to 63 % in 2000 and fell to 48 % in 2004, when up to seven attempts were made to reach the potential respondents. It is worth noting that response rates decreased also in some of the Central and Eastern European countries that used face-to-face interviews. For example, Poland registered a response rate of 96 % in 1992, but this rate decreased constantly and reached 72 % in 2004. Such decrease could be related to interview fatigue but also to an increased insecurity feeling that produced mistrust of strangers knocking at one's door, even if some countries announced the survey by letter. However, in Latin American countries, response rates are extremely high (e.g., 85 % in Mexico in 2011).

In order to improve response rates, some industrialized countries are currently testing CATI with samples that include mobile phones, and they are also introducing CAWI (computer-assisted web interview) as a new method. CAWI is sometimes combined with PAPI by inviting the persons, usually through a letter – a method that was seldom used due to its lower response rate – to fill the questionnaire either through the Internet or in a written form. Taken into account that the number of Internet users is growing constantly worldwide, there can be no doubt that the CAWI technique offers the promise of a new generation of affordable comparative crime surveys not only in industrialized nations but also in the rest of the countries. However, the results, in terms of differential nonresponse biases and pure mode effects of the few pilot surveys conducted until now with CAWI, are not satisfactory for its full implementation at the present time.

Another major problem faced by victimization surveys is related to the temporal localization of the events. Experimental research (Scherpenzeel 1992) has shown that higher annual victimization rates are found when respondents are asked directly about the victimizations suffered during the last 12 months (like in the BCS) than when they are first asked about victimizations suffered during the last 5 years and then are asked to indicate whether the event took place during the last year or before that (like in the ICVS). The NCVS, with its method of multiple interviews, also produces more accurate estimates of victimization rates and, compared to surveys conducted in other countries with a different methodology, should usually show lower levels of victimization in the USA.

Finally, small samples – as the ones usually used in Europe for the ICVS and in Latin America for the LAPOP and the Latinobarometro, which ranged between 1,000 and 2,000 households – can produce instable rates, especially for infrequent offences, which are usually the most violent ones (e.g., rape). Currently, the biggest samples are the ones used in The Netherlands (220,000 households in 2011), Mexico (90,000 households in 2011) followed by the United States (the NCVS used samples of roughly 40,000 households representing 70,000 interviews by the end of the 2000s), England and Wales (in 2010/2011, 51,000 people – including 4,000 children aged 10–15 – were interviewed for the BCS), Australia (29,000 households for the 2008–2009 MPHS), Chile (that since 2008 use a sample of 26,000 households for its annual ENUSC), and Canada (19,500 respondents for the GSS in 2009).

The best solution for all these problems is, evidently, to use the same questionnaire and the same methodology in different countries. At a continental level, the three regional victimization surveys applied in Latin America (Ecosocial, Latinobarómetro, and AmericasBarometer) use the same questionnaire but include only a few questions. In Western Europe, the ICVS has been applied usually with the same methodology of CATI in several countries. However, at the international level, the methodology used for carrying out the ICVS has not been homogeneous. Whenever the questionnaires or the methodology are fundamentally different, the best solution for comparative researchers is to concentrate on the trends shown by the surveys and not on the differences in victimization rates across countries.

Recommended Reading and References

Aebi MF, Linde A (2010) A review of victimisation surveys in Europe from 1970 to 2010. In: van Dijk J, Mayhew P, van Kesteren J, Aebi MF, Linde A (eds) Final report on the study on crime victimisation. Intervict/PrismaPrint Tilburg, Tilburg, pp D1–D76

Anttila I (1964) The criminological significance of unregistered criminality. Excerpta Criminol 4:411–414

Aromaa K (1971) Arkipaivan vakivaltaa suomessa [everyday violence in Finland]. Institute of Criminology, Helsinki

Biderman AD, Johnson L, McIntyre J, Weir A (1967) Report on a pilot study in the District of Columbia on victimization and attitudes to law enforcement. U.S. President's commission on law enforcement and administration of justice, field surveys I. U.S. Government Printing Office, Washington, D.C

Buikhuisen W (1975) Registered and non-registered criminality. Ministry of Justice/WODC, The Hague

Chambliss WJ (1994) Policing the ghetto underclass: the politics of law and law enforcement. Soc Probl 41(2):177–194

Chan W-C (ed) (2008) Support for victims of crime in Asia. Routledge, Oxon

Clinard MB (1977) Comparative crime victimization surveys: some problems and results. Int J Criminol Penology 6:221–231

Congalton AA, Najman JM (1974) New South Wales Bureau of Crime Statistics and Research, Unreported crime. Department of the Attorney

General & of Justice, NSW Bureau of Crime Statistics & Research, Sydney

Dammert L, Salazar F, Montt C, González PA (2010) Crimen e inseguridad: indicadores para las Américas. FLACSO- Chile/Banco Interamericano de Desarrollo (BID), Santiago, Chile

Ennis PH (1967) Criminal victimization in the United States: a report of a national survey. President's commission on law enforcement and administration of justice, field surveys II. U.S. Government Printing Office, Washington, D.C

Farrington DP, Langan PA, Tonry M (eds) (2004) Cross-national studies in crime and justice. Bureau of Justice Statistics, U.S. Department of Justice, Washington, pp 249–286

Lagan PA, Farrington DP (1998) Crime and justice in the United States and in England and Wales, 1981–96. Bureau of justice statistics executive summary. US Department of Justice, Bureau of Justice Statistics, Washington, DC

Lagos M, Dammert L (2012) La seguridad ciudadana: el problema principal de América latina. Corporación Latinobarómetro, Santiago de Chile

Lynch JP (2006) Problems and promise of victimization surveys for cross-national research. Crime Justice: Rev Res 34:229–287

Newport F (2010) 75 years ago, the first gallup poll. Polling matters, October 20, 2010. Retrieved online on June 4, 2012. http://pollingmatters.gallup.com/2010/10/75-years-ago-first-gallup-poll.html

President's Commission on Law Enforcement and Administration of Justice (1967) The challenge of crime in a free society. U.S. Government Printing Office, Washington, D.C

Rand M (2006) The national crime victimization survey: 34 years of measuring crime in the United States. Stat J U N Econ Comm Eur 23(4):289–301

Reiss AJ Jr (1967) Studies in crime and law enforcement in major metropolitan areas. President's commission on law enforcement and administration of justice, field surveys III, vol. 1. U.S. Government Printing Office, Washington, D.C

Sparks RF (1981) Surveys of victimization: an optimistic assessment. Crime Justice: Annu Rev Res 3:1–60

Tonry M, Farrington DP (eds) (2005) Crime and punishment in Western countries, 1980–1999 (crime and justice: a review of research 33). The University of Chicago Press, Chicago

van Dijk JJM, Steinmetz CJM (1979) The RDC crime surveys 1974–1979. Ministry of Justice/RDC, The Hague

van Dijk JJM, Mayhew P, Killias M (1990) Experiences of crime across the world. Key findings from the 1989 international crime survey. Kluwer Law and Taxation Publishers, Deventer

van Dijk JJM, Mayhew P, van Kesteren J, Aebi MF, Linde A (2010) Final report on the study on crime victimisation. Intervict/PrismaPrint Tilburg, Tilburg

Natural Crime Prevention

▶ Situational Crime Prevention and the Wild West

NCVS

▶ US National Crime Victimization Survey

Near Repeats

▶ Prediction and Crime Clusters
▶ Repeat Victimization

Near Repeats and Crime Forecasting

Shane D. Johnson and Kate J. Bowers
Department of Security and Crime Science,
University College London, London, UK

Overview

It has been recognized for some time that crime clusters at a range of spatial scales. It is also well established that offenses cluster in time such that crime occurrence is more likely at particular times of the day, week, month, or year. More recently, a growing body of evidence has started to accumulate that indicates that crime also clusters in time *and* space, such that when an event occurs at one location, there is a temporary elevation in the probability that other events will occur nearby. Where crimes occur at the exact same location, this is referred to as repeat victimization, and where they occur at nearby locations, *near-repeat victimization*. Finding that the risk of crime diffuses in space has implications for crime forecasting. However, this type of work is in its infancy and naturally, many questions remain unanswered. For example, there is something of a disconnect in methodological terms between those techniques that are used to detect crime patterns and those that are, or should be, used to generate predictive crime maps. Moreover, approaches to calibrating predictive maps so that a range of factors can be incorporated, and so that the extent to which crime risk might be diffused varies across areas, are things that have received little or no attention. Similarly, the extent to which patterns of one type of crime might affect those of another largely remains unexplored in the literature. A final point is that while crime is known to cluster in space, analyses of crime at the street segment level suggest that there exist sharp discontinuities in risk from one street to another. As most methods of crime forecasting use some form of smoothing algorithm, this issue is not accounted for in these models. The above issues and how thinking about methods of crime forecasting might be developed to accommodate them are discussed in this entry.

Introduction

Many things in life are random. Patterns of crime are not. Crime clusters at places, where places represent neighborhoods, street segments, or individual properties (for a review, see Johnson 2010). For many crimes, it is also evident that the timing of victimization is not random, with there being seasonal variation in the frequency of occurrence of offenses (Farrell and Pease 1994), and variation throughout the day (Ratcliffe 2002). Moreover, for burglary events, the elapsed time between events at the same home is typically short (Johnson et al. 1997). More recently, using (for example) techniques originally developed to detect disease contagion, research (see Table 1) has examined the association between the timing and location of crimes committed not just against the same target but also those nearby. The finding so far consistently observed is that when a crime occurs at one location, others are more likely to take place swiftly nearby. This finding applies to volume crime such as burglary, but also extends to other event types such as shootings (Ratcliffe and Rengert 2008) and insurgent activity (Townsley et al. 2008). In this entry, research concerned with this phenomenon – referred to as *near-repeat* victimization – is reviewed and consideration is given as to how this area of investigation informs methods of crime forecasting, and how it might usefully evolve. The entry is organized as follows. It begins with a discussion of methods of quantifying repeat and near-repeat victimization. Next, theoretical explanations as to why such patterns might emerge are considered. In the subsequent section, a number of unanswered research questions that might inform future work are identified. And, finally, there is a discussion of how what is already known has informed methods of crime forecasting and how such methods might be developed.

Fundamentals: Identifying and Quantifying Patterns of Near Repeats

A number of techniques have been used to examine patterns of repeat victimization of the same home (in the case of burglary) and near-repeat victimization. In the case of the former, the approach to analysis is somewhat simple. For repeat victimization proper, an incident of repeat victimization is defined simply as an offense where the victim has previously experienced an incident of the same type (or, crime more generally depending upon the researcher's interest). For this type of analysis, one may calculate the time elapsed between events and summarize this inter-event distribution. A typical finding is that the time elapsed between offenses is short, with the count of events declining logarithmically as a function of time. Figure 1 gives an example of such a time course (in this case, using fictional data).

One may also estimate whether the number of homes that were observed to have been victimized one, two, three, or more times differs from chance expectation, assuming that events are independent. The approach to analysis typically used is to estimate the number of homes that would be expected to be victimized 1 to N times, assuming the observed distribution can be explained by a Poisson process (the null hypothesis in this case). This type of analysis can be conducted assuming that the overall risk to homes

Near Repeats and Crime Forecasting, Table 1 Findings from selected studies of "near repeats" (studies in which a new method was introduced are highlighted with an asterisk)

Offense type	Study authors	City (Country)	Statistical technique	Time period	Space-time interval of effect
Burglary	Townsley et al. (2003)	Beenleigh (Australia)	Knox*	34 months	Clear effect but precise pattern unreported
	Johnson and Bowers (2004)	Liverpool (UK)	Knox, Mantel*	12 months	400 m and 28 days
	Bowers and Johnson (2005)	Liverpool (UK)	Descriptive, Mantel*	60 months	Neighbor effect, 1 week
	Johnson et al. (2007)	5 countries (AUS, NDL, NZ, UK, USA)	Monte Carlo Knox*	24 months	Varied, but always apparent for 200 m and 14 days
	Mohler et al. (2011)	Los Angeles, USA	Self-exciting point process*	24 months	100 m and 7–10 days
	Rey et al. (2011)	Mesa, USA	Markov Chains*	51 months	Adjacent locations and 1 month
Bike theft	Johnson et al. (2008)	Bournemouth (UK)	Monte Carlo Knox	12 months	300 m and 3 weeks
IED attacks	Townsley et al. (2009)	Iraq	Knox	3 months	1 km and 2 days
	Behelendorf et al. (2011)	Spain	Knox	144 months	5 miles and 2 weeks
IED attacks and counterinsurgency	Braithwaite and Johnson (2011)	Iraq	Bivariate Monte Carlo Knox*	6 months	Varied across event types
Shootings	Ratcliffe and Rengert (2008)	Philadelphia, USA	Monte Carlo Knox	13 months	400 ft and 14 days
Theft from vehicle	Johnson et al. (2009)	Bournemouth (UK)	Monte Carlo Knox	60 months	800 m and 14 days
Theft of motor vehicle	Lockwood (2010)	Lincoln, USA	Monte Carlo Knox	48 months	700 ft and 28 days

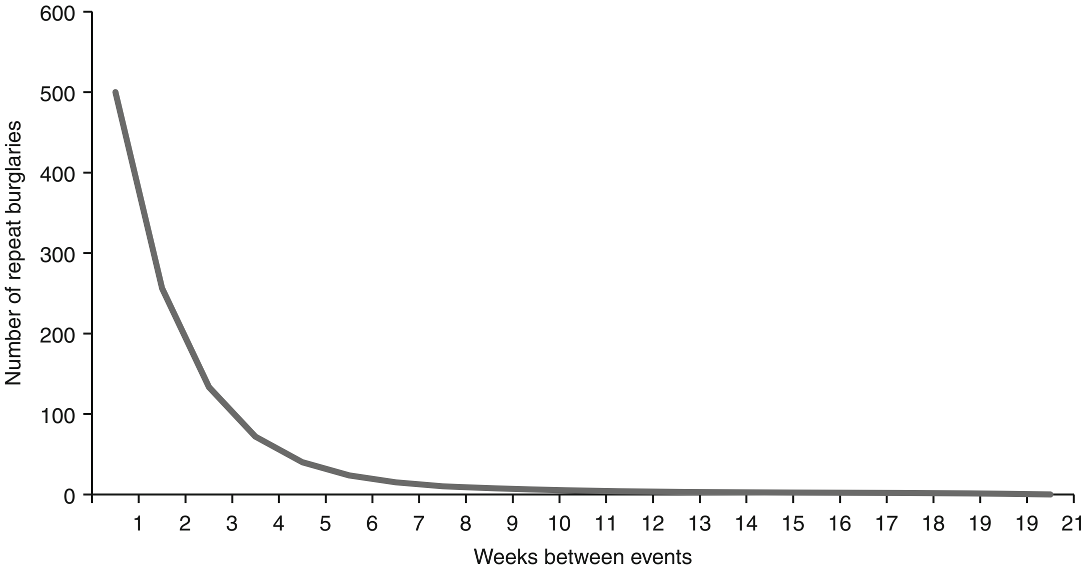

Near Repeats and Crime Forecasting, Fig. 1 Stylized illustration of the change in the risk of re-victimization risk over time is homogeneous or that there will be variation at the area (Trickett et al. 1992), neighborhood or house-level (Johnson 2008). A similar approach can also be used to estimate the expected time course of repeat victimization (Johnson 2008). In the case of burglary, what the available evidence – conducted across a variety of countries – suggests is that, relative to the pattern expected, the risk of victimization is temporarily elevated following an offense.

In terms of near-repeat victimization, the approach to analysis is slightly more complicated in that the analysis is extended to an examination of any regularity in the patterning of crimes that occur not just at the same locations, but also nearby. A number of approaches to analysis have been adopted. In one study (Bowers and Johnson 2005), using data recorded by the police (Merseyside, UK) that covered an interval of 5 years, the authors identified every street on which two or more burglaries had occurred. For each burglary, they then calculated how many doors apart that burglary had occurred from the most recent burglary that had also taken place on the same street. A distribution was then derived to indicate how many sequential burglaries had occurred one-door apart, two-doors apart, and so on. This distribution was standardized to account for the fact that because roads are not infinite in length, there would be more opportunities for burglaries to occur (say) one-door apart from each other than (say) 30 doors apart. The results clearly indicated a pattern of distance decay, such that when burglaries occurred on the same street as each other, they were most likely to occur near to each other. Further analyses also indicated that burglaries that occurred on the same street were also more likely to occur on the same side of the street as each other and near to each other in time. Thus, when a neighbor of a victimized home was burgled, this was more likely to occur within 1 week of the first offense than (say) 7–14 days apart.

Such analysis examines patterns observed on the same street, but does not consider those offenses that occur near to each other but on a different road. For this reason, other methods of analysis, originally developed to detect disease contagion, have also been used. A number of studies have used a technique developed by Mantel (1967). This test is used to see whether crimes that occur closest to each other in space are also more likely to occur close to each other in time. This particular test provides a single index of space-time clustering for a particular area, and while it has useful diagnostic value, it is a relative measure which means that what is defined as near (in time or space) is not actually specified. That is, space-time clustering would be detected even if it were the case that the events that were closest to each other were (say) two miles apart, if that is, relative to events that occurred further apart, these crimes also tended to occur closest to each other in time.

A test developed by (Knox 1964) provides an alternative approach. In this case, each event of interest is compared with every other and the absolute distance and time elapsed between them is computed. In the original formulation, a 2 × 2 contingency table was populated to enumerate how many pairs of events had occurred close to each other in space and time; close in space but not in time; close in time but not space; and neither close in space nor time. This test allows the researcher to define what is meant by "close" in both dimensions and hence provides more precise insight into observed patterns. To determine whether the latter differ from chance expectation, assuming the timing and location of events are independent, the expected counts are calculated for each cell by multiplying the total counts for the relevant row and column and then dividing by the total across all cells of the table (the same approach used to compute expected counts for a Chi-Square statistic).

Such an analytic strategy may be sufficient for estimating the extent to which a disease is contagious, and for examining general patterns in the space-time dynamics of crime, but it does not provide a particularly precise picture of crime patterns. This is so because only two categories of near and far are defined. For this reason, in recent work (Johnson and Bowers 2004a), the approach has been extended so as to include numerous categories. In general, around 10–20 different spatial and temporal bandwidths have been used. For example, Johnson and Bowers (2004a) used spatial bandwidths of 100 m and temporal intervals of 7-day periods. Patterns observed across these relatively finely resolved categories can then be compared with each other and with a more general “distant” category.

One potential weakness with the approach to statistical inference used for this test is that the test statistics are based on parametric assumptions that might not be met. For this reason, Johnson et al. (2007a) developed a nonparametric version of the test that uses a Monte Carlo simulation to estimate the statistical significance of observed patterns. As was the case for previous work, the authors found that burglary clustered in time and space more than would be expected if the timing and location of events were independent. Importantly, in that study, the authors found consistent patterns within each of two cities in each of five different countries, suggesting that the patterns are not specific to a particular locality.

A further refinement in the existing research involves a more precise study of the distribution of space-time clusters. To elaborate, for the techniques discussed above, the approach to analysis involves counting the number of pairs of events that occur within particular spatial and temporal intervals of each other and comparing these counts to expectation, assuming that the timing and location of events are independent. In an early study, Johnson and Bowers (2004b) extended the approach by examining whether pairs of events that occur close to each other in both space and time also cluster. They did this by computing the number of event pairs (hereafter, close pairs) that occurred within a given space and time interval (100 m and 1-month) of each other for a series of 151 neighborhoods, for each month of a 1-year period. Their analyses showed that in general, the number of close pairs that occurred in a neighborhood was similar for sequential monthly intervals, but that the correlation was lower and typically nonsignificant where the counts for intervals that were more than one month apart were compared. Similarly, it was evident that if a neighborhood had a high count of close pairs in a particular month, those nearby also tended to also have a high count of close pairs around the same time. There was no systematic association between the count of close pairs in neighborhoods that were a fair distance apart, regardless of whether comparisons were made for the same month or other intervals.

A slightly different approach was adopted in other work (Johnson et al. 2007b). In that study, for each burglary event, the authors calculated how many other burglaries had previously occurred nearby in the recent past. The idea of doing so was to estimate the number of events that might typically make up a *spate* (or *poly-order chain*) of burglary events. In a further study, this time concerned with Improvised Explosive Device (IED) attacks in Iraq, Johnson and Braithwaite (2009) used a Monte Carlo method to compare the observed length of spates with those expected, assuming that the timing and location of events were unrelated. In this study, relative to expectation, it was apparent that while there were significantly more spates that comprised six or more IED attacks, there were actually significantly fewer spates with seven or more offenses. That is, after six IED attacks had occurred at or near to a location shortly after each other, others were actually less likely to occur than would be expected if the timing and location of attacks were independent. Collectively, these findings suggest that when a crime occurs at a location, others are more likely to occur nearby in the near future. However, as time elapses, this increase in risk decays.

Following the development of methods to quantify patterns of near repeats, a number of studies have been conducted and Table 1 provides a summary of some of the most salient findings to date. Between the period 2003 and 2011, more than 30 studies of which the authors are aware have so far been published on this topic and so what should be apparent is that while the analysis of this type of patterning for crime events is relatively new, there now exists a considerable body of research for a range of different types of event and for different countries.

Possible Explanations for Near-Repeat Patterns

Much of the research concerned with spatial patterns of crime ignores the dimension of time, focusing exclusively on spatial patterns aggregated over some interval. Such analysis might be described as examining the stable or slow dynamics of crime. In contrast, research concerned with near repeats explicitly examines both spatial and temporal patterns, and does so at a fairly fine level of resolution. As such, this kind of analysis might be described as examining the fast dynamics of crime.

One question that remains unanswered is how the fast and slow dynamics of crime influence each other. For instance, do the fast dynamics give rise to the slow ones, or vice versa, or, is the association more complex? Given that many hotspots endure it would seem unlikely that the fast dynamics entirely define the slow ones, and so disentangling this issue is of clear importance. Controversies such as this will be discussed in a later section.

Apropos why repeat and near-repeat victimization occur, two theories have been proposed to explain observed patterns. According to the first, variation in enduring characteristics that makes targets vulnerable to crime explains the patterns described. This concept of *risk heterogeneity* (see Johnson 2008) assumes that – other than the target involved – there is no association between one offense and another. The risk of victimization across targets is seen as being time-stable and a function of a set of characteristics that many or all offenders would find appealing.

Considering offender spatial decision making, research has generally focused on the types of place that offenders might display preferences for. Underlying such models is the idea that offenders, like the rest of the population, are rational decision makers (Clarke and Cornish 1985) who weigh up the benefits, costs, and effort associated with possible courses of action. Such decision making is not seen as involving a comprehensive evaluation but is regarded as being more heuristic in nature, with offenders using “rules of thumb” to assess opportunities perceived. Of course, offenders will rarely be aware of the true costs, benefits, and rewards of any particular decision made, and so their decision making is instead considered to be bounded by the information of which they are aware, cognitive biases, and so on. However, the point of central importance is that in most of the research literature, offender decision making is typically considered in terms of spatial choices. These may be hierarchical, insofar as certain types of areas may first be selected, after which particular targets – which may confirm to cognitive templates – are selected, but they are nonetheless spatial.

Research on near repeats draws more explicit attention to the dynamic nature of offender spatial decision making. For example, the offender as forager hypothesis (Johnson and Bowers 2004b; Johnson et al. 2009b) borrows from concepts regarding animal foraging, ideas that are well established within the field of ecology. The obvious reason for the comparison being that the activities considered, be they animal foraging or offender decision making, share common underlying goals. Both seek to maximize benefits, while minimizing expended effort and the associated risks (be it detection or being eaten). Moreover, both forms of activity are subject to constraints. The actor, be it animal or offender, can only move so far and so fast per unit time, and both expend effort in doing so. In neither case would it seem to be efficient to make future choices without reference to previous ones, particularly those made most recently. Where recent choices have been rewarded, their repetition would seem reasonable, albeit up to a certain point. In the case of the offender, returning to the same location repeatedly would likely ultimately either lead to arrest or little reward – as little would be left to take. The targeting of locations near to successfully exploited targets may represent an efficient strategy as the first law of geography (Tobler 1970) states that things that are close to each other are more similar than are those that are far apart. Thus, nearby targets are likely to share important features (e.g., levels of natural surveillance, escape routes, etc.) of those previously successfully targeted that will make them good opportunities for crime, at least in the short term. The latter point is important as things may be subject to change, and memory fades.

Thus, in contrast to theories of risk heterogeneity, those concerned with *event dependency*

(Pease 1998) assume that victimization influences future risk. In the case of crimes such as burglary, where the conditions at already targeted locations are conducive to crime, it is hypothesized that the same offender (or group of offenders) will return and that they will do so swiftly.

A number of research methods have been used to test the two types of theory. The analysis of crimes detected by the police (Bernasco 2008; Johnson et al. 2009b) indicates that, relative to burglaries that occur close in space *or* time to each other, those that occur nearest to each other in both dimensions are massively more likely to be the work of the same offender(s). The same is true for theft from motor vehicles (Johnson et al. 2009b). Such patterns would be predicted by the offender as forager hypothesis but are inconsistent with theories of risk heterogeneity.

Interviews with offenders also suggest that offenders return to the same locations (Ashton et al. 1998) and those nearby (Summers et al. 2010) for reasons compatible with the offender as forager hypothesis. Of course, such research also indicates that target attractiveness plays a role; an offender can only return to an area or street they have already targeted, and so as a minimum, risk heterogeneity will play a role in determining where offenders commit initial offenses. Moreover, not all offenders will adopt foraging strategies and for such offenders, target attractiveness may be the primary factor that they consider when selecting targets.

In the case of repeat burglary victimization, it is apparent that relative to low frequency offenders, prolific offenders who are more established in their criminal careers appear to be more likely to return to the exact same properties (Everson and Pease 2001). It is unclear whether the same applies to those who commit near repeats, but this would seem to be plausible. If this is the case, then it would suggest that target attractiveness (risk heterogeneity) alone may explain the targeting strategies of those burglars who infrequently commit crime, but that those who routinely engage in such offenses, the experiences associated with the burglaries they have recently committed will significantly inform their subsequent choices.

A different approach to theory testing and falsification than has been discussed hitherto uses mathematical models or computer simulation. The detail is not important, but the central idea is to see if models that describe a particular data-generating process – such as event dependency or risk heterogeneity – are sufficient to generate space-time patterns of events that resemble those observed in crime data. This research indicates that models that include only – or completely exclude – risk heterogeneity are incapable of generating the time course of repeat victimization (Johnson 2008) or near-repeat victimization (e.g., Pitcher and Johnson 2011). However, those that include both processes generate patterns much like those observed in the research described above.

Mohler et al. (2011) have also used a class of statistical model, originally developed to examine patterns of earthquakes, to examine patterns observed for the crime of burglary. The aim of this type of statistical model, known as a *self-exciting point process*, is to estimate the contribution of (unobserved) time-stable factors that vary across space (e.g., the location and size of unknown fault lines in the case of earthquakes) and the effects of after-shocks that follow each quake. Using data for burglaries recorded by the police in Los Angeles, Mohler et al. (2011) find that for this event type at least, both types of process are important. Thus, it would seem that both types of explanation – risk heterogeneity and event dependency – have a part to play.

Before leaving this discussion, it is worth noting that one methodological issue that is sometimes discussed with respect to patterns of repeat or near-repeat victimization is that these might be explained by an artifact associated with the data. It is widely acknowledged that police-recorded crime data is by no means a perfect measure of crime, as much crime goes unreported. Moreover, from the perspective of when crimes occur, it is often difficult to pinpoint the exact time of a burglary as this will be unknown and so estimation procedures are often necessary. However, these issues are unlikely to influence the results of the types of analysis discussed unless there is a systematic pattern (in space and time) associated with them, and there is no reason to suspect that there is.

A potentially greater concern involves edge effects. In the case of near-repeat victimization, burglaries that occur at or near the boundary of a study area have more opportunity to be further away from other burglaries than do those in the middle; likewise, burglaries that occur at the start of a time period have more potential to be further apart in time from other crimes than do those that occur at the end. However, because the expected distribution of events is established using the observed data and methods such as the Monte Carlo procedure discussed above, we suggest that the approach to analysis accounts for edge effects (and other issues).

What Don't We Know About Near-Repeat Patterns?

Although much has been learned about near repeats over the last 8 years, a number of things are currently unclear. For example, while the research to date suggests that both theories of risk heterogeneity and event dependency have a part to play in explaining space-time patterns of crime, the precise contributions of each process are not well understood. For instance, as discussed, in the case of burglary, it could be that some burglars tend to select targets purely on the basis of their perceived attractiveness, whereas others select targets in a way that is consistent with the optimal foraging theory discussed above. Others still may adopt a more even mix of strategies. An alternative explanation is that all offenders adopt the same strategy, but that some burglary events encourage a swift return to the same property or those nearby, whereas in other cases, the reverse is true. For instance, if an offender is challenged by a resident while attempting to burgle a home, that offender is likely to be discouraged from returning to the same neighborhood for some time. To put it another way, at what unit of analysis – the offender or the event – do these two types or processes (risk heterogeneity and event dependency) operate?

The preceding discussion focused on the crime of burglary, but as is evident in the introduction to this entry, near repeats are observed for a range of event types and it is likely that different mechanisms will be responsible for observed patterns for some of these. In the case of similar crimes, common explanations may suffice but is the same true for very different types of events?

While this seems unlikely, there may be some commonalities associated with human mobility dynamics. For instance, most actors are subject to constraints that limit how far they can travel per unit of time and how fast they can do so. It may also be reasonable to assume that actors are more likely to engage in activities at locations with which they are familiar than those that they are not, regardless of the event type. However, while there may be some commonalities in the explanations for different event types, understanding the differences will be important if the patterns are to be truly understood.

Little research has considered how patterns, and the contribution of the two theoretical processes (risk heterogeneity and event dependency) discussed, might vary across space and/or time, but there are two exceptions. Townsley, Homel, and Chaseling (2003) found differences in patterns of repeat and near-repeat victimization across a sample of five different suburbs and found that near-repeat victimization was more likely in those suburbs with more homogeneous types of housing. They interpret this as suggesting that where replicas of a previously victimized home exist, offenders will be more likely to commit near repeats. Bowers and Johnson (2005) examined patterns of repeat and near-repeat burglary victimization for a larger sample of 118 areas in one county in the UK. They found the former to be accentuated in deprived areas, but that patterns of the latter were clearer in affluent ones. However, other than these two early studies, little research has examined how patterns vary across spatial units, or if they are stable over time. It is possible, for example, that in one hotspot of criminal activity, there may be a strong pattern of near repeats, whereas in another, there will be no such pattern. Moreover, patterns of near repeats may ebb and flow over time, perhaps with the seasons (Johnson and Bowers 2004b) or perhaps as a consequence of certain types of offenders moving into or out of an area. Studying such dynamics is likely to inform understanding and assist in efforts of crime forecasting.

Given previously well-documented offender versatility, it may be expected that the space-time distribution of different types of offense would be correlated. Only one study (Johnson et al. 2009b) has so far examined near-repeat patterns across offense types. In that study, patterns of burglary and theft from vehicle were considered and it was found that for the two types of event, patterns of victimization were *not* correlated in space and time. Further analyses of crimes detected by the police for the same area indicated that offenders who committed one type of offense typically did not commit offenses of the other kind. This latter finding explains why the space-time distributions of the two types of offenses were not found to be correlated, but are at odds with other research concerned with offender versatility. Thus, it is not yet clear whether the finding regarding space-time clustering across offense types is typical across other localities or if it holds true for other combinations of offense types.

Another issue which requires attention is how co-offending or offender cooperation might contribute to (or interfere with) patterns of repeat or near-repeat patterns. Empirical research on co-offending more generally is in its infancy (Andresen and Felson 2010) and so this may be a particularly useful area to consider. For example, it is possible that offenders may share information regarding good opportunities, or (for example) work streets together in some cases. If this is the case, then the "boost" hypothesis would still explain observed patterns, but establishing the extent to which offenders cooperate in this way would inform criminological understanding and have implications for police practice. An alternative is that where offenders work together, the patterns are quite different to when they offend alone. Those that co-offend may adopt quite different strategies to the foraging approaches discussed above. Regardless of the reality, it would be beneficial to examine this.

A point that is discussed below concerns what the appropriate units of analysis for research of this kind should be. For the Knox and similar forms of analysis, the distance between events is measured using either the Euclidian (straight-line distance) or Manhattan (the sum of the difference between two points in the horizontal and vertical dimensions) distance between them. Given that many offenses occur along a street, this may be inappropriate or it may mask important variation. For example, two homes may be (say) 100 m from each other in Euclidian space, but be separated by a much more substantial distance along the street network. Thus, a further question that has not been addressed to date is how risk spreads along the street network as opposed to within streets or across Euclidian space. Researchers might ask if risk spreads preferentially from streets that are connected to each other, if such a diffusion of risk is more likely along similar types of streets, and if there are other regularities associated with when risk spreads from one street to another and when it does not.

Predicting Future Crime Locations

While there is still much that is unknown about space-time patterns of crime, current findings have practical application and inform criminological theory. The finding that following one crime at a particular location, the risk of others occurring nearby is temporarily elevated has obvious implications for crime prevention and crime forecasting. In this regard, recent work has examined the forecasting accuracy of simple mathematical models that take account of both the timing and location of historic events, and compared this with methods that consider only where historic events occurred (regardless of when they took place). For both types of model, the risk of crime is assumed to be a function of how much crime previously occurred at that location and nearby. In the case of models that take account of when and where crimes occurred, the reason is that the risk of crime is assumed to diffuse in space much like a contagious disease. And, consequently, more recent events are seen as more important and thus given a larger mathematical weighting in the models. In the case of those models that do not consider the timing of events, the theoretical rationale for considering the location of events that occurred near to, as well as at, each location has not been explicitly articulated, but the research concerned with near-repeat victimization at least provides a justification for this.

A typical method of contrasting the accuracy of such models may be described as follows. First, a study area is represented by a grid of regular sized cells (say 50 m by 50 m in dimension). Second, predictions are generated for each cell using data for a training interval. To calculate the expected risk in each cell, a mathematical formula is used which considers how much crime occurs in each cell and within some (usually) fixed distance from it. What that distance or *bandwidth* is often is selected for atheoretical reasons, such as the visual elegance of the resulting map. However, the research concerned with near repeats provides a theoretically meaningful way of determining the bandwidth. For example, the Knox analysis can be used to determine the distance and time over which the risk of crime appears to diffuse in space and time, and hence, the distance and time over which each crime might be expected to affect the risk of crime at nearby locations. Regardless of the approach to selecting the bandwidth, all of the methods typically apply a weighting to each event that is inversely related to the distance the event occurs from the center of the cell. Methods that take account of the timing of events additionally apply a weighting that considers when the events occurred. In this case too, a bandwidth is selected and the weighting applied to events within that bandwidth is usually inversely related to the time elapsed between that event and the period for which the forecast is generated.

After generating predictions for a particular model, these are compared with observed outcomes for a series of subsequent forecast intervals. In the analyses so far conducted (e.g., Bowers et al. 2004; Johnson et al. 2007b, 2009a) for the crime of burglary, those models that take account of both the timing and location of burglary outperform those that consider only the location of previous events. In a recent study, Mohler et al. (2011) show that statistical methods that explicitly estimate the likely contributions of time-stable (estimated using crime data alone) and event-dependent processes improve the forecast accuracy of such methods.

Other work (e.g., Johnson 2008) has examined how forecasting models can be improved by considering time-stable characteristics of the environment in a different way. In that study, the researchers incorporated simple data regarding the distribution of potential targets (houses) of burglary and the street network. Again, improvements in forecast accuracy were observed relative to models that exclude such information.

Future Developments in Prediction

Work on predictive analysis is, however, embryonic. For example, information on the street network has so far been limited to an indication of where roads are physically located. As discussed above, the risk of crime may diffuse along the street network in particular ways. Even in the most basic sense, if risk diffuses along the street network rather than through geographical space, then the topology (the ways in which segments of the street network are or are not connected to each other) of the network may significantly affect how the risk of crime spreads. For example, if two cells in an area were near to each other but not connected by the street network, then it is possible that risk may not diffuse between them, or that any propagation is impeded relative to two homes that are the same distance apart but connected (all other things being equal). In addition, even if two cells are connected via the street network, while they might be close in Euclidian space, they may be far apart in topological terms. The extent to which a consideration of topology might play a role in improving the accuracy of crime forecasting is currently unknown but amenable to mathematical analysis.

As discussed above, the way in which crime risk might spread in space *and* time might well vary across different neighborhoods. In the analyses so far conducted, patterns are modeled in the same way across areas and so gains in predictive accuracy might be achieved by systematically studying how patterns vary across areas and incorporating such findings into predictive models. Similarly, patterns might vary in time. For example, near repeats might be more likely at certain times of the day or even the year. Such variation is plausible as the risk of crime more generally is known to fluctuate throughout the day (Ratcliffe 2002) and year, presumably reflecting (amongst other things) variations in people's routine activities and hence their availability to act as capable guardians against crime. In line with this suggestion, Johnson et al. (2007) find that when pairs of events occur near to each other in both space and time, they are more likely to occur during the same interval of the day than chance expectation. However, with the exception of this study, no research of which of the authors are aware has examined how the risk of near-repeat victimization might vary over the course of the day and how predictable patterns might be.

One issue not so far discussed is that the types of analyses generally conducted to detect patterns of space-time clustering focus on point patterns (the precise locations of crime events), whereas the methods used to generate predictions do so with reference to a fixed grid. The potentially important issue is that for point patterns, space-time clustering may be generated in a number of ways. For example, space-time clustering might occur if many pairs of events happen close to each other in space and time and are all located in an area surrounding a single cell – let us refer to this as cell A. Space-time clustering might also occur if sequential pairs of events happen near to each other but if, over time, the pairs of events generally move away from cell A. Figure 2 represents these possibilities. It should be clear that the risk to cell A should be modeled differently if the first type of pattern were typical than if the second type were dominant. However, it is not possible to differentiate between these two types of directional pattern using the kinds of analyses that have been traditionally used.

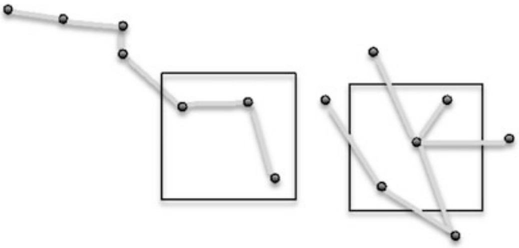

Near Repeats and Crime Forecasting, Fig. 2 Potential near-repeat point patterns surrounding a grid cell

There appears to be value in matching units of analysis used to determine space-time patterns with those used in prediction. In two recent papers (Mohler et al. 2011; Rey et al. 2011) concerned with space-time clustering, the units of analysis were grid cells rather than point locations. For example, Rey et al. (2011) divided their study area into ¼ by ¼ mile grid cells and calculated how many burglaries occurred in every cell each month for a period of 51 months. Next, using a discrete Markov Chain framework, for each month, they examined the probability with which each cell would transition from a state of experiencing no burglaries or a state of experiencing at least one to the same or the alternate state. Consistent with research that has examined point patterns, their findings clearly showed that the *state* of a cell (whether crimes occurred in it or not) at any interval was positively associated with the state of that cell and those adjacent in the previous period. Moreover, that this pattern could not be explained by risk heterogeneity alone. This, or similar approaches, may help to more precisely calibrate methods of crime forecasting that take account of both the timing and location of events.

In addition to the issue of matching units of analysis for description and prediction, there is still the underlying question of what the relevant unit of analysis should be in crime forecasting of this kind. Grid-based predictions are largely used because of computational convenience. That is, constructing a grid of regular sized cells is straightforward and the spatial relationship between each cell and every other is simple to establish. However, crimes do not occur on a grid (except where the street network is of that geometric configuration), and offenders are unlikely to engage in decision making with reference to such a representation of the world. Moreover, research suggests that there are sharp discontinuities in the distribution of crime risk, such that (for example) two roads that are very near to each other (even directly connected) may experience very different risks (Groff et al. 2010). Possible reasons for this variation are numerous but the point of central importance is that – whatever the explanation – something about the way those street segments differ is associated with variation in crime risk. This is also true of block faces in a US context which in essence are equivalent to street segments in Europe. Where risk is modeled on a grid surface, this variation is ignored as the grid cells are unlikely to reflect the geometry of the street network. Thus, the risk for the cell represents the average risk across those street segments within it, even though the places within the cell may differ considerably in factors including crime risk. Moreover, the smoothing algorithms used to derive grid-based predictions will mean that risk is assumed to diffuse from one location to any another nearby location, irrespective of the characteristics of the type of location, something that may be unreasonable.

Thus, it is suggested that future research should explore methods of crime forecasting using street networks or block faces rather than grid cells. Methods of describing and representing street networks are possible using basic techniques from a branch of mathematics known as graph theory, and that exploring this type of approach is an obvious next step in the research agenda.

In closing, a further question to consider as yet unasked in the research literature is precisely what does one want to predict with forecasting models? To make things concrete, consider those forecasting models developed to predict patterns of burglary. As discussed, the available evidence suggests that observed patterns are most likely the result of two data-generating processes: risk heterogeneity and event dependency. If, as has been tentatively suggested, these two processes reflect the dominant targeting strategies of two different types of offender, the patterns and locations of offenses generated by one group of offenders might differ from those of the other. If so, should the accuracy of forecasting models be evaluated with respect to the total amount of offenses correctly predicted, or should such assessment also consider what types of events (e.g., near-repeat victimizations) are accurately identified? One reason for asking such a question is that the types of crime reduction action that might be triggered by the two types of prediction – those that might forecast the locations of events most likely committed by those who adopt foraging strategies as opposed to those who adopt more opportunistic strategies – may well differ. If this proves to be the case, for events that may be considered most likely the result of the consequence of enduring characteristics of places (assuming such a distinction is possible), situational crime prevention interventions will likely represent the most productive approaches to crime reduction. In contrast, where events predicted are more likely to be the work of offenders who adopt optimal foraging strategies, efforts directed toward detecting (or preventing) future offenses will likely prove more fruitful. In operational terms, the development of methods of forecasting that are specifically tailored to these two types of policing may be more helpful than are those models that predict more offenses (of any kind). The extent to which this is true is, of course, an empirical question.

Related Entries

- ▶ Crime Mapping
- ▶ Crime Science
- ▶ History of Geographic Criminology Part I: Nineteenth Century
- ▶ Hot Spots and Place-Based Policing
- ▶ Prediction and Crime Clusters
- ▶ Residential Burglary

Recommended Reading and References

Andresen MA, Felson M (2010) The impact of co-offending. Br J Criminol 50(1):66

Ashton J, Brown I, Senior B, Pease K (1998) Repeat victimisation: offender accounts. Int J Risk Secur Crime Prev 3:269–280

Bernasco W (2008) Them Again? Same-Offender Involvement in Repeat and Near Repeat Burglaries. Eur J Criminol 5(4):411–431

Bowers KJ, Johnson SD (2005) Domestic burglary repeats and space-time clusters. Eur J Criminol 2(1):67

Bowers KJ, Johnson SD, Pease K (2004) Prospective hot-spotting. Br J Criminol 44(5):641

Clarke RV, Cornish DB (1985) Modeling offenders' decisions: a framework for research and policy. Crime Just 6:147

Everson S, Pease K (2001) Crime against the same person and place: detection opportunity and offender targeting. Crime Prev Stud 12:199–220

Farrell G, Pease P (1994) Crime seasonality: domestic disputes and residential burglary in Merseyside 1988–90. Br J Criminol 34(4):487

Groff ER, Weisburd D, Yang SM (2010) Is it important to examine crime trends at a local "micro" level?: a longitudinal analysis of street to street variability in crime trajectories. J Quant Criminol 26(1):7–32

Harries KD, Stadler SJ, Zdorkowski RT (1984) Seasonality and assault: explorations in inter-neighborhood variation, Dallas 1980. Ann Assoc Am Geogr 74(4):590–604

Johnson SD (2008) Repeat burglary victimisation: a tale of two theories. J Exp Criminol 4(3):215–240

Johnson SD (2010) A brief history of the analysis of crime concentration. Eur J Appl Math 21(4–5):349–370

Johnson SD, Bowers KJ (2004a) The burglary as clue to the future. Eur J Criminol 1(2):237

Johnson SD, Bowers KJ (2004b) The stability of space-time clusters of burglary. Br J Criminol 44(1):55

Johnson SD, Braithwaite A (2009) Spatio-temporal modelling of insurgency in Iraq. Criminal Justice Press, New York

Johnson SD, Bowers K, Hirschfield A (1997) New insights into the spatial and temporal distribution of repeat victimization. Br J Criminol 37(2):224

Johnson SD, Bernasco W, Bowers KJ, Elffers H, Ratcliffe J, Rengert G, Townsley M (2007a) Space–time patterns of risk: a cross national assessment of residential burglary victimization. J Quant Criminol 23(3):201–219

Johnson SD, Birks DJ, McLaughlin L, Bowers KJ, Pease K (2007) Prospective crime mapping in operational context, final report. Home Office online report, London, UK, 19, 07–08

Johnson SD, Bowers KJ, Birks DJ, Pease K (2009a) Predictive mapping of crime by ProMap: accuracy, units of analysis, and the environmental backcloth. In: Weisburd WBD, Bruinsma G (eds) Putting crime in its place. Springer, London, pp 171–198

Johnson SD, Summers L, Pease K (2009b) Offender as forager? A direct test of the boost account of victimization. J Quant Criminol 25(2):181–200

Knox G (1964) Epidemiology of childhood leukaemia in northumberland and durham. Brit J Prev Soc Med 18:17–24

Lockwood B (2012) The presence and nature of a near-repeat pattern of motor vehicle Theft. Secur J 25:38–56

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27(2 Pt 1):209

McGloin JM, Sullivan CJ, Piquero AR (2009) Aggregating to versatility? Br J Criminol 49(2):243–264

Mohler G, Short M, Brantingham P, Schoenberg F, Tita G (2011) self-exciting point process modeling of crime. J Am Stat Assoc 106(493):100–108

Mohler GO (2011) Self-exciting point process modeling of crime. J Am Stat Assoc 106:100–108. doi:10.1198/jasa.2011.ap09546

Morgan F (2001) Repeat burglary in a Perth suburb: indicator of short-term or long-term risk? Crime Prev Stud 12:83–118

Pease K (1998) Repeat victimisation: taking stock. Home Office, London, p 48, ISBN 1-84082-089-6

Pitcher AB, Johnson SD (2011) Exploring theories of victimization using a mathematical model of burglary. J Res Crime Delinq 48(1):83

Ratcliffe JH (2002) Aoristic signatures and the spatio-temporal analysis of high volume crime patterns. J Quant Criminol 18(1):23–43

Ratcliffe JH, Rengert GF (2008) Near-repeat patterns in Philadelphia shootings. Secur J 21(1):58–76

Rey SJ, Mack EA, Koschinsky J (2011) Exploratory space–time analysis of Burglary patterns. J Quant Criminol 28:1–23

Sagovsky A, Johnson SD (2007) When does repeat burglary victimisation occur? Aust N Z J Criminol 40(1):1

Short M, D'Orsogna M, Brantingham P, Tita G (2009) Measuring and modeling repeat and near-repeat burglary effects. J Quant Criminol 25(3):325–339

Summers L, Johnson SD, Rengert G (2010) The use of maps in offender interviews. In: Bernasco W (ed) Offender on Offending: Learning About Crime from Criminals. Cullompton: Willan

Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. Econ Geogr 46:234–240

Townsley M, Homel R, Chaseling J (2003) Infectious burglaries. A test of the near repeat hypothesis. Br J Criminol 43(3):615

Townsley M, Johnson SD, Ratcliffe JH (2008) Space time dynamics of insurgent activity in Iraq. Secur J 21(3):139

Trickett A, Osborn DR, Seymour J, Pease K (1992) What is different about high crime areas? Br J Criminol 32(1):81

Near-Repeat Victimization

▶ Repeat Victimization

Negotiated Guilty Pleas

▶ Plea Bargaining

Neighborhood Crime Trends

▶ Longitudinal Crime Trends at Places

Neighborhood Effects

▶ Social Network Analysis and the Measurement of Neighborhoods

Neighborhood Effects and Social Networks

Brian Soller[1] and Christopher R. Browning[2]
[1]Department of Sociology, University of New Mexico, Albuquerque, NM, USA
[2]Department of Sociology, The Ohio State University, Columbus, OH, USA

Synonyms

Communities and crime; Community organization; Neighborhood social networks

Overview

Features of neighborhood social networks are central to many of the most prominent theories of crime. Empirical tests of a number of criminological theories underscore the importance of neighbor networks in both facilitating and deterring crime, both at the individual and community levels. However, most studies that focus on the link between neighborhoods and crime are limited by the use of "perceptual" and other proxy measures that aim to capture neighbor cohesion, social capital, and other social processes thought to mediate the association between neighborhood structural characteristics (e.g., concentrated disadvantage, residential instability) and offending. Exactly which features of neighborhood social organization are the most consequential for crime remains unknown due to the paucity of relational data that capture theoretically relevant features of actual neighborhood network structures. Furthermore, the mechanisms through which social networks influence neighborhood crime remain underexplored.

Despite the significant challenges associated with collecting neighborhood-based social network data, great strides have been made in the study of neighborhood networks in recent years. This entry provides a brief overview of past research on neighborhood networks and crime and highlights notable recent advances in the conceptualization and measurement of neighborhood-based social networks in empirical research.

Whether explicit or implied, several of the most prominent criminological theories focus on the link between social networks and individual and aggregate offending patterns. For instance, theoretical constructs such as bonding, cohesion, social control, and peer influence are all central correlates of crime in criminological research (Papachristos 2010). These constructs in turn are commonly thought to represent emergent properties of the structure of affiliations among social groupings at various levels of aggregation (e.g., peer groups, neighborhoods, cities).

A number of macro- and multilevel theories also specifically focus on the direct and indirect effects of neighborhood and community networks on aggregate and individual rates of offending. Unfortunately, despite the centrality of social networks in criminological theory and empirical research, criminologists rarely employ social network analysis in research, especially when compared to researchers from other disciplines (e.g., public health and sociology; see Papachristos 2010). The use of social network analysis in community and neighborhood studies of crime occurs even less frequently. This neglect is quite noteworthy considering social processes related to neighborhood-based affiliations are central in explanations of neighborhood "effects" on crime (Sampson et al. 2002).

The disjuncture between the conceptual significance of social networks and the relatively infrequent use of network data and analysis in research on crime highlights the need for more effective integration of social network analysis into the criminological toolkit. Initial efforts toward this end have resulted in significant advances in the articulation and testing of classical and contemporary theories of crime. For instance, researchers focusing on "peer effects" (Haynie 2001) and gangs (Papachristos 2009; Radil et al. 2010) are increasingly employing network designs to test and expand upon criminological theories of diffusion, group conflict, and influence. Research on the implications of neighborhood social networks has also integrated more conceptually sophisticated social network insights into their explanations of criminal behavior. Despite the inherent difficulty of collecting data on neighborhood social networks, neighborhood scholars are developing innovative methods of data collection and analysis that may be used to study the association between neighborhood network structures and crime.

This entry presents an overview of recent advancements in the study of neighborhood social networks and crime focusing on neighborhood networks as they pertain to the dominant macro- and multilevel theories of crime and community. Those primarily interested in the association between individual level or ego networks and crime may wish to read Haynie and Soller's (this volume) chapter on "peer effects" research in criminology.

This entry proceeds as follows. First, a brief overview of classical and contemporary theories of neighborhoods and crime is presented, highlighting the theoretical associations between the structure of social relations and crime. Attending to these theoretical associations will help identify which properties of social networks are most consequential to crime. This entry then highlights the unique properties of network data while focusing on the potential application of social network analysis to the study of neighborhoods and crime. Finally, this entry describes recent studies that have employed innovative social network approaches to understanding the association between neighborhoods and crime.

Neighborhood Networks in Theory: From Yesterday to Today

Durkheim (1951) provides one of the first attempts to empirically examine the association between the structure of social relations and aggregate social problems. Durkheim proposed high levels of communal social integration promote social control over individual members, constraining individuals' natural tendency to obviate social rules in the pursuit of self-interested ends. Thus, strong affiliations among members control crime and maintain social order through social control processes.

Early Chicago School scholars such as Wirth (1938) suggested increasing urban development and division of labor has deleterious effects on the community structure and solidarity characteristic of nonindustrial societies. The linear model of development has been characterized as a prototypical example of the "community lost" paradigm, which suggests pervasive division of labor in society disrupts solidary communities, leading to the formation of fragmented networks consisting of "impersonal, transitory, and segmental ties" (Wirth 1938, p. 12). Such patterns of community organization are thought to hinder neighborhood residents' capacity to collectively organize to deal with social problems (e.g., crime and delinquency) within local communities.

Around the same time, Shaw and McKay (1942) proposed their theoretical model of community social organizational influences on crime. In this view, neighborhood structural characteristics, such as concentrated poverty, racial and ethnic heterogeneity, and residential stability, determine the extent of social interaction among neighbors and their corresponding capacity to maintain informal social control over local youth, with consequences for the prevalence of crime in the community. Building on the work of Shaw and McKay, Kornhauser (1978) developed a community control model of neighborhood social disorganization, which more precisely articulates how concentrated disadvantage, population turnover, and racial/ethnic heterogeneity shape consequential neighborhood social networks. For instance, residential instability, or the frequency with which residents move in and out in a neighborhood, is thought to limit opportunities to form trusting relations among neighbors. Racial/ethnic heterogeneity is hypothesized to limit interaction among members of different racial groups, which hinders the formation of "weak ties" (Granovetter 1973) within neighborhood settings. One potential outcome of social and cultural heterogeneity is a fragmented neighborhood network structure that hinders the collective capacity to maintain informal social control. This systemic model of crime thus proposes exogenous neighborhood structural characteristics indirectly lead to crime through shaping neighbor networks, which determine more proximate causes of crime, such as informal social control.

While Kornhauser extracted a pure control model from Shaw and McKay's theory of social disorganization, other theorists (Cloward and Ohlin 1960) propose that the criminological consequences of community social organization are dependent upon the types of individuals and cultural models that are present in the community. From this perspective, deviant "value systems" represent emergent properties of neighborhood social organization. Accordingly, structures of social ties among neighborhood residents take on additional criminological consequences by fostering the formation and maintenance of delinquent subcultures conducive to crime.

From an individual standpoint, neighborhood contexts are important in shaping peer associations, especially among youth and adolescents. Peer delinquency in turn is one of the most robust predictors of offending (Haynie 2001). Different features of neighborhood social organization also more directly influence individual offending. For instance, routine activity perspectives suggest adolescent offending is in large part a function of the time adolescents spend engaging in unstructured socializing with peers in the absence of adult authority figures (Osgood et al. 1996). Other prominent theories, such as collective efficacy theory (Sampson et al. 1997), also point to the importance of intergenerational closure (i.e., adult embeddedness in neighborhood social networks) in promoting pro-social behavioral norms, which in turn fosters adolescent development and desistence from crime.

Finally, studies on the criminological consequences of residential segregation point to the importance of intra-neighborhood social ties in determining neighborhood- and city-level crime rates. As Krivo et al. (2009, p. 1792) suggest, "city segregation contributes to neighborhood violent crime indirectly, through the way it produces isolation and structural disadvantage in predominantly minority areas, and directly, by making it difficult for separate and unequal groups to work together to foster common goals and solve shared problems." From a network perspective, racial segregation hinders social integration across neighborhoods, which in turn obstructs the formation of ties between political institutions and community investors and members of disadvantaged and isolated communities. Low levels of city-level social integration limit intergroup contact, which has deleterious effects on efforts at collective organizing to realize both the formal and informal controls over crime (Hunter 1985).

Empirical Tests of Neighborhood Theories of Crime

The aforementioned theoretical approaches present different, and at times, competing mechanisms with regard to the role neighborhood social networks play in shaping individual and aggregate rates of offending. While empirical support mounts for a number of these criminological theories, to date, few studies have tested these theoretical models with social network data.

The vast majority of studies focusing on neighborhood networks and crime instead utilize “perceptual measures” to capture aspects of neighborhood social networks. Most typically, researchers ask neighborhood residents a series of questions regarding perceptions of neighborhood social organization (e.g., intergenerational closure, friendship ties) or the extent of neighbor interactions (e.g., see Bellair and Browning 2010). Unfortunately, such measures are unable to capture features of the *structure* of neighborhood social networks (e.g., cohesion, network closure) that are most consequential for crime in the aforementioned theories. As few studies have attempted to directly measure the structure of social ties at the neighborhood level, “the structural features of locally based social networks have been inferred from the reports of individual respondents, measured indirectly, or simply presumed to operate” (Entwisle et al. 2007, pp. 1500–1501). As demonstrated below, social network data (i.e., actual relational data linking nodes in a network) better capture the social phenomena that are at the heart of many theories of community and crime.

The dearth of research that employs social network designs in the study of neighborhoods and crime is in no small part due to the significant (but not insurmountable) challenges associated with collecting data on neighborhood networks. In the following section, a brief overview of social network analysis is provided, highlighting the unique features of social network data. At the same time, the inherent difficulties associated with collecting network data across several neighborhoods that is necessary for the statistical analysis of “neighborhood effects” are discussed. However, this section also presents examples of certain forms of social network data (e.g., affiliation or two-mode network data) and network analytical techniques (e.g., network-based simulations) that may be used to enrich the understanding of the association between neighborhood networks and crime.

Some Unique Properties of Social Network Data

From a theoretical standpoint, social network analysts stress the importance of interdependence among actors (e.g., individuals, institutions, nations) to explain both individual and collective outcomes. From this standpoint, individual and collective behaviors represent emergent properties of identifiable patterns of ties between actors and institutions and other units of interaction (Papachristos 2010). At the same time, the processes that explain variation in the structural characteristics of social networks (e.g., density, clustering) are also of interest to sociologists (Entwisle et al. 2007) and criminologists (Young 2011). In this latter case, social networks themselves serve as dependent variables. As the effects of neighborhood networks on crime itself are of primary interest in this entry, networks as dependent variables are not discussed in this entry.

While an extensive review of the various types of social network data is beyond the scope of this entry, this section provides a brief overview of the constitutive elements of network data that are most relevant to research on neighborhood networks.

All forms of social network data consist of *actors*, which may be conceptualized as individuals, organizations, or any other unit of analysis that are thought be related to one another, and *ties*, which refer to specific types of relationships that link actors within the population of interest. Just as actors may be defined in any number of ways, there are several different types of ties that link actors. One of the most common types of tie that is studied among criminologists is friendship. Friendship networks are relevant to a number of criminological questions, such as, “Is adolescent delinquency related to the delinquency of one’s friends?” and “Do dense neighborhood-based friendship networks protect against crime?”

Ties need not be limited to those capturing friendship relations. Indeed, a number of different types of ties have also been considered in criminological research. For example, Papachristos (2009) used official police records to reconstruct a network consisting of Chicago gangs. In this network, two gangs were linked through a directed tie if a member of one gang killed a member of another gang. Schaefer (2012) used official police data to construct a neighborhood-based co-offending network in Maricopa County, AZ. In this network, two census tracts were linked if juvenile delinquents from two different tracts were arrested during the commission of the same crime. An emerging approach to studying neighborhood effects on delinquency and adolescent development focuses on overlap in "routine activity spaces" among neighbors to measure network structural characteristics of neighborhoods. Integrating insights from activity space approaches (which is given more attention near the end of this entry) and social network analysis, an emerging approach to measuring neighbor networks focuses on residents' overlap in routine activities. Demonstrating the utility of this approach, Browning and colleagues (2011) recently used data from Los Angeles Family and Neighborhood study (LAFANS) to construct 65 "actor-location" neighborhood networks consisting of individuals tied to block groups in which residents' routine activities took place. Findings indicate various structural features of actor-setting networks (e.g., density, clustering) vary across the census tracts in the study. The exact types of tie that can link actors in a network that are considered are determined by particular research questions.

Ties may be further distinguished by their directionality. In network terminology, directed ties are referred to as *arcs*, while ties that are undirected are referred to as *edges*. Such a distinction is important, as some relationships are inherently directional, as in the case of resource exchange between actors. Other ties may not entail directionality. One prime example of non-directed ties is ties between spouses.

Social network data are most often collected through the use of two primary research designs. *Ego network data* most typically focus on the direct ties and associates of an individual "focal" node. Ego network data are typically collected through random samples by asking subjects to identify all, or a subset, of the actors to whom they have some form of connection (e.g., friendship, economic exchanges). Respondents then report any existing ties among the identified "alters" or actors to whom the focal respondent is tied. However, ego network data cannot typically be used to measure aggregate features of social networks, which are of primary interest to neighborhood theories of crime. However, through the use of snowball sampling techniques (Sampson and Graif 2009) and respondent-driven sampling (Papachristos et al. 2012), researchers have constructed relatively complete network data from ego network data.

Whereas ego network designs focus on the structure of ties that surround single actors, *full* or *complete network* designs collect information regarding every actor's ties within a population or multiple populations of actors. For instance, the groundbreaking National Longitudinal Study of Adolescent Health or Add Health attempted to collect information on every adolescents' school-based friendship ties within several middle and high schools in the United States. Complete network data allow for the examination of a number of topics of criminological interest. For instance, as with ego network data, complete network data allow for the examination of the association between peer delinquency and individual offending. However, complete network data also allow researchers to examine how features of the larger network (e.g., cohesion) and individuals' positions within larger networks (e.g., centrality) influence offending. For instance, Haynie (2001) found that peer delinquency is more strongly associated with individual offending among adolescents who are tied to others who are also well connected in school network.

Network data are also comprised of radically different data structures. For example, bipartite networks (also known as affiliation or two-mode networks) consist of two different sets of actors that can only be linked to members of the opposite set. Browning and colleagues' (2011) actor-location network represents a bipartite network, given neighborhood residents (the first node set) are indirectly linked through sharing common activity spaces (the second node set). The data structure of bipartite networks differs from that of one-mode networks, which consist of single sets of actors and the direct ties between them. While one-mode networks are more commonplace in network research, as this entry will illustrate, certain features of bipartite network data open up new and exciting avenues for future research on neighborhood networks.

Network data capture simplified versions of the social relations between actors. However, in reality, actors are tied to one another in any number of different ways. For example, students within a school network may be tied to one another through friendships, romantic relationships, overlap in their courses, participation in extracurricular activities, neighborhood residence, or bullying relations. *Multiplex networks*, consisting of actors who are connected by two or more different types of relations, explicitly take into account the different ways in which actors are connected to one another. Importantly, multiplex networks allow for the simultaneous examination of how different types of ties overlap and create specific relational structures within networks. While most techniques of network analysis are designed for networks consisting of one type of tie, network analysts are increasingly focusing on designing methods for multiplex networks.

In sum, social network data provide unique opportunities to examine the influence of social relations on individual and collective outcomes. This property of network data is particularly attractive to researchers focused on the association between neighborhoods and crime, as the structure of social relationships among neighbors is fundamental to all of the prominent theories of neighborhoods crime.

Unfortunately, there are a number of methodological barriers to gathering network data among neighbors. For example, constructing a one-mode network consisting of friendships among neighbors within just one neighborhood would require every resident to identify each of their friends from a neighborhood roster. The likely nontrivial amount of survey nonresponse in such a study would impede reliable estimation of structural properties of the neighborhood network.

Another problem with measuring neighborhood networks with survey data stems from the "boundary specification" problem that afflicts both social network research and conventional (i.e., nonnetworked) research on neighborhoods and crime. With regard to the latter, a number of different levels of community aggregation have been considered in research on neighborhoods and crime. For example, studies have (in order of their typical size in area) examined crime and disorder at the census block level, the block group level (which in the United States typically consists of 39 blocks), the census tract level, neighborhood cluster level (i.e., two or three contiguous census tracts), and zip codes. There are two significant limitations related to all of these forms of aggregation. First, certain neighborhood boundaries are based on administrative definitions, as in the cases of census tracts and zip codes. Despite the plethora of research that identifies associations between the characteristics of neighborhoods, as defined by administrated boundaries, and individual and aggregate outcomes (see Sampson et al. 2002 for a review), there is no guarantee that administrative boundaries encompass the actual "neighborhoods" that neighborhood researchers aim to study. Second, the association between neighborhood characteristics and crime has been shown to depend on the level of aggregation. For example, Hipp (2007) demonstrates that economic resources are most strongly associated with neighborhood crime and disorder among neighborhoods that are defined at the block level, while racial heterogeneity is most strongly associated with crime and disorder when measured at the census tract level.

A similar problem of boundary specification exists in social network research. In social network analysis, the issue of boundary specification pertains in large part to both the inclusion of certain actors and types of affiliations. With regard to the latter, a researcher must decide upon which types of ties are to be measured. Decisions regarding the measurement of ties are most often driven by research questions prior to execution of the study. Conversely, identifying which actors comprise a network in question takes considerably more thought in certain situations. As Laumann et al. (1983, p.18) explain:

> because individual behavior is viewed as at least partially contingent on the nature of an actor's social relationships to certain key others. . .care must be given to specifying rules of inclusion. Such rules pertain both to the selection of actors or notes for the network *and* to the choice of types of relationships among those actors to be studied.

Accordingly, much thought must be put into identifying both the actors as types of ties that are to be included in the empirical network, as noninclusion of relevant actors may have severe consequences for the measurement of the structural properties of networks.

While researchers employing network methods and concepts continue to make novel contributions to the understanding of crime, difficulties associated with collecting network data among neighborhood samples have contributed to the relative paucity of network-based studies of neighborhood effects. In the following section, four recent alternatives to studying neighborhood networks are discussed. Each takes a distinct approach to the conceptualization and measurement of neighborhood networks. It is our contention that these approaches hold promise for advancing the understanding of how neighborhood-based social network processes shape individual and aggregate offending patterns.

New Directions

Despite the challenges facing investigation of network effects on crime, emerging directions in the conceptualization and measurement of neighborhood networks offer the potential to significantly advance the understanding neighborhood effects on crime. The remainder of this entry focuses on recent studies that have attempted to capture the structure of social relations among residents and other important actors (e.g., community organizers) in neighborhood networks. While to date, no known published studies stemming from these projects have applied social network data to examine the association between neighborhood or community networks and crime itself, studies have focused on processes that are thought to mediate the association between neighborhood networks and crime, such as trust, cohesion, and informal social control. The described approaches attempt to capture consequential neighborhood networks through innovative efforts to reconstruct the pattern of actual ties among actors (Sampson and Graif 2009) or by integrating spatial and social network information to arrive at alternative conceptualizations and measures of neighborhood networks (Butts et al. 2012; Radil et al. 2010; Browning et al. 2011).

A large share of recent research on neighborhood effects on crime has utilized data from the Project on Human Development in Chicago Neighborhoods (PHDCN). A lesser-known component of the PHDCN, the key informant (KI) study, utilized a snowball sampling design to collect information on dimensions of relationships among community leaders across religious, educational, business, law enforcement, political, and community organization domains within several of Chicago's community areas. As part of initial interviews, key informants were asked to identify up to five people they went to in order to "get things done" in the community. From these nominations, Sampson and Graif (2009) constructed community leadership networks for a subsample of the community areas, measured a number of their structural properties (e.g., density, leadership inequality, and path distance). The authors also found that network centralization in the leadership network captures the extent to which nominations from all community leaders are sent to a few key leaders to be positively associated with community residents' trust in law enforcement and trusting relations among community leaders. The authors suggest that the presence of highly central leaders may foster working trust among leaders and residents by making leaders more visible and accountable to the community. Thus, leaders and community residents may be better able to exercise social control over central leaders, creating a multilayered "enforceable trust" (Sampson and Graif 2009, p. 207) within the neighborhood network.

A number of recent studies have integrated spatial and social network approaches to arrive at theoretically relevant conceptualizations and measures of neighborhood networks. For instance, Radil et al. (2010) build on insights from geography and network perspectives to highlight the roles that the geographic distribution of gangs and the network structure of gang rivalries play in shaping patterns of gang violence within a larger community. The authors suggest that space is socially constructed in large part through processes related to geographic and network embeddedness of actors. One important feature of network embeddedness is spatialized network structural equivalence, which captures the extent to which two or more geographic areas are similarly situated in terms of positions within "spatialized" social networks.

Relying on information from gang informants and the Los Angeles Police Department, the authors reconstructed a spatialized community network for which actors represent collections of census block groups that comprise gang territories. Gang territories are in turn tied to one another through rivalries with other local gangs. Through a network analytic technique called *convergence of iterated correlations* (CONCOR), the authors identify geographic subsets that are comprised of gang territories that are similarly situated within the rivalry network. Upon mapping the territorial subsets, the authors demonstrate that the geographic distribution of the gang territories has a distinct spatial pattern and that the spatialized patterning is meaningfully related to the distribution of gang violence across the larger community. Results from Radil et al. (2010) study illustrate the importance of the structure of social relations among actors across different geographic spaces in shaping crime rates across different communities.

Butts and colleagues (2012) present an alternative, simulation-based approach to measuring community networks that relies on assumptions about the spatial organization of social ties. The authors ground their analysis on the well-established notion that the probability of a social tie between individuals increases with their geographic proximity to one another (McPherson et al. 2001). Using data on geographic distributions of individuals from the US Census, Butts and colleagues constructed simulated networks for several metropolitan (i.e., an area that includes at least one city with a population greater than 50,000 individuals) and micropolitan (areas that contain at least one city with a population between 10,000 and 50,000) areas in the USA. Through specifying and applying a spatial interaction function (SIF), which captures the "marginal probability of a tie between two randomly selected individuals at some given distance" (Butts et al. 2012, p. 83), to data on geographic location of actors, the authors constructed simulated metropolitan and micropolitan networks comprised of friendship relations and face-to-face interactions. Butts and colleagues also identified associations between the spatial distribution of actors across the units of aggregation and various structural properties of the networks. This simulation process can also incorporate several other nongeographic processes that are known to influence tie formation, such as selective mixing on race, age, and socioeconomic status (McPherson et al. 2001) to potentially simulate networks that more accurately reflect actual networks of distinct areas. While this innovative method holds much promise for understanding neighborhood network processes and crime, future research that compares the simulated networks with actual networks constructed from network-based studies will help to assess how well-simulated networks reflect neighborhood social networks.

One of the most attractive qualities of the network simulation approach of Butts and colleagues is that it enables researchers to construct social networks with readily available data across several cities and neighborhoods. This is particularly useful, as collecting data on every individual's ties across a single, let alone, several neighborhoods through conventional survey approaches is a daunting, if not impossible, task.

However, individuals may be connected to other persons through any number of ways. For instance, neighborhood residents may be indirectly tied to one another through shared participation in activity spaces. Activity spaces may be understood to encompass all of the locations individuals come into contact with as a result of their routine activities (Golledge and Stimson 1997). Incorporating activity space data into a social network analytic framework may help address the challenges facing research on neighborhood networks and crime in two primary ways. First, activity space data enable the measurement of individual exposure to specific places and settings that vary in their criminogenic properties (i.e., extent of disorder and crime). Activity space data can also capture exposure to certain types of individuals and allow for the measurement of the extent of overlap in individuals' routine activities. Finally, as illustrated below, network structural characteristics describing patterns of overlapping routine activities among neighborhood residents are likely important in the functioning of neighborhood social processes that are relevant to crime (e.g., collective efficacy, individuals' access to social capital, participation in neighborhood organizations).

Apart from capturing individuals' place-based exposures, sampling activity spaces among geographically contained populations may allow investigators to measure the extent to which neighborhood residents share activity locations within given boundaries or larger areas (e.g., entire metropolitan areas). Consistent with Jacobs (1961), the structural patterns of activity overlap may capture important features of the social organization of a neighborhood. For instance, "co-location networks" may impact a number of neighborhood processes such as social network formation as well as trust and shared expectations for informal social control. While these factors have been linked to individual offending and neighborhood crime rates, the social and ecological dynamics that are associated with such features of neighborhood organization remain elusive. Moreover, by measuring the co-location network more precisely, researchers may apply a continuously developing catalogue of network analytic techniques to characterize routine activity patterns using sophisticated global network measures (e.g., density) and characterize individuals' locations in the network (e.g., centrality). Additionally, through the use of panel and other longitudinal study designs, one may examine how co-location networks evolve over time with stochastic actor models (Ripley et al. 2012).

Browning and colleagues (2011) recently examined the association between structural features of actor-location networks among neighborhoods in Los Angeles and a number central processes in neighborhood-based studies of crime and delinquency. The authors find that triadic closure, or "clustering" (Opsahl Forthcoming), in the actor-setting network is positively associated with access to social capital as well as perceptions of network exchange (i.e., frequency of exchange of favors among neighbors), intergenerational closure among neighborhood residents, and collective efficacy. Future research that builds upon the approach of Browning and colleagues may help further identify the association between structural features of neighborhood actor-setting networks and offending patterns.

Conclusion

The primary objective of this entry was to provide an overview of network approaches to studying the association between neighborhood social organization and crime. Importantly, the approaches highlighted in this entry were situated within prominent theoretical approaches to neighborhood networks and crime. After outlining central network processes of neighborhood-based theories of crime, this entry highlighted the limitations of "perceptual" measures of community organization that are most commonly used to assess structural features of neighborhood networks. The limitations of perpetual measures were further illustrated by highlighting how relational properties of social network data allow for a more complete understanding of neighborhood social organization as it relates to individual and aggregate offending patterns.

Most importantly, this entry detailed emerging approaches to measuring structural properties of neighborhood social networks. Through the use of network simulations, snowball sampling, spatialized gang networks, and two-mode actor-location affiliation network data, recent studies have made great strides in measuring structural properties of neighborhood and community networks. Perhaps as importantly, these techniques circumvent a number of the inherent difficulties and cost restrictions of collecting neighborhood-based social network data. Future research that builds on these and similar network analytic techniques may advance the understanding of the association between neighborhood social organization and offending rates by identifying which structural properties of neighborhood networks (e.g., network density, closure) are most strongly associated with crime. In addition, considering different types of ties (e.g., friendship tie versus gang rivalry) may further demonstrate the criminological consequences of differential social organization and advance the understanding of the neighborhood context of crime.

Related Entries

- ▶ Co-offending
- ▶ Control Theory
- ▶ Gangs and Social Networks
- ▶ Neighborhood Effects and Social Networks
- ▶ Network Analysis in Criminology
- ▶ Social Control
- ▶ Social Control and Self-Control through the Life Course
- ▶ Social Network Analysis of Organized Criminal Groups
- ▶ Social Network Analysis of Urban Street Gangs
- ▶ Spatial Models and Network Analysis

Recommended Reading and References

Bellair PE, Browning CR (2010) Contemporary disorganization research: an assessment and further test of the systemic model of neighborhood crime. J Res Crime Del 47:496–521

Browning CR, Jackson AL, Soller B, Krivo L, Peterson RD (2011) Ecological community and neighborhood social organization. Paper presented at the annual meeting of the American Society of Criminology, Washington, DC, 16–19 Nov 2011

Butts CT, Acton RM, Hipp JR, Nagle NN (2012) Geographical variability and network structure. Soc Networks 34:82–100

Cloward RA, Ohlin LE (1960) Delinquency and opportunity: a theory of delinquent gangs. Free Press, Glencoe

Durkheim E (1951) Suicide: a study in sociology. Free Press, New York

Entwisle B, Faust K, Rindfuss RR, Kaneda T (2007) Networks and contexts: variation in the structure of social ties. Am J Sociol 112:1495–1533

Golledge RG, Stimson RJ (1997) Spatial behavior: a geographic perspective. The Guilford Press, New York

Granovetter M (1973) The strength of weak ties. Am J Sociol 78:1360–1380

Haynie DL (2001) Delinquent peers revisited: does network structure matter? Am J Sociol 106:1013–1057

Hipp JR (2007) Block, tract, and levels of aggregation: neighborhood structure and crime and disorder as a case in point. Am Sociol Rev 72:659–680

Hunter A (1985) Private, parochial and public social orders: the problem of crime and incivility in urban communities. In: Zald M, Suttles (eds) The challenge of social control: citizenship and institution building in modern society. Ablex Publishing, Norwood, pp 230–242

Jacobs J (1961) The death and life of great American cities. Random House, New York

Kornhauser R (1978) Social sources of delinquency. University of Chicago Press, Chicago

Kossinets G (2006) Effects of missing data in social networks. Soc Networks 28:247–268

Krivo LJ, Peterson RD, Kuhl DC (2009) Segregation, racial structure, and neighborhood violent crime. Am J Sociol 114:1765–1802

Laumann EO, Marsden PV, Prensky D (1983) The boundary specification problem in network analysis. In: Burt RS, Minor MJ (eds) Applied network analysis: a methodological introduction. Sage Publications, London, pp 18–34

McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. Annu Rev Sociol 27:415–444

Opsahl T (Forthcoming) Triadic closure in two-mode networks: redefining the global and local clustering coefficients. Soc Networks

Osgood DW, Wilson JK, O'Malley PM, Bachman JG, Johnston LD (1996) Routine activities and individual deviant behavior. Am Sociol Rev 61:635–655

Papachristos AV (2009) Murder by structure: dominance relations and the social structure of gang homicide. Am J Sociol 115:74–128

Papachristos AV (2010) The coming of a networked criminology. In: MacDonald J (ed) Measuring crime and criminality: advances in criminological theory, vol 17. Transaction Publishers, New Brunswick, pp 101–140

Papachristos AV, Braga AA, Hureau DM (2012) Social networks and the risk of gunshot injury. J Urban Health 89:992–1003

Radil SM, Flint C, Tita GE (2010) Spatializing social networks: using social network analysis to investigate geographies of gang rivalry, territoriality, and violence in Los Angeles. Ann Assoc Am Geogr 100:307–326

Ripley, Ruth M, Snijders, Tom AB, Preciado P (2012) Manual for SIENA version 4.0. University of Oxford, Department of Statistics; Nuffield College, Oxford, http://www.stats.ox.ac.uk/~snijders/siena/

Sampson RJ, Graif C (2009) Neighborhood networks and processes of trust. In: Levi M, Hardin R, Cook KS (eds) Whom can we trust? How groups, networks, and institutions make trust possible. Russell Sage Foundation, New York, pp 182–215

Sampson RJ, Raudenbush SW, Earls F (1997) Neighborhoods and violent crime: a multilevel study of collective efficacy. Science 2778:918–924

Sampson RJ, Morenoff JD, Gannon-Rowley T (2002) Assessing 'neighborhood effects': social processes and new directions in research. Annu Rev Sociol 28:443–478

Schaefer DR (2012) Youth co-offending networks: an investigation of social and spatial effects. Soc Networks 34:141–149

Shaw C, McKay HD (1942) Juvenile delinquency and urban areas. University of Chicago Press, Chicago

Wirth L (1938) Urbanism as a way of life. Am J Sociol 44:1–24

Young J (2011) How do they 'end up together'? A social network analysis of self-control, homophily, and adolescent relationships. J Quant Criminol 27: 251–273

Neighborhood Policing

▶ Community Policing

Neighborhood Policing (UK)

▶ Reassurance Policing and Signal Crimes

Neighborhood Social Networks

▶ Neighborhood Effects and Social Networks

Neighborhoods

▶ Social Network Analysis and the Measurement of Neighborhoods

Neighboring

▶ Social Network Analysis and the Measurement of Neighborhoods

Network Analysis in Criminology

Dana L. Haynie[1] and Brian Soller[2]
[1]Department of Sociology, The Ohio State University, Columbus, OH, USA
[2]Department of Sociology, University of New Mexico, Albuquerque, NM, USA

Synonyms

Adolescent contexts; Delinquency; Peer effects; Social networks

Overview

Criminological theories have focused on different aspects of social relationships to explain the role peers play in shaping delinquency. Control theory, as presented by Hirschi (1969), focuses on the strength of attachment between individuals to explain variation in delinquent outcomes. Conversely, influence theories, such as differential association theory (Sutherland and Cressey 1960), propose that delinquency is learned within intimate relationships that expose individuals to definitions favorable to law and norm violation. Unfortunately, most empirical tests of these competing perspectives rely on data from random samples that do not allow for the direct measurement of peer characteristics (e.g., peer delinquency) or the assessment of the structural characteristics (e.g., cohesion) of peer and friendship groups. Overcoming these limitations, a network approach to crime/delinquency takes into account the structure of social relations, as well as the characteristics of actors comprising the network, to explain peer influence processes. This entry outlines contemporary approaches to understanding the peer-delinquency association, focusing primarily on the ways that researchers operationalize key elements of influence and control theories, such as exposure to delinquent peers and peer group cohesion, with social network data. The entry also describes how criminologists are advancing the understanding of peer influence by studying behavioral development and network dynamics over time and by expanding the study of peer influence to include romantic partners, those to whom actors are indirectly tied, as well as associations occurring outside of school contexts.

Friend and peer influence has long been central to explanations of crime, delinquency, and other problem behaviors. Compared to children and adults, adolescents attribute greater importance to friends, spend more time socializing with friends, and are more strongly influenced by the behaviors and attitudes of their friends. While peer relationships have been less central in explaining adult offending, research documents the role of social influence from friends and romantic partners throughout the life course (Giordano et al. 2003). Not surprisingly then, the finding that individuals with delinquent friends are likely to be delinquent/criminal themselves is one of the most consistent and strongest findings in the criminological literature. Consistently robust associations between peer and individual delinquency have led some to argue that peer influence is one of the most important processes in explaining delinquent outcomes, regardless of whether the focus is on substance use, minor property offenses, status offenses, or violent crimes.

Prominent explanations of crime and delinquency disagree with regard to the mechanisms through which peers influence individual offending. While some of the debate is informed by theoretical concerns (e.g., "selection" versus "influence" as the driving mechanism), debate also stems from inadequate and inconsistent measurement of peer characteristics across research designs. In order to advance the understanding of peers' impact on delinquent outcomes, this entry provides a broad overview of past, present, and emerging techniques for measuring peer characteristics in criminological research. Particular attention is directed at social network processes and methods.

This entry begins with a discussion of two prominent theoretical perspectives on delinquency, namely, control and influence theories. Close attention is paid to the manner in which peers are purported to influence crime and delinquency in these theoretical models. Next, social network approaches to peer influence on delinquency are discussed. The Section "Measuring Peer Influence: Past and Present" describes past and current methods of measuring peer characteristics in criminological research. This section first focuses on *perceptual* or self-report measures of peer characteristics, highlighting the methodological shortcomings of this approach. Next, network approaches to measuring the *content* (e.g., peer delinquency) and *form* (i.e., structural characteristics) of adolescent peer networks are identified. In addition, this section discusses manners in which network analysis has been used to help resolve competing claims regarding the role of peers in delinquent behavior in criminological theories. In section "New Directions in Peer Influence", more recent innovations in social network approaches to peer influence are reviewed, highlighting how these methodological advancements can promote the understanding of peer influence processes. Finally, a conclusion provides a discussion of future directions in the study of peer influence that have great potential for advancing the current understanding of network processes as they relate to crime and delinquency.

Theoretical Explanations of "Peer Effects"

Control theories, such as those presented by Hirschi (1969) and Gottfredson and Hirschi (1990), and influence theories, such as those presented by Sutherland (Sutherland and Cressey 1960) and Akers (2009), are prevailing perspectives that explain the importance of peers for adolescent delinquency. These perspectives offer opposing explanations with regard to the ways in which peers influence individual delinquency. Accordingly, criminologists have focused on different features of peer contexts in order to assess the plausibility of the theoretical mechanisms in empirical research. This section briefly outlines the predominant control and influence theories of crime, and makes note of the relative importance of the *form* and *content* of peer networks in each perspective and then discusses the network approach to peer influence, and identifies how it helps advance the understanding of peer selection and influence processes.

Control Explanations of Crime

Hirschi's (1969) social control theory proposes that connections to others serve as the primary constraints against delinquent impulses and actions. With regard to peer networks, social control theory predicts that strong friendship bonds are invariably and negatively associated with adolescent delinquency. This is because strong friendships entail attachment to conventional peers that would be weak or absent among more selfish and delinquent individuals. Hirschi explains the peer-delinquency association by suggesting that delinquent youth cannot form genuine friendships with peers. This implies that social ties among delinquent youth lack high levels of attachment.

Gottfredson and Hirschi (1990) general theory of crime elaborates the process through which individuals self-select into delinquent peer groups. Gottfredson and Hirschi argue that peers have no direct influence on individual offending. Rather, individual self-control (i.e., the ability to control impulsive behavior), which is relatively stable by early adolescence, shapes how adolescents cluster together in peer settings. Delinquent adolescents with low self-control are likely to befriend other delinquents as a result of their similar levels of self-control. Apart from determining the types of friends one makes, low self-control is also a primary cause of delinquent behavior; thus delinquency and associations with delinquent others are both directly caused by low self-control.

Importantly, Hirschi and Gottfredson and Hirschi's control models propose that the association between delinquent peers and individual offending is spurious. Delinquency is instead explained by other processes such as inadequate social control caused by weak attachment to others or low self-control. These models explain any peer-delinquency association resulting from selection processes, and not because friends influence individual delinquency.

With regard to measuring peer effects on individual offending, social control theory primarily focuses on attachment within adolescent networks. Both control models also predict that delinquent peer groups will be less cohesive than those consisting of nondelinquent individuals with high self-control. Accordingly, measures that capture the strength of friendship ties, as well as the configuration of ties linking individuals within friendship groups, are crucial elements in tests of control theories. Such measures include subjective assessments of peer attachment, which capture mutual trust and involvement among friends, and network density, which measures the probability that a tie will exist between two members of friendship groups. Conversely, measures of peer characteristics, such as delinquency, are less central because the peer-delinquency association is believed to be spurious in these models. In support of both variants of control theory, weak attachments and low self-control are consistently associated with delinquency in research (Pratt and Cullen 2000).

However, research has also found that delinquent adolescents are no less attached to friends than their nondelinquent counterparts (Giordano et al. 1986). Additionally, Young et al. (2011) found self-control has a negligible impact on adolescent friendship formation after triad closure, or the tendency for individuals who share friends to become friends themselves, and other individual characteristics are taken into account. Similarly, Kreager et al. (2011), found group-level delinquency has no impact on the structural characteristics of friendship groups when group composition (i.e., average socioeconomic status and proportion male) is taken into account. Perhaps most importantly, friendships with delinquent peers is positively, and consistently, associated with delinquent behavior in criminological research (Haynie 2001), even after controlling for self-control and peer attachment (Pratt and Cullen 2000). Although friendship selection and social and self-control most likely shape delinquent behavior, peer influence continues to play an important part in explaining delinquent outcomes.

In contrast to control theories, influence theories argue that peer characteristics are consequential for individual offending. Accordingly, empirical tests of these theories have in turn largely focused on the content of social networks to explain individual offending.

Influence Perspectives

Influence theories argue that criminal and other “antisocial” behaviors are learned from intimate social relationships with others where attitudes or “definitions” favorable to law violation are acquired. The social transmission of crime/delinquency primarily occurs within peer networks through the transference of favorable attitudes and definitions that encourage criminal behavior. Akers’ (2009) extension to differential reinforcement theory emphasizes behavioral modeling and operant conditioning processes. This model assumes that the adoption of criminal behavior occurs through the imitation of friends’ behavior and the observation of its consequences, either positive or negative. Sutherland and Akers’ theories represent prototypical influence theories in that they argue that crime and delinquency, like any behavior, are learned in intimate relationships, including peer friendships.

The characteristics of peers within adolescent friendship groups have been at the forefront of empirical tests of differential association and social learning theories. For example, the principle of differential association states, “A person becomes delinquent because of an excess of definitions favorable to violation of law over definitions unfavorable to violation of law” (Sutherland and Cressey 1960, p. 78). In empirical research, differential association has been most commonly operationalized by asking respondents how many or what proportion of their friends commit deviant acts. These “perceptual” measures of peer delinquency have been remarkably strong and consistent predictors of individual offending in delinquency research (Pratt et al. 2010). Both the consistency and robustness of the association between peer and individual offending attest to the validity of peer influence theories.

While perceptual measures of peers’ criminal involvement are some of the most robust predictors of individual offending, tests of social learning and differential association theories have been hindered by oft-used self-report measures of delinquency in criminological research. Perceptual measures of peer delinquency are susceptible to report bias, which likely results from delinquent individuals ascribing their own behaviors to their friends (Gottfredson and Hirschi 1990). Indeed, recent research indicates that perceptual measures of peer delinquency both *overestimate* (Haynie and Osgood 2005) and *underestimate* (Weerman and Smeenk 2005) the actual level of delinquency among peers.

One drawback of social learning and differential association perspectives is that they offer little theoretical and empirical attention towards understanding how the form, or structural characteristics, of social networks factors into peer influence processes. These oversights are significant, given that exposure to and circulation of norms and behaviors conducive to offending are central features of the theories. For example, Sutherland proposes that differential associations vary in their “frequency, duration, priority, and intensity” (Sutherland and Cressey 1960, p. 78). Unfortunately, the exact meaning of frequency, duration, and intensity of differential associations has been underdeveloped and left largely unmeasured in empirical tests of influence theories (Haynie 2001). In contrast, the social network perspective more fully explains the mechanisms through which social relations shape the peer-delinquency relationship. Social network data, in turn, more fully capture both the form and content of peer groups than non-relational data, which enable more thorough examinations of peer influence processes.

A Social Network Perspective

The social network perspective offers a unique view of humans in its basic premise that individuals are interdependent and, more importantly, that these interdependencies have important consequences for behavior. Therefore, it emphasizes both the configuration of ties connecting individuals in a social structure and the characteristics of actors in that structure to explain delinquent outcomes. Scholars interested in applying a social network perspective to understand individual offending require data on the ties or connections among individuals within a particular setting. While these ties can take on different forms, researchers interested in adolescent behavior most often examine the form of friendship ties among adolescents. This suggests that the behavior and structure of adolescent networks offer powerful explanatory power for delinquency and other social behaviors. Network perspectives also allow researchers to test competing theoretical hypotheses of control and influence theories, while also advancing new concepts – such as centrality and density – and methods that expand our knowledge of delinquent contexts.

Friendship networks are particularly important for understanding adolescent involvement in crime/delinquency for a number of reasons. Integration into friendship groups where adolescents form bonds and spend time with peers may facilitate or inhibit delinquent behavior depending upon the norms, values, and behaviors within the network. Also, while adolescents may discount a friends' evaluation, they are less likely to discount a group evaluation. This suggests that embeddedness within friendship networks takes on additional influence as it generates expectations for behavior while reinforcing group social norms and beliefs. Adherence to group norms is especially likely during adolescence because acquiring peer acceptance is of central importance during this stage of the life course. Furthermore, friendships are crucial in establishing role identity throughout adolescence. As a result, adolescent peer networks are especially effective at directing and constraining individual members' behavior.

While a social network perspective offers a particularly useful tool for understanding peer influence, it is only recently that the full benefit of a network approach for crime/delinquency research has been revealed. The following section describes past and recent methods of measuring peer characteristics. Although conventional, non-networked, "perceptual" methods are discussed, this section primarily focuses on network approaches to assessing peer influence.

Measuring Peer Influence: Past and Present

Prior studies examining the effect of peer influence on delinquency generally ask adolescents to think about their friends and to report whether their friends have participated in a particular illegal behavior or set of illegal behaviors. Respondents usually report that "none," "some," or "all" of their friends participated in the behavior being discussed. Studies relying on such perceptual measures have found robust associations between peer delinquency and respondent offending (Pratt et al. 2010). However, the consistent association between perceived peer deviance and respondent delinquency provides only limited support for influence theories such as differential association theory for a number of reasons. First, social psychologists refer to the tendency to assume that friends behave as do you, as assumed similarity or projection. Therefore, the perceptual approach to measuring peer characteristics contains a same-source bias that inflates similarity in behavior between peers. In fact, prior research demonstrates self-report measures of peer delinquency tend to be not only inaccurate but are systematically biased (Young et al. 2011). Findings such as these support Gottfredson and Hirschi's claim that individual characteristics, such as self-control, bias perceptual measures of peer delinquency and other behaviors. Accordingly, the extent to which perceptual studies of peer delinquency can actually support influence theories is limited by the extent to which they are unbiased by individual characteristics. Furthermore, misspecification of peer delinquency is especially problematic in studies that test competing claims of influence and control theories, as biased measures of peer delinquency likely results in the upwardly biased estimates of peer effects and downwardly biased estimates of self-control effects (Meldrum et al. 2009).

Apart from mounting evidence indicating systematic misspecification of peer delinquency, perceptual measures of peer characteristics are limited in a number of other ways. For example, past research employs imprecise definitions of the friendship group in which behavioral influence is thought to occur. As a result, the number of friends considered by the adolescent when responding to questions regarding peer delinquency is often unspecified. Accordingly, the structure and composition of an adolescent's delinquent friendship network remains relatively unclear.

Perhaps as importantly, perpetual studies of peer influence fail to consider how structural properties of network – that is, how adolescents are connected to one another via friendship ties – condition peer influence processes. In doing so, past research assumes that everyone in the friendship network is affected by friends' behavior in the same manner. Non network approaches overlook individual position within the network (e.g., central versus peripheral), the cohesiveness of the network (i.e., the interconnections among network members), and the adolescent's status (e.g., popularity) within the network. These structural characteristics likely shape the degree to which adolescents are influenced by group behavior (Haynie 2001). A network perspective is guided by the assumption that both behaviors exhibited by network members and the structure of the network have important consequences for understanding subsequent behavior. With regard to delinquency, this suggests that exposure to pro- or anti-delinquent behaviors will depend upon the structure of the network, the adolescent's position within the network, and the behaviors exhibited in the network.

Measuring Peer Delinquency with Network Data

Recent data collection efforts, such as the National Longitudinal Study of Adolescent Health (Add Health), have made relational data available to criminologists interested in applying network theories and methods to the understanding of peer influence. The Add Health data are unique in that they provide an opportunity to analyze complete social networks of adolescents attending a representative sample of schools in the United States. To achieve this, researchers collected data on every student attending school on the day the questionnaire was administered. As part of this data collection effort, respondents were asked to nominate up to ten of their closest friends (five male and five female friends) from school rosters listing all schoolmates. Because every student present completed the questionnaire, it is possible to observe almost all friendship ties among students in the school. Add Health provides the largest and most comprehensive portrayal of adolescents' school-based friendship networks to date. Other smaller-scale efforts, such as the Netherlands Institute for the Study of Crime and Law Enforcement (NSCR) "School Study," the Context of Adolescent Substance Use Study, and the Promoting School-Community-University Partnerships to Enhance Resilience (PROSPER) study have utilized similar data collection techniques to directly measure peer and network structural characteristics.

With regard to the peer-delinquency association, criminologists have operationalized peer delinquency in a number of ways. Theoretical concerns should be, and for the most part, are primary in informing approaches to measuring peer delinquency. For example, the principle of differential association theory maintains that individuals become delinquent because they experience excess of definitions favorable to law violation, compared to definitions that favor observing law. In testing this claim, Haynie (2002) constructed a "relative" measure of peer delinquency which consisted of the proportion of friends that are delinquent within individuals' peer networks. Results indicated adolescents' delinquent behavior most resembled that of their peer groups when they were enmeshed in friendship networks with greatest behavioral consensus (i.e., all friends are either delinquent or nondelinquent). Another more common way of operationalizing peer delinquency is to measure the average amount of delinquency across one's peer group (Haynie 2001). This operationalization allows researchers to focus more on the consequences of total exposure, rather than relative exposure, to delinquent peers.

Recently, McGloin (2009) expanded the operationalization of peer delinquency by considering how peer deviance relative to one's own, rather than simply exposure to delinquent peers, shapes individual delinquency. McGloin points out that one aim of interactive relationships is to achieve behavioral congruence, or homeostasis with one's close associations. One hypothesis that stems from this perspective is that adolescents will seek behavioral congruence with their closest friends. In order to test this assertion, McGloin constructed a measure of *delinquency balance*, which consisted of the difference between respondents' delinquency and that of their best friends. Results indicate adolescents tend to commit less delinquency over time when their delinquency was previously higher than their best friend; conversely, respondents tend to become more delinquent when they were initially less delinquent than their best friend.

Influence theories in large part emphasize the importance of learning that takes place within intimate peer groups. Accordingly, most empirical tests and influence theories have primarily focused on the influence of peers to whom one is directly tied. However, influence theories stipulate that peer influence also takes place through vicarious learning, in which individuals observe certain behavior and others' reactions to that behavior. Individuals are more likely to perform observed behavior that is positively reinforced by others throughout social interactions. This suggests that peer influence processes likely also occur between individuals who are only indirectly tied.

Network approaches to peer influence allow for the testing of processes that extend beyond those to whom individuals are directly connected. For example, Payne and Cornwell (2007) found that the delinquency of indirect associates (i.e., friends of a friend who are not my friends) is positively associated with respondent delinquency, even after taking into account the delinquency of those to whom one is directly tied. However, the authors also found that the influence of indirect associations diminished according to the extent to which the overall delinquency of one's indirect ties diverges from that of the immediate friendship group. Similarly, Kreager and Haynie (2011), demonstrate that behavior among those to whom one is indirectly tied through a dating partner (i.e., friends of my partner who are not my friends) influence individual behavior. Interestingly, the authors find evidence that the drinking of friends of partners have stronger associations with a dater's future behavior than do his or her own friends or the romantic partner.

Social proximity to delinquent individuals also factors into peer influence processes. In their examination of peer and social network influences on adolescent substance use, Ennett et al. (2006) found that the shortest distance to the nearest substance user, as measured by the number of ties needed to link respondents to the nearest substance user within a school, is negatively associated with substance use after controlling for the substance use within friendship networks.

Recently developed methods for identifying peer "clusters" allow researchers to further examine peer influence occurring at multiple levels and between individuals who are only indirectly tied. Peer clusters consist of densely connected groups in which ties are more likely to occur between members of the same group than members of different groups. Peer clusters also represent "meso levels" of social organization, which are larger in size than micro level friendship groups, but smaller in scale than other levels of aggregation, such as schools (Magino 2009). These conglomerations of friendship ties, which have discernable boundaries, may be used to capture larger social circles or "cliques," within which peer influence likely occurs.

Clustering algorithms, such as Moody's (1999) CROWDS or Frank's (Frank et al. 2008) *Kliquefinder*, allow researchers to detect dense friendship groups in social networks. Resulting subgroups in turn may be used as distinct levels of analysis in the study of peer influence processes. For example, Frank et al. (2008) used the *Kliquefinder* algorithm to identify "local positions," represented by clusters of adolescents based on their shared academic course taking among Add Health respondents. The authors found that girls were more likely to subsequently take higher-level courses when they were previously embedded in peer group clusters that include girls who also took advanced math courses. While this finding does not directly pertain to delinquency, it nevertheless points to the likelihood that peer influence processes extend beyond individual ties. Although peer group clustering techniques have not been widely employed in delinquency research (see Kreager et al. 2011; Ennett et al. 2006 for notable exceptions), the study of peer influence processes occurring within larger peer groups represents an interesting future direction for research to pursue.

Exactly how structural characteristics of peer networks condition the peer-delinquency association has also been largely overlooked in prior delinquency research. This oversight is unfortunate, as differential association theory proposes that while criminal behavior is in large part shaped by exposure to delinquent peers, differential associations vary in their "frequency, duration, priority, and intensity" (Sutherland and Cressey 1960, p. 78). Additionally, the form, or structural characteristics of peer networks, likely shapes peer influence processes by facilitating, among other things, behavioral reinforcement, circulation of definitions of behavior and attitudes conducive to offending, and exposure to nonredundant information (Granovetter 1973), which may facilitate co-offending opportunities and contagion of delinquent behavior.

Recognizing that the form of adolescent networks likely shapes the relationship between the content of peer networks and delinquent outcomes, Haynie (2001) evaluated whether network characteristics condition the effect of peer delinquency on individual offending. Results indicated that network centrality, density, and popularity accentuated the positive association between delinquent peers and individual delinquency. Further research examining how structural features of social networks shape the association between peer characteristics and risk behavior will help advance the understanding of how the effects of differential associations on individual outcomes vary according to network characteristics.

New Directions in Peer Influence

Emerging and exciting areas related to social networks and adolescent offending have accompanied increased access to high-quality social network data. For instance, actor-based models have recently been introduced to analyze behavioral development and network dynamics over time. Developed by Snijders and colleagues (Steglich et al. 2010), the SIENA statistical package uses simulation and Markov chain models to estimate changes in friendship ties and behavior over multiple waves of network cross sections or panels. Recent applications of SIENA in criminological research are advancing the understanding of peer influence by distinguishing selection from influence effects, while also adjusting for the structural characteristics of the network.

Researchers interested in peer influence are also expanding their empirical focus to associations other than friends. Increased inter-gender contact in mid-to-late adolescence entails increased opportunities for forming romantic relationships (Warr 2002). Recognizing the potential for romantic partners to impact offending, Haynie et al. (2005) demonstrate that romantic partner's delinquency exerts a unique effect on adolescents' delinquency, over and beyond the delinquency of one's friends. Giordano et al. (2010) revisit Hirschi's "cold and brittle" relationship hypothesis with regard to delinquents and romantic involvement. Using data from the Toledo Adolescent Relationships Study, they found little difference in the quality and longevity of delinquents' and nondelinquents' romantic relationships.

While this entry has largely focused on how peers influence individual delinquency, most criminological theories recognize that delinquency often occurs in groups. Accordingly, peers also impact individual delinquency by providing co-offending opportunities, the effect of which is analytically distinct from peer influence (Warr 2002). Unfortunately co-offending remains a relatively understudied topic in criminology. Notable exceptions (McCarthy et al. 1998) provide strong evidence that co-offending is an important mechanism through which peers impact individual crime deviance.

Innovative network studies continue to advance the understanding of peer influence processes. For example, the PROSPER Peers data, emanating from Pennsylvania State University under the supervision of Osgood and colleagues (Kreager et al. 2011) follows two successive sixth-grade cohorts in 28 rural Iowa and Pennsylvania communities. Social network data were collected yearly from the 6th to the 12th grade, resulting in an impressive number of complete school networks (∼400) over a 7-year span within each school cohort. With detailed delinquency and substance use measures and SIENA analyses, studies from PROSPER and similar longitudinal network datasets promise to expand our understanding of peer selection and influence processes.

One area that remains particularly underdeveloped in peer influence research relates to the impact of nonschool friends and peers. Out-of-school peers are likely important in shaping delinquency because they are more likely to be delinquent, older, and less connected to conventional institutions than school-based peers. Understanding the content and structural properties of nonschool friendship networks, may provide important clues for the etiology of delinquency. Collecting such data requires researchers to move beyond the bounds of schools and focus on communities, which is logistically more demanding and likely to be more costly, but capable of providing new insights into the peer-delinquency association.

Papachristos et al. (2012) present an innovative approach to measuring non-school-based networks among high-risk individuals in Cape Verdean, a low-income, predominately African-American community in Boston. To construct the network of high-risk individuals, the authors first identified the population of Cape Verdean gang members who were known by the Boston Police Department (BPD). Next, the authors generated a list of gang members' immediate associates from Field Intelligence Observation (FIO) data which were collected through observations conducted by the BPD. This respondent-driven approach resulted in a network in which ties between individuals occurred when they were observed in each other's presence by the police and recorded in the FIO data. This process was repeated for the associates of the gang members. From there, the authors constructed a network of high-risk individuals that consisted of gang members, their associates, and the associates of the non-gang members. Using matched data on fatal and nonfatal gunshot injuries, the authors demonstrate that the probability of experiencing gunshot injury was related to one's distance from other gunshot victims.

The limited availability of community-based network data is largely a result of the astronomical costs of obtaining population samples needed to construct network-based measures across several communities. One innovative approach to constructing nonschool social networks entails constructing affiliation, or "two-mode" networks, which, in the context of community research, consist of ties between individuals and activities or locations. Recently, Browning (2011) constructed two-mode networks among 65 neighborhoods with data from the Los Angeles Family and Neighborhood Study (LAFANS). Browning demonstrates that triad closure within networks based on individual overlap in respondents' routine "activity spaces" is negatively associated with adolescent risk taking. One primary advantage of two-mode approaches to community networks is that affiliation network data can be collected through clustered random samples. This lowers the cost of constructing measures of global network structure and individual embeddedness within community-based networks. Frank's *Kliquefinder* algorithm may also be applied to two-mode networks to construct adolescent group clusters, within which peer influence may also take place (Frank et al. 2008).

Summary

This entry aimed to provide an overview of network approaches to the measurement of peer effects in criminological research. After outlining the merits and limitations of control and influence theories, the entry advocated a social network approach to understanding the role of peers for two primary reasons. First, the network framework described in the entry emphasized the importance of the social connections among individuals within a social setting. Capturing the relations among actors within social contexts allows criminologists to assess the interconnectedness of actors, which is an important element of both influence and control perspectives. Second, the network perspective facilitates direct measurement of peer characteristics, which allows for more stringent tests of influence and control theories. In the end, the entry intended to demonstrate the value for a network approach to the study of the peer-delinquency association which incorporates characteristics of the friendship groups in which adolescents are enmeshed.

This entry also addressed emerging areas related to social networks and peer influence. Such areas include changes in friendship structure and composition, influence from actors to whom one is indirectly tied, romantic partner influence, and co-offending. Finally, the entry identified innovative approaches to capturing nonschool social networks, including a respondent-driven approach based on police observation data, and affiliation networks based on activity-location overlap.

The network approach to adolescent delinquency provides a coherent and promising framework for investigating the variety of ways that peers shape and influence involvement in delinquency and crime. This approach is consistent with the current emphasis on the significance of social contexts (e.g., neighborhood, school) and helps understand the intricate ways in which individuals' behavior is shaped by the social networks in which they are embedded.

Related Entries

▶ Co-offending
▶ Control Theory
▶ Gangs and Social Networks
▶ Neighborhood Effects and Social Networks
▶ Social Control
▶ Social Control and Self-Control Through the Life Course
▶ Social Network Analysis and the Measurement of Neighborhoods
▶ Social Network Analysis of Organized Criminal Groups
▶ Social Network Analysis of Urban Street Gangs
▶ Spatial Models and Network Analysis

Recommended Reading and References

Akers RL (2009) Social learning and social structure: a general theory of crime and deviance. Transaction Publishers, New Brunswick

Browning CR (2011) The spatial and social embeddedness of youth activities. Paper presented at the annual meeting of the Association of American Geographers, Seattle

Ennett ST et al (2006) The peer context of adolescent substance use: findings from social network analysis. J Res Adolescent 16:159–186

Frank KA et al (2008) The social dynamics of mathematics course taking in high school. Am J Sociol 113:1645–1696

Giordano PC, Cernkovich SA, Pugh MD (1986) Friendships and delinquency. Am J Sociol 91:1170–1202

Giordano PC, Cernkovich SA, Holland DD (2003) Changes in friendship relations over the life course: implications for desistance from crime. Criminology 41:293–328

Giordano PC, Lonardo RA, Longmore MA (2010) Adolescent romance and delinquency: a further exploration of Hirschi's 'cold and brittle' relationships hypothesis. Criminology 48:919–946

Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Granovetter MS (1973) The strength of weak ties. Am J Sociol 78:1360–1380

Haynie DL (2001) Delinquent peers revisited: does network structure matter? Am J Sociol 106:1013–1057

Haynie DL (2002) Friendship networks and delinquency: the relative nature of peer delinquency. J Quant Criminol 18:99–134

Haynie DL, Osgood DW (2005) Reconsidering peers and delinquency: how do peers matter? Soc Forces 84:1109–1130

Haynie DL, Giordano PC, Manning WD, Longmore MA (2005) Adolescent romantic relationships and delinquency involvement. Criminology 43:177–210

Hirschi T (1969) Causes of delinquency. University of California Press, Berkeley

Kreager DA, Haynie DL (2011) Dangerous liaisons? Dating and drinking diffusion in adolescent peer networks. Am Sociol Rev 76(5):737–763

Kreager DA, Rullison K, Moody J (2011) Delinquency and the structure of adolescent peer groups. Criminology 49:95–127

Magino W (2009) The downside of social closure: brokerage, parental influence, and delinquency among African American Boys. Sociol Educ 82: 147–172

McCarthy B, Hagan J, Cohen LE (1998) Uncertainty, cooperation, and crime: understanding the decision to co-offend. Soc Forces 77:155–184

McGloin JM (2009) Delinquency balance: revisiting peer influence. Criminology 47:439–477

Meldrum RC, Young JTN, Weerman FM (2009) Reconsidering the effect of self-control and delinquent peers. J Res Crime Delinq 46:353–376

Moody J (1999) The structure of adolescent social relations: modeling friendship in dynamic social settings. University of North Carolina at Chapel Hill.UMI Dissertation Services, Ann Arbor

Papachristos AV, Braga AA, Hureau DM (2012) Social networks and the risk of gunshot injury. J Urban Health 89:992–1003

Payne D, Cornwell B (2007) Reconsidering peer influences on delinquency: do less proximate contacts matter. J Quant Criminol 23:127–149

Pratt TC, Cullen FT (2000) The empirical status of Gottfredson and Hirschi's general theory of crime: a meta-analysis. Criminology 38:931–964

Pratt TC et al (2010) The empirical status of social learning theory: a meta-analysis. Justice Q 27: 765–802

Steglich C, Snijders TAB, Pearson M (2010) Dynamic networks and behavior: separating selection from influence. Sociol Methodol 40:329–393

Sutherland EH, Cressey DR (1960) Principles of criminology, 6th edn. J.B. Lippincott, Philadelphia

Warr M (2002) Companions in crime: the social aspects of criminal conduct. Cambridge University Press, Cambridge

Weerman FM, Smeenk WH (2005) Peer similarity in delinquency for different types of friends: a comparison using two measurement methods. Criminology 43:499–524

Young JTN, Barnes JC, Meldrum RC, Weerman FM (2011) Assessing and explaining misperceptions of peer delinquency. Criminology 49:599–630

Networked Policing

► Family Engagement Strategies to Reduce Crime

Neurology and Neurochemistry of Crime

John Paul Wright[1,2] and Jamie Newsome[3]
[1]School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA
[2]King Abdulaziz University, Jeddah, Saudi Arabia
[3]Department of Criminal Justice, The University of Texas at San Antonio, San Antonio, TX, USA

Synonyms

Brain; Genetic expression; Neurotransmitters

Overview

Modern neuroscience has made significant and sometimes dramatic gains into understanding how the brain and central nervous system are implicated in a range of problematic, criminal, and addictive behaviors. Indeed, the growth of modern neuroscientific findings can only be described as explosive, with new discoveries being published almost daily. These findings are often produced by highly advanced technologies – technologies that, at a broad level, allow scientists to view the brain in action and under experimental conditions. This same technology also now allows scientists to directly measure brain cell activity and the strength and integrity of connections between brain cells. Never before in human history have scientists had these tools at their disposal.

Criminologists, once primarily wedded to sociological explanations of crime, are beginning to incorporate many of these findings into their understanding of the development of criminal conduct. They are also using these findings to understand the process of chemical addiction, how environmental experiences affect brain development, and how best to intervene with anti-social youths and adult criminals.

Apart from understanding brain structures and their functions, neuroscience has made tremendous progress in understanding how the chemicals in the brain, known as neurotransmitters, affect a range of personality traits and behaviors. This knowledge has led directly to the development of entire classes of psychopharmacological drugs. These drugs, known as selective-serotonin-reuptake-inhibitors (SSRI's), have been used to effectively treat a range of mood and psychological disorders. New stimulants, too, have been created that help individuals better regulate their own conduct, including those diagnosed with attention-deficit-disorder. The point is, what neuroscience reveals to us about behavior, including criminal and imprudent behavior, may lead to better, more effective interventions.

Evolutionary and Developmental Background

The human brain is the most complex computational device known to exist. It controls all necessary life functions, takes over 20 years to fully mature, and allows human beings to exercise tremendous control over their behavior and their environment. The human brain is a marvel of complexity and has been subject to evolutionary selection pressures for at least 250,000 years. Anthropological and genetic evidence indicates that the human brain has undergone at least two key evolutionary changes. The first is a mutation in the ASPD gene. The ASPD gene codes for the abnormal spindle protein homolog protein. Around 6,000 years ago a new allele of the ASPD gene emerged, resulting in progressively larger brains across humans of European, Asian, and Middle Eastern decent. Brain size is positively correlated with intelligence and is thought to allow for increases in the number of neurons and the number of connections between neurons. Growth in brain size, however, was not uniform. The area of the brain responsible for all higher-order thought processes, known as the cerebral cortex, expanded to the point where today it accounts for 77 % of total brain volume. Mutation in the ASPD gene corresponds with the advent of writing, agriculture, and science; although, it is unlikely to have directly caused them.

The second mutation also had a profound effect on human development and activity. Approximately 100,000 years ago a mutation in the FOXP2 gene occurred. Scientists believe that the mutation eventually allowed for the development of language. The mutation spread rapidly across ancestral populations, in large part because language provided early hominids a tremendous advantage in terms of survival. The area in the brain responsible for the production of language was discovered by a French physician, Paul Broca, and is known as Broca's area. The area of the brain that allows individuals to understand language was named after German psychiatrist Carl Wernicke, and is known as Wernicke's area. The development of language has allowed for tremendous advancements among human cultures. Interestingly, research has found that language deficiencies are positively correlated with antisocial behaviors.

Over 60 % of our genes code for the structure and function of the brain. In terms of nuclear material, the brain is the most expensive organ to create. The average adult brain weighs between 1,300 and 1,400 g, or about 3 lb. Because brain development is under strong genetic control, several regions of the brain appear to be strongly heritable, including regions that allow individuals to control their emotions, to plan for the future, and to engage in self-control. The heritability of brain structure and functioning necessarily means that the traits and characteristics of biological parents, both physical and behavioral, are likely to be shared and expressed by their offspring. Numerous behavioral genetic studies show that most traits, including self-control, intelligence, and specific personality features are modestly to strongly heritable. Part of the reason for the correlation between parent and offspring behavior and traits is likely to be due to similar brain architecture and functioning. Criminologists have known since the 1800s that criminal behavior and other social pathologies concentrate within some families and can show high degrees of inter-generational continuity.

The brain begins to develop almost immediately after conception. Research shows that during in-utero development the brain is forming over 250,000 connections per minute. While much of brain development is genetically controlled, evidence indicates that environmental factors can also be influential. Because the developing embryo is connected to the mother by the umbilical cord and placenta, maternal behavior can influence in-utero brain development of the fetus. Maternal nutritional intake is the clearest example. The fetus depends on receiving sufficient levels of nutrients from the mother. Insufficient nutritional intake can have devastating neurological consequences for the fetus. Moreover, maternal stress, and the accompanying release of stress hormones, has been shown to affect brain development of the fetus. While the brain of the developing fetus is protected by layers of fatty lipids, known as the blood–brain barrier, molecules from a variety of substances can pass from one side of the barrier into the developing central nervous system of the fetus. Maternal drug use and abuse, for example, has been found to potentially compromise the brain and central nervous system of the developing fetus. Cocaine molecules, as an example, easily pass across the blood–brain barrier. Moreover, lead (Pb), which can be stored in the bone of the mother, becomes bioavailable during pregnancy, passes through the blood–brain barrier, and may compromise healthy nervous system development. Prenatal lead exposure has been linked to a range of conduct problems, including delinquency in adolescence and criminal conduct in adulthood.

The rate of synaptogenesis, the creation of new nerve cell connections, corresponds directly to the increases in motor and cognitive abilities witnessed across childhood. Indeed, the brains of 2 year olds are about two-times as active as adult brains. Yet by late childhood the brain begins pruning unused and unnecessary synaptic connections. Reductions in the number of unnecessary connections eventually results in an increase in brain efficiency. The brain, in other words, does not have to use as much energy to do the same tasks in adulthood as it did in childhood. In part, this is due to the strengthening of connections through repetition; however, this is also due to the sculpting away of unused connections. While the sculpting begins in late childhood, it is accelerated during puberty when the brain undergoes a substantial period of reorganization. The onset of puberty also signals the myelination of the front part of the adolescent brain. Myelin is a fatty substance that will eventually cover axons. By covering axons in myelin the electrical impulse across the axon is accelerated by a factor of 10 over unmyelinated axons. The end result of these developmental changes is an energy efficient, highly integrated adult brain.

It is worth pointing out that although the brain undergoes substantial reorganization and pruning in the earlier stages of life, this does not mean that once one reaches adulthood the brain remains unchanged. On the contrary, the brain continues to change over time. Scientists do not yet know, however, how these changes are connected to criminal behavior.

Structures and Functions

The first part of the brain to develop is lower brain, which includes the brain stem and reticular activating system (RAS). The RAS contains a dense bundle of nerves that help direct sensory information to the correct parts of the brain. Psychologists have also linked variation in the RAS to certain personality characteristics, with some of these characteristics, such as neuroticism and extraversion, relating to offending. The lower brain controls all autonomic activities, such as breathing, that support basic life functions.

Sitting on top of the lower part of the brain is the midbrain, also known as the limbic system or the reptilian brain. The limbic system is commonly thought of as the alarm system of the brain in that it prepares the body for fight or flight when faced with a threatening situation. The limbic system controls overall neuroendocrine functioning, which regulates the release of hormones throughout the body. Containing a series of unique but highly inter-related structures, the limbic system is deeply implicated in crime and violence.

One of the more important structures, at least as it applies to understanding violence, is the amygdala. The amygdala is an almond-shaped structure responsible for human emotions and emotional memory. The term "emotion" comes from the early Greek word for motion. The Greeks understood the powerful influence of emotions in human motivation and behavior. Emotions are chemical states generated by the amygdala which can serve as powerful motivators. Several emotions have been implicated in crime and violence, such as fear, anger, disgust, and contempt. The amygdala is also implicated in social learning. Information, situations, and events that carry with them strong emotions are more likely to be stored by the brain and recalled in the future.

Other structures, such as the hippocampus, hypothalamus, thalamus, and cingulate gyrus, compose the limbic system and are also implicated in problem behavior. For example, the hypothalamus, with connections to the pituitary gland, helps to control endocrine functioning. The endocrine system regulates the flow of hormones and androgens, sex-specific hormones, in the bloodstream. Human sensations of bonding, for example, are produced by the release of vasopressin and oxytocin, while aggression has been linked to increased levels of testosterone in males and to estrogen in females. Moreover, the hippocampus is associated with the anticipation of consequences and the ability of individuals to learn from prior experience – known as conditioning. One of the hallmarks of adult criminals is their inability to learn from prior experience. Finally, the cingulate gyrus has been repeatedly linked to human aggression. The anterior cingulate gyrus influences attention, concentration, and focus. Reduced ACC volume and activity correlate with conduct disorder, obsessive thoughts, and compulsive behaviors.

Numerous studies have also implicated the hypothalamus-pituitary-adrenal (HPA) system in a range of emotional, psychiatric, and behavioral problems. The HPA axis is a complex network that regulates nervous system responses to stress. When the brain is under stress, the hypothalamus will release corticotropin-releasing hormone (CRH) and vasopressin. These molecules will be transported to the pituitary gland where they will cause the release of adrenocortictropic hormone (ACTH). ACTH will then stimulate the release of cortisol, a stress hormone, from the adrenal gland. These chemicals will then act on the hypothalamus and pituitary gland to eventually suppress CRH and ACTH production.

Extended exposure to stress can harm brain functioning and, in turn, can bring about a variety of non-adaptive stress responses. This may be why dysfunction in the HPA axis has been correlated with everything from aggression, to mood disorders, to drug-addiction and abuse. Indeed, prolonged exposure to intense stress appears capable of altering HPA axis functioning. This possibility has encouraged scholars and criminologists to examine how stress filled environments influence neurological activity. For example, some children of depressed, neglectful, or physically abusive mothers show dysregulation of the HPA axis. Blunted HPA responses to stress may occur because of repeated exposure to abuse and neglect. In turn, the affected child may be neurologically compromised in his response to threatening and difficult environmental circumstances.

The cerebral cortex sits atop the limbic structures and is divided into two hemispheres by the corpus collosum, which connects the two hemispheres through a dense bundle of fibers. The cortex has been called the key to civilization because all higher-order, uniquely human capacities are distributed across the cortex. There are approximately 10–20 billion neurons in the cerebral cortex generating between 60 and 240 trillion connections. The cortex is the last part of the brain to develop, taking approximately 20 years to fully mature, and is more susceptible to insult and injury in males than females. The limbic system, moreover, is intimately connected to the cerebral cortex.

The cortex is divided into the prefrontal cortex (PFC), the dorsolateral prefrontal cortex (DLPC) and the orbitalfrontal cortex (OFC). Collectively these areas are responsible for what are known as "executive functions." Executive functions include goal driven behavior, planning and execution of a plan, impulse control, delayed gratification, task switching, and emotional regulation. Deficits in executive functioning are of primary concern to criminologists. First, a large body of research evidence shows that criminal offenders are highly impulsive, frequently hedonistic, and have trouble delaying immediate gratification. A dominant theory in criminology refers to this syndrome as "low self-control" (Gottfredson and Hirschi 1990). Medical experts are able to assess brain function through a number of imaging techniques. A growing body of imaging evidence, including functional magnetic reasonance imaging (fMRI), positron emission tomography (PET), and single photon emission tomography (SPECT), reveals a pattern of under-activation in the in the left, but sometimes the right, PFC. The lack of activity is thought to correspond to deficits in the executive control functions of the brain. Self-control is known to be housed in the PFC, specifically the OFC. It is considered to be a relatively stable characteristic over long periods of the life-course. Activity levels in the OFC, however, can sometimes be adjusted with the use of prescribed stimulants.

Second, the PFC is the last part of the brain to fully develop, with myelination not complete until the start of the third decade of life. Because of this, adolescents appear to be more influenced by limbic impulses and emotional stimuli. Experimental studies have found that adolescents tend to over-estimate the potential emotional and social awards of a behavior, while they tend to under-estimate the actual risks associated with the behavior. Moreover, they appear to be more influenced by peer pressure than are adults. One of the developmental changes that takes place during adolescence is a shift from a reliance on initial limbic impulses to a reliance on the PFC as the primary brain mechanism used to solve problems. Interestingly, this evidence was used by the Supreme Court of the United States in *Roper v. Simmons*, 543 U.S. 551 (2005). The court held that the execution of individuals who were under the age of 18 years at the time of their capital crimes is prohibited by the Eighth and Fourteenth Amendments.

Contemporary neuroscience theorizes that aggressive and violent impulses originate in the deep limbic structures of the brain, primarily in the amygdala, where they stimulate an autonomic nervous system response that prepares the body for violence. Part of this response is an increase in the amount of hormones, such as testosterone, entering the blood stream, as well as an increase in muscle oxygenation, a loss of peripheral vision, and an increase in heart rate. Depending on the context, however, this impulse can be shunted by a sufficiently developed and interconnected PFC. The PFC causes a reduction in the initial limbic impulse and delays the behavioral actions that can accompany the impulse. This is why, for example, most people do not engage in violence even when they have been sufficiently provoked. Violent behavior can occur, however, when deficits in the PFC, either through a lack of active connectivity or insufficient connections to the limbic system, fail to filter the violent impulse.

Neurotransmission

Neurotransmission is the process by which electro-chemical messages are sent throughout the brain. Neurons are cells that facilitate this communication. There are approximately 100 billion neurons in the adult brain, and each neuron connects to between 1 and 10,000 other neurons. Neuronal activity is highly complex and is involved in every thought, emotion, and behavior engaged in by the human organism. Neuronal transmission occurs when a pre-synaptic neuron generates and sends an electrical charge down its axon, which is a tail-like extension emanating out of the cell. The electrical charge moves down the axon by passing through a series of positively and negatively charged sodium and potassium gates where it will eventually encounter the synaptic cleft. Electrical energy is then converted to chemical energy when the vesicle housing chemicals, known as neurotransmitters, releases molecules into synaptic space. Synapses, or the area between the axon of one cell and the dendrite of the another, is 1/600th the width of a human hair.

Neurotransmitters pass through synaptic space and bind to specialized receptors on the post-synaptic neuron. Once the neurotransmitters have bonded to the receptors, an electrical charge is sent back down the post-synaptic neuron and the transmission cycle is completed. The neurochemicals used in the process are reabsorbed by the neuron in a process called reuptake, or they are enzymatically degraded for future use.

Neuronal receptors are highly specialized and vary in number and sensitivity. The number of receptors located on a dendrite, and their efficiency and sensitivity, is genetically controlled and varies across individuals. Receptor sites are also specific to each neurotransmitter. For instance, dopamine receptors accept only dopamine molecules while serotonin receptors accept only serotonin molecules.

Neurotransmission frequently involves groups of neurons working collaboratively. These "neural networks" are responsible for a number of complex functions carried out within the brain. With repetitive stimulation, these networks become more efficient and send messages more quickly. In contrast, neurons that are not used or that are stimulated irregularly are pruned. This has led to the development of Hebb's rule: neurons that fire together, wire together.

Neurotransmitters can be classified as excitatory, causing an increase in cell firing, or inhibitory, which prevents cells from firing. Four neurotransmitters have been identified that are related to antisocial behavior. Dopamine is an excitatory neurotransmitter, which is involved with the reward and punishment systems in the brain, as well as cognitive functioning. Norepinephrine is another excitatory neurotransmitter. This chemical is tied to emotion and memory, and is involved in fight-or-flight responses. Gamma aminobutyric acid (GABA) and serotonin are inhibiting neurotransmitters. GABA is responsible for filtering stimulation from the environment to allow interpretation of experiences. Serotonin has been linked to mood and self-restraint. Each of these neurotransmitters has been found to be related to antisocial behavior.

Moderate levels of dopamine are associated with optimal cognitive functioning. It is known to enhance problem solving and to increase concentration. Dopamine is released in the brain when individuals participate in pleasurable activities, creating feelings of euphoria. This increases the likelihood that the behavior will be repeated. As dopamine is involved in the reward systems in the brain, it has been hypothesized that some individuals may experience reward deficiency syndrome (RDS). In these circumstances, individuals experience less neurological arousal from behaviors that would be experienced otherwise. For individuals suffering from RDS, behaviors such as eating or sex would not bring an optimal amount of pleasure. This condition is thought to lead individuals to be more impulsive, obsessive, or vulnerable to addiction. Commonly abused substances such as nicotine, caffeine, and a number of illicit drugs increase dopamine levels, and many substances such as cocaine, bind directly to dopamine receptors. As a result, dopamine is believed to play an important role in substance abuse and addiction.

Significantly high levels of dopamine have been associated with aggressive and criminal behaviors as well as reduced intelligence. A number of studies have shown that antipsychotic drugs known to reduce dopamine production have also reduced aggressive behaviors (Brizer 1988; Mollinger et al. 1962; Yudofsky et al. 1987). However, some studies have suggested that the effect of dopamine on aggressive behavior may be insignificant (Raine 1993).

Norepinephrine can be found in the cerebellum and hindbrain regions, as well as between the brain stem and cortex. It is distributed throughout the limbic system. During periods of heightened arousal norepinephrine is released, aiding individuals in responding to fright or anger. Provided that norepinephrine can be found in regions of the brain responsible for emotion and memory, these events are likely to remain in an individual's memory for a long period of time. Some drugs can block the reuptake of this neurotransmitter, resulting in an excess of this chemical in the brain. Generally, studies indicate that excessive amounts of norepinephrine are related to aggressive behavior (Brizer 1988; Fishbein 2001; Volavka 1999; Yudofsky et al. 1987); however, there is also evidence that lower levels of this neurotransmitter may be linked to antisocial behavior (Raine 1993). The available research indicates that the relationship between norepinephrine and maladaptive behaviors may be curvilinear; meaning both high and low levels of this neurotransmitter can increase problem behavior.

The most prevalent neurotransmitter is GABA (Wan et al. 1996). Primarily, GABA is used by the brain to regulate the functioning of the nervous system and cognitive functioning. This neurotransmitter works to balance excitation in the brain by inducing relaxation and sleep. Reduced levels of GABA can lead to increases in irritability, anxiety, and violent behavior. In contrast, excessive levels of this neurotransmitter depress cognitive functioning. Alcohol and the date rape drug gamma hydroxybutyrate (GHB) have effects on the brain similar to excessive amounts of GABA. The senses of the body are inhibited, and it becomes difficult to have a complete awareness of the immediate environment.

Serotonin is used by neurons found in the brain stem, the limbic system, and the frontal cortex. In addition to the influence genetics may have in the development of the serotonergic system, there are environmental factors such as seasonal changes that also seem to influence serotonin levels. This neurotransmitter has been found to play a large role in the regulation of behavior and mood. Low levels of this neurotransmitter have consistently been linked to criminal behavior among juveniles and adults (Brizer 1988; Fishbein 2001; Raine 1993). Recent research has indicated that serotonin is important in inhibiting impulsive behavior, and that impulsivity that is tied to strong emotion is more likely to result in violent outcomes (Krakowski 2003; Volavka 1999). Lower levels of serotonin have also been associated with violent forms of suicide. It has also been found to correspond with higher levels of substance abuse and a number of clinical disorders. There is even evidence that suggests that low serotonin, alcoholism, and violence may share an association.

There is clear evidence that an excess or deficiency in neurotransmitters can result in modifications in behavior. While the amount of these chemicals in the brain is influenced by the body's production of the chemical, it is also influenced by the regulation of the substance through its removal. One method of removing neurotransmitters that remain in the synaptic space after neurotransmission is through the process of enzymatic degradation. Monoamine oxidase-A (MAO-A) is an enzyme found within the brain which is responsible for metabolizing serotonin, norepinephrine, and dopamine. Having too little or too much MAO-A will interfere with neurotransmission. If an individual is deficient in MAO-A, an abundance of neurotransmitters may be left in the synaptic cleft after neurotransmission. In contrast, an overproduction of MAO-A may reduce the amount of a neurotransmitter in the transmission process. In either case, this interference can result in abnormal behavior.

Genes code for the level of MAO-A in individuals, and alleles for this enzyme are categorized as either high or low activity. Although research on the relationship between MAO-A and behavior is still in early stages, high and low levels of MAO-A have been associated with aggression and impulsivity among children and adults. There is also some evidence that suggests that varying levels of MAO-A may interact with environmental conditions producing maladaptive behavioral outcomes (Kim-Cohen et al. 2006). For example, those that possess the low activity allele and experience maltreatment during childhood may be more likely to engage in antisocial behaviors during adolescence or adulthood.

Controversies/Remaining Questions

While the available research indicates that various aspects of brain structure and function are implicated in the explanation of antisocial behavior, these relationships require further specification. For example, some studies have found that high concentrations of a neurotransmitter are related to maladaptive behaviors, but other studies demonstrate that lower concentrations of the same neurotransmitter also lead to maladaptive behaviors. The pattern of findings suggest that excesses and deficiencies result in behavior differences, but some scholars have also suggested that the relationship may be specific to types of offenses, such as property offenses, and to the nature of the offending, such as predatory or impulsive offending (Raine 1993). Either way, it is fair to say that while great advances have been made that much research is still necessary.

Neurological findings, in general, and those associated with human aggression and violence specifically, challenge many long held views about the nature of free will and choice in human action. Some scholars assume that accepting biological factors as contributing to human behavior eliminates the possibility of free will. They point, for example, to the neurological evidence showing brain deficiencies in the frontal cortex as evidence that behavior cannot be freely chosen. Moreover, other scholars have argued that since certain afflictions, such as alcoholism and drug-addiction, are highly heritable, that individual responsibility for conduct associated with these afflictions should be reduced. This is no small matter. The criminal justice system currently operates under the assumption that individual's make decisions based on free will. Neurological findings, however, have been used to question the extent to which free will exists and thus have been used to argue that the criminal justice system should more fully embrace a medical model – where the cause of criminal behavior is diagnosed and individual treatment, not punishment, provided.

Recent evidence adds to and expands this debate. Several studies have recently found that brain structure and function in the early life-course predicts brain structure and functioning in adulthood. The implications of these findings are not fully understood but they contrast with the images of rapid growth and change in brain functioning so prevalent in the research literature. It may be the case that the brain changes only within a limited range and that the changes detected in childhood through adulthood are bounded or restricted. If true, early brain activity may be a marker for problem behaviors later in life. While this information could be used to target youth for intervention, it could also be used to label youth as potential criminals. Hence, the ethics surrounding the use of neurological information in the identification and treatment of individual behavioral problems has yet to be sorted out. Do we, as a society, intervene in the lives of children and their families based on early neuro-scientific findings, or do we wait until problem behavior has emerged and someone has possibly been harmed? At this stage, the science is far ahead of our ethical understanding about the limits of the science.

Recommended Reading and References

Brizer DA (1988) Psychopharmacology and the management of violent patients. Psychiatr Clin North Am 11:551–568

Fishbein DH (1990) Biological perspectives in criminology. Criminology 28:27–72

Fishbein DH (2001) Biobehavioral perspectives in criminology. Wadsworth/Thomson Learning, Belmont

Goldberg E (2001) The executive brain: frontal lobes and the civilized mind. Oxford University Press, New York

Gottfredson M, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Kim-Cohen J, Caspi A, Taylor A, Williams B, Newcombe R, Craig IW et al (2006) MAOA, maltreatment, and gene–environment interaction predicting children's mental health: new evidence and a meta-analysis. Mol Psychiatry 11:903–913

Krakowski M (2003) Violence and serotonin: influence of impulse control, affect regulation, and social functioning. J Neuropsychiatry Clin Neurosci 15:294–305

Mollinger P, Lockner A, Sauls R, Eisenberg L (1962) Committed delinquent boys. Arch Gen Psychiatry 7:70–76

Rafter N (2008) The criminal brain: understanding biological theories of crime. New York University Press, New York

Raine A (1993) The psychopathology of crime: criminal behavior as a clinical disorder. Academic, San Diego

Rowe DC (2002) Biology and crime. Roxbury Publishing, Los Angeles

Volavka J (1999) The neurobiology of violence: an update. J Neuropsychiatry Clin Neurosci 11:307–314

Wan FJ, Berton F, Madamba SG, Francesconi W, Siggins GR (1996) Low ethanol concentrations enhance GABAergic inhibitory postsynaptic potentials in hippocampal pyramidal neurons only after block of GABAB receptors. Proc Natl Acad Sci USA 93:5049–5054

Yudofsky SC, Silver JM, Schneider SE (1987) Pharmacological treatment of aggression. Psychiatr Ann 17:397–407

Neurotransmitters

► Neurology and Neurochemistry of Crime

New Media and Crime Images

► Cultural Criminology

New Penology

Michelle Brown
Department of Sociology, University of
Tennessee, Knoxville, TN, USA

Overview

The new penology is a perspective that plots the rise of actuarial justice in understandings of crime and criminal justice. This model views crime as now normal and no longer a site for social reformation. Instead, the best that can be expected is the control of crime through techniques based upon technology and statistical calculations that enhance the identification and management of "high-risk," "dangerous," or unruly groups. This approach marks the rise of technocratic forms of knowledge no longer concerned with the rehabilitation or transformation of individual offenders but rather with the production of aggregate categories and classification schemes that locate and track actual and potential offenders. Mass incarceration and the warehousing of offenders become primary ways of managing risk in criminal justice simply by removing potentially dangerous individuals from society for long periods of time. This logic is argued to be a major influence on criminal justice agents, institutions, and policies where risk management and actuarial discourses, its proponents argue, have diffused throughout the justice system and into the community.

Fundamentals, Critiques, and New Directions

Introduction

A reaction to the decline of rehabilitation and the rise of "nothing works" discourse in the late 1970s and 1980s, the new penology is configured as a radical break from the past at a moment of penal crisis. As David Garland writes in *The Culture of Control*:

> Within the brief time it takes to progress from basic training to mid-career, a whole generation of practitioners – probation officers, prison officials, prosecutors, judges, police officers, and criminological researchers – have looked on while their professional world was turned upside down. Hierarchies shifted precariously; settled routines were pulled apart; objectives and priorities were reformulated; standard working practices were altered; and professional expertise was subjected to challenge and viewed with increasing skepticism. The rapid emergence of new ways of thinking and acting on crime, and the concomitant discrediting of older assumptions and professional orientations, ensured that many penal practitioners and academics lived through the 1980s and 1990s with a chronic sense of crisis, and professional anomie. (2001: 4)

This context of crisis became the setting for fundamental paradigm shifts in understandings of crime and punishment. In the closing moments of the twentieth century, the United States, for the first time in world history, imprisoned more people for the purposes of crime control than any other society, with a total prison and jail population of over two million people. This kind of penal expansion was in keeping with the kind of growth taking place across various forms of punishment from prison to probation. This expansion continued even as crime rates decreased. In fact, historically, there is no significant correlation between crime rates and incarceration rates; rather, offenders are more likely to commit a crime after release from prison than previous to incarceration.

For these reasons, criminologists began to look closely at the latent functions of punishment, those reasons and motives that are less apparent in rhetoric but materialize in practice. Against this unprecedented transformation, scholars began to ask a new set of questions: If the prison's function cannot be found in perhaps its most rational and logical justifications, including crime reduction, public safety, and rehabilitation, then why does the institution persist and why did it expand exponentially in the late twentieth century? How has this transformation reconfigured the implementation of justice, notions of individual and collective identity, and modes of governance? They also produced a wave of new explanatory concepts and theories – terms like "penological crisis," "incarceration binge," "prison-industrial complex," and a "new penology." The new penology, then, as articulated by its main theorists, sociolegal scholars Malcolm Feeley and Jonathan Simon, represents perhaps most fundamentally the understanding shared across much of criminology that "a different kind of historical change is unfolding in our contemporary penal practices" (Simon 1988: 452).

Background and Foundations

The New Penology

The new penology represents one of the more powerful accounts that seek to add insight to these problems and predicaments by highlighting some of the potentially permanent effects contemporary patterns in punishment may have upon the configuration of criminal justice and social life. Feeley and Simon introduced the new penology in a widely read 1992 *Criminology* article which they then followed with a body of work dedicated to revising and further developing their propositions (Feeley and Simon 1992, 1998). Their claims continue to serve as a critical point of reaction within the field of criminology. For these authors, the predominant transformation in punishment is theorized to be centered upon a "change in conception – discourses, objectives, and techniques – in the penal process... shifts that have multiple and independent origins and are not reducible to any one reigning idea" (1992: 449). This change in conception centers upon three core assertions, which they describe as follows:

1. The emergence of new discourses: In particular, the language of probability and risk increasingly replaces earlier discourses of clinical diagnosis and retributive judgment.
2. The formation of new objectives for the system: The objectives we have in mind are not simply new to the system (some of them have old antecedents) but are in some sense newly "system." We are especially interested in the increasing primacy given to the efficient control of internal system processes in place of the traditional objectives of rehabilitation and crime control. Goals like reducing "recidivism" have always been internally shaped in important ways, but in the contemporary setting, the sense that any external social referent is intended at all is becoming attenuated.
3. The deployment of new techniques: These techniques target offenders as an aggregate in place of traditional techniques for individualizing or creating equity (Feeley and Simon 1992: 450).

The emergent discourse, objectives, and techniques of the new penology express a fundamental reorganization in contemporary understandings of crime and punishment, identified in such patterns and practices of surveillance and regulation as preventive detention, profiling, intensive supervision programs, the buy-busts and drug testing practices at the heart of the drug war, the targeting of subpopulations through policies centered upon selective incapacitation (habitual, career, and "three strikes" offenders), emergent statistical and software packages that track crime, and the increase in imprisonment alongside of the rising prominence of policy and crime experts. The primary contours of this reorganization, as marked by Simon, include a shift from the individual to the aggregate, from normalization to management, and from social norms to systems. This perspective privileges utilitarianism over moral reasoning, quantitative methods over qualitative analysis, and the notion of criminal justice as a series of operations and systems over individual-focused justice and rational actor theories.

Governmentality, Risk, and Actuarialism

> It is actuarial. It is concerned with techniques for identifying, classifying and managing groups assorted by levels of dangerousness. It takes crime for granted. It accepts deviance as normal. It is skeptical that liberal interventionist crime control strategies do or can make a difference. Thus its aim is not to intervene individuals' lives for the purpose of ascertaining responsibility, making the guilty 'pay for their crime' or changing them. Rather it seeks to regulate groups as part of a strategy of managing danger. (Feeley and Simon 1998: 375)

The new penology is squarely centered in larger frameworks of governmentality, risk, and actuarialism. Governmentality, with its implied rationalities, centers upon knowledge and practices that operate across and are shared between individuals, institutions, and the state. The new penology is most formidable in that its discourses, objectives, and techniques come to operate broadly across the social body as a form of governance. The new penology thus marks a move away from one of criminology's dominant frames in understanding punishment, a disciplinary model. Disciplinary modes of punishment are organized around an intense focusing upon the individual down to the most minute of details and habits. This kind of disciplinary focus, as exemplified by Foucault, is built around surveillance and normalization of the offender, inmate, delinquent, worker, and student, embodied in the closing of the gap between a norm and its deviation, the transformation of the individual self. The new penology, which privileges an actuarial model, is no longer concerned with closing this gap between individuals and society but rather seeks an accommodation of such variables through the management and mapping of distributions of populations. As Simon argues, in part, "the movement from normalization (closing the gap between distribution norm) to accommodation (responding to variations in distributions) increases the efficiency of power because changing people is difficult and expensive" (1988: 714). In contemporary social and political contexts, "it is cheaper to know and plan around people's failings than to normalize them" (Ibid). The new penology no longer seeks to normalize populations but rather to regulate them and their differences through segregation across abstract spaces, groupings, and social distributions that ironically culminate in even deeper exclusionary divisions, including the creation of a "functional" or permanent underclass. This phenomenon is noted in the governmentality literature as governing from or at a distance.

Within this new mode of governance, crime control efforts on the part of the state are defined by an abrogation of responsibilities by institutions that are already and historically under-sourced and under-trained. These justice agencies, envisioned originally as the last resort in the line against social problems, in many ways are now the first resort in the management of the mentally ill, homeless, economically disadvantaged, etc. Such a phenomenon also seeks to responsibilize individual actors even as it restricts the discretionary decision making of frontline justice actors and then implements unwieldy mechanisms of accountability through performance indicators. This mode of governance is most compelling in its ability to move out from political centers through dense, overlapping networks in which authorities, groups, institutions, and individuals all reflect a new emphasis no longer on production but the calculation of the activities and desires of others. Such entities will direct classification via an assessment of risk, including where it is most likely and how it might most efficiently be managed. Under the new penology, risk, in the form of "dangerous" populations, is most effectively managed, not through normalization, but spatialization marked by a well-mapped exclusion and isolation through mechanisms of classification rather than the more disciplinary mode of segregation through social control institutions. This pattern is most visible when social control is organized in a manner which extends beyond justice agencies into families, communities, schools, churches, and other primary social institutions as well as the fabric of social life. For instance, increasing attention and debate has centered upon the mapping of sex offenders in communities through social policy and legislation as groups who remain intractably in need of management, grounded in a fundamental public pessimism about the possibility of individual reformation or change. In the clinical and justice discourses surrounding that management, the vocabulary of actuarialism through risk management models is strongly apparent. These patterns are significant in the manner in which they invoke penality as a way in which to generate fundamental changes in political culture, generate regimes of truth, exercise power, differentiate groups, classify individuals, and thus, in general, reorder social life. In such a context, the material and institutional structures of the new penology have social and political effects with life organized around crime control and a perpetual search for security.

Such patterns are strikingly similar to other emergent sociological concepts, including Zygmunt Bauman's paradigm of exclusion, Jock Young's exclusive society, and Loïc Wacquant's urban outcasts and "deadly symbiosis" of community and prison. In such contexts, Foucauldian disciplinary modes, with their emphasis upon normalization – and the closing of a gap between a norm and its deviation – no longer dominate. Rather, these gaps are simply accommodated and regulated largely through exclusionary practices. In contemporary social and political contexts, Simon argues "it is cheaper to know and plan around people's failings than to normalize them" (1988: 771). This marks a critical site from which the study of punishment must carefully distance itself, they imply, in order not to reproduce these exclusions and normalizations through an institutionalized myopia in its own reports and research. Rather, the place of experts is one that calls direct and critical attention, like Feeley and Simon, to these dangerous convergences. This kind of engagement with governmentality speaks to a theoretical breadth in criminology that marks not only how emergent modes of governance shape the nature of crime and punishment but of what it means to be social. Risk and actuarial logics, which often omit sociological thinking, make it more difficult for groups and individuals to build solidarity and exercise political choice. At stake is the preservation and protection of human agency from forms of regulation and classification that may, at first glance, seem all too rational, logical, and "natural." In such contexts, the privileging of actuarial techniques risks denying the moral, political, and social significance of the historical markers of vulnerability – race, class, gender, and age – instead arguing these differences are no longer problematic, closing off significant social debates a priori. Beyond this, individual and group identities are rendered flat and sterile with little critical reflexivity. As Simon argues, "Rather than making people up, actuarial practices unmake them" (Ibid: 792). Out of this emerges a new and dangerous kind of social collectivity, one that is defined not by a deep sense of shared bonds or experiences but by exclusion. This is one of the larger implications of a new penology.

In such scenarios, the weakening of social welfare frameworks for governance has led to an expansion and exportation of punitive carceral strategies by the United States, a new mode from which to regulate minorities, the poor, and the socially abandoned across the planet. Importantly, these shifts are not hidden or invisible but rather prominent, naturalized features of emergent urban landscapes. Such transformations map the manner in which sovereignty is reconfigured in a broad social and global landscape with the imprint of a distinctly American mode of punishment. Importantly, citizenship in these contexts is experienced as hierarchical, determinedly and dangerously paternalistic, and punitive in its reliance and reproduction of social and cultural oppositions of race and class, all of which flow into and out of insecurities surrounding criminality and dangerousness. In this sense, the new penology is a distinctly dangerous way of understanding crime and punishment in its capacity to feed into social practices that extend globally and foundationally into social order and life. For these reasons, the defining contours of the new penology are built around broader notions of governmentality, risk, and actuarialism, practices that:

> are so familiar and banal that it is difficult to notice them at all, let alone see them as central components of a new regime of social ordering linked to myriad exercises of social control and power, e.g., hiring, admitting, campaigning, selling, sentencing, and educating. Yet these practices are generating fundamental changes in our political culture. (Simon 1988: 771)

These practices are argued to derive from actuarial techniques that "play a central role in a proliferating set of social practices. They are at the same time a regime of truth, a way of exercising power, and a method of ordering social life" (Ibid: 772). Actuarialist models emerge as complex mixtures of styles of thought, discourse, and vocabulary as well as sets of practices, strategies, and technologies. Grounded in efficiency, they value a certain technicism where system streamlining on the basis of cost-benefit analysis and risk assessments is foregrounded. They are driven by an economy of management centered upon the distribution of aggregate numbers of individuals across the justice system and an extension of these patterns of control across the everyday life of offenders and system professionals – "a fundamental respatialization of control mechanisms, from the enclosed spaces of carceral institutions to the dispersed territory of 'the community'" (Rose 2002: 212). Their primary tools are prediction instruments, classification schemes of "at-risk" populations, distribution grids, and technology, all of which are conducive to the production of emergent, wider networks of control.

These features reflect the manner in which actuarial understandings of the world are perceived as having the potential to fundamentally transform social life, playing pivotal roles in the organization of power, expertise, and social control. In this system-based focus, justice itself is reconfigured, and, as sociologist David Garland argues, "new performance indicators tend to measure 'outputs' rather than 'outcomes,' what the organization does, rather than what, if anything, it achieves" (1996: 458; See also Garland 2001). In this regard, the new penology marks a larger shift in the history of social control, the first grounded in a "crisis of penological modernism," devoid of philosophical underpinnings and organized instead around a collapse of faith and implicit sense of futility in deterrence and rehabilitation. Its power is in the implication of a particular form of logic that facilitates other kinds of penal aims, including incapacitation and retribution.

Centered upon the prevalence of uncertainty in the discretionary patterns of justice, the new penology offers a formula for control and the semblance of predictability in terrain that has systematically denied and subverted efforts and reform and control. The center of the new penology is consequently and, not surprisingly, a paradox:

> Across its various forms and calculative techniques, one feature stands out: risk thinking seeks to bring the future into the present and make it calculable. We could say that it tries to discipline uncertainty, in the sense of making uncertainty the topic of a branch of learning and instruction. It acts as well to discipline it in a second sense, by bringing uncertainty under control, making it orderly and docile. Risk thinking tames chance, fate, and uncertainty by a paradoxical move. (Rose 2002: 214)

Its attention to risk assessment through testing instruments and statistics serves as well to automatize and conceal a host of decision-making processes by practitioners and experts, instead often responsibilizing actors for their own risk classifications. In the end, justice actors, experts, subjects, and citizens all find themselves caught up in the tensions that define the new penology. And yet, Feeley and Simon argue that this very reality is concealed by the nature of the aggregate that actuarial practices exemplify, constituting a new kind of social collectivity, "one defined neither by internal bonds, nor external experiences, but by locations on a statistical distribution" (Simon 1988: 789). These classification schemes produce particular kinds of identities and individualities, but, as Simon argues, these techniques lack a distinctive sense of subjectivity and group identity itself is treated as "singularly sterile in [its] capacity for political empowerment"

(Ibid: 789). Thus, "actuarial techniques can be used to identify people who are more or less likely to be high-rate offenders, but these variables are not integrated into a conception of the underlying subject" (Ibid: 791).... "actuarial classification, with its de-centered subject, seems to eliminate, in advance, the possibility of identity, of critical self-consciousness and of intersubjectivity" (Ibid: 792). In that sense, the new penology is a symptom of a larger crisis within governance and thus apparent in punishment, one that seeks to regulate rather than explain or change contemporary social problems. In doing so, the meanings of the social and the self are argued to be fundamentally transformed.

Critiques and New Directions

Critiques

In Feeley and Simon's own words, their "thesis has sparked vigorous debate" (1998: 148). Many have critiqued the "newness" of the new penology and instead perceive it as a repackaging to some extent of preexisting theoretical frames. For some, the new penology represents an invocation of old penological and criminological themes with its emphasis upon classical Marxist terms like the "dangerous classes." Others argue that the new penology is unable to account for the rise of populist understandings of punishment centered upon retribution and an emotive turn in law and order politics. Others wonder to what degree frontline actors and professionals in the criminal justice system have internalized the actuarial logic of the new penology. Similarly, it is not entirely clear how the new penology marks a fundamental rupture from disciplinary to actuarial modes, or from modern punishment to something postmodern or otherwise, with critics pointing to the presence of actuarial logics historically in criminal justice as well as the persistence of individualistic, rational actor orientations that privilege retribution and expressive modes of populist sentiment in contemporary contexts.

Sociolegal scholar Mona Lynch writes, "One site for examining whether the new penology is indeed emerging as a distinct penological operating system is at the place where strategy is put into penal practice: at the level of implementation. There is reason to think that the contact point between the institution (and its new penological policies) and the outside world (with its old-fashioned "modern" take on criminality) is one battleground where the sustainability of the new penology may be tested" (1998: 842). Lynch models an empirical testing of the new penology through a rigorous ethnography of a California parole field office. Her findings point to the manner in which features of old and new penology meld in the everyday settings of criminal justice agencies, with actors reaffirming conventional law enforcement roles, including an individualistic focus, even as upper management developed policies framed by actuarial logics of risk management.

Criminologist Leonidas Cheliotis makes a case as well for the unpredictable aspects of human agency within institutional frameworks as a check upon the totalizing tendencies of a new penology. In an article where he examines the ways in which the defining boundaries of the new penology foreclose other possibilities, he writes:

> In so far as criminology aspires to influence the penal policy debate to any significant extent, analyses that fail to identify 'the real human stuff of disposition, choice and action – the stuff of which society and history are actually made'... cannot but be fatalistic and nihilistic, thereby inauspiciously (one could also say 'by way of benign neglect') contributing to the reproduction of the very crises they are supposed to pre-empt or combat.... Yet, in the dearth of convincing evidence that they signify a 'marked break with the past', and given that actuarialism is neither inevitable, nor necessarily undesirable in its entirety, the concept of new penology lacks interpretive adequacy and, consequently, remains a mere hypothesis. And this is a hypothesis which, in itself, Feeley and Simon would surely wish to see eventually disproved. (2009: 331)

New Directions

Aside from these critiques, the centerpiece of the new penology is found in Simon's introduction of the idea of "governing through crime" in the volume by the same name (2007). As crime becomes the primary way in which a broad variety of social problems and social action comes to be configured, Simon finds that the penal state itself becomes a framing logic not simply for the urban centers hard-hit by crime and violence through spiraling inequalities of race, class, gender, and ethnicity but for the sites and centers for the performance of middle-class life – work, education, family, and health. Crime and punishment become the larger context for shaping and regulating the conduct of others. The construction of issues related to work performance, education, divorce and child custody, etc. as "crime" issues allows for a host of regulatory mechanisms that would not otherwise be available: drug testing and other forms of surveillance; harsher, more punitive treatment; segregation; restriction of mobility; and forced movement, such as deportation. In Simon's estimate, governing through crime exceeds the conventional parameters of criminal justice and has seeped into the fabric of social life by way of other social institutions and everyday life frameworks. In this way, governing through crime is not really about crime but is instead a useful tactic or legitimation strategy that enables actors to manage other kinds of disputes, annoyances, social problems, or personal and public issues. Its rise includes an expansion of executive power, the privileging of a new and dangerous mode of political subjectivity – the victim, the weakening of courts and judges, and the rise of mass incarceration. In this manner, governing through crime produces citizens who are not victims themselves but live in fear of being victims across complex regulatory frameworks that deny the formation of other kinds of subjectivities – the kinds of informed citizen actors upon which democracy depends.

Finally, the new penology foreshadows a growing sense of pessimism and futility in criminology and criminal justice alike. Centered in correctional managerialism, actuarial justice, and the privileging of custody, the discourses, objectives, and techniques of the new penology share key elements with what works models. Grounded similarly in the crisis of the state, a heightened perception of high crime against the futility of punishment and a hardening of popular attitudes toward criminals, both discourses speak to the limits of science. However, the new penology is marked by an expansion and elaboration of exclusionary tactics through the convergence of scientific and bureaucratic mechanisms as managerial strategies, practiced and implemented through new technological possibilities in the realm of empirical classification and social control. More significantly, the new penology, through such public and private convergences, extends outward into the production of identities and subjectivities well beyond crime across the social fabric – governance at a distance and through crime. Such a perspective relates specifically to the despair and futility of nothing works in that, as an actuarial objective, it "does not offer an answer to this perceived hopelessness. Instead, it takes for granted that no answer exists, and offers the placating alternative of at least managing this hopelessness in a less burdensome and more efficient fashion... a chance to succeed, although only through a redefinition of goals" (Kempf-Leonard and Peterson 2000: 68).

The linkages between what works discourses and the new penology are complex, but importantly many of the aims, methods, and techniques of evidence-based research are potentially and aptly directed at new penology objectives with their privileging of techniques for identifying, classifying, and managing groups assorted by levels of dangerousness. In such worlds, science can collide again with familiar terms in a way that "takes crime for granted... accepts deviance as normal... is skeptical that liberal interventionist crime control strategies do or can make a difference. Thus its aim is not to intervene in individuals' lives for the purpose of ascertaining responsibility, making the guilty 'pay for their crime' or changing them. Rather it seeks to regulate groups as part of a strategy of managing danger" (Feeley and Simon 1998: 375). In keeping with new penology claims, much of the contemporary study of punishment has concerned itself with the absence of social vision and metanarratives that might anchor the practice of punishment in democratic, humane discussions. As Adrian Cherney argues, “Any form of social policy hinges on more than a rational analysis of ‘what works.’ Instituting change in how society responds to social problems depends also on the ability to articulate a vision – a public ideal around which reform can be mobilized....” (2002: 51), a collective possibility that is constrained by the objectives, discourses, and techniques of the new penology.

Conclusion

Feeley and Simon argue that the new penology might best be envisioned not “as a theory or program conceived in full by any particular actors in the system, but as an interpretive net that can help reveal in the present some of the directions the future may take” (1992: 460). In this future, they argue that the essential elements of a new penology are powerful enough to become “permanent features of the criminal justice system” (Ibid: 470). As knowledge concerning the features and qualities of particular groups grow more distinct in public records and classification schemes, “the ideological effects of actuarial practices... render more difficult the formation” of community organization and social movements where “the effect of actuarial practices is precisely to make it more difficult for groups to intensify their solidarity or exercise political choice” (Simon 1988: 786–787). One concern then of the critics of a new penology is the preservation and protection of human agency from forms of regulation and classification that may, at first glance, seem all too rational and logical. As Simon argues in early work:

> Today’s actuarial practices presage the development of a third model of politics where neither status nor class provides the basis for engagement. Indeed, we have no real models of what aggregate politics look like, but extrapolating from current conditions leads to a disturbing picture. Actuarial practices can mobilize segments of the population and form majorities that have no basis for understanding themselves as motivated by a common cause. (1988: 793)

Mapped into dramatic changes in law and administration, across such vast sociolegal landscapes as workplace regulation, sex offender registries, immigration practice, homeland security, and the war on terror, the new penology is a prescient and cogent reminder of the present and potential face of governance now and later.

Related Entries

- ▶ Incapacitation
- ▶ Managerial Court Culture
- ▶ Postmodern Criminology
- ▶ Risk Assessment, Classification, and Prediction

Recommended Reading and References

Cheliotis L (2009) How iron is the iron cage of new penology? the role of human agency in the implementation of criminal justice policy. Punishm Soc 8(3):313–340

Cherney A (2002) Beyond technicism: broadening the ‘What Works’ paradigm in crime prevention. Crime Prev Commun Saf Int J 3(4):49–59

Feeley M, Simon J (1992) The new penology: notes on the emerging strategy of corrections and its implications. Criminology 30(4):449–474

Feeley M, Simon J (1998) Actuarial justice: the emerging new criminal law. In: Pat O’M (ed) Crime and the risk society. Ashgate, Brookfield

Garland D (1996) The limits of the Sovereign State: strategies of crime control in contemporary society. Br J Criminol 36(4):445–471

Garland D (2001) The culture of control. University of Chicago Press, Chicago (Check)

Kempf-Leonard K, Peterson ESL (2000) Expanding realms of the new penology. Punishm Soc 2(1): 66–97

Lynch M (1998) Waste managers? The new penology, crime fighting, and parole agent identity. Law Soc Rev 32(4):839–869

Rose N (2002) At risk of madness. In: Simon J, Baker T (eds) Embracing risk: the changing culture of insurance and responsibility. University of Chicago Press, Chicago, pp 209–237

Simon J (1988) The ideological effects of actuarial practices. Law Soc Rev 22(4):771–800

Simon J (1993) Poor discipline. University of Chicago Press, Chicago

Simon J (2007) Governing through crime. Oxford University Press, New York

Nodal Policing

▶ Family Engagement Strategies to Reduce Crime

Normlessness

▶ Social Disorganization and Terrorism

Nuclear Weapons and State Crime

Ronald C. Kramer
Department of Sociology, Western Michigan University, Kalamazoo, MI, USA

Overview

Since the advent of the nuclear age, only a relatively small number of lawyers and criminologists have addressed the issue of nuclear weapons from anything resembling a criminological perspective. From 1945 on, questions concerning the legality or illegality of "the bomb" were raised, and over the years, a number of legal scholars and courts have analyzed the legal status of nuclear weapons under international law (Falk 1965, 2008a, b; Lawyers' Committee on Nuclear Policy 1990; International Court of Justice 1996; Boyle 2002). One legal analyst argues "that we need to recognize the intrinsic criminality of any threat or use of nuclear weapons" (Falk 2008a, p. 43). However, the number of criminologists who have advocated for or directly engaged in criminological inquiry concerning nuclear weapons is quite small (Friedrichs 1985, 2010a, b; Kauzlarich and Kramer 1998; Kramer and Kauzlarich 1999, 2011; Kramer and Bradshaw 2011). One prominent criminologist has observed: "It is quite remarkable that the vast criminal potential in the use of nuclear weapons has been neglected by criminologists" (Friedrichs 2010a, p. 132). And he also points out that the nuclear threat is a "monumentally important issue" that should be part of what he calls a "prospective criminology" of state crime (Friedrichs 2010b, p. 77).

This entry will address the topic of nuclear weapons and state crime. Three broad areas will be examined. First, the question of the illegality and/or criminality of nuclear weapons will be assessed. Second, specific forms of crime committed by the US nuclear state since 1945 will be examined. These state crimes include the actual use of atomic bombs against Japan during World War II, the repeated threats to use nuclear weapons in various conflict situations since then, and the continued possession of nuclear weapons by the USA and others in violation of the solemn legal obligation to disarm, which in turn has lead to a variety of environmental crimes and the potential for nuclear terrorism. Finally, structural and cultural explanations that have been advanced regarding these state crimes will be briefly discussed.

The Illegality of Nuclear Weapons

Although the criminologists who work in this area could argue that the use, threat to use, and continued possession of nuclear weapons are state crimes from a broader analogous injury/social harms approach, the extant criminological analysis of these acts has thus far been carried out within a strictly legalistic framework. The use, threat to use, and continued possession of nuclear weapons are considered state crimes because they involve decisions and actions by state officials, acting as representatives of the state, that violate specific *public international laws*, such as International Humanitarian Law (IHL) and the Nuclear Non-Proliferation Treaty (NPT) of 1968. As defined by the 1945 Nuremberg Charter, these illegal state actions can be designated as *crimes against humanity* and/or *war crimes*. As Agnew (2011, p. 31) points out: "The international law is especially useful for identifying harms committed by states."

Questions concerning the legality of nuclear weapons were raised early on. The first court case to address the issue was the *Shimoda* decision, handed down by the District Court of Tokyo on December 7, 1963. In May of 1955, five Japanese nationals instituted a legal action against their government to recover damages for the injuries they had suffered as a result of the atomic bombings of Hiroshima and Nagasaki. Part of their claim for compensation was based on the assertion that the dropping of the atomic bombs as an act of hostilities was illegal under the rules of positive international law. In its important ruling on the case, the Japanese court held that the atomic bomb attacks on Hiroshima and Nagasaki did indeed violate international law, although it rejected the plaintiff claims for compensation. In his classic appraisal of the *Shimoda* case, Falk (1965, p. 776) clearly summarized the principal reasons the court gave for its decision that the atomic attacks were illegal:

1. International law forbids an indiscriminate or blind attack upon an undefended city; Hiroshima and Nagasaki were undefended.
2. International law only permits, if at all, indiscriminate bombing of a defended city if it is justified by military necessity; no military necessity of sufficient magnitude could be demonstrated here,
3. International law as it has specifically developed to govern aerial bombardment might be stretched to permit zone or area bombing of an enemy city in which military objectives were concentrated; there was no concentration of military objectives in either Hiroshima or Nagasaki.
4. International law prohibits the use of weapons and belligerent means that produce unnecessary and cruel forms of suffering as illustrated by the prohibition of lethal poisons and bacteria; the atomic bomb causes suffering far more severe and extensive than the prohibited weapons.

Taken together, the international laws that the Japanese court based its ruling on constitute what is called International Humanitarian Law (IHL) or the laws of war. Since the *Shimoda* ruling, a significant number of international law scholars have concurred with its analysis of the illegality of nuclear weapons (Boyle 2002; Falk 2008b; Lawyers' Committee on Nuclear Policy 1990). The position of these legal analysts that the use or threat to use nuclear weapons violates IHL and would therefore constitute a war crime received a tremendous boost on July 8, 1996, with the delivery of an important decision by the International Court of Justice (ICJ), also known as the World Court. In an advisory opinion, *The Legality of the Threat or Use of Nuclear Weapons*, the ICJ ruled that the threat or use of nuclear weapons would generally be illegal under international law. This decision came in response to a request from the United Nations General Assembly (UNGA) for an advisory opinion on the legality of nuclear weapons. In a complicated and controversial ruling, the ICJ stated "…the threat or use of nuclear weapons would generally be contrary to the rules of international law applicable in armed conflict, and in particular, the principles and rules of humanitarian law" (International Court of Justice 1996, paragraph 105 (2) E).

It is important to stress that the ICJ advisory opinion did not create new law. Rather, the decision clarified that existing public international law, particularly the general principles and rules of international humanitarian law applicable in armed conflict (the laws of war), did apply to the use or threat to use nuclear weapons. The Court's ruling could be characterized as, to borrow a phrase from Justice Robert H. Jackson, US Chief of Counsel at the Nuremberg Tribunal, "declaratory of existent law." The opinion supports the ruling of the Japanese court in the *Shimoda* case that the United States did violate international law by dropping atomic bombs on Japan.

As the Nuremberg Tribunal made clear, the violation of the laws or customs of war are *war crimes*, and inhumane acts committed against civilian populations are *crimes against humanity*. Thus, the use or threat to use nuclear weapons is not only illegal but also *criminal*. As one legal scholar points out concerning the ICJ ruling: "Whenever the court discusses violations of the laws and customs of war; or violations of the Hague Conventions and Protocols; or violations of international humanitarian law, etc. with respect to the threat and use of nuclear weapons, the reader must understand that such violations are not just 'illegal' and 'unlawful' but are also 'war crimes' and thus 'criminal' under basic principles of international law that have been fully subscribed to by the United States government itself" (Boyle 2002, pp. 71–72).

Additionally, one of the most important aspects of the International Court of Justice's ruling on the illegality of nuclear weapons is the call, by all 14 judges, for the abolition of these weapons. The ICJ concluded its advisory opinion with an examination of the duty of states to negotiate in good faith a complete nuclear disarmament. The relevant treaty obligation the Court relied on is that found in Article VI of the 1968 Nuclear Non-Proliferation Treaty that states: "Each of the Parties to the Treaty undertakes to pursue negotiations in good faith on effective measures relating to cessation of the nuclear arms race at an early date and to nuclear disarmament, and on a treaty on general and complete disarmament under strict and effective international control."

At the time it was negotiated, the NPT was considered a "double bargain" (Blix 2008, p. 44). The nonnuclear weapon states agreed not to develop the weapons and accepted international inspection. On the other side, the five nuclear weapon states at the time committed themselves to negotiations that would lead to general and complete nuclear disarmament. That commitment was spelled out in Article VI, and on the basis of this provision, a unanimous World Court (International Court of Justice, paragraph 105 (2) F) found that: "…there exists an obligation to pursue in good faith and bring to a conclusion negotiations leading to nuclear disarmament in all its aspects under strict and effective international control." The ICJ interpreted the NPT provision as imposing a legal obligation to achieve a precise result-nuclear disarmament in all its aspects.

State Crimes Related to Nuclear Weapons

Using this legal framework, several criminologists have argued that a number of state actions related to nuclear weapons can and should be examined as forms of state crime. As noted above, these crimes include the actual use of atomic bombs against Japan in 1945, various threats by the USA to use nuclear weapons in conflict situations, and the continued possession of these weapons in violation of the NPT. In addition, they also assert that the continued possession of nuclear weapons leads to a variety of other criminal harms.

Kramer and Kauzlarich (2011) argue that the use of atomic bombs against Hiroshima and Nagasaki were the first crimes of the "American nuclear state." However, the political, military, moral, and legal interpretations of the atomic bombings are, to this day, bitterly contested. Many basic questions concerning the decision to drop the atomic bomb can be raised. The questions that criminologists have addressed, however, are the following: Was the use of atomic weapons against Japan illegal? Was the use of the bomb a war crime and a crime against humanity? What structural and cultural factors explain the decision to "drop the bomb?"

Some argue that it is difficult to address these criminological questions about the atomic bomb without first taking into account the more general phenomenon that developed during World War II, of the movement from sporadic, selective, and tactical attacks on military and industrial targets to the bombing of cities in an effort to terrorize and kill civilians (Kramer and Kauzlarich 2011). The aerial bombardment of civilian populations in urban areas, variously referred to as "area bombing," "strategic bombing," or "total war," has been examined by numerous scholars outside of criminology. Many of them concur with the assertion that "…deliberately mounting military attacks on civilian populations, in order to cause terror and indiscriminate death among them, is a *moral crime*" (Grayling 2006, p. 4). While the main question concerns the use of atomic weapons against Hiroshima and Nagasaki, those acts cannot be analyzed separately from the aerial bombing of civilians in cities that preceded them.

Returning to the question at hand: Was the use of atomic weapons against Hiroshima and Nagasaki unlawful and criminal? Kramer and

Kauzlarich (2011) assert that they were. They argue that the atomic bombings were objectively illegal, a war crime, because they violated the rules and principles of international humanitarian law that existed in 1945, the law that the *Shimoda* decision was based on, and the law that the ICJ would rely on in its momentous decision in 1996. These international legal rules for the conduct of warfare were codified and promulgated by the United States government itself in the War Department's Field Manual 27-10, entitled *Rules of Land Warfare*, which was issued on October 4, 1940 (and amended on November 15, 1944). As one legal scholar has demonstrated, the principles of international law specified in Field Manual 27-10 were binding on US officials throughout World War II and thus prohibited the use of atomic weapons against Japan (Boyle 2002). Thus, the *Shimoda* decision, the US government's own Army Field Manual, the ICJ ruling, and the considered opinion of many international law scholars make it clear that from an objective legal perspective at least, the atomic attacks on Japan did constitute a state crime. As one legal scholar notes: "…the unavoidable legal conclusion [is] that these attacks remain unacknowledged crimes against humanity of the greatest magnitude. The use of the atomic bomb in World War II was not merely a violation of the laws of war…but was also a criminal act of the greatest severity for which the perpetrators were given impunity" (Falk 2008a, p. 42).

But what explains this great criminal act? According to Kramer and Kauzlarich (2011), the primary goal of the United States and its allies during World War II, of course, was to win the war. The "precision bombing" of military and war-related industrial targets, with its attendant collateral damage, and then the "area bombing" of enemy civilian populations to destroy their morale were two of the means to that ultimate end. But by 1944, military victory was all but assured in both the European and Pacific theatres. At this point, secondary war goals emerged to the forefront: ending the war as quickly and decisively as possible and, by accomplishing those objectives, saving the lives of Allied military personnel. The majority of American leaders came to believe that the accomplishment of these national goals necessitated a change from a sole reliance on precision bombing to an increasing use of terror bombing of enemy civilian populations, including the utilization of newly developed atomic weapons.

Some historians argue that the atomic bombings did help to shorten the war somewhat, and thus they did save some American lives (Walker 2004). After the war however, Henry Stimson, former Secretary of War, President Truman, and others who had participated in the decision to drop the atomic bombs created an elaborate mythology that the bombings of Hiroshima and Nagasaki had been the *only way* to end the war short of a costly invasion of Japan and that the lives of up to a *million soldiers* had been saved by shortening the war and avoiding the invasion. The evidence shows neither of these assertions was true. For example, Walker (2004, p. 109) demonstrates that the bomb "…was not necessary to prevent an invasion of Japan" and that it "…saved the lives of a relatively small but far from inconsequential number of Americans." Even so, according to the principle of noncombatant immunity found in both the "just war" moral tradition and the international laws of war, it is never permissible to target innocent civilians or noncombatants to accomplish war aims or save the lives of military combatants.

The narrative created by Stimson and Truman, however, served to legitimate the atomic bombings, and by extension all forms of terror bombing during the war, in the eyes of the American people. The goals of shortening the war and saving the lives of American boys were presented as self-evidently "good" and "just." Even before this, of course, the entire conflict, in a nationalistic fervor, was defined as the "Good War." Defeating what was, to most Americans, the obvious "evil" of Hitler and the fascists, whose own state crimes during the war were massive; exacting "just retribution" for the "sneak" attack on Pearl Harbor and other Japanese atrocities during the war; and defending the American ideals of freedom and democracy at home against the criminal aggression of the Axis powers were such clear moral goals that any means necessary to accomplish them came to be viewed as acceptable and legitimate to most American leaders and the public. The social construction of the goodness and morality of the war in general, and the specific objectives of shortening the bloody conflict and saving the lives of "our boys in uniform," overwhelmed and short circuited any attempt to critically evaluate the morality and legality of the terror bombing of the civilian populations of the "evil" enemy as a means to those legitimate ends (Kramer and Kauzlarich 2011).

In addition, a number of historians have documented that the United States also shared with its adversaries certain other nationalistic and expansionist motives. The war "propelled the U.S. to a hegemonic position" that provided a unique opportunity for American leaders to pursue these imperial designs (Selden 2009, p. 91). Enhancing the economic power and geopolitical position of the American empire became central goals of US wartime policies, including the policy of terror bombing.

The United States has been an imperial project from its earliest years (Zinn 1980). The Second World War provided an opportunity for the USA to greatly expand its informal empire by confronting and defeating rival imperial powers and by creating new regional and global structures. A clash of imperial ambitions precipitated "The Day of Infamy" at Pearl Harbor. This clash necessitated the decisive defeat of Japan and the complete destruction of Japanese militarism and imperialism. According to some historians, the firebombing of Japanese cities and the use of the atomic bombs were viewed as important means to accomplish these goals and exact just retribution for wartime atrocities (Gerson 2007).

As the war progressed and it became clear that the United States would be able to exercise hegemonic power in the postwar era, American leaders began to plan for the construction of new global institutions that would greatly advance their imperial designs. As Zinn (1980, p. 414) notes: "Before the war was over, the administration was planning the outlines of the new international economic order, based on partnership between government and business." This new international economic order would enhance and expand the informal "Open Door" imperialism the United States had been practicing since the early years of the twentieth century.

Even as World War II was putting the United States into a position from which it could dominate the world, American political and military leaders recognized that the Soviet Union, their wartime ally, would be their chief rival in the postwar period. The contest for power and domination between the Soviet Union and the United States that would later be dubbed "The Cold War" was already underway before the "hot" war against fascism was over. American officials increasingly came to view Stalin and the Soviets as a threat to their postwar imperial designs both in Europe and East Asia. And the perception of this threat would be an important factor in the most momentous decision of the Second World War: the decision to drop the atomic bomb.

A number of historians have argued that the decision to use the atomic bomb was motivated more by political factors related to the perceived Soviet threat than by purely military factors. Alperovitz (1995) presents persuasive evidence that American leaders dropped atomic weapons on Hiroshima and Nagasaki in an effort to impress Stalin with the power of the bomb and to intimidate the Soviet Union in the coming postwar contest for domination. He argues that Japan was on the verge of surrender in the summer of 1945 and that Truman and his advisors were well aware that alternatives to using the bomb existed. Nevertheless, Alperovitz contends, for diplomatic considerations, US leaders decided to use this powerful new weapon on Japan. To gain political leverage over the Soviets in the postwar period, the United States carried out the terror bombing of two cities using atomic weapons.

Another political factor that played a role in the decision to drop the atomic bomb was the threat of Soviet expansionism in the Far East. Once the war in Europe was over, Stalin had pledged to enter the Pacific war by attacking Manchuria, driving the Japanese out of China, and perhaps becoming involved in a prospective invasion of the Japanese homelands. Alarmed by the prospect of Soviet territorial gains in East

Asia and a shared occupation of Japan, American leaders hoped to use the atomic bomb to end the war quickly before the Soviet Union could enter the fight and become a major player in the end game in the Pacific. Again, imperial rivalry rather than military necessity seemed to drive the decision to bomb civilian populations with a new weapon of mass destruction (Kramer and Kauzlarich 2011).

A second form of state crime related to nuclear weapons that the United States has repeatedly engaged in is the threat to use these weapons in a variety of conflict situations in order to gain geopolitical and/or military advantage. As Gerson (2007, pp. 1–2) has documented: "On at least 30 occasions since the atomic bombings of Hiroshima and Nagasaki, every US president has prepared and/or threatened to initiate nuclear war during international crises, confrontations, and war-primarily in the Third World." These threats, he argues, have been used to illegally expand, consolidate, and maintain the American empire.

On some occasions, the threats to use nuclear weapons were made in an attempt to coerce a Third World government into negotiating an end to a military conflict on terms more favorable to the USA than the current conventional warfare strategy could produce. President Eisenhower's threat to use a nuclear device during the "police action" in Korea in 1953 and President Nixon's "November ultimatum" to North Vietnam in 1969 during the Vietnam War are the two primary examples of this practice (Kauzlarich and Kramer 1998). Implicit nuclear threats, which some have called nuclear terrorism, have also been used since World War II in an attempt to impose a global American imperial order and to provide a "nuclear shield" or "umbrella of U.S. power" (Chomsky 1999) to cover the use of conventional forces and make them more meaningful instruments of military and political power. Numerous presidents have used the specter of nuclear weapons in an attempt to secure "vital interests" in the Middle East, particularly with regard to Iraq and Iran, and also in East Asia (Gerson 2007). Nuclear weapons in space and missile defense systems, viewed as offensive weapons, have also increasingly been brandished as part of the nuclear shield designed to advance US hegemony in a post-Cold War world (Chomsky 2003). Finally, the threat to use nuclear weapons to reciprocate for a first strike by the former Soviet Union (a guarantee of "mutual assured destruction") was the basis for "nuclear deterrence" policies during most of the years of the Cold War.

Some criminologists point out that all of these threats by the American nuclear state to use nuclear weapons, regardless of the circumstances, are unlawful (Kramer and Kauzlarich 2011). First of all, the ICJ explicitly ruled that in all but the most extreme circumstances, the threat to use nuclear weapons would be illegal. The legal reasoning here, as *The Lawyers' Committee on Nuclear Policy* (1990, p. 19) has pointed out, is that "If a given use of nuclear weapons is judged to be contrary to the humanitarian rules of armed conflict, then logically any threat of such use should be considered contrary to the humanitarian rules of armed conflict as well." In the view of several legal scholars, these threats, as violations of international law, constitute state crimes (Boyle 2002).

In *Crimes of the American Nuclear State*, Kauzlarich and Kramer (1998) extensively analyzed the threats to use nuclear weapons by the USA during the wars in Korea and Vietnam. In addition, Gerson (2007) and Chomsky (2003) have documented numerous other situations in which the US government has engaged in criminal acts of nuclear terrorism. In these studies, the important role of imperial motives for state crimes related to nuclear threats is stressed. And just as the bombing of civilians, with either conventional weapons or atomic bombs, became a form of normalized criminal behavior over the course of World War II (Kramer 2010), so too did the threat of using nuclear weapons become a normal and acceptable criminal act in the pursuit of empire during the Cold War and beyond.

The third form of state crime related to nuclear weapons that the United States (and all of the other nuclear states) is engaged in is the continued possession of these weapons. Since the end of World War II, the US government has maintained varying stockpiles of atomic and nuclear weapons. Today, despite the end of the Cold War and substantial negotiated reductions, there are still over 10,000 nuclear weapons in the American arsenal, out of the total of 27,000 such weapons existing among the worldwide nuclear club (Cirincione 2007).

Shortly after World War II, the United Nations General Assembly was determined to eliminate the production and possession of atomic weapons through arms control treaties. Given the Cold War and the technical problems of monitoring, inspection, and enforcement, this proved impossible to accomplish at the time. As the former UN weapons inspector Hans Blix (2008, p. 42) has admitted, “. . .we have not been able to achieve rules specifically banning the production, stockpiling and use of nuclear weapons.” Does the lack of a specific rule banning the possession of nuclear weapons mean that the stockpiling of these weapons is lawful?

Since both the use and the threat to use nuclear weapons are illegal under international law, and, as Falk argues, intrinsically criminal, would not the mere possession of weapons that cannot be legally used in any practical way also be illegal and criminal? One could certainly attempt to make that legal argument. However, there is another legal agreement that can be used to evaluate the lawfulness of the continued possession of nuclear weapons under international law. That agreement is the Nuclear Non-Proliferation Treaty (NPT) of 1968, the treaty that the ICJ used in its advisory opinion where it concluded that states have a duty to negotiate in good faith a complete nuclear disarmament.

Some criminologists have concluded that by failing to engage in good faith efforts to achieve complete disarmament and continuing to possess nuclear weapons, the USA and the other nuclear states are in violation of the NPT and engaged in a form of international lawlessness (Kramer and Kauzlarich 2011; Kramer and Bradshaw 2011). This lawlessness presents a host of nuclear problems and dangers that bear on the question of the criminality of continued possession. First, the fact that the USA maintains a stockpile of nuclear weapons provides American state officials with the opportunity to illegally threaten to use them or actually use them in a future conflict situation. Second, the manufacture and assembly of these weapons by corporations in the US Nuclear Weapons Production Complex led to the massive illegal radioactive contamination of both workers and the environment at numerous sites, a form of state-corporate crime (Kauzlarich and Kramer 1998). Third, keeping the world “awash in nuclear-weapons technology” raises “the terrifying specter of a terrorist group that acquires and uses a nuclear weapon” (Schell 2007, p. 6). Finally and perhaps most importantly, by failing to disarm and thus violating the NPT, the USA encourages the proliferation of nuclear weapons among other states, making the world a much more dangerous place (Kramer and Bradshaw 2011).

The continued possession of nuclear weapons in violation of the NPT, therefore, is not only a criminal omission and an abandonment of international law, but it also creates criminogenic environments both domestically and internationally that pose a variety of clear nuclear dangers. As Schell (2007, pp. 12–13) has pointed out, the nuclear dilemma is an “indivisible whole”; that means that “proliferation and possession cannot be considered in isolation from each other.” As the world drifts toward “nuclear anarchy,” he warns that “Not since the world’s second nuclear bomb was dropped on Nagasaki has history’s third use of a nuclear weapon seemed more likely” (p. 14).

Empire, Exceptionalism, and the Nuclear State

Kauzlarich and Kramer (1998) developed a theoretical narrative to explain state crimes related to nuclear weapons that draws on an integrated framework for the study of organizational crime. This theoretical schema, an effort at theory elaboration (Vaughn 2007), links three levels of analysis, macro, meso, and micro, with three catalysts for action: motivation (goals), opportunity (means), and formal social control (sanctions). The objective is to inventory and highlight the key factors that contribute to or restrain organizational deviance at each intersection of a catalyst for action and a level of analysis. The organization is viewed as the key unit of analysis, nested within an institutional and cultural environment, and engaged in social action through the decisions of individual actors who occupied key positions within the structure of the organization. According to this schema, organizational deviance is most likely to occur when pressures for organizational goal attainment intersect with attractive and available illegitimate organizational means in the absence or neutralization of effective formal social controls.

As noted above, Kramer and Kauzlarich (2011) drew on this framework to explain that the American addiction to nuclear weapons is generated by both structural and cultural factors. They show that as organizational goals shaped by the geopolitical and economic environment (US national security and imperialistic interests) were pursued with available illegitimate means (the threat to use nuclear weapons), certain ideological or cultural forces (American exceptionalism, interpretative denial) were drawn on or developed to support and "normalize" these pursuits and disable effective social controls.

As reviewed above, US state crimes related to nuclear weapons have generally been undertaken for reasons of empire. The United States is an empire and in the post-World War II era, its imperialism has been a strong determining structural factor in its propensity to commit a wide variety of state crimes (Kramer 2011). The "bomb" has always provided an "umbrella of U.S. power" (Chomsky 1999) that allows for the pursuit of imperial domination. As Ritter (2010, p. 377) observes: "There can be no meaningful reduction of the American nuclear arsenal so long as it is attached to a national security strategy based on global dominance."

These crimes of empire have almost always been covered by some form of "interpretive denial" (Cohen 2001) that justifies or rationalizes imperial actions that cause harm or could be labeled as criminal. US state crimes, particularly those related to nuclear weapons, have been interpretively denied by reference to a broad, historical, and cultural narrative referred to as "American exceptionalism" (Kramer and Michalowski 2011). American exceptionalism generally portrays the USA as a nation of exceptional virtue, a moral leader in the world with a unique historical mission to spread "universal" values such as freedom, democracy, equality, popular sovereignty, and increasingly global capitalism. This mythic cultural construction of exceptionalism "thoroughly informs US constructs of its identity" (Ryan 2007, p. 119). World War II, "The Good War," reinforced the mythic idealism at the heart of American exceptionalism. In the end, the fight against evil and the advancement of America's exceptional ideals justify any of the means, including violent means; it selected to accomplish national goals. In the wartime environment, conformity to these cultural mandates helped to make the bombing of civilians, even with the most horribly destructive weapon that humans had ever devised, a normal and acceptable act (Kramer and Kauzlarich 2011).

Thus, denial, rationalization, and the narrative of exceptionalism have led to the normalization of the imperial crimes of the nuclear state. And as long as these structural and cultural forces remain in place, it does not matter who occupies the White House. As long as the USA remains an empire and possesses an imperial mentality, it will continue to be addicted to the power symbolized by nuclear weapons and will continue to possess these deadly weapons in violation of the legal duty to disarm imposed by the NPT. Kramer and Bradshaw (2011) argue that there is no "hope" for "change" in US nuclear weapons policy unless nuclear abolitionists and peace movement organizations challenge empire, break through the denial and normalization related to these weapons, and empower the international community to achieve the strategic goal of a concrete and verifiable nuclear disarmament treaty. The work of the scholars reviewed here and a vigorous "public criminology of state crime" (Kramer 2010) can perhaps assist both the American people generally, and criminologists specifically, to think differently about the crimes of the nuclear state and help to enhance the nuclear disarmament movement.

Related Entries

- ▶ Political Crime
- ▶ Resistance to State Crime
- ▶ State Crime
- ▶ State-Corporate Crime
- ▶ Victims of State Crime

Recommended Reading and References

Agnew R (2011) Toward a unified criminology: integrating assumptions about crime, people, and society. New York University Press, New York

Alperovitz G (1995) The decision to use the atomic bomb and the architecture of an American myth. Knopf, New York

Blix H (2008) Why nuclear disarmament matters. MIT Press, Cambridge, MA

Boyle F (2002) The criminality of nuclear deterrence: could the U.S. war on terrorism go nuclear? Clarity Press, Atlanta

Chomsky N (1999) The umbrella of U.S. power: the universal declaration of human rights and the contradictions of U.S. policy. Seven Stories Press, New York

Chomsky N (2003) Hegemony or survival: America's quest for global dominance. Metropolitan Books, New York

Cirincione J (2007) Bomb scare: the history and future of nuclear weapons. Columbia University Press, New York

Cohen S (2001) States of denial: knowing about atrocities and suffering. Polity Press, Cambridge

Falk R (1965) The *Shimoda* case: a legal appraisal of the atomic attacks upon Hiroshima and Nagasaki. Am J Int Law 59:759–793

Falk R (2008a) Non-proliferation treaty illusions and international lawlessness. In: Falk R, Krieger D (eds) At the nuclear precipice: catastrophe or transformation? Palgrave Macmillan, New York, pp 39–47

Falk R (2008b) Nuclear weapons, war, and the discipline of international law. In: Falk R, Krieger D (eds) At the nuclear precipice: catastrophe or transformation? Palgrave Macmillan, New York, pp 225–233

Friedrichs D (1985) The nuclear arms issue and the field of criminal justice. Justice Prof 1:5–9

Friedrichs D (2010a) Trusted criminals: white collar crime in contemporary society, 4th edn. Wadsworth, Belmont

Friedrichs D (2010b) Toward a prospective criminology of state crime. In: Chambliss W, Michalowski R, Kramer R (eds) State crime in the global age. Willan, Cullompton, pp 67–80

Gerson J (2007) Empire and the bomb: how the U.S. uses nuclear weapons to dominate the world. Pluto Press, London

Grayling AC (2006) Among the dead cities: the history and moral legacy of the WW II bombing of civilians in Germany and Japan. Walker, New York

International Court of Justice (1996) The legality of the threat or use of nuclear weapons (Request for Advisory Opinion Submitted by the General Assembly of the United Nations). General List, No. 95, Advisory Opinion of 8 July 1996

Kauzlarich D, Kramer R (1998) Crimes of the American nuclear state: at home and abroad. Northeastern University Press, Boston

Kramer R (2010) From Guernica to Hiroshima to Baghdad: the normalization of the terror bombing of civilian populations. In: Chambliss W, Michalowski R, Kramer R (eds) State crime in the global age. Willan, Cullompton, pp 118–133

Kramer R (2011) Curbing state crime by challenging empire. In: DeKeseredy W, Dragiewicz M (eds) Handbook of critical criminology. Routledge, London, pp 442–453

Kramer R, Bradshaw E (2011) U.S. state crimes related to nuclear weapons: is there hope for change in the Obama administration? Int J Comp Appl Crim Justice 35:245–261

Kramer R, Kauzlarich D (1999) The opinion of the International Court of Justice on the use of nuclear weapons: implications for criminology. Contemp Justice Rev 2:395–413

Kramer R, Kauzlarich D (2011) Nuclear weapons, international law, and the normalization of state crime. In: Rothe D, Mullins C (eds) State crime: current perspectives. Rutgers University Press, New Brunswick, pp 68–93

Kramer R, Michalowski R (2011) Empire and exceptionalism: the Bush administration's criminal war against Iraq. In: Rothe D, Mullins C (eds) State crime: current perspectives. Rutgers University Press, New Brunswick, pp 94–121

Lawyers' Committee on Nuclear Policy (1990) Statement on the illegality of nuclear warfare, Revised edn. Lawyers' Committee on Nuclear Policy, New York

Ritter S (2010) Dangerous ground: America's Failed Arms Control Policy, From FDR to Obama. Nation Books, New York

Ryan D (2007) Frustrated Empire: U.S. foreign policy 9/11 to Iraq. Pluto Press, London

Schell J (2007) The seventh decade: the new shape of nuclear danger. Metropolitan Books, New York

Selden M (2009) A forgotten holocaust: U.S. bombing strategy, the destruction of Japanese cities and the American way of war from the Pacific war to Iraq. In: Tanaka Y, Young M (eds) Bombing civilians: a twentieth-century history. The New Press, New York, pp 77–96

Vaughn D (2007) Beyond macro- and micro-levels of analysis, organizations, and the cultural fix. In: Pontell H, Geis G (eds) International handbook of white-collar and corporate crime. Springer, New York, pp 3–24

Walker JS (2004) Prompt and utter destruction: Truman and the use of atomic bombs against Japan, Revised edn. University of North Carolina Press, Chapel Hill

Zinn H (1980) A people's history of the United States: 1492-present. HarperCollins, New York

O

Obscenity

▶ Pornography

Occupation Crime

▶ Women and White-Collar Crime

Occupational Culture

▶ Police Culture

Offender Change in Treatment

Caleb D. Lloyd and Ralph C. Serin
Department of Psychology, Carleton University,
Ottawa, ON, Canada

Synonyms

Motivation; Offender outcomes; Rehabilitation

Overview

This entry looks at evidence that offenders change while attending correctional treatment programs. Surprisingly few research studies can demonstrate *how* offenders successfully change within correctional treatment programs. The typical research design involves using questionnaires at the beginning of a program, then repeating the same questionnaires with the same offenders once they complete the program. Although there are hundreds of studies examining pre-post-treatment change in offenders, there are fewer than 25 studies that explicitly test whether intra-individual changes (i.e., changes within the same individual between pre- and post-program assessments on knowledge and attitudes) are related to recidivism outcome. Current research is limited when attempting to describe what makes programs effective, especially in terms of the specific constructs to be measured and what methods should be used (i.e., self-report questionnaires, behavioral ratings).

Context

This entry first reviews two areas of research that provide compelling reasons that change is possible in some offenders. Based on this evidence, it is important to attempt to understand change within an individual, and to determine what processes help individuals give up crime. Simply knowing that some offenders succeed after program completion is not enough; it is necessary to understand the factors and processes that contribute to their success.

Second, this entry describes the small body of studies that have examined what changes at the individual level when offenders are successful.

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

Specifically, these studies have examined individual offenders' changes on relevant constructs (e.g., antisocial cognitions, impulsivity, and anger) and the relationship between these changes and recidivism. These studies involve general offenders, substance misusing offenders, violent offenders, family violence offenders, and sexual offenders. When considering the many types of offenders in an area of limited research, it is challenging to make strong conclusions regarding which changes are most important within treatment programs. Still, across the studies reviewed below, changes in crime-supportive beliefs were consistently significantly related to reductions in recidivism, across a variety of types of offenders. In other words, offenders who changed their attitudes or perceptions about whether crime was a positive choice during program participation were most likely to stay crime-free after program completion. This finding suggests it is important to continue refining the knowledge regarding cognitive change that occurs during all types of correctional programming.

Finally, this entry offers some considerations about how to best move forward in developing an empirical basis for understanding offender change, including how to best conceptualize and measure the change process.

Key Issues for Offender Change in Treatment

Do offenders change? Can this change be measured as it occurs?

Paying Attention to Change has Crucial Practical Importance

There are many times in the correctional system that decisions must be made about individual offenders and indications of offender change are important for lessening of sanctions. For example, the timing of release may rest upon whether an offender made progress during incarceration, whether the offender appears ready to transition into the community, and whether the offender's release plan shows positive potential versus unfavorable complications. Daily decisions within the correctional system require staff to predict success for individuals, but current research tells us very little about individual-level change and how it may be related to improved offender outcomes.

Research on rehabilitation programs has focused on the fundamental questions, such as what programs or approaches to treatment are most effective, or which offender subgroups are most receptive to treatment. This research has typically focused on issues of program fidelity (Lowenkamp et al. 2006) and attending to risk and need principles (Smith et al. 2009). Programs with higher integrity and that target higher-risk offenders and criminogenic needs yield greater effect sizes.

Nonetheless, not all offenders benefit equally, even from "good" programs. Questions remain regarding how offenders change within programs, or how unique offender characteristics and unique program characteristics together impact outcome (e.g., see Burrowes and Needs 2009; Ward et al. 2004).

In other words, rehabilitation programs would be enhanced if greater attention was paid to the treatment journey rather than just the intended destination.

Which Offenders Change?

Two areas of research support further investigation into individual-level mechanisms of change. The first area is desistance research, which uses a variety of qualitative and quantitative methodologies and samples, but is unified in its examination of crime across the life course. Desistance research has typically examined repeat offenders who have committed acquisitive crimes. Second, evaluations of correctional treatment programs, which are more narrow and specific in scope compared to desistance studies, show that even short in-prison interventions can have lasting impact on post-release behavior. These bodies of research give reason to be confident that some offenders can change, under the right circumstances.

Crime Decreases as Offenders Age

First, criminal behavior is unstable across the life course. On average, offenders increase their criminal activity to a peak during late adolescence and early adulthood and then decrease their crime frequency (i.e., an upside-down U called the age-crime curve; Blumstein and Cohen 1987). Despite variation in age of onset and cessation, it is normative for criminal activity to drift downward throughout early adulthood. For example, in a birth cohort followed to the age of 32, self-reported antisocial behavior decreased across time, with 70 % of those involved in antisocial behavior exhibiting considerably low rates or desistance by the age of 26 (Odgers et al. 2008).

Offenders who do not desist from crime are a problematic minority responsible for a majority of crimes, with each individual committing crime at high rates. This group is known as the persisters (Moffitt et al. 2002). This means not all offenders follow the same trajectory of change. The high-rate offenders with low likelihood of desistance are characterized by early life deficits and difficulties that are not as prevalent in sporadic or short-term offenders (Moffitt et al. 2002). At the later stages of a criminal career, offenders who experience particular life events are more likely to desist from crime. These life changes typically occur concurrently with changes in the offenders' priorities, goals, commitments, and social circumstances. Some key life changes related to desistance include sustained marriage, long-term employment, and residential changes (Sampson and Laub 2005). It remains to be seen if these findings will be replicated with more contemporary samples.

Thus, research findings from a life course desistance perspective offer a macro-level understanding of the age-related changes that broadly influence the process of crime reduction within offenders. Some of this research additionally has assessed psychosocial constructs (e.g., self-control) multiple times through the life course (Kazemian et al. 2009). A less demanding and more typical approach has involved measuring life events as turning points (i.e., marital status, job attainment). When researchers have investigated the internal processes, they conclude that individual-specific features such as romantic relationship quality and engagement in employment are vital aspects of this process (Sampson and Laub 1990). A unique body of studies taking a qualitative approach provides rich descriptions or narratives of ex-offenders' perspectives on their processes of change (e.g., Maruna 2001). Together, the various approaches to life course desistance research support the conclusion that offenders entering adulthood who actively embrace major life changes begin to enact new patterns of behavior and adopt noncriminal identities.

Recidivism Rates Are Lower Among Groups of Treated Offenders

The conclusion that appropriate approaches to psychosocial treatment reduce crime is now supported by upward of 40 meta-analyses (McGuire 2004). Although treatment programs and evaluations vary in quality, research overall concludes that offenders exposed to rehabilitation programs have a lower likelihood of returning to crime, on average, compared to offenders in a no-treatment control group.

Three particular features determine whether treatment has a crime-reducing effect. Andrews and Bonta (2010) specified *who* has potential to change, *what* characteristics are required to change, and the type of environment *where* change can occur. These principles have been referred to as risk/need/responsivity model, which is a guiding principle in offender programming (Andrews and Bonta 2010). Specifically, treatment can have a positive impact on higher-risk offenders, and lower-risk offenders may actually do worse if exposed to treatment with higher-risk offenders, hence only higher-risk offenders should be targeted for rehabilitation. Targeting higher-risk offenders is the *who* aspect of effective treatment.

Second, treatment programs are effective only when they focus on changing personal characteristics (i.e., criminogenic needs) known to be associated with criminal behavior. The characteristics most related to criminal behavior include impulsive personality, negative emotion, pro-offending attitudes, association with criminal friends, poor family relationships, inconsistent employment, poor use of leisure time, and misuse of substances.

Targeting these crime-related characteristics is the *what* aspect of effective treatment.

Third, treatment programs where staff match the rehabilitation to the learning styles of the offenders (i.e., cognitive-behavioral approaches that consider offender readiness) are more effective than programs offered in an unstructured format where the learning goals are unspecified or poorly adapted to offenders. Offering well-structured, cognitive-behavioral programs is the *where* aspect of effective treatment.

Adherence to these core considerations results in the most effective treatment programs currently available (Smith et al. 2009). Beyond these arguably broad strokes, there are additional considerations that remain relatively unexamined, regarding how different offenders experience treatment and exhibit unique change processes. As a result, there is increasing interest in assessing readiness to change (Day et al. 2007) and more precise aspects of responsiveness to treatment (Ward et al. 2004) in order to augment program effectiveness.

An important limitation of this research is that it primarily examines groups (offenders in treatment vs. not in treatment) and group recidivism rates. It is still unclear from current research which particular offenders are engaging in treatment and receiving the most benefit from it. Although research indicates what characteristics should be targeted in programs, researchers rarely explicitly measure whether these characteristics have actually changed at the individual level, and confirmed these changes are related to recidivism outcome.

An Overview of Within-Individual Change

The research on crime desistance and correctional programming discussed above complement each other, each using a unique vantage point to demonstrate that offenders decrease their criminal activity. In addition, both research areas provide theoretical frameworks for understanding the broad context of *why* this change occurs (e.g., aging, new engagement in conventional living, and increased knowledge and skills development). With a few exceptions, neither area of research has employed consistent measurement of within-individual variables across time. Notably, more specific examination of the mechanisms underlying *how* change occurs has been lacking.

Serin, Lloyd, Helmus, Derkzen, and Luong (2013) recently reviewed the available academic and government corrections research to compile the constructs (e.g., changes in pro-criminal attitudes, impulsivity, hostility) that show the most promise as mechanisms of intra-individual change and markers of success. The study identified 378 studies measuring change, but only 49 tests of the relationship between change scores and recidivism outcome at the individual level. It is striking that so few studies have examined within-individual change. These reviewers found 12 cognitive programs, three substance misuse programs, and two violence reduction programs that explicitly tested the association between intrapersonal change and recidivism. Of the 49 available tests for these 17 programs, 26 effect sizes (53.1 %) were significant, ranging from small ($d = 0.06$) to large ($d = 1.48$). Across all domains, change variables significantly related to recidivism included known dynamic risk factors: antisocial attitudes, antisocial beliefs, antisocial associates, antisocial personality, social support, and substance misuse. Although these studies varied in terms of program dosage, fidelity, and follow-up times, which would influence effect sizes, it is nonetheless important to begin to examine if certain constructs, when measured pre- and post-programming, are related to post-program completion recidivism. Essentially this is the answer to which offenders change, if the change indices have demonstrated predictive validity. Other reviewers have similarly found few studies focusing on domestic violence and sex offenders.

General Offender Cognition

In these 12 programs noted above, offenders attended cognitive programs and self-reported (at least twice) regarding their beliefs about crime, whether others supported their efforts to stay crime-free, negative emotion, levels of stress, life problems, connection to criminal friends, and/or symptoms of psychological distress. Some of these measures were taken at the beginning and end of prison-based programs, whereas other programs took place after the offender was already released into the community. All of these studies are unique, however, because the researchers tested if within-person changes were related to recidivism, rather than only testing if the treatment group as a whole changed, on average. This is a more fine-tuned approach because one individual may change, whereas another individual may not change even though both attend the same program. For a full listing of the programs and measures, refer to Serin, Lloyd, Helmus, Derkzen, and Luong (2013).

Overall, within these studies, individual offenders who exhibited changes in their thinking about crime were also less likely to recidivate after release. Specifically, some of the largest effect sizes were observed for measures that assessed acceptance of responsibility for crime, rejection of criminal behavior as a suitable course of action, and disregard for the law. Individual offenders who adopted more prosocial, anticriminal ways of thinking during rehabilitation were more likely to stay crime-free.

Another set of findings highlight the importance of connection to others. In particular, offenders who reported a reduced feeling of connection to criminal friends were less likely to reoffend after release. Offenders who reported increased access to individuals who supported their decision to stay crime-free were also more likely to succeed. By contrast, offenders reporting increasing problems in their family environment were more likely to recidivate.

Finally, in one study (Brown et al. 2009), two large effects were observed for changes related to the reentry environment. Specifically, recidivism was related to increases in negative emotion (negative feeling about the world) and escalating problems with reintegration back into the community. Offenders who had increasing difficulty integrating back home after release and feeling increasingly negative emotions during their reentry process were more likely to re-offend. The opposite is also true, however, where offenders who were able to clear up practical problems post-release and feel less negative moods were more likely to succeed.

Together, these cumulative results demonstrate that those offenders who experience improving life circumstances involving connection to prosocial others are effectively on a path toward staying crime-free. Similarly, those who report perceiving less value in crime as a life choice also were more likely to desist. Notably, not all studies found similar effects of change when employing the same measures, reinforcing the view that program fidelity and other factors were not controlled. This suggests certain attitudes *sometimes* appear to drive the change process. Currently, there are too few studies to more fully understand the factors that might influence this differential effect. Since many cognitive factors are interrelated (i.e., bonds to prosocial others are likely to spark reductions in criminal attitudes Andrews and Bonta 2010), another possibility worth exploring is whether there are many paths to the same outcome. In other words, it may not be that one particular type of attitude must change for success, just that prosocial attitudes *of some sort* must develop within the offender.

Substance Misuse

Few (only three) studies explicitly examined change in variables among offenders with substance misuse problems. Each of these studies examined different change factors, and each found change was directly related to reduction in recidivism. Among these, the largest effect size demonstrated that offenders who reduced their substance use during the time period shortly after release were more likely to stay crime-free in the community (Brown et al. 2009). Small effect sizes were found in the remaining two studies, demonstrating that reductions in psychiatric symptomology (e.g., anxiety, hostility, paranoia; Sullivan et al. 2008) and increases in perceived consequences of drug use were related to reductions in recidivism (Millson et al. 1995).

Further, the first study suggests individuals who become sober are less likely to commit further crime. In the latter two studies, it is difficult to draw conclusions regarding the importance of alleviation of distress symptoms and changes in drug beliefs because individuals' substance use behaviors as they participated in treatment were not explicitly measured. It would have been a stronger approach to simultaneously measure changes in substance use, rather than only measuring change in psychological variables. Still, these studies reach a similar conclusion to the studies examining cognitive programs, namely, that reductions in negative emotion (hostility) and beliefs about crime (consequences of drug use) are mechanisms that can drive reductions in criminal activity.

Violence

Two findings emerged from two studies of change among nonsexual violent offenders (e.g., assault, armed robbery, murder). These studies found medium-to-large effect sizes for reductions in negative emotion (anger, specifically; Polaschek and Dixon 2001) and self-reported increases in personal insight about life problems (e.g., offense chain; Dowden et al. 1999). Reductions in anger were associated with a lower likelihood of nonviolent recidivism. Variables reflecting that offenders learned something about their risk factors during treatment were associated with lower likelihood of violent recidivism. However, given the limited amount of evidence available, these findings need replication.

In one attempt to replicate these findings, a recent study demonstrated that violent offenders who became actively invested in changing their problem behavior during a brief motivational intervention were less likely to commit a further offence (violent or nonviolent reconvictions; Anstiss et al. 2011). This supports the finding regarding knowledge of their offense chain, and suggests violent offenders who demonstrate more motivation to actively participate in rehabilitation are more likely to change.

From these three studies, it appears reducing anger among violent offenders can reduce recidivism, but a key mechanism of change also appears to involve increasing motivation to tackle life problems in a prosocial way. Thus, offenders who seem to connect with treatment and "get it" in terms of increased personal insight and drive to change their old patterns are less likely to reoffend. In particular, Anstiss and colleagues (2011) demonstrated that reductions in recidivism occurred as a result of a treatment focused on enhancing motivation (i.e., increasing readiness to change). In other words, simply addressing motivation without attention to other skills or variables appeared to be enough to influence violent offenders toward successful change, although again this requires replication.

Family Violence

In a review of treatment approaches, typologies, and change mechanisms in the area of family violence, Scott (2004) similarly concluded research is lacking about change processes. However, her review suggested that three variables show some evidence as potential change mechanisms and deserve further consideration: anger, psychiatric symptomology, and substance use.

Changes in psychological distress and anger were recently investigated among men court-ordered to attend treatment for violence against their partners (Alexander and Morris 2008). Results showed that some men decreased their depression and anxiety during treatment, as well as increased control over their anger, but only if they reported being ready to make behavior changes at the beginning of treatment. Only a small number of romantic partners were contacted for follow-up (15 out of 210), but among this select group, men who changed during treatment (i.e., in terms of anger, depression) were less likely to inflict injury on their partner during the 6 months following treatment, compared to men who did not change. However, outcomes were only examined in terms of pretreatment scores, ignoring post-treatment and change scores. This is unfortunate given that other studies of high quality methodology are not available in the family violence treatment domain.

Research regarding change among perpetrators of family violence is particularly limited, but it is noteworthy that the limited findings are similar to those within other offender domains. Further research should attempt to replicate these results suggesting changes in negative emotion (anger) and psychological distress are important for reductions in family violence.

Sexual Offenders

Beggs (2010) recently reviewed treatment change among sexual offenders and found that recidivism was related to within-individual changes in assertiveness, empathy, anger, beliefs specifically related to offending against children (including sexual interest in children, loneliness), and general risk factors (including criminality, substance use, and treatment responsivity). In addition, Beggs and Grace (2011) demonstrated that reductions in sexual deviance, anger, and various cognitive and psychiatric symptom variables (including loneliness, intimacy, depression, and anxiety) were related to a lower likelihood of sexual recidivism. Notably, these treatment changes predicted reduced recidivism while accounting for static risk scores. By contrast, in a large sample of sexual offenders attending community treatment (Barnett et al. 2013), change was not explicitly associated with sexual and violent recidivism, but those offenders who remained in a low functioning range at the end of treatment on particular variables (including self-esteem, loneliness, assertiveness) were more likely to recidivate.

In a large sample of sexual offenders attending prison-based treatment (Wakeling et al. 2013), post-treatment scores appeared to be more important than change scores in predicting recidivism. Those who were displaying functional (i.e., higher) scores at the end of treatment reoffended the least. It is important to note, however, that some individuals displaying functional scores following treatment may not represent offenders who changed; some offenders were in the functional range at the beginning of treatment. When change scores were explicitly examined, changes in sexual interest and impulsivity were associated with lower likelihood of sexual and violent recidivism. However, this effect was not significant after accounting for static risk scores.

Summary

It is striking that across all types of offender programming reviewed, similar variables appeared to be important mechanisms of change across the different subtypes of offenders. In particular, positive in-program changes in negative emotion (e.g., anger or hostility) and offence-related cognitions (e.g., various pro-criminal beliefs) were most consistently related to post-program reductions in recidivism. Given the limited breadth of current investigations, it is difficult to discern if these variables are definitively the key drivers of the crime reduction process. At the same time, inconsistencies were observed across studies, as change in the same measure sometimes successfully, sometimes unsuccessfully predicted recidivism at the individual level.

Nonetheless, the variables identified in this entry deserve greater attention. These include changes in negative emotion (hostility, anger, depression, and anxiety), beliefs about crime (interest in and positive evaluations of criminal activities), symptoms of distress, substance use, social support, and engagement in treatment. At this time, there is little consensus about the best questionnaires for assessing each of these variables, but research is likely to benefit from subtype-specific measurement. For example, the core beliefs that support violence are different from the beliefs that support illegal drug use or sexual offending.

Conceptual and Methodological Issues About Measuring Individual Change

Of course, paying attention to process does not come without financial and operational costs, because this involves incorporating a more holistic conceptualization of change, and more sophisticated research methodologies. Some of these considerations are highlighted below. First, from a conceptual viewpoint, a complete understanding of change requires looking more inclusively beyond the treatment environment itself. For example, it is important to consider not only whether change has occurred, but also to what degree this change occurs (i.e., mild improvement or complete "remission"?). It is also important to understand whether change is occurring in one progressive, linear stretch, or with piecemeal transitions, bursts, and starts. Second, each of these three conceptual concerns has implications for methodology. Researchers should measure offenders more than twice, measure events that occur before and after the treatment program, and explicitly test whether change is real or simply due to measurement error. Each of these conceptual and methodological considerations are described below.

Complementary Approaches to the Same Broad Process

Correctional treatment provides the most concrete framework from which to investigate intra-individual change. Yet, treatment exposure is arguably only one part of an individual's journey of change. By merging macro-(desistance) and micro-level (treatment programs) investigations of change process, it is likely that reduced risk for recidivism involves both lifestyle changes *and* specific learning moments within rehabilitation programs. These processes are complementary and each contributes a piece to a complete change process. To date, these literatures exist in isolation, with limited effort to recognize that *both* can contribute to offender change.

Thus, depending on the study, it may be difficult to find treatment effects when other processes are at play, especially given that aging offenders experience additional forces outside of treatment that decrease their likelihood of future crime. Measuring changes in crime is particularly difficult because crime is a relatively rare event in an individual's pattern of behavior and is not subject to small, incremental changes across time (unlike a behavior such as reduction in smoking cigarettes). These challenges suggest it would be productive to integrate research of broad life changes with research on treatment effects. A more holistic understanding of treatment change would attempt to link an offender's pre-post changes to his or her past and current life circumstances.

From a desistance perspective, treatment should be considered just one component of a more extensive change process that is simultaneously occurring in the offender's life (see also DiClemente 2007). Being too treatment-centric limits researchers from discovering what life circumstances outside of the treatment program may have assisted clients to change (DiClemente 2007).

Recognizing change processes occurring outside of treatment does not discount the value of rehabilitation programs. Scores of research have confirmed that correctional programming can be an important turning point in certain offenders' life courses. In addition, some of the changes reviewed throughout this entry were the result of very brief, cost-effective programs, suggesting the need to investigate change both within and without the treatment environment.

To think about change processes in a holistic, interactive way requires extending current measurement strategies beyond only pre-post-treatment measurement. Additional thought needs to be put into integrating what is known about change outside of rehabilitation into questionnaires given during treatment studies. This will require expanding the range of variables examined and the range of life experiences assessed (i.e., how did a treatment episode impact future desistance during follow-up, and how did extra-treatment experiences affect the impact of the program for that individual?).

Reliable Change and Clinical Significance

To assess reliable change and clinical significance, researchers need measurement that is both precise (good psychometric properties) and meaningful (such that scores above the threshold are associated with reduced recidivism). Conceptually, this requires building knowledge about how to best define success at the end of a program, and methodologically, this requires researching quantitative cut-off scores. (See also Nunes et al. 2011, for an application of clinical significance to a study of sexual offenders).

It is difficult to know whether pre-post-treatment change is "real" or simply the result of measurement error. Jacobson and Truax (1991) argued that research should both determine which clients have changed *and* whether the change is sizable enough to have real-world impact on the individual's life. In the context of this entry, that means predicting desistance from crime. Jacobson and Truax's (1991) reliable change index is a formula for calculating whether change during therapy is greater than what would be expected due to measurement error. In other words, if offenders typically score within a plus-or-minus five-point range on a scale across short time periods, an offender who improved by four points across treatment would not be predicted to desist, as change was not reliable. Thus, the reliable change index takes this measurement error into account, identifying only those whose scores represent "true" change.

Similarly, they recommend establishing a meaningful threshold score, or a cut-off that represents improvements strong enough to impact the individual's outcome. Some individuals may change in treatment, but is this change enough to reduce their chances of recidivism? If so, the change is meaningful, and can be considered clinically significant change.

Identifying the Precise Timing of Change

Although pre-post-treatment measurement is the most typical approach in correctional psychology, in other fields of psychological therapy, researchers measure participants at every treatment session. This allows researchers to pinpoint exactly when change occurred. Such an approach requires additional cost, time, and effort. If researchers seek to describe change by employing only two assessment periods (pre-post-treatment design), they are at a disadvantage because they can only describe linear change (i.e., do offenders improve or deteriorate; Singer and Willett 2003). At least three assessment periods, but preferably more, are required to investigate non-gradual, curving processes of change (patterns of improvement and decline within the same individual). Such nonlinear change is described in desisters' narratives (Maruna 2001) and is consistent with Stages of Change research (Prochaska et al. 1992). Adding more measurement periods allows researchers to describe the timing, patterns, maintenance, and potential causes of change.

For example, among male adolescents in a secure psychiatric treatment facility, change in traits over a 6-month period showed different patterns of development (Caldwell et al. 2012). Some traits showed little initial improvement during the first 3 months, but then substantial improvement over the latter 3 months, whereas other traits changed more steadily. The timing of change was only uncovered by including a mid-treatment assessment.

A small group of high-risk burglary offenders increased their readiness to change from the beginning to end of a brief five-session motivational interviewing treatment (Austin et al. 2011). For a smaller group of participants ($n = 12$), self-report ratings were taken a third time (3–12 months later), and results indicated that change had been maintained following the brief treatment. Without the third assessment, the timing and sustainability of change would remain unknown.

Finally, Barrett and colleagues (2003) were able to describe complex patterns of change by employing four time periods, revealing that some variables evidenced curvilinear change. Variables representing different facets of motivation and treatment engagement were coded from files for 101 male sexual offenders who underwent two treatment programs, one in the institution, one in the community. For two variables, improvements were gained early and maintained into the community, whereas for three variables, gains made in the institution were lost after release into the community (an upside-down U-shaped distribution). Evidence also suggested that only change maintained into the community was related to reduced recidivism. This research also points to the possibility that the institutional program, but not the community program, was responsible for initially sparking the change.

Additional Considerations and Final Thoughts

In conclusion, although there are financial costs and operational challenges associated with investigating change in treatment outcome studies, spending these funds is also a strategic way to ensure that knowledge will be gained even from programs that are ultimately considered a nonsuccess. Limiting understanding only to whether programs "work" or not leads to blanket conclusions about discarding or keeping programs, but attending to process uncovers *why* programs positively or negatively influenced offenders. The research on correctional program fidelity, organizational readiness, and implementation (Lowenkamp et al. 2006) must also be incorporated into such an understanding, but with effect sizes less than 0.30 for even the best correctional programs, not all offenders benefit. Hence, knowing why a program did not work and *how* it failed to influence some offenders is a valuable way to recoup some costs associated with an unsuccessful program.

Despite the singular emphasis on investigating change processes in this entry, it is important to note that fundamental considerations when researching treatment programs are still critically important, no matter how sophisticated the measurement of change within the study. Treatment change should not be studied without simultaneous attention to follow-up outcomes and control groups. A program cannot be labeled as successful without investigating if recidivism is actually reduced, and it cannot be concluded that the program was responsible for offenders' success without a control group. Also, program dropout has implications for understanding the process of treatment. In particular, when multiple assessment periods are examined across the tenure of treatment, those who drop out early will be necessarily excluded from analyses. A measurement strategy that attempts to account for and explain dropout is part of understanding treatment processes.

Although attending to change within treatment is a challenging venture, it is a fruitful one. The causes, driving mechanisms, and conditions required for sustainable change are likely to be difficult to pin down, but a better understanding of these complex treatment dynamics will lead to enhanced correctional practice. Ultimately, advancing knowledge about process will allow building an appropriately complex model of offender change.

Related Entries

▶ Age-Crime Curve
▶ Career Criminals and Criminological Theory
▶ Correctional Education of Programs
▶ Correctional Workers in the Organizational Change Process
▶ Criminal Careers
▶ Desistance from Crime
▶ Examining the Effectiveness of Correctional Interventions
▶ Group-Based Trajectory Models and Developmental Change
▶ Rehabilitation

Recommended Reading and References

Alexander PC, Morris E (2008) Stages of change in batterers and their response to treatment. Violence Vict 23:476–492. doi:10.1891/0886-6708.23.4.47

Andrews DA, Bonta J (2010) The psychology of criminal conduct, 5th edn. LexisNexis, New Providence

Anstiss B, Polaschek DLL, Wilson M (2011) A brief motivational interviewing intervention with prisoners: when you lead a horse to water, can it drink for itself? Psychol Crime Law. doi:10.1080/1068316090352432

Austin KP, Williams MWM, Kilgour G (2011) The effectiveness of motivational interviewing with offenders: an outcome evaluation. N Z J Psychol 40:55–67

Barnett GD, Wakeling H, Mandeville-Norden R, Rakestrow J (2013) Does change in psychometric test scores tell us anything about risk of reconviction in sexual offenders? Psychol Crime Law 19:85–110. doi:10.1080/1068316X.2011.607820

Barrett M, Wilson RJ, Long C (2003) Measuring motivation to change in sexual offenders from institutional intake to community treatment. Sex Abus J Res Treat 15:269–283. doi:10.1177/107906320301500404

Beggs SM (2010) Within-treatment outcome among sexual offenders: a review. Aggress Violent Behav 15:369–379. doi:10.1016/j.avb.2010.06.005

Beggs SM, Grace RC (2011) Treatment gain for sexual offenders against children predicts reduced recidivism: a comparative validity study. J Consult Clin Psychol 79:182–92. doi:10.1037/a0022900

Blumstein A, Cohen J (1987) Characterizing criminal careers. Science 237:985–991

Brown SL, St. Amand MD, Zamble E (2009) The dynamic prediction of criminal recidivism: a three-wave prospective study. Law Hum Behav 33:25–45. doi:10.1007/s10979-008-9139-7

Burrowes N, Needs A (2009) Time to contemplate change? A framework for assessing readiness to change with offenders. Aggress Violent Behav 14:39–49. doi:10.1016/j.avb.2008.08.00

Caldwell MF, McCormick D, Wolfe J, Umstead D (2012) Treatment-related changes in psychopathy features and behavior in adolescent offenders. Crim Justice Behav 39:144–155. doi:10.1177/0093854811429542

Day A, Howells K, Casey S, Ward T, Birgden A (2007) Treatment readiness: an overview of Australasian work. Issues Forensic Psychol 7:21–25

DiClemente CC (2007) Mechanisms, determinants and processes of change in the modification of drinking behavior. Alcohol Clin Exp Res 31:13s–20s. doi:10.1111/j.1530-0277.2007.00489.x

Dowden C, Blanchette K, Serin RC (1999) Anger management programming for federal male inmate: an effective intervention. Research Report No. 82. Retrieved from Correctional Service of Canada website http://www.csc-scc.gc.ca/text/rsrch/reports/r106/r106_e.pdf

Jacobson NS, Truax P (1991) Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 59:12–9

Kazemian L, Farrington DP, Le Blanc M (2009) Can we make accurate long-term predictions about patterns of de-escalation in offending behavior? J Youth Adolesc 38:384–400. doi:10.1007/s10964-008-9338-z

Lowenkamp CT, Latessa EL, Smith P (2006) Does correctional program quality really matter? The impact of adhering to the principles of effective intervention. Criminol Public Policy 5:575–594. doi:10.1111/j.1745-9133.2006.00388.x

Maruna S (2001) Making good: how ex-convicts reform and rebuild their lives. American Psychological Association, Washington, DC. doi:10.1037/10430-00

McGuire J (2004) Understanding psychology and crime: perspectives on theory and action. Open University Press, New York

Millson WA, Weekes JR, Lightfoot LO (1995) The Offender Substance Abuse Pre-release Program: analysis of intermediate and post-release outcomes. Research Report No. 40. Retrieved from Correctional Services of Canada website http://www.csc-scc.gc.ca/text/rsrch/reports/r40/r40e_e.pdf

Moffitt TE, Caspi A, Harrington H, Milne BJ (2002) Males on the life-course persistent and adolescence-limited antisocial pathways: follow-up at age 26. Dev Psychopathol 14:179–206

Nunes KL, Babchishin KM, Cortoni F (2011) Measuring treatment change in sex offenders: clinical and statistical significance. Crim Justice Behav 38:157–173. doi:10.1177/0093854810391054

Odgers CL, Moffitt TE, Broadbent JM, Dickson N, Hancox RJ, Harrington H, Caspi A (2008) Female and male antisocial trajectories: from childhood origins to adult outcomes. Dev Psychopathol 20:673–716. doi:10.1017/S095457940800033

Polaschek DLL, Dixon BG (2001) The violence prevention project: the development and evaluation of a treatment programme for violent offenders. Psychol Crime Law 7:1–23. doi:10.1080/10683160108401779

Prochaska JO, DiClemente CC, Norcross JC (1992) In search of how people change: applications to addictive behaviors. Am Psychol 47:1102–1114. doi:10.1037/0003-066X.47.9.1102

Sampson RJ, Laub JH (1990) Crime and deviance over the life course: the salience of adult social bonds. Am Sociol Rev 55:609–627. doi:10.2307/2095859

Sampson RJ, Laub JH (2005) A life-course view of the development of crime. Ann Am Acad Pol Soc Sci 602:12–45. doi:10.1177/000271620528007

Scott KL (2004) Predictors of change among male batterers: application of theories and review of empirical findings. Trauma Violence Abuse 5:260–284. doi:10.1177/1524838003264339

Serin RC, Lloyd CD, Helmus L, Derkzen D, Luong D (2013) Does intra-individual change predict offender recidivism? Searching for the Holy Grail in assessing offender change. Aggress Violent Behav 18:32–53

Singer JD, Willett JB (2003) Applied longitudinal data analysis: modeling change and event occurrence. Oxford University Press, New York

Smith P, Gendreau P, Swartz K (2009) Validating the principles of effective intervention: a systematic review of the contributions of meta-analysis in the field of corrections. Vict Offenders 4:148–169. doi:10.1080/15564880802612581

Sullivan CJ, Sacks S, McKendrick K, Banks S, Sacks JY, Stommel J (2008) Modified therapeutic community treatment for offenders with co-occurring disorders. J Offender Rehabil 45:227–247. doi:10.1300/J076v45n01_1

Wakeling HC, Beech AR, Freemantle N (2013) Investigating treatment change and its relationship to recidivism in a sample of 3773 sex offenders in the UK. Psychol Crime Law 19:233–252. doi:10.1080/1068316X.2011.626413.

Ward T, Day A, Howells K, Birgden A (2004) The multifactor offender readiness model. Aggress Violent Behav 9:645–673. doi:10.1016/j.avb.2003.08.001

Offender Decision Making and Behavioral Economics

Greg Pogarsky
School of Criminal Justice, University at Albany,
State University of New York, Albany, NY, USA

Overview

This entry discusses potential applications of Behavioral Economic principles in Criminology. First it highlights aspects of the deterrence and economic approaches to offender decision making. Next, it introduces concepts and seminal findings from Behavioral Economics research outside of Criminology. Finally, potential areas of interesection between Behavioral Economics and Crime decision making are addressed.

Fundamentals

In Criminology, offender decision-making typically comes within the sub-field of deterrence. This approach shares much with economic models of decision-making and behavior (Nagin, forthcoming). First and foremost, under both perspectives, crime is sometimes a voluntary choice (Becker 1968; Bushway and Reuter 2008; McCarthy 2002). Offenders are not fully impelled toward crime by external circumstances, such as poor upbringing, financial disadvantage, or psychological infirmity, though such factors can contribute (sometimes greatly) to the likelihood of offending. Incentives relating to the potential consequences from crime impact offending decisions. Someone might refrain from crime to avoid an official sanction, such as incarceration or probation. Or, he or she might commit a crime believing the likelihood of detection is acceptably low. Each of these possibilities also pertains to various potential non-legal punishments, such as social stigma or compromised employment prospects (Grasmick and Bursik 1990; Pogarsky 2002).

Both the deterrence and economic perspectives recognize the paramountcy of perceptions. Risks and/or benefits only influence offending decisions insofar as they are perceived by potential criminal actors (Andeneas 1974; Apel and Nagin 2010). Such perceptions should reflect reality (though perhaps imperfectly) in at least two ways. First, an individual's past experiences committing crime and either experiencing or avoiding punishment should influence their sanction risk perceptions in what economic theory terms a Bayesian updating process (Anwar and Loughran 2011). As a corollary, when an individual becomes aware that someone else has committed a crime and either experienced or avoided punishment, this too should affect that individual's perceptions of the potential consequences from crime (Stafford and Warr 1993). Second, the sanctioning climate for an individual, reflected for example by law enforcement presence or arrest rates, also influences sanction risk perceptions (but compare Apel, Pogarsky, and Bates, 2009; Kleck et al. 2005; Lochner 2007). Finally, under both perspectives an individual's perceptions of the potential consequences to them from crime influence the likelihood that individual will offend (Matsueda et al. 2006).

Economic models have influenced scholarship on a range of human activities, including health, finance, public policy, and even reproduction and marriage (Becker et al. 1977). And there is little doubt about the historic and continuing influence of economic perspectives on criminological thought. Beginning in the 1970s, however, scholars have documented either systematic departures from economic predictions, or behavioral regularities that could not be accounted for by economic principles. These findings, which are not routinely applied in criminology, evolved into contemporary Behavioral Economics, the topic of this entry (e.g. Kahneman and Tversky 1979).

Some early Behavioral Economic findings involve, for example, the role of risk in decision-making. Consider the "lives saved-lives lost problem" of Tversky and Kahneman (1981: 453):

Imagine that the USA is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:

- If Program A is adopted, 200 people will be saved.
- If Program B is adopted, there is a 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved.

On average, each program is expected to save the same number of people: 1/3*600 + 2/3*0 = 200. But option B is considerably riskier, since one-third of the time, no one will be saved. In fact 72 % of individuals tend to choose option A, the less risky option that will save 200 people for sure. Next consider these reworded options:

- If Program C is adopted, 400 people will die.
- If Program D is adopted, there is one-third probability that nobody will die, and two-third probability that 600 people will die.

Notice that options C and D are functionally identical to options A and B. However, C and D are reframed in terms of how many will die (a loss) rather than how many will live (a gain). The second set of options produces a "preference reversal." That is, 78 % tend to prefer option D, the riskier option, over option C. Thus, individuals are far more comfortable taking a risk to avoid a loss than they are taking the nominally equivalent risk for gains. This finding is a profound challenge to economic theory, under which preference for risk should be invariant to the framing of choices.

Another seminal Behavioral Economic finding involves the "Endowment Effect," which was illustrated by Kahneman et al. (1990). Half of a group of students were randomly given a coffee mug from the school bookstore. Individuals with a mug were asked to report their minimum selling price, whereas the remaining participants (who could see and touch the mugs) were asked to report their maximum buying price. The mean buying price was $2.25 while the mean selling price was $4.75. In economic terms, the researchers elicited functionally equivalent information from both groups: the dollar value of a mug. Yet those who received a mug only moments earlier valued it twice as much as those who did not. As before, under economic theory, the value of the mug should not depend on the manner of valuation. But it does.

Finally consider Ellsburg's (1961) "two color" problem which explored ambiguity and choice. Respondents select one of two urns from which they will draw one ball. If the ball is red, they win a cash prize. Urn 1 contains 50 red balls and 50 black balls. Urn 2 also contains 100 red or black balls, but in unknown proportions. Although the probability of selecting a red ball and winning is identical across urns (in urn 2, the number of red balls is uniformly distributed from 0 to 100, hence, $E[x] = 50$), individuals routinely prefer Urn 1. This finding has been termed "ambiguity aversion," which cannot be accounted for under the straightforward application of economic principles.

These and other findings evolved into a body of scholarship known as judgment and decision-making research or Behavioral Economics. Among the core themes of such research is that individuals often deviate from the norms of rational choice economics *in predictable ways*. Thaler (1996: 12) has described Behavioral Economics as "economics with a higher R^2." For his pioneering scholarship on Behavioral Economics, Daniel Kahneman, a psychologist, received the 2002 Nobel Prize in Economic Science.

Just as Behavioral Economics has provided insights into other forms of human decision-making, it has great potential to elaborate crime decision-making. For example, Pogarsky and Piquero (2003) found some evidence for a "resetting effect" to explain why, contrary to deterrence and economic theory, individuals might *reduce* their estimate of the future certainty of punishment after getting caught and punished. This explanation clearly foregoes the "rational actor" assumption of deterrence and economic theory and instead relies on the Behavioral Economic principle of a gambler's fallacy in the interpretation of chance events (Gilovich 1983). Card players sometimes increase their bets after losing several consecutive hands because they feel they are due to win. Lottery participants decrease the amount wagered on a particular combination of numbers after that sequence has "hit" (Clotfelter and Cook 1993). The gambler's fallacy stems from a desire for and resultant belief in consistency. Laboring under this expectation, individuals impute an interdependence among chance events where none exists. Most offenders believe getting caught is a somewhat rare occurrence. Thus when they do get caught, under the resetting explanation, they may reduce their future estimated certainty of punishment believing they would have to be exceedingly unlucky to be caught again. If this occurs in some individuals, it is a profound challenge to both the deterrence and economic perspectives on offending decisions under which individuals should elevate their perception of the risk of punishment after getting caught.

Earlier, this entry addressed Behavioral Economic research relating to ambiguity and choice. Loughran et al. (2011) examined the interrelationship between ambiguity, sanction risks, and offending in a survey to college students with several hypothetical choice problems and data from the Pathways to Desistance study, a longitudinal investigation of serious adolescent offenders transitioning from adolescence to young adulthood. There was some evidence that individuals are "ambiguity-averse" for decisions involving losses such as criminal punishments. This means that a more ambiguous perceived certainty of punishment is a greater deterrent of some crimes than a nominally equivalent but less ambiguous one.

This finding suggests a modification of traditional deterrence and economic theory which, by omission, has presumed the irrelevance of ambiguity in perceived risk. Applying behavioral economic insights, then, generates potential policy implications not discernible from the straightforward application of deterrence and economic principles to the crime decision. In general the risk of detection for less serious crimes, such as breaking-in, stealing, theft and vandalism is quite low (e.g., Lochner 2007). Sherman (1990: 7) argued that law enforcement policies ought to exploit ambiguity aversion for these types of crimes by offering offenders "low certainty about whether the risk of punishment is high or low at any given time and place". That is, police could rotate their enforcement across both offenses and places so that the risk of punishment is far more unpredictable than it has been in the past. With the same amount of resources, an alteration of police policy by manipulating the ambiguity of the certainty of arrest – would enable them to enhance the deterrent effect they have. Thus it might be worthwhile for law enforcement to exploit any vagueness in perceptions of the risk of punishment.

This entry was intended to expose criminologists to the historical relationship between Deterrence, Economics, and Behavioral Economics. Already the few applications of Behavioral Economic principles to offender decision making have furthered our understanding of crime. These applications have helped identify policy issues not raised by the traditional approach to studying crime decisions, which is largely based on deterrence and economic principles. Further application of Behavioral Economic principles in studying crime decisions, thus, appears likely to bear even more fruit.

Recommended Reading and References

Andenaes J (1974) Punishment and deterrence. University of Michigan Press, Ann Arbor

Anwar S, Loughran TA (2011) Testing a Bayesian learning theory of deterrence among serious juvenile offenders. Criminology 49:667–698

Apel R, Nagin DS (2010) Deterrence. In: Wilson JQ, Petersilia J (eds) Crime, 4th edn. Oxford University Press, Oxford. Also, appearing In: Tonry M (ed) Handbook on crime and criminal justice. Oxford University Press, Oxford

Apel R, Pogarsky G, Bates L (2009) The sanctions-perceptions link in a model of school-based deterrence. J Quant Criminol 25:201–226

Becker G (1968) Crime and punishment: an economic approach. J Pol Econ 76:169–217

Becker GS, Landes EM, Michael RT (1977) An economic analysis of marital instability. J Pol Econ 85:1141–1188

Bushway S, Reuter P (2008) Economists' contribution to the study of crime and the criminal justice system. Crime Justice: Rev Res 37:389–451

Clotfelter CT, Cook PJ (1993) The gambler's fallacy in lottery play. Manage Sci 39:1521–1528

Ellsberg D (1961) Risk, ambiguity, and the savage axioms. Quart J Econ 75:643–669

Gilovich T (1983) Biased evaluation and persistence in gambling. J Pers Soc Psychol 44:1110–1126

Grasmick HG, Bursik RJ Jr (1990) Conscience, significant others, and rational choice: extending the deterrence model. Law Soc Rev 24:837–861

Kahneman D, Tversky A (1979) Prospect theory: an analysis of decisions under risk. Econometrica 47:313–327

Kahneman D, Knetsch J, Thaler RH (1990) Experimental tests of the endowment effect and the coarse theorem. J Pol Econ 98:1325–1348

Kleck G, Sever B, Li S, Gertz M (2005) The missing link in general deterrence research. Criminology 43:623–659

Lochner L (2007) Individual perceptions of the criminal justice system. Am Econ Rev 97:444–460

Loughran TA, Paternoster R, Piquero AR, Pogarsky G (2011) On ambiguity in perceptions of risk:

implications for decision-making and deterrence. Criminology 49:1029–1061

Matsueda RL, Kreager DA, Huizinga D (2006) Deterring delinquents: a rational choice model of theft and violence. Am Sociol Rev 71:95–122

McCarthy B (2002) New economics of sociological criminology. Annu Rev Sociol 28:417–442

Nagin DS (forthcoming) Deterrence: a review of the evidence by a criminologist for economists. Annu Rev Econ

Pogarsky G (2002) Identifying deterrable offenders: implications for deterrence research. Justice Quart 19:431–452

Pogarsky G, Piquero AR (2003) Can punishment encourage offending? Investigating the 'resetting' effect. J Res Crime Delinq 40:92–117

Sherman LW (1990) Police crackdowns: initial and residual deterrence. In: Tonry M, Morris N (eds) Crime and justice: an annual review of research, vol 12. University of Chicago Press, Chicago, pp 1–48

Stafford MC, Warr M (1993) A reconceptualization of general and specific deterrence. J Res Crime Delinq 30:123–135

Thaler RH (1996) Doing economics without homo economicus. In: Medema SG, Samuels WJ (ed) Exploring the foundations of research in economics: how should economists do economics? Russel Sage Foundation, NY

Tversky A, Kahneman D (1981) The framing of decisions and the psychology of choice. Science 211:453–458

Offender Outcomes

► Offender Change in Treatment

Offender Profiling

► Behavioral Investigative Advice
► Investigative Psychology

Offender-Based Crime Survey

► History of the Self-Report Delinquency Surveys

Offending

► Establishing Causes of Offending in Longitudinal and Experimental Studies

Offense Specialization: Key Theories and Methods

Christopher J. Schreck
Department of Criminal Justice, Rochester Institute of Technology, Rochester, NY, USA

Overview

The topic of offense specialization is one that is distinct from that of offending generally. For offending, the questions often concern how much crime that people commit and why some people do it and others do not. For specialization, the question is instead, among those who do commit crime, are there any who show a preference for a particular type of crime and, if so, then why. The question is of particular interest to policymakers, who want to be able to target their resources toward controlling, for example, the serious violent offender. Academic criminologists are interested in specialization as well, because whether it exists or not speaks to the validity of many theories of crime. The debate about the existence and value of even studying specialization has sometimes been lively.

Research on offense specialization goes back many decades and the empirical support for its patterns, and even its very existence, was long inconclusive. The development of new statistical approaches, as well as new conceptions of what specialization means, has revitalized research in this area. This more recent work has established that a small minority of offenders do specialize and even that this tendency is relatively stable over a period of time and, to some degree, predictable. Research examining how well criminological theories account for specialization, however, is in its infancy. At this point, no research has conclusively been able to locate support for any particular theory positing reasons why some offenders prefer one type of crime and other offenders another.

Introduction

In public discourse, individual offenders have commonly been described as "drug users," "sex offenders," "murderers," or "arsonists," depending on what they were arrested for, with the idea that the label characterizes the sort of crimes that they prefer and tend to commit over and over again. This, essentially, is what specialization is all about; specialists focus on a limited and specific range of offense types and thus such descriptors are appropriate. At the other end of the continuum is the idea that offenders commit a wide variety of offenses with no obvious inclination toward any one type. That is, they may take drugs but violence and property crime comprise a share of the offenses as well. Such offenders who are not specialists are described as "generalist" or "versatile." To the degree that offenders do not specialize, then attempts to classify them by the crimes they commit are pointless and targeted programs doomed to failure. The question of whether specialization even exists or not, and what causes it, is thus very important in terms of crime policy and within academic criminology.

Policymakers are particularly interested in controlling the most serious types of crime, such as violence, and so much interest in the topic of specialization reflects this. The public tends to fear violent crime, and violence produces great cost for individuals and society beyond the loss of property (if any) – injury, lost wages, death, expenses associated with attempts to improve physical security, and so on. The topic of specialization research has a long history in academic criminology as well, primarily because of its interest to policymakers as well as its relevance to many criminological theories. Theoretical concerns, however, have not generally motivated much of the research on specialization over the years; rather, scholars have invested their energies in developing innovative statistical procedures in order to address difficult methodological problems. The earliest work on specialization employed transition matrices to examine specialization and defined specialists as those who had an appropriate sequence of arrests (e.g., arrest for robbery followed by an arrest for the same crime). This type of approach, although it would dominate studies of specialization over the next two decades, only reported weak evidence of specialization and could not allow for in-depth theory testing.

Specialization has since come to include the diversity of crimes an offender commits over a period of time. Offenders who have a wide mixture of criminal activity would indicate versatility, where those who show more focus in their participation in illegal activity would indicate a specialist. Research using this conception of specialization has tended to show more consistent support for specialization. Statistical methods that define specialization in terms of the diversity of the offense repertoire have also allowed researchers to attempt to locate predictors for specialization. This research, however, is in its infancy; however, it does consistently show that groups of people do specialize in such crimes as violence.

In the following sections, a review of the major frameworks for understanding specialization will be provided as well as an examination of how researchers have attempted to not only establish the existence of specialization but also to empirically evaluate these competing theories.

Perspectives on Specialization

Notwithstanding the generally inconsistent evidence on specialization, many theories take a clear position about whether specialization exists or how it might occur. Beginning over a century ago, criminologists such as Cesare Lombroso had attempted to create typologies of offenders, classifying them based upon the sort of crimes that they tended to commit. These early schemes have largely been abandoned in favor of general theories, which are designed to encompass the processes common to all crimes.

Nevertheless, a number of perspectives have gained popularity in recent years arguing that offenders do specialize. In contrast to general theories, these approaches describe the unique process that ultimately produces a specialist offender.

The criminal career perspective. The criminal career perspective was the driving force for research on specialization between the 1980s and 1990s. This approach was not a theory, insofar as it offered no explanation for *why* individuals committed crime or specialized in some crimes over others, but rather was intended to supply a framework or language for describing the long-term development of criminal patterns. Research on specialization across the criminal career was primarily interested in the sequence of offense types committed by those who participated in crime. Those persons who switched from one type of crime to another were "versatile," whereas those who committed a series of the same offenses were defined as specialists.

The cultural deviance perspective. A more detailed account that predicts not only specialization but also explains why some offenders prefer some crime types over others can be found in the cultural deviance group of theories. The basic orientation that these theories all share is the idea that crime is something that is acquired through contact with other individuals or the prevailing culture. If an individual is embedded within a culture that values white-collar crime (and not other types of crime), for example, then those types of crime will dominate his or her offense repertoire. Specialization is not required; however, the correct combination of cultural influences can result in offenders who specialize.

A leading contemporary example of how cultural deviance theories address specialization comes from research on subcultures of violence, most notably the work of Elijah Anderson and his account of the code of the street. The code of the street is a normative system that views violence and violent retaliation as the chief means that poor African American males living in disadvantaged areas can build and assert their reputations and protect themselves. To the degree that these young males accept beliefs that acts of disrespect require a violent response, then a greater share of their crimes should be violent in nature. Similar normative systems are evident in other areas of the United States, including, for example, the Southern subculture of violence in rural Appalachia.

The control perspective. The control perspective begins with the assumption that people will naturally pursue pleasure and avoid pain; in this respect, one can understand all offenses in the same light insofar as they all bring advantage at (so far as the offender is concerned) little cost. Or, put another way, control theories typically see no distinction in the reasons offenders engage in any particular type of crime, since they all produce advantage for the offender. Gottfredson and Hirschi's self-control theory, among the most well known of control theories, explicitly rejects specialization. Self-control refers to the acquired tendency of individuals to weigh the long-term consequences of their decisions. To assume that there is specialization is to make a case that there are different causal processes that lead to the different categories of offenders. But Gottfredson and Hirschi examined the nature of crime and argued that they all share the same basic qualities. Engaging in them produces quick, easy, and certain short-term advantage and long-term negative consequences: drinking, drug use, white-collar crime, robbery, and theft. To that degree, low self-control can account for the commission of any of them and specialized theories were not necessary.

Strain theories. Strain, or anomie, theories begin with the assumption that people would like to avoid offending but are compelled by the failure to meet certain cultural expectations (e.g., the American Dream) into crime. Robert Merton's classic version of strain theory, developed in the 1930s, asserted that people adapt to strain in one of five ways: conformity, innovation, ritualism, retreatism, or rebellion. The retreatist, for example, based on their circumstances, would more likely engage in substance use. The innovator, in contrast, is an ambitious person who would break the law in order to achieve material wealth. Richard Cloward and Lloyd Ohlin's extension of Merton's theory retained the specialization argument in slightly different form, arguing that in some areas a "criminal subculture" (which would emphasize economically motivated crime) would emerge. Where opportunities for economic advancement are lacking, then a "conflict subculture" might appear; in this context, violent crime would prevail. Where offenders live in neighborhoods with diminished legitimate and illegitimate opportunities, so-called double failures, then a "retreatist subculture" is likely to emerge.

General Strain Theory, a more recent extension of Merton's theory, has begun to make arguments that indicate specialization can occur. For instance, Robert Agnew indicated that males are more likely than females to respond to strain with offending that is directed against others. Agnew's theory and research suggests that males are more expressive and retaliatory as they experience strain which increases their chances of engaging in violent crime when compared to females.

Moffitt's developmental taxonomy. Terrie Moffitt's theory is one of the leading developmental theories in contemporary criminology and it proposes that there are two distinct offender pathways, each possessing unique points of origin. The "adolescent-limited" offender is someone whose violations of the law start relatively late, in adolescence, and primarily focus on rule-breaking behaviors rather than serious crime. These offenses are perhaps best understood in terms of reactions against the rules that constrain the autonomy of adolescents, the "maturity gap," and the resultant frustration and are mainly characterized by status offending (e.g., truancy, smoking) and other minor crimes. The "life-course-persistent" offender, in contrast, begins offending fairly early in life and this offender's repertoire includes not only rule-breaking behavior but also more serious crime. This type of offender has its origins in neuropsychological deficits present at birth and extremely poor parenting. In this respect, adolescent-limited offenders tend to specialize more in minor acts of rebellion than the life-course-persistent offenders.

Methods for Studying Specialization

Although there are clearly many frameworks that allege specialization can occur, the research literature testing the validity of these is not well developed. A significant reason for this is the limitations in the availability of statistical methods with which to study specialization data, which is a problem that in recent years has begun to be better addressed. This problem suggests that the measurement of offense specialization has been a challenging issue.

In the earliest years of research on specialization, researchers have preferred to use transition matrices to explore the sequence of arrests by type. These statistical methods, used to study specialization since the 1960s, estimate the chance for an arrest of a given type at Time 2 after an arrest for a crime at Time 1. Research using this approach has often found a fairly small degree of specialization in arrests. Marvin Wolfgang and his colleagues in the early 1970s, for instance, found virtually no connection between sequential arrests, a pattern that research over the next several decades would generally replicate.

The next major innovation was the development and publication of David Farrington's "Forward Specialization Coefficient" (FSC) in the late 1980s. The FSC, which also examined specialization in the sequencing of offenses, was created in part as a response to the concern that some types of crime are quite rare, particularly those that are violent. The offenses that policymakers tended to be least concerned about – such as thefts – are in fact the likeliest crimes to occur. Not surprisingly, using a transition matrix it would be easier to detect theft specialization; however, this would not be all that meaningful given that thefts are common and could be predicted based on chance alone (evidence of specialization among rarer offense types, like violence, would be correspondingly more difficult to locate). The FSC contrasted the level of specialization that one would predict based on chance alone with that actually observed. The method produced a standardized score for the sample, with 0 indicating complete versatility (or, put another way, all evidence of specialization could be explained by chance) and 1 signifying complete specialization. The FSC in turn came to be the preferred measure for studying specialization, although not without criticism with regard to such concerns as the interpretation of the coefficient. Research seemed to indicate that adult offenders were the most apt to specialize, for example, but conclusive evidence was disappointingly elusive.

Researchers also perceived that there were disadvantages to using transition matrix approaches, including the FSC. First, many theories of crime do not appear to define specialization in terms of a sequence of events. As a result, studies examining sequences have uncertain theoretical implications because theories, such as those summarized earlier, generally do not make assertions like "a robbery arrest must be followed by another robbery arrest." Second, sequences of events create analysis problems. For instance, with the FSC and the transition matrix, a person who committed a robbery at T1 and not at T2 is not a specialist. But what if that person committed a robbery at T3? As a result of questions like this, researchers began to explore statistical methods that would capture specialization in terms of the proportion of *all* offenses that were of a given type. Third, matrix-based approaches can only report specialization in an aggregate sense. In other words, the analysis can identify the level of specialization within the entire sample, but this meant that it was extremely difficult to locate risk factors that would help academics and policymakers classify individual offenders as specialists. While the literature using transition matrixes might crudely show that adults are more prone to specialize, it was not possible to predict which individuals among these adults would be most likely to specialize without committing the ecological fallacy.

Alex Piquero and his colleagues were the first to use a "diversity index" to measure specialization. This index was originally developed by statisticians who were interested in the variety of species of animals and later applied to criminological work, and the technique offered important advantages over earlier methods insofar as it addressed the problems outlined above. Rather than considering the observed sequence of two arrests, as with transition matrixes, the index considers the entire spectrum of crimes a person has committed. Little diversity in a person's crime record would indicate a specialist offender, where a broad range of crime would mark the versatile or generalist offender. This conception of specialization offered important new advantages over the FSC. For the diversity index to show specialization, it was not necessary to find contiguous arrests because the analysis weighs the entire portfolio of offense activity over a period of time. This definition of specialization also tends to be more consistent with many of the theories designed to account for the emergence of specific types of crime.

The index also produced a score for each individual, thus allowing researchers to test the impact of theoretically relevant variables on offense diversity or specialization in a regression framework, which was a task for which transition matrixes were poorly suited. Legitimate scores would range from 0 (complete versatility) to 1 (complete specialization). Because each individual receives a diversity score, this can in turn be used as an outcome measure in a regression analysis. Tests that included both the FSC and the diversity index showed similar patterns in specialization. Tests of criminological theories using this method, often focusing on Moffitt's developmental taxonomy, has reported inconsistent results showing that adolescent-limited offenders are more apt to specialize than life-course-persistent offenders.

The diversity index, although more flexible than the FSC, also has limitations; tests of specialization often face difficult analytic problems that frequently make no single statistical method ideal. First, even if an offender committed, for example, nine offenses of the same type, if there were 10 or more crime categories then that individual will not receive the maximum diversity score. Another concern centers on the fact that the diversity index does not take into account the base rates of offending across the population. That is to say, crimes that are more common (e.g., thefts) might give the appearance of specialization when that offender's record in no way differs from that of other offenders.

Alternatively, researchers can use latent class analysis (LCA), which resembles factor analysis and is used to classify offenders rather than assess how well the component measures cluster together. Only a small number of studies have used this approach, producing some mixed findings. Some early work found that LCA could only distinguish between offenders and nonoffenders, while more recent work using offender populations and a larger number of offense categories reported the ability to classify offenders. Some scholars have noted limitations with this approach, such as the confounding between offense frequency and specialization. That is to say, offenders who commit more crimes are likely to have at least some offending of a certain type, say violence. Because of this, the determination of how much specialization there is depends upon the researcher's assessment.

A final technique that researchers have used with some frequency in recent years is the Item Response Theory (IRT)-based approach developed by Wayne Osgood and Christopher Schreck. This method employs a multilevel regression approach, with the first level assessing whether specialization exists and the second level incorporating predictors of specialization. There are additional attributes that make this approach very useful for assessing specialization. First, the approach defines specialization as the degree to which an individual deviates from the population base rates for all offenses. For instance, if we assume that 80 % of the crimes committed by everyone are nonviolent and 20 % are violent, then someone with that same distribution is not especially unique; however, a person with 70 % violent and 30 % nonviolent offenses differs quite dramatically from the population. Second, the IRT model incorporates data from all subjects, including nonoffenders and low-rate offenders; many of the approaches described in this section (e.g., the FSC, diversity index) can only employ samples of offenders, usually only those with numerous arrests. Offenders with a higher level of activity will receive greater weight in the IRT-based analyses, however. Third, the approach controls for the frequency of offending as well as for specialization. Research using this method has shown that specialization does in fact exist among a minority of offenders, that their tendencies to specialize are stable over time, and that it is possible to predict specialization in violence using theoretically relevant predictors. Jean McGloin and her colleagues, in a recently published article using this statistical method, found that subculture-of-violence measures failed to predict tendencies to specialize in violence, contrary to the theory described earlier. This body of work is in its infancy, however. The ability of the IRT method to exploit predictor variables, as with the diversity index, nevertheless leaves the field of specialization poised for significant new advances in theoretically relevant research.

As the research literature developed, however, it became clear that studying specialization introduced many challenges that make efforts to locate it more difficult. For instance, the source of data researchers use can influence whether they find specialization or not. Specialization appears more likely to emerge with self-report data, where individuals describe their own offending, than with official arrest data. This is perhaps not surprising; arrests tend to be relatively infrequent events even among those with many arrests, and many crimes that people commit go unreported to the police. Self-report data thus can provide a more comprehensive picture of what offenders are probably doing; however, there is a limit to the level of detail one can reasonably expect. For instance, with self-report data the subjects may have a difficult time, simply due to memory problems, establishing the time order of their criminal activity. Official arrest data, vetted by police and accompanied by a clear time of arrest, are more suitable for such methods as the transition matrix.

A further issue that has developed in specialization research concerns how scholars ought to measure it. That is to say, what offenses ought to get included and how ought they to be grouped together (e.g., violent vs. nonviolent, white-collar vs. street crime). The diversity index has great flexibility for incorporating multiple offense categories; however, it has a difficult time discriminating between types of offenders. For instance, two offenders with scores indicating maximum specialization might specialize in very different crimes (e.g., one in drugs, the other in rapes), with each type possibly arising from different causal processes. The IRT-based approach, in contrast, can only examine the contrast between two general categories. If theories identify three or more types of offenders, then the method cannot easily accommodate that many classifications.

Conclusion

Research on specialization is of potentially great importance in the study on crime and for policy. Those who make policy decisions have an interest in locating and targeting crime control efforts toward those who are most likely to engage in serious crime, especially violence, through practices like selective incapacitation. Academics have contentiously debated the value of research on specialization, largely because evidence (or its lack) speaks to the legitimacy of their theories. Some theories argue that violent offenders, for example, were subject to specific processes that made them violent; other processes lead to offenders who prefer other types of crime. General theories, like self-control theory, however, identify processes common to all crimes and argue that specialization cannot exist in a meaningful sense.

For decades, strong evidence of specialization (and its causes) has been elusive. In recent years, especially with the advent of research methods defining specialization in terms of diversity rather than a sequence of offenses, evidence that some offenders specialize has strengthened. Locating exactly why this is so, however, has proven difficult. This research, however, is in its early stages and can benefit from verification as well as the incorporation of additional theoretical explanations.

Recommended Reading and References

Agnew R (2006) Pressured into crime: an overview of general strain theory. Roxbury, Los Angeles, CA

Anderson E (1999) Code of the street: decency, violence, and the moral life of the inner city. Norton, New York

Blumstein A, Larson R (1969) Models of a total criminal justice system. Operations Res 17:199–232

Blumstein A, Cohen J, Das S, Moitra SD (1988) Specialization and seriousness during adult criminal careers. J Quant Criminol 4:303–345

Britt CL (1996) The measurement of specialization and escalation in the criminal career: an alternative modeling strategy. J Quant Criminol 12:193–222

Bursik RJ (1980) The dynamics of specialization in juvenile offenses. Social Forces 58:851–864

Cloward RA, Ohlin LE (1960) Delinquency and opportunity. Free Press, Glencoe, IL

Colvin M, Pauly J (1983) A critique of criminology: toward an integrated structural-Marxist theory of delinquency production. Am J Sociol 89:513–551

Deane G, Armstrong DP, Felson RB (2005) An examination of offense specialization using marginal logit models. Criminology 43:955–988

Farrington DP, Snyder HN, Finnegan TA (1988) Specialization in juvenile court careers. Criminology 26:461–487

Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford, CA

Kempf KL (1987) Specialization and the criminal career. Criminology 25:399–420

Lattimore PK, Visher C, Linster R (1994) Specialization in juvenile careers: Markov results for a California cohort. J Quant Criminol 10:291–316

Lombroso C (1911) Crime: its causes and remedies. Little, Brown, and Company, Boston

Mazerolle P, Brame R, Paternoster R, Piquero A, Dean C (2000) Onset age, persistence, and offending versatility: comparisons across gender. Criminology 38:1143–1172

McGloin JM, Schreck CJ, Stewart EA, Ousey GS (2011) Predicting the violent offender: the discriminant validity of the subculture of violence. Criminology 49:767–794

Merton RK (1938) Social structure and anomie. Am Soc Rev 3:672–682

Moffitt TE (1993) Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. Psychol Rev 100:674–701

Osgood DW, Schreck CJ (2007) A new method for studying the extent, stability, and predictors of individual specialization in violence. Criminology 45:273–312

Paternoster R, Brame R, Piquero A, Mazerolle P, Dean CW (1998) The forward specialization coefficient: distributional properties and subgroup differences. J Quant Criminol 14:133–154

Piquero AR, Paternoster R, Brame R, Mazerolle P, Dean CW (1999) Onset age and specialization in offending behavior. J Res Crime and Delinq 36:235–274

Reiss AJ, Roth JA (1993) Understanding and preventing violence, vol I. National Academy Press, Washington, DC

Sullivan CJ, McGloin JM, Pratt TC, Piquero AR (2006) Rethinking the "norm" of offender generality: investigating specialization in the short-term. Criminology 44:199–233

Wolfgang M, Figlio R, Sellin T (1972) Delinquency in a birth cohort. University of Chicago Press, Chicago

Officer Safety, Health, and Wellness

Karen L. Amendola
Police Foundation, Washington, DC, USA

Synonyms

Fatigue; Health and wellness; Safety; Shift work; Stress; Suicide

Overview

The concepts of officer safety, health, and wellness are both distinct and interrelated. This entry will focus on ongoing and emerging issues in police safety (e.g., increased violence toward the police, increased risks due to encounters with dangerous situations and persons, vehicular safety including officer pursuits, and protective measures that can be employed by agencies and law enforcement officers). In addition, issues of officer stress will be examined in terms of stress types, causes, and effects.

Recent research has also focused on health issues associated with shift work and the profession of policing more broadly, indicating greater risks for cardiovascular disease and cancers (compared to the general population), causes and sources of officer fatigue and its relationship to various negative outcomes, the higher rates of sleep disorders for police officers compared to the general population, and earlier morbidity from various causes including suicide.

Finally, policy and practice implications are discussed to promote a culture of safety, health, and wellness while also emphasizing personal responsibility. Individual officers vary significantly in the extent to which they make physical health a priority, focus on mental and emotional health (although many do not take responsibility for seeking out such assistance), and practice general safety procedures such as wearing seatbelts and body armor.

Introduction

The issue of officer safety, health, and wellness not only impacts the length and quality of officers' lives but also has significant implications for community safety, as well as departmental costs and liability. Although the issue of police stress has been examined by scientists and practitioners alike for many years, there has been little empirical research when it comes to long-term health, wellness, and safety. However, due to a number of recent national initiatives, there has been increased awareness of, and attention to, the issues of safety, health, and wellness among police officers. First, the National Institute of Justice (US Department of Justice) initiated a series of studies in the last decade focused on these issues. More recently, current US Attorney General Eric Holder committed $40 million for officer safety initiatives in 2010 and launched a law enforcement officer safety initiative in October of that year called VALOR – *Preventing Violence Against Law Enforcement and Ensuring Officer Resilience and Survivability* (http://www.mainjustice.com/2010/10/26/holder-launches-new-law-enforcement-safety-initiative/). Then in March 2011, the Attorney General issued a directive to all US Attorneys which was copied to the agency heads of the Federal Bureau of Investigation (FBI), Drug Enforcement Administration (DEA), US Marshals, and the Bureau of Alcohol, Tobacco, Firearms, and Explosives (ATF) calling on them for increased actions related to his Officer Safety Initiative (http://www.mainjustice.com/2011/03/22/holder-launches-officer-safety-initiative/).

As a result of these initiatives, two US Department of Justice offices have become involved in the issue. Specifically, the Office of Justice Programs' Bureau of Justice Assistance (BJA) and the Office of Community Oriented Policing Services (COPS Office) have, on behalf of the Attorney General, convened the Officer Safety and Wellness Working Group (OSWWG) made up of research scientists, practitioners, and policy makers in order to address these issues. This working group has focused on felonious killings and shootings of police officers, traffic accidents, and a variety of health and wellness issues, including officer suicides. Also, BJA has published a fact sheet entitled *Officer Safety Initiatives* (Edwards and Meader 2011) that includes links to a number of officer safety resources and web links (https://www.bja.gov/Publications/OfficerSafetyFS.pdf).

In addition, initiatives by the National Law Enforcement Officers Memorial and the International Association of Chiefs of Police (IACP) have focused on violence against officers and traffic safety. For example, in August 2011, the National Law Enforcement Officers Memorial Fund and the National Highway Traffic Safety Administration (NHTSA) initiated an innovative partnership to promote law enforcement officer safety on the roadways. Finally, the IACP has initiated and produced a number of safety initiatives and publications. One such initiative is *Safe Shield* which looks at all aspects of officer injuries and deaths through improved education, training, and policies as well as improvements to uniforms and equipment. As recently as September 2012, the IACP held a webinar on officer safety and wellness to announce the opening of its new Center for Officer Safety and Wellness with the goal of enhancing law enforcement's ability to be well equipped, well trained, and physically and mentally prepared to confront violence as well as other threats and dangers. In sum, it is increasingly apparent that researchers and police leaders have answered the call and become more involved in addressing concerns over officer safety, health, and wellness including the prevention of violence against officers and an improved ability to confront it.

Importantly political leaders have also taken an interest in officer safety issues. For example, in May 2012, US Senator Chris Coons (D-Del.) hosted a briefing for members of Congress and their staff to learn about and experience wearing bullet-resistant and stab-resistant vests. The *Bulletproof Vest Partnership* (*BVP*) *Grant Act of 2012*, which was introduced by Senator Patrick Leahy (D-Vt.) in May 2012 and cosponsored by Senator Coons, is a reauthorization of the competitive grant program from the Bureau of Justice Assistance that provides funding to state and local law enforcement agencies to assist in their purchasing of ballistic-resistant and stab-resistant body armor that complies with National Institute of Justice body armor standards. Since the BVP program's first authorization in 1999, it has reimbursed grantees approximately $247 million for their purchases of nearly one million vests.

Background on Officer Safety, Health, and Wellness

Issues of officer safety, health, and wellness are not new. However, due to advances in the field, state of the art, and a resurgence of interest in recent years, this entry will focus primarily on data from the past 10–15 years. The two major safety issues include assaults (including felonious killings) and vehicular accidents. Data on law enforcement officers killed and assaulted has been collected for decades. The number of law enforcement officers killed in the line of duty was collected by the FBI from 1937 to 1971 as part of the Uniform Crime Reports. However, in 1972 the FBI started to collect more detailed data associated with officers killed in the line of duty. Additionally, benefits are offered to victims' families via the Public Safety Officers' Benefits Program, a federal (Bureau of Justice Assistance), state, and local collaboration to provide to death and educational benefits to survivors. Most states offer pensions, worker's compensation, and burial support to spouses and children of fallen officers, and other organizations provide programs for survivors and/or law enforcement agencies. One example is the group Concerns of Police Survivors (C.O.P.S.) that provides resources to assist in the rebuilding of the lives of surviving families and affected coworkers of law enforcement officers killed in the line of duty, provides training to law enforcement agencies on survivor victimization issues, and educates the public of the need to support the law enforcement profession and its survivors.

As early as 1929, police departments began to concern themselves with traffic safety. For example, the Evanston, Illinois, Police Department established an Accident Prevention Bureau led by a young sergeant and student at Northwestern University (NWU) who developed a safety model that took ideas from engineering, enforcement, education, and research. Later, that sergeant became a lieutenant and went on in 1936 to become founding director of NWU's newly established Traffic Safety Institute, the first known of its kind nationwide. Today, many of the courses offered at NWU's Center for Public Safety are offered to the general public like defensive driving. However, the Center focuses on educating law enforcement officers in many areas such as police accident investigation, motorcycle operation and instruction, radar, and field sobriety instructor training, as well as many non-traffic safety areas.

Assaults and Felonious Killings. During the decade from 2001 to 2010, the FBI reported that assaults on police officers averaged about 58,500 (ranging from a low of 53,469 in 2010 to a high of 60,851 in 2007). While assaults on police officers appeared to be dropping at the end of the decade, data from the National Law Enforcement Officers Memorial Fund (NLEOMF) showed that during the 10-year period between 2002 and 2012, shootings of officers (570) surpassed vehicle accidents (470) as the leading cause of line of duty officer deaths. In 2011, there was a 20 % increase in fatal shootings of law enforcement officers over the prior year according to Executive Director Craig Floyd (2012) of the NLEOMF (http://www.nleomf.org/facts/officer-fatalities-data/causes.html and http://www.nleomf.org/facts/officer-fatalities-data/).

Data from 2011 showed that over 30 % of officers killed were not wearing body armor (http://www.cops.usdoj.gov/pdf/OSWG/Meeting_01-26-2012.pdf), in spite of the fact that since the introduction of modern material vests/body armor in the 1970s, the number of officers surviving shootings has gone up dramatically. Indeed as of March 2011, the IACP/DuPont Kevlar Survivors' Club® had documented more than 3,126 incidents where body armor protected officers from more serious physical injuries or death. In 1994, the FBI also found that the rate of death of officers shot is 14 times greater for those who do not wear body armor than for those who do (as reported in Miller 2001). Advanced material bullet- and stab-resistant vests can be costly. Since the 1970s, several new bullet-resistant fabrics, fibers, and construction methods have been developed besides woven Kevlar, such as DSM's Dyneema, Honeywell's Gold Flex and Spectra, Teijin Twaron's Twaron, Pinnacle Armor's Dragon Skin, and Toyobo's Zylon, to name several examples. While the manufacturers of many of these newer materials advertise these as thinner, lighter, and more resistant than Kevlar, they tend to be more costly.

Taylor and colleagues (2009) conducted a national survey of law enforcement agencies and found that the vast majority (87 %) issue body armor to their officers. Slightly fewer than 20 % (17.7 %) indicated that their officers purchase their own body armor, but 37 % of those provide full reimbursement, and the remaining 63 % provide reimbursements of at least 33 %. Furthermore, in an effort to increase protections to officers, the federal Bulletproof Vest Partnership (BVP) Program (which has a mandatory wear policy) has provided over $300 million to provide bullet- and stab-resistant vests to agencies for provision to their officers.

In light of the above statistics, it is both astounding and curious as to why officers would not wear vests especially because cost does not appear to be a major factor. One possible reason is that many agencies do not require that their officers wear body armor. Taylor et al. (2009) reported that just 59 % of law enforcement agencies surveyed had a mandatory wear policy for body armor and only 45 % had a written policy on this matter, even though most provided body armor to their officers. Anecdotal information suggests that some of the other reasons or circumstances in which officers choose not to wear body armor consistently include living in hot climates or during summer months, at times or in places where they do not feel at a high risk (even though any situation can become risky), or that they are cumbersome (although modern technology has resulted in significant reductions in weight, thickness, and heat retention). Most certainly these concerns ought to be outweighed by the lifesaving statistics.

Importantly, the VALOR program (above) provides training and technical assistance to officers and agencies in order to become more aware of and prevent violent encounters and ambush-style attacks on police. A number of other organizations are providing training in tactics, rapid response, and mitigation of risks and threats. Researchers examining aspects of deadly encounters and assaults as well as officer-involved shootings identified a number of offender characteristics in domestic assault cases that predicted assaults on police officers like alcohol consumption, living with the victim of domestic violence, and hostile demeanor toward officers. The provision of training to raise awareness of these risk factors could help prevent violent attacks on police.

Vehicle Accidents. Until very recently, traffic fatalities have been the leading cause of death of officers killed in the line of duty (at least for the years from 1997 to 2010). Yet, according to the NLEOMF, 21 % of officers killed in vehicle crashes were not wearing seat belts at the time of their fatal crashes (http://www.cops.usdoj.gov/pdf/OSWG/Meeting_01-26-2012.pdf). According to a 2011 report by the National Highway Traffic Safety Administration, 93 % of all law enforcement officers involved in fatal crashes were males, although Bureau of Justice Statistics' data indicates that by 2007, 12 % of all local law enforcement officers nationwide were women.

Among the most risky driving situations in policing are pursuits. In 2000, the IACP established a police pursuit database after years of recommending reforms via its "Vehicular Pursuit Model Policy" developed in 1996. That policy, in recognition of serious public safety risks in pursuits, establishes that officers should take into consideration the weather, population density, vehicular and pedestrian traffic, the capabilities of the pursuit car as well as the vehicle being pursued, the presence of others in the vehicle, and perhaps the most controversial aspect, the seriousness of the offense. Perhaps the reason for this controversy stems from what leading expert Professor Geoff Alpert has called the "competing values" of apprehending law breakers and ensuring public safety. Of course, it also comes from the increased liability to a law enforcement agency for injuries (or even death) of both bystanders and those involved and also for property damage. According to the IACP Pursuit Database, over 40 % of pursuits result from traffic-related infractions, and 35–40 % of pursuits end in a collision, and the FBI reports that 50 % of the collisions occur in the first 2 mins of the pursuit. As the area of pursuits continues to be more closely managed through greater restrictions and more specific policies set by police leaders, the risk factor for vehicular accidents and the associated outcomes involving the police and the public will continue to go down.

Summary. Current policies and procedures related to officer safety in public encounters and in vehicular operation and pursuits are reflective of the knowledge attained through statistics on the risk associated with various types of encounters such as traffic stops and domestic disturbances, as well as vehicle safety and police pursuit practices. Perhaps we are beginning to see signs of safety improvements and initiatives taking effect. As of September 15, 2012, firearm deaths of officers were down 38 % and traffic-related deaths 18 % over the same period in 2011.

Background on Officer Health and Wellness

Psychological Impacts of Stress. The issue of police stress has been studied for several decades. Originally thought to be one of the most stressful occupations in America, the assertion that policing is exceptionally stressful has not always been supported by scientific research. Early stress models emphasized a "life events" approach to stress; however, subsequent research has tended to focus on routine and organizational stressors. Emerging research on the personality dimension of resilience has provided some promise for new ways to prepare officers to better deal with job-related stress. Later in this entry the issue of wellness will include some discussion of programs related to building resilience and overcoming stress.

Stress Measures. Similarly, within occupational stress literature, and specifically that of policing, the measures of occupational stress have typically focused on either operational or organizational stressors. For example, Evans and Coman (1993) found that work stressors were either related to job content (the duties of the officers) or job context (derived from the police organizations). Similar research by Hart, Wearing, and Heady (1995) suggested that daily hassles associated with police work and the organizational bureaucracy produced greater levels of stress and affected quality of life more so than critical incidents. Most recently, McCreary and Thompson (2006) developed the Police Stress Questionnaire to assess two forms of stressors, those caused by the operational aspects (critical incidents) of policing and the other on the organizational stressors (administrative or imposed by the organization).

Officer-Involved Shootings. Research suggests that officers who are directly involved in shootings experience a variety of short- and long-term physical and psychological reactions such as tunnel vision, misperceptions of time, sound reductions, a sense of numbness, ruminations about the incidents, difficulty sleeping, sadness and associated crying, and nausea (see e.g., Artwohl 2002; Beehr et al. 2004). However, research over the past decade by Klinger (2006), in which he interviewed law enforcement officers involved in 113 separate shootings during the course of their careers, revealed that the act of shooting someone did not lead to lasting disruptions in the lives of officers. In more than half of the shootings, officers did not report negative psychological, emotional, or physical responses after a week following the incidents. After 3 months, that proportion increased to two-thirds. However, the primary lingering consequence appeared to be recurrent thoughts. Negative reactions appear to be mitigated by support from fellow officers following the incident and mandatory time off. However, family support and mandatory counseling did not appear to be associated with negative outcomes for police who engaged in shootings (Klinger 2006). However, it should be noted that research has previously demonstrated that officers' accounts typically differ from reality due to memory distortion. Nevertheless, recent research by Komarovskaya and colleagues (2011) indicated that killing or seriously injuring someone in the line of duty was significantly associated with PTSD symptoms and marginally associated with depression symptoms.

Police-on-Police Shootings. An additional stressor related to officer-involved shootings, is the issue of police-on-police shootings especially those resulting from mistaken identity. Unfortunately, these incidents often result from encounters between on-duty and off-duty, plainclothes officers. These encounters can escalate quickly due to misunderstandings either as a result of poor training, unclear protocols, or when officers are from differing jurisdictions. The New York State Task Force on Police-on-Police Shootings found that such shootings can easily traumatize police officers, families, and police departments, sometimes even polarizing entire police agencies. Importantly, the task force found "persuasive scientific evidence" that officers, like the rest of the population, share unconscious racial biases. This topic was of great enough concern that it became the focus of a 2010 "Research for the Real World Seminar Series" of the National Institute of Justice. In that series, an interview was conducted with the chair of the New York State Task Force Christopher Stone of Harvard University. Stone indicated that since 1981, 28 mistaken identity police-on-police shootings had occurred. The above data suggest that traumatic experiences may in fact increase officer stress and result in related physical and psychological consequences.

Posttraumatic Stress Disorder (*PTSD*). Typically associated with experiencing or witnessing severely traumatic incidents, PTSD is a constellation of symptoms of a psychiatric nature marked by clear biological changes and/or psychological symptoms. Psychiatric and psychological research over the past decade has demonstrated, consistent with other stress research, the routine work environment stressors are most strongly associated with PTSD symptoms, even when controlling for prior traumatic experiences, current negative life events, gender, ethnicity, and recent critical incident exposure (Liberman et al 2002; Maguen et al. 2009). Among the most common environmental stressors are lack of concern by management and supervisors, equipment safety, training, discrimination, shift work, role conflict, peers, and unfair public perceptions. This recent research suggests the need for greater attention to environmental stressors of a more routine nature that can build over time or be additive in nature.

Ongoing longitudinal research by Violanti and colleagues (see e.g. Violanti et al., 2006) has examined many health and safety factors for police officers in Buffalo, New York. One of the factors being examined is cortisol (otherwise known as the "stress hormone") in order to determine if stress is associated with risk factors predictive of more serious illnesses. Almost two decades ago, Violanti found that organizational stressors increased psychological distress by 6.3 times more than did inherent operational police stressors (Violanti and Aron 1994). And a decade prior to that, he found that police stress increases significantly in the first 13 years of an officer's career and then progressively decreases after 14 years (Violanti 1983).

Physical Health, Sleep, and Fatigue. Recent research by Violanti suggests that some longstanding patterns associated with police officers' careers (shift work, smoking, alcohol use, etc.) increase the likelihood of metabolic syndrome. Metabolic syndrome is a clustering of cardiovascular disease risk factors (obesity, elevated blood pressure, reduced HDL cholesterol, elevated triglycerides, and abnormal glucose levels). In one study Violanti (2012) found that officers with severe PTSD symptoms were three times more likely to have metabolic syndrome and therefore have higher risk for cardiovascular disease, further emphasizing the role of stress in ill-health and greater risks for disease including various forms of cancer (Vena et al. 1986).

Many of the ill effects of police work are attributed to shift work, which typically refers to those types of schedules that require working at nontraditional times such as evening or midnights. Violanti (2012) also found that the number of metabolic syndrome components was greater for younger officers working midnight shifts and that shorter sleep duration and increased overtime combined with midnight shifts may be a contributor to metabolic syndrome.

Shift Rotation. Various aspects of shift work and especially working night shifts have been demonstrated to be associated with poorer sleep and/or sleep disorders, metabolic syndrome, and other health risks. Another aspect of shift work that has been examined across industries is the extent to which shifts rotate. Shift rotation has traditionally been a common practice in law enforcement agencies, with officers rotating from day, to evening, to midnight shifts (forward rotation) or from midnights, to evenings, and then to days (backward rotation), although the frequency of rotation varies substantially as well. Typically, schedules that change twice a year or less frequently are not generally referred to as rotating shifts. Instead, a rotating shift is one in which officers rotate on a monthly basis or more typically on a much more frequent basis like every 2 weeks, week, or even everyday.

Given the extent of research suggesting negative impacts of rotating shifts, Amendola and colleagues (2011a) conducted a random national survey and surprisingly found that as recently as 2005, the use of rotating shifts was still a somewhat common practice (46 %), particularly in smaller agencies (52.1 %). However, they also found that the larger the agency, the less likely they were to employ rotating shifts (41.5 % for medium agencies and 30 % for larger agencies). When the survey was conducted 4 years later, the percentage of agencies using rotating shifts declined by almost half from 46 % in 2005 to 24.7 % in 2009. Similarly, there was an increase (about 40 %) in fixed shifts from 54 % to 75.3 %, a trend that seemed most apparent in small and medium-sized agencies. Again in 2009, the use of rotating shifts was somewhat more common among smaller and midsize agencies (25.2 % and 26.3 %, respectively), as compared to larger agencies (20.5 %).

These reductions may be due to an increasing awareness of research showing that rotating shifts typically disrupt circadian cycles and result in greater fatigue, insomnia, or other health problems (e.g., Colligan et al. 1979; Czeisler et al. 1982), and increases in worker errors and accidents (e.g., Gold et al. 1992). Furthermore, by 2009, we saw the advent of the use of both fixed and rotating schedules in agencies (4.3 %).

Shift Length Practices. While most law enforcement agencies have traditionally relied on a standard 40-h workweek consisting of five consecutive, 8-h shifts, followed by 2 days off, many agencies have used alternative shift schedules over the past several decades. In recent years, however, an increasing number of agencies have moved to some variant of a compressed workweek (CWW) schedule in which officers work four 10-h shifts per week or three 12-h shifts (or some variant). In the recent Amendola et al. (2011a) survey, researchers found that there had been a significant drop in agencies employing traditional 8-h shifts from 40.1 % in 2005 to 29.3 % in 2009. In the largest agencies (200 or more sworn officers), 10-h shifts are the most common with 35 % of them employing these shifts in 2005 and 33 % in 2009, followed by 8-h shifts and then 12-h shifts. However, small- and mid-sized agencies predominantly used 8-h shifts in 2005 (41–42 %); by 2009, these agencies had roughly equal numbers of 8- and 12-h shifts (32 % and 31 % for small agencies; 27 % and 26 % for medium).

While the trend toward CWWs had been increasing over the past decade, there had been few, if any, rigorous scientific studies examining the advantages and disadvantages associated with these work schedules for law enforcement officers and their agencies. As such, Amendola, Weisburd, and colleagues (2011b) conducted a comprehensive, randomized experiment in which shift lengths of 8, 10, and 12 h were randomly assigned to law enforcement officers in Arlington, Texas, and Detroit, Michigan, to assess their potential impacts on performance, health, safety, quality of life, sleep, fatigue, alertness, off-duty employment, and overtime among police. The study found some distinct advantages of 10-h shifts and identified some disadvantages associated with 12-h shifts that are concerning. Specifically, officers on 10-h shifts got significantly more sleep per 24-h period than those on 8-h shifts (approximately 30 min), had a higher quality of work life, and worked significantly less overtime hours. However, those on 12-h shifts did not obtain the sleep benefit, and indeed reported greater levels of fatigue and lowered alertness at work. It is important that agencies consider the safety and quality of life of their personnel and the public they serve, and therefore 10-h shifts may be optimal in this regard.

Causes and Impact of Officer Fatigue. According to research by Vila (2000), police officers on average get 6.5 h per sleep, less than that of the general population. Importantly, 40 % of officers screened positive for sleep disorders also a rate well above that of the general population. Indeed, officers with sleep disorders had an increased risk of falling asleep while driving, committing an error or safety violation attributable to fatigue, and experiencing uncontrolled anger toward a suspect.

Types, *Presence*, *and Causes of Sleep Disorders*. A recent survey showed that 40 % of police officers experience sleep disorders (Rajaratnam et al. 2011). The most common sleep disorders in police officers are obstructive sleep apnea (roughly 33 %), followed by insomnia (about 6.5 %), and then shift-work sleep disorder. Sleep apnea is a disorder characterized by abnormal pauses in breathing that can happen for a few times an hour or a few dozen times an hour. Obstructive sleep apnea is a form in which a blockage to air flow exists even though there is respiratory effort, as opposed to central sleep apnea in which there is a lack of respiratory effort.

Researchers at Brigham and Women's Hospital found the frequent use of extended shifts, long work hours, and extended work weeks can lead to increased risk for errors, unintended injuries, vehicle accidents, and other health risks (Rajaratnam et al. 2011). Indeed, officers with obstructive sleep apnea in that study were more apt to be diagnosed with diabetes and have cardiovascular disease, and those with apnea did consume more caffeine than those who did not have the condition. In addition, those in the sleep disorder category in this study also reported feeling burnout and falling asleep while driving, in addition to having more gastrointestinal and anxiety disorders and taking medication for insomnia, than those who did not have the disorder. And officers with obstructive sleep apnea were 1.5 times more likely to have depression than officers who did not. In terms of other causes, obstructive sleep apnea is sometimes caused by excess weight, and in this study population, 79.3 % of officers were overweight or obese by definition, and 33.5 % were obese. Another common symptom of obstructive sleep apnea is snoring. Violanti (2012) found that night-shift workers had greater levels of snoring even after controlling for a number of other factors.

Insomnia is a disorder in which there is difficulty falling asleep or staying asleep. It is often associated with impairment in functioning while awake like excessive sleepiness, fatigue, and psychomotor functioning. Finally shift work sleep disorder (SWSD) is associated with disruption in circadian rhythms, the normal internal clock influenced by light and dark. This disorder is most commonly experienced by those who work nights, very early days (starting at say 4:00 am), or rotating shifts. Those with SWSD are twice as likely to have work related accidents and have excessive sleepiness, reduced performance, and irritability.

Morbidity. Life expectancy of police officers is commonly believed to be lower than for the general population. Violanti's (2012) research with a current police population and cohort of deceased officers showed that mortality from all causes of death combined for police officers was significantly higher than that of the general US population. Police also had significantly higher death rates from cancer and circulatory system diseases as well, although they had a lower rate of death from nervous system and sense organ diseases, respiratory diseases, accidents, and motor vehicle accidents as compared to the general population. However, mortality due to cirrhosis of liver (caused primarily by overuse of alcohol) and suicide was elevated.

Suicide. Research by Violanti (2012) suggests that suicide ideation in police officers is highly associated with psychologically traumatic police work experiences, the development of posttraumatic stress (PTSD) in officers, and the "inordinate use of alcohol associated with this condition." Violanti suggested that police exposure to certain traumatic events increase the risk of high level PTSD symptoms, which subsequently increase the risk of high alcohol use and suicide ideation, noting that when combining the impact of PTSD and increased alcohol use increased suicide ideation risk by ten times.

Using data from the National Occupational Mortality Surveillance from 1984 to 1998, Violanti (2010) found that the police suicide rate was four times that of firefighters. Also when comparing police officer to firefighters, minority officers had 4.5 times the number of suicides, and female officers had 12 times the number of suicides. Finally, police suicides outnumbered homicides by more than 2.3 times. Although the suicide rate for police officers is often thought to be higher than that of others jobs, two meta-analyses have indicated that, after controlling for demographics, the suicide rate for police officers is actually lower than that of the population in general (Aamodt and Stalnaker 2001; Loo 2003).

Background on Officer Wellness Programs. Traditionally, most wellness programs have been limited to physical fitness, employee assistance programs, and critical incident stress debriefings (CISD). In the past, the typical focus on wellness has tended to emphasize the detection of problems and intervention such as in a CISD or an employee assistance program (EAP). Research and common wisdom regarding EAPs is that whether they are offered inside or outside the department, officers tend not to utilize them that frequently due to stigmas associated with the need for counseling or assistance, the fear that attending any sessions will be used adversely against an officer within the agency, or at a minimum that the agency will know who has participated.

However, in recent years, wellness programs have begun to emerge in law enforcement agencies across the USA. The concept of wellness today typically emphasizes physical, emotional, mental, and spiritual dimensions. Even at the Federal Bureau of Investigation, Former Supervisory Special Agent Samuel Feemster championed the concept of officer wellness from a broader holistic model in which spirituality was central including the most recent development of a program called “Beyond Survival Toward Officer Wellness” (BeSTOW) where emotional and spiritual intelligence take center stage and suicide prevention is no longer a taboo topic. Spirituality in this sense is not about religion per se, but a larger sense of meaning and purpose in law enforcement.

Many issues are driving the increased attention to employee wellness including health care costs during an officer’s career and in retirement, agency liability associated with the performance of their officers, absenteeism, stress-related problems, emotional problems, and more. According to Quigley (2008), fitness and wellness programs can benefit individuals and officers, as “officers stand to profit from an improved ability to perform job functions, reduced stress, and better physical and psychological preparation. Agencies stand to benefit in terms of efficiency as well as fiscally.”

According to Fiedler at the Office of Community Oriented Policing (COPS), US Department of Justice, promoting safety and supporting health and wellness, can be defined under four broad categories: (1) Leadership and Management, (2) Operational and Emergency Responses, (3) Training, and (4) Mental and Physical Health Wellness indicating the need for more comprehensive programs targeting policies, practices, tactics, strategies, academy and field training (including scenario based and simulation training), and physical and mental health programs (http://www.cops.usdoj.gov/pdf/OSWG/e091120401-OSWGReport.pdf).

Conclusions

The past several years have seen an increase in the attention paid to officer safety, health, and wellness issues by researchers, practitioners, and policy makers. This increase is due to such factors as including changing trends in the types and severity of situational risk factors for police officers, calls from political and governmental leaders, increasing costs and liabilities facing law enforcement agencies, and increased health and safety concerns of officers. Whether it be increasing gun violence, inherent dangers in domestic disturbance calls and traffic stops, or hostility and attacks and assaults on police officers, safety risks are ever present in policing. Due to these risks, improvements are necessary in officer vigilance, training geared toward preventing and combating intense threats via new and emerging tactics and strategies, improved information technology and sharing, enhanced safety equipment and vehicles, and increased self-compliance with personal safety practices (e.g., seat belts, body armor).

There are also policies and practices that management should continue to explore and improve in order to enhance and promote safety, particularly those practices derived from scientific research (i.e., evidence-based policies and practices). For example, information on shift scheduling practices and pursuits can be useful in developing agency-specific practices and policies to improve safety, health, and wellness. In addition, the continued research and testing of various information technology modalities (iPhones, Mobile Data Terminals, Global Positioning Systems and Crime Mapping, and other similar tools) should be continued so that agencies can make technological upgrades that will remove officer distractions while also providing clear, concise, and relevant information in a dynamic fashion.

While peers can play a role in helping to protect other officers and provide support and compassion to their colleagues, the role of personal responsibility of officers should not be underestimated. The fact that any less than 100 % of officers do not use seat belts 100 % of the time or that less than 100 % of officers do not wear safety equipment (i.e., vests) 100 % of the time indicates some resistance to or lack of understanding of the data and information about the safety facts. Some officers continue to use excessive amounts of alcohol and caffeine, smoke, engage in poor eating habits (or are obese and fail to lose weight), fail to get adequate exercise or sleep, or overuse or even abuse controlled substances. Despite the known health risks, these behaviors/conditions can shorten lives and increase health and safety risks many of these behaviors and conditions are within the control of individual officers. When these behaviors begin to be recognized as reducing health, wellness, safety, or shortening one's life, there are opportunities to receive assistance through peers, professional counselors (e.g., psychologists, psychiatrists, social workers), sleep experts, nutritional counselors, and fitness advisors. At the same time, law enforcement agency leadership and culture must be supportive, that is, leadership and organizational structure, policies, and practices should work toward promoting safety, health, and wellness if their agencies are to be successful in promoting public safety.

Related Entries

- ▶ Critical Incidents
- ▶ Evidence-Based Policing
- ▶ History of Police Unions
- ▶ Methodological Issues in Evaluating Police Performance
- ▶ Minorities Within the Police Workforce
- ▶ Police Culture
- ▶ Police Performance Measurement
- ▶ Randomized Experiments in Criminology and Criminal Justice
- ▶ Risk Management in Policing
- ▶ Victims and the Police
- ▶ Women in Policing

Recommended Reading and References

Aamodt MG, Stalnaker N (2001) Police officer suicide: frequency and officer profiles. In: Shehan D, Warren J (eds) Law enforcement and suicide. Federal Bureau of Investigation, Quantico, pp 381–398

Amendola KL, Slipka MG, Hamilton EE, Soelberg M, Koval K (2011a) Trends in shift length: results of a random national survey of police agencies. Police Foundation Reports, Washington, DC

Amendola KL, Weisburd D, Hamilton EE, Jones G, Slipka M, Heitmann A, Shane J, Ortiz C, Tarkghen E (2011a) The shift length experiment: what we know about 8-, 10-, and 12-h shifts in policing. Police Foundation, Washington, DC

Artwohl A (2002) Perceptual and memory distortions during officer-involved shootings. FBI Law Enforcment Bulletin 71(10), US Department of Justice, Washington, DC

Beehr T, Ivanitskaya L, Glaser K, Erofeev D, Canali K (2004) Working in a violent environment: the accuracy of police officers' reports about shooting incidents. J Occup Organ Psychol 77:217–235

Colligan MJ, Frockt IJ, Tasto DL (1979) Frequency of sickness absence and worksite clinic visits among nurses as a function of shift. Appl Ergon 10(2):79–85

Czeisler CA, Moore-Ede MC, Coleman RH (1982) Rotating shift work schedules that disrupt sleep are improved by applying circadian principles. Science 217(4558):460–463

Edwards S, Maeder D (2011) BJA fact sheet: officer safety initiatives. US Department of Justice, Office of Justice Programs, Washington, DC, https://www.bja.gov/Publications/OfficerSafetyFS.pdf last accessed March 17, 2013

Evans BJ, Coman GJ (1993) General versus specific measures of occupational stress: an Australian police survey. Stress Med 9(1):11–20

Floyd C (2012) National law enforcement officers' memorial fund. (http://www.nleomf.org/facts/officer-fatalities-data/causes.html http://www.nleomf.org/facts/officer-fatalities-data/). Accessed 17 Mar 2013

Gold DR, Rogacz S, Bock N, Tosteson TD, Baum TM, Speizer FE, Czeisler CA (1992) Rotating shift work, sleep, and accidents related to sleepiness in hospital nurses. Am J Public Health 82(7):1011–1014. doi:10.2105/AJPH.82.7.1011

Hart PM, Wearing AJ, Heady B (1995) Assessing police work experiences: development of the police daily hassles and uplifts scales. J Crim Justice 21(6):553–572

Hoffman D (2010) Holder launches new law enforcement safety initiative (in press). (http://www.mainjustice.com/2010/10/26/holder-launches-new-law-enforcement-safety-initiative/). Accessed 17 Mar 2013

Klinger D (2006) Police responses to officer-involved shootings. National Institute of Justice, Washington, DC, NIJ Journal

Komarovskaya I, Maguen S, McCaslin SE, Metzler TJ, Madan A, Brown AD, Galatzer-Levy IR, Henn-Haase C, Marmar CR (2011) The impact of killing and injuring others on mental health symptoms among police officers. J Psychiatr Res 45(10): 1332–1336

Liberman AM, Best SR, Metzler TJ, Fagan JA, Weiss DS, Marmar CR (2002) Routine occupational stress and psychological distress in police. Policing 25:421–439

Loo R (2003) A meta-analysis of police suicide: rates, findings, and issues. Suicide Life Threat Behav 33:313–325

Maguen S, Metzler TJ, McCaslin S, Inslicht SS, Henn-Haase C, Neylan TC, Marmar CR (2009) Routine work environment stress and PTSD symptoms in police officers. J Nerv Ment Dis 197(10):754–760

McCreary DR, Thompson MM (2006) Development of two reliable and valid measures of stressors in policing: the operational and organizational police stress questionnaires. Int J Stress Manag 13(4):494–518

Miller L (2001) Selection and application guide to personal body armor (NIJ Guide 100-01). U.S. Department of Justice, The National Institute of Justice's National Law Enforcement and Corrections Technology Center, Washington, DC

National Highway Traffic Safety Administration (2011) (http://www.cops.usdoj.gov/pdf/OSWG/e091120401-OSWGReport.pdf). Accessed 17 Mar 2013.

Officer Safety and Wellness Working Group (2012). (http://www.cops.usdoj.gov/pdf/OSWG/Meeting_01-26-2012.pdf). Accessed 17 March 2013.

Quigley A (2008) Fit for duty? The need for physical fitness programs for law enforcement officers. Police Chief 75(6) June.

Rajaratnam SM, Barger LK, Lockley SW, Shea SA, Wang W, Landrigan CP, O'Brien CS, Qadri S, Sullivan JP, Cade BE, Epstein LJ, White DP, Czeisler CA (2011) Sleep disorders, health and safety in police officers. Harvard work hours, health and safety group. JAMA 306(23):2567–78

Taylor B, Kubu B, Kappleman K, Gunaratne H, Ballard N, Martinez M (2009) The BJA/PERF body armor national survey: protecting the nation's law enforcement officers (Phase II Final Report to BJA). Police Executive Research Forum, Washington, DC

Vena JE, Violanti JM, Marshall J, Fiedler RC (1986) Mortality of a municipal worker cohort: III. Police officers. Am J Ind Med 10(4):383–397. doi:10.1002/ajim.4700100406

Vila B (2000) Tired cops: the importance of managing police fatigue. Police Executive Research Forum, Washington, DC

Violanti JM (1983) Stress patterns in police work: a longitudinal study. J Police Sci Adm 11(2):211–216

Violanti JM, Aron F (1994) Ranking police stressors. Psychol Rep 75:824–826

Violanti JM, Burchfiel CM, Miller DB, Andrew ME, Dorn J, Wactawski-Wende J, Beighley CM, Pierino K, Joseph PN, Vena JE, Sharp DS, Trevisan M (2006) The Buffalo cardio-metabolic occupational stress (BCOPS) pilot study: methods and participant characteristics. Ann Epidemiol 16(2):148–156

Violanti JM (2010) Police suicide: a national comparison with fire-fighter and military personnel. Polic: Int J Police Strateg Manag 33(2):270–286

Violanti JM (2012) Shifts, extended work hours and fatigue: an assessment of health and personal risks for police officers. Final report of award # 2005-FS-BX-0004. National Institute of Justice, Washington, DC

Woodrow S (2011) Holder launches officer safety initiative (in press). (http://www.mainjustice.com/2011/03/22/holder-launches-officer-safety-initiative/). Accessed 17 Mar 2013

One Percent Clubs/Gangs

► Motorcycle Clubs or Gangs?

Onset of Offending

Delphine Theobald[1,2] and David P. Farrington[2]
[1]Institute of Psychiatry, King's College, London, UK
[2]Institute of Criminology, University of Cambridge, Cambridge, UK

Synonyms

Adolescence-limited offending; Age-crime curve; Criminal careers; Life-course-persistent offending; Prevention; Trajectories

Overview

The age of onset of offending is the age at which the first crime is committed. Generally, this is 3–5 years earlier in self-reports than in official records. The aggregate age-crime curve conceals several trajectories and pathways. In particular, life-course-persistent offenders, who start early and persist for a long time, are distinguished from adolescence-limited offenders, who start later and have short criminal careers. An early age of onset predicts a long criminal career. The most important risk factors for early onset are similar to risk factors for offending in general and include neuropsychological problems, low intelligence, and high impulsiveness, combined with poverty and poor child rearing. However, adult onset offenders are very different, because they tend to be nervous, withdrawn and have few friends in childhood. Effective programs that can prevent or delay onset include nurse home-visiting programs, parent training, skills training, and preschool intellectual enrichment programs.

Introduction

The age of onset is defined as the age at which the first offense is committed. “Offenses” are defined as the most common types of crimes that predominate in official criminal statistics, including theft, burglary, robbery, violence, vandalism, fraud, and drug use. Ages of onset are often studied within criminal career research. A “criminal career” is defined as the longitudinal sequence of offenses committed by an individual offender (Blumstein et al. 1986). It has a beginning (onset), an end (desistance), and a length in between (duration), during which offenders commit crimes at a certain rate (frequency per year).

Criminal career research falls within developmental and life-course criminology, which is concerned with the development of offending and antisocial behavior from the womb to the tomb, as well as with risk and protective factors and life events that influence development. Crimes usually do not appear without warning; they are commonly preceded by childhood antisocial behavior (such as bullying, lying, cheating, truanting, and cruelty to animals) and followed by adult antisocial behavior (such as excessive drinking, sexual promiscuity, spouse assault, child abuse, and neglect). However, the focus in this entry is on the onset of offending.

The key questions addressed in this entry are:

1. What is known about ages of onset from official records and self-reports?
2. How does the age of onset vary with gender and race?
3. What is known about developmental sequences of onsets?
4. To what extent does the age of onset predict later criminal career features?
5. What are the most important risk factors for early onset, adolescent onset, and adult onset?
6. What are the most important theories of onset?
7. How can the onset of offending be prevented?

What Is Known About Ages of Onset from Official Records and Self-Reports?

It is well known that the aggregate (official) crime rate increases with age to a peak in the teenage years and then decreases in the twenties. It is less well known that this curve reflects variations in the prevalence of offenders rather than in the frequency of offending by offenders. A review of knowledge about the age-crime curve by Farrington (1986) concluded that the peak age of onset of offending was usually between ages 8 and 14. In the Cambridge Study in Delinquent Development (CSDD), which is a prospective longitudinal survey of over 400 London males, the peak age of the first conviction was at 14, with 5 % of the males first convicted at that age. In the CSDD, 129 males (31.9 %) were first convicted between ages 10 and 20, but only 35 (8.7 %) between ages 21 and 40, and only 6 (1.5 %) after age 40 (up to age 56; see Farrington et al. 2013).

Rather than presenting the onset rate, taking all persons in a cohort still alive as the denominator, it might be better to present a “hazard” rate. This relates the number of first offenders to the number of persons still at risk of a first offense, excluding those with a previous onset. In the CSDD, the hazard rate showed a later peak than the onset rate, at age 17, when 6 % of the males still at risk were first convicted (Farrington et al. 1990). Basically, the peak hazard rate was later and greater than the peak onset rate because of the decreasing number of males who were still at risk of a first conviction with increasing age (the denominator).

Few studies have compared self-reported and official ages of onset. In a study of 470 Montreal delinquents, Le Blanc and Fréchette (1989) found that the males had an average age of onset of 10.8 years in self-reports and 14.6 years in official records. Similarly, in the Pittsburgh Youth Study, which is a follow-up of over 1,500 Pittsburgh boys, Loeber et al. (2003) discovered that the average age of onset for self-reported serious delinquency was 11.9 years, while the first court contact for an index (more serious) offense occurred at an average age of 14.5 years. In the Dunedin study in New Zealand, Moffitt et al. (2001, p. 83) compared the ages at first arrest, first conviction, and first self-reported offense, and found that the age of onset in official records was approximately 3–5 years after the age of onset in self-reports.

Similarly, in the CSDD, Kazemian and Farrington (2005) showed that the average age of the first self-reported offense was at 11.9, while the average age of the first conviction was at 16.9. They also found that retrospective self-reports of offending (at age 32) were inadequate, compared with more contemporaneous self-reports; 41 % of prospectively admitted crimes were not admitted retrospectively, and the average retrospective age of onset was 4 years later than the average prospective age of onset. Therefore, accurate information about ages of onset cannot be obtained by retrospective questioning.

Variations by Gender and Race

In general, the official age-crime curve for African Americans is a magnified version of the curve for Caucasians, with an earlier age of onset, a higher peak prevalence of offending, and a later age of desistance. For example, Parker and Morton (2009) analyzed the criminal careers of over 800 male South Carolina delinquents and found that 32 % of African Americans, compared with 13 % of Caucasians, had an early age of onset (before age 12). Low family income and low verbal intelligence were associated with early onset and were more common among the African American delinquents.

In the Dunedin study in New Zealand, Moffitt et al. (2001) argued that gender differences in the prevalence of offending were attributable to gender differences in the number or level of risk factors such as neuropsychological problems and conduct disorder. They found little evidence for an early onset group of persistent female offenders, because females generally had fewer risk factors and because female delinquency was less persistent than male delinquency. In the Christchurch Study in New Zealand, Fergusson and Horwood (2002, p. 174) also reported that fewer females than males followed an early onset pathway, with female chronic violent offending being particularly rare. However, relative to males, proportionally more females showed early onset adolescence-limited offending (see later), and a series of explanatory factors relating to family functioning and early adjustment operated in a similar fashion for both males and females. In a multisite study in the United States, Broidy et al. (2003) showed that trajectories for male and female offending were similar in number and pattern, but the male trajectories were more highly elevated than the female trajectories, because males engaged in more delinquent behavior than females overall.

What Is Known About Developmental Sequences of Onsets?

The age of onset varies with different types of offenses. It is desirable to study sequences of onsets, to investigate to what extent the onset of one type of offense is followed by the onset of another type. In the Montreal study, Le Blanc and Fréchette (1989) discovered that shoplifting and vandalism tended to occur before adolescence (average age of onset 11), burglary and motor vehicle theft in adolescence (average onset ages 14–15), and sex offenses and drug trafficking in the later teenage years (average onset ages 17–19). In a study of over 700 adult Massachusetts sex offenders, Danielle Harris (2013) reported that their average onset ages were 15.7 for burglary, 16.0 for theft, 17.3 for robbery, 18.3 for assault, 22.4 for rape, and 25.5 for child molesting. In general, nonsex offenses occurred 6.7 years before sex offenses.

In the CSDD, Kazemian and Farrington (2005) found a typical sequence from minor to serious self-reported offenses, with shoplifting and vandalism most frequently beginning before other types of offending. In contrast, for official offenses, they found that the males were likely to be convicted for theft of vehicles and burglary before any other offense. These findings suggest that there are different onset sequences for self-reported and official offending.

How can onset sequences be explained? First, it may be that different acts are different behavioral manifestations of the same underlying construct that persists at different ages. An underlying antisocial tendency, for example, may manifest as shoplifting in early adolescence, burglary in later adolescence, and intimate partner violence and child abuse in adulthood, with no facilitating effect of an earlier act on a later one. Second, different acts may be different behavioral manifestations of the same or similar underlying constructs at different ages but also part of a developmental sequence, where one act facilitates another act, for example, cigarette smoking facilitating later marijuana use. Third, different acts may be indicators of different constructs and part of a causal sequence, where changes in an indicator of one construct cause later changes in an indicator of a different construct. For example, neuropsychological deficits may cause subsequent school failure.

The first possibility can be distinguished empirically from the second and third. When acts in a sequence are all different behavioral manifestations of the same underlying construct, preventing or changing an early act in the sequence will not necessarily affect the probability of occurrence of later acts, unless there is some change in the underlying construct. However, with developmental and causal sequences, changing an early act in a sequence will affect the probability of occurrence of later acts. In general, it is not clear whether onset sequences reflect persistent heterogeneity or state dependence (Nagin and Farrington 1992).

Rolf Loeber (1988) pioneered the study of developmental pathways and outlined five features of a developmental progression: some behaviors had an onset at earlier ages than others, there was usually escalation in the seriousness of behaviors over time, there was usually retention of earlier behaviors in the sequence when a new behavior occurred, each behavior was best predicted by the developmentally adjacent behavior, and the ordering of the behaviors in the sequence was invariant, although it was noted that many people did not progress from one particular behavior to the next one in the sequence. He proposed that there were three different developmental pathways for antisocial behavior: aggressive versatile, nonaggressive antisocial, and exclusive substance abuse.

The aggressive versatile pathway was characterized by violent, property, and drug offenses; had an onset in the preschool years; and included conduct problems, hyperactivity/impulsivity and attention deficit, educational problems, and a high rate of innovation (the occurrence of new behaviors). The nonaggressive antisocial pathway, characterized by property and drug offenses, had a later onset in late childhood or early adolescence and included only nonaggressive conduct problems, no hyperactivity/impulsivity or inattention, and a low rate of innovation. Interestingly, this latter group of children typically were involved with deviant peers. The exclusive substance abuse pathway had the latest onset (in late adolescence) and no antecedent conduct problems, but it included antecedent internalizing problems such as shyness and nervousness. Surprisingly, these categories bear a great deal of resemblance to three groups that are considered in more detail later: early onset, adolescence-limited, and adult onset offenders.

Loeber (1988) concluded that one of the factors that predicted whether there would be a transition from one behavior to the next in a pathway was the frequency of occurrence of the first behavior. This suggests that, when individuals commit trivial offenses more frequently, their behavior is likely to escalate to more serious offenses in the future. It is clear that certain types of offenses occur on average before other types and that onset sequences can be identified. There has been extensive research on Loeber's pathways, which are now termed overt, covert, and authority conflict pathways (see Farrington and Loeber 2013). The overt pathway starts with minor aggression and then progresses to physical fighting and eventually serious violence; the covert pathway starts with shoplifting and frequent lying and then progresses to vandalism and eventually serious delinquency (burglary and vehicle theft), and the authority conflict pathway starts with stubborn behavior and then progresses to defiance and eventually truancy and running away from home.

Age of Onset as a Predictor of Later Criminal Career Features

There is no doubt that an early age of onset predicts a long criminal career and many offenses. For example, in the CSDD, the boys who were first convicted at ages 10–13 had an average of 9.2 convictions and an average criminal career duration of 14.2 years, those first convicted at ages 14–16 had 6.1 convictions and a duration of 14.4 years, those first convicted at ages 17–20 had 2.7 convictions and a duration of 6.7 years, and those first convicted at ages 21–30 had 2.0 convictions and a duration of 3.8 years (up to age 56; see Farrington et al. 2013).

Less is known about whether an early onset predicts a higher rate of offending per year or whether particular types of crimes committed on the first occasion presage a long and serious criminal career. In Sweden, Robert Svensson (2002) found that theft-related offenses committed on the first occasion were most likely to be followed by long criminal careers, whereas in Australia, Mazerolle et al. (2010) concluded that a first offense of violence did not predict a distinctive criminal career (compared to a nonviolent first offense).

Terrie Moffitt (1993) distinguished between adolescence-limited (AL) and life-course-persistent (LCP) offenders. LCP offenders tend to start offending early (usually before age 13), whereas AL offenders tend to start in adolescence (at about ages 13–16). LCP offenders are relatively few in number (approximately 5–10 % of the offender population) and tend to be difficult children, often showing signs of conduct problems very early in childhood. AL offenders are highly influenced by their peers. LCP offenders have long criminal careers, whereas AL offenders often give up offending around or soon after age 20. LCP offenders are often very versatile offenders involved in both predatory and violent crime. Their offending trajectory deviates significantly from the aggregate age-crime curve, because it does not decrease dramatically after age 20. The peak in the aggregate age-crime curve in the teenage years is largely caused by AL offenders joining in with LCP offenders.

Much of the developmental research on the age-crime curve focuses on underlying trajectories of antisocial behavior and offending. There is now fairly strong empirical evidence to suggest that there are categories of offenders with trajectories that are partly concordant with Terrie Moffitt's (1993) taxonomy. Many studies have identified the two distinct offender profiles (Piquero 2008), but research has also revealed other unpredicted categories, including low-level chronics who offend at a low and stable rate across a long period of time and short-term high-rate offenders who start on a life-course-persistent path but then desist and do not show life-course-persistent characteristics by offending in adulthood (Piquero et al. 2010).

Nagin and Odgers (2010) found that four trajectories fitted the CSDD data and that there were important differences between people in each category. Other than a large nondelinquent group, these authors identified a high-rate chronic trajectory which contained the highest proportion of cases with low IQ, poor parenting, high risk taking, and parental criminality, and a low-rate chronic trajectory with fewer cases of low IQ, poor parenting, high risk taking, and parental criminality. Those in the adolescence-limited trajectory began engaging in delinquency later and ended their delinquency earlier than participants in the two chronic trajectories. A key issue is whether these trajectories differ in degree or in kind. The number of trajectories identified will be dependent on the type and size of the sample, the statistical techniques used, and the length of the follow-up period (Farrington et al. 2013).

Risk Factors for Early Onset

A number of studies have found a relationship between poor neuropsychological development and early onset life-course-persistent criminal behavior (e.g., Nagin et al. 1995; Piquero 2001). Neuropsychological deficits include delayed cognitive and motor development and poor verbal comprehension and expression. Children with neuropsychological problems are often difficult to bring up, and their parents may be those who are least able to manage a difficult child. These vulnerable children are disproportionally born into unfavorable conditions because the characteristics of parents and children, such as cognitive abilities, temperament, and personality, tend to be correlated and to some extent inherited (Moffitt 1994, p. 19). The consequences of negative interactions between the child and the parent may reinforce the problem. The impact of biological or social risk factors is likely to be affected by preexisting vulnerabilities and the timing and additive or interactional effects of other factors. Vulnerability to neuropsychological problems is probably caused by such factors as heritability, low birth weight, early brain injury, and birth complications (Piquero 2001).

Early results from the Dunedin study in New Zealand showed that young boys with both low neuropsychological test scores and adverse home environments had aggression scores more than four times greater than those of boys with only neuropsychological problems or only adverse homes. Children with poor self-control and aggressive behavior often experience rejection by adults and peers and consequently do not learn prosocial behaviors. In turn, poor socialization can cause problems at school and can result in truancy and school exclusion so that opportunities to learn basic academic and social skills are lost, and these skills become increasingly difficult to attain.

There have been very few systematic comparisons of risk factors for early versus later onset. Most research on onset compares persons who have an early onset with those who do not have an early onset (including non-offenders). Therefore, the results may reflect differences between offenders and non-offenders rather than differences between early and later onset offenders. In the CSDD, Farrington and Hawkins (1991) compared childhood risk factors for early onset (first convicted at ages 10–13) with later adolescent onset (first convicted at ages 14–20). They found that early onset was most strongly predicted by boys rarely spending leisure time with their father, high troublesomeness, authoritarian parents, and psychomotor impulsivity.

Some researchers have suggested that early onset is a behavioral manifestation of a high criminal propensity rather than a unique offender category or, in other words, that differences in onset are quantitative and continuous rather than qualitative and discontinuous. Gottfredson and Hirschi (1990) argued that the early onset of antisocial behavior was attributable to the underlying construct of low self-control that was established early in childhood because of inadequate parental management and poor socialization. They further proposed that the ordering of individuals on this trait of low self-control was stable over time and that those with the lowest levels of self-control tended to have the earliest ages of onset and high amounts of versatile criminal activity, and they desisted later than individuals with higher levels of self-control. Low self-control influences an individual's ability to delay gratification and causes criminal acts, antisocial behaviors such as promiscuous sex and alcohol and drug abuse, and substantial versatility in criminal and antisocial behavior in a range of situations and opportunities.

Risk Factors for Adolescent Onset

In contrast to those offenders who have an early onset, there is a much larger group who only offend as adolescents over a relatively short period of time. Unlike LCP offenders, these AL offenders are not handicapped by neuropsychological deficits, family problems, or socially disadvantaged backgrounds. In contrast, AL offending is often a response to changes experienced in adolescence, such as physical changes around puberty and the increasing influence of peers rather than parents. Terrie Moffitt (1993) suggested that AL offenders began offending because of a perceived "maturity gap": their inability to achieve adult goals legitimately because of their age. Delinquent acts that symbolize adult social status are common, including smoking, drinking alcohol, drug use, and promiscuous sexual behavior, demonstrating that they have left childhood behind.

Most criminal careers of AL offenders are short-lived, ending in early adulthood.

Because they have healthy neuropsychological and social development, AL offenders retain the social and academic skills to live a conventional life unless they suffer "snares," such as early disengagement from social institutions such as school and family and involvement with the criminal justice system. Breaking away from parents may occur sooner rather than later for some children, who may choose to engage with delinquent peers because of family conflict or breakdown.

Risk Factors for Adult Onset

Adult onset offenders often constitute approximately half of all adult offenders in longitudinal official data (Eggleston and Laub 2002; Zara and Farrington 2009). Terrie Moffitt (1993) denied that there were any true adult onset offenders; she thought that these offenders had previously offended but had not been caught. McGee and Farrington (2010), using data from the CSDD, found that all adult onset offenders (with a first conviction at age 21 or later) had self-reported some previous offending in childhood and adolescence, but generally for minor types of offenses with relatively low conviction rates. They concluded that about one-third of adult onset offenders were "true" self-reported delinquents who could have been convicted because of the frequency and seriousness of their offending. The adult onset offenders were disproportionally likely to commit sex offenses, theft from work, vandalism, and fraud.

It has been suggested that adult onset offending may be caused by changes in adult circumstances such as the loss of a job or a relationship or, in other words, the reduction in bonding to society (Sampson and Laub 2005). Some researchers have found that adult onset offenders are more likely than others to have mental illness (Elander et al. 2000), and findings from the CSDD suggest that these individuals are nervous, withdrawn, and had few friends in childhood (Zara and Farrington 2009). Possibly, their inhibition, nervousness, and lack of peer influence in the teenage years protected them from teenage offending but not from adult offending. Zara and Farrington (2006) systematically compared risk factors for adult onset (first convicted at age 21 or later) and earlier onset (first convicted at ages 10–20). The adult onset offenders were more nervous, less daring, and less troublesome in childhood (ages 8–10), less likely to have early sex by age 14, and less likely to truant at ages 12–14. The adult onset offenders were similar to non-offenders in their self-reported delinquency at age 14 and much less delinquent than the earlier onset offenders.

Theories of the Onset of Offending

Not surprisingly, the major theories that try to explain why people start offending, sequences of onsets, and the relation between onset and duration are the developmental and life-course theories. The important theory of Terrie Moffitt (1993) has already been discussed in some detail. Unlike Moffitt, Thornberry and Krohn (2005) did not propose that there were types of offenders, but they did suggest that different factors influenced onset at different ages. At the earliest ages (birth to 6), the most important factors were neuropsychological deficit and difficult temperament (e.g., impulsiveness, negative emotionality, fearlessness, poor emotion regulation), parenting deficits (e.g., poor monitoring, low affective ties, inconsistent discipline, physical punishment), and structural adversity (e.g., poverty, unemployment, welfare dependency, disorganized neighborhood). They also suggested that structural adversity might cause poor parenting.

According to Thornberry and Krohn, neuropsychological deficits are less important for children who start antisocial behavior at older ages. At ages 6–12, neighborhood and family factors are particularly salient, while at ages 12–18, school and peer factors predominate. Thornberry and Krohn also suggested that deviant opportunities, gangs, and deviant social networks were important for onset at ages 12–18. They proposed that late starters (ages 18–25) had cognitive deficits such as low intelligence and poor school performance but that they were protected ("cocooned") from antisocial behavior at earlier ages by a supportive family and school environment. At ages 18–25, when they became more independent, they found it hard to make a successful transition to adult roles such as employment and marriage. The theory did not postulate a single key construct underlying offending but suggested that children who started early tended to continue because of the persistence of neuropsychological and parenting deficits and structural adversity. Interestingly, Thornberry and Krohn (2005) predicted that late starters (ages 18–25) would show more continuity over time than earlier starters (ages 12–18) because the late starters had more cognitive deficits.

The key construct in Sampson and Laub's (2005) theory is age-graded informal social control, which means the strength of bonding to family, peers, schools, and later adult social institutions such as marriages and jobs. Sampson and Laub primarily aimed to explain why people did not commit offenses on the assumption that why people want to offend is unproblematic (presumably caused by hedonistic desires) and that offending is inhibited by the strength of bonding to society. The strength of bonding depended on attachments to parents, schools, delinquent friends, and delinquent siblings, and also on parental socialization processes such as discipline and supervision. Structural background variables (e.g., social class, ethnicity, large family size, criminal parents, disrupted families) and individual difference factors (e.g., low intelligence, difficult temperament, early conduct disorder) had indirect effects on offending through their effects on informal social control (attachment and socialization processes).

Sampson and Laub were concerned with the whole life course. They emphasized change over time rather than consistency and the poor ability of early childhood risk factors to predict later life outcomes. They assumed that early and later onset (and other criminal career features) depended on the strength of bonding. They focused on the importance of later life events (adult turning points), such as joining the military, getting a stable job, and getting married, in fostering desistance and "knifing off" the past from the present. Because of their emphasis on change and unpredictability, they denied the importance of types of offenders such as "life-course persisters."

The integrated cognitive antisocial potential (ICAP) theory of Farrington (2005) is mainly intended to explain offending by lower class males. No distinct types of offenders are proposed. The key construct underlying antisocial behavior is antisocial potential (AP), and there is continuity in offending and antisocial behavior over time because of consistency in the relative ordering of people on AP. Therefore, early and later onset and the relation between early onset and later criminal careers depend on levels of AP.

Long-term and short-term influences on AP are explicitly distinguished. Long-term factors encouraging offending include impulsiveness, strain, and antisocial models, while short-term (immediate situational) influences include opportunities and victims. Long-term factors inhibiting offending include attachment and socialization (based on social learning) and life events such as getting married or moving house. The theory explicitly aims to explain both the development of offending and the commission of offenses. Situational factors, motives, and cognitive (thinking and decision-making) processes are included. The theory also assumes that the consequences of offending have labeling, deterrent, or learning effects on AP.

Prevention Programs to Reduce Onset

It is particularly important to prevent the early onset of offending, since this tends to lead to LCP offending. The most effective programs are those that target key risk factors such as poor parenting, high impulsiveness, poor social skills, or school failure (Farrington and Welsh 2007). These programs include general parent education in the home, parent management training, child skills training, and preschool intellectual enrichment programs. However, evaluators have generally reported effects only on offending, not on specific features of criminal careers such as the age of onset or the duration of offending.

The best-known home-visiting program is the Nurse Family Partnership program of Olds et al. (1998). This was evaluated in a randomized trial of 400 women considered to be at high risk, such as teenage mothers, unmarried women, and those living in poverty. The results showed that the prenatal and postnatal home visits caused a significant decrease in recorded child physical abuse and neglect during the first 2 years of life, especially for those mothers who were at high risk; 4 % of visited versus 19 % of non-visited mothers at risk were guilty of child abuse or neglect. This is was an important finding, because those children who are physically abused or neglected are more likely to become violent offenders later in life. The children of visited mothers were less likely to offend later in life than the children of non-visited mothers.

A systematic review carried out by Piquero et al. (2009) concluded that early family or parent training is an effective intervention for reducing behavior problems among young children. For example, Scott et al. (2001) evaluated the "Incredible Years" program in London with 140 children aged 3–8 referred for antisocial behavior, and Gardner et al. (2006) evaluated this program in Oxfordshire with 70 children aged 2–9 referred for conduct problems. This type of research on parent training shows that it leads to improvements in child disruptive behaviors and reductions in delinquency.

In Montreal, Tremblay et al. (1995) evaluated a program that combined child skills training and parent training. Over 300 aggressive or hyperactive boys were randomly allocated at age 6 either to receive the program or not. The program reduced early offending. By age 12, the experimental boys committed less burglary and theft, were less likely to get drunk, and were less likely to be involved in fights than the controls (according to self-reports). At every age from 10 to 15, the experimental boys had lower self-reported delinquency scores than the control boys.

Preschool programs, focusing on intellectual stimulation and increasing thinking and reasoning abilities in order to increase later school achievement, have also been successful. The most famous program is the Perry project which was originally evaluated with 123 disadvantaged African American children aged 3–4. This led to reduced arrests up to age 19, and the benefits continued throughout life. A follow-up at age 40 showed that the experimental participants had significantly fewer lifetime arrests for violent crimes, property crimes, and drug crimes (Schweinhart et al. 2005). Improvements were also found in graduation from school, employment records, and annual incomes for those in the program group compared to the controls. At age 40, a benefit-cost analysis showed that the Perry project produced $17 of benefit per $1 of cost, with 76 % of this benefit being returned to the general public in the form of savings in crime, education, and welfare and increased tax revenue, and 24 % of the benefits accruing to the experimental participant.

Conclusions

The age of onset of offending is earlier in self-reports (about 10–12) than in official records (about 14–15). Because the age-crime curve is higher and wider for males and African Americans than for females and Caucasians, proportionally fewer females and Caucasians have an early onset of offending. Generally, shoplifting and vandalism have an early onset, while violence, fraud, and sex offenses have a later onset. The aggregate age-crime curve conceals different trajectories and developmental pathways. An early age of onset predicts a long criminal career, and the most famous theory (by Moffitt) that is relevant to onset distinguishes between adolescence-limited and life-course-persistent offenders. Risk factors for early onset are similar to risk factors for offending in general (and especially for serious or persistent offending), but risk factors for later adult onset are very different. There are a number of effective prevention programs, but their specific impact on ages of onset (as opposed to offending in general) is not known. Knowledge and theories about onset should be very useful in developing programs to prevent or delay onset.

Related Entries

- ▶ Age-Crime Curve
- ▶ Criminal Careers and Public Policy Responses

- ▶ Desistance from Crime
- ▶ General Theory of Crime
- ▶ Group-Based Trajectory Models
- ▶ Moffitt's Developmental Taxonomy of Antisocial Behavior

Recommended Reading and References

Blumstein A, Cohen J, Roth A, Visher CA (1986) Criminal careers and "career criminals", 2 vol. National Academy Press, Washington, DC

Broidy LM, Nagin DS, Tremblay RE, Bates JE, Brame B, Dodge KA, Vitaro F (2003) Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: a six-site, cross-national study. Dev Psychol 39:222–245

Eggleston EP, Laub JH (2002) The onset of adult offending: a neglected dimension of the criminal career. J Crim Justice 30:603–622

Elander J, Rutter M, Simonoff E, Pickles A (2000) Explanations for apparent late onset criminality in a high-risk sample of children followed up in adult life. Br J Criminol 40:497–509

Farrington DP (1986) Age and crime. In: Tonry M, Morris N (eds) Crime and justice, vol 7. University of Chicago Press, Chicago, pp 189–250

Farrington DP (2005) The integrated cognitive antisocial potential (ICAP) theory. In: Farrington DP (ed) Integrated developmental and life-course theories of offending. Transaction, New Brunswick, pp 73–92

Farrington DP, Hawkins JD (1991) Predicting participation, early onset, and later persistence in officially recorded offending. Crim Behav Ment Health 1:1–33

Farrington DP, Loeber R (2013) Two approaches to developmental/life-course theorizing. In: Cullen FT, Wilcox P (eds) The Oxford handbook of criminological theory. Oxford University Press, Oxford (in press)

Farrington DP, Welsh BC (2007) Saving children from a life of crime: early risk factors and effective interventions. Oxford University Press, Oxford

Farrington DP, Loeber R, Elliott DS, Hawkins JD, Kandel DB, Klein MW, McCord J, Rowe DC, Tremblay RE (1990) Advancing knowledge about the onset of delinquency and crime. In: Lahey BB, Kazdin AE (eds) Advances in clinical and child psychology, vol 13. Plenum, New York, pp 283–342

Farrington DP, Piquero AR, Jennings WG (2013) Offending from childhood to late middle age: recent results from the Cambridge study in delinquent development. Springer, New York

Fergusson DM, Horwood JL (2002) Male and female offending trajectories. Dev Psychopathol 14:159–177

Gardner F, Burton J, Klimes I (2006) Randomized controlled trial of a parenting intervention in the voluntary sector for reducing child conduct problems: outcomes and mechanisms of change. J Child Psychol Psychiatry 47:1123–1132

Gottfredson M, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Harris DA (2013) Age and type of onset of offending: results from a sample of male sexual offenders referred for civil commitment. Int J Offender Ther Comp Criminol (in press)

Kazemian L, Farrington DP (2005) Comparing the validity of prospective, retrospective, and official onset for different offending categories. J Quant Criminol 21:127–147

Le Blanc M, Fréchette M (1989) Male criminal activity from childhood through youth. Springer, New York

Loeber R (1988) Natural histories of conduct problems, delinquency, and associated substance use: evidence for developmental progressions. In: Lahey BB, Kazdin AE (eds) Advances in clinical child psychology, vol 11. Plenum, New York, pp 73–124

Loeber R, Farrington DP, Petechuk D (2003) Child delinquency: early intervention and prevention. Office of Juvenile Justice and Delinquency Prevention, Washington, DC

Mazerolle P, Piquero AR, Brame R (2010) Violent onset offenders: do initial experiences shape criminal career dimensions? Int Crim Justice Rev 20:132–146

McGee TR, Farrington DP (2010) Are there any true adult-onset offenders? Br J Criminol 50:530–549

Moffitt TE (1993) Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. Psychol Rev 100:674–701

Moffitt TE (1994) Natural histories of delinquency. In: Weitekamp EGM, Kerner HJ (eds) Cross-national longitudinal research on human development and criminal behavior. Kluwer, Dordrecht, pp 3–61

Moffitt TE, Caspi A, Rutter M, Silva PA (2001) Sex differences in antisocial behaviour. Cambridge University Press, Cambridge

Nagin DS, Farrington DP (1992) The onset and persistence of offending. Criminology 32:501–523

Nagin DS, Odgers CL (2010) Group-based trajectory modeling (nearly) two decades later. J Quant Criminol 26:445–453

Nagin DS, Farrington DP, Moffitt TE (1995) Life-course trajectories of different types of offenders. Criminology 33:111–139

Olds DL, Henderson CR, Cole R, Eckenrode J, Kitzman H, Luckey D, Pettitt L, Sidora K, Morris P, Powers J (1998) Long-term effects of nurse home visitation on children's criminal and antisocial behavior: 15-year follow-up of a randomized controlled trial. JAMA 280:1238–1244

Parker JS, Morton TL (2009) Distinguishing between early and late onset delinquents: race, income, verbal intelligence, and impulsivity. N Am J Psychol 11:273–284

Piquero AR (2001) Testing Moffitt's neuropsychological variation hypothesis for the prediction of life-course persistent offending. Psychol Crime Law 7:193–215

Piquero AR (2008) Taking stock of developmental trajectories of criminal activity over the life course. In: Liberman A (ed) The long view of crime: a synthesis of longitudinal research. Springer, New York, pp 23–78

Piquero AR, Farrington DP, Welsh BC, Tremblay RE, Jennings WG (2009) Effects of early family/parent training programs on antisocial behavior and delinquency. J Exp Criminol 5:83–120

Piquero AR, Sullivan CJ, Farrington DP (2010) Assessing differences among short term high rate offenders compared to long term low rate offenders. Crim Justice Behav 37:1309–1329

Sampson RJ, Laub JH (2005) A general age-graded theory of crime: lessons learned and the future of life-course criminology. In: Farrington DP (ed) Integrated developmental and life-course theories of offending. Transaction, New Brunswick, pp 165–181

Schweinhart LJ, Montie J, Zongping X, Barnett WS, Belfield CR, Nores M (2005) Lifetime effects: the high/scope Perry preschool study through age 40. High/Scope Press, Ypsilanti

Scott S, Spender Q, Doolan M, Jacobs B, Aspland H (2001) Multicentre controlled trial of parenting groups for child antisocial behaviour in clinical practice. Br Med J 323:194–196

Svensson R (2002) Strategic offences in the criminal career context. Br J Criminol 42:395–411

Thornberry TP, Krohn MD (2005) Applying interactional theory to the explanation of continuity and change in antisocial behavior. In: Farrington DP (ed) Integrated developmental and life-course theories of offending. Transaction, New Brunswick, pp 183–209

Tremblay RE, Pagani-Kurtz L, Masse LC, Vitaro F, Pihl RO (1995) A bimodal preventive intervention for disruptive kindergarten boys: its impact through mid-adolescence. J Consult Clin Psychol 63:560–568

Zara G, Farrington DP (2006) Later criminal careers: psychological influences. In: Obergell-Fuchs J, Brandenstein M (eds) Nationale und internationale entwicklungen in der kriminologie (National and international developments in criminology). Verlag fur Polizeiwissenschaft, Frankfurt, pp 109–136

Zara G, Farrington DP (2009) Childhood and adolescent predictors of late-onset criminal careers. J Youth Adolesc 38:287–300

Open Drug Market

▶ Spatial Perspectives on Illegal Drug Markets

Opioid Dependence

▶ Drug Abuse and Alcohol Dependence Among Inmates

Opioid Replacement Treatment

▶ Drug Substitution Programs and Offending

Opioid Substitution Treatment

▶ Drug Substitution Programs and Offending

Optimizing Longitudinal Studies in Offending

Magda Stouthamer-Loeber
Department of Psychiatry, Medical School,
University of Pittsburgh, Pittsburgh, PA, USA

Overview

Three aspects of potential future longitudinal studies will be discussed: (1) good practices for the execution of a study and the changes that have taken place over time; (2) opportunities to enrich the field of enquiry, thereby enabling new questions to be addressed; and (3) optimizing the use of a data set, the collection of which most likely has been very expensive.

Introduction

Longitudinal studies, with their repeated measurements spanning decades, by their nature are very expensive and should be used only for those research questions that require the study of within-individual changes over time. Thus, they should only be used when a developmental approach to offending will yield information for questions that cannot be answered by cross-sectional studies. Two topics will be discussed: the execution of longitudinal studies and increasing the yield of these studies.

Execution of Longitudinal Studies

One of the problems in executing longitudinal studies is that generally the management skills and practical experience needed to conduct such studies are rarely taught during graduate training in universities and need to be acquired by researchers by mentors or through books (Stouthamer-Loeber and van Kammen 1995). Moreover, information on how to run such studies is rarely found in articles published in scientific journals. Although we can now learn from a fair number of longitudinal studies dealing with antisocial and delinquent behavior (summarized in Loeber et al. 2008, p. 20; Killias et al. 2012), it is of great help to have a mentor who has hands-on experience and can help prevent errors.

Investigators need to be involved in all matters of data collection and management and not count on staff to run a study for them. Before researchers can withdraw behind their desks to write papers, they have to be entirely assured that the data are as good as they can possibly be. Involvement in the daily running of a study is not a distraction, but a necessity.

The optimization of the execution of a longitudinal study starts in the planning phase. Running a longitudinal study is not something one can start to think about when the grant has been awarded. The aims, budget, and plans for data collection need to be coherent. Questions that need to be answered ahead of time are the following: Can the aims be measured by the proposed data collection? Is the data collection designed to capture developmental changes? Is the sample large enough and the sample acquisition realistic? Is the timeline realistic? What efforts will be made to continue to follow up the sample? And does the budget fit all the tasks to be accomplished? Also, thought needs to be given to how the potential study results may have a social impact. Although it may be necessary to make adjustments after the project has been funded due to budget cuts, the basic framework should have been worked out in detail. This makes it also easier to discuss with a granting agency the effects of a cut and possibly argue for a lesser cut.

Longitudinal studies have changed in method of data collection and storage. In addition, issues of consent and confidentiality as well as research conditions have changed as well.

Method of Data Collection. One of the questions confronting the researcher early on is whether to contract out the sample selection and data collection or whether to keep these tasks "in-house." There are pros and cons for either decision. If the research team has no specialized knowledge of sample selection and data collection, it may be easier to use a survey organization. Also, if data collection needs to take place across the country, it may make sense to use an existing network of interviewers affiliated with a survey organization. In addition, some researchers may not have the inclination or the time to be involved in data collection and prefer to receive the data "ready-made." There are also several arguments that can be made against the use of a survey organization. First, the expense is generally larger than for in-house data collection. More important, however, is that the researchers do not have direct control over the number of contact attempts with potential participants and the quality and completeness of the data and how the participants have been treated. Survey organizations generally do their own quality control which leaves the researchers to take the quality of the final product on faith. It would be better if researchers would take some of the quality control in their own hands to ensure that the product they receive is as good as it should be. It is not wise to be just a passive recipient of a data set.

If data collection is done in-house, the greatest care should be taken in the selection and training of interviewers. Apart from finding the right persons, it is very important to be up-to-date on the current hiring and firing regulations. These regulations change from time to time and a lot of unpleasantness can be encountered by not being informed. Once the staff is on board, training needs to take place. It is advisable to have an actual written training manual which includes job responsibilities, expectations, contact with participants, and training materials for data collection. A written manual is a help for the interviewers who can consult it to refresh some information. It is also important if interviewers are trained in batches to have the same training materials. In addition, a manual can serve the purpose of evidence that rules and regulations have been laid out.

Before interviewers are allowed to do “real” interviews, they need to be thoroughly familiar with the procedures and the interview. None of the interviews done with participants should ever be used as practice or training interviews. A study is only worthwhile if the data are as close to perfect as possible. There are enough circumstances (over which a researcher has little control) that influence the quality of the data, but not enough interviewer preparation should not be one of them. Although this description generally deals with data collection by interviewing participants, similar procedures can be developed for staffs who collect and code existing data.

Once interviewers have been trained, it is helpful to have a system in place that provides the supervisor with information about quality and quantity of work done by each interviewer. This should also include information about search and contact efforts as well as address changes which can be used in the next assessment phase. If necessary, retraining can take place where procedures need to be clarified. Also errors can often be fixed if noticed early enough but not at the stage when the researcher requires the data for analyses. The information can also be used to give staff members positive feedback and even problems can be turned into learning opportunities. It is useful for interviewers to know that all aspects of their work are important enough for the researchers to be informed about on an ongoing basis. Databases should be created in which the relevant information is deposited so that reports can be created on a regular basis rather than ad hoc when one feels something is not going right.

If interviewers are working in the field, it is a good idea to have a weekly session with their supervisor. During that session, completed materials can be handed in or downloaded and problems can be discussed, including errors that still can be fixed. The supervisor can brainstorm with the interviewer about search strategies for difficult to find participants and how to deal with reluctant participants. It may be useful to convene all interviewers to share solutions to particular problems and to learn from each other’s techniques.

Various changes have taken place over the past years in the form and content of data collection. Data collection has largely moved away from unstructured interviews which required extensive training and coding afterward to interviewing with structured answer formats. The advantage of the latter method is that the questions are more likely to be asked in a uniform manner of each participant and that there is less coding to be done after the interview is completed.

Other changes have to do with the recording of the interview. Most surveys now use electronic data collection, avoiding data entry (Groves et al. 2009). An additional change has been that most studies now use multiple rather than single informants to ensure a more complete picture of the participant and his/her environment. The frequency of repeated measures has increased, and studies are continuing to cover longer periods of time, allowing a larger portion of development of offending to be studied and to move from studying persistence to also taking into account desistance. In these studies desistance from offending can be better estimated if the time window for desistance is sufficiently large to measure stable desistance. These studies also allow for the detection of cases with a late onset.

Although face-to-face interviewing will always have the advantage of ensuring a person’s identity, the ease and the reduction in cost of web-based data collection will increase over time when more people have access to online electronic devices (e.g., Celio et al. 2011; Gosling et al. 2004). Such devices do not necessarily have to be a computer but may be a phone as long as it can connect to a survey program which can provide the questions and can store the data. Various devices may also collect motion or sleep data or can alert the participant to provide some specific information (Anokwa et al. 2009). Web-based data collection makes it easier for participants to continue to participate in a study from practically all over the world. However, web-assisted interviewing still has at least three problems: getting the participant to do it, completing the survey in a confidential manner, and ascertaining the identity of the participant. Thus, face-to-face interviewing, even though it is costly, may still be the most certain way of collecting data.

Data Management. The term "audit worthiness" stands for how well data are handled, cleaned, and analyzed (Freeland and Carney 1992). Errors that may be found in an audit may not be evidence of intentional deception, but more likely the result of carelessness and/or lack of documentation. Any error reduces the usefulness of the data set, and proper data management is of the utmost importance (Stouthamer-Loeber 1993). Magnusson and Bergman (1990) lay out a set of standards on how to manage data.

With regard to data storage, we have gone from mainframe storage with backup tapes that seem to go out of date continuously to in-office servers or just desktop computers. Further changes in data storage are likely to take place. Cloud storage and computing will make it possible to use one's desktop computer just as a typing device with all the software and data stored somewhere else (Armbrust et al. 2010). Although security problems still have to be worked out, this may also be a way to make data available to certified users. Sharing data collected in large, expensive longitudinal studies is now often a requirement for obtaining funds in the first place. Cloud computing may be a way to deal with the technical, if not the confidential, end of data sharing.

Confidentiality. Robins (1976) remarked on the change that the informed consent procedures had brought to research in the United Sates. Databases that in the past could have been used to select potential participants became off bounds because prior consent was required. The more recent HIPAA law, enacted in 1996, has further restricted access to health and health-care information. Consent forms are now complicated documents. The consent and confidentiality requirements have become stricter over the years and researchers, at least in the United States, have to go to great length to explain procedures and reasons for data collection and how the data will be kept confidential. The proper execution of obtaining consent cannot be stressed strongly enough and should covered thoroughly during training. Consent forms are legal documents and no short cuts can be taken. In addition, all staff should sign a statement that assures the confidentiality of the data and identity of the participants.

Research Conditions. The increased use of answering telephone machines and the availability of the caller number has impeded the contacting of participants. Also the extensive use of cell phones makes contacting more difficult since phone numbers are not listed and cell phone numbers tend to change more rapidly than landlines. On the other hand ways of electronic searching for people have multiplied. For instance, many states have extensive databases online of incarcerated individuals. Different search and contact methods are described in Cotter et al. (2002); Haggerty et al. (2007), and Stouthamer-Loeber and van Kammen (1995).

The success of longitudinal studies depends very much on the sustained cooperation of the participants. In order for a participant to cooperate, the participant first needs to be found, contacted, and agree to another round. Although there are well-developed methods of imputation of missing data, they work best for analyzing group data (McCartney et al. 2006). However, longitudinal studies also allow the study of the developmental paths of individual participants over time, and actual data is far to be preferred over imputation. In many studies, the group of most interest (e.g., families of conduct disordered boys) requires the most attempts to find, contact, and interview (see Cotter et al. 2005; Haggerty et al. 2007). Curtailing search and contact attempts does in general yield a biased subject loss. The cost of searching for participants generally rises over time because of moves. Even though most longitudinal studies ask the participants to provide some names, addresses, and telephone numbers of people who would always know their whereabouts, this information is not always helpful because the contacts themselves may have moved or may have lost track of the participant. Therefore, it is important that the budget reflects the costs of extensive searching for participants.

Longitudinal surveys, by their nature, cannot provide extensive feedback to the participants or provide help or treatment. Thus, in general, a participant may not directly benefit from participating. The reason why people do participate in these studies is because they are curious, they are made to feel special, they are civic-minded, and/or they are paid for their time. Further factors that may influence participation are the time it takes to participate, the ease of participation, and their earlier experience in being interviewed. Since longitudinal studies are expensive and the topic of antisocial behavior and crime may be in or out of favor with funding agencies over time, it is not possible to anticipate if funds available for existing or planned longitudinal studies may fluctuate over time.

Optimizing the Content of Longitudinal Studies

Whereas the older longitudinal studies relied solely on official records, later studies also used self-reports. Considering the low proportion of crimes that are caught, the addition of self-reports is crucial. More recent studies also try to include community and socioeconomic data as well as reports of routine activities and situational information surrounding the commission of crime. Some biological information has been collected in the past such as heart rate and skin conductance; however, lately the range of biological information has increased by the collection of data on hormones, DNA, and brain activity (fMRI). The combination of individual, environmental, and biological data will make it possible to endeavor to answer complicated questions about interactions within and between these three data domains.

Content. Expansion in the array of factors included in longitudinal studies has been going on for a number of years. Measures have been developed for routine activities (Wikström 2006; Wortley and Mazzerole (2008)) and for situational circumstances surrounding the commission of antisocial behavior and crime (Wikström 2006). Also, with the enormous amount of demographic information available on the web, community information can be more fine-tuned than it was in the past, and with less effort.

Another development that will take place is that genetic databases will be formed that collect data from different studies to allow for the study of rare factors. This requires careful documentation of design and variables in the individual studies contributing to such a database and a willingness to collaborate.

The study of girls' antisocial behavior and crime is still underdeveloped. The few studies that included girls often have too small a sample to yield enough girls with problems for analyses. Very few studies have a large enough sample to focus on low base-rate behaviors (review Hipwell and Loeber, 2006). The study of girls is particularly important because they will bring up the next generation of children, often without a male in the picture. Thus, their influence on the next generation is considerable. Antisocial girls often have babies early, which may mean that they are not ready for the responsibility of motherhood. In addition, their sexual partners are likely to be antisocial themselves. This leads to the question of transmission of problem behavior across generations.

An area where progress can be made in the future is the measurement of positive outcomes. Currently in many studies the overwhelming focus is on negative outcomes and on negative factors influencing an outcome. Positive factors are often implied from the lack of negative factors. Well-measured positive factors, such as skill development and competence, and outcomes, such as jobs and stable relationships, will be of enormous help in devising treatment plans.

Most current studies have not included information on the procedures involved in moving through the justice system. So far longitudinal studies have mainly relied on official records which generally do not tell how individual cases are dealt with and how long an individual was in what kind of institution and whether treatment was applied. Such information may be very relevant to the rate of re-offending.

Longitudinal data may, in the future, be used in creative ways to look at issues that have not been extensively studied so far. With the large amount of data collected at regular intervals, modeling exercises can be undertaken to examine what the effect certain interventions could have if applied to a population. Assuming a certain level of success of an intervention, the effect on rates of later problems can be estimated. These exercises can be very useful to convince funders and politicians that certain interventions may pay off over time. Better still would be the integration of an intervention study within a survey study, something that David Farrington has advocated (2003). Such a study would have to be planned, taking into account the number of subjects that will have to be "set aside" for intervention. Such a design would make a longitudinal study even more expensive by increasing the N and instituting an intervention for a part of the sample. After the fact, not many longitudinal researchers are willing, or in a position, to reduce their N by allowing an intervention to take place with a random subsample. This would reduce the power of the survey analyses.

One puzzling question that has not yet had much attention is secular changes in the age-crime curve. Even though in general the age-crime curve has a recognizably similar shape, for different age cohorts, there is also variation is the width and/or height of the curve. Some age cohorts may start delinquency earlier and stop later or may have a higher frequency (height of curve) of offending. The variables that influence the width and the height of the curve have not been studied extensively. Information about which factors influence the age-crime curve is of great importance for public policy and for reducing future crime waves. In order to study influences on the age-crime curve, one would need several age cohorts studied over a period of time so that the ages overlap in the course of the study. If substantial differences are found in the age-crime curves of the different cohorts, then the exercise is to see if one can find variables influencing the curves. It is difficult to predict the optimal age interval between the cohorts because potential changes affecting the age-crime curve lie in the future.

Optimizing the Use of Longitudinal Studies

The expense of longitudinal studies, generally funded with tax payers' money, requires serious thought about how to optimize the use of the study.

Collaboration Within Studies. With most studies aspiring to cover many disparate areas such as genetics and environmental influences, future studies will require a team of collaborators with a lead investigator willing and able to commit a large portion of his/her professional life to keeping such a study going and attract collaborators who are experts in specific fields to work with over many years.

Since the academic system values principal investigatorship rather than co-investigatorship, a structure needs to be found that makes it worthwhile to be part of a larger team. One way to do this is to devise substudies each with their own principal investigator and funding. However, such a solution presupposes a stable line of funding for the main study.

Collaboration Across Studies or Study Sites. It is expected that in the future, there will be a better integration of biological and environmental and individual factors. This will probably be accomplished by pooling data from different studies in order to make sure the sample is large enough for small effects to show. Data pooling is often more complicated than it sounds. Presumably "comparable" data may have a slightly different format, or different collection sites may have left some questions out, leading to a cumbersome process of making a combined data set.

The funding agencies are at present the driving force for data pooling and data transfer to agencies set up to deal with data storage and data requests from researchers outside of the original group of investigators. The advantage of wider collaboration, apart from a larger sample size, is that more extensive use is made of the large investment needed for longitudinal studies. The original investigators cannot be experts on all questions that could potentially be examined with the collected data. However, the danger is that regardless of how well the data are documented, "new" researchers may misinterpret the meaning of some variables. It is, therefore, always useful to be in contact with the original investigators. Data sharing provides a challenge for the promise of confidentiality. Even if the most obvious identifiable information is removed from a file, it may still be possible to identify a participant through a combination of variables (called deductive discovery). This puts a participant at risk of being identified and information being misused. In addition, it also puts the original investigators at risk for having to pay a fine for a breach in confidentiality.

In our experience, the safest way of dealing with these confidentiality problems is to have each person wanting to use data write a small proposal that is approved by his/her university and has the appropriate IRB approval, shifting the responsibility of confidentiality to the "new" researcher who was not involved in the original data collection. Since data sharing is now more or less the norm, I expect that procedures for de-identifying data and for certifying that "new" investigators will protect the confidentiality of the data will become standard soon. De-identifying a data set, however, will reduce the number of questions that can be answered after sensitive data has been removed.

Agency Involvement. The size of future studies and the range of variables that should be included may require funding from several agencies. Collaboration between agencies and the acceptance of new thoughts in the field will require that the investigators play a very active role in preparing agencies for the scope of the planned research and for the idea of funding from different agencies. Researchers cannot just wait for a request for proposals to come out that just fits their plans or expect agencies to take the lead role in searching for collaborations with other agencies. Instead, prospective investigators need to work together with agency staff to develop a blueprint for future longitudinal studies and for finding the necessary funding.

Secondary Data Analyses. Secondary data analyses, whether it is by pooling data from different studies or analyzing data from one study for a new purpose, are a great way to increase the yield of the research investment. It is useful, however, to keep a record of secondary research projects so that researchers do not compete with each other on the same topic. That is not to say that it would not be useful to have two different views on the same research question, but researchers should be aware of other projects using the same data bordering on their areas of interest.

In summary, the complex art and science of longitudinal studies continues to be exciting and challenging. New questions can be pursued and new data collection tools spring up regularly.

Related Entries

▶ Longitudinal Studies in Criminology

Recommended Reading and References

Anokwa Y, Hartung C, Brunette W, Borriello G, Lerer A (2009) Open source data collection in the developing world. Computer 42(10):97–99, http://dx.doi.org/10.1109/MC.2009.328

Armbrust M, Fox A, Griffith R, Joseph AD, Katz R, Konwinski A, Lee G, Patterson D, Rabkin A, Stoica I, Zahara M (2010) A view of cloud computing. Commun ACM 53(4):50–58

Celio MA, Vetter-O'Hagan CS, Lisman SA, Johanson GE, Spear LP (2011) Integrating field methodology and web-based data collection to assess the reliability of the Alcohol Use Disorders Identification Test (AUDIT). Drug Alcohol Depend. doi:10.10.1/j.drugalcdep.2011.05.035

Cotter RB, Burke JD, Loeber R, Navratil JL (2002) Innovative retention methods in longitudinal research: a case study of the Development Trends Study. J Child Fam Stud 11:485–498

Cotter RB, Burke JD, Stouthamer-Loeber M, Loeber R (2005) Contacting participants for follow-up: how much effort is required to retain participants in longitudinal studies? Eval Program Plann 28:15–21

Farrington DP (2003) Developmental and life-course criminology: key theoretical and empirical

issues- The 2002 Sutherland Award address. Criminology 41:201–235
Freeland KE, Carney RM (1992) Data management and accountability in behavioral and biomedical research. Am Psychol 47:640–645
Gosling SD, Vazire S, Srivastava S, John OP (2004) Should we trust web-based studies? Am Psychol 59(2):93–104
Groves RM, Mosher WD, Lepkowski JM, Kirgis NG (2009) Planning and development of the continuous National Survey of Family Growth. Vital Health Stat 1:1–64
Haggerty KP, Fleming CB, Catalano RF, Petrie RS, Rubin RJ, Grassley MH (2007) Ten years later: locating and interviewing children of drug abusers. Eval Program Plann 31:1–9
Hipwell A, Loeber R (2006) Do we know which interventions are effective for disruptive and delinquent girls? Clinical Child and Family Psychology Review 9, 221–255
Killias M, Redondo S, Sarnecki J (2012) European perspectives. In: Loeber R, Farrington DP (eds) Transitions from juvenile delinquency to adult crime: criminal careers, justice policy and prevention. Oxford University Press, New York
Loeber R, Farrington DP, Stouthamer-Loeber M, White HR (2008) Violence and serious theft. Routledge, New York
Magnusson D, Bergman L (1990) Data quality in longitudinal research. Cambridge University Press, Cambridge
McCartney K, Burchinal MR, Bub KL (2006) Best practices in quantitative methods for developmentalists. Blackwell Publishing, Boston
Piquero AR, Farrington DP, Nagin DS, Moffitt TE (2010) Trajectories of offending and their relation to life failure in late middle age: findings from the Cambridge Study in Delinquent Development. J Res Crime Delinq 47(2):151–173
Robins LN (1976) Problems in follow-up studies. Am J Psychiatry 134:904–907
Schafer JL, Graham JW (2002) Missing data: our view of the state of the art. Psychol Methods 7:147–177
Stouthamer-Loeber M (1993) Optimizing data quality of individual and community sources in longitudinal research. In: Farrington DP, Sampson R, Wikstrom P-OH (eds) Integrating individual and ecological aspects of crime. BRÅ-report, Stockholm
Stouthamer-Loeber M, Van Kammen WB (1995) Data collection and management: a practical guide. Sage, Newbury Park
Wikström P-O (2006) Individuals, settings and acts of crime. Situational mechanisms and the explanation of crime. In: Wikström P-O, Sampson RJ (eds) The explanation of crime: Context, mechanisms and development. Cambridge University Press, Cambridge
Wortley R, Mazzerole L (eds) (2008) Environmental criminology and crime analysis. Willan, Cullompton

Order Maintenance

▶ Broken Windows Thesis
▶ Discriminant Validity of Disorder and Crime

Order Maintenance Policing

William Sousa[1] and George Kelling[2]
[1]University of Nevada, Las Vegas, Las Vegas, NV, USA
[2]Rutgers University, Newark, NJ, USA

Synonyms

"Broken windows" policing; Quality-of-life policing

Overview

Order maintenance policing is a police practice that involves managing minor offenses and neighborhood disorders in order to address community problems. Order maintenance policing is influenced by the "broken windows" hypothesis, which describes the process by which minor offenses can lead to citizen fear and the decline of neighborhoods. Order maintenance policing has been credited with crime reduction in cities across the United States, most notably in New York City during the 1990s.

Definition of Order Maintenance Policing

That officers should help to maintain order in communities has always been an expected and desired function of modern police. In the contemporary sense, order maintenance policing (also called "broken windows" policing or "quality-of-life" policing) refers to a police operational tactic that involves managing minor offenses and acts of physical and social disorder. Order maintenance is generally identified as an element of the community-policing paradigm, and as such it is closely associated with problem-oriented policing and situational crime prevention.

The merits of reducing disorder and quality-of-life offenses are self-evident (Wilson and Kelling 2006). In addition, order maintenance policing has been credited with reducing serious crime in many communities. As evidence of its effectiveness grows, more police departments and other criminal justice agencies have integrated its principles into their functions. Yet while order maintenance continues to be popular with citizens and government agencies, it is often not well understood. Order maintenance policing, for example, is often mistakenly labeled as zero-tolerance policing. The association between these terms confuses theoretical discussions and invites misinterpretation over policy and practice.

This entry begins with the origins of order maintenance policing, including a description of its theoretical underpinnings. The article then examines the approach within the context of the community-policing paradigm before turning to applications of order maintenance and evidence of its effectiveness. The entry concludes with current debates and controversies over order maintenance policing efforts.

Origins of Order Maintenance Policing

The "Broken Windows" Influence

Order maintenance policing is highly influenced by the "broken windows" hypothesis. Developed by James Q. Wilson and George Kelling in 1982, "broken windows" describes the process by which minor criminal behaviors and other types of community disorders can lead to the decline of neighborhoods. Wilson and Kelling used the term "broken windows" as a metaphor in the original hypothesis, reflecting the idea that one broken window in a building is the first sign of disorder – left unrepaired, more windows may well become broken as such behavior is overlooked or ignored.

Wilson and Kelling further argued that relatively minor offenses have the potential to generate fear among citizens, causing them to alter their own behavior to avoid confrontation with disorderly actors or conditions. As people alter their behavior to avoid disorderly places, these locations experience less informal social control and produce an atmosphere where acts of disorder appear acceptable. These places are more susceptible to serious crime because offenders will feel more at ease at locations with permissive atmospheres toward disorder and weak informal social control mechanisms.

The disorders that Wilson and Kelling refer to in the original article involve social behaviors (such as prostitution, public alcohol and drug use, public urination and defecation, aggressive panhandling, unruly gangs of youth) and physical conditions (such as vandalism, graffiti, abandoned buildings and vehicles, unkempt vacant lots, trash and litter, discarded drug paraphernalia) (see also Skogan 1990). It is important to note that disorder does not necessarily refer to illegal activities. As described by Wilson and Kelling, disorderly acts also involve behaviors that are "disreputable or obstreperous or unpredictable" (1982: 30) – although perhaps not criminal in the legal sense, these behaviors can inspire fear and apprehension in the eyes of an observer. It is also important to consider that while minor offenses and disorder can potentially impact quality of life in neighborhoods, community decline and serious crime are not inevitable. Different communities possess varying capacities to absorb disorder, and "broken windows" theorists acknowledge that numerous factors contribute to whether certain behaviors or conditions are considered problematic to citizens. For example, the time of day, the type of place (i.e., residential versus commercial versus industrial locations), the previous behaviors of disorderly persons, the condition of the observer, and the amount/accumulation of minor offenses are all factors in terms of the relative impact of disorder on a given community (Kelling and Coles 1996).

The policy implication of the "broken windows" hypothesis is clear: if police and citizens proactively maintain order by managing minor offenses, they can control a source of fear and potentially prevent more serious crime and the deterioration of quality of life in neighborhoods. Thus, “order maintenance” is a viable option for police who must address community problems involving both disorderly offenses and serious criminal activity.

Inspiration from Urban Planning, Social Psychology, and Police Practice

Three major tenets of “broken windows” (and order maintenance policing by extension) include (1) the view that citizens desire a degree of orderliness in communities, (2) the idea that minor disorder can lead to greater disorder, and (3) the belief that police are uniquely situated to support citizens in maintaining order. These three tenets were influenced by a number of works from a variety of disciplines (see Ranasinghe 2012). For example, Jane Jacobs’ *The Death and Life of Great American Cities* (1961) – a classic in the field of urban development – demonstrated the importance of order maintenance in communities by describing how citizens are strong proponents of (and can be great providers of) peace and order in public places. Experiments in social psychology – including the work of Philip Zimbardo (1970) – provided a foundation for the idea that indifference toward minor disorders can potentially lead to more serious forms of disorder. Studies in police practice, especially Egon Bittner’s (1967) seminal study of officers on skid row, demonstrated the distinction between the “law enforcement” and “peacekeeping” roles of police while revealing the police capacity for managing minor offenses in troubled communities.

Perhaps the most important empirical influences on broken windows, however, came from two sources. The first involved observations of police behavior in eight communities conducted by James Q. Wilson in the 1960s (Wilson 1968). These observations revealed three different styles of police departments – the legalistic, the watchman, and the service styles – that influence the performance and discretion of officers. The second involved the results of studies on foot patrol conducted during the 1970s. For example, researchers conducting the Newark Foot Patrol Experiment (Kelling et al. 1981) found that citizens in areas where there was foot patrol felt safer and were generally more satisfied with police services than citizens in areas where there was only motorized patrol. Drawing on these studies, Wilson and Kelling argued that foot patrol officers exercised more discretion and were more proactive in terms of regulating disorderly activities that troubled citizens. This formed the basis of the broken windows hypothesis and the idea that order maintenance performed by police can be of intrinsic value to neighborhoods.

Applications of Order Maintenance Policing

Relationship to Community Policing and Problem-Oriented Policing

Order maintenance policing and its related theory of broken windows are often associated with the origins of the community-policing paradigm. The core elements of community policing, including greater police accountability to citizens, enhanced partnerships with citizens, and proactive problem solving, began to evolve in the late 1970s in response to the failure of police strategies to adequately address urban problems and other community concerns (Kelling and Coles 1996; Kelling and Moore 1988). As foot patrol experiments informed the ideas behind broken windows, Herman Goldstein (1979) argued that the policing tactics of the mid-1900s were overly reactive in nature and focused too specifically on resolving individual incidents rather than broader community problems. To improve policing, Goldstein suggested a more proactive “problem-oriented” approach where police consider single incidents as potential symptoms of larger problems. By identifying and addressing the symptoms, officers can work to prevent future incidents and thus eliminate broader neighborhood concerns.

Since the introduction of Goldstein’s problem-oriented approach, a number of tactics have become associated with proactive activities designed to prevent crime and disorder. These include tactics that are designed to address highly problematic locations (i.e., “hot spots” policing (see Sherman and Weisburd 1995)), high-risk victims (i.e., repeat victimization strategies (see Farrell and Pease 1993)), and high-risk offenders (i.e., “pulling levers” and “ceasefire” efforts (see Kennedy 1998 and Braga et al. 2001)). Additionally, research into situational crime prevention offers police and citizens numerous options for deterring criminal activity and other community concerns (see Clarke and Eck 2005).

Order maintenance tactics can also be viewed within this problem-oriented policing framework. To the extent that community problems are associated with minor disorders and quality-of-life offenses, order maintenance offers a potential means of managing those problems. Moreover, because order maintenance tactics are generally implemented in response to public demand for enhanced police services, they represent a measure of accountability to citizens in support of the general objectives of community policing.

Specific and General Applications

Applications of order maintenance policing have ranged from microlevel initiatives to macrolevel policies. Microlevel initiatives involve programs where police seek to impact specific behaviors at targeted locations. A shopping plaza that is frequently the target of graffiti vandals, a city street that is a common stroll for prostitutes, or a public park where teenagers constantly harass and intimidate passersby could be instances where specific order maintenance interventions are appropriate. Police departments have also implemented order maintenance as a more general, problem-oriented strategy. In these cases, officers are asked to continuously analyze the impact of disorder on their beats and to be proactive in its management.

An example of a microlevel order maintenance intervention occurred in Los Angeles’ MacArthur Park in 2003 (Sousa and Kelling 2010). Over the course of decades, MacArthur Park had deteriorated from a scenic, recreational location to a hot spot for drug dealing, public drug and alcohol use, prostitution, vandalism, and various other forms of public disorder. In 2003, the Los Angeles Police Department (LAPD) began an initiative that combined elements of situational crime prevention with assertive order maintenance. By focusing resources on minor offenses, police were able to restore order to a level where citizens felt less threatened and more comfortable using the park for its intended purposes. MacArthur Park is now largely self-regulating – although LAPD continues to be mindful of the conditions in the park, the legitimate users of the park are able to maintain order on their own without substantial police intervention.

In contrast to a microlevel application, New York City provides perhaps the most well-known example of order maintenance as a macrolevel policy application. In the early 1990s, the New York City Police Department (NYPD) was facing increasing citizen demand for more effective police service as the result of high crime rates and nearly unchecked disorder in public locations and residential neighborhoods. Culturally, NYPD was a department that was generally inattentive to disorder. Officers were often discouraged from enforcing minor offenses because the department considered them to be either too unimportant or too controversial for police intervention.

Order maintenance policing in New York began in 1993 with NYPD Commissioner Ray Kelly and his much publicized efforts to have police manage aggressive “squeegeeing” – a quality-of-life offense that for many New Yorkers had become a symbol of unregulated disorderly behavior. NYPD order maintenance, however, developed further and became more closely associated with Commissioner William Bratton’s administration starting in 1994. Bratton, a strong advocate of the “broken windows” hypothesis and the principles of order maintenance, had success implementing order maintenance tactics while he was the head of the New York City Transit Police in 1990. While with the Transit Police, Bratton asked officers to more assertively manage turnstile jumping, fare evasion, and other low-level misbehavior occurring within the subway system – a tactic that was later associated with a reduction in serious crime and other forms of disorder (Kelling and Coles 1996; Bratton and Knobler 1998). When Bratton became Commissioner of the NYPD several years later, he applied the same principles on a citywide basis: NYPD officers were now encouraged to manage minor offenses and other forms of disorder that were considered problems in New York communities.

Crime in New York City declined dramatically in the 1990s following the implementation of order maintenance policing (Kelling and Coles 1996). New York's crime reduction continues to be the subject of numerous academic and journalistic commentaries. Several theories have been offered for the historic crime drop, including varying demographic patterns, shifts in economic conditions, and changes in drug-use trends – especially those associated with the crack-cocaine epidemic of the late 1980s. Other programs in the city may have played a role as well, including the influence of Business Improvement Districts (BIDs) (agreements between property owners to pool funds to provide security and maintenance for local businesses) and the establishment of the Midtown Community Court in 1993. NYPD also began additional initiatives during Bratton's administration, including the now famous Compstat system – a process that combines managerial accountability with data management and problem-oriented policing. Nevertheless, NYPD's general order maintenance policy is often considered the primary catalyst for crime reduction in New York City during the 1990s.

Order Maintenance Policing: Evidence of Effectiveness

Empirical Evidence

Researchers who study order maintenance policing will often examine the influence of the enforcement of minor offenses (a proxy for order maintenance) on instances of serious crime. One method of doing this involves the use of trend analyses that consider the impact of order maintenance tactics over time. Kelling and Sousa (2001), for example, combined data from the NYPD, the New York City Department of Education, the New York State Department of Labor, and the New York State Department of Health to conclude that order maintenance policing was the most significant factor in New York's crime reduction, suggesting that over 60,000 violent crimes were prevented during the 1990s as a result of the tactic. Corman and Mocan's (2002) time-series analysis of police and economic data also demonstrated the crime prevention benefits of order maintenance in New York, indicating that the policy produced a substantial reduction in motor vehicle theft and robbery from 1990 to 1999. Using California county-level arrest and complaint data from 1989 through 2000, Worrall (2006) found that the enforcement of minor offenses such as disorderly conduct, disturbing the peace, and public drunkenness reduced felonious assaults and burglaries.

Trend studies of order maintenance, however, have not always produced consistent results. For example, although Corman and Mocan (2002) found that order maintenance had a substantial impact on auto theft and robbery, they did not find a similar effect for other crimes such as assault and burglary. Messner et al.'s (2007) cross-sectional time-series analysis of data from New York City generally supports earlier conclusions that order maintenance policing had a significant impact on serious crime during the 1990s, but they also found that drug-use trends were important influences. Rosenfeld et al. (2007) found some crime-reduction value of order maintenance policing in New York using trend data from police, medical, and census sources, but they argue that the impact of the tactic was moderate. Similarly, Berk and MacDonald (2010) used 8 years of time-series data to evaluate the impact of an initiative that involved disorder-management activities by police in Los Angeles' Skid Row. They found that the initiative had a meaningful impact on nuisance, property, and violent crimes in the target area with some beneficial spillover effect into the surrounding areas, but they determined that the effect of the initiative was relatively modest.

Other trend/time-series analyses find few significant crime-reduction benefits of order maintenance policing. Novak et al. (1999) determined that a police initiative focusing on joyriding, loud noise, public alcohol use, and other public nuisance offenses did not reduce robbery or burglary in a community, although they admit that the project was not designed to directly impact those serious crime categories. Similarly, Katz et al. (2001) found that quality-of-life policing initiative in a section of Chandler, Arizona, did not significantly reduce calls for service for serious crime, but it did reduce calls for public morals and physical disorder offenses. Harcourt and Ludwig (2006) found no relationship between order maintenance policing and the crime decline in New York during the 1990s, although subsequent analyses failed to replicate their findings (see Messner et al. 2007).

Besides trend studies, researchers have also employed experimental designs – generally considered to be a more rigorous methodology – to study the impact of order maintenance policing. Braga et al. (1999), for example, utilized a randomized experimental design to examine the impact of a problem-oriented policing strategy at violent crime hot spots in Jersey City, New Jersey. This problem-oriented strategy, which primarily involved targeting minor offenses and forms of physical and social disorder, was successful at reducing crime at the experimental locations with little evidence of displacement. Braga and Bond (2008) used a similar experimental design to assess the effects of policing disorder at hot spots in Lowell, Massachusetts. They found significant reductions in crime in the experimental areas compared to the control areas – a result of the disorder-management strategies employed by the police.

Measuring What Matters

Studies that evaluate the impact of order maintenance policing often face several challenges in terms of measurement. One problem – particularly for trend and time-series analyses – is in the choice of the key independent variable that represents the concept of order maintenance. Researchers will often rely on a variable from agency data that measures the enforcement of minor offenses, such as officially recorded arrests for misdemeanor offenses or citations for minor legal infractions. The difficulty with indicators such as these, however, is that they do not capture the full extent of the idea of order maintenance policing. Enforcing minor offenses involves complicated interactions between officers and citizens and often includes a range of police actions that are much more complex than making arrests or issuing citations. Studies that use these measures as independent variables, therefore, are likely underestimating the extent of order maintenance policing that has been implemented. Many of the trend/time-series studies discussed above, for example, tend to rely on arrest data as a proxy for order maintenance. With only a few exceptions (see, e.g., Kelling and Sousa 2001), authors do not fully explore the limitations of such data.

A second problem with evaluations of order maintenance policing is in the choice of the dependent variable. Researchers generally choose an indicator of serious crime (i.e., felony crime) as the outcome variable because of the link between disorder and serious offenses that is hypothesized by the "broken windows" theory. This choice makes intuitive sense if the objective of the order maintenance initiative is to reduce serious crime. The difficulty, however, is that a reduction in serious crime is not always the primary goal (or even an intended goal) of an order maintenance program – sometimes the objective is to reduce disorder itself. For example, robberies may well have gone down in the New York City subway system as a result of Bratton's order maintenance program, but the initiative was originally implemented to gain control over a chaotic environment (Kelling and Coles 1996). Likewise, the order maintenance activities performed by LAPD in Los Angeles' MacArthur Park were meant to decrease disorder so that all citizens felt comfortable using the park for its intended purposes (Sousa and Kelling 2010).

Studies that fail to consider appropriate outcome variables may miss important indicators of success from order maintenance initiatives.

Thacher (2004) makes this point clear in his discussion of the value of order maintenance policing. Thacher argues that the relationship between order maintenance policing and a reduction in serious crime is *indirect* and therefore the sort of causal connection that is difficult to detect using social science methodologies. He suggests, however, that order maintenance is *directly* related to a reduction in minor offenses. As such, it is more useful for researchers to explore the importance of order maintenance policing at reducing disorder, regardless of its impact on serious offending. To the extent that disorder is itself a problem in communities, order maintenance efforts that target disorder have intrinsic value.

Recent research has indeed been more inclined to consider the impact of order maintenance policing beyond its impact on serious crime. For example, Hinkle and Weisburd (2008) use survey data from Jersey City, New Jersey, to examine the influence of an order maintenance initiative on citizens' fear of crime and perceptions of disorder. Their findings suggest that reductions in disorder can lead to reductions in fear. Interestingly, the authors also found that fear of crime may be enhanced by increased police presence, indicating that police should carefully communicate with citizens while engaging in high-visibility order maintenance activities.

Understanding Order Maintenance Policing

Controversies

Practitioners, policy-makers, and citizens commonly accept order maintenance policing as an option for improving the quality of life in neighborhoods. Within academic circles, however, order maintenance has frequently been the subject of contentious and politically charged debates. Academic criticisms of order maintenance come in various forms. Some critics question the theoretical underpinnings of order maintenance policing. Others question the appropriateness of giving officers the authority to police disorder in communities.

Critics that challenge the theoretical underpinnings of order maintenance often point to the "broken windows" hypothesis and the theorized link between disorder and serious crime. They argue that if a direct causal connection between disorder and serious crime is not strong, then efforts to manage disorder will be futile. Research is inconsistent regarding the direct link between incivilities and crime. Skogan (1990), for example, found a significant relationship between disorder and criminal activity, while analyses by Harcourt (2001) and Sampson and Raudenbush (1999) suggest that a direct link may be weak except for specific crimes like robbery. Taylor (2001) found some evidence that disorder directly leads to crime but indicates that the relationship is relatively weak and that neighborhood factors may be better predictors of criminal activity.

Two responses can be directed at the assertion that a weak direct connection between disorder and serious crime is a challenge to "broken windows" and to order maintenance policing. First, the "broken windows" hypothesis as stated by Wilson and Kelling does not propose a direct connection between disorder and serious crime – the hypothesized connection is indirect, mediated by increases in citizen fear and breakdowns in informal social control mechanisms. This explains why communities with strong informal social control can manage a level of disorder that might otherwise damage communities with weaker control mechanisms. Research that suggests a weak causal connection between disorder and serious crime, therefore, does notnecessarily disprove "broken windows" (see Gault and Silver 2008). Second, as discussed previously, there may be intrinsic value to order maintenance policing even if a strong causal connection between disorder and serious crime cannot be established (Thacher 2004). Disorder is itself a concern in communities, demonstrated by research that consistently shows a relationship between incivilities and citizen fear (see, e.g., Skogan and Maxfield 1981). A relatively weak causal connection between disorder and serious crime, therefore, does not justify the abandonment of order maintenance principles.

A second general concern voiced by critics of order maintenance policing involves the concept of "disorder" and whether it is morally appropriate for police to enforce it. Critics essentially argue that disorder is a vague concept and that one's perception of it is subjective. Police efforts to enforce disorderly behavior may therefore criminalize relatively harmless activities that some citizens consider to be acceptable (see Harcourt 2001). This is particularly a concern if police order maintenance efforts disproportionately impact people who live in poor, urban neighborhoods (see Fagan and Davies 2000). Further, critics often equate order maintenance tactics with the term "zero-tolerance" policing, suggesting that police exercise little discretion in their heavy-handed management of minor offenses (see Greene 1999).

Much of this criticism of order maintenance policing originates from academic and political sources (Sousa and Kelling 2010). These sources often differ ideologically from police administrators in terms of methods by which police should manage problems in communities. For example, although critics have warned about the misuse of order maintenance policing, few have provided empirical evidence of its misuse. Those that claim that order maintenance criminalizes harmless behavior, for instance, usually do not consider the fear-inducing qualities of disorder. Critics who argue that order maintenance practices disproportionately impact citizens living in poor, urban communities often do not consider the demand from those communities for police to deal with incivilities and minor offenses (see Skogan 1990). Finally, those that claim that order maintenance is "zero tolerance" have rarely viewed order maintenance in practice, relying instead on dramatized media accounts and agency data that are limited and often flawed.

Order Maintenance in Practice

Critics of order maintenance appropriately call attention to the moral complexities of policing disorder – and the proper deployment of order maintenance is indeed a concern. In fact, the authors of the original "broken windows" essay warned against the misuse of policies derived from the theory. Wilson and Kelling (1982) indicated, for example, that caution is necessary to ensure the fair balance between individual rights and community interests when managing disorder. Also, proper care must be taken to make certain that police use discretion appropriately when enforcing neighborhood rules and the criminal law. Like many police tools, order maintenance is a tactic that can be abused. (The history of interrogation practices demonstrates how a powerful police technique requires continuous oversight and vigilance.) Yet, like interrogation practices, order maintenance can provide a valuable option to police if properly managed.

Due to the moral complexities of policing disorder, Thacher (2004) suggests that the merits of order maintenance should be examined based on how it is deployed in communities. Some recent research has proceeded along these lines by evaluating citizens' views of order maintenance and by examining the activities of officers who practice it. Weisburd et al. (2011), for example, administered telephone surveys to citizens living in neighborhoods that had received aggressive order maintenance policing. They found that residents were not negatively impacted by police actions at those locations, suggesting that police legitimacy is not harmed by order maintenance activities. Sousa (2010) conducted observations of officers practicing order maintenance policing in New York City. Contrary to critics' claims that order maintenance is the equivalent of zero tolerance, he concluded that officers were aware of the complexities of policing disorder and often used their discretion to informally resolve situations that involved minor offenses. Importantly, the observations revealed that "maintaining order" involved making arrests and writing citations only rarely – officers were much more inclined to warn or verbally reprimand citizens for committing disorderly acts. Order maintenance activities were better categorized as officers "paying attention" to minor offenses (rather than ignoring them) – but where official action (arrest or citation) was only one of several options.

Despite the controversies, the available evidence suggests that order maintenance policing can be implemented fairly and appropriately. As with any police-involved initiative, order maintenance practices are most promising when the community accepts the intended police tactics. Therefore, to lessen controversy, supervision over officers and communication between police and the public concerning quality-of-life policing are an important part of an order maintenance program.

Related Entries

- ▶ Broken Windows Thesis
- ▶ Community Policing
- ▶ Defining Disorder
- ▶ Problem-Oriented Policing
- ▶ Theories on Policing and Communities

Recommended Reading and References

Berk R, MacDonald J (2010) Policing the homeless: an evaluation of efforts to reduce homeless-related crime. Criminol Publ Pol 9(4):813–840

Bittner E (1967) The police on skid row: a study of peacekeeping. Am Sociol Rev 32(5):699–715

Braga AA, Bond BJ (2008) Policing crime and disorder hot spots: a randomized controlled trial. Criminology 46(3):577–607

Braga AA, Weisburd D, Waring EJ, Green-Mazerolle L, Spelman W, Gajewski F (1999) Problem-oriented policing in violent crime places: a randomized controlled experiment. Criminology 37:541–580

Braga AA, Kennedy DM, Waring EJ, Piehl AM (2001) Problem-oriented policing, deterrence, and youth violence: an evaluation of Boston's Operation Ceasefire. J Res Crime Delinq 38:195

Bratton W, Knobler P (1998) Turnaround: how America's top cop reversed the crime epidemic. Random House, New York

Clarke RV, Eck JE (2005) Crime analysis for problem solvers in 60 small steps. U.S. Department of Justice, Office of Community Oriented Policing Services, Washington DC

Corman H, Mocan N (2002) Carrots, sticks, and broken windows. National Bureau of Economic Research, Cambridge, MA. Working Paper 9061

Fagan J, Davies G (2000) Street stops and broken windows: terry, race, and disorder in New York City. Fordham Urban Law Review 28(2):457–504

Farrell G, Pease K (1993) Once bitten, twice bitten: repeat victimisation and its implications for crime prevention. Home Office, Police Research Group, London. Crime Prevention Unit Paper 46

Gault M, Silver E (2008) Spuriousness or mediation? Broken windows according to Sampson and Raudenbush (1999). Journal of Criminal Justice 36:240–243

Goldstein H (1979) Improving policing: a problem-oriented approach. Crime Delinq 25:236–258

Greene JA (1999) Zero tolerance: a case study of police policies and practices in New York City. Crime Delinq 45(2):171–187

Harcourt BE (2001) The illusion of order: the false promise of broken windows policing. Harvard University Press, Cambridge

Harcourt BE, Ludwig J (2006) Broken windows: new evidence from New York City and a five-city social experiment. University of Chicago Law Review 73:271–320

Hinkle JC, Weisburd D (2008) The irony of broken windows policing: a micro-place study of the relationship between disorder, focused police crackdowns and fear of crime. Journal of Criminal Justice 36:503

Jacobs J (1961) The death and life of great American cities. Vintage Books, New York

Katz CM, Webb VJ, Schaefer DR (2001) An assessment of the impact of quality-of-life policing on crime and disorder. Justice Quarterly 18(4):825–876

Kelling GL, Coles CM (1996) Fixing broken windows. Martin Kessler Books, New York

Kelling GL, Moore MH (1988) The evolving strategy of policing. National Institute of Justice and Harvard University: Perspectives on Policing 4:1–15

Kelling GL, Pate A, Ferrara A, Utne M, Brown C (1981) The Newark foot patrol experiment. Police Foundation, Washington, DC

Kelling GL, Sousa WH (2001) Do police matter? An analysis of the impact of New York City's police reforms. Manhattan Institute, New York. Civic Report No. 22

Kennedy DM (1998) Pulling levers: getting deterrence right. National Institute of Justice Journal. US Department of Justice: National Institute of Justice

Messner SF, Galea S, Tardiff K, Tracy M, Bucciarelli A, Piper TM, Frye V, Vlahov D (2007) Policing, drugs, and the homicide decline in New York City in the 1990s. Criminology 45:385–414

Novak KJ, Hartman JL, Holsinger AM, Turner MG (1999) The effects of aggressive policing of disorder on serious crime. Policing: An International Journal of Police Strategies & Management 22(2):171–194

Ranasinghe P (2012) Jane Jacobs' framing of public disorder and its relation to the 'broken windows' theory. Theor Criminol 16(1):63–84

Rosenfeld R, Fornango R, Rengifo A (2007) The impact of order-maintenance policing on New York City homicide and robbery rates: 1988–2001. Criminology 45:355–384

Sampson RJ, Raudenbush SW (1999) Systematic social observation of public spaces: a new look at disorder in urban neighborhoods. Am J Sociol 105(3):603–651

Sherman LW, Weisburd D (1995) General deterrent effects of police patrol in crime 'hot spots': a randomized, controlled trial. Justice Quarterly 12:625–648

Skogan WG (1990) Disorder and decline: crime and the spiral of urban decay in American neighborhoods. Free Press, New York

Skogan WG, Maxfield MG (1981) Coping with crime: individual and neighborhood reactions. Sage Publications, Beverly Hills

Sousa WH (2010) Paying attention to minor offenses: order maintenance policing in practice. Police Practice and Research 11(1):45–59

Sousa WH, Kelling GL (2010) Police and the reclamation of public places: a study of MacArthur Park in Los Angeles. International Journal of Police Science and Management 12:1

Taylor RB (2001) Breaking away from broken windows. Westview Press, Boulder, CO

Thacher D (2004) Order maintenance reconsidered: moving beyond strong causal reasoning. J Crim Law Criminol 94(2):101–133

Weisburd D, Hinkle JC, Famega C, Ready J (2011) The possible 'backfire' effects of hot spots policing: an experimental assessment of impacts on legitimacy, fear and collective efficacy. J Crim Law Crim 7:297–320

Wilson JQ (1968) Varieties of police behavior. Harvard University Press, Cambridge

Wilson JQ, Kelling GL (2006) A quarter century of broken windows. The American Interest 11(1): 168–172

Wilson JQ, Kelling GL (1982) Broken windows: the police and neighborhood safety. Atlantic Monthly

Worrall JL (2006) Does targeting minor offenses reduce serious crime? A provisional, affirmative answer based on an analysis of county-level data. Police Quarterly 9(1):47–72

Zimbardo PG (1970) The human choice: individuation, reason, and order versus deindividuation, impulse, and chaos. In: Arnold WJ, Levine D (eds), Nebraska Symposium on Motivation – 1969, vol 17. 237–307, University of Nebraska Press

Organizational Change and Police Legitimacy

Jacinta M. Gau
Department of Criminal Justice, University of Central Florida, Orlando, FL, USA

Synonyms

Institutional environment; Institutional legitimacy; Institutional theory; Police legitimacy; Police organizations

Overview

Institutional theory holds that service organizations within a given field operate according to established, value-based myths. Adherence to the myths legitimizes an organization, while departure from them may spark a legitimation crisis. Police institutional legitimacy depends, in part, upon every organization's continued observance of the symbols and ceremonies that characterize the policing mythology. Importantly, these myths have been substantially decoupled from task performance, such that effectiveness and efficiency of operations are less important than is paying homage to the rituals. This entry presents suggestions for programs, policies, selection and training, and media and public relations. These suggestions are framed using the discourse of police institutional legitimacy. They are contended to be methods by which organizations can enhance task-based legitimacy without blatantly flouting myths and threatening to spark legitimation crises. It is argued that even in a value-laden institutional environment, task performance can be consonant with, if still inferior to, myth and ceremony.

The Institutional Theory of Police Organizations

Institutional theory posits that individual organizations are situated within large networks comprising other, similar organizations, interorganizational relationships, and the social and political landscapes that delineate reigning normative frameworks governing agencies' administrations and operations. The institutional climate is shaped by the socially constructed "solutions" to the (sometimes artificial) problematization of particular societal phenomena. Every organization is subsumed by and must conform to these existing social facts, or it will lose legitimacy and, possibly, perish entirely.

Under institutional theory, organizations must arrange their structures, policies, and personnel strategies to meet one or both of two ultimate ends: survival and performance (Zucker 1987).

The two goals operate in tandem in the best-case scenario, but are not always closely linked in practice (Meyer and Rowan 1977). In the private or business sector, where the free market theoretically rewards high-quality products and efficient resource management (and punishes the opposite), performance determines survival, and only those organizations that are effective and efficient endure long term. In the public sector, however, survival is often an institutional given because the organization boasts state-sanctioned legitimacy and has become a staple feature of the modern US municipal, county, or state jurisdiction (see Crank 2003 for a general discussion). State-sanctioned organizations – unlike their product-oriented business counterparts – provide services rather than readily identifiable goods (Crank 2003). State-sanctioned organizations thus cannot derive institutional legitimacy from the competitive marketplace and must, instead, actively and consistently prove their credibility, worthiness, and necessity.

Police agencies, as state-sanctioned organizations, exist in an institutional environment (see Zucker 1987). Their survival is guaranteed – police have become a fixture in modern American government, and they have no credible competitors – but their performance and legitimacy are not. Legitimacy, in turn, does not flow automatically from task performance. In policing, as in other organizations that navigate institutional environments (Meyer and Rowan 1977), effective and efficient execution of tasks is a matter quite separate from value-laden judgments about the ideal structure, function, and mission of the organization. Police organizations' legitimacy is based not on actual task performance, but on values and on institutional myths.

As explicated by Crank and Langworthy (1992), myths are socially constructed, norm-derived mandates of policing that when present, legitimize the police and, when absent, delegitimize them. Institutional myths include the image of police as crime fighters, the assumption that civil service is preferable to other hiring and evaluation schemes, and that procedural due process ensures fair and impartial outcomes. These value systems all pertain to the "ought" of policing, to the near-total exclusion of the "is"; that is to say, values are decoupled from task performance. The myths are so deeply embedded in society and in the institutional environment that they have taken on "an intrinsic quality of 'truth' or 'rightness'" (1992: 347). They are self-evident and beyond question.

Myths are observed and reinforced through the use of ceremonial rituals. The militaristic uniforms and insignia-bearing patrol cars are important symbols (Manning 1988) conveying to the public that this police agency "*looks like* or *acts like*" (Crank and Langworthy 1992: 347, emphasis in original) the archetypal police organization. The archetype is the product of a diffusion of innovation that produced isomorphism as, over time, police agencies observed the legitimacy enjoyed by successful agencies and consequently adopted their structural and operational arrangements in a process of mimicry. Isomorphism, though, can be coercive – once a standard or modal arrangement exists, nonconforming organizations may acquire a deviant status and be viewed as less competent, irrespective of the actual effectiveness and efficiency of their own model or of that to which they are being unfavorably compared (DiMaggio and Powell 1983). To establish their institutional legitimacy, then, police organizations, executives, and officers must engage in a dramaturgical (Manning 1996) display of visible activities and appearances that reaffirm their status as a "real" police organization.

That the institutional legitimacy of a police organization is predicated less on its performance than on its conformity with expectations imposed by nonexpert constituents poses obstacles for both effectiveness and legitimacy. The archetypal police organization is built upon articles of faith articulated by police leaders, union representatives, politicians, and community leaders who have opined and argued about what, precisely, police should be and do (see Wilson 1968). The result is an institutional mythology that puts police, scholars, and society alike "in the unenviable position of taking the word of those who are presumed to be knowledgeable" (Langworthy 1986: 126).

The decoupling of values from tasks and the superiority of the former over the latter create opportunities for legitimation crises. These crises occur when constituencies with competing value sets clash, and it suddenly becomes unclear just who it is the police are here to serve. Events such as corruption scandals and high-profile instances of excessive force against racial minorities can trigger these crises. The traditional methods for quelling legitimation crises at the organizational level involve the public degradation of the department and the ousting of the existing chief (Crank and Langworthy 1992). A special commission may be convened to issue observations and recommendations that will probably be ignored. The public eventually forgives (or simply forgets) the indiscretion, and the new chief – because he is isomorphic with the disgraced one due to the homogenizing influence of police hiring and promotional practices (see also DiMaggio and Powell 1983) – substantially maintains the status quo. The process of scandal and reform is complete, and time passes until the next scandal erupts.

The purpose of this entry is to propose strategies for enhancing the institutional legitimacy of police using methods that are consistent with both the legitimacy-granting myths of policing and, to the extent possible, with actual task performance. Myths, despite their flaws, must be treated with due deference. Some modern reformers have argued for radical reformation of the administration and operations of police organizations, such as delayerizing, decentralizing, and civilianizing (Maguire 1997). The institutional environment of policing, however, balks at such bold moves. Any department that attempts to alter itself in ways discrepant from institutionalized myths will likely meet the same end as the well-intentioned but ill-fated department described by Crank and Langworthy (1992). The department attempted a wholesale overhaul of its uniforms, rank structure, and even its name. As a consequence, the agency was ostracized and alienated until it finally succumbed to the pressure and reverted to looking and acting like a "real" department. The power of the institutional monolith likely helps explain why even those police organizations that pay homage to community policing have not structurally or functionally reorganized in a manner consistent with the philosophy (Maguire 1997; Mastrofski and Ritti 2000).

It is, therefore, important that the infusion of task-based performance into the extant myth-based environment be a subtle process of shift and subversion rather than a precipitous, highly visible event. The following discussion describes strategies that may be used to enhance the institutional legitimacy of the police in a manner that incorporates a discussion of task performance. "Task" is used broadly here and ranges from crime reduction to public image management; this loose definition recognizes the spectrum of ends that organizations can seek to meet in their efforts to be successful. The section subheadings contain the goal that the proposed strategy is meant to serve. The list of proposals is not by any means exhaustive, but it draws attention to some of the most prominent options for organizational and institutional legitimacy enhancement.

Programs

The existing structure, operations, and policies common across police organizations are not inevitable. They are not tied to or required by larger environmental characteristics (such as the size of the jurisdictional population) or by agency size (Langworthy 1986). Police executives have a choice in the problems they wish to target and in the programs they adopt to meet those goals. That being said, the programs discussed below are deliberately framed as being moderate in scale so as to harmonize with reigning myths and thereby avert legitimation crises. Each proposed program, and the specific problem it is suggested to address, is described below.

Community Policing to Enhance Quality of Life

Community policing arose as a backlash against the detached, forceful, strictly centralized nature of the traditional model. The community model was intended to make policing more democratic by conceptualizing local residents as stakeholders and giving them a voice in police operations. This program is thought to enhance police legitimacy by showing the public that the police are responsive to them. This entry's treatment of the subject balances two perspectives. The first involves the premise that community policing is nothing more than an elaboration of questionable assumptions about the nature of "community," about the relationship between policing and informal social control (Manning 1984), and about the extent to which police can exert actual (as opposed to apparent) control over social behavior (Manning 1988). The second is more pragmatic, if less philosophically coherent, than the first: Community policing is already institutionalized, and its absence is itself a threat to an organization's legitimacy. That being the case, the following discussion focuses on potential strategies and tactics that might lend this program an element of task effectiveness.

Community policing targets those who abide by the law rather than those who break it. One of the program's stated goals is to make police accessible and responsive to the public and to thereby reduce people's fear of crime and enhance their quality of life. The implications of community policing for (task-based) institutional legitimacy emanate from the research showing that community policing can have positive outcomes on people's emotional states, their assessments of their neighborhoods, and their satisfaction with police. Reisig and Parks (2004), for example, found that people who thought that the police worked with their neighborhoods to solve problems and that viewed their neighbors as being willing to work with the police perceived fewer incivilities and felt safer in their areas of residence. Furthermore, perceptions of community policing and coproduction nullified the deleterious impacts of neighborhood disadvantage on perceived incivilities and personal safety. It is worth noting that there was no consideration of the effectiveness of community policing at reducing crime; it was the mere act of visible police involvement that influenced citizens' quality of life. Other studies of community policing have employed various methods and variables and have arrived at mixed conclusions (see Rosenbaum 1994, for a review), likely due, in part, to a lack of uniformity in programs and methodologies.

Perhaps the empirical evidence regarding the impact of community policing on quality of life can best be summarized in three precepts. First, community residents have to know that the police are reaching out to them; the efficacy of this program depends upon citizens seeing the efforts that police are making. Second, citizens must be satisfied with what the police are doing and with the means by which they are seeking to accomplish those ends. To reduce fear, police must employ community-based efforts that inspire confidence and trust from citizens. Third, the implementation effort must be genuine and involve real structural and operational change. A halfhearted program may help an organization project the "proper" image as a department that engages in community-based practices, but it will be an entirely value-based effort with no impact on task performance (see Mastrofski and Ritti 2000).

Place-Based Problem-Oriented Policing to Reduce Crime

Problem-oriented policing (POP; e.g., Goldstein 1990) possesses the inherent advantage of conforming to the myth of the crime-fighting mandate. The merit of POP's end goal is self-evident and requires no additional rationalization to the public, politicians, or other constituencies. The worthiness of its means does require justification because they depart from – and even conflict with – the normative model of policing as driven by calls for service, or what critics have derisively called the "tyrannical force" of 911 (Goldstein 1990: 19). Repeatedly underscoring the crime-based focus of POP, though, would likely help a police organization justify the redistribution of resources and the alleviation of some officers from calls-for-service duty so that they can devote their time to the scanning, analysis, response, and assessment (SARA) model that has become the catchphrase of POP.

Problem-oriented policing furthers the task-based institutional legitimacy of police because this program seems to be promising for effectively reducing crime. Braga and Bond (2008) found that even a cursory or "shallow" (p. 585) SARA-based problem-solving effort by a municipal police department significantly reduced serious crime (as measured by calls for service) and several types of disorder (as recorded during systematic social observation) in the areas receiving the POP treatment. These reductions in the target areas were not accompanied by significant increases in problems in other beats in the city. Also, place-based situational crime reduction (e.g., fencing off vacant lots, installing street lighting or surveillance cameras) demonstrated a stronger impact on reported crime than did increases in misdemeanor arrests or social service provisions (see also Lum et al. 2011).

Policies

The previous section proposed two programmatic models that hold promise for enhancing the institutional legitimacy of police. The current section describes three organizational policies that may guide officers toward positive behaviors and away from negative ones. Policies delimit the boundaries of acceptable behavior, set expectations for performance, and prescribe penalty schedules for poor conduct. It is worth noting that although the policies described here are formal in nature, their success depends upon them being reinforced informally by police executives and managers during the day-to-day communications and operations within the organization.

Prohibiting Nonfeasance to Promote Collective Responsibility

One of the most infamous aspects of the traditional police culture is the reluctance or outright refusal of officers to come forward when they have incriminating knowledge about their workmates. Within this "code of silence" or "blue wall of silence," there is a hesitance to report instances of misbehavior and to provide information during active investigations (Klockars et al. 1997). The "bad apple" myth that ascribes total blame for misconduct to the offending officers dishonestly covers up the fact that in all likelihood, somebody (possibly several people) knew what was happening and failed to take action to stop it. A strong code of silence can contribute to legitimation crises, as the public comes to see the police badge as a symbol of nepotism, corruption, and impunity.

Police organizations should thus openly and actively address the problem of nonfeasance. It should be made clear that failing to report someone else's misbehavior is as morally and ethically indefensible as is engaging in such action oneself, and that acts of nonfeasance will be discovered and penalized. Routine investigations that are mandated by department policy (e.g., after any officer-involved shooting) and investigations into possible misconduct should seek to determine not only the propriety of the acting officers' behaviors but – if the action is deemed improper – whether other officers failed to intervene to prevent it, failed to report it, or actively assisted in its attempted concealment.

Beyond establishing formal policies penalizing nonfeasance, police leaders should seek to reshape the cultural landscape to effect rejection of the code of silence. In its stead, leaders should foster a culture of mindfulness (King 2009) that instills within every officer a sense of collective responsibility (see Harcourt 2004). Officers should be socialized to see the bad acts of fellow officers as a threat to the safety, authority, and reputation of all other officers and to the image of the department.

Interaction-Based Procedural Justice to Enhance Public Trust and Voluntary Compliance

The two foregoing policies involve intraorganizational methods of promoting positive police behavior and preventing or penalizing misconduct. A third goal involves enhancement of public trust in and deference to the police. This policy entails strategic management of the ways in which officers interact with civilians. This might be analogized to an "officer persona" that emphasizes certain key elements of officer-civilian interaction that should be present in all encounters with members of the public.

People who trust the police are less likely to believe that officers engage in illegal or unethical behavior, and they are more willing to empower them to utilize discretionary decision making (Sunshine and Tyler 2003). The need for autonomy in the day-to-day decisions made at the street level is one of the institutional myths that police have created as a means of insulating themselves from public scrutiny (Crank and Langworthy 1992). A belief in the power and authority of police, moreover, helps ensure the "continued deference of citizens to police authority in the absence of specific demands, commands, or laws" that is so central to the continued existence of police (Manning 1988: 29).

A sizable body of research demonstrates the importance of interaction-based procedural justice on people's support for the police and willingness to obey their directions and commands. Encounter-based justice refers to people's perceptions of the favorableness of the outcome they receive at the close of the encounter (e.g., a citation versus a warning) and the fairness of the decision-making process by which police arrive at the conclusion that the administered outcome is the most appropriate of all options available. Research has shown that process-based fairness rivals or, in some cases, even outweighs outcome favorableness in shaping people's judgments about police (Rosenbaum et al. 2011b; Sunshine and Tyler 2003). This has become known as the procedural justice model of policing.

Procedural justice entails two types of assessments that people make about the actions of officers (e.g., Gau 2011). The first pertains to whether officers treat citizens respectfully and allow them to voice their version of the facts under dispute. The second revolves around people's beliefs as to whether officers make decisions on the basis of facts and not upon their personal beliefs, biases, or whims. This type of procedural justice can also be conceptualized as transparency of decision making. Procedural justice is strongly linked with people's trust and confidence in police and their beliefs that police possess the moral authority to represent social order and to enforce the law as needed to regulate public spaces.

Procedural justice during police-public contacts can enhance the likelihood of respect and compliance by citizens. Studies suggest that police disrespect toward civilians (i.e., a violation of procedural justice expectations) increases citizen disrespectfulness toward officers (Dai et al. 2011) and makes it less likely that citizens will comply with officer commands (Mastrofski et al. 1996; McCluskey et al. 2000). Officer respectfulness (McCluskey et al. 2000) and consideration of the citizen's point of view (Dai et al. 2011) enhance the likelihood of compliance. A procedural justice policy, therefore, could result in better police-civilian interactions and in fewer threats to officer safety and authority.

The procedural justice model of policing represents a subtle but firm subversion of the crime-fighting myth. The latter maintains that the amount of legitimacy possessed by the institution is contingent upon clearance rates, crime rates, and deterrent impact. Tyler (2006) has called this the "instrumental perspective" (p. 3). Research shows, however, that people's trust in police is based less on instrumental concerns about police effectiveness than on the perception that police treat people fairly and respectfully during encounters (Sunshine and Tyler 2003; Tyler 2006). The myth of police as detached, legalistic crime fighters may, moreover, actually be counterproductive during face-to-face interactions. Mastrofski et al. (1996) found that use of force by an officer during an encounter made it less likely that the citizen would comply with the officer's commands. The crime-fighting myth, therefore, can actually impede effective and efficient police work, while simultaneously failing to maximize institutional legitimacy. It is time for police leaders to transition from traditional measures of effectiveness (e.g., numbers of arrests and citations) toward new assessment methods, such as community satisfaction surveys, that capture a more holistic picture of the ways in which officers interact with the public (Rosenbaum et al. 2011b).

Selection and Training

Any police organization is only as good as the people it hires and the procedures with which it trains them. Demanding selection standards help weed out those who are at risk for poor performance or serious misconduct. Training introduces new recruits to the rules, ethics, and cultural values of the organization. This section proposes two strategies for helping to ensure that personnel are of high quality. As in the foregoing sections, this list is merely an example of a few of many strategies that could and should be explored.

Education Requirements to Professionalize Policing

The rationale for recommending that police organizations implement a degree requirement rests on the distinction between a profession and a trade or craft (e.g., Wilson 1968). Professions are regulated occupations that require applicants to meet formal educational requirements before they become eligible for the job. Crafts, by contrast, have "no body of generalized, written knowledge nor a set of detailed prescriptions as to how to behave—it has in short, neither theory nor rules" (Wilson 1968: 283). This idea is analogous to the debate over whether the hiring process should seek merely to select out poor candidates or, conversely, whether the focus should be on identifying and selecting in the most qualified. A profession selects in, while a craft merely selects out.

Policing has been a craft since its inception in the United States; its hiring requirements have traditionally been low, and it has relied heavily upon on-the-job training (Wilson 1968) and has even been infused with an anti-intellectual slant (Manning 2009) that idolizes its internally developed institutional understandings of crime and offenders and rejects those derived through scientific means. Results from the 2007 Law Enforcement Management and Administrative Statistics (LEMAS) survey show that 80.1 % of agencies require only high school or GED completion, 7.9 % require some college but no degree, 9.4 % mandate a 2-year degree, and a mere 1.3 % require a 4-year degree (the remaining 1.2 % have no formal policy regarding education). Coupled with the relatively short academy training period (discussed below) that emphasizes tactical skills over more abstract concepts, low education requirements mire policing in the realm of a trade and prevent its ascension to the status of a profession.

Training to Professionalize Policing

Police scholars, leaders, and policymakers lament the short, superficial training that most of the new recruits currently experience. Academies average approximately 761 h, or 19 weeks, in duration, and classes are taught primarily by part-time instructors. The focus is almost exclusively on physical fitness, firearms competence, self-defense capability, patrol tactics, and other technical aspects of the job; relatively little attention is given to human relations, communication skills, community policing tactics, domestic disputes (Reaves 2009), and other topics that relate directly to the vast majority of actual, day-to-day police work. Most instructors, furthermore, are police officers (Reaves 2009), which raises questions about whether trainees are exposed to a wide range of ideas and information or, rather, are being given a relatively narrow glimpse into the nature, purpose, and function of the job.

Inadequate training can help spark legitimation crises. This is because poor training has two predictable outcomes: (1) Officers' job performance will be compromised, and (2) the institution will eke by on marginal amounts of legitimacy. These outcomes are obviously interrelated, the second being, in part, a product of the first. Academy and in-service training should more accurately reflect the reality of the job. This will better prepare recruits for the street. It could also help reduce the infamous tendency for field training officers to preface their socialization of new recruits with an admonition to "forget everything you learned in the academy," which pits one role orientation against another. Role ambiguity or conflict contributes to officer misbehavior (Skolnick and Fyfe 1993), so restructuring training to be more realistic and to equip officers with the skills they need in actual encounters with civilians could significantly improve officer competence and effectiveness.

Managing Media Relations and Public Images

Police organizations play an active role in shaping their public images; indeed, many of the myths that currently govern the institution were created by the police themselves as a means of convincing the public that it needs police and needs them to look and act a certain way. The myth that crime is a looming threat and that its only solution is for criminals to be caught and punished, combined with the institutional myth of police as the ones who do said catching (Crank and Langworthy 1992) and the nature of police organizations as state-sanctioned entities with a guaranteed annual cash flow (Zucker 1987), might seem to be sufficient to legitimize police in the public eye and to obviate the need for additional efforts. What is missing, though, is the guarantee that the public knows that police are displaying the symbols and carrying out the rituals that make the citizenry think that the police are doing their job correctly. Media and public relations are a "missing link" between police behavior and public knowledge about and perceptions of that behavior.

Proactive Dissemination of Information to Improve Image

Most people have very little direct contact with police. According to the 2008 Police-Public Contact Survey, only 16.9 % of persons aged 16 and older had experienced face-to-face contact with police within the past year. The modal type of contact was a traffic stop (47 % of those with any contact). Traffic stops are typically perfunctory affairs that afford civilians little opportunity to form strong or substantive opinions about officers' job performance or personal character.

The impact of vicarious (i.e., indirect) experience is thus crucial to the public's formation of attitudes about and opinions of the police. Rosenbaum et al. (2005) found that between 26 % and 58 % of respondents in their sample reported being exposed to media accounts depicting other people's direct contacts with police. White respondents were more likely than Black or Hispanic respondents to report media exposure and less likely to have had direct experiences. Media exposure is related to people's attitudes toward police (Rosenbaum et al. 2005).

Strategic utilization of media opportunities can foster a symbiotic relationship. The police rely on the media as a conduit through which the community can be reached, while reporters need crime news that is easy to arrange into a print- or broadcast-ready format (Chermak and Weiss 2005). Regular press releases apprising the media of new events and of updates to ongoing matters make a police organization a valuable source of information to the media. Reporters, not wanting to risk losing this source, may avoid printing/broadcasting vitriolic attacks on or unfounded allegations about the department or its officers.

Proactive information release also capitalizes upon the difference between an interviewer and an interviewee, a distinction that is important for appearances (Kingshott 2011). The interviewer assumes the dominant position as the gatekeeper (Surette 2001) of the types and amounts of information exchanged; by contrast, the interviewee sits passively and fashions impromptu responses to questions that might be loaded, critical, or perhaps tap a content area about which the interviewee has no knowledge. When organizational representatives are providing information, they appear competent, knowledgeable, and authoritative. When fielding questions, though, they can easily come across as ill-informed, inept, or deceitful (Kingshott 2011).

Police organizations – via PIOs and community policing officers – should also utilize the broad spectrum of technologies available for pushing information out to the community and pulling input from civilians into the department. Websites containing detailed information about the department (e.g., beat maps, functional differentiation), Web links to local social service agencies, crime data, news about upcoming or recently held community policing events, and other relevant content can help organizations be interactive with the community (Rosenbaum et al. 2011a). Twitter, Facebook, email listserves, and text messaging can likewise be utilized to maintain an active outward flow of communication, including but not limited to "reverse 911" alerts and safety tips. Similarly, agencies should strive to pull in information from residents in the form of online options for filing crime reports, contacting relevant persons or units within the department, submitting complaints against officers, and offering praise for services well rendered (Rosenbaum et al. 2011a).

Civilian Review Boards

The public has become increasingly skeptical of police organizations' ability to consistently monitor their officers, conduct honest investigations into alleged misconduct, and effectively discipline those found guilty of wrongdoing. In particular, there is serious doubt about the effort that organizations exert in investigating citizen complaints. The adage "How can we expect the police to police themselves?" reflects the deep-seated suspicion – particularly prevalent among minority groups – that so long as an investigatory body's locus of operations rests within a police organization, justice will be served sporadically and inconsistently, at best. To maintain and promote institutional legitimacy, organizations need to assemble external accountability mechanisms. Civilian review is one such option.

Civilian review faces the criticism that civilians are not trained as police officers, do not have the mental vantage point of someone who has patrolled the city streets for months or years, and cannot put themselves into the positions of police officers who thought their own or someone else's physical safety was in jeopardy. An analogous critique could be leveled at the jury system. Law is complex, actions can be ambiguous, and evidence is often not overwhelming one way or the other. It might seem odd to suggest that a panel of laypersons – unschooled in law or psychology – could be in any position whatsoever to accurately judge guilt or innocence (Skolnick and Fyfe 1993).

But finely honed truth seeking is not the fundamental purpose of either juries or civilian review boards. The rationale for their existence rests, instead, on the community's right to weigh in on issues of justice and accountability from the vantage point of "regular people." Accountability is a community matter, not merely a legal one (Skolnick and Fyfe 1993). People also like being able to see the process occurring. The most thorough of internal police investigations and the most accurate adjudication regarding guilt or innocence means little to the public demanding information about the process and not merely the outcome.

Civilian review boards may be becoming increasingly normalized within the institutional environment of policing. Adoption of civilian review protocol sends a message of transparency and a willingness to "open the doors" of the organization to a civilian audience. The existence of civilian review may have a calming effect on the public (see Skolnick and Fyfe 1993).

If implemented in such a manner so as to merely pay lip service to civilian review, however, this effort can threaten an organization's legitimacy. Police have vehemently fought civilian review (Skolnick and Fyfe 1993) and a common consequence is that civilian review procedures either afford the board minimal authority or dictate that the board contain officers or ex-officers. Both of these can violate the expectations that the presence of a civilian review mechanism creates among the public. If actual civilian review procedures fall below these expectations, complainants will be dissatisfied.

Conclusion

The institutional prospective (e.g., Meyer and Rowan 1977) as applied to policing (e.g., Crank and Langworthy 1992) holds that police organizations exist within an institutional environment governed by certain values and normative rules. These myths, originally created by police as a means of shaping public opinion and expectations about the police role, took on a life and power of their own and now operate as constraints that require every organization to conform to the archetype of what a "real" police agency looks and acts like. Police agencies rely upon skillful and strategic utilization of these myths to maintain the legitimacy of their own organizations and that of the institution as a whole. As a state-sanctioned, service-oriented institution, policing has no tangible output that can be submitted to the competitive marketplace for a test of effectiveness and efficiency. The police "product" is a service, and it is a service the field must constantly "sell" to the public in order to maintain its status as a legitimate institution.

This entry considered several proposals for enhancing police institutional legitimacy. Programs (community policing and problem-oriented policing), policies (written procedures regarding nonfeasance, early warning and intervention systems, and encounter-based justice), selection and training (education requirements and professional training), and media and public image strategies (proactive dissemination of information and civilian review) were suggested. While all of the proposed activities are themselves grounded in institutional myths that are based more on values than on task performance, it was argued that each could be used to accomplish a certain goal. In other words, certain tasks can be accomplished by certain symbols and ceremonies if those symbols and ceremonies are paired closely with a desired outcome and implemented in an honest and genuine attempt to perform that task well. This would not eliminate the tension between myths and performance, but it could close the gap somewhat while still remaining within the acceptable realm of the archetypical organization and thus effecting change while maintaining and promoting legitimacy.

Related Entries

- ▶ Community Policing
- ▶ Control of Police Misconduct
- ▶ Democratic Policing
- ▶ History of the Police Profession
- ▶ Managing Innovation of Policing
- ▶ Problem-Oriented Policing
- ▶ Procedural Justice, Legitimacy, and Policing

Recommended Reading and References

Braga AA, Bond BJ (2008) Policing crime and disorder hot spots: a randomized controlled trial. Criminology 46:577–607

Chermak S, Weiss A (2005) Maintaining legitimacy using external communication strategies: an analysis of police-media relations. J Crim Justice 33:501–512

Crank JP (2003) Institutional theory of police: a review of the state of the art. Policing An Int J Police Strat Manag 26:186–207

Crank JP, Langworthy R (1992) An institutional perspective of policing. J Crim Law Criminol 83:338–363

Dai M, Frank J, Sun I (2011) Procedural justice during police-citizen encounters: the effects of process-based policing on citizen compliance and demeanor. J Crim Justice 39:159–168

DiMaggio PJ, Powell WW (1983) The iron cage revisited: institutional isomorphism and collective rationality in organizational fields. Am Sociol Rev 48:147–160

Gau JM (2011) The convergent and discriminant validity of procedural justice and police legitimacy: an empirical test of core theoretical propositions. J Crim Justice 39:489–498

Goldstein H (1990) Problem-oriented policing. McGraw-Hill, New York

Harcourt BE (2004) Unconstitutional police searches and collective responsibility. Crim Public Policy 3:363–378

King WR (2009) Police officer misconduct as normal accidents: an organizational perspective. Crim Public Policy 8:771–776

Kingshott BF (2011) Effective police management of the media. Crim Justice Stud 24:241–253

Klockars CB, Ivkovhic SK, Harver WE (1997) Measurement of police integrity. University of Delaware, Newark

Langworthy RH (1986) The structure of police organizations. Praeger, New York

Lum C, Koper CS, Telep CW (2011) The evidence-based policing matrix. J Exp Criminol 7:2–26

Maguire ER (1997) Structural change in large municipal police organizations during the community policing era. Justice Q 14:547–576

Manning PK (1984) Community policing. Am J Police 3:205–227

Manning PK (1988) Community policing as a drama of control. In: Greene JR, Mastrofski SD (eds) Community policing: rhetoric or reality. Praeger, Westport

Manning PK (1996) Politics and the axial media event. Sociol Q 37:261–278

Manning PK (2009) Bad cops. Crim Public Policy 8:787–794

Mastrofski SD, Ritti RR (2000) Making sense of community policing: a theory-based analysis. Police Pract 1:183–210

Mastrofski SD, Snipes JB, Supina AE (1996) Compliance on demand: the public's response to specific police requests. J Res Crime Delinq 33:269–305

McCluskey JD, Mastrofski SD, Parks RB (2000) To acquiesce or rebel: predicting citizen compliance with police requests. Police Q 2:389–416

Meyer JW, Rowan B (1977) Institutionalized organizations: formal structure as myth and ceremony. Am J Sociol 83:340–363

Reaves BA (2009) State and local law enforcement training academies, 2006. Bureau of Justice Statistics, Washington, DC

Reisig MD, Parks RB (2004) Can community policing help the truly disadvantaged? Crime Delinq 50:139–167

Rosenbaum DP (ed) (1994) The challenge of community policing: testing the promises. Sage, Thousand Oaks

Rosenbaum DP, Graziano LM, Stephens CD, Schuck AM (2011a) Understanding community policing and legitimacy-seeking behavior: a national study of municipal police websites. Police Q 14:25–47

Rosenbaum DP, Schuck A, Lawrence D, Hartnett S (2011b) Community–based indicators of police performance: Introducing the Platform's Public Satisfacation Survey. Washington, DC: National Institute of Justice

Rosenbaum DP, Schuck AM, Costello SK, Hawkins DF, Ring MK (2005) Attitudes toward the police: the effects of direct and vicarious experience. Police Q 8:343–365

Skolnick JH, Fyfe JJ (1993) Above the law. The Free Press, New York

Sunshine J, Tyler TR (2003) The role of procedural justice and legitimacy in shaping public support for policing. Law Soc Rev 37:513–548

Surette R (2001) Public information officers: the civilianization of a criminal justice profession. J Crim Justice 29:107–117

Tyler TR (2006) Why people obey the law. Princeton University Press, Princeton

Wilson JQ (1968 [1976]) Varieties of police behavior. Atheneum, New York

Zucker LG (1987) Institutional theories of organization. Ann Rev Sociol 13:443–464

Organizational Crimes

▶ Crimes of the Powerful

Organizational Culture

▶ Police Culture

Organizational Liability

▶ Corporate Liability

Organizational Performance Measurement

▶ Judicial Leadership and Performance

Organized Crime

▶ Racketeering

Organized Crime and the Environment

Reece Walters
School of Justice, Faculty of Law, Queensland University of Technology, Brisbane, Australia

Synonyms

Organized eco crime; Organized environmental crime

Overview

> Interpol demands crackdown on "serious and organised" eco crime. (Johnson 2012:1)

It has been recognized for some time that the liberalization of trade policies has had deleterious impacts on the world's natural environment. The rapid expansion of globalized goods and services continues to create a human footprint with long-lasting environmental consequences (White 2010). It is a footprint that represents rapid human activity and with it has come new commercial opportunities, not only for global businesses but also for organized criminal networks. Both the acceleration and by-products of global trade have created new markets as well as underground economies. As the opening quotation reveals, transnational environmental crime must become a policing priority as organized criminal networks continue to exploit the environment with unfrequented profits (cf. Banks 2008). Organized crime syndicates reportedly earn between US$20 and 30 billion from environmental crimes (Clarke 2011). Such earnings come at substantial social, economic, and environmental expense for communities, their livelihoods, and habitats. Indeed, organized environmental crime is identified by the UN as a key factor in the impoverishment, displacement, and violent conflicts of millions of people, notably in developing societies (UNODC 2009). The theft of biodiversity and the demise of animal species and habitats have resulted not only in financial loss but in an increase in "environmental refugees," people dislocated and forced to migrate due to loss of livelihoods. This entry will explore the links between organized crime and the environment and examine the regulatory and environmentalist responses to this growing issue of global concern.

Key Issues and Controversies

Defining Organized Environmental Crime

The United Nations Interregional Crime and Justice Institute (UNICRI) has established a now widely used categorization of organized environmental crime based on various international protocols and multilateral agreements. Other organizations including Interpol, the UN Environment Programme, and G8 also adopt the following five key areas when referring to transnational and organized environmental crime. The areas are the illegal trade in endangered species and wildlife; the illegal trade in ozone-depleting substances; illegal dumping of waste and hazardous substances; illegal, unregulated, and unreported fishing; and illegal logging and trade in protected woodlands (see Hayman and Brack 2002:5).

Illegal Trade in Flora and Fauna

It is widely recognized that organized environmental crime syndicates, motivated by substantial financial rewards, continue to flourish and expand in disadvantaged societies with porous borders where corruption is widespread and regulation poor (UNODC 2009). Emerging international laws and environmental policing efforts are gradually beginning to engage with issues inextricably linked to legitimate global trade. The impacts of these emerging markets continue to decimate flora and fauna while having widespread impacts on human populations. For example, the CITES protects about 5,000 species of animals and 28,000 species of plants, yet 4,000 African elephants per year continue to be killed for illicit trade (Interpol 2011). Banks et al. (2008:7) identify that since the 1990s, "80 % of timber coming out of Indonesia was illegal, and the government has estimated that it costs the nation US$4 billion a year." In Thailand and the Malaysian peninsula, recent reports identify that the Sumatran rhinoceros has become extinct due to organized criminal poaching. Moreover, the numbers of tigers, elephants, saiga antelopes, and anteaters have become "dangerously low" (Viegas 2011). In the United States, illegal trafficking in wildlife continues to accelerate at an alarming rate and has caused the State Department to establish a Coalition Against Wildlife Trafficking comprising government, protest groups, and corporate partners to address an industry of organized crime (US State Department 2011). The Wildlife Conservation Society in New York continues to emphasize the seriousness of the issue stating "we are rapidly losing big, spectacular animals to an entirely new type of trade driven by criminalised syndicates, and the world is not yet taking it seriously" (Coghlan 2011). The situation is equally serious across the Atlantic. During 2004 alone, European Union officials reportedly made 7,000 seizures involving "more than 3.5 million wildlife specimens that were prohibited from being traded" (WWF 2011:1).

Endangered Species International concludes that "despite international laws, animal body parts of protected species are traded around the world. For example, body parts of hawksbill turtle shells, shahtoosh shawls from the Tibetan antelope, and furs from rare otters in south east Asia are found in both black and open markets. Of the estimated 350 million animals and plants being traded worldwide every year, it is believed that 25 % is carried out illegally" (ESI 2011:1). The prices for exotic and protected species vary on the black market from tens of thousands of dollars for a macaw, to just a few dollars for a giant cockroach (PETA 2011). Recent seizures in Vietnam near the Chinese border revealed over a 1,000 t of African elephant ivory smuggled in bundles of cloth (Traffic 2011). There are numerous other examples of declines or extinction of native fauna from organized illegal activity. As Antonio Maria Costa, Executive Director of the United Nations Office on Drugs and Crime, has stated,

> People are profiting from the destruction of our planet, by dumping hazardous waste, illegal logging, or the theft of bio-assets. This crime not only damages the eco-system, it impoverishes so many countries where pollution, deforestation and population displacement trigger conflict.... (Costa 2008)

While the social, economic, and political conditions of "origin nations" undoubtedly facilitate organized environmental crime, issues of demand and global trade are noticeably absent from international discourse. Bricknall (2010:7) has previously noted that EU countries take a "soft approach" with organized environmental crime, often because the perpetrators, or those involved, are reputable business personnel engaging in essential trade. She concluded, "...enforcement agencies thus are not confronted with the classical "bad guys" but ... with often highly respected players of economic life." For many, globalization has provided the context for the "easy passage of illegal goods."

Organized environmental crime continues to flourish because of trade and market demand. Countries such as the United Kingdom, which actively promote international environmental treaties to preserve and protect natural heritage, provide the markets for organized crime syndicates to dispose of their illegal merchandise. For example, Britain is the world's third largest importer of illegally logged timber. Up to 3.2 million cubic meters of timber sold in the UK and used for household furniture or garden woodchip is stolen from the Amazon rainforest and other protected habitats, and comprises a £700 million per year British industry. Moreover, imported fish is worth £4–9 billion to Britain per annum. By conservative estimates, more than 12,000 t per annum originates from illegal fishing in the offshore waters of poor countries. This illegal fishing decimates the local industry and food supply of debt-stricken countries in Western Africa while destroying marine biology. Yet, unregistered pirate vessels enter British ports unchecked, and the stolen fish are sold at London markets without question (Walters 2012).

Internationally recognized businesses often facilitate the commercial trade in illegal and endangered species. The Internet provider "Yahoo" has reportedly been providing shopping sites in Japan for the sale of more than 130 t of endangered Icelandic whale (EIA 2011). Moreover, environmental activists have lobbied eBay to ban the sale of endangered species after 7,000 wildlife products were identified on 183 websites across 11 countries (IFAW 2011).

Illegal Disposal of Waste

Not only have endangered and protected species of flora and fauna proved lucrative for crime syndicates, but so has the illegal disposal of waste. The excesses of globalized consumption and capitalism continue to produce harmful wastes that pollute and contaminate the environment. As mentioned above, the dumping and illegal transport of various kinds of hazardous waste is widely recognized by international law as an environmental crime. However, for decades, the control and enforcement of the illegal movement in waste has proved inadequate as Rebovich (1992:125) notes,

> ...it has been said that illegal hazardous waste disposal is very much like one long game of hot potato. The idea is to make as much profit as you can by being the temporary possessor of the hot potato before unloading it on some other person, organization, or place. In the end, the final recipient is the loser.

The "losers" in the illegal movement of waste are most often the poor and vulnerable. There continues to exist within political and policing circles, noticeably across Europe, complacency, and lack of awareness regarding the seriousness of illegal waste disposal and its links to organized crime. Nowhere is this more apparent than in Italy, which is worth exploring further here in some detail to understand organized environmental crime and its links with global trade and national governance.

Throughout the 1990s, international headlines reported 11 million tons of industrial waste unaccounted for and large amounts of toxic waste dumped in Italy in what came to be known as acts of the "ecomafia." Mafia-related enterprises were reported to be monopolizing waste disposal contracts from industries producing toxic residues and illegally dumping the pollutants in various areas of the Italian countryside (Ruggierio 1996). Furthermore, the mafia or "rubbish tzars" (Ruggiero and South 2010) were bidding for, and successfully obtaining, provincial contracts to clean up the very environmental mess they themselves had created. This sort of organized criminal activity is not new. Block and Scarpetti (1985) documented the ways in which the mafia in the USA monopolized the solid waste management business, eliminated industry competition, bribed officials, and routinely illegally disposed of toxic waste in the New York and New Jersey region. Interestingly, they note how the US Environment Protection Agency was obstructive in regulating toxic waste. Indeed, Szasz (1986) argues that "lax implementation and enforcement" were key factors in the expansion of organized crime monopolies over waste management contracts and the concomitant illegal activity – an insight pertinent to southern Italy's procurement of waste contracts and lack of regulatory oversight.

Italy continues to hold the worst environmental infringement record in the EU. In early 2002, a total of 125 breaches of EU environmental directives were lodged against the Italian authorities with some cases referred to the European Court of Justice.

In December 2006, the European Parliament identified that 60 environmental infringement notices remained outstanding against the Italian government (the highest in Europe), mainly for breaches of waste management. In 2009, the European Environment Commissioner Stavros Dimas stated "EU environmental law aims to prevent damage to the environment and minimise health risks to European citizens. To ensure its citizens are provided the utmost protection I urge Italy to quickly put right the shortcomings of certain of its environmental laws in line with those of the EU" (Europa 2009:1). Italy has yet to implement seven different EU environmental directives relating to water, air, soil, waste, and nature protection, and its legal regimes are often severely criticized for not harmonizing with EU law. The European Commission is pursuing legal action against the Italian government for failing to implement environmental directives into national law.

One of the reasons for Italy's noncompliance with EU environmental regulations is, as mentioned above, the widespread organized criminal activity of the "ecomafia." As South (2010:234) rightly points out, the word "ecomafia" first coined by the environmental group Legambiente is now widely accepted in Italian society to mean "organised criminal networks that profit from illegally disposing of commercial, industrial and radioactive waste." Ecomafia is big business in Italy, its worth being estimated at 20.8 billion Euro in 2008 (Legambiente 2009). Almost all criminal activities occur in the mafia strongholds of Campania, Calabria, Sicily, and Puglia. In this area alone, 31 million tons of domestic and commercial waste simply "disappeared" in 2008, dumped at sea or in local waterways. More recently, mafia groups have taken to illegally burying waste in southern Italy and then rapidly building housing estates on top. Between 2008 and 2010, 17,000 houses were illegally built on waste dumps, and 10,000 forest fires were mafia related (Legambiente 2010). The impacts of organized environmental crime in Italy also have a global reach. In 2008, it was widely reported that mozzarella cheese exported from Campania contained high levels of dioxins, the result of dairies contaminated by the illegal disposal of toxic waste (Walters 2012).

Italy's long-standing record of environmental noncompliance combined with the prolific illegalities of the ecomafia that continue to generate media headlines has necessitated a political response. Within the Carabinieri (reputedly Italy's most elite law enforcement body) has emerged a specialist policing unit to tackle environmental crime. It should be noted that Italy has a complex structure of state policing and law enforcement. There are eight separate law enforcement agencies, all combining to govern and enforce federal, provincial, and municipal law across Italy (Walters 2012).

The creation of a specialist unit within an existing police force to tackle corporate environmental crime is a first in Europe. Yet, the resources and police personnel devoted to the initiative are minuscule. Moreover, it is the provinces in the north of Italy that have witnessed the greatest political and municipal "buy-in," while the troubled areas of the south continue to struggle for interagency collaboration against a culture of suspicion and official corruption.

Emerging from Italy's eco-policing, which is still in its early days, is a story of how environmental movements can identify and target organized crime, mobilize political opinion, involve local government, and raise public awareness. In Italy, organized environmental crime is not being addressed or tackled by senior political officials, government administrators, or policing agencies; they all play a part, but the real difference is being made through Legambiente. Established in 1980, Legambiente is a left-wing environmental activist organization with 115,000 active members across 45 offices in Italy. With the use of its technologies, databases, and local intelligence, Legambiente has been instrumental in tightening waste disposal regulations and in the prosecution of mafia personnel. In a similar fashion to the ecomafia having a public identity in Italy, Legambiente is widely thought of as the "eco-police." However, Italy's four biggest mafia groups, namely, the "Ndrangheta" in Calabria, the Sacra Corona Unita in Apulia, and the Neapolitan Camorra and the Cosa Nostra in Sicily, are so embedded in the social and economic fabric of Italian society, in that organized criminal activity accounts for 7 % of the country's gross domestic product (Phillips 2008) with the ecomafia competing among themselves for waste management contracts, amidst poor and corrupt regulation. Recent endeavors by Legambiente and its allies to highlight the extent of corruption in Italy's south has intensified EU focus on what has been dubbed the "Naples rubbish crisis" where 7,200 t of rubbish is accumulating every day in the Campania region (Ruggiero and South 2010; Walters 2012).

International Perspectives

The severity of organized environmental crime has a long history. Notably, it was environmental groups in the mid-1970s that raised awareness of this activity, and it is green and social movements that are central to current efforts of policing and regulation (Block 2002). Environmental activism plays a major role alongside international instruments in combating organized environmental crime (White 2010). Such initiatives include the Interpol Environmental Crime Committee which focuses on training, data collection, and enforcement related to pollution and wildlife crimes and implements an "intelligence-led policing" model which emphasizes "operational partnerships." In 2010, during the UN's International Year of Biodiversity, Interpol's General Assembly passed a resolution that would see 188 national law enforcement agencies agreeing to collaborate with organizations such as the World Bank and environmental movements, an initiative that promises a substantial increase in policing resources to reduce organized environmental crime (Interpol 2011). In 2007, countries signatory to the United Nations Millennium Development Goals agreed to "ensure environmental stability," through, inter alia, targeting and preventing organized environmental crime (United Nations 2007). The proceeds of organized environmental crime are laundered through legitimate commercial activities. As a result, the Financial Action Task Force now officially recognized environmental crime as explicitly associated with money laundering. The Asian Regional Partners Forum on Combating Environmental Crime (ARPEC) was set up in 2005 and continues to play an active role in coordinating enforcement endeavors in the Asia Pacific region, an area renowned for illicit trade in wildlife. One important initiative, the Partnership Against Transnational Crime through Regional Organized Law Enforcement, has witnessed significant increases in wildlife seizures via coordinated policing and information exchange. Finally, in July 2011, eight countries signatory to the Lusaka Agreement on Co-operative Enforcement Operations Directed at Illegal Trade in Wild Fauna and Flora (promoted by wildlife law enforcement officers from eight Eastern and Southern African countries) met to increase enforcement resources to prevent wildlife crime (Walters 2012).

In a similar fashion to the success of citizen involvement in environmental activism in the historically progressive regulation of environmental crime in the United States (Clifford and Edwards 2012), the above international approaches expressly rely upon organizations such as Greenpeace, Endangered Species International, Environmental Investigation Agency, Environmental Justice Foundation, Legambiente, and the World Wide Fund for Nature to combat global organized environmental crime. Environmental movements are becoming central in the identification, detection, and prevention of environmental crime. Their resources, technologies, databases, and personnel are increasingly utilized by law enforcement agencies to police, regulate, and prosecute organized environmental crime. Here, environmental activism, through technology and networks of action, local alliances, as well as appeals to citizens and officials, elevates the social movement to a reliable and reputable status that is inculcated into government and regulatory structures. Environmental activism becomes not mere representative democracy but participatory democracy with both a visible presence and impact. As such, with public and political integration, it becomes a new and important form of environmental governance. The momentum created by environmental movements is a source of mobilized power, emerging from alliances with local institutions of governance, and knowledges, that have come to be relied upon as accepted and trusted regimes of truth – in turn becoming the sources of official discourse.

Future Directions

Within criminological studies, debates about organized environmental crime have emerged within discourses on state and corporate crime or "crimes of the powerful" and within the rapidly expanding area of "green criminology" (South 1998; South and Bierne 2006). South (2010:242) argues that the growing phenomenon of green crime is a product of a late-modern risk society. As a result, he identifies emerging environmental harms and injustices as requiring "a new academic way of looking at the world but also a new global politics." This includes an intellectual narrative that moves "beyond the narrow boundaries of traditional criminology and draws together political and practical action to shape public policy" (South 2010:242). This interdisciplinary approach combined with environmental activism and public policy has facilitated a green criminological perspective with three dimensions: first, scholarship that conceptualizes environmental crime; second, that devoted to exploring and uncovering various types of environmental crimes; and finally, a commitment to environmental policing and enforcement.

A substantial amount of green criminological scholarship seeks to put issues of environmental harm on the academic, political, and public radar. As this entry argues, too often when actions violate international environmental agreements or domestic laws they are referred to as "breaches" or "offenses" and not crimes. From a purely legal perspective, this is best explained by the fact that environmental offenses are often not contained within either international or municipal criminal law. As such, they are dealt with as administrative offenses and prosecuted in civil jurisdictions. Such offenses only become issues for the criminal courts when offenders fail to comply with a court sanction and are subsequently referred to a criminal court. While the language of eco-crime is used (most often by activists and NGOs), it is not expressed in such terms at international law and only in reaction to antisocial behavior within domestic law. Eco-crime is an area requiring new conceptions of offending and regulatory responses within this “new academic way of thinking and within a new global politics” (South 2010:242).

Organized environmental crime often occurs within, and is intrinsically linked, to the free market policies of state and corporate trade. For Westra (2004:309), acts by governments and corporations in pursuit of free trade that deliberatively destroys and damages biodiversity are “attacks on the human person” that deprive civilians (notably the poor) from the social, cultural, and economic benefits of their environment. As a result, eco-crime is an act of violence and should be viewed as a human rights violation as citizens are deprived of freedoms and liberties. As Halsey and White (1998) have noted, “environmental harm” is often publicly and politically accepted as necessary for maintaining human well-being. To use a Gramscian analysis, capital accumulation and the prominence of trade are preserved through ideologies of “necessity.” There is a cultural hegemony that underpins the imperatives of trade that come about through consensus or “common sense values” that cannot be undermined. As a result, it is essential to apply a political economy analysis to understand the interconnectedness between organized environmental crime and legitimate global trade.

To understand the complexities of organized environmental crime require an examination of the networks of corruption that facilitate criminal markets (Elliot 2009). Lorraine Elliott asserts that addressing this expanding and global enterprise requires “joined-up thinking” across various transnational government and nongovernment agencies. Notwithstanding the importance of this form of network analysis, it is must be recognized that policies of free trade governed by principles of market regulation provide the contexts in which organized environmental crime may flourish. The role of green criminologists must be to unpack and disentangle the ways that policies and practices of legitimate trade facilitate the opportunities and activities of organized environmental criminal networks.

Conclusion

This entry concludes that the natural environment continues to provide immense profits for organized criminal networks while decimating protected species. Moreover, it is the world’s poorest and most marginalized people most severely affected by the cultural, social, and economic consequences of transnational organized environmental crime. While trade remains an international priority within the landscapes of global economics and fiscal prosperity, organized environmental crime takes advantage of growing markets. The trade imperatives of globalized and liberal market economies provide the challenges backdrop and contexts for both domestic and international policing initiatives. On one level, policing resources are either unavailable or limited for eco-policing; at another level, a lack of political will ensure that regulation and enforcement remain a low priority. As a result, innovations in policing partnerships, which involve environmental activist groups, have an important role to play. Movements of environmental activism are emerging as the new front in the surveillance, regulation, and prosecution of organized environmental crime. Such voices must continue to be central to future green criminological perspectives that seek environmental, ecological, and species justice.

Related Entries

- ► Environmental and Human Rights
- ► Green Criminology

Recommended Reading and References

Banks D, Davies C, Gosling J, Newman J, Rice M, Wadley J, Walravens F (2008) Environmental crime: a threat to our future. Environmental Investigation Agency, London

Block A (2002) Environmental; crime and pollution: wasteful reflections. Soc Justice 29(1–2):61–81

Block A, Scarpetti F (1985) Poisoning for profit – the mafia and toxic waste in America. William Morrow, New York

Bricknell S (2010) Environmental crime in Australia. Australian Institute of Criminology, Canberra. http://www.aic.gov.au/documents/2/1/1/%7B211B5EB9-E888-4D26-AED4-1D4E76646E4B%7Drpp109.pdf

Clarke E (2011) Environmental crime and national security, United Nations environment programme, environmental crime and security workshop, Antigua, 2 March 2011

Clifford M, Edwards T (2012) Environmental crime, 2nd edn. Jones and Bartlett Learning, Burlington

Coghlan A (2011) Mob move into organised wildlife crime, New Scientist. http://www.newscientist.com/article/dn20754-mob-move-into-organised-wildlife-crime.html. 2 Aug 2011

Costa A (2008) Rule of law: a (missing) millennium development goal that can help reach the other MDGs. United Nations Commission on crime prevention and criminal justice, 17th Session, Vienna. http://www.unodc.org/unodc/en/about-unodc/speeches/2008-04-14.html. 14 April 2008

Elliot L (2009) Combatting transnational environmental crime: joined up thinking about transnational networks. In: Kangaspunta K, Marshall I (eds) Eco-crime and justice: essays on environmental crime. UNICRI, Turin

Endangered Species International (2011) Stop illegal wildlife trading project. http://www.endangeredspeciesinternational.org/project_illegaltrade.html

Environmental Investigation Agency (2011) New fin whale export by unrepentant Iceland. Yahoo! urged to stop selling the meat of endangered whales, http://www.eia-international.org/new-fin-whale-export-by-unrepentant-iceland

Europa (2009) Italy: commission pursues legal action over breaches in environmental laws. http://europa.eu/rapid/pressReleasesAction.do?reference=IP/09/419&format=HTML. Accessed 31 Aug 2009

Halsey M, White R (1998) Crime, ecophilosophy and environmental harm. Theor Criminol 2(3):345–372

Hayman G, Brack D (2002) International environmental crime: the nature and control of environmental black markets: workshop report. http://www.chathamhouse.org.uk/files/3050_riia_environmental_crime_workshop_report.pdf

Heckenberg D (2010) The global transference of toxic harms. In: White R (ed) Global environmental harm: criminological perspectives. Willan, Devon, pp 37–61

IFAW (2011) Fight illegal wildlife trade. http://www.ifaw.org/ifaw_international/join_campaigns/fight_illegal_wildlife_trade/index.php

Interpol (2011) Wildlife crime. www.interpol.int/Public/EnvironmentalCrime/Wildlife/Default.asp

Interpol (2012) http://www.interpol.int/Crime-areas/Environmental-crime/Environmental-crime

Johnson L (2012) Interpol demands crackdown on 'serious and organised' eco crime Ivory poaching and illegal logging requires needs tougher enforcement and intelligence input, says Interpol director, The Guardian, 29 March 2012

Legambiente (2009) EcoMafia. Legambiente, Rome

PETA (2011) Inside the exotic animal trade. http://www.peta.org/issues/Companion-Animals/inside-the-exotic-animal-trade.aspx

Phillips J (2008) Modern mafia operates at every level of Italian society. The Independent. http://www.independent.co.uk/news/world/europe/modern-mafia-operates-at-every-level-of-italian-society-799290.html. 22 March 2008

Rebovich D (1992) Dangerous ground – the world of hazardous waste crime. Transaction Books, New Brunswick

Ruggierio V (1996) Organised corporate crime in Europe: offers that can't be refused. Dartmouth, Aldershot

Ruggiero VA, South N (2010) Green criminology and dirty collar crime. Crit Criminol Int J 18:251–262

South N (1998) A green field for criminology: a proposal for a perspective. Theor Criminol 2(2):211–233

South N (2010) The ecocidal tendencies of late modernity: transnational crime, social exclusion, victims and rights. In: White R (ed) Global environmental harm: criminological perspectives. Willan, Devon, pp 228–247

South N, Bierne P (eds) (2006) Green criminology. Ashgate, Aldershot

Szasz A (1986) Corporations, organized crime, and the disposal of hazardous waste: an examination of the making of a criminogenic regulatory structure. Criminology 24(1):1–27

TRAFFIC (2011) Vietnamese authorities confirm more than a tonne of ivory pieces seized on Chinese border. http://www.traffic.org/home/2011/11/4/vietnamese-authorities-confirm-more-than-a-tonne-of-ivory-pi.html

United Nations (2007) Millennium development goals report 2007. www.un.org/milleniumgoals/pdf/mdq2007.pdf

United Nations Office on Drugs and Crime (2009) What does environmental crime have in common with organized crime? www.unodc.org/unodc/en/frontpage/what-does-environmental-crime-have-in-co

United States Department of State (2011) Wildlife trafficking. http://www.state.gov/g/oes/env/wlt/

Viegas J (2011) Organized crime wiping out wildlife, study finds. Discovery News. www.msnbc.msn.com/id/43987628/ns/us_news-environment/t/organized-crime

Walters R (2012) EcoMafia and environmental crime. In: Carrington K, Ball M, O'brien E, Tauri J (eds) Crime, justice and social democracy. Palgrave Macmillan, London

Westra L (2004) Eco violence and the law – supranational normative foundations of ecocrime. Transnational Publishers, Ardsley

White R (2010) Globalisation and environmental harm. In: White R (ed) Global environmental harm: criminological perspectives. Willan, Devon, pp 3–19

WWF (2011) Illegal wildlife trade. http://www.wwf.org.uk/what_we_do/safeguarding_the_natural_world/wildlife/illegal_wildlife_trade/

Organized Crime, Types of

Letizia Paoli
Institute of Criminology, University of Leuven,
Leuven, Belgium

Overview

The concept of organized crime has had a winding history, with many, partially contradictory, meanings attached to it since the expression began to be used in the United States more than a hundred years ago (see, e.g., Varese 2010; for a list of definitions, see also ▶ www.organized-crime.de/OCDEF.htm). Basically, the understanding of organized crime has shifted back and forth between two rivaling notions: (1) a set of stable organizations illegal per se or whose members systematically engage in crime and (2) a set of serious criminal activities, and particularly the provision of illegal goods and services, mostly carried out for monetary gain. (Other criminological concepts, such as white-collar, political, and professional crime, entail similar definitional problems).

Originating from the now discredited official representation of organized crime in the United States in the 1950s and 1960s, the former notion is still embraced by the media and by policymakers seeking to justify draconian measures. Although it is rejected by many scholars, such understanding is corroborated by the existence of a few large-scale and long-lasting criminal organizations worldwide, which Paoli (2002) termed "mafia-type criminal organizations." These are primarily the Sicilian Cosa Nostra and Calabrian 'Ndrangheta, the Italian-American La Cosa Nostra, the Chinese Triads, the Japanese Yakuza, and the *vory v zakone* (thieves-in-law) in the Soviet Union (USSR) and its successor states.

The second view, which is now dominant in the scientific understanding of organized crime, is most clearly exemplified by the criminal activities listed as typical of organized crime in the Organized Crime (Threat Assessment) Reports produced at regular intervals by Europol and national law enforcement and governmental agencies, such the UK Serious and Organized Crime Agency and German Federal Police Office. In this second paradigm, depending on the authors and agencies, the identification of organized crime with the provision of illegal goods and services can be complete (e.g., Van Duyne 1997) or merely partial (e.g., Edwards and Levi 2008). Prohibited psychoactive drugs (e.g., heroin, cocaine, and cannabis) are the best-known examples of illegal goods, but, in addition to them, this category also includes goods produced and traded in violation of specific regulations: human beings, if trafficked for the purposes of sexual exploitation or forced labor; cigarettes, alcohol, and gasoline, if exchanged without paying excise and other taxes; and diamonds, wood, and other natural resources, if extracted illegally or sold outside the legitimate channels. The category of illegal services is also broad, as it ranges from gambling (in those jurisdictions in which it is still prohibited) to paid sex (in jurisdictions where prostitution itself or its exploitation is prohibited) and to the entry into a rich country, in the case of human smuggling.

Organized crime official reports and scholarly studies (e.g., Edwards and Levi 2008) also often include different forms of "illegal transfers" (e.g., robberies, extortions, embezzlement, and cases of fraud). Unlike the provision of illegal goods and services, these transfers do not create value but merely transfer it from one person to another. Not least for this very reason, however, the harm they generate is more evident than the harms accruing from the provision of illegal goods and services.

This entry is structured as follows: the first section gives a summary of the debate about organized crime in the USA and Europe, providing evidence of the above-mentioned alternative understandings of organized crime but also of the growing integration of the two views. The following two sections ("Mafia-Type Organizations" and "Illegal Markets/Activities and Their Actors") are, respectively, devoted to mafia-type organizations and illegal markets and their actors, as they best symbolize the two opposing views of organized crime.

Organized Crime: A Contested Concept

The meanings of organized crime have changed considerably and repeatedly over time. Reflecting the two different notions of organized crime discussed in the overview, some authors as well as national and international policy and law enforcement agencies have emphasized the "Who," that is, the offenders and their variable partnerships, whereas others have given more relevance to the "What," that is, the criminal activities conducted (see Smith, 1991).

The organization-based approach, which was dubbed the "alien conspiracy" theory by its critics due to its emphasis on foreign criminals, was first enunciated by the Kefauver Senate Investigating Committee in the 1950s. Despite the scarcity of empirical proof, the committee set out the terms of an Italian mafia-centered view of organized crime that remained the US official standpoint for almost three decades. This identified organized crime with a nationwide, centralized criminal organization that dominated the most profitable US illegal markets, allegedly derived from an analogous parallel Sicilian organization and largely consisted of migrants of Italian (and specifically Sicilian) origin (U.S. Senate 1951, p. 131). The merger of the two concepts of organized crime and mafia was fully accomplished in 1963 when Joe Valachi testified before the Senate Permanent Subcommittee on Investigations. By recalling his experiences as a low-ranking member of an Italian-American mafia association called (La) Cosa Nostra, Valachi gave a new name to the menacing criminal organization singled out by the Kefauver Committee and provided many details about its internal composition and illegal activities (U.S. Senate 1963). Thanks to extensive television coverage, Valachi's views were popularized among the American public.

This mafia-centered view of organized crime has long dominated the (USA and international) public perception of the problem: since the 1960s, hundreds of books have been written on the topic, and dozens of movies have been made. Some of these – above all, Mario Puzo's *The Godfather* (1969) and its film adaptation by Francis Ford Coppola – have been so successful that they have profoundly shaped the general understanding of organized crime and the mafia in the United States and elsewhere. For many people, the Italian-American mafia, which is de facto identified with organized crime, *is* and *behaves* as is recounted in these romanticized novels and films.

Such an interpretation received scientific systematization from Cressey (1969): in his view, La Cosa Nostra relied upon Sicilian traditional cultural codes but was also a hierarchical and "rationally designed" organization, very close to Max Weber's ideal type of legal-rational bureaucracy and therefore capable of operating in contemporary America.

The identification of the mafia with organized crime – and thus the idea of an alien conspiracy polluting the economic and social life of the country – has been rejected by the majority of American social scientists since the 1960s. These have alternatively accused the mafia-centered view of organized crime of being ideological, serving personal political interests, and lacking in accuracy and empirical evidence (e.g., Smith 1975). Some scholars, however, have overreacted, up to the early 1980s categorically denying the existence of the Italian-American mafia as a structured and longstanding criminal organization.

Scientific attention has instead been redirected from "Who" back to "What" and upon the most visible and noncontroversial aspect of organized crime: the supply of illegal products and services. In order to eradicate ethnic stereotypes of crime and direct attention to the marketplace, several authors have put forward the expression "illicit" or "illegal enterprise" as a substitute for the ethnically loaded term "organized crime." As Smith (1975, p. 335), one of the earliest proponents of the new approach, expressed it, "illicit enterprise is the extension of legitimate market activities into areas normally proscribed – i.e., beyond existing limits of law – for the pursuit of profit and in response to a latent illicit demand."

More often, however, organized crime itself has been equated with the provision of illegal goods and services. Hence, according to Block and Chambliss (1981, p. 13), "organized crime [should] be defined as (or perhaps better limited to) those illegal activities involving the management and coordination of racketeering and vice." Organized crime has thus become a synonym for illegal enterprise. That is, the involvement in illicit market activities has become nowadays the basic requirement of virtually all definitions of organized crime in the US scientific and official discourse, and this view is shared by both supporters of the mafia-centered understanding of organized crime and its critics.

A negative side effect of this partial consensus has been that the term "organized crime" is intermittently used to refer to both sets of criminal organizations and sets of activities. In the definition quoted above, Block and Chambliss clearly present organized crime as a set of activities. The identification of organized crime with a set of organizations is instead fostered by supporters of the US official standpoint and a few independent scholars, including critics of the official understanding. According to Reuter (1983, p. 175), for example, "organized crime consists of organizations that have durability, hierarchy and involvement in a multiplicity of criminal activities . . .. The Mafia provides the most enduring and significant form of organized crime." Unsurprisingly, the frequent confusion between offender and offense often leads to circular reasoning. In 1986, for example, the (second) President's Commission on Organized Crime (1986, p. 11) concluded that drug trafficking was "the single most serious organized crime problem in the United States and the largest source of income for organized crime."

With the exception of Italy and Spain, the "illegal enterprise" approach has since the 1970s acquired a dominant position in the scientific debate in Europe and has since the late 1980s been dominant in the policy debate as well, with a resulting focus on criminal activities for gain. As early as the mid-1970s, Kerner and Mack (1975) talked about a "crime industry," and in an earlier report written in German, Kerner subscribed even more explicitly to the view of organized crime as an enterprise (1973). The emphasis on illegal market activities has remained unchallenged ever since. Hence, for example, van Duyne (1997, p. 203) asks "What is organized crime without organizing some kind of criminal trade; without selling and buying of forbidden goods and services in an organizational context? The answer is simply nothing." More recently, several scholars (e.g., Edwards and Levi 2008) have suggested getting rid altogether of the expression organized crime and focusing instead on the organization of crime for gain.

The loose understanding of organized crime in terms of profit-making criminal activities has allowed organized crime to become a successful policy term even in European countries that had no mafia problems. Concern for organized crime first spreads in the late 1980s in Germany and the Netherlands and then, to various degrees, in other European states. Given its vagueness and elasticity, the term "organized crime" could be used to express the fear that the Italian mafia could invade the rest of Europe – a fear that reached its peak in the early 1990s after the murders of Judges Falcone and Borsellino by the Sicilian Cosa Nostra – but also refer to profit-making criminal activities and entrepreneurs close to home or to the real and imagined threats coming from the East after the 1989 fall of the Iron Curtain and the 1991 implosion of the Soviet Union.

With few partial exceptions (e.g., the Netherlands and Spain), the European official (or semiofficial) definitions of organized crime draw from the illegal enterprise paradigm. The definition adopted by German State Ministers of the Interior and Justice in 1986, which has been very influential and has been adopted also by other governments (e.g., Belgium), for example, states:

> Organized crime constitutes the planned commission of criminal offenses to acquire profit or power. Such criminal offenses have to be, each or in their entirety, of a major significance and be carried out by more than two participants who cooperate

> within a division of labor for a long or undetermined time-span using a) commercial or commercial-like structures, or b) violence or other means of intimidation, or c) influence on politics, media, public administration, justice, and legitimate economy. (BKA 2012)

Analogous, market-oriented definitions of organized crime have been adopted in the United Kingdom. The most recent is bafflingly simple: "Organised crime is defined as those involved, normally working with others, in continuing serious criminal activities for substantial profit, whether based in the UK or elsewhere" (SOCA 2012).

Following up on the studies of Reuter (1983) and others, policy-making and law enforcement agencies worldwide have increasingly had to realize that stable, large-scale criminal organizations far from monopolize profit-making criminal activities (e.g., Europol 2003). Rather than discarding the pair "criminal organizations" and "profit-making criminal activities," they have increasingly watered down the definition of criminal organizations so as to include also networks, gangs, cells, and any group composed by at least three people working together for some time (e.g., UNGA 2000; SOCA 2012). For some agencies (e.g., SOCA 2012) and scholars (e.g., Felson 2009, pp. 159–160), organized crime seems to mean little more than co-offending by more than two perpetrators.

The supposed territorial scope of organized crime has also changed. Originally, organized crime was equated with racketeering, another vague notion at the core of which there is, however, extortion, an activity necessarily territorially based and usually carried out only on a local basis. Since the 1990s, instead, many researchers (e.g., Williams and Florez 1994) and an even larger number of government agencies and international organizations, ranging from the USA to the United Nations (UNGA 2000; White House 2011), have emphasized the transnational nature of organized crime.

As this brief overview already shows, the policy and scientific legitimacy and relevance of the organized crime concept have also varied considerably over time. Except for the United States and Italy, most other nations did not regard organized crime as a serious domestic problem until the late 1980s. Whereas media accounts of organized crime and mafias have always fascinated the general public, most scholars shunned the topic until the early 1990s. Only since that period has organized crime become an accepted policy and scientific concept virtually throughout the world.

Since the 1990s, the majority of governments at least in the developed world and many international organizations, including the UN, have adopted bills, decrees, action plans, and international treaties specifically targeting organized crime, usually granting extensive new prosecutorial powers to law enforcement agencies – as the United States had already done in the 1960s and 1970s. The fight against organized crime has also been used to justify far-reaching criminal justice reforms in several countries, particularly in Europe (see Fijnaut 2013a).

Despite some persisting resistance, organized crime has also become a legitimate scientific research topic, attracting considerable attention from criminologists and other social scientists during the 1990s and the early twenty-first century, particularly in Europe.

Mafia-Type Organizations

Contrary to what the critics of the alien conspiracy stated, there is nowadays indisputable evidence that a few large-scale, long-standing criminal organizations exist even in Western countries, which Paoli (2002) calls mafia-type organizations. These range from the Sicilian Cosa Nostra and the Calabrian 'Ndrangheta in Italy to the Yakuza in Japan and from the Italian American La Cosa Nostra in the United States to the Triads in China, Hong Kong, and elsewhere and to the thieves-in-law in the USSR/Russia.

Long regarded by the media and public as the prototype of organized crime, these centuries-old organizations are usually loose consortia of smaller groups. The Calabrian 'Ndrangheta and Sicilian Cosa Nostra, for example, are each composed of about 150 groups (Paoli 2013). Twenty-four mafia "families" used to compose the

American La Cosa Nostra, even though many such groups are no longer active today (Albanese 2013). According to the Japanese police, there are 22 Yakuza organizations in Japan, a number that has remained stable since the 1990s (Hill 2013). Several different triads are known to be active in Asia, America, and Australia (Chin and Liu 2013).

Although the individual units enjoy considerable autonomy, it still makes sense to regard mafia-type criminal organizations as unitary. They constitute, in fact, segmentary societies – an organizational model well known to anthropologists studying "simple societies." In segmentary societies, such as mafia-type organizations, societal boundaries are drawn by the common cultural heritage and structural organization. In several cases, a process of centralization has been built upon these segmented structures. Among the families associated with the American and Sicilian Cosa Nostra and, more recently, among those belonging to the Calabrian 'Ndrangheta, this process of centralization has led to the institutionalization of superordinate bodies of coordination (Paoli 2003). In Japan, three syndicates – the Yamaguchi-gumi, Sumiyoshi-rengo, and Inagawa-kai – have succeeded in incorporating many smaller groups and now account for 80 % of the total number of initiated members (Hill 2013). The trend, however, is not univocal. There are no signs of a centralization process going on in the heterogeneous universe of Chinese organized crime. Furthermore, even when superordinate bodies of coordination exist, their competencies are rather limited. Usually their main purpose is to minimize the visibility of the criminal organizations through the regulation of the internal use of violence. In the case of Sicilian and American Cosa Nostra cases, moreover, the centralization of power has reverted and the "commissions" are no longer active (Paoli 2013).

Mafia-type organizations are based on lifelong status and fraternization contracts, which establish ritual kinship ties among the members. It is not by chance that the basic units of the Sicilian and American Cosa Nostra are called families. Although Italian and Italian-American mafia groups clearly distinguish themselves from the blood families of their associates so much so that no women are allowed, the term evokes and, at the same time, prescribes the cohesion and solidarity of blood ties (Paoli 2003). In Japan, the relationship between a yakuza chief (*oyabun*) and lower-ranking members (*kobun*), which is the pillar of the whole association, is portrayed as the relationship between a father and his sons (Hill 2013). Relying on fictive kinship ties, mafia-type organizations enjoy a flexibility that has no parallel among contemporary businesses whose employment contracts usually promise economic benefits in exchange of specific, limited tasks. The contracts imposed by mafia-type organizations on their members are so comprehensive that the latter are expected not only to deny family and friendship bonds but even to sacrifice their own life if the group requests it. In exchange, individual members benefit from the collective action and the reputation of the organization in the pursuit of both licit and illicit businesses. The membership in a mafia-type organization group is, hence, typified by a crisscrossing of instrumentality and solidarity, of personal selfishness and unconditional involvement.

The ritual kinship ties established at the moment of the initiation ceremonies have historically been used by both the organizations as such and their members to achieve a plurality of goals and to accomplish a variety of functions. Only by sacrificing empirical evidence, it is possible to single out an encompassing function or goal that can characterize mafia-type organizations throughout their lives. In particular, although their members are heavily involved in illegal businesses today, neither the development nor the internal structure of mafia-type organizations is the product of illegal markets dynamics. Indeed, many of these organizations arose before the consolidation of modern illegal markets: the predecessors of Triads and Yakuza groups in the course of the eighteenth century and Southern Italian mafia groups in the late nineteenth century (Chin and Liu 2013; Hill 2013; Paoli 2013).

Among the many functions historically carried out by mafia-type organizations, political functions, even more than economic activities, have always had a key relevance, and it is to Diego Gambetta's credit that he brought attention to this dimension that had long been neglected in the scientific discourse on the mafia and organized crime. According to Gambetta (1993, p. 1), in fact, the Sicilian mafia is "a specific economic enterprise, an industry that produces, promotes and sells private protection" (The provision of protection is one of the most important functions historically played by mafia groups, a quintessential one, we could say, since it derives from their exercise of violence in the villages and neighborhoods where mafia groups are based. Gambetta's analysis is rather to be criticized for his one-sided emphasis on protection and his denial of the polyvalence of mafia groups. It is furthermore regrettable that out of polemical reasons, Gambetta overshadows the analogies between the mafia, as he defines it, and the state. Such a similarity was instead outlined by Charles Tilly (1985, p. 169): "if protection rackets represent organized crime at its smoothest, then war making and state making – quintessential protection rackets with the advantage of legitimacy – qualify as our largest examples of organized crime."

Although not employed on every occasion, ultimately violence constitutes the backbone of mafia-type organizations' power. These use violence foremost to secure the obedience of their own members and to punish those that have betrayed or not respected their authority. They also employ it as a means to threaten, render inoffensive, or even physically eliminate whoever endangers the power positions and the business activities of the organization or its individual units. Through the threat or the effective use of violence, mafia-type organizations have also been trying – enjoying for a long time a fairly high degree of success – to impose their rules on society at large in their territory of influence. As such, criminal organizations of mafia type can be seen as proto-states that have the ambition to impose their own legal order within a given territorial area through the threat and the use of physical force. Each family associated with either Cosa Nostra or the 'Ndrangheta, for example, claims sovereignty over a well-defined territory which usually corresponds to a village or to the district of a city.

Mafia-type organizations' political ambitions mirror the weakness of state authorities in the contexts in which they developed. The oldest among them consolidated in contexts in which the local state authorities were not able to guarantee even a minimum of security and the residents had to protect their property, women, and own lives by themselves. The American La Cosa Nostra rose in the early twentieth century when the US government still had limited authority in the Italian ethnic community and some of its representatives preferred to come to terms with, rather than prosecute, La Cosa Nostra bosses and other criminal entrepreneurs fostered by Prohibition. Founded in Stalinist prisons in the 1930s, *vory v zakone* and other Russian crime groups gained unprecedented political power and control over economic resources after the implosion of the Soviet Union in 1991 (Volkov 2013).

While criminal groups posing as proto-states rose in several parts of pre-modern Europe (e.g., Fijnaut 2013b), most of them were swept away by the consolidation of modern state structures. The current mafia-type organizations have survived in contexts in which government structures have remained weak or their representatives have been willing to enter into pacts with them. At least up to the mid-twentieth century, for example, Italian government officials came to terms with the representatives of mafia power and de facto delegated to the latter the maintenance of public order over wide areas of western Sicily and southern Calabria, where the authority of the central government was scarce and even the officials' personal safety was in danger. In Japan as well, Yakuza bosses have held close ties with members of the political establishment and specifically the Liberal Democratic Party, Japan's dominant political party since WWII (Hill 2013).

Exploiting the weakness of the state and, at the same time, their own privileged contacts with corrupt politicians and civil servants, mafia-type criminal organizations have traditionally fostered their popular legitimacy by providing income and other nonmaterial advantages to their associates and a variety of services for the general population, such as civil-service jobs, a pension benefit, or impunity for their supporters or votes for befriended politicians. Additionally, in Italy, Japan, Russia, and Chinese communities all over the world, they have occasionally controlled the workforce and carried out "dirty work" for colluding companies and landowners, for example, by protecting them from ordinary criminals or by breaking strikes or other forms of protests. Furthermore, mafia-type organizations have supplied goods and services for which there is, despite their official prohibition, a public demand. It is through the latter function that mafia-type organizations become involved in the criminal activities being considered exemplary of the alternative understanding of organized crime – thus, creating a partial overlapping between the two views. Contrary to what is often maintained, however, mafia-type organizations are far from controlling illegal markets worldwide, and indeed, according to a plurality of independent researchers, their "market share" has in recent years further declined (Chin and Liu 2013; Hill 2013; Paoli 2013).

Illegal Markets/Activities and Their Actors

Despite the exaggerations of some moral and political entrepreneurs, the growing worldwide concern about organized crime partially reflects real changes that directly impinge on illegal (market) activities and their actors – which is the second paradigmatic understanding of organized crime.

Among these changes, the most significant has been the rise of the illegal drug industry. Since the 1970s, this has become the largest and most profitable illegal market activity in virtually all Western countries, attracting the greatest number of traditional underworld figures and fostering a reorientation of professional crime toward the drug business. According to a conservative estimate, the world drug markets generated US $322 billion in retail sales in 2003, corresponding to 0.9 % of global GDP (UNODC 2005, pp. 123–143). These are substantial figures, probably much larger than the revenues of any other illegal entrepreneurial activity, although there are no reliable estimates for any of them.

The demand for illegal drugs in rich countries has been a key promoting factor of the expansion of the world drug industry, which now connects opium poppy and coca cultivation areas in Afghanistan, Burma, or Colombia to retail users several thousands of kilometers away. However, Western countries are far from being merely consumers or importers of illegal drugs. Some of them, and most notably the Netherlands, Belgium, and the United States, have also become major producers of cannabis and synthetic drugs, while others play a key role as transit countries (e.g., Spain).

Despite the reorientation of many professional criminals toward drug trafficking and dealing, several – traditional and nontraditional – profit-making criminal activities have continued to proliferate. As a matter of fact, some of these activities – ranging from car theft to robberies and the exploitation of prostitution – experienced an unexpected revival in the years immediately following the fall of the Iron Curtain, when Eastern European criminals primarily resorted to violence and ruthlessness to earn a "fast buck" in Western Europe.

A third group of entrepreneurial crime activities has also flourished since the 1980s, although they hardly are the prerogative of traditional underworld members. These activities range from fraud and other financial crimes to bid-rigging in public works tenders and the illegal wholesale trade in toxic waste, weapons, diamonds, and gold. They undoubtedly form a part of organized crime, if one accepts the loose official definitions of organized crime. Albeit difficult to estimate precisely, the harm these activities cause to society might be considerably higher than the harms of the illegal activities, such as drug and human smuggling, traditionally regarded as typical of organized crime.

Whereas the more white-collar forms of organized crime usually attract public attention only in the immediate aftermath of a big scandal, a second wave of expansion of Western illegal markets began in the mid-1980s, raising much concern in government institutions and the general public. This expansion was largely triggered by the enactment of increasingly restrictive immigration policies in most Western countries during the 1980s and 1990s, which created a large demand for human smuggling services. The number of potential customers as well as victims of veritable human trafficking also suddenly multiplied, as the liberation of Eastern Europe in 1989 and the collapse of the Soviet Union in 1991 abolished restrictions on the mobility of almost 400 million Eastern European and former Soviet citizens. Crises in other parts of the world, ranging from several African countries to Iraq, Afghanistan, and East Timor, also engorged the flow of prospective migrants, at the same time as growth and improvement of transportation facilitated their movements, by drastically reducing logistical constraints.

To meet this demand, human smuggling "companies" popped up at all the crucial borders of "Fortresses Europe and USA." Although many smugglers merely sell services desperately wanted by their customers, their prices are often extortionate and the smuggling conditions dangerous and even inhuman, as proven by the accidents all over Europe and on the US-Mexican border that cost the lives of undocumented migrants. (These incidents are caused, however, not only by the smugglers' ruthlessness but also by the rigor of Western governments' border interventions.) Moreover, this flourishing black market has opened up space for all kinds of exploitation that sometimes end up as real trafficking in human beings.

It is almost a platitude to note that public demand constitutes the main determinant for the provision of illegal goods and services. The evidence for such a statement is most compelling in the case of illegal drugs, which are willingly consumed by about 230 million people worldwide (approximately 5 % of the world population aged 15–64; UNODC 2012).

Political conditions – again primarily the strength and legitimacy of government authorities – also impact illegal markets. For example, they play a key role in determining the location of production of agriculturally based drugs (especially, heroin and cocaine). All else being equal, the cultivation of the plants from which these drugs are extracted tends to concentrate in countries with no effective enforcement of drug prohibitions and with local government or quasi-government tolerance or support for drug production and trade. Afghanistan and Burma, the world's two leading opium producers, most clearly illustrate this principle.

The draconian measures occasionally imposed by authoritarian governments and quasi-state authorities also vividly demonstrate the impact that government decisions can have, at least in the short term, on illegal markets. With a radical and concerted opium suppression campaign, the Chinese Communist led regime eliminated in the early 1950s opium consumption and production in China, which was then by far the largest market in the world. Such a campaign, however, entailed degrees of coercion that no democratic government can rightfully employ (Paoli et al. 2009). Totalitarian regimes have also been able to prevent the integration of their countries into international illegal markets. Most Western illegal drugs, for example, were unavailable in Russia until the implosion of the Soviet Union in 1991. The bulkier the product is, the higher the chances of control. As already mentioned, the Iron Curtain effectively deprived 400 million people of their right of free movement and "protected" them from falling prey to human trafficking and smuggling.

The chances of controlling illegal markets are, however, much more limited for democratic governments committed to the rule of law and a market economy. As long as these are not willing to resort to authoritarian border and law enforcement methods or to disrupt legitimate trade by, say, inspecting every container crossing their national borders, it will be almost impossible for them to stop illegal market flows. Indeed, these can be regarded as the price we need to pay for democracy and a market economy. Much like the informal economy, of which they are part, illegal markets constitute a "normal," unavoidable component of all modern societies. Contemporary governments have even less leverage than their predecessors over illegal – and legal – market flows as a result of economic globalization. The diminution of state-enforced restrictions on exchanges across borders – the core trait of globalization – has accelerated the interconnections between previously separate domestic, legal and illegal, markets and increased the mobility of goods, capital, and human beings. For illegal entrepreneurs, it has become easier than ever to move drugs and other illegal commodities from producing to consuming countries, to repatriate profits, to establish business partnerships with foreign counterparts, and even to operate in foreign countries themselves.

If the prospects of controlling global illegal market flows are quite bleak, the in-depth analysis of the world heroin market by Paoli et al. (2009) suggest that government authorities, even in democratic societies, are not powerless vis-à-vis organized crime activities. One of their main conclusions is that the strictness of governments' enforcement of prohibitions – in other words, the degree of effective illegality to which opiates (i.e., opium and its derivatives, including heroin) are subject – is the most important single factor to shape how the opiate/heroin market is organized in a particular country and the behavior of its producers and traffickers. It impacts, in particular, the size, organization, and operating methods of enterprises that produce or traffic illicit opiates.

These findings tie in well with a growing body of literature on organized crime in Western countries that show that the great majority of illegal exchanges in Western countries are carried out by numerous, relatively small, and often ephemeral enterprises. In other words, "disorganized crime" (Reuter 1983), rather than organized crime understood in terms of large-scale stable criminal organizations, predominates on the illegal markets of Western countries, reflecting the constraints deriving from the enforcement of prohibitions. These constraints have to do with the fact that illegal market entrepreneurs are obliged to operate both without and against the state.

First, since the goods and services they provide are prohibited, illegal market actors cannot resort to state institutions to enforce contracts and have violations of contracts prosecuted. Nor does the illegal arena host an alternative sovereign power to which a party may appeal for redress of injury. As a result, property rights are poorly protected, employment contracts cannot be formalized, and the development of large, formally organized, long-lasting companies is strongly discouraged (Reuter 1983).

Second, all illegal market actors are forced to operate under the constant threat of arrest and confiscation of their assets by law enforcement institutions. Participants in criminal trades will thus try to organize their activities in such a way as to assure that the risk of police detection is minimized. Incorporating illegal transactions into kinship and friendship networks and reducing the number of customers and employees are two of the most frequent strategies illegal entrepreneurs employ to reduce their vulnerability to law enforcement moves (Reuter 1983; see also Bouchard and Morselli 2013).

Due to the threat of police intervention, either in terms of seizing assets or imprisoning participants, the planning time horizons of illegal entrepreneurs are likely to be much shorter than those in legal markets. Since an illegal enterprise can hardly be sold as the entrepreneur ages, he is likely to divert an increasing share of his profits to legal assets which can be passed on to his heirs.

Finally, since they are operating against the state, illegal enterprises are prevented from marketing their products. They cannot create their own brand image and try to bind customers to it. Strong economies of scale, however, are associated with advertising, and the advantages linked to the nationwide marketing of one's own products have long been recognized as a very important factor in the rise of modern large-scale corporations. Illegal enterprises, however, are by definition excluded from the possibility of exploiting these advantages because, by doing so, they would obviously attract law enforcement attention and damage their own businesses. For the same motives, illegal enterprises in Western countries resort sparingly to violence: as Pearson and Hobbs (2001, p. 42) put it, "violence and killings attract police attention and leave traces, as well as attracting retaliation. Violence is therefore strictly 'bad for business."

For all the above reasons, it is rather unlikely that large, hierarchically organized enterprises will emerge to mediate economic transactions in the illegal marketplace of countries with effective governments. The factors promoting the development of bureaucracies in the legal portion of the economy – namely, taking advantage of economies of scale of operations and specialization of roles – are outbalanced in the illegal arena by the very consequences of product illegality. Even mafia-type organizations are subjected to the constraints deriving from the illegal status of products, and when they deal in drugs or other illegal products, they do not operate as monolithic productive and commercial units. On the contrary, their members frequently set up short-term partnerships with other mafia members, or even nonmembers, to carry out illegal transactions.

Nor are the average illegal market actors of Western countries interested in, or capable of, exercising a quasi-political power similar to that of mafia-type organizations. As Europol (2003, p. 10) recognizes, "politically, few OC groups pose a direct threat to Member States." Most organized crime groups active in the richest Western countries are simply too small and ephemeral to be able to exercise such political power.

The situation can look radically different in countries with lax or no enforcement of official prohibitions. Paoli et al.'s (2009) analysis of the world heroin market, for example, concludes that opiate enterprises can become large and stable and assume bureaucratic characteristics if they are closely linked via corrupt ties to, or coincide with, a powerful state or quasi-state authority. Under conditions of lax or non-enforcement, opiate enterprises may form oligopolies or even monopolies, if powerful and engaged state or quasi-state authorities back them up. The extent of violence and bribery also depends on the strength and involvement of state or quasi-state authorities. Violence may be high, if the supporting state or quasi-state authority is weak; it may be low, if the state or quasi-state authority is strongly institutionalized and directly involved in the drug trade.

This model also predicts that for countries that have become accustomed to lax or non-enforcement, a shift toward strict enforcement may imply a worsening of drug-related corruption, violence, and instability in the interim. What happened in Afghanistan during the first decade of the twenty-first century clearly demonstrates the potential for worsening conditions: powerful drug producing and trafficking organizations have used all available means to oppose state efforts to enforce prohibition more rigorously and fight for the "right" to continue their established businesses.

Illegality and enforcement seem to have similar effects on the size and operating methods of cocaine and cannabis trafficking enterprises as they have on opiate-related firms. For example, large and stable cocaine trafficking organizations having no qualms about openly challenging state sovereignty have consolidated in countries lacking consistent enforcement of prohibitions, such as Colombia and Mexico (e.g., Medel and Thoumi 2013; Thoumi 2013). Presidents Fox and Calderon's determined fight against Mexican drug cartels has plunged the country into an unprecedented spiral of violence. Mutatis mutandis, the model of varying effective illegality seems to be applicable to the enterprises producing and dealing with synthetic drugs and other illegal goods and services.

Conclusion

The analysis shows that the manifestations of organized crime in any single national context depend – to a considerable, albeit not exclusive, extent – on the authority of the local government agencies. A full analysis, however, would also have to weigh the benefits and costs of governmental control interventions. The US war on drugs of the past two decades may have helped keep the US drug market relatively "disorganized" but has failed to reach its main goal of reducing drugs availability and has produced huge social and financial costs: it is enough to say that a considerable proportion of the over two million people behind bars and about five million more on probation or parole have been convicted for drug offenses.

Criminal repression is undoubtedly a necessary reaction to the most serious forms of organized crime, such as a mafia-type organization challenging government authority. Nonetheless, governments should always aim at reducing the total harms associated with organized crime *and* policies and not lose sight of the fact that they also carry responsibility for these harms.

Related Entries

- ▶ Drug Trafficking
- ▶ History of Organized Crime in Mexico
- ▶ History of Organized Crime in the United States
- ▶ Italian Mafia
- ▶ Money Laundering
- ▶ Motorcycle Clubs or Gangs?
- ▶ Network Analysis in Criminology
- ▶ Organized Fraud
- ▶ Organized Crime and the Environment
- ▶ Russian Mafia
- ▶ Theories of Co-Offending
- ▶ Trafficking in Human Beings

Recommended Reading and References

Albanese J (2013) Italian American organized crime. In: Paoli L (ed) Handbook of organized crime. Oxford University Press, New York (forthcoming)

BKA, Bundeskriminalamt (2012) Lagebild organisierte kriminalität bundesrepublik Deutschland. BKA, Wiesbaden

Block A, Chambliss W (1981) Organizing crime. Elsevier, New York

Bouchard M, Morselli C (2013) Opportunistic structures of organized crime. In: Paoli L (ed) Handbook of organized crime. Oxford University Press, New York (forthcoming)

Chin K-L, Liu M (2013) Chinese organized crime. In: Paoli L (ed) Handbook of organized crime. Oxford University Press, New York (forthcoming)

Cressey D (1969) Theft of the nation. Harper and Row, New York

Edwards A, Levi M (2008) Researching the organization of serious crimes. Criminol Crim Justice 8:363–388

Europol (2003) 2003 European Union organised crime report. Office for the Official Publications of the European Communities, Luxembourg

Felson M (2009) The natural history of extended co-offending. Trends Organized Crime 12:159–165

Fijnaut C (2013a) Organized crime control policies and interventions in Europe. In: Paoli L (ed) Handbook of organized crime. Oxford University Press, New York (forthcoming)

Fijnaut C (2013b) Searching for organized crime in history. In: Paoli L (ed) Handbook of organized crime. Oxford University Press, New York (forthcoming)

Gambetta D (1993) The Sicilian mafia. The business of private protection. Harvard University Press, Cambridge

Hill P (2013) Yakuza. In: Paoli L (ed) Handbook of organized crime. Oxford University Press, New York (forthcoming)

Kerner H-J (1973) Professionelles und organisiertes Verbrechen. Versuch einer bestandsaufnahme und bericht über neuere Entwicklungstendenzen in der Bundesrepublik Deutschland und in den Niederlanden. BKA, Wiesbaden

Kerner H-J, Mack J (1975) The crime industry. Lexington Books, Lexington

Medel M, Thoumi F (2013) The drug "Cartels" in Mexico. In: Paoli L (ed) Handbook of organized crime. Oxford University Press, New York (forthcoming)

Paoli L (2002) The paradoxes of organised crime. Crime Law Soc Chang 37(1):51–97

Paoli L (2003) Mafia brotherhoods. Organized crime, Italian style. Oxford University Press, New York

Paoli L (2013) Italian organized crime. In: Paoli L (ed) Handbook of organized crime. Oxford University Press, New York (forthcoming)

Paoli L, Greenfield VA, Reuter P (2009) The world heroin market: can supply be cut? Oxford University Press, New York

Pearson G, Hobbs D (2001) Middle market drug distribution. Home Office Research, Development and Statistics Directorate, London

President's Commission on Organized Crime (1986) The impact. Organized crime today. Report to the president and the attorney general. U.S. Government Printing Office, Washington, DC

Puzo, M. 1969. The Godfather. New York: Putnam

Reuter P (1983) Disorganized crime: the economics of the visible hand. MIT Press, Cambridge

Smith D Jr (1991) Wickersham to Sutherland to Katzenbach: evolving an 'Official' definition for organized crime. Crime Law Soc Chang 16(2):138–142

Smith D Jr (1975) The mafia mystique. Basic Books, New York

SOCA, UK Serious and Organised Crime Agency (2012) Organised Crime Groups

Thoumi F (2013) Organized crime in Colombia. In: Paoli L (ed) Handbook of organized crime. Oxford University Press, New York (forthcoming)

Tilly C (1985) War making and state making as organized crime. In: Evans PB, Rueschemeyer D, Skocpol T (eds) Bringing the state back in. Cambridge University Press, Cambridge

U.S. Senate (1951) Third interim report of the special committee to investigate organized crime in interstate commerce (Kefauver Committee). 81st Cong., 2d sess. U.S. Government Printing Office, Washington, DC

U.S. Senate, Committee on Government Operations (1963) Hearings of Joseph Valachi before the permanent subcommittee on investigations of the committee on government operations. U.S. Government Printing Office, Washington, DC

UNGA, United Nations General Assembly (2000) Report of the Ad Hoc committee on the elaboration of a convention against transnational organized crime on the work of its first to eleventh session. A/55/383. 2 Nov

UNODC, United Nations Office on Drugs and Crime (2012) World drug report. United Nations, New York

Van Duyne PC (1997) Organized crime, corruption, and power. Crime Law Soc Chang 26:201–238

Varese F (2010) Introduction. In: Varese F (ed) Organized crime. Critical concepts in criminology. Routledge, London

Volkov V (2013) Russian organized crime in a transitional economy. In: Paoli L (ed) Handbook of organized crime. Oxford University Press, New York

White House (2011) Strategy to address transnational organized crime: addressing converging threats to national security. White House, Washington, DC

Williams P, Florez C (1994) Transnational criminal organizations and drug trafficking. Bull Narc 2:9–24

Organized Eco Crime

▶ Organized Crime and the Environment

Organized Environmental Crime

▶ Organized Crime and the Environment

Organized Fraud

Michael Levi
Centre for Crime, Law and Justice at Cardiff
School of Social Sciences, Cardiff University,
Cardiff, Wales, UK

Overview

The relationship between fraud, organized crime, and white-collar crime is a complex one. When Sutherland (1945) wrote his classic presidential address for the American Sociological Association toward the end of the Second World War, "White-collar crime *is* organized crime," he was pointing to the "organizedness" of the crimes committed by the otherwise legitimate corporations he had researched as well as, perhaps, appropriating some of the demonization that attaches to the "organized crime" label. Later theorists such as Shapiro (1980), arguing that the focus on offender status was too confusing, redefined white-collar crimes as crimes of deception and as characterized by interest clashes between principals and agents. Pushed to its limits, this would mean that all white-collar crimes are organized fraud (though of course, not all organized crimes are fraud). According to the UN Transnational Organized Crime Convention, one needs three or more people acting together for some undefined period of time to constitute organized crime, so some large-scale elite as well as junior opportunist "well-organized" frauds are *not* "organized crime" in this sense: if you are in charge of a large organization – business or political party – and people do what you tell them, you may not need active conspirators. In practice, as the Dutch Organized Crime Monitor has shown, "organized fraud" can be a shorthand phrase for frauds committed by career criminals rather than by respectables (e.g., van Koppen et al. 2010): the very mindset that Sutherland was seeking to undermine in his analogy, which remains a powerful folk image because part of the apparent threat of "organized crime" is that everything an "organized criminal" does threatens "legitimate society" (whereas, by omission, whatever those who are *not* organized criminals do is *less* threatening to society). Hence, the regular references to organized criminals "infiltrating" or "corrupting" commerce. Indeed, the criminal careers of white-collar criminals are more uneven than one usually finds elsewhere (Piquero and Benson 2004). Measuring "organized fraud" involves two difficult components: measuring how much of different frauds there are and measuring what proportion of them are "organized" (and by what criteria this "organizedness" is to be judged). The settings for frauds influence and reflect networks, in the context of fraud opportunities and "capable guardians," in the classical situational opportunity model.

Fundamentals of Organizing Frauds

One way of thinking about how frauds are organized is to develop a process model. When analyzing the dynamics of particular crimes and/or criminal careers, these procedural elements can be broken down further into much more concrete steps, sometimes referred to as "crime scripts" (Morselli 2008).

The Organization of Fraud: A Process Model

1. See a situation as an opportunity for fraud.
2. Obtain whatever finance/equipment/data needed for the crime.
3. Find people willing and able to offend (if conspirators are needed, inside and/or outside the victim organization), preferably those who are controllable and reliable.
4. Carry out offenses in domestic and/or overseas locations with or without physical presence in jurisdiction(s).
5. Minimize immediate enforcement/operational risks. Especially if planning to repeat frauds, neutralize law enforcement by technical skill, by corruption, and/or by legal arbitrage between jurisdictions.
6. Convert, where necessary (e.g., where goods rather than money are obtained on credit), the proceeds of crime into money or other usable assets.
7. Spend as much of the proceeds as desired.
8. Find people and places willing to store those proceeds the offenders wish to retain (and perhaps conceal their origin).
9. Decide which *available* jurisdiction(s) offers the best balance between comfort and the risk of asset forfeiture/criminal justice sanctions.

There is also a personality dimension (van Duyne 2000): some people seek out opportunities to defraud and look for facilitators like accountants or lawyers, whereas others just take what comes opportunistically. Felson (2003) has sought to account for some variations between area crime rates as the product of "crime chemistry": features of a setting that come together to foster crime, including the presence of likely offenders and suitable targets as well as the absence of capable guardianship against the offenses that would otherwise take place. Features relevant to crime levels also include three other "often-important elements": "props that help produce or prevent a crime," "camouflage that may help an offender avoid unwanted notice," and "an audience that the offender wants either to impress or to intimidate" (Felson 2003). Although Felson did not have white-collar offenses in mind, one might consider the extent to which these constructs help comprehend aspects of fraud.

Those who are connected to "organized crime" may already have in place all the steps as part of their ongoing "criminal enterprise." Criminal finance, some or all criminal personnel, or the "tools of crime" (from companies registered in "secrecy jurisdictions" to anonymous prepaid debit cards which can be used globally) may come from or go to another country, constituting "transnational" crimes, or else remain within one country. The exploitation of international regulatory and criminal justice asymmetries – for example, different levels of enforcement in the states or countries in which the fraudsters operate and victims are located – represents a positive advantage for fraud compared with most other property crimes.

Applying the sort of script found in the process model above, fraudsters seek potential victims for their schemes and develop techniques for getting them to part with their money voluntarily. In some cases, they or others may revictimize the same people by offering (for a fee) to help them get back the money they have lost. Another characteristic difference is that the offender and victim do not ever have to be in the same place at the same time. Some such offenses involve face to face contact throughout or at some stage; others are done wholly remotely, like lottery and eBay scams. An example of the latter is "419 frauds," so called after section 419 of the Nigerian criminal code (explored in detail by Schoenmakers et al. 2009). Most e-mail users have received scam emails – which usually arrive from a Yahoo! Mail, Hotmail, or Gmail address – offering them vast wealth if only they will help their previously unacquainted banker/relative of a famous deceased corrupt dictator put their "dormant account" or "unknown to the authorities but at risk" money into their own account for 25–40 % of the "take" (often £/€/$25 billion or million). The victim is later persuaded to pay "advance fees" to remove blockages in the funds transfers and may even be lured to Nigeria, South Africa, or some other country to pay out more. Nigerians turn to their advantage the stereotype of their corruption in order to make the proposition more plausible to their intended targets. Alternatively, they make religious-type appeals – starting with "Dear Brother in Christ" – to assist orphans, or some other spoof rationale for large funds being available, such as money being left to the target in an unknown relative's will.

Choice of offender or victim location is determined by convenience or by deliberate targeting by globally mobile offenders, though consumer frauds can be plotted from afar: see, e.g., Holtfreter et al. (2005, 2008). Businesspeople find it easy to transfer funds out of the company and/or to commit bankruptcy fraud, making relationships with organized crime groups usually unnecessary, though the latter may offer economies of scale and protection from the police or impose their will by threats of violence. The corporations can be substantive and real or they can be mere fronts or shells for the perpetration of fraud. But people can also commit frauds against companies and the government as outsiders or from more junior positions. Fraud permits a variety of offender organizational permutations, from Mafia-type associations to sole or small group offending.

As for the persistence of crime techniques over time, this varies depending on the countermeasures taken by potential preventers, who are mostly outside criminal justice. Fraudsters can now use payment card numbers skimmed from unsuspecting cardholders or hacked/copied from corporate databases to order hundreds of computers on the net from different suppliers, have them delivered to "drop addresses" in any credible country, and then forwarded to addresses elsewhere for resale: all of this before the cardholder or card issuer becomes aware that anything has happened to their card or account. This and other cyberfraud techniques reflect a comparative criminal advantage arising from the combination of high technological skills and high motivation because of poor opportunities in their home countries. There are also large-scale credit card and loan "bust outs" using stolen identities to obtain goods and money. Although these are often referred to as "identity theft," they are more accurately referred to as identity duplication, since the original person is not deprived of their existing identity (see Vieraitis and Copes, this volume, for a more extensive account of identity theft). The illustrations above show that behavior of victims-to-be and "capable guardians" has to be considered as part of the organization of crime. Nowadays, electronic credit histories are omnipresent in the UK and USA, but such data are less available elsewhere, especially in emerging economies where long business and individual commercial data do not exist. Even addresses in Africa are usually postbox numbers rather than physically identifiable dwellings. Many of the features upon which "identity validation" rests are not uniformly available, and this creates difficulties for legitimate people wanting to open bank or other credit accounts in the face of anti-money-laundering identification requirements developed for advanced economies.

Felson and Boba (2010) view white-collar crimes as "crimes of specialized access." However, though useful, this is an inadequate compass for all frauds. Some frauds require professional licenses, some involving lengthy training, which "organized criminals" (in the conventional usage of that term) are ill-equipped to undertake in the short run. Thus, Italian-American Mafiosi who aspired to become "penny stock" fraudsters had to find people with licenses before they could diversify – fortunately for them, some Americans of Russian origin who had licenses were willing to assist them. But other frauds, like some credit card and insurance frauds, and even corporate bankruptcy frauds, do not require specialized access but are available to new entrants who have enough initiative and some basic skills, which have become greater as the introduction (outside the USA) of cards with chip and PINs has made it difficult to use stolen cards (see Forsyth and Castro (2007) for a professional credit card fraudster's perspective). "Stealth frauds," like the illegal downloading of personal information from data sets and using it for "identity fraud," the copying of magnetic stripe data from credit cards, and illegal electronic funds transfers, may but do not always need specialized access. Shover et al. (2003) state that the sales agent in telemarketing scams generally works from a script that lays out successful sales approaches and responses. Promising contacts are turned over to a "closer," a more experienced, and a better-paid sales agent. The hierarchy of the firms and the routine of turning prospects over to more experienced closers explain why victims typically report contact with multiple salespersons.

What factors influence the choice of venue for boiler room telemarketing frauds and their modus operandi? From UK cases examined by this author (see also http://www.cityoflondon.police.uk/CityPolice/Departments/ECD/Fraud/boilerroom.htm), boiler rooms are commonly based abroad (e.g., in Spain, where police interest is low); never seek authorization by the regulator, the Financial Services Authority (FSA), which authorization is a legal requirement to sell securities in the UK; and use high-pressure sales and telephone techniques (author interviews with UK officials and police). One more sophisticated technique is to approach a small UK company not listed on the stock exchange and propose to raise capital by selling £100,000 worth of shares in that company on their behalf. Of this £100,000, the boiler room would agree to take 60 % as its fee, leaving the small company with £40,000 capital. In reality, the boiler room will "cold call" (i.e., telephone without prior contact) UK investors to sell the shares at up to 100 % over the agreed price, take their fee, and vanish. The small companies involved may become liable to refund investors the full price paid for their shares. There are several variations on the method of committing the crime:

- Complete con, where there are no shares in existence
- Different instruments used – stock, currency options (and even bull sperm)
- Restricted (e.g., US "Regulation S"), worthless, or overpriced shares
- Purported involvement in raising capital for companies
- Market manipulation where there are shares in existence and a (limited) market
- Deceptive share promotion via bulletin boards

(For a broader examination, see Shover et al. 2003; Stevenson 2008.) If the fraudsters have sufficient nerve, they can seek to become regulated in one EU country and obtain a "passport" to operate in another under EU single market regulations, using that as a base for fraud and making it difficult for local regulators to intervene to close them down. In all of these cases, what the boiler room is really selling is deceptive and worthless expectations. Similar patterns exist in the USA (Tillman and Indergaard 2005), though within the USA interstate scams constitute federal offenses: hence, the preference of some fraudsters for operating from Canada.

One common type of tax fraud in Europe is known as "carousel fraud." This may involve any type of standard-rated goods or services. Goods or services are acquired zero-rated from the EU, with the acquirer then going missing without accounting for the value-added tax (VAT) due on the onward supply. However, in the case of a British scam, the goods or services do not become available in the UK for consumption, but are sold through a series of companies in the UK and then exported or dispatched, prompting a repayment from the government (HMRC) to the exporter/dispatcher. This process can be repeated over and over again using the same goods or commodities. When this happens, it is called "carousel fraud." In some cases, no real products are traded at all but the transactions exist only on paper (Fig. 1).

Mortgage Frauds

The social construction of mortgage fraud is intriguing and has been relatively little researched (though see Nguyen and Pontell (2010), and

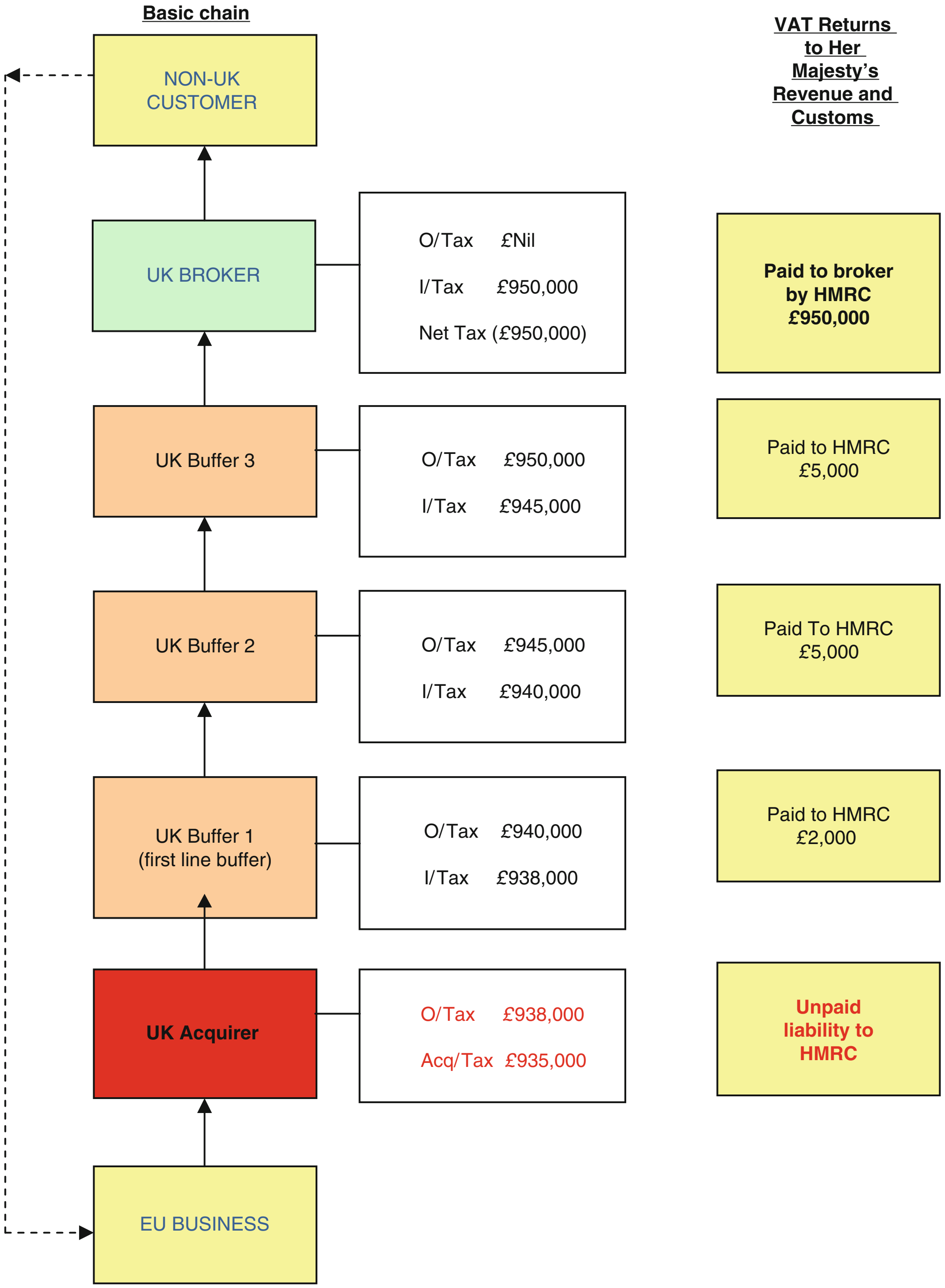
Basic chain
VAT Returns to Her Majesty's Revenue and Customs
NON-UK CUSTOMER
UK BROKER
O/Tax £Nil
I/Tax £950,000
Net Tax (£950,000)
Paid to broker by HMRC £950,000
UK Buffer 3
O/Tax £950,000
I/Tax £945,000
Paid to HMRC £5,000
UK Buffer 2
O/Tax £945,000
I/Tax £940,000
Paid To HMRC £5,000
UK Buffer 1 (first line buffer)
O/Tax £940,000
I/Tax £938,000
Paid to HMRC £2,000
UK Acquirer
O/Tax £938,000
Acq/Tax £935,000
Unpaid liability to HMRC
EU BUSINESS

*This is the acquisition tax which is also treated as input tax (there is no VAT on acquisitions).

Organized Fraud, Fig. 1 A simple missing trader intra-community (MTIC) fraud carousel scheme transaction chain their respondents and – for a quick summary – Black 2009). The FBI (2011) notes that its "perpetrators include licensed/registered and non-licensed/registered mortgage brokers, lenders, appraisers, underwriters, accountants, real estate agents, settlement attorneys, land developers, investors, builders, bank account representatives, and trust account representatives. There have been numerous instances in which various organized criminal groups were involved in mortgage fraud activity. Asian, Balkan, Armenian, La Cosa Nostra, Russian, and Eurasian organized crime groups have been linked to various mortgage fraud schemes, such as short sale fraud and loan origination schemes. . ..Mortgage fraud perpetrators have been known to recruit ethnic community members as co-conspirators and victims to participate in mortgage loan origination fraud." A later report (FBI 2012) gives a lucid overview of the process:

> Fraud has been identified throughout the loan process, which commences with the borrower providing false information to the mortgage broker and/or lender. The next layer of potential fraud – the corporate fraud – occurs with the banks, brokerage houses, and other financial institutions that package loans through the securitization process. As the housing market declined, subprime lenders have been forced to buy back a number of nonperforming loans. Many of these subprime lenders have relied on a continuous increase in real estate values to allow the borrowers to refinance or sell their properties before going into default. However, based on the sales slowdown in the housing market, loan defaults increased, the secondary market for subprime securities dwindled, and the securities lost value. As a result, publicly traded stocks dramatically decreased in value as financial institutions realized large losses due to the subprime securities they held or insured, resulting in financial difficulties and bankruptcies.

Though the subprime securitization racketeering was a US specialty, mortgage fraud still appears easy to commit in the UK, by solicitors (UK lawyers) in collaboration with property buyers. In Legal Futures (2012), a foreign lawyer registered with the English regulators to work in a law firm was sentenced to 7 years in prison as being "at the helm of a criminal gang that defrauded high street banks out of almost £8 m by taking out mortgages on properties they did not own" and indeed for consequential money laundering. A Mr Younas, using an alias, was director of Montague Mason Solicitors:

> He submitted false paperwork to Birmingham Midshires and Abbey banks in connection with 33 fraudulent mortgage applications in a two-month period in late 2008. His fellow gang members had previously identified suitable properties across the south-east to use for false mortgage applications – with the homeowners unaware of the crime. The fraudulent mortgages were finalised when false paperwork was completed by the solicitors, and signed by Younas, including the certificates of title requesting the banks release the funds. Once the deals were signed off and the funds transferred, the stolen money was quickly transferred from the solicitor's account into a network of bank accounts, controlled by Younas and the rest of the gang. The money was then withdrawn in cash.

The regulators had already closed down Montague Mason in February 2009, and he was struck off the register of foreign lawyers in 2010, incapacitating him from crimes that involve authorized lawyers in the UK, *provided that anyone checked the register*. Middleton (2005) has critically examined responses to frauds by English lawyers.

It is possible to review the dynamics between setting and criminal act for identity and other frauds, such as application frauds and insurance frauds. In mortgage frauds, fraud typically takes two forms: customers lying about their own means – that is, exaggerating their income – and/or falsifying documents, such as creating fake payslips that show they earn an amount large enough to justify the mortgage they need, even if it is a multiple of their real income. This can be done simply by printing fake payslips, if necessary, on a color printer. Self-certificated mortgages (at higher interest rates in the UK) were allowed to cater for the increasing number of self-employed persons who could not produce genuine payslips; in the USA, the brokers simply filled some invented number in for them. One incentive for mortgage introducers is that they are paid commission; one incentive for lenders is that they have sales targets to hit and performance bonuses to get, and nonpayment usually comes much later. In some cases, the borrower is told there is no way they are going to get the mortgage they want with their income and that they should leave that part of the mortgage application form blank. After they have gone, the broker inserts the false income. In the USA particularly, there have been widespread scandals relating to commission-hungry brokers lying to purchasers about the affordability of mortgages, which they discover only when the initial low rates expire (Financial Crisis Inquiry Commission 2011). In other cases, the would-be purchaser colludes with the broker. In other cases still, the broker (or lawyer) purchases the properties for themselves as beneficial owner, using the names and real or fictitious income details of clients. In a rising market, where there is demand (e.g., from students) for rental properties, fraudulent purchasers see little downside risk. In some cases, professionals appointed by lending institutions are aware that the lenders need to lend, and their judgment is swayed by this to give the valuation required to enable the mortgage to be granted: this is especially so where all the parties' desires are in the same direction (author interviews with surveyors, 1980s and 2008). However, when the market turns, as it did in 2007 (and earlier in the USA, where FBI warnings of a mortgage fraud epidemic went unheeded by politicians and regulators), these frauds are shaken out as people cannot keep up with repayments. (This was less common in Europe, though European banks – and later European taxpayers – ended up suffering because they purchased triple A-rated derivatives based on these nonperforming loans).

Fraud and Money Laundering

In any dataset of suspicious activity reports from banks, law firms, brokers, etc., a significant percentage of reports have "fraud" ticked as a reason for suspicion. Oftentimes, this is the product of hunch rather than specific knowledge, but there is so much confusing rhetoric about money laundering that I will try to bring some clarity here. There are in essence two definitions of money laundering. The first, more or less universalized following the success of the Financial Action Task Force in pressurizing countries around the world to legislate, is that anything that is done with the proceeds of crime to hide or store them becomes "laundering." Thus, all proceeds of fraud that are not directly consumed in "lifestyle" are laundered. The second – which may correspond more closely to what people think laundering is – is that all proceeds of fraud that are sought to be legitimated as "clean money" are laundered. The difference between these two categories depends on how much effort is made and expected to be made by civil creditors and law enforcement to trace funds.

Many parties can be involved in laundering, but the number can be as low as one. This insider involvement is not uncommon for large-scale financial frauds: just as Cressey (1955) argued that all accountants *could* commit embezzlement, the fraudster has exactly the skills required to also conceal the sources of his funds. Where offenders start out with a business that is being used as a medium for what looks like legitimate activity – which is not the case for credit card fraudsters, for example – then placement of funds in banks may not attract suspicion: corporate lawyers may be falling over themselves to offer well-paid services in the construction of corporate vehicles. They will not routinely suspect senior corporate staff of being major criminals, perhaps especially since they were appointed by them and would like to be paid by them in the future. Since many frauds would be unsuccessful if they did *not* look like legitimate activity, this gives them a structural advantage over other types of offenders. On the other hand, if corporations are paying bribes to public servants to defraud their countries or to private purchasing officers to defraud their companies, then more subtle mechanisms for concealment may be necessary. Such conduct would seldom be associated with "organized crime" as conventionally defined, however. Surprisingly, little research has taken place (or has been commissioned by governments around the world) on how criminals generally, or fraudsters in particular, launder their funds. In recent years, the greatest attention has been paid to how the proceeds of grand corruption are laundered: but the overlap between fraud and corruption means that many of these cases are really cases of fraud against the governments that both the bribe payers and the bribe recipients are committing. For good studies of this process of laundering the proceeds of corruption, see de Willebois et al. (2011) and Sharman (2012).

Enron

In the last year at Enron Global Finance group, managers were sometimes handed a list of Enron assets and instructed to go out and sell some to the special-purpose vehicles. A manager would pick something, from a plant to stock to a piece of a start-up company, and then discuss the deal with a team of internal lawyers and auditors. A bank or other investor lent money to the newly created company to finance the purchase. The new company, in turn, paid the money to Enron. The use of an intermediary was to make the loan belong to the new company, not Enron, and thus not to count as a debt on Enron's financial statement. Instead, it counted as income to Enron when the new company passed on the proceeds. Less debt and more income assured Enron would keep its high credit rating (making borrowing cheaper) and would keep the stock price up.

Top graduate school employees told the Houston Chronicle (20 January 2002) there were many uses for the vehicles that they considered legitimate, such as bringing in outside partners to share the risks of a particular venture: but there was little question, especially toward the end, that many had no real "business purpose" other than improving financial appearances. Say the asset was 100 shares of IBM stock. Enron would divide each share into two parts, one called a "control interest" and one called an "economic interest." Then it would sell the economic interest to a newly created special-purpose vehicle. The asset was rarely as simple as 100 shares of another company's stock. So Enron had to put a value on it. Because there was no real outside buyer, it decided the price itself and had that number approved by its auditor, Arthur Andersen.

The deal was placed with a bank, insurance company, or other major lender, which put up 97 % of the money. Sometimes the promise of Enron stock would be put up to guarantee the loan, although Enron stockholders were never told of the risk that their shares could be diluted if such new shares had to be issued. To qualify as "independent" from Enron for accounting purposes, a special-purpose vehicle had to be owned by someone else. So an outside entity would be brought in to make the required investment, perhaps a tiny percentage of the SPV's total start-up cash, sometimes illicitly lent by Enron itself. An employee told the Houston Chronicle (20 January 2002):

> Enron no longer owned the economic interest in the asset, but it did own control over it. In the sales contract with the vehicle, Enron promised always to act in the interest of the SPV. Lawyers and auditors said all this was OK. As the asset made money for the SPV – if it did, and many didn't – it made principal and interest payments to the lender and issued dividends to the outside equity partners, just like in a normal company.
>
> Enron got to report the proceeds of the sale of the asset as earnings. It had to repay the loan. . .but the debt didn't show up on Enron's financial statements.
>
> "Investors don't like to hear you say, 'Oh, I was wrong.' So you start having a yard sale to boost CFO (cash flow from operations) and net income," the employee said.

As the Enron indictments showed, there were plenty of accountants, bankers, and lawyers as well as some senior management willing to participate in criminal or marginal operations. But they had nothing to do with any minority ethnic or national communities (except for White Anglo-Saxon Protestant Americans), nor with any criminal subcultures as conventionally defined. Likewise, the many works on the savings and loans "failures" (Black 2005; Calavita and Pontell 1993) and on accounting frauds (Tillman and Indergaard 2005) emphasize the culpable involvement of elite networks rather than marginal firms, whereas there is no reason why elite *and* marginal firms (plus "organized criminals") cannot be involved in frauds and money laundering.

Conclusions

Globalization of fraud is affected by settings, with their rich and varied opportunities (reflecting patterns of business, consumer, and investment activities), the abilities of would-be perpetrators to recognize and act upon those opportunities (the "crime scripts" perspective), and their interactions with controls including law enforcement and formal regulation (touched only lightly upon here). Even in law enforcement circles (WEF 2012), the structure of groups is becoming less important than analyzing what people need from the largely illicit and largely licit worlds to go about the business of fraud. Such observations or claims about the contested and shifting nature of analysis over time complicate the already difficult question of whether fraud "itself" has changed over the years (Levi 2008a, b, 2012). It seems reasonable to reflect on two "historical" questions:

1. In what respects has fraudulent activity changed, in terms of the sorts of techniques and organization that are or can be used, in relation to the efforts made (intentionally or not) to prevent frauds?
2. In what respects has the world of fraud changed and what would the sort of people with the sort of skill sets/networks who committed frauds in the 1960s and 1970s have contemplated doing today?

In relation to the first question, although the basic techniques used by fraudsters 50 or even a hundred years ago are still available today, especially against those investors and trade creditors who make only modest enquiries, the professionalization of investor protection and credit management, interest by "consumer programs" on television and sections in newspapers, and websites and other electronic sources make the commission of such frauds harder. E-commerce, the growth of lightweight, high-value electronic products, and the technology of rapid delivery anywhere in the world have cut down decision times and opened up domestic and foreign markets to fraudsters operating within and from outside all developed economies. At the high end of insolvency frauds, however, it seems doubtful whether the more skillful abuses of insolvency by those who, for example, establish beneficially owned corporate fronts offshore and then create artificial debts to them which enable them to vote in friendly liquidators or administrators are any harder to commit or are any more likely to be punished today than they were 50 years earlier. Formal social control – the police and criminal courts – has not been particularly interested in frauds other than the more visibly harmful "widows and orphans" cases and those committed by gangsters, "professional criminals," and terrorists in whom they already have an interest irrespective of the frauds they have committed. In Australia and Europe, and even in smaller jurisdictions such as Mauritius, pushed by international technical assistance, there has been a growth in "civil recovery" regimes, applying financial investigation and asset forfeiture (irrespective of criminal conviction) to supplement post-conviction confiscation remedies. However, even if they have substantial savings rather than (e.g., as did the high-spending bankruptcy fraudsters interviewed by Levi 2008a, and did the telemarketers interviewed by Shover et al. 2003) spending "their" proceeds as they went along, few fraudsters are high-profile career criminals of a seriousness level that would interest enforcement agencies specializing in combating "organized crime." This question is connected to the second.

The dominant mode of thinking about serious crime for gain has been the "rational choice" model implicit not only in the economic analysis of crime but also in situational crime prevention "theory" (see Schuchter and Levi 2013). The selection of fraud in general or particular forms thereof in particular depends on the confidence, skills, and contact set of any given individual offenders. The presence or absence of "crime networks" known to and trusted by the willing offender makes a difference to "crime capacitation": an issue often neglected in individualized explanations of involvement in crime. Choice of crime type might also be affected by age. Those offenders who were in their 50s and over might be deterred by the technological challenges of cyber activities, and the age gap might apply to co-criminality as it does to other feature of contemporary life. So today's new generation fraudsters might gravitate toward more hi-tech forms of fraud, whereas if they were in late career, it might seem too risky to adapt in unfamiliar territory unless they can find someone younger to collaborate with. Some fraudsters display a remarkable aptitude for creativity and constant testing out of commercial systems and private individuals for signs of weakness. This focus on "criminal transferrable skills" – the set of aptitudes including social networking that individual/sets of offenders have – concentrates our attention on offender creativity, energy and social networking skills in finding co-offenders (or "turning" non-offenders into co-offenders), and in adapting techniques: many offenders (and non-offenders) lack one or all of these qualities.

With only modest sophistication, the Internet has made it easier for foreign natural and legal persons to defraud consumers and suppliers, for example, via counterfeited or cloned payment cards. Fraudsters could be involved in the theft of personal data from garbage ("bin raiding") or by hacking into data storage facilities; or they could simply send "phishing" emails offering academics invitations to conferences on "child protection and the economic crisis" to try to con them into giving their personal details for the free visa and transport arrangements that will later turn out to be illusory. They could also be involved in account manipulation by insiders, whether in call centers or elsewhere. (The offshoring of call centers led to periodic media alarm stories about blackmail and corruption in India: but it is nonsensical to think that this cannot happen in the developed world, with badly paid, high turnover staff ratios. Indeed, there may be tougher regimes in Indian call centers – staff searches and prohibitions on mobile phones – than might be allowed in Europe.) Rings of staged accidents with claims for hard-to-falsify personal injuries would be within the skill set of many (once they worked out what to do), as would organized benefit fraud and – especially – the sale of counterfeit products, whose quality digital technology has done so much to improve. For the more adventurous, scams can involve some currently fashionable musical or sporting events, or the Muslim Hajj, or a social cause such as "renewable energy" or commercialized carbon trading. The underlying concepts were available to investors centuries ago, at the time of the South Sea Bubble or the Dutch Tulip Mania, but every generation has to learn the lessons for itself. When times get hard, people may take more risks to avoid downward mobility; when times are good, everyone wants to get onto a rising bandwagon. Both circumstances offer opportunities to fraudsters.

Recommended Reading and References

Black W (2005) The best way to rob a bank is to own one. University of Texas Press, Austin

Black W (2009) The two documents everyone should read to better understand the crisis. http://www.huffingtonpost.com/william-k-black/the-two-documents-everyon_b_169813.html. Accessed 11 March 2013

Calavita K, Pontell H (1993) Savings and loan fraud as organized crime: toward a conceptual typology of corporate illegality. Criminology 31(4):519–548

Cressey D (1955) Other people's money. Free Press, New York

de Willebois, van der Does E, Emily H, Robert H, Ji Won P, Jason S (2011) The puppet masters: how the corrupt use legal structures to hide stolen assets and what to do about it. World Bank, Washington, DC

FBI (2011) 2010 Mortgage fraud report: year in review. FBI, Washington, DC

FBI (2012) Financial crimes report to the public. FBI, Washington, DC

Felson M (2003) The process of co-offending. In: Smith MJ, Cronish DB (eds) Theory for practice in situational crime prevention. Criminal Justice Press, Monsey, pp 149–167

Felson M, Boba Santos R (2010) Crime and everyday life. Sage, Thousand Oaks, CA

Forsyth N, Castro E (2007) Other people's money: the rise and fall of Britain's boldest credit card fraudster. London, Panv

Holtfreter K, Van Slyke S, Blomberg T (2005) Sociolegal change in consumer fraud: from victim-offender interactions to global networks. Crime Law Soc Chang 44(3):251–275

Holtfreter K, Reisig M, Pratt T (2008) Low self-control, routine activities and fraud victimization. Criminology 46(1):189–220

Legal Futures (2012) Struck-off lawyer jailed for £8m mortgage fraud. http://www.legalfutures.co.uk/latest-news/struck-off-lawyer-jailed-for-mortgage-fraud. Accessed 11 March 2013

Levi M (2008a) The phantom capitalists: the organisation and control of long-firm fraud, 2nd edn. Ashgate, Aldershot

Levi M (2008b) Organised fraud: unpacking research on networks and organisation. Criminol Crim Justice 8(4):389–420

Levi M (2012) Trends and costs of fraud. In: Doig A (ed) Fraud: the counter fraud practitioner's handbook. Gower Publishing, Andover, pp 7–18

Middleton D (2005) The legal and regulatory response to solicitors involved in serious fraud. Br J Criminol 45(6):810–836

Morselli C (2008) Inside criminal networks. Springer, New York

Nguyen, Pontell (2010) Nguyen, TH and HN Pontell, Mortgage origination fraud and the global economic crisis: a criminological analysis. Criminology and Public Policy 9(3)591–612

Piquero NL, Benson M (2004) White-collar crime and criminal careers: specifying a trajectory of punctuated situational offending. J Contemp Crim Justice 20(2):148–165

Schoenmakers Y, de Vries Robbé E, van Wijk A (2009) Mountains of gold: an exploratory research on Nigerian 419-fraud. SWP, Amsterdam

Schuchter A, Levi M (2013) The Fraud Triangle revisited Security Journal, 2013/02/04/online doi:10.1057/sj.2013.1

Sharman J (2012) Chasing kleptocrats' loot: narrowing the effectiveness gap. Chr. Michelsen Institute, Bergen, U4 issue August 4

Shapiro SP (1980) Collaring the crime, not the criminal: Reconsidering the concept of white-collar crime. American Sociological Review 1990;55 (3):346–65

Shover N, Coffey GS, Hobbs D (2003) Crime on the line: telemarketing and the changing nature of professional crime. Br J Criminol 43(July):489–505

Stevenson R (1998) The boiler room and other telephone sales scams. University of Illinois Press, Urbana

Stevenson RJ (2000) Boiler room and other telephone sales scams. University of Illinois Press, Champaign, Il

Sutherland E (1945) Is "white collar crime" crime? Am Sociol Rev 10(2):132–139

Sutherland E (1985) White-collar crime: the uncut version. Yale University Press, Princeton

Tillman R, Indergaard M (2005) Pump and dump: the rancid rules of the new economy. Rutgers University Press, New Brunswick

van Duyne P (2000) Mobsters are human too: behavioural science and organised crime investigation. Crime Law Soc Chang 34:369–390

van Koppen M, de Poot C, Blokland A (2010) Comparing criminal careers of organized crime offenders and general offenders. Eur J Criminol 7(5):356–374

WEF (2012) Organized crime enablers. World Economic Forum, Geneva, http://www.weforum.org/reports/organized-crime-enablers. Accessed 11 March 2013

Organizing and Supplying Forensic Science Services

Christopher Lawless
University of Edinburgh, Edinburgh, UK

Overview

The issue of how to organize and regulate the provision of forensic scientific services is a topic which has attracted increasing discussion. Concerns have been expressed about whether police control of forensic laboratories best serves the interests of justice. Elsewhere, instances of official scrutiny have made a link between the efficiency of forensic science laboratories and the capacity of police forces to combat crime. In response to ongoing controversy over the quality and timeliness of forensic science, recent decades have seen an increase in calls to reform its provision along market-led lines. Such reforms have been pursued with particular vigor in England and Wales. Developments in this jurisdiction have however raised considerable controversy, particularly in the light of the announcement in December 2010 to close the leading forensic science provider, the Forensic Science Service (FSS).

Introduction

This entry first devotes attention to exploring the arguments for and against the regulation of forensic science through competitive and market-led means. Attention is drawn upon the experience of England and Wales in order to give a flavor of the empirical consequences of such policies. Attention is also devoted to exploring how discourses of "marketization" and "liberalization" have been differentially interpreted and the resultant implications for the epistemic character of forensic science. While less data currently exists regarding developments in other jurisdictions, A brief overview is provided of those states where reform has been considered or where it poses further issues.

This entry is organized to provide information regarding the following topics:

– Arguments and counterarguments for the commercial reform of forensic science
– The empirical impact of market reforms to forensic science?
– The effect of commercialization on the epistemic character of forensic scientific knowledge
– Measuring and codifying the "value" of forensic science
– Comparing the experience of commercialization in England and Wales with other jurisdictions

The Arguments for and Against a Competitive Market for Forensic Science

Concerns over the quality of forensic science have been accompanied by calls to reconsider ways of organizing the production of scientific knowledge within criminal justice systems. Suggested reforms to the shaping and provision of forensic science have sought to redress perceived disadvantages to defendants (see ▶ Cognitive Forensics: Human Cognition, Contextual Information, and Bias). Other calls for reform have sought to improve the efficiency of analyses (see ▶ Forensic Science Effectiveness) and to help maximize the use of forensic science by police forces (see ▶ Forensic Science and Criminal Investigation).

The effect of institutional structure on the nature of forensic scientific knowledge has received particularly comprehensive attention through the work of Roger Koppl (2005, 2006, 2010). These studies of forensic science in the USA have drawn upon a series of theoretical and methodological perspectives from inter alia economics, social epistemology, and organization studies. These studies indicate that current modes of organization have led to poor-quality work being performed in laboratories, with high risks of error, bias, or even outright scientific fraud. Koppl (2005) argues that eight features of forensic science organization in US laboratories contribute to suboptimal quality: (a) the *monopoly* that each laboratory holds over the evidence it analyzes, with no other laboratory likely to analyze the same evidence; (b) *dependence* on police forces, under which they are normally organized and whom they depend on for funding; (c) weak *quality control* systems (see ▶ Forensic Science and the Paradigm of Quality); (d) ability of forensic scientists to gain *information about case circumstances* which may prejudice how they interpret evidence (see ▶ Cognitive Forensics: Human Cognition, Contextual Information, and Bias); also (e) *no division of labor between analysis and interpretation* (i.e., the scientist who performs the technical work of an analysis also interprets the results and reports them to police); (f) *lack of forensic counsel* for defendants, creating an information asymmetry; (g) *lack of competition to provide defense services*; and (h) *public ownership* of forensic laboratories which Koppl argues reinforces the dependence on police. This dependence is also seemingly compounded by performance criteria for law enforcement agencies which include numbers of arrests, prosecutions, and convictions. Koppl (2010) argues that these conditions promote bias toward the prosecution: "When making a monopolistic subjective judgment about ambiguous evidence, forensic scientists in the employ of law enforcement agencies have an incentive to interpret evidence in ways that support the interests of law enforcement" (Koppl 2010, p. 79).

In addressing these perceived shortcomings, Koppl has suggested a number of measures. Perhaps the most notable of these interventions concerns proposals to overturn the monopoly control that US forensic laboratories hold over evidential analysis. Koppl (2006) uses the term "monopoly epistemics" to describe this arrangement. Drawing upon experimental data, Koppl argues that the analysis of forensic evidence at multiple sites decreases the chance of error or bias and that the presence of multiple suppliers of information to the judge and jury improves decision-making on the part of the latter (Koppl 2005). Building on these arguments, a different arrangement has been proposed, described as "competitive self-regulation" (Koppl 2005). This involves the introduction of competition within a jurisdiction, with multiple laboratories able to analyze the same evidence. Forensic scientists in one laboratory may not know whether evidence in a specific instance is also being analyzed elsewhere, but awareness of this possibility should facilitate a higher quality of scientific work. Competitive self-regulation also involves the hiding of extraneous, circumstantial information from forensic analysts so that work proceeds in a context-free environment.

While the rigorous application of economic and organization theory to forensic science administration is to be welcomed, certain criticisms of the work of Koppl et al. can be identified. The argument that the quality of forensic scientific work improves in the absence of contextual information is not accepted by all (see ▶ Forensic Science Culture and ▶ Forensic Science and Criminal Investigation). This therefore remains a key point of contention. Even if one accepts Koppl's position, the question remains as to what exactly constitutes "contextual" information and to what extent it can be fully excluded from analysis. Koppl's arguments also do not reflect the distinction between traces (collected at a crime scene) and the subsequent production of evidence presented in court. A variety of processes may be involved in transforming traces into evidence, and forensic scientists themselves have multiple roles (forensic science culture).

Koppl's suggestions for reorganizing forensic work also do not allow for the possibility that the same individual employees may be involved in a reformed system. Previous career experience of casework could still permeate into forensic practice even in a rearranged working environment, which could perpetuate bias. Another issue concerns the involvement of multiple laboratories. While in theory this could be effective, serious practical problems could result from a lack of sufficient evidence to distribute to multiple sites of analysis. DNA evidence, for example, may only be present in trace quantities which might preclude multiple analyses. Further, if single evidential items are recovered from an incident, the nature of certain forensic tests may irreversibly alter the nature of the source evidence which might not make it amenable to subsequent tests under "blind" conditions.

Koppl et al. also argue in their work for the replacement of state dependency with a privatized mode of forensic science provision. In justifying this argument Koppl regards publicly owned laboratories as having weak financial incentives to improve the quality of work. Other commentators have also previously advocated the advantages of private firms entering forensic science markets. In an earlier assessment of forensic science in England and Wales, Gallop (1992) argued that the presence of independent firms would be advantageous to both defendants and forensic scientists working alongside the police. Gallop argued that the presence of an independent sector could remove the burden from government scientists being distracted by defense considerations (although in some jurisdictions, two scientists may respectively work for opposite sides in a case). With defendants being well served by commercial options, government scientists would be free to fully focus on prosecution work. Gallop also argued that an expansion of the sector naturally provided more career opportunities for government scientists, contributing to a healthy employment base and sense of professional collegiality.

Gallop (2003) made later arguments in favor of the route to privatization, advocating further potential benefits including (Gallop 2003, p. 59):

– Proper choice for police forces, allowing them to be more demanding in terms of quality, cost, and speed of service
– An expansion of the number of forensic laboratories
– Easier access to the forensic science community
– Better quality assurance standards

Fraser (2005) argued that privatizing reforms were necessary to improve the efficiency and turnover rate of casework. Even with a relatively modest degree of reform, Fraser argued that the speed of delivery in England and Wales compared far more favorably to publicly funded laboratories in other jurisdictions, seen to be characterized by incessant backlogs. A 2002 report of forensic science in the USA, for example, found a backlog of 500,000 cases (Fraser 2005, p. 119). More recent data from New Zealand also supports the claim that some form of economic restructuring has helped to clear backlogs (Bedford 2010).

As we shall see, market reforms in the UK were largely motivated by efficiency concerns. Not all commentators however agreed that such policies were appropriate for forensic science. In one prominent early intervention, Roberts (1996) provided a critical review of the arguments for marketization. In addressing the issue of market-based reforms, Roberts scrutinized the arguments for a free market in forensic science on its own terms. Considering the economic arguments "squarely on their merits" (Roberts 1996, p. 56), Roberts made three predictions which argued against the use of market mechanisms in forensic service provision. First, was the argument that coordination problems would prevent the production of necessary public goods, in this case research findings and datasets which could lead to improvements in forensic science as a whole. Roberts argued that the commercialization of forensic science held no incentives for actors to pool resources to promote innovation in the area, nor to provide resources of common utility or to improve standards. Second was the argument that the predominant position of the UK's FSS would block fair and true competition. Rather than creating an environment which promoted efficiency through competition, the "vast resources and unparalleled expertise" of the FSS were viewed at the time as risking driving smaller companies out of the market, leaving the FSS "with an unregulated monopoly interest" in forensic service provision. Third, and most significant according to Roberts, was that criminal justice only resembled a market in an "imperfect and attenuated sense" (Roberts 1996, p. 47). Roberts argued that a market system would fail to meet the needs of all its "consumers," particularly defendants. Access for this group to forensic scientific services was viewed as being determined not by the laws of supply and demand but by the availability of Legal Aid (Roberts 1996, p. 56). Another key consumer group, the police, was seen as constrained by budgetary limitations and the need to balance other spending priorities. Roberts argued that the emergence of charging for forensic services would risk police forces selecting cheaper (but not necessarily high-quality) services or cutting back on their use of forensic science altogether.

The empirical consequences of attempts to shape a forensic marketplace in England and Wales are outlined in the next section. Particular attention is devoted to this jurisdiction as it is here where commercialization has been pursued in a particularly systematic and comprehensive way, and for which a substantial amount of data exists. As we shall see however, commercialization has aroused considerable controversy, linked in part to the proposed plan to close the FSS.

Tracing the Emergence of Forensic Marketplaces: The Example of England and Wales

The effectiveness through which police forces in England and Wales apply scientific techniques in the course of their work became the subject of considerable official discussion from the late 1970s onward. In 1987, a survey conducted by accountants Touche Ross concluded that police management of scientific support was "generally poor" (Touche Ross 1987). It portrayed an environment in which forensic technologies had begun to significantly expand in terms of potential scope but had not been fully utilized to address a rise in serious crime. Among the many issues considered in this report was the scope for changes in the method of funding and organization of the FSS, the sole provider of forensic science to the police at the time.

The FSS was then a state agency, publicly funded by a central grant. Concluding that the police management of scientific support (both in-house and externally provided by the FSS) was "generally poor" (Touche Ross 1987, p. 3), the report recommended the appointment of managers of scientific support in each force. Each such manager would be employed by police forces (as opposed to an agency such as the FSS) and be responsible for the provision of all scientific services and the management of their own force forensic science budgets. This step represented one sign of the growing devolution of budget responsibilities to operational law enforcement actors.

The Royal Commission on Criminal Justice which reported to Parliament in July 1993 recognized the emergent policy discourses surrounding the possibility of market reform of forensic science. The Royal Commission had been established in the light of several high-profile miscarriages of justice and in direct response to the decisions by the Court of Appeal to quash the convictions of the Guildford Four and the Birmingham Six. While the Commission addressed a wide-ranging series of issues relating to criminal justice, two notable themes were advanced in its conclusions regarding forensic science (Roberts 1996). First, it provided "cautious endorsement" (Roberts 1996) of further development of free market competition for forensic scientific services. Secondly, it recognized the considerable growth in the number and variety of private firms offering scientific support to both prosecution and defense, which reflected concerns about how the quality of forensic science could be maintained (Roberts 1996, p. 39).

As the 1990s progressed, other firms had begun to emerge in direct competition with the FSS, such as LGC, Scientifics Ltd, and Forensic Alliance. During the same period, the FSS introduced product-based charging, where each product, such as a body fluid search, tool mark examination, or cannabis identification, was defined as encompassing a wider set of activities. This new definition of "product" as activity was intended to reflect more closely the actual work performed, "thus providing customers with a better understanding of the true costs of services and enabling them to make informed judgments about their value" (NAO 1998, p. 30). As the pricing strategies of forensic service providers began to become more sophisticated, sign emerged that forces were exploiting the new market conditions. Communication between the FSS and police improved, and the FSS educated police officers about the value of certain forms of evidence.

Some of these assertions about the benefits of the emerging marketplace were supported in the government-commissioned McFarland review of the FSS, completed in 2003. This review explicitly addressed the "effectiveness of the organization in meeting the needs of the criminal justice system in terms of the quality, timeliness and cost effectiveness of its services" (McFarland 3.1). McFarland also assessed the management and business structures of the FSS in the light of possible increased competition for forensic science services (McFarland 2003). Forensic science providers (FSPs) were, at this time, offering volume discounts and loyalty schemes, which convinced the McFarland review that a "truly competitive market" was "beginning to develop" in UK forensics (McFarland 2003). Furthermore, the review went on to argue that the introduction of "best value" principles "had forced the police to seek better value for money in the bought-in services" (McFarland 2003), and that the police had become "informed customers," playing off "suppliers against each other." McFarland concluded that police forces were adapting well to the new market conditions and that true competition was "beginning to develop" in UK forensics (McFarland 2003, Section 3.1). Police reported that they had begun to receive "a more personalized and responsive service" from suppliers, which was taken as a sign that competition had started to yield the kind of benefits claimed by advocates of marketization, such as "greater choice, value for money and improved service delivery" (McFarland 2003, Section 3.3). The McFarland review also recommended that the FSS should make the transition toward full privatization, although this never occurred. The FSS was however converted into a "government company" ("GovCo") in 2005, obliged to run on a profit-making basis.

At around the same time, Fraser (2003) outlined a typology of this evolving "forensic marketplace." This analysis distinguished between *strategic suppliers*, *primary niche suppliers*, *secondary niche suppliers*, *primary individual suppliers*, and *secondary individual suppliers*. Here, *strategic suppliers* were regarded as providing a comprehensive range of police support services and able to supply these services in an integrated way. These organizations had the potential however to be relatively inflexible in terms of the products and services they offered and potentially highly bureaucratic. *Primary niche suppliers* were perceived to display a similar level of experience in the forensic sector to strategic suppliers, but to differ in that they were seen to provide a far narrower range of products or services, albeit to a high standard. *Secondary niche suppliers* were to seen differ in that the main functions of the organization may *not* be to provide forensic science, but that this may be one of a portfolio of different scientific services. The knowledge of the broader context of criminal justice that secondary niche suppliers held could potentially be highly variable. *Primary individual suppliers* were described as individuals who offer particular forms of forensic scientific expertise. This may be highly specialized but might be vulnerable to questions of competence and quality. Finally, *secondary individual suppliers* described individuals who held specific scientific expertise which might appeal to some of the needs of police forces but not exclusive to forensic science. The experience of criminal justice contexts of such individuals might vary and again were regarded as potentially vulnerable to questions of competence.

Fraser's typology reflects how the modern marketplace for forensic science in England and Wales had encompassed a diverse array of FSPs with a considerable degree of variance in terms of their size, scope, and product range. Fraser's overview also highlighted the ongoing concerns over the consistency and quality of scientific work produced within this marketplace. Key oversight bodies such as the National Policing Improvement Agency (NPIA) sought to standardize the level of service provision from FSPs. In August 2008, 12 FSPs, including the FSS, agreed to participate in the National Forensic Framework Agreement (NFFA), which sought to function as an organizing mechanism for forensic science procurement by most police forces in England and Wales. A growing proportion of forensic scientific work remained in the hands of the police forces themselves, however, through the presence of in-house facilities.

The NFFA represented a notable reinforcement of the market system. A BBC radio documentary broadcast in December 2009 however identified serious issues with the commercialization process. It was reported that forensic scientists often disagreed with police over the best course of action during casework, although they were constrained by the contractual arrangements imposed by the NFFA. The program gave an example of one forensic scientist who had contributed to the successful investigation of a case only by ignoring police instructions and pursuing analyses which had not been originally commissioned.

More seriously still, the same documentary reported that the FSS was experiencing severe financial difficulties and had become dependent on government bailouts. By this time the FSS had implemented a "transformation program" which included the closure of three of its regional laboratories, which in turn contributed to speculation over the future of the organization. The announced closure of the FSS by UK Government in December 2010 nonetheless surprised many. It also stimulated a high degree of controversy, with many commentators expressing fears about the future viability of scientific support to police forces. In response the House of Commons Select Committee on Science and Technology conducted an inquiry into the proposed closure, which also focused on the state of the forensic market in England and Wales as a whole.

Although the Committee experienced significant difficulties assessing the exact condition of the forensic science market, the findings indicated that the overall size of the market was in serious decline (House of Commons 2011, p. 15). Police expenditure on external forensic science provision has decreased markedly in previous years. Police forces had, however, invested increasingly in their own "in-house" forensic scientific provision over the same period. While figures for 2010–2011 show a slight decrease in overall police spending on both internal and external provision, the former continued to considerably outweigh the latter. In 2010–2011, the total expenditure of police forces in England and Wales on internal provision was £181 million, compared to £138 million for external services over the same period (House of Commons 2011).

The inquiry heard conflicting evidence to explain the shrinkage of the market. The NPIA claimed that the reduction in external spend had been caused by the NFFA, leading to forces enjoying considerable savings on their external spend. Other witnesses to the inquiry pointed out however that forces may have viewed in-sourcing as a cheaper option in the light of wider budget cuts (House of Commons 2011, p. 17). The House of Commons Committee criticized the trends toward in-sourcing. Such a shift, with the customer becoming a competitor to external FSPs, was viewed as undermining the very idea of a fully functioning market for forensic science. The Committee also reported "almost unanimous alarm" (House of Commons 2011, p. 35) about forensic science in police laboratories, which were not subject to the same quality accreditation procedures as those of external FSPs. The inquiry concluded by expressing concerns "about the risks to impartiality of forensic evidence produced by nonaccredited police laboratories" (House of Commons, p. 43). It also highlighted what it perceived to be a more serious risk, namely, "the introduction of bias based on selective forensic examination of exhibits, arising from the need to make savings." The Committee concluded that a reliance on in-sourced forensic practitioners would increase the risk of the collection of evidence being influenced by immediate policing concerns rather than working in the more balanced interests of justice.

Commercialization and the Epistemic Character of Forensic Scientific Knowledge

The possibility of increasing in-sourcing returns us to some of the concerns expressed by Koppl et al. regarding the effect of organizational context on forensic reasoning. The question of how commercialization has affected the epistemological foundations of forensic science is one that has been recently explored by the author and colleagues (Lawless 2011; Lawless and Williams 2010). These studies have explored how liberalizing policies have been interpreted by actors in different ways, manifesting themselves in varying infrastructural arrangements for the market-led provision of forensic science. Differing initiatives can be seen to construct the epistemic character of forensic science in differing ways.

Initiatives such as the Case Assessment and Investigation (CAI) model, originally developed on behalf of the FSS, arose partly due to the challenges posed by the introduction of charging. CAI, a framework for evaluating evidence incorporating principles of Bayesian probability, was aimed at reforming practices of evidence interpretation and first introduced during 1990s (see ► Probability and Inference in Forensic Science). At the same time, it was also intended to help police optimize value for money from forensic science, as the newly introduced market conditions began to take hold. The type of forensic knowledge produced via CAI displays a markedly distinct epistemic character (Lawless and Williams 2010). Here, the evaluation of evidence is reported in probabilistic terms, allowing for both the contextual interpretation of "prosecution" and "defense" positions. This reflects the intent to shape a more balanced view of forensic evidence. CAI consciously facilitates the reporting of interpretative data in a way which conveys its utility to a specific investigation *and* the financial cost of its processing. The initiative therefore represents a manifestation of marketized forensic knowledge which seeks to manage the relationship between the "producers" and "consumers" of forensic science, in a way which accounts for both epistemic risk and budgetary concerns.

The kind of forensic epistemology promoted by CAI has had to compete however with contending appropriations of commercialization. As described above, the NFFA served to formalize the manner in which the various police forces enter into service agreements with external forensic science providers (FSPs). It aimed to guarantee to police forces a standardized form of scientific support across a range of criminal investigative functions. These functions, or "work packages," range from routine scientific analyses in areas such as DNA analysis, fire investigation, or footwear analysis to far more complex "casework" packages in which forensic scientists may act in a more investigative role. Each work package is highly specified and costed. The NFFA has however received significant criticism for constraining the investigative contribution forensic scientists can make to criminal casework. Research by the author (Lawless 2011) indicates that NFFA is regarded as having severely challenged the professional autonomy of forensic scientists. Elsewhere, NFFA is reported as having maximized the "commodification and disaggregation of supply" (House of Commons 2011) with a number of different providers handling different tests in a single investigation. This has prevented the systematic exchange of "contextual information and scientific results as anticipated by CAI" (House of Commons 2011). The NFFA is perceived to have seriously reduced the ability of scientists to use their initiative in the course of investigations, giving more relative power to the police. Although some scope exists within the NFFA for FSPs to explicitly provide evidence interpretation services, these are costed more highly than basic analytical work. One witness to the inquiry alleged that too often, police took the cheaper option: "Getting a DNA profile does not necessarily solve a crime but is a lot cheaper than interpretation of how the DNA got there, which is the more important aspect of successfully solving a crime" (House of Commons 2011, p. 22).

CAI and NFFA represent different ways in which marketization has been interpreted to reform forensic science. Scientists may prize scientific objectivity, but police may simply demand want technical support and may be less keen on the investigative contribution scientists might want to contribute. The presence of contending appropriations of commercialized forensic science exposes the coordination problems facing a heterogeneous set of stakeholders. These actors have experienced considerable difficulty in mediating between differing expectations of forensic science. Disagreements manifest themselves concerning precisely what constitutes the "products" in the forensic marketplace. Uncertainty about the nature of the final "product" also impacts upon problems concerning how the overall quality of these "products" can be judged and ultimately how they should be priced. Judging the "value" or "effectiveness" of forensic science has therefore become a highly complicated matter.

Likewise, measuring the quality of forensic science is fraught with a number of complications. Firstly, forensic science encompasses a considerably wide array of specialisms, some of which are employed more often than others. Establishing generic forms for assessing all of these provides a significant challenge. Second, the definition of quality as "fitness for purpose" raises problems in a criminal justice context. Multiple purposes are apparent. For example, forensic science can be used to eliminate an individual from suspicion as well as helping to secure a conviction. Third, complications are apparent if one considers that the *production* of forensic scientific data can be separated from its *interpretation* in the course of specific case circumstances. In the adversarial context, poor-quality scientific work may still be able to help secure a conviction. Likewise, the products of high-quality scientific labor may still fail to satisfy the court (Gallop 1992). Efforts to measure the quality of forensic science are complicated by questions about exactly *who* is the final “customer” of forensic science (York 2011) – with inter alia police, defendants, lawyers, and victims of crime all raising a claim but with different needs and priorities.

While England and Wales has the most extensive experience of the marketization of forensic science, other jurisdictions have responded in different ways to the challenge of organizing and procuring forensic science. In the next section, a brief overview is provided of developments elsewhere.

Forensic Marketplaces in Other Jurisdictions

In Scotland and Northern Ireland, forensic science is provided via single organizations working alongside the police, and in the Republic of Ireland, forensic science is still controlled by the country’s police force, the Garda Siochana. In the USA, it has long been recognized that the commercial sector played an important role in advancing the development and use of forensic science, particularly recent technologies like DNA profiling. However, while commercial science firms have supplied equipment, and in some cases consulting services, most forensic laboratories remain under public control, via law enforcement agencies.

US forensic science has been regarded as having existed in a largely unregulated environment (Koops et al. 2010). Partly in response to concerns about the status of forensic science, the US National Academy of Sciences submitted a report to the Department of Justice in 2009. Among other shortcomings, the report highlighted an array of serious resource disparities across local, state, and federal law enforcement agencies (see ▶ Critical Report on Forensic Science, NAS 2009, p. 5). Severe inequalities in skilled personnel, equipment, and funding were identified. Although activities at the federal level were found to be relatively well-supported, state and local jurisdictions, the sites of the vast majority of criminal law enforcement, were often found to be “sorely lacking” in the kinds of resources needed to facilitate and maintain suitably robust forensic laboratory infrastructures (NAS 2009, p. 6). The report concluded that these inequalities strongly contributed to a substantial variation in the “depth, reliability, and overall quality” of information arising from forensic examinations (ibid). Resourcing problems were also found to continue to perpetuate severe backlogs (page ref).

The NAS proposed a solution to these types of issues which significantly contrasted with the reforms implemented in England and Wales. Rather than promoting further market competition, the NAS argued for the establishment of a publicly funded “National Institute of Forensic Science” (NIFS), an independent federal body to oversee the development of forensic science at all jurisdictional levels within the USA. Overall however, the question of the economics of forensic science received relatively little attention in the report compared to other concerns such as standards and interpretation issues.

Marketization has yet to be implemented in most other jurisdictions, although there are some examples where a concern with the economics of forensic science is evident. In New Zealand, for example, forensic scientific support is provided solely by Environmental Science and Research (ESR), a government-owned research institute which since 1992 has operated on a “user-pays” system similar to the direct charging model associated with the early reforms to forensic science in England and Wales. ESR currently operates on a “fee for service” billing model where every product or service is priced (Bedford 2010). Although most prices are unit-based (“units” referring to, e.g., a DNA test, a drug identification, and a screening procedure), some pricing is time based. This brings with a considerable accounting overhead including the need for a sophisticated laboratory information management system (LIMS) and financial management system (Bedford 2010).

There are some signs that some jurisdictions are beginning to seriously consider implementing competitive market systems. Most notably, Koops et al. (2010) have recently carried out an assessment of the forensic scientific landscape in the Netherlands. In their study they explicitly address the question of whether some degree of competition should be introduced in this jurisdiction. In their exploration of this question, they tentatively conclude that competition may be an option for some specialisms within forensic science, provided the quality of providers can be guaranteed and competition is introduced on a gradual basis. In their survey, however, they found opinions to be divided about whether competition could be introduced and at what pace.

Little data is currently available on attitudes to the marketization of forensic science in emerging or developing economies. De Ungria and Jose (2010) do however provide some illumination about the economics of forensic science in the Philippines. They point out that in this kind of economy, the establishment of technologies such as DNA databases involves difficult political decisions to allocate a significant proportion of resources from an already limited national budget (De Ungria and Jose 2010, p. 324). Such technologies have to compete for budget priority with a raft of fundamental concerns such as those relating to basic healthcare, education, and poverty amelioration. At issue here are matters of tradeoffs between short – and long-term national benefits. Efforts to address health, education, and welfare issues hold tangible, perceptible benefits. Investment in DNA databasing *may* carry long-term benefits in terms of crime reduction and improved national security but represents something of a calculated risk. The devotion of resources to expensive technologies where the future expected utilities are uncertain is a particularly difficult decision for developing economies. Further, the successful implementation of advanced forensic technologies naturally involves a well-trained and motivated workforce and requires a suitably robust management and technical infrastructure, which comes at a price. These kinds of overheads may place an extra burden on the crime and policing budgets of developing nations. Some countries may opt to rely on the expertise in university science sectors, which has been suggested by some English commentators (Mennell 2006).

Given the risks of the success of such technology, and the need to provide common goods, it is difficult to envisage what clear incentives exist for commercial firms to enter such markets. Possible intervention from such firms may come about only with considerable demands attached. Research in science policy has indicated that multinational firms may be otherwise reluctant to transfer technological knowledge and knowhow to developing economies (Pavitt and Patel 1999). It may be that the costs for commercial firms to invest heavily in forensic science in the developing world may outweigh the benefits. Much more research however is needed before any concrete conclusions can be reached.

Conclusion: Future Directions

Further possibilities for research can be gleaned from this exploration of forensic marketplaces. For example, more work is needed on how context shapes the epistemic character of forensic science (see ▶ Forensic Science Culture, 9). The work of Koppl and colleagues provides some very useful data to highlight the consequences of poor-quality forensic science and of the possibilities of competitive regulation. More research in the area of forensic science administration needs to be done however to understand the way in which forensic scientific reasoning is contextually shaped in actual criminal investigative contexts (see ▶ Forensic Science and Criminal Investigation). In particular, work in this area could focus on further comparisons of the epistemic character of forensic science performed by internal and external providers to the police. This could help meet one pressing issue raised by the recent House of Commons inquiry concerning the possible differences in quality of scientific work carried out in external laboratories and in-sourced facilities. Comparative work could also focus on how certain items are selected for forensic analysis and how forensic traces are collected at the crime scene.

There is also a pressing need for further research on actual or potential forensic marketplaces in other jurisdictions. While the experience of England and Wales provides a useful case study, strong caution must be exercised with regard to the conclusions that can be drawn from a single example. While developments in other jurisdictions across Europe, North America, and Australasia remain of interest, it is equally important to consider how the organization and provision of forensic science is progressing in non-Western and emerging economies. It is in these latter jurisdictions where interesting questions may be raised regarding how the provision of forensic science, possibly through commercialized means, interacts with broader questions regarding economic development and human rights.

The economic outlook across Europe and indeed globally has been highly uncertain. How the character of forensic science may be affected by an adverse (and largely unforeseen) economic climate is only starting to be understood. Many governments have had to make considerable cuts to public spending. In the UK, these spending cuts may strongly impact upon the police, and further cuts to force budgets are likely to continue to affect their spend on forensic science. Police forces have to balance a series of spending priorities, including personnel, transport, equipment, and trained animals, and studies suggest that forensic science may lie low on their priorities (Roberts 1996). The experience of the employees of FSPs under market conditions has not inspired confidence so far, and the future of these firms, and the market system in England and Wales as a whole, is far from certain. At the same time, it is unclear whether forces will continue to invest in further in-house capacity. This may be a difficult decision for forces, as the development of such resources themselves comes with considerable start-up costs and overheads, which appear to be increasing given the need to meet a wider and more stringent range of standards (see ► Forensic Science and the Paradigm of Quality).

This brief overview is nevertheless sufficient to indicate the uncertainties which lie ahead for forensic marketplaces in England and Wales and elsewhere. Predicting the future arrangement of forensic science in an era of rapid technological, political, and economic flux arguably represents the most pressing challenge to researchers.

Recommended Reading and References

Bedford K (2010) Forensic science service provider models – is there a “best” option? Conference paper presented to the 20th International Symposium on the Forensic Sciences, Sydney, 5–9 Sept 2010

De Ungria MC, Manguera J (2010) Forensic DNA profiling and databasing: the Philippines experience. In: Hindmarsh R, Prainsack B (eds) Genetic suspects: global governance of forensic DNA profiling and databasing. Cambridge University Press, Cambridge

Fraser J (2003) Delivery and evaluation of forensic science. Sci Justice 43:249–252

Fraser J (2005) Who guards the guards? Sci Justice 45:119

Gallop A (1992) Points of view. J Forensic Sci Soc 32:59–67

Gallop A (2003) Forensic science coming of age. Sci Justice 43:55–59

House of Commons Select Committee on Science and Technology (2011) The forensic science service. HMSO, London

Koops B-J, Leukfeldt R, Hoogenboom B, Stol W (2010) Marktwerking in het forensisch onderzoek. Een veldverkenning. Tilburg Institute for Law, Technology and Society, Tilburg

Koppl R (2005) How to improve forensic science. Eur J Law Econ 20:255–286

Koppl R (2006) Democratic epistemics: an experiment on how to improve forensic science. Max Planck Institute of Economics, Jena, Papers on Economics and Evolution. No.0609

Koppl R (2010) Organization economics explains many forensic errors. J Inst Econ 6:71–81

Lawless C (2011) Policing markets: the contested shaping of neoliberal forensic science. Br J Criminol 51:671–689

Lawless C, Williams R (2010) Helping with inquiries, or helping with profits? The trials and tribulations of a technology of forensic reasoning. Soc Stud Sci 40:731–755

McFarland R (2003) Review of the forensic science service. HMSO, London

Mennell J (2006) The future of forensic and crime scene science Part 2: A UK perspective on forensic science education. Forensic Science International 157S

National Academy of Sciences (2009) Strengthening forensic science in the United States: a path forward. National Academies Press, Washington, DC

National Audit Office (1998) The forensic science service. HMSO, London

Pavitt K, Patel P (1999) Global corporations and national systems of innovation: who dominates whom? In: Archibugi D, Howells J, Michie J (eds) Innovation policy in a global economy. Cambridge University Press, Cambridge, p 94

Roberts P (1996) What price a free market in forensic science services? The organization and regulation of science in the criminal process. Br J Criminol 36:37–60

Touche Ross (1987) Review of scientific support for the police. The Home Office, London

York CA (2011) ANZFSS symposium. Aust J Forensic Sci 43(2–3):85–88

Origins of Modern Policing

▶ Democratic Policing

Outdoor Serious Violence: The Role of Place

Lucia Summers
School of Criminal Justice, Texas State
University-San Marcos, San Marcos, TX, USA

Overview

Like many other crime types, outdoor serious violence is not randomly distributed in space. This article critically considers why this may be the case from the perspective of environmental criminology. Special consideration is given to routine activity theory, which argues that for a crime to occur three elements need to coincide in space and time: a motivated offender, a suitable target, and the absence of a capable guardian. In the case of outdoor serious violence, both targets and guardians are often mobile. Therefore, to understand the geographic distribution of this type of crime, the movement patterns of offenders, victims, and guardians all need to be considered. But there are different types of outdoor serious violence, each with its own "optimal" place and set of circumstances. For example, predatory violence (e.g., sexual- or robbery-related) tends to take place in secluded locations such as quiet streets, parking lots, and parks. In contrast, unplanned or expressive attacks, such as those resulting from alcohol-related altercations, are more common in crowded spaces, often near entertainment venues. A third type is gang-related violence, which also has a strong geographical element due to the territorial nature of gang activity. While considering these scenarios from a routine activity perspective, a number of theoretical and methodological issues arise, such as the role and significance of guardianship and the difficulties associated with measuring the ambient population. The article concludes with the implications these issues may have for research and policing.

Introduction

This entry examines the influence of "place" on the geographical distribution of outdoor serious violence (hereafter, OSV) by considering two separate but interrelated questions. The first considers whether OSV concentrates spatially, either through the formation of clusters (i.e., hot spots) or through common characteristics across disperse crime locations (i.e., the offenses are not clustered geographically but the types of places where they occur are similar in some way). The second question considers the level of specialization of OSV locations; in other words, does OSV occur in places that are different from where other types of crime – or no crime – occur?

The entry is structured as follows. The first section sets the context by providing a working definition of OSV, a rationale for its study, and an overview of the theoretical framework employed. Three main types of OSV are identified and these are considered separately in the second section, which summarizes the state of the art. The third part of the essay highlights areas of controversy and suggests further research.

Background Description

For the purposes of this entry, OSV is defined as the unlawful killing, or near killing, of another person which takes place in an open public space (usually a street). As such, it incorporates unlawful homicides and "near misses" where an individual sustains near-lethal injuries (e.g., attempted murder, aggravated assault).

Although researchers often provide summaries of the type of location where homicides occur, it is not common for the type of offense location (such as "outdoor") to be used as a classification variable per se. Instead, homicides tend to be defined in terms of motive (e.g., sex murder), the offender-victim relationship (e.g., stranger homicide), or the victim characteristics (e.g., child homicide). While these variables will no doubt be informative, knowledge about the immediate social and physical setting should also be considered when generating typologies, particularly if the focus is on understanding the crime event itself – as opposed to, for example, an individual's propensity to engage in criminal activity.

This is precisely the focus of environmental criminology. Within this approach, it is not just the offender's predisposition to engage in criminal activity that is considered but also the roles of victims and bystanders, as well as a range of other situational variables that may influence the likelihood of a crime occurring. Theories of environmental criminology include routine activity theory (Cohen and Felson 1979), which argues that for a crime to occur three elements need to coincide in space and time: a motivated offender, a suitable target, and the absence of a capable guardian.

Open public spaces are particularly interesting from a routine activity perspective. As the definition implies, they are usually accessible to anyone at any time, which means potential victims and offenders can come together without any access restrictions. Because individuals tend to move through these spaces or remain there only for small time periods, the nature of the environment is in a constant state of flux. Public spaces are also seen as singular in that they belong to everyone and no one at the same time, which results in space users being less inclined to regulate them (e.g., than they would be to regulate their own yard or store; Newman 1972). As Campbell (1986: 120–121) put it,

> There is no starting- and finishing-time for street life. It is an ongoing party that changes its character constantly as individuals join and leave in the course of the day. The streets are above all public places. Unlike in someone's home, those present are not beholden to their host, nor are they required to constrain their behaviour out of respect for someone else's property or territorial rights. The street therefore represents an area that is "up for grabs" in terms of control and territorial rights.

In this entry, environmental criminology and, more specifically, routine activity theory provide the theoretical framework for discussing why OSV may occur in some places (under certain circumstances) but not in others. The next section summarizes what we already know in this respect.

State of the Art

When considering whether OSV happens in some places but not in others, one question should be whether OSV clusters geographically. While the identification of geographical clusters for homicide and other forms of serious violence is neither uncommon nor new (e.g., Bullock 1955), only a very small number of authors have reported disaggregated figures based on the type of place where the crime occurred. One such author is Per-Olof Wikström, who uses a contextual classification of violence – originally developed by McClintock in 1963 – based on both victim-offender relationship and crime location type.

In a study published in 1994 (cited in Wikström 1995), he reported that outdoor homicides were highly concentrated in space and that these hot spots were fairly stable over a four-decade period (1951–1991). These outdoor homicide hot spots highly overlapped with hot spots of street violence identified in a previous study (Wikström 1991). This seems to indicate that either these two offense types are similar in nature or that they are enabled – or facilitated – by the same kind of environment(s).

Both outdoor homicides and outdoor violence were concentrated in the central business district (CBD). From a routine activity perspective, this makes sense. CBDs are places where vast amounts of people congregate, hence making it more likely that a motivated offender and a suitable target will come together in time and space. At the same time, most of these people are strangers to each other, thus creating an environment low in social order, where individuals are less inclined to act as guardians (see Wikström 1995).

A Need to Disaggregate

At a more micro level, however, different types of OSV are likely to be associated with different types of places. At least two main types are worth considering: (1) planned or predatory violence and (2) unplanned or expressive attacks. Where an offense is planned (e.g., robbery-related, sexual murders, retaliatory attacks), the offender is likely to have chosen the attack location a priori based on specific criteria, such as the area being isolated and thus lacking a capable guardian. In many cases, predatory offenders would not have planned to go out and commit an offense but rather take advantage of an opportunity as it presents itself. However, these cases as still regarded as "planned" in that thought would have been given in advance as to how to commit the crime and which circumstances and/or locations would be appropriate. Because isolated areas will not only have fewer potential witnesses but also a reduced victim pool, predatory offenders often need to identify their victims at a busier, nearby location (i.e., the encounter site) and follow them to a location where they feel comfortable assaulting the victim (i.e., the attack site). As a result, isolated street segments in close proximity of busy areas may be particularly risky locations (Rossmo 2000).

Unplanned violence such as dispute-related attacks, on the other hand, may occur in busy places where friction is more likely. These offenses are more expressive in nature which means offenders may not be dissuaded by the presence of potential guardians. Outdoor dispute-related violence often occurs near alcohol outlets (e.g., Nelson et al. 2001). This is especially the case if these facilities are in close proximity of each other and have similar closing times, as this results in what has been termed a "potentiation effect" (see Rossmo 1995). In such circumstances, crowds of intoxicated individuals try to navigate the street network and compete for available resources (e.g., taxis, fast food), often leading to disorder, crime, and violence. Additional "crime generators" for outdoor serious violence, albeit to a lesser extent, include retail stores (Nelson et al. 2001) and neighborhood parks (Groff and McCord 2011).

Offender Awareness Spaces

Whatever their motives and levels of planning, offenders "consistently commit crime in neighbourhoods they personally know well or that are very similar in physical, social and economic characteristics to their home neighbourhoods" (Brantingham and Brantingham 1995: 13). In predatory violence, these are areas where they feel comfortable as familiarity provides them with a degree of control (e.g., they know how frequented the street is, what escape routes are available). Predatory offenders are familiar with these areas because they have come to know them while engaging in noncriminal activities or, less commonly, through targeted spatial exploration (i.e., they have familiarized themselves with these areas with the intention of committing crimes there). In unplanned violence, the crime usually occurs *while* the offender is engaged in everyday routine activities (Rossmo 2000).

Because the behavior of offenders – just like everyone else's – is influenced by the least effort principle (Zipf 1949), the areas offenders frequent during the course of their routine activities tend to be close to their residence (or another significant anchor point such as work or entertainment). This translates into crime sites being close to the offender's home, a finding that has been used to explain spatial distributions of crime and to prioritize suspects during crime investigations (i.e., geographic profiling; Rossmo 2000).

When targets are stationary, examining the distances between the crime sites and the offender's home (referred to as journey-to-crime; JTC, hereafter) seems appropriate. But when targets are mobile, a greater understanding can be achieved by also considering the location of the victim's home, as this may be different from that of the crime site. Tita and Griffiths (2005) have devised a mobility-based spatial typology of homicide, based on how these three locations (i.e., the crime site, the offender's home, and the victim's home) relate to one another. Five separate types of homicide can be identified in this way, as follows:

1. internal, where all three locations lie within the same census tract;
2. predatory, where the crime site and the victim's home are in the same tract while the offender lives elsewhere;
3. intrusion, where the crime site and the offender's home are in the same tract while the victim lives elsewhere;
4. offense mobility, where the offender and the victim live in the same tract but the crime occurs elsewhere; and
5. total mobility, where each location is located in a separate census tract.

Outdoor homicides are less likely to be "internal" or "predatory" and, more likely, to be "intrusions" or have "total mobility," which led the authors to conclude that "events occurring outside are predicated primarily on the mobility of the victim" (Tita and Griffiths 2005: 303).

Gang-Related Violence: A Special Case

Although OSV may be loosely divided into planned (predatory) and unplanned (expressive, altercation) attacks, gang-related violence is worth special attention. This is because gang-related violence is often turf-related in nature, and, as such, it has a strong spatial component. Gangs tend to claim territories within which they engage in drug-selling and other criminal activities or simply hang out. Although the territories claimed can be as big as a whole neighborhood, gang members tend to congregate in smaller areas defined by Tita et al. (2005: 280) as the gang's "set space":

> . . .a gang may "claim" an entire neighborhood as its domain or "turf," but set space is the actual area within the neighborhood where gang members come together as a gang. Thus, just as sets are part of a larger gang, set space is a subset of a larger gang turf or territory.

Set spaces are usually located in open public spaces, leading to most incidents of gang-related violence taking place outdoors. Where rival gangs occupy territories that are in close proximity of each other, this can lead to an increased prevalence of violence. As described by Valdez et al. (2009: 303) in their study of Mexican American street gangs,

> Geographic proximity of gang territories emerged as an important characteristic in explaining these gang homicides. The fact that most of the homicides were between members of gangs bordering the same neighborhood or area may provide an understanding of the escalation of violent acts between these groups of youth in these neighborhoods. The spatial proximity of two gangs results in frequent contact and high visibility that create a volatile environment susceptible to aggression and violence. These findings are similar to the findings of others that gang homicide was more often turf related than drug related.

George Tita and colleagues have carried out extensive work on how the geographic distribution of gang-related homicide – and violence in general – can be explained in relation to the location of the "set space" of different gangs in relation to one another (e.g., rivals vs. neutral vs. friendly; see Tita and Radil 2011 for a recent example).

Gang-related attacks may be planned or unplanned. Planned attacks are usually retaliatory (as part of ongoing violence between rival gangs or as a punishment for not showing "respect") or used as a means of solving disputes arising during the course of illegal activities such as drug-selling. Unplanned or spontaneous attacks tend to occur in unclaimed territories or within gang turf if infiltrated by a rival gang member. An example of the latter scenario is described by Valdez et al. (2009: 295–6):

> Me and RadioMan were walking over by the wall and didn't notice we went into their neighborhood. It was around Christmas time. At first I thought they were firecrackers, but then I saw they were shooting at us from the alley. I got shot in the stomach. When I turned around, RadioMan got shot too. All I remember was seeing him lying there, and he died in the ambulance.

Routine Activity Theory and Outdoor Serious Violence

Routine activity theory provides a useful framework for explaining the spatial distribution of OSV. The routine activities of both offenders and victims dictate where crime occurs, although

Outdoor Serious Violence: The Role of Place, Table 1 Variables influencing the activity spaces of motivated offenders, suitable targets, and capable guardians in relation to predatory and expressive outdoor serious violence

	Offender	Victim	Guardian
Predatory (planned)	Routine activities	Routine activities	Routine activities
	Active exploration		Formal guardianship
Expressive (unplanned)	Routine activities	Routine activities	Limited influence

some predatory offenders may also purposively familiarize themselves with an area not normally frequented with a view to committing crimes there (i.e., active spatial exploration; see Table 1). Guardianship may play a lesser role in OSV as compared to violent incidents taking place in other locations. This is because, as argued earlier, the public nature of the outdoor environment can make bystanders less willing to intervene, particularly when exposed to expressive violence (Felson 1993).

Most predatory violent offenders look for a particular type of victim (as opposed to a particular individual; Rossmo 2000); hence, the location of predatory attacks is mainly determined by the offender's routine activities and the absence of a capable guardian which may be a bystander or some form of formal guardianship (e.g., CCTV camera). Where the victim has been predetermined a priori, the victim's routine activities will be the most decisive factor.

Expressive attacks occurring outdoors tend to be either alcohol- or gang-related. As described earlier, the former usually take place in close proximity of alcohol outlets and are overrepresented in downtown areas. Gang-related OSV, on the other hand, is often reported in gang turf or in unclaimed areas nearby.

Controversies and Open Questions

Despite extensive research on the spatial concentration and specialization of serious violence, consensus is still to be reached in relation to some issues. These issues, as well as the impact of recently emerging techniques, are considered in this section.

Should Homicide Be Considered a Separate Offense or Is It Just an Outcome?

As Gottfredson and Hirschi (1990: 34) put it, "the difference between homicide and assault may simply be the intervention of a bystander, the accuracy of a gun, the weight of a frying pan, the speed of an ambulance or the availability of a trauma centre." In a similar vein, Block (1977: 10) argues that "most killings are the outcome of either an aggravated assault or a robbery which somehow progressed beyond the degree of harm intended by the offender" and that this should prompt us to regard lethal and near-lethal attacks as similar in nature. In this way, homicide and near-lethal attacks are seen as a single form of crime albeit with outcomes varying in severity and which can be positioned along a continuum.

A contrasting view is offered by Felson and Messner (1996: 520), who state that "a substantial proportion of homicide offenders really do intend to kill their victims and not merely injure them" and claim this is an indication that homicides and near-lethal violence are different entities. The fact of the matter is that intent to kill is difficult – if not impossible – to measure and that, when an assumption is made as to whether intent to kill is present, this is imperfectly associated with the outcome of the offense (i.e., in manslaughter, there is no intent to kill but a fatal outcome, while attempted murder results in a nonfatal outcome despite the intent being lethal).

Although most authors tend to favor the former position, most homicide research still considers just homicide data, as opposed to, for example, also including attempted murder and aggravated assault. A number of studies have included both in an attempt to identify risk factors associated with lethality, which has resulted in some interesting findings (e.g., Weaver et al. 2004). The question to consider here is if lethal and near-lethal violent acts are essentially of the same nature, should future research aim to include both in their analyses? This approach might be advantageous in that it may provide a fuller picture, while also increasing sample sizes, which can be particularly helpful when studying low incidence crimes such as homicide.

Does Alcohol-Related Homicide Really Cluster Around Alcohol Outlets?

While most research findings point to an association between the spatial distribution of alcohol-related homicide (and violence in general) and that for alcohol outlets, contradictory results have also been reported. For instance, Block and Block (1995) described how hot spots for "liquor-involved" homicide did not overlap those for alcohol outlets (i.e., taverns and liquor stores) or even hot spots for crime within such premises.

But Block and Block – like many other authors – included in their analysis all instances of homicide where alcohol was thought to have played a role, regardless of where the incident took place. Including residence-based offenses in the analysis has the potential of obscuring any associations that may exist. The situational context in such cases is markedly different, which means different crime facilitators may be at play: while the disinhibiting effects of alcohol may apply in all cases, the environment within and around taverns is thought to present additional risk factors (e.g., crowds, competition for limited resources) which are tied – geographically – to the taverns themselves.

Michael Townsley and colleagues at Griffiths University in Australia have recently carried out an observational study into violence around taverns. What is interesting about this research is that it specifically examines crowding and waiting in line as precipitating factors (see Wortley 1998). One of the findings is that it is not so much overcrowding that leads to aggression, but how pedestrians' direction of travel (as influenced by the street network and surrounding land use) increases the likelihood of individuals running into each other (Townsley and Grimshaw 2013).

Studies like this where the microenvironments of violence are examined are essential in helping us understand when a potential offender may be motivated enough to physically attack another individual. As stated earlier, guardianship is thought to have a limited role in such cases, and the roles of victims and offenders are often interchangeable. This means that research may be more informative – from a preventative viewpoint – if it focuses on the circumstances that increase an individual's propensity to violence and his/her ability to inflict serious harm.

Interpreting Journey-to-Crime Distances

As discussed, offenders tend to commit (violent) crime in familiar areas close to their home. This has been reported to be particularly the case for violent offenses, as compared to property crime. But these figures should be interpreted with caution, as analyses often involve offenses that are committed in the offender's own home. When these cases are excluded, the mean (or median) distances traveled are likely to be on a par with those for property crime.

To illustrate, the JTC mean distance for the 363 homicides in Pizarro et al. (2007) study was 1.8 miles; if the homicides taking place at the offender's home are excluded (67 in total), the mean JTC increases to 2.2 miles. Unsurprisingly, the biggest differences are in domestic homicide: when considering all domestic cases (71), the mean JTC is 0.6 miles; if zero JTC distances are excluded, the mean JTC increases to 2.9 miles.

Of course it is important to determine which proportion of offenses takes place within the offender's home, but these should be examined separately when providing summary statistics.

Assessing Victimization Risk

Dangerous places are not necessarily those where many incidents of outdoor serious violence occur. This is because dangerousness relates to the odds of being victimized and, as such, it is mediated by a denominator that represents the population at risk (i.e., the number of potential victims). The number of residents is the most commonly used denominator when calculating violent crime rates. This may be appropriate for some crime types, such as domestic violence, which tend to occur in the home. For crimes that occur outdoors, however, a better denominator would be the number of people on the street, often referred to as the ambient population. This is particularly important when calculating rates for central business districts, which tend to have busy streets but small resident populations.

While resident population statistics can usually be obtained from national census polls, ambient population figures are not readily available, meaning researchers are forced to generate these data themselves. This is usually done through observation, where a fieldworker stations him/herself at a "gate" (i.e., a specific point within the street network) and counts the number of pedestrians (or vehicles) crossing such a gate over a particular period of time. Data may be collected at the same gate at various times of the day, week, or year, to determine whether the ambient population varies accordingly.

For instance, Chainey and Desyllas (2006) used a pedestrian model to calculate victimization rates of street robbery and "snatch" theft in central London, England. The model was calibrated with relevant key variables (i.e., land use patterns, availability of public transport, sidewalk capacity, and straight-line visibility for easier navigation) and observation data from 231 block segments in 107 different locations. The application of the model to a new area for testing revealed a strong association between estimated and actual pedestrian counts ($R^2 = 0.74$).When the spatial distributions of the two crime types were examined, it emerged that the maps displaying the victimization rates based on estimated pedestrian counts bore little resemblance to those where the resident population was used as a denominator instead. Both types of rates led to distributions that were also different from simple crime density hot spot maps.

But studies of this nature are rare. Collecting data to measure the ambient population can be incredibly resource intensive, especially for larger areas. This has led a number of authors to question whether the changes potentially observed in the new crime rates (i.e., as compared to the rates where the resident population is used as the denominator) are sizeable enough to justify the cost (e.g., Cohen et al. 1985).

In recent years, however, at least two different methods of estimating the ambient population have become available, namely, LandScan and Space Syntax. LandScan is a geographic data set developed by the Oak Ridge National Laboratory which provides ambient population estimates worldwide (see Dobson et al. 2000). These estimates are calculated on the basis of census information, physical geography data, lighting intensity, and other spatial data sets. Despite the level of resolution (approximate one-square-kilometer or 30-arcsecond square cells) still not being fine enough for microlevel analyses, LandScan has already been demonstrated to improve our understanding of the spatial distribution of homicide and other violent crime (e.g., Andresen 2011). At the time of writing, LandScan data at an improved 90m resolution (3-arcsecond) were already available for the US and select OCONUS areas, with plans to extend this further.

A second methodological advance in this area comes from the Space Syntax literature (Hillier and Hanson 1984). Space Syntax provides a means for systematically measuring the configuration of street networks. It does this by representing the street network as an inverted graph where the nodes are the block segments and the links (between them) denote whether the nodes intersect. Using this inverted graph and formulae adapted from graph theory, a number of measures are generated for each segment. Examples of these measures are integration, which indicates how accessible a block segment is within a particular area, and choice, which is the number of times a segment is passed through when generating all shortest paths. These measures have been shown to correlate with pedestrian flows and vehicular flows (Hillier and Iida 2005).

Unlike LandScan, Space Syntax is not normally used to generate pedestrian counts that can then be used as denominators in the calculation of victimization rates (although there is no reason why this should be avoided). Instead, Space Syntax studies tend to examine the relationship between integration (and other graph-derived measures) and crime levels. These tend to indicate that it is more segregated (isolated) segments that are associated with higher crime counts. This applies to residential burglary, street robbery (e.g., Hillier and Sahbaz 2008), and lower level violence (Friedrich et al. 2009). In contrast, Summers and Johnson (2011) found outdoor serious violence – including homicide – to be associated with more integrated (busier) segments.

A problem with Space Syntax is that the proxy measures it generates are atemporal. Values are assigned purely based on the structure of the street network; because this is constant over time, so too are the measures generated. But pedestrian and vehicular flows are known to vary dependent on the time of the day, week, or year. Efforts have been made to address this issue by disaggregating crime data by time of offense (e.g., Hillier and Sahbaz 2008), but more work needs to be done, especially in relation to incorporating land use data into the analyses.

In regard to LandScan, daytime and nighttime ambient population data have been made available for the USA (see LandScan USA). It is expected that these estimates will be further disaggregated in the near future and extended to other geographical areas.

A developing field awash with contradictory findings, combined with a lack of research in relation to OSV, calls for more studies to be performed in this area before a full understanding of the spatial distribution of this crime type can be achieved. Research incorporating various measures of the ambient population, different crime types, and complementary data sets (e.g., land use, census) would be particularly helpful in clarifying inconsistencies.

Situating the Crime Event Within the Neighborhood Context

To achieve a better understanding of the context of the crime event, it is important to establish not just how many people are around but also what these people are like. In recent years, researchers have used multilevel analyses and spatial econometric statistical models to examine the joint effect of microlevel variables (e.g., number of bystanders) and neighborhood characteristics (e.g., deprivation) on the spatial distribution of crime. For instance, Bernasco and Block (2009) employed such an approach with robbery data for the city of Chicago to examine what affected offender target choices. Their results suggested that offenders avoided neighborhoods high in collective efficacy and those where the predominant racial or ethnic group was different from theirs. Collective efficacy is defined as "social cohesion among neighbors combined with their willingness to intervene on behalf of the common good" (Sampson et al. 1997: 918). Proximity measures for a lack of collective efficacy include ethnic heterogeneity and high levels of residential mobility.

Most studies of this kind have focused on residential burglary and other property crime. However, recent research by Summers and Johnson using homicide and attempted murder data for London, England, seems to indicate that neighborhoods of similar racial profiles are also favored by offenders engaging in OSV (Summers and Johnson 2011). Another factor found to be associated with higher crime rates was proximity to the offender's home area. Though promising, more research needs to be carried out in this area to advance understanding of how the interaction between offender characteristics and micro- and meso-level situational variables influence the spatial distribution of OSV.

Concluding Thoughts

The aim of this entry was to provide an overview of what we already know about OSV and to highlight the areas within this field where controversy and gaps in knowledge remain. As previously stated, our understanding of this phenomenon is severely restricted by the fact that research findings are seldom disaggregated by location type. Outdoor crime events are qualitatively different in two main ways. First, because individuals tend to move through these spaces or only remain there for limited time periods, the environment is highly fluid. The second difference between outdoor and indoor violence is that guardianship is severely impaired in the former, as its public nature makes bystanders less willing to intervene.

Still, the presence of bystanders is important for some types of OSV, namely predatory attacks. Offenders committing such offenses are likely to be deterred by potential guardians. As a result, the placement of these offenses will be different to that of other forms of OSV, as predatory attacks will tend to take place in more isolated areas (or at times of the day where it is more isolated, e.g., late at night).

An understanding of where and under what circumstances OSV might occur is important from a preventative perspective. Police in many cities are aware of the issues leading to violence in downtown areas and have in some cases implemented interventions that are informed by such knowledge, including taxi marshals and staggered closing times for taverns and nightclubs. More recent research (e.g., Townsley and Grimshaw 2013) aims to improve understanding and lead to further preventative responses, such as modifying the street layout and land use to minimize pedestrians running into each other. Crime prevention efforts by police – and other agencies – in outdoor settings are likely to be markedly different from what may be required in other environments, which again calls for researchers to treat such offenses as a separate entity.

To sum up, there is much we already know in relation to homicide and other forms of serious violence, but it is important to separate events taking place in different settings. This will aid our understanding of the drivers within such localized environments, which will in turn help us to develop more fitting preventative strategies.

Related Entries

- ▶ Agent-Based Models to Predict Crime at Places
- ▶ Applied Geographical Profiling
- ▶ Crime Location Choice
- ▶ Crime Mapping
- ▶ Forecasting of Shootings Using Risk Terrain Modeling
- ▶ History of Geographic Criminology Part I: Nineteenth Century
- ▶ Hot Spots and Place-Based Policing
- ▶ Longitudinal Crime Trends at Places
- ▶ Near Repeats and Crime Forecasting
- ▶ Offender Decision Making and Behavioral Economics
- ▶ Spatial Models and Network Analysis
- ▶ Spatial Perspectives on Illegal Drug Markets

Recommended Reading and References

Andresen MA (2011) The ambient population and crime analysis. Prof Geogr 63(2):193–212

Bernasco W, Block R (2009) Where offenders choose to attack: a discrete choice model of robberies in Chicago. Criminology 47:93–130

Block R (1977) Violent crime. Lexington Books, Lexington

Block RL, Block CR (1995) Space, place and crime: hot spot areas and hot places of liquor-related crime. In: Eck JE, Weisburd D (eds) Crime and place, vol 5, Crime prevention studies. Criminal Justice Press, Monsey, pp 145–183

Brantingham PL, Brantingham PJ (1995) Criminality of place: crime generators and crime attractors. Eur J Criml Policy Res 3(3):1–26

Bullock HA (1955) Urban homicide in theory and fact. J Crim Law Criminol Police Sci 45(5):565–575

Campbell A (1986) The streets and violence. In: Campbell A, Gibbs JJ (eds) Violent transactions. Blackwell, Oxford, pp 115–132

Chainey S, Desyllas J (2006) Modelling pedestrian movement to measure on-street crime risk. In: Liu L, Eck J (eds) Artificial crime analysis systems: using computer simulations and geographic information systems. Idea Group Inc., Hershey

Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44(4):588–608

Cohen LE, Kaufman RL, Gottfredson MR (1985) Risk-based crime statistics: a forecasting comparison for comparison for burglary and auto theft. J Crim Justice 13(5):445–457

Dobson JE, Bright EA, Coleman PR, Durfee RC, Worley BA (2000) LandScan: a global population database for estimating populations at risk. Photogramm Eng Remote Sens 66(7):849–857

Felson R (1993) Predatory and dispute-related violence: a social interactionist approach. In: Clarke RV, Felson M (eds) Routine activity and rational choice, vol 5, Advances in criminological theory. Transaction Press, New Brunswick, pp 103–126

Felson RB, Messner SF (1996) To kill or not to kill? Lethal outcomes in injurious attacks. Criminology 34(4):519–545

Friedrich E, Hillier B, Chiaradia A (2009) Anti-social behavior and urban configuration: using space syntax to understand spatial patterns of socio-environmental disorder. In: Koch D, Marcus L, Steen J (eds) Proceedings of the 7th international space syntax symposium. KTH, Stockholm

Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Groff E, McCord ES (2011) The role of neighborhood parks as crime generators. Secur J 25(1):1–24

Hillier B, Hanson J (1984) The social logic of space. Cambridge University Press, Cambridge

Hillier B, Iida S (2005) Network and psychological effects in urban movement. In: Cohn AG, Mark DM (eds) Proceedings of spatial information theory: international conference, COSIT 2005, Ellicottsville, 14–18 Sept 2005. Springer, Berlin

Hillier B, Sahbaz O (2008) An evidence based approach to crime and urban design. In: Cooper R, Boyko C, Evans G, Adams M (eds) Urban sustainability for the 24 hour city. Routledge, London

McClintock FH (1963) Crimes of violence. Macmillan, London

Nelson AL, Bromley RDF, Thomas CJ (2001) Identifying micro-spatial and temporal patterns of violent crime and disorder in the British city centre. Appl Geogr 21:249–274

Newman O (1972) Defensible space. Macmillan, New York

Pizarro JM, Corsaro N, Yu SV (2007) Journey to crime and victimization: An application of routine activities theory and environmental criminology to homicide. Vict Offenders 2:375–394

Rossmo DK (1995) Strategic crime patterning: problem-oriented policing and displacement. In: Block CR, Dabdoub M, Fregly S (eds) Crime analysis through computer mapping. Police Executive Research Forum, Washington, DC, pp 1–14

Rossmo DK (2000) Geographic profiling. CRC Press, Boca Raton

Sampson RJ, Raudenbush SW, Earls F (1997) Neighborhoods and violent crime: a multilevel study of collective efficacy. Science 277(15):918–924

Summers L, Johnson SD (2011) The influence of the street network on the spatial distribution of outdoor serious violence. Paper presented at the 20th international symposium of environmental criminology and crime analysis (ECCA), Durban, 18–21 July 2011

Tita GE, Griffiths E (2005) Traveling to violence: the case for a mobility-based spatial typology of homicide. J Res Crime and Delinq 42:275–308

Tita GE, Radil SM (2011) Spatializing the social networks of gangs to explore patterns of violence. J Quant Criminol 27(4):521–545

Tita GE, Cohen J, Engberg J (2005) An ecological study of the location of gang “set space”. Soc Probl 52(2):272–299

Townsley M, Grimshaw R (2013) The consequences of queueing: Crowding, situational features and aggression in entertainment precincts. Crime Prev and Community Safety 15(1):23–47

Valdez A, Cepedea A, Kaplan C (2009) Homicidal events among Mexican American street gangs: a situational analysis. Homicide Stud 13(3):288–306

Weaver GS, Wittekind JEC, Huff-Corzine L, Corzine J, Petee TA, Jarvis JP (2004) Violent encounters: a criminal event analysis of lethal and nonlethal outcomes. J Contemp Crim Justice 20:348–368

Wikström P-OH (1991) Urban crime, criminals and victims: the Swedish experience in an Anglo-American comparative perspective. Springer, New York

Wikström P-OH (1995) Preventing city-center street crimes. Crime Justice 19:429–468

Wortley R (1998) A two-stage model of situational crime prevention. Stud Crime Crime Prev 7:173–188

Zipf GK (1949) Human behavior and the principle of least effort. Addison-Wesley Press, Oxford

Outlaw Motorcycle Clubs/Gangs

► Motorcycle Clubs or Gangs?

Overturning Wrongful Convictions and Compensating Exonerees

Myrna Raeder
Southwestern Law School,
Los Angeles, CA, USA

Overview

DNA has provided incontrovertible proof that individuals can be convicted for crimes they did not commit. The DNA exonerations now stand at 302, with 18 released from death row. Since DNA evidence is estimated to be present in only a small fraction of cases, at best 10–20 %, it is likely that the major systemic issues that are responsible for those wrongful convictions (mistaken eyewitness testimony, faulty forensic evidence, false confessions, lying informants, government misconduct, and ineffective defense counsel) also infect at least some of the cases for which proof of innocence is confined to more traditional sources such as confessions by actual perpetrators, recantation of witnesses, and newly discovered evidence. For example, one study found that 196 non-DNA exonerations occurred from 1989 to 2003 (Gross et al. 2006), and the Death Penalty Information Center's website indicates that since 1973 a total of 140 people sentenced to death have been released from death row with evidence of their innocence (Innocence and the Death Penalty, accessed February 5, 2012, http://www.deathpenaltyinfo.org/innocence-and-death-penalty). The most recent report in June 2012 of all exonerations based on innocence since 1989 identified the number as 891 (Gross and Shaffer 2012). However, this excludes an estimated 1170 defendants whose convictions were dismissed in 13 "group exonerations" resulting from major police scandals. For example, the police scandals in such places as Los Angeles (California) and Tulia (Texas) produced more than 200 exonerations, with the vast majority of overturned convictions resulting from guilty pleas, rather than trials. Wrongful convictions cast doubt on the reliability and fairness of the criminal justice system and expose public safety failures because perpetrators, who include serial rapists and murderers, remain at large to pursue new victims. Thus, all of us, not just wrongfully convicted defendants, are harmed by these systemic breakdowns. Yet, because the criminal justice system depends on finality of judgment, it was exceedingly rare to obtain postconviction relief on the grounds of innocence prior to the advent of DNA testing.

While statutes now afford more opportunity to challenge convictions based on DNA testing, old attitudes die hard, and Chief Justice Roberts pointed out in *District Attorney's Office for the Third Judicial District v. Osborne*, 129 S. Ct. 2308 (2009), that the mere existence of new technology cannot mean that no case is final. Moreover, for cases without biological evidence, the ability to obtain relief has changed little. This entry will present an overview of postconviction remedies involving claims of innocence (Raeder 2009) and the requirements that must be satisfied to receive compensation for being wrongfully incarcerated (Raeder 2008).

Postconviction Requests for DNA Testing to Establish Innocence

Appellate remedies in the American criminal justice system are designed to ensure that defendants receive a fair trial, not to second-guess jury verdicts in the absence of insufficient evidence supporting a conviction. Finality concerns have long dictated very short time windows for new trial motions. Such statutes of limitations often expire less than 1 year from conviction, standing in stark contrast to 13.5 years, which is the average length of time served by DNA exonerees. With the advent of forensic DNA testing came the recognition of the inequity of a criminal justice system that had no apparent remedy to free individuals who were clearly factually innocent. To remedy this problem, the Innocence Protection Act (IPA), 18 U.S.C. § 3600, was passed in 2004 to permit DNA testing for federal inmates who assert their innocence under penalty of perjury and meet a number of detailed criteria, including:

- The evidence was not previously tested, and the applicant did not knowingly or voluntarily waive the right to request testing.
- The evidence was tested, but current methods are substantially more probative.
- The evidence to be tested is in the possession of the government, has been subject to a chain of custody, and retained under conditions sufficient to ensure that it has not been substituted, contaminated, tampered with, replaced, or altered.
- The evidence supports a theory that is not inconsistent with any affirmative defense at trial.
- The proposed testing would provide new material evidence that would raise a reasonable probability that the applicant did not commit the offense.
- Identity was an issue at trial.
- The applicant has provided a DNA sample for purposes of comparison with profiles in the FBI database.
- The request is timely, with detailed criteria for determining rebuttal presumptions of timeliness and untimeliness.

The court can appoint counsel for the applicant, issue a preservation order to the government, and direct the government to pay the cost of testing for indigent applicants. Inculpatory results have a number of potentially negative consequences for the applicant who may be held in contempt, receive a sentence of 3 years consecutive to the current sentence if convicted of making false assertions in a proceeding, be required to pay for the testing, lose good time, be denied parole, and if relevant have the results sent to state officials.

In contrast, exculpatory results excuse any time bar that would otherwise preclude a new trial or resentencing motion. The new trial motion is granted if the DNA test results, considered with all other evidence in the case (regardless of whether such evidence was introduced at trial), establish by compelling evidence that a new trial would result in an acquittal. The statute also indicates it is not an exclusive remedy and does not count as a habeas corpus motion for purposes of evaluating the existence of successive motions. A writ of habeas corpus is the typical way that prisoners challenge the lawfulness of their detention when they have exhausted the right to directly appeal from their conviction. Beyond being a model for states, the IPA also actively encouraged states to follow its dictates by providing DNA grants to law enforcement in states that adopt "comparable" postconviction DNA testing statutes (118 Stat. 2285).

By 2011, 48 states had adopted such statutes (Brooks and Simpson 2011). However, some states have enacted restrictions that hinder applicants from obtaining testing, such as prohibiting applications by individuals who plead guilty, who admit guilt to obtain parole, or whose attorneys did not request testing; limiting the crimes for which relief can be sought; applying only to individuals sentenced to death; requiring applicants to establish a "likelihood" rather than a "possibility" that the testing will be exculpatory, or clear and convincing evidence that the new results would be significantly more discriminating than the results of previous testing; and failing to provide adequate safeguards to preserve biological evidence. A few states even retain a statute of limitations in their DNA testing statutes.

Even the federal statute is limited to cases in which identification was an issue at trial and contains chain-of-custody requirements that if interpreted literally may be virtually impossible to meet. Judge Higginbotham reversed a district court's denial of testing that was based on a claim of insufficient chain of custody, explaining "we cannot place upon the defendant the burden of proving history while it is held in government custody. To do so would create an entrance gate so difficult to enter as to frustrate the core objective of the statute" (*United States v. Fasano*, 577F.3d 572, 577 (5th Cir. 2009)). In addition, Garrett (2008a, b) found that 20 % of exonerees did not obtain DNA testing because of the failure of defense counsel to request it. Although this failure could bar an applicant from obtaining testing, it may not meet the test established in *Strickland v. Washington*, 466 U.S. 668 (1984), for incompetent counsel, if a strategic reason for the failure can be articulated.

In *District Attorney's Office for the Third Judicial District v. Osborne*, the Supreme Court refused to recognize a constitutional right to obtain postconviction DNA. *Osborne* recognized that "DNA testing can provide powerful new evidence unlike anything known before" (129S. Ct. at 2916) and noted the dual role of DNA in exonerating wrongly convicted people while confirming the convictions of "many others" (Garrett 2008a, b). However, the court could not overcome its apprehension about the effect of DNA on finality. The *Osborne* majority ceded DNA postconviction relief to state and federal legislators, claiming for the most part that these legislators had already enacted statutes with varying requirements to provide relief. In addition to finding no procedural due process claim, *Osborne* also rejected an invitation to create an "untethered" substantive due process right to DNA evidence for purposes of testing. Both finality concerns and a "reluctan[ce] to enlist the Federal Judiciary in creating a new constitutional code of rules for handling DNA" (Garrett 2008a, b) doomed the plea for a specific right to obtain DNA testing.

The court cited the prompt response by states to the "challenge" to the criminal justice system posed by DNA technology as reason to reject the court's preempting the field and forestalling public debate about appropriate remedies.

While *Osborne* recognized that DNA has confirmed that the system "cannot be perfect" (Garrett 2008a, b at 2323), it was not ready to craft a judicial remedy for the systemic flaws exposed by DNA. More recently, in *Skinner v. Switzer*, 131S. Ct. 1289 (2011), the court held that a convicted state prisoner seeking DNA testing of crime-scene evidence could assert a procedural due process claim in a civil rights action under 42 U.S.C. § 1983 challenging the Texas postconviction DNA statute. Success in his suit for DNA testing would not "necessarily imply" the invalidity of his conviction since results might prove inconclusive or they might further incriminate Skinner. Therefore, the court found that he did not have to proceed by habeas corpus action, which had significant barriers to relief, though constraints imposed by the Prison Litigation Reform Act of 1995 (PLRA), 110 Stat. 1321–66, still apply.

Other than by enacting new statutory remedies, the best way to avoid the barriers to DNA testing may lie with prosecutors, not the courts, even though some prosecutors refuse to admit that they could have been responsible for convicting an innocent person (Orenstein 2011). A 2008 amendment to the ABA Model Rules of Professional Responsibility 3.8(g) provides that in cases where there is "a reasonable likelihood" that a convicted defendant did not commit the offense, prosecutors should (1) promptly disclose that evidence to an appropriate court or authority and, unless a court authorizes delay, promptly disclose that evidence to the defendant and (2) undertake such further inquiry or investigation as may be necessary to determine whether the defendant was convicted of an offense that the defendant did not commit. Rule 3.8(h) was added to apply when a prosecutor knows of clear and convincing evidence establishing that a defendant was convicted of an offense that the defendant did not commit. In that case, the prosecutor shall seek to remedy the conviction. These rules may be key in cases where testing has taken place or other evidence establishing innocence is present, such as credible recantations and/or third-party confessions. However, they leave open how prosecutors should react before testing. In this regard, Comment 1 added to 3.8(g) lends support for proactive action even in non-DNA cases by explicitly referencing the prosecutor's role as a minister of justice that carries obligations to see "that special precautions are taken to prevent and to rectify the conviction of innocent persons." At a minimum, this comment cautions against objecting to testing requests without good cause.

However, the practical difficulty with testing requests is determining whether evidence has been preserved that can lead to establishing innocence. While the existence of such evidence is particularly problematic in older cases, even new preservation statutes may be limited to evidence in only certain crimes, permit disposal of evidence prematurely, or do not have adequate penalties for unauthorized destruction of evidence. The ABA has urged legislation be drafted to generally ensure the preservation of material evidence for postconviction review (Resolution 111 F, approved August 2004), and the ABA Standards on DNA Evidence, Standard 2.6(b), provides for the preservation of DNA evidence until the convicted defendant has completed the sentence.

Obtaining Habeas Release Based on Innocence

AEDPA Restrictions

Because collateral attacks on convictions typically depend on evidence not presented at trial, they often are raised in petitions for habeas corpus. While statutory remedies concerning DNA testing and reliance on § 1983 eliminate some of the need for habeas, when relief is sought for a wrongful conviction, habeas will still be the likely remedy. In addition, new trial motions are not usually very helpful in non-DNA cases, because of their strict statutes of limitations, limited access to discovery, and high legal and evidentiary thresholds (Medwed 2005). For example, *Farrar v. People*, 208 P.3d 702 (Colo. 2009), held that a complaining witness's postconviction recantation of testimony that raises a reasonable doubt about the defendant's guilt is not enough to justify a new trial unless the court finds the recantation credible and would probably have led to an acquittal. As a result, a teenager's recantation of her uncorroborated accusations of sexual abuse by her stepfather was not enough to justify a new trial for the stepfather.

The complexities of habeas corpus practice are beyond the scope of this entry. However, timeliness and exhaustion questions often arise in the context of innocence claims. Title 28 U.S.C. § 2244(d)(1)(A) of the Antiterrorism and Effective Death Penalty Act (AEDPA) of 1996 created a 1-year limitation period for petitions for writ of habeas corpus brought pursuant to 28 U.S.C. § 2254 by state prisoners. While the time during which a properly filed application for state postconviction or other collateral review is pending does not count toward any period of limitation (28 U.S.C. § 2244(d)(2)), AEDPA effectively blocks many claims.

The law concerning an innocence exception to the timeliness requirement of AEDPA is in complete disarray, with some federal circuits recognizing the exception, some rejecting it, and a few refusing to reach the issue because the petitioner had not met the high burden of proof needed to establish the innocence claim. Most recently, *Lee v. Lampert*, 653F.3d 929 (9th Cir. 2011) (*en banc*), adopted an equitable exception to AEDPA based on a credible showing of actual innocence. To the extent that courts permit equitable tolling of AEDPA's statute of limitations, *Lawrence v. Florida*, 549 U.S. 327 (2007), provides that the petitioner must establish both due diligence and that extraordinary circumstances prevented timely filing. However, *Martinez v. Ryan*, 132S. Ct. 1309 (2012), clarified that inadequate assistance of counsel at initial-review collateral proceedings may establish cause for a prisoner's procedural default of a claim of ineffective assistance at trial and that AEDPA did not bar a prisoner from using ineffectiveness of his postconviction attorney to establish "cause" for his procedural default.

In October 2012, the Supreme Court granted certiorari in *McQuiggin v. Perkins* to determine two questions: (1) whether there is an actual innocence exception to AEDPA's requirement that a habeas petitioner must show an extraordinary circumstance that prevented timely filing from its 1-year limitation period and (2), if so, whether there is an additional actual innocence exception to the requirement that a petitioner demonstrate that "he has been pursuing his rights diligently."

While this should resolve those discrete issues, it will not end the disarray fueled by the Supreme Court's continued refusal to explicitly recognize a federal constitutional right to be released upon proof of "actual innocence." As Chief Justice Roberts admitted in *Osborne* "[w]e have struggled with it over the years, in some cases assuming, arguendo, that it exists while also noting the difficult questions such a right would pose and the high standard any claimant would have to meet.... In this case too we can assume without deciding that such a claim exists, because even if so there is no due process problem" (Garrett 2008a, b at 2321–22).

The rejection of his second habeas petition in *In re Davis*, 565F.3d 810 (11th Cir. 2009), led Troy Davis, a death row inmate, to file an original petition for habeas corpus in the Supreme Court claiming actual innocence to excuse his untimely filing. This route was taken because the Supreme Court had agreed in *Felker v. Turpin*, 518 U.S. 651(1996), that one of AEDPA's provisions prevented it from reviewing a Court of Appeals' order denying leave to file a second habeas petition. *Felker* held, however, that the Supreme Court was not deprived of appellate jurisdiction because AEDPA did not remove the court's authority to entertain an original petition for habeas corpus. As a result, in *In re Davis*, 130S. Ct. 1 (2009), the court took what Justice Scalia characterized as "the extraordinary step-one not taken in nearly 50 years-of instructing a district court to adjudicate" a claim of innocence (*id.* at *2). The court's unsigned one-paragraph decision ordered the district court to "receive testimony and make findings of fact as to whether evidence that could not have been obtained at the time of trial clearly establishes petitioner's innocence" (Garrett 2008a, b at 1). In contrast, the dissent of Justice Scalia, which is joined by Justice Thomas, would appear to allow an innocent petitioner to be put to death so long as that individual had received a fair trial.

Davis was convicted of the 1989 murder of an off-duty Savannah police officer, but claimed he was not the shooter. Seven key witnesses later recanted, and several people identified the main prosecution witness as the shooter. Davis' request for a pardon had been denied before he petitioned the Supreme Court. Ultimately, the trial court held a hearing and concluded that his evidence of innocence consisted largely of "smoke and mirrors" (*In re Davis*, No. CV 409–130, 2010 WL 3385081 *59 (S.D.Ga. Aug. 24, 2010)). The court also denied a certificate of appealability (COA) on the grounds that any appeal must be directed to the Supreme Court, a decision affirmed by the Eleventh Circuit in *Davis v. Terry*, 625F.3d 716 (11th Cir. 2010), *cert. denied sub nom*, *Davis v. Humphrey*, 131S. Ct. 1787 (2011). Finally, in *Davis v. Humphrey*, 131S. Ct. 1788 (2011), the Supreme Court denied both Davis' direct appeal and request for certiorari from the district court's decision. Thus, the Supreme Court was ultimately unwilling to reach the merits of Davis' innocence claim. After this, Davis was again denied a pardon and was executed in September of 2011.

Gateway Claims that Excuse Procedural Defaults

Generally, a procedural default may be excused in habeas if the petitioner can show that failure to address the claim on the merits would lead to a fundamental miscarriage of justice, often referred to as the "actual innocence" exception (*Wainwright v. Sykes*, 433 U.S. 72, 90 (1977)). Pre-AEDPA, in *Schlup v. Delo*, 513 U.S. 298 (1995), the court discussed these so-called gateway claims of innocence, which simply permit the federal court to hear a constitutional claim brought by a federal habeas petition that was procedurally defaulted in state court. *Schlup* held that to state a claim of actual innocence sufficient to excuse procedural default, a defendant must show that, in light of all the evidence, "it is more likely than not that no reasonable juror would have convicted him in light of the new evidence" (Garrett 2008a, b at 314). *Schlup* requires the petitioner to establish factual innocence rather than mere legal insufficiency and explicitly limited the miscarriage of justice test to "extraordinary" cases:

> To be credible, such a claim requires petitioner to support his allegations of constitutional error with new reliable evidence–whether it be exculpatory scientific evidence, trustworthy eyewitness accounts, or critical physical evidence–that was not presented at trial. Because such evidence is obviously unavailable in the vast majority of cases, claims of actual innocence are rarely successful. (Garrett 2008a, b at 324)

Pro se litigants can hardly be expected to meet this criterion, but even inmates who find lawyers willing to champion their innocence are hard-pressed to find this quality of evidence in the absence of DNA, because most other evidence comes from recantations by accomplices or witnesses, which are viewed skeptically by judges. In the 10 years after *Schlup* was decided, less than 10 % of decisions citing that case resulted in consideration of otherwise barred claims, and of those 31 cases, only 20 were resolved in favor of the petitioner. Thus, 11 individuals with enough evidence to suggest their probable innocence were denied relief (Segal 2008).

House v. Bell, 547 U.S. 518, 536–37 (2006), was the first Supreme Court case to satisfy the *Schlup* standard, which it described as follows: "Prisoners asserting innocence as a gateway to defaulted claims must establish that, in light of new evidence, it is more likely than not that no reasonable juror would have found petitioner guilty beyond a reasonable doubt." *House* also reaffirmed that the reviewing court must make a "holistic judgment" about all the evidence and its likely effect on reasonable jurors applying the reasonable-doubt standard (Garrett 2008a, b at 539). The difference between the 5–3 decision favoring House and the 5–4 decision in *Osborne* rejecting a right to DNA testing was Justice Kennedy's swing vote in both cases, which may have been influenced by the strong facts favoring innocence in *House* and the much weaker factual showing in *Osborne*. The exculpatory evidence in *House* included DNA, which confirmed the semen on the victim's underwear belonged to her husband and not to House. This undermined the state's argument at trial that House's motive to kill the victim was to mask that he had sexually assaulted her. Expert evidence also indicated that incriminatory bloodstains actually resulted from a spill from House's reference sample. Finally, new evidence indicated that the victim's husband had confessed to the crime.

Even with this evidence, three justices did not believe that House met the innocence threshold necessary to establish a gateway claim. Chief Justice Roberts' dissent faulted the majority for disregarding the district court's evaluation of discounting the reliability of House's new evidence. He also opined that it was more likely than not that at least one juror acting reasonably would vote to convict. In other words, he argued that the threshold was not simply reasonable doubt but "such compelling evidence of innocence that it becomes more likely than not that no single juror, acting reasonably, would vote to convict him" (547 U.S. at 571).

Because *House* did not find actual innocence but only permitted litigation of an otherwise barred procedural claim of ineffective assistance of counsel, the case was remanded. In other words, being more likely than not that no reasonable jury would convict him would not have saved House from execution unless his claim of constitutional procedural error was granted. Without that finding, he could not be freed, let alone granted a new trial. As difficult as House's journey to freedom proved, his petition did not raise any thorny AEDPA issues because it was his first federal filing, so not barred as a successive claim. Because *Schlup* was decided a year before the enactment of AEDPA, its relationship to AEDPA is not clear, meaning that many habeas litigants will have to surmount the previously mentioned AEDPA barriers, as well as satisfying innocence criteria. Ultimately, House was granted a conditional writ of habeas corpus that did not prohibit the state from retrying him. Although the local District Attorney had vowed he would retry House, charges were dropped shortly before the new trial date in 2009. House was originally convicted of capital murder in 1986 and was not released from prison until July 2008 after additional DNA testing and investigation further bolstered his claim of innocence, although even that decision generated a dissent. While his long road to freedom proved the system can rectify its errors, it reinforces how difficult the process is.

"Newly" Discovered Evidence and New Trial Motions

What constitutes "new" evidence for new trial motions and gateway claims is currently in dispute, with some circuits allowing evidence not previously presented, even if it technically would have been available, while others consider only evidence that could not have been discovered prior to trial through the exercise of due diligence (Nelson 2008). Only the broader rule encompasses evidence that is often relevant in cases alleging innocence, such as (1) evidence excluded because of actions of the defense counsel, whether or not they reach the level of incompetent counsel, or (2) the defendant's own testimony, which may have been absent because the defendant failed to take the stand or plead guilty. Similarly, in federal court, a majority of circuits exclude a codefendant's posttrial offer of exculpatory testimony from the definition of "newly discovered evidence" (Brennan 2008).

Another problem with the definition of newly discovered evidence is demonstrated by the developing scientific consensus about fire investigation, which suggests that there may be as many as 200 people convicted of setting fires that were actually accidents, based in part on expert testimony that would currently be subject to exclusion (Wolf 2009). Is this newly discovered evidence? And would it reach the level of a freestanding innocence claim? It may not be helpful in a gateway claim, since it is difficult to claim incompetence of counsel or prosecutorial misconduct when the field would not have excluded the evidence at the time it was offered. These questions are not academic since an influential article in *The New Yorker* argues that Cameron Todd Willingham, who was executed in Texas in 2004 after spending 12 years on death row for killing his three children by arson, was innocent (Grann 2009). Willingham always proclaimed his innocence and refused a deal that would have given him a life sentence. Expert reports that challenged the finding of arson in his case years after his 1992 trial did not prevent Willingham's execution. A 2009 report of an independent expert contracted by the Texas Forensic Science Commission rejected the conclusion the fire was arson. The only other evidence of guilt was by a jailhouse informant whose evidence was even viewed skeptically by the prosecution. The commission never reached a decision on the merits of Willingham's innocence claim after the governor replaced several of its members. More recently, *Han Tak Lee v. Glunt*, 667F.3d 397 (3d Cir. 2012), treated the advances in arson science as newly discovered evidence that could raise a due process claim of innocence in habeas based on the unreliability of expert testimony originally given in1990.

However, many courts are hesitant to permit review long after trial in the absence of DNA. In such cases, finality concerns are more significant because innocence cannot typically be demonstrated by a neutral source, there may be circumstantial evidence of guilt, and second-guessing juries could result in retrials with stale evidence and upset victims without a guarantee that justice will be done. In some states like California, the bias against newly discovered evidence is pronounced and in the habeas context does not warrant relief unless it points unerringly to innocence (Medwed 2007). Yet, since the early days of our republic when "dead" people reappeared after defendants were sentenced to death for murder, there have been several hundred exonerations without DNA (Gross and Shaffer 2012; Gross et al. 2006). Statutory changes to facilitate non-DNA innocence claims are long overdue. Medwed makes a number of suggestions for specific reforms such as extending the time frame for bringing new trial motions, only demanding evidence that "would have probably changed the outcome at trial," allowing direct appeal of a superior court's denial of habeas claims, not applying the abuse of discretion standard to summary denials of newly discovered evidence, and not sending newly discovered evidence claims to the original judge. Medwed also proposes developing a single remedy for newly discovered non-DNA evidence claims, resembling New York's approach, which combines attributes of new trial, habeas corpus, and writs of coram nobis, which can be used to correct fundamental errors of fact and law.

Freestanding Constitutional Right to Innocence

Innocence claims brought in habeas petitions typically allege violations of due process under the Fifth or Fourteenth Amendments and claim cruel and unusual punishment under the Eighth Amendment. In *Herrera v. Collins*, 506 U.S. 390, 417 (1993), the U.S. Supreme Court assumed without deciding that "in a capital case a truly persuasive demonstration of 'actual innocence' made after trial would render the execution of a defendant unconstitutional, and warrant federal habeas relief if there were no state avenue open to pursue such a claim." *Herrera* refrained from announcing a standard for evaluating a freestanding innocence claim, but cautioned that due to the disruptive effect on finality and the "enormous" burden on the state in having to retry cases on stale evidence, the threshold of such a claim would be "extraordinarily high" (Gross et al. 2006). The court characterized Herrera's showing as "falling far short" of satisfying this standard (Gross et al. 2006). His affidavits were submitted shortly before the scheduled execution, which was more than 10 years after the conviction, and asserted the defendant's brother who died 6 years earlier had committed the murder.

Justice O'Connor's concurrence, which was joined by Justice Kennedy, observed "executing the innocent is inconsistent with the Constitution" and would be "a constitutionally intolerable event" (Gross et al. 2006 at 418). The dissent of Justice Blackmun, joined by Justices Stevens and Souter, agreed with Justice O'Connor, declaring "the Constitution forbids the execution of a person who has been validly convicted and sentenced but who, nonetheless, can prove his innocence with newly discovered evidence" (Gross et al. 2006 at 431). Thus, five justices agreed that the execution of an actually innocent person would violate the Constitution.

In *House*, the court once again left open whether a truly persuasive freestanding innocence claim in a capital case would warrant federal habeas relief if no state avenues of relief remain available. In dicta, *Osborne* again assumed without deciding that an actual innocence claim could be brought in habeas, but noted "the high standard any claimant would have to meet" to succeed with such a claim (129S. Ct. at 2321). Given such language, it is not surprising that *Herrera* relief was denied to five people whom DNA later proved innocent, including one who had actually presented initial DNA results that excluded him as the perpetrator (Garrett, Claiming Innocence 2008). However, some circuits do not even recognize freestanding claims of actual innocence on federal habeas review (*see*, e.g., *In re Swearingen*, 556F.3d 344 (5th Cir. 2009)). Garrett suggests that judicial reliance on finality in this debate is misplaced since state DNA testing statutes have abandoned this rationale (Claiming Innocence 2008).

The concurring decision of Justices Scalia and Thomas in *Herrera* suggested that "[w]ith any luck, we shall avoid ever having to face this embarrassing question again, since it is improbable that evidence of innocence as convincing as today's opinion requires would fail to produce an executive pardon" (506 U.S. at 428). But some question whether the standard is set too high. As previously mentioned, despite the Supreme Court's direction that the district court hold a hearing concerning Troy Davis's innocence, the Supreme Court ultimately refused to review the merits of his claim. Moreover, even if a circuit permits a freestanding innocence claim, some categories of non-DNA cases appear never able to meet the test. For example, in *Mills v. Hill*, 330 Fed. Appx. 135, 136 (9th Cir. 2009), the court noted that "[w]e have rejected freestanding claims of innocence based on the affidavit of a mental health expert hired by the defense, reasoning that '[b]ecause psychiatrists disagree widely and frequently on what constitutes mental illness, a defendant could always provide a showing of factual innocence by hiring psychiatric experts who would reach a favorable conclusion.'" Until the Supreme Court provides more guidance about the existence and scope of actual innocence claims, the case law will likely continue to favor finality over innocence.

Compensation for Exonerees

Individuals who are exonerated in postconviction litigation are precluded by absolute immunity from successfully suing judges (*Pierson v. Ray*, 386 U.S. 547 (1967)) or prosecutors who are acting either as advocates or in their administrative roles (*Van de Kamp v. Goldstein*, 555 U.S. 335 (2009)). Law enforcement personnel are protected by qualified immunity (*Hunter v. Bryant*, 502 U.S. 224 (1997)), which also defeats many claims. As a result, even when a wrongful conviction is overturned, exonerees must rely on compensation statutes, section 1983 suits for violations of constitutional rights, state tort claims or private bills to obtain any monetary relief for their wrongful incarceration (Ad Hoc Innocence Committee 2006; Bernhard 2009; Bernhard 2004).

To date, less that 30 states have adopted compensation statutes. Without this avenue of relief, many unjustly convicted individuals who have had their lives virtually destroyed by their lengthy incarceration are unable to obtain any monetary reimbursement or services. While they can never recover their lost years, for some spent housed on death row or labeled as sexual predators, many states have yet to acknowledge a moral obligation to provide them with compensation and social services. Even some jurisdictions that have enacted statutes appear to remain hostile to providing full relief. For example, when such statutes are adopted, they are oftentimes quite miserly, with very low dollar caps, and may require proof of economic injury or have very short statutes of limitations (Ad Hoc Committee 2006; Raeder 2008). Few of them provide for any services, whether medical, psychological, educational, or assistance in obtaining housing or employment (Raeder 2008). In contrast, probationers and parolees are entitled to more assistance and social services than those who are released because they did not commit any crime. The lack of resources for exonerees poses a serious problem for individuals, many of whom who suffer from post-traumatic stress disorder from their lengthy prison ordeals (Raeder 2008), as well as for those who were incarcerated before the technological revolution vastly changed the nature of skills needed for employment.

Some of the existing statutes also require that exonerees obtain a pardon despite the fact that DNA may have clearly excluded the person as the wrongdoer. A more significant hurdle to recovery is that a number of statutes disallow compensation for those who are claimed to have contributed to their convictions (Ad Hoc Committee 2006). This can occur because the defendant pled guilty, as was evident in the nearly 200 cases reversed in the wake of the police scandals in Los Angeles and Tulia. Similarly, in cases involving false confession, many of those who confess are youths or have mental disabilities or low IQs. Yet, some of the statutes literally prohibit recovery per se, rejecting a case-by-case basis which would look at the surrounding circumstances (Ad Hoc Committee 2006).

Jurisdictions can differ about the details of what an appropriate statute should look like, but when unrealistically high barriers are enacted, the relief provided by such statutes becomes illusory. In the absence of any statute, exonerees are placed at the mercy of convincing a legislator to pass a private bill for compensation, unless they can find a lawyer who is willing to argue that their incarceration resulted from a constitutional violation. When constitutional violations are established, multimillion dollar verdicts can be returned, though recovery can be a lengthy process. In too many other cases, an exoneree comes away empty handed. Like the lottery, people who have spent decades in prison for crimes they did not commit get to see which one will receive $3 million, $300,000, or more typically nothing.

This lack of uniformity in providing compensation encourages arbitrary results (see Fact Sheet, Innocence Project, Compensating the Wrongly Convicted, accessed February 5, 2012, http://www.innocenceproject.org/Content/Compensating_The_Wrongly_Convicted.php). One possible model for legislation is the Federal Innocence Protection Act, 28 U.S.C.A. § 2513(e). The IPA sets compensation at $100,000 a year for those who have served time on death row and $50,000 a year for others. Several states have increased their compensation since the federal legislation was enacted.

Innocence awards may also be subject to taxation. Unlike California, very few state statutes exempt wrongful conviction compensation awards from state taxation. Moreover, a real question exists about whether the awards are also exempt from federal taxation. Efforts to pass a Wrongful Convictions Tax Relief Act failed in 2010 and 2007. To date, no federal tax legislation has been enacted. In 2010, a memorandum from the Office of Chief Counsel Internal Revenue Service on Tax Treatment of Compensation to Exonerated Prisoners clarified that an exonerated prisoner could exempt his compensation from federal taxation because he had suffered physical injury and sickness (Number: 201045023, Release Date: 11/12/2010, http://www.innocenceproject.org/docs/IRS_Compensation_Ruling.pdf). However, *Stadnyk v. C.I.R.*, 367 Fed. Appx. 586 (6th Cir. 2010), explicitly states that physical restraint alone does not provide the physical injuries or physical sickness necessary to exempt an award from taxation and notes that compensation for mental distress is also not exempted. Freeing the innocent requires more than releasing them to the streets penniless and with skills that atrophied a decade ago. Today many exonerees must rely on the fortuity of having family or friends to provide them with a home and a basic standard of living when they cannot obtain compensation.

Conclusions and Future Directions

In the absence of DNA exclusions, postconviction claims of innocence tend to be discounted because of our inability to readily identify individuals who were wrongfully convicted. *Arizona v. Youngblood*, 488 U.S. 51, 58–59 (1988), poses the most telling example of this dilemma. *Youngblood* denied relief where the state had not preserved evidence that the defendant claimed could exonerate him and also rejected any constitutional remedy in the absence of bad faith destruction of such evidence by the government. The decision did not even require any showing of good cause for the destruction. Yet some 12 years later, evidence was found that when submitted to newer DNA analysis exonerated Youngblood. In other words, the Supreme Court downplayed the possibility of Youngblood's innocence in setting a high threshold in a case where the defendant was factually innocent. That *Youngblood* is not an isolated example of our failure to accurately determine after conviction who is innocent is demonstrated by Garrett's study of exonerees, which revealed that the reason for denial of relief in one-third of the cases with written decisions on direct appeal was harmless error (Garrett, Judging Innocence 2008). He found that in half of the written decisions, courts referred to the likely guilt of the defendant, and 10 % referred to the evidence of guilt as overwhelming. Thus, appellate review not only failed to rectify the wrongful conviction, but judges also incorrectly identified innocent defendant as guilty and thereby disregarded the strength of the procedural errors during trial. Given the lack of a viable constitutional theory to raise claims of innocence on direct appeal, it is not surprising that of the 16 exonerees who sought a new trial based on newly discovered evidence of their innocence, none received relief prior to obtaining DNA testing. Similarly, the five who raised *Herrera* claims had them denied.

To respond to this disconnect, North Carolina has set up a commission that inquires into individual cases and actually reviews evidence not submitted at trial (The North Carolina Innocence Inquiry Commission, accessed February 5, 2012, http://www.innocencecommission-nc.gov/index.html). This type of commission resembles the model used in the United Kingdom and Canada (Ad Hoc Committee 2006; Griffin 2001; Scheck and Neufeld 2002). While such an approach may be considered too bold at present for most jurisdictions to adopt, commission review is no guarantee that innocence will be established, although three of the four cases currently decided resulted in exonerations.

Thus, until technology or forensic science is able to provide the equivalent of DNA exclusions in cases not involving biological evidence, such postconviction claims of innocence will understandably continue to be treated skeptically. Dedicated prosecutors and defense counsel are the best protectors of innocence, but we also need to be more open to reviewing policies, practices, and legal doctrines across the criminal justice system to ensure that procedures that appear to create high risk of error are changed to reduce the risk of wrongful convictions. In addition, more attention should be paid to providing fair and speedy compensation and services to individuals who have satisfied the high standard necessary to establish their factual innocence.

Recommended Reading and References

Ad Hoc Innocence Committee to Ensure the Integrity of the Criminal Process of the ABA Criminal Justice Section (2006) Compensation for the wrongfully convicted. In: Giannelli P, Raeder M (eds) Achieving justice: freeing the innocent, convicting the guilty. American Bar Association, Washington, DC

Bernhard A (2004) Justice still fails: a review of recent efforts to compensate individuals who have been unjustly convicted and later exonerated. Drake Law Rev 52:703–738

Bernhard A (2009) A short overview of the statutory remedies for the wrongly convicted: what works, what doesn't and why. Boston Univ Pub Int Law J 18:403–417

Brennan ME (2008) Interpreting the phrase "newly discovered evidence": may previously unavailable exculpatory testimony serve as the basis for a motion for a new trial under rule 33? Fordham Law Rev 77:1095–1145

Brooks J, Simpson A (2011) Blood sugar sex magik: a review of postconviction DNA testing statutes and legislative recommendations. Drake Law Rev 59(Spring):799–866

Garrett BL (2008a) Claiming innocence. Minn Law Rev 92:1629–1719

Garrett BL (2008b) Judging innocence. Colum Law Rev 108:55–142

Grann D (2009) Trial by fire, did Texas execute an innocent man? The New Yorker. http://www.newyorker.com/reporting/2009/09/07/090907fa_fact_grann?currentPage=all

Griffin L (2001) The correction of wrongful convictions: a comparative perspective. Am Univ Int Law Rev 16:1241–1308

Gross SR, Shaffer M (2012) Exonerations in the United States, 1989–2012: report by the national registry of exonerations. http://www.law.umich.edu/special/exoneration/Documents/exonerations_us_1989_2012_full_report.pdf

Gross SR, Jacoby K, Matheson DJ, Montgomery N, Patil S (2006) Exonerations in the United States 1989 through 2003. J Crim Law Criminol 95:523–560

Medwed DS (2005) Up the river without a procedure: innocent prisoners and newly discovered non-DNA evidence in state court. Ariz Law Rev 47:655–718

Medwed DS (2007) California dreaming? The golden state's restless approach to newly discovered evidence of innocence. UC Davis Law Rev 40:1437–1480

Nelson J (2008) Facing up to wrongful convictions: broadly defining "new" evidence at the actual innocence gateway. Hastings Law J 57:711–757

Orenstein A (2011) Facing the unfaceable: dealing with prosecutorial denial in postconviction cases of actual innocence. San Diego Law Rev 48:401–446

Raeder MS (2008) Introduction to wrongful convictions symposium. Southwest Univ Law Rev 37:745–761

Raeder MS (2009) Postconviction claims of innocence. Crim Justice 24:14–25

Scheck BC, Neufeld PJ (2002) Toward the formation of "innocence commissions" in America. Judicature 86:98–105

Segal B (2008) Habeas corpus, equitable tolling, and AEDPA's statute of limitations: why the *Schlup v Delo* gateway standard for claims of actual innocence fails to alleviate the plight of wrongfully convicted Americans. Univ Hawaii Law Rev 31:225–256

Wolf MP (2009) Habeas relief from bad science: does federal habeas corpus provide relief for prisoners possibly convicted on misunderstood fire science? Minn J Law Sci Technol 10:213–254

P

Parametric Sample Selection Models

William Rhodes
Abt Associates Inc., Cambridge, MA, USA

Overview

James Heckman's work (1976, 1979) regarding the consequences and solutions for sample selection bias has influenced three decades of social science research. For this and related contributions, he won a Nobel Prize in economics. Heckman's original solution requires strong parametric assumptions. When wrong, those assumptions can lead to parameter estimates with mean-squared errors worse than those of estimators that ignore selection bias.

Nevertheless, the parametric approach remains popular. Discussions appear in econometrics textbooks (Davidson and MacKinnon 1993, pp. 542–545; Cameron and Trivedi 2005, pp. 546–569; Greene 2008, pp. 863–903) and technical papers; estimation routines appear in standard econometrics software; and the approach is frequently used in criminal justice research (Bushway et al. 2007). Criminologist and criminal justice researchers cannot disregard the parametric approach.

All estimators require some parametric assumptions, so what distinguishes estimators that are parametric from another class? Estimators considered in this entry require strong distributional assumptions about error terms in a regression and are distinguished thereby from estimators that impose weaker assumptions.

Without being comprehensive, this entry identifies six cases where selection bias occurs in criminal justice research:

- Sample selection due to unobserved variables
- Estimating treatment effects using observational data
- Breakdowns in randomized controlled trials
- Survival models with censoring and competing events
- Endogenous stratification/choice-based sampling
- Missing data

This entry shows how parametric assumptions can overcome bias but also how the approach can fail. It illustrates points with Monte Carlo experiments.

Theoretical Background

To motivate a discussion, consider a *data generation process* (DGP) comprising two equations. The first is the *outcome equation*: It explains the outcome of interest in terms of explanatory variables. **The objective is to estimate one or more parameters of this outcome equation**. The second is the *selection equation*: It explains how data get sampled for analysis. Selection bias *may* occur when the same factors affect both sampling and outcome (Rosenbaum 2002). After introducing notation, this entry explains when selection bias occurs and how parametric models correct for bias.

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

When writing the two equations, use the notation:

Y Denotes the outcome measure: the dependent variable in the first equation. A subscript denoting the ith observation is implicit.

S Denotes the selection measure: the dependent variable in the second equation. Its interpretation will differ across the six cases.

X This is an *observed* row vector of variables that affect both Y and S. Observed means that the researcher has measured these variables and uses them in the analysis. X typically includes a constant.

W This is an *unobserved* row vector of variables that affect both Y and S. The researcher has not measured these variables or else omits them from the analysis. Assume that unobserved variables are distributed independent of other explanatory variables appearing in the outcome equation; consequences of assuming otherwise are discussed subsequently.

V This is a vector of observed variables that affect Y but not S.

Z This is a vector of variables that affect S but not Y. When observed, the vector comprises *instrumental variables* used to impose *identification restrictions*.

β This represents a conformable column vector of parameters appearing in the outcome equation. Subscripts will denote whether these pertain to X, W, or V. Conformable means the column vector β has as many elements as the row vectors X, W, and V.

α This represents a conformable column vector of parameters appearing in the selection equation. Subscript will denote whether these parameters pertain to X, W, or Z.

e This represents an error term pertaining to the first equation.

u This represents an error term pertaining to the second equation.

For simplicity, assume the error terms are mean-zero, identically and independently distributed within each equation, independently distributed across equations, and independent of X, W, V, and Z. Assuming otherwise would add complications without enhancing insight.

Given this notation and limiting the immediate discussion to linear models, a general data generation process (DGP) is written:

$$Y = X\beta_X + W\beta_W + V\beta_V + e \tag{1}$$

$$S = X\alpha_X + W\alpha_W + Z\alpha_Z + u \tag{2}$$

A DGP is the *actual* stochastic process generating the data. The researcher necessarily omits unobserved variables from the analysis, so a model of the DGP is expressed as:

$$Y = X\beta_X + V\beta_V + e^* \tag{3}$$

$$S = X\alpha_X + Z\alpha_Z + u^* \tag{4}$$

The error terms change to e* and u*:

$$\begin{aligned} e^* &= e + W\beta_W \\ u^* &= u + W\alpha_W \end{aligned} \tag{5}$$

Some additional notation will be useful:

$\sigma^2_{e^*}$ This is the variance of e*.

$\sigma^2_{u^*}$ This is the variance of u*.

$\sigma_{e^*u^*}$ This is the covariance of e* and u*.

Although e is independent of X in the population, the following section will show that e* is not independent of X in the sample used for estimation. Consequently, a regression based on model (3) will violate a fundamental assumption: errors must be independent of explanatory variables. Standard regression procedures yield biased and inconsistent estimates of β_X.

Is the DGP encapsulated by Eqs. 1 and 2 plausible or just abstract mumbo-jumbo? In fact behavioral models often translate into DGPs with counterparts to Eqs. 1 and 2. Based on perceptions of self-interest, individuals elect to participate in activities (Eq. 2) studied by researchers (Eq. 1). Alternatively, based on perceptions of community interests, agents (prosecutors, judges, probation officers, etc.) perform selection to enhance community safety. Thus, the DGP and statistical models cover situations common to criminology/criminal justice.

Correctly specified parametric models eliminate selection bias. One approach uses maximum likelihood or partial likelihood to estimate parameters of the joint distribution of e* and u* in the sample. The joint distribution is written as

$$\phi_b(e^*, u^*|\text{sample}) \quad (6)$$

The α, β, and σ parameters are implicit arguments in this distribution. The notation indicates that the distribution depends on the sampling procedure, and different sampling procedures lead to various distributions (Maddala 1983), only a few of which are discussed in this entry.

A second approach is to estimate the expected value of e* in the sample and introduce this expected value into the model. Equation 3 now becomes

$$Y = X\beta_X + V\beta_V + \varsigma E[e^*|\text{selection}] + v \quad (7)$$

The new error term *v* is mean-zero in expectation but heteroscedastic, so computed standard errors require adjustment (Cameron and Trivedi 2005, p. 567; Greene 2008, p. 888). Modern software applies these corrections.

Some parametric assumptions are required. Although other distributions might be used, typically analysts assume that e* and u* are distributed as bivariate normal, perhaps after transformations to increase model flexibility and perhaps after converting the σ from scalars to functions of explanatory variables. Inferences are asymptotic, meaning that valid inferences require large samples.

Econometricians have introduced semiparametric approaches for estimating E[e*|selection] (Vella 1998). These approaches are not commonly used; they do not appear in standard computing software; and (Greene 2002, pp. E23–2) for one is skeptical that semiparametric approach will reach mainstream application. This entry does not discuss semiparametric approaches.

In practice Eqs. 6 and 7 will not produce the same results. Results may be very different if the distributional assumptions are wrong. Furthermore, Eq. 6 can be adapted for models that are nonlinear in the outcome equation. This is unclear for Eq. 7 (Cameron and Trivedi 2005, p. 192). Many models are nonlinear, so Eq. 6 seems preferable, but Eq. 7 has heuristic and practical value.

The form provided by Eq. 7 is especially useful for interpreting results. Let x represent a single component of X and let β_x represent the corresponding parameter. If the researcher's interest is how x changes Y in the population, then the answer is β_x. But often the researcher's interest is how Y changes with x given a selection mechanism that has behavioral interpretations. Then the answer is

$$\frac{\partial Y}{\partial x} = \beta_x + \varsigma \frac{\partial E[e^*|\text{selection}]}{\partial x} \quad (8)$$

The correct interpretation depends on a well-specific research question (Heckman and Vytlacil 2007a, b).

The difficulty of applying these solutions depends on an as yet unspecified assumption about the selection process. When population data are available for estimating Eq. 4 but only selected data are available for estimating Eq. 3, the data are *censored*. When only sampled data are available for estimating both Eqs. 3 and 4, the data are *truncated*. This entry deals exclusively with censored data although similar concepts apply to truncated data. Unfortunately there is room for confusion, because the censoring problem is sometimes referenced as *incidental truncation* to mean that the data for estimating Eq. 3 are truncated by the process described by Eq. 2.

This entry is focused on selection rules that bifurcate the population, but other selection processes occur. Especially in criminal justice applications, where justice administration acts as a sieve, selection may involve a sequence of decisions (Vella 1998, p. 153). Also in criminal justice applications, the selection rules may put individuals into ordered categories (low risk, medium risk, high risk) or into nominal categories. There exist suitable econometric models (Vella 1998), some of which have been programmed into conventional computing software, but discussion of alternative selection rules goes beyond the scope of this entry.

Six Cases Illustrating Selection Bias

Sample Selection Due to Unobserved Variables

Assume that the selection equation determines which data enter into estimation of the outcome equation. Specifically, a data point is selected when S is larger than a threshold value; the threshold is 0 here. Elaborate the data generation process:

$S \geq 0$ implies selection into the sample used to estimate the outcome equation.

$S < 0$ implies exclusion from the sample.

From Eq. 2, sample selection occurs when $S = X\alpha_X + W\alpha_W + Z\alpha_Z + u \geq 0$ or when $W\alpha_W \geq -X\alpha_X - Z\alpha_Z - u$

When the linear combination $X\alpha_X$ is very large in the population, any value of the linear combination $W\alpha_W$ will likely pass the test for inclusion in the sample; when $X\alpha_X$ is very small, only large values of $W\alpha_W$ will likely pass the test. This induces a negative correlation between $X\alpha_X$ and $W\alpha_W$ *in the sample*. The reason is that the linear combination $W\beta_W$ is an element of e* in the model represented by Eq. 3. Given their common elements, $W\alpha_W$ and $W\beta_W$ are likely correlated, in which case elements of X will be correlated with the error term e* in the sample.

The same argument establishes that X will be correlated with Z in the sample. This does not matter, however, because Z is not part of the outcomes equation. V appears in the outcomes equation, but V does not affect e*. The entire explanation for selection arises because the unobserved W appears in both the outcome equation and the selection equation (Rosenbaum 2002). The classification of variables as X, W, V, and Z depends on the model, an observation important for model building (Pearl 2000; Gerring 2012), but those implications are not discussed here.

What are the solutions? An obvious solution is to assure that X comprises variables that affect both the selection and outcome equations so there are no W variables. That may be impossible and the solution is uncertain because there is seldom reason to justify that X is inclusive. A researcher who relies on this reasoning will typically face validity challenges. Nevertheless, this assumption underlies all estimation procedures that assume *selection on the observables*.

In contrast, Heckman's solution assumes selection on the unobserved W. As noted, estimation requires assuming a joint distribution of e* and u*. Assuming that e* and u* are bivariate normal is problematic. Even if e and u are independent normal, bivariate normality of e* and u* depends on $W\alpha_W$ and $W\beta_W$ being distributed as normal in the population. Although possible, it is a strong assumption, and if wrong, $\hat{\beta}$ will be inconsistent.

Criminologists and criminal justice researchers have used the parametric model to deal with selection bias (Bushway et al. 2007). For example, in an early application in criminal justice research, the author used Heckman's approach to estimate how risk factors predict pretrial misconduct; the sample selection mechanism entered because the worst pretrial risks were detained pending trial so they were omitted from the sample for estimating the outcome equation. That application illustrates a point: Eq. 4 can be estimated consistently because the analyst observes release/detention decisions for every offender. Selection bias affects the estimation of Eq. 3 because the outcomes are observed for only those offenders released pending trial, but the inference is about risk in the population of offenders. The application motivates two-step estimation: Estimate the pretrial release decision first and then estimate the pretrial misconduct equation.

Estimating Treatment Effects Using Observational Data

In the second case, the selection equation determines who enters treatment. An explanation requires modest changes to Eq. 1 through Eq. 4. Let

T is a dummy variable that denotes treatment when $T = 1$ and no treatment when $T = 0$

δ a vector of parameters (the treatment effect), which is the objective of estimation

Then

$S \geq 0$ implies selection into treatment ($T = 1$).

$S < 0$ implies no treatment ($T = 0$).

Treatment potentially changes all the parameters in Eq. 1, so rewrite Eq. 1 as

$Y = X\beta'_X + W\beta'_W + V\beta'_V + e$ if treated

$Y = X\beta_X + W\beta_W + V\beta_V + e$ otherwise

Equivalently the DGP can be rewritten as a single equation with interactions

$$\begin{aligned} Y =& X\beta_X + W\beta_W + V\beta_V \\ &+ [X\delta_X + W\delta_W + V\delta_V]T + e \end{aligned}$$

The model becomes

$$\begin{aligned} Y =& X\beta_X + V\beta_V + [X\delta_X + V\delta_V]T \\ &+ [W\beta_W + WT\delta_W + e] \\ =& X\beta_X + V\beta_V + [X\delta_X + V\delta_V]T + e^{**} \end{aligned}$$

Where

$$e^{**} = W\beta_W + WT\delta_W + e$$

Often selection into treatment depends on whether the person being treated his or her agent perceives treatment to be beneficial. This theory motivates specification of the selection equation. It is common to assume that the δ are zero except for a constant—that is, $X\delta_X + V\delta_V = \delta$ and $WT\delta_W = 0_.$ This assumption has negative consequences if wrong. Nevertheless, for simplicity, treat δ as a scalar.

T is not orthogonal to e**. The reasoning is the same as before. T is partly determined by W; also, e** is correlated with W. Hence, T is correlated with e**. An analyst could not estimate δ consistently in a regression that fails to account for selection.

Simple solutions follow when selection on observables holds (Imbens and Wooldridge 2009). Otherwise, estimation rests on strong parametric assumption about the joint distribution for e** and u**. As a practical matter, parameter identification requires that Z be observed and included in the model. Identification is discussed later. For example, in the criminal justice literature, the author and his or her colleagues (Rhodes et al. 2001) used the parametric approach to evaluate within prison substance abuse treatment.

Breakdowns in Randomized Controlled Trials

RCT imposes selection by design. Unless the design is perverse, random assignment typically solves the selection bias problem. There are three situations. First, selection is purely random: The selection mechanism contains no variables in common with the outcome equation. Estimation is straightforward: the mean difference between the treated and control group is an unbiased estimate of the average treatment effect; a regression might be used to increase efficiency or to learn about heterogeneous treatment effects.

Second, selection is random conditional on X (exogenous sampling). Stratified sampling obviously qualifies, as do most complex survey designs. There is no sampling bias provided the evaluator uses an estimator that accounts for X. This might be done with weighting, but that topic goes beyond the scope of this entry.

Third, selection is random conditional on W. This perverse case happens. Consider a multisite study where the researcher samples randomly from within a site but does not select randomly across sites – perhaps because many sites decline to participate in the study. Or, consider the problem where some subjects assigned to treatment refuse the treatment or some subjects assigned to the control condition cross over to the treatment condition, or both. Random assignment involves selection bias in these common cases. Evaluators sometimes use the terms internal and external validity to describe this perverse case.

Thus, well-crafted and implemented RCTs are free of selection bias. However, true RCTs are difficult to design and implement, and selection bias occurs when design/implementation is faulty. Selection may be present in RCT (Berk 2005; Sampson 2010), and furthermore, the RCT may be unable to answer important policy questions about the distribution of treatment effects (Heckman and Vytlacil 2007a; Heckman et al. 1997). Meta-analysis (Borenstein et al. 2009) often integrates experimental and quasi-experimental studies to understand selection bias and better estimate treatment effects.

Survival Models with Censoring and Competing Events

Survival analysis has a long history in the study of criminal recidivism. Advances have incorporated competing events into the study of criminal recidivism. Selection bias can occur when estimating survival models for the reasons discussed in 2.2 through 2.4, but survival models sometimes raise special selection problems.

Equations 1 and 2 will again be useful but now recast those equations by assuming there are two events, A and B. Event A might represent the occurrence of criminal recidivism, defined as an arrest for a new crime while under community supervision. Event B might represent the occurrence of a revocation for a technical violation of the conditions of supervision, treated here as a competing event. Y and S represent the logarithm of the time until events A and B, respectively. The error term e is the distribution of log-time until an arrest conditional on X, W, and V; the error term u is the distribution of log-time until a revocation conditional on X, W, and Z. The analyst can observe the timing for whichever event occurs first: A or B. Consequently, one can treat B as "selecting" what is observable about A.

Almost always, analysts assume that the timing of event A (the outcome of interest) is conditionally independent of the timing of event B (competing events). This is equivalent to assuming selection on observables. If selection on observables is wrong, traditional survival analysis will result in biased parameter estimates.

Given censoring because of unobservables, the solution is also familiar. The analyst needs to model the joint distribution of e* and u*, but accounting for the error structure is complicated (van den Berg 2008). Moreover, commonly available software lacks estimation routines. Nevertheless, the maximum likelihood approach introduced earlier provides a solution requiring parametric assumptions; the author used this approach to study criminal recidivism under community supervision.

To extend this observation, the reader might note that survival models are simply censored regressions. Therefore, the same selection problems that affect survival models also affect Tobit models, which are frequently used in criminal justice research.

Endogenous Stratification/Choice-Based Sampling

Many samples depend on the occurrence of some event. For example, data from the Arrestee Drug Abuse Monitoring (ADAM) survey requires a subject to be arrested and booked. Samples used in the criminal careers literature require a subject to acquire a criminal history. Such samples involve endogenous stratification or choice-based sampling.

To illustrate the problem, return to Eq. 1 and think of Y as a count variable – zero arrests, one arrest, two arrests, and so on occurring from a Poisson process. The research question is to estimate the arrest rate during a period of interest.

Return to Eq. 2 and think of S as being the event: arrested and booked on the day that ADAM arrestees were sampled. Selection bias occurs because offenders with high crime commission rates have relatively high sampling probabilities. The sample will not represent the general population of offenders.

Notice a subtle difference between endogenous stratification and the other selection problems: Selection into the sample depends on the process determining the outcome. To illustrate, assume a subject is sampled into the study depending on the outcome, so:

$Y > 0$ implies selection into the sample

$Y \leq 0$ implies no selection into the sample

The outcome equation is written as

$$Y = X\beta_X + V\beta_V + e^{***}$$

The expected value for e^{***} *in the sample* is

$$E[e^{***}] = E[e|e > -X\beta_X - W\beta_W - V\beta_V]$$

The error term e^{***} will be correlated with X and V in the sample.

Endogenous stratification raises complex estimation issues (Cameron and Trivedi 2005, pp. 822–829). Solutions may be best developed for Poisson and negative binomial models (Cameron and Trivedi 1998), which is fortunate given the ubiquitous role of Poisson models in criminal justice modeling. The criminal careers literature has sometimes dealt with the issue; early illustrations come from the Rand Corporation and others. The author has used a negative binomial model with endogenous stratification to estimate the arrest rate for drug-involved offenders in the community based on a sample of arrestees (Rhodes et al. 2007).

Missing Data

Analysts often deal with missing data (Schaefer 1997; Little and Rubin 2002). For example, within a data set, sex may be missing for some subjects, age may be missing for others, and both age and sex may be missing for still others. The *variable* is observed (it is an X variable), but its *value* is missing for some observations.

One way to deal with missing data is to exclude cases where any data are missing. This is called list-wise or case-wise deletion. This depletes the sample and thereby reduces power; it can bias parameter estimation.

An alternative approach is to impute values for the missing data, and imputation software is available for common problems. Imputation and estimation with imputed data is complicated. Consider a simple problem: A single variable x from the vector X has missing values. Assume that x is the Kth element of X and that the other K-1 elements are never missing. To focus attention on x, rewrite Eq. 1 equivalently as

$$Y = X_{1:K-1}\beta_{X_{1:K-1}} + x\beta_{X_K} + W\beta_W + V\beta_V + e \tag{9}$$

Interpret the selection equation as determining when x is observed ($S > 0$) and missing ($S < 0$). To emphasize the role of x, rescale so the $\alpha_K = 1$ and rewrite Eq. 2 as

$$S = X_{1:K-1}\alpha_{X_{1:K-1}} + x + W\alpha_W + Z\alpha_W + u \tag{10}$$

This implies that x is observed when

$$x \geq -X_{1:K-1}\alpha_{1:K-1} - Z\alpha_Z - W\alpha_W - u \tag{11}$$

As before, the motivation is to estimate the parameters in the model:

$$Y = X_{1:K-1}\beta_{X_{1:K-1}} + x_K\alpha_K + V\beta_V + e^* \tag{12}$$

Consider three cases. The first is when elements of α in Eq. 11 are 0 except perhaps for the constant. Selection, which depends only on u, is called *missing completely at random*. Missing data will not bias the parameter estimates for Eq. 12 because the error term e* will not be correlated with X, x, or V. The second situation is when α_W is zero but other α parameters are not zero. Selection, which now depends on one or more of the observed variables, is called *missing at random*. Missing data will not bias parameter estimates because the error term e* will be uncorrelated with X, x, and V. The third situation is where $\alpha_W \neq 0$. Selection, which now depends on unobservable variables, is called *missing not completely at random*. This leads to biased parameter estimates because the missing data will be correlated with the error term e*.

There are parametric solutions to the first two cases and approximate solutions to the third case. The solution requires specifying x as a function of observables. A specification might be

$$x = Y\varsigma_Y + X_{1:K-1}\varsigma_{X_{1:K-1}} + V\varsigma_V + Z\varsigma_Z + R\varsigma_R + v \tag{13}$$

There are some surprises in Eq. 13. One is that the outcome variable Y appears in the specification. The intuition is that x predicts Y, so Y predicts x, and predicting x is the purpose of Eq. 13. A second surprise is inclusion of a vector of variables R that did not appear elsewhere in the DGP. Again, the instrumental objective is to predict x, and it seems plausible that some set of variables R might predict x yet not explain Y. Finally, v is an error term *whose distribution is known* except for some unspecified parameters. The solution is parametric because of the linear specification and the assumed distribution for v. Estimation of Eq. 13, including the parameters of the distribution of v, leads to a predicted (posterior) distribution for missing values for x.

Estimation of Eq. 9 proceeds by sampling values of x from this posterior distribution and using those sampled values to estimate the parameter appearing in Eq. 9. The approach is conceptually simple but estimation rests on the specified distributional assumptions, is computer intensive, and requires specialized software.

In the first two cases—missing completely at random and missing at random—Bayesian logic leads to consistent estimates for the parameters in Eq. 9. This is not true in the third case. There the analyst has two choices. The first is to construct an imputation model that is explicit about the selection mechanism; this is not frequently done. The second is to use imputations from Eq. 13 with an expectation that these will reduce mean-squared errors in the β. This is plausible because the imputations are based on additional information that would otherwise be ignored by case-wise deletion.

The single imputation problem is simple compared with a multiple imputation problem where all of the X and V variables may have missing values. In that case, the starting assumption is that X and V are distributed as multivariate normal perhaps after transformations and after conditioning on other variables such as Y and R. This solution is highly parametric and probably unrealistic. Nevertheless, experience shows that estimation is insensitive to distributional assumptions although one would be dubious of applying this approach to data that have a high proportion of missing to known values.

As an illustration, the ADAM survey randomly samples arrestees, questions them about drug use, and requests a urine specimen to test for recent drug use. Not all arrestees are able or willing to comply with the request, leading to missing results for urine tests. ADAM analysts have developed procedures used to impute urine test results conditional on self-reports about recent drug use.

Identification

The first five cases share a common problem: The arguments assume that W is orthogonal to X and V. This assumption was convenient because it meant that in the absence of any selection bias, an analyst could consistently estimate the β parameters.

What if the distribution of W is not orthogonal to X and V? Then when W is omitted from the outcome equation, the β parameters would not be identified even absent selection bias. The problem is omitted variable bias, which is discussed in introductory textbooks and is widely appreciated by researchers. One can nevertheless think of Eq. 3 as estimating a *conditional mean* for the population (Angrist and Pischke 2009). Provided the analyst considers his or her estimates to pertain to a conditional mean function, the argument regarding selection bias does not require any assumptions about independence between W and either X or V, so one can focus on selection bias as *the* problem.

Putting this discussion about orthogonality aside, the parametric solution to selection bias requires an assumption about the joint distribution between e* and u*. Suppose the assumption is correct. Is this sufficient to identify the β? The answer is a qualified *yes*, but asymptotic justification for the parameter estimates and their sampling distributions may require very large samples.

Equation 7 provides some intuition. Absent Z, $E[e^*|\text{sample}]$ will be nearly collinear with X. The relationship will not be perfectly collinear because the nature of the joint distribution ϕ_b will induce nonlinearities, but the departure from collinearity will be slight in most real-world applications. Consequently, the standard errors for β_X would be large and inference tenuous. Few analysts would feel comfortable relying on identification resting on nonlinearities induced by correct specification of the joint distribution of e* and u*.

A more compelling case for identification requires the presence of Z. Intuition is again useful. The larger is the explanatory power of Z in Eq. 4, the smaller is the correlation between X and $E[e^*|\text{sample}]$. The availability of Z mitigates the collinearity problem. There is an extensive literature on instrumental variables (Hahn and Hausman 2003; Stock et al. 2002). This entry is not the place to summarize that literature except to review some properties of a good instrument.

First, to be an instrument, Z must appear in the selection equation but not in the outcome equation. Second, to be a *useful* instrument, Z must account for a good deal of the variance in S. A weak instrument (e.g., one that accounts for a negligible amount of the variance in S conditional on X) can provide an estimate with worse mean-squared error properties than estimates from models that ignore selection bias. Unfortunately, good instruments are difficult to find and to justify, a fact that greatly limits the appeal of the Heckman estimation procedures.

This is a good point to pose the question: If correcting for selection bias can worsen estimates, can an analyst tell if correcting for selection bias is necessary? The answer is a tentative yes: The test is equivalent to testing whether $\sigma_{e^*u^*} = 0$. The practical problem is that without a strong instrument, this test will lack power. An analyst would frequently fail to reject the null hypothesis of no selection bias when, in fact, selection bias is serious. There are tests of whether the parameters are identified by nonlinearities or identification restrictions, but these tests require correct parametric assumptions and, anyway, do not answer the question of whether selection bias exists.

A final point closes this section: Much of the argument regarding instrumental variables pertains to linear regression models and does not translate readily into the nonlinear models often used in criminology/criminal justice research. For useful discussions, see Davidson and MacKinnon (1993), pp. 224–226 and Cameron and Trivedi (2005), pp. 192–199. This is a difficult problem. Angrist and Pischke (2009) advise researchers to assess if the research question is answerable without actually estimating the β. The author (Rhodes 2010) adopted that recommendation in a justice context.

Monte Carlo Illustrations

A Monte Carlo experiment illustrates points based on the first case (section "Sample Selection Due to Unobserved Variables"). There are single X, W, V, and Z variables. X, V, and Z are distributed as independent normal with mean of zero and standard deviation of 1; W has different distributions across the experiments. The α and β parameters all equal 1. The error terms e and u are distributed as normal with mean of zero and standard deviation of 2. The simulation generates 1,000 observations before selection (about 500 after selection), and the experiment is replicated 5,000 times.

In the first experiment (see Table 1), W is distributed as normal with mean of 0 and standard deviation of 2. This is ideal for the Heckman estimator because e* and u* are distributed as joint normal. The table shows parameter estimates using an OLS regression, Heckman's maximum likelihood, and Heckman's two-step procedure. The table reports the average of the three β parameters across 5,000 replications (β_C is the constant), the "true" standard error across those 5,000 replications (based on the 5,000 replications) and the average of the estimated standard errors across those 5,000 replications.

OLS with a robust covariance estimator yields biased and inconsistent parameter estimates for β_X and β_C. The estimate for β_V is not biased. The maximum likelihood Heckman procedure yields consistent parameter estimates at the expense of inflating standard errors. The two-step procedure yields consistent parameter estimates although the standard errors are even larger. The Heckman estimators perform well in this ideal case.

The second exercise changes the distribution for the instrument Z. Although still normally distributed with mean of zero, it now has a standard deviation of 0.01. This simulates a weak instrument. Everything else in the simulations remains the same. Results are reported as exercise 2 in Table 1.

The table shows no results for the maximum likelihood procedure, which often fails to converge on a solution. The two-step estimator is less demanding and makes the intended points. First, the two-step estimator yields seriously biased parameter estimates, which should give pause to those who seek to rely upon nonlinearity to identify parameters. Second, the mean-squared error properties of the two-step estimator are worse

Parametric Sample Selection Models, Table 1 A Monte Carlo demonstration of selection bias

		Least squares			Maximum likelihood			Two-step		
		Estimate	True standard error	Estimated standard error	Estimate	True standard error	Estimated standard error	Estimate	True standard error	Estimated standard error
Experiment 1	β_X	0.728	0.123	0.122	0.993	0.163	0.160	1.002	0.176	0.174
	β_V	0.999	0.118	0.118	0.999	0.117	0.117	0.999	0.117	0.117
	β_C	2.077	0.121	0.121	1.024	0.416	0.407	0.990	0.487	0.485
Experiment 2	β_X	0.694	0.122	0.121				0.859	1.023	1.406
	β_V	1.001	0.118	0.117				1.001	0.118	0.146
	β_C	2.145	0.120	0.120				1.522	3.925	5.365
Experiment 3	β_X	0.872	0.112	0.112	1.025	0.175	0.153	1.006	0.151	0.150
	β_V	1.000	0.107	0.106	1.000	0.106	0.105	1.000	0.107	0.105
	β_C	1.306	0.110	0.113	0.806	0.436	0.353	0.867	0.351	0.343
Experiment 4	β_X	−0.034	0.397	0.373				1.638	3.941	5.708
	β_V	1.002	0.375	0.351				1.006	0.374	0.499
	β_C	3.432	0.416	0.403				−3.744	17.346	24.586
Experiment 5	β_X	0.973	0.090	0.090	0.999	0.121	0.119	0.999	0.121	0.120
	β_V	0.999	0.086	0.086	0.999	0.086	0.086	0.999	0.086	0.086
	β_C	1.576	0.088	0.089	1.499	0.244	0.240	1.498	0.242	0.241
Exercises:										

1. W has a normal distribution and the instrument is strong
2. W has a normal distribution and the instrument is weak
3. W has a lognormal distribution with variance of 1.02 and the instrument is strong
4. W has a lognormal distribution with variance of 1.02 and the instrument is weak
5. W has a binomial distribution with probability equal to 0.5

than those of the OLS estimator: 0.329 for OLS and 1.033 for Heckman.

The third exercise is the same as the first but changes the distribution for W. Previously distributed as normal, it is now distributed as lognormal. This is easily accomplished by taking the exponential of the original W multiplied by 0.7. The distribution of W now has a mean of 1.278 and a standard deviation of 1.02. The exercise subtracts 1.278 from the simulated values of W to center the new distribution of W on zero. The assumption of a lognormal distribution for W violates the assumptions that e* and u* are distributed as bivariate normal. Nevertheless, the maximum likelihood and two-step procedures perform tolerably given a strong instrument.

The fourth exercise is the same as the third but weakens the instrument by changing Z to have a normal distribution with standard deviation of 0.01. As before, convergence is a problem for the maximum likelihood solution, but the two-step approach provides a solution. Estimates are biased with large sampling variances.

The fifth and final exercise uses the original W but now changes it to a dummy variable coded 1 if W exceeds 0 and coded 0 otherwise. The instrument is strong. As before, the Heckman procedure provides tolerable solutions especially for the slope coefficient despite the faulty distributional assumptions.

Simulations cannot cover all situations, but generalizing based on the above leads to two conclusions: First, the Heckman procedure is fairly robust to departures from bivariate normality provided the instrument is strong. Second, the Heckman procedure provides poor estimates when instruments are weak regardless of whether distributional assumptions are correct.

Conclusions

Application of the parametric approach is demanding, but this is equally true of partially parametric approaches. An applied researcher should provide a sober, informed consideration of whether his or her data can answer the research question. The following observations might inform that consideration.

First, when the researcher has access to a rich set of data, which includes all variables relevant to the outcome equation according to substantive theory, the researcher might assume selection on the observables. In general, assuming selection on the observables is difficult to justify, but justification is advanced when the setting allows repeated observations (difference-in-differences) and exogenous interruptions (regression discontinuity) provided the researcher is content with estimating a subset of the β identified by these designs.

Second, when selection on the observables is untenable, the researcher needs strong instruments to impose identifying restrictions. Good instruments are difficult to find and justify, but examples exist in criminal justice research. Although nonlinearities can provide identifying restrictions, these are not reliable.

Weak instruments are a problem. All the estimators discussed in this entry provide estimates of β that are biased but consistent provided distributional assumptions hold. The weaker the instrument, the larger the sample required to reduce the bias to an acceptable level. Asymptotic justification can require samples in the tens of thousands when instruments are weak.

Third, it is common for a researcher to estimate two competing models that (1) do not adjust for selection bias and (2) adjust for selection bias using parametric assumption. If the two approaches provide different results, the researcher might conclude that the second model is preferred and even necessary. Exercise 2 demonstrated that this is a poor test that provides no basis for concluding that the Heckman model is an improvement.

Along this line, this entry does not cover an important body of work that uses RCT to estimate a treatment effect (case 2 from above) and then applies the Heckman estimator to observational data to estimate ostensibly the same treatment effect. Researchers interpret the difference between the two estimates as demonstrating failure of the Heckman approach, but this comparison does not accurately quantify the bias. Rather, observational data have different coverage and consequently provide estimates for a different population (Heckman et al. 1998). Thus, the comparison does not demonstrate failure of the parametric approach to produce useful estimates of treatment effectiveness.

One final point ends this entry: This entry has treated estimation of the outcome equation as providing the estimates of scientific interest. This is not always the case. Especially when making policy recommendations, the researcher may be interested in estimation and interpretation of the selection equation, or the researcher might be interested in the distribution of treatment effects across the population, or both (Heckman and Vytlacil 2007a, b). Even when the selection equation has little scientific interest, careful thought about the selection equation is essential for applying the parametric approach. This thought exercise is a design consideration driven by substantive theory. Statistical theory alone cannot provide a magic solution.

Related Entries

- ▶ Causes
- ▶ Econometrics of Crime
- ▶ Interrupted Time Series Models
- ▶ Longitudinal Studies in Criminology
- ▶ Propensity Score Matching
- ▶ Randomized Experiments in Criminology and Criminal Justice
- ▶ Recent Perspectives on the Regression Discontinuity Design
- ▶ Sample Selection Models
- ▶ Sample Selection Problems

Recommended Reading and References

Angrist J, Pischke J (2009) Mostly harmless econometrics: an empiricist's companion. Princeton University Press, Princeton

Berk R (2005) Randomized experiments as the bronze standard. J Exp Criminol 1:417–433

Borenstein M, Hedges L, Higgins J, Rothstein H (2009) Introduction to meta-analysis. Wiley, West Sussex

Bushway S, Johnson B, Slocum L (2007) Is the magic still there? The use of the Heckman two-step correction for selection bias in criminology. J Quant Criminol 23:151–178

Cameron A, Trivedi P (1998) Regression analysis for count data. Cambridge University Press, Cambridge

Cameron A, Trivedi P (2005) Microeconometrics: methods and applications. Cambridge University Press, Cambridge

Davidson R, MacKinnon JG (1993) Estimation and inference in econometrics. Oxford University Press, New York

Gerring J (2012) Social science methodology: a unified framework, 2nd edn. Cambridge University Press, Cambridge

Greene W (2002) LIMDEP version 8.0: econometric modeling guide, volume 2. Econometric Software, Plainview

Greene W (2008) Econometric analysis, 6th edn. Prentice Hall, Upper Saddle River

Hahn J, Hausman J (2003) Weak instruments: diagnostics and cures in empirical econometrics. Am Econ Rev 93:118–125

Heckman J (1976) The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Ann Econ Soc Meas 5:475–492

Heckman J (1979) Sample selection bias as a specification error. Econometrica 47:153–161

Heckman J, Vytlacil E (2007a) Econometric evaluation of social programs, part I: causal models, structural models and econometric policy evaluations. In: Heckman J, Leamer E (eds) Handbook of econometrics volume 6B. North Holland Press, Amsterdam

Heckman J, Vytlacil E (2007b) Econometric evaluations of social programs, part II: using the marginal treatment effect to evaluate social programs, and to forecast their effects in new environments. In: Heckman J, Leamer E, Leamer E (eds) Handbook of econometrics volume 6B. North-Holland Press, Amsterdam

Heckman J, Smith J, Clements N (1997) Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts. Rev Econ Stud 64:487–535

Heckman J, Ichimura H, Smith J, Todd P (1998) Characterizing selection bias using experimental data. Econometrica 66(5):1017–1098

Imbens G, Wooldridge J (2009) Recent developments in the econometrics of program evaluation. J Econ Lit 47(1):5–86

Little R, Rubin D (2002) Statistical analysis with missing data, 2nd edn. Wiley, Hobokan

Maddala G (1983) Limited-dependent and qualitative variables in econometrics. Cambridge University Press, Cambridge

Pearl J (2000) Causality: models, reasoning and inferences. Cambridge University Press, Cambridge

Rhodes W (2010) Estimating treatment effects and predicting recidivism for community supervision using survival analysis with instrumental variables. J Quant Criminol 26:391–413

Rhodes W, Pelissier B, Gaes G, Saylor B, Camp S, Wallace S (2001) Alternative solutions to the problem of selection bias in an analysis of federal residential drug treatment programs. Eval Rev 25:331–369

Rhodes W, Kling R, Johnson P (2007) Using booking data to model drug use arrest rates: a preliminary to estimating the prevalence of chronic drug users. J Quant Criminol 23(1):1–22

Rosenbaum P (2002) Observational studies, 2nd edn. Springer, New York

Sampson R (2010) Gold standard myths: observations on the experimental turn in quantitative criminology. J Quant Criminol 26:489–500

Schaefer J (1997) Analysis of incomplete missing data. Chapman & HALL/CRC, Boca Raton

Stock J, Wright J, Yogo M (2002) A survey of weak instruments and weak identification in generalized methods of moments. J Bus Econ Stud 20:518–529

van den Berg G (2008) Competing risks models. In: Durlauf S, Blume L (eds) The new palgrave dictionary of economics, 2nd edn. Palgrave, MacMillan

Vella F (1998) Estimating models with sample selection bias: a survey. J Hum Resour 33(1):127–169

Parole

▶ Desistance and Supervision

▶ Early Release from Prison

▶ Front Door and Backdoor Sentencing

▶ Mentally Disordered Offenders Under Community Supervision

▶ Probation and Community Sanctions

Parole and Prisoner Reentry in the United States

Joan Petersilia
Stanford Law School, Stanford University,
Stanford, CA, USA

Overview

The US prison and parole population declined in 2009 for the first time in 30 years, perhaps indicating the beginning of the end of America's long commitment to mass incarceration. But current prison population declines may well be reversed if we do not do a better job of planning for the reentry of prisoners who have finished their sentences. At release, many prisoners are unable to find jobs and suitable housing. Some will be legally barred from voting, receiving public assistance, obtaining a driver's license, or retaining custody of their children. If the past is any lesson, many (upward of two-thirds) eventually return to crime and prison, where the cycle begins again.

Faced with these realities, what can we do? Fortunately, there is much that *can* be done. We must focus attention on parole policies and practices. Parole – which refers to both a release mechanism and a method of community supervision – is central to crime control and prisoner reintegration. If parole is effective, dangerous offenders remain in prison, and those who are released are better prepared for reentry. If parole resources are misdirected, community safety is threatened as prisoners return home with few resources and little surveillance. Which scenario proves true will depend on realistic expectations and an understanding of parole's history, current operations, needs of the parole population, and "what works" in reentry programming.

Fundamentals of Parole and Reentry

Definition and Functions of Parole

People often confuse probation and parole. *Probation* is a judge's sentence that allows a convicted offender to continue to live in the community after criminal conviction, with restrictions on activities and with supervision for the duration of the sentence. *Parole* refers to offenders who have spent time in prison and are released to complete the remainder of their sentence under community supervision. Parole is usually granted from authorities in the correctional system (i.e., a parole board), since responsibility for offenders passes from the judicial system to the correctional system upon imprisonment. Parolees are technically still in state custody; they have merely been granted the privilege of living in the community instead of prison. If parolees or probationers violate the rules of their release, they can be returned to incarceration.

Inmates are released from prison mandatorily or discretionarily. *Mandatory release* is release after a specified period of time, as required by law, and occurs in jurisdictions using determinate sentencing. In determinate sentencing, the offender is given a set amount of time to serve by the court, although these sentencing structures may still incorporate a degree of discretion. Determinate sentencing eliminates parole boards, although the exact requirements vary by state. *Discretionary release* is at the paroling authority's discretion, within boundaries established by the sentence and by law.

In those states that permit discretionary release, state laws give parole boards the authority to change, within certain limits, the length of a sentence that is actually served. Parole officials may also change the conditions under which convicted offenders are supervised and may release offenders from prison to supervision in the community or to an outside facility. Parole authorities can also issue warrants revoking parole and reincarcerating offenders who violate parole conditions. For jurisdictions with determinate sentencing and no discretion for the timing of release, the paroling authority may still determine conditions of release. They thus can have a direct effect on prison management. For example, they can increase the number of prisoners required to be on post-prison supervision, or they can decrease, by policy, the number of parole revocations returned to prison. It is this gatekeeper role that makes paroling authorities so central to current debates about prisoner reentry and prison crowding.

With parole, like probation, information about an offender is gathered and presented to a decision-making authority, and that authority has the power to release the offender under specific conditions, which are articulated in a contract signed by the offender. Contractual conditions may be standard (applicable to all parolees) or tailored to particular offenders. Standard parole conditions are similar throughout most jurisdictions and include payment of supervision fees, finding employment, not carrying weapons, reporting changes of address and employment, not committing crimes, and submitting to search by the police and parole officers. Examples of special conditions include periodic drug testing for substance abusers and registration for sex offenders and arsonists.

The principal responsibility of the parole officer is to monitor this court-imposed contract and conditions. If offenders fail to live up to their conditions, they can be revoked and returned to jail or prison to serve out the remainder of the original sentence or to serve a new sentence. Parole can be revoked for two reasons: (1) the commission of a new crime or (2) the violation of the conditions of parole (a technical violation). Technical violations pertain to behavior that is not criminal, such as the failure to refrain from alcohol use or remain employed. In either event, the violation process is rather straightforward. Given that parolees are technically still in the legal custody of the court or prison authorities, their constitutional rights are severely limited. When parole officers become aware of violations of the parole contract, they notify their supervisors, who make a recommendation to the parole authorities and can easily return a parolee to prison. Two US Supreme Court cases, *Morrissey v. Brewer*, 408 U.S. 471 (1972) and *Gagnon v. Scarpelli*, 411 U.S. 778 (1973), established minimum requirements for the revocation of parole, giving parolees some rights (such as written notice of the violation and the opportunity to confront the accuser) in revocation proceedings.

The goals of probation and parole supervision are identical: to protect the community and help rehabilitate offenders. These dual functions are referred to as the law enforcement function, which emphasizes surveillance of the offender and close control of behavior, and the social work function, which attempts to provide supportive services to meet offenders' needs. Both have always been a part of community corrections, and debating which should be of higher priority has always caused strain. Currently, the social work function has given way to the law enforcement function, and probation and parole officers are often less interested today in treating clients than in controlling their behavior. Higher risk clients, combined with shrinking resources, has caused many agents to prioritize offender monitoring. Parole agents are equipped with legal authority to carry and use firearms; to search places, persons, and property without the requirements imposed by the Fourth Amendment (i.e., the right to privacy); to order arrests without probable cause; and to confine without bail.

Monitoring parolee behavior and delivering services is managed through caseloads (the number of parolees assigned to a single parole agent). Higher-risk parolees are placed on smaller caseloads, which facilitate more intensive services and surveillance. Caseload assignment is usually based on a structured assessment of parolee risk and an assessment of the needs or problem areas that have contributed to the parolee's criminality. By scoring personal information relative to the risk of recidivism and the particular needs of the offender (i.e., a risk/need instrument), a total score is derived, which determines the particular level of parole supervision (e.g., intensive, medium, regular, administrative). Each jurisdiction has established policies that dictate the contact levels (times the officer will meet with the parolee). Officers may also contact family members or employers to inquire about the parolee's progress. The purpose of the contacts is to make sure that parolees are complying with parole conditions.

Parole and Reentry

The term *prisoner reentry* began affecting community corrections in early 2000. Unlike probation and parole, prisoner reentry is not a legal status or program, but a new conceptual framework for thinking about the processing of criminal offenders, from sentencing to parole discharge. Prisoner reentry is defined as the process of leaving incarceration and returning to society and includes all activities and programming conducted to prepare ex-convicts to return safely to the community and to live as law-abiding citizens (Petersilia 2003). Reentry is an inevitable consequence of incarceration. As Jeremy Travis (Travis 2005, p. xxi) reminds us, "reentry is not a form of supervision, like parole. Reentry is not a goal, like rehabilitation or reintegration. Reentry is not an option. Reentry reflects the iron law of imprisonment: they all come back."

Refocusing the justice system around a reentry perspective represents a fundamental paradigm shift that impacts decisions about the timing of release, the procedures for making the release decision, the preparation of the prisoner for release, supervision after release, and the linkages between in-prison and post-release activities.

Except for those who die naturally or are executed in prison, 95 % of all state prisoners will eventually leave prison (Petersilia 2003). Estimates from the Department of Justice show that nearly one-half of all state inmates will be released within 1 year and three-quarters will be released within 5 years (U.S. Government Accountability Office 2001). Parolees represent 11 % of the 7.2 million people "under correctional control" (incarcerated or on community supervision) (Glaze and Bonczar 2010). Regardless of whether ex-prisoners are on formal parole supervision, they all return home forever changed, often facing isolation, stigma, and a narrowed array of life chances.

Today's parole population is mostly male (88 %), although the number of females has risen steadily over the past decade. The average parolee is in his or her mid-thirties, with a median age of 34. For both males and females, most parolees are members of racial or ethnic minorities (37 % are black and 19 % are Hispanic or Latino) (Glaze and Bonczar 2008). The Bureau of Justice Statistics reported that 37 % of *all* adults on parole in 2007 were convicted of drug crimes (Glaze and Bonczar 2008). Many states are reconsidering the harsh drug laws passed in the 1980s and 1990s, rolling back mandatory sentences and accelerating releases for some drug offenders. The anticipated accelerated pace of drug-involved prison releases will strain parole and community resources even further.

Most of those released from prison today have serious social and medical problems. More than three-fourths of all prisoners have a history of substance abuse (one-fourth have histories of injection drug use), and one in six suffers from mental illness (Ditton 1999). Yet less than a third of exiting prisoners have received substance abuse or mental health treatment while in prison. And while some states have recently provided more funding for prison drug treatment, the dramatic increase in the prison population has resulted in a decline in the percentage of state prisoners participating in such programs that have been declining, from 25 % in the 1990s to about 10 % in 2001 (Lynch and Sabol 2001).

A significant share of the prison population also lives with an infectious disease. Two to three percent of state prisoners are HIV positive or have AIDS, a rate five times higher than that of the US population (Hammett et al. 2002). According to the Centers for Disease Control, about 25 % of *all* individuals living with HIV or AIDS in the United States had been released from a prison or jail that year (National Commission on Correctional Health Care 2002). This proportion may continue to increase as more drug offenders, many of whom engage in intravenous drug use, share needles, or trade sex for drugs, are released from incarceration. Prisoners with health problems may have a more difficult reentry process than others, as they are additionally confronted with the tasks of managing their health problems, obtaining health care, and keeping up with medications or appointments.

Few inmates have marketable skills or sufficient literacy to become gainfully employed at release. A third of all US prisoners were unemployed at their most recent arrest, and just 60 % of inmates have a GED or high school diploma (compared to 85 % of the US adult population). The National Adult Literacy Survey established that 11 % of inmates, compared with 3 % of the general population, have a learning disability and 3 % are mentally retarded (National Center for Education Statistics 1994). Again, despite evidence that inmates' literacy and job readiness has declined, fewer inmates are participating in prison education or vocational programs. Just over 25 % of all those released from prison in 2001 had participated in vocational training programs, and about a third of exiting prisoners will have participated in education programs – both figures down from the previous decade (Lynch and Sabol 2001).

In 2007, the nation spent about $47 billion in state general funds on corrections (which was 7 % of all state general fund spending), yet spending on treatment equaled just 6 % of the annual cost of housing a prisoner (Pew Center on the States 2009). The need for services for substance-abuse treatment and educational programming in prison has never been greater, but the percentage of prisoners receiving these services has declined. More punitive attitudes, combined with diminishing rehabilitation programs, mean that more inmates spend their prison time "idle." Ironically, as inmate needs have increased and in-prison programs decreased, parole supervision and community services have also decreased for most returning prisoners.

Current Issues and Controversies

The Decline of Rehabilitation and Discretionary Parole Release

The pillars of the American corrections systems – indeterminate sentencing coupled with parole release, for the purposes of offender rehabilitation – collapsed during the late 1970s and early 1980s. Attacks on indeterminate sentencing and parole release centered on three major criticisms. First, *there was little scientific evidence that parole release and supervision reduced subsequent recidivism*. Robert Martinson and his colleagues published the now-famous review of the effectiveness of correctional treatment and concluded, "With few and isolated exceptions, the rehabilitative efforts that have been reported so far have had no appreciable effect on recidivism" (1974, p. 25). Of the 289 studies they reviewed, just 25 (9 %) pertained to parole, and yet Martinson's summary was interpreted to mean that parole supervision (and all rehabilitation programs) did not work. Once rehabilitation could not be legitimated by science, there was nothing to support the "readiness for release" idea and, therefore, no role for parole boards or indeterminate sentencing.

Second, *parole and indeterminate sentencing were challenged on moral grounds as unjust and inhumane*, especially when imposed on unwilling participants. Research at the time showed there was little relationship between in-prison behaviors, participation in rehabilitation programs, and post-release recidivism. If that were true, then why base release dates on in-prison performance? Prisoners too argued that not knowing their release dates held them in "suspended animation" and contributed one more pain of imprisonment.

Third, indeterminate sentencing permitted parole authorities to use a great deal of *uncontrolled discretion in release decisions*, and these decisions often were inconsistent and discriminatory. Since parole boards had a great deal of autonomy and their decisions were not subject to outside scrutiny, critics argued that it was a hidden system of discretionary decision making that led to race and class bias in release decisions.

It seemed as if no one liked indeterminate sentencing and parole in the early 1980s, and the time was ripe for change. Crime control advocates denounced parole supervision as being largely nominal and ineffective; social welfare advocates decried the lack of meaningful and useful rehabilitation programs. A political coalition resulted, and soon incapacitation and "just deserts" replaced rehabilitation as the primary goal of American prisons.

With that changed focus, indeterminate sentencing and parole release came under serious attack, and calls for "abolishing parole" were heard in state after state. In 1975, Maine became the first state to eliminate parole. The following year, California and Indiana established determinate sentencing and abolished discretionary parole release. By 2002, 16 states had abolished discretionary parole release for nearly all offenders. In 19 other states, parole authorities had discretion over a small and decreasing number of parole-eligible inmates. Likewise, at the federal level, the Comprehensive Crime Control Act of 1984 created the US Sentencing Commission and phased out discretionary parole for most federal prisoners in 1997 (Petersilia 2003).

Proponents hoped that determinate sentencing with mandatory parole would make sentencing more consistent across offenders and offenses – and it has (Marvell and Moody 1996). It was also thought that abolishing parole would lengthen the time inmates spent behind bars. After all, parole release was widely regarded as letting them out early. Solomon et al. (2005) did find that prisoners released to post-prison supervision served nearly a year less than those released without supervision (mandatorily), but it is not clear whether the longer prison terms served were the result of discretionary parole systems or the nature of the inmate's crime or institutional misconduct.

Opponents of parole also assert that there is no evidence that placing offenders on parole supervision helps reduce their recidivism rates. A controversial study by the Urban Institute "Does Parole Supervision Work?" found no difference in the rearrest rates of offenders released from prison with and without parole supervision. After statistically controlling for the offenders' demographic characteristics and criminal histories, the researchers found that 61 % of mandatory parolees (those *without* supervision) were rearrested, compared with 57 % of discretionary (*with* supervision) parole releasees. Solomon et al. (2005, p. 37) concluded that "Parole has not contributed substantially to reduced recidivism and increased public safety." This is not to say that parole supervision *could not* reduce recidivism, only that at the aggregate level in which it was studied, there was no evidence that it reduced rearrests.

Rehabilitation Programs, the Cost of Parole Supervision, and Civil Disabilities

Although 70 % of all persons under correctional control are on probation or parole, nearly nine out of ten correctional dollars goes to funding prisons. Nationally, the 2008 average annual cost of a year in prison was $29,000. In contrast, the average annual cost of parole supervision was $2,750. In California, with the nation's largest prison and parole population, the figures are dramatic: for every dollar California spends on prisons, it spent 15 cents on parole (Pew Center on the States 2009). Parole officers complain of growing paperwork and diminishing resources devoted to treatment of parolees. Some parolees do not participate in *any* parole programming at all. It is no wonder that recidivism rates are so high. In a sense, we get what we pay for, and we have never chosen to invest sufficiently in parole or reentry programs.

In addition to few treatment resources, parolees are also subject to a number of statutory restrictions or "civil disabilities" when they return home. Many restrictions are statutory, stemming from a common-law tradition that people who are incarcerated are "civilly dead" and have lost all civil rights (Travis 2005). Their criminal record may preclude them from voting or retaining their parental rights and be grounds for divorce, and they may be barred from serving on a jury, holding public office, and owning firearms. In 11 states in 2004, ex-prisoners were permanently denied the right to vote (Manza and Uggen 2006). Employers are also increasingly forbidden from hiring parolees for certain jobs and are mandated to perform background checks for many others. The most common types of jobs with legal prohibitions against parolees are in the fields of childcare, education, security, nursing, and home health care – exactly the types of jobs that are expanding. Since the mid-1980s, the number of barred occupations has increased dramatically.

Even if a parolee is not legally barred from a particular job, research shows that ex-offenders face bleak prospects in the labor market, with the mark of a criminal record representing an important barrier to finding work (Pager 2003). More than 60 % of employers claim that they would not knowingly hire an applicant with a criminal background (Holzer et al. 2002). Overcoming the barriers that ex-offenders face in finding a job is critical to successful reintegration, since employment helps ex-prisoners be productive, take care of their families, develop valuable life skills, and strengthen their self-esteem and social connectedness. Research has also empirically established a positive link between job stability and reduced criminal offending. Lipsey's (1995) meta-analysis of nearly 400 studies found that the single most effective factor in reducing re-offending rates was employment.

Revolving Door Justice: Inmates Release, Recidivism, and Prison Return

Staying out of prison is a lot harder than getting out. The landmark BJS study of prisoners released in 15 states in 1994 found that fully two-thirds (67 %) were rearrested and just over one-half (52 %) were back in prison, serving time for a new prison sentence or for a technical violation of their release, within 3 years (Langan and Levin 2002). The "two-thirds rearrest rate" after 3 years has been documented in the United States for about 35 years, ever since Daniel Glaser conducted his classic follow-up study of prisoners in *The Effectiveness of a Prison and Parole System* (1969). The Langan and Levin (2002) study also found that certain characteristics were associated with parole failure: black parolees had higher rearrest rates, as did males in general, those with fewer prior arrests, those incarcerated for property offenses, and those who were younger at release (under 25). The failure rate for prisoners who had been previously released on parole and were now being re-released was much higher (64 %) than for prisoners being released to parole for the first time (21 %) (Travis and Lawrence 2002).

Rosenfeld et al. (2005), using data from the BJS recidivism study, calculated arrest probabilities by month for each of the 36 months after release. They adjusted the probability of arrest by subtracting out persons who were in jail, in prison, or dead during the month and therefore not eligible for arrest. Their results show that the probability of arrest declines with months out of prison and that the probability of arrest during the first month out of prison is roughly double than that during the fifteenth month. Arrest probabilities also differ by type of crime: prison releasees arrested on property or drug offenses are more likely to be arrested early in the post-release period than those arrested for violent offenses.

Clearly, the first few months after prison release represent high crime risks, but Binswanger et al. (2007) also found that death rates for new prison releases – within the first days and weeks – are much higher than for matched demographic groups in the general population. The higher prisoner-release death rates (12 times the average for the general population) were caused by high rates of homicide and drug overdoses. These results have led to calls to front-load parole services and surveillance so as to reduce these and other negative outcomes during the first 6 months after release (National Research Council 2008).

While individual recidivism studies have always shown high failure rates, until recently there was little attention paid to the overall amount of crime returning prisoners were responsible for. Rosenfeld et al. (2005) reanalyzed the BJS data and estimated that parolees accounted for 10–15 % of all violent, property, and drug arrests between 1994 and 1997 and the share of total arrests attributable to released prisoners grew as general crime rates declined during the 1990s. In 1994 the arrests of prisoners released in the previous 3 years accounted for 13 % of all arrests. By 2001 that figure had increased to more than 20 %. If former prisoners are accounting for nearly one-fifth of all the nation's arrests, then investing in prisoner reentry is unquestionably a matter of public safety. Such evidence has encouraged law enforcement organizations such as the International Association of Chiefs of Police (IACP) and the National District Attorneys Association to develop policies, training, and other tools that support effective prisoner reentry (La Vigne et al. 2006).

Parole Violators and Their Impact on Prison Populations

The BJS reports that about 50 % of released prisoners will be returned to prison in the 3 years following release – and most of them will eventually be re-released, and the revolving door process will continually be repeated. Prisoners refer to it as "doing a life term on the installment plan." The constant churning of parolees in and out of incarceration is a major contributor to the growing US population. As a percentage of all admissions to state prisons, parole violators more than doubled from 17 % in 1980 to nearly 35 % in 2006 (Sabol and Couture 2008).

In these budget-strapped times, many question the public safety benefits of violating parolees for technical violations, such as missing a meeting with a parole agent, being out after a curfew time, and, increasingly, failing a drug test. Processing admissions of parole violators takes as much time and costs as much money as processing admissions of new convictions and for offenders who often will be in prison for only a few months. Many advocates argue that technical violations should not result in a return to prison and can be handled in the community with less-drastic intermediate sanctions. The contrary view is that some violations, even of a procedural nature, are significant signs that a parolee is not respecting the terms of the parole contract and should suffer consequences. This can also be a sign that even if the current violation is minor, if left unattended, it might lead to more serious ones. Several states have begun to use parole violation decision-making instruments to respond systematically to violations. The goal of such instruments is to differentiate technical violations and the response to them by the severity and risk posed by the offender, ultimately decreasing prison admissions for technical violations and reducing disparities in parole decision making.

Responding appropriately to parolee recidivism is one way to manage correctional resources better, but preventing parolees from failing in the first place is a more proactive strategy. But that begs the all-important question of whether we can implement reentry programs to increase the odds of success.

Future Directions

No one believes that the current parole and reentry system is working, but the $64,000 question is, can we do better? Can we improve the outcomes for people returning from prison so that they are less likely to be rearrested? Fortunately, the last decade has seen an explosion of interest in parole and prisoner reentry, and the "what works?" literature has improved. After extensively reviewing all of the literature, the National Research Council (2008, p. 82) concluded, "'Nothing works' is no longer a defensible conclusion from assessment of program effects on re-entry outcomes."

What Works in Reentry Programming?

Using a variety of techniques, researchers have developed a list of principles of effective intervention and found that programs adhering to these principles significantly reduced recidivism, sometimes by as much as 30 % (Andrews and Bonta 2006; Lowenkamp et al. 2006). Petersilia (2004) summarized these principles and their applicability to reentry programs:

- Treatment services should be behavioral in nature; interventions should employ the cognitive behavioral and social learning techniques of modeling.
- Reinforcements in the program should be largely positive, not negative.
- Services should be intensive, lasting 3–12 months (depending on need), and occupy 40–70 % of the offender's time during the course of the program.
- Treatment interventions should be used primarily with higher-risk offenders, targeting their criminogenic needs (dynamic risk factors for change).
- The best strategy for discerning offender risk level is to rely on actuarial-based assessment instruments.
- Conducting interventions in the community as opposed to an institutional setting will increase treatment effectiveness.
- In terms of staffing, there is a need to match styles and modes of treatment service to the learning styles of the offender (specific responsivity).

It is impossible to know the extent to which these evidence-based principles are being used in current parole and reentry programs. Interest in prisoner reentry over the last decade has fueled the development of hundreds of programs across the United States. Some reentry programs are small and administered by community- and faith-based organizations, whereas other programs are large statewide initiatives, often administered by the state's corrections department. Some programs start inside the prison and continue into the community, whereas others begin when the prisoner has returned home. Some programs are residential, others involve day reporting, and still others involve a meeting every month or two. They rely on existing and volunteer community treatment services, primarily Alcoholics Anonymous and Narcotics Anonymous. The academic "what works" literature barely touches these programs, but these programs are delivering vital services and should not be dismissed, as they are the foot soldiers of the reentry movement.

Many other reentry programs are part of federally funded reentry initiatives implemented in recent years. The "what works" literature is influencing their design and evaluations. The largest effort to implement evidence-based reentry principles systematically is the federal government's Serious and Violent Offender Reentry Initiative (SVORI). In 2003, 69 agencies representing all 50 states received more than $110 million in federal funds to develop programs to improve the outcomes of serious and violent prisoners coming home (Lattimore et al. 2005). The federal government's SVORI goals were to improve a variety of outcomes, including family relationships, work, health, community integration, housing, and reduced crime. Programs were also encouraged to focus specifically on serious and violent offenders, or both. Most programs used a formal risk assessment tool to identify risks and needs, as recommended in the "what works" literature. Early outcome results appear promising. SVORI participants were more likely to receive services and participate in needed programs, and SVORI participants are doing better across a wide range of outcomes. Recidivism outcomes are not yet reported, but other outcomes, including housing, employment, mental health, and substance abuse, have improved. As Lattimore and Visher (2009) reported to Congress, "In most cases, the difference in outcomes between those participating in SVORI programs and the comparison subjects indicates that SVORI program participation resulted in an improvement in outcomes." It is important to note that most SVORI participants were repeat serious offenders (the males had an average of 13 prior arrests), so that even relatively small reductions in recidivism rates may be quite cost-beneficial.

In sum, it is possible to reduce offender recidivism in a cost-effective manner. The answer lies in investing in reentry programs that incorporate "what works" principles, targeting those programs on specific offenders who can most benefit, and continually evaluating and revising program models as the science accumulates.

Conclusions and Future Research

Parole and prisoner reentry issues have captured the nation's attention, and the US parole system is entering another chapter in its history. Attacks on parole and community-based programs have virtually disappeared and have been replaced with calls for investing *more*, not less, in parole and prisoner reentry. Where there was little scholarly attention paid to parole just 10 years ago, the volume and visibility of work around parole and prisoner reentry issues has grown to such an extent that it is now commonly referred to as a full-fledged movement (Petersilia 2009). And the reentry momentum is likely to continue. Voters are tiring of our dependence on mass incarceration as a response to crime and more willing to embrace a more balanced system of punishment.

For all that has been learned about parole and reentry in recent years, a number of important research questions remain. We now know a good deal about the needs of returning prisoners and have some evidence about the services they receive. However, much less research exists on the effectiveness of particular approaches to delivering services, and only a small number of studies have used rigorous scientific methods to test promising practices. The highest priority for future research is more credible program evaluations. To date, no studies have analyzed the differences among low-, medium-, and high-risk offenders using an experimental design. A recent panel of the National Research Council (2008, p. 82) concluded that while there is a great deal of experiential and practitioner knowledge with regard to the apparent efficacy of reentry programs, "the challenge now is to subject these promising practices to rigorously designed evaluations." Rigorous program evaluations should accompany every significant reentry initiative, and outcomes for these studies should focus not solely on recidivism but also on other behavioral outcomes, such as sobriety, stability in housing and employment, and attachment to families and communities. Our studies must also disaggregate the characteristics of the offender population so that we can design better programs for specialized populations, such as women, the elderly, sex offenders, or the mentally ill. A higher proportion of parolees in the future are likely to be composed of one of these distinct population groups, and we know very little about how to deliver services to meet their specific needs.

We also need a better understanding of how neighborhood characteristics affect the reintegration of offenders. Hipp et al. (2010) recently reported that the presence of more social service providers nearby (within two miles) led to lower recidivism rates and that this protective effect was particularly strong for African American and Latino parolees. They also found that parolees living in neighborhoods with higher levels of concentrated disadvantage experienced greater rates of recidivism, even after taking into account the individual characteristics of these parolees. This study highlights the importance of social context for successful reintegration. Research must move beyond simple statistical models that attempt to explain parolees' return to prison solely as a function of the parolees' background and behavior, since the characteristics of their agent, supervising agency, and community may be significant predictors as well.

Finally, researchers must study successful parolees in order to uncover the factors that encourage offenders to shift from formal social controls to informal social controls. Ultimately, parolees who make it shift from being accountable to programs and criminal justice agencies (e.g., police, parole) to being accountable to more informal social controls (e.g., families, neighbors). Ideally, formal criminal justice sanctions should act as "presses" to increase social bonds to law-abiding family members and conventional institutions. Ethnographic studies can identify how to promote positive social bonds between ex-convicts and community members.

We are now witnessing the start of a new ideological pendulum shift in US punishment policy. A declining economy that is pressuring states to reduce incarceration, combined with a growing body of evidence identifying effective reentry programs, has created a window of opportunity for change. Parole and prisoner reentry may well serve as a major conceptual framework for reorganizing criminal justice policy in the twenty-first century.

Related Entries

- ▶ Aging Correctional Populations
- ▶ Aging Prison Population: Factors to Consider
- ▶ Crime and the Racial Composition of Communities
- ▶ Criminal Careers and Public Policy Responses
- ▶ Deterrent Effect of Imprisonment
- ▶ Drug Abuse and Alcohol Dependence Among Inmates
- ▶ Drug Courts
- ▶ Health Issues in Prison Reentry Models
- ▶ History of Corrections
- ▶ History of Probation and Parole in the United States
- ▶ Intensive Probation and Parole
- ▶ New Penology
- ▶ Offender Change in Treatment
- ▶ Probation and Parole Practices
- ▶ Probation Officer Decision-Making
- ▶ Punishment as Rehabilitation
- ▶ Re-entry
- ▶ Reentry Courts
- ▶ Sex Offenders and Criminal Policy

Recommended Reading and References

Andrews D, Bonta J (2006) The psychology of criminal conduct, 4th edn. Anderson, Cincinnati

Binswanger I, Stern M, Deyo R, Heagerty P, Cheadle A, Elmore J, Koepsell T (2007) Release from rison – a high risk of death for former inmates. N Engl J Med 356:157–165

Ditton P (1999) Mental health and treatment of inmates and probationers. Bureau of Justice Statistics, Washington DC

Glaze LE, Bonczar T (2010) Probation and Parole in the United States, 2009. Bureau of Justice Statistics, Washington, DC

Glaze LE, Bonczar TP (2008) Probation and Parole in the United States, 2007 – statistical tables. Bureau of Justice Statistics. U.S. Department of Justice, Washington, DC

Hammett T, Harmon P, Rhodes W (2002) The burden of infectious disease among inmates and releasees from correctional facilities. In: National Commission on Correctional Healthcare (ed) The health status of soon-to-be-released inmates. National Institute of Justice, Washington DC

Hipp J, Petersilia J, Turner S (2010) Parolee recidivism in California: the effect of neighborhood context and social service agency characteristics. Criminology 48(4):947–979

Holzer H, Raphael S, Stoll M (2002) Can employers play a more positive role in prisoner reentry? Urban Institute, Washington, DC

Langan P, Levin D (2002) Recidivism of prisoners released in 1994. Bureau of Justice Statistics, Washington, DC

Lattimore PK, Visher CA, Winterfield L, Lindquist C, Brumbaugh S (2005) Implementation of prisoner reentry programs: findings from the serious and violent offender reentry initiative multi-site evaluation. Justice Res Policy 7(2):87–109

Lattimore PK, Visher CA (2009) The multi-site evaluation of the serious and violent offender reentry initiative. The Urban Institute. Available at http://www.urban.org/publications/412075.html

Lipsey MW (1995) What do we learn from 400 research studies on the effectiveness of treatment with Juvenile Delinquency? In: McQuire J (ed) What works: reducing reoffending. Wiley, West Sussex

Lowenkamp CT, Latessa EJ, Smith P (2006) Does correctional program quality really matter? The impact of adhering to the principles of effective intervention. Criminol Public Policy 5(3):575–594

Lynch JP, Sabol WJ (2001) Prisoner reentry in perspective. Urban Institute, Washington, DC

Manza J, Uggen C (2006) Locked out: felon disenfranchisement and American democracy. Oxford University Press, New York

Martinson R (1974) What works? Questions and answers about prison reform. Public Interest 35(Spring):22–35

Marvell TB, Moody CE (1996) Determinate sentencing and abolishing parole: the long-term impacts on prisons and crime. Criminology 34(1):107–128

National Center for Education Statistics (1994) Literacy behind prison walls. U.S. Department of Education, Washington, DC

National Commission on Correctional Health Care (2002) The health status of soon-to-be-released inmates: a report to congress. National Institute of Justice, Washington, DC

National Research Council (2008) Parole, desistance from crime, and community integration. National Academy of Sciences, Washington, DC

Pager D (2003) The mark of a criminal record. Am J Sociol 108(5):937–75

Petersilia J (2003) When prisoners come home: parole and prisoner reentry. Oxford University Press, New York

Petersilia J (2004) What works in prisoner reentry? Reviewing and questioning the evidence. Fed Probat 68(2):4–8

Petersilia J (2009) Transformation in prisoner reentry: what a difference a decade makes. Afterword in When Prisoners Come Home by Petersilia, Oxford University Press

Pew Center on the States (2009) One in 31: the long reach of American corrections. The Pew Charitable Trusts, Washington, DC

Rosenfeld R, Wallman J, Fornango R (2005) The contribution of ex-prisoners to crime rates. In: Travis J, Visher CA (eds) Prisoner reentry and crime in America. Cambridge University Press, New York

Sabol WJ, Couture H (2008) Prison inmates at midyear 2007. Bureau of Justice Statistics, Washington

Solomon AL, Kachnowski V, Bhati A (2005) Does parole work? Analyzing the impact of postprison supervision on rearrest outcomes. Urban Institute, Washington, DC

Travis J (2005) But they all come back: facing the challenges of prisoner reentry. The Urban Institute, Washington, DC

Travis J, Lawrence S (2002) Beyond the prison gates: the state of parole in America. Urban Institute, Washington, DC

U.S. Government Accounting Office (2001) Prisoner Releases: trends and information on reintegration programs, Washington, DC

Vigne L, Nancy G, Solomon A, Beckman K, Dedel K (2006) Prisoner reentry and community policing: strategies for enhancing public safety. Urban Institute, Washington, DC

Pathways to Delinquency

Rolf Loeber
Department of Psychiatry, University of
Pittsburgh, Pittsburgh, PA, USA

Overview

Modern criminological theories aim to explain crime, but they do so with a different emphasis on pathways to crime and ranges of explanatory factors (e.g., Farrington 2005; Thornberry and Krohn 2003). Although theories of antisocial and delinquent behavior often have several factors in common (e.g., juveniles' relationships with parents and peers), they differ in their relative emphasis on domains, settings, and details of explanatory factors, and the ways that these factors are developmentally interrelated (see above sources and chapters in Farrington 2005, and in Lahey et al. 2003). The models also much differ in the extent to which they are two-dimensional (i.e., on one surface) on paper or can be visualized in a better ways. This entry concerns advances in the conceptualization and visualization of development of behavioral stages, leading to serious forms of delinquency and explanatory factors that account for individual differences in how far individuals progress in their offending.

Issues in the Construction of Models to Explain Delinquency and Crime

Criminological theories (Thornberry and Krohn 2003) almost always share three themes with the goals of explaining: (a) antisocial and delinquent behavior over the life-course, particularly in terms of prevalence, frequency, and severity of delinquent acts; (b) individual differences in antisocial/delinquent behavior and developmental changes in these differences; and (c) nonoffending or low-level offending. The key is for researchers, practitioners, and policy makers to understand how the combination of (a), (b), and (c) explains why some individuals and not others become serious property offenders or violent offenders and understand this both on the population level (e.g., all youth in a particular city) and on the individual level (i.e., a particular juvenile minor offender at risk of becoming a serious offender). Along these lines, there is a need for better life-course models that incorporate the development and *accumulation* of risk *and* protective factors.

To date, the integration of these components in a model has taken place almost exclusively through the following methods. Firstly, model specification is usually visually illustrated in a two-dimensional plane of variables in boxes (representing usually a single outcome of delinquency, such as violence or property crime) and their interrelationships are expressed by arrows between the boxes. Secondly, virtually, all model specification rests on multivariate statistical tests of the hypothesized relationships between independent measures also called risk factors (the boxes) and an outcome, such as serious delinquency. Risk factors are defined here as events or conditions that are associated with an increased probability of serious forms of delinquency, distinguishing risk factors from protective factors, which are factors associated with lowered risk of delinquency.

However, there are at least five limitations to the models tested to date (Loeber et al. 2006). Firstly, the empirical results of the multivariate models have been enormously diverse. The second disadvantage is that most models have not delineated developmental pathways to serious forms of antisocial behavior (such as violence) by the specification of developmental antecedents to such serious outcomes (e.g., nonviolent forms of aggression). Third, current models do not specify differences between individuals in their exposure to an *accumulation* of specific risk (or protective) factors with development. A fourth disadvantage has to do with the fact that the majority of studies focus only on risk factors as a way of explaining antisocial behavior and either neglect to consider protective factors or the combined effect of risk and protective factors on antisocial outcomes.

We argue that criminological theories can be enhanced in at least two other ways, firstly by incorporating past development, and secondly, by incorporating possible future development. In the first category are differences in individuals' *past history of offending* (and history of behavior problems that are precursors to delinquency), *and* differences in the *individuals' exposure to earlier risk factors* (e.g., prenatal exposure to toxins or child abuse during the preschool period) *and protective factors* (e.g., a good relationship with an adult). The second understudied area has to do with two aspects of individuals' futures: firstly, the possible course of subsequent delinquent development (or pathways) as examined by longitudinal studies of youth who have been followed up into adulthood; and secondly, expected exposure to risk and protective factors based on the observation of such factors in older populations of youth. Examples of "new" risk factors that individuals may be exposed to during adolescence include: gang membership, violent victimization, and alcohol intoxication. In summary, we propose that developmental models should benefit from knowledge of pathways and exposure to risk and protective factors based on past studies which can then provide a framework in which to place the development of specific individuals. In that sense, population data can describe both the stages in pathways to serious delinquency (the latter being the outer, more severe boundary) and the gradual accumulation of risk and protective factors that can maximally take place in populations (another instance of an outer boundary). Knowledge of these two types of outer boundaries will make it possible to describe the past history of an individual's antisocial and delinquent behavior, his/her past exposure to risk and protective factors, and possible future exposure to ongoing or new risk and protective factors.

One of the key advantages of such a model is that, in contrast to current statistical models, it is likely to provide valuable information about choices and targets of intervention. This is very much in line with contemporary principles of prevention and intervention, with its focus on behavioral development and an emphasis on reducing exposure to risk factors while at the same time increasing exposure to protective factors (Pollard et al. 1999).

We argue that the complex of factors described above is easier to understand by policy makers, informed lay-persons, and practitioners when visualized in a model of a three-, rather than a two-dimensional space or model specified as a multivariate formulae. To construct such a visual model, it is necessary to review five key topics: (1) pathways representing individuals' escalation toward serious property offenses and violence; (2) differences between individuals in their exposure to risk and protective factors at birth (called start-up factors) and subsequently during childhood, adolescence, and early adulthood; (3) developmental accumulation of risk and protective factors over time; (4) the existence of a dose–response relationship between the number of risk/protective factors and later serious delinquency and violence; and (5) additive and compensatory effects of protective and risk factors. Finally, we integrate these five aspects into a single, three-dimensional cumulative, developmental model.

Escalation and the Formulation of Developmental Pathways

Youth who commit acts of violence or serious theft rarely do so *de novo*. Instead, many criminologists, child psychologists, and psychiatrists agree that the majority of youth who commit acts of violence or serious theft practiced less serious forms of delinquency earlier in life (e.g., Le Blanc 2002; Warr 2002). Thus, the onset of violence is usually preceded by a history of escalation in the severity of aggression, which often, but not always, starts in childhood (Moffitt 1993).

The findings of several studies focus on a single escalation pathway. However Loeber et al. (1993) challenged this and tested a triple pathway model (Fig. 1), which better fitted the data. The three pathways are as follows: (a) An *Authority Conflict Pathway* prior to the age of 12, that starts with stubborn behavior, and has defiance as a second stage, and authority avoidance (e.g., truancy) as a third stage; (b) *A Covert Pathway* prior to age 15 which starts with minor covert acts, has property damage as a second stage, and moderate to serious delinquency as a third stage; and (c) *An Overt Pathway* that starts with minor aggression, has physical fighting as a second stage, and more severe violence as a third stage. It is possible that within this last stage, homicide constitutes a separate, most serious component (Loeber et al. 1993). Recent longitudinal research on the Pittsburgh Youth Study (Loeber et al. 2011) indicates that 94 % of later homicide offenders have displayed violence earlier in life. The pathways are hierarchical in that those who have advanced to the most serious behavior in each of the pathways have usually displayed persistent problem behavior characteristics at the earlier stages in each pathway. By a process of selection, increasingly smaller groups of youth become at risk for the more serious behaviors. The pathways are also related to neighborhoods. For example, a higher

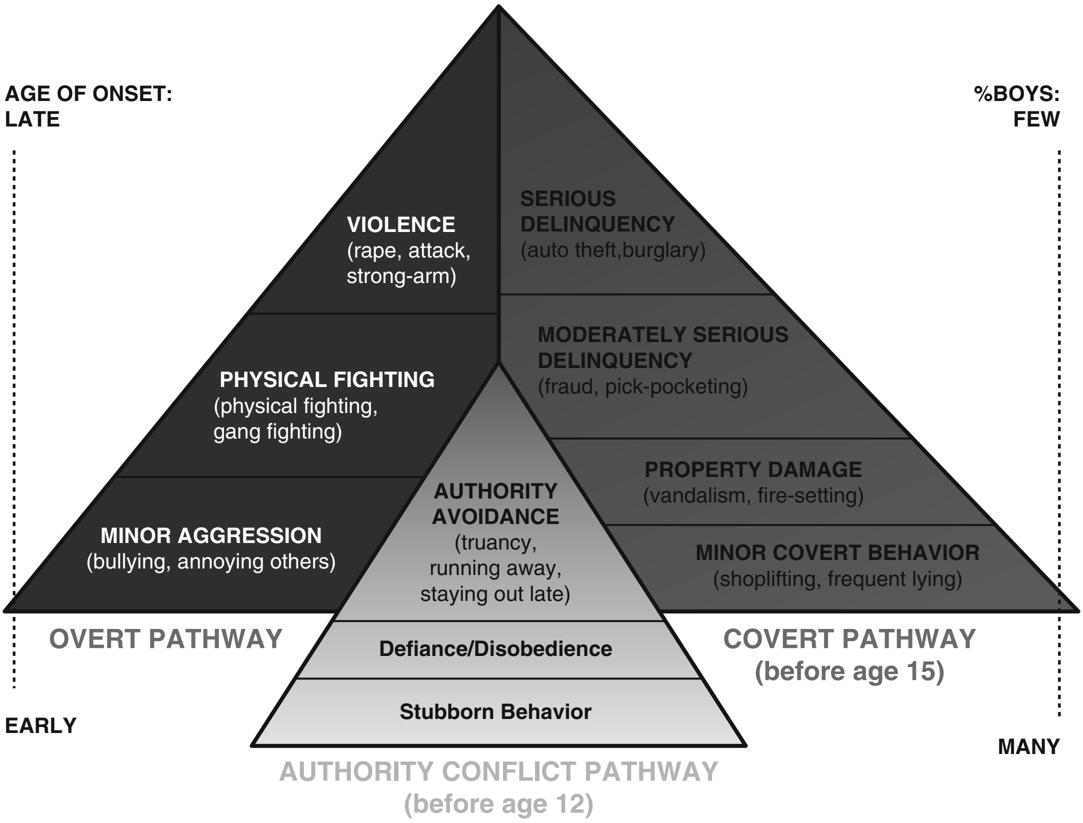
AGE OF ONSET:
LATE
EARLY
%BOYS:
FEW
MANY
VIOLENCE
(rape, attack, strong-arm)
PHYSICAL FIGHTING
(physical fighting, gang fighting)
MINOR AGGRESSION
(bullying, annoying others)
SERIOUS DELINQUENCY
(auto theft,burglary)
MODERATELY SERIOUS DELINQUENCY
(fraud, pick-pocketing)
PROPERTY DAMAGE
(vandalism, fire-setting)
MINOR COVERT BEHAVIOR
(shoplifting, frequent lying)
AUTHORITY AVOIDANCE
(truancy, running away, staying out late)
Defiance/Disobedience
Stubborn Behavior
OVERT PATHWAY
COVERT PATHWAY
(before age 15)
AUTHORITY CONFLICT PATHWAY
(before age 12)

Pathways to Delinquency, Fig. 1 Developmental pathways to serious delinquency and violence (Loeber et al. 1993)

percentage of youth in the most disadvantaged neighborhoods escalates from minor aggression to violence than youth living in more advantaged neighborhoods (Loeber and Wikström 1993). The three-pathway model has been subsequently replicated across two other samples in the Pittsburgh Youth Study. In addition, several other studies have replicated the pathways model (Loeber et al. 1999; Tolan et al. 2000).

Risk Factors

The majority of theories about the causes of crime and antisocial behavior are generally based on risk factors alone, although different terms for such factors have been used (e.g., Gottfredson and Hirshi 1990; Moffitt 1993). Risk factors can be distinguished according to the different domains and contexts in which they operate. Overall, researchers and theoreticians agree (with some variation) to distinguish risk factors in the following domains: individual, family, and peer group. In addition, some risk factors are associated with children's exposure to risk factors in specific contexts, including characteristics of the school attended and the neighborhood in which the children reside or spend their time (Howell 2003; Loeber and Farrington 1998; Wikström and Loeber 2000). Data from prediction studies indicate that risk factors from each of the domains (individual, family, peers, schools, and neighborhoods) contributes to the explanation of why some individuals and not others progress from minor problem behavior such as bullying, to physical fighting, to violence (e.g., Farrington 1997; Loeber et al. 2011). Individuals' exposure to an accumulation of risk factors in multiple domains rather than in a single domain heightens the probability of later adverse outcomes. For example, Deater-Deckard et al. (1998) find that each of four domains (child characteristics, sociocultural, parent, and peer experiences) contributes to the prediction of externalizing problems.

Many researchers have been relatively silent about the developmental saliency and priority of different domains of risk factors (individual, family, peer, etc.), other than noting that children's exposure to family factors usually precedes their exposure to peer factors. As children develop, the configuration of problem behaviors changes, with children becoming more mobile and becoming also more active in selecting settings with their own risk and protective factors, when children become exposed to risk and protective factors of peers, school, and, eventually, work. Thus, with development, there is a reconfiguration of risk domains with new domains and new risk factors being introduced over time.

Developmentally graded Risk Factors. Among the risk factors, some are thought to be present at birth or become manifest shortly afterward, while other risk factors emerge later in individuals' lives. Before discussing each category of risk factors, our overall strategy is to view risk factors from a developmental angle in that children's exposure to different risk factors is gradual, that with development, children's exposure to different risk domains increases, and that many risk factors persist over time and, consequently, are "stacked" over time. It would be clearly untenable to insist that all the risk and protective factors mentioned in the preceding section are present in juveniles' lives from a very young age onward. Instead, it is much more likely that selection processes operate in which certain categories of children and youth are incrementally exposed to certain risk and protective factors as they grow into adults (see below).

Loeber et al. (2006) reviewed the emergence of risk factors during childhood and adolescence, starting with risk factors present at birth, and those emerging during the following periods: preschool, elementary school, middle/high school, and early adulthood. Emergence is defined as the probable earliest exposure of children to risk factors. Systematic knowledge in this respect is still wanting, and as a result, we have made estimates (knowing that some of the onset of risk factors, such as in the case of poor executive functioning, partly depends on the state of measurement in that area). Note that emergence should not be taken too literally and may apply differently from one child to another. However, we maintain that an accumulation of risk factors may take place over many years, and that risk factors at birth (e.g., low IQ, negative emotional temperament) may be followed by exposure to "new" risk factors emerging during early childhood, such as language problems, which in turn may be followed by other novel risk factors emerging during middle to late childhood, such as poor social skills, or poor academic achievement. During adolescence, novel risk factors may consist of heavy substance use or delinquent victimization. These examples are risk factors in the individual domain, but a similar accumulation of new risk factors may occur in other domains as well, such as the family, peers, schools, and neighborhood (including work). It should be noted that, to some extent, risk factors emerging in one life period may persist across another life period and that even when they are time-limited (e.g., child abuse), they may influence the probability of serious delinquency in a cumulative manner.

Loeber et al. (2006) concluded that the total number of risk factors associated with antisocial behavior and delinquency to which children *can* be exposed is about equally divided at or close to birth (15), emerging in the preschool years (12), and emerging in the elementary school age years (18). Much lower numbers of risk factors probably first appear during the middle and high school years (5). Thus, the most salient risk window of children's exposure to risk factors is prior to adolescence. This must be true for early-onset cases of antisocial behavior and delinquency. Even within the category of early-onset cases, the model is flexible in that we assume that none of the early risk factors are necessary (i.e., absolutely required) for the emergence of antisocial behavior and delinquency. In that sense, there can be some substitution of risk factors that apply to some and not to other early-onset cases. In summary, the evidence suggests that individual risk factors continue to increase in numbers and in different configurations after childhood and constitute a substantial proportion of all known risk factors after early childhood.

Dose–Response Relationship Between Risk Factors and Delinquency

Studies agree that the higher the number of risk factors, the greater the likelihood that individuals will be affected by a negative outcome or other deviant behavior such as serious delinquency. This association is usually called a dose–response relationship and has been demonstrated for the full range of indicators of antisocial behavior: sociopathy, externalizing problems (Deater-Deckard et al. 1998), conduct problems (Fergusson and Woodward 2000), serious delinquency (Smith et al. 1995), and violence (Loeber et al. 2011).

As an example (Fig. 2), a prediction index in the Pittsburgh Youth Study was constructed on the basis of the 11 strongest predictors of violence (Loeber et al. 1993): truancy, low school motivation, onset of delinquency before age ten, cruelty to people, depressed mood, physical aggression, and callous/unemotional behavior, low family SES, family on welfare, high parental stress, and bad (i.e., disadvantaged) neighborhood (parent report). Figure 2 also shows that the dose–response relationship applies to the number of predictors and the probability of later violence (Odds Ratio (OR) = 6.0 for four or more risk factors). Remarkably, the range of probabilities in the Pittsburgh data is from 3 % at zero risk factors to 100 % at 11 or more risk factors. In summary, these results (and those of several other studies) indicate a robust association between the number of risk factors and the probability of later violence. Remarkably, the association holds even when slightly different risk factors are measured from study to study.

Stability. Implied is that some risk factors may persist over time. Investigations of risk factors often ignore the fact that such factors, similar to the outcomes that they purportedly predict, may vary in their temporal stability over time. Yet few studies have addressed the issue of stability of risk factors. An exception is Sameroff et al. (1998), who found that environmental risk factors correlate .77 over a 5-year period (ages 13–18), which is about the same level of stability as intelligence. Loeber et al. (2000) found that the year-to-year stability coefficients for the interactions between parents and their boys between ages 6 and 18 average.66 for bad parent–child relationship,.70 for poor parent–child communication, and lower for poor supervision (.56) and physical punishment (.46). Examination of absolute stability reveals that this is higher for bad relationships and poor communication, but decreases with age for poor supervision and physical punishment. It should be noted, however, that despite level changes in risk factors with development, many of the known risk factors continue to predict later negative outcomes, including delinquency, at different ages of children (e.g., Loeber and Farrington 1998; Loeber et al. 2008).

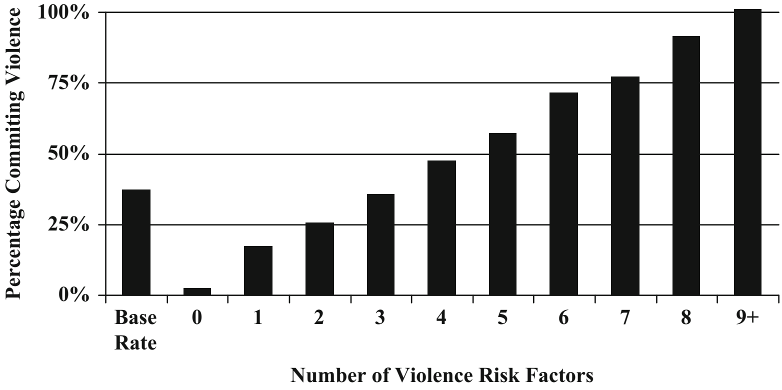

Pathways to Delinquency, Fig. 2 Proportion of boys committing violent offenses for different levels of risk (Based on Loeber et al. 2005)

It is plausible that the presence of some risk factors set in motion a cascade of other risk factors. For example, the toxic teratogenic effect of maternal smoking may increase the probability of several other risk factors in the offspring, including poor executive functioning, poor academic achievement in school, low motivation to attend school, truancy, and, subsequently, delinquency. As another example, the presence of several risk factors (e.g., parents' exposure to enduring stress and parents' substance abuse) may increase the chance of antisocial behavior in the offspring at home, which in turn, may set the scene for heightened risk of problem behavior outside of the home. Yet, as Howell (2003) points out, it remains to be seen when and how risk factors operate in a "sequential causal chain" (p. 110). It is clearer that the prevalence of different risk factors is higher in disadvantaged compared to advantaged neighborhoods (Stouthamer-Loeber et al. 2002; Wikström and Loeber 2000).

Protective Factors

Most criminology and psychopathology studies with a public health slant have routinely neglected the study of protective factors (Lösel and Bender 2003). Protective factors are associated with the likelihood of reduced antisocial behavior/delinquency and/or increased positive outcomes, including positive adjustment and positive mental health.

Developmentally graded Protective Factors. There is an increasing body of research on protective factors (see review by Lösel and Bender 2003), but that body is still miniscule in comparison to the number of publications on risk factors pertaining to antisocial behavior and delinquency. In particular, we know very little about the developmental aspects of the accumulation of protective factors through the life-course. However, the framework proposed by Hawkins and colleagues (Catalano and Hawkins 1996) and ourselves (Stouthamer-Loeber et al. 2002) is unusual in criminology because each postulates changes within protective factors to explain individual differences in the development of offending. At the risk of being speculative, we postulate that some protective factors, like risk factors, may be present at birth, but that other factors emerge during the first decades of life. Loeber et al. (2006) shows protective factors mentioned in the research literature (or factors that are probably protective according to our view) as being relevant to antisocial behavior and delinquency (based on Loeber and Farrington in press; Lösel and Bender 2003; Pollard et al. 1999; Sampson and Laub 1990; Stouthamer-Loeber et al. 2002, 2004). Most of the protective factors are the inverse of risk factors, in that many, but not all, of the protective factors represent the other pole of risk factors. However, we do not advocate that protective factors are the complete inverse of risk factors (see Stouthamer-Loeber et al. 2004 for details). Research shows that there are several protective factors that do not have a risk factor equivalent (Loeber et al. 2008). Examples of these unique protective factors are low Attention-Deficit Hyperactivity Disorder symptoms, and low anxiety. As with the development of risk factors, we assume that there is a developmentally graded emergence of protective factors (Loeber et al. 2006). Some of the protective factors are thought to be present at birth (e.g., moderate to high intelligence, absence of prenatal exposure to toxins), while other factors become manifest later. It is also thought that there is a high degree of temporal stability of protective factors, although the data on this have not yet been reported. In the Pittsburgh Youth Study (Stouthamer-Loeber et al. 2004), accountability, perceived likelihood of getting caught for delinquent acts, and low physical punishment are recurring factors that are associated with desistance in persistent serious delinquency at different ages. Relatively little is known about increases in protective factors with development. Stouthamer-Loeber (2002) found that the proportion of significant associations between protective factors and delinquency increases with age. This conclusion, however, is limited by the reason that the age comparisons are based on different cohorts, and, therefore, require replication to be proven.

It is likely that protective factors, such as risk factors, are correlated, in that one protective factor may set the scene for another protective factor to emerge. For instance, association with prosocial peers may generate new prosocial behavior, which, once adopted by an individual child, may further decrease the probability that that child will engage in delinquent acts. Hawkins and his colleagues (2003) are among the few who have specified prosocial pathways in the delinquency research that can be conceptualized as a string of protective factors. They postulate that opportunities for prosocial behavior set the scene for interpersonal involvement, which in turn produces rewards, improving bonding to others and promoting belief in a moral order. Research shows that the prevalence of different protective factors is highest in the most advantaged neighborhoods (Stouthamer-Loeber et al. 2002; Wikström and Loeber 2000).

Inverse Dose–Response Relationship Between Protective Factors and Delinquency

Is there an inverse dose–response relationship between the number of protective factors and later deviance (the higher the number of protective factors, the lower the probability of deviance)? There are very few research studies addressing this question. Outside the area of delinquency, Sameroff et al. (1998) found that the higher the number of protective factors, the lower is the probability of behavior problems, with the results mirroring the results of the risk factors: "The more risk factors, the worse the outcomes; the more protective factors, the better the outcomes" (p. 172). Similar results have been reported by Wikström and Loeber (2000), thus lending support to the notion that the accumulation of protective factors in juveniles' lives counters the likelihood of later delinquent involvement.

Do Protective Factors Offset the Impact of Risk Factors?

What is the evidence that protective factors buffer the risk for later antisocial behavior and delinquency? Stouthamer-Loeber et al. (2002) showed that even at the level of the number of risk or protective domains (child behaviors, child attitudes, school, peers, family, demographic characteristics), the sum of risk and protective domains (where the number of protective domains is deducted from the risk domains) linearly predicts persistent serious delinquency (Fig. 3). The higher the number of risk domains to which a youth is exposed and the lower his/her exposure to protective domains, the greater the likelihood of later persistent delinquency. Conversely, the higher the number of protective domains and the lower the number of risk domains, the smaller the likelihood of later persistent serious delinquency. This was replicated for the youngest and oldest samples; for different degrees of neighborhood advantage and disadvantage, for specific risk and protective factors, and for early and late onset forms of delinquency (Loeber et al. 2008; Stouthamer-Loeber et al. 2002; Wikström and Loeber 2000). Not all researchers, however, have found that the relationship between risk and protective factors is linear. For example, Pollard et al. (1999), studying such diverse outcomes as substance use, arrest, and attack to hurt, found a nonlinear effect with the effect of risk factors greater as the level of risk increased. However, it is plausible that once neighborhood context is taken into account, nonlinear associations occur. For example, Wikström and Loeber (2000) report that, "the overwhelming majority of boys with a high risk score were involved in serious offending regardless of the socioeconomic context of their neighbourhood" (p. 1130). However, neighborhood matters greatly for those with a balanced score of risk and protective factors.

Toward a Three-Dimensional Model of Developmental Pathways and Developmentally Graded Risk and Protective Factors

To summarize firstly, we have reviewed developmental pathways from minor to serious delinquent behavior. Secondly, we have discussed

Pathways to Delinquency, Fig. 3 The higher the number of risk domains (and the lower the number of protective domains), the higher the risk of later persistent serious delinquency (Stouthamer-Loeber et al. 2002)

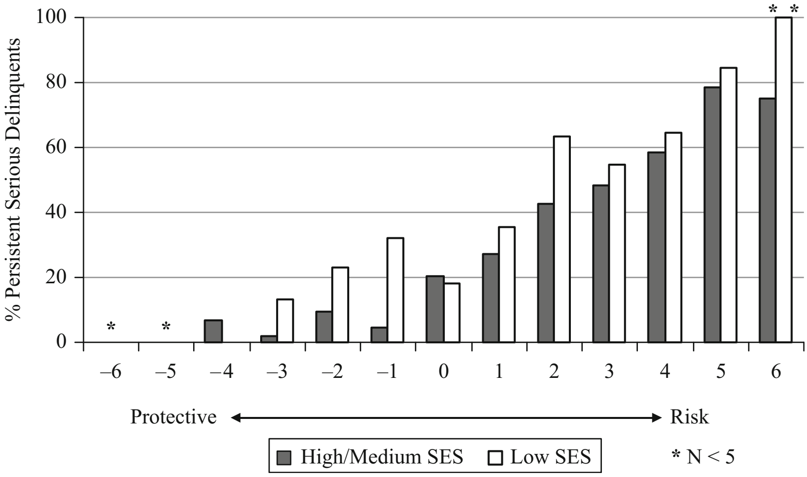

developmentally graded, cumulative onset of risk factors and developmentally graded, cumulative onset of protective factors (Loeber et al. 2006). Thirdly, we have seen that there are dose–response relationships between the number of risk factors and the probability of later delinquency and violence, and an inverse dose–response relationship between the number of protective factors and the reduced probability of later delinquency and violence. Finally, we have seen that knowledge of the proportion of risk and protective factors is more important than knowledge of either. The question then is how to bring these different strands of evidence together? At this point, we want to take a rather unconventional route. In science, the verification of models is usually accomplished by mathematical formulae. We are not aware of formulae that describe the different aspects that we have stressed. An alternative or rather a step toward such a composite model is to display the above interrelationships visually in a three-dimensional model.

Before discussing this model, we will present each component. Figure 4 shows a version of two of the pathways depicted in Fig. 1, which has been reoriented into the horizontal plane. These two pathways are overt and covert behavior (leaving aside for the moment the authority conflict pathway), with the overt pathway triangle superimposed on the covert pathway triangle. Figure 5 represents the developmentally graded,

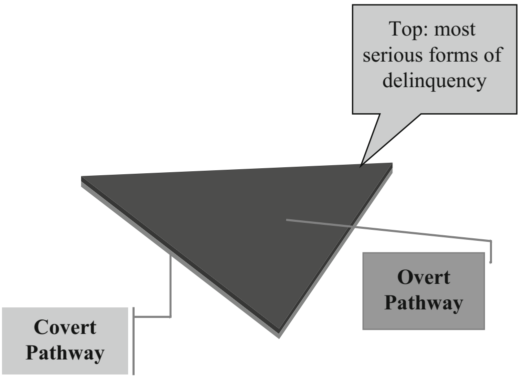

Pathways to Delinquency, Fig. 4 Pathway model flipped horizontally (overt pathway on the top and the cover pathway underneath)

cumulative onset of risk factors into a number of vertical panels, each of the same shape, but with ribs of different sizes, representing different degrees of individuals' exposure to risk factors, so that they would fit on the horizontal pathway picture (Fig. 4). In the central and highest risk rib, the number of risk factors present at birth is thought to be higher than in the adjoining ribs, the accumulation of risk factors is highest, and the length of the rib is shown to be longest to represent the higher degree of severity of delinquency. In contrast, the adjoining smaller ribs represent fewer risk factors at birth and a smaller accumulation of later risk factors. Figure 6 represents the distribution of "start-up" risk factors at birth in the form of a long triangle.

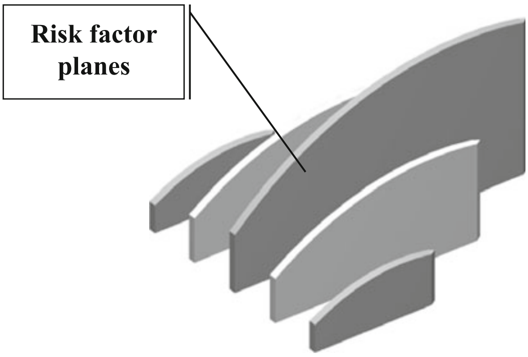

Pathways to Delinquency, Fig. 5 Accumulation of risk factors at different levels of accumulation

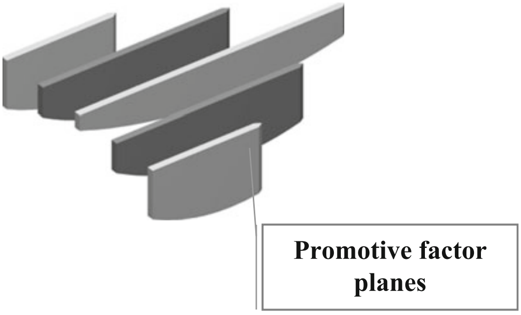

Pathways to Delinquency, Fig. 7 Accumulation of promotive factors at different levels of accumulation

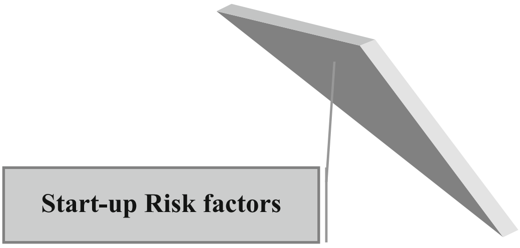

Pathways to Delinquency, Fig. 6 Distribution of risk factors at birth

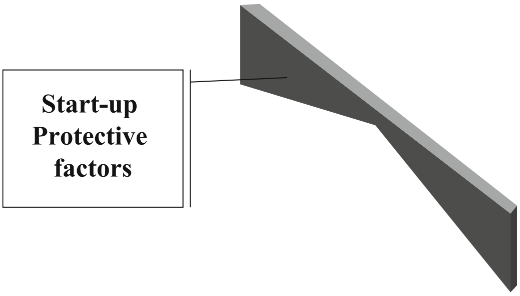

Pathways to Delinquency, Fig. 8 Distribution of promotive factors at birth (start-up promotive factors)

The central top of that triangle indicates a category of children with the highest number of risk factors at birth and at the extremes are those with the fewest number of risk factors at that time. Figure 6 represents developmental rather than chronological time.

Figure 7 represents ribs of the accumulation of protective factors in an inverse way to the risk factors. However, the central rib is the least tall because those who advance to the most serious acts are usually exposed to the lowest number of protective factors. In contrast, individuals who do not advance to serious delinquency outcomes tend to start out in life with and accumulate more protective factors than those in the central rib. Figure 8, in the form of half a butterfly, may clarify the “start-up” protective factors which are thought to be lowest in the center and highest at the extremes.

Finally, Fig. 9 shows how each of the above elements is put together. Individuals who advance to the most serious delinquent acts on the horizontal pathways (the flat, double triangle) are thought to be exposed to the highest number of risk factors over time *and* the lowest number of protective factors over time. At birth, they also tend to have the highest number of risk factors (see the start-up risk panel) and the lowest number of protective factors (see the start-up protective factors). To clarify this cumulative, developmental model, Fig. 10 shows how it looks from the front, while Fig. 11 shows how it looks from underneath.

Relevance of the Model for Assessments and Interventions

What is the relevance of the cumulative, developmental model for assessment and the evaluation of preventive and remedial interventions?

Assessments. Advances have been made in the past decades in the area of more precise assessment of juveniles’ problem behavior in terms of

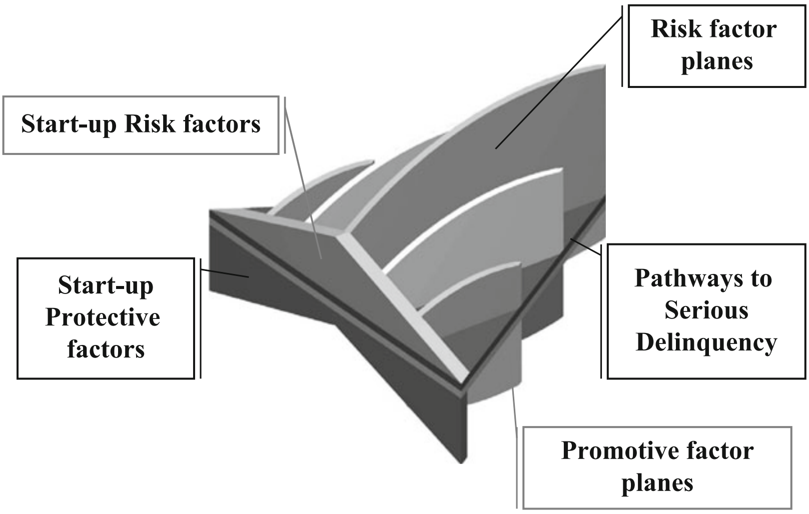

Pathways to Delinquency, Fig. 9 The full model of developmental pathways as a function of the accumulation of risk and promotive factors (shown from above)

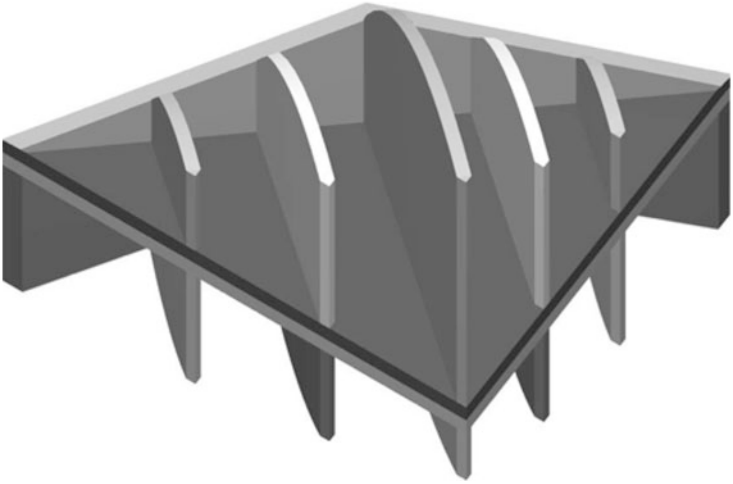

Pathways to Delinquency, Fig. 10 The full model of developmental pathways as a function of the accumulation of risk and promotive factors (shown from the front)

the types of behavior, their severity, and prognostic validity. In addition, screening instruments are now available to go beyond the range of problem behaviors of juveniles that also include the past history of risk factors to which juveniles have been exposed. Based on the information provided in this entry, we argue however that assessments can probably benefit from two other components: (a) an appraisal of current and past protective factors to which juveniles are exposed, because it is the mixture of risk and protective effects that appears most crucial in determining the future risk of serious offending as well as the probability of full desistance or lower-level offending. We also argue that the next generation of risk assessment devices could potentially also benefit from (b) an appraisal of juveniles' expected future exposure to risk and protective factors based on knowledge from longitudinal survey studies. Specifically, assessments of developmentally graded exposure to risk factors, based on longitudinal survey studies, could be employed to calculate the probability that juveniles will be exposed in the future to risk and protective factors typically emerging for that age-group.

One might argue that cross-sectional studies contain such information. There are several reasons why this is not the case. Firstly, in contrast to cross-sectional studies, longitudinal studies have

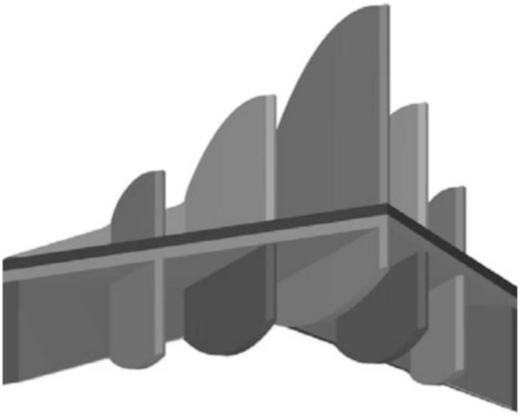

Pathways to Delinquency, Fig. 11 The full model of developmental pathways as a function of the accumulation of risk and promotive factors (shown from the bottom)

the power to specify selection processes determining which groups of youth are most likely to be exposed to future risk and protective factors. The second reason has to do with protective factors associated with stable nonoffending or a de-escalation from serious to minor or offending-nonoffending. Since juveniles' engagement in delinquency varies from year to year, it is important to establish which protective factors are associated with persistent nonoffending or, alternatively, with a stable de-escalation in the severity of offending. Longitudinal studies, rather than cross-sectional studies, are the best methods by which to identify protective factors.

Preventive and Remedial Interventions. We agree with Howell (2003) that a developmental approach to offending needs a reconceptualization of when interventions, preventive or remedial, can best take place. He has been a primary proponent of the influential Comprehensive Strategy that addresses the full continuum of antisocial development with two foci for implementing change: prevention and early intervention, and graduated sanctions. We believe that this orientation, which is based on knowledge of risk and protective factors (called protective factors in this case), can also greatly benefit from the assessment of future risk through developmentally graded expectations of exposure to risk and protective factors. We argue that this is especially important because of the fact that even the most effective interventions still do not reduce the risk of recidivism by more than 40 % (Lipsey and Wilson 1998) and that the reoffending rate of high-risk populations of youth still remains extraordinary high (Loeber and Farrington 1998).

It should be understood that knowledge of risk and protective factors (Loeber et al. 2006) does not mean that we have available the tools to change each of them, and of course some are not malleable (e.g., single parenthood). However, we hope that our list of the risk and protective factors and our model will stimulate the development of other forms of interventions that can augment the overall efficacy of interventions currently available. It is our impression that most treatment programs lack components that focus on increasing or strengthening protective factors. We foresee that interventions that mobilize or enhance protective factors (together with the reduction of risk factors) eventually are likely to improve treatment efficacy and, possibly, the maintenance of treatment effects. We agree, however, with Pollard et al. (1999) that a sole focus on interventions enhancing protective factors only, because of the compensatory function of risk and protective factors, is likely to be inadequate. Another possible beneficial result of our model specification is the further individualization of interventions tailored to an individual's past, present, and possible future exposure to risk and protective factors.

Finally, we should stress that our design of a three-dimensional model of antisocial behavior has certain limitations. Firstly, we are mostly focusing on the cumulative processes of risk and protective factors rather than the specifics of mediation and moderation within and between risk and protective factors. Also, we do not focus on possible reciprocal processes between risk and protective factors and negative or positive child behaviors, or the fact that as children's mobility in the community increases with age, they become more able to select settings and situations in which novel risk and protective factors may present themselves. We are also aware that our model may have different components depending on the subject population of study (girls vs. boys, different ethnic groups), but we think it is generic enough to serve as a model to explain a range of forms of maladjustment other than serious delinquency. Moreover, the model depicts escalation in the severity of antisocial behavior better than de-escalation to lower severity levels, but de-escalation could perhaps be incorporated in the next iteration of the model. Also, there is a need to transfer the basic mathematical properties of our model into future model testing. It is clear that many tasks await, and we hope that this text will inspire others to improve on where we are at this point.

Related Entries

- ▶ Age-Crime Curve
- ▶ Career Criminals and Criminological Theory
- ▶ Desistance from Crime
- ▶ Identification Issues in Life Course Criminology
- ▶ Longitudinal Studies in Criminology
- ▶ Optimizing Longitudinal Studies in Offending
- ▶ Risk Assessment, Classification, and Prediction

Recommended Reading and References

Catalano RF, Hawkins JD (1996) The social development model: a theory of antisocial behavior. In: Hawkins JD (ed) Delinquency and crime: current theories. Cambridge University Press, New York, pp 149–197

Deater-Deckard K, Dodge KA, Bates JE, Pettit GS (1998) Multiple risk factors in the development of externalizing behavior problems: group and individual differences. Dev Psychopathol 10:469–493

Farrington DP (1997) Early prediction of violent and nonviolent youthful offending. Eur J Crim Policy Res 5:51–66

Farrington DP (2005) Integrated developmental and life-course theories of offending, vol 14, Advances in criminological theory. Transaction, New Brunswick

Fergusson DM, Woodward LJ (2000) Educational, psychosocial, and sexual outcomes of girls with conduct problems in early adolescence. J Child Psychol Psychiatry 41:779–792

Gottfredson MR, Hirshi T (1990) A general theory of crime. Stanford University Press, Stanford

Hawkins JD, Smith BH, Hill KG, Kosterman R, Catalano RF, Abbott RD (2003) Understanding and preventing crime and violence: findings from the Seattle social development project. In: Thornberry TP, Krohn MD (eds) Taking stock of delinquency: an overview of findings from contemporary longitudinal studies. Kluwer Academic/Plenum Publishers, New York, pp 255–312

Howell JC (2003) Preventing and reducing juvenile delinquency, a comprehensive framework. Sage Publications, Inc., Thousand Oaks

Lahey BB, Moffitt TE, Caspi A (2003) Causes of conduct disorder and juvenile delinquency. Guilford, New York

Le Blanc M (2002) The offending cycle, escalation and de-escalation in delinquent behavior: a challenge for criminology. Intern J Comp Appl Crim Justice 26:53–83

Lipsey MW, Wilson DB (1998) Effective intervention for serious juvenile offenders: a synthesis of research. In: Loeber R, Farrington DP (eds) Serious and violent juvenile offenders: risk factors and successful interventions. Sage Publications, Thousand Oaks, pp 313–346

Loeber R, Drinkwater M, Yin Y, Anderson SJ, Schmidt LC, Crawford A (2000) Stability of family interactions from ages 6 to 18. J Abnormal Child Psychol 28:353–369

Loeber R, Farrington DP (1998) Serious and violent juvenile offenders: risk factors and successful interventions. Sage, Thousand Oaks

Loeber R, Farrington DP, Stouthamer-Loeber M, White HR (2008) Violence and serious theft: development and prediction from childhood to adulthood. New York: Routledge

Loeber R, Farrington DP (in press) Protective and risk factors pertaining to violence in four longitudinal studies. Am J Prev Med 43:S24–S27.

Loeber R, Wikström PO (1993) Individual pathways to crime in different types of neighborhood. In: Farrington DP, Sampson RJ, Wikström PO (eds) Integrating individual and ecological aspects of crime. Liber Forlag, Stockholm, pp 169–204

Loeber R, Keenan K, Lahey BB, Green SM, Thomas C (1993) Evidence for developmentally based diagnoses of oppositional defiant disorder and conduct disorder. J Abnorm Child Psychol 21:377–410

Loeber R, Keenan K, Zhang Q (1997) Boys' experimentation and persistence in developmental pathways toward serious delinquency. J Child Fam Stud 6:321–357

Loeber R, Wei E, Stouthamer-Loeber M, Huizinga D, Thornberry T (1999) Behavioral antecedents to serious and violent juvenile offending: joint analyses from the Denver youth survey, Pittsburgh youth study, and the Rochester development study. Stud Crime Crime Prev 8:245–263

Loeber R, Slot NW, Stouthamer-Loeber M (2006) A three-dimensional, cumulative developmental model of serious delinquency. In: Wikström POH, Sampson R (eds) The explanation of crime: contexts and mechanisms. Cambridge University Press, Cambridge, pp 153–194

Loeber R, Farrington DP, Stallings R (2011) The Pittsburgh youth study. In: Loeber R, Farrington DP (eds) Young homicide offenders and victims. Springer-Verlag, New York, pp 19–36

Lösel F, Bender D (2003) Protective factors and resilence. In: Farrington DP, Coid J (eds) Early prevention of adult anti-social behaviour. Cambridge University Press, Cambridge

Loeber R, Wung P, Keenan K, Giroux B, Stouthamer-Loeber M, Van Kammen WB, Maughan B (1993) Developmental pathways in disruptive child behavior. Dev Psychopathol 5:101–132

Moffitt TE (1993) Adolescence-limited and life-cycle-persistent antisocial behavior: a developmental taxonomy. Psychol Rev 100:674–701

Pollard JA, Hawkins JD, Arthur MW (1999) Risk and protection: are both necessary to understand diverse behavioral outcomes in adolescence? Soc Work Res 23:145–158

Sameroff AJ, Bartko WT, Baldwin A, Baldwin C, Seifer R (1998) Family and social influences on the development of child competence. In: Lewis M, Feiring C (eds) Families, risk, and competence. Lawrence Erlbaum Associates, Mahwah, pp 161–185

Sampson RJ, Laub JH (1990) Crime and deviance over the life course: the salience of adult social bonds. Am Sociol Rev 55:609–627

Smith C, Lizotte AJ, Thornberry TP, Krohn MD (1995) Resilient youth: identifying factors that prevent high risk youth from engaging in delinquency and drug use. In: Hagan J (ed) Current perspectives on aging and the life cycle, vol 4, Delinquency and disrepute in the life course. JAI, Greenwich, pp 217–247

Stouthamer-Loeber M, Loeber R, Wei E, Farrington DP, Wikström PO (2002) Risk and protective effects in the explanation of persistent serious delinquency in boys. J Consult Clin Psychol 70:111–123

Stouthamer-Loeber M, Wei E, Loeber R, Masten AF (2004) Desistance from persistent serious delinquency in the transition to adulthood. Dev Psychopathol 16:897–918

Thornberry TP, Krohn MD (2003) Taking stock of delinquency: an overview of findings from contemporary longitudinal studies. Kluwer Academic/Plenum Publishers, New York

Tolan PH, Gorman-Smith D, Loeber R (2000) Developmental timing of onsets of disruptive behaviors and later delinquency of inner-city youth. J Child Fam Stud 9:203–230

Warr M (2002) Companions in crime: the social aspects of criminal conduct. Cambridge University Press, Cambridge, UK

West DJ (1982) Delinquency: its roots, careers and prospects. Heinemann, London

Wikström PO, Loeber R (2000) Do disadvantaged neighborhoods cause well-adjusted children to become adolescent delinquents? A study of male juvenile serious offending, risk and protective factors, and neighborhood context. Criminology 38:1109–1141

Patrol

► Preventive Patrol

Patterns in Youthful Offending

► Specialization in Juvenile Offending

Peer Effects

► Network Analysis in Criminology

Penal Abolitionism

Vincenzo Ruggiero
School of Law, Middlesex University,
London, UK

Overview

Abolitionism is not only a strategy or a set of demands aimed at the reduction (or suppression) of custody, it is also a perspective, a philosophy, an approach. Penal abolitionism challenges conventional definitions of crime and the law, while defying official views of the meanings and effects of punishment. This entry examines the origin, philosophy, and achievements of penal abolitionism, collecting and discussing the key arguments and views expressed by its leading proponents. It sets off with reference to Tocqueville's view that prisoners lose their right to humanity and Foucault's remarks, during his visit to Attica, about the "consumption and elimination" of the excluded enacted by prison institutions. It then discusses the degree to which the abolitionist stance can be deemed utopian and presents the classic arguments which constitute the components of that stance. Three areas are addressed: crime, the law, and punishment, namely, the key areas in which abolitionism elaborates its own philosophy while distancing itself from conventional sensibilities and, at the same time, responding to its mainstream critics.

Introduction

Abolitionism fights against the "curious mechanism of circular elimination" identified by Foucault during his visit to Attica, where he noted that society excludes certain specific types of its members, and by sending them to prison, it breaks them up, crushes them, physically eliminates them. Then

> The prison eliminates them by "freeing" them and sending them back to society; the state in which they come out insures that society will eliminate them once again, sending them to prison. Attica is a machine for elimination, a form of prodigious stomach, a kidney that consumes, destroys, breaks up and then rejects, and that consumes in order to eliminate what it has already eliminated. (Simon 1991, p. 27)

Similarly, observing the American prison system, de Tocqueville (1956) noted that offenders released from custody remained among humans, but they lost their rights to humanity; people fled them as impure, and even those who believed in their innocence abandoned them. Once released, they could go in peace, with their live generously left to them, but a life worse than death.

Against this strategy of elimination, the primary challenge for penal abolitionists is "to construct a political language and theoretical discourse that disarticulates crime from punishment" (Davis 2008, p. 3). However, such strategy is accompanied by alternative conceptualizations of crime, critical analyses of law, and radical thinking around the very nature, function and philosophy of punishment.

According to one of the most representative figures of this multifaceted school of thought, abolitionism is a manifestation of

> the general human urge to do away with and to struggle against those phenomena or institutions of a social, political or religious nature that at a given time are considered to be unjust, wrong or unfair. (Bianchi 1991, p. 9)

The "general human urge" in the definition provided by Bianchi is the same urge that gave strength to the ancestors of contemporary abolitionists, namely, the women and men who fought against slavery, and later the campaigners who battled, and continue to do so, for the abolition of the death penalty. But the battle also continues on other fronts. Abolitionism posits that the criminal justice system as a whole constitutes a social problem in itself. Some forms of abolitionism, however, are already in place. Surely, there is an abolitionist stance in the proposition that the state-centralized administration of penal justice should be replaced by decentralized forms of autonomous conflict regulation. Echoes of abolitionism are also heard when critics warn that, if we want to reform the penal system, we have to start a process of collective conversion away from the traditional and conventional grammar that characterizes it. It should be reiterated, therefore, that abolitionism does not advocate the immediate suppression of all custodial institutions, a program which, on the other hand, would enjoy the support of many a critic who find the conditions of imprisonment all over the world shameful. Abolitionism t is not a "program," but an approach, a perspective, a methodology, and most of all a way of seeing. By observing the concepts and the cultural matrix which give shape to abolitionism, one may discover that this school of thought sits very comfortably within Western popular culture, which guides the conduct of many and which many could mobilize to justify their conduct.

Utopia?

If the activist ancestors of abolitionism are the men and women who fought against slavery and the death penalty, their philosophical ancestors are harder to identify with precision. But let us start with a general, preliminary characterization. Abolitionism, I would submit, is grounded in a variety of social philosophies which are primarily concerned with discussing processes of social development that can be viewed as pathological or as "misdevelopments" (Honneth 2007).

Among these philosophies are those expressing the view that societies should support a rich plurality of activities, each valuable in its own right, and that each person should be treated as an end, not as a mere means to the ends of others. This Aristotelian view would deny "that a society can be flourishing as a whole when some members are doing extremely badly" (Nussbaum 2000, p. 106).

There is nothing utopian in attempts to redress "remediable injustices": abolitionists do not pursue perfect justice, rather, they aim at enhancing justice. Their focus on social interactions rather than institutions, on precise settings in which people live rather than official norms and extraneous professionals, locates them in a specific political and philosophical tradition. According to a distinction suggested by Amartya Sen (2009), there are contractarian approaches and comparative approaches to the idea of justice. The former establish general, universal, principles of justice and are concerned with setting up "just institutions." For such institutions to function, total compliance of people's behavior is required. The latter assess the different ways in which people lead their lives, actually behave and interact. A contractarian approach is described by Sen as "transcendental institutionalism," in that it searches for the ideal institutions capable of forging a perfectly just society. By contrast, a comparative approach is led by the search for social arrangements that satisfy people in their concrete collective life.

> When people across the world agitate for *more* global justice – and I emphasise here the comparative word "more" – they are not clamouring for some kind of "minimal humanitarianism". Nor are they agitating for a "perfectly just" world society, but merely for the elimination of some outrageously unjust arrangements to enhance global justice. (Sen 2009, p. 26)

Comparison entails information, which in turn presupposes proximity to the actors involved in the process of forging ideas of justice. Abolitionists propend for this type of approach.

Some forms of human suffering may be unavoidable, and perhaps cannot be remedied in some particular place or at some particular time. Correspondingly, however, there are forms of suffering that are needless at particular times and places. According to Gouldner (1975), it is the sociologist's job to give special attention to the latter. Penal suffering is avoidable, particularly if its ineffectiveness can be proven, and the prisoners' standpoint deserves to be heard not because they have any special virtue and not because they alone live in a world of suffering.

> A sociology of the underdog is justified because, and to the extent that, his suffering is less likely to be known and because – by the very reason of his being underdog – the extent and character of his suffering are likely to contain much that is avoidable. (Gouldner 1975, p. 37)

Similarly, Nils Christie often reminds us that all he wants to do is to reduce the amount of suffering in the world: "I have never known someone who wants to increase it" (personal communication). Christie may have been particularly fortunate in his encounters, but his views and those of other abolitionists sit comfortably with a variety of insights found in sociology, politics, and philosophy.

Problematic Acts

Are crimes exceptional events? Louk Hulsman's reply to this rhetorical question takes the form of an excursus into the criminal justice system as a "special" set of procedures addressed to a "not so special" category of individuals. We are accustomed, he argues, to regarding the criminal justice system as a rational apparatus specifically devised to respond to crime and to control or neutralize its effects. We are also accustomed to interpreting crimes as exceptional events, "events which differ to an important extent from other events which are not defined as criminal" (Hulsman 1986, p. 63). In conventional perceptions, criminal conduct is at odds with the average conduct in that it deviates from the type of interactions supposedly established by the majority. Such deviation, statistically exceptional, is thus deemed to trigger the special responses offered by official institutions: special events justify the special nature of the reaction against them.

According to abolitionist analysis, conducts classified as criminal are only a small proportion of a variety of similar conducts which escape criminalization. On the one hand, therefore, only certain types of events are selected as specific, undeniable, self-evident, representatives of what we understand as crime. On the other hand, the events selected are brought together in a discreet category, as if they constituted a homogenous group of conducts. "Within the concept of criminality a broad range of situations are linked together. Most of these, however, have separate properties and no common denominator" (Hulsman 1986, p. 65). What do violence within the family and street violence have in common? Do shoplifting, drug selling, and armed robbery belong to the same rubric of behavior? What makes dangerous driving similar to fencing? And political violence to pollution? Neither the motivation of those involved nor the techniques required, let alone the consequences of those specific conducts, displays a precise set of common components. If we are led to analyze them by referring such conducts to a common logical and moral framework, it is because the framework we adopt is inspired by the criminal justice system: "All these events have in common is that the criminal justice system is authorised to take action against them" (Hulsman 1986, p. 65).

This is not denying that human interactions may cause considerable degrees of suffering to those involved: abolitionism would only question why some types of suffering mobilize the intervention of institutional agencies while others do not. Moreover, in an imaginary scale of personal hardship, the distress caused by some interactions of a criminal nature would not score particularly high when compared to those of a quotidian, apparently prosaic, nature. "Matrimonial difficulties, difficulties between parents and children, serious difficulties at work and housing problems will, as a rule, be experienced as more serious." In brief, there is nothing that intrinsically distinguishes, for those directly involved, criminal events from a variety of other unpleasant events. Most episodes causing hardship are normally dealt with informally by the individuals affected, at times with the mediatory participation of people or groups belonging to the community in which such episodes occur. Many conducts that generate serious harm do not elicit responses from the criminal justice system, while those involved in problematic situations mostly attempt to find solutions within the context in which they arise: the family, the group of friends or neighbors, and the work place. All this means, Hulsman concludes, "that there is no 'ontological reality' of crime" (Hulsman 1986, p. 66). Some examples may provide support to this conclusion.

A murder occurs in a town inhabited by some 300,000 people, and the alleged murderer is a university lecturer. A few days earlier, the lecturer had delivered a public speech and no one in the audience noticed anything unusual in the speech and in the speaker. The event is incomprehensible and the investigators feel the necessity to entrust a team of psychiatrists with the explanation of the mystery. Let us now imagine another murder occurring 200 years before in a small town.

> If we had lived in that small town at that time, we would probably have found it ridiculous to call in an expert of the mind to explain why the killer had killed. Ridiculous because we all knew why he killed. After all, we would not have been surprised, and we would have agreed among friends that this was exactly what we all might have expected all along, (Nils Christie, personal communication)

The difference between the experience and perception of the two murders resides in the nature and amount of information the residents share about one another. So many people live in a middle-sized modern town that it is impossible to know them all. In addition, life is organized in ways that only allow us to hold a superficial or segmented knowledge of other beings. We have a narrow basis for predicting behavior outside the specific group to which we belong.

Another example comes from the Norwegian valleys, where the traditional institutional figure of the *lensmann* is still operative. A local sheriff of sort, the *lensmann* performs a variety of civil tasks and his/her ability to function is dependent upon the support received by fellow citizens. The *lensmann* may direct auctions, help mothers get money from absent fathers, and deal with crime. When Christie interviewed one such institutional figure, he was told that there was no crime in the district. But during the course of the interview, several things happened that testified to the contrary.

> The telephone rang, a lady had lost her purse: the *lensmann* asked his assistant to drive down to the close-by café; the purse was found and brought back to the lady. So was the young man who was using the purse. He happened to be the lady's son. A report came in on breaking and entering into a store of weapons. The *lensmann* jumped into his car, drove far up into the mountains in the direction of the store, met a car high up there, stopped the car, found Ole drunk as usual, with a carload of guns he had stolen just to irritate his farther. He brought Ole home and took the guns to a safer depot. What a cliff-hanger story lost for the mass-media! Helicopters and anti-terrorist police might have been engaged in the crime-hunt of the century. Now, it was Ole. An old story of misery and family quarrels. (Christie 1982, pp. 73–74)

These cases and the one following show that definitions of crime are based on "ignorance" or lack of familiarity with the events at hand.

A man administers some drugs to his wife and causes her death by suffocation. He then writes to the police, saying that he has committed murder and that he is going to drown himself. The letter reaches the police 2 days after the man's suicide. The flat where the coupled lived is visited by the officers, who find the dead woman. She had Alzheimer's and, before killing her, the man had looked after her with deep love. "To some this is a story of Romeo and Juliet. To others, it is one of plain murder" (Christie 2004, p. 1). From numerous stories and situations such as this, abolitionists derive the argument that crime is a shallow concept, too imprecise to allow the subtle distinctions and understandings we need. The alternative concept of "trouble" may be a better starting point, trouble being something which is widely experienced and simultaneously caused by people interacting. Moreover, trouble requires responses, solutions, lest it causes harm and suffering. "The danger is too hastily to define troubles as crime." The next step is, therefore, to examine in detail which acts are beneficial and which harmful.

> Then follows an analysis of these acts perceived as bad – a classificatory scheme with categories as irritation, unpleasantness, disgust, sin - and then, but only as one among so many alternatives – crime. When crime is the last concept in the line, it is easier to raise the analytical question: What are the social conditions for acts to be designated as crimes? Crime does not exist. Only acts exist, acts often given different meanings within various social frameworks. (Christie 2004, pp. 2–3)

Let us take a final example of how social frameworks determine the meaning of acts. A man carrying a bag-load of beer cans arrives in a park. He starts drinking one can after the other, and because he acts "funny," he attracts the attention of the children playing nearby who surround him and enjoy the show. Suddenly, the man goes behind a bush with the intention of relieving himself and the children, who are carried away by the entertaining situation, follow him. It is summer, and many people are enjoying the sun from their windows or balconies in two adjacent blocks of flats. Christie names the two buildings as, respectively, The House of Perfection and The House of Turbulence. The former was built by an efficient and reliable company, the tenants moved in smoothly, on the agreed day, and to their satisfaction they found everything they expected perfectly in place. The latter was built by a company that went bankrupt, causing delays and distress among buyers, who organized themselves with the aim of claiming some form of compensation. The inhabitants of The House of Turbulence, in other words, knew each other fairly well, had experienced some form of collective action, and identified some common understanding of their needs. Now, a situation in which a man relieving himself while surrounded by children lends itself to highly different interpretations. For the inhabitants of the House of Turbulence, the case was clear.

> The man is Peter, son of Anna. He had an accident when he was little, behaves generally a bit strangely, but is as kind as the midsummer night is long. When he drinks too much, it is just to phone his family and someone comes to take him home. (Christie 2004, p. 5)

On the contrary, in The House of Perfection nobody knows Peter: he is just a man urinating in front of children. The police must be called, because this is a case of indecent exposure. In Christie's analysis, the inhabitants of this house never had a chance to get to know each other, nor were they ever forced to establish forms of cooperation with their neighbors. In conclusion, a "limited amount of knowledge inside a social system opens the possibility of giving an act the meaning of crime." (Christie 2004, p. 5).

Substantial Law

Moving on to the abolitionist analysis of the law, it is necessary to start from a general, conventional premise. Juridical equality may be described as everybody's right to mobilize state institutions for the protection and safeguard of their wellbeing. In this perspective, it is, therefore, the right to mutual coercion. Disrespect for the liberty of others amounts to the denial of their freedom. The state intervenes to deny that denial and restore the initial situation. Coercion is therefore legitimate in that it denies an act which has denied the freedom of others. Abolitionists would retort that such arguments might only be suitable for societies in which equal access to the law is complemented with equal access to resources. Their critique of the law and the criminal justice system is addressed to the iniquitous societies in which we live, while critical responses are also provided to other key assumptions, namely, that the law addresses individual rather than collective actors, and that liability and non-liability can be scientifically assessed.

Institutional intervention into problematic situations aims to obliterate an ethics of shared responsibility for conduct while affirming an ethics of individual responsibility. In this way, institutions can establish their monopoly over the power to punish or pardon, namely, their right to destroy or "repair" the individual responsible. The weakening of any networks contesting this right, be they neighborhood, religious, political, or cultural aggregations of people claiming their own right to deal with problematic situations and responsibility, is crucial for that monopoly to be accomplished. When the official power to punish or pardon solidifies, penal sanctions as chief characteristics of the criminal law emerge.

> It is only then that crime – as defined by the state itself – becomes possible. It is only then that the continuity of crime in some form becomes fundamental to the very existence of the state. And this form is determined by what the state chooses to sanction penally. (Kennedy 1976, p. 63)

The abolitionist stance, here, echoes some aspects of conflict theory. Conflict theorists propose that law in the books and law in action favor the interests of special groups, and that, therefore, rather than being concerned with the explanation of crime, theory should engage in explaining the function of criminal law (Vold 1958; Chambliss 1969, 1975; Turk 1969; Quinney 1974; Akers 1997). Similarly, abolitionists do not ask why some people commit crime, but rather why some acts are defined as criminal? As succinctly stated by Quinney (1970, p. 18): "criminal definitions describe behaviour that conflicts with the interests of the segments of society that have the power to shape public policy." From another, though adjacent, perspective, the criminal law itself is described as devoid of rationality and replete with conflicting principles, as its central features derive from struggles taking place in previous epochs. Such features took shape amid social and political conflicts and, inevitably, bear the mark of previous historical events: "The central principles of the law are the site of struggle and contradiction... The fate of law as a rationalising enterprise is tied up with the nature of law as a social, historical force" (Norrie 1993, p. 9).

Elements of Marxist analysis are detectable in these formulations, and in general in all conceptions of law as expression of group conflicts. The law, according to Marx, allows states to deal with "social imperfections" as if they were "evil dispositions of the poor" rather than products of specific economic and political arrangements.

Marx, however, defines the law as the form in which the individuals of a certain ruling class assert their common interests.

Abolitionist analysis, however, develops some central tenets of conflict theory in an original manner. Where conflict theorists seem to limit themselves to the critique of the criminal justice system as an expression of antagonistic values and interests operating from above, abolitionists reappropriate the very notion of "conflict" and turn it into a critical tool to be utilized from below. Nils Christie (1977, p. 1), for instance, remarks that conflicts are hijacked by the criminal justice system and that criminology lends a helpful hand in the process: "conflicts have been taken away from the parties directly involved and thereby have either disappeared or become other people's property." Law specialists are not the only ones to "steal" conflicts; conflict theorists in criminology, in effect, do the same, though in a different fashion: they turn interpersonal conflicts into class conflicts, thus again taking them away from those directly involved.

Comparing two opposite scenes, an adjudication case in a small village in Tanzania and a case dealt with in a juvenile court in England, Christie provides the following description. In the former, the contending parties, a man and a woman, occupy the center of the judicial scene, they make their claims, while their respective friends and relatives take part without "taking over," and the audience intervenes with questions, information, and jokes. In the latter, the predominance of professionals is visible, while the young defendant appears to be confused by those surrounding him, whose role he does not understand and who do not take any notice of him. "The truth is that, for the most part, the business of the criminal courts is dull, commonplace, ordinary and after a while downright tedious" (Christie 1977, p. 2).

Christie (1998, p. 19) provides two pictures of how rules are created. In the first, Moses descends from the mountains carrying the rules engraved in granite. "Moses was only a messenger, the people – the *populus* – were the receivers, controlled from far above." The other picture sees a group of women gathering at the fountain, around the well, or along a river, who while fetching water exchange information and evaluations.

> The point of departure of their conversation will often be concrete acts and situations. These are *described*, *compared* to similar occurrences in the past or somewhere else, and *evaluated* – right or wrong, beautiful or ugly, strong or weak. Slowly, some common understanding of the occurrences might emerge. This is a process whereby norms are *created*. It is a classical case of horizontal or *equalitarian justice*. (Christie 1998, p. 119)

The act whereby conflicts are hijacked by professionals entails a specific construction of reality hinging on an incident, narrowly defined in time and space, whereby the individual is separated from the context in which the action takes place. According to the abolitionist critique of legal professionalism, those who act, in this way, are isolated from the very incident involving them, from the environment in which it occurs, from their "friends, family, and the material substratum of their life world" (Hulsman 1991, p. 23). Also, they become separated from those people who, in that specific situation, feel victimized by certain acts. Hulsman focuses on the culture and the organizational make up that extrude people from their social context and artificially sets them against one another: "In this sense the cultural organisation of criminal justice creates 'fictitious individuals', and a 'fictitious' interaction between them" (Hulsman 1991). The core function of such an organization is the apportioning of blame, whereby events are assembled and sanctions catalogued on the basis of seriousness and gravity, respectively. A hierarchy is, therefore, postulated and patterns of acts and institutional responses to them are artificially established. Events, in their turn, are chosen among a limited range of conducts and interactions, namely, those which the criminal justice organization feels more comfortable in explaining and evaluating. "No comparison is made with events and behaviours outside that range." Patterns and hierarchies are drawn in a universe which is far removed from the context in which events took place, so that what appears to be consistent within the criminal justice profession is experienced as inconsistent by those who constitute the objects of that profession. Hulsman remarks, in this respect, that "values and perceptions in society are not uniform." And in this way he echoes conflict theorists such as Sellin (1939), who emphasizes "culture conflicts," that is the differences between social groups, their aspirations and behavioral models, leading to acts defined as crime. But he adds that the "program" for blame allocation typifying the criminal justice system is a true copy of the doctrines of "the last judgment" and "purgatory" developed in certain varieties of Western Christian theology. The criminal justice system, in his view, maintains the features of "centrality and totalitarianism which belong to those theological doctrines. Naturally, those origins – this 'old' rationality – are hidden behind new words: 'God' is replaced by the 'law' and the consensus of the people" (Hulsman 1991, p. 24).

The notions of separateness and isolation, applied by abolitionism to the offender, the victim, and the problematic situation characterizing their interaction, are also mobilized by conflict theorists, when their analyses focus on the law as the realm of individualist philosophies and as the reflection of egoistic interest. Abolitionism enriches this approach with detailed descriptions of contexts and events, suggesting that these are subjected to a process of abstraction or outright obliteration. It is the very cultural organization of the criminal law that fosters this process, a process whose "frame of reference," as Hulsman terms it, gives no room "for active participation and guidance for the parties involved."

> When the police operate within a criminal justice frame they tend not to be directed by the wishes and desires of the complainant, but by the requirements of the legal procedure which they are preparing. The complainant becomes a guide for their activities as "witness". A witness is mainly a tool to bring legal proceedings to a successful end. In a comparable way the frame of court proceedings precludes – or makes it anyway specially difficult – for the victims to express freely their view on the situation or to enter in an interaction with the person who is standing as a supposed offender before the court. (Hulsman 1991, p. 24)

If classical conflict theory would mainly examine structures and agencies while omitting to look in any detail into their specific social composition, abolitionism devotes particular attention to such composition, particularly to the increasing separate specialization characterizing it. In brief, it is argued that each institutional agency develops its own criteria for action, its own ideology and culture which may lead it to collide with other agencies. Agencies are bureaucracies pursuing internal goals, such as expand, attenuate internal problems, monitor the wellbeing of their members, and ultimately ensure their own survival. "The process of bureaucratisation and professionalisation within the criminal justice system makes it a soulless machinery" (Hulsman 1982, pp. 56–57).

Limits to Pain

Among the most critical traits characterizing abolitionist theories and practices are those concerned with the nature, function, and philosophy of punishment. The infliction of punishment might be justified if it contributed to the maximization of happiness rather than the maximization of pain. Hence the need for any theorizing on punishment to consider its social consequences. Those who defend institutional coercion in the form of punishment may advocate rehabilitative treatment, may value its general or individual deterrent effect or its function of incapacitation aimed at prolific offenders. Abolitionists argue against such philosophies and their practical outcomes. In Mathiesen's (1990) view, by contrast, prison has no defense whatsoever, particularly when inspired by penal retribution, whereby offenses are said to possess a "punishment value" that can be translated into a specific amount of "time." Mathiesen argues that time is only measurable subjectively, and that the perception of its entity depends on one's proximity to those serving a prison sentence.

After crossing almost four centuries, Mathiesen (1990, p. 33) identifies four major components in the rehabilitative ideology which have been remarkably constant through time. "Work, school, morality and discipline have run through the centuries as main pillars of thinking." While the emphasis may constantly shift from one to the other component, rehabilitation itself, understood as "return to competence" is rarely achieved. Evidence of this failure is found by Mathiesen in a variety of studies grouped in three main areas. First, he refers to prison treatment studies, which yield largely uniform findings: the effects of treatment, in terms of preventing recidivism, and regardless of strategy or intensity, are generally poor. He then considers the notion of treatment in relation to the actual environment in which it is expected to be carried out. The organization of prison life, the authoritarian principles governing it, along with the bureaucratic arrangement stifling creative projects, provide a setting which is conducive to everything but treatment. Moreover, prisons are usually "overcrowded, run-down, and more or less dangerous places to those who inhabit them," and if anything, they exacerbate the conducts of those they purport to treat (Mathiesen 1990, p. 41). Finally, he looks at a range of studies devoted to prison as a social organization or focused on prisoners' communities, arguing that inmates become soon "prisonized," that is they internalize the values and rules of a violent and coercive environment and adopt a culture which is impervious to change (Clemmer 1940). Such culture, which protects the inmates from the very setting that they inhabit, makes prison perform the function of a crime school, and prisoners more or less immune to treatment or readjustment programs.

Moving on to the discussion of deterrence as general prevention, Mathiesen (1990) suggests that the failure of this alleged function of imprisonment should prompt a dramatic reduction of the use of imprisonment as a whole. Research into the actual effectiveness of general prevention is so inconclusive that, when seen against the harmful outcomes of custodial punishment, it would be wise to limit the use of custody altogether, until its effectiveness is soundly proven. General deterrence consists of a message addressed to the general population, but it would appear that something goes horrendously wrong in the communicating process: "general prevention functions in relation to those who do not 'need' it. In relation to those who 'need' it, it does not function" (Mathiesen 1990, p. 69). It is for moral reasons, therefore, and due to lacking evidence of the actual working of general deterrence, that the use of imprisonment should be reduced to a minimum. There is, however, a supplementary argument that would strengthen Mathiesen's reductionist stance, namely, that deterrence makes no difference between guilty and innocent individuals, in the sense that if non-offenders were punished its goal would still be achieved. By punishing people at random, irrespective of whether or not they have committed crimes, one may well assume that the general population is discouraged from violating the law. But, although acceptable from a utilitarian perspective, punishing the innocent would run counter the Kantian moral principle whereby individuals must not be treated as mere instruments for the achievement of a social good: all individuals should be treated as ends in themselves, in accord with their dignity and worth as persons.

Considering how ideas of general deterrence have increasingly replaced notions of treatment, Christie (1982, p. 29) notes that at least the latter "had a formally clear target: those receiving treatment." The former, instead, chooses as a target a vague entity described as the general population, which is composed of people who, if deterred from one type of offense may resort to another type; people who, in the face of a punitive message launched in one country, may decide to move to another country; or finally, people who might find it hard to receive the punitive message altogether. "Conceptually, as well as empirically, ideas of general prevention or deterrence are thus more cumbersome to handle than ideas of treatment" (Christie 1982, p. 29).

As for individual deterrence, the belief that imprisonment amounts to a form of training for criminal activities to come is persuasive enough, in the abolitionist view, to discard this specific official function of prison as inappropriate. Schools of crime teach resentment and hostility rather than respect for norms and others. In addition, it could be stressed that deterrent effects are notoriously difficult to establish, "since they involve counterfactuals: would these individuals have acted differently had the threat of criminal penalties been other than it was?" (Duff and Garland 1994, p. 25).

The history of punishment may be interpreted as an evolutionary process leading to distinct stages characterized by increasing leniency. Some commentators may describe the development occurring as one bringing a gradual decrease in pain, a progressive shift from brutality to sweetness. Of course, there is difference between gruesome public executions and incarceration, between suffocating reclusion and custody in a single room, with cold and hot water and a TV set. But does this indicate an objective reduction of the pain caused by imprisonment?

> I just do not know. Each form would have to be evaluated according to its own time, by those receiving the pain, in the framework of their usual life and other people's life, and in the light of what they saw as their sins. I do not see how a scale could be established. (Christie 1982, p. 9)

Law texts establish when punishments have to be inflicted, while judges decide how long they should last; neither examines their effects on bodies and minds, the suffering produced, how it feels. The belief in slow progressive humanization accounts for the lack of this type of information (Hulsman 1986).

Criticizing the humanization thesis, Hulsman (1986) notes that the overall amount of punishment does not tend to decline, as the number of convicted people per 100,000 of population shows a strong cyclic movement. The trend of convictions, in many countries in the industrialized world, has in fact for some time moved upward. Hulsman finds the assessment of the qualitative aspects of punishment more difficult. It is true that the application of the death penalty has been greatly reduced in recent centuries, and in normal times has been abolished in many countries. The same can be said, he adds, of many forms of corporal punishment. It may also be true that some progress has been made in improving the regime in prison systems. He, nevertheless, suggests caution when judging qualitative amelioration and humanization. The amount of suffering incorporated in legal penalties cannot be measured on a scale of absolute values. "To a large extent it consists of the difference between the normal living situation of people and that which is created by the intervention of the criminal justice system." Hulsman is alluding to the concept of less-eligibility whereby the conditions in prison must be worse than the worst social condition of people in liberty. Because the prison system "has always drawn its clientele mainly from the most disadvantaged sections of the population," conditions in prison will reflect the lowest standard of living experienced by this social sector. Now, as "the living standards of those same sections have in Europe improved considerably in recent years," one may presume that prison conditions have improved accordingly, but the reality is that "improvement inside prisons during the last 30 years do not appear to have kept pace." The conclusion is that "if this supposition is correct, then the degree of suffering from the penal sanction has in a sense increased" (Hulsman 1986, pp. 64–65).

The "evolution" of pain, moreover, is alleged to have rendered physical punishment obsolete, which is far from reality. Prison still entails forms of corporal affliction: it degrades the body, it deprives of air and light, it imposes humiliating sanitary conditions, it causes diseases, and it produces sterile suffering.

> Not all suffering is bad; some is beneficial, in that it makes our consciousness develop while opening up new existential paths, making us better humans and getting us closer to the others. Imprisonment is a type of suffering that does not create anything, that does not generate any meaning. (Hulsman 1982, p. 59)

There are, finally, problems caused by institutional intervention per se: custodial and non-custodial measures alike erect.

> obstacles to the development of social solidarity and respect for differences that manifest themselves in "lived life"... Rather than alternatives to custody we should devise alternative approaches by which information is collected about the needs of those involved in problematic situations. (Hulsman 1982)

Intellectual Biographies

As remarked above, abolitionism does not possess one single theoretical or political source of inspiration, but a composite backdrop from which, wittingly or otherwise, it draws its arguments and proposals for action (Ruggiero 2010, 2011). The intellectual biography of Louk Hulsman, for example, should be related to some crucial passages in the Bible where mercy is advocated while judgment and retribution are rejected. The Gospels of Mark, Luke, and Paul seem to provide an apposite theological underpinning for Hulsman's abolitionism, which can also be assimilated to Saint Francis' ecumenism and his view that thieves are not those who steal, but those who do not give enough to the needy. Radical theology or the theology of liberation also provide significant sources of inspiration. With Bakunin's anarchism, Hulsman shares the belief that the realization of freedom requires that political action be conducted religiously. In some pages of Marx, Engels, Tolstoy, and Hugo an echo is felt of Hulsman's concepts of redemption of punishment, self-government, mercy, and pietas. Hulsman's system of thought, in brief, is shown to display a high degree of syncretism.

An equal if not a higher degree of syncretism shows the intellectual trajectory of Thomas Mathiesen, whose materialist approach is punctuated by arguments for a pluralistic, interdisciplinary, sociology of law. The writings of Marx and Engels constitute an ideal background for an understanding of Mathiesen's work, which on the other hand draws on a number of other theoretical sources. The focus on offenders and prisons, and on social movements traditionally excluded from orthodox notions of class struggle, makes Mathiesen's stance an implicit critique of classical Marxism. What constitutes the originality of Mathiesen's work is the coalescence of research, action, and theorizing that characterizes his entire career, as an academic as well as an activist. Action is inherent in the research method adopted, and those researched are the prime subjects involved in research as well as in action. His radical analysis, therefore, is a tool which constantly translates knowledge about conflict into collective praxis for those producing it.

In his work, Nils Christie stands in favor of simplicity and intelligibility. He says that when we write we should keep our favorite aunt in mind, like Kropotkin said that anarchist literature had to keep in mind the workers to whom it was addressed. Christie's system of thought echoes anarchist theories of law and authority. His critique of legal professionalism is akin to libertarian arguments against the proliferation of laws, while his appreciation of conflict as a resource brings to mind the anarchist idea that problems within communities can only been resolved if those involved possess sufficient autonomous resources to do so. One of Christie's arguments is that communities and groups, irrespective of their dimension, may find abolitionist experimentation possible only if the interactions within them are highly frequent and intense. In this sense, he expresses the purest of anarchist notions, namely, that a better social life is experienced when communities develop social feelings and, particularly, a collective sense of justice that grows until it becomes a habit.

Conclusion

On the nature and concept of crime abolitionism engages in a debate that our ancestors Plato, Aristotle, and Spinoza saw as the distinction between good and evil. Far from exhausted, such debate continues in an array of disciplines ranging from theology to psychology, from political philosophy to social theory. The abolitionist critique of the law contains some Kantian moral principles, in that existing law is seen as using people as an instrument, a chance to signal the strength of transcendental and worldly authority. Abolitionists hold the view that the law performs a transferral of vendetta from the sacred into the judicial system of the state, and that it is mainly addressed to scapegoats, chosen violators among many, who are at the same time "evil" and "sacred," in that they reproduce social life through sacrifice.

In abolitionist analysis, issues such as the designation of what constitutes crime, the intervention of the law, and the infliction of penalties that take place in institutional settings, which are distant from the situations, are addressed. Within such settings, punishment is the outcome of a cognitive process whereby the institutions make sense of events and claim that their response is based on the knowledge of such events. Abolitionists, instead, claim that knowledge is mainly achieved within precise contexts in which problematic situations arise, and it is for those involved to "repair" such situations through the development of restorative collective capabilities.

Related Entries

- ► Juvenile Diversion
- ► Monitoring and Evaluation of Restorative Justice
- ► Restorative Justice
- ► Restorative Justice and Practice

Recommended Reading and References

Akers RL (1997) Criminological theories: introduction and evaluation. Roxbury Publishing Company, Los Angeles

Bianchi H (1991) Abolitionism in the past, present and future. In: Lasocik Z, Platek M, Rzeplinska I (eds) Abolitionism in history: on another way of thinking. Institute of Social Prevention and Resocialisation, University of Warsaw, Warsaw

Chambliss W (1969) Crime and the legal process. McGraw-Hill, New York

Chambliss W (1975) Criminal law in action. Hamilton, Santa Barbara

Christie N (1977) Conflicts as property. Br J Criminol 17:1–15

Christie N (1982) Limits to pain. Martin Robertson, Oxford

Christie N (1998) Between civility and state. In: Ruggiero V, South N, Taylor I (eds) The new European criminology: crime and social order in Europe. Routledge, London

Christie N (2004) A suitable amount of crime. Routledge, London/New York

Clemmer D (1940) The prison community. Holt, Rinehart and Winston, New York

Davis AY (2008) History is a weapon. The challenge of prison abolition: a conversation between Angela Y. Davis and Dylan Rodriguez. http://www.historyasweapon.com/defcon/davisinterview.html

de Tocqueville A (1956) Democracy in America. Mentor Books, New York

Duff RA, Garland D (eds) (1994) A reader on punishment. Oxford University Press, Oxford

Gouldner AW (1975) For sociology: renewal and critique in sociology today. Penguin, Harmondsworth

Honneth A (2007) Disrespect: the normative foundations of critical theory. Polity, Cambridge

Hulsman L (1982) Peines perdues: le système pénal en question. Le Centurion, Paris (interview with J. Bernat de Célis)

Hulsman L (1986) Critical criminology and the concept of crime. Contemp Crises 10(1):63–80

Hulsman L (1991) Alternatives to criminal justice: decriminalization and depenalization. In: Lasocik Z, Platek M, Rzeplinska I (eds) Abolitionism in history: on another way of thinking. Institute of Social Prevention and Resocialization, Warsaw

Kennedy MC (1976) Beyond incrimination: some neglected facets of the theory of punishment. In: Chambliss W, Mankoff M (eds) Whose law? What order? A conflict approach to criminology. Wiley, New York

Mathiesen T (1990) Prison on trial. Sage, London

Norrie A (1993) Crime, reason and history: a critical introduction to criminal law. Weidenfeld and Nicolson, London

Nussbaum M (2000) Women and human development: the capabilities approach. Cambridge University Press, Cambridge

Quinney R (1970) The social reality of crime. Little, Brown, Boston

Quinney R (1974) Critique of the legal order. Little, Brown, Boston

Ruggiero V (2010) Penal abolitionism. Oxford University Press, Oxford

Ruggiero V (2011) An abolitionist view of restorative Justice. Int J Law Crime Justice 39(2):100–110

Sellin T (1939) Culture, conflict and crime. Social Science Research Council, New York

Sen A (2009) The idea of justice. Allen Lane, London

Simon JK (1991) Michel Foucault on Attica: an interview. Soc Justice 18(3):23–37

Turk A (1969) Criminality and the legal order. Rand-McNally, Chicago

Vold GB (1958) Theoretical criminology. Oxford University Press, New York

Penal Justice and Social Injustice

Matt Matravers
Morrell Centre for Toleration,
University of York, York, UK

Overview

The connections between retributive justice, broadly conceived, and distributive or social (in) justice have long troubled criminologists, moral and legal philosophers, and criminal justice practitioners. One reason for this is that the economically deprived make up a disproportionate number of criminal offenders (and victims), a fact made even more striking in many places when overlaid with statistics about race. Another is that inequalities within liberal democratic societies have risen since the 1970s; dramatically in some places such the USA and UK. This rise has accompanied a dramatic rise in punishment and incarceration. Assuming inequality is a proxy for distributive injustice, then, the world is becoming less just and, in some places, more punitive. These facts compel our attention in (at least) two ways. First, there is a practical public policy question. If crime rates are correlated to poverty and/or inequality, then one way to address crime is through social justice and the redistribution of income, wealth, and opportunities. This is no doubt sensible, but it is not the subject of discussion here. Rather, second, juxtaposing criminal and distributive justice brings into focus a normative question of whether social justice is a precondition of legitimate criminal justice, or more minimally, whether – and if so in what ways – social injustice renders the imposition of criminal liability and sanctions morally problematic.

It is this normative issue that is discussed below. Certainly, some philosophers concerned with the justification of punishment have held that conditions of social injustice – such as those prevalent in much of the developed world – render punishment at best problematic and at worst morally impermissible. Thus, Antony Duff (1986, p. 294) argues that punishment "will not be justifiable unless and until we have brought about deep and far-reaching social, political, legal, and moral changes in ourselves and our society". Similarly, Jeffrie Murphy (1973) claims that existing social injustice fatally undermines the only successful theory through which punishment can be made permissible. And David Bazelon (1976, p. 385) that "there can be no truly just criminal law in the absence of social justice."

This entry will consider the arguments for the proposition that social injustice renders the imposition of criminal liability and sanctions morally problematic. It does not deal with the question of whether *lesser* punishments are merited by those who are disadvantaged by injustice, but rather with the moral legitimacy of punishment as a whole in circumstances of distributive injustice. The first set of arguments focuses on theories of punishment; the second on the responsibility of the offender; the third on whether an unjust society has the standing to call offenders to account. In all cases, it is assumed that the background is one in which the law is not egregiously unjust. That is, the discussion does not encompass societies in which, for example, the law formally discriminates against non-white persons with respect to political and property rights. Such societies clearly raise issues about whether obedience is owed by those against whom discrimination is practiced and about the moral and legal liability of those who break the law. However, the assumption here is that the system has no such formal rules, but that the social and economic distribution of goods is unequal and unjust (which will have implications for the value of the formal rights enjoyed by all). In short, the focus is on societies such as the USA and UK, not on those such as South Africa in the era of Apartheid.

Ideal Theories and Actual Practices

There are many theories that purport to justify or render permissible state punishment. Such theories tend to be ideal; that is, they develop philosophical accounts of the justification of punishment assuming certain ideal conditions. The question then arises as to whether those theories remain compelling when applied to the actual world. It is not possible to discuss each and every such theory here, but consider some general approaches and how they might be affected by social injustice.

Consequentialist Theories

Consequentialist theories justify punishment by appeal to its good consequences. A system of punishment is justified if it brings about better overall consequences than any alternative system including the option of doing nothing. In general, consequentialists claim that punishment is justified because it secures good consequences through individual and general deterrence, rehabilitation, and incapacitation. These are empirical claims. If punishment were to fail to realize good consequences, or fail to realize as good a set of consequences as could be attained by an alternative policy, then it is not justified on this basis.

Clearly issues of social justice will as a matter of fact have an effect on any such empirical claims. It may be best to divert resources away from criminal justice and toward social justice if that will be more effective in reducing crime (or realize other good consequences that outweigh any resulting bad consequences). It may be that rehabilitation is better realized by providing social and economic opportunities rather than by punishment, and so on. These matters can only be settled by empirical enquiry, not by philosophical armchair reflection.

Varieties of Desert

According to one variety of desert-based retributive justice, punishment is justified if and only if it restores the fair balance of benefits and burdens that the original crime disturbed. This so-called fair-play theory was pioneered by Herbert Morris (1968) and Jeffrie Murphy (1973).

In Murphy's formulation, the argument relied on a hypothetical social contract. In return for security and assurance, self-interested persons would accept an authority whose purpose was to ensure those goods by upholding and enforcing the agreed rules. That is, the citizen benefits from the rule of law and all that comes with it and in return undertakes to maintain society's rules through his obedience to them. Failure to abide by the rules is a form of "free-riding"; the offender gains the advantages of the system without having to "pay" for them through his own self-restraint. In being punished – in (hypothetically) agreeing to the system of punishment under which he is now made to suffer – the offender pays a different price, one designed to remove the extra benefit he has attained and thus restore the *status quo ante*.

However, as Murphy was quick to point out, the contract here is reciprocal. The authority of the law – and so the duty of obedience on the part of the citizens – stems from its being part of a system of rules to which reasonable self-interested persons could agree. Thus, if the system of rules is unjust – if, for example, it enforces fundamentally unjust distributive arrangements or discrimination between persons on the basis of their skin color – then citizens no longer have a duty of obedience and the state no longer has the moral authority to punish breaches of the rules.

This account of punishment has been subject to criticism and few (even among its early adherents) continue to adhere to it (for a sympathetic contemporary account, see Dagger 1993; for a summary of the main criticisms, see Matravers 2000, pp. 52–72), but the contractualist worry is a genuine one. To the degree that society is conceived of as a cooperative endeavor in which reciprocally agreed rules are enforced and obeyed, there is a concern that the state's failing to keep its side of the bargain, so to speak, renders illegitimate its enforcing of those rules.

A different, but related worry, affects the censure account of retributive punishment that replaced the "fair-play" account as the leading desert theory in the literature. According to Antony Duff and others, the justification of punishment lies wholly, or in part, in its calling the alleged offender to account and, should the alleged offender prove to have performed the wrongful act without excuse or justification, in its expressing censure of the offender on behalf of the polity. In imposing a penal sanction, the polity provides a vehicle for this censure and a potential vehicle for the offender's penitential response.

In his early work (1986, pp. 291–299), Duff is clear that social injustice renders this account inapplicable to the real world. Such injustice is a symptom, and cause, of the absence of a genuine moral community on whose behalf, and in whose name, the offender is called to account. If there are no shared values, then the offender cannot be censured in their name and cannot in turn use punishment to restore herself to those values and so to the community. Like Murphy, Duff argues that the gap between the ideal and the actual is so great as to render the ideal unusable. However, in later writings, Duff becomes less pessimistic although he still poses the question of whether those who have been systematically excluded from their fair shares of society's goods can be thought to be bound by the laws of that society, a question he thinks likely to be answered in the negative for many offenders (2001, pp. 183–184). This motivates a different worry, considered below, which is that social injustice undermines the standing of the society to call the offender to account.

Consequentialist, fair-play, and censure-based theories of punishment, then, are each in their own way threatened by conditions of social injustice. In all these cases, the fact of injustice threatens the viability or applicability of the philosophical theory that purports to justify punishment in the real world. However, desert theorists (in particular) also have a different worry about social injustice; this is not that it renders the theory – that punishment is justified because and to the degree that it is deserved – inapplicable, but that it undermines the deservingness of individual offenders; it renders previously unjustly treated citizens less blameworthy and so less deserving of punishment.

Blameworthiness and Distributive Injustice

The arguments of the section above suggest that social injustice might challenge the legitimacy of the law, the account of political obligation, or the justification of the system of punishment as a whole. A (perhaps) more familiar worry is that in some way or other, defendants from indigent, or unjustly disadvantaged, backgrounds are less to blame than are others. This section considers this worry in relation to cases of excuse and of justification.

Rotten Social Background

Perhaps the most famous debate in relation to doing penal justice in conditions of social injustice arose out of a 1973 case, *US v Alexander*. Two black defendants were charged with the murder of two white victims. The defendants had shot the victims following an altercation at a fast-food restaurant during which the victims had called the defendants "God-damned niggers." On behalf of one of the defendants, who was not mentally ill according to any recognized diagnostic categories, a psychiatrist nevertheless testified that the defendant suffered from an "emotional illness" that was rooted in his upbringing in a very poor and deprived neighborhood. This upbringing had left the defendant with an acute sensitivity to racial discrimination such that when he had been subject to racist abuse, he had felt "an irresistible impulse to shoot" (for useful accounts of the issues here, see Green 2011; Heffernan 2000). The judge allowed the jury to consider the mental capacity of the defendant, but mindful of the possibility that the jury might be moved by pity or sympathy for the defendant given his upbringing, the judge went on to say to the jury that it ought not to be concerned "with the question of whether a man had a rotten social background." The defendants were convicted.

The court of appeals upheld the conviction, but one of its judges, David Bazelon, in part dissented. He argued that there was a relevant causal claim linking the defendant's rotten social background and the defendant having no "meaningful choice when the racial insult triggered" his response in the restaurant (*United States v Alexander*, 471 F.2d 923: 960). The truth of such a causal claim would need to be considered by the jury, but Bazelon wrote confidently of the social science evidence in its support and of the need to consider what to do with such defendants should they prove innocent, but dangerous (holding out the options of therapy or preventive detention).

The idea of a defense of rotten social background was taken up by Richard Delgado (1985). Delgado argued that the law should recognize a new excuse based on extreme poverty and deprivation. As an excuse, the claim would not be that the conduct was justified. Rather, the defendant would have to prove that at the time of acting, her conduct was the result of behaviors shaped by her background such that she met (an extended understanding of) normal excusing conditions: For example, she was unable accurately to understand the nature and consequences of her conduct; she was suffering from involuntary rage; or was otherwise unable to control her conduct.

It is important to note that neither Bazelon nor Delgado argues that the fact of an (unusual) causal history to the agent's conduct is enough to excuse. All actions have causes, some are more unusual than others and some evoke sympathy, but as Stephen Morse insists (2000, p. 130), "causation is neither an excuse per se nor the equivalent of hard choice (so-called compulsion), which is an excusing condition." Rather, the argument is that the particular background of deprivation is such as to render some defendants excused at the time of their conduct because it interferes with those ordinary capacities on which the law depends when holding persons responsible. In this, a defense of rotten social background would come close to the defense of "battered woman syndrome," in which it is argued that a history of violent abuse can render a woman incapable of (legally) responsible behavior even at some later point when her conduct seems otherwise to be chosen.

Justification

Proponents of the excuse of rotten social background accept that the defendant's conduct was wrongful; they merely argue that it may be excused. However, another response to social injustice is to regard some conduct performed by the disadvantaged as justified. Consider a case imagined by Jeremy Waldron (2000, pp. 104–105). A homeless, unemployed, destitute man in a society with little or no welfare assistance scavenges in a park. He comes across a half-eaten hamburger and begins to eat it. However, the person who had bought the hamburger, having eaten half, was planning to feed the rest to the birds. He complains to the police and the destitute man is arrested and charged with theft.

The defendant claims the justification of necessity. Before considering this, it is worth noting how this differs from the claim of excuse in rotten social background. In the case of excuses, the conduct is wrong, but excused. In the case of justification, the conduct itself is not wrongful. Thus, in rotten social background cases, there is no obvious restriction on the possible conduct to which the excuse might attach or on the range of victims of the conduct. As is clear from the *Alexander* case, the excuse potentially covers serious wrongs such as homicide and it would have made no difference to the case whether the victim was poor or rich. In the justification case under discussion, it is clear that the conduct that is justified is severely restricted. The defendant might claim justification for picking up the hamburger, but he could not do so for killing the owner so as to get the hamburger (no matter how hungry he was). Similarly, the agent could not claim justification for taking the last morsel of food from a similarly situated indigent victim. These differences, of course, reflect the different structures of excuses and justifications.

Although Waldron is clear that there is a moral case for the justifiability of the destitute man's actions, he is confident that any legal system dedicated to the protection of property rights will reject the necessity defense in this instance. To establish necessity in the common law tradition, the defendant must show imminent threat; that the conduct would be effective in ameliorating that threat; that no alternative legal means to do the same was available; that the conduct was less harmful than the harm threatened; and that the circumstances that gave rise to the threat are not attributable to her.

The threshold for the use of the necessity defense is very high. To rebut it, the prosecution would only have to show that the indigent eater of the hamburger was not in danger of death or serious injury, or that he could have gone from the park to the local charitable shelter, or that his condition was due to his culpable failure to find work. In short, necessity and long-term poverty do not map easily onto each other. Moreover, as Waldron insists, systems of law protect the orderly and predictable governance of property, an order that cannot be subject to disruption whenever someone feels the need to appropriate something of someone else's.

The arguments in this section demonstrate the difficulties in matching concerns about social justice with issues in criminal justice. In one way, this is right; not all poor people commit crimes (and not all crimes are committed by the poor) and it would be demeaning to large sections of the population to pretend otherwise. If poverty or disadvantage is to be relevant to judgments of the offender's liability, then it has to be shown either that the offender's background relevantly impaired him at the time of acting, or that his actions were legally justified. It is likely that these hurdles will not in fact be often cleared.

However, this seems to leave the issue unresolved. It *does* matter that society is distributively unjust and it does trouble us that those who systematically suffer most from injustice also disproportionately appear in the criminal justice system as victims and offenders.

The State's Standing to Hold the Offender to Account

The focus has so far been on the applicability of theories of punishment to the real world and on the offender. Social injustice raises genuine questions for the first, but the focus on the blameworthiness of the offender seems either to lead in directions covered by the existing criminal law (with respect to excuses and justifications) or to risk widening the scope of the enquiry in ways that potentially demean the poor and disadvantaged. Returning to the original issue, however, suggests an alternative approach in which the focus is not on the offender, but on the state.

Assume – reasonably enough in most cases – that the offender acts culpably. The law prohibiting the conduct is reasonable, the offense is wrongful, and the offender is blameworthy for it. To take an example, a disadvantaged offender defrauds the state by claiming unemployment benefits while working on the black market for cash.

In response to cases like this, Antony Duff asks not whether the offender is responsible, but *to whom* he is responsible. That is, criminal responsibility is a relational concept; an alleged offender is responsible *for* some conduct *to* someone or some institution. Thus, a teacher is responsible for marking her students' essays and that responsibility is owed to her students. Moreover, the teacher can call her students to account for failing to submit their essays, and so on. In short, with respect to essays (and no doubt other things), the teacher has *standing*.

The question Duff poses is whether a state's maintaining of chronic and systematic injustice removes its standing to call offenders to account. As the paper of one of his papers suggests – "'I Might Be Guilty, But You Can't Try Me': Estoppel and Other Bars to Trial" (2003), Duff is skeptical that unjust societies retain the standing to call to account those whom they have treated unjustly.

This is an attractive response to the issues. It does not deny the responsibility of indigent offenders, and it surely captures a plausible intuition (and mirrors many real life situations in which persons may think, "I may be wrong, but who are you to criticize?"). There is surely something wrong in being held to account by an institution that is itself in violation of the values to which it appeals in its calling the alleged offender to answer for her actions (cf Sykes and Matza 1957 on "The Condemnation of the Condemners").

However, as stated, the account is underdeveloped. Two questions arise: First, does the unjust state lose the standing to judge all offenders and offenses (even those unrelated to the injustice)? Second, does the unjust or immoral state lose all standing across all dimensions? To see this, first contrast the unemployed offender mentioned above who defrauds the state by cheating on his social security payments and the same person who beats his wife. It may be thought that the state lacks the standing to call him to account for fraud, given a background of systemic injustice, but why should this affect its standing to call him to account for assault? Second, consider the standing of the UK state to call a well-off defendant to account for insurance fraud, given recent revelations that the same state had deliberately misled its citizens in the run-up to the second Iraq war and had, when fighting terrorists in Northern Ireland in the 1970s, connived in the extrajudicial execution of (at least one) political agitator (for a summary of these criticisms, see Matravers 2006).

In short, the argument from standing seems to do too much. It is in that sense reminiscent of Murphy's argument that legitimate punishment is impossible in a real world in which distributive injustice is prevalent. Yet, it undeniably captures something of the worry that many people have when confronted by the facts of distributive and retributive justice and, if nothing else, it should evoke in all citizens what Duff (2003, p. 259) calls "a properly humble and cautious spirit which recognizes how far from clean our collective hands and consciences are" with respect to penal justice.

Conclusion

Not all the poor suffer from the same injustices, not all of them commit crimes, and not all of those who commit crimes do so in similar circumstances (for a nuanced discussion of a "case-by-case" approach to social injustice in relation to the criminal law, see Green 2011, pp. 364–373). Nevertheless, the poor and disadvantaged are more likely to be victimized by crime, they are more likely to engage in criminal activity, and they appear disproportionately among those arrested and convicted (Green 2011, pp. 354–355; Western 2006). If one were to visit any prison throughout much of the world, one would be confronted with populations predominantly made up of young men from poor and disadvantaged backgrounds (in the USA, young black men from the same) (Wacquant 2009). At the same time across much of the developed world, the inequalities between the richest and the poorest in societies are growing. These facts compel attention not just in themselves, but because in punishing the poor, "we run the risk of compounding the sins of socio-economic injustice with those of retributive injustice" (Green 2011, p. 376).

The purpose of the arguments above is to show that confronting this possible double injustice from within the existing structures of the criminal law is deeply problematic. Some disadvantaged offenders will suffer from "abnormality of mind" (or similar) consequent upon their disadvantage, but this will be a small minority and, even among those, difficult to establish in court. Few, if any, will be able to claim a justification such as necessity. Ideal theories of punishment may not apply to the world we confront, but absent some other form of social organization, the necessity to punish will persist.

Perhaps the route out of these difficulties is to look beyond the resources of the criminal law and of philosophies of punishment and to political philosophy more generally. Much contemporary discussion of justice revolves around the question of how we might distinguish between those things – actions, omissions, character features, and so on – for which the agent is responsible and on the basis of which he might legitimately claim the benefits or be made to bear the burdens and those things that are merely chance or circumstance – things that, so to speak, simply happen to her – on the basis of which she may deserve assistance or compensation. Of course, while in many cases, poverty and injustice simply happen to people, the decision to commit an offense may fall into the category of choice rather than chance. Nevertheless, asking "when is it fair to make people bear the burdens or enjoy the benefits of their conduct?" may open up avenues of enquiry that might inform the criminal law and, if not that, at least our attitudes to those who fall liable to it (for an initial development of such a strategy, see Matravers 2007: Chap. 3; Scanlon 1998: Chap. 6). That said, given the extent of disagreement among political philosophers as to the answers to these questions, criminal law issues such as those discussed above remain both alive and worthy of attention.

Recommended Reading and References

Bazelon D (1976) The morality of the criminal law. South Calif Law Rev 49:385–405

Dagger R (1993) Playing fair with punishment. Ethics 103:473–488

Delgado R (1985) "Rotten Social Background": should the criminal law recognize a defense of severe environmental deprivation?. Law Inequal 39:12–23

Duff RA (1986) Trials & punishments. Cambridge University Press, Cambridge

Duff RA (2001) Punishment, communication, and community. Oxford University Press, Oxford

Duff RA (2003) "I Might Be Guilty, But You Can't Try Me": Estoppel and other bars to trial Ohio State. J Crim Law 1:245–259

Green S (2011) Just deserts in unjust societies. In: Duff RA, Green S (eds) Philosophical foundations of the criminal law. Oxford University Press, New York, pp 352–376

Heffernan WC (2000) Social justice/criminal justice. In: Heffernan WC, Kleinig J (eds) From social justice to criminal justice: poverty and the administration of criminal law. Oxford University Press, New York/Oxford, pp 47–83

Matravers M (2000) Justice and punishment: the rationale of coercion. Oxford University Press, Oxford

Matravers M (2006) ‘Who’s Still Standing?’ A comment on Antony Duff’s preconditions of criminal liability. J Moral Philos 3(3):320–330

Matravers M (2007) Responsibility and justice. Polity Press, Cambridge

Morris H (1968) Persons and punishment. Monist 52:475–501

Morse SJ (2000) Deprivation and desert. In: Heffernan WC, Kleinig J (eds) From social justice to criminal justice: poverty and the administration of criminal law, Practical and professional ethics series. Oxford University Press, New York/Oxford, pp 114–160

Murphy J (1973) Marxism and retribution. Philos Public Aff 22:17–43

Scanlon TM (1998) What we owe to each other. Harvard University Press, Cambridge, MA

Sykes GM, Matza D (1957) Techniques of neutralization: a theory of delinquency. Am Sociol Rev 22(6):664–670

Wacquant LJD (2009) Punishing the poor: the neoliberal government of social insecurity. Duke University Press, Durham

Waldron J (2000) Why indigence is not a justification. In: Heffernan WC, Kleinig J (eds) From social justice to criminal justice: poverty and the administration of criminal law. Oxford University Press, New York/Oxford, pp 98–113

Western B (2006) Punishment and inequality in America. R. Sage Foundation, New York

Penal Paradigms: Past and Present

Leslie Sebba
Institute of Criminology, Faculty of Law,
The Hebrew University, Mt. Scopus,
Jerusalem, Israel

Overview

Accounts, critiques, and theoretical conceptualizations of penal systems reveal that the characteristics of these systems may vary in very significant ways. Penological histories show how the “classical school” of the late eighteenth and early nineteenth centuries (epitomized in the famous essay of Cesare Beccaria’s *Of Crimes and Punishments*) believed that the primary rationale for the penal system was general deterrence and advocated that this should be reflected in sentences proportionate to the seriousness of the offence. They then describe how this ideology was in the latter half of the nineteenth century replaced by individualized sentencing – which in turn gave way a century later to a “just deserts” philosophy (again proportionate). These successive approaches are sometimes referred to as **paradigms**.

A focus on the discourses prevalent in the literature on contemporary criminal justice, however, indicates that in fact a variety of competing **models** are identified or advocated – perhaps most famously illustrated by Herbert Packer’s seminal article outlining two apparently alternative models – “due process” and “crime control.” This entry will attempt to provide an introductory overview of these developments – both the historical and the contemporary. The usefulness of these approaches will be considered, as well as confusions arising from the terminology employed, given that neither the terms “paradigm” nor “model” are always used consistently.

Introduction

While penal systems, whether formal or informal, appear to constitute an integral part of every society (and thus presumably, as argued by Durkheim, fulfill some significant social function), they differ from each other in almost every conceivable aspect: the forms of conduct punished, the procedures for determining guilt, the aims of punishment, its modalities, and its severity. Moreover, individual penal systems fluctuate and indeed undergo radical changes over time. In elaborating on these characteristics, traditional penological texts (e.g., Barnes and Teeters 1959) presented largely descriptive accounts of the historical development of penal systems, mainly in Europe and the Americas – while there has also been a fascination with ancient and “primitive” systems (Diamond 1971). Social theorists since Durkheim have offered generalized theoretical explanations for the form and severity of punitive practices – a development analyzed by

Garland in *Punishment and Modern Society* and applied to contemporary society in his *Culture of Control* (Garland 1990, 2001). More recently, there has been a wave of research endeavoring to explain the wide differences among nations in the severity of their sanctioning policies (Tonry 2007).

What many of both the descriptive accounts of penal systems and the more critical analyses share in common is that they tend to view the penal system as – indeed – a "system" with a coherent purpose and philosophy, thereby implying a degree of interconnectedness and compatibility among its different component elements and shared values among the various actors. It is also common to identify somewhat dramatic ideologically driven policy transformations believed to have taken place at certain periods of time. Thus, historical accounts will invariably refer to the eighteenth century as the era of classical criminology (and in particular penology), the late nineteenth and most of the twentieth century as the era of positivism, with the 1970s onward as being dominated by the "just deserts" model. The penal systems identified under these different headings are shown to diverge in fundamental ways, in particular having different sentencing rationales, thereby giving rise to their designation as alternative "penal paradigms." One purpose of this entry will be to elaborate on and evaluate such historical presentations.

It is suggested here that such analyses may be viewed as being complemented by contemporary discourses relating to the characteristics and ideologies of the present-day criminal justice system, which sometimes echo the typification of past ideologies but with a tendency to describe a multiplicity of trends or ideologies operating simultaneously or in competition.

Such a connection between past and present is indeed made by Jock Young in his much-cited essays "Thinking Seriously about Crime" and "Incessant Chatter: Recent Paradigms in Criminology" (Young 1981, 1994) which incorporate critiques of the traditional presentation of the historical paradigms while using this concept to analyze a number of more recent theoretical approaches. However, Young also uses the term "model" in this context, and this is the term more frequently adopted in current penological discourse when referring to alternative penal options. Much of the discourse relating to criminal justice models, however, seems to have been inspired by the dichotomy developed by Herbert Packer (1964) in his well-known law review article "Two Models of the Criminal Process," and indeed the relevant literature frequently acknowledges such a debt.

This entry will provide a brief survey both of the historic "paradigms" and of the contemporary "models," followed by some reflections on the nature of their contribution and related conceptual issues. The survey will be preceded, however, by a discussion of the terminology employed here and elsewhere in the literature in this field.

Paradigms and Models in a Criminological-Penological Context

The term paradigm was developed in the context of the hard sciences to denote the prevailing understanding or interpretation of natural phenomena. It was popularized by Thomas Kuhn in his *Structure of Scientific Revolutions* (Kuhn 2012) when analyzing how **changes** come about in such understandings on the part of the scientific community and came to be applied also to the social sciences. While Kuhn's analysis is essentially restricted in its application to phenomena which may be studied empirically and is thus not applicable to many concepts and ideas pertaining to punishment, it has now a more widespread usage, connoting "a cluster of concepts such as assumptions, values, practices and methodologies shared by a community of researchers in a given discipline" (New World Encyclopedia, online). There may, however, be competing paradigms at any given time, where the evidence is equivocal and/or the scientific community divided in their views and interpretations.

In this context it should be noted that while paradigms relate to "scientific findings" and are thus closely linked to positivist conceptions, each paradigm is rather a social construct, describing the collective understandings of a given community as perceived by an observer (or participant). In the social sciences where "evidence" tends to be less hard than in the natural sciences (and the role of values more salient), it is probable that there will be less consensuality, thereby increasing the possibility of "competing paradigms."

The term "model" has similarly acquired a wider usage than its original meaning of a (usually small-scale) replication of a product for the purposes of demonstrating or testing its operation, or "providing insight into the consequences of a decision" (Chaiken et al. 1975: v). It is simply used to describe the main features and mode of operation of a system – whether existing or proposed. Both the expressions "paradigm" and "model" implicitly refer to a system or framework which comprises a number of components operating as a cohesive whole – and, where relevant, sharing the same values and ideology. Only the paradigm, however, has a connotation of exclusivity – although also that such exclusivity may be ephemeral.

Other related terms used in the penological context include *penal strategies* (Garland 1985) and *styles of social control* (Black 1976). Young (1981) not only used both the terms "models" and "paradigms" but also Weber's concept of *ideal types*, which suggests an idealized account of the institution in question which is unlikely to be found in that form in real life. Recent analyses of Packer's models have considered their relationship with Weber's ideal types (see below).

It should also be noted that the first two paradigms to be discussed here, the "classical" and the "positivist" paradigms, have generally been described in the literature as *schools*. Finally, certain distinct perspectives on criminological issues today may be referred to as *criminologies*, such as "feminist criminology" and "green (environmentalist) criminology." The relative significance and usefulness of these terms has yet to be fully explored.

Historic Paradigms

The three historical paradigms to be considered here are set in somewhat different theoretical frameworks. While the classical school, at least as reflected by its most notable representative, Cesare Beccaria, focuses almost exclusively on the formal institutions of the criminal justice system (in this respect resembling the contemporary models to be considered later in this entry), the positivist school is much wider in its concerns, being equally – or indeed primarily – concerned with the etiology of crime. By contrast, the third of our historical paradigms, just deserts, has the narrowest focus, being concerned exclusively with the purposes and procedures of sentencing.

The Classical Paradigm

This paradigm is identified with the European enlightenment movement and the emergence of the ideas which were to become identified with liberal theory and English constitutionalism (Young 1981) as an antidote to the monarchical despotism prevailing in much of continental Europe. Building on Montesquieu's doctrine of the separation of powers and Rousseau's social contract theory, Beccaria (1764/1995) in his seminal essay advocated a strict separation of powers and equality before the law in order to counter the arbitrary practices prevalent in certain contemporary European regimes which occurred as a result of executive intervention in trials and sentencing. Beccaria advocated uniform penalties, proportional to the seriousness of the offence, which would strengthen their deterrent power: rational persons would refrain from committing crimes attracting high penalties, especially if law enforcement was systematic, certain, and immediate. The attribution of free will and rationality to potential offenders and correspondingly the importance of deterrence as the primary aim of punishment are thus perceived as identifying characteristics of the classical paradigm.

This paradigm may also be seen to be linked to the development of the modern prison. While Beccaria was advocating the prison as a more effective deterrent than the death penalty, his contemporary John Howard was documenting the practices of the then local prisons, which led the way to the enactment in 1779 of the Penitentiary Houses Act in order to establish a new and rationalized form of prison – the penitentiary.

As is well known, this system, based upon uniform treatment in identical cells which it was assumed would lead to penitence and reform, was ultimately pioneered in the United States a few decades later. Another key figure identified with this paradigm was Jeremy Bentham, also famous for developing his own "rational" model of prison development – the Panopticon (Semple 1993).

The Positivist Paradigm

Concepts of equality and uniformity in penal policy were challenged during the second half of the nineteenth century by the rise of the social sciences and the claim of Auguste Comte that in the course of human development "metaphysical" or abstract thinking was to be replaced by the "positivist" stage – and the application of purportedly scientific methodology and empiricism. One outcome of this was Lombroso's theory of delinquent man. The focus on human differences resulted in a penchant among positivist criminologists and their disciples in the field for the classification of offenders for treatment purposes. Such techniques derived encouragement from the mid-century experiments with the "progressive stage system" (pioneered in Ireland) whereby prisoners advanced through the prison system and into the community in the light of their conduct.

Under the influence of positivism, the reformation of the offender became the declared purpose of punishment rather than deterrence, and the ideal of standardized penalties proportional to the seriousness of the offence as posited by Beccaria came to be replaced by the individualization of punishment, as expounded in the treatise bearing this name written by Raymond Saleilles (1911). The new approach to sentencing was epitomized by the principle of the indeterminate sentence widely adopted in the USA whereby terms of imprisonment imposed by the courts were open-ended, the actual duration of the sentence to be determined by a parole board in the light of the prisoner's progress.

This "individual treatment model" of sentencing, as it developed towards the end of the nineteenth century and throughout much of the twentieth century, was characterized by the creation of specialized penal and judicial institutions ostensibly adapted to the needs of the offenders in question, such as reformatories and juvenile courts for younger offenders and institutions for dangerous psychopaths, the involvement of behavioral experts such as psychiatrists and probation officers in decision-making, and the increasing recognition of the helping professions as agents of offender rehabilitation in both closed institutions and in the community. Moreover, while the determinist conceptualization of human behavior and the focus on the individual led to the application of more flexible criteria of criminal responsibility, some versions of the Social Defense ideology (a branch of the positivist movement which focused on penological issues) proposed also to dispense with the legal requirement of a "guilty mind" (cf. Ancel 1965). The threshold requirement for treatment would be the personality of the alleged offender – the main criterion being his or her dangerousness.

The Just Deserts Paradigm

The positivist ideology with its emphasis on rehabilitation meshed well with the welfarist ethos that was widely supported during the period following the Second World War. However, if the individualized treatment paradigm was most widely accepted (at least in the prevailing rhetoric) in the USA, it was there too that it was the most forcefully rejected during the 1970s. Critiques of this model focused on the inequality and arbitrariness of the system, in particular in the context of parole decision-making (American Friends Service Committee 1971). Critics included not only radicals and liberals but also conservatives (cf. Blomberg and Lucken 2000). Adding fuel to the fire was the critique of sentencing disparities deriving from the wide discretionary powers held by the judges (Frankel 1973), and – perhaps most influential of all – the empirical claim that the effectiveness of the treatment methods on which the system was based was unproven (Martinson 1974).

The structure which was to take account of these critiques, as formulated most articulately in *Doing Justice*, the report published by Andrew von Hirsch (1976), was a return to fixed and proportional sentencing. While outwardly this bore some resemblance to the system advocated by the Classical School (see above), its underlying philosophy was not deterrence, or indeed any other consequentialist purpose – as none had been shown to be effective by the empirical evaluations, but desert, a contemporary version of the historic aim of retribution. According to this doctrine, the penalty would match the seriousness of the offence as measured by the harm inflicted and the culpability of the offender – and thus become standardized. The mechanism whereby this was to be achieved was the establishment of a sentencing commission which would issue detailed guidelines to the sentencing judges, whose discretion would be severely curtailed. Since sentences were to be for fixed terms, the discretionary power of the parole boards would be abolished.

Support for such reforms in the USA was such that they were adopted in principle by the federal system and around half of the states – as well as some overseas countries, as documented in other entries. The concept of "structured sentencing," whether by guidelines or otherwise, attracted even wider support. The "individual treatment model" as the foundation of the penal system had in general been rejected.

Historical Paradigms: A Critique

The preceding overview suggests that in the context of this entry, the term "paradigm" is most appropriately (or least inappropriately) used in the context of the positivist school, since the policies advocated in accordance with this approach were the most rooted in the empirical understanding on the part of its adherents of the world about them. It was indeed this characteristic that ultimately contributed to the demise of this paradigm when research findings suggested that the ability of decision-makers to make accurate prognoses and identify successful treatments was very limited.

There are, however, other problems with what Young has described as the typical textbook presentation of a linear development of a succession of paradigms. Young is correct in pointing out that despite the enormous impact of positivist thinking on the twentieth century penological discourse, criminal justice structures – particularly in European countries – remained throughout the twentieth century strongly influenced by the tenets of eighteenth century classicism. The nature and seriousness of the offence remained the primary consideration in determining the sentence, with the characteristics of the offender (other than past convictions) playing a secondary role. Contemporaries therefore referred to the model adhered to in Europe during the course of the twentieth century Europe as "neoclassical."

Today too it is hard to identify a dominant paradigm. As noted above, the parameters of the third of the paradigms referred to, the desert paradigm, were somewhat narrowly drawn (being confined to sentencing practices), and in any case, most contemporary western systems have been strongly influenced by consequentialist ideologies – incapacitation, deterrence and even rehabilitation. It should also be recalled that the concept of the paradigm refers in principle not directly to the practices of the actors in the field (in this case the criminal justice agencies) but to the **perception** of these practices – or perhaps of the ideologies which are thought to motivate them – on the part of the relevant disciplinary group. Thus, radical criminologists such as Young, who perceive the criminal justice system primarily in terms of control, if not exploitation, of the powerless by the powerful, have since the 1960s been offering alternative paradigms to explain its operation.

It may always have been the case that it was more problematic to identify monolithic trends (such as are implicit in the concept of the paradigm) in one's own era, when different views or nuances will be more apparent than when looking back in history. However, this problem has surely been aggravated by the dichotomization of social scientists into positivists and social constructivists, further compounded by the fragmentation caused by postmodern developments – leading Garland and Sparks (2000) to opine that further paradigms in the (consensual) Kuhnian sense can no longer be anticipated. These authors point to the salience of crime in the contemporary politics of neoliberal societies. This has had the effect of reducing the relevance of the prevailing views of academia and the professions, which have perhaps themselves become more fragmented and politicized.

Criminal Justice Models

The Legacy of Herbert Packer

Unlike the paradigm, the concept of a model does not imply consensus and exclusivity but rather variation and experimentation. While criminologists have sometimes used them in the context of empirically oriented work (e.g., Feeley 1973; Chaiken et al. 1975), much of the modern discourse on criminal justice models has been devoted to normative and/or ideological issues, having been inspired, as noted above, by Herbert Packer's seminal article "Two Models of the Criminal Process" (Packer 1964 – subsequently expanded into a book – Packer 1968). The article followed a period of activism in the area of civil rights and criminal justice on the part of the US Supreme Court headed by Chief Justice Warren, but Packer was doubtless concerned by the minority dissenting views which were to become the majority in the era of Chief Justice Burger. It seemed to Packer that the Court and indeed other criminal justice agencies were being required to choose (or at least prioritize) between two alternative types of process which he identified as due process and crime control, respectively, representing alternative sets of values and priorities.

A system based upon due process he perceived as being analogous to an obstacle course, in that it sought to ensure protection of the rights of the innocent (and to risk the acquittal of the guilty) by imposing restraints on the prosecution. Under the crime control model, on the other hand, the system was comparable to an assembly line in a factory – designed to maximize the numbers who would be processed through the system to conviction and punishment in minimum time. This was to be achieved by focusing on the investigation stage and offering plea bargains to the "factually guilty." Although Packer did not perceive these models as mutually exclusive, his dichotomization has penetrated deeply into (at least) the academic criminal justice culture.

Critiques of Packer's Models

Packer's Models have been the brunt of a number of critiques, some trenchant, with Griffiths (1970) being a notable early example, and most recently and comprehensively by Macdonald (2008). These critiques have focused on three main aspects of Packer's work: (a) his conceptualization of the model, (b) the validity of the due process and the crime control models and the relationship between them, and (c) the need for additional models having (in the view of their proponents) greater validity or irrelevance.

(a) Conceptualization

The lack of precision in the concept of the model in the contemporary literature of the social sciences – on one view its "indiscriminate use" (Macdonald 2008) – has already been referred to. Zedner (2004: 116) has observed that "It is often difficult to determine whether models are intended as analytical devices or as normative statements of ideals to which the process should aspire" – an ambivalence conveyed by Packer too. In the wake of Damaska's discussion as to whether Packer's models are consistent with Weber's ideal types, Macdonald notes that Weber distinguished between ideal types and ideals, the former constituting an analytical construct based upon accentuated tendencies which was in his view how Packer conceptualized his models, without fully developing them. (Whether in fact the models were quite unconnected with value judgments, as Weber described his ideal types, is another question).

Macdonald sought, as the title of his article indicates, to "learn from Packer's mistakes" and to "succeed… where Packer failed" (Macdonald 2008: 269), for which purpose he found necessary to differentiate between the investigative and the trial stages of the process, while arguing the need for a multidimensional model. On similar grounds Ashworth and Redmayne (2005: 38ff.)

reject the "models" approach as an unsatisfactory basis for the evaluation of the criminal process. These writers find models based on simple dichotomies to be too crude – tending "to reduce the world to black and white" (Zedner 2004: 119–120).

Additional conceptual issues arise in the context of the evaluation of the two models as outlined by Packer, and the relationship between them.

(b) The Two Models and the Relationship Between Them

A number of criticisms have been directed at the specific models described by Packer, as well as suggesting that the dichotomy may be a false one (see, e.g., Macdonald 2008: 264ff.). Both Damaska (1973) and Smith (1997) find fault with due process as an objective (in Damaska's words "obstacles without a course"), rather than a principle imposing restraints on the pursuit of other objectives. In this context Macdonald (2008: 289) refers to the work of the Swedish jurist Jareborg who differentiated between a "defensive model" and an "offensive approach" to criminal policy. Zedner (2004: 118) has pointed out that speed and efficiency – characteristics attributed by Packer to the crime control model – are also important for due process and protection of the defendant's liberty. McBarnet (1981), echoing the literature of radical criminology, has argued that rather than constituting an alternative approach, due process serves to legitimize the goals of (repressive) crime control, thus giving rise to "due process. . . **for** crime control."

(c) Additional Models

With Packer's dichotomy as a "given," other writers have proffered additional (often designated "third") models. Griffiths' "family model" (which clearly falls into Zedner's normative category) was a proposal for a more therapeutic procedure. This, too, has been the direction of at least some of the recent writings seeking to accommodate a role in the system for the victim. Thus, while Beloof (1999) has proposed a Victim Participation Model within a due process framework, Roach (1999) has portrayed the alternatives of a punitive and a non-punitive model of victims' rights, and it is the latter – designated a "Circle Model" and based upon restorative justice – which he seeks to promote.

A unique attempt to incorporate a broader model approach into an empirical analysis of the criminal justice system was undertaken some years ago by King (1981). King identified six alternative models of the system for which accounts could be found – whether by name or implicit – in the criminological or criminal justice literature. To Packer's duo (due process and crime control), King added: the medical (or treatment) model, and three further models deriving from criminological research or sociological critiques – the bureaucratic model, the status passage model, and the power model. He then identified the "features of the court" which one would expect to find under each model. The validity of these models could then be tested (in the instant case, in his study of English magistrates' courts) by observing to what extent these features were actually found in the courts. "As in the physical and social sciences, the term model is used here to denote a hypothetical but coherent scheme for testing the evidence" (p. 12).

There is no doubt that the integration of a multi-model approach into a field study served to enrich this discourse. However, there is a measure of disjuncture here in the construction of the models in that the first three, which King calls "participant" models, reflect the idealized expectations of a certain category of participant (thus, again, falling into the category of normative models), while the other three which he refers to as "social" models are based on academic critiques of the existing system. On the other hand, the attribution to each model of a "social function" bestows upon the models a unifying ideology or rationale such as generally characterized the paradigms discussed in the first part of this entry.

It should be noted that the "penal process" to which the literature reviewed in this section relates is that of common law jurisdictions (primarily the USA). Models relating to other systems will be very different, but some "Packer scholars" have dwelt on the relationship between the common law and developments in continental Europe's "inquisitorial system": see, e.g., Damaska (1973).

Special Categories of Offender

Variations in the manner in which criminal justice systems operate tend to be augmented when dealing with offenders with distinctive traits or pathologies – or whose offences are perceived as particularly threatening or distinctive in some other way. Examples would be sex offenders, offenders against state security laws ("terrorists"), offences related to drug use, minors, or the mentally unstable. In such cases different paradigms or models may be identified. Thus, the modification of human rights protections when dealing with security offences (including punishment without trials) may be such that the "war" model is arguably the most appropriate metaphor to attach in this case.

By contrast, purportedly softer alternatives to the prevailing model(s) may be adopted when dealing with minors. Identifying the optimal policy for his purpose, however, has been the subject of intense debate over the past century. It was in relation to juvenile offenders that the treatment model of the positivist school assumed its most comprehensive application, in particular via the vehicle of the juvenile court, first established (in Chicago) in 1899. The treatment of juvenile offenders became virtually indistinguishable from "welfare" cases such as children in need of care and supervision. They were thus deprived of the guarantees provided under the US Constitution's Bill of Rights. Juvenile courts in other common law jurisdictions, however, retained somewhat more of a rights framework – while the US Supreme Court case *Re Gault* in 1967 (387 U.S. 1) reversed the welfarist trend in the American system. If for some decades the juvenile justice system has seen these two approaches – the "welfare" model and the "justice" or "legal" model – in combat (or fusion), recent decades have seen strong claims, particularly in the antipodes, for their replacement by a system of restorative justice. Three distinct approaches to juvenile justice have thus emerged. As observed by Borowski and O'Connor (1997: 229): "Unlike other aspects of the criminal justice system, it is possible to identify coherent theoretical positions reflected in legislation, policy and practice."

O'Connor's review of these models considers how each model responds to such underlying questions as follows: What is the cause of crime? How should it be viewed? How should it be defined in law? How should the state respond? What is the relationship between the individual and the state? What is the purpose of the sanction? O'Connor's models arguably present both an analytical and a normative framework for the comparison of these three ideologies, while his approach bears a resemblance to Young's analysis of criminological paradigms referred to in the Introduction.

Conclusions

This entry has analyzed the use made of two concepts – the paradigm and the model – in the course of the history of the criminal justice system. It seems to the present writer that the paradigm has been used primarily to describe changes in the main orientation of the system at different periods, while the model has been more extensively used in contemporary settings. Both terms have had a usage somewhat different from their original meaning (in the case of the model, a multiplicity of usages), and partly for this reason, the discourses in which they have been incorporated have been challenged.

The departure point for this entry was the concept of the paradigm – perhaps the more interesting concept because of the connotations conveyed by this term in a penological context of (a) being unequivocally linked to an ideology, (b) indicating an interpretation uniquely accepted by the relevant disciplinary group, and (c) having played a key role in the history of penology and the penal system. Yet while the brief historical overview presented at the beginning of this entry may have been sufficient to indicate the dramatic contrast between the main historic paradigms (in particular classicism and positivism), Young is almost certainly correct in denying the occurrence of a "paradigm shift" as a momentary historical event, the reality having been considerably more complex. Further, the idea that the "community of [criminology] scholars" will identify with a common vision of penal policy has become increasingly unlikely from generation to generation. Even less likely is the prospect of shared perceptions between the "criminology community" and policy-makers, such as Garland suggests may have occurred in the era of "penal welfarism" and "government criminology." As observed by Garland and Sparks (2000), in an era of ideological fragmentation and politicization, paradigms – like cultures – are likely to be multiple.

The relatively fluid concept of the model is free from such connotations, but its conceptual ambiguities have given rise to much confusion. This confusion came to the fore in the aftermath of the publication of Herbert Packer's famous two models of the criminal process – the crime control model and the due process model, and the numerous critiques thereof. Nevertheless, Packer's models have become an inseparable part of the discourse of criminal justice. Indeed, the two concepts considered in this entry are likely to survive and enhance both critical analyses and normative debates on the many facets of penality – while methodological inquiries as to their validity and applicability may serve to clarify underlying epistemological issues.

Acknowledgement The author would like to thank Prof. Lucia Zedner for her helpful comments on a previous draft.

Related Entries

- ▶ Historical Criminology
- ▶ History of Corrections
- ▶ History of Juvenile Justice
- ▶ Legal Control of the Police
- ▶ Problem-Solving Courts
- ▶ Prosecution and Wrongful Convictions

Recommended Reading and References

American Friends Service Committee (1971) The struggle for justice. Hill & Wang, New York

Ancel M (1965) Social defence. Routledge & Kegan Paul, London

Ashworth A, Redmayne M (2005) The criminal process, 3rd edn. OUP, Oxford

Barnes HE, Teeters NK (1959) New horizons in criminology, 3rd edn. Prentice-Hall, Englewood Cliffs

Beccaria C (1764/1995) On crimes and punishments and other writings (trans ed: Bellamy R). Cambridge University Press, Cambridge

Beloof DE (1999) The third model of criminal process: the victim participation model. Utah Law Rev 2:289

Black D (1976) The behavior of law. Academic, New York

Blomberg T, Lucken K (2000) American penology: a history of control. Aldine de Gruyter, New York

Borowski A, O'Connor I (1997) Models of juvenile justice. In: O'Connor I, Borowski A (eds) Juvenile crime, justice and corrections. Addison Wesley Longman, Melbourne (Chap 12)

Chaiken J et al (1975) Criminal justice models: an overview. Rand, Santa Monica

Damaska M (1973) Evidentiary barriers to conviction and two models of criminal procedure: a comparative study. Univ Pa Law Rev 121:506

Diamond AS (1971) Primitive law: past & present. Methuen, London

Feeley M (1973) Two models of the criminal justice system – an organizational perspective. Law Soc Rev 7:403

Frankel ME (1973) Criminal sentences: law without order. Hill & Wang, New York

Garland D (1985) Punishment and welfare. Gower, Aldershot

Garland D (1990) Punishment and modern society. Clarendon, Oxford

Garland D (2001) The culture of control. OUP, Oxford

Garland D, Sparks R (2000) Criminology, social theory, and the challenge of our times. In: Garland D, Sparks R (eds) Criminology and social theory 1. Clarendon, Oxford

Griffiths J (1970) Ideology in criminal procedure or a 'third model' of the criminal process. Yale Law J 79:359

http://www.newworldencyclopedia.org/entry/Paradigm. Accessed 22 Jan 2013

King M (1981) The framework of criminal justice. Croom Helm, London

Kuhn T (2012) The structure of scientific revolutions. 4th edn, University of Chicago Press, Chicago

MacDonald S (2008) Constructing a framework for criminal justice research: learning from Packer's mistakes. New Crim Law Rev 11:257

Martinson R (1974) What works? Questions and answers about prison reform. Pub Interes 35:22

McBarnet DJ (1981) Conviction: law, the state and the construction of justice. Macmillan, London

Packer H (1964) Two models of the criminal process. Univ Pa Law Rev 113:1

Packer H (1968) The limits of the criminal sanction. Stanford University Press, Stanford

Roach K (1999) Four models of the criminal process. J Crim Law Criminol 89:671

Saleilles R (1911) The individualisation of punishment. Little, Brown, Boston

Semple J (1993) Bentham's prison: a study of the panopticon penitentiary. Clarendon, Oxford

Smith DJ (1997) Case construction and the goals of criminal justice. Br J Criminol 37:319

Tonry M (ed) (2007) Crime, punishment and politics in comparative perspective. U of Chicago Press, Chicago

von Hirsch A (1976) Doing justice: the choice of punishments, Hill & Wang, New York

Young J (1981) Thinking seriously about crime: some models of criminology. In: Fitzgerald M (ed) Crime and society: readings in history and theory. Routledge, London

Young J (1994) Incessant chatter: recent paradigms in criminology. In: Maguire M et al (eds) The Oxford handbook of criminology. Clarendon, Oxford, p 69

Zedner L (2004) Criminal justice. OUP, Oxford

Penal Philosophy and Sentencing Theory

Matt Matravers
Morrell Centre for Toleration
University of York, York, UK

Overview

Many penal philosophers think that offering a justification of the state's right to punish is particularly urgent (and difficult) precisely because punishment involves doing things to people that would, outside the practice of punishment, be gross rights violations. That is, punishment typically involves imposing some deprivation – for example, the removal of property or freedoms – on a person for committing an offense and this stands in need of justification. Yet, despite this motivation, many theories of the justification of punishment say surprisingly little about the detail of sentencing and penalties. This entry considers the various rationales for sentencing with a view to showing both how different justifications of punishment have different implications for sentencing theory and practice and how reflecting on sentencing matters can shed light on the justification of punishment. The entry considers first consequentialist theories – that is, theories that justify punishment by reference to the net benefits that accrue through punishment in comparison with any other form of social control (or doing nothing) – and then desert-based, retributive, theories. Finally, it considers some questions of the relationship between theory and practice.

Punishment and Sentencing

Punishment

Punishment – or at least state punishment – involves the intentional infliction on an offender by a proper authority of "pain or other consequences normally considered unpleasant" for an offense against legal rules (Hart 1968, pp. 4–5). In addition, it involves the expression of moral condemnation or censure (Duff 2007; Feinberg 1970). When a court passes sentence, it gives shape to the particular consequence to be visited on the offender. In practice, the options available to the court may be limited by guidelines or statute and the court may not be explicit in justifying the use of one penalty rather than another. However, the use of punishment in general, and of particular penalties, requires justification and those justifications tend to fall into two broad categories: consequentialist and retributive.

For consequentialists, punishment is justified because its consequences – specifically, having the system of punishment with its associated penalties – can be expected to yield higher net benefits than having an alternative system of social control or having no such system. This may be because punishment reduces future instances of criminal behavior through deterrence, rehabilitation, or incapacitation.

For retributivists, punishment is justified because it is deserved by the offender for his offense. It may be that good consequences also follow, but the justification of punishment does not appeal to these consequences. What, then, do these broad theories of punishment imply for sentencing theory (insofar as we can distinguish sentencing theory and penal philosophy) and practice?

Deterrence

Deterrence is typically split into "special" and "general." Special deterrence is aimed at the particular offender and is sometimes discussed in relation to the giving of more substantial punishments to repeat offenders. However, it is not advanced as a general theory of sentencing. This is perhaps a reflection of the lack of empirical evidence when it comes to special deterrence and of the fact that a sentencing scheme that was genuinely tied to special deterrence would have to allow specific, individualized sentences for each offender in relation to his particular likelihood of reoffending. Such a sentencing scheme would lack all consistency and would, in any case, be unworkable.

General deterrence addresses not the particular offender in court, but potential offenders throughout the population. If the general justifying aim of punishment is to secure good consequences by reducing criminal behavior, then it is argued that the mechanism by which it does that is general deterrence. The origins of the modern approach are most often associated with the great utilitarian philosopher Jeremy Bentham. Bentham thought the infliction of pain was, other things equal, an evil, but the threat – and where that threat failed, the infliction – of pain could be justified if it secured better future consequences than would be achieved by any alternative action or policy. Assuming that the addressees of the system (that is, ordinary citizens) were motivated by the pursuit of pleasure and the avoidance of pain, Bentham argued that criminal sanctions should be set at just that level required to change the outcome of the "hedonic calculus" for each individual in favor of what would be conducive to the maximization of the general good. Moreover, Bentham went on, penalties should be such as to "induce a man to choose always the least mischievous of two offenses" by ensuring that "where two offenses come into competition, the punishment for the greater offense must be sufficient to induce a man to prefer the less" (Bentham 1970, pp. 168). That is, he developed an account of penalty scales (Bentham 1843).

In thinking of people as rational calculators set on maximizing their own interests and susceptible to changes in the payoffs of different actions, Bentham's theory has much in common with contemporary proponents of the economic theory of law such as Richard Posner (Posner 1985). However, critics of the position have pointed to flaws both in the empirical claims made on behalf of general deterrence and in its normative foundations.

Empirical Questions Around General Deterrence

Empirical questions around general deterrence revolve around the overall claim that the threat of legal punishment reduces offending and the claim that *marginal* changes in the severity of sanctions has an effect on rates of offending. The difficulty with the first of these claims is in establishing that it is the threat of punishment rather than anything else that stops people from breaking the law. Moreover, insofar as the theory posits an effect that happens because people reason about what to do as rational, self-interested calculators, the account may explain some people's behavior and not that of others. It may simply not be true that certain kinds of offenders do accurately balance the factors for and against their offending in advance (even when they know, for example, what penalty they are likely to be given and the likelihood of being caught). In short, general deterrence may be at best a selective explanation of law-abiding behavior because many offenders do not think as Bentham and modern economists believe they think (Doob and Webster 2003).

A similar problem arises with the empirical evidence in relation to the effects of marginal changes in sentencing severity or probability of conviction. In a review of research, Andrew von Hirsch and his coauthors identified the factors that would need to be true for such marginal changes to have an effect on a potential offender's behavior. That offender, they argued, would need to know about the change; he would need to include the increased or decreased risks of suffering punishment in his decision as to whether to offend; he would need to believe in a realistic possibility that he might be caught and that the penalty would be applied to him; and he would need to be willing to adapt his behavior in the light of these new considerations (von Hirsch 1999). The conclusion of the study is that while there is some evidence of a link between offending and the certainty of punishment, there is far less evidence of a link between offending and the severity of punishment (see also Bottoms and von Hirsch 2011).

Normative Questions Around General Deterrence

Insofar as deterrence theories are embedded in, and result from, consequentialist accounts of punishment, they are subject to the moral criticisms that are often leveled at those accounts. The problem is a general one: Consequentialist accounts value the aggregate (sometimes the average) good achieved in the future. We are to do what will, in Bentham's famous phrase, achieve "the greatest happiness of the greatest number." But, in being concerned for the greatest happiness (or utility), the theory leaves out questions of the distribution of utility; it fails to pay attention to what John Rawls (1971, p. 27) calls, "the distinction between persons." For this reason, it is alleged, if the calculation dictates it, it may be not merely morally permissible, but morally obligatory, to frame and punish an innocent person (McCloskey 1968) or to punish an offender far in excess of what is usual to make an example of him.

Consequentialist writers have responded to this kind of criticism by insisting that the theory is best thought of as addressing rules and institutions and not individual actions (Goodin 1995; Rawls 1955/1999). In that case, they argue the system of rules in relation to criminal justice and the range and type of penalties that best promote long-term good will be much the same as is found in current liberal systems. That is, the rule "only punish those (believed to be) guilty and not the innocent" will better promote the good than a rule that allows the innocent to be sacrificed each and every time that particular act will have better long-term consequences than any alternative. However, few have been convinced by this argument in part because of doubts about the claim that the usual procedural rights and protections of (suspected) offenders will ensure long-term utility, and in part because even if it delivers the right answer (not to hang the innocent man), it does so for the wrong reasons. The innocent man deserves to be spared because it would be wrong to kill him, not because a rule against hanging innocents secures better consequences than its opposite (Matravers 2000, pp. 17–23).

The idea that the justification of punishment and of the penalties it imposed must be grounded in the requirement "to preserve and increase the welfare of the state" (Michael and Adler 1933, p. 340) was prevalent for much of the century prior to the 1960s (Tonry 2011b, pp. 14–18). It seemed to some writers to be the only "rational" basis for sentencing (Walker 1969). However, from the 1970s onward, the theory came under sustained criticism. In moral theory from neo-Kantian writers including, and inspired by, John Rawls and in penal theory by the "Just Deserts" movement. Yet, we continue to believe – and to find evidence for – the claim that the whole criminal justice system must surely have *something* to do with reducing future crime rates and that taken as a whole, it does deter crime (Robinson and Darley 2003, 2004). Moreover, statutes and judges continue to refer to deterrence at the point of sentencing. So, whatever its philosophical difficulties, it is clear that deterrence is not yet moribund in practice.

Incapacitation and Rehabilitation

At the high point of utilitarian welfarist thinking about punishment (in roughly the middle of the Twentieth Century), it seemed obvious to writers on the criminal law that a rational approach to punishment would eschew retributivism as "fallacious" (Michael and Adler 1933, p. 341) and barbaric. Instead, it seemed clear that if the welfare of the state was to be increased, then antisocial behavior, and those who committed it, needed to be dealt with "scientifically." Offenders would be "treated" under the careful eye of experts and until "cured" would generally be held away from the general population. For the influential social and legal thinker, Barbara Wootton, the criminal law was irrational in its concern with intentions and mental states prior to conviction. Rather, *mens rea* considerations mattered only "after a breach of law has been proved" because of "the light which they throw on the likelihood of [the offender] offending again, and upon the most hopeful way of dealing with him" (Wootton 1978, p. 224). Sentences should be indeterminate and offenders released when, or if, their individualized treatment plans proved successful.

There were individual voices of dissent. For example, CS Lewis's (1949) "The Humanitarian Theory of Punishment" offered the argument that the treatment model reduced persons from choosing agents to manipulable things. By the late 1960s, those voices had increased and by the 1970s become a clamor (Frankel 1972; Kleinig 1973; Morris 1974; von Hirsch and Committee for the Study of Incarceration 1976). The charge was normative in that utilitarianism treated people as mere objects to be manipulated; political in that indeterminate sentences were racially biased and unfair; and empirical in that increasing doubts were expressed about the efficacy of treatment.

Remnants of both incapacitation and rehabilitation nevertheless remain in contemporary penal philosophy and sentencing theory. With respect to incapacitation, governments (of all persuasions) have responded to the risks created by so-called dangerous persons – often sex offenders – by allowing incapacitative sentences or post-sentence restrictions such as sex-offender registers. However, the policy has largely drifted free of any theoretical foundation in a theory of punishment. Rather, it exists as a bolt-on public policy that philosophers have approached as a discrete problem. The question is, can – and if so, how should – a liberal society restrict the freedoms of some for purely or primarily preventive reasons? The best answer to that is that any such policy is justifiable only if: (1) the potential harm was sufficiently grave; (2) the prediction technology was sufficiently accurate; (3) the preventive response was maximally humane and minimally intrusive under the circumstances; and (4) if the preventive action was preceded by adequate due process (Morse 1999, p. 297). For many critics, condition (2) is seldom met given the unreliability of current predictive judgments (Monahan 2004).

Similarly, rehabilitation is not regarded by most contemporary penal philosophers as providing an overall theory of punishment or criminal justice, but that is not to say that it is redundant in sentencing decisions. Two different motivations have sustained rehabilitative programs: first, a humanitarian desire to help those in incarceration many of whom have mental health issues, drug issues, and basic training needs; second, increased frustration with rates of recidivism that spawned the "What Works?" movement – a political movement keen to subject policy to evidence-based evaluation – which revitalized interest in techniques for anger control, the use of cognitive behavioral therapy, and other techniques of behavioral modification. This focus on the "good done" by legal processes (including arrest, trial, and punishment) has developed into a small but significant movement under the title of "therapeutic jurisprudence", which (as the name suggest) aims to assess criminal justice institutions against a standard of therapeutic usefulness (Wexler 2008; Winick 1997).

In short, incapacitation and rehabilitation remain important parts of public policy and sentencing practice. But, they do so largely as discrete elements and not as the result of an overall theory of punishment in which the welfare of society is the only desideratum.

Retributive Theories

Retributive theories of punishment go back (at least) to the great German philosophers Immanuel Kant and Georg Hegel (see ▶ Traditional Retributivism). Following the assault on indeterminate sentencing in the 1970s, and the more general philosophical critique of consequentialist thinking at around the same time, retributivism came back into fashion for the rest of the century. Retributivists hold that punishment is justified when, and because, it is deserved (although the exact sense of desert is contested (Matravers 2011)). More precisely, the argument is that those who culpably commit criminal offenses deserve censure and that censure should take the form of a penal sanction – so-called hard treatment. Most importantly for the development of a retributive sentencing scheme, that hard treatment must be *proportionate* both in the sense of *cardinal* proportionality (the penalty should be fitting or commensurate with the wrong) and of *ordinal* proportionality (crimes of comparable seriousness ought to receive penalties of comparable severity).

The precise way in which retributivists defend their theories varies. For Michael Moore (Moore 1987, 1997), punishment is good when and because it is deserved and this judgment and the associated penalties are a matter of what is dictated by a correct account of morality. For Antony Duff, it is censure that is deserved and hard treatment is merely the vehicle through which that censure is communicated and through which the offender can show penance (Duff 1986, 2007). For Andrew von Hirsch and Andrew Ashworth, retributivism is a mixed theory in which censure addresses us as moral beings and hard treatment offers a "prudential supplement" encouraging us to keep to the law (von Hirsch 1993; von Hirsch and Ashworth 2005).

Proportionality

The principle of proportionality is very attractive. If punishment is an exercise in blaming, then the penalties must reflect this in the sense both that the penalty must be at an appropriate level and that more serious wrongs deserve more blame and so more severe punishment. This also captures a very important commitment to equality and justice often glossed as "treat like cases alike and different cases differently." That is, if two offenders commit crimes of equal seriousness with equal culpability, then they should be punished equally. Moreover, once an account can be given of the anchoring points of the scales (the least and most serious offenses and the least and most severe penalties), it appears to offer a workable translation of a theory of punishment into a sentencing scheme.

However, the proportionality account of punishment is not without its critics. For "pure" censure-based theorists, the addition of hard treatment as a prudential supplement treats citizens as things to be coerced rather than agents to be reasoned with, and threatens to "drown out" the moral message of the law (see the exchange between Duff and von Hirsch in Matravers 1999). However, even those who accept the necessity of hard treatment as a legitimate part of punishment have queried the usefulness of proportionality as a guiding principle. The criticisms in part focus on the theory's workability and in part on its implications for sentencing practice.

A punishment scheme based on proportionality presumes that we can accurately measure offense seriousness and culpability, and that we can then match that to a point on a penalty severity scale. These things pose a significant challenge to proportionality theorists. In relation to offense seriousness, what is needed is a way to put together on one scale crimes such as assault by an individual and commercial pollution. In relation to culpability, in principle, the theory needs to assess the subjective state of the offender, but in practice may have to rely on objective measures such as the absence or presence of previous convictions. Proportionality theorists are aware of these (and other challenges) and have tried to meet them (Ashworth 2010; von Hirsch and Ashworth 2005), but not everyone is convinced that they have been successful (for a helpful summary of the criticisms, see Tonry 2011b, pp. 224–34).

Perhaps even more significant is the need to anchor the scale of penalty severity. The bottom anchoring point – the penalty to be associated with the least serious offense – may be reasonably easy to fix in that many jurisdictions may have something like a police caution. However, what of the upper anchoring point? Should this be capital punishment (or worse), life imprisonment without the possibility of parole, 21 years, or less?

Without an answer to this question, the scheme is inadequate.

Proportionality theorists have reached for a variety of principled, and socially contingent, ways of fixing the upper anchoring point. As noted above, one might worry, for example, that too severe punishments will "drown out" the moral message conveyed by the law. The problem the account faces is that insofar as the penalty conveys proportionate blame, different societies, with different histories and practices, will think different overall levels of suffering needed to do this accurately.

Parsimony

In contrast to the equality and proportionality required by the proportionality account, Norval Morris (1974) and others have argued for a principle of parsimony. By "parsimony" it is meant that punishment levels ought in general to be set such as that offenders receive "the least severe sentence consistent with the governing purposes of sentencing" (Tonry 1992, p. 81). In developing this account, Morris and others have come to think of desert as a "limiting" principle (Frase 2004) such that offenders should receive the least severe "not undeserved" sanction from a range in which the upper point is consistent with proportionality, but the lower bounds allow for individualization of sentences and parsimony. Such an account is still hostage to a theory of cardinal desert, but its relative vagueness would, its proponents suggest, at least allow greater justice to be done to each particular offender; justice that would be sacrificed by the fetishizing of proportionality. Exactly that discretion, though, is what bothers proponents of proportionality (perhaps mindful of the way it was used to disadvantage minority groups in the past).

Theory and Practice

Following the breakdown of the consequentialist consensus in the final third of the twentieth century, broadly retributive theories of punishment took center stage and, by-and-large, remain there (although see the essays in Tonry 2011a). The associated "just deserts" movement, and its various political analogues like the campaign for "truth-in-sentencing," was dominant, but for a shorter period and not without challenge. Throughout the same period, and in particular after the mid-1980s, many sentencing policies that explicitly violated proportionality were passed. Mandatory minimum sentences for drugs offenses, sexual offenses (particularly against children), and possession offenses (particularly of child pornography) have meant that these crimes attract penalties far above those given for what might be thought of as roughly equally serious violent crimes. Another example is California's notorious three-strikes law, which means that relatively trivial third offenses could attract very severe penalties, and there are many others.

As noted above, other less retributive policies – grounded in therapeutic jurisprudence or restorative justice – also flourished in corners of the criminal justice system often because they were perceived to "work"; in particular to reduce recidivism.

Very few of these policies can be said to flow directly from a fully worked-out philosophical position on the justification of punishment. Yet, it would be absurd to say that "theory does not matter." Policies do change as a result of theoretical reflections and will no doubt continue to do so. However, sentencing – perhaps more than any other part of the criminal justice system – is political and pluralistic. It is political in that it is a highly visible and politically controversial. To be "soft on crime" is, in many political systems, to court political disaster (Lacey 2008). It is pluralistic in that, *at* the point of sentencing, judges in systems where there is discretion will often call upon many considerations (an intuitive notion of desert, the need to express the community's blame, deterrence, the risk posed by the offender, the offender's potential for reform, and so on). Such pluralism can sit ill with philosophers of punishment who seek solutions which are logically consistent and morally coherent and defensible.

That said, precisely because crime is a politically sensitive issue, and because in many jurisdictions, judges are limited in their discretion by mandatory sentences or restrictive sentencing guidelines, it is vital that moral and penal philosophers continue to subject sentencing rationales to critical analysis. And, in doing so, they will inevitably call on overall accounts of the general purpose, and justification, of punishment even if those accounts do not always translate easily into particular sentencing outcomes.

Recommended Reading and References

Ashworth A (2010) Sentencing and criminal justice, 5th edn. Cambridge University Press, Cambridge

Bentham J (1843) Principles of penal law. In: Bowring J (ed) The works of Jeremy Bentham, 1st edn. William Tait, Edinburgh

Bentham J (1970) In: Fred R, Philip S (eds) An introduction to the principles of morals and legislation. Oxford University Press, Oxford

Bottoms A, von Hirsch A (2011) The crime preventive impact of penal sanctions. In: Cane P, Kritzer H (eds) The Oxford handbook of empirical legal research. Oxford University Press, Oxford

Doob AN, Webster CM (2003) Sentence severity and crime: accepting the null hypothesis. Crime Justice 30:143–195 (ArticleType: research-article/Full publication date: 2003/Copyright © 2003 The University of Chicago Press)

Duff RA (1986) Trials and punishments. Cambridge University Press, Cambridge

Duff RF (2007) Answering for crime: responsibility and liability in the criminal law. Hart Publishing, Oxford

Feinberg J (1970) The expressive function of punishment. In: Doing and deserving: essays in the theory of responsibility. Princeton University Press, Princeton, pp 95–118

Frankel ME (1972) Criminal sentences: law without order. Hill and Wang, New York

Frase R (2004) Limiting retributivism. In: Tonry M (ed) The future of imprisonment. Oxford University Press, New York, pp 83–120

Goodin R (1995) Utilitarianism as a public philosophy. Cambridge University Press, Cambridge

Hart HLA (1968) Punishment and responsibility: essays in the philosophy of law. Oxford University Press, Oxford

Kleinig J (1973) Punishment and desert. Springer, New York

Lacey N (2008) The prisoners' dilemma: political economy and punishment in contemporary democracies. Cambridge University Press, Cambridge

Lewis CS (1949) The humanitarian theory of punishment. 20 Century Aust Q Rev 3(3):5–12

Matravers M (1999) Punishment and political theory. Hart Publishing, Oxford

Matravers M (2000) Justice and punishment: the rationale of coercion. Oxford University Press, Oxford

Matravers M (2011) Is twenty-first century punishment post-desert? In: Tonry M (ed) Retributivism has a past: has it a future? Oxford University Press, New York

McCloskey H (1968) A non-utilitarian apporach to punishment. In: Bayles M (ed) Contemporary utilitarianism. Doubleday, New York, pp 239–259

Michael J, Adler M (1933) Crime, law and social science. Harcourt, Brace, New York

Monahan J (2004) The future of violence risk management. In: Tonry M (ed) The future of imprisonment. Oxford University Press, New York

Moore M (1987) The moral worth of retribution. In: Schoemann F (ed) Responsibility, character and the emotions. Cambridge University Press, New York, pp 179–219

Moore M (1997) Placing blame: a general theory of the criminal law. Oxford University Press, Oxford

Morris N (1974) The future of imprisonment. University of Chicago Press, Chicago

Morse SJ (1999) Neither desert nor disease. Legal Theory 5:265–309

Posner RA (1985) An economic theory of criminal law. Columbia Law Rev 85:1193–1195

Rawls J (1955/1999) Two concepts of rules. In: Rawls J, Freeman SR (eds) Collected papers. Harvard University Press, Cambridge, pp 20–46

Rawls J (1971) A theory of justice. Harvard University Press, Cambridge

Robinson P, Darley J (2003) The role of deterrence in the formulation of criminal law rules: at its worst when doing its best. Georgetown Law J 91:949

Robinson P, Darley J (2004) Does criminal law deter? A behavioural science investigation. Oxf J Legal Stud 24:173–205

Tonry M (1992) Proportionality, parsimony, and interchangeability of punishments. In: Duff A et al (eds) Penal theory and penal practice. Manchester University Press, Manchester

Tonry M (2011a) Retributivism has a past: has it a future? Oxford University Press, New York

Tonry M (ed) (2011b) Why punish? How much?: a reader on punishment. Oxford University Press, New York

von Hirsch A (1993) Censure and sanctions. Clarendon Press, Oxford

von Hirsch A (1999) Criminal deterrence and sentence severity : an analysis of recent research. Hart, Oxford/Portland

von Hirsch A, Ashworth A (2005) Proportionate sentencing : exploring the principles. Oxford University Press, Oxford/New York

von Hirsch A, Committee for the Study of Incarceration (1976) Doing justice: the choice of punishments: report of the Committee for the Study of Incarceration, 1st edn. Hill and Wang, New York

Walker N (1969) Sentencing in a rational society. Allen Lane/The Penguin Press, London
Wexler DB (2008) Rehabilitating lawyers: principles of therapeutic jurisprudence for criminal law practice. Carolina Academic Press, Durham
Winick B (1997) The jurisprudence of therapeutic jurisprudence. Psychol Public Policy Law 3(1):184–206
Wootton B (1978) Crime and penal policy: reflections on fifty years' experience. George Allen & Unwin, London

Penalties

▶ Sentencing Research

Perceptions of Risk

▶ Rational Choice, Deterrence, and Crime: Sociological Contributions

Perceptual Updating

▶ Bayesian Updating and Crime

Performance Management

▶ Judicial Leadership and Performance

Perpetrator Characteristics

▶ Estimator Variables and Eyewitness Identification

Persistent Offenders

▶ Career Criminals and Criminological Theory

Personal Control: Criminal Behavior, Social Control, Self-Control, Models, Constraints

▶ Control Theory

Personalized Policing

▶ Community Policing

Philosophical Basis of the Forensic Process

Camilla Hald
Danish National Police,
Research & Development, Copenhagen,
Denmark

Overview

In recent years controversies have arisen over the definition and demarcation of science in relation to forensic science. These controversies are among other expressed in debates following a report from the National Academy of Sciences in the US, questioning the scientific status of forensics and its interpretations of evidence before courts of law. The controversies highlight an apparent incommensurability of the two fields of knowledge straddled by the forensic process: science and law. These controversies are however as old as science and law, respectively, and can be traced in the major philosophical principles' underpinning, shaping, and informing the origins and the characteristics of the forensic process. The philosophical basis of the forensic process consists of a meshwork of ideas. It originates from and is influenced by a range of different disciplines, models of inquiry, and forms of validating truth claims, which have been decisive in bringing the discipline about. Some major ones runs from divination, classical rhetoric, theology, law, philology, philosophy, history and the arts, over the model of Galilean science, medicine, psychotherapy, art connoisseurship, the infant social science and statistics, to natural sciences such as biology and chemistry, and, new philosophies of science born from these disciplines. Following the trajectories of thought illuminates how there is a philosophical basis to the controversies. The forensic process is informed by ideas from two different models of science. One is a model of science based on identifying and interpreting "signs", the other is one based on observation and verification of general facts. Each model plays a significant part in the workings of the forensic process. The lack of acknowledgement of the contributions from each, however, plays a major part in the controversies. Hence the way forward may prove to be related to attaining a more nuanced understanding of the philosophical roots of the forensic process as well as in incorporating each model equally into the theoretical and operational framework of the discipline.

Key Issues/Controversies

The forensic process is the process of inquiry belonging to forensic science. Forensic science is most often defined as an applied science, whose raison d'être is to provide scientific analysis of evidence in the service of law. The forensic process is part of the investigative inquiry carried out within the legal process which is oriented toward providing answers to questions essential to the legal deliberation and arbitration of cases. This however is not a straightforward business, since the questions asked and the methods of supplying answers differ markedly between law and science.

Put crudely, the ultimate goal of legal inquiries is to establish "what happened" (how, when, where, and why) and "who did it" in a given case under scrutiny. The goal of inquiry of the sciences is however rather to establish "what a given phenomenon is" and "how it works." Also the methods of inquiry, and for validating "truth" differ between law and science. Legal inquiry makes use of qualitative methods, testimonial evidence, and a judgment of and deliberation between, probable arguments derived here from, whereas the sciences use observation and experiments, favoring the quantifiable, reproducible, and empirically verifiable.

Differences between law and science are grounded in a difference between the object and goal of inquiry of each discipline. The legal process deals with questions of a social, moral, and cultural kind and in the deliberation of individual cases in which truth is of an approximate kind as versed in the dictum "beyond reasonable doubt." Natural science (with which forensic science is most commonly associated and from which the most influential forensic techniques and forms of analysis originates: physics, molecular biology, chemistry, etc.) deals with questions about nature, its laws, and general characteristics to the point of establishing the truth about it in a more absolute sense. The knowledge sought by science is one which may be reproduced independent of the particular researcher or situation, and from which it is possible to predict results based on true premises. Thus for law the goal of inquiry is the most reasonable explanation for particular events, while the goal of scientific inquiry is the universal truth about "what is." These two fields of knowledge is straddled by the forensic process, in which legal questions must be translated into scientific questions in order to provide scientific answers which may inform legal argumentation and deliberation (see Fig. 1). Forensic science and its role in legal deliberation is in other words a matter of using science in the service of law and legal deliberation, to help establish "matters of fact" which may aid the process of deliberating between "probable arguments".

Evidence produced through the forensic process is thus also submitted to different forms of validation in different contexts and phases of the process. Within the laboratory, evidence is evaluated according to the soundness of analysis and integrity of the process of analysis (an uncompromised chain of custody and accuracy

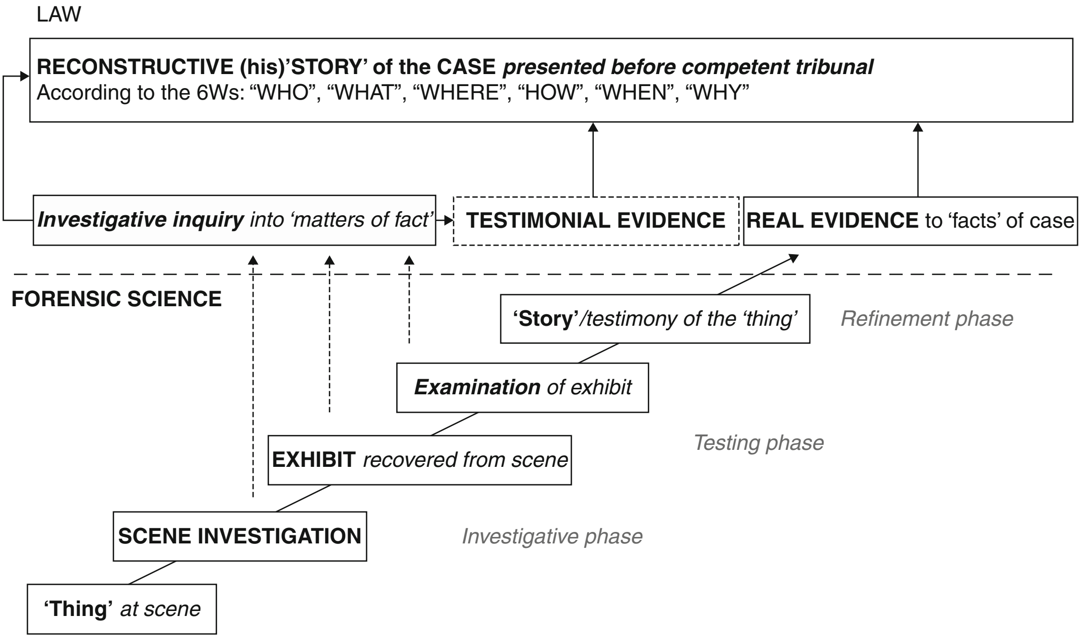
LAW
RECONSTRUCTIVE (his)'STORY' of the CASE presented before competent tribunal
According to the 6Ws: "WHO", "WHAT", "WHERE", "HOW", "WHEN", "WHY"
Investigative inquiry into 'matters of fact'
TESTIMONIAL EVIDENCE
REAL EVIDENCE to 'facts' of case
FORENSIC SCIENCE
'Story'/testimony of the 'thing'
Refinement phase
Examination of exhibit
EXHIBIT recovered from scene
Testing phase
SCENE INVESTIGATION
Investigative phase
'Thing' at scene

Philosophical Basis of the Forensic Process, Fig. 1 The forensic process – science in service of law (From Tilstone et al. 2012)

of instruments and methods applied) in order to assure the scientific "rigor" of the results through recognized methods of analysis and scientific procedures for testing. Within the legal process, the evidence is in addition to the assurance of the integrity of analysis and evidence, evaluated for their *meaning* in relation to the legal argumentation presented before a court of law. The former demands that what is infered from the evidence is verifiable and stand up to tests in accordance with scientifically acknowledged theories and methods. The latter demands that evidence is seen in the context of the case and the circumstances surrounding the actions or event in question and is based on evaluation of possible or probable causes or sources of particular forms of evidence.

The evidential value of the information provided through scientific analysis and the professional judgment of the analyst is, in other words, ultimately decided in the context of the case and the legal process of deliberating between the arguments presented by the advocates of prosecution and defense.

This fundamental condition of the forensic process has resulted in intense debates about the role of the forensic scientist in providing service to law by analyzing and interpreting physical evidence. Questions such as: How this is done; where the boundaries of professional judgments in interpreting the evidence, are set and how far the forensic analyst may stretch in order to help answer the questions of the legal process – and how the authority and integrity of the scientist should be assessed – are however subject to different interpretations. Within forensic science itself, tensions are evident in the way that different schools have arisen, especially within criminalistics, each school emphasizing the role of the forensic analyst differently, as well as the degree to which criminalistics should be defined by the needs and methods of law versus those of the pure sciences (Inman and Rudin 2001).

These inherent tensions have become more obvious in the United States, in the wake of a report from the National Academy of Sciences: *Strengthening Forensic Science in the United States: A Path Forward* (2009), evaluating the field of forensic science. The report raises many issues and questions echoing debates, which have a long history – both within forensic science disciplines and beyond. These are issues such as: that the current forensic science system is not based on science; that the analysis is frequently subjective; and that the same technique used by different analysts can lead to different results, just as the same technique used by the same analyst on the same sample at different points in time and in different contexts can lead to a different result, and that anecdotal information, phrased like "In my training, knowledge, and experience. . . ." makes its way into the courtroom. Reactions to the NAS report have been swift and all across the board from those that deny that there is anything wrong, to those who have adopted the NAS report and endeavored to suggest new modes of operation to improve on the lack of scientific rigor and culture of forensics, especially within criminalistics (e.g., Mnookin et al. 2011).

The report does however not deal with the fundamental philosophical basis of the forensic process, Therefore there is no recognition of the philosophical roots of this debate, and the fact that the issues have existed in different forms since well before the development of the different disciplines of forensic science. In order to understand the controversies a proper historical and theoretical context is needed. The purpose of the entry is therefore to explore the major philosophical principles that have inspired and shaped the forensic process. Given that the word "forensic" in contemporary use has two meanings: (a) as a form of legal evidence and (b) as a short term for forensic science or criminalistics (forensics), this entry will inspect the philosophical history along two trajectories. Firstly, through a visitation of the history of legal rhetoric – the predecessor of modern law – secondly, through the history of the growth of science and the scientific method and the philosophical "goods" being, respectively, embraced and discarded in the process. Ultimately, it is in the combined history of these two trajectories of thought that the philosophical basis of the forensic process is to be found.

Fundamentals of Legal History

The Art of Rhetoric: The Foundation of Modern Law

Etymology is a proper place to start when digging for the roots and the philosophical basis of the forensic process. The word "forensic" originates from the Latin adjective *forensis*, meaning "of or before the forum." In Roman times, a criminal charge meant presenting the case before a group of public individuals in the forum. Both the person accused of the crime and the accuser would be represented by an advocate giving speeches on their behalf, presenting their side of the matter. The individual with the best argument and delivery would succeed. This origin is the source of the modern usage of the word "forensic" to denote a form of legal evidence (forensic evidence) and as a short representation of "forensic science," that is, science related to and of relevance to courts (forensics). Historically, courts of law (forums or public assemblies) were intimately connected to rhetoric. In classical Greece, rhetoric was taught to citizens so that they could defend themselves in the court and deliberate in the assembly. In Rome, advocates (called logographs) used oratory to plead on behalf of their clients before the forum. Law, as a subject of its own, did not emerge until the twelfth century (Schoeck 1983), and law students continued to be educated in rhetoric well into the Early Modern period (Balkin 1996). A key element of the art of rhetoric was how to find and present arguments in a case to be presented before a court of peers. To guide the search for arguments and proofs to their fact, which would persuade the court to find the case presented by the rhetorician the most believable and worthy of support, the Ancient Greeks developed a meticulous system or inventory of possible forms and modes of argument to this end, also known as *topoi* or in English "topics." The following presents a brief history of the origins of the topics and their incorporation into modern legal procedure and the forensic process of which the topics form the backbone.

The Topics of Classical Rhetoric

The first attempts at setting out guidelines for those who were to investigate and present cases before a tribunal of decision makers seems to be dated back to one of the deemed founders of the modern art of rhetoric, Hermagoras of Temnos (300 BC). Hermagoras divided the materials of rhetoric into two parts: *thesis* and *hypothesis*. A *thesis* involves an abstract, general question, whereas a *hypothesis* involves a question concerning concrete particulars. Hermagoras provided a list of circumstances which needed to be addressed in order to form hypotheses and to be able to demonstrate their truth-value, in the form of a list of questions. He contended that no hypothetical question or questions involving particular persons and acts can arise without reference to these circumstances, and no demonstration of such a question can be made without using them.

The questions defining Hermagoras' circumstances are known only in Latin because they arose as part of the theological writings of St. Augustine recognizing Hermagoras as the original source. St. Augustine quoted Hermagoras' questions in Latin as follows: *quis, quid, quando, ubi, cur, quem ad modum, quibus adminiculis* (who, what, when, where, why, how, in what way, and by what means/which aids) (Robertson 1946). Hermagoras' list of questions for producing hypothesis and proof is a mnemonic device intended to provide a systematic methodology to the task of producing rhetorical evidence (proof). This form of mnemonic belongs to a part of the methodology of the discipline of rhetoric, derived from Aristotle, known as *topoi* – or "the topics" – of which he was the first to write systematically about.

The word "topic" originates from the Greek word *topos* (pl. *topoi*), meaning "place," and refers to the "places" (on the mind's horizon) where one can go to look for and discover the relevant arguments (i.e., get ideas for argumentation). In Latin, the topics are referred to as *loci* (places) or *loci communes* – meaning the "common places." The use of the word *topos*/*topoi* may refer to the ancient mnemonic principle of using "spatial locations" or "activities" by which to remember even complex or huge amounts of information. One may for instance remember a long list of dinner guests by imagining them around the table. Generally, the word "topics" is commonly used to refer both to the actual "places" where one can go to find material for arguments and to the methodology of the process of searching these places. The topics have since their very conception in classical rhetorical theory been a very confused concept, not easily defined and incorporating a bewildering diversity of meanings. Initially, "Topics" is the title of a treatise of Aristotle on the art of dialectic – the invention and discovery of arguments in which the propositions rest upon *commonly held opinion* or what he termed *endoxa* (Aristotle 384–322 BC). These commonly held beliefs, the *topoi* or "topics," are not merely popular notions held by any given individual in a city or culture about any and all subjects. Rather, they are elements of reason upon which those who conscientiously dispute agree in principle, i.e., that which is firmly embedded within opinion or belief among those who engage in disputation. The topics belong according to Aristotle to the art of rhetoric, in which they inform the invention and discovery of arguments (Aristotle and Lawson-Tancred 1991).

Aristotle's theory of the topics emphasized a particular aspect of evidence related to forming and proving hypotheses. Hypotheses was thought to belong to the form of inference pertaining to "practical knowledge" (techné and phronesis) which was opposed to "theoretical knowledge" (episteme). Aristotle contended that when lacking the certainty and necessity of the true premises described in the syllogism and logic, one needed to take point of departure in commonly held beliefs which could be viewed as "law-like" in nature because of their general acceptance as "true." The concept of topics to Aristotle was thus *both* (a) a minutely devised inventory or system of cases, in which one might profitably take point of departure when devising the argument and providing for them and (b) the *theory of hypothesis building and testing* within the area of rhetoric, to which they pertained. Thus when a "belief" is considered of a high level of generality, i.e., being "common sense", it belongs to the *topoi*.

The Philosophical History of the Topics

The ideas of rhetoric conceived by the Ancient Greeks were spread and became known to the rest of the European continent up through the Middle Ages and during the Renaissance. They were distributed through the works of Roman advocates and orators and the treatises they drafted in the language of church and scholars: Latin. A key figure in the history of European rhetoric and philosophy, linking Classical philosophy and Medieval and Renaissance thinking, was the Latin rhetorician Cicero (106–43 BC) and his works *Topica, De Oratore* and *De Inventione*.

Cicero divides his work on the circumstances into two parts: attributes of the person (*personae*) and attributes of the act (*factum*). As attributes of the person, he lists *nomen, naturam, victum, fortunam, habitum, affectionem, studia, consilia, facta, casus* and *rationes* as the topics to consider in building a credible statement of the case, but more importantly, to develop the arguments which should make the main body of the text or speech, and according to which the logical arguments of the case should be elaborated. Attributes of the action are of three kinds: those concerning the performance of the action itself, those related to the action, and those consequent upon the action; only the first of these reflects the system of Hermagoras. In relation to the action itself, Cicero names *locus, tempus, modus, occasio* and *facultas* (Robertson 1946).

The most important discussion of the topics, considering their relation to legal rhetoric, is made by Anicius Manlius Severinus Boethius (ca. 480–524 AD) in the early Middle Ages. Boethius, in his influential writings *De differentiis topicus*, makes the seven topics (or circumstances) fundamental to the arts of prosecution and defense. The *loci* proper to rhetoric are confined to the attributes of the person who is called for judgment and those of the act or statement for which he is to be judged. All arguments, either in the defense of the person or for his prosecution, must be taken from these materials. Thus, Boethius associated the topics with the process of determining the extent of guilt of a person accused of violating legal rules or customs, which may be significant to the adoption of these by papal forums and priests in their work with ascertaining as accurately as possible the magnitude of each sin confessed to them.

The topics arise in a new guise in the pastoral and theoretical theology in the thirteenth century, as Mnemonic verses consisting of variations of Hermagoras questions. This is evident in the 21st canon of the Fourth Lateral Council arranged by Pope Innocent III, which stipulated that confessors should diligently seek out *et peccatoris circumstantias et peccati* (the circumstances of the sin and of the sinner) – resembling to no small extent Cicero's *person* and *act personis et negotiis* – in order that they might justly weigh the sins confessed to them and administer suitable remedies. This appropriation of the topics to theological matters and theological inquisitorial processes can be attributed the theological writings of St. Augustine (Augustine of Hippo) and Thomas Aquinas. Verses made of the questions in different variations, afforded a flexible instrument for interrogation. It allowed the priest to associate conventional theological divisions of the sins with the questions so that they afforded a frame of reference against which to place his theological knowledge. In this way they provided a practical device for spontaneous and quick analysis (Robertson 1946, pp. 6–7).

The topics were then revived by Renaissance rhetoricians, such as Thomas Wilson, *The Art of Rhetorique*, who made an English verse of them: *"Who, what, and where, by what helps, and by whose: Why how, and when, doe many things disclose*" (cit. in Robertson 1946, p. 14). A principle of such fundamental importance to both rhetoric and theology soon spread to other areas of literary work and studies. Italian writers adapted them to historiography. One of the important figures in Renaissance rhetoric Giambattista Vico (1668–1744), trained in jurisprudence, but widely read in Classics, philology, and philosophy, all of which informed his highly original views on history, historiography, and culture, revisited the topics in his mature work: *Scienza Nuova* (or The New Science). In this work he argued for a central role of the topics in all disciplines including the sciences – even mathematics.

Vico's point was that "topics" is not to be learned from a book but is more akin to a disposition whereby one has analogies and arguments at one's hand. Thus for the doctor the topics provide ready mental access to the stock of medical precedents that have historically enabled successful diagnosis and treatment, and to the mathematician pressing the boundaries of knowledge further, the topics consist of the accepted theories preceding the development of new hypotheses. Vico found that the principle described in the use of the topics was to be considered a necessary part of making judgments about everyday life in general, and that their cultivation is required for creative discovery and quick action in *all* disciplines, from physics to law.

In his own time, Vico was relatively unknown, but from the nineteenth century onward where the methodology and philosophical content of rhetorics and linguistics converged with other developments in the history of ideas giving rise to new critical theories in the philosophy of science, his views found a wider audience. Today his influence is widespread in the humanities as well as in the social sciences. Before delineating this turn, tying the knot between the two histories of ideas by describing how they inform the growth of forensic science and the philosophical basis of the forensic process as it looks today, a brief history of the fundamentals in the development of science is called for.

Fundamentals of the Philosophical History of Forensic Science

Science and the Conjectural Model: The Foundation of Forensic Science

Forensic science is intimately related to investigative inquiry through the forensic process. Investigation comes from Latin *investigationem* meaning "a searching into, a searching for." The noun stems from *investigare* "to trace out, search after" which again is composited from *in* "in, into" and *vestigare* "to track, trace" from *vestigium* "footprint, track." The etymology of the word "investigate" carries the seed to an understanding of the philosophical roots of crime investigations and the forensic process, by referring to "tracks and traces" and "footprints," since the roots of science and of medical, legal, and forecasting sciences in particular, may well be rooted in the first efforts of interpreting footprints in the sand or the tracks of wildlife in the forests, with the aim to hunt it down and slay it. The point being that the philosophical basis of the forensic process has a deep-rooted connection to semiotics and other sciences informed by what may be termed the "semiotic paradigm" – which we shall get back to shortly. Firstly, let us briefly return to etymology as a starting point for the philosophical basis of the forensic process stemming from the history of science.

The root of the word "science" is the Latin *scire*, to know, and hence *scientia*, or knowledge. Before the late sixteenth and early seventeenth centuries, the word denoted any kind of knowledge, undiscriminating between different forms or types of knowledge. However, with the Enlightenment and the introduction of scientific method plus the resurgence of learning based on classical theory during the Renaissance, the word *scientia* came to mean knowledge of the natural world and became increasingly associated with natural science.

The Italian historian Carlo Ginzburg traces the trajectory of what he terms the "conjectural or evidential paradigm" throughout European history, attempting to excavate the historical lineage of "disciplines based on reading the evidence" (Ginzburg 1980, p. 14). Since this history is rarely researched, the exposition of this part of the philosophical history of the forensic process is resumed from his seminal piece: *Morelli, Freud and Sherlock Holmes: Clues and Scientific Method* (1980). The roots of this paradigm, Ginzburg argues, can be found well beyond the civilizations of Ancient Greece with the earliest hunters and the skills they developed for reading *signs* in nature to track down the game and in the Mesopotamian cultures different forms of divination. These skills he argues, played a significant role in the invention of writing and the capacity for abstract thinking, which is based on the use of purposefully crafted signs to represent inherent meanings. The common characteristics of these roots of skills leading onward to modern science, are that, both traditions, hunting and divination, "require minute examination of the real, however trivial, to uncover the traces of events which the observer cannot directly experience" (Ginzburg 1980, p. 13). The way hunters used their senses – sniffing, listening, and observing – to give meaning and context to the slightest trace, "in the shadowy and treacherous clearing," Ginzburg argues, endowed them with the ability to make complex calculations in an instant. Hunters inferred from part to whole, effect to cause. The Mesopotamian model of divination gradually intensified this tendency to generalize from basic facts, thus strengthening this mode of inferring cause from effect. The model for explanation and divination decipherable in Mesopotamian texts is one which could be oriented toward past, present, or future, depending on which form of knowledge was called upon. Seeking signs to form knowledge oriented toward the future, was divination proper. Orientation toward the past was related to jurisprudence or the body of legal knowledge. Divination oriented towards the past, present, and future in concert was that of medical knowledge of symptoms. The latter contained all of the orientations by way of its double character of diagnosis, explaining past and present, and prognosis, suggesting the likely future following from the illness running its course.

The conjectural paradigm lies at the root of modern science in general. It is, however, the development into scientific disciplines of these original practical *arts* (hunting, divination, medicine, and law) of deciphering various signs and symptoms as well as individual cases to infer generalized meanings, which serves to characterize the philosophical basis of the forensic process in the Modern Age. It is however not well researched nor generally acknowledged how these artistic aspects of producing knowledge inform the growth of science and ultimately forensic science. Retracing this history is however the goal in the following.

Ancient Greek Philosophy: The Cradle of Modern Science

What characterized the Greek culture and the city-states, of which Modern European legal culture and science are heirs, was the emergence of a group of disciplines which all "depended on the deciphering of various kinds of signs, from symptoms to writing" (Ginzburg 1980, p. 14). Passing on to the civilizations of ancient Greece, these disciplines changed considerably, giving rise to new lines of study such as history and philology while acquiring also independence from older disciplines, such as medicine. In and through these disciplines, one finds the seeds of the development of a general scientific stance or method. Through these disciplines the "body" (e.g., medicine), "speech" (e.g., rhetoric, philology), and "history" were all for the first time subjected to dispassionate investigation, excluding the possibility of divine intervention. An important part of this change can, according to Ginzburg, be ascribed to a model of knowledge acquisition based on symptoms and signs – the evidential or "conjectural model" of knowledge.

This development is most evident in the case of Hippocratic medicine. Hippocrates' model of medicine based its central methods on the concept of "the symptom", in Greek *semeion* (hence semiotics). The central tenet in the Hippocratic model was the idea that by carefully observing and registering every symptom, it was possible to establish precise histories of each disease, despite the fact that the disease as an entity would remain intangible. The insistence on the evidential nature of medicine, prevalent with the Hippocratic model, probably stemmed from a distinction between the certainty of divine knowledge and the provisional, conjectural nature of human knowledge, expounded by the Pythagorean doctor, Alcmaeon of Croton in his work *Peri Physeos* (or On nature). This was the first scientific treatise devoted to the art of medicine. The work itself has not been preserved, but its contents and chief findings have been passed on by commentators and collectors of ancient knowledge such as Aristotle, Theophrastus, and Aetius. From these it is known that Alcmaeon held that if reality was not necessarily clear – then by implication – it was right to proceed by building up knowledge of the whole from the parts, using conjecture.

Proceeding by using conjecture was in fact the most common approach in a number of spheres of activity in Ancient Greece, where physicians, historians, politicians, potters, joiners, mariners, hunters, fishermen, and women in general were held to be adept in the vast areas of conjectural knowledge. This conjectural or semiotic paradigm, continued however to be merely implicit up through the ages, since it was completely overshadowed by Plato’s theory of knowledge, holding sway in more influential circles, hence shrouded in more prestige. Plato’s theory of knowledge emphasized the existence of an ideal (i.e., ideational) world of universal untainted truth existing independently of the human species and lost to the human subject. To Plato, knowledge could only be regained or recollected (*anamnesis*) through inspired contemplation and philosophy.

The Science of the Universal and the Problem of the Individual

While Ancient Greece and the developments of disciplines of dispassionate investigation of various subjects and forms, may be the cradle of modern science, the genesis of *scientific method* is generally acknowledged to be the Middle Ages. The emergence of a truly scientific paradigm is commonly associated with Galileo (1564–1642). The scientific paradigm based on, but outliving, Galileo, provided significant changes in the concepts of “rigor” and “science.” The model of Galilean science was fundamentally different from any of the disciplines using conjecture to form knowledge of their subject. None of these would meet the criteria of scientific inference set out by the Galilean approach. The ancient disciplines were concerned with the qualitative, the individual and the case particular, i.e., the individual body, document or situation itself. Galilean physics were altogether different. It used mathematics and experimental method to gain knowledge of the universal order of nature. This was in stark contrast to the conjectural model of the Ancient disciplines in the results of which there was always an element of chance (Ginzburg 1980, p. 15). In the science that grew from the Galilean model, chance was sought tamed by a dependence on measurement and repetition of results. Creating knowledge by induction from experimental design and observation of results, science increasingly became associated with the need to measure and repeat phenomena in order to draw rigorous conclusions about the phenomenal world. Paradigms following the Galilean tradition were concerned alone with knowledge of the universal and generalizable, while being unable to say anything about the individual and particular. The rise of the scientific paradigm resulted in a division of fields of knowledge which would meet the criteria of scientific inference set by and essential to the Galilean approach, and those who could not. A host of the classical disciplines could not meet or make do with the criteria set for science by the new scientific paradigm. However, one discipline in particular diverged from the rest, proclaiming its scientific status while drawing on a model of inquiry stemming from the conjectural or semiotic model – medicine.

The philosophical history of forensic science leads in many ways directly to an origin in medicine. This is generally acknowledged not least in the fact that subjects of medicine such as the autopsy and toxicology are some of the first forms of knowledge enrolled in the service of law to decide on the issues of identifying and discriminating between a criminal act and a legitimate or natural one. However, there are more subtle philosophical roots to be discerned from the relation between medicine and the forensic process as a process of inquiry and theory of knowledge.

The Rise of the Semiotic Model

The trajectories of thought and ideas informing the move from hunter to the scientific detective are manifold and diverse. It is therefore not possible to delineate them in detail here. However, their different results may be indicated by zooming in more directly on the common denominator through which they all at some point pass – medical science and the paradigms informing it. The roots of this paradigm may, following Ginzburg, be traced to the contemporaries of Galileo and described through a historical event in which the medical expertise was enrolled to decide on the nature of a new phenomenon: a two-headed calf born outside of Rome in April 1625. The event became object for discussion among a group of naturalists in the Vatican Belvedere Garden, which was documented in detail by Giovanni Faber (a friend of Galileo and secretary of the Lincei Academy, a center of a thriving intellectual group founded in 1603 by Frederico Cesi). The discussion was attended besides Faber, by another friend of Galileo, a Cardinal, Pope Urbano VIII, and the pope's chief medical man, a physician from Siena, Giulio Mancini (Ginzburg 1980, p. 20).

Mancini was besides being a brilliant diagnostician renowned for his ability to divine the seat of disease by a rapid look at the patient, also a keen intellectual and the author of a book about art recognition titled: *Considerazioni sulla pittura* (or Some considerations concerning painting as an amusement for a noble gentleman and introducing what needs to be said) (Mancini ed. Marucchi 1956–1957 ref. in Ginzburg 1980). Part of this book was devoted to the recognition of paintings, and set out a method for telling originals from copies, by recognizing the particular strokes of the hand of the master in areas of the painting that were swiftly executed. The underlying assumption of Mancini's views, were never stated, but were nevertheless important to the later development of *connoisseurship* in the century to follow and its role in transmitting the semiotic model to other areas of European cultural life and knowledge, Ginzburg argues. The core assumption of deep relevance to the rise of the conjectural paradigm was that there are ineradicable differences between a canvas painted by the hand of Rafael, for example, and any copy of it (Ginzburg 1980, p. 20).

The discussion between the parties participating in the event represents a historical example of differing perspectives of different disciplines on the individual case and its relation to the general, as well as a moment in the history of a converging of differing paradigms, embodied in the practitioner of medicine and art connoisseur Giulio Mancini. The first question addressed by the group was whether the animal should count as one animal or two. For the physicians the feature distinguishing the individual was the brain, whereas for followers of Aristotle it was the heart.

To some extent, Mancini in the discussion of the two-headed calf represents a point of contact between the divinatory approach of the conjectural paradigm (as a diagnostician and a connoisseur) and the generalizing model of the Galilean approach to physics (as anatomist and naturalist). He represents the emergence of a generalizing paradigm within medicine while at the same time being the skilled artistically inclined diagnostician, reading the signs of the disease in the bodies of the patient or identifying the hand of the master versus that of the copyist in a piece of art. By adopting the role of the physician, Mancini was ascertaining the character of the individual as a "type" (i.e., within a scheme of classification) rather than of the character of the individual calf itself. The physician, in other words, participated with a view to arrive at a more accurate definition of what was normal – and therefore *repeatable* for the individual of the species, i.e., establishing the "common character" of it rather than the one peculiar to it (Ginzburg 1980).

The medical paradigm that developed through the eighteenth and nineteenth centuries thus consisted of two paradigmatic models – the anatomical model and the semiotic model (Foucault and Sheridan 1973). The first aspired for a system of knowledge which could be generalized and ascertained through empirical observation and classification (continuing the natural history starting with Aristotle) and making use of the organ associated with the sciences of the time, which could not be reduced to numbers and mathematical calculation: *sight*. Vision and observations became synonymous with "objectivity" – the defining character of empirical science (Daston and Galison 2007).

The second model of medicine aspired for an effective diagnostics which were able to identify and treat illnesses in individual cases (Foucault and Sheridan 1973) and thus keep the subject alive, healthy, and maintain a productive population to the benefit of the state (Foucault and Sheridan 1979).

The conjectural or semiotic model of medicine was, however, intimately connected with knowledge of the general, in the form evoked by the anatomical model. The latter was based on the dissection of corpses and the description of the anatomy of the human body. This practice was used to create knowledge of the general character of the human body, which in turn could be used as a general norm against which pathology could measure the particular body's deviance or abnormalities to achieve new knowledge of pathologies and their effects and signs in the human body. The anatomical model, and its knowledge of the general achieved from observations and generalizations of many individual bodies was promoted and spread through the medical atlases, which became a significant artifact through which medical science flourished (Daston and Galison 1992). It did so in coexistence and mutual development with a semiotic model of medicine, based in the clinic, and the experience and knowledge of the emergence, courses, symptoms, and effects of different illnesses, which arose from the diagnostician's effort to device the proper treatment in each individual case (Foucault and Sheridan 1973). In a similar way, the discussions and the tiny incisions made in the animal to reveal the internal organs representing the roots of the modern scientific autopsy, which developed during the Renaissance, represents the meetings of divination and science in the anatomical model of medicine. In a review article of the history of the autopsy, King and Meehan (1973) find its roots in former practices of divination using animal organs, the so-called *haruspicy*. Autopsy, which literally means "seeing for oneself," reflects the central role of sight referred to above, as the key component of the scientific method of "observation" and scientific objectivity. "Seeing," however, here has roots of a different kind, namely, that of ancient diviners looking for signs to foretell the future by the reading of organs and tissues of dead animals. The medical atlases of the nineteenth century thus also had a predecessor in Babylonian models of livers with diagrammatic markings used for the instruction of diviners.

The predicament of medical science, that is, its apparent lack of absolute scientific rigor and certainty while aspiring to be a science proper, can be said to have two basic reasons. On the one hand, there is a discrepancy between theory and practice. Theoretically adequate descriptions of particular diseases are not necessarily so in practice, where diseases can present itself in different ways in different bodies. On the other hand, knowledge of disease in the *living* body remains indirect and conjectural, since it is per definition out of reach (Ginzburg 1980, p. 21).

Medicine and the Birth of the Detective

It was however from the diagnostic branch of medicine that the ideas giving birth to both detective fiction and forensic science originated. In the late nineteenth century, there was a general upsurge in the semiotic paradigm, giving birth to the practice of art connoisseurship, the theories of psychotherapy, and the detective novel. There are actual biographical and inspirational links between the key figures representing each of the three, in the form of Giovanni Morelli, Sigmund Freud, and Arthur Conan Doyle. The connecting factor between them is *medicine*. Freud was a doctor, Morelli had a degree in medicine, and it is generally acknowledged that the inspiration for Sherlock Holmes was Sir Arthur Conan Doyle's esteemed professor in medical school (he was a doctor before he started writing) Professor Joseph Bell. Bell was renowned for his diagnostic abilities, and there are key passages in the stories of Sherlock Holmes, for example. *The Cardboard Box* (1892), where Holmes uses the actual methods of Morelli. Furthermore the peculiar similarity of the activities of Holmes and Freud have been discussed in literature (e.g., Marcus 1976).

At the same time however, they represent a point in history where the semiotic model (or the conjectural paradigm) resulted in different ideational developments, all however sharing a common feature, namely, that they were based on an interpretive method where one takes the evidence at hand as "signs" which stands for something else otherwise inaccessible for observation. In psychotherapy, art connoisseurship, and detective fiction, methods were devised which made tiny details the key to a deeper reality or knowledge hereof that are inaccessible by other methods. These details may be *symptoms* for Freud, *clues* for Holmes, or *features* of paintings for Morelli, but in all three cases they are seen as evidence to inferences about the nature and characteristics of something hidden to the inquirer: the psyche, the identity of the criminal and the structure of the criminal event, or the identity of the painter and the characteristics of his paintings, behind a given piece of art. Sir Arthur Conan Doyle's character Sherlock Holmes has been deemed inspirational for one or more of the pioneers in forensic science in the area of criminalistics, such as Bertillon and Locard. Whether or not this is a historical fact, we encounter also in the foundation of the field of forensics the medical common factor, as Locard's teacher and mentor was Alexandre Lacassagne, a physician and chair of Médecine Légale de la Faculté de Lyon, and, Bertillon the son of a medical professor.

However, developments within the natural and social sciences occurring at the same time were also influenced by the conjectural paradigm and the problem of the individual. Thus the semiotic or conjectural paradigm generated forensic science from several trajectories of thought, which in concert gave rise to the idea of the scientific detective and the forensic process.

The Social Sciences and the Rise of Technologies of Government

The problem of the individual in the generalizing Galilean model of science was encountered from another field of knowledge, namely, that of the emerging social sciences. During the eighteenth and nineteenth centuries, a tendency evolved for the state to impose a close-meshed net of control on society: one based on knowing its individual members and submitting them to forms of surveillance that became increasingly more subtle and that was based on knowledge of its population, statistical as well as biographical (Foucault 1970; Foucault and Gordon 1980; Foucault et al. 2007, 2008; Foucault and Sheridan 1979).

With the emergence of a capitalist economy based on production in the late eighteenth and early nineteenth century, new concepts of property and the introduction of a still greater number of punishable offenses and punishments of more severity were introduced. Thus, prison systems were built up to contain the resultant growing population of criminals. With the rise in the number and types of crimes came the problem of identifying repeat offenders. This problem became a specific subset of the more general one of keeping a complete and general track on the whole of society, that is, keeping order through knowledge. It also represents the point where the models derived from Galilean physics start to impinge on and interact with the semiotic model of medicine, in the development of new forms of control and methods for identification, devised in the same period of the 1800s as the detective novel, psychoanalysis, and connoisseurship evolved.

This was evident in two competing systems for identification initiated or developed by Francis Galton (1822–1911) and Alphonse Bertillon (1853–1914), respectively. In both cases, however, science meets conjecture to give rise to the methods of scientific investigation based on classifications of the individual and interpretations of effects of specific causes in particular cases.

In the first case, the system of Bertillionage, a classification scheme for identifying repeat offenders through anthropometric measures plus the description and classification of physical characteristics, was based on the idea of probability statistics used within astronomy and physics to get a measure of errors, introduced to the study of society and crime by Adolphe Quetelet (1796–1874) in his conceptualisation of a "social physics". Quetelet's work provided a major inspiration to Bertillon and the anthropometric method, by suggesting that while single measures might be just probative, the higher number of specific measures made per individual the higher the probability of identification.

In the second case, the system of fingerprint identification introduced by Francis Galton, yet not originally discovered by him (for a history of fingerprint identification, see Cole 2001; Sengoopta 2004), was based on the idea that the individual has an identity that may be recognized in his every characteristic, even the most imperceptible and the slightest, and that there will be an internal norm or "typus," which maintains the variety of each species within its limits. The philosophical input, as well as the study of fingerprints as the empirical example of the rule, was made by a physiologist founding the subject of study of organ tissues: Purkyné in a work in 1923.

Bertillon, being the son of a medically educated statistician and inspired by an influential figure in the emerging science of society especially criminology developed by Adolphe Quetelet, represents the meeting point of the conjectural model derived from medical semiotics and that of divining probability through the science developed to this end: statistics. Galton's discoveries of the individuality of the fingerprint and its possible practical implications, on the other hand represent the point of contact between old divinatory traditions; the problem of identification of the complex society and the risks posed by the transgressor of laws and regulations, and the anatomical model of medicine and its attempt at understanding the relation between individual and general type, in order to diagnose individual symptoms and secure the proper treatment in each individual case.

Forensic science arose from the interconnections of these ideas and trajectories of thought spanning detective fiction, graphology, art connoisseurship, statistics, sociology, criminology, administrative practices, psychology, linguistics, history, etc., as they were developed and promoted in the works of Hans Gross, Alexandre Lacassagne, Edmond Locard, Alphonse Bertillon, Francis Galton, and many more in a range of subdisciplines to emerge within the field.

New Philosophies of Science: Conjecture and "The Scientific Man"

In the late nineteenth century and early twentieth century, the philosophy of science saw new developments which were influenced by processes rooted in the conjectural or semiotic paradigm. One of the ways in which semiotics was developed into scientific theories and new approaches to knowledge were through the ideas of American Pragmatism. In addition to being an exemplification of the broader upsurge and explication of *semiotics* as a theoretical discipline, (which itself developed into a range of different philosophical theories, for example, critical theory, post-structuralism, hermeneutics, reflexivity, etc.) this particular theoretical paradigm has overt links to the developments of forensic science. The credited founder of Pragmatism Charles Sanders Peirce (1839–1914) was deeply inspired by the detective figure in his exposition of his thoughts on reasoning and the "scientific man", which again was part of a more general theory of signs and sign processes as a distinct theory of knowledge (Eco and Sebeok 1983). Besides, Peirce was and is generally acknowledged as a natural scientist, above being a philosopher, although his philosophy of science is seminal.

Based among other on the figure of the scientifically reasoning detective fashioned by Conan Doyle, Peirce developed a new concept of inference – abduction – which served to integrate the approach to science initiated with Galileo and the necessity of using incomplete knowledge and conjecture – not least in face of lack of enough data – to form hypotheses and predictive conditionals to be tested as a means of approaching what could qualify as secure beliefs. Abduction is Peirce's designation of an inference which is made from an observation to its possible explanation; that is, from evidence to a hypothesis or conclusion – where the phenomenon to be explained, through the evidence of it observed, itself constitutes a significant part of the reason to presume that the explanation is correct (Lipton 2000).

Peirce held that all inquiry, including science, is to be described as a struggle to move from irritating, inhibitory doubt born of surprise, disagreement, and the like, toward the reach of a secure belief. Belief being defined as that on which one is prepared to act, rather than as the pursuit of truth per se. Scientific inquiry, to Peirce, was thus just one part of a broader spectrum and spurred, like inquiry generally, by genuine doubt, rather than by mere verbal or hyperbolic doubt, which he held to be fruitless. Having grouped all forms of inquiry under the same pursuit, he also devised a theory of inquiry and logic, which introduced a new logic of inference to the description of science.

Peirce held that abduction, induction, and deduction made a complementary whole which together constitute scientific method. None of the modes of inference could be dispensed with – since abduction to Peirce is the only way new knowledge can arise. Logic was based on all the modes of inference, including abduction, which he perceived as a precondition for the development of science. Peirce also replaced the conception of science as a collection of certified truths or systematized knowledge with a definition of science as an inquisitorial attitude, the defining characteristics of science to be found in the approach to inquiry and the controls put on it rather than in the contents of its canons.

Being deeply inspired by Aristotle, whose concept of *apagoge* seems to entail all the things that Peirce describes as retroduction or abduction and who may be the original inspiration for his thinking (Aliseda 2006), he picks up on the same theme that Vico (as mentioned earlier) raised during the Renaissance: that all actions, whether practical decisions, medical diagnosis or scientific inquiry make use of the principle of the topics (i.e., those culturally formed beliefs, habits of experience and expectations to the order and workings of the empirical world) to form judgments and beliefs on which to act whether practically or theoretically (Miner 1998). One can legitimately speak of pragmatism as a reinterpretation of Greek philosophy about knowledge and thus a rekindling with the roots of the development of the sciences and their conjectural roots.

While Peirce is not a prominent figure in the canons of forensic science – if at all present, the practices, which inspire his philosophical positions, are indeed those of the forensic practitioner. It therefore seems in keeping with the micro-history of science and the conjectural paradigm delineated by Ginzburg, to point to the pragmatist philosophy as an example of trends in the philosophy of science, which are informed by the conjectural and semiotic paradigm running from Ancient Greek philosophy to some of the developments in thinking about science making way from the nineteenth century and onward, and which are in keeping – however subtle and mutely – with the philosophical basis of the forensic process.

Conclusion and Further Perspectives

The Forensic Process: Conjecture *and* Science or Conjectural Science?

The philosophical basis of the forensic process is not to be discerned as a linear development of thought but rather as a meshwork of ideas. However, the influence of the ideas delineated is clear considering the principles brought to bear on the forensic process by the founding fathers and key figures in forensic science, for example: (a) the principle of transfer: that two physical objects coming into contact with each other will result in minute traces, imprints or effects of this contact (also known as the principle of exchange) commonly attributed to Edmond Locard (1928, 1930), but by him attributed to the inspiration from Alexandre Lacassagne; (b) the goal of individualization based on the identification and/or classification of the physical or chemical nature of the evidence (Saferstein 1998); (c) the idea of classification/individualization and the attempt to determine the source of evidence (Kirk 1963); (d) the idea of association linking person to crime scene and leading to inferences of source and target (Osterburg 1968), and (e) the purpose of reconstruction of past events (DeForest et al. 1983) (for an overview description, see Inman and Rudin 2001). All the central ideas of forensic analysis are based on a combination of topical inspiration (i.e., attempts at answering the 6W's essential to legal deliberation) and the foundational concepts of the semiotic paradigm that infers causes from effects and which on the basis of classifications of the nature of the physical world discerns individual characteristics of the particular situation, object, or document. Characteristics which in turn serves to discern the history, meaning, and causes of the phenomenon under scrutiny, (i.e., of the physical evidence left over from an event (possibly criminal) passed).

In much the same way as the clinician of early modern medicine, the forensic practitioner may be said to embody a point of contact between two differing yet cognate paradigms – that of semiotics and that of Galilean natural science. Each has differing goals and criteria for valuing truth claims: one related to abstract disinterested truth and one related to practical effects and rationality. Like the physicians of the late nineteenth century, the forensic practitioner is a "diagnostician" and "expert connoisseur" with respect to assisting the investigative inquiry and translating the topics and questions of law into scientific questions. He is also a "naturalist" and "empiricist scientist" with respect to analyzing artifacts believed to adhere to an incident under legal scrutiny and providing conclusions regarding their identity based exclusively on knowledge obtained through observable and verifiable data. The first form of knowledge tends to be unspoken and does not easily lend itself to formalization, since nobody learns how to be a connoisseur or diagnostician simply by applying the rules.

The allusion to one of the key works in Modern scientific philosophy by Karl R. Popper: *Conjectures and refutations: The growth of scientific knowledge* (1963) in the heading above is intended as a pun on words which may serve as the key points of further perspectives, since the question is not whether conjecture is part of what informed the growth of forensics but rather: How much conjecture is allowed and under what circumstances for forensic science to qualify as *science?* The key concern of the forensic sciences and thus the key philosophical problem with which it grapples is that of *individualization* (Kirk and Thornton 1974). Forensic science – arising out of the conjectural paradigm and the semiotic model of medicine – is in many ways the science of the individual *par excellence*. Seeking from beginning to end of the forensic process and through the fields of knowledge informing the practical process of *forensic inquiry* (i.e., the forensic process) answers, which may provide individuation of both people and cases – of "who" and "what" of the Ciceroean topics. In this way, the forensic process is heir not only of the key topics identified by Cicero for the presentation of a legal case but also of the quest of the sciences build on the Galilean model and the problems facing these in their attempt to explain and describe the particular and the characteristics of the individual, and the ways that phenomena known through theoretical classifications may play out in particular situations, hereby ascertaining the span of characteristics to be associated with the phenomena in a given situation and thus ascribed to it.

This problem is becoming increasingly acute to the forensic sciences, due to emergence of debates in scientific circles of is scientific status. Yet in the philosophical history of the forensic process of inquiry, can also be identified ideas and principles stemming from a semiotic paradigm, which has grown large in the human and social sciences, and which holds promise of a new formulation of a model of science, which may provide an alternative route to the problems faced, by introducing a concept of scientific rigor which is more "elastic."

Forensic sciences is finding itself in a position where new developments are imminent, if not already in process. Developments which may prove decisive for the future role and status of forensic evidence, and which springs exactly from the problem having faced sciences developed in the image of Galilean physics, since the beginning of the Modern Era. This dilemma is coined by Ginzburg (1980) as that of choosing between: (a) achieving significant results from a scientifically weak position, or (b) to establish a strong scientific position, but get meager results.

Rethinking and revisiting the philosophical basis of the forensic process, and turning to key thinkers within this paradigm for new approaches to the dilemma may provide an alternative route bypassing the deadlock of the debates. The theories, models, and scientific content of the conjectural model that might inspire new ways of theorizing and conceptualizing the foundational and defining principles of the forensic process are still however largely unexcavated, even though it provides the philosophical basis and seed of development for a broad range of contemporary disciplines in all areas of science.

Related Entries

- ▶ Forensic Science and Criminal Inquiry
- ▶ Forensic Science Culture
- ▶ Forensic Science Effectiveness
- ▶ Forensic Science in the Nineteenth and Twentieth Centuries
- ▶ History of Criminal Investigation
- ▶ History of Forensic Science in Policing
- ▶ Philosophy of Forensic Identification
- ▶ Prosecutorial Discretion
- ▶ Scientific Basis of the Forensic Process
- ▶ Scientific Evidence in Criminal Prosecutions

Recommended Reading and References

Aliseda A (2006) Abductive reasoning: logical investigations into discovery and explanation. Springer, Dordrecht

Aristotle, Lawson-Tancred H (1991) The art of rhetoric. Penguin Books, London

Aristotle (384–322 BC) Topics: (trans: Gaza T, CR Editors)

Balkin JM (1996) A night in the topics: the reason of legal rhetoric and the rhetoric of legal reason. In: Brooks P, Gewirtz P (eds) Law, stories: narrative and rhetoric in the law. Yale University Press, New Haven

Cole SA (2001) Suspect identities: a history of fingerprinting and criminal identification. Harvard University Press, Cambridge/London

Daston L, Galison P (1992) The image of objectivity. Representations, No. 40, Special issue: seeing science. pp 81–128

Daston L, Galison P (2007) Objectivity. Zone Books, New York/Cambridge

DeForest P, Lee H, Gaensslen R (1983) Forensic science: an introduction to criminalistics. McGraw Hill, New York

Eco U, Sebeok TA (1983) The sign of three: Dupin, Holmes, Peirce. Indiana University Press, Bloomington

Foucault M (1970) The order of things: an archaeology of the human sciences. Tavistock, London

Foucault M, Gordon C (1980) Power-knowledge: selected interviews and other writings, 1972–1977. Harvester Press, Brighton

Foucault M, Sheridan A (1973) The birth of the clinic: an archaeology of medical perception. Tavistock Publications, London

Foucault M, Sheridan A (1979) Discipline and punish: the birth of the prison. Penguin, Harmondsworth

Foucault M, Senellart M, Davidson AI (2007) Security, territory, population lectures at the Collège de France, 1977–78. Palgrave Macmillan, Basingstoke/New York

Foucault M, Senellart M, Ewald F, Fontana A, Burchell G, Collège de France (2008) The birth of biopolitics: lectures at the Collège de France, 1978–79. Palgrave Macmillan, Basingstoke/New York

Giambattista V (transl. Marsh D 1999) 1744 New Science. Penguin Books, London/New York/Victoria/Ontario

Ginzburg C (1980) Morelli, Freud and Sherlock Holmes: clues and scientific method. Hist Workshop J 9(1):5–36

Hookway C (1985) Peirce. Arguments of the Philosophers series. Routledge, London/New York

Inman K, Rudin N (2001) Principle and practice of criminalistics: the profession of forensic science, Protocols in forensic science. CRC Press/Taylor & Francis Group, Boca Raton/London/New York

King LS, Meehan MC (1973) A history of the autopsy: a review. Am J Pathol 73(2):514–544

Kirk PL (1963) The ontogeny of criminalistics. J Crim Law Criminol Police Sci 54:235–238

Kirk PL, Thornton JI (1974) In: Thornton JI (ed) Crime investigation, 2nd edn. Wiley, New York/London

Leff MC (1983) The topics of argumentative invetion in Latin rhetorical theory from Cicero to Boethius. Rhetorica J Hist Rhetoric 1(1):23–44

Lipton P (1991) Inference to best explanation. Routledge, London/New York

Locard E (1928) Dust and its analysis. Police J 1:177

Locard E (1930) The analysis of dust traces Am J Police Sci 1: pp. 276–298, 401–418, 496–514

Markus S (1976) Introduction to A. Conan Doyle. In: Conan Doyle A (ed) The adventures of Sherlock Holmes: a facsimile of the stories as they were first published in the Strand magazine. Schocken Books, New York, pp x–xi

Miner RC (1998) “Verum-factum” and practical wisdom in the early writings of Giambattista Vico. J Hist Ideas 59(1):53–73

Mnookin JL et al (2011) The need for a research culture in the forensic sciences. UCLA Law Rev 58:725–779

National Academy of Sciences (2009) Strengthening forensic science in the United States: a path forward. National Academies Press, Washington, DC

Osterburg JW (1968) The crime laboratory: case studies of scientific criminal investigation. Indianna University Press, Bloomington

Popper KR (1963) Conjectures and refutations: the growth of scientific knowledge. Routledge, London

Robertson DW (1946) A note on the classical origins of "Circumstances" in the medieval confessional. Stud Philol 43(1):6–14

Saferstein R (1998) Criminalistics: an introduction to forensic science, 6th edn. Prentice Hall, Englewood Cliffs/New York

Schoeck RJ (1983) Lawyers and rhetoric in sixteenth-century England. In: Murphy J (ed) Renaissance eloquence: studies in the theory and practice of renaissance rhetoric. University of California Press, Berkeley

Sengoopta C (2004) Imprint of the Raj: how fingerprinting was born in colonal India. Pan Books, Basingstoke/Oxford

Tilstone W, Hastrup M, Hald C (eds) (2012) Fisher's techniques of crime scene investigation. First international edition. CRC Press/Taylor & Francis Group, Boca Raton

Toulmin SE (1958) The uses of argument. Cambridge University Press, Cambridge

Victorinus M (1863) Explanationum in rhetoricam M. Tullii Ciceronis libri duo. In: Halm C (ed) Rhetores Latini minores. Teubner, Leipzig

Philosophy of Forensic Identification

Ton Broeders
Faculty of Law, Institute for Criminal Law
and Criminology, Leiden University, Leiden,
The Netherlands
Maastricht University, Maastricht,
The Netherlands

Overview

In most if not all criminal investigations, the collection, examination, and interpretation of physical evidence plays a major role. Material traces, whether they are of a physical, chemical, biological, or digital nature, may serve to suggest or support plausible scenarios of what might have happened at a possible scene of crime. Ultimately, they may be instrumental in distinguishing between the rival scenarios that the judge or jury will have to consider in arriving at the final verdict: Was the crime committed by the suspect in the way described by the prosecution, was it someone else who did it, or was no crime committed in the first place?

Individualization of physical traces, like DNA, handwriting, or fingerprints, logically amounts to a process of inference of identity of source between a crime scene trace and some reference material whose origin is known. While its aim is to uniquely identify the source of the trace and many criminalists, most notably the dactyloscopists, have long proceeded as though individualization of traces is a routine affair, there is no scientific basis for this claim or for the methods by which individualization or *source attribution* is supposedly achieved. It is largely as a result of the advent of forensic DNA analysis and the conceptual framework associated with the interpretation of DNA evidence that the forensic community has come to appreciate the "individualization fallacy" (Saks and Koehler 2008) and the consequent need to adopt what has come to be known as a logical or Bayesian approach to the interpretation of technical and scientific evidence.

A comprehensive model for the assessment and interpretation of forensic evidence along Bayesian lines was developed in Britain by Ian Evett and colleagues at the Forensic Science Service (Cook et al. 1998a, b; AoFSP 2009; Evett 2011). It defines a hierarchy of propositions to be addressed in casework which is composed of three levels and extends over two domains. The first domain, that of the forensic expert, involves propositions at the level of source and at the level of activity. Propositions at the level of source address the question of the origin of the trace, while propositions at the activity level relate to the question how and when the trace arose. Propositions formulated at the third level, termed the level of offense, belong exclusively to the domain of the trier of fact and relate to the ultimate, legal issue whether an offense was committed and if so, whether it was committed by the defendant.

While the Bayesian approach provides an excellent framework for the evaluation and interpretation of expert evidence in that it helps define the relevant questions and draws a sharp line between the domain of the expert and that of the trier of fact, it is not an intuitively easy approach. A clear and transparent exposition of the method used by the expert to determine the weight of the scientific evidence is therefore of the essence. Recent research as well as court decisions suggest that if this information is lacking judges and juries will be hard put to assess the evidence at its true value.

Criminalistic Processes: Individualization Versus Identification

In order for material traces like DNA to be able to contribute to criminal investigations in the form of technical or scientific expert evidence, there are various questions that need to be addressed. One question seeks to determine the *nature* or *substance* of the trace material. Are we dealing with blood, gunshot residue, MDMA, or hair? The process by which this type of question is answered is called *identification.*

A second process, *classification*, takes the first process one step further. It is undertaken to determine whether the crime scene trace can be assigned to a particular group or class of similar objects or substances. Examples of classification processes in the forensic sciences are: the classification of blood into human versus animal blood; of hair into animal versus human hair and, beyond that, the (sub)classification into body, head, and pubic hair; the classification of fibers into man-made versus natural fibers; the classification of firearms and cartridges in terms of their make and caliber, or that of shoes according to size, make, model, sole pattern, etc.

A third process seeks to identify the *origin* or *source* of the trace material. Is it possible to relate the trace or the carrier of the trace to a particular person or object? This process, the determination of the unique source of a trace, is technically called *individualization* and is often seen as the process that lies at the heart of the criminalistic effort. Although, in common parlance, the term "identification" may also be used for the process of source determination, as in a phrase like fingerprint identification or indeed in the title of this entry, the proper technical term for this process is individualization.

A fourth question concerns the relation of the trace material with the crime by trying to establish *how* and *when* the trace was generated. This process, which focuses on the nature of the *activity* by which the trace material arose, is often referred to as *association.*

All four processes, identification, classification, individualization, and association, are essential elements of the larger process of *reconstruction* (Inman and Rudin 2000): the attempt to create coherent scenarios of the chain of events that constitutes the incident under examination on the basis of the available evidence.

Particularly in the context of DNA evidence, both "source level" questions of the type "Whose DNA is it?" and "activity level" questions of the type "How and when did the cell material get there?" are of paramount importance, as it is only on the strength of the combined answers to these questions that a link can be established between a particular person or object and a possible crime.

Conclusions in Individualization

> Criminalistics is the science of individualization (Kirk 1963, p. 235)

Individualization may be defined as the process undertaken to establish the unique source or origin of a particular material trace. Put simply, it tries to answer questions like "Whose DNA/handwriting/fingerprint, etc. is this?" According to Paul Kirk (1902–1970), one of the founding fathers of criminalistics, it is individualization that marks off criminalistics as a science from other sciences.

While the goal of criminalistics is always the same, there are considerable differences in the methodology used and the conclusions formulated by the practitioners of the various forensic disciplines to achieve this aim. By and large, three groups may be distinguished in terms of the way in which the conclusions of a comparative trace examination are expressed. First, the conclusions may directly address *the probability of the source hypothesis* given the evidence. This tends to be the case in dactyloscopy and typically also still holds for most other types of source level trace examinations. An important exception is DNA, where the conclusions expressed relate to *the probability of the evidence* under a particular hypothesis rather than to the probability of the hypothesis given the evidence. Second, unlike dactyloscopists, who will generally express their conclusions in categorical terms, forensic experts in most other fields of trace examination will tend to express their conclusions in probabilistic terms, in either a verbal format or in a quantitative format, as in DNA typing.

Schematically, where C stands for categorical, P for probabilistic, E for statement of the probability of the evidence, and H for statement of the probability of the hypothesis, the following distinctions may be noted:

Dactyloscopy (fingerprints)	C	H
Other physical traces (e.g., *handwriting, fibers, shoe prints, tool marks)*	P (and verbal)	H
DNA (biological traces)	P (and quantitative)	E

For example, the dactyloscopist will conclude that a finger mark does or does not originate from a particular (person's) finger, thereby making a categorical statement of the probability of H, the source attribution hypothesis. But the handwriting expert will – traditionally anyway – typically conclude that the questioned writing *probably/very probably/with a probability bordering on certainty* does – or does not, as the case may be – originate from the writer of the reference material, thereby making a (verbal) probabilistic rather than a categorical statement of H, the source hypothesis.

By contrast, the DNA expert will first determine whether the source material *could* originate from the person whose reference profile is being compared with that of the crime scene sample. If the profiles of the crime scene material and the person under investigation match, the expert will state that the crime scene material *may* originate from this person. To indicate the significance of this finding, the expert will add how likely the evidence, i.e., a matching profile, is if the crime scene material originates from the "matching" person under investigation as opposed to a random member of the population. Alternatively, the expert may report the estimated frequency of the profile in the relevant population. A third way to report the weight of the evidence is in terms of the so-called random match probability of the profile: The probability that a randomly chosen member of the population who is not related to the donor of the cell material or the matching suspect will match the crime scene profile.

It appears therefore that while the experts in the diverse fields are answering the same question, that of the origin of the trace, they will tend to use different conclusion formats to express their findings. The notions underlying these differences are explored below.

The Classical Approach to Individualization: The Underlying Principles

The traditional forensic approach to trace individualization is based on four – partly implicit – principles:

1. The principle of transfer of evidence
2. The principle of the divisibility of matter
3. The uniqueness assumption
4. The individualization principle

The Transfer Principle: "*Every Contact Leaves a Trace*"

The first of these, the *transfer principle*, is captured in the phrase "Every contact leaves a trace." It expresses the notion that in the commission of criminal acts, invariably some traces will be left on the scene. It provides a theoretical basis for the generation of traces and a principled argument for the examination of the crime scene for traces. This insight was formulated most clearly by the Frenchman Edmond Locard (1877–1966), who was the first director of the *Laboratoire de Police Scientifique* in Lyon, France, in 1912 and is universally acclaimed as one of the fathers of criminalistics. While Locard was primarily thinking of the transfer of microscopic traces like dust, dirt, nail debris, or fibers, as left in the commission of the more violent type of crime, Locard's exchange principle, as the principle is also commonly referred to, is also applicable to traces that arise in the context of less violent crimes as well as to latent (hidden), patent (visible to the naked eye), or plastic impression evidence, such as finger marks, tool marks, footwear marks, or striation marks on bullets and cartridge cases. In the latter case, it is not so much the physical matter that is deposited at or taken away from the crime scene that is of interest but the patterns or shapes that are transferred from donor (object) to recipient (object). In addition, the principle increasingly applies to so-called *contact traces*, as in DNA evidence. Here, terms like "trace DNA" or "touch DNA" are used to refer to cell material and debris transferred through skin contact, which may arise as a result of regular use, as on a watch, from a single, firm contact, as on a tie wrap or strangulation cord, or from a single touch, as on a glass surface (Raymond et al. 2004).

As Locard observed at the time, transfer may go either way. Traces like glass fragments, blood stains, hairs, or fibers are left at the crime scene by a *donor* and may be picked up from there by a *receptor*. If (part of) the material collected is subsequently left by the receptor and picked up by a second receptor, we speak of *secondary transfer*. For example, fibers or DNA material picked up by A from B's clothing may be transferred from B's clothing to chair C, and eventually end up on the clothing of D. In principle, forms of tertiary and quarternary transfer might also occur.

The Divisibility of Matter: *"Matter Divides Under Pressure"*

The second underlying principle of criminalistics, that of the *divisibility of matter*, was explicitly defined only relatively recently by the American DNA experts Keith Inman and Norah Rudin (2000). It explains *why* transfer can play such an important role in the generation of traces. Although the principle as such is fairly obvious, it is of considerable importance for a proper understanding of the relation between the way traces arise and their interpretation. In an article entitled "*The origin of evidence*," Inman and Rudin (2002: p. 12) describe the process of the division of matter and its results as follows:

> Matter divides into smaller component parts when sufficient force is applied. The component parts will acquire characteristics created by the process of division itself and retain physico-chemical properties of the larger piece

This mechanism has important implications for the relation between traces and their sources. They are:

> *Corollary 1* Some characteristics retained by the smaller pieces are unique to the original item or to the division process. These traits are useful for individualizing all pieces to the original item.
>
> *Corollary 2* Some characteristics retained by the smaller pieces are common to the original as well as to other items of similar manufacture. We rely on these traits to classify them.
>
> *Corollary 3* Some characteristics from the original item will be lost or changed during or after the moment of division and subsequent dispersal; this confounds the attempt to infer a common source. (Inman and Rudin 2002, p. 12)

While the two principles discussed so far relate to the creation or *generation* of traces, the following two principles are central to the *interpretation* of traces, as viewed in the traditional approach to trace individualization.

The Uniqueness Assumption: *"Nature Never Repeats Itself"*

The first of these is the uniqueness assumption. It is nicely captured in the phrase "Nature never repeats itself" and essentially simply states that no two objects are identical. Or, as Kirk and Grunbaum (1968, p. 289) put it:

> Now most students believe that all items of the universe are in some respect different from other similar items, so that ultimately it may be possible to individualize not only a person but any object of interest. This effort is the heart of criminalistics.

The uniqueness assumption was probably most vigorously championed by the fingerprint fraternity. However, before the fingerprint was discovered as a means to verify a person's identity and subsequently came to be used for forensic purposes, the same principle provided a basis for forensic anthropometry, which was developed by the Frenchman Alphonse Bertillon (1853–1914).

Anthropometry. Anthropometry was developed by Bertillon primarily to identify repeated offenders. The method is based on an assumption derived from the Belgian astronomer and statistician Adolphe Quetelet (1796–1874) that no two human bodies are equal. A founder of modern quantitative sociology, Quetelet is not only believed to be the inspiration for the frequently cited phrase that nature does not repeat itself but must also be credited with the definition of the Quetelet index, which, since 1972, has come to be more widely known as the BMI or Body Mass Index.

The anthropometric method consisted in recording the dimensions of an arrestee's body in terms of seven and later 12 measurements of a fixed set of parts of the body, including total physical height, the length and width of the head, the right ear, and the left foot, which were believed to be constant for adult members of the human race. In 1883, Bertillon succeeded in identifying a repeated offender by comparing his measurements with anthropometric data recorded earlier. Later in his life, Bertillon was the first to make a fingerprint identification in a murder case on the European Continent (Thorwald 1965, p. 83).

The assumption of uniqueness, together with the temporal stability of fingerprints, or friction ridge patterns as they are more technically called, is often adduced as a theoretical ground for the justification of the use of categorical conclusions of origin, as has typically been common if not universal practice in dactyloscopy. However, as Saks and Koehler (2005, p. 892) put it, in formulating these categorical conclusions, dactyloscopists in fact rely on a flawed notion of "discernible uniqueness." The real issue in source attribution is not whether all possible sources can be distinguished from each other in principle, which is what the uniqueness assumption – presumably correctly – implies. The crucial question is whether a trace, which, due to the factors described under the divisibility principle, will inevitably differ to some extent from its particular source, can be attributed – with certainty, or, failing that, with any reliable degree of probability – to that source.

The Individualization Principle: *"That Can't Be a Coincidence"*

The fourth principle in the traditional approach to trace individualization states that a conclusion of (probable) common origin of a trace and reference material – as in a comparative examination of a questioned handwriting sample and a reference sample from a known person – may be arrived at if there are so many similarities of such significance that their occurring together by chance may be practically excluded. As the American handwriting expert R.A. Huber (1959–1960, p. 289) put it in his definition of what he called "the principle of identification":

> When any two items have characteristics in common of such number and significance as to preclude their simultaneous occurrence by chance, and there are no inexplicable differences, then it may be concluded that they are the same, or from the same source.

The suggestion here is that a criterion, i.e., "of such number and significance," may be defined which will provide a principled and objective way to determine that the possibility (or probability) that two objects meet this criterion by chance can be excluded. Such a criterion is not only not feasible in practice, as it begs the question what number and what degree of significance is required, but it also lacks a theoretical basis in that it ignores the essentially inductive nature of the individualization process.

The reason why individualization is problematic from a theoretical point of view is that any attempt to identify the unique source of physical, biological, or pattern evidence like finger marks, footwear marks, DNA, or handwriting is typically frustrated by the induction problem. We cannot, solely on the strength of even an extreme degree of similarity between trace and reference material, conclude that a particular trace must have originated from some specific reference material to the exclusion of all other possible sources, unless we have been able to examine all these alternative sources and eliminate them categorically.

To begin with, the population from which a finger mark or a questioned handwritten text actually originates is typically indefinite in size and frequently largely unavailable for examination. But even if it were possible to examine a large number of potential writers or fingers, we could not exclude finding one or more whose reference handwriting or fingerprint would show a similar or even greater degree of similarity with the questioned handwriting or finger mark than did the reference sample of the suspect. Since this possibility cannot be excluded, the individualization problem tends to be impossible to solve and in that sense is strongly reminiscent of that of Popper's white swans: We cannot conclude that all swans are white unless we have been able to examine all swans (Popper 1959). Nor for that matter can we conclude that the trace *probably* originates from the reference material with which it shares many features. Indeed, although the observed degree of similarity between trace and possible source will tend to make the hypothesis of identity of source *more probable* in relative terms than it was before the comparative examination was carried out, similarity is neither a sufficient nor even a necessary condition for identity of source in absolute terms.

It is interesting that the individualization criterion as captured in the phrase "That cannot be a coincidence" is essentially similar to that used in traditional, standard textbook statistical significance testing. In this approach, the result of a hypothesis test is termed significant if it is unlikely to have occurred by chance. More specifically, the so-called null hypothesis that the result is due to chance may be rejected if the obtained result is less likely to occur under this hypothesis than a predetermined threshold probability of – frequently – 1 or 5 %. Like the traditional identification paradigm, this approach is also coming in for more and more criticism, partly because it also fails to take account of the probability of the result under the alternative hypothesis.

Class Characteristics Versus Individual Characteristics

In an attempt to overcome the induction problem traditional forensic identification, experts frequently rely on the distinction between class characteristics and individual characteristics. For example, all firearms of a particular make and caliber may leave the same markers on a cartridge or bullet, thereby making it possible to identify the type of weapon on the basis of the class or system characteristics that the particular type of firearm is known to leave. However, a certain configuration of striation marks left on a bullet may be distinctive for a particular weapon. As Thornton and Peterson (2002) put it:

> Class characteristics are general characteristics that separate a group of objects from a universe of diverse objects. In a comparison process, class characteristics serve the very useful purpose of screening a large number of items by eliminating from consideration those items that do not share the characteristics common to all the members of that group. Class characteristics do not, and cannot establish uniqueness.
>
> Individual characteristics, on the other hand, are those exceptional characteristics that may establish the uniqueness of an object. It should be recognized that an individual characteristic, taken in isolation, might not in itself be unique. The uniqueness of an object may be established by an ensemble of individual characteristics. A scratch on the surface of a bullet, for example, is not a unique event; it is the arrangement of the scratches on the bullet that mark it as unique.

Unfortunately, the definition of individual characteristics is circular. They are defined as characteristics that are – collectively – capable of establishing uniqueness. But whether an ensemble of individual characteristics is unique is itself an inductive question: We can never be sure that a feature or combination of features is unique, until we have observed all relevant objects, which is impossible.

What practitioners of traditional forensic identification sciences really do is perhaps best described by Stoney (1991), who used the image of the "leap of faith" as the mechanism whereby the forensic scientist actually establishes individualization, as in dactyloscopy:

> When more and more corresponding features are found between the two patterns scientist and lay person alike become subjectively certain that the patterns could not possibly be duplicated by chance. What has happened here is somewhat analogous to a leap of faith. It is a jump, an extrapolation, based on the observation of highly variable traits among a few characteristics, and then considering the case of many characteristics. [] In fingerprint work, we become subjectively convinced of identity; we do not prove it.

Ultimately then, in traditional identification disciplines, in reaching a conclusion about the probable or categorical origin of the trace material, the expert delivers an essentially subjective opinion. When informed by adequate levels of training, experience, and expertise, this conclusion will frequently be correct, but it must be clear that the expert becomes convinced of the (probable) origin of the trace. He does not "prove" it: There is no logical basis for the conclusion.

The Logical Approach to Individualization: The Concept of the Likelihood Ratio

By contrast, in the logical approach, the expert does not primarily seek to determine the probability of the source or activity hypothesis. Instead, the purpose of the comparative examination is to determine the likelihood of the evidence under two mutually exclusive propositions or hypotheses. Suppose we find a size 14 shoeprint at a crime scene and a suspect emerges who takes size 14. It will be clear that the shoe size information by itself gives us insufficient basis to say that it was the suspect who left the print rather than one of the other shoe size 14 wearers in the area (or beyond). The mere fact that the suspect wears size 14 shoes does not make him more suspect than anybody else with this size shoes. At the same time, it is clear that the finding looks incriminating. Or is it a mere coincidence?

The likelihood ratio (LR) provides a principled way to address this question. To calculate it, we need to determine the ratio of the likelihood of a size 14 turning up at the crime scene under two rival hypotheses: (1) The shoe mark was left by the suspect versus (2) the mark was left by a random member of the relevant potential donor population. We know that the likelihood of finding a size 14 mark if the suspect left it (and assuming he does not occasionally or otherwise wear a different size shoe) is 1, or 100 %. To assess the likelihood of finding a size 14 shoe mark under the alternative hypothesis that a random member of the relevant population left the print, we need to know what percentage of that population takes a size 14 shoe. Suppose we know this figure to be 4 %. We can now calculate the likelihood ratio of the evidence under these two competing hypotheses, which in this case would amount to 100/4 = 25.

We can paraphrase this result by saying that the footwear evidence is 25 times more likely if the suspect left the mark than if a random member of the population left it. We can also say that the evidence makes it (25 times) more likely that the suspect left the mark than we believed was the case before we obtained the evidence. But we cannot on the basis of the shoe mark evidence alone pronounce upon the probability of the hypothesis that the suspect left the mark in absolute terms. Alternatively, what we can say in a case like this is that an LR of 25 reduces the group of potential suspects by a factor 25, leaving just one potential suspect on average for every 25 potential suspects considered before the shoe print evidence became available.

Similarly, when applied to DNA evidence, the likelihood ratio is a measure of the weight of the evidence and may be seen as an indication of the extent to which the uncertainty about the source hypothesis is reduced by the evidence. If, for example, the matching profile is known to have a frequency of 1 in a million in the relevant population, the LR of the matching DNA evidence may be reported as one million: It is the ratio of the likelihood of the evidence under the hypothesis that the person under investigation is the donor of the DNA material, i.e., 1 (or 100 %) divided by the likelihood of the evidence under the alternative hypothesis that the DNA originates from a random member of the population, i.e., 1/1,000,000. The conclusion implies that the evidence is a million times more likely if the crime scene sample originates from the donor of the reference material than if a random member of the population were the donor of the crime scene sample. It is only if there is no match that the DNA expert may make a – negative – categorical statement of the probability of the hypothesis.

Discrete Versus Continuous Variables

In the examples involving DNA and shoe sizes discussed above, the variables of interest are discrete or categorical in nature and the similarity between trace and reference material could be said to be either complete or to be entirely absent: The DNA profiles and the shoe sizes are the same or they are different. If they are different, the suspect's shoe or body can be eliminated as the source of the trace. However, in many cases, the correspondence between trace and source will not be perfect. Many variables are not discrete (like DNA markers or shoe size) but continuous. Examples are quantitative variables such as length, weight, or shape. For these variables, there will always be a difference between trace and reference materials because in both cases, we are dealing with samples that will at best only approximate the "true," i.e., average, population value for the feature of interest.

Again, handwriting analysis may serve as a case in point, but the same holds for any other type of trace with continuous properties like tool marks, fibers, speech, glass, or paint. Both the questioned writing and the reference writing must be seen as samples taken from the indefinitely large population of handwriting productions that the writers of these samples are capable of producing. This means that any marker that is examined, like the shape and execution of the letter *t* or the figure *8*, will exhibit a certain degree of variability even within samples originating from the same source. However, traces originating from a single source will typically show a relatively small degree of *within-source* variation, while samples of material originating from different sources will typically show a relatively larger degree of *between-source* variation. As a result, in these cases, the numerator of the likelihood ratio will not be 100 % or 1, but say 80: A degree of similarity as great as that found between trace material and reference material will then be found in 80 % of cases if both originate from the same source and in say only 10 % of cases if trace material and reference material originate from different sources. The likelihood ratio would then be 80/10 = 8.

If, however, the degree of similarity found would be expected in only 20 % of cases under the hypothesis of common origin and in 40 % of cases under the hypothesis of different origin, the likelihood ratio would be 20/40 = ½, and the evidence would actually weaken the common source hypothesis rather than strengthen it. In the first instance, the likelihood ratio would exceed 1 and the evidence would (weakly) support the hypothesis of common origin; in the second instance, the likelihood ratio would be less than 1 and the evidence would – again weakly – support the alternative hypothesis, that the trace did not originate from the same source as the reference material.

The Likelihood Ratio and the Diagnostic Value

The concept of the likelihood ratio is similar to that of the *diagnostic value*, a measure which has found wide acceptance in fields like medicine and psychology as a way to express the value of a diagnostic test result. It is arrived at by dividing the relative number of correct (positive or negative) results of the test in question by the relative number of false (positive or negative) results of the test. In more technical language, by dividing the sensitivity of the test by (1 – the specificity). The sensitivity of the test is the percentage of correct positives it produces, the specificity the percentage of correct negatives. Applied to an HIV test, if the sensitivity of the test is 98 % and the specificity is 93 %, this would mean that 98 % of those infected would (correctly) test positive and 7 % of those not-infected would (incorrectly) also test positive. The diagnostic value of a positive result for the test would then be 98/(100 − 93) = 14 and the probability of the patient being infected would be 14 times greater now that the test result is known than whatever it was estimated to be before the test result was known. The diagnostic value of a negative test result would in this case be 93 (the relative number of correct negatives) divided by (100 − 98) = 2, the relative number of false negatives, i.e., 93 divided by 2 or 46.5.

Scientific evidence may be said to have *diagnostic value* in much the same way as a medical test such as an HIV test, which will not provide absolute proof of infection or otherwise but, depending on the result and its diagnostic value, will make infection more or less probable. Similarly, evidence may be more or less likely under one of two rival hypotheses. To the extent that the evidence favors, or better fits one hypothesis rather than another, it may be said to lend more support to that hypothesis.

Bayes and the Prior Probability

The likelihood ratio may be seen as a measure of the extent to which the hypothesis of interest is more probable or less probable after the scientific evidence is known than it was before the evidence was known. Another way of putting it is to say that we can update what we saw as the prior probability of the hypothesis with the evidence that has become available to arrive at the posterior probability of the hypothesis, in which the weight of the new evidence has been taken into account. By means of the Bayes' Rule, the odds form of Bayes' theorem, so called after the Rev. Thomas Bayes (1702–1761), we can calculate the posterior probability of a hypothesis by multiplying its prior probability with the likelihood ratio:

$$\begin{aligned}&\text{Prior probability} \times \text{Likelihood ratio}\\&= \text{Posterior probability}\end{aligned}$$

A simple example may illustrate the application of the rule. Suppose a prisoner is found dead and we may safely assume that one of the 99 remaining fellow prisoners in the ward is the perpetrator. As it happens, one of these 99 prisoners, S, admits to being the killer. His DNA profile is obtained and found to match with a partial profile obtained from the nail debris secured from the victim. The estimated frequency of the partial profile in the relevant population is 1 in 1,000.

If the cell material in the nail debris originates from S, the likelihood of the evidence, i.e., the matching profile, would be 1, or 100 %. If, on the other hand, the cell material belonged to a random member of the population, the likelihood of a match would be 1 in 1,000 or 0.001. The likelihood ratio of the DNA evidence would therefore be 1 divided by 0.001 = 1,000.

To determine the prior probability, we assume that all 99 remaining prisoners are equally likely to be the perpetrator. In that case, the prior probability of any one of them being the donor is 1/99, or, expressed in odds, 1 to 98. We can now calculate the prior probability of S being the donor of the cell material in the nail debris by applying Bayes' rule:

$$\begin{aligned}&\text{Prior probability} \times \text{LR} = \text{Posterior probability}\\&(1 \text{ to } 98) \times 1,000 = 1,000 \text{ to } 98.\end{aligned}$$

The odds form of the posterior probability may be converted to a fraction, i.e., 1,000/1,098, and from that into a percentage, i.e., 91.1 %. This means that on the basis of the DNA evidence alone, the probability of the suspect being the donor of the cell material is 91.1 %, assuming that the donor is one of the remaining 99 prisoners.

Of course, apart from the DNA evidence, the fact that S admitted to killing his fellow inmate is also relevant to the determination of the ultimate issue whether S is the perpetrator. However, this type of evidence as well as possible eyewitness accounts from fellow prisoners or prison staff, and any other non-DNA evidence, clearly extends beyond the domain of the DNA expert. Other experts might be able to assign a particular weight to a spontaneous confession or an eyewitness identification, in the form of a likelihood ratio. For example, empirical research may be available on which an estimate may be based of the diagnostic value (= likelihood ratio) of a confession. This could take the form of a statement that a confession is on average seven times more likely if the suspect is the perpetrator than if he is not. This evidence in turn could then be used to further update the hypothesis of guilt.

When Numbers Are Lacking

It is worth noting that the concept of the likelihood ratio may also be applied to evidence types where the frequency of relevant markers cannot be estimated in numerical terms, for example, because there are no suitable reference databases.

This would in fact hold for many types of trace evidence, including handwriting, glass, paint, fibers, toolmarks, and firearms. The logical framework may then be applied to the verbal statements that are used in these fields. In response to the logical objections raised to the use of traditional probability scales in forensic identification, various proposals have been made in recent years for the introduction of logically correct verbal probability scales. One such format, developed at the Netherlands Forensic Institute (NFI) and used primarily in the various forensic identification disciplines, looks as follows:

> The findings of the comparative examination are
> equally likely/
> more likely/
> much more likely/
> very much more likely
> under the prosecution hypothesis that the suspect is the source of trace material as/than under the defence hypothesis that a random member of the population is the source of the trace material

Note that, true to the logical format, it is the likelihood of the evidence that is addressed. While the verbal phrases used clearly have a probabilistic basis, the probabilities are not based on quantitative empirical data but informed by the analyst's experience and may be seen as internalized frequency estimates.

In those exceptional cases where the forensic scientist arrives at a subjective conviction that the trace material originates from a particular source (as in physical fits of torn paper, or qualitatively superior shoe prints or tool marks), he or she might express his or her subjective conviction, emphasizing that this is precisely that – a subjective conviction, not a scientific fact.

Words Versus Numbers

In the United Kingdom as well as in some Continental European countries, conclusions in forensic identification are increasingly expressed in a slightly different logical format. According to the "standards for the formulation of evaluative forensic science expert opinion," compiled by the Association of Forensic Science Providers (AoFSP 2009), "the evidential weight (...) is the expression of the extent to which the observations support one of the two competing propositions. The extent of the support is expressed to the client in terms of a numerical value of the likelihood ratio (where sufficiently robust data is available) or a verbal scale related to the magnitude of the likelihood ratio when it is not." (AoFSP 2009: p. 63) For this purpose, the following scale is used:

Value of likelihood ratio	Verbal equivalent
>1–10	Weak support for proposition
10–100	Moderate support
100–1,000	Moderately strong support
1000–10,000	Strong support
10,000–1,000,000	Very strong
>1,000,000	Extremely strong

In the case of the DNA evidence with a likelihood ratio of 1,000, the expert would either report the numerical value of the likelihood ratio as such or would report that the DNA evidence provides "moderately strong" to "strong" support to hypothesis 1 (the suspect is the donor). The decision to use the numerical or the verbal form would depend on the extent to which the expert considers the data underlying the calculation of the likelihood ratio to be robust. This sounds like a sensible criterion. However, there are those who advocate the blanket use of verbal terms, i.e., even if perfectly valid qualitative empirical data are available, arguing that the use of a uniform reporting format is vastly preferable and adding that quantitative data are generally too complex for nonscientists to grasp.

The Prosecutor's Fallacy

Regardless of the use of words or numbers, statements of the probability of the evidence under a particular hypothesis, as made in the Bayesian approach, are often prone to misunderstanding and may strike the recipient as counterintuitive. The finding that the similarities observed between the handwriting of the writer of an anonymous letter and the reference material produced by the suspect are *much more likely if* the suspect wrote the questioned sample than if it was written by a random member of the population is often taken to mean the converse: that it is *much more* *likely that* the suspect wrote the questioned handwriting sample than that it was written by a random member of the population.

However, if we do this, we are guilty of making a fundamental logical error which, in the judicial context, has come to be referred to as the prosecutor's fallacy (Thompson and Schumann 1987). Although the term would seem to suggest that prosecutors are particularly prone to this fallacy, it is in fact an example of a more general type of error that is often made in the context of probability statements or inverse reasoning, where it is known as the "fallacy of the transposed conditional." In its simplest form, it is easily spotted: if an animal is a cow, it is very likely to have four legs. However, the converse clearly does not hold: if an animal has four legs, it is not very likely to be a cow.

Transposed conditionals or prosecutor's fallacies are very frequently encountered in the context of DNA evidence. Suppose a partial profile is obtained from a crime scene sample whose frequency in the relevant population is estimated to be smaller than 1 in 100,000. If the expert subsequently reports the probability that a random member of the population has the same profile as the crime scene sample as being smaller than, say, 1 in a 100,000, this statement is frequently understood to mean that *the probability that* the DNA does not originate from the suspect is smaller than 1 in a 100,000. However, the former is clearly a statement of the probability of the evidence (i.e., a match with the partial profile), while the latter is a statement of the posterior probability of the source hypothesis. Mathematically, the latter statement would be correct in a situation where the prior probability was set at 50 %, or 1 to 1. This would be the case if, in addition to the matching suspect, only one person would equally qualify as a possible donor, which will not frequently be a reasonable assumption to make. If, for the sake of the argument, the prior probability were set at 200,000 (e.g., the size of the adult male population of a large town), application of the odds form of the Bayes' rule would yield a posterior probability of 33.3 %: (1 to 200,000) × 100,000 = 1 to 2, or 33.3 %.

Transposed conditionals may also occur in the context of cause and effect arguments. While it is correct to say that the street will be wet if it has been raining, the converse is clearly not necessarily true. The single observation that the street is wet does not allow us to infer that it must have been raining, or even that it has probably been raining. Alternative explanations are possible: The street may have got wet when the police used water cannon to break up a demonstration, or it is wet because somebody has just been washing his car. So we can make a statement about the likelihood of a particular finding (e.g., a wet street) under a particular hypothesis ("it has been raining") but not about the probability of this same hypothesis merely on the basis of the finding that the street is wet. To determine the posterior probability that it has been raining, we need to combine the evidence of the wet street with the prior probability of rain. In England, the prior would be high, but in a country like Dubai, it would presumably be very low. The same evidence combined with vastly different priors may lead to very different posterior probabilities.

Controversy

The Bayesian approach is not uncontroversial. Its opponents frequently view its advocates as "believers" (Risinger 2012). More specifically, critics of the Bayesian approach object to the subjective nature of the prior probability, as well as to the use of likelihood ratios which lack an empirical basis. While the former criticism is clearly valid, it might be argued that the explicit consideration of the prior probability and the formulation of alternative hypotheses is a virtue in that it helps identify the relevant questions and prevents tunnel vision. The conclusion format propagated by the predominantly British Association of Forensic Science Providers is not unproblematic either, in that a phrase, like "the examination provides strong support for the proposition that X originates from Y," will almost invariably be interpreted to mean that it is very likely that X originates from Y. Without due warning, logically correct conclusions of this type will almost inevitably tend to be mistaken for the logically flawed ones they are meant to replace.

A further problem is highlighted by a decision of the English Court of Appeal in R v T (2010). In it, the judges express sharp criticism of the report of a footwear expert, who, quite in line with the policy of the (now defunct) UK Forensic Science Service, had phrased his conclusion in verbal terms without making it clear to the judge or the jury that the verbal conclusion was based on a quantitative estimate of the frequency of the characteristics size, pattern, and wear as exhibited by the shoe mark found at the crime scene. The expert's conclusion was formulated as: "(. . .) there is at this stage a moderate degree of scientific evidence to support the view that [the Nike trainers recovered from the appellant] had made the footwear marks." However, as later appeared from case notes that were not added to the original report, the expert had calculated a likelihood ratio of 100, which number he had subsequently converted into the verbal phrase "moderate support," in accordance with the scale presented above. In addition, counsel for the appellant pointed out to the Court of Appeal that in his testimony in court in response to questions by the defense, the expert had mentioned estimates of the relevant characteristics of the shoe mark which, when combined, would lead to a likelihood ratio of no less than 26,400, in which case the footwear evidence would have been much more incriminating (in verbal terms expressed as "very strong support"). The Court of Appeal's judgment is perfectly clear:

> The process by which the evidence was adduced lacked transparency. (. . .) it is simply wrong in principle for an expert to fail to set out the way in which he has reached his conclusion in his report.

The Court of Appeal continues:

> (. . .) the practice of using a Bayesian approach and likelihood ratios to formulate opinions placed before a jury without that process being disclosed and debated in court is contrary to principles of open justice.

The ruling in R v T has led to various reactions, many (Berger et al. 2011; Evett 2011; Redmayne et al. 2011) but not all (Risinger 2012) in defense of the Bayesian approach. While one may disagree with some of the views expressed in the Court of Appeal judgment, recent research conducted in the Netherlands suggests that there is at least one other problem with the use of these "logically correct" conclusion formats. It appears that evaluative opinions expressed in the Bayesian format are likely to be misunderstood not only by defense lawyers and judges but also by forensic experts themselves. Participants in the study were asked to indicate for a variety of statements whether they were correct paraphrases of the Bayesian style conclusions that were used in a fictitious report. The study shows that a proper understanding of statements involving likelihood ratios by jurists is alarmingly poor. In order to be able to compare actual versus supposed understanding, participants were also asked to indicate how well they understood the Bayesian style conclusions of the reports on a scale from 1 (*"I do not understand it at all."*) to 7 (*"I understand it perfectly."*). The most worrying finding to emerge from the study is no doubt that not only did judges, defense lawyers, and forensic experts alike tend to interpret the conclusions of the submitted reports incorrectly but they combined their lack of understanding with a high degree of overestimation: They believed they understood the conclusions much better than in fact they did. This suggests that the continued use of Bayesian style conclusion formats or likelihood ratios requires a major educational effort if structural miscommunication between experts and triers of fact is to be avoided. Aitken et al. (2010) and Puch-Solis (2012) may be seen as attempts to meet this demand.

Conclusion

The findings of a comparative examination undertaken with a view to establishing the source of a particular trace or set of traces do not strictly allow the type of probabilistic source attributions, be they of a quantitative or verbal nature, that until recently were used by the vast majority of forensic practitioners. With the advent of forensic DNA analysis over the last decades and the widespread use of the conceptual framework associated with the interpretation of DNA evidence, awareness among forensic practitioners of other identification disciplines of the inadequacies of the traditional evidence evaluation paradigm has grown rapidly. Increasingly, this is leading to attempts to apply a logically correct way to express the value of the findings of a source attribution examination of trace material other than DNA by expressing the weight of the trace evidence in a way similar to that used in forensic DNA analysis. This takes the form of a so-called *likelihood ratio*. The concept is similar to that of the *diagnostic value*, a measure which has found wide acceptance in fields like medicine and psychology as a way to express the value of any diagnostic test result.

The concept of the likelihood ratio requires the consideration of the probability of the evidence under two competing hypotheses, one based on a proposition formulated by the police or the prosecution, the other based on an alternative proposition which may be based on a scenario put forward by the defense. As such, the Bayesian approach may be seen as a remedy against suspect-driven investigations, in which the police tend to focus on collecting evidence that will confirm the suspect's involvement in the crime and ignores alternative explanations. By contrast, in a crime-driven investigation, the investigators base the direction of the investigation on the clues provided by the crime rather than by the person of the suspect and develop one or more scenarios based on the evidence rather than make the evidence fit a particular scenario.

Technical or scientific evidence derived from material traces such as DNA, finger marks, handwriting, fibers, footwear marks, or digital data derived from a harddisk or a mobile telephone may be incompatible with a particular hypothesis that is central to a larger scenario and then effectively eliminate that scenario. More frequently, scientific evidence may be more or less likely to be found in one scenario than another and in this way may help discriminate between various scenarios. It is the consideration of the totality of evidence, both direct, witness and scientific evidence if available, which forms the basis for the ultimate decision made by the trier of fact.

Related Entries

- ▶ Automated and Manual Forensic Examinations
- ▶ Bayesian Updating and Crime
- ▶ Cognitive Forensics: Human Cognition, Contextual Information, and Bias
- ▶ Crime Science
- ▶ Criminal Investigative Analysis
- ▶ DNA Profiling
- ▶ DNA Technology and Police Investigations
- ▶ Fingerprint Identification
- ▶ Forensic Science in the Nineteenth and Twentieth Centuries
- ▶ Identification and the Development of Forensic Science
- ▶ Identification Technologies in Policing and Proof
- ▶ Scientific Basis of the Forensic Process
- ▶ Scientific Evidence in Criminal Prosecutions

Recommended Reading and References

Aitken C, Roberts P, Jackson G (2010) Communicating and interpreting statistical evidence in the administration of criminal justice: fundamentals of probability and statistical evidence. In criminal proceedings – Guidance for judges, lawyers, forensic scientists and expert witnesses, Royal Statistical Society, London

Association of Forensic Science Providers (2009) Standards for the formulation of evaluative forensic science expert opinion. Sci Justice 49:161–164

Berger CEH, Buckleton J, Champod C, Evett IW, Jackson G (2011) Evidence evaluation: a response to the court of appeal judgment in R v T. Sci Justice 51:43–49

Broeders APA (2006) Of earprints, fingerprints, scent dogs, cot deaths and cognitive contamination: a brief look at the present state of play in the forensic arena. Forensic Sci Int 159(2–3):148–157

Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA (1998a) A model of case assessment and interpretation. Sci Justice 38(3):151–156

Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA (1998b) A hierarchy of propositions: deciding which level to address in casework. Sci Justice 38(4):231–239

de Keijser J, Elffers H (2010) Understanding of forensic expert reports by judges, defense lawyers and forensic professionals. Psychol Crime Law 18(2):191–207

Evett IW (2011) Expressing evaluative opinions: a position statement. Sci Justice 51:1–2

Inman K, Rudin R (2000) Principles and practice of criminalistics: the profession of forensic science. CRC, Boca Raton

Inman K, Rudin R (2002) The origin of evidence. Forensic Sci Int 126:11–16

Kirk PL (1963) The ontogeny of criminalistics. J Crim Law Criminol Police Sci 54:235–238

Kirk PL, Grunbaum BW (1968) Individuality of blood and its forensic significance. Leg Med Annu 289–325

Popper KR (1959) The logic of scientific discovery. Hutchinson, London

Puch-Solis R, Roberts P, Pope S, Aitken C (2012) Communicating and interpreting statistical evidence in the administration of criminal justice: 2. Assessing the probative value of DNA evidence. Royal Statistical Society, London

Quetelet A (1870) Anthropométrie ou Mesure des Différentes Facultés de l'Homme. Muquardt, Bruxelles

R v T (2010) EWCA Crim 2439

Raymond JJ, Walsh SJ, Van Oorschot RA, Roux C (2004) Trace DNA: an under-utilized resource or Pandora's box? – A review of the use of trace DNA analysis in the investigation of volume crime. J Forensic Identif 54(6):668–686

Redmayne M, Roberts P, Aitken CGG, Jackson G (2011) Forensic science evidence in question. Crim Law Rev 5:347–356

Risinger DM (2012) Reservations about likelihood ratios (and some other aspects of forensic 'Bayesianism'). Law Probab Risk 0:1–11

Saks MJ, Koehler JJ (2005) The coming paradigm shift in forensic identification science. Science 309:892–895

Saks MJ, Koehler JJ (2008) The individualization fallacy in forensic science evidence. Vanderbilt Law Rev 61:199–219

Stoney DA (1991) What made us ever think we could individualize using statistics? J Forensic Sci Soc 31(2):197–199

Thompson WC, Schumann EL (1987) Interpretation of statistical evidence in criminal trials: the prosecutor's fallacy and the defence attorney's fallacy. Law Hum Behav 11:167–187

Thornton JI, Peterson JL (2002) 'The general assumptions and rationale of forensic identification' § 24–2.0–2.1. In: Faigman DL, Kaye DH, Saks MJ, Sanders J (eds) Modern scientific evidence: the law and science of expert testimony. West Publishing Co, St. Paul

Thorwald JT (1965) The century of the detective. Harcourt, Brace & World, New York, first published in German as *Das Jahrhundert der Detektive*, Droemersche Verlagsanstalt: Zürich 1964

Physical Punishment

► History of Corporal Punishment

Place-Based Randomized Trials

Robert Boruch[1], David Weisburd[2,3], Richard A. Berk[4] and Breanne Cave[5]
[1]Graduate School of Education, University of Pennsylvania, Philadelphia, PA, USA
[2]Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA
[3]Faculty of Law, The Hebrew University, Mt. Scopus, Jerusalem, Israel
[4]Department of Criminology, University of Pennsylvania, Philadelphia, PA, USA
[5]George Mason University, Fairfax, VA, USA

Overview

Place-randomized trials are an important vehicle for generating evidence about "what works" in criminology. Place-based randomized trials are a form of cluster randomization that involves identifying a sample of places (for instance, crime hot spots) and randomly allocating these locations to different police or community interventions. Random allocation assures a fair comparison among the interventions, and when the analysis is correct, a legitimate statistical statement of confidence in the resulting estimates of their effectiveness can be made.

This entry provides basic definitions and practical counsel about the use of such trials in generating evidence in crime prevention. It also identifies issues, ideas, and challenges that might be addressed by future research. Finally, it discusses the special analytic difficulties that may occur in the development of place-randomized trials as opposed to more traditional trials where individuals are the units of allocation and analysis.

Places that Are Randomized: Theory and Units of Randomization

Place-randomized trials depend on a clear understanding of the role of place. A place can be an entity in itself; for instance, a business establishment that has a legal status separates from that of owners or employees. Or a place can be an organizational convenience for smaller units within it. One may, for instance, become interested in the effect of neighborhood context on the likelihood of individual victimization. Inasmuch as individual habits can also contribute to risk of victimization, attention may also be directed toward both the attributes of neighborhoods as well as to the individuals who live within the neighborhood.

Whether the place is an entity per se or a receptacle for lower-level observational units can dramatically affect how the randomized trial is designed and analyzed. The key definitional point in a "place-randomized trial," however, is that the random assignment occurs at the place level. The implication is that the methodological benefits of random assignment are realized across places, not the units within them. Statistical analysis at the place level therefore conforms to well-understood and accepted statistical practices. Analysis attempted at the level of the nonrandomized units which are nested within places can be complex and controversial.

The "units of randomization" in a place-randomized trial may vary considerably. Weisburd and Green's (1995) study, for instance, operationalized "drug hot spots" as street segments rather than institutions such as housing developments, schools, or business units. The broad theory underlying the trial posited that focusing police and other resources on these hot spots would reduce crime, rather than leading to no effect or a migration of criminal activity into other areas. An earlier crime hot spot trial by Sherman and Weisburd (1995) in Minneapolis similarly defined the unit of analysis as a single street segment from intersection to intersection.

Other places have been targeted for different types of interventions, such as saloons in the context of preventing violence and glassware-related injuries (Warburton and Sheppard 2000). Private properties, including apartment houses and businesses, have been targeted in a study of the effects of civil remedies and drug control (Mazerolle et al. 2000). Convenience stores, crack houses, and other entities have also been randomly allocated to different interventions.

These examples invite a number of basic questions. How can such trials be deployed well? Can better theory be developed about what should happen as a consequence of an intervention at high levels of units: province or county, city or village, institution or housing development, or crime hot spot? Can theories be developed to guide thinking about change or rate of change at the primary aggregate level – the places – and below it? What new statistical problems emerge from randomization of places? These examples also invite questions about how to learn about other trials of this sort, involving yet other units of allocation and analysis.

Relationships and Agreements

People get place-randomized trials off the ground through agreements between the trialist's team and prospective partners in the place-based trial. Here, "partners" mean individuals or groups whose cooperation and experience are essential in deploying both the intervention and the trial. Weisburd (2005) emphasizes the need to develop personal relationships that lead to trust and willingness to experiment on innovations that might work better than conventional practice. In his Jersey City experiment, for instance, the strong involvement of the senior police commander as principal investigator in the study played a critical role in preventing a breakdown of the experiment after 9 months.

In Jersey City, the Deputy Chief who administered the interventions was strongly convinced of the failures of traditional approaches and the need to test new ones. The commander took personal authority over the narcotic unit and used his command powers to carefully monitor the daily activities of detectives in the trial. This style of work suggests the importance of integrating "clinical" and research work in criminal justice, much as they are integrated in medical experiments. It also reinforces the importance of practitioner "belief" in the necessity of implementing a randomized study. The Kingswood experiment described by Clarke and Cornish (1972) illustrates how doubts regarding the application of experimental treatment led practitioners to undermine the implementation of the study.

Place-randomized trials do not only involve the practitioners and researchers who are directly involved with implementation and analysis. For instance, in policing, federal funding agencies, such as the National Institute of Justice and the Bureau of Justice Assistance; accreditation agencies; and professional organizations such as the International Association of Chiefs of Police, the Police Executive Research Forum, or the Major Cities Chiefs Police Association can all influence the willingness of individual police departments to adopt place-based practices and implement place-based trials. The capacity of researchers and practitioners to carry out place-randomized trials is therefore often influenced by the preferences and priorities of these institutional stakeholders.

Developing relationships in place-randomized trials as in many other kinds of field research depends on reputation and trust, of course. The topic invites attention to questions for the future. How can better contracts and agreements in networks of organizations, public and private, ones that permit us to generate better evidence about the effects of an innovation, be developed? How can "model" contracts and memorandums of understanding be developed and made available to other trialists and their potential collaborators?

Justifications for a Place-Randomized Trial

For many social scientists, an important condition for mounting a randomized trial on any intervention that is purported to work is that (a) the effectiveness of a conventional practice, policy, or program is debatable and (b) the debates can be informed by better scientific evidence. In the crime sector, police of course are local theorists, and they often disagree about what could work better. Crime experts have also disagreed about what approaches might be effective in high-crime areas. More generally, of course, people disagree with one another about what might work in the policy sector, and there is, at times, some agreement that better evidence would be helpful.

For instance, Weisburd (2003) points out that one of the major justifications for random trials is disagreement among experts about the effectiveness of an intervention. This is an important factor in justifying a randomized trial using individuals or places as the units of random allocation. For instance, the Cambridge-Somerville Youth Study, one of the most famous experiments in youth crime prevention, illustrated that even well-planned and implemented intervention programs may have no effect on offending behavior or even have a "backfire" effect (McCord 2003). Experimental studies of the effectiveness of random beat patrol and foot patrol likewise provided surprising results regarding the effectiveness of widely accepted standard practices in criminal justice in reducing crime (Kelling et al. 1974). In these cases, experimental research concerning the effectiveness of commonly accepted practices spurred criminologists and criminal justice agencies to revise and improve intervention strategies.

The scientific justification for place-randomized trials is the assurance that if the trial is carried out properly, there are no systematic differences between groups of places randomized, which in turn carries a guarantee of statistically unbiased estimates the intervention's effect. It also assures that chance and chance imbalances can be taken into account and that a legitimate statistical statement of one's confidence in the results can be made. Further, Weisburd (2005) and others have pointed out that the simplicity and transparency of the idea of fair comparison through a randomized trial has a strong appeal for policy makers and practitioners who cannot understand and do not trust complex model-based analyses of data from nonrandomized studies. For the abiding statistician, the crucial aspect of simplicity is that the statistical inferences as to the effect's size relative to chance need not depend on econometric, statistical, or mathematical models. The randomization feature permits and invites less dependence on such speculation. And modern methods permit the use of randomization tests.

The empirical evidence on the vulnerability of evidence from nonrandomized trials in comparison to the evidence from randomized trials has been building since at least the late 1940s. Assuring that one does not depend on weak and easily assailable evidence when stronger evidence can be produced is an incentive at times in parts of the policy community. Assuring that one does not needlessly depend on heroic assumptions to produce good estimates of effect, assumptions often required in the nonrandomized trials, is an incentive for the scientific and statistical community. Boruch (2007) summarizes these concepts in the field of health care, employment and training, and economics.

Shadish and colleagues (2008) provided a persuasive illustration that is especially compelling because comparisons between a randomized experiment and an observational study were anticipated as part of their research design and before the data were collected. Their results suggest comparability, as opposed to major differences, if the quasi-experiment is designed well in a particular domain. Theirs is an interesting and potentially important specific case. More generally, the biases in estimating an intervention's effect based on the quasi-experiments can be very large, small, or nonexistent (Weisburd et al. 2001). The variance in estimates of effects appears to be typically larger in the quasi-experiments than in the randomized tests. So far, and with some narrow exceptions, there is no way to predict the directionality or magnitude of such biases or in the variances of the estimates of the intervention's effect based on a nonrandomized trial.

It remains to be seen whether similar methodological studies on aggregate level analyses using randomized versus nonrandomized places, clusters of individuals, or groups yield similar results, that is, uncover serious biases in estimating effects or the variance of the estimates or both. But it is reasonable to expect biases here also. Bertrand, Duflo, and Mullainathan (2002), for instance, focused on biases in estimates of the standard error of effects assuming no effect at all using conventional difference in differences methods and found type I error rates that were nine times the error rate presumed (0.05) in using conventional statistical tests. This was partly on account of serial correlation. More methodological research, however, needs to be done on the quasi-experimental approaches to aggregate level units so as to understand when the biases in estimates of effect appear, when the biases in estimates of their standard errors appear, and how large the mean square error is, relative to place-randomized trials.

It is important to identify dependable scenarios in which bias and variance of estimates generated in nonrandomized trials are tolerable. Doing so can reduce the need for randomized trials (Boruch 2007). It is also not easy, as yet, to identify particular scenarios in which bias in estimates of effect or variance will be small, as was the case in some work by Shadish et al. (2008). Berk's (2005) more general handling of the strengths and weaknesses of randomized controlled trials at the individual level is pertinent to the analysis of data generated in place-randomized trials also.

The scientific justifications that are identified here are important in the near term. In the long term, it would be good to understand what other incentives are and to make these explicit at different levels, for example, policy, institution (agencies), and individual service-provider levels. Incentives for better evidence may differ depending on whether the stakeholders are members of the police force at different levels, the mayor's office, and the community organizations that have a voice, and so on. Many police executives, for instance, want to improve policing and produce evidence on whether things do improve and also usually want to keep their jobs. The two incentives may not always be compatible if a city council or mayoral preferences are antagonistic toward defensible evidence generated in well-run field tests. Sturdy indifference to dependable evidence of any kind is, of course, a problem in some policy sectors.

Deploying the Intervention: Implementation, Dimensionalization, and Measurement

Justifications and incentives are essential for assuring that places, and influential people within them, are willing to participate in a randomized trial. Understanding how to deploy a new program requires expertise at a ground level, as well as substantial cooperation between researchers and mid- and upper-level management. It also requires constant communication between researchers and street-level practitioners to ensure that their daily practices conform to the treatment that they intend to occur.

The Drug Market Analysis Program, which fostered a series of randomized experiments on crime hot spots, suggests that "ordinary" criminal justice agencies can be brought on board to participate in experimental study if there is strong governmental encouragement and financial support that rewards participation (see Weisburd et al. 2006). A similar experience in the Spouse Assault Replication Program (SARP) reinforces these observations. Joel Garner (2002), who served as program manager for SARP, noted that he knew that the program was a success the "day that we got 17 proposals with something like 21 police agencies willing to randomly assign offenders to be arrested" (Weisburd 2005, p. 232).

Day-to-day implementation of place-randomized trials also requires significant forethought on the part of both the researchers and middle managers who are involved with the experiments. For instance, during Lum and colleagues' (2010) trial of the effectiveness of license plate recognition systems, the research team made an effort to directly involve both the officers who would be implementing the intervention and the middle managers who would be responsible for directly supervising the officers in the process of identifying hot spots, developing mechanisms for reporting officer patrol, and overseeing implementation. Given that the intervention involved the participation of two separate police departments, it was also important to understand and account for different practices in supervising officers, in the responsibilities of officers to continue to respond to calls for service, and other functions. Gaining an understanding of the experience and obligations of these practitioners allowed the researchers to revise plans for patrol deployment and ensure that the experiment could realistically be implemented given the normal practices of the two police departments.

The reports on deploying programs and the randomized trials that are published in peer-reviewed journals can be excellent, but they are typically brief. The brevity invites broad questions about how the authors' experience *in detail* can be shared with others, that is, through web-based journals, reports without page limits, workshops, and so on. It invites more scientific attention to the question of how one can dimensionalize implementation and the engineering questions of how to measure or observe implementation level inexpensively and how to establish a high threshold condition for implementation.

Resources for the Trial's Design, Statistical Problems, and Solutions

Some of the important technical references for trial design and model-based data analysis include Hayes and Moulton (2009) on cluster-randomized trials in different countries, mainly in the health arena. Murray's (1998) book on group-randomized trials focuses more on individuals within groups and considers diverse applications in the USA. Bryk and Raudenbush's (2002) text considers in detail model-based statistical analyses of multilevel data that may or may not have been generated in randomized trials. The models are complex and entail assumptions that the analyst may not find acceptable.

In the context of place-randomized, cluster-randomized, or group-randomized trials, there are few publications covering the simplest and *least* model-dependent approaches to analyzing data from such a trial. Such approaches fall under the rubric of randomization tests or permutation tests. A simple randomization test, for instance, involves computing all possible outcomes of the trial, ignoring the actual allocation to intervention or control conditions, and then making a probabilistic judgment about the dependability of the effect detected based on the distribution of possible outcomes so generated. There is no dependence on linear or other parametric models.

Given the relatively small sample sizes involved in place-randomized trials (typically less than 200 are allocated to one or another intervention) and given contemporary computing capacity, generating the distribution of possible outcomes in these approaches is relatively straightforward. The basic concepts were described by Sir Ronald Fisher in the 1950s for trials in which individuals or plants were the independent units of random allocation and analysis. The idea is directly relevant to trials in which places are randomly allocated and inferences are made about the effects of intervention in places.

The Campbell Collaboration organized conferences on place-randomized trials in 2001 and 2002 to update researchers on the design and conduct of place-randomized trials. A special issue of the *Annals of the American Academy of Political and Social Sciences* (Vol. 599, 2005) resulted from these. After 2002, the William T. Grant Foundation funded workshops and software development to enhance the technical capacity of researchers to design such trials and analyze results. At the time of this writing, there was no information on more specialized efforts sponsored by the National Institute of Justice or private foundations in criminology.

One of the major scientific challenges in designing a place-randomized trial is assuring that the size of the sample of places is large enough to detect the relative effects of interventions. A statistical power analysis is essential to this planning process and allows the researcher to determine the number of places that are needed given assumptions about the expected effect size, randomization levels, and specific statistical tests of hypotheses and related procedures. User-friendly software for estimating the model-based statistical power of a trial under various assumptions about sample size and other factors is available on the William T. Grant Foundation's website (http://www.wtgrantfdn.org).

There are a number of common mistakes made during analysis of place-randomized trials. First, researchers have often wrongly employed the number of people in places for hypothesis testing and power analysis *rather* than employing the number of places randomly allocated (see, e.g., Kelling et al. 1974). Second, simple linear regression models are often used to analyze data even though many of the modeling assumptions are not necessarily true, as the "uncertainty" in randomized trials stems from the assignment process, rather than unobservable disturbances that are assumed to be the source of variation in regression models (see Freedman 2008). It is very difficult to determine how misleading the results of such regression analyses are likely to be in multilevel contexts. Analyses based on regression models with categorical, count, or time-to-failure outcomes create additional problems (Freedman 2008).

Place-based trials often involve the study of people within places, and thus, a very small number of independent units (places) may be present at the randomized level and a large number of related units (people) may be nested within each place. A key issue in this scenario is that individuals or other units within a place were *not* randomly assigned to the intervention or control conditions. Randomization at the place level ensures that preexisting differences between treatment and control groups in the study are not systematic; this is due to the matching and subsequent randomization process. However, since all individuals within each place are automatically assigned the same condition, intervention, or control, the analysis is riskier to carry out at the individual level (Raudenbush and Bryk 2002; Spybrook et al. 2006).

Prior techniques for clustered data involved either aggregating all information to the group level as in a randomization test or a "t-test" of mean differences between place level outcomes or have involved disaggregating all group-level traits and confining attention to all the individuals within the intervention and control conditions. For Raudenbush and Bryk (2002), the problem with the first method is that all within group information is wasted; it is omitted from the analysis. The problem with the second method is that the observations are no longer independent, as all individuals within a certain intervention or control condition will have the same value on a certain variable and will not be independent from one another on account of the intra-class correlation among people within place. To account for the dependence that arises when using samples of individuals nested within places, researchers can use several statistical methods to correct for clustering, such as hierarchical linear modeling techniques (i.e., HLM) and generalized least squares estimation techniques with robust standard errors.

There are several options for analyzing the results of place-based randomized experiments in a largely model-free manner. Small and colleagues (2008), for instance, emphasize randomization/permutation tests with different kinds of adjustments for covariates, not relying on any form of HLM. Imai, Kind, and Nall (2009) advanced the state of the art by showing how matched pair designs in place-randomized trials can often enhance precision in estimates of the effects of interventions, increase statistical power of analysis in detecting effects, and better assure unbiased estimates of variability of the effect with a small number of places. Matched pair designs can also be used in this context, but the concern is that matching may unnecessarily reduce degrees of freedom and statistical power of the studies.

As a practical matter, the work by Imai and his colleagues (2009) shows that the number of units within a place are an important matching variable that can reduce costs and increase statistical power in a trial. Bloom and Riccio (2005) advance the state of the art in another important respect by coupling conventional comparison of randomized places in a time series analysis of data from places. While the coupling of randomized and nonrandomized approaches to estimating effects of interventions is not a new concept (see, e.g., Boruch 1985), incorporating these ideas thoroughly into design and analysis is unusual.

Another novel contribution relating to designing place-randomized trials lays in the idea of "step wedge" and "dynamic wait list" designs. The wait list design assigns a random half sample of eligible places to the intervention for some period of time. The remaining half sample receives the intervention after this period has ended, under the assumption that all units will eventually receive the intervention. While this approach satisfies ethical concerns that may arise if a beneficial treatment is not administered to certain places, it may be problematic if there are treatment consequences for a delay in treatment. To address such concerns about the sequence of treatment, places may also be divided into subsets which are periodically given treatments of a given intensity at a given time period. In a related area of work, Hussey and Hughes (2006) develop the idea of step wedged designs, with the "steps" being a point at which a subset of places is randomly assigned a treatment.

The step wedge and dynamic wait list designs are innovative and have attractive features. However, the power and design analysis of these approaches have relied on hierarchical linear models with unusual and sometimes untestable assumptions. There may be alternatives that are less model dependent.

Registers of Place-Randomized Trials

Learning about place-randomized trials can be difficult. Reports may be published in a variety of scientific journals and may appear in unpublished (gray) literature. Relying on web-based searches may not be as effective as hand searches in terms of locating randomized trials in social sciences. Taljaard et al. (2009), for instance, found that fewer than 50 % of published place-randomized trials in health are appropriately classified in titles and abstracts.

There are a number of data sources concerning randomized trials that can help to reduce these difficulties. For instance, the Cochrane and Campbell Collaborations, which are concerned with the fields of health care and social sciences, respectively (including psychology, education, criminal justice, and sociology), publish reviews of results of randomized trials in which references to original sources can be accessed through their websites. In the United States, one can learn about trials in the health sector, including place-randomized trials, at http://www.clinicaltrials.gov. Often the interventions tested or the outcome variables examined in these trials are pertinent to criminologists as well as health-care researchers and practitioners. Finally, in 2007 David Greenberg and Mark Schroder developed a new register of randomized trials oriented toward tests of economic interventions under the auspices of the Social Sciences Research Network (SSRN).

Standards for reporting randomized trials where individuals are the unit of allocation were the product of the late 1990s. The Consolidated Standards of Reporting Trials (CONSORT) statement, which provided guidance on the composition of a report on randomized trial in health care, has been modified to provide guidance on reporting place-randomized trials. For instance, studies should provide information on the rationale for a randomized design and measure outcomes on all levels of sampling and statistical inference. As of this writing, there are no uniform standards for reporting randomized trials in criminal justice. These problems have been raised by scholars of criminology who have emphasized that "reporting validity" is critical for the development of experimental science in criminology (Perry et al. 2009; Gill 2009; Farrington 2003). A lack of clear and uniform standards in this area of research means that reanalysis of experimental studies is difficult and often impossible. While some funding agencies have tried to make policy to assure that independent analysts have access to record data in research that they sponsor, clear criteria such as that included in the CONSORT statement are critical for advancing experimental criminology.

Ethics and the Law

As of this writing, no professional society or government agency has promulgated explicit statements about the ethical propriety of place-randomized trials. Applying contemporary standards to place-randomized trials can be awkward and imperfect (see, e.g., Boruch 2007), and at times, standards of ethics and government regulations are not clearly relevant to place-randomized trials (see, e.g., Taljaard et al. 2009, for a discussion of these issues in health care). For instance, ethical standards established for the treatment of human subjects are arguably inappropriate for randomly allocated treatment in a city's crime hot spots. The ethical issues surrounding place-based trials in criminology are complex. For instance, researchers may confront concerns about equality of treatment, as in research surrounding hot spot policing, wherein residents were concerned that concentrated patrol in one area would result in reduced services in other locations (see Weisburd 2005). Or officers could feel that sitting for extended periods of time reduces their capacity to protect the public (Weisburd et al. 2006). Finally, residents of areas that are designated as crime hot spots may feel as if they are being unfairly targeted by the police. In these cases, the implementation of a place-randomized trial can undermine the relationship between the practitioners involved in the trial and the clients whom they serve (see also Clarke and Cornish 1972).

Regardless of the ethical dilemmas engendered by place-based trials over the last decade, there appear to have been no serious challenges in the US courts to the conduct of place-randomized trials. For instance, the former director of the Institute for Education Sciences (IES), Russ Whitehurst, reported in a personal communication (2008) that he had encountered no court challenges as a consequence of the IES's sponsoring many such trials in education during 2002–2008. Similarly, we are aware of no judicial challenges in the context of place-randomized trials in the crime sector, such as the trials on crime hot spots, bar room violence, convenience store vulnerability to holdups, and so on.

Conclusions and Implications for the Future

Place-randomized trials have become important not only in criminology but also in a broad range of other disciplines because they employ substantive theory about the effects of intervention at the place level. The experience of people who have been involved in the design, execution, and analysis of place-randomized trials is an important source of intellectual and social capital. This experience must be exploited by future researchers and incorporated into graduate and postgraduate education, as should the knowledge of researchers in disciplines such as medicine and education who have substantially advanced the state of knowledge about more generally cluster-randomized trials. The statistical armamentarium for design and analysis of place-randomized trials is fundamental and lies on the simple idea of randomization rather than complex statistical models. Future development of place-randomized trials should better explore the implications of randomization of different interventions when the expected effects of the interventions are in "equipoise" and advance investigations into the ethics of this type of research. Further, standardized reporting of design, execution, and analysis of place-randomized trials should be advanced in criminology to allow for reanalysis and secondary analysis, particularly of controversial studies.

Walter Lippmann, an able social scientist and newspaper writer, had a strong interest in cops and crimes by adults and adolescents and was familiar with political ambivalence about or opposition to sound evidence. He was a street-level criminologist, remarkable writer, and good thinker. In the 1940s, Lippmann said: "The problem is one for which public remedies are most likely to be found by choosing the most obvious issues and tackling them experimentally...the commissions of study are more likely to be productive if they can study the effects of practical experimentation." Nowadays, trialists in criminology would have little difficulty in subscribing to Lippmann's counsel.

Related Entries

- ▶ Cambridge-Somerville Youth Experiment
- ▶ Criminology of Place
- ▶ History of Geographic Criminology Part I: Nineteenth Century
- ▶ History of the Statistics of Crime and Criminal Justice
- ▶ Hot Spots and Place-Based Policing
- ▶ Randomized Block Designs
- ▶ Randomized Experiments in Criminology and Criminal Justice

Recommended Reading and References

Berk R (2005) Randomized experiments as the bronze standard. J Exp Criminol 1(4):417–433

Bertrand M, Duflo E, Mullainathan S (2002) How much should we trust difference in differences estimates? Working paper 8841. National Bureau of Economic Research, Cambridge

Bloom HS, Riccio JA (2005) Using place random assignment and comparative interrupted time-series analysis to evaluate the jobs-plus employment program for public housing residents. Ann Am Acad Polit Soc Sci 599:19–51

Boruch RF (2007) Encouraging the flight of error: ethical standards, evidence standards, and randomized trials. N Dir Eval 133:55–73

Boruch RF, Wothke W (1985) Seven kinds of randomization plans for designing field experiments. N Dir Prog Eval 28:95–113

Bryk AS, Raudenbush SW (2002) Hierarchical linear models. Sage, Thousand Oaks

Clarke RVG, Cornish DB (1972) Controlled trial in institutional research – paradigm or pitfall for penal evaluators. The Home Office, London

Farrington DP (2003) Methodological standards for evaluation research. Ann Am Acad Polit Soc Sci 587(1):49–68

Freedman DA (2008) Randomization does not justify logistic regression. Stat Sci 23:237–249

Gill C (2009) Reporting in criminological journals. Report for seminar on advanced topics in experimental design. Graduate School of Education and Criminology Department, University of Pennsylvania, Philadelphia

Hussey MA, Hughes JP (2006) Design and analysis of stepped wedge cluster randomized trials. Contemp Clin Trials 28(2):182–191

Imai K, King G, Nall C (2009) The essential role of pair matching in cluster randomized experiments, with application to the Mexican universal health insurance evaluation. Stat Sci 24:29

Kelling GL, Pate T, Dieckman D, Brown CE (1974) Kansas City preventative patrol experiment: a technical report. Police Foundation, Washington, DC

Lippmann W (1963) The young criminals. In: Rossiter C, Lare J (eds) The essential Lippmann. Random House, New York, Originally published 1933

Lum C, Merola L, Willis J, Cave B (2010) License plate recognition technology (LPR): impact evaluation and community assessment. Accessed from http://gemini.gmu.edu/cebcp/lpr_final.pdf
Mazerolle L, Price J, Roehl J (2000) Civil remedies and drug control: a randomized trial in Oakland, California. Eval Rev 24(2):212–241
McCord J (2003) Cures that harm: unanticipated outcomes of crime prevention programs. Ann Am Acad Polit Soc Sci 587(1):16–30
Murray D (1998) Design and analysis of group randomized trials. Oxford University Press, New York
Perry AE, Weisburd D, Hewitt C (2009) Are criminologists reporting experiments in ways that allow us to access them? Unpublished manuscript/report
Raudenbush SW, Bryk AS (2002) Hierarchical linear models: applications and data analysis methods. Sage, Thousand Oaks
Shadish WR, Clark MH, Steiner PM (2008) Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. J Am Stat Assoc 103(484):1334–1356
Sherman LW, Weisburd DL (1995) General deterrent effects of police patrol in crime "hot spots": a randomized, controlled trial. Justice Q 12(4): 625–648
Spybrook J, Raudenbush SW, Liu X, Congdon R (2006) Optimal design for longitudinal and multilevel research: documentation for the "Optimal Design" software. Accessed from http://rmcs.buu.ac.th/statcenter/HLM.pdf
Taljaard M, Weijer C, Grimshaw J, Bell Brown J, Binik A, Boruch R, Brejhaut J, Chaudry S, Eccles M, McRae A, Saginur R, Zwarenstein M, Donner A (2009) Cluster randomized trials: rational design of a mixed methods research study
Warburton AL, Sheppard JP (2000) Effectiveness of toughened glass in terms of reducing injury in bars: a randomized controlled trial. Injury Prev 6:36–40
Weisburd D (2003) Ethical practice and evaluation of interventions in crime and justice: the moral imperative for randomized trials. Eval Rev 27(3): 336–354
Weisburd D (2005) Hot spots policing experiments and lessons from the field. Ann Am Acad Polit Soc Sci 599:220–245
Weisburd DL, Green L (1995) Policing drug hot spots: the Jersey City drug market analysis experiment. Justice Q 12(4):711–735
Weisburd D, Lum C, Petrosino A (2001) Does research design affect study outcomes? Ann Am Acad Polit Soc Sci 578:50–70
Weisburd D, Wycoff L, Ready J, Eck JE, Hinkle JC, Gajewski F (2006) Does crime move around the corner? A controlled study of spatial displacement and diffusion of crime control benefits. Criminology 44(3):549–591

Plea Bargaining

Ronald F. Wright
Wake Forest University, Winston-Salem, NC, USA

Synonyms

Negotiated guilty pleas

Overview

As criminal justice systems increase in volume, they turn to negotiated guilty pleas as a primary method to dispose of cases. The prosecution offers reduced sentencing risk to defendants in exchange for more certain and inexpensive convictions. Although such negotiated outcomes are unpopular with the general public, the full-time practitioners who operate the system think of plea bargains as a necessary practice in busy courts. Negotiated guilty pleas have become by far the most common disposition of criminal charges in the United States, and similar practices are becoming more prevalent in other countries. Plea bargains create some uncertainty about the accuracy of convictions in the system and alter the balance of power among criminal justice actors by making judges and juries less influential by comparison to prosecutors. While external legal limits on plea bargains have mostly proven ineffectual, prosecutor offices routinely place internal administrative limits on the use of plea bargains.

Introduction

Criminal defendants in the United States usually decide to plead guilty. Very often, defendants negotiate with the prosecution to receive particular benefits as a precondition to entering a guilty plea – so often, in fact, that these so-called plea bargains have become more common and more important than criminal trials. In the United States, more than 9 out of every 10 felony convictions (and more than 99 out of every 100 misdemeanor convictions) derive from guilty pleas, and the overwhelming majority of guilty pleas occur after the parties negotiate the terms of that plea. Plea negotiations have been slower to arrive in other countries, but criminal court systems around the world often deal with increased volume by allowing some form of discounted sentence in return for a negotiated settlement. Plea bargains, in some form or another, are critical to the operation of high-volume criminal justice systems around the world today.

The dominance of plea bargaining has predictable effects on the institutions of criminal justice in the United States. Most immediately, it makes criminal juries less influential. Negotiated pleas also shift power away from the trial judge toward the parties; they particularly strengthen the hand of the prosecutor. The prosecutor does not merely present legal and factual arguments to the judge on an equal footing with the defense attorney. Instead, the prosecutor hears arguments from the defense and then selects from the available charges and evidence thus substantially impacting the ultimate sentence. The judge, in effect, reviews these prosecutor decisions for possible error in exceptional cases. This shift of power in favor of prosecutors is more pronounced in "guideline sentencing" states that link the sentence more closely to the charge of conviction and to particular factual findings about the crime.

Plea bargaining also affects outcomes in the criminal courts. For one thing, system volume goes up. A system that relies heavily on negotiated pleas can process more cases than a jurisdiction that resolves more of its cases through trials. Negotiated plea bargains also probably trade the quality of convictions for a higher quantity of cases. As systems get busier, the risk of erroneous convictions increases.

External limits on the terms of negotiation between prosecution and defense are available in some systems but are still unusual and have shown only limited practical effects. Internal limits on bargaining practices imposed within the bureaucratic hierarchy of the prosecutor's office are still the most common limits on individual sentencing discretion of a prosecuting attorney.

This entry begins with a review of the most common topics for bargaining between the defense attorney and the prosecutor in criminal cases. It then discusses the prevalence and systemic effects of those party negotiations, both domestically and internationally. The entry closes with a review of the external and internal controls that legal systems use to structure prosecutorial discretion.

Topics for Plea Negotiations

During plea negotiations, the prosecution can offer the defendant two categories of benefits in exchange for a plea of guilty. First, under a "sentence bargain," the prosecutor agrees to recommend to the judge a sentence below the maximum available, without amending the charges. A variation on this type of agreement involves a "fact bargain." In this situation, the prosecution and defense agree to represent to the court that certain facts were either present or absent in a case. Under the sentencing rules of some jurisdictions, these facts trigger specific sentencing consequences.

The second major category of negotiated plea aims for a "charge bargain." In this situation, the prosecutor agrees to reduce the most serious charge to some lesser offense or to reduce the total number of counts in the indictment or information. The effect of amending the charges is to reduce the defendant's potential exposure to more severe sentence outcomes.

Sentence Bargains

The parties develop detailed knowledge about the strength of the evidence, the expectations of the victims, the most likely outcomes of any legal issues presented, and the normal sentences imposed in similar cases. The judge, meanwhile, does not get involved in the details of the case as early as the parties and never becomes as familiar as the parties with the specifics of the crime and the offender's background. When the court faces a crowded docket and the parties assure the judge that a given sentence is appropriate – particularly if that recommendation falls roughly into a normal range of sentences for this crime – the judge usually accepts the recommendation (Scott and Stuntz 1992).

Sometimes the parties want a level of certainty about the sentencing outcome that is not possible based on sentencing recommendations alone. The criminal procedure rules in some jurisdictions allow the defendant to enter a plea of guilty conditioned on the judge's acceptance of the agreed-upon sentence. In other words, the defendant may withdraw the plea of guilty if the court later rejects the specified sentence. Some scholars favor sentence bargains that give the judge something more than a "take it or leave it" option; more substantial input for the judge in the selection of the sentence leads, according to this view, to a more balanced sentencing system (Alschuler 1976; Wright and Miller 2002).

Before accepting a plea of guilty, the trial court judge must be convinced that the defendant's plea is "knowing and voluntary" and that the prosecution holds a sufficient "factual basis" to prove each element of the crime beyond a reasonable doubt. Other facts may be relevant to the sentence, as well, even if they are not elements of the offense. For instance, in the federal system, a defendant who plays a "minimal role" or "minor role" in a group offense receives a lower sentence.

Given this opportunity to control the available sentence through the facts they plan to present, the parties naturally turn to negotiations about the proof of these non-element facts; they agree in some cases to tell the judge that the fact is present or absent from the case. The parties cannot lie to the judge and cannot compel the judge to make the factual finding that forms the basis for their agreement. But the prosecutor and defense attorney do exercise serious practical influence when they present a united front to the judge.

This subtype of sentence bargaining is sometimes known as "fact bargaining." It is especially pertinent in jurisdictions with sentencing guidelines or other statutes that attach specific sentencing consequences to particular facts. For instance, some state codes authorize a 5-year increase in a sentence based on the use of a weapon during commission of the crime. The prosecution and defense might negotiate an agreement to stipulate to the judge that no weapon was used during the crime.

The prosecutor sometimes goes beyond the role of advocate and plays a strong gatekeeper role for some non-element facts. In these special situations, the prosecutor does not merely request a factual finding from the judge; instead, the prosecutor herself exercises the power to increase or decrease a sentence automatically. For example, in the federal system (and in some states), the offender is eligible for a sentence discount after cooperating in the government's investigation of further crimes only if the government certifies that the defendant provided "substantial assistance." The government's willingness to file such a motion is one of the important bargaining chips available to the prosecutor. Some states give the prosecutor a similar gatekeeper function over certain sentencing enhancement facts, such as the proximity of a narcotics sale to school grounds or the fact that a defendant's prior record makes him or her a "habitual felon." The defense attorney and prosecutor routinely bargain to determine whether the prosecutor will file the allegations that trigger such a sentencing enhancement.

Charge Bargains

Criminal codes in both the federal and state systems give prosecutors a generous menu of options in the selection of charges. Many common fact scenarios could support criminal charges under multiple sections of the criminal code, each leading to different potential sentencing outcomes.

After the initial filing of charges and the assignment of defense counsel to a case, the parties often negotiate over possible amendments to those charges. The defense attorney might request a dismissal of the most serious charge to be replaced by a less serious charge (a "vertical" charge bargain). Alternatively, the defense lawyer might ask for a dismissal of some charges in a multi-count indictment (a "horizontal" charge bargain). Both types of amendments reduce the maximum sentence that a judge could legally impose on the defendant.

Negotiations between the parties about revised charges take on even greater importance in the context of more highly structured sentencing systems. Sentencing laws that structure or reduce judicial discretion, including "mandatory minimum sentences" and "presumptive sentencing guidelines," make the charge of conviction a more important predictor of the sentence. Once the defendant is found guilty of a particular crime, these highly structured sentencing laws give judges relatively few options for how to sentence the defendant (and perhaps no options at all). Because a structured sentencing environment makes visible the sentencing consequences of charging decisions, the parties can isolate the impact of a charge reduction and negotiate based on a more certain prediction of the outcome.

Mandatory minimum sentencing laws offer a clear-cut example of the impact of charge bargains on sentencing outcomes. These laws create "cliff effects" that dramatically affect the potential sentence, depending on whether the parties agree to the mandatory sentence versus a similar crime with no mandatory penalty attached. Mandatory penalties could lead to wholesale increases in sentence severity in a jurisdiction. Empirical studies confirm, however, that prosecutors selectively mitigate the impact of specific mandatory sentencing laws by dismissing or reducing charges in some cases (Bjerk 2005).

Presumptive sentencing guidelines tie sentencing options to the charge of conviction, much like mandatory minimum sentencing laws do. Systems that link the charge of conviction to a narrow range of sentence outcomes limit the options for the sentencing judge and empower the prosecutor, who controls both the filing of charges and the proof of the defendant's conduct.

Empirical studies have examined what prosecutors in guideline jurisdictions actually do about charge reductions. For the least serious cases, studies found only small shifts from sentence bargains to charge bargains. Apparently, busy prosecutors in high-volume systems do not change their customary negotiation practices across the board.

Charge bargains increased more markedly, however, in more serious cases. In cases where conviction on the original charge would result in a presumed prison sentence and conviction on a lesser charge would allow probation or a shorter jail sentence, charge bargains happened more often after the arrival of sentencing guidelines in a jurisdiction. The amended rules gave the parties greater control over the sentencing outcomes in those cases, and they took advantage of that new power (Miethe 1987; Frase 2005). Research on the charging and plea bargaining practices of federal prosecutors under the federal sentencing guidelines also suggests that prosecutors frequently exercise their charge bargaining discretion to reduce sentences, particularly in drug and weapon possession cases (Bowman and Heise 2001). On the whole, the empirical evidence suggests that sentencing guideline systems make charge bargains somewhat more attractive than they are in discretionary sentencing regimes.

There are also questions about the equal distribution of charge bargain benefits. The presence of legally relevant facts is the most important determinant of the charge of conviction (i.e., the charge that ultimately forms the basis for the conviction). Some nonlegal factors, however, also have some effect on the charge of conviction. Factors such as the race of the defendant appear to have a relatively small but persistent impact on the outcomes (Steffensmeier et al. 1998).

Defense Concessions

While the prosecutor can offer reduced sentences to the defendant, the defense attorney also holds certain bargaining chips during plea negotiations. Each of the defendant's concessions involves the removal of procedural hurdles from the prosecutor's path.

The defendant's power to make the government's work easier begins with the investigation. If the defendant cooperates in an ongoing investigation, the government might be able to convict additional defendants. This form of assistance is especially important in the federal

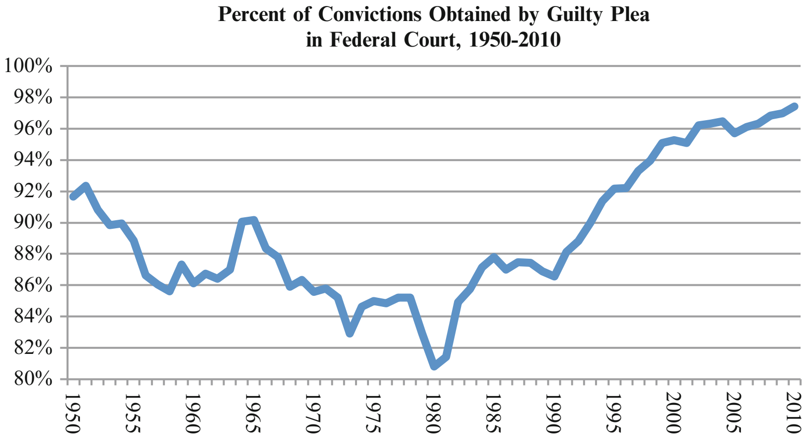

Plea Bargaining, Fig. 1 Percent of convictions obtained by guilty plea in federal court, 1950–2010

system. The discounts available for "substantial assistance" lead to some difficult anomalies in the sentences among defendants who commit their crimes as part of a group, with the largest discounts awarded to the most blameworthy (and knowledgeable) organizers of the criminal enterprise (Maxfield and Kramer 1998).

Another procedural hurdle that the defendant can remove for the prosecution is discovery and disclosure. The prosecution has a constitutional duty to disclose all material exculpatory information in its possession and a duty under state statutes or procedural rules to respond to any discovery requests for certain types of inculpatory evidence. The defendant, however, can remove most of those obligations from the prosecution by pleading guilty.

The most valuable concessions that defendants make in plea negotiations are waivers of pretrial hearings and the trial itself. These waivers might extend to all the procedural rights at trial, including the right to a jury, confrontation of adverse witnesses, counsel, and so forth. In some jurisdictions, defendants can obtain some benefits by offering to waive the jury trial in favor of a bench trial (Schulhofer 1984).

The defendant can also offer the government certainty and finality of outcomes by waiving the right to appeal or to file post-conviction collateral attacks on the conviction. Most state and federal courts have concluded that a defendant may explicitly waive the right to appeal a conviction as part of a plea agreement. One empirical study of federal cases found that nearly two-thirds of the cases settled by plea agreement included a waiver of appeal rights and three-quarters of the defendants who waived appeal also waived collateral review (King and O'Neill 2005).

Prevalence of Plea Negotiations

The concessions that the prosecutor and the defense attorney offer each other during plea negotiations have become the norm in criminal justice: plea agreements account for the great majority of convictions in every jurisdiction in the USA. The proportion of negotiated pleas changes over time and across different systems. Negotiated outcomes were not at all common in the mid-nineteenth century in state systems (Fisher 2003). In the federal system, less than 8 % of convictions were obtained through guilty pleas in the early 1970s; that number rose inexorably through the decades to reach the current level, above 97 % (Wright 2005). See Fig. 1.

A high volume of cases within a court system – both the civil and criminal dockets – creates the conditions for plea bargaining to thrive (Bibas 2012). Courts that resort to plea bargaining, in turn, increase their caseloads even further. Routine practices and expectations also play a role in sustaining the institution of plea bargaining: some courts with relatively uncrowded dockets still depend almost entirely on negotiated pleas.

The United States is unusual in an international context in its heavy reliance on party negotiations to resolve criminal proceedings and the tight connection it promotes between plea negotiations and sentencing outcomes. For example, only a generation ago, Germany was known as the "land without plea bargaining" (Langbein 1979). Over time, however, many European nations have faced a higher volume of criminal cases; crowded court dockets have led to various innovations that allow the prosecutor to designate some cases for summary treatment.

In some countries, party negotiations to achieve these more streamlined criminal adjudications have become more common. Parties not only negotiate guilty pleas but also dismissals and diversion into alternative punishment or restitution programs (Jehle and Wade 2006; Luna and Wade 2010). In the past, these negotiated summary dispositions were restricted to less serious cases, but that limitation is disappearing over time (Langer 2004). Plea bargaining is somewhat more prevalent in Israel, Canada, Australia, England, and Wales than in European Continental countries.

Negotiation Effects

It is difficult to say whether plea bargains reduce the typical sentence imposed on a defendant over the long run. On the one hand, defendants who plead guilty to a crime receive a definite "discount" for waiving trial rights, when compared to defendants facing the same charges during the same time period. On the other hand, as the courts rely more heavily on plea bargains and prosecutors file more charges, the legislature tends to pass new criminal laws that increase the risk for defendants who refuse to plead guilty. That is, the new laws authorize higher sentences that defendants potentially face if they insist on a trial and lose. Over the long run, reliance on plea bargains does not correlate with reduced sentences (Pfaff 2011).

Other potential effects of plea bargains go beyond the sentences imposed, raising issues about the accuracy of system outcomes and the influence of different system actors.

Questions About Accuracy

Established constitutional doctrine declares that the courts may not punish a criminal defendant simply for exercising the constitutional right to a trial. Daily practice in every American criminal court, however, contradicts this doctrine. It is clear that defendants routinely receive more severe sentences after trial than they would receive if they were to plead guilty (Brereton and Casper 1981). Judges resolve this apparent conflict between constitutional doctrine and daily reality by declaring that they reduce sentences in guilty plea cases based on the defendant's cooperative attitude and prospects for rehabilitation, and not based simply on the fact that the defendant waived trial rights.

The exact size of the "trial penalty" – or to put it more politely, the "guilty plea discount" – is difficult to measure. Studies of the trial penalty that attempt to control for the seriousness of the offense and other variables find a substantial gap between posttrial sentences and post-plea sentences. One study found a wide range of differences – between 13 % and 461 % – depending on the crime and the jurisdiction involved (King et al. 2005).

When defendants face such a large trial penalty, concerns start to mount that some defendants with valid defenses nevertheless plead guilty. Although there is a reasonable prospect that such defendants would be acquitted at trial, they dare not risk the large increase in the sentence that happens after a conviction at trial (Wright 2005). Behavioral economics offers some reasons to believe that defendants will undervalue the long-term impact of a felony conviction, leading defendants to accept a guilty plea too easily (Bibas 2004). On the other hand, there is evidence that some innocent defendants resist highly attractive plea offers, even when it might be rational to accept the offer in light of the risk of a wrongful conviction (Gazal-Ayal and Tor 2012).

The risk of coercing innocent defendants into pleading guilty is built into some basic structural features of American criminal justice. Criminal codes that offer more bargaining options to the prosecutor make it possible to increase the trial penalty and to pressure some defendants into waiving potentially effective defenses. The same holds true for increased sentence severity. When the law authorizes higher maximum sentences for a wide range of offenses and retains low potential sentences for lesser-included offenses, defendants face an enormous range of risk. Particularly in those states with sentencing guidelines or other limits on judicial sentencing discretion, the defendant can control some of that risk through a charge bargain. It would be an overstatement to claim that sentencing guidelines lead directly to more inaccurate convictions. They can, however, contribute to a coercive environment for defendants.

Finally, the structure of criminal justice institutions increases the risk of coercive and inaccurate guilty pleas. The public does not invest in enough judges, courtrooms, prosecutors, and public defenders to try a substantial proportion of the cases filed each year. The "working group" dynamic that develops in most courtrooms places the highest value on agreements that will move cases more quickly through the system (Nardulli et al. 1985).

Balance of Power Among Sentencing Actors

Sentencing guidelines and other structured sentencing laws that began to proliferate in the 1970s were designed to regulate judicial discretion in sentencing. As a result, some critics of these laws expressed concern that the reforms would transfer power from judges to prosecutors. Compared to traditional indeterminate sentencing schemes, the more structured systems would concentrate sentence authority in one branch rather than allowing one institution to check and balance the other (Alschuler 1991).

The transfer of power hypothesis finds some tentative empirical support. For instance, studies of guideline systems confirm that charge bargains become more common for at least some crimes, and charge reductions determine an important component of the sentence actually served (Frase 2005). But this shift in charging practices may not result in a provable transfer of authority from judges to prosecutors since judges might find other methods to influence sentences (Engen 2008).

At the very least, fears that prosecutors would entirely usurp the judge's sentencing authority in presumptive guideline jurisdictions appear to have been overstated. For one thing, judges in an indeterminate sentencing system do not, in reality, control the sentence that an offender actually serves. The judge announces one sentence, but parole authorities could later reduce that sentence. Since the judge in an indeterminate system never held actual power over the sentence to be served, guidelines do not take that power away from the judge. It is also true that presumptive sentencing guidelines leave important zones of discretion available to judges in the selection of sentences.

Limits on Prosecutorial Discretion

Legal institutions place some controls on the negotiation of guilty pleas. Those constraints originate both from outside the office of the prosecutor and from the internal workings of a local prosecutor's office.

External Constraints

If the legislature defines crimes narrowly and sets penalties at modest levels, it reduces the risks of inaccurate convictions and increases the power of different sentencing actors to check each another. This approach to crime legislation, however, does not thrive in the American political climate. Voters expect prosecutors to take the lead in addressing crime, and they expect legislators to give them the legal tools to do the job. Legislators respond with broadly worded criminal laws and multiple statutes (each with a different corresponding punishment) that could apply to a single common factual scenario.

Although legislators do not seriously constrain prosecutors through the terms of the substantive criminal law, statutes in a few jurisdictions do limit the timing of plea negotiations or limit the size of the charge reduction that a prosecutor can offer to dispose of a case (at least for some high priority categories of crime). While these legislative directives can be meaningful, their current impact is small. Statutory limits on the timing of negotiations simply push plea bargains into earlier phases of the proceedings.

Judges also refuse, for the most part, to monitor and control the negotiation of guilty pleas. Granted, judges hold the power to accept or reject guilty pleas, along with the plea agreements that the parties present to them. These judicial powers, however, operate within a system of mass justice. The caseload would become overwhelming if judges balked regularly at proposals to remove a case from the trial docket, or even took the time regularly to investigate this possibility.

The law in some jurisdictions also limits the role of the judge during plea negotiations. Statutes, rules of criminal procedure, rules of judicial ethics, and judicial opinions in more than half of the states instruct the judge not to "participate" at all in the plea discussions. The judge's only role is to evaluate the legitimacy of the guilty plea after the parties finalize their agreement. A smaller number of states discourage judges from participating in plea negotiations, but they do not ban the practice outright. The laws in these states allow the judge to comment on the acceptability of charges and sentences that the parties themselves propose or to participate in the negotiations only upon the invitation of both parties.

The rules of professional responsibility as enforced by state licensing authorities are also a potential source of limits on the plea bargaining behavior of prosecutors. For instance, a prosecutor who withholds discoverable information from the defense during plea negotiations might violate the specialized ethical obligations of prosecuting attorneys. Again, however, we get limited accountability from these regulators. State bar authorities rarely discipline prosecutors, and the penalties are usually not severe (Zacharias 2001).

Although institutions external to the prosecutor's office do not exert much power over the discretion of line prosecutors, the chief prosecutor is directly accountable to the voters. Prosecutors in the United States are normally elected. Given the distaste among voters for the practice of plea bargaining, elections in theory should limit the prosecutor's ability to reduce charges or to recommend lower sentences as part of a plea negotiation. In practice, however, the influence of voters over the plea bargaining policies of the prosecutor's office is limited. The heavy advantage of incumbents in prosecutorial elections makes this a weak accountability mechanism (Wright 2009).

Internal Constraints

While legal institutions outside the prosecutor's office do not fully meet the need for checks and balances, internal regulation has a substantial constraining effect. Forces within the prosecutor's office can produce plea policies that remain true to declared sources of law, in keeping with current public priorities in the enforcement of that law, applied with reasonable consistency across cases (Bibas 2012; Miller and Wright 2008).

These internal regulations take several forms. First, the managers in a prosecutor's office sometimes arrange the flow of cases to encourage line attorneys to interact dispose of cases through plea negotiations. This might involve the use of "horizontal" prosecution for some crimes, with different attorneys or units in the office making decisions as a file moves up through the system. It is also common, particularly in larger offices, to require supervisor approval for any plea agreement that dismisses a charge in priority cases, such as homicide and domestic violence.

Second, chief prosecutors can promote consistency and fidelity to public values among their line attorneys by creating *written* guidelines for the disposition of cases (Podgor 2012). The United States Attorney's Manual is one such resource. The guidelines typically declare that they do not carry the force of law and are not enforceable in judicial proceedings, but they nevertheless exert some control over the behavior of prosecutors (Abrams 1971). Through the use of such guidelines, prosecutors have from time to time banned the use of plea bargains for certain classes of cases, although such guidelines require serious monitoring and enforcement to remain effective.

While there is much promise in the power of chief prosecutors to hold their line prosecutors accountable, the system also depends on professional tradition, informal office culture, peer pressure, and individual conscience to achieve just results. At the end of the day, each prosecutor must remain individually committed to the ideal of responsible prosecution.

Policy Challenges

The most significant checks on the use of plea bargains come from inside prosecutors' offices. Prosecutors in the United States are profoundly decentralized: the state courts operate over 2000 separate prosecutors' offices, with no effective hierarchical control over the local offices in most states. As a result, the internal controls on plea bargains are often nontransparent and inconsistent from place to place. Policy makers who hope to promote enforcement of the criminal law that is consistent with legal values and with current popular priorities must find ways to make these local internal policies and practices more visible and subject to evaluation by the public and by other legal actors. A balance of power, in a transparent environment, would lead to the most responsible use of negotiated guilty pleas.

Related Entries

- ▶ Prosecution and Wrongful Convictions
- ▶ Prosecutorial Discretion

Recommended Reading and References

Abrams N (1971) Internal policy: guiding the exercise of prosecutorial discretion. UCLA Law Rev 19:1–58

Alschuler AW (1976) The trial judge's role in plea bargaining (pt. 1). Columbia Law Rev 76:1059–1154

Alschuler AW (1991) The failure of sentencing guidelines: a plea for less aggregation. Univ Chic Law Rev 58:901–951

Bibas S (2004) Plea bargaining outside the shadow of trial. Harv Law Rev 117:2463–2547

Bibas S (2012) The machinery of criminal justice. Oxford University Press, New York

Bjerk D (2005) Making the crime fit the penalty: the role of prosecutorial discretion under mandatory minimum sentencing. J Law Econ 48:591–625

Bowman FO, Heise M (2001) Quiet rebellion? Explaining nearly a decade of declining federal drug sentences. Iowa Law Rev 86:1043

Brereton D, Casper J (1981) Does it pay to plead guilty? Differential sentencing and the functioning of criminal courts. Law Soc Rev 16:45–70

Engen RL (2008) Have sentencing reforms displaced discretion over sentencing from judges to prosecutors? In: Worral JL, Nugent-Borakove ME (eds) The changing role of the American prosecutor. State University of New York Press, Albany

Fisher G (2003) Plea bargaining's triumph: a history of plea bargaining in America. Stanford University Press, Palo Alto

Frase RS (2005) Sentencing guidelines in Minnesota, 1978–2003. Crime Justice: A Rev Res 32:131–219

Gazal-Ayal O, Tor A (2012) The innocence effect. Duke Law J 62:339–401

Jehle JM, Wade M (2006) Coping with overloaded criminal justice systems: the rise of prosecutorial power across Europe. Springer, Berlin

King NJ, O'Neill M (2005) Appeal waivers and the future of sentencing policy. Duke Law J 55:209–261

King NJ, Soule DA, Steen S, Weidner RR (2005) When process affects punishment: differences in sentences after guilty plea, bench trial, and jury trial in five guidelines states. Columbia Law Rev 105:959–1009

Langbein JH (1979) Land without plea bargaining: how the Germans do it. Mich Law Rev 78:204–225

Langer M (2004) From legal transplants to legal translations: the globalization of plea bargaining and the Americanization thesis in criminal procedure. Harvard Int Law J 45:1–64

Luna E, Wade M (2010) Prosecutors as judges. Washington Lee Law Rev 67:1413–1532

Maxfield LD, Kramer JH (1998) Substantial assistance: an empirical yardstick gauging equity in current federal policy and practice. United States Sentencing Commission, Washington, DC

Miethe TD (1987) Charging and plea bargaining practices under determinate sentencing: an investigation of the hydraulic displacement of discretion. J Crim Law Criminol 78:155–176

Miller ML, Wright RF (2008) The black box. Iowa Law Rev 94:125–196

Nardulli PF, Flemming RB, Eisenstein J (1985) Criminal courts and bureaucratic justice: concessions and consensus in the guilty plea process. J Crim Law Criminol 76:1103–1131

Pfaff JF (2011) The myths and realities of correctional severity: evidence from the National Corrections Reporting Program. Am Law Econ Rev 13:491–531

Podgor ES (2012) Prosecution guidelines in the United States. In: Luna E, Wade M (eds) The prosecutor in transnational perspective. Oxford University Press, New York
Schulhofer SJ (1984) Is plea bargaining inevitable? Harv Law Rev 97:1037–1107
Scott RE, Stuntz WJ (1992) Plea bargaining as contract. Yale Law J 101:1909–1968
Steffensmeier D, Ulmer JT, Kramer JH (1998) The interaction of race, gender, and age in criminal sentencing: the punishment cost of being young, black and male. Criminology 36:763–798
Wright RF (2005) Trial distortion and the end of innocence in federal criminal justice. Univ PA Law Rev 154:79–156
Wright RF (2009) How prosecutor elections fail us. Ohio State J Crim Law 6:581–610
Wright RF, Miller ML (2002) The screening/bargaining tradeoff. Stanford Law Rev 55:29–118
Zacharias FC (2001) The professional discipline of prosecutors. North Carol Law Rev 79:725–743

Plural Policing

▶ Family Engagement Strategies to Reduce Crime
▶ Legal Frameworks for Third-Party Policing

Poaching

▶ Crimes Against Animal Life

Police

▶ Privatization of Policing in an International Context
▶ Police and the Excessive Use of Force

Police Accountability

▶ Control of Police Misconduct

Police and the Excessive Use of Force

Robert J. Kane
Program in Criminal Justice, Drexel University,
Philadelphia, PA, USA

Synonyms

Brutality; Coercion; Force; Police

Overview

The excessive use of force by police is often a difficult phenomenon to identify, as well as a difficult concept to define. Unlike corruption that is motivated by profit – which is generally impermissible by its very nature – excessive use of force often resides at the marginal end of the *acceptable* use of force continuum. That is, because society authorizes the police to use physical coercion to carry out their mandate, it is often unclear when, along the use of force continuum, the acceptable use of force becomes excessive. Perhaps this is why excessive force, or brutality, often seems to occur unchecked in communities that are unable to advocate for themselves, making it difficult to hold police officers accountable for the use of unauthorized violence.

Several classic police scholars have developed occupational templates to help describe the circumstances under which some police officers engage in brutality. These perspectives try to account for police-suspect interactions, the police working environment, and dangers therein; while others locate the causes in the personalities of the officers, or the agencies in which they work. By understanding some of the causes of police brutality, society and policy-makers have a better chance of developing prevention strategies that can minimize the use of excessive force, particularly in the most vulnerable communities. Currently, these prevention strategies are rooted in the areas of recruitment and training, as well as the proper supervision of officers in the field.

Introduction

When two officers of the Los Angeles Police Department were shown on tape beating the motorist Rodney King with batons – after their field supervisor had already exposed King, with little effect, to the Taser, a conductive energy device designed to "stun" suspects into submission – even while he was rolling on the ground and seemingly defenseless – few observers were left with any doubt that they had just witnessed an example of police brutality. Some experts (e.g., Skolnick and Fyfe 1993; Toch 1996), however, who viewed the entire tape – as opposed to simply its last 7 s – saw a more nuanced escalation of force that ultimately became excessive. Rodney King had broken a primary subcultural rule in policing: When initially contacted by the team of California Highway Patrol officers who attempted to stop his vehicle for speeding on Interstate 210 in the San Fernando Valley area of Los Angeles, he failed to pull over for them. Indeed, King would later report that he fled from the CHP cruiser because he was driving while intoxicated and feared that an arrest would violate the provisions of his parole (the result of an earlier robbery conviction) (Skolnick and Fyfe 1993). For his failure to submit to police authority in ways that satisfied the five LAPD officers at the scene, King was "Tazed" once, struck with police batons 56 times (several of which were blows to the head), and kicked six times by officers (Independent Commission 1991). To date, the beating of Rodney King represents the archetype of police brutality, and it encapsulates multiple conceptual perspectives – from the levels of the social interactionist, workgroup, and cultural – aimed at explaining why some police officers resort to excessive force during their encounters with members of the public.

Excessive use of force by police is a tricky phenomenon to identify and explain. Even in the most "obvious" and infamous examples of police brutality, it is frequently the case that at some point during the violent encounter, some force on the part of police was justified – perhaps even necessary – to preserve officer and/or public safety. As such, excessive force by police is substantively distinct from other forms of misconduct, such as profit-motivated corruption because, while the later is usually wrong on its very face (e.g., police officers are not permitted to extort money from drug dealers or motorists; and they are not permitted to steal jewelry off corpses they encounter during health-and-welfare checks), they are allowed, and in many cases even expected, to use coercive force as part of their role in society. In fact, it as a virtual axiom in policing that the use of force is a "tool," whose use should be regulated but not taken away.

Several leading scholars have developed frameworks to explain *why* the excessive use of force occurs and *how* it may foster in some police departments. Van Maanen (1978) identifies the "asshole" as one lens through which to view the use of force encounter. He argues that when a suspect commits an "affront" to a police officer's authority, the officer – under certain circumstances – may respond with violence as a way of (re)establishing control over the police-suspect event. Muir (1977) takes a more individualistic approach, observing that some police officers have an "enforcer"-type personality borne out of their inabilities to (1) empathize with people whom they contact under the worst of circumstances, and (2) communicate in ways that may de-escalate potentially violent encounters. Skolnick and Fyfe (1993) view the excessive use of force from an organizational-level, noting that certain police departments develop and foster a so-called "siege mentality" in which officers are made to feel outnumbered, outgunned, and unwelcome in the communities they serve. Indeed, the siege mentality – which stems from the top levels of police department administration – views officers as a collective occupying force deployed in dangerous communities inhabited by dangerous people. As a result, officers in such locations are often made to believe that the only way for them to maintain order is through the use of highly aggressive enforcement strategies that may translate into excessive force.

This entry examines the use of excessive force by police through the perspectives offered above while also noting the difficulty and complexities involved in defining police brutality. As alluded to above, identifying excessive force is made all-the-more difficult by virtue of the use of physical coercion that is a large part of how the police fulfill their mandate in society. In addition to discussing these conceptual issues, the entry also describes how legal mechanisms, such as judicial review and the US Code, have made the excessive use of force difficult to prove at an acceptable legal standard, which may inadvertently stifle efforts to redress abusive police practices and reform police departments in need of reforming. The entry closes with a discussion of how to prevent and control the excessive use of force.

The Excessive Use of Force Prism

Before trying to define the *excessive* use of force by police, the reader might first start with a more fundamental question: *Why do we have the police in our society?* It is common knowledge that at times, some police officers use more force than is necessary on suspects they encounter; some engage in profit-motivated misconduct and then lie on the witness stand to protect their corrupt enterprises; and these practices often take place in the most socially and economically vulnerable communities (Kane 2002). So, again, while most people understand that some police officers, as well as some departments, have the propensity to engage in brutality, the question is again begged: *Why have the police in the first place?*

American citizens enjoy – perhaps more than those of other developed nations – almost unfettered access to their court system. They have the ability to report crimes directly to their local District Attorney's Offices, and they have the right to request protective orders and other forms of injunctive relief directly from magistrates in open court. In short, American citizens under the best of circumstances have the ability to mobilize government on their behalf in ways that could theoretically bypass the police entirely, and yet, most Americans – even those who have had bad experiences with them – would scarcely advocate dissolving the police as a public institution. Why? As cited by Klockars (1985), we maintain a collective police force because of what Egon Bittner wrote in 1970: "something-ought-not-to-be-happening-about-which-something-ought-to-be-done. . .*NOW!!*" That is, although it is *possible* to mobilize the government without using the police, it is often not *feasible* to do so. In highlighting this reality, Klockars (1985) notes that if a crazy resident goes into the yard of his neighbor to chop down a prized apple tree with an axe, seeking an injunction against such an intrusion likely will not prevent the destruction of the apple tree. In such events, therefore, the point of the police is to stop – or at the very least – freeze a moment in time in order for the courts to gain the opportunity to properly adjudicate the case. The police accomplish this mandate by exercising their "general right to use coercive force" (Klockars 1985) – a right bestowed upon them by the American public through the Executive Branch of government. And it is this general right to use coercive force that makes possible both the best and worst practices of policing. At their best, the police use their coercive authority to improve the life chances of people they contact – often under the most abject social and economic conditions imaginable. At their worst, police use their coercive authority in ways that abuse, denigrate, and otherwise "tread on the human dignity" (Carter 1985, p. 322) of people they encounter.

Although the American public grants police the general right to employ coercive force, it does so with few guidelines as to the boundaries of its proper use, making it difficult for evaluators of police behaviors to determine – for example – when appropriate force may have become excessive, or when excessive force was used when no force was necessary. The following serve as instructive examples:

- **Incident 1**: Two police officers on foot see a man on a public sidewalk who matches the description of a robbery suspect. They attempt to contact the man, but as they summon him their way, the man turns and runs. Both officers engage the suspect in chase. The man leads them on a 5-min foot pursuit through alleys, over fences, and across a vacant lot until they finally catch him in an open field. At the point of contact, the man physically resists the officers' efforts to subdue and handcuff him. He swings and kicks at them, making glancing contact, until one of the officers makes two baton strikes to the man's body, sending the suspect hard to the ground. One of the officers is bleeding slightly from a scratch on his face, the other suffered a hyperextended knee. As the suspect writhes on the ground from the pain of the baton strikes, one of the officers kneels to handcuff him, while the other officer kicks the man twice in the ribs. Even after the suspect is securely cuffed and clearly incapacitated, the other officer punches him with a closed fist in the kidney two times just after the other officer pulled him from the ground.

- **Incident 2**: A police officer working traffic duty sits in his marked cruiser at mid-day. The car is parked in a somewhat concealed position relative to the corner of an intersection governed by a four-way stop. When he observes a pickup truck roll confidently through the intersection without stopping, the officer ignites his engine, activates the light bar, and pulls quickly away from the curb to initiate a vehicle stop on the truck. After a few hundred yards, and two siren "chirps," the driver of the pickup truck finally pulls to the curb. The officer exits his vehicle, walks to the truck, and instructs the driver to put down his window. The officer observes that the driver and his one passenger are teenage boys, most likely between the ages of 16 and 18 years. The officer waits with expectancy while the driver – who is wearing a smirk on his face – finally begins to slowly roll down his window. The officer, losing patience, asks for the driver's operating license. After a pause, the driver says, "Why do you need that? I am just taking my buddy home." The officer again requests the young man's license, and although the driver complies, he does so slowly. Suddenly, the officer reaches in through the window, grabs the young man by the hair, and swiftly pulls him through the window and out onto the street. The officer then drags the young man – still by the hair – around the front of the truck to the sidewalk, where he slams the young man's head against the hood of the truck. The young man screams in surprise and pain, while the officer repeats the motion of bashing his head against the hood. The officer then drags the young man back around the front of the truck, opens the driver's side door, and roughly deposits the young man back into the truck, slamming the door closed. The officer then advises, "The next time a police officer asks for your driver's license, I suggest you give it to him the first time." Then officer then turns, walks back to his patrol car, enters, and drives away.

The above scenarios indicate two types of situations during which excessive force may be used by police, and both have their own definitional difficulties associated with them. In the case of incident #1, it is clear that the two officers involved in the foot chase were justified – even compelled – to use force in the apprehension and control of the suspect. The suspect, in fact, used physical force to prevent his apprehension, swinging, kicking, and striking the officers as they attempted to subdue him. The officers were probably justified in the use of their batons while the suspect was fighting them, but what about the subsequent kicks and blows once the suspect was handcuffed and effectively restrained? At what point did the appropriate use of force by the police become excessive? Different people – e.g., police officers, police administrators, community residents, and promoters of law and order – may have different answers to this question. For some, the baton strikes may have been regarded as excessive, considering there were two officers and the suspect was not at all armed with a weapon. To others, the kicks and punches delivered by officers at the conclusion of the encounter were not excessive, but rather "just desserts" visited upon a "victim" who should have cooperated with the police in the first place.

The events described in incident #2 suggest perhaps a less ambiguous evaluation of the use of excessive force. Although the driver of the pickup truck was not physically combative with the officer who stopped him, he was defiant, and he failed to comply with the lawful order of police officer, namely, to produce his operator's license. In doing so, the driver of the pickup truck left the officer with two categories of choice: (1) make a tactical retreat, perhaps to his vehicle where he could have radioed for a supervisor to join him at the scene in an effort to avoid the use of force; and (2) increase the dosage of coercion in order to secure the driver's compliance. From the description of events in incident #2, the officer clearly chose the latter by using physical coercion to neutralize the young man's defiance. And again, commentators of police behaviors may disagree as to the appropriateness of the officer's actions. Some would conclude that the force used by the officer was unnecessary, excessive, and perhaps even illegal; others may conclude that the officer rightly punished a smart-aleck kid who needed to be taught to respect authority.

Despite potential differences in assessments of the excessive use of force across incidents 1 and 2, the hypothetical events share a common link: From the perspective of responding officers, both victims may have "deserved" it. That is, in both cases described above, an argument can be made that both suspects-turned-victim forced a confrontation with the police, which led to the excessive force. This is another feature of many use of force incidents that makes identifying brutality or excessive use of force difficult. As Walker (2010) has argued in his criticism of drug-war policing, when police officers use excessive force during their encounters with the public, they usually do so on socially marginalized and/or otherwise "undesirable" suspects, often rendering it unlikely that the victim of brutality can successfully hold officers accountable for their actions. Thus, despite the procedural and legal constraints (discussed in the following section) placed on officers' authority on the use of physical coercion, the sociology of use of force encounters often challenges society's definition of the excessive use of force, particularly when the force may have been used on someone who at least initially defied the police.

The following section discusses several of the legal issues associated with defining the excessive use of force, which adds further difficulty of successfully defining excessive force and brutality.

Excessive Force Through Legal and Social Lenses

Moving more thoroughly into an evaluation of the excessive use of force by police requires some definitional clarity. In general, police use of force can be categorized into four types: (1) appropriate, (2) deadly, (3) excessive, and (4) unnecessary. Once these types of force are defined, the entry focuses on excessive force, also commonly referred to as police brutality (e.g., Skolnick and Fyfe 1993).

The appropriate use of force occurs when police officers use physical coercion that is proportional to a suspect's resistance to gain compliance and/or diffuse a violent or potentially violent encounter. Though perhaps commonly believed, police officers are not required to match a suspect's resistance incrementally with force that follows a use of force continuum (*Graham v. Connor* 1989). Rather, police officers are legally (and usually administratively) permitted to "skip" steps on the use of force continuum in order to overwhelm a suspect – within reason – to achieve custody. Society and the courts do not require police officers to engage in fist fights with suspects in order to gain compliance. Indeed, most police officers are trained to control many suspects – whether armed or not – from a safe distance, often via the threat or use of chemical spray, a conductive energy device (e.g., Taser), or a firearm. To provide a framework for assessing the appropriateness of the use of force, the Supreme Court applied an "objective reasonableness" standard (via the Search and Seizure Clause of the Fourth Amendment) in the case *Graham v. Connor* (1989), holding that the legality of the use of force must be evaluated based on the judgment of a "reasonable" officer at the scene of the event rather than through a "retrospective" lens. The Court reasoned that when police officers become involved in violent or potentially violent encounters, they frequently have just seconds to decide how to apply coercion, often in self-defense and/or in the defense of others. As such, the fairest assessment of the legality (and administrative permissibility) of the application of force must be made from the perspective of someone at the incident who would have had to make a force decision in the heat of the moment – just as the actual officer did.

Deadly force – i.e., "force that kills or is likely to kill" (Fyfe 1979) – and which is usually, but not always, the result of a firearm discharge – is governed largely by the same objective reasonableness standard as nonlethal force (the standard appears to change to a test of "deliberate indifference," as governed largely by the Eighth Amendment, if deadly force is used on a suspect while in a custody facility), though with an extra restriction. In the case *Tennessee v. Garner* (1985), the Supreme Court abolished the common law custom that had historically allowed police officers to use deadly force against fleeing suspects who posed no threat to life. In abolishing this practice – which was known as the "fleeing felon rule" – the Court reasoned that using deadly force to apprehend a person suspected of having committed a felony, but who seemed to pose no imminent threat to life – constituted an unreasonable seizure of the person under the Fourth Amendment. Despite this landmark decision that made the "defense of life" standard the law of the land, the Court in *Garner* failed to outline circumstances under which deadly force *was* permissible. Thus, *Garner* allows officers to determine the meaning of "life" threatening: The risk to life need not be *real* at the time of the incident; it must be *apparent*.

Among all the dimensions of the use of force construct, "unnecessary" force is perhaps the most difficult to assess and the most controversial to condemn. Unnecessary force usually represents the proper use of force at the immediate moment of application, but which could have been avoided had officers not allowed, or caused, the incident to escalate (Fyfe 1986). Thus, rather than representing the product of malice or the inclination to "punish," unnecessary force generally results from police officer's professional incompetence. A typical example of this occurs when police respond to a man with a gun call. They might arrive on the scene to find an agitated man pacing in his front yard with a pistol in hand. Because officers neglected to find cover and concealment (i.e., failing to position themselves behind parked cars or out of direct line of fire), they shoot the man when he makes a furtive movement. When focusing on the last point in the encounter – i.e., the instance immediately preceding the shooting – the force used may be justifiable since the police were compelled to protect their lives and perhaps the lives of bystanders. When considering, however, the events that led to the shooting, professional incompetence left the officers at risk of being shot in the first place. Had they found proper cover, they might have been able to wait and persuade the man to drop the gun, which would have averted the use of violence. Although the examination of unnecessary force is largely beyond the scope of this entry, it is important to at least identify because (1) it represents a component of the overall use of force construct, and (2) likely happens more frequently than does excessive force, often without the subsequent application of accountability remedies (Fyfe 1986).

As the above discussion should suggest, police brutality is physical force that breaches the confines of legally permissible physical coercion. Because police brutality is often committed out of malice, disrespect, and/or officers' desires to reassert their authority over a suspect (e.g., Skolnick and Fyfe 1993), several academic researchers often regard excessive force as an example of police abuse of authority (e.g., Carter 1985) rather than a mere "excess" of force. This is because police brutality often stems from the same theoretical processes that lead to verbal chastising and psychological abuse during police interrogations (e.g., Carter 1985) – both of which are abusive but not physical.

Although most state criminal codes in the United States include statutes that specifically make police brutality under "color" of state authority illegal, the most common legal remedy for parties attempting to redress the alleged excessive use of force by police is Title 42, Section 1983 of the US Code. This is the Federal civil rights statute that allows persons alleging harm by police (and any representative of the State who may have acted under color of authority) to seek damages from both the individual officer(s) and the government agency that employed them. The reason why Section 1983 is likely the most common legal mechanism invoked in excessive use of force claims is because the statute relaxes the usual requirement placed on plaintiffs that they exhaust all lower administrative and legal outlets before filing a claim in Federal court. Although the statute does not require proof that a police officer intended to harm the victim (that is, the plaintiff need not prove an officer's psychological state of mind), it does require the plaintiff to successfully demonstrate that the government entity (e.g., municipal police department) that employed the officer maintained official policies and directives that directly resulted in the harm caused, or that the entity had established informal customs and practices that led to the harm. For this reason, Section 1983 lawsuits are often colloquially referred to as "patterns and practices" suits and generally allege that the brutality (or other problematic behaviors) occurred because the police agency failed to train and/or supervise the officers properly and/or negligently retained the officers as employees.

Thus, integrating the Section 1983 statute with Supreme Court decisions results in plaintiffs gaining ready access to Federal court for the purposes of attempting to hold police officers and the agencies/governments that employed them accountable for brutality, though trial courts must adhere to *Graham's* "objective reasonableness" requirement of evaluating the use of force from the standpoint of a reasonable officer at the scene. The net effect has been that, while it is relatively easy to sue police officers and police departments, it is highly difficult to win such cases due to the "reasonable officer at the scene" requirement (Skolnick and Fyfe 1993). Moreover, as research has shown, even when plaintiffs are generally successful in proving police brutality at an acceptable legal threshold, juries historically have been reluctant to decide against police officers due to generalized public support for the police, and the fact than many plaintiffs who allege brutality occupy dubious social positions relative to the officers they are accusing. Perhaps the best example in modern history was the Rodney King case in which all officers accused of brutality were found not guilty in the first trial.

Classic Templates to Explain Excessive Use of Force

As noted in the opening pages of this entry, police brutality – or excessive use of force – occurs for myriad reasons across multiple levels of analysis: the psychological, sociological, social contextual, and organizational. Since William Westley (1970) published his classic ethnography of police in Indiana (which was actually published 20 years after his dissertation work that produced the findings) in which he described for the first time the importance officers and police organizations placed on making felony arrests, several researchers have replicated Westley's methodology in an effort to highlight several aspects of the police occupation that previously had been generally unknown. The result was the publication of classic statements on police coercion that described the processes by which police officers may commit, and why police departments sometimes attempt to hide, police brutality. This entry considers these perspectives as occupational templates and describes the most important of these below.

Starting at the individual-level, in his largely ethnographic study of the police in "Laconia" (a pseudonym), Muir (1977) interviewed and then observed a group of police officers in an effort to understand how they responded to "critical" incidents in their work settings. It is now commonly known that "Laconia" was Oakland, CA, a large city that employed 800 police officers at the time of Muir's research, and which was characterized by high levels of racially concentrated economic resource deprivation and crime. During his research, Muir observed four police officer personality "types" that were differentiated by two elements: the abilities to (1) empathize with those whom they encountered on the street, and (2) morally reconcile need to use force in certain circumstances. For example, the archetypal officer in Muir's typology was the "professional" – i.e., a collective group of officers who could place themselves psychologically in the life circumstances of the suspects they encountered, and although professionals did not care to use force to resolve most conflicts during their engagements with members of the public, they used force when necessary to stop conflict, protect lives, and restore order. Once the situational exigencies were neutralized, professionals ceased their use of physical force. A key attribute of professionals was their ability to communicate effectively with suspects in ways that often minimized, de-escalated, or even eliminated the need to use force during many incidents.

In contrast to the professionals were the "enforcers:" Officers who had the highest propensity to use excessive force against suspects because they (1) generally lacked the capacity to fathom the often "tragic" conditions that characterized the lives of many people they encountered, and (2) were quick to resort to violence when they perceived minimal threats to their safety or when they felt compelled to assert police authority. Muir (1977) noted that, unlike the professionals, enforcers usually lacked the communication skills necessary to de-escalate and/or otherwise avoid the need to use force in certain circumstances. As a result, enforcers often created situations where other officers near them were forced to use violence against suspects because of the degree to which enforcers provoked violence during their contacts with the public. Thus, and as Muir notes, enforcers were problematic to police workgroups partly because they frequently engaged in excessive force, and partly because they often caused *potentially* violent encounters to escalate into *actual* violent encounters.

Perhaps the most interesting and important element of Muir's (1977) typology was the recognition that officers might slide between one personality type and another based on cues in the social environment. For example, officers regarded as professionals – and who usually conducted themselves appropriately during most police-citizen encounters – sometimes became enforcers during certain incidents. Muir, for example, identified one "professional" officer among those he studied who was highly communicative and restrained in virtually all street encounters except when he responded to sexual or domestic assaults against female victims. These events tended to trigger an uncharacteristic response in the officer – perhaps due to family history or other personal experience – leading him to take on the attributes of the enforcer: He was quick to use violence as an initial response when contacting an alleged domestic or sexual assault offender. Thus, although Muir's typology was largely psychological, it recognized the importance of social context when analyzing street-level behavior of officers.

Although no identified studies have found reliable empirical support for Muir's typology (see, for example, Hochstedler (1981)), none has tested the typology in a setting that replicates that in which Muir conducted his research. It may be that Muir's findings hold primarily for postindustrial port cities with large urban African-American populations that live predominantly in poverty conditions, socially isolated from the larger urban matrix. Such are the conditions that most often allow officers to engage in abuse of authority – including excessive force – with little fear of recourse (e.g., Kane 2002; Kubrin and Weitzer 2003). Moreover, despite scant support for the typology itself, at least one study has found that – like Muir's findings – officers who exhibit certain personality characteristics (diffuse though they may be) often "slide" between profiles as the social contexts in which they work vary (Hochstedler 1981). This finding supports Muir's observation that even so-called professional police officers may become "enforcers" under the circumstances that elicit particularly strong responses, and perhaps where local social constraints are wanting. Such findings demonstrate the importance of the social ecological contexts in which officers work as determinants of their behaviors. At the very least, Muir's typology provides a useful taxonomy for efficiently describing certain officer characteristics that are easily recognized – i.e., enforcer.

Residing one level up from Muir (1977) on the social ecological progression is Van Maanen's (1978) "asshole" perspective. This framework is so-named for the colloquial usage of the term by police officers as a shorthand for suspects who – through speech or gesture – "deny" a police officer's "definition" of the power structure that characterizes their interactions with the public – namely, that officers are in charge. Van Maanen argued that the line officers he studied tended to classify members of the public into three groups: "suspicious persons" (suspected criminal offenders), "assholes" (people who challenge police authority), and "know nothings" (people who do not fit the previous two categories and who are also not police officers). Whereas police officers usually have little or no cause to use coercion against know nothings, and they generally use coercion only as necessary to control and/or gain the compliance of suspicious persons, they may use coercion – i.e., force – against those labeled assholes as a retaliatory mechanism to redress an affront to an officer's authority (Van Maanen 1978).

Van Maanen (1978) observed that a suspect's initial response to a police officer's inquiry determined the officer's categorization of him (Van Maanen never clarifies as to whether the template applies to female suspects), as well as the subsequent police reaction. Suspects who appeared to reject the officer's definition of the situation – thereby challenging the officer's authority and legitimacy – were labeled assholes and were often subjected to a violent response as a form of punishment or reassertion of authority. From an officer workgroup standpoint, police officers who labeled a suspect an asshole were justified, and perhaps even expected, to use violence to reestablish control.

Recall Incident #2 – the dramatized use of force scenario – in the "Excessive Force Through Legal and Social Lenses" section of this entry. The officer's response to the youthful driver's verbal resistance was a quintessential example of how a person may become labeled an asshole under Van Maanen's model. Importantly – and as shown in Incident #2 – when officers label a suspect an asshole, and to the extent that the officers use brutality to redress the affront, they often do so to impose police authority typically at the expense of formal sanctioning (e.g., arrest). As such, some suspects – particularly those engaged in illegal behavior when contacted by the officer – may accept the brutality in lieu of a formal sanction because if they reported the brutality, they would also have to report their activities that led to their initial contact with the police.

Though it is the case that occupational templates-perspectives often fail to yield empirical support, at least one recent study has offered some empirical validation of Van Maanen's (1978) "asshole" thesis. In a study of how the occupational goals of police may sometimes conflict with rule of law practices, Kane and Cronin (2011) examined the extent to which police officers used force in situations that may have been characterized by the "asshole" process. Through the use of discriminant function analysis, Kane and Cronin reported that during coercive exchanges with suspects (i.e., situations that ultimately ended in arrest), suspects who demonstrated "verbal antagonism" toward officers – but not physical resistance – were likely to be "punched" and "kicked" during the incident. Interestingly, when suspects demonstrated physical resistance, but not necessarily verbal antagonism, they were usually subjected to more "normal" types of force, such as chemical sprays, baton strikes, or basic joint locks (Kane and Cronin 2011). Verbal antagonism appeared to elicit "asshole" responses in the form of punches and kicks – use of force techniques that usually do not exist on a police use of force continuum.

Whereas Van Maanen (1978) operationalized the "asshole" perspective as one rooted primarily in the social interactionist tradition, there is evidence that the asshole process also may be driven in part by police subcultural pressures. In his observational study of how the Los Angeles Police Department controls space within its vast jurisdictional boundaries, Herbert (1998, p. 347) adds a conceptual dimension to Van Maanen (1978) with his identification of normative orders: a collective subcultural belief system "…oriented around a common value." Herbert argues that police officers working in small groups develop informal work and productivity standards within the context of six normative orders that include law, bureaucratic control, adventure/machismo, safety, competence, and morality. Herbert (1998, p. 347) observes that officer workgroups place differential values on the normative orders in ways that "provide guidelines and justifications for" workgroup enforcement activities. For present purposes, Herbert notes that normative orders – such as law and safety or law and competence – may conflict with one another as officers use their coercive authority to satisfy one normative order (e.g., competence), perhaps at the expense of another (e.g., law or bureaucratic control).

Herbert's normative order system, which he referred to a police subculture, may be integrate with occupational templates because it explains how situational circumstances may trigger police responses designed to satisfy occupational goals. For example, officers confronted by a suspect they label as "dangerous" will likely invoke the normative order of *safety* to minimize the threat of violence. Similarly, officers who enter into a coercive encounter with a bellicose suspect – an "asshole" in Van Maanan's (1978) taxonomy – may seek to demonstrate their *competence* to other officers in their workgroup by using excessive force or brutality, which may be regarded as situationally justified to reassert police authority over a noncompliant suspect. Again, such an application of coercion may conflict with rule of law orders, but may be regarded as occupationally necessary if officers achieve the objectives of their primary normative value.

Finally, to understand how police agencies themselves may create a culture that fosters, if not encourages, the excessive use of force by their officers, it is important to consider the research on police organizations. In Los Angeles, the Christopher Commission assigned to investigate patterns of excessive violence in the LAPD in the wake of the 1991 beating of motorist Rodney King identified a "siege mentality" that was prevalent among officers of that department (Independent Commission 1991, p. 95). This siege mentality, the Commission argued, led officers to view their relationship with the community at large as adversarial (the "us vs. them" perspective) and appeared to legitimize the excessive use of force by police against suspects in an environment that all but guaranteed immunity from accountability (Skolnick and Fyfe 1993). The confidence that officers placed in their ability to freely "punish" suspects who committed affronts against police authority was perhaps best illustrated by the routine exchanges among officers via their mobile data terminals in which anecdotes describing violent activities – often in racist terminology – were openly exchanged (see: Independent Commission (1991)).

Appealing to occupational templates to help explain the processes by which excessive force may be used and legitimized by police officers is instructive because they offer a social ecological context within which to interpret police brutality that is often absent from the empirical literature that examines excessive force. Templates, such as Muir's (1977) extortionate transaction model, Van Maanen's (1978) "asshole" perspective, Herbert's (1997) identification of workgroup "normative orders," and Skolnick and Fyfe's (1993) elaboration of the "siege" mentality, show that police brutality may be influenced at every level of a police officer's contextual occupational life. Moreover, they offer insight into how some officers who may be disinclined to use brutality on suspects may operate in layered contexts that favor, or even expect, them to use excessive against certain suspects under certain circumstances.

Some Consequences of Police Brutality

Perhaps the most obvious consequences of police brutality are the direct effects endured by victims. Although there are no national data on the subject, it seems clear that people who experience police brutality can suffer physical harms ranging from minor aches and pains to serious injury, and in some cases, death. It is also likely that people who experience physical abuse at the hands of police may suffer emotional trauma. In the United States, however, police agencies are not required to report use of force – let alone excessive force – incidents to any centralized authority, leading researchers and policy-makers to make guesses about the frequency and severity of police brutality. This problem could be ameliorated by the implementation of a national reporting system on the police use of force.

In addition to the direct effects of police brutality to those who experience such abuse, there are often broader sociological consequences that may result when police engage in the excessive use of force, particularly when the force is systemic in nature and occurs in communities whose residents perceive that they have been socially and legally marginalized by their municipal governments. Among these, perhaps the gravest consequence is the loss of legitimacy police forces can experience when officers are viewed as brutal or abusive, and when community residents perceive little or no recourse to control them.

Tyler (1990) has argued that societies are more likely than not to comply with the law when they believe in the morality of the law and in the legitimacy of the government making the law. For a society to accept law as moral and the government as legitimate, its members must believe in the fairness of the law-making process (Tyler 1990). This model of procedural justice generally argues that the way a state achieves and maintains legitimacy is by distributing its authority through consensus-based processes without regard for the social and economic statuses of societal subgroups. If the public becomes polarized over issues related to the distribution of legal authority, the government may lose legitimacy and find it difficult to fulfill its “regulatory role” in society (Sunshine and Tyler 2003, p. 515) without having to resort to the frequent use of physical coercion. This issue is important for municipal authorities, as some members of the public, particularly those residing in communities characterized by racial segregation and/or systemic economic resource deprivation, may perceive inequities in the distribution of police resources in violation of procedural justice expectations (Jacob 1971).

Kubrin and Weitzer (2003) were among the first researchers to find social scientific evidence supporting the above arguments. In their study of retaliatory homicide in St. Louis, Kubrin and Weitzer found that certain types of homicides predicted subsequent homicides in the form of retaliation, suggesting that, had the initial homicide been stopped, the net reduction in many cases would have been two, given that the retaliatory homicide also would have been negated. In the qualitative portion of their study, Kubrin and Weitzer found through unstructured interviews that the initial homicides were allowed to occur in large part because community residents refused to share information related to criminal activity with the police due to their distrust of law enforcement authorities. This distrust stemmed from perceived abuses – including brutality – of community residents by police over a number years. Kubrin and Weitzer’s (2003) findings were largely consistent with Anderson’s (1999) observations of the so-called Code of the Street. Anderson argued that the Code – in which respect was valued as a social good, earned and maintained by the threat and/or use of violence – prevailed in the north Philadelphia community he studied in part because the police had virtually no legitimacy due to their perceived abuses of authority.

Kane (2005) made similar findings in New York City in his examination of the consequences of “over-policing” and police misconduct (including violence) on violent crime. For present purposes, Kane found that in economically stable communities, police misconduct had no effect on violent crime. In economically disadvantaged, and “extremely” economically disadvantaged, communities, police misconduct (including violence) led to subsequent increases in violent crime. Kane attributed this relationship to the same processes driving Kurbin and Weitzer’s (2003) findings: As residents of socially/economically marginalized communities perceived increases in police abuses, they likely ceased sharing crime-related intelligence with authorities, allowing violent crimes to occur that may have been otherwise prevented.

In his classic lecture, "Politics as a Vocation," Max Weber (see: Weber et al. 2004, pp. 78–79) argued that because the State "claims the monopoly on the legitimate use of force" to enforce its laws, politicians who hope to gain the "obedience" of those whom they govern need not simply *do* the right thing; politicians must *appear* to do the right thing. Weber's arguments are made salient by Kubrin and Weitzer (2003) and Kane (2005): Failure of the police to "govern" effectively the territories they serve may lead to noncompliance on the part of people who reside in communities characterized by the perceptions of police abuse of authority. In the "everyday" sense, citizen's noncompliance may manifest in the form of people failing to readily yield to police authority during traffic stops and stop-and-frisk events – both of which may create a context of hostility and distrust on the part of the police and members of the public. In the broader and more extreme sense, generations of police abuse, including unchecked or redressed brutality, may lead to urban violence. Indeed, two of the most destructive urban riots in American history have been attributed to long-standing patterns of police abuses, including unchecked brutality.

On August 11, 1965, a white California Highway Patrol officer stopped an African-American motorist in the Watts section of Los Angeles for suspected Driving While Intoxicated. Even at that time, Watts was an area of Los Angeles, characterized by racially concentrated economic resource deprivation, which had a difficult history with the police. The motorist – Marquette Frye – resisted being taken into custody, which led to an escalating confrontation with police (Barnhill 2010). As more officers arrived on scene, the crowd of bystanders also grew until someone finally began throwing bottles and other objects at the officers, who continued their struggle to subdue Frye (Barnhill 2010). In the process of taking Frye into custody, officers also arrested his mother and brother for unlawfully interfering with the arrest, which led the crowd to switch from throwing bottles to throwing rocks and other large items at police (Barnhill 2010). This inciting incident led to 6 days of rioting in Watts, leading to 34 deaths, 1,032 injuries, and over 3,000 arrests (Barnhill 2010). Although post-riot analyses generally concluded that poor living conditions and the social marginalization of African-Americans in Los Angles represented the root causes of the Watts riot (e.g., California Governor's Commission on the Los Angeles Riots 1965), some researchers concluded that years of police abuse and brutality in Los Angeles – particularly during the tenure of Police Chief William H. Parker, during which the LAPD transformed into highly aggressive paramilitary police force – was largely responsible for the riot (Skolnick and Fyfe). Indeed, the McCone Commission, empanelled just a few months after the riots and charged with identifying the causes, noted from its many interviews of African-American Watts residents:

> "Police brutality" has been the recurring charge. One witness after another has recounted instances in which, in their opinion, the police have used excessive force or have been disrespectful and abusive in their language or manner (Governor's Commission 1965, p. 40)

The social and economic costs of the 1965 Los Angeles riot were eclipsed only by those of the 1992 Los Angeles riot, which occurred in the immediate aftermath of two white police officers being acquitted for beating motorist Rodney King. During the 1992 riot, the LAPD and the California National Guard killed 53 people. More than 2,000 people were injured, and over 1,000 buildings were destroyed by fire (Wood 2002). The fact that both riots occurred in Los Angeles was no surprise to Jerome Skolnick and James J. Fyfe – two of the nation's leading scholars of police authority and accountability – who argued that since William Parker was appointment police chief of Los Angeles in 1950, the LAPD became increasingly insular, aggressive, and unaccountable to the mayor and city council (Skolnick and Fyfe 1993). The unaccountability was particularly pronounced under Chief Daryl Gates, who enjoyed civil service protection and indemnity from lawsuits (Skolnick and Fyfe 1993).

It seems clear that, while people subjected to direct forms of police brutality suffer the most immediate harm from such abuse, those residing in communities – even cities – in which the police engage in the excessive use of force with little accountability also experience the resulting harms. Violent incidents in Los Angeles (1965 and 1992), Detroit (1967), Chicago (1968), Liberty City-Miami (1980), Cincinnati (2001) – even Tulsa in 1921 – were not so much riots as they were urban rebellions. In each event, the triggering incident was related to some form of police abuse, and all post-riot analyses identified abusive police practices, largely in communities characterized by racially concentrated economic resource deprivation and social marginalization, as major causal factors (see: Independent Commission 1991; Kerner Commission 1968; Skolnick and Fyfe 1993).

Preventing Excessive Use of Force

As previously alluded to, people who experience police brutality or the excessive use of force have recourse through the conventional mechanisms of police accountability: the citizen complaint process, mediation, and the courts. Each of these avenues of recourse has varying chances of success, depending on the several factors related to the event, including the social desirability of the accuser, the degree to which the incident occurred in the public realm and/or in view of credible witnesses, the extent to which the accuser may have attempted to initially evade or who otherwise resisted police officers, and in many cases the quality of the accuser's legal counsel. While individual plaintiffs may prevail in their attempts to hold police accountable for the use of excessive force, most would probably argue that best recourse would be the prevention of police brutality before it occurs. The most immediate methods by which to control and/or prevent the excessive use of force by police are probably rooted in the recruitment of police officers, training and supervision, and the establishment of organizational cultures within police departments that value the protection of life over the enforcement of law as a primary function.

The personnel processes police departments use to recruit and hire officers play a crucial role in shaping subsequent organizational behavior, but in ways that are perhaps more subtle than they are apparent. Police department hiring practices are generally designed to screen out undesirable applicants, which is substantively different from trying to screen *in* desirable applicants. These so-called screening-out practices are rarely evidence-based; and they are often not followed consistently over time or across police recruiters or background officers. Moreover, to the extent that the excessive use of force is a relatively rare event within police organizations, it is difficult to accurately predict who might have become "problem" officers over the course of their careers were they allowed to join the force. Although emerging evidence supports the practices of screening out applicants on the basis of employment and criminal histories to prevent general misconduct (see: Kane and White 2012), there is virtually no evidence that "violence-prone" (Toch 1996) police officers can be identified and effectively screened out through the normal recruiting process.

Screening methods that favor recruits whose education, experience, cultural identities, and world views would create police department cultures not dominated by a single gender or ethnic group, and it would raise the possibility that members of all subgroups would contribute to the development of a sensitive workforce tolerant of people across race and class. Empirical findings support this claim: In their study of career-ending misconduct in the New York City Police Department, Kane and White (2012) found that from 1975 to 1998, as the NYPD became increasingly racially and ethnically diverse, it also became better behaved. Rates of misconduct decreased as diversity increased, independent of external factors that often influence organizational rates of misconduct. Although the NYPD over that time period may not have systematically recruited persons with diverse characteristics or personal histories, the results of bringing diverse pools of recruits into policing are clear: Organizational misconduct declines.

In addition to recruitment, supervision of police officers is also critical in preventing excessive use of force – particularly in communities that historically have had difficult access to the conventional mechanisms of police accountability. For example, in a study of police misconduct in New York City police precincts, Kane (2002) found that the traditional antecedents of social disorganization – particularly, structural disadvantage – predicted increased rates of career-ending police misconduct. Precincts characterized by the highest rates of racially concentrated structural disadvantage also had the highest rates of police misconduct. This finding led Kane (2002, p. 891) to conclude that, "The very communities likely in need of the most protection by the police. . .(are also) in need of the greatest protection from the police. . ." One way to offer that protection is through the supervision of officers.

In many American police departments, the average staffing ratios call for one sergeant for every nine patrol officers (Walker and Katz 2010), and while scant research highlights what supervisors actually *do*, at least one study suggests that low spans-of-control (i.e., high ratio of supervisors to officers) decreased excessive force (Terrill 2001). Although modern organizational police theory generally argues that low spans-of-control are inefficient to the administration of police departments (making the assumption that "flatter" organizations function more efficiently than "tall" organizations) (e.g., Walker and Katz), in communities where the risk of misconduct may be higher than average, it seems that lower spans-of-control would reduce the risk of brutality as supervisors have more contact with line officers in the field. Indeed, under such conditions, what may be sacrificed in terms of organizational efficiency may be gained in effectiveness.

Finally, recruitment and supervision to limit police and the excessive use of force seem to have their highest probabilities of success if the overall police department culture supports the structural efforts in place to guard against brutality. The tone must be set at the top through the implementation of progressive policies that strive for transparency in the disciplinary review process and promote model tactics and strategies. In addition, evidence suggests that when line officers are involved in the development of use of force policies and field tactics designed to reduce police officer violence (i.e., Fyfe 1997), then the use of force within departments tends to decrease, partly because police officers view themselves as contributors to the policy development process, rather than simply as people who need to be controlled.

Once the chief administrator implements an infrastructure of progressive policy and gains the "buy-in" of line officers, the message of progressive policing must filter through all layers of the police organization. One method by which to accomplish the process of communicating the department's value system is via the field training officers (FTOs). Toch (1996) argued that for the purposes of controlling police violence, field training officers were instrumental in bridging the gap for new officers between the messages they received in the training academy and the messages they received from their peers once on the street. Toch argued that FTOs serve a crucial organizational role because, among other functions, they help probationary police officers practice their recently acquired skills in live settings while under highly structured supervision. During this formative period in a police officer's career, FTOs can help new officers interpret street encounters in ways that allow them to learn and practice good policing through the metaphorical lenses of a protection of life mandate. Once they complete their field training programs, police officers may continue to practice good policing and the protection of life in both patrol and specialized settings.

Successfully preventing the use of excessive force by police likely requires a multilayered approach that considers the social ecology of the total police organization. While the chief administrator is an important part of the process within the organization, it must be noted that the chief also represents the will of the community. Chief administrators for police departments are usually chosen by the elected bodies of local governments (e.g., city councils, county boards of supervisors, etc.), often for their stated positions on certain issues, such as violent crime, fear of crime, urban disorder, etc. Excessive use of force will be minimized only if governing boards select chief administrators who are actively interested in controlling and preventing excessive use of force. Moreover, the governing boards themselves need to share the same commitment to police accountability if excessive force is to be controlled and prevented in a police organization.

Conclusions

Evaluating the use of excessive force by the police tends to be much more difficult than other forms of misconduct, such as profit-motivated corruption. Unlike other forms of police deviance, which tend to be wrong by their very occurrence, the roots of police brutality reside in the legitimate use of force – a tool society gives police to accomplish its mandate. Therefore, when an officer is accused of brutality, it is often difficult, if not impossible, to determine where the justified use of force may have become excessive. Even in the infamous, and unquestioningly brutal, Rodney King case, the police officers who contacted King had cause to use some physical coercion in their initial attempts to bring him into compliance. This entry identified classic occupational templates of the police use of force in an effort to identify context within which the excessive use of force may be considered. Although these templates have received little empirical validation, they are meaningful because in many cases, they offer a shorthand for describing the conditions under which police officers may resort to brutality. Although police departments can use recruitment as a tool to prevent the excessive use of force by their officers, the best prevention is likely based on a multi-level approach that involves local governing bodies selecting chief administrators with strong commitments to transparency and the virtues of "good" policing.

Related Entries

- ▶ Biased Policing
- ▶ Control of Police Misconduct
- ▶ Crime and the Racial Composition of Communities
- ▶ Democratic Policing
- ▶ Drug Enforcement
- ▶ Hot Spots and Place-Based Policing
- ▶ Law of Police Use of Force
- ▶ Minorities Within the Police Workforce
- ▶ Police and the Military Nexus
- ▶ Police Corruption
- ▶ Police Culture
- ▶ Police Discretion and Its Control
- ▶ Police Integrity
- ▶ Police Leadership Styles
- ▶ Police Legitimacy and Police Encounters
- ▶ Police Line-Level and Agency Accountability
- ▶ Police Selection
- ▶ Police Use of Firearms
- ▶ Risk Management in Policing
- ▶ Sex Offenders and Criminal Policy

Recommended Reading and References

Anderson E (1999) Code of the street: decency, violence, and the moral life of the inner city. New York: W.W. Norton & Company

42 U.S.C. § 1983

Barnhill JH (2010) Watts riots (1965). In: Danver SL (ed) Revolts, protests, demonstrations, and rebellions in American history, vol 3. Santa Barbara, ABC-CLIO

Carter D (1985) Police brutality: a model for definition, perspective, and control. In: Neiderhoffer A, Blumberg A (eds) The ambivalent force. Holt, Rinehart and Winston, New York

Fyfe J (1979) Administrative interventions on police shooting discretion: an empirical examination. J Crim Justice 7:309–323

Fyfe J (1986) The split-second syndrome and other determinants of police violence. In: Anne T. Campbell, John J. Gibbs (eds) Violent Transactions. Oxford: Basil Blackwell, pp 207–225

Fyfe J (1997) Good policing. In: Dunham R, Alpert GP (eds) Critical issues in policing: contemporary readings, 5th edn. Waveland, Prospect Heights, pp 194–213

California Governor's Commission on the Los Angeles Riots (1965) Violence in the city: an end or a beginning?

Graham v. Connor, 490 U.S. at 388 (1989)

Herbert S (1997) Policing space: territoriality and the Los Angeles Police Department. University of Minnesota Press, Minneapolis

Herbert S (1998) Police subculture revisited. Criminology 36:343–369

Hochstedler E (1981) Testing types: a review and test of police types. J Crim Justice 19:451–466

Independent Commission (1991) Report of the Independent Commission on the Los Angeles Police Department. The Commission, Los Angeles

Jacob H (1971) Black and white perceptions of justice in the city. Law Soc Rev 6:69–89

Kane R (2002) The social ecology of police misconduct. Criminology 40:867–896

Kane R (2005) Linking compromised police legitimacy to violent crime in structurally disadvantaged communities. Criminology 43:469–498

Kane R, Cronin S (2011) Maintaining order under the rule of law: occupational templates and the police use of force. J Crime Justice 34:163–177

Kane R, White M (2012) Jammed up: bad cops, police misconduct, and the New York City Police Department. New York University Press, New York

Kerner Commission (1968) Report of the National Advisory Commission on civil disorders. Bantam Books, New York

Klockars C (1985) The idea of police. Sage, Beverly Hills

Kubrin C, Weitzer R (2003) Retaliatory homicide: concentrated disadvantage and neighborhood culture. Soc Probl 50:157–180

Muir WK (1977) Police: streetcorner politicians. University of Chicago Press, Chicago

Skolnick J, Fyfe JJ (1993) Above the law: police and the excessive use of force. Free Press, New York

Sunshine J, Tyler T (2003) The role of procedural justice and legitimacy in shaping public support for policing. Law Soc Rev 37:513–547

Tennessee v. Garner, 471 U.S. 1 (1985)

Terrill W (2001) Police coercion. LBF Scholarly Publishing, New York

Toch H (1996) The violence-prone police officer. In: Geller W, Toch H (eds) Police violence: understanding and controlling police abuse of force. Yale University Press, New Haven, pp 94–112

Tyler T (1990) Why people obey the law. Yale University Press, New Haven

Van Maanen J (1978) The asshole. In: Manning PK, Van Maanen J (eds) Policing: a view from the street. Goodyear, Santa Monica, pp 221–238

Walker S (2010) Sense and nonsense about crime, drugs, and communities, 7th edn. Wadsworth, Belmont

Walker S, Katz C (2010) Police in America: an Introduction. McGraw-Hill, New York

Weber M, Owen D, Strong TB (2004) The vocation lectures: "Science as a Vocation" and "Politics as a Vocation," (trans: Livingstone R). Indianapolis: Hackett Publishing Company

Westley W (1970) Violence and the police. MIT Press, Cambridge

Wood DB (2002) L.A.'s darkest days. Christian science monitor. Retrieved 2 Nov 2012

Police and the Military Nexus

Otwin Marenin
Department of Criminal Justice and
Criminology, Washington State University,
Pullman, WA, USA

Synonyms

Constabularization; Militarization

Overview

The police and military (armed forces), the two coercive agencies of the state, are restricted by law, tradition, and policy to two domains in which they exercise authority. The police protect the domestic order against risk, threats, and crimes by persuasive communications, the enforcement of laws, and, ultimately, force (or its potential use), while the military defends the nation and the state against aggression from without. These analytical and policy distinctions between two agencies (or systems of agencies) which are authorized to employ legitimate internal and external use of force to deal with threats have become blurred in practice (Easton et al. 2010). The argument and worry that the police in the USA and other democratic countries are increasingly becoming militarized, hence will loose the civil and democratic orientation to their work, the use of force, and the collections of intelligence, has been raised by a number of scholars who study the police and by the police themselves (Kraska 2001, 2007; Kraska and Kappeller 1997).

Much of the concern that the blurring of distinctions between the policing and the military work will undermine civil and democratic policing rests on two limited understandings about the police, namely, that policing has never been militarized to any degree and by comparing two "ideal types" (yet legitimate conceptions) of both the police and the military rather than their historical configurations, relationships, and work.

The blurring of policing and military work and tasks is not new; it only has increased and taken many new forms in recent times, specifically by recent domestic and international developments (transnational crime, fears of terrorism, civil violence in failed states, peacekeeping and peacebuilding interventions, the weakening of border controls, technological advances in information acquisition, processing, and distribution). The distinctions drawn between military and police forces have always been blurred in practice. Police and military have worked in the same occupational space – the protection of society against threats; they have been bestowed by their societies with the right to use force, and the core values enshrined in their occupational cultures (internal authority, discipline, closed ranks, loyalty, courage, masculinity) are similar (den Boer et al. 2010, 225–226).

Ideal Typing the Police and the Military

Ideal Typing Democratic, Civil Policing

The classic image of the democratic police officer is the patrol officer, the constable walking the beat, known by and knowing the community and the village; approachable, protective, and friendly when encountering citizens in need; committed to service, the protection of human rights, and the rule of law; doing her/his work with integrity; and ensuring the effective protection of public order and the security of property and lives of the populace against threats and crimes. Democratic, civil police retain substantial domains of discretion in how to respond and to adjust their decisions to situational exigencies, guided by law, professional norms, public demands, and democratic political oversight (OSCE 2006). Since full service tasks are most common work done by police in democratic countries, this image is not a false one.

The essence of democratic policing is captured in six aspects of their work: authority to control and suppress illegal conduct is limited by and large to the domestic arena; the composition of police personnel should approximate the distribution of salient identity groups in society; orientation to public service; the constrained use of force; transparency in actions; and accountability to civic society by various mechanisms.

The occupational culture of democratic police officers and the organizational policies which shape their work and guide their discretion stress that policing is a service to the public and not to the state. Their use of force should be strictly limited by law, professional norms, and organizational directives and used only as a last resort when dealing with all people they encounter. The police have a responsibility to explain when asked, and the public has a right to know what the police are doing (except for legitimate secrecy in pending cases and investigative aspects of their work). The police cannot hide behind the shield of expertise and secrecy related to their general work, nor deny the media, the public, and the political leaders' access to the information in their possession, unless legitimately protected. Lastly, if allegations of misconduct, abuse of power, or corruption are made against them, the police have an obligation to submit to external oversight, and the public has the right to require their submission to civic reviews. They do not have the exclusive authority and power to investigate allegations of misconduct in their ranks by themselves and take corrective actions. As government officials having great power over the routines of living in society, they cannot claim impunity for their conduct.

Ideal Typing the Professional Soldier

The military are an armed, professionally trained, and hierarchically controlled organization in which lower ranks are expected to respond quickly and without questioning to commands from the top. The military protects the nation, its people and territory, and the "national interest" by using its ability to detect and suppress external threats. It uses its capacity to exercise force to capture, disable, or kill enemies of the nation in an efficient, fast, and effective manner. The military collects information in secret via sophisticated and complex technologies and human spies and evaluates and distributes information as "intelligence" which is typically kept secret by national security justifications and tactical contingencies. Transparency is severely limited by the need to keep the country safe from aggression which, were the nature of the intelligence and the means by which it was collected be known, would jeopardize the ability to plan and execute the collection of information. Accountability of the military for its strategic and tactical conduct is first to itself, with civilian oversight by elected leaders the last overriding democratic option. By law and tradition, the military in democratic societies only intervene in domestic security when national and local policing systems are unable to maintain control.

Distinctions

In short, ideal-type descriptions of police and military emphasize fundamental distinctions between the two security agencies authorized to use force in order to protect the nation-state, its government, its economic and political ways of life, and its people and properties against domestic and external threats.

Specifically, the police see the people they have to deal with primarily as customers, innocents, and suspects, while the military sees real and potential enemies. The police are committed to maintaining public order, provide services, and enforce domestic laws while the military seeks the defeat of enemies and victory. The police use force only as the last resort in encounters, while the military resorts to force immediately to disable, capture, or kill enemies. The police tend to work as individuals (or limited teams) while teamwork is essential to the military. The police at the lowest ranks of the organizational hierarchy have substantial discretion in specific actions and encounters, a discretion that is circumscribed by situational exigencies and the best judgment of the officers on what is the appropriate, effective, and balanced action in the situation they are faced with. The military limits discretion at the lowest ranks through intensive training, a commitment to authoritative interpretations of what needs to be done in a situation ("that's an order, soldier"), and loyalty to group. The discretion of the police is influenced by personal morality and ideological beliefs, while the military seeks to socialize new recruits and experienced soldiers to an ideology that eliminates personal views as guides to actions.

The police are accountable to public oversight and the courts; the military has its own legal system, code of conduct, and "criminal justice" institutions (military police, administrative tribunals and courts, investigative units). Civilian law enters only at the appeals level, after military institutions have reached a decision on alleged criminal misconduct or violations of the code of conduct. Control over actions by a professional military in a democracy is exercised largely internally, supported by professional norms which accept that the military is ultimately subject to civilian leadership (by the "Commander in Chief," or some such term). Unlike the police, which are always guided and judged by laws made through an open, participatory political process external to the police, the military judges itself by its own laws and rules with external domestic political oversight and international conventions on the conduct of war as the ultimate backstops.

The argument predicts that the "militarization" of the culture, policies, and actions of police – that is, the adoption by civilian police forces "of militaristic practices, hardware, technology, values, language, and ideology" (Kraska 1997: 299) – will be a danger to democratic norms and professional ethics of the police. Rather than follow the rule of law, professional codes of conduct, or commitment to public service and the protection of human rights, the police will drift toward a professional ideology which stresses the need to protect national and local security against threats and risks without interference or oversight by legal institutions, political oversight, or public demands. The necessary democratic balance between the rights of people and the practical need for security will tilt strongly toward security.

This argument rests on four main assertions. First, the police will alter their orientation to the use of force and the identification of categories of people against whom force can and should be used quickly and easily, justified by appeals to an "enemy" and "threat" language. Second, the police will shift from an orientation to public service toward a suppression of crime and disorder mentality and compliance with political definitions of threats to public order and safety. Third, the police will be drawn into the secretive collection of intelligence and categorical risk assessments which invade the privacy of people and assume threats, risks, and criminal intentions based on an individual's membership in categories of persons (e.g., young male members of minority or ethnic groups, bikers, Roma, Muslims, travelers from certain countries) as these categories are defined by the police and intelligence agencies. Perceived threats and risks will have to be deterred and prevented by proactive police and intelligence collection policies before they can be executed. Fourth, transparency will decline. The police will become less accountable via civic oversight of their actions.

There is evidence to support these arguments but less support for the conclusions that democratic policing is threatened and may be replaced by new forms of aggressive, forceful policing which undermine basic privacy rights and equal treatment of people and communities.

Beyond Ideal Typing: Existing Overlaps in History and Space

The reality is that the police have always done a variety of work in nondemocratic ways and that the military has always worked with the police on specific security and order maintaining tasks. In practice, historically and currently, the police and the military have exhibited overlapping characteristics and work habits.

Varieties of Policing

Being visible, available, and responsive to the needs of people, though crucial to any conception of democratic policing, is not the only type of work done by any police force. Tasks specified for the police have always included multiple goals and roles ranging from civil policing to state security policing (Brodeur 1983). All police forces anywhere include functional specializations which approach military-style tactics and mentalities: riot control, specialized units confronting dangerous situations (armed response teams, paramilitary police units), border policing, state security police protecting the political regime against threats to its rule, and the complete exception to democratic norms of colonial policing. Specialized police agencies and units tend not to share the mentalities, priorities, and practices which are claimed to be the defining traits of civil, democratic, and accountable police force.

Control of riots and demonstrations is not targeted primarily against individual malfeasance but is deployed as an organized, forceful police response to public, frequently violent, disorders (Della Porta and Reiter 1998). In most countries, should the police fail to control disorders, the military (e.g., the National Guard in the USA) stands in the background ready to enter the fray and restore order.

Distinct units trained in and employing military-style tactics (PPUs, Paramilitary Police Units) have infiltrated normal policing, a process which Kraska and Kappeller (1997) have argued is expanding. PPUs are expensive; not using them wastes public money; they have high status within the police profession; hence, their arms, group cohesion, and tactics are attractive to other police.

Persons entering a country at recognized border crossing points, whether legally or illegally, have restricted democratic rights based on the overarching need to keep the country secure against threats and risks. Seeking to cross outside of legal control points will bring swift and forceful reactions from border guards who can be police, military, or integrated units (Andreas 2009; Caparini and Marenin 2006).

At the US land borders, Marine units are deployed along the Mexico-US border, nominally for exercises but also to occupy and preempt territory through which threats and illegal migrants might enter the USA (Dunn 1996). At the Canada-US border, IBETs (Integrated Border Enforcement Teams), which include Canadian and US police, military, and intelligence agents, are a more recent development dealing with security threats (IBETs, web). Also, at that border, military drones have been called upon to assist local law enforcement in detecting fleeing felons and have responded successfully. The Coast Guard, an arm of the armed forces, has been responsible for the protection of sea borders. In most countries, including the European Union member states, border control is shared with police agencies or done exclusively by the military.

State security police tasked with protecting the state, the ruling regime, and the powerful against challenges to their rule and status have few qualms about abiding by legal and democratic norms, nor do state security police take their orders from or are responsive to public concerns. Their job is to keep those in power safe. This is true in democratic countries and even more so in authoritarian, transitional, or military states. The prototype of state security police agencies were the KGB in the Soviet Union and other socialist and authoritarian countries, but they can also be found in democratic countries, as during the Franco dictatorship in Spain when the Guardia Civil (a constabulary force) was the tool used to terminally eliminate political and intellectual opponents of the regime, or in lesser and more benign forms, such as transnational threat assessment and anti-terrorism units in police departments or specialized security agencies, to arrest suspects before they can act (Brodeur 2000).

In the history of state development (Bayley 1985; Tilly 1990) and the creation of security forces, there has always been a significant overlap between police and military work. That overlap has been strengthened in recent years by the shift in organized violence from traditional forms of aggression (requiring a military response) toward diffused, asymmetrical, nonconventional, substate forms – civil wars, guerilla insurgencies, identity group-based riots, and killings; the emergence in many parts of the world of powerful organized transnational crime groups who exert influence on states and societies by corruption, fear, and violence; and the growth in privatized policing and international policing and security providing companies. These developments have been enabled and smoothed by new technologies for sharing information and intelligence on a (almost) real-time basis. This new world (dis) order (Kaldor 2001; Oakley et al. 1998) has led to a transformation in how all state-based, and even non-state, actors (e.g., private risk assessment and security companies) can be engaged in an integrated manner against new threat and risk dynamics.

This convergence is most apparent in internationally sanctioned peacekeeping and peacebuilding operations. Police in UN and regional peacekeeping interventions, where local security forces have collapsed (e.g., Haiti, Timor Leste, Kosovo), have had to assume semi-military suppression and control functions as well as “normal” policing operations (Bayley and Perito 2010; Goldsmith and Harris 2010; Greener 2009).

The biggest exception to democratic, nonmilitarized policing have been colonial policing systems. When European colonial powers developed police organizations for their colonies, they did not export policing systems deemed appropriate for their own, that is, civilized people, but constabulary forces to protect their economic, political, and “civilizing” interests in their colonies. As the majority of the world’s territories and states are former colonies, at independence they inherited nondemocratic and militarized policing systems which since have been quite resistant to changes, despite massive efforts to make them more democratic in their performance.

Constabulary forces are semi-militarized police units created in different political, organizational, and cultural forms. In France, the gendarmerie developed as an occupying force in outlying areas of what is now France, in regions which opposed the imposition of central rule from Paris. In consequence, until very recently, the gendarmerie was located within the Ministry of Defense, even though it had become the police force which protected small towns and rural areas in France.

British policy on policing England versus policing the colonies exemplifies the use of constabulary forces. The British practiced how to control restless populations during their centuries-long and brutal rule of the Irish. After Irish independence, policing in Northern Ireland, which remained part of the United Kingdom but was beset by ethnic, religious, and political violence and conflicts, continued the constabulary tradition via the Royal Ulster Constabulary, which was used to protect Irish loyalists in Northern Ireland often in cooperation with the British military. The lessons learned by the British in their colonial rule of Ireland became the model for policing their colonies (Brogden 1987).

The Spanish and Portuguese colonizers exported their security systems to Latin and Central America which closely combined military and policing functions under one overarching organization. The police in most Latin American countries remained part of the military until fairly recently but still retain their titles (e.g., the *Policia Militar* in Brazil) and their outlook on citizens and criminals, captured nicely in the title of Husain's (2007) study of police reform in Rio de Janeiro, "Those Who Die in War Are Not Innocent."

The USA established constabulary forces in Central America and the Caribbean to help protect American investments and personnel against "revolutionary" agitations. The constabularies became converted into (para)military forces once local control took over. In Nicaragua, the armed forces of the Somoza government, which engaged in a protracted war with the Sandinista Revolutionary Front to try to keep the government in power, were drawn mainly from that constabulary and were supported by illegally funded paramilitaries (the contras) by the USA. US police assistance under the aegis of OPS (Office of Public Safety), in the late 1960s and early 1970s, stressed counterinsurgency over democratic norms and roles which were easily picked up by local police forces in Southeast Asia and Latin America, the recipients of the majority of OPS funding and mentoring (Huggins 1987).

Colonial policing had few attributes that one would characterize as civil or democratic policing. The police in colonies were used mainly by the colonial administrators to "pacify" the restless native, engage in military-style campaigns to punish local groups which resisted the imposition and demands (pay taxes, conscripted work) of colonial rule, protect tax collectors and the persons and property of commercial enterprises and missionaries, and suppress local riots and sabotage. There was no pretense that the police were to serve the population in any meaningful way; their job was to enable and protect colonial rule and all its activities, privileges, and laws.

The modification of imposed policing system after decolonization has been slow and incremental, and much of the occupational cultures and priorities of the police have not been responsive to demands for reforms. For example, the most populous democracy in the world, India, still uses the colonial Police Act, passed in 1861, as the basis for defining the jobs of the police which, despite numerous constitutional, ideological, and legal statements, still shape the way in which state and federal police understand their job (Verma 2011). In Morocco, military officers still occupy the commanding positions in the gendarmerie, the police force which patrols the roads and services the rural areas of the country.

In short, democratic policing, operationally defined as being available and willing to be of service to people in need, has been a rare policing model in history or is currently across the globe. Still, democratic policing as practiced by developed countries, currently in the guise of some form of community policing, has become the goal of police reforms promoted through international assistance programs. Militarization is seen as threatening progress toward that goal.

Reforms of policing systems in countries where policing and justice systems have collapsed increasingly include the institutional separation of the police from the military and argue for the demilitarization of the occupational culture of the police. For example, the Chapultepec Peace Accords, signed in 1992, which ended the civil war in El Salvador, dissolved the National Police, then under the control of the military, and reconstituted a National Civil Police with detailed instructions on the organization, tasks, recruitment, training, and democratic responsibilities. Reform and democratic governance were deemed not possible unless the police were reconstituted and removed from the organizational authority of the military.

Yet one must be careful when drawing a distinction between the organizational home of "policing agencies," such as the gendarmerie, which are housed within the armed forces, and organizational and occupational cultures. Being under the administration of the military does not prevent democratic policing, nor does removing them from military administrative oversight create more democratic policing. The police can be quite forceful and abusive whether they work under military or civilian organizational umbrella. What matters are the organizational and occupational norms and cultures which are the guiding doctrines and "recipe rules" for the police.

Constabularization of the Military

The military, in its various institutional divisions, has been increasingly drawn into what used to be considered "policing" work in a variety of ways: assisting in border controls, intelligence sharing with domestic law enforcement agencies, counter-insurgency strategies to reach the hearts and minds of local populations, United Nations or regionally authorized peacekeeping operations which include military and police contingents, and "nation-building" after multilateral and regional interventions in failed and conflicted-ridden states. These trends undermine the traditional restrictions of the armed forces to war-fighting and war-ending work by involving them in tasks for which they have little training, few skills, and limited willingness (Friesendorf 2010; Friesendorf and Kempel 2011; Perito 2011).

Training for military and police being deployed in international interventions, and increasingly for public order policing domestically, is conducted in a language which blurs traditional lines, which has led to both the constabularization of military goals and cultures and the paramilitarization of policing ideologies (Moelker 2010), and which will continue to undermine the bright line between police and military organizational cultures and practices and the distinction between domestic and external authority to exercise coercion.

This convergence is not by design but forced by the nature of current international and domestic security climates. For example, the explosion of transnational crimes which affect domestic security (drugs, human trafficking, arms trade) requires that police now work in other countries, in cooperation with local police, militaries, and intelligence agencies to control crime. Preventing convergence, or alternatively ensuring a continued separation and distinction between police and military work, will be difficult to maintain.

Beyond Dichotomies: Theorizing the Police-Military Nexus

How to theorize the new security structures? The most promising theoretical approach is the notion of a security sector or system (SS) and corresponding conceptualizations of reform (SSR) and governance (SSG). The security sector, in its leanest definition, includes the armed forces, police, border guards/police, and intelligence agencies as the core state security providers. In more expansive conceptions, the SS includes the criminal justice system, legal aspects of security, and non-state providers (Bryden and Hänggi 2004; OECD Organization for Economic Co-operation and Development 2007).

Security sector theorizing stresses the interconnectedness of the four core state security agencies. Reforms of the security sector, or specifically the police, to establish more democratic professional and accountable norms and actions cannot be simply focused on the police but must take their connections to other state-based security and private security providers into account. Security is provided by many actors, and that requires that they cooperate and work together domestically and internationally.

The political nature of policing, and other security policies, is central to the notion of a security sector and its governance. In addition to demilitarizing the police, reformers argue that the police be de-linked from centralized, political control of operational policies and tactics.

Concluding Comments

Concerns about militarizing the police are not illusory. Civil policing would change significantly if military cultures, norms, and practices began to infiltrate democratic cultures and styles of policing.

At the same time, militarization is not the greatest threat to democratic policing. More serious is the involvement of the police in secretive intelligence gathering, often along stereotypical categorizations of terrorist, transnational crime groups, or local "troublemakers," such as "suggestions" from the FBI to local police to map the distribution of mosques in their communities or efforts by the antiterrorist unit in the New York Police Department to infiltrate and keep tabs on Islamic leaders. The police who conduct such secret invasions and information gathering of people's normal, and legal, activities by the justification of local and national security will loose, once their police activities become known, political support and legitimacy.

The secret gathering of intelligence is just as likely to change the norms and cultures of policing as does militarization. The loss of transparency in collecting intelligence about normal activities disconnected from actual criminal threats and the convergence of domestic and international work from both the police and the military perspectives, both driven by political and national security justifications and the changing global world, are the historical and current reality which will not disappear. For example, President Reagan declared, by his authority, international drug trafficking a national security threat, not just a crime, thereby authorizing crime control and intelligence collecting actions which were legally prohibited to domestic police.

The militarization of the police happens for diverse reasons, be it by the promotion of that style of defining their work by police themselves; by the acceptance of a frustrated public fed up with disorder, crime, and fears which beset their lives and willing to give up some rights for greater protection; by depictions in the mass and entertainment media which portray armed police tactics as the normal response to increasingly dangerous situations in normal police work; and by the exploitation of risks and threats by political polemicists.

The issues are whether the militarization of policing and the constabularization of the military will become dominant norms and ideologies and whether the in-creeping of military styles into democratic policing will move that style permanently way from its service and order protection goals. It seems an unlikely possibility, as democratic policing is promoted worldwide, as part of larger political changes, by reformers, police officials, and progressive political leaders and also demanded by populations, as long as effectiveness levels are maintained sufficient to keep populations willing to grant legitimacy to democratic styles which balance crime control with justice. In the end, the greater threat to democratic policing is not militarization but the politicization of policing even in democratic countries, of which militarization is only one piece of evidence, and the increasing closure of the police to public knowledge and oversight linked to an increasing disregard for the rights of people and due process.

Related Entries

- ▶ Conceptualizing of Police
- ▶ Democratic Policing
- ▶ Policing of Peacekeeping

Recommended Reading and References

Andreas P (2009) Border games: Policing the U.S.-Mexico divide, 2nd edn. Cornell University Press, Cornell

Bayley DH (1985) Patterns of policing. A comparative international analysis. Rutgers University Press, New Brunswick

Bayley DH, Perito R (2010) The police in war. Fighting insurgency, terrorism, and violent crime. Lynne Rienner, Boulder

Brodeur JP (1983) High policing and low policing: Remarks on the policing of political activities. Soc Probl 30:507–520

Brodeur JP (2000) Cops and spooks: the uneasy partnership. Police Pr Res 1:299–321

Brogden M (1987) The emergence of the police – the colonial dimension. Brit J Criminol 27:4–14

Bryden A, Hänggi H (eds) (2004) Reform and reconstruction of the security sector, Baden-Baden. Nomos Verlagsgesellschaft, Germany

Caparini M, Marenin O (eds) (2006) Borders and security governance. Managing borders in a globalised world. LIT Verlag, Vienna

Della Porta D, Reiter H (eds) (1998) Policing protest. The control of mass demonstrations in western democracies. The University of Minnesota Press, Minneapolis

den Boer M, Janssens J, Vander Beken T, Easton M, Moelker R (2010) Epilogue. Concluding notes on the convergence between military and police roles. In: Easton M, den Boer M, Janssens J, Moelker R, Vander Beken T (eds) Blurring military and police roles. Eleven International, The Hague, pp 223–228

Dunn TJ (1996) The militarization of the U.S.-Mexico border, 1972–1992. Texas University Press, Austin

Easton M, den Boer M, Janssens J, Moelker R, Vander Beken T (eds) (2010) Blurring military and police roles. Eleven International, The Hague

Friesendorf C (2010) The military and law enforcement in peace operations. Lessons from Bosnia-Herzegovina and Kosovo. Geneva Centre for the Democratic Control of Armed Forces, Geneva

Friesendorf C, Kempel J (2011) Militarized versus civilian policing: problems of reforming the Afghan national police. Peace Research Institute, Frankfurt (PRIF Report No. 102)

Goldsmith A, Harris V (2010) Police-military cooperation in foreign interventions: Timor-Leste and the Solomon islands. In: Lemieux F (ed) International police cooperation. Emerging issues, theory and practice. Willan, Cullompton, pp 221–237

Greener B (2009) The new international policing. Palgrave Macmillan, New York

Huggins MK (1987) U.S.-supported state terror: a history of police training in Latin America. Crime Soc Justice 2728:149–171

Husain S (2007) Those who die in war are not innocent (‘Na Guerra, quem morre Não É innocente): human rights implementation, policing, and public security in Rio de Janeiro. Rozenberg, Brazil/Amsterdam

Integrated Border Enforcement Teams (IBETs). http://www.rcmp-grc.gc.ca/ibet-eipf/index-eng.htm

Kaldor M (2001) New & old wars. Organized violence in a global era. Stanford University Press, Stanford

Kraska PB (1997) The military as drug police. Exercising the ideology of war. In: Gaines LK, Kraska PB (eds) Drugs, crime, and justice. Contemporary perspectives. Waveland, Prospect Heights, pp 297–320

Kraska PB (ed) (2001) The militarizing of the American justice system. The changing roles of the armed forces and the police. Northeastern University Press, Boston

Kraska P (2007) Militarization and policing – its relevance to 21st century police. Policing 1:501–513

Kraska P, Kappeller VE (1997) Militarizing American police: the rise and normalization of paramilitary units. Soc Probl 44:1–18

Moelker R (2010) Cultures converging upon constabularization. In: Easton M, den Boer M, Janssens J, Moelker R, Vander Beken T (eds) Blurring military and police roles. Eleven International, The Hague, pp 151–169

Oakley RB, Dziedciz MJ, Goldberg EM (eds) (1998) Policing the new world disorder: Peace operations and public security. National Defense University Press, Washington, DC

OECD (Organization for Economic Co-operation and Development) (2007) OECD handbook on security system reform. Supporting security and justice, 2007 Edition. OECD Publishing, Paris. www.oecd.org/dataoecd/43/25/38406485.PDF

OSCE (Organization for Security and Co-operation in Europe) (2006) Guidebook on democratic policing. Author: Vienna

Perito RM (2011) The Iraq federal police. U.S. police building under fire. Washington, DC, United States Institute of Peace, Special Report 29x

Tilly C (1990) Coercion, capital and European states, AD 990–1990. B. Blackwell, Cambridge, MA

Verma A (2011) The new khaki. The evolving nature of policing in India. CRC Press, New York

Police Corruption

Sanja Kutnjak Ivković
School of Criminal Justice, Michigan State University, East Lansing, MI, USA

Overview

Police corruption is a form of police misconduct or police deviance typically defined through the motivation to achieve personal gain. Police corruption includes many heterogeneous forms of behavior that could be classified on the basis of several criteria, including the motivation for corruption (i.e., economic corruption v. noble-cause corruption), regularity of payments (i.e., pads v. scores), consequences (i.e., distortive v. non-distortive corruption), and the level of aggressiveness (i.e., grass-eaters v. meat-eaters). Barker and Roebuck developed a typology of police corruption which recognizes corruption of authority, kickbacks, opportunistic theft, shakedowns, protection of illegal activities, the fix, direct criminal activities, and internal payoffs. Punch added flaking (i.e., planting of evidence) or padding (i.e., supplementing of evidence) as the ninth type of corruption.

Theories explaining the causes of police corruption could be classified into four larger groups: (1) theories oriented toward individual police officers and their characteristics (i.e., "individualistic theory" or "rotten apple theory"), (2) theories examining the relation between opportunities for corruption and police corruption ("occupational theories"), (3) theories focusing on the relation between the characteristics of a specific police agency and police corruption ("organizational theories"), and (4) theories concerning the relation between the society at large and police corruption ("social structure theories"). Despite the existence of extensive literature on the theories of police corruption, research that tests the theories is very limited. Police corruption could be measured using several different methodologies. A relatively simple, yet very inaccurate way of measuring corruption is through the official data such as the arrest rates, complaint rates, and conviction rates. Somewhat more promising attempts to assess the actual level of corruption are public opinion surveys, police officer surveys, field studies, and investigations by independent commissions.

Police Corruption

Definition

Police corruption is a form of police misconduct or police deviance typically defined through the motivation to achieve personal gain (e.g., Barker and Carter 1986; Goldstein 1975; Klockars et al. 2000; Kutnjak Ivković 2005; Sherman 1974). The definition contains several essential elements.

First, corrupt behavior could be defined as a violation of the penal codes, administrative agency rules, or the codes of ethics. Like other citizens, police officers could violate norms of federal and state criminal codes. In addition, federal and state codes establish certain crimes which only public officials can commit (e.g., bribery of public officials and witnesses, Title 18 of the U.S. Code, Chapter 11, Section 201, 1999; extortion by public officials, Title 18 of the U.S. Code, Section 872, 1999; deprivation of civil rights, Title 18 of the U.S. Code, Section 242, 1999). Police agencies, particularly large municipal agencies, have administrative rules and policies compiled in the standard operating procedure manuals. These administrative rules channel the use of discretion, describe appropriate conduct of police officers, prohibit inappropriate conduct, instruct officers to complete written reports after critical incidents, and require supervisory oversight (see, e.g., National Research Council 2004). Even in the agencies with extensive rules, the rules could be ambiguous or fail to regulate certain aspects of corruption. Lastly, the International Association of the Chiefs of Police has developed the code of ethics for the US police officers, which explicitly asks of police officers to promise that they would not engage in corruption, bribery, or gratuities themselves nor condone such conduct by their fellow police officers.

Second, while corruption is typically viewed as corrupt behavior (thus suggesting that something is being actively done), corruption can include both active (i.e., act) and passive (e.g., omission) forms. A police officer who fixes a felony ticket in exchange for a bribe provides an example of an action, while a police officer who does not write a ticket in exchange for a bribe from a motorist caught speeding provides an example of omission.

Third, various definitions discuss the gain resulting from corrupt activities in terms of explicit personal gain (e.g., Goldstein 1975; Kutnjak Ivković 2005; Moore 1997) or implied personal gain (i.e., the definition rests on the idea that the money/goods from the corrupt act are to be consumed by the individual officers; e.g., Barker and Wells 1981; Roebuck and Barker 1974). Several authors (e.g., Bracey 1995; Carter 1990) included organizational gain in their definitions as well.

The goal typically has a monetary value; it typically involves "some tangible object, either cash, services, or goods that have cash value" (Barker 1996, p. 25). However, it is also possible that the gain is nonmonetary. While monetary value could usually be attached to, or calculated for, most of the items, in some instances it is difficult, or borderline impossible, to determine it. For example, fixing a speeding ticket for a sergeant's niece could have different price tags attached to it, depending on the extent of sergeant's gratitude and the ways in which he is ready and willing to express them. It can potentially lead to preferential treatment by the sergeant and the early promotion, thus yielding thousands of dollars in gain.

The size of the gain may be an important factor. The discussion in the literature about the minimum amount, that is, the amount below which the gifts of this value would not be considered as corruption, is intense. The Code of Ethics is explicit in prohibiting the acceptance of gratuities ("...never accepting gratuities"; 2011). Although the value of these gratuities may be small in each individual case, if such gifts are given on a regular basis (e.g., Ruiz and Bono 2004) and/or to a large number of police officers (e.g., Pennsylvania Crime Commission 1974), their total value vastly increases their individual value. Furthermore, the purpose of the gratuities may be to influence the police officers' decisions, as was illustrated on the examples uncovered by independent commissions (e.g., Knapp Commission 1972; Mollen Commission 1994; Pennsylvania Crime Commission 1974). Barker and Wells (1981, p. 11–12) argue:

> Police corruption for many officers often begins with the shared belief among the police peer group that "policemen have a right to a break," and the progression along the continuum of corruption is often so gradual that an officer is deeply involved before he realizes it.

The acceptance of gratuities could start "the slippery slope of corruption," leading from the less serious forms of corruption toward the more serious ones; "... [i]t is claimed that the acceptance of small gratuities such as free cups of coffee by police officers will increase the likelihood of, or lead by degrees to, or is not significantly different from, corruption of the worst kind (Kleinig 1996, p. 174). There are several steps in the process of "becoming bent," as Sherman describes it (1985), from the acceptance of minor "perks" and free drinks to regular payoffs and the involvement in the distribution and use of narcotics. As Michael Dowd (a former NYPD police officer who stole money and drugs, participated in larger drug rings, became a drug dealer, and was eventually caught, tried, and sentenced to 14 years in prison) emphasized in his testimony before the Mollen Commission, he would test the new officers with initial temptations which were but trivial violations of the NYPD's policy (Mollen Commission 1994). Once the officers crossed the line between the allowed and the forbidden, it was easier for them to continue and justify to themselves their further involvement in various violations of the Department's policy.

Proponents of the acceptance of gratuities (e.g., Kania 1988, 2004) argue that the acceptance of gratuities contributes toward the development of a friendly bond between the police and the citizens. Kania (1994, p. 2) continues that, although the "cheerful waves, warm greetings and welcome smiles," and letters of appreciation may result in better service just as likely as may the free drinks, police administrators obviously do not want to discourage such friendly gestures.

Typology of Police Corruption

Police corruption is a group of heterogeneous activities, from the acceptance of a bribe for not issuing a speeding ticket to leaking the information about the upcoming raid to the drug dealers. Literature has used several different ways of classifying these diverse activities.

Economic v. Noble-Cause Corruption

Corruption described in this entry is economic corruption or material-reward corruption. Its primary motivation is the material reward, be it in the form of money or other goods and/or services. Noble-cause corruption is different; it is defined as "corruption in the name of the moral rightness of good ends" (Crank and Caldero 2000, p. 5). When police officers behave in a manner consistent with noble-cause corruption, their actions are no longer regulated by the law; instead, they "act if they *are* the law" (Crank and Caldero 2000, p. 75). Police officers attempt to rationalize this "noble-cause corruption" as follows (Moore 1997, p. 63):

> I did something wrong, but justice demanded it, not tolerated it but demanded it, because I could put the guy away who otherwise wouldn't be successfully prosecuted. I, the police officer, wouldn't gain personally from it; I didn't get anything from it. I only acted for the community in the community sense of justice to accomplish this goal.

The relation between the police and the law changes in the case of noble-cause corruption (e.g., Crank and Caldero 2000). Police officers violate the law to achieve higher purposes or ends of policing. The causes of economic corruption and noble-cause corruption are different, as are the efforts used to control them.

Pads v. Scores

In terms of the regularity of the payments, corrupt activities could be divided into "pads" and "scores." If the payment takes place just once, as the opportunity presents itself, it is called a "score." The Knapp Commission (1972, p. 66) described the "score" as "a one-time payment that an officer might solicit from, for example, a motorist or a narcotics violator. The term is also used as a verb, as in 'I scored him for $1,500.'" On the other hand, if there is an established arrangement between a citizen and a police officer and the payments occurs on a regular basis, they are called "pads." The Knapp Commission (1972, p. 66) described "pads" as "regular weekly, biweekly, or monthly payments, usually picked up by a police bagman and divided among fellow officers. Those who make such payments as well as policemen who receive them are referred to as being 'on the pad.'"

Distortive v. Non-distortive Corruption

This classification focuses on the outcome of the police officer-citizen interaction. If, compared to the interaction without corruption, the outcome of the interaction with corruption is different, then such corruption is called "distortive corruption." For example, a police officer stops a motorist for speeding and wants to write a speeding ticket. If the citizen does not offer a bribe, the police officer will write the ticket. If, on the other hand, the citizen offers a bribe and the police officer accepts it, the police officer will not write the ticket. Thus, the outcome of the interaction (i.e., ticket v. no ticket) is different, depending on whether the bribe was offered and accepted. In cases of distortive corruption, the police officer does something that he was not supposed to do (e.g., revealed an undercover operation to a drug dealer, issued a permit to a nonqualified applicant) and does not do something he was supposed to do (e.g., did not arrest a citizen caught violating the law, did not issue a license to a qualified applicant).

If the outcome of the interaction is the same, regardless of whether the bribe was given or not, then such corruption is called "non-distortive corruption." For example, a police officer issues a passport or a liquor license to a qualified applicant. The police officer does what he is supposed to be doing or does not do something that he is not supposed to be doing. The reason for the bribe is that the citizen wants to secure something that he has a legal right to but may need to speed up the bureaucratic machine (hence the "grease money" term used for this type of corruption).

Grass-Eaters v. Meat-Eaters

According to the Knapp Commission (1972, p. 4, 65), dishonest, corrupt police officers can be classified as either "grass-eaters" or "meat-eaters." "Meat-eaters" are police officers who aggressively misuse their police power for personal gains, while "grass-eaters" simply accept the payoffs the circumstances of police work throw their way (Knapp Commission 1972, p. 4). A police officer like Michael Dowd who beats up drug dealers and steals their money and drugs (e.g., Mollen Commission 1994) would be a clear example of a "meat-eater." A police officer who is on the pad and who accepts the money that the bagman is giving him every 1st of the month (see, e.g., Knapp Commission 1972) would be an example of a "grass-eater."

A strong support of the code of silence and reluctance of honest police officers to report their corrupt colleagues, on the one hand, and an aggressive misuse of police powers by "meat-eaters," on the other hand, provide an atmosphere in which it is very easy and natural for "grass-eaters" to accept payoffs. The Knapp Commission (1972, p. 65) suggested that "grass-eaters" continued to accept the gain from corrupt activities out of their feeling of loyalty to their fellow officers and that the "grass-eaters" are "the heart of the problem" since "[t]heir great numbers tend to make corruption 'respectable'" (Knapp Commission 1972, p. 4). According to the Knapp Commission (1972, p. 65), unlike "meat-eaters," "grass-eaters" are not willing to take considerable risks in order to obtain illegal gains, and their behavior is considerably more likely to be affected by a change of atmosphere/attitudes in the department.

Barker and Roebuck's Typology

The most frequently used typology of police corruption has been developed by Barker and Roebuck (1973; Roebuck and Barker 1974). In their joint work, Barker and Roebuck (1973; Roebuck and Barker 1974) point out that corruption takes many forms. Based on several dimensions (e.g., acts and actors involved, norms violated, support from peer group, organizational degree, police department's reaction), they specify eight types of corruption: corruption of authority, kickbacks, opportunistic thefts, shakedowns, protection of illegal activity, the fix, illegal criminal activity, and internal payoffs.

The first type involves cases of **corruption of authority**. Barker and Roebuck (1973, p. 21) describe it along the following lines: "the officer's authority is corrupted when he receives officially unauthorized, unearned material gain by virtue of his position as a police officer without violating the law *per se*." Thus, the actions are not violations of criminal law, but they do violate departmental policies. Examples of such activity would include acceptance of free drinks, free food, discounts on merchandise, or providing property protection for a fee. According to the findings of the Knapp Commission (1972) and Barker and Roebuck's study (1973, pp. 22–23), the practice seemed to be widespread (31 % of the businessman in a study *openly* acknowledged providing favors to the police), and the corruptors, typically respectable citizens, according to Barker and Roebuck, give these gifts for a reason. Police officers perceive these gratuities as informal rewards and approve of and support them within the peer group; police officers who refuse such gifts are perceived by their fellow officers as deviant.

The second type of police corruption involves **kickbacks**. They are defined as the acceptance of goods, services, or money for referring business to various businesses and service providers, including towing companies, ambulances, garages, lawyers, and doctors (Barker and Roebuck 1973, p. 24). These activities do not violate criminal laws, but they do violate departmental policies. According to Barker and Roebuck (1973, pp. 24–25), the corruptors are legitimate businessmen and professionals whose purpose is to develop and maintain a good working relationship with the police. These payments are perceived by police officers as clean and are, therefore, supported by the peer group (Barker and Roebuck 1973, p. 25), whereas the departments typically either condone or overlook these kickbacks, provided that they are made by legitimate businesses discreetly. In the case of goods and services, the department's reaction ranges from acceptance to mild sanctions, with cash rewards resulting in a more severe punishment.

The third type of police corruption involves **opportunistic theft** from arrestees, victims, crime scenes, and unprotected property (Barker and Roebuck 1973, p. 26). These behaviors violate both the departmental rules and the criminal rules. The disciplinary measures vary from mild to very serious (including criminal charges and dismissal).

The fourth type of police corruption involves **shakedowns**. Shakedowns occur upon the police officer's discovery of both the criminal violation and the violator, and they result in the police officer's acceptance of a bribe in exchange for not making an arrest (Barker and Roebuck 1973, p. 27). One of the characteristics of shakedowns is that the victim is rather unlikely to complain, because the victim and the police officer could be found guilty of another crime – bribery. These actions are, of course, violations of both the departmental rules and the criminal codes.

The peer culture and the department itself may react differently to the "clean" money than to the "dirty" money, which are distinguished by the source of the money. Officers who are publicly exposed receive severe punishment, including potentially dismissal, and may face a criminal prosecution (Barker and Roebuck 1973, pp. 28–29).

The fifth type of police corruption involves **protection of illegal activities**. Individuals involved in illegal activities reward police officers in order to be able to operate without police harassment (Barker and Roebuck 1973, p. 29). The corruptors may be citizens with long criminal records, as well as legitimate businesses operating illegally. Some of the activities protect the illegal services and goods which citizens desire or perceive as necessary. This form of police corruption involves a high degree of organization; police members must be coordinated and know which places enjoy police protection. Departmental reaction depends on the "degree of its own involvement with criminal organizations or legitimate businesses that operate illegally, informal definition of *clean* money, identity of the corruptor, and whether or not there is public disclosure of flagrant violations" (Barker and Roebuck 1973, p. 33).

The sixth type of police corruption involves **the fix**. The fix may include "the quashing of prosecution proceedings following the offender's arrest and . . .the taking up (disposal of record) of traffic tickets" (Barker and Roebuck 1973, p. 34). The corruptors are arrestees who want to avoid the court action in their case, and the police officer involved typically "fails to request prosecution, tampers with the existing evidence, or gives perjured testimony" (Barker and Roebuck 1973, p. 34). The peer approval depends on whether the case to be fixed is a felony, misdemeanor, or a traffic case. Departmental reaction to the fixing of criminal cases is generally severe (Barker and Roebuck 1973, p. 35).

The seventh type of police corruption is the police officers' involvement in **direct criminal activities**. Barker and Roebuck (1973, pp. 35–36) argued that "policemen directly commit crimes against the person or property of another for material gain, acts which are clear violation of both departmental and criminal norms." Because the profit coming from this transaction is perceived to be "dirty" money, the fellow police officers would typically provide very little support for this type of police corruption and even the departments that may tolerate other forms of police corruption, will react severely by firing the police officers and pressing criminal charges (Barker and Roebuck 1973, p. 36). Some organization is typically required for this type of police corruption to be carried out, and it involves small groups of police officers.

The last, eighth type of police corruption involves **internal payoffs**. In the case of internal payoffs, both the corruptors and the corrupted are police officers who sell or buy assignments, off-days, holidays, promotions, etc. (Barker and Roebuck 1973, p. 36). Peer groups either do not support this type of police corruption or perceive it as necessary and inevitable if they are engaged in other forms of police corruption (Barker and Roebuck 1973, p. 37). In the departments in which other forms of police corruption flourish, this form exists as well and is highly organized. Possible reactions by the departments may range from informal approval to dismissal and pressing criminal charges (Barker and Roebuck 1973, p. 38).

Punch (1985) added **flaking** (i.e., planting of evidence) or **padding** (i.e., supplementing of evidence) as the ninth type of corruption. According to Punch (1985), this type of corruption is particularly evident in drug-related cases.

In a nationwide study of more than 3,000 police officers from 30 diverse US police agencies, Klockars and colleagues (2000) found that shakedowns and opportunistic thefts were evaluated to be the most serious types of corruption in all 30 agencies, while the acceptance of gratuities – be it on a regular basis or only for the holidays – was viewed as the least serious form of corruption in all 30 agencies, with the cases of internal corruption and kickbacks lying somewhere between these two extremes. These results are very consistent with results of surveys of police officers from 13 other countries as diverse as Croatia, Finland, Japan, Pakistan, and South Africa (Klockars et al. 2004).

Causes of Police Corruption

The National Research Council (2004, p. 271) points out that "[t]he research literature [on causes of police corruption] is long on theory and short on evidence about what causes police corruption." The existing literature proposes several different approaches toward understanding why police officers engage in corrupt behavior, ranging from police officer individual characteristics to characteristics of the police agency itself and its larger social environment.

The first group of studies focuses on individual police officers and their characteristics (e.g., Muir 1977). The literature in this area tries to ascertain the features which make police officers prone to corruption, such as their prior criminal record and weak moral values. The Knapp Commission (1972) called this the "rotten apple approach." Over time, scholars discovered that psychological screening tests, traditionally used to prevent future "rotten apples" from entering the police organization, are not accurate predictors of future behavior.

The second group of studies examines the relation between opportunities for corruption and police corruption. By its nature, policing is viewed as an occupation rife with opportunities for corruption (Klockars et al. 2000). Crank and Caldero (2000, p. 63) provide an example:

> When we think of police corruption, graft typically comes to mind. The police, in their day-to-day pursuits, are exposed to great temptations, and there are few observers to watch what they do. Imagine this. You're a street officer. You've just made an incredible drug bust. Your reputation in the department is assured, and you're feeling charged! There's a pile of money on the floor. All you have to do is reach down, scoop up a handful, and put it in your pocket, and you can put your kid through college.

The opportunities for corruption vary assignments, ranks, units, and police agencies. Detectives, particularly those assigned to narcotic units, have especially extensive opportunities for corruption (General Accounting Office 1998), as do police officers in charge of laws without moral consensus and vague-defined laws (Knapp Commission 1972).

The third group of studies analyzes the relation between the characteristics of a specific police agency ("rotten barrels" or "rotten orchards," Punch 2009) and police corruption. This approach argues that police agencies have the dominant role in addressing police misconduct by creating systems that establish rules, enforce rules, detect corruption, and control the code of silence (e.g., Klockars et al. 2000; Kutnjak Ivković 2005; Sherman 1974; 1978). The police chief and his top administrators have critical roles (e.g., Goldstein 1975; Knapp Commission 1972; Kutnjak Ivković 2005; Pennsylvania Crime Commission 1974), while the roles of first-line supervisors (e.g., Knapp Commission 1972; Mollen Commission 1994) and peers (e.g., Chen 2003; Klitgaard 1988; Kutnjak Ivković 2005; Stoddard 1974) are not negligible either.

The fourth group of studies examines the relation between the society at large and police corruption. The police agency is part of the larger environment and is influenced by the legal norms (e.g., Knapp Commission 1972) and public expectations (e.g., Goldstein 1975; Sherman 1977). Sherman (1977) proposed that communities differ greatly in their expectations, from "communities with a more public-regarding ethos" (like Charlotte, North Carolina, Kansas City, Missouri, and Portland, Oregon) to communities with more "private-regarding" (like New York City and New Orleans, Louisiana). The same argument could be used to explain differentiation in corruption rates across the world; in the study of International Crime Victim Surveys, Kutnjak Ivković (2003, p. 612) reveals that "the countries with the reputation in the international business community of being more corrupt, as indicated by a low score on the 1999 Corruption Perception Index (CPI) . . . appear also to have a higher percentage of the respondents who said that they had been asked to pay a bribe to a police officer last year."

Measuring the Extent of Police Corruption

The measurement of the extent of police corruption could be attempted at different levels (Fig. 1).

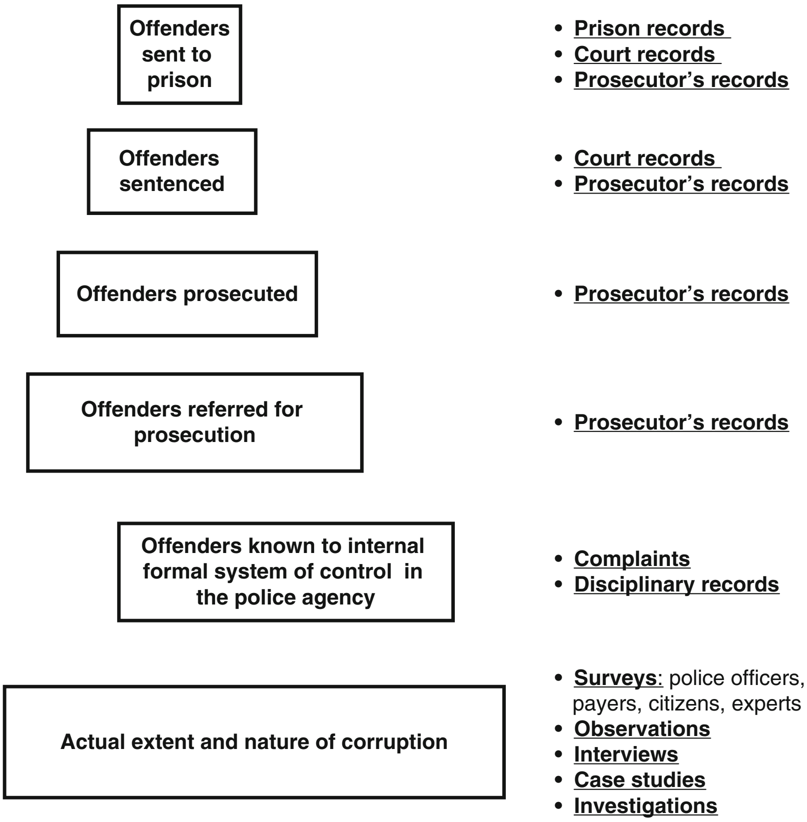
Offenders sent to prison
Offenders sentenced
Offenders prosecuted
Offenders referred for prosecution
Offenders known to internal formal system of control in the police agency
Actual extent and nature of corruption
• Prison records
• Court records
• Prosecutor's records
• Court records
• Prosecutor's records
• Prosecutor's records
• Prosecutor's records
• Complaints
• Disciplinary records
• Surveys: police officers, payers, citizens, experts
• Observations
• Interviews
• Case studies
• Investigations

Police Corruption, Fig. 1 The funnel of police corruption and the data collection methods. source: Kutnjak Ivković 2003

The data sources seeking to assess the actual level of police corruption include citizen and police officer surveys. The results of citizen surveys portray a heterogeneous picture of the extent of police corruption across the country; in the 1960s, fewer than 2 % of Caucasian respondents nationwide perceived that most of the police were corrupt (President's Commission on Law Enforcement and Administration of Justice 1067b), in contrast to 93 % of New Yorkers in the 1990s who perceived corruption to be widespread (Kraus 1994).

The public opinion surveys indicate that, on the one hand, the public tends to have a relatively positive opinion about police honesty, while, on the other hand, the public perceives that police officers frequently engage in corruption. A 1987 survey of Philadelphia citizens (Moore 1997, p. 62) revealed that the public provided very positive ratings of the police service, although one-third of the respondents thought that police officers *often* took bribes. Similarly, a 1994 survey of New York citizens showed that, while 93 % of the surveyed citizens perceived that corruption is either "widespread" or "limited," about one-half of these same citizens estimated that the police are doing a "good" or an "excellent" job (Kraus 1994).

On the comparative front, the International Crime Victimization Survey asked the respondents whether they have paid a bribe and, if so, who the recipient of the bribe was. About one percent of the respondents or fewer from Western democracies, including the USA, reported paying a bribe to the police, while the corresponding percentages were dramatically higher (between 10 % and 20 %) in some East European, Asian, and Latin American countries (Kutnjak Ivković 2003). In addition, the Gallup International 50th Anniversary Survey (1996) reports that approximately one-third of the respondents in the West European countries and Israel, and more than two-thirds of the respondents in the East European countries, the Far Eastern countries, and the

Central and South American countries assessed that police corruption was widespread in their countries.

According to the results of the police officer surveys, it seems that the perceived frequency of occurrence of corruption was negatively related to perceived severity of corruption. In the 1994 Illinois study (Martin 1994, p. 33), fewer than 0.4 % of the respondents said that they saw a police officer accepting a bribe, stealing property, or purchasing stolen merchandise in the past year, while 81 % of the police officers said that they saw a police officer accepting free coffee or food from a restaurant. Similarly, Ohio police officers participating in the 1996 survey (Knowles 1996) reported observing serious types of police corruption (acceptance of a payment to overlook illegal activity, purchase of stolen merchandise for personal use or gain) very infrequently (less than 0.6 % observed it last year; less than 4.7 % observed it during their careers), while they reported observing police officers accepting free coffee or food from restaurants quite frequently (71 % observed it over the course of the last 12 months; 87.3 % observed it during their careers).

Surveys of citizens and police officers alike are burdened with the methodological problems (see, e.g., Kutnjak Ivković 2003; National Research Council 2004, p. 269). For various reasons, neither police officers nor citizens (participants or potential witnesses) have motives to report a corrupt transaction; in fact, they have motives *not* to report it (e.g., Klockars et al. 2000; Kutnjak Ivković 2003; Kutnjak Ivković 2005; Stoddard 1974). In addition, most of the existing surveys focus on one country and are conducted locally, usually with the purpose of surveying the population of a particular city.

Another source includes independent commission report. While providing the results of an in-depth investigations, these independent commissions are limited because they refer to a specific police agency, time period, and are bound by the resources and powers granted to the commission. The Knapp Commission (1972) and the Pennsylvania Crime Commission (1972) find widespread corruption in the NYPD and Philadelphia Police Department, respectively, and the presence of a strong code of silence. In contrast, the Mollen Commission (1994) reports that most police officers in the NYPD are honest but still found pockets of police officers aggressively seeking opportunities to obtain money and drugs.

A potential source of information about corruption includes field studies of the police. Typically, they focus on a small number of police agencies and involve a combination of methods (e.g., observation, interviews, analyses of documents). The President's Commission on Law Enforcement and the Administration of Justice made a push for field studies of the police in the 1960s. Reiss's study of the police (1971), originating as one of the studies submitted to the President's Commission on Law Enforcement and the Administration of Justice, also contained the data about the frequency with which police officers engage in police misconduct. The rates of corrupt behavior per 100 police officers, based on the observations by the researchers, self-reports by police officers, and allegations of misconduct by others (calculated from the data provided in Reiss's study 1971) were 22.8 in City X, 20.5 in City Y, and 15.6 in City Z (see Kutnjak Ivković 2003).

Another level of measuring corruption is through the official data, be it arrest rates or the complaint rates. At the federal level, there were between 83 and 150 officers convicted annually in the period from 1993 to 1997 (General Accounting Office 1998, p. 11). A comparison of the official data with the reports by independent commission implies that official data may be just the tip of the iceberg. When the Knapp Commission (1972) reported widespread corruption n the NYPD, the prosecutors filed charges in only about 30 cases of corruption annually (Kutnjak Ivković 2003). Similar problems could be expected for the police agency's internal records of corruption complaints. At the same time when the Knapp Commission (1972) discovered widespread corruption in the NYPD, the complaint rate in the NYPD was less than 1 per 100 officers.

The most novel approach is to measure the extent of police integrity instead. Developed by

Klockars and Kutnjak Ivković (Klockars et al. 2000), it measures the level of police integrity and avoids the methodological problems associated with the direct measurement of police corruption. A study of 30 police agencies (Klockars et al. 2000) revealed that police agencies varied considerably in the contours of their police integrity.

Related Entries

- ▶ Control of Police Misconduct
- ▶ Law of Police Searches
- ▶ Law of Police Seizures and the Exercise of Discretion
- ▶ Law of Police Use of Force
- ▶ Police and the Excessive Use of Force
- ▶ Police Culture
- ▶ Police Discretion and Its Control
- ▶ Police Integrity
- ▶ Police Lying and Deception

Recommended Reading and References

[Mollen Commission] New York City Commission to Investigate Allegations of Police Corruption and the Anti-Corruption Procedures of the Police Department (1994) Commission report. Mollen Commission, New York

Barker T, Carter DL (1986) A typology of police deviance. In: Barker T, Carter DL (eds) Police deviance, 3rd edn. Anderson Publishing, Cincinnati

Barker T, Roebuck J (1973) An empirical typology of police corruption. Charles C. Thomas Publisher, Springfield

Barker T, Wells RO (1981) Police administrators' attitudes toward the definition and control of police deviance. FBI Law Enforc Bull 51(3):8–16

Bracey DH (1995) Police corruption. In: Bailey WG (ed) The encyclopedia of police science, 2nd edn. Garland, New York, pp 545–549

Carter DL (1990) Drug-related corruption of police officers: a contemporary typology. J Crim Justice 18(2):85–98

Commission to Investigate Allegations of Police Corruption and the City's Anti-Corruption Procedures (1972) Knapp Commission report on police corruption. G. Braziller, New York

Crank JP, Caldero MA (2000) Police ethics: the corruption of noble cause. Anderson Publishing, Cincinnati

Gallup International (1996) Gallup international 50th anniversary survey. Unpublished manuscript. On file with the author

General Accounting Office (1998) Law enforcement: information on drug-related police corruption. General Accounting Office, Washington, D.C

Goldstein H (1975) Police corruption: a perspective on its nature and control. The Police Foundation, Washington, D.C

Ivković SK (2003) To serve and collect: measuring police corruption. J Crim Law Criminol 93(2–3): 593–649

Ivković SK (2005) Fallen blue knights: controlling police corruption. Oxford University Press, New York

Kania RR (1988) Should we tell the police to say 'Yes' to gratuities? Crim Justice Ethics 7(2):37–49

Kania RR (1994) Invalidating the 'Dowd Test'. Ethics Roll Call 2(1):2–4

Kania RR (2004) The ethical acceptability of gratuities: still saying "Yes" after all these years. Crim Justice Ethics 1:54–63

Klitgaard R (1988) Controlling corruption. University of California Press, Berkeley

Klockars CB, Ivković SK, Harver WE, Haberfeld MR (2000) The measurement of police integrity. Research in Brief. U.S. Department of Justice, Office of Justice Programs, National Institute of Justice. Government Printing Office, Washington, DC

Klockars CB, Ivković SK, Haberfeld MR (eds) (2004) The contours of police integrity. Sage, Newbury Park

Knowles JJ (1996) The Ohio police behavior study. Office of Criminal Justice Services, Columbia

Martin C (1994) Illinois municipal officers' perceptions of police ethics. Illinois Criminal Justice Information Authority, Statistical Analysis Center, Chicago

Moore M (1997) Epilogue, Police integrity: public service with honor. U.S. Department of Justice. Available at https://www.ncjrs.gov/pdffiles/163811.pdf

Muir W (1977) Police streetcorner politicians. University of Chicago Press, Chicago

National Research Council Committee to Review Research on Police Policy and Practices (2004) Committee on Law and Justice, Division of Behavioral and Social Sciences and Education. In: Skogan W, Frydl K (eds) Fairness and effectiveness in policing: the evidence. The National Academies Press: Washington

Pennsylvania Crime Commission (1974) Report on Police corruption and the Quality of Law Enforcement in Philadelphia. Pennsylvania Crime Commission, Saint Davids

President's Commission on Law Enforcement and Administration of Justice (1967) Task force report: the police. GPO, Washington, D.C

Punch M (1985) Conduct unbecoming. Tavistock, London

Punch M (2009) Police corruption: deviance, accountability and reform in policing. Willan Publishing, Portland

Reiss AJ Jr (1971) The police and the public. Yale University Press, New Haven

Roebuck JB, Barker T (1974) A typology of police corruption. Soc Probl 21:423–437

Sherman LL (1974) Becoming bent: moral careers of corrupt policemen. In: Sherman L (ed) Police corruption. Anchor Press, Garden City

Sherman LL (1977) Police corruption control. In: Bayley D (ed) Police and society. Sage, Thousand Oaks
Sherman LL (1978) Scandal and reform. University of California Press, Berkeley
Stoddard ER (1974) A group approach to Blue-Coat crime. In: Sherman LW (ed) Police corruption: a sociological perspective. Anchor Press, Garden City

Police Culture

Eugene A. Paoline III
Department of Criminal Justice, University of Central Florida, Orlando, FL, USA

Synonyms

Occupational culture; Organizational culture; Police styles; Police subculture

Overview

The job of a police officer can be both mentally and physically exhausting. As such, officers rely on one another for emotional and physical support. Terms like *brotherhood*, *thin blue line*, and *blue code of silence* are common illustrations of the cultural bond that officers share. Most often connotations of police culture (or police subculture to some) are negative. For example, when officers misuse their coercive power, it is often police culture that gets blamed (for endorsing a no nonsense aggressive approach to dealing with citizens). When organizational reforms, like community policing, are met with officer resistance, the culture is cited as the primary barrier to soft policing innovations brought on by "out of touch" administrators. At the same time, police culture can operate as a powerful positive mechanism in helping officers mutually buffer the various tensions of their job.

Unfortunately, a great deal of vagueness comes with the concept of police culture. That is, if you asked people to define and explain it, you are liable to receive 100 different responses from 100 different people. The most popular depiction of police culture views it as an occupational phenomenon shared by all police no matter where (or for whom) they work. As such, there is more of a focus on the commonalities across police personnel over the differences. Even among those who endorse the occupational view, there is a great deal of variation in describing the facets of police culture. Cumulatively, this is problematic because it limits our ability to fully understand, and comprehensively research, this important aspect of policing.

In illustrating the ways of conceptualizing police culture, consider the occupation of criminal justice professor. Undoubtedly, across the various universities that house professors, there are commonly shared cultural ways of handling the work-related strains created by students as primary clientele (e.g., class size and preparation, advising, grade appeals, plagiarism) and administrative supervisors (e.g., expectations for peer-reviewed publications, procuring funded research, service). At the same time, it is reasonable to expect that such demands, and subsequent ways of coping, vary depending on the type of institution where one works (e.g., large research-oriented university versus small private teaching college). Moreover, these strains might be met differently based on the working style of the professor (e.g., pro-research versus pro-instructor). Even among criminal justice professors housed within the same institution, concerns and demands might be interpreted differently across the ranks of assistant, associate, and full professor. Relatedly, for those in the upper tiers, cultural variation might also be a function of changing assignments (e.g., research coordinator, departmental chair, associate dean, dean). One's background (e.g., gender, race, educational experiences) might also work to produce differences in the ways in which criminal justice professors deal with the pressures of their work. If one could effectively tease out the various influences, across these divergent conceptualizations, it would undoubtedly produce a deeper comprehension of criminal justice professor culture(s). A similar approach is taken in this chapter in understanding police culture.

The Occupational Culture of Policing: The Monolithic Model

The foundation for understanding police occupational culture dates back to the seminal work of William Westley (1970) who, in the 1950s, conducted one of the first sociological studies of policing. In doing so, Westley highlighted important themes of the police occupation. More specifically, he found that the police he studied were a secretive tight-knit group that often faced violence on the streets in performing their duties. In detailing these informal aspects of policing, Westley was painting the first picture of its kind. In essence, the message delivered was that cops believed they have to stick together in their interactions against potentially dangerous and hostile citizens.

The other significant contribution to our initial impressions of police culture was the research of Jerome Skolnick (1966), who explained how the occupation actually worked to shape a distinct police personality. Similar to Westley (1970), Skolnick detailed life on the streets for the police officer as a dangerous endeavor, one where officers wield a tremendous amount of authority over citizens. Skolnick also tapped into a third primary feature that helped form the police personality, the need to appear efficient to superiors, which introduced an equally dangerous and hostile work environment – the police organization.

Both of these ethnographic accounts of police work highlighted many of the informal aspects of the job while also providing a loosely defined template for scholars in comprehending the occupational culture of police. More specifically, these two studies started the process of thinking about the ways in which police officers *collectively cope*, *via their attitudes*, *values*, *and norms*, *with the strains created by their work* – which is precisely what defines an occupational culture (Paoline 2001). Unfortunately, unlike foundational explanations of crime causation that tended to clearly (and parsimoniously) posit sets of interrelated propositions, full accounts of police occupational culture have been as vague as the concept of love. That is, you tend to know it when you see it but can mean different things to different people. From a theoretical and empirical standpoint, this can be both confusing and frustrating.

The works that followed Westley (1970) and Skolnick (1966) described and analyzed several different components of culture that tapped various dimensions of the street and organizational environments. What was lacking though was a clear comprehension of how these individual elements, that research identified, contributed to an overall understanding of the police culture process. That is, what exact features are police dealing with in their primary work environments that are in need of occupational collectiveness? In addition, what are the so-called collective responses that officers utilize to manage the strains of these environments? Finally, what are the consequences of the strains of the work environments and the coping mechanisms that are used to deal with such concerns? While the elements for such an understanding were present in the volumes of research (e.g., an explanation of police loyalty, the crime fighting mandate, citizen danger), a concise roadmap of these correlates, in explaining exactly what a police culture entails, was not readily existent. The monolithic model of police culture, presented by Paoline (2003), helped clarify these concerns.

Paoline's (2003, p. 201) description of the monolithic occupational culture of policing presents, based on extant research, a way to conceptualize the stressful factors of the work environments that operate to produce collective coping mechanisms and outcomes. In a path-like presentation, he identifies interactions on the street with citizens (i.e., *occupational*) and those with supervisors in the department (i.e., *organizational*) as the two primary work environments of the police. As a starting point for understanding what produces the coping mechanisms of the monolithic culture, the occupational street environment is described as *dangerous*, with the defining mandate to display one's *coercive authority* over citizens. The model also details the organizational environment as being characterized by uncertain *supervisor scrutiny* of police decisions (i.e., watchful and punitive superiors) and *role ambiguity* whereby officers are expected to perform all police functions equally yet really only recognized for crime fighting duties. Both the occupational and organizational environments produce equally intense stress and anxiety that is relieved through the collective coping mechanisms found within the occupational police culture.

In illustrating the second principal feature of the monolithic police culture model, Paoline (2003) identifies *suspiciousness* (in dealing with danger) and *maintaining the edge* (in properly displaying coercive power) as the primary coping mechanisms that officers employ in handling the strains created on the street with citizens. In terms of dealing with supervisors within the equally hostile organizational environment, the culture prescribes that officers *lay low/c.y.a.* (from watchful and punitive superiors) and embrace the *crime-fighter orientation* (in minimizing role ambiguity). These prescriptive coping mechanisms of the police occupational culture are transmitted across officers via a socialization process that begins in the training academy and continues throughout one's career (Van Maanen 1974). The final stage of the model highlights the consequences of the strains of the work environments and the coping mechanisms prescribed by the culture, which include a *socially isolated* occupational group that is extremely *loyal* to one another.

Collectively, the work that helped produce this model paints a caricature sketch of police as a socially isolated group that are distrustful and suspicious of their primary clientele, as they continually attempt to maintain the upper hand in utilizing their coercive authority. Moreover, officers approach police work solely in crime fighting terms while laying low from supervisors, choosing to only trust their immediate peers. While this is the dominant portrayal of police culture, other research points to important ways in which the occupational group may be fragmented.

Sources of Variation in Police Culture

The preceding discussion regarding *a* police culture rests on the assumption that the strains that police face on the streets and within the department are the same across the United States and therefore officers' responses should be similar as well (Crank 1998). By contrast, there are other lines of research that cast doubt on the notion of homogeneity among police by pointing out important sources of cultural variation and segmentation.

Organizations

Although often used synonymously, occupational and organization cultures are not the same phenomenon. As opposed to a universally shared set of responses to the internal and external strains of the job (for all police) that make up occupational accounts of culture, organizational cultures represent ways that police deal with specific concerns from the various places where they are housed. A principal difference between occupational and organizational accounts of police culture lies within the locus of influence. Occupational cultures are usually formed and maintained by lower level personnel (Van Maanen and Barley 1984), while organizational cultures are created by upper level management and imposed downward through the ranks (Schein 1992). Irrespective of such nuances, this suggests that police organizations, embedded within the overall occupation, exert cultural influence over police officers. James Q. Wilson's (1968) seminal study accentuates such points.

Wilson's (1968) examination of organizational culture, which he believed was indicative of a departmental style, was based on research conducted in eight communities. Wilson identified three organizational styles that differed in terms of their priorities toward core role orientations of enforcing the law and maintaining order. The author asserted that it was the top police administrator that defined the given style of the department based on their interpretation of the primary needs of the community being served.

According to Wilson (1968), some departments are situated within urban environments characterized by greater crime concerns, and thus the agency will embrace more of a formal crime fighting detached approach to police work. In such *legalistic* style departments, arrests and tickets will be more frequent compared to other organizational styles. By contrast, Wilson's *watchman* style departments are more common in low crime rural areas, focusing primarily on maintaining public order. Watchman style departments are less likely to formally respond to citizen transgressions unless they are serious in nature. Wilson asserted that the relational distance between the police and the public in watchman areas would be much smaller than that found in legalistic style departments. Finally, *service* style departments are described by Wilson as those where law enforcement and order maintenance is not an overall priority, but would be handled as needed. These departments focus on providing assistance and are more likely to be situated in suburban areas with less crime and disorder. Wilson explained that service style departments are closely connected with citizens, choosing to intervene frequently (when needed) but not formally (i.e., arrests and tickets).

Wilson's (1968) research suggests that organizational environments vary, and as such, it highlights the fact that individual police responses to different work conditions might also vary. Interestingly, the legalistic style department he identified comports closely to the organizational environment found in explanations of the monolithic occupational culture. By contrast, the watchman and service style organizational environments represent vastly different internal work arenas where intense supervisor scrutiny and role ambiguity (noted in the occupational account) would be much less common. Likewise, Wilson's work also points out rather stark variation in the occupational/street environments, as watchman and service style areas are not characterized as overly hostile and crime ridden. As such, cumulative coping mechanisms where officers are suspicious of (and maintain the edge over) citizens, strictly endorsing the crime-fighter image, while covering their ass from supervisors might be functional in a legalistic style department but certainly would be out of place (and probably not tolerated) in both watchman and service style agencies.

Recent empirical inquiries have provided additional support for the ways in which policing functions and philosophies differ across organizations embedded in urban, rural, and suburban contexts (Crank 1990; Liederbach 2005). This line of research casts doubt on occupational accounts of culture that suggest that all organizations are the same while also acknowledging that management can impact police culture.

Rank

Occupational accounts of police culture focus heavily on the homogeneity of officers' attitudes, values, and norms. The socialization process transmits the culture across occupational members and is often described as intense, especially for new personnel (Van Maanen 1974). Because all police at some point in their career are assigned to patrol functions, the supposition is that culture originates at the lower ranks. This is understandable given that patrol officers are those most likely to deal with citizens on the streets and with supervisors in the department. What is much less clear is the role that culture plays in officers' lives once they move beyond entry level positions. Does the occupational culture still buffer the strains of the work environments for sergeants, lieutenants, captains, majors, deputy chiefs, and chiefs the same way(s) that it did when these officers were assigned to patrol, or does culture change by rank? The works of Reuss-Ianni (1983) and Manning (1994) help answer such questions.

Reuss-Ianni (1983), based on research conducted in the NYPD, asserts that there are two distinct cultures in policing. The *street cop culture*, which Reuss-Ianni delineates in a series of codes, concentrates on the "here and now" in policing and embodies many of the values outlined previously as part of the police occupational culture. Street cops are found at the patrol level and tend to rely on their developed craft, as well as their loyal peers, in surviving on the street and controlling crime at the local level (i.e., beat and precinct). By contrast, *management cop culture* focuses on city-wide long-term concerns (e.g., crime control, citizen responsiveness, organizational efficiency), taking into account political, social, and economic factors. Importantly, Reuss-Ianni believes that temporal changes in the dynamics of policing (e.g., officer composition, resource competition, accountability concerns) have contributed to a weakening of the street cop culture of the "good old days" when organizational leaders could still utilize the codes of the street in running police departments.

The importance of Reuss-Ianni's (1983) work lies in her recognition that rank contributes to differences in police culture. This makes sense given that duties and concerns do change as one moves upward from patrol with the organization. This might also help explain some of the tension noted between officers and their supervisors in the occupational account of police culture. While it is clear that street cop culture originates and protects occupational members at the patrol level, the author fails to inform the reader exactly when police enter management cop culture. That is, it is easy to deduce that top commanders are embedded in the management cop culture, but what about ranks like sergeants and lieutenants (i.e., middle managers), who are in organizational limbo between the very bottom and the very top. Manning's (1994) work helps alleviate such concerns.

Manning (1994) agrees that police culture is hierarchically segmented, although he differentiates rank in a three-tiered manner based on a series of themes and meta-themes. At the first tier are *lower participants* (i.e., patrol and street sergeants) whose culture, like Reuss-Ianni's (1983) street cops, focuses on the immediate aspects of "real" police work. The author's second tier of culture focuses on *middle management* (i.e., some sergeants up to department brass) who emphasizes supervisory control but also concentrates heavily on buffering concerns of line members of the street (i.e., the first tier) and top police officials (i.e., the third tier). Finally, *top command* culture (i.e., commanders, superintendents, deputy chiefs, chiefs) represents the third tier. Here, there is a concentration on dealing with the politics of running a police department internally while also buffering the organization from external audiences.

Interestingly, in terms of middle managers, Manning (1994) points out that sergeants assigned to street functions can rest within the first tier (i.e., lower participant culture), while middle managers with other responsibilities beyond patrol may be part of the second tier. Here, the author allows for behavioral freedom for sergeants who, as supervisors, might still align more with the line level versus management. Although not addressed by Manning, the same argument can be made for first tier lower participants, who depending on their assignment (or aspirations) might differentially align with police culture(s). For example, those who have jobs with duties that are based on more traditional aggressive crime fighting mandates (e.g., criminal investigations, tactical, K-9, anti-street crimes, SWAT, emergency response units) might be more apt to orient themselves with the lower participant culture. By contrast, those in positions that focus more on long-term solutions to controlling crime with increased officer participation (e.g., crime prevention, community outreach, school resource, community policing, victim advocate, homeland security, tourism units) might orient themselves more like upper ranking officials.

Both Reuss-Ianni (1983) and Manning (1994) highlight the dynamic nature of police culture as changing and adapting to insulate group members' issues and concerns unique to their position in the organizational hierarchy. As such, their work suggests that police culture is not monolithic, but instead comprised of multiple cultures based on rank.

Style

Yet another way that police culture can be segmented is by the style of the officer. In a series of studies published in the 1970s, researchers constructed typologies of the police based on their orientations toward various features of their occupational/street and organizational environments (Broderick 1977; Brown 1988; Muir 1977; White 1972). Of interest is the fact that these researchers were utilizing different dimensions to construct their typologies and were working independent of one another across various departments, areas, and time – yet they concluded with almost identical types/styles of officers. This was evidenced by Worden's (1995) synthesis of American police typology research into five distinct officer styles. Reiner (1985) also summarized typology research, utilizing some of the studies that Worden (1995) included (i.e., Broderick 1977; Muir 1977), although he incorporated work done on British and Canadian police. Interestingly, both authors concluded with very similar syntheses of officer types.

Worden (1995) provides a thumbnail sketch of the five primary styles of police officers. Worden's *tough-cop* approaches police work in a cynical fashion, often conflicting with citizens and supervisors in their aggressive approach to selectively fighting serious crime. These officers do not want to be bothered with trivial police matters (e.g., maintaining public order or providing service), and they place a premium on their street experience over formal college education. This style of officer would be your stereotypical gruff "cop's cop" that is often portrayed in television and movies, as one that is not afraid to push (or perhaps even exceed) the limits of police power if deemed necessary.

A second style, *clean-beat crime-fighter*, resembles the tough cop in their cynical approach, strong orientations toward crime fighting, and conflicting relationships with supervisors. What separates this style of officers from the former is their undying pursuit of enforcing all laws, not just the serious ones (in keeping a clean beat), and their belief that police should follow the procedural rules (i.e., not violate citizen rights).

Avoiders are described as a cynical group that, as their name implies, avoid as much work as possible. This style of officer has a very narrow approach to police work, as they take "laying low" to extremes. Their orientations toward all aspects of the job are rather detached, as they attempt to just do their time.

Problem solvers are those officers with very favorable attitudes toward citizens and the service style of policing. Problem solvers are the group that is least oriented toward crime fighting and aggressive policing tactics. As opposed to traditional ways of handling situations, problem solvers focus on outcomes, desiring to see the problems they are called upon to deal with through to their resolution. To many traditional police officers, the approaches endorsed by the problem solver would appear as "soft."

Professionals embody the values of the professional reform movement and are characterized as the most positive of the five policing styles. Officers in this group hold favorable orientations toward all aspects of their street and organizational environments. Professionals are portrayed as the most well-rounded group in terms of performing multiple functions (i.e., broad role orientation) in a manner in which citizens are not treated aggressively and where supervisors are pleased.

Police typology research illustrates that there are a variety of ways in which officers cope, via their style, with the demands of their job. Interestingly, Worden's (1995) tough-cop orientation comports closely to the attitudes and values associated with occupational accounts of culture, as well as Reuss-Ianni's (1983) street cop culture and Manning's (1994) lower participant culture. The other four styles summarized by Worden suggest the type of ideological differentiation that forms the basis for occupational subcultures (Van Maanen and Barley 1985). While research has yet to empirically validate these working styles, one thing is clear – there is more than just a single way for officers to deal with the strains of policing.

Officer style research illustrated cultural fragmentation among police during a time when officers were relatively demographically homogenous (i.e., White males with a high school education). Changes in the overall composition of police (i.e., more females, non-Whites, and college educated), as well as policing philosophies (i.e., community policing), in recent decades prompted reexaminations of police culture(s) (Paoline et al. 2000). Moreover, recent empirical approaches have utilized advanced statistical classification techniques, such as cluster analysis and discriminant function analysis, to examine more than the two or three attitudinal dimensions that were used for previous police typology inquiries (Cochran and Bromley 2003; Jermier et al. 1991; Paoline 2001). Similar to the research that produced evidence of policing styles, studies that utilized quantitative classification schemes also found multiple groups of officers with varying alignment to a single occupational culture. Because much of this (and typology) research is based primarily on officers' attitudinal orientations, it begs the question as to the applicability of such approaches on explanations of officer behavior. The age old attitude-behavior link has stymied social scientists for decades, as empirical connections have failed to verify intuitive expectations that one's attitudes affect one's behavior (Frank and Brandl 1991). Paoline's (2001) classification scheme provides one of the few exceptions.

Paoline (2001) utilized survey data collected as part of the Project on Policing Neighborhoods (POPN) to analyze several prominent attitudinal features of prior police culture research. Paoline identified seven groups of officers with varying cultural orientations. Five of the groups very closely resembled those Worden described in the framework he presented in 1995: tough cops (i.e., *traditionalists*), clean-beat crime-fighters (i.e., *law enforcers*), avoiders (i.e., *lay lows*), problem solvers (i.e., *peacekeepers*), and professionals (i.e., *old pros*). The other two officer groups were *anti-organizational street cops*, who were distinguished in terms of their strong negative views of supervisors and their strong favorable views of citizens, and *dirty harry enforcers*, who were distinguished in terms of their strong beliefs in aggressively fighting crime and disorder, even if it meant violating the rights of citizens.

Although Paoline's (2001) seven groups were attitudinally and empirically distinguishable from one another, there was no genuine way to rank order them (for comparison purposes) in terms of a continuum of police culture commitment. The groups did provide the author with a conservative manner in which to categorize them in terms of positive orientations toward culture (i.e., pro culture), negative outlooks or antithesis of cultural expectations (i.e., con culture), and those that fell somewhere in the middle of the two extremes (i.e., mid culture). Paoline's trichotomized classification scheme of police has been utilized to empirically examine the connection between cultural attitudes and theoretically related behaviors of citizen searches (Paoline and Terrill 2005) and the use of force (Terrill et al. 2003). In both instances, statistically related differences were noted between officers' cultural alignment and their behavior. More specifically, those who held attitudes that were strongly congruent with the traditional police culture (i.e., pro culture) or mixed (i.e., mid culture) searched citizens more often during traffic stops and used force more frequently (and at higher levels) than those who ardently resisted cultural attitudes (i.e., con culture).

Background Characteristics

Besides the organization, rank, and policing style, officer background represents another potential source of cultural variation. As previously mentioned, the foundation for understanding the monolithic occupational account of police culture was built during a time when policing was largely a demographically homogenous occupation (i.e., made up almost exclusively of high school-educated White males). Researchers have detailed the integration struggles and resistance that female (e.g., Martin and Jurik 1996), non-White (e.g., Bolton 2003), and gay/lesbian (e.g., Miller et al. 2003) officers have faced in a male-dominated masculine occupation. To a lesser extent, college-educated officers have been chided for their "book smarts" over the preferred "street smarts," in learning the craft of policing (Fielding 1988). The idea here is that those with different backgrounds from the modal officer are excluded from the broader occupational culture.

As departments continue to diversify their personnel in terms of their individual characteristics (Hassell and Brandl 2009), it is certainly reasonable to expect that these previously excluded policing members will, in some manner, contribute to a less cohesive occupational group. To date, we lack a concrete understanding of how such advancements in officer diversity affect police culture(s). That is, despite qualitative (and detailed case study) accounts of cultural dissociation, background characteristics of police provide limited power in explaining cultural perceptions among police (Paoline et al. 2000). In fact, the impetus for some of the recent officer classification schemes, in revisiting working styles, was built on the notion that diversification of personnel should produce differences in cultural alignment (Paoline 2001). Even among these studies that utilize advanced statistical techniques with larger sample sizes, there is little (if any) systematic connection between officer background characteristics and their group membership. In the end, such work suggests that cultural variation may have little to do with one's background.

The lack of a strong empirical connection between officer characteristics and alignment with police culture(s) could be a function of a couple of factors. It could be that demographic changes that are occurring are still too recent, and thus, we have yet to see the full cultural effects of such transitions in membership. A second explanation could be that the forces of cultural socialization as so strong, as some traditionalists would assert (Van Maanen 1974), that individual differences (via one's sex, race, education, sexual orientation, etc.) are washed away as officers collectively deal with the strains of the occupation. As such, the primary environments of policing are what shape the culture or subcultural styles, irrespective of what officers bring to the job.

Conclusion and Future Research

Police culture has been a topic of study for over half a century. Although the monolithic depiction still tends to dominate popular conceptions, this chapter illustrates that police culture is a multidimensional concept that is molded and shaped at various levels (i.e., organization, rank, style, assignment, background). Police scholars, in attempting to understand this police phenomenon, have wrestled with such complexities. At the same time, police culture is not just an academic concept, as practitioners recognize its power, vagueness, and malleability. As noted by the Community Relations Service of the United States Department of Justice's (2002, p. 9) handbook on "Police Use of Excessive Force," in order to change police culture, "one must analyze and understand the currently existing culture." While this chapter summarizes many of the advancements in detailing the pathways and contours of police culture, much more work is needed. As Fielding (1988, p. 185) accurately contends "if occupational culture is to serve as an empirically satisfactory concept as well as theoretically necessary one, the sense of its internal variations and textures must be brought out in the same fashion as have conceptions of culture in relation to delinquency." What follows are a few suggestions for future research inquiries.

Ideally one would want a totally comprehensive study, which would require starting at the occupational level and working through differences between and among organizations, ranks, styles, assignments, and backgrounds. Simply put, this may be asking way too much at this point. A more reasonable request starts by considering the various sources of influence on police culture.

The monolithic account of *the* police culture was created based on early ethnographic studies interested in capturing themes across officers. It is not surprising then that cultural similarities were the focus over cultural segmentation. Even if we acknowledge the sources of cultural variation, typology (and recent officer classification scheme) research did find one style of officer that deals with the strains of the primary work environments in the same manner presented by monolithic characterizations. Paoline (2003) presented a model for understanding this monolithic version of culture, and while parts of it have been researched, the entire model has yet to be empirically tested. As a logical starting point, it would be interesting to gather baseline data regarding the extent to which this traditional culture model is endorsed among contemporary police. This is especially salient given that policing philosophies are revisiting more bottom line crime fighting goals (e.g., COMPSTAT) at the cost of many community policing initiatives that have permeated agencies for the last 20 years. At the occupational level, we still do not have a firm empirical grip on the things that officers might share versus the exact points where they might be divided.

Regarding the influence of the organization, researchers should continue efforts aimed at examining departmental style as it impacts police culture(s). In doing so, empirical work can be done to deduce whether organizational environments, via agency style, are similar or dissimilar from that noted by Wilson (1968). For example, the rural and suburban areas that made up the smaller watchman and service departmental styles might be organized and operating in different ways today as criminogenic forces (e.g., gangs, unincorporated areas for producing and distributing drugs) are not just reserved for legalistic urban areas. As such, the directives and goals espoused as part of today's watchman and service organizational styles may look more similar to the legalistic style than that noted in the 1960s.

With respect to officer styles, efforts should be made to examine the connected (or disjointed) way(s) in which individual level approaches to dealing with the strains of policing operate within organizational level, top command-driven, styles. Jermier et al. (1991) found differences between subcultural groups of officers and the legalistic style crime fighting official organizational culture. This begs the question – what about other departmental styles and potential organizational survival of officer styles? For example, the crime fighting approaches of Worden's (1995) "tough cop" and "clean-beat crime-fighter" would certainly seem out of place compared to a "problem solver" or "professional" in a service style suburban department, as opposed to a legalistic style organization where the former would be more of a fit than the latter.

Regarding rank-related sources of variation, disentangling officer assignment and formal designation within the police hierarchy would be useful in understanding how culture(s) differ across Manning's (1994) three primary tiers. For example, can lower level participants, with assignments that call for increased officer participation and long-term orientations, align more like middle managers and their culture than their own? Conversely, can captains that are assigned to aggressive crime fighting street duties orient themselves more like lower level participants? In addition, do officer styles differ across the three tiers or is the culture monolithic by rank? If policing styles, as proxies for subcultures, develop at the patrol level, how and when do they change as one advances through the organizational hierarchy? In what way(s) do supervisory styles impact the development and maintenance of subordinate styles? These are but a few questions whose answers would add greatly to our police culture knowledge base.

Finally, despite the lack of consistent statistical associations, the demographic changes that have occurred in policing should continue to be part of police culture research. In doing so, racial (and ethnic) groups other than dichotomous White and non-White (predominantly comprised of African Americans) classifications should be part of such inquiries (e.g., Hispanic, Asian, Pacific Islander, American Indian, Alaskan Native). As such, previously excluded groups continue to populate police organizations, especially above the patrol level, cultural (and sub cultural) differences might be more visible. The same holds for those with divergent educational backgrounds and sexual orientations that differ from the traditional blue-collar aggressive crime fighting officer depicted in characterizations of the police culture "brotherhood." Moving beyond descriptive accounts of the struggles of those with divergent backgrounds and focusing on quantitative multivariate modeling will enhance our understanding of the potential independent effects of such factors on police culture(s).

The individual pieces of the multidimensional puzzle have been presented here in an attempt to detail the various intricacies of police culture. If scholars, practitioners, and funding agencies are serious about the overall importance of understanding (and possibly changing) police culture, it is high time to devote the empirical effort required to fully comprehend the way(s) in which officers deal with their work environments.

Related Entries

▶ Communities and the Police
▶ Community Policing
▶ COMPSTAT
▶ Control of Police Misconduct
▶ History of the Police Profession
▶ Managing Innovation of Policing
▶ Minorities Within the Police Workforce
▶ Police and the Excessive Use of Force
▶ Police Corruption
▶ Police Discretion in Law Enforcement
▶ Police Integrity
▶ Police Legitimacy and Police Encounters
▶ Policing the Police
▶ Role and Function of the Police
▶ Women in Policing

Recommended Reading and References

Bolton K Jr (2003) Shared perceptions: black officers discuss continuing barriers in policing. Pol Int J Police Strateg Manag 26:386–389

Broderick JJ (1977) Police in a time of change. General Learning Press, Morristown

Brown MK (1988) Working the street: police discretion and the dilemmas of reform, 2nd edn. Russell Sage, New York

Cochran JK, Bromley ML (2003) The myth(?) of the police sub-culture. Pol Int J Police Strateg Manag 26:88–117

Crank JP (1990) The influence of environmental and organizational factors on police style in urban and rural environments. J Res Crime Del 27:166–189

Crank JP (1998) Understanding police culture. Anderson Publishing, Cincinnati

Fielding NG (1988) Joining forces: police training, socialization, and occupation competence. Routledge, London

Frank J, Brandl SS (1991) The police attitude-behavior relationship: methodological and conceptual considerations. Am J Police 58:83–103

Hassell KD, Brandl SS (2009) An examination of the workplace experiences of police patrol officers: the role of race, sex, and sexual orientation. Police Q 12:408–430

Jermier JM, Slocum JW Jr, Gaines J (1991) Organizational subcultures in a soft bureaucracy: resistance behind the myth and facade of an official culture. Organ Sci 2:170–194

Liederbach J (2005) Addressing the elephant in the living room: an observational study of the work of suburban police. Pol Int J Police Strateg Manag 28:415–434

Manning PK (1994) Police occupational culture: segmentation, politics, and sentiments. Unpublished manuscript, Michigan State University

Martin SE, Jurik NC (1996) Doing justice, doing gender. Sage, Thousand Oaks

Miller SL, Forest KB, Jurik NC (2003) Diversity in blue: lesbian and gay police officers in a masculine occupation. Men Masc 5:355–385

Muir WK Jr (1977) Police: streetcorner politicians. University of Chicago Press, Chicago

Paoline EA III (2001) Rethinking police culture: officers' occupational attitudes. LFB Publishing, New York

Paoline EA III (2003) Taking stock: toward a richer understanding of police culture. J Crim Justice 31:199–214

Paoline EA III, Myers SM, Worden RE (2000) Police culture, individualism, and community policing: evidence from two police departments. Justice Q 17:575–605

Paoline EA III, Terrill W (2005) The impact of police culture on traffic stop searches: an analysis of attitudes and behavior. Pol Int J Police Strateg Manag 28:455–472

Reiner R (1985) The politics of the police. St. Martin's Press, New York

Reuss-Ianni E (1983) Two cultures of policing. Transaction, New Brunswick

Schein EH (1992) Organizational culture and leadership. Josey Bass Publishers, San Francisco

Skolnick JH (1966) Justice without trial: law enforcement in democratic society. John Wiley, New York

Terrill W, Paoline EA III, Manning PK (2003) Police culture and coercion. Criminology 41:1003–1034

U.S. Department of Justice (2002) Police use of excessive force: a conciliation handbook for the police and the community. Community Relations Service, Washington, DC

Van Maanen J (1974) Working the street: a developmental view of police behavior. In: Jacob H (ed) The potential for reform of criminal justice. Sage, Beverly Hills, pp 83–130

Van Maanen J, Barley SR (1984) Occupational communities: culture and control in organizations. In: Staw BM, Cummings CC (eds) Research in organizational behavior, vol 6. JAI Press, Greenwich, pp 287–365

Van Maanen J, Barley SR (1985) Cultural organization: fragments of a theory. In: Frost PJ, Moore LF, Louis MR, Lundberg CC, Martin J (eds) Organizational culture. Sage, Beverly Hills, pp 31–53

Westley WA (1970) Violence and the police: a sociological study of law, custom, and morality. MIT Press, Cambridge, MA

White SO (1972) A perspective on police professionalization. Law Soc Rev 7:61–85

Wilson JQ (1968) Varieties of police behavior: the management of law and order in eight communities. Harvard University Press, Cambridge, MA

Worden RE (1995) Police officers' belief systems: a framework for analysis. Am J Police XIV:49–81

Police Discretion and Its Control

Gary Cordner[1] and Michael S. Scott[2]
[1]Department of Criminal Justice, Kutztown University, Kutztown, PA, USA
[2]Law School, University of Wisconsin, Madison, WI, USA

Overview

Police officers exercise a tremendous amount of discretion in carrying out their functions. That is, they make many choices from a range of possible actions or inactions available to them, and their choices are not specifically prescribed by law, policy, procedure, or rule. This simple notion, which seems self-evident to some and controversial to others, lies at the heart of many issues of policing in democratic societies.

That the police do exercise discretion was only recognized and openly acknowledged beginning in the 1960s (J. Goldstein 1960; H. Goldstein 1963; LaFave 1965; Davis 1969). The conventional view prior to that time, and persisting among some long thereafter, was that the police function was entirely a ministerial one: that the police only took actions that were specifically authorized or mandated by legislative bodies. Under this view, policing was understood to be simply a matter of enforcing the laws on the books. But a number of pioneering research studies of policing in action found that the law was silent on many important matters involving police action and ambiguous on others, and police officers did not always adhere to what the law prescribed, even where the law was clear and specific (Banton 1965; Skolnick 1966; Wilson 1968; Westley 1970; Reiss 1971; Brown 1981).

The exercise of discretion is nearly inevitable in policing (H. Goldstein 1977). Some laws are practically unenforceable because they are outdated, or widely unpopular, or unconstitutional, or lack enforceable sanctions. Legislatures pass laws, and fail to abolish others, for a variety of purposes, only one of which is to establish clear expectations and guidelines for enforcement through police action. Moreover, the police seek to achieve various objectives in carrying out their duties, and at times, those objectives conflict with one another. In such instances, the police must decide which objectives take precedence over others. For example, during a public demonstration held in the streets, the police may find that the objective of keeping traffic avenues clear conflicts with the objective of safeguarding citizens' rights to peaceful protest.

Even absent conflicting objectives, police and other criminal justice resources are far too limited to allow the police to enforce all laws exhaustively. Most communities would not tolerate full enforcement of the law, even if resources would allow it, preferring a degree of police tolerance, especially for minor legal transgressions. The very capacity of the criminal justice system to continue functioning in many communities depends to a great extent on the police not fully enforcing the law. Sudden increases in police arrest activity can seriously challenge the capacity of the legal system to process the resultant cases.

Discretion is exercised in policing at all levels of the police hierarchy. In contrast with other occupations and professions, the greatest amount of discretion in policing is exercised at the line level by patrol officers and detectives, but supervisors and policy makers also exercise large amounts of discretion.

Fundamentals of Police Discretion

Perhaps the most profound types of discretionary decisions made in policing are the decisions to use force and to arrest. In regard to the use of force, Bittner (1970: 46) famously described the core of the police role as "the distribution of non-negotiably coercive force employed in accordance with the dictates of an intuitive grasp of situational exigencies." What Bittner meant is that the police are regularly called to the scenes of trouble and sometimes find it on their own. These situations are often confusing, if not chaotic. In these situations, police officers are authorized to use reasonable force, if necessary, in order to quell the trouble. The decision about whether the use of force is necessary, and if so, how much force to use, is discretionary. The decision is made by the officer or officers on the scene, based on their "intuitive grasp" of the circumstances (the "situational exigencies") as they unfold.

Police decisions about whether to arrest are discretionary too. Generally, the police are authorized to arrest (1) with a warrant issued by a court, (2) whenever a person commits any crime in their presence, and (3) whenever they have probable cause to believe that a felony (serious) crime has been committed and that a particular person has committed it, even if the probable cause is based on secondhand or thirdhand information. In regard to warrantless arrests (types 2 and 3 above), which are the most commonplace arrest situations, law and official policy rarely say that the police *must* arrest. Most often the law says, or is interpreted as meaning, that the police *may* arrest, leaving discretion in the hands of ordinary police officers.

Perhaps the most ubiquitous discretionary situations encountered by the police are the decisions whether to stop drivers who violate traffic laws and, when traffic stops are made, whether to cite the drivers and, further, to search the vehicle for evidence or contraband. Many observed traffic violations are minor in nature (such as driving 1 mile over the speed limit) and therefore frequently ignored by the police or, at most, addressed with a warning. By the same token, poor driving causes many deaths and injuries every year and a huge amount of property loss, so traffic policing, and its associated discretionary decision making, should not be dismissed as a trivial matter. As well, when officers decide to issue citations to drivers, fines are routinely in the hundreds of dollars today and can result in the loss of driving privileges, which in turn can have serious consequences for employment and income. Moreover, the police commonly engage in traffic enforcement as a method of more serious crime control: to locate illegal guns and drugs, to locate persons with outstanding arrest warrants, to disrupt illegal drug and prostitution markets, to deter burglars and thieves from prowling neighborhoods, and so forth.

There exists some research data about the frequency of these three major discretionary situations – use of force, arrests, and traffic enforcement. Clearly, the police frequently choose not to use force, or have no justification to do so – estimates are that the police use physical force (beyond handcuffing) in only about 20 % of all arrests (Terrill et al. 2008), in only 2–3 % of all encounters with the public (Friedrich 1980), and in less than 1 % of all traffic stops (Eith and Durose 2011). Similarly, the police often choose not to arrest – one study estimated that the police make arrests in about one-half of situations when it would be lawful to do so (Black 1980), and another study found that arrests are made in only 5 % of total police encounters with the public (Whitaker 1982). This still results in a large number of arrests though – 13 million in 2010, of which 2.2 million were for violent or property offenses, 1.6 million for drug offenses, and 1.4 million for drunken driving (FBI 2011). The picture is similar regarding traffic enforcement, with just over half (55 %) of 18 million drivers stopped in 2008 reporting that they were given a citation (Eith and Durose 2011).

Beyond use of force, custodial arrests, and traffic citations, the police make many other types of discretionary decisions, including decisions about which laws to invoke when an arrest or other form of detention is made (e.g., whether to add a charge of resisting arrest when an offender does not come along peacefully); whether to refer matters to other agencies; what tactics to adopt in mounting proactive operations; what conduct to investigate, what investigative techniques to apply, and how intrusive those techniques are; what level of resources to commit to various activities, places, and problems; whether to secure prior authorization for certain actions (e.g., whether to apply for search warrants or other court orders); what level of urgency to give to various duties; whether to grant permission for certain activities to take place (e.g., where the police are responsible for issuing parade permits); and so forth. Many of these discretionary decisions are made by patrol officers and detectives, but some are also made by supervisors and commanders.

Variations in Discretion

The range of choices open to police officers varies in different types of situations, as does the degree of latitude that they have in making those choices. So, for example, if an officer is dealing with a person experiencing a mental health crisis, it matters if there is a specialized unit that can respond and take over, or if there is a mental health crisis center to which the person can be taken, or a psychiatric unit within the local hospital. If none of those options is available, then the number of choices open to the officer is more limited – to perhaps not much more than curbside counseling or an arrest for disorderly conduct.

In some jurisdictions, while the law may authorize the police to arrest for certain offenses, that option may not really be available. For example, the county jail may be completely full, causing the jail administrator or sheriff to advise local agencies not to make any arrests except in extreme cases. In this situation, police officers may still have the option to issue a citation to an offender, but the option of physical or custodial arrest would be severely curtailed. In another scenario, it is common for prosecutors to inform police departments that, because of heavy caseloads facing the court, arrests should not be made for particular categories of minor offenses. This situation would discourage even the issuance of a citation, unless the accused had the option of paying a fine in lieu of trial, thus preventing any additional workload burden on the court.

Another real-world constraint on police discretion is the number of police officers available to handle other police matters should an officer take some action that removes him or her from the streets for an extended period. Whether in small agencies where there is only one police officer on duty or in larger agencies where police reinforcements are a long distance away, an officer might have to weigh the benefits of making an arrest against the costs of being tied up for several hours processing the arrestee, leaving no or few other officers to police the area. Both the public's and any remaining officers' safety must be taken into account. An added dimension could be that the county jail, hospital, or other facility is many miles away, building even more processing time into the detention and literally taking the officer out of the jurisdiction. These kinds of practical constraints may narrow the real range of choices available to officers when handling various situations, thereby further constraining their discretion.

In regard to the use of force, not all police officers and police departments have the same equipment available to them, which can affect their range of options in some situations. For example, conducted energy devices (e.g., Tasers) have become commonplace, but some officers do not have them and thus do not have one of the more effective less-lethal weapon options in confrontations with combative subjects. Similarly, when a high-speed pursuit is underway, the range of options is affected by whether the officers out ahead of the pursued vehicle have stop sticks or similar technology. Likewise, in a hostage or barricaded-person scenario, some police officers might have sophisticated communications, eavesdropping, or breaching technology available, while others might not, affecting the tactical choices open to them.

From a comparative perspective, the 18,000 chief police executives in the United States probably have broader discretion than their counterparts in most other countries, because of the highly decentralized structure of American policing. A US police chief has the authority to choose the type of pursuit-driving policy to implement, whether to place a higher priority on drug enforcement or traffic enforcement, whether to require applicants to have a college degree, how to address various types of policing incidents and problems, and so forth. In a country with a unitary (national) police system, these and other decisions are usually made at the national level, leaving much less discretion in the hands of local commanders. Sheriffs in the United States may have the broadest executive-level discretion of all, since they are locally elected and are primarily accountable to the electorate rather than to other government executives.

Differing legal systems may also affect police discretion. Within the United States, local and state police officers generally exercise wider discretion than do federal law enforcement officers, whose investigations, in particular, are guided to a larger degree by prosecutors (US attorneys) than is common at the state and local levels. Similarly, in countries that have civil law or inquisitorial legal systems, judges and/or prosecutors often direct investigations, reducing the degree of discretion available to police. In some countries, frontline police personnel (the equivalent of patrol officers in the United States or constables in the UK) do not have the same broad authority or discretion in the use of force, arrest, and investigation described in the previous section. The actual legal authority of ordinary police officers is more restricted, or organizational rules and procedures require them to get permission from superior officers before taking significant actions. Of course, it is also true that in some countries the authority of the police to detain, search, and interrogate is even broader (under the law or simply in practice) than discussed here, effectively increasing the amount of discretion available to those permitted to exercise the authority, whether it is frontline police personnel or superior officers.

Factors That Affect Police Use of Discretion

An important theme in police research has been the study of police decision making. Besides documenting the existence of discretion, as described above, many studies have looked for patterns and correlations in police decisions to use force, stop vehicles and pedestrians, make arrests, issue citations, and conduct searches. Factors affecting police decision making and the use of discretion by police officers fall into four general categories: situational characteristics, police officer characteristics, police organization characteristics, and neighborhood/jurisdiction characteristics.

Consistent with the finding that the police exercise broad discretion and also consistent with Bittner's (1970) argument that police behavior is guided by "situational exigencies," characteristics of the situations in which the police encounter the public have been found to have the greatest influence on police decisions to arrest, cite, and use force (Skogan and Frydl 2004). Among these, studies have found that the most influential are legal factors – the seriousness of the offense committed by the suspect, the suspect's previous offending history, the amount of evidence available, and the degree of threat posed by the suspect toward the officer. This finding – that police use of discretion is affected more by legal factors than any others – is reassuring. A related quasi-legal factor that also has some influence is the stated preference of the victim, especially when the victim asks the officer not to arrest the suspect.

Compared to these legal factors, a variety of extralegal situational factors have been found to have mixed and inconsistent effects on police use of discretion. These extralegal factors include the demeanor of the suspect and the suspect's race, sex, and social class. Each of these factors has been found to influence police decision making in some studies, but not always in the same direction, plus many other studies have found no effects. Of these factors, the one that has gotten the most empirical support, especially in more recent studies, is race (Kochel et al. 2011). Studies of so-called racial profiling consistently find that people of color are disproportionately stopped by the police, whether as drivers or as pedestrians, and once stopped are more likely to be searched. The key question, of course, is whether these police decisions to stop and search people of color are made based on the extralegal factor of the race of the person, or whether they are based on legal factors such as the seriousness of the person's behavior and the amount of evidence that is available. Unfortunately, research on racial profiling has not generally been able to answer that key question.

Police decision making and use of discretion might also be influenced by the personal characteristics of the police officer, such as his or her race, sex, and level of education. However, the cumulative findings from many studies indicate that these factors do not have large or consistent effects on police officer behavior. Similarly, neither police officers' general orientations toward their role nor their levels of authoritarianism, cynicism, or job satisfaction have been found to have any consistent influence on their use of discretion.

A third set of factors that might affect police officer decision making is organization-level characteristics. For example, officers in small police departments might use their discretion differently than officers in larger departments, officer decision making might vary between different types of agencies (e.g., municipal police departments versus sheriff's offices), or decision making might be influenced by the degree to which a police organization is centralized, specialized, or formalized. The internal climate or culture of a police organization might also lead to variations in police behavior, including decision making. While a good case can be made that organizational characteristics such as these ought to affect the use of discretion by organization members, the kinds of multiorganization studies that would be needed to verify such effects have simply not been conducted.

It has also been argued that police use of discretion varies according to characteristics external to the police, such as between rural, suburban, and urban communities, or between poor neighborhoods and affluent ones. Once again, a theoretical case can be made that community characteristics should influence police behavior, including decision making in discretionary situations, but relatively few studies have been able to incorporate multiple neighborhoods and/or multiple jurisdictions, and the results from those few studies have not been consistent (Varano et al. 2009). As was true for officer characteristics and organizational characteristics, community characteristics have not been found to have as much impact on police officers' use of discretion as situational characteristics, and within the latter, legal factors seem to have the greatest influence.

Consequences of Police Discretion

As discussed above, police discretion seems inevitable given the broad mandate of the police, the incredible variety of situations that the police encounter, the tendency of legislatures to pass laws that no one expects will be fully enforced, the limited capacity of jails and courts, and numerous other philosophical and practical considerations. So the police have broad discretion, though more in some situations than others. Also as noted, studies indicate that situational legal factors affect police officers' decisions more than any other factors, although future studies using better methodologies may reveal that extra-legal factors (such as the race of the suspect) and the characteristics of police organizations and communities have more influence than can presently be demonstrated.

Given the existence of police discretion, what are its consequences? The most obvious immediate result is that the police often choose not to enforce the law, thus softening the impact of the criminal law on people's lives. With discretion, officers get to choose how to handle various situations, and at least half the time when an arrest could be made or a citation issued, they choose under-enforcement or some other version of leniency. Because of discretion, officers' actions can be affected by sympathy and empathy, and officers have the opportunity to choose what they consider to be the most fair or just resolution of the situation, even if that entails not enforcing the law. Discretion enables the police to treat people *as people*, taking into consideration a person's circumstances, motives, intent, remorse, and promises – many would say that discretion allows the police to humanize the application of the criminal law.

This phenomenon of under-enforcement enabled by discretion might be a cause for more serious concern if it reduced the deterrent effect of arrest and punishment, emboldening potential offenders and creating a sense of impunity within society. Such a concern is less warranted though, since the police exercise their discretion mainly in response to minor crimes and disorders. It is generally the person caught drinking in public, driving 10 miles over the speed limit, or shoplifting a small item who benefits from an officer's decision to handle a matter informally, not the burglar, robber, or murderer. With that said, however, it is also true that detectives and prosecutors have discretion in making "deals" even with serious crime suspects, as when they offer to reduce or drop charges in return for confessions, incriminating evidence or testimony against codefendants, and other types of cooperation. The exercise of discretion by detectives has not been studied as closely as the discretionary arrest and use of force decisions made by patrol officers.

Concern about the consequences of under-enforcement and leniency cannot be ignored, despite the evidence that the police have been making lots and lots of arrests. "Zero tolerance" campaigns remain very popular, at least in the United States. It is not usually necessary to declare zero tolerance for murder or robbery – rather, zero tolerance campaigns usually target such lesser offenses as street prostitution and panhandling, or else they define a specific geographic area within which all criminal laws will be enforced, no matter how minor. In the public forum and in police strategy circles, zero tolerance is most commonly presented as a strict crime-control policy, leading to debates about whether it is fair, sustainable, and effective in reducing crime. However, zero tolerance is also a direct assault on police discretion. A zero tolerance policy takes discretion away from police officers since it eliminates every option other than enforcement.

In light of the concern that police discretion might result in too much leniency, it is ironic to note that, at least in the United States over the last 20 years, jails and prisons have been full, if not overflowing. In other words, the police have had broad discretion, and the evidence indicates that they have regularly used it to avoid making arrests – and yet, court dockets are crowded and jails are full. One might wonder what the situation would be if not for police discretion.

There is one aspect and consequence of this discretion/leniency phenomenon that deserves special consideration. Police discretion sometimes extends to deciding whether a reported event counts as a crime, whether to record it as a serious crime or a lesser crime, whether to thoroughly investigate it, and whether to recommend it for prosecution (Lum 2011). If the victim of a reported crime is perceived by an officer as unworthy of serious concern, the crime might be discounted or ignored. For example, this might occur to a woman domestic violence victim or sexual assault victim if the police officer identifies more with the other party involved, or if the officer believes that the victim somehow contributed to her own misfortune. Similarly, an officer might fail to act in response to a reported crime if the victim is someone whose status is not fully respected, such as a recent immigrant or a person whose sexual orientation is considered deviant by the officer. In these examples, discretion allows the police to act leniently toward offenders in situations in which their victims are held in low regard. It can be argued that these are simply more examples of discretion enabling the police to pursue justice and fairness rather than mandatory enforcement, but it is at least equally possible that, in these types of cases, discretion allows officers to act on the basis of their own stereotypes and biases to the disadvantage of deserving victims.

Another potential problem with discretion and under-enforcement/leniency arises if the police are more lenient toward majority groups and higher-status individuals than others. The police might seem lenient if they only arrest 30 % of the people they stop for drinking in public, or if they only conduct searches in conjunction with 10 % of their traffic stops. Analyses often indicate, however, that members of minority groups are disproportionately arrested for minor offenses and disproportionately subject to searches following traffic stops. In other words, the overall picture is one of under-enforcement, but often the enforcement that *is* undertaken is differentially applied to persons of color. As explained above, this pattern of police discretionary decision making, often called racial profiling, might still be accounted for by situational legal factors, but it might also be a reflection of conscious or unconscious bias.

The inherent and most fundamental challenge associated with discretion is simply inconsistency. Discretion gives police officers leeway in deciding how to handle encounters and confrontations based on an "intuitive grasp" of each situation. Ultimately, some people caught with a marijuana cigarette get arrested and some do not; some drivers who fail to signal their turns get citations and some do not. This brings into focus two competing conceptions of justice and fairness – one perspective says that fairness is treating every rule breaker the same, while the other perspective says that fairness is treating each person *as a person*, according to their unique circumstances. Many of us subscribe to the first viewpoint as long as it is applied to others, but want the second approach taken when we are the ones against whom the rules and laws might be applied.

Guiding and Controlling Police Discretion

While police discretion may be inevitable, it is widely accepted that it should not be unfettered: that it needs to be guided and controlled in order to avoid the kinds of negative consequences discussed above, including inconsistency, undue leniency, and discrimination. State statutes and local laws can and do guide police discretion to a certain degree. For example, some legislatures have sought to restrict police discretion in specific types of cases, notably in the realm of domestic violence through laws mandating arrest of offenders (Phillips and Sobol 2010). Such efforts sometimes merely shift the locus of discretion, however. If the law requires the police to arrest whenever they have probable cause to believe that a domestic assault has occurred, but officers in some situations do not think that arrest is the best response, they may become more likely to decide that the evidence falls short of probable cause, or that no crime has occurred at all. In some cases this kind of decision making may also be driven by a victim's stated preference for no arrest or by knowledge that the prosecutor is only willing to take certain kinds of cases into court. This is not to say that mandatory laws have no impact on police discretion and decision making, but it is important to recognize that the police and other criminal justice decision makers tend to react in ways that preserve their discretion (Green and Kelso 2010).

The supervising authority of the courts, including specific court rulings and the risk of civil liability, also has the potential to influence the exercise of police discretion, at least at the margins. The US Supreme Court's *Tennessee v. Garner* (1985) decision is a case in point. Prior to *Garner*, the police had the discretion to use deadly force in situations in which a fleeing or escaping felon could be captured only by shooting them as they fled. The Supreme Court ruled that this authority was too broad and effectively revoked the "fleeing felon rule," thus narrowing police discretion in the use of deadly force.

By themselves, however, laws and court decisions are generally inadequate for guiding and controlling police discretion. As noted, often they merely shift the locus of discretion to another point in the process. Also, court decisions typically resolve one specific issue but leave others unaddressed. The *Garner* case, for example, eliminated one ancient common law justification for police use of deadly force, but did not posit a new rule in its place or seek to clarify other justifications for police use of deadly force. Moreover, one of the key legal principles underlying police use of force, including deadly force, is "reasonableness" – the police can use any level of force that is reasonable to defend themselves or others from serious threats of bodily harm or to carry out their other lawful responsibilities, such as enforcing the law. While the *Garner* case ruled that deadly force was *un*reasonable in one particular set of circumstances, the reasonableness standard remains and it is inherently vague and subjective. An inevitable consequence is that the police have significant discretion in deciding whether to use force and how much force to use.

In large part because of these limitations of the law, written policies and procedures are increasingly used by police agencies to set out the parameters of officers' discretion in certain types of cases and to provide contextual guidance for the proper exercise of it. In the case of domestic violence, for example, police department policies often reinforce and tighten legal requirements by (1) articulating limited exceptions to mandatory enforcement, (2) requiring written justification whenever nonenforcement is chosen in an applicable situation, (3) identifying decisions that require supervisory approval, and (4) specifying penalties for failing to adhere to legal and policy mandates. Similarly, in the case of deadly force, police department policies frequently define an even narrower range of justifications than anything found in the law, they usually caution officers about the sanctity of human life, and they invariably specify elaborate processes that must be followed in the aftermath of any incident in which the police exercise force that could lead to death, regardless of whether anyone dies or is even injured. In neither case, domestic violence nor use of force, do written policies completely eliminate discretion, but they constrain and guide it more thoroughly than is accomplished by law alone.

Somewhat less common, but no less important, are written policies that provide guidance for the exercise of discretion in general. One example policy identifies factors that officers should and should not take into consideration when deciding whether to make an arrest (Scott 1995):

> In general, police officers, using sound professional judgment, may take the following factors into consideration when deciding whether or not to arrest a citizen: (a) the seriousness and nature of the offense (generally, the more serious the offense, the more likely arrest is the preferred course of action); (b) the potential that arrest will effectively resolve a conflict; (c) the availability of legal alternatives to arrest that would adequately resolve the conflict or problem; (d) the likelihood that the citizen will be deterred from future violations by warning and education; (e) the officer's belief that the citizen made an honest mistake in violation of the law; (f) the victim-witness's interest in prosecution; (g) the likelihood of formal prosecution of the offense; (h) the potential that arrest will create more serious breaches of the peace or other problems (e.g., inciting riot); (i) legitimate competing priorities for police resources.
>
> The following factors are among those that are improper for a police officer to consider in deciding whether or not to make an arrest: (a) the citizen's economic status, race, ethnicity, gender, or other status for which the law prohibits legal discrimination; (b) the revenue likely to be generated by fines or penalties imposed upon conviction; (c) the personal or professional relationship that the citizen has to the police officer or to other influential citizens; (d) the personal advantage to the officer for processing or avoiding processing of the arrest (e.g., overtime compensation, desire to finish tour of duty, avoidance of paperwork, etc.).

Police organizations utilize several other methods of guiding officers in the exercise of discretion. Police agency accreditation standards, while not typically dictating the substance of most discretionary judgments, can help create an organizational structure that supports administrative rule making. Police training programs are essential for improving officers' decision-making skills in the application of policies and procedures. The development of a body of professional and scientific knowledge, grounded in research and practice, about how the police can effectively and fairly address public safety problems also holds promise for shaping important discretionary decisions (Weisburd and Neyroud 2011; Center for Problem-Oriented Policing 2012). In addition, many police agencies are increasingly looking to citizens to provide them with guidance on a range of discretionary matters, from what public safety problems to focus on to what means to use in addressing them. Citizen input can be provided at the policy-making level as well as through systematic input to line officers (Livingston 1997).

Years ago, Jerome Skolnick (1966) observed that the police, like members of most organizations, are affected by the need to appear productive. Thus, officers' discretionary decisions about whether to arrest or issue citations may be influenced by supervisory or command-level expectations about how much "activity" a productive officer should generate. While not new, these pressures seem to be greater today in the era of "metrics" and heightened accountability. Many police departments now use some form of Compstat, a system designed to help top-level executives use crime analysis and statistics to hold mid-level commanders accountable for targeting and reducing crime in their areas of responsibility. These mid-level commanders often press their subordinates to increase enforcement in order to show that they are doing everything possible to reduce crime. The net effect is to reduce discretion at the lowest levels in order to satisfy pressures from above. Another effect, perverse and unintended, may be to encourage officers to use their discretion to downgrade or ignore some reported crimes, thus making it appear that crime reduction is occurring even if it is not (Eterno and Silverman 2012).

Modern technology is also affecting police discretion. Thanks to mobile radios, cell phones, in-car computers, and automatic vehicle locator systems, today's police supervisors and commanders can keep track of their subordinates more closely than was the case a decade or two ago. Even more intrusive are in-car cameras and body cameras (not to mention the ubiquitous cell phone cameras in the hands of the public) that record everything that police officers do, thus magnifying the potential for after-the-fact review. While these surveillance technologies may not directly narrow the discretion that officers have in making their decisions, they probably have a substantial indirect effect, since officers must give more consideration to how their decisions will look and be judged in retrospect. In the near future, this technological impact is likely to be magnified further when the feeds from in-car and body cameras go "live" back at police headquarters. When that time comes, supervisors and commanders will have a greatly expanded opportunity to direct officer decision making in the field. This is likely to shift the locus of some discretionary decision making away from police officers and onto the managers back in the electronic control room. This kind of technological breakthrough could produce a fundamental change to our traditional understanding of police work and the centrality of street-level police discretion. It will be quite interesting to see whether it actually does have this profound effect, or if, once again, it merely causes some shifting in the forms and locus of police discretion.

Related Entries

- ▶ Biased Policing
- ▶ Order Maintenance Policing
- ▶ Police Discretion in Law Enforcement
- ▶ Police Discretion in Providing Services and Assistance
- ▶ Police Legitimacy and Police Encounters
- ▶ Role and Function of the Police

Recommended Reading and References

Banton M (1965) The policeman in the community. Basic Books, New York

Bittner E (1970) The functions of the police in modern society. GPO, Washington, DC

Black D (1980) The manners and customs of the police. Academic, New York

Brown MK (1981) Working the street: police discretion and the dilemmas of reform. Russell Sage, New York

Center for problem-oriented policing (2012) POP Guides. Online at http://www.popcenter.org/guides/

Davis KC (1969) Discretionary justice. Greenwood Press, Westport

Eith C, Durose MR (2011) Contacts between police and the public, 2008. Bureau of Justice Statistics, Washington, DC, Online at http://www.bjs.gov/content/pub/pdf/cpp08.pdf

Eterno JA, Silverman EB (2012) The crime numbers game: management by manipulation. CRC Press, Boca Raton

FBI (2011) Crime in the United States: 2010. Federal Bureau of Investigation, Washington, DC, Online at http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2010/crime-in-the-u.s.-2010/persons-arrested

Friedrich RJ (1980) Police use of force: individuals, situations, and organizations. Ann Am Acad Polit Soc Sci 452:82–97

Goldstein J (1960) Police discretion not to invoke the criminal process: low-visibility decisions in the administration of justice. Yale Law J 69:543–589

Goldstein H (1963) Police discretion: the ideal versus the real. Public Admin Rev 23:140–148

Goldstein H (1977) Policing a free society. Ballinger, Cambridge, MA

Green T, Kelso C (2010) Police discretion in domestic violence incidents. Am Behav Social Sci J 14:24–44

Kochel TR, Wilson DB, Mastrofski SM (2011) Effect of suspect race on officers' arrest decisions. Criminology 49:473–512

LaFave W (1965) Arrest: the decision to take a suspect into custody. Little Brown, Boston

Livingston D (1997) Police discretion and the quality of life in public places: courts, communities, and the new policing. Columbia Law Rev 97(3):551–672

Lum C (2011) The influence of places on police decision pathways: from call for service to arrest. Justice Quart 28:631–665

Phillips SW, Sobol JJ (2010) Twenty years of mandatory arrest: police decision making in the face of legal requirements. Criminal Just Policy Rev 21:98–118

Reiss AJ Jr (1971) The police and the public. Yale University, New Haven, CT
Scott MS (1995) Lauderhill (Florida) police department policy manual. Personal files of the author
Skogan W, Frydl K (2004) Fairness and effectiveness in policing: the evidence. National Academies Press, Washington, DC
Skolnick J (1966) Justice without trial. Wiley, New York
Tennessee v. Garner. (1985). 471 U.S. 1
Terrill W, Leinfelt FH, Kwak D-H (2008) Examining police use of force: a smaller agency perspective. Policing Int J Police Strat Manag 31:57–76
Varano SP, Schafer JA, Cancino JM, Swatt ML (2009) Constructing crime: neighborhood characteristics and police recording behavior. J Criminal Just 37:553–563
Weisburd D, Neyroud P (2011) Police science: toward a new paradigm, New perspectives in policing. National Institute of Justice, Washington, DC
Westley WA (1970) Violence and the police. M.I.T. Press, Cambridge, MA
Whitaker GP (1982) What is patrol work. Police Stud 4:13–22
Wilson JQ (1968) Varieties of police behavior. Harvard University, Cambridge, MA

Police Discretion in Law Enforcement

Robert E. Worden[1,2] and Sarah J. McLean[2]
[1]School of Criminal Justice, University at Albany, State University of New York, Albany, NY, USA
[2]John F. Finn Institute for Public Safety, Inc., Albany, NY, USA

Overview

Coercive authority is central to the police role, and the discretionary use of police authority is a decision-making process, as officers evaluate the situations in which they intervene and choose a course of action from among a set of alternatives. This entry approaches the study of police discretion from this decision-making perspective. It is concerned with several forms of authority: stops, frisks and searches, arrests, tickets, and the use of physical force.

Police authority is infrequently invoked, as a proportion of all police-citizen contacts. Officers consider their authority as a resource in "handling" situations, and often they are able to handle situations without resorting to the use of formal authority. Research has illuminated to a degree the circumstances under which authority is used, from which inferences are drawn about the influences on police decision-making, but research has not for the most part formed the basis for normative judgments about the discretionary use of authority – when it has been used well or poorly.

Explaining the police decision-making process can be approached like any decision-making process. Behavior can be predicted or explained when the premises for decision are known. Police officers' decisions turn on factual premises, which concern the consequences of alternative police actions and the nature of the situation from which the consequences can be projected, and value premises, about the goals of police intervention and the desirability of alternative sets of consequences. Explaining police decision-making is, however, complicated by the ambiguity and uncertainty of police officers' task environments.

Many different elements of police-citizen encounters have been analyzed as hypothesized influences on police behavior, but only a few of them have been consistently found to affect the use of police authority. Several of those are legal factors: the seriousness of the offense, the strength of evidence, and the preferences of complainants for disposition. Among extralegal factors, only the demeanor of suspects has emerged as a fairly consistent predictor of police action.

The discretionary exercise of police authority varies across individual officers. It is fairly clear that this variation is patterned by officers' length of service and also by officers' outlooks and personality traits. The influence of other individual factors, such as educational background, is undetermined.

The use of police authority is subject to regulation by departmental policies, but the circumstances under which policy regulation can have maximum impact are seldom realized in policing. The use of deadly force can be successfully restricted. Other uses of police authority are probably less susceptible to policy guidance.

Among the many open questions about police use of their authority are three addressed in this entry: the influence of citizens' race, modeling police decision-making, and how research might contribute to improvements in the quality of police decision-making.

Introduction

Coercive authority is central to the police officer's role; it is a unique occupational prerogative that enables police to "handle" urgent situations. The situations may be crimes, with clear and serious violations of criminal law, the response to which is the apprehension of perpetrators, or the situations could be disorders, with ambiguities stemming from whether the law applies, who bears how much responsibility for the situation, and what should be done about it. But all of these situations share an element of urgency, in that prompt action is needed, whether it is a burglary in progress, a drunk driver, a loud party, a domestic conflict, or a blocked driveway (see especially Bittner 1974; Muir 1977).

In each instance, police exercise discretion, which is the capacity to choose among courses of action based on one's judgment. Police officers typically work outside of direct supervision and often in private settings in which only the officer(s) and the parties directly involved have knowledge of the circumstances and the actions, if any, that police take. So while thick books of policies and procedures exist in many agencies, in practice, officers are left to define the situation and fashion a response, thus determining whether and, if so, which policies and procedures apply.

The exercise of police discretion is a decision-making process, as officers evaluate the situations in which they intervene and choose a course of action from among a set of alternatives; if their involvement is not at the request of a citizen, they also make decisions about whether to intervene. This entry approaches the study of police discretion from this decision-making perspective. In the next section, the entry discusses the forms of authority that police exercise and how they have been analyzed; then it explains how a decision-making framework can be applied to police discretion in their use of authority. In the following section, the entry summarizes what is known about the explanations of officers' discretionary choices. Then several open questions are addressed.

Police Authority and Discretionary Decision-making

Police Authority

Police exercise several forms of authority. An officer may stop and detain someone when he/she has reasonable, articulable suspicion that the person has committed, is committing, or is about to commit a crime. An officer may frisk, or pat down, an individual if she/he has reasonable suspicion that the person might be armed. An officer may conduct a more extensive search of a person incident to that person's arrest and may conduct a warrantless search of a person, a vehicle, or even a building under some other circumstances. An officer may take someone into custody for the purpose of charging them with a crime; if the suspected offense is a felony, an officer needs only probable cause to believe that a crime was committed, while for most offenses of lesser seriousness, the officer may act only if she/he witnesses the offense or a citizen signs a complaint. For some types of offenses, officers may issue a summons that directs the alleged offender to appear in court, rather than make a custodial arrest; this is of course common for traffic violations and also for some other less serious offenses (such as public drinking). Finally, police may use the physical force that is reasonably necessary to perform their duties; the US Supreme Court ruled in *Tennessee v. Garner* that police may even use deadly force if an "officer has probable cause to believe that the suspect poses a threat of some physical harm, either to the officer or others." The forms of physical force that police may use has multiplied with the development and adoption of less-lethal weaponry, such as pepper spray and conducted energy devices, in the last 10–20 years.

Social science has focused mainly on the exercise of authority by patrol officers, and it has normally conceived, measured, and analyzed these forms of police authority one by one – that is, some studies examine arrest, while others analyze the use of physical force, and still others analyze stops, etc. In most instances, officers' use of police authority consists of a discrete act – making an arrest, conducting a search; the main exception is the use of physical force, which is commonly broken down into forms of force that vary in their seriousness, such as physical restraint, impact methods (punches or kicks), and the use of weapons. These are not mutually exclusive forms of behavior, of course; to the contrary, an officer might well stop and detain, frisk, arrest, and physically restrain an individual in a single encounter. One effort to form a single scale of authority weighted the individual components, but empirically it was not much different from an arrest/no arrest dichotomy.

In general, police authority is infrequently invoked, as a proportion of all police-citizen contacts. Many years ago, James Q. Wilson (1968) observed that the tendency of the police is to *underenforce* the law, declining to take legal action even when the circumstances authorize them to take legal action. Thus it was not uncommon for police to release in the field offenders about whose culpability police had evidence sufficient to warrant arrest. Wilson pointed out that police see their authority as a resource in accomplishing their objectives, such that the application of their authority is not an end but a means, and in many instances their objective is merely to "handle the situation" – that is, to restore order and to prevent immediate violence. When an arrest was made, the "formal charge justifies the arrest but is not the reason for it" (Bittner 1974, p. 27); the reason for such an arrest is that the situation could not be handled in some other, informal fashion. Similarly, studies of the use of force have found that police seldom use physical force, and some research has found that police frequently refrain from using physical force when they could legally do so or use less physical force than the citizen's resistance would justify. Sometimes police presence by itself is sufficient to restore order. And even if presence alone is not enough, police may be able to resolve matters by talking: persuading, cajoling, exhorting, negotiating, or mediating.

Since Wilson made his observation, with changes in the (semiprofessional) status of police, criminal procedure, and civil liability, the view of the police may be somewhat different, but not dramatically different. Even far into (and perhaps beyond) the "community era" of policing, authority is central to the police role. Police may now supplement their use (or threatened use) of authority with other approaches, but even when they perform problem-oriented policing, authority remains a resource on which they can and often do draw. In addition, the difficulty of reorienting police to a more community- and problem-oriented role – getting officers to think in terms of groups of related incidents rather than individual incidents and to consider unconventional solutions to those problems – suggests that the use of coercive authority has not changed dramatically in the last 40–50 years. Arrest and physical force remain infrequent acts. Police proactivity – contacts with citizens initiated by the police – is almost certainly at a higher level now, however, and it has occasioned some controversy about its benefits and the social distribution of its costs.

Another change over time might lie in the public's expectations for police effects on social conditions. In the early 1970s, Egon Bittner could say that police are "empowered and required to impose or . . . coerce a *provisional* solution upon emergent problems . . ." (1974: 18; emphasis added), suggesting that the solution extended no further than the immediate situation. Yet we evaluate the police in terms of more – for example, research on policing spouse assault assessed police interventions in terms of recidivism by assailants over a 6-month post-intervention period, not in terms of more "provisional" solutions.

Aside from hypothesized biases in the use of authority, to be discussed below, research has tended *not* to address the normative dimensions of police authority, that is, whether an arrest or the use of physical force was a good choice or a bad choice. The situations that police handle are complex, and so a normative judgment about the use of police authority is contingent on many factors. Moreover, the criteria against which such a judgment should be made are not at all clear. Officers' choices could be judged in terms of immediate outcomes, such as whether agitated citizens are calmed, or in terms of longer-term outcomes, such as whether a domestic assailant recidivates. None of these have been much considered in research on police use of authority.

Another potential criterion is a professional one of "workmanship." Bittner (1983) distinguishes legality from workmanship as standards against which to judge police performance. The former involves the application of "explicitly formulated schemes of regulation," while the latter is concerned with officers' "ability to call upon the resources of knowledge, skill, and judgment to meet and master the unexpected within one's sphere of competence." Bittner goes on to explain that the standard of workmanship would subject officers' performance to the scrutiny and critique of others authorized to assess their work after the fact: peers, who are subject to the same review, and the community, on whose behalf the work is done. Similarly, Klockars (1996) proposes a "craft standard" against which the use of force could be compared, that is, what a skilled police officer would have done under the circumstances. A skilled police officer would take all possible steps to structure the situation into which she/he intervenes in such a way that either resistance will not be offered or the resistance can be overcome with a minimum of force. Few efforts have been made to develop such standards conceptually or empirically, however.

As a general proposition, social scientists seem agreed that authority is a resource that police should use sparingly. Muir (1977) describes a "good" officer as one who is morally comfortable with the use of coercive power as needed but who uses it *only* as needed. Moore (2002) allows as how police are more efficient when they use a minimum of authority. The use of police authority, he points out, depletes the "stock of private liberty" (2002: 26) that the public enjoys, and the "net value" of policing must take account of not only the crime and disorder that police prevent but also the costs of policing, one of which is the use of authority.

In the last 10 years or so, research has reflected a better appreciation for how the quality of police performance might have as much (or more) to do with *how* authority is exercised as it does with what form of authority is exercised. Surveys of citizens have repeatedly shown that they are more satisfied with their contacts with the police when they believe that police treated them with dignity and respect, gave them an opportunity to explain their situations and listened to what they had to say, and explained what police have done and/or will do, so that it is clear that officers are taking account of people's needs and concerns and basing their decisions on facts. Moreover, these patterns hold even for people who experience unfavorable outcomes, such as being ticketed. Unfortunately, however, research on police behavior has only scratched the surface in its examination of the behavioral manifestations of procedural justice.

Decision-Making

The discretionary use of police authority is a decision-making task, as officers must make choices – numerous choices – in their encounters with citizens. Among their choices, the uses of their authority are prominent: to stop and question, to frisk or search, to arrest, to cite, and to use any of a variety of forms of physical force. As an encounter with a citizen unfolds over time, the choices may recur.

Explaining this decision-making process, and the behavior that ensues therefrom, can in principle be approached like any decision-making process. Behavior can be predicted or explained when the premises for decision are known. Police officers' decisions turn on factual premises, which concern the consequences of alternative police actions and the nature of the situation from which the consequences can be projected, and value premises, about the goals of police intervention and the desirability of alternative sets of consequences (see Worden and Brandl 1990, pp. 302–303). Most research on police behavior has, in effect, treated as hypothesized decision premises the characteristics of police-citizen encounters, such as the nature of the offense, the demeanor of suspects, and the preferences of complainants for a disposition. Several studies have also examined as potential influences the characteristics of police organizations; organizational policies, procedures, and informal norms may serve as decision premises. Some research has, in effect, allowed for variation in decision-making among classes of decision-makers, for example, male and female officers, novice and experienced officers, and college-educated and less-educated officers.

Explaining police decision-making in terms of situational and organizational cues is complicated by the ambiguity and uncertainty of police officers' task environments. Situational cues are ambiguous in that the meaning or significance of any cue, or set of cues, is subject to different interpretations. The credibility of citizens' testimony (which often is contradicted by other citizens' testimony), the severity of injuries, the degree of citizens' sobriety, and many other features of police-citizen encounters are all matters of officers' judgment, and these judgments can vary from situation to situation and from officer to officer. The same cue can be interpreted differently across multiple events or even ignored as irrelevant in some of them. The meaning and significance even of readily recognizable cues is ambiguous, subject to different interpretations by different officers. Similarly, organizational policies and group norms are ambiguous, requiring that officers interpret them as they are applied in individual cases.

Explanation is further complicated by uncertainty about cause-and-effect relationships. Once an officer interprets the meaning of situational cues and the applicability of organizational policies, she/he must choose a course of action based partly on the projected consequences of those alternatives, but no body of scientific or other technical knowledge guides officers in making such projections, and even if there were, they would be in any case only probabilistic. Officers may be socialized into some common working rules, but they also learn through their own idiosyncratic experiences, drawing from them some equally idiosyncratic lessons for practice. Thus we can posit that the meaning and also the implications of situational and organizational cues are heterogeneous across officers, which would attenuate the explanatory power of situational and organizational variables.

Prior research on police decision-making, for the most part, takes one of two approaches: ethnographic inquiry and quantitative analyses of decision outcomes (e.g., police decisions to arrest or to use force) in terms of situational factors that are the hypothesized stimuli to which decision-makers respond. The former have been very illuminating, though they are limited in their generalizability. The latter are scientifically rigorous, exploiting numerical data on the characteristics of the immediate decision task to estimate (often through regression analysis) their independent effects on decision outcomes. Researchers draw inferences about the forces that shape decision-making from the regression coefficients, treating the process by which informational inputs are interpreted and judgments are made to reach decisions as a "black box." But these analyses are limited in their explanatory scope to factors that are of a priori significance, and we know that these factors fall far short of explaining police decisions.

Moreover, much of the research has been oriented not to explaining discretionary decision-making as such, but rather to determining whether and to what extent the application of police authority is influenced by "extralegal" factors, such as race, sex, and age. Thus, the analyses seek to control for the legal factors that legitimately shape police behavior, whereupon a determination can be made about whether residual variation is related to extralegal factors.

Influences on Discretionary Decision-Making

Most of the factors that have been analyzed as parts of an explanation of discretionary decisions by police have been classified into one of several sets, each a different level of explanation. As noted above, some of these factors, and those at the lowest level of explanation, are features of the police-citizen encounters in which the discretionary decisions are made; these are often referred to as situational factors. A second set, and the next higher level of explanation, are comprised of the characteristics and outlooks of individual officers. A third set, and a still higher level of explanation, are characteristics of the neighborhoods that form the immediate social context of police-citizen encounters. The fourth set include characteristics of police organizations, including both the formal, structural features of police agencies – rules, regulations, and specialization – and the informal organization: the peer culture(s) and norms. Finally, some research has attended to another, still higher level of organization: the requirements and constraints of the law, especially that of criminal procedure. This entry concentrates on three of these sets: situational, officers, and organizational.

Situational Factors

Situational factors are external to the decision-maker and form the immediate decision environment. The most succinct statement of this explanatory perspective is Bittner's, who asserts that "the role of the police is best understood as a mechanism for the distribution of non-negotiably coercive force employed in accordance with the dictates of an intuitive grasp of situational exigencies" (1970: 46). Situational factors represent these exigencies. A more conventional sociological formulation is that of Black and Reiss (1967), who posit that police action turns on the social status, situational status (e.g., as complainant, suspect), and "subversive capability" of citizens. Most research of this genre, however, has been less theoretically driven in its assessments of situational exigencies.

Many different elements of police-citizen encounters have been analyzed as hypothesized influences on police behavior, but only a few of them have been consistently found to affect the use of police authority. Several of those are legal factors: the seriousness of the offense, the strength of evidence, and the preferences of complainants for disposition. Among extralegal factors, only the demeanor of suspects has emerged as a fairly consistent predictor of police action; the estimated effects of sex and age have been mixed, as have those of race, to which the entry returns below.

It should come as no surprise that the use of authority by American police is affected by legal factors, but research confirms that police are more likely to use their authority when they have evidence of criminal offending – and the stronger the evidence, the more likely is the use of authority – and when the offense is more serious. It is perhaps remarkable, however, that the effects are not stronger still. Certainly, police use of authority is not *determined* by legal factors; as noted above, police often do not take legal action even when they have grounds for legal action.

Furthermore, when complainants articulate a clear preference for or against legal action, police tend to comply, especially when the preference is for lenience. This tendency "gives police work a radically democratic character," as Black (1971) points out, and as a result, police apply a standard of justice that varies with the moral standards of complainants, not a uniform standard of justice. However, as Mastrofski et al. (1995) observe, the cooperation of complainants is likely to affect prosecution, and so when complainants prefer that police not take legal action, police might reasonably treat that as an "instrumental" legal consideration. The frequency with which victims of domestic violence are reluctant to cooperate has, given concerns that these complainants are not exercising choice freely, prompted changes in laws and policies that are intended to attenuate the effect of complainant preference in such cases.

Both qualitative and quantitative studies have found that police are more likely to respond punitively to suspects who display disrespect toward the police. Empirical support for this relationship is nearly uniform in research on police behavior, even in the absence of a single, widely accepted definition and operationalization of demeanor, and notwithstanding whatever random measurement error has been contained in indicators of demeanor. This relationship is open to one or both of two interpretations: police penalize people who flunk the "attitude test" with arrest, ticketing, or a "thumping" (Van Maanen 1978), or police often give "breaks" to deferential suspects but seldom to those who are hostile or disrespectful. Some support can be found for both of these causal mechanisms.

Klinger (1994) questioned this "criminological axiom", arguing that demeanor had been improperly conceived and measured in previous research and also that previous research had failed to control (adequately or at all) for crime committed by suspects during their encounters with police, including especially assaults on police officers. However, reanalyses of then-extant data, using measures that took account of Klinger's critique, affirmed the previous conclusions, and other elements of Klinger's argument have also been disputed (Worden et al. 1996). But Klinger's critique surely revealed shortcomings and inconsistencies in the conceptualization and measurement of demeanor, and subsequent studies have exhibited the greater care that Klinger's critique demanded.

Even so, theoretical ambiguity remains. Research now differentiates "resistance" from "demeanor," yet many forms of (especially passive or only verbal) resistance are entirely legal, and we have reason to believe that many forms of (both legal and illegal) resistance are interpreted by police as failures of the attitude test, such that we should interpret their effects as the effects of demeanor (see Worden et al. 1996). So long as research is interested in explaining the discretionary use of police authority, and not merely in forming post hoc judgments about its propriety, conclusions about the effects of demeanor must turn on how officers interpret the various manifestations of disrespect.

More generally, the theory (such as it is) and much of the empirical analysis of situational effects has been based on a sociological framework, and one that makes an implicit assumption that the effects are not contingent on the backgrounds or outlooks of individual officers. This assumption is not entirely without merit, inasmuch as early research on police emphasized the shared working conditions, experiences, values, and norms among police; hence assuming (albeit implicitly) that officers attend to the same cues, interpret them similarly, consider the same set of alternative courses of action, and choose based on the same criteria is not unreasonable. Furthermore, some results seem to suggest that, on some dimensions of decision-making, officers approach their choices similarly: the arrest decision is shaped by the seriousness of the offense, the strength of evidence, the preferences of complainants, and the demeanor of suspects, and while there may be individual differences on these scores, the differences are not stark, and across the board it appears that these variables have a substantial degree of explanatory power. In other choices, especially choices among informal responses, however, the effects of situational factors have not been so robust (Worden 1989), and the use of authority other than arrest may be more susceptible to individual differences. Several qualitative studies in the 1970s (e.g., Brown 1981; Muir 1977) showed that officers do not all see their occupational environment the same way and that they develop distinct "operational styles"; different studies using different frameworks nevertheless described a set of officer types that were broadly congruent

Officer Characteristics and Outlooks

While it is clear that officers' discretionary choices are structured by the characteristics of the incidents that they handle, it is also clear that these situational elements do not completely determine officers' responses, as much of the variation in their behavior is unexplained by situational variables. Because officers' decisions are of "low visibility" to their superiors and to the legal system, scholars have sometimes speculatively attributed the unexplained variation in behavior to officers' attitudes and values. This perspective directs attention to factors internal to decision-maker, which might affect how situational cues are interpreted and evaluated and how options are assessed. For example, Toch (1980) found that violence-prone officers are especially sensitive to citizens' challenges to their authority. Such an account has intuitive appeal, and it is of more than theoretical significance, inasmuch as it underlies initiatives to change the composition of or otherwise "upgrade" police personnel, since it is supposed that minority officers, female officers, or college-educated officers have different outlooks and therefore perform the police role differently. It also underlies efforts to inculcate different outlooks through training.

Systematic research has not always confirmed these intuitive propositions. One of the earliest and most striking findings from observational research on police was that, while many officers professed to be prejudiced against African-Americans, officers did not act on that prejudice in their encounters with citizens (Reiss 1971). A later and more general examination of attitude-behavior consistency (Worden 1989) found only weak relationships between officers' occupational attitudes (for example, their conceptions of the police role, attitudes toward citizens, and attitudes toward legal restrictions) and their behaviors (for example, proactivity in field interrogation and dispute resolution). Such results perhaps should not be surprising, in view of the ambiguity and uncertainty of police work. The ambiguous situations into which police intervene are subject to different interpretations, and they are not governed by well-established cause-and-effect relationships between police actions and desired outcomes, so officers' choices among alternative courses of action may be only loosely structured by their attitudes and values. Officers with the same occupational outlooks might judge a situation differently, reading the informational cues in a situation in different ways, or perceive different ways of reaching the same objectives.

Be that as it may, analyses in several agencies have shown that small numbers of officers account for disproportionately large fractions of citizen complaints and use-of-force reports. Furthermore, Brandl et al. (2001) found that less experienced officers are disproportionately represented among officers with multiple complaints about the use of excessive force. Scrivner (1994) discovered five groups of officers among those referred to police psychologists due to their use of excessive force, including officers with personality disorders; officers whose job-related experiences – for example, traumatic incidents such as police shootings – put the officers at risk for abusing force; young and inexperienced officers who were also "highly impressionable and impulsive"; officers who develop inappropriate patrol styles; and officers with personal problems. Harris (2010) showed that officers differ in their career "trajectories" of misconduct.

Further, more recent findings have provided more support for the hypothesized effects of officers' outlooks, mainly because it has analyzed attitudes as bundles in terms of which subsets of officers can be differentiated, rather than estimating the additive effects of attitudes one by one. For example, Terrill et al. (2003) found that officers whose occupational attitudes conform more closely to the tenets of the traditional police culture are more prone to the use of their coercive authority. Similarly, Paoline and Terrill (2005) found that such officers are more likely to conduct searches during traffic stops. Cuttler and Muchinsky (2006) found that personality traits and work history predict "dysfunctional job performance." Other characteristics of officers – their race, sex, and educational background – have all been hypothesized to affect how officers do their jobs, though the evidence on these hypotheses is mixed and inconclusive (National Research Council 2004).

Thus it is fairly clear that the discretionary exercise of police authority varies across individual officers. It is also fairly clear that this variation is patterned by officers' length of service and also by officers' outlooks and personality traits. The influence of other individual factors, such as educational background, is undetermined.

Organizational Factors

The hypothesized effects of organizational factors, and particularly the formal, structural characteristics of rules, regulations, standard operating procedures, and the hierarchical chain of command through which they are enforced, are central to common ideas about police accountability. Police departments are widely known as quasi-military organizations, with many of the trappings of such organizations (such as ranks and insignia), as well as the aura of a tight chain of command. The nature of police tasks, however, which makes discretion an inherent element of police work, also circumscribes the effects of these organizational factors on the use of police authority.

One mechanism by which police departments might influence the use of force by their officers is, of course, through rules and regulations that set limits on when police may use force and how much force they may use. Research on the use of deadly force suggests that this mechanism can be effective, under some circumstances. In 1972, well in advance of the *Garner* ruling by the Supreme Court (in 1984), the New York City Police Department (NYPD) modified its policy about the use of deadly force, imposing much tighter restrictions, providing that "every other reasonable means will be utilized for arresting, preventing, or terminating a felony or for the defense of oneself or another before a police officer resorts to the use of his firearm" (quoted in Fyfe 1979). A study of the effects of this policy change showed that it had a substantial effect on officers' use of their firearms (Fyfe 1979). The policy had an especially pronounced effect on what Fyfe (1988, p. 185) characterized as "elective" shootings, that is, situations in which officers could have chosen not to shoot at no risk to themselves or other parties. Furthermore, the reduction in shootings was achieved with no increase in officer injuries or deaths. The effectiveness of this policy almost certainly turned to a large extent on its enforcement. The NYPD established a Firearms Discharge Review Board to investigate and adjudicate all firearm discharges, holding officers accountable for their use of deadly force. Other research also found that restrictive deadly force policies are effective only if they are vigorously enforced.

No one should suppose that departmental policies that regulate the use of nonlethal force are or can be equally effective, however. The circumstances under which nonlethal force is justified are more heterogeneous and also more ambiguous; thus policy prescriptions concerning nonlethal force cannot achieve equivalent clarity. Use-of-force continua and matrices specify different kinds and degree of force, and (sometimes) the degrees of citizen resistance to which forms of force correspond, but the resistance can take widely varied forms and the behavioral lines that conceptually demarcate different levels of resistance are far from clear in practice; ambiguity is inescapable. Moreover, the use of nonlethal force is probably less reliably reported by officers, and even if such force is infrequent as a proportion of police-citizen contacts, it happens with sufficient frequency that police superiors may not be able to routinely investigate every use of nonlethal force in order to determine officers' compliance with policy. Furthermore, even when investigations can be conducted, one can anticipate that many will be inconclusive, with evidence limited to the contrasting accounts of the officer and the citizen against whom force was used. Thus the enforcement of such policies would almost inevitably lack the teeth associated with NYPD's reviews of firearm discharges (where, as Fyfe points out, the population density is so high that one might reasonably suppose that few firearms discharges could be safely concealed). This is not to imply that policies would be altogether ineffective in limiting and guiding the use of nonlethal force, however; one recent study suggests that some types of policies are more effective than others – and presumably, better than no policy at all – in promoting the use of force that is proportional to citizen resistance (Terrill et al. 2011).

Similar expectations might be formed with respect to policies that are intended to promote or even mandate arrest in cases of domestic violence. The situations are ambiguous, and so the applicability of policy is subject to officers' interpretations in the field, and enforcement of policy requirements would be hindered by the cost and difficulty of determining whether officers complied with the policy in individual cases, which are numerous. Even so, some research indicates that pro-arrest policies yield increases in arrests in domestic assault cases, though examples of resistance and noncompliance can be found.

Open Questions

The open questions about police use of their authority are far too numerous to even list here, much less consider them. Three questions are addressed: the influence of citizens' race, modeling police decision-making, and how research might contribute to improvements in the quality of police decision-making.

Research on the police consistently finds that African-Americans are disproportionately represented among the people against whom police use their authority. It has inconsistently found that, other things being equal, race affects at the margin the application of police authority. The question of racial (and ethnic) bias in policing dates to at least the 1960s, when the grievances of urban minority populations against the police were cited as contributing factors to civil unrest. It has been especially salient in the aftermath of incidents in which police have been suspected of abusing their authority in their treatment of minority citizens, of which the Rodney King incident in 1991 may be the most widely known example. But since the mid-1990s, concerns about racial profiling by police – that is, the use of a drug courier profile that included race as a feature, and more generally stopping and searching African-Americans and other people of color based on their race or ethnicity – have been prevalent and sustained, even prompting many states and localities to mandate that police departments track the race of the people who are stopped. Unfortunately, the volume of analysis on this question has not been matched by its theoretical sophistication, and bedeviled by what is sometimes known as the benchmarking problem, the studies can seldom provide persuasive evidence about whether the disparities are or are not attributable to legitimate practices, and moreover, they are not oriented toward explanation and offer little insight into the role of race in police decision-making.

The mixed findings on the role of race may reflect the complex, contingent nature of the effects of race, but they might also be symptomatic of drawbacks to the conventional approach to analyzing police decision-making. This approach, described above, rests on an assumption that police decisions, like structural regression models, are a weighted sum of the postulated decision cues: legal seriousness, strength of evidence, complainant preference, suspect demeanor, and the like. It is unlikely that the process of police decision-making resembles this computational model, however. And the regression models seldom explain more than a small fraction of the variation in behavior. Different kinds of data and a different kind of model might yield deeper insights into the forces that shape police decisions.

A different approach, that of "protocol analysis" or "process tracing," promises to shed further light onto decision-making by opening the black box of police officers' cognitive processes (see Worden and Brandl 1990, and the sources cited therein). For research based on protocol analysis, decision-makers are asked to think aloud as they perform decision tasks, or they might be asked to recount their thinking as soon as possible after performing a decision task. Research subjects' verbal reports of their thinking are data on their decision processes. Protocol analysis of police has not been conducted often, but several studies demonstrate that it is both feasible and illuminating. For example, Stalans and Finn (1995) discovered that the perceptual and cognitive processes that officers apply in interpreting and evaluating domestic violence incidents vary by experience: less experienced ("novice") officers tend to evaluate domestic violence situations in terms of blame, while their more experienced colleagues are more pragmatic, putting more emphasis on the sparing use of arrest (and their time) and on an assessment of the immediate potential for further violence and the prospects for prosecution. Both novice and experienced officers, however, tended to assess situations as typical or atypical. Stroshine et al. (2008) enumerate informal "working rules" that serve to guide officers in performing their jobs, and they observe that "different officers look for different things and respond to suspects and situations differently" (2008: 335). They describe the cues to which officers attend in forming suspicion (e.g., time and place, and citizens' appearance and behaviors), the rules of thumb that they apply (e.g., with respect to the seriousness of violations, or the citizen's demeanor), and more general approaches that they take to their work (seeking out or avoiding opportunities for traffic enforcement). The potential of protocol analysis might be best realized through the construction of process models, which resemble flow charts. Compared to the structural models that are normally estimated, process models are probably more accurate representations of how people make complex decisions, especially under conditions of ambiguity and uncertainty.

Finally, an overriding question revolves around the normative dimensions of police decision-making and the use of police authority. Research has little to say about the extent to which, and the circumstances under which, police use their authority well. Even research on the use of physical force has, for the most part, lost a focus on the improper use of force. Research will be far more useful if and when it can draw empirically grounded conclusions about when police should use what authority and why. Research will also be more useful if and when it can establish the connections between what the police do – what authority they exercise and how they exercise it – and how it is subjectively experienced by citizens; survey research has dwelt on citizens' perceptions, but it has not estimated the relationships between citizens' perceptions and overt police acts.

Recommended Reading and References

Bittner E (1970) The functions of the police in modern society. National Institute of Mental Health, Bethesda, MD

Bittner E (1974) Florence nightingale in pursuit of Willie Sutton: a theory of the police. In: Jacob H (ed) The potential for reform of criminal justice. Sage, Beverly Hills

Bittner E (1983) Legality and workmanship: introduction to control in the police organization. In: Punch M (ed) Control in the police organization. MIT Press, Cambridge, MA

Black D (1971) The social organization of arrest. Stanford Law Rev 23:1087–1111

Black D, Reiss AJ Jr (1967) Patterns of behavior in police and citizen transactions. In: Studies of crime and law enforcement in major metropolitan areas, vol II, Sec I. GPO, Washington, DC

Brandl SG, Stroshine MS, Frank J (2001) Who are the complaint-prone officers? an examination of the relationship between police officers' attributes, arrest activity, assignment, and citizens' complaints about excessive force. J Criminal Just 29:521–529

Brown MK (1981) Working the street: police discretion and the dilemmas of reform. Russell Sage, New York

Brown MK (1988) Working the street: police discretion and the dilemmas of reform. Russell Sage, New York

Cuttler MJ, Muchinsky PM (2006) Prediction of Law enforcement training performance and dysfunctional Job performance with general mental ability, personality, and life history variables. Criminal Just Behav 33:3–25

Fyfe JJ (1979) Administrative interventions on police shooting discretion. J Criminal Just 7:309–323

Fyfe JJ (1988) Police use of deadly force: research and reform. Justice Quart 5:165–205

Harris CJ (2010) Pathways of police misconduct: problem behavior patterns and trajectories from two cohorts. Carolina Academic Press, Durham

Klinger DA (1994) Demeanor or crime? Why 'Hostile' citizens are more likely to be arrested. Criminology 32:475–493

Klockars CB (1996) A theory of excessive force and its control. In: Geller WA, Toch H (eds) Police violence: understanding and controlling police abuse of force. Yale University Press, New Haven

Mastrofski SD, Worden RE, Snipes JB (1995) Law enforcement in a time of community policing. Criminology 33:539–563

Moore MH (2002) Recognizing value in policing: the challenge of measuring police performance. Police Executive Research Forum, Washington, DC

Muir WK Jr (1977) Police: streetcorner politicians. University of Chicago Press, Chicago

National Research Council (2004) In: Committee to Review Research on Police Policy and Practices, Skogan WG, Frydl K (eds) Fairness and effectiveness in policing: the evidence. Committee on Law and Justice, Division of Behavioral and Social and Behavioral Sciences and Education. Washington, DC: National Academies Press

Paoline, Eugene A. III, William Terrill (2005) The Impact of Police Culture on Traffic Stop Searches: An Analysis of Attitudes and Behavior. Policing: An International Journal of Police Strategies and Management 28(3):455–472

Reiss AJ Jr (1971) The police and the public. Yale University Press, New Haven

Scrivner EM (1994) Controlling police Use of excessive force: the role of the police psychologist. National Institute of Justice, Research in Brief, Washington

Stalans LJ, Finn MA (1995) How novice and experienced officers interpret wife assaults: normative and efficiency frames. Law Soc Rev 29:287–321
Stroshine M, Alpert G, Dunham R (2008) The influence of ‘working Rules’ on police suspicion and discretionary decision making. Police Quart 11:315–337
Terrill W, Paoline EA III, Ingram J (2011) Assessing police use of force policy and outcomes. Michigan State University, East Lansing
Terrill W, Paoline EA III, Manning PK (2003) Police culture and coercion. Criminology 41:1003–1034
Toch H (1980) Violent men, Revth edn. Schenkmen, Cambridge, MA
Van Maanen J (1974) Working the street: a developmental view of police behavior. In: Jacob H (ed) The potential for reform of criminal justice. Sage, Beverly Hills
Van Maanen J (1978) The asshole. In: Manning PK, Van Maanen J (eds) Policing: a view from the street. Goodyear, Santa Monica, CA
Wilson JQ (1968) Varieties of police behavior: the management of Law and order in eight communities. Harvard University Press, Cambridge, MA
Worden RE (1989) Situational and attitudinal explanations of police behavior: a theoretical reappraisal and empirical assessment. Law Soc Rev 23:667–711
Worden RE, Brandl SG (1990) Protocol analysis of police decision making: toward a theory of police behavior. Am J Criminal Just 14:297–318
Worden RE, Shepard RL, Mastrofski SD (1996) On the meaning and measurement of Suspects’ demeanor toward the police. J Res Crime Delinq 33:324–332

Police Discretion in Providing Services and Assistance

John D. McCluskey
Department of Criminal Justice, Rochester
Institute of Technology, Rochester, NY, USA

Overview

Consideration of police discretion as it relates to the provision of services to the public is an area of much speculation but little hard data. It is widely recognized that police are called upon to do much more than fight crime and enforce the law (e.g., Cumming et al. 1965; Guyot 1991). Patrol officers, for example, spend little time arresting individuals for crimes in any given work shift and are often observed not arresting under circumstances when authorized to do so by the law. Nevertheless, the academic literature focuses heavily on the decision to arrest and the decision to use force, because they address issues that are at the core of democratic governance: How does the state use its coercive power? This leaves a substantial gap in understanding what police choose to with regard to offering help and assistance as necessary and as directed by the public. The gap is especially apparent with regard to enumerating and explaining the actions police undertake outside the scope of their legally mandated and more clearly authorized law enforcement duties (Skogan and Frydl 2004).

The issues surrounding police discretion in helping and serving the public will be discussed along several dimensions. First, the history of police and their role, particularly in the United States, will be considered. The mandate of police, as a social control mechanism, offers an opportunity to understand broader patterns of police behavior over the long term and how police relate to and define their work. Second, service and helping behaviors will be defined. Since service and helping constitutes, in the most rigorous analysis, a variable, it is necessary to locate reasonable boundaries and dimensions upon which service and helping can be arrayed. No precise definition will be offered, as technological changes and task requirements continuously shape police work, but a set of considerations will be explored. Third, measurement of service including attempts and future directions will be considered. Fourth, a set of preliminary propositions regarding police exercise of discretion in allocating or engaging in helping behaviors will be outlined. Fifth, the transformation of the police mandate toward recognition, if tacit, of the importance of service activity is examined in light of problem-oriented policing and the community policing movements and the influence on discretion regarding helping behaviors. Sixth, the service and help behaviors of police will be related to the specific client population of mentally ill citizens whom which they very frequently encounter. Finally, the entry is summarized and future directions for consideration in the areas of service and helping behavior and police discretion are offered.

History

US municipal police forces, according to some historians, were established as general service provision agencies and had wide-ranging responsibilities including dealing with lost children, inspecting boilers, and feeding and housing the homeless. Lane (1967) documents the primary role that the Boston police, for example, played in running soup kitchens and providing temporary lodging for the homeless. This service was based on the observation that "misery bred crime and soup helped to prevent it" (Lane 1967: 193). Data gathered by historians also indicate that activities involving lost children, for example, were recorded with some care, thus illustrating the important nature of that task. Monkonnen (1981: 119) argued that the centralized communication capacity of the police made them the primary agency in assisting citizens with their problems. Others have argued, more cynically, that police were more easily understood as paid agents of the dominant order than a public agency aimed at help and service (Harring 1983).

Historical analysis derived from a Boston, MA police officer's diary entries in the year 1895 indicates the panoply of service provided to constituents and the rarity of activity that explicitly involved legal channels (Von Hoffman 1992). Rather daily work would more closely resemble regulatory attention to maintaining social order and largely via service. Even crime-related entries represented mobilizations by citizens who desired the service of the officer in maintaining security or apprehending wrongdoers. The focus on service is likely due in part to some problems not being easily defined as legal or criminal problems and the cost that is borne, in terms of time lost, when an officer has to take the time to invoke arrest, especially on foot. Nevertheless it is imperative to recognize that police have traditionally been asked to address a wide variety of problems as a general service agency, which was probably cast best by Egon Bittner's (1990) hyphenated summary of what becomes police business: "something-that-ought-not-to-be-happening-and-about-which-someone-had-better-do-something-now." Thus, there is an expansive definition, broadly shared among the public, of what constitutes a situation that police ought to handle, even if among the police some of those tasks are seen as not police work or treated with disdain.

Kelling and Moore (1988) make a cogent argument that three eras of US policing, the political, the reform, and the community problem-solving eras, can be mapped to changes in the authorization and the function of the police. In the political era, in which municipal police agencies were formed, the police derived their legitimacy from the community and their function was to serve that community's needs. In the reform era, the authorization of the police was narrowed to matters of law, and the police function was dominated by formal law enforcement. In the community and problem-solving era, which Kelling and Moore were forecasting, it was expected that there would be a return to service in the community and recognition of community needs as authorizing police actions. Clearly, service and helping is at the root of much police work, but its emphasis has undergone dynamic change as agencies have grappled with the role of police in a changing society. In some ways, these differing emphases of police agencies represent discretion writ large, as managers have defined, refined, and redefined "what police do" over time. Such institutional variation over time foreshadows contemporary interorganizational differences in coming to grips with what services should be provided to the public.

Fundamentals

Definition: What Is Service, Help, and Assistance?

How to define police service and helping behaviors is a fundamental question that confronts police researchers and police managers alike. Perhaps it is useful to consider what is not service: Arresting a wrongdoer or using force against a suspect, for example, do not appear to be service. However, under careful scrutiny, this quickly turns on an issue of perspective. If police are acting at the behest of another when applying an arrest or forcefully evicting an agitated spouse from a household, then clearly such actions constitute service, at least from the perspective of the requesting citizen. Mastrofski and colleagues (2000) did much to outline police helping behaviors, especially those in which police were mobilized to use their legal powers against another citizen. Furthermore, arrests of individuals who may injure themselves (e.g., the intoxicated or mentally ill) might, broadly defined, constitute service and help.

The definition of service and help would seem to hinge on *using police power or expertise to advance the well-being of a citizen or a community through formal and informal mechanisms*. Service and help should not be understood as a strict and formal application of the law, law enforcement, or deterrence powers police possess. Rather service should be considered to be dominated by, but not exclusively, the informal application of police power. As noted above, however, helping and service is dependent upon perspective and can occasionally emanate from arrests, citations, and other enforcement actions.

A second approach to defining police service and helping behaviors is to generate illustrations by way of examples. Police researchers, for example, have engaged in studies of how police spend their time and document what police do *for* citizens. These actions include taking reports; providing assistance to motorists; making referrals to agencies (such as a juvenile in need of supervision, mentally ill person in need of psychiatric care, victim in need of counseling, or other person in need of assistance outside of the police expertise), transportation (e.g., for juveniles out after curfew), and directed actions (banish, counsel, arrest) on behalf of a requesting citizen; attending to traffic accidents and directing traffic; counseling citizens about self-protection or self-help actions for crime prevention; and providing comfort and first aid to victims. The list is extensive, illustrates great variety, but it is certainly not exhaustive. Police agencies, however, spend little time measuring or systematically rewarding the extent to which police engage in such activities. Thus, little is known about the extent and variety of ways that the public engages the police in service.

A national snapshot, derived from the National Crime Victimization Survey's sampling frame, has been collected periodically in the form of the Police Public Contact Survey (PPCS). The PPCS contacts approximately 60,000 citizens over age 18 and asks them about contacts with the police during the previous year. The survey has been administered most recently in 2008, where it was determined that police had approximately 40 million contacts with citizens aged 16 and older. Three categories of contact have broad overlap with police service activities: police response to a traffic accident (12 % of total), police contact involving the reporting of a crime or problem (21 % of total), and police contact emanating from a request for assistance (6 % of total). In all these, three contact types represent almost 16 million or 40 % of the police-citizen contacts estimated from the survey results (Eith and Durose 2011). Though ill-defined, these contacts represent a substantial proportion of the public investment in public safety, yet little is known about the content of these contacts.

Measuring Service and Help

What can be measured regarding police service and helping behaviors? This question requires a commitment to a level of analysis, and for current purposes, it would be prudent to consider face-to-face contacts between citizens and the police as the unit for consideration. This allows for aggregation to higher levels, such as neighborhoods and organizations for comparisons, and allows for a more thorough consideration of the encounter between police and the client as the unit of work or service delivery.

This level of analysis leads to consideration of two aspects of helping behavior, one quantitative, the other qualitative. With respect to quantitative help, clearly one is required to count the presence or absence of a report taken, a referral made, a citizen given first aid, or a child given transportation to his family's home. The quality of help (service with a smile?) is less amenable to counting but perhaps more influential in how citizens experience police service. That one aspect of service is amenable to counting makes the complex appear deceptively simple. What should be counted? Mastrofski and colleagues (2000) examined how police respond to requests for help in terms of not fulfilling, partially fulfilling, and completely fulfilling requests by citizens asking police to control another person. Thus an enumeration of requests and whether they are fulfilled could be a measure of police *reactive* service to the public. However, much of what police do may in fact be *proactive* or at least self-directed service to, or in consultation with, the public. For example, police may follow up a case and refer a domestic violence victim to a legal service, a shelter, or some other services as necessitated by situational dictates, victim preferences, and available resources. This proactive approach to service speaks to both counting as something done in a particular case and also the variety of actions that an officer may take.

Capturing variety requires deeper consideration than can be accommodated here, but knowing the possible referrals, resources available to be accessed, and the needs of the individual would be a useful starting point. Thus, accounting for individual instances of services rendered would require attention to both possibilities (referrals to agencies is a good example, since the proliferation and quality varies across cities and even neighborhoods) and appropriateness. Such an approach to measurement would depend on a needs assessment among the service population under study and would be onerous except in the most narrowly conceived execution, such as the services provided by a crisis intervention team to mentally ill individuals encountered. As a general approach to measuring service, it would appear to be too cumbersome to match what police did with what they should have or possibly could have done. Thus, only in specific and well-defined incident or client population types (e.g., the aforementioned provision of service to a mentally ill citizen) would this be a feasible approach. In most efforts to measure the quantity of help, it would appear that counting actions by the police (referrals, assistance offered, aid given) would be the extent to which help could be quantified. Questions of appropriateness and whether needs were addressed become value questions, which would require researchers to know more about what a person needs than may be possible except in all but a few research settings.

As Mastrofski (1999) has noted, the *quality* of service delivery is often as important as the outcome itself. A larger body of procedural justice literature points to quality of treatment and quality of decision-making as being vital elements for understanding police-citizen contacts as antecedents of legitimacy and cooperation. Thus quality of service delivery is separate from whether requests are fulfilled or appropriate referrals are made. Comfort of victims (Foley and Terrill 2008), for example, could be considered a qualitative aspect of the service that police offer to victims of crime. Similarly, respectful treatment of citizens (Mastrofski et al. 1996) is an important part of the quality of service delivery. Measurement of the quality of police-citizen interactions (whether police listen to citizens, whether they inform them of their decision-making thought process, whether citizens are allowed to participate in police decision-making, whether police are fair in their decision-making, and so on) has been conducted both via in-person observations and through post contact surveys. The burgeoning literature on the quality of treatment during police-citizen contacts, even when restricted to service type activities, indicates that more just treatment yields citizens who cooperate and comply with the police. Thus, although service contacts are little studied, they represent an important interface where quality policing can have substantial consequences for future support and cooperation with the police.

Clearly there is variation in the nature, extent, quality, and variety of services that police can offer to citizens in their everyday encounters. The simplest way to conceive of the encounter is that a person was in need of something and police either did nothing or they did something to ameliorate the situation. As the preceding considerations illustrated, measuring help or service is more complex and opens questions of appropriateness, alternative courses of action, extent of help offered, and the manner in which services are delivered. These dimensions should be considered when the measurement of service is undertaken. One may surmise that measuring service, at the encounter level, is difficult to do well and impossible to do perfectly. However, the challenge should not inhibit researchers or police managers, since this aspect of police work is extensive and a deeper understanding would aid in recognizing it, targeting it, explaining it, and making it a more productive aspect of policing.

Explaining Service and Helping

If one had a measure of service and help directed toward citizens, what might explain the choices police make in the delivery of services and help? Since the allocation of services by police, in some sense, represents a legal (if informal) intervention, the literature on how police allocate legal outcomes such as arrests or tickets will be considered as a background for understanding discretionary decision-making in the allocation of service. Service could then be understood as an extension of this behavior, and hypotheses developed to explain arrest, use of force, or police issuance of citations would provide a useful framework when extended to service. Thus, the discussion below adopts the following distinctions drawn from this more general literature: need, legal factors and situational factors, officer characteristics and attitudes, neighborhoods, and organizations. These are arrayed from proximal (at the encounter level) to the distal influences (neighborhood routines, organizational resources) and discussed below.

Need

Mastrofski and colleagues (2000) conducted an analysis involving 396 police-citizen encounters in which citizens requested police help dealing with another person at the scene. An important area of consideration, regarding whether police took actions, was the perceived need of the citizen. This research focused on youthful or elderly status, gender of the requestor, gender of the target, whether the target was intoxicated, and seriousness of the situation as indicators of need. Interestingly, in the analysis of citizen-directed police assistance, only whether the target was intoxicated had an impact on whether police fulfilled a citizen's request.

Regardless of these initial empirical results, several elements of need are considered here as imperative for discussing, generally, what may influence an officer's decision to help a citizen. The seriousness of the situation has dual meanings as a motive for police helping behaviors. One is the seriousness related to legal standards, treated below. The second meaning of seriousness would entail the social or physical consequences or harms facing a citizen, which may have no relation to legal seriousness. Not precisely a "legal" variable, the police would likely tend to ameliorate serious problems via helping. Intoxication of citizens and apparent mental illness are thus likely indicia of need that would motivate help or service from the police.

The mobilization of the police may be aligned with need as well and may be important for predicting the level of service received in an encounter. Predictions of greater service in proactive encounters or reactive encounters, holding other elements constant, are not straightforward. Where police are mobilized by citizens (i.e., reactive encounters), their presence has greater legitimacy for intervention, especially legal intervention (e.g., Reiss 1971). The reactive encounter may also signal a greater need for resolution in the form of help or service to the citizens. There does not appear to be a clear foundation for whether this legitimacy translates into more formal solutions or more service solutions. Proactive encounters, in contrast, may be more likely to yield services as police initiate such encounters to help, assist, or otherwise provide service to citizens in need. More simply, mobilization of police is likely linked with need, but it is unclear if mobilization has a clear relationship to whether police provide service.

Legal and Situational Factors

Several legal and situational factors likely shape whether police engage in helping or service behaviors in a particular instance. First, to the extent there is more evidence of wrongdoing, police likely feel activated and obligated to demonstrate a response. In work by Mastrofski and colleagues (2000), when there was greater evidence of wrongdoing, police tended to respond to requests made by citizens desiring a consequence for the wrongdoer. Second, with regard to more serious legal implication, police are likely to help, to a point, and at the highest level of seriousness, they are likely to have less discretion to bring informal sanctions or solutions to bear on the problems they confront. Thus one may speculate an inverted "u" response shape, whereby at low and very high seriousness police have little discretion. In the middle ranges where discretion is high, they likely would engage in more helping and assisting behaviors. Similarly, the prior history of an individual, in terms of misbehavior, would likely influence whether police engage in helping behaviors or revert to more formal handling of situations. In the modern age of technology, the requestor is likely to be "run" against computerized files and evaluated in terms of history as well, both as a complainant and as a potential target against whom which help is mobilized.

Situational factors cover a wide variety of variables which may influence discretion. For purposes of this discussion, three are considered: race, social class, and demeanor of the target.

Race and ethnicity may be linked with whether police offer more services or help to citizens. Research on arrests, for example, indicates that minorities are more likely to be arrested. In light of those findings, one might expect that minority citizens will receive less informal help and service from the police.

Social class is also linked to criminal justice outcomes, and as such, we might expect that lower social class individuals are less likely to receive services relative to their more well-off counterparts. Clearly neither race nor class should matter, but research on other facets of discretionary police decisions lead to hypotheses regarding race and class influencing decisions to give services to individuals.

Demeanor of the citizen who is seeking or is in need of service is likely more important than either race or social class. Demeanor refers to whether a citizen is respectful or disrespectful toward the police. In many settings, across eras, demeanor has shown to be a relatively consistent and moderately strong predictor of what legal agents do, in the sense that those citizens with disrespectful demeanors tend to get more negative outcomes (more disrespect from the police, more likely to be arrested, and so on). Thus it is expected that citizens with disrespectful demeanor will receive less service from the police, especially if that service is conceived of as a reward such as provision of service or help.

Individual Officer Characteristics

Evidence regarding the effect of officer characteristics on discretionary decisions is mixed but certainly it points to no more than a modest to weak effect on legal outcomes. With regard to how officer discretion might be influenced by personal characteristics, we focus on four: education, experience, gender, and officer's attitudes. Officer education has been shown to have limited effect on performance. With regard to service, it might be expected that officers with greater educational attainment might be more inventive and expansive in their approaches to problems requiring help or service as a possible solution. Research examining whether officers comfort citizens indicates, however, that officers with college degrees provided significantly less comfort to victims when compared to those without college degrees (Foley and Terrill 2008). Similar to the prediction regarding college education, experience, which is arguably the best teacher, should allow for more inventive and helpful behaviors to be routinely developed. These would manifest as possibilities for solving presenting situations and become routine for officers with longer times on the job. Again, however, the research of Foley and Terrill (2008) found officers with more experience less likely to comfort victims. Gender is controversial, in the sense that associating female officers with a greater capacity for helping appears, on its face, sexist. However, Foley and Terrill (2008) found that female officers were no more likely to comfort victims than their male counterparts. Nevertheless, gender may be a proxy for attitudes regarding the expansiveness of the police mandate and reflect both socialization and attitudinal patterns.

Officer attitudes toward work vary and would appear to be important for determining how individuals choose to execute facets of their work. For example, Michael Brown (1988) argued that selectivity and aggressiveness were two key dimensions for typifying officer styles. Among these styles some officers would be much more likely to narrowly define the appropriate tasks for "police work." In fact, Brown argued that there was an identifiable "service style" in which some officers adopted a more helping role rather than strictly law enforcement roles. Similarly, William K. Muir's observations of officers' orientations toward service and helping, in the 1970s, found officers who defined some work as civil issues (e.g., landlord-tenant disputes) and explicitly sought to avoid such situations. The emergence of the community policing model in the 1980s and 1990s more broadly defined police mandates and thus provides cover for officers to adopt more helping orientations toward their work.

Neighborhood Context

The neighborhoods in which police encounter citizens provide the backdrop or setting for face-to-face interactions. Aspects of neighborhoods have been shown to influence what police do during encounters. Those neighborhoods that are characterized by higher poverty, often measured by researchers as concentrated disadvantage which combines indicators of race, poverty, and employment, have been linked with greater likelihood of police arrests and use of force against citizens in prior research. These same neighborhoods are likely to demand greater service from the police. In the aggregate they may receive greater numbers of police helping services. However, at the level of individual interactions between police and citizens, it is likely that those in disadvantaged neighborhoods receive less service relative to those in wealthier or more advantaged neighborhoods. In part this could be a consequence of police formal authority undermining many of the sources of local social regulation. Or it could be that there is an absence of collective efficacy, upon which police can draw informal solutions that can be enforced by informal social control networks in these neighborhoods. In either scenario the literature leads to the prediction that police will, when confronted with similar situations, exercise less helping behavior in the more disadvantaged neighborhoods.

Organizational Context

In a classic study of police at the organizational level, James Q. Wilson's *Varieties of Police Behaviors* examined how departments molded their responses to the political environment. It is telling that in low-crime, more wealthy locales, the police were considered primarily a service function. Though the research is more than 40 years old, the contrast between Oakland, CA, and Nassau County in Long Island and the legalistic style (defining work in terms of application of the law) and the service style (defining work in terms of what the police can do for the public and what the public wants) anticipated the movement toward community policing and the explicit recognition of "what the public wants." This research also indicates that organizational commitment to service likely influences officers' discretion in their day-to-day work.

Contemporary organizational adaptation to servicing public needs may be most comprehensively studied in Chicago, IL, (Skogan 2006) where neighborhood policing was specifically aimed at understanding community needs and trying to accommodate them. The Chicago Alternative Policing Strategy (CAPS) explicitly tapped the community to assess needs and targeted police activity toward problems such as graffiti abatement projects and efforts at removing abandoned cars from neighborhoods. This menu of responses is a far departure from the strictly conceived law enforcement mandate and illustrates how variation at the organizational level is likely to influence police discretion. The mechanism by which this operates is the expanded mandate of what police should do coupled with an effort to understand community needs. Officers working under such regimes are likely to develop a larger palette of helping responses when compared to those working in more legalistic environments.

At the organizational level, one could thus measure variations in how organizations array themselves across a legal/service continuum. Organizational commitment to service could be measured in a variety of ways including information sources collected in assessing needs, varieties of services offered by police, interfaces by police agencies with other service agencies, and the extent to which performance measures are guided by service roles of the agency (i.e., does the agency count "problems solved").

It should be clear that police managers and researchers alike have little direct experience in measuring service and predicting "who gets what." For example, among those variables considered to be important characteristics of officers, experience and college education are both hypothesized to relate to greater levels of helping. However, at least with respect to comfort to victims, Foley and Terrill (2008) drew the opposite conclusion. In more recent research, Rossler and Terrill (2012) examined police responses to citizen requests to file reports, act on their behalf with an agency, provide information, or provide physical help. In these analyses, officers' college education had no impact on whether or the extent to which officers fulfilled the request. But college-educated officers were less inclined to provide an explanation when denying a request. This is a sign that caution should be exercised in forecasting how police may exercise their ability to help in face-to-face encounters and a clear indicator for more data collection, measurement, and theory development. Although some propositions are introduced here regarding need, situational and legal factors, officer characteristics, and neighborhood and organizational context, none of them have been sufficiently tested against observations to a point where one could make statements of any empirical regularity with confidence. The absence of this knowledge is welcome territory for future scholarship but currently a barren landscape from which to make suggestions for police managers, policy for organizations, or to yield strong theory from which to begin analyses.

Current Issues

Considering Problem-Oriented Policing, Community Policing, and Service Responsibilities

Herman Goldstein (1990) observed that when the unit of work patrol for patrol was defined as a "case," then police often missed larger problems that are frequently the underlying generator of calls for service. Such an underlying problem can generate many cases which are resolved by a response, a report, and perhaps an arrest. A drug market, for example, might generate calls of nuisance behaviors and police may respond and disperse suspected dealers, take a report from a complainant, or perhaps make an arrest of an individual for possession, then return to duty. If the unit of work is the call, then this call is satisfied. If the unit of work is defined as a problem, then the problem has not been addressed by the actions described. The problem-oriented policing approach outlines a variety of sources of problems such as homelessness, bar hours, public health hazards such as litter and trash, and a broader palette of incivilities both physical and social that may be considered within the sphere of problems. These larger categories may have an ambiguous relationship to the legal power of police and therefore are easy for organizations and officers to avoid, if so inclined. However, once engaged in problem-oriented policing that is consistent with the scanning, analysis, response, and assessment approach outlined by Eck and Spelman (1987), the likelihood that services and helping behaviors and referrals will be accessed increases. For example, interagency partnerships can form the backbone of these approaches as illustrated by the Beat Health program in Oakland, CA. Here the Oakland police department addressed decay and disorder via partnerships with other public agencies as well as private place managers such as landlords. In this case the service provided is creating a more ordered and civil public space in which conventional citizens may engage in legitimate daily pursuits. Conversely, confronting the problems in the local spaces discouraged the illegitimate use of spaces by, for example, street-level drug dealers. Thus a legal problem was solved by a service approach. So the gray area between civil and criminal was straddled in the Oakland experiment as documented by Mazerolle and Roehl (1999), whereas, two decades earlier, Muir (1977) working with the Laconia (a pseudonym) department found that officers tended to deny responsibility for cases that were civil in nature.

The blending of the civil and criminal authorities within local law enforcement, from a legal perspective, may be another area for future consideration. The nature of public regulation, compliance, and obedience is dramatically different in the 2010s than in the 1970s when the theoretical foundations for much criminal justice research on police and their behavior was being established (Reiss 1971; Brown 1988; Muir 1977). This transformation is undoubtedly related to the community policing movement and its explicit desire to deactivate the legalistic mindset that defined police work in terms of criminal law (e.g., Kelling and Moore 1988). More expansive definitions of police work will broaden what police can do and will also provide opportunities to study and understand why police make choices to help or provide service in some instances, but not others.

Populations in Need of Services: The Mentally Ill

Lurigio and colleagues (2008) have outlined the contemporary scene with regard to the extent to which police process the mentally ill and point out that the jail systems in Chicago, New York (Rikers Island), and Los Angeles represent the largest mental health facilities in the United States. Consideration of the mentally ill clients as persons in need of help mirrors some of the civil/criminal tension that has characterized the consideration of police service and helping behaviors. The police contact with the mentally ill is not a new subject for academic study as these contacts have long been considered an important part of police repertoire and a very difficult problem to handle with only legal tools such as arrest and coercion (Muir 1977). In many ways, bridging service and legal processing for this population has been a continuous challenge for street-level policing. Adoption of crisis intervention teams, collaboration built around psychiatric services and police, and emphasis on training officers for the difficulties inherent in encounters of the mentally ill point to efforts to shape discretion with respect to this particular population. Interestingly the movement has first recognized that police are often the first line of contact for mentally ill individuals, and to handle this exigency, agencies have trained and developed resources targeted specifically toward this population's needs. Equipping police with greater structure and resources shapes discretion but broadens possibilities of what police can do.

Summary and Future Directions

Police have always engaged in helping and service behaviors. Despite this, modern scholars and managers alike are hard pressed to measure, theorize, and evaluate the effectiveness of these actions. This stems from recognition that the mandate upon which helping behaviors rests is clearly not as strong as the law enforcement mandate, which is rooted in law and linked to fundamental issues involving the exercise and control of coercive force. Thus systematic study of helping has been largely overlooked, unmeasured, and, unfortunately, unrewarded. As the public comes to expect more service from police agencies, especially in times of economic hardship and social consequences that accompany it, there is likely to be greater pressure to systematically measure and evaluate this aspect of police work. Among police researchers, similarly, there is likely to be a movement to test propositions about "who gets what" from police in terms of services rendered. As noted here it is possible to speculate about how neighborhood characteristics, citizen race and ethnicity, and citizen social class relate to service delivery patterns; in truth, these are open research questions yet to be thoroughly addressed. Nevertheless, efforts to adopt the community and problem-solving approaches have driven police organizations and the officers within them to embrace a greater service role, absent information on how such service is delivered. Given this situation it is likely that much research will be targeted at best practices of service delivery, especially with regard to services targeted at special populations, such as the mentally ill.

Related Entries

- ▶ Community Policing
- ▶ Measuring Police Performance
- ▶ Police Discretion and Its Control
- ▶ Police Discretion in Law Enforcement
- ▶ Police Family Violence Services
- ▶ Police Legitimacy and Police Encounters
- ▶ Policing Special Populations

Recommended Reading and References

Bittner E (1990) Florence Nightingale in pursuit of Willie Sutton: a theory of the police. In: Bittner E (ed) Aspects of police work. Northeastern University, Boston

Brown MK (1988) Working the street: police discretion and the dilemmas of reform. Russell Sage, New York

Cumming E, Cumming I, Edell L (1965) Policeman as philosopher, guide and friend. Soc Probl 12(1):276–286

Eck JE, Spelman W (1987) Problem-solving: problem-oriented policing in Newport news. Police Executive Research Forum, Washington, DC

Eith C, Durose MR (2011) Contacts between police and the public, 2008. Special report. Bureau of Justice Statistics, Washington, DC

Foley T, Terrill W (2008) Police comfort and victims. Vict Offender 3(2):192–216

Goldstein H (1990) Problem-oriented policing. McGraw-Hill, New York

Guyot D (1991) Policing as though people matter. Temple University Press, Philadelphia

Harring S (1983) Policing a class society. Rutgers University Press, New Brunswick

Kelling GL, Moore MH (1988) The evolving strategy of policing. Perspectives on policing, vol 4. National Institute of Justice, Washington, DC

Lane R (1967) Policing the city: Boston 1822–1885. Harvard University Press, Cambridge, MA

Lurigio AJ, Smith A, Harris A (2008) The challenge of responding to people with mental illness: police officer training and special programmes. Police J 81:295–322

Mastrofski SD (1999) Policing for people. Police Foundation, Washington, DC

Mastrofski S, Snipes J, Supina A (1996) Compliance on demand: the public's response to specific police requests. J Res Crime Delinq 33:269–305

Mastrofski S, Snipes J, Parks R, Maxwell C (2000) The helping hand of the law: police control of citizens on request. Criminology 38:307–342

Mazerolle L, Roehl J (1999) Controlling drug and disorder problems: a focus on Oakland's beat health program. National Institute of Justice, Washington, DC

Monkkonen E (1981) Police in urban America, 1860–1920. Cambridge University Press, New York

Muir WK (1977) Police: streetcorner politicians. University of Chicago Press, Chicago

Reiss AJ Jr (1971) The police and the public. Yale University Press, Connecticut

Rossler MT, Terrill W (2012) Police responsiveness to service-related requests. Police Quart 15(1):3–24

Skogan W (2006) Police and community in Chicago: a tale of three cities. Oxford University Press, New York

Skogan W, Frydl K (eds) (2004) Fairness and effectiveness in policing: the evidence. The National Academies Press, Washington, DC

Von Hoffman A (1992) An officer of the neighborhood: a Boston patrolman on the beat in 1895. JSoc Hist 2(Winter):309–330

Wilson JQ (1968) Varieties of police behavior: the management of law and order in eight communities. Harvard University Press, Cambridge, MA

Police Family Violence Services

Robert C. Davis and Sarah Greathouse
RAND Corporation, Arlington, VA, USA

Overview

In recent years a number of programs have been developed in which the initial police response to a family violence incident is followed by a visit from a trained team of officers or by an officer and a social worker. These "second responders" attempt to help victims find long-term solutions to recurring abuse (e.g. see Dean et al. 2000; Mickish 2002).

Second responder programs are based on the premises that family violence often recurs and that victims are likely to be especially receptive to crime prevention opportunities immediately following victimization. That is, there is a "window of opportunity" during the first hours or days after a crime during which victims feel vulnerable and are willing to seriously consider behavioral and lifestyle changes

(Davis and Smith 1994; Anderson et al. 1995). The response often takes the form of a home visit but other times consist of a phone call to the victim. The team provides the victim with information on services and legal options and (in some models) may warn those perpetrators present at the follow-up of the legal consequences of continued abuse. The purpose of working directly with the victims is to reduce the likelihood of a new offense by helping them to understand the cyclical nature of family violence, develop a safety plan, obtain a restraining order, increase their knowledge about legal rights and options, and provide shelter placement or other relocation assistance. A secondary aim of the intervention with victims may be to establish greater independence for victims through counseling, job training, public assistance, or other social service referrals. The purpose of conversations with abusers is to ensure that they understand that assaulting an intimate is criminal and that further abuse will result in (additional) sanctions.

Key Issues

History of Family Violence Officers

The role of police family violence services has evolved over time. Prior to the 1970s, law enforcement's response to family violence was significantly limited. Departments rarely contained specialized teams to respond to incidents of family violence. Responses by general patrol to family violence incidents were often slow or nonexistent (Ford 1983; Manning 1988). When police did respond to service calls, departmental policies often dictated that officers should only attempt to diffuse the immediate situation and subsequently refer the individuals to social service agencies; arrests of the perpetrator were rarely made and follow-ups to the initial call for service were rarely conducted (Hutchison et al. 1996; Parnas 1967). This lack of attention to family violence issues by law enforcement reflected both societal beliefs about family violence and traditional law enforcement culture. Family disputes were commonly viewed as private affairs, and it was believed that police's role in handling family disputes should be one of mediators and references to social services (Breslin 1978). As such, police officers often viewed responses to domestic disputes as not "real" police work (Buzawa and Buzawa 2003). Surveys of police officers' attitudes toward domestic violence, specifically, showed that police believed arrests to be futile because victims would rarely follow through to press charges and that arresting the male head of household would negatively impact the family as a whole. Police also endorsed the commonly held but incorrect belief that responses to domestic disputes placed the officer in greater danger compared to other calls (Buzawa and Buzawa 2003). As a result, law enforcement devoted little attention to family violence issues; even less attention and resources was devoted to taking actions that reduced further incidents of violence.

The law enforcement response to family violence began to become significantly more proactive in the 1970s and 1980s. Several events were catalyst for the change. First, victim's rights advocacy groups and feminist groups began to challenge the police's lack of forceful responses to domestic disputes and their failure to protect female victims of domestic violence (Buzawa and Buzawa 2003). Second, a series of legal cases in the 1970s and 1980s ruled that police departments could be held liable and face financial penalties if they failed to protect victims of family violence (*Bruno v Codd* 1977; Tracey Thurman et al. 1984). Finally, empirical research around this same time began to demonstrate that the actions taken by police had the potential to reduce incidents of family violence.

One of the most influential empirical studies of police response to incidents of family violence was an evaluation of police departments that employed violence family crisis intervention units (FCIU's) (Bard 1975). Originally developed by psychologists, officers in this program received training in mediating and diffusing domestic disputes. While the officers still responded to all calls for service, they were specifically called in to handle family violence calls. The National Institute of Justice sponsored a training and demonstration program that employed 10 family crisis intervention units in police departments across the country. The evaluation of these programs reported some positive effects of FCIU's, including decreases in arrests and officer injuries (Bard 1975). Although this evaluation also found that calls for service to family violence incidents increased, this finding was explained as evidence of the FCIU's increasing victims' faith in the ability of the police to effectively handle incidents of family violence. As a result of the reported success of these programs, many major US police departments began adopting either FCIU's or specialized units comprised of law enforcement and social service workers (Liebman and Schwartz 1973). Often funded with federal funds, these officers were specially trained in handling victim needs and problems and in keeping themselves safe during home visits. However, responses to FCIU's were not all positive. Liebman and Schwartz (1973) reported that New York City's highly praised FCIU's program had resulted in increased homicides and domestic disputes. Critical research findings, coupled with strategic difficulties in implementing specialized units, led many departments to completely cut or severely limit the role of family violence response units.

In the 1980s, police departments shifted focus from specialized family violence policing services to arrest and punishment of those individuals that perpetrate family violence. Several research studies were major forces in changing the manner in which police responded to family violence. The most prominent of these studies was an experimental evaluation of police responses to domestic violence conducted in Minneapolis, Minnesota (Sherman and Berk 1984). In this study, police who responded to domestic violence calls were randomly assigned to handle the dispute in one of three ways: (1) separate the parties (2) mediate the dispute (3) make an arrest. The researchers reported that making an arrest resulted in significantly fewer further domestic disputes between the couple. Although the study received some criticism from the scientific community (Binder and Meeker 1988; Lempert 1989), the results from this experiment were highly publicized and spurred mandatory arrest laws in many jurisdictions and arrest oriented policies in police departments across the country. This focus on deterring future violent behavior through arrest took precedent over employing specialized family violence units or secondary responder programs.

Although the criminal justice system has been largely focused on reducing repeated incidents of family violence through arrest and prosecution, in recent years there has been recognition that these policies may not be enough to effectively address and reduce incidents of family violence. Research studies have not always reported positive effects for pro-arrest policies (Dunford et al. 1990; Hirschel et al. 1992). In addition, despite a strong emphasis on pro-arrest policies, prior research reports a high chance of revictimization for victims of domestic violence, particularly immediately after the original incident of abuse (Lloyd et al. 1994). This research raised recognition that more needed to be done to assist and empower victims of family violence. As a result, a growing number of police departments began to adopt or re-adopt specialized family violence service units that provide secondary responses to victims following the initial call for service. Although the total number of such programs is not known, we do know a number of specialized programs across the United States and Great Britain have been implemented and evaluated. The research examining the effects of these specific programs will be discussed later in this article.

In recent years, law enforcement has also become increasingly aware of the need for specific units to address a special type of family violence, elder abuse. Traditionally, elder abuse was not considered a serious criminal problem. Law enforcement officers did not often receive specialized training to handle reports of elder abuse, and departments rarely seriously pursued elder abuse cases (Plotkin 1988). Recent research has demonstrated that incidents of elder abuse are much more prevalent than previously believed

(Tatara et al. 1997). As a result, a growing number of departments have established specialized units to handle cases of elder abuse (Heisler 2000). Because cases of elder abuse are often complex, specialized law enforcement often coordinates with local prosecutors and social service agencies to investigate reports of elder abuse and work with the victims.

Training of Family Violence Officers

Historically, police officers have received very minimal training about how to handle calls to incidents of family violence. A survey of law enforcement's training practices during the 1970s revealed that most often police officers received a limited amount of training on handling family violence during a half-day to 1 day training on responding to calls with disturbed individuals (Buzawa and Buzawa 2003). This module of training addressed a variety of situations and didn't solely focus on handling family disputes. Many of the training programs also simply reinforced the idea that police officers should only play a minimal role in handling incidents of family violence. Once on the job, officers rarely received continued training on handling family violence. The most common method of educating young officers came through observations of more experienced officers' responses to incidents of family violence.

Today, more structured training programs focusing on intervention strategies, arrest policies, and attitudes toward domestic/family violence are in existence. The International Chiefs of Police and other national organizations have produced recommended requirements for domestic violence training, and an increasing number of states are requiring their law enforcement officers to receive family violence training (Miller 1997). Although officer training on domestic violence may be more detailed, the amount of training officers typically received has not significantly increased. One survey found that on average officers receive 10 h of initial domestic violence training. Once in the field, most officers are not required to shadow officers who are experienced in handling family violence, and the quality of on-going training programs can vary significantly, from short videos to in-class trainings with role playing activities (Buzawa and Buzawa 2003). A growing number of states are also requiring officers to receiving training on elder abuse (Heisler 2000). Although national organizations such as the Police Executive Forum have distributed recommended practices for training officer on elder abuse, little information exists on the training practices of specific departments. Little information also exists on the typical methods used to train individuals within any specialized family violence or secondary responder units. Because there are not standardized methods for operating specialized family violence service units, methods of training these team members vary by department.

Functions of Domestic Violence Officers

Redlands, CA Police Department had a robust domestic violence officer program that was the subject of an NIJ-funded evaluation. Consequently, detailed information on what these officers did are contained in the evaluation report. The following description of the functions of family violence officers draws heavily on the report by Davis et al. (2010).

In Redlands, a team of officers, including a trained female domestic violence detective, attempted to visit households within either 24 h or 7 days of a domestic complaint depending on the severity of the incident. According to the officers conducting the second responses, contact was made with the victims at their homes in 84 % of the cases in which it was attempted. In cases where the home visit attempt was not successful, literature was left with information about community services.

The visits typically lasted 30–45 min, depending on the victim's receptiveness to assistance. The goals of home visits were to ensure that the victim had information about and access to resources and services, to answer any questions they had about the complaint or the justice process, and to encourage a sense of trust in the police and the criminal justice system as a whole.

A written protocol guided the officer or officers making home visits. The visits began by the officer talking to the victim about the recent incident and any immediate safety concerns that she had. The officer discussed with the victim the nature of domestic violence and the very real possibility that the incident she experienced would recur if no action was taken. The officer tried to make the victim understand that the police department took the matter seriously and was there to assist her. She also asked the victim a series of questions about her relationship with the abuser, history of abuse, and the presence of children and weapons in the home.

Once preliminaries were taken care of, the second response officer tried to ensure that the victim had information about resources and services; offered practical assistance; worked with the victim to develop a safety plan; and instructed the victim in how to document future abusive or stalking behaviors. Before leaving, the officer provided the victim with a written description of local resources to assist domestic violence victims, including housing relocation, counseling, domestic violence shelters, medical help, civil legal assistance, information about the criminal justice process, aid in applying for an order of relief, and emergency financial assistance. Referrals were most often made to counseling programs or parenting classes; smaller numbers of victims were referred to shelters, civil legal assistance, assistance in obtaining a restraining order, and district attorney victim advocates.

One concern about conducting second response visits was that the visit might trigger anger in the perpetrator. To avoid this, there was an intentional practice to call ahead to make sure perpetrators were not present when the officers came to the home. However, since officers were not always able to reach victims by phone ahead of time, there were some instances in which the perpetrator was there when the officers arrived. While partners were home during just a handful of visits, nearly half (46 %) of victims said in subsequent interviews that their partners were aware that the visit occurred. Of these, approximately one in four (28 %) reported that their partner had a negative reaction to the visit.

The Chicago Police Department has a similar second responder program for victims of elder abuse. Each Chicago police district has a designated elder abuse officer whose job it is to provide services to those victims. During the course of their home visits, the elder abuse officers ensure that victims' medical and physical needs (food and shelter) are being met. When they are not, officers arrange for temporary shelter. They make referrals to social service programs, including counseling, independent living services, Meals on Wheels, and programs for Alzheimer's patients. Officers also enroll victims in emergency identification bracelet program so police responding to future incidents can quickly grasp history and information on victim, assist victims in completing forms to receive state compensation, and provide assistance getting to court. Finally, a report is made to adult protective services.

Effectiveness

There have been a number of high quality evaluations of second responder programs yielding conflicting results. Several randomized field trials of second responder programs were conducted in New York City public housing projects. Each tested the same intervention model: persons who reported family violence to the police were randomly assigned to receive or not to receive a follow-up visit from a domestic violence police officer and a social worker. This follow-up visit was not immediate, as is the case with most second responder programs, but occurred an average of 2 weeks later.

The sample for one of the studies (Davis and Taylor 1997) included instances where someone had called the police in response to a family violence incident (this could be violence between romantic intimates, sibling violence, elder abuse, or other forms of violence between persons related or living under the same roof). The incidents were minor in nature (only 7 % of the incidents resulted in arrests and just 14 % of victims reported any form of injury).

Four hundred and thirty-five victims were randomly assigned to receive a home visit as a follow-up to the patrol response. The control group received only the initial police patrol response. Additional calls for police services were tracked for both groups over the next 6 months. At the end of the tracking period, researchers interviewed victims to ask about new abuse, about satisfaction with the police response, and about victims' knowledge and use of social services.

According to law enforcement records, households that received the home visit intervention were more likely to call the police during the subsequent 6 months than households that did not receive the interventions. Yet, according to victim survey data, there were no differences between the two groups in abuse during the 6 months following the trigger incident. In the literature on the effectiveness of arrest on curbing violence, victim reports and calls to the police usually are both treated as imperfect indicators measuring an underlying construct of actual violence. However, the two measures clearly are not synonymous. Many family violence victimizations are not reported to the police. Davis & Taylor interpreted this pattern of results to mean that the experimental interventions did not affect actual violence levels but did increase victims' confidence in the police and made victims more willing to report violence when it occurred.

A second experimental investigation (Davis and Medina 2001) of the same intervention was conducted several years later, this time using a sample of 402 public housing residents who had reported elder abuse incidents to the police. Like the cases in the first field test, incidents in this study were also relatively minor (5 % of the abusers were arrested, just 4 % of victims reported any injuries, and in only 22 % of the cases was a crime alleged to have occurred). Once again, law enforcement records for these households were tracked for the next 6 months. As in the first experiment, it was found that victims who received the home visit intervention called the police sooner and more often than controls.

Pooled analyses of these and a third unpublished experiment indicated that the interventions were associated with an increase in reporting of new abusive incidents not only to authorities (which could indicate simply greater confidence in the police), but also to research interviewers (Davis et al. 2006). The New York field tests suggested that second response programs might actually increase the likelihood of new abuse.

Going into these studies, it had been assumed that the effects of the interventions would be to empower victims through information about their situation, available services, and legal options. The program logic model posited that new abuse would decline as victims extracted themselves from self-defeating relationships or worked with social services and criminal justice staff to develop strategies to end the abuse while staying in the relationship. However, researchers in the New York studies found no evidence that those who received the interventions were more likely to avail themselves of social or legal services, so the intervention could not have worked – at least not in the way intended.

Some evaluations of other second responder programs found results similar to the New York field tests. A subsequent study by one of the authors of the New York evaluations, this one conducted in Redlands, CA, found that households that received a second response had worse outcomes on seven measures of new abuse, although the results were not statistically significant (Davis et al. 2010). Two studies conducted in New Haven, CT by Stover and her associates (Stover et al. 2009, 2010) found between them that persons who received a second response were more likely to call the police again, but less likely to report new abuse on victim surveys. A study by Hovell et al. (2006) found more reports of abuse to the police among victims who received a second response.

Other studies however, reported a positive effect of second responder programs. Greenspan et al. (2003) found that victims who received a second response in Richmond, VA were less likely to report victimization on a subsequent

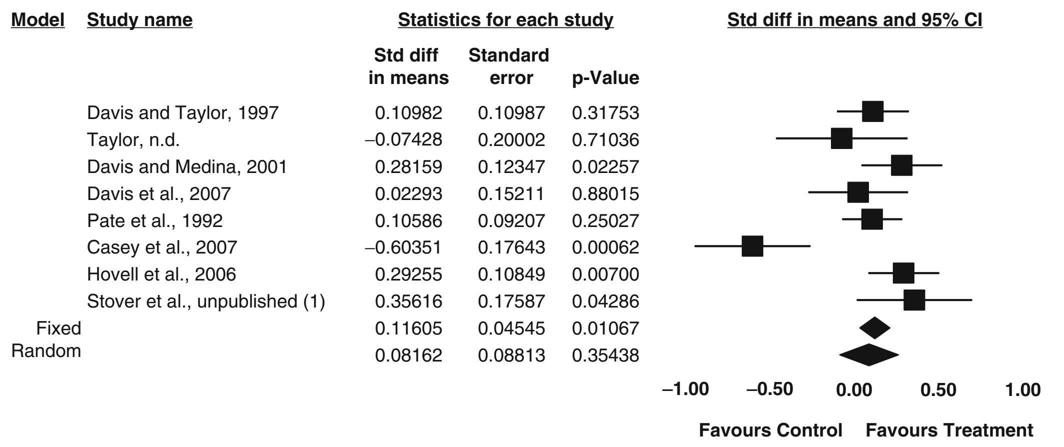

Model	Study name	Statistics for each study		
		Std diff in means	Standard error	p-Value
	Davis and Taylor, 1997	0.10982	0.10987	0.31753
	Taylor, n.d.	−0.07428	0.20002	0.71036
	Davis and Medina, 2001	0.28159	0.12347	0.02257
	Davis et al., 2007	0.02293	0.15211	0.88015
	Pate et al., 1992	0.10586	0.09207	0.25027
	Casey et al., 2007	−0.60351	0.17643	0.00062
	Hovell et al., 2006	0.29255	0.10849	0.00700
	Stover et al., unpublished (1)	0.35616	0.17587	0.04286
Fixed		0.11605	0.04545	0.01067
Random		0.08162	0.08813	0.35438

Police Family Violence Services, Fig. 1 Fixed and random effects models for reports of abuse to the police: experimental and quasi-experimental designs

survey. Pate et al (1992) also found a decrease in subsequent violence reported on a survey following a second response in Dade County, FL. Casey et al. (2007) reported fewer calls made to the police among victims who received a second response in New Haven, CT.

Davis et al. (2008) conducted a meta-analysis of second responder programs. The analysis concluded that the odds of reporting new abuse to the police were about 1-1/4 times higher for households assigned to a home visit treatment. The meta-analysis found no difference between treatments in reports of new abuse on research surveys (see Fig. 1 below).

In sum, then, the weight of the evidence does not indicate that second responder programs reduce new instances of abuse and may, in fact, increase subsequent calls to the police – possibly because the intervention generates more actual abuse or possibly because people who receive a second response have more confidence in the police.

Related Entries

▶ Domestic Violence
▶ Measuring Police Unit Performance
▶ Randomized Experiments in Criminology and Criminal Justice

Recommended Reading and References

Anderson D, Chenery S, Pease K (1995) Biting back, tackling repeat burglary and car crime. Crime Detection & Prevention Series Paper 58. Home Office, London

Bard M (1975) Role of law enforcement in the helping system. In: Coffey AR, Renner JE (eds) Criminal justice as a system: readings. Prentice-Hall, Englewood Cliffs, pp 56–66

Binder A, Meeker JW (1988) Experiments as reforms. J Crim Justice 16:347–358

Breslin WJ (1978) Police intervention in domestic confrontation. J Police Sci Adm 6:293–302

Bruno v. Codd, 396 N.Y.S. 2nd 974, NY Sup Ct (1977)

Buzawa ES, Buzawa CG (2003) Domestic violence: the criminal justice response. Sage, Newbury Park

Casey RL, Berkman M, Stover CS, Gill K, Durso S, Marans S (2007) Preliminary results of a police-advocate home-visit intervention project for Victims of domestic violence. J Psychol Trauma 61:39–49

Davis RC, Medina J (2001) Results from an elder abuse prevention experiment in New York City. National Institute of Justice Research in Brief. U.S. Department of Justice, Washington, DC

Davis RC, Smith B (1994) Teaching victims crime prevention skills: can individuals lower their risk of crime? Crim Justice Rev 19:56–68

Davis RC, Taylor B (1997) Evaluating a proactive police response to domestic violence: the results of a randomized experiment. Criminology 35:307–333

Davis RC, Maxwell C, Taylor B (2006) Preventing repeat incidents of family violence: analysis of data from three field experiments. J Exp Criminol 2:183–210

Davis RC, Weisburd DW, Taylor B (2008) Effects of second responder programs on repeat incidents of family abuse. Campbell Collaboration Systematic

Review. http://db.c2admin.org/doc-pdf/Davis_2ndresp_review.pdf
Davis RC, Weisburd D, Hamilton EH (2010) Preventing repeat incidents of family violence: a randomized field test of a second responder program. J Exp Criminol 6:397–418
Dean CW, Lumb R, Proctor K, Klopovic J, Hyatt A, Hamby R (2000) Social work and police partnership: a summons to the village strategies and effective practices. A report of the Charlotte-Mecklenburg Police Department and the University of North Carolina at Charlotte. North Carolina Governor's Crime Commission, Raleigh
Dunford FW, Huizinga D, Elliot DS (1990) The role of arrest in domestic assault: the Omaha police experiment. Criminology 28:183–206
Ford DA (1983) Wife battery and criminal justice: a study of victim decision-making. Fam Relat 32:463–475
Greenspan R, Weisburd D, Lane E, Ready J, Crossen-Powell S, Booth WC (2003) The Richmond/Police Foundation domestic violence partnership. Police Foundation, Washington, DC
Heisler CJ (2000) Elder abuse and the criminal justice system: new awareness, new responses. Generations 24:52–58
Hirschel JD, Hutchison IW, Dean CW (1992) The failure of arrest to deter spouse abuse. J Res Crime Delinquency 29:7–33
Hovell MF, Seid AG, Liles S (2006) Evaluation of a police and social services domestic violence program: empirical evidence needed to inform public health policies. Violence Against Women 12:137–159
Hutchison IW, Hirschel JD, Pesackis CE (1996) Family violence and police utilization. Violence Vict 9:299–313
Lempert R (1989) Humility is a virtue: on the publicization of policy relevant research. Law Soc Rev 23:145–161
Liebman DA, Schwartz JA (1973) Police programs in domestic crisis intervention: a review. In: Snibbe JR, Snibbe JM (eds) The urban policeman in transition. Charles C. Thomas, Springfield, pp 421–472
Lloyd S, Farrell G, Pease K (1994) Preventing repeated domestic violence: a demonstration project on Merseyside, vol 49, Crime prevention unit paper. Home Office, London
Manning P (1988) Symbolic interaction: signifying calls and police response. MIT Press, Cambridge, MA
Mickish JE (2002) Domestic violence. In: Byers BD, Hendricks JE (eds) Crisis intervention in criminal justice/social services, 3rd edn. Charles C. Thomas, Springfield, pp 77–118
Miller N (1997) Domestic violence legislation affecting police and prosecutor responsibilities in the United States: inferences from a 50-state review of state statutory codes. Institute for Law and Justice, Alexandria
Parnas RI (1967) The police response to the domestic disturbance. Wis Law Rev 2:914–930
Pate A, Hamilton EE, Anan S (1992) Metro-Dade spouse abuse replication project technical report. Police Foundation, Washington, DC
Plotkin MR (1988) A time for dignity: police and domestic abuse of the elderly. American Association of Retired Persons (AARP), Washington, DC
Sherman LA, Berk RA (1984) The specific deterrent effects of arrest for domestic assault. Am Sociol Rev 49:261–272
Stover CS, Poole G, Marans S (2009) The domestic violence home visit intervention: impact on police reported incidents of repeat violence over twelve months. Violence Victims 24(5):591–606
Stover CS, Berkman M, Desai R, Marans S (2010) The efficacy of a police-advocacy intervention for victims of domestic violence: 12-month follow-up data. Violence Against Women 16(4):410–425
Tatara R, Kuzmeskus MA, Duckhorn E (1997) Domestic elder abuse information series. National Center on Elder Abuse, Washington, DC
Taylor B (n.d.) Do home visits reduce repeat domestic abuse calls to the police? Unpublished paper
Tracey Thurman et al. v. City of Torrington, Connecticut, 595 F. Supp. 1521 (D. Connecticut 1984)

Police Federations

► History of Police Unions

Police in Nazi Germany

Herbert Reinke[1] and Thomas Roth[2]
[1]Bergische Universität Wuppertal, Wuppertal, Germany
[2]NS-Dokumentationszentrum der Stadt Köln, Köln, Germany

Overview

Dictatorial regimes, not only in Europe but in many other countries outside Europe as well, have used and still use today the police for terrorizing and for controlling political opponents and for maintaining the power of the regime. The Nazi movement went far beyond this when getting hold of the police after their seizure of power.

While the Nazi regime particularly at its beginning drew on traditional police forces and started from authoritarian goals and strategies of police work, it rapidly generated a radical new kind of police organization, with a specific corps of officers, committed to Nazi ideology, and a radical concept of policing, focused on racial categories and the goal of systematically cleansing society of all groups labeled as deviant.

During the Nazi years, the maintenance of "law and order" and crime control and the persecution of political opponents became increasingly intertwined with policies of racial, often total, exclusion. A police force evolved not only supporting but promoting politics of genocide.

Police in Nazi Germany: Institutional Contexts

Since the beginnings of modern policing in Germany during the nineteenth century, and due to federalism as a core characteristic of the German political system, the police had always been under the command of the federal states. There has never been, except for the Nazi period, a large-scale and centralized national police. "Police" meant, and still means primarily, the polices of the Länder, federal states (i.e., before 1933 Prussia, Bavaria, Saxonia, etc.). This institutional context remained unchanged until the end of the Weimar Republic. Although demands were put forward during the 1920s for more centralization, in particular for fighting crime more effectively, a centralized police apparatus did not emerge before 1933.

In July 1932, the conservative German Reichsregierung, the national government, had the middle-left Prussian government on the basis of an emergency decree dismissed, constituting a kind of prelude to the institutional changes, which took place after the Nazi seizure of power. In 1932, a presumed incapacity for maintaining law and order was given as a reason for the replacement of the Prussian government (senior police officers included) by commissioners, installed by the national government.

After the Nazi seizure of power at the end of January 1933, a leading and prominent figure of the Nazi party, Hermann Göring, was commissioned to direct the Prussian Ministry of the Interior. This position served as a basis for getting hold of the Prussian uniformed police and the Kriminalpolizei (Kripo, Criminal Police) as power instruments for consolidating the Nazi regime. Two emergency decrees, issued in the weeks after the Nazi seizure of power, suspended the basic constitutional rights of the citizens, thus providing the Nazi regime with the capacities for prosecuting ruthlessly its political opponents. These two emergency decrees issued in February 1933 served until the end of the Nazi regime as a pseudo-legal basis for the unlimited police power of the regime.

Another leading and prominent figure of the Nazi party, Heinrich Himmler, became head of the Bavarian Political Police soon after the Nazi seizure of power. Thereafter, until the first months of 1934, Heinrich Himmler gained the command of most of the political police departments of the Länder and finally the command of the Prussian Political Police as well, which had been labeled since the end of November 1933 as "Geheime Staatspolizei." In November 1934, all political polices, existing by that time in Germany, were collected under the directorship of Himmler. In a parallel process, the command over the polices was transferred from the Länder to the national government. This development toward a centralized national police came to a central point, when Himmler was given in June 1936 the title of "Reichsführer SS and Chef der Deutschen Polizei," indicating his dual directorship as head of the SS and the German police. In the aftermath of this entitlement, the centralization and the "Vereichlichung" (nationalization) of the different police branches were further enhanced; under Himmler, a national police apparatus had been established, which had not existed before.

But the centralization of the police was only one aspect on the police agenda of the Nazi regime. A further, very important development was the merging of the police apparatus with the SS, in order to create a novel kind of security force effectively and vigorously serving the ideological ends of the Nazi regime. This meant a growing detachment of the police from its traditional institutional context, while it was more and more interleaved with the organizations of the Nazi movement. Under the organizational roof of the "Reichssicherheitshauptamt," a sort of holding for the different security forces of the Nazi regime, which was established in 1939 after the beginning of the war, this merging of the police with the SS apparatus did find its institutional imprint (Topographie des Terrors 2010; Wildt 2009).

Police in Nazi Germany II: From Dictatorial State Police to Ideological Security Force

After the Nazi movement got hold of the police in 1933, it took them only a few years to establish a new approach of policing, clearly exceeding the politics of persecution and suppression known from other dictatorships and authoritarian regimes being in power in Europe at that time.

Under the Nazis, the police's functions were no longer focused on "traditional" target groups only, such as criminal perpetrators and political opponents. The police became instead increasingly, and in a growing proactive manner, focused on specific categories of people, who, based on racial (Jews, Roma, and Sinti) and to a significant extent on criminal-biological criteria (social outsiders, habitual criminals), were labeled as enemies of the racially defined ethnic community of Germans, the "Volksgemeinschaft" (Herbert 2011).

Policing was conceptualized as a policy of social cleansing, meant to rid the "Volksgemeinschaft" of political and criminal "enemies of the people" ("Volksfeinde") and of all "community aliens" ("Gemeinschaftsfremde"), not fitting into or not willing or capable to adapt to the racial and biological scheme of the Nazi regime. This conceptualization was expressed in the image of the police as a "doctor" cutting out all symptoms of sickness out of the "social body." Racism, concepts of "racial hygiene," and criminal biology, increasingly popular since the 1920s, now penetrated nearly all layers of police work. The identification and persecution of members of "foreign races" and the stigmatization and exclusion of "hereditary" or habitual criminals and of other categories of people labeled as deviant became a core component of a police-driven, for those concerned increasingly deadly social engineering. The new police concept constituted the basis for an active participation of the police forces in the murderous and genocidal politics of the Nazi regime.

Policing during the Nazi regime rested upon a familiar yet specifically radicalized division of labor: While the State Police (Geheime Staatspolizei, Gestapo) organized the repression of political opponents and took part in the elimination of "racial aliens," such as Jews, Poles, or Russians, the Criminal Police (Kriminalpolizei, Kripo) became responsible for the prosecution and the elimination of criminals, social outcasts, and the group of Roma and Sinti. The uniformed police (since 1936 under the label Ordnungspolizei, Orpo), responsible for the overall maintenance of law and order in Nazi society, was involved in the surveillance of everyday life and played an important role in ensuring the populace's conformity and identifying deviants. It thus provided essential assistance to the Security Police making a significant contribution to the Nazi regime's political persecution and social exclusion.

Phases of Development

The historical research on the police during the Nazi regime has identified five phases for summarizing the growing involvement of the police in the Nazi regime. During the first phase, primarily during the months after the Nazi seizure of power, the efforts of the Nazi regime were much focused on getting hold of the police as a power instrument, indispensible for suppressing political opponents and for maintaining the Nazi regime during its first months of being in power. The second phase, consisting more or less of the years 1934 to

1936, could be described as the phase when foundations for a merging of the police and the SS apparatus of the Nazi movement were laid. This merging was started by putting Heinrich Himmler, the head of the SS apparatus, and other SS leaders at the head of the German police; by linking the Kriminalpolizei and the Gestapo under the label of a national socialist Security Police (Sicherheitspolizei); and by placing the Security Police and the Security Service of the SS, the Sicherheitsdienst (SD), under a joint institutional roof. The third phase, that is, the years from 1936 to 1939, was very much marked by specific forms of radicalizations of police practices. During these years, the Kripo and the Gestapo began to implement the racial-biological ideologies of the Nazi regime into police work and started to select systematically those groups labeled as enemies of the "community" for special police measures, for new forms of confinement, and eventually for deportation into the expanding concentration camp system. With the beginning of the war in September 1939 and the establishment of the "Reichssicherheit-shauptamt" in that year until late 1944, the police in its fourth phase of its involvement in the Nazi regime turned into the most important instrument for the completion of the Nazi racial ideologies: By its participation in the killing units of the "Einsatzgruppen," the police engaged as main actor in the mass murder of the European Jews in Europe. So did the units of the uniformed police, especially after the war against the Soviet Union had begun (1941). And back home within Germany, it was the police which summoned the German-Jewish population of the cities and the countryside for the deportation into the concentration camps and which accompanied as guards the trains and transported the victims to the death camps. By the same time, the Kripo again radicalized its practices by sending in more and more people into concentration camps, thus occupying next to the Gestapo a central position in the overall strategy for the cleansing of the "community" from its enemies. When carrying out this, the Gestapo and the Kripo were not only instruments used by the Nazi regime but engaged in it beyond obeying orders by putting forward own initiatives for improving the efficiency of the cleansing of the "Volksgemeinschaft," that is, of mass murder. During the fifth and final phase, which lasted from 1944 to the end of the war and the end of the Nazi regime in 1945, the radicalization of police practices intensified once again. During this final phase, the Security Police focused its control efforts on the foreign workers, who lived by that time in enormous number in Germany and had mostly been forced to work for the "Third Reich." Moreover, indicators on increasing discontent and disloyalty among the Germans within the "Reich" turned into a major concern for the Security Police. Starting in the last months of the war and the regime and continuing in its final days, the Security Police tried to cope with the decreasing power of the regime and the increasing disloyalty of the population with a dramatically growing brutalization, resulting in the mass murders of foreign workers and political opponents and selected killings of ordinary Germans no longer loyal to the regime during the final days of the war.

The Security Police and Its Personnel

In order to turn the Security Police into the police of the "Volksgemeinschaft," it became organizationally realigned, centralized, and provided with new hierarchical structures, guidelines, and standards. The transformation and renewal of the staff played a key role as well. While the personnel of the Kripo with about 13,000 employees was mainly transformed with the normal turnover of staff, the Gestapo was established fundamentally new. On the basis of the political police of the Weimar Republic, with its approximately 1,000 officers, the Gestapo grew to an apparatus with over 30,000 employees until the end of World War II.

Senior positions within the Gestapo and the Criminal Police were held since 1933 by men who openly showed their loyalty to the Nazi regime or had even been active in the Nazi movement for a long time already. The few police officers, who had declared themselves against the

Nazis and had represented a decidedly republican, democratic police during the Weimar years, were dismissed or demoted. New police recruits had to indicate, at least through membership in the Nazi party, that they were supporting the regime permanently. The police training combined political-ideological indoctrination and professional police instruction. Finally, with the inclusion of police officers in the SS, and with joint spheres of action and social contacts between the Security Police and the SS Security Service (SD), close relations between police and the Nazi apparatus emerged (Banach 1998; Browder 1996).

In this way, the Sipo established the model of a new type of police officer, combining expert knowledge and professional ambition with ideological radicalism and the commitment of a "political soldier." This new type of police officer became a dominating pattern inside the head office of the Gestapo and the Criminal Police in Berlin, as well as among the senior officers of the regional headquarters of the Sipo branches. This leadership group consisted mainly of younger cadres with middle-class background, well-educated, academically trained, often active in extreme right-wing organizations before 1933 and striving for implementing the societal models of the Nazi regime (Wildt 2009). Another form of "Nazification" took place among the lower ranks of the Security Police, particularly within the Gestapo, which, due to its growing staff requirements, recruited members of Nazi organizations with low-level formal education only and without specific police skills. These activists played an important part in the implementation of the Nazi movement's political and racial stereotypes and its violence into police practice.

Despite these fundamental changes, patterns of regular, traditional police work remained present in the Sipo (Dams and Stolle 2012; Roth 2010). Established procedures and routines of criminal investigation and administration were not abolished but pursued for the purposes of Nazi regime. In addition, older senior detectives, who had entered the police service during the German Empire or after World War I, played a significant role even after 1933. Although the Nazi leadership had promised to purge the police of politically "unreliable" officers after the seizure of power, the rate of dismissals in 1933/1934 was relatively low (<5 %), partially due to the lack of candidates for replacing the officers to be dismissed. Thus, detectives, who were professionally trained and socialized in the Weimar Republic, had a significant impact on the apparatus of the Sipo, in particular during the early years of the Nazi regime. In the middle and lower ranks of the Sipo, the local posts, and particularly inside the Kripo, they remained an influential and supporting element until the end of the regime. With their experience and technical knowledge, they contributed significantly to the functioning and effectiveness of the Security Police.

Reasons for Support and Initiative

The support that the Nazi regime gained from various groups of police officers was due to political indoctrination, opportunism, and careerism; instead, shared values and professional expectations were key factors in that respect as well. Nationalism, antiliberalism and anticommunism, radically stereotyped views on political outsiders or marginalized groups, authoritarian conceptions of state intervention, and social order ("law and order") had been part of the mainstream police philosophy during the 1920s. The rejection of parliamentary democracy and of the rule of law, demands for extensive and harsh control of political opponents, social outcasts and criminals, and the expansion, unification, and centralization of the police forces came upon a broad consensus within the police apparatus. The Nazis reacted positively on the demands and expectations circulating inside the police by suspending or abolishing existing legal restrictions and by giving the police the means for realizing their ideas of radical social control and social cleansing. In the "fight against criminality," against "enemies of the Volksgemeinschaft" or against "antisocial elements," the police and Nazi movement, senior detectives, and the new leadership of the Sipo shared the goals to pursue (Paul and Mallmann 1996; Wagner 1996).

As far as the cultural integration of the police in the Nazi state was concerned, celebrations or public events played a role, too. Of particular importance was that the Sipo was given a new image, which was perceived positively by most police officers. While public criticism of the police nearly disappeared after the suppression of the free press, the Nazi regime and senior Sipo officers developed an image of the police, which depicted the detectives as modern, scientifically trained, effective investigators and provided the Gestapo with the myth of an omnipresent and omnipotent power apparatus.

For the participation of Kripo and Gestapo officers in the terror of the Nazi regime, other factors and processes were important as well: increasing routine and familiarity with practices of exclusion and violence, the dehumanization of the victims through propaganda and bureaucratic procedures, and division of labor and responsibilities and social distance to those persecuted, but also the emergence of a specific Nazi 'cop culture' that was based on harshness and soldierly masculinity, camaraderie, and group pressure.

The Security Police: Methods and Instruments of Power

The alignment of the Sipo with the overall policies of the Nazi regime and the emergence of a concept of "racial policing" was accompanied by an expansion of instruments of power and a radicalization of police methods. Limitations of police work by legal bonds and external controls were abolished, and fundamental civil and human rights were suspended. Those affected were exposed to nearly unrestricted police arbitrariness.

While the Gestapo and Kripo detectives continued to work as a part of the law enforcement process and as an investigative administration for the public prosecutor, an independent police penal justice steadily expanded since 1933.

Its base was the introduction of an independent police custody, the so-called protective custody (Schutzhaft) used by the Gestapo and the "preventive custody" in the case of the Kripo. Both instruments allowed for detention without formal criminal justice proceedings and judicial review and resulted in permanent confinements of regime opponents and offenders in prisons or concentration camps. Those affected by these forms of confinement could be exposed to permanent drill, exhausting labor, hunger and disease, corporal punishments, and mortal danger. "Protective" and "preventive custody" enabled the police to punish, even on the ground of mere suspicion and poor evidence. Both forms of custody were used to "correct" court decisions and became instruments used massively for the intimidation and elimination of political opponents and people labeled as deviant. By applying this police custody, Gestapo and Kripo developed a policy of social exclusion that, until 1945, took several hundred thousand people into the concentration camps.The radicalization of police methods went even further within the Gestapo. While the use of physical violence against suspects by the Criminal Police was rather the exception, the Gestapo had the right to use violence for extorting statements. Against the backdrop of World War II, the Gestapo implemented its own prison camps, the so-called corrective labor camps (Arbeitserziehungslager), where prisoners were usually held for several weeks. In the 1940s, groups such as Jews, Poles, or Russians were completely removed from the realm of the judiciary and brought completely under the control of the Gestapo. Toward the end of the war, local and regional Gestapo offices were finally empowered to execute prisoners.

Arbitrary arrests and deportations, penetrations of the private sphere, extensive surveillance measures, and the use of informers (V-Männer) were the main features of the work of the Security Police, but routine police methods of investigation, detection, and forensic collection were continued as well, being the basis for the deployment of police terror. Despite the image of being omnipresent and omnipotent, the police force was dependent on the support of other institutions of the Nazi state. When heading for a comprehensive control of the "Volksgemeinschaft," the Sipo's own resources and personnel were insufficient (Gellately 1990;

Paul and Mallmann 1996). But the Security Police could count on the cooperation of other instances of social control: Together with subunits of the Nazi party, the uniformed police, government and municipal administrations, health and employment services, the tax administration, postal and railroad services, as well as the customs authorities and large parts of the economic sector, a cooperative network for enhancing and enforcing the policies and the terror of the Nazi regime was created. Significant support came also from the public. For controlling and suppressing core political opposition groups such as communists, or for controlling so-called criminals, the Gestapo and the Kripo could fall back on their own investigative work, but for the discovery of small individual deviances and minor violations, for the observation of everyday life, for the penetration into the privacy of suspected people, and for the pervasion of so-called closed milieus, the Sipo had to rely on denouncers and informers. For the discovery of a spontaneous protest, for the exclusion of the Jews from the "German" society, and for the close observation of minorities, input from the population and party organizations was essential. The power of Gestapo and Kripo derived not only from the results of its own investigations but from a network of individual and institutional supporters.

Key Aspects of Gestapo Work

Central to the work of the Sipo were permanent dynamics of radicalization (Dams and Stolle 2012; Mallmann and Paul 2000). When heading for an overall cleansing of the "Volksgemeinschaft" by eliminating political opponents' "racial enemies" and all sorts of deviance within the population, new enemy groups and accordingly new security requirements were continuously constructed. Initially the most important activity of the Secret State Police was the suppression of the left-wing workers' movement, in which the regime saw the main potential for opposition and unrest. The Gestapo's attack was directed primarily against communists, moreover against socialists, social democrats, and representatives of the Free Trade Unions. By 1933, tens of thousands of opponents have been arrested, mistreated and tortured, handled by criminal courts, or transferred to – by that time – early forms of concentration camps. Other measures were repeated raids, searches, and the confiscation of literature, printing machines, and illegal publications; the ban of organizations and newspapers; and the closure of the workers' movement's associations. Although the left opposition was constantly trying to reorganize illegal groups, the Gestapo managed to dismantle their resistance until 1936/1937.

Additionally, the Gestapo also developed a supporting role in the control of the Christian churches and religious groups. The aim was to stem criticism from the churches, to promote the adaptation of the believers to the Nazi ideology, to force the churches out of the public, and to reduce their impact on the population. Here, the repressive approach of the Gestapo was clearly more differentiated than toward communist resistance and was aimed mainly at integration into the "Volksgemeinschaft." While in Protestant-dominated areas, the Gestapo only monitored a critical minority within the church; in Catholic areas, where the Gestapo suspected a widespread reluctance to Nazism, the State Police developed a wide range of measures: It reached from the observation of meetings and church services and larger propaganda campaigns against the church to the ban on Christian associations and the punishment of individual representatives of political Catholicism and dissident clerics.

Another important field of Gestapo work was the control of public opinion. In addition to the suppression of critical statements from the labor movement and the churches, the communication of "ordinary" citizens should be regulated. Based on new decrees and laws putting the criticism of the representatives and actions of the government and the Nazi party under punishment as "treachery" (Heimtücke), the Gestapo initiated tens of thousands of investigations. Since the beginning of the war, expressing doubts about the Nazi propaganda or pessimistic statements about the war were persecuted more intensely and partly criminalized as "undermining of the military strength" (Wehrkraftzersetzung). The same was true of the media consumption of the population, as listening to foreign radio stations and distributing demoralizing information could be persecuted since 1939 under the label "radio crime" (Rundfunkverbrechen).

Since the mid-1930s, the State Police exerted increased pressure on nonconformist juveniles who rejected national socialist cultural hegemony by insisting on their particular subcultures and by turning against the Nazi youth organization. In addition, the Gestapo initiated wide-ranging prosecution actions against homosexuals, whose behavior was considered a threat to Nazi population policy. It punished violations of economic policy and labor discipline as well and was responsible for criminal proceedings against deviant members of the Nazi movement. On all these fields, the Gestapo sometimes acted in competition but usually in close consultation with the Kripo.

In several of these fields of action and against occasional violations of the norms of the community the Security Police, especially during the prewar years, acted selectively (Gellately 2001; Roth 2010). So it not only applied hard, negative sanctions but used warnings and lighter penalties to intimidate and discipline violators, in order to make them get in line with the "Volksgemeinschaft."

As far as the German-Jewish population was concerned, the police aimed at a radical marginalization and exclusion. While the political, social, and economic exclusion of Jews had been enhanced during the first years after the Nazi seizure of power very much by Nazi party subunits, the Gestapo thereafter took more and more the initiative. Since legislations in 1935, which turned German Jews into second-class citizens, the Gestapo went into organizing the discrimination, isolation, and exclusion of this population group. The police conducted the registration of German Jews, restricted their freedom of movement, were continuously involved in the criminalization of relations between Jews and non-Jews, and also contributed to the economic expropriation of Jewish citizens. As of November 1938, the Gestapo began with the systematic exclusion from German society. During and after the November 1938 pogroms, the Gestapo throughout Germany arrested more than 30,000 German-Jewish men and deported them into concentration camps, where the victims were released only when they had agreed upon abandoning their properties and upon leaving Germany. Just a few weeks earlier, in a nationwide campaign, the Gestapo had arrested about 17,000 Polish Jews and deported them over the German-Polish border – a test run for the mass deportations of the 1940s.

From the spring of 1941, the Gestapo organized the final spatial segregation of the remaining Jews within Germany. In the cities Jews were evicted from "Aryan" houses and sent to special "ghetto houses" or local camps. People of Jewish origin, who were kept in asylums, prisons, and other institutions, were separated from other inmates and handed over to the Gestapo. From autumn 1941, the State Police organized the deportations from the "Reich" to the ghettos and death camps in occupied Eastern Europe. By 1943 some 150,000 people were deported from Germany, while only a few were able to survive as "Mischlinge" (crossbreeds), partners in so-called mixed marriages or in hiding.

The Work of the Kripo

Unlike the Gestapo, the Criminal Police was initially focused more on traditional police activities such as criminal investigation and crime detection. During the 1930s, however, its practice was more and more determined by the methods and goals of a police of the "Volksgemeinschaft" (Browder 1996; Wagner 1996). In late 1933, the senior ranks of the Kripo opened a new field of activity under the label of "preventive fighting against crime" (vorbeugende Verbrechensbekämpfung). Under the pretext of reducing crime, the Kripo rendered recidivists into "preventive detention" by deporting them for an indefinite time into concentration camps, even in the absence of a specific offense or criminal court proceedings. Since 1934, the policemen were also enabled to subject known criminals to a "systematic observation" ("planmäßige Überwachung") and rigid rules of conduct.

At the beginning, the policy of preventive fighting against crime targeted at a relative small group of a few thousand criminals, which should be kept from further criminal activity by deterrence and confinement. With the vision of a "Volksgemeinschaft" without crime in mind, the Kripo constantly enlarged its net. According to a nationwide campaign of arrests against 2,000 so-called professional and habitual criminals ("Berufs-" and "Gewohnheitsverbrecher") in March 1937, the local Kripo offices started systematically detecting recidivists and sending them in very large numbers to concentration camps. The criminal deportations soon were directed not only against professional burglars and violent or sexual offenders but met more often petty criminals. At the same time, the persecution of homosexuals now shifted from the Gestapo to the Kripo. Even when acting against other social outcasts, the Kripo officers acted with increasing harshness. By doing so, the Criminal Police took up and radicalized the strategies and practices, pursued already by labor, social, and health administrations. As early as 1933, the Kripo had initiated numerous raids against prostitutes, beggars, and vagrants, demonstrating that social "disorder" and public "indecency" should no longer be openly tolerated. The next step came when the practice of "crime prevention" in December 1937 was extended to marginalized groups, and the Kripo was empowered to apply indefinite confinement on all individuals "endangering the general public by asocial behavior." Thereupon the concentration camp became a common measure against outsiders, while the Kripo turned into the dominant actor of social exclusion (Ayaß 1995; Gellately and Stoltzfus 2001).

In 1938, the police leadership prompted two nationwide sweeps against so-called work shy, in the course of which the Criminal Police arrested more than 9,000 people (the Gestapo about 1,500). Since then the local Kripo posts consistently put more and more "asocials" into "preventive detention" – next to prostitutes, homeless and vagrants welfare recipients, alcoholics, men neglecting their family duties, women labeled as promiscuous and women with venereal diseases, dealers, peddlers, etc. Even biographically, "crime prevention" radicalized: Since the early 1940s, 1,000 "criminal and antisocial youths" were interned, terrorized, and biologically examined in special camps. Especially involved was the female unit of the Kripo (Weibliche Kriminalpolizei), which not only served as kind of welfare police for (female) adolescents but was responsible for the elimination of "criminal offspring." Until the end of the Nazi era, the Kripo under the label of "crime prevention" delivered about 80,000 offenders to the terror of the concentration camps (Wagner 1996).

In addition, there was the fight against "alien races." The Criminal Police assisted the Gestapo in its anti-Jewish policy, for example, by sending Jews with criminal records to camps or persecuting relationships between "Aryans" and "non-Aryans" as "race defilement" ("Rassenschande").

To core projects of the Kripo belonged the exclusion of the Roma and Sinti, who, as "gypsies," had been particularly discriminated against and placed under special surveillance long before 1933 (Zimmermann 1996). Since the mid-1930s, they were deprived of civil rights, prohibited further mobility, separately registrated, examined by racial experts, and marked as "racial aliens." After May 1940 the Kripo started several deportations, taking thousands of German and Austrian Roma and Sinti to the annexed and occupied Poland, where most of them lost their lives. From March 1943, about 23,000 "gypsies" from the "Reich" and occupied Europe were transported to the extermination camp at Auschwitz. As the practice of Gestapo, the work of the Kripo since the 1940s was closely associated with national socialist mass murder.

Gestapo and Kripo During World War II

The beginning of World War II provided additional rationalizations for the radicalization of policing and increased the willingness of policemen to act harshly and murderously against political, social, and racial "enemies." With the expansion of the Nazi regime into large parts of Europe, new spaces of social control and eliminatory rule were created, while millions of people who were defined as dangerous by the Nazi regime fell in the hands of the Gestapo and the Criminal Police (Mallmann and Paul 2000; Topographie des Terrors 2010).

The further radicalization of the Sipo occurred in the occupied territories, where the Sipo exercised unprecedented and unbound physical violence. Since September 1939, more and more Gestapo and Kripo detectives had to leave their offices within Germany and put themselves at the service of Nazi occupation and warfare. Some of them participated in the "Einsatzgruppen," who entered the occupied territories behind the Wehrmacht to "cleanse" the conquered areas of indigenous elites, oppositional groups, Jews, and minorities like gypsies and the disabled. The "Einsatzgruppen," already established as mobile killing units in the Polish campaign 1939 and considerably extended since the invasion of the Soviet Union in 1941, have been responsible for the direct killing of hundreds of thousands of people especially in Eastern Europe (Klein 1997). In these units, and the offices established in the occupied territory, the institutional merging of police and SS was further promoted.

Members of the Security Police were involved at all levels in Nazi crimes: in the invention of killing techniques for the gas vans and death camps and the "administrative execution" of genocide; in the surveillance of ghettos and economic plundering of the occupied countries; in reprisals against political opponents, mass executions of Jews, political officials, prisoners of war, or "partisans"; and in the killing of "criminals," "asocials," or "gypsies." In the violent environment of the East, it could be practiced without restraint and directly, what the police in the "Reich" before the war could achieve only indirectly through deportation and imprisonment into concentration camps: the annihilation of those categories of people defined as enemies of the "Volksgemeinschaft." When doing this, the Security Police was not only supported by the Germany army or the civil administrations of the occupied countries. An important element of German police power in Europe was the collaboration of local police forces, radical nationalist organizations, and auxiliaries, which contributed significantly to the execution of mass murder.

The Security Police on the "Home Front"

In the course of World War II also inside Germany, the Sipo radicalized its scenarios of crucial enemies and possible dangers (Mallmann and Paul 2000). Political opposition and social deviance were now seen as a threat to the Nazi war effort and the stability of the "home front." But while this resulted in a further tightening of sanctions, gaps in the system of police control were widening, as many officers were withdrawn from the "internal front," the police functions were continually extended, and the prolonged war and ongoing air raids, especially in big cities, lead to a decline of law-abidance and loyalty to the regime. The impending loss of control in the war did not lead the Security Police to a strategy of de-escalation but was answered with a further intensification of terror. To this added, that police officers who had returned from the occupied territories brought their experiences of violence and concepts of "enemy combat" with them to the "internal front." Due to the expansion of control requirements and the social distractions of war society, the Kripo focused more and more on the disciplining of those refusing to work for the war community, on surveilling the conduct of young people and single women, on halting the spread of sexually transmitted diseases among prostitutes and soldiers, or on the persecution of the black market, "looting" and other war-related crimes (Roth 2010; Wagner 1996). These control efforts also met conservative and bourgeois opponents of the regime – as the mass arrests after the attempt to assassinate Hitler in 1944 illustrate.

However, on the "home front" within Germany, racial politics stood at the center of the State Police's activity. In addition to the exclusion of the Jewish population, the control of "foreign workers" became a main task of the Gestapo – especially since the intensification of the use of foreign workers in the German Economy (Ausländereinsatz) and the massive recruitment of forced laborers from the occupied Soviet Union since 1942. The Gestapo not only had to monitor the work discipline of foreign workers and to ensure the functioning of the war economy, it also was to suppress resistance and sabotage and punish forbidden relations between Germans and foreigners. That was mainly applied to workers from Poland and Soviet Russia: They were classified as "racial aliens" and placed at the bottom of the racial hierarchy of the Nazi system, were regarded as a special threat to the order of the "internal front," and were almost entirely subject to the power and violence of the State Police. Since the middle of the war, the measures of the Gestapo even included the killing of East European laborers. Numerous Polish, Ukrainian, or Russian civilian workers, who have had intimate relations with locals, were publicly executed due to transgression of "racial barriers" (Gellately 2001). This indicates what State Police practice characterized in the last months of the war: A policy of open terror, which brought mass murder to the "home front." In the final phase of the war, the Gestapo inside the "Reich" increasingly followed the models and methods of occupation. Hierarchies, bureaucratic procedures, and the formal allocation of responsibilities became less important, as much of the work of the State Police was taken by Sonderkommandos (special task forces) operating with high mobility and flexibility. They worked increasingly independent to the central offices in Berlin, with almost unlimited instruments of terror. In the last months of the war, executions, initially dependent upon authorization from Berlin, could be imposed by regional commanders of the Security Police and eventually the local offices themselves. Before the collapse of the Nazi system, the work of the Gestapo in the "Reich" was marked by mass arrests, targeted killings, and "combat missions." They were directed primarily against "foreign workers," young people, deserters, and criminals who were in hiding in the ruins of the big cities. Since the Gestapo with its flexible organizational structures and unlimited powers assumed nearly all of the Security Police's competencies, the officers of the Criminal Police lost importance and influence in the last months of the Nazi regime. However, at least some of the Kripo officers participated in arbitrary killings and supported the "final battle" of the Gestapo.

The Uniformed Police: Everyday Practices and Participation in the Holocaust

During the Weimar Republic Period, the uniformed police tried to change the image it had in the public. While an image, deriving from the police before World War I, lingered on, which portrayed the police very much as authoritarian, militaristic, and emphasizing the social distance between the uniformed police and the general public, senior police officers tried to change this image for a portrait, which displayed a police much closer to the public, a police, which was no longer primarily a protector of the state but a protector of the "people" as well. After the Nazi seizure of power, the leadership of the police attempted to give this portrait even more contours by popularizing a label, which showed "The police, your friend and helper." The uniformed police organized during the 1930s and even during the first years of the war a number of "Tage der Polizei" (days of the police), which, as a public relations action, were to propagandize the label. Other activities, inaugurated by the police during the 1930s, focused on improving the traffic circulation in the cities and on preventing the growing risk of traffic accidents, by training the general public, in particular children, how to behave adequately on the roads of the city. All this was intended to make the slogan "The police, your friend and helper" popular, but this label covered only partially the reality of the uniformed police during the Nazi regime. The uniformed police was "friend and helper" only to those, who were members of the "Volksgemeinschaft." Jews, for example, were not members of this community, which comprised only Germans along very narrowly defined social, racial, and biological criteria. Some of those, who were opponents of the new regime, or were not members of the "Volksgemeinschaft," experienced immediately, others very soon a new repressive side of the uniformed police and a new specific, selective approach to law and order, influenced by the political strategies of the Nazi regime. The uniformed police did not intervene, when political opponents of the Nazi regime were arrested, maltreated, or even killed but conceded instead to a partition of labor with the police doing the work, which looked professional and members of the paramilitary groups of the Nazi movement, often in the position of auxiliary police personnel, doing the "dirty" work, such as torture and killings of political opponents. And in 1933, soon after the Nazi movement had come into power, the police did not intervene, when the Nazi movement staged a boycott against German-Jewish shopowners and merchants, thus indicating law and order might not pertain to German Jews any longer. These attitudes became increasingly enhanced, when a growing loading of police training with patterns of Nazi ideology and anti-Semitic contents occurred (Matthäus 2003; Westermann 2005).

Institutionally and organizationally, the uniformed police experienced a drawback, when in 1936, after a organizational reconstruction, which had started very early after the Nazi seizure of power already, the units of the uniformed police were transferred to the German army to increase the army's numerical strength. This weakened the organizational resources of the uniformed police and reduced the number of its personnel to about fifty percent of its original strength. Until the end of the 1930s, the uniformed police gradually managed to increase again the number of its officers and to reestablish police units. The personnel of these newly established units, later named as police battalions (Polizeibataillone), was comprised to large extents of police reserve recruits without any professional police background instead of professional police officers. These police units were used alongside the military, when the Nazi regime annexed Austria (1938) and Czechoslovakia (1938/1939). These participations of units of the uniformed police still ran under the heading of securing the occupied territories, but with the start of the war against Poland in 1939, the uniformed police turned into an instrument executing the racial and genocidal ideologies of the Nazi regime. The uniformed police finally became a murderous key institution in the Holocaust. Already during the starting phase of the war against Poland, the units of the uniformed police, following the advancing German army, committed atrocities among the Polish population, Jewish and non-Jewish, thus becoming part of the racial strategies the Nazi regime started pursuing already during the first phase of the war. Especially the violent transfer of populations, the settlement of "ethnic Germans" in the occupied parts of Western Poland and the brutal eviction of Polish and Polish-Jewish inhabitants of these parts of the country into the occupied central Poland (the "Generalgouvernement") became one of the first elements of the participation of the uniformed police in the geostrategical and racial policies of the Nazi regime.

With the beginning of the war against the Soviet Union in June 1941, the involvement of the units of the uniformed police gained unprecedented genocidal dimensions. The so-called Einsatzgruppen, operating in the hinterland of the advancing German army within the Soviet Union, killed more than 500,000 victims until the end of 1941, most of them Jews. These mobile killings units not only were composed of members of the Security Police and the Security Service of the SS (SD) but could rely to large parts on military SS-formations (Waffen-SS) and members of the uniformed police. These mobile killing units advanced within the Soviet Union as far as the German army, that is, into the territory of the former Baltic states, and from there close to Leningrad, they covered Belorussia and moved forward into the vicinity of Moscow, covered the whole of Ukraine and proceeded as far as the northern fringe of the Caucasus. Most of the direct killing actions and massacres on the Eastern Front, as the killing of about 33,000 Jews from Kiew in Babi Jar, were carried out by the "Einsatzgruppen" and supporting units.

While the overall number of members of the "Einsatzgruppen" remained relatively small (there were up to 3,000 men in the "Einsatzgruppen" following the advancing German army in the war against the Soviet Union), the overall figures for the members of the units of the uniformed police participating in war, occupation and genocide were much higher: In occupied Poland and in the occupied parts of the Soviet Union, approximately 50,000 men served in different units of the uniformed police, the number of these units amounting to 90. This high figure of men serving in mobile units of the uniformed police was due to the fact that the Order Police was not limited to its participation in the murderous actions of the "Einsatzgruppen." The units of the uniformed police carried out a great number of executions of their own, quite often when the Ghettos in the cities of Eastern Europe, where the Jewish population was forced to live in, were dissolved and its inhabitants being sent into the death camps. But the radius of action of the uniformed police was not limited to Eastern Europe. In most of the countries occupied by Nazi Germany, units of the uniformed police held a significant position in the occupation apparatus, often charged with a more general maintenance of the security in the occupied countries. In the later phases of the war, units of the uniformed police became heavily involved in fighting guerrilla activities against the German occupation ("Partisanen-" or "Bandenbekämpfung," terms often used for masking executions of Jewish men, women, and children; the elimination of resistance groups; and the brutal suppression of the local population). During the final phases of the war, units of the uniformed police took part in the indiscriminate mass killings of hostages all over Europe, leading to the extinction of whole village populations (Oradour-sur-Glane).

The overall death toll, for which the uniformed police of the Nazi regime could be held responsible, is difficult to calculate precisely. But calculations in different publications, older ones (Gutman 1990) and more recent ones (Curilla 2011), count approximately 1,200,000 victims, who were either murdered by the units of the uniformed police or by members of this police as participating in mass executions as members of other units ("Einsatzgruppen").

After 1945: Research and Institutional Reflections

Elaborated research about the role and the functions of the police as part of the Nazi regime did set in late in Germany. The few studies available until the 1980s either gave limited descriptions of the police's institutional settings and the development of this settings during the years from the Nazi seizure of power until the end of the war, or they focused on the Gestapo, emphasizing a picture of the Gestapo as omnipresent and omnipotent, terrorizing political opponents and the so-called racial enemies of the German people and, since the beginning of the war, covering the whole of Nazi-occupied Europe with its murderous activities. This early research concentrated very much on a top-down perspective, by looking primarily at the main goals, leadership, and institutional context of the Gestapo's activities without analyzing the whole personnel involved and without detailing the Gestapo's impact and societal background. This changed for a bottom-up perspective, when the social history turn within historiography touched upon the writing about the Gestapo as well. This meant primarily to find out in detail what the Gestapo had meant for its victims. But next to the attempts of giving answers to this question, another feature turned up during the 1970s and 1980s: A new social history of the Gestapo revealed how important input from other institutions and the population had been for the work of the police during the Nazi regime, in particular input in the form of denunciations and input deriving from all sorts of informers. This led some historians to postulate, that Germany under the Nazis had been a self-policing society (Gellately 1990; Paul and Mallmann 1996). More recent research after the year 2000 however has refocused the attention on the terror exercised by the Gestapo and emphasized the initiative and specific political-ideological radicalization of the Gestapo corps.

At the same time earlier, descriptions of the Gestapo focusing mainly on Himmler and Heydrich or paradigmatic figures like Eichmann were replaced by more complex and differentiated analyses of the personnel, emphasizing factors like generational background, collective values, organizational culture, and institutional dynamics (Mallmann and Paul 2004; Paul 2002; Wildt 2009).

If the research about the Gestapo did set in relatively late already, critical studies of the other polices, that is, about the criminal and the uniformed police, did set in even later. The long-lasting silence surrounding the Kripo's crimes during the Nazi period had much to do with the fact that those victims affected by national socialist "crime fighting" were continuously marginalized after 1945, mostly remained objects of repressive policing and were not seen as "legitimate" victims of the "Third Reich." This changed during the 1970s as critical social scientists and "grassroots historians" developed a more comprehensive approach toward the exclusionary politics of the Nazi regime and claimed recognition for the formerly "forgotten victims" (Ayaß 1995; Wagner 2002). The research on the Kripo, which was clearly expanding since the 1990s, has given its attention on the politics of the "fight against criminality" and the Kripo's growing inclusion into the biological and racial strategies of the Nazi regime. While it elaborated the characteristics of crime fighting in the "racial state" of the "Third Reich," this research yet also made clear in what extent police work during the Nazi regime adapted values, images, policies, and techniques from pre-Nazi times – and what continuities of policing can be identified between the German Empire and the Weimar Republic – the Nazi regime, and the postwar years.

The interest, research has taken into the history of the uniformed police, has grown significantly in recent years. In particular its involvement in the mass executions of the Jewish population in Nazi-occupied Eastern Europe has been analyzed in great detail. Due to this focus, research on ordinary policing of the uniformed police in German cities is only gradually emerging. A core issue of the research about the uniformed police is asking for the factors, which turned policemen from members of the police force, loyal to the Nazi regime, into murderous perpetrators. There is currently a consensus among researchers that a whole spectrum of factors has to be taken into account for explaining this shift, ranging from the effects of political indoctrination to the habituation to murderous violence, from psychological factors such as group pressure to situational time- and location-based factors (Browning 1992; Welzer 2005).

In 1945, when Nazi Germany was defeated, the victorious allied forces declared the Gestapo, together with a number of Nazi organizations, a criminal organization, while the Kripo and the uniformed police were left out from being accused of a specific affinity to the Nazi regime. In those parts of Germany, occupied by the Americans, the British, and the French, purges of the police took place, but the overall continuity remained relatively high. Very many of the detectives and the officers of the uniformed police continued to work in the police service (Fürmetz et al. 2001). As the public – at least during the first decades of (Western) postwar Germany – was reluctant to a detailed and critical examination of the Nazi regime, its leading actors, and supporters and as strong political, institutional, and legal barriers were impeding systematic investigations against Nazi perpetrators, only few of the former members of the Gestapo, Kripo, and Orpo were accused and sentenced for their participation in racial politics and genocide (Mallmann and Andrej 2009; Ullrich 2011).

It has taken the German police considerable time to deal with the legacies of its Nazi past. First projects did set in during the early 1980s already. These projects, which quite often had the form of joint teams of police officers and academic historians, were much driven by so-called "critical" members of the police force, who saw police history as being part of a police reform strategy. This strategy aimed at a more civil, service-orientated, and self-reflecting police. But as this reform turned into a contested field, which did not find undivided acceptance among members of the police force, some of these early police history enterprises came under the critic from the police, especially as not everyone among senior ranks of the police was already "open" for critical views on the history of the police during the Nazi period. More or less a decade later, during the 1990s, a second wave of police history activities, involving big city police forces, did set in. The – in many aspects – most prominent among these big city police force histories was carried out under the auspices of the Cologne police (Buhlan and Jung 2000). A major result of this project on the Cologne police had been to show how deeply and intensively everyday policing had been involved in pursuing the exclusionary, racial policies of the Nazi regime. Further projects on the police of major German cities during the Nazi period confirmed these findings. In 2007, the Bundeskriminalamt (BKA) initiated a history project, which was to deal with the history of the BKA during the 1950s and 1960s, thereby focussing on the question to what extent continuities existed within the BKA: "Continuities" either meant asking for the members of the Bundeskriminalamt, who had served already in the security forces of the Nazi regime, or it meant trying to find out what impact (if at all) these characteristics of the BKA-personnel had on the practical work, the practices of the Bundeskriminalamt (Baumann et al. 2011). Since very recently (2011), projects dealing with the Nazi legacies within the Bundesnachrichtendienst (Federal External Secret Service) and the Bundesamt für Verfassungsschutz (Federal Internal Secret Service) are on the way as well. The initiatives for these projects are very much driven by the following logics: Firstly, dealing with the Nazi legacies of an institution such as the police, even if highly problematic aspects might emerge, is increasingly seen as important for a reflective police culture. Secondly, the German public generally recognizes positively efforts to deal with the Nazi past. And thirdly, and last but not least, within many institutions in the Federal Republic, a self-understanding has evolved, which sees the development of the institutional infrastructure of the Federal Republic as part of a democratic success story – which is the more successful, as this success had evolved on the background of a terrible past. That is why many of the research questions put forward in this context are not only directed on scandalizing continuities (personnel, practices) beyond the 1945 line within these institutions but to show how successful the respective institutions had moved toward democracy and "Rechtsstaat," the rule of law, despite the burdens of the Nazi past.

Related Entries

- ▶ Comparing Police Systems Across the World
- ▶ Conceptualizing of Police
- ▶ German Police Until 1918
- ▶ History of Criminal Investigation
- ▶ History of Criminological Theories: Causes of Crime
- ▶ Police in the Police State
- ▶ Political Crime

Recommended Reading and References

Ayaß W (1995) 'Asoziale' im Nationalsozialismus. Cotta, Stuttgart

Banach J (1998) Heydrichs Elite. Das Führerkorps der Sicherheitspolizei und des SD 1936–1945. Schöningh, Paderborn

Baumann I, Reinke H, Stephan A, Wagner P (2011) Schatten der Vergangenheit. Das BKA und seine Gründungsgeneration in der frühen Bundesrepublik. Luchterhand, Köln

Browder GC (1996) Hitler's enforcers. The Gestapo and the SS Security Service in the Nazi revolution. Oxford University Press, New York/Oxford

Browning CR (1993) Ordinary men: reserve police battalion 101 and the final solution in Poland. Harper Collins, New York

Buhlan H, Jung W (eds) (2000) Wessen Freund und wessen Helfer? Die Kölner Polizei im Nationalsozialismus. Emons, Köln

Curilla W (2011) Der Judenmord in Polen und die Ordnungspolizei. Paderborn, Schöningh

Dams C, Stolle M (2012) Die Gestapo. Herrschaft und Terror im Dritten Reich, 3rd edn. Beck, München

Fürmetz G, Reinke H, Weinhauer K (eds) (2001) Nachkriegspolizei. Sicherheit und Ordnung in Ost- und Westdeutschland. Ergebnisse, Hamburg

Gellately R (1990) The Gestapo and German society. Enforcing racial policy 1933–1945. Oxford University Press, New York, Oxford

Gellately R (2001) Backing Hitler. Consent and coercion in Nazi Germany. Oxford University Press, New York/Oxford

Gellately R, Stoltzfus N (eds) (2001) Social outsiders in Nazi Germany. Princeton University Press, Princeton/Oxford

Gutman I (ed) (1990) Encyclopedia of the Holocaust. Macmillan, New York

Herbert U (2011) Best. Biographische Studien über Radikalismus, Weltanschauung und Vernunft. 1903–1989, 5th edn. Dietz Nachfolger, Bonn

Klein P (ed) (1997) Die Einsatzgruppen in der besetzten Sowjetunion, 1941/42. Die Tätigkeits- und Lageberichte des Chefs der Sicherheitspolizei und des SD. Edition Hentrich, Berlin

Mallmann K-M, Andrej A (eds) (2009) Die Gestapo nach 1945. Wissenschaftliche Buchgesellschaft, Darmstadt

Mallmann K-M, Paul G (eds) (2000) Die Gestapo im Zweiten Weltkrieg. 'Heimatfront' und besetztes Europa. Wissenschaftliche Buchgesellschaft, Darmstadt

Mallmann K-M, Paul G (eds) (2004) Karrieren der Gewalt. Nationalsozialistische Täterbiographien. Wissenschaftliche Buchgesellschaft, Darmstadt

Matthäus J (2003) Ausbildungsziel Judenmord? 'Weltanschauliche Erziehung' von SS, Polizei und Waffen-SS im Rahmen der 'Endlösung'. Fischer, Frankfurt/M

Paul G (ed) (2002) Die Täter der Shoah. Fanatische Nationalsozialisten oder ganz normale Deutsche? Wallstein, Göttingen

Paul G, Mallmann KM (eds) (1996) Die Gestapo. Mythos und Realität. Wissenschaftliche Buchgesellschaft, Darmstadt

Roth T (2010) 'Verbrechensbekämpfung' und soziale Ausgrenzung im nationalsozialistischen Köln. Kriminalpolizei, Strafjustiz und abweichendes Verhalten zwischen Machtübernahme und Kriegsende. Emons, Köln

Topographie des Terrors (2010) Gestapo, SS und Reichssicherheitshauptamt in der Wilhelm- und Prinz-Albrecht-Straße. Eine Dokumentation. Stiftung Topographie des Terrors, Berlin

Ullrich C (2011) 'Ich fühl' mich nicht als Mörder'. Die Integration von NS-Tätern in die Nachkriegsgesellschaft. Wissenschaftliche Buchgesellschaft, Darmstadt

Wagner P (1996) Volksgemeinschaft ohne Verbrecher. Konzeptionen und Praxis der Kriminalpolizei in der Zeit der Weimarer Republik und des Nationalsozialismus. Hamburg, Christians

Wagner P (2002) Hitlers Kriminalisten. Die deutsche Kriminalpolizei und der Nationalsozialismus zwischen 1920 und 1960. Beck, München

Welzer H (2005) Täter. Wie aus ganz normalen Menschen Massenmörder werden. Fischer, Frankfurt/M

Westermann EB (2005) Hitler's police battalions enforcing. Racial war in the East. UP of Kansas, Lawrence

Wildt M (2009) An uncompromising generation. The Nazi leadership of the Reich Security main office. University of Wisconsin Press, Madison

Zimmermann M (1996) Rassenutopie und Genozid. Die nationalsozialistische 'Lösung der Zigeunerfrage'. Christians, Hamburg

Police in the Police State

Justice Tankebe
Institute of Criminology, University
of Cambridge, Cambridge, UK

Overview

"Police state" is a translation of the German word *polizeistaat*. The first citation of the term by the *Oxford English Dictionary* comes from the *Times* (London) of 1851: "Austria has become more of a police state than before." The *War Illustrated* followed suit in 1939: "spies are everywhere; indeed, Germany is the modern exemplification of 'the police state' in action." One way to understand the category of the police state is to frame it in terms of the question of the relationship between the "police" and the "state" (Neocleous 2000). Two, albeit, crude distinctions can be drawn; on the one hand, there are regimes in which the relationship is intimate to the point where the police institution is merely a handmaiden of the state and concerned almost exclusively with the task of executing regime orders; the aim is to protect the regime rather than serve citizens. On the other hand are regimes that claim to operate in such a way as to maintain a distance between the state and the police institution. These two types of regimes correspond, respectively, with non-liberal and liberal political cultures. In liberal political cultures, the police institution is seen as identical with civil society. For example, it has been argued that the source of legitimacy for the British police is a strong moral and emotional identification between the police and the British people: Thus, Reith (1956, p. 287) stressed what he considered to be "the historic tradition that the police are the public and that the public are the police." This alleged independence of the police from the state is the converse of what happens in police states or totalitarian states.

This entry summarizes three main issues: First, it examines what is known about the nature of police states. Second, how do police states emerge? Third, it discusses what the police do and how well they do it in the police state. For example, what is the nature of crime in the police state, and what are the technologies of crime control? How reliable is it to assess the effectiveness of the technologies used? Fourth, how do citizens respond to the police state? Were the actions of the police considered "legitimate," or police states lie outside the consideration of legitimacy? Did citizens live in fear of the police or they expressed popular support for the police?

Problems of Definition[1]

The emergence of the modern state in Europe was associated with the birth of what has been termed the "well-ordered police state" in the sixteenth century. It is not that police states were unknown prior to this period. Indeed, some scholars make a distinction between the *traditional police state* and the *modern police state*, the latter referring to the well-ordered police state and the former describing the state prior to the era of the modern state.

Why are certain regimes classified as "police states"? Stated differently, what is the "police state"? Prima facie, the answer to both questions is straightforward, and many will easily cite examples of such regimes. However, a careful consideration of the literature leads to a more cautious appraisal of the concept, that one cannot simply make a binary categorization of regimes into police states and nonpolice states. Rather it is more useful to speak of a *continuum of police statehood*, with different regimes (both democratic and nondemocratic) positioned on different points of the continuum: totalitarian regimes at one extreme end of the continuum and (liberal) democratic regimes at the other end. One of the few attempts to engage conceptually with the idea of the police state is Brain Chapman (1970). He set out a number of criteria for measuring the applicability of the police state model.

The first is *politicization*. This is not to be equated with the idea that the police are political. All police forces, whether in totalitarian or liberal democracies, are political in the sense that they are created by the state and used by the state to maintain law and order. By politicization, Chapman is referring to police involvement in partisan politics and to a situation where police actions are dictated by partisan political considerations rather than the rule of law. Under such circumstances, the vicissitudes of police legitimacy are closely tied up with politics. Second is *militarization* of the police; rather than relying on the army for armed response where necessary, the police organization chooses to augment the capacity of its riot unit by providing it with armored vehicles and other equipments to enable it to operate independently of the army. This is a deeply problematic criterion, especially when Chapman argues that a state is on the road to becoming a police state if its police force weans itself from the control of the army and operates as an independent state institution (p. 119). There is no liberal democracy in which police operations are under the control or supervision of the army.

Chapman's third criterion is the *Centralization* of police services. It is hard to see how can be seen as a distinctive feature of the police state. For example, the structural configuration of a police force does not necessarily correspond to the character of the political culture in that country. Countries such as Belgium and Sweden have centralized police forces, yet they are scarcely the kind of countries we might describe as being totalitarian. A decentralized police institution does not guarantee against the emergence of localized despots, whether such despots are police officers acting on their own or

[1] The essay draws heavily on the work of Arendt (1968) and Evans (2005).

under the dictates of local politicians. The Royal Commission on Police in Britain considered and rejected this approach to defining the police state. The Commission argued that the proper criterion for the police state should be "whether the police are answerable to the law," and that in the police state, "the government acknowledges no accountability to a democratically elected parliament, and the citizen cannot rely on the courts to protect him" (1962, p. 45). But even this test of police statehood is not without difficulty. For example, it fails to differentiate between formal or procedural accountability and substantive accountability. Police states such as Nazi Germany had a parliament (the *Reichstag*) which, it was claimed, represented the will of the German public and passed laws to legitimate Nazi actions often in retrospect (The question of legality and its legitimacy will be considered in final part of this entry).

Finally, Chapman discusses what he labels *penetration*. This, he asserts, entails an encroachment into the judicial domain, with the police obtaining powers of arrest, supervision, and detention, and a right to inflict penal sanctions outside the control of the normal judicial machinery. It is very hard to claim that this characterization is typical of totalitarian regimes and that such encroachment is alien in liberal democracies, even in the twenty-first century; police forces in all societies are vested with powers of arrest and detention, although the duration of detention before trial varies widely. Nor is it unknown in democratic societies for police to have powers to administer penal sanctions without recourse to the courts. Take the example of Britain: Criminal Justice Act 2003 effectively granted police officers the power of punishment. Specifically, the Act allows officers to issue cautions to certain categories of offenders, a legal provision that has been criticized for its procedural and substantive unfairness (Brownlee 2007).

While dictionary definitions should not drive criminological analysis, a look at the dictionary can often illuminate our grasp of certain concepts. According to the *Oxford English Dictionary* the meaning of a police state is "a totalitarian state run by means of a national police force, using repressive methods such as covert surveillance and arbitrary arrest and imprisonment to control the population." Here again, the idea of a national or centralized police force is set out as a feature of totalitarianism, yet it is never entirely clear why centralization should be anathema to democratic governance. Even the notion of "repressive methods" is not without difficulties. Although a liberal democracy, the United Kingdom is considered to be one of the most closely surveilled countries in the world. For example, the City of London has 69 closed circuit television (CCTV) cameras per 1,000 population; Wandsworth, an area of 4.6 miles, has a total of 1,113 CCTV cameras which is far more than the CCTV cameras of police departments of Boston (USA), Johannesburg (South Africa), and Dublin City (Ireland) combined (BBC 2009). This level of surveillance was unparalleled by the experience of Nazi Germany or any other police state.

In the end, all regimes target certain sections of their populations (for example, migrants and suspected terrorists), and embark on elaborate surveillance against them. It may be a bold claim but it may not be delusional to suggest that it is not the *fact* of particular actions or modes of operation of particular regimes that ipso facto make those regimes totalitarian or police states; many liberal democracies share with totalitarian regimes the use of intensive covert surveillance, imprisonment as a means of control of some sections of their population, and encroach upon the powers of the judiciary. Far more important might be *how* these technologies are employed and what mechanisms exist for substantive democratic accountability. In other words, deploying the category of police state to label non-liberal states may not help to illuminate the difference between them and liberal states (Neocleous 2000). It is far more useful to pay attention to how the police in both liberal and non-liberal societies treat citizens during everyday encounters. As a dictum, claims to democratic governance "would hardly be allowed to go unchallenged if the police severely restricted public meetings and political demonstrations or resorted readily to physical force and intimation in order to prevent crime" (Bayley 1969, p. 11). Nor is it unreasonable to observe that even within democratic societies, the experiences and perceptions that some sections of society have of the police is akin to what citizens of police states experience. Put colloquially, one person's democratic state might be another person's police state.

Emergence of Police States

The preceding comments about the conceptual challenges in making sense of the category of police state do not mean that such a category is without merit. It has been argued that concepts are the gateway to the empirical world of study for empirical science, and therefore, the effective functioning of concepts is a matter of decisive importance (Blumer 1954). However, one cannot hope to deploy concepts effectively without-clarity in their meaning. That is not a task necessarily attempted here; the objective so far has been more modest, and it is to draw attention to the need for further work in order to understand more fully its analytical utility. Nonetheless, there are examples of states in recent history widely described as police states: Nazi Germany and Stalin's Russia.

Under what conditions do police states emerge? It is difficult to answer this question without the risk of contextless generalizations. The coming to power of Stalin and Hitler occurred under completely different conditions. Russian was a relatively backward country, with a weak economy and history of despotic rule. Germany, on the other hand, was a well-developed economy; it had a highly educated population with reasonably strong institutions and a burgeoning democracy. Indeed, as discussed below, Hitler came to power through elections. It is important to expatiate on the Germany situation further, because it demonstrates that economic development and education do not necessarily suppress the emergence of totalitarian rule (Fukuyama 1992).

Toward the end of the nineteenth century, Germany had embarked upon its first experiment with democratic governance, and the experiment had gone spectacularly wrong. Parliament was toothless, unable to check executive power. The rights of workers were not fully recognized and industrialists had mounted strong opposition against moves toward unionized labor, arguing that unionized labor was an impediment to economic growth. Nor could it restrain encroachment upon the civil liberties of some sections of the population, in particular Catholics and Jews, both of which were considered "enemies of the Reich" (Evans 2005). At the same time, extremist political parties were increasingly gaining prominence. These parties tapped into the anxieties and resentment of ethnic Germans who felt increasingly disillusioned by the major economic and social transformation in German society at the time. In many Western societies today, migrants and asylum seekers are easy targets for such anxieties (Bauman 2004). In Germany, Jews were the target; they were seen as the source of all societal problems. Extremist politicians argued for their civil liberties and economic activities to be restricted. Mainstream political parties capitulated. Finally, defeat in the First World War brought economic burdens, and there was also increased violence. The judiciary was a tainted institution, widely perceived as partisan in favor of reactionary elements.

It was within this context that Nazis came to power under the leadership of Adolf Hitler "who possessed one great gift: the ability to move crowds with his rhetoric" (Evans 2005, p. 7). The first electoral success for the Nazis was in September 1930 and July 1932, largely with support from the middle classes which had felt threatened by the election promises of the Communists to restrain capitalism. After various political maneuvers in parliament, Hitler was appointed the head of a new government in 1933. In the initial stages of the development of the police state, there is often a struggle for power and the focus of the regime is to liquidate all forms of organized resistance, both open and secret. Nazi Germany was not an exception. It established an extensive surveillance system that tracked down, arrested, and punished people opposed to it. So effective was this pursuit that the regime succeeded in crushing completely all forms of organized opposition. The Nazis banned all political parties, transforming Germany from a fledging multi-democratic state into a one-party state under the leadership of Hitler. In 1934, a nationwide plebiscite approved a law that made Hitler the Leader and Reich Chancellor; his authority was "total and all-embracing. . .and subject to no checks or controls" (Evans 2005, p. 44).

One of the features of the police state is the dominance of the police in domestic politics. The military tends to occupy peripheral positions. There are at least two interrelated reasons for this state of affairs. First, regimes in police states tend not to have confidence in the military to make the necessary cognitive shift from its raison d'être of dealing with external threats to that of considering its own people as if they were foreign combatants. Second, police states aspire eventually to establish a world government (Arendt 1968). Consequently, they tend to approach the victims of their foreign aggressive adventures as though they were rebels, and therefore prefer to govern them with the police rather than the military.

The Police State and Crime Control

To understand the effectiveness of the mechanisms of crime control, it is prudent to first consider what *law* and *crime* mean under the police state. A key feature of law in democratic societies is that it brings a certain degree of certainty and predictability into aspects of relationships among citizens. The law is also applied prospectively, thereby allowing those subject to power to know in advance when they will be subject to coercion and thus avoiding needless interference in their lives. The opposite is true of police states; unpredictability and uncertainty were the main features of the law both in Germany under Hitler and Russia under Stalin. The law and procedures for its application are always in "continuous flux."

In Nazi Germany, two states existed contemporaneously: the *normative state* and the *prerogative state* (Fraenkel 1941). The former was the formal institutions that existed prior to Hitler coming to power, and were bounded by long-standing rules, laws, and procedures. The Prerogative state was an extralegal institutional arrangement that derived its legitimacy from Hitler. Although initially characterized by conflict and sometimes accommodation, the relationship between the two eventually became one of subjugation; the spirit of the prerogative state permeated the normative state as it abandoned its legal procedures and gave approval to hitherto illegal state actions. Hitler's word was law, and the legally correct procedure was as he deemed it. At any rate, the courts of the normative state were dominated by Nazi sympathizers; the Nazis disregarded the law, including laws enacted by themselves if this suited their convenience. Within this context is to be expected shifting definitions of crime and their punishments.

Criminologists have long recognized that crimes are social constructions and that there is a great deal of variation between societies in perceptions of crime seriousness. Thus, the kind of behavior, utterance, or dissent that would be tolerated in liberal democratic societies is viewed as criminal in police states. For example, under the Malicious Gossip Law of 1934 in Nazi Germany, it was an offense to make "spiteful or provocative statements" in public against the governing party, to criticize its policies, or complain about suppression of freedoms and civil liberties. Whether the offense was for malicious gossip or not, the social identity of the offender was always an important consideration in determining the swiftness of police response and the severity of sentences by the courts. Thus, Jews, for instance, were punished more severely for criminal offenses that were otherwise overlooked or attracted very lenient sentences if committed by non-Jews.

Once the police state has been established, there is a shift from dealing with the "suspect" to dealing with the "objective enemy" (Arendt 1968). The objective enemy is unlike the suspect in many ways. Suspects are people whose previous and present deeds and "dangerous thoughts" give reason for the state to be suspicious about what they might do. They are people who oppose and desire to overthrow the government, or are believed to have committed crimes. The objective enemy, on the other hand, is the product of government policy irrespective of the orientations of those so defined. The objective enemy is seen as a "carrier of tendencies" considered a threat to the state. The government never runs out of objective enemies because new ones are easily found depending on the circumstances (Arendt 1968). For example, the Nazi regime had well-developed plans for the extermination of the Polish people when the extermination of Jews was near completion (Arendt 1968). This was to involve, among other things, attempts to impose regulations strikingly similar to what had been done before implementing "the final solution": for example, change of names and the death penalty for racially undesirable marriages, such as those between Germans and Poles (Evans 2005).

The secret police constitute the main channel through which orders of the government are transmitted. In addition to the change from dealing with suspects to attending to objective enemies is a corresponding change from *suspected offense* to *possible crime*:

> . . .every thought that deviates from the officially prescribed and permanently changing line is already suspect, no matter in which field of human activity it occurs. Simply because of their capacity to think, human beings are suspects by definition, and this suspicion cannot be diverted by exemplary behavior, for the human capacity to think is also a capacity to change one's mind. Since, moreover, it is impossible ever to know beyond doubt another man's heart . . .suspicion can no longer be allayed if neither a community of values nor the predictabilities of self-interest exist as social realities. (Arendt 1968, p. 430)

The result of this is to create *a system of ubiquitous spying*, infusing hitherto healthy social relationships with mutual suspicion. Each person feels under constant surveillance; what was once methods employed exclusively for dealing with the population is now the methods that ordinary citizens employ to deal with their neighbors in everyday encounters. As discussed later, however, scholars differ on whether fear, terror, and intimidation were as pervasive as it is often portrayed in popular accounts, or it was the case that citizens freely consented to the decisions and directives of the police.

In the initial years of the establishment of the police state, the focus is primarily upon eliminating all sources of organized opposition. Concern with suspects of possible crimes assumes prominence when the regime is fully established. As the police state enters its last and fully totalitarian stage, it abandons the category of objective enemy and possible crime. The new category of "undesirables" replaces them. Here victims are chosen at random and officially declared unfit to live. In the well-documented case of the Nazis, undesirables included the mentally ill and persons with certain kinds of diseases. This introduction of arbitrariness means that the police state becomes far more efficient in suppressing freedoms; both the innocent and the guilty share the fate of being undesirable. But more importantly, this change in the concept of crime and criminals comes with new methods for law enforcement. Consistent with their task of executing executive orders, it is the responsibility of the police to ensure criminals are punished and undesirables disappear, often without a trace. The only trace victims leave behind are the memories of those who knew them; a challenging task for the secret police is to ensure that even this trace will disappear with the victims. The Russian secret police, for instance, employed elaborate network analysis of families, friends, and acquaintances of suspects with the aim to eliminating those believed to possess "dangerous memories" of the disappeared. Hannah Arendt has captured this well:

> In totalitarian countries all places of detention ruled by the police are made to be veritable holes of oblivion into which people stumble by accident and without leaving behind them such ordinary traces of former existence as a body and a grave. Compared with this newest invention for doing away with people, the old-fashioned method of murder, political or criminal, is inefficient indeed. The murderer leaves behind him a corpse, and although he tries to efface the traces of his own

> identity, he has no power to erase the identity of his victim from the memory of the surviving world. The operation of the secret police, on the contrary, miraculously sees to it that the victim never existed at all. (Arendt 1968, pp. 434–435)

In common with their counterparts in democratic societies, the lifeblood of police work in the police state is information from the public. Two decades of empirical analyses of police legitimacy have shown that cooperation with police forces rests mainly on the perceived legitimacy of the police (Tyler 1990). Police states and their secret police forces were fully aware of this role of legitimacy in securing the flow of intelligence, and they devoted much energy and resources cultivating legitimacy among their publics. In Nazi Germany, the Gestapo relied upon the assistance of the German public, and there is strong evidence to show that many Germans responded positively, denouncing their neighbors, colleagues, and relatives to the secret police. The police in the police state also rely heavily on extensive networks of informers, both *impressed* and *voluntary* informers (Chapman 1970). The use of informers is not exclusive to police states; it is a technique that is also widely used by law enforcement agencies in democratic societies (Natapoff 2009). But informers are not always reliable sources of intelligence. The police therefore seek to operate their own sources of intelligence gathering by infiltrating areas of social and political life. Again, the strategy of infiltration is as much a practice of police states as it is of democratic states.

The police in Nazi Germany were effective in destroying initial opposition to the regime and subsequently perceived enemies and undesirables such as Jews and social outsiders. Levels of petty criminality were however comparable to those that pertain in democratic societies. Johnson (2011) has shown that a large proportion of Germans committed crimes on a frequent basis, without much fear of being detected and punished. The reason for that situation was not ineffectiveness by the police; on the contrary, it was because the police prioritize other crimes over petty street-level offenses.

Citizens and the Police State

How do citizens react to police states? This question divides historians of Germany under Hitler and Russia under Stalin, the two quintessential cases of police state in the twentieth century. There are two broad schools of thought.

One school of thought holds that citizens consented voluntarily to the police state. Some studies have produced evidence to show that the Gestapo, for example, was a much smaller organization than it was often thought. Hitler was "so immensely popular among most Germans that intimidation and terror were rarely needed to enforce loyalty" (Johnson and Reuband 2005, p. 329). Gellately (1990) cites election results in support of a view that the Nazi regime was popular among the German people. The coercion the regime applied was aimed at only a small proportion of minorities but that was even done with the approval of the majority of the population. Apart from election results, the other reason for the claim that citizens did not experience the police state as an "all knowing, all powerful, and omniscient" is the size of the secret police. The ratio of secret police to citizens is estimated at 1 per 10,000 residents in the cities of the Third Reich, and usually with extremely limited presence in rural areas (Johnson 2011). The secret police, it is said, were after all not always well resourced, and therefore could not be said to have relied systematically on terror and coercion to induce citizen compliance. As further evidence, scholars argue that the reaction of the ordinary population was that of voluntary supply of information through denunciation of neighbors (Johnson 2011).

A second school of thought contends that police organizations of police states were omnipresent and omnipotent, and that they employed arbitrary powers, repressive tactics, and unimaginable brutality against citizens. Consequently, citizens lived in fear and terror, mindful of the potentially ruthless repercussions from the police. Elections, it is argued, lacked integrity and therefore a poor measure of popular attitudes toward the police state. Far from being a society of "self-policing," terror was widespread.

The Nazis regularly publicized executions, court proceedings, and sentences. The aim of the publicity was to deter potential offenders from underestimating the risk of being caught and punished (Evans 2005).

It is fair to say that police states thrived upon a mixture of coercion and terror, and popular support both from within the security forces and the general population. There is no evidence throughout history to show that any regime relied exclusively on its ability to physically intimidate its subjects or citizens into obedience, not even the most "unjust and blood-minded dictatorship" (Fukuyama 1992, p. 16). The Nazis, for instance, "did not just seek to batter the population into passive, sullen acquiescence. They also wanted to rouse it into positive, enthusiastic endorsement of their ideals and their policies, to change people's minds and spirits and to create a new German culture that would reflect their values alone" (Evans 2005, p. 118). The question that rises is whether one can speak of *legitimacy* under totalitarian regimes. It is a question that recalls a long-standing debate among scholars. Social scientists approach legitimacy (that is, recognition of the moral rightness of power) within the particular historical societies rather than universally. They are fully aware that what makes power legitimate in one society may differ from others, and that the conditions for legitimacy in one may be repudiated by another. This approach does not require the social scientist to make any judgment about the appropriateness or otherwise of the social order she investigates. A clear implication of this line of thought is that it is possible for a researcher to conclude that a police organization is legitimate in the empirical sense (that is, it finds wide moral acceptance among citizens) and yet for that researcher to believe that that organization is deeply unjust or even "evil."

It is on this conclusion that moral philosophers and some political scientists depart from Max Weber. For them, legitimacy relates to whether by some objective standards of ethical evaluation, a claim to legitimacy can be recognized as valid. The contention is that one cannot simply reduce legitimacy to a matter of fact, the fact that citizens hold a certain belief about a regime. On the contrary, the concept should signify a normative evaluation of the correctness of the procedures, the justification for decisions, and the fairness with which regimes treat their subjects (Grafstein 1981). The full implication of this approach is that there is a need to connect analysis of legitimacy to theories about justice (Bottoms and Tankebe 2012). That is a discussion beyond the scope this entry. The point to note here, however, is that a focus on the public actions or behavior of citizens as the measure of their reactions to a regime, especially one that exercises unrestrained power, is unlikely to tell the whole story about power relations; indeed, it is likely to lead to the erroneous conclusion that "subordinate groups endorse the terms of their subordination and are willing, even enthusiastic, partners in that subordination" (Scott 1990, p. 4). What is known so far about the reactions of citizens to police states offers no assurance against this error.

Conclusion

It is commonplace to read from newspapers, and even from scholarly work, that some democratic societies are on a pathway to becoming police states. Such claims often arise from evidence that those states have undertaken certain measures to facilitate a greater surveillance of the population or a section of that population (e.g., the US Patriot Act and the UK CONTEST strategy for counterterrorism) and other forms of intrusion into private liberties of citizens. By any account, these are poor indicators of the police state; no state, democratic and totalitarian, can hope to respond effectively to threats in the modern world without some level of surveillance of (some) its population.

The absence of conceptual clarity regarding the category of police state suggests that the greatest immediate need in assessing the utility of that category has to start with attempts to clarify its meaning. It appears that criminology can contribute toward that endeavor from two interrelated standpoints. First is to return to

Edwin Sutherland's well-known definition of criminology as the study of "the processes of law-making, of law-breaking, and of reacting to the breaking of laws" (Sutherland 1939, p. 1). At the heart of this triumvirate is the question of power, and how it is exercised. A second and related issue is for criminology to engage with the political science literature with the aim of procuring a proper understanding of democracy. That literature suggests that democratic societies are characterized by a universalistic ideological claim about human equality, that all human beings are of equal worth and therefore entitled to equal respect and treatment. Consequently, any exercise of power must do so with respect for the principles of individual liberty and equality. Democratic societies attempt to achieve this through mechanisms of democratic accountability of the police, including external oversight of police institutions that is independent of government. How effectively such oversight works is open to debate. What is true is that democracies differ from police states on this check on police power, and it is here that the search for conceptual clarity should begin.

Related Entries

- ▶ Conceptualizing of Police
- ▶ Counterterrorism
- ▶ Democratic Policing
- ▶ Electronic Monitoring
- ▶ False Confessions and Police Interrogation
- ▶ Policing of Peacekeeping
- ▶ Surveillance Technology and Policing

Recommended Reading and References

Arendt H (1968) The origins of totalitarianism. Harcourt, London

Bauman Z (2004) Wasted lives: modernity and its outcasts. Polity Press, Cambridge

Bayley DH (1969) The police and political development in India. Princeton University Press, Princeton

Blumer H (1954) What is wrong with social theory? Am Sociol Rev 19:3–10

Bottoms AE, Tankebe J (2012) Beyond procedural justice: a dialogic approach to legitimacy in criminal justice. J Crim Law Criminol 102:119–170

Browder GC (1990) Foundations of the Nazi police state: the formation of Sipo and SD. University Press of Kentucky, Kentucky

Brownlee ID (2007) Conditional cautions and fair trial rights in England and Wales: form versus substance in the diversionary agenda? Crim Law Rev 129–140

Chapman B (1970) Police state. Pall Mall Press, London

Coicaud J-M (2002) Legitimacy and politics: a contribution to the study of political right and political responsibility. Cambridge University Press, Cambridge, UK

Evans RJ (2005) The third Reich in power: how the Nazis won over the hearts and minds of a nation. Penguin, London

Fraenkel E (1941) The dual state: a contribution to the theory of dictatorship. Oxford University Press, New York

Fukuyama F (1992) The end of history and the last man. Penguin Books, London

Gellately R (1990) The Gestapo and German society: enforcing racial policy, 1933–1945. Oxford University Press, Oxford

Giddens A (1985) The nation-state and violence. Polity Press, Cambridge

Grafstein R (1981) The failure of Weber's conception of legitimacy: its causes and implications. J Polit 43:456–472

Johnson EA (2011) Criminal justice, coercion and consent in 'totalitarian' society: the case of national socialist Germany. Brit J Criminol 51:599–614

Johnson EW, Reuband K-H (2005) What we knew: terror, mass murder, and everyday life in Nazi Germany: an oral history. Basic Books, Cambridge, MA

Lukes I (1996) The birth of a police state: the Czechoslovak ministry of the interior, 1945–48. Intell Natl Sec 11:78–88

MacCormick N (2007) Institutions of law: an essay in legal theory. Oxford University Press, Oxford

Natapoff A (2009) Snitching: criminal informants and the erosion of American justice. NYU Press, New York

Neocleous M (2000) The fabrication of social order: a critical theory of police power. Pluto Press, London

Raeff M (1983) The well-ordered police state: social and institutional change through law in the Germanies and Russia, 1600–1800. Yale University Press, New Haven/London

Reith C (1956) A new study of police history. Oliver & Boyd, London

Scott JC (1990) Domination and the arts of resistance: hidden transcripts. Yale University Press, New Haven/London

Sutherland EH (1939) Principles of criminology. Lippincott, Philadelphia

Tamanaha BZ (2004) On the rule of law. Cambridge University Press, Cambridge

Tyler T (1990) Why people obey the law. Yale University Press, Yale

Weber M (1978) Economy and society: an outline of interpretive sociology (Roth G, Wittich C, eds), 2 vols. University of California Press, Berkeley

Zureik E, Salter MB (2007) Global surveillance and policing: borders, security, and identity. Willan Publishing, Cullompton

Police Integrity

Sanja Kutnjak Ivković
School of Criminal Justice, Michigan State
University, East Lansing, MI, USA

Overview

Police integrity and police misconduct are closely related; traditionally, the discussion of police integrity typically would focus on the events that signal the *lack* of integrity (e.g., forms of police misconduct and the ways of controlling them). Since the mid-1990s, the focus has shifted on police integrity, which, in turn, allows for the exploration of a broader concept – the development of a high-integrity police agency. Police integrity is defined as "the normative inclination among police to resist temptations to abuse the rights and privileges of their occupation" (Klockars et al. 2006).

Although police integrity could be associated with the moral virtues of individual police officers, starting with the writings by Herman Goldstein in the 1970s, the idea has started to develop that this "bad apple" theory of police corruption is inadequate and that a more organizational approach should be used instead. Klockars and colleagues (Klockars and Kutnjak Ivković 2004; Klockars et al. 1997) built upon this view and proposed their theory of police integrity. The theory features four dimensions: organizational rulemaking, detection, investigation, and discipline of rule violations, curtailing the code of silence, and the influence of public expectations and agency history on police integrity. The accompanying methodological approach has been utilized to survey police officers and to detect the contours of police integrity in nearly 20 countries across the world. Other research on police integrity explores the mechanisms used to enhance integrity and accountability, such as the early warning systems and citizen reviews.

Police Integrity

Introduction

In the 1990s, a wave of police misconduct scandals shook the country, from the police corruption scandal in New York, which prompted the establishment of the Mollen Commission, the Rodney King beating, which resulted in the establishment of the Christopher Commission, the Abner Luima sexual violence case in New York, which resulted in $8.75 million dollar settlement, and the Rampart Division scandal in Los Angeles. In 1997, the Office of Community Oriented Policing Services (COPS) and the National Institute of Justice (NIJ) held a national symposium bringing together policy makers and social scientists to address this emerging problem. A natural tendency would be to anchor the conference on the events that signal the *lack* of integrity. That is, the discussion would center on various forms of misconduct and the ways of controlling them. As Hickman and colleagues (2004, p. 1.1) put it, "[a]ll too often we infer integrity from its demonstrated absence. That is to say, research that focuses on corruption infers ideas about integrity absent its measurement." Instead, the COPS/NIJ team decided to use the term *police integrity* as the guiding theme for the conference (Greenberg 1997, p. v):

> The focus on "police integrity" opened a whole new domain. Although previous research, study, and experimentation had focused on critical issues such as corruption and excessive use of force, these approaches, had, in fact, revolved around single dimension. As a result, the solutions were constricted in that they were derived out of a need to control unwanted behaviors of individuals. In comparison, police integrity guided the focus on

> the broader domain of developing a healthy organization that would serve to reinforce and maintain the good character and constructive motivations of many of the individuals joining the ranks of law enforcement.

In July of 1996, about 200 police leaders, politicians, lawyers, and researchers attended a three-day National Symposium on Police Integrity in Washington, D.C. (Gaffigan and McDonald 1997). Among the participants who tried to define police integrity, the preferred approach was to envision a police officer with high integrity and list the characteristics (e.g., virtues, values, and character traits) that separate this police officer apart from other police officers of lower integrity. The characteristics explicitly listed at the Symposium were allegiance, courage, honor, honesty, prudence, trust, effacement of self-interest, intellectual honesty, justice, morality, principled behavior, responsibility, and dedication to mission (Vicchio 1997). Mark Moore was in charge of summarizing the findings of the symposium; he wrote that "[w]hat we mean by integrity and professionalism is law-abiding character, technical confidence, neutrality, distance – in Steve Vicchio's wonderful phrase, 'the effacement of personal interest' – and probably some notion of courtesy and client responsiveness" (Moore 1997, p. 63). Similarly, Hickman and colleagues wrote in their introductory remarks to a co-edited book on police integrity that "police integrity refers to the underlying values and ethical attachment of the police and how those values and ethics affect police behavior" (Hickman et al. 2004, p. 1.1).

These lists of characteristics contain expected individual police officers' character traits. As Klockars and colleagues argue (2006, p. xiv), "[t]he fact is that *any* virtue – charity, compassion, decency, faith, loyalty, passion, patience, and perspective, to add but a few to the list mentioned at the Symposium – is at least arguably as crucial to the integrity of a professional police officer as any other and this approach gives us no guidance whatsoever in prioritizing this litany." Yet, some of the officers who exhibit these characteristics may decide not to act in accordance with high integrity, just like the police officers who do not possess many of these characteristics may decide to act in accordance with high integrity. Finally, these character traits are features of individuals and they close the possibility that police integrity could be an organizational feature as well.

Since 1996, the issue of police integrity has been discussed at several conferences ("the 1999 NIJ/NYU Seminar on Police Integrity and Democracies; the 1999 *Strengthening Police-Community Relationships* conference; the 2000 biannual conference *Policing in Central and Eastern Europe: Ethics, Integrity, and Human Right* (Pagon 2000)*;* the 1998 *Sixth International Conference on Ethics: Integrity at the Public-Private Interface* (Huberts and van den Heuvel 1999")), and continues to be explored bi-annually at corruption conferences such as the *Global Forum against Corruption and Safeguarding Integrity* and the Transparency International's *International Anti-Corruption Conference*. The US Department of Justice provided a publication addressing the *Principles for Promoting Police Integrity* (2001). The co-edited book *Police Integrity and Ethics* (Hickman et al. 2004) contains chapters addressing various aspects of police integrity. However, the understanding of what police integrity is, and what should be covered under this topic, varies substantially across conference organizers, presenters, and authors. A substantial portion of the literature seeks to develop integrity-enhancing mechanisms or mechanisms that would control police misconduct (Pope 2000; Transparency International 2001; U.S. Department of Justice 2001; See ▶ Control of Police Misconduct).

Yet, it seems that even the basic debate regarding what constitutes police integrity has not been resolved. Sam Walker, a leading expert on police accountability, points out that "[t]he issue of police integrity is extremely important and has received an increasing amount of public attention among policy makers and the general public. Unfortunately, the academic literature has not adequately addressed this very important subject" (2004, p. vii).

Definition

Klockars and colleagues (Klockars et al. 2006) have provided a definition of police integrity that goes beyond the traditional focus on individual police officers' character traits and allows for the organizational approach toward it. They have also accompanied the definition with a theory of police integrity and the methodological approach toward its measurement. Klockars and colleagues (2006) define police integrity as "the normative inclination among police to resist temptations to abuse the rights and privileges of their occupation." According to the authors (Klockars et al. 2006), this definition has six critical components.

Normative – The normative aspect of the definition emphasizes that integrity is based on the moral norms of right and wrong (Klockars et al. 2006, p. 1) that tell people what they should or should not do. Thus, integrity combines a belief in moral values with an inclination to behave in accordance with that belief; "[j]ust as a belief in honesty inclines one to avoid lying and a belief in fidelity obliges one to be faithful, integrity requires not only a belief that certain behaviors are right or wrong, but also actions that are in accord with those beliefs" (Klockars et al. 2006, p. 2).

Inclination to Resist – This part of the definition emphasizes the difference between attitudes and behavior. The authors do not expect that the person of high integrity will always behave in accordance with these values. Rather, they discuss a general problem that people who believe in honesty sometimes lie; people who believe in fidelity sometimes are unfaithful; and people of integrity sometimes do things they know are wrong. Three dimensions of the relation between police attitudes of integrity and police misconduct are addressed.

The first point the authors emphasize is that, while integrity describes the normative inclination to resist temptations, integrity is not the only source (Klockars et al. 2006, p. 3). Other possible candidates include the lack of imagination, lack of opportunity, fear of discovery and public humiliation, shame, punishment, or a simple unfavorable risk/reward calculus as potentially critical in ensuring that police officers behave honestly. The second point the authors emphasize is that that attitudes of integrity put at least some pressure on police officers who share them to actually avoid wrongful behavior. Because these attitudes of integrity may lead police officers to comply with the rules and, at the same time, may not necessarily be the driving force behind compliant behavior, the authors have decided to refer to integrity as the *inclination* to resist rather than the actual resistance to temptations. The third point the authors emphasize is that the direction of the causal relation between attitudes of integrity and behavior of integrity is not always absolutely clear. In particular, while attitudes may cause behavior, behavior may also cause attitudes. Requiring of police officers to behave in accordance with integrity may lead them to adopt the attitudes of integrity as well.

Police – Klockars et al. (2006) emphasize that integrity is a characteristic of the "police," without specifying that this is a characteristic of individual police officers or police agencies. This reflects the view that integrity may be a feature of an individual police officer, a group of police officers, a subunit of a police agency, a whole police agency, or police agencies. The authors caution that the discussion about police integrity easily lapses into the discussion of integrity at the police officers' individual level. Yet, police integrity is also an organizational feature – a feature of police agencies, organizations, and institutions – and we can discuss organizational cultures of police integrity. However, the discussion of the dynamics and the relevant correlates of police integrity will differ depending on the level of integrity under discussion (e.g., individual, organizational); "[h]ow one understands and explains the psychology of integrity of an individual police officer will most certainly differ from the understanding and explanation of the evolution of a culture of integrity in a police agency" (Klockars et al. 2006, p. 3).

Temptations – The motivation or the reasons why police officers engage in misconduct could vary dramatically across forms of police misconduct, individual police officers, and police agencies. The most obvious of these temptations is *gain*, the monetary or non-monetary reward for type of police misbehavior, a typical feature of traditional police corruption. However, other forms of police misconduct are not necessarily motivated by gain; different forms of police misconduct are by no means the product of a similar, singular, or even ignoble temptation. The example of use of excessive force is particularly illustrative (Klockars 1995, p. 17):

> . . .[the excessive force] need not (and usually will not) be the product of malicious or sadistic behavior. It can spring from good intentions as well as bad, mistakes and misreading, lack of experience, overconfidence, momentary inattention, physical or mental fatigue, experimentation, inadequate or improper training, prejudice, passion, an urge to do justice or demonstrate bravery, misplaced trust, boredom, illness, a specific incompetence, or a hundred other factors that might influence an officer to behave in a particular situation in a less than expert way.

Klockars et al. (2006) conclude that the methods used to control misconduct should differ depending on the type of temptation and argue that the methods used to control corruption could be quite different for the methods used to control the use of excessive force. They also infer not only that the contours of integrity may be very different in different police agencies, but also that the contours of police integrity within the same police agency could be different depending on the form of police misconduct. They argue that police integrity need not be a uniform phenomenon. Consequently, "it is not difficult to imagine a police organization or subculture that was highly intolerant of officer theft, soliciting bribes, taking kickbacks and other acts of corruption and at the same time was much more accepting of discourtesy, excessive force, perjury, forging records, fabricating evidence, or unwarranted or illegal searches" (Klockars et al. 2006, p. 4). The authors proceed to note that one could also expect that police officers who succumbed to the temptations of one variety (e.g., gain) may also be more likely to succumb to the temptations of another variety (e.g., prejudice).

Abuse – A crucial element of the definition is the concept of abuse. In many situations, the abuse of office is obvious (e.g., a police officer accepts a bribe in exchange for letting the citizen caught violating the speeding limit). However, the discussion of whether particular behavior is abusive is confounded by two possible arguments. The first argument flatly denies that the behavior in question is abusive. For example, the police officers' acceptance of discounts, holiday gifts, free food, and liquor could be viewed as gestures of goodwill, hospitality, and/or gratitude. The second argument recognizes that the behavior could be abusive, but justifies and excuses it as an expression of "street justice." As Klockars and colleagues (Klockars et al. 2006, p. 5) elaborate, "[a] common theme is that police are 'human' and cannot be expected to behave without normal human emotions in situations in which they are insulted, defied, assaulted, deceived, shocked, repulsed, disgusted, or horrified by the conduct of those they police."

The Rights and Privileges of Their Occupation – Policing is a highly discretionary, coercive activity that routinely takes place in private settings, out of the sight of supervisors, and before witnesses who are often regarded as unreliable. Policing as an occupation is rife with opportunities for misconduct, as independent commission reports and scholarly studies clearly demonstrate (see, e.g., Christopher Commission 1991; Knapp Commission 1972; Manning and Redlinger 1993; Mollen Commission 1994; Pennsylvania Crime Commission 1974).

Causes of Police Integrity

Until relatively recently, the prevailing administrative view of integrity (at least in the United States) was to associate integrity with the moral virtues of individual police officers; consequently, the tendency was to fight corruption, for example, by carefully screening applicants for police positions, pursuing defective officers aggressively, and removing them from their police positions before their behavior spreads throughout the agency (see, e.g., Knapp Commission 1972). In the 1970s, Herman Goldstein (1975) embarked on the pioneering work and argued that this “bad apple” theory of police corruption is inadequate and that a more organizational approach should be used instead.

Klockars and Kutnjak Ivković and their colleagues (Klockars and Kutnjak Ivković 2004; Klockars et al. 1997) built upon this view and proposed their theory of police integrity. This approach stresses the importance of four distinct dimensions, each of which is profoundly organizational in nature.

Organizational Rulemaking – The first dimension of the organizational theory of police integrity addresses organizational rule making (see, e.g., Klockars and Kutnjak Ivković 2003; Klockars et al. 1997) and relates police misconduct to the ways police agencies create, teach, and enforce rules explicitly prohibiting misbehavior (Klockars and Kutnjak Ivković 2003, p. 1.4). In countries with highly decentralized police organization, as is the case with the United States, police agencies could differ dramatically in the nature of the behavior they prohibit. First, agencies could differ in terms of whether they have official rules at all. Some agencies may have no written rules, particularly if they are very small (fewer than 10 police officers), while others, particularly if they are very large (with 500 sworn officers or more), may have extremely lengthy official rules (e.g., standard operating procedures, official rules). Second, when they do have the official rules established, the behaviors those rules allow and prohibit may differ substantially from agency to agency. This is particularly the case for less serious forms of misconduct such as *mala prohibita* corrupt behavior (e.g., receipt of favors, gratuities, small gifts, free meals, and discounts, off-duty employment).

The problem of organizational rulemaking is further complicated by the fact that in many agencies, while an agency’s official policy formally prohibits certain activities, the agency’s unofficial policy tolerates them (see, e.g., Knapp Commission 1972; Mollen Commission 1994). The obligations of rulemaking require police agencies not only to develop both formal and informal rules that specify agency expectations of integrity, but also to teach these rules and enforce them. As Klockars and colleagues conclude (2006, p. 9), “[i]n a police agency of integrity police officers ought to know the agency’s integrity relevant rules, understand the agency’s rationale for them, and believe in the rightness of both.”

Detecting, Investigating, and Disciplining Rule Violations – The second dimension of the organizational theory of police integrity focuses on both the creation and maintenance of activities that permit detection, investigation, and discipline of misconduct (see, e.g., Klockars and Kutnjak Ivković 2003; Klockars et al. 1997). The activities are heterogeneous by nature. They include not only reactive and proactive investigations by police agencies, but also inspections, audits, early-warning systems, external reviews, reception of citizen complaints, and integrity testing. The extent to which these activities have been implemented and actually used by a police agency varies substantially across the agencies. According to Klockars and colleagues (2006, p. 9), in a police agency of integrity the occupational culture of the agency will support the introduction and management of the activities employed to detect, investigate, and discipline rule-violating behavior.

Circumscribing “The Code” – The third dimension of the organizational theory of police integrity emphasizes the police agency’s obligation to circumscribe the code of silence (i.e., “the blue curtain,” “the code”). The code of silence refers to the norm of the police culture which prohibits reporting of misconduct by fellow police officers (see, e.g., Klockars and Kutnjak Ivković 2003; Klockars et al. 1997). However, although potentially viewed as a single norm that could apply to all police officers, the reality is that the code of silence could vary substantially. First, exactly what behavior – in terms of its nature and seriousness – is covered by the code varies enormously across police agencies. The code of silence in some agencies may cover only the least serious forms of corruption (e.g., the acceptance of gratuities and small gifts), while in others it can cover even the most serious forms of corruption (e.g., thefts from crime scenes, the acceptance of bribes from motorists caught speeding). Second, to whom the benefit of the code's protection is extended could also vary extensively. In some agencies, the code could be limited to protect only misconduct by partners, viewed a testimonial immunity that would be a mirror image of traditionally privileged relationships (husband and wife, physician and patient, lawyer and client). In other agencies, the code could protect misconduct not only by police partners, but also by all police officers working on the same shift, section, or unit. Klockars and colleagues (2006, p. 9) write about the close relation between the code of silence and integrity:

> Many police administrators probably understand that circumscribing both whom and what The Code covers should be an administrative priority . . . However virtually all police administrators were line officers at some point in their careers, and thus they have at least an appreciation, if not an affection, for the bonds of collegial loyalty and fraternal support that are part of the subculture of policing. To the extent that circumscribing The Code requires the weakening of those bonds of loyalty and support, it is a task that not a few police administrators approach with ambivalence. A police agency of integrity is one in which the occupational culture is intolerant of those who abuse the rights and privileges of their office.

The Influence of Public Expectations and Agency History on Police Integrity – The fourth dimension of the organizational theory of police integrity relates to the influence of the social and political environment in which police institutions, systems, and agencies operate (see, e.g., Klockars and Kutnjak Ivković 2003; Klockars et al. 1997). This theory argues that the integrity of a police agency is affected by the larger social and political environment in which it operates (Klockars and Kutnjak Ivković 2003). Although this understanding – the idea that the police agency is affected by its environment – is the underlying assumption of virtually all historical studies of police, Reiss and Bordua (1967; Reiss 1971) were the pioneers who provided the first systematic exploration of the topic.

When the society at large expects ethical behavior of its officials, police agencies are also more likely to set high expectations and expect ethical behavior from its employees. Not only do the expectations of integrity differ enormously across the world (see, e.g., Transparency International Corruption Perceptions Index), but they could also differ within the same country. Using the United States as an example, Klockars and colleagues (2006, p. 10) argue that some parts of the country have long and virtually uninterrupted traditions of persistent police corruption (e.g., Chicago, New Orleans, Key West), some areas have equally long traditions of integrity (e.g., Milwaukee, Kansas City), and other areas have undergone repeated cycles of scandal and reform (e.g., New York, Philadelphia, Oakland). In the end, Klockars and colleagues conclude that "not only public expectations about police integrity exert vastly different pressures on police agencies in different areas, but also police agencies of integrity may effectively resist such pressures" (2006, p. 10).

Measuring the Extent of Police Integrity

Klockars and Kutnjak Ivković (2003) designed a questionnaire that seeks to measure the extent of police integrity in systematic, standardized, and quantitative manner. The respondents were provided with a letter asking them to assume that the officer described in the scenarios had been a police officer for 5 years, had a satisfactory working record, and had not been disciplined in the past. The first version of the questionnaire contains 11 hypothetical scenarios focusing mostly on corruption and ranging from those merely giving an appearance of a conflict of interest to those describing incidents of bribery and theft (Klockars and Kutnjak Ivković 2004; Klockars et al. 1997). The second version of the questionnaire went beyond seeing police integrity as the opposite of police corruption and includes a range of police misconduct such as the use of excessive force, failure to execute an arrest warrant, and falsification of official record, as well as police corruption. Each scenario is followed by the same set of seven questions that ask about police officers' knowledge of official rules, their opinion about the seriousness of particular rule-violating behaviors, the discipline these behaviors should receive and would actually receive, and their estimates of how willing they would be to report such behavior.

The first version of the questionnaire has been distributed across 30 police agencies in the United States (Klockars et al. 1997, 2000), characterized with a decentralized police. The authors reported substantial differences across the agencies. Although the sample of police agencies is not nationally representative, but a convenience sample, it nevertheless illustrates the point that the level of police integrity could vary substantially across the police agencies. Klockars and colleagues (2000) created a ranking of police agencies based on their levels of integrity. The authors (Klockars et al. 2006, p. 37) note:

> In some police agencies in our sample police officers found nearly half of the behaviors described in our sample to be of sufficient seriousness to merit dismissal. In those agencies officers also claimed that they and their police colleagues were highly likely to report all but the least serious forms of misconduct. In other agencies only theft from a crime scene was, in the opinion of a plurality of respondents, sufficient grounds to fire a police officer and not even that offence would motivate the majority of officers in such agencies to break the Code of Silence and report the misconduct of a colleague.

A comparison of the results from the two large municipal police agencies, one ranked at the top and one at the bottom of the police integrity scale, shows large differences in how serious police officers perceive misconduct, what discipline they expect and approve, and how wiling they seem to be to report the fellow officers' misconduct. With one exception, police officers from the agency at the bottom of the police integrity scale would not expect the police officer to be dismissed for any of the behaviors described in the questionnaire, while the police officers from the agency at the top of the police integrity scale expected dismissal in four most serious cases. In addition, Klockars and colleagues (Klockars et al. 2006) found stark contrasts in the extent of the code of silence between the two agencies: in the agency ranked at the top of the integrity scale, the majority of police officers reported that their colleagues would report their fellow officer who engaged in the behavior described in all cases. On the other hand, the majority of police officers from the agency ranked at the bottom of police integrity scale did not expect their fellow officers to report misconduct in any of the described cases.

However, as the subsequent study of three large municipal police agencies (identified to be in the top third of the police integrity scale) demonstrates, police agencies can be quite diverse in the ways they create and maintain their integrity (Klockars et al. 2006). The authors studied each of these three agencies – Charleston, South Carolina; Charlotte-Mecklenburg, North Carolina; and St. Petersburg, Florida – for two years and have developed their integrity profiles.

Since 1996, nearly two dozen studies have used this first police integrity survey. Three studies (Micucci and Gomme 2009; Burbach Raines 2009; Marche 2009) analyzed the 30-agency data collected by Klockars and colleagues, and another four studies (Chappell and Piquero 2004; Hickman, et al. 2004; Schafer and Martinelli 2008, Gottschalk 2009) used the police integrity questionnaire as the basis for their surveys of other US police agencies. In addition to the US agencies, the same questionnaire has been used in nearly 20 countries across the world, including Austria, Bosnia and Herzegovina, Canada, Croatia, the Czech Republic, Finland, Hungary, Japan, the Netherlands, Pakistan, Poland, Slovenia, South Africa, Sweden, and the UK (see Klockars et al. 2004).

Although the list of the countries participating in the survey is far from exhaustive, the findings clearly show that the levels of police integrity differ substantially and significantly, analogously to the tremendous variation in the levels of police integrity across the 30 US police agencies. The nature and extent of police integrity measured in the countries such as Finland and Sweden is quite different from the nature and extent of the police integrity measured in countries such as South Africa and Pakistan (see Klockars et al. 2004). The differences are visible not only in the police officers' perceptions of seriousness and willingness to report, but also in their views of appropriate and expected discipline (Klockars et al. 2004, p. 13):

> It appears that in each country the seriousness of officers' misconduct is, in large part, determined by the absolute level of discipline the organization is expected to visit on the offending officer. In almost every case, when the police organization is expected to punish a offense very severely, officers regard that offense as serious. Conversely, when organizations do not punish misbehavior severely, as is the case in Hungary, Pakistan, and South Africa, officers seem to have little ability to distinguish among the levels of seriousness with regard to misconduct. Perhaps the most dramatic finding that emerges from examining the contours of integrity concerns the worldwide prevalence of the code of silence.

The authors continue to note that, in five out of 14 countries included in the book "Contours of Police Integrity," the code of silence would cover misconduct described in *every* scenario (Klockars et al. 2004, p. 17). In addition, the code of silence would protect behaviors such as the acceptance of a bribe from a person caught speeding – a violation of the penal code – in nine out of 14 countries.

The heterogeneity of police integrity is influenced by a host of reasons; a police agency's local social, political, economic, and legal environments influence its level of police integrity. Research indicates that countries that belong to the same category along one dimension (e.g., economic development, geographic location) still have different levels of police integrity. This is the case for both developed democracies (e.g., the USA and the UK) and countries in transition (e.g., Croatia and Hungary), as is the case for countries from Northern America (e.g., the USA and Canada) and Europe (e.g., Croatia and the Netherlands).

The second version of the questionnaire, exploring the resistance to various sources of temptations (not only police corruption), has been developed as well. The initial work by Klockars and colleagues explored the extent of police integrity in three US police agencies using both versions of the questionnaire (see Klockars et al. 2006). The comparative efforts are underway, with the first results coming from Croatia (Kutnjak Ivković 2009), South Korea (Kutnjak Ivković and Kang 2011), and South Africa (Kutnjak Ivković and Sauerman 2012).

Other Research on Police Integrity

The issues of police integrity and police misconduct are closely interconnected. Please see separate entries addressing different forms of police misconduct (e.g., "police corruption," "use of excessive force," "police lying"). A few writings *explicitly* connect police misconduct (or any of its specific forms) with police integrity. For example, Garner and colleagues (2004) studied the patterns in the police use of force as a way of measuring police integrity. The authors (Garner et al. 2004, p. 6.119) conclude that, although the average extent of force used by the police during a typical arrest is about equal across racial categories (and thus should be taken as an indication of police integrity), the extent of force used by the police during a typical arrest of male suspects is larger than the quantity of force used during a typical arrest of female suspects (and thus should be taken as an indication of the lack of police integrity).

In 2001, the US Department of Justice published *Principles of Promoting Police Integrity*. The publication, available in print and on the Internet, lists the "best practices" for promoting integrity. The first part of the publication explores the use of force, complaint and misconduct investigations, training, recruitment, hiring, and retention, as well as general principles of promoting accountability and effective police management. The second part of the publication contains examples of promising police practices and policies, as well as the research projects on police integrity funded by the US Department of Justice. Many of the practices that should be used to enhance integrity (e.g., complaint misconduct investigations, the early warning systems, citizen reviews), at the same time, are mechanisms of accountability. Publications like that, focusing on the specific mechanisms or practices that should be used to enhance integrity, have significantly grown in numbers since the late 1990s. A separate body of research focuses on these mechanisms. Please see the entry "Control of Police Misconduct."

Related Entries

- ▶ Control of Police Misconduct
- ▶ History of Police Unions
- ▶ Police and the Excessive Use of Force
- ▶ Police Corruption
- ▶ Police Culture
- ▶ Police Discretion and Its Control
- ▶ Police Lying and Deception

Recommended Reading and References

Burbach Raines J (2009) Ethics in policing: misconduct and integrity. Jones and Barlett, Sudbury

Chappell AT, Piquero AR (2004) Applying social learning theory to police misconduct. Deviant Behav 25:89–108

Gaffigan SJ, McDonald PP (eds) (1997) Police integrity: public service with honor. U.S. Department of Justice, Washington, DC. http://www.ncjrs.org/pdffiles/163811.pdf. Accessed 8 Apr 2006

Garner JH, Maxwell CD, Heraux C (2004) Patterns of police use of force as a measure of police integrity. In: Hickman M, Piquero AR, Greene JR (eds) Police integrity and ethics. Wadsworth/Thomson Learning, Belmont, p vii

Goldstein H (1975) Police corruption: a perspective on its nature and control. The Police Foundation, Washington, DC

Gottschalk P (2009) Crime-based survey instrument for police integrity measurement. Policing Int J Police Strateg Manag 33(1):52–68

Greenberg SF (1997) Executive summary. In: Gaffigan SJ, McDonald PP (eds) Police integrity: public service with honor. U.S. Department of Justice, Washington, DC. http://www.ncjrs.org/pdffiles/163811.pdf, p. v. Accessed 8 Apr 2006

Hickman M, Piquero AR, Greene JR (eds) (2004) Police integrity and ethics. Wadsworth/Thomson Learning, Belmont

Huberts LWJC, van den Heuvel JHJ (eds) (1999) Integrity at the public-private interface. Shaker Publishing B.V, Maastricht

Klockars, C. (1995). A Theory of Excessive Force and Its Control. In William A. Geller & Hans Toch (Eds.), And Justice for All. Washington, D.C.: Police Executive Research Forum, pp. 11-29.

Klockars CB, Kutnjak Ivković S (2003) Measuring police integrity. In: Hickman MJ, Piquero AR, Greene JR (eds) Police integrity and ethics. Wadsworth, Belmont

Klockars CB, Kutnjak Ivković S, Harver WE, Haberfeld MR (1997) The measurement of police integrity. Final Report Submitted to the U.S. Department of Justice, Office of Justice Programs, National Institute of Justice

Klockars CB, Kutnjak Ivković S, Harver WE, Haberfeld MR (2000) The Measurement of Police Integrity. Research in Brief. U.S. Department of Justice, Office of Justice Programs, National Institute of Justice. Government Printing Office, Washington, DC

Klockars CB, Kutnjak Ivković S, Haberfeld MR (eds) (2004) The contours of police integrity. Sage, Newbury Park

Klockars CB, Kutnjak Ivković S, Haberfeld MR (2006) Enhancing police integrity. Springer, New York

Kutnjak Ivković S (2009) The Croatian police, police integrity, and transition toward Democratic policing. Policing Int J Police Strateg Manag 32(3):459–488

Kutnjak Ivković S, Kang W (2011) Police integrity in South Korea. Policing Int J Police Strateg Manag 35(1):76–103

Kutnjak Ivković S, Sauerman A (2012) Curtailing the code of silence among the South African Police. Policing: Int J Police Strateg Manag (forthcoming)

Manning PK, Redlinger L (1993) The invitational edges of police construction. In: Klockars C, Mastrofski S (eds) Thinking about police. McGraw Hill, New York, pp 398–412

Marche GE (2009) Integrity, culture, and scale: an empirical test of the big bad police agency. Crime Law Soc Change 51:463–486

Micucci AJ, Gomme IM (2009) American police and subcultural support for the use of excessive force. J Crim Justice 33:487–500

Moore M (1997) Epilogue, police integrity. In: Public service with honor. U.S. Department of Justice. https://www.ncjrs.gov/pdffiles/163811.pdf. Retrieved on Oct 10 2012

Pagon M (ed) (2000) Policing in central and Eastern Europe: ethics, integrity, and human rights. College of Police and Security Studies, Ljubljana

Pope J (2000) Confronting corruption. The elements of a national integrity system (TI Source book 2000). Transparency International, Berlin. http://www.transparency.org. Accessed 8 Apr 2006

Reiss AJ Jr (1971) The police and the public. Yale University Press, New Haven

Reiss AJ Jr, Bordua DJ (1967) Environment and organization: a perspective on the police. In: Bordua D (ed) The police: six sociological essays. Wiley, New York

Schafer JA, Martinelli TJ (2008) First-line supervisor's perceptions of police integrity. Policing Int J Police Strateg Manag 31(2):306–323

Transparency International (2001) The national integrity system. Concept and practice. A report by Transparency International (TI) for the global forum II on fighting corruption and safeguarding integrity. Transparency International, Berlin

U.S. Department of Justice (2001) Principles for promoting police integrity; examples of promising police practices and policies. U.S. Department of Justice, Washington, DC. http://www.ncjrs.gov/pdffiles1/ojp/186189.pdf. Retrieved on 8 Apr 2006

Vicchio SJ (1997) Ethics and police integrity: some definitions and questions for study. In: Gaffigan SJ, McDonald PP (eds) Police integrity: public service with honor. U.S. Department of Justice, Washington, DC

Walker S (2004) Forward. In: Hickman M, Piquero AR, Greene JR (eds) Police integrity and ethics. Wadsworth/Thomson Learning, Belmont, p vii

Police Interrogation

► Law of Police Interrogation

Police Leadership Styles

Joseph A. Schafer
Department of Criminology and Criminal Justice, Southern Illinois University, Carbondale, IL, USA

Overview

Leadership is of central importance to the operation of effective, efficient, and equitable police organizations, yet it remains an elusive concept. It is clear that good leadership is vital to ensure that an agency operates in a manner that preserves public legitimacy and trust. Among the compelling evidence in support of this assertion is the fact that a crisis of leadership is often associated with major scandals in police agencies. Despite recognizing that strong leadership is of vital importance, scholars and police organizations have struggled to measure and understand a number of key questions, including what leadership styles are commonly used by police supervisors, when and how supervisors can influence subordinate performance and output, and what distinguishes between effective and less effective leaders.

This entry considers the topic of police leadership styles. Literature considering leadership in police organizations can be separated into two major approaches. First, some write about how general theories of leadership and organizational behavior might be used to explain aspects of police organizations (Adlam and Villiers 2003). This is often the approach taken in textbooks and professional literature. The second approach examines the styles and methods supervisors use in seeking to influence subordinate personnel (Engel 2001). The entry begins by briefly reviewing the dominant implications arising from these two traditions. It continues with an examination of what research evidence suggests about the ability of police leaders to influence unit productivity and output. The entry concludes by considering how different generations of police personnel and leaders might respond to different styles of leadership.

Introduction

At the onset it is important to distinguish between police management and police leadership. Police supervisors are often interchangeably referred to as managers and leaders. Management is the control of routine processes and the maintenance of the status quo. Leadership emphasizes changes that will improve processes, personnel, and organizations. Effective supervisors tend to excel at both management and leadership, because both skill sets are important. However, they are not of the same quality, and policing has tended to emphasize management far more than leadership to its own detriment. It should also be noted that leadership is a behavior that can be independent of formal rank or authority. Some very prominent and influential leaders in police organizations are front-line personnel. Much of what has been written about police "leadership" is actually describing aspects of managing, administering, and supervising police organizations and personnel.

The emphasis of this entry is on true leadership, not what is often described in policing literature (including academic scholarship, which has tended to confuse management and leadership). Leaders innovate, take calculated (and often necessarily) risks, encourage and develop personnel, and are willing to challenge tradition and status quo. This does not imply that leaders blindly challenge anything, everything, and everyone in the organization. Rather, they have the courage to confront the pointless, mindless, counterproductive aspects of their organizations and its practices. This tendency is sorely lacking in many aspects of the modern public sector, particularly policing.

Though much has been written about leadership in broader settings, especially in corporate, military, political, and athletic contexts, policing scholars have given limited empirical consideration to matters of leaders and leadership. The traditional perspective found in police literature tends to cast leaders as using a narrow range of directive and controlling styles; as a consequence, supervisor-subordinate relationships are characterized by tension and animosity (Rowe 2006; Van Maanen 1984). More recently, research evidence has suggested an alternative perspective. Leaders have increasingly been framed as using a range of styles and approaches when interacting with those they seek to influence. These clusters of leadership approaches go beyond the traditional, authoritarian, and autocratic supervisory approaches that once dominated policing (Kuykendall and Unsinger 1982). It is quite likely this change reflects a very real transformation in leadership styles observed in policing.

Major Theories Relevant to Police Leadership

A wide array of theories has been offered to describe and explain the behavior of individuals and organizations. Some of the earliest studies of leaders and leadership tended to focus on the "great man/great woman" tradition. Well-regarded leaders were studied in a biographical format in the hope of deriving an understanding of the behaviors and actions that contributed to their perceived success. Over time, studies expanded this tradition by examining samples of recognized leaders. Modern research continues to seek an understanding of the casual links between leader traits and subsequent efficacy; what characteristics, habits, and behaviors differentiate leaders who are most effective from their peers?

Because early studies were concerned with the traits and habits of leaders, they tended to pay less attention to followers. The latter were often seen as liabilities and risks that organizations and supervisors needed to manage through control, policy, procedure, reporting requirements, and similar of mechanisms. This was based on what Douglas McGregor (1960) referred to as Theory X assumptions, and it led to a very distinctive, controlling, directive form of management. In contrast, McGregor described Theory Y approaches that were predicated on the belief that people can achieve good work outcomes when they are given the proper motivation and environment. As a result, Theory Y management approaches emphasize creating the optimal environment in which employees can be engaged and empowered, rather than seeking to regulate and constrain their choices and behavior.

James Burns observed a similar duality in the nature of leader-follower interactions, writing about "transactional" and "transformational" approaches to leadership (1978). Transactional leaders were seen as emphasizing exchanges and quid pro quo arrangements with the personnel they supervised. Followers were provided with an understanding of what they were to accomplish and how it was to be achieved; they also understood what they would receive for adhering to those expectations and the consequences of violating accepted practices. The role of leaders in such arrangements was not to provide leadership; in actuality, they were there to monitor and manage followers, grant rewards when earned, and issue discipline when needed. Followers were given "carrots" (rewards) for good behavior and faced the risks of "sticks" (punishments) if they did not produce the expected outcomes in the expected way.

In contrast, transformational leadership was characterized as a process of seeking to improve organizations, operations, and personnel. Leaders truly lead rather than just managing and maintaining the status quo. To accomplish this outcome, transformational leaders sought to inspire employees to embrace a shared vision and empowered them to use discretion and creativity in the pursuit of that outcome. Followers were given the freedom to make choices, the authority to develop new ideas, and the independence to identify and implement ways to pursue the leader's vision. In other words, followers were viewed as being smart enough and creative enough to be entrusted to solve problems, pursue new ideas, and act in a way that was consistent with the vision of the leader. This was done without requiring followers to engage in burdensome reporting requirements, bureaucratic permission seeking, and mindless compliance checks.

Though not universally embraced, there is also a sense in more contemporary perspectives that the models, styles, and traits that generate leadership efficacy are situational. This idea was first expressed by Tannenbaum and Schmidt (1958) in writing about management in the corporate world. They arrayed leadership approaches on a continuum from "boss centered" to "subordinate centered," noting that a major challenge for (then) modern managers was deciding where to fall on the continuum in a given situation. In policing contexts, this suggests that "what works" for a leader is rarely universal, instead varying based on context, culture, position, objectives, the environment, the strengths and predispositions of a leader, and the preferences and characteristics of those the leader seeks to influence. An officer or leader who is ineffective in one agency or assignment might perform quite well in a different set of circumstances or vice versa. It has also been argued that leadership is not the same at all levels of an organization. The behaviors that make a leader an effective patrol sergeant might not yield the same outcomes for the leader when promoted to serve as a lieutenant overseeing the records division. Leaders need to be conscious of the fact that the approaches used in a given situation and/or to achieve a given end might not yield universal success.

Leadership is a challenging process, but it can be the element that distinguishes organizations that are "great" from those that are merely "good." In addition, leadership is a behavior, not a position within an organization. Leaders are people who demonstrate certain skills and habits, regardless of their formal rank or office. Too often it is presumed that all supervisors are leaders and that only supervisors can lead. Increasingly there is awareness and celebration of the idea that every officer must be a leader, at least in some contexts. Informal leadership is very difficult to measure in any type of organization, but it can have a profound influence on personnel and operations (Schafer 2001). Though the influence of informal police leaders is generally evident in policing scholarship, research has not been able to empirically define when and how it matters.

Theoretical perspectives and explanations regarding leadership extend far beyond the brief ideas reviewed in this passage. What is of relevance for this discussion are the ideas that leadership is situation but that modern perspectives tend to emphasize involving employees in the decision-making process. It is sometimes observed that leadership is something done *with*, not *to*, others. Further, it should be recognized that good leadership is actually quite difficult to achieve. Effective leaders continually seek ways to improve their performance as they confront new and different challenges (Schafer 2012).

Leadership Styles Displayed in Policing

Despite the presumed importance and influence of informal leaders in policing context, studies of police leadership have focused on studying formal supervisors. As a consequence, what is known about leadership styles and their influence (or lack thereof) is based on studies of those who have been awarded a rank. Such supervisors may or may not demonstrate actual leadership behaviors as they carry out their duties. Far less is known about leaders in police organizations, who may or may not hold an official rank in their agency (Van Maanen 1983). This section reviews literature seeking to describe the management and leaders styles police supervisors have displayed in prior studies. There is a distinction between what scholars found when they first began to study police supervisors and what is more often seen in contemporary agencies.

Early efforts to study the management styles of police supervisors did not paint a favorable picture of the level of actual leadership taking place in organizations. Kuykendall and Unsinger (1982) administered a leadership inventory to police managers. Their results suggested that most respondents used a variety of styles in the course of performing their supervisory duties. The most commonly reported management styles were selling, telling, and participation. There was little evidence that supervisors used actual leadership or other forms delegation as a component of their managerial style, though their data was collected in the late 1970s before these types of approaches had entered the discourse of police management. In other words, managers reported a preference to use styles that were oriented more toward "safe" methods emphasizing control and direction of personnel, rather than responses that might be considered more "risky" styles, such as delegating important tasks to subordinate personnel.

Mayo (1985) contended supervisors tended not to trust the skills, loyalty, and judgment of personnel in their agencies. This resulted in the observed practice of police supervisors engaging in micromanagement. Girodo (1998) surveyed police executives from around the world and found most reported leaning toward what might be considered "Machiavellian" approaches. End objectives were achieved not through democratic and participative approaches but through the manipulation of subordinate personnel. This is reflective of the tendency for police agencies to seek to operate as paramilitary organizations that exert top-down control over personnel, communication, and decision making. These traditional styles of police supervision have been subject to extensive criticism because they do not encourage employees to be committed, productive, and creative members of the workforce.

Allen (1982) found the type of supervision (i.e., the use of incentives) was likely more influential than the volume of supervision. Quality was more important than quantity (more contact), though his measures were quite limited. It should also be noted his research was conducted on a 1970s group of front-line police personnel. The education, traits, experiences, and expectations of modern patrol officers are likely quite different than what was seen in earlier eras.

Johnson (2006) studied supervisor influence on traffic and drunk driving enforcement in a collection of agencies in the Cincinnati metropolitan region. His results support that leadership matters, both in terms of the priorities of agency executives and also the priorities of patrol supervisors. Though leadership was deemed important, it was not the only factor influencing enforcement productivity. Organizational factors such as policies, supervision, and bureaucratic requirements do influence when and how officers enforce the law but so do situational factors, officer preferences, and environmental considerations. In other words, even when leadership matters, it is not the only force or factor influencing officers and the decisions they make.

Engel (2001) developed four broad categorizations of supervisory styles that help frame an understanding of different leadership approaches in policing. "Traditional" leaders were highly task oriented and expected aggressive enforcement of the law from subordinates; they were focused on traditional policing outcomes and presumably would have leaned toward transactional relationships with followers (e.g., officers making a lot of arrests and issue a lot of citations would presumably be ranked higher by traditional leaders). "Innovative" leaders were situated further onto the transactional end of the leadership continuum; they shared power with subordinates, sought to have friendly and productive relationships with subordinates, and emphasized community relations. "Supportive" supervisors sought to be a buffer between front-line personnel and top agency leaders; they place less emphasis on accomplishing traditional policing tasks. Finally, "active" supervisors were themselves highly active in front-line policing matters; they had positive views of their subordinates, tended to exemplify the idea of "leading by example," and did not see themselves as radically different than those they supervised. Active supervisors demonstrated the most influenced in the two agencies included in Engel's study.

Beginning in the 1980s, there has been a greater emphasis on democratic and participative approaches to police supervision and leadership (Wycoff and Skogan 1994), though limited advancements have been realized. The promise of such approaches is multifaceted. A leader who engages subordinates in making decisions can tap into the creativity, experience, and intelligence of employees. It is expected that participative approaches can make organizations more adaptable and responsive in the face of shifting and dynamic social environments. Leaders who use these types of approaches do not make all choices unilaterally. They engage subordinates to identify problems, determine priorities, and derive solutions. This does not suggest that all choices and actions are subject to full participation by all employees. Certainly there are aspects of police operations that require command-and-control authority and decision making, but such circumstances are the exception, not the rule. Though there is limited evidence assessing these types of practices, experiences to date suggest promising implications (Steinheider and Wuestewald 2008).

Survey data from police officers finds personnel often express a preference for supportive and participatory leadership styles (Witte et al. 1990). Officers do not like directive and controlling management styles. Instead, they prefer to be led in a process that grants them a voice and input into organizational decisions. This does not suggest that officers do not understand and support the needs of leaders to have ultimate authority. Rather, it implies that, when possible, officers prefer to be able to participate in organizational decision making in an open and democratic manner. Tentative evidence suggests that many police executives and leaders are similarly open-minded in employing nontraditional systems (Steinheider and Wuestewald 2008; Tannenbaum and Schmidt 1958). It is less clear, however, whether those expressions of support translate into actual changes in supervisory style as perceived by employees.

Consideration of supervisory styles in policing should not overlook the role of poor or ineffective leadership. Barbara Kellerman (2004) offered one of the best considerations of this topic in leadership literature. She argued that too often, professional and scholarly discussions of leadership frame leader behavior as benevolent and positive behavior. She and others have argued consideration of leadership needs to acknowledge leaders sometimes use less-than-ideal methods. Some supervisors achieve their objectives while using poor leadership techniques. Some leaders are quite successful but pursue ignoble or evil objectives (Adolf Hitler is the common example of this outcome). People who are generally very good and effective leaders occasionally make mistakes. Leaders who have been successful sometimes “derail” (McCall and Lombardo 1983). Consideration of police leadership styles should not overlook how poor and ineffective approaches can hamper the ability of leaders to achieve their objectives.

The Influence of Leadership

It is generally accepted that police supervisors (who are formally in positions of leadership) can shape various organizational and individual outcomes produced by employees (Engel 2001). Supervisors are thought to have the ability to influence the tone of police operations, helping shape the style, tactics, and outcomes. In reality, it is far easier to believe that supervisors have such a strong and direct influence on officers and the organizations than it is to prove that belief is reality. This is partially a function of the limited amount of quality research studying supervisory and leadership influences in policing. It is also a reflection of the nature of the police organization itself.

In many organizational contexts, supervisors have routine and direct contact with their subordinates. The task environment is also such that employees often have limited discretion, established accountability mechanisms, and routine reporting demands that serve to structure, regulate, and formalize the decisions they make. Police departments are often characterized as a quasi-military bureaucracies characterized by command-and-control relationships between supervisors and subordinates. This tends to support the use of directive supervisory styles, which are predicated on the ability of a manager to closely monitor employee performance.

In reality, officers are frequently separated from their supervisors by both time and space. The nature of police work also means that officers enjoy a wide degree of discretion in performing their duties, particularly the low seriousness events that typify much of police work (Brown 1988; Van Maanen 1983). Additionally, police operations and personnel are influenced by cultural and external forces that serve to condition where, when, how, and to what end police decisions are made. The result is that it has been suggested police agencies are "loosely coupled organizations" (Maguire and Katz 2002). Police managers do not have the time or capacity to effectively monitor the actions of officers on the street. The unpredictable, dynamic, and complex nature of police work also makes it quite difficult for agencies to craft effective policies to govern and guide how officers make decisions in the field. As a result, police personnel are not nearly as restricted by administrative impositions.

On paper and in textbooks, agencies are orderly and control is clear; reality is quite different. This makes leadership (particularly transformational and democratic approaches) all the more important. Directive, controlling, manipulative, and micromanaging approaches actually function to push officers away from organizational objectives and discourage them from supporting the mission and vision of executives. Real leadership (as opposed to management) can help engage, empower, and motivate personnel to support and pursue organizational objectives. Because officers are often performing their duties with no direct supervision and little reliable evidence verifying when, where, and how they perform their duties, having engaged personnel is all the more important. Creating work environments that treat officers in accordance with McGregor's (1960) Theory Y assumptions would, in the aggregate, be expected to produce better outcomes and results for the organization.

As a result of these factors, it is reasonable to expect limitations on the ability of police leadership to influence police personnel. Many of these challenges are far more acute in policing than in other occupational and organizational settings. Police personnel operate in environments with limited direct supervision, handle a diverse range of tasks, and make myriad high-discretion decisions (Allen 1982; Engel 2001; Van Maanen 1983, 1984). Though supervisors and organizations seek to impose policies, protocols, structures, and mandates to condition and direct discretionary behavior, officers enjoy a considerable degree of latitude in performing their duties. This totality of circumstances suggests that leaders will have a limited, but not nonexistent, ability to influence officers. Police supervisors and leaders are able to exert at least a limited degree of influence on personnel.

Much of the research studying supervisory influence has focused on examining traditional policing outcomes, such as enforcement behavior, use of force, and officer misbehavior (Brown 1988; Huberts et al. 2007). Though such considerations are important, they reflect but a narrow range of the tasks to which officers attend and the expectations the public has of their police. Equally important, they tend to focus on the ends, rather than the means, of policing. A small number of studies have attempted to examine issues such as the nature of police encounters with the public and self-initiated efforts officer undertakes to address specific problems. Taken as a whole, researchers have achieved mixed assessments of whether supervisors can influence the behavior of subordinate personnel (cf, Allen 1982; Brown 1988; National Research Council 2004). It should be noted that many of these studies were limited in the types of influence they were studying and the methods that were used to measure specified outcomes.

It has conventionally been assumed that bringing about change in police organizations is akin to "bending granite" (Rowe 2006). In recent years, increasing experience and empirical evidence have begun to suggest this may not always be the case. Jermier and Berkes (1979) studied officers in a Midwestern agency, finding that officers were supportive of leadership approaches that emphasized participation. Officers reported they were less inclined to respond to directive and controlling management approaches. Case studies of community policing have highlighted the role of leadership as a mechanism to facilitate rapid (and at times, radical) organizational transformation (Wycoff and Skogan 1994). Though critics have dismissed the roles of William Bratton and COMPSTAT as causal forces influencing New York City's crime decline, there can be little doubt that his leadership style in NYPD (and elsewhere) has brought about rapid change in climate and culture. Thus, leadership can change police culture, though it is still difficult to determine how and why that outcome is achieved in some instances and not in others.

Several studies based on data collected for the Project on Policing Neighborhoods (POPN) tend to offer some provisional evidence that supervisors can influence front-line personnel (Engel 2001). Importantly, these studies attempted to assess not just the general idea of supervisory influence. Instead, supervisors were categorized based on their general management style. Findings based on the POPN data offer a possible explanation for the mixed results achieved in prior studies by suggesting that variation in supervisory influence is partially a product of dissimilarity in supervisory styles. Not all supervisors use the same approaches in seeking to control, influence, and lead. The methods and styles used by supervisors vary based on the circumstances, timing, and audience, as well as the supervisor's perceptions of various situational exigencies.

Far less is known about the measurement of leadership efficacy beyond considerations of officer performance, output, and conduct. What other metrics might be used and who should provide input on the performance of a given leader? These efforts can be challenging, as they require the development of broader definitions and measures of leader efficacy, which tends to invoke a number of methodological difficulties. Huberts et al. (2007) studied Dutch police officers. They found leadership approaches did influence integrity violations by officers. In particular their results suggested that officers were influenced by strong role modeling by their leaders. When leaders took a strong stance against integrity violations and were perceived to perform their duties with a high moral standard, officers were more inclined to emulate that approach. This is certainly a favorable representation of what might be achieved through leadership, at least based on some important outcomes of relevance to policing.

As a result of this body of literature, several important observations and conclusions can be made about leadership and influence in police organizations. Supervisors can influence front-line personnel, at least in the production of some outcomes using some management styles. Poor leadership practices can also generate real and negative consequences in the workplace, including poor productivity, dissatisfaction, stress, attrition, and absenteeism, among other concerns (Kelloway et al. 2005). Though traditional police leadership approaches were primarily transactional (stick and carrot) approaches, transformational strategies seem more promising in their ability to favorably influence personnel and behavior. There is evidence that followers might prefer such participatory approaches and that leaders are open to their use. To date, however, research has not been able to conclusively contrast the outcomes of transactional and transformational leadership approaches in policing.

Generational Effects

There may be generational and experiential influences that shape the approaches used by supervisors. In the 1960s and 1970s, there was a renewed emphasis on college education for police personnel. Pursley (1974) compared police chiefs based on education, experience, and involvement in professional activities (belonging to national professional associations, publishing in professional periodicals); he analyzed their responses to a variety of leadership inventories. He found that educated chiefs had less experience (both as officers and chiefs, suggesting they were also likely to be younger), were more connected with professional groups and activities, and were more receptive to more transformational leadership approaches (though he did not use this term).

Younger leaders were more willing to support delegating authority within the workforce, allowing subordinates to participate in decision making, embracing more democratic approaches, and supporting less authoritarian and controlling leadership approaches. Thus, when considering how generations influence policing and police leadership, it is not just a matter of different mentalities across groups of different officers. Supervisors from different generations might also have different ideas about how to lead personnel.

Cohen (1980) studied command personnel (captains and above) in the New York City Police Department based on data collected in 1972. He found two primary management styles. Tradition-oriented leaders generally had no education beyond high school. Reform-oriented leaders generally had attended or were attending college. Tradition-oriented leaders were focused on power and authority, while reform-oriented leaders were more participative and emphasized community relations and service. Cohen found that reform-oriented leaders had received lower ratings on internal performance assessments, but he attributed this to the probability that tradition-oriented leaders were evaluating most reform-oriented leaders. Thus, it is would not be surprising to see that the disjuncture in their orientations toward leadership and the broader role of the police might result in less positive performance evaluations.

This holds important implications for understanding leadership. First, it reinforces that leadership is situational; at the very least "good" leadership is a time-bound concept. Cohen's work was being done during a time of transition and reform in NYPD. Second, leadership practices and police personnel might both experience different generations. Several factors contribute to this situation. In general, different generations are distinct from one another based on variation in life experiences, culture, and broader society during their formative years. During eras of transition in policing (the movement out of the traditional model in the 1970s, the proliferation of community policing in the 1980s and 1900s, the emergence of various data-driven and evidence-based policing practices in the 2000s), there is likely to be tension throughout all ranks of police agencies. Personnel who started their career one (or two) ideological generations earlier might be less inclined to adapt their beliefs, values, and practices. Making matters more complicated, the role of leadership in periods of reform and transition might be a function both of a leader's generation and position within the organization. Top executives likely are driving forces behind implementing these reforms and innovations, so their support might be expected. In contrast, the drag or tension may be among mid-level supervisors. Sergeants, lieutenants, captains, and others who feel no ownership over the new idea have little reason to embrace new perspectives. They may resist the change or evenly actively obstruct its emergence because they view the change as a threat to both their perspective on leadership and their continued advancement in the organization.

Andreescu and Vito (2010) found contemporary managers expressed strong support for ideals consistent with transformational leadership approaches. These included inclusive and human-oriented styles of leadership, the articulation of an organizational vision, caring for the well-being of employees, and setting an example for employees. Transformational approaches remain a challenge for police organizations. Though support for this style of leadership is apparent among police personnel and leaders (Witte et al. 1990), there is also a belief that such circumstances as participatory management are not actually found in most organizations. Stamper (1992) found a disjuncture between how executives perceived themselves and how members of their executive staff perceived them. Executives believed they lead their organization in one fashion. This self-perception tended to differ from how those around the executives perceived their leadership style and influence.

Intuitively it makes sense that younger officers might be more likely to support transformational, participatory, and democratic approaches to leadership. These youth are more oriented toward social interaction, social involvement, and being allowed a voice in aspects of organizational operations. Scholars and professionals continue to speculate that there might be a generational preference for transformational leadership approaches, but there is still an absence of research evidence assessing this matter in policing contexts. As the labor force becomes increasingly filled with those from generations desiring more transformational leadership, it will be increasingly important that police leadership adapt to reflect that situation. If organizations are to maintain long-term viability, it will be essential to create work environments that attract and retain newer generations of officers.

Conclusions

Leadership remains a vital issue in modern police organizations. Its importance cannot be refuted, and the need for leadership (not simply strong management) is quite apparent within the profession. What remains more elusive is a clear understanding of when and how leadership matters as a force influencing the values, beliefs, attitudes, and behaviors of police personnel. Though there is a natural appeal toward transformational leadership styles, the potential benefits of these approaches remain presumed rather than proven in police organizations. Issues of leadership will most likely continue to become increasingly important in the future as police organizations need to be more adaptive in responding to ever-changing criminal threats and community expectations. This will be compounded by work forces increasingly composed of officers from more recent generations, who may well require a very different style of leadership and a very different organizational environment. If police organizations are going to remain viable in confronting crime and disorder, real leadership will be of increasing importance. Both the research community and the profession itself must support the development of greater knowledge of the role leadership will play in these dynamic future environments.

Related Entries

- ▶ Managing Innovation of Policing
- ▶ Measuring Police Performance
- ▶ Methodological Issues in Evaluating Police Performance

Recommended Reading and References

Adlam R, Villiers P (eds) (2003) Police leadership in the twenty-first century: philosophy, doctrine and developments. Waterside Press, Winchester

Allen DN (1982) Police supervision on the street: an analysis of supervisor/officer interaction during the shift. J Crim Justice 10:91–109

Andreescu V, Vito GF (2010) An exploratory study of ideal leadership behaviour: the opinions of American police managers. Int J Police Sci Manag 12:567–583

Brown MK (1988) Working the street: police discretion and the dilemmas of reform. Russell Sage, New York

Burns JM (1978) Leadership. Harper & Row, New York

Cohen B (1980) Leadership styles of commanders in the New York City police department. J Police Sci Adm 8:125–138

Engel RS (2001) Supervisory styles of patrol sergeants and lieutenants. J Crim Justice 29:341–355

Girodo M (1998) Machiavellian, bureaucratic, and transformational leadership styles in police managers: preliminary findings of interpersonal ethics. Percept Mot Skills 86:419–427

Huberts LWJC, Kaptein M, Lasthuizen K (2007) A study of the impact of three leadership styles on integrity violations committed by police officers. Policing Int J Police Strateg Manag 30:587–607

Jermier JM, Berkes LJ (1979) Leader behavior in a police command bureaucracy: a closer look at the quasi-military model. Adm Sci Quart 24:1–23

Johnson RR (2006) Management influences on officer traffic enforcement productivity. Intl J Police Sci Mgt 8:205–217

Kellerman B (2004) Bad leadership: what it is, how it happens, why it matters. Harvard Business School Press, Boston

Kelloway EK, Sivanathan N, Francis L, Barling J (2005) Poor leadership. In: Barling J, Kelloway EK, Frone MR (eds) Handbook of work stress. Sage, Thousand Oaks, pp 89–112

Kuykendall JL, Unsinger PC (1982) The leadership styles of police managers. J Crim Justice 10:311–321

Maguire ER, Katz DM (2002) Community policing, loose coupling, and sensemaking in American police agencies. Justice Quart 19:503–536

Mayo L (1985) Leading blindly: an assessment of chief's information about police operations. In: Geller W (ed) Police leadership in America. American Bar Foundation, Washington, DC

McCall MW Jr, Lombardo MM (1983) Off the track: why and how successful executives get derailed. Center for Creative Leadership, Greensboro

McGregor D (1960) The human side of enterprise. McGraw-Hill, New York

National Research Council (2004) Fairness and effectiveness in policing: the evidence. Committee to review research on police policy and practices. In: Skogan W, Frydl K (eds) Committee on law and justice, division on behavioral and social sciences and education. The National Academies Press, Washington, DC

Pursley RD (1974) Leadership and community identification attitudes among two categories of police chiefs: an exploratory inquiry. J Police Sci Adm 2:414–422

Rowe M (2006) Following the leader: front-line narratives on police leadership. Policing Int J Police Strateg Manag 29:757–767

Schafer JA (2001) Community policing: the challenges of successful organizational change. LFB Scholarly Publishing, New York

Schafer JA (2012) Effective leadership in policing: successful traits and habits. Carolina Academic Press, Durham

Stamper NH (1992) Removing managerial barriers to effective police leadership: a study of executive leadership and executive management in big-city police departments. Police Executive Research Forum, Washington, DC

Steinheider B, Wuestewald T (2008) From the bottom-up: sharing leadership in a police agency. Police Pract Res 9:145–163

Tannenbaum R, Schmidt WH (1958) How to choose a leadership pattern. Harv Bus Rev 36:95–101

Van Maanen J (1983) The boss: the American police sergeant. In: Punch M (ed) Control in the police organization. MIT Press, Cambridge, MA, pp 275–317

Van Maanen J (1984) Making rank: becoming an American police sergeant. Urban Life 13:155–176

Witte JH, Travis LF, Langworthy RH (1990) Participatory management in law enforcement: police officer, supervisor and administrator perceptions. Am J Police 9(4):1–23

Wycoff MA, Skogan WG (1994) The effect of a community policing management style on officers' attitudes. Crime Delinq 40:371–383

Police Legitimacy

► Organizational Change and Police Legitimacy

Police Legitimacy and Police Encounters

John D. McCluskey[1], Robert E. Worden[2,3] and Sarah J. McLean[3]
[1]Department of Criminal Justice, Rochester Institute of Technology, Rochester, NY, USA
[2]School of Criminal Justice, University at Albany, State University of New York, Albany, NY, USA
[3]John F. Finn Institute for Public Safety, Inc., Albany, NY, USA

Overview

Research on police legitimacy is an area of growing interest among researchers and practitioners. The concept of legitimacy ties together issues of compliance, obedience, satisfaction, trust, and order. The fundamental recognition of its importance rests on the idea that police will more easily maintain order when civilians view the police as legitimate. Regimes, organizations, and individual officers lacking legitimacy, conversely, face crises based on undermined public support for their authority. The discussion that follows considers how legitimacy is influenced by individual encounters with authorities and in a reciprocal sense, provides a context for understanding how citizens relate to police in terms of future cooperation and compliance.

Fundamentals

Definition of Legitimacy

Legitimacy is an outlook, with cognitive and affective components as well as behavioral predispositions. People who believe that the police are legitimate trust the police to exercise their authority appropriately, identify with the police (Tyler and Huo 2002), have confidence in the police to do the right thing, and feel an obligation to obey the police. The sense of obligation to obey has been central to discussions of legitimacy, because insofar as legitimacy has behavioral implications, it points the way toward broader and more economical ways for the police to cultivate public support, cooperation, and compliance. More generally, we might say that when police are perceived as legitimate, they are thought to be entitled to have their authority reciprocated with obedience.

From the observation that legitimacy encompasses a sense of obligation to obey, it does not follow that obedience ensues reliably from legitimacy; the association between legitimacy – an outlook – and obedience (a behavior) is imperfect. Many studies of attitude-behavior consistency have reported only rather modest associations, and against that empirical backdrop, the reported relationships between legitimacy and obedience appear quite robust, as we summarize below, but they are far from perfect, and we might surmise that the strength of the relationship depends in part on the nature of the behavior.

Legitimacy, to the extent it is an attitude toward police, likely shares a high correlation with other attitudes, such as satisfaction with the police, which have been studied more extensively. Furthermore, a variety of research points to the likelihood that legitimacy is conceptually multidimensional, comprised of trust in motives, obligation to obey, and identification with the authority. This suggests that legitimacy is a composite of both affect toward authorities and their motives (e.g., trust and identification) and perhaps what might be considered an "action"-oriented component based on whether the authority's directives should be followed (obligation to obey) or can form the basis for citizen authorization (empowerment).

We would add that legitimacy is a property of different types of authority – of political systems; discrete political institutions, such as Congress or the Supreme Court; legal systems; specific legal institutions or agencies; and even private employers. We focus here on the legitimacy of the police, but we caution that inasmuch as the police exercise a mandate based, arguably, in law enforcement, police legitimacy is difficult to separate from the legitimacy of the legal order in which police are embedded. The police are, thus, tied directly to the concept of legitimacy as they represent the legal order and simultaneously are reliant on deference to that order as a reservoir of public cooperation, obedience to their authority, and trust. The relationship between these concepts and legitimacy is explored below.

Sources of Legitimacy

Wilbur Miller's (1977) comparative history of the New York City police and the London "bobbies" illustrates the tie between legal order and legitimacy. In England, for example, the police were viewed as impersonal representations of the established legal order, which was primarily divided along the lines of social class in the relatively homogeneous English society. As such the police had impersonal authority that stemmed from a general civilian deference to the larger political structure in 1800s England. Conversely, in New York, during the 1800s, there was no such easily agreed upon legal order, given the diversity of religious, ethnic, and social classes occupying the city. Thus, in this environment, police authority was believed to emanate from personal authority that was essentially earned respect and deference from the community which accumulated over the course of face-to-face encounters.

Legitimacy of one's role and action that ties directly to the law would appear to be advantageous to police as the bobbies enjoyed a reluctant respect even among the lower classes with whom they had a more adversarial relationship. However, this simplification glosses over several differences between the US police system and those of much of the remainder of the world. First, centralized authority in England allowed for uniform lawmaking and centralized police forces. The USA, conversely, has a proliferation of thousands of agencies, each which has to build its own reservoir of authority from a local (or State) body of law. Thus, fragmentation and decentralization push legal legitimacy to local levels in the US political system. While one might argue this is a weakness and causes police to diverge from a strictly legal mandate, there is some evidence that police function in such an environment, in the long term, may offer some advantages.

First, no local police agency in the USA represents the legitimacy of law in the United States. Hence, we should distinguish the legitimacy of the legal system, in the more abstract sense, from the legitimacy of police or of individual police agencies in a more concrete sense; the latter is much more proximate to the day-to-day work of police officers. The fragmented and decentralized nature of US policing, instead, gives police a local legitimacy and authority. By contrast, in the Fall 2005, French police chased several Muslim youths, and two were accidentally electrocuted. This situation boiled into many days of widespread riots against French police, since, in the more centralized governmental system, the police represent all of French government in some respects. As one can observe, the source of legitimacy in a political system has trade-offs and costs, some of which are not readily apparent in the day-to-day function of police.

Clearly, in the USA, with a proliferation of local police departments, legitimacy is homegrown by departments and the men and women who serve within those organizations. To paraphrase Tip O'Neill's keen observation, "Much policing is local." Thus, important questions to ask are as follows: Where does legitimacy come from, and what are the consequences of legitimacy? As such one might consider legitimacy to be both a dependent variable (something caused by other forces, such as respect for political institutions and the behavior of those institutions) and an independent variable (one that predicts changes in some outcomes, such as citizens' cooperation with authorities).

Where Does Legitimacy Come from?

The encounter between police and the public has been studied as a petri dish for the creation of legitimacy among the public. Observations of police-citizen contacts as well as post-contact surveys of citizens have formed the basis for a substantial body of literature pointing to the antecedents of legitimacy. These antecedents emanate from the treatment that citizens receive or perceive receiving in those encounters they have with authorities, such as police. The term *procedural justice* is used to describe treatment that is fair, respectful, and generally thought to lead to a sense that the processing one receives is appropriate and that the authority behind it is legitimate. *Satisfaction* or the extent to which citizens are content with the encounter is also correlated with legitimacy. These linkages are explored below.

Tom Tyler (2003, 2004) and other researchers have done extensive research to outline the relationship between processing, procedural justice, and legitimacy. Procedural justice has been conceived as encompassing two broad elements: *quality of treatment* and *quality of decision-making*.

Quality of treatment encompasses the tenor of interaction between citizens and authorities and includes elements such as *voice*, or the notion that people believe that they should be given an opportunity to tell their side of a story, explain their situation, and communicate their views, and *quality of interpersonal treatment*, the notion that people believe that they should be treated with dignity and respect.

Quality of decision-making comprises a set of signals that pass from authorities to citizens about how decisions are being made in any particular instance and whether they appear to be fair. Indicia of quality decision-making include how authorities demonstrate *trustworthy motives*. For example, people believe that authorities should care about their well-being and consider their needs and concerns, and they draw inferences about that when authorities explain their decisions and justify and account for their actions. Similarly, *neutrality* is a marker for quality of decision-making, since people believe that decisions should be made evenhandedly and with proper consideration of objective facts.

While it may be helpful for heuristic purposes to describe these elements of procedural justice as distinct constructs, they are conceptually and empirically intertwined with one another. Empirically, these features of police-citizen encounters tend to be strongly or at least moderately correlated, yielding only one (Reisig et al. 2007; Gau 2011) or two (Tyler 1990) factors in factor analysis. These patterns of association are open to (at least) two interpretations: It might be that authorities' behaviors are correlated (e.g., officers who are polite also tend to listen to citizens and explain themselves) and that citizens' perceptions are correlated (citizens who judge officers' actions as respectful also tend to perceive them as attentive), or some combination of both sets of forces.

Survey research has repeatedly shown that the procedural justice that people subjectively experience in their encounters with the police is directly related to their satisfaction with the encounters and, moreover, to the legitimacy that they attribute to the police. People are more satisfied with their encounters with the police when they perceive that police acted in a procedurally just fashion. Furthermore, people have greater trust in the police, and feel a greater obligation to obey police instructions, when they have experienced procedurally just encounters with the police. This is the sense in which legitimacy can be "created" by police in their day-to-day interactions with the public and also illustrates its close tie to citizen satisfaction. Satisfaction and legitimacy can also be eroded by police, if they act with procedural injustice, if they are impolite or disrespectful, and if they interrupt citizens, ignore what citizens have to say, or do not permit citizens to explain themselves. The core linkage posited is that elements of processing which leave citizens unsatisfied (e.g., disrespect, bias) undermine legitimacy, while those which amplify satisfaction (e.g., respectful treatment, explaining decisions) similarly enhance legitimacy.

For example, in one of the earliest studies, Tyler and Folger (1980) surveyed a random sample of Evanston (Illinois) residents in 1979, identifying among the respondents those who reportedly had a personal experience with Evanston police in the previous 5 years, by either calling for assistance or being stopped by police. About each type of contact, they asked respondents to indicate the outcome (police had or had not taken care of the problem, and police issued or did not issue a citation, respectively), whether police treated them fairly, and how satisfied they were with the contact. They found that both outcome and perceived fairness affects citizens' satisfaction in both types of contacts, and also that perceived fairness affects citizens' more general evaluations of the police. Similarly, Tyler (1988) surveyed a sample of Chicago residents, nearly half of whom reportedly had a direct experience with Chicago police and/or courts in the preceding year. He found that procedural fairness affects citizens' satisfaction with both their outcomes and their treatment in the particular case, and also that procedural fairness affects citizens' more general evaluations of and support for the authorities.

Moreover, this research makes it clear that citizens' subjective experience is not determined entirely or even primarily by the outcomes that they receive. Put differently, neither substantive justice (whether one deserves what one objectively receives) nor distributive justice (whether one receives what others who are similarly situated receive) is as important to attitude formation as procedural justice (the manner in which justice is delivered). Even citizens whose outcomes are unambiguously unfavorable – who, for example, are cited for a traffic violation – may be satisfied with their experience if police are procedurally fair. It is people's normative expectations about how authority should be exercised, more than their instrumental considerations about the favorability of outcomes, that drive their assessments of their own experiences.

Since much of the research on this topic is based on cross-sectional survey data, the analysis of which forms the basis for only rather tentative inferences about cause and effect relationships; caution in drawing conclusions is advised. Several studies suggest that citizens' prior attitudes toward police shape both their subjective experiences, including perceptions of procedural justice, and their more general attitudes toward the police, including legitimacy, thereby contributing to a partially or wholly spurious relationship between procedural justice and legitimacy. But when these relationships have been examined in panel surveys (e.g., Tyler 1990; Tyler and Fagan 2008), so that the temporal ordering of the variables can be properly specified, the results have shown that even while prior attitudes do influence judgments about procedural justice and legitimacy, procedural justice has independent effects on subsequent legitimacy.

The structure and context of police-citizen encounters vary considerably, with respect to visibility, the nature of the issue that prompted police involvement, the "situational status" of the citizen (as complainant, suspected offender, witness), and the neighborhood setting of the encounter. The structure and context of interaction between police and citizens likely contribute both to objective features of the encounter (the routines of interaction or typical procedure) and how the interaction is experienced by the citizen. One structural feature that has an important bearing on citizens' subjective experience is how the encounter is initiated – at the request of a citizen or on the authority of the officer. Citizen requests may serve, at least to a degree, to legitimate the involvement of the police, as the officer(s) is responding to a request for police assistance. For at least one of the citizens who participate in such an encounter, the encounter is voluntary. When police intervene on their own initiative, however, as they do when they stop traffic violators or other people who arouse police suspicion, citizens' participation is not voluntary but obligatory, and citizens may be more apt to question the propriety of police involvement (Reiss 1971). Such "obligation encounters" (Moore 2002) pose greater challenges for police in meeting citizens' expectations for procedural justice. Skogan (2006) found among Chicago residents a significant disparity in satisfaction between those whose contacts with the police were self-initiated (78 % satisfied) and those whose contacts were initiated by police (58 % satisfied). Furthermore, procedural justice criteria appear to have an even greater bearing on citizens' satisfaction in police-initiated contacts than they do in citizen-initiated contacts. Tyler and Folger (1980) found that procedural fairness has a larger effect when citizens are stopped than when citizens call for assistance. Wells (2007) surveyed citizens involved in any of three types of contacts with the Lincoln (NE) police: victims of crimes, drivers involved in vehicle crashes, and drivers who received citations. He found that in each type of situation, citizens' judgments about procedural justice affect their ratings of overall police performance in the encounter, though procedural justice mattered less for crime victims, who typically call for assistance, than for motorists who are stopped for traffic violations.

We would also note that the effects of satisfactory experiences many not be as powerful as the effects of unsatisfactory experiences. Skogan (2006) stresses the asymmetry in the effects of positive and negative experiences with the police, finding in survey data on Chicagoans that their confidence in the police is very weakly or not at all related to positively evaluated contacts, but negatively evaluated contacts have a substantial detrimental effect on confidence. Skogan (2005) also found that citizens' satisfaction with their contacts is shaped by procedural justice. More specifically, whether police were polite and whether they were helpful were the most important predictors of satisfaction with citizen-initiated encounters, and whether police were polite and whether they were perceived as fair were the most important predictors of satisfaction in police-initiated encounters. Satisfaction was also enhanced if police explained what action they would take. These beneficial effects of procedural justice did not, however, translate "up" to confidence in the police. Other studies have also found asymmetrical effects, though not to the degree that satisfactory contact had no effect on more general attitudes (e.g., Reisig and Parks 2000).

No treatment of police legitimacy would be complete without acknowledging the attitudinal disparities across race and ethnicity: African-Americans hold less favorable views of the police than whites do, with Latinos typically in between. For example, Weitzer and Tuch (2006) conducted a national survey and found variations in satisfaction with police across race and ethnicity, as 86 % of White, 80 % of Hispanic, and 73 % of Black respondents reported being satisfied or very satisfied with the police department in their city. More interestingly, perhaps, they also found that the effect of race on satisfaction was mediated by what citizens perceived as police actions. For instance, those believing police were involved in misconduct, such as improper stops, verbal abuse, excessive force, or corruption, were less satisfied with police. Further, they found that beliefs about misconduct were shaped not only direct, personal experiences but also by vicarious experiences (those communicated to individuals by others). Thus, the police-citizen encounter can have far-reaching effects that ripple through a community.

The overriding racial issue in policing for the past 15 years has been racial profiling, with countless police departments across the US collecting and analyzing data on the race of the people stopped by their officers in order to address concerns about racial bias. The analyses are seldom conclusive with respect to bias, but some research suggests that procedural justice plays an important role in shaping citizens' perceptions of bias. For example, Tyler and Wakslak (2004) surveyed residents of Oakland and Los Angeles who had been stopped by police. Those who thought that police were polite, treated them with dignity and respect, and showed concern for their rights were much less likely to believe that they had been profiled. Other elements of procedural justice had similar, albeit more modest effects on profiling attributions. More recently, Gau and Brunson (2010) interviewed 45 youths from low-income areas of St. Louis, MO, and reported findings consistent with the theme that perceptions of unfair treatment undermined youths' support for the police.

Thus, legitimacy appears to be causally linked back to the basic concept outlined above: those situations where there is an obligation to obey the authority, which can be boiled down to several simple aspects of police encounters. Were the police invited to intervene (as compared to proactively inserting themselves into a situation)? Did the police clearly explain the rationale behind their decisions? Did police allow the individual to tell his or her side of the story? Did the police listen to the individual? Did the police act courteously and show respect to the individual? These types of questions, when answered affirmatively, illustrate the procedural justice of police encounters; they also represent important antecedents of legitimacy, and, not surprisingly, they are closely correlated with citizen satisfaction.

Legitimacy is intrinsically important, but it is important also because it is consequential. Legitimacy affects citizens' acceptance of police decisions, their cooperation with the police, and even their compliance with the law more generally. We turn now to a consideration of these outcomes.

Legitimacy and Outcomes: Support and Citizen Self-Regulation

Police rely heavily upon citizens for support and self-regulation. Supporting citizen behavior and attitudes is important to the police for the maintenance of order and the control of crime. Police effectiveness depends upon citizen involvement in community activities, reporting crimes and giving information to the police, which would be considered as *cooperation*. *Empowerment* of the police is also a form of (active or passive) support, in the sense that citizens who grant police the capacity to exercise authority with fewer rather than more restrictions, and are supportive of (or do not oppose) police tactics and strategies, offer important if implicit consent from the governed to the authorities.

Self-regulation, that is, obeying laws and complying with police requests, represents another area in which police rely on citizens. Thus, if police request citizens to leave a scene or to discontinue making noise at a party, they are relying on citizen self-regulation (Mastrofski et al. 1996). In the larger sense, police rely on everyday self-regulation or self-control in refraining from committing minor (e.g., littering) and major (e.g., robbery) crimes.

One might think that citizens comply and cooperate due to the overall legitimacy of the legal system in which police-citizen contacts are embedded. This, for example, might help us understand why an English citizen would obey a bobby but someone in New York would be reluctant to obey a police officer in the 1880s. Thus, in that larger sense, legal legitimacy can predict cooperation in face-to-face encounters. Analogously, one might look across US agencies and examine the relationship between the police and the community as an indicator of overall police legitimacy. Thus, in locales where police-community ties are strong, one might expect greater cooperation and compliance from citizens in encounters with police.

Legitimacy and Support

It is a well-understood fact that police rely on the public for information and mobilization to identify and solve problems. Police rely on citizens to report crimes and to act as witnesses and information providers on scene. One question that arises is whether there is a link between legitimacy and the willingness to cooperate with the police. If so, then police that engage in promoting legitimacy will build public support and cooperation.

Evidence linking legitimacy and cooperation has slowly accumulated from the body of research on procedural justice. Tyler and Fagan (2008) used a panel data survey design with 830 respondents to test the relationship between legitimacy and cooperation. This allowed them to measure legitimacy at two points in time and determine how contact with police influenced subsequent legitimacy and self-reports of intention to cooperate with the police. Consistent with other research, procedurally just treatment yielded increases in perceived legitimacy of the police. More specifically, those respondents reporting they had been treated fairly and with respect had measureable increases in legitimacy, controlling for whether they received a positive or negative outcome from the police.

To measure cooperation, the researchers asked citizens how likely they were to call the police and report a crime, help find a criminal, and report suspicious activity. A parallel scale regarding helping the community fight crime was constructed from similar measures asking whether the respondent was likely to volunteer time to help the police, to patrol streets with others, and to attend community meetings about crime. In multivariate analyses, taking advantage of the panel design capacity for causal analysis, Tyler and Fagan (2008) found that those who reported police as being more legitimate were more likely to report intending to cooperate directly with the police and also cooperate with others in efforts to deal with community crime.

Sunshine and Tyler (2003) have examined the relationship between empowerment and legitimacy among a panel of New York City respondents. Empowerment represents the extent to which citizens recognize a greater scope for police action and discretion. The findings from that research indicated that citizen perceptions of legitimacy and to a lesser extent distributive fairness predicted respondents' extent of empowerment.

Legitimacy and Self-regulation

Research on self-regulation has focused on whether citizens obey the commands of police immediately as well as in the longer term with respect to both minor offending (e.g., traffic or other violations) and serious crime (e.g., violence). The research on immediate compliance has been derived from observations of what police do in interactions with citizens and whether citizens comply with police requests for self-control. For example, McCluskey (2003) studied face-to-face interactions between police and 939 suspects in St. Petersburg, FL, and Indianapolis, IN. The results of that research indicated that, with regard to immediate compliance, it appears that citizens who are given opportunities to tell their side of the story and are treated respectfully are significantly more compliant than citizens not afforded such courtesies during the police-citizen encounter. This type of treatment and behavioral outcome lends support to the linkage between treatment by authorities, legitimacy, and compliance.

Sunshine and Tyler (2003) collected survey data regarding the link between citizens' reports of compliance and the legitimacy of police. That research indicated a similar effect of perceptions of legitimacy, which provided a robust explanation of citizens' self-reported compliance with the law. Thus, observed and self-reported behaviors, among the general public, are influenced by legitimacy. More simply, as legitimacy increases either as measured by perception of citizens, or by the fair procedures consistent with legitimacy, compliance increases. It is important to note that the observed police-citizen contacts and self-reports involve typically more trivial disputes or lawbreaking. Thus, one is obliged to ask: Does the perceived legitimacy of the police and law influence serious misbehaviors in a similar fashion?

With regard to legitimacy's relationship to longer-term patterns of compliance among serious offenders, several studies involving arrestees for spousal assault, intoxicated drivers, and serious delinquents indicate substantial links. Paternoster and his colleagues (1997) were the first to study the influence of procedural justice on serious criminal activity. Using data collected for the Milwaukee Domestic Violence Experiment, they found among a sample of males arrested for spouse assault that perceptions of procedural justice influenced recidivism. All of these offenders experienced an unfavorable outcome (arrest), but some of them perceived their treatment as procedurally fair, and those who did were significantly less likely to recidivate. Moreover, the recidivism rate for those who were arrested and who felt that the process was fair was similar to the recidivism rate for those who had been warned only. These findings suggest that legitimacy impacts long-term legal compliance to the extent that perceived fair treatment reported by these offenders is related to their sense of fair and legitimate treatment.

Based on a sample of offenders arrested for drunk driving in Australia, Tyler and his colleagues (2007b) evaluated the impact of reintegrative shaming and procedural justice on recidivism. Recidivism was based on self-reports in a 2-year follow-up period and on police records for a 4-year follow-up period. The study tested the hypotheses that procedural justice and restorative justice conferences (as opposed to traditional court proceedings) would foster more positive assessments of the law and lead to higher levels of law abidingness. First, they found that the treatment (restorative justice conference) was not associated with lower recidivism (measured in terms of self-reported drunk driving or police records), though they did detect a direct effect on self-reported efforts to curb driving while drunk. With respect to the influence of legitimacy on behavior, they found that those who viewed the police as more legitimate were less likely to recidivate. This held true when behavior was measured in terms of self-reported levels of drunk driving and when it was measured using police records. The extent to which the initial experience was viewed as procedurally just shaped subsequent views of legitimacy, which in turn led to lower levels of recidivism.

Bouffard and Piquero (2010) analyzed the Philadelphia Birth Cohort data follow-up for 212 juveniles with police contacts. Those characterizing their initial contact as unfair had substantially higher frequency and longer periods of offending compared to those who did not. Fagan and Piquero (2007) examined a sample of 1,355 juveniles referred to courts in Philadelphia and Phoenix and found that assessments of procedural justice had a strong direct impact on perceptions of legitimacy among these serious offenders. In turn, legitimacy exercised a substantial direct impact on self-reported aggressive and income offending.

In sum, across multiple methodologies, law-abiding and lawbreaking samples, and less serious and serious behavioral outcomes, it appears that where police legitimacy and behaviors consistent with procedurally fair policing are greatest, citizen self-regulation, both immediately and in the longer term, is greater. More specifically, citizen self-regulation is positively associated with police legitimacy.

The Policy Consequences of Legitimacy and Future Research

As the notion of procedural justice (Tyler 1990) and legitimacy came to the forefront of academic consideration, it was recognized that at least some of what police might accomplish required thinking about the issue of legitimacy and how citizens come to see authorities as legitimate.

This sparked research on procedural justice and has led to the recognition that legitimacy is to a significant degree in the hands of police as they relate to the public in everyday encounters. Clearly police misbehavior, corruption, and rudeness can do much damage to legitimacy. But the more optimistic side of the research indicates that police can repair and create legitimacy via politeness, taking time to explain their decisions and treating individuals in a fair manner. This is a very economical solution to the police-community relation issue which has at times seemed intractable and dominated by *us vs. them* rhetoric.

Field stops by police have attracted considerable attention and generated considerable controversy, especially in New York City, where police have in recent years documented hundreds of thousands of "stop, question, and frisk" contacts annually. Proponents tout the value of such stops in confiscating illegal firearms and, more generally, reducing crime. Critics hold that the stops exhibit a racially biased pattern and, moreover, detract from the legitimacy of the police, as stops for low-level offenses may be viewed by the citizens who are stopped – and by their family, neighbors, and friends – as harassment (Gau and Brunson 2010; Warren 2011). Extant research does not suffice to resolve this controversy, even if it could be resolved with reference to empirical evidence, but we would note that several questions are important to consider in this connection: (1) How often are stops made? (2) For what reasons are stops made? (3) Are stops concentrated in places and at times at which crime occurs? (4) Are the stops conducted by officers in conformance with principles of procedural justice?

Some have called for a "procedural justice model" of policing (Schulhofer et al. 2011; Meares 2009). A procedural justice model of policing does not hold that police platoons become grin-and-wave squads. The coercive authority of police is, as Bittner (1970) observed long ago, their unique occupational prerogative, and it enables them to fulfill their role in society: negotiating or imposing solutions upon emergent problems. Procedural justice is not about whether but how authority is exercised. Police who operate according to a procedural justice model will of course need to exercise their authority, but they do so sparingly, judiciously, and in accordance with principles of procedural justice: They honor legal limits on their authority; they treat the people with whom they interact – even the people whom they arrest – with dignity and respect; they allow – even invite – people to explain themselves and their situations; and they explain their actions.

Police departments that adopt such a model would, we suppose, establish and enforce expectations that their officers will exercise their authority in these ways. Their chief executives make procedural justice an explicit priority. They embody their expectations in department policies and procedures. For example, the Redlands (CA) police revised their traffic stop procedures, as their chief explains, "instead of starting out the interaction between the motorist and the police officer, where we typically walk up and say, 'Let me see your license and registration, please,' then we walk back to our car, and the next time you see us is when we're handing you a ticket or whatever. Turning that around and saying, 'Good afternoon, my name is Officer Bueermann, and I've stopped you because you were speeding. Is there some reason you might have been speeding?' And then we give you the opportunity to explain to us what was going on in your head, or why you were doing . . ." (quoted in Tyler 2011). They train their officers in proper police-citizen interaction (see, e.g., Rosenbaum and Lawrence 2012). They monitor the available indicators of police performance, such as complaints and uses of force, and recognizing the limits of these indicators, they make supervisors responsible for spot-checking the quality of police-citizen encounters. They might even develop more systematic measures of such performance. And they treat officers with the same procedural justice that they demand of officers in their encounters with citizens, thereby nurturing the legitimacy of the organization and its rules in the eyes of its human resources (Tyler et al. 2007; Wolfe and Piquero 2011).

Beyond these basic suppositions, however, extant research does not provide much guidance for police administrators who are persuaded that their agencies would benefit from systematic efforts to engage in procedurally just policing. Hardly any empirical evidence is available for these purposes. One study showed that police disrespect toward citizens – a form of procedural injustice – is most likely a response to citizen disrespect toward the police, but it is also affected by a citizen's lack of self-control (e.g., in the form of intoxication) and social status (Mastrofski et al. 2002). The same study also detected some differences across jurisdictions that might be attributable to differences in department administration. But at this time the translation of research on procedural justice and police legitimacy into the practice of procedurally just policing is largely a matter of trial and error.

As research on legitimacy and procedural justice yields more insights regarding the specific content of just and fair processing and its consequences, it becomes imperative to begin developing key elements into the content of everyday police-citizen contacts. Answering questions such as what dosage level and content of procedurally just processing are related to the formation of a "sense of legitimacy" will help to tune police action to desired outcomes. If the current research trajectory on legitimacy matures into an action research agenda, then perhaps the police will find they have always had the tools for building legitimacy, cooperation, and compliance at their disposal. This would have an interesting symmetry with the idea of personal authority and the difference between cops and bobbies outlined at the beginning of this entry. More simply, the creation of legitimate police authority, in a decentralized republic, does in fact lie in the day-to-day contact between police officers and citizens. Thus, legitimacy is indeed local and continuously recreated or undermined in those everyday encounters.

Related Entries

- ▶ Causes of Police Legitimacy
- ▶ Organizational Change and Police Legitimacy
- ▶ Police Self-legitimacy
- ▶ Police-Led Interventions to Enhance Police Legitimacy
- ▶ Procedural Justice and Cooperation
- ▶ Procedural Justice, Legitimacy, and Policing

Recommended Reading and References

Bittner E (1970) The functions of the police in modern society. Oelgeschlager, Cambridge, MA

Bouffard LA, Piquero NL (2010) Defiance theory and life course explanations of persistent offending. Crime Delinq 56(2):227–252

Fagan J, Piquero A (2007) Rational choice and developmental influences on recidivism among adolescent felony offenders. J Emp Legal Stud 4(4):715–748

Gau JM (2011) The convergent and discriminant validity of procedural justice and police legitimacy: an empirical test of core theoretical propositions. J Crim Justice 39(6):489–498

Gau JM, Brunson RK (2010) Procedural justice and order maintenance policing: a study of inner city young men's perceptions of police legitimacy. Justice Q 27(2):255–279

Mastrofski S, Snipes J, Supina A (1996) Compliance on demand: the public's response to specific police requests. J Res Crime Delinq 33:269–305

Mastrofski SD, Reisig M, McCluskey JD (2002) Police disrespect toward the public: an encounter-based analysis. Criminology 40:519–552

McCluskey J (2003) Police requests for compliance. LFB Scholarly Press, New York

Meares T (2009) Barrock lecture on criminal law; the legitimacy of the police among young African American men. Marquette Law Rev 92:651–666

Miller W (1977) Cops and Bobbies. University of Chicago Press, Chicago

Moore MH (2002) Recognizing value in policing: the challenge of measuring police performance. Police Executive Research Forum, Washington, DC

Paternoster R, Brame R, Bachman R, Sherman L (1997) Do fair procedures matter? Law Soc Rev 31:163–204

Reisig MD, Parks RB (2000) Experience, quality of life, and neighborhood context: a hierarchical analysis of satisfaction with police. Justice Q 17(3): 607–630

Reisig MD, Bratton J, Gertz MG (2007) The construct validity and refinement of process-based policing measures. Crim Justice Behav 34:1005–1028

Reiss AJ Jr (1971) The police and the public. Yale University Press, Connecticut

Rosenbaum DP, Lawrence DS (2012) Teaching respectful police-citizen encounters and good decision making: results of a randomized control trial with police recruits. Center for Research in Law and Justice, University of Illinois at Chicago, Chicago. www.nationalpoliceresearch.org

Schulhofer SJ, Tyler TR, Huq A (2011) American policing at a crossroads: unsustainable policies and the procedural justice alternative. J Crim Law Criminol 101:335–374
Skogan WG (2005) Citizen satisfaction with police encounters. Police Q 8(3):298–321
Skogan WG (2006) Asymmetry in the impact of encounters with police. Police Soc 6:99–126
Skogan W, Frydl K (eds) (2004) Fairness and effectiveness in policing: the evidence. The National Academies Press, Washington, DC
Sunshine J, Tyler T (2003) The role of procedural justice and legitimacy in shaping public support for policing. Law Soc Rev 37:513–548
Tyler TR (1988) What is procedural justice? Criteria used by citizens to assess the fairness of legal procedures. Law Soc Rev 22:103–135
Tyler TR (1990) Why people obey the law. Yale University Press, New Haven
Tyler TR (2003) Procedural justice, legitimacy, and the effective rule of law. In: Tonry M (ed) Crime Justice 30:431–505
Tyler TR (2004) Enhancing police legitimacy. In: Skogan WG (ed) Ann Am Acad Polit Soc Sci 593:84–99
Tyler T (2011) Race, police legitimacy and cooperation with the police. National Institute of Justice, research for the real world seminar. Transcript Accessed on 20 May 2013 http://www.nij.gov/nij/multimedia/presenter/presenter-tyler/data/resources/presenter-tyler-tr anscript.htm
Tyler TR, Fagan J (2008) Legitimacy and cooperation: why do people help the police fight crime in their communities? Ohio State J Crim Law 6:231–275
Tyler TR, Folger R (1980) Distributional and procedural aspects of satisfaction with citizen-police encounters. Basic Appl Soc Psychol 1(4):281–292
Tyler TR, Huo YJ (2002) Trust in the law. Russell Sage, New York
Tyler TR, Wakslak CJ (2004) Profiling and police legitimacy: procedural justice, attributions of motive, and acceptance of police authority. Criminology 42(2):253–281
Tyler TR, Callahan P, Frost J (2007a) Armed and dangerous(?): motivating rule adherence among agents of social control. Law Soc Rev 41:457–492
Tyler TR, Sherman LW, Strang H, Barnes GC, Woods DJ (2007b) Reintegrative shaming, procedural justice, and recidivism: the engagement of offenders' psychological mechanisms in the Canberra RISE drinking-and-driving experiment. Law Soc Rev 41(3):553–586
Warren PY (2011) Perceptions of police disrespect during vehicle stops: a race-based analysis. Crime Delinq 57:356–376
Weitzer R, Tuch S (2006) Race and policing in America: conflict and reform. New York, Cambridge
Wells W (2007) Types of contact and evaluations of police officers: the effects of procedural justice across three types of police-citizen encounters. J Crim Justice 35:612–621
Wolfe SE, Piquero AR (2011) Organizational justice and police misconduct. Crim Justice Behav 38:332–353

Police Line-Level and Agency Accountability

Michael E. Buerger
Bowling Green State University, Bowling Green, OH, USA

Overview

The exercise of discretion – the ability to choose among options for resolving a situation – is bounded by officers' "intuitive understanding of situational exigencies" (Bittner 1970). Police officers are employees of the public and agents of the state, so there is a general expectation that they will observe the law, treat all citizens equally and with basic respect (even those whose behavior has been the antithesis of respectful), and generally work for the betterment of the community as a whole.

Society must trust the officer's instincts to make correct decisions in thousands of dissimilar incidents. Law enforcement agencies must create and maintain proper means for shaping and verifying those decisions along ethical lines, to increase the likelihood that officers will choose well. It must also have a proportionate way to respond to misjudgments, whether based on simple error, lack of proper preparation, or some more venal animus.

The good work of policing has been tarnished by a long series of public scandals exposing abrogation of responsibility and lingering distrust generated by police misconduct: incompetence, callousness, brutality, corruption, and criminal acts while in uniform dominate the list of offensive behaviors. At the organizational level, systemic biases – lack of services to a particular population, racial or other profiling, tolerance of brutality, and corruption of the legal process – have been the dominant problems, along with a persistent unwillingness of some agencies to investigate properly any allegations of wrongdoing by employees.

The antidotes to sins of omission and commission by police officers and agencies have been parallel efforts to upgrade police service and improve supervision. The broad philosophical push toward ***professionalism*** is one such effort: it is manifested in the hiring, training, and supervisory functions, reflecting articulated values. A narrower bureaucratic focus on ***rules and regulations*** complements professionalism by clearly identifying proper and unacceptable behaviors. Each has positive and negative aspects, and sometimes an incisive eye is required to distinguish true adherence to professional precepts from linguistic smoke screens designed to protect less desirable forms of policing (Klockars 1991). These approaches are combined at the agency level in the convention of state or national ***accreditation***.

In addition, police work continues to be subject to external scrutiny and judgment: prosecutorial and judicial review of individual cases and ***political oversight*** of agency operations and results. State and federal oversight occasionally is brought to bear at the organizational level in the form of ***consent decrees*** when agencies are or appear to be unresponsive to other corrective measures (Walker and Macdonald 2009). Over the last quarter century, ***citizen review*** of police activity has been promoted as a local-level control of police behavior, with mixed results.

Professionalism

For years, "professionalism" was claimed by the police because of a single attribute: the desire to govern their own affairs autonomously, as had been the case for the three classic professions of law, medicine, and the clergy. The catchphrase that "only a cop can judge another cop" was passed from generation to generation of police as a cultural meme, despite a lack of supporting evidence. For those generations, the other two attributes of a profession – a unique body of knowledge acquired through years of study and a service orientation – were more honored in the breach than the observance, despite the extensive community service provided by individuals.

Since the crisis of confidence in the police that marked the 1960s, the professional model has made great strides toward filling that void. While the nature of the criminal justice system makes it unlikely that the police will ever be free of external review, the understanding of the relationship of professional conduct to *police legitimacy* in the eyes of political sovereigns is now widely recognized and serves as a basis for both professionalization and bureaucratic mechanisms. Cultivating professionalism begins with hiring the right people, then training them with a proper orientation to the work to be done, and continuing to reinforce that orientation with adequate supervision and performance feedback.

Hiring

The hiring process is integral to creating and sustaining professional orientation and demeanor. In the best case, it involves a recruiting process that presents a clear articulation of the qualities expected of new employees and of the nature of the job to be performed. A good hiring process will select candidates who are intelligent, both well educated and attuned to important concepts like cultural diversity and changes in the social and legal environment. Critical thinking skills, a sound ethical grounding, flexibility, physical fitness, and mental toughness are all important attributes for future police officers. In the case of lateral transfers (experienced police officers from other departments joining a new agency), demonstrated good performance and ethical behavior in prior assignments are critical.

The hiring process has its own variance and weaknesses. The hiring pool for a particular agency may be weak, forcing agencies to hire the best candidates from a thin field of applicants. In some cases, too rapid an expansion of a force under political duress leads to cutting corners, trading integrity for numbers (see, e.g., and Harrison and Flaherty 1994, et. seq.). Written tests vary widely in both their rigor and their applicability to policing, and finances or politics may dictate the choice available: civil service tests for larger municipalities may test only for basic city employment, not specific police needs. Physical agility tests tend to be basic fitness gauges and may not indicate mental toughness for more rigorous training or more complex motor skills needed for police work.

Training

Though the best-case scenario for hiring is often limited by resources available to the agency, the process is now bolstered by a formal external control: state certification of police officers under Police Officer Standards and Training (POST) Board standards or other state peace officer requirements. The judgment of the hiring agency must be reinforced by a judgment of fitness by a neutral body with broad responsibilities. The long-standing model for certification and licensing began with the agency hiring a candidate, then putting the rookie through academy training, followed by an FTO period, and then annual retraining and firearms certification. Under that model, the background check preceded hiring and training.

Financial pressures created a second route to police employment, known as preservice certification. Individual candidates make themselves more marketable by paying for their own certification, either at a community college program or by attending a regular state academy. The community college option often involves earning an Associate's Degree as well as police certification but takes a longer period of time. The regular academy process tends to be shorter, between 480 and 650 hours, though many municipal and State Police academies are longer. In both models, elements of professionalism are embedded in all aspects of the training regimen.

Attaining police certification and licensing demonstrates a basic level of physical skills and intellectual knowledge about police tasks and expectations. It does not necessarily provide insight into a person's mental orientation to the job, aside from an ability to give "the right answer" on a test. Background checks are intended to fill the void but may take a backseat to the financial benefit of hiring an already-certified officer who can go to work after an abbreviated orientation period. A similar flaw attends the lateral-hiring route, where problematic officers may leave a department with a "golden handshake" agreement that will mask their troublesome behavior from a casual background inquiry.

Field Training

New hires are provisional employees in most cases, serving a probationary period with no guarantee of employment, civil service protection, or union eligibility until they have been more fully tested and observed. Field training pairs new officers with an experienced officer who serves a dual role as coach or tutor and evaluator. Where possible, rookies will work with more than one field training officer (FTO) in order to gain additional perspective and avoid the appearance of personal bias.

The FTO period places the rookie in a different testing environment, gauging his or her reactions and applying skills and knowledge to real-world incidents. Performance is graded, and feedback provided, during the experience, always with the intent to reinforce good behavior while correcting substandard work. If a rookie's work and attitudes do not meet acceptable levels at the conclusion of the FTO period, additional training may be offered to bring performance up to standard. However, if serious deficiencies remain, a sub-performing officer may be separated from the agency at any time prior to the completion of the probationary period.

Supervision and Performance Review

The agency's responsibility to its employees extends through their careers and takes many forms. Providing training and education to update and upgrade skills is essential to performance levels and opportunities for promotion or other career development. Providing feedback and supervision validates proper behaviors and corrects deviations from acceptable norms of effort, attitude, and exercise of discretion. Organizational expectations may be articulated in one of two main ways: as organizational *values* or as a series of specific rules and regulations.

Rules and Regulations. The traditional form of organizational direction of its employees is The Rulebook. A definitive set of do and don't regulations are codified in a growing register.

The Rulebook is provided to each new employee, and they are expected to read it, understand it, and internalize its requirements on their own.

The basic problem with The Rulebook is the same problem that leads to the need for discretion: it is impossible to anticipate every possible need the organization will have. Rules and regulations are also amenable to strict interpretation. New rules tend to be added to the book when an employee does something that is clearly unethical or unprofessional, but is not covered by a specific existing rule; the void is then filled retroactively, against future misconduct of that one particular type. Few, if any, of these rules have a specific empirical basis: even though no evidence exists that the conduct would have been widespread without the rule, the lack of subsequent violations is considered sufficient justification to maintain the rule.

Because police officers tend to be concrete thinkers, they are fairly good at following the rules, with the mind-set that "As long as I follow the rules, they [the administration] can't punish me." In public discourse, this is referred to as "being in compliance" with organizational expectations. Rule-bound behavior is not always effective at fulfilling organizational missions, however, and pushing the limits of specific rules can extend as far as breaking the law (see Crank and Caldero 2010).

While most issues of error or minor misconduct are handled informally, persistent problems are addressed through a system of *progressive discipline*, which applies to relatively low-level infractions or performance deficiencies: major problems, such as criminal conduct on or off duty, may lead to immediate suspension or dismissal. For the lesser offenses – usually violations of in-house rules – progressive discipline almost always begins with a verbal warning from a superior officer. No record is kept, as the presumption is that the employee will understand and comply with the correction (the supervisor may make a note of the date and nature of the warning in her or his own records, but the matter is not part of the employee's personnel file). If misbehavior or substandard behavior persists, a written warning may be issued: typically, these warnings are entered into the employee's file for a limited period of time and may factor into the next annual performance review. Letters of reprimand are the next step, followed by suspension with or without pay, and finally a move to dismiss the officers if the problem behavior remains uncorrected or worsens.

Progressive discipline is a process of documenting both the employee's errors or infractions and the agency's response to them. Like the use of force continuum in a different context, it constitutes a measured response with graduated penalties but can quickly be asserted at higher or lower levels depending upon the employee's response. The documentation that an employee has been offered additional counseling and guidance, and has failed to respond positively, is an essential element to support more drastic measures such as suspension without pay or dismissal.

The tension between being a public servant (an agent of the state) and a police officer's Fourth Amendment property right in employment is salient in the disciplinary process. As a general rule, agencies may compel their employees to answer questions about misconduct truthfully; if the employee refuses to answer or fails to answer honestly, they can be dismissed. However, the Fifth Amendment guarantees a right against self-incrimination in criminal matters. The 1967 Supreme Court decision of *Garrity v. New Jersey* established that "information which is compelled by government" as a condition of employment cannot be used in criminal proceedings. The Garrity Rule was expanded in the 1968 *Gardner v. Broderick* decision, which indirectly affirmed the right to dismiss employees for not answering "relevant questions about official duties" but reaffirming the protection against subsequent use of any incriminating information thus revealed.

Values. An alternative to the narrowly defined specifics of rules and regulations is a values orientation: broad declarations of expectations, such as observance of the law, fair and ethical treatment of all citizens, and the like. The emerging trend called "transformational leadership" rests in large part upon just such an orientation:

desirable conduct is clearly visible as a large picture, not a complex tapestry of small, narrow rules. Values are subject to interpretation, which is a cause of concern for unions and other representatives, but they constitute a broad platform that avoids the "if it's not in The Rulebook" approach that hinders effective response to clearly unacceptable behavior.

Performance Measures

Most departments have a formal means of assessing an employee's overall work performance, usually administered on an annual basis. Quantifiable outcomes such as arrests, traffic citations, calls handled, and occasional contributions to other efforts (providing critical information to detectives to help solve a case, for instance) will be recorded and evaluated. Areas of outstanding, acceptable, and inadequate job performance are scored on a scale and become part of the employee's permanent record, used for decisions about promotion and requests for special duty (Mastrofski 2003).

The disadvantages of performance evaluations can outweigh their advantages if the organization's oversight is lax. Some supervisors may evaluate all their employees in the middle range, avoiding the need to justify "why he was rated higher than I was." Others may give maximum marks to all subordinates ("everyone's a swan" in the words of one commander), especially if others in the organization are known to do the same: few supervisors will knowingly disadvantage their subordinates' chances for promotion or other improvements. Even without those attitudinal limits, transfers and promotions may present new supervisors with the responsibility to evaluate subordinates whose work they have had only a week or two to observe.

Some performance indicators are administered by the agency but are linked to the external certification. The most visible of these, and arguably the most important, is the firearms proficiency requirement. Any officer who fails to maintain his or her proficiency with regulation firearms is ineligible for street duty. The same may be true for other basic skills, including physical fitness standards: not only is every officer at risk of intense physical activity, their ability to control violently resisting individuals is critical to the safety of their brother and sister officers.

Performance measures may or may not embody all of a department's expectations, however. Certain standards may be known informally, but never appear in print. Officer-initiated activity, whether motor vehicle citations or stop-and-frisk checks, is unlikely to have formal declarations, in order to avoid complaints of bias. While "quotas" are avoided as a negative image, "performance expectations" will be unofficially encouraged and formally enforced, even if nebulously articulated.

In like manner, not all internal controls are paired to The Rulebook or to other formal elements of correction. Transfer to an undesirable shift or other type of assignment is a clear signal of organizational disfavor but without any rebuttable criticism that could be cause for further action. At more intimate unit levels, denial of requests for days off or special coverage serves a similar purpose.

Internal Affairs (IAD)

Most agencies of any size have some office like the old Internal Affairs Unit: police officers who investigate complaints against other officers. More likely to be called the Office of Professional Responsibility or similar title today, officers in these units investigate complaints originating from citizens and from within the department. Citizen complaints usually center on biased behavior, brutality, or complaints of criminal behavior such as theft of personal property. Because many of these complaints boil down to "he said/he said" situations, officers are often exonerated for lack of evidence. Some may trigger more extensive observation of the officer's behavior, particularly if a pattern of credible but not conclusive evidence emerges over multiple complaints. In such cases, the investigators are careful to discern between legitimate complaints from citizens and those which appear to be retaliatory complaints from criminals seeking to neutralize righteous police pressure on illegal activities.

Over the course of time, IAD investigations are more successful in validating internal complaints: dereliction of duty, calling in nonexistent stops to avoid unpleasant calls, abuse of sick leave, misuse of departmental equipment, and outright crime such as theft of drugs from the evidence room. IAD investigations are referred to the administrative chain of command for action: hearings are held at the command level for complaints that are substantiated; those which could lead to dismissal are usually at the chief's decision.

External Constraints

Departmental discipline is not a self-sufficient exercise: it is often bound by decisions made by other bodies. Disciplinary actions can be appealed to civil service boards or reviewed by outside arbitrators, depending upon the state, the municipal situation, and any applicable union rules. Even nominally "at will" employees have recourse for wrongful actions under some circumstances. The appeals process is intended to be a part of checks and balances against biased or arbitrary actions on the part of the agency.

There are a number of high-profile cases where employee "for cause" dismissals that were clearly proper have been reversed by politically distinct agents or entities, saddling the organizations with undesirable employees who seem to be impervious to the usual internal controls.

The civil courts provide the final resort for employees appealing adverse decisions. For agencies that have had to reemploy an undesirable worker, the options are relatively few: restrict the duties and thus the opportunities for harmful conduct ("hide the problem child in a closet") or return the person to regular duty under close scrutiny and begin the documentation process again ("give him enough rope to hang himself").

Depending upon the state, decertification may also be a route for dealing with problematic officers. In some states, only the employing agency can initiate a request for decertification for cause. In others, citizens or other entities may also initiate the process, within narrowly defined statutory criteria. There are distinctions between decertification for oversights, which are correctable, and decertification for cause, which effectively ends a person's ability to work as a police officer in that state. Recertification in another state remains possible, but the International Association of Directors of Law Enforcement Standards and Training (IADLEST) maintains a National Decertification Index (NDI) that tracks decertification cases in many (though not all) states.

Agency Accountability

By definition, accountability for agencies lies with external sovereigns. State legislatures and the Congress define the powers for police acting in their jurisdictions. Local, state, and federal appellate courts define and refine the specific applications of those powers. Municipal, county, and state political bodies control the budgets for the agencies, as well as the power of appointment of the chief in most cases.

Agencies are accountable for at least two levels: for the overall accomplishments of the department, usually but not universally measured by the crime rate, and for the manner in which police services are delivered. Even rising crime rates may be tolerated if the police response to the problems are considered to be positive: the influx of a new drug, dramatic shifts in local economics (both a downturn and a sudden boom), or other new circumstances may exceed a department's capacity for a short period of time. If a proper foundation has been laid, and communication with the public open and candid, a temporary inability to produce dramatic results need not be negative.

The second level is the activities and attitudes of the agency's employees as perceived by a significant constituency of the public. Where the agency has little choice but to back its officers in he said/he said complaints on an individual basis, citizens take a broader view of all interactions. Common to almost every complaint about police interaction with citizens (up to and including complaints of excessive force or brutality) is the phrase “no respect.” Complaints about officer attitudes are difficult to resolve, but the enduring effects of widespread perceptions of disrespect by officers can be toxic.

Political Control

Most law enforcement agencies are not autonomous units but rather are a department within a larger government organizational unit: a cabinet-level department in the case of many federal agencies, part of a justice department or working under the state’s attorney general’s office at the state level, and subject to the control of mayor, council, city manager, or other executive power at the local level. Where county police departments exist, they are answerable to the county commissioners or the county executive (see, e.g., Skogan and Frydl 2004, pp. 196–202).

Sheriffs are an exception to the general rule above. In almost all states, the county sheriff is an office authorized by the state constitution. As such, the sheriff is answerable only to the voters of the county he or she serves. In some cases, sheriffs have been voted back into office despite federal convictions for various offenses.

In Hawaii, the sheriff’s division is a part of the Department of Public Safety. Alaska has no counties, and thus no county sheriffs (Alaska State Troopers provide law enforcement services where no local police department exists). Connecticut eliminated the constitutional office of sheriff in 2000 in response to a series of corruption scandals. Because the office of county sheriff was written into the state constitution, a constitutional amendment was required. By a margin of almost two to one, voters in Connecticut eliminated the office of sheriff, transferring its duties to the existing Judicial Department and to a newly created office of marshal, who were to be overseen by a State Marshals Commission.

As a rule, when agencies are underperforming or malfunctioning in the eyes of their political constituents, the most common response is to replace the chief (by whatever title). The old “Chief for Life” convention is largely gone, replaced by a renewable contract for fixed terms, usually between 3 and 5 years (Mastrofski 2002). Since most chiefs serve at the pleasure of the administrative unit (whether an individual or a board), the contract preserves their property right, but not their tenure: a severance agreement must be reached if the chief is dismissed for political reasons, such as a change of mayor or board, though a dismissal for cause usually voids the contract.

It is a commonplace in law enforcement that if a chief leaves for another post or retirement when an agency is performing well, her replacement will come from within the department. If there is dissatisfaction with the overall performance of the agency, a replacement will be chosen from outside, with a mandate to institute organizational change.

Even if that perception is based on a generally reliable fact pattern, it overestimates the effectiveness of the chief executive. Like replacing the coach of a losing team, the malfunctioning players remain. If the problems of the organization stemmed from defective leadership, a new approach from the top position may very well turn performance around. If the root problem is an entrenched cultural one, change will be more difficult to achieve. Incompetence can be addressed by the infusion of new leadership and training resources; malevolent or self-serving cultures are more difficult.

In some instances, police agencies are simply disbanded, though that is much easier to do for smaller jurisdictions than larger cities. Scandal is often not the reason, though it has accounted for some instances. Budget difficulties are more often the reason, sometimes complicated by a desire to eliminate union rules or defined-benefit pension obligations, as is alleged in the 2012 elimination of the Camden (NJ) Police Department.

When such instances occur, law enforcement duties are transferred to another agency. Camden is incorporating with the county police; many small towns contract with their county’s sheriff’s office or rely upon State Police for the infrequent problems that arise. Contract cities are relatively common in the American Southwest, with Sheriff’s Departments providing a specific number of hours or standard police presence, in addition to emergency responses and criminal investigations as needed.

Accreditation

The agency-level articulation of professionalism is embodied in a set of national standards articulated by the Commission on Accreditation for Law Enforcement Agencies (CALEA). Using an accreditation model similar to that used in higher education, CALEA promulgates "gold standard" sets of agency requirements across a broad spectrum of issues, from rules and regulations to equipment, training, personnel standards, bona fide occupational standards for specialty positions, and others (CALEA n.d.). CALEA specifications are tailored to meet the different situations found in the broad range of law enforcement agencies, including size and available resource. After revising their operations and documenting the results, agencies apply for an independent evaluation of their adherence to these standards.

Accreditation is an attempt to formalize legitimacy of police operations and is often offered as a hypothetical defense against liability. Whether it actually serves that purposes remains an open question; few serious challenges have been mounted. The national accreditation process has also been criticized as time-consuming and expensive. The latter condition has led to the creation of state-level accrediting bodies (see, e.g., California CALEA n.d.).

Citizen Oversight

At the local level, dissatisfaction with the organization's internal controls may give rise to a call for citizen review of complaints. When this occurs, the cause is often not dissatisfaction with overall results but with a perceived lack of effective action taken against malfeasant officers (Skogan and Frydl 2004, pp. 202–204).

Citizen oversight – often called "civilian review" – is one of the responses to the belief that "the police protect their own" rather than serve the broader interest of the public. Different models of direct citizen oversight have been tried and some abandoned over time. One grants citizens an advisory role in the disciplinary process, but no direct input. A second limits citizen input to a general review of cases over time, again without direct input into individual cases. Yet another allows for citizen participation in the investigatory process, including the right to be present on case review boards, but without any real power: the results of the investigation are referred to the agency's internal process for adjudication. The most controversial form, and most desired by community groups, is a direct role in investigating complaints, with a power to recommend discipline. Even in this most stringent form, however, the final decision almost always rests with the police chief or other internal body.

Police fear that citizens cannot fully understand what police officers face in confrontational, potentially dangerous situations and thus are likely to be overly critical, with unrealistic expectations about police capacities. For their part, citizens consider themselves considerably more realistic than police give them credit for: not only do they deal with the same difficult individuals the police do, but they do so for longer periods of time and under a wider set of circumstances. Moreover, they are also capable of distinguishing between good police service and poor service.

The citizen review option is not a panacea. Selection of the oversight board's members can become a highly politicized process, with police advocates contending against police critics for the slots. Police themselves – fearing a "stacked deck" of longtime police critics, who are often the persons pushing the reform – may attempt to game the system by encouraging the candidacy of those who believe the police can do no wrong, ever.

In most cases, however, civilian review boards have faced the same dilemma that confronts police internal processes: the lack of definitive evidence of an officer's conduct, as well as the conduct of the accuser. That state of affairs has undergone a rapid and radical shift in recent years, however, with the proliferation of cell phone cameras and other visual media. The "low visibility environment" that once protected police misconduct in the field is now almost fully illuminated by citizen witness video. From

George Halliday's incomplete video of the Rodney King incident to the pepper spray incidents at U-Cal Davis and the G-20 protests in Toronto, individual video records have verified complaints of misconduct that in prior years would have been dismissed. Video is a neutral witness, however: dashboard cameras in police cruisers, GPS tracking systems, and even private security videos have exonerated police officers from false complaints of misconduct.

Consent Decrees

In a relatively small number of cases, the United States Department of Justice has stepped in where police agencies demonstrate a "pattern and practice" of abusive or derelict performance. The most famous case is perhaps the consent decree with the state of New Jersey over claims of racial profiling in the 1990s, complaints that were validated by the State Police records and training materials.

The intent of a consent decree is remedial more than it is punitive, though the restrictions placed upon the agency are often perceived to be punitive: the loss of autonomy is itself a punishment. Toll-free complaint lines, independent auditors, onerous reporting duties, and other requirements are generally part of a comprehensive plan to force a new manner of performing police duties, under scrutiny. The consent decree is also a signal to the political entities responsible for the agency that drastic change must be created internally. Many of the avenues of recourse described above may also come into play at the local level while the consent decree is in place.

Related Entries

▶ Biased Policing
▶ Causes of Wrongful Convictions
▶ Community Policing
▶ COMPSTAT
▶ Control of Police Misconduct
▶ Cross-National Performance in Policing
▶ Democratic Policing
▶ False Confessions and Police Interrogation
▶ History of Police Unions
▶ History of the Police Profession
▶ Law of Police Interrogation
▶ Law of Police Searches
▶ Law of Police Seizures and the Exercise of Discretion
▶ Law of Police Use of Force
▶ Measuring Police Performance
▶ Measuring Police Unit Performance
▶ Methodological Issues in Evaluating Police Performance
▶ Organizational Change and Police Legitimacy
▶ Police Corruption
▶ Police Culture
▶ Police Discretion and Its Control
▶ Police Discretion in Law Enforcement
▶ Police Discretion in Providing Services and Assistance
▶ Police Integrity
▶ Police Legitimacy and Police Encounters
▶ Police Lying and Deception
▶ Police Performance Measurement
▶ Police Selection
▶ Police Self-Legitimacy
▶ Police-Led Interventions to Enhance Police Legitimacy
▶ Policing the Police
▶ Procedural Justice, Legitimacy, and Policing
▶ Role and Function of the Police
▶ Role and Function of the Police Manager

Recommended Reading and References

Associated Press (2012) UC agrees to pay $1M to settle pepper-spray suit. New York Times, 26 Sept. http://www.nytimes.com/aponline/2012/09/26/us/ap-us-occupy-pepper-spray.html?ref=us. Accessed 26 Sept 2012

Bittner E (1970) The functions of police in modern society: a review of background factors, current practices, and possible role models. Crime and delinquency issues. National clearinghouse for mental health information. National Institute of Mental Health, Chevy Chase

CALEA (Commission on Accreditation for Law Enforcement Agencies) (n.d.) http://www.calea.org/

California CALEA (n.d.) http://www.chp.ca.gov/programs/calea.html

Crank JP, Caldero MA (2010) Police ethics: the corruption of noble cause, 3rd edn. Anderson, Cincinnati
Gardner v. Broderick 392 U.S. 298 (1968)
Garrity v. New Jersey 385 U.S. 493 (1967)
Harrison KA, Flaherty MP (1994) D.C. Police paying for hiring binge. The Washington Post, 28 Aug. http://www.washingtonpost.com/wp-srv/local/longterm/library/dc/dcpolice/94series/trainingday1.htm. Accessed 30 Sept 2012
http://hawaii.gov/psd/law-enforcement/sheriffs-division
http://www.ballotpedia.org/wiki/index.php/Connecticut_Eliminate_County_Sheriffs,_Question_1_(2000)
Klockars CB (1988) The rhetoric of community policing. In: Greene JR, Mastrofski SD (eds) Community policing: rhetoric or reality. Praeger, New York, pp 239–258
Mastrofski SD (2002) The romance of police leadership. In: Waring E, Weisburd D, Sherman LW (eds) Theoretical advances in criminology: crime & social organization. Transaction Publishers, New Brunswick, pp 153–196
Mastrofski SD (2003) Personnel and agency performance appraisal. In: Geller WA (ed) Local government police management, 2nd edn. International City Management Association, Washington, DC, pp 447–486
Skogan W, Frydl K (eds) (2004) Fairness and effectiveness in policing: the evidence. National Academies Press, Washington, DC
Stinson PM Sr, Liederbach J, Freiburger TL (2010) Exit strategy: an exploration of late-stage police crime. Police Q 13(4):413–435
Walker S (2001) Police accountability: the role of citizen oversight, Wadsworth professionalism in policing series. Wadsworth, Belmont
Walker S, Macdonald M (2009) An alternative remedy for police misconduct: a model state pattern or practice statute. George Mason Univ Civil Rights Law J 19(2008–2009):479–552

Police Lying and Deception

Jennifer C. Hunt
Montclair State University, Montclair, NJ, USA

Overview

This entry explores some police lying and deception that arises in different social contexts including among patrol officers, in interrogations by police detectives, during undercover operations, and by superiors of different rank. Some scholars have utilized epistemological approaches to the study of lying that assume that an objective reality exists that is independent of interpretation and retrospective construction. Such works assume that most lying that violates law and procedure is deviant or malicious and unjustifiable and that police administrators are not aware of the endemic nature of lying and want reform. This entry touches on the works of these scholars but highlights the perspectives of police officers who view the police organization in a different way. Police officers believe that cops and detectives lie, that supervisors expect them to lie and lie themselves, and that managers use deception to manipulate the press and public.

Introduction

"If a perp 'falls' in a stationhouse where there are no civilians around to see it happen, did it happen at all?" This statement was made by NYPD police officer during a standup comedy routine delivered at a New York club. It revised a classic philosophical question to comment on the secrecy that surrounds police activity and the lies that accompany certain violations of law, in this case the beating of a suspect. This entry addresses the subject of police deception and lying, the exploration of which circles back to the question the NYPD officer raises. It may be impossible to separate fact from interpretation and narrative construction in the police organization because it is grounded in deception from the bottom to the top of the chain of command.

Police lying has been subject to special scrutiny in the scholarly literature because of the social and legal contexts in which it takes place, the power police have in determining suspects' fates, and the public assumption that police should be honest and trustworthy (Alpert and Noble 2009; Leo 2008). This entry begins by exploring different types of lying with reference to the scholarly literature, most of which takes an outsider's perspective that there exists a world of fact that is separate from interpretation and retrospective narrative reconstruction. It goes on to examine how lying is understood and constructed from the perspective of rank and file officers who work inside the police world. The entry concludes with a summary of the areas that need to be addressed in future research in order to develop an understanding of lying within the police organization as a whole.

Lying is defined as an act of symbolic communication that is known to be false and is meant to deceive its audience (Hunt and Manning 1991, p. 52; Manning 1974). Lying is a feature of everyday life in a variety of subcultures and organizational and occupational environments besides the police (Aquino and Becker 2005; Pershing 2003; Manning 1974). It can involve a simple omission of facts; an odd arrangement of adjective, subject, and verb; or the formulation of completely fictitious narrative. While formally lying may be depicted as wrong and shameful, informally it is recognized as a tool of survival in a world in which truth can result in punishment, loss of peer support, or accusations of insensitivity, social ineptitude, or mental illness. Meltzer (2003, p. 1) goes so far as to suggest that "Human life is thus only a perpetual illusion.... [and] Human society is founded on mutual deceit."

In view of the illusion that people should be truthful and organizations operate according to an established and mostly visible set of rules, deception may evoke moral dilemmas for participants, depending on the social context in which it is used. As a result, individuals who routinely lie in the course of their work may experience emotional distress (Aquino and Becker 2005). It thus becomes necessary to develop accounts that excuse and justify acts of lying, neutralize guilt, minimize stress, and maintain participants' sense of moral worth without the introduction of paralyzing moral qualms that could interfere with work (Aquino and Becker 2005; Hunt and Manning 1991; Scott and Lyman 1968; Sykes and Matza 1957).

Police Lying and Ethnographic Fieldwork

Hunt (1984; Hunt and Manning 2000) was first introduced to police lying when she was doing a year of fieldwork in mostly high crime precincts in the New York City Police Department in the late 1970s. She recalls accompanying two veteran officers into the large apartment of an elderly couple who were living in a neighborhood in which crime was becoming a serious problem. The couple explained that some young men had thrown a rock at their window and nearly hit one of them. They pointed to two rocks the size of a man's fist and a lake of shattered glass on the floor. After the cops took a report, they began to give the husband and wife "a hand job," a term police use to describe a lie that is intended to reassure in situations in which nothing can really be done. The officers promised to watch the building. They then applied scotch tape to one of the rocks in order to "take fingerprints" that they claimed would help detectives investigate the case. Desk officers have been known to "marry" and "divorce" people in busy precincts in the Bronx, Brooklyn, and Harlem.

Hunt's understanding of police lying in the NYPD helped her develop rapport with police in the Philadelphia P.D. when she went there to do 18 months of participant observation (Hunt 1984). The department was under investigation for corruption and brutality, including the homicide division whose members had been accused of using torture during interrogations. Initially, police officers believed she was a spy for the Justice Department or Internal Affairs. During her first tour, she rode with an officer who insisted that she was wearing a "wire" (recording devise) and refused to talk to her for the first four of their 8-h shift. In Hunt's second day on patrol, her veteran partner took her into a restaurant while his peers picked up his jobs, an arrangement she believed they made in advance. She and the officer sat in the diner a long time without alerting radio communications regarding their whereabouts. Hunt proceeded to converse with her police partner and make up cover stories to account for their absence from the street. In doing so, she revealed that she understood informal norms surrounding self-protective lying. When they returned to the precinct at the end of 4 p.m. to midnight shift, his colleagues asked him if Hunt had learned anything. "I think she knows it already," her partner replied (Hunt 1984). Subsequent tours went smoothly as she continued to try and negotiate trust by showing, in small symbolic gestures, that she understood and would follow cops' rules, including those that surrounded deception. An NYPD officer told a joke that illustrated this basic principle of lying: "How many cops does it take to screw in a light bulb? Four – one to screw in the bulb and the other three to say: "I didn't see nothing!"

Hunt quickly learned that "he fell down the stairs" was the most common explanation that followed the beating of suspects who emerged, bruised and swollen, from the backroom of a police facility. Lies like this were not simply addressed to her as a civilian but also to cops who did not witness events. It was commonly understood that knowledge could be a burden in the event that an investigation took place. By then police trusted that she knew how to read between the lines of the explanations that typically accompany legally problematic events. "I don't know nothing" became her standard reply when she was questioned during internal inquiries into suspicious incidents. These sorts of responses led to nods of approval, more exposure to violence than she would have liked, and an invitation from homicide detectives to accompany them on jobs. "You'll have to ride with us sometimes," one said in the wake of an investigation into why a suspect had sustained burns on his legs while in police custody (Hunt 1984).

Police Interrogations

Police lie in a variety of social contexts. The use of torture during interrogations has largely been replaced by methods that claim to have scientific support in psychology (Leo 2008). Detectives may attempt to manipulate suspects by claiming they have evidence they don't have, displaying evidence that does not belong to the case being discussed, and playing good and bad cop roles that allow one officer to break the suspect down and the other to build rapport. In "Fluffing up the Evidence and Covering Your Ass: Some Conceptual Notes on Police Lying," Barker and Carter (1990) organize police deception into a three fold taxonomy that includes accepted, tolerated, and deviant lying. Deception by interrogators is categorized as accepted lies because they adhere to procedure and law and fulfill "reasonable" organizational goals. Indeed, detectives receive formal training in deceptive interrogation techniques.

Alpert and Noble (2009) construct a continuum of deception that ranges from lies that are acceptable to those that are "malicious" and "unjustifiable." From their perspective, lies by police interrogators fall within the acceptable range. Police officers generally concur with this view. In their eyes, such techniques are not only necessary but can be rationalized because most suspects who are subject to even prolonged psychological interrogation techniques are guilty or have a record of guilt that justifies harsh treatment. Furthermore, detectives would argue, psychological methods do not result in physical harm, and thus injury can be denied.

Leo (2008) is critical of the use of deception on the part of police interrogators. He notes that detectives falsely believe that they can assess when a suspect is untruthful by certain behaviors, such as the movement of eyes. They also believe that sorting mistakes are rarely made and that innocent people cannot be induced to confess unless they are subjected to "the third degree" or are mentally ill. The use of deception in interrogations is harmful to our adversarial system of justice because it violates the right against self-incrimination of mostly poor and minority citizens and sometimes forces innocent people to confess. Alpert and Noble (2009) acknowledge the problem of false confessions. Feeling trapped by a flood of information that police insist prove guilt and convinced that a false admission might result in a sentence reduction, suspects can be induced to confess to acts they did not do. It is notable that in some countries, including Germany, there are stricter rules surrounding the use of deception in police interrogations and result in suspects making statements that conflict with their self-interest (Ross 2008).

Undercovers

Both Alpert and Noble (2009) and Barker and Carter (1990) view the acts of deception and trickery that accompany undercover work as accepted and/or justifiable within their taxonomy or continuum of police lying. Officers who work undercover must manage impressions in such a way that they appear to be someone else. In large departments, some undercovers may work several cases at once and have to keep their lies consistent when they work or when "clientele" call to talk to them on their cell phones. Undercover officers typically take pride in their acting skills, enjoy the expertise the work involves, and appear to have little problem neutralizing the use of deception. After all, they would argue, the job results in the arrest of serious felons including gun or drug dealers and does not focus on the single buyers who are at the bottom of the "food chain."

While legal, deception in undercover operations may have unintended consequences, some of which police do not recognize. Gary Marx (1989) suggests that officers who use deception in sting operations sometimes border on entrapment and/or encourage the commission of crimes that otherwise would not occur, criminalizing those who might not have otherwise violated the law. Police undercover work is also physically and psychologically dangerous for the officers involved. Police organizations are more cognizant now than in the past of the psychological hazards of deep cover operations that operate for lengthy periods of time. Today there is frequent contact between undercovers and their peers and bosses, resulting in the reinforcement of police identity and bolstering resistance to "going native" in the criminal world. Still officers who work in deep cover narcotics operations may be put in positions in which they feel compelled to take drugs or blow their cover. Undercovers who are assumed to be criminals may be killed by "other" criminals. Friendly fire incidents have been known to result in injuries of officers working in plainclothes operations in units like Anticrime. Several years ago, two undercovers were murdered during a gun buy. Their colleagues did not believe they had been "made" as cops although the district attorney wanted their commanding officer to testify that he believed the officers had been made. This case suggests that lying by police is sometimes encouraged by other players in the criminal justice system and may be systemic in the larger organizations with which the police interact. An additional negative side effect of chronic lying by undercover police is that the practice can contaminate other aspects of their lives, including family relationships that are based on mutual trust.

Lying Among the Ranks of Uniformed Patrol

Patrol officers lie routinely during the course of their daily work. Sometimes they lie to make people feel good in situations in which there is little else they can do. The fingerprint story provided an example. Carl Klockers (1984) calls this type of lying "police placebos" and notes that placebos are created for the benefit of members of the public, not the officers involved. He mentions death notification as an example. When police inform family members that a relative has died in an accident or as a result of crime, they may say that the death was instant and caused no pain regardless of the circumstances involved. Most "placebos" would be categorized as acceptable or tolerated according to the scholars' taxonomy or continuum (Barker and Carter 1990; Alpert and Noble 2009).

Patrol officers also use deceptive ruses in situations in which they are suspicious of peoples' claims and want to put them to the test. In this case, the lie is not intended to make the person feel good but to increase discomfort so he or she reveals information that will allow officers to better assess "the facts." One New York City patrol officer recalled that, in the late 1990s, the precinct was getting a rash of reported robberies by European tourists. Suspicion was aroused when a man from Spain walked into the precinct and told the officer that he had been robbed near Columbus Circle in Manhattan and needed a police report for insurance. During the interview, the victim claimed that the "perp" had ordered him to "Stick'em up!" Stifling his laughter and assuming there was a language problem involved, the officer asked the victim if those were the suspect's exact and literal words. The victim said they were. The officer took the victim back to the scene of the crime and then pretended to ask "Central" (central communications – or police radio) to download the "nonexistent" videotapes from the "nonexistent" cameras affixed to the buildings within scanning view of the location. The officer explained what happened after that:

> The victim starts to freak and back out of his story, claiming he had to get to the airport. When we started pressing him, he admitted the story was BS and he was subsequently arrested for filing a false report. The fake video cameras became SOP (standard operating procedure) in (the area covered by the precinct) for a while and several people were arrested. Word must have spread because when I (transferred to a different unit), the fake reporting seemed to have died down.

Cops also use deception in lieu of physical force to punish "assholes" (Van Maanen 1978) who challenge their authority. This sort of lying would probably not be included in the realm of acceptable or justifiable lying in the schemas of most police scholars but is viewed as reasonable and necessary by cops because it helps restore the proper imbalance of power – that cops have more – and also evens the score. A video recording of a one-man theatrical show by a retired NYPD officer provides an example. In the tape, two police officers are redirecting traffic in Central Park in Manhattan because a rainstorm has led to significant flooding in a wide and deep pocket of the road that directly connects the west and east sides. When one of the officers politely attempts to redirect the driver of a car with diplomatic plates, the driver becomes argumentative and insists on taking the direct route. The officer backs down and tells him to go ahead, while informing his partner that he does not want to cause an international incident. The diplomat soon finds himself alone and immobilized in a deep river of dirty water that covers a large section of the street.

Patrol officers as well as those from other units routinely lie to supervisors to create narratives that benefit their personal goals, such as extending a sick day from one to three or whatever the maximum is before a visit to a police surgeon is required. While scholars might view such lies as unacceptable because they involve "stealing departmental time," cops believe that bosses expect such lies to arise. After all, bosses were once cops who know how the game is played. Such acts are rationalized with reference to "common sense" and the scarcity of rewards. Veteran cops think it is stupid not to take advantage of available opportunities unless a respected boss needs them for an unusual event.

Police also believe that bosses "don't want to know" about much of what they do behind superiors' backs. One detective in a first responder unit explained that he and his teammates had had a party during an all-night tour of duty inside. When the lieutenant comes to work in the morning and discovered that all of them were drunk, he angrily ordered them home. In response to the order, one of the officers explained that he could not go home because he was "too drunk to drive." Furious, the lieutenant went to the captain and told him what his officers had done. The captain asked the lieutenant if he planned to do anything. When he said no, the captain responded: "Then why are you telling me? There are some things I don't have to know" (Hunt 2010a, p. 337). A patrol cop provided an example of a lie he told his sergeant that he believed the sergeant would have approved. Indeed, the officer justified the lie as necessary to prevent his boss from involvement in a troublesome incident that might result in accusations that he had not properly done his job.

> One time I sprained my ankle while leaving my girlfriend's apartment. Problem was, I'm on duty but way off post. I'd gone down to court to process an arrest and her place was on the way back to the stationhouse. Now, there's no way I'm telling my squad sergeant where this happened because it's going to open a whole can of worms and drag him into a mess, failure to supervise and shit like that. So I sprained my ankle down the block from the precinct. What's the difference anyway? My ankle is still sprained. Right? It's not like I'm dragging myself to work after spraining it playing softball! Now that would be fucked up!

While Barker and Carter's (1990) taxonomy of accepted, tolerated, and deviant lying is useful in some respect, it takes an outsider perspective that does not illuminate how rank and file officers construct their world of deception in contrast to high-ranking members of the police organization. Alpert and Noble (2009) appear to assume the perspective of administrators who recognize that lying can have negative consequences and some forms should be subject to punishment, including the dismissal of the officers involved. Cops would argue that such administrators only surface when lying by rank and file officers and detectives becomes a source of embarrassment for the department.

Solutions to Police Lying and the Reproduction of Deception

Alpert and Noble (2009) advocate a number of solutions to the problems they see in acceptable lying that does not violate law and procedure and malicious lying that does. These include the incorporation of materials on ethics and the consequences of deceptive interrogation processes in police training. If liability issues did not interfere, it would also be helpful to include materials on what street cops do and how they neutralize deceptive practices as this instruction might help undermine recruits' ability to effectively use the relevant accounts while preparing them for what is to come. In order to avoid problems that could ensue if former recruits got into trouble and claimed they learned to lie in the police academy while reading and discussing papers on what police really do, instructors typically isolate recruits from veteran colleagues while indoctrinating students about the formal rules as if policing is done by the books (see Moskos 2008; Hunt 1985). In a way, academy training thus creates the deception that "the job is on the level," a reality that diminishes with time on the street. This phrase refers to the belief of naive police or non-police persons that the police organization is what it claims to be and deception is not endemic.

Street Cops Perspectives

Police categorize as normal particular types of lying that violate law and procedure. Such culturally sanctioned lies are also viewed as troublesome because they could subject officers to legal scrutiny and punishment. The classic case of a troublesome lie that went amok occurred in the O. J. Simpson case when detective Mark Fuhrman claimed that he had not used the word "nigger" in at least 10 years. When proof was introduced that Fuhrman had used the word "nigger" more recently, including during taped interviews with a writer, working on a screenplay, it discredited the rest of his testimony and bolstered the defense's suggestion that he had planted a bloody glove on Simpson's property (Slobogin 1996). Although Fuhrman did not go to jail for perjury, his actions undermined the prosecution's case and embarrassed the department.

Once police leave the formal world of the academy, they learn to distinguish normal lies from those that are culturally unacceptable in their world. They also learn how to construct normal lies and account for their use in ways that are shared with veteran peers. The ability to skillfully lie in appropriate social contexts helps define the boundaries between police and "the other" that does not understand or belong in their world. From the cynical perspective of street cops, "the other" includes scholars who believe police administrators' claim that they, themselves, do not engage in deception, do not want cops to tell lies that violate law and procedure, and desire serious reform that extends beyond removing the few rotten applies who float to the surface and embarrass their department. The phrase "the job is not on the level" refers to street cops' belief that the police organization is built on deception and informed by politics and that those at the top of the chain of command are complicit in maintaining the status quo.

Case Lies

One type of troublesome lies that are recognized by patrol officers is case lies that are presented in court testimony or on paper to facilitate the conviction of suspects (Hunt and Manning 1993). Cops routinely lie in court to invent probable cause when it is weak or absent if they are fully convinced that the suspect is guilty. Typical cases involve minor issues that do not result in serious punishment to defendants, such as disorderly conduct by juveniles that result in complaints. Cops also lie in the construction of case stories by creating narratives of events that manipulate the perceived sentiments of judge or jury. Among the accounts cops use to neutralize internal conflict in the wake of lying in court are claims that court is “a game” in which each player has to top the other’s story (everyone lies), the suspect is guilty and deserves what he gets (denial of the victim), the punishment is minor (denial of injury), and “I made the arrest at the request of superiors who told us to clear the corner” (denial of responsibility) (Hunt and Manning 1993).

Cover Stories

Cover stories are the second type of lies that police categorize as normal and acceptable even when they arise in a “troublesome” context in which officers could be exposed. Police use cover stories to avoid revelations that could result in punishment in situations in which they have violated departmental procedure and/or law. Cover stories routinely arise in innocuous contexts such as when patrol officers miss a radio call because they are running a personal errand or when the sergeant needs an explanation for why subordinates were not on the street when he or she came by to sign their logbook. Police lie to supervisors to create narratives for personal benefit such as extending a sick day.

In one case, a supervisor and captain got together to create a coherent narrative about an accidental discharge of weapons that did not result in injury. Neither one of the bosses believed the cop’s account of the event. The sergeant tried but failed to persuade the cop to alter his statement to one that made sense to him. To avoid questions from superiors, the sergeant and captain got together and wrote a report that revised the officer’s oral statement and created a “believable” account that would minimize additional inquiry.

The *NYPD Patrol Guide* and similar documents in other departments contain an abundance of formal procedures, some of which make the job hard to do expediently. The result is that even the best cops violate one or another rule at various times during the course of their work. This means that supervisors who view particular cops as troublesome can always “get something on them” if they want to do them harm. For this reason, officers believe that cover stories are necessary to avoid persecution by political enemies within the department. Officers who have good relationships with supervisors are convinced that bosses expect them to violate some of the rules and to prepare cover stories in the rare event that an inquiry comes to pass.

Officers routinely get together to create cover stories in instances in which suspects are beaten or physically abused. Typically such occasions include such trumped up charges as resisting arrest and assault on an officer. Police also sometimes invent cover stories in the wake of ambiguous use of lethal force incidents. Hunt and Manning (1993) provide examples of how officers got together to construct an imaginative cover story that fit observable facts in what they recognized was a “bad shooting” of a child. Police define a bad shooting as one that is not culturally sanctioned or legally justified. Waegel (1984) provides an example in which an officer claimed that his “bad shooting” was the result of an “accidental discharge.” It is interesting that the military uses the term “negligent” rather than “accidental” discharge. In most cases, the only way an officer can “accidentally” fire his gun is if he has his finger on the trigger when he does not have a reason to shoot. This practice violates weapons’ training and is dangerous to civilians and cops.

The frequency of lying by police in instances in which lethal force is used is confirmed by Paul Chevigny (1995) who notes that police often lie in shooting incidents when lying is not necessary to make their case or protect themselves. Indeed, such lies can result in inconsistencies in police testimony and accusations of perjury. Hunt (2010a) suggests that it is not surprising that police officers frequently get together to make up cover stories in the wake of shootings in view of the expectation that their memories should be clear and intact. Studies have shown that stress can affect memory and there is little correspondence between “accuracy and confidence of recollection” in the reporting of critical incidents (Hunt 2010a, pp. 327–327; Artwohl and Christensen 1997; Alpert et al. 2011).

Some scholars have labeled cover stories as deviant or malicious and unjustifiable (Barker and Carter 1990; Alpert and Noble 2009). Police officers have a different point of view. Police only view cover stories as "bad" when they are not composed with the skill that is required to have them pass as truth. Cover stories are also seen as only as "moral" or acceptable as the incident that they are trying to protect. If the act underlying the lie is viewed as corrupt or brutal and cannot be neutralized with reference to culturally sanctioned accounting practices, then the cover story is also stigmatized. Prohibitions against ratting out peers that ordinarily maintain cover stories intact may be neutralized in instances in which the underlying act is heavily frowned upon.

This was the case when police officer Justin Volpe sodomized Haitian immigrant, Abner Louima, in the bathroom of the "7-0" (70th) precinct in Brooklyn, New York, in 1997. The majority of officers in the NYPD did not believe Volpe had committed the crime until he confessed at trial because they could not face the fact that a fellow cop could commit such a heinous act (Hunt 2010a, pp. 162–166). Instead, even the most cynical cops preferred to believe the absurd story propagated by the defense that Louima hurt himself while participating in rough but consensual, homosexual sex. The irony here is that even cops, who routinely lie and believe that they can assess when others are telling the truth, were convinced that Louima's injuries were self-inflicted. Officers had no doubt that Louima was beaten in the patrol car on the way to the precinct if Volpe (mistakenly) believed Louima "sucker punched" him while they were standing in front of the club. However, most police could not believe that Volpe had perpetuated a sadistic, homosexual assault. Such an act did not fit cultural notions of how far even corrupt and brutal cops will go.

A number of the 7–0 precinct officers who saw Louima before or after the assault came forward and confessed what they had seen. At first glance their actions appear to have been a violation of the "blue code" in which police officers unite to protect themselves (Manning 1978). However, other police did not shun those who testified for the prosecution. From their perspective Volpe's crimes neutralized the code. Not only did he commit a horrible act but he also made cops ashamed to be cops (Hunt 2010a, pp. 162–166). It is notable that the statement "he's good with his hands" is not a compliment. Most officers do not want to work with partners who resort to physical force at the drop of a hat (see Hunt 1985).

Deception Up the Chain of Command

Despite the relative abundance of works on police lying compared to deception in other occupations, the current scholarly literature tends to focus on deception that occurs at the level of rank and file officers in uniformed patrol and in the detective division among interrogators. Police scholars therefore have a limited understanding of the nature of lying up the chain of command and the degree of collusion that exists at the level of middle and upper management. As the phrase "the job is not on the level" suggests, police view deception as endemic to the organization. It is difficult to assess the nature of managerial involvement in deception that violates law and procedure. An examination of official reports suggests that there is some degree of collusion in terms of issues that relate to high-ranking officers' desire to protect the image of the organization and demonstrate loyalty to the police commissioner or chief.

Official reports are typically sanitized accounts of complex events that strive to eliminate the confusion and ambiguity that sometimes accompany decisions made on the street. Such accounts tend to draw a neat line between officers' actions and their justification in law and departmental procedure, regardless of the confusion that greeted the real events (Van Maanen 1980; Hunt 2010a). Collusion by members of top brass in the creation of sanitized police incident reports may be more common than we know. In one recent case in a large urban police department, the attendance and experience of a particular group of police officers were left out of an official account of a police shooting in order to avoid liability in the event that the victim's relatives filed a lawsuit. While the shooting was technically legal, it could have been avoided according to the group who were left out of the report. Top brass knew the group had been at the scene and had a different take about what went on.

They also knew that the group was excluded from inclusion in the final report.

There is a growing body of evidence that suggests the collusion of high-ranking members of the NYPD in the falsification of crime reports through encouraging the underreporting or downgrading of crime (Silverman and Eterno 2012; Hunt 2010b, c; Rayman 2010). It appears that pressure exerted from the top of the chain of command is partly responsible for police officers and supervisors' manipulations of crime statistics. Recent reports in the *New York Post* indicate that nine police precincts are showing an "alarming spike in crime" (Bennett et al. 2011).

The NYPD recently removed Deputy Inspector Jose Navarro from his position as commanding officer of the three to four precinct because he could not "control" the rise in crime. The *Post* article also states: "NYPD Chief Joseph Esposito and Deputy Commissioner Patrick Timlin have launched a get-tough policy with precinct commanders to demand they get crime under control" (Bennett et al. 2011). There are limitations in how much police can control crime through altering deployment strategies in view of the limited manpower available. Apparently, management does not want to acknowledge that punishing precinct commanders for problems they cannot always control has the same result as encouraging the manipulation of crime statistics to create the appearance that crime is not on the rise. One NYPD police officer recalled a conversation in which his commanding officer (a captain or deputy inspector) pressured him to change a "61" (name of the type of report) from a burglary to a larceny "because a building in a park can't really be considered permanent, right?"

There appears to be less transparency in the NYPD today under Mayor Michael Bloomberg and Police Commissioner Raymond Kelly than in past few administrations. The existence of police operations that appear to violate the constitution or other legal standards is routinely denied by Paul Browne, the NYPD's Deputy Commissioner, Public Information. Prior to the publication of a documented article about the NYPD Demographics Unit by the Associated Press, Browne denied that the Intelligence Division contains the seventeen-member unit (Appuzo and Goldman 2011; Demographic Unit, undated). The unit is composed of undercover officers ("rakers") who are tasked with spying on members of Muslim communities in the tri-state area. Rakers "visited Islamic bookstores and cafes, businesses and clubs [that attracted relevant minorities]....moniter current events, keep an eye on community bulletin boards inside the houses of workship and look for 'hot spots' of trouble. ... They played cricket and eavesdropped in the city's ethnic cafes and clubs....when the CIA would launch drone attacks on Pakistan, the NYPD would dispatch rakers to Pakistani neighborhoods in New York, New Jersey and Connecticut...." (Apuzzo and Goldman 2011). After the article was published, Browne changed his statement but claimed that "We do not employ undercovers...unless there is information indicating the possibility of unlawful activity." The department has also denied CIA collusion in the activity of the NYPD through a former member of the Agency is on the department payroll (Apuzzo and Goldman 2011). Browne did not explain what legal grounds justified the actions described in the AP article or in other newspapers describing the NYPD's efforts to spy on Muslim college students.

Conclusion

A number of excellent studies exist that explore police lying and deception during their daily activities, including the interrogation of suspects; patrol jobs involving hard-to-handle human problems; undercover detective work including sting operations, in court testimony; and to avoid punishment in situations in which officers make mistakes, abuse their authority or knowingly participate in acts of corruption that peers would not support. However, scholars know little about deception and lying by police officers at the top of the chain of command. One arena in which such lies appear to arise involves the construction of official reports of critical incidents that might make the department look bad in the eyes of the public or subject the department to lawsuits. Another involves implicit or explicit encouragement of the falsification of crime reports. A third area that would be worth pursuing is the manipulation of the media by police entities that manage the public and the press, such as the Office of the Deputy Commissioner, Public Information (DCPI) in the NYPD. The hidden activities of the secret undercover unit in the NYPD Intelligence Division remain just one of several units that should be of particular concern for researchers in terms of issues of deception, transparency, and the trampling of the US constitution. The activities of such units are authorized by high-ranking members of the service and subject to lies and distortions by commanders including those in the Office of the Deputy Commissioner, Public Information. Such lies are aimed at deluding the public into thinking that the police are only concerned with protecting them against terrorism and do not engage in illegal activities or others that waste financial resources and deplete police precincts of officers dedicated to serving the community and managing crime.

In terms of the issue of lying, the separation between management and street cop views is not as distinct as some would have us think (Hunt 2010a, pp. 305–308). Managers know that cops lie in a variety of legal and not so legal social contexts because they generally did the same when they were cops. Rank and file officers and supervisors believe that "there are some things" bosses don't want to know. Cops thus believe that supervisors and police executives, including top brass, expect them to lie or keep certain types of information to themselves. Only when something arises that threatens to spiral out of control and into the arms of the press do superiors want cops to tell them the truth of what is going on so they can put out the fire before careers are undermined and the administration goes up in flames. Then managers will scramble the facts to create a version of events that shifts the blame to the lowest ranking participants and leave the organization intact. If what cops believe is true is, in fact, true, the police department is an organization that is sustained on lies and deception but strives to maintain the illusory image that managers encourage the opposite. Alternatively, high-ranking superiors have come to believe that the illusion of truth that they began to perpetuate as they moved up the ranks actually reflects the reality of an organization that is founded on deceit (see Meltzer 2003).

Related Entries

▶ COMPSTAT
▶ Control of Police Misconduct
▶ Police and the Excessive Use of Force
▶ Police Corruption
▶ Police Culture
▶ Police Integrity
▶ Police Self-Legitimacy
▶ Policing the Police

Recommended Reading and References

"The Demographics Unit." (NYPD DU) Internal NYPD Document, undated

Alpert GP, Noble JJ (2009) Lies, true lies, and conscious deception: police officers and the truth. Police Q 12:237–254

Alpert G, Rivera J, Lott L (2011) Working towards the truth in officer-involved shootings memory, stress and time. FBI Law Enforc Bull (forthcoming)

Apuzzo M, Goldman A (2011) NYPD monitored where Muslims ate, shopped and prayed. Associated Press, 8/31

Aquino K, Becker TE (2005) Lying in negotiations: how individual and situational factors influence the use of neutralization strategies. J Organ Behav 26:661–679

Artwohl A, Christensen LW (1997) Deadly force encounters: what cops need to know to mentally and physically prepare for and survive a gunfight. Paladin Press, Boulder

Barker T, Carter D (1990) Fluffing up the evidence and covering your ass: some conceptual notes on police lying. Deviant Behav 11:61073

Barnes JA (1994) A pack of lies: towards a sociology of lying. Cambridge University Press, Cambridge

Beehr TA, Ivanitskaya L, Glaser K, Erofeev D, Canali K (2004) Working in a violent environment: the accuracy of police officers' reports about shooting incidents. J Occup Organ Psychol 77:217–238

Bennet C, Simeone J, Mixson C, New York Post. "Dangerous 'hoods." 22 Aug, 2011. http://www.nypost.com/p/news/local/dangerous_hoods_gaOd77vgw3QC7xRrunsnDM

Chevigny P (1995) Edge of the knife: police violence in the Americas. New Press, New York

Hunt JC (1984) The development of rapport through the negotiation of gender in fieldwork among police. Hum Organ 43:283–296

Hunt JC (1985) Police accounts of normal force. Urban Life 13:315–341

Hunt JC (2010a) Seven shots: an NYPD raid on a terrorist cell and its aftermath. University of Chicago Press, Chicago

Hunt JC (2010b) Schoolcraft tapes and the current investigation in the NYPD. Archive for Sep 2010 http://www.seven-shots.com/2010/10/

Hunt JC (2010c) A dip in crime or the height of spin. Archive for Nov 2010 http://www.seven-shots.com/2010/11/

Hunt JC (2011) The CIA and Detainee Abuse, 1 July. http://www.seven-shots.com/blog/

Hunt JC, Manning PK (1991) The social context of police lying. Symb Interact 14:51–70

Van Maanen J (1978) The asshole. In: Manning PK, Van Maanen J (eds) Policing a view from the street. Goodyear, Santa Monica

Klockers CB (1984) Blue lies and police placebos. In: Klockers C (ed) American Behavioral Scientist 27:411–544

Leo RA (2008) Police interrogation and American justice. Harvard University Press, Cambridge, MA

Manning PK (1974) Police lying. J Contemp Ethnogr 3:283–294

Manning PK (1977) Police work. MIT Press, Cambridge, MA

Manning PK (1978) Lying, secrecy and social control. In: Manning PK, Van Maanen J (eds) Policing: a view from the street. Goodyear, Santa Monica

Marx GT (1989) Undercover: police surveillance in American. University of California Press, Berkeley

Pershing JL (2003) To snitch or not to snitch? Applying the concept of neutralization techniques to the enforcement of occupational misconduct. Sociol Perspect 46:149–178

Rayman G (2010) The NYPD tapes: inside Bed-Stuy's 81st precinct. Tuesday, 3 May, *Village Voice*. http://www.villagevoice.com/2010-05-04/news/the-nypd-tapes-inside-bed-stuy-

Ross J (2008) Do rules of evidence apply (only) in the courtroom? Deceptive interrogation in the United States and Germany

Secret NYPD Demographic Unit Revealed,, 01 Sep 2011 http://rt.com/usa/news/nypd-demographic-ap-muslm-571/

Silverman E, Eterno J (2011) The crime numbers game: management by manipulation. CRC Press/Taylor and Francis Group, London, p 2011

Slobogin C (1996) Testilying: police perjury and what to do about it. Univ Colo Law Rev 1037:1–16

Sykes GM, Matza D (1957) Techniques of neutralization: a theory of delinquency. Am Sociol Rev 22:664–670

Waegel WB (1984) How police justify the use of deadly force. Soc Probl 32:144–155

Police Officers in Schools

Chongmin Na[1] and Denise C. Gottfredson[2]
[1]School of Human Science and Humanities, University of Houston-Clear Lake, Houston, TX, USA
[2]Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA

Synonyms

School resource officers (SROs)

Overview

Criminal offending, victimization, and disciplinary problems at school have been a great concern in the USA, and millions of dollars are being spent on improving school safety. Unsafe school conditions can have detrimental effects on students' academic and behavioral outcomes either directly or indirectly by influencing the nature of social relationships within schools and school climate. Not surprisingly, public concerns about school crime and violence have led to a variety of responses to increase the safety of schools, which include the use of security and surveillance practices. One of the strategies gaining popularity to enhance school safety is having uniformed police officers stationed on the school grounds. In the past 12 years, the rising concerns about high-profile school violence and shootings led to a dramatic increase in the use of school resource officers (SROs), and the number of schools with full-time SROs still continues to grow rapidly, especially in large, urban, and secondary schools.

While most key stakeholders (especially school administrators and SROs) believe that the presence of SROs increases school safety, such positive impressions are often not corroborated when more objective measures of school safety are used to assess the effectiveness of

SRO programs. Both quantitative and qualitative research suggests that increased use of SROs might lead to the over-criminalization of disciplinary problems and increased reliance on harsh discipline and exclusionary practices to deal with problematic students. The presence of SROs might also disproportionately affect specific segments of population (e.g., racial/ethnic minorities, special education students).

This entry begins with a discussion of the recent trend towards increasing reliance on SROs and the reasons why SROs are being used extensively in schools within the context of larger historical, structural, and political shifts. This is followed by a discussion of the possible reasons for and consequences of increased use of police in schools, including both anticipated and unexpected effects. Then extant research on the effectiveness of SRO programs is summarized. This entry ends with a discussion of implications for future research and policy.

The SRO Concept and Trends

The concept of SRO first emerged during the 1950s in Flint, Michigan, as part of the implementation of community policing (Girouard 2001). The concept grew during the 1960s and 1970s, primarily in Florida, although it did not spread nationally until the mid-1990s, when legislation such as the Safe Schools Act of 1994 and a 1998 amendment to the Omnibus Crime Control and Safe Streets Act of 1968 encouraged partnerships between schools and law enforcement. The US Department of Justice "COPS in Schools" grant program dramatically increased the use of SROs in schools beginning in 1999. SROs are typically uniformed, armed officers who have been trained for their role as school-based officers. Their duties typically involve patrolling the school, investigating criminal complaints, handling student rule/law violators, and trying to minimize disruptions. They are also often involved with educational and prevention-related programming, such as counseling students and providing DARE instruction. Although the specific goals of SRO programs may vary across time and space, the federal "COPS in Schools" program has two primary objectives: to "encourage working relationships between police and schools, thus bringing the principles and philosophy of community policing directly into the school environment" and to "assist communities in focusing leadership and resources on the issues related to creating and maintaining a safe school environment" (Girouard 2001).

According to the most recent School Crime Supplement to the National Crime Victimization Survey (NCVS), the percentage of students aged 12–18 who reported the presence of security guards and/or assigned police officers at their schools was 69 % in 2007 (Dinkes et al. 2009). A recent New York Times article (January 4, 2009) reported that more than 17,000 police officers are now placed in the nation's schools.

The use of police in schools has not always been so high. In 1975, principals in only 1 % of the nation's schools reported police stationed in the school (National Institute of Education 1978). Non-city schools and elementary schools almost never had police stationed in them. Only between 10 and 20 % of senior high schools had police. The School Survey on Crime and Safety (SSCS) data shows that by the 2003–2004 school year, principals in 36 % of schools reported police stationed in the school and by 2007–2008, the percentage had risen to 40 %. Other data sources concur. Data collected from a nationally representative sample of local police departments (from the Law Enforcement Management and Administrative Statistics [LEMAS] survey) show the number of SROs placed in public schools grew from 9,400 in 1997 to 14,337 in 2003 (Bureau of Justice Statistics 2000, 2006) and the percentage of students aged 12–18 who reported the presence of security guards or assigned police officers at their schools increased from 54 % in 1999 to 69 % in 2007, according to the NCVS.

Reasons for the Increase

Why is the use of police in schools increasing in the USA? The simplest answer is that the federal government has been increasing funding for police in schools. The Department of Justice Office of Community Policing Services (COPS) initiated the "COPS in Schools" (CIS) grant program in 1999, just after the highly publicized shootings at Columbine High School. As of July, 2005, COPS has awarded in excess of $753 million to more than 3,000 grantees to hire more than 6,500 SROs through the CIS program and more than $10 million to hire approximately 100 SROs through the Safe Schools/Healthy Students program. In 2004 the CIS program provided an additional $1.5 million in federal funding for SROs in conjunction with the Office of Justice Program's Gang Reduction Project. States often subsidize federal funding to hire additional SROs and provide funding to support school security, sometimes requiring that schools employ SROs to qualify for certain state money

The increased funding for police in schools was a highly visible response to increasing rates of juvenile crime throughout the 1980s and the numerous school shootings that occurred during the 1990s, culminating in the Columbine event. Between 1984 and 1994, the homicide rate for adolescents doubled and nonfatal victimizations increased nearly 20 % (Elliott et al. 1998). Rates of victimization at school were also high during this period, with 56 % of juvenile victimizations occurring at school in 1991 (Elliott et al. 1998). These realities created an urgency to do something about the problem. But why police in schools? Hirschfield (2008) places this response in larger historical, structural, and political context, tracing the origins of the trend towards "criminalization of school discipline." The placement of police in schools is but one element of a larger shift towards more formal treatment of student discipline. Legal reforms have mandated that certain offenses (such as drug and weapon possession) be referred to the police when they occur on school property. Other reforms have increased surveillance by using a variety of security technologies including metal detectors and security cameras and have broadened the conditions under which student searches are conducted.

This trend, according to Hirschfield (2008), was in part a delayed response to the student rights movement during the 1960s and 1970s that resulted in several judicial rulings limiting the discretion of school personnel to exclude students from school for disciplinary reasons. Teacher unions and associations and the national school principals associations, seeking to limit their constituents' liability for disciplinary actions, strongly supported more defined roles for teachers and principals with respect to school discipline responses in general and zero tolerance policies in particular. With increasing youth violence and highly publicized school shootings, the passage of the Gun-Free Schools Act of 1994 requiring that schools adopt a "zero tolerance" approach to weapons in schools became politically feasible. Many schools extended zero tolerance policies to apply also to the use of drugs and alcohol. As of 1998, 91 % of school principals reported that their schools automatically or usually (after a hearing) expelled or suspended students for possession of a gun, drugs, alcohol, or a knife (Gottfredson and Gottfredson 2001). This shift away from school personnel discretion and towards formalization of school responses to school discipline sets the stage for the more widespread use of police in schools that would soon follow.

Consequences of Increased Police Presence

Increasing police presence in schools may have made sense as a reaction to increasing rates of youth violence and school shootings, but these events cannot explain why police continue to be stationed in school buildings today. Since 1993, schools have enjoyed a strong downward trend in crime of all types that mimics the downward trend in overall youth victimization. Based on the NCVS data, Cook et al. (2010) report that the victimization rates of youths aged 12–18 at and away from schools continued to decline between 1992 and 2005. For theft and violence, the 2005 figures were about one-third of the peak in 1993. Yet the use of police in schools continued to rise as school crime rates declined.

In all likelihood, schools continue to use SROs because these officers are widely regarded as effective for maintaining school safety. Proponents believe that SROs contribute to school safety not only through their surveillance and enforcement functions but also because they create bonds of trust with students, who are then more likely to report potential crimes to them (McDevitt and Panniello 2005). SROs might also contribute to improved relations between youth and police. Of course, the presence of police in schools also provides readily available first responders in the case of real emergencies, and they help school administrators determine if certain behaviors constitute law violations.

Others share a less optimistic view of the consequences of keeping police in schools. The cost of adding SROs to schools is high, not only in personnel costs but also in extra costs related to formal processing of misbehaviors that would otherwise be handled in the school. Some research suggests that students attending schools with SROs are more likely to be arrested, especially for minor offenses such as disciplinary infractions (Theriot 2009). This becomes more problematic considering the "invisible" costs and consequences that such over-criminalization might produce (e.g., negative impact on the youths while they are held in adult detention centers before cases are dismissed or when they are actually processed through juvenile courts and labeled as criminals). In addition, one of the most troubling consequences is that SROs can shape the school discipline climate in ways that could potentially harm students. The findings from qualitative analysis of SRO effectiveness (e.g., Kupchik 2010) suggest that increased use of police officers facilitates the formal processing of minor offenses and harsh response to minor disciplinary situations. That is, school principals tend to rely on the officer as a legal adviser when there is an uncertainty about the relevant rules of law to apply. Police officers are more likely to resort to legal definitions and formal processing, especially when they have an obligation to take legal action under the zero tolerance policies. To the extent that minor behavioral problems are redefined as criminal problems and teachers are expected to rely on police in dealing with disciplinary problems, discipline responsibilities tend to be shifted away from teachers, administrators, and other school staffs to the SROs.

There are also civil liberties issues to be considered. A recent inquiry about civil rights violations related to the use of SROs highlighted another potential downside of the program. As reported in the New York Times (January 4, 2009), an A.C.L.U. inquiry into school-based arrests in Hartford, Connecticut, found that the presence of SROs disproportionately affected minority youths. This accords with a larger body of research showing that the use of suspension, especially long-term suspension, has a disproportionate impact on minority and special education populations, whose behavior places them more at risk for suspension. Civil liberties advocates have long argued that zero tolerance policies rob youths of their right to a public education. Unfortunately, these possible negative and positive consequences of increasing police presence in schools still remain untested by rigorous studies.

Research on SROs

To assess the effect of placing SRO officers in schools on the level of school safety, it is necessary to compare a reliable and objective measure of school safety pertaining to a period during which SROs worked in the schools to a suitable measure representing the counterfactual condition – for example, no SRO officers. The counterfactual measurement might be based on a reliable estimate of the outcomes taken from comparable schools with no SRO or from a time period before placement of the SRO. In either case, the number of observations of both the treatment and control conditions must be sufficient to generate stable estimates for each condition, and the outcome measurement must not be influenced by the placement of the officers in the school as it would be, for example, if the officers' own incident reports were used.

No evaluation of SROs to date meets this standard. National assessments of SRO programs supported by the National Institute of Justice (e.g., Travis and Coon 2005; Finn and McDevitt 2005) focus exclusively on the roles played by SROs, factors related to these roles, and how the SRO programs have been implemented. When they discuss the program effects, they either present descriptive statistics or simply rely on perceptions of campus safety as outcomes. Several other evaluations of SRO programs have also asked key stakeholders such as SROs or school administrators to report on their perceptions of the effectiveness of the SRO programs for increasing school safety. Not surprisingly, almost all (99 %) SROs report that their presence has increased school safety (Trump 2001) and most school administrators also report generally positive impressions of the SRO programs (e.g., May et al. 2004). However, it is well known that positive impressions of the effectiveness of an intervention are often not corroborated with more objective measures.

Other studies rely on surveys of students in schools with SROs to assess the likelihood of reporting crimes to the SRO officer, perceptions of safety, opinions about the SRO officer, and frequency of interactions with SRO officers (e.g., McDevitt and Panniello 2005). While providing useful information about youth impressions of SRO officers, these studies do little to inform us about program effectiveness because they cannot compare the experiences of students exposed to SROs with those of students not exposed. For example, McDevitt and Panniello (2005) report that students feel comfortable reporting crimes to SROs and that they feel safe at school. The important question, though, is whether students in schools with SROs feel safer than students in schools without SROs and whether they are more likely to report crimes to an adult in schools with SROs than in schools without SROs.

The first published evaluation of an SRO program to go beyond stakeholder impressions (Johnson 1999) also used cross-sectional self-report data collected from SROs, program administrators, and school principals in five schools in Birmingham, Alabama, all of which had SRO officers. But the evaluation also included a comparison of suspension counts from the year before the SROs were placed in the cities' schools until the semester after they were placed for all 18 schools that received SROs. Although the evaluation concluded that the placement of SROs into the schools was effective for reducing suspensions, the lack of a non-SRO comparison group, the reliance on a single time point of pretreatment data for schools with SROs, and the use of suspensions counts rather than rates meant that the study was not sufficiently rigorous to enable confident conclusions to be drawn regarding the effectiveness of SRO programs on youth behavior.

Subsequent evaluations have also failed to meet the standard necessary for drawing causal conclusions about program effectiveness. Only three studies have compared SRO schools with non-SRO schools. One (Theriot 2009) found that the presence of SROs increased rates of arrest for disorderly conduct but decreased rates of arrest for more serious assault and weapons charges. Another (Gottfredson et al. 2002) reported no effects of SRO presence on students' beliefs about the acceptability of offending or on their perceptions of the police, but students in the SRO schools were *less* likely than controls to report that they would be identified if they were to participate in delinquent activities. The contributions of these studies are limited because they were based on small numbers of nonrepresentative schools and nonrepresentative samples of students within the schools (Gottfredson et al. 2002), lacked comparable non-SRO schools or a sufficiently long pretreatment assessment period (Jackson 2002; Theriot 2009), or lacked measures of actual student behaviors or perceptions of school safety (Jackson 2002). Also, Theriot (2009) compared the SRO condition with non-SRO schools that employed law enforcement officers who are not trained in school-based policing but focus exclusively on law enforcement functions, making the results less interesting for our purposes.

The most recent study of SROs (Na and Gottfredson, forthcoming) used a nationally representative sample of US public schools to compare schools that added SROs or other sworn law enforcement officer to schools that did not. They examined the extent to which the addition of police in schools is associated with changes in levels of school crime and schools' responses to crime. They presented both cross-sectional comparisons and comparisons based on a longitudinal sample that allowed for more complete controls for preexisting differences among schools than were available in previous research. Outcomes included principal reports of the number of school crimes, the percentage of those school crimes that were reported to the police, and the percentage of offenses for which the offending student was removed, transferred, or suspended.

Unlike studies that have reported on key stakeholders' perceptions of the effectiveness of the SRO programs for increasing school safety, this study found no evidence suggesting that SRO or other sworn law enforcement officers contribute to school safety. That is, for no crime type was an increase in the presence of police significantly related to decreased crime rates. The preponderance of evidence suggested that, to the contrary, more crimes involving weapons possession and drugs are recorded in schools that add police officers than in similar schools that do not. The analyses also showed that as schools increase their use of police officers, the percentage of crimes involving nonserious violent offenses that are reported to law enforcement increases. These findings are consistent with the conclusions from a previous qualitative research (Kupchik 2010, p. 115) which found that the presence of police officers helps to redefine disciplinary situations as criminal justice problems rather than social, psychological, or academic problems and accordingly increases the likelihood that students are arrested at school. Adding police, however, does not increase the reporting of serious violent crimes or crimes involving weapons and drugs to law enforcement, probably because the rates of reporting of these crimes to law enforcement are already very high.

Contrary to speculations that the presence of SRO officers may unjustly rob students of their right to a public education through increased use of suspension and expulsion or may contribute to civil rights violations by disproportionately impacting minority or special education youth, Na and Gottfredson (forthcoming) found that students in schools that add police officers are no more likely to be removed, transferred, or suspended from school as a result of an offense than are students in schools that do not.

Future Directions

The findings from the studies discussed here need to be interpreted carefully due to their weak design and inherent limitations. The most important limitations in the extant research are that (1) the measurement of school crime may be influenced by the placement of a police officer in the school, (2) the research designs used in the studies have been incapable of ruling out selection artifacts as an alternative explanation for observed associations, and (3) the mechanisms through which SROs might influence school crime and safety have not been studied.

A substantial number of violent and nonviolent crimes committed within schools are never reported. Having SROs stationed at schools would tend to increase arrest rates simply by increasing surveillance and detection of crime even when the number of crimes remains the same or is reduced. Similarly, the number of crimes reported by the principal may reflect the actual level of crime in the school, the accuracy of the recording of school crime, or both. To the extent that police officers increase the accuracy of reporting in the school in which they are stationed or to the extent that police officers redefine ambiguous situations to conform to legal definitions of law violations, the observed increase in recorded school crime due to police officer presence observed in some of the studies may reflect a change in measurement practices rather than an increase in actual crime. Future studies of the effects of police officer presence on school crime should use crime measures that cannot be influenced by changes in official recording, such as student self-reports of victimization and offending in school.

The second major limitation of the existing evaluation studies is that most are based on the cross-sectional data that makes it difficult to determine the true causal directionality of the associations observed. To the extent that SROs are placed in the most troubled schools (as was demonstrated in Na and Gottfredson, forthcoming), a positive association will be observed between the presence of SROs and crime. Longitudinal data is needed to control for preexisting school conditions. Even using longitudinal data, available pretreatment measures may fail to adequately control for preexisting differences between schools with and without police officers. Future evaluations of the effects of placing police in schools should randomly assign schools to have police officers stationed therein or not.

Finally, no studies have examined the mechanisms through which SRO presence might increase or decrease school safety and crime outcomes. As noted by Girouard (2001), SROs are expected to serve a multifaceted role (e.g., law enforcement officer, counselor, educator). Through such a variety of duties, SROs can have a substantial impact not only on school safety itself but also on school environment in general, which in turn affect school safety (Cook et al. 2010). Scholars have hypothesized about the effects that the presence of police officers may have on the social climate of the school. While some evidence suggests that police contribute to a more positive school climate by encouraging student trust, other evidence suggests that increased law enforcement presence in schools erodes the school's sense of community, diminishes students' willingness to confide in school staff, and creates confusion about the roles of school administrators and the police in resolving disciplinary incidents. These social climate outcomes should be measured in future SRO research.

Implications for Policy

Any intervention strategy that adds new personnel to a system is bound to be very costly. Programs that station police in schools is no exception. In the USA, federal and local tax dollars pay these costs. In addition to the apparent costs of the program are hidden costs related to large increases in the formal processing of youthful offending in the schools in which police are placed. Note that the studies discussed in this entry suggest that students involved in less serious crime are more likely to be either formally arrested or reported to law enforcement in schools in which police are stationed.

Like many social programs that are motivated by a sense of urgency to do something about a perceived crisis situation, SRO programs have grown dramatically without benefit of scientific evaluation. As noted earlier, compared to 1975 when only 1 % of the nation's schools had police stationed in them, as of the 2007–2008 school year, 40 % of schools had police stationed in them. No rigorous study to date has demonstrated that placing police in schools promotes school safety. To the contrary, scholars have suggested that increased police presence may have the unintended effect of increasing school crime via school climate changes described earlier.

The bottom line is that more rigorous research on this topic is absolutely essential. The possibility that placement of law enforcement officers in schools increases referrals to law enforcement for crimes of a less serious nature and results in systematic construal of ambiguous situations as law violating behavior requires us to assess more carefully the school climate and school safety outcomes related to this popular and costly practice. Needed are studies involving enough schools to provide sufficient statistical power to detect important differences on the outcomes of interest, using a research design that can effectively rule out selection effects and using objective measures that are not likely to be influenced by the presence of police in the schools. It would also be desirable if the studies had sufficient statistical power to detect differences by type of school or community in the effectiveness of SRO programs. Hirschfield (2008), for example, provides a rationale for anticipating that the functions of SROs will differ in suburban and urban schools. The effectiveness of SRO programs may also differ depending on the perceived level of crime in the school and community.

In the meantime, a more cautious approach to maintaining order in schools would be to rely on approaches that have been demonstrated in research to reduce school crime. There is no shortage of such evidence-based practices. Several narrative reviews and meta-analyses of school-based interventions aimed at reducing conduct problems and delinquent behavior have been published in the last 10 years (e.g., Gottfredson et al. 2002; Hahn et al. 2007; Wilson and Lipsey 2007). All of these sources identify numerous school-based programs and practices that have been demonstrated in high-quality research to enhance school safety. Many of these effective practices are also known to be cost-effective (Drake et al. 2009). Until the effectiveness of the practice of placing police officers in schools can be demonstrated, schools are encouraged to make more extensive use of these effective practices.

Notes and Acknowledgments This entry is based on a more detailed chapter (Cook et al. 2010) on school crime control and prevention. Portions of this work are also included in Gottfredson et al. (2012) as well as in Na and Gottfredson (forthcoming).

Related Entries

▶ School Security Practices and Crime

Recommended Reading and References

Bureau of Justice Statistics (2000) Local police departments 1997. Office of Justice Programs, US Department of Justice, Washington, DC

Bureau of Justice Statistics (2006) Local police departments 2003. Office of Justice Programs, US Department of Justice, Washington, DC

Cook PJ, Gottfredson DC, Na C (2010) School crime control and prevention. In: Michael T (ed) Crime and justice: a review of research, vol. 39. University of Chicago Press, Chicago, pp 313–440

Dinkes R, Kemp J, Baum K (2009) Indicators of school crime and safety: 2008 (NCES 2009–022/NCJ 226343). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice, Washington, DC

Drake EK, Aos S, Miller MG (2009) Evidence-based public policy options to reduce crime and criminal justice costs: implications in Washington state. Victims and Offenders 4:170–196

Elliott DS, Hamburg BA, Williams KR (1998) Violence in American schools. Cambridge University Press, New York

Finn P, McDevitt J (2005) National assessment of school resource officer programs. Final Project Report, Cambridge. Available at https://www.ncjrs.gov/pdffiles1/nij/grants/209273.pdf

Girouard C (2001) OJJDP fact sheet: school resource officer training program. U.S. Department of Justice, Washington, DC

Gottfredson GD, Gottfredson DC (2001) What schools do to prevent delinquency and promote safe environments. J Educ Psychol Consult 12(4):313–344

Gottfredson DC, Wilson DB, Najaka SS (2002) School-based crime prevention. In: Sherman LW, Farrington DP, Welsh BC, MacKenzie DL (eds) Evidence-based crime prevention. Routledge, London, pp 56–164

Gottfredson DC, Cook PJ, Na C (2012) School-based crime prevention. In: Welsh BC, Farrington DP (eds) The Oxford handbook on crime prevention. Oxford University Press, New York, pp 269–290

Hahn RA, Fuqua-Whitley DS, Wethington H, Lowy J, Crosby A, Fullilove M, Johnson R, Liberman A, Moscicki E, Price LS, Snyder S, Tuma F, Cory S, Stone G, Mukhopadhaya K, Chattopadhyay S, Dahlberg L (2007) A review of the effectiveness of universal school-based programs for the prevention of violence. Am J Prev Med 33:S114–S129

Hirschfield PJ (2008) Preparing for prison? The criminalization of school discipline in the USA. Theor Criminol 12(1):79–101

Jackson A (2002) Police-school resource officers' and students' perception of the police and offending. Policing: Int J Police Strateg Manag 25:631–650

Johnson IM (1999) School violence: the effectiveness of a school resource officer program in a southern city. J Crim Just 27(2):173–192

Kupchik A (2010) Homeroom security: school discipline in an age of fear. New York University Press, New York

May DC, Fessel SD, Means S (2004) Predictors of principals' perceptions of school resource officer effectiveness in Kentucky. Am J Crim Just 29:75–93

McDevitt, J, Panniello J (2005) National assessment of school resource officer programs: survey of students in three large SRO programs. Available at http://www.ncjrs.gov/pdffiles1/nij/grants/209270.pdf

Na C, Gottfredson DC (forthcoming) Police officers in schools: effects on school crime and the processing of offending behaviors. Justice Q

National Institute of Education (ed) (1978) Violent schools – safe schools: the safe school study report to congress. U.S. Government Printing Office, Washington, DC

Theriot MT (2009) School resource officers and the criminalization of student behavior. J Crim Just 37:280–287

Travis L, Coon J (2005) Role of law enforcement in public school safety: a national survey. U.S. Department of Justice, National Institute of Justice, Washington, DC
Trump KS (2001) 2001 NASRO school resource officer survey. National Association of School Resource Officers, Boynton Beach
Wilson SJ, Lipsey MW (2007) School-based interventions for aggressive and disruptive behavior: update of a meta-analysis. Am J Prev Med 33:S130–S143

Police or Officer Discretionary Decision-Making

► Biased Policing

Police Organizations

► Organizational Change and Police Legitimacy

Police Performance Measurement

Craig D. Uchida
Justice and Security Strategies, Inc.,
Silver Spring, MD, USA

Overview

This entry discusses systems of performance measurement for police agencies. It briefly examines trends in measurement from the early years of policing to the twenty-first century. Common themes that run through the entry include the following: What are the purposes for measurement? Who is concerned about the measure (i.e., are there external or internal pressures or concerns?)? Have the measures been implemented and to what degree or level of success? Have researchers conducted evaluations? What are the results?

The entry focuses on four major areas: (1) fundamental or traditional measures, (2) the rise of CompStat as a performance management tool, (3) the use of early intervention or early warning systems to measure police misconduct, and (4) comprehensive systems of performance measurement.

Fundamentals

What should be measured by police? What is the appropriate measure of success? Is it crime reduction, clearance rates, complaints, calls for service, or use of force incidents? Is it the number of community meetings attended by officers or the quantity and quality of problem-oriented policing projects? In most instances, answers to these questions depend on the mission and goals of the agency but are further complicated by the political problem of whether they should be measured and the technical problem of how to measure them. Selecting and using appropriate measures is a problem shared by administrators in most public agencies that have no single, definable outcomes (Maguire and Uchida 2000; Bayley 1994; Moore and Braga 2003a).

It is well established that law enforcement officers engage in many different activities. They make arrests, process offenders, serve warrants, quell disturbances, respond to emergencies, work with the community, solve problems, form relationships with the community, and perform many other things in many jurisdictions. These activities are their outputs. Most attempts to measure and explain these outputs have relied on arrest, clearance, response times, and crime statistics.

Since the inception of municipal police departments in the early to mid-1800s, police agencies have gauged their performance primarily through arrest rates and reported crimes, though historians have shown that they also kept track of lost children and other noncrime activities (Monkkonen 1981). The rise of professionalism in the 1900s led the International Association of Chiefs of Police (IACP) and the FBI to focus on crime fighting rather than noncrime services. This marked a profound shift in how police characterized their work and led to the collection of national data through the Uniform Crime Reports (UCR) (Uchida 2010).

Reformers like Chief August Vollmer of the Berkeley Police Department and the IACP urged departments to maintain records related to crime and operations. The IACP Committee on Uniform Crime Records was formed in 1927 to create a system for collecting uniform police data. Standardized definitions were formulated to overcome regional differences in the definitions of criminal offenses. By 1929, the committee finalized a plan for crime reporting that became the basis for UCR. The committee chose to obtain data on offenses that come to the attention of law enforcement agencies because they were more readily available than other reportable crime data. Seven offenses, because of their seriousness, frequency of occurrence, and likelihood of being reported to law enforcement, were initially selected to serve as an index for evaluating fluctuations in the volume of crime. These Crime Index offenses were murder and nonnegligent manslaughter, forcible rape, robbery, aggravated assault, burglary, larceny/theft, and motor vehicle theft. By congressional mandate, arson was added as the eighth Index offense in 1978.

There were two primary purposes for uniform crime reporting. First, police leaders saw the value of better data for the improvement of management and operation strategies. Second was the desire for data that could combat the perception of crime waves created by the media. Police managers were concerned that such crime waves were seen as a reflection of their departments' inadequacies. They believed that uniform crime data would dispel the notion of crime waves and, therefore, demonstrate the value of police efforts to control crime. The tension between developing a system that would represent a true measure of crimes known to police and developing one that would prove crime was not increasing has been present in UCR since the beginning (Wellford 1982).

Since 1930 the UCR has evolved from a relatively small data collection effort into a large-scale national effort. In its first year of operation, the program collected data from 400 cities in 43 states, representing about 20 million people. In 2011 UCR collected data from nearly 17,000 jurisdictions in all 50 states, representing about 300 million people. UCR remains a voluntary reporting program, with city, county, and state law enforcement agencies reporting monthly to the FBI the number of part 1 offenses and part 1 and part 2 arrests that have occurred within their jurisdictions. In addition to monthly tallies of offenses and arrests, additional data are captured on particular offenses, and data on age, sex, race, and ethnicity are collected for arrests.

Another important measure is response time, the seconds or minutes that it takes for a patrol car to respond to a call for service. This measure has been used by police to show the public how efficient and effective they are in dealing with crime. But research has shown that rapid response does not lead to apprehension of suspects immediately nor is it an expectation of those who call the police (Spelman and Brown 1981). Changes in available resources have led police to streamline their response strategies and in the process make measurement more specific. Police have employed alternative strategies, known as differential police response (DPR) where priorities are established regarding calls, reports are taken over the telephone by civilians, or complainants are asked to use the Internet and online reporting. Since the 1980s, departments often use "priority" call types. For example, Priority 1 may include crimes in progress or more serious violent crimes (homicides, rapes, robberies), while Priority 2 and 3 calls may include those crimes that require nonimmediate responses (burglaries, alarms, etc.).

Changes in Measures

With the discovery of police discretion in the 1960s, researchers and police administrators realized that order maintenance and other noncrime activities were also important to measure, as officers were not engaged in crime fighting all of the time. But it was not until the 1990s with the community policing movement did researchers and police begin to recognize that this new paradigm should include new performance measures. Alpert and Moore (1993) recommended the adoption of nontraditional measures that were tied to essential goals of community policing. Yet they still focused primarily on police-related data and measures, including number of traffic tickets, crimes and arrests cleared by conviction, and police misconduct. Some of the nontraditional measures included time and dollars allocated to problem-solving initiatives, types of assistance provided to victims, and resources for strengthening police-community relationships. From this point forward, researchers and police began to develop appropriate measures that covered the spectrum of traditional policing to community policing and problem-oriented policing activities and outcomes.

The Rise of CompStat

In 1994 Commissioner William Bratton and Jack Maple introduced the New York City Police Department (NYPD) to CompStat (comprehensive computer statistics), a management tool that emphasized using appropriate resources in the right place at the right time to reduce crime. A harsh and oftentimes unforgiving method, CompStat was known for publicly questioning precinct commanders about their knowledge of crime in their areas and what they were doing about the problem. Using technology and computer-generated mapping, CompStat generates crime statistics for each precinct and uses them to hold precinct commanders accountable for their areas (Bratton and Knobler 1998).

CompStat became a popular mechanism for crime reduction, and police departments across the country quickly adopted the methodology. Indeed, by 2000 nearly one-third of the country's largest police agencies had implemented a CompStat-like program (Weisburd et al. 2001). This occurred even though evidence of crime reduction was not validated by research efforts. The explosive growth in computer power (in terms of speed and storage capacity relative to cost), the standardization of formats across software programs, and the increase in user accessibility (and decrease in cost) to these software programs among other factors facilitated the dramatic spread of CompStat across police agencies. In addition, cities, counties, and at least one state (Maryland) have established CitiStat, CountyStat, and StateStat (respectively) to assist in decision making and accountability.

Though some scholars express doubts about the success of CompStat (especially for crime reduction), it is generally accepted that Bratton mobilized the NYPD and, later, the Los Angeles Police Department (LAPD) to raise arrest productivity significantly.

From an administrative perspective, CompStat represents a significant change from the past. Moore and Braga (2003a) see CompStat as setting the standard for innovation for police management and as a unique system for internal and external accountability. They emphasize that as a measurement system it is "behaviorally powerful" in that it drives the NYPD and other departments in the following ways:

- It is aligned with organizational units that hold managers accountable for performance.
- The measures are simple and continuously used so that performance can be observed over time.
- The measures are aligned with those who oversee the department externally.
- Accountability is frequent, so that managers are attentive.
- Managers recognize that CompStat affects their current and future standing, pay, and promotional opportunities.
- Reviews of performance are public.
- Comparisons can be made across situations and managers.

While the original model of CompStat has evolved since 1994, its basic principles of accountability and measurement remain intact.

Systems to Account for Police Behavior

The problems of police-community relations identified four decades ago by the President's Commission on Law Enforcement and Administration of Justice (1967) continue to exist. Concern over use of force by police officers led Congress in 1994 to order the US Department of Justice (DOJ) to acquire data about the use of excessive force by law enforcement officers and to publish an annual summary of the data acquired (McEwen 1996). The Bureau of Justice Statistics served as the repository for such data.

Concern over the practice of racial profiling by police officers making traffic stops, a long-time concern of minority communities, led to calls for measurement on the severity of the problem. During the early part of the twenty-first century a number of large and small departments across the country began to implement methods to measure officer encounters with the public as a result. More than 20 states have passed legislation prohibiting racial profiling and/or requiring jurisdictions within the state to collect data on law enforcement stops and searches. According to Northeastern University's Racial Profiling Data Collection Resource Center, "hundreds of law enforcement agencies" are now collecting such data (see http://www.racialprofilinganalysis.neu.edu/index.php).

Early Warning Systems

In addition to collecting racial profiling data, the US Department of Justice (DOJ), through the Violent Crime Control and Law Enforcement Act of 1994, imposed systems of accountability on law enforcement for "patterns or practices" of policing that violate the Constitution (Chanin 2012). Since 1994 at least 25 law enforcement agencies have been investigated and/or reached settlements with DOJ regarding practices that violate citizens' federal rights. These agencies often agree to operate under consent decrees to improve policing practices to avoid further litigation. A common requirement of these consent decrees is the implementation of an early warning system (EWS) or early intervention (EI) system designed to identify and intervene with officers at risk for future misconduct. EI systems are promoted as a specific example of practices designed to promote police integrity (United States Department of Justice 2001). One of the major assumptions behind these systems is "that a small number of police officers are responsible for a disproportionate amount of problematic police behavior." The long-term goal of such a system is to "create a culture of accountability in the agency" and thus reduce citizen complaints and litigation (USDOJ 2001, p. 10).

EI systems are expected to improve the capacity of police agencies to police themselves by concentrating their efforts on the people that generate the most problems. By generating "red flags" on persons who engage in more misbehavior than their colleagues, police agencies can intervene more effectively in reducing that behavior. Evaluations of these systems have shown mixed results; some show success in reducing officer misconduct, while others raise questions about their efficacy.

An example of an early intervention system is that of TEAMS II in the Los Angeles Police Department. In TEAMS II (Training Evaluation and Management System 2nd phase), data are collected from 14 separate systems and compiled for all 10,000 sworn officers in the LAPD. This information is compiled in the Risk Management Information System (RMIS) and is used in two ways: employee performance assessments/continued monitoring and the automated risk management system. Like other EI systems, RMIS is designed to examine employee outcomes across five domains: use of force, citizen complaints, claims and lawsuits, preventable vehicle crashes, and vehicle pursuits. After one of these events has occurred, RMIS conducts an automated assessment of the officer's performance against patterns established by his or her peers. These measures are then compared to the performance of the peer group defined by similar work functions. An action item or AI is generated if the measure for the individual officer is substantially different from his/her peer group.

Once an action item is generated, the report is automatically forwarded to the employee's supervisor for review. The supervisor is then required to notify the employee of the action and to conduct a review of all relevant information to identify whether the employee's actions (both the current activity as well as prior activity) indicate a persistent pattern of at-risk behavior. If evidence suggests the possibility that at-risk behavior may be occurring, the supervisor engages in an investigation of the available information in the system reports. The supervisor documents the results of this examination by summarizing the various reports, determining whether there is a pattern of conduct, provides a comparison of the employee's performance to similar employees, explains any significant differences between the employee's performance and that of similar employees, selects and justifies a disposition (including a decision to take no action), and provides a brief summary of supervisor's discussion with the employee about this report. Cases that are canceled, judged to require no action, or end with informal counseling receive no further review up the chain of command, but those receiving other actions are reviewed up the chain to the commanding officer level.

A recent Office of Inspector General (OIG 2012) audit found that TEAMS II corrective actions are rare. Only 64 of 1,384 AIs generated from August 1, 2010, to January 31, 2011, yielded any corrective action disposition. This audit also revealed that approximately 72 % of the AI investigations were completed within 60 days of their initiation (OIG 2012).

As part of its accountability process, LAPD has integrated the TEAMS II process into its CompStat process. This ensures that the middle and upper levels of LAPD's hierarchy are kept informed of progress in reducing undesired or questionable patterns of police practice and makes it more likely that the organization will attend to these aspects of agency as well as individual officer performance.

Studies of Early Warning/Intervention Systems

Since 2000 a handful of studies have examined EI systems similar to the LAPD's. These include a study of Pittsburgh (Davis et al. 2002); a national evaluation and three case studies by the Police Executive Research Forum (PERF) and the University of Nebraska Omaha (UNO) (Walker et al. 2001); Worden et al. (in press) in a northeastern city; and a study of a southeastern police department by Bazley et al. (2009).

In 1997 the Pittsburgh Police Bureau became the first police organization to be issued a consent decree (Davis et al. 2002). The consent decree required implementation of an early warning system (EWS) and it stipulated both the specific information to be included (citizen complaints, training, transfers, arrests, etc.) as well as the categories that should be capable of being examined (individual officer, squad, shift, unit, etc.). This EWS was designed to compare officers to their peer groups using standard deviation calculations. In this way, for example, officers working the day shift in a low-crime area were not directly compared to officers working the night shift in a high-crime area.

The evaluation looked at the effects of the consent decree by conducting in-depth interviews, examining trends in departmental procedures, and administering a community survey (Davis et al. 2002). The researchers did not directly analyze the effect of the EWS on officer performance, but instead reported on interviews with officers and supervisors and provided data on trends in police performance and disciplinary actions before and after the consent decree was issued. Overall the interviews showed a mix of views about the EWS. Some supervisors appreciated the system, while others said it did not provide any new information. Supervisors complained about the extra time and paperwork and said that they had less time to respond to serious incidents and to supervise their squads on the street. Officers voiced concerns about being afraid to do their jobs, stating that it was safer to do nothing.

Ultimately, the researchers concluded that "the early warning system . . . all but ensures that officers heading for trouble will be identified and efforts made to straighten them out." They also recognized problems like system redundancies and the extra time for filling out paperwork and commented that "most of the potential problem officers identified by the early warning system may already be known to supervisors" (Davis et al. 2002, p. 62). Lastly, they noted that in the absence of a more thorough methodology, it is difficult to determine what effects the EWS may have had on the performance, attitudes, and values of officers in the department and what effects can be attributed to the imposition of the consent decree.

Walker et al. (2001) conducted a national survey of police agencies to determine the prevalence of early warning systems. They found that 27 % of surveyed agencies had an EWS in 1999, with 12 % planning to institute a system. Larger departments were more likely than smaller agencies to create an early warning system. The researchers also evaluated the effectiveness of three different early intervention systems: Miami-Dade, Minneapolis, and New Orleans. They found that in all three departments the "program appeared to reduce problem behaviors significantly" and that they "encourage changes in the behavior of supervisors as well as of the identified officers" (Walker et al. 2001).

More recently, Harris (2009, 2010) and Worden and his colleagues (Worden et al. in press) examined officer misconduct and evaluated the effects of an early intervention system in a large police department in the northeastern United States. In evaluating the overall effects of the EI, Worden et al. found that "rates of citizen complaints, personnel complaints more generally, and secondary arrests all declined." Yet, matched controls in the department "exhibited very similar patterns, suggesting that these changes should not be attributed to the Officer-Citizen Interaction (OCI) School intervention." They also found a small negative effect on arrests, particularly proactive arrest, stemming from the classes. Finally, they found a marked and problematic potential for false positives in early identification systems. If false positives occur at too high a rate, this could seriously jeopardize the efficacy of these systems as well as their legitimacy among officers and supervisors.

Bazley et al. (2009) examined use of force reports as an indicator for one large (over 1,000 sworn officers) urban southeastern police department's EI program. In 2000, when the study was conducted, 33 officers qualified for the program under use of force criteria. Departmental data were analyzed to determine whether either of their independent variables – total number of officer use of force reports and weighted officer force factor value (calculated by subtracting force level from resistance level) – had a statistically significant effect on officer qualification for the early intervention program. They found that both variables were statistically significant, but only the number of use of force reports was in the direction expected (more use of force reports indicated a greater likelihood of qualification). Weighted officer force factor, on the other hand, indicated that the program actually "flagged officers who clearly displayed a tendency to use less force when confronting higher levels of resistance" (p. 119). Moreover, the system did not flag any of the 22 officers in the department with negative weighted force factor values that indicated a tendency "to use more force than resistance encountered" (p. 120). Ultimately, then, Bazley et al. state that "these findings would seem to bring into question whether the qualifying use of force criteria, i.e., tracking the numbers of higher level use for force incidents quarterly and annually, are actually identifying the problematic officers in this department" (p. 121).

Comprehensive Systems of Performance Measurement

In the twenty-first century, researchers and practitioners have sought appropriate and comprehensive measures for police performance. The search for meaningful measures included national organizations – the Commission on Accreditation for Law Enforcement Agencies (CALEA) and the Police Executive Research Forum (PERF) working with researchers and police executives. Ultimately, they wanted to broaden the scope of police activities and goals and attached specific measures to them. The work of Maguire (2004), Moore and Braga (2003b), and Milligan et al. (2006) established the theoretical and research-based groundwork for CALEA and PERF. These works showed the multidimensionality of police work and demonstrated the need for multiple measures and diverse data collection methods and analyses.

Maguire (2004) traced the development of performance measures within the context of police organizations. He stressed the need for formulating broad dimensions of police performance and the use of specific measures and tools. Citing previous research by O'Neill et al. (1980), Hatry et al. (1992), Moore (2003), and Mastrofski (1999), Maguire provides examples of the dimensions of police activities. In brief, they include such areas as crime prevention, crime control (apprehension of offenders), fear reduction and the feeling of security, fairness and courtesy, police administration (financial resources), use of force, and satisfaction of customer service. Measurement tools include the use of general community surveys, specific contact surveys, employee surveys, direct observations, and independent testing or simulations for collecting information about an agency's performance. Moreover, he included methods of analysis, weighting of measures, and a detailed section on how to implement comparative performance measures.

Similarly, at PERF Milligan et al. (2006) identified seven dimensions of policing that warranted measurement: (1) community safety and security; (2) perceptions of safety and security; (3) confidence, trust, and satisfaction; (4) response to crime; (5) prevention of crime; (6) enhancement of safety and security; and (7) community health. The authors found that the Prince William County (VA) Police Department followed these dimensions and had also established appropriate performance measures.

Building on all of these studies, Davis et al. (2010) and CALEA developed a pilot project to establish useful and useable performance measures. Researchers and a group of police executives drafted performance indicators to assess nine dimensions of policing: (1) delivering quality service; (2) fear, safety, and order; (3) ethics/values; (4) legitimacy/customer ervice; (5) organizational competence/commitment to high standards; (6) reducing crime and victimization; (7) resource use; (8) responding to offenders; and (9) use of authority.

Each of these categories contained multiple indicators that applied to a spectrum of large and small agencies, police departments, and sheriff's offices. Four CALEA-accredited agencies – Dallas, TX; Knoxville, TN; Kettering, OH; and Broward County, FL – developed and tested 28 specific measures that were part of the nine dimensions. Data were collected using five instruments, including:

- *A self-assessment instrument* that asks agencies to provide traditional data – basic crime rates, clearance rates, arrests, citizen complaints, and budgetary information
- *Community surveys and business surveys* to gauge people's and local retailers' perceptions of police effectiveness
- *Contact surveys* that assess the quality of interaction between police and citizens who request police assistance or who are stopped by the police
- *Officer surveys* that measure a broad array of perceptions of police including morale, integrity, agency leadership, and knowledge of policies

After successfully implementing the pilot phase of the project, five participants were added to the original four – Raleigh, NC; Avon, CT; Boca Raton, FL; and Las Vegas, NV, Police Departments and the Arapahoe County, CO, Sheriff's Office.

The results from the analysis of the nine agencies show the value of this comprehensive set of measures. As the authors indicate, the set of measures or subsets can be used for a variety of purposes (with appropriate caveats). Some examples include:

- Comparing performance across patrol areas to determine "whether some areas stand apart from the others in how satisfied citizens are with the handling of a crime complaint or a traffic stop"
- Making year-to-year comparisons of agency of patrol area performance to determine if performance is trending in a positive direction
- Comparing performance with other agencies to provide a benchmark against which to assess how an agency is doing in various areas

Lastly, the authors provide detailed information about their measures, particularly the survey instruments. They give specific instructions about implementing the surveys and describe their results in the nine agencies. Further, they provide the range of scores for the specific measures that they use. These explanations and tables serve as a guide as to whether an agency's performance on each dimension is within the bounds established by agencies involved in the field test. The value of the research allows another agency or researcher to use the data collection instruments and analyze the data appropriately. The results from an individual department can be compared to the nine CALEA agencies to obtain an overall context for its performance.

Key Issues/Controversies

While a range of performance measures for police has been discussed, created, and implemented for over 150 years, adequate measures have yet to be fully agreed upon and validated by researchers and practitioners. This entry has attempted to identify those approaches and to describe research and evaluation that have been conducted on those measures. This section discusses key issues and ongoing controversies about measurement.

On Traditional Measures

Traditional measures of arrests, crimes known, clearance rates, and response times have been criticized for their lack of comparability and validity. In addition, using crime rates as a measure of police success ignores other factors such as the social, economic, and political influences within jurisdictions. Reported crime rates also do not reflect the number of actual crimes that occur. Underreporting of crime is commonplace, especially among high-risk populations, immigrant groups, those fearful of retaliation, and those who do not believe that police can do anything about a particular crime. Arrests and clearance rates are also problematic. Legal definitions of arrests vary across jurisdictions, making them difficult to compare. Clearance rates are indicators of investigative effectiveness but oftentimes can be manipulated and subject to measurement error (Cordner 1989).

On CompStat

As indicated above, CompStat has been a popular method as a management tool and to measure performance. From a measurement perspective, CompStat relies on the same crime data that have been criticized as inadequate measures of police performance for the past 30 years. The difference in their use today is that computer-generated maps, trend analysis, and other visualizations assist decision makers in "seeing" what is occurring on a week-to-week, month-to-month, and year-to-year basis.

CompStat has received accolades and criticism from various circles, including researchers and practitioners. Evaluations of CompStat focus primarily on its effectiveness (or lack thereof) regarding crime reduction (see Willis et al. 2007). Evaluations of CompStat as a management tool are virtually nonexistent.

On Early Intervention Systems

Early warning/intervention systems focus primarily on officer misconduct, though the LAPD's system includes performance measures of employees. Researchers have found mixed results about the effectiveness of EWS/EIs in deterring that misconduct. Problems with the studies abound and are self-reported by the researchers themselves. For example, Walker et al. could not determine the most effective aspects of intervention or whether certain aspects are more effective for certain types of officers. This problem occurs with the studies by Davis et al. and Bazley et al. Another problem is the inability of the early warning systems to collect data on positive police officer performance, so measuring the deterrent effect of an early intervention is difficult to assess. Third, researchers could not disentangle the effect of the general climate of rising standards of accountability on officer performance from the effect of the intervention program itself.

On Comprehensive Performance Measurement Systems

While researchers and practitioners recommend different designs for comprehensive performance measurement systems, the actual implementation of those systems is rare. O'Neill et al. (1980) and Hatry et al. (1992) suggested multiple dimensions and multiple measures, but those were not implemented. Maguire and Moore and Braga's systems were adopted in part by CALEA and used in nine agencies (Davis et al. 2010). It remains to be seen whether more agencies will implement the system.

The CALEA/Davis et al. approach (2010) is noteworthy for its use of multiple surveys and data instruments, but the difficulties lie in their actual use. A representative sample is necessary to obtain accurate perceptions of the jurisdiction and decisions need to be made about the sampling frame and sample size. Similarly, the use of contact surveys with victims of crime or those who were involuntarily stopped by police include problems with sample size, administering the survey, and completion rates. If these can be overcome, however, the information gathered from the surveys are particularly valuable for the police and the jurisdiction that they serve.

Future Directions

Despite the problems associated with performance measurement, the future of this topic is bright. Police agencies are striving to become more efficient and effective in their work because of the budget crises and economic downturns in their jurisdictions. In addition, police executives are looking for ways to continue to reduce crime, particularly violent crime. Measuring performance and justifying or maintaining their budgets are important to these executives. As a result, using CompStat-like methods, obtaining feedback from residents and businesses, keeping track of officer performance, and continuing to make managers accountable for their areas, divisions, and precincts are methods that will continue to be used and expanded.

Technological developments will assist in the ease of use of measurement tools. Traditional data of crimes, arrests, clearance rates, and response times will be managed more efficiently because of improvements in computer systems and will be analyzed and visualized more quickly than before. Computer-generated "dashboards" that show charts, graphs, maps, and tables of crime are available now and will be used extensively by executives, commanders, and managers for a variety of purposes including performance measurement. Tablets, iPads, smart phones, software applications, and the Internet will enable agencies to conduct surveys of officers, citizens, and businesses on a routine basis.

Related Entries

- ▶ COMPSTAT
- ▶ Control of Police Misconduct
- ▶ Development of the UCR and the NCVS
- ▶ History of the Statistics of Crime and Criminal Justice
- ▶ Measuring Police Performance
- ▶ Measuring Police Unit Performance
- ▶ Methodological Issues in Evaluating Police Performance
- ▶ Police and the Excessive Use of Force

Recommended Reading and References

Alpert GP, Moore MH (1993) Measuring police performance in the new paradigm of policing. In: DiIulio JJ, Alpert GP, Moore MH, Cole GC, Logan C, Petersilia J, Wilson JQ (eds) Performance measures for the criminal justice system. Discussion Papers from the BJS-Princeton Project. US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics, Washington, DC

Bayley DH (1994) Police for the future. Oxford University Press, New York

Bazley T, Mieczkowski T, Lersch K (2009) Early intervention program criteria: evaluating police use of force. Justice Quart 26(1):107–124

Bratton W, Knobler P (1998) Turnaround: how America's top cop reversed the crime epidemic. Random House, New York

Chanin JM (2012) Negotiated justice? The legal, administrative, and policy implications of 'pattern or practice' police misconduct reform. Research report submitted to the U.S. Department of Justice

Cordner GW (1989) Police agency size and investigative effectiveness. J Crim Just 17(3):145–155

Davis RC, Ortiz CW, Henderson NJ, Miller J, Massie MK (2002) Turning necessity into virtue: Pittsburgh's experience with a federal consent decree. Vera Institute of Justice, New York. Retrieved from http://www.vera.org/content/turning-necessity-virtuepittsburghs-experience-federal-consent-decree

Davis RC, Cordner G, Hartley C, Newell R,Ortiz C (2010) Striving for excellence: a guidebook for implementing standardized performance measures for law enforcement agencies. CALEA, Washington, DC. Retrieved from http://www.calea.org/content/striving-excellence-guidebook-implementing-standardized-performance-measures-law-enforcement

Eck J, Maguire ER (2000) Have changes in policing reduced violent crime: an assessment of the evidence. In: Blumstein A, Wallman J (eds) The crime drop in America. Cambridge University Press, New York, pp 207–265

Harris C (2009) Exploring the relationship between experience and misconduct: A longitudinal examination of officers from a large cohort. Police Quarterly 12(2):192–212

Harris C (2010) Problem officers? An analysis of problem behavior patterns from a large cohort. J Crim Just 38(2):216–225

Hatry HP, Blair LH, Fisk DM, Greiner JM, Hall JR Jr, Schaenman PS (1992) How effective are your community services? Procedures for measuring their quality, 2nd edn. Urban Institute and International City/County Management Association, Washington, DC

Maguire ER (2003) Measuring the performance of law enforcement agencies – Part 1. *CALEA Update*, issue 83, September. Retrieved October 2011 http://www.calea.org/caleaupdate-magazine/issue83/measuring-performance-law-enforcementagenciespart1of-2-oartarticl

Maguire ER (2004) Measuring the performance of law enforcement agencies – Part 2. *CALEA Update*, issue 84, February. Retrieved October 2011 http://www.calea.org/calea-update-magazine/archive/2004-02

Maguire ER, Uchida CD (2000) Measurement and explanation in the comparative study of American police organizations. In: Duffee D (ed) Criminal justice 2000, vol 4, Measurement and analysis of crime and justice. National Institute of Justice, Washington, DC

Mastrofski SD (1999) Policing for people. Ideas in American policing. U.S. Department of Justice, Police Foundation, Washington, DC

McEwen T (1996) National data collection on police use of force. NCJ 160113. U.S. Department of Justice, Bureau of Justice Statistics and National Institute of Justice, Washington, DC

Milligan SO, Fridell L, Taylor B (2006) Implementing an agency–level performance measurement system: A guide for law enforcement executives. US Department of Justice, Washington, DC NCJ 214439

Monkkonen E (1981) Police in urban America: 1860–1920. Cambridge University Press, Cambridge

Moore M (2003) The bottom line of policing: What citizens should value (and measure!) in police performance. Police Executive Research Forum, Washington, DC

Moore MH, Braga A (2003a) The bottom line of policing: what citizens should value (and measure!) in police performance. Police Executive Research Forum, Washington, DC

Moore MH, Braga A (2003b) Measuring and improving police performance: the lessons of compstat and its progeny. Policing Int J Police Strateg Manag 26(3):439–453

O'Neill MW, Needle JA, Galvin RT (1980) Appraising the performance of police agencies: the PPPM (police program performance measures) system. J Police Sci Admin 8(3):253–264

Office of the Inspector General (2012) Training evaluation and management system (TEAMS) II audit, phase I. 2010–2011. Office of the Inspector General. Los Angeles Police Commission

President's Commission on Law Enforcement and Administration of Justice (1967) Task force report: the police. U.S. Government Printing Office, Washington, DC

Spelman W, Brown D (1981) Calling the police: citizen reporting of serious crime. Police Executive Research Forum, Washington, DC

Uchida CD (2010) The development of the American police: an historical overview. In: Dunham RG, Alpert GP (eds) Critical issues in policing: contemporary readings, 6th edn, Prospect heights. Waveland Press, Prospect Heights

United States Department of Justice (2001) Principles for promoting police integrity: examples of promising police practices and policies. Retrieved from https://www.ncjrs.gov/pdffiles1/ojp/186189.pdf

Walker S, Alpert GP, Kenney DJ (2001) "Early warning systems: responding to the problem police officer." Research in Brief. National Institute of Justice, Washington, DC

Weisburd D, Mastrofski SD, McNally A, Greenspan R, Willis JJ (2001) COMPSTAT and organizational change: findings from a national survey. Police Foundation, Washington, DC

Wellford CF (1982) Redesigning the uniform crime reports. Am J Police 1(2):76–93

Willis JJ, Mastrofski SD, Weisburd D (2007) Making sense of Compstat: A theory based analysis of organizational change in three police departments. Law & Society Review 41:147–188

Worden RE, Kim M, Harris C, McGreevy M, Catlin S, Schlief S (in press) Intervention with problem officers. An impact evaluation of one agency's EIS. Crim Justice Behav

Police Promotional Practices

Cassi L. Fields
Fields Consulting Group, Inc., McLean,
VA, USA
LeT Corporation

Overview

Controversy surrounds promotional process in American police departments. This is true despite the fact that promotional tests are valid and useful tools when constructed consistently with professional and legal guidelines. Candidates and management share some concerns about promotional processes, but candidates' concerns focus mostly on the degree to which promotional processes are high quality, transparent, consistent, and fair. Management's concerns focus more on whether the promotional process identifies the best candidates, and minimize costs and grievances. The research on promotional tests is informative in the best practices to construct tests and promotional processes. In an effort to avoid direct and indirect costs, departments often avoid utilizing best practices. When the result of cost savings is an invalid test, the consequences are severe including litigation, low employee morale, and ineffective leadership succession plans, whereas the investment in a well-designed promotional process creates buy-in, promotional opportunities, employee development, and a clear leadership succession plan.

Key Issues/Controversies

Introduction

Promotional processes in American police departments are considered high-stakes tests because their outcome affects candidates, department management, customers (i.e., citizens), and potential other departments within a jurisdiction including human resources, budget, and legal. These tests may include a simple multiple-choice job knowledge test, assessment center, structured interview, performance evaluation, or some combination of these components. Relative to other test types, a multiple-choice test is inexpensive. However, it is a limited tool. Multiple-choice tests do not require candidates to directly demonstrate their knowledge and have often been shown to result in adverse impact against minority candidates (Hunter and Hunter 1984; Rushton and Jensen 2005).

Assessment centers are a combination of simulations such as oral presentations, writing exercises, and emergency incidents that measure candidates' knowledge, skills, and abilities in different testing formats. The research has shown that assessment centers are valid measures of knowledge, skills, and abilities (Arthur et al. 2003) and are a more direct measure than a job knowledge test.

Interviews and performance evaluations are often used in place of, or to augment, other types of tests. An interview can be structured to be a more direct measure of knowledge, skills, and abilities than a multiple-choice test. Structured interviews have high validity, but they do not measure written communication skills.

Performance evaluations are generally unreliable measures. Unless properly validated, they should not be incorporated into a promotional process, and even then, they are routinely subject to successful legal challenges. The job analysis of the target rank determines what knowledge, skills, and abilities should be measured. Whenever possible, the best promotional practices incorporate multiple assessment tools in order to measure a broad set of important knowledge, skills, and abilities (indirectly and directly) and to reduce subgroup differences whenever possible.

Controversy in Promotional Processes

Controversy surrounds promotional processes in American police departments. The causes of the controversies vary. Typically, these controversies stem from the candidates who take the tests; less frequently, they stem from management.

Candidates who do not get promoted after they take a promotional process often must wait

1–3 years, and in some departments longer, to go through another promotional process. These candidates believe that this *waiting period* delays their salary increases, their career progression, and, potentially, their pensions.

Those that do get promoted are critical to the success of the department because they must be able to supervise and lead officers who are often assigned to decentralized positions. They must ensure that patrol officers protect the rights of citizens, protect victims, properly handle illegal drugs and money, and perform many other critical tasks. Without effective supervision and leadership, there is unlimited opportunity for intentional and unintentional corruption.

One of the most challenging aspects of using promotional exams in police departments is ensuring that they are valid *and* produce a diverse group of candidates. This result can be achieved, but it takes a deliberate and thorough plan to succeed. Further, creating a valid test that is defensible if legally challenged is difficult and expensive. Recently, it has become more difficult to defend a challenged test because proving that the test is valid may be insufficient (Johnson v. City of Memphis 2006). Today, it seems that the defense must also demonstrate that they considered all equally valid alternatives and ultimately administered the test with the least adverse impact. However, finding or developing a valid alternative that produces acceptable diversity remains elusive due to:

1. Testing outcomes (i.e., population subgroups may not perform equally)
2. Strict laws that govern testing in employment settings

Who Writes the Test?

There are no statistics showing how many police departments outsource their promotional exams compared to those who write them internally. Since the 1970s, there has been insufficient internal departmental staff to validate promotional exams consistent with the *Uniform Guidelines on Employee Selection Procedures* (*EEOC*1978), the *Principles for the Validation and Use of Personnel Selection Procedures* (2003), and case law precedent. There is a tendency for departments who write their exams internally to continue to do so until there is a specific challenge. Those that use external consultants have experienced mixed results, but there seems to be a pattern of better success in defending a test with consultant services than without.

When insufficient resources are devoted to promotional tests, an invalid test may result and the consequences are often severe. A civil rights challenge (Civil Rights Act 1964, as amended in 1991) can stop a promotional process for several years because the litigation is lengthy and expensive. The ripple effects from a suspended promotional process affect the ability to fill vacancies in other ranks, competence in supervision and leadership within the department, and individuals seeking promotion but who cannot compete due to the litigation.

Candidate Protests

Candidates express concerns about promotional processes that contain poor-quality test content and lack transparency, consistency, and fairness.

Test Quality

Researchers have analyzed test bias and test flaws and the manner in which they may contribute to test performance disparities. This line of research presumes that test bias is either partially or entirely the cause of test performance disparities. These biases and flaws include, but are not limited to, test unreliability, invalid cognitive load, and invalid passing or weighting schemes (Hunter and Hunter 1984; Sackett et al. 2001). In general, these theories have not led to the creation of perfect tests as evidenced by continued internal grievances and external litigation.

Furthermore, candidates are not likely to accept a test containing questions with typographical errors, those that ask about *minutia*, those that are academic in nature, or those that are relevant solely to a subset of the population (e.g., questions that investigators need to know when it is a first-line supervisory test). The degree to which these types of tests negatively affect subgroup populations determines what types of actions candidates take. If, for example, minority candidates are excluded from promotion based upon a test, they might sue their department or jurisdiction. If all candidates believe the test was poorly constructed, they might internally appeal, grieve, or verbally protest. In cases such as these, it is likely management will know about the problems. However, management rarely takes action unless pressure comes from an outside source such as the US Department of Justice, the Equal Employment Opportunity Commission, a lawyer, or the union. The irony is that a successful promotional process, without that unwanted interference, is developmental. It creates a highly ethical environment, and it is an automatic part of a succession plan.

Transparency

There is little research on promotional process transparency. Transparency can be defined as "informing candidates about their promotional process *before* the process is administered to include, but is not limited to, test dates, times and locations, test components, what will be measured, how it will be administered, how it will be scored, how the results will be published, and how long the results will be used for promotions." There has been some debate over whether the scoring criteria for a test can be determined after test administration. This debate is contentious because the reasons for assigning passing scores or test component weights after test administration are generally unethical. For example, this system would allow the department to select a specific number of candidates, select specific individuals, or select a specific group of individuals (e.g., minorities, females, nonminorities). Each of these reasons negates transparency. Candidates, in general, prefer that the test process and rules be announced in advance of the test so that any potential for manipulation of the results is eliminated. Unless the leaders of the department have ulterior motives, they too tend to prefer a completely transparent promotional process.

Test score banding falls under candidates' transparency concern. Banding is a process in which candidates' test scores are placed into bands or groups. For example, candidates scoring between 90 and 100 are place in one band and candidates scoring between 80 and 89 are placed in another. Banding can be unpopular because candidates often believe that the highest-scoring candidate should be promoted first. This belief is based upon the theory that a test is a perfect predictor of future job success. A valid test is not a perfect measure of performance nor is it complete. It is a predictor; it predicts who is most likely to succeed in the target rank. A separate article could be devoted to banding because it is controversial (e.g., Bobko and Roth 2004). This author has had success in using administrative bands which are used to place candidates' test scores into predefined (fixed) groups (e.g., rule of 10) and then creating a valid measure to select from the band. This procedure can assist with promoting a diverse workforce, and if the rules are published in advance, it is transparent.

Consistency

It is well known among test developers that a test's validity is limited by its reliability (Cascio 1998). This means that if a test is unreliable, it cannot measure what it is intended to measure. Three examples of unreliability are (1) inconsistent test administration, (2) unreliable test content, and (3) inconsistent ratings, if ratings are used. When a test is administered inconsistently to candidates, the candidates' scores are likely to reflect that inconsistent administration. It is impossible to identify whether the inconsistent administration caused those differences in candidates' test scores or the test content caused those differences. This invalidates the test.

It is also well known that when test content is unreliable, the test cannot be valid. For example, when the wording of a question is indecipherable, it is unreliable. If the responder cannot interpret the question correctly, then the response is unreliable and therefore invalid.

Often, raters score part of a promotional test. If these assessors cannot agree on a score for the candidate's performance, the test will also be considered unreliable. The rating criteria must assist assessors rate candidates consistently.

If the raters cannot do so, then the rating criteria are unreliable, and therefore the test is invalid.

Candidates generally recognize an unreliable test because inconsistent test administration and indecipherable test content are obvious. Unreliable rating criteria are not as obvious to candidates unless they receive feedback about their test performance. Typically, the feedback about their performance will cause candidates to question the quality of the rating criteria.

Fairness

Decades of research have focused on the differences in test outcomes between population subgroups, with particular emphasis on race differences in test performance. Subgroup populations that have been compared and found to differ with regard to test performance include, but are not limited to, African Americans, Asians, Hispanics, and Whites (Cullen et al. 2004; Neisser et al. 1996; Nguyen 2003; Roth et al. 2001; Rushton and Jensen 2005; Schmidt and Hunter 1998). The extent of the gap between African American and White cognitive ability test performance has been shown to be between 0.5 and 1.5 standard deviations (e.g., Hunter and Hunter 1984; Rushton and Jensen 2005), with White test takers scoring higher than African American test takers. Research examining African American-White test performance differences seems to have dominated the research and continues to be a dominant concern.

Some researchers consider this consistent group difference evidence of a difference in intelligence levels between subgroups. This intelligence is also referred to in the research literature as *g*, shorthand for general mental ability (Gottfredson 1997). In military studies, path models of training and job proficiency have shown that *g* strongly predicts success in training and acquiring job knowledge, which in turn strongly predicts task proficiency (Ree et al. 1995). Studies throughout the decades, that include both nonmilitary and military samples, have shown that cognitive ability test performance, driven by *g*, is one of the strongest predictors of job performance and training outcomes (Schmidt and Hunter 1998).

Researchers continually examine whether cognitive ability test performance is a valid predictor of job performance. The research repeatedly finds that cognitive ability tests are valid predictors of training and job performance with some exceptions. For example, Campbell (1996) states that there is no substantive definition of intelligence which becomes problematic as we attempt to measure it. This means, because it is so difficult to define *intelligence*, it may be difficult to reliably measure it.

Campbell (1996) says that achievement-based knowledge may predict individual differences in management performance better than measures of intelligence. Intuitively, this makes sense in that testing a more specific content domain (such as promotional testing) would help determine differences in job performance better than testing a content domain that may be too general (such as cognitive ability). “If test scores represent the relative mastery of a particular domain of response capabilities, then a ‘score’ must be a function of the amount of exposure to the domain, the amount of attention paid to it, the motivational forces that control attention, the amount and kind of practice, and the retention of the specified response capabilities” (Campbell 1996, p. 131).

Regarding fairness, the findings, in general, are that cognitive ability tests do not differentially predict African American and White test performance (e.g., Neisser et al. 1996; Rushton and Jensen 2005; Schmidt and Hunter 1998, p. 272). This has led to the conclusion that tests are not culturally biased. Campbell (1996) wonders if the importance is not whether tests and criteria are culturally biased but whether bias is appropriate. In other words, should observed group differences be accepted as fact?

The concepts of cultural bias and differential prediction are two definitions of fairness, but not everyone agrees. For example, some individuals define fairness as a test that produces equal, or near equal, subgroup outcomes. This philosophical debate occurs during test litigation and in informal settings more so than in the scientific literature. In the scientific literature, the preponderance of the research concludes that cognitive ability is a valid and fair predictor of job performance.

In sum, the scientific evidence concludes that tests, in general, produce different subgroup outcomes even though they do not contain cultural bias nor do they differentially predict subgroup job performance. However, many in American society have not accepted these conclusions. As a result, plaintiffs have brought employment discrimination lawsuits, litigation has ensued, and laws have been enacted.

Why Do Subgroup Differences Exist? Another Explanation

Current researchers who study subgroup differences in test performance are focusing on differences in education and socioeconomic status that occur well before test takers enter the workforce (Campbell 1996; Campbell et al. 2001). "If two groups differ in mean IQ, culture-only theorists conjecture either that the lower scoring group has been exposed to one or more deleterious experiences or been deprived of some beneficial environmental stimuli or that the tests are not valid measures of their true ability" (Rushton and Jensen 2005, p. 240).

What cultural influences might cause group disparities in test performance? There are a host of environmental factors including, but not limited to, the type and quantity of tests individuals were exposed to during primary education and their formative years. For example, one might hypothesize that better school systems utilize greater numbers of tests, greater number of source material for those tests (e.g., textbooks), a greater amount of formal test preparation training, more extensive instruction, and that the tests, textbooks, formal test preparation training, and instruction are of higher quality.

If subgroup test performance is based upon early life cultural influences, then test professionals have a potential method to build systems to compensate for what was not provided in early life. According to Ryan (2001), test perceptions (stemming from environmental influences) include anxiety, self-efficacy, motivation, belief in tests, and perceptions of a specific test's fairness and job relatedness. Ryan (2001) noted that there has been a recent "focus on how differences between African Americans and whites may be a function of the way they approach the cognitive ability testing situation itself" (Ryan 2001, p. 45). Any examination involving moderators of the race-employment test performance relationship implies that if the moderators of that relationship can be controlled, group differences should disappear.

Some common moderators, discussed next, are test-taking skills, test motivation, test anxiety, and stereotype threat.

Test-Taking Skills

Test preparation training assumes that certain population subgroups were improperly or incompletely prepared to take tests during their primary education. Test preparation can influence test-taking skills which incorporate test-wiseness, practice, and preparation. Test-wiseness includes the ability to utilize information within the test to assist the individual in identifying the correct answers, independent of the construct being measured (Sarnacki 1979). An example of this would include the ability to eliminate alternatives in order to identify the only correct alternative, or understanding the level of detail that must be displayed in order to increase one's test score (e.g., in a writing test or in an oral communication test). Test-taking preparation includes organizing, learning, and memorizing the study material. Clause et al. (2001) found that differences in the effectiveness of study methods and learning processes could explain differences in individual test performance. Test-taking practice is repeated exposure to the type of test questions that the individual will respond to in the test. Test-taking practice should positively affect test-wiseness and test perceptions. However, "mere experience in testing does not guarantee future success on tests" (Sarnacki 1979, p. 264).

Hough, et al. (2001) reviewed the research on test preparation courses (test orientation sessions, test coaching), and concluded that test preparation courses have mixed results in improving cognitive ability test performance. The major problems with this line of research are:

- That, in each case, the test preparation course is not clearly described;
- Test preparation participation is not mandatory meaning only those who are motivated to do well, or perhaps high performers in any case, attend, and
- No control groups to compare performance of those who attended test preparation to those who did not.

There is a clear pattern in law enforcement that better test performers attend test preparation classes and seek out test-related feedback more than those that do not perform well. To address this issue, more departments are making test preparation courses a mandatory part of the promotional process. This mandatory requirement has the added benefit of increasing the transparency of the promotional process for all candidates, thus decreasing the chance of protests about "not being aware" of some part of the process. It also permits challenges in advance of the process that can be addressed before a problem multiplies.

Test Motivation

Test motivation research has focused on whether test takers' motivation influences their test performance. Hough, et al. (2001) summarized research comparing subgroup test perceptions and test motivation. A summary of their research concludes that differences in test attitudes do in fact explain subgroup test score differences. Clause et al. (2001) found that individual test-taking motivation and self-efficacy impacted an individual's meta-cognitive preparation activity. In essence, those who were most motivated and most aware of their own strengths and weaknesses as they relate to test content were most likely to structure their test preparation time to maximize their own test performance.

The most common motivation-test ability interaction stems from Vroom's (1964) expectancy theory, which indicates that, when motivation is low, both high and low ability individuals demonstrate similar low levels of performance. When motivation is high, however, variability in performance due to individual differences is more apparent (Kanfer and Ackerman 1989). Overall, test-taking motivation can be divided into two general categories: (1) the motivation to succeed on the test based upon the expectation that it will be instrumental in achieving desired outcomes (e.g., selection), and (2) the motivation to exert effort throughout the entire testing program.

Several studies have observed that motivation affects individual performance in several ways (Kanfer and Ackerman 1989). Ployhart and Ehrhart (2002) found that differences in test-taking motivation across groups could be reduced through development of more face valid tests. An individual who is unmotivated to perform in the target position may also be unmotivated to extend the effort necessary to study and learn the testing (selection) process and associated materials. Furthermore, test takers who do not believe they can succeed in a testing process may lack the motivation to succeed. Those who have witnessed numerous testing program failures may not be motivated to succeed in a testing program since they believe the test results are likely to be unusable. This lack of motivation will then negatively affect test takers' performance on the test, or their desire to exert effort from beginning to end of a long testing process.

Sanchez et al. (2000) found a positive relationship between expectancy and actual test performance, indicating the motivational influence of expected outcomes and test performance. Additionally, when compared to a pretest, final test-taking motivation is different among participants from different subgroups even after controlling for race and performance on the pretest (Chan et al. 1997). Test preparation has been shown to increase test motivation indicating that preparation is the mechanism through which motivation affects test performance (Clause et al. 2001).

Test Anxiety

Test anxiety research focuses on a priori test anxiety that individuals bring to a test situation and that which candidates experience while taking their tests. A priori test anxiety can be attributed to fear of the unknown (the test format/content) and fear about test performance. Tobias (1985) examined anxiety that candidates experience while taking the test, specifically the extent to which test anxiety interferes with retrieval of prior learning. More specifically, fill-in the blank tests (such as assessment center exercises), are more of a problem for test-anxious students (Benjamin et al. 1981). When test takers were trained in test-taking skills, their anxiety levels were reduced, resulting in higher performance (Tobias 1985).

Stereotype Threat

Another topic of research in subgroup test performance disparities has been stereotype threat. Stereotype threat has been shown to partially explain minority test performance and cognitive ability. If stereotype threat is low, the test performance of minority applicants improves (McKay et al. 2003). Steele and Aronson (1995) hypothesize that this stereotype threat may interfere with intellectual functioning. Additionally, race differences in cognitive ability tests were significantly lowered (Brown and Day 2006) when stereotype threat was low. It is important to note that stereotype threat is not theorized to be responsible for the gap, but it is believed to widen existing gaps in test performance (Nguyen et al. 2003).

Management Concerns

Management's concerns about promotional processes include identifying the most qualified candidates, the indirect and direct cost of giving them, and the number of complaints, grievances, appeals, and/or litigation that result. They typically want the most qualified candidates to succeed in the promotional process, and they want a diverse candidate pool from whom to choose. They are frequently concerned with the cost of testing in terms of direct dollars and human capital time. In some cases, as discussed earlier, they will administer lesser quality tests in order to save money and time. If test grievances and litigation abound, the department's leadership succession plan may be compromised. Grievances and litigation also negatively impact the integrity of the department and employee morale.

Why Use Promotional Processes?

As stated, the research on promotional tests has generally shown that they are valid predictors of job performance (e.g., Schmidt and Hunter 1998). For this very reason, American police departments are large buyers and users of promotional examinations. Other reasons for their reliance on promotional exams include their desire for objectivity and their need to evaluate large candidate pools. Promotional tests, in particular, meet all of these requirements.

Conclusions

It is possible to create a valid promotional process that produces a diverse candidate pool and that withstands candidates' and managements' protests. The research suggests that a comprehensive promotional testing program will include at least:

1. A test developer trained in test validation
2. A thorough test preparation session that makes the promotional process transparent and provides exposure to the test format and test preparation strategies
3. A valid test that measures a representative sample of the important knowledge, skills, and abilities (KSAs) needed for the target rank
4. Multiple methods to measure those KSAs so that all adult test takers can successfully demonstrate their KSAs (e.g., verbally, in writing)
5. Informing candidates of their scores (while maintaining their right to privacy)
6. Administrative banding of test scores
7. Detailed candidate feedback

The up-front costs of a well-designed promotional process far outweigh the costs of a poor promotional processes.

Recommended Reading and References

Arthur W, Day EA, McNelly TL, Edens PS (2003) A meta-analysis of the criterion-related validity of assessment center dimensions. Pers Psychol 56(1):125–154

Benjamin M, McKeachie W, Lin Y, Hollinger D (1981) Test anxiety: deficits in information processing. J Educ Psychol 73:816–824

Bobko P, Roth P (2004) Personnel selection with top-score-referenced banding: on the inappropriateness of current procedures. Int J Sel Assess 12:291–298

Brown RP, Day EA (2006) The difference isn't black and white: stereotype threat and the race gap on Raven's advance progressive matrices. J Appl Psychol 91(4):979–985

Campbell JP (1996) Group differences and personnel decisions: validity, fairness, and affirmative action. J Vocat Behav 49:122–158

Campbell FA, Pungello EP, Miller-Johnson S, Burchinal M, Ramey CT (2001) The development of cognitive and academic abilities: growth curves from an early childhood educational experiment. Dev Psychol 37(2):231–242

Cascio WF (1998) Applied psychology in human resource management, 5th edn. Prentice Hall, Upper Saddle River

Chan D, Schmitt N, DeShon RP (1997) Reactions to cognitive ability tests: the relationships between race, test performance, face validity perceptions, and test-taking motivation. J Appl Psychol 82(2):300–310

Civil Rights Act of 1964, 42 U.S.C. §1971 et seq. (1964), as amended in 1991 Pub. L. 102–166, (1991)

Clause CS, Delbridge K, Schmitt N, Chan D, Jennings D (2001) Test preparation activities and employment test performance. Hum Perform 14(2): 149–167

Cullen MJ, Hardison CM, Sackett PR (2004) Using SAT-Grade and ability- job performance relationships to test predictions derived from stereotype threat theory. J Appl Psychol 89(2):220–230

Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, Department of Justice (1978) Adoption of Four Agencies of Uniform Guidelines on Employee Selection Procedures, 43 Federal Register 38, 290–38, 315 (August 25, 1978)

Gottfredson LS (1997) Why *g* matters: the complexity of everyday life. Intelligence 24(1):79–132

Hough LM, Oswald FL, Ployhart RE (2001) Determinants, detection and amelioration of adverse impact in personnel selection procedures: issues, evidence and lessons learned. Int J Sel Assess 9(1–2):152–194

Hunter JE, Hunter RF (1984) Validity and utility of alternative predictors of job performance. Psychol Bull 96(1):72–98

Johnson v. City of Memphis, Case No. 00–2608, 04–2017 and Billingsley v. City of Memphis, Case No. 04–2013 (2006)

Kanfer R, Ackerman PL (1989) Motivation and cognitive abilities: an integrative/aptitude-treatment interaction approach to skill acquisition. J Appl Psychol 74(4):657–690

McKay PF, Doverspike D, Bowen-Hilton D, McKay QD (2003) The effects of demographic variables and stereotype threat on black/white differences in cognitive ability test performance. J Bus Psychol 18(1):1–14

Neisser U, Boodoo G, Bouchard TJ, Boykin AW, Brody N, Ceci SJ, Halpern DF, Loehlin JC, Perloff R, Sternberg RJ, Urbina S (1996) Intelligence: knowns and unknowns. Am Psychol 51(2):77–101

Nguyen HD, O'Neal A, Ryan AM (2003) Relating test-taking attitudes and skills and stereotype threat effects the racial gap in cognitive ability test performance. Hum Perform 16(3):261–293

Ployhart RE, Ehrhart MG (2002) Modeling the practical effects of applicant reactions: subgroup differences in test-taking motivation, test performance, and selection rates. Int J Sel Assess 10(4):258–270

Ree MJ, Carretta TR, Teachout MS (1995) Role of ability and prior job knowledge in complex training performance. J Appl Psychol 80(6):721–730

Roth PL, Bevier CA, Bobko P (2001) Ethnic group differences in cognitive ability in employment and education settings: a meta-analysis. Pers Psychol 54(2):297–330

Rushton JP, Jensen AR (2005) Thirty years of research on race differences in cognitive ability. Psychol Publ Policy Law 11(2):235–294

Ryan AM (2001) Explaining the black/white test score gap: the role of test perceptions. Hum Perform 14:45–75

Sackett PR, Schmitt N, Ellingson JE, Kabin MB (2001) High-stakes testing in employment, credentialing, and higher education: prospects in a post-affirmative-action world. Am Psychol 56(4):302–318

Sanchez RJ, Truxillo DM, Bauer TN (2000) Development and examination of an expectancy-based measure of test-taking motivation. J Appl Psychol 85:739–750

Sarnacki RE (1979) An examination of test-wiseness in the cognitive test domain. Rev Educ Res 49(2):252–279

Schmidt FL, Hunter JE (1998) The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 85 years of research findings. Psychol Bull 124(2):262–274

Society for Industrial and Organizational Psychology Inc (2003) Principles for the validation and use of personnel selection procedures, 3rd edn. Author, College Park

Steele CM, Aronson J (1995) Stereotype threat and the intellectual test performance of African-Americans. J Pers Soc Psychol 69:797–811

Tobias S (1985) Test anxiety: interference, defective skills, and cognitive capacity. Educ Psychol 20(3):135–142

Vroom VH (1964) Work and motivation. Wiley, New York

Police School Services

Justin Turner[1] and David May[2]
[1]Old Dominion University, Norfolk, VA, USA
[2]Mississippi State University, MS, USA

Synonyms

School police; School resource officers

Overview

Since the time when formal policing began, police officers have provided services to communities in a number of contexts. Components of the police officer role have included traffic control, crime investigation and apprehension, social service referrals, and a wide variety of other tasks that prospective officers do not normally consider when contemplating employment in policing. In the early 1970s, the role of the police took on a new, highly visible context when police begin working regularly in schools as Drug and Alcohol Reduction Education (D.A.R.E.) teachers. This move, coupled with existing relationships between some school districts and police departments that had assigned police officers to patrol the school setting, began a relationship that has blossomed into a field that has now become known as police school services.

The most narrow (and perhaps most commonly used) definition of police school services would define this concept as placement of sworn police officers in school environments to assist with (1) investigation and apprehension of criminal offenders in the school setting and (2) delivery of law-related educational programs to reduce student involvement in crime. Nevertheless, this definition would neglect a wide variety of services that are provided by law enforcement to students in the school setting. First, and perhaps most importantly, many officers providing supervision in police roles in school settings are not sworn police officers. Although these officers do not have official arrest powers, they are an important component of police school services. Second, police also provide coverage for all schools in their jurisdictions, whether they have an officer formally "assigned" to the school or not. These officers often serve as members of safety planning committees, volunteer coaches, and in a wide variety of other roles that assist the students, staff, and teachers in the school setting.

As Brown (2006) states, it is difficult to determine when police school services actually began. Although some trace the origin of school resource officers (SROs) (and thus the beginning of formal police school services) to Flint, Michigan, in the 1950s, there is evidence that police officers were in schools in the United States several decades before that. Police school services are not only present in the United States; police officers are in schools in the United Kingdom and South Korea and other countries throughout the world. Given that the most extensive evidence about school police services is limited to the United States, the following discussion draws primarily from that literature.

Fundamentals

Although police school services vary by state, by county, and often by school, the three most prominent police school services are school resource officers (SROs), D.A.R.E. officers, and officers that deliver Gang Resistance and Education and Avoidance Training (GREAT officers). Each of these programs is discussed in detail below.

SROs

An SRO is officially defined by Part Q of Title I of the Omnibus Crime Control and Safe Streets Act of 1968 and amended in 1998 as a "career law enforcement officer, with sworn authority, deployed in community-oriented policing, and assigned by the employing police department or agency to work in collaboration with school and community-based organizations" (United States Department of Justice 1999, p. 1) although the term is often applied to any sworn law enforcement officer working in an educational environment. SROs are thus sworn officers whose primary function is law enforcement in schools. Nevertheless, this is not the SRO's only function.

The image of the "typical" police officer is a uniformed and armed individual. The SRO is not the typical police officer; they work in a unique and unusual environment. In fact, the National Association of School Resource Officers (NASRO) recommends that, in addition to the role of law enforcement, SROs should also (a) act as a liaison between the school, the police, and the community; (b) teach law-related education classes; and (c) counsel students (United States Department of Justice 2001).

Although there is a National Association of School Resource Officers (NASRO) with a number of state chapters, it is difficult to accurately estimate the prevalence of SROs assigned full- and part-time to work in schools. The 2004 census of law enforcement agencies reported that 3,517 officers were deployed in public school districts (Reaves 2007, p. 7), although the National Association of School Resource Officers website reported that there were 6,000 officers who were members of that organization (NASRO 2010). In a recent report, Wald and Thurau (2010) observed that there were an estimated 17,000 SROs within the United States (p. 1).

SRO programs have existed for over 60 years. In 1948, the Los Angeles School Police Department was started. Their website states it was created "to create a safe and tranquil environment for the students, teachers and staff of the Los Angeles Unified School District" (Los Angeles School Police Department 2012). This created a school police department that is based in a school district. Other law enforcement scholars have the first SRO program originating in Flint, Michigan, in the 1950s (Brown 2006, p. 592). As the LASPD website has shown, there were SROs around of different kinds for many, many years. Despite the debate around the origin of the SRO program, there is no debate that the presence of SROs in schools in the United States increased after the school shootings in the 1990s (Theriot 2009), and the Community Oriented Policing Services (COPS) funding that began in 1994 created a number of SRO positions throughout the United States. In 2006, the National Association of School Resource Officers reported that the school resource officer was the fastest growing area of law enforcement (Theriot 2009).

School resource officers are trained to work in their primary role of law enforcement officer, and they often answer directly to their agency (municipal police chief or county sheriff) chain of command (Brown 2006, p. 591). They work in a school building filled with teachers, kids, and administrators. Brown states that school resource officers "serve a multifaceted role which incorporates the duties of law enforcement officer, counselor, teacher, and liaison between law enforcement, schools, families, and the community" (Brown 2006, p. 593). As referenced above, SROs are commonly said to follow the triad model. The triad model encompasses three areas: law enforcement, teaching, and advising/mentoring (Brown 2006; Finn 2005; NASRO 2010).

In general, most SROs engage in at least some activities that fall under each part of the triad. They may partake in various kinds of law enforcement activities, from patrolling school facilities to issuing citations. They may advise, mentor, and counsel school staff, students, or families. SROs often teach students about drugs, legal issues, safety education, crime awareness, and conflict resolution. Nevertheless, some evidence suggests that SROs may not be fulfilling all the elements of that triad model equally. May and his colleagues (May et al. 2004; Ruddell and May 2011) have determined that, rather than a triad, a more accurate conceptualization of SRO duties would be a quadrant model, with two of the quadrants devoted to law enforcement (for a total of half their time dedicated to law enforcement duties) and one each to law-related counseling and law-related teaching.

Despite the large number of SROs throughout the United States, there is limited evidence examining the type of training that SROs receive and the effectiveness of that training.

Unpublished research by Wheaton and May (2012) suggests that the most common source of SRO training is NASRO. They offer Basic Training, which is a 40-h, 5-day week, and an Advanced Courses, which is a 3-day, 24-h block of training (NASRO 2010). A number of states (e.g., Kentucky, Georgia, Minnesota, and Mississippi) have developed their own basic training for SROs. Other than the description posted on the NASRO website, however, little is known about the type of training SROs receive. Nevertheless, many states also report that no formal training is offered to SROs in their state and none is required (Wheaton and May 2012).

Despite the abundance of SROs throughout the United States, there is limited research examining their effectiveness. The limited evidence examining the effectiveness of SROs suggests (1) principals at schools where SROs are assigned generally value their presence and feel they make a valuable contribution to school safety and (2) the presence of SROs at schools often leads to increases in student arrests for less serious crimes, particularly disorderly conduct. The research is mixed on how the presence of SROs impacts perceptions of students about safety and whether SROs truly make schools a safer environment (Brown 2006; May et al. 2004; Theriot 2009). These results have caused some to question the use of SROs in schools and others to argue they should be removed entirely (Justice Policy Institute 2011).

Nevertheless, Finn (2005) has highlighted four benefits of SRO programs. These benefits include (1) reduced law enforcement workload by decreasing the amount of time that patrol officers have to spend at the school by the SRO handling calls that would previously have been handled through a traditional 911 calls, (2) SROs can improve the image of law enforcement among juveniles, (3) the SRO creates a better relationship between the schools and police, and (4) the SRO can enhance the police agency's image in the community.

Although SROs continue to be a controversial addition to the school environment, both critics and advocates agree that more research is needed regarding their effectiveness in making schools safer environments. Brown (2006) suggests a number of methodological approaches to answer these questions. Until researchers uncover these answers, SROs will continue to be one of the most visible (and controversial) elements of police school services.

D.A.R.E. Officers

The largest and longest active drug prevention program developed specifically for school-age children is the Drug Abuse Resistance Education (D.A.R.E.) program (DARE Annual Report 2007). D.A.R.E. was developed in 1983 by the Los Angeles Police Department and is a police officer-led series of classroom lessons with the goal of educating children on ways to resist conforming to peer pressure by providing students tools to live a drug- and violence-free life (D.A.R.E. 2012).

Initially, D.A.R.E. was developed to focus on fifth- and sixth-grade students. The belief was that this was the age group most responsive to D.A.R.E. training, and if police officers could instill the abilities to resist pressures to use drugs of students at this age, those students would go on to lead drug-free lives (Rosenbaum et al. 1994). The D.A.R.E. program has now been expanded to cover kindergarten through 12th grade and has even expanded into educating families and communities, creating a national nonprofit organization in D.A.R.E. America.

D.A.R.E. has trained literally hundreds of thousands of police officers and has educated millions of children about drug prevention strategies in 43 countries. Each year, a day in April is set aside as National D.A.R.E. Day (Berman and Fox 2009). While the D.A.R.E. program focuses on teaching the skills needed to stay drug-free, it is different than other drug prevention programs because it utilizes specially trained police officers to teach and interact with students in a school environment. This provides an opportunity for the D.A.R.E. officer to develop a relationship with his/her students that goes beyond the normal crime and punishment aspect most perceived as the cops' job. Through the D.A.R.E. program, police officers become "humanized" and "young people can relate to officers as people" (dare.com). D.A.R.E. helps to foster a community policing relationship by developing an open line of communication between school, law enforcement, and families (Carter 1995).

Unlike the limited research around the effectiveness of SROs, widespread evidence indicates that D.A.R.E. does not reduce drug use among its participants (Pan and Bai 2009). Although D.A.R.E. does not reduced drug use, D.A.R.E. does appear to improve perceptions of police among the parents and students involved in the program (Lucas 2008). In determining the effectiveness of D.A.R.E. officers, not only are the children's and parents' opinions necessary but so are the perceptions of the educators. Little research has been committed to the task of exploring the importance of D.A.R.E. officers in the classroom. Donnermeyer's (1998) study of the educator's perspective of D.A.R.E. officers in 300 public schools throughout Ohio's school districts found that overall, there was a positive perception of the work that the officers were doing. More than two in three teachers and principals involved in the study found that the "graduation ceremony, the content of the curriculum, teacher/officer interaction, the student workbook, and the role playing exercises" were excellent. Nevertheless, only 20 % of the teachers and principals believed that the students' change to a more positive attitude was because of the D.A.R.E. program (Donnermeyer 1998).

D.A.R.E. requires the officers to go through a specific and detailed training program before entering a classroom. Dusenbury and colleagues (2002) found that one of the characteristics needed for a drug abuse prevention program to be successful was determining how the officer was prepared based on the training that was required of the officer. Hensley (2002) found that the specific training for D.A.R.E. officers required 80 h of instructor training in topics such as classroom management, teaching strategies, communication skills, adolescent development, drug information, and instruction of the D.A.R.E. lessons. An additional 40 h of training is geared toward teaching a high school curriculum to the officers. As a result of this training, the instructors are prepared to teach a structured curriculum consisting of hourly sessions with the officer in the classroom over the length of 17 weeks. Due to the supervisory position that is required of a D.A.R.E. officer, it becomes necessary to understand how important the role actually is. The supervisor must be able to relate to their students in a positive and supportive relationship (Kadushin 1992).

Despite the limited effectiveness of D.A.R.E. in reducing drug use, it appears that the D.A.R.E. program is in no danger of losing funding in the near future. D.A.R.E. is supported by numerous corporate, private, and government organizations, and according to the D.A.R.E. website, over 200 communities have started programs in the last 3 years (www.dare.org). In fact, D.A.R.E. is lauded for its innovative training strategies, its permanence and effect on police-child relationships in local communities, and its widespread impact across most school districts in the United States and many throughout the world (Berman and Fox 2009). Thus, despite the limited success in reducing drug use (and unlike SROs reviewed earlier), most researchers, policymakers, police officers, and local community leaders accept D.A.R.E. as an important police school service.

G.R.E.A.T. Officers

The first Gang Resistance Education and Treatment (G.R.E.A.T.) program was developed in 1991 by a law enforcement agency in Phoenix, Arizona. In response to the local gang problem in Phoenix, the initial goal of this program was to educate school-age children about gang violence. The G.R.E.A.T. program was created to emulate the D.A.R.E. program described earlier and follows a similar structure and curriculum as that of D.A.R.E. (Esbensen et al. 2011). Where the D.A.R.E. program brings in police officers to the schools to teach students about how to resist drugs and educate them on drug abuse, the G.R.E.A.T. officers deliver an antigang message within the schools. The semblance of the curriculum's structure and content to that of the D.A.R.E. program can be seen from the influx of the D.A.R.E. officers into the G.R.E.A.T. program (Winfree et al. 1999). The skills that the program intends to create are based on various life skills that are needed in order to reduce negative attitudes toward law enforcement and reduce gang involvement and subsequent gang violence and other negative ramifications of gang membership (Humphrey and Baker 1994).

Originally, the program consisted of eight lessons over a 9-week period that was targeted mainly to middle school students. Based on evaluations over the last decade, the program has been extended to a 13-session curriculum for middle schools, as well as an elementary school curriculum, a summer program, and a family training program. As the G.R.E.A.T. program has expanded, so has its coverage. What was intended to be a locally based program has now expanded nationwide, consisting of four regional training centers for law enforcement across the country (GREAT.com).

The G.R.E.A.T. program is designed to help youths avoid gangs and youth violence and delinquency. The G.R.E.A.T. program has four primary components – an elementary school curriculum, a middle school curriculum, a summer program, and a families training component. Although officers may offer up to four of the programs at the G.R.E.A.T. program sites, the required, core component of the program is the middle school curriculum, which teaches students problem-solving strategies and life skills with the end goal of helping them avoid violence, delinquency, and gang membership (Promising Practices Network on Children and Communities 2012).

Each of the two school components (elementary and middle school) is delivered by uniformed police officers in the school setting, while the summer program component and the families program component are led by facilitators trained in those components (usually uniformed officers as well). The 13-week middle school curriculum consists of 13 lessons that are 45–60 min long and are designed to produce changes in behavior and attitude through active learning, including positive-behavior rehearsal and other role-playing scenarios. The elementary school curriculum is also a skills-based curriculum designed to be delivered to fourth- and fifth-grade students. In addition to curriculum designed to reduce delinquency and gang membership, the curriculum consists of six 30–45-min lessons covering topics such as bystander intervention in bullying situations and the importance of involving adults when problems arise. The summer program extends the lessons taught to the students through the curriculum delivered in the school setting and works to enhance students' cognitive and social skills by providing alternatives to gang involvement in the summer months. The families program is designed for youths age 10–14 and their families and consists of six sessions that work in conjunction with the other curriculums to engage entire families in cooperative lessons whose goal is to improve both communication and decision-making skills among family members (Promising Practices Network on Children and Communities 2012). From its inception, G.R.E.A.T. has focused on the importance of training and program integrity (Dusenbury et al. 2003). Dusenbury et al. (2003) listed four areas that were important to fidelity for programs like G.R.E.A.T.: officer/instructor training, officer/instructor characteristics, the support the officer received from the host organizations (i.e., school and national G.R.E.A.T. program), and the integrity of the program itself (e.g., manual and curriculum). G.R.E.A.T. officers receive two different types of required training; separate trainings are designed and delivered in one- and 2-week training sessions. The variations in training are based on classroom experience. The 1-week training requires the officer to have experience in classroom management or to have a teaching certificate. The 2-week training is required for all officers that do not meet the requirements of having classroom experience.

Esbensen et al. (2011a) completed a comprehensive study of the fidelity of the G.R.E.A.T. Officer Training (G.O.T.) 2-week training period (the 80-h training). The authors studied three specific areas based on Dusenbury et al.'s (2003) four characteristics of fidelity: how the officer was prepared, the commitment the officer gave to delivering the program to students, the support that was received, and the value of the G.R.E.A.T. program. As outside observers of the training, Esbensen and colleagues found that the G.O.T. sufficiently prepared officers with the knowledge and skill to execute the G.R.E.A.T. program effectively in a classroom environment. To prepare the officers, the G.O.T. requires them to practice public speaking daily, beginning with 2- to 3-min presentations and eventually building up to a presentation of an entire G.R.E.A.T. lesson. Along with presentations, the officers went through lessons in a variety of teaching styles and attended discussions of pedagogy led by educational specialists who also showed officers examples of gang crime and gang research.

In assessing their second area of fidelity (the support that was received), the authors surveyed school personnel. In general, school personal were supportive of programs such as G.R.E.A.T. They generally felt that these types of programs could encourage kids to stay away from drugs and gangs. While most school personal supported the G.R.E.A.T. program, few continued or reinforced the education of the program after the officers left. Their evaluation found that the school personnel were still "untapped resources" and that school personnel should have a greater involvement in program delivery and implementation in the future than they had in the past.

Finally, Esbensen and colleagues (2011) also examined the quality of program delivery. In this analysis, the researchers focused on the management of time and classroom, the teacher involvement in the program, and the overall quality of the program. Officers were found to be competent in all categories. In the category of teacher involvement, Esbensen and his colleagues suggested that better coordination between teacher, student, and officer equaled better program quality.

The G.R.E.A.T. program has received a "promising" rating from at least one agency evaluating its effectiveness membership (Promising Practices Network on Children and Communities 2012). The evaluators cite two evaluations of the program that suggest that students receiving G.R.E.A.T. training were less involved in both total and minor delinquency than their counterparts who had not participated in G.R.E.A.T. Although the evaluators cite methodological weaknesses of the evaluations and accurately highlight that no significant evidence of reduced gang involvement was found in either evaluation, Esbensen et al. (2001) suggest that the G.R.E.A.T. program may have a delayed (rather than an immediate) impact on participants. Esbensen et al. (2001) determined that while there were no significant differences in outcome variables between control and treatment groups after 1 year, there were some significant differences at 1-, 2-, and 3-year follow-ups.

A more recent evaluation provides even more promising results. Esbensen and his colleagues (2011) conducted an outcome evaluation using a longitudinal panel of approximately 3,800 students from 31 middle schools in 7 cities where G.R.E.A.T. was being used. This evaluation determined that (1) the vast majority of school personnel that were familiar with G.R.E.A.T. in schools where it was being used had positive attitudes about the program and (2) the officers implementing G.R.E.A.T. did so with high program fidelity. Additionally, the evaluators also determined that G.R.E.A.T. students were significantly more likely than non-G.R.E.A.T. students to report lower rates of gang membership, more positive attitudes about police, less positive attitudes about gangs, more frequent use of refusal skills, and greater resistance to peer pressure. G.R.E.A.T. students were just as likely as non-G.R.E.A.T. students to engage in general delinquency and did not show improvements over their counterparts who were not exposed to the G.R.E.A.T. curriculum on measures of empathy, risk-seeking, and conflict resolution (Esbensen et al. 2011b).

Despite the somewhat ambiguous evidence around the effectiveness of G.R.E.A.T., particularly in reducing involvement in gangs, the support for G.R.E.A.T. continues among funding agencies, law enforcement officers delivering the program, and personnel working in schools where G.R.E.A.T. is delivered. Of the three most well-known police school services, G.R.E.A.T. has the most empirical support for its effectiveness and thus will continue to be delivered in communities throughout the United States.

Future Directions

This entry has reviewed the three most common police school services. Despite the differences in the program goals, missions, and delivery, there are a number of similarities among the programs. First, and most importantly, each of the three programs needs more rigorous evaluation to determine its effectiveness. In fact, the current evidence suggests the three programs could be placed on a continuum of effectiveness with D.A.R.E. on one end as largely ineffective (at least when it comes to reducing drug use) and G.R.E.A.T. on the other end, with the majority of the empirical evidence suggesting that the current program is meeting its goals of reducing involvement in gangs. Given the lack of rigorous evaluations, SRO programs would best be placed between the two on this continuum. A second similarity between the three programs is that, by and large, in schools where the program is being delivered, school personnel (and, to some extent, students) have positive things to say about the program. In each case, the harshest critics of the police school service programs are not within the school, but those outside the school. Finally, none of the three programs appears to be in danger of extinction in the near future. For each of the three programs, funding at the national level has been reduced in the last 5 years with minimal impact on program delivery, suggesting that local communities are so supportive of the police school services in their community that they are willing to find funding to keep those programs afloat. In fact, if anything, the future may hold further expansion of police school services, as improved relations between school administrators and law enforcement agencies may bring about further expansion of the role of law enforcement in the school environment.

Related Entries

▶ Communities and the Police
▶ Community Policing
▶ Impacts of Community-Oriented Policing
▶ Methodological Issues in Evaluating Police Performance
▶ Order Maintenance Policing
▶ Police Officers in Schools

Recommended Reading and References

Berman G, Fox A (2009) Lessons from the battle over D.A.R.E.: the complicated relationship between research and practice. Bureau of Justice Assistance and Center for Court Innovation, Washington, DC

Brown B (2006) Understanding and assessing school police officers: a conceptual and methodological comment. J Crim Justice 34:591–604

Carter D (1995) Community policing and D.A.R.E.: a practitioner's perspective. Community policing and D.A.R.E.: a practitioner's perspective. https://www.ncjrs.gov/txtfiles/comdare.txt Accessed 2 Aug 2011

D.A.R.E. (2012) Home page. http://www.dare.com/home/default.asp

D.A.R.E. America (2007) D.A.R.E. America Annual Report. Author, Inglewood

Donnermeyer JF (1998) Educator perceptions of the D.A.R.E. officer. J Alcohol Drug Educ 44:1–17

Dusenbury L, Brannigan R, Falco M, Hansen WB (2003) A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. Health Educ Res 18:237–256

Esbensen F, Osgood DW (1999) Gang resistance education and training (G.R.E.A.T.): results from the national evaluation. J Res in Crime Delinquency 36(2):194–225

Esbensen F, Osgood DW, Taylor TJ, Peterson D, Freng A (2001) How great is G.R.E.A.T.? Results from a longitudinal quasi-experimental design. Criminol Public Policy 1(1):87–118

Esbensen F, Matsuda KN, Taylor TJ, Peterson D (2011a) Multimethod strategy for assessing program fidelity: the national evaluation of the revised G.R.E.A.T. program. Eval Rev 35:14–39

Esbensen F, Peterson D, Taylor TJ, Freng A, Osgood DW, Carson DC, Matsuda KN (2011b) Evaluation and evolution of the Gang Resistance Education and Training (G.R.E.A.T.) program. J Sch Violence 10:53–70

Finn PB (2005) School resource officer programs FBI. Law Enforcement Bull 75(8):1–13

Hensley PH (2002) The value of supervision. Clin Superv 21(1):97–110

Humphrey KR, Baker PR (1994) GREAT program: gang resistance education and training. FBI Law Enforcement Bull 63:1–4
Justice Policy Institute (2011) Education under arrest. Author, Washington, DC
Los Angeles School Police Department (2012) Home page. http://www.laspd.com/
May DC, Cordner G, Fessel SB (2004) School resource officers as community police officers: fact or fiction? Law Enforcement Exec Forum 4:173–188
National Association of School Resource Officers (2012) Home Page. http://www.nasro.org/mc/page.do;jsessionid=C4FD6DE03F0295D4E58D5E0E8B18E43A.mc1?sitePageId=115971
Pan W, Bai H (2009) A multivariate approach to a meta-analytic review of the effectiveness of the D.A.R.E. program. Int J Environ Res Public Health 6:267–277
Promising Practices Network on Children and Communities (2012) Gang resistance education and training (G.R.E.A.T.). http://www.promisingpractices.net/program.asp?programid=112#findings
Reaves BA (2010) Local police departments, 2007. Bureau of Justice Statistics, Washington, DC
Rosenbaum DP, Flewelling RL, Bailey SL, Ringwalt CL, Wilkinson DL (1994) Cops in the classroom: a longitudinal evaluation of drug abuse resistance education (DARE). J Res Crime Delinquency 31:3–31
Ruddell R, May DC (2011) Challenging our perceptions of rural policing: an examination of school resource officers in rural and urban Kentucky schools Kentucky. J Anthropol Sociol 1(1):5–18
Theriot MT (2009) School resource officers and the criminalization of student behavior. J Crim Justice 37:280–287
United States Department of Justice (1999) A national assessment of school resource officer programs. National Institute of Justice, Washington, DC
United States Department of Justice (2001) COPS: on the beat. Office of Community Oriented Policing Services, Washington, DC
Wald J, Thurau L (2010) First do no harm: how educators and police can work together more effectively to keep schools safe and protect vulnerable students. The Charles Hamilton Houston Institute for Race and Justice, Cambridge, MA
West SL, O'Neal KK (2004) Project D.A.R.E. outcome effectiveness revisited. Am J Public Health 94(6):1027–1029
Wheaton A, May DC (2012) A national consideration of school resource officers' prevalence and training. Unpublished paper presented at the Academy of Criminal Justice Sciences Meeting, New York
Winfree LT, Lynskey DP, Maupin JR (1999) Developing local police and federal law enforcement partnerships: G.R.E.A.T. as a case study of policy implementation. Crim Justice Rev 24:145–168

Police Selection

Michael G. Aamodt
DCI Consulting Group, Inc., Washington, DC, USA

Overview

Over the past 30 years, great progress has been made in developing a process for selecting police officers. There are four distinct phases to the selection process: conducting a job analysis, determining which applicants have the competencies to be good officers, investigating the background of the applicant, and conducting a psychological evaluation to determine if there is something about the applicant that would be a threat to the safety of the officer or to others. Each of these four phases is conducted by professionals with very different skills and training. This entry will describe each of the four phases and summarize the research findings regarding what methods best predict police officer job performance.

Conducting a Job Analysis

The first step in developing a system to hire police officers is to conduct a job analysis. The goal of a job analysis is to identify the tasks that are performed, the conditions under which they are performed, and the competencies needed to perform the tasks. Although there are many job analysis methods, the typical job analysis begins with the job analyst observing the job being performed (*ride alongs*) and then conducting interviews with incumbents. These interviews are conducted with individual officers, groups of officers at one time, and their supervisors. When the interviews are conducted with groups of incumbents and supervisors, they are called subject-matter expert (SME) conferences.

Once the observations and interviews are concluded, the next step is to develop a task inventory and a competency inventory. The task inventory is a list of the tasks performed by the officers, and the competency inventory is a list of competencies (e.g., knowledge, skills, abilities) needed to perform the tasks. Next to each task are scales where incumbents will rate the frequency with which a task is performed as well as the importance of the task. Some tasks, such as writing traffic citations, will be rated as being important and frequently occurring, whereas others, such as shooting a gun, will be rated as occurring with low frequency but of having high importance.

Once the task and competency inventories have been developed, they are administered to a sample of police officers who will rate the frequency and importance of the tasks and competencies. These ratings are then summarized to determine which tasks and competencies are *critical* to the performance of the job.

Although the critical tasks identified in the job analysis are often similar across law enforcement agencies, the frequency with which they occur can vary tremendously. That is, while every police officer makes arrests and writes traffic citations, the frequency of those activities might be different in a large city with a high crime rate than it would in a small rural town.

Developing the Test Battery

Now that the critical tasks and competencies have been identified, the next step is to decide how to measure whether an applicant has the necessary competencies to perform the essential tasks. Some competencies, such as knowledge of the law, will be taught in the academy and thus will not be a part of the selection system. For competencies that cannot be easily learned in the academy, it is essential to test for them prior to hiring the officer. Although the term *test* often conjures up the image of a paper-and-pencil test, psychologists and the courts use the term to describe any technique used to evaluate someone. Thus, selection methods such as interviews and background checks are considered to be tests. Common tests used in police selection include minimum qualifications, interviews, cognitive ability tests, personality inventories, and physical ability tests.

Minimum Qualifications

Every law enforcement agency has minimum qualifications that an applicant must have to even be considered for the job. These minimum qualifications, often determined by state regulation, include such requirements as being 21 years of age, having at least a high school diploma/GED, possessing a valid driver's license, and having never been convicted of a felony. As an example, in 2012, the minimum requirements in Arizona are:

1. Be a United States citizen
2. Be at least 21 years of age; except that a person may attend an academy if the person will be 21 before graduating
3. Be a high school graduate or have successfully completed a General Education Development (G.E.D.) examination
4. Undergo a complete background investigation
5. Undergo a medical examination within 1 year before appointment
6. Not have been convicted of a felony or any offense that would be a felony if committed in Arizona
7. Not have been dishonorably discharged from the United States Armed Forces
8. Not have been previously denied certified status, have certified status revoked, or have current certified status suspended
9. Not have illegally sold, produced, cultivated, or transported for sale marijuana
10. Not have illegally used marijuana for any purpose within the past 3 years
11. Not have ever illegally used marijuana other than for experimentation
12. Not have ever illegally used marijuana while employed or appointed as a peace officer
13. Not have illegally sold, produced, cultivated, or transported for sale a dangerous drug or narcotic
14. Not have illegally used a dangerous drug or narcotic, other than marijuana, for any purpose within the past 7 years

15. Not have ever illegally used a dangerous drug or narcotic other than for experimentation
16. Not have ever illegally used a dangerous drug or narcotic while employed or appointed as a peace officer
17. Not have a pattern of abuse of prescription medication
18. Undergo a polygraph examination
19. Not have been convicted of or adjudged to have violated traffic regulations governing the movement of vehicles with a frequency within the past 3 years that indicates a disrespect for traffic laws or a disregard for the safety of other persons on the highway
20. Read the code of ethics and affirm by signature the person's understanding of and agreement to abide by the code

Interviews

Interviews used in police selection will differ in structure and style. The structure of an interview is determined by the source of the questions, the extent to which all applicants are asked the same questions, and the structure of the system used to score the answers. A *structured interview* is one in which (1) the source of the questions is a job analysis (job-related questions), (2) all applicants are asked the same questions, and (3) there is a standardized scoring key to evaluate each answer.

An *unstructured interview* is one in which interviewers are free to ask anything they want (e.g., Where do you want to be in 5 years? What was the last book you read?), are not required to be consistent in what they ask of each applicant, and may assign numbers of points at their own discretion. Interviews vary in their structure, and rather than calling interviews structured or unstructured, it might make more sense to use terms such as highly structured (all three criteria are met), moderately structured (two criteria are met), slightly structured (one criterion is met), and unstructured (none of the three criteria are met). The research is clear that highly structured interviews are more reliable and valid than interviews with less structure (Huffcutt and Arthur 1994).

The style of an interview is determined by the number of interviewees and number of interviewers. *One-on-one interviews* involve one interviewer interviewing one applicant. *Serial interviews* involve a series of single interviews. For example, the lieutenant might interview an applicant at 9:00 a.m., the captain interviews the applicant at 10:00 a.m., and the chief interviews the applicant at 11:00 a.m. *Return* interviews are similar to serial interviews with the difference being a passing of time between the first and subsequent interview. For example, an applicant might be interviewed by the HR manager and then brought back a week later to interview with the chief. *Panel interviews* have multiple interviewers asking questions and evaluating answers of the same applicant at the same time. Panel interviews are the most commonly used interviews in police selection.

Questions asked in job interviews for law enforcement positions often focus on two major areas: the applicant's background and their situational judgment. Background questions might focus on whether the applicant has used illegal drugs, why they want to be a police officer, and the extent to which they have engaged in violent behavior (e.g., fights at school, yelling at a girl/boyfriend). Situational judgment questions ask the applicant how he or she would handle a hypothetical situation. For example, an applicant might be asked, *If you saw a fellow officer accept a bribe, what would you do?* or *If a motorist would not roll down their window during a traffic stop, what would you do?*

Interview questions often vary tremendously across law enforcement agencies. For example, questions asked of applicants for the Michigan State Police focus on building trust, adaptability, decision making, work standards, initiating action, stress tolerance, continuous learning, customer focus, communication, and job fit, whereas questions asked in one New Mexico agency covered such topics as history, current events, and hobbies.

Cognitive Ability Tests

In one form or another, cognitive ability tests are commonly used in law enforcement selection. This category of tests includes a wide variety of tests ranging from those tapping general intelligence to those tapping such specific aspects of cognitive ability as reading, math, vocabulary, and logic. From a content validity perspective, cognitive ability tests are thought to be important in law enforcement selection as they are related to the ability to perform such tasks as learning and understanding new information, writing reports, making mathematical calculations during investigations, and solving problems.

Cognitive ability tests can be placed into four categories on the basis of where they were developed and the extent to which they are commercially available: Publisher developed general cognitive ability tests, nationally developed law enforcement tests, tests developed by the federal government, and locally developed civil service exams.

Publisher Developed General Cognitive Ability Tests. This first category includes cognitive ability tests developed by national test publishers. These tests are available for a wide variety of uses and can be purchased directly from the publishers. Though these tests were not specifically designed for law enforcement selection, many are certainly compatible with constructs related to law enforcement performance. Tests from this category that are commonly used in law enforcement selection include the Wonderlic Personnel Test, Nelson-Denny Reading Test, Shipley Institute of Living Scale, and the Wechsler Adult Intelligence Scale.

Nationally Developed Law Enforcement Tests. The second category of cognitive ability tests includes those developed by consultants or trade organizations for specific use with law enforcement agencies. Examples of tests from this category include the Entry-Level Police Officer Test, National Police Officer Selection Test (POST), and the Law Enforcement Candidate Record (LECR). The Entry-Level Police Officer Test, published by the International Public Management Association for Human Resources (a trade organization), is a 100-item test that measures observation and memory, ability to learn police material, verbal and reading comprehension, and situational judgment and problem solving. It also contains a noncognitive component that measures interest in policing as a career. In contrast, the POST, published by Stanard and Associates (a private company), is a 75-item test that measures arithmetic, reading comprehension, grammar, and incident report writing.

Tests Developed by the Federal Government. The third category of tests was developed by the federal government for use either with the military or with general employment testing. The major test in this category used in law enforcement selection was the Army General Classification Test. Although these tests were commonly used for police selection in the 1960s and 1970s, they are seldom used today outside of the federal government.

Locally Developed Civil Service Exams. The fourth category consists of cognitive ability tests developed by various municipalities and Civil Service Commissions for their own use. Normally, the municipality hires an outside consulting firm to develop these tests. Though it is not unusual for these tests to be shared with other agencies, they are not commercially available.

The use of cognitive ability tests in law enforcement selection is controversial. Proponents of these tests cite an abundance of research demonstrating that cognitive ability tests are excellent predictors of performance in a wide variety of jobs (Schmidt and Hunter 1998) as well as good predictors of how a police cadet will perform in the academy (Aamodt 2004). Opponents argue that cognitive ability tests result in adverse impact against Blacks and Hispanics (Roth et al. 2001) and that there are alternative methods with comparable validity but result in lower levels of adverse impact.

Because of the high levels of adverse impact that occur with cognitive ability tests, it is essential that a law enforcement agency conducts a validation study to ensure that the people who score highly on these tests actually perform better on the job than those that do not. In the typical validation study, the test is administered to newly hired police officers, and their scores are correlated with such criteria as academy grades, supervisor ratings of on-the-job performance, number of commendations received, and number of disciplinary incidents.

In recent years, applicants have challenged the legality of cognitive ability tests based on the score needed to pass the test. That is, should the applicants with the highest scores always be given preference over people with lower scores or should the law enforcement agency set a passing score (e.g., 70 %) such that anyone who achieves that score is considered equally qualified for the job? Most law enforcement agencies have adopted the passing score approach.

Personality Inventories

Personality inventories are commonly used methods to predict performance in law enforcement settings (Weiss 2010). Although there are hundreds of personality inventories available, they generally fall into one of two categories based on their intended purpose: measures of psychopathology and measures of normal personality.

Measures of Psychopathology. Measures of psychopathology (abnormal behavior) determine if individuals have serious psychological problems such as depression, bipolar disorder, and schizophrenia. Commonly used measures of psychopathology used in law enforcement research include the Minnesota Multiphasic Personality Inventory 2 (MMPI; MMPI-2), Millon Multiaxial Clinical Inventory (MMCI-III), Personality Assessment Inventory (PAI), and the Clinical Assessment Questionnaire (CAQ). Such measures are designed to *screen out* applicants who have psychological problems that would cause performance or discipline problems on the job. Measures of psychopathology are not designed to *select in* applicants and are seldom predictive of job performance (Aamodt 2004, 2010).

Measures of psychopathology are considered to be medical exams. As a result, to be in compliance with the Americans with Disabilities Act (ADA), they can only be administered after a conditional offer of employment has been made. Prior to the passage of the ADA in 1990, these tests were routinely administered to all applicants.

Measures of Normal Personality. Tests of normal personality measure the traits exhibited by normal individuals in everyday life. Examples of such traits are extraversion, shyness, assertiveness, and friendliness. Though there is some disagreement, psychologists today generally agree there are five main personality dimensions. Popularly known as the Big Five, these dimensions are openness to experience (bright, adaptable, inquisitive), conscientiousness (reliable, dependable, rule oriented), extraversion (outgoing, friendly, talkative), agreeableness (works well with others, loyal), and emotional stability (calm, not anxious or tense). Commonly used measures of normal personality in law enforcement selection include the California Psychological Inventory (CPI), M-PULSE, 16PF, NEO PR-R, Inwald Personality Inventory (IPI), and Edwards Personal Preference Schedule (EPPS). In contrast to measures of psychopathology, measures of normal personality are used to *select in* rather than *screen out* applicants.

Though personality inventories are commonly used in police selection, in part because racial groups tend to score similarly (Foldes et al. 2008), a meta-analysis of the validity of personality inventories indicates that, in general, they are not good predictors of police performance (Aamodt 2004, 2010).

Physical Ability Tests

Most law enforcement agencies require applicants to pass a physical ability test either prior to hire or at the completion of the academy. These physical ability tests usually come in one of two formats. In the first format, applicants perform a variety of exercises related to stamina and strength. Such exercises often include sprints, push-ups, and sit-ups. In the second format, applicants perform a job-related simulation involving physical agility. Such a simulation might involve getting out of a car, running a short distance, leaping over an obstacle, climbing a fence, going through a window, and dry firing a pistol at the end of the obstacle course.

Because physical ability tests often result in adverse impact against women, it is essential that law enforcement agencies establish the validity of their physical ability tests. Best practices include using simulations rather than individual tests (e.g., push-ups) and using cutoff scores that are realistic and appropriate for the work being performed.

Below is an example of the physical ability test used by the Oakland (CA) Police Department:

- *Cone Maze*: While running through the cone maze, you may not knock over or move any of the cones from its original position. If you do, you will be asked to put the cone back to its original position. The additional time you use to do so will be counted toward your overall qualifying time.
- *Fence Climb*: You will have three (3) chances to climb over the 5-ft fence. If you fail to go over the fence after three attempts, the proctor will stop the test and give you a failing grade. While climbing the fence, you may not use the support on either side of the fence to assist you.
- *Ditch Jump*: You will have three (3) chances to jump over the simulated 4-ft ditch (rubber mats). Your foot must not come in contact with the ditch area at anytime. If you fail to complete this event after three attempts, the proctor will stop the test and give you a failing grade.
- *Stair Climb/Window Entry*: When climbing the stairs and going through the window, you must step on each stair on both the front and backside of the simulated window frame. You may use any part of the window frame to brace yourself in order to help facilitate your climb through the window frame.
- *Dummy Drag*: In order to pass this event, you must successfully drag the dummy around the designated cone and drop it after it crosses over the black tape marker. Then you will run as quickly as you can to the handcuff simulator platform. Your timed run will end when you touch the handcuff simulation platform.
- *Handcuffing Simulator*: This is the last event of the physical ability test. You are allowed three (3) attempts to complete. You must grasp each end of the bar and bend the bar until both ends touch. You must hold the bar in this position for 30 s. You may not interlock your fingers and your hands must not touch your chest at any time. If you let the ends of the bar separate, even slightly, your time will be restarted and you must begin the event again. Remember: The time begins when the ends of the bar touch.

The applicant must complete the first five events within 2 min and 35 s. To pass the handcuff simulation, the applicant must perform the activity for 30 s while maintaining a proper posture.

Background Investigation

Once it has been determined that an applicant has the competencies needed to be an effective police officer, the department then conducts a background investigation to determine if there are problems that were missed in the interview and testing process. The background investigation begins with the applicants completing a lengthy set of questions regarding their education, work history, credit history, driving record, arrest and conviction record, and previous places of residence. The department will then verify whether the applicants have the degrees they claim to have earned, worked at the jobs listed in their application, and listed all arrests or convictions. From there, the investigators will often contact the applicants' former professors, employers, and neighbors to determine if the applicant acted in a way that would reflect potential problems. For example, an employer might indicate that an applicant was verbally aggressive with customers, a neighbor might mention the loud music and noise from wild parties that came from the applicant's apartment, and a professor might indicate that the applicant occasionally *seemed stoned* during her 9:00 a.m. class. The last step in the background investigation is a polygraph test in which the applicant is asked questions about such activities as drug use, theft, and aggressive acts.

At the conclusion of the background investigation, the investigator will compile all of the information and make a recommendation about the fitness of the applicant. This overall recommendation is important because seldom does a single piece of information (e.g., number of traffic citations, being fired from one job) predict how an officer will perform on the job. Instead, it is a *pattern* of problematic behaviors that is most predictive of potential problems.

Psychological and Medical Evaluation

The final step in the selection process is for the applicant to be evaluated by a licensed physician and a licensed psychologist. The physician is given a job description and asked to evaluate whether the applicant has any medical condition that would keep him or her from performing the duties of a police officer. The clinical psychologist is asked to determine if the applicant has a psychological disorder that might make them a potential danger to themselves or to others. The Americans with Disabilities Act requires that the medical exam and the psychological exam be administered after a conditional offer of hire. That is, the applicants are told that they will be hired as long as they pass the medical and psychological evaluations.

The nature of the psychological exam will depend on the needs of the department. If the department already has a comprehensive selection system such as that described in this entry, the role of the psychological exam is limited to determining whether the applicant has some sort of psychopathology that will limit his or her ability to be an effective police officer. This limited role is because the law enforcement agency has already tested and interviewed the applicant and believes that the applicant has the competencies needed to perform the job. What the agency needs from the psychologist is certification that there is not some form of psychopathology that will make the applicant a danger to themselves or to others.

If, however, the applicant has applied to a small police department that does not use extensive testing, the agency might ask the psychologist to determine not only whether the applicant has some form of psychopathology but also whether the applicant has the basic competencies (e.g., personality, cognitive ability) to effectively perform the job. In some states, such as New Mexico, an applicant can go directly to a clinical psychologist to get certified and then take that certification to the department as proof that they are ready to be hired and sent to the academy.

Though clinical psychologists will differ in how they conduct psychological evaluations, there are usually three main parts to the evaluation: assessment of background information, a written test of psychopathology (e.g., MMPI-2), and a clinical interview. The clinician will use all three components to arrive at an overall recommendation of whether the applicant is psychologically fit to be a law enforcement officer.

From these assessments, Trompetter (1998) believes that the clinical psychologist should determine whether the applicant:

1. Can accept criticism
2. Can adapt to change
3. Be assertive when appropriate
4. Be interpersonally sensitive
5. Is motivated to achieve
6. Be objective
7. Be persuasive
8. Be vigilant
9. Can conform to rules and regulations
10. Has good impulse control
11. Demonstrates integrity
12. Demonstrates practical intelligence
13. Demonstrates reliability
14. Demonstrates social concern
15. Exhibits a positive attitude
16. Can manage his/her anger
17. Can tolerate stress
18. Can work in a team

The Academy

Once the applicant has been hired, he/she will still need to complete a law enforcement academy to become certified as a police officer, sheriff's deputy, state trooper, or other law enforcement professional. Graduation from an accredited academy is mandatory in order to be certified as a peace officer. Each state sets a minimum number of training hours for an academy, but each academy can offer more extended training. The average academy lasts 761 h (19 weeks) at a cost per cadet to complete the academy of $16,100

Police Selection, Table 1 Topics learned in law enforcement academies

Topic	% of academies offering	Median hours of instruction
Operations		
Report writing	100	20
Patrol	99	40
Investigations	99	40
Basic first aid/CPR	99	24
Emergency vehicle operations	97	40
Computers/ information systems	58	8
Weapons/self defense		
Self-defense	99	51
Firearms skills	98	60
Nonlethal weapons	98	12
Legal		
Criminal law	100	36
Constitutional law	98	12
History of law enforcement	84	4
Self-improvement		
Ethics and integrity	100	8
Health and fitness	96	46
Stress prevention/ management	87	5
Basic foreign language	36	16
Community policing		
Cultural diversity/ human relations	98	11
Basic strategies	92	8
Mediation skills/ conflict management	88	8
Special topics		
Domestic violence	99	14
Juveniles	99	8
Domestic preparedness	88	8
Hate crimes/bias crimes	87	4

Source: Reaves (2006)

(Reaves 2006). Louisiana has the fewest required hours (520), and West Virginia has the most (1,582; Rojek et al. 2007). Approximately 14 % of cadets who enter the academy fail to complete the requirements needed to become a certified police officer (Reaves 2006). As shown in Table 1, there are a wide range of topics studied in the typical academy.

After completing the academy, the final step toward becoming a law enforcement professional is to complete supervised field training. Some academies include this field training as part of the academy training, whereas others have the agency that hired the cadet conduct the training. The typical department requires 520 h (13 weeks) of field training (Rojek et al. 2007).

Conclusion

As you can see from this entry, law enforcement agencies often expend tremendous effort and finances to ensure that their agencies are staffed with high-quality employees. To survive a legal challenge, each stage of the selection process must be shown to be job related.

Related Entries

- ▶ Minorities Within the Police Workforce
- ▶ Officer Safety, Health, and Wellness
- ▶ Police Promotional Practices
- ▶ Women Police

Recommended Reading and References

Aamodt MG (2004) Research in law enforcement selection. BrownWalker, Boca Raton

Aamodt MG (2010) Predicting law enforcement officer performance with personality inventories. In: Weiss PA (ed) Personality assessment in police psychology. Charles C. Thomas, Springfield

Foldes HJ, Duehr EE, Ones DS (2008) Group differences in personality: meta-analyses comparing five U.S. racial groups. Pers Psychol 61(3):579–616

Huffcutt AI, Arthur W (1994) Hunter and Hunter (1984) revisited: interview validity for entry-level jobs. J Appl Psychol 79(2):184–190

Reaves BA (2006) State and local law enforcement training academies, 2006. Bureau of Justice Statistics, Washington, DC

Rojek J, Kaminski RJ, Smith MR, Scheer C (2007) South Carolina law enforcement training survey: a national and state analysis. University of South Carolina, Columbia

Roth PL, BeVier CA, Bobko P, Switzer FS, Tyler P (2001) Ethnic group differences in cognitive ability in employment and educational settings: a meta-analysis. Pers Psychol 54(2):297–330

Schmidt FL, Hunter JE (1998) The validity and utility of selection methods in personnel psychology: practical

and theoretical implications of 85 years of research findings. Psychol Bull 124(2):262–274
Trompetter P (1998) Fitness-for-duty evaluations: what agencies can expect. Police Chief 60:97–105
Weiss PA (2010) Personality assessment in police psychology. Charles C. Thomas, Springfield

Police Self-Legitimacy

Justice Tankebe
Institute of Criminology, University of Cambridge, Cambridge, UK

Overview

In many ways, these are the best times for legitimacy in criminology. For decades, legitimacy had been peripheral in criminological research although it was a central concept in both political science and sociology. In many ways, this inattention to legitimacy was due to a general historical tendency for criminologists to avoid normative issues. The situation today is different; legitimacy is now an established topic in criminological studies, with an explosion in empirical research about its antecedents and consequences. The impetus for the remarkable change of fortune for legitimacy can undoubtedly be traced to the pioneering work of Tom Tyler, beginning with his *Why People Obey the Law*. Tyler's procedural justice arguments have been studied in different contexts, focusing principally on everyday interactions between criminal justice agents – such as prisons and police officers – and citizens. The results from these studies consistently show that public perceptions of police legitimacy shape general compliance with the law, cooperation with legal authorities, recidivism in spousal assault, and even support for vigilante self-help.

Notwithstanding this important advancement in our understanding of legitimacy, some areas of legitimacy remain underdeveloped. One of these is the police's views of their own legitimacy. To date, the dominant approach to legitimacy is from the standpoint of citizens; thus, various empirical analyses attempt to assess the conditions associated with citizens' views of the legitimacy or otherwise of the police. Yet one must equally be attentive to the belief and the confidence that officers have in their own legitimacy because such a belief can have important implications for how they carry out their work and for the cultivation of legitimacy among citizens. Max Weber, to whom we owe much for the stature of legitimacy in contemporary social science, emphasized their duality of legitimacy (see further below).

This dimension of legitimacy from the standpoint of power-holders is what has been described as *power-holder legitimacy* or *self-legitimacy* (Bottoms and Tankebe 2012). Self-legitimacy refers to a belief on that part of power-holders, such as the police, that the positions they occupy and their attendant roles are morally acceptable and justified to themselves. It is, in other words, about the self-recognition of entitlement to power. Unlike organization legitimacy, which focuses on power-holders' views of the organization, self-legitimacy is concerned with power-holders' views of their own individual legitimacy. What relationship exists, if any, between self-legitimacy and organizational legitimacy is yet to be empirically examined. This entry presents a brief theoretical analysis of self-legitimacy, hypothesizing a number of factors that might shape officers' self-legitimacy. It also discusses the consequences of police self-legitimacy.

Understanding Police Self-Legitimacy

A surfeit of definitions of legitimacy exists in the social sciences. However, many of these definitions are one-dimensional, in that they approach legitimacy from the standpoint of those subject to power; the views and beliefs of those who govern are often not taken into account. There are however a few exceptions, the most instructive of which is perhaps that offered decades ago by Dolf Sternberger: "legitimacy is the foundation of such governmental power as is exercised both with a consciousness on the government's part that it has a right to govern and with some recognition by the governed of that right" (1968: 244). There are many attractive elements in this definition, of which three are particularly noteworthy. First is its emphasis on the normative character of legitimacy that it is a positive *recognition* of the moral *right* to be in a position of power. Second, it is necessarily *conditional;* legitimacy of any power-holder or institution can be squandered or "deliberately nourished." Third, it extends both to the *power-holders* and to their *audiences*. The focus of this entry is on how power-holders, in particular police officers, legitimate their authority in their own eyes.

As with all contemporary analysis of legitimacy in the social sciences, the starting point of self-legitimacy is Max Weber. Weber argued that power-holders have a need to cultivate legitimacy of their power and positions not only for the purposes of securing the cooperation of their power-subjects but for their own personal consumption. As he put it, the powerful have a need to "persuade *themselves* that their fates are deserved and therefore rightful" (Kronman 1983: 41). In Weber's conception, self-legitimacy is a necessary precondition for claiming legitimacy among citizens. "To the extent that he anticipates and understands the criticism of those who are less fortunate, the man of good fortune must already be a critic himself" (Kronman 1983: 41).

Weber did not develop his analysis of this dimension of legitimacy as fully as he did the audience dimension. Nonetheless, the authors argue that because of the central place that Weber assigned to formal legality in his treatise on audience legitimacy, it might be reasonable to suspect that it would have been the grounds on which power-holders would seek to justify the rightness of their power to themselves. In other words, police officers will believe in their own legitimacy if and only if they ensure that the positions they occupy, the powers they wield, and the manner in which such powers are exercised on a day-to-day basis are formally and legally correct. Yet, legality is not sufficient to establish legitimacy; in addition to legality, those in power must also be able to convince themselves that their claims to legitimacy is justified in terms of a society's shared values and beliefs. Dennis Wrong has emphasized this need for reference to the shared beliefs held both by rulers and by those who are governed, when he notes that power-holders have:

> a need to believe that the power they possess is morally justified, that they are servants of a larger collective goal or system of values surpassing mere determination to perpetuate themselves in power, [and] that their exercise of power is not inescapably at odds with hallowed standards of morality. (1995: 51)

However, analysis of power-holder legitimacy in the political science literature has tended to focus almost exclusively on the ruling elite, thereby potentially missing the significant role in social order played by more junior power-holders (such as basic-grade police and prison officers), who are in direct contact with citizens and often exercise a significant degree of local power on a daily basis. Such local power-holders of course play crucial roles in all law enforcement agencies. Nonetheless, "as dominated dominators or, more precisely, as dominated parties within the field of power" (Bourdieu as cited in Frow 2000: 313), such junior power-holders are a special group. For example, patrolling police officers are "the state made flesh. [. . .] they are the most direct representatives of the state for citizens given their visible, uniformed, 24-h presence on the streets and their crucial involvement in social intervention and law enforcement" (Punch 2000: 322); but simultaneously, they can sometimes be the least powerful group within a large criminal justice agency. The decisions of police managers undoubtedly set limits and create opportunities for ordinary officers. Similarly, however, the nature and outcome of everyday police–public interactions may lead to outcomes that necessitate a recalibration of the modes of external legitimacy and a transformation of police practices and procedures. Given this, it would seem that the problem of ordinary police officers' cultivation of self-confidence in their own moral entitlement to exercise power is a question that has to be taken seriously. Indeed, for street-level officers, it might be hypothesized that the further one climbs downwards on the rungs of organizational structure, the greater the energy, time, and intensity of legitimation for the confirmation of claims to authority.

A central conceptual problem for self-legitimacy concerns the "disconnected" power-holder who has lost touch with the public he or she serves and takes an "I am always right" position (Bottoms and Tankebe 2012). Unfortunately, under certain circumstances members of police services can easily slide into this kind of attitude, as seen in, for example, the too-frequent abuse of power in policing, in so-called noble cause corruption, where officers subscribe to the view that it is appropriate to manufacture evidence against a suspect because "he is clearly guilty anyway." In such instances, officers are implicitly making claims to possession of a higher normative validity than that which the State represents; adherence to such a norm is, in their view, a necessity for society to survive. By contrast to these approaches, a healthier view of self-legitimacy asserts that "means and ends are not separate; the things we care about profoundly affect *how* we honour [them]" (Archer 2000: 84). In a criminal justice context, self-legitimacy is therefore best understood as the cultivation of self-confidence in the moral rightness of power-holders' authority, within a framework of both official laws and regulations, and societal normative expectations (Bottoms and Tankebe 2012).

As noted previously, police researchers have not yet paid sufficient attention to self-legitimacy in their empirical analyses of legitimacy. It is nonetheless important to note that the main thrust of this dimension of legitimacy is at least implicit in Muir's (1977) in-depth field on good policing, as well as an empirical description of ways in which police deviate from it. A core element of Muir's argument is that police officers are moral animals and that they wrestle with justifying their actions in moral terms. That is, they need to construct moral arguments about "the rightness of their cause," which is at the core of the idea of self-legitimacy. Muir famously identified four "types" of officers – *professionals, enforcers, reciprocators*, and *avoiders* – each with different orientations to their work but all seeking some way to justify their actions in particular and their work styles in general. Officers with a greater sense of self-legitimacy approximate Muir's professional officers:

> The professional response never involved an indefensible violation of the law. Any apparent illegality, if there was one, was always put in an understandable and acceptable light, openly and publicly justified. Nor did the professional response amount to the naked assertion of the law. The law was invoked after careful preparation of a foundation of knowledge, or fearfulness, or both. (1977: 144–145)

What Factors Promote Self-Legitimacy?

Practically and theoretically, the problem of understanding the conditions that help to create and sustain police self-legitimacy is of immense importance. However, empirically, not much is known about those conditions. Part of the reason is that this dimension of legitimacy has been undeveloped. However, it is possible to hypothesize a number of such factors that might relate to officer self-legitimacy. That is the aim of this section of the entry. Specifically, it will consider the possible role of procedural justice, relational social capital, and experiences of misconduct, and effectiveness in crime prevention.

Procedural Justice

Jack Barbalet has argued that "feelings of confidence arise from acceptance and recognition in social relationships" (2001: 87). Within this context, procedural justice becomes potentially useful in analysis of officers' feelings of confidence in their own legitimacy. Procedural justice is now an established area of criminological research. It refers to the perceived fairness of the procedures used to determine specific outcomes. In more developed work, procedural justice has itself been shown to comprise two separate elements: namely, *quality of decision-making* (e.g., did the citizen have a fair opportunity to state his/her point of view?) and *quality of treatment* (i.e., how far was the citizen treated as a person with human dignity and respect?).

There is a growing body of empirical studies that show procedural justice is the predominant determinant of citizens' perceptions of police legitimacy. Among power-holders, a study by Tyler, Callahan, and Frost (2007: 476) found that assessments of legitimacy made by both law enforcement agents and soldiers were significantly influenced by procedural justice. Tyler and his colleagues were concerned with evaluations of organizational legitimacy rather than the individual officers' self-confidence in their own legitimacy. Yet it is reasonable to hypothesize that procedural justice can be important in promoting self-legitimacy among police officers. Indeed, in his study in Ghana, Tankebe (2007) reported that officers who perceived that they had been treated with fairly by their supervisors were those who expressed greater confidence in their self-legitimacy.

Relational Social Capital

According to Coleman (1988: 101), "a group within which there is extensive trustworthiness and extensive trust is able to accomplish much more than a comparable group without that trustworthiness and trust." It is a group characterized by the so-called relational social capital, and is considered "an aid in accounting for different outcomes at the level of individual actors" (Coleman 1988: 101). The general notion of social capital is not new to criminology; there is a strong body of evidence demonstrating the role of conventional social capital suppresses offending and protects against victimizations (McCarthy et al. 2002), while criminal social capital facilitates it (Hagan and McCarthy 1997). Among police officers, the police subculture literature suggests that relational social capital among officers can be an important resource for getting police work done as much as it can facilitate misconduct among officers. The concern of the present analysis is to extend the notion of relational social capital to the legitimacy discussion by exploring its influence on officers' understandings of their own legitimacy. As previously argued, following Barbalet (2001), feelings of confidence have their foundation in social relations; thus, we expect officers' self-confidence in their legitimacy to arise from situations of high relational social capital. This is supported by Tankebe's (2007) study in Ghana, which found that officers with strained relationships with their colleagues were those who expressed lesser degrees of self-legitimacy. This is consistent with Muir's (1977) study, which emphasized the role of social capital (the attachment to and views of one's peers) to the cultivation, maintenance, and reproduction of officers' self-legitimacy.

The salience of relational social capital in promoting self-legitimacy would be even stronger in contexts in which there is less regularity in interactions between management or supervisors and frontline officers. One of the remarkable developments in policing over the last several decades is the use of technology. It has allowed police managers and supervisors to communicate to, and to access information (e.g., data on performance targets) from frontline officers without much warrant for recurrent face-to-face interactions. As Cain (1973) found in her detailed ethnographic work in England in the United Kingdom, low visibility and infrequent interactions with senior officers were the seedbed of trust and camaraderie among officers. It is thus the quality of relationships among these officers rather than the quality of interactions with "virtual supervisors" that might prove more crucial in promoting self-legitimacy among police officers.

Effectiveness in Crime Prevention

In the *Leviathan*, Hobbes argued that the raison d'être for the state and its agencies is the procurement of security and safety for citizens. This assertion of the foundational role of effectiveness to legitimacy is widely repeated by political theorists. Rothschild (1977: 488–489) puts it this way: a legal and political order, "though fully legitimate, is subject to erosion and delegitimation over the long run if it becomes chronically incompetent or ineffective" and continues that "no system stands still, forever legitimate, without being effective." Criminologists have, of course, had long-standing recognition of the momentous influence of effectiveness in reducing crime, disorder, and fear on the moral standing of criminal justice agencies. For example, the appearance of effectiveness in crime prevention was a key mode of legitimation for Robert Peel's "New Police" in Britain. More recently, survey-based studies have provided support for the effectiveness–legitimacy nexus, even if the effects of effectiveness are consistently found to be subservient to the role of procedural justice.

However, the focus of those studies has been on public views; it can be suggested that this can be extended to the question of the police's self-legitimacy. It is reasonable to expect that positive assessments of levels of safety and security in their jurisdictions will increase or strengthen self-confidence in legitimacy, but negative assessments will undermine it by engendering disillusionment among officers as they begin to cast doubt on the moral validity of their continuous claims to the exclusive exercise of power. What moral self-justification is left for legal authorities if they appear overwhelmed by insecurity, the prevention of which constitutes the rationale of their authority? In his discussion of vigilantism in Nigeria, Harnischfeger (2003: 25) reports that not only did criminals operate with confidence and impunity but also "in most cases, [the police] would run away whenever and wherever they sighted them." Anecdotal information from southeast Nigeria indicates that some officers even jettisoned the organization to join the vigilante groups that had emerged to fight the upsurge in crime. As Bauman has noted, such instances are expressive of a "tangible weakening, perplexity or sheepishness of power" (cited in Lewis 1984:15) and a diminution of the last strand of self-justification of the rightness of the authority vested in the police. Perceptions that they, or the police organization as a whole, are doing well in promoting neighborhood safety and security may therefore exert a strong influence upon a feeling self-legitimacy.

Police Misconduct

Previous studies have shown that police misconduct, including corruption, can undermine the external legitimacy of both political and legal institutions. It is therefore reasonable to argue that officers' experiences and perceptions of misconduct within the police organization can also be important for the level of confidence they express in their own legitimacy. As Goldstein (1977: 199) argues, "[a]n officer who sees the processing of hundreds of petty offenders through a city's minor courts cannot help but be struck by the futility of the procedure – the lack of justice, the lack of dignity, and the ineffectiveness of the criminal justice process [. . .]." Such cynicism can be detrimental to power-holder legitimacy in much the same way that public experiences and perceptions of corruption under perceived legitimacy of legal and political institutions.

Personal Characteristics

In his study in Ghana, Tankebe (2007) found that the personal characteristics of officers were largely unrelated to self-legitimacy. Only the rank of officers had a significant effect on confidence in self-legitimacy in Ghana; in particular, the results showed that corporals and sergeants expressed greater confidence in their legitimacy than did inspectors. It is important for future studies to examine more fully how self-legitimacy among officers differ by age, gender, rank, length of service, and ethnicity. In the countries such as the United Kingdom and the United States with ethnically diverse personnel, there is urgent need to test potential ethnic differences. This is against the backdrop of evidence suggesting that the experiences of ethnic minority officers within police forces in those countries have often been negative (Bland et al. 1999; Holdaway and Barron 1997).

Consequences of Self-Legitimacy

Why should one take self-legitimacy seriously? What practical implications does it have for everyday police work? Again, on these important questions, we currently have almost no answers. Yet it is possible to speculate on a few potentially important consequences, which might be the subject of future empirical analysis. This section focuses, in particular, upon three consequences: use of deadly force, audience or external legitimacy, and organizational commitment.

First, self-legitimacy can help to explain why some officers are more prone to the use of deadly than others. According to Archer (2003: 139), people with different identities will "evaluate the same situations quite differently and their responses will vary accordingly." If that is correct, it would seem reasonable to expect individual differences in officers' beliefs about their self-legitimacy, as well as the social and institutional context, to influence how they perceive, evaluate, and respond to situations (Bottoms and Tankebe 2012). In Muir's work, he found that unlike the "professional," the "enforcer" was "more aggressive . . . more impatient and unenlightening, unresponsive to the possible changes going on inside the citizen's head and heart" (1977: 145). Similarly, officers with differential self-legitimacy will lead to different uses of power and therefore differential quality of interactions with citizens, including the tendency to use (deadly) force. It has been hypothesized that a lack of confidence in self-legitimacy among officers can translate into indiscreet, impulsive, and unreflective handling of interactions with citizens. Bottoms and Tankebe (2012: 145) have therefore argued that, properly developed, self-legitimacy should result in "a *critical self-awareness* by police officers of the importance of the ways in which they view themselves and use power."

The second possible consequence for self-legitimacy is its implication for the successful cultivation, maintenance, and reproduction of external or audience legitimacy. Unless officers are able to cultivate belief in the moral rightness of their own legitimacy, it might be difficult for them to convince others that they are entitled to the positions they occupy and the accompanying powers. Bottoms and Tankebe (2012) have recently argued that legitimacy is best conceptualized as involving a continuous *dialogue* between power-holders and audiences, with the former (e.g., police officers) making claims, the latter (e.g., citizens) responding, power-holders then perhaps adjusting their claims, and so on. If that is correct, then one might argue that the initiator of the dialogue has to be clear on its necessity, before she can convince others to participate.

Finally, organizational commitment has been defined as "the relative strength of the individual's identification with, and involvement in, a particular organization, [. . .] characterized by three factors: (a) a strong sense of belief in, and acceptance of, the organization's goals and values; (b) a readiness to exert effort on behalf of the organization; (c) a strong desire to remain a member of the organization" (Porter et al. 1974, p. 604). Although there are various studies that examine the correlates of organizational commitment, the role of self-legitimacy has so far not been sufficiently considered. An exception is Tankebe (2010) survey-based study, based on data from Ghana. In that study, the author found that self-legitimacy had an indirect influence upon organizational commitment.

Conclusion

Previous empirical analysis of police legitimacy had approached it from two standpoints. On the one hand, there are studies that focus on the perceptions of citizens regarding the moral rightness of police exercise of power. On the other hand are studies that examine the perceptions of law enforcement personnel on the legitimacy of their organizations. This entry sought to draw attention to a third and complementary standpoint, namely, officers' perceptions of their own individual legitimacy. This dimension of legitimacy has been described as self-legitimacy or power-holder legitimacy. This is in accordance with Holmes' (1993: 39) contention that legitimacy "is not merely about mass/popular attitudes toward a regime and/or system. Nor is it merely concerned with the beliefs of the leaders and/or the staffs. It is about all of these." In focusing on this aspect of legitimacy, the object of this entry has been to extend prior knowledge on the subject.

There have been only a handful of studies that have sought to understand, empirically, the conditions that create and sustain self-legitimacy; others have looked at the consequences of self-legitimacy (Muir 1977; Tankebe 2007, 2010). Although important in seeking to extend the boundaries of legitimacy result, these studies have their own limitations. There is therefore a need for more empirical studies that examine the various hypotheses discussed above. For example, there is the need for longitudinal studies that attempt to measure changes in the levels of self-legitimacy over time and the conditions that might be antecedent to stability and change in levels of self-legitimacy through time. Second, and as noted earlier, there are prior studies on employees' perceptions of organizational legitimacy. These studies have identified procedural justice as consistent primary antecedent of perceived organizational legitimacy. To what extent might self-legitimacy relate to assessments of organizational legitimacy? There are presently no systematic attempts to address this question, but it might be worthy to investigate it because the answer can offer an additional pathway to enhance organizational legitimacy. Thirdly, might levels of self-legitimacy help to shed light on officers' readiness to use force against citizens? To accomplish this will require an obvious need for methodological diversity since survey-based data alone cannot help satisfactorily to investigate these questions. It would be necessary for future tests to use data from systematic social observations (Mastrofski et al 1998) and interviews of the kind Rydberg and Terrill (2010) employed in their investigation of the effects of educational attainment on police behavior.

Related Entries

- ► Causes of Police Legitimacy
- ► Police Corruption
- ► Police Legitimacy and Police Encounters

Recommended Reading and References

Archer MS (2000) Being human: the problem of agency. Cambridge University Press, Cambridge, UK

Archer MS (2003) Structure, agency and the internal conversation. Cambridge University Press, Cambridge

Barker R (2001) Legitimating identities: the self-presentation of rulers and subjects. Cambridge University Press, Cambridge, UK

Barker R (2003) Legitimacy, legitimation, and the European union: what crisis? In: Craig P, Rawlings R (eds) Law and administration. Oxford University Press, Oxford

Bottoms EA, Tankebe J (2012) Beyond procedural justice: a dialogic approach to legitimacy in criminal justice. J Crim Law Criminol 102:119–170

Cain M (1973) Society and the policeman's role. Routledge & Kegan Paul, London

Coicaud J-M (2002) Legitimacy and politics: a contribution to the study of political right and political responsibility. Cambridge University Press, Cambridge, UK

Coleman J (1988) Social capital in the creation of human capital. Am J Sociol 94:95–120

Crank JP (1998) Understanding police culture. Anderson, Cincinnati

Ericson RV, Haggerty KD (1997) Policing the risk society. University of Toronto Press, Toronto

Frow J (2000) Economies of Value. In: Machor JL, Goldstein P (eds) Reception study: from literary theory to cultural studies. Routledge, London, pp 294–318

Goldstein H (1977) Policing a free society. Ballinger, Cambridge, MA

Hagan J, McCarthy B (1997) Mean streets: youth crime and homelessness. Cambridge University Press, New York

Harnischfeger J (2003) The Bakassi boys: fighting crime in Nigeria. J Mod Afr Stud 41:23–49

Herbert S (1998) Police subculture reconsidered. Criminology 36(2):343–370

Herbert S (2006) Tangled up in blue: conflicting paths to police legitimacy. Theor Criminol 10:481–504

Holdaway S, Barron AM (1997) Resigners? The experience of black and Asian police officers. Macmillan, Basingstoke

Holmes L (1993) The end of communist power: anti-corruption campaigns and legitimation crisis. Polity Press, Cambridge, UK

Kronman A (1983) Max weber. Edward Arnold Publishers, London

Lewis PG (1984) Legitimation and political crises: east European developments in post stalin period. In: Lewis PG (ed) Eastern Europe: political crisis and legitimation. Croom Helm, Sydney

Mastrofski SD, Parks RB, Reiss AJ, Worden RE, DeJong C, Snipes JB (1998) Systematic observation of public police: applying field research methods to policy issues (NCJ Publication No. 172859). National Institute of Justice, Washington, DC

McCarthy B, Hagan J, Martin MJ (2002) In and out of harm's way: violent victimization and the social capital of fictive street families. Criminology 40:831–865
Muir WK (1977) Police: streetcorner politicians. University of Chicago Press, Chicago
Porter LW, Steers RM, Mowday RT, Boulian PV (1974) Organizational commitment, job satisfaction and turnover among psychiatric technicians. J Appl Psychol 59:603–609
Punch M (2000) Police corruption and its prevention. Eur J Crim Policy Res 8:301–324
Reisig MD, Bratton J, Gertz M (2007) The construct validity and refinement of process-based policing measures. Crim Justice Behav 34:1005–1028
Rothschild J (1977) Observation on legitimacy in contemporary Europe. Polit Sci Q 92(3):487–501
Rydberg J, Terrill W (2010) The effect of higher education on police behavior. Police Q 13(1):92–120
Sunshine J, Tyler T (2003) The role of procedural justice and legitimacy in shaping public support for policing. Law Soc Rev 37(3):513–348
Tankebe J (2007) Policing and legitimacy in a post-colonial democracy: a theoretical and empirical study of Ghana. Ph.D. dissertation. University of Cambridge, Cambridge, UK
Tankebe J (2009) Self-help, policing and procedural justice: vigilantism and the rule of law in Ghana. Law Soc Rev 43(2):245–269
Tankebe J (2010) Legitimation and resistance: police reform in the (un)making. In: Cheliotis LK (ed) Roots, rites and sites of resistance: the banality of good. Palgrave, London, pp 197–219
Tyler T (1990) Why people obey the law. Yale University Press, Yale
Tyler T, Callahan P, Frost J (2007) Armed, and dangerous (?): motivating rule adherence among agents of social control. Law Soc Rev 41:457
Tyler T, Schulhofer SJ, Huq A (2010) Legitimacy and deterrence effects in counter-terrorism policing: a study of Muslim American. Law Soc Rev 44:365–402
Weber M (1978) Roth G, Wittich C (eds) Economy and society: an outline of interpretive sociology, 2 vols. University of California Press, Berkeley/Los Angeles
Wrong D (1995) Power: its forms, bases and uses. Transaction Publishers, London

Police Styles

► Police Culture

Police Subculture

► Police Culture

Police Use of Firearms

Jyoti Belur
University College London, London, UK

Overview

Police forces around the world are mostly routinely armed. Even those countries, such as the UK, New Zealand, and Norway, where police officers are unarmed, exceptions exist in the form of special armed units. Police use of firearms is always under scrutiny, especially in the aftermath of a high-profile incident of shooting where the police are accused of either overusing or abusing their power to legitimately employ deadly force. Nevertheless, there is a growing demand for paramilitarization of the police in a world increasingly perceived as being at risk from terrorism and growing insecurity. Further arming of the police as a result of increasingly risky working conditions, in turn, multiplies risk for police officers themselves, as well as affecting police-public relations adversely. It engenders distrust and suspicion of the police, especially among ethnic and minority communities. The resulting public disorder and antipolice demonstrations increase the gulf between the police and the community it serves and are detrimental to police fulfilling their primary functions.

Explanations for police overuse or misuse of firearms are provided at the individual, organizational, and structural levels in the literature. Accordingly, policy recommendations for eliminating misuse or abuse of firearms are focused on better training, stricter administrative, and legal policies and increasing public awareness. The movement of a civil police force towards becoming increasingly armed and paramilitarized seems inexorable and is irreversible. Following the terrorist attacks of 9/11, even unarmed police in the UK are increasingly moving away from a "restraint paradigm," which focused on minimum use of force, to a "military paradigm," adopting a shoot-to-kill counterterrorism policy. How the changing face of the civil police will affect police-public relations is a matter for police leaders and policy makers to incorporate in their vision for policing in the future.

Police Use of Firearms

Modern policing was introduced in Britain and in the USA based on the premise that in opposition to the military, the police would intentionally be an intensely civil and democratic institution. However, while the police in Britain were routinely unarmed, paradoxically, the gun as the most overt symbol of police power became synonymous with the American style of policing (Waddington 1999). The colonial model, a distinctive militarized form of policing, where the main role of the police was to subjugate the rebellious population by force is the precursor to many of the policing models in ex-colonies (Waddington 1999). Despite the type of policing model envisaged, police work, almost universally, is characterized by features such as danger, authority, and the mandate to use coercive force that is nonnegotiable (Bittner 1975; Skolnick and Fyfe 1993). As the governmental law enforcement agency, the police see themselves as the "thin blue line" that separates anarchy from order (Skolnick 1975), and the use of violence to maintain that order is deemed legitimate and necessary. Deadly force as the most extreme form of police violence has attracted its fair share of attention from criminological scholars and the public alike. Deadly force can be employed either through the use of firearms and other lethal and nonlethal weapons or through the improper use of holds or restraining techniques, but here the focus is limited to the use of firearms.

The police use firearms in the course of their official duty, either in a situation of direct confrontation between the police and alleged criminal(s) or in a riot control or public disorder situation. The circumstantial and situational factors precipitating the use of deadly force in both these scenarios are quite different. Policing civil disorders often engenders fear, anger, frustration and heightened anxiety for police officers who are too close to the actions to be objective; additionally, heightened emotions on the part of the public make the situation more volatile. Therefore, stricter supervision, command, and control of such operations are required (Waddington 1991). Only when the crowd turns uncontrollably violent does the police resort legitimately to the use of deadly force, ideally and theoretically, in a controlled and methodical manner. In contrast, a sudden confrontation between police and "suspects" can be more fluid, with greater discretion on the part of individual officers to use firearms. Though admittedly, some riot situations may flare up, and some confrontations with criminals may be planned operations. However, principles that ought to govern the use of firearms by the police are universal and not contingent upon the situation under which it has to be employed.

Punch (2010) describes four facets of police use of firearms:

1. *Shoot to prevent* – traditional understanding of the use of firearms in the UK and the USA is to prevent or incapacitate and not to kill. Thus, the police are mainly operating under the "restraint paradigm" where the governing principle is "minimum force necessary." However, Waddington (1991) raises concerns about shooting to injure since shots intending to injure are likely to miss as limbs are more difficult to hit than the torso; a wounded person is unlikely to be incapacitated totally or immediately; and as the result of the first two, officers would be inclined to shoot at a much lower level of threat; the result often is an increase in serious injury and death.
2. *Shoot to kill* – informal cop code advises that a firearm should not be deployed unless the intention is to kill. If the intention is stopping and immediate incapacitation – and if in the process the person dies, then the officer must be prepared for that. A failure to totally incapacitate can prove costly to the police or other innocent people.
3. *Shoot to kill as murder* – police use of firearms in some countries and in some circumstances amounts to illegal executions. In some contexts, shoot to kill has come to mean summary executions, such as in Northern Ireland

(Waddington 1991), India (Belur 2010), various countries in Latin America (Chevigny 1995), and the Caribbean (Bowling 2010).

4. *Shoot to eliminate* – where the aim is not only to kill but also to shoot until all possibility of recovery or retaliation is eliminated. This is especially the case during covert counterinsurgency operations and is a clear move from the restraint paradigm to a military paradigm where the civil police operate with the intention of eliminating the target as a matter of policy.

The principle governing any use of force is that it should be necessary and reasonable. The question then arises regarding who determines what is necessary and reasonable and what is the benchmark for judging this. For example, the US Supreme Court ruled in the *Graham v. Connor* (1989) case that the use of force must be "objectively reasonable" and went on to specify that every instance of use of force will be judged to be within the constitutional limits on a case-by-case basis from the perspective of a reasonable officer on the scene rather than retrospectively on hindsight (Klinger and Brunson 2009).

Geller and Scott (1992) examine use of deadly force as a result of police shootings of civilians and shootings of police officers by civilians and other police officers in the USA. They acknowledge problems associated with police shootings, such as whether incidents occurred while an officer was on duty or off duty and whether they were officially or personally motivated. Geller and Scott (1992) suggest that the most common type of police shooting involves an on-duty, uniformed, white, male officer and an unarmed black, male civilian between the ages of 17 and 30 in a public location within a high-crime precinct at night in connection with a suspected armed robbery or a "man with a gun" call.

Researchers have found that in Western democracies race was a crucial element in shooting incidents and that black people were more likely than their white counterparts to be involved in police-related shootings (cf. Belur 2010, Chapter 2). Some research studies found this to be rooted in systematic racism. Others suggest the possibility that blacks and Hispanic minorities were disproportionately involved in violent crimes and therefore were represented in higher numbers in police shootings. They also displayed disproportionately higher unemployment, thus likely to spend their time on the streets, thereby exposing them selectively to confrontation with the police and involvement in shootings All these factors reflect wider structures of racial inequality. Other researchers observed that race was not a controlling factor in a patrol officer's decision to shoot, nor were there significant differences in the race of the victim given similar situational factors. Various findings regarding the connection between police shootings and race are highly contradictory and do not conclusively show any correlation between the two.

Fyfe (1981) describes a continuum from elective (the officer decides whether he wants to shoot or not) to nonelective shootings (where the officer has no choice but to shoot). Research shows that officers' reasons for shooting range from gun use threat to use of threat of other deadly weapons, fight without other resistance, other reasons for intentionally shooting, accidents, mistaken identity, a stray bullet, warning shots, shots to summon assistance, and felonious shootings (Geller and Scott 1992). However, the worst abuses occur when the police, impatient with the workings of the courts, take it upon themselves to dispose of suspects in bogus "shootouts" (Chevigny 1995).

Clearly, gender, race, and age of the police officer and of the "suspect" as well as the situational factors and circumstances that lead up to the incident appear to be important. Also important are structures of race and class inequality and the culture of racial antagonism that flows from this.

Explanations for Police Use of Firearms

The only legally acceptable and morally justifiable reasons for police shootings are self-defense or defense of another person's life. In some countries, it is legally permissible for the police to use firearms while arresting a dangerous and armed suspect who is avoiding arrest. The police in these countries are thus legally entitled to use force in the pursuit or arrest of a dangerous "criminal," but not in the UK or the USA, where the Supreme Court ruled in the *Tennessee v. Garner* (1985) case that the police may not use deadly force in pursuance of a fleeing felon if there is no imminent danger to a person's life.

Three main types of explanation for police violence available are individual and situational, organizational and subcultural, and sociological and structural (Green and Ward 2004). Theoretical explanations for police violence are pitched at different levels and can be classified as micro, meso, and macro theories and are interconnected.

Micro-level theories explain police violence in terms of individual officers' psychological makeup and/ or the situational exigencies in which they have to exercise their judgement. There are two types of individual theories. The "rotten apple theory" states that the attitudes and personal characteristics of some officers make them prone to use more violence than others, and it rests on the assumption that a majority of police officers are not violence prone but work within the limitations of law. The other strand of individual theories, the "fascist pigs theory," suggests that only people with certain dispositions, such as authoritarian personalities, are attracted to police work due to its nature so that a majority of officers are violence prone (Uldricks and van Mastrigt 1991). The individual theory espouses that "violence-prone" or "problem" officers manifest a *propensity* to use force and account for a majority of use of excessive force incidents (Toch 1996).

Meso-level theories explain police violence in terms of the organizational culture that encourages or tolerates such use. In turn, the organization and its culture, rules, procedures, and allegiance to rule of law and accountability affect an individual officer's attitudes and "dispositions" to act in certain ways, varying according to individual officer's personality. Organizational level theories are grounded in the explanation that the police, like other organizations, have their own subculture, which on the one hand can create a "violent" officer and on the other protect the officer's actions from external censure. Key cultural characteristics of police officers have been identified as follows: mission, action, cynicism, suspicion, machismo, isolation, solidarity, loyalty, pragmatism, and conservatism (Reiner 2000). These characteristics are core elements of the central police culture, caused by the structural features of police work such as authority, danger, and pressure for results, which gives rise to a certain "working personality" (Skolnick 1975).

The nature of the organization and the emphasis on factors such as loyalty and secrecy protect officers who are violence prone. "Closing ranks" or erecting a "blue wall of silence" are techniques by which the organization shields its officers from being under scrutiny or being prosecuted or punished by outside investigating agencies. It can also be argued that certain police organizations appear to either have an overt or a tacit policy or have specialist units that support the use of force by police officers in certain situations.

Macro-level theories, essentially sociological and structural explanations, take into account social, political, economic, and cultural factors that create circumstances that allow and legitimize police use of force. Sociological theories aim to develop understanding of the problem of police violence beyond the micro processes of individual action and organizational culture that ask the question: "Why do the police do it?" These theorists aim to answer the question "why are they allowed to do it?" by referring to a broader sociohistorical framework. Police violence tends to increase in proportion to the elite's fear of disorder, and the more fearful the elite, the more likely they are to tolerate illegal violence against potentially dangerous groups. The police use of violence as informal punishment for defying police authority or as "street justice" or vigilante justice has been reported by researchers (Skolnick and Fyfe 1993; Chevigny 1995). This penal character of police shootings is of particular concern especially when it enjoys a degree of popular support in societies where such summary justice is not considered inappropriate (Chevigny 1995; Belur 2010).

Another social trend in some Western democracies has been the growth of paramilitary policing to control social disorder. The demand for and subsequent creation of specialist elite units, whose training and culture emphasizes features of the police culture most conducive to violence: the perception of danger, fear of outsiders, isolation, secrecy, intense group loyalty, and pleasure in "warrior fantasies" (Green and Ward 2004). The increased threat of terrorism globally has meant that this trend has intensified. Historically, it has been demonstrated that measures originally justified on the basis of countering terrorism are quickly absorbed and translated into everyday policing (Green and Ward 2004).

Thus, the reasons an officer shoots or does not shoot result from the interplay of personality traits, proximal and distal situational and structural factors, and within a complex matrix of social forces such as the police department, legal context, and larger social culture (Scharf and Binder 1983).

Policing involves situations where the use of force is legitimate and necessary. However, there are other situations that are perceived as requiring the use of dirty means, including the use of force as a last resort in order to achieve a good end. Another way of approaching the problem of understanding use of force decisions is by viewing it as a "classic police dilemma" – the "Dirty Harry problem" (Klockars 1991). The Dirty Harry problem originates from the film *Dirty Harry* (1971) where Inspector Harry "Dirty Harry" Callahan is placed in a series of situations where he has to make decisions about whether "bad" means can justifiably be used to achieve "good" ends. The troublesome aspect of this problem is not whether a right choice can be made, but that any choice is between two wrongs and in choosing either, the policeman is inevitably tainted or tarnished (Klockars 1991). Thus, by choosing either to act or not act, the police officer is guilty of wrongdoing. The policeman often tries to solve the dilemma by denying the dirtiness of its means, justifying its use for achieving good means. The only way to put an end to the moral problem is by punishing officers who adopt dirty means regardless of the ends they aim to achieve, though this by no means is an easy moral choice either (Klockars 1991).

While police officers in some countries like the UK may be held individually accountable for serious events, even while they were acting under orders of their superior commanders (Punch 2010), officers in Nordic countries are exempt from criminal liability when they use firearms in the exercise of "lawful authority" (Knutsson and Noree 2010). The police are often in effect allowed to get away with blatant abuse of force. Prosecutions of officers may be rare even in cases of excessive use of deadly force in many countries, and convictions rarer still (Uldricks and van Mastrigt 1991; Geller and Scott 1992; Belur 2010). Factors responsible for low rates of culpability for police officers in many instances of excessive use of force are as follows: frustration with the criminal justice system which appears to provide a magical cloak of immunity for police officers; the relatively small number of complaints made against the police; difficulties in substantiating complaints; complete control of investigation by the police themselves; the "code" of silence that ensures officers go to great lengths to protect fellow officers; the greater credibility commonly attached to a police officer's account of events as opposed to that of an accused criminal; and jurors frequently feeling more sympathetic to an officer than the complainant and sometimes even intimidation of witnesses, lawyers, and magistrates. One of the most effective ways in which the aftermath of a controversial police shooting can be managed is a credible independent investigation, whether it is by an internal or external agency (Geller and Scott 1992).

Proper use and supervision of police use of force is essential not only to maintain state order and legitimacy but also as it affects public perception, attitude, and behavior towards the police and the government.

Police Use of Firearms and Police-Public Relations

The traditional policing model based on the restraint paradigm that emphasized containment, minimum casualties, and apprehension of suspects for criminal justice processing rather than summary execution is increasingly being replaced by a militarized version of mainstream policing based on secrecy, overwhelming force, uncompromising tactics, and military solutions post-9/11 (Squires and Kennison 2010). The fatal nature of shootings ensures their prominence in any discussion about police violence. Public discourse on shootings has focused on the strict legality of the shootings, their necessity, their moral and ethical parameters, the training of police in the use of force, and the speculation of racism in a culture of violence (Coady et al. 2000).

Police shootings can be socially divisive, threatening to damage the fabric of trust, legitimacy, and accountability which are the founding principles of civil policing (Squires and Kennison 2010). Many communities have been torn asunder following a perceived abuse of force incident resulting in urban riots and profound political unrest. The police, as custodians of the state's monopoly to legitimately use force, are entitled to use deadly force. However, as the executive arm of the law and government, their actions involving deadly force cause considerable disquiet and are under close scrutiny from the public, the media, politicians, and NGOs as well as the organization itself, at times.

Often scrutiny of police actions results in one or more of the following 5 reactions: public protests, public riots, hearings by government bodies, organizational reform, and civil litigation (Klinger and Brunson 2009). Each of these is potentially problematic for building public confidence and trust in the police. Racial disparity is often the main cause of such discontent, particularly where white officers allegedly target black or minority ethnic men disproportionately, even in the use of deadly force. These racial disparities in shooting statistics might be explained by differential involvement in criminal activity, that is, more black men are likely to commit serious crimes and are therefore more likely to be shot either by white of minority ethnic officers. Locke's (1996) examination of relevant research fails to establish an unequivocal relationship between race and police use of excessive force; however, this does not deny the racially linked outcomes for law enforcement: outcomes such as complaints and civil damage suits, which cause significant problems for police-minority community relations. Minority citizens have especially charged the police with abusing their power in the use of deadly force, an issue which has often polarized "blacks and whites, liberals and conservatives, and police officers and civilians" (Scharf and Binder 1983).

It has been recognized that public attitudes towards the police use of force can be sometimes pretty raw, often ill-informed, occasionally easily influenced, subject to personal biases and prejudices, and more likely to be fueled by sensationalist reporting than by an accurate reflection of the overall picture. Since the media pays too much attention in instances where the police have allegedly used excessive force, these comparatively rare events have nonetheless prompted a perception of police volatility, thus encouraging renewed reform of policies, training, supervision, and accountability structures (Alpert and Dunham 2004). Not all use of force incidents are deemed newsworthy and get more than a passing coverage in the media. Lawrence (2000) proposes that only when journalists find "critical story cues" do they deem a story to be newsworthy. Some of these critical cues are "accounts of family, witnesses and other sources that challenge the police or official version"; coroner's report or when a legal proceeding is initiated against the officers; availability of evidence of wrongdoing; when there is a strong citizen reaction; racial identities of the officer and citizen, especially white on black incidents; when officials make a public pronouncement or initiate special investigation; or when police policies are reformed.

It is unclear whether these critical cues reinforce media interest or whether media interest reinforces critical cues that maintain the newsworthiness of a story. Managing public perception and engaging with the media in productive ways is one of the biggest challenges facing police leaders in the aftermath of a police shooting. If handled poorly at this stage, it can have serious consequences for the organization and police officers involved.

Police-public relations are worst affected not only when the police are ill-trained and ill-disciplined to the extent that they injure or kill the innocent but also when they, due to their incompetence, fail to protect the public from dangerous offenders. So the police are damned if they do [shoot the "wrong" person] and damned if they do not [shoot the "dangerous" offender]. The trick is in the balancing act – shooting those who cannot be dealt with in any other way and protecting life in situations where the use of firearms is uncalled for. The problem is that the decisions regarding who is the right and wrong person are always made on hindsight, which can be problematic.

Public reaction to police use of firearms is ambiguous, depending upon whether the use of force is perceived as being legitimate and justified or not. There have been many early attempts to measure public attitude towards the police, asking questions in public surveys about "satisfaction" with the police or "approval" for local police and even "confidence" in the police. By and large, most of these studies have found that a majority of the public are generally satisfied with the police, but there are differences based on the characteristics of the citizens, the characteristics of the neighborhoods, and the frequency and type of citizen-police contacts (Flanagan and Vaughan 1996).

Since use of deadly force incidents are relatively rare, public attitudes towards them are reflected more in the immediate protests, or public disturbances following such an incident, than in the large-scale public opinion surveys on confidence in the police. Especially since it has long been recognized that public opinion is very sensitive to events and pressures, each survey should be taken to reflect the immediate opinion at that particular point of time. Flanagan and Vaughan (1996) cite the example of public opinion polls taken in various cities in the USA which showed that public opinion regarding police brutality and excessive use of deadly force fluctuated sharply within and between various ethnic communities depending upon whether it followed a much publicized police brutality case or not. They conclude that public opinion definitely acts as a vehicle of social control on abusive police practice and that regardless of whether public perception that the police use excessive force commonly is justified or not, there is a need to educate the public about the realities of police work and concerns about use of deadly force as well as introduce police reforms for the benefit of both.

Controlling Police Use of Firearms

Prescriptive studies on police use of deadly force offer solutions to curb the phenomenon at the individual, organizational, situational, or social level depending upon the particular theoretical approach adopted. Scharf and Binder (1983) suggest that various solutions are guided by different theories of how the interventions will work and are based on certain assumptions about police behavior. For example, those who assume that excessive shooting occurs because officers are poorly or inadequately trained advocate better training to allow officers to use finer discrimination and tactics to avoid using deadly force. Similarly, those who advocate an intensification of shooting review procedures and increased accountability assume that the police use excessive force because of low levels of sanctions in the organization and society. Thus, if scrutiny and punishments are increased, avoidable shootings will be reduced.

Reforms at the level of the individual officer include attitude and psychological tests as part of the recruitment process to ensure only the most "suitable" candidates are chosen to be able to use firearms. Better training and sensitizing programs are also part of the solution to ensure that individual officers are given an opportunity to renew their commitment to the rule of law and human rights (Grant and Grant 1996). However, Worden (1996) found little consistency between both, officers' attitudes and behaviors, and an individual officer's behavior from one incident to another. Thus, solutions that sift out "unsuitable" candidates at the recruitment stage (via psychological tests) and emphasize training and sensitizing officers though useful are not the whole solution to curbing excessive use of force.

Research that locates the source of violent behavior of officers in the police subculture and organizational ethos either put forward solutions that emphasize training and placing administrative controls to ensure no abuse of force is tolerated or suggests that real change can come about only when police culture can be changed to ensure that tolerating or encouraging deviant behavior is prohibited (Reiner 2000). Some studies concluded that introduction of administrative controls via new guidelines and procedures succeeded to a large extent by bringing down use of deadly force in New York, Kansas City, and Atlanta (Fyfe 1978; Sherman 1983). Sherman's (1983) research showed that the reduction in shooting was only in dubious cases, but shots fired in more serious, life-threatening circumstances (what might be termed as "nonelective" shootings in Fyfe's terminology) remained the same. There is, nevertheless, general agreement among policing scholars that cultural change possibilities are very limited without structural change.

Studies that focus on analyzing the situational aspect of the violent police-citizen interaction advocate solutions whereby officers have better control on the nature of the encounter and are able to process information better to enable them to take adequate precautions to ensure that the outcome of a chance encounter is not violent (Scharf and Binder 1983). Fyfe (1986) suggests that in order to reduce unnecessary violence, the police role must be defined as one of a diagnostician, and they must learn that role thoroughly as well as use principles of tactical knowledge and concealment to reduce the likelihood of having to resort to deadly force, before they actually confront someone who may be armed and dangerous. By keeping things simple, involving as few people as possible, and not getting too close, cops can avoid unnecessary shootings.

Tackling the problem of police abuse of deadly force has been attempted by developed countries like the USA, the UK, and Canada as well as developing democracies like Brazil, Argentina, and South Africa. The difference in the approach has been at the various levels at which the problem is perceived and addressed. For example, in the USA, the UK, Canada, and Australia, which have more established democratic traditions, accountability mechanisms exist, but the emphasis is on fine-tuning them or making them more refined to suit the purpose. In developing countries in Latin America or South Africa, the problems are more fundamental, of propagating a culture of democratic accountability to the rule of law and the people, of setting up proper accountability mechanisms in the first place, and also of reforming management and supervision practices to ensure greater independence from political interference and ensure more professionalism. Such reforms would require changes in the wider social context and problems facing the police.

Current Issues and Controversies

There are 4 issues that are currently the subject of research and controversy in the area of police use of firearms:

1. *Establishing a standard of "reasonableness" for use of firearms decisions*: Shooting incidents are considered suspicious either because the police fired too many shots, police accounts conflicted with each other or with the facts, or witness accounts contradicted the official version of events. Furthermore, officer decisions to shoot are generally judged on the basis of hindsight, by people with inadequate technical know-how and often under pressure from public opinion and dissatisfied minority ethnic communities.

   However, research into use of lethal force by police officers has demonstrated that perception, judgement, and reaction times of police officers in stressful situations can be hugely complex (cf. Belur 2010, Chapter 2). Testing officer reaction times under laboratory conditions demonstrated that the more stressful the situation, the slower a person is likely to respond to changes in the situation. Thus, once officers begin shooting under perceived stressful conditions, while doing many other activities simultaneously and sequentially, their reaction time to stop shooting will be longer This might explain the large number of bullets fired in some of these cases.

   Research has indicated that hindsight bias (i.e., perception of an incident is influenced by outcome knowledge) plays an important role in the way an incident is perceived by the actors involved and others observing or commenting on it. In incidents where the "victim" was not found in possession of a weapon, observers perceived less danger to the police officers on the scene, and police use of deadly force tends to be considered less justifiable under those circumstances. Research has also lent support to the contention that stressful situations such as shooting incidents can distort officer's perceptions in different ways, leading to accounts conflicting with each other and with the facts. Additionally, graphic visualization techniques and expert testimony were used to explain consistency between officer's accounts and forensic facts in particularly contentious shooting incidents. More work needs to be done in this area in order to place officer's decisions to shoot in perspective given the stressful nature of the experience and concomitant distortions in perception and decision making this can induce.

2. *Arming the unarmed police*: The British police have traditionally been unarmed, and there continues to be significant support for this position. However, Punch (2010) traces developments in arming the British police from what he calls hazardous amateurism in the early 1950s to the emergence of armed force support units in some police forces in the 1970s, to the growth of armed response vehicles nationally in the late 1980s, towards a more professional deployment of armed units, culminating in the shoot-to-kill counterterrorism policy called Operation Kratos following the London bombings of July 2005. The main reasons why police have become armed is because they have felt unsafe and vulnerable to attack in certain areas so that they have readily available means at hand to deal with armed crime and terrorism and more recently as a public reassurance exercise. The creation of SWAT teams, first by the LAPD, followed by other elite squads of armed and trained officers for special operations has become a trend in various parts of the world, including the UK, showing a move towards the militarized end of the spectrum (Waddington 1999).

   Research in Australia showed that in areas with comparable rates of violent crimes the risk of police officers being killed are significantly higher in areas where the police are armed than in those where the police are unarmed (Hawkins and Ward 1970). Another unintended consequence of paramilitarization of policing is the transfer of police issued weapons into the hands of antisocial elements. The Maoist-inspired insurgents in India have become a formidable guerrilla force to reckon with after they began arming themselves with weapons looted from government armories, state police, and security forces from the 1980s onwards (Ramana 2006).

   Given the "incident-driven" nature of the introduction of greater arming of the British police, there is a danger that a particularly horrific incident can escalate arming of the police, but there are few events that can prompt a compensating de-escalation especially given the rising threat of terrorism (Waddington 1991). The decision of arming an unarmed police force requires careful consideration as this can only be a one-way path – the experience of many countries, for example, postwar Germany and Japan, shows that disarming an armed police force remains no option once the level of societal violence has escalated to accommodate an armed police force (Phelps 2010).

3. *Use of alternative nonlethal force*: Given the pressure of performing under increasingly challenging conditions as well as from the human rights groups regarding the use of firearms, police forces are exploring alternative means of less than lethal force in order to achieve incapacitation of target without loss of life. Innovations in weaponry such as stun guns, incapacitation sprays, disorientation instruments, directed energy devices, and acoustic technologies which are being tried and tested are themselves not totally guaranteed to be nonlethal or free of controversy (Summers and Kuhns 2010). Many impact, chemical or electrical weapons have limitations in terms of accuracy, distance, or effect. They all work if used within the parameters placed upon them by their

manufacturers. However, police work involves fluidly dynamic and totally unexpected situations often precluding the manufacturer's parameters, in which case they are often useless. Less lethal technologies are still in the stage of development and testing. Police officers are often reluctant in placing their trust in such equipment to overcome someone with a gun, though the first steps are being taken. Clearly there is need for further research not only in the area of development of less lethal technology but also in their performance under a wider range of situational and circumstantial factors, as well as their impact on police-public relations more broadly. Policy analysts, police leaders, practitioners, technologists, and theorists need to work together in developing more acceptable and less lethal weapons of force fit for purpose.

4. *Terrorism and public insecurity*: The gap between the police and the military is being significantly jeopardized by terrorism (Waddington 1999). The threat of terrorism and aftermath of 9/11 and subsequent terrorist attacks in London, Madrid, Bali, and Mumbai have led to police forces in many countries adopting a military approach or a widened criminal justice model (Greene and Herzog 2009) towards terrorism by raising specialized units with training, capacity, and equipment to deal with terrorist threats of the most violent kind. The growth of special tactical groups and the resulting paramilitarization that are targeted towards countering possible terrorist activity have increased police violence in Australia (McCulloch 2000). Involvement of the civil police in covert counterterrorist activities might lead to an increase in the distrust and suspicion between the community and the police and drive a wedge in police-public relations. This has been demonstrated in Israel where shift in emphasis from low (community) policing to high (covert, antiterrorist) policing, where among other covert activities, special units have been empowered to shoot a suspected terrorist dead to prevent an attack, has led to a gradual distancing of the civil police from the community it polices (Weisburd et al. 2009).

Conclusions and Future Directions

Police use of firearms has been subject to intense debate, controversy, and scrutiny in many parts of the world. While the police are armed in a majority of countries, there is very little clarity on the subject of when and how they are to wield firearms uncontroversially. The law usually only permits police officers to use firearms to kill in self-defense or to protect someone else's life, or in a few countries while arresting an armed and dangerous offender. A police officer's decision to use firearms in real-life situations is thus based on her/his judgement of the situation and the danger they perceive to their own or someone else's life. However, this decision is subsequently subject to review with the benefit of hindsight, and often the officer's judgement is questioned or found wanting. Police shootings have caused public discontent and anger especially among ethnic minority communities who perceive the disproportionately high number of shooting incidents involving minority ethnic citizens to be racially motivated. Public riots, disorder, and demonstrations against perceived misuse of or excessive use of lethal force put a strain on police-community relations, which in turn has a detrimental impact on the police role of maintaining law and order. We need further research to understand and improve police decision making under stressful situations. This would also be beneficial in-educating the public about the dynamics of the decision-making process and also explain why police actions that appear to be racially prejudiced might not actually be so in some cases.

Various explanations have been put forward to explain police propensity to abuse or use excessive force. These explanations account for police violence at the individual level, at the organizational level, and at the social or structural level. To maintain harmonious police-public relations, police use of force – especially firearms – has to be made transparently accountable and subject to administrative control and review. We need further research to evaluate the effectiveness of these various interventions aimed at different levels and under different circumstances and contexts in order to understand what works best, under which circumstances, and why.

Growing terrorist threats and increasing public insecurity have led the civil police in many countries to adopt a more paramilitary role, with specialized units armed and trained to deal with emerging terrorist threats. The inevitable and inexorable paramilitarization of the civil police under these circumstances has placed additional strain on police-public relations. We need to understand how police counter terrorist responses can be inclusive of community policing rather than alienating the community via “high policing” tactics. A related issue is the demand for arming of unarmed police especially in the UK. The agenda for arming the police, or specialized police units, is driven by a perception of wider social violence in general but more specifically by particular incidents. We need a better understanding of what the outcomes of the incident-driven nature of further arming the police might be with an eye to the possibility of embarking upon an ever spiraling cycle of violence, given the fact that disarming an armed police force is rarely ever an option.

Recommended Reading and References

Alpert G, Dunham R (2004) Understanding police use of force: officers, suspects and reciprocity. Cambridge University Press, Cambridge

Belur J (2010) Permission to shoot? Police use of deadly force in democracies. Springer, New York

Bittner E (1975) The functions of the police in modern society. Jason Aronson, New York

Bowling B (2010) Policing the caribbean. Oxford University Press, Oxford

Chevigny P (1995) Edge of the knife: police violence in the Americas. New Press, New York

Coady T, James S, Miller S, O’Keefe M (eds) (2000) Violence and police culture. Melbourne University Press, Carlton South

Flanagan T, Vaughan M (1996) Public opinion about police use of force. In: Geller W, Toch H (eds) Police violence: understanding and controlling police abuse of force. Yale University Press, New Haven, pp 113–128

Fyfe J (1978) Administrative intervention on police shooting discretion. J Crim Justice 7:309–323

Fyfe J (1981) Toward a typology of police shootings. In: Fyfe J (ed) Contemporary issues in law enforcement. Sage, Beverly Hills

Fyfe J (1986) The split second syndrome and other determinants of police violence. In: Campbell A, Gibbs J (eds) Violent transactions. Basil Blackwell, Oxford

Geller W, Scott M (1992) Deadly force: what we know. Police Executive Research Forum, Washington, DC

Grant D, Grant J (1996) Officer selection and the prevention of abuse of force. In: Geller W, Toch H (eds) Police violence: understanding and controlling police abuse of force. Yale University Press, New Haven

Green P, Ward T (2004) State crime: governments, violence and corruption. London/Sterling, Pluto Press

Greene J, Herzog S (2009) The implications of terrorism on the formal and social organization of policing in the US and Israel: some concerns and opportunities. In: Weisburd D, Feucht T, Hakimi I, Mock L, Perry S (eds) To protect and to serve. Springer, New York

Hawkins G, Ward P (1970) Armed and disarmed police: police firearms policy and levels of violence. J Res Crime Delinq 7(2):188–197

Klinger D, Brunson R (2009) Police officers’ perceptual distortions during lethal force situations: informing the reasonableness standard. Criminol Public Policy 8(3):117–140

Klockars C (1991) The Dirty Harry Problem. In: Klockars C, Mastrofski S (eds) Thinking about police: contemporary readings, 2nd edn. McGraw-Hill, New York

Knutsson J, Noree A (2010) Police use of firearms in the Nordic Countries: a comparison. In: Kuhn J, Knutsson J (eds) Police use of force: a global perspective. Praeger, Santa Barbara, pp 115–123

Lawrence R (2000) The politics of force: media and the construction of police brutality. University of California Press, Berkeley

Locke H (1996) The color of law and the issue of color. In: Geller W, Toch H (eds) Police violence: understanding and controlling police abuse of force. Yale University Press, New Haven, pp 129–149

McCulloch J (2000) Keeping the peace and making war: the police and the military – rhetoric and reality. In: Coady T, James S, Miller S, O’Keefe M (eds) Violence and police culture. Melbourne University Press, Carlton South, pp 182–204

Phelps J (2010) Lessons in developing democratic police from post-second world war successes. Int J Comp Appl Crim Justice 34(1):119–137

Punch M (2010) Shoot to kill: police accountability, firearms and fatal force. Policy Press, Bristol

Ramana PV (2006) The maoist movement in India. Def Secur Anal 22(4):435–449

Reiner R (2000) The politics of the police, 3rd edn. Oxford University Press, Oxford

Scharf P, Binder A (1983) The badge and the bullet. Praeger, New York

Sherman L (1983) Reducing police gun use: critical events, administrative policy and organisational change. In: Punch M (ed) Control in the police organisation. MIT Press, Cambridge, MA, pp 98–125
Skolnick J (1975) Justice without trial: law enforcement in democratic society, 2nd edn. Wiley, New York
Skolnick J, Fyfe J (1993) Above the law: police and the excessive use of force. The Free Press, New York
Squires P, Kennison P (2010) Shooting to kill? Policing, firearms and armed response. Wiley, West Sussex
Summers D, Kuhns J (2010) Currently available less-than-lethal alternatives and emerging technologies for the future. In: Kuhn J, Knutsson J (eds) Police use of force: a global perspective. Praeger, Santa Barbara, pp 188–198
Toch H (1996) The violence-prone police officer. In: Geller W, Toch H (eds) Police violence: understanding and controlling police abuse of force. Yale University Press, New Haven, pp 94–112
Uldricks N, van Mastrigt H (1991) Policing police violence. Kluwer, Dordrecht
Waddington PAJ (1991) The strong arm of the law. Clarendon, Oxford
Waddington PAJ (1999) Armed and unarmed policing. In: Mawby R (ed) Policing across the world: issues for the twenty-first century. UCL Press, London, pp 151–166
Weisburd D, Jonathan T, Perry S (2009) The israeli model for policing terrorism: goals, strategies and open questions. Crim Justice Behav 36(12):1259–1278
Worden R (1996) The causes of police brutality: theory and evidence on police use of force. In: Geller W, Toch H (eds) Police violence. Yale University Press, New Haven

Police-Led Interventions to Enhance Police Legitimacy

Sarah Bennett[1], Jacqueline Davis[1] and
Lorraine Mazerolle[2]
[1]Australian Research Council Centre of
Excellence in Policing and Security (ARC),
Institute for Social Science Research (ISSR),
The University of Queensland, St. Lucia, QLD,
Australia
[2]Institute for Social Science Research,
The University of Queensland, St. Lucia,
QLD, Australia

Synonyms

Procedural justice

Overview

A growing body of policing research demonstrates that procedurally just approaches to policing – including satisfaction, trust, and confidence – improve citizen perceptions of police, thereby influencing perceptions of police legitimacy and facilitating the capacity of police to maintain order and control crime. This entry draws from a systematic search of police legitimacy and presents a qualitative overview of 104 evaluations of policing interventions that sought to improve police legitimacy. Mazerolle and colleagues (2013) used content analytic methods to synthesize the extant literature, examining studies that used either a process- or instrumental-based approach to enhance perceptions of legitimacy. In this entry, it is shown that some interventions, notably those that are police-only type interventions, seek to influence legitimacy primarily through their efforts to be more effective. In contrast, interventions that involve partnership approaches tend to directly influence perceptions of legitimacy by increasing efforts to enhance positive perceptions of police. Overall, it is shown that different process- and instrumental-type policing strategies seek to improve police legitimacy in a variety of ways.

Fundamentals

An extensive and growing body of international research focuses on police legitimacy: that is, the expectation that the police will be obeyed, over and beyond the threat of punishment or expectation of reward (Tyler 2006). Several theoretical pathways have been proposed for the relationships between police activity and police legitimacy. The process model of police legitimacy holds that when police use the key ingredients of procedural justice – allowing citizens to participate in decision-making, treating people with dignity and respect, maintaining neutrality, and acting with transparent motives – this leads to enhanced citizen perceptions of police, including trust, confidence, and satisfaction with police, and thereby increased levels of police legitimacy

(Hinds and Murphy 2007; Mastrofski et al. 1996; McCluskey 2003; Reiss 1971; Sunshine and Tyler 2003; Tyler and Fagan 2008; Wells 2007). Increased police legitimacy then encourages compliance with both police directives and the law, thus indirectly affecting crime and disorder (Tyler and Fagan 2008; Tyler and Huo 2002). The instrumental model of police legitimacy, in contrast, posits that police effectiveness in controlling crime and disorder is a key antecedent to police legitimacy, and thereby perceptions of effectiveness influence citizen perceptions such as satisfaction with, and trust and confidence in, the police (Sunshine and Tyler 2003).

The causal pathways proposed in these legitimacy models show that police use a number of different intervention strategies – which are either process-focused or outcome-focused – in their efforts to enhance perceptions of legitimacy. One way to situate the variety of police interventions that seek to enhance police legitimacy is to use a typology developed by Weisburd and Eck (2004). These authors describe policing strategies on two dimensions: diversity of approaches and level of focus. They identify four broad categories of police interventions: first, community policing, broadly defined as those interventions that involve police partnerships with the community, employ a highly diverse portfolio of approaches, and are loosely focused around general crime problems; second, problem-oriented policing, broadly defined as those interventions that, again, involve police partnerships but are highly focused on discrete problem people or places; third, hot spot policing, defined as those police interventions that are highly focused on problem people or places but have little diversity and few partnerships with other agencies; and finally, the standard model of policing, defined as police interventions that neither involve diverse partnerships nor a specific focus on problem people or places.

In this entry, the heuristic of the Weisburd and Eck (2004) typology of policing interventions is used to compare and contrast different policing approaches used to enhance citizen perceptions of police legitimacy, drawing from a systematic search of the extant literature that identified 163 studies out of over 20,000 abstracts reviewed that met the review criteria (see Bennett et al. 2009). Content analytic methods were used to group strategies, identifying a subset of 104 studies that included evaluation information that could be used in a narrative review. These interventions are examined, based on their area of focus and diversity of approaches, in order to illustrate the varying impact of process-focused or outcome-focused models of police legitimacy.

Key Issues

Scholarly interest in police–citizen interactions, perceptions of police treatment, and the legitimate authority of the public police is not new (e.g., Bayley and Mendelsohn 1969; Bellman 1935; Decker 1981; Parratt 1938; Reiss 1971). However, it is the contemporary stream of research by Tom Tyler and his colleagues that spearheaded renewed interest in police legitimacy and continues to shape the field today (see Tyler 1990, 2003, 2004, 2006; Tyler and Fagan 2008; Tyler and Huo 2002). This recent body of research finds that citizens are more likely to comply with police directives when they view police as legitimate. Tyler and his colleagues also show that police legitimacy encourages law-abiding behavior not just *during* a police–citizen encounter but also *outside* of encounters (i.e., in everyday life, such as abiding by traffic rules) (Sunshine and Tyler 2003; Tyler and Huo 2002). As Tyler (2004), p. 85 suggests, unless the police are “widely obeyed” by the public, the capacity of police to maintain order is compromised (see also Tyler 1990). Arguably, one of the most important contributions of Tyler’s research over the last 20 years is the deepening of our understanding as to how the efficiency and effectiveness of the order maintenance role of the public police is dependent upon the public’s support for, and perceived legitimacy of, the police.

Legitimacy is defined by Tyler (2006), p. 375 as “a psychological property of an authority, institution, or social arrangement that leads those connected to it to believe that it is appropriate, proper, and just." The key defining feature of a legitimate authority is that people feel obliged to voluntarily comply with the authority's directives as opposed to compliance out of fear of punishment or expectation of reward (Tyler 2006). In policing, legitimacy reflects a "social value orientation toward authority and institutions" (Hinds and Murphy 2007, p. 27), and evidence shows that it is a person's belief in the legitimacy of the authority or institution issuing a command that "leads people to feel that the authority or institution is entitled to be deferred to and obeyed" (Sunshine and Tyler 2003, p. 514).

Procedural justice refers to the use of fair procedures in the decision-making process (Hinds and Murphy 2007; Thibaut and Walker 1975). Scholars across many different contexts and disciplines (e.g., taxation compliance and organizational behavior) have studied the impact of a person's treatment in a range of different decision-making forums. In policing, renewed academic interest in procedural justice emerged during the late 1980s and early 1990s at a time of high-profile incidents of pervasive police corruption and police misconduct (e.g., racial profiling, excessive force) (Reiner 1985) and concerns about police inadequacies in dealing with upsurges in crime (Maher and Dixon 1999; Weisburd and Braga 2006), which led to a general loss of confidence in traditional police responses to crime (Weisburd and Eck 2004). These concerns created fertile ground for the study of police legitimacy and the concomitant study of police–citizen encounters based on fair and respectful processes and procedures. Tyler and colleagues identify four factors that define "procedural justice" in interactions between police and the public: citizen *participation* in the proceedings prior to an authority reaching a decision, perceived *neutrality* of the authority in his/her decision, whether or not the authority showed *dignity and respect* throughout the interaction, and whether or not the authority is perceived to have *trustworthy motives* (Tyler 2004; Tyler and Lind 1992).

Researchers have sought to differentiate between the effects of police legitimacy on subjective outcomes, such as citizens' perceptions of police, or intent to comply, and objective outcomes, such as actual compliant behaviors or reoffending. The difference between citizens' perceptions and actions is a crucial consideration in the design of police interventions and evaluations of these interventions. While it is neither the purpose of this entry to explore the complex relationship between legitimacy and procedural justice further nor to compare operational definitions and measurement methods, this is a highly important area of research. In this entry, authors' identification of legitimacy and procedural justice is accepted and the distinction between process outcomes (such as citizen satisfaction) and instrumental outcomes (such as reoffending) is used to classify intervention strategies according to their underlying logic.

The Systematic Search Approach

The results of a systematic search conducted on behalf of the UK National Policing Improvement Agency (NPIA) (Bennett et al. 2009) were used to obtain a population of studies evaluating police-led interventions to improve legitimacy. The systematic search covered a range of databases for published journal articles, as well as a search of the gray (unpublished) literature including reports and dissertations. Search keywords focused on police legitimacy, effectiveness, and procedural justice. The systematic search identified over 20,000 abstracts on police legitimacy and related topics. Readers are directed to the NPIA report for a full explanation of the systematic search and document retrieval strategy (see Bennett et al. 2009).

Information on any evaluation of police-led interventions was gathered from the 20,000 plus abstracts. These interventions had to either (1) explicitly aim to increase legitimacy or, (2) in the absence of an explicit legitimacy aim, use at least one of the principles of procedural justice: citizen participation in decision-making, police neutrality, dignity and respect, and transparent motives. Information from descriptive studies was also recorded to gain an

Police-Led Interventions to Enhance Police Legitimacy, Fig. 1 A framework for categorizing legitimacy policing interventions (Source: Weisburd and Eck (2004))

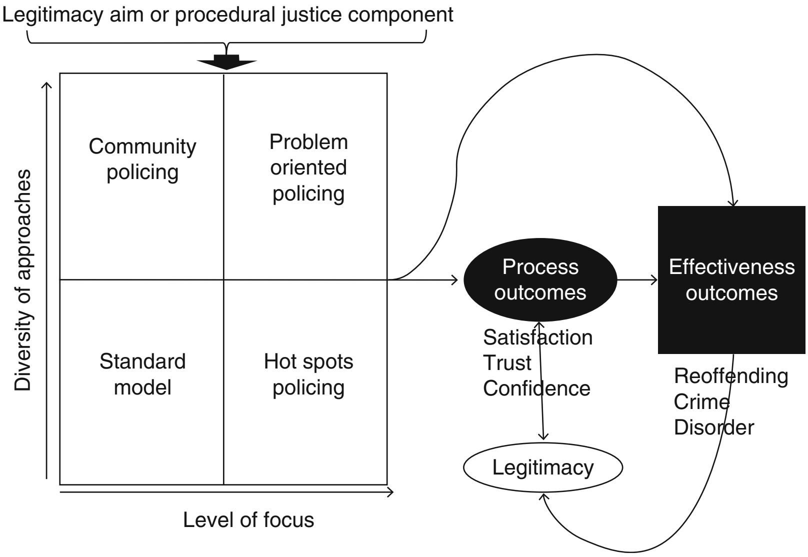

understanding of initiatives that had been tried, even if they had not been evaluated. A pictorial explanation of the organizational framework (presented in Fig. 1) was used to guide coding and synthesis of the extant literature.

The systematic review of the identified abstracts revealed 163 sources that met the above criteria. Text analytics software (Leximancer: Leximancer software examines a body of text and produces a ranked list of concepts on the basis of word frequency and co-occurrence of usage (see https://www.leximancer.com/)) was used to create a map of the key concepts contained within the documents and to rank the most prominent concepts relating to policing within the body of text. The outputs from this analysis were used to compare whether, in the body of literature, the focus on the concept of legitimacy in the text matched the focus on the outcome of legitimacy in the evaluations. The conceptual structure of the 163 sources on police legitimacy interventions is presented in Fig. 2, below. As shown in Fig. 2, the central concept in the corpus of text was legitimacy, as would be expected given the goals of the search. The concept of legitimacy links directly to a large number of concepts, including the concept of police. The map suggests that within the body of studies, the concept of procedural justice is indirectly linked to legitimacy through the concept of policing. The semantic analysis showed that "crime" was the most prominent concept relative to the concepts of "police" and "policing." The second most prominent concept related to policing was the concept of "trust," with "justice" and "procedural" as the third and fourth most prominent terms, respectively. The most prominent compound term in the literature on police was "procedural justice," with a relative frequency of 29 % and strength of 38 %. The content analysis thus demonstrates the centrality of legitimacy as a concept within this particular body of literature.

Systematic Search Findings

Of the 163 selected studies, 104 of the studies reported evaluation findings. All of the interventions included in this subset of studies articulated that the intervention sought to improve police legitimacy. The studies referred to in the following section, and the complete list of 163 studies, can be found in the Campbell Collaboration systematic review on legitimacy in policing (Mazerolle et al. 2013). The subset of studies included in our narrative review is organized using the typology articulated by Weisburd and Eck (2004) as follows.

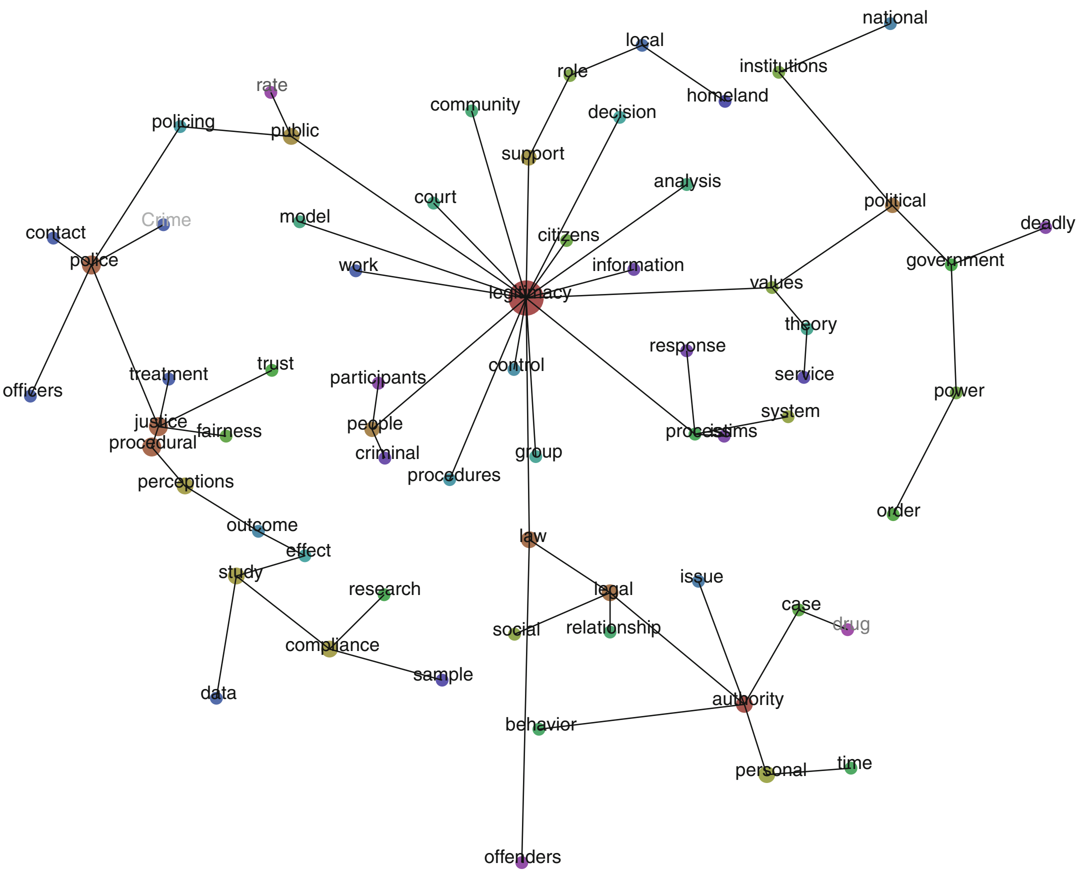
national
local
institutions
role
rate
homeland
community
decision
policing
public
support
analysis
court
political
Crime
model
contact
citizens
deadly
police
work
information
government
legitimacy
values
theory
response
control
trust
treatment
participants
service
officers
power
system
justice
fairness
people
procedural
process
victims
criminal
group
procedures
perceptions
order
outcome
law
effect
study
issue
research
legal
case
drug
social
relationship
compliance
sample
data
authority
behavior
personal
time
offenders

Police-Led Interventions to Enhance Police Legitimacy, Fig. 2 Concept map for legitimacy policing literature. The larger concepts are more prominent within the text

Community-Based Policing Strategies

Community policing strategies fall into the top left quadrant of the Weisburd and Eck (2004) typology, characterized by a high diversity of approaches but low level of place or person focus (see Fig. 1). The review uncovered a number of legitimacy-enhancing strategies that fell into this category, including community policing, but also included interventions that were not strictly community policing interventions. These approaches tended to focus on the process model of increasing legitimacy directly through procedurally fair operations. Evaluations of these strategies therefore tended to focus on "soft" outcomes such as citizen confidence, satisfaction, and/or trust in police; perceptions of procedural justice; and police legitimacy. The following section discusses each strategy grouping in turn.

Community Policing

Community policing refers to a group of strategies that emphasize high-quality police–citizen interaction, especially in nonemergency situations, where police and community share responsibility for crime prevention and control (Bell 2004). Community policing is hypothesized to provide benefits over traditional policing methods, including increased public trust, satisfaction, and confidence in police; higher reporting of crime; more effective crime prevention; and better police–community relations (Bell 2004).

Community policing has been variously described as a philosophy, a strategy, or a series of tactics. For the purpose of the review, a number of elements commonly found in community policing strategies were drawn out. While most community policing strategies use one or more of these elements, not all elements are required for a strategy to be termed community policing. Skolnick and Bayley (1988) separate community policing elements into four categories. First, community-based *crime prevention* includes initiatives such as neighborhood watch, police safety seminars, and supplying security material to citizens (Bell 2004). Second, an emphasis on *policing nonemergency situations* includes routine motor, foot, or bicycle patrols and door-to-door activities. Third, increasing *police accountability* strategies include the development of specialist liaison officers for communication with minority groups, and programs that allow citizens to observe the police at work. Finally, *decentralizing command* within the police force and giving more power to officers actively policing the community is a key element of community policing. Other authors have identified additional elements, including the long-term assignment of officers to beats (geographic areas), community consultation to define community problems, reducing fear of crime, extended foot patrols, and an emphasis on nonconfrontational officer–citizen interactions (Fielding 2002).

The systematic search located 70 articles and reports on community policing-type strategies that fit within the Weisburd/Eck typology of *diversity of approaches/low level of focus*. The strategies included community policing grants and directives, reassurance policing, beat policing, proximity policing, contact patrols, ombudsman policing, mini-police stations and shopfronts, and neighborhood watch programs.

Overall, community policing was found to have a positive effect on a range of process outcomes, including citizens' confidence in police, satisfaction with police, and perceptions of police effectiveness. Community policing initiatives also affected police officers' views and internal processes such as increased levels of officer monitoring and accountability and holding positive views of mini-police stations. The successful creation of partnerships with community agencies was another benefit identified in the literature.

Not all evaluations were positive, however. For example, one researcher found that the introduction of proximity policing resulted in a decrease in confidence in police, due in part to an increase in administrative tasks to the detriment of time spent on physical patrols.

Evidence was mixed regarding instrumental outcomes for community policing. In terms of positive effects, several studies found a decrease in citizens' perceptions of crime overall, disorder, and fear of crime. Others found decreases in perceptions of particular crimes. For example, some scholars found a decrease in the perceived levels of vandalism and graffiti, home burglaries, and drug dealing. Some social observational studies found a decrease in citizen noncompliance behaviors with police and also that citizen cooperation was higher for encounters with beat police compared to traditional patrols. In a series of evaluations of the Weed and Seed community policing grants, researchers found decreases in citizens' self-reported revictimization in nearly all program sites, a finding paralleled in an evaluation of reassurance policing in English cities. Other measures of community policing effectiveness included citizens taking steps to improve their home security.

Again, however, not all studies found positive instrumental effects for community policing. One study found an increase in perceived crime levels, a decrease in perceptions of police fairness, and a large decrease in perceived safety. Other community policing evaluations found that community policing had no effect on either perceptions of safety or levels of official reported crime. One study found that an increase in official crime reports at sites with police shopfronts was due to the increased willingness of shopkeepers to report crimes, rather than an increase in victimization.

Finally, the experimental literature is sparse and ambiguous regarding the effect of community policing on police legitimacy. Some authors have found no effect of community policing on legitimacy, while others have found a decrease in perceptions of police legitimacy after the implementation of a community policing intervention. Notably, these findings are contrary to nonexperimental survey research on procedural justice and legitimacy such as that of Sunshine and Tyler (2003).

School-Based Interventions

A second category of interventions that fell under the *high diversity of approaches/low level of focus* was school-based policing interventions that seek to reduce delinquency, antisocial behavior, and truancy by increasing police legitimacy through both process and instrumental paths. The review located six such interventions. Two of these described police officers whose activities were based primarily within one or more schools and who formed partnerships with the school community to enhance law enforcement, education, and counseling. Two sources described a program in which police officers visited schools and gave a talk on resisting gangs. One intervention focused on schoolchildren interacting informally with police officers in a nonconfrontational situation to increase children's perceptions of police legitimacy, thereby increasing their willingness to assist police in the short and longer term. Finally, one source described a nonpunitive truant recovery program. Few implementation problems were reported for school-based interventions.

The overall evidence for school-based interventions was mixed. School-based interventions tended to measure success using instrumental outcomes, including willingness to assist police, grades and truancy, and police contacts and arrests. The truant recovery program decreased the number of fail grades, unexcused absences, school sanctions, and disciplinary reports for children in the program. However, school-based interventions had no significant impact on willingness to assist police, police contacts, or arrests. Further, schools assigned a police officer tended to have greater prevalence of truancy than schools without one, although the causal direction of this effect is not explicated in the group of studies collected. The informal contact program found a null effect on students' perceptions of police legitimacy.

Interagency Collaboration

Another group of interventions falling into the *high diversity of approaches/low-focus* quadrant are those focusing on interagency collaboration, where police used formal arrangements with another organization (often a social service) to improve service delivery. These included police collaborating with a community mental health unit, with social workers following up on police domestic violence investigations; a multi-jurisdictional task force established for broad-based service provision; outreach workers working with sex workers at risk of violence; and a localized initiative aimed at reducing the prevalence of street walkers.

While interventions involving collaboration between multiple agencies frequently reported problems in implementation arising from differences between the agencies involved – such as different reporting policies and information retrieval systems – several studies reported improved information sharing and closer relationships between agencies as a result of the intervention.

The studies evaluating interagency collaboration effects on legitimacy tended to focus on instrumental outcomes rather than process outcomes. The intervention aimed at reducing the number of street walkers in an area was seen to be largely successful in achieving its goals, as well as improving police–community relations, and decreasing recorded street crime and fear of crime. Households that received a follow-up call from a social worker after a domestic violence incident were no less likely to experience violence than households that did not receive the call, but they were more likely to report new incidences of violence. The intervention encouraging sex workers to report violent clients achieved an increase in reporting and contributed to the arrest and conviction of two violent clients.

Restorative Justice Conferencing

Given their explicit focus on enhancing legitimacy (Daly et al. 2006) and improving both victims' and offenders' perceptions of procedural justice, restorative justice conferencing programs were included in the systematic review of police legitimacy. According to Marshall (1999), p. 5, “Restorative justice is a process whereby parties with a stake in a specific offence collectively resolve how to deal with the aftermath of the offence and its implications for the future.” When interventions are facilitated by police, improved perceptions of police procedural justice can lead to improved perceptions of police legitimacy.

Most empirical research on restorative justice processes has been conducted on those processes that involve a face-to-face meeting between the offender, victim, and supporters. Although conferences can be mediated by either a police officer (the Wagga model: Moore and O’Connell 1994) or a nonpolice officer trained for the purpose (the New Zealand model: Daly et al. 2006), the systematic review focused on the former, more clearly police-led interventions because the police, as conveners, have direct power to enforce the outcomes of the conference (Moore and O’Connell 1994).

The systematic search located a number of articles reporting on studies of police-run restorative justice conferences, as well as restorative cautioning, and a program that trained police officers in mediation skills and gave them the opportunity to refer parties to a local mediation facility during an incident.

The review found that, overall, police-run conferences had a positive impact on process outcomes for both offenders and victims. The police-led conferencing model has also been observed to prompt a shift from a retributive, adversarial policing culture to one focused on restoration and harm prevention (Moore and O’Connell 1994), although this was not observed in all cases. One study measured the effects of restorative justice interventions on perceptions of police legitimacy and found a substantial effect across participants and crime types.

The effect of restorative justice on instrumental outcomes is less clear. While some scholars have found that conferenced offenders had lower recidivism than court-processed offenders, the difference was not large enough to be significant and varied greatly across different types of offence. Others have similarly found that individuals who were cautioned using restorative justice procedures were less likely to reoffend than individuals cautioned using traditional police methods.

Problem-Oriented Policing

Problem-oriented policing strategies fall into Weisburd and Eck’s (2004) top right quadrant of *high diversity of approaches/high level of focus on a particular problem* (see Fig. 1) and generally fall into the instrumental model of police legitimacy. The systematic search uncovered a diverse range of problem-oriented policing interventions in the context of police legitimacy, including large-scale initiatives focused on violent crime in general, gun violence, youth gun violence, and juvenile curfews. Other interventions included community alarm systems, specialized task forces for drug reduction, concentrated traffic enforcement, crime prevention through environmental design, training for liquor licensees and bar staff to recognize fake identification, grants to enhance gender-based violence prevention efforts, and vehicle safety programs. Several sources described general problem-oriented policing programs, referred to variously as problem-oriented policing, problem-solving policing, and risk-focused policing.

Study authors identified a range of issues with the implementation of problem-oriented policing interventions, including lack of a single responsible managing agency, citizen anxiety about police and social services becoming involved in their personal lives (e.g., community alarm system case), difficulty with sustaining a high level of focus over time, inability of police to fulfill their commitments according to the program, and difficulty enforcing collaboration between police and partner agencies.

The review showed that only a few problem-oriented policing strategies used the process model of legitimacy, as evidenced by a lack of attention to process-related outcomes in the group of interventions. These included community support, citizen perceptions of procedural justice, interagency communication, and job satisfaction.

Instead, problem-oriented policing strategies identified in the review that articulated measures of police legitimacy tended to measure outcomes related to effectiveness. In particular, the studies reported measures that captured the problems the interventions were trying to address. These measures included improving police legitimacy through improving police effectiveness and impact on violent crime, gang crime, homicides, severe road accidents, speed-related accidents, traffic violations, street crime, burglary, arrests of juveniles, self-reported delinquency, reoffending, calls for service, and citizen perceptions of safety.

Hot Spot Policing

Hot spot policing strategies fall into the *high focus/low diversity of approaches* typology. This type of strategy is generally implemented by police alone in a small geographic area, with the aim of reducing crime and improving police legitimacy through improved effectiveness. The search located only two evaluations of interventions that could be described as hot spot policing aimed at enhancing legitimacy. Both interventions implemented increased enforcement rates and penalties for disorder crimes in a small geographic area. One evaluator commented that the program, while successful in reducing crime, had created a large number of fugitives from the justice system. The other evaluation found that enforcing disorder crimes had no effect on levels of more serious crime (aggravated burglary and robbery) in the target area.

Clearly, a much larger literature exists on the effectiveness of hot spot policing interventions. However, the sufficiently different language and terminology used in the literature around legitimacy policing and the literature around hot spot policing led to the exclusion of most of these types of interventions from the review. Moreover, either the researchers or police have not viewed hot spot policing as the type of intervention that could enhance police legitimacy. This does not mean that hot spot policing *cannot be used* to enhance perceptions of legitimacy. Rather, the studies are silent on how hot spot policing could be used to enhance legitimacy.

Standard Model of Policing

The standard model of policing includes interventions that are *low in diversity of approaches/ low in focus*. Several such strategies were found that authors linked to police legitimacy. Standard model policing interventions tended to focus broadly on instrumental outcomes and police effectiveness, although this varied among interventions.

Specialized Teams

Several studies described police units that had created specialized teams of officers focused on a particular type of crime, including child protection, domestic and family violence, mental health, drug crime, homicide and violent crime, gang crime, disorder crimes, and automatic number plate recognition intercept teams for traffic violations.

Evaluations of specialized teams often commented on the importance of the commitment of individual officers, especially the commitment of senior officers in their support of the specialized team in accessing information and resources. Interventions in which the team consisted of members from multiple agencies sometimes encountered problems in communication. This was largely due to differences in management structures and decision-making styles between police departments and other agencies, leading to a low level of honesty and trust in working relationships within the teams. Other evaluators commented that the goals of national-level programs did not always match the goals of individuals implementing the intervention in the field. One study found that reassigning officers into structured and response teams was confusing for the officers involved due to a lack of clearly stated objectives for the two groups. Another study found that a particular issue for specialized units was the potential for the unit members to legitimize their role by creating and reinforcing myths concerning the problem they were established to address.

Specialized team interventions generally focused on increasing legitimacy through improving instrumental outcomes, although some interventions did report process outcomes. Several evaluations found that specialized teams were positively received by community members, clients, and community leaders. One study found that it increased positive attitudes of police officers toward their jobs.

Evaluations of specialized teams also tended to emphasize instrumental outcomes, with mixed results. For example, a proactive multiagency mental health unit had a substantial impact on the number of arrests and involuntary hospitalizations made in mental health-related calls for service. In geographic areas assigned a specialized homicide and violent crime task force, rates of violent crime and homicide decreased, although the design of the evaluation prevented authors from attributing this effect directly to the intervention. Domestic violence suspects visited by a specialized domestic violence unit had a lower prevalence of reoffending than suspects visited by regular patrol officers, although frequency and severity of reoffending were not affected for suspects who did reoffend. Less positively, one evaluation found a higher rate of reoffending in suspects visited by a special family violence response team, compared to controls. Another evaluation of a crisis intervention team for domestic violence found that the crisis team generated more arrests than regular officers in domestic violence cases but that victim cooperation was lower in cases visited by the specialized team than in other cases.

Targeted Enforcement

Targeted enforcement covers a range of strategies aimed at reducing a particular problem by increasing detection of the problem and enforcement of penalties. Targeted enforcement falls under the standard model of policing and can be separated from problem-oriented policing because of its low diversity of approaches and sole implementation by police agencies, with no collaboration between police and citizens or other agencies. Enforcement strategies tended to target particular crimes such as sales of alcohol and tobacco to minors. One targeted enforcement strategy used enhanced forensic science techniques in investigations.

The systematic search found few evaluations of this type of intervention that articulated to the legitimacy theme. As such, little information was gathered on the process of implementing these interventions. The standard policing interventions generally focused on instrumental outcomes, to the exclusion of process-type outcomes. The evidence for the effectiveness of these strategies was not very positive. For example, in one case in which enhanced forensic science techniques were implemented as a strategy, the researchers found that rates of case closure were no higher than in other cases, and the special forensic science technology was very expensive to use. A strategy targeting tobacco sales to minors found that although targeted communities improved their compliance with tobacco sales laws faster than control communities, adolescents' self-reported access to, and use of, tobacco was not affected. Another intervention involving undercover police checks of licensed premises failed to produce any significant deterrent effect on future sales of alcohol to minors.

Officer Training

One of the primary sources for improved police legitimacy is citizens' day-to-day interactions with the police, independent of any large-scale initiative. Police can implement training interventions to increase the likelihood of a positive outcome in officers' encounters with members of the public. Although evaluations of this type of intervention are rare, Hails and Borum (2003) suggest that internal training to improve legitimacy is carried out by many police forces. Legitimacy training initiatives are distinct from other types of police training focusing on combat or responding to hostile situations. The search discovered several different citizen-focused police training initiatives, including community policing training, diversity training, leadership training, life skills training, crisis intervention training, and victim-focused training.

The overuse of force and arrest against particular sections of society by police (e.g., ethnic minority groups), and the overrepresentation of these populations within the prison population compared to society in general, represents a key weakness of the criminal justice system and a significant challenge to its legitimacy in many countries. Legitimacy training may thus include specific training to help officers respond to the needs of populations they are not familiar with, for example, victims of domestic violence or people with a physical or mental disability. It is hypothesized that specialized training for particular population interactions enables police officers to respond to the needs of these populations in crisis situations, thereby reducing the officers' need to resort to arrest or force (Hails and Borum 2003).

Implementation issues identified in police training generally focused on the occasional discrepancies between officers' training and expectations of their performance in the field. For example, a community policing training program evaluation found that although the training was successful in teaching officers about community policing, their ability to implement their training in practice was constrained because of performance indicators that focused on instrumental outcomes, such as arrests and tickets, rather than the process-type outcomes promoted in community policing.

Police training interventions tended to focus on the process model of legitimacy and generally reported positive results for process-type outcomes, particularly those relating to police perceptions. Officers trained in crisis intervention tended to give themselves more positive ratings for handling cases than did officers who did not receive the training. Similarly, command staff who participated in diversity training reported an increased ability to communicate effectively; police who received victim-focused training reported more favorable attitudes and behavioral intentions toward victims than did untrained officers; and officers who underwent training in life skills reported increases in self-confidence and lower dogmatism compared to untrained officers. An exception to the generally positive results was a study evaluating police leadership training which found no effect of the training on police leadership styles.

Officer training programs had mixed impacts on process outcomes regarding citizen perceptions of police. Clients of the crisis intervention training program were happier with the service provided by trained officers than that of untrained officers, but an evaluation of victim-focused training found no effect on victim perceptions of police or fear of crime. Instrumental outcomes were not generally measured by police training interventions. The crisis intervention training program produced a null effect on the rate of return calls to the same address.

Discussion and Conclusion

Police legitimacy is important to maintaining order and ensuring compliance with laws and police requests. If people feel the police have legitimacy, they will voluntarily obey police, freeing police resources from enforcing compliance based on fear of punishment or anticipation of reward. While there are multiple pathways to improving police legitimacy, including improving police performance and enhancing the nature of police–citizen encounters, the focus of this review was on strategies that sought to improve police legitimacy either (1) explicitly (as identified by the authors of papers and reports on the intervention) or (2) implicitly through the use of procedural justice elements in policing.

The systematic review found 163 studies that evaluated policing interventions articulated as approaches seeking to enhance perceptions of police legitimacy, of which 104 studies with sufficient data are used in this essay to show the breadth of research in this area. For the population of 163 studies, text analysis shows that the term "legitimacy" is indeed a key concept within the policing literature. A map of the conceptual structure of the body of police legitimacy evaluation literature revealed that legitimacy was the central underlying concept, with strong links to the concepts of policing and procedural justice.

However, while legitimacy was central in the text analysis, it was clearly not an outcome routinely measured or considered to be of primary importance by study authors. Interventions tended to measure effectiveness using process-type outcomes, such as citizens' perceptions of the fairness of police decisions, or instrumental outcomes, such as citizens' perceptions of police effectiveness, but rarely did they measure the apparently central concept of legitimacy. Thus, this review concludes that the enthusiasm for police legitimacy as a construct in the theoretical and survey literature is not matched in the evaluation literature.

The preponderance of these 163 legitimacy-enhancing police interventions was in the two quadrants that Weisburd and Eck (2004) identify as *diversity of approaches*, including community- and problem-oriented policing approaches. Community policing types of interventions that sought to enhance legitimacy tended to use the process model of legitimacy, in which citizens' perceptions of police legitimacy are influenced by their perceptions of the process of policing, rather than police effectiveness. In contrast, problem-oriented policing types of interventions tended to use an instrumental approach to legitimacy, in which increased police effectiveness leads to increased legitimacy.

This overview suggests that while many policing interventions, including community policing, problem-oriented policing, hot spot policing, and standard model policing, may be useful for improving police legitimacy, these approaches are rarely explicitly articulated as legitimacy-enhancing interventions. In theory, previously evaluated interventions may have influenced legitimacy in multiple ways, through improving citizens' perceptions of police processes or through improving police effectiveness. However, in practice very few evaluations measure legitimacy as an outcome. It is suggested, therefore, that evaluators need to carefully focus on measuring a wider range of outcomes in addition to the more commonly reported process and instrumental outcomes.

Related Entries

- ▶ Causes of Police Legitimacy
- ▶ Community Policing
- ▶ History of Restorative Justice
- ▶ Hot Spots and Place-Based Policing
- ▶ Monitoring and Evaluation of Restorative Justice
- ▶ Police Legitimacy and Police Encounters
- ▶ Police Self-Legitimacy
- ▶ Problem-Oriented Policing
- ▶ Procedural Justice, Legitimacy, and Policing
- ▶ Reassurance Policing and Signal Crimes

Recommended Reading and References

Bayley DH, Mendelsohn H (1969) Minorities and the police: confrontation in America. Free Press, New York

Bell J (2004) The police and policing. In: Sarat A (ed) The Blackwell companion to law and society. Blackwell Publishing, Blackwell Reference Online. Retrieved from http://www.blackwellreference.com/subscriber/tocnode?id=g9780631228967_chunk_g978063122896710

Bellman A (1935) A police service rating scale. J Crim Law Criminol (1931–1951) 26(1):74–114

Bennett S, Denning R, Mazerolle L, Stocks B (2009) Procedural justice: a systematic literature search and technical report to the National Policing Improvement Agency. ARC Centre of Excellence in Policing and Security, Brisbane

Daly K, Hayes H, Marchetti E (2006) New visions of justice. In: Goldsmith A, Israel M, Daly K (eds) Crime and justice: a guide to criminology, 3rd edn. Lawbook Company, Sydney, pp 439–464

Decker SH (1981) Citizen attitudes toward the police – a review of past findings and suggestions for future policy. J Police Sci Adm 9(1):80–87

Fielding NG (2002) Theorizing community policing. Brit J Criminol 42(1):147–163

Hails J, Borum R (2003) Police training and specialized approaches to respond to people with mental illnesses. Crime Delinq 49(1):52–61

Hinds L, Murphy K (2007) Public satisfaction with police: using procedural justice to improve police legitimacy. Aust N Z J Criminol 40(1):27–42

Maher L, Dixon D (1999) Policing and public health: law enforcement and harm minimization in a street-level drug market. Br J Criminol 39(4):488–512

Marshall TF (1999) Restorative justice: an overview. Home Office Research, Development and Statistics Directorate, London

Mastrofski SD, Snipes JB, Supina AE (1996) Compliance on demand: the public's response to specific police requests. J Res Crime Delinq 33(3):269–305

Mazerolle L, Bennett S, Davis J, Manning M (2013) Legitimacy in policing. The Campbell Collaboration Library of Systematic Reviews. http://campbellcollaboration.org/lib/project/141/

McCluskey JD (2003) Police requests for compliance: coercive and procedurally just tactics. LFB Scholarly Publishing, New York

Moore DB, O'Connell T (1994) Family conferencing in Wagga Wagga: a communitarian model of justice. In: Alder C, Wundersitz J (eds) Family conferencing and juvenile justice: the way forward or misplaced optimism? Australian Institute of Criminology, Canberra, pp 15–44

Parratt SD (1938) How effective is a police department? Ann Am Acad Pol Soc Sci 199:153–164

Reiner R (1985) The politics of the police, 3rd edn. Oxford University Press, Oxford

Reiss AJ (1971) The police and the public. Yale University Press, New Haven

Skolnick JH, Bayley DH (1988) Theme and variation in community policing. Crime Justice 10:1–37

Sunshine J, Tyler TR (2003) The role of procedural justice for legitimacy in shaping public support for policing. Law Soc Rev 37(3):513–548

Thibaut JW, Walker L (1975) Procedural justice: a psychological analysis. Lawrence Erlbaum Associates, Hillsdale

Tyler TR (1990) Why people obey the law. Yale University Press, New Haven

Tyler TR (2003) Procedural justice, legitimacy, and the effective rule of law. Crime Justice 30:283–357

Tyler TR (2004) Enhancing police legitimacy. Ann Am Acad Pol Soc Sci 593:84–99

Tyler TR (2006) Psychological perspectives on legitimacy and legitimation. Annu Rev Psychol 57:375–400

Tyler TR, Fagan J (2008) Legitimacy and cooperation: why do people help the police fight crime in their communities? Ohio State J Crim Law 6:231–275

Tyler TR, Huo YJ (2002) Trust in the law. Russell Sage, New York

Tyler TR, Lind EA (1992) A relational model of authority in groups. Adv Exp Soc Psychol 25:115–191

Weisburd D, Braga AA (eds) (2006) Police innovation: contrasting perspectives. Cambridge University Press, Cambridge

Weisburd D, Eck JE (2004) What can police do to reduce crime, disorder, and fear? Ann Am Acad Pol Soc Sci 593(1):42–46

Wells LE (2007) Type of contact and evaluations of police officers: the effects of procedural justice across three types of police–citizen contacts. J Crim Justice 35(6):612–621

Policing

▶ Drug Enforcement
▶ Privatization of Policing in an International Context

Policing in Britain c.1750–1829

David G. Barrie
Discipline Group of History, The University of Western Australia, Crawley, WA, Australia

Overview

The last few decades have seen a dramatic change in the manner in which police history in Britain has been analyzed and interpreted and in the level of interest it has provoked. Once the almost exclusive preserve of former officers and civil servants, the development of the so-called "new" police in the nineteenth century has increasingly drawn the attention of professional scholars from the fields of history, sociology, criminology, and legal studies. In so far as the police serve as a mechanism for enforcing the law and utilizing state force, the emergence of the "policeman state" must be regarded as one of the most profound institutional developments in recent history (Gatrell 1992). The introduction of modern policing was crucial not just to the evolution of modern criminal justice systems but also for developing an understanding of the construction, distribution, and operation of power – social, racial, and gender – in government and society (Reiner 2000; Westmarland 2002). As this entry shows, the fundamentals of modern policing had been established in Britain long before central government introduced the 1829 Metropolitan Police Act – once widely regarded as representing the birth of modern policing (Critchley 1967).

Key Issues in Police Historiography

In older, institutional police histories, the 1829 Metropolitan Police Act – which established a full-time, uniformed police force to prevent crime in the metropolis of London – was a watershed in law enforcement (Ascoli 1979). Introduced in response to an unprecedented rise in crime, the inefficiency, inadequacy, and corruption of existing law enforcement arrangements and the pioneering influence of police reformers, the "new" police that took to the streets were a marked improvement on their predecessors and quickly became an example for provincial centers, counties, and other countries to follow. What distinguished the new police from their predecessors, according to this view of police history, was the high standards of professionalism and efficiency to which they were subject: Metropolitan police officers were carefully selected, full-time, salaried, organized, and disciplined, while law enforcement arrangements were rationalized and coordinated. Officers were the embodiment of a "new science of policing," charged with preventing crime through their uniformed presence and a beat patrol system of policing rather than merely responding to it. In such histories, police reform is viewed in a favorable, inherently progressive light – as an unproblematic, linear development beginning with the earliest police reformers, like Henry Fielding and Patrick Colquhoun, and culminating with the architect of the Metropolitan Police Act, Sir Robert Peel.

In the 1960s and 1970s, this conservative, consensual model of police history was challenged by a revisionist interpretation (Silver 1967; Storch 1975). Far from being a marked improvement on their predecessors, the "new" police were shown to have suffered from the same problems and weaknesses as the "old" police, as chief constables struggled to instill discipline and professionalism in a profession undermined by high turnover rates and low pay (Taylor 1997). What was "new" about the new police introduced in London in 1829 and then in provincial towns and counties over the next two decades was not the police's enhanced efficiency, discipline, and professionalism but rather their role and function in society (Hay and Snyder 1989). The new police represented a new, bureaucratized form of social management designed to penetrate civil society with the values of the state through continuous surveillance of working-class society. Influenced by Marxist thought, this view argued that the new police were introduced to meet the needs of a class-based capitalist society and to respond to threats of political unrest and industrial militancy. They were the product not of growing criminality but rather of class antagonisms, changing middle-class perceptions of the sprawling masses, greater demands for urban order, and the erosion of traditional forms of social control that accompanied the transition to urban, industrial society. Fighting crime, in other words, was less important to police reformers than keeping the lower orders disciplined and under close surveillance, and ultimately safeguarding private property, the status quo, and the interests of one section of society over the other.

In recent years, a plethora of studies has nuanced and challenged both of these "problem-response" interpretations of police history. Scholars have questioned whether the police were introduced on a scale capable of disciplining and controlling the working classes, the degree of consensus that existed among men of property as to the merits of reform, and just how far the "new" police were introduced in response to the failures of the old system. Indeed, much scholarship has pointed to the degree of innovation in law enforcement *prior* to the so-called birth of modern policing in London in 1829. Far from heralding a new style of policing, 1829 was, in many ways, the culmination of a long period of reform from at least the mid-eighteenth century in London and provincial English centers. Moreover, the major cities and towns of Ireland and Scotland have been shown to have been at the forefront of innovation, introducing police acts some time before their larger and more widely acclaimed neighbor (Palmer 1988; Barrie 2008). Predating, in many cases, the full-scale introduction of capitalist working practices and industrialization, advances in law enforcement before

1829 point to the need to appreciate more fully broader social, demographic, intellectual, gender, and civic influences behind the birth of modern policing than the pioneering traditional/consensual and revisionist/conflict models – for all their merits – have hitherto done. Recent studies have also situated reform within wider changes to criminal justice systems, addressed the impact of Enlightenment thought, and emphasized the multi-faced nature of police work and the differing police typologies under which police officers function (See, for instance, Emsley 2007).

Fundamentals of Premodern Law Enforcement

There was no single model of law enforcement in mid-eighteenth-century Britain. There were variations between and within national borders, which were determined by long-standing national and local charters and customs, demography, and urban–rural practices. Differing legal systems also affected how the law was enforced. Scotland, for instance, had a public prosecutor who would carry out preliminary investigations of reported crimes to establish whether a case was worth prosecuting; in England and Wales, the overwhelming majority of prosecutions were brought by victims themselves. There were also, however, a number of similarities in the different parts of the British Isles. In most small parishes, amateur parish constables carried out policing responsibilities under the authority of local justices and magistrates. Drawn predominantly from men of the middling ranks, and elected by local boards and trustees, parish constables were expected to serve for 1 year and would often carry out their duties on a part-time basis in conjunction with other employment. They received no formal training or, in most parishes, annual salary, although many constables were able to make a decent living by collecting rewards and expenses. Those who refused to serve as constables were expected to pay for a substitute or face formal prosecution. Constables tended to be reactive rather than proactive. They would not patrol or seek to prevent crime by their visible presence and would only carry out extensive criminal investigations if the victim was willing, and had the financial means, to pay for it. In most cases, victims would have to take the time and trouble to investigate crime and identify offenders themselves before constables would get involved.

In urban centers, the office of constable was sometimes supplemented by watch forces that would patrol the streets at night. In certain parts of the country, local male householders would, like the office constable, perform this task on an unpaid basis as part of their civic duty to the local community. The system, though, was inconsistently and unevenly applied, with local justices invoking ancient statutes that governed “watching and warding” during periods of heightened tension with law and order. In the large centers, local authorities, or men of property, sought to put arrangements on a more permanent footing by employing paid watch forces. The various parishes and districts of London in the late eighteenth century employed several thousand watchmen between them (Paley 1989; Reynolds 1998), Dublin several hundred (Palmer 1988), and Edinburgh relied upon an Old Town Guard to police the dimly lit nighttime streets of Scotland’s capital. The role of such forces, though, was fairly limited: they would protect property, assist magistrates and constables in the exercise of their duties, and apprehend suspicious persons or those committing public order offenses, but they did not investigate and detect crime.

Influenced by the eighteenth- and early nineteenth-century criticisms of police reformers, older police histories perceived such arrangements as being inefficient, inadequate, wretched, and, in some cases, corrupt. It is certainly the case that many watch forces were staffed by old, sometime infirm, ex-soldiers and struggled to attract good recruits because of low pay. It is also the case that some constables, such as the notorious Jonathan Wild in London, were involved in criminal activities. Moreover, as constables and watchmen lived in the communities they served, there was often a reluctance to enforce unpopular legislation or laws that would provoke the wrath of the general populace. Whether or not a crime was investigated, or indeed taken seriously, by local law enforcement officials was often dependent upon the character, social background, and gender of the victim; the nature of the alleged offense; and, above all, the capacity of the victim to fund a criminal investigation and prosecution.

For the most part, though, mid-eighteenth-century policing arrangements functioned adequately. Far from being unreliable and of suspect character, research has shown that the majority of constables were sober, conscientious, and civic-minded individuals who took their responsibilities seriously (Morgan and Rushton 1998). Although in principle an amateur system of law enforcement, arrangements were not necessarily unprofessional in their workings. In some parishes, the position of constable appears to have become almost semipermanent and salaried. Constables were often retained in office after their annual service was up, thereby increasing their experience and expertise. As John Beattie's work on the policing of the City of London, 1660–1750, has shown, the authorities pioneered a number of policing initiatives in an effort to turn this area of historic civic importance into a prosperous symbol of security in an expanding commercial world (Beattie 2001). A perceived rise in property crime, sensationalized in a burgeoning print culture, combined with a desire better to control and regulate the "night" heralded significant reforms to the quality of London's watch system. Moreover, in the overwhelming majority of Scottish towns, local elites had such confidence in existing policing arrangements that they excluded watching and constabulary provisions from local police and improvement legislation for the bulk of the eighteenth century.

To the pioneering scholars of police history, writing over 200 years later, such arrangements appeared backward and inefficient when compared to modern policing practice, but they were perfectly suited to meet the needs of the time, the existing social and religious power structures, and the ethos of criminal justice systems. This was a system of law enforcement that was valued by local justices because it was cheap to administer and it conveyed upon all the legal participants a degree of discretion and influence. While a number of histories have shown that all sections of society made use of local courts (King 2000; Gray 2009), it was also common for disputes to be resolved informally or through extra-judicial means. Indeed, recourse to the law was often the last resort – and an unaffordable and burdensome one at that. This was especially the case in Scotland where local criminal courts in many parts of the country would sit infrequently and at irregular intervals (Barrie and Broomhall 2012a). Indeed, Scots were more likely to be brought before a Church Court – where the local Kirk Session would be actively involved in regulating the behavior and morality of parishioners – than a criminal court in the eighteenth century.

Judicial discretion and exemplary punishment, rather than certainty of prosecution, were the hallmarks of the unreformed eighteenth-century criminal justice systems in both Scotland and England. In pre-Enlightenment Britain, punishing a small number of offenders in a severe, exemplary manner was, it was believed, the best safeguard against future criminality. This, along with a long-standing unwillingness to pay for full-time police forces that might threaten individual liberty, and the status, power, and authority of local elites (Philips 1980), underpinned the workings of eighteenth-century law enforcement. Men of property would respond to periods of disorder and rioting through paternalism or by recourse to military or auxiliary forces, or to periods of heightened concern with crime by voluntarily subscribing to watch forces or associations for the prosecution of felons. But, for day-to-day law enforcement, the mid-eighteenth-century system of parish constables delivered what it was expected to – it responded to reports of crime and brought suspects before local justices, rather than seeking to prevent, investigate, and, in most cases, prosecute crime.

This was a system of policing, though, that was suited to the modest expectation of law enforcement at that time and to small, sparsely populated preindustrial parishes. It was not a system that was able to meet the changing needs of a rapidly urbanizing, commercial society, the challenges brought about by large-scale political protest and unrest, and a growing desire among the middle ranks for better protection of private property and uniformity in practice and performance. As the subsequent section examines, these social, economic, intellectual, and political drivers would stimulate a wide array of local policing innovations and experiments throughout Britain, which would ultimately lay the foundations for greater central government direction from London in the nineteenth century.

Key Issues: Policing Innovations and the Drivers Behind Reform, c.1750 to 1829

Given its size and importance as Britain's leading center of commerce and governance, it should come as no surprise that London pioneered innovations in law enforcement on a more regular basis from the mid-eighteenth century onward. John Beattie's recent work on the Bow Street Runners has shown that the system of criminal investigation established in 1750 by the Middlesex magistrate Henry Fielding, and later by his half-brother John Fielding, developed the Bow Street Magistrates' Office into a leading center for policing and prosecution in the metropolis (Beattie 2012). The officers appointed, under the control of Bow Street magistrates, were detectives in all but name, highly skilled in investigating crime, in hunting down and arresting offenders, and in securing convictions in London's main criminal court, the Old Bailey. Introduced in response to ongoing concerns about property crime, the work of the Fieldings, and the officers they (employed) was important to the way that policing in London developed over subsequent decades. It helped to spread the idea that crime could be reduced through efficient and quick detection, and in doing so, it offered an alternative vision of how to control crime other than through exemplary punishment (Rawlings 2002, p. 96).

In provincial areas, criminal detection in the second half of the eighteenth century was facilitated by the growing use of thief takers, who would pursue suspected offenders for financial reward (for more on the issues discussed in this paragraph, see the collection of essays in Hay and Snyder 1989). Also important was the rise in print advertising offering rewards for information on crime, which was greatly facilitated by the expansion of the provincial press, and the spread of associations for the prosecution of felons, which raised funds through subscriptions to defray the costs of investigating and prosecuting crime. These initiatives, which often developed in response to perceived rises in criminal activity, were symbolic of the fact that among the rising middle ranks there was a growing intolerance of property crime and a growing willingness to seek legal redress rather than settle matters informally. The reimbursement of expenses for successful prosecutions in England in the mid-eighteenth century provided a further stimulus to legal action, while the courts' declining use of capital punishment relative to the number of capital convictions in the second half of the eighteenth century – and the emergence of new forms of punishment such as transportation – encouraged more victims to seek legal redress safe in the knowledge that they would not have to carry the burden of having someone executed for a relatively minor crime.

These developments were carried out against an intellectual backdrop that was slowly beginning to question not only traditional forms of policing and law enforcement but also the ethos and purpose of the criminal justice system itself and how its institutional structures should function. Arguably, the most famous treatise of the time, the Italian philosopher Beccaria's essay *On Crimes and Punishment* (1767), argued that certainty rather than severity of punishment was a more effective deterrent to crime (Draper 2000). Beccaria's theories were quickly incorporated into English penal theory debate and utilitarian thinking and permeated to some degree middle-class discourse on the hallmarks of civilized society. Although by no means universally accepted, his ideas, and the way in which they were picked up and developed by police and penal reformers in England, helped to stimulate debate and lay the intellectual foundations for new thinking on crime control. Ideas about the merits of preventative policing and controlling crime through patrol and surveillance started to become more popular in London in the 1780s (Beattie 2012). This new wave of thinking was perhaps best exemplified by Patrick Colquhoun's *Treatise on the Police of the Metropolis* (1796), which went through seven editions by 1806 and stimulated discussion on policing in the capital. In it, Colquhoun advocated a "new science" of policing, shifting emphasis away from apprehending and severely punishing criminals once a felony had been committed to preventing it from happening in the first place. He called for the creation of a centralized police force in London to prevent and detect crime, as well as the establishment of a public prosecutor in order to relieve victims of the trouble and expense of prosecuting criminals. Ward constables were to maintain a "general superintendence" of their districts, and there was to be a separation between police and judicial powers. Crime, he argued, was a product of indigence, ignorance, and immorality and needed to be tackled through the poor law and education, as well as police reform.

However, Colquhoun's plan was never fully implemented due in no small part to local opposition to his centralized model and the threat it would pose to the power and status of vested propertied interests and because (as is revealed below) the existing system was working much better than Colquhoun claimed (Gray 2009). Nonetheless, his influence was much more important than recent historians have acknowledged. *Treatise on the Police* was widely and enthusiastically acclaimed in the press and Parliament, had a significant impact upon middle-class thinking, and put the issue of police reform on the political agenda. Not long after it was published, Colquhoun was invited by local businessmen to establish the privately funded Thames River Police to prevent dockland crime – a topic close to Colquhoun's heart from his time in Glasgow, where he established the world's first Chamber of Commerce in 1783. The Thames scheme was established in response to a perceived rise in dockland crime, but its creation was also indicative of the crucial role of the redefinition of crime and the reformulation of employer-employee relations played in shaping police reform. The Thames police officers were largely engaged in clamping down on the customary practice of taking perquisites from workplaces as a supplement to their wages – a practice that was being challenged in business circles and by business leaders seduced by the free-market notions of Enlightenment thinkers, not least another influential Scot, Adam Smith, in his famous *Wealth of Nations* (1776). The "new science of policing" that was slowly evolving was not just about crime and its control; it was about responding to wider structural changes in the economy and society and to the interests of the propertied.

This was a period of significant innovation in policing in the metropolis. In the early 1790s, seven police officers under the control of stipendiary magistrates were established in Middlesex, which was an important milestone for expanding the capacity of summary justice to deal with more offenders and forged an important link between the police, reported offenses, and the criminal courts (Paley 1989). According to Gray, city officials, and not just the Bow Street Runners, became more proactively engaged in investigating and detecting crime in this period (Gray 2009). Moreover, night watch forces – both in London and the larger provincial towns of England – were professionalized over the course of the late eighteenth and early nineteenth centuries in response to mounting fears of property crime and disorder, especially during the periods of economic downturn that followed the end of the American Revolutionary (1775–1783) and the Napoleonic (1793–1815) Wars when concerns with law and order intensified. Far from being an inadequate, wretched mess as police reformers argued, many of these forces not only performed preventative police functions through beat patrolling – the hallmark of Peel's so-called new police – they also introduced a tiered command structure and put in place a plan for the "new" model which would be initiated in 1829. Indeed, according to Reynolds, Sir Robert Peel's 1829 Metropolitan Police Act centralized policing arrangements in the capital rather than created them (Reynolds 1998). Modern policing in the capital therefore evolved out of the old watch system through the efforts of local officials rather than politicians in central government. Moreover, as Palmer has pointed out, Peel's blueprint for the Metropolitan Police drew heavily on his years serving as secretary of state for Ireland, where full-time, professional, armed policed forces were introduced in Dublin in 1786 and in provincial Ireland in 1787 in response to concerns with Irish nationalism (Palmer 1988).

Urban and commercial expansion, allied with an increasingly free-market economy, not only loosened the social and paternal bonds that had provided an important safety valve and form of control in many parts of preindustrial English society, it also provided an environment in which more crime, immorality, and drunken disorder could flourish. However, whether crime was actually on the increase in the late eighteenth and early nineteenth centuries is debatable. There was certainly a disproportionate rise in recorded indictable property crime relative to population growth in this period, but this is likely to have reflected to some degree the growing willingness of victims to prosecute due to legal and administrative reforms, changing expectations of law enforcement and a hardening attitude among the emerging middle ranks toward property offenses (Gatrell 1992). Moreover, the expansion of middle-class newspaper readership and the annual publication of criminal statistics from the early 1800s fuelled middle-rank concerns about crime and its control (Lemmings and Walker 2010). They made crime appear more visible and helped to strengthen not just the resolve of police reformers to implement change but also the case for it – something that was ruthlessly exploited by Peel in London and provincial police reformers throughout England. As Emsley has pointed out, it was the perception that crime was on the increase that was important, not whether it actually was or not (Emsley 1996). At the same time, social and cultural habits were changing in the late eighteenth and early nineteenth centuries. Towns, especially those in the first phase of industrial expansion, were becoming larger, more rowdy, and more difficult to control at a time when the rising middle classes were demanding higher levels of urban order and discipline from the expanding urban masses. It was a demand for order in civil society that would result in watch and police forces being introduced and professionalized across in the great municipal towns of England and time-honored, popular working-class pastimes criminalized as a desire to control public space and the built environment became a preoccupation for men of property.

Scotland, with own distinct legal system and legislative tradition, was also at the forefront of policing innovations. In the 1780s, the rapidly expanding mercantile and manufacturing center of Glasgow was home to Patrick Colquhoun, who served as Lord Provost of Glasgow between 1782 and 1783 and as a Lanarkshire Justice of the Peace for much of the 1780s. Protecting private property and controlling urban rioting were a major concern for Colquhoun and his civic counterparts, and a number of policing proposals were put forward and experiments implemented to effect these aims. In 1788, magistrates appointed an "intendent of police," a police clerk, and a number of "police officers" "in order to detect and prevent crimes" and divided the city of Glasgow into nine districts, each with "ward superintendents" – a concept that would resurface in Colquhoun's *Treatise*. The initiative was innovative and forward thinking in seeking to prevent crime by establishing a degree of community supervision over local residents and was indicative, in many ways, of Colquhoun's notion (espoused in his famous *Treatise*) of establishing a "general superintendence" of local districts. Its underlying premise was that a successful scheme of police required locally appointed ward superintendents with intimate knowledge of suspicious characters and crimes. The scheme coincided with mounting concerns about crime, vagrancy, and rioting in the city, but underlying them was a growing recognition that a more professional system for investigating and detecting crime was necessary for such a large and prosperous city. It was accompanied by a sustained campaign to acquire a statutory local police act which illustrated just how far the new thinking of preventative policing had permeated the discourse of the city’s business community. In discussing a proposed police bill in 1789, the Trades’ House – which represented craft interests in the city – argued that the law enforcement clauses would prove ineffective as they were designed only “for the detection and punishment of crimes once committed” and fell “exceedingly short of that much more important object of preventing the increase, and removing the possible cause of the crimes” (Barrie 2008) The ensuing discourse on law enforcement in the city exposed not only a growing intolerance of crime but also the growing problems of investigating and detecting it. There was growing resentment in middle-rank circles that victims were expected to take the initiative in bringing legal action against offenders, and new ways were being sought to defend private property more effectively by making it easier to bring criminals to justice.

In his famous *Treatise*, Colquhoun drew inspiration from policing systems in France and Dublin, as well as his experiences serving as a magistrate in London. It is also possible, though, that developments in his native city provided an intellectual influence on his thinking. Colquhoun came from a background that was questioning the existing system of criminal investigation and prosecution and looking at new ways better to protect private property. A number of the ideas he expressed in *Treatise* bore a striking resemblance to the political discourse surrounding police reform in Glasgow in the late eighteenth century and to a wider intellectual culture of Scottish Enlightenment thought on the rule of law, jurisprudence, and civic governance. How much influence Colquhoun was able to exert over policing arrangements in Glasgow is impossible to say. What is clear, though, is that Colquhoun’s self-proclaimed “new science of policing” was being discussed and debated at length in middle-rank circles in Glasgow several years before it made its appearance in *Treatise*, which raises the possibility that his Scottish heritage and the intellectual and philosophical discourse which emanated from the Scottish Enlightenment influenced, to some degree, subsequent thinking in London.

In 1800, Glasgow finally acquired its first police act after years of protracted debate and political infighting among urban elites over issues of control and cost. The city’s lead was followed by a large number of Scottish burghs, including Greenock (1801), Edinburgh (1805), Paisley (1806), and Dundee (1824). Indeed, 23 Scottish burghs in total introduced police acts between 1800 and 1826. These statutes were similar to, and influenced by, eighteenth-century English improvement legislation in that they covered a wide array of powers relating to the general well-being and good ordering of local communities and included provision for full-time, paid watch forces. They did not represent the triumph of a new police model but rather the expansion of a model that had been introduced in English towns and was familiar on continental Europe (Dodsworth 2004). “Police” in early nineteenth-century urban Scotland, in keeping with the concept’s European origins, was defined broadly as a form of civic administration for the common good. The intention behind these acts was to levy assessments, elect commissioners, enlarge judicial boundaries, and obtain and extend powers for regulating the civil and criminal affairs with the intention of promoting cleanliness, health, security, and good order.

However, although public amenity provisions featured prominently in these acts, it would be wrong to regard them as having little significance for the evolution of modern policing or being merely an offshoot of English practice. In many cases, provision was also made for police forces under a hierarchical command structure for the prevention of crime under the management of locally elected police commissioners. This new, bureaucratic structure of law enforcement was an important step in the professionalization of law enforcement and in associating “police” with crime control. Moreover, in the larger cities, police courts were set up to expand the capacity of summary justice with senior police officers taking on the role of public prosecutor for minor crimes, offenses, and misdemeanors – a role the “new police” in England would not take on until the second half of the nineteenth century (Barrie and Broomhall 2012a). These courts came to be integral to the smooth running of towns, the maintenance of urban order, and the development of the Scottish criminal justice system. They were the principal local courts before which people were likely to be summoned and interact with the law and were instrumental in bringing about a massive increase in the number and type of crimes and offenses that were dealt with and punished in a summary manner. They were therefore an essential component of the emergence of the "policeman state" (Gatrell 1992).

The economic and social profile of "improving" burghs consisted mainly of expanding manufacturing centers, such as Glasgow, Paisley, and Airdrie; buoyant sea ports, such as Port Glasgow and Leith; and flourishing commercial centers, such as Dumfries. Many had western coastal locations, which suggest that police reform in urban Scotland was closely associated with urban growth, demographic change, and the development of trans-Atlantic trading. As in England, urban growth and changing work practices brought growing concerns with crime and its control. There were acute periods of heightened tension with property crime, vagrancy, and urban order, especially during periods of economic downturn. The post-Napoleonic years were also marked by political radicalism and industrial militancy. Relations between employers and employees became increasingly fraught in the face of changing work practices and declining wage levels, which accompanied the move toward a more market-orientated system of production. As in England, middle-class attitudes toward crime and the expanding, industrial urban masses were hardening. Men of property were becoming increasingly intolerant of urban disorder – partly because it was becoming more prevalent as rapid urban expansion and economic takeoff in expanding Scottish burghs brought more and more migrants and social problems and partly because an effective system of police was deemed essential to the progress of civil society. The 1800 Glasgow Police Act, for instance, was introduced not long after a major disturbance in the city, while the 1805 Edinburgh Police Act of 1805 was framed within the context of heightened anxiety over public safety, growing intolerance of public begging, and mounting criticism of the problems victims faced in bringing offenders to justice.

The inclusion of watching provisions in local police acts, however, also reflected an expansion of late-night entertainment and a growing reluctance among men of property to carry out their time-honored watching duties. Indeed, some police acts included a provision for watching by proxy in lieu of personal attendance. Other towns embraced reform because their neighbors had done so. What is particularly striking about the pattern of police development in Scotland in the first few decades of the nineteenth century is its clustered nature. Burghs within the same region tended to introduce improvement and police legislation around the same time. Acquiring statutory police powers set new standards that were likely to have made provisions in neighboring towns appear backward, as what was once acceptable quickly became unacceptable in light of competing models. Indeed, far from London providing the main driver for reform, provincial elites throughout the United Kingdom often looked to each other's models for inspiration. It was not uncommon for improvement commissioners in English northern towns to look to Scottish practices or police commissioners in Glasgow and Edinburgh to look to policing arrangements in Manchester or Liverpool whenever the issue of reform was on the agenda. Indeed, in 1806, commissioners in Edinburgh ruled out looking at London as a model for police reform on the basis that its size and significance were unique in a United Kingdom context.

There were similarities here with the policing of rural England. As Philips and Storch have shown, traditional forms of law enforcement in the second quarter of the nineteenth century came to be viewed as inefficient and ineffective when compared with the Metropolitan Police, and in the face of mounting criticism of its deficiencies in government circles, even though, in practice, they were functioning fairly effectively (Philips and Storch 1999). Luddite disturbances, rural riots, the erosion of paternal authority, a perceived rise in crime, and changing expectations of the criminal justice system caused men of property in rural England to think of better ways to defend their property and resulted in the introduction in the 1820s of new policing initiatives based on the underlying principles and ethos of the old constabulary system (Philips 1980). Indeed, far from being a radical departure from what had gone before, opposition to the introduction of the new police from the propertied ranks in rural England in the 1830s and 1840s was somewhat reduced because it was not fundamentally new but rather a product of earlier innovation and experimentation. In many English towns and counties, the "new police" were often staffed by the same men who had served in the "old police," while in other parishes, constables continued to play an important role in the apprehension of criminals long after their so-called demise. The transition from "old" to "new" in rural England after 1829 was not, therefore, necessarily a product of the failure of the system but rather has to be seen in the context of long-term transformations of the local state and a changing social outlook and administrative philosophy of the ruling elite – a philosophy which prioritized a bureaucratic system of stipendiary magistrates and paid police in order to meet the new expectations of law and order associated with the Metropolitan Police and to respond to heightened sensitivity about crime, order, and public decorum.

Therefore, even although it is difficult to define exactly what constitutes modern policing, it is highly unlikely to have begun in 1829. Many of its recognizable characteristics – such as crime prevention and detection, police-led prosecutions, and police courts to administer summary justice – were in operation in different forms and in different parts of mainland Britain long before Peel's police officers took to the streets. Modernization had a long history and many layers – a history of debate, experimentation, and reform which is extremely difficult to periodize in any precise manner and which did not evolve in a vacuum from wider thinking of other branches of the legal system. Police culture was, and indeed continues to be, heavily influenced by traditions, customs, and practices (Williams 2010).The 1829 Metropolitan Police Act was still of historic importance and would, according to Philips and Storch, become the main model for law enforcement against which others were measured in England following its introduction (Philips and Storch 1999). But it was not the historic turning point that was once claimed in traditional/conservative histories, especially in Scottish towns where it had little significant impact at all beyond providing – in the form of the Metropolitan Police – a pool of officers to recruit when required.

In the large, expanding commercial and industrial cities of mainland Britain, structural changes within society and the economy, and concern with industrial militancy, disorder, and radicalism, would come to play an increasingly important role in strengthening the resolve of men of property to improve law enforcement arrangements as the nineteenth century progressed (Carson and Idzikowska 1989; Storch 1975). But, there were sufficient innovations in nonindustrial areas to caution against overstating social-control, class-conflict revisionist models that locate the emergence of police to the problems thrown up by industrial society. In some of the smaller Scottish burghs, police reform, and the wide-ranging municipal powers that were acquired under police legislation, was indicative of changing attitudes associated with civic consciousness and urban refinement. Indeed, for a town to seek statutory policing powers was by no means always an admission of social problems; in some cases, it was indicative more of the concept of a new civil society beneath which many forms of improvements could be introduced and administered. Similarly, eighteenth-century English improvement commissions, with their wide-ranging municipal powers, were rooted in Enlightenment notions of improvement for the common good, even if, as in Scotland, men of property had a self-serving notion of what constituted the common good. This points to the need to locate police reform in British cities more fully within the context of the long history of municipal governance, Enlightenment thought, and changing attitudes toward, and expectations of, the built environment and the criminal justice system and what these should deliver, rather than merely the viewing it as being the knee-jerk response to a specific problem. Concerns with crime, disorder, and unrest might have helped strengthen the resolve of men of property to introduce reform, but, in many cases, it had been on the local agenda for a considerable period of time and for a variety of different reasons.

Conclusions: Future Directions on the Evolution of Modern Policing

While recent years have witnessed the emergence of major studies of policing in nineteenth-century urban Scotland and provincial England, more attention needs to be devoted to the policing of eighteenth-century English towns, rural Scotland, and Wales (which has been well served by David Jones's studies on nineteenth-century police and crime but less so for the eighteenth century; Jones 1992). Such concentration will not only address major gaps, it also has the potential to further understanding of an historiography (pre-1829) that has been largely preoccupied with London and which might have skewed a wider national picture. It is certainly the case that London's problems had a disproportionate role in shaping national policy and government thinking on English criminal justice matters, but eighteenth-century provincial towns and cities were by no means incapable of independent action. Indeed, as Peter King recently pointed out in his study of the English criminal law and courts, eighteenth- and nineteenth-century justice was often made from the margins by local elites rather than by Parliament in London (King 2007). Throughout Europe, states and cities studied each other's police systems (Emsley 2007), but how much intellectual transfer there was in municipal and rural policing in eighteenth century in Britain is yet to be fully investigated. As in the early nineteenth century, it is possible that civic rather than metropolitan emulation was a fundamental driver of reform – a fact which has significant implications not just for police reform per se but also for a wider understanding of the construction and dissemination of police typologies, civic governance, and the relationship between the center and the localities.

While a number of studies have explored the relationship between the eighteenth-century summary courts (Gray 2009), much more, surprisingly, is needed on the police courts that were introduced in the nineteenth century. What work has been done in a British context has, with a few exceptions (Barrie and Broomhall 2012a; Godfrey 2008), been London centered, with attention focusing on the reform to the Bow Street Police Offices in the 1790s and the extent to which London police courts provided a poor man's system of justice (Davis 1984). But these courts have captured the imagination of historians far less to date than the eighteenth-century English "Bloody Code," despite the crucial role they played in the expansion of summary justice and the police's regulation of nineteenth-century urban societies. Examining the relationship between police courts, police forces, and police magistrates will provide a fuller understanding of the extent to which police development was interconnected with wider changes in the criminal justice system and how the different branches of police regulated urban and industrial society.

Recent years have also seen greater attention being devoted to the historical relationship between police, gender ideologies, and power structures, and especially how they have impacted on police culture, the experience of female officers (Jackson 2006), and how ideas about masculinity have shaped police culture, practice, policy, and institutional organization over time (Barrie and Broomhall 2012b). Yet, much more needs to be done, especially on the ways in which policing reflects, sustains, embodies, and enforces ideas about masculinity given the gender imbalance that still exists within police forces. Indeed, the paucity of studies which have stressed the importance of gender in shaping eighteenth- and nineteenth-century systems of policing relative to class-based interpretations is surprising given that police institutions not only incorporate changing models of male authority but also are closely intertwined with the distribution and operation of power within society. As Barrie and Broomhall have recently argued, a greater appreciation of masculinity has significant implications for the historian's understanding of the birth of modern policing and police typologies. In historical studies of policing that prioritize gender as a key focus of analysis, conceptualizations of "old" and "new" police models cannot be sustained as all models of policing – including the eighteenth-century old parish system and the reformed bureaucratized model that replaced it in the nineteenth century – have linked their justification and practice of control over communities to concepts of masculinity. Indeed, rather than holding up the Metropolitan model as a template for modern policing practices throughout the world as many studies have done, Barrie and Broomhall have stressed the importance of different models in shaping international institutional policing structures and practices – models that were shaped by conceptions of masculinity and which have been subject to change in different historical and geographical settings (Barrie and Broomhall 2012b).

Lastly, critiques of traditionalist and revisionist accounts of police history as outlined above have encouraged many historians to abandon overarching structural interpretations of reform in favor of local variation. There has, in recent years, been a move away from attempting to pinpoint a particular incident, individual, or social interest as the main driver for police reform (Emsley 1996). However, although recent empirical research and local histories have been extremely valuable in highlighting the complex ways that modern policing evolved, there is a danger, as Philips and Storch warn, of failing to give sufficient historical context for wider, national moves for reform and how these relate to wider changes in the economy, society, and governance of British towns and cities (Philips and Storch 1999). Preoccupation with local diversity, while providing a useful corrective to oversimplified explanations, runs a risk of becoming too specific and neglecting not only the wider context in which reform emerged but also the common problems found in different circumstances (Taylor 1997). The major future advancements in the field will not come from scholars who engage in myopic empiricism alone but rather from those who seek to locate their research in wider theoretical frameworks and ideologies of, among others, class, gender, race, and ethnicity; from those who are attentive to national, intellectual and political developments; and from those who seek to explain diversity rather than using it as the basis for abandoning general, theoretically informed explanations.

Recommended Reading and References

Ascoli D (1979) The queen's peace: the origins and development of the metropolitan police, 1829 to 1979. H. Hamilton, London

Barrie DG (2008) Police in the age of improvement. Police development and the civic tradition in Scotland, 1775–1865. Willan Publishing, Cullompton

Barrie DG, Broomhall S (2012a) Public men, private interests? The development, structure and practice of police courts in Scotland, 1800 to 1833. Continuity Change 27(1):449–494

Barrie DG, Broomhall S (eds) (2012b) A history of police and masculinities, 1700–2010. Routledge, Abingdon

Beattie JM (2001) Policing and punishment in London, 1660–1750: urban crime and the limits of terror. Oxford University Press, Oxford

Beattie JM (2012) The first english detectives. The bow street runners and the policing of London, 1750–1840. Oxford University Press, Oxford

Carson K, Idzikowska H (1989) The social production of Scottish policing, 1795–1900. In: Hay D, Snyder F (eds) Policing and prosecution in Britain, 1750–1850. Clarendon, Oxford, pp 267–300

Critchley TA (1967) A history of police in England and Wales, 900–1966. Constable, London

Davis JS (1984) A poor man's system of justice? The London police courts in the second half of the nineteenth Century. Hist J 27(2):309–335

Dodsworth FM (2004) "Civic" police and the condition of liberty: the rationality of governance in eighteenth-century England. Soc Hist 29(2):199–216

Draper AJ (2000) Cesare Beccaria's influence on english discussions on punishment, 1764–1789. Hist Eur Ideas 26:177–99

Emsley C (1996) The english police: a political and social history, 2nd edn. Longman, London

Emsley C (2007) Crime, police and penal policy. European experiences 1750–1940. Oxford University Press, Oxford

Gatrell VAC (1992) Crime, authority and the policeman-state. In: Thompson FML (ed) Cambridge social history of Britain, 1750–1950. Cambridge University Press, Cambridge, pp 243–310

Godfrey B (2008) Changing prosecution practices and their impact on crime figures, 1857–1940. Br J Criminol 48:171–189

Gray DD (2009) Crime, prosecution and social relations in London: the summary courts of London in the late eighteenth century. Palgrave MacMillan, Basingstoke

Hay D (1975) Property, authority and the criminal law. In: Hay D, Linebaugh P, Thompson EP (eds) Albion's fatal tree: crime and society in eighteenth-century England. Allen Lane, London, pp 17–63

Hay D, Snyder F (eds) (1989) Policing and prosecution in Britain, 1750–1850. Clarendon, Oxford

Jackson L (2006) Women police: gender, welfare and surveillance in the twentieth century. Manchester University Press, Manchester

Jones DJV (1992) Crime in nineteenth-century Wales. University of Wales Press, Cardiff

King P (2000) Crime, justice and discretion in England, 1740–1820. Oxford University Press, Oxford

King P (2006) Crime and the law in England, 1750–1840. Cambridge University Press, Cambridge

King P (2007) Crime and the Law in England, 1750–1840: remaking justice from the margins. Cambridge University Press, Cambridge

Lemmings D, Walker C (2010) Moral panics, the media and the law in early modern England. Palgrave MacMillan, Basingstoke

Morgan G, Rushton P (1998) Rogues, thieves and the rule of law. The problem of law enforcement in North-East England, 1718–1800. Routledge, London

Paley R (1989) "An imperfect, inadequate and wretched system?" policing London before Peel. Crim Justice Hist 10:93–123

Palmer SH (1988) Police and protest in England and Ireland, 1780–1850. Cambridge University Press, Cambridge

Philips D (1980) A new engine of power and authority: the institutionalisation of law enforcement in England, 1780-1830. In: Gatrell VAC, Lenman B, Parker G (eds) Crime and the law. The social history of crime in Western Europe since 1500. Europa, London

Philips D, Storch RD (1999) Policing provincial England, 1829–56: the politics of reform. Leicester University Press, London

Rawlings P (2002) Policing: a short history. Willan Publishing, Cullompton

Reiner R (2000) The politics of the police, 3rd edn. Oxford University Press, Oxford

Reynolds EA (1998) Before the bobbies: the night watch and police reform in metropolitan London, 1720–1830. Stanford University Press, Stanford

Silver A (1967) The demand for order in civil society: a review of some themes. In: Bordua DJ (ed) The police: six sociological essays. Wiley, New York

Storch RD (1975) "The plague of blue locusts": police reform and popular resistance in Northern England, 1840–57. Int Rev Soc Hist XX:69–90

Storch RD (1989) Policing rural Southern England before the police: opinion and practice, 1830–56. In: Hay D, Snyder F (eds) Policing and prosecution in Britain, 1750–1850. Clarendon, Oxford

Taylor D (1997) The new police in nineteenth-century England: crime, conflict and control. Manchester University Press, Manchester

Westmarland L (2002) Gender and policing: sex, power and police culture. Willan Publishing, Cullompton

Williams C (2010) Policing the populace: the road to professionalisation. In: Kilday A-M, Nash D (eds) Histories of crime: Britain. Palgrave MacMillan, Basingstoke, pp 1600–2000

Policing in Developing Democracies

Nathan W. Pino
Department of Sociology, Texas State University-San Marcos, San Marcos, TX, USA

Overview

Throughout the global South, countries are struggling to build new democracies. Under the liberal democratic model, the police are regarded as central to the establishment of social order and to the preservation of fundamental democratic rights and freedoms. Unfortunately, developing democracies tend to exhibit certain characteristics that pose challenges for developing democratic forms of policing, and the police in developing democracies often act as they did under their colonial period, engaging in abusive and repressive behaviors that threaten security and political stability. Forces largely outside a developing democracy's control such as neoliberal globalization and the hegemonic behavior of powerful states (involving the domination of weaker states without direct military force) present further challenges for reform and are often not taken into account by academics or practitioners seeking to reform the police in these young democracies. When powerful Western states or organizations attempt to reform police systems in developing democracies, their efforts often fail for a number of reasons, including lack of adequate knowledge of the country receiving aid, the coercive implementation of aid resisted by the recipient country, reforms that benefit the donor more than the recipient, weak recipient capacity or willingness to implement reforms sustainably, and the inadequate participation of local actors in reform planning and implementation. Suggestions provided in the academic literature for overcoming the above challenges are presented and discussed.

Introduction

Since the great wave of democratization in the late twentieth century, many countries are in various stages of democratic development. Democracies at a minimum have a constitution (including uncodified ones as in the United Kingdom), fair electoral systems, multiple political parties, and mechanisms of accountability (Plattner 2008). Developing democracies that survive exhibit good governance in part by implementing sound economic policies that benefit the majority of their citizens, who then view the government as legitimate (Kapstein and Converse 2008). Liberal democracies have certain attributes that set them apart from simple electoral democracies, such as a free press, a respect for civil liberties, and adherence to the rule of law. In order to gain legitimacy in established liberal democracies, the police are expected to be accountable and transparent, uphold the rule of law, preserve social order, protect fundamental democratic rights and freedoms, and understand that their authority derives from their fellow citizens. Developing democracies are also expected by the international community to instill in their police these liberal democratic ideals while changing police policies regarding political activity, political prisoners, and the rights of the accused. Unfortunately, developing democracies tend to have characteristics that present numerous challenges to police development and reform. Democratic societies have an authority granted by law to maintain a democratically determined order, but if the power of the police is misused, this abuse of power is potentially harmful or even destructive to the structure and processes of democratic societies.

Characteristics of Developing Democracies and How They Present Challenges for Policing

Democratic institutions and norms in numerous developing democracies have been eroding. Many countries that gained independence from the barbarism of colonization emerged with illiterate, rural populations with few opportunities for personal advancement (Pinkney 2003). In addition, these transitional and developing societies are experiencing rapid change associated with modernization and globalization, particularly in urban areas. These rapid changes are in part responsible for the numerous social problems these countries face and are less able to contend with, such as economic inequality; rapid rural to urban migration; the influx of refugees; increases in the proportion of adolescent and young adult males in the population; decreases in job opportunity; inadequate health and educational services; rapid changes in family and neighborhood structures; the proliferation of organized crime; and the legacies of ethnic and other forms of conflict (Neild 1999). The police cannot be expected to control crime or contend with these social problems alone, but the presence of these problems in the absence of a strong democratic state increases the likelihood that the police will resort to repressive measures and support their regime rather than their fellow citizens.

Developing democracies tend to suffer from weak democratic institutions, including those meant to institutionalize police accountability and responsiveness, a lack of commitment to the rule of law, corruption, fraudulent elections, high levels of crime and societal instability, significant levels of poverty and inequality, and poor police-community relations (Hinton and Newburn 2009). Many of these countries still have authoritarian provisions in their constitutions with too few safeguards against control by militaries or the elite, and parliaments and legislatures may not act independently from the executive branch. Governments will therefore be less responsive to citizen demands, allow higher levels of police indifference and abuse, and offer fewer chances for economic development or trust in government. In other words, while these countries are engaging in democratic transitions, they are not necessarily ripe for democracy. Countries ripe for democracy possess, among other things, a stable middle class; an active, engaged, and educated civil society; and elite commitment to democratizing institutions (Pinkney 2003).

A number of transitional and developing societies could be considered weak states, which are countries without social cohesion or the state capacity or willingness needed to protect citizens as they live their daily lives (Goldsmith 2003). Weak states can contribute to the erosion of citizen safety by giving regime stability and narrow sectional interests based on ethnicity, religion, or some other category more attention than considerations of general public safety. In post-conflict situations, weak states can collapse without developing democratic institutions when crime rates are high (Neild 1999).

Transitional countries ideally seek to balance the need to encourage economic growth on the one hand and the need to discourage economic concentration and other forms of inequality on the other. Unfortunately, the capacity of emerging democracies to balance these needs is hampered by low levels of education among their citizens and little if any independent development opportunities owing to the marginal positions of these countries in the global economic system. Economic concentration occurs with regularity in these countries, which weakens the capacity and participation of civil society, corrupts politicians, and influences the police to act as protectors of the rich by aggressively policing the poor. Deepening inequality can lead to increases in violent crime and terrorism committed by concentrations of young unemployed males. State sovereignty and capacity are also weakened by the privatization of state enterprises, reduced regulations on business activity, and the ceding of authority to global institutions of governance such as the World Trade Organization. These neoliberal activities allow for the development of illicit economies and opportunities for criminals to participate in government, increasing corruption and weakening legal, judicial, and policing systems (Aas 2007). In highly unequal countries, citizens are more willing to accept authoritarian rule and politicians are more easily able to violate human rights and the law as well as weaken justice systems. In poor countries with high crime rates, the police are politicized and are used to crack down on minority groups and the lower classes (Hinton and Newburn 2009).

Violent and protracted ethnic conflicts are more likely to occur after colonial rule and this is another reason why post-colonial states have difficulty transitioning democratically (Stavenhagen 1996). Police violence, human rights violations, and other forms of police deviance, particularly by special units, increase as a result of these conflicts, as we have seen in Somalia, Cambodia, El Salvador, East Timor, Kosovo, Macedonia, and elsewhere. When ethnic conflict is a central component of social life in developing democracies, ethnicity itself becomes a politically mobilizing force and ethnically based political parties often develop (see Stavenhagen 1996).

High rates of conventional street crime are also typical of developing democracies, particularly in the low-income countries of Africa, Latin America, the Caribbean, and the Middle East, particularly among their minority populations (Neild 2001). Sources of crime in transitional societies are as complex as in developed societies, and internal as well as regional factors must be taken into account when assessing causes of crime in any particular country. Authoritarian elements in government create social control systems designed to maintain the authority and control of the regime, leading to police training that fails to prepare police to contend with conventional street crimes adequately. Increases in crime and feelings of insecurity coupled with a weak or repressive police response can weaken police legitimacy and lead to the privatization of security, death squads, and vigilante actions by citizens (Neild 2001).

Characteristics of the Police in Developing Democracies

Most developing democracies were once colonies, and the legacies of colonialism impact policing just as they impact other social and governmental institutions. While scholars have observed some similarities among colonial police systems, such systems were still unique due to local factors and relations with the colonial state. For example, some police systems were organized around the annexation of territories for the purpose of settlement, resulting in civilian police systems (North America, Australia), while others were developed in order to subdue colonies annexed primarily for trade, leading to paramilitary policing on the continents of Africa, Asia, and Central and South America (Cole 1999). In the latter case, police leaders were often recruited from imperial armies, and police forces and militaries worked alongside one another. Sometimes private companies were established to maintain law and order, defend colonial territories, suppress political activities, supply cheap labor, and to hold rights over land and people.

In post-conflict or post-colonial situations, it is expensive and time consuming to create a new and inexperienced police force from scratch. Therefore, the composition of the police in a transitional period might be composed of former police and military officers from the colonial period as well as members of ex-insurgent or guerrilla forces, some of whom might have committed human rights violations in the past (Neild 2001). In post-conflict environments, institutional weaknesses, authoritarian nationalist cultures, apathetic or weak civil society groups, and the difficulty of imparting skills to new officers all work to hinder the police reform process (Neild 2001). Police institutions are not well defined and the lines separating police and military forces are blurred (Hills 2000). Citizens often view the police as weak and ineffective, and public demands for order can motivate repressive policies that erode citizen rights and militarize the police. In addition, political leaders in many of these countries from previous autocratic periods have converted themselves into supposed democratic leaders during democratic transitions, leading to a continued lack of legitimacy for both governments and their police (Hills 2000).

Regime policing – policing oriented toward the regime rather than citizens – easily flourishes in weak post-colonial states that are under-resourced, and contain authoritarian and paternalistic institutions, including paramilitary police agencies (Neild 1999). During democratic transitions, these post-colonial or post-conflict police agencies tend to shield themselves from democratic accountability, and even though Western countries have provided police reform aid, police systems have generally not evolved in any meaningful way (Hills 2000). Police officers tend to experience low socioeconomic status and are often viewed by citizens as no different from criminals.

Neild (1999) traces how security debates tend to unfold in post-conflict societies. After a period of regime policing as described above, human rights and civil society groups engage in collective action to criticize policing and call for the democratization of policing. In some cases, police reforms following this activity center around building security and stability, but in most unstable democracies, more time is devoted to civil-military relations than to changing police institutions and practices. Next, as crime and crime control become more pressing concerns, the legacies of regime policing reappear in order to crack down on crime, though this threatens the democratic transition process. In response to these previous developments, some countries engage in reform efforts that create new models of democratically based security, but many developing democracies have not yet reached this point, leading to a continuation of the status quo of militarized regime policing, vigilantism, and the proliferation of private policing. Marginalized groups, such as trade unionists, the unemployed, students, children and adolescents, and the homeless, remain targets of police activity. Thus, even after reforms are instituted, demands for order-restoration lead to the return to old practices, and crime rates in the poor countries rise as ineffective and under-regulated police forces engage in old tactics (Neild 1999).

Police-community relations in much of the developing world are characterized by violence and repression, and policing is often used as an instrument of state violence and corruption.

The socioeconomic inequalities in these countries are taken advantage of by state actors as certain sectors of the population are criminalized and therefore brutalized by the state. There is little-to-no community involvement with the militarized police and there are few if any genuine accountability measures. Newly democratic governments, such as those in Hungary, South Africa, and Brazil, have increased rather than decreased police powers such as the ability to detain suspects longer, to search and seize property, and to conduct undercover surveillance (Stone and Ward 2000).

The way that other components of criminal justice systems evolve or get reformed in democratic transitions can also hinder police reform efforts. For example, judicial reform tends to occur more slowly than police reform, and if the justice system fails to engage in legal tactics such as issuing warrants, the police might resort to illegal actions to carry out regular functions (Neild 2001). Various nation-building activities, legal aid, and good governance initiatives have not promoted development or created a legal framework that supports democratic policing or criminal justice reform in these countries (Hills 2000). Police and judicial processes remain abusive, corrupt, and ineffective in the developing world even though many of these countries have gone through democratic transitions. Neild (1999) describes these democratic countries with abusive justice systems and state failure to provide protection and services as uncivil or low intensity democracies.

Ethnic, sectarian, and other schisms negatively impact police behavior, recruitment, and training as well. Numerous ethnic recruitment policies occur around the world, and those organizations that engage in those policies are more likely to have paramilitarized forces, where training is in a military style, numerous rapid deployment forces that protect regime control are operating, and police and military roles are blurred even though civilians tend to head police services (Hills 2000). Centralized police forces are often less able to contend with local problems and priorities, and accountability and effectiveness are made more difficult due to a lack of quality statistical data and inadequate or politically motivated resourcing (Hinton and Newburn 2009).

Given the economic problems these countries face, the police suffer from a lack of resources, faulty-equipment, low salaries, and poor working conditions (Hills 2000; Neild 1999). These conditions undermine morale and self-efficacy (belief in one's competence); and increase the likelihood of corrupt behaviors that weaken their legitimacy (Hills 2000; Uildriks and Van Reenen 2003). Low levels of satisfaction with the police among populations in developing democracies cannot be underestimated. Citizens often find the police illegitimate, unaccountable, ineffective, brutal, and violent. In the developing world, victim dissatisfaction with the way in which police handle reported cases of victimization and the ways in which the police control crime measure higher than in the developed world (Zvekic and del Frate 1995). Police in these countries continue to be distanced from the civilian population, leading to a mutual lack of trust and lowering the chances that genuine democratic policing can emerge or that victims will report crimes. Citizens indicate that their distrust stems from perceptions of corruption, unprofessionalism, unwillingness to protect citizens, and abuse of authority, which limits the willingness of citizens and civil society groups to work with the police (Uildriks and Van Reenen 2003). Human rights groups, for example, are reluctant to work with police forces in ways other than establishing watchdog groups on police abuses. Police officers in turn feel mistrusted by the public, forsaken by the government, and distrusted by their supervisors and other criminal justice system agents.

One major consequence of crime is fear, which has important reciprocal effects on communities and neighborhoods. Violence and the fear of it limits economic and physical well-being, restricts movement and access to jobs and schools, weakens social ties and interaction, deters investment and reduces economic output, weakens legitimacy in social institutions, and threatens the rule of law (Neild 1999). Fear of crime and lack of police legitimacy lead to the increased use of under-regulated private police forces sometimes staffed by demobilized soldiers, which increases inequalities by physically separating the rich from the non-rich and can lead to various kinds of repressive policing (Neild 2001). Private security is pervasive in the developing world and generally has a negative influence on the conventional police and justice system to fulfill their missions. Even public policing can become privatized in the sense that the wealthy and politicians can buy police presence while the poor suffer further as the non-policed areas in which they live become taken over by gangs, death squads and organized crime syndicates, all of which are often linked with the police (Hinton and Newburn 2009).

Other Challenges to Establishing Democratic Policing

Policing scholars and practitioners are often guilty of failing to place policing in developing democracies in its proper historical, economic, political, and social contexts, treating policing as if it occurs in a vacuum and leading to one-size-fits-all reform strategies that frequently fail to meet expectations. There are a number of factors that may hinder a developing democracy's ability to implement a democratic policing strategy. In varying degrees, these factors include underdevelopment, a lack of sufficiently democratic institutions (including criminal justice agencies), a weakened and unempowered civil society, low levels of social and other forms of capital, pressing crime and other security problems, social and cultural conflict, political instability, a lack of legitimacy in social and political institutions, and international pressures that weaken the capacity of states (Pino and Wiatrowski 2006). In addition, pressure from foreign states, a lack of social cohesion, shortages of resources and functioning equipment, and political instability threaten the democratization of policing (Goldsmith 2003; Neild 2001). These various problems promote abusive, corrupt, neglectful, indifferent, and ineffective policing by negatively impacting training, morale, and operations.

Neoliberal globalization, championed by Western capitals and global cities, has had a decidedly negative impact on police reform efforts. In neoliberal thought, markets govern social institutions and a culture of market-oriented behavior in individuals is promoted to weaken the influence of states in the economic and political sphere. Capital flows, labor markets, and international trade are deregulated and then reregulated in a business-friendly way, and social and public services are privatized, enhancing the influence of transnational corporations over state policies (McMichael 2008). The USA has prescribed a set of policies to developing countries, known as the Washington Consensus, which spreads neoliberal practices globally (Williamson 2000). These policies include reducing public expenditures that subsidize social needs such as infrastructure, health care, and education; weakening labor, environmental, and legal protections; liberalizing trade and foreign direct investment; privatizing public enterprises; reducing marginal tax and interest rates; extending private property rights; and deregulating economic activity.

International Monetary Fund (IMF) and World Bank structural adjustment policies are integral to the neoliberal agenda and the Washington Consensus, and these policies have had damaging effects on economies by increasing inequality within and between nations, leading numerous scholars to argue that neoliberal globalization has replaced older forms of imperialism to make developing countries economically dependent on Western powers (James 1997; McMichael 2008). IMF-mandated public spending cuts often reduce police officer salaries, spawning further corruption, and both retired and active-duty police officers might supplement their incomes by working as private security guards or participating in death squads (Hinton and Newburn 2009).

In addition, by promoting economic insecurity and therefore fear of crime, neoliberal globalization compels states to increase coercive powers while taking a more limited role in the economy, creating a market for policing and other security reforms. In fact, police reform is increasingly conducted by private non-state actors

(Kempa et al. 1999), leading to the weakening of state policing institutions across the globe. While precise numbers are difficult to come by, it is now known that private security personnel are more numerous than state police (Aas 2007). Policing is less rooted in the nation state because of privatization, transnationalization, and the development of a transnational police subculture. These developments, including intelligence sharing between countries, moving some forms of police authority in Europe from nation-states to the European Union, and other formal and informal agreements, threaten the democratic nature, accountability, and legitimacy of policing (Loader 2002).

Since the end of the cold war, international financial institutions such as the IMF and powerful states – particularly the USA – have been increasingly showing interest in security sector reforms based on Western models, and coercing governments in developing democracies to adopt Western models and practices. For example, owing to the fact that police can hinder development by ignoring the rule of law and acting independently from their governments, the IMF and World Bank have been making the strengthening of governance one of their mandates for securing loans (Ball 2001). Police aid has become a major global export (Bayley 2006; Brogden 2005; Ellison 2007), involving private corporations who train community policing officers, private consultants who market themselves as gurus or experts, international governmental and nongovernmental organizations, and Western overseas development agencies. Numerous agencies such as the United Nations Development Program (UNDP), the US Agency for International Development (USAID), the Department for International Development (DFID), the International Criminal Investigative Training Assistance Program (ICITAP), and various nongovernmental organizations (NGOs) work to diffuse police concepts and practices between nation-states and supra-nationally above individual states in the case of Interpol (Goldsmith and Sheptycki 2007).

The USA has played a major role in these developments. As an important component of its foreign policy, the country has sent more police abroad, trained more police, and has signed more bilateral law enforcement treaties than any other country (Andreas and Nadelmann 2006). The USA usually dictates the processes and practices of these agreements and the contents of laws, and while elites in developing countries participating in these agreements claim they are now meeting international standards, they can utilize the agreements to meet various domestic agendas (Andreas and Nadelmann 2006).

One cannot assume, therefore, that the aid provided by Western experts is solely for the recipient's benefit. During the cold war, for example, institutions and public order policies implemented abroad were intended to protect pro-US military regimes instead of the citizenry, leading to militarized police forces that engaged in abuses and repression (Goldsmith 2003; Neild 1999). When paramilitary units are created instead of democratic forms of policing in order to defeat political enemies, drug traffickers, or terrorists, it contributes to goal deflection, the overdevelopment of special units, undemocratic and dependency-driven reform and training, and insufficient participation by local actors and civil society groups. Local actors and stakeholders such as NGOs, local governmental leaders, and the citizenry are usually not consulted or part of the process of police reform in any meaningful way (Hills 2000). Donors attempt to instill Western values and principles where they may not apply on the ground, and during times of reform, there is a strong emphasis on nonpartisan, high-tech, information-driven policing as a way to shield police from democratic forms of accountability (Neild 2001).

Police reform within nation-states rarely involves meaningful forms of civilian oversight, resulting in police control over reforms. Since many populations seeking changes are marginalized and lack access to participating the process, the chances for genuine reform can appear highly unlikely. While activism and resistance by citizens, politicians, external donors, and civil society groups can influence reform efforts, poverty and inequality reduce the efficacy of these efforts since many citizens are busy with subsistence and survival concerns, and politicians can easily purchase votes with jobs, cash, and various services (Hinton and Newburn 2009). NGOs have difficulty collaborating in reforms because of their tenuous relationships with governments, and if many reform changes take place within police management, it is unlikely NGOs will have much access to or input on those decisions (Neild 1999). NGOs tend to be inexperienced in a number of countries and need more assistance if they are to participate. Police reform tends to be slow, erratic, and partial, and as NGOs attempt to assist the police while monitoring them, there is more opportunity for weakened relations between the police and NGOs. In addition, the obstacles to female citizen participation in co-productive activities with the police are rooted in the varied patriarchal structures of societies (Pino and Wiatrowski 2006). There are low levels of female participation in decision making in general and women suffer distinct harms disproportionately from crime (domestic violence, rape, etc.) and social and ethnic conflict. Low female participation fails to tap into the local knowledge of crimes and other problems that confront them in their communities. Because women are vulnerable to victimization in many developing democracies, their children are also at risk, limiting social activities that can be used to improve livelihoods.

A common problem is that donors have been largely uninformed about a recipient country's history, political and ethnic problems, and cultural scripts for negotiating and implementing agreements, leading donor agencies to act with incompetence regardless of the quality of the plans put in place. In many endeavors around the globe, donor governments and their various participating agencies have not followed-through on their commitments and have failed to cooperate well (Muehlmann 2008). Many people hired to work on the ground with local actors lack expertise or knowledge of the country and culture, and when multiple agencies are involved in providing reforms, as in Iraq and Afghanistan, there are blunders in coordination and failures to meet funding guarantees. Planning and implementation are often rushed, and as Bayley (2006) has mentioned, many donors conduct most of their work via simple fly-bys and drop in and out visits. At a conference in the Caribbean, the author heard more than one local scholar call those that engage in such practices "tourist criminologists."

Another donor-related problem is that Western democracies do not always agree on appropriate police standards. There appears to be no global consensus emerging on appropriate policing practice, use of force, or effectiveness (Hills 2000). When supposed universal policing standards are implemented, they often do not convincingly match the rhetoric due to local political, historical, and cultural dynamics that are not properly accounted for in preparation for or during reform efforts. If state-builders retreat from democratization efforts for short-term gain, lower their expectations, fail to accept responsibility, or over-emphasize local capacity and ownership from the beginning before capacity is developed, it will likely lead to ineffective service and higher levels of corruption and abuse (Kahler 2009).

The most popular policing export from the West to developing democracies is community-oriented policing (COP), but the effects of COP in developing democracies have been mixed at best. On the positive side, there have been numerous improvements in the treatment of female victims, police visibility, and citizen perceptions regarding fear of crime, but there are serious problems that have hindered successful implementation (Brogden 2005; Pino and Wiatrowski 2006). It is questionable whether COP is suitable for export to foreign countries in the first place owing to the problems associated with implementation in the USA. While the concepts and theories behind COP are popular among academics and practitioners alike, it is conceptualized and practiced differently in every community it is implemented, the status quo is often maintained in terms of police behavior, co-productive activities between the police and citizens are usually limited to nonexistent, and those that need police reform and improved police-community relations the most often receive the least of them (Pino and Wiatrowski 2006). Many of the problems associated with community policing in the West also occur in transitional societies, such as a lack of participation and social capital to sustain reforms, lack of assessment and other forms of evaluation, continued police abuses, lack of organizational change in police departments, and others. Furthermore, when donors state they will implement community policing but engage in goal-deflecting activities such as creating paramilitarized special units to fight terrorism, battle well-armed drug traffickers and the like, or do not take the time to learn the local context and other realities, any hope for developing democratic forms of policing can be squandered (Pino and Wiatrowski 2006).

Another factor to consider is the willingness and capacity of the country receiving aid to actually carry out the reforms. Many recipients are willing if they see an economic or political benefit from doing so, but the capacity of countries to implement reforms, even if they are willing, varies. For example, owing to the legacies of colonialism and conflict, there are few local experts in many countries, such as policy makers, academics, NGOs, and other civilians with expertise that can assist with reforms, leading to reform efforts being dictated solely by donors, and in spite of the numerous pressures donors place on countries to engage in reform, there are various ways that recipients resist this pressure (Andreas and Nadelmann 2006; Neild 1999). Recipient countries can decide to implement only the aspects of reform they like after donors leave, resources can be diverted, and reform can be used as catalysts for political gain.

Reform requires political stability and favorable relations between federal, regional, and local governments, but these conditions do not exist in many countries. In numerous developing democracies, police are under the influence of individual politicians instead of being institutionally accountable (Hinton and Newburn 2009). Police commissioners are replaced regularly and tenure is often most dependent on loyalty to a political party or regime. In these cases, corruption is rampant, accountability mechanisms are weak, and police officers engage in violence and other forms of crime. Bureaucratic organizations are penetrated by neopatrimonial politics that make little-to-no distinction between the public and private sector, creating the personalization of political power into one individual who engages in clientalist politics that encourage corruption and the use of state resources and coercion to enhance authority and legitimacy (Hills 2000).

We do have to keep in mind, however, that sometimes the process of creating a new force works fairly well. Police reform in post-conflict societies making democratic transitions is more successful when the international community supports the peace process and provides assistance to public security reforms (Neild 1999). In countries such as El Salvador and Guatemala, local communities appeared at the time to welcome these new police forces and see them as an improvement over previous forces (Neild 2001). Still, very quickly, these forces can be corrupted, unless they are placed in a framework that emphasizes their role in terms of a democratic society.

Suggestions in the Literature for Overcoming Barriers to Sustainable Police Reform

The literature has offered some advice on how practitioners can improve policing in developing democracies in terms of practical and technical issues, but also in formulating ways to contend with the daunting challenges presented by neoliberal globalization on police reform efforts. Some argue that scholars and practitioners should pragmatically embrace and work within the realities of neoliberal globalization and American foreign policy objectives by utilizing top-down strategies and Western expertise, but to do so with better preplanning and coordination within and between countries, agencies, and private contractors involved in providing the aid, improving oversight and regulation of private actors, increasing transparency, and engaging in some collaboration with local stakeholders (see Bayley 2006).

Other scholars such as Loader and Walker (2007) question the underlying policy objectives of the USA and other Western countries and emphasize strengthening state capacity to manage economies and provide security democratically without relying on private contractors. In their view, states should be allowed to develop their economies independently and provide what Loader and Walker (2007) call thick security, which they define as security that is seen as a communal public good and provides a civilizing effect on political discourse, therefore strengthening democratic institutions. Others such as Goldsmith and Sheptycki (2007) want to help develop a transnational ethic for the provision of security that promotes peace and security in a democratic fashion rather than based on dictates from powerful states in a neocolonial fashion. This type of effort compliments the desire for thick security and independent development.

Yet another set of scholars argue for strengthening the capacity of civil society groups and the development of bottom-up strategies for reform, drawing from indigenous knowledge and the strength of NGOs to emphasize local needs (see for example Hills 2000; Pino and Wiatrowski 2006). While bottom-up strategies ensure that local actors are able to participate in the process, one must keep in mind that local leadership and capacity may be too weak to spearhead all initiatives depending on the context, and reliance on the private sector or civil society groups for security would not be an effective replacement for state-managed police forces due to the internal conflicts, inequalities, availability of small arms, and other difficulties that normally require states with the adequate capacity to manage (Goldsmith 2003). Civilian review boards are a popular method for involving citizens in the police reform process, influencing internal police discipline, and providing analysis, but success requires strong leadership, analytical capabilities, political support, resources, and minimal police resistance (Neild 1999). Female participation is needed as well to ensure that the needs and concerns of women and their children are addressed and because women have valuable local knowledge about local problems that can inform police reform and crime reduction efforts.

Scholars are still struggling with these issues involving police reform and neoliberal globalization, the proper roles and obligations of donors and recipients, and the extent to which civil society groups and private actors participate. Nevertheless, it appears from the above review that at the very least reform efforts require effective planning and coordination among all actors; significant local participation, including that of civil society groups; adherence to funding and other commitments made in agreements; an equal emphasis on both human rights and security; accountability and transparency among all actors involved; and an emphasis on long-term goals and planning. However, before any planning or coordination begins, certain conditions should be in place if sustainable reform is to be achieved. Reforms should only be attempted in relatively stable, non-repressive states ripe for democracy with the resource and capacity potential to maintain reforms sustainably and provide essential services. Reforms should be mutually sought rather than coerced or paternalistically dictated, and donor states or organizations should only utilize competent actors with knowledge of and expertise in the country receiving assistance. Reform also requires strong and legitimate police leadership eager and willing to carry out reform efforts and to participate in the process. Finally, police reforms should be part of a comprehensive approach that seeks to strengthen multiple democratic institutions simultaneously, such as other criminal justice systems, the educational system, militaries, and political and economic systems.

Related Entries

▶ British Police
▶ Causes of Police Legitimacy
▶ Communities and the Police
▶ Control of Police Misconduct
▶ Cross-National Performance in Policing
▶ Democratic Policing
▶ French Colonial Police
▶ Organizational Change and Police Legitimacy
▶ Police Corruption
▶ Police Culture

- ▶ Police in the Police State
- ▶ Policing of Peacekeeping
- ▶ Policing the Police
- ▶ Privatization of Policing in an International Context
- ▶ Role and Function of the Police
- ▶ Women in Policing

Recommended Reading and References

Aas KF (2007) Globalization and crime. Sage, London

Andreas P, Nadelmann E (2006) Policing the globe: criminalizationand crime control in international relations. Oxford University Press, New York

Ball N (2001) Transforming security sectors: the IMF and World Bank approaches. Confl Secur Dev 1(1):45–66

Bayley D (2006) Changing the guard: developing democratic police abroad. Oxford University Press, New York

Brogden M (2005) Horses for courses and thin blue lines: community policing in transitional society. Police Q 8(1):64–98

Cole BA (1999) Post-colonial systems. In: Mawby R (ed) Policing across the world: issues for the 21st century. UCL Press, New York, pp 88–108

Ellison G (2007) Fostering a dependency culture: the commodification of community policing in a global marketplace. In: Goldsmith A, Sheptycki J (eds) Crafting transnational policing: police capacity-building and global policing reform. Hart Publishing, Oxford, UK

Goldsmith A (2003) Policing weak states: citizen safety and state responsibility. Polic Soc 13(1):3–21

Goldsmith A, Sheptycki J (2007) Introduction. In: Goldsmith A, Sheptycki J (eds) Crafting transnational policing: police capacity-building and global policing reform. Hart Publishing, Oxford, pp 1–27

Hills A (2000) Policing Africa: internal security and the limits of liberalization. Lynne Rienner, Boulder

Hinton MS, Newburn T (2009) Introduction. In: Hinton M, Newburn T (eds) Policing developing democracies. Routledge, New York, pp 1–27

James P (1997) Postdependency? In: Darby P (ed) At the edge of international relations. Pinter, New York, pp 61–83

Kahler M (2009) Statebuilding after Afghanistan and Iraq. In: Paris R, Sisk TD (eds) The dilemas of statebuilding: confronting the contradictions in postwar peace operations. Routledge, New York, pp 287–303

Kapstein EB, Converse N (2008) The fate of young democracies. Cambridge University Press, New York

Kempa M, Carrier R, Wood J, Shearing C (1999) Reflections of the evolving concept of 'private policing'. Eur J Crim Policy Res 7(2):197–223

Loader I (2002) Governing European policing: some problems and prospects. Polic Soc 12(4):291–305

Loader I, Walker N (2007) Locating the public interest in transnational policing. In: Goldsmith A, Sheptycki J (eds) Crafting transnational policing: police capacity-building and global policing reform. Hart Publishing, Oxford, pp 111–146

McMichael P (2008) Development and social change: a global perspective, 4th edn. Pine Forge Press, Los Angeles

Muehlmann T (2008) Police restructuring in Bosnia-Herzegovina: problems of internationally-led security sector reform. J Interv Statebuilding 2(1):1–22

Neild R (1999) From national security to citizen security: civil society and the evolution of public order debates. International Center for Human Rights and Democratic Development, Montreal

Neild R (2001) Democratic police reforms in war-torn societies. Confl Secur Dev 1(1):21–43

Pinkney R (2003) Democracy in the third world. Lynne Rienner, Boulder

Pino N, Wiatrowski MD (2006) Democratic policing in transitional and developing countries. Ashgate, Aldershot

Plattner MF (2008) Democracy without borders? Global challenges to liberal democracy. Rowman & Littlefield, Lanham

Stavenhagen R (1996) Ethnic conflicts and the nation-state. St. Martin's Press, New York

Stone CE, Ward HH (2000) Democratic policing: a framework for action. Polic Soc 10:11–45

Uildriks N, Van Reenen P (2003) Policing post-communist societies: police-public violence, democratic policing, and human rights. Intersentia, Antwerp

Williamson J (2000) What should the bank think about the Washington Consensus? World Bank Res Obser 15(2):251–264

Zvekic U, Alvazzi del Frate A (1995) Criminal victimization in the developing world. UNICRI publication number 55

Policing of Peacekeeping

B. K. Greener
Politics Programme, School of People, Environment, and Planning, Massey University, Palmerston North, New Zealand

Overview

Policing has become an important part of peace and stability operations. This entry briefly outlines the rise of policing in peacekeeping before highlighting some of the difficulties met in efforts to deploy civilian police personnel to peace and stability operations. Alarmingly, such difficulties have helped motivate increased military and Formed Police Unit (FPU) involvement in policing tasks. The implications of this trend are wide-ranging, suggesting that limits for military, FPU, and civilian police involvement in operations need to be more clearly defined.

Key Issues

Policing has become an integral part of contemporary peace and stability operations. In 2011, for example, there were: approximately 14,500 UN police (UNPOL) deployed in places as diverse as Timor Leste and Chad (UNDPKO 2011); EU police (EUPOL) could be found operating in the Palestinian territories, Afghanistan, and elsewhere; a Participating Police Force (PPF) drawn from the Pacific Islands Forum countries were still policing in the Solomon Islands; and national contingents engaged in other bilateral or multilateral arrangements all around the world. Ideally such policing would be carried out by qualified civilian police personnel, but at times, military personnel have been tasked with components of the police function in some of these operations, and more militarized police units are increasingly being utilized in such operations too. There are a number of practical reasons behind this trend, but there are also weighty normative reasons to be cautious about such developments.

This entry sketches the rise of the importance of policing in peace operations and considers some of the problems that have developed with the rise of international policing. It then outlines two reasons as to why military personnel and Formed Police Units have increasingly been involved in such tasks: quantity issues and the mission environment. The entry then assesses some of the implications of this particular development, concluding that, despite growing proponents of military involvement in policing, care needs to be taken to avoid any moves toward the militarization of policing in peace operations and more needs to be done to address this situation.

International Policing

The current international policing agenda has a fairly recent history if we exclude the role of colonial policing. Initial post War policing efforts in occupied Japan and Germany focused on rebuilding and tweaking structures that had existed prior, while Cold War peacekeeping policing tasks focused on fairly uncontroversial Support, Monitoring, Advising, Reporting, and Training (SMART) type activities. Policing in peace operations during this phase tended to revolve around monitoring of security situations and ceasefires, as exemplified by the UN deployment to Cyprus. Post-Cold War scenarios, however, saw the UN and other regional or bilateral policing agencies becoming involved in ever more complex arrangements as the need for the restoration of law and order has come to be seen as vitally important for post-conflict peacekeeping, peace enforcement, and peacebuilding (Greener 2009).

The move toward peacebuilding and statebuilding, the convergence of security and development agendas, and increasing challenges to the nonintervention principle have all underwritten developments in the burgeoning field of international policing. This new policing agenda has therefore been driven by the complex demands of peace operations as they have evolved past initial peacekeeping concepts throughout the post-Cold War era. Increasingly comprehensive and extensive attempts to prevent a resurgence of violence in post-conflict sites, and the drive to rebuild states in particular, have helped amplify demands for the implementation of the rule of law and the restoration of law and order – demanding involvement that goes beyond the mere provision of a security pause. In addition to this, post-9/11 demands to prevent or respond to the failure of states, to counter terrorism (and potentially related transnational crime), and to shore up controls over illicit and insurgent activity have all contributed to an increasing demand for police as well as military involvement in international security affairs.

It is due to such developments that policing has become an increasingly important part of peace operations. Instances of such policing may involve either more "active" roles particularly in the stability phase of operations (roles such as those which occur where there is an executive policing mandate whereby external police have the power of arrest in situations where local law and order has broken down), or it can involve a range of other less obviously interventionist roles in later or more permissive stages (roles such as monitoring, training, mentoring, capacity building, or programs to "reform, rebuild, and restructure" existing police capabilities). Executive missions have become somewhat less prevalent but capacity-building programs currently abound – both through UN or regional auspices as well as through bilateral arrangements.

There have been a number of relevant concerns raised about the rise of this policing agenda. There are salient concerns about how police personnel from outside a jurisdiction could possibly hope to successfully police a population or reform a local police service in an area that they are not connected to nor necessarily very familiar with. Further complicating this picture is the fact that one of the major problems with international policing efforts so far has therefore been the quality of that policing provided. Variations in civilian police policies, practices, and abilities have also caused difficulties, as individual police services around the world vary greatly in their expertise, ethos, and capabilities. Problems with a lack of cultural sensitivity and language barriers, as demonstrated in Timor Leste (Peake 2009) or Solomon Islands (Allen 2006), are significant enough, even without the concern that police operating overseas may both be getting paid very well while feeling less ethical constraints and being less accountable to oversight mechanisms. Experiences in Bosnia, Sudan, and elsewhere have demonstrated that the ideals of human rights centered democratic policing where police "protect and serve" the local population have not always been adhered to, with a number of charges of abuse of office, unwillingness to carry out policing tasks, or corrupt, immoral, or illegal behavior being reported.

In spite of these particular concerns, however, international policing efforts seem set to continue well into the future (Bayley 2001). This notion is borne out by contemporary conversations as to possible mission deployments to Libya. Concerns over the international policing project itself have often been elided over as internal and external security agendas increasingly merge (Rubenstein 2010), but more specific criticisms about the way in which such policing efforts are undertaken are being taken into account. Efforts are therefore underway to try to standardize training, improve policing standards to help offset current critique (Rotmann 2009; Goldsmith and Scheptycki 2007). The UN Police Division, in particular, has set about trying to improve the current policing practice of UNPOL personnel through standardizing some pre-deployment training and drawing up new professional standards documents. Questions of local versus international ownership are also being discussed in more detail (Hansen 2008).

These efforts may help mitigate some of the concerns about particular aspects of international policing, such as the quality of the policing that is delivered. However, as noted below, in addition to this quality control problem, there also remain concerns about attaining enough *quantity* of personnel to actually undertake those policing missions abroad. (This concern over quantity is in fact slightly ironic given that concerns over quality of policing available in international operations have often led police managers to suggest a better job could be done with less but better quality international police personnel (i.e., police personnel who are ethical, professional, trained in the task at hand and fully accountable for their actions)). And, as we will see below, this particular issue has had a significant knock-on effect.

The Quantity Issue

Although numbers of UNPOL personnel have increased exponentially, and a number of national governments such as Australia and regional institutions such as the EU have created specialized policing deployment pools or groups in response to the demand for international policing capabilities, there are still difficulties in deploying enough civilian police to post-conflict settings. Put simply, there are never enough civilian police to deploy on international operations in a timely fashion. In Africa, in 2009, for example, the UN Security Council (UNSC) had authorized the use of approximately: 6,400 police for the UN–AU Hybrid Operation in Darfur (UNAMID); 1,400 for UN Organisation Mission in the Democratic Republic of Congo (MONUC); 1,200 for the UN Mission in Liberia (UNMIL); 1,200 for the UN Operation in Cote d'Ivoire (UNOCI); and 700 for the UN Mission in the Sudan (UNMIS), but most of these were operating with major shortfalls in numbers (i.e., around 3,000 of the 6,400 were operating in Darfur, while only approximately 600 were operating in the Democratic Republic of Congo) (Williams 2009). And although a number of efforts have been made in recent years to boost international policing capacity in Africa itself, it has been said that:

> Police in general are difficult to recruit to international missions, because they are needed at home. African police are no exception. Many African countries suffer from high crime rates, making it difficult for them to "spare" police for external missions. (Levine 2008, p. 35)

The issue of getting enough police on the ground, particularly within short time frames, therefore continues to bedevil the planners of peace missions and is a product of the place and primary purpose of civilian police. The fact that police are typically being used in a domestic setting, and therefore much planning and resourcing must be in place to release such police, means that those attempting to organize deployments continue to struggle to get enough police into the field.

This issue impacts upon the ability to police in peace operations with civilian police rather than alternative agencies. It is partly this issue of releasing enough numbers of individual civilian police officers for timely international deployment, combined with other mission factors that will be described below, that has encouraged the use of more militarized Formed Police Units (FPUs) or even military personnel for the taking up of policing-type roles as these are institutions that have the ability to deploy mobile units of personnel quickly.

FPUs are often drawn from constabulary forces such as the French *gendarmerie* or the Italian *carabinieri*, forces which are paramilitary in terms of their organization and weaponry. The "Multinational Specialist Units" (MSUs) that were developed by NATO for use in Bosnia in order to help bridge military and civilian police capabilities demonstrate the key roles of FPUs as these particular MSUs consisted of:

> Police Forces with military status who perform duties including civil disturbance operations. This type of force focuses on the civilian population, employs minimum force and often employs small units to accomplish the mission. However, the MSU is not a replacement for the [Bosnia and Herzegovna] Police Forces and does not conduct criminal investigation beyond the scope of the SFOR mandate. (NATO 2004)

Indeed this quote highlights the fact that some FPUs, such as those drawn from the European continent, have additional policing capabilities such as investigative capabilities. Some other FPUs, however, are predominantly riot control or public order management units only. All generally have more "hardware" immediately at hand than civilian police in that such units have access to a range of riot equipment such as shields, batons, capsicum spray, "flashbangs," tear gas, water cannon, rubber bullets, or shotgun bean bag rounds. FPUs are trained in a range of public order management techniques and can operate as a complete, mobile, expeditionary unit.

The use of FPUs in international peace operations since their introduction in Kosovo and Bosnia has increased markedly. In the UN case, for example, FPUs were only brought into UN missions just prior to 2000, by April 2002, about 1,000 of 7,000 UNPOL deployed were members of FPUs, while in April 2009, that proportion had risen to approximately 7,000 of the 17,000 UNPOL deployed in peace operations (Carpenter 2010). This is in part due to the easily deployable and flexible nature of such units. For example, the EU has developed a specialist group out of four continental forces from Spain, France, Italy, and the Netherlands called the European Gendarmerie Force (EUGENDFOR). The EUROGENDFOR can deploy 800 officers in 30 days, and given its nature "can carry out all types of missions in crisis management operations, taking part in the military phase of a crisis, acting during the transition period from a military to a civil operation, and participating in prevention missions" (Ministeria de Defensa de Espana n.d.) In particular, EUROGENDFOR's police function can "be developed under civilian or military chain of command," whereby: during the initial phase of the operation the EGF could "enter the theatre along with the military force in order to perform its police tasks"; during the transitional phase could "continue its mission alone or together with a military force, facilitating co-ordination and co-operation with local or international police units"; and during the military disengagement phase could "facilitate the handing over of responsibilities to civilian authorities and agencies taking part in the co-operation efforts, if necessary" (EUROGENDFOR n.d.). This means that the EU can deploy 800 personnel quickly, with the ability to utilize them in different roles.

In addition to the rise in use of FPUs, the use of military personnel for policing-type activities in international operations has some historical precedent but again, due in part to their deployability and the demands for policing capabilities, has also increased dramatically in recent years. In Panama, for example, "the US military, as the only entity in the country with the capacity to provide the level of stability required, had to deal with the lack of a viable and functioning [host nation] security and [law enforcement] apparatus," (Jayamaha et al. 2010, p. 21). In Kosovo, the USA used their military police as a constabulary type force. The involvement of military personnel, particularly MPs, in traditional policing roles has continued in recent years, with the US military playing "a key role in law enforcement and related issues, even if not specifically tasked with a law enforcement mandate" (Jayamaha et al. 2010, p. xiii) and with undertaking the tasks of police training, mentoring, and institutional reform – mainly in Iraq and Afghanistan (Keller 2010).

Difficulties in getting civilian police to international deployments have therefore helped encourage the use of FPUs or military personnel in policing peace operations, as both of these types of organizations deploy personnel in larger units that have expeditionary capabilities which are more readily released for service than their civilian police counterparts. However, the complex nature of post-conflict environments has also been a motivating factor in the increased involvement of military and FPU personnel in policing or policing-type activities.

The Permissibility Factor

The complex nature of peace operations has also created concerns about whether or not civilian police can in fact operate in more difficult or "hot" security situations (Bayley and Perito 2010). This concern over the safety of civilian police has at times seemed to require military involvement in policing, particularly if state or private police are unable or unwilling to operate in "nonpermissive" environments (Keller 2010). Perhaps more markedly, the concerns over the ability of civilian police to operate in difficult security environments have also, in particular, again also promoted the use of FPUs in policing peace operations.

Contemporary peace operations can involve significant levels of civil disorder, often involving armed protagonists. The problem with such scenarios is that a "capability gap" is perceived to exist between civilian police and the military when it comes to dealing with this kind of situation. Dziedzic (2003, p. 2) argues that:

> Military forces are ill suited to engage in confrontations with civilians because, with the exception of constabulary or military police units, they are generally not trained in the measured use of force, control of riots, negotiating techniques, or de-escalation of conflict. Individual [civilian police] are not capable of handling such large-scale, strategic challenges, either.

Instead, FPUs are thought to be a suitable halfway house in helping to police peace operations in that:

> International peacekeeping operations have in the past been labelled *constabulary* missions, calling for gendarme rather than military capabilities, although throughout the Cold War they were conducted primarily by military forces. Since 1996, beginning with the Multinational Specialised

> Unit in Bosnia, forces like the Italian Carabinieri and the Dutch Marechaussee have been in demand to deal with the combination of organised crime and ethnic extremism that complicates protracted social conflicts. (Mendee and Last 2008, p. 18)

Constabulary forces, then, can seem appealing in certain situations as they are seen as capable of performing both "cooler" military and "hotter" police functions and can be assigned "in either a military or civilian capacity" (Perito 2004, p. 47). EU member states therefore promoted the creation of the EUROGENDFOR as filling a perceived capability gap as well as helping the EU redefine and extend its contributions to peace and stability operations (de Weger 2009). And reports and observation from places such as Timor Leste have suggested that such units, in this case the Guarda Nacional Republicana (GNR), have at times been key in the restoration of security and order due to their ability to respond quickly and decisively to small crises to avoid escalation of localized conflict (see Lemay-Hebert 2009). Formed Police Units and military personnel are therefore being used with increased frequency to undertake policing in peace operations for a number of reasons. However, although the status of Formed Police Units is that of "police," there remain a number of salient concerns both about the increased reliance on FPUs and about the use of military police or other military personnel in some policing-type roles.

Concerns Over the Militarization of Policing

As Alice Hills has noted, although FPUS "make sense: they work" in terms of: their abilities to control crowds; the fact that they allow the UN in this case to expand the number of contributing countries and to get closer to gender targets in that there are all-female units that can deploy to help fill UN quotas; and because they are cheaper per person, the use of FPUS can also be problematic as "they are paramilitary and their composition and operations blur police/military boundaries" (Hills 2009, p. 309). This is particularly problematic in post-conflict situations where the ideal is to clearly demarcate policing and military roles to help demilitarize society and to exclude military from internal security roles. This creates a difficult situation. The security environment may require a "harder" approach to policing society so as to maintain order, but the transition to a less militarized environment becomes more difficult to achieve with such approaches in place.

Moreover, in addition to the provision of basic security needs, the institutionalization of a civilian police force that supports, protects, and serves the local population well is also key to broader democratic or human rights–centered approaches to the political and social situation at hand. After all, as Mercedes Hinton and Tim Newburn note in their work on policing developing democracies "if people can have trust and confidence in rules, institutions, and authorities, they are likely to believe that their long-term personal interests will be well served by voluntary compliance with the laws of the state" (2009, p. 4). If more aggressive paramilitary styles of policing inhibit this trust and confidence, the whole peacebuilding or statebuilding project becomes more tenuous.

However, one possible way to work around this problem could be to utilize FPUs for specialist roles in post-conflict sites that are *known temporary* measures for filling a capacity gap that utilize international units rather than locally sourced personnel. This has been suggested as a way to avoid besmirching the role of local civilian police who are ideally attempting to build legitimacy and consent-based policing models, therefore:

> It may be necessary to create a unique police force that can both provide security and also promote a transition to a more stable and accountable environment, thus allowing conventional police forces to focus on developing according to democratic values. . .. They will not make local police responsible for security functions that would otherwise result in their para-militarisation and the use of extra-legal means for controlling crime and restoring order. (Wiatrowski et al. 2008, pp. 1 and 11)

However, this may also be problematic due to the potential for alienation of the population should the timing or operationalization of this phase be misapplied. The fact that FPUs are typically under military command can have ramifications for command and control, and for how they are therefore related to the issue of governance of the mission though, as noted above, they can also be deployed under civilian control.

In addition to these particular concerns about the heavy use of FPUs in peace operations, additional significant disquiet regarding the use of military in law enforcement-type tasks also remains as while they may be in situ with useful capabilities, policing activities are *not* the core business of the military. Policing skills and mindsets are developed over time and then consolidated on a daily basis through the activities, interactions, and ongoing training of police personnel – such skills or mindsets cannot be internalized quickly, rather this process takes some time to embed to result in a maturity of approach toward complex issues. There therefore remain a number of important distinctions between police and military personnel. In a recent American study, which outlined a number of these important differences between street cops and combat troops, for example, it was found that:

> Although both jobs are in protective service occupations, they differ significantly in terms of specific tasks and expectations. While police officers patrol and investigate wrongdoing, combat rifle crewmembers destroy or capture enemies; while police officers inspect, warn and arrest; rifle crewmembers fortify, camouflage, and repair weapons; while police officers report hazards, disperse unruly crowds, and write daily activity reports, rifle crew members place antitank mines and fire machine guns, and so on. (Campbell and Campbell 2010, p. 339)

It is due to these differences in training and ethos that the authors also argue that when military units are required to undertake policing roles, they focus primarily on developing relevant skills, while police taking on more military-oriented tasks tends to focus less on skill development and more on attitude adjustment (2010, p. 343).

These constitute reasons to be wary of military personnel playing a wider role in training and mentoring police forces too (Mobekk 2005, pp. 5–6). Overly militarized approaches to training local police forces (including the mentoring of police officers by members of the military) can lead to the police force itself taking on militaristic features and militaristic priorities (Sedra 2006, p. 95; Bayley and Perito 2010, p. 4). Much of the literature therefore demonstrates some level of concern about the involvement of military forces in police training and reform, yet this involvement is still occurring and being encouraged in some situations – particularly in Iraq and Afghanistan. This is in large part capability-driven, in that the military tends to have the capabilities or the funds to undertake such a mammoth task, but unless there were significant changes to military forces – changes which would inevitably have a negative impact upon their ability to play their core roles – there will remain a reason to maintain the police–military divide both at home and in operational contexts. This principle is blurred somewhat by the security gap issue that commonly arises in post-conflict situations, and it therefore seems plausible that some military roles might usefully be played *in support of* law enforcement. However, such roles need to be clearly defined and understood prior to deployment, and clear plans made to withdraw military or more militarized policing units from civilian policing tasks at the first possible moment.

Conclusions

In summation, policing has become a key part of international, regional, and bilateral efforts to stabilize and improve post-conflict situations. Policing in peace and stability operations has also become much broader and deeper in terms of the types of tasks and roles police play. There are a number of critiques about the rise of the international policing agenda, but concerted efforts are underway to try to improve the operational aspects of such policing. The ability to generate enough good civilian police for international deployments has constituted one part of such recent efforts, as the lack of available civilian police has at times helped prompt the use of FPUs and military personnel in undertaking some policing functions which has brought its own concerns.

FPUs are, by most definitions, at least a form of "police" and operational experience suggests that they can provide a useful capability, provided they are used wisely. However, there remain some reservations about their use in post-conflict peace operations in particular, and the use of military personnel to perform policing functions too brings a number of additional problems. In addition to simple capability differences (i.e., the training, equipment, situational awareness, etc., will differ for police versus military personnel), there are a number of broader ramifications. In particular, a major goal of most peace operations is the demilitarization and de-escalation of the local environment, as well as the embedding of particular political contexts; therefore, the use of military (or highly militarized FPUs for that matter) for policing functions can blur the democratic ideal of military as external actor and police as internal actor.

The debate over the practical and ethical issues of military or militarized police units undertaking civilian policing in peace and stability operations therefore looks set to continue until a more ideal division of labor can be clearly conceptualized of, internalized by the agencies involved, adequately resourced but, most importantly of all, sited within a more general conversation about the broader ethical and normative shifts that this trend embodies. It is likely that planners of peace and stability operations will look to utilize three main security actors – military personnel (including military police), FPUs, and civilian police – in a more systematic way in future. The limitations, both practical and ethical, of using such agents in particular roles need careful forethought. FPUs deployed for public order maintenance roles, for example, may have a place in plugging the capability gap in less permissive environments, but should be removed as soon as possible to encourage demilitarization of society. Military police may be preferable to infantry for training police in nonpermissive environments, but if military police rather than civilian police are the ones providing training for this reason then the question must then also be asked as to whether or not the situation can actively support a civilian police force at that point in time. Creating a highly militarized police force in a post-conflict or potentially still conflict-ridden environment may simply be creating another destabilizing force in country.

There are practical concerns about securing civilian safety in peace and stability operations that need to be met. However, there are also larger concerns about keeping the military out of internal security roles in burgeoning democracies, or about demilitarizing a traumatized post-conflict environment, that mean that decisions to use agents other than civilian police for internal security must be weighed and measured carefully. Future research needs to look at what patterns of policing have worked in what types of environments, and needs to probe beyond practical concerns to the larger political and normative background against which such technical and operational tenets sit.

Acknowledgments Some of the research for this entry was undertaken as part of the “Civil-Military Interactions in Peace Operations” project that was funded by the Asia Pacific Civil Military Centre of Excellence, Queanbeyan, Australia, and part of this entry draws on an initial working paper: B.K. Greener and W.J. Fish, *The Development of Civilian Policing: Lessons for Contemporary Post Conflict Operations*, Occasional Paper 4/2011, Canberra: APCMCoE. However, the views expressed in this piece are those of the author alone. The author would also like to thank all of those police and military personnel who have given so willingly of their time to discuss their experiences as part of this project.

Related Entries

- ▶ Democratic Policing
- ▶ Gendarmerie Policing
- ▶ Police and the Military Nexus
- ▶ Order Maintenance Policing
- ▶ Policing in Developing Democracies

Recommended Reading and References

Allen M (2006) Dissenting voices: local perspectives on the regional assistance mission to the Solomon Islands. Pac Econ Bull 21(2):194–201

Bayley D (2001) Democratizing the police abroad: what to do and how to do it. US Department of Justice, Washington, DC

Bayley D, Perito R (2010) The police in war: fighting insurgency, terrorism and violent crime. Lynne Rienner, Boulder

Campbell DJ, Campbell KM (2010) Soldiers as police officers/police officers as soldiers: role evolution and revolution in the United States. Armed Forces Soc 36(2):327–350

Carpenter A (2010) UN police peacekeeping: its different from the day job. Powerpoint presentation. Retrieved 8 Apr 2011 from www.docstoc.com/docs/49053445/UN-Police-Presentation

De Weger M (2009) The potential of the European gendarmerie force. Clingendael, Netherlands Institute of International Relations, The Hague

Dziedzic M (2003) The public security challenge and international stability police units. Perceptions, 8 Dec 2003. Retrieved 10 Oct 2009 from http://www.sam.gov.tr/perceptions/Volume8/December2003/mDziedzicbasimicin.pdf

EUROGENDFOR (n.d.) Mission tasks. Retrieved 8 Apr 2011 from http://www.eurogendfor.org/egfpages/missiontasks.aspx

Goldsmith A, Scheptycki J (2007) Crafting transnational policing: police capacity-building and global policing reform. Hart, Oxford/Portland

Greener BK (2009) The new international policing. Palgrave, Houndsmills

Hansen A (2008) Local ownership in peace operations. In: Donais T (ed) Local ownership and security sector reform. Geneva Centre for the Democratic Control of the Armed Forces, Geneva

Hills A (2009) The possibility of transnational policing. Policing Soc 19(3):300–317

Hinton MS, Newburn T (2009) Introduction: policing developing democracies. In: Hinton MS, Newburn T (eds) Policing developing democracies. Routledge, Milton Part

Jayamaha D, Brady S, Fitzgerald B, Fritz J (2010) Lessons learned from US government in international operations. Strategic Studies Institute US Army War College, Carlisle

Keller DE (2010) US military forces and police assistance in stability operations: the least worst option to fill the US capability gap. Strategic Studies Institute US Army War College, Carlisle

Lemay-Hebert N (2009) UNPOL and police reform in Timor-Leste: accomplishments and setbacks. Int Peacekeeping 16(3):393–406

Levine DH (2008) African civilian police capacity for international peacekeeping operations. Stimson Centre, Washington, DC

Mendee J, Last D (2008) Whole of government responses in Mongolia: from domestic response to international implications. The Pearson Papers 11(2) Fall 2008

Ministeria de Defensa de Espana (n.d.) European gendarmerie force. Retrieved 8 Apr 2011 from http://www.mde.es/en/politica/seguridad-defensa/contexto/fuerzas/gendarmeria/

NATO (2004) ‘Multinational specialist unit’, SFOR factsheet, August 2004. http://www.nato.int/sfor/Factsheet/msn/t040809a.html. Accessed 4 Aug 2004

Peake G (2009) Police reform in timor-leste. In: Hinton MS, Newburn T (eds) Policing developing democracies. Routledge, Milton Park

Perito R (2004) Where is the lone ranger when we need him? America’s search for a Postconflict Stability Force. USIP, Washington, DC

Rotmann P (2009) First steps towards a UN police doctrine for peace operations 2001–2006. Geneva Centre for the Democratic Control of the Armed Forces, Geneva

Rubenstein RA (2010) Peacekeeping and the return of imperial policing. Int Peacekeeping 17(4):457–470

Sedra M (2006) Security sector reform in Afghanistan: the slide towards expediency. Int Peacekeeping 13(1):94–110

United Nations Department of Peacekeeping Operations (2011) Monthly summary of contributions (police, military, experts on missions and troops). http://www.un.org/en/peacekeeping/contributors/2011/sept11_1.pdf. Accessed 10 Sept 2011

Wiatrowski MD, Pino N, Pritchard A (2008) Policing and formed police units during democratic transitions. J Secur Sect Manage 6(3):1–14

Williams P (2009) Peace operations in Africa: seven challenges, any solutions? Conflict Trends. Retrieved 10 Sept 2011 from www.kms1.isn.ethz.ch/serviceengine/Files/ISN/109805/. . ./2009347f. . ./1.pdf

Policing Special Populations

Garrett Grothoff
School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA

Overview

Society expects that the police manage a wide variety of problems. Compounding these expectations are special populations that create unique situations that the police must deal with. The nature of these populations as small subsets of society makes them an understudied area in criminal justice research. However, they present important issues for law enforcement, policy makers, and academics. As such, this entry addresses a number of these populations, including domestic violence offenders, persons with mental illness, juveniles, and immigrants, specifically focusing on the control of discretion associated with them. This entry begins with a discussion of the factors that influence discretion in cases of domestic violence and follows by overviewing limitations to discretion during these interactions. Next, it describes the interaction of the police with the mentally ill and the impact of discretion on their decision-making in this context. Then the discussion of juveniles highlights both the formal and informal decisions influenced by discretion. Finally, the interaction between police and immigrants is overviewed in the context of profiling and police perception.

Fundamentals

Domestic Violence

In the United States, domestic violence became a topical issue in the late 1970s. At this time, feminists and women rights advocates started to draw attention to domestic violence as it related to the lack of attention by the police. These groups wanted a response from the criminal justice system to address the concerns they had with police choosing not to arrest. This sparked a push from many police departments to take an aggressive stance toward domestic violence, creating mandatory arrest policies rather than the more traditional responses of advising, mediating, or separating disputants (Robinson and Chandek 2000).

Prior to mandatory arrest and aggressive policies being commonplace, research shows mixed results between the police's use of discretion and domestic violence cases. In an attempt to reduce the problems associated with domestic violence, academics attempted to identify the causes of police discretion. Traditionally, serious violent crimes have been one of the strongest predictors of arrest. However, the police approach domestic violence as a private family issue and have been very reluctant to make an arrest and intervene. This unwillingness to arrest coupled with the belief that domestic violence is a private issue led to the development of a leniency thesis by academics and advocate groups. The leniency thesis proposes that the police are less likely to arrest spouses who assault their wives, compared to other domestic violence victim-offender relationships.

Currently, there is no consensus on whether the leniency thesis reflects reality. However, there are consistent findings that police officers are unlikely to make an arrest in domestic violence cases. Although research on the leniency thesis is mixed, it has helped to shed light on the overall amount of discretion that is exercised by officers in domestic violence cases.

Critics contend that attempts by the police to limit discretion are based on more than just the goals of reducing violence and protecting victims. The police also consider crime control, liability issues, and public opinion. To provide a clear picture of the factors that influence the decision to arrest, it is best to categorize them into demographic, attitudinal, and situational characteristics.

There are three primary demographic characteristics that research focuses on: race, age, and sex. These demographic characteristics are assessed for the responding officer, victim, and offender. The literature on race has not been able to draw firm conclusions. In certain instances, the findings show that minority victims are less likely to have an arrest made on their behalf compared to whites (Black 1971; Ferraro 1989; Smith 1987). However, research does not produce consistent findings when assessing the race of the suspect.

The literature on age and sex characteristics shows that on average, females are less likely to make an arrest than males and older officers are less likely to make an arrest compared to younger officers (Martin 1993; Stalans and Finn 1995). The findings produced by this body of research show that these characteristics might influence officer discretion and should be included in future analyses.

Two primary factors influence the attitudes and discretion of police officers in cases of domestic violence: likelihood of the victim to drop the charges and whether the victim prefers an arrest (Robinson and Chandek 2000). Logic should dictate that if a victim prefers an arrest that the police will be more likely to arrest. The research confirms this hypothesis; however, other situational factors can influence this decision. If the victim used drugs or displayed a negative demeanor, officers were less likely to follow a request (Buzawa and Buzawa 1993). In addition, a reduced likelihood of arrest is replicated in cases where the officer believes the victim will drop the charges.

Previous research on situational characteristics of domestic violence shows that legal rather than extralegal (e.g., attitudinal or demographic characteristics) variables are more important predictors of arrest. This research shows that the use of a weapon during the incident, seriousness of the offense, additional witnesses, use of alcohol or drugs, whether or not the reporter was a female, whether or not the victim was injured, and the incidence of repeat violence significantly increase the chance of an arrest (Avakame and Fyfe 2001; Berk and Loseke 1981; Ferraro 1989; Robinson and Chandek 2000; Worden and Pollitz 1984; Worden 1989). In contrast, if the victim and offender are married, or the suspect leaves the scene, the chance of an arrest decreases.

Attempts to control discretion started from a push to find new ways to reduce domestic violence. The Minneapolis Domestic Violence Experiment (MDVE) was the first attempt to address this by determining whether mandating arrest decreased future domestic violence in a controlled experiment. The Minneapolis experiment, conducted from 1981 to 1982, compared three police options to handle domestic violence situations. The officers were randomly assigned to arrest, remove, or counsel the suspect involved. These procedures were the three standard methods of police response to domestic violence at that time. The experiment was limited to domestic violence cases where both the suspect and victim were present. The results of the experiment indicated that among the three methods, arrest was the most effective at reducing future domestic violence. State legislatures, as a fast fix to the domestic violence problem, quickly adopted the results from Minneapolis. However, the replication of this study in different cities soon showed that the results were short lived.

Researchers began to find in the replicated studies that the deterrent effect of mandatory arrest would soon diminish after a short lag in time. The results after the lag proved contrary to the original finding in Minneapolis; mandatory arrest only increased recidivism rates. Overall, the results of the experiment in Minneapolis and its replications suggest that arrest can reduce domestic violence in some cities but not in others, that it reduces domestic violence among the employed but increases it for the unemployed, and that it has a short-term reduction effect but can increase domestic violence long term. A final point that persisted throughout the domestic violence experiments is the fact that the police in many circumstances can identify those people at risk of suffering future violence. However, the value of privacy and personnel freedom prevents proactive action (Schmidt and Sherman 1993). These studies are an example of an attempt to control police discretion during incidents of domestic violence. Their findings show that this is not an easy task. Controlling or limiting discretion can actually increase the frequency and severity of many problems.

Mental Illness

Police encounters with the mentally ill have increased due to the closure of psychiatric hospitals, overcrowded prisons, and growing population of inmates diagnosed with a mental illness. This population is a serious threat to the police based on the symptoms associated with many mental illnesses. Impaired judgment, disorganized thought, impulsivity, and an impaired ability to perceive risk and protect their person can increase levels of violence during police interactions. Reducing the potential for dangerous encounters is why it is important that the criminal justice system improve the understanding of the police's interactions with the mentally ill. The ability to understand how decisions are made and the use of discretion in these situations can help protect the police, the public, and persons suffering from mental illness. Ultimately a better understanding can lead to better policies and practices that reduce violence and dangerous encounters.

Prior research identified the police as a form of "street-corner psychiatrist" that describes their role as gatekeeper for both the criminal justice and mental health systems. Describing persons with mental illness as a special population refers to their vulnerability and unpopular stigma. Persons with mental illness in general are more prone to have substance use disorders, homelessness, and victimization. Research suggests that persons with mental illness suffer victimization from procedural justice due to their diminished capacity to understand and invoke the criminal justice system on their behalf, demonstrated by the lack of legal resources they are able to obtain. This highlights the duty and value of police officers as first responders to help persons with mental illness find the resources they need within the community (Gur 2010).

Egon Bittner, in the late 1960s, determined that the police, after contacts with the mentally ill, were more likely to use hospitalization or a psychiatric referral in the most severe cases. In general, however, research finds that the police take an informal approach to deal with mentally ill suspects, such as attempting to calm the person or take them home. These tactics used by police in the 1960s lead to the description of police contacts with the mentally ill as "psychiatric first aid" (Bittner 1967). The police were in a position to handle the immediate situation, but had few solutions to solve the larger problem of mental illness.

Two established principles that guide police discretion in interactions with persons with mental illness are the duties of removing dangerous people from the community and protecting citizens that are unable to help themselves. Police officers are primarily responsible for protecting the community first, before the individual from themselves. The police do not have a legal code that dictates how to resolve situations with the mentally ill. That is, there is no rule of law that determines the degree of mental illness and guides officer discretion. It is known that arrest is a rare disposition and informal dispositions are more common. Therefore, the police base their informal decisions on social and structural factors that occur during a situation, rather than the degree of psychiatric health of the suspect (Teplin and Pruett 1992).

Generally, after accounting for necessary control variables, the police are no more likely to arrest persons with mental illness than without. Applying this to encounters with the mentally ill, one finds that most encounters are order maintenance, which increases the amount of discretion used by the officer. Community context is a factor that should not be ignored. High levels of violence and availability of mental health services can affect the discretion used by officers when deciding whether to make an arrest. Compounding this problem, research shows that the mental health system is tougher for police officers to work with than the criminal justice system. In cases where the suspect appears dangerous, mental health officials are reluctant to provide services. Psychiatry research reports higher rates of arrest for persons with mental illness. In these cases, the discretion of an officer to recognize the symptoms of mental illness may be increasing arrests due to misidentification; police officers do not recognize that the suspect is mentally ill (Engel and Silver 2001).

A framework has emerged to describe the discretion used by police officers during encounters with the mentally ill. Morabito (2007) and Bittner (1967) explain that three "horizons," the scenic, temporal, and manipulative, influence police discretion. These three horizons include characteristics of the community, offender, and incident. The development of these horizons has indicated that research has focused primarily on mental illness as the sole predictor of arrest. Research in the field of criminal justice shows that there are many factors that influence police discretion when deciding to arrest. Researchers must also consider environmental and organizational factors when trying to determine the independent impact of mental illness on the decision to arrest. Not controlling for these other factors has oversimplified police discretion, a problem hopefully overcome by using the horizon framework to account for them.

There is a difference in the actions of police officers depending on who initiates the contact with a mentally ill suspect. A contact initiated by the police officer can involve more discretion due to the fact that the police officer's perceptions and attitudes can play a larger role because there is less oversight from supervisors or third-party citizens. In situations when the police respond to a call from a citizen, the wishes of the citizen must be recognized along with an understanding of the behavior that has already taken place (Lamb et al. 2002).

Researchers have also raised the question as to the extent of which mental illness should matter in the decision to arrest, especially in cases of serious crimes. Others maintain that persons with mental illness are not all dangerous. Persons with mental illness commit crimes with differing degrees of seriousness, just like persons free of mental illness. The challenge to the mental health and criminal justice systems is to hold these individuals responsible for their crimes while also providing treatment for those in need.

Extant research suggests that approximately 6–10 % of suspects encountered by the police suffer from a mental illness. The results show that a growing method to reduce police discretion with the mentally ill is the development of Crisis Intervention Teams (CITs). The primary focus of CITs rests on police officer characteristics and training, the goal of which is to reduce the number of mentally ill persons that enter the criminal justice system due to a lack of a better alternative. However, CIT models also incorporate organizational and environmental characteristics that help increase the available resources for police officers when confronting persons with mental illness. Research shows that these teams have a positive, short-term impact. However, for a better understanding of how this model controls discretion, these results must be strengthened; no direct test of the CIT model currently exists (Watson et al. 2008).

Explorations into the control of discretion reveal that mobile crisis teams consisting of police officers, mental health officials, or both can reduce the number of persons with mental illness from entering the criminal justice system. The police are essential in recognizing when persons with mental illness need attention and assisting those people to help find treatment resources. In some circumstances, the use of alcohol or drugs can mask the symptoms of mental illness. This can lead to police officers making an arrest on legal factors that may not be the true cause of the behavior. Compounding this problem even further is the likelihood that a person with mental illness is using alcohol or drugs, possibly intensifying their symptoms. Situations that cause the police to take physical action and subdue a suspect can also hide symptoms of mental illness. Any violence that is displayed during an interaction with the police will increase the likelihood of an arrest. A lack of alternatives leads the police to use an arrest out of mercy and as a final resolution for the problem. Still, when the police do determine that a suspect should be taken to a mental hospital, the mental health officials may present problems, including long waits for treatment that prevent the police from attending to other duties, questioning the judgment of the police that the person is mentally ill, and that the mental health facility may quickly release a suspect that the police view as a danger to society.

Police officers have three choices when handling mentally ill suspects; they can make an arrest, take the suspect to a mental hospital, or solve the problem informally. The situation increases the amount of discretion used by a police officer when making the decision to respond formally or informally. A formal response is either an arrest or hospitalization. Research shows that in most circumstances the first choice is to use an informal method to resolve the situation. However, mental disorder is associated with a higher likelihood of poor demeanor, and police are more likely to arrest a suspect with a poor demeanor. The criminal justice system is not set up for easy entry into the mental health system. Once a record is established, jail or prison is a more likely outcome. Unfortunately, jails and prisons are not the best environment for effective treatment for mental illness (Teplin 2000).

Mental illness is becoming more visible within the community. Although the number of mentally ill is increasing in communities, they are not responding with an increase in acceptance.

Discretion of police officers is limited through legal and bureaucratic restrictions. The mental health bureaucracy can complicate and impede the referral of suspects to service. Many psychiatric programs are cautious to admit dangerous suspects or those with many prior hospitalizations. The criminal justice system does not have the luxury of saying no to dealing with mentally ill suspects. Jails are turned into pseudo-mental health facilities. An increase in police training to help identify suspects that exhibit symptoms of mental illness will help to provide the best service to the right people. Negotiating with hospitals to reduce the number of declined suspects will also help. The best approach may be to take the least restrictive alternative to help control discretion and the criminalization of the mentally ill (Teplin 1984).

Juveniles

Juveniles are becoming a growing problem for the entire criminal justice system. In addition, the point of first contact with the criminal justice system for most youth is with the police. This places the police, as a gatekeeper, in the position to make the first decision on whether formal action should be taken against juveniles in the form of arrest or detention. However, the tradition of the criminal justice system is to treat juveniles more leniently due to a feeling that they are not culpable for their actions. This approach is further perpetuated when the courts choose to release juveniles to their parents, making it difficult for the police to remove troubling youth from the streets. Due to this lenient treatment, in many cases, it does not make taking formal action against juveniles worthwhile for police if there is an alternative solution available.

Coupled with the lenient approach from the system, recent focus on crime control strategies and a push to reduce crime have increased the courts willingness to sentence juveniles that are arrested as adults. The blend of crime reduction strategies along with a traditional lenient approach toward juvenile behavior creates unique problems for the police. The police are now forced to make the decision of whether formal or informal action is more appropriate to solve the situation. The officer making this determination exercises a large amount of discretion and can ultimately decide whether the juvenile will have a record or not.

Previous research on police encounters with juveniles shows that there are far more contacts between the two than what ultimately end up in juvenile court. Few data sets to date have collected information specifically on the interaction between the police and juveniles. Research past and present has relied on data collected on police encounters with juveniles during observational studies with patrol officers, as well as interviews with juveniles.

Encounters with juveniles tend to resemble encounters with adults (Pilivian and Briar 1964; Black and Reiss 1970; Worden and Myers forthcoming). The nature of these encounters is typically not serious, does not result in an arrest, and does not involve any action taken by police that is not based on observed legal factors. Officer-initiated encounters primarily occur as the result of officers seeing wanted youths, in response to reported offenses, or direct observation of law violation and suspicious behavior (Pilivian and Briar 1964). Citizens tend to initiate the majority of police-juvenile encounters. Police contacts with juveniles can also occur through referrals from school officials, parents, and others.

A growing body of policing literature demonstrates a wide latitude of discretion that exists during encounters with juveniles. Discretion as it relates to police encounters with juveniles often depends on the officer making the contact (Pilivian and Briar 1964). Patrol officers exercise different discretion than juvenile officers. Officers that specialize in juvenile cases take a stance that formal action, like an arrest, may do more harm than good. On the other hand, patrol officers are concerned with the demands of the complainant and the requirements needed to remedy a situation. The juvenile system also stresses that the officer uses individual characteristics as a determining factor when deciding to make an arrest. This shift from the offense to the offender increases the amount of discretion an officer has by adding more information.

Both situational and organizational factors limit the control of discretion (Lundman et al. 1978). Legal factors of the situation determine the majority of decisions made by the police when encountering juveniles. However, extralegal factors, like demeanor, still play an important role, especially when a disrespectful demeanor is shown toward the police. Organizational factors that limit police discretion result from departmental policies that require police officers to treat juveniles in a more lenient fashion. Departmental policy can also dictate the number of encounters with juveniles by taking a proactive approach. A proactive approach emphasizes the number of contacts the police have with juveniles, attempting to reduce the ratio of juvenile encounters with citizens and police.

There has been limited research on the control of police discretion pertaining to juvenile encounters. This is partly due to the nature of the juvenile court philosophy that more emphasis should be placed on the character and life situations of the juvenile rather than actual behavior. On the other hand, research shows that when the police do decide to apply formal sanctions to juveniles, they base these decisions on legal factors, such as seriousness of the crime and prior record. Therefore, the most discretion used by the police when encountering juveniles is the decision to invoke the formal criminal justice system or not (Pilivian and Briar 1964).

The recent adoption of community-oriented policing (COP) changes the dynamic between the police and juveniles. COP offers the police a new model to address juvenile crime. Community policing officers approach juveniles with a different attitude. This attitude is more proactive and places a greater emphasis on improving the situations of young people at the neighborhood level. The goals of these officers are to enhance prevention, further diversion efforts, and increase advocacy and parental support (Bazemore and Senjo 1997). However, research is unclear on whether community-oriented policing will result in more diversions or arrests of juveniles.

Another issue relevant to the control of police discretion during juvenile contacts involves the decision to interrogate (Feld 2006). Juvenile interrogations present the police with two problems: a lack of the understanding of the law by the juvenile and the ability to understand their Miranda rights. Psychological research examining juveniles competency of the law shows that youths 15 years of age and younger exhibit the greatest lack of understanding compared to adults (Bonnie and Grisso 2000). Psychology literature also reports that juveniles are more likely to defer, or waive, their Miranda rights due to the social pressure of cooperation with police officers and the misunderstanding that their Miranda rights that are granted to them can be taken away. The United States Supreme Court's rulings in *Yarborough v. Alvarado* and *Fare v. Michael C*. draw the line at 16 years of age for juveniles to be responsible to make the decision to waive their Miranda rights. Technology has increased the control of discretion and subsequently officer actions during interrogations. The taping or video recording of the questioning in interrogations serves as oversight to help protect the rights of juveniles. Some research also recommends that the length of the interviews be limited (Feld 2006). Limiting the length of the interrogation reduces the amount of stress the police can generate to coax a confession.

Immigrants

The legal and illegal immigrant populations in the United States over the past few decades have surged. This has created enclaves within communities with their own specific issues, forcing federal, state, and local law enforcement to adapt new methods of policing to attend to their needs. Census estimates show that Latinos are now the largest minority population in the United States, making up approximately 14 % of the population. However, immigration in the United States includes more than just Latino or Hispanic immigrants. Areas of immigrant populations are found across the United States comprised of people from around the world. Immigrants as a special population present the police with unique challenges, specifically with the use of discretion. There are many interactions with immigrants where the police can exercise discretion. These include racial profiling, search and seizure, and more recently the power to enforce federal immigration laws.

Profiling based on immigrant status and race is a continuing issue that the criminal justice system struggles to reduce. Research on race and police decision-making has shown that race is one of many factors that affect discretion. However, even though racial disparities exist in the criminal justice system, controlling for variables like officer and suspect characteristics, organizational structure, and situational and contextual variables diminishes any direct effects of race. The events surrounding September 11, however, have changed the United States' landscape affecting the use of police discretion toward immigrant populations.

In an effort to prevent another terrorist attack, after those of September 11, law enforcement at all levels is increasing their enforcement of immigration laws. These attacks have also increased the amount of discretion available to law enforcement, mainly through the USA PATRIOT Act of 2001. Local law enforcement now has a role in national security. This discretion is seen firsthand by many immigrants as their privacy in the United States has diminished. However, the extent to which local law enforcement has embraced this increase in discretion is minimal. Many cities with large immigrant populations have adopted mandates that explicitly state immigration enforcement is a federal mandate. Police departments cite many reasons for the adoption of these mandates that include a lack of resources and diminished trust among immigrant groups (Harris 2006).

With immigrant populations, the police may place a stronger emphasis on race. An essential part of any profile includes the race of an individual. The Supreme Court has upheld that the police can use race in combination with other factors to develop a profile. This enhances the amount of discretion afforded to the police when making decisions to stop and search Muslims in airport terminals or legal Hispanic citizens living along the US and Mexican border.

An increase in discretion and closer attention from law enforcement has stressed an already fragile police-immigrant relationship. Immigrant populations are already less likely than whites or blacks to report crimes to the police. This lack of reporting can lead to further reductions in police officers' ability to control crime. In addition, many immigrants carry with them distrust for the police from negative experiences in their home countries. For example, authoritarian governments and countries plagued by high levels of violence can influence immigrants' perceptions of the police in a negative fashion (Davis et al. 2001). These perceptions have led researchers to explore further the potential causes of negative perceptions in hopes of strengthening the police-immigrant relationship. As a result, research has identified three factors that affect the perceptions of immigrants toward the police: social networks, prior homeland experiences, and contact with US immigration officials (Menjiver and Bejarano 2004).

Perceptions of immigrants toward the police are important in how the police use their discretion toward immigrants. A negative perception can manifest into a negative demeanor in police-immigrant interactions. Demeanor is one factor that can persuade a police officer to use a formal rather than informal disposition, such as making an arrest versus writing a warning. Survey research targeting Chinese immigrants supports the claims that perception is important not only to the police but also to the immigrants. In general, these surveys find that a quality police contact was more important than the quantity of contacts. The views of the Chinese immigrants were that attempts to improve the quality of police services, increase police officers understanding of the language and culture, would increase immigrants' satisfaction with the police (Chu et al. 2005).

The police may have a difficult time addressing these views. Many attempts to improve police officer relationships with immigrant populations are accomplished through community-oriented policing strategies. Community-oriented policing strategies place the officers closer to the public and encourage interaction to address community problems directly. However, community-oriented policing increases the amount of discretion used by police officers, discretion that many immigrant populations are fighting to limit.

Perception and overcoming profiling are not the only immigrant concerns for the police. Language many times can force police officers to use more discretion with immigrant populations than with English-speaking citizens. This presents severe constitutional problems with the interpretation and understanding of suspect rights when an immigrant that does not speak or understand English is read the Miranda warning. However, police departments and officers have taken steps to account for the language barrier. Along the US-Mexico border, Miranda warnings are read in English and Spanish. Many departments in areas with large immigrant populations actively recruit bilingual officers. These have been accompanied with changes to interrogation policies and the presence of an interpreter if necessary.

Conclusion

Police discretion is ever present in the criminal justice system. Special populations of citizens present unique circumstances and situations that increase and decrease the amount of discretion police officers use. These populations include cases of domestic violence, persons with mental illness, juveniles, and immigrants. Research in these areas has unfortunately found that there is no easy way to control police discretion. In some circumstances, findings show that a certain amount of discretion is a positive. Discretion allows the police to invoke informal rather than formal sanctions that reduce the number of people entering the criminal justice system. Discretion also allows the police to direct offenders in need of special healthcare services to resources that can provide alternatives to jail or prison. In other instances, discretion has led to the potential of police infringing on citizens' constitutional rights. Finally, discretion leaves the police with a choice, one that in cases of domestic violence can leave a victim in a dangerous situation.

In all, the influences on police decision-making are both positive and negative making police discretion hard to control. The special populations here demonstrate these points. Future research should focus on managing discretion in a manner that best serves the public's interests. The police, guided by more research, must overcome the hurdle of managing discretion in a way that does no harm to the public or impedes officers from accomplishing their goals.

Related Entries

▶ Community Policing
▶ Police Discretion and its Control

Recommended Reading and References

Avakame F, Fyfe JJ (2001) Differential police treatment of male-on-female spousal violence: additional evidence on the leniency thesis. Violence Against Women 7:22–45

Bazemore G, Senjo S (1997) Police encounters juveniles revisited: an exploratory study of themes and styles in community policing. Policing Int J Police Strategy Manag 20:60–82

Bittner E (1967) Police discretion in emergency apprehension of mentally ill persons. Soc Probl 14:278–292

Black DJ, Reiss AJ Jr (1970) Police control of juveniles. Am Sociol Rev 35:63–77

Castellano-Hoyt DW (2003) Enhancing police response to persons in mental health crisis: providing strategies, communication techniques, and crisis intervention preparation in overcoming institutional challenges. Charles C. Thomas, Springfield

Chu D, Huey-Long Song J, Dombrink J (2005) Chinese immigrants' perceptions of the police in New York City. Int Crim Just Rev 15:101–114

Decker S (2003) Policing gangs and youth violence, 3rd edn. Wadsworth, Belmont

Eitle D (2005) The influence of mandatory arrest policies, police organizational characteristics, and situational variables on the probability of arrest in domestic violence cases. Crime Delinq 51:573–597

Engel RS, Silver E (2001) Policing mentally disordered suspects: a reexamination of the criminalization hypothesis. Criminology 39:225–252

Feder L (1996) Police handling of domestic calls: the importance of offender's presence in the arrest decision. J Crim Justice 24:481–490

Feld B (2006) Police interrogation of juveniles: an empirical study of policy and practice. J Crim Law Criminol 97:219–316

Fyfe JJ, Klinger DA, Flavin JM (1997) Differential police treatment of male-on-female spousal violence. Criminology 35:455–473

Gur OM (2010) Persons with mental illness in the criminal justice system: police interventions to prevent violence and criminalization. J Police Crisis Negot 10:220–240

Harris DA (2006) The war on terror, local police, and immigration enforcement: a curious tale of police power in post-9/11 America. Rutgers Law J 38:1–60

Johnson KR (2000) The case against race profiling in immigration enforcement. Wash Univ Law Q 78:531–592

Kane RJ (1999) Patterns of arrest in domestic violence encounters: identifying a police decision-making model. J Crim Justice 27:65–79

Kenney J, Fuller D, Barry R (1995) Police work with juveniles and the administration of juvenile justice, 8th edn. C.C. Thomas, Springfield

Klinger DA (1995) Policing spousal assault. J Res Crime Delinq 32:308–324

Laml HR, Weinberger LE, DeCuir WJ (2002) The police and mental health. Psychiatr Serv. doi:10.1176/appi.ps.53.10.1266

Lundman RJ, Sykes RE, Clark JP (1978) Police control of juveniles. J Res Crime Delinq 15:74–91

Menjiver C, Bejarano CL (2004) Latino immigrants' perceptions of crime and police authorities in the United States: a case study from the phoenix metropolitan area. Ethnic Racial Stud 27:120–148

Morabito MS (2007) Horizons of context: understanding the police decision to arrest people with mental illness. Psychiatr Serv 58:1582–1587

National Research Council (2004) Fairness and effectiveness in policing: the evidence, committee to review research on police policy and practices. In: Skogan W, Frydl K (eds) Committee on Law and justice, division of behavioral and social sciences and education. The National Academies Press, Washington, DC

Pilivian I, Briar S (1964) Police encounters with juveniles. Am J Sociol 70:206–214

Ratcliffe J (2008) Intelligence-led policing. Willan, Cullompton

Robinson AL, Chandek MS (2000) The domestic violence arrest decision: examining demographic, attitudinal, and situational variables. Crime Delinq 46:18–37

Smith DA (1987) Police response to interpersonal violence: defining the parameters of legal control. Soc Forces 65:767–782

Taylor T, Turner KB, Esbensen F, Winfree LT Jr (2001) Coppin' an attitude: attitudinal differences among juveniles toward police. J Crim Justice 29:295–305

Teplin LA (1984) Criminalizing mental disorder: the comparative arrest rate of the mentally ill. Am Psychol 39:794–803

Teplin LA (2000) Keeping the peace: police discretion and mentally ill persons. Natl Inst Justice J 244:1–8

Teplin LA, Pruett NS (1992) Police as street corner psychiatrist: managing the mentally ill. Int J Law Psychiatry 15:139–156

Waddington PAJ, Stenson K, Don D (2004) In proportion: race, and police stop and search. Brit J Criminol 44:889–914

Watson AC, Morabito MS, Draine J, Ottati V (2008) Improving police response to persons with mental illness: a multi-level conceptualization of CIT. Int J Law Psychiatry 31:359–368

Worden RE, Myers SM (forthcoming) Police encounters with juvenile suspects

Worden RE, Pollitz AA (1984) Police arrests in domestic disturbances: a further look. Law Soc Rev 18:105–120

Policing Terrorism and Legitimacy

Tal Jonathan-Zamir
Institute of Criminology, Faculty of Law,
The Hebrew University of Jerusalem,
Jerusalem, Israel

Overview

Over the past decade police agencies in the democratic world have increased their involvement in counterterrorism. This new focus, in turn, has generated much debate (and limited empirical research) on the unintended outcomes of policing terrorism, such as its effects on the ability of the police to address crime and carry out other "classic" policing duties, and their relationship with the public. At the same time, discussions about the effects of counterterrorism policing on police-community relationships have focused primarily on minority, Muslim or Arab groups, while much less attention has been paid to *majority*, non-Muslim communities.

Nevertheless, the few studies that have tackled the effects of policing terrorism on the relationship between the police and majority groups indicate that there are important issues that should be addressed in this context. First, as reviewed below, research reveals that terrorism threats and police involvement in this domain, in the long run, bring about a drop in majority communities' assessments of the police. With regard to the impact of policing terrorism on the processes by which public evaluations of the police are formed, studies show that majority groups remain very much concerned with procedural fairness and police legitimacy, even when faced with intense security threats or when considering counterterrorism policies and cooperation with the police in their efforts to combat terrorism. These communities were also found to be aware of the potential negative outcomes of police focus on counterterrorism, such as neglect of "classic" policing duties and negative effects on the relationship between the police and Muslim or Arab communities, and, finally, research reveals that majority communities remain sensitive to procedural injustice in policing terrorism, even when directed at Muslim communities. Overall, and perhaps contrary to intuitive thinking, it appears that majority groups *do not* abandon considerations of fair policing in the context of counterterrorism and, while expecting security from the threat, remain equally concerned with procedural justice and place much value on the legitimacy of the police.

Background

Since the terror attack of September 11, 2001, police forces in the United States, as well as in other Western democracies, have become significantly more involved in counterterrorism and other security tasks in comparison to previous decades (Bayley and Weisburd 2009; International Association of Chiefs of Police 2005; National Research Council 2004). Because such responsibilities were relatively new for local police agencies in countries like the United States and not necessarily perceived as "natural" police roles (International Association of Chiefs of Police 2005; Weisburd et al. 2009a), this new focus on counterterrorism gave rise to much debate, speculation, and some empirical research about the effects these tasks may have on the police, their organization, priorities, performance and character, as well as their role in society more generally (e.g., Weisburd et al. 2009b).

One major concern raised by scholars focused on the potential effects of police involvement in counterterrorism on their relationship with the public. Bayley and Weisburd (2009), for example, have argued that extensive homeland security responsibilities may change the focus of the police from "low" to "high" policing (Brodeur 1983), which may affect both their goals and strategies and their character and interactions with citizens. The focus of "high policing" is on macro-level issues rather than local problems. The methods of this policing style are characterized by covert tactics such as surveillance and intelligence gathering, which are often less transparent and accountable, and associated with violations of human rights and procedural justice. Thus, they argue, "high policing" may change the orientation of the police from providing service and viewing citizens as clients to controlling the public and viewing citizens as suspects.

Mastrofski (2006) raised similar concerns in the context of the future of community policing. He suggested that police involvement in counterterrorism may subordinate the community to the police and undermine police accountability and responsiveness to the public. He further argued that focusing on counterterrorism may result in close association between the police and military or intelligence agencies, which may contribute to the militarization of the police, which, in turn, may hamper recent efforts to bring the police and the community closer together. Similar arguments and concerns were raised by other scholars as well (e.g., Hasisi et al. 2009; Murray 2005; Thacher 2005; Weisburd et al. 2009a).

Importantly, however, such concerns were generally raised with regard to minority Arab or Muslim communities. Scholars and practitioners warned that this sector may be perceived as the "enemy within" due to ethnic or national links to the source of the terrorism threat (Hasisi et al. 2009; Weisburd et al. 2009a), and thus their relationship with the police is particularly vulnerable (Henderson et al. 2006; Thacher 2005; Tyler et al. 2010). Much less attention has been paid to the effects of policing terrorism on *majority/non-Muslim communities* and their relationship with the police, possibly because of the common assumption that members of these communities and the police are "on the same side" in the struggle against terrorism, and thus there is no reason to suspect that the policing of terrorism would impact their partnerships. However, as detailed below, there are important reasons to suspect that this is not necessarily the case, and in fact numerous questions and concerns could be raised in this context. Studies that have examined some of these issues, and have indeed identified various effects (or lack of effects) of policing terrorism on public preferences, expectations of and general attitudes toward the police among majority communities, are reviewed in the following sections followed by an integrative discussion, conclusions, and suggestions for future research.

The Study of the Relationship Between the Police and Majority Communities in the Context of Policing Terrorism

What are the effects of police involvement in counterterrorism on the relationship between the police and majority, non-Muslim or non-Arab communities? What would this relationship look like in places or in times of intense terrorism threats or where the police devote much of their resources to counterterrorism? These broad questions may be broken down into many, more specific inquiries. One may wonder, for example, about the correlation between policing terrorism and public evaluations of the police. Does extensive police involvement in counterterrorism impact the image of the police in the public eye? If so, do evaluations improve or weaken? A different set of questions concern the effects of security threats and police involvement in counterterrorism not on absolute levels of support or trust in the police, but rather on the *processes* by which such evaluations are formed. When assessing police legitimacy in situations of threat, or when considering cooperating with the police in their counterterrorism efforts, do citizens take into account the same factors they would in the context of fighting crime or when there are no specific security threats in the background? Do they give these factors the same weight?

One may also ask if majority communities, while experiencing fear and vulnerability in the face of ongoing security threats, are also aware of the potential costs of police focus on counterterrorism, such as reduced emphasis on crime control or change in the nature of policing services they are receiving. A related question concerns the responses of majority communities to perceived mistreatment of minorities. If the police are viewed as treating minorities in an unfair or biased fashion while engaging in counterterrorism, would this impact their image in the eyes of *majority* communities? These issues, which have been examined by researchers, make up the current state of knowledge about the relationship between the police and majority communities in the context of policing terrorism. The studies that have tackled these questions are reviewed below, followed by an integrative discussion. It is important to emphasize, however, that the picture is far from complete, as many more questions could be asked and examined in this context. Some directions for future research will be discussed in the last section of the entry.

What Is the Relationship Between Police Involvement in Counterterrorism and Public Evaluations of the Police over Time?

Is there a relationship between the level of terrorism threats (and thus police involvement in counterterrorism) and public attitudes toward the police? If so, does focusing on counterterrorism improve or alternatively weaken public support? These questions were examined by Jonathan (2010) within the Israeli context, where the national police force is responsible for and treats all terrorist threats or attacks within Israeli borders. As reviewed by this author, prior to empirical examination both hypotheses appear plausible. On the one hand, scholars have argued that policing terrorism may weaken public assessments of the police, both as a result of change in the nature and character of policing and because the preoccupation with security tasks may come at the expense of handling local crime and disorder problems (e.g., Bayley and Weisburd 2009; Hasisi et al. 2009; Weisburd et al. 2009a, 2010).

On the other hand, there are also reasons to suspect that extensive involvement in counterterrorism could actually *improve* the public image of the police: in times of acute security threats, by focusing on counterterrorism the police may be perceived as responding to the problem that is mostly troubling the public; terrorism threats may encourage collaboration between the police and community members/organizations; security-related matters are often viewed as "high status" and "prestigious" by majority communities and may thus improve the overall image of the police; and, finally, the police may appear highly efficient and professional in their rapid, full-scale counterterrorism responses, which may enhance assessments of police competence (see Weisburd et al. 2009a, also see review by Jonathan 2010).

Following these mixed hypotheses, Jonathan (2010) examined fluctuations in numerous types of public attitudes toward the Israel National Police between 1998 and 2007 and, specifically, compared assessments across three time points, each representing a different situation of terrorism threats around the "Second Palestinian Intifada": 2000, before the outbreak of the Intifada; 2002, the height of the terrorism threats of the Intifada; and 2007, when terrorism threats returned to pre-Intifada levels. This author identified a rise in public assessments of the police following the rise in terrorism threats and a peak in public support corresponding with the peak in threat levels. However, once the threat began to drop, so did evaluations of the police, often to levels lower than those measured at the beginning of the decade, prior to the high-threat period of the Intifada.

These findings are attributed to the tendency for internal cohesion in the face of an external threat (e.g., Stein 1976) and, specifically, to the "Rally 'Round the Flag" Effect (Mueller 1970), according to which, in times of international crises, support for the national leader and for other public institutions increases, but only for a limited period of time. Once the crisis has passed or, alternatively, drags on longer than expected, support typically returns to earlier levels. The apparently conflicting hypotheses mentioned earlier also seem to be supported: in the short term, assessments of the police among majority communities improve as a result of police involvement in counterterrorism. In the long term, however, the negative effects appear to take their toll and bring about a drop in public evaluations. Importantly, similar fluctuations in attitudes were identified in the USA around the 9/11 terror attack (see Shaw and Brannan 2009).

Do Security Threats Impact the Factors Citizen Consider when Forming Evaluations of Police Legitimacy?

A second important question examined by scholars in the context of the relationship between the police and majority community in situations of threat concerns the factors citizens consider and value when assessing police legitimacy. The legitimacy of the police in the eyes of the public is frequently viewed as a key outcome of policing in democratic societies (National Research Council 2004). By "police legitimacy" scholars often refer to personal, subjective beliefs that the agency or the rule ought to be obeyed (see Weber 1968), or to "the belief that the police are entitled to call upon the public to follow the law and help combat crime, and that members of the public have an obligation to engage in cooperative behaviors" (Tyler 2004, pp. 86–87).

The critical importance of police legitimacy lays in both normative and instrumental reasons. In democracies, where the police receive their authority and powers from the public, it is expected that they would be perceived as an expression of the community, serving the public interest, rather than alien and distant from citizens and their needs. From the practical perspective, police legitimacy was found to result in numerous desirable outcomes, such as compliance with the law, acceptance of police decisions, and willingness to collaborate with the police, provide information, assist in solving crime and empower the police (e.g., Reisig et al. 2007; Sunshine and Tyler 2003; Tyler 2004, 2009; Tyler and Fagan 2008; also see review by the National Research Council 2004).

Given the importance of police legitimacy, scholars have devoted much attention to identifying the factors that shape legitimacy assessments. These studies have repeatedly shown that the extent to which the police are perceived as legitimate depends primarily on evaluations of *the fairness of the processes* by which they exercise their authority, or "procedural justice" (National Research Council 2004; Reisig et al. 2007; Tyler 2004, 2009; Tyler et al. 2010). A second important factor was found to be evaluations of the ability of the police to "bring results" – catch rule breakers, for example, or control crime. However, although playing an important role in predicting police legitimacy, these instrumental considerations were consistently found to be *secondary* to procedural justice in their importance (Sunshine and Tyler 2003; Tyler 2004, 2009).

Nevertheless, according to Jonathan-Zamir and Weisburd (2013), it should not be taken for granted that "procedural justice" would remain the primary predictor of police legitimacy when the public is faced with acute, ongoing security threats. They argue that under such circumstances, when citizens are experiencing insecurity, fear and vulnerability, they may be more concerned with end results, i.e. with the police handling the threat and providing security, and less worried about fair processes. Thus, under threat, the role of police performance in shaping legitimacy may rise, while the importance of procedural justice may drop. What is more, assessments of police performance may overtake procedural justice and become the primary antecedent of police legitimacy. Similar suggestions were raised by Huq et al. (2011), who argued that in the context of counterterrorism policing, non-Muslims may view the threat of terrorism as particularly dangerous or morally odious, or believe that only Muslim or Arab communities will bear the costs of policing terrorism, and thus these non-Muslim communities may become less sensitive to procedural fairness.

Jonathan-Zamir and Weisburd (2013) examined these hypotheses by comparing predictors of police legitimacy across two groups of communities in Israel: one facing immediate, severe security threats (missile attacks originating from the Gaza Strip) and the other facing no specific threats at the time. They found that while assessments of the performance of the police did rise in importance for the public in situations of threat, evaluations of the fairness of police processes *did not* decline in importance and, what is more, remained the primary predictor of police legitimacy in both conditions. These results suggest that security threats impact public expectations regarding the performance and efficiency of the police (which become more important in assessing police legitimacy). At the same time, they also reveal what *does not* change in situations of security threats: under such circumstances, majority communities *do not* become less concerned with fairness in police treatment, and these assessments remain the primary predictor of police legitimacy.

An analysis of survey data from New York City carried out by Huq et al. (2011) reveals similar results. Among non-Muslim communities, assessments of procedural justice in the *implementation* of counterterrorism policy were the primary predictor of police legitimacy, followed by evaluations of procedural justice in the *formation* of counterterrorism policy. Other potentially important factors, such as the extent to which the police make the respondent feel safe from terrorist threats or the effectiveness the police show in fighting terrorism, were found to be either statistically nonsignificant in predicting police legitimacy or significant but less influential than procedural justice assessments. These analyses clearly show that the importance of procedural fairness for the public, which has been repeatedly demonstrated in the context of crime control, does not decline for non-Muslim communities in the context of policing terrorism.

Which Factors Impact the Willingness of Majority Communities to Cooperate with the Police in Their Efforts to Fight Terrorism?

As mentioned above, the legitimacy of the police in the eyes of citizens was found to have important effects on their willingness to engage in cooperative behaviors, at least on the declarative level. Sunshine and Tyler (2003), for example, found among two samples of New York City residents that evaluations of police legitimacy were associated with willingness to report crimes to the police, provide information, take part in police-community activities, and empower the police. Similarly, Tyler and Fagan (2008) identified that individuals who viewed the police as legitimate were more likely to state they would cooperate with the police by fighting crime in their neighborhoods. These conclusions are supported by other researchers as well (e.g., Mastrofski et al. 1996; Tyler and Wakslak 2004).

At the same time, and similarly to the previous question about antecedents of police legitimacy, it cannot be taken for granted that assessments of police legitimacy would have the same impact on willingness to cooperate with the police in the context of combating terrorism. As argued by Huq et al. (2011), majority or non-Muslim communities are likely to perceive terrorism threats as particularly salient or harmful, and may hold different normative and ideological beliefs about terrorism (in comparison to crime). Thus, it is reasonable to expect that factors such as perceptions regarding the seriousness of the terrorism threat or the effectiveness of the police in addressing it, would become particularly salient in shaping cooperative behaviors.

Nevertheless, and similarly to studies carried out in the context of crime control, the analysis reported by Huq et al. (2011) reveals the critical importance of police legitimacy in shaping cooperation among non-Muslims in the context of policing terrorism as well. Assessments of police legitimacy were found to be the most important factor predicting the willingness of these communities to alert the police to potential terrorism risks, more influential than evaluations of police effectiveness in combating terrorism, the extent to which the police make the respondent feel safe from terrorism threats, or sociodemographic factors (although views regarding the seriousness of the terrorism threat were almost as important as assessments of police legitimacy). When predicting the willingness of non-Muslims to cooperate with the police by taking part in educational efforts about the dangers imposed by potential terrorists and by encouraging community members to cooperate with law enforcement officials in their efforts to combat terrorism, legitimacy assessments were found to be the second strongest predictor, surpassed only by race.

Assessments of procedural justice in both the formation and implementation of counterterrorism policies were also found to have direct and strong effects on overall willingness to cooperate with the police in their efforts to fight terrorism. It should be noted, however, that perceptions regarding the seriousness of the terrorism threat were the primary predictor in this analysis. Nevertheless, similarly to the results reviewed above regarding antecedents of police legitimacy under threat, the legitimacy of the police in the eyes of the public, as well as the fairness of the processes by which they exercise their authority, remains critical in shaping willingness to cooperate with the police among majority communities in the context of policing terrorism.

Do Majority Communities Consider the Potential Negative Implications of Police Involvement in Counterterrorism?

The research findings reviewed above raise another important question about majority communities and the involvement of local police agencies in counterterrorism. It is often assumed that majority communities desire and support harsh responses to terrorism threats (see, e.g., Gordon and Arian 2001; Huddy et al. 2005), and as reviewed above, there is evidence suggesting that members of these communities evaluate the police more positively in times of severe security threats, at least in the short term (Jonathan 2010). But at the same time, do they also consider the potential costs of extensive police involvement in counterterrorism, such as reduced focus on crime control, order maintenance and other "classic" police responsibilities, or change in the nature and character of policing in the direction of "high policing?" (see Bayley and Weisburd 2009).

Jonathan and Weisburd (2010) have examined this question using a large-scale community survey in Israel, and found that, overall, majority communities are indeed aware of and consider at least some negative outcomes of extensive police involvement in counterterrorism. Many respondents to their survey believe that policing terrorism negatively affects the relationship between the Israeli police and the public, particularly Arab communities. Respondents also expressed strong agreement with the notion that policing terrorism in Israel comes at the expense of other police responsibilities, such as fighting crime and enforcing traffic regulations. Moreover, many hold that counterterrorism is often used by the police as an excuse for weak performance in fighting crime.

These results suggest that majority communities are more sophisticated in their expectations from the police than is often assumed, and their desires are complex. While possibly supporting harsh counterterrorism responses (e.g., Gordon and Arian 2001; Huddy et al. 2005), these citizens are also well aware of the potential, unintended negative outcomes. Thus, it appears that in order to maintain public trust in the police and police legitimacy, even in the face of security threats the police cannot focus solely on providing rapid, forceful responses and abandon adequate performance in "classic" domains, fair processes and appropriate interactions with the public.

Are Majority Communities Affected by the Way Minority Groups Are Treated by the Police?

As reviewed above, majority Jewish communities in Israel were found to be very much aware of the potential negative effects of policing terrorism on the relationship between the police and minority Arab communities. However, Jonathan and Weisburd (2010) have not addressed the question of whether differential treatment of these minority communities affects the relationship between the police and *majority* groups. In other words, are majority communities sensitive to unfair treatment when it is directed at what may be perceived as an out-group, possibly associated with terrorism threats?

This effect, which was observed by Tyler and Wakslak (2004) in the context of crime control and termed the "spillover effect," was examined by Huq et al. (2011) in the context of policing terrorism. Survey data collected by these authors in New York City revealed that evaluations that the police are treating Muslims disrespectfully in their counterterrorism efforts, or that they are increasingly targeting Muslim communities, had significant and negative effects on assessments of both procedural justice and police legitimacy among majority, non-Muslim communities. Interestingly, beliefs regarding heightened suspicion of Muslims had no significant effects. Thus, simply viewing Muslim communities as a more risky group was not, in and of itself, a problem in the eyes of majority communities. However, disrespectful treatment or increased targeting of Muslims had a negative effect on how majority groups view the police. Finally, evaluations of procedural justice and police legitimacy among non-Muslims were also significantly and negatively affected by assessments of "public police intrusions" directed at minorities (public police activities, such as searching bags at train stations).

Huq et al. (2011) conclude that when the police, as part of their counterterrorism efforts, behave toward minority groups in a way that is perceived to be unfair, they lose legitimacy among *majority* communities as well. In other words, in the context of policing terrorism (as is the case with policing crime), it appears that the relationship between the police and majority communities is influenced not only by direct effects, such as reduced focus on crime control or a more militaristic style of policing, but also by perceptions regarding how the "other" group is being treated, even if that group is viewed as a potential source of threat. These results again highlight the importance of fair treatment in the context of policing terrorism.

Discussion and Conclusions

How does policing terrorism impact the relationship between the police and majority groups? What should local police agencies take into consideration when thinking about their partnership with these communities in "the new era of counterterrorism policing?" Before addressing the specific effects (or lack of effects) identified by scholars, it is important to stress that the studies reviewed above, first and foremost, draw attention to the importance of thinking about and investigating the relationship between the police and *majority* communities in the context of policing terrorism. To date, most discussions and empirical research on counterterrorism and police-community relationship have focused on minority, Muslim or Arab groups. While clearly important, propositions and hypotheses raised by scholars, as well as findings from the few studies that *have* included majority communities, highlight the importance of focusing on this group as well.

With regard to the specific implications of counterterrorism policing for majority groups and their partnerships with the police, studies have approached the topic from different angles and thus provide different answers to this question. At the same time, as detailed below, commonalities do emerge. First, majority communities' assessments of the police do appear to be affected by both the terrorism threat itself and by police involvement in this domain. In the short term, public support for the police and evaluations of their performance and legitimacy improve. In the long term, however, these assessments return to previous levels and, what is more, often drop to levels lower than those measured prior to the high-threat period.

One of the explanations offered for these fluctuations is that, over time, the negative, unintended outcomes of policing terrorism begin to take their toll and impact majority communities' attitudes toward the police. The survey specifically addressing perceptions of the costs of policing terrorism reinforces the premise that majority communities consider, or are at least aware of, some of the prices paid for extensive police involvement in counterterrorism. Whether it is reduced emphasis on crime control or a more militaristic, "high policing" style of policing, these costs appear to impact the views of majority communities, as evidenced by the drop in the public image of the police over time. Perhaps, the most critical is the drop in public trust in the police, which, in addition to being a cost in its own right, may significantly hamper public cooperation and acceptance of police authority (e.g., Tyler 2004, 2009). In this context, it should also be stressed that any immediate improvement in public attitudes toward the police in times of intense security threats should be viewed cautiously, as it is, in all likelihood, fragile, temporary by nature, and in large part a result of the threat itself rather than specific police responses. Such peaks in support may also result in members of the public temporarily ignoring or undervaluing police mistakes or mistreatment.

The second approach taken by scholars to the study of the relationship between the police and majority groups in the context of policing terrorism utilized the "legitimacy" or "procedural justice" model, as specified by Tyler (2004, 2009) and others (e.g., Huq et al. 2011; Reisig et al. 2007; Sunshine and Tyler 2003; Tyler and Fagan 2008). The main question here is whether the basic relationships of the model, as identified in previous studies in the context of crime control, hold for majority communities in the context of counterterrorism policing. Specifically, would "procedural justice" retain its importance as the primary antecedent of police legitimacy? Would legitimacy assessments, in turn, remain a major predictor of willingness to cooperate with the police? Finally, would evaluations of the way the police treat Muslim minorities impact the legitimacy of the police and assessments of their fairness in the eyes of majority groups?

As reviewed above, there are compelling reasons to suspect that the legitimacy model would work differently for majority communities in the context of policing terrorism, primarily because terrorism threats appear to bring about much fear, uncertainty and vulnerability, as well as ideological perspectives that are different from those generated by crime threats. Moreover, majority communities may believe that counterterrorism policing and any procedural injustice associated with it would not be directed at them, but at the minority out-group linked to terrorism threats. Thus, it is reasonable to suspect that considerations regarding the fairness of police processes or the legitimacy of the police more generally would decline in importance, while more instrumental, outcome-oriented concerns such as the magnitude of the threat or successes the police show in addressing it, would have the most salient impact on citizens.

At the same time, studies carried out both in Israel and in the USA reveal that, overall, this is not the case. Even when facing intense security threats or when asked specifically about counterterrorism policies, assessments regarding the fairness of police processes remain the most important antecedent of police legitimacy. While the performance or effectiveness of the police may rise in importance in situations of threat, it nevertheless remains secondary to procedural justice in predicting police legitimacy, as do other factors such as the sense of security the police are providing or sociodemographic factors. Police legitimacy, in turn, remains a major predictor of willingness to cooperate with the police in their counterterrorism efforts, including willingness to alert the police to potential terrorist threats and take part in educational efforts about counterterrorism (although other, instrumental considerations were also found to be important in this context).

Research findings also reveal that majority communities are not indifferent to the way the police treat Muslim minority groups, even if they are perceived as potentially linked to terrorism threats. In addition to being aware of the negative effects of policing terrorism on the relationship between the police and minority Muslim or Arab communities, beliefs that the police are treating minorities disrespectfully, or are increasingly targeting them, hamper the views of majority groups regarding police use of fair processes, as well as their legitimacy more generally.

In sum, it appears that, contrary to intuitive views, when faced with security threats or when considering counterterrorism in particular, majority groups *do not* put aside daily concerns with "ordinary policing" and fair processes and become solely interested in forceful, rapid responses. Importantly, this is not to say that instrumental considerations do not matter. The studies reviewed above reveal that majority communities *do* value police performance more under threat; support and evaluate the police more positively during high-threat periods; and, when considering cooperating with the police in their counterterrorism efforts, *do* value instrumental considerations, such as the seriousness of the terrorism threat.

At the same time, majority groups are also well aware of the costs of policing terrorism and their negative effects on fighting crime and police-community relationships, and their evaluations of the police, including perceptions of trust, procedural justice and performance, dramatically drop over time, perhaps, at least in part, because of these negative outcomes. Majority communities also desire and value fair processes *more* than end results when considering police legitimacy, even under situations of acute security threats or when specifically considering counterterrorism policing. Moreover, assessments of police legitimacy remain a major predictor of majority communities' willingness to cooperate with the police in fighting terrorism, and, finally, it appears that majority groups are highly sensitive to procedural injustice, even when directed at minority Muslim communities. Thus, it appears that majority groups are not willing to "give the police an exemption" in the face of terrorism just as long as they provide protection from the threat. Members of these communities still expect the police to address crime and other "classic" problems and, importantly, insist on fair, polite, respectable processes.

It should be noted, however, the picture is far from complete, as many more questions could be asked and examined in this context. For example, the relationships between the different components of the legitimacy model, as well as public assessments of the costs of policing terrorism, have been tested using public surveys, and thus, the findings reviewed above are based solely on citizens' statements. It is therefore not clear if, for example, we would actually find more cooperation with the police in their counterterrorism efforts in places or in times where they are perceived as more legitimate. We also know little about how citizens form their evaluations of police use of procedural justice in countering terrorism or assessments regarding the prices paid for police focus on counterterrorism. Are these evaluations based on personal experiences, the experiences of friends or family members, or media coverage of policing terrorism?

Hypotheses regarding the outcomes of policing terrorism and, in turn, their effects on police-community relationships also need to be tested. For example, does focusing on counterterrorism indeed change the nature and character of policing in the direction of "high policing?" Does it affect crime control? While there is some evidence supporting these propositions (e.g., Weisburd et al. 2010), they clearly need to be examined further. In turn, if these outcomes are indeed identified, do they actually affect the partnership between the police and the pubic in the real world, and, if so, in what ways? Finally, questions could be asked about mitigating the potential negative outcomes of policing terrorism. Given that local police agencies sometimes have no choice but to focus on counterterrorism, at least for limited periods of time, what measures can they take in order to preserve, and perhaps even strengthen, their relationships with majority groups while still engaging in effective counterterrorism policing?

Conclusions

While often not the focus of discussions or research, policing terrorism bears important implications for the relationship between the police and *majority* groups. Perhaps most significant are the research findings indicating that, contrary to intuitive views, even in the face of major security threats or in the context of policing terrorism more generally, and even when considering the relationship between the police and minority Muslim communities, majority groups *do not* abandon their concerns with fair processes and legitimate policing. Any preconceptions that when facing terrorism threats, majority communities become only interested in security and support forceful, rapid police responses no matter the cost are misleading, and if utilized as the basis for counterterrorism policies are likely to have negative effects on the relationship between the police and majority groups.

Related Entries

- ▶ Causes of Police Legitimacy
- ▶ Policing of Peacekeeping
- ▶ Procedural Justice, Legitimacy, and Policing

Recommended Reading and References

Bayley D, Weisburd D (2009) Cops and spooks: the role of the police in counterterrorism. In: Weisburd D, Feucht T, Hakimi I, Lois M, Perry S (eds) To protect and to serve: policing in an age of terrorism. Springer, New York, pp 81–99

Brodeur J (1983) High and low policing: remarks about the policing of political activities. Soc Probl 30(5):507–520

Gordon C, Arian A (2001) Threat and decision making. J Confl Resolut 45(2):197–215

Hasisi B, Alpert G, Flynn D (2009) The impacts of policing terrorism on society: lessons from Israel and the U.S. In: Weisburd D, Feucht T, Hakimi I, Lois M, Perry S (eds) To protect and to serve: policing in an age of terrorism. Springer, New York, pp 177–202

Henderson NJ, Ortiz CW, Sugie NF, Miller J (2006) Law enforcement and Arab American community relations after September 11, 2001: technical report. Vera Institute of Justice, New York

Huddy L, Feldman S, Taber C, Lahav G (2005) Threat, anxiety, and support of anti-terrorism policies. Am J Polit Sci 49(3):610–625

Huq A, Tyler T, Schulhofer S (2011) Why does the public cooperate with law enforcement? The influence of the purposes and targets of policing. Psychol Public Policy Law 17(3):419–450

International Association of Chiefs of Police (2005) Post 9-11 policing: the crime-control-homeland security paradigm – taking command of new realities. International Association of Chiefs of Police, Alexandria

Jonathan T (2010) Police involvement in counterterrorism and public attitudes towards the police in Israel: 1998–2007. Br J Criminol 50:748–771

Jonathan T, Weisburd D (2010) How do majority communities view the potential costs of policing terrorism?: Findings from a community survey in Israel. Polic J Policy Pract 4(2):169–181

Jonathan-Zamir T, Weisburd D (2013) The effects of security threats on antecedents of police legitimacy: findings from a quasi-experiment in Israel. J Res Crime Delinq 50(1):3–32

Mastrofski SD (2006) Community policing: a skeptical view. In: Weisburd D, Braga AA (eds) Police innovation: contrasting perspectives. Cambridge University Press, New York, pp 44–73

Mastrofski SD, Snipes JB, Supina AE (1996) Compliance on demand: the public's response to specific police requests. J Res Crime Delinq 33:269–305

Mueller JE (1970) Presidential popularity from Truman to Johnson. Am Polit Sci Rev 64:18–34

Murray J (2005) Policing terrorism: a threat to community policing or just a shift in priorities? Police Pract Res 6(4):347–361

National Research Council (2004) Fairness and effectiveness in policing: the evidence. Committee to review research on police policy and practices. In: Skogan W, Frydl K (eds) Committee on law and justice, division of behavioral and social sciences and education. National Academies Press, Washington, DC

Reisig MD, Bratton J, Gertz M (2007) The construct validity and refinement of process-based policing measures. Crim Justice Behav 34:1005–1028

Shaw GM, Brannan KE (2009) The polls – trends: confidence in law enforcement. Public Opin Q 73(1):199–220

Stein A (1976) Conflict and cohesion: a review of the literature. J Confl Resolut 20(1):143–172

Sunshine J, Tyler TR (2003) The role of procedural justice and legitimacy in shaping public support for policing. Law Soc Rev 37:513–548

Thacher D (2005) The local role in homeland security. Law Soc Rev 39(3):635–676

Tyler TR (2004) Enhancing police legitimacy. The annals of the American academy of political and social science, May, 593–608

Tyler TR (2009) Legitimacy and criminal justice: the benefits of self-regulation. Ohio State J Criminal Law 7:307–359

Tyler TR, Fagan J (2008) Legitimacy and cooperation: why do people help the police fight crime in their communities? (Paper No. 06-99). Columbia Law School, Public Law & Legal Theory Working Paper Group, New York

Tyler TR, Wakslak CJ (2004) Profiling and police legitimacy: procedural justice, attributions of motive and acceptance of police authority. Criminology 18:267–274

Tyler TR, Schulhofer S, Huq A (2010) Legitimacy and deterrence effects in counterterrorism policing: a study of Muslim Americans. Law Soc Rev 44(2):365–401

Weber M (1968) Economy and society. Roth G, Wittich C (Eds) New York: Bedminster

Weisburd D, Jonathan T, Perry S (2009a) The Israeli model for policing terrorism: goals, strategies and open questions. Crim Justice Behav 36(12):1259–1278

Weisburd D, Feucht T, Hakimi I, Mock LF, Perry S (eds) (2009b) To protect and to serve: policing in an age of terrorism. Springer, New York

Weisburd D, Hasisi B, Jonathan T, Aviv G (2010) Terrorism threats and police performance: a study of Israeli communities. Br J Criminol 50(4):725–747

Policing the Police

► Control of Police Misconduct

Policy Exchange

► International Policy Transfer

Policy Transfer

► International Policy Transfer

Political Action

► Social Disorganization and Terrorism

Political Crime

Jeffrey Ian Ross
School of Criminal Justice, College of Public Affairs, University of Baltimore, Baltimore, MD, USA

Overview

Political crime is often ignored or given minimal attention by criminologists especially in the traditional American criminology/criminal justice university curriculum. It is frequently and easily confused with issues dealing with how politics affects the passage of criminal laws or how politics shapes public policy in matters of criminal justice. Unfortunately, this interpretation is unnecessarily naïve, narrow, and safe with respect to understanding the relationship between politics and crime and criminal justice. In order to introduce the reader to the corpus of research on political crime, this entry contextualizes English language scholarly (and where appropriate popular) literature bearing on the subject of political crime looking at the categories that have been published. Despite their occasional utility, the review ignores the copious government reports that have been published in the form of senate or congressional inquiries, Royal Commissions, public prosecutions, and the like. This summary also does not include the numerous reports produced by activist organizations, professional groups, and consulting organizations that make their way into the public domain.

Definitional and Conceptual Issues

Before reviewing the concept of political crime, it is wise to settle on a relatively respectable definition of political crime. Although a handful of recognized scholars have proffered definitions of political crime, one of the best was provided by Beirne and Messerschmidt (1991, p. 240).

> . . . crimes against the state (violations of law for the purpose of modifying or changing social conditions). . .[and]crimes by the state, both domestic (violations of law and unethical acts by state officials and agencies whose victimization occurs inside [a particular country]) and international (violations of domestic and international law by state officials and agencies whose victimization occurs outside the US).

If anything remains true about the study of political crime and the production of scholarly knowledge on this subject, it is that the parts are infinitely greater than the whole. In other words, the majority of research on the phenomenon of political crime deals with its subcomponents (e.g., sedition, treason, state-corporate crime), rather than attempts to consolidate these disparate foci. Thus, we have numerous articles and chapters in scholarly books on espionage/spying and state-corporate crime, and book-length studies on treason and on political terrorism, but few attempts have integrated all political crimes. This observation is extended to political crimes' two major categories: oppositional political crime or state crimes.

Indeed, there are exceptions to this state of affairs. In terms of broad overviews that attempt to deal with the subject of political crime in a comprehensive manner, ten works stand out: Proal's *Political Crime* (1898/1973), Schafer's *The Political Criminal* (Schafer 1974), Turk's *Political Criminality* (1982), Kittrie and Wedlocks' *The tree of liberty* (1986), Ingraham's *Political Crime in Europe* (1979), Hagan's *Political Crime* (1997), Kittrie's *Rebels without a cause* (2000), Ross' *Dynamics of Political Crime* (2003), Head's *Crimes against the state* (2011), and Ross' *An Introduction to Political Crime* (2012). Although some of these treatments approach the subject of political crime with a rather narrow lens limiting their gaze to oppositional political crime, or a handful of prominent states (i.e., typically advanced industrialized democracies of one shape or form), others take a more broad brush approach.

Likewise, even though all of these books look specifically at oppositional political crime, all but a handful integrate the subject of state crime into their analyses. The disproportionate focus on oppositional political crime is predictable. Despite constitutional safeguards, especially the right to free speech engrained in most advanced industrialized democracies, many political activists, journalists, and scholars are downright timid about criticizing their government or assuming that the state can commit crimes against its own citizens.

As a reaction, since the early 1990s, some criminologists (typically critical criminologists) have produced books on state crime. Some of these contributions looked at the subject matter in the context of edited books with original chapters (e.g., Barak 1991; Ross 1995/2000, 2000; Chambliss et al. 2010). One of the earlier treatments was a collection of previously published journal articles (e.g., Friedrichs 1998a, b). More recently, books which have attempted to provide an integrated approach to the subject matter of state crime have been published (e.g., Rothe 2009).

Unfortunately, a considerable amount of scholarship in this field fails to build upon the work that has already been done. For example, much has been published in the English language legal journals on the subject of political crime but is frequently ignored by criminologists of political crime and vice versa because of unexplored and unexplained reasons, criminologists, sociologists of crime, and legal scholars rarely consult each others body of research. This scholarship is certainly relevant to the study of political crime, but it is not framed in the political crime context.

Assuming there is consensus on the previously mentioned definition of political crime and its subcomponents, then the following review attempts to look at the current state of research on this subject matter. With some notable exceptions (Alexander 1992a, b; Merton 1964; Moran 1974; Schafer 1974; Ingraham 1979; Turk 1992; Ross 2003, 2012), little of the research on political crime places the concept into a theoretical perspective.

Dominating the research on political crime are descriptive case studies, especially those focusing on political crime in advanced industrialized countries. The following reviews this work.

Case Studies

Occasionally articles and books about individuals charged and convicted of treason and sedition have been published. By far the greatest proportion of the literature (primarily written by journalists directed towards a popular audience) takes the form of case studies of individuals considered spies who have engaged in espionage. Numerous books covering both the activities of these unique political criminals operating in the United States (e.g., Ethel and Julius Rosenberg, Aldrich Aimes, Harry Gold, Robert Hanssen), Canada (e.g., Igor Gouzenko), and the United Kingdom (e.g., Anthony Blunt, Kim Philby) have been written, as have in-depth analyses of infamous political terrorists (e.g., Illich "Carlos the Jackal" Ramirez) and organizations such as law enforcement agencies, at both oppositional and state level, that have engaged in political crime.

In the United States, for example, many case studies examining the deviance, abuse of power, and/or crimes of the state criminogenic agencies have been produced. Federal agencies such as the Internal Revenue Service (IRS), the Federal Bureau of Investigation (FBI), the Central Intelligence Agency (CIA), and other prominent and selected national security/intelligence agencies/organizations have more generally been singled out for intense investigation.

In Canada, the United Kingdom, and Australia, we have a similar pattern with a smattering of articles and books focusing on intelligence/national security agencies and their misdeeds that have been published.

Country-Specific Studies

Whether because of access to data or simply an issue of differences and difficulties with understanding foreign languages, the majority of research on political crime focuses on its presence in the advanced industrialized countries. Countries that are the locus of political crime research are typically Australia, the United Kingdom, Canada, and the United States. This makes sense because looking at the country as an individual unit is more manageable.

United States

As with the larger study of political crime, there are few comprehensive studies (e.g., books) on the subject focusing on its occurrence in America. Two efforts stand out. The first is Roebuck and Weeber's *Political Crime in the United States* (1978). This book, though dated in content, provides a historical introduction to the subject matter. The other is Tunnell's, edited book, *Political Crime in contemporary America* (1993). This publication consisting of eight chapters looks at both theoretical issues, actions (e.g., terrorism, hate crime, worker safety), and controversial criminal justice policies and practices (e.g., death penalty) that have been unduly influenced by political concerns. Unlike Roebuck and Weeber, the collection emphasizes both oppositional and state crimes.

Other than these two books, there has been a disproportionate (but necessary) focus by scholars on most subcomponents of political crime. To begin with, a considerable amount of research has been conducted by reporters and scholars on the problem of oppositional political terrorism in the United States. In addition to a couple of comprehensive treatments, the work to date has covered domestic right-wing groups such as the Ku Klux Klan, Aryan Nations, and the Covenant, the Sword, and the Arm of the Lord. Books about the left-wing terrorist group Weather Underground have also been published, as have autobiographies by members of the organization.

Although a handful of broad-based historical treatments of state crime in the United States (though not labeled as such at the time of their writing) were produced during the 1970s, far more studies have been conducted on the subcomponents of these controversial government actions. Perhaps because of the dominance of the right to free speech, there have been numerous analyses of censorship.

Though not typically framed in the context of political or state crime, a considerable amount of literature exists on police activities of a questionable nature. Some of this research has focused on what some scholars call political policing (i.e., the use of illegal domestic surveillance and repression) (e.g., Theoharis 1978), the use of excessive force, police use of deadly force, and police corruption.

A focus on corruption both in federal government agencies, at the executive level (e.g., president), especially in connection with particular incidents (e.g., Watergate Affair, Iran-Contra) and with municipal police departments, has been dominant in the American literature. A handful of well-researched books, written by respected journalists, have told the stories of individuals accused of and/or caught up in corruption probes and/or periods of police corruption in major law enforcement agencies. Some of these incidents and studies have been the basis for movies produced by Hollywood studios for consumption by the general public.

Given the high levels of criminal violence that has historically plagued the United States, it is no surprise that political violence has been a subject of great concern. In addition to studies of political assassins and assassination, issues of oppositional political terrorism have been a staple of the political crime research in the United States.

The focus on the subcomponents of political crime is as true with oppositional crime as it is in state crime research. Much of the research focuses on case studies. The largest component has been the case studies of state-corporate crime (e.g., Michalowski and Kramer 2006), covering incidents such as the Ford Pinto design problems (1977), the Challenger space shuttle explosion (1986), the fire at the Imperial Chicken plant in Hamlet North Carolina (1991), nuclear weapons production, the crash of ValuJet 592 (1996), and the Bridgestone and Firestone tire controversy (2010).

Case studies and analyses of deviance and corruption in our correctional facilities, including the treatment of political prisoners and the use of excessive force in American jails and prisons, have also been important. Many of these studies have been idiographic/illustrative and not comprehensive in nature. Not only have jails and prisons been singled out as places where state crime has occurred, but in the past decade, America's participation in Abu Ghraib and Guantanamo has been selected for intense analysis.

Periodically the issues of slavery, racism, and genocide have been used as conceptual categories in which to analyze state treatment of its minorities, typically African Americans and American Indians.

United Kingdom

The problem of controversial police activities, especially corruption and use of excessive force, has been staples of contemporary British political crime research. Some of this has focused on the general topic of the policing of industrial disputes, while other scholarship has concentrated more specifically on the miners' strike (1984) and police involvement and reactions to race-related riots during the 1980s.

The problem of state crime in the United Kingdom has been dealt with in the context of illegal actions by state intelligence agencies (e.g., Thurlow 1994). One notable full-length treatment is Doig (2011). More prominent, however, are case studies of state-corporate crime (Tombs and Whyte 2003).

Canada

Unlike the United States, with the exception of a couple of chapters appearing in edited books, there are no book-length studies of political crime in Canada or oppositional political crime or state crime in this country. This is not because it does not exist, but more so because of the supply and demand kinds of dynamics in the country. Four areas seem to be the focus on research on political crime in Canada. Like the United States, Canadian legal scholars have examined issues connected to freedom of expression, censorship, and emergency legislation.

Also prominent concerns the issue of political corruption at the provincial and federal level. Both articles and book-length treatments of the subject have been done. The other area has been police abuse in terms of violence/excessive force and/or illegal domestic surveillance particularly looking into the Royal Canadian Mounted Police (RCMP). Other research has examined the frequency of political terrorism through the construction and analysis of events data bases. Some nascent literature looks at state crime in Canada (Corrado and Davies 2000).

Australia

Although no particular study focuses solely on political crime in Australia, Head's book uses many examples from history of political crime. As a corollary, the Grabosky's book *Wayward Governance* (1989) is a collection of case studies of well-known acts of state crime in Australia. He also provides a review of the literature on state crime and some initial thoughts on a theory of state crime and its control. Far more numerous are the case studies of state crime appearing as articles in journals or chapters in scholarly books. For example, because of its large aboriginal population, studies have been conducted on the fate of aboriginal deaths behind bars (e.g., Hazelhurst 1991).

Political Crime in Nonwestern Countries

By far the greatest shortcoming in terms of political crime scholarship has that which has occurred in nonwestern countries. This may be attributable to a number of factors including that studies may exist, but their accessibility to western researchers is poor and the fact that they are not written in English may preclude their integration into the larger body of work in English on political crime. One of the best collections of scholarship including nonwestern countries is C.E. S. Frank's *Dissent and the State* (1989). Although outdated, three chapters reviewed dissent in Eastern Europe (e.g., Braun 1989) and another on Latin America (Facher and Fitzgibbons 1989). By far the literature on political crime that has been available in English has disproportionately been in the field of political terrorism where we have analyses of terrorists, terrorists' organizations, and the causes and responses to this kind of political crime (Ross 2012).

Otherwise, linguist and political activist Noam Chomsky has produced a considerable number of books that look at state crime in countries and regions such as Indonesia and the Middle East. This has been done in the context of a critique of the questionable aspects of American foreign policy.

Again, case studies of numerous state crimes have been conducted by scholars and journalists, ranging from massacres to incidents of corruption to state-corporate crime.

Cross-National Studies

With the exception of the previously mentioned classic studies, few broad cross-national studies of political crime, whether oppositional or state, have been done. Attempts to provide this sort of thing have been done by authors of edited books (Barak 1991; Ross 1995/2000, 2000; Chambliss et al. 2010) in the concluding chapters of their books. Cross-national studies of a subcomponent of political crime, such as corruption, human rights violations, and oppositional political terrorism, are more common.

Explanations and Theories

With a subject that is as conceptually diffuse as political crime, it is not surprising that few experts have attempted to develop a holistic theory that explains the causal dynamics of this type of criminal/political behavior. In many respects, understanding how these actions emerge, who is involved, and why they take part is elusive. Needless to say, four attempts can be singled out. First, Merton (1938, 1964, 1966) provided one of the earliest explanations that, in part, touches on political crime. According to his anomic theory of deviance (i.e., strain theory), individuals live in societies that have a considerable amount of "structural dysfunctionalism." This, in turn, leads people to experience an ends/means discrepancy. These processes combined together create stress. In order to minimize the discomfort, individuals have five options, one of which is rebellion (nominally a type of political crime).

Second, Moran (1974) describes "sequential stages which in successive combination might account for the development of a political criminal" (p. 73). The first are what he calls "predisposing conditions or background factors, the conjunction of which forms a pool of potential political criminals. These conditions exist prior to an individual's decision to commit a political crime and by themselves do not account for his behavior" (pp. 73–74). The aforementioned conditions include the concept of strain and "a political problem solving perspective." The latter consists of "situational contingencies which lead to the commission of political crimes by predisposed individuals" (p. 74). Moran advocates a five-stage "developmental model" consisting of the following steps: (a) strain, (b) "political problem solving perspective," (c) a turning point event, (d) commitment to act, and (e) engaging in the political crime.

Third, Turk (1982) posits that although power and inequality are important factors in explaining political crime, the cultural gap between offenders and authorities is the primary factor that leads to the commission of political crime.

Merton's, Moran's, and Turk's theories are useful in describing, and in some cases explaining, various types of political crime, but they are not very helpful in accounting for all types of this phenomenon. The dynamic nature of such activities needs to be more thoroughly explored, and furthermore, the macro- and microlevel processes in political crime should be linked.

Thus, (Ross 2003, 2012), anchoring his explanation in Sutherlands' Differential Association Theory and Conflict Theory, developed the ISOR explanation. He argues that although many political crimes are committed by groups that are formally or loosely structured, whether oppositional in nature or formed by state organizations, these activities are, in the final analysis, committed by individuals. These people are working within the structural confines of both informal and complex organizations, political systems, political economies, and different cultures. They make decisions and act while often denying that any wrongdoing has occurred. In sum, according to Ross, political crime is the result of a complex interplay among individuals (I), situations/opportunities (S), organizations (O), and resource adequacy (R). These combined factors have been identified as the ISOR explanation. Ross argues that the ISOR relationship is not simply a resource mobilization theory or a crime prevention method brought into alignment with environmental design theories, but a more comprehensive explanation than previous attempts.

Research Methods

Similar to the previous comment made about theories of political crime, there is a shortage of diverse methods that criminologists (or other interested social scientists) investigating political crime have utilized. With the exception of a small but growing subset of studies of oppositional political terrorism using increasingly sophisticated quantitative techniques, the majority of research on political crime has used qualitative research methods, particularly single case studies. This is as true for journalistic accounts as it is those done by seasoned scholars.

Numerous reasons exist for this state of affairs. Political crime is complex and often requires cross-disciplinary knowledge. Additionally, no single database on political crime exists, nor will one be created in the near future. Thus, opportunities to test hypotheses, by mining an appropriate database are nonexistent. Moreover, the types of actions of political crime are so diverse and no single agency or organization collects this kind of information, so this presents numerous collection problems. Additionally, either scholars are timid about submitting proposals for funding to funding agencies or the agencies have not been enamored with the proposals they have received to agree to fund studies of political crime. Thus, what we have is a smattering of research studies that tend to be descriptive in nature with a shortage of empirical research methods.

Conclusion

Undoubtedly, this review has been limited by studies that have been conducted in English. It would be helpful for researchers to start moving to more integrative studies, beyond case studies of individuals and organizations and trying at the very least to do some cross-national studies. Similar critiques have been made by a collection of individuals who write in the field of state crime (Rothe et al. 2009). The production of this scholarship should not be interpreted in such a manner as to suggest that one country has more or less political crime than the other. Much of the choice of what gets researched, what gets submitted to journals, and what gets published is connected to the personal predilections of researchers, journal editors, and acquisitions editors with respect to what they think is interesting, their perceptions about the audience for their ideas, and the sophistication of the product submitted.

Related Entries

▶ Crimes of the Powerful
▶ Nuclear Weapons and State Crime
▶ Resistance to State Crime
▶ State Crime
▶ State-Corporate Crime

Recommended Reading and References

Alexander DW (1992a) Applying Merton’s theory of anomie to political criminality. Paper presented to the American Society of Criminology Annual Meetings, New Orleans

Alexander DW (1992b) Political crime: an application of Merton’s theory of social structure and anomie. Master’s thesis. Virginia Polytechnic Institute and State University

Barak G (ed) (1991) Crimes by the capitalist state: an introduction to state criminality. State University of New York Press, Albany

Beirne P, Messerschmidt J (1991) Criminology. Harcourt Brace Jovanovitch, New York

Braun A (1989) Dissent and the state in Eastern Europe. In: Franks CES (ed) Dissent and the state. Oxford University Press, Toronto, pp 111–137

Chambliss WJ, Michalowski R, Kramer RC (eds) (2010) State crime in the global age. Willan Publishing, Cullompton, Devon

Chomsky N (1983) The fateful triangle. South End, Boston

Corrado RR, Davies G (2000) Controlling state crime in Canada. In: Ross JI (ed) Varieties of state crime and its control. Criminal Justice Press, Monsey, pp 59–88

Doig A (2011) State crime. Willan Publishing, Abingdon

Franks CES (1989) Dissent and the state. Oxford University Press, Toronto

Facher P, Fitzgibbons K (1989) Dissent and the state in Latin America. In: Franks CES (ed) Dissent and the state. Oxford University Press, Don Mills, pp 138–168

Friedrichs DO (1998a) State crime, vol 1. Ashgate, Aldershot

Friedrichs DO (1998b) State crime, vol 2. Ashgate, Aldershot
Grabosky PN (1989) Wayward governance: illegality and its control in the public sector. Australian Institute of Criminology, Canberra
Hagan F (1997) Political crime: ideology and criminality. Allyn & Bacon, Boston
Hazelhurst KM (1991) Passion and policy: aboriginal deaths in custody in Australia, 1980-1989. In: Barak G (ed) Crimes by the capitalist state. State University of New York Press, Albany, pp 21–48
Head M (2011) Crimes against the state. Ashgate, Farnham
Ingraham BL (1979) Political crime in Europe: a comparative study of France, Germany, and England. University of California Press, Berkeley
Kittrie NN (2000) Rebels with a cause. Westview, Boulder
Kittrie NN, Wedlock ED Jr (1986) The tree of liberty: a documentary history of rebellion and political crime in America. Johns Hopkins University Press, Baltimore
McCulloch J, Tham J-C (2005) Secret state, transparent subject: the Australian security intelligence organisation in the age of terror. Aust NZ J Criminol 38(3):400–415
Merton RK (1938) Social structure and anomie. Am Sociol Rev 3:672–682
Merton RK (1964) Anomie, anomia, and social interaction: contexts of deviant behavior. In: Clinard MB (ed) Anomie and deviant behavior. Free Press, New York, pp 397–408
Merton RK (1966) Social problems and sociological theory. In: Merton RK, Nisbet RA (eds) Contemporary social problems. Harcourt Brace, New York, pp 775–823
Michalowski RJ, Kramer RC (eds) (2006) State-corporate crime: wrongdoing at the intersection of business and government. Rutgers University Press, Piscataway
Moran R (1974) Political crime. Doctoral dissertation, University of Pennsylvania
Proal L (1898/1973) Political crime. Paterson Smith, Montclair
Roebuck J, Weeber SC (1978) Political crime in the United States. Praeger, New York
Ross JI (ed) (1995/2000) Controlling state crime, 2nd edn. Transaction Books, New Brunswick
Ross JI (ed) (2000) Varieties of state crime and its control. Criminal Justice Press, Monsey
Ross JI (2003) The dynamics of political crime. Sage Publications, Thousand Oaks
Ross JI (2012) An introduction to political crime. The Policy Press, Bristol
Rothe DL (2009) State criminality: the crime of all crimes. Lexington Books, Lanham
Rothe DL, Ross JI, Mullins CW, Friedrichs D, Michalowski R, Barak G, Kauzlarich D, Kramer RC (2009) That was then, this is now, what about tomorrow? Future directions in state crime studies. Critic Criminol Int J 17(1):3–13
Schafer S (1974) The political criminal. Free Press, New York
Theoharis A (1978) Spying on Americans: political surveillance from Hoover to the Huston Plan. Temple University Press, Philadelphia
Thurlow R (1994) The secret state: British internal security in the twentieth century. Blackwell, Cambridge
Tombs S, Whyte D (2003) (eds) Unmasking the crimes of the powerful: Scrutinizing states and corporations. Peter Lang, New York
Tunnell KD (ed) (1993) Political crime in contemporary America: a critical approach. Garland, New York
Turk AT (1982) Political criminality. Sage Publications, Beverly Hills

Political Crimes

▶ Restorative Justice and State Crime

Political Violence

▶ Homegrown Terrorism in the United States

Polyvictimization

▶ Repeat Victimization

Population Size Estimation

▶ Capture Recapture to Estimate Criminal Populations

Porn

▶ Pornography

Pornography

Timbre Wulf-Ludden and Abby Vandenberg
School of Criminology and Criminal Justice,
University of Nebraska at Omaha, Omaha,
NE, USA

Synonyms

Internet pornography; Obscenity; Porn; Violent pornography

Overview

This entry discusses pornography and its importance in a variety of areas. First, a definition of pornography is discussed, followed by an examination of obscenity; the legal guidelines intended to differentiate between legally restricted, obscene materials; and pornographic materials that are protected under the First Amendment's protection of free speech. Conflicting feminist perspectives are also discussed. These perspectives largely focus on the question: Is pornography inherently harmful to women, or can it have a positive, potentially empowering influence on women, as well as other pornography viewers? Finally, the role of violence in pornographic material is discussed in depth followed by a brief note on Internet pornography.

Introduction

The production and sale of pornography is a highly lucrative industry in American society today (Eberstadt 2009). Due to modern technology, especially the Internet, pornographic material is available to a wide variety of individuals (Eberstadt 2009). In fact, pornography is the "single most searched-for item on the Internet," making it a pervasive part of American society (Eberstadt 2009, p. 4). The following sections discuss the features that differentiate pornographic materials from obscene materials, anti-pornography feminist perspectives, pro-pornography feminist perspectives, the relationship between violence and pornography, and the influence of the Internet on the development and dissemination of pornography.

What Is Pornography?

Pornography can be defined, at the most basic level, as explicit sexual material (Lindgren 1993). This definition, however, is overly broad. Such a simple definition would encompass a range of materials from classic Roman art to hard-core pornographic rape films. As such, this type of definition of pornography leaves many questions unanswered: Can a nude or sexual image be artistic, not pornographic? If so, under what circumstances does this occur? Is all pornography legal? When do the images presented in pornography or other obscene materials reach a level that is unacceptable to community standards?

In reality, pornography is a vague concept and one that is difficult to define. There is no consistent definition of pornography in existing empirical literature. Likewise, there is no universally accepted legal definition of pornography in the criminal justice system of the United States. Indeed, much pornographic material is not subject to regulation by criminal justice officials. Instead, much of the relevant literature and legal precedent related to the topic focuses on defining and regulating obscenity, not pornography.

Obscenity

While obscenity is a crucial concept in the discussion of pornography, pornography and obscenity are not synonymous (Lindgren 1993), as not all pornography is obscene. Thus, while obscene materials can be legally regulated, pornography, in general, is not considered illegal (Hoffman 1985). This is, in part, because the First Amendment guarantees the right to freedom of speech. The First Amendment guarantees not only the right to literally say things freely but it also guarantees the right to express one's self in a variety of ways. Although verbal speech, the printed word, art, and other forms of expression are protected by the First Amendment, there are limits to the types of expressions that may be afforded First Amendment protection. Specifically relevant to the discussion of obscenity and pornography is the United States Supreme Court's ruling in *Roth v. United States* (1957). In *Roth v. United States*, the Court ruled that obscene materials are a form of "unprotected speech." In other words, anything (e.g., literature, photographs, videos) that can be considered obscene is not protected under the First Amendment and is subject to legal regulation (Hoffman 1985). While the Court's ruling in *Roth* effectively outlawed obscene materials, a significant question was left unanswered: What materials can be considered obscene?

For more than a decade, the Supreme Court left this question unanswered. Finally, in *Miller v. California* (1973), the Court established specific standards that would be used to differentiate obscene material from constitutionally protected material. Three basic criteria are used to determine whether specific materials are obscene. The Court determined that a work could be considered obscene if:

> (a) . . . 'the average person, applying contemporary community standards' would find that the work, taken as a whole, appeals to the prurient interest,. . . (b) . . . the work depicts or describes, in a patently offensive way, sexual conduct specifically defined by the applicable state law, and (c) . . . the work, taken as a whole, lacks serious literary, artistic, political, or scientific value (Miller v. California 1973).

Under these guidelines, many forms of pornography are not considered obscene and are, thus, perfectly legal (Hoffman 1985). With regard to pornography, most "soft-core" less explicit materials (i.e., those that do not depict graphic sexual acts) are easily identifiable as constitutionally protected materials. On the other hand, extremely explicit "hard-core" pornography can be classified as legally obscene and subject to regulation (Hoffman 1985).

Even with the Court's three guidelines, however, it is difficult to definitively, and consistently, classify pornographic materials as being either obscene or legally acceptable. One of the biggest reasons for this difficulty is because the three guidelines provided in *Miller* v. *California* (1973) are inherently subjective. For example, with regard to the first criterion, contemporary community standards may vary from one place to another. Thus, what legally protected pornographic material in one state may be deemed obscene and, therefore, subject to regulation in another. Furthermore, images that may have artistic value to some may depict acts that offend state law in other areas. In sum, pornography is not a simple concept with an all encompassing definition. While obscene pornography is not subject to the protections of the First Amendment, there is a very fine line between pornographic materials that are obscene and those that are constitutionally protected.

Controversies in the Literature: Varying Feminist Perspectives on Pornography

Even beyond disagreements concerning the definition and identification of pornography versus obscenity, other controversies focus on the intrinsic value of pornography. Debate among feminist scholars concerning the potential harms (or benefits) created by pornography began to gain momentum in the 1970s and 1980s (Hoffman 1985). Many feminist scholars argued that pornography was inherently harmful to women because it encouraged the objectification and abuse of women and girls (Eaton 2007; MacKinnon 1989). Other feminist scholars argued that pornography is not always harmful to women. These researchers noted that pornographic materials can, at times, be positive and empowering (see, e.g., Rubin 1993). Despite decades of extensive discussion and empirical study into these issues, the debate surrounding the potential harms and benefits of pornographic materials has not been resolved (Eaton 2007). The following sections address several common arguments presented by scholars on either side of the ongoing debate.

Anti-pornography Perspectives

As mentioned above, some feminist scholars argue that pornography in inherently degrading and harmful to women (see, e.g., Eaton 2007;

MacKinnon 1989). These scholars have come to be identified as “antiporn feminists” (Eaton 2007). Antiporn feminists offer a variety of criticisms of pornography. One argument is that pornography is not focused on healthy sexual expression. Rather, it is a tool used to assert dominance and power (Hoffman 1985). More specifically, these feminists posit that pornography is used to assert males’ dominance and power over women (Hoffman 1985). MacKinnon (1989), for example, notes that pornography is a tool used to socially construct sexuality. She notes that pornography is primarily created for men and, as a result, illustrates the male ideal of sexuality (MacKinnon 1989). For example, she notes that pornography often depicts women as subordinate, less than human, and as creatures that exist to serve men and their sexual desires. In most cases, depictions of women in pornographic materials show them being violated and victimized, while men are shown as dominant and powerful (MacKinnon 1989). MacKinnon and other antiporn feminists seem to agree that

> pornography endorses [gender inequity] by representing women enjoying, benefiting from, and deserving acts that are objectifying, degrading, or even physically injurious and rendering these things libidinally appealing on a visceral level. And, as any advertiser will tell you, making something sexy is among the most effective means of endorsement (Eaton 2007, p. 682).

In sum, the perspective of antiporn feminists holds that pornography reinforces existing patriarchal societal values; pornography depicts male dominance as accepted and even desired by both women and men (Hoffman 1985; MacKinnon 1989). In this way, pornography is harmful to women because the patriarchal values depicted therein perpetuate assumptions that women are somehow less than men. Because women in pornography are often portrayed as enjoying themselves, pornography sends a message that women like to be used and objectified and that they welcome male power and control (Hoffman 1985). Furthermore, this perspective posits that pornography supports assumptions that women are supposed to be victimized and that female subordination is depicted as natural, normal, and even pleasurable (MacKinnon 1989). Such depictions undermine efforts to obtain equal power and equal rights for women and men.

Since their development in the 1980s, anti-pornography feminist perspectives have continued to evolve into the present day. For example, Eaton (2007) attempted to summarize and revise past antiporn feminist arguments into a “reasonable,” modern, antiporn perspective. Eaton (2007) acknowledged that some pornography may not be harmful in the way earlier scholars believed, such as being inherently violent or explicitly abusive toward women. Rather, Eaton (2007) focuses her attention on what she calls “inegalitarian pornography.” Inegalitarian pornography can be defined as “sexually explicit representations that, as a whole, eroticize relations (acts, scenarios, or postures) characterized by gender inequity (676).” Eaton (2007) argues, therefore, that while the effects of this type of pornography are different from those put forth by early antiporn feminist scholars, they are still damaging to women, though in more subtle ways. To elaborate, inegalitarian pornography may still reinforce inequitable gender stereotypes and promote the power and dominance of men over women (Eaton 2007). Eaton (2007) acknowledges that there are other factors that contribute to societal level gender inequality, but she maintains that inegalitarian pornography makes an important contribution. She points out that inegalitarian pornography may play a part in a dangerous “feedback loop.” Indeed, the images, attitudes, and actions depicted in inegalitarian pornographic materials may be rooted in existing gender inequality, but such pornography also promotes and contributes to the continuance of gender disparities (Eaton 2007).

Pro-pornography Perspectives

In contrast to the antiporn feminists, another body of feminist scholars argues that pornography, in and of itself, is not problematic. For example, while Rubin (1993) acknowledges that anti-pornography feminists identify important concerns about sexist attitudes and behaviors toward women, she argues that pornography is a “scapegoat.” In other words, Rubin (1993) posits that pornography is identified among antiporn feminists as causing problems it does not ultimately create. Specifically, she contests the perspective of anti-pornography feminist scholars who argue most pornography is violent. There are a majority of pornographic materials that, while sexually explicit, do not depict any acts of violence (Rubin 1993). To elaborate, the vast majority of pornography depicts common sexual activities more often than any form of "kinky" or violent sexual acts (Rubin 1993). Indeed, only specialized subgenres of pornography (e.g., sadomasochistic materials) depict violent acts. Furthermore, it is important to note that these subgenres make up a clear minority of all pornographic material (Rubin 1993). Rubin (1993) further notes that even in graphically violent pornographic materials, such as in depictions of sadomasochism, one should not assume that the individuals are victimized:

> Since SM often involves an appearance of coercion, it is especially easy to presume that the people doing it are victims. However . . . this is a false stereotype and does not reflect social and sexual reality. Sadomasochism is part of the erotic repertoire, and many people are not only willing but eager participants in SM activity (Rubin 1993, p. 248).

Rubin (1993) argues that pornography is no more violent than any other form of mass media and that individuals are much more likely to see the promotion of patriarchal values and violence against women (including rape and other sexual assaults) in mainstream television and movies than in pornographic material. Rubin (1993) concludes that it is the anti-pornography activists, themselves, who perpetuate harmful stereotypes. The assumption that pornography is inherently wrong breeds contempt for sex workers and reinforces assumptions that sex work is not legitimate or acceptable (Rubin 1993). Therefore, Rubin (1993) argues that feminist scholars should work to legitimize sex work and empower sex workers.

Other feminist scholars make arguments similar to Rubin's (1993) regarding the positive effects and female empowerment that pornography can provide, when utilized properly. For example, Sherman (1995) argues that pornography, as a whole, cannot automatically be dismissed as degrading and/or obscene. He argues that some types of pornography can have a positive and empowering influence for minority groups. Specifically, Sherman (1995) argues that homophobia is rampant in our society. He believes that both young gay men, as well as the society at large, are led to believe that homosexual encounters are dirty, debasing, and undesirable (Sherman 1995). His study of pornography created for gay male audiences, however, revealed that pornography can be a significant source of information about sex for gay males. Indeed, rather than reinforcing the homophobic notions prevalent in society, gay pornography seemed to demonstrate that sexual interactions between two gay men can be positive, healthy, and expressive (Sherman 1995). Thus, in a society steeped in homophobic values and images, pornography may offer the only positive depictions of gay sexual activity that a young gay male may have to rely on (Sherman 1995). Such positive imagery allows gay youth to learn about and accept their sexuality in a healthy, safe way (Sherman 1995). By lumping all pornography together and arguing that any form of pornography is harmful and degrading, anti-pornography feminists unwittingly affirm homophobic messages (Sherman 1995). In conclusion, Sherman (1995) argues that it is important that scholars not assume that all pornography is necessarily harmful. Instead, he argues that they must remain open to the possibility that pornography can be positive and empowering.

Pornography and Violence

Central to the arguments of both antiporn feminists and pro-porn feminists is the concept of harm. It is clear that there is a considerable debate regarding whether pornography is damaging to women. The most blatant form of harm that pornography may produce is actual physical and/or sexual violence against women. A number of scholars have examined whether pornography contributes to violence against women.

Pornography Has a "Cathartic" Effect

Although one of the most controversial debates over the last few decades has been the relationship between pornography and violence, the research findings regarding the link between pornography and aggression are conflicting, at best. Some scholars have posited that merely viewing pornography does not influence the attitudes that male and female research participants have about the acceptability of violence against women. One of the main arguments put forth by those who have found that pornography does not increase negative attitudes or behavior toward women is that pornography has a "cathartic effect" on viewers (see, e.g., Ben-Veniste 1971). The argument made by these scholars is that because pornography provides individuals with an outlet for sexual release, they are less likely to engage in acts of coercion or sexual violence in order to achieve sexual gratification (Ben-Veniste 1971). Thus, according to this perspective, an individual's exposure to pornography does not increase his or her likelihood to engage in acts of violence or overt aggression; rather, it decreases the probability of such behavior (see, e.g., Kutchinsky 1971). Empirical support for this contention has been found in a number of studies.

Two of the most important works relating to the "cathartic effect" of pornography, however, were conducted by Ben-Veniste (1971) and Kutchinsky (1971). The results from these analyses indicated that there was an overall decrease in the amount of sex crimes in Denmark after the government repealed laws that restricted the sale of pornography. Future studies, however, revealed that crime rates in Denmark only dropped for offenses such as voyeurism and exhibitionism, while rape and other sexual crime rates remained constant.

Pornography Enhances Negative Attitudes Toward Women

Because of these inconsistent findings, it is important to consider the equal number of researchers who have reported that exposure to pornography does make male viewers more likely to have negative attitudes toward women. These scholars are more likely to adhere to a social learning perspective, which posits that exposure to pornographic materials influences individual attitudes regarding appropriate sexual behavior as well as stereotypical male and female traits (Bergen and Bogle 2000). For example, it is no secret that women are frequently objectified in pornography. The images presented in pornographic materials focus explicitly on the woman's sexual organs and leave viewers with the impression that she is no more than an object to be used in order for the male to obtain sexual gratification. After repeated exposure to pornography, viewers may come to believe that women are merely sexual objects and, consequently, begin to treat them as such (Bergen and Bogle 2000; Flood and Pease 2009). Therefore, the overall "findings of this research support the idea that there is a relationship between pornography and violent against women" (Bergen and Bogle 2000, p. 232).

Pornography Increases Beliefs in Rape Myths

One of the most significant influences that pornography has is on attitudes is one's willingness to accept "rape myths" (e.g., women instigate or deserve their rape experiences because of the way they acted or the manner in which they were dressed, most women have a subconscious desire to be raped or engage in violent sexual activities) (see, e.g., Flood and Pease 2009; Russell 1988). Scholars have noted that men who watch pornography, especially if it features some element of sexual violence, are much more likely to believe rape myths than those who do not use pornographic materials. Mock trial studies have also presented interesting information regarding the treatment of women by those who view pornography (see, e.g., Russell 1988). These experiments generally expose one group of participants to sexually explicit films and one group to neutral films (such as psychology or nature videos), then present them with testimony or transcripts from actual rape cases. These studies consistently show that those individuals who were exposed to pornography rated the female victim as less believable, more responsible for her actions, and more worthless as a human being than did the individuals in the control group (Russell 1988). Furthermore, male acceptance of violence against women was shown to increase significantly when the woman was portrayed as unwilling to participate in the sex act, but was subsequently shown to be enjoying the experience (Weisz and Earls 1995).

Attitudes Are Linked to Aggression: Pornography Develops Aggressive Attitudes

It is important to consider the attitudinal changes that develop because of one's exposure to pornography because one's attitudes are linked to his or her future actions. For example, a number of experimental studies have shown that exposure to pornography increases aggression, and attitudes that are favorable to aggression, in males (see, e.g., Flood and Pease 2009). Over time, or after prolonged exposure to the messages presented in pornographic materials, individuals will eventually attempt to recreate the sexual acts they have seen. This is problematic when the acts they attempt to reenact include violence and/or the submission of an unwilling partner. Over the past few decades, research findings indicate that a significant number of males have reported that they either wanted, or had tried, to imitate acts that they saw portrayed in pornographic materials (for further discussion, see Russell 1988). Other studies have consistently revealed that between one-third and one-half of male research subjects admit that they would attempt to rape a female if they knew there was no chance of getting caught (Bergen and Bogle 2000). These findings are especially concerning because these attitudes are considerably enhanced in males exposed to sexually violent pornography (Flood and Pease 2009).

Pornography with Sexual Violence Is Particularly Damaging to Attitudes

Researchers studying sexual violence in pornography research have generally defined it as the overt use of physical force or psychological coercion to obtain sexual gratification; any nonconsensual sexual contact; the degradation, dehumanization, or objectification of one or more participants; and/or any form of copulation that is accompanied by acts of physical aggression or violence, such as whipping, punching, and/or bondage (Monk-Turner & Prucell 1999). Although scholars disagree about the prevalence of violence in pornography, they have consistently found that rape is the most common form of sexual violence depicted in these materials and that violent pornography creates more negative attitudes of, and behavior toward, women than nonviolent pornography. This is especially true among men who are high-frequency users of pornography. Studies have revealed that after viewing violent pornography, men who were not previously aroused by such images subsequently began fantasizing about, and becoming sexually aroused by, rape (see, e.g., Malamuth 1981).

With all of these studies, however, it is important to note that the most significant factor in pornography that shapes viewers' attitudes and beliefs is not the erotic nature of the content but the dehumanization and the degradation of women (Gray 1982; Russell 1988; Weisz and Earls 1995). In other words, when films contain acts of sexual violence, viewers' negative attitudes toward women are shaped more by the interpersonal violence that occurs in these films, rather than the erotic subject matter. A number of researchers have studied this phenomenon by exposing groups of viewers to films with neutral content, violent content, sexual content, and sexually violent content (see, e.g., Weisz and Earls 1995). These studies consistently find that the attitudes toward women were similar among those who viewed the violent content and those who viewed the sexually violent content. Furthermore, individuals in both of those groups had significantly more negative attitudes toward women than those who viewed either type of nonviolent film (Weisz and Earls 1995).

While these studies report that violent pornography increases aggression in most people, other studies have found that pornographic material containing sexual violence only increases aggression in individuals with higher predispositions toward aggressive behavior, or those who already feel anger toward a particular target (Gray 1982). This becomes dangerous when individuals are unable to distinguish whether it is the abuse and violence triggering their arousal or the sexual content of the pornography. Scholars have noted a high correlation between sexual fantasies based on deviant acts, such as rape, and subsequent deviant behaviors (Gray 1982). Not only, then, does sexually violent pornography reinforce preexisting ideas within individuals, but it may also provide them with new ideas (Russell 1988; Cramer et al. 1998). As noted above, many males admit that they have tried, or have been tempted to try, things that they have seen contained in pornographic materials (Russell 1988). However, this becomes problematic when their sexual desires include elements of unwanted sexual contact. Interviews with female rape and sexual assault victims revealed that the women who experienced the most frequent, and severe, cases of violence were those whose abusers asked or forced them to view or imitate scenes from pornographic material, to pose for pornographic pictures, or to create pornographic videos (see, e.g., Cramer et al. 1998). Therefore, it appears that in some contexts, pornography may teach individuals how to commit acts of sexual assault.

Pornography Influence on Youth

Although the link between pornography and violence is complex and multifaceted, there are noteworthy correlations that are hard to ignore. For example, a number of studies have focused on the impact of pornography exposure to youth. Some have found that a significant number of convicted rapists reported being exposed to more sexually explicit pornography before the age of 10, but reported less exposure to these materials than the average male during their adolescence and adulthood (Gray 1982). Furthermore, scholars have noted that in addition to less exposure to pornographic material, juvenile sex offenders also had parents who were more sexually restrictive. Goldstein and Kant (1974) posit that "pornography performs an educational function for men during their formative years; deprived of information about sex, rapists and pedophiles have few stimuli which portray society's definition of the 'normal sex act'" (as cited in Gray 1982, p. 392). Therefore, because sex offenders are largely denied access to pornography during puberty, they must rely on the pornography they viewed during their youth in order to form conceptions of "typical" adult sexual activity. However, because age influences one's comfort level with, interest in, and ability to fully understand material that is presented to them, these offenders will have inaccurate conceptions of "normal" sexual activity and, therefore, will be much more inclined to engage in deviant and/or violent sex acts (for further discussion, see Gray 1982).

Even though not all youth who view pornography will become sex offenders, pornographic materials still have an important influence on the development of individuals' later attitudes toward women and sex. There is evidence to suggest that frequent use of violent pornography is linked to high levels of sexual aggression, aggressive behavior, belief in rape myths, and fantasy about rape among adolescent males (see, e.g., Flood 2009). Not all youth who view pornography, however, become sex offenders, and there is evidence to suggest that a great number of youth are troubled when they encounter pornographic materials. Girls are much more likely to report feeling embarrassed or disgusted by pornography, while boys are more likely to become sexually aroused by these images (Flood 2009).

Although many would argue that it is a small minority of youth who are exposed to "actual" pornography, one must be aware that the same outcomes occur in studies of youth reactions to, and the attitudes that develop from, exposure to mainstream sexual media. As noted earlier, sexualized themes presented in advertising campaigns, television programs, and movies disseminate many of the same stereotypical and sexist messages as pornography, most strongly the idea that women are sexual objects (Flood 2009; Rubin 1993). These messages teach both girls and boys that a young woman's worth should be equated with her physical attractiveness which, in turn, influences how young people treat and respond to each other in their social interactions (Flood 2009).

Pornography Influence on Females

While most studies focus on the impact of pornography on male attitudes and beliefs, it is important to consider the effect that pornographic materials have on females. Overall, most studies find that, relative to women who are not exposed to pornography, females who view pornography are not more likely to develop negative attitudes toward other women, deepen their beliefs in rape myths, have negative body images, have lower self-esteem, or show more acceptance of violence against women (see, e.g., Linz 1989). In fact, most women were less accepting of rape myths and interpersonal violence after viewing sexually violent pornography than they were before (Linz 1989).

However, some studies have found that women who use pornography are not likely to engage in acts of physical sexual assault, but they may be just as likely as males to engage in acts of sexual coercion (see, e.g., Kernsmith and Kernsmith 2009). The reason for this may be because of the messages that pornography disseminates about "appropriate" actions and manners of behavior for women. Rarely are women in pornographic materials more physically aggressive than their male partner(s). More often, females use seduction or other manipulative behaviors to initiate sex. Therefore, women who are exposed to pornography may attempt to replicate these behaviors in order to achieve their own sexual conquests (Kernsmith and Kernsmith 2009).

Pornography Influence on Adult Sex Offenders

Studies have shown that pornography use is especially high among sex offenders (for more detailed discussions, see Gray 1982). In contrast to their rates of exposure to pornography during youth and adolescence, adult sex offenders report higher rates of pornography usage and masturbation than the general adult male population (Gray 1982; Marshall 1988). Approximately one-third of both rapists and child molesters report that they have been, on at least one occasion, prompted to commit a sexual offense based on what they observed in pornography (Gray 1982). Furthermore, nearly one-third of rapists and over half of child molesters noted that they specifically used pornography in preparation for their offense (see, e.g., Marshall 1988).

In sum, the scholarly research results concerning the link between pornography, aggression, and violence is unclear, at best. The findings presented in this section provide empirical support for the arguments made by both antiporn and pro-porn feminists. Some scholars have found that pornography had no influence on shaping attitudes toward women or one's likelihood to engage in violent behavior. Others have found that pornography does lead to negative attitudes toward women among men, women, and children, while still others report that pornography's influence on attitudes and actions depends on one's personal predispositions toward violence and male dominance. Therefore, it appears that this decades-long, multifaceted debate will not be resolved any time in the near future.

Current Issues: The Role of the Internet

Any discussion of the modern debates surrounding, however, pornography would be incomplete without some discussion of the Internet. Pornography has changed and thrived with the advent of the Internet. The overwhelming majority of Americans have an Internet-capable computer in their home or have Internet access via another source (e.g., a public computer, a school or work computer, a friend or family member's computer). The pornography industry has thrived on the Internet because it serves as a mechanism that provides more people than ever before with quick and easy access to pornographic materials (Eberstadt 2009). The ease of access to pornography and quick distribution to a wide demographic that the Internet provides has resulted in a greater social acceptance of pornography (Eberstadt 2009).

However, the Internet not only facilitates the distribution of traditional pornographic material, it also allows for the development or expansion of new types of pornography. For example, Paasonen (2010) discusses the Internet's facilitation of amateur pornography. While amateur pornography existed before the widespread use of the Internet, it was much less frequently consumed and distributed (Paasonen 2010). With the advent of the Internet, however, it was possible for nonprofessionals to not only create but also share their own pornographic material with a huge audience (Paasonen 2010). The Internet also made available alternative porn, or specific niches of porn available with great ease to a wide audience. Alternative porn was developed for specific subgroups (i.e., Goth or indie pornography) and depicts sexual lifestyles not often seen in traditional pornography (Paasonen 2010). In sum, the Internet not only made pornography available to a wider audience than ever before, but it also created an opportunity for the production and dissemination of new kinds of porn created by professionals and nonprofessionals alike. The Internet pornography industry will no doubt continue to develop and expand as technology further advances.

Conclusion

Pornography is a complex topic; its definition and legal standing continue to be debated. In addition, while scholars have long debated the merits of pornography, there is no consensus to date as to whether or not pornography is empowering, harmful, or even a catalyst for violent behavior. Such debate among scholars will no doubt continue as pornography continues to change and develop along with advancing technology and changing societal expectations.

Related Entries

- ▶ Child Pornography
- ▶ Feminist Theory in the Context of Sexual Violence

Recommended Reading and References

Bergen RK, Bogle KA (2000) Exploring the connection between pornography and sexual violence. Violence Vict 15(3):227–234

Ben-Veniste R (1971) Pornography and sex-crime: the Danish experience. Technical reports of the commission on obscenity and pornography, vol 8. Government Printing Office, Washington, DC

Cramer E, McFarlane J, Parker B, Soeken K, Silva C, Reel S (1998) Violent pornography and abuse of women: theory to practice. Violence Vict 13(4):319–332

Eaton AW (2007) A sensible antiporn feminism. Ethics 117(4):674–715

Eberstadt M (2009) Is pornography the new tobacco? Policy Rev 154:3–18

Flood M (2009) The harms of pornography exposure among children and young people. Child Abuse Rev 18:384–400

Flood M, Pease B (2009) Factors influencing attitudes to violence against women. Trauma Violence Abuse 10(2):125–142

Gray SH (1982) Exposure to pornography and aggression toward women: the case of the angry male. Soc Probl 29(4):387–398

Hoffman E (1985) Feminism, pornography, and law. Univ Pennsylvania Law Rev 133(2):497–534

Kernsmith PD, Kernsmith RM (2009) Female pornography use and sexual coercion perpetration. Deviant Behav 30:589–610

Kutchinsky B (1971) Towards and exploration of the decrease in registered sex crimes in Copenhagen. Technical reports of the commission on obscenity and pornography, vol 8. Government Printing Office, Washington, DC

Lindgren J (1993) Defining pornography. Univ Pennsylvania Law Rev 141(4):1153–1275

Linz D (1989) Exposure to sexually explicit materials and attitudes toward rape: a comparison of study results. J Sex Res 26(1):50–84

MacKinnon CA (1989) Sexuality, pornography, and method: "Pleasure under patriarchy". Ethics 99(2): 314–346

Malamuth N (1981) Rape proclivity among males. J Social Issues 37:138–157

Marshall WL (1988) The use of sexually explicit stimuli by rapists, child molesters, and nonoffenders. J Sex Res 25(2):267–288

Miller v. California, 413 U.S. 15 (1973)

Monk-Turner E, Prucell HC (1999) Sexual violence in pornography: how prevalent is it? Gender Issues 17(2):58–67

Paasonen S (2010) Labors of love: netporn, Web 2.0 and the meanings of amateurism. New Media Soc 12(8):1297–1312

Roth v. United States, 354 U.S. 476 (1957)

Rubin G (1993) Misguided, dangerous and wrong: an analysis of anti-pornography politics. In: Assiter A, Carol A (eds) Bad girls and dirty pictures: the challenge to reclaim feminism. Pluto Press, London, pp 244–251

Russell DEH (1988) Pornography and rape: a causal model. Polit Psychol 9(1):41–73

Sherman JG (1995) Love speech: the social utility of pornography. Stanford Law Rev 47(4):661–705

Weisz MG, Earls CM (1995) The effects of exposure to filmed sexual violence on attitudes toward rape. J Interpers Violence 10(1):71–84

Post-Conflict Justice

▶ Post-Conflict Traditional Justice

Post-Conflict Traditional Justice

Tim Allen[1] and Anna Macdonald[2]
[1]Department International development,
London School of Economics and Political
Science, London, UK
[2]Department of War Studies, King's College
London, London, UK

Synonyms

Customary justice; Post-conflict justice; Transitional justice

Overview

Since the mid-1990s, there has been a proliferation of attempts to adapt and institutionalize forms of traditional justice as part of post-conflict policy. This has occurred in places as diverse as Timor Leste and Sierra Leone, Rwanda, and Afghanistan. While anthropologists have long been interested in traditional justice, and have emphasized it in studies of rapid social change and post-conflict reconciliation, it is a relatively new arena for transitional justice. Indeed, the foregrounding of traditional justice as a possible alternative to new international mechanisms seems to have come as a bit of a surprise to those promoting truth commissions, criminal courts, and tribunals. There are certainly paradoxical aspects to it, given that the shift of interest towards local accountability mechanisms is occurring at the same time as international criminal law is expanding its reach. However, both trajectories may also be viewed as being part of the same process in that they seek forms of viable justice that are less directly connected with the formal authority of sovereign states – authority which may be very partial and compromised in politically fragile post-conflict circumstances.

The actual content of the traditional justice category is rather vague. Other adjectives such as customary, informal, community based, grass roots, indigenous, and local are all sometimes used interchangeably. To a large extent, it has become a catch-all designation to describe procedures in those places that other kinds of justice provision cannot reach and also as an explanation for why more formal judicial mechanisms introduced in post-conflict setting seem to have such limited effects. It has been explicitly linked to the promotion of more relevant and grounded transitional justice, although the desire for a holistic approach – one that strikes a balance between meaningful customary practices and universal principles – is essentially an aspiration whose applicability and efficacy has rarely been tested. Aid agencies, human rights activists, and local power brokers are finding in traditional justice ways of furthering their diverse agendas. Yet, despite some grand claims, in reality, we still know remarkably little about the role and impact of informal justice processes in post-conflict situations (Kelsall 2009; Huyse and Salter 2008; Shaw et al. 2010). As so often in discussions of justice, normative notions of what is inherently believed to be right shape perceptions, rather than evidence about what has been occurring. One consequence is a tendency to misleadingly generalize about traditional justice as if it is some sort of cohesive and homogenized alternative to formal systems. Much more rigorous, nuanced and systematic research is required. Highlighted below are certain characteristics of the way in which traditional justice is currently discussed within the transitional justice literature and among transitional justice practitioners, followed by comments on important controversies.

Fundamentals

To begin with, it is helpful to place interest in the local dynamics of justice in conflict and post-conflict situations in a broader context. The World Bank's World Development Report 2011 draws attention to the paucity of information about what is happening on the ground in war affected and politically fragile locations. Meanwhile, other World Bank publications have been drawing attention to the need to take better account of cultural and ethnic diversity when designing and implementing development programs, including practices like witchcraft, spirit possession, and spiritual healing (Marc 2010, p. 4). Such a perspective remains controversial within the World Bank and in other development organizations such as the UK's DFID. Nevertheless, it is evidently the case that throughout the global south, there are vast regions in which the power and authority of state law is "nominal rather than operational" (Falk Moore 1986, p. 150). In 2007, the Organisation for Economic Cooperation and Development noted that as many as 80 % of the people in today's fragile states rely on non-state actors for various forms of justice and security (OECD 2007). In Sierra Leone, it has been estimated that some 85 % of the population does not have access to formal justice and relies upon traditional measures (Sriram 2007, p. 598). In Afghanistan, in those areas not controlled by the Taliban, an estimated 80–90 % of all disputes are mediated in the customary system (Wojkowska 2006). Local justice tends to be more accessible to the poor, relatively quick and cheap, and, crucially, the arbiter of issues of great social and economic concern, namely, land and family/lineage issues. In Kenya, for example, where land is frequently a source of private and communal disputes, traditional institutions are widely held to be more reliable in resolving conflicts than the state (WDR 2011, p. 134). In fact, customary tenure is said to cover 75 % of land in most African countries, affecting 90 % of land transactions in Mozambique and Ghana (Wojkowska 2006, p. 12).

Traditional and indigenous processes are currently receiving ever more attention in both state building and counter-insurgency policy (Branch 2011), while putative traditional governance systems are being foregrounded in a manner that has not happened since the era of colonial indirect rule. It is believed that embedding orthodox peace building approaches in local culture will enhance their legitimacy and efficacy, thereby providing an authentic and familiar environment through which popular participation might begin to flourish (Branch 2011). As one UNDP report notes: "Existence of these systems cannot be overlooked. We need to develop strategies to take advantage of the benefits of informal systems" (Wojkowska 2006, p. 13). In a similar way, the World Bank World Development Report 2011 notes that supplementing formal justice with traditional community systems can be a "best-fit" but with a revealing caveat: "the lesson here appears to be to use a process of recognition and reform to draw on the capacities of traditional community structures and to 'pull' them gradually in the direction of respect for equity and international norms" (WDR 2011, p. 167). Selective support for traditional justice here provides a sort of indigenous anchor: a means by which the broader, donor-supported accountability agenda can be grounded, authenticated, and legitimized.

The interest in traditional justice is also linked to perceived limitations in the initial formulation of the transitional justice concept. Transitional justice emerged largely as a socio-legal policy response to the so-called "dirty" wars in South America and was associated with transitions from authoritarian and oppressive states to democratic states (Arthur 2009). These were "states characterized by relatively high levels of horizontal and vertical institutionalization" (De Grieff 2011, p. 1). To a considerable extent that model applied in the South African case too, but what about northern Uganda or Southern Sudan, or the Democratic Republic of Congo? In those places, it has not been clear that the violence has been *mostly* linked to formal government forces, nor is it clear that there is a transition from oppressive authority to democracy. Rather these are territories characterized by hybrid authority structures, and the prospect of achieving stable, accountable, and representative governance is remote. In 2005, a rather poignant confessional was delivered by David Crane, the former Prosecutor of the Special Court for Sierra Leone: "Our perspectives are off kilter... we consider our justice as the only justice... we don't create mechanisms by which we can consider the cultural and customary approaches to justice within the region" (cf. Kelsall 2009, p. 11). This shortcoming had been acknowledged a year earlier by the UN Secretary General, Kofi Annan, in a report to the Security Council titled "The rule of law and transitional justice in conflict and post-conflict societies," in which he observed that: "due regard must be given to indigenous and informal traditions for administering justice or settling disputes, to help them to continue their often vital role. . ." (UNSC 2004, p. 12). Support for traditional justice provides much needed diversity in each context, guarding against what scholars have disparagingly termed the "templatization" and standardization of transitional justice or what the UN Security Council refers to as "one-size-fits-all" solutions (UNSC 2004, p. 1).

It is striking that the emergence of traditional justice as an alternative within a framework of transitional justice has actually been invigorated by the creation of the International Criminal Court (ICC). While it is the case that traditional justice was held up as an alternative to other international instruments, notably the International Criminal Tribunal for Rwanda (ICTR) and the Special Court for Sierra Leone (SCSL), the ICC cannot escape engagement with it. The reason for this lies in the wording of the Rome Statute. Multiple references to the requirement of the court to act "in the interests of justice," without explanation of what that means, have enabled lobbying groups to demand serious consideration of alternative conceptions. Furthermore, the Rome Statute has allowed space for arguments to be made about traditional justice in relation to the requirement of the court to act in a way that is complementary with local procedures. The ICC itself has been on something of a learning curve in terms of how to handle such pressure. A reflection of this are ICC statements like the one that appeared in a 2003 ICC policy paper, which declares that the prosecutor "will take into consideration the need to respect the diversity of legal systems, traditions and cultures" (Allen 2006, p. 129).

Many activists and some scholars believe that traditional justice is not just an alternative or possible supplement to more established processes. Rather they take the view that it is better, or at least that a fully integrated approach is the best option, one in which conventional legal processes are not privileged, and "multiple pathways to justice" can be "interwoven, sequenced and accommodated" (Roht-Arriaza 2006, p. 8).

The view is premised on an acceptance that not only formal trials but also truth commissions are insufficiently attentive to social integration and reconstruction. The latter have often been portrayed as somehow more culturally embedded, but critics have argued that they can be equally remote from local realities. As Priscilla Hayner notes, "indigenous national characteristics may make truth-seeking unnecessary and undesirable, such as unofficial community based mechanisms that respond to recent violence or a culture that eschews confronting reality directly" (Hayner 2010, p. 186). From Peru to Cambodia to Sierra Leone, scholars have highlighted the danger of what one termed the "tyranny of total recall" (Theidon 2009). These societies, it is argued, are characterized in varying degrees, by social ideas of forgiveness. Rosalind shaw, meanwhile, has traced the genealogy of truth commissions and finds their genesis in a Western tradition of confession that has no immediate resonance in contexts such as Sierra Leone, where "a factually accurate depiction of the past is less important to reconciliation than achieving a cool heart" (Shaw 2007; Kelsall 2009, p. 14).

The problem for transitional justice scholars and practitioners then is that internationally sponsored judicial and non-judicial processes and decisions appear to be making little sense and garnering very limited support from the very constituencies they are supposed to be benefitting. The appeal of a more locally oriented justice, in contrast, is claimed to lie in its potential to repair and restore communal relationships via familiar, locally grounded processes that all community members can associate with (Alie 2008; Latigo 2008). Traditional justice is laudable, so the argument goes, because it is culturally relevant. It draws upon authentic indigenous identities and rituals and "taps into profound spiritual worlds" based on non-Western concepts of community harmony and well-being (Roht-Arriaza 2006, p. 12). It is also suggested that justice built on established customs of reconciliation and compensation is more appropriate and pragmatic in close-knit community settings, where people remain dependent on continuous social and economic relationships with their neighbors

(PRI 2002). Thus, James Otto, the head of Human Rights Focus in northern Uganda, graphically expressed opposition to the ICC's intervention by explaining that: "there is a balance in the community that cannot be found in the briefcase of the white man" (Allen 2006, p. 134).

To some extent too, support for traditional justice has been pragmatic. The vast scale of atrocity crimes in places where transitional justice currently operates makes it very hard to hold every suspected perpetrator accountable. The *gacaca* system that emerged in Rwanda in 2001 to deal with the aftermath of the 1994 genocide was partly a national response to this kind of logistical dilemma. At least 800,000 people had been killed during the violence and the country's jails were reaching bursting point with 120,000 alleged perpetrators and only 15 judges able to oversee their trials. The United Nations Security Council had set up the International Criminal Tribunal for Rwanda (ICTR) in Arusha, Tanzania, in November 1994, but there was widespread frustration with the court: it was seen as too slow, too expensive, and too far removed from Rwanda. The Rwandan government's response was to adopt and adapt a traditional community conflict resolution system, the *gacaca*, and to train more than 250,000 community members to serve on panels in 11,000 jurisdictions (Clark 2010, p. 3). As Phil Clark explains, the Gacaca Law was enacted with the aim of expediting justice for genocide crimes by relieving the national courts and the ICTR of the vast numbers of low-level suspects and allowing them to focus on the more senior accused. The gacaca system was also intended to pursue the broader reparative goals of social healing and reconciliation (pp. 63–64).

The focus on traditional justice has certainly gained momentum since the Gacaca courts were set up and there are now numerous programs aimed at supporting it, but this has not resulted in a formalized typology in any international agreement. There is also diversity in state recognition of post-conflict traditional justice processes, ranging from de facto rejection to full incorporation. In Burundi, for example, the National Council of Bashingantahe was created by constitutional fiat to mediate disputes, including interethnic massacres and violence occurring since 1993. This occurred with foreign support in 2005, but the government has little enthusiasm for a revival of a precolonial decentralized system of adjudication that endowed the king and his chiefs with significant power at the local level (Uvin and Nee 2010). Meanwhile, in Mozambique, the government has been quietly tolerant of traditional accountability and reconciliation rituals. Ordinary people have been conducting *magamba* spirit ceremonies to create a socio-cultural environment conducive to engagement with the past and communal repair (Igreja and Dias-Lambranca 2008). However, there has been no formal engagement or endorsement of these practices, and the official line, premised on the allocation of impunity to known perpetrators of terrible acts, is to try to forget what happened.

In contrast, in some other post-conflict places, such as Sierra Leone and East Timor, traditional justice has been officially recognized and sanctioned. The 2000 Sierra Leone Truth and Reconciliation Act authorized the Truth and Reconciliation Commission (TRC) to "seek assistance for traditional and religious leaders to facilitate its public sessions and in resolving local conflicts arising from past violations of abuses in support of healing and reconciliation" (TRC Act Part 3(2)). Incorporation of traditional justice into the workings of the TRC was, however, rather weak (Huyse and Salter 2008). In East Timor, the government incorporated a more extensive range of customary law into their Reception and Reconciliation Commission (CAVR) community hearings. Three quarters of the reconciliation hearings involved a local dispute resolution practice named *nahe biti boot* (Drexler 2009; Stanley 2009). The hearings also incorporated long-established processes of *adat* or *lisan* to build local participation. In both Sierra Leone and East Timor, then – albeit to varying degrees – traditional justice has been used to supplement and legitimize more "formal" transitional justice processes. However, Rwanda is the only country where an adapted traditional accountability mechanism has been made wholly part of the official post-conflict justice policy and granted a central role as part of the formal state system (Wojkowska 2006).

Key Issues/Controversies

As noted in the introduction, literature promoting traditional justice as an aspect of transitional justice tends not to be focused on measuring the effectiveness of such processes or on understanding how such processes are experienced on the ground (Weinstein 2011). The result is a knowledge gap which has "produced decision making based on weak data, ex-ante evaluation and speculation" (Huyse and Salter 2008, p. 6). However, while the literature remains small and partial, it is becoming gradually more nuanced. In the following subsections, we comment on some of the key debates that have emerged.

Does State Capture Matter?

The term traditional justice usually ends up referring to a range of qualities found in local procedures which are in some way similar to those associated with conventional judicial processes, or established notions of transitional justice, or with generalized ideas about forgiveness. However, this kind of perception may mean that other qualities of local procedures may be entirely overlooked. Kimberly Theidon, for example, has warned of "the facile embrace of the local or community" as the "the realm of solution" (Theidon 2009, p. 296). Just because a process or an institution is nominally traditional does not insulate it from interference from various kinds of public authority, including the state. Furthermore, as anthropologists have shown, local customs relating to accountability can be highly dynamic and remarkably adaptable; they are rarely static and timeless. This is partly because they are mostly not written down but are endlessly negotiated. To codify or regulate them changes them.

These are issues that have been raised about the *gacaca* courts in Rwanda. International NGOs such as Amnesty International and Human Rights Watch, as well as non-Rwandan scholars have argued that the modern *gacaca* courts are controlled by the Rwandan government and have been used by an increasingly oppressive and authoritarian state to regulate reconciliation and justice processes in the peripheries (Ingelaere 2008). The argument follows that the state has interfered in the hearings in order to collectivize the guilt of all Hutu and, in doing so, has coerced Rwandans into publicly sharing the details of the genocide, thus violating a cultural and pragmatic inclination towards silence. Thus, legislation has transformed the original *gacaca* institution into something qualitatively different: legalistic procedures, state control, and forced participation for the population mean that the current process bears only partial resemblance to that which it was originally modeled on.

Other analysts have taken a more complex position, accepting that there has been a good deal of state capture, but noting that the *gacaca* system is not homogenous, and that the government's controlling role is not as pervasive as has been suggested. The *gacaca* system they maintain has a degree of autonomy and continues to resonate with local custom, even if the local courts are not quite what they were before. Phil Clark in particular has highlighted the elasticity and dynamism of the *gacaca* courts in his important finding that *gacaca* in one village could differ significantly from *gacaca* in another just a short distance away in terms of conduct, vibrancy of debate, and "societal impact" of hearings (Clark 2010). In his analysis, arguments about the government's role in *gacaca* tend to neglect the "importance of individual and communal agency in *gacaca* and the vital role of the general population in running and shaping the institution, often with highly unpredictable results" (87).

These autonomous and varied aspects of *gacaca* must be accepted. It is still, however, the case that the Rwandan government effectively used the system to institutionalize the allocation of blame. Where the state can, perhaps, most effectively "capture" traditional justice or at least most successfully manipulate it is not in the battle for direct control of these processes but rather in the battle of perceptions of wrongdoing. This has been a fundamental problem with the focus on the ritual of *mato oput* in Uganda. The focus on a custom associated with just one group, the Acholi, implies that the Lord's Resistance Army insurgency was a local affair, when in fact it was a conflict underpinned by national and international dimensions (Allen 2006, 2010). Elizabeth Drexler comes to a similar conclusion in East Timor, where she observes that the "excessive localization" of transitional justice processes risks "horizontalizing conflicts," positing them as "conflicts between different groups in society, rather than between a state and its citizens" (Drexler 2009, p. 50). In both these cases, a preoccupation with local justice (unintentionally or otherwise) protects crimes allegedly perpetrated by government officials and soldiers from scrutiny and accountability. It actually makes national political justice more elusive (Branch 2011).

Is Traditional Justice Restorative?

Related to the above discussion is the long-standing question as to whether justice should be restorative or retributive. As has already been indicated, some enthusiasts of traditional approaches assert that they are essentially restorative and reparative. Although the evidence base for this is open to question, it does seems that communal reconciliation and social repair are the apparent goals of reintegration rituals in many situations. This has been described for lower level ex-soldiers and perpetrators in countries such as East Timor and Sierra Leone (Shaw et al. 2010).

However, too much has been made of the almost intangible innateness of non-Western impulses towards forgiveness and restoration of social harmony. Such arguments reached their rhetorical height at the time of the South African Truth and Reconciliation Commission. As Pierre Hazan notes, Archbishop Desmond Tutu successfully "constructed a spiritual dimension to the process, linking Christian forgiveness and African mysticism to the goal of reconciliation" (Hazan 2010, p. 36). The resulting narrative was the notion that the "third way" (amnesty without oblivion) was somehow the African way. Invoking the concept of *ubuntu* which is largely a romantic expression of the "rural African community," Tutu said in one interview, "Ubuntu says I am human only because you are human… you must do what you can to maintain this great harmony which is perpetually undermined by resentment, anger, desire for vengeance. That is why African jurisprudence is restorative rather than retributive" (cf. Wilson 2001, p. 9). But this is misleading. We only need to look at the legacy of the South African TRC to understand that a single prescription of how to deal with South Africa's past was oppressive, and dissenting voices, although present at the time, were largely drowned out. Studies reflecting on the success of the TRC have found that earlier support for amnesty was "a reluctant, contingent concession that coexisted with a basic interest in seeing at least a degree of accountability," and that victims involved in hearings stressed that the desire for acknowledgement of wrongdoing and learning new information was the priority: it was truth, not reconciliation or forgiveness, that was paramount (Backer 2010, p. 453).

Despite the emphasis on a restoration of social harmony, there is a clear accountability component to most of the documented reconciliation rites (Huyse and Salter 2008). Reconciliation ceremonies in Uganda, Mozambique, Rwanda, Sierra Leone, and Burundi that have been adopted by activists tend to contain the requirement that the offender must acknowledge his or her guilt in order to be redeemed. In northern Uganda, for example, it is obvious to anyone who had read historical studies and early sources on the region that claims about the Acholi people forgiving offenders and accepting compensation were overblown. Punitive measures were common. It depended on the crime, who had committed it, and who was arbitrating (Allen 2006, 2010). In situations in which a crime is locally understood to be heinous, such as certain kinds of witchcraft, punishment could be very severe. In many places that remains the case, and violent responses to alleged witches and sorcerers, have been reported from numerous parts of Africa. In several countries, including South Africa, there have also been concerted efforts to incorporate trials for what might be called traditional or customary crimes into the formal system and to ensure custodial sentences for those found guilty. In Central African Republic, a UN prison study found that more than half of those being held had been accused of witchcraft.

In East Timor meanwhile, it is true that the Community Reconciliation Hearings were successful in reintegrating low-level combatants. The Commission for Reception, Truth, and Reconciliation in East Timor is said to have gained more than just legitimacy by promoting the hearings among traditional leaders. It secured widespread participation (Stanley 2009). The CAVR undertook 216 community reconciliation hearings for 1,379 perpetrators and it is also estimated that up to 40,000 people attended (CAVR 2005; Ch. 1.5:126). Such achievements are impressive, but it is important not to conflate the characteristics of a process with its outcomes. Widespread frustration has been reported among participants, who lamented their inability to challenge Indonesian impunity (Stanley 2009; Drexler 2009).

The reality is that Western justice systems and indigenous dispute resolution systems pursue the similar objectives to different degrees and the restorative/retributive dichotomy is exaggerated and essentializing. Usually the biggest difference between formal and informal approaches is the actual procedures employed to reach the various objectives (PRI 2002). Although there are exceptions, informal systems tend to draw more on ritual elements and emphasize the community dimension of criminal behavior over individual accountability (Huyse and Salter 2008). This is where local processes have very clear benefits. In complicated situations, the “guilty,” “not guilty,” “victim,” and “perpetrator” dichotomies can be misleading and even harmful. Violent conflicts characterized by moral “grey zones” in which different forms of guilt and innocence are mixed are complicated territory for criminal law and render the delivery of clear verdicts a difficult exercise (Shaw et al. 2010; Hinton 2011). These are situations in which people inhabit shifting “perpetrator-victim” identities: the child soldier, abducted from its family and forced to commit brutal crimes in the course of conflicts in Sierra Leone and Uganda, provides a case in point (Baines 2010). Courtrooms are not usually capable of dealing with the subtlety needed to address such complexities, and moreover, the shaping of these complexities into simple legal categories risks misrepresenting the situation and can make post-conflict reconciliation ever harder (Shaw et al. 2010; Hinton 2011). As one scholar notes “a combination of palavers, the African way of prolonging discussions, and ritual events create in principle, more opportunities for exploring issues of accountability, innocence and guilt that are integral to the legacy of violent conflict” (Huyse and Salter 2008, p.15).

Does It Matter That Traditional Justice Can Be Discriminatory?

From Afghanistan to Sierra Leone, studies have highlighted the persistent ethnic, religious, generational, and gender hierarchies and divisions that complicate and limit the effectiveness of traditional practice from a transitional justice perspective. This is a point implicitly recognized in Kofi Annan’s observation to the UN Security Council that the rule of law and transitional justice must be “in conformity with both international standards and local tradition” (UNSC 2004, p. 12).

In African contexts, for example, social accountability is closely connected with hierarchies, ones which may systematically subordinate and repress particular groups, notably women. Probably the most frequently mentioned cause of concern is that tradition-based systems of dispute resolution are likely to be dominated by men. As Huyse and Salter find in their edited collection on post-conflict traditional justice, in Mozambique, only the spirits of men killed during the civil war are allowed to return to the realm of the living to claim justice; in Burundi, women are not allowed to become members of the Ubushinganthe; they can only participate in the proceedings as the wife of the widow of a member, and the traditional justice system in Sierra Leone exhibits a clear prejudice towards married women, although some provision is made for female representation (Huyse and Salter 2008). It has been documented that in Somalia, a woman who is raped is often forced to marry her attacker, while customary practices of wife inheritance as an aspect of ritual cleansing continue in parts of Kenya (Wojkowska 2006). In Afghanistan, women can be exchanged in compensation for criminal offences. This is regarded as preferable to the alternative: a blood feud that might escalate into full-blown tribal conflict (Schmeidl 2011). A Commission on Conflict Mediation set up in Khost Province in 2007 attempted to integrate formal and informal justice processes, and while it has been regarded as relatively successful, the Commission still follows the practice of only allowing men to litigate traditional justice (Schmeidl 2011, p. 162). In East Timor, women were sidelined in the community reconciliation hearings because their participation faced resistance from male family members and also because they were constrained in their ability to attend due to family and home duties. In a study by Elizabeth Stanley, CAVR staff commented that this was probably to be expected: "We are living in a patriarchal society so patriarchy is bound to be reflected in the collation of testimonies" (Stanley 2009, p. 117). Despite efforts by regional workers to access women's stories she finds that the CAVR "framed women out of stories" (Stanley 2009, p. 117).

The male dominance of local justice mechanisms can also be compounded by overtly patriarchal characteristics too. Rituals may be co-opted by male elders to further their own interests. In Sierra Leone, for example, customary law is the domain of senior men and this creates concern among younger men that their land and other possessions may be removed from them through "notoriously arbitrary and excessive fines" (Shaw et al. 2010, p. 16). It has been observed that although the TRC reconciliation rituals may have been effective in helping to reintegrate high-ranking ex-combatants, they also "retrenched young men's subordination to 'big men', a situation that had watered the roots of Sierra Leone's armed conflict in the first place" (Shaw et al. 2010, p. 17). It has also been noted that elders in northern Uganda have deployed traditional justice techniques as a way of disciplining returning LRA combatants, particularly young men and women (Branch 2011). Sometimes traditional justice seems to be focused on the *re*constitution of *pre*-conflict structures. The kind of senior male authority that is legitimated in customary law is largely upended during conflict and war. Family dislocation, mass displacement into refugee camps, and mass migration (particularly of youth) into the cities have disrupted the "natural biotope" of traditional practice and undermined the status enjoyed by customary leaders (Huyse and Salter 2008, pp. 185–186). But is it socially progressive to reinstate the social order? If that social order was linked to the outbreak of violent conflict, it is not at all clear that such an agenda is appropriate. Also many will resist it if they can, especially women and young people who have found new opportunities during times of upheaval (Branch 2011). Sometimes, it is in such circumstances that accusations of witchcraft proliferate, with arbitration procedures being co-opted by senior men to assert authority, and vulnerable women often ending up being targeted.

There are, however, some interesting ways in which traditional mechanisms are being used to challenge traditional values and social orders. In Afghanistan, the *jirga*, traditionally an ad hoc forum for Pashtun elders to assemble and discuss a particular issue of concern, has been redesignated to describe any gathering aimed at consultation with the general population. In 2010, Afghan Civil Society groups established the *Victims Jirga for Justice* in response to the flawed, non-inclusive government run *Peace Jirga*. The *Victims Jirga* included over 100 victims from all over Afghanistan and provided the first truly national articulation of a transitional justice agenda, including demands for prosecutions, truth-seeking, and reparations (Kouvo and Mazoori 2011). By loosely adopting the traditional *jirga* framework for discussion, the meetings provided a familiar and supportive space to recount the abuses that had been endured and to formulate policy proposals to the government (Kouvo and Mazoori 2011). Similar "curative" spaces have been created by Women's Courts in Guatemala and Columbia. In Columbia, women have been holding regional tribunals in preparation for the launch of the permanent Columbian Women's Court Against Forgetting and Re-existence (Quest 2008). A hybrid of legal and nonlegal procedures, the tribunals include rituals of apology and judgment by a panel of "wise women." In the case of the Victims Jirga and women's courts, traditional processes are being used by excluded populations to facilitate testimony and to formulate recommendations based on these testimonies to the government. Nevertheless, while promising, these kinds of measures are far from widespread, and the underlying problem remains. In practice, customary justice can elevate the goal of community harmony above individual rights and freedoms in worrying ways.

Is Traditional Justice Appropriate to Deal with Mass Crimes?

Probably the important outstanding question in any given instance where traditional processes are a focus of post-conflict policies is whether the measures being promoted are really capable of dealing with large-scale war crimes, genocide, and crimes against humanity, sometimes committed over long-time periods and often involving the destruction of the very social and material systems upon which indigenous processes depend. As has already been alluded to, it is unclear whether community-based processes can resolve intercommunal problems as their scope and legitimacy tends to be limited. A big concern among international lawyers and human rights NGOs is that customary tools do not respect the duty under international law to prosecute mass atrocities. Large NGOs such as Human Rights Watch and Amnesty International are outspoken defenders of this position and accept no curtailment of the international obligation to prosecute. Legalist critics apply strict criteria in their assessment of community-based justice processes, and as such, institutions like the *gacaca* have come under widespread criticism for an apparent failure to ensure due process, including professional representation and rules of evidence.

Analysts with a less narrowly legalistic approach are more open to the possibilities of informal procedures. Indeed, among such commentators, there is a tendency to begin with a kind of ideal model of transitional justice in which traditional processes are complementary to more formal post-conflict justice processes. Although findings remain rather vague, inconclusive, and anecdotal, studies of places as varied as Peru, Burundi, and Afghanistan tend to suggest that traditional processes are "partially effective" and "partially legitimate" in addressing post-conflict justice issues (Huyse and Salter 2008, pp. 188–190) and that they could effectively combine with other strategies for dealing with accountability and reconciliation. However, even where this is being attempted, we do not have a clear enough understanding about how different processes actually function together in practice and with what outcomes. In so far as there is evidence, it would appear the relationship is characterized by suspicion, friction, and sometimes incompetence rather than productive cooperation.

Competition of this kind is evident in Rwanda, where international, national, and localized courts comprise judges and lawyers with "divergent interpretations of the role and objectives of transitional justice" resulting in a "stratified and at times competitive set of criminal courts" (Palmer 2012, p. 3). In Sierra Leone, the TRC and the SCSL worked in parallel for 18 months, and the precise nature of their relationship was never clarified. The two bodies ended their period of parallel operation with tension over testimony of indicted prisoners, but it is still up for debate whether the court and the TRC, generally, coexisted happily (Schabas 2004). Certainly the lack of clarity between the two institutions confused Sierra Leoneans, and one widespread theory was that there was an underground tunnel linking the two, through which information given to the TRC was immediately leaked to the SCSL (Nkansah 2011). Meanwhile, in East Timor, the failure of collaboration between the CAVR process and the Special Crimes Unit has been commented on by victims who suggest that "good" progress in the former was "downgraded" by the failure of the latter to bring cases to court and that there remains "a considerable amount of unfinished business – a significant caseload that falls in between the two procedures" (Stanley 2009, p. 122). Finally, Tim Kelsall (2009) has shown how the kinds of "supernatural evidence" common to ritual and informal processes fared during the hearings of the Special Court for Sierra Leone. He finds that rather than clash directly, international legal and local norms appeared to simply elude one another. He argues that the court's decision to sidestep the issue of magic and the occult during the trial was an ethnocentric mistake which made little sense to local populations.

Does It Matter That Traditional Justice Is Sometimes Largely Invented?

As elsewhere in Africa, research in northern Uganda has shown how rituals and ceremonies are used to interpret the spirit world and the experience of misfortune and to reestablish or make manifest social relations. Ceremonies and rituals that become important at any particular time are by no means always old ones that are taken "off the peg," but rather ideas about old models are used to help shape new ones. Despite this, attempts *have* been made to try and codify rituals into an ostensibly coherent form of traditional justice. During the time of the Ugandan Protectorate, the British administrators incorporated selected tribal customs into the indirect system of government through chiefs and other local agents. More recently in northern Uganda, in the context of efforts to establish peace and reconciliation during the 20-year conflict between the Government of Uganda and the Lord's Resistance Army (LRA), there have been moves to codify community-based rituals, particularly those that draw on traditional Acholi values and institutions. For the powerful coalition of sympathetic international agencies, activists, traditional leaders, and religious leaders advocating this, the reasons were twofold: firstly, the war had led to a breakdown in traditional social values that needed to be restored and secondly, decisions about peace, justice, and reconciliation should be made by the victims and not by the Ugandan government or a foreign institution, such as the ICC (Allen 2006, 2010).

This approach was subsequently articulated vigorously in the peace negotiations between the Ugandan government and the LRA in Juba and ended up being formalized in the 2007 Agreement on Accountability and Reconciliation signed between the Government of Uganda and the Lord's Resistance Army, which listed both Acholi and non-Acholi rituals as part of the broader transitional justice effort in Uganda, alongside a new War Crimes Division in the High Court and a possible truth commission. During these developments, the local justice mechanisms were promoted without much understanding of local circumstances. Indeed serious concerns had emerged about the chieftaincy system's capacity to implement the agenda that was being proposed. Critical examinations concluded that traditional structures were weak and fragmented, that many of the elders were themselves not sure how to carry out traditional rituals, and that there was widespread disagreement about who the *real* traditional leaders were (Bradbury 1999). It was also noted that there were tensions between elders over the possible financial benefits and that there were concerns that the external support for traditional chiefs was just another way of trying to bring the region under closer government control without contributing to improved education and economic development (Bradbury 1999). Other research in the region undertaken in 2004 and 2005 found that northern Ugandan populations were critical and circumspect about the value and potential of traditional justice solutions and that few people considered the traditional structures a key priority (Allen 2006).

Does this matter if the newly codified practices prove to be helpful? Probably it does. The problem with codifying selected local practice, as noted above, is that it takes them out of the contexts in which they have been used and adapted flexibly to specific circumstances, and it reifies them. If they are categorized and institutionalized into semiformal judicial systems, they will inevitably be very different to what they were to start with. They will lose their flexibility and will no longer have all the many resonances and associations of lived ritual actions.

But crucially, they will have a status that is at least partly based on their externally supported authority. They will become privileged rites and most likely the preserve of certain figures of male authority recognized by the international community or the government. Interestingly, one supportive NGO report conceded that elders will need to be trained on traditional practices and the younger generation "do not even know how to be Acholi" (Liu Institute 2000, p. 22). It is revealing that on one occasion, the USAID funded Northern Ugandan Peace Initiative (NUPI) organized for elders to explain Acholi forgiveness rituals to representatives of "the youth." At the time, the paramount chief admitted that he did not know how to perform the traditional *mato oput* ceremony (Allen 2010). It seems odd that it is now up to non-Acholi experts and outsiders to help revive those traditions among the Acholi.

This was most likely not the intention of many of the advocates of traditional justice, but it is certainly symptomatic of what Adam Branch has described as the "ethnojustice" agenda (Branch 2011). This approach mistakenly views traditional systems of justice as a "single, coherent and positive system... universally, consensually and spontaneously adhered to by all members of that culture" (Branch 2011, p. 163). It goes without saying that the reality is more nuanced than this. In a study of the Kpaa Mende, southern and eastern Sierra Leone, Joe Alie found that certain Mende customary practices are only applicable to and resonant with certain elements of the community and, since there has been a great deal of intermarriage with other ethnic groups over the years, these practices may not be suitable for settling disputes between Mende and non-Mende people within the community (Alie 2008). In northern Uganda, the Acholi *Mato Oput* receives the most attention, but the Langi, Teso, and Madi, also affected by the 20-year conflict, have their own rituals. An even more neglected fact, even though it may sound like a truism, is that attitudes towards traditional approaches vary *within* ethnic groups too. There is no integrated system of traditional justice among the Acholi, for example. Traditional approaches are less relevant and less acceptable to some – this is especially true for young people who have grown up during a time of war with restricted opportunities to experience or participate in such practices. As Branch points out, and it is surely a point that is applicable across contexts: rather than Acholi traditional justice, we should talk about Acholi *traditions* of justice (Branch 2011, p. 177).

Conclusion

It is difficult to conclude with a general statement about traditional justice, because it is not clear exactly what is being discussed. Measures associated with social accountability vary widely within population groups as well as between them, and the kinds of mechanisms selected to be called traditional justice by advocates are rarely more than a selection of activities that conform with normative ideals, usually linked to the notion that they ought to be restorative. As we have indicated here, research reveals that local judicial measures may be linked to state interests or may have qualities that are highly problematic from an international perspective. However, problems of legitimacy, exclusion, gender bias, and politicization are also manifestly evident in formal national justice systems dealing with post-conflict accountability, and these problems also emerge in various ways at the level of international tribunals and courts. So the turn to the local and traditional for a better approach is likely to persist, whatever the controversies involved.

There is to date little detailed knowledge of the effects of projects and programs that have sought to promote putatively traditional systems. The findings of research that has been carried out suggest that they may be helpful in some instances, but overall results are mixed. In some instances, counter-productive consequences have been noted. Much more adequate assessment is required and much better monitoring. There is no doubt that local rituals and customs are important for populations caught up in violent conflict and dealing with its aftermath. However, there is also no doubt too that those local rituals and customs do not form a coherent alternative to formal national and international processes. Traditional justice cannot be harnessed to the transitional justice agenda in a straightforward way. The situation will vary radically from place to place, and where it occurs, the local mechanisms will take on hybrid qualities. Indeed to call them traditional will almost inevitably become a misnomer. They will change, and how they change needs to be closely observed to ensure positive outcomes.

For an extended version of this text with a full and more extended bibliography please see: http://www2.lse.ac.uk/internationalDevelopment/research/JSRP/JSRP%20Papers/JSRP-Paper-3.aspx (last accessed 16/04/13).

Recommended Reading and References

Alie J (2008) Reconciliation and traditional justice: tradition-based practices of the Kpaa Mende in Sierra Leone. In:Huyse L, Salter M (eds) Traditional justice and reconciliation after violent conflict: learning from African experiences. IDEA

Allen T (2006) Trial justice : the international criminal court and the lord's resistance army. Zed Books, London

Allen T (2010) Bitter roots: the 'invention' of acholi traditional justice. In: Tim A, Koen V (eds) The lord's resistance army: myth and reality. Zed Books, London, 2010

Arthur P (2009) How transitions reshaped human rights: a conceptual history of transitional justice. Hum Rights Q 31(2):321–67

Backer D (2010) Watching a bargain unravel? A panel study of victims' attitudes about transitional justice in cape town, South Africa. Int J Transit Justice 4(3):443–456

Baines E (2010) Spirits and social reconstruction after mass violence: rethinking transitional justice. Afr Aff 109:409–430

Benda-Beckmann, Von F (2002) Who's afraid of legal pluralism. J Leg Pluralism Unofficial Law I47:37–82

Bradbury M (1999) An overview of initiatives for peace in Acholi, northern Uganda, Cambridge, collaborative for development action

Branch A (2011) Displacing human rights: war and intervention in northern Uganda. Oxford University Press, Oxford

CAVR (2005) Report of the commission for reception, truth and reconciliation in Timor Leste *Chega!*

Clark P (2010) The Gacaca courts, post-genocide justice and reconciliation in Rwanda : justice without lawyers. Cambridge University Press, Cambridge

De Grieff P (2011) Some thoughts on the development and present state of transitional justice. Anuario de Derrechos Humanos, No. 7, Santiago, Chile

Drexler E (2009) Addressing the legacies of mass violence and genocide in Indonesia and East Timor. In: Hinton A, O'Neil L (eds) Genocide: truth, memory and representation, Duke University Press. Durham

Hayner P (2010) Unspeakable truths: transitional justice and the challenge of truth commissions, 2nd edn. Routledge, New York: T & F Books

Hazan P (2010) Judging war, judging history: behind truth and reconciliation. Stanford University Press, Stanford

Hinton A (ed) (2011) Transitional justice: global mechanisms and local realities after genocide and mass violence. Rutgers University Press, New Brunswick

Huyse L, Salter M (2008) Introduction: tradition-based approaches in peace-making, transitional justice and reconciliation policies. In: Huyse L, Salter M (2008) Traditional justice and reconciliation after violent conflict: learning from African experiences. International Institute for Democracy and Electoral Assistance, Stockholm

Igreja V, Dias-Lambranca B (2008) Restorative justice and the role of magamba spirits in post-civil war Gorongosa, central Mozambique. In: Huyse L, Salter M (eds) Traditional justice and reconciliation after violent conflict: learning from African experiences. International Institute for Democracy and Electoral Assistance, Stockholm

Ingelaere B (2008) Does the truth pass across the fire without burning? Locating the short circuit in Rwanda's Gacaca courts. J Mod Afr Stud 47(4):507

Kelsall T (2009) Culture under cross-examination: international justice and the special court for Sierra Leone. Cambridge University Press, Cambridge

Kouvo S, Mazoori D (2011) Reconciliation, justice and mobilization of war victims in Afghanistan. Int J Transit Justice 5(3):492–503

Latigo JO (2008) Northern Uganda: tradition-based practices in the Acholi region. In: Huyse L, Salter M (eds) Traditional justice and reconciliation after violent conflict: learning from African experiences. International Institute for Democracy and Electoral Assistance, Stockholm

Liu Institute for Global Issues, Gulu District NGO Forum and Ker Kwaro Acholi (2005) Roco Wat I Acholi, restoring relationships in Acholi-land: traditional approaches to justice and reconciliation. Liu Institute for Global Issues, Vancouver,

Marc A (2010) New frontiers of social policy: delivering services in multicultural societies. World Bank, Washington, DC

Moore SF (1986) Social facts and fabrications: 'customary' law on Kilimanjaro 1880–1980. Cambridge University Press, Cambridge

Nkansah LA (2011) Restorative justice in transitional Sierra Leone. J Public Adm Governance 1:1

Nouwen S (2011) Complementarity in Uganda: domestic diversity or international imposition? In: Stahn C,

Zeidy ME (eds) The international criminal court and complementarity: from theory to practice. Cambridge University Press, Cambridge, 2011
Organisation for Economic Cooperation and Development (2007) Enhancing the delivery of justice and security, governance, peace and security. OECD, Paris
Palmer N (2012) Transfer or transformation: a review of the rule 11 bis decisions of the international criminal tribunal for Rwanda. African Afr Journal of International and Comparative Law 20:1–21
Penal Reform International (2002) Access to justice in Sub-Saharan Africa: the role of traditional and informal justice systems. Penal Reform International, London
Quest S (2008) Justice reform and gender, DCAF, OSCE/ODIHR, UNINSTRAW
Roht-Arriaza N (2006) Transitional justice in the twenty-first century : beyond truth versus justice. Cambridge University Press, Cambridge, UK
Schabas W (2004) A synergistic relationship: the Sierra Leone truth and reconciliation commission and the special court for Sierra Leone. Crim Law Forum 15:3–54
Schmeidl S (2011) Engaging traditional justice mechanisms in Afghanistan: state building opportunity of dangerous liaison? In: Mason W (ed) The rule of law in Afganistan: missing in inaction. Cambridge University Press, Cambridge
Shaw R (2007) Memory frictions: localizing the truth and reconciliation commission in Sierra Leone. Int J Transit Justice 1(2):183–207
Shaw R, Waldorf L, Hazan P (2010) Localizing transitional justice: interventions and priorities after mass violence. Stanford University Press, Stanford
Sriram CL (2007) Justice as peace? Liberal peacebuilding and strategies of transitional justice. Glob Soc 21(4):579–591
Stanley E (2009) Torture, truth and justice: the case of Timor Leste. Routledge, London
Theidon K (2009) Editorial note. Int J Transit Justice 3(3):295–300
UN Security Council (2004) The rule of law and transitional justice in conflict and post-conflict societies. Report of the Secretary-General, New York, 23 Aug 2004
Uvin P, Nee A (2010) In: Shaw R et al (eds) Localising transitional justice. Stanford University Press, Stanford
Weinstein H (2011) Editorial note: they myth of closure, the illusion of reconciliation: final thoughts on five years as Co-editor-in chief. Int J Transit Justice 5(1):1–10
Wilson R (2001) The politics of truth and reconciliation in South Africa: legitimising the post-apartheid state. Cambridge University Press, Cambridge
Wojkowska E (2006) Doing justice: how informal systems can contribute. UNDP, Oslo
World Bank (2011) World development report 2011: conflict security and development. World Bank, Washington, DC

Postmodern Criminology

Bruce A. Arrigo and Heather Y. Bersot
Department of Criminal Justice and Criminology, University of North Carolina at Charlotte, Charlotte, NC, USA

Overview

Postmodern criminology is a critically-animated theoretical framework that humanistically accounts for the problems of crime and the possibilities of justice at the self/society divide. Mindful of the development of postmodernist-inspired criminology, this entry succinctly reviews four issues. First, it describes the philosophical evolution of the theory, emphasizing its negative and fatalistic; affirmative, constitutive, and integrative; and ultramodern strains of heterodoxy. Second, it presents several key concepts refined throughout postmodern criminology's historical development. These pivotal concepts include the following: (1) discourse; (2) meaning; (3) knowledge; (4) power; and (5) subjectivity/agency. Third, it delineates at some greater length postmodern criminology's "state-of-the art" philosophy. This delineation includes attention to current (and novel) directions in theory, method, and praxis. The entry concludes by specifying a number of conceptual controversies that warrant further philosophical elaboration. These controversies are characterized as "open questions" that await postmodern criminology's consideration.

Origins of Postmodern Criminology

Throughout history, people have sought ways to make sense of human behavior and social phenomena. During the Middle Ages, theologically-based concepts were employed to explain human nature and the seemingly ongoing battle between good and evil. When such accounts were deemed inadequate, people endeavored to better understand themselves and the world in which they lived through more scientific methods. This modern age, also known as the Enlightenment period, ushered in a new philosophical approach to examining peoples' thoughts and actions. Humans were now perceived as logical beings capable of exercising deductive reasoning, employing free will, and choosing their own desires (Best and Kellner 1991). In this way, absolute truths purportedly were made discernable. This modern age "held the extravagant expectation that the arts and sciences would further not only the control of the forces of nature but also the understanding of self and world, moral progress, justice in social institutions, and even human happiness." (Habermas 1980, pp. 162–163.)

The phrase, *Cogito, ergo sum – I think, therefore I am*, is often utilized to describe the modernist's logical and fully autonomous subject. Since the Enlightenment period, this belief in a reasoning agent has endured. As such, the concept of the rational individual has provided the foundation for numerous criminological theories employed by researchers to study and explain wayward behavior. For example, rational choice theory rests on the notion of a self-made person who purposefully chooses his or her actions. The theory asserts that a criminal makes use of deductive reasoning while weighing the benefits and costs of committing an offense. If the perceived benefits of committing the crime (e.g., money) outweigh the perceived costs (e.g., incarceration), the criminal will engage in the unlawful act. Thus, these and other modernist-inspired theories support a number of deterrence and retributive-based responses to crime and delinquency.

While modernism seemingly promises systematic and robust ways to explain and predict human behavior, some question its ability to adequately address problems that plague contemporary society (Best and Kellner 1991). The philosophy of postmodernism emerged in response to the modernist contentions regarding human behavior and social phenomena. Its proponents charge that science fails to sufficiently account for the human experience. In other words, while empirical methods may explain the biology of human beings, they cannot fully explicate the anomalies of being human differently. After all, contradictions and inconsistencies, spontaneities and absurdities are also a part of our humanities.

Further, while supporters of the modernist agenda argue that truth is absolute and knowledge is certain, postmodern thinkers assert that commitments to totalizing truths or universalizing epistemologies are the fictions of the Enlightenment era (Arrigo et al. 2005). As Foucault (1977) insightfully proclaimed, knowledge consists of only "claims to truth." A number of postmodern commentators contend that such claims have contributed to the establishment of harmful movements that egregiously advance the ideological interests of imperialism, sexism, racism, and class oppression (Henry and Milovanovic 1999). Thus, as one of the foremost postmodernists, Jean-Francois Lyotard (1984) declared, "Let us wage a war on totality; let us be witnesses to the unpresentable; let us activate the differences and save the honor of the name" (p. 82). The "honor" to which he alludes, is our own dignity lived differently and celebrated dynamically.

Negative and Fatalistic

Some of the earliest postmodern thinkers consisted of Jacques Lacan, Roland Barthes, Gilles Deleuze and Felix Guattari, Jean Baudrillard, Jacques Derrida, and Julia Kristeva (Arrigo et al. 2005). Although he did not formally promote a postmodern perspective, French philosopher Michel Foucault is traditionally hailed as one of the leading exponents of this approach. The intellectual stance that these post-Enlightenment luminaries maintained with respect to truth and knowledge has led some to describe them as negative and fatalistic in philosophical orientation (e.g., Schwartz and Friedrichs 1994). Indeed, according to "first wave" postmodern scholars, total objectivity cannot be established, final truths neither exist nor are they fully discoverable, and the foundations of any proposed structure (whether in art, reason, science, and/or their cultural artifacts) can be disassembled and reassembled in countless ways. Thus, negative and fatalistic postmodernists seek to challenge all claims to certainty, exactitude, and completeness. In doing so, they endeavor to expose underlying assumptions in text-making that seemingly support such finite declarations.

To debunk absolute claims to truth, postmodern commentators engage in critique. This approach to uncovering fragments, fictions, and falsehoods in knowledge production is not simply rooted in criticism but also is sourced in *deconstruction*. As method, deconstruction systematically examines that which appears to be true and reveals contradictions, inconsistencies, oppositions, and interdependencies lodged within presumably neutral and impartial texts (Derrida 1970, 1981). As such, those who endorse this method of inquiry do not insist on the existence of alternative truths. This reasoning would undo the logic of deconstruction. Indeed, from the negative and fatalistic perspective of postmodernism, to insist on an epistemological standpoint entails that one make a categorical truth claim. Assertions such as these represent a landscape of inexhaustible deconstructive inquiry. This is why the early postmodernists have been classified as skeptics, nihilists, and anti-foundationalists (Henry and Milovanovic 1996, 1999). Under these de-stabilizing conditions, what purpose could the making of meaning serve?

Affirmative, Constitutive, and Integrative

Where the negative and fatalistic approach to postmodernism seeks to make explicit the "undecidable" nature of texts, the affirmative, constitutive, and integrative perspective (*for applications see*, Arrigo and Milovanovic 2010) also endeavors to *reconstruct* them. This reconstruction entails claims to conditional, positional, and relational truth and knowledge (Arrigo 1995). As an intellectual transition in theory development, attention is directed toward understanding how an "edifice [a text or system of thought] is built, and how it stands, in spite of opposition [but also] how it can be rebuilt or built differently" (Einstadter and Henry 1995, pp. 280–281). These replacement constructions acknowledge that humans co-produce and reify the structures, organizations, and institutions of which they are a part, while, simultaneously and interdependently, these material forces contribute to defining the subject's very own possibilities. This is the meaning of the constitutive enterprise (e.g., Henry and Milovanovic 1996).

While negative and fatalistic postmodernists contend that humans are susceptible to the limit-setting influences of these structural intensities (e.g., the "death" of the subject), second wave proponents of postmodernism demonstrate how people can nevertheless create new social worlds and emergent social selves (Henry and Milovanovic 1999). Along these lines, criminologists in support of actualizing these potentialities have called for greater interdisciplinary academic border-crossing. This would entail synthesizing insights derived from such disparate fields as chaos theory, psychoanalytic semiotics, literary criticism, cultural studies, existential phenomenology, and critical dialogical pedagogy (Arrigo and Milovanovic 2010).

Ultramodern

Recently, a third transition within the philosophy of postmodernism has developed. This new wave of thinking is known as the ultramodern. Consistent with constitutive theory, this perspective asserts that the self/society mutuality is co-produced by the human agent and the structural/organizational intensities that influence/shape the subject. To this notion, ultra-modernists contribute the view that four mutually supporting spheres of interdependent influence exist. These influencing flows include circumscribed images (Symbolic realm), privileged texts (Linguistic realm), embodied inscriptions (Material realm), and their replications and disseminations (Cultural realm) (Arrigo et al. 2011; Arrigo and Milovanovic 2009; Milovanovic 2011). The relationship between the four spheres and the twin dynamics (i.e., self/society mutuality) is one in which each force contributes to the maintenance of the others. Ultramodern commentators charge that these spheres of influence currently perpetuate and sustain harms of reduction (limits to being) and harms of repression (denials of becoming). This harm, both material and existential in composition, extends its captivity beyond offenders (the kept) to include their keepers, regulators and watchers. In the ultramodern age, this captivity is criminal.

The ultramodern version of postmodernism seeks to specify how the ubiquity of this captivity renders human/social progress no more than a *shadow* of what it could be or could become for one and about all. To transcend harms of reduction/repression, ultra-modernists employ both deconstructive and reconstructive techniques. The most promising reconstructive technique emerges from within a virtue ethic that celebrates human/social flourishing and the different and transformative possibilities that pursuing this ethic implies. Through ongoing engagement of this sort, ultra-modernists maintain that a new type of citizenship (i.e., personal well-being; individual excellence) and a new quality of justice (i.e., for one, about all, and collectivist) can be seeded and nurtured in provisional, positional, and relational contexts (Arrigo et al. 2011).

Background on Postmodern Criminology Concepts

Postmodern criminology is an ever-evolving philosophy. Indeed, there are numerous contributors to its theoretical development and research trajectory (Arrigo and Milovanovic 2010). Notwithstanding this rich intellectual history, several shared notions have been pivotal to its evolution. These notions include: discourse, meaning, knowledge, power, and subjectivity/agency. In what follows, each of these concepts is briefly described.

Discourse

From the modernist perspective, discourse is neutral. That is, the act of speech and the act of writing are instruments by which individuals communicate rationally formulated expressions. However, postmodernists regard discourse as a powerful vehicle through which humans construct identity (e.g., self, others, situations). Drawing primarily upon Lacan's (1977) groundbreaking work on discourse construction and analysis, postmodernists view human subjects as "formed by and through their use of language and by the inherent meaning that language use creates and invokes" (Henry and Milovanovic 1996, p. 8).

Lacan's schematizations on discourse are featured prominently in the theory and method of deconstructionism (Arrigo et al. 2005). This latter concept is something of a fulcrum for a number of ideas on which postmodernism rests. Indeed, postmodern thinkers "consider all narrative accounts to be texts, including books, stories, descriptions, reports, accounts, even non-verbal communication such as demeanor and gestures. Thus, all discourse of any kind is a text, as are all phenomena and events" (Henry and Milovanovic 1996, p. 5; *see also* Rosenau 1992). And, as previously explained, the undecidable nature of all texts (i.e., that they are filled with multiple readings), means that truth, even when specified as such, can never fully or exhaustively be discerned. For example, juridical language is a text through which judges convey their opinions on legal matters. However, court declarations are only one rendition of a given reality. The text of "legalese" produces meaning that is consistent with black letter law; however, lived experience is purified by way of legal language. Thus, "court talk" cannot fully communicate or embrace the non-sanitized experiences of those to whom or about whom it is directed. This "cleansing" process is what makes juridical discourse totalizing in its stigmatization.

In the illustration above, we see how human interactions and social phenomena are connected through language, including the system of language (i.e., the text) that is the law. But, given the postmodernist perspective on discourse, any text can be disassembled to make more explicit how it (unwittingly) gives preference to majority "voices" (e.g., those of status and standing) over minority "voices" (e.g., those lacking social, economic, political capital). This is why "[d]econstruction tears a text apart [and] reveals its contradictions and assumptions" (Rosenau 1992, p. xi). Legal and criminological postmodernists, particularly those who endorse the affirmative and ultramodern perspectives, contend that these texts can then be reassembled so that they more completely affirm the voices (the identities) of the excluded and powerless (e.g., Arrigo and Milovanovic 2009, 2010).

Meaning

People construct their identities, their realities, through communication (the written or spoken word), and by assigning meaning to the language they utilize. Postmodernists frequently employ phrases such as “language games,” “regimes of signs,” “discursive formations,” “rhetorics,” and “linguistic coordinate systems” when referencing the specialized meaning situated within a given text (Henry and Milovanovic 1996, p. 8). In the lexicon of postmodernism, words are often referred to as signifiers. That is, words are unstable, value-laden, and freighted with meaning. To illustrate, consider Henry and Milovanovic’s (1996) assessment of the language employed in the Fourteenth Amendment’s provision that “no person. . .shall be deprived of life, liberty, or property without due process of the law” (p. 9). As they contend, terms such as “person,” “life,” “liberty,” and “law” are signifiers. These signifiers have a specific meaning within the realm of jurisprudence. The process of meaning-making by way of signifiers, then, is a critical component to how the subject constructs identity, including their social worlds and social selves (Lacan 1977, *see also* Arrigo et al. 2005).

Knowledge

From the modernist perspective, truth can only be retrieved by relying on robust, scientifically-sourced methods. In doing so, “Absolute Postulates” emerge in which “all other ‘facts’ can be explained [utilizing] linear, deductive logic” (Milovanovic 1997, p. 14). With such truth claims, a preferred system of knowledge is established. It is a system of thought, of knowing, that demands compliance if conventional meaning is to be assured. Stated differently, this system of knowing is one that demands that others think, communicate, and live similarly rather than differently.

In contrast, postmodern adherents argue that since texts are poly-vocal, knowledge about such renditions of reality and identity are, at best, communicated in fragments and through impressions. This is why there is “no independent reality other than in the minds and practices of those who create them and recreate them” (Henry and Milovanovic, 1999, p. 5). Indeed, truth and the knowledge claims about its legitimacy are social constructions or fictions that temporarily anchor our existences in circumscribed and incomplete ways (Arrigo and Williams, 2006; Best and Kellner 1991). In other words, specialized types of knowledge exist for individuals and/or groups but they can only ever be historically contingent, awaiting careful textual exegeses. Thus, while extant systems of modernist knowledge contend that empirical inquiry provides the most exacting path by which to explain and predict human behavior, postmodernists point out that science’s mantle of privilege and respectability often discounts, ignores, or dismisses other worthwhile avenues for knowing and communicating the difference that is the human/social experience. After all, “science,” too, is a social construction, a way of rendering reality.

Power

Featured prominently in postmodern thinking is the role of power. According to two of its leading exponents, Deleuze (1988) and Foucault (1977), the concept of power is present in all relations. A number of postmodernists posit that crime is the power to harm and deny others their humanity (Arrigo et al., 2011; Arrigo and Milovanovic 2009; Henry and Milovanovic 1996). In other words, crime is “an expression of energy to make a difference over others, to the exclusion of those others who, in the instant, are rendered powerless to make [and be] their own difference” (Henry and Milovanovic 1996, p. x). This power is legitimized through systems of thought within such “disciplines” as education, medicine, law, and penology. This disciplining is then institutionalized by way of indoctrinated agents (the keepers and regulators of the kept) (Arrigo and Milovanovic 2009; Arrigo et al. 2011). From the postmodernist perspective, however, such systems of knowing “advance only specialized truths that serve to establish or maintain oppressive power relations" (Arrigo and Williams 2006, p. 13).

The effects of such disciplining by way of privileging some texts (i.e., slavishly obeying circumscribed ways of knowing, of making meaning, and of constructing identity) is the power to harm. As captivity, this harm extends to all (the kept, and their keepers, managers, and watchers). Indeed, this is the *maddening* power to harm by way of total confinement, experienced both existentially and materially to, from, and about us all (Arrigo 2012). These harms, as crimes, involve denials of being (i.e., the subject who seeks recovery) and limits on becoming (i.e., the subject who seeks transformation) (Arrigo and Milovanovic 2009).

Harms of reduction occur when an individual's ability *to make a difference productively* is hindered through the actions of some system's agent or representative. For example, a recovering addict may be prevented from pursuing palliative health-enhancing activities because of the ex-user's history of relapse addiction. The person is rendered less than what he or she could be (i.e., as in further restored, recovered, or healed). Harms of repression occur when an individual's ability *to be different dynamically* is hindered by the actions of some system's agent or representative. Here, the subject's transformative possibilities are limited because of one's status as a recovering addict (i.e., rendered finite given one's already marginalized standing) or are circumscribed (i.e., defined by the terms and conditions of others given that one's consequent life already has been decided and determined). This is the powerful realm of harm in which one's societally-enforced and/or self-inscribed text disciplines subjectivity (the individual is forever branded as deviant, diseased, dangerous, or other). Moreover, both forms of harm extend to the manifold incarnations of subjectivity. Indeed, the keepers, managers, and watchers (i.e., the anesthetized general public) also are texts that write their own captivity as subjects, especially when they teach, proscribe, govern, enforce, and/or inter-relate by way of harms of reduction and repression.

Subjectivity/Agency

As noted previously, the modernist perspective asserts that humans are rational. Following this episteme, people are "conscious, whole, self-directing, reflective, unitary, and transparent" subjects (Milovanovic 1997, p. 9). They are purposeful as in criminology's rational choice theory, and centered (i.e., in control, stable) as in law's juridic subject developed in the "reasonable woman or man" standard (Arrigo et al. 2005).

The postmodernist perspective argues that the subject, the acting agent, is de-centered. This is an individual who is open to transitions; fluid and in-process; and whose humanity is more fully expressed and embodied through the chaotic (i.e., orderly disordered) forces of contradiction, inconsistency, spontaneity, and even absurdity. For postmodernists, these aspects of our existences must be extolled rather than confined. In other words, unlike the predictable, linear, and static subject of modernity, post-modernity maintains that our identities are not so easily reducible to finite categorizations or cause-effect rationales. This is how captivity in texts and in systems of thought prevails. To illustrate, when we define individuals through an "either/or" logic and view them as "deviants or conformists; law violators or law abiders; villains or heroes" (Arrigo 1995, p. 454), we fail to comprehend that their (our) more complete subjectivity is located in the *mutuality* that these binaries signify. Restoring this agency (i.e., the recovering subject) and revolutionizing it (i.e., the transforming subject) is the ultramodern journey that awaits all subjects (e.g., Lecercle 1985; Milovanovic 2011).

State of the Art Postmodern Criminology

Drawing upon the concepts established in the first two transitions of postmodern thinking, ultramodern commentators have developed an innovative model for assessing the madness (the captivity) manufactured, reified, and legitimized by way of the criminal justice apparatus.

This is a reference to institutional decision making that dangerously and fearfully supports the technologies of risk/threat avoidance, hypervigilance, and panoptic surveillance, notwithstanding considerable evidence to the contrary on their capacity to grow citizenship and advance justice. Among other examples, three specific institutional practices have been examined in the recent literature and are worth noting: (1) cognitively impaired juveniles waived to the adult system, (2) psychiatrically disordered inmates placed in long-term disciplinary solitary confinement, and (3) formerly incarcerated sexually violent predators placed in civil detention followed by multiple forms of community inspection and reentry monitoring (Arrigo et al. 2011).

In each instance, the precedent-setting and/or prevailing case law was identified and reviewed. These three systems of thought (i.e., texts of mental health law), were then deconstructed at the level of judicial temperament (i.e., attitudes and perceptions) and jurisprudential ethic (i.e., moral philosophy informing juridical disposition). At issue were the ways in which jurists talk about and render decisions affecting troubled, vulnerable, and distressed offenders; those victimized by their actions whether in free society or behind prison walls; and the communities that bind all of these constituencies together. In this qualitative study, the investigators demonstrated how a careful re-reading (i.e., deconstruction) of these mental health law texts could further explain how state-regulated institutions, by way of their respective agents, operate as “totalizing apparatuses” that sustain harms of reduction and repression (Arrigo 2004, p. vii; see also Arrigo et al. 2011). Stated differently, virtue ethics and its promise of excellence for one and about all were not a part of these systems of thought; character did not write these texts. Thus, decision making that spoke of courage, compassion, generosity, mercifulness, etc., for one and about all was altogether lacking in and throughout the case law narratives. However, these virtues, when written about and lived, are the habits of character that help to seed and grow excellence in human/social flourishing (Aristotle 2000; Levinas 2004). The madness of *total* captivity sustains itself when identities are constructed, meaning is made, and subjectivity is categorized in the absence of such mutating self/society excellence (Arrigo 2012). In what follows, the relevance of these observations will be tentatively explored in relation to further developing ultramodern theory, method, and praxis. These comments provide some grounding for conceiving of ultramodern philosophy as an emerging and postmodernist-informed human/social science of criminology.

Developments in Theory

How does the power exercised through harms of reduction and repression operate? In other words, how do these limits and denials unconsciously emerge, take up residence in texts, materially and existentially discipline self/social identities, and guarantee the culturalized replication of each whose co-productive effects reify the normalization of violence? As described previously, four interactive and mutually supporting spheres of influence co-shape the self/society binary. These flows, with their corresponding frequencies and fluctuating intensities, include the symbolic, linguistic, material, and cultural realms. Presently, the interdependent and constitutive effects of these forces resonate as domains of captivity, as conditions of control.

The self/society mutuality consists of the human agent and the structural and/or organizational intensities to which the person is directed and captured and from which the subject co-shapes these flows. The relationship among the four spheres and the twin dynamics (i.e., self/society) is porous in nature. In other words, each force is dependent on and contributes to the composition of the other spheres. Because harms of reduction and repression are existentially and materially ubiquitous, ultramodern theorists argue that their power must be de-stabilized by way of deconstruction. To accomplish this, some further commentary is needed on how these dynamically supporting spheres interact with the twin dynamics such that these flows function as conditions of control reifying and legitimizing captivity for one and all.

The Symbolic sphere involves the consumption of prevailing, incomplete, and circumscribed images (i.e., pictures or scenes in our minds). These images symbolize an aesthetic about and for crime and those who commit acts of violence and victimization. But this inadequate symbolization extends to others as well. Indeed, an aesthetic sensibility also is conjured about victims and how their injury is managed and/or corrected by way of social institutions and their agents. Thus, the space that the symbolic spheres inhabits, is one that commits or assigns us to "the consumption of a particular and dominant aesthetic regarding vulnerable, troubled, and distressed individuals; those professionals whose expertise includes treatment, corrections, and societal reentry; and the interventions exercised to ameliorate offenders and the offended" (Arrigo et al. 2011, p. 164).

Once they are spoken, the shadows that these images cast about the self/society mutuality become an oral history. Eventually, these spoken images become a written text. This is the realm that represents the influencing force of the linguistic sphere on the twin dynamics. Because the image-crafting at the symbolic level lacks an aesthetic that more completely signifies our dynamic and transformative humanity, the story that it tells (for example, about offenders) while one that communicates a type of narrative coherence, can only read as a summary representation (Cicourel 1981; Knorr-Cetina 1981) of, for, and about the incarcerate. As such, the knowledge that is constructed by way of these representations constitutes a false epistemology; it tells a fictional story through its incompleteness. Indeed, these texts are fashioned by way of partial truths about those who injure wrongly, those who are victimized painfully, those who keep and manage both purposefully, and those who watch on hypnotically and docilely (i.e., uncritically). The currency in identity construction and in meaning making regarding these truths is sourced in politics – the politics of rendering silent the texts and images of social selves and social relations that could be dynamically restorative and vibrantly transformative for one and about all. Nurturing and sustaining this lack of experimentation and innovation in storytelling, is how harms of reduction and repression become reified texts.

When the fictional story that is told about crime and criminals becomes a lived narrative, then it is materially embodied as virtuous. This is the influencing force of the material sphere that produces bodies of knowledge or systems of thought whose fluctuating intensities discipline the self/society mutuality. This disciplining extends to the use of productive technologies of control that ensure compliance – whether, for example, in education, medicine, law, or economics. These technologies (e.g., threat/risk assessment protocol, actuarial penology, and "evidence-based" science) endorse renditions of reality, of self/society identities, whose deployment often demands total and slavish-like obedience to these regimes of captivating truth and normalizing regimens of human/social existence. These regimes and regimens dangerously "inscribe the twin dynamics and problematically render. . .silent manifestations of dissent" (Arrigo et al. 2011, p. 165). Here, too, this captivity and normalization extend to and from the kept, as well as their keepers, managers, and watchers.

The fluctuating intensities of the symbolic, linguistic, and material spheres are culturalized by way of their continuously replicated multimedia-based derivatives. The immediacy and reproduction of iterative information, of manufactured reality, by way of the Internet's global village, 24/7 news cycles, Hollywood dramatizations and re-enactments, and even the sites and sounds of Las Vegas' "other worldly" spectacles or Disney World's animatronic simulations, makes identities instantly, although incompletely, accessible. This is because these identities about the self, others, and of situations are nothing more than illusions. These copies of reality, as fictions, are problematic in that they can be interpreted as more authentic and true than the reality from which they are derived (Baudrillard 1983a, b). This *cosmopolitan* influence is the force of the cultural sphere. It imitatively perpetuates and sustains the status quo in image-crafting, text-making, and material disciplining. The cultural sphere is both the source and product of violence normalized, of harms experienced reductively and repressively, and of humanity rendered an ontological shadow of what it could be or could become.

Developments in Method

As method, the ultramodern builds on the deconstructive approach. The postmodernist critique that ensues emphasizes a re-consideration of the text in question. This questioning consists of a re-examination of the reality rendered to which the self/society is directed and from which this mutuality co-produces that reality. This rendering of reality is informed by its constituent conditions of control with their interdependent flows, intensities, fluctuations, and effects. These conditions, as spheres of influence, are prisms through which the disassembling critique is filtered in historically contingent although manifest form. The manifestations of reality, then, are the images, texts, embodiments, and cultural reproductions of each that exercise power existentially and materially. This is the power to set limits on being and impose denials of becoming for one and about all. This harm, as the crime of captivity, is what ultra-modernity seeks to specify and renounce. The criminal justice apparatus is replete with texts whose rendering of reality is both totalizing and maddening. These renditions await ultramodern critique (e.g., Arrigo et al. 2011).

Overcoming the madness reified and legitimized by the systems of thought that are criminal justice entails yet another methodological journey. This excursion rebuilds or builds differently the text in question. This is the activity of reconstruction. At issue are the assemblages of images, texts, embodiments, and cultural reproductions of each that celebrate, both virtuously and excellently, how one can make a difference productively (the recovering subject) and how one can be different dynamically (the transformative subject). Undertaking this reconstructive journey is how the shadows of our selves can be quieted and the captivity of all subjectivities can be quelled. Methodologically, ultramodernists have not yet sufficiently charted the constituents of this passage. To be clear, what is envisioned here is a revolution in, a re-writing and re-reading of, the scientific method. This observation notwithstanding, the criminal justice apparatus is replete with texts awaiting alternative renderings of its realities (e.g., Arrigo et al. 2011; Arrigo 2012). These renditions would be provisional, positional, and relational; otherwise, they, too, would become totalities and this condition is anathema to the evolution in ultramodern philosophy and postmodernist-informed criminological science.

Developments in Praxis

The deconstructive/reconstructive ultramodern method is also relevant at the level of engaged praxis. This is the realm of thinking about and doing habits of character so that the possibilities of vibrantly living virtuously become more realizable for one and about all. In other words, in an era of captivity that engenders harms of totalizing madness, "ideas like "dignity," "the therapeutic," "restorative," "fair-mindedness," "injury," "community," "health," "self," "society," and the like are all filtered through the intemperance that this captivity most assuredly guarantees" (Arrigo et al. 2011, p. 169). The presence of this captivity is how a dynamic praxis is both deferred and denied. However, this far-from-equilibrium activity is the condition that grows a transformative citizenship informed by emergent visions of justice. This is why the symbolic, linguistic, material, and cultural meanings that attach to the *practices* of this self/society mutuality must be dis- and reengaged. This is how ultra-modernity's conception of deconstruction and reconstruction operates at the realm of lived praxis.

To illustrate this perspective, let us consider the following. What images do we conjure (the symbolic realm), what texts do we write (linguistic realm), what oral histories do we live and embody (the material realm), and what imitations of each do we reproduce (cultural realm) when we honor an offender's dignity, heal a victim's injury, and restore a community's conscience in the wake of violence and victimization? Mindful of ultramodern praxis, constructs such as dignity, health, restoration, and community must be critiqued in order to discern the potentially concealed aesthetical, epistemological, ethical, and ontological preferences that reductively and repressively co-produce harm. In essence, then, the ultramodern reminds us that thinking about and doing virtue as dynamic and transformative praxis is an embodied ethic that questions the rendition of reality, of justice, from which such possibilities are advanced and to which they are directed. Indeed, for whom and under what conditions is justice served by any particularized rendering?

Controversies in Postmodern Literature

The proposed developments in ultramodern theory, method, and praxis as constituents of a postmodernist-informed science are controversial in their originality and experimentation. Undertaking this journey – of developing and growing this uncultivated criminology – is an exercise in renovation and innovation for a people yet to come (Deleuze and Guattari 1984, 1987). But transcending the society of captives and overcoming the shadows of our selves is a *strange* expedition. It is an awaiting revolution (Arrigo and Milovanovic 2009). This revolution begins as critique – from the symbolic to the linguistic, from the material to the cultural, and unfolds in historically contingent although manifest form. This critique seeks to restore and transform all those who experience harms of reduction and repression as crime and violence normalized so that they (and all of us) can make a difference productively and become a difference dynamically. This is the critical pedagogy by which captivity, in all of its manifestations, is studied and revolutionized. Each of the above notions – especially in their relationships to ultramodern philosophy – contributes to the contentious postmodern landscape.

Open Questions

What is the nature of authentic human/social interconnectedness in the age of the ultramodern?

This is a quality of citizenship and of social justice whose mutuality remains problematically under-developed. To illustrate, the global village of the Internet and its hyper-real progeny normalize violence (re-imagine, re-tell, and re-inscribe) through limit-setting and denial-imposing meaning-making. This is identity construction that re-conceives such human/social phenomena as "intimacy," "courage," and even "friendship." In the digital age, identity that celebrates the virtual rather than the visceral, the imitative rather than the authenticated, is privileged. Text-messaging, myspacing, facebooking, and skyping are proliferating forms of social networking. But how are these forms of relatedness renditions of reality that replicate human/social citizenship excellently? In other words, what does the ubiquity of this immediately accessible interconnectivity teach us about intimacy, courage, and friendship? If it is found by way of the previously and provisionally enumerated methodology that these digitized encounters are lacking in a mutating authenticity, then overcoming the harms that they seed and nurture (i.e., excessive investments in liquid identities and virtual realities that reduce/repress humanity) is a revolutionary journey that must be deconstructively and reconstructively pursued.

If liquid identities and virtual realities are artifacts of the ultramodern, how do we overcome the death of experience that similarly disappears along with them?

The conditions of a society at its center, tell us a great deal about the nature of behavior at its fringes. The fringe behavior at issue here is the breadth and depth of experience that the power to harm reifies and captures by way of its systems of thought. The tendencies of such captivity are to discipline and to normalize. The maintenance of these sanitizing tendencies or central conditions is the power to harm teleologically. The death of experience alluded to here encompasses the pseudo-realities individuals both construct and encounter, and they consist of their replicas and façades that function to supplant matter and materiality (Arrigo 2012). Overcoming such finalizing of experience requires that we shatter and increase our language so that we can shatter and increase our realities. These activities return us to the excavation work of the unconscious and the images that await symbolization and activation for a people yet to come. Specifying the constituents of this mobilization, of this awaiting revolution, is a strange and uncharted journey of becoming other (Levinas 2004). The activity of becoming other – of being present to and for another always already more courageously, generously, compassionately, mercifully, etc., – is the promise of character. Living this promise excellently is captivity's release. This release is how the retrieval and transformation of experience appears in consciousness, travels through texts, inhabits histories, and culturally manufactures the constitutive reproductions of each.

How do the death of experience and the inauthenticities to and from which liquid identities and virtual realities sustain their mutuality, make possible an experiential conversion (an overcoming) in images, texts, embodiments, and replications?

The possibilities for an experiential conversion that renovates and innovates the human agency – social structure mutuality, are located in the vast reservoir that is the unconscious. The co-productive flows, intensities, frequencies, and fluctuations that shape and are shaped by extant identities and realities, currently write the subject, the social, and their constitutive interdependencies as a text of totalizing confinement. This is a systemic madness because the recursive effects of such captivity impede citizenship's potential (i.e., in its images, texts, and embodied practices). These harm-generating impediments thwart the possibility of conceiving of citizenship as dynamic difference lived ever more virtuously and humanly. Additionally, the effects of this systemic pathology produce circumscribed visions of justice for one and about all whose consequent limits and denials derivatively manufacture mere shadows of justice. However, when we inhabit the space of otherness – dwell joyfully, lovingly, virtuously, and excellently within this strange, uncertain basin – this act of will mobilized to power helps to spark the awaiting revolution for a people yet to come. Occupying this space is much like becoming a jazz player. The cascading conditions, fluctuating intensities, and corresponding frequencies that extol such otherness, such difference, remain a source of future postmodernist-inspired ultramodern inquiry.

Conclusion

The problem of explicating human behavior and social phenomena in ways that respectively grow citizenship and advance justice undoubtedly will persist. One facet of this persistence is in recognizing how the power that harm exercises, reifies crimes of reduction and repression. The modernist approach employs quantitatively-sourced and empirically tested scientific methods to explain and predict criminal and delinquent behavior. According to postmodernists, this criminological approach is less than sufficient and, at worst, inflicts existential and material harms on all those within its fearfully hypervigilant and dangerously panoptic ambit. This captivity is totalizing. It is a madness that limits being and denies becoming for the kept, and their keepers, regulators, and watchers. Sustaining the influencing shadows of this captivity is how the normalization of violence, of crime, prevails. Its co-productive effects extend from the aesthetical to the epistemological, from the ethical to the ontological. A postmodern criminology informed by ultramodern philosophy seeks to overcome the reductive and repressive power to harm by way of restorative and transformative flows, frequencies, fluctuations, and intensities. These influencing forces celebrate spheres of human/social flourishing. Cultivating this ethic is an invitation to further develop theory, method, and praxis. This is the ultramodern challenge of igniting a revolutionary science of criminology.

Related Entries

- ▶ Cultural Criminology
- ▶ Fear and Punishment
- ▶ Feminist Criminological Theory
- ▶ Modern Marxist and Radical Theory

- ▶ Social Control Theory of Sexual Homicide Offending
- ▶ Theories of Punishment

Recommended Reading and References

Aristotle (2000) Nicomachean ethics (trans: Crisp R). Cambridge University Press, Cambridge, UK (Original work composed 4th century, B.C.E)

Arrigo BA (1995) The peripheral core of law and criminology: on postmodern social theory and conceptual integration. Justice Q 12(3):448–472

Arrigo BA (ed) (2004) Psychological jurisprudence: critical explorations in law, crime, and society. SUNY Press, Albany

Arrigo BA (2012) The ultramodern condition: on the phenomenology of the shadow as transgression. Hum Stud J Philos Soc Sci 35(3):429–444

Arrigo BA, Milovanovic D (2009) Revolution in penology: rethinking the society of captives. Rowman & Littlefield, Lanham

Arrigo BA, Milovanovic D (eds) (2010) Postmodernist and post-structuralist theories of crime. Ashgate Press, Farnham, Surrey

Arrigo BA, Williams CR (eds) (2006) Philosophy, crime, and criminology. University of Illinois Press, Urbana/Chicago

Arrigo BA, Milovanovic D, Schehr RC (2005) The French connection in criminology: rediscovering crime, law, and social change. State University of New York Press, Albany

Arrigo BA, Bersot HY, Sellers BG (2011) The ethics of total confinement: a critique of madness, citizenship, and social justice. Oxford University Press, New York

Baudrillard J (1983a) Simulations. Semiotext(e), New York

Baudrillard J (1983b) In the shadow of the silent majorities. Semiotext(e), New York

Best S, Kellner D (1991) Postmodern theory: critical interrogations. Guilford Press, New York

Cicourel AV (1981) Notes on the integration of micro- and macro-levels of analysis. In: Knorr-Cetina K, Cicourel AV (eds) Advances in social theory: towards an integration of micro- and macro-sociologies. Routledge and Kegan Paul, London

Deleuze G (1988) Spinoza: practical philosophy (trans: Hurley R). City Light Books, San Francisco

Deleuze G, Guattari F (1984) Anti-oedipus. University of Minnesota Press, Minneapolis

Deleuze G, Guattari F (1987) A thousand plateaus. University of Minnesota Press, Minneapolis

Derrida J (1970) Structure, sign and play in the discourse of the human sciences. In: Macksey R, Donato E (eds) The structuralist controversy. The Johns Hopkins Press, Baltimore

Derrida J (1981) Positions. University of Chicago Press, Chicago

Einstadter WJ, Henry S (1995) Criminological theory: an analysis of its underlying assumptions. Harcourt Brace, Fort Worth

Foucault M (1977) Discipline and punish: the birth of a prison. Allen Lane, Harmondsworth

Habermas J (1980) Modernity: an unfinished project. In: Passerin M and Benhabib S (eds) *D'Entreves* (trans: Walker N). Polity Press, Cambridge, UK

Henry S, Milovanovic D (1996) Constitutive criminology: beyond postmodernism. Sage, London

Henry S, Milovanovic D (1999) Constitutive criminology at work: applications to crime and justice. State University of New York Press, Albany

Knorr-Cetina K (1981) Introduction: the micro-sociological challenge of macro-sociology: towards a reconstruction of social theory and methodology. In: Knorr-Cetina K, Cicourel A (eds) Advances in social theory and methodology: toward an integration of macro- and micro-sociologies. Routledge and Kegan Paul, London

Lacan J (1977) Ecrits. Norton, New York

Lecercle JJ (1985) Philosophy through the looking glass. Hutchinson, London

Levinas E (2004) Otherwise than being. Duquesne University Press, Pittsburgh

Lyotard JF (1984) The postmodern condition: a report on knowledge. University of Minnesota Press, Minneapolis

Milovanovic D (ed) (1997) Chaos, criminology, and social justice: the new orderly (dis)order. Praeger, Westport

Milovanovic D (2011) Justice-rendering schemas: a typology for forms of justice and a prolegomenon for transformative justice. J Theor Philos Criminol 3(1):1–56

Rosenau PM (1992) Postmodernism and the social sciences: insights, inroads, and intrusions. Princeton University Press, Princeton

Schwartz MD, Friedrichs DO (1994) Postmodern thought and criminological discontent: new metaphors for understanding violence. Criminology 32(2):221–246

Poverty, Inequality, and Area Differences in Crime

Travis C. Pratt and Bryan D. Eisentraut
School of Criminology and Criminal Justice,
Arizona State University, Phoenix, AZ, USA

Overview

American criminological research and theory has predominantly focused on individual offenders. Following the growth and refinement of survey research methods in the 1960s, which had a strict focus on the individual as the unit of analysis, a resurgence of macro-level criminological research has taken place in recent decades. Two of the most commonly assessed ecological correlates of crime rates include poverty (absolute deprivation) and inequality (relative deprivation). This entry takes stock of this body of literature that has examined the effects of poverty and inequality to explain area differences in crime rates.

Introduction

Since its inception, American criminological research and theory has predominantly focused on individual offenders. Criminology grew largely out of the classical school and positivist notions of human behavior in the early 1900s. Both of these perspectives focused on the individual correlates of crime and deviance. While the classical school emphasized the rational decision-making processes of individuals, the positivist school looked for physical and/or psychological explanations of criminal behavior. Despite the different assumptions underlying the classical and positivist schools of criminology, it was their shared focus on the individual that guided much of the subsequent theoretical development in criminology. As such, most of the major criminological theories covered in books, edited volumes, and introductory level textbooks are individual-level theories. The majority of empirical research in criminology also treats the individual as the unit of analysis. Consider, for example, the ubiquity of research employing individuals' responses to self-report scales in academic journals.

In part as a response to criminology's emphasis on the individual, Shaw and McKay's work and the anomie tradition began by Merton directed attention toward macro-level factors in the mid-twentieth century. This is not, of course, to say that the approaches taken by Shaw and McKay and by Merton were merely reactions to the previously articulated individualistic theories of crime. Rather, their work started from a different premise altogether: that human behavior was guided primarily by social conditions and/or structural constraints, as opposed to inherent biological drives and inhibitions. These formulations outlined aggregate-level conditions of geographic regions (e.g., urban characteristics) and sociocultural characteristics (e.g., a society's collective emphasis on material success) that may lead to increased *rates* of crime. But as Bursik and Grasmick (1993, p. ix) note, "with the refinement of survey approaches to data collection and the increased interest in social-psychological theories of control, deterrence, learning, and labeling, the focus of the discipline significantly began to shift from group dynamics to individual processes during the 1960s and 1970s." Even so, beginning in the late 1970s and early 1980s, macro-level (or "ecological") theory and research reemerged and has since earned sustained criminological attention. Indeed, Bursik and Grasmick (1993, p. ix) argue that "the pendulum has begun to swing in the other direction, and there has been a relatively recent acceleration in the number of studies that have been conducted with an explicit focus on [macro-level] dynamics."

At least in part, this resurgence of interest in macro-level approaches has been encouraged by Cohen and Felson's (1979) development of routine activity theory, the seminal work of Blau and Blau (1982) on inequality and violent crime, the rediscovery in the 1980s of Shaw and McKay's social disorganization theory by scholars such as Bursik (1988) and Sampson and Groves (1989), and the renewed interest in deterrence-rational choice theory as well as the rise of conflict theory at the macro level during the 1970s and 1980s (see, Pratt and Cullen 2005).

Two of the most commonly assessed macro-level correlates of crime include poverty (or absolute deprivation) and inequality (or relative deprivation). Numerous tests of theoretical propositions surrounding poverty and inequality have been conducted in the sociological and criminological literature (see Pratt and Cullen 2005) over the last 50 years. The purpose here is to take stock of this body of literature. In particular, this entry begins with a review of the spectrum of theoretical perspectives that have specified the importance of either poverty or inequality in both defining and explaining crime. The discussion then turns to an assessment of the empirical evidence concerning the effects of both poverty and inequality on macro-level crime rates. The entry concludes with an overview of the new directions that have been taken in the poverty-inequality criminological literature.

Poverty and Crime

Poverty and the Definition of Crime

Both social disorganization and anomie/strain theories, for example, point to the structural characteristics that may allow criminal activity to flourish. While these theories differ in their propositions regarding how certain social conditions may produce high crime rates, they are in tacit agreement that "crime," as a social phenomenon, is an objective social reality. In other words, these theories – and most others – implicitly assume that a rough consensus exists within societies about what types of behaviors should, and should not, be considered "criminal."

Thus, neither perspective devotes much attention to how certain behaviors come to be labeled by society as "criminal." Indeed, although the bulk of the theories that are typically covered in criminology textbooks share the contention that groups of individuals are mired in a system of social stratification, they overlook the degree to which power differentials between the relatively powerful and powerless groups in society may shape not only individuals' behaviors, but also the social perceptions of those behaviors. Discussions of the relationship between poverty and crime, which are often guided by conflict theory, attempt to fill this theoretical void.

At its most basic level, conflict theory draws on the Marxian tradition and is most concerned with "[focusing] attention on struggles between individuals and/or groups in terms of power differentials" (Lilly et al. 1995, pp. 132–133). In essence, conflict theory sees crime as a socially constructed label that powerful groups are able to place on the behaviors of groups or individuals who hold, in comparison, less social power and/or political authority. Crime, therefore, is viewed through the lens of conflict theory as being morally relativistic. Since, although both members of the upper and lower social classes may engage in morally questionable behaviors, the legal system – which is assumed by conflict theorists to be a tool wielded only by the upper classes – tends to punish only the deviant activities of the lower classes, particularly those seen as threatening the privileged position of the powerful (Bonger 1916[1969]).

Variations and Critiques of Conflict Theory's Views on Poverty

It is important to note that conflict theory places its iron in multiple fires. Some versions of conflict theory are concerned only with the dynamics of conflict between social groups. For example, early work by Vold (1958) focused on how social groups in conflict tended to maintain an equilibrium that eventually led to social stability and order. Other versions of conflict theory attempt to specify methods for which conflict may be eliminated – such as through trading in the evils of capitalism for the bliss of communism (Marx and Engels 1848[1992]).

The plurality of approaches falling under the heading of conflict theory has, to a certain extent, contributed to its somewhat shaky empirical foundation and to its theoretical murkiness. In particular, many conflict theorists have focused explicitly on "political" crimes (e.g., protest movements, conflicts between labor and management) as indicators of how powerful groups may shape and define substantive law (Vold 1958). Other conflict theorists attempt to marshal in their defense laws against "victimless" crimes, such as "vagrancy," as evidence that the "normal" everyday activities of politically disenfranchised groups of individuals are routinely demonized by the rule-making upper classes (Chambliss 1964). Granted, each of these approaches may be useful for illustrating how "law" may be used as a method of social control. The central problem facing conflict theorists, however, in one of attempting to convince the criminological community that certain violent offenses, such as rape or murder, are socially constructed labels placed on the preferred behaviors of the lower classes (Akers 1997). Certain conflict theorists, however (see, e.g., Reiman 1995), argue that even with regard to violent crimes, members of the upper classes are typically viewed as less responsible for their actions, and tend to receive relatively favorable treatment from the criminal justice system. For example, negligent business practices that result in multiple employee deaths (e.g., mining accidents) are often couched in the language of "liability" instead of "homicide." Reiman (1995) goes on to note how this may be treated as de facto evidence that the wealthy are able to "get away with murder."

The debates surrounding these elements of conflict theory are unlikely to be resolved any time soon. There are, however, other versions of conflict theory that are relatively unconcerned with the more postmodern endeavor of delineating the "definitions" of criminal and noncriminal behavior. Indeed, certain versions of conflict theory attempt to provide an explanation, setting aside "value judgments" regarding the "definition" of crime, as to the causal mechanisms by which social conflict leads to high rates of crime (Bonger 1916[1969]). It is this final variant of conflict theory – which overlaps substantially in its intellectual focus with a host of other ecological perspectives on crime causation (e.g., strain/anomie and social disorganization theories) – that is the focus of the remainder of this section since its primary concern is the *explanation* (and therefore prediction) of how poverty influences crime rates.

Conflict Theory and Crime Causation

In its simplest terms, conflict theory views the "causes" of crime in the following manner. First, a capitalist economic structure is likely to produce a condition of widespread poverty, where large groups of people will experience significant "resource deprivations" in terms of money and property. This absolute deprivation, in turn, heightens and reinforces the animosity harbored by the groups living in conditions of poverty toward those in the leisure class (see, e.g., Veblen 1889[1934]). Finally, the condition of poverty may be criminogenic, by itself, in two ways.

First, poverty may *directly* cause crime among the "subordinate classes" as members of such groups seek daily survival. In other words, certain criminal offenses (e.g., theft) may be necessary for some individuals to simply "get by" (see, e.g., the discussion by Bonger 1916[1969]; see also Lilly et al. 1995). This position is similar to Merton's (1938) notion of deviant/criminal adaptations to strain, where such instrumental offenses would simply be the result of creative ways of engaging in an economy to which those in the lower classes are disenfranchised. Second, poverty may lead to crime *indirectly* as members of the poverty-stricken groups may eventually come to question the validity of the social arrangement they have been handed. Feeling as though they have been given a "raw deal," these groups "would then be more likely to organize and to bring the conflict out in the open, after which there would be polarization and violence" (Lilly et al. 1995, p. 134). A somewhat more nuanced take on this process comes from Shaw and McKay's social disorganization perspective, which holds that conditions of economic deprivation (poverty) provide the context for a conflict of values to take place. In economically-deprived communities, some groups will harbor values in opposition to the law, which may challenge traditional middle-class values and, in turn, may influence the next generation of would-be delinquents through the transmission of such values. Thus, regardless of the theoretical tradition one is drawing upon (e.g., conflict, strain/anomie, social disorganization), they all share the common proposition that impoverishment itself is criminogenic, and that there should be an empirical link between variables which proxy economic conditions, such as poverty rates, and crime rates.

Inequality and Crime

Along with discussions of the criminogenic effects of absolute deprivation, criminologists have also focused on the importance of understanding the consequences of relative deprivation, or inequality. Dating all the way back to

Merton (1938) and continuing up to the present day, inequality has long held the interests of scholars in the field. And in one of the more important works in this tradition, in their landmark study of metropolitan structure and violent crime, Blau and Blau (1982) set forth an explanation as to the "cause" of violent crime that differed from the absolute deprivation/conflict theory model. In short, Blau and Blau (1982, p. 116) noted that the "poverty" model of crime argues that urban slums tend to create a particular subculture where youths value "toughness, smartness, excitement, and fatalism" which, in turn, "bring young persons into contact with the law." Thus, the macro-level perspective on absolute deprivation/conflict theory "interprets delinquency not in terms of individual poverty but in terms of the shared cultural values that tend to develop in the impoverished conditions of urban slums" (Blau and Blau 1982, p. 116). This position is similar to the urban subcultural perspective developed by Fischer (1975), who argued that the population density associated with rapid urbanization increased the likelihood that individuals with values in opposition to the law will run into each other and form their own deviant groups.

Blau and Blau (1982), however, viewed this position as problematic. Specifically, based on Blau's (1977) previously articulated general macrosocial theory, Blau and Blau (1982) held that a number of Marxian perspectives and theories of opportunity are at least implicitly concerned with the effects of economic inequality, or "relative deprivation," on crime rates. For example, early works by Bonger (1916[1969]) and the more contemporary writings of Quinney (1974) both focus on the exploitation of the poor by the rich in terms of property relations and living conditions and on the inevitability of crime as a result of such inequalities. Blau and Blau even noted that Merton highlighted how inequalities – particularly those based on ascribed characteristics (e.g., race/ethnicity) – tend to breed high levels of resentment and anger. Even more explicit is the statement by McDonald (1976, p. 22) that "Inequalities in power, economic or political, were ultimately responsible for the nature of the criminal law established, its enforcement, and the pattern of criminal behavior appearing."

As a first step toward differentiating the effects of inequality on crime rates relative to the effects of poverty, Blau and Blau (1982) analyzed the violent crime rates from a sample of 125 of the largest American SMSAs. Using measures of socioeconomic inequality between racial groups and economic inequality in general, the results suggested considerable support for the inequality/relative deprivation hypothesis. In particular, after controlling for the effects of inequality, poverty – the key variable for absolute deprivation/conflict theory – was not significantly related to total violent crime rates or to disaggregated rates of murder, rape, or assault. While poverty was significantly related to rates of robbery, the magnitude of its effect was substantially weaker than the effects of inequality (the beta weight values for inequality and poverty were .49 and − .30, respectively). Thus, the main theoretical proposition being made by Blau and Blau (1982, p. 126) is that "aggressive acts of violence seem to result not so much from lack of advantages as from being taken advantage of, not from absolute but from relative deprivation."

Theoretically, then, there is a plurality of reasons that have been set forth as to why poverty and/or inequality should be related to crime rates at the macro level. While some perspectives focus explicitly on either poverty or inequality, the majority of these theories appear to recognize the importance of both absolute and relative levels of economic deprivation when it comes to explaining crime. And though there is wide variation in the issue of why either poverty or inequality should "matter," nearly all of the criminological theories addressing these issues highlight how such ecological conditions provide fertile ground for anger and frustration on the part of the citizens. And when citizens are angry and frustrated, bad things tend to happen, including crime (Agnew 1999).

The Empirical Evidence

Poverty and the Definition of Crime/Enforcement of the Criminal Law

Although purely economic variables tend to be the preferred vehicles of poverty-related research, other relationships have been specified by within this tradition. For example, much of this work does not treat "crime rates" as the dependent variable. Rather, a number of studies grounded in the poverty-crime literature have examined the "threat hypothesis" of crime control (Liska et al. 1981). In this body of work, researchers have drawn on the proposition that impoverished members of racial and economic minority groups are viewed as a threat by the dominant groups in a social system. A corollary assumption is that the "powerful groups and strata are able to translate their perceptions of threat into public policy and thereby affect the size and administration of crime control apparatuses" (Chamlin 1989, p. 355). Typically, these studies examine the effect of the relative size of a social aggregate's poor racial minority population on its capacity to provide crime control – often through police expenditures (see, Pratt and Cullen 2005).

Chamlin (1989) took this perspective a step further and examined the relationship between "threat hypothesis" variables and a particular type of crime – rates of police killings. Using state-level data, four measures were constructed to measure the presence of threatening groups: the percentage of families below the poverty level, income inequality (measured as the Gini index of income concentration), the percentage of blacks, and the percentage of individuals with Spanish surnames. After controlling for arrest rates (for index offenses), total index crime rates, police size, and the divorce rate, the threat hypothesis variables were consistently related to rates of police killings. Further, it was not uncommon for the threat hypothesis variables to be the strongest predictors of police killings in the regression models.

Poverty as a Cause of Crime

Aside from this work on the racial threat hypothesis, most of the empirical studies within this tradition have examined the effects of absolute economic deprivation variables, such as "poverty rates," on crime rates. The measures of "poverty rates" across studies are fairly consistent: The percent below the poverty threshold as defined by the Social Security Administration – which has been adopted by the US Census Bureau – is typically used. Other approaches have, however, been taken. For example, Loftin and Hill (1974) used a "structural poverty index" (SPI) that is based on aggregate indicators such as the percentage of children living with one parent, and the percentage of the population failing the Armed Forces Mental Test. Also, Messner and South (1986) used an annual household income of $4,000 as a cutoff point for "absolute poverty." This estimate closely resembled the official poverty level for a family of four ($3,698) during the time of their data collection. Other methodological characteristics, however, exhibit considerable variation across studies. For example, studies of poverty and crime have been conducted at the neighborhood level, at the city level, on standard metropolitan statistical areas (SMSAs), and on countries (Pratt and Cullen 2005).

Studies also differ in terms of the types of offenses comprising the dependent variable. Studies of poverty and crime have attempted to predict general rates of violent crime, homicide rates alone, or multiple rates of different types of violent and/or property crime rates. Overall, Pratt and Cullen's (2005) meta-analysis of the macro-level criminological literature indicated that the effect of poverty on crime has been well tested, with over 200 studies examining the effect of poverty on crime. The aggregated mean "effect size" (or predictive capacity) of poverty variables on crime is both strong and robust across a number of different methodological specifications.

Inequality and Crime

Similar to the body of empirical work on the relationship between poverty and crime, studies of the relationship between absolute deprivation/ inequality variables and crime rates differ considerably in terms of the methodological approaches taken since the study conducted by Blau and Blau (1982). In particular, researchers have employed a number of different measures of the central concept of "inequality." General inequality is typically measured by the Gini coefficient for family income. This measure is generally computed for heads of households that are families and is based on the combined income of all family members (see Blau and Blau 1982). The Gini index includes wages or salary, income from self-employment, social security, public assistance funds, retirement benefits/pensions, and all other types of income (e.g., stock dividends, interest).

Even so, studies vary as to whether racially homogeneous or heterogeneous measures of inequality are investigated. This may pose a potential problem, some argue, since racial inequality, as opposed to inequality in general, may be the chief predictor of rates of urban violence (see, e.g., the discussions by Blau and Blau 1982; Sampson and Wilson 1995). Accordingly, researchers have also argued that racially disaggregated crime rates should be used in studies of inequality and crime, since the social and economic histories of different racial groups (e.g., blacks and whites) may have conditioned the relationship between inequality and crime (Messner 1983; Messner and Golden 1992).

Studies of the inequality-crime relationship have also focused on different levels of aggregation. Analyses have been conducted at the neighborhood level, at the city level, using SMSAs as the unit of analysis, at the state level, and at the national level. Studies of inequality and crime also vary in terms of what types of crime rates are being predicted. For example, some studies of the inequality-crime link focus on rates of violent crime in general, homicide rates, or rates of individual types of violent and/or property crime rates. In assessing all of this literature, Pratt and Cullen's (2005) analysis revealed that, similar to its absolute deprivation cousin, the relationship between relative deprivation/inequality and crime has also been well tested empirically. The overall mean effect size estimate for inequality is also quite strong, which places it in the top tier of predictive strength among all macro-level predictors of crime.

Conclusions

It is clear that both poverty and inequality are strong predictors of crime rates across units of analysis. Of course, some scholars contend that separating poverty and inequality amounts to creating a false dichotomy (Land et al. 1990), since both constructs tap into the broader issue of resource deprivation. The bulk of the empirical evidence, however, reveals that across units of analysis (from neighborhoods to nations) and across different indicators of crime rates (from various violent to property offenses), poverty and inequality seem to exert independent effects on crime (see Pratt and Cullen 2005). It is equally possible, however, that poverty and inequality operate in an interactive way. For example, it could be the case that, when it comes to predicting crime, inequality matters more when conditions of poverty are also present – that is, perhaps inequality is not so criminogenic if those on the bottom are doing alright in an absolute sense. Either way, it remains to be seen how future empirical work addresses this question across various units of analysis and across various crime measures – something that has yet to be addressed meaningfully in the criminological literature.

Even so, macro-level concepts such as poverty and inequality are now being extended into new areas of criminological research. In particular, studies of offender recidivism are now being couched in the both the language and theory of poverty and inequality (Reisig et al. 2007), as is work regarding the "institutional efficacy" of correctional treatment programs (Wright et al. 2012). These are good starts, but there are other directions for future work in this tradition that are equally important and have thus far been all but ignored by criminologists. For example, it would be incredibly useful theoretically for future empirical work to address the intervening mechanisms that operate between poverty, inequality, and crime. Indeed, studies that have included direct measures of these intervening processes in macro-level research are extremely rare in the literature (see Sampson and Groves 1989; see also Lowenkamp et al. 2003), and if work in this vein were to continue, many of the debates concerning *why* poverty and/or inequality should or should not matter could be resolved. In addition, future work should focus on how other institutional arrangements – such as state welfare programs, reentry services for released prison inmates, and investments in public education – may ameliorate the potentially harmful social consequences of poverty and inequality. All in all, poverty and inequality have a strong presence in the criminological literature that shows no signs of slowing down.

Related Entries

- ▶ Defining Disorder
- ▶ Disadvantage, Disorganization and Crime
- ▶ Discriminant Validity of Disorder and Crime
- ▶ Disorder: Observational and Perceptual Measures
- ▶ Early Chicago School Theory
- ▶ General Strain Theory
- ▶ Informal Social Control
- ▶ Social Control
- ▶ Social Disorder and Physical Disorder at Places

Recommended Reading and References

Agnew R (1999) A general strain theory of community differences in crime rates. J Res Crime Delinq 36:123–155

Akers RL (1997) Criminological theories: introduction and evaluation, 2nd edn. Roxbury, Los Angeles

Blau PM (1977) Inequality and heterogeneity. The Free Press, New York

Blau PM, Blau JR (1982) The cost of inequality: metropolitan structure and violent crime. Am Sociol Rev 47:114–29

Bonger W (1916[1969]) Criminality and economic conditions. Indiana University Press, Bloomington

Bursik RJ (1988) Social disorganization and theories of crime and delinquency: problems and prospects. Criminology 26:519–51

Bursik RJ, Grasmick HG (1993) Economic deprivation and neighborhood crime rates, 1960–1980. Law Soc Rev 27:263–83

Chambliss WJ (1964) A sociological analysis of the law of vagrancy. Soc Probl 12:67–77

Chamlin MB (1989) Conflict theory and police killings. Deviant Behavior 10:353–68

Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activities approach. Am Sociol Rev 44:588–608

Fischer CS (1975) Toward a subcultural theory of urbanism. Am J Sociol 80:1319–1341

Land KC, McCall PL, Cohen LE (1990) Structural covariates of homicide rates: are there any invariances across time and space? Am J Sociol 95:922–63

Lilly JR, Cullen FT, Ball RA (1995) Criminological theory: context and consequences, 2nd edn. Sage, Thousand Oaks

Liska AE, Lawrence JJ, Benson M (1981) Perspectives on the legal order: the capacity of social control. Am J Sociol 87:413–26

Loftin C, Hill RH (1974) Regional subculture and homicide: an examination of the Gastil-Hackney thesis. Am Sociol Rev 39:714–24

Lowenkamp CT, Cullen FT, Pratt TC (2003) Replicating Sampson and Groves's test of social disorganization theory: revisiting a criminological classic. J Res Crime Delinq 40:351–373

Marx K, Engels F (1848[1992]) Communist manifesto. Bantam, New York

McDonald L (1976) The sociology of law and order. Westview Press, Boulder

Merton RK (1938) Social structure and anomie. Am Sociol Rev 3:672–682

Messner SF (1983) Regional and racial effects on the urban homicide rate: the subculture of violence revisited. Am J Sociol 88:997–1007

Messner SF, Golden RM (1992) Racial inequality and racially disaggregated homicide rates: an assessment of alternative theoretical explanations. Criminology 30:421–47

Messner SF, South SJ (1986) Economic deprivation, opportunity structure, and robbery victimization: intra- and interracial patterns. Soc Forces 64:975–91

Pratt TC, Cullen FT (2005) Assessing macro-level predictors and theories of crime: a meta-analysis. In: Tonry MH (ed) Crime and justice: a review of research, vol 32. University of Chicago Press, Chicago

Quinney R (1974) Critique of the legal order: crime control in a capitalist society. Little, Brown, Boston

Reiman JH (1995) Rich get richer and the poor get prison: Ideology, crime, and criminal justice. Allyn and Bacon, Needham Heights

Reisig MI, Bales WD, Hay C, Wang X (2007) The effect of racial inequality on black male recidivism. Justice Q 24:408–34

Sampson RJ, Groves WB (1989) Community structure and crime: testing social disorganization theory. Am J Sociol 94:774–802

Sampson RJ, Wilson WJ (1995) Toward a theory of race, crime, and urban inequality. In: Hagan J, Peterson RD (eds) Crime and inequality. Stanford University Press, Stanford

Veblen T (1889[1934]) The Theory of the leisure class. Macmillan, New York

Vold GB (1958) Theoretical criminology, 3rd edn. Oxford University Press, New York

Wright KA, Pratt TC, Lowenkamp CT, Latessa EJ (2012) The importance of ecological context for correctional rehabilitation: understanding the micro- and macro-level dimensions of successful offender treatment. Justice Q 29:775–798

Prediction and Crime Clusters

Graham Farrell[1] and Ken Pease[2]
[1]Institute for Canadian Urban Research Studies and School of Criminology, Simon Fraser University, Burnaby, BC, Canada
[2]Jill Dando Institute, University College London, London, UK

Synonyms

Crime concentration; Crime spates; Crime sprees; Hot dots; Hot products; Hot smudges; Hot spots; Hot targets; Near repeats; Repeat offending; Repeat victimization; Repeats; Risky facilities; Risky professions; Risky routes; Virtual repeats

Overview

Predicting crime is a necessary condition for its prevention, and crime is most predictable along those dimensions in which it is concentrated. The most established forms of crime's tendency to concentrate or cluster are repeat offending, repeat victimization, and geographical hot spots, with complementary concepts including supertargets, hot products, hot places, hot targets, risky facilities, risky routes, and crime sprees and spates. This entry charts the relationship between such clusters, observing how a broad conception of "near repeats," incorporating crimes with similar situations and characteristics, is a useful unifying concept. Metrics of nearness, or conversely the difference between crime events, may inform efforts to understand and disperse crime clusters. The theories of repeat victimization and hot spots are shown to be overlapping and compatible and suggest a unified theory of clusters should result from greater conceptual integration in this field. The overall aim of such integration should be to inform cluster-busting crime prevention efforts.

Definitions and Concepts

In what follows the term "crime clusters" is used to refer to crime's tendency to concentrate in time, space, and other dimensions along which it is measured. A range of related concepts and terms has emerged. *Repeat victimization* can be thought of as crime against the same target, however defined. That can be the same person, household, business, vehicle, place, or other target. Crime follows crime such that a small proportion of targets and places experience a vastly disproportionate amount. One study found that 1 % of people experience 59 % of personal crime including violence and that 2 % of households experience 41 % of property crime (Pease 1998). This is largely responsible for spatial clustering wherein a small percentage of *hot spot* locations accounted for a significant proportion of all calls for police service (Sherman et al. 1989; Andresen and Malleson 2011). The most chronically victimized targets are termed *supertargets*, with many studies finding two or so percent of potential targets and locations accounting for around half or more of crime. Hot spots may or may not be synonymous with *high-crime areas*, depending on geographical scales, which are also disproportionately composed of repeats.

A common feature of retail crime and personal theft is the *hot product*. These are typically high-value portable electronic goods including – at the time of writing – smartphones, laptops, and SatNav-GPS systems, or low-value items easy to dispose of to unscrupulous retail outlets, such as razor blades or cosmetics. Locations which host crimes of the same type – for example, if pharmacies suffer more robberies than other stores – have been termed *hot targets* (Velasco and Boba 2000), and those hosting multiple crimes of the same and different types have been explored as *hot places* (Block and Block 1994), *crime generators* and *crime attractors* (Brantingham and Brantingham 1995), and *risky facilities* (Eck et al. 2007). This can include bars, retail establishments, and land use of different types. Banks are risky facilities with, crucially for the present discussion, some chains and some branches experiencing many repeat robberies, and a few schools are victimized far more than others (Lindstrom 1997).

Crime is common for passengers on some bus routes (Tompson et al. 2009). The term "hot routes" is used in the urban planning literature to refer to roads with high traffic flow (Li et al. 2007), so perhaps *risky routes* distinguishes those with disproportionate crime. The *risky professions* tend to be those that provide care or services to the public such as nurses, customer services staff, and emergency services staff. Some taxi drivers are unusually likely to be robbed (Smith 2005), while firefighters, police, and paramedics often find themselves in the firing line. Within those professions, crime is concentrated on a small percentage of oft victimized staff who experience a lot of the crime. One study of police officers found that around half were assaulted but 3.5 % experienced a quarter of all assaults – likely a conservative estimate. But such statements can be frustratingly unnuanced, in that one needs to know, within the general observation, the times, circumstances, and personalities around which events cluster.

Evolving from work on repeat victimization, the term *near repeats* (Morgan 2001) is used to refer to similar crimes occurring a short time and distance from an earlier crime. The phrase near repeats has mostly been used as synonymous with spatiotemporal proximity. Neighboring and nearby households are more likely to be victimized after a break-in (Townsley et al. 2003; Johnson et al. 2007a, b). Ratcliffe and Rengert (2008) found repeat shootings in Philadelphia were more likely within a city block and two weeks. To return to an earlier issue, repeats and near repeats suggest why crime clusters spatially. Levy and Tartaro note:

> In 1996, repeat locations became known as hot dots (Pease & Laycock, 1996; Townsley, Homel, & Chaseling, 2000). "Hot dots" are locations within a hot spot that are known to have a high incidence of victimization – a repeat victimization location. (Levy and Tartaro 2010, p. 300)

Studying car theft, Levy and Tartaro use the concept of repeat victimization locations as a useful way of examining a combination of repeats and near repeats, hot dots and hot spots:

> Repeat victimization locations are defined as any street segment on which more than one auto theft occurred during the study period. In Atlantic City the average length of a city block is approximately 150 ft. (Levy and Tartaro 2010, pp. 304–305)

The same offenders are more likely to be those committing repeats, and such returners are more likely to be prolific (Everson 2003). This implies that a crime *spree* (from the offender's perspective) or crime *spate* (from the victim's perspective), typically defined as more than two similar crimes in a short time period, is usually composed of quick and near repeats. The conceptualization of near repeats is developed further later in this entry.

Most of the clustering concepts referred to so far relate to crime recurring over relatively short periods of time. Yet crime is also common against certain targets over a longer period. Averdijk (2010) explored the area of victimization over the life course, or "victim careers," identifying a clear need for further research. Similar terminology had previously been developed in relation to the criminal careers of places (see chapters in Eck and Weisburd 1995). In a similar vein, the onset, frequency, duration, and decline in the criminal careers of stolen products have been explored (Mailley et al. 2008). The trajectory of a product theft career depends on the extent to which it is *craved*, that is, the extent to which it is concealable, removable, available, valuable, enjoyable, and disposable (Clarke 1999; Wellsmith and Burrell 2005). Heavy valueless items have little if any theft career.

Hence, the study of crime clusters, concentration, or repeats has evolved on a piecemeal basis.

One result is the diverse lexicon of sometimes overlapping terms and concepts described so far. Naming specific clusters is useful because it provides the handles with which to grasp how crime against certain targets, times, and places can be predicted and thereby prevented. Yet it is clear that there are many common threads emerging that are naturally leading the field toward increased conceptual and theoretical integration, and the bulk of the remainder of this entry outlines how this is occurring.

Near Repeats as a Broad Inclusive Concept

The concept of near repeats has been discussed so far as referring primarily to *spatial* and *temporal* proximity. However, the concept of nearness can be applied more widely. Spatial proximity is perhaps the easiest form of nearness of which to conceive, but a crime may also be a near repeat in terms of any quantifiable characteristic or combination of quantifiable characteristics. Quantification here includes classification. For example, hot product theft is a form of near repeat along the dimension of product similarity. Drive offs from service stations may reveal a form of near repeat in terms of forecourt layout or oil company concerns. A few professions are disproportionately targeted, and within those a few individuals with common characteristics (such as the particular population or area they serve) may be disproportionately victimized. Linking profession and time of victimization yields, for example, insights about the victimization of emergency room nurses in the late evening. Near repeats may be inferred from similarities between types of recreational establishment or location, staffing levels, and client surveillability. Route characteristics in public transport will reveal forms of near repeat either directly or in interaction with time of day and day of week.

Two further points about the near repeats terminology are warranted. First, near repeat provides a useful and distinct term because it embodies the *mechanism* by which crime is patterned, namely, their nearness along at least one dimension. Nearness may vary, and there are degrees of nearness because some crimes are more similar than others, an issue returned to later. However, a broad concept of near repeats includes a range of types of crime cluster where similarity between event, participants, and contexts is key. Second, it could be suggested that terms such as “hot” or “risky” are equally good as near repeats. They are catchy terms and this can be a surprisingly important component of academic concepts and theories – think of Buckyballs or of chaos theory, the latter of which was frequently contested (because the name inaccurately represents the field) until its popularity brought credibility. However, unlike the term near repeats, they do not embody the notion of why repetition occurs. Near repeats seems a preferable general term because it embodies the notion of similarity and difference between crimes across any dimension or combination of dimensions. This is critical because it is similarity that defines all forms of crime cluster. Defining how concentration is distinguished from a non-concentration of crime is avoided here, but a simple definition as a rule of thumb would be two standard deviations from the mean.

One way of thinking about the distinctiveness of near repeats is to consider it as a reflection of offender targeting practices, summed across offenders. Target selection will, in the aggregate, be represented in the pattern of near repeats. Locally, departures from the aggregate pattern will reflect the signature of prolific individuals.

Mapping the Relationships

This section describes the matrix of crime clusters shown as Table 1 which builds on Farrell (2005) and Farrell and Pease (2008). Following routine activity theory, the columns relate to targets, locations, or offenders. The rows denote four dimensions along which individual crimes cluster, described further below: spatial unit, time, crime type, and modus operandi. The matrix is a heuristic device that allows different types of crime cluster to be viewed simultaneously in terms of their key characteristics,

Prediction and Crime Clusters, Table 1 Matrix of crime clusters

	Target	Location	Offender
Spatial repeat	Same target in same or nearby place. *Hot dots*	Near repeat at same/nearby place (not necessarily same offender or crime type). Includes *repeat victimization locations*, *attractors*, *generators*, *hot smudges and hot spots*, *risky facilities*, *risky routes*	Same offender reoffending at same/similar place (not necessarily same crime type). Linked to *repeat offending*
Temporal repeat	Quick repeat against same target, e.g., much domestic violence; *risky professions*. Longer term: *victim careers*	Quick repeat at same or nearby place (not necessarily same offender), e.g., looting of a prone store. Longer term: *criminal careers of places*	Quick repeat by same offender (not necessarily same crime type or place) – a *spree* or *spate*. Longer term: *criminal careers*
Crime-type repeat	Same crime type against same targets, e.g., robbery in same location	The same crime type in the same or similar place, e.g., continued drug dealing in local area	Same offender commits same crime type, e.g., identity theft from different persons. Linked to *serial offending* and *offence specialization*
Tactical repeat	*Virtual repeat* using same tactic (modus operandi) leads to *hot products*	Same tactic facilitated by same place, e.g., theft and pickpocket at a street market	Repeat offending using same *modus operandi*; linked to *offence specialization*

though of course it is a simplification of reality because in practice many of the cells overlap and some of the possible characterizations of near repeat clusters are excluded.

The first column of Table 1 identifies crime against the same or similar (including perceived similar) targets. Such crime may be in the same or a nearby place, but not all repeat targets are spatially alike: Consider a victim who is robbed, harassed, and assaulted at different places, perhaps on the basis of ethnicity, sexual orientation, ostentatious wealth, or perceived vulnerability. The cell in the second row of the first column identifies the role of time. Many repeats and near repeats occur quickly, particularly when committed by the same offenders. Continuing down the first column, many repeats are of the same crime type. The row for tactic or modus operandi identifies the fact that crime is often repeated using the same method, particular quick repeats of the same crime type by the same offender, as when a burglar reenters a home through the same rear patio door.

In practice, crime is usually common along many dimensions. The limiting case here is crime against the same target in the same place shortly after the previous crime and committed by the same offender using the same tools and tactics. In many instances, however, nearness is greater in some dimensions than others. Crimes concentrated upon the same victims and product types but not necessarily temporally concentrated have been noted already in the context of the criminal careers of places, victims, and products.

Just as the definition of target was broad, the second column of Table 1 defines repeats at the same location, one nearby or with similar characteristics. The term location or place is scalable. It could cover a region, a country, city, neighborhood, building, apartment, street corner or intersection, or other precise geographical or virtual coordinate. Consequently, the intersection with spatial repeats (the top cell in the second column) reflects the importance of possible variation in spatial unit. A street or neighborhood which hosts a cluster of burglaries, or a shopping mall suffering a cluster of robberies, may be the same location for many crime reduction purposes.

Within those locations, however, there are specific targets where crime is repeated – particular houses, retail outlets, street corners, or alleyways. Levy and Tartaro are worth quoting at length in relation to unit of analysis:

> Other aspects of the event should be studied to better understand the repeat victimization phenomenon. One such concept is the "unit of analysis" for the initial criminal event. As the unit of analysis changes, the way in which police and researchers view the crime may also change. Using auto theft as an example, if the unit of analysis is the car, and the car is stolen and dismantled, there is no chance of that car being stolen again. However, if the unit of analysis is the location from which the car is stolen, the fate of the first car has little to do with whether another car could be stolen from that lot. In this situation, there is a chance for repeat victimization, another car could be stolen from that lot and the lot would then be victimized twice. Similarly, the location may be repeatedly victimized, but the owner may not be. If the location is a parking lot, many cars can be stolen from the lot, but they may not belong to the same owner. An owner can be yet another type of repeat victim. The owner could have a different car stolen at different points in time, from the same or different locations. Though the same car has not been victimized again, the owner has still been victimized repeatedly. (Levy and Tartaro 2010, p. 301)

It may be necessary to consider other unit types for analysis. The recommended practice for police seeking to prevent repeat victimization has long been to use that unit of analysis which worked best to facilitate prevention, although that may be easier with hindsight.

It is also necessary to consider the composition and nature of exact and near repeats, as these may inform how crime prevention resources are best allocated. Hence, Table 1 includes concepts relating to attractors and generators, hot spots, and risky facilities. Spatially defined repeats may also vary in the time to repetition, in the type of crime and the modus operandi, or some combination of these relating to the criminal career of the place.

Column three of Table 1 refers to repeats by the same offenders. Progressing down the column, these may be at the same or a similar location, may recur quickly, and may be the same type of offence committed by the same tactic. The same offender committing further crimes shortly afterward could result in a *spree* or *spate*, usually defined as more than two crimes in quick succession, and which may or may not involve the same crime types and tactics. However, crimes of the same type may be indicative of offence specialization. It should be noted that using offender characteristics to identify commonality is only possible after detection. The more usual causal direction involves taking event characteristics to suggest likely offenders.

The matrix provides an overview of a few of the key dimensions of the various clusters and helps clarify their relationships. Since patterning is a necessary condition of nonrandom prediction, the matrix may well provide information and insight that assists in the development of cluster-busting preventive responses.

Measuring the Relationships

In offender profiling, behavioral linkage analysis (BHA) seeks to determine whether, in the absence of physical evidence, unsolved crimes were committed by the same serial offender. For example, with a burglary where there is no DNA or other evidence, the behavioral characteristics of the crime (e.g., type of property targeted, means of access and egress, type of items stolen, the presence of gratuitous mess) are used to develop a behavioral profile. The characteristics of different crimes can then be compared to determine, within parameters, whether they may plausibly have been committed by the same offender. There are multiple comparison techniques which use metrics of the goodness of a match between case characteristics. Some crimes are more similar than others. Two of the more prominent measures are Jaccard's similarity coefficient and the taxonomic similarity index. Thus, behavioral linkage analysis has been used to determine whether, in the absence of physical evidence, unsolved crimes were committed by the same offender, based on across-crime similarity coefficients (Melnyk et al. 2010).

BHA is predicated on two assumptions (Canter 2004). The first is behavioral stability – the assumption that offenders are reasonably consistent in their offending so that a series of their crimes has common characteristics. The second is behavioral distinctiveness – the notion that each offender's crime series embodies particular characteristics that effectively serve as their behavioral signature.

The application of linkage analysis and similarity coefficients to a crime *event* profiling, rather than *offender* profiling, sits well with evolving approaches to crime cluster analysis. Crime clusters can be linked by factors other than the offender. They have other similar characteristics and similar underlying causal mechanisms relating to the opportunity structure (e.g., the characteristics of frequently stolen products). Crime event linkage analysis is thus predicated on the supposition that similarity across crimes is a key factor underpinning clusters of crime irrespective of the dimensions across which common features are held. The orientation of such analysis should be to inform crime understanding and consequent prevention efforts. That could involve the detection of serial offenders but seems more likely to involve designing out crime.

Most concepts relating to crime clusters hinge on the similarity between criminal events. This can be gauged via an index of similarity. Consider how crimes differ. One crime is only perfectly identical to itself. Thus, "identical" is the theoretic maximum level of commonality. Two entirely unrelated crimes of, say, different types in different countries and eras are in effect completely dissimilar (other than that they both involve human behavior and breach of a criminal code). This theoretically maximum difference defines the theoretically minimum level of similarity. All other criminal events lie on a spectrum somewhere in-between.

Quick repeat burglaries by the same offender against the same household are very similar crimes, as is repeated domestic violence involving the same partners, household, and situation. They are sufficiently similar to have been termed "exact" repeats (Summers 2010), differing primarily in time of occurrence. Likewise, simultaneous attacks on two computer networks by the same hacker using the same modus operandi differ primarily in the geographical location of the target. Hence, they are also exact repeats insofar as they differ primarily in one dimension (of course, the violence may result in different injuries, and different goods may be stolen in repeat burglaries, but they are assumed similar for present purposes).

One step further away is a spatially near repeat burglary of a neighboring household shortly after a burglary. This differs in two elements – time and target – from the prior burglary, although only by a relatively small amount in each case. Other forms of crime commonality are typically very similar or "near" in relation to other characteristics. Frequently stolen hot products such as some types of smartphones, laptops, and GPS-SatNavs may be stolen by the same offenders using the same modus operandi (such as sneak theft) but differ in terms of time and place. Racist attacks against members of the same ethnic group by the same perpetrator have different specific targets. However such targets are similar in key characteristics whether that is portability and high resale value or perceived ethnicity. For present purposes, the target is the same and they are forms of near repeat victimization. Whereas the near repeat burglary was defined by spatial proximity, the near repeat theft or assault is defined by the nearness of interchangeability of the target. That is, as noted earlier, the concept of "nearness" is not confined to the spatial and temporal variables. Here the broad inclusive definition of near repeats is important because it demonstrates that it is the degree of similarity between crimes that is the key to crime clusters. For present purposes, spatially near repeat burglaries might be considered equally similar as the targeting of hot products which are near repeat thefts. A hot spot has crimes that are similar in spatial terms but perhaps in little else. This means a hot spot warrants closer scrutiny to determine if it is composed of repeat victimization of the same target or victimization of different targets and crime types. The important point, however, is the following: Similarity is the key to prediction and informed crime prevention. The greater the commonality across crimes, the potentially more informed the response and the greater the preventive scope. The predictability of a repeat crime depends on the type of crime and the context. Hence, quick repeats of the same crime type against the same target, by the same offenders using the same tactic, present the greatest potential for crime prevention. Such precise repeats should be easiest to prevent because the maximum amount of information is available. Other forms of crime commonality are derivatives with less similarity or, conversely, a higher index of dissimilarity.

The appropriate terminology is likely to be that of *difference* rather than similarity. Difference might be preferred when metrics are developed because it allows lower values to refer to more similar crimes – that is, it is more appropriate to refer to a *low* difference score. Similar indices exist in various fields including the Hamming distance, and the Sorensen Index and various metrics for DNA fingerprint matches. Critical elements of crime event similarity are the *target*, the *location*, the *time*, the crime *type*, the *modus operandi*, and the *offender*. Measures of the extent of difference *within* each variable require more sophisticated indices. However, there is a likely benefit to simplicity in a crime event difference index as it would be useful primarily to the extent it served as a heuristic device for crime prevention.

Toward a Crime Cluster Theory

Theories of repeat victimization and hot spots emerged around the same time and from a similar epidemiological tradition. As key areas of crime commonality where theory is established, if they prove compatible, then this may provide a platform for further integration of both theory and crime control practice. Hence, while recognizing the need for a wider repertoire of variables to be considered in understanding crime patterns, this section of the entry reverts to the more familiar ones because they are the most established and readily comprehended.

Cohen and Felson, in the landmark study that developed routine activity theory, anticipated more recent theories of crime clustering when they noted

> [T]he effects of the convergence in time and space of these elements [suitable targets, likely offenders, and the absence of capable guardianship] may be multiplicative rather than additive. That is, their convergence by a fixed percentage may produce increases in crime rates far greater than that fixed percentage. (Cohen and Felson 1979, p. 604)

There are two main theories of hot spots. These are embodied in the concepts of crime attractors and generators. Generators are places with high flows of people which yield spatial concentrations of crime, even though one crime may be no more related to another than elsewhere. Attractors are places that gain a reputation for crime and thereby attract likely offenders. The same place can be a generator and an attractor, because a generator is likely to attract would-be offenders due to the rich supply of possible targets. Brantingham and Brantingham (1995, pp. 7–8) observe that

> Crime generators are particular areas to which large numbers of people are attracted for reasons unrelated to any particular level of criminal motivation they might have or to any particular crime they might end up committing. Typical examples might include shopping precincts; entertainment districts; office concentrations; or sports stadiums. . . . Crime generators produce crime by creating particular times and places that provide appropriate concentrations of people and other targets (Angel, 1968) in settings that are conducive to particular types of criminal acts. Mixed into the people gathered at generator locations are some potential offenders with sufficient general levels of criminal motivation that although they did not come to the area with the explicit intent of doing a crime, they notice and exploit criminal opportunities as presented (either immediately or on a subsequent occasion).

There is much in common here with Cohen and Felson's notion of multiplicative interaction effects. Further, the Brantinghams state that

> Crime attractors are particular places, areas, neighbourhoods, districts which create well-known criminal opportunities to which strongly motivated, intending criminal offenders are attracted because of the known opportunities for particular types of crime. Examples might include bar districts; prostitution areas; drug markets; large shopping malls, particularly those near major public transit exchanges; large, insecure parking lots in business or commercial areas. The intending offender goes to rough bars looking for fights or other kinds of 'action.'

These definitions are consistent with those in the more recent work of Kinney et al. (2008). With respect to repeat victimization, Farrell and Pease (1993, p. 13) offered three theories. The first is that

> It is possible that a first victimisation does not increase the probability of repeat victimisation, but merely flags the high prior probability of victimisation, which is attested again by its suffering a repeat burglary.

which has become known as the risk heterogeneity or flag theory. The second is that

> The same offenders may return to take things they had forgotten the first time, or for which they now have fencing opportunities. Since they know house layout and exit points, problems are less numerous than for the first offence.

which has become known as the event dependency or boost theory. A successful crime boosts the chances of repetition, so that, for example, bank robbers who escape with a good haul are more likely to return to the same bank branch. The third theory is that

> The offenders on the first occasion may tell others of the remaining goods, and those whom they tell return to commit offences.

which could be termed the buddy theory. Drawing upon this work, a theory of the relationship between repeat victimization and high-crime areas was proposed independently to the work on attractors and generators but which also drew on Cohen and Felson's (1979) notion of multiplicative interactions. The theoretical model draws on routine activity and has three principal steps. The first observes that linear increases in interactions between suitable targets, potential offenders, and conducive environments (the absence of guardianship) produce nonlinear increases in crime, the effect on repeat victimization depending upon the specifics of the respective changes. Thus, this interactor is similar to a crime generator. The second step adds a contagion or boost component wherein victimized targets have a heightened risk of experiencing further crime. Such risk is known to increase with each victimization so that a small proportion of targets progress to become chronically victimized supertargets. This explains why repeat victimization accounts disproportionately for crime in high-crime areas as found elsewhere (Johnson et al. 1997). The interactor-contagion model was developed in Farrell et al. (1994, 2005).

The interactor-contagion model offers a stepping-stone from repeat victimization to hot spots. Some hot spots are defined solely in terms of repeat victimization. Others include multiple targets that may or may not be repeatedly victimized. The mechanism that underlies generators is common to these theoretical perspectives. The interactor-contagion model incorporates increases in targets and suitable environments plus the same offenders returning, in addition to the increase in likely offenders that is the characteristic of attractors.

This brief review suggests theories of repeat victimization, and hot spots are compatible and to a large extent already integrated, at least implicitly, though differences in terminology and emphasis mean this not always obvious. Interactor-generator notions are common to both. Statistically, these result in a random chance or Poisson distribution of victimization across targets in the absence of a boost or attractor effect, and this has been a mainstay of research on repeat victimization since Sparks et al. (1977). Boosts, buddies, and attractors offer similar but distinct explanatory mechanisms. Boost denotes increased likelihood of *the same* offenders returning, buddies denotes increased likelihood of criminal *associates* returning, while attractors denote increased likelihood of *other* offenders returning.

The links between the theories bodes well for integration of theory and concepts relating to crime clusters. It seems that perhaps crime concentration of all types take place by the same basic mechanisms. Hence, a unified theory of crime clusters would include flags, interactor-generators, boosters, buddies, and attractors. Such a theory appears to account for changes in the number and interactions of suitable targets, suitable environments, and likely offenders, and thereby for variation in both individual and area-level risk. It might be applied more universally across clusters so that, for example, hot products flag their attractiveness and are found more frequently at interactor-generators, with their theft producing boost, buddy, and attractor effects.

Conclusion

Crime clusters are central to prediction, and absent prediction there is no prevention. The term clusters was used as a shorthand for, or stepping-stone to, a broad definition of near repeats incorporating the various dimensions and types of crime commonality and concentration.

A diverse set of crime clusters and commonality was reviewed, and crime event difference indices may assist in the exploration of how similarity, or nearness, is key to crime's clustering in all its forms. A comparison of the theories of key crime clusters showed them to be overlapping and compatible and suggests that the unification of theory in this area may promote cluster-busting crime prevention practice.

Related Entries

- ▶ Agent-Based Models to Predict Crime at Places
- ▶ Crime Location Choice
- ▶ Crime Mapping
- ▶ Hot Spots and Place-Based Policing
- ▶ Multiple Victims and Super Targets
- ▶ Repeat Victimization

Recommended Reading and References

Andresen MA, Malleson N (2011) Testing the stability of crime patterns: implications for theory and practice. J Res Crime Del 48(1):58–82

Averdijk M (2010) Individuals victimization patterns over time. Haveka, Alblasserdam

Block RL, Block CR (1994) Space, place and crime: hot spot areas and hot places of liquor-related crime. In: Eck J, Weisburd D (eds) Crime and place. Crime prevention studies, vol 4. Criminal Justice Press, New York

Brantingham PJ, Brantingham PL (1995) Criminality of place: crime generators and crime attractors. Eur J Crim Pol Res 3(3):1–26

Canter DV (2004) Offender profiling and investigative psychology. J Investig Psych Offender Profiling 1:1–15.

Clarke RV (1999) Hot products: understanding, anticipating and reducing demand for stolen goods. Police research series paper 112, Home Office, London

Cohen L, Felson M (1979) Social change and crime rate trends: A routine activity approach. Am Sociol Rev 44:588–605

Eck JE, Clarke RV, Guerette RT (2007) Risky facilities: crime concentration in homogenous sets of establishments and facilities. In: Farrell G, Bowers KJ, Johnson SD, Townelsey M (eds) Imagination for crime prevention: essays in honour of Ken Pease. Crime prevention studies, vol 21. Criminal Justice Press, New York

Eck JE, Weisburd D (1995) Crime and place, volume 4 of crime prevention studies. Criminal Justice Press, Monsey

Everson S (2003) Repeat victimisation and prolific offending: chance or choice? Int J Police Sci Manag 5:180–194

Farrell G (2005) Progress and prospects in the prevention of repeat victimisation. In: Tilley N (ed) The handbook of crime prevention and community safety. Willan, Cullompton, pp 145–172

Farrell G, Pease K (1993) Once bitten, twice bitten: repeat victimisation and its implications for crime prevention. Police Research Group, Crime Prevention Unit Paper 46. Home Office, London

Farrell G, Pease K (2008) Repeat victimization. In: Wortley R, Mazerolle L (eds) Environmental criminology and crime analysis. Cullompton, Willan

Farrell G, Clark K, Ellingworth D, Pease K (1994) High crime rates, repeat victimization, and routine activities. Paper presented to the Cropwood conference. University of Cambridge, Institute of Criminology, Cambridge

Farrell G, Clark K, Ellingworth D, Pease K (2005) Of targets and supertargets: a routine activity theory of high crime areas. Internet J Criminol pp 1–25

Johnson SD, Bernasco W, Bowers K, Elffers H, Ratcliffe J, Rengert G, Townsley M (2007a) Space-time patterns of risk: a cross national assessment of residential burglary victimization. J Quant Criminol 23:201–219

Johnson SD, Birks DJ, McLaughlin L, Bowers KJ, Pease K (2007b) Prospective crime mapping in operational context. Home Office, London

Johnson SD, Bowers KJ, Hirschfield A (1997) New insights into the spatial and temporal distribution of repeat victimisation. Br J Criminol 37(2):224–241

Kinney JB, Brantingham PL, Wuschke K, Kirk MG, Brantingham PJ (2008) Crime attractors generators and detractors: land use and urban crime opportunities. Built Environ 34(1):62–74

Levy MP, Tartaro C (2010) Repeat victimization: a study of auto theft in Atlantic city using the WALLS variables to measure environmental indicators. Crim Just Policy Rev 21:296–318

Li X, Han J, Lee JG, Gonzalez H (2007) Traffic density-based discovery of hot routes in road networks. In: Proceedings of the 10th international symposium on spatial and temporal databases, Boston. Available at http://www.xiaolei.org/research/pubs/sstd07.pdf
Lindstrom P (1997) Patterns of school crime: a replication and empirical extension. Br J Criminol 37(1):121–130
Mailley J, Garcia R, Whitehead S, Farrell G (2008) Phone theft index. Secur J 21(3):212–227
Melnyk T, Bennella C, Gauthiera DJ, Gauthier D (2010) Another look at across-crime similarity coefficients for use in behavioural linkage analysis: an attempt to replicate Woodhams, Grant and Price (2007). Psychol Crime Law 17(4):359–380
Morgan F (2001) Repeat burglary in a Perth suburb: indicator of short-term or long-term risk? In: Farrell G, Pease K (eds) Repeat victimization. Criminal Justice Press, Monsey, pp 83–118
Pease K (1998) Repeat victimisation: taking stock. Police Research Group, Home Office, London
Ratcliffe JH, Rengert GF (2008) Near repeat patterns in Philadelphia shootings. Secur J 21:58–76
Sherman LW, Gartin PR, Buerger ME (1989) Hot spots of predatory crime: routine activities and the criminology of place. Criminology 27(1):27–55
Smith MJ (2005) Robbery of taxi drivers. Problem-Oriented Guides for Police Problem-Specific Guide Series No. 34. Department of Justice, Washington, DC
Summers L (2010) Virtual repeats and near repeats. In: Fisher B, Lab S (eds) Encyclopedia of victimology and crime prevention. Sage, Los Angeles, pp 1043–1047
Tompson L, Partridge H, Shepherd N (2009) Hot routes: developing a new technique for the spatial analysis of crime. Crime Mapp J Res Prac 1(1):77–96
Townsley M, Homel R, Chaseling J (2003) Infectious burglaries: a test of the near repeat hypothesis. Br J Criminol 43:615–633
Velasco M, Boba R (2000) Tactical crime analysis and geographic information systems: concepts and examples. Crime Mapping News 2(2):1–4
Wellsmith M, Burrell A (2005) The influence of purchase price and ownership levels on theft targets: the example of domestic burglary. Br J Criminol 45:741–764

Predictive Policing

Craig D. Uchida
Justice and Security Strategies, Inc., Silver Spring, MD, USA

Synonyms

Forecasting crime or crime forecasting

Overview

Predictive policing is a new concept for law enforcement in the twenty-first century. While still in its infancy and relatively untested, predictive policing has the potential to change the way in which law enforcement deals with crime and victims. This entry describes predictive policing in terms of its definition and roots, the theories and models that have been developed, applications in law enforcement, and the issues that surround it.

Conceptually, predictive policing involves the use of data and predictive analytics to predict or forecast where and when the next crime or series of crimes will take place. The concept has engendered new terminology in law enforcement. “Predictive analytics,” “data mining,” “nonobvious relationships,” and “predictive spatial analysis” are among the new phrases used by chief executives, policy makers, and researchers to describe aspects of predictive policing. These and other phrases will be described and discussed.

Definition

Bill Bratton et al. (2009) initially defined predictive policing as “any policing strategy or tactic that develops and uses information and advanced analysis to inform forward-thinking crime prevention.” The Los Angeles Police Department and Uchida (2009) refined the definition and said, predictive policing is a “multi-disciplinary, law enforcement-based strategy that brings together advanced technologies, criminological theory, predictive analysis, and tactical operations that ultimately lead to results and outcomes – crime reduction, management efficiency, and safer communities.” Other definitions have emerged as well. Pearsall (2009) sees predictive policing as a “generic term for any crime fighting approach that includes a reliance on information technology (usually crime mapping data and analysis), criminology theory, predictive algorithms, and the use of data to improve crime suppression on the streets.” The National

Institute of Justice's Geospatial Technical Working Group (Wilson et al. 2009) broadly discussed predictive policing as including risk-based assessments, problem solving, crime prevention, and "as a strategy by which uncertainties are minimized and expectations are maximized."

Fundamentals

Using the LAPD/Uchida definition (Uchida 2009), predictive policing is built on multiple pillars including advanced technologies and analytics from the business world, statistical forecasting, criminological theory, and tactics and operations that are based on problem-oriented, intelligence-led, and hot spot policing. Understanding each of these components and how they converge into predictive policing is the central focus of this entry.

Advanced Technologies and Analytics: Business Intelligence

Predictive policing is based in part on business intelligence models and data analysis tools used by retailers like Wal-Mart and Amazon, insurance companies, or credit card companies like American Express. Netflix uses predictive analytics and data mining as the basis for its business model. They establish their regional centers, manage their inventory, and market their services based on the predicted behaviors of its customers.

In sports, baseball statisticians developed models and use analytics to predict performance. The Society for American Baseball Research (SABR) founded by Bill James in 1977 includes members who use different statistical models to predict or forecast the performance of individual ballplayers and teams. Though largely ignored by Major League Baseball executives for almost 25 years, sabermetricians became popular when Oakland A's general manager Billy Beane successfully showed the value of data and analytics to propel the A's to division championships in 2002 and 2003. Beane's story was chronicled in *Moneyball* (Lewis 2004). As a result of Beane's success, every baseball team now uses some form of statistical analysis to measure performance.

Other sports-related enterprises have followed this trend in using statistics and predictive algorithms. The New England Patriots of the National Football League was one of the first football teams to use analytics to evaluate college players and the Leicester Tigers in the United Kingdom, a two-time European champion in rugby, uses predictive tools to reduce injuries.

The commonalities of these diverse businesses are their use of data mining and predictive analytics. Data mining entails the analysis of data to identify trends, patterns, or relationships among the data. Basically, data mining is about "practical application – application of the algorithms developed by researchers in artificial intelligence, machine learning, computer science, and statistics" (Williams 2011).

Data mining includes the development of a predictive model. For example, Wal-Mart relied upon data from its supply chain to predict future demand. Analysts mined data seeking relationships between weather patterns and customer needs. They found that in anticipation of a large weather event like a hurricane or tropical storm, products like duct tape, bottled water, plywood, and Pop-Tarts are sent to an affected area. While water, tape, and plywood are typical necessities, Pop-Tarts (specifically strawberry ones) were discovered through data mining techniques and serve as an example of a "non-obvious relationship" (Beck and McCue 2009).

Predictive analytics is a fairly broad phrase that describes a variety of statistical and analytical techniques used to develop models for prediction. The form of these predictive models varies, depending on the behavior or event that is being predicted. Most predictive models generate a score, with a higher score indicating a higher likelihood of the given event occurring. Predictive analytics rely on increasingly sophisticated statistical methods, including multivariate analyses like time series and advanced regression models. These techniques enable the analyst to determine trends and relationships that may not be readily apparent or are "nonobvious" relationships.

Criminological Research and Theory

Criminologists have applied forecasting and predictive methods for nearly a hundred years. Berk (2008) notes that research on crime trends in the nineteenth century paved the way for explicit forecasts on both aggregate patterns of crime and on the behavior of individuals. As data and statistical tools have improved over time, a wide variety of work on forecasting and prediction have occurred. He cites 18 examples of forecasting studies between 1928 and 2007 (Berk 2008: 221), many of which forecast behavior of offenders after they have been released from prison and those that forecast prison populations. He cautions that many of the studies were not successful in accurately predicting behavior or crime and that the forecasts have not been properly evaluated.

In addition to Berk's examination of forecasts, the subtext for predictive policing lies within a number of macro-level, quantitative criminological research studies, including those that use repeat victimization, social disorganization, collective efficacy, strain, economic deprivation, routine activity, deterrence or rational choice, social support, and subcultural theories. These theories use "predictor variables" to forecast and predict crime locations and victims.

The theory of repeat victimization posits that for every type of crime studied (with the exception of homicide), the risk of victimization increases following an initial event, that a small proportion of victims accounts for a large proportion of crime, and where repeat victimizations occur, they usually do so within a brief period of time, allowing for a small window of opportunity for intervention (Johnson et al. 2009).

Theories of social disorganization and collective efficacy are also relevant in identifying and predicting where crime may flourish or not. Sociological literature over the last 30 years establishes the relationship between community characteristics and problem behavior. Areas characterized by high levels of economic disadvantage have long been associated with crime and other forms of disorder. Bursik and Grasmick (1993) argued that concentrated disadvantage and other neighborhood-level indicators of social disorganization create conditions that hamper the development of meaningful pro-social institutions to flourish. Institutions such as schools, churches, formal/informal community groups play an important role in regulating behavior among its residents. Sampson and Groves (1989) found that disorganized neighborhoods were also characterized by low organizational participation, sparse friendship networks, and large numbers of unsupervised youth. Neighborhoods with the most extreme forms of social disorganization also tended to experience greatly reduced levels of collective efficacy. Collective efficacy, which relates to the capacity of residents and other groups to exert levels of social control, has important implications for how neighborhoods are informally managed by residents. Research shows that neighborhoods with higher levels of collective efficacy generally experience lower levels of violence. With poor community organization, the capacity of the communities to develop formal and informal mechanisms of legal social control was diminished. From these findings, one can predict that violent crime may be higher in areas of low collective efficacy and lower in areas of high collective efficacy.

Other criminological theories can be applied to predictive policing models. Environmental criminology, which includes routine activity, rational choice, and situational crime prevention, is the broad, general approach that is most germane. Routine activity theory explains the components of a criminal incident by breaking it down to three basic elements: (1) a likely offender, (2) a suitable target, and (3) the absence of a capable guardian. It is only when these three elements converge in time and space that a crime occurs (Cohen and Felson 1979). This perspective established the spatial and temporal context of criminal events as an important focus of study.

Situational crime prevention involves crime prevention strategies that are aimed at reducing the criminal opportunities that arise from the routines of everyday life. Situational crime prevention assumes that crime is a rational choice by offenders and that crime can be prevented by hardening targets to increase the risks and reduce the rewards. British scholars, led by Ronald Clarke, explored the practical application of this theory in the United Kingdom. They saw that "opportunity" was at the core of crime, and rather than trying to reform offenders, they sought to reduce the criminal opportunities available to the criminal. Thus, they sought to change the environment through target hardening, improving surveillance of areas that might attract crime, and deflecting potential offenders from settings in which crimes might occur.

Criminological theory is important in predictive policing – because it assists in determining the appropriate unit of analysis for prediction, guides the specification of the predictive algorithms, and determines the measures of an intervention or tactic employed by the police.

Problem-Oriented, Intelligence-Led, and Hotspot Policing

Predictive policing is tied directly to principles, research findings, and tactics and operations established in problem-oriented, intelligence-led, and hot spot policing. Central to the notion of problem-oriented policing is the idea that police agencies need to become proactive rather than reactive, moving away from simply reacting to citizen calls-for-service, and moving toward identifying and solving the problems that generate excessive calls (Goldstein 1990). For example, research in Minneapolis showed that only 3.3 % of addresses and intersections were responsible for 50.4 % of all calls for which a police car was dispatched (Sherman et al. 1989). Solving the chronic, recurring problems underlying these "hot spots" might reduce crime, fear of crime, and calls-for-service, and improve police-community relations. While spatial proximity is one of the ways that individual incidents or calls to the police can be grouped into larger problem types, there are many others. For instance, police might group individual incidents together into larger problem types based on similarities in victim characteristics, offender characteristics, modus operandi, or temporal patterns.

Likewise, hot spot policing implicitly and directly includes predictive components. Weisburd's notion of "the law of crime concentrations at places" refers to research that shows "irrespective … of the city, or even country examined, … 20 to 25 % of crime is found at only 1 % of the places in a city" (Weisburd 2012). By extension and implication, these are "predictable" locations.

Intelligence-led policing is also closely associated with predictive and problem-oriented policing because they all fit into the community-oriented policing paradigm. Ratcliffe defines intelligence-led policing as "a business model and managerial philosophy where data analysis and crime intelligence are pivotal to an objective, decision-making framework that facilitates crime and problem reduction, disruption and prevention through both strategic management and effective enforcement strategies that target prolific and serious offenders" (Radcliffe 2008).

Predictive Models

A number of models currently exist for predicting crime within the geography of crime literature (Groff and La Vigne 2002) and in texts and articles on statistical and mathematical methods (Berk 2008, 2009; Williams 2011). Prediction models may use a variety of regression models, including linear regression (OLS), partial and stepwise regression, logit or probit regressions, and regression splines. Regression models describe the relationship between the dependent variable (in predictive situations, the variable to be predicted) and independent, explanatory variables. Regression implies some causation and then tests how well the regression model fits the specified relationship. However, the statistical techniques used in predictive analytics are largely untested and have not been rigorously evaluated.

Geography of Crime

Geographic information systems (GISs) have been used to develop models for forecasting and predicting crime. Over the past decade, these models vary from simple to complex. The simplest forecasting method is common to Compstat, where the hot spots of last week or last month are the hot spots of next week or next month. Crime analysts prepare maps of crime that have already occurred and those maps are used to determine where to deploy officers and where to intervene. Other methods include the use of raster GIS, a method included in ArcView's Spatial Analyst that interpolates a surface of crime and looks similar to a weather map.

Risk Terrain Modeling (Caplan et al. 2011) uses multiple layers of raster maps relying on the "dynamic interaction between social, physical and behavioral factors that occurs at places." Risk Terrain Modeling (RTM) will identify particular risk factors for crime. The model will assign a value to every place within a specific geographic area and will signify the presence, absence, or intensity of each risk factor. When multiple layers are combined into a risk terrain map, it provides a composite risk value to particular places. The higher the risk value, the greater the likelihood of a crime occurring at that location. RTM has been applied to shootings, violent crimes, and burglaries in Newark and Philadelphia (Caplan et al. 2011).

Another example of predictive mapping is ProMap, a prospective mathematical model that predicts where burglaries would next occur. Using repeat victimization theory, Johnson et al. (2009) determined that specific grid locations were predicted to have the greatest imminent risk where the greatest number of burglaries clustered in time and space both recently and nearby. The researchers theorized that risk of burglaries diminished over time and used a short time period of 1 week for their temporal interval. They compared the predictive ability of ProMap to two methods – Kernel Density Estimation (KDE) and to maps generated by analysts using police beat geography and data from Merseyside, UK. The ProMap model consistently performed better than the KDE method in identifying 10 %, 25 %, 50 %, and 75 % of burglaries. The mapping approach used by analysts was worse than the KDE method as it "failed to exceed chance for almost very trial" (Johnson et al. 2009).

Swatt (2003) attempted forecasts of burglary and robbery in Omaha on a monthly basis for Census blocks using a Hierarchical Linear Model framework. He found that overdispersed models were overwhelmed by the large number of zero event counts and thus could not accurately forecast either crime. Gorr and Olligschlaeger (2002) compared simple univariate, OLS regression, and neural networks strategies for forecasting calls-for-service in Pittsburgh. Their results did not convincingly demonstrate an advantage for the neural network method. Brown and Kerchner (2000) developed a methodology for forecasting criminal events based on the "feature space" of offender site selection, a set of spatial, temporal, and site characteristics that influence an offender's decision to engage in crime. This strategy, as well as the strategy outlined by Liu and Brown (2000), outperformed a simple model for predicting breaking and entering, with only limited improvement for auto theft. Recently, Fox and colleagues (2012) have extended this model by incorporating it within a Hierarchical Bayesian Model to introduce spatial and temporal effects. Their results suggest that although the extended model outperforms the original Brown and Kercher (2000) strategy for predicting short-term variation in assaults at the Census block group level, the added computation time limits the applicability of this strategy.

Statistical Learning Models

"Statistical learning" includes methods that experiment with many different models for data. Statistical learning has a strong focus on data mining and machine learning traditions of data analysis. In describing the use of statistical learning, Berk (2008) begins with regression splines and moves to other models, particularly those that make use of decision trees or classification and regression trees (CARTs). Decision trees or CARTs are the building blocks of data mining. Statisticians have found that using one decision tree model is overly simplistic, so combining multiple models into a single ensemble of models – a forest of trees – works better in prediction (Williams 2011). "Bagging," "boosting," and "random forests" are among the algorithms used within statistical learning methods that have been developed for predicting and forecasting.

The random forest algorithm holds substantial promise as a method of forecasting crime using statistical learning. The random forest algorithm is an extension of the decision tree or CART method. The random forest algorithm (see Berk 2008 and Williams 2011) grows a large number of decision trees using samples (500 or more, Berk 2008) from a database. At each stage in CART, a random sample of predictors is used (this is referred to as "bagging," as a random sample of observations is placed in a "bag"). After the trees are formed, classifications are found by majority vote across trees where the case in question was in the bagged data. In other words, in CART, each tree will develop its own predictive value. When there are 500 trees in a random forest, the majority of trees will determine the prediction. If 300 of 500 trees predict that a crime will occur at a location tomorrow, then it is expected that there is a high likelihood (60 %) that the crime will occur at the location.

The random forest algorithm appears to offer a number of advantages over other methods. First, data do not need to be preprocessed or normalized for use. For example, outliers (defined as extreme deviations in the data) do not need to be expunged or normalized. Second, if there are many input variables, specific selection of those variables does not need to take place because the random forest model targets the most useful variables. Third, "overfitting" of data (when a statistical model describes random error or noise instead of the underlying relationship) is less of a problem because classifications are averaged across a large number of trees (Berk 2008). The random forest model is yet to be tested with police data, so it is not yet known whether and how it can be used for predictive policing. Berk (2011), however, demonstrates its use and value in forecasting parole or probation failures.

Applications

One of the first applications of predictive policing methods occurred in the Richmond (VA) Police Department (RPD) in 2003. With the assistance of vendors, RPD developed a system that included a combination of predictive crime analysis, data mining, geographic information system (GIS) capabilities, and reporting to RPD officers. Data from the RPD records management system were integrated and analyzed using SPSS's predictive analytic.

Richmond analysts created a model that automatically improved itself and avoided the manual refreshing of variables. Time, day, holidays, weather, moon phases, city events, paydays, and crime records were among the data that were analyzed. RPD developed a custom solution using several different technologies. Information Builders' WebFocus software included business intelligence capabilities that display criminal activity every 4 hours. Officers received the information at police stations and in squad cars, and real-time alerts were sent by email and text messages. GIS mapping from ESRI and aerial photography from Pictometry provided maps and pictures of the reported incident locations and the surrounding neighborhoods. RPD used SPSS's Clementine and Predictive Enterprise Services products for data mining capabilities that examine how current crime reports relate to data on past, present, and projected actions (Harris 2008).

Data mining tools and predictive analytics were the key elements in the Richmond effort. Using data mining techniques, which basically churned through data, analysts found "hidden patterns and relationships" (McCue and Parker 2003). According to McCue, hypotheses were created, models were developed, and "what if" questions were posed. The data mining tools allowed RPD to conduct a multitude of activities focusing on crime fighting and efficient use of manpower. The developers of these tools asserted that they added value to (1) deployment, (2) tactical crime analysis, (3) behavioral analysis of violent crime, and (4) officer safety.

In using this approach, the Richmond police deployed officers at different hours of the day, during different seasons, and for varying types of offenses. The tools provided them with unanticipated factors that they would not ordinarily find without data mining. While initially this approach showed promise in Richmond, a new police administration has opted to move away from it and return to more traditional strategies.

LAPD and Santa Cruz Applications

In 2011, the Los Angeles Police Department (LAPD) and the Santa Cruz Police Department (SCPD) made news headlines for their use of predictive analytics developed by mathematicians at UCLA and Santa Clara University. Using methods from seismology, Dr. George Mohler and Dr. Martin Short developed math algorithms that have the potential to predict crime across time and space (Mohler et al. 2011). The algorithms were initially tested using core data from the LAPD. The math model is based on "self-exciting point processes" and repeat victimization theory. That is, crime tends to display similar statistical patterns whereby an initial crime event is followed close in time and close in space by additional crimes. In the case of property crimes, offenders tend to return to the same, or a nearby location to commit more crime in a short period of time to replicate the successes of the prior offense (Townsley et al. 2003). In the case of many violent crimes, repeat offenses occur in short successions in the same places as part of cycles of reprisals and counter-reprisals.

The self-exciting point process model describes the rate at which crime occurs at a given spatial location (e.g., street address) and does so in terms of a background rate and a self-exciting component. The background rate is connected to general features of the local environment, which are thought to be constant relative to the time scales at which criminals make decisions. For example, houses targeted by burglars generally do not change fast enough in their characteristics to deter a burglar who has recently committed a crime there. By contrast, the self-exciting component takes into account the notion that recent crimes at a given location or neighboring locations tend to increase the rate at which crime occurs there. Burglars often learn of goods that they could steal during the process of breaking into and burglarizing a home. They act on this information by returning to victimize that home again a few days later.

In Los Angeles and in Santa Cruz, the math algorithm is tested on burglaries and burglaries from motor vehicles on 3 years of data. The expectation is that there is a background component to burglary rates as well as a strong self-exciting component. Locations targeted for burglary from motor vehicles typically offer both a large numbers of targets (e.g., parking lots, high-density street parking), with high variance in target types (e.g., car types) and goods to steal (e.g., laptops, cell phones, GPS units). Offenders specializing in burglary from motor vehicles will tend to return to the same or nearby locations to replicate previous successes and they will do so over short time intervals, consistent with a self-exciting process. The researchers quantified these background and self-exciting components for burglary from motor vehicles and developed forecasts based on these quantitative models. Forecasts take the form of probabilistic assessment of the likelihood that different locations will experience burglaries. The volume of burglary from motor vehicle crimes and the greater stationarity of the target base suggest that forecasts may be possible at the scale of days and perhaps even finer temporal scales.

The self-exciting point process is being tested in the field in five divisions in Los Angeles. The algorithm identifies 500 by 500 ft areas of probable criminal activity. Each week, 60 areas are designated as predictive policing areas. Using an experimental design, the police captain within each division directs officers to 30 randomly assigned areas to prevent the crimes from occurring and leaves 30 areas as control areas where routine patrol activities take place. Measures include the time spent in each predictive area by "extra patrol" and by regular patrol. Crimes are also measured to determine how many have been prevented. Preliminary findings show that nonviolent crime has dropped in the areas.

A number other agencies are in the process of using predictive analytics to predict crime and locations. These include: Cambridge (MA), Charlotte-Mecklenburg (NC), Charleston, SC. Chicago, Indio (CA), Lincoln (NE), Memphis, Miami-Dade, New York City, Redlands (CA), Salt Lake City, and Shreveport (LA).

Key Issues/Controversies

This section examines three major issues: questions about "Minority Report," civil liberties, and intelligence gathering; importance of data; and untested predictive analytic methods.

"Minority Report," Civil Liberties, and Intelligence Gathering

The Steven Spielberg production of "Minority Report" starring Tom Cruise was released in 2002 and tells the story of a futuristic "Department of Pre-Crime" that works out of the US Department of Justice. The film depicts the use of "pre-cogs" who predict crimes as they float in an indoor swimming pool and send telepathic messages that name potential victims and suspects to Justice Department police. Using advanced technology and having access to untold databases, the police investigate the crime before it happens. The police seemingly ignore the 4th Amendment by entering houses without warrants, use biometrics to confirm identities, surveil and enter apartments using spider-like devices that "read" corneas for identification, and when they apprehend a suspect before the crime takes place they arrest him for a "pre-crime."

The actions in the film led many policymakers, researchers, and law enforcement to ask: "Is predictive policing similar to Minority Report? Are we headed in that direction?" The questions raise a number of issues related to civil liberties, the 4th Amendment's right to be free from unreasonable searches and seizures, and a concern for the type of information that is or could be collected by law enforcement.

Police departments and executives are keenly aware of the negative implications raised by the film and its application to predictive policing. In Los Angeles, the chief responds publicly and privately that "you can't break the law to enforce the law" (OurWeekly 2010) and that the police "practice Constitutional policing." More importantly, the concern for civil liberties has led to discussions with the American Civil Liberties Union (ACLU), the development of policies and procedures, and acknowledgement of the need for training on all facets of predictive policing. As a result of these discussions, the focus of the LAPD effort is centered on preventing specific types of crime and reducing crime in specific locations rather than predicting offender behavior.

Data

As with all uses of analytics, data are among the most critical aspects of predictive policing. The validity and legitimacy of any predictive model depends on the quality and quantity of data available. In law enforcement, data are usually in abundance – calls-for-service, incident reports, arrest reports, field interviews, and other data are collected routinely by most large- and medium-size departments across the country. But these data vary by police agency in terms of consistency and uniformity, and while they are abundant, their quality is unknown. Data need to be cleaned and validated to ensure that the results from predictive analytics are accurate.

While the amount of data may be sufficient for use in a predictive model, the information is often stored on legacy systems that may not be compatible with systems running predictive analytics software. Converting data on these legacy systems to a useable format is a challenge.

Untested Methods

The statistical techniques used in predictive analytics are largely untested and have not been rigorously evaluated. Risk Terrain Modeling, ProMap, neural networks, feature space, statistical learning models, the self-exciting point process, and predictive analytics developed by private vendors (IBM SPSS, SAS, IBI, and others) are yet to be fully tested in the field by independent evaluators.

Conclusion

Though predictive policing remains an untested concept, many police departments have adopted ways to begin to predict crime using mapping, business, and statistical analytic methods. The reasons behind the use of predictive analytics are many, including the need to increase efficiency and maximize police resources. In cities like Los Angeles, where resources are limited because of the downturn in the economy, cost-effective information technology solutions are an essential management tool. Increasing the ability of police agencies to effectively analyze existing and new data may provide broader crime reduction opportunities. This in turn will lead to new deployment strategies, different methods for measuring prevention, and ultimately, safer communities.

Related Entries

▶ Applied Geographical Profiling

▶ Crime Mapping

▶ Forecasting of Shootings Using Risk Terrain Modeling

▶ Hot Spots and Place-Based Policing

▶ Intelligence-Led Policing

▶ Near Repeats and Crime Forecasting

▶ Problem-Oriented Policing

▶ Routine Activities Approach

▶ Situational Crime Prevention

Recommended Reading and References

Beck C, McCue C (2009) Predictive policing: what can we learn from Wal-Mart and Amazon about fighting crime in a recession? The Police Chief 76(11)

Berk R (2008) Forecasting methods in crime and justice. Ann Rev Law Soc Sci 4:219–238

Berk R (2009) Statistical learning from a regression perspective. Springer, New York

Berk R (2011) Asymmetric loss functions for forecasting in criminal justice settings. J Quant Criminol 27:107–123

Bowers K, Johnson S, Pease K (2004) Prospective hot spotting: the future of crime mapping? Brit J Criminol 44(5):641–658

Bratton W, Morgan J, Malinowski S (2009) The need for innovation in policing today. Unpublished manuscript, Harvard Executive Sessions, October

Brown DE, Kerchner SH (2000) Spatial-temporal point process models for criminal events. Paper presented at the National Institute of Justice Annual Crime Mapping Conference. Washington, DC. Included in Brown DE (2002) Final report: predictive models for law enforcement. U.S. Department of Justice, Washington, DC

Bursik RJ, Grasmick HG (1993) Neighborhoods and crime: the dimensions of effective community control. Lexington Books, Lanham

Caplan JM, Kennedy LW, Miller J (2011) Risk terrain modeling: brokering criminological theory and GIS methods for crime forecasting. Justice Quarterly 28 (2):360–381

Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44(4):588–605

Fox JS, Huddleston H, Gerber M, Brown DE (2012) Investigating a Bayesian hierarchical framework for feature-space modeling of criminal site-selection problems. Paper presented at the Midwest artificial intelligence and cognitive science conference 2012, Cincinnati, 21–22 Apr 2012. Retrieved 29 May 2012, from http://ceur-ws.org/Vol-841/submission_31.pdf

Goldstein H (1990) Problem-oriented policing. McGraw Hill, New York

Groff ER, La Vigne NG (2002) Forecasting the future of predictive crime mapping. Crime Prevention Studies 13:29–57

Gorr W, Olligschlaeger A (2002) Crime hot spot forecasting: modeling and comparative evaluation. Final report. U.S. Department of justice. National Institute of Justice. Washington, D.C

Harris C (2008) Richmond, Virginia, police department helps lower crime rates with crime prediction software. Retrieved 10 Dec 2009, from http://www.govtech.com/gt/print_article.php?id=575229

Johnson SD, Bowers KJ, Birks DJ, Pease K (2009) Predictive mapping of crime by ProMap: accuracy, units of analysis, and the environmental backcloth. In: Weisburd D, Bernasco W, Gerben JN, Bruinsma J (eds) Putting crime in its place. Springer, New York

Lewis M (2004) Moneyball: the art of winning an unfair game. W.W. Norton, New York

Liu H, Brown DE (2000) A new point process transition density model for space-time event prediction. In: Brown DE (ed) (2002). Final report. Predictive models for law enforcement. U.S. Department of justice. Washington, DC

McCue C, Parker A (2003) Connecting the dots: data mining and predictive analytics in law enforcement and intelligence analysis. Police Chief 10(10): 115–124

Mohler GO, Short MB, Brantingham PJ, Schoenberg FP, Tita GE (2011) Self-exciting point process modeling of crime. J Am Stat Assoc 106(493):100–108

OurWeekly (2010). Retrieved 5 May 2012, from http://ourweekly.com/los-angeles/los-angeles-police-chief-charlie-beck

Pearsall B (2009) Predictive policing: the future of law enforcement? NIJ J/Issue No. 266, at 16 (National Institute of Justice)

Radcliffe J (2008) Intelligence-led policing. Willan, London

Sampson RJ, Groves W (1989) Community structure and crime: testing social disorganization theory. Am J Sociol 94(4):774–802

Sherman LW, Gartin P, Buerger M (1989) Hot spots of predatory crime. Criminology 27(1):27–56

Swatt M (2003) Short-term forecasting of crime for small geographic areas. Unpublished PhD dissertation, University of Nebraska at Omaha, 216 pp

Townsley M, Homel R, Chasling J (2003) Infectious burglaries: A test of near repeat hypothesis. Brit J Criminol 43:615–633

Uchida CD (2009) Predictive policing in Los Angeles: planning and development. A white paper published by Justice & Security Strategies, Inc., December

Weisburd D (2012) Bringing social context back into the equation. Criminol Public Policy 11(2):317

Williams G (2011) Data mining with rattle and R: the art of excating data for knowledge discovery. Springer, New York

Wilson R, Smith SC, Markovic JD, LeBeau JL (2009) Geospatial technical working group: meeting report on predictive policing. US Department of Justice, National Institute of Justice

Predictive Sentencing

Andreas von Hirsch[1,2] and Lila Kazemian[3]
[1]Cambridge University, Cambridge, UK
[2]Goethe University Frankfurt, Frankfurt, Germany
[3]John Jay College of Criminal Justice, New York, NY, USA

Overview

Prediction research in criminology has, by and large, focused on characteristics of offenders: various facts about criminals are recorded – their age, previous arrests and convictions, social history, and so forth. It is then statistically determined which of these factors are most strongly associated with subsequent offending (see Gottfredson 1967). The result is a "selective" prediction strategy: among those convicted of a given category of offense, some will be identified as bad risks and others will not. Under an incapacitative strategy, those convicted offenders who constitute bad risks would be imprisoned or imprisoned for longer periods.

Traditional Prediction Methods

Traditional statistical prediction techniques pursued this selective approach. Generally, they found that certain facts about an offender (principally, his previous criminal history, drug habits, and history of unemployment) were to a modest extent indicative of increased likelihood of recidivism on his part (Gottfredson 1967).

These techniques did not, however, distinguish between serious and lesser forms of reoffending. Both the offender who subsequently committed a single minor offense and the individual who committed many serious new crimes would be lumped together as potential recidivists. Moreover, the technique offered no promise of generating reduced net crime rates, as they did not attempt to estimate such incapacitative strategies' aggregate crime-preventive effects. Locking up the potential recidivist thus assured only that he or she would be restrained; since other criminals remained at large, it did not necessarily diminish the overall risk of victimization. These limitations eventually reduced penologists' interest in traditional prediction techniques.

"Selective Incapacitation"

Surveys of imprisoned offenders conducted in the USA in the early 1980s found that a small number of active offenders admitted responsibility for a disproportionate number of serious offenses. If that minority of dangerous offenders could be identified and segregated, perhaps this could reduce crime rates after all. These surveys thus generated a renewed interest in prediction research.

The most notable product was a 1982 RAND Corporation study by Peter W. Greenwood (1982). He named his prediction strategy "selective incapacitation." His idea was to target high-rate, serious offenders – namely, those offenders likely to commit frequent further acts of robbery or other violent crimes in the future. For that purpose, he took a group of incarcerated robbers, asked them how frequently they had committed such crimes, and then identified the characteristics of those reporting the highest robbery rates. From this, he fashioned a predictive index, which identified potential high-rate offenders on the basis of their criminal records and histories of drug use and unemployment. Greenwood's predictive factors were (i) prior convictions of instant offense type, (ii) incarceration for more than half the preceding 2 years, (iii) conviction before age 16, (iv) time served in a state juvenile facility, (v) drug use during the preceding 2 years, (vi) juvenile druguse, and (vii) employment for less than 50 % of the preceding 2 years. He defined "high-risk" offenders as those for whom at least four of these seven factors were present (Greenwood 1982).

Greenwood also devised a novel method of projecting the aggregate crime reduction impact of this prediction technique. On the basis of offender self-reports, he estimated the average annual rate of offending of those robbers who were identified as high risks by his prediction index. He then calculated the number of robberies that would be prevented by incarcerating such individuals for given periods. By increasing the prison terms for the high-risk robbers while reducing the terms for the others, he concluded it would become possible to reduce the robbery rate by as much as 15–20 % – without causing prison population to rise (Greenwood 1982).

Queries About Effectiveness

While Greenwood's study initially attracted much interest, certain difficulties became apparent. One difficulty was making the predictions hold up when official data of the kind that sentencing courts have available are relied upon. The objective of selective incapacitation is to target the potential high-rate serious offenders and to distinguish them from recidivists who reoffend less frequently or gravely. To make this distinction, the RAND studies, including Greenwood's, relied upon offender self-reports. A sentencing court, however, is seldom in the position to rely upon defendants' willingness to supply the necessary information about their criminal histories. The court will need to rely instead on officially recorded information about offenders' adult and juvenile records, and such documentation is too meager to make the distinction adequately. When Greenwood's data were reanalyzed to see how well the potential high-risk serious offenders could be identified from the information available in court records, the results were disappointing. The officially recorded facts –arrests, convictions, and information about offenders' personal histories – did not permit the potential high-rate robbers to be distinguished from (say) the potential car thieves. The factors in the self-report study that had proved the most useful – such as early and extensive youthful violence and multiple drug use – were not reflected in court records (Chaiken and Chaiken 1984). To make the predictions work, the courts would need to obtain and rely on information in school and social-service file – with all the problems of practicability and due process that would involve.

Questions became apparent, also, in the projections of preventive impact. Greenwood based his crime reduction estimates on the self-reported activities only of incarcerated robbers and then extrapolated those estimates to robbers generally. Incarcerated robbers, however, are scarcely a representative group: they tend to offend more frequently than robbers generally in the community. (It is like trying to learn about the smoking habits of smokers generally by studying the self-reported smoking activities of patients in a lung cancer ward.) When this kind of extrapolation is eliminated, the projected crime reduction impact is reduced by about half (see von Hirsch 1985, Chap. 10).

Other problems with the projections also became apparent. Greenwood assumed, for example, that his high-rate robbers would continue offending for a long time. When shorter (and perhaps more realistic) residual criminal careers are assumed instead, the estimated preventive effect shrinks dramatically. The accuracy of the forecasts was also disappointing: comparing predicted with actual offending rates, Greenwood's analysis showed that only about half of the respondents were categorized accurately the by predictive scale.

These doubts were confirmed by a report of a National Academy of Sciences' Panel on criminal careers that appeared 4 years after the Greenwood study (for text of the Panel's report, see Blumstein et al. 1986). The Panel included several noted advocates of predictively based sentencing – and the report endorsed the idea of predictive strategies (within certain limits) so long as these could be shown to be effective. Nevertheless, the Panel's conclusions regarding the crime-preventive effects of selective incapacitation were quite skeptical. After recalculating Greenwood's results and thus scaling its initial preventive estimates down considerably, the Panel noted that even those revised estimates would shrink further were the scale drawn from a broader and potentially more heterogeneous population than persons in confinement andwere to utilize officially recorded rather than self-reported information. Even those reduced projected effects would largely disappear, moreover, if the estimated length of the residual criminal career were scaled down (see von Hirsch 1988).

Prospects for Improving Predictive Techniques

Could these difficulties be overcome? Greenwood's research has been only a beginning, and future selective incapacitation research may eventually produce more convincing results. The obstacles are considerable, however. If the aim is to distinguish potential high-rate, serious offenders from lesser potential criminals, this task remains difficult to achieve using the scant official records courts have at their disposal. Records of early offending might become somewhat more accessible with a change in the law concerning the confidentiality of juvenile records – but such records notoriously suffer from incompleteness and inaccuracy. Social histories, such as drug use and unemployment, will be even more difficult to ascertain accurately.

Estimation of the impact of selective incapacitation on crime rates also involves difficult problems of sampling. Analyses of convicted or incarcerated offenders' criminal activities suffer from the difficulty mentioned already: it is not clear to what extent these person's activities are representative of the activity of offenders in the community. Samples drawn from the general population are free from such bias but may contain too small a number of active offenders.

The most troublesome issue, however, remains that of estimating the length of offenders' residual criminal careers. The serious offenders who are the targets of selective incapacitation policies ordinarily would be imprisoned in any event; the most salient policy issue is that of the length of their confinement. The strategy is to impose longer terms on the supposed high-risk offenders, but that assumes they will continue their criminal activities for a protracted period. Little prevention is achieved if the bad risks who are confined are those whose active careers will terminate fairly soon. This means that selective incapacitation, to succeed, needs not merely to pick out high-risk offenders during their initial periods of high-rate offending, but rather those who are likely to go offending for substantial time. But how much do we know about forecasting residual careers?

A number of recent criminal-career studies have attempted to make estimates of the duration and intensity of offenders' residual criminal careers (see, e.g., Kazemian and Farrington 2006). These studies confirm that residual career length and frequency of offending decline significantly with age. Moreover, offenders' scores on risk-assessment indices – when based mainly on information included in official records – were significantly but only modestly associated with the extent of their remaining criminal careers. These findings also suggest that incapacitative benefits will decline significantly during the remainder of a predictively based sentence – thus limiting the crime-preventive payoff of selective incapacitation strategies.

Even with improved predictive efficacy, there would be no assurance as to how much confining the higher-risk offenders will prevent offenses from taking place. With group crimes and network offenses, for example, the offender may be replaced easily by other criminal participants. Moreover, it also needs to be born in mind that individuals who are admitted into prisons must eventually be released. Sampson and Laub's (1997) theory of cumulative disadvantage, for example, emphasizes the "*indirect* role of incarceration in generating future crime," which may occur through severance of bonds to conventional social institutions.

Proportionality Problems

Selective prediction strategies – whether the traditional sort or more sophisticated methods such as Greenwood's – must confront also an important ethical question: their apparent conflict with the requirements of proportionality. The conflict stems from the character of the factors relied upon to predict. Those predictive factors tend to have little relation to degree of reprehensibleness of the offender's criminal choices (which is the basis of proportionality judgements – see more fully von Hirsch and Ashworth 2005, Chap. 9).

Proportionality requires that penalties be based chiefly on the seriousness of the crime for which the offender currently stands convicted. The offender's previous criminal record, if considered at all, should have a secondary role and the offender's social status is largely immaterial to the penalty he or she deserves. With selective prediction, the emphasis necessarily shifts away from the seriousness of the current offense. In fact, Kazemian and Farrington found that offense type was not a significant predictor of residual career length or residual number of offenses. Since the aim is to select the higher-risk individuals from among those convicted of a given category of crime, the character of the current offense cannot have much weight.

Traditional prediction indices largely ignored the gravity of the current offense and concentrated on the offender's earlier criminal and social histories. Selective incapacitation techniques have similar emphasis: of Greenwood's predictive factors, three do not measure criminal activity of a significant nature at all but, instead, measure the offender's personal drug consumption and lack of stable employment. Of the four other factors, only two reflect the offender's recent criminal record, and none measures the heinousness (e.g., the degree of violence) of the offender's current offense (see discussion above).

When aggregate preventive effects are taken into account, the proportionality problems become more worrisome still. Selective incapacitation techniques, by their own proponent's reckoning, could promise significant crime reduction effects only by infringing proportionality requirements to a very substantial degree. Greenwood's projection of a significant reduction on the robbery rate was made on the assumption that robbers who score badly on his prediction index would receive about 8 years imprisonment, whereas better-scoring robbers would receive only 1 year in jail (Greenwood 1982). This means a great difference in severity – in Greenwood's study, about 800 % – in the punishment of offenders convicted of the same type of offense, and one that can scarcely begin to be accounted for by distinctions in the seriousness of the offender's criminal conduct. When this punishment differential is narrowed – when high-risk robbers receive only modestly longer terms than robbers deemed lower risks – the crime reduction payoff (even by Greenwood's methodology) shrinks to slender proportions.

Conclusion

Where does this leave us? A limited capacity to forecast risk has long existed: persons with criminal records, drug habits, and no jobs tend to recidivate at a higher rate than other offenders, as researchers have known for decades. However, the limitations in that forecasting capacity must be recognized – for selective incapacitation as well as more traditional forecasting techniques. Identifying high-risk, serious offenders will be impeded by the quality of information available (or likely to become available) to sentencing courts. The potential impact of selective incapacitation on crime rates is far below proponents' initial estimates and is likely to be modest. Considerations of proportionality limit the inequalities in sentence that may fairly be visited for the sake of restraining high-risk offenders, and limiting these permissible inequalities will, in turn, further restrict the technique's impact on crime. In order for this sentencing model to be effective, some important empirical and ethical caveats associated with selective incapacitation policies must be addressed.

Related Entries

- ▶ Crime Preventive Effects of Incapacitation
- ▶ Estimating the Effects of Incapacitation
- ▶ European Perspectives on Long-term Imprisonment
- ▶ Extent of Imprisonment: Global View
- ▶ Incapacitation
- ▶ Unintended Effects of Imprisonment

Recommended Reading and References

Blumstein A, Cohen J, Roth J, Visher C (eds) (1986) Criminal careers and career criminals. National Academy of Sciences, Washington, DC

Bottoms AE, von Hirsch A (2010) The preventive effect of criminal sanctions, Ch. 3. In: Cane P, Kritzer H (eds) Oxford handbook of empirical legal studies. Oxford University Press, Oxford

Chaiken M, Chaiken J (1984) Offender types and public policy. Crime Delinq 30:195–226

Gottfredson D (1967) Assessment and prediction methods in crime and delinquency, in president's national commission on law enforcement and criminal justice, task force report: juvenile delinquency and youth crime. Washington, DC

Greenwood PW (1982) Selective incapacitation. RAND Corporation, Santa Monica

Kazemian L, Farrington DP (2006) Exploring residual career length and residual number of offenses for two generations of repeat offenders. J Res Crime Delinq 43:89–113

Sampson RJ, Laub JH (1997) A life-course theory of cumulative disadvantage and the stability of delinquency, Ch. 4. In: Thornberry T (ed) Developmental theories of crime and delinquency: advances in criminological theory. Transaction Publisher, New Brunswick

von Hirsch A (1985) Past or future crimes: deservedness and dangerousness in the sentencing of criminals. Rutgers University Press, New Brunswick

von Hirsch A, Ashworth A (2005) Proportionate sentencing: exploring the principles. Oxford University Press, Oxford

Zimring FE, Hawkins G (1995) Incapacitation. Oxford University Press, New York

Predictive Validity

▶ Self-Reported Offending: Reliability and Validity

Predictors of Gang Membership

▶ Risk Factors for Gang Membership

Predictors of Successful Investigations

Robert T. Meesig
Melbourne, FL, USA

Overview

The continuing tension in our society between community and individual interests has shaped police investigation styles since the 1840s. In the context of this tension, the relative ineffectiveness of the current traditional, reactive, and case-oriented investigation style is examined in terms of four predictors that affect whether or not the investigation of a crime may be successful. Recent developments in community policing and terrorism matters indicate how to increase investigative successes, but they can also alter the balance between community and individual interests in ways that may not have yet been fully resolved.

Introduction

When a crime is committed in the United States, what is the likelihood that it will be investigated successfully? Apparently, based on the available information, it does not appear to be very likely indeed. In this entry, four major predictors of police criminal investigation successes are examined in separate sections. They are as follows:

- Does the crime become known to the police?
- Does the crime survive the case screening process of the police?
- What factors affect the police investigation?
- What is a successful investigation?

For each predictor, the entry describes significant influences on investigative outcomes and identifies some of the major issues involved. It concludes with a look at what the future may hold for police investigations.

Does the Crime Become Known to the Police?

In the United States, one of the more common ways to categorize various types of crimes is as follows (Cole and Smith 2008):

- Visible crimes – including murder, rape, robbery, assault, burglary, larceny, motor vehicle theft, and arson and lesser disorders such as vagrancy, shoplifting, vandalism, and violations of local ordinances and codes
- Victimless crimes – crimes such as drug offenses, prostitution, and gambling that involve illegal exchanges of goods or services
- Organized crimes – crimes committed by social frameworks organized to commit crimes
- Occupational crimes – crimes committed within a legal business or profession
- Political crimes – crimes such as terrorism, treason, sedition, and espionage

It is only in the Visible Crime Category that we have national crime statistics programs that annually document the number of crimes committed in the United States, and there are two such programs. The National Crime Victimization Survey (NCVS) is one program and is discussed below. The Uniform Crime Report (UCR) is the other and is addressed in the next section.

The NCVS is a national survey effort in which about 60,000 people aged 12 and over are interviewed twice during a year's period to learn about crime and victims of crime. It generates estimates of victimization rates for the violent crimes of rape or sexual assault, robbery, and aggravated and simple assault and the property crimes of burglary, motor vehicle theft, and household theft. It is sponsored by the United States Department of Justice Bureau of Justice Statistics (BJS) (BJS 2010b).

One might think that when a crime is committed, it is only natural to report it to the police. However, upon examination, it becomes obvious that this is not the case. In fact, it seems there are often more reasons for not reporting crimes than for reporting them. The NCVS is the only major national program that gives a clear indicator from the victim's perspective as to whether or not crimes are reported to police, and it shows that fewer than half of the visible crimes that it tracks are actually reported. When the self-identified victims of those crimes were asked why they did not report to the police crimes that were committed against them, their reasons included the crime being a personal or private matter, the severity of the offense, fear of reprisal, and to protect the offender (Hart and Rennison 2003).

For the remaining four major crime categories (victimless, organized, occupational, and political), it is much more difficult to determine the number of crimes committed, as there are no national crime statistics programs that document them, and there are reasons for this. With regard to victimless crimes, for example, many of the people involved are willing participants (e.g., drug dealers and users) who are typically strongly motivated to not report their crimes to the police or anyone else for fear of incriminating themselves. Often, in organized crimes involving gangs, (city gangs, the Mafia, etc.), even when there are clear victims, such as in extortion cases, the victims may not report crimes because of the very real threats of reprisal made by the gangs. Occupational crimes committed in corporations (cybercrime, industrial espionage, internal fraud, tax evasion, etc.) may not be reported for fear of damaging corporate reputations and the potential harm that such exposure to the public could have on their business and financial status. Political crimes involving terrorists and spies (treason, espionage, etc.) are often not reported openly to the public for security reasons and also for many of the same reasons that crimes in the other categories may not be reported, including guilty involvement, fear of reprisal, and national reputations (Cole and Smith 2008).

Issues

It is clear that most crimes committed in the United States never become known to the police and are never subjected to police investigations. Therefore, one of the primary predictors of the successful investigation of a crime is that it first becomes known to the police.

It seems that the decision of individuals to report crime is often based on whether or not they adjudge it to be in their best self-interest to do so. If they, as victims or community members, are motivated to seek justice, or to maintain or regain freedom from crime, they will likely report crime to the police to achieve those ends. However, if they view reporting crime as a threat to their personal interests or status, then they will likely not seek recourse or justice through the criminal justice system (May et al. 2008).

This theme of how community vs. individual interests influences the reporting of crime information to the police is mirrored in how information is processed through the criminal justice system. Two models have been used to describe the process – crime control and due process. While both models emphasize personal freedom, crime control applies an assembly line approach to repress crime by subjecting criminal matters to a routine, informal administrative and discretionary process in order to maximize disposition. This includes filtering out cases that do not contain sufficient information to act upon. This model essentially represents the community interest in maintaining freedom by dealing with as many crime problems as possible in a practical and efficient way. Due process, on the other hand, emphasizes individual interests and liberty. It advocates an obstacle course approach that relies on legal and procedural restraints to test the reliability of evidence regarding criminal matters in a formalized adversarial process. Its aim is to protect the innocent from undue persecution, even at the expense of allowing the guilty to not be held to account (Cole and Smith 2008). As will be seen in the following sections, the American criminal justice system appears to operate generally along the lines of the crime control model, but with increasingly more due process constraints built into it as an accused person is processed through it.

Does the Crime Survive the Case Screening Process of the Police?

If most crimes are not reported to the police, then how effectively do the police deal with the ones that are in fact reported to them? In addressing this question, the UCR, mentioned previously, provides some answers. It should be noted here that the terms “agency” and “agencies” are used in this entry to refer to governmental law enforcement organizations in the United States in general, and the term “police” includes state, county, and municipal law enforcement agencies.

The UCR is a compilation of data reported by more than 18,000 police agencies regarding the types and numbers of crimes reported to them. Police document in the UCR 28 categories of crimes reported to them. However, they report the crimes that they subsequently resolve, or clear, for only eight categories. A “cleared” crime, or “clearance,” generally means that an offender has been arrested, charged with the commission of an offense, and turned over to the courts for prosecution. The eight crimes for which police report clearance rates are the violent crimes of murder, forcible rape, robbery, and aggravated assault and the property crimes of burglary, larceny, motor vehicle theft, and arson. These eight crimes are known as the UCR Index crimes, and they overlap to some extent with the six crimes tracked by the NCVS. However, whereas the NCVS counts crimes that victims say were committed against them, the UCR reflects only the crimes that police say are reported to them, and the UCR Index crimes reflect the numbers of those reported crimes that the police say they cleared. The UCR is published annually by the Federal Bureau of Investigation (FBI) (FBI 2010).

The UCR data help to understand the police case screening process. When a crime is reported to the police, a police patrol officer usually responds to the complainant or crime scene and conducts a preliminary investigation. The term "patrol officer" is used generically here and includes law enforcement officers (police officers, sheriff deputies, state troopers, etc.) whose primary duties are related to patrol. Based on the preliminary investigation, the reported crime is documented in the agency UCR crime report.

If the crime is a non-Index crime, its further disposition is not documented in the UCR. If the crime is an Index crime and the patrol officer's preliminary investigation clears it, then it is counted in the UCR as a cleared crime. If sufficient information is not developed during the preliminary investigation to clear a crime, it will usually be referred to investigators to determine whether a follow-up investigation is warranted. Detectives in about 60 % of agencies use written criteria called case solvability factors as guides to determine this. (The terms "investigator" and "detective" are used interchangeably.) If the case is "screened" out (not further pursued) or if it is pursued in a follow-up investigation but not resolved, the likelihood of it ever being solved is minimal. If the case is pursued and subsequently cleared, it is documented as such in the UCR. According to the UCR, the police investigation and case screening process result in clearing only about 20 % of all the reported Index crimes (BJS 1988; Eck 1983; Greenwood et al. 1977; Maguire and Pastore 2001).

The seminal study on the police criminal investigation process, called the Rand Report, was conducted during the 1970s (Greenwood et al. 1977). Although it surveyed only about 300 large municipal and sheriff agencies, it found that investigators comprised an average of 17 % of an agency's sworn officers and that they generally investigated only the more serious crimes. It also found that the primary investigative task of investigators was talking to people and that they actually spent less than one-third of their time investigating unsolved cases. They spent the remainder of their time on non-case (45 %) and post-arrest activities (26 %).

During the 1970s and 1980s, a number of other research studies were conducted in an effort to improve agency investigative performance. The following points represent the general consensus regarding what they found:

- The public is by far the largest source of crime information for the police.
- The police detect by themselves only about 5 % of the Index crimes they deal with.
- The patrol officer plays a critical role in the investigation process, in that the preliminary investigation either clears the case or determines whether the crime will be screened out or further investigated.
- More than 80 % of the cases cleared by the police are the result of the identification of perpetrators when a crime is initially reported, on-scene arrests made by patrol officers, other patrol officer investigative activity, or information provided spontaneously by the public.
- Detectives play a critical role in conducting follow-up investigations and post-arrest activities, and many of their duties require highly specialized skills.
- The amount of information available to police about reported crimes is mainly a function of the type of crime and the circumstances in which they occur.
- For most reported crimes (i.e., burglaries, larcenies), the amount of information that can be provided by the public to responding patrol officers is very low, and therefore, the capacity of the police to solve crime based on the reported circumstances is generally quite limited.
- Physical evidence is collected in less than 10 % of police investigations, and much of the collected evidence is never submitted for analysis. Thus, in the great majority of cases, physical evidence plays a relatively minor role.
- A number of investigative reforms to improve agency clearance rates were tested in field studies since the 1970s, but they showed at best only a marginal effect, as investigators had only a relatively minor impact on agency arrest and clearance rates.
- Case resolution rates are much more heavily influenced by cooperation between citizens and the police and patrol officer activities.

The research essentially agreed on the relative ineffectiveness of what has become known as the traditional, reactive, and case-oriented style of police investigation (respond to reports of crime, conduct preliminary investigation, screen out cases with insufficient information, conduct follow-up investigation as appropriate, and either clear or cease further investigation – hereinafter referred to as the "traditional" investigation style) (Eck 1983; Greenwood et al. 1977; Horvath and Meesig 1996; Skogan and Antunes 1979).

Over the past four decades, the average UCR Index crime rates have fluctuated only about 2 % (from lows of just under 4,000 crimes per 100,000 inhabitants in the 1970s to highs of over 5,900 per 100,000 in the 1990s) and, in recent years, have hovered below the 1970s crime rates. Additionally, average clearance rates have consistently ranged about 2.6 % around a 20 % average, despite changes in the nature of crime, the police, and our society (FBI 2010; Maguire and Pastore 2001). This suggests that successful (defined here as cleared) investigations may be influenced at least as much by the traditional investigation style itself as other factors. This lends support to the Rand Report and other research that the amount of information available regarding reported crimes is normally very low and that the constrained traditional investigation style produces consistent case outcomes over time.

Issues

If the first major predictor of a successful investigation is a crime becoming known to the police, then clearly the second major predictor is for the crime to survive the police case screening process. Together they exclude from consideration by the rest of the criminal justice system (prosecution, courts, corrections) the great bulk of the visible crimes committed in our society. This is in addition to the unknown amounts of victimless, organizational, occupational, and political crimes that are eliminated because they are never reported to, investigated, or cleared by the police.

The police case screening process illustrates to some degree how the community and crime control interests vs. the individual due process interests are applied in the criminal justice system, and how they influence investigative outcomes. Cases and offenders are subjected to a winnowing screening process that supports community interests in repressing crime by routinely producing cleared cases that will best support a prosecution. While these cleared cases may be considered successful investigations, the system at the same time protects individuals by screening out most of the cases due to the lack of sufficient evidence. While these might in turn be considered unsuccessful investigations, they do provide for an orderly way to safeguard individual rights. To better understand how this has impacted the police, it is worth taking a brief look at how and why the police investigation process has evolved in the United States.

What Factors Affect the Police Investigation?

In the early nineteenth century, the powerful forces unleashed by the Industrial Revolution brought with them burgeoning problems of social disorder in the dynamic and expanding cities of the United States. Based on the general model of local government police organizations that were formed in England to deal with similar problems in the 1830 s, new public city police forces began to appear in American cities in the 1840s. The focus of detectives in these agencies was essentially offender-oriented, in that their job was mainly to associate with the criminal underworld, either covertly or openly, in order to obtain information about criminal activities.

Described as Secretive Rogues, they often worked proactively with minimal central control or supervision. Corruption, abuse of authority, privacy issues, and political influence were common problems in police agencies, including among the Secretive Rogue detectives. This has been described as the "Political Era" of policing (Kuykendall 1986; Langworthy and Travis 1999).

In the early twentieth century, the increasing public outcry against widespread government corruption and abuse eventually brought about broad societal change, including the reorganization and centralization of the police. The resulting new "Reform Era" of policing "professionalized" the police by reducing political influence, corruption, and abuse and by increasing police accountability (Trojanowicz and Carter 1988).

Detectives also experienced changes regarding how they performed their duties during this era. Initially, the Secretive Rogue proactive offender-oriented style was restructured into an Inquisitor style of reactive case-oriented investigations. Detective autonomy and access to the underworld was constrained by closer supervisory controls and, rather than working to detect crime, they were assigned to investigate criminal cases that had already been committed and reported to the police. In order to compensate for the loss of their own direct contacts with criminals in the underworld, detectives often resorted to the inquisitorial tactics of coercion to develop clandestine informers to collect information for them, and they focused on strong interrogation methods utilizing physical brutality and psychological pressure to elicit confessions from identified suspects (Kuykendall 1986).

By the 1940s, however, the Inquisitor investigator style began to evolve into an administration-oriented Bureaucrat who practiced an even more confined style of investigation. The increasing police controls over investigators and the burgeoning caseloads forced agencies to emphasize stronger and more organized case management in order to permit prosecutors and courts to deal with cases more efficiently. The goal of investigators changed from solving crime by any means to a more bureaucratic, legally guided, and orderly management of information designed primarily to process and dispose of cases, whether they were solved or not (Kuykendall 1986).

In the 1960s, again amid widespread civil unrest and rising crime rates in urban areas, the police once again became the target of strong public disapproval. The police were often viewed as both hostile and repressive, and the professional policing style came under heavy criticism for having isolated the police from the social environment in which they operated. In response to this, the federal government funded a series of research programs to better understand and deal with this issue. This marked the first time that any significant empirical research (including the Rand Report and other studies reviewed above) was conducted in this area (Kuykendall 1986; President's Commission on Law Enforcement and Administration of Justice 1968; Langworthy and Travis 1999).

The "Era of Community Policing" essentially evolved out of much of the research regarding the police/public relationship in the 1970s and 1980s. Community policing has been described as a fundamental shift away from traditional policing and toward a broader, more comprehensive philosophy of crime prevention. It is arguably the dominant model of policing in the United States today, in that at least two-thirds of all local agencies practice some form of community policing. In general terms, it promotes the "Broken Windows" approach by using patrol officers as Community Policing Officers to build close working relationships with the community and to use problem-solving methodologies with community partners to prevent crime. However, the effect of community policing on the traditional investigation style of the Bureaucrat investigator was not clear (BJS 2010a; Trojanowicz and Carter 1988; Wilson and Kelling 1982).

In 2001, the first-ever nationally representative survey of police agencies regarding the police criminal investigation process was conducted to update and expand the Rand Report (Horvath et al. 2001). A total of 3,123 agencies of all sizes were surveyed, and 1,746 (56 %) responded. The respondents employed more than half of the full-time sworn officers (over 350,000) in the country, about 16 % (50,000) of which were investigators. Overall, the survey findings revealed that the police investigation process seemed to have been relatively uninfluenced by changes in policing over the prior 30 years. Although there were some promising developments, the traditional Bureaucrat style of investigations remained predominant. Beyond unique case-specific problems, the overwhelming majority (90 %) of agencies did not identify any legal issues as important problems affecting investigations. This included the use of operational techniques such as searches, informants, undercover activities, sting operations, surveillance, covert listening devices, interviews and interrogations, coercion, and arrests. Further, they did not identify any significant problems regarding their relationships with prosecutors. They reported that the top three factors that would improve their investigation clearance rates were more personnel, technology, and training. They indicated no significant plans to change the investigation process in the future, and they did not project any clear vision of how their role in community policing related to their investigative function.

In 2004, an analysis of the data in the 2001 survey was conducted to determine specifically whether community policing had affected police practices in investigating and solving crime (Meesig 2004). The study found that general community policing practices in agencies have a significant positive effect on murder and robbery, and provided clear support for the effective integration of community policing and investigations to improve police clearance rates. However, as in the 2001 survey, no significant evidence was found of any substantive internal agency management efforts among most agencies to do so.

In addition to community policing, a more recent factor has affected the police investigations process – terrorism. After the destruction of the twin towers of the World Trade Center in New York City, and the attack on the Pentagon in the Washington, D.C. area, by terrorists on September 11, 2001, this has become a national priority. While the act of a terrorist incident itself is designed to take place in a public venue, the planning and preparation for such an attack often occurs in the homes and neighborhoods of our communities. And while federal and military agencies (more than 65 agencies, 85,000 sworn officers) may have primary jurisdiction over many such attacks, it is the people who live in our communities who stand to bear the brunt of most of the attacks, and it is the much larger force of municipal and sheriff agencies (more than 15,000 agencies and 600,000 sworn officers) that are already situated in the communities that have the greatest access to potentially terrorist-related information over the longest periods of time (Meesig et al. 2002). Also, as the primary goal of terrorism investigations is to prevent the terrorist act from occurring, they frequently require intrusive offender-oriented investigative tactics and extensive information collection and collaboration between federal and local agencies (Horvath et al. 2001; Meesig 2004).

To facilitate the investigation of terrorism, various laws have expanded the legal and organizational authorities of agencies to collect and use information to prevent terrorism (Bulzomi 2003). Additionally, federal and local investigators have out of necessity resorted to many of the methods employed in the past by the Secretive Rogue and Inquisitor detectives (covert/undercover activities, the use of clandestine sources, etc.) to neutralize terrorist plots before they can be carried out. This has apparently been done to good effect to date, as at least 40 significant terrorist plots against the United States have reportedly been thwarted since 2001 (Carafano and Zuckerman 2011).

Issues

As can be seen from a historical perspective, a third predictor of successful investigations is an investigation style that is acceptable to both community and individual interests. The various aspects, or factors, of investigation styles that have created controversy in the past have included the corruption, abuse, privacy issues, and political influence that were prevalent in the Secretive Rogue style and the coercion and brutality factors in the Inquisitor style. It appears that by minimizing these factors, in conjunction with implementing less aggressive case-oriented and screening factors, the current traditional Bureaucratic style has become more accepted by community and individual interests.

While community policing and terrorism investigations have demonstrated positive results and support community crime control interests in preventing and repressing crime, they employ more intrusive and offender-oriented methods than the traditional investigation style, and they clearly raise issues regarding the protection of individual rights and due process. One need only recall the NCVS survey findings regarding victims' stated reasons for not reporting crimes (privacy matters, severity of offense, fear of reprisal, offender relationships, etc.) to wonder how long the public will continue to tolerate these offender-oriented investigations in the future, even though they are legally sanctioned.

What Is a Successful Investigation?

How one describes a successful investigation can depend on one's view. In this section, several different perspectives of success are examined.

An unfounded crime is essentially a report to the police that a crime occurred, but subsequent investigation determines that the crime actually did not occur. The investigation is successful in that it resolved the false crime report. However, it is generally not counted as a success because it is not counted in the UCR (FBI 2010).

Whereas the term unfounded focuses on the status of a crime, the term cleared focuses more broadly on the status of a person. As previously mentioned, a reported crime is documented as cleared in the UCR if at least one person has been arrested, charged with the crime, and turned over to the court for prosecution (FBI 2010). Cleared crimes are commonly considered to be successful investigations, but there are some definitional issues:

- The UCR only counts the number of crimes that are cleared, not the number of persons arrested.
- In some cases, an offender may have been identified and located and sufficient evidence gathered to support an arrest, but circumstances beyond an agency's control precluded the agency from making the actual arrest (offender's death, extradition issues, etc). In such cases, the crime is considered cleared by exception in UCR reporting, even though no one has been arrested or prosecuted.
- In matters involving a juvenile offender, a case may be considered cleared if the juvenile is cited to appear before juvenile authorities, even though no arrest was made.

So using the UCR definition of a cleared crime as a criterion of a successful investigation, one or more persons may be arrested or not arrested (exceptional and juvenile clearance) and then turned over or not turned over (again exceptional and juvenile clearance) to the courts. That statement may seem a bit confusing, but it illustrates the nuances involved in using clearance rates to identify successful investigations.

Aside from the definitional issues regarding unfounded crimes and clearance rates, there are a number of other issues, three of which are mentioned here:

- Some agencies do not report crimes to the UCR, and others may either accidentally or knowingly report erroneous data (BJS 1999).
- Sometimes key information regarding an offender's identification or arrest warrants is not entered in a timely and reasonable manner into state-wide and national databases that make such information available to law enforcement agencies throughout the country (BJS 1999).
- Technical advances, such as the Automated Fingerprint Identification System (AFIS) and deoxyribonucleic acid (DNA) analysis, which become available after a case is closed without being cleared, may help to identify suspects from evidence previously collected. They may also help to clear people convicted of crimes they did not commit and to identify the actual offenders (Cromett and Thurston-Myster 2005; DNA Institute 2010; Moses et al. 2010).

These reporting- and technology-related issues are often beyond the purview of individual investigators to directly influence, but they can affect whether or not an investigation is considered successfully cleared because they may provide key pieces of information impacting on investigation outcomes.

It can be argued that an investigation should not be considered successful unless it leads to the prosecution of an offender. However, determinations to prosecute can also depend upon circumstances beyond police control. In about one-third of the cleared police investigations, the offenders are juveniles and are referred to juvenile authorities and often not prosecuted.

In the remaining two-thirds of cleared cases, the adult offenders are referred to a prosecutor, but only about 55 % of them are actually accepted for prosecution. The remaining 45 % are not prosecuted for a variety of reasons. For example, the police tend to use what has been described as a "probable cause" standard to refer cases for prosecution. In essence, once they make the determination that there is probable cause (e.g., sufficient information has been developed) to indicate that a specific person is likely to be guilty of a crime, they make the referral. However, a prosecutor is required to prove the guilt of a person "beyond a reasonable doubt" in court and may decide that insufficient evidence existed to meet that higher standard. This is reported to be the situation for about 50 % of referred-but-not-prosecuted cases. Other reasons for non-prosecution may include witness issues (about 20 % of cases) or other factors (BJS 1988; Cole and Smith 2008; Forst 1995).

It can also be argued that the clearance and prosecution of a case are still insufficient to be considered a successful investigation. Rather, success should depend on whether or not an offender is actually convicted of a crime. The majority of cases brought into court by prosecutors result in guilty pleas. However, guilty pleas are often offered based on negotiations to reduce the nature of the charges brought against an offender or for consideration of lighter sentencing decisions, rather than solely on the strength of the investigation. In those very few cases that actually go to a jury or bench trial, most result in convictions. But even they can be appealed and reversed on grounds which may be beyond the control of the investigation process (May et al. 2008; National Center for State Courts 2003).

From a broader perspective, the various filters and outcomes regarding the police investigation process have been described in terms of a criminal justice funnel, as follows (Cole and Smith 2008; May et al. 2008):

- Out of 1,000 serious crimes, only about 500 are reported to the police, and about 100 result in arrests.
- Roughly 35 of the 100 arrests involve juvenile offenders, who are referred to juvenile authorities for disposition.
- About 65 of the 100 arrests are adults that the police refer to prosecutors, and only about 35 are accepted for prosecution.
- Five of the 35 accepted cases are not prosecuted for reasons beyond the control of the prosecutor, and the remaining 30 go to trial.
- Twenty-three of the 30 plead guilty, four are convicted during trial, and three are acquitted.
- Of the 27 convicted offenders, a few may subsequently have their convictions reversed on appeal for legal, technical, or other reasons.

Issues

A fourth predictor of successful investigations is how success is defined. UCR definitions, reporting and technical issues, and prosecution and conviction rates all judge success somewhat differently, and they all include at least some factors that are beyond the ability of investigators to influence. Further, proactive investigations that prevent crimes (e.g., neutralizing serial offenders, drug interdiction, and terrorist activities) are often not included in conventional descriptions of success.

From a community crime control perspective, the criminal justice funnel figures – only 27 convictions (less than 3 %) out of 1,000 crimes committed – may be sobering. However, from the perspective of individuals, the funnel serves as a series of important due process safeguards against the wrongful investigation and conviction of innocent people.

The Future

In sum, four major predictors have been identified as influencing successful police investigations. First, a crime must become known to police, and second, it must not be screened out by them. Third, the investigation style factors used by the police must be legal and accepted by the public. And finally, the different criteria used for evaluating investigative outcomes can alter the definition of investigative success considerably.

The tension between community and individual interests in American society is reflected in the crime control vs. due process description of the criminal justice system and affects the manner in which police investigate crime. The research, clearance rates, and the criminal justice funnel all indicate that the current traditional investigation style is a relatively ineffective approach toward addressing most crimes in the nation, yet the police report no significant problems regarding that style. This suggests a relative balance between community and individual interests regarding the manner in which police deal with crime.

Recent developments regarding community policing and terrorism show that changes to the traditional investigation style can lead to improved Index crime clearance rates and prevent some types of crime. Although the police have used these developments effectively in some areas, they have not fully integrated them into their overall investigative function, and there may be several reasons for this. The changes focus more on the offender-oriented methods that were in the past associated with factors that engendered public acrimony regarding their effects on community and individual interests. Additionally, the downward trend in Index crime rates and the absence of any significant terrorist attacks on American soil during the past few years may tend to minimize the perceived fear of crime by the public over time and heighten public sensitivity to further infringements on personal liberties.

In the past, nationally publicized incidents reflecting negatively on the police have driven public demands to constrain police activities, and it is quite possible that this could occur again in the future (Lasley 1994; Lynch 2001). But, conversely, another catastrophic event like the terrorist attacks of September 11, 2001, could once again suddenly increase demand for even greater community security. In such a situation, it is not hard to imagine an evolution from the Bureaucrat investigator to a more modernized "Community Policing" investigator. This new approach could conceivably expand the traditional investigation style from merely reacting to reported crime to combine a community policing crime prevention approach with more offender-oriented investigation methods. However, while such an approach might improve the overall success of the police criminal investigations process, it would necessarily have to be done in a legally sanctioned and publicly accepted way.

In conclusion, the balance between community and individual interests will probably continue to shape the police criminal investigation process in the future in much the same manner as it has historically. At this point in time, the absence of any significant problems or demand for change indicates that the traditional investigation style will continue as the predominant investigation model. The concept of the Community Policing investigator that was described above will likely remain for the most part relegated to the paperbound confines of research studies, and not become a greater reality on the street until there is a stronger public mandate for it. If and when this should happen, hopefully the guiding research for the development and implementation of more successful investigation strategies will be as prescient going forward as it has been in the past.

Related Entries

- ▶ Communities and the Police
- ▶ Community Policing
- ▶ Criminal Investigative Analysis
- ▶ Evidence-based Policing
- ▶ False Confessions and Police Interrogation
- ▶ FBI Influence on State and Local Police
- ▶ Forensic Science
- ▶ History of Criminal Investigation
- ▶ Information Technology and Police Work
- ▶ Intelligence-Led Policing
- ▶ Interview and Interrogation Methods Effects on Confession Accuracy
- ▶ Managing Innovation of Policing
- ▶ Organizational Change and Police Legitimacy
- ▶ Policing Terrorism and Legitimacy

- ▶ Predictive Policing
- ▶ Problem-oriented Policing
- ▶ Scientific Evidence in Criminal Prosecutions

Recommended Reading and References

Bulzomi MJ (2003) Foreign intelligence surveillance act: before and after the USA patriot act. U.S. Department of Justice, Washington, DC, Report No. NCJ-200902

Bureau of Justice Statistics (1988) Report to the nation on crime and justice. U.S. Department of Justice, Washington, DC, Report No. NCJ-105506

Bureau of Justice Statistics (1999) Bridging the gaps in police crime data. U.S. Department of Justice, Washington, DC, Report No. NCJ-176365

Bureau of Justice Statistics (2010a) Local police departments, 2007: law enforcement management and administrative statistics (LEMAS) (Report No. NCJ-231174). U.S. Department of Justice, Washington, DC, http://bjs.ojp.usdoj.gov/

Bureau of Justice Statistics (2010b) Violent and property crime rates declined in 2009, continuing the trend observed in the last ten years. U.S. Department of Justice, Washington, DC, http://bjs.ojp.usdoj.gov/

Carafano JJ, Zuckerman J (2011) 40 Terror plots foiled since 9/11: combating complacency in the long war on terror. Backgrounder No. 2604. The Heritage Foundation, Washington, DC, http://report.heritage.org/bg2604

Cole GF, Smith CE (2008) The American system of criminal justice, 12th edn. West/Wadsworth, Belmont

Cromett MF, Thurston-Myster SM (2005) Work of an innocence project: forensic science and wrongful conviction (Report No. NCJ-211808). http://www,ncjrs.gov/app/publications/abstract.aspx?ID=233270

DNA Initiative United States (2010) A prosecutor's notebook (Report No. NCJ-219372). U.S. Department of Justice, Washington, DC

Eck JE (1983) Solving crimes: the investigation of burglary and robbery. Police Executive Research Forum, Washington, DC

Federal Bureau of Investigations (2010) Crime in the United States, 2010: FBI releases 2010 crime statistics. U.S. Department of Justice, Washington, DC

Forst B (1995) Prosecution and sentencing. In: Wilson JQ, Joan P (eds) Crime, 2nd edn. ISC Press, San francisco

Greenwood PW, Chaiken JM, Petersilia J (1977) The criminal investigation process. D.C. Heath and Company, Lexington

Hart TC, Rennison C (2003) Reporting crime to the police, 1992-2000 (Report No. NCJ-195710. U.S. Department of Justice, Washington, DC

Horvath F, Meesig R (1996) The criminal investigation process and the role of forensic evidence: a review of empirical findings. J Forensic Sci 41(6):963–969

Horvath F, Meesig R, Lee YH (2001) A national survey of police policies and practices regarding the criminal investigation process: twenty-five years after rand. U.S. Department of Justice, Washington, DC, Report No. NCJ-202902

Kuykendall J (1986) The municipal police detective: an historical analysis. Criminology 24(1):175–201

Langworthy RH, Travis LP III (1999) Policing in America: a balance of forces, 2nd edn. Prentice Hall, Upper Saddle River

Lasley JR (1994) Impact of the Rodney King incident on citizen attitudes toward police. Haynes (the John Randolph) and Dora Haynes Foundation, Los Angeles, Report No. NCJ-148757

Lynch T (2001) No confidence: an unofficial account of the Waco incident, executive summary. Cato Institute, Washington, DC, Report No. NCJ-191599

Maguire K, Pastore AL (eds.) (2001) Sourcebook of criminal justice statistics. http://www.albany.edu./sourcebook

May DC, Minor KI, Ruddell R, Matthews BA (2008) Corrections and the criminal justice system. Jones & Bartlett, Sudbury

Meesig R (2004) An exploratory study regarding the effects of community policing on index crime clearance rates in local agencies with investigators. Michigan State University, East Lansing, Unpublished doctoral dissertation

Meesig R, Lee YH, Horvath F (2002) The police investigation process. In: David L (ed) Encyclopedia of crime and punishment. Sage, Thousand Oaks, pp 503–509, Vol. 2

Moses KR, Higgins P, McCabe M, Probhakar S, Swann S (2010) Fingerprint sourcebook-chapter 6: automated fingerprint identification system (AFIS). National Institute of Justice/NCJRS, Rockville, Report No. NCJ-225326

National Center for State Courts (2003) Examining the work of state courts, 2002: a national perspective from the court statistics project. http://www.ncjrs.gov/app/publications/abstract.aspx?ID=201846

President's Commission on Law Enforcement and Administration of Justice (1968) The challenge of crime in a free society. Avon Books, New York

Skogan WG, Antunes GE (1979) Information, apprehension, and deterrence: exploring the limits of police productivity. J Crim Justice 7:217–241

Trojanowicz R, Carter D (1988) The philosophy and role of community policing. National Neighborhood Foot Patrol Center, Michigan State University, East Lansing

Wilson JQ, Kelling GL (1982) Broken windows: the police and neighborhood safety. Atlantic Mon 38:29–36

Pre-lineup Instructions

- ▶ Eyewitness Research

Prenatal and Postnatal Preventive Interventions Based on Risk Factors

Lynette Feder[1] and Marianne Junger[2]
[1]College of Health and Public Affairs, University of Central Florida, Orlando, FL, USA
[2]Twente University, Enschede, The Netherlands

Synonyms

Prenatal intervention; Prevention

Overview

Research from a wide array of disciplines indicates that specific factors in a child's life are associated with an increased likelihood of delinquency and other poor outcomes. These correlates, collectively known as risk factors, operate in the multiple domains in which the child interacts including his family, schools, peers, and community. These factors have been found to operate not simply in an additive manner but instead cumulatively and interactively, thereby producing higher likelihood, severity, and frequency of negative life outcomes with the increasing numbers of risk factors the child faces. Additionally, research finds that the children from these disadvantaged households demonstrate a high and stable trajectory of disruptive and disturbing behaviors that continue and even escalate into adolescence. As many of these risk factors can be identified prenatally or early in the child's first few years of life, there is the opportunity to intervene preventively. Prevention programs seek to compensate or correct for factors placing the child's developmental course at risk by targeting the child, their caregivers, and/or their communities. This entry provides an overview of three preventive programs implemented during pregnancy, infancy, or early childhood that have been rigorously tested and found effective in lessening the likelihood of these poor outcomes for children deemed to be at high risk. Included are home visitation programs, parent training programs, and early enriched educational programs.

Research from a wide array of disciplines shows surprising similarity in their findings regarding childhood variables associated with a higher likelihood of negative life outcomes. These variables are collectively known as risk factors and have been found to exist in the multiple domains in which a child interacts including his family, peers, school, and community.

Current research has found that many different disorders share the same risk factors. That is, predictors of any one specific negative outcome, such as delinquency, are associated with a wide array of problematic outcomes such as disruptive and defiant behavior, poor school adjustment, academic failure, drug use and alcohol abuse, mental illness, risky sexual behaviors, and suicide, among others. Importantly, these risk factors appear to operate similarly across different racial and cultural groups. Therefore, carefully and well-designed preventive programs have the potential to affect multiple health and behavioral problems simultaneously thereby proving highly cost-effective.

Many of these risk factors can be identified prenatally or early in the child's first few years of life. Longitudinal research has found several factors associated with maternal characteristics and behavior during pregnancy that are related to an increased likelihood of early childhood disruptive behaviors. These include mother's socioeconomic status, young age at first birth, low educational status, and smoking while pregnant. In the months and years following the child's birth, the parent's inability to effectively socialize has been found to be a powerful predictor of a child's long-term poor outcomes.

While research indicates that all young children engage in disruptive and physically aggressive behaviors, most children learn to control these impulses by the time they enter school. However, studies have found that a small percentage continue to demonstrate a high and stable trajectory of disruptive and disturbing behaviors that continue well into adolescence. Data from a large longitudinal study conducted in America (Campbell et al. 2006) found that a problematic trajectory is associated with high risk for poor academic and social functioning. These results have been confirmed by data from six sites in three countries finding that disruptive behaviors in children entering school are one of the best predictors of adolescent and adult criminality, including nonviolent and violent offending (Broidy et al. 2003).

In fact, children coming from households where they have been ineffectively parented are likely to arrive at school with multiple deficits, chief among them an inability to control their behavior as indicated by impulsive, oppositional, and/or aggressive behaviors. These children will then be at high risk for poor classroom outcomes thereby increasing the likelihood that they will experience negative teacher-child interactions. Additionally, their bad behavior will alienate their peers leading to rejection by their classmates. This exclusion from normative peers will serve as another risk factor by further feeding their anger while curtailing opportunities to learn pro-social behaviors.

The escalating spiral of aggressive and aversive interactions with teachers and peers, along with the additional learning difficulties accrued as these children fall further behind, increases the likelihood that they will engage in even more disruptive and inattentive behaviors. This, in turn, makes it more likely that the child will fail at school and either drop out or be expelled. Research has found that failing to graduate exposes the child to a new array of risks outside of the school setting including unemployment, teen parenthood, low income, delinquency, drug use, and alcohol abuse. As such, a troubling school performance is both highly related to conduct problems at school entry and strongly predictive of later and more serious problematic behaviors within as well as outside of the school setting.

This "chain of cumulative continuity" (Moffitt 1993) captures the idea that with each stage the individual diminishes his probabilities of enjoying future legitimate prospects. This relationship between early disruptive behavior and a winnowing of conventional life opportunities has been found in studies using large samples in other countries. Simply put, the child's inability to behave pro-socially provokes dysfunctional transactions with parents, peers, and teachers which then set him upon a path of decreasing opportunities to learn to behave pro-socially. This may largely explain why studies find that correcting bad behaviors is so much more challenging than preventing them.

An understanding of risk factors has obvious implications for criminology. If an individual's antisocial behavior is stable from preschool to adulthood, then there is the ability to look for its roots early in life and based on factors that are present before or soon after birth. In fact, this is the focus of developmental criminology – to study retrospectively and prospectively the earliest factors associated with a high risk for delinquent behavior and its continuation into adulthood. From this perspective, randomized controlled trials with long-term follow-ups embedded in evidence-based preventive interventions occurring prenatally and postnatally are the necessary next step.

As these preventive interventions are based on a risk factor approach, a description of these factors and the developmental sequence associated with these problems is provided first. General categories of preventive interventions that have been implemented prenatally and in the first few years of an individual's life and found effective in changing these negative trajectories are then discussed followed by a brief overview of some of the most widely respected programs that serve as models in each of these categories.

Background Description

Though longitudinal studies from infancy to adolescence are rare, they demonstrate that there is a constellation of correlates associated with a child's long-term unfavorable developmental trajectory. These include mother's antisocial behaviors as indicated by such things as having children at a young age, achieving low educational status, and smoking while pregnant. While disruptive and aggressive behaviors are widespread in the first few years of life, by age three or four, children are learning to regulate their behavior. For those who do not learn, their problematic behaviors serve as a risk marker for later poor outcomes.

The term prevention is used broadly to refer to a wide range of programs provided during pregnancy, infancy, and/or early childhood. Prevention programs typically seek to compensate or correct for factors placing the child at risk for any number of bad outcomes. Alternately, prevention can seek to enhance protective factors already in existence. Either way, these programs target the child, their caregivers, and/or their communities.

Prevention programs have now been implemented in countries around the world and have demonstrated both short- and long-term success in changing the developmental trajectories of those born into high-risk disadvantaged households. Unfortunately, most programs to prevent delinquency intervene only once the disruptive and physically aggressive behaviors and responses to them from parents, schools, and their communities have become entrenched making it less likely for them to succeed.

Many researchers are now finding that risk factors impacted early are more likely to be successfully modified. Recently, researchers have combined developmental studies and experimental interventions with economic modeling to analyze skill formation in young children and its relationship to life outcomes. Results indicate that achievement is determined by cognitive abilities, as well as noncognitive skills like motivation, self-control, and perseverance. These skills influence positive outcomes, such as steady employment and high income, as well as deviant behavior like delinquency, teenage pregnancies, and drug use. Researchers have further found evidence that well-designed preventive interventions can positively impact these adult outcomes thereby proving more cost effective (Cunha and Heckman 2009). These findings have led James Heckman, a Nobel laureate economist, to conclude, “In an era of tight government budgets, it is impractical to consider active investment programs for all persons. The real question is how to use the available funds wisely. The best evidence supports the policy prescription: invest in the very young and improve basic learning and socialization skills” (Heckman 2000, 8).

The categories of the various promising single-component early prevention strategies for reducing risks and/or increasing protective factors, along with a brief description of its intervention, intended outcomes, and effectiveness, are discussed below. This includes (a) home visitation, (b) parent training, and (c) enriched early education. The next section will then provide a state-of-the-art example falling under each category.

Home Visitation. Home visitation programs typically target mothers-to-be and new mothers. They are growing in popularity and presently number in the thousands across the United States. These programs are built upon research indicating that parents play a significant role in their child’s development, most especially prenatally and postnatally. Additionally, home visitation programs have found that the best way to reach disadvantaged new mothers is to bring services to them and their newborns rather than expecting that they will consistently keep office appointments.

While home visitation programs are popular, they do not represent one specific intervention. Instead, they are a strategy for providing services to mothers of young children with the home visitor fulfilling any number of roles including case manager, parent trainer, and/or family consultant. Typically their goal is to improve the child’s well-being by positively affecting pregnancy outcomes and/or the mother’s ability to properly parent her child. Home visitation programs vary in the staff that is used to fulfill this function. Some programs use paid volunteers, others use paraprofessionals, and a few use community health nurses.

Even as home visitation programs are growing in numbers across the world, the research on their effectiveness is inconsistent. One review of six rigorously evaluated home visitation programs found these programs to have some positive impact on the mothers (better parenting practices and improved attitudes and knowledge), but the benefits to children (in terms of their health, development, and rates of abuse and neglect) were elusive (Gomby et al. 1999). Another review of 20 home visitation programs exclusively using nurse visitors revealed significant and positive outcomes in terms of both the child (as measured by mental development, mental health, and physical growth) and mother (as indicated by depressive symptoms, employment, education, and nutrition) (Ciliska et al. 2001). But a systematic review and meta-analysis conducted for the Cochrane Collaboration identifying 11 distinct experimental studies found no significant overall differences in terms of maternal (as indicated by depression, anxiety, stress, parenting skills, or child abuse risk) or child (as measured by preventive health-care visits, psychosocial health, language development, behavior problems, or number of accidental injuries) outcomes (Bennett et al. 2007).

It may be that who delivers the intervention or, alternately, the rigor of the methodology used to evaluate these preventive interventions, which accounts for the differences observed in the effectiveness of home visitation programs.

Parent Training. Those interventions aimed at teaching caregivers to more properly parent are generally referred to as parent training. The rationale underlying these programs is based on research indicating that antisocial youth have parents who engage in negative practices that promote children's bad behavior. Parent training assumes that if poor parenting practices have created antisocial behavior, one only needs to change the parenting style – teach the parent how to correctly supervise, monitor, reward, and non-punitively correct misbehaviors – to accomplish positive behavior changes in the child. These training programs have been applied to infants and toddlers, as well as young children and adolescents.

There are a wide array of different parent training programs though most share the following common characteristics: (a) minimal or no contact between the therapist and the child; (b) didactic instruction in social learning principles; (c) training parents to identify, define, and observe problem behaviors in their children; (d) training parents to effectively respond to their child's problematic behaviors; and (e) providing parents with the opportunities to see and practice these parenting techniques.

Parent training is one of the most thoroughly evaluated interventions for the treatment of antisocial behavior in children and adolescents. Several of these studies have used experimental and quasi-experimental methods in their evaluations of these programs' effectiveness. In three meta-analyses of various parent training programs (Durlak and Wells 1997; Barlow 2000; Serketich and Dumas 1996) and three systematic reviews (Berkowitz and Graziano 1972; Farrington and Welsh 2003; Weisz and Simpson Gray 2008), all but one (Durlak and Wells 1997) indicated positive effects for the children of parents who received parent training versus those in the control condition who did not receive this program. Furthermore, these studies have found treatment effectiveness maintained up to 4.5 years post-intervention.

Finally, evaluation of different treatment characteristics and delivery approaches used in parent training has found that the program typically works better with parents of younger rather than older children and that therapist training and skill, as well as duration of treatment (with shorter interventions less effective), may affect treatment outcomes. There are inconsistent findings as to whether parent training works better in individual versus group sessions, with some arguing that group sessions provide the additional opportunity of receiving support from other parents.

Early Enriched Education. One of the most widely used prevention programs in the United States today is a structured and enriched educational day care or preschool for at-risk children. As has already been discussed, quite a lot is known about children at risk. Many are from families that are themselves at risk. The stressors these families face, due to social and economic disadvantage, may lead to problems in their parenting behavior. This then increases the likelihood that the child will enter school demonstrating disruptive and highly problematic behavior. A person who has not learned to control his behavior will have difficulties in the classroom setting, including problematic relations with teachers and peers. His disruptive and inattentive behavior, along with his growing aversion for school, increases the likelihood of learning difficulties.

Additionally, because of limited interactions with the parent, the child may not have received sufficient cognitive stimulation in the early years leading to his entering school with significant deficiencies. Researchers are increasingly recognizing how critical the first few years of life are for a child's attainment of developmental milestones. So, for instance, entering school with a language deficiency leads to a greater likelihood that the child will fail in school, again placing him at greater risk for a negative school experience and, with that, yet another path towards a problematic life outcome. Given the strong relationship between cognitive and academic impairments and conduct disorder, these deficiencies become a natural target for prevention.

As with the other preventive programs already discussed, these enriched early education programs do not represent one single entity but rather should be viewed as a vehicle for delivering services to children who are at risk of arriving at school with deficits that will interrupt the learning sequence. They are typically delivered in the preschool years though there is great variety in the rigor of the program as well as its duration. Some programs are center based, others are provided out of the home, and some use a combination of the two. Similarly, many of these programs serve the child directly, while others seek to accomplish their goals by targeting the child's caregivers, and still others use both modalities. All, though, are based on research indicating that conception through the first few years of life provides the foundation for long-term physical, mental, and cognitive development. Therefore, they seek to build a strong base starting very young so as to increase the likelihood that the child will not follow a negative developmental trajectory inside as well as outside of the school setting.

Data from a number of earlier studies on enriched preschool programs found that while IQ scores and academic performance might initially increase, this was not maintained over time. However, there were significant and positive long-term social and academic benefits to children attending these programs. And a systematic review and meta-analysis reported by the Cochrane Collaboration found that out-of-home day care for preschool children (note that this did not necessarily include an intensive educational component), singularly or in combination with an additional preventive component, had significant beneficial effects in terms of the child's IQ, school performance, and behavior (Zoritch et al. 2009). Additionally, these beneficial social and cognitive effects were more pronounced for children from high-risk families (Anderson et al. 2003).

The Abecedarian Project specifically evaluated the long-term effects of preschool education for at-risk children. In that evaluation, infants from high-risk families were randomly assigned into an intensive preschool intervention (vs. control condition) and then randomly assigned again into a school-age intervention (or control condition), thereby creating four cohorts who were followed until age 21. Findings indicated that while the school-age intervention only had weak long-term effects, the preschool treatment led to significant and long-lasting differences academically and socially for study children. Specifically, those receiving the intensive early education program had higher cognitive test scores as well as achieving higher reading and math scores in comparison to their control counterparts. Those in the early experimental intervention also attained more education, were more likely to attend a 4-year college or university, and were less likely to become teen parents (Campbell et al. 2002). Finally, while there were no significant differences in rates of employment, experimental children were more likely to be in a skilled job than those in the control condition (Clarke and Campbell 1998).

The policy implications from this review are clear and are very much in line with Heckman's recommendations. That is, we need to invest in the young with well-designed and rigorously evaluated preventive interventions that improve children's basic skills and socialization.

State of the Art

As space is limited, examples of successful programs from each of the three categories are presented below. This is followed by an example of a hybrid program combining these separate preventive intervention components. Each of the programs highlighted has been rigorously tested and found effective when delivered to high-risk young children and/or their families.

Home Visitation. David Olds' Nurse-Family Partnership (NFP) (see also http://www.nursefamilypartnership.org/) provides one of the most highly regarded and rigorously evaluated home visitation programs. This preventive program targets high-risk first-time mothers-to-be from the time they are in their second trimester through the child's second birthday using nurse visitors who provide frequent and regular (weekly and then biweekly) contact with the woman. NFP focuses on maximizing (1) healthy behaviors in mothers-to-be during pregnancy to increase the likelihood of a positive birth outcome, (2) sensitive and caring parenting to increase the mother-child bond and decrease the likelihood of neglect and abuse, and (3) positive life course for the mother to increase her long-term outcomes.

Olds originally tested the NFP program in Elmira, New York, using a randomized controlled trial (RCT) with a large sample of women. Results 4 years post-program completion indicated positive and significant effects for the mothers and their children. More home-visited mothers returned to and graduated from high school, demonstrated higher employment rates, had lower rates of subsequent pregnancies, were more involved with their children, and displayed improved coping around their parenting in comparison to the control women who received treatment as usual. Children from nurse-visited homes demonstrated fewer behavioral problems, lived in homes with fewer hazards, and made fewer emergency room visits in comparison to controls. Unfortunately, no significant differences emerged in terms of rates of child abuse or neglect or child's intellectual functioning (Olds et al. 1988).

Whereas the RCT in Elmira included predominantly white high-risk women living in a rural area, NFP was next tested with a largely African-American high-risk urban population in Memphis, Tennessee. As with Elmira, Olds and his colleagues found some success. Specifically, women visited by nurses had lower pregnancy-induced hypertension, fewer health-care encounters for their children due to injuries or ingestion, and a lower rate of second pregnancies. However, home-visited mothers and their children did not differ significantly from their non-visited counterparts in terms of the children's birth weight, rates of immunization, mental development, and behavioral problems or the mother's education and employment (KItzman et al. 1997).

In their latest trial conducted in Colorado, Olds tested nurse- versus paraprofessional-delivered home visitation, and both were then compared to a no-treatment control group to see if program effects would maintain when NFP was delivered by well-trained non-nurses to a disproportionately Hispanic sample. Whereas effects for women and their children who were visited by paraprofessionals were small and typically not significantly different than those for control women, those visited by nurses continued to show important differences on both maternal (fewer subsequent pregnancies, higher employment rates, and higher rates of interaction with their infants) and child (improved language and emotional development) outcomes. However, there were limitations with what NFP was able to achieve. Nurse-home-visited women and their children did not differ with controls on the mother's educational achievement and use of welfare or the children's temperament and behavioral problems (Olds et al. 2002).

Though the results across these three sites were mixed in terms of outcomes, a 15-year follow-up of the original Elmira sample showed that the children of mothers served in NFP also benefited from this program over time. Specifically, the children of women enrolled in NFP demonstrated significantly lowered rates of running away, cigarette and alcohol use, and arrests and reported having fewer lifetime sexual partners (Olds et al. 1998).

Olds' Nurse-Family Partnership (NFP) is now implemented in more than 100 sites in 31 states in the United States in addition to half a dozen sites worldwide. Today, a national NFP office handles information dissemination and ensures that there is high fidelity to the model. The largest concern with NFP to date continues to be that its effectiveness has yet to be independently evaluated against a no-treatment control group by researchers who are not connected to the program. However, an independent Dutch replication study using an experimental design is currently underway and initial results seem promising (http://www.voorzorg.info/voorzorg/download/20120324_Factsheet_VoorZorg_mrt_2012_final[1].pdf).

Parent Training. Parent-child Interaction Therapy (PCIT) (see also http://pcit.phhp.ufl.edu/Literature.htm) provides an excellent example of a widely used and rigorously evaluated parent training program. Developed by Sheila Eyberg, PCIT has been recommended as a best practice by several state and federal agencies and has now been implemented worldwide. PCIT was initially developed for children between the ages of 2 and 7, but recent evidence has demonstrated its efficacy with children as young as 18 months (Bagner et al. 2010). PCIT has been found to improve the quality of the parent-child relationship, aid parents in developing appropriate child management skills, and significantly decrease parenting stress (Zisser and Eyberg 2010). PCIT is comprised of two distinct phases.

In the Child-Directed Interaction (CDI) phase, the program focuses on establishing a warm relationship between the parent and child by teaching parents to follow the child's lead which helps them become more responsive to their children. It is thought that this will establish a more secure and nurturing relationship from which to move to the second phase of PCIT, the Parent-Directed Interaction (PDI). The goal of PDI is to teach the parent how to set limits, clearly communicate these limits to the child, and then consistently and firmly use discipline when the child does not comply. In both phases of PCIT, the mother is actively involved in the learning process by being coached and provided with moment-to-moment feedback by the psychologist as she plays with her child.

Based on several randomized controlled trials, PCIT has been shown to be an effective intervention with children demonstrating disruptive behavioral problems, children with developmental disabilities, and families from racially and ethnically diverse backgrounds, among others. PCIT has also been shown to be effective in a group format provided in a primary care setting and is currently being tested with infants. A meta-analysis based on nine PCIT experimental and quasi-experimental evaluations (Thomas and Zimmer-Gembeck 2007) reported a moderate to large effect size ($d = 1.21 - 1.57$) leading to the conclusion that PCIT met criteria for a well-established intervention. Importantly, PCIT's effects have been found to maintain up to 6 years post-intervention (Hood and Eyberg 2003). Just as critically, PCIT has been evaluated not only by Eyberg and her colleagues but additionally by others who are not connected to the program's development.

Enriched Early Education. The Early Training Program (ETP) provides one of the first examples of this type of intervention. Begun in 1962 by Susan Gray in a rural town in Tennessee, this preventive intervention was highly influential in gaining support for programs like Head Start. In the summer of 1962, 65 disadvantaged black children (ages 3–4) were enrolled in this study and randomly assigned to either the experimental intervention – a three-summer or two-summer intensive enrichment program combined with weekly visits by a preschool teacher to the child's home to meet the parents in educating her child – or control (treatment as usual) conditions. Their results demonstrated an increase in performance scores initially with most of these differences dissipating by the time the child reached the fourth grade (Gray and Klaus 1970).

As the sample size for this and similar research on early enriched educational programs was small, and the results somewhat confusing, the data from the ETP plus ten other well-known studies were pooled to form the Consortium on Longitudinal Studies. With a fairly low attrition rate (20 %), the Consortium followed these individuals in 1975 (with ETP individuals now between 16 and 17 years of age). Results from the Consortium indicated that IQ scores increased for approximately 3–4 years after the study ended but were not maintained. However, they discovered that these programs had a significant and positive long-term impact. Specifically, the children who received the preventive intervention demonstrated reductions in both special education placement and grade retention, and they demonstrated more positive attitudes towards achievement when compared to their control counterparts (Lazar et al. 1982).

With research surfacing showing a gender gap in benefits accruing to children receiving these early enriched educational programs, the Consortium data was disaggregated to separately study program effects for males and females. Analyses revealed that while cognitive benefits faded over time, there was a clear pattern of treatment effects by gender. Specifically, girls who were exposed to the ETP showed improved rates in high school graduation, college attendance, marital rates, and overall economic well-being. These women also demonstrated lower rates of criminal behavior and drug use. There was limited evidence for positive treatment effects for the men who received this program (Anderson 2008).

Hybrid Preventive Interventions. While there are many hybrid interventions combining two or more of the previously mentioned three program types, like single-component preventive interventions, most have not been rigorously evaluated with follow-up of subjects over a lengthy period of time to validly investigate its effects. However, there are a few notable and noteworthy exceptions.

An excellent example is the Perry Preschool Project (PPP) which randomly assigned 123 disadvantaged African-American preschoolers (3–4 years of age) in Ypsilanti, Michigan, either to the control condition (treatment as usual) or to a high-quality early education program that emphasized the child’s intellectual and social development. In addition to the education component, the Perry Preschool Project also provided home visitation where teachers met weekly with parents to work with them on parenting skills. (Though the Early Training Program had a home visitation component, the preschool teachers did not work on parent training but rather attempted to get parents allied with their teaching goals.) In fact, this combination of an enriched early education for the child combined with a parent training component for the caregivers has been found to be one of the most successful models for preventive interventions.

By age 19, researchers were finding significant benefits for the Perry Preschool Project in terms of both academic and social outcomes. As with the Early Training Project, children in the experimental group originally demonstrated higher IQ scores that diminished with the ending of the program. But other results maintained and even increased over time. Children receiving the preventive intervention had fewer special education placements and improved attitudes towards schooling, better grades, increased rates of high school graduation, and more postsecondary education. Socially, these children had higher rates of employment and self-sufficiency and lower rates of welfare, self-reported misconducts, and arrests (Weikart and Schweinhart 1992).

A follow-up conducted when these individuals were 40 years of age revealed that the social benefits continued to accumulate for those attending this hybrid preventive program. Specifically, these individuals had fewer arrests (including arrests for violent crimes) and were less likely to have a history that included a jail or prison sentence. They were also more likely to have a savings account. As with the ETP, researchers noted that the Perry Preschool Project benefited girls academically and socially more so than boys and that this held true for both short- and long-term results (Schweinhart 2005).

Possible Controversies in the Literature

Most criminologists focus on adolescents or adults who are getting in trouble with the law. This seems logical as these are the individuals who, when disruptive, are perceived as a danger to the public. But prevention scientists instead attend to infants, toddlers, and young children who, though not currently violating legal standards, are at high risk for poor outcomes. As these young children are not currently a threat to the public, given their limited physical abilities to harm others through their disruptive behaviors, some think it foolish to focus interventions on this population. But as we have seen, one of the best predictors of antisocial behavior in adolescents is a child's chronic physical aggression during the preschool years. Adding to this, evaluations of interventions implemented once bad behaviors have begun have shown limited effectiveness. In the end, therefore, criminology may find it fruitful to take a lesson from public health. They have found prevention to be more successful and cost effective than treatment.

Individuals might also be worried that prevention programs, by targeting high-risk youngsters, will negatively label these individuals early on, thereby causing the bad behaviors that it was intended to prevent. However, if these programs are offered as a way to assist parents in parenting, help build strong bonds in families, or ready a child for entrance to school, then individuals are not being labeled as predelinquents. While each of these programs has one or the other aims, we have seen through research how strengthening parenting practices or building a stronger foundation for educational success also significantly lowers the likelihood of a delinquent outcome.

Two current controversies in the criminological literature deal with the methodology that is implemented in prevention science. The first is the use of longitudinal versus cross-sectional data for testing causal hypotheses, and the second is the utilization of randomized controlled trials. Each is discussed in more detail below.

There are some who argue that longitudinal data is unnecessary, as it serves no purpose that cannot be obtained by cross-sectional data. Prevention scientists, on the other hand, argue that longitudinal studies are necessary (though not sufficient) as it allows the study of an individual's development over time. Criminologists have recently recognized the need to go beyond studying the causes of crime exclusively. Instead, they are finding that it is more informative to study the correlates associated with onset, continuation, and desistance of crime. To do this, longitudinal data is necessary as factors associated with onset may not be the same as those correlated with continuation or desistance.

However, while it is understood that longitudinal studies can reveal correlates associated with delinquency's onset, persistence, or desistance, they cannot determine if these correlates are causative. Therefore, prevention scientists also advocate for these interventions to be embedded in randomized controlled trials and for individuals to then be followed over lengthy periods of time. This design would not only demonstrate the intervention's effect on the child's short- and long-term behavior in comparison to the control group; it also allows the testing of causal hypotheses.

The use of randomized controlled trials in criminology has also led to some suggestions that it is unethical to use human subjects to test interventions. But the fact is that these interventions are being implemented all the time – it is just that they are usually being implemented without any control and therefore their true effects cannot be determined. Others argue that it is unethical to deny individuals an intervention based on what amounts to a flip of a coin. However, this misses the point of conducting randomized controlled trials. If the full effects of the intervention were known, then there would be no need to rigorously test it. Those in prevention science note that even well-intended programs can have harmful effects, and therefore not to test these programs using the most rigorous methods possible is truly what should be thought unethical.

Conclusion

While the majority of violence "prevention" programs target adolescent youth, many of these are really corrective interventions. However, we now have a number of longitudinal studies from early childhood to adolescence that clearly point to specific risk factors associated with poor outcomes such as delinquency and crime. These include coming from a background of disadvantage as indicated by having low socioeconomic status and a young and undereducated mother who proves to be deficient in her parenting skills. As many of these variables can be easily measured at the child's birth or shortly after, this provides the opportunity to intervene preventively in the early years. If that chance is missed, these children can be easily screened upon entrance to preschool or kindergarten as they are most likely to display highly impulsive, disruptive, and/or aggressive behaviors. A great deal of research indicates that failure to do so increases the likelihood that the individual will experience deeply troubled and troubling lives once they leave school.

Though the field is undoubtedly in its infancy, there are now specific preventive interventions that have been rigorously tested and replicated and found effective in significantly changing this negative trajectory. In this review of prenatal and postnatal single- and multiple-component preventive interventions representing a wide range of program strategies, better interventions started early in a child's life, used a risk factor approach that intervened in multiple domains of high-risk families, and then rigorously evaluated the program's effectiveness using long follow-up periods so as to assess the full range of effects. There is a need, however, for an additional suggestion. As human behavior is influenced by a multitude of factors (biology, family, school, peer, community, etc.), using a collaborative interdisciplinary approach when addressing prevention is deemed necessary to achieve diversity in expertise and breadth of knowledge. Taken together, these recommendations would go far in building the knowledge base for better addressing the social ills that our society currently faces.

Related Entries

- ▶ Babies Behind Bars
- ▶ Bullying
- ▶ Compliance and Corporate Crime Control
- ▶ Identification Issues in Life Course Criminology
- ▶ Police Family Violence Services
- ▶ Preventive Order: Civil Law
- ▶ Propensity Score Matching
- ▶ Randomized Experiments in Criminology and Criminal Justice
- ▶ Risk Assessment, Classification, and Prediction
- ▶ Role and Function of the Police Manager
- ▶ School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis

Recommended Reading and References

Anderson M (2008) Multiple inference and gender differences in the effects of early intervention: a reevaluation of the abecedarian, Perry Preschool, and early training projects. J Am Stat Assoc 103(484):1481–1495

Anderson LM, Shinn C, Fullilove MT, Scrimshaw SC, Fielding JE, Normand J et al (2003) The effectiveness of early childhood development programs: a systematic review. Am J Prev Med 24(3, Suppl 1):32

Bagner DM, Sheinkopf SJ, Vohr BR, Lester BM (2010) Parenting intervention for externalizing behavior problems in children born premature: an initial examination. J Dev Behav Pediatr 31:209–216

Barlow C (2000) Parent-training programmes for improving maternal psychosocial health (Cochrane Review). The Cochrane Library (Issue 4). Update Software, Oxford

Bennett C, Macdonald GM, Dennis J, Coren E, Patterson J, Astin M, Abbott J (2007) Home-based support for disadvantaged adult mothers. Cochrane Database Syst Rev (3). Art. No.: CD003759. doi:10.1002/14651858.CD003759.pub2

Berkowitz, Graziano (1972) Training parents as behavior therapists: a review. Behav Res Ther 10(4):297–317

Broidy LM, Nagin DS, Tremblay RE, Bates JE, Brame B, Dodge KA et al (2003) Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: a six-site, cross-national study. Dev Psychol 39(2):222–245

Campbell FA, Ramey CT, Pungello E, Sparling J, Miller-Johnson S (2002) Early childhood education: young adult outcomes from the Abecedarian project. Appl Dev Sci 6(1):42–57

Campbell SB, Spieker S, Burchinal M, Poe MD, The NICHD Early Child Care Research Network (2006) Trajectories of aggression from toddlerhood to age 9 predict academic and social functioning through age 12. J Child Psychol Psychiatry 47(8):791–800

Ciliska D, Mastrilli P, Ploeg J, Hayward S, Brunton G, Underwood J (2001) The effectiveness of home visiting as delivery strategy for public health nursing interventions to clients in the prenatal and postnatal period: a systematic review. Prim Health Care Res Dev 2:41–54

Clarke S, Campbell F (1998) Can intervention early prevent crime later? The abecedarian project compared with other programs. Early Child Res Q 13(2): 319–343

Cunha F, Heckman JJ (2009) The economics and psychology of inequality and human development. J Eur Econ Assoc 7:320–364

Durlak J, Wells A (1997) Primary prevention mental health programs for children and adolescents: a meta-analytic review. Am J Community Psychol 25(2): 115–152

Farrington DP, Welsh BC (2003) Family based prevention of offending: a meta analysis. N Z J Criminol 36:127–151

Gomby D, Culross P, Behrman R (1999) Home visiting: recent program evaluations, analysis and recommendations. Future Child 9(1):1–26

Gray SW, Klaus RA (1970) The early training project: a seventh-year report. Child Dev 41(4):909–924

Heckman J (2000) Policies to foster human capital. Res Econ 54:3–56

Hood KK, Eyberg SM (2003) Outcomes of parent-child interaction therapy: mothers' reports of maintenance three to six years after treatment. J Clin Child Adolesc Psychol 32(3):419–429

http://pcit.phhp.ufl.edu/

http://www.excellence-earlychildhood.ca/home.asp

http://www.highscope.org/content.asp?contentid=219

http://www.nursefamilypartnership.org/

http://www.preventionaction.org/people/skills-beget-skills-nobel-laureate-updates-view-early-intervention

http://www.wsipp.wa.gov/default.asp

Kitzman H, Olds D, Henderson C, Hanks C, Cole R, Tatelbaum R, McConnochie K, Sidora K, Luckey D, Shaver D, Engelhardt K, James D, Barnard K (1997) Effect of prenatal and infancy home visitation by nurses on pregnancy outcomes, childhood injuries, and repeated childbearing. J Am Med Assoc 278(8):644–652

Lazar I, Darlington R, Murray H, Royce J, Snipper A, Ramey C (1982) Lasting effects of early education: a report from the consortium for longitudinal studies. Monogr Soc Res Child Dev 47(2/3):1–142

Moffitt T (1993) Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. Psychol Rev 100(4):674–701

Olds DL, Eckenrode J, Henderson CR, Kitzman H, Powers J, Cole R et al (1998) Long term effects of home visitation on maternal life course and child abuse and neglect: fifteen-year follow-up of a randomized trial. J Am Med Assoc 278:637–643

Olds D, Robinson J, O'Brien R, Luckey D, Pettit L, Henderson C, Sheff K, Korfmacher J, Siatt S, Talmi A (2002) Home visiting by paraprofessionals and by nurses: a randomized, controlled trial. Pediatrics 110(3):486–496

Schweinhart LJ (2005) The high/scope Perry preschool study through age 40: summary, conclusions and frequently asked questions. High/Scope Educational Research Foundation, Ypsilanti

Serketich WJ, Dumas JE (1996) The effectiveness of behavioral parent training to modify antisocial behavior in children: a meta-analysis. Behav Ther 27:171–186

Thomas R, Zimmer-Gembeck M (2007) Behavioral outcomes of parent-child interaction therapy and triple P-positive parenting program: a review and meta-anlaysis. J Abnorm Child Psychol 35:475–495

Weikart DP, Schweinhart LJ (1992) High/scope preschool program outcomes. In: McCord J, Tremblay RE (eds) Preventing antisocial behavior: interventions from birth through adolescence. Guilford, New York, pp 67–86

Weisz JR, Simpson Gray J (2008) Evidence-based psychotherapy for children and adolescents: data from the present and a model for the future. Child Adolesc Mental Health 13(2):54–65

Zisser A, Eyberg SM (2010) Treating oppositional behavior in children using parent-child interaction therapy. In: Kazdin AE, Weisz JR (eds) Evidence-based psychotherapies for children and adolescents, 2nd edn. Guilford, New York, pp 179–193

Zoritch B, Roberts I, Oakley A (2009) The health and welfare effects of day-care: a systematic review of randomised controlled trials. Soc Sci Med 47(3): 317–327

Prenatal Intervention

▶ Prenatal and Postnatal Preventive Interventions Based on Risk Factors

Pretrail Detention

▶ History of Bail

Pretrial Release

▶ History of Bail

Preventing of Terror at Shopping Malls

Robert C. Davis
RAND Corporation, Arlington, VA, USA

Synonyms

Emergency response; Malls; Private security; Risk assessment; Terrorism

Overview

For the most part, malls and other soft targets that are part of US homeland security concerns are protected, not by public police but by private security. The events of 9/11 thrust private security officers into a new and important role. This entry examines issues involved in providing security in shopping malls against terrorist attack. It begins by examining the threat environment. It then describes some of the unique problems involved in trying to protect against terrorist attack in a "mass private space" that is open to public access. It discusses two major studies that have examined the preparedness of shopping malls both to prevent and to respond to a terrorist attack. Lastly, it examines how Israel – where the threat of terrorism is far more pervasive than in the USA – has approached the issue of mall security. The entry concludes that malls are better prepared today than prior to 2001 to prevent and respond to terrorist attack. It also argues that, while terrorist attack against a mall is unlikely in the current threat environment, some of the same measures that would be effective against terrorism – risk assessments, drills with emergency responders, counterterrorism training for security staff, and enhanced partnerships with the private sector – would also be effective against random violence, natural disaster, and other emergencies that malls are more likely to face.

Introduction

Since the events of September 11, 2001, security concerns have figured prominently in the national agenda. Government officials and the public now recognize a wider array of potential terrorist targets extending beyond military installations. These "soft targets," or areas with public access, include transit hubs, schools, and mass private spaces like amusement parks and sports arenas.

One type of soft target that has received too little attention is the retail mall. With all the other soft targets that exist (e.g., transit systems, schools, hospitals), why should citizens be concerned about attacks against shopping malls? One reason is that the nature of malls makes them very vulnerable: there are multiple entrances and exits, and they are open to the public. Large numbers of people come and go, making it easy for potential terrorists to blend in unnoticed. Many of the visitors carry large parcels that could hide a bomb or other weapon. There are multiple ways to attack a mall, ranging from automatic weapons to car bombs to bombs placed inside the mall, even to an attack using a biological or chemical agent.

Moreover, the consequences of an attack could be quite serious. In the case of an attack using a biological or chemical agent, or a bomb blast resulting in structural collapse, the casualties could be very high. An attack could also produce insurance and job losses. A coordinated series of attacks against malls would almost certainly result in long-term lost business and serious regional or national economic consequences, as we saw in the airline industry following 9/11.

In fact, malls and the retail sector in general have been attacked in various parts of the world for the past several decades. Israel has experienced or thwarted attacks against malls on ten occasions since the start of the Intifada in the West Bank in the mid-1990s. Countries as disparate as Turkey and Finland have had attacks against malls in recent years. England suffered attacks against retail stores by the Irish Republican Army as far back as the 1970s (Bilefsky and Zimmerman 2002). A RAND

Corporation study identified 62 terrorist attacks against shopping malls between 1998 and 2005 (LaTourette et al. 2006).

In the United States, several terrorist plots against malls have been uncovered by law enforcement:

- In 2003, the FBI arrested a man on charges that he intended to blow up a Columbus, Ohio, shopping center. The man, a Somali immigrant who allegedly traveled to Ethiopia to obtain terrorist training, was a friend of a man convicted of conspiring to blow up the Brooklyn Bridge. The Columbus suspect was sentenced to 10 years in prison (CBS News 2009).
- In 2007, the FBI warned, based on information from a source considered reliable, of a plot by al Qaeda to disrupt the US economy through a series of holiday attacks against shopping malls in Los Angeles and Chicago (Esposito and Walter 2007).
- In 2009, Federal prosecutors charged a 27-year-old Muslim convert in Boston with plotting to kill politicians and wage violent jihad by randomly shooting people at shopping malls (Bryan 2009).

The Challenge of Securing Shopping Malls

Like gated communities and sports arenas, shopping malls are an example of what Stenning and Shearing termed “mass private property” (Stenning and Shearing 1980). These are large tracts of public-access, privately owned space which have traditionally fallen outside of the domain of public police. In these mass private spaces, the line between what is public and private property – and who is responsible for policing public and private space – is fuzzy. Private mall security officers are the first line of defense against criminal behavior and other threats to the public in shopping malls. But, in cases of serious criminal behavior or other serious threats to the public, public police and other first responders have authority once they arrive on the scene so that public police and private security agencies must develop cooperative relationships and emergency plans. Moreover, mall security officers are typically responsible for what goes on in the mall’s common spaces, while individual stores may have their own security personnel and systems in place. The shared responsibility for security means that coordination between the different entities is essential for an effective response to emergencies.

Shopping malls are also a challenge to secure against terrorist attacks because they have multiple entrances and exits. Typically, there are a number of main entry ways plus entry through large anchor department stores. It would be difficult to screen persons coming and going through so many portals. Moreover, to do so would risk turning customers off and nullifying the efforts of mall owners to create an environment that represents a safe and secure getaway from people’s day-to-day worries and problems.

Finally, mall security is challenging because security staff are not well paid and turnover is rampant. There are exceptions, to be sure, but typically mall security officer positions require little education or experience and salaries reflect that. High turnover rates discourage mall owners from investing in extensive training programs for their officers.

Research on Mall Preparedness for Terrorist Attacks

There have been two major empirical studies on the preparedness of shopping malls to prevent and respond to terrorist attack. Washington, D.C.’s Police Foundation conducted an assessment of the level of security in large enclosed shopping malls as well as the associated issues of training and legislation of private security forces (Davis et al. 2006). The core issue addressed in the Police Foundation report was the degree to which malls had become better prepared to respond to terrorist attacks in the aftermath of 9/11. The investigation included surveys with

33 state homeland security advisors to get their views on mall preparedness as well as surveys with 120 security directors of the nation's largest indoor retail malls. The researchers conducted site visits to eight US malls and two Israeli malls to gain greater insight into how they are dealing with security preparedness and response to disasters. They also conducted a state-by-state analysis of legislation regulating the hiring and training of private security.

The report painted a picture of what malls were doing in the areas of risk assessments, preventive measures, emergency preparedness plans, training, and coordination with state and local government. The comprehensive picture that emerged of the state of security in large retail malls suggested that (a) there were significant gaps in preparedness, (b) there were relatively inexpensive steps that could be taken to fill those gaps, and (c) state homeland security officials and local police as well as mall owners have a role to play in filling those gaps.

The Police Foundation assessment found that malls have taken some steps to improve security. Most security personnel now get several hours of antiterrorism training. Still, half of the mall security directors answering the survey nonetheless felt that their staff could use more training. The report noted as well that a large majority of mall security directors said they have emergency management plans to define actions that security staff should take in the event of a disaster.

However, the surveys and site visits found that, outside of areas receiving specific Department of Homeland Security grants, very little money is being spent to upgrade security. Only a few states had changed their statutes to require background checks, minimum hiring standards or training, and few malls had upgraded hiring standards of their own accord. Risk assessments were rare and, when they were performed, were instigated by the Department of Homeland Security programs and state homeland security officials. Emergency management plans were too often developed without the participation of local first responders or mall store owners and their security staff. Drills to test the security staff's knowledge of what to do in emergencies – when done at all – were seldom rigorous, seldom done with first responders, and were usually done without clear standards to measure their success. Many malls did not even have plans to limit access to sensitive areas in times of heightened alert. Many state homeland security offices had not taken an active interest in working with large malls to enhance security.

The Police Foundation report made four recommendations, including the following:

- Conduct formal risk assessments and take steps to mitigate known risks on a cost-benefit basis.
- Develop and rehearse detailed and coordinated emergency response plans and involve stakeholders.
- Standardize antiterrorism training courses.
- Enhance partnerships with the public sector.

The RAND study examined physical security approaches to reducing the risk of terrorist attack at commercial shopping centers (LaTourette et al. 2006). Unlike the Police Foundation study that examined both reducing risk of an attack and responding effectively in the aftermath of an attack, the RAND study focused solely on reducing the risk of an attack. It used a modeling approach developed on the basis of information obtained from three shopping centers to identify and prioritize the effectiveness of various security options at increasing the odds of thwarting different types of terrorist attacks. The study was intended to help guide shopping centers in designing and implementing strategies aimed at reducing risk.

The RAND analysis included 17 types of terrorist attacks, ranging from sniper attacks to hostage taking to bombs to biological or chemical attacks. The different types of attacks were rated according to their likelihood and according to the extent of damage they were likely to effect, both based on historical data. The responses examined included implementation of 32 different security measures ranging from installation of bollards to searches of people and vehicles to arming security staff to chemical and biological weapon detectors. Responses were rated in terms of their cost and likely effectiveness in thwarting various types of attacks.

The study found that implementing traditional security approaches such as installation of bollards, searching bags, encouraging reporting of suspicious packages, and searching vehicles were likely to be the most effective ways to reduce the risk and cost of an attack. The report also noted that some of these responses might have the effect of driving some customers away.

How Israeli Malls Handle the Terrorism Threat

In Israel, where the threat of terrorist attack is high, malls have been far more stringent in their approach to terrorism than US malls. The Israeli approach to mall security is based on the concept of concentric circles. At the outermost circle, there are roving patrols of one or two security officers and vehicle inspection points. Bollards and retractable barriers are utilized to restrict the possibility of vehicles being driven through the checkpoints. All vehicles entering mall parking areas and pedestrians are subjected to search by officers equipped with explosive detection technology (Story 2011). In addition, visitors are assertively questioned by security officers in an effort to determine whether they pose a possible security threat.

The Police Foundation study reported that approximately 40 % of the operational budget of Israeli malls was devoted to security, compared to 3–5 % in the USA. The high cost of security was driven, in large part, by government standards regarding numbers of security guards, vehicle checkpoints, and barricades, all of which must be met in order for malls to keep their operating licenses.

In Israel, mall security represents a close partnership of private and public sectors. The district police license all armed and unarmed security candidates. They conduct frequent on-site inspections and observe many of the drills being conducted by the mall security staff. Joint exercises team mall security in drills with the district fire brigade, ambulance system, and the entire police district. In addition, there is a system of transparency and open intelligence sharing between mall security and local law enforcement. Finally, mall security and local law enforcement share interoperable communication systems allowing communication over a shared radio band.

Israeli malls are required under governmental regulation structures to conduct periodic risk assessments, using independent contractors. The experts produce a comprehensive security plan for each mall. Once the plan is completed, it is handed over to the district police for approval.

Israeli malls have comprehensive emergency response plans for various contingencies, also required by the government. Each security officer is given a duty under the plan including a subset of officers assigned to an emergency response team. Plans include preventing outsiders from gaining access to the mall as well as possible evacuation – a decision that is entirely up to the security director. An immediate reaction squad meets in a special control room to monitor and direct security actions until the police arrive.

Malls provide monthly training for all officers, almost entirely focused on recognizing and responding to terror threats. Training is highly repetitive, both to engrain the procedures in the minds of the officers and to counteract the effects of high security staff turnover, a problem as acute in Israel as it is in the USA. One advantage that Israel has is that, while turnover is high, many security officers come to the job with recent military training as a result of the country's policy of compulsory military service.

According to the Police Foundation study, Israeli malls usually conduct about 50 drills per month. The drills range from minor procedural drills to covert drills during which false bombs are planted and attempts are made to bring them into the mall.

For a further comparison of antiterrorist efforts in Israel and the USA, see Hasisi, Alpert, and Flynn (2009).

Conclusion

The level of security in Israeli malls is striking and, in fact, even during the last Intifada, not in any ten bomb attempts did the attacker actually penetrate a shopping mall. Israeli malls clearly have done an excellent job of creating as secure an environment as possible under very difficult circumstances.

There are those who would argue that US malls should take a lesson from Israel and do much more to secure themselves against a terrorist attack. A 2010 article on malls predicted that

> . . .the mall in many ways symbolizes the United States to people across the world, acting as a kind of American Horn of Plenty. . .attacks will come and they will be ugly. (Dunn 2010)

The author's recommended solution to the problem he foresaw was to set up armed patrols of ex-soldiers and police officers or to organize mall workers who own guns to bring their firearms to work.

This seems like a serious over-reaction. A recent industry survey found that just 13 % of shoppers thought that malls were unsafe (Smith 2011). Moreover, a recent National Public Radio story examined the practice of "security interviews" conducted by mall staff of people deemed to be suspicious. The story documented instances where shoppers were visibly upset by the questioning and questioned the practice of sharing this information with local and federal law enforcement agencies (Schulz et al. 2011).

No reasonable person would suggest importing the level of security used in Israeli malls to the USA in the present security environment. Additional terrorist attacks feared after 9/11 have not materialized. People would not stand for queuing to go through metal detectors, and there is no reason to ask them to do so. As a society, we have a strong predisposition against the kind of ethnic profiling that is standard practice in Israeli malls. There is no justification for malls to spend nearly half of their operating budgets on security.

However, while there is no reason to take the extreme security measures that have been adopted in Israeli malls, the level of preparedness in a post-9/11 environment ought to be substantially higher than previously. New programs have been put in place by mall owners since the publication of the 2006 Police Foundation study, including bollards placed at strategic points, expanded use of security cameras, more extensive counterterrorism training of security staff, and emergency preparedness drills conducted with local first responders (Smith 2011).

Even if the possibility of an attack by ideologically inspired terrorists is discounted, the same kinds of actions that would help secure malls against that kind of threat also would largely serve to protect against other disasters as well. Risk assessments, emergency management plans, and drills with local first responders can help mitigate the impact of the sort of random acts of violence by individuals that have occasionally occurred in shopping malls (Heffter et al. 2005), as well as the effects of fires and other natural disasters.

Moreover, as Hasisi et al. (2009) have pointed out, the best plans and strategies will only be effective if they are carried out in active partnership with the community. Here, the USA can take a lesson from Israel, where the police have educated the public in identifying suspicious persons, vehicles, and objects. Public education efforts have also focused on how to behave in case of emergency so they reduce the likelihood of harm to themselves and provide help to the police rather than hindering investigations.

Related Entries

- ▶ CCTV and Crime Prevention
- ▶ Counterterrorism
- ▶ Critical Incidents
- ▶ Effectiveness of Situational Crime Prevention
- ▶ Policing Terrorism and Legitimacy

Recommended Reading and References

Bilefsky D, Zimmerman A (2002) Can shoppers be kept safe? – bomb threat against Ikea in Europe spurs stores, malls to reassess security measures. Wall Str J

Bryan J (2009) Homegrown terror plot to attack U.S. malls? ABC News. http://abcnews.go.com/WN/homegrown-terrorism-man-charged-plot-attack-malls/story?id=8881844. Accessed 21 Sept 2009

CBS News (2009) Feds allege Ohio mall terror plot. http://www.cbsnews.com/stories/2004/06/15/terror/main623119.shtml. Accessed 10 Sept 2009

Davis RC, Ortiz C, Rowe R, Broz J, Rigakos G, Collins P (2006) An assessment of the preparedness of large retail malls to prevent and respond to terrorist attack. The Police Foundation, Washington, DC

Dunn JR (2010) Terror at the mall? American Thinker. www.americanthinker.com/archived-articles/../2010/01/terror_at_the_mall.html

Esposito R, Walter V (2007) Exclusive: FBI: Al Qaeda may strike U.S. shopping malls in LA, Chicago. http://abcnews.go.com/blogs/headlines/2007/11/exclusive-fbi-a/. Accessed 8 Nov 2007

Hasisi B, Alpert G, Flynn D (2009) The impacts of policing terrorism on society: lessons from Israel and the U.S. In: Weisburd D, Feucht T, Hakimi I, Lois M, Perry S (eds) To protect and to serve: policing in the years of terrorism, and beyond. Springer, New York, pp 177–202

Heffter E, Sommerfeld J, Carter M (2005) Man arrested in Tacoma mall shooting. The Seattle Times. http://seattletimes.nwsource.com/html/localnews/2002636254_webmallshooting20.html. Accessed 20 Nov 2005

LaTourette T, Howell DR, Mosher DE, MacDonald J (2006) Reducing terrorism risk at shopping centers: an analysis of potential security options. RAND Corporation, Santa Monica

Schulz GW, Zwerdling D, Becker A, Williams M (2011) Shoppers entangled in war on terrorism. www.npr.org/2011/09/07/140246129/shoppers-entangled-in-war-on-terror. Accessed 7 Sept 2011

Smith D (2011) Prepared for the worst: malls improve technology and interaction with government authorities to deter terrorist attacks. Shopping Centers Today. www.icsc.org/srch/sct/sct1105/feat_security_prepared.php

Stenning P, Shearing C (1980) The quiet revolution: the nature, development, general legal implications of private security in Canada. Crim Law Q 22:220–248

Story DW (2011) How should malls address terrorism? Security Management. www.securitymanagement.com/article/how-should-malls-address-terrorism

Prevention

► Bullying Prevention: Assessing Existing Meta-Evaluations

► Onset of Offending

► Prenatal and Postnatal Preventive Interventions Based on Risk Factors

Preventive Order: Civil Law

Daniel Ohana
The Hebrew University, Jerusalem, Israel
Tel Aviv University, Tel Aviv, Israel

Overview

What legal mechanisms are available to control an individual whose behavior reveals a propensity for harm? For the greater part of the twentieth century, with the exception of civilly committing the mentally ill, the state acted through the criminal sanctioning process to counter actors who presented a risk of criminal conduct. Law enforcement authorities could thus immediately perform an arrest and initiate a prosecution; if a conviction followed, the court could impose sanctions to protect the public, including measures to rehabilitate or incapacitate the offender. Recent decades, however, have been marked by the growing use in Anglo-American jurisdictions of preventive mechanisms beyond the criminal law to cope with the likelihood that an actor will commit a crime. Significantly, these preventive orders are deemed civil rather than criminal in nature and may be granted in respect of an individual who has not actually perpetrated an offence. Furthermore, the state need not meet all of the heightened procedural and evidentiary barriers which apply in a criminal trial. Since the trend of using preventive orders has been especially pronounced in the United Kingdom, the following discussion largely focuses on the developments that have marked the English landscape of criminal justice.

Historical Background

For the greater part of the twentieth century, legislative practice in Anglo-American jurisdictions proceeded on the assumption that, in coping with the risk posed by an individual whose behavior suggests a propensity for harm, preventive measures, such as arrest and detention or the seizure of tools intended to be used for criminal purposes, could only be taken if the actor actually engaged in conduct which has been criminalized. However, the past three decades have been characterized by a growing trend which involves the use of preventive orders in lieu of, or in conjunction with, the machinery of the criminal law, which do not necessarily require the performance of conduct which amounts to a criminal offence. These preventive orders subject individuals who present a risk of harm to judicial control. The court may impose restrictions on movement, contacts with specified individuals, and access to designated materials or tools. These judicial powers are not without antecedents (Ramsay 2012). Common-law courts have historically had the power to "bind over" an individual likely to breach the peace and enjoin her to enter into a recognizance to keep the peace and specify reasonable conditions to secure her good conduct (Williams 1953). Similarly, the traditional civil injunction process granted courts the power to impose preventive measures by requiring individuals who pose a risk of harm to refrain from performing certain actions. So-called "status" offences criminalizing such activities as begging and vagrancy also commonly served as a vehicle to subject to public control individuals whose behavior was believed to evince a criminal proclivity (Williams 1955). Inchoate offences that criminalize preparatory actions and other remote risks of harm now serve a similar preventive function (Dubber 2002), although it should be noted that, in some cases, the criminalization of conduct performed with the intention of laying the groundwork for the commission of a substantive offence may be justified based on retributive principles of justice (Ohana 2007). Modern forerunners of the preventive order can also be traced back to the *ante delictum* security measures which emerged in a number of European countries in the early and middle twentieth century. These measures were championed by members of the "Social defense" movement who were opposed to the classical model of criminal law and punishment guided by the purposes of deterrence and retribution (Digneffe 2008). Viewing crime primarily as a symptom of an "*état dangereux*" (a dangerous condition), they argued that conviction of a criminal offence should not necessarily be required in order to authorize preventive and rehabilitative action by the criminal justice system and that once satisfactory indications were forthcoming that an individual is in a state of dangerousness, steps should be taken to avoid further deterioration in her condition (Ancel 1987). *Ante delictum* security measures failed to fulfill the expectations of their supporters, inter alia, because the measures were exploited by authoritarian regimes to exclude social deviants and repress political opponents (Pradel 2002). A vestige of the *ante delictum* preventive measures can still be found in the Italian system of *misure di prevenzione*, which is now used to control individuals associated with Mafia organizations and terrorist groups (Molinari and Papadia 2002).

In American jurisdictions, the civil injunction has most notably been used in recent decades to protect against spousal abuse (Finn and O'Brien Hylton 1994) and gang-related criminal activity (O'Deane 2012). This legislative practice forms part of a larger trend, whereby a variety of civil remedies, such as civil monetary penalties and civil forfeiture, are pressed into service to address criminal behavior (Cheh 1991). This trend has given rise to the formation of a "middleground jurisprudence" (Mann 1992) between the criminal law and the civil law. In Canadian law, preventive measures, which are not predicated on the commission of a criminal offence, have been statutorily framed as an exercise of the Common-law judicial power to "bind over" an individual likely to breach the peace (Criminal Code of Canada, R.S.C., 1985, c. C-46, sections 810–810.2; R. v. Budreo, 104 C.C.C. (3d) 245). These measures may be imposed on an individual who is likely to engage in spousal or child abuse, a criminal organization or terrorism offence, sexual misconduct towards a minor, or a serious personal injury offence.

Although preventive instruments allowing for the imposition of protective measures in respect of an individual who presents a risk of harm can be found in the legislation of a variety of

Common-law countries, the tendency of using preventive orders in lieu of, or in conjunction with, the machinery of the criminal law has been especially pronounced in the United Kingdom. The preventive order first rose to prominence with the enactment of the Crime and Disorder Act 1998, which empowered courts to extensively restrict the movement of individuals found to have acted in an antisocial manner by subjecting them to an "Anti-Social Behaviour Order." The Prevention of Terrorism Act 2005 provided another impetus for the entrenchment of the preventive order within the English landscape of criminal justice. The Act empowered the Secretary of State for the Home Department to impose a "Control Order" significantly curtailing the activities of individuals suspected of involvement in terrorism-related activity. Both of these preventive orders proved to be controversial. As explained below, these preventive orders have lately been the subject of reform initiatives by lawmaking authorities. In this context, the Prevention of Terrorism Act 2005 (and, with it, the Control Order) was abolished and replaced by the Terrorism Prevention and Investigation Measures Act 2011, which introduced a new preventive mechanism, namely, the Terrorism Prevention and Investigation Measure. Moreover, a program for comprehensive reform of the Anti-Social Behaviour Order was recently announced.

Even with these efforts to reform the preventive orders available to protect against antisocial behavior and terrorism-related activity, legislative schemes authorizing the imposition of a preventive order on an individual due to the likelihood of criminal conduct on her part have become commonplace in the British statute books. Indeed, the scope of the preventive order has greatly expanded as of late. Most tellingly, the Serious Crime Act 2007 significantly widened the range of criminal activities in respect to which a preventive order may be sought. The Act provides that a Serious Crime Prevention Order may be issued to prevent drug trafficking, human trafficking, prostitution, armed robbery, money laundering, fraud, tax evasion, corruption and bribery, blackmail, intellectual property offences, and environmental crimes (Serious Crime Act 2007, Schedule 1). Furthermore, a Serious Crime Prevention Order may also be made to prevent any offence which "in the particular circumstances of the case, the court considers to be sufficiently serious to be treated for the purposes of the application or matter as if it were so specified" (Serious Crime Act 2007, section 2(2)(b)).

Characteristics and Functions of the Preventive Order

According to Andrew Ashworth and Lucia Zedner, "the essence of [preventive] measures is that they involve (i) restrictions on individual liberty of action; (ii) in order to prevent harm or risk of harm and (iii) are backed by threats of coercive sanctions" (Ashworth and Zedner 2010, p. 61). There is a major structural difference between the preventive order and the machinery of the criminal law. The preventive order is explicitly intended to be preventive, not punitive. Although conviction of a criminal offence may constitute a prerequisite for the making of a preventive order, it is fundamentally oriented towards the future. The actor's prior conduct is only taken into consideration to assess the likelihood that she will engage in the commission of a crime. Setting in motion the criminal sanctioning process, on the other hand, hinges on the past conduct of the offender, rather than on her personal dangerousness. Whereas punishment of offenders may fulfill an important function in protecting the public, the fact that the machinery of the criminal law can only be triggered by the commission of an offence endows it with an essentially backward-looking dimension, which does not exist in the preventive order.

A typical legislative scheme which allows for the issuance of a preventive order contains rules pertaining to most or all of the following: the specific harm to be averted, the preventive restrictions available, the conduct prerequisites, the principles of procedure and evidence applicable, the penal consequences of a breach of the order, the competent authority, the parties qualified to file an application, the possibility of obtaining an interim order pending decision on the merits, and the duration, variation, and discharge of the order. There is considerable diversity across statutory arrangements in the specifics of each of these components.

A standard preventive order places a number of obligations on the targeted person, and the violation of these obligations may constitute a criminal offence subject to a maximum penalty of 5 years of imprisonment. Hence, the preventive order has been dubbed a "two-step prohibition" (Simester and Von Hirsch 2006, p. 174). As onerous as some of the obligations may be, they involve restrictions on, rather than deprivations of, liberty (Ashworth and Zedner 2010). The conditions attached to a preventive order typically enjoin the actor to refrain from engaging in specified forms of activity, such as accessing certain areas, coming into contact with potential accomplices or victims, and possessing designated instruments. For instance, the Sexual Offences Act 2003 empowers a court to grant a Sexual Offences Prohibition Order that might prohibit having unsupervised contact with anyone under the age of 18, being present in certain places such as schools or playgrounds or accessing the Internet. Significantly, the court must be satisfied that "it is *necessary* to make an order for the purpose of protecting the public or any particular members of the public from serious sexual harm from the defendant" (Sexual Offences Act 2003, section 104, emphasis added). Some statutes set a lower threshold: Under the Serious Crime Act 2007, a Serious Crime Prevention Order may be issued restricting the person's place of residence, working arrangements, business dealings, access to specified premises, travel, and means of communication and association with others (Serious Crime Act 2007, section 5) provided that the court has "*reasonable grounds to believe* that the order *would* protect the public by preventing, restricting or disrupting involvement by the person in serious crime" (Serious Crime Act 2007, section 1(1)(a)) (emphasis added).

Since the issuance of a preventive order may result in stringent restrictions of liberty, the function of adjudicating the application process is confined in principle to the judiciary. Some legislative schemes, however, provide for special arrangements. The Terrorism Prevention and Investigation Measures Act 2011 gives the Secretary of State the power to *decide* to impose a Terrorism Prevention and Investigation Measure (hereinafter, TPIM) on an individual who is involved in terrorism-related activity. Before imposing the measure, the Secretary of State must obtain the permission of the court, except in cases of urgency, where the measure must be immediately referred to the court for confirmation. The court may refuse permission if the Secretary of State's decision is "obviously flawed" (section 6(3)(a)). In practice, the courts subject the decision of the Secretary of State to a relatively high level of scrutiny both with respect to the finding that the individual concerned is involved in terrorism-related activity and with respect to the necessity for each of the restrictions imposed to prevent further such activity (Walker and Horne 2012).

Under nearly all legislative schemes governing the making of a preventive order, it is required that the applicant prove that the person concerned engaged in a course of conduct defined in the statute. There are some preventive orders, such as the Sexual Offences Prohibition Order under the Sexual Offences Act 2003, that may almost only be imposed if the actor was convicted of a designated offence (Sexual Offences Act 2003, section 104). Under a number of statutes, inappropriate conduct that does not necessarily amount to a criminal offence may provide grounds for the issuance of a preventive order. Under the Sexual Offences Act 2003, a Risk of Sexual Harm Order can be made against an individual who has, on at least two occasions, engaged in such conduct as sexual activity involving a child or in the presence of a child; causing or inciting a child to watch a person engaging in sexual activity or to look at a moving or still image that is sexual; giving a child anything that relates to sexual activity or contains a reference to such activity; and communicating with a child, where any part of the communication is sexual (Sexual Offences Act 2003, section 123(3)).

Since the application proceeding for a preventive order is classified as civil, the hearsay rule does not apply as a rule. Indeed, the fact that hearsay evidence may be used in application proceedings for an Anti-Social Behaviour Order is viewed as representing a significant advantage for public authorities, since it enables them to overcome the problem of witness intimidation. That the preventive order is viewed as a civil mechanism also entails that the criminal ("beyond a reasonable doubt") standard of proof does not need to be met. The Terrorism Prevention and Investigation Measures Act 2011 provides for a relatively low evidentiary threshold: It is sufficient that the Secretary of State "reasonably believes" that the individual concerned is involved in terrorism-related activity (Terrorism Prevention and Investigation Measures Act 2011, section 3) (Walker and Horne 2012). The vast majority of preventive measures rest on a civil proof standard ("balance of probabilities"). The House of Lords, however, ruled in *R* (*McCann and Others*) v *Manchester Crown Court* [2002] 4 All E.R. 593 that, although the application proceeding for an Anti-Social Behaviour Order should be classified as civil, it must be proven beyond a reasonable doubt that the individual acted in an antisocial manner.

To conclude, it is not always the case that the application process for the issuance of a preventive order may only be initiated by the police or the prosecution authorities. Some legislative schemes allow other state and non-state actors to set the process in motion. This is not surprising, perhaps, given that the protective measures involved are deemed civil rather than criminal in nature. For instance, the Crime and Disorder Act 1998, which lays down the legislative framework for the imposition of an "Anti-Social Behaviour Order" on an individual who acts in an antisocial manner, specifies that this preventive order may be sought by the council for a local government area as well as by a person who by statute provides or manages any houses or hostels in a local government area (Crime and Disorder Act 1998, section 1A). Under some statutes, a private individual may seek a preventive order. Thus, the Family Law Act 1996 states that any person may apply for a non-molestation order prohibiting another person previously associated with her in marriage, civil partnership, cohabitation, or an intimate physical relationship of significant duration from harassing the applicant or a relevant child (Family Law Act 1996, section 42).

Current Issues and Controversies

The preventive order has proven to be a controversial mechanism of legal control amongst criminal law theorists and criminologists. Although commentators generally agree that the state has the obligation to take steps to protect citizens from an individual who poses a risk of harm, they argue that when the state resorts to the preventive order, rather than the criminal law, to achieve security, it does not give sufficient weight to the autonomy and liberty of the individual concerned and to her expectations of legal certainty. It is also argued that the preventive order does not provide for enough procedural safeguards and that it undermines the principle of proportionality in punishment and the separation of powers doctrine. As explained below, various aspects of this preventive instrument have provoked concerns and trenchant criticisms, based on considerations of both principle and policy.

First, it has been argued that the preventive order is inconsistent with the principle that the state should treat its citizens as responsible agents, because it allows for restrictions to be set on an individual based on a prediction about her future conduct, thereby violating her freedom to autonomously decide whether to abstain or not from breaking the law. A similar claim has been made from a communitarian standpoint, namely, that the preventive order runs counter to a mode of governance which befits dealings between fellow members of the polity, because it unjustifiably singles out a given actor as an untrustworthy citizen, whose commitment to the values underpinning the norms of the criminal law is put in doubt (Duff and Marshall 2006; Simester and Von Hirsch 2006).

Second, the hybrid character of the preventive order has been attacked for failing to concur with a liberal paradigm of criminal justice, in which respect for the rights of suspects and defendants constitutes a central value. Since it is categorized as a measure that is civil rather than penal in nature, the preventive order enables the state to impose intrusive restrictions while circumventing principles of due process that apply as a rule in a criminal trial. Indeed, "[t]he principal reason why some governments have been devising forms of civil proceedings. . .is that the criminal justice system is thought to present too many barriers to swift and effective outcomes" (Ashworth 2004, p. 273). It has been argued that an application for a preventive order should not be classified as a civil proceeding simply because the order is made with preventive rather than punitive objectives. Rather, attention should also focus on the impact and potentially detrimental consequences of the prohibitions attached to an order for the dignity and freedom of movement of the individual concerned (Ashworth 2004). The boundaries between the civil and the criminal law are not only blurred by the fact that the restrictions which attach to a preventive order may be onerous but also by the fact that a breach of the order may constitute a criminal offence (Simester and Von Hirsch 2006). Given the stringent character of some of the restrictions which may be imposed on an individual targeted by a preventive order, some commentators have argued that the core procedural protections guaranteed to defendants in a criminal trial should also apply in a proceeding leading to the making of such an order, even if the proceeding is formally categorized as civil (Ashworth and Zedner 2010).

Third, the preventive order has been criticized for not according with a desert-centered conception of criminal liability and punishment. Specifically, critics note that a violation of the terms of a preventive order without reasonable excuse suffices to establish liability for a breach and that no evidence need be adduced to prove the mental state of the defendant (Simester and Von Hirsch 2006). It is also argued that the issuance of a preventive order can generate outcomes which are at odds with the principle of proportionality in punishment, since restrictions may be imposed on an offender even after completion of sentence if these are deemed necessary to protect the public (Ohana 2010). The fact that, as noted below, in practice courts sometimes impose disproportionately harsh sanctions (including imprisonment) in response to a breach strengthens this objection.

Fourth, it has been argued that unduly harsh criminal sanctions are imposed for a violation of the terms of a preventive order. Significantly, the breach of a preventive order is often threatened with a maximum penalty of 5 years imprisonment. The preventive order has been likened to a Trojan horse, with the civil law serving as a vehicle for setting in motion the machinery of the criminal law once a breach has taken place (Simester and Von Hirsch 2006). Given that the prohibitions can be wide-ranging and poorly targeted, and that little rehabilitative assistance is provided to the individual concerned, it has been argued that the preventive order involves significant potential for "net-widening" – that is, it risks entangling actors who would not otherwise have been prosecuted for dangerous or harmful conduct constitutive of a criminal offence (Crawford 2009). This criticism has been raised most forcefully in respect of the Anti-Social Behaviour Order. Research has found that the prohibitions contained in Anti-Social Behaviour Orders made by magistrates are often loosely defined, and thus not only often entirely bar the person concerned from a wide range of public areas, such as parks and shopping malls, but also heighten the risk of a breach later on (Von Hirsch and Shearing 2000; Hoffman and Macdonald 2010). More importantly, approximately half of all Anti-Social Behaviour Orders are breached, and approximately half of those convicted receive a custodial sanction (Home Office 2010). These findings reveal that the Anti-Social Behaviour Order is not entirely successful in its preventative function. More generally, these findings reveal how the preventive order can blur the boundaries between the civil and criminal law and detract from the potency of the condemnatory message and sanctions of the criminal law and from its standing as a distinct moral voice within the community.

Fifth, critics have taken issue with the imprecision which often characterizes the conduct prerequisite in some of the statutes authorizing the imposition of a preventive order (Ashworth 2004). Specifically, it is argued that there are often vague statutory formulations, which run afoul of the principle of legal certainty, failing to provide citizens the guidance needed to steer themselves in such a way as to remain free of state intervention. For instance, the Crime and Disorder Act 1998 defines antisocial behavior for the purpose of obtaining an Anti-Social Behaviour Order as acting in "a manner which causes or is likely to cause harassment, alarm or distress to one or more persons not of the same household" as the actor concerned (Crime and Disorder Act 1998, section 1(1)(a)). The Serious Crime Prevention Order provides another example of the ways in which, as Lucia Zedner has observed, "uncertainty now provides the grounds for action against individuals who are believed, but not known, to be involved in harmful behaviours[.]" (Zedner 2009, p. 51) It may be made against an individual who has been involved in serious crime, which is defined as including cases where the individual has conducted herself in such a way *that was likely to* facilitate the commission by herself or another person of a serious offence. In assessing whether the conduct prerequisite is established, the court is to disregard the intentions of the actor at the time of the act in question, as well as acts which are shown to have been reasonable under the circumstances (Serious Crime Act 2007, sections 4(2)-(3)).

Sixth, it has been argued that the proliferation of preventive orders within the English landscape of criminal justice is problematic because unelected judges are given the power to proclaim which norms of conduct must be complied with, on pain of criminal sanction for failure to do so, in the absence of detailed legislative guidance (Ashworth and Zedner 2010). This arguably runs counter to the fundamental principle of democratic governance, according to which the power to prohibit conduct by means of the criminal law should reside in the legislature, which is composed of the elected representatives of the citizenry and which offers a forum for public debate (Simester and Von Hirsch 2006).

Reforms and Proposals

Following this review of the objections leveled at the preventive order, it should be noted that although academic experts in the criminal justice field have faulted the preventive order on many grounds, it has not encountered significant hostility amongst actors within the criminal justice system, such as policymakers, judges, law enforcement officials, and local governmental offices (Ramsay 2009). Moreover, as mentioned previously, in recent years, steps have been taken towards reforming some varieties of the preventive order while taking into account a number of relevant criticisms. The Terrorism Prevention and Investigation Measures Act 2011 abolished the controversial "Control Order" which had been introduced under the Prevention of Terrorism Act 2005 and replaced it with a new system of TPIMs. This new system arguably represents an attempt to strike a more measured balance between counterterrorism security measures and individual liberties by tightening the normative framework for imposing preventive restrictions on persons suspected of involvement in terrorism (Walker and Horne 2012). The new Act requires that the Secretary of State have a reasonable *belief* that the individual is or has been involved in terrorism-related activity, whereas earlier, only a reasonable *suspicion* was needed to make a Control Order. And whereas Control Orders could previously remain in force for 12 months and be renewed an indefinite number of times, TPIMs are now subject to a 2-year time limit, unless there is new evidence of further involvement in terrorism-related activity. The Terrorism Prevention and Investigation Measures Act 2011 is similarly distinguished from its predecessor by exhaustively listing the prohibitions, restrictions, and requirements which may attach to a TPIM. In some respects, the measures have been made less onerous. Most notably, 16-h curfews have been replaced with overnight residence requirements.

There has also been a recent governmental proposal to reform the Anti-Social Behaviour Order. A White Paper published by the Home Secretary in May 2012 aims to enhance the effectiveness of the legal powers available to respond to antisocial behavior. The most far-reaching proposal is to abolish the criminal sanction for a breach and relabel the Anti-Social Behaviour Order as a “Crime Prevention Injunction” (Home Office 2012). This proposal differs importantly from current law in that positive requirements as well as prohibitions could be attached to the order. Specifically, positive obligations could require the offender to participate in rehabilitative programs to address underlying causes of antisocial behavior, such as drug dependency. In addition, only a civil, rather than criminal, standard of proof would have to be met within application proceedings. This change in the law would have the effect of offsetting the previously mentioned ruling of the House of Lords in the *McCann* case that, although application proceedings for an Anti-Social Behaviour Order are civil in nature, it must be established beyond a reasonable doubt that the individual engaged in antisocial behavior. A breach of the Crime Prevention Injunction would be subject to civil sanctions by triggering liability for civil contempt. Lastly, the Crime Prevention Injunction would exist alongside a “Criminal Behaviour Order,” the breach of which would be subject to criminal sanctions. This preventive order would be reserved for actors convicted of an offence and would be imposed at sentencing.

Although some criminal law theorists have called for the outright abolition of the Anti-Social Behaviour Order (Ramsay 2010), suggestions for reform have been advanced in the academic literature. Andrew Cornford has proposed that the Anti-Social Behaviour Order be reformed by narrowing the remit of the definition of antisocial behavior. In his view, the Anti-Social Behaviour Order should only be available in respect of individuals whose conduct causes others to experience justifiable anxiety about the safety of their local community (Cornford 2011). As regards more generally the appropriate uses which may be made of the preventive order as a mechanism of legal control in lieu of the criminal sanctioning process, Daniel Ohana has examined the possibilities of revamping the preventive order in order to allow for moderate preventive intervention vis-à-vis an actor who has taken concrete steps with the intention to commit a crime (Ohana 2006). He has suggested that the preventive order could serve to avoid the criminalization of minor preparatory offences which are not sufficiently blameworthy to justify the invocation of criminal sanctions, without neglecting public protection concerns. Ohana has argued that the infringement on liberty imposed by a preventive order can be mitigated by following a similar approach to that used by situational crime prevention practitioners, who draw upon empirical “script” (Cornish 1994) analyses of crime-commission processes to provide tailor-made solutions to particular crime problems. More specifically, Ohana has articulated how script studies can be used to pinpoint the restrictions imposed by a preventive order and thus ensure that the hardship suffered by the individual concerned is not disproportionate to the relatively low level of seriousness of the preparatory actions performed and that their ability to lead a reasonably normal life is not unduly hampered.

The Rise of the Preventive Order and the Sociopolitical Context of Late Modernity

Many commentators have inquired why it is that particularly in recent decades the preventive order has attained such a high level of prominence in the English landscape of criminal justice. Examining the wider historical, political, and socioeconomic context of the late twentieth century and early twenty-first century, they have explored and drawn out the various ways in which the preventive order is related to ideas and practices of governance that have gained currency in the present era. Given that the preventive order allows for state intervention to avert harmful behavior well beyond the traditional boundaries of the criminal law, Ashworth and Zedner have identified its ascent as emblematic of the rise of the "preventive state" (Ashworth and Zedner 2008, p. 40; Steiker 1998). This institutional "ideal type" is characterized by a temporal shift in the focus of governance, whereby the state is preoccupied with preventing future harm no less than with punishing past crimes. Preventive orders are infused by the logic of precaution and arguably represent an extension of the so-called precautionary principle (Ewald 2002) which was first developed in the field of environmental protection to address the risk of serious, irreversible damage to natural resources (Crawford 2009; Zedner 2009). According to the precautionary principle, it is incumbent upon public authorities confronted with a possible threat of harm to take action immediately – even where uncertainty prevails as to the nature, intensity, and scope of the risk at hand. In this context, uncertainty concerning whether a source of potential danger will actually prove harmful is no excuse for adopting a posture of nonintervention. On the contrary, active, widespread anticipatory actions are required, so as to ensure safety and avert harm. From this perspective, the rapid growth and entrenchment of the preventive order as a tool of crime control dovetails with the emergence of the "risk society" (Beck 1992) in which institutional goals and strategies are focused on identifying, ordering, and responding to risks and assuaging fear and anxiety by providing assurances that various risks are under control (Ashworth 2004).

Peter Ramsay has argued that the preventive order has found favor with English lawmakers over the past two decades because it serves as a tool for the protection of what he calls "vulnerable autonomy" (Ramsay 2008). More specifically, the preventive order enforces a demand that members of the polity not "fail to reassure" others. Ramsay locates the source of this normative demand to respect vulnerability in political theories which have gained preponderant influence in the United Kingdom, namely, the Third Way, communitarianism, and neoliberalism. Moreover, Ramsay observes that this depiction of citizens as vulnerable "views the law generally and the criminal law in particular, as having failed in relation to anti-social behaviour and fear of crime." Ramsay's thesis echoes Lord Steyn's comments in the *McCann* case, which concerned the Anti-Social Behaviour Order:

> It is well known that. . .young persons, and groups of young persons, cause fear, distress and insecurity to law-abiding and innocent people by outrageous anti-social behaviour. . ..Sometimes the conduct falls short of cognisable criminal offences. . .In recent years this phenomenon became a serious problem. There appeared to be a gap in the law. The criminal law offered insufficient protection to communities. Public confidence in the rule of law was undermined by a not unreasonable view in some communities that the law failed them.

The Anti-Social Behaviour Order, from this viewpoint, forms part of a "political strategy self-consciously aimed at restoring "public confidence" in the exercise of state power" (Ramsay 2009, p. 175). Indeed, several commentators have noted that, beyond its instrumental capacity to provide security, the preventive order acts to assuage the anxiety of a public bewildered by the erosion of the state's power to prevent crime. Specifically, it has been said that the preventive order reassures the public by manifesting the readiness of law enforcement authorities to act decisively and by conveying that, as Adam Crawford as has put it, "problems have reached such a point as to require drastic and exceptional action" (Crawford 2008, p. 774). Ohana has also emphasized the function fulfilled by the preventive order in restoring the trust of the public in the authority of the criminal law while attempting to bring to light the various ways in which the public's trust rests on its belief in the readiness of citizens to abide by the norms of the criminal law (Ohana 2010). Specifically, Ohana has scrutinized the conduct prerequisites of the preventive order and has argued that it bolsters confidence in the criminal justice system by verifying the law-abidingness of citizens. By initially eliciting awareness of suspicious behavior that tarnishes the perceived authority of the criminal law by upsetting social expectations of responsible behavior, the preventive order sets in motion an inquiry in which the competent authority tries to ascertain whether this behavior is actually symptomatic of a lack of law-abidingness. In this manner, the practices and procedures involved in the making of a preventive order constitute a ritual of verification which reassures the public that, aside from those untrustworthy actors who have actually been made subject to a preventive order, the criminal law largely continues to hold sway within society.

Related Entries

- ▶ Civil Remedies
- ▶ Law, Diversion and Community Sanctions in Juvenile Justice
- ▶ Penal Philosophy and Sentencing Theory
- ▶ Situational Crime Prevention
- ▶ Third Party Policing

Recommended Reading and References

Ancel M (1987) Social defense: the future of penal reform. Rothman, Littleton

Ashworth A (2004) Social control and "anti-social behaviour": the subversion of human rights? Law Quart Rev 120:263–291

Ashworth A, Zedner L (2008) Defending the criminal law: reflections on the changing character of crime, procedure, and sanctions. Crim Law Philos 2:21–51

Ashworth A, Zedner L (2010) Preventive orders: a problem of undercriminalization. In: Duff RA, Farmer L, Marshall SE, Renzo M, Tadros V (eds) The boundaries of the criminal law. Oxford University Press, Oxford

Beck U (1992) Risk society: towards a new modernity. Sage, London

Cheh M (1991) Constitutional limits on using civil remedies to achieve criminal law objectives: understanding and transcending the criminal-civil distinction. Hastings Law J 42:1325–1413

Cornford A (2011) Criminalising anti-social behaviour. Crim Law Philos 6:1–19

Cornish D (1994) The procedural analysis of offending and its relevance for situational prevention. Crime Prev Stud 3:151–196

Crawford A (2008) Dispersal powers and the symbolic role of anti-social behaviour legislation. Mod Law Rev 71:753–784

Crawford A (2009) Governing through anti-social behaviour. Br J Criminol 49:810–831

Digneffe F (2008) L'école positive italienne et le mouvement de défense sociale. In: DeBuyst C, Digneffe F, Pires AP (eds) L'histoire des sciences sur le crime et la peine, tome 2: la rationalité de la peine et la naissance de la criminologie. Larcier, Bruxelles

Dubber M (2002) Policing possession: the war on crime and the end of criminal law. J Crim Law Criminol 91:829–980

Duff A, Marshall S (2006) How offensive can you get? In: von Hirsch A, Simester AP (eds) Incivilities: regulating offensive behaviour. Hart, Oxford

Ewald F (2002) The return of descartes's malicious demon: an outline of a philosophy of precaution. In: Baker T, Simon J (eds) Embracing risk. University of Chicago Press, Chicago

Finn P, O'Brien Hylton M (1994) Using civil remedies for criminal behavior: rationale, case studies, and constitutional issues. U.S. National Institute of Justice, Washington, DC

Hoffman S, Macdonald S (2010) Should ASBOs be civilized? Crim Law Rev 6:457–473

Home Office (2010) Anti-social behaviour order statistics – England and Wales

Home Office (2012) Putting victims first: more effective responses to anti-social behaviour

Mann K (1992) Punitive civil sanctions: the middleground between criminal and civil law. Yale Law J 101: 1795–1873

Molinari PV, Papadia U (2002) Le misure di prevenzione, 2nd edn. Milano, Giuffrè

O'Deane M (2012) Gang injunctions and civil abatement: using civil remedies to curb gang-related crimes. CRC Press, Bocaraton

Ohana D (2006) Responding to acts preparatory to the commission of a crime: criminalization or prevention. Crim Just Ethics 25:23–39

Ohana D (2007) Desert and punishment for acts preparatory to the commision of a crime. Can J Law Jurisprudence 20:113–142

Ohana D (2010) Trust, distrust and reassurance: diversion and preventive orders through the prism of feindstrafrecht. Mod Law Rev 73(5):721–751

Pradel J (2002) Droit pénal comparé, 2nd edn. Dalloz, Paris

Ramsay P (2008) Vulnerability, sovereignty, and police power in the ASBO. In: Dubber MD, Valverde M (eds) Police and the liberal state. Stanford University Press, Stanford

Ramsay P (2009) The theory of vulnerable autonomy and the legitimacy of the civil preventive order. In: McSherry B, Norrie A, Bronitt S (eds) Regulating deviance. Hart, Oxford

Ramsay P (2010) Substantively uncivilized ASBOs. Crim Law Rev 10:761–763

Ramsay P (2012) The insecurity state: vulnerability, autonomy, and the right to security in the criminal Law. Oxford University Press, Oxford

Simester AP, von Hirsch A (2006) Regulating offensive conduct through two-step prohibitions. In: von Hirsch A, Simester AP (eds) Incivilities: regulating offensive behaviour. Hart, Oxford

Steiker C (1998) Foreword: the limits of the preventive state. J Crim Law Criminol 88:771–808

Von Hirsch A, Shearing C (2000) Exclusion from public space. In: von Hirsch A, Garland D, Wakefield A (eds) Ethical and social perspectives on situational crime prevention. Hart, Oxford

Walker C, Horne A (2012) The terrorism and investigation measures act 2011: one thing but not much the other? Crim Law Rev 6:421–438

Williams G (1953) Preventive justice and the rule of law. Mod Law Rev 16:17

Williams G (1955) Police control of intending criminals. Crim Law Rev 66

Zedner L (2009) Fixing the future? The Pre-emptive turn in criminal justice. In: McSherry B, Norrie A, Bronitt S (eds) Regulating deviance. Hart, Oxford

Preventive Patrol

Kenneth J. Novak
Department of Criminal Justice and
Criminology, University of Missouri,
Kansas City, MO, USA

Synonyms

General patrol; Patrol; Routine patrol

Overview

Preventive patrol is the most basic and oldest function of the police and historically has been the primary method that the police seek to prevent crime, maintain peace, and provide general policing services. However, strong evidence suggests that traditional preventive patrol has very limited crime prevention benefit, so over time the police have supplemented this function with other refined interventions that are more strategic, and proactive in nature, including foot patrol, directed patrol, crackdowns, and hot spot policing. These approaches vary considerably across three dimensions, including dosage, geographic and crime/problem focus, and type of intervention.

Patrol is an indispensable service that plays a leading role in the accomplishment of the police purpose. It is the only form of police service that directly attempts to eliminate opportunity for misconduct; it also checks the development of desire for misconduct by destroying unwholesome stimuli, by actively creating wholesome ones, and by favorably influencing individual and group attitudes in its routine daily association with the public.

Insofar as patrol fails to eliminate desire and belief in opportunity, misconduct results. Patrol is then immediately available to investigate offenses, apprehend offenders, and recover stolen property. Constant availability is important because time is of the essence in most police work (Wilson and McClaren 1963, 320).

Patrol has been referred to as the backbone of policing because the majority of sworn officers in most departments engage in this activity (Bayley 1994). Preventive patrol is perhaps among the oldest forms of police technology, dating back to the Bow Street Runners and early formulation of the Metropolitan Police Force in London, but could in fact be traced to thirteenth-century Hangchow, China (Kelling et al. 1974). Preventive patrol relies on the assumption that the omnipresence of the police will deter criminal activity and availability of officers will promote rapid response to crime emergencies, and as such, it continues to be the primary vehicle for the delivery of police services. It is the most basic of police activities. More sworn officers are assigned to patrol than other functions within the department (Bayley 1994), and a significant amount of personnel and capital resources are dedicated to preventive patrol. Typically preventive patrol is the first assignment for rookie officers, where cultural values and production norms are learned and where officers demonstrate their police competency. Patrol affords officers ample opportunity to interact with the public and make high-discretion decisions. As gatekeepers of the criminal justice system, patrol officers can initiate the criminal justice process. However, patrol is no longer the “only form of police service that directly attempts to eliminate” criminal opportunity, and while all police departments continue to engage in preventive patrol, there are many strategies that the police have developed over time to effectively and efficiently realize their public safety goals. Many of these contemporary strategies have evolved from the traditional patrol function.

Patrol seeks to accomplish a variety of goals. Chief among these goals is public safety; however, it is not clear whether this goal is consistently achieved. Other important but secondary goals may include crime prevention, preserving the peace/order maintenance, deterrence, being available to respond to crimes or other emergencies, apprehending suspects, increasing citizens' perceptions of safety, improving citizens' attitudes toward the police, traffic enforcement, ensuring roadway safety, identifying and responding to community problems, or responding to any number of service requests from the public. However, preventive patrol – i.e., nondirected, random patrolling by officers when not handling citizen requests or other emergencies – continues to occupy a considerable amount of the typical officer's time (Famega et al. 2005).

Crime prevention interventions can extend beyond uniformed officers randomly driving around geographic areas waiting to react to citizens' requests via 911 calls for service. Patrol functions can also include high levels of crime, problem or geographic focus, as well as a variety of interventions from law enforcement to problem solving. This essay examines traditional preventive patrol and its adaptations and will critically examine the impact of patrol through the lens of effectiveness and efficiency. A variety of methods of patrol will be introduced, including random preventive patrol, foot patrol, directed patrol, crackdowns, and hot spot policing. This essay will demonstrate that the best practices for public safety have evolved from the standard, traditional, reactive, passive model of patrol to a more flexible, data-driven, intelligence-led, proactive approach.

Random Preventive Patrol

Typically police jurisdictions are geographically divided into patrol beats or zones, and officers are assigned to these beats in order to conduct patrol activities. Beats deliberately vary in size based on a variety of factors, including residential population, commercial business density, number of calls for service, and crime rates in order to equalize officer workload. The result is some beats are very large (many square miles) while some are very small. These factors, when taken together, have an impact on officer workload and activities (Herbert 1997).

The most influential early study on the impact of preventive patrol was conducted in Kansas City (MO) in the early 1970s. This large-scale experiment involved manipulating the level of patrol allocation (or patrol dosage) across 15 beats in the South Patrol Division of the KCPD for an entire year, with the goal of determining what effect patrol has across a variety of outcomes. Each of the beats was assigned among three treatment conditions: proactive beats were assigned 2 to 3 additional patrol units, reactive beats were assigned no patrol units, and control beats were assigned the normal dosage levels. These beats were matched across a variety of crime and social characteristics permitting meaningful comparisons. Results indicated that assignment to one of the treatment conditions had little significant, or inconsistent, impact on a variety of suspected outcomes. Specifically, victimization surveys indicated no differences across residential and nonresidential burglaries, auto thefts (or thefts from autos), robberies, or vandalism; rates in which crimes were reported to the police did not vary; citizens' attitudes toward the police, satisfaction with police encounters, and fear of crime did not vary; citizens' protective measures (e.g., target hardening) did not vary; business owners had similar attitudes toward the police; response times were not faster in proactive beats; and the rate of traffic accidents did not vary (Kelling et al. 1974). Additionally, the project observed that officers had a significant level of uncommitted time; in fact, less than 20 % of an officer's time was dedicated to crime-related activities, whereas up to 80 % was spent on various service and peacekeeping activities. About 60 % of officers' total time was "uncommitted." These conclusions suggest that traditional preventive patrol may not be very effective at achieving objectively measured performance measures (e.g., crime suppression) or subjectively measured performance measures (e.g., citizens' perceptions).

The Kansas City Preventive Patrol Experiment (KCPPE) remains among the most influential and important examinations of the police function in America. The size and scope of this project was remarkable, and at its time, it was truly unique and, in fact, has not been fully replicated. Yet it is challenging to draw firm conclusions about the impact of the project. The research endeavor was truly path marking; however, methodological and theoretical limitations cast doubt on its findings (see Bayley 1994; Sherman and Weisburd 1995; Weisburd and Eck 2004). Nevertheless, minor adjustments to normal patrol allocations across beats probably do not increase a potential offender's perception of being caught. Beats in urban environments are often many square miles, so doubling police presence within those areas is likely undetectable. Also, the deterrence model assumes a rational offender is able to calculate the certainty of being caught, where in reality many crimes are the function of impulse, passion, or opportunity, which would break down any inherent value of police presence. Many crimes are not suppressible by patrol, including crimes committed indoors and outside the observation of the police (or offender). Finally, traditional preventive patrol may be too passive and reactive. Assuming that crime is prevented when randomly driving within large patrol beats and waiting for calls for service places the police at a crime prevention disadvantage, as little of these resources are dedicated to anticipating or preventing criminal events (Bayley 1994). Recognizing that officers have significant levels of uncommitted time presents an opportunity to guide proactive activities of officers strategically to increase the deterrent value of routine police activities. As Sherman and Eck (2002, 307) note, "there is little evidence that changing the number of officers patrolling beats has a consistent and measurable impact on crime." However, although these limitations may undermine the effects of patrol generally, the more focused and strategic uses of patrol may increase the likelihood that patrol officers are in the right places at the right times to deter or interrupt crimes (even those involving impulse or passion) and that their regular presence in hot spots will tend to drive would-be offenders away from areas that provide the greatest opportunities. Also, greater familiarity between police and citizens in high-risk locations may encourage cooperation and trust between these groups and facilitate higher clearances for crimes that do occur.

Another assumption of traditional preventive patrol is that officers will be out in the community and available to respond to citizens' requests for emergency assistance, and this activity will enhance public safety. While this may be an important service the police provide to the public, rapid response to emergencies has not been closely tied to crime prevention or public safety. The idea is that quick response by officers increases the likelihood of apprehending criminals, and this process provides a general crime deterrence (National Research Council 2004). However, empirical examinations note a weak relationship between rapid response and public safety (Spelman and Brown 1981) mostly because typical calls for service involve discovery crimes, where citizens "discover" the crime long after the offender has fled the scene or that citizens delay in contacting the police even after crimes involving contact with an offender(s). It is unlikely that rapid response will reduce crime (Sherman and Eck 2002), but because it is symbolically important to citizens, many police departments continue to monitor and track response times as an indication of effectiveness.

Why, despite the evidence and conventional wisdom that indicates preventive patrol is not as effective as intended, do police departments continue to utilize it as a core function? Perhaps one reason is citizen demand. Patrol can be viewed as an institutionalized component that is so engrained in American policing that the police may feel they have to minimally conform to citizens' desires, regardless of whether it is effective. Citizens expect to see uniformed officers patrolling streets in marked vehicles, and police risk criticism if they fail to comply with these traditions. Preventive patrol is a fixture within the institutional environment of policing. Crank and Langworthy (1992, p346) noted that “even though the rituals of 911 systems and random preventive patrol may be neither effective nor efficient as law enforcement or crime prevention strategies, they provide ceremonial evidence that a police department behaves as it should. . .Failure to sustain these important rituals may result in the de-legitimation of a police department.” Therefore, it really does not matter whether patrol “works” (in terms of crime control); the police must do it anyway. Furthermore, preventive patrol and rapid response provides ample opportunity for the police to interact with the public. These encounters are important components of the police–public relationship as it provides opportunities for the police to deliver important service (which is often unrelated to crime). Another reason may simply be habit or organizational inertia. Police organizations continue to assign vast personnel and capital resources to preventive patrol simply because they always have. It is a paradox that citizens may demand the police engage in a particular technology while also holding them accountable for public safety. Tying these concepts together, Mastrofski and Willis (2010) note that traditional preventive patrol is the “main course” that citizens crave – the standard or traditional model of policing provides citizens an opportunity to request police services and summon officers to them for a face-to-face encounter. Perhaps there is nothing more democratic than the ability of citizens to call the police (as agents of government) and know that they will respond, and in fact, citizens are more likely to have an encounter with a police officer than any other governmental representative (Langan et al. 2001). Preventive patrol and rapid response, even if it is ineffective and/or inefficient as crime control, provide other material and symbolic value to both the police and citizens.

But opportunity emerges from this reality. Recognizing the limits of preventive patrol along with the fact that officers often have large amounts of uncommitted time (Kelling, et al. 1974), the police have complemented this strategy with other tactics. Preventive patrol is viewed as the standard model (Weisburd and Eck 2004; National Research Council 2004) of policing, and while preventive patrol may not ever be completely eliminated or replaced, increasingly resources that could be used on patrol are being siphoned away toward strategies that have higher levels of focus and greater diversity in approaches. Supplemental patrol approaches, such as foot patrol, directed patrol, crackdowns, and hot spot policing, seek to focus resources in smaller geographic areas or on specific types of problems, sometimes using real-time, locally relevant crime information. These approaches are presented here within the context of patrol because they do not necessarily indicate a paradigm shift in the philosophy of policing, as problem-oriented policing or community policing may.

Beyond preventive patrol, crime prevention approaches may be characterized across at least three different dimensions. These dimensions borrow heavily from existing conceptualizations of police crime prevention strategies (see Weisburd and Eck 2004; Lum et al. 2011). The first dimension is officer *dosage* or overall allocation of personnel resources. Officer staffing levels may remain the same, and existing personnel can be reallocated in a more strategic manner with no additional resources. Shifting existing personnel from patrol to other, more focused strategies represents greater efficiency than adding additional personnel beyond current staffing levels (e.g., hiring more officers or more likely paying officers overtime to conduct special projects like crackdowns). Second is *focus*, including geographic and problem/crime focus. Using data, intelligence, and local knowledge, geographic areas may be identified that have specific types of crime problems, including guns and drugs. Police focus their attention in smaller geographic areas to address these specific types of problems. Third is the type of *intervention* or the activities that officers emphasize. Officers may be engaged in proactive, zero-tolerance style of law enforcement, or problem-solving approaches, or offender-focused approaches, or other activities that seek to address the problem identified earlier (e.g., trying to remove guns from the street to reduce gun-related crimes). Because crime prevention strategies tend to vary considerably across these and other dimensions, it can be challenging to draw broad conclusions on the best practices, and likewise, it is challenging to label specific activities as being a crackdown rather than directed patrol. What is clearer is that these progeny represent alternatives or supplements to traditional preventive patrol and with that in mind will be discussed collectively.

Foot Patrol

Even since the advent of motorized patrol, assigning officers to foot patrol has remained widely popular for decades. In fact, it is estimated that 55 % of all police departments regularly schedule officers to foot patrol, and utilization is most frequent in larger jurisdictions (Reaves 2010). Because motorized officers spend more time isolated in their vehicles, automobile patrol separates and creates a barrier between the police and the public, but foot patrol or bicycle patrol is positioned to encourage nonconfrontational interactions between officers and citizens, which in part can improve communication and relationships, as well as aid in peacekeeping and service delivery. But despite the widespread adoption of this strategy (which involves significant expenditure of personnel resources), the actual effectiveness of foot patrol is less clear. Evaluations have generally found citizens' satisfaction with the police and perceived safety increase, but there is a lack of consensus whether foot patrol influences actual crime rates (particularly violent crime rates). Two comprehensive evaluations of foot patrol revealed that although foot patrol may result in a slight reduction in crime, it primarily reduces citizens' fear of crime and changes the nature of police–citizen interactions toward more positive and non-adversarial exchanges. The major studies were the Newark Foot Patrol Experiment, which included data on foot patrol in Newark and 28 additional cities (Police Foundation 1981), and the Neighborhood Foot Patrol Program, which was conducted in 14 neighborhoods of Flint (MI) (Trojanowicz 1982). Later reports on the Flint study discovered an increased positive acceptance of the program and confidence in police services (Trojanowicz and Banas 1985). The size of a foot patrol beat should be small enough that it can be covered at least once or twice per shift to improve police–community interactions and reduce fear of crime (Payne and Trojanowicz 1985). At best, evaluations elsewhere demonstrated little-to-modest crime prevention benefit, which may have relegated foot patrol (as a strategy) as an add-on or supplement to real public safety interventions or conceptualized it as a public-relations tool.

Examinations of foot patrol may have been biased toward null results for several reasons. First, patrol areas for which officers were responsible may have been too large, creating dosage problems or diluting any possible benefit. This is analogous to the observations of early patrol experiments (Sherman and Weisburd 1995), and crime prevention and problem solving is more effective when it focuses on geographically small micro-places (Sherman and Weisburd 1995; Weisburd and Eck 2004). Second, the content of what officers actually do also matters. Increases in the mere presence of officers may yield modest crime prevention benefits; however, there is evidence that when coupled with specific problem-solving tactics, the impact may increase (Braga et al. 1999; Weisburd and Eck 2004). Early examinations of the effectiveness of foot patrol did not use rigorous research methodologies, and stronger research strategies are necessary to inform best practices in police crime prevention strategies.

Recent research conducted within the Philadelphia Police Department (PPD) paints a different picture of the effectiveness of foot patrol. The PPD identified micro-places within the city which could be considered hot spots for violent crime (particularly homicide, aggravated assaults, and outdoor robberies). These foot beats were small high-crime areas that averaged about 14.7 intersections and 1.3 miles of surface streets. The intervention involved areas that were amenable to foot patrol, without diluting dosage to the community level. A quasi-experimental design was employed to assign half to a treatment condition (foot patrol) with the other half serving as a control or comparison area. Violent crime in foot patrol areas declined significantly, while no measurable impact was observed in control areas, even after taking possible geographic displacement into account. The greatest crime prevention benefit was observed in the most crime-prone foot beats – those beats above the 60th percentile in their violent crime rate experienced less violent crime than their comparable control areas (Ratcliffe et al. 2011). This proactive, focused, micro-place strategy resulted in a successful outcome, which is consistent with results from evidence-based policing (EBP) studies, in that "police strategies are more effective when they are place-based, proactive and focused" (Lum et al. 2011, p. 5). Foot patrol can be implemented using existing resources, as evidenced in Philadelphia. However, because they can be so labor-intensive, if implemented over a large number of areas, then staffing and personnel increases may be necessary.

Directed Patrol and Crackdowns

Directed patrol represents a more strategic approach to use of patrol resources. While officers may in fact have considerable amount of discretionary time during preventive patrol, this uncommitted time can be interrupted by a call for service at inopportune times, making it challenging for officers to engage in meaningful, strategic crime prevention. Additionally, randomly patrolling throughout assigned beats in order to provide general deterrence may not be an efficient or effective use of resources. Directed patrol seeks to mitigate this by relieving some officers from responding to calls for service in order to provide them with uninterrupted time and further focuses their presence in smaller areas that have higher incidents of crime problems. This can also provide officers an opportunity to tailor interventions to particular areas. For example, an early application of directed patrol occurred in Pontiac (MI), where officers performed aggressive activities at strategic locations during uncommitted time. In analyzing the relationship between directed patrol activities and target crime specifically robberies, burglaries, auto thefts, and larcenies from automobiles, the number of arrests during program periods increased significantly. Results suggested that intervention while on directed patrol had more of an effect on crime than simply how many police were on patrol or how much time was devoted to directed patrol. Proactive patrol activities, such as arrests, vehicle stops, and field interrogations, decreased "at least some categories of reported crime" (Cordner 1981, 52), and Sherman and Eck (2002, 310) explain, "while there is very little empirical support for the effectiveness of adding more police across the board, there is a consistent body of strong scientific support for concentrating police patrols at places and times where crime is the most likely to occur. Patrolling low crime areas is unlikely to improve safety, but more patrolling in high crime places can be effective."

Intensive, sudden, short-term increases in officers' presence and proactive arrests for specific or all offenses in particular areas are known as crackdowns. Police crackdowns have been used to focus the attention of officers in particular areas for a variety of place-specific problems and as such may be considered a more intense form of directed patrol. This approach assumes that crackdowns increase the perceived risk of apprehension while simultaneously focusing on high-risk offenders within targeted areas, achieving crime prevention through deterrence and short-term incapacitation. Evidence indicates that crackdowns can achieve initial deterrence for some offenses as well as crime reduction after the crackdown has ended (e.g., residual deterrence). Crackdowns tend to have positive results – in a comprehensive review of 43 evaluations of crackdowns, Scott (2004) identified 24 that reported effectiveness, while 5 others reported mixed results. Case studies indicated that short-term crackdowns incur less deterrence decay after the conclusion of the strategy than long-term crackdowns, which suggests that personnel resources can be saturated in areas for relatively short periods of time, withdrawn, and crime prevention will linger for periods following the withdraw. This can be an effective and efficient practice, if strategically undertaken. For example, an examination of a 30-day crackdown in Richmond (VA) revealed a 92 % reduction in reported crime, relative to a comparison area. The crackdown produced some residual deterrence over the subsequent 6-month period; however, shortly thereafter some crime had returned to pre-project levels in some parts of the targeted zone (Smith 2001). An intensive crackdown for gun-related offense demonstrated similar benefits. Officers saturated a specific patrol beat in Kansas City (MO) for 29 weeks, during which time they proactively concentrated on removing guns from the street. Gun seizures increased by 65 %, while gun-related crime decreased by 49 %. There was no indication of crime displacement to adjacent areas (Sherman et al. 1995). The effect is likely because the intervention was geographically focused to address a specific crime, rather than offering a mere general deterrent effect of increased presence. A similar initiative in Indianapolis that involved directed police patrol in high-crime areas yielded results when employing a targeted offender/specific deterrence strategy on suspicious people and problem places within beats to reducing violence (McGarrell et al. 2001).

There are limitations with crackdowns that are worth noting. First, crackdowns typically are expensive and labor-intensive, which compromises its efficiency as a crime prevention strategy. It is promising that short-term crackdowns yield the greatest residual deterrence; however, intense saturation of patrol in small areas may be cost prohibitive as a general public safety model. This is particularly true if crackdowns are implemented as a special project where officers are paid overtime for participation in the crackdown. In the crackdowns described above, Smith (2001) indicated that 4,812 additional man-hours were realized to implement the crackdown; the Kansas City Gun Experiment was executed with an additional 4,512 officer hours (Sherman et al. 1995); the Indianapolis directed patrol intervention involved 2,905 additional officer hours (McGarrell et al. 2001). Crackdowns, unlike other strategies like the intervention describe above in Pontiac and hot spots policing (to be discussed later) may not be easily implemented with existing resources. Second, crackdowns can compromise police–public relationships. Proactive, zero-tolerance arrests within areas increase the possibility of negative, confrontational police–public contacts that can interrupt the ability of these parties to engage in meaningful coproductive partnerships. Crackdowns that focus on street-level activity in socially disorganized areas disproportionately impact poor citizens, which could not only exacerbate police–public relationships but also heighten fear among residents. It is wise to consult with citizens in advance of implementing crackdowns to determine appropriate and ethical target selection and officer behavior (Sherman, et al. 1995; McGarrell, et al. 2001). Related, because crackdowns involve aggressive law enforcement efforts, there is an increased potential for abuse of citizens' rights by the police, which further underscores the importance of careful planning and constant monitoring of activities by administration. Crackdowns may be best suited as short-term, quick fixes to emerging problems implemented in consultation with residents rather than a long-term strategic public safety solution, particularly since crackdowns rarely address the underlying causes of crime (Scott 2004).

Hot Spots of Crime and Disorder

Within patrol beats there may be micro-places that account for a disproportionate amount of police activities. Calls for service, crimes, or disorder tend to be concentrated spatially among a few parcel addresses, street corners, or city blocks and further may be concentrated temporally. An analysis of calls for service in Minneapolis revealed that very few places accounted for a disproportionate amount of recurring problems requiring police attention. In fact, over the course of a year, most places in the city produced one call for service or less, whereas 3.3 % of all places accounted for over 50 % of all calls for service. This was also true for predatory crimes, such as robbery, rape/sexual conduct, and auto theft (Sherman et al. 1989). Places with the highest rate of crime and calls for service which required police attention are referred to as "hot spots" of crime. This has direct implications for patrol allocation in that local information can be used to identify places within the city where crime and calls for service are so frequent that they almost become predictable, and this is further concentrated among certain "hot times" during the day or week. Recognizing that random preventive patrol has limited effectiveness to begin with, and coupled with the reality that much of a patrol officer's day includes uncommitted time, it becomes possible to strategically allocate officers to hot spots during hot times to effectively impact crime and disorder without significant alterations to personnel resources. Hot spot policing differs from directed patrol or crackdowns primarily in that its geographic focus is much smaller.

There is compelling evidence that hot spot policing can be an effective and efficient strategy to impact crime. Geographically focused policing strategies such as this have been identified as a promising strategy for increasing the effectiveness of the police (National Research Council 2004). Braga (2007) systematically examined existing published empirical examinations and concluded that 7 of the 9 studies reported significant reductions in crime and disorder and therefore will be explored more fully in another entry of the current work (see Telep and Weisburd *citation here*). The Minneapolis Hot Spots Patrol Experiment was an early examination of the efficacy of this spatially focused approach and was designed to examine the crime prevention effects of patrol officer presence at hot spots during hot times. Police and researchers identified 110 hot spot places and randomly assigned half of these places to a treatment condition. Officers were instructed to be present at one of the targeted hot spots during uncommitted time; thus, this examination was not so much concerned about specific officers' activities as it was focused on additional officer presence to deter criminal activity. Results indicated that increased police presence in hot spots yielded significant reductions in calls for crime. Calls for service regarding disorder experienced the largest decline, and researchers conducting systematic social observations of hot spots observed less disorder in target hot spots relative to their counterparts. The deterrent effect of hot spot policing may be different from what was observed in the KCPPE because police focused on place-specific "micro-deterrence" (Sherman and Weisburd 1995, 646), thus strategically focusing officers on identifiable, small problem places so as to not dilute their deterrent effect across large beats. Subsequent analyses revealed that the amount of time officers are at a hot spot does not need to be too long to experience measurable reductions in disorder. Officers need to minimally be at a hot spot for 10 min to yield beneficial results, with the optimal time being 14 to 15 min for each hot spot per hot spot. Time spent at a hot spot beyond 15 min begins to yield diminishing returns which compromise the efficiency of personnel resources (Koper 1995).

Lessons from the Minneapolis examination continue to have relevance for everyday patrol work. These results demonstrate that police departments can use timely, locally relevant data to focus patrol activities in very specific locations. Famega et al. (2005) noted that typically 75 % of officers assigned to preventive patrol have uncommitted or highly discretionary time but further note that very little of that unassigned time is spent on proactive strategies such as problem solving. Rather, they note that officer typically engage in routine patrol, back up other officers, or wait to be dispatched. Opportunity arises from this in that patrol officers can target hot spots as part of their normal routine. Internal crime analysts can utilize local crime data, as done in Minneapolis, to continually redefine hot spots, examine the persistence of crime in these micro-places, and reassign officers to emerging hot spots as necessary. This reality, perhaps more than others, has the greatest potential to improve police effectiveness while utilizing existing organizational resources.

Hot spot policing strategies can have notable shortcomings. First, the overall utility of hot spots policing would be mitigated if crime and disorder were merely pushed around the corner – displacing crime and disorder from a hot spot to another location yields a zero-sum gain for public safety, and there exists a serious risk of creating new hot spots in locations that were previously crime-free. Recognizing this risk it is important to monitor areas contiguous to hot spots (i.e., catchment areas). Measuring and monitoring displacement can be complicated, as crime can be displaced territorially (from one location to another), temporally (from one time of the day to another), tactically (changing the method of crime, e.g., auto theft decline but carjacking increase), by target (different victims in the same area), functionally (offenders commit different offenses because the risk of committing the original is too high), or by perpetrator (some offenders are deterred, but new offenders emerge to replace them) (Lab 2010). Previous studies tend to demonstrate minimal territorial displacement to surrounding catchment areas, and in some cases, crime in surrounding areas actually decreases. A diffusion of benefits occurs when areas not targeted for police intervention nevertheless experience residual benefit and crime decreases (see Weisburd and Green 1995). The weight of the evidence on various hot spot policing interventions suggests that the police can impact crime, calls for service, and disorder within localities without significant risk for crime displacement. Second, like crackdowns, hot spot policing may demonstrate short-term impacts with little long-term benefits unless the strategy is coupled with meaning problem solving aimed at addressing the underlying social problems that permitted crime to flourish in the first place (i.e., what made this spot "hot" to begin with?). Third, this approach also runs the serious potential for alienating citizens. If citizens view hot spot policing as police officers saturating the area with aggressive enforcement activities, then the police may inadvertently compromise police–public relations. This can result in labeling or stigmatizing areas (and the people in those areas) as undesirable. To avoid these and other unanticipated consequences, Rosenbaum (2006) encourages comprehensive analyses using a variety of different information sources to identify hot spots; employ a variety of alternatives in the areas to facilitate short- and long-term success; build meaningful partnerships between the police and other governmental, private, and citizens groups who are hidden allies in facilitating public safety; and pay attention to preventing crime across individual, family, and community levels (and not just the micro-spatial level).

Summary

Preventive patrol remains a dominant activity function of the police; however, the evidence suggests that it has limited effectiveness and efficiency in promoting public safety. Over time a number of approaches have developed to counterbalance this reality, and while it remains challenging to suggest that a single evolving approach is superior to another, it is safe to say that they represent a dramatic improvement to and hold more crime prevention promise than preventive patrol.

The evolving nature of street-level activities of police officers to achieve effective and efficient crime prevention comes with it a number of challenges and considerations for policy makers. These considerations are directly related to the dimensions previously developed. First is a strategic allocation of resources. Effective and efficient crime prevention often must be accomplished within fiscal realities; thus, it is important to determine whether officers traditionally assigned to preventive patrol duties can be engaged to participate in the supplements such as foot patrol, directed patrol, crackdowns, or hot spot policing. Will officers be tasked with performing activities during periods of uncommitted time, or will they be untethered from responding to calls for service in order to engage in strategic problem solving? Hot spot policing or some version of directed patrol may be suitable for generalist officers who are still responsible for calls for service, whereas foot patrol and crackdowns may be most easily implemented by either assigning officers to blocks of time where they are not responding to calls to implement this strategy. Interventions that are executed through the use of overtime assignments may be doomed to be inefficient and too episodic in nature to have a lasting public safety benefit, and thus may be reserved for unusual, exigent circumstances rather than business as usual.

Second, identification and focus become paramount importance. Rather than conceptualizing geographic areas as large beats, neighborhoods, or some other arbitrary quasi-permanent boundary, focus would have to be defined as smaller areas within patrol beats. Related, the type of crime, problem, or disorder occurring within the jurisdiction may also vary from place to place. Some areas may be prone to gun-related violence, whereas others are plagued by drugs. Some areas may be hot spots of disorder and public nuisances, whereas others may be prone to violence. Even after disaggregation by these means, further analysis and local knowledge may be necessary. Regardless, greater geographic and crime/problem specificity is necessary to more fully identify locations in need of supplemental police attention and intervention and must be monitored on a regular basis in order to determine whether problems emerge in new areas (territorial displacement) at different times (temporal displacement) or whether old problems are replaced with new problems (tactical displacement).

Third, interventions will vary and will need to be place and crime specific. Mere presence for general deterrence, problem solving, focused activities, and aggressive law enforcement all can have public safety benefits; however, they must be suited to the geographic and crime/problem focus of different places. Strategic, data-driven, intelligent policing coupled with monitoring of objective and subjective outcomes holds public safety promise beyond the traditional preventive patrol approach.

Fourth, preventive patrol and other interventions will need to make an effort to continually engage the public in a collaborative manner. An unanticipated consequence of random motor patrol is that it separated the police from the public, created a barrier for nonconfrontational interactions, and created social distance between the police and citizens. A conscientious approach to facilitating a partnership between these groups will avoid misconceptions, help the police to strategically identify places of greatest concern and the underlying problems associated with these places, and facilitate effective order maintenance and delivery of police services.

Related Entries

- ▶ Deterrence
- ▶ Evidence-Based Policing
- ▶ Hot Spots and Place-Based Policing
- ▶ Intelligence-Led Policing
- ▶ Order Maintenance Policing
- ▶ Predictive Policing
- ▶ Theories on Policing and Communities

Recommended Reading and References

Bayley DH (1994) Police for the future. Oxford University Press, New York

Braga AA (2007) The effects of hot spots policing on crime. Campbell Collaboration. doi:10.4073/csr.2007.1

Braga AA, Weisburd DL, Waring EJ, Mazerolle LG, Spelman W, Gajewski F (1999) Problem-oriented policing in violent crime places: a randomized controlled experiment. Criminology 37:541–580

Cordner GW (1981) The effects of directed patrol: a natural quasi-experiment in Pontiac. In: Fyfe JJ (ed) Contemporary Issues in law enforcement. Sage, Beverly Hills

Crank JP, Langworthy R (1992) An institutional perspective of policing. J Crim Law Criminol 83:338–363

Famega CN, Frank J, Mazerolle L (2005) Managing police patrol time: the role of supervisor directives. Justice Quarterly 22:540–559

Foundation P (1981) The Newark foot patrol experiment. Police Foundation, Washington, DC

Herbert S (1997) Policing space: territoriality and the Los Angeles police department. University of Minnesota Press, Minneapolis

Kelling GL, Pate T, Dieckman D, Brown CE (1974) The Kansas City preventive patrol experiment: a technical report. Police Foundation, Washington, DC

Koper CS (1995) Just enough police presence: reducing crime and disorderly behavior by optimizing patrol time in crime hot spots. Justice Quarterly 12:649–672

Lab SP (2010) Crime prevention: approaches, practices and evaluations, 7th edn. Anderson, New Providence

Langan PA, Greenfeld LA, Smith SK, Durose MR, Levin DL (2001) Contacts between police and the public: findings from the 1999 national survey. Bureau of Justice Statistics, Washington, DC

Lum C, Koper CS, Telep CW (2011) The evidence-based policing matrix. J Exp Criminol 7:3–26

Mastrofski SD, Willis JJ (2010) Police organization continuity and change: into the twenty-first century. Crime and Justice: A review of research 39:55–144

McGarrell EF, Chermak S, Weiss A, Wilson J (2001) Reducing firearms violence through directed police patrol. Criminology and Public Policy 1:119–148

National Research Council (2004) Fairness and effectiveness in policing: the evidence. In: Skogan W, Frydl K (eds) Committee to review research on police policy and practices. National Academies Press, Washington, DC

Payne DM, Trojanowicz RC (1985) Performance profiles of foot versus motor officers. Michigan State University, East Lansing

Ratcliffe JH, Taniguchi T, Groff ER, Wood JD (2011) The Philadelphia foot patrol experiment: a randomized controlled trial of police patrol effectiveness in violent crime hotspots. Criminology 49:795–831

Reaves BA (2010) Local police departments, 2007. U.S. Department of Justice, Washington, DC

Rosenbaum DP (2006) The limits of hot spot policing. In: Weisburd D, Braga AA (eds) Police innovation: contrasting perspectives. Cambridge University Press, Cambridge

Scott MS (2004) The benefits and consequences of police crackdowns, Problem-Oriented Guides for Policing. U.S. Department of Justice, Washington, DC

Sherman LW, Eck JE (2002) Policing for crime prevention. In: Sherman LW, Farrington DP, Welsh BC, MacKenzie DL (eds) Evidence-based crime prevention. Routledge, London, pp 295–329

Sherman LW, Weisburd D (1995) General deterrent effects of police patrol in crime hotspots: a randomized controlled trial. Justice Quarterly 12:625–648

Sherman LW, Gartin PR, Buerger ME (1989) Hot spots of predatory crime: routine activities and the criminology of place. Criminology 27:27–56

Sherman LW, Shaw JW, Rogan DP (1995) The Kansas City gun experiment, research in brief. National Institute of Justice, Washington, DC

Smith MR (2001) Police-led crackdowns and cleanups: an evaluation of a crime control initiative in Richmond, Virginia. Crime Delinq 47:60–83

Spelman W, Brown DK (1981) Calling the police: a replication of the citizen reporting component of the Kansas City response time analysis. PERF, Washington, DC

Trojanowicz RC (1982) An evaluation of the neighborhood foot patrol program in Flint, Michigan. School of Criminal Justice, Michigan State University, East Lansing

Trojanowicz RC, Banas DW (1985) The impact of foot patrol on black and white perceptions of policing. Michigan State University, East Lansing

Weisburd D, Eck JE (2004) What can the police do to reduce crime, disorder and fear? The Annals 593:42–65

Weisburd D, Green L (1995) Policing drug hot spots: the Jersey City drug market analysis experiment. Justice Quarterly 12:711–735

Wilson OW, McClaren RC (1963) Police administration, 3rd edn. McGraw-Hill, New York

Principles of Effective Correctional Intervention

▶ Actualizing Risk-Need-Responsivity

Principles of Effective Intervention

▶ Actualizing Risk-Need-Responsivity

Prior Victimization

▶ Repeat Victimization

Prison

▶ Drug Abuse and Alcohol Dependence Among Inmates

Prison Aversion Programs

▶ Scared Straight Programs

Prison Warden Stress and Job Performance

Cassandra A. Atkin-Plunk[1] and
Gaylene S. Armstrong[2]
[1]Sam Houston State University, Huntsville, TX, USA
[2]College of Criminal Justice, Correctional Management Institute of Texas, Sam Houston State University, Huntsville, TX, USA

Synonyms

Correctional administrators; Leadership

Overview

While many studies examine factors relating to job satisfaction and job stress of correctional officers and correctional treatment staff, the scholarly literature on job satisfaction and job stress of wardens is scarce. Of the studies conducted, many determinants have been found to impact job satisfaction and job stress of wardens, such as individual-level attributes (e.g., age, race, gender, educational level, and tenure status) and work environment/organizational features (e.g., career experiences, organizational conditions, nature of the work role, correctional approaches, intrinsic and extrinsic rewards, organizational justice, organizational structure, social support, work-family conflict, and job stress). This entry provides an examination of the determinants of job satisfaction and job stress of correctional officers as well as wardens. It also examines the impact that different leadership styles could have on job satisfaction and job stress.

Fundamentals

Despite substantial growth of the United States correctional system, limited research exists on the leaders who are responsible for the effective and efficient functioning of correctional organizations. Typically, prison wardens are responsible for a wide array of tasks and must answer to constituencies both internal and external to their correctional facility. While ensuring the priority of a safe and secure prison environment for both staff and inmates, wardens must effectively and efficiently respond to macrolevel budget and legislative issues, human resource and staffing issues, facility functioning, as well as concerns of inmates and their families (Ruddell and Norris 2008). To be successful in these and other efforts, wardens must be effective leaders who can successfully convey and inspire their staff with a shared vision for their organization (Heaton and Atherton 2008) without suffering negative affect from intensive job tasks. Yet, the dynamic prison environment presents daily challenges for wardens as many wardens perceive that they are insufficiently prepared (McCampbell 2002).

Insight into promising management practices for prison wardens has been gained in a number of areas including overcrowding, prison gangs, inmate amenities, physical location of the warden's office, and managerial attitudes. Less attention has been paid to the successful functioning of prison wardens insofar as identifying individual or organizational attributes related to a positive work experience and job-related stress (Cullen et al. 1993; Flanagan et al. 1996).

Only in related literature that has assessed correctional officers have researchers examined antecedents of job stress and related coping mechanisms. Job stress has been typically defined in the job stress literature as "the harmful physical and emotional responses that occur when the requirements of the job do not match the capabilities, resources, or needs of the worker" (National Institute for Occupational Safety and Health 1999, p. 6). Stress typically results from exposure to stressors that develop as a result of the workplace and associated duties (Cullen et al. 1985; Lambert 2004). Experiencing high levels of stress especially in a prolonged or regular manner can have damaging outcomes on an individual. These negative outcomes can include poor job performance as well as increased levels of social problems, burnout, divorce, mental health problems, and illness (Griffin et al. 2010; Salami et al. 2010).

Given the importance of prison wardens in the everyday functioning of correctional organizations, a warden's inevitable exposure to job stress may have a damaging effect on both the individual as well as the organization (Hom and Griffeth 1995; Lambert et al. 2007). Negative work experiences such as job stress can result in higher levels of turnover (Shaw 2011) which can be especially concerning when it encompasses leaders of an organization. As Shaw (2011) and others have noted, both negative proximal (e.g., safety, productivity, efficiency) and distal (e.g., organizational performance) impacts of job stress could cause eventual organizational disruption. Consequently, it is important for researchers to consider correlates of prison warden job stress given the significant level of organizational capital held by a person in this position as well as coping mechanisms that may offset job stress.

As noted earlier, the consequences of job stress can be damaging to both an individual employee and the correctional organization as a whole (Lambert et al. 2007). While extensive research in this arena has been conducted in other occupational fields, within corrections the job stress literature has primarily focused on custodial staff, such as correctional or detention officers and to a limited extent noncustodial, correctional treatment staff (see Armstrong and Griffin 2004). Insufficient attention has been paid to prison wardens and job stress, with a very limited literature that has focused on job satisfaction (Cullen et al. 1993; Flanagan et al. 1996). Existing studies have demonstrated that correctional officers report high levels of job stress due in part to individual perceptions of their work environment, which is contemporaneously influenced by their personal characteristics (Mitchell et al. 2001). Given the demanding responsibilities of prison wardens, including maintaining a safe and secure prison environment, managing staff resources, disciplinary incidents and grievance processes, fiscal constraints, upholding public relations, and maintaining professionalism (Ruddell and Norris 2008), it is reasonable to anticipate high levels of job stress exists among prison wardens.

Job stress levels among prison wardens are a unique concern because everyday operations of prisons and their environment could be negatively affected by a warden who experiences significant job stress. If high job stress levels are the norm, prison wardens may be unduly exposed to higher risks of poor decision making, job performance, or turnover (Hulsheger et al. 2010). While turnover is not always a drawback (i.e., some degree of staff turnover can be healthy), when it occurs during a period of organizational change or fiscal conservation, it can be internally damaging to an organization. Alternatively, it could be the case that prison wardens may not experience significant job stress due to unique characteristics of that individual or adaptive coping methods successfully employed. Attrition of those correctional staff may have already occurred for individuals that perceive significant job stress or fail to successfully cope with existing stress, or at minimum those individuals have not successfully navigated and promoted through the correctional system. Stated another way, individuals who perceive the correctional environment to be stressful may not seek or receive career advancement opportunities or be "washed out" at an earlier career point prior to appointment as a warden. Thus, a determination of the level of prison warden job stress, as well as its corollary conditions, is important and significantly understudied.

Correlates of Correctional Officer Job Stress

Factors previously examined in the correctional officer job stress literature have included individual-level attributes (e.g., gender, race, age, education level, tenure status, and position) as well as organizational or workplace characteristics (e.g., organizational justice, conditions of confinement, work roles, organizational and co-worker support, quality of supervision) (see Armstrong and Griffin 2004; Cullen et al. 1985, 1993; Flanagan et al. 1996; Hepburn and Knepper 1993; Lambert 2004; Lambert et al. 2006a, b, 2007; Mitchell et al. 2004; Triplett et al. 1999). While some studies have concurrently examined job stress and job satisfaction (Lambert et al. 2006a, 2007), the distinct influence of the two concepts has been difficult to disentangle.

Within the correctional officer job stress literature, a number of inconsistencies in determining individual attributes that correlate with job stress persist. Specifically, the racial background of an officer has been found to both influence as well as be unrelated to correctional officer perceptions of job stress. For example, Lambert et al. (2007) found correctional officers who were White reported significantly higher job stress (see also Cullen et al. 1985), yet Armstrong and Griffin (2004) found no significant relationship between officer race and job stress (see also Lambert et al. 2010). Gender has been a more consistent indicator of job stress, albeit an interactive effect that is influenced by characteristics of the prison environment. Specifically, female officers as compared to male officers have tended to report significantly more job stress (Armstrong and Griffin 2004; Lambert et al. 2007; Cullen et al. 1985). Regarding gender, Triplett and colleagues (1999) found significant interactive effects between gender and various aspects of the work environment. For example, female correctional officers experienced greater work-home conflict, higher levels of contact with prisoners, and an increased perception of job dangerousness. In turn, these factors were positively correlated with job stress. In comparison, with male officers factors such as tenure, quantitative work overload, and perceived job dangerousness were related to higher levels of job stress.

Interestingly, while officers with greater tenure or “time on the job” have been found to also have higher levels of job stress (Armstrong and Griffin 2004; Lambert et al. 2007; Cullen et al. 1985), beyond number of years of correctional experience, a more textured assessment of prior experience has not been explored. Prior literature has argued that important differences in job stress may exist as well as distinct perspectives between custodial and treatment staff (see Armstrong and Griffin 2004), and these may play a role in how a person experiences subsequent positions. Prison wardens may promote from either type of experiential background; thus, type of prior experience is an important extension in the measurement of the tenure variable.

In addition to individual characteristics, the attributes of the correctional environment have been found to influence correctional officer stress (Mitchell et al. 2004). Studies have found that officers who perceived their job as “dangerous” (Armstrong and Griffin 2004; Cullen et al. 1985; Triplett et al. 1999) or were employed at a higher security level prison (Cullen et al. 1985) also perceived higher levels of job stress. These types of facility effects may extend to the composition of the inmates in a facility in terms of the prison being a single gender or co-ed facility as additional security concerns are presented or minimized as a result.

In contrast to attributes associated with higher levels of job stress, researchers have also advanced an understanding of protective factors thought to be related to lower levels of correctional officer job stress. To date, researchers have studied various forms of social support including co-worker or peer support hypothesized to decrease job stress. Results have not demonstrated a clear linkage such that some studies have found supervisory and family support to be negatively correlated with work stress, while peer support was positively correlated with job stress (Cullen et al. 1985). Other studies have found that co-worker or peer support reduced job stress (Armstrong and Griffin 2004).

While the perspective and daily routines of prison wardens differ from a typical custodial correctional officer, the work environment itself remains the same. That is, the prison warden remains exposed to the same environmental factors of the correctional facility as the correctional officers including the prison’s size, custody level of the population held, and gender of the inmates (i.e., male, female, or co-ed population); therefore, these organizational attributes may influence a prison warden, but the warden's unique perspective as well as additional factors should be considered.

Transformational Leadership as a Protective Factor Against Job Stress

Research has consistently found that correctional officers who perceived the supervision they receive to be of high quality also tended to report lower levels of job stress (Armstrong and Griffin 2004; Cullen et al. 1985; Waters 1999). Despite the importance of supervision quality from the perspective of the correctional officer, good quality leadership is not necessarily an innate attribute of all prison wardens. According to Heaton and Atherton (2008), "becoming a leader in any organization usually involves a process of years of personal and professional development" (p. 14). Moreover, Heaton and Atherton (2008) suggested that a successful leader must understand the challenges confronting staff, maintain good relationships with co-workers, have a balanced life and good personal health, be energetic, have an appropriate emotional outlet, effectively communicate the purpose of the correctional organization to stakeholders, be a mediator, clearly articulate performance expectations, and be invested in their relationships with staff members. What type of leadership style might encompass such positive attributes? The organizational literature on leadership suggests a transformational leader exhibits similar characteristics.

Transformational leaders are individuals who aim to increase their organization's awareness of appropriate tasks and further motivate organization members to perform beyond basic expectations (Bass 1985). A growing body of research on transformational leadership supports the suggestion that a transformational leadership style has a direct, positive impact on performance outcomes and the behavior or experiences of subordinates (Arnold et al. 2007). For example, Mullen et al. (2011) examined transformational leadership as compared to an alternative form of passive or uninvolved leadership, which "are generally considered to be the most ineffective styles of leadership" (p. 42). Their results demonstrated that subordinates of safety managers who consistently demonstrated transformational leadership skills engaged in greater levels of safety compliance and safety participation. Additionally, using an experimental design Bono and Ilies (2006) found that "charismatic leaders enable their followers to experience positive emotions" (p. 331). Sosik and Godshalk (2000) specifically found that transformational leadership led to less job stress as a result of the increase mentoring received by subordinates. Thus, from existing literature, it appears as if transformational leadership can affect both individual well-being of subordinates in the organization as well as performance outcomes. This twofold effect of transformational leadership on performance outcomes as well as individuals was evident in Chin's (2007) meta-analysis of 28 independent studies of transformational school leadership. In the study, Chin examined the effects of transformational leadership on teachers finding higher levels of transformational leadership coincided with higher rates of teacher job satisfaction, school effectiveness as perceived by the teachers, and student achievement.

While the literature on the direct effects of exhibiting transformational leadership on subordinates is well researched, the relationship between transformational leadership skills and the leader's well-being is less clear. In the limited existing research, one study by Ram and Prebhakar (2010) found that within a sample of managers from the telecom industry, the greater the extent to which managers reported having transformational leadership skills, the lower levels of reported job stress. Yet, the mechanism through which this reduction in job stress is achieved is not explained. Arguably those who engage in a transformational leadership style experience more successful organizational functioning overall (i.e., a reduced role stressor) and potentially more support from their subordinates, thereby individually reporting lower levels of job stress. As a first step, however, it must be first considered whether transformational leadership has significant relationship with job stress within a prison warden population.

Recommended Reading and References

Armstrong GS, Griffin ML (2004) Does the job matter? Comparing correlates of stress among treatment and correctional staff in prisons. J Crim Justice 32:577–592

Arnold KA, Turner N, Barling J, Kelloway EK, McKee MC (2007) Transformational leadership and psychological well-being: the mediating role of meaningful work. J Occup Health Psychol 12:193–203

Bass BM (1985) Leadership and performance beyond expectation. Free Press, New York

Bono JE, Ilies R (2006) Charisma, positive emotions and mood contagion. Leadersh Q 17:317–334

Chin J (2007) Meta-analysis of transformational leadership effects in school outcomes in Taiwan and the USA. Asia Pac Educ Rev 8:166–177

Cullen FT, Link BG, Wolfe NT, Frank J (1985) The social dimensions of correctional officer stress. Justice Q 2:505–533

Cullen FT, Latessa EJ, Kopache R, Lombardo LX, Burton VS (1993) Prison wardens' job satisfaction. Prison J 73:141–161

Flanagan TJ, Johnson WW, Bennett K (1996) Job satisfaction among correctional executives: a contemporary portrait of wardens of state prisons for adults. Prison J 76:385–397

Griffin ML, Hogan NL, Lambert EG, Tucker-Gail KA, Baker DN (2010) Job involvement, job stress, job satisfaction, and organizational commitment and the burnout of correctional staff. Crim Justice Behav 37:239–255

Heaton J, Atherton G (2008) Leadership for the 21st century. Correct Today 70:14–18

Hepburn JR, Knepper PE (1993) Correctional officers as human services workers: the effect on job satisfaction. Justice Q 10:315–337

Hom PW, Griffeth RW (1995) Employee turnover. South-Western, Cincinnati

Hulsheger UR, Lang J, Maier G (2010) Emotional labor, strain, and performance: testing reciprocal relationships in a longitudinal panel study. J Occup Health Psychol 15:505–521

Lambert EG (2004) The impact of job characteristics on correctional staff members. Prison J 84:208–227

Lambert EG, Hogan NL, Camp SD, Ventura LA (2006a) The impact of work-family conflict on correctional staff: a preliminary study. Criminol Crim Justice 6:371–387

Lambert EG, Paoline EA, Hogan NL (2006b) The impact of centralization and formalization on correctional staff job satisfaction and organizational commitment: an exploratory study. Crim Justice Stud 19:23–44

Lambert EG, Hogan NL, Griffin ML (2007) The impact of distributive and procedural justice on correctional staff job stress, job satisfaction, and organizational commitment. J Crim Justice 35:644–656

Lambert EG, Altheimer I, Hogan NL (2010) Exploring the relationship between social support and job burnout among correctional staff. Crim Justice Behav 37:1217–1236

McCampbell SW (2002) Making successful new wardens. Correct Today 64:130–134

Mitchell O, MacKenzie DL, Gover AR, Styve G (2001) The influences of personal background on perceptions of juvenile correctional environments. J Crim Justice 29:67–77

Mitchell O, MacKenzie DL, Styve GJ, Gover AR (2004) The impact of individual, organizational, and environmental attributes on voluntary turnover among juvenile correctional staff members. Justice Q 17:333–357

Mullen J, Kelloway EK, Teed M (2011) Inconsistent style of leadership as a predictor of safety behavior. Work Stress 25:41–54

National Institute for Occupational Safety and Health (1999) Stress at work (DHHS Publication No. 99–101). Department of Health and Human Services, Cincinnati

Ram P, Prebhakar GV (2010) Leadership styles and perceived organizational politics as predictors of work related outcomes. Eur J Soc Sci 15:40–55

Ruddell R, Norris T (2008) The changing role of wardens: a focus on safety and security. Correct Today 70:36–39

Salami AO, Ojokuku RM, Ilesanmi O (2010) Impact of job stress on managers' performance. Eur J Sci Res 45:249–260

Shaw JD (2011) Turnover rates and organizational performance: review, critique, and research agenda. Organ Psychol Rev 1:187–213

Sosik JJ, Godshalk VM (2000) Leadership styles, mentoring functions received, and job-related stress: a conceptual model and preliminary study. J Organ Behav 21:365–390

Triplett R, Mullings JL, Scarborough KE (1999) Examining the effect of work-home conflict on work-related stress among correctional officers. J Crim Justice 27:371–385

Waters JE (1999) The impact of work resources on job stress among correctional treatment staff. J Addict Offending Couns 20:26–34

Prison Without Bars

▶ Electronic Monitoring

Private Policing

▶ Privatization of Policing in an International Context

Private Security

▶ Preventing of Terror at Shopping Malls
▶ Privatization of Policing in an International Context

Privatization

▶ Privatization of Policing in an International Context

Privatization of Policing in an International Context

Mark Button and Alison Wakefield
University of Portsmouth, Portsmouth, UK

Synonyms

Police; Policing; Private policing; Private security; Privatization

Overview

It is now widely recognized within criminology that "policing" does not simply equate with the work of "the police," with private security personnel, police auxiliaries, and voluntary bodies featuring among the multiplicity of agencies playing a role in keeping communities safe. This can be seen as part of a general, worldwide trend in which security agencies are expanding and diversifying.

Notable among these developments has been the substantial rise in demand for private security in Western democratic countries, beginning after the Second World War and gaining momentum through the latter decades of the twentieth century in response to a host of social and economic forces. These include rising crime in many countries, especially from the 1970s to the 1990s; increasing prosperity, resulting in more private property and consumer goods needing protection; the expansion of privately controlled, publicly accessible spaces ranging from hypermarkets to airports with their own distinct policing needs; a general growth in the subcontracting of security functions within both the public and private sectors, in order to limit spending and gain from the economies of scale achieved by specialist providers; and increased safety concerns on the part of companies, public institutions, and private individuals in an increasingly risk averse world. Alternative arrangements to the police are also commonplace in poorer parts of the world: in many instances private security or voluntary provision may be a preferred substitute for police forces that lack legitimacy with the public and may be ineffective in the face of high crime levels, compelling rich and poor alike to take matters into their own hands.

Such factors have promoted a growing governmental and social acceptance of private security, as it has become increasingly recognized that the private sector can make a significant contribution to the security of organizations and citizens. At the same time, a prospering security industry has been able to develop products and services better tailored to the needs of its customers, capitalizing on benefits ranging from economies of scale to technological advancement. Criminologists need to recognize these changes and adapt their focus accordingly; hence, the aim of this entry is to analyze and evaluate the restructuring of policing in late modern society with particular reference to privatization trends.

Police or Policing?

It is important to define what is meant by "policing" as it is a term that is often misunderstood. Policing is often seen as a little more than what *the police* do; indeed, it is common to find many books and articles on policing dedicated solely to the police. It is, however, only relatively recently that the two concepts have become so closely associated in the English language. Before the mid-eighteenth century, "policing" encompassed a much broader concept of the regulation of government and the morals or economy of society (Johnston 1992). It was only after the mid-eighteenth century that the word "police," borrowed from French, came to be commonly associated with the maintenance of order and crime prevention, as Sir Robert Peel's model for London's Metropolitan Police was exported throughout Britain, as well as influencing the development of municipal policing in the United States, Europe, and parts of the British Empire.

A problem that has emerged as a consequence is the assumption that solely "the police" can or should undertake "policing." It is important that these terms are differentiated because there are many other agencies engaged in this process. The noun refers to a particular organization, while the verb describes a social process of which the police *and other phenomena* are a part (Reiner 1994). It has indeed been recognized that the vast bulk of policing is carried out by people and organizations other than the police.

Reiner (1994) argues that the essence of policing is those activities aimed at ensuring the security of a particular social order. This distinguishes policing from the broader criminological concept of social control, in that the latter encompasses all those processes and activities that promote and enforce a society's norms and values. Policing is an aspect of social control, encompassing systems of surveillance combined with the threat of sanctions for breaches of criminal law, for the purpose of maintaining a society's security. Reiner defines it as:

> an aspect of social control processes which occurs universally in all social situations in which there is at least the potential for conflict, deviance, or disorder. It involves surveillance to discover actual or anticipated breaches, and the threat or mobilization of sanctions to ensure the security of order. The order in question may be based upon consensus, or conflict and oppression, or an ambiguous amalgam of the two, which is usually the case in modern societies (1994: 722).

In delivering these functions, the police typically conduct patrols, respond to incidents, undertake investigations, apprehend offenders for breaches of the criminal law, and refer cases with sufficient evidence to prosecutors for prosecution. Those found guilty by the courts are then subjected to punishment. Yet there are also many other bodies engaged in these processes. Some citizens form voluntary groups to patrol the streets, in some cases acting as vigilantes to apprehend offenders and hand suspects over to the authorities or even to initiate their own brand of "justice." Many private security firms patrol areas, conduct investigations, and refer some of the offenders they find to the police. There are also other public bodies engaged in these processes that enforce particular areas of the law, conducting investigations and bringing prosecutions, such as the Health and Safety Executive (HSE) in the UK. Policing can alternatively be carried out by new technology controlled by humans, closed circuit television (CCTV) being a good example whereby the presence of such cameras, in town centers, private premises, and even on the roads against speeding drivers, may deter deviant behavior, its recordings able to be used for investigating and prosecuting criminals. The secret activities of intelligence agencies, targeted against espionage, terrorism, or organized crime, form another important dimension of these processes. Thus, when the broader concept of policing is dissected, the wide range of bodies – in both the public and private sectors – engaged in these activities becomes apparent. This entry will focus upon the commercial aspects of this.

History of Privatization?

The history of policing in most countries is largely dominated by private and voluntary forms of policing. It is only with the emergence of modern industrial countries towards the end of the nineteenth century that the public police became the dominant model in the delivery of policing (Johnston 1992). Private and voluntary forms of policing did not disappear but rather were eclipsed by the dominant public arrangements. The emergence of commercial private security companies, however, is relatively new. In the USA such companies can be traced to the late nineteenth century and the rapid emergence of company towns surrounding the coal, iron, and steel industries.

There is a body of literature that is very critical of the emergence of private security companies at this time in the USA. This "radical" perspective views the growth of private policing as an inevitable consequence of the crisis of capitalism whereby the state draws in the private sector to strengthen its legitimacy (Button 2008). The development of the coal, iron, and steel industries took place in rural areas in which the state was not well established. Industrial militancy threatened the corporations of the time, and as a result companies resorted to private policing to maintain control. These measures went well beyond keeping the peace, being employed to ensure the working classes remained obedient. The companies would draw upon their own private forces but also made use of contractors for investigations and policing industrial disputes, most famously Pinkerton, a firm still in operation today across a range of countries as part of the Securitas group of companies.

Companies such as Pinkerton provided a range of services that included general property protection services and a range of strikebreaking services such as labor espionage, strikebreakers, strike guards, and strike missionaries (those who were paid to convince strikers to go back to work) (Weiss 1978). The policing practices that emerged, particularly during industrial action in this period, led to serious confrontations between the forces of capital and labor. There were many disputes in which the civil rights of striking workers were violated, and in some cases individuals were even killed. Such were the problems throughout this period that "private police systems" were the subject of many congressional reports (see United States Committee on Education and Labour 1971, a publication of a 1931 report). As the 1931 Congressional report argued:

> The use of private police systems to infringe upon the civil liberties of workers has a long and often blood stained history. The methods used by private armed guards have been violent. The purposes have usually been to prevent the exercise of civil rights in the self-organisation of employees into unions or to break strikes either called to enforce collective bargaining or to obtain better working conditions for union members.

This extract is typical of a widespread view in America at that time that conceived private security as involving private armies/spies and viewed this as a threat to the public interest, as a danger to society, and as a phenomenon that should be restricted in the roles it undertook.

It is the post-Second World War period, however, that has witnessed the emergence of the modern private security industry and its expansion to the extent that in many countries there are now more private security officers than police officers. Indeed, Johnston (1992) has described the changes as the "rebirth of private policing." For example, G4S, the largest security company in the world employing over 635, 000 personnel in more than 125 countries, emerged during this period. In 1951 Plant Protection Ltd was formed in Macclesfield, England. By 1958 it had over 200 employees and was undertaking its first cash-in-transit work. The name of the company was changed to Factory Guards in 1963, and the corporate member companies Store Detectives Ltd and Securitas Alarms were founded. In 1968 the company's name was changed yet again, to the famous Group 4 Total Security Ltd, and the company has since through mergers and acquisitions grown into the largest in the the world (George and Button 2000), adopting the G4S brand identity in 2006.

What Is Privatization?

Privatization, and what constitutes it, is the subject of much academic and political debate. It can vary from very wide understandings relating to the shrinking of the state to the more precise definition of replacing public sector workers with identical private sector personnel. The underlying theme, however, is the reduction in the role of the state and a priori belief that the private sector is more efficient and effective at providing goods and services (Button 2008). Butler (1991) has identified a range of privatization policies, starting with the most complete – the sale of state assets to the private sector. Secondly, there is deregulation, whereby the regulatory burden on industries is reduced or removed, which may as a consequence lead to greater private sector involvement. Thirdly, there is contracting out, whereby public authorities contract out to the private sector to undertake functions that previously they provided directly. Finally, Butler distinguishes vouchers, where the government gives recipients of their service vouchers to shop around for the best service. Privatization could even involve the removal or reduction of subsidies or the provision of tax incentives (Johnston 1992). Where direct privatization has not been possible, attempts have been made at introducing market and private sector practices in the public sector. This led to the rise of what has become known as the "new public management" or "managerialism" (Loveday 1999). It is also important to note that privatization does not always mean an increasing role for private companies vis-à-vis the state. There are other non-state organizations that have also benefited (Johnston 1992), such as charities, voluntary organizations, and consumer groups, by replacing the public sector in fulfilling governmental functions such as areas of market regulation or the delivery of welfare or justice services.

Privatization and the Police

In applying forms of privatization to the police, a range of different policies have been pursued in different countries (Chaiken and Chaiken 1987; Johnston 1992; Forst and Manning 1999). They range from the complete replacement of police departments with private security staff (O'Leary 1994) to hiring out officers to private organizations, to the charging of fees for services provided (Johnston 1992), for example, the policing of public events. Before international developments in police privatization are assessed, it is useful to examine scholars' attempts to classify applications of these measures.

In a major study of how public policing could be privately provided in the USA, Chaiken and Chaiken (1987) identified four mechanisms. The first they called **default transfer**, whereby the police are unable for various reasons to meet public demands leading the private sector to step in to fill the gap. The second they called **accommodation and cooperation**, whereby the police informally allow private security personnel to carry out tasks they do not wish to do, in return for which they provide additional services. In an example of this policy, they illustrate how the policing of the homeless in bus and railway stations in some areas has been handed over to the security personnel employed at those venues. In return, if required to deal with an incident at a station, the police will respond immediately. The third type of privatization they termed **legislation**, whereby private security personnel have been given specific roles and powers in statute. For instance, in some US states retail security personnel have been given special powers of arrest and prosecution, meaning the police do not have to become so involved. Finally, they distinguished **contracts**, whereby the government has contracted with bodies to carry out specific tasks. The main criticism that can be leveled against Chaiken and Chaiken's classification is that it does not embrace all the policies of privatization that have been applied to the police. For instance, as Johnston's classification reveals, it does not reflect how the police have embraced commercial practices such as sponsorship and charging fees.

Johnston (1992) identifies three trends exemplifying what he terms the "privatization mentality" that has been applied to policing. Firstly, there is **load shedding**, directly and indirectly. This is where the police relinquish roles directly to the private or voluntary sectors, or these effectively usurp the police because the latter are unable to provide the service the public want. Secondly, there is **contracting out**, whereby government bodies or even the police themselves contract with a private organization to undertake a function for which they remain in ultimate control. Finally, Johnston identifies **charging fees and selling services**, whereby the police have begun increasingly to act in a commercial manner by offering their services for a fee, potentially in direct competition with private security. Johnston's classification can be strengthened by including, within load shedding, Chaiken and Chaiken's example of **accommodation and cooperation**. Further, by widening Johnston's final category to the **embracing of private sector management practices**, the broader range of private sector strategies increasingly being adopted by the police can also be assessed. This entry will now use Johnston's amended trichotomy to evaluate the impact of privatization on policing across the world.

Load Shedding

This is where the funding mechanism as well as the delivery of the service is moved from the public to the private or voluntary sectors. It can happen directly, whereby the police actively abandon certain functions or indirectly whereby, because of constraints on resources, they are unable to supply a service that is wanted and the private or voluntary sector steps in to fill the gap.

One of the most significant areas in which the private sector has sought to provide services when the police are unable to meet demand is in the provision of patrol (see Noaks 2000). One of the best documented examples of the private sector successfully filling this gap is the emergence of the private security firm Intelligarde in Toronto, Canada, to police public housing estates. Rigakos (2002) shows how these deprived estates were largely abandoned by the police to disorder and crime. Intelligarde, modeled as the "parapolice," with a tough, enforcement-focused approach, reestablished a degree of security that the police were unable to guarantee.

Canada is not the only country in which other actors have filled the gaps that the police are unable to address. In South Africa, beset by high crime levels and a historically poor relationship between the police and the community, it is commonplace among rich and poor alike to take matters into their own hands through the hiring of private security or community initiatives (Wood and Shearing 2007; Steinberg 2008; Marks and Bonnin 2010). As Marks and Bonnin (2010) have documented, these can take a number of guises, adopting such descriptions as street committees, neighborhood associations, street patrols, community policing forums, and neighborhood watches depending on their emphases and objectives.

There have also been examples of more purposeful, direct load shedding. Over the last 50 years, the history of the British police has been one of incremental reduction of tasks. The police no longer check unoccupied properties or escort cash-in-transit vehicles, which they once did routinely among many other functions. There are other areas of policing where load shedding has been prominent in more recent years. Public order policing in the UK is a significant dimension in which the police have sought to reduce their role. At football matches up to the late 1980s, it was common for many police officers to be deployed within the grounds to undertake safety and public order policing functions. As a consequence of the Hillsborough disaster, when a human crush resulted in the deaths of 96 people, there were a number of reports into the safety and security strategies at football matches, the most important of which was the *Taylor Report* published in 1990. Following its publication, football clubs agreed to undertake a greater role in safety with the police concentrating on crime, public order, and emergency management. As a consequence, police presence at most routine matches has declined significantly and been replaced by stewards and safety officers. Indeed, there are some Premier League football matches that take place with no police officers within the ground. To illustrate this change, in 1989 Nottingham Forest Football Club would typically have had 150 police officers supported by 75 stewards in the ground. In 1995 this had fallen to 22 police officers along with around 250 stewards inside the ground (Frosdick and Sidney 1997).

A further area where the police have effectively shed their role in the UK is the investigation of fraud. During the 1950s and 1960s, if an organization suffered a fraud internally, it would be common for the police to undertake the investigation. Such is the extent of fraud in many organizations today that the police will normally require prima facie evidence before they become involved. They may, indeed, even expect the investigation to have been completed by private means. Indeed, the recent Fraud Review Team (2006: 69) in the UK uncovered the following response from a chief constable to an organization that had reported a £100,000 employee fraud:

> The investigation of fraud is extremely expensive in terms of hours spent obtaining statements and preparing a prosecution case. The Constabulary is required under the Crime and Disorder Act to produce a crime reduction strategy. Our strategy identifies priority areas and police resources are directed to those priority areas. Fraud is not one of them.

As a consequence of this – and often a reason to keep control of the investigation – many organizations have turned to the private sector to undertake the initial and sometimes the whole fraud investigation (Fraud Review Team, 2006). This type of load shedding has exhibited both direct and indirect tendencies. The police have been able to avoid having to investigate some frauds, which can be labor intensive, but may capitalize on the benefits of a successful investigation by the private sector.

At the international level, the limitations of military provision in unstable regions of the world, as well as corporate advancement into emerging markets including areas of high security risk, have also opened up substantial opportunities for the private security sector. O'Reilly (2010) notes how the longer established of the international security consultancies, such as the British firm Control Risks Group, emerged in the 1970s as many corporations began to extend their reach into high-risk regions of the world and faced, in particular, the threat of executive kidnap and ransom. With respect to this problem in Latin America, he explains, "By making this threat manageable, they also made it insurable" (p. 185). Since that time, the relationship between such consultancies and the insurance sector has been fundamental to their existence and development. O'Reilly observes that these early successes began to assuage initial corporate fears about moving into hazardous environments and encouraged further ventures that, in turn, required additional specialist services and cross-selling of tailored insurance products.

One element of the security industry's international expansion that has evoked particular controversy is the provision of military-related services. Avant (2009) notes how the accelerated expansion of this sector can be observed to have occurred in two distinct phases, associated with Africa's adjustment to a post-Cold War, postcolonial era in the 1990s, and the conflict in Iraq in the new millennium. The earlier phase saw the sector and its activities constructed as a matter of growing political concern, whereas the latter has seen it establish itself as serving an essential role in an era of uncertainty, albeit with acknowledged and often serious limitations still remaining. The downsizing by many countries of their military forces after the end of the Cold War has provided extensive opportunity, indeed an expressed need, for the private sector to play a role in defense, security, peacekeeping, and reconstruction. With respect to Iraq, Isenberg (2009) identifies the three main roles of private security companies as being as follows: personal security for senior civilian officials, nonmilitary security of buildings and infrastructure, and nonmilitary logistics security, often through subcontracting by US government or Coalition Provisional Authority (CPA) contractors or employment by Iraqi or foreign companies. It should be noted that, for the security industry today, Iraq is a somewhat unique case in that the demand for private security is closely tied to the Western military intervention in that country. Of the two main leading supplier countries of private security, the USA and the UK, British industry spokespersons Bearpark and Schulz (2007) argue that most UK private security providers along with other European companies largely refrain from providing frontline services in conflict zones, placing more reliance on private sector clients who are protective of their corporate reputations. Thus, the UK security sector is gaining much ground in another area of burgeoning marketplace demand: global maritime security.

Contracting Out

In this form of privatization, a governmental body retains responsibility for the funding of the activity but contracts with a third party to deliver the service. The Confederation of European Security Services (CoESS), a joint initiative of European trade associations for the security industry, observed that the UK is the country that has gone furthest in transferring traditionally public security functions to the private sector (Palmer and Button 2011), applying not only to police organizations but the justice system as a whole. Its immigration detention functions have been contracted out since the 1970s, and a wave of subsequent outsourcing commenced in 1992 with the privatization of prisoner escort services, certain prisons (reported to number 11 of 130 at the time of writing of the CoESS report), and some security functions within public prisons, such as the monitoring of offenders under curfew, the security of the lower courts of law (magistrates' courts), and the securing of many police custody suites or police premises as a whole. In the case of police organizations, such practices have gradually extended from relatively uncontroversial functions such as catering, cleaning, and maintenance to "back office" tasks hitherto undertaken by police officers (Button, Williamson and Johnston 2007).

Embracing Private Sector Management Practices

The third aspect to privatization is the embracing of private sector management practices. In the UK throughout the Conservative periods of government of the 1980s and 1990s, and continuing under the Labour governments from 1997 to 2010, there has been a gradual application of market principles within the public sector, including the police. This process has been termed "managerialism" or "new public management" (NPM) (see Loveday 1999). Most public sector organizations were subject to a Financial Management Initiative which emphasized the "three Es": economy, efficiency, and effectiveness. The rudimentary basis of this initiative was the setting of objectives and determining priorities and, ultimately, the encouragement of modes of working more characteristic of the private sector. These principles were further cemented with the encouragement of what later became known as "policing by objectives" in the issuing of a Home Office Circular (114/1983), a means of offering central government "guidance" to police forces on a wide variety of operational and policy issues that came to be regarded as binding by most forces due to the authority held by the Home Office in police governance structures.

By the early 1990s the pace of reform had intensified. The Police and Magistrates Courts Act 1994 shifted the balance of power in the "tripartite" structure of police governance (comprising the chief constable, the Home Office, and the police authority, made up of locally elected councilors and magistrates), still further in favor of central government. It emphasized the "chief executive" role of the chief police officer (in the UK known as the chief constable or, in the two London forces, the commissioner), who was responsible for managing the police service rather than administering it. The Act also reconstituted chief constables as budget holders, enabled the Home Secretary to set national performance targets, and introduced greater business interest into the police authorities, by reducing the number of democratically elected councilors at the expense of independents (expected to be from the business community) (Loveday and Reid 2003). Another significant event was the Sheehy Inquiry into Police Rewards and Responsibilities, tasked with critically examining the thorny issues of the police rank structure, remuneration, and conditions of service and remuneration. Following considerable police resistance to the recommendations, these were largely rejected, although some changes, such as performance-related pay for top-ranking officers, were implemented. Some constabularies appointed accountants to key positions in the hierarchy, and in Merseyside the chief constable's team became known as the "Board of Directors" (Loveday 1999).

In the government plan for the criminal justice system, more secondments in the private sector for senior managers in the criminal justice system (including the police) were advocated and have now began to be implemented through the Workforce Modernisation Agenda.

The application of NPM to the public police has been the subject of much research (Loveday 1999; Loveday and Reid 2003); however, the focus of this entry will be the selling of policing services, charging of fees, and pursuit of sponsorship. The police have long charged for "special services," particularly for the officers they provide at football matches and public order events (Gans 2000). Under the Police Act 1964 (now section 25 of the Police Act 1996):

> The chief officer of police of any police force may provide, at the request of any person, special police services at any premises or in any locality in the police area for which the force is maintained, subject to the payment to the police authority of charges on such scales as may be determined by the authority.

Under section 18 of the same Act, the police authority can also supply goods and services to local authorities. It is not the aim here to explore the debate over what constitutes "special policing services," rather to illustrate some of the services the police have provided. As mentioned, it has been a long-standing practice for the police to charge for the services they provide at football matches and other public events on private land.

Such practices extend much further in the USA, where Reiss (1988) found three models of employment of the public police for private purposes. The first was where the officer found secondary employment and charged for his/her services directly. The second was through police unions who coordinated their secondary employment. Finally, a model exists where the police department contracts with the organization for secondary employment, for which they receive fees and then pay the police officer.

Through sponsorship, police forces can also raise an additional 1 % of their budget. An expose by the British national newspaper the *Sunday Times* illustrated some of the more bizarre sponsorship deals many forces would apparently consider. Posing as executives from a fictional firm called "Keystone Security" (Keystone Cops!), for the appropriate fee several forces offered to emblazon their name on their vehicles. One force offered to name an anti-burglary campaign after the company, and in another it was suggested their name could be displayed on the wall of a new police station shared with the local authority. Many forces have indeed been successful in acquiring money in return for advertising. London's prestigious department store Harrods has in the past sponsored a police patrol car, and a liquor store chain called Thresher sponsored a police van for £10, 000.

Linked to the charging of fees and selling of services, Bryett (1996) has also identified four methods in which privately owned non-police resources have been expended to the public police. At its simplest level, monies and physical resources have been given to the police. Clearly, the donation of large sums of money or resources raises concern over the independence of the police should an investigation into the donor ever become necessary. At a second level, another donation is that of space, which can also be a form of physical resource. Bryett provides one example from the USA where the McDonald's food chain gave part of one of its stores as premises for a police station. A third type of private sector aid is giving time. This most frequently takes the form of private individuals offering their time as police auxiliaries, termed "special constables" in the UK. At another level it might be helping the police in a search for a missing person or for evidence. The pursuit of cooperative ventures between the public and private sectors is the fourth means of cooperation. For instance, in Montgomery County in the USA, the local police department cooperated with IBM to produce a sophisticated computer disaster and security capability. In the UK the Private Finance Initiative has also enabled some forces to contract with the private sector to design, finance, build, and manage police facilities. Some of these have included police stations and firearms ranges.

Assessing the Impact of Privatization

The emergence of a substantial private security industry undertaking functions that, in modern times, came to be associated with the state has not been without its problems. Before some of these are examined, however, it is important to note that the outlook is not entirely negative. Private security personnel provide services that are essential to public safety and which, if the private sector did not provide them, would in many cases not be supplied at all. The private security industry can be a strong partner to the state to enhance the overall resilience to crime and terrorism, as in the case of Project Griffin, an initiative launched by the UK's City of London and Metropolitan forces in 2004. Its approach is to train private security officers to provide their own companies with better protection against terrorism, as well as to assist the police and other organizations through information sharing and securing cordons in the event of an emergency. The project has subsequently been extended around the UK, and the model exported to a number of other countries, including Canada and Australia. There is also the strong argument, based on some of the examples of outsourcing provided above, that in some circumstances the private sector can supply an equal or even better service to the state at a lower cost. This brings benefits to society as a whole, particularly in an era of fiscal constraint.

Yet it would be wrong to ignore some of the challenges associated with private security. In some countries there has been evidence of significant problems arising in association with the burgeoning size and role of the private security (Palmer and Button 2011). These have included, among others, criminal infiltration of the industry, little or no training for security personnel, abuses of authority including the excessive use of force, and generally low standards of professionalism. Some of these will be briefly considered.

There are a number of countries that do not regulate private security, or do not regulate all the private security activities in an industry, or where the enforcement of regulation is minimal. As a consequence, in some countries there are problems with individuals of inappropriate character, namely, significant criminal records, working in the private security industry. At a more serious level, on occasions, organized crime has captured control of security companies enabling them to use company resources and staff for their criminal means (SEESAC 2005). The collapse of the Soviet Union and its satellite countries has also led to the dominance of security companies by former state security services operatives in that region. In Russia between September 1991 and June 1992, over 20, 000 KGB officers left or were discharged. Many of these personnel moved to the private sector to work for and establish private security operations. Indeed, Volkov (2000) estimated that nearly 50, 000 former officers of state security and law enforcement were working for private security agencies. Many of these personnel have taken the culture and tactics of their old employers to their new private masters.

Some of the most disturbing examples of abuse and excessive force by private security personnel have been documented in South America. At the most disturbing end of the spectrum, private security has been implicated in repressive social control, including organized death squads, most notably in the slums of Brazil. In São Paulo during the last 3 months of 1999, of the 109 citizens killed by the militarized police, 69 were recorded as having been dispatched by off-duty officers working part time in private security. As Huggins (2000: 121) has argued.

> In Brazil today, death squads continue to operate, paralleled by a growing 'rent-a-cop' industry, with each social control entity directly 'serving' different population segments to the benefit of one over the other. The poor are treated as undifferentiated 'criminals' by death squads and 'rent-a-cops', and richer Brazilians are protected from the poor by their private security forces.

A major critique that has been raised by leading security scholar Clifford Shearing and a number of his collaborators is that the expansion of private security has highlighted and exacerbated security inequality (see Wood and Shearing 2007).

They do not discount the value of private security in certain circumstances, and the potential for poorer sectors of society to harness its resources in circumstances in which public provision is lacking. Yet they also observe how, in developing countries, the rich are able to purchase their own security measures to supplement slender state provision, while poorer areas are often left with a security deficit.

In addition, in many countries the private security industry also is characterized by poor quality security personnel, lowly paid, poorly trained, and lacking motivation (Button 2008). This has highlighted the need for effective regulation and control of the private security industry. Most countries do regulate the industry but the standards vary considerably (Palmer and Button 2011). The need to raise standards has invited the concern of a number of international bodies, most notably the United Nations. Work by the United Nations Office for Drugs and Crime in 2011 culminated in an intergovernmental expert group making preliminary draft recommendations on the principles of a model regulatory system for the private security industry.

Conclusion

In Western democratic countries, policing is rapidly evolving against a backdrop of significant and ongoing political and economic change. This is creating a need for police forces to establish new ways of working in order to meet the burgeoning demands upon them, as well as substantial opportunities for commercial security companies to address areas of deficit. In financially austere times and under the shadow of global terrorism, more active partnership between the public and private sectors is becoming a necessity, the counter terrorism initiative Project Griffin being a key example as to how the public sector can harness private security resources.

In parts of the developing world, police forces may be ineffective, lack legitimacy with the public, or even present a threat to citizens' security, making private or community security provision a necessity. Business opportunities in emerging economies, more politically and economically volatile than traditional markets since their legal and institutional systems are less well developed, are promoting the security industry's international expansion as it serves the needs of enterprise. A substantial, and growing, international private security industry is now an established fact of life in a globalized world.

These changing drivers of insecurity, and responses to them, illustrate that neither the providers of security nor those who study them can stand still. Criminologists need to recognize the changes and their implications in terms of the functions of policing and who delivers these, which will vary from one jurisdiction to the next. They must also take account of their ramifications for the oversight and accountability of policing agencies, both public and private, as well as the partnerships they are entering into. The pace of social change is presenting enormous challenges for policy makers, practitioners, and researchers and the need for innovative solutions.

Related Entries

- ▶ Crime on Public Transport
- ▶ Policing the Police
- ▶ Preventing of Terror at Shopping Malls

Recommended Reading and References

Avant D (2009) The emerging market for private military services and the problems of regulation. In: Chesterman S, Lehnardt C (eds) From mercenaries to market: the rise and regulation of private military companies. Oxford University Press, Oxford

Bearpark A, Schulz S (2007) The future of the market. In: Chesterman S, Lehnardt C (eds) From mercenaries to market: the rise and regulation of private military companies. Oxford University Press, Oxford

Bryett K (1996) Privatisation - variation on a theme. Polic Soc 6:23–35

Butler S (1991) Privatisation for public purposes. In: Gormley WT (ed) Privatisation and its alternatives. University of Wisconsin Press, Madison

Button M (2008) Doing security: critical reflections and an agenda for change. Palgrave, Basingstoke

Button M, Williamson T, Johnston L (2007) 'Too many chiefs and not enough chief executives'. Barriers to the development of PFI in the police service in England and Wales. Criminol Crim Justice 7:287–306

Chaiken M, Chaiken J (1987) Public policing – privately provided. National Institute of Justice, Washington, DC

Forst B, Manning PK (1999) The privatisation of policing. George Washington University Press, Washington, DC

Fraud Review Team (2006) Final Report. London: The Legal Secretariat to the Law Offices

Frosdick S, Sidney J (1997) The evolution of safety management and stewarding at football grounds. In: Frosdick S, Walley L (eds) Sport and safety management. Butterworth-Heinemann, Oxford

Gans J (2000) Privately paid public policing: law and practice. Polic Soc 10:183–206

George B, Button M (2000) Private security. Perpetuity Press, Leicester

Huggins MK (2000) Urban violence and police privatisation in Brazil: blended invisibility. Soc Justice 27:113–34

Isenberg D (2009) A government in search of cover: private military companies in Iraq. In: Chesterman S, Lehnardt C (eds) From mercenaries to market: the rise and regulation of private military companies. Oxford University Press, Oxford

Johnston L (1992) The rebirth of private policing. Routledge, London

Loveday B (1999) Managing the police. In: Horton S, Farnham D (eds) Public management in Britain. Macmillan, Basingstoke

Loveday B, Reid A (2003) Going local. Policy Exchange, London

Marks M, Bonnin D (2010) Generating safety from below: community safety groups and the policing nexus in Durban. S Afr Rev Sociol 41(1):56–77

Noaks L (2000) Private cops on the block: a review of the role of private security in residential communities. Polic Soc 10:143–161

O'Leary D (1994) Reflections on police privatisation. FBI Law Enforcement Bulletin, 21–25 September

O'Reilly C (2010) The transnational security consultancy industry: a case of state–corporate symbiosis. Theor Criminol 14(2):183–210

Palmer R, Button M (2011) Civilian private security services: their role, oversight and contribution to crime prevention and community safety. Expert group on civilian private security services. United Nations Office on Drugs and Crime, Vienna

Reiner R (1994) Policing and the Police. In: Maguire M, Morgan R, Reiner R (eds) The Oxford handbook of criminology. Oxford University Press, Oxford

Reiss AJ (1988) Private employment of public police. National Institute of Justice, Washington, DC

Rigakos GS (2002) The New Parapolice. University of Toronto Press, Toronto

SEESAC (2005) SALW and private security companies in South Eastern Europe: a cause or effect on insecurity. South Eastern Europe Clearinghouse for the Control of Small Arm and Light Weapons, Belgrade

Steinberg J (2008) Thin blue: the unwritten rules of policing in South Africa. Jonathan Ball Publishers, Cape Town

United States Committee on Education and Labour (1971) Private police systems. Arno Press and the New York Times, New York

Volkov V (2000) Between economy and the state: private security and rule enforcement in Russia. Polit Soc 28:483–501

Weiss R (1978) The Emergence and Transformation of Private Detective and Industrial Policing in the United States, 1850-1940, Crime and Social Justice, 1:35-48

Wood J, Shearing CD (2007) Imagining security. Willan, Cullompton

Probability and Inference in Forensic Science

Franco Taroni and Alex Biedermann
School of Criminal Justice, Institute of Forensic Science, University of Lausanne,
Lausanne, Switzerland

Overview

Various members of the justice system encounter uncertainty as an inevitable complication in inference and decision-making. Inference relates to the use of incomplete information (typically given by results of scientific examinations) in order to reason about propositions of interest (e.g., whether or not a given individual is the source of an evidential trace). In turn, judges are required to make practical decisions which represent a core aspect of their professional activity (e.g., deciding whether or not a given suspect is to be considered as the source of a given crime-related trace). Both aspects, inference and decision-making, require a logical assistance because unaided human reasoning is known to be liable to bias. From a methodological point of view, these challenges should be approached within a general framework that includes probability and (Bayesian) decision theory.

Introduction

Since the early 1960s, the forensic science community started to take a more explicit position with respect to the existence of a problem of interpretation and evaluation of data. In a now widely known quote, Kirk and Kingston from the University of California, Berkeley, note:

> When we claim that criminalistics is a science, we must be embarrassed, for no science is without some mathematical background, however meagre. This lack must be a matter of primary concern to the educator [. . .]. Most, if not all, of the amateurish efforts of all of us to justify our own evidence interpretations have been deficient in mathematical exactness and philosophical understanding. (Kingston and Kirk 1964, pp. 435–436)

Today, interpretation and data evaluation are still held as a neglected area, mainly in fields that involve so-called physical evidence. This is acknowledged, for instance, by institutions such as the National Research Council (2009) (see ► Critical Report on Forensic Science). In its report from 2009 (at pp. 6–3), the Council notes that "[t]here is a critical need in most fields of forensic science to raise standards for reporting and testifying about the results of investigations." In many contexts, this perception is reinforced by the fact that scientists' assessments of evidential value consist of a largely "subjective" component and a connotation of "arbitrary." This should be contrasted, however, with an alternative interpretation that views the term "subjective" as a coherent state of personal belief held by an individual about something uncertain (see also Sects. 5 and 6).

As mentioned by Kirk and Kingston (1964), it was indeed rare at their time of writing – and it probably still is often so today – that a scientist's opinion was based on a quantitative study. Besides some notable exceptions, such as forensic DNA (that is beyond the scope of this text), this is still so today even though the presentation of a general model for inference was already put forward by Darboux et al. (1908) and rediscovered later by Kingston (1965). In a more formal sense, that model is known in current days as the "Bayesian perspective" to reasoning and inference in forensic science. It constitutes a main part of an approach to accepting that forensic science as well as science in general should abandon the expression of certainty. By moving away from the notion of certainty, it becomes a logical necessity to determine the degree of belief that may be assigned to a particular event or proposition of interest. In this context, the general branch of statistics can offer a valuable contribution to science. When uncertainty exists, and data are available, then it offers the possibility to assess and measure the uncertainty based upon a precise and logical line of reasoning (de Finetti 1993). Although statistics can offer a viable contribution to data evaluation by providing methods and techniques, it is not sufficient only by itself. Statistics concentrates primarily on data, whereas the retrospective meaning of an observation relies on the more general concept of inference which focuses on the notion of uncertainty. A concise expression of this is due to an influential writer in forensic interpretation, I. W. Evett, who noted that "[p]eople call this statistics – [but] it is not actually statistics, it's inference, it's reasonable reasoning in the face of uncertainty" (I. W. Evett, "Clarity or confusion? Making expert opinions make sense," presentation held on June 24, 2009, at the event of his award of a Doctorate *Honoris Causa*, issued as part of the celebration of the 100th anniversary of the School of Criminal Justice of the University of Lausanne).

In view of these limiting aspects of contemporary uncertainty management in forensic science, this entry will cover six arguments underlying the evaluation of forensic science data. First, Sect. 3 emphasizes the inevitability of uncertainty in day-to-day scientific work, followed in Sect. 4 by an outline of the necessity to deal with such a constraint through the use of a model that helps reasoners in a legal context – be they scientists or recipients of expert information – to manage the uncertainty about issues they are faced with. Next, probability and Bayes' theorem are introduced as the essential ways of learning from experience, a process interpreted here as the updating of beliefs about an event of interest in the light of newly acquired information (Sect. 5). Some discussion is developed on how to interpret the notion of probability as a consequence of the use of the laws of probability. Section 6 introduces the personalistic (subjective) viewpoint which finds applications also beyond law and forensic science, such as economics and other scientific disciplines. A particular aspect of the probabilistic approach allows scientists to provide a measure of the value of forensic science data that a Court of Justice may incorporate in a more general framework of reasoning about ultimate target propositions. This is described in Sect. 7, followed by a general conclusion in Sect. 8. It reiterates the fundamental benefit of an approach (Bayesian) to deal with uncertainty as an indissociable feature of human activities in general.

This entry avoids a dense presentation in terms of a list of mathematical formulae to deal with different forensic scenarios and scientific data as they may be encountered with, for example, DNA, shoe marks, glass and paint fragments, and textile fibers. Full treatises as given by, for instance, Aitken and Taroni (2004) and Buckleton et al. (2005), offer detailed formal developments of this kind. The perspective here is broader and seeks to offer a more general account on the relationship between inference, probability, and forensic science.

Uncertainty in Forensic Science

Human understanding of the past, the present, and the future is inevitably incomplete. This implies what is commonly referred to as a state of uncertainty, that is, a situation encountered by an individual with imperfect knowledge. Uncertainty is an indissociable complication in daily life, and the case of forensic science epitomizes this.

The criminal justice system at large, along with forensic science, is typically confronted with unique/singular events that happened in the past, which cannot be replicated. Due to spatial and temporal limitations, only partial knowledge about past occurrences is possible. Throughout history, this circumstance substantially troubled, and continues to do so, both scientists and other actors of the criminal justice system. This discomfort clearly illustrates the continuing need to give this outset a careful attention.

Despite this substantial drawback, “the problem of uncertainty” may not necessarily be considered as unsurmountable. Indeed, the careful reader may invoke the fact that past events leave one with distinct remains and vestiges, in particular, as a result of criminal activities. These may generate tangible physical entities such as blood stains, patterns, signals, glass fragments, or textile fibers that scientists may discover, seize, and examine. Results of such examinations have a potential for informing about and retracing past and ongoing events. Moreover, one may invoke developments in science and technology that offer vast capabilities for providing scientists with valuable information. While systematic analytical testing and observation may indeed produce abundant quantities of information, much of this information may substantively lack in qualities sufficient to provide an evaluative opinion that would be needed to entail (or make necessary) particular hypotheses that are of interest to legal reasoners.

A technocratic view of forensic science in terms of machinery and equipment under controlled laboratory settings is largely artificial. Actually, forensic science operates under a general framework of circumstances that makes testing and analyses a rather challenging undertaking. It suffices to note that material related to crimes and many other real-world events may often be affected by degradation or contamination. In addition, it may also be present in only low quantities so that only limited methods or measurements can be applied. A further aspect is that specimens submitted for laboratory analysis may be of varying relevance. In short, there is uncertainty about a true connection between the material at hand and the offense or the offender. In summary, thus, questions about laboratory performance are undoubtedly important, but they are not a general guard against the rise of uncertainty in attempts to reconstruct the past.

Generating data, observations, measurements, and counts constitute a core part of forensic science practice, but it is equally important to inquire about the ways in which such information ought – in the light of uncertainty – be used to modify a reasoner's states of belief in a specific case. Legal actors have a natural interest in this inferential topic because it represents a vital step to guarantee that scientific information meaningfully serves the purpose of a particular application. This connects closely to the agreed requirement according to which procedures for learning about past events in the light of new information should be in some sense rational and internally consistent.

Scientific Inference and Induction

Learning from experience represents as much a fundamental problem of scientific progress as it presents one in everyday life. In part, experience leads to knowledge that is merely a description of what has already been observed. However, one can take this outset a step further and consider, based on past experience, prediction of future experience, referred also to as induction. Stated otherwise, on the basis of what one sees, one seeks to evaluate the uncertainty associated with a particular event of interest, even though this kind of reasoning for learning new things provides only an incomplete basis for a conclusion. In inductive reasoning, conclusions contain elements not present in the premises, and go beyond these premises, and this means that one's knowledge is extended and the inference is said to be ampliative.

Humans tend to be well accommodated with operating under and dealing with situations involving uncertainty. Thus, the mere presence of uncertainty does not hinder the practice of science in principle. Accordingly, debates should focus on uncertainty explicitly and seek to distinguish between what is more likely and what is less likely, rather than attempting to endorse a concept of certainty. With that being said, the very relevant need of science is that of learning how to deal quantitatively with "probabilities of causes" (Poincaré 1896). Indeed, a main aim of the experimental method is that of discriminating between events of interest, or "causes," in the light of particular acquired information.

The so-called Bayesian theory (Sect. 5) claims that the inferential steps involved in inductive reasoning must be accomplished by means of probability, and in particular by means of Bayes' theorem that shows how the probability assessments of future events are to be modified in the light of observed events. This is said to represent the meaning of the phrase "to learn from experience" in a mathematical formulation of induction (de Finetti 1972).

A Model for Inference: Bayes' Theorem

As argued so far, learning from experience is central to scientific inference. According to a now widely accepted view, scientific reasoning essentially amounts to reasoning in accordance with the laws of probability theory (Howson and Urbach 1993) with Bayes' theorem providing a solution to the general problem of induction. The latter is seen as an appropriate logical scheme for characterizing inferences designed to establish scientific case hypotheses. To state this more formally for a finite case setting, start by considering a set of mutually exclusive and exhaustive hypotheses $H_1,\ldots,H_n$, the collection of background knowledge B, and a set of experimental results or effects E. Bayes' theorem then says that the probability of a hypothesis of interest H_i, given E and B, is obtained as follows:

$$Pr(H_i|E,B) = \frac{Pr(E|H_i,B)Pr(H_i|B)}{Pr(E|H_1,B)Pr(H_1|B)+\ldots+Pr(E|H_n,B)Pr(H_n|B)}. \quad (1)$$

Broadly speaking, Bayes' theorem tells one how to update prior beliefs (i.e., prior to data acquisition) about hypotheses in light of data. This leads one to posterior beliefs, that is, a state of belief after that data has been acquired.

It is worth emphasizing at this juncture that the conceptual meaning of a posterior belief is not that of a rectification. On this issue, de Finetti noted the following: "For convenience, brevity, and seeming clarity, I have just spoken of opinion as being 'modified'. It would not be exact, however, to interpret this as 'corrected'. The probability of an event conditional on, or in the light of, a specified result is a different probability, not a better evaluation of the original probability. [. . .] [p]robabilities are not corrections of the first one; each is the probability relative to a specific state of information [. . .]" (de Finetti 1972, at pp. 149–150).

In the above equation, the notation Pr represents "probability" and the vertical bar | denotes the conditioning. Thus, $Pr(H_i|E,B)$ expresses the probability that H_i occurs, given that data E has occurred and background information is as described. This conditional probability of the given hypotheses H_i, after obtaining the data, is usually referred to as the posterior probability of H_i. In turn, $Pr(H_i|B)$ is the corresponding prior probability of H_i, that is, before considering data E. Each probability is conditioned on the background knowledge B.

It is possible to consider an effect E also in terms of its influence on the belief about the truth or otherwise of a hypothesis, say H, that is:

- E *confirms* or *supports* H when $Pr(H|E,B) > Pr(H|B)$.
- E *disconfirms* or *undermines* H when $Pr(H|E,B) < Pr(H|B)$.
- E *is neutral* with respect to H when $Pr(H|E,B) = Pr(H|B)$.

One can thus interpret the difference $Pr(H|E, B) - Pr(H|B)$ as a measure of the degree with which E supports H, and Bayes' theorem constitutes a logical scheme to understand how the observation of an effect supports or undermines hypotheses. On a conceptual account, Jeffrey (1975) has noted the following:

> Bayesianism does not take the task of scientific methodology to be that of establishing the truth of scientific hypotheses, but to be that of confirming or disconfirming them to degrees which reflect the overall effect of the available evidence - positive, negative, or neutral, as the case may be. (at p. 104)

For applications in forensic science, it is common to consider hypotheses in pairs. For example, H_1 may denote the hypothesis proposed by the prosecution, while H_2 denotes the hypothesis proposed by the defense. In such a setting, an alternative version of Bayes' theorem, known as the odds form, is often used. In that version, H_2 denotes the complement of H_1 so that $Pr(H_2) = 1 - Pr(H_1)$. Then, the *odds in favor of* H_1 are $Pr(H_1)/Pr(H_2)$, denoted $O(H_1)$, and the *odds in favor of* H_1 given E are denoted $O(H_1|E)$. Note that the conditioning on B is omitted for simplicity of notation. The odds form of Bayes' theorem then is

$$O(H_1|E) = \frac{Pr(E|H_1)}{Pr(E|H_2)} \times O(H_1).$$

The left-hand side of this equation is the odds in favor of the "prosecution" hypothesis H_1 after the scientific evidence has been presented, $O(H_1|E)$. This is referred to as the *posterior odds*. The odds $O(H_1)$ are the *prior odds* as held by a reasoner before considering the observation. The ratio which converts prior odds to posterior odds is the fraction

$$\frac{Pr(E|H_1)}{Pr(E|H_2)},$$

known as the *Bayes' factor*, or often also called *likelihood ratio*. It is commonly denoted by the letter V, short for "value." It can take values between 0 and ∞. A value greater than 1 lends support to the prosecution's hypothesis H_1 and a value less than 1 lends support to the defense's hypothesis H_2. Effects for which the value is 1 are neutral, that is, it does not allow one to discriminate between the two competing hypotheses. Note that if logarithms are used, the relationship becomes additive. This has the very appealing and intuitive interpretation of weighing evidence in the scales of justice. In statistical literature, the logarithm of the Bayes' factor became known as the *weight of evidence* and is generally acknowledged to Good (1950).

In forensic and legal settings, the value of scientific data is most beneficially assessed by finding a value for the Bayes' factor, a task commonly assigned to forensic scientists. In turn, it is the role of judge and jury to assess the prior and posterior odds. In this view, the scientist can assess how prior odds are altered by the evidence, but he cannot actually assign a value to the prior or posterior odds. In order to assign such a value, all the other evidence in the case also needs to be considered. Although the adequacy of such a perspective continues to be debated in both literature and practice, this tends to be less the case today than during the past decades (Redmayne et al. 2011). Probably, it is also fair to say that there currently appears to be no other method that is of better overall assistance to evidential reasoning in judicial contexts. This is maintained, for example, by Ian W. Evett, a pioneer forensic statistician, who has been quoted as saying (Joyce 2005, at p. 37), "That framework – call it Bayesian, call it logical – is just so perfect for forensic science. All the statisticians I know who have come into this field, and have looked at the problem of interpreting evidence within the context of the criminal trial, have come to see it as centring around Bayes's Theorem." It is not claimed, however, that Bayesian inference is a panacea for all problems in legal reasoning. But it can be maintained that it is the best available model for understanding the interpretation of scientific evidence.

Besides discrete propositions, forensic scientists may also need to deal with continuous parameters. Examples include the mean of a continuous variable or a proportion. For such applications, Bayes' theorem as defined in Eq. (1) takes a slightly different form. In the context, an unknown parameter is commonly written Θ and data available for analysis as $\mathbf{y} = (y_1,\ldots,y_n)$. In the case of a continuous parameter, beliefs are represented as probability density functions. Denoting the prior distribution as $\pi(\Theta)$ and the posterior density as $\pi(\Theta|\mathbf{y})$, Bayes' theorem for a continuous parameter is

$$\pi(\Theta|\mathbf{y}) = \frac{f(\mathbf{y}|\Theta)\pi(\Theta)}{\int f(\mathbf{y}|\Theta)\pi(\Theta)d\Theta}.$$

Subjective Probabilities in the Bayesian Model

Independent of the way in which probabilities are interpreted, they take values in the range between 0 and 1. The value 0 corresponds to an event whose occurrence is impossible. An event whose occurrence is certain has probability 1. In a common, and for various reasons appropriate, interpretation of the theory, all probabilities are considered as subjective – in the sense of "personal" – expressions of degree of belief, held by an individual. Probabilities thus reflect the extent to which an individual's knowledge is imperfect, and it is important to acknowledge that such personal belief is graduated. One can believe in the truth of an event more or less than at another one at a given time, essentially because observations can accumulate to modify the perception of the probability of its happening.

A main aspect of this interpretation is that the degrees of belief held by an individual, in order to be said "rational," need to conform to the rules of probability. As such, probability represents the quantified judgment of a particular individual. This is in opposition to other interpretations of probability, known, for example, as the "classical" or "frequentist" interpretations that consider probability in terms of a long-run frequency. These latter views may encounter problems in applicability when settings refer to non-repetitive or singular situations (Press and Tanur 2001). Indeed, the frequentist view strictly supposes a relative frequency obtained in a long sequence of trials that are assumed to be performed in stable conditions and physically independent of each other. This is typically incompatible with events and parameters encountered in such diverse fields such as history, law, economy, forensic science, and other contexts where inductive inference plays a central role. In the latter branches, the entities of interest are usually not the result of repetitive or replicable processes. On the contrary, they are singular and unique, and this makes the idea of repeating the course of time over and over again, and to note the number of occasions on which some happening in the past occurred, inconceivable.

Such complications do not arise with the personalistic interpretation because it does not consider probability as a feature of the external world, but as a notion that describes the relationship between a person issuing a statement of uncertainty and the real world to which that statement relates, and in which that person acts. For the belief-type perspective, it is therefore perfectly reasonable to assign probability to non-repeatable events. This feature renders probability as a measure of belief a particularly useful concept for judicial contexts.

An direct implication of considering probability as personal degree of belief is that the knowledge, experience, and information on which the individual at hand relies become important. That is, more explicitly formulated, an individual's assessment of the truth of a given statement or event (i) depends on information, (ii) may change as the information changes, and (iii) may vary among individuals because different individuals may have different information or assessment criteria.

Although this perspective most closely embraces the actual situation faced by an individual in a situation of uncertainty, personal probabilities are sometimes viewed cautiously. As a concept, it may appear abstract and nontrivial to capture, and typically scientists may be irrationally suspicious of articulating, yet admitting to hold genuine subjective probabilities. In the same context, people may also be reluctant to express probabilities numerically and suggest that this approach to probability is both arbitrary and, from a practical point of view, an inaccessible concept. Such perceptions are restrictive because they disregard the fact that personal degrees of belief can be elicited and investigated empirically. One possibility to effectuate this is to measure probabilities maintained by an individual in terms of bets that the individual is willing to accept. For example, an individual's probability of a proposition can be elicited by comparing two lotteries of the same price.

In order to illustrate the usefulness of such a procedure for quantifying judgments, consider a situation in which it is of interest to find a probability for rain tomorrow (Winkler 1996, e.g.; at 18). For this purpose, define the following two offers:

- Lottery A: Winning € 100 with probability 0.5, and winning € 0 with probability 0.5
- Lottery B: Winning € 100 if it rains tomorrow, and winning € 0 if it does not rain tomorrow

With this outset, one can assume that one would choose that offer which presents the greater chance of winning the price. Clearly, if one prefers Lottery B, then this signifies that one considers the probability of rain tomorrow to be greater than 0.5. By analogy, choosing Lottery A would imply that one's belief in rain tomorrow is lower than 0.5. Moreover, in a case in which one is indifferent between the two gambles, one's probability for rain tomorrow would equate the probability of winning the price in Lottery A. Therefore, one can conceive of a procedure in which one adjusts the chance of winning in Lottery A so that the individual, whose probability for a proposition of interest is to be elicited, would be indifferent with respect to Lottery B. Similarly, one can elicit an individual's personal probability for any event of interest.

Another question relates to the "appropriateness" of a set of probabilities held by a particular individual. In this context, the possibility of representing subjective degrees of belief in terms of betting rates is often forwarded as part of a line of argument to require that subjective degrees of belief should satisfy the laws of probability. This line of argument takes two steps. The first one of these proposes that betting rates should be coherent, in the sense that they should not be open to a sure-loss contract. The second one is given by the fact – established by twentieth-century writers in probability (e.g., Ramsey 1990) – that a set of betting rates is coherent if and only if it satisfies the laws of probability. Thus, if an individual assigns degrees of belief in such a manner that, as a whole, they do not respect the laws of probability, then their beliefs are not coherent. In the context, such incoherence is also called "logical imprudence."

The Bayesian Model and the Desiderata for Evidential Assessment and Interpretation

In the context of framing scientific evaluation in legal contexts, scientists and jurists wish their analytical thoughts and behavior to conform to several practical precepts, and interestingly, the

Bayesian model allows to conceptualize these precepts explicitly. It is common to refer to essentially six desiderata upon which the majority of current scientific and legal literature as well as the practice converge in their opinion, namely, balance, transparency, robustness, added value, flexibility, and logic. These notions represent desirable properties of an evaluative procedure for scientific opinions. They have been advocated and contextualized, to a great extent, by some quarters in forensic science and from jurists from the so-called New Evidence Scholarship.

For an inferential process to be *balanced* or, in the words of some authors, impartial, attention cannot be restricted to only one side of an argument. Evett (1996) has noted, for instance, that "[. . .] a scientist cannot speculate about the truth of a proposition without considering at least one alternative proposition. Indeed, an interpretation is without meaning unless the scientist clearly states the alternatives he has considered." The requirement of considering alternative propositions is a general one that equally applies in many instances of daily life (Lindley 1985), but in legal contexts, its role is fundamental. The Bayesian framework allows one to cope with this requirement in an integral way as it provides for an explicit treatment of a collection of rival propositions within a coordinated whole.

The proposal of considering alternatives should not be taken to mean, however, that scientist should address propositions in terms of probability. This is a task left to other legal actors as mentioned earlier in Sect. 5. A likelihood ratio-based approach for ascribing evidential value suggests that forensic scientists should primarily be concerned with their observations and results and not with the competing propositions that are put forward to account for such data. This distinction is crucial in that it provides for a sharp demarcation of the boundaries of the expert's and the court's areas of competence. Failures in recognizing that distinction are at the heart of pitfalls of intuition that have caused – and continue to cause – much discussion throughout the judicial literature and practice since (e.g., Koehler 1993).

Besides balance, a forensic scientist's evaluation should also comply with the requirement of *transparency*. This amounts to explain in a clear and explicit way what scientists have done, why they have done it, and how they have arrived at their conclusions. Closely related to this is the notion of *robustness*, which challenges a scientist's ability to explain the grounds for his opinion, together with his degree of understanding of the particular data type. These elements should help work towards *added value*, that is, a descriptor of a forensic deliverable that contributes in some substantial way to a case. Often, added value is a function of time and monetary resources, deployed in a way such as to help solve or clarify specific issues that actually matter with respect to a given client's objectives. These desiderata characterize primarily the scientist, that is, his attitude in evaluating and offering opinions, as well as the product of that activity.

The degree to which the scientist succeeds in meeting these criteria depends crucially on the chosen inferential framework, which may be judged by the two criteria *flexibility* and *logic*. These descriptors relate to properties of an inferential method rather than behavioral aspects of the scientist as mentioned in the previous paragraph. In a broad sense, flexibility is a criterion that demands a form of reasoning to be generally applicable, that is, not limited to particular subject matter (Robertson and Vignaux 1998). In turn, logic refers to a set of principles that qualify as "rational." There is a broad agreement among legal and forensic researchers and practitioners that the Bayesian framework for reasoning acceptably complies with these requirements nowadays. On a conceptual account, this view is supported by the fact that the paths of reasoning used to evaluate propositions in judicial contexts can be reconstructed as inferences in accordance with Bayes' theorem.

General Appreciation and Outlook

As a key point, probability is shown to be a formal quantitative framework with solid logical grounds that usefully assists legal actors, including forensic scientists, to make plain inferential subtleties that common language and unaided intuitive reasoning cannot achieve and to explain these issues. In this general framework, it is preferable to employ probability as degrees of belief, a standpoint commonly known as the subjectivist (or personalist) interpretation of probability. Such degrees of belief are personalized assessments of credibility formed by an individual about something uncertain, given information that the individual knows or can discover. This usage of the probability apparatus implements a concept of reference in which personalized weights may be assigned to the possible states of selected aspects of real-world problems. This closely embraces the actual situation encountered by legal practitioners. Even though uncertainty is inevitable in principle, the position of legal actors is one which allows at least some further evidence to be gained by enquiry, analysis, and experimentation. As a consequence of this, some means is required to adjust existing beliefs in the light of new information.

As a second main element, this entry favors the requirement of adjusting personal beliefs according to Bayes' theorem, essentially because it is a logical consequence of the basic rules of probability. This perspective is normative, in the sense that if one accepts the laws probability as constraints to which one's beliefs should conform, then one also needs to accept the ensuing properties – among which is Bayes' theorem. Interpreting these rules as a norm, or a standard, allows for a prescription of how a sensible individual ought to reason, to behave, or to proceed. Although practical experience shows that people in general, including scientists, may not conform well in their reasoning with the laws of probability – in fact, descriptive studies actually show that people's intuitive judgments diverge from the probability standard – the argument is that the concept *should* guide in the inferential processes in which one engages. It provides a defensible framework that allows one to understand how to treat new information, how to critically analyze, and how to revise opinions in a coherent way. Arguably, probability as a measure of uncertainty takes the role of fundamental quality criteria in forensic science. The benefits of a probabilistic framework are both immediate and uncontroversial because of the direct effect that uncertainties may have on a line of reasoning which matters in some way to an evaluator or decision maker.

From a historical perspective, Bayes' (1702–1761) theorem came about 250 years ago, but the attribute "Bayesian" as a descriptor of a particular class of inference methods appears to have gained more widespread use only since the middle of the twentieth century (Fienberg 2003). In the legal context, there now is a growing community that has spent many years applying Bayesian principles to infer from evidence. This interest in Bayes' theorem is reinforced because of its supportive capacity in explaining the concept of evidence and proof in both science and law and other inferential contexts, but also because formal procedures can be easier to explain and justify than informal ones.

Still within a historical perspective, practical applications of patterns of reasoning corresponding to a Bayesian approach can be found, for example, as early as the beginning of the twentieth century (Taroni et al. 1998). For example, at the Dreyfus's military trial held in 1908, Henri Poincaré invoked Bayes' theorem as *the only* way in which the court ought to revise its opinion about the issue of forgery. Subsequently, Bayesian ideas for inference entered legal literature and debates more systematically only in the second half of the twentieth century. Kingston (1965), Finkelstein and Fairley (1970), and Lindley (1977) are some of the main reference publications from that period. Later, within the 1990s, specialized textbooks from Aitken and Stoney (1991) and Robertson and Vignaux (1995) appeared. During the past decade, further textbooks along with a regular stream of research papers focusing on Bayesian evaluations of particular categories of evidence, such as glass or DNA (Buckleton et al. 2005) were published. But yet, if one looks at science more generally, the interest in Bayesian methods to solve problems relating to data analysis appears to be relatively scarce, with forensic science being no exception (Taroni et al. 2010).

Although much of the current discussion essentially focuses, in one way or another, on the evaluation and manipulation of personal beliefs held in propositions about which one is uncertain, reality shows that what one believes is essential, but not only in its own right. Along with a reasoner's preferences, personal beliefs – typically expressed in terms of probabilities – are a fundamental ingredient of considerations that seek to analyze what one ought to decide. Indeed, for the criminal justice system, reasoning about propositions of interest is not the end of the matter. It is commonly understood that, ultimately, decisions must be taken and that these are at the heart of judicial proceedings. For instance, a court may have to decide if it finds a suspect guilty of the offense for which he has been charged, with the obvious aim that such decisions are made accurately (Kaplan 1968). In this broader perspective, probability can be taken as a preliminary to decision-making. On a formal account, this logical extension is embedded in decision theory, which provides the means for dealing with the key elements of decision analysis processes. Since both forensic scientists and recipients of expert information encounter the recurrent need to *decide* on matters that relate to their respective areas of competence, Bayesian decision analysis represents a topic of ongoing interest. Although its relevance from a legal point of view has already been recognized decades ago (e.g., Kaplan 1968; Lempert 1977), it is only more recently that genuine forensic applications (Taroni et al. 2005; Biedermann et al. 2008) have been approached.

Related Entries

- ▶ Forensic Science and the Paradigm of Quality
- ▶ Forensic Science Culture
- ▶ Forensic Science
- ▶ Philosophical Basis of the Forensic Process
- ▶ Philosophy of Forensic Identification

Recommended Reading and References

Aitken CGG, Stoney DA (1991) The use of statistics in forensic science. Ellis Horwood, New York

Aitken CGG, Taroni F (2004) Statistics and the evaluation of evidence for forensic scientists, 2nd edn. Wiley, Chichester

Biedermann A, Bozza S, Taroni F (2008) Decision theoretic properties of forensic identification: underlying logic and argumentative implications. Forensic Sci Int 177:120–132

Buckleton JS, Triggs CM, Walsh SJ (2005) Forensic DNA evidence interpretation. CRC Press, Boca Raton

Darboux JG, Appell PE, Poincaré JH (1908) Examen critique des divers systèmes ou études graphologiques auxquels a donné lieu le bordereau. In: L'affaire Dreyfus – La révision du procès de Rennes – Enquête de la chambre criminelle de la Cour de Cassation. Ligue française des droits de l'homme et du citoyen, Paris

de Finetti B (1972) Probability, induction and statistics. Wiley, New York

de Finetti B (1993) The role of probability in the different attitudes of scientific thinking. In: Monari P, Cocchi D (eds) Bruno de Finetti, Probabilità e induzione. Bibliotheca di STATISTICA, Bologna, pp 491–511, 1977

Evett IW (1996) Expert evidence and forensic misconceptions of the nature of exact science. Sci Justice 36:118–122

Fienberg SE (2003) When did Bayesian inference become "Bayesian"? Bayesian Anal 1:1–41

Finkelstein MO, Fairley WB (1970) A Bayesian approach to identification evidence. Harv Law Rev 83:489–517

Good IJ (1950) Probability and the weighing of evidence. Griffin, London

Howson C, Urbach P (1993) Scientific reasoning: the Bayesian approach, 2nd edn. Open Court, La Salle

Jeffrey RC (1975) Probability and falsification: critique of the popper program. Synthese 30:95–117

Joyce H (2005) Career story: consultant forensic statistician. Communication with Ian Evett. Significance 2(34–37):2005, March 2005

Kaplan J (1968) Decision theory and the factfinding process. Stanf Law Rev 20:1065–1092

Kingston CR (1965) Application of probability theory in criminalistics – II. J Am Stat Assoc 60:1028–1034

Kingston CR, Kirk PL (1964) The use of statistics in criminalistics. J Crim Law Criminol Police Sci 55:514–521

Koehler JJ (1993) Error and exaggeration in the presentation of DNA evidence at trial. Jurimetrics J 34:21–39

Lempert RO (1977) Modeling relevance. Mich Law Rev 75:1021–1057

Lindley DV (1977) Probability and the law. Statistician 26:203–220

Lindley D (1985) Making decisions, 2nd edn. Wiley, Chichester

National Research Council (2009) Strengthening forensic science in the United States: a path forward. National Academy Press, Washington, D.C

Poincaré H (1896) Calcul des probabilités. Leçons professées pendant le deuxième semestre 1893–1894. Gauthier-Villars, Paris
Press SJ, Tanur JM (2001) The subjectivity of scientists and the Bayesian approach. Wiley, New York
Ramsey FP (1990) Truth and probability. In: Mellor DH (ed) Philosophical papers. Cambridge University Press, Cambridge, pp 52–109, 1926
Redmayne M, Roberts P, Aitken CGG, Jackson G (2011) Forensic science evidence in question. Crim Law Rev 5:347–356
Robertson B, Vignaux GA (1995) Interpreting evidence. Evaluating forensic science in the courtroom. Wiley, Chichester
Robertson B, Vignaux GA (1998) Explaining evidence logically. N Law J Expert Witn Suppl 148:159–162
Taroni F, Champod C, Margot P (1998) Forerunners of Bayesianism in early forensic science. Jurimetrics J 38:183–200
Taroni F, Bozza S, Aitken CGG (2005) Decision analysis in forensic science. J Forensic Sci 50:894–905
Taroni F, Bozza S, Biedermann A, Garbolino G, Aitken CGG (2010) Data analysis in forensic science: a Bayesian decision perspective. Wiley, Chichester
Winkler RL (1996) An introduction to Bayesian inference and decision. Probabilistic Publishing, Gainesville

Probation

▶ Behavioral Management in Probation
▶ Desistance and Supervision
▶ Mentally Disordered Offenders Under Community Supervision

Probation and Community Sanctions

Pascal Décarpes[1] and Ioan Durnescu[2]
[1]University of Greifswald, Greifswald, Germany
[2]Faculty of Sociology and Social Work, University of Bucharest, Bucharest, Romania

Synonyms

Conditional conviction; Parole; Suspended sentence

Overview

Both terms "probation" and "community sanctions" referred during more than a century to different realities and practices according to the country where it proceeded. Nevertheless, it is striking to notice that these two concepts became essential to criminal policies following similar developments based on the same grounds. Whereas probation consists in an institution supervising an offender who has been sentenced to certain obligations and constraints, community sanctions are a practical extension of it and can be defined as a penal punishment situated between imprisonment and financial penalties ("intermediate sentence"; Bottoms et al. 2001). Whereas custodial sentence is served inside prison structures, community sanctions are executed in the free society along any form of social interaction. As long as the common sanction was of a custodial nature, its implementation was fulfilled by the prison services. But as soon as other types of sanction appeared in criminal justice systems, a new form of penal punishment had to emanate. Thus, the link between probation and community sanctions goes back to the nineteenth century at a time where modern modalities of sentence execution had to be combined with the emergence of supervising agencies.

In order to understand how it arose in this frame, one way to consider a community sanction considered as "normal" is to expect it to comply with the following criteria: It should not be custodial; it has to reach a balance between harshness and laxity (also proportionality, see Bagaric (1999)); it should not be too onerous and has to be transferred to social services; it has not to be supervised by over-skilled staff; it should be simple and quick in its implementation, and the community sanction and its supervision shall have a control and disciplinary aspect per se (Normandeau 1979). This definition has been shaped over decades through different types of classification. One approach differentiates three types of rationality in which the probation system has prospered: the discipline rationality during the 1850–1940s period; then the welfare rationality along the 1950–1970s, also called the "probation's golden years"; and lately the security rationality as from the 1980s (da Agra 2010).

Beyond that, a more differentiated and comprehensive step-by-step approach highlights the evolution of probation work. The early days were heavily influenced by the church and charity organizations from origins until 1878 with religious missionaries. The years of professionalization with focus on personality and training arose during the first half of the twentieth century: Probation became a professional undertaking rather than a mission. The third wave is the years of casework including good relationship and diagnosis around the 1960s and consisting in a direct method of intervention relying on a strong and open relationship between the probationer and a trained probation officer. A backlash nourished some years of pessimism and to the necessity to help in the client's terms (1970 and 1980s): high rates of recidivism conducted to a more personalized but also control-oriented supervision. A way out was then meant as rehabilitation revival insisting on "what works" along with limited "transfer" into the practitioners' daily work (1990s): The policy emphasis was on procedure and technicality. The current phase concentrates on desistance and other emerging approaches, especially with the return to morality (2000s): Desistance is placed at the intersection between structure, agency, and reflexivity (Durnescu 2012).

The development of community sanctions has followed a parallel path since they are inherently interdependent with probation services, the former being generally supervised by the latter. An attempt to analyze its emergence underlines six rather consecutive phases starting with the origins at a time when community sanctions were considered as privileges. After World War I (1914–1918), supervised community sanctions were introduced. It took half a century to introduce intermediate sanctions characterized by intensive surveillance (1970–1980s), developing then restorative community sanctions in the 1980s, technological community sanctions as of 1990s and postmodern community sanctions with extended supervision as of the end of the 1990s (Durnescu 2009).

Brief History of Probation Services

Probation, despite all differences among countries around the world, has one "common penal cornerstone" represented by the prison, and its development, besides various independent socio-political characteristics, is contingent to critical opinions toward the use of imprisonment and its effectiveness, and "the normative aspirations originally invested in the prison were transferred symbolically (if not always in substance) to non-custodial measures and, in particular, probation in its various forms" (Vanstone 1998, p. 751).

Still, the first practical steps of probation were due to philanthropy, and religious and humanitarian concerns that led to the first probation model in England (Jarvis 1972; van Kalmthout and Durnescu 2010). However, social control and the maintenance of social order remained primordial in the definition and implementation of probation and community sanctions. Criminology at the end of the nineteenth century fed with its new concepts the ideology of promising penal policies while permitting welfare at the same time that it kept control over poor and marginalized social groups (Vanstone 2008). Besides, the post-first-world-war period was an exciting time for new penal ideas in such a way that two major scientific reviews were launched: the British "Probation Journal" in 1929 and the US "Federal Probation" in 1936.

Until the end of the eighteenth century, there was no kind of help or support foreseen for released prisoners, also because deprivation of liberty as a penal punishment, (conditional) release, and aftercare measures were not regulated by criminal law. In countries where these types of sanctions existed, charity organizations took care of prisoners and, incidentally, of the released ones. The origins of probation go back to England and the philanthropist John Howard (1726–1790) as one of its key-thinkers. He believed in the importance, even the necessity for a society to provide released prisoners with law-abiding resources to survive. His concept led shortly after his death to a law dealing with patronage (1792). In many countries, charitable clubs and societies were the only bodies officially recognized to deal with detained or released offenders by giving them material and spiritual (often Christian) support (van Kalmthout and Durnescu 2010). Their work involved a strong moral component and focused deeply on the offender's personality.

Two Founding Fathers of Probation

Matthew Davenport Hill, at that time a young British lawyer, assisted in a court in 1820 at a new way for magistrates to prevent juvenile criminals going to prison. The idea was to condemn to solely 1 day of imprisonment, provided that the juvenile has to be given back to his father or tutor with the obligation to be kept and surveyed more carefully (United Nations 1951). Later in 1840, as a judge at the Birmingham court, he kept practicing this symbolic 1-day sentence under the condition that the offender remains under supervision of an adult. M. Hill regarded individual cases as hopeful, because he thought that most of the people are not entirely criminals and it is always possible to "correct" them. Moreover, the supervision by responsible and willing persons who are in position to commit themselves to support young offenders shall, in Hill's mind, give better reintegration opportunities than custodial facilities. There was no system of legal surveillance, but Hill requested the police head officer to proceed to steady inquiries on offenders' behavior. This can be seen as an early form of police surveillance in the area of community sanctions. After 7 years of practicing this experience, the judge Hill presented his results as follows: "[. . .] young offenders who kept to the straight and narrow are much more numerous among those who have been treated this way than those who haven't had this privilege and served prison sentences (United Nations 1951). Another judge, Edward William Cox, developed the concept with the designation of an inquiry agent who was in charge of supervising several offenders. This task was then taken over in 1876 by the Church of England Temperance Society that appointed "police court missionaries" who tried to keep offenders away from consuming alcohol and help them to resettle (van Kalmthout and Durnescu 2008, p. 4). The first mentioning of probation is to be found in the 1886 "Probation of First Offenders Bill," definitely structuring the field with the 1907 "Probation of Offenders Act" creating probation services all around the country.

A second character has also shaped the probation history (Vanstone 2008) since John Augustus (1784–1859) is considered as being the first probation officer. Born in Woburn, Massachusetts, he was a thriving shoemaker living in Boston. Taking part as a spectator at a police court audience in 1841, he asked the judge to guarantee a patronage to a drinking offender during a period of probation. After extending his supervision activities to other types of offense, he published a sum of his achievements under the title "A report of the Labors of John Augustus, for the last 10 years, in aide of the unfortunate." He instituted during this period some working criteria for modern probation services such as the selection of probationers (he favored first-time offenders), social enquiries with criminal background, and personality check, he controlled individuals' behavior and made sure that they would go to school or to work (United Nations 1951). In the wake of Augustus' experiences, Massachusetts became the first US State to pass a law on probation in 1878. The development in other states was so dynamic that within 60 years, both the federal State and 34 US states had also introduced probation laws.

Creation of Probation Services in Other Countries

Probation was introduced in national legislations around the world mostly between 1878 and 1920 (Vanstone 2008) under the influence of modern criminal law that emphasized on individualization of the sentence execution and reintegration of the offender. New penal instruments arose: conditional and suspended sentence, conditional release and parole, bail, house arrest, etc. (van Kalmthout and Durnescu 2010).

In the Netherlands, the establishment of probation in 1823 took example to the English system. It offered school training and religious support, conducted with private funding and voluntary workers. In 1910, the State undertook the tasks of probation, focusing on the criminal's treatment in order to protect the society and involving a national association dealing with offenders (Jäger 2010). The probation officers gained professionalization with a 1947 regulation (*Reclaseringsregeling*) that structured their activities and curriculum. After some back and forth, the probation associations and the state justice agencies found in the mid-1980s a balance between autonomy and control.

In Germany, influenced by the French and Belgium systems, the first practice of probation (*Bewährung*) took place at the end of the nineteenth century in the frame of supervision for juveniles (Schöch 2003). Even if some associations already accompanied released prisoners, today's German probation structure was experimented in 1951 and then implemented in 1953 in both penal code and juvenile court law (Jäger 2010, p. XIV). Confined to a control aspect of the probationers, probation duties extended 1969 to further help and support activities, connecting with the extension of suspended sentences (up to 2 years) that came into force the same year. Federal statistics show since 1963 and their first data an increase of probationers, from 27,401 up to 131,381 in 1990 and 180,074 in 2010, which represents a growth of 37 % within the last 20 years, much more than the general population evolution with 2.6 % between 1990 and 2010. Though, the lack of performance or efficiency evaluation has remained controversial (Dünkel and Spiess 1983).

In France, probation is a recent institution in relation with other countries. Even if conditional release was already adopted in 1885 and supervised freedom (*mise en liberté*) in 1908, special structures for released prisoners were introduced only after the Second World War in 1946. Called "comities for help and placement of released prisoners" and placed under the hierarchy of the court president, they supervised conditional released and supported released prisoners having served the complete sentence. The contemporary probation scheme was formed by a 1958 governmental law creating the enforcement judge responsible for the execution of sentences (*Juge de l'Application des Peines* (JAP)), the suspended sentence with obligations (*Sursis avec Mise à l'Epreuve* (SME)), and the comity of probation and help to release prisoners (*Comité de probation et d'assistance aux libérés* (CPAL)). The professionalization of educators (1966) was a further step toward more numerous, and better trained staff, assimilating then in 1967 the probation functions of social workers to educators delegated to probation.

In Switzerland, albeit the Fribourg canton organized actions to support released prisoners as early as 1842, the probation movement appeared in some districts (so in Basel or Zurich) only at the end of the nineteenth century, during which societies acted to protect released prisoners from banishment – hence its appellation *Schutzverein* (protection association). At the same time, a priest named Meyer, who was during years intervening in prisons, published a brochure entitled "Call and project of status for the creation of a patronage society in favor of released prisoners," which took some decades until its first creation in 1888. Eventually, this private body was integrated and transformed into a public service in 1941 and structured in its contemporary shape in 1967, taking its current denomination of "probation service" in 2005.

On the other side of the Atlantic, directly inspired by the 1887 British "Probation of First Offenders Act," probation has been developed in Canada as of 1889 through the law enabling conditional release for first-time offenders with neither surveillance nor time limit. The first Canadian criminal code from 1892 stipulated that offenders convicted of a petty crime and sent to prison could have their sentence suspended and be released with the obligation to sign an agreement in which he commits himself to behave properly. The Canadian Parliament added in 1921 the notion of surveillance operated by a person nominated by the court. Ontario was the first province to legislate on the creation of probation services in 1922, later followed by British Columbia in 1946, which was still not the case for Quebec in 1956, even if there were already functioning help associations for released prisoners and further social agencies to control offenders under supervision. It is only in 1967 that the Ministry of Justice hired the first probation officers, embedded in a 1969 law on probation and custody. The evolution of Quebec probation services can be observed aid of the number of probation officers: 28 in 1970, then a considerable increase within 4 years (140 in 1974) and 335 in 2011.

Probation Services Today

Probation services become a reality in almost every jurisdiction in the world, especially in Europe. The comments in this section are based on the survey conducted by van Kalmthout and Durnescu in 2008 (where the source is different, it is indicated accordingly). Depending on the country size, the administrative structure, or the judicial traditions, probation services are organized either by the prosecution service (The Netherlands, Luxembourg), by the social service (Scotland), or independently close to the court system. A special group of countries have their probation service located in the third sector (The Netherlands, Austria). In their case although the probation activity is delivered by the NGOs, they are contracted and monitored by the Ministry of Justice. At the national level, some of them are coordinated by independent departments within the Ministry of Justice (Portugal, Romania) or by the prison and probation departments (France, Denmark, Estonia, England and Wales, some Federal States in Germany). Looking at the recent history, one could claim that as a European trend, probation services tend to be more and more associated with the prison departments. Most of the time, reasons for amalgamating these two structures have to do with reducing the administrative costs and with the need for a better and closer relationship between these two institutions.

The federal states however tend to display a more mixed picture. In Switzerland, for instance, the probations services are organized at the canton level. In Germany, probation service is structured around the regional courts in some Landers while in others, it is incorporated within the social services but still under the coordination of the Ministry of Justice. In the USA, probation activities are delivered by hundreds and hundreds of organizations and structures, sometimes at the federal or state level but other times even at the municipal level. As a result, probation systems in the USA are highly fragmented and local oriented with no standards and no coherence at the federal level (Burrell 2010).

Under the umbrella of "probation," a lot of activities are delivered or enforced. What they all seem to have in common is the desire to prevent re-offending. Some of these activities are based on a court decision (supervision, enforcement, programs, electronic monitoring, etc.). Some other activities are based on the law (victim assistance, half-way houses) or on a more general vision that probation service should serve the community (prevention in schools or other hot spots). There are also countries that are more and more involved in diversionary measures or even provide mediation services (Czech Republic, Slovakia, etc.). Probation services in Scotland and England and Wales deliver bail advice and supervision to prevent pretrial detention. Having admitted the huge variety of tasks that probation service is called upon, there are core probation activities that can be found in almost all the jurisdictions. They are decision support (presentence reports, social inquiry reports, etc.), supervision of offenders in the community, and they provide assistance to cover the so-called criminogenic needs. Since the early 1970s, more and more countries have introduced community service orders as an intermediate sanction (see the chapter of Dünkel and Lappi-Seppälä on community service in this volume). The organization and supervision of community service belongs to the tasks of the probation services (as far as it is not transferred to private organizations of the third sector).

Aiming at changing the offending behavior, most of the probation services recruit probation staff among the graduates of social/behavioral sciences. In some countries, probation staff has to graduate from a social work school (Denmark, Germany, Italy, the Netherlands, Norway, Scotland, etc.) or other related fields. In England

and Wales, there is a special Diploma on probation that trains probation officers in a blended model: education and practice.

One major difficulty admitted by a large majority of the European jurisdictions is work overload. Depending on geographical distribution, judicial practice, or the season of the year, the workload varies between 35–50 clients (in Denmark, Estonia, etc.) and 50–100 clients (Switzerland, Luxembourg, etc.). In most cases, the maximum number of cases that one probation worker could supervise is not regulated firmly in any regulation. As it was shown in the literature, work overload has a significant impact on probation effectiveness (see Petersilia, this volume). Another side effect of probation is the one noticed especially in the USA where intensive supervision was introduced. It seems that the more surveillance probation service provides, the more recall into prison it produces (see for instance Solomon et al. 2005).

A Future for Probation?

Looking at the history of probation around the world, one can project a future of it obviously with a high level of generalization and oversimplifications.

As suggested in the first section of this entry, probation services were called initially to "save the souls" of offenders. Increasingly, over the years, probation services started to treat and to supervise sick or unadapted deviants. Probation orders and latter suspended sentences were invented locally and diffused around the world with some local adaptations. What is almost common around Europe is that in the last 20 years, the supervision element of probation became more and more prevalent. Electronic monitoring and extended supervision are only two examples of how surveillance and controlling captured the probation discourse in some Western countries. In other countries, like in the German-speaking or Scandinavian states, probation service is still seen predominantly as a rehabilitation agency. Central and Eastern European countries seem to develop probation services according to their local needs and to what it seems to be fashionable in the West. One example of this mood is the risk assessment tools that were exported massively and sometimes uncritically from the West to the Central and Eastern European countries. Sometimes, this policy transfer comes within the evidence-based rhetoric (see the ▶ RNR Model), but sometimes, it comes with the package in different EU programs. There are reasons to believe that penal policy transfer will intensify on a short and medium run.

Apart from the national or regional diversity drivers, there are also more and more forces acting toward Europeanization. The European Council adopted in 2008 a framework decision (2008/947/JHA) aiming at facilitating rehabilitation through transferring supervision among EU countries. The Conference European of Probation (CEP), an intergovernmental association of probation, also fights for a common set of principles and values for probation in Europe. The EU Commission funds many projects that support states to develop probation systems with the help of other states. An Erasmus project is also running, aiming at setting up a European probation curriculum for probation. These are only a few examples of how harmonization forces work toward a European probation system.

Once electronic monitoring was introduced either as a stand-alone measure or as a tool to implement other measures (like house arrest), it was created the premises of involving the private sector into the control industry. In most of the European countries, handling of the electronic monitoring equipment is in the hands of the private sector (see the entry on electronic monitoring of Haverkamp in this volume).

There are countries where electronic monitoring is delivered fully only by the private sector (see England and Wales). Within the context of neoliberalism and facilitated by the budgetary austerity, the private sector becomes more and more visible in delivering probation activities. For instance, Serco, a security company, in partnership with London Probation Trust, has started in October 2012 to carry out "community payback" sentences (or community service or unpaid work, in other jurisdictions). Although this probation privatization is not necessarily a general trend, it might be an indication of how probation and probation activities could look like in the future.

Fortunately, as it was illustrated above, developments in the probation world are, sometimes in spite of the harmonization drivers, incremental and local. It took decades from the moment suspended sentence was introduced in France and Belgium till the moment of upgrading it to the suspended sentence under supervision in almost all over Europe. Electronic monitoring was also introduced after several years of piloting. The time between the introduction and testing of a new penal technology until this technology is fully implemented is shorter and shorter, but still one cannot hope for revolutionary changes in the probation world. Every change will be introduced slowly, with a lot of resistance from many levels and in the end the core idea of probation – to rehabilitate offenders – will probably rise again and again in different shapes and forms.

Related Entries

- ▶ Community Service in Europe
- ▶ Early Release from Prison
- ▶ Imprisonment
- ▶ Law, Diversion and Community Sanctions in Juvenile Justice
- ▶ Recall

Recommended Reading and References

Bagaric M (1999) Suspended sentences and preventive sentence: illusory evils and disproportionate punishments. UNSW Law J 22(2):535–563

Bottoms A, Gesthorpe L, Sue R (2001) Community penalties: change and challenges. Willan, Cullompton

Burrell WD (2010) Probation in United States. In: Herzog-Evans M (ed) Transnational criminology manual, vol 3. Wolf Legal Publishers, Nijmigen

Da Agra C (2010) La probation et ses contextes socio-historiques. http://www.cepprobation.org/uploaded_files/Pres%20GA%2010%20Agr.pdf

Dünkel F, Spiess G (1983) Alternativen zur Freiheitsstrafe – Strafaussetzung zur Bewährung und Bewährungshilfe im internationalen Vergleich. Freiburg, MPI

Durnescu I (2012) What matters most in probation supervision: staff characteristics, staff skills or programme? Crim Justice Criminol 12(2):193–216

Durnescu I (2009) Community sanctions and measures in Europe. "Some results from a recent study" in EuroVista. Probation and Community Justice 1(1): 10–19.

Jäger B (2010) Die sozialpädagogische Betreuung von straffälligen Menschen in der Bewährungshilfe. Ein Ländervergleich. Universitätsverlag Brockmeyer, Bochum

Jarvis F (1972) Advise, assist and befriend: a history of the probation and after-care service. National Association of Probation Officers, London

Normandeau A (1979) Faut-il transformer ou abolir la probation ? Notes de lecture. Criminologie 12(2): 89–105

Schöch H (2003) Bewährungshilfe und humane Strafrechtspflege. Bewährungshilfe 50(3):211–225

Solomon AL, Kachnowski V, Bhati A (2005) Does parole work? Analyzing the impact of postprison supervision on rearrest outcomes. Urban Institute, Washington, DC

United Nations (1951) Probation and related measures. United Nations, New York

van Kalmthout AM, Durnescu I (2008) Probation in Europe. Wolf Legal Publishers, Nijmegen

Vanstone M (2008) The international origins and initial development of probation. An early example of policy transfer. Br J Criminol 48(6):735–755

Probation and Human Rights

Rob Canton
Community and Criminal Justice, Health and
Life Sciences, De Montfort University,
Leicester, UK

Overview

Human rights are those rights that all people have in virtue of our common humanity. They include *liberties* and *claims*, being used in political debate to remind governments both of the limits of their powers over their citizens (liberties) and of their obligations to create circumstances in which people can thrive and prosper (claims). International treaties and conventions have been established to clarify rights, to set standards, and to attempt to make all countries accountable to the international community. Human rights are especially important in criminal justice and punishment, where the coercive powers of the state are so manifest. Punishment may even be defined as a deprivation or suspension of rights. Yet hard questions arise about which rights are forfeit and which should be protected. The rights of serving prisoners are especially vulnerable, but there must also be concern for those subject to community sanctions and measures. At the same time, the rights of other people, in particular actual and potential victims, must also be part of an adequate ethical account. Principled ways must be found of dealing with circumstances where rights are in conflict, although such conflicts may be less common than political debate seems to assume. Some contemporary developments have particular implications for human rights – notably the policy in many countries to make community punishments as "prison-like" as possible in a quest for punitive credibility, the political predominance of risk, and preventive interventions. Such trends must be scrutinized with special care because of their potential not only to deprive offenders of their rights but to trespass on the rights of other members of the community without adequate justification. While debate often centers on offenders' liberties, their claim-rights must also be affirmed. The discourse of human rights offers a way of framing guiding ethical principles for probation and could constitute a foundation for policy and practice.

What Are Human Rights?

Human rights are those rights that we all have by virtue of our common humanity. Since human rights are intimately bound to humanity and personhood (Griffin 2008), to fail to give weight to someone's rights is to risk treating them as less than human. Earlier philosophers sometimes referred to *natural rights* – a conception that was found by critics (e.g., Jeremy Bentham) to be obscure and perhaps incoherent. Even so, the idea that people have rights that are not granted by the state – and which consequently may not be taken away by the state – implies the critical possibility of an ethical position from which the laws and practices of government can be appraised, circumscribed, and directed. In these ways, the language of rights can be deployed to remind governments of their limits and of their obligations, that ends may not be assumed to justify means, that individuals and minorities may not simply be disregarded in a proclaimed pursuit of the state's conception of the general welfare, and that it is a duty of government to strive to create circumstances in which all may flourish.

Like most serious moral questions, the matter of which rights we have remains open, indeterminate, and inherently contested, and the rights that people have asserted for themselves and for others are many and diverse. Historically and politically, rights have been invoked in two main ways: to take a stand against cruelty and injustice and to articulate claims – aspirational statements about the social and economic conditions that should be created so that people may thrive (Gearty 2006). Most notably, for example, the United Nations Universal Declaration of Human Rights (UDHR) (1948) first sets out articles prohibiting oppression and cruelty, then articulates claims to human well-being.

This corresponds to a philosophical distinction between *liberties* and *claims*. Liberties include the right to life, a prohibition on torture, on inhuman and degrading treatment, on cruel or unusual punishment, bans on slavery, forced labor, and servitude. Claims express a commitment to human flourishing, to propitious circumstances and meaningful opportunities – for example, rights to education, employment, and participation in government – and to an adequate standard of living. These are claims that require more than the state's forbearance and call for positive action by government.

By no means all rights are absolute – probably very few are. The European Convention on Human Rights, for example, distinguishes absolute rights (which may never be taken away), limited rights (which may only be compromised in explicitly identified, specific circumstances), and qualified rights (where individual rights must be considered alongside broader social and community interests). To assert a right, then, even convincingly, does not by itself entail that it would be unjustifiable to infringe that right, though it does mean that there is a need for moral judgement. But a human right may not be ignored or merely set aside as inconvenient: if it must be infringed (because other moral considerations turn out to be more compelling), then this calls for moral justification and should take place with a sense of regret and with an attempt to mitigate the implications for the right-bearer. This should also stimulate a search for other solutions which better respect the right.

International Perspectives

The modern promulgation of human rights has been accomplished through international charters and conventions. The best known of these – and the foundation of most others – is the United Nations Universal Declaration of Human Rights (UDHR). In its Preamble, the Declaration begins with a clear statement that the "recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world." Drawn up as it was in the aftermath of the Second World War, it goes on to assert that "disregard and contempt for human rights have resulted in barbarous acts which have outraged the conscience of mankind."

The International Covenant on Civil and Political Rights (ICCPR), monitored by the United Nations Human Rights Committee, developed the principles of UDHR and, in particular, delimits the powers of states with regard to the rights of liberty and security of persons (article 9). It also affirms rights to justice and to a fair trial (article 14). Like the UDHR, the ICCPR is signed and ratified by countries from all parts of the world. Over time, international regions have drawn up their own conventions, notably the African Charter on Human and Peoples' Rights, the American Convention on Human Rights, and the European Convention for the Protection of Human Rights and Fundamental Freedoms. These seek to apply UDHR and ICCPR principles to the specific characteristics of each region, taking account of differences of culture and tradition. The charters and conventions are guarded by international courts – the European Court of Human Rights and the Inter-American Court of Human Rights (a protocol establishing the African Court on Human and Peoples' Rights is in force, but this Court, to be merged with the African Court of Justice, has yet to be established). Courts adjudicate, hearing and ruling on allegations of violations referred to them, but may also advise: the Inter-American Court of Human Rights, for example, advises on points of legal interpretation referred to it by the member states.

Signatories to these conventions have sometimes incorporated the same articles, or at least their principles, into their domestic law. There are still many reasons why an international framework remains essential. In particular, where there are conflicts between human rights and a government's conception of the common good, it is not clear that the state is an even-handed judge in its own cause. The protection of the law is accordingly required and accountability to an international community is a crucial bulwark.

Punishment and Human Rights

Human rights are an especially important consideration for criminal justice and punishment, where the coercive powers of the state are always near to the surface and often manifest. The censures and "hard treatment" of punishment necessarily involve at least a degree of deprivation of some rights. Rights that states are in other circumstances expected to protect vigilantly are taken away or compromised – not only contingently but deliberately, as punishment. But which rights are forfeit and which retained? How is punishment to be determined and administered in a principled way that respects the offender's humanity?

Prison is a total institution and the conditions of imprisonment impact on all aspects of prisoners' lives. The complete dependence and vulnerability of prisoners make their rights especially precarious and much of the literature on human rights and punishment accordingly focuses on imprisonment (for summary and discussion, see van Zyl Smit 2007). Yet developments in a number of countries over the past 40 years or so have started to raise hard questions about the human rights implications of probation and parole as well.

Parole, Probation, and Intermediate Sanctions

To guard against suspicions that early release constitutes an inappropriate reduction in the length of the prison term, parole is sometimes characterized as the continuation of a sentence, though now in the community. The emphasis of resettlement/reentry interventions is accordingly on tight supervision requirements, risk management, and strict enforcement. If parolees are seen as serving prisoners, the disciplines of imprisonment can legitimately be invoked and many of the rights enjoyed by other citizens may be withheld. In some parts of the United States, for example, parolees are not entitled to vote. The voting entitlement of convicted offenders varies considerably between countries, and indeed among the states of the USA, representing different views about the standing and rights of convicted offenders.

As well as post-release measures, changes in the character of community sanctions have raised comparable questions. Traditionally, probation was seen as a measure of support, assistance, and rehabilitation: in England and Wales, for most of its history, probation was *instead of* a sentence – not a punishment in its own right. But since the late 1980s, community sanctions have often been presented as *punishment in the community*, with the restrictive and punitive aspects of these sanctions deliberately emphasized. The idea is that tough community sanctions may reduce the prison population by persuading courts and the public that these are credible alternatives to custody. This aspiration, however, has a disappointing history: it is hard to find times or places when the introduction or toughening of community sanctions can be shown to have made significant reductions in the numbers of people in prison.

Similarly, the USA has seen a range of *intermediate sanctions* – sanctions, that is, intermediate between prison and community punishment. Michael Tonry (1998) attributed this development to three factors: the loss of confidence in the effectiveness of rehabilitation; the rise of a penal philosophy of *just deserts*, under which all court disposals had to be seen to be punitive in proportion to the offense(s); and the increased politicization of crime and punishment that led politicians to be very wary of an accusation that they might be *soft on crime*. These influences combined to lead to the emergence of a range of sanctions and measures that make clear demands on people's time and (in the case of community service/unpaid work) their labor and can constrain their freedoms – for example, by requiring them to be at (or stay away from) particular places at certain times, to participate in programs, and to submit to tests for drug or alcohol use.

It should be emphasized that the treatment philosophy of an earlier time had quite as much potential to trespass upon individual rights. One of the reasons why America and other countries, partly under US influence, repudiated the treatment model of corrections was because of its disregard for due process and human rights constraints: interventions that were ostensibly designed to offer treatment (and could therefore be argued to be in the offender's best interests) could countenance intensive and disproportionate interventions. While treatment has been disavowed in many countries, and even though *just deserts* should set limits, attempts to demonstrate punitive credibility have warranted increasing levels of intervention.

In short, the more "prison-like" community supervision aspires to be, the more likely it is to impinge upon offenders' human rights. Liberal criminology's opposition to imprisonment must remain principled, however: human rights matter here too and it is not true that anything goes so long as it is not prison.

Specific Issues for Community Punishment

Trends in the *enforcement* of community supervision and parole licenses are a matter for particular disquiet. In some countries, increasing numbers of people are in prison not because of the substantial harms they have been responsible for, nor even on the basis of assessed risk, but because of their failure to comply with the requirements of supervision (Padfield and Maruna 2006). This has implications for proportionality in sentencing and could also raise concerns about the processes of recall and the grounds on which such decisions are taken.

Another example is *disclosure*. There are complex questions about the proper boundaries between disclosures of personal information about offenders who are believed to pose a risk to the safety of potential victims and the rights of such (ex-)offenders to privacy and safety from harassment. Provisions like Megan's law (laws in the United States requiring law enforcement authorities to notify the public about registered sex offenders) raise these questions starkly. It can also be argued plausibly that exposing these offenders to opprobrium and perhaps to vigilante reprisal will lead to some very dangerous people "going underground," concealing their whereabouts from those authorities who need to monitor them. It is increasingly recognized that offenders' cooperation and compliance depends on *legitimacy*, on their perception that they are being dealt with fairly. Due regard for their human rights is one indispensable component of this legitimacy.

While some infringements of rights are deliberate and intrinsic to the lawfully determined penalty, punishment normally has other consequences besides. Walker (1991) uses the term *incidental* punishments to refer to those consequences that are not intrinsic to the punishment, but side effects, as it were. For example, intensive community supervision (perhaps a requirement to report daily to probation) could interfere with the legitimate pursuit of employment – and in the process delay desistance from offending. This is not an intentional result of the lawfully determined punishment. Where rights are avoidably (even if not deliberately) infringed, it is arguable that these infringements should be recognized, minimized, and maybe compensated.

One immediate difficulty here, however, is that punishment, notoriously, has many purposes, including a just imposition for the wrong that has been done, the reduction of further offending (by rehabilitation, deterrence, or incapacitation) and reparation. In many countries, these purposes are often not defined, so that it is hard to tell the loss of rights that is intended to be part of the punishment and which is "incidental." It might be possible to argue that a right is justifiably removed if, but only if, its removal is a necessary consequence of the sentence and contributes to a legitimate purpose of punishment. This is probably an incomplete safeguard: claims about public protection and an exaggerated confidence in the efficacy of deterrence could continue to countenance punitive excesses. Meanwhile the test of proportionality – that punishment must be in proportion to the crime – is probably the single most secure safeguard against undue infringement of the rights of offenders (van Zyl Smit and Ashworth 2004).

Walker (1991) also notes that punishment commonly has effects that fall upon other people – for example, offenders' children, whose own rights to a family life and general well-being may well be prejudiced by the imprisonment of a parent. Stigma, surveillance and disclosure are likely to have their impact not only upon offenders but upon their families and associates as well. Their human rights too must be respected and not merely seen as part of the punishment of the offender.

Prevention, Risk, and Human Rights

Several scholars have drawn attention to and debated the significance of a shift in many countries from the traditional criminal justice tasks of the just punishment and rehabilitation of offenders towards an emphasis on crime prevention (e.g., Garland 2001). The identification, assessment, and management of risk have become a uniting objective for the several agencies of criminal justice and a priority for each of them. In the penal system, this has led to a range of preventive sanctions and measures that are less a response to what an individual has done than a preemptive intervention in anticipation of what it is believed that s/he might do. This same trajectory has led to modifications and compromises to the safeguards of due process (see, e.g., Ashworth and Zedner 2008 and references there cited).

In practice preemptive interventions in the penal system are taken mostly against those with a (typically significant) previous record of offending, so that it is not always easy to distinguish what is retributive (due punishment for wrongs done) and what preemptive. But plainly these developments raise hard questions about proportionality. Proportion used to be understood to relate to an appropriate fit between the crime and the sanction. There was a guiding principle of *limiting retributivism* such that the punishment may never exceed the seriousness of the crime (although in some circumstances it may be reduced because of mercy or mitigation). But this is now significantly complicated by the emphasis on prevention. One argument is that proportionality may be exceeded to prevent crimes – perhaps even the term *proportionality* should be understood to mean proportionate not only to wrong done but to wrong anticipated. If the term is extended in this way, however, it becomes a much less effective constraint upon excesses of punishment.

Probation is centrally implicated in these debates. An influential probation model, arguably dominant in most English-speaking countries, focuses on risks, needs, and responsivity (RNR) in assessment and intervention. The risks in question are mainly risks of reoffending; the relevant needs are known as criminogenic needs (those associated with offending); responsivity is the principle that interventions with offenders are to be delivered in a manner that engages them, encourages full participation, and takes account of their individuality. Among criminogenic needs are problems like homelessness, unemployment, substance misuse, and poverty. Since these are amenable to change, they have been sometimes redefined as *dynamic risks*, with the result that social exclusion and disadvantage could come to be seen not as reasons to provide support and fair access to resources but as an occasion for more intrusive intervention. Where RNR prevails, risk assessment and management are at the center of probation practice. Assessed risk is now most likely to determine the nature and intensity of intervention, which again has the potential to challenge the limits of proportionality.

In targeting potential offenders, preventive justice (Steiker 1998) goes beyond the penal system and has largely developed without the due process safeguards of criminal justice or legal regulation (Janus 2010). Since the category of potential offenders could be limitless, preventive strategies make a difference to everyone. Ways of thinking about crimes and victims and technologies of crime prevention pervade civil institutions – families, schools, and the workplace – to an extent where states may be said to be *governing through crime* (Simon 2007). Modern technologies offer unprecedented opportunities of surveillance, monitoring, and data collection and collation, and the "commodification" of security adds a commercial impetus to these developments. Crime prevention thus lends legitimacy to actions that would otherwise be seen as an unwarrantable trespass on people's human rights.

When Human Rights Conflict

Proponents of preventive justice commonly defend their position by setting the rights of actual and potential victims and/or "community safety" in rhetorical opposition to the rights of offenders. Yet many offender rights – including almost all rights relating to the realities of sentencing implementation – do not in any way infringe the rights of victims or the community. The habit of supposing that the rights of potential or actual offenders must be in tension with the rights of victims or with community safety should be resisted. Such an assumption too readily slides into a belief that any enhancement of offender rights must entail some diminution of the rights of victims. Nor may it be assumed that practices that infringe human rights make for a safer community.

Where there are conflicts of rights, they should be considered carefully and managed in a principled manner. There is no easy formula – like "the rights of victims always outweigh the rights of offenders" – that will enable a decision to be taken summarily: it will depend upon the rights in question and of all those affected (see Ashworth 1996). As we have seen, asserting a right is unlikely to settle a moral problem – and indeed any way of talking about ethics that purports to dissolve substantial moral dilemmas should be viewed with suspicion. Meanwhile it is necessary to guard against the possibility that offender rights are gratuitously disregarded because they turn out to be inconvenient or are deliberately withheld as additional punishment – beyond, that is, the lawful imposition of the court – on an unexamined pretext that this is necessary to protect the community.

Challenges and Difficulties

It has been suggested that some misgivings about human rights – for example, that all rights are absolute, that offender and victim rights are intrinsically in opposition – rest on misunderstandings. Other objections need to be addressed. The first of these is that human rights, at least in penal affairs, are too general or even vague. Ethical principles ought to be action-guiding but can human rights offer sufficient guidance? Unless it is clear what these very general rights really amount to, if they cannot be deployed to guide policy and practice, they are at risk of becoming not much more than rhetorical claims.

Perhaps most importantly, court judgements establish a body of precedent that gives authoritative clarification to the articles of human rights conventions. But since this is a reactive (and slow) process, there have been attempts to develop instruments that apply the articles and principles of the conventions to the specific context of criminal justice and punishment (for discussion of the development of the international regulation of punishment in these ways, see van Zyl Smit forthcoming). The ICCPR takes significant steps in this direction. Again, the UN Standard Minimum Rules for Non-custodial Measures (*The Tokyo Rules*) seek to specify the implications of the UDHR for the position of defendants and offenders, while the European Prison Rules apply the European Convention to the particular circumstances of prisoners. European Probation Rules have also now been adopted (Canton 2010). In principle, such rules *draw out* what it will mean in practice to respect the rights of offenders, victims, and the community – and how to resolve conflicts of rights in a principled way. Rules thus contribute to a solution to this challenge of generality and remoteness.

A second objection is that rights are minimal: they are what is left when all else fails and are therefore of limited assistance in guiding the strategies and practices of a civilized and decent criminal justice system (see Nellis and Gelsthorpe 2003). Here, however, the legal principle of *positive obligations* becomes salient: both victim and offender rights can call for positive action, not merely forbearance – claims as well as liberties. This idea has the potential to challenge and to stretch the state's duties. For example, in *E. and others versus United Kingdom*, the court held that the UK had violated Article 3 (prohibition of inhuman and degrading treatment) by failing to accord E. adequate protection from abuse at the hands of her mother's cohabitant (Hofstotter 2004). In this way, a basic liberty-right may progressive evolve into something much stronger.

This is a principle with clear and significant implications for – in particular – the rights of victims of crime (as in the case of E.), but the principle should apply to offenders as well. Many offenders have experienced substantial social and economic disadvantage, and while their crimes entail a liability to punishment, their rights to fair opportunities should also be affirmed. Penal intervention should leave people better fitted to live well and without offending, and this could be conceptualized as a right with a corresponding duty that the state should assume. Once the sentence has been fulfilled, the state should accept a responsibility to create opportunities that the punishment may have curtailed – for example, an ex-prisoner may have a right to support to seek employment and accommodation. Rehabilitation through resettlement/reentry/aftercare could be seen as a right – not just a continuation of a punishment (Rotman 1994; Lewis 2005).

This reaffirms one of probation's traditional missions of advocacy on behalf of offenders, encouraging the community to accept its responsibilities towards those of its members who have committed crimes and not only to insist on their putative claims against them. The role of interventions, programs, and services can be invaluable in helping people to learn how to manage the problems that beset them without recourse to crime, to develop the skills they need to thrive. Yet desistance studies show that motivation and skills may not be sufficient: people need genuine opportunities to develop and to lead lives in which offending has no place. Reintegration, after all, implies not only a motivated individual but also a community willing to believe in the possibility of change and to respect the legitimate interests of ex-offenders. As it is, many offenders continue to experience significant stigma and disadvantage long after the completion of the lawfully imposed punishment. At the least, probation should advocate for ex-offenders to have fair opportunity and access to services enjoyed by others in the community.

Conceptualizing rehabilitation as a right, then, lends ethical support to the findings of desistance research, which emphasizes the importance of social capital and social inclusion in giving people genuine opportunities to thrive. The origins of offending are bound up in the social order and responses to it cannot just rely on the mechanisms of criminal justice. Probation has a role here – in advocacy, in the negotiation of access to inclusive services – though the success of its work here will depend on wider socioeconomic factors.

A further objection to be considered is the accusation that the discourse of human rights is culture-bound: an attempt by the West to universalize their own ethical standards and impose them on others. Yet, while recognizing the force of this allegation, Clapham replies that

> Human rights were invoked and claimed in the context of anti-colonialism, anti-imperialism, anti-slavery, anti-apartheid, anti-racism, and feminist and indigenous struggles everywhere. Western governments may recently have dominated the discourse at the highest international levels, but the chanting on the ground did not necessarily take its cue from them, nor did it sing to the West's tune. (2007, p. 19)

To assert rights that belong to everyone, rights that do not have to be "deserved" and are not contingent on gender, race, age, ability, and socioeconomic status, sounds like a promising beginning in challenging oppression.

Human Rights and the Values of Probation

It has already been remarked that human rights have been criticized as too remote and general to guide action. Yet a conception of human rights based on principles of personhood (Griffin 2008) – autonomy, choice, the ability, and liberty to act – could generate some specific guiding values for probation. The list below is an attempt to specify some such principles which could be derived from general rights as they apply to the work of probation. These could be considered as ethical parameters within which probation should be developed or perhaps as a benchmark for current policy and practice:

1. Believing in the possibility of personal change, probation should help to create opportunities for people to lead fulfilling and crime-free lives, enable them to acquire the skills needed to avail themselves of those opportunities, and support them in their motivations to change.
2. Recognizing people's responsibilities and rights, probation should address offenders as responsible, self-determining people.
3. As far as possible, offenders should be fully involved in (informed and consulted about)

any decisions that affect them. They should be actively involved in assessment, planning, interventions, and evaluation.

4. Probation seeks to gain the consent of offenders. Even when probation staff are obliged to take action against someone's wishes, they should always try to secure the offender's understanding of and, so far as possible, consent to any decisions that affect them.
5. The purpose of intervention is to reduce the likelihood of reoffending. Interventions must accordingly be constructive and not punitive in character.
6. Probation should countenance no imposition or limitation of rights unless this can be justified by some legitimate penal purpose.
7. Probation must always have regard to the rights and interests of offenders, even when the rights and interests of others must take priority.
8. Some limitation of rights is a necessary consequence of punishment but any such limitation of rights should be *either* regarded as part of the lawful sentence *or* minimized. Any restrictions on offenders' rights must be proportionate to the offense.
9. Offenders are members of the community some of whose rights have, for the duration of the penalty, been taken away or compromised. Once the penalty has been served, these rights should be fully reinstated (indeed, this is one meaning of the term *rehabilitation* – McWilliams and Pease 1990). In this respect, communities have responsibilities towards those of their members who have committed crimes.
10. Social inclusion is a requirement of social justice and a key guiding principle in probation practice. Since people who do not have fair and reasonable access to the services and institutions of civil society (social exclusion) are more likely to offend, this principle also helps to reduce offending.
11. To achieve the social inclusion of offenders, probation must work in close partnership with the agencies of civil society. Meeting the complex needs of many offenders calls for coordinated and complementary interagency work. Probation should use its expertise to help other agencies to make their services genuinely accessible to offenders. These principles attempt to connect the work of probation to the wider society. They go beyond organizational claims about the advantages of working in partnership to say that the services and institutions of civil society should be genuinely available to offenders as a matter of social justice and their legitimate claim-rights.
12. All probation practices must respect the interests and rights of victims of crime, in work with them or on their behalf, as well as working with offenders to help them to recognize the harm that they have caused.
13. Probation should explain its work and its significance to the public. Its values, policies, and practices must be open and must command the confidence and trust of the community. This is a requirement of legitimacy.
14. Probation policy and practice should be evidence-led. Rigorous research and evaluation should guide the work of probation agencies. If probation's objectives are worth achieving, the extent to which they are indeed being achieved must be appraised.

These principles may be in conflict with other ethical considerations. For example, while interventions should be constructive and not punitive, there are interventions – for instance, constraints on some offenders who seek to abuse or exploit the vulnerable – that will indeed be experienced as punitive and restrictive but are necessary to safeguard the rights of others. Similarly, while completion of the lawful sanction ought normally to restore people to their full rights, the continuing risk posed by some people means that their rights may have to continue to be circumscribed. These principles, then, are best understood as morally relevant considerations that ought always to be marked and allowed for when decisions are to be taken, even where they must give way to other moral considerations.

Conclusions

Probation research and policy have been dominated by the question of *what works*. Effective practice could indeed be seen as a right, shared by offenders (who should be enabled to participate in interventions that will help them to lead law-abiding lives) and by actual and potential victims, to the extent that further offending is thus reduced. A focus on human rights, however, is a reminder that probation should construct an ethical foundation for its policies and practices.

Probation and criminal justice in general give expression to a society's values and are not reducible to their instrumental functions. It has been argued that a respect for people's rights brings a legitimacy to practice that enhances individual offenders' cooperation and compliance. A repressive approach to risk management can lead to strategies that tend to increase risk through stigmatizing and exclusionary approaches. This can endanger a community by undermining the trust and reciprocity on which community safety and well-being ultimately depend. In this way, it may be that a rights-based approach to probation practice could turn out to be its most effective foundation.

Related Entries

- ▶ Forced Migration and Human Rights
- ▶ Human Rights Violations in Criminal Court
- ▶ Victims' Rights in the Criminal Justice System

Recommended Reading and References

Ashworth A (1996) Crime, community and creeping consequentialism. Crim Law Rev 43:220–230

Ashworth A, Zedner L (2008) Defending the criminal law: reflections on the changing character of crime, procedure, and sanctions. Crim Law Philos 2:21–51

Canton R (2009) Nonsense upon stilts? Human rights, the ethics of punishment and the values of probation. Brit J Community Justice 7(1):5–22

Canton R (2010) European probation rules: what they are, why they matter. EuroVista 1(2):62–71

Cavadino M, Dignan J (2007) The penal system: an introduction, 4th edn. Sage, London Chap. 2, 5 and 11

Clapham A (2007) Human rights. Oxford University Press, Oxford

Duff RA (2001) Punishment, communication, and community. Oxford University Press, New York

Faulkner D (2006) Crime, state and citizen, 2nd edn. Waterside, Winchester

Foucault M (1977) Discipline and punish: the birth of the prison. Penguin, Harmondsworth

Garland D (2001) The culture of control: crime and social order in contemporary society. Oxford University Press, Oxford

Gearty C (2006) Can human rights survive? Cambridge University Press, Cambridge

Griffin J (2008) On human rights. Oxford University Press, Oxford

Hofstotter B (2004) European court of human rights: positive obligations in E. And others v. United Kingdom. Int J Constitutional Law 2(3):525–560

Hudson B (2001) Human rights, public safety and the probation service: defending justice in the risk society. Howard J 40(2):103–113

Janus E (2010) The preventive state: when is prevention of harm harmful? In: Nash M, Williams A (eds) Handbook of public protection. Willan, Cullompton

Lewis S (2005) Rehabilitation: headline or footnote in the new penal policy? Probat J 52(2):119–135

McWilliams W, Pease K (1990) Probation practice and an end to punishment. Howard J 29(1):14–24

Nellis M, Gelsthorpe L (2003) Human rights and the probation values debate. In: Chui WH, Nellis M (eds) Moving probation forward: evidence, arguments and practice. Pearson, Harlow

Padfield N, Maruna S (2006) The revolving door at the prison gate: exploring the dramatic increase in recalls to prison. Criminol Crim Justice 6(3): 329–352

Rawls J (1967) Two concepts of rules. In: Foot P (ed) Theories of ethics. Oxford University Press, Oxford

Rotman E (1994) Beyond punishment. In: Duff A, Garland D (eds) Reader on punishment. Oxford University Press, Oxford

Simon J (2007) Governing through crime. Oxford University Press, New York

Steiker C (1998) The limits of the preventive state. J Crim Law Criminol 88(3):771–808

Tonry M (1998) Intermediate sanctions. In: Tonry M (ed) The handbook of crime and punishment. Oxford University Press, New York

van Zyl Smit D (2007) Prisoners' rights. In: Jewkes Y (ed) Handbook on prisons. Willan, Cullompton

van Zyl Smit D (forthcoming) Punishment and human rights. In: Simon J, Sparks R (eds) Handbook of punishment and society. London, Sage

Walker N (1991) Why punish? Oxford University Press, Oxford

Ward T (2009) Dignity and human rights in correctional practice. Eur Probat J 1(2):112–127

Probation and Parole Practices

Melissa Alexander[1], Christopher T. Lowenkamp[2] and Charles Robinson[2]
[1]US District Court, North Carolina Middle Probation Office, USA
[2]Administrative Office of the U.S. Courts, Washington, DC, USA

Overview

Community supervision involves the monitoring of offenders living in the community, either following an institutional sentence or in lieu of incarceration. Historically offenders have been supervised via monitoring of "conditions" they must meet (i.e., obtaining or maintaining employment, reporting any changes in address/employment, abstaining from drugs and/or alcohol) and perhaps referral to community agencies for substance abuse treatment or other types of treatment. Research into the effectiveness of these strategies has been limited but for the most part suggests little to no impact on recidivism rates. At the same time, emerging research has suggested a new model of supervision (Risk-Need-Responsivity), which focuses on risk of recidivism, offender needs directly related to recidivism, and individual offender issues. Importantly, part of the changing focus has been on the specific officer-offender interaction and how officers might incorporate cognitive-behavioral and social learning principles into their work. Additionally, research has focused on incorporating the principles of motivational interviewing (MI) in criminal justice supervision. The potential impact of this supervision model has been promising, with initial studies suggesting 25-50 % relative reductions in recidivism for some strategies. However, the implementation of these strategies has been challenging, both from an organizational standpoint and the specific training of individual officers. As these are skill-based strategies, it is likely that significant training, coaching, and follow-up are needed in order for officers to become competent in the skills. A number of additional issues have yet to be researched, including which offenders might benefit most from the strategies, how strategies are best learned/incorporated into practice, and optimal levels of use of the various strategies.

Introduction

Community supervision is a broad term that has been used to capture work done by probation and parole officers within a community setting. Offenders on probation have generally served little to no time in an institution, instead being allowed to remain in the community. In contrast, offenders on parole have received institutional sentences but are released early and allowed to complete their sentences in the community. In both instances, offenders are supervised by an officer and are required to abide by certain "conditions" such as obtaining or maintaining employment, reporting any changes in address/employment, abstaining from drugs and/or alcohol, and not committing additional crimes. Offenders may be supervised for varying lengths of time (several months to several years or more) and may be required to report to an officer with varying frequency (i.e., weekly, monthly, or less). Traditionally, the primary mission of community supervision officers has been to monitor the behavior of offenders under their watch, to ensure they abide by any conditions imposed and remain law abiding. If violations occur, supervision may be revoked and the offender sent to an institution.

The effectiveness of community supervision in reducing recidivism has been called into question, with numerous studies suggesting that traditional approaches to supervision have little or no impact on recidivism rates (such as Bonta et al. 2008). At the same time, criminal justice research has focused on learning "what works" with offenders and has distilled this knowledge into what has become known as the Risk-Need-Responsivity (RNR) model (see most recently Andrews and Dowden 2007). Risk involves focusing on those offenders most likely to recidivate, Need incorporates identifying the specific factors related to reoffending, and Responsivity involves delivering treatment and interventions in an individualistic way. A large body of research has indicated that implementation of the RNR model can significantly impact recidivism rates (see Andrews and Bonta 2010 for a review of this research and Lowenkamp et al. 2010 for a review of its application to supervision based programs). A recent meta-analytic review (Andrews and Bonta 2010) demonstrated that adherence to all three principles leads to the greatest reductions in recidivism, while nonadherence with these principles leads to increases in recidivism. Unfortunately, analysis of conversations between officers and offenders demonstrated there was little adherence to the RNR model (Bonta et al. 2008).

The Importance of the Officer-Offender Interaction

To date most research and meta-analyses have focused on the translation of Risk-Need-Responsivity (RNR) into formal treatment programs, with little emphasis on how the model should be implemented within the officer-offender interaction itself, nor what occurs during the actual contact between an officer and offender. In fact, research on officer-offender interactions has suggested that officers focus on monitoring versus changing behavior, and training of officers has typically focused on understanding local laws, policies, and procedures. However, beginning in the early 1990s, more emphasis has been placed on the content of officer-offender interactions, beginning with a focus on risk assessment. Since that time, there has been a move from the traditional monitoring role of community supervision agents to a change-oriented role. The shift has been informed by social learning theory and cognitive-behavioral research on effective ways to engage people in behavior change, as well as research on the nature of helping relationships. Social learning theory emphasizes that people learn not only from their own experiences but also by observing the behavior and resulting consequences of other people's behavior, whether it be actual behavior seen, a description of behavior/consequences, or what they see in television, movies, etc. Additionally, social learning theory notes that the internal rewards someone may experience (pride, satisfaction, etc.) may be just as important, or even more important, than rewards received from others. This focus on the internal thought processes coincides with cognitive-behavioral theory, which emphasizes that our behavior is a result of our thoughts and feelings about an event, rather than the external event itself. Under this theory, behavior change occurs when someone focuses on their internal analysis of an external event (what I thought, how I felt) and the active changing of those thoughts and resultant feelings. Finally, such behavior change must occur within the context of an empathic helping relationship. The translation of these theories into the criminal justice arena has resulted in what has become known as Core Correctional Practice (CCP) (e.g., Andrews and Dowden 2004). These practices include areas such as focusing on high-risk situations (likely to lead to criminal behavior) and analysis of cognitions that support criminal behavior. For example, trained officers frequently respond to positive drug screens by examining the antisocial thoughts of the offender ("I can get away with it," "I won't be tested soon," or "I'm not hurting anybody") and working collaboratively to identify thoughts that reduce the likelihood of the behavior in future high-risk situations. Likewise, the CCP-trained officer looks to identify cognitive responses to prosocial behavior. Trained officers reinforce prosocial behavior by moving beyond simplistic verbal praise (i.e., "good job"), instead asking the offender to identify immediate and long-term internal and external rewards. Identifying the cognitions ("my kids are going to be proud of me," "if I continue to do this I can build toward what I really want") that support the prosocial behavior helps develop meaningful thoughts that can be used in future high-risk situations. The CCP-trained officer also focuses on teaching offenders skills using a structured and graduated learning process complete with rehearsal and practice. Finally, in addition to the special attention paid to the thinking of the offender, the CCP-trained officer deliberately builds a working alliance that focuses on the acquisition and development of skills that reduce risk in potentially risky situations. Utilizing these types of interventions can fundamentally change the work of probation officers. The traditional case management approach, whereby the officer "manages" the case through monitoring and referral, is transformed to one where the officer is now a "change agent" (Bourgon et al. 2011). In this new role, officers actively work to facilitate the change process through their interactions with offenders, rather than relying on interventions by other professionals.

Research on the impact of these behaviorally based interactions is promising. Trotter (1996) found that using empathy, problem-solving, and a prosocial approach reduced recidivism. Taxman (2004) outlined a process of proactive community supervision (PCS) which focused on comprehensive assessment, case planning, setting clear expectations, reinforcement, desistance, and the use of behavioral contracts. Recidivism rates of offenders supervised by those trained in this model were about 30 % (relative risk) lower than offenders supervised by officers not trained in PCS. Bonta (2010) utilized the Strategic Training Initiative in Community Supervision (STICS) curriculum to teach and evaluate CCP by categorizing various officer actions/techniques into four areas: structuring skills includes items such as "checking in" with the offender, reviewing homework, and prioritizing needs; relationship building includes skills such as role clarification, active listening, and effective feedback; behavioral techniques focus on modeling, the effective use of reinforcement and disapproval, problem solving, and self-management skills; and cognitive techniques, including specific targeting of procriminal attitudes, the application of a behavioral model, and cognitive restructuring. Officer training in these skills included a 3-day training followed by monthly maintenance sessions which allowed officers to discuss the use of skills, receive feedback on their use of skills, and provide "refresher" training. Officers trained in STICS used the skills more often, and their offenders had lower recidivism rates, than those officers who did not receive training. Subsequent studies using STICS have found similar increases in officers' focus on procriminal attitudes, the use of cognitive interventions, and subsequent lower recidivism rates. A similar curriculum, Staff Training Aimed at Reducing Rearrest (STARR), has also been developed. The STARR skills include specific strategies for active listening, role clarification, effective use of authority, effective disapproval, effective reinforcement, effective punishment, problem solving, and teaching, applying, and reviewing the cognitive model. For each strategy, skill cards outline the specific activities officers need to do to successfully deliver the strategy. A fundamental focus for each skill is the internalization of strategies so that offenders begin to learn and apply the strategies on their own. Initial research on STARR demonstrated comparable results to STICS: officers used the skills significantly more than officers not trained in STARR, and offenders had a 50 % relative reduction in recidivism (Robinson et al. 2011). Importantly, in each of these studies, interactions remained relatively brief (around 20 min) yet still impacted outcomes. In total, the research on the use of CCP by community supervision officers has spanned 15 years, three continents, and four very different locations with results consistently yielding a 25-50 % relative risk reduction in recidivism rates.

An additional area of officer training which has received significant attention is motivational interviewing (MI). MI is a communication style that involves strategic use of questions and statements to help clients find their own reasons for change (Miller and Rollnick 2002). MI borrows from client-centered counseling in its emphasis on empathy, optimism, and respect for client choice. MI also draws from self-perception theory, which says that people become more or less interested in change based on how they talk about it. Thus, an offender who talks about the benefits of change is more likely to make that change, whereas an offender who argues and defends the status quo is more likely to continue in the present behavior. Finally, MI is also logically connected to the Stages of Change model, which says that people go through a sequence of stages when considering change (Prochaska et al. 1992). Because MI is a communication style, it is usually introduced as a set of stylistic principles: (1) express empathy, which involves a sincere attempt to understand the offender's point of view; (2) roll with resistance, which emphasizes avoiding arguments whenever possible and finding other ways to respond when challenged; (3) develop discrepancy, which means working to elicit the offender's own reasons for change; and (4) support self-efficacy, which emphasizes positive language and an environment that is supportive of change. From its beginnings in addiction counseling, MI has been translated into a number of behavior-change areas. MI currently has strong research support in areas such as alcohol and drug use, smoking cessation, medication compliance, HIV risk behaviors, and diet/exercise. Two meta-analyses of more than 70 MI outcome studies in different areas suggest an overall significant and clinically relevant effect (Hettema et al. 2005; Rubak et al. 2005).

Many criminal justice agencies have embraced the idea of MI, and numerous articles have been written encouraging its use by community corrections agencies (e.g., Alexander et al. 2008; Walters et al. 2007). However, few studies have focused on the effectiveness of MI within criminal justice settings. McMurran (2009) identified 13 published studies and six dissertation abstracts evaluating the use of MI within correctional settings, with only ten studies using randomized controlled trials. Due to the lack of studies and the significant variation within existing studies, McMurran concluded that "no overall definitive conclusion about the effectiveness of MI with offenders can be drawn." A recent randomized trial of MI within a probation setting indicated no impact on outcome (Walters et al. 2010), though it is noted that the level of MI skill demonstrated by officers may have been below accepted proficiency levels.

Effective Implementation of Core Correctional Practices

Despite the large body of research supporting the RNR model, and the emerging research on officer-offender interactions, the model has not yet become entrenched in the everyday practice of community corrections. The problem of translating research into practice is not something new; other behavioral fields (e.g., medicine, substance abuse, mental health) have experienced similar issues. As such, many have begun focusing on how to effectively implement practices so they do become routine. Klein and Knight (2005) suggest that six key factors will determine if an intervention is effectively implemented. These include (1) the quality and quantity of training; (2) a positive organizational climate; (3) administrative support; (4) financial resources; (5) learning orientation that supports staff skill development, competence, and growth; and (6) patience with the inevitable stumbling blocks that will occur during implementation. We focus on the issue of training, while noting that the other factors are equally important in leading to full-scale adoption of the practice.

Skills Training

Historically, training of officers has focused on brief (1–3 day) trainings with little to no follow-up. Unlike trainings focused on imparting knowledge (i.e., facts about substance use), the interventions noted above all involve the development of a new *skill* to be used. Research on skill development in practitioners has demonstrated that ongoing feedback and coaching is essential if new skills are to be translated into everyday practice. A review of implementation research consistently demonstrates that trainings focused simply on knowledge transfer, with no skill training or follow-up, do not lead to changes in everyday practice (Fixsen et al. 2005). For example, Joyce and Showers (2002) found that only 5 % of participants translated skill use from training to the practice environment if no follow-up was included, while 95 % of participants subsequently used the skills if they receive on-the-job coaching. Research specific to MI suggests that officers may go through as many as eight stages in learning MI (Miller and Moyers 2006). Failure to provide officers the necessary time and support to develop through these stages will likely lead to officers failing to become proficient in the skill. Preliminary research suggests that many officers who only receive workshop training and little or no follow-up/coaching do not meet minimum proficiency levels for MI (Lowenkamp 2010). Similarly, research on training in cognitive-behavioral therapy suggests that the use of follow-up/coaching is essential to increased knowledge and use of the skill (Cully et al. 2010). Importantly, it is noted that several studies of cognitive therapy have shown that outcomes improve as therapist skill level improves (Shafran et al. 2009).

Current Issues

The above research covers a number of different skills that could be used in a multitude of combinations. Consistent with the Responsivity aspect of the RNR model, all skills/interventions may not be needed by all offenders. Further research is needed to determine which skills/interventions may be critical for most, if not all, offenders and which skills may be less important. Additionally, interventions may be effective for some offenders, but not others. Further understanding the nuances in these skills will be critical to moving these practices forward. Similarly, it is unclear whether all officers can, or should, be able to learn all of these skills. Anecdotal evidence suggests that some skills, like MI, may be more difficult for officers to learn than others. Officer orientation, prior training, and experience will all likely influence an officer's interest and ability to learn these skills.

Research in other areas suggests that having a mentor and participating in discussions with others regarding new practices is more likely to lead to integration of new techniques into daily practice (Cook et al. 2009), pointing to the importance of developing an environment of colleagues that support the new intervention. In order to ensure successful skill transfer from the classroom to the community, agencies must also have access to trainers who can provide not only classroom training but also on-the-job coaching (Cully et al. 2010; Miller et al. 2004). Additionally, agencies must provide staff with the time necessary to practice skills and receive feedback. In an era of budget cuts and increasing caseloads, it becomes difficult for officers to focus on spending time on skill development.

In order to determine effectiveness of these intervention skills, actual use of the skill must be documented. Two facets must be considered: adherence and competence. Adherence suggests that the skill is being done, while competence indicates *how well* the skill is performed. While such ratings can initially be done through role-play situations, in order to be truly meaningful ultimately, the ratings must be on actual offender-officer interactions. Miller and Rose (2009) note "We know of no reliable and valid way to measure MI fidelity other than through the direct coding of practice samples. Clinicians' self-reported proficiency in delivering MI has been found to be unrelated to actual practice proficiency ratings by skilled coders (Miller and Mount 2001; Miller et al. 2004), and it is the latter ratings that predict treatment outcome." Coding of interactions can be accomplished through either direct observation or audio/videotaping of interactions. Only recently has the practice of audiotaping and coding interactions between officers and offenders been introduced (e.g., Bonta et al. 2008, 2010; Robinson et al. 2011). Prior to initiating audiotaping of interactions, several legal issues must be considered, including whether offenders have the right to refuse to be taped, who will have access to tapes, whether tapes will become part of an offender's official record, and whether tapes can be used in future legal proceedings. Since the purpose of taping is the skill development of the officer, we strongly advocate for tapes being accessible only to the extent that they are needed and used for training purposes. If tapes become a part of the legal case against the offender, the ability to obtain tapes may become difficult, and the usefulness of tapes will likely be diluted.

Conclusions and Future Research

The impact probation officers can have through their direct intervention with offenders is promising and supports the transition from compliance monitoring case management to a more comprehensive approach focused on the development of new skills that reduce risk. This finding is important given the incredible costs of recidivism, the typical responses to it (prison), and the costs typically associated with treating offenders. In an environment with seemingly always decreasing resources, it is important that we maximize the resources we continue to have available to us. Intentional CCP-based officer-offender interactions may support offender learning and skill mastery that is occurring via formal treatment programs, or may be stand-alone interventions. The purposeful skill-focused interaction allows agencies to extend the therapeutic environment beyond the traditional and often less effective treatment setting. This is important given the number of hours typically necessary to facilitate the change process for many high-risk offenders and the ability of most agencies to meet the demand. In order to ensure that these promising practices become routine, there must be a significant focus on effective implementation. To date, it is unclear exactly how long skill development will take, how many coaching sessions should occur, etc. Future research needs to focus on delineating these issues so the field has a clear vision of how to achieve effective practice.

Related Entries

- ▶ Actualizing Risk-Need-Responsivity
- ▶ Behavioral Management in Probation
- ▶ Correctional Workers in the Organizational Change Process
- ▶ Desistance and Supervision
- ▶ Effective Supervision Principles for Probation and Parole
- ▶ History of Probation and Parole in the United States
- ▶ History of Staff Skills in Probation
- ▶ Intensive Probation and Parole
- ▶ Offender Change in Treatment
- ▶ Offender Decision Making and Behavioral Economics
- ▶ Risk Assessment, Classification, and Prediction

Recommended Reading and References

Alexander M, VanBenschoten S, Walters S (2008) Motivational interviewing in criminal justice: development of a plan. Fed Probat 72(2):61–66

Andrews DA, Bonta J (2010) The psychology of criminal conduct, 5th edn. LexisNexis, New Providence

Andrews DA, Dowden C (2007) The risk-need-responsivity model of assessment and human service in prevention and corrections: crime-prevention jurisprudence. Can J Criminol Crim Justice 49:439–464

Bonta J, Bourgon G, Rugge T, Scott T, Yessine AK, Gutierrez L, Li J (2010) The strategic training initiative in community supervision: risk-need-responsivity in the real world. Public Safety Canada, Ottawa. ISBN 978-1-100-15750-4

Bonta J, Rugge T, Scott T, Bourgon G, Yessine AK (2008) Exploring the black box of community supervision. J Offender Rehabil 47(3):248–270

Bourgon G, Gutierrez L, Ashton J (2011) The evolution of community supervision practice: the transformation from case manager to change agent. Ir Probat J 8:28–48

Clark MD, Walters ST, Gingerich R, Meltzer M (2007) Motivating offenders to change: a guide for probation and parole. National Institute of Corrections, NIC Accession Number 022253

Cook JM, Schnurr PP, Biyanova T, Coyne JC (2009) Apples don't fall far from the tree: influences on psychotherapists' adoption and sustained use of new therapies. Psychiatr Serv 60(5):671–676

Cully JA, Teten AL, Benge JF, Sorocco KH, Kauth MR (2010) Multidisciplinary cognitive-behavioral therapy training for the veterans affairs primary care setting. Prim Care Companion J Clin Psychiatry 12(3)

Dowden C, Andrews DA (2004) The importance of staff practices in delivering effective correctional treatment: a meta-analysis of core correctional practices. Int J Offender Ther Comp Criminol 48(2):203–214

Fixsen DL, Naoom SF, Blase KA, Friedman RM, Wallace F (2005) Implementation research: a synthesis of the literature. University of South Florida, Louis de la Parte Florida Mental Health Institute, National Implementation Research Network, Tampa, http://nirn.fmhi.usf.edu/resources/publications/Monograph/index.cfm

Hettema J, Steele J, Miller WR (2005) Motivational interviewing. Annu Rev Clin Psychol 1:91–111

Joyce B, Showers B (2002) Student achievement through staff development, 3rd edn. Association for Supervision and Curriculum Development, Alexandria

Klein KJ, Knight AP (2005) Innovation implementation: overcoming the challenge. Curr Dir Psychol Sci 14(5):243–246

Lowenkamp CT (2010) Impact of motivational interviewing on offender outcomes: results from a pilot trial in US Probation. Unpublished data

Lowenkamp CT, Flores AW, Holsinger AM, Makarios M, Latessa EJ (2010) Intensive supervision programs: does program philosophy and the principles of effective interventions matter? J Crim Justice 38(4): 368–375

McMurran M (2009) Motivational interviewing with offenders: a systematic review. Leg Criminol Psychol 14(1):83–100

Miller WR, Mount KA (2001) A small study of training in motivational interviewing: does one workshop change clinician and client behavior? Behav Cogn Psychother 29:457–471

Miller WR, Moyers TB (2006) Eight stages in learning motivational interviewing. J Teach Addict 5:3–17

Miller WR, Rollnick S (2002) Motivational interviewing: preparing people for change, 2nd edn. The Guilford Press, New York

Miller WR, Rose GS (2009) Toward a theory of motivational interviewing. Am Psychol 64(6):527–537

Miller WR, Yahne CE, Moyers TB, Martinez J, Pirritano M (2004) A randomized trial of methods to help clinicians learn motivational interviewing. J Consult Clin Psychol 72(6):1050–1062

Prochaska JO, DiClemente CC, Norcross JC (1992) In search of how people change: applications to addictive behaviors. Am Psychol 47(9):1102–1114

Robinson C, VanBenschoten S, Alexander M, Lowenkamp CT (2011) A random (almost) Study of Staff Training Aimed at Reducing Rearrest (STARR): reducing recidivism through intentional design. Fed Probat 75:57–63

Rubak S, Sandboek A, Lauritzen T, Christensen B (2005) Motivational interviewing: a systematic review and meta-analysis. Br J Gen Pract 55(513):305–312

Shafran R, Clark DM, Fairburn CG, Arntz A, Barlowe DH, Ehlers A, Freeston M, Garety PA, Hollon SD, Ost LG, Salkovskis PM, Williams JMG, Wilson GT (2009) Mind the gap: improving the dissemination of CBT. Behav Res Ther 47:902–909

Taxman FS (2004) Tools of the trade: a guide to incorporating science into practice. Maryland Department of Public Safety and Correctional Services

Trotter C (1996) The impact of different supervision practices in community corrections: cause for optimism. Aust N Z J Criminol 29(1):29–46

Walters ST, Clark MD, Gingerich R, Meltzer M (2007) Motivating offenders to change: a guide for probation and parole. National Institute of Corrections, U.S. Department of Justice, Washington, DC

Walters ST, Vader AM, Nguyen N, Harris TR, Eelsl J (2010) Motivational interviewing as a supervision strategy in probation: a randomized effectiveness trial. J Offender Rehabil 49:309–323

Probation Officer Decision-Making

Arthur J. Lurigio
Department of Psychology, Loyola University Chicago, Chicago, IL, USA

Overview

Imposed in lieu of incarceration, probation is the most common sentence in the United States for convicted adults. A probation sentence orders convicted offenders to adhere to conditions of release that are specified by statute (mandatory conditions) or leveled in response to the particular circumstances of the case (special conditions). The former require probationers, for example, to report to their probation officers and to seek permission from the court to move from the sentencing jurisdiction. The latter require probationers, for example, to pay restitution to victims or to obtain treatment for drug or alcohol addiction. Probation officers are agents of the court responsible for enforcing the mandatory and special conditions of probation.

In the initial appearance stage of the court process, probation officers evaluate arrestees for bail. These evaluations are based on arrestees' charges as well as their criminal histories, ties to the community, and previous records of court appearances. Those with the most severe charges and criminal histories – especially those involving sex or other types of violent crimes – and records of failing to appear in court are the least likely to be recommended for bail or for a low amount of bail, ensuring that they will be detained in a jail until their cases are adjudicated. Before a sentence is rendered, probation officers conduct presentence investigations that inform the court in rendering the most appropriate sentence in light of a defendant's criminal and social histories. Officers typically recommend a sentence to probation for those who can be supervised safely in the community and benefit from social and behavioral healthcare services.

As part of their duties, probation officers also engage in assessments that guide their case management strategies at the post-adjudication stage of the court process. Based on an assessment of risk (i.e., the likelihood of continued criminal activity while on probation), officers decide on the level and frequency of contacts for each member of their caseloads. Higher-risk offenders are seen more often in the office and in home visits compared with lower-risk offenders who may be allowed to report by phone, mail, or computer. Based on an assessment of need (i.e., the rehabilitative and treatment services required to help offenders reintegrate into the community as law-abiding citizens), officers decide on the nature and intensity of services brokered in the community or provided in-house. Other decisions rendered by probation officers include the filing of petitions to the court requesting to revoke probation for offender rule breaking or to terminate an offender early for rule adherence. Whether officers are more inclined to monitor or counsel their cases depends in part on their backgrounds (e.g., law enforcement or social work) and styles of supervision.

Introduction

Of the numerous correctional reforms and innovations that emerged in the nineteenth century, few were as broadly or readily embraced as probation. Rooted in common law practices, such as suspended sentences and judicial reprieves, the first use of probation in the courts is widely attributed to the voluntary efforts of an altruistic Boston boot maker, John Augustus (1784–1859), who is considered the father of probation. Between 1841 and 1859, Augustus acted as advisor, advocate, and surety for nearly 2,000 people. Augustus was an early and ardent proponent of rehabilitation, declaring that the purpose of the law is to reform criminals and to prevent crime, not to punish criminals for the sake of retribution or vengeance. He was known to rebuke police officers, judges, and others who challenged his correctional philosophy; many officials regarded Augustus as an interloper in the court system and interfered with his efforts to persuade judges to release defendants to his guardianship. Augustus's progressive, humanitarian approach to supervision was designed to encourage offenders to repudiate their destructive habits (rehabilitation) and return to the community as law-abiding citizens (reintegration).

Augustus assisted lawbreakers by counseling and supervising them during periods of release from court custody. According to his records, among the first 1,100 probationers under his authority, only one forfeited bond. If any of his charges were too poor to pay court costs, Augustus advanced them a loan and provided them with lodging and subsistence. Rudimentary assessment strategies appear in his informal evaluations of offenders' intentions to remain sober and to become upright, law-abiding citizens. During Augustus's time, other concerned Bostonians emulated his efforts by putting up bail for defendants and offering employment to those who were jobless. These voluntary endeavors greatly impressed the court and were instrumental in leading Massachusetts to become the forerunner in probation legislation. Almost 20 years after Augustus's death, the Probation Act of 1878 called for the appointment of a paid probation officer (PO) to serve the courts, thereby creating the basic model of probation that remains in operation today. POs were given authority to investigate cases and recommend probation for people who could be reformed without punishment.

Probation, a sentence in lieu of incarceration that orders convicted offenders to adhere to conditions of release, is the most common sanction in the United States. At the end of 2009, more than four million adults were on probation. In general, the conditions of probation can be categorized as mandatory, punitive, or service oriented. All people sentenced to probation are subject to mandatory conditions as dictated by state statute, such as reporting to a PO and refraining from criminal activity and gun ownership. Mandatory conditions also prohibit offenders from leaving the jurisdiction without the express permission of the sentencing judge. Punitive and service-oriented conditions – known as special conditions – are tailored to the specific circumstances of the case and the special needs of probationers. The former include allowing unannounced home visits, fulfilling community service hours, paying fines and restitution, and obeying a curfew. The latter include being referred to treatment for psychiatric or substance use disorders and participating in vocational or educational programs. Conditions of probation must be reasonable and constitutional and serve a legitimate purpose in terms of rehabilitating offenders and protecting community safety. Throughout the period of supervision, the court retains the right to modify the conditions of probation and to resentence the offender when those conditions are violated.

Probation Officers' Roles

POs are agents of the court, responsible for a variety of activities before sentencing (pre-adjudication), as well as monitoring and enforcing the conditions of probation after sentencing (post-adjudication). To serve the court, POs must make several decisions that affect probationers throughout the criminal justice process. Like most other criminal justice professionals, these officers exercise discretion in rendering decisions about defendants and offenders. However, also like most other criminal justice professionals, their decisions must be bound by the law and in accordance with departmental policies, standard operating procedures, and professional standards.

The responsibilities and case management decisions of POs and parole agents are highly similar; they both strive to protect community safety by monitoring the conditions of release and reintegrating offenders into the community by brokering services. Unlike probation, which is a sentence in lieu of incarceration, parole is early release from prison (post-sentence). Probation officers enforce the mandates of the court system; parole agents enforce the mandates of the corrections system. Although parole agents are never involved in pretrial or presentencing decisions, they are involved in making decisions pertaining to assessment, supervision, and revocation. Unlike probation officers, parole agents struggle to help formerly incarcerated persons with the many and varied challenges of prisoner reentry.

Pretrial Decisions

The Federal Bail Reform Acts of 1966 and 1984 significantly changed the mechanisms for setting bail and were enacted to protect the constitutional rights of all defendants, especially the indigent who were languishing in jail because they were unable to afford the bond amount needed to secure their freedom before their cases were heard. An experiment by the Vera Institute of Justice, known as the Manhattan Bail Project, played a significant role in the bail reform movement and was instrumental in POs' participation in the bail decision-making process. The project examined whether judges could release defendants safely without monetary bail (on their own recognizance), contingent upon their providing the court with verifiable information about their families, jobs, and associates (criminal and noncriminal). The study found that defendants who had been released on their own recognizance were more likely to appear in court than defendants released with monetary bail, setting the stage for POs' evaluations of defendants for pretrial release and bond amounts.

The Bail Reform Act of 1966 enjoined the court to first consider releasing defendants on their own recognizance and to use a cash bond only in cases in which a recognizance release was deemed insufficient to guarantee a defendant's appearance in court. In addition, the court was required to select bail release options that were responsive to the needs of each defendant. The Act also defined the factors that should be considered in pretrial release decisions: the nature and circumstances of the current offense; the weight of evidence against the defendant; the defendant's social history and personal characteristics, including the defendant's character, mental health problems, family ties, employment status, and financial resources; the defendant's length of residence in the community; the defendant's criminal history and previous failures to appear before the court; and whether the defendant was on probation or parole or had other pending cases at the time of the arrest.

The Bail Reform Act of 1984 amended the previous reform act in order to permit the court to consider preventive detention as a means to protect the public from high-risk defendants. In such instances, a defendant could be denied bail and detained if the defendant was likely to pose a serious risk to public safety or to obstruct justice or intimidate witnesses or jurors. Bail also could be denied to defendants with lengthy criminal histories or to those accused of a felony, or a violent or drug offense that carried a potential life sentence or the death penalty.

At the pre-adjudication or pretrial level, POs present judges with information that helps them determine whether to release a defendant on bond. The process of collecting and presenting information to the court is based, in part, on bail legislation and research. POs' recommendations also inform judges in setting the type and amount of bond ordered. Officers conduct a comprehensive evaluation of defendants to arrive at these recommendations. The first area of investigation involves a defendant's risk of continued criminal activity and threat to public safety. To render decisions in this area, POs examine a defendant's current charge(s) and criminal history. In general, those who pose the greatest risk for committing a future crime have lengthier criminal records and are charged with more than one offense. Other risk factors include gang membership and addiction to drugs or alcohol.

The second area of investigation involves the risk of absconding (the failure to appear on the designated court dates), which is based on an assessment of community ties or connectedness. Defendants who have ties to social institutions (school, family, job) are considered a lower risk of absconding than defendants with weak or no ties (single, unemployed, and out of school). POs' decisions take the form of a recommendation to the court to release defendants on bail or keep them confined to jail pending court hearings or trial. A decision to recommend bond also can include a recommendation of bond type (cash, deposit, or recognizance) and dollar amount (high, moderate, or low).

Several studies have created and tested statistical tools for bail decision-making, which consist of factors that are related to pretrial outcomes. POs assess defendants for bail by scoring them on each factor – for example, current charges, criminal histories, and community ties – for the purpose of determining the likelihood that the defendant will appear in court and refrain from criminal activity while awaiting trial. The scores are added, and the sum equates to the level of risk. Defendants who score higher on such assessments usually are considered a higher risk for committing a future crime or absconding. In such cases, POs are likely to recommend detention or a high bond amount. Inversely, defendants who score lower are considered a lower risk. In such cases, POs are likely to recommend release to the community or a low bond amount. POs have the discretion to override the score and recommend a decision to the court that is based on their informed judgment of the defendant's suitability for release.

The Vera Institute of Justice introduced its "points system" of bail decision-making in the Manhattan Bail Project. In this system, more points predict lower risk. Specifically, defendants are awarded points for having no criminal record, living with immediate family in the area, being employed, or attending school. Defendants who have lived at same residence for several years or are cooperative during the bail interview also are awarded more points in the system.

The most common factors incorporated in bail assessment tools include prior court appearances, nature of the offense, and criminal history. Courts also examine defendants' employment and financial status, family and community ties, psychiatric and substance use disorders, length of residency, and evidence of dangerousness to the community. In addition, the court may order defendants to participate in pretrial supervision programs, which were implemented to relieve jail overcrowding by monitoring defendants in the community. POs' decisions about the level or closeness of pretrial supervision are based on their evaluation of roughly the same factors considered in POs' decisions about bond. Defendants who present the highest risk for criminality and absconding are monitored at the highest level of supervision, and vice versa.

Presentence Decisions

After a defendant has been convicted, POs shift the focus of their decisions from the pretrial to the presentence level, at which POs conduct a thorough evaluation of convicted offenders, or presentence investigation (PSI) – usually required for offenders convicted of a felony. The PSI consists of interviews with offenders and crime victims, as well as reviews of court and criminal records. POs compile this information into a PSI report, covering the offender's criminal, social, educational, military, and employment histories; current charges and the circumstances surrounding his or her arrest, such as the use of a weapon and victim harm; current living arrangements; and previous experiences in prison and on probation or parole. POs decide whether to recommend probation, and, if so, they prepare a supervision plan that specifies how the offender should be monitored and serviced. POs' decisions to recommend probation are highly important in cases in which a judge may sentence an offender to a term of either probation or imprisonment, that is, when the statute imposes no mandatory prison sentence for a given charge (Class X Felony) or a given offender (repeat felon statutes or "three strikes" laws).

Postsentence Decisions

POs' postsentencing decisions relate to the case management of offenders, which involves monitoring offenders in order to minimize their risk of continued criminal activity and responding to offenders' needs for services and treatment in order to assist them in living more productive and crime-free lives. POs also are interested in helping offenders to reintegrate into the community and restore their good citizenship. As the first step in the case management process, risk and needs assessment guide officers in rendering decisions about the nature and frequency of offender contacts, as well as the provision or brokerage of services. POs' assessment strategies have evolved through three generations of decision-making protocols.

Risk assessment is the cornerstone of effective offender management. The first generation of risk assessment was based on subjective or clinical judgments in which officers trusted their "gut instincts" and experiences in the formulation of case management plans. Early studies of PO and parole officers' supervisory decisions demonstrated that they used stereotypes and beliefs that crystallized through repeated contacts with offenders and were situated in the professional climates and cultures in which officers functioned. For example, one study showed that POs stored their information about probationers into stereotypic categories known as schemata, which consisted of information about offenders' past criminal behaviors, social histories, pathways to crime, and drug and alcohol use. Schemata are abstract cognitive representations of organized prior knowledge, extracted from experiences with specific cases. Types of probationer schemata have included "the gang member," "the professional burglar," "the white-collar criminal," "the drug addict," "the career criminal," and "the violent man." Experienced POs produced fewer but richer (more detailed) schemata than inexperienced POs. Research has shown that stereotypic categories, for example, "dangerous men," "criminals," and "sincere clients," also are used by parole agents to guide supervisory strategies, treatments, and prognoses, and to render decisions about the nature and frequency of offender contacts.

Decisions about schematic cases usually are made more quickly, easily, and confidently than decisions about non-schematic cases. Despite these advantages, the use of informal or stereotyped decision-making is fraught with limitations and errors. Clinical judgments – especially those that involve predictions of future risk – are mostly invalid (inaccurate) and unreliable (inconsistent).

Such professional judgments are unstructured and unstandardized. In addition, they are subject to personal bias and difficult to correct because of their deeply practiced and entrenched nature. Furthermore, professional assessments are difficult to study because each PO's decision-making procedure is idiosyncratic, and individual judgments have never been systematically tracked to determine their predictive value.

The second generation of risk assessment was based on actuarial (statistical) assessments of risk. Since the 1950s, research has found consistently that actuarial or statistical methods of risk assessment that are empirically derived predict criminal behavior more accurately than those that are clinically derived. As a result of their demonstrated predictive superiority, actuarial assessment and classification models have been implemented by correctional agencies in order to guide supervision practices. Based on prediction tools that were developed to gauge offender risk for parole release (e.g., the Salient Factor Score Model), probation departments created case classification systems that assigned probationers to levels of supervision, establishing a fairer and a more rational and effective use of department resources than could be achieved with clinical approaches to offender assessment. With such statistical models, officers' time could now be properly allocated to monitor the highest-risk and highest-need offenders. More intensive surveillance and services could be dedicated to manage offenders at higher levels of risk and with greater needs for services. A uniform classification system can also be used to describe the client population, to allocate department resources, and measure the effectiveness of community supervision practices.

The development of second-generation risk assessment tools often involves a multistep process of research/analysis. A sample of closed cases is selected randomly from a recently terminated population of probationers, yielding information about the predictive or outcome variable, such as rearrests during the probation period and status at termination (successful or unsuccessful fulfillment of the conditions of the sentence). The sample is split evenly between probationers who had a positive (no rearrests, completion of conditions) or a negative (rearrest or violation of conditions) outcome. Information on the outcome and a variety of predictor variables (e.g., age at intake, number of previous arrests, juvenile record) are collected and coded from the case files.

The sample is then divided in half, preserving in each half equal numbers of successful and unsuccessful probationers. The data in the first half of the sample, called the construction sample, are analyzed to identify factors that are statistically related to the outcome.

The data in second half of the sample, called the validation sample, are analyzed to test how well the factors can predict the outcome with a new group of cases. In this stage of the analysis, the accuracy of the variables is determined by comparing the percentages of correct and incorrect classifications, which include false-positive errors (predicting a failure when the case was actually successful) and false-negative errors (predicting a success when the case was actually a failure). Other measures of the accuracy of a prediction tool are its sensitivity (its power to select true positives) and specificity (its power to predict true negatives).

If the tool has a high rate of correct classifications (a low rate of errors), the next step in the process is to scale the factors and place them in an instrument. In a pilot test of the tool, POs score a sample of probationers on each variable and add up the scores to determine the probationer's level of risk: high, medium, or low. The pilot test results suggest a cutoff score for each of the risk categories.

In the final step of the process, POs are trained to use the tool and asked to complete questionnaires about its practicality and effectiveness. They also are also asked to track the amount of time they spend supervising offenders in each of the risk categories and to recommend improvements to the tool. Such feasibility studies are critical in encouraging the proper use of the tool and ensuring that changes are made before mistakes and poor practices become embedded in everyday usage.

Undertaken by the National Institute of Corrections (NIC) in the early 1980s, the Risk Classification Initiative encouraged the adoption of case classification systems in probation. NIC offered training and technical assistance to probation agencies nationwide in an effort to implement case classification systems as tools for accurate assessment and efficient resource allocation. The NIC Model Risk Classification system consisted of both risk and needs assessment instruments, which were developed in Wisconsin. The risk component of the assessment evaluates the likelihood that probationers will commit an offense while on probation. The Wisconsin Risk Assessment tool contained items such as age at first conviction, number of prior periods of probation and parole supervision, number of prior felony convictions, and convictions/adjudications for a violent crime.

The needs component of assessment evaluates the nature and extent of offenders' problems and the amount of time necessary for supervision and resource brokerage for cases with varying degrees of problems. The needs scale contained the following items: academic/vocational skills, employment problems, financial management, marital/family relationships, companions, emotional stability, alcohol use, other drug use, mental ability, health, sexual behavior, and officers' impressions of offender problems. Based on scores from the risk and needs assessment scales, offenders are assigned to maximum, medium, or minimum levels of supervision. POs place probationers in the highest supervision category indicated by either scale. However, with approval from their supervisors, POs can override the level of supervision dictated by the scales if individual circumstances warrant the reassignment of a probationer to a higher or lower level of monitoring. Maximum supervision probationers exhibit either a high potential for continued criminal activity or a substantial need for services, or both. Medium supervision probationers exhibit either a moderate potential for continued criminal activity or a moderate need for services, or both. Minimum supervision probationers exhibit little potential for continued criminal activity and few problems and therefore are likely to successfully complete probation.

In tandem with the Wisconsin risk and needs instruments, the NIC model used the Client Management Classification tool (CMC), which was a structured interview that helped POs formulate case plans for working with different types of offenders. The CMC identified four treatment/supervisory modalities for different types of probationers: selective intervention, environmental structure, casework/control, and limit setting.

Although second-generation actuarial prediction tools contained risk factors that are related statistically to outcomes, these factors consisted mostly of static or historical factors, which are unchangeable. For example, juvenile record, history of conduct disorder, and parental involvement in crime are all strong predictors of early adult offending. However, they also are immutable. In other words, such variables can explain criminality and reoffending but are unresponsive to probation officers' interventions. In contrast, criminogenic needs are dynamic factors, which form the building blocks of third-generation risk assessment batteries. When properly addressed by POs, dynamic factors can reduce offender risk. For example, if an offender is unemployed (dynamic risk factor), then the PO can help the offender find a job. Examples of other dynamic factors are educational status, faulty cognitions, deviant sexual fantasies, family relationships, leisure activities, criminal associates, and substance use disorders. The third-generation approach to risk assessment is known as "structured professional judgment."

According to research, the eight major risk factors associated with criminal conduct are antisocial/procriminal attitudes, values, and beliefs; procriminal associates and isolation from prosocial people; temperament and personality factors, such as impulsivity, adventurousness, and pleasure-seeking tendencies; history of antisocial behavior; family criminality or lack of family caring or cohesiveness; low levels of educational, vocational, or financial achievement; lack of prosocial leisure activities; and abuse of drugs and alcohol. All of these considered critical predictors but the first four, as also known as the

"Big Four," and are deemed the most predictive risk factors among the set. Third-generation tools often contain many of these factors.

The Level of Service Inventory-Revised (LSI-R) is the best-known and most widely adopted third-generation assessment tool. Like its second-generation predecessors, the LSI-R is used to evaluate offenders for case management decisions. Unlike earlier models of risk assessment, the LSI-R focuses on criminogenic needs; it also has a theoretical basis and considerable empirical support, attesting to its accuracy, reliability, and usefulness in several areas of correctional practices. In addition to the principles of risk and needs, the tool incorporates the principle of responsivity: tailoring interventions to each offender's motivational level, abilities, strengths, and learning style. The tool consists of 54 items, divided into ten dimensions or subscales: criminal history, employment/education, finance, accommodations, family/marital, leisure/recreation, friends/associates, drug/alcohol use, mental health, and attitude/orientation. Information to score the LSI-R is drawn from a thorough case review and semi-structured interview. Offenders who score higher on the LSI-R have a higher risk of recidivism, and vice versa.

Caseload Management Decisions

Offender case management strategies are predicated on discretionary decisions about the nature of the officer-offender relationship and the status of probationers during the supervisory period. These decisions cover the areas of supervision, treatment, early termination, and revocation. The surveillance or control aspect of the supervisory decision consists of two components. The first component is the frequency with which probationers report to their officers. The frequency of contacts is modified in accordance with the officer's assessment of the offender's level of risk or likelihood of continued criminal behavior. Probationers can report monthly (the most common frequency), bimonthly, or weekly. Higher-risk offenders report more often than lower-risk offenders.

The type or mode of supervision is the second component of the decision. POs can monitor a probationer through office visits, telephone contacts, mail-in reports, or computer-screen reporting at a kiosk. The offender's assessed level of risk can also determine the supervisory mode. For example, felony probationers are generally required to visit the probation office regularly, whereas less serious offenders are allowed mail-in or kiosk reports. Related to the determination of the supervisory mode are officers' decisions to assume a particular professional posture with different members of their caseloads.

POs adapt their attitudes, focus, and emotional tone to different offenders. Specifically, officers' supervisory styles are influenced by offenders' reporting demeanor; willingness to cooperate in the rehabilitation process; and desire to lead a productive and law-abiding life, for example, by finding a job, finishing school, and refraining from gang-related activity. These factors comprise what is known as the probationer's "attitude." Offenders who are honest in their self-disclosures and accept the conditions of their sentence are viewed as progressing satisfactorily and possessing a "positive attitude." A "negative attitude," on the other hand, is displayed in a probationer's continued belligerence, indifference, sarcasm, or blatant attempts to curry favor with an officer. Such behaviors are indicative of a poor adjustment to probation. In short, offenders with positive and negative attitudes are treated differently during office and home visits.

At any point during the probation period, POs can contact probationers' spouses, parents, teachers, friends, employers, or other professionals with whom the probationers are in contact. These contacts are often initiated to verify information, such as residence, employment, and the fulfillment of special conditions. POs also attempt to enlist the aid of others in efforts to control, rehabilitate, and reintegrate offenders. If POs suspect that a probationer is involved in illegal activities, they can obtain reports about subsequent arrests.

The second discretionary decision consists of an assessment of offenders' major problems, for example, emotional, medical, interpersonal, and financial. During initial meetings with a probationer, POs search for signs of drug/alcohol abuse, symptoms of serious psychological disorders (e.g., disorientation, bizarre ideation, or odd behaviors), intellectual deficits, or lack of social or vocational skills in order to evaluate probationers' needs for counseling or other services. Using the assessment tools described above, POs also can rely on their own sensitivity and experience to identify offender needs and to devise problem-solving strategies accordingly. Counseling sessions with officers follow a didactic, instructional style that fits better within a guardian-ward model than a therapist-client model. Officers' educational backgrounds rarely prepare them to conduct therapy with probationers who have mental health or substance use problems. In such cases, POs broker services, referring offenders to community-based programs that treat a variety of specific disorders. POs link offenders to these services, ensuring that they obtain access to interventions for their most pressing problems. Given their large caseloads, POs can spend only limited time counseling or advising offenders. The choice of available referrals is dictated by situations or factors outside the POs' control, for example, economic conditions that affect the number of job referrals and budget cuts for social service programs. Hence, officers must choose probationers for referrals who are most in need of treatment and the most likely to benefit from it. Offenders who request services, have minor criminal records, and display a "positive attitude" are generally considered the best candidates for counseling and other programs.

POs can recommend early terminations for offenders who pose no risk to the community. Early termination is warranted for probationers who have complied with their conditions and are on track for continued success in school or at work; for example, early termination can be recommended for an offender who asks to leave the state in order to accept a job. In contrast, for noncompliant probationers, officers can ask judges to extend periods of supervision up to the maximum sentence allowed by statute for the original offense.

Early termination is an incentive for "good behavior" and a demonstration to other probationers that cooperation and compliance with rules are rewarded. A key factor in the decision to recommend early termination is consistent offender reporting, which increases the likelihood that a case will be reviewed for early termination. Indeed, the "best probationers" are those who routinely report at their scheduled times. If a cancellation is unavoidable, these probationers promptly call their officers to inform them about the circumstances that prevented or that will prevent them from reporting at their scheduled times. POs take a dim view of probationers who are frequently late, skip appointments, and who call at the last minute or after the fact with implausible excuses for failing to report. When queried about the progress of a case, POs are likely to respond with a quick tally of the number of times a probationer missed a report day.

Release on probation is conditional and subject to compliance with the court-ordered terms or conditions of release. When probationers violate those terms, POs can initiate revocation proceedings. In most circumstances, the commission of a new crime during the probation term, known as a law violation, results in the filing of a violation of probation (VOP) petition, which is especially likely for a felony arrest. Technical VOPs involve no new offenses and are considered less serious infractions. In such cases, POs exercise discretion when deciding whether to file a VOP petition with the court. They do so by evaluating the infraction as well as offenders' performance on probation, criminal history, attitude, family relationships, and employment status. Probationers with a consistent record of reporting to their POs and of gainful employment are usually allowed to continue on probation for relatively minor transgressions.

After filing a VOP petition, POs must decide to recommend the imposition of a more serious sentence or a continuation of probation with more stringent conditions of release. This determination is based on the same set of factors as the decision to file a VOP. Officers might recommend a jail term for probationers during which they can receive more intensive services and contemplate the value of their freedom. The court can reject POs' recommendations and dismiss the VOP petition, thereby allowing the offender to continue serving the term under the original conditions of release. Nonetheless, as in the case of early termination, the court often abides by officers' recommendations to revoke probation and to impose a more serious penalty. In three landmark cases, the Supreme Court ruled that probationers and parolees retain their due process rights in revocation hearings (Morrissey v. Brewer, Gagnon v. Scarpelli, and Mempa v. Rhay).

Probation Officer Typologies and Styles

POs wear many hats in pursuing the primary goals of probation. Different typologies describe the various roles of officers, such as the punitive/law enforcement officer, the social worker, the civil servant, and the synthetic officer. Punitive/law enforcement POs are concerned overwhelmingly with exercising their legal authority and enforcing the conditions of probation. These POs regard probation as a privilege, not as a right. The second type of officer is the social worker or therapeutic agent, who strives to rehabilitate and reintegrate offenders into the community. Social workers cultivate a helping relationship with offenders by formulating plans for treatment and services. In contrast, civil servants or timeservers exhibit little concern for the welfare of the community or the probationer. They concentrate on maintaining or advancing their position in the probation bureaucracy with the ultimate goal of retirement, pension, or entry into another field, for example, law or police work. Civil servants observe rather than initiate behavioral changes in their clients. Synthetic/protective officers' supervisory styles reflect the basic orientations of the rule enforcer and social worker, recognizing the importance of both the treatment and control components of probation. Synthetic/protective officers integrate monitoring with rehabilitation. A more fundamental distinction has been made between the probation officer as a law enforcement agent and as a service broker and client advocate. In most cases, POs shift their emphasis to one role or the other depending on their assessment of each probationer's risk and needs as well as his or her progress on probation.

Related Entries

- ▶ Effective Supervision Principles for Probation and Parole
- ▶ History of Probation and Parole in the United States
- ▶ Probation and Parole Practices
- ▶ Risk Assessment, Classification, and Prediction

Recommended Reading and References

Abadinsky H (2002) Probation and parole: theory and practice. Prentice Hall, Englewood Cliffs

Augustus J (1972) John Augustus: the first probation officer. Patterson Smith, Montclair

Austin J (2006) How much risk can we take? The misuse of risk assessment in corrections. Fed Probat 70:58–63

Baird C, Heinz RC, Benus B (1982) The Wisconsin classification/staff development project: a two-year, follow-up report. American Correctional Association, College Park

Bonta J, Wormith SJ (2008) Risk and need assessment. In: McIvor G, Raynor P (eds) Developments in social work with offenders. Jessica Kingsley, London, pp 131–152

McCleary R (1978) Dangerous men: the sociology of parole. Sage, Beverly Hills

Petersilia J (2002) Reforming probation and parole. American Correctional Association, Lanhan

Travis J, Petersilia J (2001) Reentry reconsidered: a new look at an old question. Crime Delinq 47:291–313

Problem Solving Policing

▶ Problem-Oriented Policing

Problem-Oriented Courts

▶ Problem-Solving Courts

Problem-Oriented Policing

Anthony A. Braga
School of Criminal Justice, Rutgers University, Newark, NJ, USA
Kennedy School of Government, Harvard University, Cambridge, MA, USA

Synonyms

Problem Solving Policing

Overview

Problem-oriented policing is an alternative approach to crime reduction that challenges police officers to understand the underlying situations and dynamics that give rise to recurring crime problems and to develop appropriate responses to address these underlying conditions. Problem-oriented policing is often given operational structure through the well-known SARA model that includes a series of iterative steps: Scanning, Analysis, Response, and Assessment. Police officers often find it difficult to implement problem oriented policing properly with deficiencies existing in all stages of the process. The existing evaluation evidence shows that problem-oriented policing generates noteworthy crime and disorder reduction impacts. These crime reduction impacts are generated even when problem-oriented policing is not fully implemented; this confirms the robustness of the problem-oriented approach in addressing crime and disorder problems.

Introduction

Most police departments have historically engaged incident-driven crime prevention strategies. In dealing with crime, these departments were aimed at resolving individual incidents instead of addressing recurring crime problems (Eck and Spelman 1987). Officers responded to repeated calls and never looked for the underlying conditions that may be causing like groups of incidents. Officers often became frustrated because they answered similar calls and seemingly made no real progress. Citizens also became dissatisfied because the problems that generate their repeated calls still existed (Eck and Spelman 1987). In a seminal article that challenged existing police policy and practice, Herman Goldstein (1979) proposed an alternative; he felt that police should go further than answering call after call and search for solutions to recurring problems that generate the repeated calls. Goldstein (1979) described this strategy as the "problem-oriented approach" and envisioned it as a department-wide activity. Problem-oriented policing is now a common police crime prevention and control strategy.

Problem-oriented policing seeks to identify the underlying causes of crime problems and to frame appropriate responses using a wide variety of innovative approaches (Goldstein 1979). Using a basic iterative approach of problem identification, analysis, response, assessment, and adjustment of the response, this adaptable and dynamic analytic approach provides an appropriate framework to uncover the complex mechanisms at play in crime problems and to develop tailor-made interventions to address the underlying conditions that cause crime problems (Eck and Spelman 1987; Goldstein 1990). Since the publication of Goldstein's article, many police departments have experimented with the approach and the available evaluation evidence suggests that problem-oriented policing is a fundamentally sound approach to controlling crime and disorder problems (Skogan and Frydl 2004; Braga 2008; Weisburd et al. 2010).

The Principles of Problem-Oriented Policing

Beginning in the 1940s and continuing through the emergence of community policing, police departments followed what many have come to call the "standard" or "professional" model of policing (see e.g. Skogan and Frydl 2004) that was characterized by rigorous professional standards, militaristic organizational structures, the use of technology, and other important historical reforms. Under this model, police departments attempted to prevent serious crimes by advancing three operational strategies: preventive patrol, rapid response, and investigation of more serious cases by specialized detective units. The limits of these strategies are, by now, well known (Skogan and Frydl 2004). Research studies found that varying levels of preventive patrol did not reduce crime, rapid response to calls for service did not increase the probability of arrest as very few crimes are reported in progress, and follow-up investigations solved only a relatively small proportion of reported crimes.

The findings of these studies had a strong impact on a generation of police scholars and practitioners. By the early 1990s, there was a broad consensus among criminologists and police scholars that crime was a product of larger social forces, and the police could do little if anything to impact upon crime or crime rates (Gottfredson and Hirschi 1990; Bayley 1994). The police as "crime fighters" might have been a popular idea in the media and among the public, but the idea that the police could do something about crime had little credence in the universities or research institutes that were concerned with policing.

In 1979, Herman Goldstein, a respected University of Wisconsin law professor and former aide to Chicago police chief O.W. Wilson, made a simple and straightforward proposition that challenged police officers to address problems rather than simply respond to incidents. According to Goldstein (1979, 1990), behind every recurring problem there are underlying conditions that create it. Incident-driven policing never addresses these conditions; therefore incidents are likely to recur. Answering calls for service is an important task and still must be done, but police officers should respond systematically to recurring calls for the same problem. In order for the police to be more efficient and effective, they must gather information about incidents and design an appropriate response based on the nature of the underlying conditions that cause the problem(s) (Goldstein 1990).

It is important to note here that Herman Goldstein (1979, 1990) intended problem-oriented policing to be a general approach that could be applied to a wide range of police business problems. This includes non-crime problems such as personnel issues, budgetary concerns, and police-community relations. Most problem-oriented policing research and practical experience, however, has focused on applying the approach to addressing crime and disorder problems. As such, this entry examines our existing knowledge base on police use of the problem-oriented approach to tackle recurring crime and disorder problems.

The Process of Problem-Oriented Policing

The problem-oriented policing approach was given an operational structure in Newport News, Virginia. Researchers from the Police Executive Research Forum (PERF) and a group of officers selected from the various ranks of the Newport News Police Department crystallized the philosophy into a set of steps known as the SARA model (Eck and Spelman 1987). The SARA model consists of these stages: *Scanning*- the identification of an issue and determining whether it is a problem; *Analysis*- data collection on the problem to determine its scope, nature, and causes; *Response*- information from the analysis is used to design an appropriate response which can involve other agencies outside the normal police arena; and *Assessment*- the response is evaluated and these results can be used to re-examine the problem and change responses or maintain positive conditions (Eck and Spelman 1987). In practice, it is important to recognize that the development and implementation of problem-oriented responses do not always follow the linear, distinct steps of the SARA model (Capowich and Roehl 1994; Braga 2008). Rather, depending on the complexity of the problems to be addressed, the process can be characterized as a series of disjointed and often simultaneous activities. A wide variety of issues can cause deviations from the SARA model, including identified problems needing to be re-analyzed because initial responses were ineffective and implemented responses that sometimes reveal new problems (Braga 2008).

Scanning. The process of scanning involves the identification of problems that are worth looking at because they are important and amenable to solution. Herman Goldstein (1990) suggests that the definition of problems be at the street-level of analysis and not be restricted by preconceived typologies. Goldstein further clarifies what is meant by a problem by specifying the term as: “a cluster of similar, related, or recurring incidents rather than a single incident; a substantive community concern; or a unit of police business” (1990, p. 66).

Analysis. The analysis phase challenges police officers to analyze the causes of problems behind a string of crime incidents or substantive community concern. Once the underlying conditions that give rise to crime problems are known, police officers develop and implement appropriate responses. The challenge to police officers is to go beyond the analysis that naturally occurs to them; namely, to find the places and times where particular offenses are likely to occur, and to identify the offenders that are likely to be responsible for the crimes. Although these approaches have had some operational success, this type of analysis usually produces directed patrol operations or a focus on repeat offenders. The idea of analysis for problem solving was intended to go beyond this. Goldstein (1990) describes this as the problem of “ensuring adequate depth” in the analysis.

Situational crime prevention has further developed the methodology of analyzing problems, and provided important examples of how crime problems may be closely analyzed. Situational crime prevention measures are tailored to highly specific categories of crime. As Clarke (1997) describes, distinctions must be made not between broad crime categories such as burglary and robbery, but between the different kinds of offenses that comprise each of these categories. For example, in their analysis of domestic burglary in a British city, Poyner and Webb (1991) revealed that cash and jewelry burglaries tended to occur in older homes near the city center, while burglaries of electronic goods, such as TVs and VCRs, generally occurred in newer homes in the suburbs. Analysis further revealed that offenders on foot committed cash and jewelry burglaries. In the electronic goods burglaries, offenders used cars that had to be parked near to the house, but not so close that they would attract attention. The resulting crime prevention strategies differed accordingly. To prevent cash and jewelry burglaries in the city center, Poyner and Webb (1991) recommended improving security and surveillance at the burglar’s point of entry; in contrast, to prevent electronic good burglaries in the suburbs, they suggested improving the natural surveillance of parking places and roadways in the area.

Response. After a problem has been clearly defined and analyzed, police officers confront the challenge of developing a plausibly effective response. The development of appropriate responses is closely linked with the analysis that is performed. The analysis reveals the potential targets for an intervention, and it is at least partly the idea about what form the intervention might take that suggests important lines of analysis. As such, the reason police often look at places and times where crimes are committed is that they are already imagining that an effective way to prevent the crimes would be to get officers on the scene through directed patrols. The reason they often look for the likely offender is that they think that the most effective and just response to a crime problem would be to arrest and incapacitate the offender. However, the concept of problem-oriented policing as envisioned by Herman Goldstein (1990), calls on the police to make a much more “uninhibited” search for possible responses and not to limit themselves to getting officers in the right places at the right times, or identifying and arresting the offender (although both may be valuable responses). Effective responses often depend on getting other people to take actions that reduce the opportunities for criminal offending, or to mobilize informal social control to drive offenders away from certain locations.

Assessment. The crucial last step in the practice of problem-oriented policing is to assess the impact the intervention has had on the problem it was supposed to solve. Assessment is important for at least two different reasons. The first is to ensure that police remain accountable for their performance and for their use of resources. Citizens and their representatives want to know how the money and freedom they surrendered to the police are being used, and whether important results in the form of less crime, enhanced security, or increased citizen satisfaction with the police has been achieved. A second reason assessment is important is to allow the police to learn about what methods are effective in dealing with particular problems. Unless the police check to see whether their efforts produced a result, it will be hard for them to improve their practices.

In general, problem-oriented police should strive to conduct more rigorous assessments of their responses with due consideration to time and resource constraints. Depending on the availability of funds, police departments should consider partnering up with independent researchers to conduct systematic evaluations of their efforts. In the absence of such a partnership, Clarke (1998) suggests that police should take care to relate any observed results to specific actions taken, develop assessment plans while outlining the project, present control data when available and reasonably comparable to the subject(s) of the intervention, and, as will be discussed further, measure crime displacement. While the degree of rigor applied to the assessment of responses may vary, what must not be sacrificed is the goal of measuring results. This will keep the police focused on results rather than means, and that is one of the most important contributions of the idea of problem-oriented policing.

Community Policing and "Problem Solving"

During the late 1980s and throughout the 1990s, problem-oriented policing and community policing were heralded as revolutionary alternatives to the professional model. The terms are sometimes referred to as essentially the same strategy (Kennedy and Moore 1995), however others maintain a distinct separation between the two concepts (Eck and Spelman 1987; Goldstein 1990). Problem-oriented policing is typically defined as focusing police attention on the underlying causes of problems behind a string of crime incidents, while community policing emphasizes the development of strong police-community partnerships in a joint effort to reduce crime and enhance security (Moore 1992). Indeed, community-oriented police officers use problem solving as a tool and problem-oriented departments often form partnerships with the community.

The term "problem solving" is often conceptualized as what an officer does to handle small, recurring beat-level problems, and it is distinguished from problem-oriented policing based on its rudimentary analysis of the problem and lack of formal assessment (see, e.g., Cordner 1998). In short, some observers suggest the term "problem solving" does not adequately capture the substance of problem-oriented policing as envisioned by Goldstein (1990). Scott (2000) reports that Goldstein himself has been especially careful to avoid the term "problem solving" because many, if not most, problems the police confront are too complex for anything approaching a final solution; reducing harm, alleviating suffering, and/or providing some measure of relief are ambitious enough aims for the police.

Situational Crime Prevention and Supporting Theoretical Perspectives

The field of situational crime prevention has supported the problem-oriented policing movement since its genesis in the British Government's Home Office Research Unit in the early 1980s (Clarke 1997). Instead of preventing crime by altering broad social conditions such as poverty and inequality, situational crime prevention advocates changes in local environments to decrease opportunities for crimes to be committed. Situational crime prevention techniques comprise opportunity-reducing measures that are: "(1) directed at highly specific forms of crime (2) that involve the management, design, or manipulation of the immediate environment in as systematic and permanent way as possible (3) so as to increase the effort and risks of crime and reduce the rewards as perceived by a wide range of offenders" (Clarke 1997, p. 4). Like problem-oriented policing, the situational analysis of crime problems follows an action-research model that systematically identifies and examines problems, develops solutions, and evaluates results. Applications of situational crime prevention have shown convincing crime prevention results when applied to a variety problems ranging from obscene phone callers to burglary to car radio theft. This simple but powerful perspective is applicable to crime problems facing the police, security personnel, business owners, local government officials, and private citizens.

Problem-oriented policing and situational crime prevention draw upon theories of criminal opportunity, such as rational choice and routine activities, to analyze crime problems and develop appropriate responses (Clarke 1997; Braga 2008). Most criminological research focuses on why some people become persistent offenders. However, as Eck (2000) observes, by the time a problem comes to the attention of the police, the questions of why people offend are no longer relevant. The most pressing concerns are why offenders are committing crimes at particular places, selecting particular targets, and committing crimes at specific times (Eck 2000). While police officers are important entry points to social services for many people, they are best positioned to prevent crimes by focusing on the situational opportunities for offending rather than attempting to manipulate socio-economic conditions that are the subjects of much criminological inquiry and the primary focus of other governmental agencies (Braga 2008). Theories that deal with the "root causes" of crime focus on interventions that are beyond the scope of most problem-oriented policing projects. Theories that deal with opportunities for crime and how likely offenders, potential victims, and others make decisions based on perceived opportunities have greater utility in designing effective problem-oriented policing interventions.

The rational choice perspective assumes that "crime is purposive behavior designed to meet the offender's commonplace needs for such things as money, status, sex, and excitement, and that meeting these needs involves the making of (sometimes quite rudimentary) decisions and choices, constrained as these are by limits of time and ability and the availability of relevant information" (Clarke 1995, p. 98). Rational choice makes distinctions between the decisions to initially become involved in crime, to continue criminal involvement, and to desist from crimina offending, as well as the decisions made to complete a particular criminal act. This separation of the decision-making processes in the criminal event from the stages of criminal involvement allows the modeling of the commission of crime events in a way that yields potentially valuable insights for crime prevention.

The emphasis of the rational choice perspective on concepts of risk, reward, and effort in criminal decision making has been used to inform the development of problem-oriented policing and situational crime prevention strategies that seek to change offender appraisals of criminal opportunities (Clarke 1997; Braga 2008). Of particular importance, the decision processes and information utilized in committing criminal acts can vary greatly across offenses; ignoring these differences and the situational contingencies associated with making choices may reduce the ability to effectively intervene (Clarke 1995). For example, a robber may choose a "favorite" spot because of certain desirable attributes that facilitate an ambush, such as poor lighting and untrimmed bushes. One obvious response to this situation would be to improve the lighting and trim the bushes.

Rational choice is often combined with routine activity theory to explain criminal behavior during the crime event. Rational offenders come across criminal opportunities as they go about their daily routine activities and make decisions whether to take action. The source of the offender's motivation to commit a crime is not addressed (it is assumed that offenders commit crimes for any number of reasons); rather, the basic ingredients for a criminal act to be completed are closely examined. Routine activity posits that a criminal act occurs when a likely offender converges in space and time with a suitable target (e.g., victim or property) in the absence of a capable guardian (e.g., property owner or security guard) (Cohen and Felson 1979).

The routine activity approach was used to demonstrate that increases in residential burglary in the United States between 1960 and 1970 could be largely explained by changes in the routine activities of households (Cohen and Felson 1979). During this time period, the number of empty homes during the day increased as the number of single-person households and female participation in the workforce grew. At the same time, households increasingly contained attractive items to steal, such as more portable televisions and other electronic goods. Burglary increased, as fewer capable guardians were present in the home to protect the new suitable targets from burglars. These kinds of analytical insights on the nature of crime problems are very valuable to problem-oriented police officers seeking to develop appropriate responses.

The Practice of Problem-Oriented Policing

Although the problem-oriented approach has demonstrated much potential value in improving police practices, research has also documented that it is very difficult for police officers to implement problem-oriented policing strategies (Eck and Spelman 1987; Clarke 1998; Braga and Weisburd 2006). Cordner (1998) identifies a number of challenging issues in the substance and implementation of many problem-oriented policing projects. These issues include: the tendency for officers to conduct only a superficial analysis of problems and rushing to implement a response, the tendency for officers to rely on traditional or faddish responses rather than conducting a wider search for creative responses, and the tendency to completely ignore the assessment of the effectiveness of implemented responses (Cordner 1998). Indeed, the research literature is filled with cases where problem-oriented policing programs tend to lean towards traditional methods and where problem analysis is weak. In his review of several hundred submissions for the Police Executive Research Forum's Herman Goldstein Award for Excellence in Problem-Oriented Policing, Clarke (1998) laments that many recent examples of problem-oriented policing projects bear little resemblance to Goldstein's original definition and suggests this misrepresentation puts the concept at risk of being pronounced a failure before it has been properly tested.

Deficiencies in current problem-oriented policing practices exist in all phases of the process. During the scanning phase, police officers risk undertaking a project that is too small (e.g., the lonely old man who repeatedly calls the police for companionship) or too broad (e.g., gang delinquency) and this destroys the discrete problem focus of the project and leads to a lack of direction at the beginning of analysis (Clarke 1998). Some officers skip the analysis phase or conduct an overly simple analysis that does not adequately dissect the problem or does not use relevant information from other agencies (such as hospitals, schools, and private businesses) (Clarke 1998). Based on his extensive experience with police departments implementing problem-oriented policing, Eck (2000) suggests that much problem analysis consists of a simple examination of police data coupled with the officer's working experience with the problem. In their analysis of problem-oriented initiatives in 43 police departments in England and Wales, Read and Tilley (2000) found that problem analysis was generally weak with many initiatives accepting the definition of a problem at face value, using only short-term data to unravel the nature of the problem, and failing to adequately examine the genesis of the crime problems.

Given the limited analysis that many crime problems, it is not surprising that the responses of many problem-oriented policing projects rely too much on traditional police tactics (such as arrests, surveillance, and crackdowns) and neglect the wider range of available alternative responses. Read and Tilley (2000) found that officers selected certain responses prior to, or in spite of, analysis; failed to think through the need for a sustained crime reduction; failed to think through the mechanisms by which the response could have a measurable impact; failed to fully involve partners; narrowly focused responses, usually on offenders; as well as a number of other weakness in the response development process. Finally, Scott and Clarke (2000) observed that the assessment of responses is rare and, when undertaken, it is usually cursory and limited to anecdotal or impressionistic data.

Reflecting on these practical issues, Eck (2000) comments that the problem-oriented policing that is practiced by many police department diverges significantly from the original concept that was envisioned by Goldstein. Braga and Weisburd (2006), however, find value in the imperfect implementation of problem-oriented policing. They argue that there is much evidence that what might be called "shallow" problem solving responses can be effective in combating crime problems. Apparently, weak problem-oriented policing is better than none at all. This being the case, Braga and Weisburd (2006) question whether the pursuit of problem-oriented policing as it has been modeled by Goldstein and others, should be abandoned in favor of the achievement of a more realistic type of problem solving. While less satisfying for scholars, it is what the police have tended to do, and it has been found to lead to real crime prevention benefits.

"Ideal" Applications of Problem-Oriented Policing

Herman Goldstein (1979, 1990) originally suggested that problem-oriented policing efforts should be located within a headquarters unit rather than assigned to police officers in operational units. The decentralization of the approach to street police officers may have reduced the quality of routine problem-oriented policing efforts as busy officers handled too many problems and had little time to conduct the extensive analysis and search for appropriate responses. Ideal applications of problem-oriented policing tend to involve larger scale problems, the involvement of academic researchers and crime analysis units, and the solid support of the police command staff to implement alternative responses. Two examples of these types of "ideal" problem-oriented projects are briefly reviewed here: The Boston Police Department's Operation Ceasefire intervention to prevent gang violence (Braga et al. 2001) and the Charlotte-Mecklenburg Police Department's program to reduce theft from construction sites (Clarke and Goldstein 2002).

The Boston Gun Project was a problem-oriented policing enterprise expressly aimed at taking on a serious, large-scale crime problem — homicide victimization among young people in Boston. Like many large cities in the United States, Boston experienced a large sudden increase in youth homicide between the late 1980s and early 1990s. The Boston Gun Project proceeded by: (1) assembling an interagency working group of largely line-level criminal justice and other practitioners; (2) applying quantitative and qualitative research techniques to create an assessment of the nature of, and dynamics driving, youth violence in Boston; (3) developing an intervention designed to have a substantial, near-term impact on youth homicide; (4) implementing and adapting the intervention; and (5) evaluating the intervention's impact (Kennedy et al. 1996). The Project began in early 1995 and implemented what is now known as the "Operation Ceasefire" intervention, which began in the late spring of 1996.

The trajectory of the Boston Gun Project and of Operation Ceasefire is by now well known and extensively documented. Briefly, the working group of law enforcement personnel, youth workers, and researchers diagnosed the youth violence problem in Boston as one of patterned, largely vendetta-like ("beef") hostility amongst a small population of chronic offenders, and particularly among those involved in some 61 loose, informal, mostly neighborhood-based groups. These 61 gangs consisted of between 1,100 and 1,300 members, representing less than 1 % of the city's youth between the ages of 14 and 24. Although small in number, these gangs were responsible for more than 60 % of youth homicide in Boston.

The Operation Ceasefire "pulling levers" strategy was designed to deter violence by reaching out directly to gangs, saying explicitly that violence would no longer be tolerated, and backing up that message by "pulling every lever" legally available when violence occurred. Simultaneously, youth workers, probation and parole officers, and later churches and other community groups offered gang members services and other kinds of help. The Ceasefire Working Group delivered this message in formal meetings with gang members, through individual police and probation contacts with gang members, through meetings with inmates at secure juvenile facilities in the city, and through gang outreach workers. The deterrence message was not a deal with gang members to stop violence. Rather, it was a promise to gang members that violent behavior would evoke an immediate and intense response. If gangs committed other crimes but refrained from violence, the normal workings of police, prosecutors, and the rest of the criminal justice system dealt with these matters. But if gangs hurt people, the Working Group concentrated its enforcement actions on their members.

A large reduction in the yearly number of Boston youth homicides followed immediately after Operation Ceasefire was implemented in mid-1996. A US Department of Justice (DOJ)-sponsored evaluation of Operation Ceasefire revealed that the intervention was associated with a 63 % decrease in the monthly number of Boston youth homicides, a 32 % decrease in the monthly number of shots-fired calls, a 25 % decrease in the monthly number of gun assaults, and, in one high-risk police district given special attention in the evaluation, a 44 % decrease in the monthly number of youth gun assault incidents (Braga et al. 2001). The evaluation also suggested that Boston's significant youth homicide reduction associated with Operation Ceasefire was distinct when compared to youth homicide trends in most major US and New England cities (Braga et al. 2001).

To many observers, the analysis phase is the critical step in the problem-oriented policing process as it unravels the nature of recurring problems and points police towards innovative responses that go beyond traditional enforcement activities. Clarke and Goldstein (2002) document the vital role played by innovative crime analysis in a problem-oriented policing project undertaken by the Charlotte-Mecklenburg Police Department to address a sharp increase in the number of kitchen appliances stolen from new houses under construction. A detailed analysis of security practices and risks for theft among 25 builders in one police service district was conducted. The analysis led to the recommendation that the installation of appliances should be delayed until the new owners moved into the residence. Removing the targets of theft was found to be an effective response: as appliance theft declined markedly in the police service district and there was no evidence of displacement to surrounding district (Clarke and Goldstein 2002).

A key moment in the analysis of the theft problem occurred when the crime analyst discovered that a "certificate of occupancy" had to be issued by the county before a new owner could move into the residence. Compared to the building permits that had been used in earlier iterations of the problem analysis, these certificates provided a better measure of when a house was ready to be occupied and, therefore, a timelier basis for calculating the risk of theft. Building permits measured only planned construction. Builders may obtain 100 permits to build houses, but only actually build a fraction in a given year. That is why building permit data could not be used to accurately assess the stage when a house was completed and, thus, at-risk for theft of newly installed appliances.

Crime Prevention Effects of Problem-Oriented Policing

There is a growing body of evaluation evidence that problem-oriented policing generates noteworthy crime control gains. The US National Academy of Sciences' Committee to Review Research on Police Policy and Practices recently concluded that problem-oriented policing is a promising approach to deal with crime, disorder and fear and recommended that additional research was necessary to understand the organizational arrangements that foster effective problem solving (Skogan and Frydl 2004; Weisburd and Eck 2004). Several published volumes on problem-oriented policing case studies provide a good sense for the work being done as well as the strengths and weaknesses of some of the better problem-oriented efforts. Indeed, the widespread use of problem-oriented policing as a central crime prevention and control strategy in police agencies across the world is a strong indicator of the practical value of the approach.

However, the strongest empirical evidence available in support of problem-oriented policing comes from three Campbell Collaboration systematic reviews. Formed in 2000, the Campbell Collaboration Crime and Justice Group aims to prepare and maintain systematic reviews of criminological interventions and to make them electronically accessible to scholars, practitioners, policy makers and the general public (www.campbellcollaboration.org). The Crime and Justice Group requires reviewers of criminological interventions to select studies with high internal validity such as randomized controlled trials and quasi-experiments.

David Weisburd and his colleagues (2010) recently completed a Campbell Collaboration systematic review of the crime prevention effects of problem-oriented policing on crime and disorder. Despite reviewing a very large number of empirical studies on the approach, they identified only ten problem-oriented policing studies that used more rigorous randomized experimental and quasi-experimental evaluation designs. Given the popularity of problem-oriented policing, Weisburd et al. (2010) were surprised by the small number of rigorous evaluations studies that examined the crime prevention benefits of the approach. A meta-analysis of these ten evaluations revealed that problem-oriented policing programs generated a modest but statistically-significant impact on crime and disorder outcomes. These results were consistent when Weisburd et al. (2010) examined randomized experiments and quasi-experiments separately.

The Campbell problem-oriented policing review also reported on the crime prevention effects of simple pre/post comparison evaluation studies. While these studies did not include a comparison group and were less methodologically rigorous, Weisburd et al. (2010) found that they were far more numerous and identified 45 pre/post evaluations. 43 of these 45 evaluations reported that the approach generated beneficial crime prevention effects. These studies also reported much larger crime reduction impacts associated with problem-oriented policing when compared to the effects reported by the more rigorous research designs.

Policing crime hot spots represents an important advance in focusing police crime prevention practice (Braga and Weisburd 2010). Since crime hot spots generate a bulk of urban crime problems, it seems commonsensical to address the conditions and situations that give rise to the criminal opportunities that sustain high-activity crime places. The available evaluation evidence also suggests that problem-oriented policing holds great promise in addressing the criminogenic attributes of specific places that cause them to be crime hot spots. A recently updated Campbell Collaboration review of 19 rigorous evaluation studies found that hot spots policing generates modest crime reductions and these crime control benefits diffuse into areas immediately surrounding targeted crime hot spots (Braga et al. 2012). A moderator analysis of the types of hot spots policing programs found that problem-oriented policing interventions generate larger crime control impacts when compared to interventions that simply increase levels of traditional police actions in crime hot spots.

The Campbell hot spots policing review also reported that the problem-oriented policing interventions generated larger diffusion of crime control benefits into areas immediately surrounding the targeted hot spots areas (Braga et al. 2012). Many of the problem-oriented policing interventions used to control crime hot spots were described as suffering from superficial problem analyses, a preponderance of traditional policing tactics, and limited situational crime prevention responses. However, these generally "shallow" problem-oriented policing programs still generated crime reduction gains. This finding supports the assertion made earlier by Braga and Weisburd (2006) that limited problem-oriented policing, also known simply as "problem solving," is a stronger approach to crime prevention when compared to traditional police crime prevention strategies.

A number of US jurisdictions have been experimenting with new problem-oriented policing frameworks, generally known as pulling levers focused deterrence strategies, to understand and respond to serious crime problems generated by chronically offending groups, such as gun violence among gang-involved offenders (Kennedy 2008). These approaches include the well-known Boston Gun Project and its Operation Ceasefire intervention (Braga et al. 2001) discussed earlier in this entry and typically represent carefully implemented problem-oriented policing projects. Another recently completed Campbell Collaboration systematic review of focused deterrence strategies found that 10 out of 11 rigorous evaluations reported significant crime reduction effects associated with this problem-oriented approach (Braga and Weisburd 2012). A meta-analysis of these programs reported that focused deterrence strategies were associated with an overall statistically-significant, medium-sized crime reduction effect. This review provides additional evidence that the general problem-oriented policing approach can inform innovative violence reduction strategies and generate impressive crime control gains.

Conclusion

Problem-oriented policing represents an important innovation in American policing. Indeed, the advocates of this young and evolving approach have accomplished much since Herman Goldstein first presented the concept in 1979. Early experiences in Madison, London, Baltimore County, and Newport News demonstrated that police officers could greatly improve their handling of crime problems by taking a problem-oriented approach. Since then, many police agencies in the United States, the United Kingdom, Canada, Scandinavia, Australia, and New Zealand have continued to implement problem-oriented policing, to apply it to a wide range of crime and disorder problems, and to change their organizations to better support problem-oriented policing (www.popcenter.org).

The practice of problem-oriented policing sometimes falls short of the principles suggested by Herman Goldstein (1979, 1990). While the approach is more than 30 years old, it is important to recognize that problem-oriented policing is still in its formative stages and its practice is still developing. Progress in policing is incremental and slow, and that does not make problem-oriented policing unrealistic. Police departments should strive to implement problem-oriented policing properly but recognize that even weak problem solving can be beneficial when applied to recurring crime problems. Within police departments, it seems like a balanced problem-oriented policing agenda would include both a commitment among officers in the field to apply problem-solving techniques to address problems that they encounter on a routine basis and a commitment to maintaining a centralized problem-oriented policing unit capable of conducting high-quality analyses of larger and more persistent problems and developing more creative responses to reduce them.

In closing, it also seems important to point out that the demonstrated crime reduction efficacy of problem-oriented policing is a striking result considering the large body of research that shows the ineffectiveness of many police crime prevention efforts. The robustness of the problem-oriented policing is underscored by the observation that, even when it is not implemented properly, the approach still generates noteworthy crime reduction gains (Braga and Weisburd 2006). It is tantalizing to think that had the police more fully implemented the problem-oriented approach and took a more specific, more focused approach to crime and disorder problems, crime control benefits might have been greater. Of course, this requires the development of such skills from both "trial and error" experience of problem solving on the street and additional training in the problem-oriented model, particularly in the area of problem analysis. The investment in the acquisition of these skills could be well worth the effort.

Related Entries

- ▶ Community Policing
- ▶ Compstat
- ▶ Evidence-Based Policing
- ▶ Hot Spots and Place-Based Policing
- ▶ Intelligence-Led Policing
- ▶ Order Maintenance Policing
- ▶ Predictive Policing
- ▶ Situational Crime Prevention
- ▶ Third Party Policing and School Truancy

Recommended Reading and References

Bayley D (1994) Police for the future. Oxford University Press, New York

Braga AA (2008) Problem-oriented policing and crime prevention, 2nd edn. Lynne Rienner Publishers, Boulder

Braga AA, Weisburd DL (2006) Problem-oriented policing: the disconnect between principles and practice. In: Weisburd DL, Braga AA (eds) Police innovation: contrasting perspectives. Cambridge University Press, New York

Braga AA, Weisburd DL (2010) Policing problem places: crime hot spots and effective prevention. Oxford University Press, New York

Braga AA, Kennedy DM, Waring EJ, Piehl AM (2001) Problem-oriented policing, deterrence, and youth violence: an evaluation of Boston's operation ceasefire. J Res Crime Delinq 38:195–225

Braga AA, Papachristos AV, Hureau DM (2012) Hot spots policing effects on crime. Campbell Systematic Reviews 2012:8. doi:10.4073/csr.2012.8

Braga AA, Weisburd DL (2012). The effects of focused deterrence strategies on crime: a systematic review and meta-analysis of the empirical evidence. J Res Crime Delinq 49:323–358

Capowich G, Roehl J (1994) Problem-oriented policing: actions and effectiveness in San Diego. In: Rosenbaum D (ed) The challenge of community policing: testing the promises. Sage Publications, Thousand Oaks

Clarke RV (1995) Situational crime prevention. In: Tonry M, Farrington D (eds) Building a safer society: strategic approaches to crime prevention. University of Chicago Press, Chicago

Clarke RV (ed) (1997) Situational crime prevention: successful case studies, 2nd edn. Harrow and Heston, Albany

Clarke RV (1998) Defining police strategies: problem solving, problem-oriented policing and community-oriented policing. In: O'Connor T, Grant AC (eds) Problem-oriented policing: crime-specific problems, critical issues, and making POP work. Police Executive Research Forum, Washington, DC

Clarke RV, Goldstein H (2002) Reducing theft at construction sites: lessons from a problem-oriented project. In: Tilley N (ed) Analysis for crime prevention. Criminal Justice Press, Monsey

Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44:588–605

Cordner G (1998) Problem-oriented policing vs. zero tolerance. In: O'Connor T, Grant AC (eds) Problem-oriented policing: crime-specific problems, critical issues, and making POP work. Police Executive Research Forum, Washington, DC

Eck JE (2000) Problem-oriented policing and its problems: the means over ends syndrome strikes back and the return of the problem-solver. Unpublished manuscript. University of Cincinnati, Cincinnati

Eck JE, Spelman W (1987) Problem-solving: problem-oriented policing in Newport news. National Institute of Justice, Washington, DC

Goldstein H (1979) Improving policing: a problem-oriented approach. Crime Delinq 25:236–258

Goldstein H (1990) Problem-oriented policing. Temple University Press, Philadelphia

Gottfredson M, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Kennedy DM (2008) Deterrence and crime prevention. Routledge, London and New York

Kennedy DM, Moore MH (1995) Underwriting the risky investment in community policing: what social science should be doing to evaluate community policing. Justice System Journal 17:271–290

Kennedy DM, Piehl AM, Braga AA (1996) Youth violence in Boston: gun markets, serious youth offenders, and a use-reduction strategy. Law Contemp Probl 59:147–197

Moore MH (1992) Problem-solving and community policing. In: Tonry M, Morris N (eds) Modern policing. University of Chicago Press, Chicago

Poyner B, Webb B (1991) Crime free housing. Butterworth-Architecture, Oxford

Read T, Tilley N (2000) Not rocket science? Problem-solving and crime reduction. Crime Reduction Series Paper 6. Policing and Crime Reduction Unit, Home Office, London

Scott M (2000) Problem-oriented policing: reflections on the first 20 years. Office of Community Oriented Policing Services, U.S. Department of Justice, Washington, DC

Scott M, Clarke RV (2000) A review of submission for the Herman Goldstein excellence in problem-oriented policing. In: Brito CS, Gratto E (eds) Problem oriented policing: crime-specific problems, critical issues, and making POP work, vol 3. Police Executive Research Forum, Washington, DC

SkoganW, Frydl K (eds.) (2004) Fairness and effectiveness in policing: the evidence. Committee to Review Research on Police Policy and Practices. The National Academies Press, Washington, DC

Weisburd DL, Eck JE (2004) What can police do to reduce crime, disorder, and fear? Ann Am Acad Pol Soc Sci 593:42–65

Weisburd DL, Telep CW, Hinkle JC, Eck JE (2010) Is problem-oriented policing effective in reducing crime and disorder? Findings from a Campbell systematic review. Criminology & Public Policy 9:139–172

www.campbellcollaboration.org

www.popcenter.org

Problem-Solving Courts

David B. Rottman[1] and Jordan Bowman[2]
[1]National Center for State Courts, Williamsburg, VA, USA
[2]William & Mary School of Law, Williamsburg, VA, USA

Synonyms

Alternative courts; Problem-oriented courts; Specialty courts; Therapeutic courts

Overview

The label "problem-solving courts" refers to a movement of court specialization that began in 1989 and subsequently dominated criminal court reform discussions and agendas in the United States. The label encompasses a legal theory of adjudication and a set of practices that have been applied in pursuit of better individual and societal-level outcomes in court cases involving offenders with underlying psychological, public health, and social problems perceived as hindering a desistance from crime. Problem-solving courts began by addressing the problem of substance abuse. In turn, the problem-solving court movement expanded to address mental health, domestic violence, and, most recently, the difficulties military veterans are experiencing in adjusting to civilian society. Other examples include the creation of reentry courts (to support offenders released from prison due to prison overcrowding or cost-saving programs of early release) and community courts (responding to persistent low-level offenses that are seen as harmful to the quality of life of residents and businesses). Problem-solving courts have also been created to deal with problems associated with guns, drunk driving, tobacco, and noise abatement (environmental courts).

By the end of 2009, there were an estimated 3,649 operational problem-solving courts, two-thirds of which were drug courts (Huddleston and Marlowe 2011:1). This movement began with the opening of a single court, the Miami (Florida) Drug Court. The early problem-solving courts were in many respects judicial innovations, the creations of a few judges who were dissatisfied with what they could accomplish by adhering to the traditional judicial role and using conventional sentencing options. Wider interest from state and federal policy-makers followed almost immediately. The rapid diffusion of problem-solving courts and the fervor of its advocates can be compared in some ways to the launch of the juvenile court movement 100 years prior. It is only recently, and only definitively for adult drug courts, that research has confirmed that the impacts such as recidivism in problem-solving courts improve upon what is typical in misdemeanor courts.

The problem-solving concept has also been well received outside of the United States. A diverse range of countries – including Australia, Canada, Great Britain, the Netherlands, and South Africa – embraced the concept. By the most recent available international count (for 2007), drug courts are the most prevalent (in eight countries), followed by community courts (in seven countries), domestic violence courts (in four countries), and mental health courts only in Australia (Nolan 2009). The adoption of American problem-solving court principles has not been uncritical; modifications have been made to reflect local culture. Therefore, "in the United States, the courts are characterized by enthusiasm, boldness and pragmatism; whereas in the other five countries, [cited above] the contrasting disposition is one of moderation, deliberation, and restraint" (Nolan 2012:156).

Definition

Properly understood, problem-solving courts consist of court dockets, a courtroom, and a judge that, for a period of time ranging from a half-day per week to a 5 full days a week, exclusively adjudicates cases that have been selected for special processing. There are only a handful of stand-alone drug courts or any other type of problem-solving court. This distinction is not always recognized in discussions of problem-solving courts. In some respects, problem-solving courts are a continuation of one of the oldest approaches to court reform, caseflow management, in which types of cases are grouped together in order to secure more efficient and effective case processing.

There is no authoritative definition of what constitutes a problem-solving court. Most would agree, however, with this formulation: a problem-solving court must demonstrate "(1) close and ongoing judicial monitoring, (2) a multidisciplinary or team-oriented approach, (3) a therapeutic or treatment orientation, (4) the altering of traditional roles in the adjudication process, and (5) an emphasis on solving the problems of individual offenders" (Nolan 2010:32).

Problem-solving courts can be distinguished from the types of courts and dockets that emerged in earlier periods to specialize adjudication in criminal court cases. The distinctiveness stems, first, from a package of practices and beliefs on how to wean substance abusers and others from persistent offending associated with an underlying psychological or physical problem and, second, from a new legal theory that justifies their practices. The difference is one of degree, but it is consequential. By a similar logic, some commentators exclude domestic violence courts from the ranks of problem-solving courts: there is little expectation that the actions of the court can reduce recidivism in such cases. The distinction is well captured in the standard court terminology which refers to adults as defendants or offenders, while in a problem-solving court, we hear talk of "clients," "participants," and "graduations," rather than defendants, offenders, and sentences.

Origins

Today's problem-solving courts are products of several trends that came together in the late 1980s. Those trends included reformist judges, a legal theory supporting a more expansive judicial role, enthusiastic government support and funding, and the emergence of a new profession.

First, a significant minority of United States judges were frustrated by their inability to intervene in such a manner as to reduce persistent re-offending by addressing the underlying psychological or social problems that prevented progress. Substance abuse was the primary cause of dissatisfaction. Nothing worked. The same judges often were dissatisfied with the limitations imposed by the traditional role of the judge and looked for a way of defining the judicial role that they would find more personally and professionally satisfying (Baum 2011:116). In key respects, problem-solving courts in the United States are products of the judiciary itself: they represent the most recent effort to create specialized courts to achieve faster and more effective responses to particular types of cases. Outside of the United States, the introduction of problem-solving courts tends to be the product of government initiatives, working with the courts.

Second, there was a legal theory available to justify deviation from the judicial role as defined by the adversarial process. The emerging theory came from the field of therapeutic jurisprudence developed by United States law professors David Wexler and Bruce Winnick. Therapeutic jurisprudence asks "whether the law's antitherapeutic consequences can be reduced, and its therapeutic consequences enhanced, without subordinating due process and other justice values" (Wexler and Winnick 1996:xvii); see also (Hora et al. 1999).

The concept of therapeutic jurisprudence quickly attracted a following among judges in other common law countries, easing the diffusion of problem-solving courts to those jurisdictions.

Third, the problem-solving court movement had important advocates in high places. With the vocal support came access to government funding and sponsorship. The first problem-solving court, the 1989 Dade County Drug Court, was a joint project by then Dade County prosecutor Janet Reno and Judge Herbert Klein. When Reno became attorney general, drug courts became incorporated into Department of Justice policy. Community courts followed as an early addition to the Department of Justice's programs supporting drug courts, with funding made available to support visits to what were designated as "demonstration projects." Thereafter, all federal budgets have made significant funding available for drug courts and subsequently other types of problem-solving courts as they arose. Additionally, problem-solving courts were not entering a crowded field with powerful interests likely to obstruct a newcomer promising to contend with problems like substance abuse. There was a general recognition that previous approaches to the problem had failed.

Finally, problem-solving courts were nurtured by a very active professional association, the National Association of Drug Court Professionals (NADCP), which was founded in 1994. NADCP's annual conferences bring together judges, social workers, and probation officers but also representatives from federal and state government agencies, judicial organizations, US senators and congressmen, rock stars, and movie stars. NADCP also has a research arm, the National Drug Court Institute. The NADCP has taken the lead in developing ten key components for drug courts, a formulation other problem-solving courts have modified. In recent years, the NADCP has also taken a leadership role in determining what configurations make for effective drug courts and developing standards a drug court must meet. As it matured, NADCP has proved willing to abandon some of its original assumptions and policy prescriptions through the application of evidence-based practices. The NADCP has increasingly broadened its focus to include other types of problem-solving courts and judges, court staff, and public officials from other countries.

Types of Problem-Solving Courts

The seven most common types of problem-solving courts are distinctive applications of problem-solving legal theory and practice. Listed by the order in which they emerged, these are adult drug treatment courts, juvenile drug courts, community courts, domestic violence courts, mental health courts, reentry courts, and veterans' courts. Short descriptions of their origins and distinctive features supplement longer presentations elsewhere in the encyclopedia. A guide to the key elements and the main dimensions of the most common types of problem-solving courts, along with a description of how they vary, is available (Casey and Rottman 2005). It should be noted that in Australia and several other countries, indigenous sentencing courts were established to accommodate traditional non-adversarial models of justice, resulting in courts that have clear problem-solving court features (King et al. 2009:178–183).

Drug Courts

In response to an increase in case filings in the 1980s and 1990s, drug courts began as diversion and treatment programs designed to get at the "root cause" of drug abuse and decrease recidivism. Today, drug courts integrate alcohol and drug treatment services as offenders are processed through the justice system. Drug courts emphasize a non-adversarial approach, where prosecution and defense counsel collaborate on treatment options while protecting participants' due process rights. Offenders are subject to frequent alcohol and drug testing and have access to rehabilitation services and other related treatment. Frequently, drug courts emphasize ongoing judicial interaction, where offenders meet regularly with a judge for updates on progress. Interaction with the community is also an important aspect of drug courts, as they frequently establish partnerships with agencies or community-based organizations in order to increase effectiveness. Over 2,600 drug courts are in operation in the United States today; most (58 %) follow a model in which a defendant must first plead guilty before being admitted to the program (Huddleston and Marlowe 2011:1). One 2005 estimate posited that 53,000 defendants/offenders were enrolled in drug courts. To put that presence in perspective, estimates show that drug courts are serving about one-half of offenders meeting current criteria for enrollment but serving only about 3.8 % of all offenders who use or are addicted to illegal substances (Bhati et al. 2008:34). Drug courts have themselves become specialized, notably through the establishment of juvenile and family drug courts.

Community Courts

Community courts are, in part, a reaction to the centralization and consolidation of trial courts, which was perceived as creating a gap between courts and communities. They were also a response to the prominence of the "broken windows" approach to responding to low-level crime in urban areas. New York City's Midtown Community Court opened its doors 1993. As of 2009, there were 38 community courts in cities across the United States. There are nearly as many community courts outside of the United States, located in Australia, England, Scotland, and South Africa (Turgeon 2011). England and South Africa in particular have embraced the concept. In Australia, only one community court (named a Neighborhood Justice Center) has been established. Community courts focus on replacing jail and fines with community service and alternative sanctions. Offenders in community court generally receive more attention than do those in the general court system. An increase in court time and resources devoted to minor misdemeanors, coupled with extensive gathering of information about defendants and strict monitoring of compliance, is designed to reduce recidivism. Offenders in community courts have access to a comprehensive package of social services and treatments through a mix of government and nonprofit agencies. Ultimately, community courts are committed to a dual approach: changing the lives of individuals and improving the quality of life in communities. As the subject of four comprehensive evaluations, community courts are among the most studied type of problem-solving court.

Mental Health Courts

Mental health courts began as a response to both high recidivism and high rates of homelessness, unemployment, alcohol and drug use, and physical and sexual abuse among persons with a mental illness. The first mental health court was founded in 1997 in Broward County, Florida. Like the over 288 other mental health courts in the United States, it is centered on diverting defendants with a mental illness into treatment, instead of simply incarcerating them (Huddleston and Marlowe 2011:40). Mental health courts emphasize voluntary participation, early identification and intervention, and a therapeutic environment to reduce the trauma that is often experienced by persons with a mental illness in the criminal justice system. Part of this approach includes implementing practices that reduce the stigma that often accompanies mental illness. Like drug courts, mental health courts use a dedicated team approach that involves the judge, counsel, and court and treatment professionals in a less-formal setting. A treatment plan will be developed that is centered on the individual's specific needs, with regular status hearings and assessments. Mental health courts have not as proved popular in other countries compared to other types of problem-solving courts. In Australia, for example, the primary objective of government policy is to avoid involving mentally ill offenders with the criminal justice system (King et al. 2009:150).

Domestic Violence Courts

Created in order to promote victim safety and make offenders more accountable, domestic violence courts typically emphasize a collaborative approach through cooperation with community-based organizations. Domestic violence courts typically have a dedicated judge and staff who receive ongoing training in domestic violence dynamics. Victims in domestic violence courts typically receive specialized intake services and receive early access to advocacy, while efforts are made to ensure that defendants understand court conditions. Domestic violence courts in the United States also closely monitor compliance pre- and post-disposition, a practice not followed in other countries such as Australia (King et al. 2009: 156). The first domestic violence court was established in Dade County (Miami), Florida, in 1992. Public funding for domestic violence courts became available in 2000, and there are now more than 200 such courts nationally (Huddleston and Marlowe 2011:40). Many observers, however, doubt the applicability of the label "problem-solving" to these specialized dockets because there is no consensus on what, if any, treatment can predictably intervene successfully in the cycle of abuse. Domestic violence courts, instead, focus on providing protection and services to the victims of family violence rather than the treatment of the offender.

Reentry Courts

In 2000, the US Department of Justice, Office of Justice Programs, introduced a Reentry Court Initiative (RCI) to address the rise in incarceration during the previous decades. Specifically, the RCI adapted the drug court model to create community-based programs to oversee the transition from prison to parole supervision and to "establish a seamless system of offender accountability and support services throughout the reentry process" (Lindquist et al. 2003). Nine pilot programs were established. Their common elements were assessment and planning, active oversight, accountability to community, management of support services, graduated and parsimonious sanctions, and incentives for success. A reentry court's jurisdiction over offenders can begin as early as the sentencing process, and in some cases offenders remain in a reentry program until post-release supervision is terminated. The first reentry courts became active in 2000. In 2011, 26 reentry courts and 29 reentry drug courts – reentry courts that are combined with drug courts – were in operation within the United States (Huddleston and Marlowe 2011). Although a few program evaluations have been conducted, definitive evidence that reentry courts reduce recidivism or are cost-effective is not available.

Veterans' Courts

A more recent addition, veterans' courts are a response to the frequency of veterans who return from service with mental illness and/or substance addictions. The first veterans' treatment court began in Buffalo, New York, in 2008. This court has been followed by more than 80 of its kind (Shevroya 2011; Center for Court Innovation 2011). Veterans' courts are an amalgam of practices from drug courts and mental health courts. Similar to the process in mental health courts, offenders who are veterans are identified early in judicial proceedings, and participation in a veterans' court is voluntary. Veterans' courts integrate alcohol, drug treatment, and mental health services with justice system case processing. A comprehensive treatment plan is decided on through a non-adversarial approach in which the prosecution and defense work together to promote public safety while preserving participants' due process rights. Abstinence is monitored by frequent alcohol and drug testing, and judges continue to interact with offenders throughout the program. Veterans' courts frequently monitor and evaluate effectiveness, and staff often engage in interdisciplinary education in order to achieve maximum effectiveness. Like drug courts, veterans' courts form partnerships with public agencies such as the Veterans Administration and community-based organizations that support veterans.

Research Evidence on the Effectiveness of Problem-Solving Courts

A large body of research evidence, much of it using quasi-experimental designs, demonstrates the ability of problem-solving courts to generate better, more cost-effective outcomes than traditional courts. There are several notable features of this body of evidence. First, nearly all of the research has been carried out on adult drug courts, with only a small number of studies available for community courts and mental health courts. We can generalize from what we know about drug courts to the universe of problem-solving courts with only modest expectations of accuracy. Second, as one observer notes, "scholarly and quasi-scholarly writing on problem-solving courts is heavily tinged with advocacy, especially by their proponents" (Baum 2011:27). For drug courts, and to some extent for community courts, the extant research can address two questions. Do problem-solving courts work? And, if they work, why (how?) do they work?

Do They Work as Intended?

This summary of the available research does not seek to be comprehensive. Other entries in this encyclopedia give comprehensive descriptions of what the available evidence says about the efficacy of specific types of problem-solving courts.

Initially, research on problem-solving courts was inconclusive. The rapid expansion in the number of problem-solving courts took place before definitive research was available to demonstrate their efficacy. Early research on problem-solving courts tended to be single court studies. By 2005, sufficient evidence had accumulated for proponents of problem-solving courts to report significant success in the form of reduced recidivism and compliance with court requirements. The state of the evidence as of 2005 was the following: "There have been more than 100 research studies about adult drug courts, and if you look at the best, most rigorous 25 of those, you probably come to the conclusion that drug courts reduce criminal offending by 15–20 %. And what that means is that if you have 30 % of your population you would expect to see rearrested in the next year or so, then instead they go through drug court, it's more like 24 %" (Bathi et al. 2005).

A 2012 comprehensive evaluation of the Red Hook Community Justice Center compared recidivism rates at that court and its downtown trial court counterpart. The evaluation used quasi-experimental methods and matching using propensity scores to conduct a survival analysis following offenders over a 3-year period (Lee et al. 2013). Recidivism rates at Red Hook were 10 % less than those found at the downtown court.

Attention is also placed on the costs and benefits that accrue from problem-solving courts relative to traditional misdemeanor courts. The basic finding here is that problem-solving courts are more costly to operate, but from the point of view of the taxpayer, they are cost-effective compared to the processing of such cases in regular courts once reduced recidivism and savings in the health-care system are taken into account. One estimate (Bathi et al. 2008:xv) identifies an annual cost of $515 million to operate drug courts, while the reduction in recidivism results in more than $1 billion in savings.

How and Why Do They Work?

The initial compelling evidence of positive impacts did not lead to a clear understanding of which drug court features delivered the best results. The studies were also unable to answer the question of *why* drug treatment courts generate better outcomes. Some possible factors have not been significant in studies of drug courts and community courts. The unexpected finding from research to date is that neither the treatment provided nor any deterrent effect could be identified as a cause for the reduction in recidivism rates.

The explanation appears instead to lie in the nature of the interaction between the judge and the offender that occurs in the context of a problem-solving court. The difference in interaction styles may be attributable both to the type of judges who seek assignment to a problem-solving court or to the way in which problem-solving courts structure that interaction. The latter explanation is more plausible as an individual-level explanation seems unlikely to produce the pattern of findings in the available research.

A study of the Baltimore Drug Treatment court provided some of the initial evidence pointing to an "interaction" explanation. The researchers concluded that "the DTC [Drug Treatment Court] program, especially the judicial hearings, contributes to an offender's perception of fairness and due process, thereby increasing his or her willingness to fulfill his or her part of the negotiated DTC agreement" (Gottfredson et al. 2007:28). This was consistent with the perspective of procedural fairness. Procedural fairness is a theoretical approach to explaining satisfaction and compliance with decision-makers, buttressed by a very large supportive body of research. It is most closely associated with social psychologist Tom Tyler. Procedural fairness is present when people perceive that they are experiencing a decision-maker who treats them with respect, neutrality, allows participation, and is trustworthy (Tyler 2004:443–447). Some decision-makers and decision-making forums are perceived as following a more fair process than others.

Judge-to-offender interaction was highlighted as the advantage inherent in problem-solving courts by National Institute of Justice's *Multi-Site Adult Drug Court Evaluation* (Rossman et al. 2011). This study found that drug court participants across 23 sites were more likely than a matched comparison sample to perceive their treatment by the judge as fair; the study then found that these more positive perceptions of the judge comprised an influential factor in explaining why the drug court sample was less likely than the comparison sample either to commit further crimes or to use drugs during the follow-up period (Rossman et al. 2011).

Findings from evaluations of community courts also point to a procedural fairness explanation. A study of defendants at the Red Hook (Brooklyn, New York City) Community Justice Center found that defendant perceptions of procedural fairness were stronger at RHCJC than among a comparison group of defendants processed in a centralized court (Frazer 2007). The study found that Red Hook defendants were more likely to perceive the court process as fair, but the proportion of defendants perceiving the centralized court as fair was also high. However, at Red Hook, positive perceptions of the judge appeared to influence overall perceptions of the court process. In another community court study, it was noted that "in the community court, the judge spoke directly to the defendant in 45 % of the observed appearances, while in the traditional court this occurred in only 19 % of appearances" (Frazer 2007:22).

A more recent Red Hook Community Justice Center evaluation reinforces the focus on procedural fairness. The reduction in misdemeanor defendants' likelihood of recidivism was not attributable to deterrence effects, treatment and social services, or to connections forged with the local community. Based on interviews conducted as part of the ethnographic component and other research, procedural fairness appears to be the mechanism that makes a difference (Lee et al. 2013).

The core finding is that the advantage problem-solving court processing has over traditional low-level criminal court processing is due to the judge – or, more specifically, to the nature of the interaction between the judge and offender in the courtroom.

Controversies

Critics and proponents of problem-solving courts have displayed comparable levels of fervor in making their arguments. The most influential and persistent critic from the academic work has been James Nolan (Nolan 2001, 2009, 2010, 2012). Within the judicial world, Denver trial judge Morris Hoffman has also engaged in sustained criticism (Hoffman 2002). Among national organizations, the National Association of Criminal Defense Lawyers has consistently criticized the ethical status of problem-solving courts (National Association of Criminal Defense Lawyers 2009). Proponents of problem-solving courts are too numerous for it to be sensible to single out a few individuals. It is easier to focus on organizations, noting the large volume of material produced by the Center for Court Innovation, National Drug Court Institute, and NPC Research. Both sides are well represented in a special issue of the *Fordham Urban Law Journal* that appeared in 2002.

Do Problem-Solving Courts Extend the Reach of Government?

A common controversy surrounding problem-solving courts is whether they are a social control mechanism that, whether by design or not, increases the reach of the government into peoples' lives. There are several dimensions to this concern. One criticism is that drug courts wrongfully politicize health and addiction, becoming wrongfully engaged in the process of a "management of pleasures" (Mackenzie 2008). Another dimension is that of "widening the net so people are prosecuted when they otherwise wouldn't be" (Miller 2004; Hoffman 2002). Others argue that, while there is no evidence that net widening is taking place, net widening would be a good thing because it would result in people getting the treatment that they need (Berman and Gulick 2003).

Yet another dimension is the tendency of problem-solving courts to keep offenders under court supervision for longer periods of time than if they had been processed through a traditional misdemeanor court. Most misdemeanor defendants receive short jail sentences, fines, or "time served." In most circumstances, they leave the court with little or no supervision. Offenders processed in problem-solving courts spend longer periods of time under court supervision and are exposed to enhanced risks of failure because the court's expectations are more demanding.

Do Problem-Solving Courts Weaken the Adversarial Process?

To some United States critics, problem-solving courts relax the adversarial process to the extent that they are no longer courts. To one critic, "They are not concerned with due process or adjudicating guilt, either by plea or trial. They are correctional agencies and as such might be an excellent new model of correctional agency. But as post-adjudicative bodies, they must be called what they really are, and it is *not* courts" (McCoy 2006).

The majority of drug courts today are indeed post-plea. But whether pre- or post-plea, problem-solving courts relax certain aspects of the traditional role of a judge in deciding a case. There are also some shifts in the balance of power in the courtroom. In a problem-solving court, the judge is the best-informed person in the courtroom about an offender and supplants the power traditionally held by the prosecutor in a plea bargaining context.

Critics also find fault with various aspects of problem-solving courts that raise ethical issues for judges and attorneys. First, being processed through a problem-solving court exposes a defendant to a greater risk of failure because of the strict criteria for success. Failure may be met with a punishment more severe than if they had opted for a traditional misdemeanor court for processing. Second, problem-solving courts accumulate far more information on defendants that would be the case in a traditional court, and that information is shared among a wide group of stakeholders. Third, entering a problem-solving court generally requires that the defendant waive the right to discovery. Fourth, the decision to enter a drug court must be made within a shorter time frame than is usually available before a trial in order to get the offender immediately into treatment, potentially preventing counsel from fully investigating the case. Fifth, the use of teams in the decision-making process can often lead to ex parte communication about a case in the absence of the defendant or their defense counsel. Sixth, defending a client in a problem-solving court imposes additional workload burdens on public defenders. In many problem-solving courts, there are an increased number of post-plea hearings and "team" meetings in which consequential decisions about problem-solving court participants are made. In order to properly represent their client, defense counsel must not just be present, but be prepared and involved during these meetings (National Association of Criminal Defense Lawyers 2009:31–32).

A major criticism of problem-solving courts (really, drug courts) is that a defendant is denied zealous advocacy. It has been claimed that problem-solving courts unduly limit the time that defense counsel has to investigate or prepare a defense because in the United States but not elsewhere, they are moving in a post-plea direction (Finigan et al. 2007). The ongoing nature of cases may create an extra burden for public defenders, thus not allowing them to adequately represent their clients. The participant in a problem-solving court is under pressure to be "team player" and advance team goals, yet the participant's short-term interest and drug court team's goals may diverge (Meekins 2007:108). There is also concern that firmly asserting a client's rights in a team environment could lead to economic repercussions for the lawyer who may become less likely to be appointed to future cases in that court (Meekins 2007:110).

In order to be balanced, claims that problem-solving courts are not adversarial should be considered in reference to the often feeble nature of the adversarial process as practiced in traditional misdemeanor courts based on plea bargaining (Hessick and Saujani 2002). One important difference may be that in traditional courts the pressure to move forward with the case comes from the prosecutor in the form of decisions to pursue charges and other trial/violation considerations, whereas in a problem-solving court, the most powerful player in the courtroom is invariably the judge. In a problem-solving court, judges make use of their authority to promote compliance with court orders while the prosecutor helps to address participants' problems. Also, it can be argued that in the post-adjudication phase of monitoring, the definition of appropriate advocacy may differ from that during adjudication. Nonetheless, concerns persist concerning the degree to which the very nature of problem-solving courts raises genuine ethical issues for judges, defenders, and prosecutors (Leben et al. 2013). Concerns over the possible deviations from a strict adversarial model have not proved as problematic outside of the United States.

Who Belongs in a Problem-Solving Court?

Another controversy is the issue of who belongs in a problem-solving court. Early drug courts and government policy reserved space for offenders with relatively minor involvement in drug use – users but not addicts. By 2010, the consensus among researchers and policy-makers is that drug courts should be reserved for the more serious offenders, as indexed by their prior criminal history, involvement in substance abuse, and mental health. Policy-makers in other countries such as Australia focused from the start on more serious offences and addicts, perhaps because of the existing pre- and post-plea dispositions available for minor offenders.

The new consensus as led by the National Drug Court Institute is that drug courts should be focused on offenders who "are dependent on alcohol or other drugs and are also at risk for failure in standard correctional rehabilitation programs" (Marlowe 2009). Under this approach, admitting low-level offenders to drug courts is to be discouraged. Individuals who are not substance dependent, but who are substance users and have a low risk for failure in standard settings, are unsuitable for drug courts. Participation in drug courts could expose this population to substance abusers and antisocial values while pulling them away from school or work (Marlowe 2009).

What Should Be a Problem-Solving Court's Goal?

There is a degree of American exceptionalism in the world of problem-solving courts. Nolan argues that while in the United States problem-solving courts seek "total abstinence" or "demand reduction," other common law countries have a philosophy of "harm reduction" or "harm minimization" (Nolan 2010). This explains, in part, the general aversion of drug courts in the United States to making use of medication-assisted treatment. If the goal is to eliminate the harm, there is likely to be opposition to promoting reliance on any type of medication, even if the purpose is to ease or speed the withdrawal process for substance users upon entry into the program, inhibit cravings to use illicit substances or the effects of such use.

Conclusion: The Future of Problem-Solving Courts

Overwhelmingly, what we know about problem-solving courts is based on studies of adult drug courts in the United States. This limits our ability to draw conclusions about problem-solving courts in general, or even about the future for juvenile drug courts. Much depends on whether evaluations of mental health courts, veterans' courts, and community courts are as positive in their findings as the research on drug courts. Even now, however, there are a few indicators of what is to come.

For one thing, it is likely that in the near future, problem-solving courts will start to look different than the ones in the current landscape. The emphasis on evidence-based practices is steering drug courts into new directions. Other problem-solving courts are likely to follow suit. Long-standing assumptions on the types of offenders who should be processed through a problem-solving court are being undone by research evidence and debate over whether harm abstinence or minimization should be the goal of the court. The National Association of Drug Court Professionals is developing standards all drug courts will be expected to follow. There has even been talk of accreditation. The willingness to change the nature of problem-solving courts is nearly as impressive as the sheer doubling and tripling of their numbers experienced in recent decades.

There is also uncertainty on whether government policy in the United States will continue to foster more problem-solving courts and apply them to more "problems." In one scenario, problem-solving courts will become the norm as the mainstream criminal courts are encouraged to incorporate problem-solving theory and practices. Problem-solving courts might then simply dissolve into the larger court system. In another scenario, the large-scale underwriting of the problem-solving concept by state and federal governments might slow to a point where problem-solving courts can no longer be sustained.

Outside of the United States, the prospects also are mixed. Only in the United States have problem-solving courts significantly altered the landscape of how lower-level criminal cases are processed. Such courts are not proliferating even in the common law countries that adopted them. To one observer, a critic of problem-solving courts but also the author of the most significant comparative work on them, this is to be expected: "To import problem-solving courts is to import elements of the particular culture out of which the programs first emerged–and from which they are not easily extricated" (Nolan 2012:164). This might be an unduly negative reading of the potential for the spread of problem-solving courts internationally. After all, the problem-solving courts in Australia and in other countries have adapted the basic model to fit their distinctive legal and other circumstances. Today's problem-solving courts, on balance, seem likely to remain influential features in criminal justice policy and practice in the United States and elsewhere.

Related Entries

- ▶ Community Courts
- ▶ Drug Courts
- ▶ Drug Courts' Effects on Criminal Offending
- ▶ Evolving Judicial Roles
- ▶ Judicial Leadership and Performance
- ▶ Mental Health Courts
- ▶ Reentry Courts
- ▶ Therapeutic Jurisprudence

Recommended Reading and References

Baum L (2011) Specializing the courts. University of Chicago Press, Chicago

Belenko S (2001) Research on drug courts: a critical review – 2001 update. National center on addiction and substance abuse. Columbia University, New York

Berman G, Gulick A (2003) Just the (unwieldy, hard to gather, but nonetheless essential) facts, ma'am: what we know and don't know about problem-solving courts. Fordham Urban Law Rev 30(3):1027–1054

Bhati AS, Roman JK, Chalfin A (2008) To treat or not to treat: evidence on the prospects of expanding treatment to drug-involved offenders. Urban Institute, Washington, DC

Boldt RC (1998) Rehabilitative punishment and the drug court movement. Washington Univ Law Quart 76:1205–1306

Burns SL (2010) The future of problem-solving courts: inside the courts and beyond. Univ Maryland Law J Race Relig Gender Class 10:73–87

Casey P, Rottman D (2005) Problem-solving courts: models and trends. Just Syst J 26(1):35–56

Center for Court Innovation (2011) What is a community court? Author, Washington, DC

Dorf MC, Sabel CF (2000) Drug treatment courts and emergent experimentalist government. Vanderbilt Law Rev 53(3):831–883

Feeley MM (1992) The process is the punishment: handling cases in a lower criminal court. Russell Sage, New York

Finigan MW, Carey SM, Cox A (2007) Impact of a mature drug court over 10 years of operation: recidivism and costs. NPC Research, Portland

Frazer MS (2007) Defendant perceptions of fairness at the Red Hook Community Justice Center. Center for Court Innovation, New York

Goldkamp JS, White MD, Robinson JB (2001) Do drug courts work? Getting inside the drug court black box. J Drug Iss 31:27–72

Gottfredson DC, Kearley BW, Rocha CM (2007) How drug treatment courts work: an analysis of mediators. J Res Crime Delinquency 44:33–35

Hessick AF, Saujani R (2002) Plea bargaining and convicting the innocent: the role of the prosecutor, the defense counsel and the judge. BYU J Public Law 16:189

Hoffman MB (2002) Therapeutic jurisprudence, neo-rehabilitations, and judicial collectivism: the least dangerous branch becomes most dangerous. Fordham Urban Law J 29(5):2063–2098

Hora PF, Stalcup T (2008) Drug treatment courts in the twenty-first century: the evolution of the revolution in problem-solving courts. Georgia Law Rev 42:717–811

Hora P, Schma WG, Rosenthal JTA (1999) Therapeutic jurisprudence and the drug treatment court movement: revolutionizing the criminal justice system's response to drug abuse and crime in America. Notre Dame Law Rev 74(2):439–538

Huddleston W, Marlowe DB (2011) Painting the current picture. National Drug Court Institute, Washington, DC

Jensen EL, Mosher C (2006) Adult drug courts: emergence, growth, outcome evaluations, and the need for a continuum of care. Idaho Law Rev 42:443

King M, Freiberg A, Batagol B, Hyams R (2009) Non-adversarial justice. The Federation Press, Sydney

Klein R (2004) Due process denied: judicial coercion in the plea bargaining process. Hofstra Law Rev 32:1349–1423

Leben S, Rottman D, Casey P (2013) Legal and ethical considerations. In: Koetzle D, Guastaferro W (eds) Drug courts and the criminal justice system. Lynne Rienner, Boulder (in press)

Lee C, Cheesman F, Rottman D, Swaner R, Lambson S, Rempel M, Curtis R (2013) A comprehensive evaluation of the red hook community justice center. National Center for State Courts, Williamsburg

Lindquist C, Hardison J, Lattimore P (2003) Reentry courts process evaluation, final report. National Institute of Justice, Washington, DC

Mackenzie R (2008) Feeling good: the ethopolitics of pleasure, psychoactive substance Use and public health and criminal justice system governance: therapeutic jurisprudence and the drug courts in the United States of America. Social Legal Stud 17(4):513–533

Marlowe DB (2009) Evidence-based sentencing for drug offenders: an analysis of prognostic risks and criminogenic needs. Chapman J Crim Just 1:167–201

McCoy C (2006) Review of "good courts: the case for problem-solving justice", by Greg Berman and John Feinblatt. Law Polit Book Rev 16(12):964–969

Meekins T (2007) Risky business: criminal specialty courts and the ethical obligations of the zealous criminal defender. Berkeley J Crim Law 12:75–126

Miller EJ (2004) Embracing addiction: drug courts and the false promise of judicial interventionism. Ohio State Law J 65(6):1479–1576

Mirchandani R (2008) Beyond therapy: problem-solving courts and the deliberative democratic state. Law Soc Inquiry 33(4):853–893

National Association of Criminal Defense Lawyers (2009) America's Problem-solving courts: the criminal costs of treatment and the case for reform. Author, Washington, DC

Nolan JL Jr (2001) Reinventing justice: the American drug court movement. Princeton University Press, Princeton

Nolan JL Jr (2009) Legal accents, legal borrowing: the international problem-solving court movement. Princeton University Press, Princeton

Nolan JL Jr (2010) Harm reduction and the American difference: drug treatment and problem-solving courts in comparative perspective. J Health Care Law Policy 13:31–47

Nolan JL Jr (2012) Problem-solving courts: an international comparison. In: Petersilia J, Reitz KR (eds) The oxford handbook of sentencing and corrections. Oxford University Press, New York, p 150

Reisig M (2002) The difficult role of the defense lawyer in a post-adjudication drug treatment court: accommodating therapeutic jurisprudence and due process. Crim Law Bull 38:216–224

Rossman SB, Roman JK, Zweig JM, Rempel M, Lindquist CH (2011) The multi-site adult drug court evaluation: executive summary. Urban Institute, Washington, DC

Saum CA, Scarpitti FR, Robbins CA (2001) Violent offenders in drug court. J Drug Iss 31(1):107–128

Shevroya K (2011) Why veterans should get their own courts. The Atlantic monthly. Atlantic Monthly Group, Boston

Turgeon C (2011) Community courts around the world. Washington, DC: Center for Court Innovation. Retrieved 13 Sept 2012 from http://www.courtinnovation.org/research/community-courts-around-world?mode=4&url=research%2F4%2Farticle

Tyler T (2004) Procedural justice. In: Sarat A (ed) The Blackwell companion to law and society. Blackwell, Oxford

Urban Institute (2005) Are drug courts a solution to the drug problem? Retrieved 13 Sept 2012 from http://www.urban.org/publications/900803.html

Weinstein SP (2000) Ethical considerations for prosecutors in drug courts. Crim Justice 15:26–29

Wexler DB, Winnick BJ (1996) Law in a therapeutic Key. Carolina Academic, Durham

Willrich M (2003) City of courts: specializing justice in progressive era Chicago. Cambridge University Press, Cambridge

Wolf RV (2008/2009) Race, bias, and problem-solving courts. Natl Black Law J 21(1):27–52

Wright M (2006) Reversing the prison landscape: the role of drug courts in reducing minority incarceration. Rutgers Race Law Rev 8(1):79–106

Procedural Fairness

▶ Procedural Justice, Legitimacy, and Policing

Procedural Justice

▶ Police-Led Interventions to Enhance Police Legitimacy

Procedural Justice and Cooperation

Tom R. Tyler[1], Jonathan Jackson[2] and
Ben Bradford[3]
[1]Yale Law School, New Haven, CT, USA
[2]Department of Methodology and Mannheim
Centre for Criminology, London School of
Economics, London, UK
[3]Centre for Criminology, University of Oxford,
Oxford, UK

Overview

Recent discussions of the relationship between legal authorities and the people within their communities emphasize the benefits to legal authorities of gaining voluntary deference and willing cooperation from the people with whom they deal. A key element in gaining such cooperation is being viewed as legitimate. Legitimacy is based primarily upon the fairness of the manner in which legal authorities exercise their authority, i.e., procedural justice. If legal authorities exercise their authority fairly, they build legitimacy and increase both willing deference to rules and the decisions of the police and courts, as well as the motivation to help with the task of maintaining social order in the community.

Procedural Justice and Cooperation

In the United States the dominant model for the exercise of legal authority is deterrence. Its goal is to encourage public compliance with the law. The mechanism for achieving this goal is through the threat or use of punishment for rule breaking. The traditional objective of the law and of the actions of legal authorities is to gain first public compliance with the law, and second acceptance of the particular decisions made by legal authorities such as judges and police officers. In both cases, public compliance cannot be taken for granted.

The problems involved in obtaining compliance in everyday life are exhibited inter alia in traffic laws, drug laws, illegal immigration, and the payment of taxes. While most people comply with the law most of the time, legal authorities are confronted with sufficient noncompliance to challenge the resources normally devoted to social control. In situations such as the illegal use of drugs and the downloading of music and illegal copying of movies, levels of noncompliance are so high as to make effective regulation very difficult.

Similar problems arise with the decisions of legal authorities. Studies of personal encounters with the police demonstrate that people often resist and even defy legal and judicial orders (Mastrofski et al. 1996). Compliance can never be assumed. The ability to manage encounters with members of the public without provoking conflict and resistance is a core competency issue for police officers. As Mastrofski et al. (1996, p. 272) note: “Although deference to police authority is the norm, disobedience occurs with sufficient frequency that skill in handling the rebellious, the disgruntled, and the hard to manage – or those potentially so – has become the street officer’s performance litmus test.” Based upon observational data, these authors estimate that the general noncompliance rate is around 22 %.

The willingness of litigants to accept the decisions made by judges is also a long-standing issue in the courts. A major motivation for the alternative dispute resolution movement, which uses non-adjudicative forums such as mediation instead of traditional adjudication, is to find a way to increase the willingness of the involved parties to accept the decisions made by third-party authorities. Alternative procedures do so by replacing judges with mediators and adjudication with a more informal and cooperative problem-solving procedure. Within all types of adjudicative settings, judges have struggled to find ways to make decisions that are acceptable to the parties who bring cases into court.

In recent decades the exercise of legal authority to obtain compliance has become primarily associated with the use of threat and punishment aimed at deterring people from engaging in criminal behavior (Nagin 1998). Irrespective of the type of case involved, the traditional means of obtaining compliance is via attempts to deter rule breaking by threatening to punish wrongdoing. The argument driving this perspective is threefold: that fear of possible future punishment leads to compliance with the law, that risk calculations are partly shaped by both the anticipated likelihood of punishment and by judgments about its severity, and that the focus is (and should be) on the power of legal authorities and institutions to shape behavior by threatening to deliver (or by actually delivering) negative sanctions for rule breaking. Within legal circles, this way of viewing the relationship between legal authorities and citizens is referred to as the "deterrence" or "social control" model. It is this model of human behavior that – for better or worse – currently dominates law and public policy.

In this entry we argue that a deterrence model of legal authority is not only expensive and minimally effective; it also undermines forms of social capital that promote long-term public commitment to the law and, crucially, the public cooperation on which legal authorities fundamentally rely. The exercise of authority via the application of fair process strengthens the social bonds between individuals and authorities. Procedural justice promotes normative modes of compliance and cooperation that are both more stable and more sustainable in the long run. Using the example of the UK, we show how specific policing contexts can modulate – but not fundamentally alter – these associations. We consider moral alignment with the London Metropolitan Police – one of a broader range of motivations to cooperate with legal authorities that are based not on material interests but on social values and normative commitment.

Deterrence-Based Legal Policy

There are two aspects of deterrence policy. First is the idea that people's law-related behavior is shaped by their expectations about the likely punishment that will result from rule breaking. According to this perspective, people are influenced by their estimates of the likelihood of punishment, the severity and certainty of punishment, or both. Second, deterrence models suggest that if people are caught and punished, the probability of punishment shapes the likelihood of post-punishment wrongdoing.

Judges, for example, attempt to influence people's acceptance of their decisions by threatening fines or jail time for failure to comply. Similarly, police officers carry guns and clubs, and they are empowered to threaten citizens with physical injury and incapacitation, among other penalties. The goal is to establish legal authority. As Reiss (1971, p. 46) points out: "The uniform, badge, truncheon, and arms all may play a role in asserting authority" in the effort to "gain control of the situation." The police thereby seek to gain control over the individual's behavior "by manipulating an individual's calculus regarding whether 'crime pays' in any particular instance" (Meares 2000, p. 396).

More generally, agents of the legal system who are charged with producing compliant behavior concern themselves with shaping environmental contingencies in such a way that citizens will be faced with the prospect of heavy losses (e.g., fines, arrest, incarceration) that are intended to outweigh the anticipated gains of engaging in criminal behavior. The deterrence model dictates that the responsibility of lawmakers is to decide which acts should be prevented and then to specify sufficiently strict penalties, generally fines or prison term, so that the prohibited behavior is rarely enacted.

At the heart of an institutional model of compliance is rational choice theory, which is derived from neoclassical economics (Blumstein et al. 1978). It is assumed that people calculate expected utilities by multiplying the probability of an outcome (e.g., getting caught for armed robbery or drunk driving) by its valence (very, very bad), and then balance the result against the benefits of crime (Paternoster 2006). Rational self-interest is the motivational engine of the deterrence/social control model. If laws and sanctions are well calibrated, people will arrive at the desired conclusion that they should follow the law. It follows that to regulate behavior, decision-makers should adjust criminal sanctions to the needed level, so that the expected losses associated with lawbreaking will minimize the likelihood that people will break the law.

This model has several problems. First, it is a costly and minimally effective system of social control. The high cost of the system stems from the need to create and maintain a credible threat of punishment. People will only change their behavior when they feel that there is a reasonable risk of being caught and punished for wrongdoing. And, of course, they will try to hide their illegal behavior, so a system of surveillance will be needed to identify wrongdoing. Such surveillance is often not possible.

Second, even when surveillance is possible, the magnitude of influence found is at best weak. There is some evidence that variations in the perceived certainty of punishment do affect some people's compliance with the law – at least to some degree, in some circumstances, at some points in time (Paternoster 2006). People's behavior is sometimes – though certainly not always – shaped by their estimate of the likelihood that, if they disobey the law, they will be caught and punished (see Nagin and Paternoster 1991; Paternoster 2006). But, overall, the body of evidence is mixed. According to Piquero et al. (2011, p. 1; see also Paternoster 2006): "…some studies finding that punishment weakens compliance, some finding that sanctions have no effect on compliance, and some finding that the effect of sanctions depends on moderating factors."

Crucially, when perceptions of the likelihood of being caught and punished do influence people's behavior, the effect seems to be relatively small. Consequently, social control strategies that are based exclusively on a deterrence model of human behavior have had at best limited success. An example is the analysis by MacCoun (1993), which shows that variations in the certainty and severity of punishment for drug use account for approximately 5 % of the variance in drug use behavior. This is a finding consistent with the suggestion of Paternoster (1987, p. 191) that "perceived certainty [of punishment] plays virtually no role in explaining deviant/criminal conduct."

In summary, deterrence is a high-cost strategy that at best yields identifiable but weak, results. A system for detecting wrongdoing must be created and maintained, but this is itself unlikely to be effective unless vast sums are spent on it. Further, it is not realistic to substitute draconian punishments for a yet more costly system that creates credible risks of being detected while engaging in wrongdoing. For this reason, as Meares (2000, p. 401) notes, the effectiveness of "instrumental means of producing compliance always depend[s] on resource limits." The relevant questions are the following: how much in terms of financial and other benefits and burdens authorities are willing to expend in order to control crime? And how much power to intrude into their lives are citizens willing to grant to the authorities?

Deterrence and Social Capital

In addition to being costly and minimally effective, the deterrence model is problematic from a social capital perspective. Social scientists have coined the term "social capital" to refer to the reservoir of psychological (positive values, attitudes, identities) and sociological (dense social networks) factors that predispose communities to be able to work collectively to address community issues more effectively. In this vein, this entry focuses on the psychological predispositions that promote compliance and facilitate cooperation, i.e., the values, attitudes, and identities that lead people in communities to more willingly work with others to address the needs of their community (Tyler 2011).

From a social capital perspective, the deterrence model is problematic because threat and punishment do not build positive values, attitudes, and identities. The central negative consequence of deterrence approaches is that they define people's relationship to law and legal authorities as one of risk and punishment. This dampens peoples' focus upon other aspects of their connection to society – such as shared values and concerns – and encourages people to act in ways that are linked to personal gains and losses.

A further negative consequence of deterrence approaches is that, because people associate law and legal authorities with punishment, the instrumental relationship between the public and the legal system is antagonistic. People become more likely to resist, more likely to avoid legal authorities, and less likely to cooperate with them. Social psychologists have long pointed out that punishments and rewards do not create positive psychological dispositions.

Strategies based upon external contingencies rely upon the continuing ability to deliver credible threats, to enforce rules with punishments, and – when incentives are promised – to deliver those incentives (Tyler 2011). As a consequence, when authorities manage people by surveillance, there is no basis for building a sense of trust. For example, employees who have been given the opportunity to follow rules for internal reasons demonstrate to workplace authorities when they do so that they can be trusted. Subsequently, authorities are more comfortable allowing those individuals to work without supervision. However, when authorities are constantly present, they have no basis for trust and can suspect that the moment they leave, people will stop following the rules. Hence, their very behavior of surveillance creates the conditions requiring future surveillance.

Recent research findings have moved beyond the statement that the deterrence model does not build supportive attitudes and values. Scholars have argued that the policies and practices associated with this model may actually undermine existing social capital. Crucially, the use of surveillance systems has deleterious effects on the social climate of groups. The use of surveillance implies distrust, which decreases people's ability to feel positively about themselves, their groups, and the system itself (Kramer and Tyler 1996). Furthermore, people may experience intrusions into their lives as procedurally unfair, leading to anger and other negative emotions often associated with perceptions of injustice (e.g., Gurr 1970).

Whether surveillance works or not, it is often demotivating. Surveillance introduces new costs in terms of distrust and perhaps even paranoia in subsequent social interaction. Such costs are borne by groups, organizations, and societies to which people belong, as they lose the gains that occur when people are willing to cooperate with each other. Research suggests that the increasing use of deterrence strategies and social control has exerted precisely this type of negative influence on the US social climate.

These issues can be illustrated by decisions about child custody and child support (Bryan 2006). Judges can issue orders concerning custody and child support payment and can back them up in a variety of ways. However, to maximize the favorability of the climate within which children are raised, it is important to make decisions that will be willingly followed by both fathers and mothers. Further, to the degree to which it is possible, it is important to create positive post-separation dynamics in which both parents take responsibility for supporting their children financially and emotionally (Bryan 2006). Hence, family law cases involving child custody and child support have been a particular focus of efforts to find ways to create acceptable decisions. And those efforts point to the potentially destructive role played by adversarial dynamics, which can lead to a negative relationship among parents who need to cooperate to raise their children.

Deterrence has also created an adversarial relationship between legal authorities and members of the communities they serve. This is especially so with respect to racial and ethnic minority group members (Tyler and Huo 2002). An adversarial relationship leads the public to grow less compliant with the law and less willing to help the police to fight crime (Sunshine and Tyler 2003). Those people who feel under surveillance by the police or who have been stopped while driving or walking to be questioned and potentially arrested have more negative feelings about the police and the law (Tyler and Fagan 2008; Bradford et al. 2009).

An analysis of people's connection to law in terms of general principles of human motivation further suggests that if people comply with the law only in response to coercive power, they will be less likely to obey the law in the future because acting in response to external pressures diminishes internal motivations to engage in a behavior (Tyler and Blader 2000). This follows from the well-known distinction in social psychology between intrinsic and extrinsic motivation. Research on intrinsic versus extrinsic motivation shows that when people are motivated solely by the prospect of obtaining external rewards and punishments (i.e., extrinsic motivation), they become less likely to perform the desired behavior in the absence of such environmental reinforcements (e.g., Deci 1975). On the other hand, if people are motivated by intrinsic reasons for behaving in a certain way, then their compliance becomes much more reliable and less context-dependent.

Studies of regulatory authorities demonstrate that seeking to regulate behavior through the use of threat serves to undermine people's commitment to rules and authorities. From a motivational perspective, instrumental approaches are not self-sustaining and require the maintenance of institutions and authorities that can keep the probability of detection for wrongdoing at a sufficiently high level to constantly motivate the public through external means (i.e., the threat of punishment). Over time it becomes more and more important to have such external constraints in place, for whatever intrinsic motivation people originally had is gradually "crowded out" by external concerns.

And while the focus of the rest of this entry is on legitimacy, the normative commitment to cooperate and the psychological aspects of social capital, it is equally important to recognize that deterrence approaches undermine the sociological elements of social capital by damaging communities. Because of the widespread belief that crime is deterred by the threat and/or the experience of punishment, a massive number of American citizens have been convicted and sentenced to serve time in American prisons. Today the USA is a world leader in the proportion of citizens it holds in prison. The United States has the highest prison population rate in the world, 756 per 100,000 national population in 2007, much higher than those in equivalent countries such as England and Wales (153 per 100,000), France (96 per 100,000), and Germany (89 per 100,000) (Walmsley 2008). This imprisonment has a strong impact upon American communities, especially urban communities and especially among members of racial and ethnic minority groups that are overrepresented in the prison system. Families and the communities in which they live are undermined as social units by the loss of large numbers of young and predominantly male members of these communities.

Focusing upon Voluntary Cooperation

Discussions of all types of social institutions have focused upon trying to understand how institutions can best encourage citizens to act pro-socially, and these discussions are also relevant to public cooperation with the police and courts. The focus of study has broadened to include not just compliance – motivated by concerns about costs and gains – but willing deference to the law and the decisions of legal authorities and the general willingness to cooperate with legal authorities. The legal system benefits when people voluntarily defer to regulations and continue to defer over time. In the context of personal experiences with police officers or judges, the legal system is more effective if people voluntarily accept the decisions made by legal authorities. Absent such acceptance, legal authorities must engage in a continuing effort to create a credible threat of punishment to assure long-term rule-following/decision acceptance. These types of voluntary activity are not effectively motivated by the risk of punishment. Threats can sometimes compel obedience. But they do not motivate voluntary deference.

The Importance of Legitimacy

How can legal authorities generate voluntary deference? Studies show that people cooperate when they believe that agents of the law are rightful holders of authority and when they imbue the legal system with the corresponding duty to obey (Tyler 2006a, b). Feelings of identification with the police and courts – and a willingness to defer to their instructions – generate the belief that authorities have the right to dictate appropriate behavior and that authorities are justified in expecting feelings of obligation and responsibility from citizens. Importantly, legitimacy activates self-regulatory mechanisms: people defer to, and cooperate with, legitimate authorities because they feel it is right to do so (Tyler 2006a, b).

The legitimacy of the law and of legal authorities resides most fundamentally in the recognition that the criminal justice system has the right to exist and dictate behavior, and use its authority to determine the law, use force, and punish those who act illegally in justifiable ways. While definitions of legitimacy vary widely, a key feature of many is that it confers the right to command and to dictate behavior and that it promotes the corresponding duty to obey (Weber 1968; Tyler 2006a). Modern discussions of legitimacy are usually traced to the writings of Weber (1968) on authority and the social dynamics of authority. Weber, like Machiavelli and others before him, argued that successful leaders and institutions use more than brute force to execute their will. They strive to win the consent of the governed so that their commands will be voluntarily obeyed (Tyler 2006a).

Kelman and Hamilton (1989) refer to legitimacy as "authorization" to reflect the idea that a person authorizes an authority to determine appropriate behavior within some situation, and then feels obligated to follow the directives or rules that authority establishes. As they indicate, the authorization of actions by authorities "seem [s] to carry automatic justification for them. Behaviorally, authorization obviates the necessity of making judgments or choices. Not only do normal moral principles become inoperative, but – particularly when the actions are explicitly ordered – a different type of morality, linked to the duty to obey superior orders, tends to take over" (Kelman and Hamilton 1989, p. 16).

Legitimacy, according to this general view, is a quality that is possessed by an authority, law, or institution that leads others to feel obligated to accept its directives. It is "a quality attributed to a regime by a population" (Merelman 1966, p. 548). When people ascribe legitimacy to the system that governs them, they become willing subjects whose behavior is strongly influenced by official (and unofficial) doctrine. They also internalize a set of moral values that is consonant with the aims of the system. And – for better or for worse – they take on the ideological task of justifying the system and its particulars.

As Kelman (1969) puts it: "It is essential to the effective functioning of the nation-state that the basic tenets of its ideology be widely accepted within the population.... This means that the average citizen is prepared to meet the expectations of the citizen role and to comply with the demands that the state makes upon him, even when this requires considerable personal sacrifice" (p. 278). Widespread voluntary cooperation with the state and the social system allows authorities to concentrate their resources most effectively on pursuing the long-term goals of society. The authorities do not need to provide incentives or sanctions to all citizens to get them to support every rule or policy they enact.

Perceptions of legitimacy are based on feelings of obligation that are disconnected from substance and material interest. Legitimacy is linked not to the authorities' possession of instruments of reward or coercion but to properties of the authority that lead people to feel it is entitled to make decisions and be obeyed. People will abide by the law and follow police directives even if they disagree with specific guidelines and instructions. In an increasingly pluralistic and diverse society, it is important that legal authorities enjoy a sort of social influence that transcends particular moral values. Remaining salient in situations where authorities and subordinates disagree, legitimacy is an important source of governance, especially in complex societies. For example, the moral beliefs of anti-abortion activists directly conflict with the views of the Supreme Court. But the legitimacy of a Supreme Court ruling on abortion is still conceded. Because legitimacy is based upon the fairness of the procedures used by authorities to govern, rather than upon the substance of their decisions, legitimacy allows authorities to make decisions that are widely accepted even by those who disagree with them.

Procedural Justice

Why do people confer legitimacy to legal authorities? Research shows the importance of procedural justice. The most immediate context of procedural justice is the direct experience that citizens have with police and court officials. The primary factor shaping decision acceptance (when legal authorities make decisions concerning the individual in question) is the procedural justice of the process through which a decision was reached (Tyler 2006a, b). This factor is approximately seven times as important as either the favorability or the fairness of the outcome. Similar findings emerge when we consider why people have positive or negative views about the authorities involved. Again procedural justice is the key antecedent. In particular, procedural justice shapes views about the overall legitimacy of the law and the legal system. Research indicates that people's view that the legal system is legitimate increases following a negative outcome, as long as people experience the procedures used by authorities as being fair (Tyler and Fagan 2008).

In studies of the general population, people are also found to regard the police as legitimate if they believe that the police exercise their authority through fair and impartial means (Sunshine and Tyler 2003; Jackson et al. 2012a, b). Importantly, available evidence suggests that the perceived procedural fairness of the police (the belief that the police would treat me with respect and dignity if I were to come into contact with officers, e.g., and the belief that the police generally treat people with respect and dignity) procedural justice judgments are more central to judgments of legitimacy than are such factors as the perceived effectiveness of the police in combating crime. To the extent that people perceive law enforcement officials as legitimate, they are significantly more willing to comply with the law in general (Sunshine and Tyler 2003; Tyler 2006a).

The procedural basis of legitimacy is especially strong with respect to public opinion concerning political and legal institutions. Studies of the presidency, the legislature,, and the Supreme Court all suggest that when citizens are evaluating government institutions, they focus primarily on the fairness of the procedures by which the institutions make policies and implement its decisions. Research on work organizations also shows that perceived legitimacy has a strong procedural basis (Tyler and Blader 2000).

Procedural justice also plays a critical role in securing compliance over time (Paternoster et al. 1997). It is by now clear that people's reactions to law and legal authorities are heavily influenced by their assessments of the fairness of legal procedures. The most reliable way of attaining legitimacy and maintaining support for legal institutions and authorities is by establishing and protecting procedural safeguards. Indeed, the need for procedural safeguards is one of the strongest arguments for the constitutional separation of executive, representative, and judicial branches of government. To the extent that procedures for insuring genuine fairness are compromised, the system will begin to lose legitimacy and – over time – fail to inspire the kind of cooperation and deference that is often taken for granted during periods of stability.

What is particularly striking about procedural justice judgments is that they shape the reactions even of those who are on the losing side of cases. If a person who does not receive an outcome that they think favorable or fair feels that the decision was arrived at in a fair way, they are more likely to accept it. Longitudinal studies show that people continue to adhere to fairly arrived at decisions over time, suggesting that their acceptance of those decisions is genuine and not simply the result of fear or coercion (Tyler et al. 2007). Further, people who experience procedural justice in court rate the courts and court personnel more favorably, indicating higher levels of trust and confidence in the courts and the court system.

What Makes a Procedure Fair?

As understood in US communities, procedural justice is defined in terms of four issues (also see Murphy, this volume). First, people want to have an opportunity to explain their situation or tell their side of the story in a conflict (voice). Second, people react to evidence that the authorities with whom they are dealing are neutral.

Third, people are sensitive to whether they are treated with dignity and politeness and to whether their rights as citizens are respected. Finally, people focus on cues that communicate information about the intentions and character of the legal authorities with whom they are dealing ("their trustworthiness").

Legitimacy and trust matter because they motivate voluntary deference both with the decision of legal authorities and with the law in everyday life. When people feel responsibility to defer to legal authorities, they do so because it is the right thing to do, not out of fear of punishment or expectation of reward. Hence, legitimacy is a key to self-regulation.

Cooperation

Beyond deference, legal authorities are increasingly focusing upon the benefits of active public cooperation with the police, the courts, and the law. The legal system – and particularly the police – needs voluntary help from the public (Tyler and Fagan 2008). Studies of crime and policing make clear that the police benefit when members of the community voluntarily work with them to manage social order in their community. First, the police need community members who are familiar with the neighborhood and its residents to report crime and criminals to the police. Second, the police need community residents to join them in town meetings and neighborhood watch organizations to identify and deal with community problems.

The shift from compliance to cooperation requires us to expand the scope of our instrumental model – a model that can potentially involve both costs and benefits. While the legal system is based upon highlighting the cost of rule breaking, law and legal authorities also offer benefits that can strengthen social capital. For example, Tyler and Huo (2002) found that the primary reason for people having contact with the police and courts is having gone to them for help. Legal authorities can create benefits by their performance in resolving disputes and maintaining social order. Unlike the relationship of the police to regulation, in which they are primarily the dispensers of punishment, when we consider cooperation, we can also see the police as a group able to provide people with desirable outcomes, by helping them to solve problems and address problems in their communities. The police can provide desirable rewards such as high performance in solving crime, maintaining order, or addressing public problems. Members of the public can – and should – have an active interest in police success that does not rest on sanctions directed at them.

An extension of the traditional view of legitimacy is to also consider moral alignment – the belief that the police share the moral values of individuals, i.e., they believe that justice institutions have a just and moral purpose (which creates a normative justifiability of police power and authority). When people feel aligned with the moral values of an authority in a group setting, they will act in ways that support the group that the authority represents. Institutions can encourage cooperation by activating the ethical motivations that lead people to adhere to group rules and to act on behalf of the group (Tyler 2006a, b; Tyler and Huo 2002).

The powerful influence of procedural justice, legitimacy, and moral alignment on public cooperation with legal authorities is just one example of a broader set of motivations based on social connections rather than material interests (Tyler 2011). Studies demonstrate the importance of five types of normative commitment to cooperate with legal, managerial, and political authorities. These are attitudes, values, identity, procedural justice, and motive-based trust.

Individuals also cooperate with organizations when those organizations serve the social function of providing individuals with a favorable identity and a positive sense of self (Blader and Tyler 2009). Pride and respect within the group are important. If people feel pride in the group and believe that they are accorded respect, then their motives will be transformed from the personal to the group level. If individuals identify with the group that the police represent (i.e., the society and nation), then they merge their sense of self with the group. Defining themselves in terms of their group membership, they will be more willing to act cooperatively on behalf of that group. The goals of the police become one's own goals.

The effect of these two values – legitimacy and moral alignment – on willingness to cooperate with police and criminal courts has been demonstrated in recent UK-based work (Jackson et al. 2012a, b). First, legitimacy was an important predictor of cooperation with the London Metropolitan Police. People who voluntarily deferred to the police, who felt obligated to obey the decisions and directives of officers and who felt the police had a right to determine appropriate behavior, also reported being willing to call the police, to come forward as witnesses, and to identify suspects in court. The correlation between legitimacy and cooperation was particularly strong among a specific subset of the sample (young male from particular ethnic minorities) who tend to have a more fractious, situated, and confrontational relationship to the police.

But the same study also identified a second factor linked to cooperation. People who felt aligned with the moral values of the police (cf. Sunshine and Tyler 2003; Jackson and Sunshine 2007) also reported being motivated to aid the police and courts. In turn, procedural justice was a strong predictor of moral alignment. Participants in the London study valued fairness, decency, and transparency (in interpersonal treatment and decision-making) over instrumental concerns (cf. Tyler and Huo 2002; Gau and Brunson 2010). Social psychological research shows that people want to feel valued members of social groups and they derive self-relevant information through the quality of their interactions with group representatives. Fair, decent, and respectful treatment fosters motive-based trust and identification with the group (Tyler and Blader 2000), as well as a sense of shared moral values.

In Jackson et al. (2012a), moral alignment was also related to a particular collection of relational concerns and social characteristics of the neighborhood in which individuals lived. The level of collective efficacy in a neighborhood – the sense of trust and shared values, and the willingness of residents to act upon those values on behalf of the common good – and disorder in that community (i.e., signs of the lack of community oversight and the inability of officials to maintain low-level social order) explained a good deal of variation in moral alignment with the police (cf. Jackson and Sunshine 2007; Jackson and Bradford 2009). By contrast, neighborhood levels of crime and fear of crime were not related to moral alignment with the police. When subtle, informal social controls are strong, community members believe the police are aligned with their values. Conversely, poor social conditions may encourage the feeling that the police are not representing and defending group norms, values, and identities.

Overall, these findings support the idea that moral alignment with the police is not an instrumental motivation – based on material interests and personal concerns about crime and risk – but a normative motivation rooted in social connections and identification. Moral alignment seems to strengthen the motivation to cooperate because (a) the goals accord with one's moral values, (b) the interests of the group become more of one's own interests, and (c) two people value the social connection they have with the police and the group the police represents. This seems to be an additive effect, on top of the effect of legitimacy on cooperation.

Widening the Definition of Legitimacy?

Is legitimacy both obligation and moral alignment (Jackson et al. 2012a, b)? Individuals may confer the right of the police to dictate appropriate behavior when they believe that the police are legitimate in the authority they possess and legitimate in the moral values they defend and represent. People may obey the police, even if they disagree with the specific context of the instructions, because of the authority the police possess and because of a shared moral purpose. This is the idea floated recently in Europe (Jackson et al. 2011; Hough et al. 2013).

Including moral alignment in the definition of legitimacy certainly makes sense in England and Wales, given the specific symbolism and the legitimizing foundation of the British police. The cultural symbolism of the police is rooted in the foundational myths of the London Metropolitan Police and the "bobby on the beat." When considering the history of British policing, it is commonplace to refer to Robert Peel's principles of policing operation for the Metropolitan Police. While doubts exist as to the primary source of these principles, which Lentz and Chaires (2007) suggest may be somewhat the "invention" of twentieth-century textbooks, their theme and spirit remain important. Phrases like "the police are the public and the public are the police" and "the ability of the police to perform their duties is dependent upon public approval of police actions" (e.g., Reith 1952, p. 154) speak to a close social connection between what were then "subjects of the crown" and the police.

In the eyes of many citizens, the continued justification of the existence and moral authority of a police force in England and Wales may rest on the idea that the police reflects and defends a common sense of shared values. People look to the police to defend and typify moral values; they assess its right to determine appropriate behavior partly on the basis of shared values. It follows that the legitimacy of the police rests in part on a moral link between people and police – a link that may remain strong despite the decline in public trust in the police over the past few decades.

Defining legitimacy as duty to obey and moral alignment also accords with the work of David Beetham (1991). Beetham insists that legitimacy is not simply a property of an authority or something which exists only in the subjective experiences of citizens. It is also *constituted by* normative assessments of authorities and the actions of those they govern. When citizens grant legitimacy to the police, they do so on the basis that the relationship between individual and authority is founded on common shared values. When individuals act in ways congruent with recognition of this commonality, their behavior serves to confirm and reproduce the legitimate role of the authority. Beetham outlines three dimensions of legitimacy, each of which must be fulfilled for an authority to be considered legitimate. First is conformity to a set of rules. Second is the justifiability of these rules in terms of shared beliefs. Third is the expressed consent of those governed or otherwise affected by the authority. In the case of the psychological state of legitimacy, these will be beliefs about the lawfulness and legality of an authority, a sense of shared moral values, and a positive duty to obey.

Whether moral alignment is important in other countries, and whether it can be reasonably thought of as legitimacy, remains an open and empirical question (although see Hough et al. 2013). In many societies the police may be seen less as a connected to shared moral values and more an agent of last resort. Such an institution requires from citizens obligations and feelings of responsibility, but does not draw its authority from any kind of value congruence.

And the difficulty with viewing moral alignment as an aspect of legitimacy may arise particularly in diverse multicultural societies. If the people within a political entity do not share a common set of moral values, then it will be harder for government authorities such as the police or the courts to represent shared values, and harder for them to invoke values as a source of authority. It is also harder for legal authorities to enforce laws that do not reflect moral values, shared or not. When the police are trying to enforce laws against using drugs or illegally downloading music and films, they will have difficulty drawing upon shared value structure because many people do not view such behaviors as immoral. As with parking violations or speeding, many people do not view breaking such laws as wrong, and adopt a more instrumental "try to catch me" attitude toward their behavior.

It is for future empirical research to assess, but it may be that people's sense of "moral alignment" is based primarily on procedural justice, namely, the values that police express by the actions and decision-making of officers. Extant research shows extremely strong associations between trust in police fairness and moral alignment (Jackson et al. 2012a, b), suggesting that procedural justice is a central way in which authorities demonstrate their moral values. If this is so, then legitimacy may still not have to depend too strongly on consensual views among different communities. Procedural justice sits above diverse value positions; two parties in an interaction (indeed, in an argument) do not have to agree with each other on the specific issue at hand to treat each other with respect and dignity, allow each other a voice, and develop a sense of mutual trust. Fair treatment and fair decision-making may demonstrate to citizens that police officers respect their views and values and are thus working to a broad set of appropriate moral standards that recognize the worth of citizens and which give then officers a rightful claim to power and authority.

Final Words: Social Capital, Regulation, and Citizenship

As the results of the UK study indicate, when the police can draw upon shared values, moral alignment provides an important additional normative force beyond legitimacy that aids their efforts. The British example underlines the extent to which the primary aim of policing – the reproduction of normative social order – can be secured via coproduction involving police and community. When police behave in a procedurally fair manner, people are encouraged to feel ready and able to cooperate with officers. They are more likely to feel it is right and proper to support the police, and they are more willing to engage with them to address the problems they face in their communities. They do so not when they are incentivized by aggressive policing, nor necessarily when they see the possibility of a direct material return, but when such cooperation becomes valid in itself, because it seen as an effort to work toward a common goal.

Furthermore, individual and collective propensity to cooperate with the police can itself be a form of social capital. Despite the growth in forms of private policing, in the UK as elsewhere (Jones and Newburn 2002), for most of the population, the police are the only option. Those who feel unable or unwilling to draw on the services of the police at times of need – perhaps because they have experienced police unfairness in the past – are denied a core element of the assistance the state can provide to them. Attempting to deal personally with problems that might indicate a need for the police may incur significant economic, social, and personal costs.

It may also cause problems in the community when people take the law into their own hands (Huq et al. 2011). Police acts that are unfair damage citizen's sense of alignment with police and trigger withdrawal from an important form of social support, a form of capital all citizens should feel able to draw upon when needed. When this happens, police are less likely to be able to solve problems in the communities in which they live in an effective and equitable manner. This may serve to initiate a downward spiral of failure, as social cohesion is undermined, disorder is felt to increase, and police legitimacy is further undermined.

Across the social sciences, there has been a widespread recognition that it is important to understand how to motivate cooperation on the part of people within group settings. This is the case irrespective of whether those settings are small groups, organizations, or communities (Tyler 2011). Studies in management show that work organizations benefit when their members actively work for company success. Within law, research shows that crime and problems of community disorder are difficult to solve without the active involvement of community residents. Political scientists recognize the importance of public involvement in building both viable communities and strong societies. And those in public policy have identified the value of cooperation in the process of policy making – for example, in stakeholder policy making groups. Hence it is important to ask whether the actions of legal authorities can aid in the production of social capital that might more generally engage cooperation.

Recent discussions of legal authority focus on a broader conception of the relationship between community residents and legal authorities. This broader framework suggests a new and broader conception of citizenship. Rather than concerning simply deferring to authorities, citizenship is increasingly seen as connected to the willing and active cooperation with authorities in creating and maintaining order in the community. For example, Tyler and Fagan (2008) identify two distinct forms of cooperation that benefit the police. The first is cooperation in solving particular crimes, for instance, by reporting crime in the community. The second is cooperation in co-policing the community, by working with the police in discussions about community issues and problems and through general willingness to work with the police, for example, by joining neighborhood watch and helping to patrol areas in one's neighborhood. Tyler et al. (2010) make a similar distinction in their study of anti-terror policing in the Muslim American community. They note that the same two types of cooperation are valued by legal authorities: reporting potentially dangerous actions and generally working with the police to help police the community against terror threats.

Both of these forms of cooperation are essentially voluntary. Noncooperation risks little or no sanction. In particular, *not* reporting threats to the community is something that is virtually undetectable. People's willingness to do so is therefore linked to a broader loyalty or sense of duty to the community and its authorities. Such cooperative behaviors are therefore general civic activities akin to the more traditionally studied forms of political participation. The issue that is being raised in respect to crime, in other words, is the same type of issue raised with regard to any matter affecting a community. How can citizens be motivated to become involved in deciding how to manage their community and in engaging in behaviors oriented toward doing so?

It is also clear that the actions of legal authorities have an impact on people's views about government. Because the actions of legal authorities generalize to views about society and government, it should be possible to develop strategies of law enforcement that are more socially beneficial by building identification with government and society as well as the type of supportive values that lead to feelings of obligation. For example, the police might be able to help build the values that would lead people to pay their taxes or act as volunteers. Further, they might build the type of connections that lead people to work willingly and enthusiastically in their communities. In other words, rather than being viewed as a (necessary) cost that has negative implications for society, the legal system could generate supportive attitudes and values that would resonate more broadly through society.

Legal authorities can produce the benefits of building public support, benefits which include success in fighting crime and also general support for the community. People are more likely to come to and revisit communities in which they feel that they will be well treated by the representatives of government they are most likely to encounter – the police. This benefits communities economically because people more willingly come to them to work and to shop, as tourists, and for entertainment and sporting events. Hence, the police play a central role in creating the reassurance that makes a community inviting and desirable to the general public. In other words, the law provides a framework for building vibrant, alive, successful communities. If people feel reassured by the presence of the police and believe that they will be protected and, if they need it, helped, then they will be encouraged to engage in their communities socially and economically. When people engage in such behaviors, they build social capital and the sense of efficacy that we have already noted has broad social value. If people engage in their communities, they come to know others and to know how to work with them when problems arise in the community. They build trust in others and develop the police that others can and will join together to address issues when they arise. By providing a framework of reassurance, the police are creating the climate that allows the community to develop valuable psychological and sociological characteristics.

Underlying these comments is the question of how legitimacy, trust, and values such as collective efficacy are related. It may be that strong government and legitimate institutions help foster collective efficacy (LaFree 1998; Kochel 2012). Equally, work in the UK suggests that people infer the moral values of the police from the strength of the values and social ties they experience in their community. We would suggest that to some extent, both of these are true. A feedback loop may exist between collective efficacy and institutional legitimacy.

Tyler and Blader (2000) explore the relationship between people and groups in the context of work organizations. They demonstrate that identification with authorities and institutions is central to motivating the development of supportive attitudes and values, as well as to motivating cooperative behavior. Hence, to the degree that the police can build identification with legal authorities and with the community itself, they promote supportive public attitudes and voluntary cooperative behaviors. Studies in work settings widely suggest that, as with law, when authorities exercise their authority in fair ways, they motivate employees to identify with their companies, and this identification in turn leads to voluntary cooperation with management. Our argument is that the police and the courts can similarly build identification with society and social institutions and through that identification can motivate members of the community to more actively work on its behalf.

A value-based perspective on human motivation suggests the importance of developing and sustaining a civic culture in which people more willingly abide by the law because they feel that legal authorities are legitimate and ought to be obeyed. For this model to work, society must create and maintain public values. Psychological social capital must be created. Political scientists refer to this set of values as a “reservoir of support” for government and society. But however labeled, a value-based model is consistent with a social psychological understanding of how authorities can effectively regulate citizen behavior, maintain social order, and promote an effective, well-functioning society – by developing and maintaining a culture of supportive social values that will be internalized by the citizenry (Tyler 2011).

As we have noted, such values include feelings of obligation and responsibility to the state. They also include judgments about the degree of moral alignment that people feel with the police, the courts, and the law. And they include broader types of social connections to society, for example, identification with authorities and institutions. All of these values potentially provide an alternative basis for the effective operation of the legal system.

Recommended Reading and References

Beetham D (1991) The legitimation of power. Macmillan, London

Blader S, Tyler TR (2009) Testing and expanding the group engagement model. J Appl Psychol 94:445–464

Blumstein A, Cohen J, Nagin D (eds) (1978) Deterrence and incapacitation: estimating the effects of criminal sanctions on crime rates. National Academy of Sciences, Washington, DC

Bradford B, Jackson J, Stanko E (2009) Contact and confidence: revisiting the impact of public encounters with the police. Polic Soc 19(1):20–46

Bryan PE (2006) Constructive divorce: procedural justice and sociolegal reform. American Psychological Association, Washington, DC

Deci EL (1975) Intrinsic motivation. Plenum, New York

Gau JM, Brunson RK (2010) Procedural justice and order maintenance policing: a study of inner-city young men’s perceptions of police legitimacy. Justice Q 27:1–25

Gurr TR (1970) Why men rebel. Basic Books, New York

Hough M, Jackson J, Bradford B, Myhill A, Quinton P (2010) Procedural justice, trust and institutional legitimacy. Polic J Policy Pract 4(3):203–210

Hough M, Jackson J, Bradford B (in press) The governance of criminal justice, legitimacy and trust. In: Body-Gendrot, S Lévy R, Hough M, Snacken S, Kerezsi K (eds) The Routledge handbook of European criminology. Routledge, Oxon

Huq A, Tyler TR, Schulhofer S (2011) Why does the public cooperate with law enforcement: The influence of the purposes and targets of policing? Psychology, Public Policy and Law, 17(3):419–450

Jackson J, Bradford B (2009) Crime, policing and social order: on the expressive nature of public confidence in policing. Br J Sociol 60(3):493–521

Jackson J, Sunshine J (2007) Public confidence in policing: a Neo-Durkheimian perspective. Br J Criminol 47(2):214–233

Jackson J, Bradford B, Hough M, Kuha J, Stares SR, Widdop S, Fitzgerald R, Yordanova M, Galev T (2011) Developing European indicators of trust in justice. Eur J Criminol 8(4):267–285

Jackson J, Bradford B, Hough M, Myhill A, Quinton P, Tyler TR (2012a) Why do people comply with the law? Legitimacy and the influence of legal institutions. Br J Criminol 52(6):1051–1071

Jackson J, Bradford B, Stanko EA, Hohl K (2012b) Just authority? Trust in the police in England and Wales. Routledge, Oxon

Jones T, Newburn T (2002) The transformation of policing? Understanding current trends in policing systems. Br J Criminol 41(1):129–146

Kelman HC (1969) Patterns of personal involvement in the national system: a social-psychological analysis of political legitimacy. In: Rosenau J (ed) International politics and foreign policy. Free Press, New York, pp 276–288

Kelman HC, Hamilton VL (1989) Crimes of obedience. Yale, New Haven
Kochel TR (2012) Can police legitimacy promote collective efficacy?. Justice Q 29:384–419
Kramer RM, Tyler TR (eds) (1996) Trust in organizations. Sage, Thousand Oaks
LaFree G (1998) Losing legitimacy: street crime and the decline of social institutions in America. Westview, Boulder
Lentz SA, Chaires RH (2007) The invention of Peel's principles: a study of policing "textbook" history. J Crim Justice 35:69–79
MacCoun RJ (1993) Drugs and the law: A psychological analysis of drug prohibition. Psychological Bulletin 113:497–512
Mastrofski SD, Snipes JB, Supina AE (1996) Compliance on demand: the public's response to specific police requests. J Res Crime Delinq 33(3):269–305
Meares TL (2000) Norms, legitimacy, and law enforcement. Or Law Rev 79:391–415
Merelman RJ (1966) Learning and legitimacy. Am Polit Sci Rev 60:548–561
Nagin DS, Paternoster R (1991) The preventive effects of the perceived risk of arrest. Criminology 29:561–585
Nagin DS (1998) Criminal deterrence research at the outset of the twenty-first century. In: Tonry M (ed) Crime and justice, vol. 23. University of Chicago Press, Chicago, pp 1–42
Pagliaro S, Ellemers N, Barreto M (2011) Sharing moral values: anticipated in group respect as a determinant of adherence to morality-based (but not competence-based) group norms. Personal Soc Psychol Bull 37(8):1117–1129
Paternoster R (1987) The deterrent effect of the perceived certainty and severity of punishment: a review of the evidence and issues. Justice Q 4:173–218
Paternoster R (2006) How much do we really know about criminal deterrence? J Crim Law Criminol 100:765–824
Paternoster R, Brame R, Bachman R, Sherman LW (1997) Do fair procedures matter? The effect of procedural justice on spouse assault. Law and Society Review 31:163–204
Piquero AR, Paternoster R, Pogarsky G, Loughran T (2011) Elaborating the individual difference component in deterrence theory. Annu Rev Law Soc Sci 7:335–360
Reiss AJ (1971) The police and the public. Yale, New Haven
Reith C (1952) The blind eye of history: a study of the origins of the present police era. Faber and Faber, London
Sunshine J, Tyler TR (2003) The role of procedural justice and legitimacy in shaping public support for policing. Law Soc Rev 37:513–548
Tyler TR (2006a) Psychological perspectives on legitimacy and legitimation. Annu Rev Psychol 57:375–400
Tyler TR (2006b) Why people obey the law. Princeton University Press, Princeton
Tyler TR (2011) Why people cooperate. Princeton University Press, Princeton
Tyler TR, Blader S (2000) Cooperation in groups: procedural justice, social identity, and behavioral engagement. Psychology Press, Philadelphia
Tyler TR, Fagan J (2008) Legitimacy and cooperation: why do people help the police fight crime in their communities? Ohio State J Crim Law 6(1):231–275
Tyler TR, Huo YJ (2002) Trust in the law: encouraging public cooperation with the police and courts. Russell-Sage Foundation, New York
Tyler TR, Lind EA (1992) A relational model of authority in groups. In: Zanna MP (ed) Advances in experimental social psychology, vol 25. Academic, San Diego, pp 115–191
Tyler TR, Sherman LW, Strang H, Barnes GC, Woods DJ (2007) Reintegrative shaming, procedural justice, and recidivism: the engagement of offenders' psychological mechanisms in the Canberra RISE drinking-and-driving experiment. Law Soc Rev 41(3):553–586
Tyler TR, Schulhofer S, Huq A (2010) Legitimacy and deterrence effects in counter-terrorism policing: a study of Muslim Americans. Law Soc Rev 44:365–401
Walmsley R (2008) World prison population list, 8th edn. International Centre for Prison Studies, London
Weber M (1968) Economy and society. In: Roth G, Wittich C (eds) University of California Press, Berkeley

Procedural Justice, Legitimacy, and Policing

Kristina Murphy
Griffith University, Brisbane, QLD, Australia

Synonyms

Procedural fairness

Overview

A significant amount of police work involves face-to-face interactions with the public. The nature of these interactions can vary considerably from encounter to encounter. Police are often faced with situations in which they have to deliver unfavorable outcomes to the individual or group they are dealing with. This can sometimes lead to overt resistance or aggressive behavior. Citizen compliance with police directives is therefore important and can help police to resolve issues more efficiently. Research in the procedural justice field has shown that if police treat citizens respectfully and make decisions in a fair way, then it can enhance public perceptions of police legitimacy. This in turn has been shown to shape compliance-related behavior during and after a police encounter and has even been shown to promote long-term compliance with the law. This entry provides a summary of procedural justice research that has been conducted in the policing context and highlights future directions and controversies that have emerged in the field.

Fundamentals

A long-standing debate has existed in the criminology literature which contrasts the work of those who believe that individuals will comply with authorities and the law only when confronted with harsh sanctions and penalties and those who believe that alternative models that seek to encourage voluntary compliance will be more successful (see Ayres and Braithwaite 1992). Traditional crime control models that place an emphasis on deterring potential offenders from breaking the law have tended to dominate policy making and enforcement approaches in the criminal justice field, including policing. The basic premise of such models is that people are rational actors who behave in a manner that will maximize their expected utility. In other words, individuals assess opportunities and risks and disobey the law when the anticipated fine and probability of being caught are small in relation to the profits to be made through noncompliance. The view is that if an individual is a rational decision-maker whose aim is to maximize expected gains, authorities should respond by deterring them from acts of noncompliance by ensuring the benefits to be obtained through noncompliance are lower than those obtained through compliance. Advocates of this perspective therefore suggest that compliance can be best achieved in two ways: (1) by increasing the probability of detecting noncompliers and (2) by increasing sanctions to the point where noncompliance becomes irrational (Becker 1968). According to supporters of this deterrence-based approach, authorities need to find an appropriate balance between these two measures to make compliance behavior the rational choice.

Although some research supports the basic premise of the deterrence model and its ability to deter people from noncompliance, other research also suggests that these factors have, at best, a minor influence on people's law-related behavior. Some studies suggest that such estimates do not independently influence behavior when the influence of other factors – such as socialization processes, social norms, or moral reasoning – is considered. Critics of the deterrence model also suggest that such an approach to crime control is costly to maintain. Further, they suggest that it cannot satisfactorily explain the high levels of voluntary compliance observed in situations where the probability of detection is small. If people were simply rational actors motivated purely by self-interest, one would expect that compliance with the law would be significantly lower than what has currently been observed. Finally, some research has also demonstrated that deterrence-based enforcement strategies can be counterproductive in the long term and can undermine the relationship between legal authorities and those they regulate. For example, research has shown that the use of threat and legal coercion, particularly when perceived as illegitimate, can produce the opposite behavior from that sought. Such actions have been found to result in further noncompliance, creative compliance, criminal behavior, or overt opposition. In fact, a growing body of research has shown that individuals comply with the law for reasons other than an instrumental calculation of benefits and risks of offending. This research shows that most people obey most laws most of the time because they believe it is the right thing to do.

In the 1980s, therefore, many scholars began to question the value of deterrence alone in regulating behavior. Criminologists began to focus their attention on researching compliance rather than deterrence and began to realize the importance of using alternative approaches for securing compliance. One such approach can be found in the procedural justice literature. According to *procedural justice* models of compliance, institutions responsible for enforcing the law can play a major role in influencing citizens' compliance behavior through using procedurally fair practices.

Procedural justice has been shown to be extremely effective in shaping people's compliance-related behaviors because procedural justice can increase people's perceptions of an authority's *legitimacy* (Tyler 1990). If people view an authority as legitimate, they agree that the authority has the right to govern them. Providing an authority with legitimacy transfers to them the authority and right to define what constitutes acceptable forms of behavior. As such, citizens will be more willing to defer to a legitimate authority's directives or rules. While procedural justice scholarship has been conducted across many different contexts (workplaces, courts, taxation, policing, etc.), this entry focuses specifically on the role that the institution of policing can play in encouraging citizens to comply with both police directives and the law more generally.

Scholarly interest in perceptions of police treatment and the legitimate authority of the public police is not new. However, it is the contemporary stream of policing research by Tom Tyler and his colleagues that has shaped the field today. Since the publication of Tyler's (1990) influential book *Why People Obey the Law*, there has been a rapidly expanding literature around the importance of procedural justice to effective policing practice and in particular how the police can use procedural justice principles to encourage citizens to defer to their authority and comply with the law.

In the policing literature, procedural justice has usually been defined in terms of four issues: (1) neutrality, (2) respect, (3) trustworthiness, and (4) voice. First, people react positively to evidence that the police with whom they are dealing are neutral. Neutrality involves making decisions based upon consistently applied legal rules and principles and the facts of the case, not upon personal opinions and biases. Transparency or openness about how decisions are being made facilitates the belief that decision-making procedures are neutral when it reveals that decisions are being made in rule-based, principled, and unbiased ways. Second, people are sensitive to whether they are treated with dignity and politeness and to whether their rights as citizens are respected. The issue of interpersonal treatment consistently emerges as a key factor in reactions to dealings with police, with many complaints being made about police involving inappropriate verbal conduct by an officer. People believe that they are entitled to treatment with respect and react very negatively to dismissive or demeaning interpersonal treatment. Third, an authority's trustworthiness is also an important component for determining whether they have acted in a procedurally fair manner. People focus on cues that communicate information about the intentions and character of the legal authorities with whom they are dealing (i.e., "their trustworthiness"). People react favorably to the judgment that the police officers with whom they are interacting are benevolent and caring and are sincerely trying to do what is best for the people with whom they are dealing. Authorities communicate this type of concern when they listen to people's accounts and explain or justify their actions in ways that show an awareness of and sensitivity to people's needs and concerns. Finally, voice has also been shown to be important to people. People want to have an opportunity to explain their situation or tell their side of the story in a conflict. They are interested in having a forum in which they can tell their story. For procedural justice to take place, the opportunity to make arguments and present evidence should occur before a police officer makes a decision about what to do.

Early theorizing postulated that procedural justice was important to people because it was able to maximize their instrumental gains. For example, having the opportunity to express one's opinion through voice can increase one's control over outcomes and may maximize the probability of a favorable outcome. This instrumental view of procedural justice can be traced back to the seminal work of Thibaut and Walker (1975). Thibaut and Walker's control theory suggested that people do not focus directly on the favorability of the outcomes they receive from authorities. Instead, they focus on the degree to which they are able to exert influence over an authority's decisions. If people feel they have control over decisions, they will believe the procedure has been fair. In contrast, if they feel they lack control over the final outcome, they will believe the process has been unfair.

More recent social justice theories, however, have steered away from such an instrumentally focused view of individuals to suggest that procedural justice matters to people for reasons above and beyond self-interest concerns and the outcomes they receive. Lind and Tyler's (1988) Group Value Model has dominated theorizing in the procedural justice field. The Group Value Model suggests that procedural justice conveys a symbolic message that carries self-relevant implications. Drawing on psychological theories of social identity, the group value approach suggests that people are motivated to develop a positive sense of self. Unlike traditional social identity theories, however, which propose that people develop a strong sense of self through acquiring membership in high-status groups, the Group Value Model perspective suggests that one's status within a group can also add to the sense of self. According to this view, individuals will assess their status within groups by evaluating the extent to which important group representatives (such as police) treat them fairly. When people feel they are treated with procedural fairness, their sense of self-worth is bolstered, and their attachment to the group is reaffirmed. Unfair treatment, in contrast, signals marginality and exclusion and conveys to the recipient that they are not a valued member of the group.

Scholars who have examined the effects of procedural justice in a policing context have consistently found that people are much more likely to view police as legitimate when they feel they have been or will be treated in a fair, respectful, and impartial manner by police. Conducted primarily in the United States, this research shows that building police legitimacy by using procedural justice can lead to greater public respect for legal authorities and stronger felt obligation to obey the law (Tyler and Huo 2002; Sunshine and Tyler 2003). This research also shows that when police are viewed as more legitimate, citizens will be more likely to comply with the law and will be more likely to defer to police directives as a result (Murphy et al. 2008). By having legitimacy, police can encourage law-abiding behavior not just during a police-citizen encounter but also outside of encounters through everyday compliance with the law. Social theorists argue that legitimacy is effective in shaping compliance behavior because it reflects an important social value that is distinct from self-interest concerns. Tyler (1990) makes the point that when behavior flows from one's own internal values about what one *ought to do*, rewards and punishments or self-interest concerns are not the central determinants of behavior. It has therefore been argued that compliance which is garnered through a procedural justice-based approach will be economically more viable and more stable over time than instrumental deterrence-based approaches to regulation.

Inherent in most social regulation is the delivery of negative outcomes, and policing is no exception. Policing sometimes involves limiting people's ability to behave as they wish, or imposing sanctions on those who behave inappropriately. While receiving an unfavorable outcome can sometimes lead to feelings of resentment and possible defiance toward police, research consistently finds that procedural justice is more important to citizens than obtaining outcomes that they regard as fair or favorable to themselves. Research by Tyler and Huo (2002) shows just how important police use of procedural justice can be to citizens. Tyler and Huo (2002) examined survey responses from a sample of Americans who were stopped by the police. While people were somewhat more willing to accept police decisions that were favorable or that had fair outcomes, Tyler and Huo found that they were most strongly influenced by procedural fairness. People were about 15 %

more willing to accept decisions that were favorable, compared to those that were unfavorable. However, people were about 70 % more willing to accept decisions when they received fair treatment as opposed to when they received unfair treatment. This difference was found irrespective of whether the outcome was good or bad. So while both factors mattered, fairness of treatment dominated people's reactions to personal encounters with the police.

Police stops are not the only way that people have contact with the police. In fact, the most typical form of police-citizen interaction other than police stops occurs when people seek help from the police. In this type of situation, the issue is not compliance or deference to police directives but satisfaction with police efforts to help. Researchers have found that procedural justice can also be important for shaping general attitudes toward police and satisfaction with an encounter. Tyler and Huo (2002) examined people's reactions to police actions after requests for help, and the findings are essentially the same as those already noted concerning stops. People's satisfaction with police actions in response to requests for help is greater if the police solve their problems, but the primary factor shaping satisfaction is the fairness of police treatment.

Future Directions

Procedural justice researchers consistently find that citizens are more likely to comply with police directives and with the law more generally when they view police as legitimate. They also find that legitimacy is shaped predominantly by procedurally fair treatment. A number of recent developments in this field of research have occurred in the past 5–7 years which appear to be shaping scholarship in the field. These developments include (1) more sophisticated theorizing around the concept of legitimacy, (2) a divergence away from studying compliance with police and the law to understanding citizen willingness to cooperate and engage with the police, (3) understanding the contingencies surrounding procedural justice effects, and (4) implementing procedural justice and legitimate policing in applied settings. While procedural justice research in the nonpolicing field has led to other developments not mentioned here, the developments highlighted here are the ones that have emerged in the policing literature and warrant further discussion in this entry.

Conceptualizing Legitimacy

As already highlighted, procedural justice shapes satisfaction with police and compliance with police directives and the law because it increases the perceived *legitimacy* of police. Legitimacy is a judgment people make about the status of an organization as an authority that has the right to command others and be obeyed. Police legitimacy has traditionally been conceptualized in the procedural justice literature as reflecting two judgments. The first is public trust and confidence in the police. Such confidence involves the belief that the police are honest and that they try to do their jobs well and are able to protect the community against crime and violence. Second, police legitimacy reflects the public's willingness to defer to the law and to police authority. Where police are judged to be legitimate, people feel that they ought to defer to their decisions and rules, cooperate with them, and follow them voluntarily out of obligation rather than out of fear of punishment or anticipation of reward.

Tyler's empirical work in policing has dominated research in the field. His work has tended to focus specifically on the legitimacy of the police. In other words, Tyler's research has tended to conceptualize legitimacy as originating in the innate properties of the police that leads people to trust the police and feel obligated to cooperate or comply with them. Only recently, Murphy, Tyler, and Curtis (2009) have argued that past procedural justice research has overlooked how people may also perceive the legitimacy of the *laws and rules* that the police are enforcing (i.e., what they termed *legal legitimacy*). Legal legitimacy is the belief that laws should be obeyed, regardless of whether the content of those laws align with one's own moral beliefs about what is right or wrong. Murphy and colleagues suggest that individual police officers or the institution of policing more generally may be seen to have legitimate authority (with people feeling obligated to follow police directions), but the legitimacy of the policies, rules, and laws that the police enforce can be called into question. They argued that if people question the legitimacy of the laws they are being asked to obey, then compliance with laws or voluntary cooperation with police may be less likely (see also Murphy and Cherney 2012).

Using Australian survey data collected across three different regulatory contexts (including policing), Murphy and her colleagues (2009) found that the perceived legitimacy of the law did indeed influence people's compliance with the law and/or their willingness to cooperate with an authority. Those who questioned the legitimacy of the law were found to be less likely to comply. More importantly, however, perceptions of legal legitimacy were found to moderate the effect of procedural justice on compliance. Procedural justice was found to be *more* important in shaping people's compliance-related behaviors in a positive way when they questioned the legitimacy of the law than when they saw the law as legitimate. These findings demonstrate the importance of considering a broader definition of legitimacy in procedural justice research.

Scholars from the United Kingdom have also recently recognized the importance of legal legitimacy for understanding compliance with police (see Jackson et al. 2011). However, while Jackson and his colleagues distinguish between police legitimacy and legal legitimacy, they also expand on the traditional definition of police legitimacy. They have recently defined and measured police legitimacy as a multidimensional concept with three interlinked elements: (1) obligation to obey, (2) moral alignment, and (3) legality. Obligation to obey is consistent with Tyler's definition of legitimacy, with a legitimate police force being able to garner obedience from the public. Moral alignment reflects the belief that the police and the public share broadly similar moral positions about appropriate law-abiding behavior. Jackson and his colleagues argue that legitimacy needs a moral foundation; otherwise, compliance with authority directives is blind. Jackson's third conceptualization of police legitimacy reflects whether police themselves follow their own rules. If police are seen by the public to be acting in an ethical manner or exercising their authority according to established principles, then they will be seen to be legitimate. Research examining these new notions of legitimacy has only just commenced (see Jackson et al. 2012). By exploring empirically the link between the different forms of police legitimacy and legal legitimacy, future research can begin to tease apart the separate roles that each might play in explaining compliance-related behaviors.

From Compliance to Cooperation

To be effective police need to ensure that citizens comply with the law. To maintain social order the police must be widely obeyed, and this obedience must occur in both personal encounters with police and in people's everyday compliance with the law. At the same time, however, effective policing also requires the ongoing support and *voluntary cooperation* of the public. Police cannot be everywhere at all times. Their ability to detect and deal with social disorder and crime is also dependant on the publics' willingness to cooperate and assist the police by reporting crimes and passing on information about suspicious persons, events, and other information. While the police can use strategies to enforce legal compliance or compliance with police directives, as can be seen from the information presented earlier in this entry, strategies such as these can have negative consequences for people's subsequent willingness to cooperate with police in the future. Coercive approaches can destroy altruistic cooperation almost completely, with research showing that in disadvantaged communities where police are often perceived to be heavy-handed with locals, residents generally resist talking to police, or fear contacting them to report a crime or even one's own victimization (Weitzer and Brunson 2009). The benefit to both police and society of a public that is willing to assist and support them in collaborative crime control efforts is obvious.

In a review of policing and policing practices in the United States, it was suggested that understanding how to motivate public *cooperation* with the police is the most important future research topic in policing (Skogan and Frydl 2004). Identifying approaches that can intrinsically motivate and encourage people to want to voluntarily cooperate with police has therefore been identified as an important avenue for future procedural justice research.

A recent set of studies show that fair treatment of citizens by police is particularly important for shaping such voluntary behavior. Sunshine and Tyler's (2003) study was the first empirical study to demonstrate a link between procedural justice and people's willingness to cooperate with police. Sunshine and Tyler utilized longitudinal survey data collected from 1,600 New Yorkers. They found that procedural justice shaped people's compliance with the law, their willingness to cooperate with police, and their willingness to empower police with a wider range of discretion in their duties through shaping their views of the legitimacy of the police.

Since the publication of Sunshine and Tyler's study, the emphasis on studying legal compliance in the procedural justice literature has decreased in favor of studying the impact of procedural justice and police legitimacy on willingness to voluntarily cooperate with police (Murphy et al. 2008). Such cooperation and engagement could include contacting the police to report a crime, a general willingness to assist police if asked or required, and participation in crime prevention programs. Tyler and his colleagues have also recently found that procedural justice is the primary factor shaping the Muslim's willingness to cooperate and work with US police in anti-terror policing (Tyler et al. 2010; see also Cherney and Murphy forthcoming). This emerging body of research suggests that if police adhere to principles of procedural justice in their dealings with the public, then they may be able to successfully engage people and shape their willingness to cooperate with police in a range of different matters. Identifying the situations and circumstances surrounding when people would be willing to cooperate with police, work with police in crime prevention, or report crimes to their local authorities is therefore emerging as a fruitful and important avenue for future procedural justice and police legitimacy research.

Contingencies Surrounding Procedural Justice Effects

In the past 10 years, there has been a growing amount of scholarly interest in procedural justice theory and its relevance to policing. A key concern of this body of work has been to understand the antecedents of police legitimacy and people's willingness to comply or cooperate with police. Much of the published research on procedural justice to date has aimed to establish whether Tom Tyler's model of procedural justice can be replicated. For example, researchers have begun to assess the applicability of Tyler's procedural justice model in different countries. Tankebe (2009a) made the point that the majority of procedural justice and police legitimacy scholarship has occurred in the West. Considering that the history of public-police relations can differ significantly across different countries, it begs the question of whether procedural justice may have different effects across particular sociopolitical and cultural settings. Similarly, very little procedural justice research has been conducted across different groups within the United States or other Western societies (e.g., new migrants, youth, victims vs. offenders) or whether the effectiveness of procedural justice may relate to certain forms of compliant behavior. Whether procedural justice proves to be more or less effective across these different contexts or groups remains largely untested. While the research in this area generally supports the conclusion that Tyler's model can be generalized across various groups or countries, a handful of recent studies also indicates that Tyler's model may not hold true across all conditions and contexts, or across different groups. For example, Tankebe (2009a) has shown that in the African context, instrumental factors and personal outcomes are far more important in shaping police legitimacy than procedural justice-based concerns. Murphy and her colleagues have also recently shown in the Australian context that procedural justice can be counterproductive with some ethnic minority groups (Murphy and Cherney 2011, 2012). Murphy's research has also shown that procedural justice matters more to people in police-initiated encounters but matters less when they initiate contact with police. Here, instrumental factors, such as whether the police fix a problem or solve a crime, appear to be more important (Murphy 2009). Research with victims of crime has also found that reutilization of police services is shaped by whether victims have received a favorable outcome from police in the past, not by whether they had received procedural justice from police during the investigation (Hickman and Simpson 2003). While not a complete repudiation of Tyler's model, this work does indicate that attention needs to be given to the possibility that procedural justice effects vary across groups or under certain conditions or contexts. These contingencies have only just begun to be explored and open up a major avenue for future research in the field.

Operationalizing Procedural Justice into Policing Practice

The fourth major development in the police legitimacy and procedural justice field relates to the testing and evaluation of procedural justice-based policing approaches in applied settings. The overwhelming majority of procedural justice and police legitimacy studies rely on survey methodology (for a review, see Mazerolle et al. 2010). Almost no studies in the policing context use experimental field designs that attempt to ascertain the relevance of procedural justice theory to policing practice. Hence, research that seeks to examine how procedural justice can be operationalized into police practice and that also seeks to evaluate the effectiveness of such practice on citizen behavior offers another potential and important avenue for future procedural justice work. Only a handful of studies purporting to do this have emerged in the literature.

A recent Australian example is perhaps the most impressive of the studies that has been conducted to date (Mazerolle et al. 2012a). In partnership with the Queensland Police Service in Brisbane, Australia, the study's researchers implemented a change in police procedure which adopted procedural justice elements. The study varied police actions during encounters with drivers in Random Breath Testing contacts. In Australia, a police stop for the purposes of random breath testing for alcohol is the most common reason people have contact with police. The study used a randomized field trial design to examine whether introducing elements of procedural justice into the breath testing encounter could increase perceptions of procedural justice and police legitimacy. Police in the experimental condition were provided with a scripted procedural justice protocol. This protocol was longer than the standard control protocol. In the procedural justice protocol, officers were trained to focus on four procedural justice elements: voice, neutrality, trustworthiness, and respect.

To provide voice the officer gave the stopped driver a newsletter with recent crime news from their area. They asked if the driver had any questions or suggestions about crime issues or how to conduct policing in their area. To communicate neutrality the officers explained the procedures being followed. They told the driver that they had not been singled out for the random drink driving test. Instead, they had been stopped randomly. They also explained that the purpose of the police procedure was to reduce alcohol-related accidents. To develop trustworthiness officers communicated concern for people in the community. They noted that their actions were motivated by concerns about the driver and the community. Officers indicated the number of deaths from drunk driving incidents and noted that police hated to have to tell a family that someone they care about had died. They asked the driver to help the police reduce accidents by driving carefully. After the driver took the breath test, the officers were instructed to try to end the interaction on a respectful note with a gesture of courtesy toward the driver (e.g., "thank you for maintaining your car"). The officer then invited all drivers to complete a short survey of their experiences of the encounter and to return it in a provided reply-paid envelope in their own time. In the control condition, in contrast, drivers were pulled over by a police officer and simply instructed to provide a specimen of their breath for alcohol content testing. If the recording came back negative, drivers were handed a survey and were allowed to continue with their journey. A total of 21,000 drivers were involved in the study.

Survey responses from 1,645 drivers exposed to the experimental condition and 1,102 drivers from the control condition were received. Evidence from the study suggests that the procedural justice approach yielded a number of desirable benefits. Relative to the control group who did not receive the procedural justice protocol, it increased the view that the police are fair and trustworthy and left citizens more satisfied with the encounter. It also raised levels of confidence in the police and increased subsequent willingness to cooperate with police. Hence, the police were able to implement an enforcement action while building legitimacy through procedural fairness (see also Mazerolle et al. 2012b).

While the Queensland Random Breath Testing Study involved a specific police-citizen interaction, building police legitimacy need not be linked to such a specific encounter. Police programs that aim to build police legitimacy can also apply to general police efforts to engage with the community. For example, the police can conduct meetings with people in the community to hear about community concerns. They can also communicate with the public through newsletters, bulletins, or through the media. A recent study from the United Kingdom has found that the police can communicate concerns effectively and improve confidence in police by creating and disseminating newsletters (see Hohl et al. 2010). The study found that trust and confidence in police in London was much higher among people living in suburbs that received newsletters about police engagement in their neighborhoods. For people living in neighborhoods that did not receive the newsletters, perceptions of police legitimacy were lower. While both these examples support the idea that police legitimacy can be enhanced by direct police communication showing concern for the public, more research is needed in this space to ascertain whether procedural justice-based policing can be put into practice effectively for different groups of people, or in different contexts.

Controversies

The discussion presented above suggests that procedural justice-based policing shows much promise for shaping public perceptions of police legitimacy and for shaping people’s willingness to cooperate with police and comply with the law. However, despite the apparent success of procedural justice as an effective policing strategy, research in this field is not without its limitations. Several continuing problems or controversies with this body of research remain. Three of the main issues will be highlighted here.

First, Reisig et al. (2007, p. 1006) have noted that much of the survey research in this area has inconsistently operationalized key constructs and has done little to evaluate the construct validity of existing scales. They noted that measures of procedural justice and police legitimacy have differed from one study to the next. They state that this is an important concern given the interest in procedural justice policing, the number of police-citizen interactions every day, and the need for police to secure citizen cooperation and compliance. They argued that given the lack of attention by researchers to develop reliable and consistent scales across different studies, it raises a concern as to whether the measured variables across different studies reflect the theoretical constructs of interest. If measures are inconsistent or not measuring what is intended, it may lead to misleading and inaccurate conclusions across different studies. To rectify this issue, Reisig and his colleagues suggest that more needs to be done in future research to refine the scales commonly used to study police-citizen relations and to test the key propositions of the procedural justice-based model of policing using these refined measures. If the results continue to be consistent and continue to reveal the importance of procedural justice to shaping police legitimacy and compliance-related behaviors, then practitioners and researchers can be more confident in the findings.

A second major limitation of the procedural justice literature is the tendency for researchers to (1) overemphasize the relationship between procedural justice and compliance, not only in the short term but also in the long term, and (2) to underemphasize the role of instrumental factors such as outcome favorability or deterrence in shaping behavior (for a detailed critique of the procedural justice policing literature, see Tankebe 2009b). Teasing apart whether compliance stems from procedural justice or instrumental factors in any given police-citizen encounter is difficult.

Tankebe (2009b) uses research from the field of restorative justice to highlight the first issue. He notes that one of the more robust findings from the restorative justice literature is that offenders who have experienced a restorative justice conference, relative to those who have experienced traditional court procedures, tend to evaluate the experience more positively on aspects of procedural justice. However, studies that have evaluated the long-term effectiveness of restorative justice conferences on offending behavior have tended to find that they reduce recidivism in the short term but that the effect on long-term law-abiding behavior is less conclusive. If procedural justice leads to long-term behavioral improvements, then one should expect to see conclusive evidence for both short- and long-term reduction in repeat offending among participants of restorative justice conferences.

Relevant to the second issue, it has been suggested that procedural justice scholars have tended to underemphasize the importance of instrumental factors in shaping people's views and behaviors toward authorities. By being neutral across all situations, police may be able to enhance the impression that they are being fair, but such a strategy can sometimes seem off-putting to citizens who expect a more assertive response to a problem. For example, Herbert (2006) cited the example of citizens who were calling on police in Seattle to deal with absentee landlords who failed to adequately screen or regulate their tenants and thus enabled ongoing street level crime in the area. Police officers reminded the concerned citizens that they were unable to police private property arrangements. Such a response did not allay citizens' concerns and in fact led to police being viewed as unresponsive and illegitimate. Hence, Herbert argued that in such situations, outcomes may be what matters to people and that on some occasions a procedural justice-based policing approach can "perhaps imperil police legitimacy as much as enhance it" (p. 498). As noted earlier, a growing number of studies have also shown that under some circumstances, outcomes may matter more to people than fair treatment. Understanding the context surrounding when this may be the case is important and should not be underemphasized by future research studies.

Finally, the third major criticism that has been raised in relation to the procedural justice literature is the preoccupation with the utilitarian effect of procedural justice on compliance-related behaviors. It has been noted that police practitioners and researchers often overemphasize the value of procedural justice for securing cooperation and compliance, rather than promoting it as a good in itself (Tankebe 2009a, b). The police have an intrinsic obligation to exercise their authority in a procedurally fair way, irrespective of any instrumental benefit such an approach may have in facilitating their role in maintaining social order. Both researchers and police practitioners need to keep this in mind when advocating a procedural justice-based policing approach.

Despite these criticisms, however, it is clear that procedural justice-based policing has an important role in policing practice. As the review above suggests, such an approach can serve to improve police-citizen relationships; it can improve the public's willingness to cooperate and comply with the law while at the same time being of intrinsic value in and of itself.

Related Entries

▶ Procedural Justice and Cooperation

Recommended Reading and References

Ayres I, Braithwaite J (1992) Responsive regulation. Oxford University Press, New York

Becker GS (1968) Crime and punishment: an econometric approach. J Polit Econ 76:169–217

Cherney A, Murphy K (forthcoming) Policing terrorism with procedural justice: the role of police legitimacy and law legitimacy. Aust N Z J Criminol (in press)

Herbert S (2006) Tangled up in blue: conflicting paths to police legitimacy. Theor Criminol 10:481–504

Hickman L, Simpson S (2003) Fair treatment or preferred outcome? The impact of police behavior on victim reports of domestic violence incidents. Law Soc Rev 37:607–634

Hohl K, Bradford B, Stanko E (2010) Influencing trust and confidence in the London metropolitan police. Br J Criminol 50:491–513

Jackson J, Bradford B, Hough M, Murray K (2011) Compliance with the law and policing by consent: notes on legal legitimacy and cynicism. In: Crawford A, Hucklesby A (eds) Legitimacy and compliance in criminal justice. Routledge, London

Jackson J, Bradford B, Hough M, Myhill A, Quinton P, Tyler TR (2012) Why do people comply with the law? Legitimacy and the influence of legal institutions. Br J Criminol 52:1051–1071

Lind EA, Tyler TR (1988) The social psychology of procedural justice. Plenum, New York

Mazerolle L, Bennett S, Davis J, Manning M (2010) Legitimacy and policing. Campbell Systematic Reviews, The Campbell Collaboration

Mazerolle L, Antrobus E, Bennett S, Tyler TR (2012a) Shaping citizen perceptions of police legitimacy: a randomized field trial of procedural justice. Criminology. Advanced online access doi:10.1111/j.1745-9125.2012.00289.x

Mazerolle L, Bennett S, Antrobus E, Eggins L (2012b). Procedural justice, routine encounters and citizen perceptions of police: main findings from the Queensland Community Engagement Trial (QCET). J Exp Criminol. Advanced online access doi:10.1007/s11292-012-9160-1

Murphy K (2009) Public satisfaction with police: the importance of procedural justice and police performance in police-citizen encounters. Aust N Z J Criminol 42:159–178

Murphy K, Cherney A (2011) Fostering cooperation with the police: how do ethnic minorities in Australia respond to procedural justice-based policing? Aust N Z J Criminol 44:235–257

Murphy K, Cherney A (2012) Understanding cooperation with police in a diverse society. Br J Criminol 52:181–201

Murphy K, Hinds L, Fleming J (2008) Encouraging public cooperation and support for police. Polic Soc 18:138–157

Murphy K, Tyler TR, Curtis A (2009) Does procedural justice help authorities when people question the underlying legitimacy of the law? Regul Gov 3:1–26

Reisig MD, Bratton J, Gertz MG (2007) The construct validity and refinement of process-based policing measures. Crim Justice Behav 34:1005–1028

Skogan WG, Frydl K (2004) Fairness and effectiveness in policing: the evidence. Committee to Review Research on Police Policy and Practices. Committee on Law and Justice, Division of Behavioral and Social Sciences and Education, National Research Council, The National Academies Press, Washington, DC

Sunshine J, Tyler TR (2003) The role of procedural justice and legitimacy in shaping public support for policing. Law Soc Rev 37:513–547

Tankebe J (2009a) Public cooperation with police in Ghana: does procedural fairness matter? Criminology 47:1265–1293

Tankebe J (2009b) Policing, procedural fairness and public behaviour: a review and critique. Int J Police Sci Manag 11:8–19

Thibaut J, Walker L (1975) Procedural justice: a psychological analysis. Lawrence Erlbaum, Hillsdale

Tyler TR (1990) Why people obey the law. Yale University Press, New Haven

Tyler TR, Huo YJ (2002) Trust in the law. Russell Sage, New York

Tyler TR, Schulhofer S, Huq A (2010) Legitimacy and deterrence effects in counter terrorism policing: a study of Muslim Americans. Law Soc Rev 44:365–402

Weitzer RJ, Brunson RK (2009) Strategic responses to the police among inner-city youth. Sociol Q 50:235–256

Professional Crime

▶ Women and White-Collar Crime

Profiling

▶ Criminal Investigative Analysis

Profiling Arson

Katarina Fritzon, Rebekah Doley and Ryan Bell
Australian Centre for Arson Research and Treatment, Bond University, Robina, QLD, Australia

Overview

Criminal psychological profiling is the forensic technique of inferring personal, psychological, demographic, and behavioral characteristics of offenders based on crime scene evidence. While the majority of research concerning criminal profiling has been focused on the investigation of crimes of sexual violence such as murder and rape, criminal psychological profiling is frequently described as being applicable to the investigation of serial arson crimes, and the frequency with which psychological profiling has been used in the investigation of arson crimes has been growing steadily over the past 30 years (Drabsch 2004; Kocsis 2004, 2006; Turvey 1999). This current entry reviews the growing body of literature in relation to arson profiling, especially focusing on approaches that have published a specific methodology that applies to arson, as opposed to approaches that have a generalized methodology that is not necessarily differentiated for arson.

Investigating Arson

The increased interest in the application of profiling techniques to assist arson investigation may be due, in part, to the difficulty associated with the detection and prosecution of arson offenses; in the 2003 Uniform Crime Report, the United States Department of Justice states that arson has one of the lowest clearance rates of any major crime (FBI 2003). Within the United States only 18.2 % of known arson incidents were cleared by arrest in 2010 (FBI 2011), which is comparable to other developed nations including Canada (16.4 %; Statistics Canada 2011) and New Zealand (19.6 %; Statistics New Zealand 2011). The ability, then, to classify arson offenders into discrete offender profiles is innately appealing to authorities tasked with the responsibility of investigating and apprehending arson offenders, allowing fire investigators to narrow down a suspect pool according to assumed characteristics of an offender.

Approaches to Profiling

Historically, the classification of arsonists has often been arbitrary, with categories based on simple characteristics of a fire-setting group (e.g., children) or on the basis of psychological diagnoses (Lewis and Yarnell 1951) or apparent motive (Bradford 1982; Icove and Estepp 1987). It was only in the 1990s that, following the methods and procedures utilized by the Behavioral Science Unit (BSU) of the United States FBI, a typology-based profiling system of fire-setting behavior was proposed that purported to profile offenders based on a relationship between crime scene features and various personal characteristics of the offender (Douglas et al. 1992). While the approach described by the BSU was (and remains) the most commonly used profiling framework employed by police and firefighters to classify arson offenders by investigators around the world (Stauss 1998), it has been criticized by some researchers as lacking theoretical grounding (Canter and Heritage 1990). To address this perceived gap, Canter and Fritzon (1998) attempted to apply a more rigorously scientific approach to the profiling of arson offenders, identifying statistically significant relationships between crime scene behaviors and offender personal characteristics across four main behavioral themes. Kocsis and Cooksey (2002), utilizing a procedure similar to the method employed by Canter and Fritzon, also identified significant relationships between crime scene behaviors and offender characteristics among serial arson offenders only, yielding four additional behavioral themes.

While many classification systems have been described in the literature to typify arson offenders, these three classification systems remain the only widely disseminated frameworks that purport to profile arson offenders by inferring personal and psychological characteristics based on crime scene evidence. While the three frameworks canvassed within this entry have a great deal of similarity, it is important to point out that differences between these classification systems may represent differences inherent in the theoretical background and methodology employed by the authors of each framework and may not reflect strengths or weaknesses when compared to each other. In spite of the research described, there remains a surprising dearth of empirical studies on the topic of profiling arson offenders or arson crimes from criminal investigations, and there has been little progress toward a definitive method of profiling arson. Further research is needed to validate a comprehensive classification framework that can adequately represent the range and depth of arsonist behavior and characteristics.

Motivation-Based Profiling

While criminal profiling efforts have been described in the available literature as occurring during World War II and through the 1950s, psychological profiling was only added as an adjunct to the normal instructional programs of the Behavioral Science Unit at the Federal Bureau of Investigation in the United States in 1980 (Depue 1986). In 1981, the BSU developed the Violent Criminal Apprehension Program, a system to collect, organize, analyze, and disseminate information about violent crimes, which was incorporated with the BSU in 1984, forming the National Center for the Analysis of Violent Crime (NCAVC) within the FBI (Depue). While the principal aim of the NCAVC was the profiling of sexual and violent offenders, the methodologies employed by this program were subsequently applied to the criminal profiling of arson offenses. The NCAVC is presently responsible for the largest body of research examining arson for the purpose of criminal psychological profiling; the profiling methodologies employed by the ATF and FBI in the United States and a number of law enforcement agencies abroad are advised by research produced by the NCAVC (Douglas et al. 1992; Rossmo 2000).

The central theme of the research produced by the NCAVC was the development of discrete offender typologies based on motive categories. According to the NCAVC, recognizing the motivation of an offender can allow an investigator to recognize personal traits and characteristics of an offender; these characteristics can form the basis of discrete offender profiles which can be used to help identify arson offenders or rule out non-offenders (Douglas and Olshaker 1999; Douglas et al. 1986). This research culminated in the description of six motive classifications that appeared to be most effective in identifying offender characteristics: *vandalism*, *excitement*, *revenge*, *crime concealment*, *profit*, and *extremist*. These six categories are detailed in the NCAVC's Crime Classification Manual (Douglas et al. 1992).

Vandalism-motivated arson arises from a malicious and mischievous desire to cause destruction or damage. According to Douglas et al. (1992), the typical crime scene of a vandalism-motivated arsonist will reflect the spontaneous nature of the offense. This can include evidence of poor planning (e.g., leaving evidence at the scene, such as footprints and fingerprints) and the use of direct, handheld ignition to materials available at the scene. Sapp et al. (1995) note that targets of vandalism-motivated arsonists tend to be educational facilities; however, targets tend to be those of opportunity and can include other residential properties and vegetation, which will generally fall within a small geographical area which will be known to the offender. Among fires set in secured structures, offenders will frequently gain access via windows, and evidence will often show mechanical breaking of glass. Other signs suggesting vandalism can include graffiti left at the scene, general destruction of property, and/or property missing from the scene. Time of offending generally occurs in the hours after school or work and on weekends.

Douglas et al. (1992) describe the typical offender as a juvenile male with 7–9 years of formal education. The typical offender will have a record of poor academic performance and vocational history and is likely to be known by police and have an arrest record. It is also likely that the offender will be known to school authorities as being disruptive and have problems with authority. Compared to other motivations, vandalism arson frequently involves multiple offenders although, if the fire is set by a group, one personality tends to be the leader or instigator of the group.

Excitement-motivated arson deals with fires set to resolve a need for psychological excitement that is satisfied by fire setting and can include crimes committed for histrionic/heroic needs, to satisfy sexual feelings, to satisfy a need for stimulation, or as a result of psychotic delusion. A defining feature of this category is that the offender rarely intends the fire to harm people; this is frequently reflected in the type of property reflected (e.g., grass, brush, woodland, deserted construction sites, or vacant residential properties). As this category of fire setting is motivated by excitement, offenders frequently select a location that offers a good vantage point and will frequently interject themselves into the response or investigation of the fire.

Excitement-motivated fires are frequently set on properties adjacent to areas that are known to be frequently occupied. Fires are usually set using materials on hand and, among older offenders, accelerants are frequently used. Douglas et al. (1992) note that a small number of excitement fires are motivated by sexual perversion; finding ejaculate, fecal deposits, and/or pornographic material can help identify the motivation of the fire setter.

Douglas et al. (1992) describe the typical excitement arsonist as a juvenile or young adult male with 10 or more years of formal education. The typical offender within this category is described as socially inadequate, which may contribute to their inability to attain psychological excitement via pro-social means. Serial arson offenders frequently belong to this motivational category as, frequently, the psychological needs met through the fire setting are recurring (Sapp et al. 1994). A history of police contact for nuisance offenses is prevalent, with older offenders typically having a longer record. Excitement arsonists tend to remain at the scene, although those who leave frequently return to observe the results of their fire.

Revenge-motivated arson refers to those arson offenses that are set in retaliation for any real or imagined injustice, perceived by the offender. While this type of arson can span both well-planned, one-time events or serial events, this type of arson, unlike others, is always committed in response to a real or imagined stressor. Consequently, when this motive is suspected, a detailed and accurate analysis of the relationship between the victim and the offender is crucial (Sapp et al. 1995). Douglas et al. (1992) point out that the person or property victimized by a revenge arsonist will generally have a history of interpersonal or professional conflict with the offender. The NCAVC recommends that this motive be immediately considered when a low-risk individual or target suffers arson, particularly when the arson does not appear to be part of a series.

The Crime Classification Manual lists several subtypes of revenge arson that further identify the motivation of the offender: personal retaliation, societal retaliation, institutional retaliation, group retaliation, and others. The targeted property often varies with the subtype of revenge arsonist; an arsonist from the personal retaliation subcategory, for example, is more likely to burn items of significance to their victim such as clothing, bedding, or personal effects (Douglas et al. 1992). Arsonists seeking revenge against society, however, may exhibit displaced aggression by choosing targets at random, while other offenders may retaliate against institutions such as churches, government facilities, and universities. Compared to other categories, revenge arson attacks are more likely to use accelerant and, when accelerant is used, are more likely to use excessive amounts. According to some research (e.g., Sapp et al. 1994), revenge arsonists are the most prevalent among serial arsonists (Sapp et al. 1994).

Revenge arsonists are typically adult males with 10 or more years of formal education.

If employed, they typically hold blue-collar jobs and are of lower socioeconomic status. Douglas et al. (1992) describe a revenge arsonist as having close relationships, but state that these relationships tend to be unstable or short term. A revenge arsonist will often have prior law enforcement contact for property offenses such as burglary, theft, or vandalism. Alcohol use is common prior to committing this offense. The Crime Classification Manual specifies that the revenge arsonist typically acts alone and seldom returns to the crime scene, instead preferring to establish an alibi. Importantly, the offender will frequently express a short-lived sense of relief and satisfaction following the fire.

Crime concealment arsons are arson offenses committed as a secondary criminal activity to conceal a primary criminal activity. The targeted property for this offense can vary depending on the nature of the concealment. While it is tempting to assume that an arson committed for the purpose of crime concealment can demonstrate prior planning or premeditation, fire is notoriously efficient at memorializing evidence and inefficient at destroying biological evidence such as flesh and bone. Consequently, this type of fire is rarely of the organizational or sophistication level of the primary crime and is often set during periods of great stress and emotion which can interfere with an individual's ability to calmly think through their actions.

Liquid accelerant is commonly used when committing crime concealment arson, and the fire is usually set on or at the site of the primary crime. An investigator should expect to find more physical evidence, and a fire set in a more disorganized fashion, than with other arsons; this can lead to an increased risk of incorrectly assuming more than one offender, particularly if the disorganized fire follows an apparently organized burglary or homicide (Douglas et al. 1992).

The Crime Classification Manual states that crime concealment offenders tend to be young adults who live within the surrounding community and tend to be highly mobile. Alcohol and recreational drug use are common among this type of arsonist, as is a lengthy criminal history, frequently spanning a wide range of offenses. An arsonist who uses a fire to conceal burglary or auto theft is frequently accompanied by coconspirators and virtually always leaves the scene immediately and does not return. Generally this type of arson is seen as a one-time event and does not involve serial arson, although Sapp et al. (1994) found that crime concealment accounted for 5 % of the motives reported by serial arsonists.

Profit-motivated arson refers to any arson set for the explicit purpose of achieving material or financial gain. While other motivations of arson are differentiated in that they represent an expression of a psychological need, profit-motivated arson is unique in that it is committed for financial reasons. Consequently, Douglas et al. (1992) point out that, among the motivations for arson, profit-motivated arsonists exhibit the least amount of passion for their crime. This typology describes individuals who set fires to commit fraud (e.g., to collect insurance, to conceal loss, or to liquidate inventory) as well as individuals who set fires for hire (to commit fraud for a third party).

Because the destruction of property is the goal of profit-motivated arson and because this motivation is considered the least passionate, the crime scenes of a profit-motivated arson tend to demonstrate a more organized style, containing less physical evidence and more sophisticated incendiary devices. Crime scenes typically evidence an excessive use of accelerant, and multiple seats are evident. Items of value are often removed, or substituted with items of lower quality. The Crime Classification Manual suggests that fire investigators count the clothes hangers in the closets of burned residences to determine whether items have been removed beforehand.

As with revenge-motivated arson, establishing the victimology is critical in identifying this motive. Examining who is benefiting from the commission of the crime in an apparently motiveless arson could help contribute to an offender profile, and the Crime Classification Manual suggests that an offender's pre-offense conversations may provide indications of premeditation, such as warning co-workers not to attend work that day or a sudden change in an insurance policy. Icove and Estepp (1987) report that profit-motivated arson is less suitable to traditional profiling assistance, as no psychological needs are being met or are driving this behavior. Nevertheless, Douglas et al. (1992) suggest that the typical offender in this category is an adult male with 10 or more years of formal education. This offender likely has no police records, although if the fire is set by a "torch for hire," the primary offender will likely have a prior arrest record which may include a previous arson (demonstrating experience in committing this sort of crime). Offenders typically live more than one mile from the crime scene and are frequently accompanied to the crime scene by coconspirators; most leave the scene and do not return.

Extremist-motivated arsons are committed to further a sociopolitical cause. This includes arsons set for reasons of terrorism or civil disturbance and represents a very narrow range of arson offenses committed. Although few studies to date have found a significant population of arsonists for whom extremism represents a motivation, current NCAVC research suggests that extremist-motivated arson is on the increase, particularly following the terrorist attacks of Sept. 11, 2001.

According to the Crime Classification Manual, analysis of the targeted property is essential in the determination of this motivation category (Douglas et al. 1992). In order to qualify as an extremist-motivated arson incident, the target must represent the antithesis of the offender's beliefs. Examples include research laboratories and fur stores targeted by animal rights groups, abortion clinics targeted by extremist anti-choice groups, or businesses targeted by unions. As with profit-motivated arson, the crime scene typically reflects an organized and sophisticated attack with offenders frequently employing incendiary devices and accelerants, leaving little forensic information. While profit-motivated arson is clandestine, however, extremist arson is frequently committed to deliver a symbolic message and will often be accompanied by some form of message left at the scene or via a communique delivered to the media or local authorities claiming responsibility.

Extremist-motivated arson is difficult to profile due, in part, to the rarity of the crime in comparison to other arson offenders. Extremist arsonists are often more sophisticated offenders and will typically be well educated, literate, and politically active or aware. Of particular relevance to an arson investigator, the extremist offender is frequently readily identified with a cause or group, and interviews with the friends and family of a suspected arsonist may identify membership into a group. Offenders may have previous police contact for civil disobedience, trespassing, criminal mischief, or civil rights violations (Douglas et al. 1992).

A central premise of the profiling research described by the NCAVC is that inferences concerning an offender's motivation can be made from crime scene evidence and that a psychological profile can then be constructed from this motivation (Douglas and Burgess 1986; Douglas and Olshaker 1999). While this assumption has remained implicit within the profiling literature produced by the NCAVC, there is little empirical evidence suggesting that offender motivation can be successfully inferred from crime scene behaviors or that a psychological profile of an offender can be generated from an offender's motivations (Kocsis 2006). Additionally, while the motivational categories (and associated profiles) of arsonists are presented by Douglas et al. (1992) as representing discrete offender profiles, research investigating the clinical utility of these motivational categories suggests that arsonist motivations are frequently multidimensional. In a review of the motivations of 212 convicted serial arsonists, Sapp et al. (1994) found that 12 % of offenders reported multiple motivations spanning two or more motivational categories described by the NCAVC, and similar research has suggested that individual motivations for fire setting tend to be complex and multifaceted: in one landmark study, Lewis and Yarnell (1951) reported that elements of revenge were present in some form in nearly all cases of arson! While the motivations of an offender carrying out a crime would certainly be of interest to an investigator attempting to develop a profile, in light of the multidimensional nature of offender motivations, the number of inferences that can be drawn from this information may be limited.

The Action Systems Approach to Arson Profiling

According to Canter and Heritage (1990), the basic tenet of profiling is that offenders differ in the behaviors they exhibit when committing a crime and that there are associations between offender characteristics and the thematic focus of their crime. Within this framework, criminal profiling is developed solely on the classification of offense behaviors independent from any inferred motivations. Implicit within this framework is the assumption that offenders will show consistency between the nature of their crimes and other characteristics they exhibit in other situations. This is referred to by Canter (e.g., Canter and Youngs 2009) as the A-C equation and is a central tenet of the Investigative Psychology approach to offender profiling. While a single criminal action may be insufficient to make inferences concerning an offender characteristic, a group of actions together may represent a strong predictor of an offender's style.

Using a multivariate statistical methodology, Salfati and Canter (1999) examined the crime scene actions of a sample of 82 homicide offenders who victimized strangers and compared them to known characteristics of those offenders and found that those offenders who made an effort to conceal their crime (e.g., by hiding or transporting the body or theft of non-identifiable property) were more likely to have a previous custodial sentence. Profiling studies utilizing this model thus seek to thematically link crime scene actions to offender characteristics, to develop criminal profiles of offenders.

Canter and Fritzon (1998) explored the crime scene behaviors from 175 solved cases of arson in the United Kingdom for evidence of behavioral themes indicative of discrete offender profiles. Using a combination of multidimensional Scaling and content and thematic analysis, four behavioral themes emerged, and these were interpreted according to Shye's (1985) Action System Model of Functioning. The four themes were correlated with background characteristics of the known offenders. Using this method, four profiles were identified, which varied according to the apparent motive of the fire setting (instrumental or expressive) and the target of the fire setting (object or person). Thus, the four profiles described by Canter and Fritzon were termed *instrumental-object* (arson committed in order to satisfy general criminal goals), *instrumental-person* (arson committed in order to extract revenge against another person), *expressive-object* (arson committed to provide emotional comfort or satisfaction), and *expressive-person* (arson committed to attract attention).

Instrumental-object arson is described by Canter and Fritzon (1998) as any arson set as an "attempt to change aspects of the object where the change will be of direct benefit to the fire setter." Fires set to conceal other crimes would fall under this category; examples include a burglar setting fire to a residence to hide forensic evidence or a business owner setting fire to company records to hide corporate malfeasance. Aspects of this type of arson have a great deal in common with the crime concealment motivational category described by Douglas et al. (1992) as well as the arson-for-profit classification described by Barnett (1992) and Vreeland and Lewin (1980). These arsons are seen as opportunistic, in that the decision to commit the crime may only be arrived at when the individual recognizes the environmental possibilities. The offense is viewed as merely fulfilling general criminal goals or satiating a need to commit damage or vandalism and thus has little personal meaning to the offender.

This arson profile is associated with arson that is set by multiple offenders. It frequently occurs on uninhabited properties and co-occurs with other property crimes such as theft or vandalism. As the goal of the arson tends to be property motivated (as in the destruction of property for gain is the explicit goal), it is rarely associated with fires set in a manner where death or injury is likely to occur. Spree arsonists (i.e., those who set fire at three or more separate locations with no emotional cooling-off period between them) are associated with this profile.

Instrumental-person arson is described as arson that occurs "as a reaction to frustration by another person which the fire setter wishes to hurt or remove" (Canter and Fritzon 1998). This type of arson generally emerges as a consequence of a dispute between the offender and another person, and, as a result, a specific triggering event can usually be identified by the offender. In this case, fire serves as a method with which the offender can retaliate, in order to address feelings of injustice on the part of the offender. Consequently, this type of arson is frequently committed against items of symbolic significance to the victim such as important property, heirlooms, or photographs.

Unsurprisingly, this arson profile is associated with arson that occurs following an argument and following threats of arson. As this arson occurs in response to an argument or conflict, the victim is usually known to the offender and is frequently a close friend or romantic partner. Vehicular arson is associated with this arson profile as is evidence of planning including the use of accelerants.

The third profile of arson described by Canter and Fritzon (1992) is termed *expressive-object* and refers to arson which involves "the demonstration of aspects of the arsonist on the external world." Canter and Fritzon draw a comparison between this category and Geller's (1992) description of fire setting as a means of emotional acting out and suggest that fire setters within this profile will choose targets with symbolic, emotional significance. This type of fire is more likely to proceed in a serial manner, suggesting an inherent fascination with fire or the trappings of fire. This hypothesis is further supported by the observation that fire setters within this profile are more likely to remain at their crime scene to observe their fires.

As this profile is associated with the use of fire for emotional release, it is associated with offenders who remain at the scene. It is also associated strongly with individuals who commit multiple fires (Canter and Fritzon 1998). As this profile is hypothesized to represent offenders who set fires for emotional reasons, nonspecific but identifiable triggers are associated with this profile as offenders may set fires to cope with emotionally charged events. Additionally, this profile is associated with the target of commercial buildings and hospitals which may cause more attention to be focused on the individual or on the event, corresponding to the emotional symbolism the fire may hold to the offender.

The final profile described by Canter and Fritzon (1998) is termed *expressive-person* and refers to arson which is "turned inward to lead to the disintegration of the fire setter him/herself." Broadly, this category refers to arson committed for the purpose of self-harm or suicide or arson committed for the purpose of seeking attention from family or authorities. Canter and Fritzon point out that, typically, this style of arson is addressed in a therapeutic context rather than a forensic context and is thus seldom described within the available forensic literature. It can frequently occur in the context of mental illness, particularly depression or psychosis.

This profile is associated with the use of fire as an avenue to commit some sort of self-harm so, unsurprisingly, it is associated with suicide notes left at the scene of the crime, is frequently directed at the self, and frequently occurs in the offender's own home. Due to the mortal intent of this profile, offenders frequently set fires in a manner where injury or death is likely to occur, and offenders are likely to remain at the scene (presumably to ensure they succeed in their intent).

In a pair of follow-up studies, the four-profile model proposed by Canter and Fritzon has been validated among a sample of adult prisoners (Almond et al. 2005) and a sample of juvenile fire setters (Santtila et al. 2003). Kocsis and Cooksey (2002) point out, however, that the model proposed by Canter and Fritzon was conducted using a sample containing both serial and one-time arson offenders. While the study had enormous implications to the field of arson profiling, several researchers (e.g., Geberth 1996) have pointed out that the technique of profiling has been found to have much more utility in the investigation of serial offenses in general, as non-recidivistic crimes are typically solved via regular investigative procedures.

The Profiling of Serial Arson Offenders

Kocsis and Cooksey (2002) examined 148 incidents of arson committed in New South Wales and Victoria in Australia, in order to determine the extent to which offense behaviors could be used to classify offenders into discrete offender profiles from which offender characteristics could be inferred. Although the methodology, data analysis, and aims were similar to the study conducted by Canter and Fritzon (1998), all cases of arson examined by Kocsis and Cooksey were each part of a series of three or more arsons committed by a serial arson offender. It was thus a methodological aim of Kocsis and Cooksey to develop offender profiles that could be applied to recidivistic arson, on the assumption that development of these profiles would provide greater clinical and investigative utility than the profiles developed through previous research efforts. Using multidimensional scaling techniques and cluster analysis to examine 71 crime scene behavior variables, a total of four offender profiles were generated: *thrill*, *anger*, *wanton*, and *sexual* fire setters.

Kocsis and Cooksey (2002) describe the *thrill pattern* as encompassing those arsons that are set to satisfy nonsexual needs of stimulation and excitement. Within this profile, offenders are not represented to hold any animosity toward their target or victim; rather, they are attracted to the element of risk associated with the act (hence the label of "thrill" for this pattern). This profile can be likened to the "Excitement" profile of arson described by Douglas et al. (1992) although, while Douglas et al. group arsons which sublimate sexual needs in this category, Kocsis and Cooksey specify that sexually motivated arson represents a separate and distinct psychological profile and use the term "recreational fantasy" to describe the theme found within this arson profile. This profile will include arsonists such as firefighters who set fires in order to increase activity for themselves, teenagers who set fires in order to increase sensory arousal, and individuals who set fires in order to interject themselves with the firefighting effort.

Common crime scene behaviors associated with this pattern include variables associated with behavioral sophistication and planning. Offenders will usually be quite mobile, often choosing their target beforehand and travelling more than one mile to set their fire. While accelerants and materials will frequently be brought to the scene in order to create a larger fire, because fire tends to be the goal (rather than the destruction of property or harming another party), targets tend to be bush, grassland, or vegetation where risk of harm to other people will be minimized. As with the excitement profile described by Douglas et al. (1992), offenders within this profile will frequently set their fire in highly visible locations and will frequently remain on the scene (and may be subsequently identified and apprehended). A series of fires set by an offender within this profile will usually have less crime scene behaviors in common from scene to scene than a series committed by other profiles, as offenders within this profile are more likely to vary their choice of targets, methods of ignition, and other crime scene behaviors.

A number of offender characteristics are associated with this arson profile. Thrill arsonists are more likely than other arson offenders to set fires with co-offenders, are more likely to be socially affluent (e.g., they are less likely to live alone and more likely to be involved in social and romantic relationships), and are more likely to be employed. This follows from the observation that this pattern of arson is often committed to create excitement for a group of offenders (as with the example of a fire fighter setting a fire to provide action for his department or a teenager setting a fire with a group of friends). Interestingly, Kocsis and Cooksey (2002) point out that arsonists following this pattern frequently have poor dental work and some type of outstanding physical feature such as scarring according to information contained within the offenders police files. Kocsis and Cooksey surmise that, because these offenders are not physically attractive, their fires may represent the sublimation of possible sexual drives.

The second profile described by Kocsis and Cooksey (2002) is termed the *anger pattern* and comprises arsons in which animosity or rage forms the principal motivation for the offense. In contrast to the thrill pattern, the fire is not the principal focus of this offense; rather, the fire is a vehicle through which the offender can facilitate personal harm to a specific victim. Consequently, offending behaviors associated with this pattern demonstrate an intent to cause specific and personalized harm to a victim. This profile is similar to the *instrumental-person* profile described by Canter and Fritzon (1998) and the *revenge-motivated* arsonist described by Douglas et al. (1992). Kocsis and Cooksey point out, however, that both the instrumental-person and revenge profiles described in previous studies require an identifiable relationship between the offender and the victim. Within the anger profile, the relationship between the offender and the victim is based more on cognitive knowledge or familiarity. Within this profile, offenders may not harbor any previous animosity toward the victim (or have a relationship at all!), but may attack due to a perceived relationship. This pattern, thus, is analogous to sexual murderers or serial rapists, who may choose their victim by type or perceived relationship. The common element between these offenders and arsonists within the anger profile is the presence of an unfocused internal rage. Offenders who commit arson in retaliation for a perceived slight or serial arsonists who target potential romantic partners who they perceive would reject them would both be included within this pattern.

Crime scene behaviors indicative of this pattern include evidence that the offense was designed to cause personal harm to the victim. Common within this profile is the tendency for offenders to physically destroy possessions of the victim prior to committing the offense, particularly items of personal significance to the victim or offender (e.g., photographs, souvenirs, gifts from other individuals). Targets of this pattern of arson commonly include residential properties or motor vehicles, which is unsurprising given the personal nature of this offense. It is also likely that the intended target of the offense will evidence a break-in and that the fire will be started from within the structure. The intent to cause harm will also be evident in the choice of conveyance for the fire; offenders within this profile will frequently use trailers to ensure a thorough spread of the fire.

Kocsis and Cooksey (2002) noted that, compared to other profiles, offenders within this profile tend to be foreign nationals who are more likely to be bi- or trilingual and to possess a foreign accent. As the goal of this pattern of arson is generally the causing of harm to someone else rather than any fascination with the fire itself, offenders within this profile rarely remain on the scene. While Kocsis and Cooksey do not comment on the expected proportion of serial arsonists that may belong to each profile, other researchers (e.g., Sapp et al. 1994) have noted that anger and revenge appear to be the most frequently cited motive among serial arsonists studied.

The third profile described by Kocsis and Cooksey (2002) is labelled the *resentment pattern* and includes arsons set in a generalized sentiment of animosity. Like the anger classification, this profile contains an element of displaced aggression; however, the anger displayed within this category will typically not be directed at a specific target – rather, it will be directed at a vague classification. This profile has some thematic similarity to the vandalism-motivated arson described by Douglas et al. (1992) and the instrumental-object profile outlined by Canter and Fritzon (1998), in that all three profiles refer to an unfocused nature of attack. Consequently, many of the observations made of offenses committed within this profile are similar to the observations made of vandalism and instrumental-object offenders. Examples of offenders who would be included within this profile are juveniles setting fire to a school or areas of bushland.

Within the model proposed by Kocsis and Cooksey (2002), the principal crime scene behaviors indicative of fires falling within this profile are the choice of target. School properties, state-owned properties, and commercial properties are strongly associated with this profile, which is consistent with target choices within the vandalism profile outlined by Douglas et al. (1992).

This choice of targets may be reflective of a general sentiment of displaced aggression against public institutions, which would support the hypothesis that fires within this profile represent a sublimation of anger and resentment toward society. Examination of the MDS plot generated by Kocsis and Cooksey suggests that the majority of behaviors associated with this pattern of arson are deliberate and planned and are committed against more secured targets with access restricted.

Few individual offender characteristics are associated with this profile. Offenders within this profile typically possess a prior criminal record, which follows the hypothesis that these offenders will possess a general animosity toward society, and offenders may present with a pattern of authority issues and general antisocial actions. As with vandalism-motivated arson, individuals following the animosity pattern will light fires on weekends. While vandalism-motivated arson is described by Douglas et al. (1992) as largely being committed by juveniles, Kocsis and Cooksey (2002) do not comment on the anticipated age of this sort of offender, although they note that "school fire arson" (in which educational facilities are attacked by juveniles) is overrepresented within this profile. While the researchers do indicate that an element of animosity or resentment is a necessary ingredient for an offender to be included in this profile, Kocsis and Cooksey suggest that this profile warrants further study, in order to determine *why* this animosity develops.

The final profile outlined by Kocsis and Cooksey (2002) was termed the *sexual profile*; this embodies an offense style associating the ignition of fire with sexual excitement and gratification. The inclusion of this pattern is consistent with early psychodynamic literature (e.g., Freud 1932; Macht and Mack 1968) which typified recidivistic arson as being principally motivated by repressed sexual drives. The most distinguishing feature of this pattern of arson is evidence of sexual activity by the offender at or near the scene of the crime. This profile of arson offending is difficult to reconcile with the profiles furthered by earlier research, which may be attributed to the choice of offenders sampled for Kocsis and Cooksey's study (i.e., only serial arsonists, rather than serial and one-time arsonists). The authors do note, however, that the sexual profile does bear some resemblance to Cantor and Fritzon's (1998) *expressive-object* pattern, in that both profiles are concerned with committing an offense to achieve emotional relief. Additionally, while the excitement profile described by Douglas et al. (1992) describes arsonists who offend in order to achieve excitement, sexual excitement is included as a subclassification of excitement arsonists in the Crime Classification Manual; it may be the case that sexually excited offenders from within this profile are more likely to reoffend, resulting in their being represented in greater numbers within Kocsis and Cooksey's sample.

As expected, the crime scene behaviors with the strongest association with this profile are any evidence of sexual activity such as semen left at the scene, sexual aids (e.g., pornographic material), or fecal matter. Common targets include state-owned property on easily accessible premises or outdoor settings. This is consistent with the hypothesis that offenders will commit this crime in order to attain sexual release; offenders will typically choose targets which are easily accessible and highly visible, so that they can be watched to attain the required level of stimulation. Consequently, offenders will frequently remain at the crime scene in order to observe the fire or participate in its extinguishing. As the focus tends to be on the fire itself rather than the destruction of property, these arson attacks tend to be small in size and rarely escalate to major attacks. Common targets include trash receptacles, post office boxes, and public toilets.

Few offender characteristics are noted for this profile although Kocsis and Cooksey (2002) do point out that, compared to other offender profiles, arsonists following this pattern tend to have dark-colored hair and eyes – although the authors point out that these variables were included in the analysis in order to provide richness of data and were not designed to racially stereotype any group of offenders. Offenders are also likely to have a history of domestic travel, although they typically do not travel far to commit their offense, instead choosing to remain in areas familiar to them.

While the research conducted by Kocsis and Cooksey (2002) identified four major profiles associated with recidivistic arson offenders, a novel finding of their research was a cluster of arson crime scene behaviors that seemed to be common among all arson offenses committed by the offenders within their sample. Termed the *common behavior profile*, this profile provides a core description of the characteristic behaviors common to patterns of serial arson. While the four profiles outlined by Kocsis and Cooksey provide important clues to differentiate serial arson offenders from each other, the common behavior profile can provide important clues to differentiate serial arson offenders from one-time or non-arson offenders.

A central element of all serial arson offenses reviewed by Kocsis and Cooksey (2002) is the elements of planning, an observation matched by the finding that a relationship between the victim and the offender existed in nearly all cases of arson committed in the sample. This finding is inconsistent with the profiles outlined by Douglas et al. (1997) (notably the vandalism and crime concealment profiles) and the instrumental-object profile furthered by Canter and Fritzon (1998), with both studies reporting that planning was not an essential component of an arson offense. Kocsis and Cooksey address this incongruence, however, by pointing out that planning may be an element differentiating serial arsonists in particular and that crime scenes demonstrating planning of the offense may indicate the work of a serial arsonist. Kocsis and Cooksey also point out that serial arsonists across all profiles commonly enter their targets and steal items of value before setting their fire, a constellation of behaviors suggesting a greater disregard for the risk of apprehension or detection. This is in contrast to other crime modalities, such as murder or rape, in which offenders are typically deterred by greater risk of apprehension.

Future Directions and Conclusions

Though underdeveloped, the field of profiling arson offenders appears to hold a number of exciting possibilities for future advancement. In spite of the profiling systems described to date, however, no system has been developed that adequately represents the range and depth of behaviors and characteristics of arsonists. This may be partly attributed to a number of methodological and practical considerations that limit investigation into arson profiling. Within this section, we will outline what we believe are the most significant problems in this area and make suggestions for further advancement within this field.

A concern inherent in the study of arson profiling (and, in fact, arson literature in general) is an innate sample bias. As mentioned previously, arson has a low clearance rate, and what is known about arsonists pertains only to those fire setters who are identified and apprehended and conclusions drawn from such a sample may not be truly representative of the arsonist population. It is worth noting that the profiles developed by Canter and Fritzon (1998) and Kocsis and Cooksey (2002) were derived from an analysis of solved arson offenses in the United Kingdom and Australia, respectively, and the profiles generated by the NCAVC were developed based on case reviews of solved arson offenses (Douglas et al. 1992). While using solved cases represents the only method that can be used to marry crime scene details to offender characteristics, it is impossible to determine the extent to which these profiling efforts can be generalized to unsolved offenses. Future research may address this concern by collecting data from individuals imprisoned for other offenses but who may have fire-setting histories for which they were not apprehended. As investigative techniques evolve, improving the clearance rate of arson and contributing to a larger (and more representative) sample size of solved arson cases, the present frameworks may be found to have inadequate empirical depth to account for the relationships between offender characteristics and crime scene behaviors.

It is important to consider, also, the origins of any classification system developed to profile offenders. Doley (2003) points out that asking questions such as who collected the data and why, the type of sample the information was drawn from, and for what purpose the typology was originally intended for are important for identifying any potential weaknesses that are inherent within that system. While the profiles developed by Kocsis and Cooksey (2002) differ in several fundamental ways from the profiles generated by Canter and Fritzon (1998), the authors point out that this is likely due to their decision to only include serial arson offenders in their sample and the typologies generated by this approach are only applicable to the profiling of recidivistic arson offenders. While this methodological concern may limit the extent to which Kocsis and Cooksey's approach can be generalized to the profiling of arson offenders generally, it also suggests potentially fruitful avenues of future research, for example, further studies investigating the extent to which non-serial arson offenders can be profiled using this framework. Similarly, while the profiles generated by the BSU have been criticized by some researchers as lacking theoretical grounding (Canter 1989, 1994), the methodology used to generate these profiles is the same as the methodology used to generate all profiling frameworks currently used by the United States FBI (Dietz 1987; Rossmo 2000). Certainly, these profiles do result in a series of testable hypotheses and further research could be conducted to determine the extent to which these profiles can be used to profile offenders.

While the research canvassed within this entry represents a series of admirable attempts to provide a framework for the profiling of arson offenders, understanding of this field is underdeveloped and limited to only a handful of published studies. Given the amount of theory proliferation in other areas of offender profiling in recent years, this dearth in the literature is especially remarkable. In spite of a handful of published frameworks purporting to profile arson offenders, research validating these frameworks is virtually nonexistent and crime scene investigators, professional profilers, and mental health professionals are forced to employ profiling methodologies which are not empirically validated. Criminal profiling is frequently cited as being helpful in the criminal investigation of violent and sexual crimes by criminal investigators (Geberth 2006; Holmes and Holmes 1992, 2000, 2008; Turvey 1999). Thus, it is likely that professionals would benefit substantially from further development in the field of arson profiling.

Related Entries

- ▶ Behavioral Investigative Advice
- ▶ Criminal Investigative Analysis
- ▶ Criminal Profiling
- ▶ Investigative Psychology

Recommended Reading and References

Almond L, Duggan L, Shine J, Canter D (2005) Test of the arson action system model in an incarcerated population. Psychol Crime Law 11:1–15

Barnett W (1992) Psychology and psychopathology of firesetting, 1955–1991 – a review. Fortschr Neurol Psychiatr 60(7):274–286

Bradford JMW (1982) Arson: a clinical study. Canadian Journal of Psychiatry, 27:188–192

Canter D (1989) 'Offender Profiling', The Psychologist, Vol. 2, pp. 12–16

Canter D (1994) Criminal shadows. Harper Collins

Canter D, Fritzon K (1998) Differentiating arsonists: a model of firesetting actions and characteristics. Leg Criminol Psychol 3:73–96

Canter D, Heritage R (1990) A multivariate model of sexual offence behaviour: developments in 'offender profiling. J Forensic Psychiatry 1(2):185–212

Canter D, Youngs D (2009) Investigative Psychology: Offender profiling and the analysis of criminal action. Chichester: John Wiley & Sons

Depue RL (1986) An American response to an era of violence. FBI Law Enforcement Bull 55(12):2–5

Dietz PE (1987) Sex offender profiling by the FBI: A preliminary conceptual model. In Ben-Aron MH, Hucker SJ, Webster CD (eds) Clinical criminology: The assessment and treatment of criminal behaviour. Toronto: University of Toronto

Doley R (2003) Making sense of arson through classification. Psychiatry, Psychology and Law 10(2):346–352. DOI:10.1375/pplt.2003.10.2.346

Douglas JE, Burgess AW (1986) Criminal profiling: a viable instigative tool against violent crime. FBI Law Enforcement Bull 55(12):9–12

Douglas J, Burgess AW, Burgess AG, et al. (1997) Crime Classification Manual: A Standard System for Investigating and Classifying Violent Crimes (revised edn). Jessey Bass

Douglas JE, Olshaker M (1999) The anatomy of motive. Scribner, New York

Douglas JE, Ressler RK, Burgess AW, Hartman CR (1986) Criminal profiling from crime scene analysis. Behav Sci Law 4(4):401–421

Douglas JE, Burgess AW, Burgess AG (1992) Crime classification manual. Lexington Books, New York

Drabsch T (2004) Arson. New South Wales parliamentary library, briefing paper no 2/03

Federal Bureau of Investigation (FBI) (2003) Crime in the United States 2002: uniform crime reports. U.S. Department of Justice, Washington, D. C

Federal Bureau of Investigation (FBI), United States (2011) The uniform crime reporting statistics. Department of Justice webpage. http://www.ucrdatatool.gov/. Retrieved 14 Feb 2012

Freud S (1932) The acquisition of power over fire. International journal of psychoanalysis, 13:405–410

Geberth VJ (1996) Practical homicide investigation, 2nd edn. CRC Press, New York

Geberth VJ (2006) Practical Homicide Investigation: Tactics, Procedures, and Forensic Techniques, 4th edn. Boca Raton, FL: CRC Press

Geller JL (1992) Arson in review: from profit to pathology. Clin Forensic Psychol 15:623–645

Holmes R, Holmes S (1992) Understanding mass murder: A starting point. Federal Probation, 56, 53–61

Holmes R, Holmes S (2000) Murder in America, 2nd edn. Thousand Oaks, CA: Sage

Holmes R, Holmes S (2008) Profiling Violent Crimes: An Investigative Tool, 4 edn. Thousand Oaks: Sage Publications, Inc

Icove DJ, Estepp MH (1987) Motive-based offender profiles of arson and fire-related crimes. FBI Law Enforcement Bull, April, 1987

Kocsis RN (2003) Criminal psychological profiling: validities and abilities. Int J Offender Ther Comp Criminol 47(2):126–144

Kocsis RN (2004) Psychological profiling of serial arson offences: an assessment of skill and accuracy. Crim Justice Behav 31(3):341–361

Kocsis RN (2006) Criminal profiling: principals and practice. Humana Press, Totawa

Kocsis RN, Cooksey RW (2002) Criminal profiling of serial arson crimes. Int J Offender Ther Comp Criminol 46(6):631–656

Lewis NOC, Yarnell H (1951) Pathological firesetting (Pyromania). Nervous and mental disease monograph. Coolidge Foundation, Nicholasville, vol. 82

Macht LB, Mack JE (1968) The firesetter syndrome. National Emergency Training Center

Rossmo DK (2000) Geographic profiling. CRC Press, Boca Raton

Salfati CG, Canter D (1999) Differentiating stranger murders: profiling offender characteristics from behavioural styles. J Behav Sci Law 17:391–406

Santtila P, Hakkanen H, Alison L, Whyte C (2003) Juvenile firesetters: crime scene actions and offender characteristics. Leg Criminol Psychol 8:1–20

Sapp AD, Gary GP, Huff TG, James S (1994) Motives of arsonists aboard naval ships. J Police Crim Psychol 10(1):8–13

Sapp AD, Huff TG, Gary GP, Icove DJ (1995) A motive based offender analysis of serial arsonists. National Center for the Analysis of Violent Crime, Quantico

Shye S (1985) Non-metric multivariate models for behavioural action systems. In: Canter D (ed) Facet theory: approaches to social research. Springer Verlag, New York

Statistics Canada (2011) Juristat crime statistics. http://www.statscan.gc.ca/pub/85-002-w/index-eng.htm. Accessed 14 Feb 2012

Statistics New Zealand (2011) Crime in New Zealand: 1996–2005. Tatauranga Aotearoa, Wellington

Stauss, E. (1998). A death sentence for four victims: How Keller manipulated the fire service. Fire and Arson Investigator, 48(3):16–22

Turvey BE (1999) Criminal profiling. Academic, San Diego

Vreeland RG, Lewin BM (1980) Psychological aspects of firesetting. In: Canter D (ed) Fires and human behaviour. Wiley, New Work, pp 31–46

Propensity Score Matching

Ben Grunwald and John M. MacDonald
Department of Criminology, University of Pennsylvania, Philadelphia, PA, USA

Overview

Multivariate regression-based methods dominate the research literature in criminology for estimating the effects of causal variables. As many scholars have noted, however, regression suffers from a number of methodological and practical limitations (Berk 2004). On its own, regression provides a poor analytic framework for conceptualizing causal inference, offers limited tools to address nonlinear relationships, and encourages researchers to cherry-pick models in a manner that may bias the results of a study. Despite these issues, criminologists have failed to take full advantage of other approaches to observational research developed in the fields of statistics and biostatistics.

One alternative is propensity score matching (Rosenbaum and Rubin 1983), an analytic method closely associated with the Rubin Causal Model (RCM; Rubin 2008; Holland 1986). Though a number of criminological studies have used propensity score approaches (e.g., Haviland et al. 2007; MacDonald et al. 2007; Paternoster and Brame 2008), it remains an underused strategy for causal estimation. This entry draws heavily from the work of Rubin and Rosenbaum to outline the RCM and propensity score matching and to demonstrate their application to theoretical and applied criminology (see, e.g., Rosenbaum and Rubin 1983; Rosenbaum 2002, 2010; Rubin 2008).

This entry begins with a brief overview of the RCM and propensity score matching. It then describes an implementation of propensity score matching to answer a traditional question of criminology: what is the effect of gang membership on violent offending? The final section provides a methodological assessment of the advantages of propensity score matching relative to traditional regression-based methods. For many applications, the former is a superior approach that provides substantial benefits over the regression-based methods typically used in criminology.

The General Framework of the Rubin Causal Model and Propensity Score Matching

For most of the twentieth century, the frameworks guiding experimental (Fisher 1935; Cochrane and Cox 1950) and observational studies (Blalock 1964; Kenny 1979; Cochran 1983) were distinct from one another. Social science methodologists developed quasi-experimental approaches to generating causal inference in observational studies, but these involved little mathematical formalization (Campbell and Stanley 1963). The two domains of experimental and observational research were formally unified in the 1970s in the field of statistics by the development of the RCM (Rubin 1974). For current purposes, the RCM contains two relevant parts: (1) the potential outcomes framework and (2) the concept of an assignment mechanism (Rubin 2008).

The potential outcomes framework provides a conceptual approach for understanding causal inference. A causal effect is understood as the difference between the value of an outcome variable subject to different treatments. For example, the causal effect of gang membership on violent offending for some adolescent *i* is the difference between the number of violent crimes that adolescent would commit if he were part of a gang $(Y_i|T_i = 1)$ and the number of violent crimes the same adolescent would commit if he were not part of a gang $(Y_i|T_i = 0)$. It is expressed formally:

$$(Y_i|T_i = 1) - (Y_i|T_i = 0) \qquad (1)$$

Unfortunately, for any single individual, one can only observe the outcome variable subject to the treatment assignment that he or she actually receives; one cannot observe the counterfactual outcome, the outcome one would have observed had the individual not received the treatment. Without observing the counterfactual outcome variable, it is impossible to calculate the treatment effect for any particular individual.

Though calculating the counterfactual outcome for any single individual is out of the question, it is possible to estimate the mean of the counterfactual outcome variable for a sample of individuals: the mean of the outcome variable in the control group can serve as the mean of the counterfactual outcome variable for the treatment group. By calculating the difference between the mean outcome variable for subjects who receive the treatment $(E(Y_i|T_i = 1))$ and the mean outcome variable for subjects who do not receive the treatment $(E(Y_i|T_i = 0))$, one can estimate the mean causal effect of the treatment. It is expressed formally:

$$E(Y_i|T_i = 1) - E(Y_i|T_i = 0) \\ = E([Y_i|T_i = 1] - [Y_i|T_i = 0]) \tag{2}$$

Whether Eq. 2 produces an unbiased estimate of the mean treatment effect depends upon a strong assumption that, aside from treatment assignment, subjects in the treatment and control groups do not differ systematically on variables that influence the outcome. Imagine, for example, that drug use increases offending independently of gang membership and that drug use is twice as common among gang members as non-gang members. Under these circumstances, the treatment and control group differ systematically on an important pretreatment covariate, and the control group cannot provide an unbiased estimate of the mean counterfactual outcome variable for the treatment group.

Randomized experiments are the strongest method for ensuring that the treatment and control group are balanced on pretreatment covariates. In an experiment, the assignment mechanism – the process by which subjects are assigned to the treatment or control group – is explicitly manipulated by the researcher. The researcher randomly assigns subjects to a treatment or control to ensure that subjects' background characteristics do not influence their probability of treatment assignment. In a typical experiment with one treatment and one control group, subjects are given a p probability of assignment to the treatment, and a $1 - p$ probability of assignment to the control. It is expressed formally:

$$\pi_i = \Pr(Z_i = 1) = p \tag{3}$$

where π_i represents some constant probability p of receiving the treatment ($Z_i = 1$) for all subjects i in the study. Where Eq. 3 holds, the treatment and control groups are balanced on all pretreatment covariates, and Eq. 2 provides an unbiased estimate of the treatment effect.

Importantly, Eq. 3 is a sufficient but not necessary condition for unbiased causal estimation. The key feature of a randomized experiment is not that all subjects share the same π_i, but rather that the researcher *knows* the π_i for all subjects in the sample. As long as π_i is known, it is easy to perform statistical adjustments that account for variations in π_i across subjects in the sample, and balance the treatment and control group on pretreatment covariates. For example, if heavy drug users are twice as likely as non-users to receive the treatment, then the influence of heavy drug users in the treatment group can be down weighted by one-half, assuring that heavy drug use will be balanced in the treatment and control groups for purposes of estimating the treatment effect. A number of other ways to balance the treatment and control groups are available including perfect matching, case control designs that match subjects on different groups, and propensity score matching (demonstrated in this entry).

Observational research refers to empirical studies in which the researcher does not control the treatment assignment through randomization. Instead, subjects are assigned to the treatment or control based upon the natural course of their lives. Observational research thus faces two related problems. First, it is unlikely that all subjects share the same probability of treatment assignment, π_i. It is expressed formally:

$$\pi_i = \Pr(Z_i = 1|\mathbf{x}_i, \mathbf{u}_i) \tag{4}$$

where $\mathbf{x}_i$ and $\mathbf{u}_i$ refer, respectively, to the set of observed and unobserved covariates that render subjects more or less likely to receive the treatment. Equation 4 implies that Eq. 3 may not hold. As a result, without statistical adjustment, an estimate of the treatment effect in Eq. 2 may be biased. The second and related problem faced in observational research is that the researcher does not know π_i for all subjects in the study. Given that Eq. 3 may not hold, without information about π_i, it is difficult to perform the statistical adjustments necessary to compute an unbiased causal estimate.

The RCM conceives of observational studies as randomized experiments in which information about the assignment mechanism, π_i, is lost to the researcher (Rubin 2008). The RCM divides observational research into two key stages. In the design stage, the researcher attempts to recover the lost information about the assignment mechanism by modeling the selection process into the treatment and control groups. In the analysis phase, the researcher uses the model of the selection process to make statistical adjustments to the analysis of the outcome variable, thereby ensuring that imbalances on observed pretreatment covariates $\mathbf{x}_i$ do not bias the estimate of the treatment effect.

A number of methods are available for implementing the design and analysis phases of the RCM. Following Rosenbaum and Rubin (1983), this entry focuses on one popular method often called propensity score matching. Propensity score matching has two key steps. In the first step, logistic regression is used to estimate the relationship between violent offending and observed pretreatment covariates $\mathbf{x}_i$ for the purpose of computing $\hat{\pi}_i$, the estimated probability of treatment assignment. In the context of propensity score matching, $\hat{\pi}_i$ is often called a propensity score. In the second step, subjects in the treatment group are matched with subjects in the control group that share the same propensity score, $\hat{\pi}$. One common strategy is to pair one unique subject in the treatment group with one unique subject in the control group. It is expressed formally:

$$\hat{\pi}_j = \hat{\pi}_k = \Pr(Z = 1|\mathbf{x}_i, \mathbf{u}_i) \qquad (5)$$

where subjects j and k are matched subjects in the treatment and control group, respectively, and $\mathbf{x}_i$ and $\mathbf{u}_i$ refer, respectively, to the set of observed and unobserved covariates that render subjects more or less likely to receive the treatment. Other constant matching ratios, such as 2:1 or 3:1, and even variable matching ratios are possible. All subjects in the sample that cannot be matched are excluded from the dataset for the remainder of the study.

Of course, matching subjects on $\hat{\pi}_i$ does not ensure that matched pairs share precisely the same set of pretreatment covariate values. How, then, does matching on $\hat{\pi}_i$ balance the treatment and control group on pretreatment covariates $\mathbf{x}_i$? Matched pairs share the same propensity score, $\hat{\pi}_i$. Thus, differences in pretreatment covariate values among matched pairs do not help predict which subject is more or less likely to receive the treatment. Both theory and practice suggest that, as a result, the distribution of pretreatment covariates $\mathbf{x}_i$ will be roughly balanced in the treatment and control group in the matched dataset (Joffee and Rosenbaum 1999).

The analysis phase of the RCM proceeds by analyzing the matched dataset from the design phase as though it were generated in a randomized experiment. A t-test or Wilcoxon ranked sum test of violent offending in the treatment and control groups can check for statistically significant differences between the groups.

How well can propensity score matching overcome systematic bias from imbalances on pretreatment covariates? The power of propensity score matching is limited by the availability of observed data on pretreatment covariates that influence the true propensity score, π_i. Propensity score matching can only remove systematic bias that arises from imbalances on observed pretreatment covariates, $\mathbf{x}_i$, and not from unobserved pretreatment covariates, $\mathbf{u}_i$. The assumption that differences between the treatment and control group on unobserved variables $\mathbf{u}_i$ do not bias the estimate of the treatment effect is often referred to in the literature as the ignorability assumption.

Some diagnostic methods are available to assess concerns about systematic bias caused by unobserved covariates, $\mathbf{u}_i$. Rosenbaum (2009), for example, describes a useful method for assessing Γ, the level of bias that $\mathbf{u}_i$ would need to produce in order to reverse the findings of the study. This design sensitivity analysis is described in more detail below.

Implementing Propensity Score Matching: An Example

This section describes the application of propensity score matching to a traditional question of criminological theory: what is the effect of gang membership on violent crime? No substantive theoretical conclusions should be drawn from the analysis. The goal is to demonstrate the role of propensity score matching within the RCM, rather than to conduct a rigorous test of theory.

A two-wave dataset on 294 male and 48 female adolescents from Los Angeles County, California, is used to test the effect of gang membership on violent offending. The data were originally collected for the RAND Corporation's Adolescent Outcomes Project (AOP) (Morral et al. 2004). The AOP was designed to assess the effects of a community-based substance use treatment program on adolescents who were assigned to a group home by a probation official. The eligibility criteria for the research subjects are described in detail by Morral et al. (2004: 259).

The observed pretreatment covariates, $\mathbf{x}_i$, are taken from a baseline survey conducted when the subjects were 12 and 13 years old. Demographic data on gender and race were collected. Data on general satisfaction, orientation toward problems, substance use, quality of health, days worked in the last 90 days, and days spent in jail in the last 90 days were also collected. Two additional indexes were collected that measure peer delinquency and family criminality. Finally, two variables capturing the number of violent crimes (robbery, physical fight with and without injuries to victim, sexual assault) and property crimes (destruction of property, forgery, theft, car theft, breaking and entering) committed in the previous 90 days, and a variable indicating subjects' gang membership in the previous 90 days were collected. All of these pretreatment covariates are included as predictors in a logistic regression on gang membership.

Follow-up interviews were conducted 72 months after baseline. Only two variables from that wave are used in the analysis. First, subjects were asked whether they were involved in a gang in the previous 90 days. However, since the primary aim of the discussion is to demonstrate an implementation of the RCM, the data is treated as though it covers the entire 72-month period for clarity of presentation. Subjects that reported gang involvement in the previous 72 months are assigned to the treatment group, and subjects that did not report involvement are assigned to the control group. Second, a variable measuring the number of violent crimes committed in the previous 90 days serves as the outcome variable.

The Design Phase

The primary purpose of the design phase in the RCM is to model the assignment mechanism for subjects in the sample. Once estimated, this information helps remove systematic bias from the estimate of the treatment effect by balancing the treatment and control group on observed pretreatment covariates, $\mathbf{x}_i$. This section demonstrates how propensity score matching implements the design phase in two steps. First, propensity scores, $\hat{\pi}_i$, are estimated by modeling gang membership between baseline and follow-up on observed pretreatment covariates, $\mathbf{x}_i$, collected at baseline. Second, each gang member is matched with one unique non-gang member that shares the same $\hat{\pi}_i$. This matching strategy balances the distribution of $\hat{\pi}_i$ in the treatment and control groups and tends to balance the distribution of $\mathbf{x}_i$ as well.

Though the technical details associated with propensity score matching are somewhat complex, the procedure is easy to implement in a number of statistical packages. Rosenbaum (2009), for example, provides an excellent explanation of how propensity score matching can be implemented in R.

Part I: Estimating Probability of Treatment Assignment

Here, $\hat{\pi}_i$ is estimated through multivariate logistic regression, though other parametric and nonparametric approaches to modeling treatment assignment are available (see Hirano and Imbens 2001; McCaffrey et al. 2004 for other estimators). The following model is used:

$$\log\left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right) = \alpha_i + \mathbf{b}_i\mathbf{x}_i \qquad (6)$$

where $\hat{\pi}_i$ is the estimated propensity score of gang membership for subject i, $\mathbf{x}_i$ is a set of observed pretreatment covariates theoretically linked with gang membership, and $\mathbf{b}$ is a vector of logistic coefficients.

Table 1 displays descriptive statistics on key measures used in the logistic regression. Two kinds of Wilcoxon ranked sum tests were used to compare the treatment and control group on covariates $\mathbf{x}_i$. First, the unadjusted ranked sum test compares the distribution of a covariate x in the treatment ($n = 80$) and control group ($n = 262$). Since each treatment subject will be matched with one unique control subject, $262 - 80 = 182$ cases in the control group will be dropped from the study. As a result, statistical power to detect differences between the groups will drop dramatically after matching. It is important to assess the marginal increase in balance after those cases are dropped, but such a comparison using statistical tests is only meaningful if the tested sample sizes are similar. It can, therefore, be helpful to perform a second adjusted ranked sum test in order to compare the marginal increase in balance on $\mathbf{x}_i$ after those cases are dropped. This can be accomplished by selecting 500 samples, each consisting of 80 randomly selected cases (with replacement) from the control group. A Wilcoxon ranked sum test is then conducted on covariate $\mathbf{x}_i$ for each of the 500 samples, comparing them against the treatment group. The final p-value for covariate x is computed by the mean of the 500 p-values from the 500 ranked sum tests. The absolute standardized mean difference between the treatment and control group for $\mathbf{x}_i$ is displayed later in Table 3.

Propensity Score Matching, Table 1 Descriptive statistics on pretreatment covariates x_i for gang members (treatment) and non-gang members (control) between baseline and 72-month follow-up

Covariate	Treatment mean (SD)	Control mean (SD)	Unadjusted/ adjusted p-values
Male	0.96 (0.19)	0.83 (0.38)	**0.003/0.025**
Race			
Black	0.14 (0.35)	0.16 (0.36)	0.68/0.584
Hispanic	0.62 (0.49)	0.53(0.5)	0.123/0.296
Other	0.15 (0.36)	0.14 (0.34)	0.778/0.582
General satisfaction	2.33 (0.8)	2.22 (0.73)	0.199/0.367
Problem orientation	0.78 (1.57)	0.97 (1.69)	0.379/0.495
Substance abuse	0.47 (0.3)	0.52 (0.32)	0.203/0.393
Days of work	9.41 (22.13)	9.57 (20.21)	0.767/0.59
Days in jail	40.49 (30.52)	36.68 (27.15)	0.512/0.546
Violent crime	5.35 (14.4)	3.11 (9.06)	0.513/0.536
Property crime	11.91 (27.36)	10.78 (28.31)	0.516/0.536
Delinq. peers	2 (0.92)	1.71 (0.86)	**0.011**/0.091
Family criminality	0.31 (0.23)	0.26 (0.22)	0.067/0.214
Gang membership	0.52 (0.5)	0.26 (0.44)	**0/0.004**
Sample size	88	270	

Table 1 provides evidence of imbalance on the pretreatment covariates. Perhaps most strikingly, nearly twice as many subjects in the treatment group were involved in gangs at baseline than the subjects in the control group. Moreover, 13 % more of the treatment group is male than of the control group, and members of the treatment group are associated with having more delinquent peers. Though not statistically significant, the treatment group also has more violent crime and higher family criminality. These differences suggest that failing to balance the treatment and control group on pretreatment covariates may lead to a biased estimate of the effect of gang membership on violent crime. Propensity score matching is an effective method for balancing a large number of pretreatment covariates.

Propensity Score Matching, Table 2 Results of the logistic regression predicting gang membership between baseline and 72-month follow-up

Variable	Odds ratio	Confidence interval	*P* value
Intercept	0.02	(0, 0.1)	**0**
Male	5.38	(1.77, 23.7)	**0.009**
Race			
Black	0.73	(0.23, 2.41)	0.6
Hispanic	1.27	(0.52, 3.47)	0.623
Other	1.35	(0.45, 4.31)	0.599
General satisfaction	1.17	(0.78, 1.77)	0.44
Problem orientation	1.03	(0.84, 1.26)	0.74
Substance abuse	0.32	(0.1, 0.96)	**0.045**
Days of work	1	(0.99, 1.01)	0.956
Days in jail	1	(0.99, 1.01)	0.999
Violent crime	1.02	(0.98, 1.05)	0.321
Property crime	0.99	(0.98, 1.01)	0.488
Delinq. peers	1.5	(1.05, 2.16)	**0.027**
Family criminality	2.63	(0.79, 8.75)	0.113
Gang membership	2.3	(1.26, 4.23)	**0.007**

Propensity Score Matching, Table 3 d_x and d_{xm} for propensity scores and pretreatment covariates

	d_x	d_{xm}	*P* value
Propensity Score	0.82	0.03	0.928
Male	0.45	0	1
Race			
Black	−0.05	0	1
Hispanic	0.2	−0.08	0.623
Other	0.04	0	1
General satisfaction	0.14	0.03	0.705
Problem orientation	−0.12	0.28	**0.025**
Substance abuse	−0.16	0.14	0.361
Days of work	−0.01	0.12	0.547
Days in jail	0.13	0.05	0.985
Violent crime	0.19	0.18	0.883
Property crime	0.04	0.2	0.209
Delinq. peers	0.32	0.19	0.234
Family criminality	0.23	−0.08	0.606
Gang membership	0.55	−0.03	0.876
Total mean of absolute values	0.23	0.09	0.665

Equation 6 generates an estimate of subjects' propensity scores, $\hat{\pi}_i$. Table 2 reveals that several variables in $\mathbf{x}_i$ are significant predictors of gang membership. Males, for example, are more likely to be gang members than females (OR = 5.4; 95 % CI = 1.8, 23.7). Quite unsurprisingly, adolescents with peers that engage in delinquency are more likely to be gang members (OR = 1.5; 95 % CI = 1.1, 2.2), as are adolescents who use illegal substances (OR = 0.32; 95 % CI = 0.1, 0.96). Prior gang membership is also a strong predictor of current gang membership. Adolescent gang members at baseline are over twice as likely to be gang members between baseline and follow-up (OR = 2.3; 95 % CI = 1.3, 4.2).

A comparison between the treatment and control group on $\hat{\pi}_i$ confirms the expectation that $\hat{\pi}_i$ is not constant for all subjects in the study. Figure 1 shows that $\hat{\pi}_i$ in the treatment group is substantially higher than in the control group. Any estimate of the treatment effect that fails to take into account the observed pretreatment differences might suffer from substantial systematic bias. Fortunately, there is sufficient overlap between the two distributions to find credible matches in the control group for subjects in the treatment group.

Part II: Matching Treated with Controls

In this section, each treated subject is matched with one unique control on $\hat{\pi}_i$. The basic matching strategy proceeds as follows. First, a distance matrix is constructed with 282 columns and 80 rows. Each column is associated with one unique subject in the control group, and each row is associated with one unique subject in the treatment group. Each cell in the matrix represents the distance in estimated propensity scores, $\hat{\pi}_i$, between the $80 * 282 = 22,560$ possible pairs of treated and controls in the sample. The distance, $\hat{d}_{ij}$, is computed as the squared difference in $\hat{\pi}_i$:

$$\hat{d}_{ij} = (\hat{\pi}_{mT} - \hat{\pi}_{nC})^2 \qquad (7)$$

where $\hat{\pi}_{mT}$ refers to the estimated propensity score for subject m in the treatment group and

Propensity Score Matching, Fig. 1 Propensity scores in treatment, matched control, and control groups

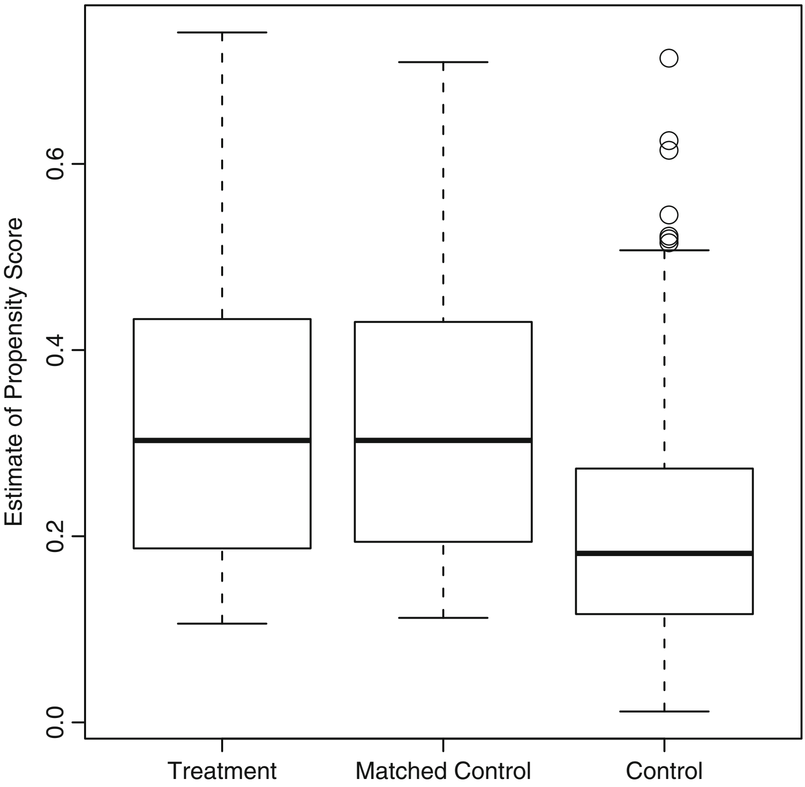

$\hat{\pi}_{nC}$ refers to the estimated propensity score for subject n in the control group.

In 1-to-1 matching, each subject in the treatment group is matched with one unique subject in the control group to minimize the total distance, $\sum \hat{d}_{ij}$, between the matched pairs. The 80 controls selected as matches for subjects in the treatment group are referred to as the matched control group. The remaining 182 unmatched controls are set aside for the remainder of the study.

Having outlined the basic strategy for propensity score matching, other methods are available to improve pretreatment balance including (1) an alternative distance matrix, (2) the rank-based Mahalanobis distance, and (3) exact match adjustments.

Matching on $\hat{\pi}_i$ tends to achieve balance between the treatment and control group on pretreatment covariates, $\mathbf{x}_i$, without ensuring any similarity in $\mathbf{x}_i$ within matched pairs. An alternative distance matrix improves upon traditional $\hat{\pi}_i$ matching by requiring that pairs with suitably similar $\hat{\pi}_i$ also have similar $\mathbf{x}_i$ (Rosenbaum and Rubin 1985). An alternative matrix executes this more sophisticated matching strategy by making adjustments to the original distance matrix described above. In the alternative matrix, all cells with distances exceeding some specified upper bound – referred to as the caliper – are set to infinity, ∞. Pairs associated with cells set to ∞ cannot be matched. Rosenbaum (2010) suggests that one fifth of the standard deviation of $\hat{\pi}_i$ is an adequate caliper value. The present study uses a caliper of 1/20 of the standard deviation of $\hat{\pi}_i$ as it produces closer balance on $\mathbf{x}_i$.

After having set to ∞ all cells in the original distance matrix that exceed the caliper, an alternative distance matrix calculates the Mahalanobis distance for all remaining cells that are not set to ∞. The Mahalanobis distance has two key features: first, a standardized difference of a given size across all uncorrelated variables in $\mathbf{x}_i$ has the same effect on the

Mahalanobis distance. Second, the Mahalanobis distance accounts for correlation between variables. In calculating distances between pairs, two variables that are highly correlated (e.g., with an $r^2 = 0.95$) effectively count as one variable (Rosenbaum 2010).

To calculate the Mahalanobis distance for any pair of individuals, the covariance matrix for observed pretreatment covariates, $\Omega(\mathbf{x}_i)$, is first computed. Second, the vector $(\mathbf{x}_{mT} - \mathbf{x}_{nC})$, where $\mathbf{x}_{mT}$ and $\mathbf{x}_{nC}$ refer to the vectors of pretreatment covariates for treatment subject m and control subject n, is computed. And third, the inverse of $\Omega(\mathbf{x}_i)$ is pre- and post-multiplied by the vector $(\mathbf{x}_{mT} - \mathbf{x}_{nC})$:

$$d_{mahal} = (\mathbf{x}_{mT} - \mathbf{x}_{nC})^T \Omega(\mathbf{x}_i)^{-1} (\mathbf{x}_{mT} - \mathbf{x}_{nC}) \quad (8)$$

Not all data employed in the present study is distributed normally. To address this issue, the rank-based Mahalanobis distance is used (Rosenbaum 2010). The rank-based distance replaces the pretreatment covariate values with ranks to avoid the possibility of unusual behavior in the Mahalanobis distance for variables with extreme outliers.

The alternative distance matrix is an advance over the original distance matrix because it ensures greater similarity in $\mathbf{x}_i$ within matched pairs. Under some circumstances, however, it is desirable to place an even more stringent restriction of a perfect match on some subset of variables. Exact matching can be executed by adding a penalty to any cell in the alternative distance matrix in which both individuals do not share the same value. Perfect matching is not used on any pretreatment covariates in this example.

After computing an alternative Mahalanobis distance matrix, one critical step remains. Each subject in the treatment group must be matched with one unique subject in the control group in such a way that minimizes the sum of the distance between matched pairs, $\sum \hat{d}_{ij}$. Best-first or greedy algorithms rarely minimize $\sum \hat{d}_{ij}$. In contrast, Bertsekas' (1981) optimal pair match algorithm provides an efficient method to identify the set of matched pairs which minimize $\sum \hat{d}_{ij}$. Bertsekas' algorithm can be implemented in R using Hansen and Klopfer's (2006) "optmatch" package.

After forming the matched control group, it is important to assess the balance achieved on $\hat{\pi}_i$ and on $\mathbf{x}_i$. Figure 1 reveals impressive balance between the treatment and matched control group on $\hat{\pi}_i$. Welch's two-sample t-test ($p = 0.861$) and the Wilcoxon ranked sum test ($p = 0.928$) detect no statistically significant difference between the two distributions.

To assess post-match balance on pretreatment covariates, $\mathbf{x}_i$, one can calculate the mean of covariate x in the treatment group, m_{tx}, the control group, m_{cx}, and the matched control group, m_{mcx} (Rosenbaum and Rubin 1985; Haviland et al. 2007). One should also calculate the standard deviation of x in the treatment group, s_{tx}, and the control group, s_{cx}. Finally, an overall standard deviation, s_x, for all subjects in treatment and control is also calculated:

$$s_x = \sqrt{\frac{s_{tx}^2 + s_{cx}^2}{2}}. \quad (9)$$

Balance on covariate x is evaluated using d_x and d_{xm}. d_x represents the absolute standardized mean difference between the treatment group and control group on variable x. An x that is well balanced between the treatment and control group will have a d_x value close to zero:

$$d_x = \frac{|m_{tx} - m_{cx}|}{s_x} \quad (10)$$

d_{xm} represents the absolute standardized mean difference between the treatment group and matched control group on variable x. An x that is well balanced between the treatment and matched control group will have a d_{xm} value close to zero:

$$d_{xm} = \frac{|m_{tx} - m_{mcx}|}{s_x} \quad (11)$$

Table 3 provides data on d_x and d_{xm} for the observed pretreatment covariates, $\mathbf{x}_i$. The matching strategy increases balance on all but two of the covariates. Problem orientation has the greatest post-match imbalance, but even this variable is imbalanced by only a quarter of a standard deviation $(0.28 * s_x)$. Notably, matching decreased the mean of the absolute value of all the d_x values by over 50 % (from 0.23 to 0.09). Wilcoxon ranked sum tests detect only one significant difference in the distribution of $\mathbf{x}_i$ in the treatment group and matched control group. One statistically significant difference among 10–15 covariates would be expected, even in an experiment where assignment to the treatment and control group is fully randomized.

The above analysis suggests that propensity score matching has effectively balanced the treatment and control group on matched pretreatment covariates, $\mathbf{x}_i$.

The Analysis Phase

In this section, the matched dataset produced in the design phase is analyzed. First, a new variable, *dif*, is created by subtracting the outcome variable of each treatment subject from the outcome variable of the respective matched pair in the control group. Second, a parametric estimate of the treatment effect is calculated by the mean of *dif* and by conducting a t-test to assess statistical significance. The parametric estimate indicates that gang membership increases violent offending by an average of 1.3 in a 90-day period. This is a relatively large treatment effect, but no statistically significant difference is detected because of the small sample size and the long-tailed structure of the distribution (p-value = 0.202).

Given the dramatic deviation from the normal distribution, the treatment effect is also estimated with the Hodges-Lehmann estimator and the Wilcoxon signed rank test (Hollander and Wolfe 1999). The Hodges-Lehmann estimator finds a larger and statistically significant effect size of 2.5 violent crimes in a 90-day period (p-value = 0.002).

It is worth wondering what results would have been produced by more traditional methods of multivariate regression. Controlling for the same set of pretreatment covariates, multivariate ordinary least squares regression estimates that gang membership significantly increases violent offending by 1.8 ($p = 0.006$). Propensity score matching, thus, estimates a causal effect that is nearly 40 % larger than the linear regression model.

Similarly, controlling for all the pretreatment covariates, Poisson regression estimates that gang membership multiplies the expected count of violent offending by (B = 0.995; p-value = 0.003) (e^B=) 270 %. Using the matched pair dataset in which all pretreatment covariates are balanced, the Poisson regression estimates that gang membership multiplies the expected count of violent offending by (B = 0.594; p-value = 0.278) (e^B=) 181 %. Assuming the propensity score approach models gang membership correctly, traditional multivariate regression shows clear evidence of biased coefficients, either understating (ordinary least squares) or overstating (Poisson) the size of the gang membership effect on violent offending.

Sensitivity Analysis of Unobserved Bias

If all pretreatment covariates that affect the probability of joining a gang are observed in $\mathbf{x}_i$, it is likely that the design phase matched subjects in the treatment group with subjects in the control group that share the same true probability of treatment assignment, π. In most observational studies, however, it is naïve to assume that all relevant pretreatment covariates are observed. Unobserved variables, $\mathbf{u}_i$, may influence the probability of treatment assignment, and as a result, matched pairs will not share the same probability of treatment. Under these circumstances, some or all of the observed treatment effect may be explained by systematic omitted variable bias rather than the effect of the treatment. It is, therefore, important to consider how large the systematic bias would need to be in order to reverse the statistically significant findings of the study.

Rosenbaum (2009) describes a design sensitivity analysis for assessing the resistance of empirical findings to systematic bias. Γ refers to the ratio between the odds of assignment to the treatment group for a matched pair. If a matched pair shares the same π, then Γ is equal to 1. If Γ equals 1 for all matched pairs, then the treatment and matched control groups are balanced and the observed treatment effect is unbiased. The design sensitivity analysis operates by determining the highest level of Γ at which the findings cannot be reversed by systematic bias. For example, if $\mathit{\Gamma} = 2$ is assumed, then the sensitivity analysis calculates every possible p-value that could be produced when $\mathit{\Gamma} = 2$. If none of the p-values are greater than 0.05, then the observed effect of the treatment is resistant to a level of bias represented by $\mathit{\Gamma} = 2$. See Rosenbaum (2010) for a more in-depth and technical treatment of the procedure.

Here, the observed effect of gang membership on violent offending remains statistically significant for a value of $\mathit{\Gamma}$ as large as 1.5. Thus, the findings are moderately resistant to systematic bias: even if members of the treatment group are 50 % more likely to receive the treatment than their matched pair, the observed effect of gang membership on violent offending remains statistically significant. The observed findings are not resistant to a level of bias greater than $\mathit{\Gamma} = 1.5$.

Assessing the Propensity Score Matching and Regression

This section compares the advantages and disadvantages of propensity score matching and traditional regression-based methods in criminology. For the current purposes, the category of traditional regression-based methods will refer primarily to the parametric regression models that dominate the criminological literature.

The advantages are divided into two categories: methodological and practical. Methodological advantages arise from the inherent mathematical properties of the analytic methods; these are advantages that accrue to a particular method when it is implemented appropriately and in accordance with its own assumptions. The discussion of practical advantages takes into consideration that analytic methods are not always implemented appropriately. Using propensity score matching in the context of the RCM provides the practical advantage of discouraging deviations from accepted scientific practice.

Methodological Advantages of Propensity Score Matching

Propensity score matching has a number of methodological advantages over conventional regression-based methods.

First, propensity score matching uses a stronger method of controlling for confounding covariates. Regression uses indirect adjustment. The adjustment is indirect because regression estimates the effect of confounding covariates and conditions out those effects while estimating the effect of the causal variable of interest. As a simple example, an ordinary least squares regression of gang membership on violent offending would simultaneously estimate how average offending covaries with gang membership and drug use. To estimate the effect of gang membership holding drug use constant, regression simultaneously estimates the effect of drug use on gang membership. In contrast, propensity score matching uses direct adjustment, the same general strategy as in randomized experiments: rather than controlling for the effects of confounding covariates, direct adjustment balances the distribution of those covariates in the treatment and control groups. As long as there are an equal number of heavy drug users in the treatment and control group, for example, heavy drug use is balanced and will not bias the estimate of the effect of gang membership. Direct adjustment is stronger than indirect adjustment because, rather than adjusting for the *unknown and estimated effects* of observed covariates, it adjusts the *known distribution* of those covariates.

Second, direct control enables propensity score matching to better account for nonlinear relationships between controlled covariates and the outcome variable. Regression models can account for nonlinear relationships through functional transformations, but it is difficult to determine whether all nonlinearities have been appropriately modeled. In contrast, nonlinear confounders that are balanced in the treatment and control group do not bias the observed effect of the treatment. This is a major benefit of propensity score matching: whereas regression only accounts for nonlinear relationships that are known and are accurately modeled, propensity score matching accounts for all nonlinearities associated with observed pretreatment covariates when balance is achieved through matching.

Third, the strategy of indirect control renders linear regression susceptible to problems of multicollinearity that are not faced by propensity score matching. Adding multiple correlated covariates into a regression model can inflate standard errors and diminish the detection of statistically significant differences. Propensity score matching circumvents this methodological challenge by avoiding the need to disentangle and estimate the causal effects of specific covariates, and instead simply removes them from the estimated treatment effect by balancing the covariates themselves.

Fourth, propensity score matching can be combined with efficient causal estimators through the use of nonparametric estimation and statistical tests. With some exceptions, regression-based methods in criminology typically entail parametric assumptions about constant and normally distributed errors. Nonparametric regression-based methods are available, but they are rarely used. Some criminological data satisfy these strong assumptions, and other data can be transformed to satisfy them. For these data, the nonparametric Wilcoxon signed rank test is 95 % efficient relative to the t-test commonly used in regression (Lehmann 1998; Haviland et al. 2007). The Wilcoxon signed rank test is substantially more efficient for other distributions, especially skewed distributions typical of criminal offending. Since the true conditional distribution of the outcome measure is rarely known, it makes sense to use an estimator with good performance across different kinds of distributions. Propensity score matching can be combined with nonparametric estimators and statistical tests that perform well across different kinds of distributions.

Fifth, propensity score matching provides a simple method for estimating different kinds of causal treatment estimates, or perhaps more precisely, for estimating causal estimates for treatments in different subpopulations. This is particularly valuable in criminology where the kind of treatment effect estimated is often not considered. Calculating the treatment effect on the treated with regression, for example, requires more computation than simply using the coefficient of the treatment indicator. Propensity score matching can estimate the *effect of the treatment on the treated* (TOT) by retaining all subjects in the treatment group and matching them to one or more subjects in the control group. TOT analyses are especially useful for calculating the causal effect of a treatment that is currently experienced by small and nonrepresentative subsets of the population. Gang membership, for example, is a common experience among young, urban, and historically disadvantaged ethnic minorities, but is uncommon for much of the rest of the population. It is, therefore, perhaps more interesting to examine the effect of gang membership on this specific segment of the population than on the entire population. On the other hand, sometimes it is more important to study the effect of the treatment on individuals who did not receive the treatment. The *effect of the treatment on the untreated* can be calculated by retaining all subjects in the control group and by matching them with one or more members of the treatment group. Though rare, this kind of causal analysis can be helpful when extending a treatment for a subpopulation to the rest of the population. Finally, the *effect of the treatment on the population* (TOP) can be roughly estimated by balancing the treatment and control group to an estimate of the population mean. Linear regression can compute these different kinds of causal estimates, but not without careful efforts in designing the structure of the study. Propensity score matching, on the other hand, calculates these causal estimates with ease by adjusting the matching criteria.

Aside from the modeling advantages already described, propensity score matching has two further methodological advantages. The central problem in observational research is omitted variable bias: the researcher is always left wondering whether some unknown confounding variable leads to unobserved bias in the causal estimate of the treatment. Propensity score matching offers a design sensitivity analysis, which measures the level of unobserved bias necessary to reverse the findings of the study. When that level is high, the data provide more confidence that differences observed between the treatment and control groups are the result of the true effect of the treatment, and not the result of omitted variable bias. Research on smoking offers a strong example of the value of design sensitivity analyses. It is difficult to study the long-term effects of smoking on health through a randomized experiment; thus, observational research is the main avenue for exploring the effects of smoking. Observational studies on smoking come with the usual caveats about unobserved systematic bias. Some observed effect sizes associated with smoking, however, are so strong that it would be implausible to attribute them entirely to unobserved bias. Cornfield et al. (1959), for example, observe that the level of bias necessary to reverse the effect of smoking on cancer would be extremely high (Rosenbaum 2009). Effect sizes in criminology are rarely so large, but that only substantiates the need for design sensitivity analyses. Greater awareness that systematic bias can easily explain some research findings may provide a healthy reality check to a research community focused on small effect sizes and statistical significance.

Finally, propensity score matching permits researchers to control for noncausal variables such as race or gender without making assumptions about the causality of covariates. Zuberi (2001) has argued that in most social science research, race cannot be conceptualized as a causal variable. This is not a trivial methodological problem. Race and other demographic factors correlate strongly with a wide range of crucial social variables, and it is important to ensure that differences in the distribution of those variables do not bias causal estimates in observational research (e.g., Berk et al. 2005). Propensity score matching provides a practical solution to the problem. By controlling for differences in racial groups without measuring the "effect of race," propensity score matching enables social scientists to control for differences across race without problematic assumptions about the "causality" of noncausal variables.

Practical Advantages of Propensity Score Matching

The prior section argued that the RCM and propensity score matching provide important methodological advantages over traditional regression-based approaches. This section suggests that the RCM and propensity score matching also encourage researchers to adhere more closely to sound scientific practice.

First, propensity score matching makes fewer assumptions than the traditional forms of regression that dominate the criminological literature. Propensity score matching, for example, does not necessarily assume (1) linear effects between individual control variables and the outcome variable, (2) normally distributed errors, or (3) homoscedasticity (constant errors). Propensity score matching also requires no test for multicollinearity. Finally, using a nonparametric statistical estimator largely obviates the need to test for high-leverage points. Articles in criminology journals rarely indicate whether key assumptions of regression models are satisfied, but they are critical for estimating unbiased parameters. All of the assumptions needed for regression are less important for propensity score matching. In matching, the researcher primarily needs to check for balance between the treatment and control group on the estimated propensity scores, $\hat{\pi}_i$, and on pretreatment covariates $\mathbf{x}_i$.

Second, propensity score matching implemented in accordance with the RCM is more resistant to the influence of researcher bias. It is well known that a researcher's beliefs and attitudes can influence the findings of a study. Three features of regression-based methods leave the door wide open to these biases (Greiner 2008). First, assessing the goodness of fit of a regression model is partially a subjective judgment, and researchers must often rely upon their own expertise and judgment. Second, adjustments to the model intended to improve fitness – such as the removal or addition of variables – can result in substantial changes in the results of the analysis. Third, the researcher sees the results of the analysis when assessing model fit. As a researcher modifies the model to improve fitness, she sees the effect of those modifications on the results of the study. Because the selection of the "fittest" model is largely subjective, it is easy for the researcher to justify the selection of one model over another. Regression can, thus, encourage a researcher to cherry-pick a model specification that supports a particular hypothesis or theory. Greiner explains:

> A model's fit is never perfect. At each stage of this process of exploration and assessment, the substantive result, the... answer [to the empirical question at hand], stares the analyst in the face. Only the superhuman can completely disregard the temptation to lean towards a result favorable to a chosen side, consciously or no. (Greiner 2008, p. 544)

The structure of the RCM and propensity score matching strongly discourages this form of model pruning for desired results and, if followed strictly, renders it nearly impossible. The RCM divides the development of the design and the analysis of the data into two separate phases. Since the researcher selects the design before estimating the treatment effect, the RCM precludes undesirable forms of model pruning that may bias the findings of the study. In fact, Rubin (2008) suggests deleting the outcome variable from the dataset during the design phase in order to remove any temptation to see the results before finalizing the design. Of course, it is possible to follow all the rules of the RCM and then start over again because the results are undesirable. But at the very least, the RCM and propensity score matching discourage less conscious forms of model pruning that bias the findings of studies.

Disadvantages of the Rubin Causal Model

Propensity score matching has two key practical disadvantages relative to the traditional regression approach.

First, matching subjects in the treatment and control group often results in a decrease in sample size. Subjects in one group who lack a comparable subject in the other are effectively dropped from the study in order to achieve covariate balance. This is less of a problem than it seems: keeping noncomparable subjects in a regression model for the purpose of maintaining a large sample size may not be valuable either and may bias causal estimates if the model is not properly specified.

Second, propensity score matching increases the difficulty of examining the causal effect of quantitative variables that are not binary. Regression is flexible with regard to the structure of all independent variables: they can be nominal, ordinal, or quantitative. Propensity score matching is equally flexible for balancing pretreatment covariates, but not for estimating treatment variables. Because propensity score matching typically separates the sample into two or more discrete treatment and control groups, it has traditionally not been used to measure the effect of a quantitative treatment variable. The same limitation exists in randomized experiments, where the subjects must be grouped into discrete treatment and control groups. However, more recent work by Hirano and Imbens (2004) has extended propensity score matching to continuous treatment variables, thus providing some of the benefits of the generalized linear model in regression.

What Problems Remain Unsolved by the Rubin Causal Model and Propensity Score Matching?

There are several methodological problems associated with all observational research, including propensity score matching.

First, propensity score matching does not solve the problem of omitted variable bias. Excluding a covariate that affects the outcome variable may cause unobserved bias in the estimated treatment effect. Propensity score matching, however, offers a transparent design sensitivity analysis, a tool for assessing the plausibility that unobserved bias has caused the differences observed in the treatment and control groups.

Second, like regression, propensity score matching assumes the stable unit treatment value assumption (SUTVA). SUTVA asserts that the assignment of subjects to the treatment or control group does not affect the outcomes of other subjects (Cox 1958). This is often referred to as "interference" in the experimental literature (Rosenbaum 2010; Sampson 2010). Unfortunately, the problem of interference is difficult to avoid in criminological research – whether experimental or observational – because it is likely that treatment assignment influences social interactions between persons in the study.

Third, like linear regression, propensity score matching does not automatically account for two- or three-way interactions that affect selection into the treatment and control groups. Imagine that using drugs and dropping out of high school independently increase offending. In propensity score matching, one would, of course, need to balance the treatment and control group on these two variables before estimating the effect of gang membership on offending. Imagine further that using drugs and dropping out collectively exert a larger effect on offending than their respective independent effects. Under these circumstances, balancing the treatment and control group on drug use and family structure separately may not balance their co-distribution: it is possible for the treatment or control group to have a substantially larger number of subjects who are both drug users and dropouts. To ensure balance on the co-distribution of the two variables, an interaction variable in the logistic regression model is needed to properly estimate propensity scores for subjects in the sample. Still, one might argue that in some cases, matching better accounts for omitted two- and three-way interactions than linear regression: unless the correlation between two interacting variables is different among those who receive the treatment and those who do not, balancing the independent distributions of the variables may very well balance their co-distribution as well.

Conclusion

The RCM and propensity score matching provide a number of methodological and practical advantages over the more traditional regression-based methods in the criminological literature. Those benefits include a more reasonable strategy for controlling for confounding variables, a better method for accounting for nonlinear relationships, and the straightforward application of nonparametric estimators and statistical tests. Propensity score matching also relaxes the assumptions of linear regression and uses a two-part research design that discourages the practice of model pruning, which may bias the findings of observational research. Criminologists would do well to adopt the propensity score matching in many theoretical and applied research studies.

Before concluding, one note is in order. This entry has suggested that, if criminologists are to choose between regression and propensity score matching, they should often choose the latter. But regression and propensity score matching are not fully at odds with each other. Criminologists can combine both. One option is to match and then regress. In fact, Rubin has argued this two-step strategy will sometimes decrease bias in the estimate of the treatment effect (Cochran and Rubin 1973). Another option is to include the estimated propensity scores as a predictor variable in a regression model (e.g., Berk and Newton 1985). Unlike in propensity score matching, however, this strategy requires the researcher to explicitly model the functional relationship between the propensity scores and the outcome variable. One other option is to use the estimated propensity scores as weights in a regression model (Hirano and Imbens 2001). In some circumstances, weighting is an undesirable approach because it can underestimate standard errors and bias estimates of the treatment effect (Freedman and Berk 2008).

Acknowledgements The data used in this case study were originally collected as part of a grant funded by the National Institute on Drug Abuse (NIDA) (grant R01DA16722) to Andrew Morral at the RAND Corporation. The opinions expressed in this entry are those of the authors and do not represent the official positions of NIDA or the RAND Corporation.

Related Entries

- ▶ Desistance from Gangs
- ▶ Examining the Effectiveness of Correctional Interventions
- ▶ Identification Issues in Life Course Criminology

Recommended Reading and References

Berk RA (2004) Regression analysis: a constructive critique. Sage, Thousand Oaks

Berk RA, Newton PJ (1985) Does arrest really deter wife battery? An effort to replicate the findings of the Minneapolis spouse abuse experiment. Am Sociol Rev 50:253–262

Berk RA, Rauma D (1983) Capitalizing on nonrandom assignment to treatments: a regression-discontinuity evaluation of a crime-control program. J Am Stat Assoc 78:21–27

Berk RA, Li A, Hickman L (2005) Statistical difficulties in determining the role of race in capital cases: a re-analysis of data from the state of Maryland. J Quant Criminol 21:365–390

Bertsekas DP (1981) A new algorithm for the assignment problem. Math Program 21:152–171

Blalock H (1964) Causal inference in nonexperimental research. University of North Carolina Press, Chapel Hill

Campbell D, Stanley J (1963) Experimental and quasi-experimental designs for research and teaching. In: Gage NL (ed) Handbook of research on teaching. Rand McNally, Chicago

Cochran W, Cox G (1950) Experimental designs. Wiley, New York

Cochran WG, Rubin DB (1973) Controlling bias in observational studies. Sankhya Indian J Stat 35:417–446

Cochran W (1983) Planning and analysis of observational studies. Wiley, New York

Cornfield J, Haenszel W, Hammond E, Lilienfeld A, Shimkin M, Wynder E (1959) Smoking and lung cancer: recent evidence and a discussion of some questions. J Natl Cancer Inst 22:173–203

Fisher R (1935) The design of experiments. Oliver and Boyd, Edinburgh

Freedman DA, Berk RA (2008) Weighting regressions by propensity scores. Eval Rev 32:392–409

Greiner JD (2008) Causal inference in civil rights litigation. Harv Law Rev 122:533–598

Hansen BB, Klopfer SO (2006) Optimal full matching and related designs via network flows. J Comput Graphical Statist 15: 609–627

Haviland A, Nagin DS, Rosenbaum PR (2007) Combining propensity score matching and group-based trajectory analysis in an observational study. Psychol Methods 12:247–267

Hirano K, Imbens GW (2001) Estimation of causal effects using propensity score weighting: an application to data on right heart catheterization. Health Serv Outcomes Res Methodol 2:259–278

Hirano K, Imbens GW (2004) The propensity score with continuous treatments. In: Gelman A, Meng XL (eds) Applied Bayesian modeling and causal inference from incomplete-data perspectives. Wiley InterScience, West Sussex, pp 73–84

Holland P (1986) Statistics and causal inference. J Am Stat Assoc 81:945–960

Hollander M, Wolfe DA (1999) Nonparametric statistics methods, 2nd edn. Wiley, New York

Joffee MM, Rosenbaum PR (1999) Invited commentary: propensity scores. Am J Epidemiol 150:327–333

Lehmann EL (1998) Nonparametrics: statistical methods based on ranks. Prentice Hall, Saddle River

MacDonald JM, Stokes R, Ridgeway G, Riley KJ (2007) Race, neighborhood context, and perceptions of injustice by the Police in Cincinnati. Urban Studies 44:2567–2585

McCaffrey DF, Ridgeway G, Morral AR (2004) Propensity score estimation with boosted regression for evaluating causal effects in observational studies. Psychol Methods 9:403–425

Morral AR, McCaffrey DF, Ridgeway G (2004) Effectiveness of community-based treatment for substance-abusing adolescents: 12-month outcomes of youths entering phoenix academy or alternative probation dispositions. Psychol Addict Behav 18:257–268

Paternoster R, Brame R (2008) Reassessing race disparities in Maryland capital cases. Criminology 46(4): 971–1008

Rosenbaum P, Rubin D (1983) The central role of the propensity score in observational studies for causal effects. Biometrika 70:41–55

Rosenbaum P, Rubin D (1985) Constructing a control group using multivariate matched sampling incorporating the propensity core. Amer Statist 39:33–38

Rosenbaum PR (2002) Observational studies. Springer, New York

Rosenbaum P (2009) Design of observational studies. Springer, New York

Rosenbaum PR (2010) Design of observational studies. Springer, New York

Rubin DB (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol 66(5):688–701

Rubin DB (2008) For objective causal inference, design trumps analysis. Ann Appl Stat 2:808–840

Sampson R (2010) Gold standard myths: observations on the experimental turn in quantitative criminology. J Quant Crim 26:489–500

Westreich D, Lessler J, Funk MJ (2010) Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. J Clin Epidemiol 63:826–833

Zuberi T (2001) Thicker than blood: how racial statistics lie. University of Minnesota Press, Minneapolis

Proportionality

Jesper Ryberg and Thomas Søbirk Petersen
Department of Philosophy and Science Studies,
Roskilde University, Roskilde, Denmark

Overview

Proportionality denotes a relation between crime and punishment, namely, between the seriousness of a crime and the severity of the punishment. As a normative principle for the distribution of punishment – standardly referred to as the principle of proportionality – proportionality implies that the severity of a punishment should fit the seriousness of the crime that has been committed. More precisely, this idea comprises two different aspects of the relation between crime and punishment.

The first aspect of proportionality concerns the requirement that a punishment should reflect the seriousness of the crime, in the sense that its severity should comport with the severity of punishments for other crimes. For instance, if theft is a less serious crime than burglary, then *ceteris paribus* the thief should be punished more leniently than the burglar. And, if two persons are convicted of equally serious cases of criminal conduct, they should be allotted equally severe punishments. This purely comparative requirement is usually referred to as "ordinal proportionality" (von Hirsch 1993; Ryberg 2004; von Hirsch and Ashworth 2005).

The second aspect of proportionality concerns the way punishments non-relatively comport to specific crimes. For instance, a sentencing system which imposes a minor fine for a rape or several years of imprisonment for a parking offense might well be regarded as imposing grossly disproportionate punishments irrespective of what constitute the penal levels for other crimes. This idea of a non-comparative relation between crime and punishment is known as "cardinal proportionality" (von Hirsch 1993; von Hirsch and Ashworth 2005).

This entry concerns the ethical aspects of proportionality in punishment (the enactment of the proportionality principle in penal laws and in relation to court decisions in different countries is not dealt with here (see, for instance, Clarkson and Morgan 1995; Bagaric 2001)). The first section concerns the justification of proportionality. The second section deals with the ranking of crimes in gravity and punishments in severity. The final section concerns the so-called anchor problem, that is, the challenge of how scale of crimes and punishments should be linked.

Justifying Proportionality

Why should the severity of the punishment be proportionate to the seriousness of the crime? Answers to this question can be found in both of the two main rival positions that have dominated penal theoretical considerations of the justification of state punishment, namely, utilitarianism and retributivism.

An early example of a forward-directed utilitarian justification was given by C. Beccaria who held that, if crimes of unequal seriousness – such as assassination, poaching, and forging – were punished equally harshly, this would undermine people's ability to distinguish between the seriousness of these crimes (Beccaria 1964). A corresponding argument was presented by J. Bentham who devoted a whole section of his *The Principles of Morals and Legislation* to considerations of proportionality. For instance, Bentham argued that, in the absence of proportionality, potential criminals would not be deterred from committing serious crimes any more than minor crimes and, therefore, would just as readily commit them. As an illustration of this consideration, he held that "If then, for giving you ten blows, he is punished no more than for giving you five, the giving you five of these ten blows is an offence for which there is no punishment at all: which being understood, as often as a man gives you five blows, he will be sure to give you five more, since he may have the pleasure of these five for nothing"

(Bentham 1988, pp. 181–2, note 3). In the modern utilitarian literature, the main focus has been on the undesirable consequences – e.g., in terms of antipathy or distrust toward legal institutions and practices – that might follow if the legal system is perceived as not observing the conception of justice that seems to lie inherently in the idea that a more serious crime should be punished more harshly than a less serious crime (Bagaric 2001, pp. 185–86). However, whether a utilitarian punishment system should distribute punishment in a proportionate manner depends on whether this sort of penal distribution scheme is in fact the one that maximizes well-being, which is a purely empirical question and a question that has been the object of criticism (von Hirsch 1985).

The main modern discussion of the justification of proportionality has taken place within the framework of retributivist penal theory. The renaissance of Kantian and Hegelian thoughts on punishment – and the corresponding decline of the utilitarian penal theory – that took place in the 1970s not only marked a turning point in penal theoretical thinking but also, as emphasized by A. von Hirsch, led to a change of focus from the question "Why punish?" to "How much?," thereby providing an opening for new and more elaborate thoughts on proportionality (von Hirsch 1991). Modern retributivism – a deontological position according to which punishment is justified in terms of desert – has been presented in many different versions, leading to correspondingly different approaches to the justification of a proportionate penal distribution. The overall idea of the retributivist justification is that a criminal deserves a more severe punishment, the more serious the crime committed is. The differences between the various expositions of this idea lie in various views on why a criminal deserves a punitive response for having committed a crime.

A first and simple answer is to hold that what a criminal deserves is to *suffer* and, hence, that he or she deserves to suffer more, the greater the seriousness of the crime. Since a punishment is more severe the more suffering it inflicts, it follows that a criminal should be punished more severely, the more serious his or her crimes. However, as has been pointed out, this way of justifying proportionality in punishing is vulnerable to the objection that suffering may befall a person in many other ways than by being punished (Ezorsky 1972; Ryberg 2004). For instance, a perpetrator may be racked by guilt, feel anxious about the prospect of imprisonment, or he or she may simply suffer from some sort of disease. However, if there are several ways in which a criminal may suffer after his crime is committed, then it seems that, in comparison with other criminals, who have committed equally serious crimes but who have not undergone some sort of nonpunitive suffering, the criminal has already paid part (or all) of his or her desert debt. But this implies that in this case, equally serious crimes should not be equally severely punished after all. Thus, it becomes less obvious that a proportionate punishment scheme can be justified on the grounds of the view that criminals deserve to suffer.

Another and theoretically more elaborate attempt to justify proportionality within a retributivist framework has been presented by what has become known as the "fairness theory" (Davis 1992). This theory is often presented on the ground of the broader view that, in a cooperative venture, involving costs and benefits of all parties, there should be an equitable distribution of those costs and benefits. Applied to the question of law and punishment, the view is that a criminal, on the one hand, gains a benefit from other people's obedience to the law but, on the other, gains an extra benefit (an unfair advantage) by not restricting his or her action as does the law-abiding person. On this view, the criminal deserves a punishment as a way of outweighing the extra benefit which he or she has unfairly gained. Following this outlook, proportionality has been justified on the ground of the claim that a more serious crime provides the criminal with a greater unfair benefit and that a more severe punishment therefore is required in order to restore the equilibrium of benefits and burdens. As part of the task of justifying proportionality, two of the main challenges facing the fairness theory have been to clarify whether it is always the case that a greater unfair benefit is gained from having committed a more serious crime and, also, whether an appropriate burden can befall a criminal in nonpunitive ways or only be properly imposed as part of a punishment (Burgh 1982; Sher 1987; Davis 1992; Ryberg 2004).

A third retributive approach to the justification of proportionality has been provided by the expressionist theory of punishment. According to this theory, a punishment should not be seen merely as the infliction of suffering. Rather a punishment is perceived as a way of expressing or communicating an appropriate condemnatory message to the criminal (von Hirsch 1993; Duff 1986, 2001). What the criminal deserves is to be appropriately censured or blamed and the punishment is the means to this goal. The step from the expressionist penal theory to a proportionality in punishing has been presented by A. von Hirsch in the following way: “1) The state’s sanctions against proscribed conduct should take a punitive form; that is, visit deprivations in a manner that expresses censure or blame. 2) The severity of the sanction expresses the stringency of the blame. 3) Hence, punitive sanctions should be arrayed according to the degree of blameworthiness (i.e. seriousness) of the conduct (von Hirsch 1993, p. 15).” One of the questions which the expressionist view and its justification of proportionality has given rise to is why, if it is a fact that censure and blame can be expressed or conveyed to someone in many different ways, the state’s sanction should take the form of a punishment. This “why hard treatment” question has been answered in different ways by modern adherents of expressionist approaches to punishment (Matravers 1999).

Ranking Crimes and Punishments

If the severity of punishment should be determined by reference to the seriousness of the crime committed, then the tasks of clarifying what makes one crime more serious than another and what makes one punishment more severe than another become crucial. For instance, unless it is possible to tell whether a rape is more serious than a burglary or whether theft is more serious than reckless driving, the idea of proportionality in punishing becomes vacuous and unable to provide any practical guidance on penal distribution. Similarly, proportionality presupposes that it is possible to determine whether one punishment, say a prison term of a certain length, is more or less severe than another punishment, say, a fine of a certain amount.

Though there does not exist theoretical agreement on how crimes should be ranked in terms of seriousness, a standard view in the modern retributivist tradition is that crime gravity should be determined on the ground of *harm* and *culpability* (von Hirsch 1993; von Hirsch and Ashworth 2005). With regard to the harm dimension, the view is that the relative gravity of a crime increases with the degree of harmfulness. Assault is a more serious crime than theft because it causes much more harm to the victim. Though this seems intuitively appealing, there may be other cases in which it is less clear how to compare the relative harm of different crimes. Thus, one of the theoretical tasks has been to specify what should be understood by criminal harm and to provide some sort of guidelines for the computation of harm (see von Hirsch and Jareborg 1991).

One of the things that complicates estimates on the harm of crimes is that there are several crimes which do not, at least not in any straightforward way, involve harmful conduct. A standard example is conduct which only risks or attempts harm. For instance, even if someone drives hazardously in a crowded street, there may be no one who is actually harmed. The same is the case with the inchoate crime of attempt. For instance, a planned crime may not succeed for the simple reason that the person did not commit all the acts necessary to bring it about. Or the person did all the intended acts but nevertheless did not succeed in bringing about a desired result. One way of dealing with such cases has been to suggest that what matters is not the actual outcome (harm) of a criminal act but rather that the criminal ought to be held liable for what he or she intended, no matter whether this was realized by the crime (Ashworth 1984). Another way of accounting for non-realized harms has been to draw on judgments of risk. On this view, what matters in the determination of crime gravity is the risk-adjusted harm, that is, the product of the harm intended and the risk that it would be realized by the crime that was committed (von Hirsch and Jareborg 1991).

The other dimension of crime gravity is the culpability of the criminal. The first aspect of culpability is the mental state or attitude a person holds when a crime is performed. The law uses the term "mens rea" (the guilty mind) to denote these mental elements. The traditional *mens rea* distinctions are between: intention, knowledge, recklessness, and negligence (and often in the last case no criminal culpability at all). The idea is that different kinds of *mens rea* imply different degrees of culpability. For instance, it is generally agreed upon that intentional harm *ceteris paribus* implies a higher degree of culpability than recklessly caused harm. As H. L. A. Hart illustratively remarks, it seems worse to break someone's Ming china intentionally than to knock it over while walzing wildly round the room not thinking of what might get knocked over (Hart 1968).

The other aspect of culpability concerns what is usually referred to as "excuses" which may take several forms. For instance, a person may be excused if a harmful effect is caused involuntarily, under constraints from defects of knowledge or defects of will, or if a person in different ways lacks sufficient capacity to make judgments (e.g., due to intoxication, infancy, or insanity). A standard view on excuses is that they instantiate an underlying view on responsibility and that different degrees of responsibility imply different degrees of culpability. However, what constitutes the most plausible theory of criminal responsibility is a matter of ongoing dispute.

The fact that computation of the seriousness of a crime is a poly-dimensional enterprise obviously complicates such judgments. One way of illustrating how harm and culpability affect the assessment of crime gravity is to perceive the gravity of a particular crime as the product C times H where H is the harm done or risked, while C is the culpability of the criminal, indicated by the numerical values from zero to one (see Nozick 1981). This formula implies that culpability is a matter of degree and that when there is no culpability – i.e., when $C = 0$ – the gravity of the crime is zero, which implies that the defendant does not deserve a punishment. However, in real life, numerical judgments are not an option. And the question as to how one should rank two crimes in gravity, when the first crime is more harmful than the second, while the second is performed with a higher degree of culpability, remains a theoretically open question (see Ryberg 2004).

The ranking of punishments in severity is generally perceived as a theoretically less challenging task. For instance, it seems reasonable to hold that the severity of a prison term varies with its duration and that severity of fines varies with the amount of money. However, motivated, for instance, by the contention that a punishment system which offers only a few punishment options will often punish criminals either too severely or too leniently relative to the crime committed, there has been a growing interest in intermediate sanctions – including, e.g., home detention, community service, electronic monitoring, etc. – thereby prompting a discussion of how such punishments should be compared in terms of seriousness. The obvious answer is that sanction gravity should be determined on the ground of the amount of suffering, deprivation, or inconvenience which a punishment inflicts on the punished. One of the things that complicates this suggestion is that punishments which are the same in name – say, 1 year in prison – may affect two persons very differently in terms of suffering or deprivation. This fact has led to discussion of whether proportionality implies that two persons, who have committed the same crime, should be punished with objectively different punishments, if there are differences in two persons' vulnerability to a certain type of punishment, or whether an objectivist account of punishment severity – which does not include considerations of differences in the ways a punishment is felt or experienced – is the more plausible (von Hirsch 1993; Kolber 2009; Ryberg 2004, 2010).

Specifying Proportionality

How far can the principle of proportionality guide us when it comes to the distribution of punishment? The idea of proportionality in punishing – as we have just seen – presupposes that it is possible to rank different crimes in seriousness and to rank punishments in severity. Thus, another way of posing the question is to ask how should the scale of punishment and the scale of crime, once they have been constructed, be linked? This question, which has been considered almost exclusively within a retributivist framework, opens up a number of different answers.

One possible answer might be to hold that the scales should be linked by letting the punishment directly mirror the character of the crime that has been committed. This view – referred to as *lex talionis* – is often expressed in the classical biblical formulation "an eye for an eye, a tooth for a tooth." Though this principle was, in the historical context, understood not as prescribing exact retribution but rather as a way of restraining disproportionate cruel retaliation – "several eyes for an eye" – it is obvious that a literalistic interpretation of "like for like" would constitute a simple device for meting out appropriate punishments for different crimes. However, in penal theory, this interpretation has had very few adherents. One of the classical retributivists who came closest to defending this view is I. Kant who insists that a principle of retaliation "can assign both the quality and quantity of a just penalty" (Kant 1996). However, even he recommends a number of punishments, which violates the definiteness of *lex talionis*. One of the reasons for this is probably that *lex talionis* faces the simple practical challenge that there are a large number of cases in which it is impossible to reverse a crime against the criminal himself/herself. For instance, it is hard to see what punishment one should inflict on a blackmailer, a forger, a dope peddler, a multiple murderer, a smuggler, or "a toothless fiend who has knocked somebody else's tooth out" (Kleinig 1973, p. 120).

A possible reaction to this problem might be to interpret *lex talionis* somewhat more loosely by holding that what matters in penal distribution is not that materially the same should be done to the criminal as was done to the victim, but rather that it is the harm that a crime caused to the victim that should be reversed against the criminal himself/herself. This view was defended by G. W. F. Hegel who held that it is only in respect of the "form that there is a plain inequality between theft and robbery on the on hand, and fines, imprisonment, &c., on the other. In respect of their 'value', however, i.e. in respect of their universal property of being injuries, they are comparable" (Hegel 1967, p. 72). By interpreting proportionality as equivalence of harm, the problem as to how one should punitively respond to blackmail, forgery, and several other nonreversible crimes is resolved. However, this is not tantamount to the fact that there are no problems of application. One of the reasons why most modern retributivists do not subscribe to *lex talionis* – in either of the two interpretations – is that the principle seems to ignore the complexity of judgments of what determines the seriousness of a crime. As indicated above, few, if any, modern retributivists would hold that all that matters in the comparison of crime seriousness is the harm that is caused. The culpability of the criminal is usually considered a seriousness-determining factor as well. However, if this is so, then it is no longer clear what "like for like" means. For instance, if person A and person B have each committed a crime that caused the same degree of harm to their respective victims, but it turns out that while A was fully culpable for his/her crime while B was not fully culpable (e.g., because he/she was only partially responsible), then the crime committed by A would be more serious than the crime committed by B. However, this is not captured by an interpretation of proportionality which requires that a crime is either literally reversed against the criminal or reversed in the form of harm-for-harm equivalence. Conversely, if one insists that proportionality should be interpreted as a like for like principle, then it becomes obscure what constitutes the equivalent punishment of a crime for which the perpetrator was only partially culpable.

In fact, this point is what some of the very early opponents of retributivism had in mind in their criticism of the idea of proportionate punishment. For instance, more than a half a century ago, W. G. Maclagan argued that retributivism is impossible in application because it presupposes a notion of equivalence which itself is impossible because "the two things that are to be measured against each other are in their very nature incommensurable" (Maclagan 1939, p. 290). The same objection was presented by H. Rashdall and E. Ferri more than a century ago. However, in the modern penal philosophical discussion, this criticism has had less resonance, for the simple reason that several retributivists themselves reject the idea of an equivalence between crime and punishment. But if proportionality is not interpreted as a "like for like" principle, then how should one determine what constitutes the proportionate punishment for a certain crime? Several attempts at answering this question have been presented in the modern retributively oriented era of penal theory.

One proposal, which has been defended by J. Kleinig and adopted by several other penal theorists, is to link the scales of crimes and punishments by the use of "anchor points" (Kleinig 1973). The proposal is based on the assumption that the two scales are not simply ordinal scales. That is, crimes are not merely ranked in terms of being more or less serious than other crimes. Rather the idea is that differences in the relative seriousness of different crimes should be reflected in the way crimes are ranked. To take a simple example, if out of three crimes, ranked in ascending order of seriousness – say, theft of 50€, embezzlement of 100€, and rape – the second is only slightly more serious than the first while the third is considerably more serious than the second, then these relative differences in gravity between the crimes should be reflected in the way the three crimes are being punished. In more technical terms, there should be an interval matching of the scale of crimes and the scale of punishments. On the ground of this assumption, the proposal is that the poles of the respective scales should serve as anchor points. The idea is presented by D. E. Scheid in the following way: "The main strategy . . . is to consider a scale that includes all crimes and a scale that includes the full range of acceptable punishments, and then anchor the scale to each other at two points. We anchor the most severe punishment to the most serious crime (say, life imprisonment for murder) and the least severe punishment for the least serious crime (say, a fine for petty larceny), with the other crimes falling in between" (Scheid 1997, p. 494).

In the penal theoretical discussion, this way of giving content to the idea of proportionate punishing has given rise to several objections (Ryberg 2004). For instance, it has been held that it is far from clear how one should identify the upper anchor points. First, it is not obvious what should be regarded as the most serious crime. Should this be murder? Why not the murder of ten persons or perhaps of 50? Or what about genocide? The challenge is that for any crime, it always seems possible to imagine another one (in fact sometimes reality lives up to imagination) that is even more serious. Second, the idea of determining the most severe of punishments has also been questioned. There is no straightforward way of making sense of the idea of the most severe of possible punishments. One proposal has been that the upper limit of punishment should be determined at the point at which a punishment would violate human dignity (Kleinig 1973). Critics, for their part, have been skeptical as to whether it is possible to develop a plausible account of dignity on the ground of which it will be possible to determine that, say, the length of a prison term becomes morally unacceptable because it violates human dignity (Ryberg 2004). Moreover, as a further critical comment, it has been held that the outlined method for the anchoring of the two scales is in need of an argument as to why the scales should be pegged at the end points of the scales. Why is it that the most serious crime should be pegged to the most severe punishment? Why not peg it to the punishment just below the most severe on the scale? In the absence of an argument to this effect – the criticism goes – the suggested anchor theory ends up as morally arbitrary (see von Hirsch 1993, p. 39; Ryberg 2004, p. 142).

Another approach to the question as to how the scale of crimes and the scale of punishments should be linked has been developed by A. von Hirsch. This theory combines considerations of desert, prevention, and parsimony in punishing (von Hirsch 1993). The main idea is that a punishment system should start by ordering the crime scale, from the least to the most serious, and by indicating a punishment level which corresponds to the current level in the society. However, von Hirsch further suggests that one should follow a "principle of parsimony," according to which state-inflicted suffering should be kept to the minimum necessary to achieve that purpose of the intervention. On the ground of this principle, one should follow a "decremental strategy" by a *pro rata* reduction of the penalties. Von Hirsch, finally, suggests that this progressive diminution of the punishment levels should stop at a point determined on the ground of considerations of crime prevention. This proposal has led to discussions of the role which crime prevention plays within the overall retributively minded penal theory and of whether this way of anchoring the scale succeeds in providing a sufficient guideline for how crimes of different gravity should be punished (Ryberg 2004). The discussion of this as well as of the former proposal illustrates that, even among penal theorists who subscribe to the principle of proportionality, there exists no general current consensus on how the proportionate punishment for different crimes should be determined.

Related Entries

- ▶ Punishment
- ▶ Traditional Retributivism

Recommended Reading and References

Ashworth A (1984) Sharpening the subjective element in criminal liability. In: Duff RA, Simmonds N (eds) Philosophy and the criminal law. Steiner, Wiesbaden

Bagaric M (2001) Punishment and sentencing. Gavendish, Great Britain

Beccaria C (1964) On crimes and punishment. Oxford University Press, Oxford

Bentham J (1988) The principles of morals and legislation. Prometheus Books, USA

Burgh RW (1982) Do the guilty deserve punishment? J Philos 79:193–210

Clarkson C, Morgan R (1995) The politics of sentencing reform. Clarendon, Oxford

Davis M (1992) To make punishment fit the crime. West View Press, Boulder

Duff RA (1986) Trials and punishment. Cambridge University Press, Cambridge

Duff RA (2001) Punishment, communication, and community. Oxford University Press, New York

Ezorsky G (1972) Philosophical perspectives on punishment. State University of New York Press, Albany

Hart HLA (1968) Punishment and responsibility. Oxford University Press, New York

Hegel GWF (1967) Philosophy of Right (trans: Knox TM). Oxford University Press, New York

Hirsch v (1985) Past and future crimes. Rutgers University Press, New Brunswick

Kant I (1996) The Metaphysics of Morals (trans: Gregor M). Cambridge University Press, Cambridge

Kleinig J (1973) Punishment and desert. Martinus Nijhoff, The Hague

Kolber A (2009) The subjective experience of punishment. Columbia Law Rev 109:182–236

Maclagan WG (1939) Punishment and retribution. Philosophy 14:281–298

Matravers M (1999) Punishment and political theory. Hart, Oxford

Nozick R (1981) Philosophical explanations. Harvard University Press, Cambridge

Ryberg J (2004) Proportionate punishment: a critical investigation. Kluwer, Dordrecht/New York

Ryberg J (2010) Punishment and the measurement of severity. In: Ryberg J, Corlett JA (eds) Punishment and ethics. Palgrave Macmillan, Basingstoke

Scheid DE (1997) Constructing a theory of punishment, desert, and the distribution of punishments. The Can J Law Jurisprud 10:441–506

Sher G (1987) Desert. Princeton University Press, Princeton

Tadros V (2011) The ends of harm. The moral foundations of criminal law. Oxford University Press, Oxford

Tonry M (1995) Proportionality, parsimony, and interchangeability of punishments. In: Duff RA (ed) Penal theory and penal practice. Manchester University Press, Manchester

Tonry M (2011) Retributivism has a past. Has it a future? Oxford University Press, New York

von Hirsch A (1991) Proportionality in the philosophy of punishment: from "Why Punish?" to "How Much?". Isr Law Rev 25:549–579

von Hirsch A (1993) Censure and sanctions. Oxford University Press, Oxford

von Hirsch A, Ashworth A (2005) Proportionate sentencing. Oxford University Press, Oxford

von Hirsch A, Jareborg N (1991) Gauging criminal harm: a living standard analysis. Oxf J Leg Stud 11:1–38
Walker N (1991) Why punish? Oxford University Press, Oxford
White MD (2011) Retributivism. Oxford University Press, New York

Prosecution and Wrongful Convictions

Rashaan A. DeShay and John L. Worrall
School of Economic, Political, and Policy Sciences, The University of Texas at Dallas, Richardson, TX, USA

Synonyms

Exonerations; Miscarriages of justice

Overview

This entry explores the relationship between prosecution and wrongful convictions. It begins with an overview of the problem, and then it introduces the means by which wrongful convictions occur. Prosecutorial causes for wrongful convictions are then reviewed, as are pertinent court cases in the area. This entry concludes with policy recommendations aimed at minimizing the role of prosecution in wrongful convictions.

Introduction

In June of 2007, Durham County District Attorney Michael Nifong was disbarred and stripped by the North Carolina State Bar Disciplinary Committee of his license to practice law (*North Carolina State Bar v. Michael B. Nifong*, 06 DHC 35 (2007)). Nifong's disbarment was based on his actions during the infamous and highly publicized Duke University lacrosse scandal, during which he committed a number of professional missteps, including going public with false accusations and refusing to hear exculpatory evidence against the players identified as suspects in the case. According to the disciplinary committee, no other action short of disbarment would have been appropriate under the circumstances. Although the case was not typical, it served as a shining example of what can happen when prosecutors overstep their authority.

Prosecutors represent the government in criminal proceedings. According to the American Bar Association's Code of Professional Responsibility, "The responsibility of a public prosecutor differs from that of the usual advocate; his duty is to seek justice, not merely to convict." Prosecutors also enjoy a significant amount of power in handling their duties, especially considering the fact that they choose which cases to pursue and whether to take cases to trial or to settle matters through plea agreements (Worall 2008). The control they have even plays a role in how cases are handled at each step of the process. Finally, research has shown that prosecutors are often instrumental in case outcomes, including as to whether a wrongful conviction takes place (Huff et al. 1986; Scheck et al. 2000; Schehr and Sears 2005).

Despite the image of prosecutors as hardworking professionals who represent the government and play a major role in the criminal process (Worall 2008), wrongful convictions owing to their actions place a black eye on the justice system. Such injustices undermine people's faith in the ability of the justice system to identify and hold accountable people who break the law (Huff et al. 1986; Ricciardelli et al. 2009). For instance, knowing that wrongful convictions occur may result in jurors and other individuals calling into question the guilt of those being accused of crimes (Huff et al. 1986). Further, they may doubt that convictions secured by prosecutors are valid (Huff et al. 1986). In light of this concern, this entry begins with an overview of the problem of wrongful convictions, and then it introduces the causes and contributing factors of wrongful convictions. Prosecutorial causes for wrongful convictions are also reviewed, as are pertinent (primarily US Supreme Court) cases in the area. This entry concludes with policy recommendations aimed at minimizing the role of prosecution in wrongful convictions.

The Problem of Wrongful Convictions

Wrongful convictions have been defined as "those cases in which a person was convicted. . ., but later is found to be innocent beyond a reasonable doubt, generally due to a confession by the actual offender, evidence that had been available but was not sufficiently used at the time of conviction, new evidence that was not previously available, [or] other factors," (Huff et al. 1986, p. 519). This definition highlights a number of means by which a person can be convicted for a crime he or she did not commit. It also highlights particular features of a case that a wrongfully convicted individual may seize upon in order to challenge the trial verdict.

Conservative estimates suggest that wrongful convictions occur in about one-half of one percent of all felony convictions (Huff et al. 1986). Other estimates suggest that the rate of wrongful convictions reflect the particular type of crime at issue. For example, Gross (2008) estimated that the rate for capital murder, rape, and robbery could be from 2 % to 5 %. Focusing on capital rape murder, Risinger (2007) estimated that wrongful convictions occur in between 3 % and 5 % of such cases.

It is not clear which miscarriage of justice does the most harm. Is it worse to wrongfully incarcerate an individual for several years than tarnish a prosecutor's reputation because he/she wrongfully convicted someone with no malicious intent whatsoever? In truth, each and every instance of wrongful conviction is an injustice, making it difficult to rank them from least to worst based on harm. That being said, most would agree that wrongfully executing an individual is the ultimate miscarriage of justice.

A defendant who is wrongfully convicted may eventually be exonerated. An exoneration is "an official act declaring a defendant not guilty of a crime for which he or she had previously been convicted" (Gross et al. 2005, p. 524). Identifying the total number of exonerations in this country is difficult, but evidence suggests that approximately 600–700 exonerations have occurred over the past 35 years (Gross 2008). Of the exonerations examined by Gross and his colleagues (2005), 96° % were for either murder or rape (or sexual assault).

The reasons for which the wrongfully convicted are exonerated also vary. For example, as of September 24, 2011, there were 273 individuals exonerated through the use of deoxyribonucleic acid (DNA) evidence in the United States, 17 of whom were sitting on death row prior to being exonerated (Innocence Project n.d.).

According to Gross (2008), it should come as no surprise that wrongful convictions occur. No system is perfect. Indeed, one can argue that trials are not supposed to be perfect, only fair. Even so, some people have a difficult time believing the criminal justice system can make such egregious mistakes. Were it not for the press attention high-profile exonerations receive, the problem might go relatively unnoticed. The sheer incidence of exonerations has increased from "an average of 12 a year from 1989 to 1994, to an average of 42 a year since 2000" (Gross et al. 2005, p. 527), making the problem of wrongful convictions all the more apparent.

The Death Penalty and Wrongful Convictions

Wrongful convictions in death penalty cases are especially worrisome and arguably the most serious type of mistake the justice system can make. For example, former Illinois Governor George Ryan placed a moratorium on the death penalty in 2000 because at one point, 13 individuals on the state's death row had been exonerated (Zalman 2006). Since that time, seven more death row inmates have been exonerated, bringing the grand total to 20 (Center on Wrongful Convictions n.d.). These 20 exonerations amount to a rate of more than 6° %, which is the highest exoneration rate among states with the death penalty (Center on Wrongful Convictions n.d.). Partly in response to this problem, Illinois Governor Pat Quinn signed a bill into law during 2011 that abolished the death penalty in Illinois (Schwartz and Fitzsimmons 2011).

Unfortunately, researchers have been able to identify cases in which individuals were executed in spite of compelling claims of innocence (Bedau and Radelet 1987). Once a person is executed, there is obviously no going back to correct mistakes; the only way to correct the mistake would be to issue a posthumous exoneration, which would do nothing for the wrongfully convicted person who has been executed.

Notwithstanding the prospect of executing an innocent individual, there is also great concern about forcing the wrong person to sit on death row for an extended period of time. Death row exonerees have pointed to the "fear, anger, loneliness, and feelings of inhumanity that come with facing execution for crimes they did not commit" (Westervelt and Cook 2010, p. 266). Exonerees also speak about the fear that comes along with watching a fellow death row inmate be removed from his cell to face execution, the issues that come along with having one's execution date set, and the anxiety and stress that come from coming within hours of actually being executed (Westervelt and Cook 2010).

The Innocence Movement

Barry Scheck and Peter Neufeld helped draw attention to the issue of wrongful convictions by launching the Innocence Project in 1992. Housed in the Benjamin N. Cardozo School of Law at Yeshiva University in New York, the Innocence Project focuses on helping exonerate the wrongfully convicted, primarily through the use of DNA testing (Innocence Project n.d.). The Innocence Project is able to pursue its mission with the help of the funding it receives from individuals, foundations, its annual benefit dinner, corporations, and the Cardozo School of Law (Innocence Project n.d.).

The Innocence Project is not alone. Other such projects across the country conduct similar work, investigating cases and fighting for individuals they have reason to believe were wrongfully convicted.

The Innocence Movement, as it has been called (Hughes 2011; Zalman 2006), has taken on a significant amount of power with the creation of the Innocence Network. The network is an "affiliation of organizations dedicated to providing pro bono legal and investigative services to individuals seeking to prove innocence of crimes for which they have been convicted and working to redress the causes of wrongful convictions" (Innocence Network n.d.). Each year, the Innocence Network holds an annual conference that brings together member organizations and provides training and information for attorneys working on wrongful conviction cases. In addition, these conferences welcome exonerees from all over the country and invite them to participate, usually through the provision of special activities in which exonerees are able to take part.

Causes of Wrongful Convictions

Several causes of wrongful convictions have been identified. Some are incidental to prosecutorial misconduct. Others are not. Often, what contributes to the problem of wrongful convictions is a series of actions involving the prosecutor and other individuals inside and outside the criminal justice system.

Eyewitness misidentification has been noted as the leading cause of wrongful convictions (Innocence Project n.d.). According to the Innocence Project, eyewitness misidentification was a factor in 75° % of the convictions overturned through DNA evidence (Innocent Project n.d.). Although not necessarily identified as a cause, race has been identified as a factor involved in wrongful convictions, as well; plenty of evidence suggests a defendant's race is associated with whether the individual will be convicted (Bedau and Radelet 1987; Gross et al. 2005).

Judicial errors have also proven responsible for wrongful convictions. They include instances in which judges "allow incompetent defense attorneys, overzealous police and prosecutors, and questionable forensic evidence to permeate the courtroom" (Ramsey and Frank 2007, pp. 446–447). This could include judges improperly allowing prosecutors to make improper and inflammatory statements during the trial.

So-called junk science has been linked with wrongful convictions, as well. This refers to scientific evidence that is either faulty or analyzed improperly and is introduced into evidence. It also refers to forensic testing that is "inaccurately conveyed in trial testimony" (Innocence Project n.d.).

Another cause of wrongful convictions is "snitch testimony," or statements made by a witness who has an incentive to lie. "Snitches" are usually either jailhouse informants who receive some sort of leniency in exchange for their testimony or killers who are trying to divert suspicion from themselves (Warden 2004). According to Warden (2004), snitch testimony is the leading cause of wrongful convictions in capital cases.

Yet another cause is confirmatory bias, which is "the tendency to emphasize and believe experiences which support one's views and to ignore or discredit those which do not" (Mahoney 1977, p. 161). This can be especially troublesome when prosecutors, those working in law enforcement, or even jurors allow confirmatory bias to guide their decision-making.

Still other causes of wrongful convictions include juror bias, juror neglect, community pressure for conviction, errors made by medical examiners and forensic science experts, false confessions, the mental capacity of the accused, and inadequate defense representation (Huff et al. 1986; Scheck et al. 2000; Schehr and Sears 2005).

In general, a single wrongful conviction case will involve more than one of the aforementioned causes (Huff et al. 1986). An example of this can be found in the case of Charles Chatman, who, before he was exonerated in 2008, served over 26 years of a 99-year sentence for a rape he did not commit. The factors that contributed to his wrongful conviction included inadequate defense representation and eyewitness misidentification (Innocence Project n.d.).

Prosecutorial Misconduct

Prosecutorial misconduct involves "the intentional use of illegal or improper methods for attaining convictions against defendants in criminal trials" (Lucas et al. 2006, p. 97). The prevalence of prosecutorial misconduct has been documented in cases of wrongful convictions. For example, in an examination of DNA exonerations conducted by the Innocence Project, researchers found that in 63° % of cases, police and prosecutorial misconduct played a significant part in the wrongful conviction (Scheck et al. 2000).

Prosecutorial misconduct occurs by several means. One common form of misconduct has been identified as withholding exculpatory evidence from the defense (Innocence Project n.d.), in violation of *Brady v. Maryland* (373 U.S. 83 [1963]). In *Brady*, the Supreme Court held that "the suppression by the prosecution of evidence favorable to an accused upon request violates due process where the evidence is material either to guilt or to punishment, irrespective of the good faith or bad faith of the prosecution" (p. 87). The Supreme Court has since explained that *Brady* only applies to pretrial disclosure and does not extend to the post-conviction context (see *District Attorney's Office for the Third Judicial District v. Osborne*, [129S.Ct. 2308, 2319 (2009)]).

How serious is the problem of withholding exculpatory evidence? In their study of 340 wrongful capital convictions, Bedau and Radelet (1987) determined that prosecutor errors occurred in 50 of them. Of those 50, 35 involved the suppression by the prosecutor of exculpatory evidence.

Other methods of prosecutorial misconduct include deliberately mishandling, mistreating, or destroying evidence, allowing witnesses they know or should know are not truthful to testify, pressuring defense witnesses not to testify, relying on fraudulent forensic experts, and making misleading arguments that overstate the probative value of testimony (the Innocence Project n.d.). For example, consider the case of Steven Linscott, who was wrongfully convicted of murder and sentenced to 40 years in prison. During his trial, the prosecutor overstated the strength and importance of forensic evidence and also mischaracterized the state expert witness's testimony (Innocence Project n.d.; West 2010). Or consider the case of Curtis McCarty, who was wrongfully convicted and sentenced to death for a crime he did not commit (Innocence Project n.d.; West 2010). The prosecutor in his case failed to disclose evidence to the defense in a timely manner, made improper comments about facts that were not in evidence, expressed his personal opinion about the guilt of the defendant, and made comments that were found to be improper and inflammatory (West 2010). McCarty served 21 years in prison before he was released (Innocence Project n.d.).

Causes of Prosecutorial Misconduct

Identifying why prosecutorial misconduct occurs could be helpful in preventing future problems. It has been suggested that prosecutorial misconduct may result from insufficient training coupled with a focus on winning, as opposed to a focus on doing the right thing (Ridolfi and Possley 2010). In addition, it is possible that the use of a lower evidentiary standard in the early stages of a case (such as anything other than the "proof beyond a reasonable doubt" that is required at trial) can make it easier for prosecutors to convict someone who is innocent (Forst 2008).

Researchers have also called attention to the high occurrence of procedural errors that take place during the course of prosecutions of serious crimes (Bedau and Radelet 1987; Lucas et al. 2006). This may be as a result of the significant pressure that prosecutors face when dealing with difficult cases, leading them to engage in misconduct partly because they have convinced themselves that these defendants are guilty, thereby justifying their behavior (Lucas et al. 2006).

It has even been suggested that prosecutors engage in misconduct because of the heavy burdens and responsibilities placed on them and not because of any sinister intention to secure a conviction by any means necessary. For example, the expectation of securing convictions may lead prosecutors to naturally view certain evidence in an inculpatory light, thereby preventing them from being able to see the exculpatory value of that evidence. Again, it is not necessarily the case that each instance of prosecutorial misconduct can be traced to intentional acts on the part of the prosecutor.

The Courts and Wrongful Convictions

Appellate courts are especially important in the discussion of wrongful convictions and prosecution because if an innocent individual is convicted, his or her next course of action is to file an appeal. The decisions handed down by appellate courts, especially the US Supreme Court, greatly influence how cases will be decided at the appellate level.

The US Supreme Court has decided a number of cases dealing with all facets of wrongful convictions. These cases have considered a number of issues, including DNA testing, the proper claims to file when seeking certain types of relief, and whether or not it is unconstitutional to execute an innocent person. In addition to Supreme Court cases, the North Carolina State Bar held a disciplinary hearing that was extremely important in reference to the behavior of prosecutors during their quest to obtain convictions. Though not a Supreme Court case, it ranks highly among key wrongful conviction cases.

Key US Supreme Court Cases

Under the doctrine of harmless error, it is possible for an appellate court to affirm a person's conviction, even though prosecutorial misconduct or some other errors were present, so long as the reviewing court believes that the outcome of the case was not affected by errors (Ridolfi and Possley 2010). This was the rule announced in *Chapman v. California* (386 U.S. 18 [1967]), a case in which the Supreme Court decided that "there may be some constitutional errors in a conviction which, in the setting of a particular case, are so unimportant and insignificant that they may, consistent with Federal Constitution, be deemed harmless and which will not require automatic reversal of the conviction" (p. 22). The *Chapman* decision has had a significant effect on those seeking to prove their innocence on appeal. Why? If a number of errors were made at trial, someone who is wrongfully convicted may have serious trouble asserting his or her innocence if the errors made were "harmless." In fact, these errors deemed as harmless may have actually contributed to the person being wrongfully convicted (West 2010).

Another Supreme Court case that has a serious impact on those that are innocent and appealing their convictions is *Strickland v. Washington* (466 U.S. 668 [1984]). *Strickland* dealt with the issue of ineffective assistance of counsel, which, as discussed earlier, is one of the causes of wrongful convictions. In *Strickland*, the Supreme Court set forth the two-prong test for appellate courts to use in analyzing a convicted individual's claim that trial counsel was ineffective. A convicted defendant must show (1) that counsel's performance was deficient, meaning that counsel made errors so serious that it was not functioning as the "counsel" guaranteed by the Sixth Amendment and (2) the deficient performance prejudiced the defense, meaning that counsel made errors so serious as to deny the defendant a fair trial. This test can make it difficult for convicted defendants to prevail on appeal because appellate courts often defer to the judgment of the trial counsel (see, e.g., *Premo v. Moore*, 131S.Ct. 733 [2011]). Therefore, even if defense counsel is ineffective at trial, that attorney's actions may be seen as merely strategic and therefore treated with deference, meaning that a court of appeals will not likely overturn a guilty verdict.

In *Herrera v. Collins* (506 U.S. 390 [1993]), the Supreme Court considered the issue of whether, based on the Eighth Amendment's prohibition against cruel and unusual punishment and the Fourteenth Amendment's due process guarantee, it is unconstitutional to execute someone who was actually innocent of the crime for which he or she is convicted. In that case, Herrera filed a petition for a writ of habeas corpus claiming that, based on new evidence, he was innocent and that his brother, who was deceased at the time the case was heard, actually committed the murders. Although the court ruled against Herrera, holding that his claim of actual innocence did not entitle him to federal habeas relief, it did not clearly articulate whether it would be unconstitutional to execute an innocent person, thereby leaving the key issue unresolved.

In *District Attorney's Office v. Osborne* (129S.Ct. 2308 [2009]), the Supreme Court held that the defendant had no constitutional right to obtain post-conviction access to the state's evidence for DNA testing. Osborne's goal was to obtain restriction fragment length polymorphism (RFLP) DNA testing on the forensic evidence used in his case, as his attorney had failed to do so during trial. Even though Osborne had filed a request under Alaska's post-conviction relief statute, the state would not allow for the evidence to be tested, claiming that the statute did not apply to DNA testing that had been available at trial. The Supreme Court sided with the state. The practical result of the court's decision is that individuals seeking to prove their innocence in posttrial review are at the mercy of state law.

In *Skinner v. Switzer* (131S.Ct. 1289 [2011]), the petitioner was sentenced to death and was trying to obtain DNA testing of previously untested evidence by filing a civil rights action under 42 U.S.C. § 1983. The narrow issue the court decided concerned whether this avenue for relief was proper. It held that a post-conviction claim for DNA testing is properly pursued in a § 1983 claim, meaning that Skinner was permitted to pursue his claim in the federal courts. This provides convicted individuals with the opportunity to file a civil rights claim to seek post-conviction DNA testing, which at the very least allows for a remedy other than filing a writ for habeas relief. Even so, post-conviction review is not guaranteed in light of this decision.

Arizona v. Youngblood (488 U.S. 51 [1988]) offers yet another interesting look into the Supreme Court and the issue of wrongful convictions. *Youngblood* dealt with the issue of a prosecutor/police department's duty to preserve evidence. The police in that case improperly stored evidence, causing it to be degraded. This prevented them from obtaining information that could have identified the offender. Youngblood argued that the destruction of evidence, which he claimed was potentially exculpatory, violated his due process rights. The Supreme Court, however, held that the prosecution and/or police's failure to preserve potentially useful evidence did not amount to a denial of due process of law unless the defendant was able to prove bad faith on the part of the prosecution and/or police. The *Youngblood* decision was handed down in 1988, but in 2000, Youngblood's attorneys were able to persuade the police to test the evidence using a new type of DNA technology. This testing revealed that Youngblood had in fact been wrongfully convicted, which led to his eventual release from prison. In addition, the DNA testing disclosed the identity of the true assailant, which led to that individual's conviction.

The North Carolina Experience

The matter of *The North Carolina State Bar v. Michael B. Nifong* (06 DHC 35 (2007)) is also extremely important to the discussion of prosecution and wrongful convictions. The decision, which was rendered by the North Carolina State Bar, came on the heels of inappropriate actions by then-Durham County District Attorney Michael Nifong in connection with the infamous Duke lacrosse scandal, mentioned at the beginning of this entry. Although the case did not involve a wrongful conviction per se, it is relevant in the present context because, before the North Carolina State Bar become involved, it was a wrongful conviction waiting to happen.

The matter began when an alleged victim reported to the police that she had been sexually assaulted at a party attended by a number of players for the Duke lacrosse team. Early on in the investigation, Nifong decided to prosecute the case himself, and it was believed he had ulterior motives in that he was trying to take advantage of the media attention the case was bound to receive (*The North Carolina State Bar v. Michael B. Nifong*, 06 DHC 35 (2007)). Indeed, it was determined that Nifong continued to proceed with the case and secured indictments against three members of the lacrosse team, even though there was a significant amount of evidence suggesting that the defendants were innocent. Such evidence included (1) questions about the alleged victim's credibility, (2) inconsistent statements made by the alleged victim, (3) being told by law enforcement officials that there was no evidence a rape had occurred, (4) failure of the alleged victim to identify her attackers even though she had viewed two photo arrays, (5) testing from a rape kit that yielded no semen, blood, or saliva evidence in spite of the fact that the alleged victim claimed her attackers did not wear a condom, (6) the alleged victim identifying her alleged attackers on her third try under questionable circumstances, (7) and further DNA testing revealing DNA from four separate males, all of which were inconsistent with members of the lacrosse team.

In spite of all of this evidence, Nifong made repeated statements to the media, speaking about the fact that he was certain there was a rape and that it was racially motivated, talking about the harm and pain suffered by the alleged victim, and described the attack as "gang-like rape activity" (North Carolina State Bar v. Michael B. Nifong and 06 DHC 35 2007, p. 6). The disciplinary committee also found that Nifong had intentionally withheld exculpatory evidence and made false representations to the court – and to opposing counsel. As a result of his actions, the committee found that the only proper punishment was disbarment and ruled that Nifong would have to surrender his license to practice law in North Carolina. In hindsight, it appears that Nifong had every intention of continuing to pursue prosecution, regardless of evidence that the accused lacrosse players were innocent. He went beyond that, however, and even engaged in a campaign to withhold evidence, smear the defendants, and lie to the court. The facts of this case suggest that it had the potential to be a wrongful conviction and demonstrate what can happen when a prosecutor engages in a gross abuse of discretion.

Improving Prosecution

The US Supreme Court noted the following in its majority opinion in *Berger v. U.S.*:

> The [prosecutor] is the representative not of an ordinary party to a controversy, but of a sovereignty whose obligation to govern impartially is as compelling as its obligation to govern at all, and whose interest, therefore, in a criminal prosecution is not that it shall win a case, but that justice shall be done. As such, he is in a peculiar and very definite sense the servant of the law, the two-fold aim of which is that guilt shall not escape or innocence suffer. He may prosecute with earnestness and vigor – indeed, he should do so. But, while he may strike hard blows, he is not at liberty to strike foul ones. It is as much his duty to refrain from improper methods calculated to produce a wrongful conviction as it is to use every legitimate means to bring about a just one. (295 U.S. 78, 88 [1935])

This language highlights the important role of the prosecutor in criminal cases. Although the prosecutor may fight to win the case, the goal is to bring about a just result. In the process, prosecutors must avoid improper conduct that is intended to bring about a wrongful conviction.

As previously discussed, prosecutors enjoy a great deal of power. This is not to say, however, that all prosecutors abuse this power or that the majority of prosecutors do not care about whether they convict an innocent person. Instead, the point is that prosecutors possess a great deal of control in how cases are handled. This points to the fact that their power and discretion can be utilized to ensure that cases are handled properly and in accordance with all ethical guidelines and legal requirements. If prosecutors do make the decision to focus their efforts on seeking justice rather than thinking only about conviction rates, this could change the culture of district attorneys' offices around the country. Such a cultural change could foster an environment in which future wrongful convictions are seen as both unacceptable and preventable. Indeed, there are two key means by which prosecutors' offices can ensure that they seek justice and avoid contributing to wrongful convictions: the establishment of conviction integrity units and by support for post-conviction DNA testing.

Conviction Integrity Units

In January of 2007, newly elected Dallas County District Attorney Craig Watkins was sworn into office. After realizing that Dallas County led the nation in wrongful convictions, he established a Conviction Integrity Unit (Dallas County District Attorney n.d.). Working with the Innocence Project of Texas, the unit now reviews and reinvestigates legitimate post-conviction claims of innocence. It also reviews and prosecutes old cases, those that are DNA and non-DNA related, when there is evidence identifying different or additional perpetrators. To date, the Conviction Integrity Unit has reviewed more than 300 cases and has helped contribute to the running total of more than 20 wrongly convicted inmates coming out of Dallas County (Dallas County District Attorney's Office n.d.).

Dallas County's Conviction Integrity Unit conducts investigations of valid claims, which may require the review of case files, interviewing witnesses, and conducting DNA testing. The county also maintains an open-file policy, which allows defense attorneys to have access to all case files, even at the post-conviction stage (Dallas County District Attorney's Office n.d.). In a number of instances, investigations have led the district attorney to even support the exoneration and release of individuals who have been wrongfully convicted (Dallas County District Attorney's Office n.d.). In addition to exonerating those who have been wrongfully convicted, the unit has also reported instances of confirming the guilt of previously prosecuted individuals.

The Dallas County experience has encouraged similar reforms around the country. For example, in March of 2010, Manhattan District Attorney Cyrus R. Vance, Jr. created a Conviction Integrity Program with the goal of addressing claims of actual innocence and preventing future wrongful convictions. In addition to reinvestigating cases it deems as having meaningful claims of innocence, the unit also utilizes senior members of the staff in order to ensure that the policies of the office are proper in terms of preventing errors, such as false confessions and eyewitness misidentifications. The program also solicits advice from criminal justice experts who serve on the unit's policy advisory panel and who offer advice on the best practices for minimizing wrongful convictions (New York County District Attorney's Office n.d.).

Post-conviction DNA Testing

Prosecutors should refrain from challenging requests that have been filed by convicted defendants seeking post-conviction DNA testing, especially when these requests are properly filed. In addition, prosecutors should support these requests in the sense that they are an opportunity to determine the truth. Although some prosecutors chose not to oppose requests early on, some exonerees report having faced resistance when they attempted to obtain DNA testing to prove their innocence (Garrett 2008).

Prosecutors have become more open to post-conviction DNA testing as a result of changes in state law (Garrett 2008). Forty-eight states currently have some sort of legislation that allows for access to post-conviction DNA testing (Innocence Project n.d.). Some of the elements of these statutes (depending on the state) involve appointing counsel to indigent defendants, setting forth the requirements to qualify for testing, providing information on which laboratory will be used for testing, and specifying who will shoulder the costs of testing (Swedlow 2002).

Despite the benefits associated with such statutes, a number of deficiencies need to be overcome. Several have been identified: (1) not allowing for DNA testing if the person seeking relief entered a guilty plea, (2) not making allowances for individuals that are no longer in custody, (3) not making allowances for individuals who want to challenge one conviction in spite of the fact that they are serving time for a second conviction, (4) not taking into consideration issues that may arise concerning the timing of when these requests should be filed, (5) not allowing successive petitions, and (6) failing to set forth what the proper procedural steps will be if the DNA testing exonerates an individual (Swedlow 2002).

Regardless of their potential limitations, DNA testing statutes play a pivotal role in the criminal process. Without such laws, individuals seeking DNA testing would be at the mercy of law enforcement (Kreimer and Rudovsky 2002). Moreover, access to DNA testing could be blocked, such as out of a desire to uphold the finality of a court's decision, because of testing cost concerns, or due to a belief that a guilty verdict is solid and should be upheld (Kreimer and Rudovsky 2002).

Other Policy Recommendations

To the extent that prosecutors' offices have a role in the incidence of wrongful convictions, they should take aggressive steps to ensure that actual perpetrators are held accountable. Of course, prosecutors are not solely to blame for wrongful convictions; additional justice system reforms are necessary. Several such reforms are readily identifiable.

First, prosecutors should ensure that evidence is being properly preserved, even after an individual is convicted. It has been noted that DNA evidence is not preserved in a large percentage of cases (Garrett 2008). A lack of DNA evidence creates a roadblock for convicted persons who wish to prove their innocence, making it more difficult or even impossible to secure an exoneration. Further, DNA testing has resulted in finding the actual perpetrator in almost 40 % of the DNA exonerations (Innocence Project n.d.). The failure to preserve evidence could prevent the police and prosecution from finding the actual offender once the innocence of the exoneree is proven.

Second, prosecutors should support police reforms that have been based on research identifying factors leading to wrongful convictions. Consider lineups. As previously mentioned, eyewitness misidentification has long been identified as a leading cause of wrongful convictions. Research has suggested that "double-blind" lineups should be utilized so that police officers cannot be suggestive in any manner during lineup administration (Gould and Leo 2010). Double-blind lineups reduce the likelihood of suggestion because neither the officer nor the eyewitness knows the identity of the suspect (Gould and Leo 2010). Prosecutors should push for the implementation of lineup reform in police departments across the country. Doing so can only serve to improve the evidence with which they get to work.

Third, encouraging prosecutors to play some sort of role in the innocence movement could be beneficial. They need to further grasp the extent of the wrongful conviction problem and gain additional awareness of the role of prosecution in past wrongful convictions. Requiring prosecutors to "participate" in the innocence movement to a certain degree could also add to its credibility. This participation could be as simple as establishing a program or policy that resembles the Conviction Integrity Unit in Dallas County. Since prosecutors will likely have access to special resources that defense attorneys may not, establishing a specialized unit within the prosecutor's office could result in not only uncovering miscarriages of justice but also confirming the guilt of convicted individuals.

Fourth, the importance of leadership in prosecutors' offices cannot be overemphasized (Scheck et al. 2000). District attorneys and assistant district attorneys in positions of leadership could set the example for other prosecutors in their offices. They could make it clear that their office respects the law and intends to adhere to ethical guideline*s* and all applicable legal requirements. This could serve as a catalyst in creating the appropriate culture in prosecutors' offices that may be lacking such qualities. At the very least, instilling in prosecutors the notion that although convictions are important justice is only served when the correct person is convicted for the crime could play a major role in improving some of the aforementioned problems associated with prosecution and wrongful convictions.

Fifth, in an effort to address some of the concerns associated with prosecutorial misconduct, prosecutors' offices around the country should consider implementing written policies to ensure that misconduct will not be tolerated. These policies could include guidelines that set forth monitoring and disciplinary procedures that offices rely on when misconduct is suspected or occurs (Ridolfi and Possley 2010).

Sixth, prosecutors' offices and law enforcement agencies could implement written policies that dictate how exculpatory evidence will be handled. These policies should include proper training for prosecutors and law enforcement officials. In addition, they should provide guidelines regarding the disclosure of evidence to ensure that law enforcement is making prosecutors aware of exculpatory evidence and, in turn, that prosecutors hand that evidence over to the defense (Ridolfi and Possley 2010). Written policies, along with proper training, could assist in preventing or reducing the failure of prosecutors to hand over exculpatory evidence.

Seventh, the immunity protection enjoyed by prosecutors should be narrowed, allowing for lawsuits to be filed based on intentional misconduct (Scheck et al. 2000). This would allow for wrongfully convicted individuals to file civil suits against prosecutors that, for instance, let a witness testify against a defendant when the prosecutor knows he or she is lying or intentionally conceals evidence that demonstrates a defendant's guilt (Scheck et al. 2000). Permitting these lawsuits to proceed would prevent prosecutors from asserting immunity as a defense in order to get a suit dismissed. The filing of such lawsuits could serve as a deterrent to "the most outrageous practices, while not interfering with conscientious officials" (Scheck et al. 2000, p. 181).

Lastly, in the interest of justice, prosecutors should support and encourage the provision of adequate assistance to exonerees upon their release from prison. The post-incarceration needs of exonerees and the impact of wrongful convictions have been amply documented. Exonerees often face difficulty finding housing, lack sufficient resources, and are often unable to obtain reintegration services that would have been available to them had they been paroled (Scheck et al. 2000). In addition, exonerees have to deal with feelings of anger, may become anxious when out in public, may suffer from post-traumatic stress disorder, and have to deal with the fact that personal relationships were damaged or destroyed as a result of them being sent to prison (Denov and Campbell 2005). According to the Innocence Project, only 27 states currently have compensation packages that offer monetary (and possibly other types of assistance) to exonerees once they are released from prison, and some of these packages are extremely limited (Innocence Project n.d.). At a minimum, prosecutors could acknowledge that wrongful convictions occur and that exonerees are in need of assistance once they are released from prison. As such, prosecutors should support both the creation of statutes in states that have none and the improvement of existing statutes that are lacking in the services provided to exonerees.

Conclusion

Wrongful convictions have garnered considerable attention over the past few decades, and the role that prosecutors may play in them has been of increasing interest. Prosecutors enjoy considerable authority and autonomy in the criminal justice system. Unfortunately, their authority and autonomy are sometimes abused. At the same time, however, the tremendous power and discretion that prosecutors enjoy can be used to ensure that each case is handled properly and that justice is sought.

Fortunately, progress is being made. DNA testing statutes, court decisions clarifying rights and setting forth certain paths for individuals to take to seek relief, and the implementation of innovative policies like Dallas County's Conviction Integrity Unit are designed to ensure that those who are wrongfully convicted (or even think they are) have some mechanism for challenging their convictions. With continued efforts on the part of everyone involved in the process, including prosecutors, positive change will continue to take place.

Related Entries

- ▶ Causes of Wrongful Convictions
- ▶ Criminal Defense Profession
- ▶ False Confessions and Police Interrogation
- ▶ Forensic Science and Miscarriages of Justice
- ▶ Innocence Commissions
- ▶ Legal Rules, Forensic Science and Wrongful Convictions
- ▶ Measuring Wrongful Convictions
- ▶ Overturning Wrongful Convictions and Compensating Exonerees
- ▶ Scientific Evidence in Criminal Prosecutions

Recommended Reading and References

Bedau HA, Radelet ML (1987) Miscarriages of justice in potentially capital cases. Stan Law Rev 40:21–179

Center on Wrongful Convictions (n.d.) Exonerations in Illinois capital cases. http://www.law.northwestern.edu/wrongfulconvictions/exonerations/ilIndexdp.html. Retrieved 23 Oct 2011

Dallas County District Attorney's Office (n.d.) Conviction integrity unit. http://www.dallasda.com. Retrieved 13 Oct 2011

Dallas County District Attorney's Office (n.d.) Conviction integrity unit's investigation, victim's recantation clear Dale Duke of 1992 aggravated sexual assault of a child. http://dallasda.co/webdev/?p=779. Retrieved 19 Nov 2011

Dallas County District Attorney's Office (n.d.) The honorable Craig Watkins. http://www.dallasda.com/Craig_Watkins.html. Retrieved 13 Oct 2011

Denov MS, Campbell KM (2005) Criminal injustice: understanding the causes, effects, and responses to wrongful conviction in Canada. J Contemp Crim Justice 21:224–249

District Attorney's Office for the Third Judicial District et al (2009) v. Osborne, 129 S.Ct. 2308

Forst B (2008) Prosecution policy and errors of justice. In: Worrall JL, Nugent-Borakove ME (eds) The changing role of the American prosecutor. State University of New York Press, Albany, pp 51–71

Garrett BL (2008) Judging innocence. Colum Law Rev 108:55–142

Gould JB, Leo RA (2010) One hundred years later: wrongful convictions after a century of research. J Crim Law Criminol 100:825–868

Gross SR (2008) Convicting the innocent. Ann Rev Law Soc Sci 4:173–192

Gross SR, Jacoby K, Matheson DJ, Montgomery N, Patil S (2005) Exonerations in the United States 1989 through 2003. J Crim Law Criminol 95:523–560

Huff CR, Rattner A, Sagarin E, MacNamara DE (1986) Guilty until proven innocent: wrongful conviction and public policy. Crime Delinq 32:518–544

Hughes E (2011) Innocence unmodified. N C Law Rev 89:1083–1124

Innocence Network (n.d.) About the innocence network. http://www.innoncencenetwork.org/. Retrieved 10 Oct 2011

Innocence Project (n.d.) About the innocence project. http://www.innocenceproject.org/about/. Retrieved 10 Oct 2011

Kreimer SF, Rudovsky D (2002) Double helix, double blind: factual innocence and post-conviction DNA testing. Univ Pennsylvania Law Rev 151:547–617

Lucas JW, Graif C, Lovaglia MJ (2006) Misconduct in the prosecution of severe crimes: theory and experimental test. Soc Psychol Quarter 69:97–107

Mahoney MJ (1977) Publication prejudices: an experimental study of confirmatory bias in the peer review system. Cogn Therapy Res 1:161–175

North Carolina State Bar v. Michael B. Nifong, 06 DHC 35 (2007) http://www.ncbar.com/Nifong%20Final%20Order.pdf. Retrieved 17 Sept 2011

Ramsey RJ, Frank J (2007) Wrongful conviction: perceptions of criminal justice professionals regarding the frequency of wrongful conviction and the extent of system errors. Crime Delinq 53:436–470

Ricciardelli R, Bell JG, Clow KA (2009) Social attitudes toward wrongful conviction. Can J Criminol Crim Justice 51:411–432

Ridolfi K, Possley M (2010) Preventable error: a report on prosecutorial misconduct in California 1997–2009. Northern California Innocence Project, Santa Clara

Risinger DM (2007) Innocents convicted: an empirically justified factual wrongful conviction rate. J Crim Law Criminol 97:761–806

Scheck B, Neufeld P, Dwyer J (2000) Actual innocence: five days to execution, and other dispatches from the wrongly convicted. Doubleday, New York

Schehr RC, Sears J (2005) Innocence commissions: due process remedies and protection for the innocent. Crit Criminol 13:181–209

Schwartz J, Fitzsimmons EG (2011) Illinois governor signs capital punishment ban. The New York Times. http://www.nytimes.com/2011/03/10/us/10illinois.html. Retrieved 20 Nov 2011

Swedlow K (2002) Don't believe everything you read: a review of modern "post-conviction" DNA testing statutes. Calif West Law Rev 38:355–387

Warden R (2004) The snitch system: how snitch testimony sent Randy Steidl and other innocent Americans to death row. Northwestern University School of Law, Center on Wrongful Convictions. http://www.innocenceproject.org/docs/SnitchSystemBooklet.pdf. Retrieved 19 Nov 2011

West EM (2010) Court findings of prosecutorial misconduct claims in post-conviction appeals and civil suits among the first 255 DNA exoneration cases. Innocence Project, New York. http://www.innocenceproject.org/docs/Innocence_Project_Pros_Misconduct.pdf. Retrieved 13 Oct 2011

Westervelt SD, Cook KJ (2010) Framing innocents: the wrongly convicted as victims of state harm. Crime Law Soc Change 53:259–275

Worall JL (2008) Prosecution in America: a historical and comparative account. In: Worrall JL, Nugent-Borakove ME (eds) The changing role of the American prosecutor. State University of New York Press, Albany, pp 3–27

Zalman M (2006) Criminal justice system reform and wrongful conviction: a research agenda. Crim Justice Policy Rev 17:468–492

Prosecutorial Discretion

Kay Levine
Emory University School of Law, Atlanta,
GA, USA

Overview

This entry addresses the nature of prosecutorial discretion in the United States, with a particular focus on the discretionary behavior of state-level prosecutors (as opposed to federal). After describing the two most prominent forms of discretionary decision-making by prosecutors – filing and plea bargaining – the entry addresses external institutions' ability and willingness to constrain prosecutorial discretion and the role of the prosecutor's office in enabling and shaping discretionary choices by individual employees. Lastly, the entry discusses the case- and offender-level characteristics that seem to influence prosecutorial decisions, and the risk that these considerations lead to discrimination and reinforce stereotypes of crimes, criminals, and victims.

Introduction

Among nations that formally subscribe to an adversarial model of the criminal process, the prosecutor has come to be regarded as the most powerful player in the criminal court. Interest in prosecutorial power is nowhere more prevalent than in the United States, a nation that has witnessed the explosion of its prison population at the same time as its crime rates have shrunk. The responsibility of the prosecutor for mediating between crime and punishment thus has inspired countless investigations of prosecutorial decision-making, including the factors that influence individual choices about the pursuit of criminal charges and the availability of structural constraints to limit those choices in furtherance of equalized treatment of similarly situated offenders.

The Source and Nature of Prosecutorial Discretion

As members of the executive branch of government, prosecutors enforce criminal laws by initiating and pursuing criminal cases against alleged violators. This enforcement duty is not a mandate, though; in most jurisdictions for most crimes, prosecutors have discretion to decide which alleged offenders merit prosecution – and ultimately punishment – by the state. Beyond simply initiating a case against a suspect, prosecutors have the authority to decide how many charges to file for a given episode or transaction, and whether those charges should be felonies (subjecting the offender to more than 1 year incarceration), misdemeanors (subjecting the offender to no more than 1 year incarceration), or merely infractions (subjecting the offender to monetary penalties but no incarceration). For those suspects deemed not worthy of prosecution, the prosecutor has three options: decline to file altogether, pursue the crime as a probation or parole violation if the offender was already subject to probation or parole, or divert the offender into some form of rehabilitation program. (Rarely do prosecutors affirmatively declare a suspect innocent; they simply abstain from prosecuting those cases that seem unmeritorious.) This initial filing decision forms the essence of what is meant by prosecutorial discretion.

In some civil law jurisdictions, prosecutors have no such discretion; once they conclude that a reasonable amount of evidence exists to support the charges, they must initiate a criminal case. (In these nations, undesired consequences of overly broad laws are meant to be handled by code amendments rather than by selective charging strategies.) Whether prosecutors in compulsory jurisdictions vigorously pursue all the cases they file is another matter, however. They might allow a weak case to lapse or fail to diligently conduct further investigation, thereby forcing the court to dismiss the case at some later date. Given these tendencies, even jurisdictions that formally subscribe to compulsory prosecution manifest variations in prosecutorial decision-making (Levine and Feeley 2001: 12227–8 (and cites therein)).

In discretionary jurisdictions, filing is not the only form of discretionary decision-making by prosecutors. Prosecutors also exercise significant discretion when making a plea bargain offer to a defendant, a process that involves asking the defendant to waive his right to trial, plead guilty to certain charges in the indictment, and in return receive a reduced punishment. The prosecutor's discretionary choice at the plea offer stage is twofold: she must first decide what sort of punishment reduction is appropriate for the defendant in light of the facts, the defendant's background, and other relevant circumstances, and secondly, she must decide which of the filed charges the defendant must admit in order to receive this reduced sentence. (The defendant can accept the deal, decline the prosecutor's offer, or make a counteroffer, which the prosecutor may then accept, decline, or counter, pursuant to her discretion.)

In both the filing and plea offer contexts, prosecutors tend to be guided by instrumental objectives. They determine how much the case is "worth" in terms of penalty and then file charges (or make plea offers) to achieve that result. The penal codes of most jurisdictions offer a wide menu of options for prosecutors to choose from, because a single behavior can be characterized as any number of different crimes, each of which carries its own set of penalties, and the prosecutor is not legally obligated to choose either the least or the most serious alternative. For that reason, the prosecutor deciding whether and how to charge a case, or whether and how much to deal a case, pays as much attention to his intuitions about the "right" outcome for the defendant as to the provisions of the substantive criminal law. The relationship between justice and the "right" outcome is not always clear or consistent; the two might be in tension with each other, or justice might be viewed as an integral component of the "right" result.

Given the extraordinary ability of the prosecutor to shape the outcome of a criminal case through his discretionary decisions, one naturally wonders about the ability of other criminal justice institutions to constrain this power. That is the subject of the next section of this entry.

External Mechanisms That Influence Prosecutorial Discretion

This section offers a quick assessment of the external sources of regulation on prosecutorial discretion – legislatures, judges, juries, state bar authorities, and voters. As the literature has documented, these outside institutions are remarkably ill-equipped or unwilling to constrain prosecutorial discretion in state criminal justice systems. Formal legal controls over the institution of prosecution are relatively weak, and voters are able to exert only muted control over prosecutors due to information deficits and the relative absence of contested elections.

The first place to look for regulations on prosecutorial behavior is the criminal code, crafted by the state legislature. By defining crimes narrowly and setting penalties at modest levels, legislators can confine the discretion of prosecutors in the jurisdiction. When it comes to criminal regulatory approaches in the United States, though, parsimony does not flourish. Voters expect prosecutors to take the lead in addressing crime, and they expect legislators to give them the legal tools to do the job. Legislatures do exactly that. Over time they tend to multiply – not reduce – the legal tools available to prosecutors, amending criminal codes to cover more behavior, to increase the range of punishments that apply to criminal behavior, and to intensify the overlap between criminal code provisions (Stuntz 2001). Legislatures thus augment the discretionary authority of prosecutors at both the filing and plea offer stages by providing them with a dizzying array of choices and few requirements.

Though legislators do not regulate prosecutorial discretion through the terms of the substantive criminal law, they might motivate or restrict prosecutors through a combination of carrots and sticks, such as grant programs or direct instructions for how to perform their work. Such techniques gain popularity with legislatures who want to appear responsive to victim complaints and press reports of under-prosecution for certain crimes, as these programs tend to raise the priority status of the targeted crimes vis-à-vis others that occur in the community. For example, the state budget might include funding for local offices to hire extra prosecutors to pursue designated crimes that offices might otherwise choose to ignore. Alternatively, legislatures might instruct local prosecutors to file more charges for specified crimes or to reduce the percentage of dismissals in previously filed cases that allege certain law violations.

While these legislative directives have the potential to be meaningful regulatory devices, their current impact is small due to their exceptional and tentative nature. Budgetary line items that direct prosecutors to devote resources to one type of crime rather than another are unusual, and they represent only a small fraction of an office's annual budget. Most of the money that arrives in the local prosecutor's office each year does not have strings attached, meaning the chief prosecutor can allocate his funds to meet locally identified priorities without having to explain his decisions to state lawmakers (for a discussion of how local priorities produce variation among federal prosecution offices, see O'Neill 2004:1486). As for the crime-specific "no drop" laws, which purport to be hard and fast rules limiting prosecutorial discretion, they leave plenty of room for the individual prosecutor to exercise judgment and to manage her caseload. For example, before filing a "no drop" case the individual prosecutor must still determine whether there is a minimum factual basis for the charge, and she is likely to have little trouble fitting cases into exceptions when she wants to reduce or dismiss a case that has already been filed.

Like the legislature, the judiciary has assumed a muted role in monitoring and controlling the discretion of criminal prosecutors. When judges are asked to override prosecutor choices about the selection or pretrial disposition of charges, they apply separation of powers doctrine and conclude that such decisions are executive choices. Applying the formal law of due process, the judge considers only whether the charges have some minimal factual support in the available evidence. Courts will, however, overturn a prosecutorial decision that is based on race, gender, or some other prohibited ground (Oyler v. Boyles 1962).

Judges do have the statutory authority to accept or reject plea agreements that the parties present to them, and judges concerned about inappropriate deals will question the advocates about their terms. But these judicial powers operate within a system of mass justice where institutional competencies differ and power arrangements shift over time. With respect to competency, the parties typically know more than the judge about both the alleged crime and the defendant's background; for that reason, it is rare for the judge to override a plea deal that both parties support. On the power front, the introduction of mandatory minimum sentence requirements and sentencing guideline schemes has led to a decrease in judicial authority relative to prosecutorial authority (Miethe 1987; Engen and Steen 2000). Legislative innovations intended to thwart judicial leniency thus have merely shifted the critical decision point and the power balance in the criminal court, as prosecutorial charge determinations have become squarely determinative of sentence outcomes in a wide range of cases. This has led scholars to describe the current framework as an "administrative" rather than adjudicatory, criminal justice system (Lynch 1998).

While judges have lost some of their ability to rein in prosecutors, juries have retained their theoretical role as a check on prosecutorial discretion through their power to nullify. Jury nullification denotes the power of the jury to return not guilty verdicts despite strong evidentiary support for the crimes charged. By exercising this power, juries can signal to prosecutors that the charges brought against a particular defendant ought not to have been filed or that the criminal code provision is at odds with community values or priorities. But the jury's nullification power is rarely more than theoretical in light of courthouse realities. Fewer than 5 % of criminal cases are heard by juries (most are plea bargained out of the system at the pretrial stage), and most judges do not allow juries to learn of their power to nullify – either through court-authored jury instructions or attorney argument. As a result, few criminal defendants can count on being saved from prosecutorial excess by a jury with a conscience (or an ax to grind).

Prosecutors as attorneys are also subject to the rules of the legal profession, which instructs and regulates its own members pursuant to disciplinary codes and bar tribunals. But state bar organizations have brought relatively few disciplinary proceedings against prosecutors, and those that have been brought tend to result in only light punishment. Moreover, the courts have routinely found that prosecutors are entitled to absolute immunity from civil damages for all professional decisions made in the course of prosecuting a case, so a criminal defendant who challenges the prosecutor's discretionary filing or bargaining strategy in his case would face an absolute bar to recovery in civil court.

In the absence of robust standards stemming from the legislature, judiciary, jury, or state bar, perhaps state-level elections keep the chief prosecutor's discretion within acceptable boundaries. In contrast to the 93 appointed chief prosecutors who populate the federal criminal justice system (aka United States Attorneys), the more than 2,300 chief prosecutors in the much higher-volume state systems must run for office. Many chief prosecutors are elected on a countywide or citywide basis and must live in the community that elects them. (Occasionally a chief prosecutor is appointed by the governor to fill a recently vacated post in between electoral periods.) For that reason, the pressures of the ballot might subject the elected prosecutor to the wishes of the population she serves or at least force her to pay attention to those wishes when exercising her discretion.

Yet elections of prosecutors deliver less than they promise by way of accountability due to incumbency advantage and the content of electoral discourse. Incumbents enjoy reelection rates of 95 % or higher in most jurisdictions, and about 85 % of them run unopposed (Wright 2009). Aside from the benefit that arises from their position, incumbents in contested prosecutorial elections rarely encounter meaningful public scrutiny of their policies or priorities for the office. Instead, elections turn on generic claims about "competence," familiar but unhelpful measures (like the "conviction rate" of the office), and – most common of all – claims about high-profile cases (both successes and failures). Election rhetoric does not highlight ideological or policy differences between candidates nor does it describe the nature of discretionary decision-making by either the chief or her assistants.

Taken together, these external institutions are sure to have some cumulative regulatory effects on prosecutorial discretion. They do part of the work to ensure that prosecutor choices comply with the law and with current public priorities. They do not, however, explain why prosecutors make the decisions they make in individual cases. To fully understand prosecutorial discretion, one should consider practices that originate from inside the prosecutor's office as well.

Internal Mechanisms That Influence Prosecutorial Discretion

In light of the gaps left by external regulatory approaches, what are the prospects for *internal* regulation of prosecutorial discretion? Some legal commentators considering the sources of internal regulation have focused on the value of one's internal moral compass, concluding that a prosecutor's professional conscience keeps her individually committed to the ideal of responsible behavior in the exercise of her discretion (Cassidy 2006). But this portrayal of the prosecutor as a self-determined island is vastly incomplete. For one thing, the professionals who work together routinely in a criminal courtroom – the prosecutor, the defense attorney, the judge, the clerks, and so forth – influence one another in predictable ways. Some prominent scholarship has explored the impact of these "working groups" of courtroom actors, noting that the members of such groups tend to prioritize the maintenance of their relationships over individual goals and that these relationships affect prosecutorial choices in given cases (Nardulli et al. 1988).

Working groups within the courtroom are not the only relevant group in a prosecutor's world: the home office can also significantly shape a prosecutor's understanding of the justice system, public priorities, and even what qualifies as a crime. Sometimes this office's influence is explicit, as when the chief prosecutor announces office policies that restrict the discretion individual line prosecutors may exercise for certain crimes (like the imposition of compulsory prosecution for crimes involving firearms) or at certain times (like the imposition of limits on how much plea bargaining will be tolerated after filing). The leadership of an office can also create monitoring and enforcement mechanisms within the office to learn about departures from the announced policies and to discourage them. These features are commonplace in federal prosecution offices (called United States Attorneys' Offices), which are all organized under the US Department of Justice. Such formal bureaucratic controls are also a routine part of the prosecutorial services in many other countries (Wright and Miller 2010; Johnson 1998).

Beyond the explicit instructions or direct oversight received from the elected prosecutor or her chief assistant, line prosecutors take their cues from the unarticulated values at work in the office. These often unspoken principles can affect prosecutorial discretion in a variety of ways. For example, they might address the best treatment of a particular type of offense or offender, or they might involve general attitudes about the importance of consistency in handling all types of cases. Like the official policies announced by the chief, these unofficial, word-of-mouth standards have the capacity to both guide and constrain discretionary choices by prosecutors in their individual caseloads.

Finally, structural aspects of the home office may guide prosecutorial decision-making. Various organizational features of the prosecutor's office – size of the workforce, status hierarchies, hiring strategies, job assignments, promotion ladders, physical arrangement of offices, access to the boss, the existence of vertical prosecution units, and the like – structure the professional identities and work lives of the attorneys in the office (see Levine and Wright 2012). These aspects of the social architecture inside the office combine with aspects of the office's external professional environment – such as the proximity to the courthouse, relationships with the defense bar, relationships with the judiciary, relationships with local law enforcement, stability of courtroom personnel, and docket size and diversity (Utz 1978; Jacoby 1977) – to produce the overall institutional environment in which the individual prosecutor works. While the prosecutor may not explicitly recognize environmental factors as controlling when he is considering whether to file charges or to offer a plea deal, their influence is nonetheless observable. For example, scholars have found that offices display certain overall tendencies that distinguish them from other offices in terms of willingness to bargain with the defense bar or to challenge the local bench. Moreover, in most jurisdictions and with respect to most crimes, local law enforcement officers retain significant discretion about which cases to forward to the prosecutor's office in the first place, thereby affecting the diet of potential cases that prosecutors even see. Hence, prosecutorial willingness and ability to press forward certain kinds of cases may depend significantly on law enforcement willingness and ability to properly investigate those cases.

In sum, while external institutions have yielded a lot of ground to prosecutors in the battle over the need for limits on filing and plea strategies, forces inside the prosecutor's own office – explicit policies, behavioral norms, and structural arrangements – tend to keep most individual prosecutors within reasonable (or at least fairly predictable) bounds when it comes to exercising their individual case management choices. Of course, if an office's internal constraints are inadequate, misguided, or corrupt, as when an office fails to properly train its employees regarding their Constitutional disclosure obligations, uses of discretion become less predictable and more risky in terms of legality.

The final section of this entry moves away from the structural perspective to address case- and offender-level characteristics that seem to influence prosecutorial decisions. It also considers scholarly treatment of the risks that such decisions will discriminate against, or further stereotypes about, crimes, offenders, and victims.

Case and Offender Traits That Influence Prosecutorial Discretion

Prosecutors in the USA are under no legal imperative to achieve (or to prioritize) uniformity in case processing; unless internal office policy demands otherwise, they need not worry about trading consistency for individualized treatment of suspects accused of similar crimes (Remington 1979; cf. Johnson 1998, discussing the consistency imperative in Japan). The backstop is the Equal Protection Clause of the Constitution, which prohibits filing or plea strategies based on the offender's membership in a protected class (Oyler v. Boyles 1962). Because the Constitution can tolerate variation that falls short of this kind of intentional discrimination (and because intentional discrimination is exceptionally difficult to prove), empirical work on prosecutorial discretion has centered on identifying the components of individualized treatment, with a normative inquiry into whether certain factors are legally appropriate to include in the filing or punishment calculus. Accordingly, work in this area tends to compare the impact of legally relevant variables with that of legally irrelevant variables on prosecutorial decision-making and case outcomes (see, e.g., Adams and Cutshall 1987; Williams et al. 2007).

Most studies, whether based on qualitative or quantitative data, find that legally relevant variables strongly correlate with both filing decisions and case outcomes. For example, prosecutors assert (and independent review of case files demonstrates) that offenders with significant criminal histories are more likely to suffer adverse consequences in the criminal justice system at every level – to have charges filed against them, to have serious charges filed against them, to have higher bail amounts requested and granted, to be convicted of at least one crime charged in the indictment, and to receive sentences involving incarceration. There is also a positive correlation between these outcomes and evidence strength or crime seriousness (judged by witness number and credibility, level of injury, quantity of drugs, amount of dollar loss, availability of forensic reports, and the like).

For many scholars of prosecutorial discretion, the influence of these factors is expected and defensible; after all, the legally relevant components of a crime and of the offender's background should affect professionals' judgment about the need for criminal justice intervention, particularly when resources are scarce and trial outcomes are uncertain. For others, facially neutral charging and case-handling strategies mask stereotypes about crimes, criminals, and victims that lead criminal justice officials "to focus on some offenses and offenders more than others" (Dawson 2004:106), which in turn causes them to reproduce and reinforce societal biases in the courtroom (Frohmann 1997; see also Spears and Spohn 1997). It may also cause them to underenforce certain crimes, as they implicitly read into the statute extra requirements (elements that the legislature did not explicitly include) in order to conserve enforcement resources or to avoid presenting juries with ambiguous cases (McPhail and Jenness 2005; Levine 2006).

Likewise, the empirical salience of legally irrelevant, or "extralegal," factors in determining case outcomes gives rise to concerns about the exercise of official discretion in the absence of meaningful external controls. Criminal justice scholars have long questioned whether prosecutors exercise their discretion – file charges, request high bail amounts, recommend longer sentences – in ways that disadvantage men, people of color, the poor, nontraditional victims, and other minorities. For example, numerous studies have shown that race is a significant variable in cases involving interpersonal violence (see, e.g., Walsh 1987), although sometimes it is the race of the victim, rather than the race of the defendant, that produces a statistically significant effect (see, e.g., Kingsworth et al. 2002: 562). Findings regarding the impact of offender race on prosecutorial treatment of nonviolent offenses have been more mixed (cf. Engen et al. 1999 with Steffensmeier and Demuth 2000).

Geographic variation within districts that are all subject to the same formal law has also come under scrutiny, with scholars asking whether offenders receive different treatment based on which office files the case against them. Intra-district variation is troubling but not unexpected, given the chief prosecutor's ties to his or her electoral community and the dimensions on which communities vary. Scholars have found a correlation between community poverty levels, community political conservatism, and prosecutorial dismissals (Franklin 2010) and between community demographics and prosecutorial willingness to file charges with mandatory minimum sentences (Ulmer et al. 2007). Geography and defendant characteristics also may not be mutually exclusive influences when it comes to the filing decision. For example, Franklin found that prosecutors in different counties did not use traits like age, race, sex, and prior criminal history in the same way; rather, "the degree to which these factors influenced prosecutorial decision-making (in the initial filing decision) varied depending on the jurisdiction" (Franklin 2010: 697).

The problems with treating similarly situated defendants in a disparate manner become particularly acute in the death penalty context. There has been quite a bit of scholarly interest in the prosecutor's decision to file a murder case as a capital murder case, given that most state statutes authorize capital murder prosecutions in a variety of settings but do not require even the most heinous murders to be pursued in this fashion. For example, scholars have documented significant disparities between counties located in the same state, applying the same formal law, when it comes to their rate of death penalty filings per "death eligible" murder case (American Bar Association 2006), leading to policy calls for more transparency and consistency in the process in order to assure that one's chances of ending up on death row do not arbitrarily depend on county boundaries. Others have found that, irrespective of jurisdiction, the race of the victim has a strong influence on whether a homicide is pursued as capital murder as opposed to something less (Williams et al. 2007). Because those who kill whites are statistically more likely to be prosecuted for capital murder than those who kill African-Americans, prosecutors have been accused of unfairly devaluing the lives of minorities relative to whites. So far, courts have refused to consider these sorts of statistical disparities as legally actionable under the Equal Protection Clause, given the absence of proof of intent to discriminate.

Conclusion

The discretionary decision to charge a suspect with a crime or to offer a defendant a deal prior to the start of trial results from a complex network of factors, some explicit, some implicit in the mind of the prosecutor. The relative absence of external controls on the prosecutor's decision-making process has led to calls for more oversight and accountability, particularly in light of the decrease in judicial authority that has resulted from sentencing guideline schemes and mandatory minimum requirements. While a certain amount of variation is perhaps inevitable given disparities in local needs, resources, priorities, and personalities, institutional mechanisms are necessary to ensure that prosecutorial choices are thoughtfully made, wisely and consistently implemented, and subject to evaluation. Without such efforts, the power of the prosecutor is likely to grow stronger, fed by the public's fear about rising crime rates, politicians' inclination to expand rather than contract the web of criminal statutes, and courts' reluctance to impose doctrinal limits on prosecutorial behavior that falls short of intentional discrimination.

Related Entries

▶ Evolving Judicial Roles
▶ History of Probation and Parole in the United States
▶ Jury Impartiality in the Modern Era
▶ Juvenile Diversion
▶ Plea Bargaining
▶ Prosecution and Wrongful Convictions
▶ Race and the Likelihood of Arrest
▶ Victimization, Gender, and the Criminal Justice System

Recommended Reading and References

Adams K, Cutshall C (1987) Refusing to prosecute minor offenses: the relative influence of legal and extralegal factors. Justice Q 4:595

Albonetti C (1987) Prosecutorial discretion: the effects of uncertainty. Law & Soc'y Rev 21:291

American Bar Association (1996) Evaluating fairness and accuracy in state death penalty systems: The Georgia death penalty assessment report. American Bar Association, Washington, DC

Cassidy RM (2006) Character and context: what virtue theory can teach us about a prosecutor's ethical duty to "seek justice". Notre Dame Law Rev 82:635

Davis KC (1969) Discretionary justice: a preliminary inquiry. Louisiana State University Press, Baton Rouge

Davis A (2007) Arbitrary justice: the power of the American prosecutor. Oxford University Press, New York

Dawson M (2004) Rethinking the boundaries of intimacy at the end of the century: the role of victim-defendant relationship in criminal justice decisionmaking over time. Law & Soc'y Rev 38:105

Engen RL, Steen S (2000) The power to punish: discretion and sentencing reform in the war on drugs. Am J Sociol 105:1357

Engen RL, Gainey RR, Steen S (1999) The impact of race and ethnicity on charging and sentencing processes for drug offenders in three counties of Washington State. Washington State Minority and Justice Commission

Franklin TW (2010) Community influence on prosecutorial dismissals: a multilevel analysis of case- and county-level factors. J Crim Justice 38:693

Frohmann L (1997) Convictability and discordant locales: reproducing race, class, and gender ideologies in prosecutorial decision making. Law & Soc'y Rev 31:531

Jacoby J (1977) The prosecutor's charging decision: a policy perspective. US Department of Justice, law Enforcement Assistance Administration, National Institute of Law Enforcement and Criminal Justice, Washington

Johnson D (1998) The organization of prosecution and the possibility of order. Law & Soc'y Rev 32:247

Kingsworth RF, MacIntosh RC, Sutherland S (2002) Criminal charge or probation violation? Prosecutorial discretion and implications for research in criminal court processing. Criminology 40(3):553

LaFree G (1989) Rape and criminal justice: the social construction of sexual assault. Wadsworth, Belmont

Levine KL (2006) The intimacy discount: prosecutorial discretion, privacy, and equality in the statutory rape caseload. Emory Law J 55:691

Levine KL, Feeley MM (2001) Prosecution. In: The international encyclopedia of the social and behavioral sciences. Elsevier, London, pp 12224–12230

Levine KL, RF Wright (2012) Prosecution in 3-D. J Crim Law Criminol 102(4):1119–1180

Lynch G (1998) Our administrative system of criminal justice. Fordham Law Rev 66:2117

McPhail B, Jenness V (2005) To charge or not to charge? – That is the question: the pursuit of strategic advantage in prosecutorial decision-making surrounding hate crime. J Hate Stud 4:89

Miethe T (1987) Charging and Plea bargaining practices under determinate sentencing: an investigation of hydraulic displacement of discretion. J Crim Law Criminol 78:155

Miller M, Wright RF (2008) The black box. Iowa Law Rev 94:125

Nardulli P, Eisenstein J, Flemming RB (1988) The tenor of justice: criminal courts and the Guilty Plea process. University of Illinois Press, Urbana

O'Neill ME (2004) Understanding federal prosecutorial declinations: an empirical analysis of predictive factors. Am Crim Law Rev 41:1439

Oyler v. Boyles (1962) 368 U.S. 448, 456

Remington F (1993) The decision to charge, the decision to convict on a Plea of Guilty, and the impact of sentence structure on prosecution practices. In: Ohlin L, Remington F (eds) Discretion in criminal justice: the tension between individualization and uniformity. State University of New York Press, Albany

Spears J, Spohn C (1997) The effect of evidence factors and victim characteristics on prosecutors' charging decisions in sexual assault cases. Justice Q 14:501

Steffensmeier D, Demuth S (2000) Ethnicity and sentencing outcomes in US federal courts: who is punished more harshly? Am Sociol Rev 65(5):705

Stuntz W (2001) The pathological politics of criminal law. Mich Law Rev 100:505

Ulmer JT, Kramer JH (1998) The use and transformation of formal decision-making criteria: sentencing guidelines, organizational contexts, and case processing strategies. Soc Probs 45:248

Ulmer JT, Kurlychek M, Kramer JH (2007) Prosecutorial discretion and the imposition of mandatory minimum sentences. J Res Crime Delinq 44(4):427

Utz P (1978) Settling the facts: discretion and negotiation in criminal court. Lexington Books, Lexington

Vera Institute of Justice (1981) Felony arrests: their prosecution and disposition in New York City's courts. Vera Institute of Justice, New York

Walsh A (1987) The sexual stratification hypothesis and sexual assault in light of changing conceptions of race. Criminology 25:153

Williams MR, Demuth S, Holcomb J (2007) Understanding the influence of victim gender in death penalty cases: the importance of victim race, sex-related victimization and jury decision making. Criminology 45(4):865

Wright RF (2009) How prosecutor elections fail us. Ohio State J Crim Law 6:581

Wright RF, Miller M (2002) The screening/bargaining tradeoff. Stanf Law Rev 55:29

Wright RF, Miller M (2010) The worldwide accountability gap in criminal prosecution. Wash & Lee Law Rev 67:1587

Prostitution

Ronald Weitzer
Department of Sociology, George Washington University, Washington, DC, USA

Overview

Most research focuses on street prostitution and female sex workers, with much less attention devoted to indoor prostitution, male and transgender workers, clients, and third parties including managers. Furthermore, most of the literature examines prostitution where it is illegal, neglecting nations where it is legal and regulated by the government. This entry summarizes theory and research and argues that research on neglected topics will enrich our understanding of contemporary prostitution.

Introduction

Prostitution is one of the "vices" that is distinguished by moral ambivalence: It is considered both deviant and enticing or pleasurable at the same time. Unlike predatory crimes with a clear victim, prostitution involves exchanges between willing buyers and sellers, which leads some members of the public to view it with a measure of tolerance even if they hold an otherwise negative view of it. The same moral ambivalence applies to pornography, gambling, and marijuana use.

The research literature on prostitution is sizeable but remains deficient in some important respects. Much more attention has been paid to street prostitution than to indoor prostitution and to female prostitutes much more than other actors (male and transgender workers, clients, managers). This entry highlights key aspects of contemporary prostitution and then draws attention to several under-examined issues. The focus is mostly on Anglo-American societies.

Competing Theories

Three main theoretical perspectives have been applied to sex work. One of these, the *oppression paradigm*, holds that prostitution reflects and reinforces patriarchal gender relations. Advocates of this paradigm argue that exploitation, subjugation, and violence against women are *intrinsic to and ineradicable from* sex work (Barry 1995; Jeffreys 1997), and these advocates typically use dramatic language to highlight the plight of workers ("sexual slavery," "prostituted women," "survivors"). Such terminology is meant to emphasize the notion that prostitutes are victims and that prostitution is not something that can be chosen. Likewise, clients are called "prostitute users" and "sexual predators."

Some of these claims are not amenable to verification or falsification, and the very definition of prostitution as inherently oppressive is one dimensional. Oppression theorists typically describe *only* the worst examples of sex work and treat them as representative, and they tend to ignore counterevidence that challenges the theory. Despite being a fairly extreme paradigm, it remains one of the three main theoretical perspectives on sex work.

The *empowerment paradigm* is radically different. It focuses on the ways in which sexual services qualify as work and may be potentially liberating for workers (Chapkis 1997; Delacoste and Alexander 1987). This paradigm holds that there is nothing inherent in sex work that would prevent it from being organized like any other economic transaction for the mutual gain of buyers and sellers alike. Prostitution can benefit sellers economically and can, under the right circumstances, provide greater control over working conditions than many traditional jobs. Writers who adopt the empowerment model also argue that most of the tenets of the oppression model reflect the way in which *some* sex work manifests itself when it is *criminalized*.

Most empowerment theorists argue that sex work is *potentially* empowering, not that it is necessarily so. They tend to neglect sex workers who labor under harsh conditions and instead highlight success stories in prostitution. It can be liberating for those who are "fleeing from small-town prejudices, dead-end jobs, dangerous streets, and suffocating families" (Agustin 2007, p. 45).

Both of these paradigms are one dimensional. While exploitation and empowerment are certainly present in prostitution, there is sufficient diversity across time, place, and sector to demonstrate that prostitution cannot be reduced to a single phenomenon. An alternative perspective, the *polymorphous* paradigm, holds that there is broad variety of occupational arrangements, power relations, and experiences of those who sell and buy sex. This paradigm is sensitive to complexities and to the structural conditions shaping the uneven distribution of agency, subordination, and job satisfaction (Weitzer 2010). The polymorphous perspective is superior to the other two paradigms because it is supported by a wealth of social science evidence (see Shaver 2005; Vanwesenbeeck 2001; Weitzer 2010).

Types of Prostitution

Prostitution is defined as the exchange of sexual services for material compensation. There is a great deal of *variation* in prostitution within and between societies and across different echelons. Prostitutes differ in their reasons for entry, number and type of clients, freedom to refuse clients, access to resources for protection, relationships with colleagues, dependence on third parties, experiences with the authorities, public visibility, and impact on the surrounding community.

Street Versus Indoor Settings

In street prostitution, the initial transaction occurs in a public place (sidewalk, park, truck stop), and the sex act takes place in either a public or private setting (alley, car, park, hotel, etc.). Indoor prostitution takes place in brothels, massage parlors, bars, hotels, private premises, and other places. In many countries, indoor prostitution is much more prevalent than street prostitution. On many dimensions of prostitution, street and indoor sectors differ substantially. It is not the sheer fact of being indoors that distinguishes indoor from street prostitution; instead, certain features of indoor settings are preconditions for a work environment that can differ substantially from street-level work. Such indoor features include the capacity for better screening of clients, less accessibility to street predators, and greater safety if one works indoors with other prostitutes or with a manager. In addition, prostitutes who work indoors generally come from at least somewhat more advantaged socioeconomic backgrounds than those who work on the streets.

Individuals enter prostitution through different paths. Some are recruited by pimps or otherwise coerced into the trade; others drift into prostitution gradually and tentatively, often at the encouragement of friends. Some initially worked in other branches of the sex industry (e.g., strip clubs, phone sex) and later decided to experiment with prostitution. Many street prostitutes are runaways who end up in a new locale with no resources and little recourse but to engage in some kind of criminal activity. Economic motives predominate throughout the trade, ranging from survival to a desire for financial independence or upward mobility.

A segment of the street population is involved in "survival sex": They sell sex out of dire necessity. Many street workers have a history of childhood abuse (violence, incest, neglect). Many work and live in crime-ridden areas; are socially isolated and disconnected from support services; use addictive drugs; are in poor health; are exploited and abused by pimps; and are vulnerable to being assaulted, robbed, raped, or killed on the streets. Off-street work is less hazardous. Studies that compare street and indoor workers, using well-constructed, purposive samples, find substantial and sometimes huge differences in victimization rates. In addition to street-indoor differences in *ever* being victimized, similar disparities have been documented in the *frequency* and *severity* of victimization.

This does not mean that off-street work is risk-free: Indoor work in the Third World, for example, is typically conducted under harsher conditions than in developed countries. Still, there is no doubt that indoor settings are *generally* safer than the streets: Indoor workers and their managers are in a better position to screen clients prior to contact; they may work in a place with video surveillance and alarm systems and if they work with others, can rely on colleagues to intervene in the event of trouble with a client.

Street and indoor prostitutes also differ in the *services* they provide. Street workers spend little time with customers and conversation is kept to a minimum. Indoor interactions are typically longer and richer and include conversation, foreplay, nonsexual intimacy, and a semblance of romance. Studies also show that escorts and brothel workers are much more likely than street workers to receive a massage, manual stimulation, or oral sex from their clients (Lever and Dolnick 2010; Woodward et al. 2004). Some indoor workers *expect* such conduct from clients as a routine part of the encounter. For them, the experience revolves around *both* party's sexual pleasure.

Like other jobs, prostitution does not have a uniform effect on workers' psyches and self-images. Research indicates, as one would expect, that street prostitutes are much more likely than both nonprostitutes and indoor sex workers to exhibit psychological disorders and low self-esteem. The stress and danger associated with street work contribute to psychological problems.

A comparative study of 75 streetwalkers and 75 call girls in California and 150 legal brothel workers in Nevada found that 97 % of the call girls and 50 % of the brothel workers reported an *increase* in self-esteem after they began working in prostitution, compared only 8 % of the streetwalkers (Prince 1986, p. 454). And an Australian study found that half of the 102 call girls and 103 brothel workers interviewed felt that their work was a "major source of satisfaction" in their lives, while 7 out of 10 said they would "definitely choose" this work if they had it to do over again (Woodward et al. 2004, p. 39). A majority of indoor workers in other studies similarly report that they enjoy the job, feel that their work has at least some positive effect on their lives, or believe that they provide a valuable service. Self-esteem and job satisfaction are higher at the mid and upper levels of prostitution because of workers' higher education, greater income, control over working conditions, and better screening of clients.

Sex workers do not necessarily confine themselves to one type of work for their entire career. There is some occupational mobility. Some strippers and porn performers, male and female, engage in paid sexual encounters with their fans. It is rare, however, for workers to experience substantial upward or downward mobility. Transitioning from street work to the middle and top echelons is quite rare, because most street workers lack the education and social skills associated with upscale work.

The studies reviewed here provide strong evidence contradicting some popular myths as well as the oppression paradigm in academia. While certain factors are universal (minimizing risks, managing client behavior, coping with stigma), other work-related experiences, as well as the harms typically associated with prostitution, vary greatly. The prostitution market is *segmented* between the indoor and street sectors – marked by major differences in working conditions, risk of victimization, and job satisfaction and self-esteem.

Male and Transgender Prostitution

Most theory and research centers on female prostitutes. Transgender prostitutes rarely have been studied, whereas male prostitutes have attracted somewhat more attention from scholars. The limited research points to both similarities and differences between the three types. Like women on the street, young male street prostitutes often enter the trade as runaways or to support a drug habit. Like mid- and high-end female workers, men working at those levels have at least some regular clients and may develop emotional attachments to them. And, like female workers in the mid- and upper-level tiers, similarly situated men are more likely than their street counterparts to view their work positively (Minichiello et al. 2001; West 1993). Uy et al. (2007) found that a sample of male escorts felt desired and validated as a result of being generously paid for sex; their self-confidence and body images improved the longer they sold sex.

Gender makes a difference in the following areas. Compared to women, men:

- Are less likely to be coerced into prostitution, to have pimps, and to experience violence from customers because males are able to exercise greater physical power over customers
- Exercise greater control over their working conditions
- Are less stigmatized within a sector of the population (the gay community) but more stigmatized in the wider society because of the combination of prostitution and homosexual behavior

Too little is known about transgender sex workers to draw even tentative conclusions along these bulleted dimensions. However, transgenders do face greater challenges than both female or male prostitutes: They have higher HIV infection rates, "usually have the least desirable prostitution location, make the least money, and are stigmatized and ridiculed by non-transvestite male and female prostitutes" (Boles and Elifson 1994, p. 85). Transgender sex workers are also more likely than males to be assaulted or raped while at work. Many male transgenders conceal from customers that they are not biological women, increasing the chances of altercations with deceived clients. Other customers expressly seek out transgenders who appear to be women but are anatomically male – something the customer may find "kinky" or thrillingly transgressive and allowing them to have the "best of both worlds" physically and sexually. This dynamic is unique to transgender sex workers and their customers.

A Brazilian study reported that for transgender sex workers, prostitution was the only sphere of life that enhanced their self-image: Prostitution gave them a "sense of personal worth, self-confidence, and self-esteem" (Kulick 1998, p. 136). They sold sex not only for the money but also for emotional and sexual fulfillment. In a San Francisco study, researchers discovered that "sex work involvement provided many young transgender women of color feelings of community and social support, which they often lacked in their family contexts." Another advantage was that sex work gave these individuals a "sense of independence and non-reliance on others (i.e., managers, co-workers) who might express discrimination or harassment" (Sausa et al. 2007, p. 772). Sex work was one of the few arenas in which they could shield themselves from societal rejection.

Customers

A significant number of men have bought sex. The General Social Survey reports, in ten polls from 1991 to 2008, that 3–4 % of American men had paid for sex in the past year and 15–18 % saying they had paid for sex at some time in their life. Similar figures are reported for Australia (16 %) and the average within Europe (15 %) (Rissel 2003). The real numbers are likely higher given the stigma involved in revealing such conduct. The majority of customers appear to have bought sex only once or a few times, while a minority are repeat or frequent clients.

Customers vary demographically (e.g., age, social class, marital status) just like the wider male population: A study comparing a large sample of customers with a nationally representative sample of American men found few differences between them (Monto and McRee 2005). Most clients do not act aggressively or violently: There is "no evidence to suggest that more than a minority of customers assault prostitutes" (Monto 2010, p. 243).

Customers buy sex for very different reasons:

- They desire sex with a person with a certain demeanor or physical appearance (e.g., physique, race, transgender).
- If they have a regular partner, they may be unsatisfied with the sexual dimension of their relationship and seek satisfaction elsewhere.
- They have difficulty finding a partner for a conventional relationship.
- They find this transgressive conduct thrilling.
- They seek to exercise control over or violence against women and act out this desire by targeting prostitutes because of the ease of access and belief that they will not report abuse to the police.
- They want to be free of the obligations of a conventional relationship.
- They seek an emotional connection in addition to sex.

A survey of 1,342 customers in three cities found that 41 % sought "a different kind of sex than my regular partner" wants, 43 % said that they were "excited by the idea of approaching a prostitute," 42 % said they "like to have a -variety of sexual partners," 42 % said they "like to be in control when I'm having sex," and 28 % said they did not want the responsibilities of a conventional relationship (Monto 2010).

Some clients are motivated by the possibility of forming a short- or long-term relationship with a sex worker. This is particularly evident in tourist locations, such as Thailand, the Philippines, and the Caribbean, where Western men meet dancers or bargirls at nightclubs and pay a bar fine to leave the club with a woman. They may spend a night or several days together (perhaps visiting tourist sites) with the man paying for the woman's time or for all expenses. Some of these men become boyfriends and enter into long-term or serial relationships, sending gifts and money from overseas and reuniting on return visits. Many of the women and some of the men are consciously seeking a long-term relationship, and some end up marrying. The women see such associations as a means of securing upward social mobility. A similar dynamic occurs among some gay tourists and male prostitutes, where paid sex can evolve into a protracted boyfriend relationship (Padilla 2007).

Some recent research sheds light on customers' *experiences* during paid-sex encounters. In addition to those Internet websites where workers advertise (sites listing services and prices, biographical sketches, photos, etc.), some websites contain message boards for clients. These sites provide novices with information on what to expect in prices and services, client "reviews" of a specific worker (appearance, behavior), location of establishments, and other useful information and accounts of personal experiences. The sites provide unique insight into customer expectations, justifications, and conduct norms – dimensions addressed only partially in previous interviews and surveys. Derogatory comments about certain providers are made in some online entries but others lavish praise on specific workers. And many of the cyber exchanges among men draw from a code of ethics for consumers: discussing inappropriate behavior toward sex workers, with warnings for misogynists, those seeking underage workers, and other deviants. Internet forums offer a unique window into customers' experiences, the meanings they attach to prostitution, and a new subculture.

Many clients of indoor workers seek more than sex. They also want a provider to be attentive and friendly; to engage in conversation; and to kiss, cuddle, and be otherwise affectionate. This kind of treatment has long been sought by customers of indoor workers, but recently, it has acquired a new label – the "girlfriend experience" (GFE), which borders on what one would expect from a girlfriend or the "boyfriend experience" for clients of male prostitutes. One study found that customers' online entries focused on nonsexual aspects of the GFE to a greater extent than on a sex worker's physical attributes or sexual performances (Sharp and Earle 2003). For some clients, the GFE is more physical and means that the sex is not mechanical but instead is experienced as "making love." Many customers value reciprocity: They are interested not only in satisfying themselves but also in giving a sex worker pleasure. Such clients, as well as those who seek emotional intimacy and companionship, challenge the oppression paradigm's assumption that commercial sex inherently involves objectification. Indoor clients in one study did not view sex workers as "targets of sexual conquest" or "simply as bodies"; instead, these men were paying for a meaningful, personal connection with a woman (Sanders 2008, p. 98).

Research shows that many have had very good paid-sex experiences and feel that their encounters have enhanced their lives. Others, however, have had negative feelings: They fear discovery, feel shame or guilt, or are dissatisfied with a particular encounter because of awkwardness, performance problems, being rushed, or feeling that they did not get what they paid for. Others feel embarrassed about paying for sex, fear contracting a disease, or fear being caught by wives, girlfriends, or the police.

Like any service occupation where workers spend an extended amount of time with clients, indoor sex workers view different clients differently. They like some clients tremendously and despise others and are indifferent to yet others. Some customers are demanding, belligerent, and difficult to handle; this does not appear to be the norm among clients of mid- and upper-echelon providers because they can be more selective. In one study, "most call girls have not had bad experiences, and more often than not, they have positive things to say about their customers" (Perkins and Lovejoy 2007, p. 112). As one call girl stated, "The only way I can sustain regulars is if I actually like them and I may like them for different reasons.... A number of my clients are intelligent men who are well informed and can carry on a stimulating conversation" (quoted in Perkins and Lovejoy 2007, p. 63). Bernstein (2007, p. 103) describes this as "authentic (if fleeting) libidinal and emotional ties with clients, endowing them with a sense of desirability, esteem, or even love." This is certainly not true for all indoor prostitutes, as some maintain strict affective boundaries to avoid developing strong feelings for their clients: They engage in "counterfeit intimacy" – nothing more than a manufactured emotional connection that only *appears* to be a GFE.

It might be assumed that all customers would prefer to buy sex services indoors because these settings are associated with greater safety, hygiene, a more pleasant atmosphere, and a much lower risk of being arrested by the police than what is typical of street-level sex.

Some indoor places, such as brothels and saunas, offer other kinds of recreation in addition to sex, such as a bar, karaoke singing, stripping, and gambling – attractions that clients may value as much as the sex for sale. Many clients indeed feel this way, but others prefer the streets. Streetwalkers may be preferred because of easy access and low prices, and some men get a thrill out of cruising for sex. These men are not seeking a GFE; the brevity of street transactions means that conversation is minimal and an emotional connection either impossible or rather truncated. Other clients would never consider buying sex on the street, because they are seen as dangerous places.

What about female customers? Little is known about this but it is a much smaller market than that catering to male clients. Vacation spots are one setting in which women may buy sex from men. In the Caribbean, for instance, affluent European and North American female tourists meet young local men on the beaches and at clubs (Phillips 1999; Pruitt and LaFont 1995; Taylor 2001). The similarities between male and female sex tourism include the economic inequality between buyer and seller. In both cases, local people may have few other options for earning the kind of money they can get from selling sex, while the foreign client comes from much more advantaged circumstances. Aside from this fundamental structural pattern, it is not known whether female sex tourists differ significantly from the male tourist/female provider type in terms of objectification of the worker, amount or type of exploitation, and the exercise of control over the workers. Finally, almost no research has been done on commercial sex transactions in which both parties are women. What differences, if any, are there between this type of exchange and transactions where both parties are men? Comparative research would help to clarify the ways in which gender differences affect the dynamics of commercial sex exchanges.

Third Parties

Most research focuses on either sex workers or customers, ignoring third parties. Much more research is needed on the dynamics of recruitment, socialization, surveillance, exploitation, coercion, and trafficking. The findings will contribute to a more comprehensive picture of diverse power relations, ranging from those types where workers experience extreme domination to those where they exercise substantial control and are free of mistreatment.

While many prostitutes work independently, others are managed by or involved with third parties. A manager is someone who exercises control over a worker and extracts some or all of the profit. This includes street pimps and those who run brothels, massage parlors, and escort agencies. Some management practices are consistent with those of any other business, while other aspects are unique to a sexually oriented and marginalized business. Workers generally expect managers to screen customers and act as allies, defending them against problem clients. But managers vary in the extent to which they live up to these expectations. Some provide the bare minimum of amenities to their workers, are lax about health and safety, favor certain employees over others, or treat all of their workers very poorly; others have collegial relations with their employees, rigorously screen customers, and generally take pains to ensure safe and healthy working conditions.

These different patterns are illustrated in a small but important body of research, including Heyl's (1977) study of the training of novice workers in one American brothel, Zheng's (2009) ethnography of erotic karaoke bars in a Chinese city, Trotter's (2008) ethnography of bars in the harbor areas of South Africa, where the bar girls cater to foreign sailors, Steinfatt's (2002) analysis of bar prostitution in Thailand, and Perez-y-Perez's (2003) participant observation in three New Zealand massage parlors. Each researcher documents the ways in which managers control, support, or exploit workers, and act as brokers between workers and customers. Like other businesses, the managers also deal with external forces, such as the police, political officials, and local residents when they attempt to intervene in the affairs of the establishment – for example, demanding bribes or free sex, imposing new restrictions on how the business operates, or calling for the place to be shut down. Middlemen are a different and unique third party, connecting sex workers with clients. They offer a vital service where direct client access to workers is hampered. For instance, middlemen in towns along transnational highways get paid by long-distance truck drivers who need interpreters and brokers to find a local prostitute; the middlemen recommend specific women based on the drivers' preferences and protect both the clients and the workers from robbery, assault, or reneging on agreed-upon services. These are some of the ways that third parties organize and manage prostitution.

Legal Systems

The legal context governing prostitution can be considered a critical variable, but most research has been conducted in nations where prostitution is prohibited. This means that what we think we "know" about prostitution may be distorted by the heavy focus on criminalized prostitution. The few existing studies of legal prostitution systems indicate that many of the harms often associated with prostitution are partly due to its illicit status. Where it is illegal, prostitution is set apart from legitimate work, workers are marginalized, and the authorities provide little if any protection. Legalization has the potential of alleviating these problems. And the fact that prostitution is legal and regulated in several nations shows that there are alternative prisms through which a population can view prostitution, perhaps including somewhat different moral universes within which this and other kinds of vice are regarded. It is possible, therefore, for this type of deviance to become at least somewhat normalized, as is true for some other vices, in the aftermath of legalization.

Certain kinds of prostitution are legal and regulated by the government in many more nations than is commonly thought, including the Netherlands, New Zealand, Australia, and Germany. It is technically illegal yet tolerated and even regulated in some other places, such as Belgium and Thailand. Regulations vary from place to place, but a common objective is harm reduction. Depending upon the nation, the law may give workers labor rights and provide for the screening of brothel owners and the licensing and taxing of their businesses; regulations may govern how establishments operate, dictate their size and their signage, and impose safe sex and other health requirements (Weitzer 2012).

The available evidence suggests that prostitution, when legalized and regulated by the government, can pay dividends although this is certainly not guaranteed. Nevada legalized prostitution in 1971 and currently has about 30 brothels scattered around its rural counties (but not in Las Vegas and Reno). These legal brothels "offer the safest environment available for women to sell consensual sex acts for money" (Brents and Hausbeck 2005, p. 289). Research on other legal prostitution regimes finds similarly that they tend to increase sex workers' safety (Weitzer 2012). In most of these cases, brothels have implemented screening procedures, surveillance, panic buttons, or listening devices that reduce the chances of abuse by customers and allow for rapid intervention in case of trouble.

Legalizing prostitution is one thing; implementing and enforcing regulations in accordance with the new law is another, and it has presented serious challenges in many nations post-legalization. Some of the common problems include difficulties in getting prostitutes and business owners to comply with the law, eliminating parasitical third parties such as pimps and traffickers, and dealing with continuing opposition from a segment of the public that wishes to tighten restrictions or repeal the law entirely. We are just beginning to identify regulations that work well and those that are problematic, and policymakers continue to grapple with unforeseen challenges in governing a legal prostitution sector. There are some exceptions to these teething problems, such as Nevada and New Zealand, where the aftermath of legalization has been relatively smooth, but elsewhere (e.g., Germany, Australia, the Netherlands), the implementation of legal arrangements has been buffeted by a host of problems, including the simultaneous growth of illegal prostitution in tandem with the legal sector – a two-tiered situation that has been difficult to resolve (see Weitzer 2012). Such problems are not unique to prostitution post-legalization but apply to other vices as well, as illustrated by the recent challenges posed by legalized medical marijuana in the United States.

Conclusion

The literature is heavily weighted toward studies of street prostitution, female sex workers, and illegal prostitution. To facilitate a more balanced and comprehensive picture, a major shift in research is needed – investigating indoor sex workers of all types, male and transgender providers, managers, and legal prostitution systems. Research in these areas will enhance our understanding of sex work and its polymorphous character.

Related Entries

- ▶ Feminist Criminological Theory
- ▶ Pornography
- ▶ Victimization, Gender, and the Criminal Justice System

Recommended Reading and References

Agustin L (2007) Sex at the margins: migration, labor, and the rescue industry. Zed, London

Barry K (1995) The prostitution of sexuality. NYU Press, New York

Bernstein E (2007) Temporarily yours: intimacy, authenticity, and the commerce of sex. University of Chicago Press, Chicago

Boles J, Elifson KW (1994) The social organization of transvestite prostitution and AIDS. Soc Sci Med 39:85–93

Brents BG, Hausbeck K (2005) Violence and legalized brothel prostitution in Nevada. J Interpers Violence 20:270–295

Chapkis W (1997) Live sex acts: women performing erotic labor. Routledge, New York

Delacoste F, Alexander P (eds) (1987) Sex work. Cleis, Pittsburgh

Heyl BS (1977) The madam as teacher: the training of house prostitutes. Soc Probl 24:545–555

Jeffreys S (1997) The idea of prostitution. Spinifex, Melbourne

Kulick D (1998) Travesti: sex, gender, and culture among Brazilian transgendered prostitutes. University of Chicago Press, Chicago

Lever J, Dolnick D (2010) Call girls and street prostitutes: selling sex and intimacy. In: Weitzer R (ed) Sex for sale: prostitution, pornography, and the sex industry. Routledge, New York, pp 85–100

Minichiello V, Marino R, Browne J (2001) Knowledge, risk perceptions, and condom usage in male sex workers from three Australian cities. AIDS Care 13:387–402

Monto M (2010) Prostitutes' customers: motives and misconceptions. In: Weitzer R (ed) Sex for sale: prostitution, pornography, and the sex industry. Routledge, New York, pp 233–254

Monto MA, McRee N (2005) A comparison of the male customers of female street prostitutes with national samples of men. Int J Offender Ther Comp Criminol 49:505–529

Padilla M (2007) Caribbean pleasure industry. University of Chicago Press, Chicago

Perez-y-Perez M (2003) Discipline, autonomy, and ambiguity: organizations, markets, and work in the sex industry, Christchurch, New Zealand. Ph.D. dissertation, University of Canterbury, Christchurch

Perkins R, Lovejoy F (2007) Call girls: private sex workers in Australia. University of Western Australia Press, Crawley

Phillips J (1999) Tourist-oriented prostitution in Barbados: the case of the beach boy and the white female tourist. In: Kempadoo K (ed) Sun, sex, and gold. Rowman and Littlefield, Lanham, pp 183–200

Prince DA (1986) A psychological study of prostitutes in California and Nevada. Doctoral dissertation. US International University, San Diego

Pruitt D, LaFont S (1995) For love and money: romance tourism in Jamaica. Ann Tourism Res 22:422–440

Rissel CE (2003) Experiences of commercial sex in a representative sample of adults. Aust N Z J Public Health 27:191–197

Sanders T (2008) Paying for pleasure: men who buy sex. Willan, Cullompton

Sausa L, Keatley J, Operario D (2007) Perceived risks and benefits of sex work among transgender women of color in San Francisco. Arch Sex Behav 36:768–777

Sharp K, Earle S (2003) Cyberpunters and cyberwhores: prostitution on the Internet. In: Jewkes Y (ed) Dot. Cons: crime, deviance, and identity on the internet. Willan, Portland, pp 36–52

Shaver F (2005) Sex work research: methodological and ethical challenges. J Interpers Violence 20:296–319

Steinfatt T (2002) Working at the bar: sex work and health communication in Thailand. Ablex, Westport

Taylor JS (2001) Dollars are a girl's best friend: female tourists' sexual behavior in the Caribbean. Sociology 35:749–764

Trotter H (2008) Sugar girls and seamen: a journey into the world of dockside prostitution in South Africa. Jacana, Auckland Park

Uy JM, Parsons JT, Bimbi DS, Koken JA, Halkitis PN (2007) Gay and bisexual male escorts who advertise on the internet: understanding the reasons for and effects of involvement in commercial sex. Int J Mens Health 3:11–26

Vanwesenbeeck I (2001) Another decade of social scientific work on prostitution. Annu Rev Sex Res 12:242–289

Weitzer R (2010) Sex work: paradigms and policies. In: Weitzer R (ed) Sex for sale: prostitution, pornography, and the sex industry. Routledge, New York, pp 1–43

Weitzer R (2012) Legalizing prostitution: from illicit vice to lawful business. NYU Press, New York

West D (1993) Male prostitution. Haworth, Binghamton

Woodward C, Fischer J, Najman J, Dunne M (2004) Selling sex in Queensland. Prostitution Licensing Authority, Brisbane

Zheng T (2009) Red lights: the lives of sex workers in postsocialist China. University of Minnesota Press, Minneapolis

Proximity Policing

▶ Community Policing

Psychiatric Autopsy

▶ Psychological Autopsy of Equivocal Deaths

Psychological Autopsy of Equivocal Deaths

Michael R. Davis and James R. P. Ogloff
Centre for Forensic Behavioural Science,
Monash University, Melbourne, VIC, Australia

Synonyms

Equivocal death analysis; Psychiatric autopsy; Reconstructive mental state evaluation; Reconstructive psychological evaluation

Overview

Sometimes it is important for legal purposes to determine a decedent's mental state at the time of death. The psychological autopsy was developed in the late 1950s by mental health professionals associated with the Los Angeles Suicide Prevention Center (SPC) to assist with clarifying the mode of death in equivocal cases (Litman et al. 1970; Shneidman 1981; Shneidman and Farberow 1970). An equivocal death is one in which the *cause* of death may be known – but the *mode* of death is unclear (Shneidman 1981). The mode of death is classified by way of the NASH acronym: natural causes, accident, suicide, or homicide. For example, if an individual is found floating in a river, the cause of death may be asphyxiation due to drowning, but the mode of death may be unknown with three possibilities arising: an accidental drowning, a suicide, or even a homicide. Another example would be someone found dead at the base of a building; the cause of death is multiple crushing injuries, but was this an accidental fall, a suicidal jump, or a case of the deceased being pushed? Shneidman (1981), who coined the term *psychological autopsy*, noted that as many as 20 % of all deaths may initially be deemed equivocal. In such deaths, the mode of death cannot be determined by conventional medical or investigative procedures. What is needed is a retrospective evaluation of the decedent's mental state and the surrounding facts in the case to help determine the mode of death. This entry begins with a historical summary of the psychological autopsy and an examination of its relationship with other forms of analysis, its legal admissibility, and its criticisms of the technique. This will be followed by a discussion of guidelines for conducting psychological autopsies, including appropriate qualifications for practitioners, decision-making frameworks, and the reporting of results. Finally, future directions for the psychological autopsy will be outlined.

Historical Foundations

The psychological autopsy was developed when Drs. Edwin Shneidman, Norman Farberow, and Robert Litman from the SPC were invited by then Los Angeles County Coroner, Dr. Theodore Curphey, to assist in cases of equivocal suicide. The coroner named this group the Death Investigation Team. Shneidman and Farberow (1970) reported that the team

> focused primarily on the *personality* elements associated with suicide, such as suicidal intention, subtle communications relating to suicidal intent, reactive or psychotic depression, and schizophrenia... the approach of the team was to reconstruct the life style and personality of the deceased, mainly by interviewing the spouse, grown children, parents, physician, and others who knew him well. On the basis of these psychological data, and any other physical evidence of suicide, accident, homicide, or natural death that might be found, the Death Investigation Team attempted to make informed extrapolations over the last days of the victim's life and then to report these conclusions and recommendations to the coroner for final integration with his other findings (p. 497).

It should be noted that Dr. Avery Weisman subsequently adapted the psychological autopsy for use in helping to understand the precipitating causes related to unequivocal deaths. To this end, Shneidman (1974) wrote that "Dr. Weisman has creatively extended the reach of the psychological autopsy to deaths that are unequivocal as to mode, pursuing such additional piercing questions as: why did this (clearly physically, accidental, or suicidal) death occur at *this* time? What psychological events preceding the dying period might be related to the death itself? How can we understand this death better?" (p. xi). Weisman (1974) described the application of the "Omega Version" of the psychological autopsy within hospitals, which included information obtained from "actual contact with the patient" (p. 35).

In contemporary times, the psychological autopsy is frequently used as an approach to help understand suicides. In fact, a database search of the term psychological autopsy will invariably result in more references referring to the use of the technique in unequivocal cases than the equivocal cases for which it was originally developed. Accordingly, LaFon (1999) sensibly suggested that the equivocal death psychological autopsy be differentiated from the suicide psychological autopsy. The focus of this entry will be on the equivocal death psychological autopsy, although invariably some literature and research pertaining to the suicide psychological autopsy are still relevant to the investigation of equivocal deaths as the two variants of the technique are clearly conceptually and practically related.

It should also be noted that since the introduction of psychological autopsies, a variety of terms have been used to describe the use of the psychological autopsy in equivocal deaths. For example, some authors have used the term *psychiatric autopsy*, although this has usually been used as a synonym for the same technique. Bendheim (1979) attempted to differentiate the psychiatric autopsy from the psychological autopsy based on the differences in training between the two types of mental health professionals. However, Ogloff and Otto (1993) cogently noted that "this distinction appears strained and smacks of guild interests more than significant differences in technique" (p. 608). Accordingly, the broader term, psychological autopsy, seems to have been used by most mental health professionals who research or carry out this type of work.

Equivocal death analysis (EDA) was a term coined by behavioral scientists at the Federal Bureau of Investigation to describe their analyses of equivocal deaths for law enforcement (Ault et al. 1994; Poythress et al. 1993). Canter (1999) noted that an EDA "is carried out by law enforcement officers, usually FBI agents, who only examine the crime scene material and other information directly available to the police inquiry" and suggested that "an EDA is open to many more biases and distortions than a 'full' psychological autopsy" (p. 128). However, recent literature from EDA practitioners suggests that there is perhaps little actual difference between the two terms (Lacks et al. 2008), although assessors may not always conduct the required interviews themselves. Given the variety of terms in use and the various permutations of the technique, Ogloff and Otto (2003) suggested the umbrella term *reconstructive mental state evaluation* to "include techniques or a group of techniques whereby a mental health professional attempts to describe or discern the mental state of a deceased or missing person at some prior point in time" (p. 1186). It should be noted that they stressed this definition and did not suggest that nonmental health professionals were unqualified to conduct such analyses in certain circumstances.

Relationship to Other Forms of Analysis

While seemingly esoteric, psychological autopsies clearly have some parallels with other forms of behavioral and psychological analysis. As the prior mental state of the person is one of the key considerations, clinical assessments for the purposes of an insanity or mental impairment defense are conceptually related. While both techniques involve the retrospective analysis of collateral information, insanity evaluations generally involve knowledge of the behavioral outcome and are concerned with determining the person's mental state and intent at the time of the offense (Ogloff et al. 1993). In comparison, while psychological autopsies also require a retrospective assessment of the individual's mental state and intent, an opinion is also required regarding the person's actions (i.e., to determine the mode of death). Indeed, in the case of the mental state at the time of the offense evaluation, the mode of death is known. Furthermore, the major difference is obviously that for insanity evaluations the person being assessed is available for examination, although his or her mental state may well have changed from what it was at the time of the offense.

Some authors have made a conceptual link between the psychological autopsy and offender profiling, a technique in which the likely characteristics of a person are inferred from their offense behaviors. This link likely reflects the fact that some authors and practitioners have been engaged in both activities (e.g., Ault et al. 1994; Canter 1999). However, these are considerably different techniques. Offender profiling is essentially a form of retro-classification that provides the broad characteristics of an unknown offender. In contrast, the psychological autopsy provides more fine-grained and definitive opinions regarding the intentions of a known individual, right down to classifying their death into one of four nominal categories. Where the two activities do overlap is in regard to their basic rationale. According to Shneidman (1970), psychological autopsies are theoretically based on the principle that dying behaviors are an integral part of the lifestyle of the individual. In this way, a detailed examination of the individual's life will shed light on their death. Similarly, offender profiling is often based on the premise that behavior reflects personality (Douglas et al. 1992), such that aspects of offence behavior will reflect behavior in other aspects of the offender's life. Thus, both activities can be conceptualized as elaborations of routine activity theory (see Canter 2000) but from different directions and with very different levels of focus.

Perhaps the most closely related technique to the psychology autopsy is that of indirect personality assessment. As the name suggests, this involves an assessment of an individual's personality without actually meeting the person. These include "psychiatric profiles" of world leaders used by governments for intelligence purposes, such as the well-known psychiatric analyses of Adolf Hitler conducted in the 1940s (e.g., Langer 1972). A seemingly more common application is in criminal investigations, where evaluations of an individual suspect and their likely reaction to questioning or other investigative approaches are discerned from collateral information (Ault and Hazelwood 1995). This is similar to a psychological autopsy, except for the fact that the person is still alive, although, as with the psychological autopsy, the person conducting the assessment does not have the benefit of directly accessing the subject of the evaluation. The other difference lies again in the focus, which is on likely future behavior rather than likely past behavior (i.e., reason and cause of death).

Legal Admissibility of Psychological Autopsies

Evidence based upon psychological autopsies, or similar reconstructive evaluations, has been admitted by courts in various jurisdictions, although this has primarily been in the civil arena. In the United Kingdom, attempts to admit psychological autopsies in criminal cases have been unsuccessful. For example, in the case of *R v. Gilfoyle*, a man was convicted of murdering his pregnant wife who died from hanging. A psychological autopsy conducted by a professor of psychology suggested that suicide was a likely mode of death, but this was deemed inadmissible (see Canter 2005). However, evidence of a similar nature was admitted in the unpublished Northern Ireland case of Torney (1996, as cited in Gudjonsson and Haward 1998). In this case a police officer was charged with the murders of his wife, son, and daughter. He claimed that his son had gone "berserk" and killed his mother and sister before committing suicide. A psychologist reviewed the case material and testified that the scene was not consistent with this scenario. Rather, they suggested that a carefully planned execution was more likely. This evidence was admitted and the defendant was convicted of all three murders. While this evidence does not appear to have been described as a psychological autopsy, the parallels with this technique are obvious.

In regard to American courts, Ogloff and Otto (1993, 2003) noted that the technique has been chiefly admitted within civil courts, but added that there has been reluctance to permit testimony regarding mode of death. Rather, testimony limited to the decedent's state of mind prior to death is more likely to be allowed. Ogloff and Otto also noted that psychological autopsies have been allowed in cases of insurance and workers compensation, but not in regard to testamentary capacity or intestate succession, which suggests that the courts are uncomfortable with psychological autopsies that opine on the ultimate issue. Ogloff and Otto also found that criminal courts have been less willing to admit such evidence, and on the rare occasions in which they have, the ultimate issue is not directly addressed. Indeed, Ormerod (2001) described the 1996 case of *USA v. Jean* in which a "psychological expert was permitted to testify that none of the indicators of suicide normally associated with a person who commits suicide were present" (p. 22) but was not permitted to state that therefore the deceased was murdered.

Criticisms of Psychological Autopsies: The USS Iowa Controversy

The use of psychological autopsies in equivocal deaths was soundly criticized following the widely publicized and now infamous USS Iowa incident in 1989. This involved an explosion aboard a US naval ship that killed 47 sailors. A Naval Investigative Service (NIS) investigation initially ruled out the possibility of an accidental explosion. The NIS focused their attention on two sailors: one who died in the explosion and a friend who was set to be the beneficiary of the other's life insurance policy. They asked behavioral scientists at the FBI's National Center for the Analysis of Violent Crime to give an opinion, based upon the evidence collected during the investigation, regarding the likelihood of three outcomes:

(a) Mass murder by the surviving sailor
(b) A suicide-homicide attempt by the dead sailor to get revenge on the surviving sailor for getting married
(c) Suicide by the dead sailor due to a variety of personal reasons

The FBI analysts formed the latter opinion. This caused considerable controversy, with some writers critical that the agents had not considered the possibility of an accident (which was later determined to be the cause of the explosion when the case was reopened; see Lacks et al. 2008). This is perhaps an unfair criticism because an accident had been ruled out by the referring agency at the time of the referral. Nevertheless, a "peer review" panel of 14 psychologists and two psychiatrists was somewhat critical overall. The co-chair of the House Armed Services Committee reportedly commented that "the FBI flunked this peer review" (see Poythress et al. 1993, p. 10).

However, a more balanced evaluation of the review panel's conclusions is provided in an empirical study of their reports (Otto et al. 1993). This revealed that the members of the peer review panel were largely in agreement with the FBI agents' inferences regarding the dead sailor's adjustment and psychological functioning, with agreement on 70 % of such information. However, the panel was not in agreement with the agents regarding the dead sailor's culpability and risk for suicide, with the FBI agents providing what were described as more negative and stronger conclusions than those of the panel members. There was also some suggestion that competing hypotheses were ignored. On the basis of this case, some members of the panel recommended that the users of psychological autopsy techniques "should not assert categorical conclusions about the precise mental state or actions of the deceased" (Poythress et al. 1993, p. 12) and should also make appropriate qualifying statements for any opinions rendered. This is a very sound suggestion that should be followed in psychological autopsies and other equivocal death analyses since they force the person or people conducting the review to explicitly consider and weigh evidence both in favor of and against all reasonable alternatives.

Qualifications Necessary for Conducting Psychological Autopsies

Those conducting psychological autopsies would appear to require a high degree of expertise and knowledge from a variety of disciplines and areas. Given the nature of the practice, it is recommended that the assessor be a registered mental health professional (i.e., a psychologist or psychiatrist is most likely), that they have training and experience in the assessment of suicide and risk factors for suicide, and that they have some familiarity with viewing and interpreting death scene materials and photographs. As noted above, a deficit in one of these areas would not necessarily obviate an individual from conducting psychological autopsies, so long as the person benefitted from the expertise of colleagues who have the requisite level of expertise in a complimentary area. Indeed, FBI behavioral analysts and detectives trained in the FBI's methods conduct such analyses without being registered mental health professionals; ideally, however, they have access to psychological or psychiatric experts with whom they can discuss the case and obtain the necessary assistance and support. Regardless, these proposed skills suggest that some degree of specialist multidisciplinary knowledge is required. Similarly, Ritchie and Gelles (2002) suggested that mental health professionals conducting psychological autopsies should receive "basic didactic information" in the areas of crime scene investigation, toxicology, and forensic pathology. This advice would seem particularly necessary for psychologists who need not be medically trained and may not have a background in biology or chemistry.

Lack of Standard Guidelines

Despite the fact that psychological autopsies have been conducted in cases of equivocal death since the late 1950s – and likely long before without being recognized as such – there is a remarkable lack of standard guidelines for the technique. Shneidman (1981), one of the originators of the technique, advocated a flexible procedure but noted that there were 16 areas of information which "might be included" in a psychological autopsy. These were extended and greatly elaborated by Ebert (1987) to 26 categories of applicable information regarding the deceased individual. Ebert's list of information has been deservedly influential, as subsequent guides have essentially modified and slightly extended his initial list. All of these approaches provide useful information for ensuring that data collection is thorough. However, none describe how one should subsequently use the information to make a final conclusion (Otto et al.1993). Ogloff and Otto (1993) argued that without a standard technique, terms such as psychological autopsy "are really describing a goal of an inquiry as much as a particular technique" (p. 611). This is concerning because opinions expressed in psychological autopsies can ultimately form the basis of very important decisions. For example, Jobes et al. (1986) reported that, in an experimental paradigm, the inclusion of brief psychological autopsies had a statistically significant impact on medical examiners' subsequent opinions regarding mode of death.

An attempt at providing an actuarial (i.e., mechanical and algorithmic) approach to one of the decisions involved in psychological autopsies is the *Empirical Criteria for the Determination of Suicide* (ECDS; Jobes et al. 1991). This is a 16-item tool, derived from an earlier 55-item Death Investigation Checklist, which can provide an indication as to whether a death is likely to be a suicide or an accident. The ECDS involves three sections. Section A refers to self-infliction and intention and consists of 13 items (e.g., toxicological evidence indicates self-inflicted death, decedent had suicidal thoughts). Section B refers to self-infliction only and includes two items (i.e., decedent had experienced general instability in immediate family, decedent had recent interpersonal conflict). Section C includes one item relating to intention only (i.e., decedent had history of generally poor physical health). A total self-infliction score is obtained by adding sections A and B, and a total intention score is obtained by adding sections A and C. The case is determined to be a suicide if the score on both totals is three or more. The case is determined to be an accident if at least one of the total scores is two or fewer.

The ECDS was found to correctly classify 100 % of suicides and 83 % of accidents in the cases reviewed, producing an overall level of correspondence with the professional medicolegal judgment of death 92 % of the time. However, and perhaps appropriately, the authors cautioned that "it must be emphasized that the ECDS do not provide absolute and definitive determinations of the manner of death independent of professional medicolegal judgment. Indeed, the ECDS should only be understood (and used) as a tool for investigation which augments professional judgment" (Jobes et al. 1991, p. 254). Similarly, Simon (2002) described the ECDS as "useful as an adjunctive tool to professional clinical judgment" (p. 150). It should also be noted that the ECDS was developed to differentiate between suicide and accident modes of death, so it is not applicable in cases of potential homicide.

A Decision-Making Framework for Psychological Autopsies

As noted above, the various published lists of information required to inform a psychological autopsy vary in the degree and nature of information required, and they do not describe how one should subsequently use the information to make a final conclusion of accident, suicide, homicide, or natural causes. Shneidman (1981) provided a rare example in which a widow revealed to him her late husband's suicide note that she had not shown the police, but such "smoking gun" variables are likely to be very rare in equivocal death cases. Ebert's (1987) pioneering guidelines suggested that practitioners make a chart divided into the four nominal categories (natural causes, accident, suicide, homicide) and record data that supports each mode of death. Analysts were also encouraged to consider the possible reasons why the deceased may have committed suicide and the possible reasons why they may have been murdered. The listing of evidence for and against the four modes of death (or three, as some authors focus only on accident, suicide, and homicide where appropriate) has also been advocated by subsequent authors (e.g., Napier and Baker 2005; Ritchie and Gelles 2002). However, this process should not simply result in an additive procedure wherein the person counts the number of items in each category and chooses the one with the highest frequency count. Rather, Napier and Baker (2005) stated that "the goal is to arrive at a preponderance of factors listed under one of the headings" but added that "from this listing it becomes evident that some factors weight more heavily than others" (p. 624). The final determination is made therefore by the evaluator by some unknown weighting of the various pieces of evidence adduced.

Given the paucity of guidelines for assisting practitioners in making the inferential leap of determining accident, suicide, or homicide in psychological autopsies, Davis (2004, 2010) proposed that a scenario-based hypothesis-testing approach that forms part of a structured decision-making and report-writing framework may be useful. The first section of these guidelines is the *Checklist of Equivocal Death Information*. This brings together more than 30 areas of information included in previous psychological autopsy publications, as well as other potentially useful information, and is loosely divided into six domains: Personal, Health, Family, Relationships, Preceding Months and Days, and Death Scene. The assessor conducting a psychological autopsy is interested in the longitudinal status of information within each of these domains as well as any recent changes. The *Personal* domain covers the deceased's education and employment history, military service, religious beliefs, legal history, victimization, writing, language, reading material, feelings regarding death, familiarity with methods of death, and hobbies and interests. The *Health* domain covers physical health, alcohol and substance use, and mental health (including general psychological history, mood, and personality). The *Family* domain addresses the quality of relationships with family members, any familial history of substance use or mental illness, conflicts within the family, and any history of suicides or recent deaths in the family, along with the deceased's reaction to any such deaths. The *Relationships* domain covers the deceased's history of intimate and non-intimate relationships, as well as the details of any current romantic relationships. The *Preceding Months and Days* domain draws the assessor's attention to psychosocial stressors, any pre-suicidal behavior, recent changes in habits or routine, and any evidence of future life plans. An estimation of the mental status examination prior to death (i.e., orientation, mood, affect, psychotic symptoms, and judgment) is made, and the events of the day prior to death are reconstructed with particular focus on any departures from the deceased's usual routine. The final domain is the *Death Scene*. This involves an examination of the police report, death scene photographs, pathology and other forensic reports, any suicide notes, and the significance of the death scene to the victim. Particular attention is paid to any evidence of other people at the death scene, any signs of a struggle, the possibility of a staged crime scene, or evidence of sexual activity.

There is some unavoidable overlap between these various domains. However, the grouping into six categories facilitates consideration of the deceased's personality, behavior, mental state, and interpersonal environment over time, with the final two domains providing an eventual focus on the preceding days and the death itself. The checklist, like other previously published psychological autopsy checklists, provides an aide-memoire for the assessor to ensure that a comprehensive assessment takes place, thereby reducing errors of omission. However, a thorough assessment will also be sensitive to the idiographic features of the case. For example, if the death potentially involved dangerous autoerotic practices, more information regarding the deceased's history of such behavior would undoubtedly be required. Hazelwood et al. (1981) identified five characteristics of the autoerotic death scene which assessors should consider in such situations: evidence of a physiological mechanism for obtaining sexual arousal that is dependent upon a self-rescue mechanism or the victim's judgment to discontinue its effect, evidence of solo sexual activity, evidence of sexual fantasy aids, evidence of prior dangerous autoerotic practice, and no apparent suicidal intent.

To use this wealth of information to answer the questions required of a psychological autopsy, Davis (2004, 2010) has argued that it is important to adopt a method in which the totality of the data is considered in context. While not a case of equivocal death, Canter (1999) alluded to the use of a hypothesis-testing approach in investigating the cause of a fire. This is not unlike the "scenario"-based approach advocated for the assessment and management of recidivism risk in offenders and sex offenders (Hart et al. 2003), although that approach is used to speculate about the future rather than the past. Nonetheless, such an approach appears relevant to the analysis of equivocal deaths, as it promotes consideration of multiple variables in context, rather than simply adding variables together in an abstract mechanical fashion.

Using the above approach, Davis (2004, 2010) proposed that assessors develop a series of scenarios that may explain the death in each modality (accident, suicide, and homicide). The assessor then lists the evidence for and against each competing scenario. It is important to explicitly list and evaluate evidence that does not support one's preferred scenario, as failing to acknowledge conflicting information may contribute to the bias of the assessor and has been a previous criticism of psychological autopsies (Poythress et al. 1993). In essence, each scenario is conceptualized as a hypothesis that has to be subjected to systematic critical scrutiny and potential falsification.

In evaluating these scenarios, relevant empirical information should be consulted where appropriate. For example, significant changes in routine are an important consideration in suicide cases (Litman et al. 1970). When formulating homicide scenarios, it is important to also consider the deceased's level of risk for murder and to determine how much their lifestyle and activities placed them at risk of being a victim of violence. When formulating accident scenarios, the lethality of the deceased's behavior should be taken into account, such as how likely an accident was to result in death.

At the end of this process, Davis (2004, 2010) proposed that assessors rate the likelihood that each scenario occurred. This can be expressed in probabilistic terms, or perhaps more appropriately, using ordinal ratings of low, moderate, or high. This is important, because as mentioned above, making outright conclusive opinions has previously been questioned in the literature (Poythress et al. 1993). It would be prudent to consider relevant crude base rates where possible (although they will presumably all be quite low) and also to consider which is the most parsimonious explanation out of the competing scenarios. When making their final determination as to likely mode of death, it may also be helpful for the assessor to provide an indication of their confidence in the decision being made. Indeed, Douglas and Ogloff (2003) have found that determinations of future risk made by forensic mental health evaluators when they are confident about their decision are more likely to be accurate than when they are less confident about their determinations of risk for violence. The option of "unclear" should always be available, as the assessor should not feel that a conclusive opinion must be made. It is an inevitable reality of this inexact and complex practice that some cases will simply have to remain equivocal.

Communicating Psychological Autopsy Results

The most detailed and prescriptive report-writing guidelines for psychological autopsies have been provided by the United States military. These guidelines involve numerous subheadings in an attempt to carefully avoid speculation in the report. Part of a proposed peer review process is to determine whether the examiner has followed the prescribed structure when writing their reports (see Ritchie and Gelles 2002). Employing a somewhat different approach, Davis (2004, 2010) suggested that the exact format of the report will likely vary from case to case; however, a useful structure should include the following components: sources of information and caveats, case summary, personal history of the deceased, stressors, evaluation of different scenarios in each modality (the core of the report) and a final opinion of which scenario is most likely, and the assessor's degree of confidence about the opinion.

Future Directions

The psychological autopsy has been in existence for more than 50 years. However, it is still hampered by questions of practically unknown reliability and validity. Indeed, Ogloff and Otto (2003) observed that it is difficult to find another area of psychological expert testimony that has been admitted as evidence with such little empirical foundation. It should be noted that reliability has received some limited research attention in regard to the psychological autopsy of unequivocal suicides. Werlang and Botega (2003) developed a semi-structured interview protocol to assist in conducting such analyses. They found that this resulted in a high level of inter-rater reliability in regard to assessors' final ratings on each scale of precipitators and stressors, motivation, lethality, and intentionality. Similar studies are needed to determine the inter-rater reliability of psychological autopsies in cases of equivocal death, particularly in regard to the various modes of death. Structured guidelines, such as those described above, have the potential to maximize inter-rater reliability over that of more intuitive approaches.

Establishing the validity of psychological autopsies is extremely difficult, which may explain why it has not yet been accomplished or even attempted. By definition, validity is impossible to determine because the focus of the analysis is deceased. The very factors that make a death equivocal and in need of a psychological autopsy also obviate any knowledge of the "true" mode of death by which to compare the outcome of the analysis. However, Ogloff and Otto (1993) suggested almost two decades ago that researchers could identify cases in which the correct answer is known and then remove the most obvious information, such as a suicide note, prior to presenting the case. Analysts using the psychological autopsy could then be compared to the known "true" outcome to provide some measure of predictive validity. However, such studies have not yet been conducted. Accordingly, the psychological autopsy currently remains a promising, yet unavoidably subjective, form of forensic analysis in cases of equivocal death.

Related Entries

- ▶ Behavioral Investigative Advice
- ▶ Criminal Investigative Analysis
- ▶ Investigative Psychology

Recommended Reading and References

Ault RL, Hazelwood RR (1995) Indirect personality assessment. In: Hazelwood RR, Burgess AW (eds) Practical aspects of rape investigation, 2nd edn. CRC Press, Boca Raton, pp 205–218

Ault RL, Hazelwood RR, Reboussin R (1994) Epistemological status of equivocal death analysis. Am Psychol 49:72–73

Bendheim OL (1979) The psychiatric autopsy: its legal application. Bull Am Acad Psychiatry Law 7:400–410

Canter DV (1999) Equivocal death. In: Canter DV, Alison LJ (eds) Offender profiling series vol. 2. Profiling in policy and practice. Ashgate, Aldershot, pp 123–156

Canter DV (2000) Offender profiling and criminal differentiation. Leg Criminol Psychol 5:23–46

Canter DV (2005) Suicide or murder? Implicit narratives in the Eddie Gilfoyle case. In: Alison L (ed) The forensic psychologist's casebook: psychological profiling and criminal investigation. Willan Publishing, Devon, pp 315–333

Davis MR (2004) Equivocal death analysis: professional guidelines. School of Psychology, Psychiatry, and Psychological Medicine, Monash University. Unpublished manuscript

Davis M R (2010 January) The psychological autopsy: structured guidelines for the analysis of equivocal deaths. Keynote address presented at the 9th international investigative psychology conference, London, UK

Douglas JE, Burgess AW, Burgess AG, Ressler RK (1992) Crime classification manual: a standard system for investigating and classifying violent crime. Simon and Schuster, New York

Douglas KS, Ogloff JRP (2003) The impact of confidence on the accuracy of structured professional and actuarial violence risk judgements in a sample of forensic psychiatric patients. Law Hum Behav 27:573–587

Ebert BW (1987) Guide to conducting a psychological autopsy. Prof Psychol Res Pract 18:52–56

Gudjonsson GH, Haward LRC (eds) (1998) Forensic psychology: a guide to practice. Routledge, London

Hart SD, Kropp PR, Laws DR (2003) The risk for sexual violence protocol (RSVP): structured professional guidelines for assessing risk of sexual violence. Pacific Psychological Assessment Corporation, Victoria

Hazelwood RR, Burgess AW, Groth AN (1981) Death during dangerous autoerotic practice. Soc Sci Med 15E:129–133

Jobes DA, Berman AL, Josselson AR (1986) The impact of psychological autopsies on medical examiners' determination of manner of death. J Forensic Sci 31:177–189

Jobes DA, Casey JO, Berman AL, Wright DG (1991) Empirical criteria for the determination of suicide manner of death. J Forensic Sci 36:244–256

Lacks RD, Westveer AE, Dibble A (2008) Equivocal death investigation: case study analyses. Vict Offenders 3:150–164

LaFon DS (1999) Psychological autopsies for equivocal deaths. Int J Emerg Mental Health 1:183–188

Langer WC (1972) The mind of Adolf Hitler: the secret wartime report. Basic Books, New York

Litman RE, Curphey T, Shneidman ES, Farberow NL, Tabachnick N (1970) The psychological autopsy of equivocal deaths. In: Shneidman ES, Farberow NL, Litman RE (eds) The psychology of suicide. Science House, New York, pp 485–496

Napier MR, Baker KP (2005) Criminal personality profiling. In: James SH, Nordby JJ (eds) Forensic science: an introduction to scientific and investigative techniques, 2nd edn. CRC Press, Boca Raton, pp 615–634

Ogloff JRP, Otto RK (1993) Psychological autopsy: clinical and legal perspectives. Saint Louis Univ Law J 37:607–646

Ogloff JRP, Otto RK (2003) Psychological autopsy and other reconstructive mental state evaluations: clinical and legal perspectives. In: Schultz IZ, Brady DO (eds) Psychological injuries at trial. American Bar Association Press, Chicago, pp 1186–1230

Ogloff JRP, Roberts CF, Roesch R (1993) The insanity defence: legal standards and clinical assessment. Appl Prev Psychol 2:163–178

Ormerod D (2001) Psychological autopsies: legal applications and admissibility. Int J Evid Proof 5:1–31

Otto RK, Poythress N, Starr L, Darkes J (1993) An empirical study of the reports of APA's peer review panel in the congressional review of the USS Iowa incident. J Personal Assess 61:425–442

Poythress N, Otto RK, Darkes J, Starr L (1993) APA's expert panel in the congressional review of the USS Iowa incident. Am Psychol 48:8–15

Ritchie EC, Gelles MG (2002) Psychological autopsies: the current department of defense effort to standardize training and quality assurance. J Forensic Sci 47:1–3

Shneidman ES (1970) Suggestions for revision of the death certificate. In: Shneidman ES, Farberow NL, Litman RE (eds) The psychology of suicide. Science House, New York, pp 551–552

Shneidman ES (1974) Foreword. In: Weisman AD (ed) The realization of death: a guide for the psychological autopsy. Jason Aronson, New York, pp ix–xi

Shneidman ES (1981) The psychological autopsy. Suicide Life Threat Behav 11:325–340

Shneidman ES, Farberow NL (1970) Sample psychological autopsies. In: Shneidman ES, Farberow NL, Litman RE (eds) The psychology of suicide. Science House, New York, pp 497–510

Simon RI (2002) Murder, suicide, accident, or natural death? Assessment of suicide risk factors at the time of death. In: Simon RI, Shuman DW (eds) Retrospective assessment of mental states in litigation: predicting the past. American Psychiatric Publishing, Washington, DC, pp 135–153

Weisman AD (1974) The realization of death: a guide for the psychological autopsy. Jason Aronson, New York

Werlang BSG, Botega NJ (2003) A semi-structured interview for psychological autopsy: an inter-rater reliability study. Suicide Life-Threat Behav 33:326–330

Psychological Profiling

► Behavioral Investigative Advice
► Criminal Investigative Analysis

Psychology of False Confessions

Jennifer T. Perillo
John Jay College of Criminal Justice and
The Graduate Center, The City University of
New York, New York, NY, USA

Overview

A false confession is an admission by an innocent person to a crime he or she did not commit, often accompanied by a narrative indicating the particulars of how the crime was committed and why. As first noted by Hugo Munsterberg in 1908, innocent people may confess for a variety of reasons, both dispositional and situational. Munsterberg exhibited an understanding of false confessions ahead of his time and discussed different types of false confessions that are prompted in different ways. In the decades since Munsterberg's historical work, research has revealed much about the various motivations, tactics, and issues that may result in false confessions, as well as the consequences of those confessions.

False confessions create serious problems for the criminal justice system, especially when the consequences for wrongful convictions are considered. Confession evidence is enormously powerful in court and has a strong impact on judges and juries. Even when contradicted by other evidence and lacking corroboration, confessions produce high rates of conviction. This is partially because confessions can corrupt other evidence, making exculpatory evidence seem artificially weaker while at the same time making inculpatory evidence seem stronger.

To safeguard the criminal justice system, researchers have proposed a number of reforms to the practices of interrogation – including a ban on tactics that put innocent people at risk and the requirement that all interrogations be recorded in their entirety. Additionally, questions have risen concerning the admissibility of expert testimony to educate jurors on why innocent people would confess and the possible impact of those confessions on other evidence. If the rate of false confessions can be curtailed and if lay jurors can become better informed about the limits of confession evidence, the conviction of innocent people may be significantly reduced.

Fundamentals of False Confessions

Prevalence of False Confessions

Although false confessions pose a problem for the criminal justice system, there is no way to derive or estimate the frequency of false confessions. Often guilt and innocence cannot be known, and many confession cases are never adequately scrutinized. Examples throughout history, however, suggest false confessions are not a rare occurrence. Dating back to the Salem witch trials of 1692, numerous false confessions have been revealed through various means.

In the past, false confessions could only be identified only when the confessed crime was proved to be impossible (e.g., the alleged victim was still alive) or the real perpetrator was apprehended. The advent of DNA testing in recent years has expanded the identification of the wrongfully convicted, predominantly in cases involving rape and murder. In aggregated databases, it appears that 15–25 % of those wrongfully convicted had confessed to their crimes.

Confession experts note that these cases are the tip of the iceberg. Indeed, the databases of wrongful convictions do not include those whose false confessions were discovered before trial, those that were not publicized, those that were not contested by the confessor, or those that resulted in a plea bargain. In addition, there remain numerous inmates protesting their innocence who are barred access to DNA testing, some of whom may be innocent. Finally, these databases concentrate on US populations – despite documentation of false confessions all over the world – and within the criminal justice system, thus excluding false confessions that occurred within the private sector (e.g., loss prevention cases) and in military intelligence settings.

Types of False Confessions

Although it is not possible to determine the overall prevalence of false confessions, it is possible to distinguish among different types of false confessions. Researchers recognized that false confessions occur in different ways and for different reasons and created a taxonomy that distinguishes between three types of false confessions: voluntary, compliant, and internalized (Kassin and Wrightsman 1985).

Voluntary False Confessions

Voluntary false confessions occur when individuals willingly and purposefully claim responsibility for crimes they did not commit. These admissions are made in the absence of police pressure or encouragement. People may volunteer false confessions for any number of reasons: to garner attention, fame, or recognition; to satisfy a need for self-punishment in the face of perceived transgressions and feelings of guilt; the inability to distinguish fact from fantasy, which is a common feature of certain psychological disorders; or to protect someone else. Especially a problem with juveniles, voluntary false confessions often occur to protect a parent, child, or other close relative or friend.

Compliant False Confessions

Compliant false confessions occur in response to police pressure during interrogation. In these cases, a suspect is induced to confess in order to escape the stress of interrogation, to avoid a threatened or implied punishment, or to gain a promised or implied reward. The social pressure acting in the interrogation room serves to convince innocent individuals that the short-term benefit of confession (i.e., escape from the interrogation room) outweighs the long-term costs (i.e., prosecution and possible conviction). As a result of isolation, stress, fatigue, or deprivation of needs, innocent individuals may acquiesce to demands for confession, despite knowing they are innocent, as an act of public compliance.

Internalized False Confessions

Sometimes, innocent suspects actually come to believe they have committed a crime for which they are innocent. Typically internalized false confessions involve suspects whose memories are rendered vulnerable to manipulation and who are presented during interrogation with highly suggestive and misleading "evidence" of their guilt. In these cases, the suspect becomes confused, comes to believe that he or she may have committed the crime, and even, at times, confabulates detailed false accounts of what happened. The innocent suspects' internalized beliefs in their guilt can actually lead to the formation of false memories of what they allegedly did, how they did it, and why.

Risk Factors for False Confession

The risk of false confessions is associated with three sets of factors: (1) dispositional characteristics that make some suspects more vulnerable than others, (2) situational factors associated with the stresses of custody and interrogation, and (3) the phenomenology of innocence, whereby innocents waive their rights.

Dispositional Risk Factors

Dispositional risk factors refer to personal characteristics of suspects that increase their risk of a false confession relative to the general population. Not everyone is equally vulnerable to influence during an interrogation. Instead, factors such as personality, cognitive impairments, psychopathology, and youth can place some individuals at a disproportionate risk of falsely confessing.

Compliance and suggestibility appear to be two important personality risk factors (Gudjonsson 2003). Compliance refers to a desire to please others and avoid confrontation and conflict. Compliant individuals are particularly susceptible to those in a position of authority, with whom confrontation would be much more stressful, putting them at risk in an adversarial interrogation. In contrast, suggestibility concerns the strength and stability of one's memory. Overall, individuals who score high in suggestibility have poorer overall memories, lower self-esteem, high anxiety, and low levels of assertiveness.

The Gudjonsson Suggestibility Scale (GSS) is the typical measure used to assess interrogative suggestibility (Gudjonsson 1984). This measure involves presenting a narrative to a subject who must recall the story immediately and then after a delay. The subject then answers questions about the contents of the paragraph, and some of the questions include misleading information. The test administrator provides negative feedback indicating several answers are incorrect, and then the subject is retested. The scale provides subscale measurements of shift, the tendency to change one's memory over time, as well as yield, the tendency to cede to misleading questions and incorporate misinformation into memory. In studies of crime suspects, those who had confessed and later retracted their confessions scored higher in suggestibility than the general population, whereas resistors, those who maintained their innocence throughout interrogation, obtained lower scores. Sleep deprivation and alcohol withdrawal can also increase an individuals' suggestibility as tested by the GSS.

Individuals suffering from mental retardation or other intellectual deficits are also at risk in an interrogation setting (Gudjonsson 2003). Mental retardation is associated with heightened levels of both compliance and suggestibility. Individuals with mental impairments are also less likely to fully comprehend their *Miranda* rights, more likely to acquiesce to authority figures, and more susceptible to leading questions, misinformation, and other forms of suggestive questioning. Research shows that individuals with mental retardation rely more on authority figures for answers to problems, accept blame for negative outcomes, and have memory deficits.

Research has also identified mental illness as a risk factor (Gudjonsson 2003). Many psychological disorders are associated with distorted perceptions of reality, including difficulty with reality monitoring or separating real life from fantasy. Other common symptoms of psychological disorders include impaired judgment, debilitating anxiety, mood disturbance, and lack of self-control. Each of these factors alone can increase the likelihood of false confession and, in combination, may create a strong risk factor for offering misleading information, including false confessions, during interrogation. Individuals suffering from high anxiety may also be more likely to false confess simply to escape the stress of the interrogation room.

Youth is a particularly strong dispositional risk factor (for a review, see Drizin and Colgan, 2004). Overrepresented in the cases of known false confessors, juveniles aged 18 and younger comprise over 30 % of the sample. Juveniles tend to be more compliant toward authority figures, and they are more likely to trust what those authority figures tell them. Juveniles are more suggestible than adult witnesses, so they are more likely to confabulate details for false accounts under suggestive questioning. Juveniles also have less ability to consider long-term consequences of their actions. This means that juveniles are more likely to confess to obtain the short-term goal of leaving the interrogation room without considering the long-term consequences of such an action. Juveniles also show poorer comprehension of their *Miranda* rights than adults, so they are less likely to invoke their rights. Research also shows that juveniles are overall more likely to confess than adults.

Situational Risk Factors

Situational risk factors refer to aspects of the custodial setting and police interrogation tactics that can lead innocent people to confess. The Reid technique, the leading method for interrogating suspects used by interrogators across the United States and widely used across the world, uses psychological tactics to produce confessions from suspects (Inbau et al. 2013). The basic approach of the Reid technique is to increase the anxiety associated with continued denial of an offense and decrease the anxiety associated with confession to make it easier for a perpetrator to admit involvement in a crime. To that end, the Reid technique uses many tactics to increase the anxiety of the interrogation setting, some of which may increase the vulnerability of innocent suspects. In particular, researchers have focused on three basic aspects of the interrogation process: isolation, confrontation, and minimization.

Interrogators are trained to question suspects in the police station, away from any support system or familiar surroundings. The Reid technique offers specific instructions on the layout of interrogation rooms, recommending that they be sparse and free of any visual distractions to increase feelings of isolation. Basic social psychology research has shown that isolation can have profound effects on people, who are social beings, particularly in times of stress. Isolation reduces feelings of self-control, autonomy, certainty about the future, anxiety, and, in turn, incentive to confess in order to escape the aversive situation.

The anxiety stemming from isolation builds up over time, which highlights the issue of interrogation time as a risk factor. Observational and survey research indicate that the typical interrogation lasts under 2 h. In contrast, interrogations resulting in known false confessions last an average of more than 16 h, with over one-third of false confessions lasting between 12 and 24 h (Drizin and Leo 2004). It stands to reason that these lengthy interrogations are associated with increased anxiety, fatigue, and sleep deprivation, which can impair decision-making and increase the tendency to confess as matter of expediency.

In the second stage of interrogation, interrogators confront suspects with strong assertions of guilt. The goal is to convince suspects that resistance is futile and that confession is their best option. Interrogators are specially trained to overcome suspects' denials, refute their alibis, point to inconsistencies in their stories, and convince them that proof of the suspect's guilt is certain. These claims are often bolstered by the presentation of false evidence. One common method of presenting false evidence is through use of the polygraph. After a suspect takes a polygraph, the interrogator informs the suspect that he or she has failed, irrespective of the actual outcome of the test. This tactic is designed to convince suspects that their guilt is obvious and conviction is certain, so confession is the only reasonable option. The presentation of false evidence serves to make innocent suspects feel trapped by the evidence. In this situation, innocent suspects sometimes feel they have no choice but to confess in order to obtain leniency for a crime for which they do not believe they can prove their innocence. There have been numerous exoneration cases involving the presentation of false evidence, and false evidence has been shown to consistently increase the rate of false confession in laboratory experiments as well.

The third stage of the interrogation process is to minimize the crime. Once suspects feel trapped by the weight of alleged evidence, interrogators provide suspects with moral justifications and face-saving excuses for the crime. Minimization increases the tendency to confess by minimizing suspects' perceptions of the consequences of confession. Interrogators are thus trained to develop elaborate "themes" for why a suspect committed a crime – themes that minimize their moral culpability. Interrogators may suggest that the crime was accidental, for example, provoked by the victim, unplanned, well intended, or the result of a drug or alcohol problem or peer pressure. Although minimization statements do not offer explicit promises of leniency, research shows that they lead people to infer that leniency will be forthcoming following confession.

Innocence as a Risk Factor

By observing that innocent suspects behaved differently from guilty suspects in the interrogation room, Kassin (2005) proposed the provocative theory that innocence itself could be a risk factor for confession. Innocent suspects tend to believe the truth will prevail and their innocence will be obvious to interrogators. As a result, innocent suspects put themselves at risk by waiving their *Miranda* rights to silence and to counsel, agreeing to speak with the police, and cooperating fully (Kassin and Norwick 2004). Believing they have nothing to fear or hide, innocent suspects fully answer questions about a crime, even when the answers include information that could be seen as incriminating. Research suggests that innocent suspects do not use self-presentation "strategies" about their narrative content when talking to police, so their stories do not vary according to the questioning process (e.g., Hartwig et al. 2005). Of course, innocent suspects are often initially unaware that they are suspects, instead believing that they are aiding in the investigation.

Being forthright can have unintended consequences for the innocent suspect who is interrogated. Research shows that police presume guilt when questioning suspects and that this guilt bias can lead them to engage in more aggressive interrogations with innocent suspects who vigorously deny involvement (Kassin et al. 2003). These aggressive interrogations may include more problematic techniques that increase risk of false confession. When interrogators use the false-evidence ploy, for example, innocent suspects may feel trapped and confess as a means of minimizing consequences. Even when interrogators do not use the false-evidence ploy, the practice of increasing the anxiety associated with denial during interrogation can lead innocent, as well as guilty, suspects to search for a way out of interrogation. When police officers bluff about evidence – assert evidence has been found but not yet tested – innocent suspects may still falsely confess to escape the stressful situation, believing the evidence would later exonerate them (Perillo and Kassin 2011).

Empirical Research on False Confessions

False confessions have been studied using a variety of methods: individual and aggregated case studies, naturalistic observations of police interrogations, self-report data from police officers and prison inmates, and laboratory and field experiments.

Two main laboratory paradigms exist in the false confession field. In the first, participants are accused of hitting a forbidden key during a computer-typing task and crashing a computer (Kassin and Kiechel 2004; Kassin and Kiechel 1996). In the original study of this paradigm, researchers also manipulated participant vulnerability by varying the speed in which the participants typed the letters. Some participants were also presented with false evidence in the guise of a confederate who claimed to have seen the participants hit the key they were instructed to avoid. The false-evidence ploy increased rates of false confession, internalization, and confabulation, especially among those when rendered vulnerable with the fast typing speed. This computer-crash paradigm has been used all over the world to replicate and extend these original findings.

A second paradigm tested false confessions by varying guilt and innocence within a cheating scenario (Russano et al. 2005). Participants were paired with a confederate to solve a series of problems, which varied between solo and joint sessions. In the guilty condition, the confederate requested help from the participant during a solo session, which was a violation of the experimental rule. In the innocent condition, the confederate worked silently and did not request help. All participants were later accused of cheating by the experimenter. In a test of the impact of minimization, the experimenter then promised leniency, made minimizing remarks, used both tactics, or used no tactics when asking for a confession. Results showed that minimization substantially increased the rate of false confession, similar to the increase caused by an explicit promise of leniency. This paradigm has also been replicated and extended a number of times and offers the advantage of eliciting from participants true and false confessions to willful actions of consequence.

Current Issues and Controversies

The Identification of Suspects

The Reid technique operates through a two-part structure. Interrogators are trained to first conduct a pre-interrogation interview, during which investigators identify who should be a suspect. After identification of suspects, the second stage begins: the interrogation.

The pre-interrogation interview is designed to be a nonconfrontational information-gathering interaction during which the interrogator establishes rapport, develops an understanding of a suspect's baseline of behavior, and attempts to determine whether the suspect is being deceptive. Using the Behavior Analysis Interview, or BAI, interrogators are trained to ask behavior-provoking questions and then observe the suspect's responses. The responses to the provocative questions are believed to offer indications of innocence or guilt. Interrogators are taught to focus on nonverbal cues (e.g., body language, fidgeting, eye gaze, posture) as well as verbal cues (e.g., qualified denials, pauses) to decide who is lying or telling the truth. If the interrogator is convinced of truthfulness, the individual is dismissed and sent home. If the interrogator is convinced of deception, the suspect is interrogated.

The Reid technique, both in the manual and in actual training sessions, claims that interrogators can be trained to make these judgments of guilt and innocence at a high level of accuracy, yet the research in the field of deception detection is highly contradictory to these claims. Research has consistently shown that laypersons can only detect deception at around chance levels (i.e., 54 % accuracy) and, furthermore, that training does not appreciably improve performance (Bond and DePaulo 2006). Research studies specifically using the BAI have shown similarly dismal rates of accuracy in deception detection and identification of guilty suspects. There is no evidence that the behavior-provoking questions recommended by the Reid technique actually have any diagnostic value. Furthermore, there is no evidence that suggests that the verbal and nonverbal cues upon which interrogators are trained to rely have any diagnostic value. Research on deception detection has consistently shown that there are few valid cues to deception and that those that do exist are often too weak to be used for discriminatory purposes in an interrogation (Hartwig and Bond 2011). Instead of improving performance, training simply increases and often inflates confidence in one's decision.

The pre-interrogation interview sets the stage for interrogations as a guilt-presumptive process. Investigators enter into interrogations with a strong belief in the guilt of their suspects, so the goal of each interrogation is to extract a confession. This guilt-presumptive belief impacts how investigators conduct and perceive their interrogations. Research in basic psychology has long shown that once people form a belief, they seek out information that confirms that belief and ignore or misinterpret evidence that contradicts it. Known as confirmation bias, this tendency means it is difficult for a suspect to convince an interrogator of his or her innocence once a suspicion of guilt has formed. The presumption of guilt can also create a self-fulfilling prophecy, whereby the interrogator's aggressive behavior invokes anxiety in the suspect, which in turn confirms the interrogator's belief in guilt. Research has shown that interviewers who believe a suspect is guilty ask more guilt-presumptive questions and place more pressure on suspects to confess, which increases risk for false confession.

The Consequences of Confession

The criminal justice system is designed to provide safety nets to prevent false confessions from resulting in prosecution and wrongful convictions. For these safety nets to work, attorneys, judges, and juries must be able to distinguish between true and false confessions. The narrative content of confessions, however, as well as their corrupting influence on other evidence makes it difficult for people inside and outside the system to make this distinction.

The Impact on Juries

Confession evidence is considered something of a gold standard for evidence. It is difficult for laypersons to accept the idea that individuals may falsely confess to crimes they did not commit; therefore, juries tend to accept confessions at face value. Research has shown that juries are capable of distinguishing between coercive and noncoercive interrogations, but that distinction does not impact their use of the confession in their verdicts. Juries are just as likely to use highly suspect confessions derived from coercive interrogations to determine verdicts as they are to use one stemming from a noncoercive interview.

Another reason why juries may have a hard time dismissing false confessions is their often-elaborative content. Examination of the narrative content of false confessions has shown that false confessions tend to be highly detailed, often containing accurate nonpublic information about the crime (Garrett 2010). False confessions also tend to include descriptions of motives and feelings during and after the crime (Appleby et al. 2011). Combined with the tendency to rely on the (erroneous) common sense that innocent people do not falsely confess, these crime details may serve to convince jurors that these confessions are true and accurate statements from actual perpetrators.

Corrupting Influence of Confessions

Confessions can also taint the interpretation of other evidence. Once it is known that a suspect confessed, trained professionals as well as laypeople may view other evidence as consistent and indicative of guilt – especially when that evidence is ambiguous. This effect has been observed in judgments made about polygraph charts and eyewitness identifications (Elaad et al. 1994; Hasel and Kassin 2009). Research has also shown that context effects have similarly been observed in judgments made from fingerprints and certain complex DNA evidence (Dror and Charlton 2006; Dror and Hampikian 2011). These findings are bolstered by a recent analysis of DNA exonerations from the Innocence Project, where forensic science errors are found in nearly two-thirds of false confession cases (Kassin et al. 2012).

Reforms

Research has shown that confessions may sometimes improperly influence the investigation, prosecution, and review of cases. To that end, interrogation reform is needed to follow the best practices that have been identified in the literature. Research must also continue to address other potential reforms at other stages in the criminal justice process.

The PEACE Model

Reforms to interrogation practices have been enacted in Great Britain. Beginning with the Police and Criminal Evidence (PACE) Act of

1984, interrogation practices have shifted from a focus on psychologically coercive tactics designed to elicit confessions to a more investigative, fact-finding approach. In 1993, the Royal Commission on Criminal Justice created the PEACE model, which instituted training for police officers and further reduced reliance on psychologically coercive techniques. PEACE refers to the five components of the new interviewing technique: "preparation and planning," "engage and explain," "account," "closure," and "evaluate." Studies show that certain Reid techniques – including minimization and the false-evidence ploy – are now rarely used in Great Britain, yet the rate of confessions has not dropped (for a review, see Kassin et al. 2010). Given the current research findings, it appears investigative interviewing may present a useful alternative to the confrontational methods of the typical American interrogation.

Expert Testimony

Getting false confession experts admitted into court has proven difficult in many jurisdictions. In general, expert testimony on false confessions has been excluded for two reasons: (1) questions about the science behind the study of false confessions and (2) the idea that false confession research is known to jurors as a matter of common sense, rendering expert testimony unnecessary. Research has shown that jurors may be able to distinguish between coercive and noncoercive tactics, but they still do not believe that coercive tactics will produce false confessions. Recent research combined with anecdotal evidence suggests that jurors believe that the ends justify the means, in that it is acceptable to use highly coercive tactics to obtain confessions, strengthening presumptions that highly coerced confessions may be nonetheless accepted as true (Blandón-Gitlin et al. 2011). This recent research has also suggested that expert testimony can successfully educate jurors as to how these coercive tactics can produce false confessions, but much more research on the potential benefits of expert testimony is needed.

Videotaping of Police Interrogations

The most widespread proposed reform, which has recently been adopted in many states and jurisdictions, concerns the videotaping of interrogations in their entirety. Videotaping of interrogation is actually the leading recommendation of the official American-Psychology Law Society White Paper on false confessions (Kassin et al. 2010). The strengths of videotaping benefit both the police and the suspect. With a complete recording of an interrogation, the truth about the use of egregious interrogation tactics is known. This record may serve the suspect by reducing interrogators' reliance on those tactics, as well as providing an accurate record about the use of those tactics for court. This practice may also serve interrogators, however, because that same record could be used to refute illegitimate claims of police intimidation or coercion.

It is a well-known finding in psychology that memory is not perfect. Memories are subject to forgetting and distortions, which impair the accuracy of a suspect's or interrogator's memory of an interrogation. With a videotape of the interrogation, there is no question of who said what and when. It would be possible, for example, to review a tape and see where evidence "only the perpetrator could know" actually came from – the suspect or the interrogator. Anecdotal evidence has shown that police interrogators sometimes make source monitoring errors and remember suspects as first admitting details that they had in fact confronted them with earlier. Review of videotaped interrogations could clarify cases when incriminating accounts by a suspect were based on information first offered by the interrogator and, just as importantly, affirm investigators' accounts when the suspect indeed offered unique information "only the perpetrator could know."

Conclusions

False confessions weaken the integrity of our criminal justice system. Many recent examples have garnered strong media attention and interest, including the Amanda Knox case. Research has identified a number of risk factors that place innocent suspects at risk for falsely confessing, including dispositional and interrogative risk factors as well as innocence itself. Research has also been undertaken to understand the consequences of confession in hopes that the impact of false confessions may one day be mitigated, if not prevented.

Still, more research is needed in the realm of false confessions. Investigative interviewing presents a first step for interrogative reform, but there remains a need to identify interrogation tactics that have genuine diagnostic value – the ability to elicit true confessions without also eliciting false confessions. Research also must further investigate the consequences of confessions – how confessions impact the criminal justice process and how that impact can be minimized – and the extent that expert testimony can serve as a remedy to the consequences of confession. These avenues of research have been largely untouched and need to be explored.

By offering erroneous evidence that a crime has been "solved," false confessions interfere with our justice system's ability to capture the actual perpetrators of crime. Once police have a confession from the person they believe committed the crime, all avenues of evidence toward other suspects – including, in cases of false confessions, the actual perpetrator – are no longer explored, and any further evidence pointing another direction is often either disregarded or misinterpreted to further implicate the innocent confessor. Current practices must be evaluated and reforms undertaken to secure convictions from the actually guilty while avoiding the wrongful convictions of innocent persons.

Related Entries

- ▶ Detecting Deception with fMRI
- ▶ False Confessions and Police Interrogation

Recommended Reading and References

Appleby SC, Hasel LE, Kassin SM (2011) Police induced confessions: an empirical analysis of their content and impact. Psychol Crime Law. Advance online publication

Blandón-Gitlin I, Sperry K, Leo R (2011) Jurors believe interrogation tactics are not likely to elicit false confessions: will expert testimony inform them otherwise? Psychol Crime Law 17(3):239–260

Bond CF, DePaulo BM (2006) Accuracy of deception judgments. Pers Soc Psychol Rev 10(3):214–234

Drizin DA, Colgan B (2004) Tales from the juvenile confession front: A guide to how standard police interrogation tactics can produce coerced and false confessions from juvenile suspects. In Lassiter GD (ed.) Interrogations, confessions, and entrapment (pp. 127-162). New York: Kluwer Academic/Plenum

Drizin DA, Leo RA (2004) The problem of false confessions in the post-DNA world. North Carolina Law Review 82:891–1004

Dror IE, Charlton D (2006) Why experts make errors. J Forensic Identif 56(4):600–616

Dror IE, Hampikian G (2011) Subjectivity and bias in forensic DNA mixture interpretation. Sci Justice 51:204–208

Elaad E, Ginton A, Ben-Shakhar G (1994) The effects of prior expectations and outcome knowledge on polygraph examiners' decisions. J Behav Decis Mak 7(4):279–292

Garrett BL (2010) The substance of false confessions. Stanford Law Rev 62(4):1051–1119

Gudjonsson GH (1984) A new scale of interrogative suggestibility. Pers Individ Diff 5(3):303–314

Gudjonsson GH (2003) The psychology of interrogations and confessions: a handbook. Wiley, Chichester

Hartwig M, Bond CF Jr (2011) Why do lie-catchers fail? A lens model of human lie judgments. Psychol Bull 137(4):643–659

Hartwig M, Granhag PA, Strömwall LA, Vrij A (2005) Detecting deception via strategic disclosure of evidence. Law Hum Behav 29(4):469–484

Hasel LE, Kassin SM (2009) On the presumption of evidentiary independence: can confessions corrupt eyewitness identifications? Psychol Sci 20(1):122–126

Home Office (1985) Police and Criminal Evidence Act 1984. HMSO, London

Home Office (2003) Police and Criminal Evidence Act 1984 Codes of Practice A-E revised edition. HMSO, London

Inbau FE, Reid JE, Buckley JP, Jane BC (2013) Criminal interrogations and confessions, 5th edn. Jones & Bartlett Learning, Burlington

Kassin SM (2005) On the psychology of confessions: does innocence put innocents at risk? Am Psychol 60(3):215–228

Kassin SM (2012) Why confession trump innocence. Am Psychol 67(6):431–445

Kassin SM, Gudjonsson GH (2004) The psychology of confessions: a review of the literature and issues. Psychol Sci Public Interest 5(2):33–67

Kassin SM, Kiechel KL (1996) The social psychology of false confessions: Compliance, internalization, and confabulation. Psychological Science 7(3):125–128.

Kassin SM, Norwick RJ (2004) Why people waive their Miranda rights: the power of innocence. Law Hum Behav 28(2):211–221

Kassin SM, Wrightsman LS (1985) Confession Evidence. In: Kassin SM, Wrightsman LS (eds) The psychology of evidence and trial procedure, 3rd edn. Sage, Beverly Hills

Kassin SM, Goldstein CC, Savitsky K (2003) Behavioral confirmation in the interrogation room: on the dangers of presuming guilt. Law Hum Behav 27(2):187–203

Kassin SM, Leo RA, Meissner CA, Richman KD, Colwell LH, Leach A-M, La Fon D (2007) Police interviewing and interrogation: a self-report survey of police practices and beliefs. Law Hum Behav 31(4):381–400

Kassin SM, Drizin SA, Grisso T, Gudjonsson GH, Leo RA, Redlich AD (2010) Police-induced confessions: risk factors and recommendations. Law Hum Behav 34(1):3–38

Kassin SM, Bogart D, Kerner J (2012) Confessions that corrupt: evidence from the DNA exoneration files. Psychol Sci 23(1):41–45

Münsterberg H (1908) On the witness stand. Doubleday, Garden City

Perillo JT, Kassin SM (2011) Inside interrogation: the lie, the bluff, and false confessions. Law Hum Behav 35(4):327–337

Russano MB, Meissner CA, Narchet FM, Kassin SM (2005) Investigating true and false confessions within a novel experimental paradigm. Psychol Sci 16(6):481–486

Psychology of Risk

► Fear of Crime and the Psychology of Risk

Psychopathic Personality Disorder

► Psychopathy and Antisocial Personality Disorder

Psychopathy

► Psychopathy and Antisocial Personality Disorder

Psychopathy and Antisocial Personality Disorder

James R. P. Ogloff and Michael R. Davis
Centre for Forensic Behavioural Science,
Monash University, Melbourne, VIC, Australia

Synonyms

Dissocial personality disorder; Psychopathy; Psychopathic personality disorder

Overview

The history of psychiatric nosologies has been scattered with a variety of terms used to describe individuals with serious personality pathology that leads them to engage in ways considered repugnant to the social mores of the time. Manie sans delire, moral insanity, moral imbecility, degenerate constitution, congenital delinquency, constitutional inferiority, sociopathy, antisocial personality disorder, and psychopathy are among the many semantic variations of the main theme. While the diagnostic labels have continued to evolve for more than 200 years, the personality characteristics and behaviors that they denote have remained relatively unchanged. The most commonly ascribed characteristics include a pronounced disorder of affective or moral functions, accompanied by irresponsibility, impulsivity, and a propensity to engage in behavior that brings them into conflict with society – and often the law. However, the disorder is typically absent of any appreciable alteration in cognitive functioning, such as perception, judgment, imagination, or memory.

The term psychopathy has been most consistently used in the psychological and psychiatric research and clinical literature; therefore, that is the term that will be used in this entry. The entry begins with a brief overview of the history of psychopathy and its clinical and behavioral implications before turning to a discussion of the relevance of psychopathy for criminology.

What Is Psychopathy?

First and foremost, psychopathy is a serious personality disorder. Personality disorder is defined as "an enduring pattern of inner experience and behavior that deviates markedly from the expectations of the individual's culture, is pervasive and inflexible, has an onset in adolescence…is stable over time, and leads to distress or impairment" (American Psychiatric Association 2000, p. 629). Personality disorders are manifested by difficulties in at least two of the following four areas: cognitions (capacity to think and reason), affectivity (expressed emotion), interpersonal functioning (ability to engage in meaningful and stable relationships), and impulse control (the ability to adapt and control behavior). Psychopathy is a personality disorder that is characterized by deficits in interpersonal functioning, expressed emotion, and behavior. Characteristics of the disorder in the interpersonal domain include an elevated sense of importance, proneness to pathological lying and manipulation, and superficiality. The deficits associated with expressed emotion include having a flat level of affect, limited capacity for empathy, remorse, and feelings of guilt. Low levels of anxiety are also noted. The behavioral manifestations of the disorder include an elevated need for stimulation, proneness to boredom, impulsivity, irresponsibility, and a failure to take responsibility for their own actions. Such individuals tend to lack conventional social values which repeatedly bring them into contact with authority figures and, oftentimes, the law.

Millon and colleagues (1998) noted that "psychopathy was the first personality disorder to be recognized in psychiatry. The concept has a long historical and clinical tradition, and in the last decade a growing body of research has supported its validity…" (p. 28). Early on, the term psychopathy was used to refer to a range of personality disorders ("psychopathic personalities") that were seen to be extreme forms of normal personality. General descriptions of people who share the characteristics found in the modern conceptualization of "psychopathy" (see Cleckley 1941) have been found in ancient writing. Scholars and practitioners alike have debated the merits of conditions such as those characterized by so-called antisocial features for as long as attempts have been made to classify mental disorder. For example, the term "manie sans délire" (i.e., mania without confusion of the mind), which Pinel coined in the 1700s to describe patients whose affective faculties were disordered, was criticized as early as 1866 for only having use in court (Falret 1866). As noted at the outset, over the years, a number of different labels have been used for the condition; all of the terms that have been used are pejorative, and there is little doubt that each would conjure up very negative images for people. The term with the longest clinical tradition is "psychopathy." As such, it has been the subject of considerable research and scholarly writing. While sharing characteristics with psychopathy, the contemporary condition antisocial personality disorder is much broader and is based more on behavioral features than on some of the traditional personality characteristics associated with psychopathy.

In 1941, the American psychiatrist Hervey Cleckley described the condition of psychopathy in his now classic book the *Mask of Sanity*. He identified 16 characteristics of psychopathy which he drew from the literature and his clinical experience:

1. Superficial charm and good intelligence
2. Absences of delusions and other signs of irrational thinking
3. Absence of "nervousness" or psychoneurotic manifestations
4. Unreliability
5. Untruthfulness and insincerity
6. Lack of remorse or shame
7. Inadequately motivated antisocial behavior
8. Poor judgment and failure to learn from experience
9. Pathological egocentricity and incapacity for love
10. General poverty in major affective reactions
11. Specific loss of insight
12. Unresponsiveness in general interpersonal relations
13. Fantastic and uninviting behavior, with drink and sometimes without

14. Suicide rarely carried out
15. Sex life impersonal, trivial, and poorly integrated
16. Failure to follow any life plan

Initially this disorder was known as psychopathy or psychopathic personality; however, by the introduction of the second edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-II 1968), the condition was labeled "Personality Disorder, Antisocial Type," although the diagnostic criteria still closely resembled those described by Cleckley (1941).The term "antisocial personality disorder" was first introduced in 1980 in the DSM-III (American Psychiatric Association 1980) and has remained with the most recent edition of the DSM, published in 2000 (DSM-IV-TR; APA 2000). Beginning with the introduction of DSM-III in 1980, the DSM criteria began to focus almost exclusively on behavioral features. The current criteria for antisocial personality disorder are as follows:

- Evidence of conduct disorder before age 15
- A pervasive pattern of disregard for, and violation of, the rights of others since the age of 15, as indicated by three or more of the following:
  1. Failure to conform to social norms with respect to lawful behaviors, as indicated by repeatedly performing acts that are grounds for arrest
  2. Deceitfulness, as indicated by repeated lying, use of aliases, or conning others for personal profit or pleasure
  3. Impulsivity or failure to plan ahead
  4. Irritability and aggressiveness, as indicated by repeated physical fights or assaults
  5. Reckless disregard for safety of self or others
  6. Consistent irresponsibility, as indicated by repeated failure to sustain consistent work behavior or honor financial obligations
  7. Lack of remorse, as indicated by being indifferent to or rationalizing having hurt, mistreated, or stolen from another

Ogloff (2006) noted that while the current diagnostic conceptualization of antisocial personality disorder is based largely on behaviors, traditionally the focus of psychopathy was on the interpersonal and affective deficits. Regrettably, the disorder has become a diagnostic category for behavioral difficulties pertaining to criminality. Moreover, as will be discussed below under implications, far more people (particularly prisoners) meet the criteria for a diagnosis of antisocial personality disorder than is warranted.

To complicate matters further, the tenth edition of the International Classification of Diseases (ICD-10; World Health Organization 1992) uses both personality traits and behaviors for the diagnosis of dissocial personality disorder, conceptually similar to psychopathy. The criteria for this disorder are:

1. Callous unconcern for the feelings of others and lack of the capacity for empathy
2. Gross and persistent attitude of irresponsibility and disregard for social norms, rules, and obligations
3. Incapacity to maintain enduring relationships
4. Very low tolerance to frustration and a low threshold for discharge of aggression, including violence
5. Incapacity to experience guilt and to profit from experience, particularly punishment
6. Marked proneness to blame others or to offer plausible rationalizations for the behavior bringing the subject into conflict with society
7. Persistent irritability

As compared to antisocial personality disorder, dissocial personality disorder places more emphasis on traditional psychopathic personality features. In particular, dissocial personality disorder emphasizes deficits of affect or expressed emotion, which have traditionally been seen as among the central personality features of psychopathy.

How will antisocial personality disorder and psychopathy fare in the upcoming revision to the DSM, the DSM-5? The proposed changes to the DSM-5 include substantial shifts in how personality disorders are defined and organized (see www.dsm5.org). Included in these changes are significant modifications to the criteria for antisocial personality disorder. The proposed DSM-5 revisions, referred to as antisocial personality disorder/dissocial personality disorder available at the time of writing, shift the diagnostic focus more toward the traditional interpersonal and emotional aspects of psychopathic personality and rely less heavily on the behavioral, or criminal actions of the individual. While the revision process is yet to be completed, clarification and differentiation between antisocial personality disorder and psychopathy are warranted lest the confusion will continue.

The Operationalization of Psychopathy

The Hare Psychopathy Instruments

Although the criteria for antisocial personality disorder have changed with each edition of the DSM, Dr. Robert Hare developed the Psychopathy Checklist (PCL-R; Hare 1980, 2003) to reliably and validly assess the traditional construct of psychopathy described by Cleckley (see Fig. 1).

The PCL-R is a symptom-construct rating scale comprised of interpersonal, affective, lifestyle, and antisocial features. It consists of 20 items, 18 of which comprise a two-factor, four-facet model (the two additional items do not load on either factor but do contribute to the total score). Each of these items is coded on a three-point ordinal scale (0 = does not apply; 1 = partially applies; 2 = definitely applies), such that total scores on the instrument range from 0 to 40. Scores on this instrument measure the extent to which an individual matches the prototypical "psychopath." The PCL-R is a dimensional measure of psychopathic traits, and while it is true that many people may possess some of the characteristics to some extent, very few possess enough of the traits to be considered "psychopathic" (i.e., with scores above 30 on the instrument). Indeed, approximately 15 % of North American male prisoners, 7.5 % of North American female prisoners, 10 % of male forensic psychiatric patients, fewer than 3 % of psychiatric patients, and an estimated 1 % or less of the general community would have PCL-R scores of this magnitude.

The PCL-R is recognized as the "gold standard" in the assessment of psychopathy. It has been adapted to produce a screening version (Psychopathy Checklist: Screening Version, PCL:SV) and a youth version (Psychopathy Checklist: Youth Version, PCL:YV). The PCL-R has been found to be strongly related to re-offending and violence and is widely used in clinical practice and research in forensic settings. It has proven useful in predicting recidivism,

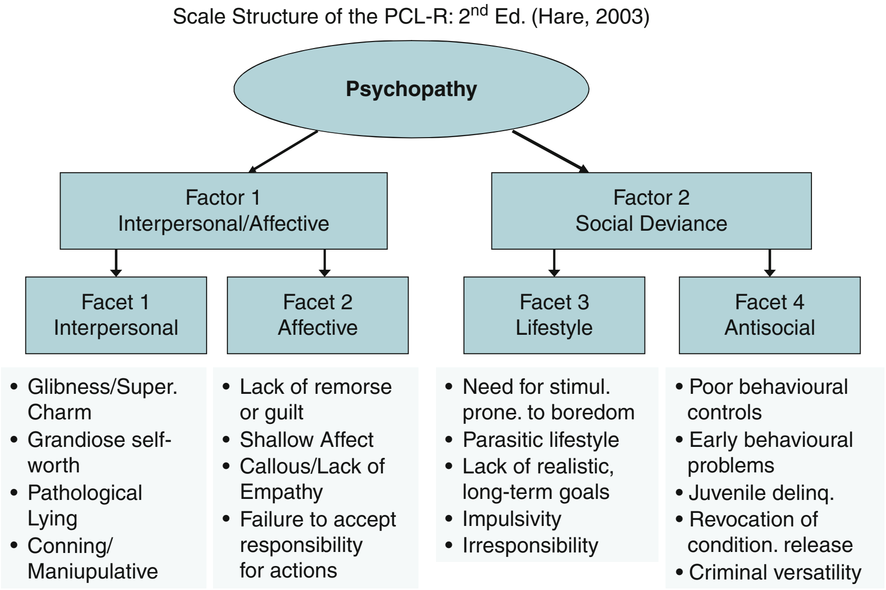
Scale Structure of the PCL-R: 2nd Ed. (Hare, 2003)
Psychopathy
Factor 1
Interpersonal/Affective
Factor 2
Social Deviance
Facet 1
Interpersonal
Facet 2
Affective
Facet 3
Lifestyle
Facet 4
Antisocial
• Glibness/Super. Charm
• Grandiose self-worth
• Pathological Lying
• Conning/ Maniupulative
• Lack of remorse or guilt
• Shallow Affect
• Callous/Lack of Empathy
• Failure to accept responsibility for actions
• Need for stimul. prone. to boredom
• Parasitic lifestyle
• Lack of realistic, long-term goals
• Impulsivity
• Irresponsibility
• Poor behavioural controls
• Early behavioural problems
• Juvenile delinq.
• Revocation of condition. release
• Criminal versatility

Psychopathy and Antisocial Personality Disorder, Fig. 1 The psychopathy checklist-revised (PCL-R) poor treatment outcome, and as a strong predictor of violence. There are now well over 500 research articles and chapters on the PCL measures.

Despite the widespread use of the PCL-R, there has been heated scholarly debate in regard to the inclusion of the fourth facet – antisocial behavior – as a diagnostic feature of psychopathy. Cooke and Michie (2001) argued on the basis of factor analytic results that a three-factor model for psychopathy, essentially the first three facets of the PCL-R, was appropriate. They argued that the antisocial features of psychopathy were behavioral manifestations or consequences of these core features. Nonetheless, they still emphasized "the necessity of continuing to use the full PCL-R for risk assessment and other applied purposes" (p. 186). The suggestion of a three-factor model was subsequently disputed by Hare and colleagues who developed the four-facet model in addition to the previous two-factor model. Nevertheless, the importance of antisocial behaviors as features of psychopathy continues to be vigorously debated in the literature, often in quite caustic exchanges (e.g., Hare and Neumann 2010; Skeem and Cooke 2010a, b). In any event, the PCL-R currently remains the instrument of choice for the assessment of psychopathic personality features, although other instruments are being developed.

Self-Report Measures of Psychopathy

Some self-report measures have been developed to assess similar factors identified in the PCL-R. Although limited in their utility given the characteristics of manipulation and pathological lying associated with psychopathy, the measures have utility for research and some clinical applications that do not require a comprehensive assessment of the individual.

The Levenson Self-Report Psychopathy Scale (LSRP; Levenson et al. 1995) is a brief, 26-item measure that was designed to assess psychopathic features in subclinical samples. The Self-Report Psychopathy-II Scale (SRP-II) is a 60-item scale that was originally designed to be a self-report version of the PCL-R. In comparing the SPR-II to the PCL-R in forensic settings, Hare (1991) found that the total scores of each measure had a correlation of $r = .54$. In recent years, the SRP-II was revised to form the SRP-III in response to criticisms regarding an excess of anxiety-related items and minimal inclusion of antisocial behavior items within the SRP-II.

The Psychopathic Personality Inventory (PPI; Lilienfeld and Andrews 1996) is a 187-item scale that was developed for use in noncriminal populations. The PPI was originally conceptualized as having eight lower-order factors and a total score. However, subsequent factor analyses suggested that the items formed two factors conceptually related to those on the PCL-R (Lilienfeld and Fowler 2006). A more recent version of the PPI was developed to lower the reading level, decrease cultural bias, and reduce the number of items to 154 (Lilienfeld and Windows 2005). The PPI-R is reported to perform similarly to the original measure, but additional research is warranted.

What Causes Psychopathy?

Despite the long-standing interest in psychopathy, research and knowledge pertaining to the etiology of psychopathy is in its infancy. Some evidence suggests that psychopathy has a genetic influence. Although it is difficult to know precisely how much of the psychopathic personality is attributed to genetics, it is likely to be in the order of 30 % of the variance of personality. This shows that while genetics is an important contributing factor to the etiology of psychopathy, it is not the full story.

Discussions regarding the causes of psychopathy are often tied to speculation and theories about variants or subtypes of psychopathy. Researchers have proposed that psychopathy is not a homogeneous construct, but rather a heterogeneous construct with different etiological factors associated with different subtypes of psychopathy. Most often, these subtypes are referred to as primary and secondary psychopathies (Karpman 1948; Lykken 1995). Each of these subtypes has its own hypothesized etiology and associated characteristics.

Traditionally, the primary psychopath has been described as incapable of experiencing emotions, thus not capable of feelings such as guilt, empathy, or remorse. The primary psychopath is cold and calculating and engages in behavior not hampered by anxiety. In contrast, the secondary psychopath can experience a high level of negative emotions, such as anxiety and guilt, and displays behavior that is more aggressive and impulsive.

It has been suggested that the affective deficit underlying primary psychopathy occurs as a result of nature (e.g., genetics), whereas the affective disruption of the secondary psychopath is the result of nurture (e.g., environmental insults) (Karpman 1948). Environmental factors responsible for the later development of secondary psychopathy have been suggested to include unresolved conflicts resulting from harsh punishment and parental rejection (Karpman 1948), poor and ineffective parenting (Lykken 1995), and early childhood abuse or abandonment (Porter 1996). Not only does the speculated etiology of primary versus secondary psychopathy differ, but so does expectations regarding treatment outcome for the two subtypes. Given the presence of anxiety and learned emotional adaptation to environmental abuses, the secondary psychopath is seen as more amenable to treatment than the primary psychopath (see Ogloff and Wood 2010).

Research from Skeem and colleagues (2007) also lends empirical support to distinct subtypes of psychopathy. Skeem and her colleagues (2007) found support for two homogeneous groups that parallel classic primary and secondary psychopathy formulations. They found that violent psychopaths who scored high on a measure of psychopathy were distinguishable based on their level of anxiety. Skeem and colleagues (2007) noted that "secondary (high-anxious) psychopaths were most drastically distinguished from primary (low-anxious) psychopaths by their (a) emotional disturbance (anxiety, major mental and substance abuse disorders, borderline features, impaired functioning), (b) interpersonal hostility (irritability, paranoid features, indirect aggression), and (c) interpersonal submissiveness (lack of assertiveness, withdrawal, avoidant and dependant features)" (p. 405).

A considerable amount of research has been devoted to the study of emotional, attentional, and cognitive processing in psychopathic personality disorder. This research shows that psychopathy is not readily interpreted in terms of damage to specific areas of the brain; but any brain abnormalities that do exist are subtle and likely more functional than structural. For example, while brain anomalies in psychopaths are largely found in the amygdala and orbital-frontal cortex, individuals with physical damage to these areas show more global deficits than do psychopaths. As such, psychopathy is not apparently the product of structural brain damage or deficits.

While *structural* brain differences of psychopaths compared to non-psychopaths are limited, considerable research exists to show that there are some significant *functional* differences in the brain functioning between psychopaths and others. Generally, the differential processing of attention and emotional information by psychopaths has implicated anomalies in the functioning of the amygdala, the orbital-frontal cortex (OFC), and differences in cerebral asymmetry. For example, the amygdala is considered to modulate sensory processing and the processing of emotional stimuli, as well as influencing goal-directed behavior, all of which appear impaired in psychopathy (Blair et al. 2005). Studies have shown impaired learning from punishment when there is potential for reward (gambling tasks), reduced autonomic response to distress in others, impaired naming of fearful and sad expressions (but not other emotions), and reduced physiological responses to anticipation of fear- or pain-inducing stimuli (see Blair et al. 2005; Hiatt and Newman 2006).

Substantial advancement has been made toward understanding the neurobiological aspects of psychopathic personality, but a coherent picture or etiological model that can explain all aspects of these anomalies or how they may give rise to either the personality or behavioral characteristics of psychopathy is yet to be devised. The most widely inclusive theory is posited by Blair and colleagues (2005), who argue that genetic anomalies disrupt amygdala functioning from an early age, leading to impairment in emotional learning as a core dysfunction of psychopathy. This impaired emotional learning would not necessarily lead to the development of the full psychopathy syndrome (such as antisocial and impulsive behaviors), but increases the probability that an individual will learn to employ antisocial means to achieve goals depending on social environment and learning history (i.e., environmental risk factors). Blair and colleagues argue that instrumental aggression may result from impairment to the amygdala functioning (impaired socialization, affect processing, and empathy), while reactive aggression may be the result of OFC impairment (inability to modify behavior to achieve goals when environmental cues change, leading to frustration and then aggression). However, their theory fails to account for additional brain abnormalities in psychopathy, such as cerebral asymmetry. In addition, the heavy focus of research on male adult prison inmates (where psychopathic personality is likely to be easily available and identifiable) makes it difficult to separate deficits associated with the core personality features (Factor 1) and antisocial behavior more generally. There is a need to examine potential neurological and cognitive abnormalities among "successful" psychopaths in the community, as well as in alternative cultural, ethnic, gender, and age groups.

Criminal Correlates of Psychopathy

As noted previously, less than one percent of people in the general population are thought to be psychopaths. However, research shows that between 15 % and 20 % of people in prisons have high scores on the PCL-R (Hare 2003). Literally hundreds of studies, and at least six meta-analyses, have explored the relationship between psychopathy and recidivism; they have uniformly found that psychopaths are at greater risk for offending and violent offending. While the specific rates of offending differ somewhat across the samples, as compared to non-psychopaths, at 1-year follow-up, psychopaths have been found to be approximately three times as likely to offend and four times as likely to violently re-offend than non-psychopathic offenders. Indeed, Hemphill (2007) reported that across three separate meta-analyses, the unweighted correlation between PCL-R scores and violent recidivism was "in the .20–25 range" and admitted that these were conservative values (i.e., the effect sizes were larger when weighted by sample size). Furthermore, Monahan and colleagues (2001) studied a range of risk factors for violence among civil psychiatric patients in the United States and found that the PCL:SV had the highest correlation with subsequent violence of all of the 134 risk factors ($r = .26$).

While not originally developed to be a risk assessment measure, psychopathy has emerged as a robust risk factor for criminal recidivism, and many have advocated its use in any assessment of risk for offending or violence. For example, Hart (1998) stated that "psychopathy is such a robust and important risk factor for violence that failure to consider it may constitute professional negligence" (p. 133). Accordingly, several formal risk assessment schemes require that one include or consider scores on the PCL-R or the PCL:SV.

Psychopathic offenders differ in many ways from their non-psychopathic criminal counterparts. Compared with other offenders, psychopaths are at greater risk for callous, cold-blooded behavior and crimes, institutional infractions, general recidivism, violent recidivism, and sexual recidivism (Hare 2003). Psychopaths also engage in a greater number and variety of criminal behaviors when compared to other repeat offenders, begin committing crimes at an earlier age than other career criminals, and display a different pattern across the lifespan. While criminal behavior tends to decrease over time, Hare and colleagues found that only about half of the psychopaths studied demonstrated a significant decrease in criminal behaviors beginning around ages 35–40 years. However, this reduction largely applied to nonviolent offences and minimal reduction was seen in violent behavior.

Woodworth and Porter (2002) studied 125 homicide offenders and found that the majority of homicides committed by psychopaths were "instrumental" murders (i.e., cold-hearted murders that are conducted for revenge, thrill seeking, or some material gain) rather than "reactive" (i.e., in the heat of the moment without secondary gain). By contrast, the types of murders committed by non-psychopaths were fairly evenly split between instrumental and reactive. Among sexual homicides, psychopathic perpetrators exhibit significantly higher levels of gratuitous and sadistic violence than non-psychopathic sexual homicide perpetrators (Porter et al. 2003). Overall, violence by psychopaths is more likely to be motivated by material gain, revenge, or thrill seeking (i.e., instrumental aggression) rather than momentary heightened emotional arousal (reactive aggression).That said, psychopaths are likely to engage in violence in response to both motivations, whereas non-psychopath violent offenders typically do not engage in instrumentally motivated violence (Blair et al. 2005).

Psychopathy and Criminal Investigation

Given the prevalence of psychopathic personality features in the criminal justice system, particularly in regard to serious violent offenders, it is perhaps unsurprising that psychopathy is an important construct for behavioral scientists who provide consulting services to law enforcement agencies. Such activities are known variously as criminal investigative analysis, behavioral investigative advice, and offender profiling (Depue, → Criminal Investigative Analysis; Rainbow, Gregory, & Alison, → Behavioural Investigative Advice). These can involve making inferences about an unknown offender's likely characteristics based on the behaviors exhibited during their crimes, as well as recommendations for investigative or interview strategy (Ault and Hazelwood 1995). Impulsivity and sensation-seeking behavior, glibness and superficial charm, conning and manipulative offence behavior, and predation and instrumental violence may all suggest the presence of psychopathic personality features in an unknown offender (see O'Toole 2007). While far from a clinical diagnosis, such hypotheses, when viewed in conjunction with other inferences from the offender's behavior, can assist in the development of appropriate investigative and interview strategies.

The White-Collar Psychopath

While most research on psychopathy has been conducted on criminals, it has long been speculated, therefore, that not all psychopaths engage in criminal or violent behavior resulting in imprisonment and that individuals with psychopathic personality can be found in a variety of successful trades and professions in the general community. Hare (1993) argued that psychopaths are "to be found nearly everywhere – in business, the home, the professions, the military, the arts, the entertainment industry, the news media, academe, and the blue-collar world" (p. 115). It has also been argued that certain psychopathic traits (fearlessness, dominance, superficial charm) may prove valuable in some professions such as law, politics, and leadership positions in business and the military (Lykken 1995). However, the limited research evidence has been based on anecdotal clinical descriptions, with almost no quantitative study (e.g., Babiak 2007). Cleckley's seminal book (1941) described several case studies of people who possessed the "core" personality features observed in criminal psychopaths, but manifested those traits in ways that did not result in frequent criminal activity (or at least, arrest). Nonetheless, the case studies that have been discussed have shown that high psychopathy scores in the business world are associated with manipulation of co-workers and bosses, circumventing human resources, finance, and security systems, and fast-tracked promotion to management ranks and successful careers despite limited experience or professional dedication (Babiak 2007).

Descriptions of psychopaths in the community and business have highlighted terms such as "successful," "sub-clinical," and "white-collar" psychopaths, but conceptual difficulties have been raised in the differentiation between such individuals and their criminal counterparts. For example, Hall and Benning (2006) question whether the etiology of psychopathy among noncriminals is the same as for psychopathic criminals (i.e., do community psychopaths differ in degree or kind from criminal psychopaths?). Should "successful" psychopathy be viewed as dimensionally sub-clinical, as a moderated manifestation of the same level of traits, or as a separate clinical entity with different etiological development? These are questions that will be examined in future research on corporate and other community psychopaths.

Conclusions and Future Directions

Psychopathy is an important clinical diagnosis that has significant implications for the criminal justice system. Although few in number, psychopaths are more likely than others to engage in offending and are more likely than other offenders to re-offend or commit violent offences. The specific terms used to identify this group of people have varied over time – with terms such as antisocial personality disorder, sociopathy, and dissocial personality disorder being used in varying ways to conceptualize this population. As the information presented and discussed in this entry shows, despite the longstanding interest and the burgeoning research on the topic of psychopathy, relatively little is still known about the etiology of this serious personality disorder.

With advanced knowledge of genetics and developing technology in brain imaging, it is known that a portion of the variance in the development of a psychopathic personality can be attributed to genetic factors. Similarly, research shows that functional and even some significant structural brain differences exist among people who are psychopathic that distinguish them from others; however, the differences are relatively small and are not yet diagnostic. With ongoing advances in technology and methodology, however, this area shows considerable promise for research. At present, it appears that the brain differences that do exist pertain to areas of the brain associated with affect and emotions and other areas that relate to behavioral control and impulsivity. Importantly, social development and environmental factors play a significant role in the development of psychopathic personality traits and the type of behavior that manifests. Taken together, psychopathy is best construed as a disorder that is influenced by interactions of genetic, biological, and social factors which are moderated by the environment.

Although psychopathy has been discussed in the broad population and among people who are not necessarily criminals (e.g., Cleckley 1941), the interest surrounding this disorder has been largely driven by the criminal correlates of psychopathy. As noted above, it is not surprising that people who are not influenced by society's rules and laws, who do not have the capacity for affective features like empathy and guilt, and who do not learn from punishment have increased levels of offending and recidivism. Indeed, the prevalence of psychopathy in criminal (15–20 %) and even forensic mental health (10 %) samples is much greater than what is found in the community (less than 1 %). People with high levels of psychopathic personality features have been found to have higher rates of all types of offending and violence. The offences that psychopaths commit are more likely to be instrumental, or for secondary gain, than for other offenders. The violence that they inflict is typically more extreme and callous.

Taken together, psychopathy is an important construct for those who work in criminology and related fields. There is a large and growing body of research that has enabled an enhanced understanding of psychopathy and its correlates. At the time of writing this entry, at least two significant developments were occurring in the field. First, the DSM-5 was in the final stages of development with the Personality Disorders Working Group having finished its work. It is anticipated that the DSM-5 will be published in May 2013 so one will need to wait to see exactly how antisocial personality disorder is defined at that time and whether the traditional personality features of psychopathy will be prominently featured. Second, a new measure of psychopathy, the Comprehensive Assessment of Psychopathic Personality (Cooke et al. 2004) was being developed and validated. This instrument shows promise as a measure of psychopathy that builds upon the literature. It is designed to be comprehensive and dynamic with the ability to assess possible changes in the severity of symptoms over time. Preliminary research has indicated that the CAPP shows comparable predictive validity to the PCL:SV in regard to violent and nonviolent crime (Pedersen et al. 2010). Nonetheless, it will be some time before the research base is developed to know how effective the measure will be in identifying psychopathic personality features or whether it will be best utilized as a stand-alone measure or an adjunct to the PCL-R. It is clear, therefore, that despite the long history and research interest in psychopathy, there is still much to learn.

Related Entries

- ▶ Behavioral Investigative Advice
- ▶ Career Criminals and Criminological Theory
- ▶ Criminal Investigative Analysis
- ▶ Genes, Crime, and Antisocial Behaviors
- ▶ Psychopathy and Offending
- ▶ Risk Assessment, Classification, and Prediction

Recommended Reading and References

American Psychiatric Association (1968) Diagnostic and statistical manual of mental disorders, 2nd edn. Author, Washington, DC

American Psychiatric Association (1980) Diagnostic and statistical manual of mental disorders, 3rd edn. Author, Washington, DC

American Psychiatric Association (2000) Diagnostic and statistical manual of mental disorders, 4th edn, text revision. Author, Washington, DC

Ault RL, Hazelwood RR (1995) Indirect personality assessment. In: Hazelwood RR, Burgess AW (eds) Practical aspects of rape investigation, 2nd edn. CRC Press, Boca Raton, pp 205–218

Babiak P (2007) From darkness into the light: psychopathy in industrial and organizational psychology. In: Herve H, Yuille JC (eds) The psychopath: theory, research, and practice. Lawrence Erlbaum, Mahwah, pp 411–428

Blair J, Mitchell D, Blair K (2005) The psychopath: emotion and the brain. Blackwell, Malden

Cleckley H (1941) The mask of sanity. St. Mosby, Louis

Cooke DJ, Michie C (2001) Refining the construct of psychopathy: towards a hierarchical model. Psychol Assess 13:171–188

Cooke DJ, Hart SD, Logan C, Michie C (2004) Comprehensive assessment of psychopathic Personality – Institutional Rating Scale (CAPP-IRS). Unpublished manuscript

Falret J (1866) La folie raisonnante ou folie morale. Ann Méd Psychol XXIV:382–396

Hall JR, Benning SD (2006) The “successful” psychopath: adaptive and subclinical manifestations of psychopathy in the general population. In: Patrick CJ (ed) Handbook of psychopathy. Guilford, New York, pp 459–478

Hare RD (1980) A research scale for the assessment of psychopathy in clinical populations. Personal Individ Differ 1:111–119

Hare RD (1991) The psychopathy checklist – revised. Multi-Health Systems, Toronto

Hare RD (1993) Without conscience: the disturbing world of the psychopaths among us. Guilford, New York

Hare RD (2003) The psychopathy checklist – revised, 2nd edn. Multi-Health Systems, Toronto

Hare RD, Neumann CS (2010) The role of antisociality in the psychopathy construct: comment on Skeem and Cooke (2010). Psychol Assess 22:446–454

Hart SD (1998) The role of psychopathy in assessing risk for violence: conceptual and methodological issues. Leg Criminol Psychol 3:121–137

Hemphill JF (2007) The Hare psychopathy checklist and recidivism: methodological issues and guidelines for critically evaluating empirical evidence. In: Hervé H, Yuille JC (eds) The psychopath: theory, research, and practice. Lawrence Erlbaum, Mahwah, pp 141–170

Hiatt KD, Newman JP (2006) Understanding psychopathy: the cognitive side. In: Patrick CJ (ed) Handbook of psychopathy. Guilford, New York, pp 334–352

Karpman B (1948) The myth of the psychopathic personality. Am J Psychiatry 104:523–534

Levenson MR, Kiehl KA, Fitzpatrick CM (1995) Assessing psychopathic attributes in a non-institutionalized population. J Pers Soc Psychol 68:151–158

Lilienfeld SO, Andrews BP (1996) Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal populations. J Pers Assess 66:488–524

Lilienfeld SO, Fowler KA (2006) The self-report assessment of psychopathy: problems, pitfalls, and promises. In: Patrick CJ (ed) Handbook of psychopathy. Guilford, New York, pp 107–132

Lilienfeld SO, Windows MR (2005) Professional manual for the psychopathic personality inventory-revised: (PPI-R). Psychological Assessment Resources, Lutz

Lykken D (1995) The antisocial personalities. Erlbaum, Hillsdale

Millon T, Simonson E, Birket-Smith M (1998) Historical conceptions of psychopathy in the United States and Europe. In: Millon T, Simonson E, Birket-Smith M, Davis RD (eds) Psychopathy: antisocial, criminal, and violent behaviour. Guilford, New York, pp 3–31

Monahan J, Steadman HJ, Silver E, Appelbaum P, Robbins PC, Mulvey EP et al (2001) Rethinking risk assessment: the MacArthur study of mental disorder and violence. Oxford University Press, Oxford

O'Toole ME (2007) Psychopathy as a behavior classification system for violent and serial crime scenes. In: Hervé H, Yuille JC (eds) The psychopath: theory, research, and practice. Lawrence Erlbaum, Mahwah, pp 301–325

Ogloff JRP (2006) The psychopathy/antisocial personality disorder conundrum. Aust NZ J Psychiatry 40:519–528

Ogloff JRP, Wood M (2010) The treatment of psychopathy: clinical nihilism or steps in the right direction. In: Malatesti L, McMillan J (eds) Responsibility and psychopathy: interfacing law, psychiatry, and philosophy. Oxford University Press, Oxford, pp 155–181

Pedersen L, Kunz C, Rasmussen K, Elsass P (2010) Psychopathy as a risk factor for violent recidivism: investigating the Psychopathy Checklist Screening Version (PCL:SV) and the Comprehensive Assessment of Psychopathic Personality (CAPP) in a forensic psychiatric setting. Int J Forensic Mental Health 9:308–315

Porter S (1996) Without conscience or without active conscience? the etiology of psychopathy revisited. Aggress Violent Behav 1:179–189

Porter S, Woodworth M, Earle J, Drugge J, Boer D (2003) Characteristics of sexual homicides committed by psychopathic and nonpsychopathic offenders. Law Hum Behav 27:459–469

Skeem JL, Cooke DJ (2010a) Is criminal behavior a central component of psychopathy? conceptual directions for resolving the debate. Psychol Assess 22:433–455

Skeem JL, Cooke DJ (2010b) One measure does not a construct make: directions toward reinvigorating psychopathy research – reply to Hare and Neumann (2010). Psychol Assess 22:455–459

Skeem JL, Johansson P, Andershed H, Kerr M, Eno Louden J (2007) Two subtypes of psychopathic violent offenders that parallel primary and secondary variants. J Abnorm Psychol 116:395–409

Woodworth M, Porter S (2002) In cold blood: characteristics of criminal homicides as a function of psychopathy. J Abnorm Psychol 111:436–445

World Health Organization (1992) International classification of diseases and related health problems (10th revision). Author, Geneva

Psychopathy and Offending

David J. Cooke
Glasgow Caledonian University, Glasgow, UK
University of Bergen, Bergen, Norway

Synonyms

Antisocial personality disorder; Dyssocial personality disorder

Overview

Psychopathic personality disorder (PPD) is a form of personality disorder; it influences how an individual typically thinks, feels, and behaves. Those who suffer from this disorder have difficulty forming and maintaining attached relationships with others, they are often suspicious and intolerant, their emotional experiences are impaired, their view of their self is distorted, being grandiose and self-centered, and their behavior is frequently reckless and aggressive.

Historically, PPD has been linked to offending in both preclinical and early clinical writing. In recent years research on PPD has been dominated by the Psychopathy Checklist-Revised (PCL-R), an instrument developed in criminal populations and used to make a lifetime diagnosis of the disorder. Meta-analytic studies using the PCL-R, and its derivatives, suggest a small to moderate relation between the PCL-R and future reoffending. Care must be taken in the interpretation of these results because PCL-R ratings focus, both explicitly and implicitly, on the past criminal behavior of the person being assessed.

The development of new measures of psychopathy suggests that the construct of psychopathy, as distinct from the PCL-R measure, may have predictive utility. In practical contexts, for example, risk management, a clear formulation of an individual's profile of PPD symptoms, and its relevance for future violence and risk management, is likely to have utility. Future research should focus on the relationships among specific facets of PPD and specific facets of offending behavior.

Psychopathic Personality Disorder and Offending

Personality disorders are forms of mental disorder that are chronic in nature, starting in adolescence or early adulthood; they affect how an individual thinks, feels, and behaves, the consequences being chronic disturbance in the individuals relations with self, others, and their environment. This chronic disturbance leads, in turn, to subjective distress and/or a failure to properly fulfil social roles and obligations (American Psychiatric Association 2000). In psychology traits are inferred, "latent," constructs that are hypothesized to be causal entities that underpin behavioral dispositions. Traits are inherently explanatory; they can predict a range of consequential outcomes. The symptoms of personality disorders are regarded as being extreme personality traits that lead to disorder and dysfunction.

Psychopathic personality disorder (PPD) is a specific form of personality disorder that is characterized by dysfunction in six domains of functioning. In terms of the interpersonal features, those who suffer from PPD are generally pathologically dominant in their relationships with others; they are frequently antagonistic, deceitful, and manipulative. They have difficulty forming – and maintaining – attachments to family, friends, and intimate partners; typically, they are detached, uncommitted, and unempathic. Cognitively they are frequently suspicious, intolerant, and inflexible; their emotional function is impaired because they lack appropriate anxiety, emotional depth, and empathy. Those suffering from PPD suffer a distorted view of self – an unusual form of psychological disturbance; in particular, they may display a sense of being invulnerable, unique, or entitled; they are self-centered and engage in self-justification. Behaviorally, they are unreliable, reckless, and aggressive (Cooke et al. 2012).

The concept of psychopathy, and its link to offending, has been the subject of debate and controversy: there are significant conceptual and empirical challenges in understanding the link. In this entry the association between PPD and offending will be considered from its historical roots to recent empirical research. Some of the challenges confronting the clear explicating of both the strength and quality of the association between PPD and offending will be outlined; the practical implications that this association has for the effective risk management of certain offenders will be considered.

Psychopathic Personality Disorder Crime and Antisocial Behavior

Quick reflection will reveal that it would not be surprising if the personality features enshrined in the definition of PPD alluded to above led to criminal behavior; indeed, PPD is frequently associated with diverse types of criminal behavior. Skeem et al. (2011) illustrated this diversity: the corporate psychopath who swindles investors with elaborate Ponzi schemes, the con artist who dupes others with ease and consummate skill, and even the serial killer.

Current empirical evidence suggests that those who suffer from PPD have distinctive criminal careers, starting early in life and desisting much later than other offenders – if at all – and they engage in a diverse range of types of criminal activity (Hart and Hare 1997). Their criminal conduct is often motivated by factors that differ from those which drive other offenders (e.g., the need to dominate and control; instrumental rather than expressive needs); also, the behavioral topography of their criminal acts (e.g., victim type and modus operandi) varies from those who do not suffer from this disorder. For example, those who suffer from PPD are more likely to have male strangers among their victims as compared with those who do not suffer from PPD.

General theories of crime, in terms of psychosocial and cultural factors, may have less traction in terms of explanation for the offending of this group. While PPD is regarded as a mental disorder, within the framework of Anglo-American legal theory, those with PPD are legally sane; they retain both the cognitive ability to appreciate that criminal acts are morally reprehensible and sufficient volitional control to inhibit their antisocial and criminal acts – should they choose to do so (Hart and Hare 1997).

PPD has long been associated with criminal behavior whether in ancient sources including the Bible, Chaucer's Canterbury Tales, and the Icelandic Sagas or in the early clinical descriptions of mental disorder. In the early clinical descriptions, three strands linking personality pathology and offending can be discerned. Clinical writers in the early part of the nineteenth century linked repeated acts of violence to behavioral dyscontrol (e.g., recklessness and impulsivity) in patients who suffered neither psychotic symptoms nor impaired intellectual functioning but rather suffered from inchoate forms of personality pathology (e.g., Pinel, Partridge, Prichard). Early twentieth century nosologists emphasized a second strand, an interpersonal aspect, characterized by persuasiveness and charm, self-confidence, and social assertiveness which was linked to crimes including swindling and fraud (e.g., Cleckley, Kraepelin, Schneider). A third strand that can be discerned in early clinical writings is the emotionally cold, predatory exploitativeness linked to instrumental violence (e.g., Schneider, Pinel, Rush). Thus clinicians identified three distinct aspects of PPD – interpersonal, affective, and behavioral – each of which might be linked to offending.

Moving from clinical conceptualizations to empirical studies, there is little doubt that the dominant approach to measurement of PPD over the last three decades has been the Psychopathy Checklist-Revised (PCL-R; Hare 2003). The measure was developed in the 1970s and was based explicitly on Cleckley's clinical description of psychopathy. Although designed to measure Cleckley's concept of psychopathy, the PCL-R may be regarded as imperfect as it is overly saturated with measures of criminal acts and impulsivity and omits key Clecklian features such as absence of nervousness (Skeem and Cooke 2010). Given its preeminence in the literature, the PCL-R will be considered in some detail.

The Psychopathy Checklist and Offending

The PCL-R is a 20-item scale that is completed by a trained clinician who has reviewed the subject's institutional and criminal records and has carried out a thorough interview concerning the person's history and personality features. The information about the subject is compared to operational definitions in the PCL-R manual, and the clinician decides whether there is evidence that the subject matches these operational definitions. Broadly speaking, the PCL-R items can be grouped into a number of conceptual groups: (1) Interpersonal (e.g., Grandiose Sense of Self-Worth, Glibness/Superficial Charm, Pathological Lying), (2) Affective (e.g., Shallow Affect, Callous/Lack of Empathy, Lack of Remorse or Guilt), and (3) Behavioral (e.g., Impulsivity, Irresponsibility, Need for Stimulation/Proneness to Boredom). These three aspects, which reflect the clinical condition, form the core of the disorder (Cooke and Michie 2001; Cooke et al. 2007). However, in addition, the PCL-R contains aspects that have been termed "Antisocial" (e.g., Criminal Versatility, Revocation of Conditional Release, Juvenile Delinquency); it also includes items pertaining to having many relationships and sexually promiscuity.

There can be little doubt that over the last three decades, the PCL-R has established itself both in the history of personality disorder research and in the armamentarium of applied and research psychologists. It is widely used. Hare (2003) contended that "Perhaps the PCL-R *saves* lives . . . because it helps to keep very dangerous people in prison" (p. 16, emphasis in original). Others have suggested that the application of the PCL-R can have profound prejudicial effects when it is used in certain court proceedings (Bersoff 2002).

What might the PCL-R tell us about the link between psychopathy and offending?

A common paradigm is to assess prisoners using the PCL-R and then, following their release from prison, prospectively assess their rate of reoffending. Meta-analytic studies suggest a low to moderate association between PCL-R scores and future offending including general recidivism, violent recidivism, and sexual reoffending (e.g., Leistico et al. 2008). Yang et al. (2010), looking specifically at violent offending, found that the PCL-R scores were weakly or moderately associated with future offending, but critically, the interpersonal and affective features of the disorder – the core of the disorder – failed to contribute to the prediction, the observed association being due to impulsivity and history of criminal activity. Indeed, empirical measures that quantify the chronicity of criminal behavior or criminal lifestyle appear to perform as well as the PCL-R in the prediction of both general and violent recidivism (Skeem and Cooke 2010).

While the empirical fact of an association between PCL-R scores and risk of reoffending is clear, the meaning of that association is less clear. To understand the meaning it is necessary to consider, in some more detail, what the PCL-R may actually measure.

Psychopathy and Criminal Behavior: Measurement Issues

The latent structure of the PCL-R has been the focus of some debate regarding the relationship between psychopathy and crime and offending (Cooke et al. 2007; Hare 2003; Skeem and Cooke 2010). Cooke and Michie (2001) demonstrated that PCL-R data, and other measures of the psychopathy construct, could be best represented by a three-factor model in which the superordinate construct of psychopathy was underpinned by three distinct latent traits described as: (1) Arrogant and Deceitful Interpersonal Style, (2) Deficient Affective Experience, and (3) Irresponsible Behavioral Style. Hare (2003) proposed a number of "four-factor" models in which he added what he termed an "Antisocial facet" to the Cooke and Michie three-factor model. Although described as antisocial the facet is operationalized, to a considerable degree, by consideration of criminal behavior including convictions (see above).

There are empirical and theoretical reasons why the four-facet model is, in my view, unsustainable: these issues have been considered in detail elsewhere (Cooke et al. 2007; Skeem and Cooke 2010). In brief, the preferred model of Hare (2003 et seq.) appears to be one in which the Cooke and Michie three factors are shown to be correlated with the "Antisocial facet" (Cooke et al. 2007, figure 5). In essence this correlated model merely says that criminal or antisocial behavior is correlated with the three primary traits of PPD, a perhaps unsurprising finding (Cooke et al. 2007). It does not indicate that the four facets are underpinned by a common cause, i.e., psychopathy (see Smith et al. 2003 for a detailed discussion of the important differences between compound traits and multifaceted traits). Attempts to fit an appropriate model to test the assumption that criminal and antisocial behavior are underpinned by a common cause (Cooke et al. 2007, figure 3) have been generally unsuccessful.

Why might this matter? The majority of empirical research concerning the relationship between psychopathy and offending has been carried out using the PCL-R, or its derivatives, to operationalize the construct of PPD. Our understanding about the association between psychopathy and offending – or lack of understanding – is critically linked to the qualities of the PCL-R. Putting to one side the effect of mono-method bias on our understanding of the association, there are other method and conceptual factors that require consideration.

Psychopathy and Offending: Avoiding the Tautology

Clearly, if we wish to properly understand the relationship between PPD and offending, we must begin by measuring the two constructs separately. Blackburn (2007) explained the error of mixing aspects of personality pathology with aspects of criminal and antisocial behavior within one measure. Blackburn noted that such a practice "confounds different conceptual domains because personality deviation and social deviance belong in different universes of discourse. Personality deviation is defined within the framework of interpersonal norms, whereas social deviance represents departures from legal or moral rules. To define the former in terms of the latter precludes any understanding of the relationship" [between the two] (Blackburn 2007, p. 145, citations omitted).

It is perhaps not surprising that the measurement of psychopathy in the PCL-R is confounded with the measurement of criminal behavior. The PCL-R was developed as a research tool for use in correctional populations (Hare 2003): certain items are explicitly founded upon criminal acts (e.g., Criminal Versatility, Juvenile Delinquency, and Revocation of Conditional Release) while others draw heavily upon offenders' discussion of their crimes and victims (e.g., Callous/Lack of Empathy, Parasitic Lifestyle, Failure to Accept Responsibility for Own Actions).

Failure to disaggregate the measurement of the two constructs of interest – PPD and offending – will inevitably lead to tautological thinking, a problem perhaps most eloquently expressed by Ellard (1988); thus, "Why has this man done these terrible things? Because he is a psychopath. And how do you know that he is a psychopath? Because he has done these terrible things" (p. 387).

This is not merely of theoretical importance. Walters et al. (2008) considered whether the interpersonal, affective, and behavioral facets of the PCL-R actually added to the predictive power of the "Antisocial" facet. They found that the first three facets added little incremental predictive power over and above the "Antisocial" facet. As Walters et al. (2008) observed, "If Cooke and his associates are correct in their contention that criminality is not a key component of the psychopathy picture, then the present results have portentous implications for the psychopathy construct and offer serious challenge to Hare's (1998) assertion that psychopathy is the 'single most important clinical construct in the criminal justice system' (p. 99)" (p. 403).

Given the above, should we abandon consideration of the construct of psychopathy in relation to offending or not? It can be argued that such a course would be premature given the historical, theoretical, and empirical links between PPD and offending. Blackburn's (2007) prescription of separating the measurement of personality deviance from the measurement of criminal and antisocial behavior is a first step; clearer theorizing about the risk processes entailed by the disorder is a second step. Both these steps will be considered now.

A New Model of Psychopathy

A new model of psychopath has recently been developed – the Comprehensive Assessment of Psychopathic Personality (CAPP; Cooke et al. 2012). The CAPP model is an attempt to explicate the construct of psychopathy, to describe the conceptual domain of interest. The CAPP model is distinct from measures of psychopathy but informs the development of measures designed to map the conceptual domain. Perhaps, the two signal features of the model are first, it entails symptoms defined in terms of personality pathology – stylistic variations among people – and excludes symptoms that are departures from cultural or moral standards, such as specific criminal or antisocial behaviors. Second, it is dynamic; in keeping with contemporary psychological thinking, traits are not viewed as fixed but rather they are dynamic constructs that can account for both stability and change in a range of behaviors including criminal activity, in particular, the age-crime curve (Blonigen 2010).

The CAPP model was developed by reviewing the extensive literature on the construct of PPD and by consulting subject matter experts in a systematic manner. Information about putative symptoms was subjected to a lexical analysis that led to the identification of 33 symptoms of the disorder. Symptoms are defined in atomistic terms, rather than high-level, complex, or blended terms. On rational grounds these symptoms were allocated to six conceptual domains that are consistent with past research on basic dimensions of personality pathology, namely, *Attachment* (e.g., detached, unempathic, uncaring), *Behavioral* (lacks perseverance, restless, disruptive), *Cognitive* (suspicious, lacks concentration, lacks planfulness), *Dominance* (domineering, deceitful, manipulative), *Emotional* (lacks remorse, lacks anxiety, lacks emotional depth), and *Self* (self-centered, sense of uniqueness, self-aggrandising). These domains where thought to possess heuristic value and have recently been supported by empirical results (Cooke et al. 2012).

Research on the CAPP-IRS (CAPP- Institutional Rating Scale) is still at the preliminary stage; however, results so far suggest grounds for cautious optimism. Specifically, in terms of predictive validity in relation to recidivism, Pedersen et al. (2010) showed that the CAPP-IRS was comparable to the PCL:SV (a short form of the PCL-R), the predictive validity being moderate. Thus, a method which operationalizes the PPD construct without reference to a history of offending performed as well as the PCL:SV. In a post-dictive study, the capacity of the six CAPP domain scores and three primary PCL-R facets scores to predict Facet 4 of the PCL-R, an index of antisocial and criminal behavior, was evaluated. Stepwise regression analyses demonstrated that the CAPP-IRS Behavioral domain score had superior predictive power over any of the three primary PCL-R facet scores; this is particularly interesting given that the CAPP ratings were based on the previous 6 months while the PCL-R ratings were based on lifetime ratings (Cooke et al. 2012).

These results echo results from other measures of psychopathy which do not depend, to any degree, on ratings of criminal behavior, e.g., The Psychopathic Personality Inventory (PPI-R; Lilienfeld et al. 2005). Skeem et al. (2011) noted that the PPI-R has predictive validity in relation to institutional misconduct, violence, and general offending in a range of populations including adults and young offenders, and community and forensic psychiatric residents. These findings are suggestive that even when construct-criterion contamination is removed, it is still possible to demonstrate a link between PPD and offending.

From Groups to Individuals

The fundamental driver for the renaissance of interest in the concept of PPD over the last few decades has been concern about identifying those who may be of risk of committing violence in the future. The irresistible rise of risk assessment has promoted interest in PPD not least because measures such as the PCL-R, and its derivatives, have been shown to be *relatively* strong risk factors – compared with other risk factors – for future violence. Two broad approaches to risk assessment can be discerned; the actuarial, an approach in which scores on instruments like the PCL-R are used to make a numerical prediction about the likelihood of reoffending, and Structured Professional Judgement (SPJ) approaches where comprehensive knowledge of the individual, and their circumstances, are used to plan a risk management plan designed to obviate and perceived risk. The actuarial approach, although popular, is fundamentally misleading. One foundation problem is the assumption that knowledge about an individual's PCL-R score can provide meaningful predictions about that individual's future offending. Cooke and Michie (2010) demonstrated that the prediction interval (i.e., confidence interval associated with a point prediction) associated with predictions of future offending were enormous – typically between 1 % and 99 % – indicating that such predictions have no value when making decisions about individual cases.

From Statistical Thinking to Psychological Thinking

Should we then dismiss and ignore information about PPD, or might such information have some heuristic or clinical value or relevance. A risk factor can be said to have clinical relevance if it either contributes to the understanding of why an individual might go on to offend violently, or because it contributes to our understanding of why the individual might be difficult to engage in risk management through treatment, monitoring, supervision and victim safety planning.

One of the reasons that the PCL-R has been associated with future violence – *at the aggregate level* – is because it taps into many symptoms (and sequelae) of psychopathy that either increase risk of violence or interfere with effective risk management (Cooke 2010). PCL-R scores are silent on cause; their relevance to risk processes in the individual case has to be explicated. In order to understand future risk, the assessor must disaggregate these risk processes and identify those that apply to the particular individual that they are managing. To illustrate, the nature of the risk – and the risk management plan – of an individual who poses a violence risk because he is pathologically dominant possesses a sense of entitlement and construes others as having malevolent intent would be different to the management strategy for someone who is callous, reckless, and impulsive. Each individuals might be risky because of their particular pattern of psychopathic traits; however, the psychological processes underpinning their risk – and thereby the interventions required – could be quite different.

Psychopathy has been described as a "mini-theory" of violence (Hart 1998); indeed, distinct aspects of psychopathy may be conceptualized as *drivers*, *disinhibitors*, or *destabilizers* that act to increase the likelihood of violence (Cooke 2010). The status needs of psychopathic individuals may drive *violence* because these needs frequently lead to suspiciousness and a tendency to construe maleficent motives underpinning the behavior of others. Those with psychopathic personality disorder are often overly sensitive to challenges to their status. Aggression is a common response to the fear of losing face. Specific forms of violence, e.g., sadism, may be founded on the desire to control, demean, and humiliate others which is a consequence of status needs and other forms of self pathology, including possessing a sense of entitlement.

The emotional deficits that characterize PPD including lack of empathy, guilt, and fear have relevance for criminal behavior, in general, and violent behavior, in particular. The lack of emotional depth, remorse, or empathy all may be considered as *disinhibitors* with the absence of these affective responses resulting in the failure to inhibit antisocial cognitions and actions. If you lack anxiety ahead of a violent act, or you lack remorse for a violent act that you have committed, there is neither inhibition nor cost associated with that act. The behavioral features of the disorder including disruptiveness, recklessness, unreliability, and absence of perseverance, together with attachment deficits including being uncaring and uncommitted, can act as *destabilizers*. These features cause the breakdown in reliable social structures, supports, and routines, and this creates the chaotic circumstances in which violent behavior is more likely (Cooke 2010).

In the applied world, the clinician has to take a fundamentally nomothetic approach to the understanding of an individual offender and their particular configuration of symptoms of PPD. As Millon noted recently, "It is only the unique way in which the personality construct is seen in real patients that is ultimately of clinical value." (Millon 2011; p. 296). An individualized formulation has to be developed in which an understanding of the nature and etiology of an individual's difficulties is developed in a precise and systematic manner, communicated to the client, and others, in the form of a narrative written in natural language. Models such as the CAPP model of PPD can be applied to provide a precise and systematic description of how the symptoms of PPD might influence an individual's likelihood of engaging in future offending, but also, his manageability. The clinician's role is to use this, and other, information in order to establish a risk management plan that entails strategies of treatment, supervision, monitoring, and victim safety planning designed to obviate future violence risk and which takes into account the particular pattern of PPD traits displayed by the client.

Looking to the Future

While historically, PPD and offending have been linked; contemporary research leaves many key questions unanswered, are current empirical findings the result of construct-criterion contamination, i.e., tautological thinking, and are particular aspects or facet of PPD of particular relevance to future offending, in general, or violent offending, in particular; what moderating influence do other key variables, e.g., gender, age, and psychiatric status, have on the putative link between PPD and offending?

In order to answer these – and other – questions, it is necessary to improve our measurement of both PPD and of offending; most studies have depended on the PCL-R and on simple recidivism, the latter criterion being an impoverished account of a complex phenomenon. Future research should endeavor to explore the specificity of relationships between aspects of PPD and aspects of offending. This paradigm is common in other areas of quasi-experimental research. The specific role of cigarette smoking in the etiology of lung cancer only became apparent when the relationship between all types of smoking and all disease was refined. Similarly, it could be conjectured that certain aspects of PPD might be linked with certain types of violence, e.g., instrumental versus reactive.

Conclusions

PPD and offending have long been associated. The empirical evidence goes some way to support a link but methodological problems cloud the true nature of the relationship. From a theoretical perspective new approaches to measurement are required to clarify key issues. From a practical perspective detailed studies, including systematic case studies, are required to clarify the why question: why do particular symptoms, or configurations of symptoms, of PPD increase the likelihood of offending?

Related Entries

- ▶ Establishing Causes of Offending in Longitudinal and Experimental Studies
- ▶ Evolutionary Theories of Criminal Behaviors
- ▶ Psychopathy
- ▶ Risk Factors for Prison Recidivism

Recommended Reading and References

American Psychiatric Association (2000) Diagnostic and statistical manual of mental disorders: text revision. American Psychiatric Association, Washington, DC

Bersoff D (2002) Some contrarian concerns about law, psychology and public policy. Law Hum Behav 26(5):565–574

Blackburn R (2007) Personality disorder and antisocial deviance: comments on the debate on the structure of the psychopathy checklist-revised. J Personal Disord 21(2):142–159

Blonigen DM (2010) Explaining the relationship between age and crime: contributions from the developmental literature on personality. Clin Psychol Rev 30:89–100

Cooke DJ (2010) Personality disorder and violence: understand violence risk and risk processes. J Personal Disord 24(5):539–550

Cooke DJ, Michie C (2001) Refining the construct of psychopathy: towards a hierarchical model. Psychol Assess 13:171–188

Cooke DJ, Michie C (2010) Limitations of diagnostic precision and predictive utility in the individual case: a challenge for forensic practice. Law Hum Behav 34:258–274

Cooke DJ, Forth A, Hare RD (1998) Psychopathy: theory, research and implications for society. Kluwer, Dordrecht

Cooke DJ, Michie C, Skeem JL (2007) Understanding the structure of the psychopathy checklist revised: an exploration of methodological confusion. Br J Psychiatry 190(Suppl 49):39–50

Cooke DJ, Hart SD, Logan C, Michie C (2012) Explicating the construct of psychopathy: the development of the Comprehensive Assessment of Psychopathic Personality (CAPP) model. Paper Under Review

Ellard J (1988) The history and present status of moral insanity. Aust N Z J Psychiatry 22(4):383–389

Hare RD (2003) Manual for the revised psychopathy checklist, 2nd edn. Multi-Health Systems, Toronto

Hart SD (1998) The role of psychopathy in assessing risk for violence: conceptual and methodological issues. Leg Criminol Psychol 3(121):137

Hart SD, Hare RD (1997) Psychopathy: assessment and association with criminal conduct. In: Stoff DM, Maser J, Brieling J (eds) Handbook of antisocial behaviour. Wiley, New York

Leistico AR, Salekin RT, DeCoster J, Rogers R (2008) A large-scale meta-analysis relating the Hare measures of psychopathy to antisocial conduct. Law Hum Behav 32:28–45

Lilienfeld SO, Widows MR, Staff P (2005) Psychopathic personality inventory TM-revised. Soc Influ (SOI) 61(65):97

Millon T (2011) Classifying personality disorders: an evolution-based alternative to an evidence-based approach. J Personal Disord 25(3):279–304

Patrick CJ (2006) Handbook of psychopathy. The Guilford Press, New York

Pedersen L, Kunz C, Rasmussen K, Elsass P (2010) Psychopathy as a risk factor for violent recidivism: investigating the psychopathy checklist screening version (PCL: SV) and the Comprehensive Assessment of Psychopathic Personality (CAPP) in a forensic psychiatric setting. Int J Forensic Ment Health 9(4):308–315
Skeem JL, Cooke DJ (2010) Is criminal behavior a central component of psychopathy? Conceptual directions for resolving the debate. Psychol Assess 22:433–445
Skeem JL, Polaschek DLL, Patrick CJ, Lilienfeld SO (2011) Psychopathic personality bridging the gap between scientific evidence and public policy. Psychol Sci Publ Interes 12(3):95–162
Smith GT, Fischer S, Fister S (2003) Incremental validity principles in test construction. Psychol Assess 15(4):467–477
Walters GD, Knight RA, Grann M, Dahle KP (2008) Incremental validity of the psychopathy checklist facet scores: predicting release outcome in six samples. J Abnorm Psychol 117(2):396
Yang M, Wong S, Coid J (2010) The efficacy of violence prediction: a meta-analytic comparison of nine risk assessment tools. Psychol Bull 136(5):740–767

Public Criminology

► So What Criminology?

Public Defenders

Matthew Kleiman and Cynthia G. Lee
National Center for State Courts, Williamsburg, VA, USA

Synonyms

Indigent defense

Overview

The United States Constitution guarantees the right to be represented by attorney to every person accused of a crime whose life or liberty is at stake. When a defendant cannot afford to hire his own attorney, the government is obligated to provide him not merely with token representation but with effective assistance of counsel. In recent years, indigent defense systems have been characterized as overloaded, and the constitutional promise of effective legal representation for the poor in state courts has been perceived as "broken" (Justice Policy Institute 2011; American Bar Association 2004). Among the primary challenges are excessive caseloads and a lack of adequate funding for defense services. In order to function properly, the adversarial system used in American courts relies upon the presence of capable counsel for both the state and the accused. Without appropriate resources, the ability of the defense attorney to provide effective assistance of counsel is compromised, jeopardizing the constitutional rights of the accused and the integrity of the criminal justice system. According to former US Attorney General Janet Reno, "[i]ndigent defense is an equally essential element of the criminal justice process, one which should be appropriately structured and funded and operating with effective standards. When the conviction of a defendant is challenged on the basis of inadequate representation, the very legitimacy of the conviction itself is called into question. Our criminal justice system is interdependent: if one leg of the system is weaker than the others, the whole system will ultimately falter" (American Bar Association 2004). This entry explores the constitutional foundations of the right to counsel, provides an overview of the various means by which states provide indigent defense services, outlines the numerous tasks and functions that public defender attorneys and staff must perform in order to provide effective assistance of counsel, and examines the implications of an underfunded indigent defense system.

The Right to Counsel

The Sixth Amendment of the United States Constitution provides that "[i]n all criminal prosecutions, the accused shall enjoy the right . . . to have the Assistance of Counsel for his defence." Until the middle of the twentieth century, the Sixth Amendment was held to apply only to the federal government, not to the states, and was interpreted to require the government only to allow the defendant to hire his own lawyer, not to appoint one for him. A number of states did provide counsel in certain cases, either by statute or by virtue of a right created under the state constitution, but the appointment of counsel was typically limited to capital cases (Wice 2005).

In 1932, the Supreme Court held for the first time that a state government was obligated to provide the defendant with counsel in a capital case. Basing its decision upon the Fourteenth Amendment right to due process of law rather than the right to counsel enumerated in the Sixth Amendment, the Court explained that

> The right to be heard would be, in many cases, of little avail if it did not comprehend the right to be heard by counsel. Even the intelligent and educated layman has small and sometimes no skill in the science of law. He lacks both the skill and knowledge adequately to prepare his defense, even though he had a perfect one. He requires the guiding hand of counsel at every step in the proceedings against him. Without it, though he be not guilty, he faces the danger of conviction because he does not know how to establish his innocence. (Powell v. Alabama 1932)

The Powell v. Alabama decision recognized a defendant's right to appointed counsel only in capital cases tried in state courts. In 1938, the Supreme Court held that the Sixth Amendment required the federal government to provide counsel for a defendant facing federal noncapital felony charges who was unable to obtain counsel on his own, unless the defendant had "intelligently waived" the right to counsel (Johnson v. Zerbst 1938). Just four years later, however, the court in Betts v. Brady declined to require states to appoint counsel in all noncapital felony cases, finding that the Sixth Amendment right to counsel was not a "fundamental right" that applied against the states through the due process clause of the Fourteenth Amendment (Betts v. Brady 1942).

In 1963, the Supreme Court overturned Betts v. Brady in the case of Gideon v. Wainwright, finally recognizing a right to appointed counsel in all felony cases prosecuted in state courts. The state of Florida had charged Clarence Earl Gideon with the felony offense of breaking and entering a poolroom with intent to commit a misdemeanor. Unable to afford an attorney, Gideon asked the trial court to appoint one for him. The court denied his request, and after representing himself at trial, Gideon was convicted and sentenced to five years in prison. Gideon petitioned the Florida Supreme Court for a writ of habeas corpus, claiming that his conviction violated his rights under the US Constitution. After the Florida Supreme Court denied relief, Gideon appealed to the Supreme Court of the United States, which agreed to hear his claim and appointed a prominent attorney to represent him on the appeal. The Court went on to hold that the Sixth Amendment right to appointed counsel was indeed fundamental in nature and thus applied against the states under the Fourteenth Amendment's due process clause, declaring it an "obvious truth" that "in our adversary system of criminal justice, any person haled into court, who is too poor to hire a lawyer, cannot be assured a fair trial unless counsel is provided for him" (Gideon v. Wainwright 1963).

Following its decision in Gideon, the Supreme Court interpreted the right to appointed counsel to apply in juvenile delinquency proceedings (In re Gault 1967) and in misdemeanor cases in which a sentence of incarceration is actually imposed (Argersinger v. Hamlin 1972). The Sixth Amendment right to counsel attaches at "a criminal defendant's initial appearance before a judicial officer, where he learns the charges against him and his liberty is subject to restriction," and counsel must be provided "during any 'critical stage' of the post-attachment proceedings" Rothgery v. Gillespie County (2008). The Supreme Court has also interpreted the Fourteenth Amendment to require the appointment of counsel for a defendant's first appeal as of right, as well as for a discretionary appeal by a defendant who has pled guilty (Douglas v. California 1963; Halbert v. Michigan 2005). Finally, the court has held that the due process and equal protection clauses of the Fourteenth Amendment require the state to furnish an indigent defendant with expert witnesses relevant to material issues (Ake v. Oklahoma 1985).

The mere fact that an attorney appears on behalf of the defendant is not sufficient to fulfill the Sixth Amendment right to counsel. When a defendant cannot afford to hire his own attorney, the government is obligated to provide him not just with counsel but with effective assistance of counsel. Effective assistance of counsel requires that the attorney exercise reasonable professional judgment under the circumstances of the case (Strickland v. Washington 1984). Attorneys in every state are also bound to abide by the state's rules of professional conduct. Many states pattern their ethical rules for attorneys on the American Bar Association's Model Rules of Professional Conduct, which require the attorney to provide "competent" and "diligent" representation to every client. These requirements include a responsibility to control the lawyer's workload "so that each matter can be handled competently." The attorney must also promptly communicate with the client regarding significant developments in the case, such as the prosecution's offer of a plea bargain, and meaningfully involve the client in decision-making regarding the case (American Bar Association 2007). The American Bar Association's Standards For Criminal Justice: Defense Function, although not binding upon attorneys, also establish norms for the performance of defense attorneys. Standard 4-1.3(e) specifies that "[d]efense counsel should not carry a workload that, by reason of excessive size, interferes with the rendering of quality representation, endangers the client's interest in the speedy disposition of charges, or may lead to the breach of professional obligations." The standards also set expectations for the defense attorney's performance in specific activities such as investigating the facts of the case and exploring the possibility of disposition without trial.

Indigent Defense Systems

An indigent defendant may be represented by an attorney from a public defender's office, assigned counsel, or a contract attorney. In a public defender program, salaried attorneys represent the majority of indigent defendants within a jurisdiction. Public defender offices exist primarily in larger communities, and public defender attorneys typically work together in a central office that is headed by a director and supported by additional staff (Wice 2005). Public defenders may be government employees, and the public defender office may be an independent nonprofit agency such as the Legal Aid Society of New York. In smaller jurisdictions, a system of assigned counsel is typically used. Under this model, private attorneys are selected from a list, either formal or informal, to take cases and are paid either by the case or by the hour (Lefstein and Spangenberg 2009). Many states cap the fees or hours an attorney may bill per case; some states allow judges to waive these caps upon request, but others do not. Roughly 2,800 out of 3,100 counties in the United States rely exclusively on assigned counsel to provide indigent defense services (Wice). Jurisdictions with public defender offices also use assigned counsel in cases involving conflicts of interest, such as those with multiple defendants. Finally, under the contract model, private attorneys bid to provide representation for all or a percentage of indigent defendants in a particular jurisdiction. Contract attorneys may be paid a flat fee to accept all assigned defendants or may be paid according to their caseload.

Indigent defense systems are paid for by state governments, local governments, or a combination of both. As of fiscal year 2008, indigent defense services in 30 states were supported primarily or entirely through state funds; these states include Florida, Maryland, Missouri, New Hampshire, New Mexico, and Virginia. Indigent defense systems in 18 states (including Alabama, California, Georgia, Michigan, Nevada, New York, and Texas) rely on counties for more than half of their funding; in Utah and Pennsylvania, indigent defense services are funded entirely by county governments (Lefstein and Spangenberg 2009). County-funded systems have been criticized because they create the potential for varying levels of funding, access, and quality depending upon where a defendant lives (Justice Policy Institute 2011).

Public Defender Tasks and Functions

The popular image of a criminal attorney centers upon the exchange of evidence and arguments in the confines of the courtroom. On one side is the prosecution; on the other, the defense. Each attorney calls witnesses, presents evidence to the judge and jury, and summarizes the case in closing arguments. For attorneys who represent indigent defendants, however, the popular view of the role of a criminal attorney is extremely myopic. Instead, the work of the defense team extends well beyond the courtroom, with defenders alternately wearing the hats of investigator, social worker, appellate attorney, and litigator.

Recent work by the National Center for State Courts highlights the diverse set of case-related and non-case-related tasks and functions that indigent defense attorneys and their support staff perform (Kleiman and Lee 2010; Steelman et al. 2007; Ostrom et al. 2005). Case-related activities include all tasks and functions directly associated with the handling of individual cases, from the time of appointment through posttrial and appellate activity. Both attorneys and staff must also devote time to certain core activities that are not directly related to a particular case, such as continuing legal education, attending training and conferences, and handling administrative matters such as budgets and personnel issues, community outreach, and staff meetings.

Attorney Activities

The basic case-related duties and responsibilities of defense attorneys can be sorted into a series of complementary functional areas. The first function is pretrial activities and hearings. This function encompasses all of the in-court and out-of-court work associated with pretrial proceedings, including legal research, preparation, participation in pretrial hearings, and drafting and arguing pretrial motions. Attorneys prepare for and participate in a host of bond and detention hearings, arraignments, pretrial conferences, status conferences, competency hearings, motion hearings, and other pretrial hearings. Additionally, defense attorneys are responsible for preparing and arguing motions regarding the evidence to be used at trial (e.g., motions to suppress, motions in limine) and motions for funds to secure expert witnesses. A second critical function of the defense attorney is negotiating plea alternatives with the prosecution. More than 90 % of criminal cases are resolved without trial through plea bargaining. In most cases, the defendant agrees to plead guilty in exchange for a reduction of the charges or sentence; other cases are resolved without a conviction through programs such as diversion.

A crucial – and often time-consuming – function for defense attorneys is client contact. The American Bar Association's Ten Principles of a Public Defense Delivery Service assert that high-quality delivery of public defense necessitates that "counsel is provided sufficient time and a confidential space within which to meet with the client." The defense attorney is obligated to inform the client of his or her rights at the earliest opportunity, build a rapport with the client that instills trust and confidence, and interview the client about the events associated with the charges. Attorneys are also responsible for keeping the client informed of case developments, responding to all client correspondence and inquiries, promptly explaining proposed plea bargains, and engaging the client in meaningful plea discussions. It is not unusual for an attorney to spend part of each day responding to numerous messages from clients calling to inquire about hearing dates and case status. Meeting with in-custody clients can be especially time-consuming, as the attorney must travel to the jail, contend with security procedures, and wait for the client to be produced.

In order to provide an effective defense, public defenders must perform investigative and discovery activities. Attorneys are responsible for directing the activities of investigative staff, visiting the crime scene, and, when necessary, conferring with social workers and independent experts (e.g., forensic experts). Attorneys are also responsible for interacting with the prosecution regarding discovery and for preparing and submitting discovery requests. In short-staffed offices, attorneys often perform duties that could otherwise be handled by non-attorney investigative staff. These include reviewing audio and video recordings of hearings and interviews, interviewing prosecution and defense witnesses, visiting the crime scene to take measurements and photographs, preparing summary reports of investigations, and serving witness subpoenas.

A more "traditional" function of the defense attorney is participating in bench and jury trials. In addition to in-court work, trials may require a substantial amount of out-of-court preparation. Trial-related activities include preparing for and participating in jury selection, drafting proposed jury instructions, preparing and presenting the opening statement, cross-examining prosecution witnesses, and preparing and presenting the defense case and closing argument. Following a guilty verdict at trial, the defense attorney is responsible for sentencing and posttrial activities, including activities related to sentencing, posttrial motions, probation violations, appeals, and collateral proceedings such as petitions for habeas corpus. Specific post-conviction tasks include researching and preparing posttrial motions (e.g., motion for new trial), reviewing presentence reports, sentencing memoranda and the defendant's social history, addressing the admissibility of the prosecution's sentencing evidence, preparing for sentencing hearings, reviewing and correcting sentencing orders, investigating the circumstances surrounding probation violations, and preparing for probation revocation hearings. Tasks related to appeals and collateral proceedings include preparing petitions and briefs, reviewing the trial record, legal research, and preparing for oral arguments. Appeals and collateral proceedings may be handled by the attorney or office that originally represented the defendant at trial or by an attorney or office specializing in appellate work.

Finally, some attorneys are responsible for social work/sentencing advocacy tasks. Such tasks include identifying alternative sanction options; finding program placements; compiling the defendant's medical, educational, and family history; and coordinating mental health evaluations and reviews.

Staff Activities

Effective assistance of counsel depends on the combined work of a defense team comprising both attorneys and non-attorney support staff. Public defender staff efficiently and cost-effectively perform many functions integral to the effective assistance of counsel, from investigating the facts of the case to locating alternative placement options to maintaining complete and accurate case files. Without sufficient support staff resources, public defenders' ability to provide competent and diligent representation to indigent clients may be seriously compromised. Inadequate staff support also reduces the efficiency of the public defender office, as attorneys are forced to take time away from legal work to answer telephones, type documents, and maintain case files themselves.

Support staff tasks can be classified into two broad categories. The first set of staff functions is administrative in nature and includes activities such as client intake, records management, and secretarial services. Client intake includes tasks such preparing file folders, verifying claims of indigency, handling intake paperwork, scheduling initial client appointments, identifying potential conflicts of interest, and obtaining charging documents from court files. Records management includes entering data into case management systems, archiving and retrieving files, and handling billing for contract attorneys. Finally, secretarial services include tasks such as taking telephone messages, managing attorneys' calendars, typing documents such as motions and subpoenas, and transcribing witness interviews.

The second category of support staff functions consists of professional and paraprofessional tasks directly associated with the facilitation of legal representation, including investigative services, legal research, direct attorney support, and social work/sentencing advocacy. Investigative services are typically performed by an investigator who is responsible for locating and interviewing defense and prosecution witnesses, reviewing the offense report and discovery package, visiting the crime scene to take measurements and photos, viewing and obtaining evidence, preparing summary reports of investigations, and testifying in court when necessary. Legal research is a function typically handled by a legally trained staff member, such as a paralegal, who reviews case files to identify legal issues, prepares legal memoranda, and performs legal research related to procedural and evidentiary issues. Direct attorney support includes in-court and out-of-court activities that provide the attorney with logistical support. Out of court, staff locate clients (e.g., identifying the jail in which a defendant is being held) and determine whether a client is a defendant in other criminal proceedings in the current jurisdiction or in other jurisdictions. Staff members also provide direct support to attorneys in the courtroom during all types of court proceedings, such as arraignments, bail reviews, preliminary hearings, motion hearings, trials, and sentencing hearings.

Finally, a specialized group of support staff provide defense attorneys with information related to a client's medical, psychiatric, educational, and family history. Social work/sentencing advocacy staff develop mitigating information for sentencing, match defendants with appropriate drug treatment programs and other alternatives to incarceration, investigate the facts surrounding probation violations, coordinate witness appearances, and directly assist defendants with matters such as jail issues and obtaining legal identification.

Caseload Standards

To ensure that public defenders are able to fulfill their constitutional obligation to provide effective assistance of counsel, a set of national caseload standards were developed in 1973. Caseload standards are important because public defenders do not exercise direct control over the number of defendants they are assigned to represent and therefore the amount of work for which they are responsible. The National Legal Aid & Defender Association (NLADA) points out that public defender workloads "rise or fall due to a convergence of societal trends and decisions made by legislatures, police departments and prosecutors which are completely beyond the control of indigent defense providers" (NLADA 2008). ABA Principle 5 proclaims that national caseload standards should not be exceeded, and the workload of an attorney "should never be so large as to interfere with the rendering of quality representation or lead to the breach of ethical obligations, and counsel is obligated to decline appointments above such levels."

Developed in 1973 by the National Advisory Commission on Criminal Justice Standards and Goals, the existing national caseload standards represent the maximum number of cases that a defense attorney should handle in a year, assuming that the attorney handles only a single case type. The commission recommended that caseloads should not exceed 150 felonies, 400 misdemeanors (excluding traffic), 200 juvenile court cases, 200 Mental Health Act cases, or 25 appeals per attorney per year. In recent years, several states, including Virginia, Maryland, and New Mexico, have developed their own state-specific caseload standards for attorneys and support staff in public defender offices. These efforts have been undertaken in response to the perceived limitations of the 1973 national caseload standards. Criticisms of the existing standards include the facts that the standards were not founded upon empirical research, do not allow for the varying complexity of different felony case types (e.g., homicide, violent felonies, and nonviolent felonies), do not capture the variability between states in criminal justice procedures and practices, and predate the introduction of information technology resources commonly used by today's attorneys and support staff. Furthermore, the standards are silent on the level of staff support attorneys require in order to provide effective representation.

The new state-specific standards are expressed as case weights, which represent the average amount of time needed for an attorney or staff member to handle effectively each case of a particular type, from appointment through post-disposition activity. Separate case weights are calculated for different types of cases of varying levels of complexity (e.g., homicide, violent felony, misdemeanor, driving under the influence of alcohol, juvenile delinquency). For example,

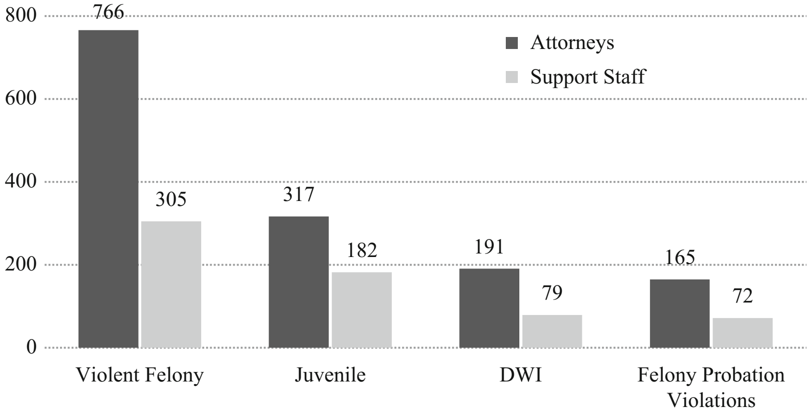

Public Defenders, Fig. 1 Attorney and support staff case weights (minutes) in Virginia

in Virginia, a violent felony case requires, on average, 766 minutes (12.8 hours) of hands-on attorney time and 305 minutes (5.1 hours) of staff time from appointment through post-disposition activity. In comparison, a nonviolent felony case in Virginia requires, on average, 433 minutes (7.2 hours) of attorney time. Case weights are typically constructed from empirical data collected during a "time study" that measures current practice, then reviewed by panels of expert attorneys and support staff to ensure that the weights allow sufficient time to provide effective representation. The case weights can be used in combination with caseload data to calculate total attorney and staff workload for a public defender's office, along with the number of attorneys and support staff members needed to handle this workload effectively.

An Example from Virginia

Figure 1 displays a subset of the case weights developed for public defender offices in Virginia in 2010. Each case weight represents the average number of minutes of attorney or staff time required to defend one case of the specified type. Figure 1 illustrates two important concepts about the nature of public defense cases. First, cases handled by public defenders vary in complexity, as measured by the amount of attorney and staff time needed to provide effective assistance of counsel. For example, an average violent felony case requires twice as much attorney time as a juvenile case and four times as much as a DWI case or a felony probation violation. Second, effective assistance of counsel requires the combined efforts of attorneys and support staff. For example, a typical juvenile case requires approximately 5 h of attorney time and 3 h of staff time.

Figure 2 illustrates a third important concept: the relative amount of time attorneys (and staff) devote to different tasks and functions varies dramatically among case types. For example, in Virginia, approximately one-half of the attorney work required to defend violent felony cases is expended on pretrial proceedings, investigation, discovery, and plea negotiations. This results from factors such as the large amount of evidence typically associated with violent crimes and the complexity of plea negotiations in these high-stake cases. In felony probation violation cases, the largest share of attorney time is devoted to contact with the client and the client's family. In DWI cases, on the other hand, a larger proportion of time is spent on legal research.

Implications of Excessive Caseloads

The Washington Public Defender Association asserts that "caseload levels are the single biggest predictor of the quality of public defense representation. Not even the most able and industrious lawyers can provide effective representation when their workloads are unmanageable. A warm body with a law degree, able to affix his or her name to a plea agreement, is not an acceptable substitute for the effective advocate

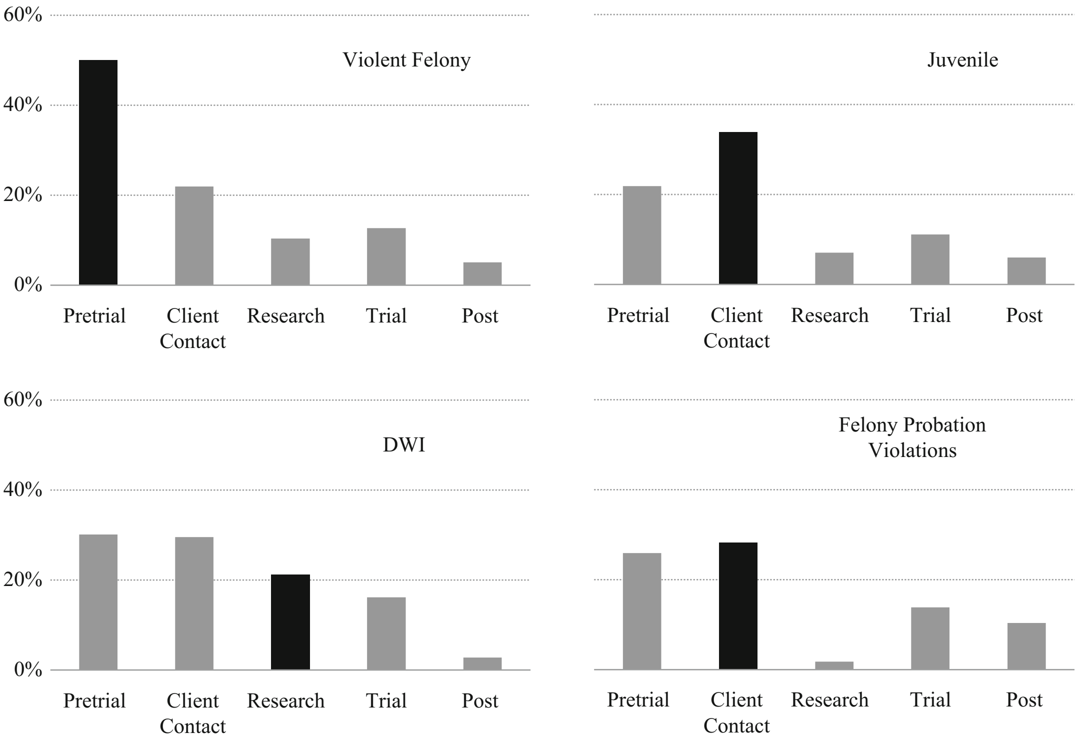

Public Defenders, Fig. 2 Proportion of Virginia public defender attorney time spent on tasks and functions by case type

envisioned when the Supreme Court extended the right to counsel to all persons facing incarceration." Similarly, a recent report by the Justice Policy Institute argues that "[p]roper resources and adequate time may not guarantee that every public defender will provide quality representation, but high caseloads and limited resources almost guarantee that this will not happen for all clients" (Justice Policy Institute 2011). In many jurisdictions across the nation, excessive caseloads currently jeopardize the quality of indigent defense services (Lefstein 2011). For example, in 2006, public defenders in Clark County, Nevada, had average per-attorney caseloads of 364 felony and gross misdemeanor cases (Lefstein and Spangenberg 2009). Although this figure includes felonies and excludes less serious misdemeanor cases, it is much closer to the 1973 recommendation for misdemeanor-only caseloads (400 per attorney per year) than the corresponding recommendation for felony caseloads (150 per attorney per year). Similarly, a recent census of public defender offices by the Bureau of Justice Statistics found that 15 state programs exceeded the 1973 recommendations for the maximum numbers of felony and misdemeanor cases per attorney (Langton and Farole 2010). Resource constraints have been exacerbated by the recent fiscal crisis, which has resulted in attorney and staff reductions for many public defender systems. For example, the Minnesota Board of Public Defense cut 15 % of its attorney positions in 2008 in response to budget shortfalls. Similarly, the Maryland Office of the Public Defender has recently reduced its workforce by nearly 20 % due to budget cuts (DeWolfe 2011).

Faced with excessive caseloads, public defenders are often forced to adopt a "triage" approach, rationing their time and focusing scarce resources on the most serious or urgent matters. Under such conditions, attorneys may not have sufficient time or opportunity to enter an appearance promptly during the critical early stages of the case, prepare thoroughly for all proceedings, communicate directly and in a timely manner with the accused, fully interview the accused and other witnesses to explore pertinent facts in a case, conduct a thorough investigation of all relevant facts and circumstances in each case, duly focus on each case and provide effective representation at trial, and prepare for and meaningfully participate in the sentencing process (Ostrom, Kleiman, Ryan).

ABA Principle 8 articulates the need for "parity between defense counsel and the prosecution with respect to resources." Without sufficient resources to effectively handle indigent defense caseloads, guarantees of a fair and balanced adversarial process are compromised, with significant costs for individuals, the criminal justice system, and society as a whole. First, the failure to provide effective assistance of counsel can result in the unnecessary imprisonment of innocent defendants or in the imposition of unduly long prison sentences due to errors in scoring sentencing guidelines or failure to present mitigating evidence. The Justice Policy Institute proclaims that "[w]ithout a functioning adversarial justice system, everyday human error is more likely to go undiscovered and result in the tragedy of innocent people being tried, convicted and imprisoned" (American Bar Association 2004). Erroneous convictions and excessive sentences impose extra costs on taxpayers and the criminal justice system in terms of additional jail and prison bed days. Unnecessary incarceration is profoundly disruptive not only to the defendant himself but to his family and employer as well. A wrongful conviction also leaves the true perpetrator free to commit additional crimes, with corresponding societal costs. Second, ineffective assistance of counsel can lead to an increase in the number of appeals and post-judgment challenges to convictions, resulting in increased costs to the criminal justice system (Williams and Bost 1971). Third, when the defense attorney has insufficient time to prepare for and appear at court proceedings early in the life of a case (e.g., bail review hearing, arraignment), the defendant may remain in pretrial detention longer than necessary, or the opportunity for an early resolution of the case may be lost, resulting in increased jail, attorney, and court system expenditures. Finally, a lack of quality representation for indigent defendants can adversely impact perceptions of fairness and public trust and confidence in the criminal justice system. The National Legal Aid & Defender Association states that "public confidence in the integrity of the system is lost when the community perceives that inadequate representation creates a system that metes out justice differently to the rich and the poor" (NLADA 2004). It is not only high-profile reversals of wrongful convictions that can undermine public opinion regarding the quality of indigent defense; damage is also done when defendants, their families, jurors, and other participants in the criminal justice process perceive indigent defense attorneys to be overworked, ill prepared, or unresponsive to the needs and concerns of individual defendants.

Conclusion

Public defenders and other attorneys who represent indigent criminal defendants are essential to the due process of law in our nation's criminal courts. The effective delivery of indigent defense services requires the joint effort of attorneys and support staff, whose work extends well beyond the confines of the courtroom. Recent studies have echoed a growing concern that "[t]he defender systems that people must turn to are too often completely overwhelmed; many dedicated defenders simply have too many cases, too little time and too few resources to provide quality or even adequate legal representation" (Justice Policy Institute 2011). In his 2005 State of the Judiciary address, former New Mexico Chief Justice Richard Bosson observed that "the fiscal needs of the [New Mexico] Public Defender are so dire, their situation seems so hopeless, that many times prosecutions cannot go forward due to lack of sufficient personnel." Chief Justice Bosson went on to describe the criminal justice system with the oft-used metaphor of the "three-legged stool" supported by the prosecution, the defense, and the court: "When one leg is weakened, you know what happens; you end up on the floor. Well, we are not on the floor yet, but we are not far off." As the current fiscal crisis threatens further reductions to already strained budgets for indigent defense services, the three-legged stool is wobbling more visibly in many states across the nation. Empirically based workload standards for public defenders can help to ensure that caseloads are manageable, enabling attorneys to make good on the constitutional guarantee of effective assistance of counsel and helping to preserve the integrity of the criminal justice system.

Recommended Reading and References

Ake v. Oklahoma (1985) 470 US 68

American Bar Association Standing Committee on Legal Aid and Indigent Defendants (2004) Gideon's broken promise: America's continuing quest for equal justice. American Bar Association, Chicago

American Bar Association (2007) Model Rules of Professional Conduct, Chicago, IL

Argersinger v. Hamlin (1972) 407 US 25

Betts v. Brady (1942) 316 US 455 472

Criminal Justice Standards Committee American Bar Association (1993) ABA standards for criminal justice: def funct 3rd Edn, Chicago, IL

Davies ALB, Worden AP (2009) State politics and the right to counsel: a comparative analysis. Law Soc Rev 43:187–220

DeWolfe PB (2011) Fiscal year 2011, annual report with strategic plan Maryland office of the public defender. http://www.opd.state.md.us/Index%20Assets/Annual%20Report%20FY2011.pdf

Douglas v. California (1963) 372 US 353

Gideon v. Wainwright (1963) 372 US 335

Halbert v. Michigan (2005) 545 US 605

Hansen RA, Hewitt WE, Ostrom BJ, Lomvardias C (1992) Indigent defenders get the job done and done well. National Center for State Courts, Williamsburg

In re Gault (1967) 387 US 1

Johnson v. Zerbst (1938) 304 US 458

Justice Policy Institute (2011) System overload: the costs of under-resourcing public defense July 2011

Kleiman M, Lee CG (2010) Virginia indigent defense commission: attorney and support staff workload assessment. National Center for State Courts, Williamsburg

Langton L, Farole D Jr (2010) State public defender programs, 2007. Bureau of Justice Statistics September, NJC 228229

Lefstein N (2011) Excessive public defense workloads: are ABA standards for criminal justice adequate? Hast Const Law Q 38(4):949–982

Lefstein N, Spangenberg RL (2009) Justice denied: America's continuing neglect of our constitutional right to counsel: Rep of the National Right to Couns Comm

McIntyre L (1987) The Chicago public defender. University of Chicago: University of Chicago Press, Chicago

National Legal Aid and Defender Association (2004) In defense of public access to justice: an assessment of trial-level indigent defense services in Louisiana 40 years after Gideon. Washington, DC

National Legal Aid and Defender Association (2008) Evaluation of trial-level indigent defense systems in Michigan. National Legal Aid and Defender Association, Washington, DC

Ostrom BJ, Kleiman M, Ryan C (2005) Maryland attorney and staff workload assessment, 2005. National Center for State Courts, Williamsburg

Platt A, Pollock R (1974) Channeling lawyers: the careers of public defenders. Issues Criminol 9(1):1–31

Powell v. Alabama (1932) 287 US 45

Rothgery v. Gillespie County (2008) 554 US 191

Standing Committee on Legal Aid and Indigent Defendants, American Bar Association (2002) ABA ten principles of a public defense delivery system

Steelman DC, Tallarico SK, Kleiman M, Fanflik PL (2007) A workload assessment study for the New Mexico trial court judiciary, New Mexico district attorneys' offices, and the New Mexico public defender department. National Center for State Courts, Williamsburg

Strickland v. Washington (1984) 466 US 668

Vagenas GN, Bethke JD, Dattel M, Desilets R, Newhouse DJ, Spangenberg R (2009) Indigent defense chapter 11 in the state of criminal justice. American Bar Association Criminal Justice Section

Washington Defender Association WDA standards for public defense services standard three: caseload limits and types of cases. http://dev21.npowerseattle.org/wda/resources/standards/wda-standards-for-public-defense-services/standard-three-caseload-limits-and-types-of-cases/

Wice PB (2005) Public defenders and the American justice system. Praeger, Westport

Williams PH, Bost WM Jr (1971) The assigned counsel system: an exercise of servitude? Miss Law J 42:32–59

Public Housing and Crime Patterns

Elizabeth Griffiths
School of Criminal Justice, Rutgers University,
Newark, NJ, USA

Synonyms

Subsidized housing

Overview

Since the Department of Housing and Urban Development (HUD) initiated programs to address the shortage of low-income housing in the *U.S. Housing Act of 1937*, the provision of housing assistance to the needy has taken a variety of forms from high-density project-based developments to smaller scale scattered-site complexes and tenant-based voucher assistance. Regardless of form, central concerns around the association between low-income, subsidized housing and local patterns and trends in crime tend to dominate the public imagination. The extant research shows a moderately strong positive relationship between the location of public housing projects and crime hotspots in urban centers. Some studies identify a distance-decay effect whereby rates of delinquency, crime, and violence diminish with distance from the housing complex whereas other research finds crime and violence to be high but relatively contained within the project. There is additional evidence suggesting that residence in public housing can elevate both levels of offending for young male inhabitants and risk of criminal victimization for all tenants. To remediate the concentration of poverty and elevated crime risks associated with large-scale public housing projects located in major American cities, various federal initiatives, including Section 8, MTO, and HOPE VI programs, have gained support in recent decades. Current controversies over the importation of crime into destination communities, ostensibly at the hands of former project residents, serves as a new and promising avenue of scholarly research.

Crime and Public Housing

Scholarly literature investigating the relationship between densely concentrated public housing and crime provides considerable evidence of at least some positive association at both the neighborhood and individual levels-of-analyses. First, there is a moderately strong relationship between the location of public housing developments and crime hotspots in urban areas, although the size of the effect varies by the scale of housing and the number and type of control variables included in the analysis (Galster et al. 2002; McNulty and Holloway 2000; Suresh and Vito 2007). Second, opportunities for involvement in gang violence and drug sales, among other kinds of offending, have been found to be more readily available to youth who reside in public housing developments, compared to those who live outside (Popkin et al. 2000; Venkatesh 2000). In fact, the *same* youth report reductions in criminal involvement after leaving public housing; for example, arrests for violent crime and, in particular, robbery decline by half after juveniles move from project-based public housing into low-poverty neighborhoods (Ludwig et al. 2001; see also Kling et al. 2005a). Finally, public housing residents experience elevated levels of criminal victimization relative to their nonpublic housing counterparts, prompting significant and credible fear of crime (Griffiths and Tita 2009; Holzman et al. 2001; Kling et al. 2005b). Taken together, these findings suggest that twentieth century federal housing policy effectively anchored the poor in highly dangerous environments and isolated them from the economic opportunities that traditionally serve as the building blocks of social and residential mobility.

Part of the rationale for recent shifts away from project-based toward scattered-site and especially tenant-based housing assistance is founded on the premise that the dispersion of public housing tenants across city neighborhoods will deconcentrate urban poverty and dilute the negative effects of related social and economic disadvantages. The logic inspiring the construction of new mixed-income communities on the sites of demolished developments is reflected more broadly in tenant-based housing policies. According to the tenant-based narrative, dispersing public housing tenants across more advantaged neighborhoods and attracting market-rate owners and renters to newly constructed units on the sites of former distressed developments should both deconcentrate poverty and reshape cities to reflect a patchwork of thriving mixed-income communities. The resulting redistribution of low-income families is expected to alleviate the deleterious causes and consequences of concentrated poverty (Massey and Kanaiaupuni 1993). As yet, however, little is known about whether and how tenant-based housing assistance affects the redistribution of crime across the city.

A Brief History of Federal Housing Policy

Federal policies on housing for low-income families in the United States can be distinguished into three distinct eras, each associated with a different ideology and approach to the provision of subsidized housing. Beginning in the 1930s, the first phase "mainly helped two-parent families displaced temporarily by the Depression or in need of housing following the end of World War II" (Wilson 2000, p. ix). More specifically, the *U.S. Housing Act of 1937* integrated housing subsidies into the social welfare system and the *Servicemen's Readjustment Act of 1944* (commonly known as the G.I. Bill) made homeownership loans and access to subsidized rental housing available to returning veterans. Families resided in public housing units for brief periods as a temporary respite and the mix of tenants varied considerably in economic circumstance as a consequence of flexible income ceilings. In this stage, public housing was conceived of, and operated as, "transitional" housing.

The second era of housing policy, targeting slum clearance and urban renewal, can be traced to the passage of the *Housing Act of 1949* supported by President Truman. In this stage, extremely low-income families, many of whom were headed by single mothers, were dislocated from economically disadvantaged crime-ridden neighborhoods marked for slum clearance to high-density public housing developments. In light of the growing population in need of low-income housing, local Housing Authorities in major US cities lowered income ceilings to meet housing demand for the most economically destitute. Thus, eligibility requirements for, and accessibility to, project-based housing tightened considerably. Critics of this second stage of housing policy contend that project-based public housing contributed to a host of urban problems including the concentration of extreme poverty in neighborhoods with public housing, the social and physical isolation of public housing tenants, ineffective and inefficient management of funds for regular maintenance of existing public housing structures, vulnerability to the infiltration of gangs in developments, and elevated levels of crime and violence in public housing communities (Fagan and Davies 2000; Griffiths and Tita 2009; Popkin et al. 2000; Venkatesh 2000). These developments no longer served as a temporary or transitional stopgap, but instead stranded extremely poor residents in rapidly deteriorating social and physical environs. According to critics, the strict eligibility requirements likewise served as disincentives toward self-sufficiency and indirectly encouraged intergenerational tenancy.

Opposition to the siting of these developments was vociferous in well-organized communities of homeowners who feared erosion in property values and the influx of crime (Galster et al. 2002; Santiago et al. 2003). As such, large-scale, high-density public housing developments tended to be clustered within less desirable areas of major cities, where opposition was weak or nonexistent. Indeed, various court decisions over the past few decades, including *Thompson v HUD* (1994), confirm that cities like Baltimore, "with the approval of HUD, acted in concert over many decades to create a deeply segregated system of public housing, with project siting decisions largely driven by community opposition in white neighborhoods, in the context of a central city housing authority with limited jurisdiction over housing outside its own city limits" (Poverty & Race Research Action Council 2005, p. 1). In locating projects in close proximity to one another, tenants were physically isolated from basic amenities by major thoroughfares, were racially and economically segregated, and were socially marginalized from their nonsubsidized counterparts.

Amid mounting concerns over the physical deterioration of buildings and infrastructure, substantial health and safety concerns surrounding the disrepair of tenant units and semipublic spaces (especially hallways, stairwells, and elevator shafts), security concerns cultivated by gangs and gang-related violence, police "sweeps" and "swarms" that represented civil liberties violations, the mismanagement of resources that undermined public trust, and intense fear of crime among residents that thwarted collective efforts at informal social control, a new era of federal public housing policy has once again transformed the administration of housing assistance in the United States (Popkin et al. 2000; Venkatesh 2000). Beginning in 1961, an addition of Section 23 to the *Housing Act* (entitled the Leased Housing Program) allowed qualified low-income renters to reside in privately owned apartment buildings maintained by their local Housing Authority. The building's owner would receive some portion of the market rent from the tenant and the remainder from the Housing Authority itself. In the past 20 years, the shift away from project-based assistance toward tenant-based assistance in the form of alternative programs like Section 8 housing (or, more recently, Housing Choice Vouchers) has grown markedly with the introduction of programs like Housing Opportunities for People Everywhere (HOPE VI) and Moving to Opportunity (MTO).

In 1992, HUD unveiled its HOPE VI vision, originally known as the Urban Revitalization Demonstration (Katz 2009). The blueprint for revitalizing public housing communities under this program would center on a series of major reforms suggested in *The Final Report of the National Commission on Severely Distressed Public Housing*. First, the Commission recognized the need for capital investment in rehabilitating and modernizing public housing stock that had fallen into disrepair, but was not beyond salvaging. Second, the Commission recommended demolishing the 86,000 "severely distressed" public housing units in which residents: (a) lack ready access to social services, support programs, and employment opportunities; (b) live in physically deteriorated buildings; (c) suffer from formal and informal mismanagement of the environment; and (d) reside in developments and neighborhoods marked by excessive crime, drug crime, and violent crime rates relative to the citywide averages. Finally, the Committee emphasized the necessity of access to social services, employment opportunities, and relocation support for those displaced from distressed public housing projects slated for demolition.

Cities such as Atlanta, Baltimore, Boston, Chicago, Los Angeles, and New York – areas in which some of the most distressed public housing developments are located – have been extremely active in realizing the most recent re-visioning of federal housing assistance policy. In various ways, local Housing Authorities in these and other urban centers have aggressively pursued HOPE VI revitalization grants to demolish severely distressed public housing developments, participated in the Moving to Opportunity program which offered Section 8 housing vouchers and housing counseling to 4,600 families who volunteered to relocate to neighborhoods with less than 10 % in poverty (Briggs et al. 2010), and explored ways to expand access to Housing Choice Voucher programs. This movement away from project-based assistance toward tenant-based assistance is intended to benefit public housing residents by offering mobility out of highly disadvantaged crime-ridden communities while, at the same time, serving to deconcentrate poverty, "cool" crime hotspots in urban areas, and encourage the establishment of mixed-income communities.

Crime and Project-Based Developments

Both qualitative and quantitative studies on crime in and around large and densely populated public housing projects have shown elevated levels of property crimes, violent crimes, and drug-related offenses, even compared to surrounding and generally disadvantaged neighborhoods (Griffiths and Tita 2009; Kling et al. 2005b; Popkin et al. 2000; Venkatesh 2000; Weatherburn et al. 1999). Griffiths and Tita (2009) demonstrate, for example, that distressed public housing complexes in Southeast Los Angeles neighborhoods such as

Watts have markedly higher homicide rates relative to surrounding disadvantaged neighborhoods. In some cases, the high concentrations of crime on public housing properties have been shown to diffuse outward from those developments; McNulty and Holloway (2000) find evidence of distance-decay "spread effects" of crime from public housing developments in Atlanta, GA, to the surrounding neighborhoods and Fagan and Davies (2000) report a significant concentration of violent crimes within a 200 yard radius of public housing in Brooklyn, NY. From these patterns, it is clear that crime is concentrated in and around the sites of high-density public housing in major cities. Even when distance-decay effects are identified, however, researchers cannot conclude that public housing residents are necessarily responsible for the elevated crime in the neighborhoods surrounding their developments (Griffiths and Tita 2009).

Although some studies show disparities in offending behavior between public housing residents and their nonpublic housing counterparts, these differences are generally not large. For example, Ireland and colleagues (2003) find that offending rates for youth in public housing in Rochester, NY, are not statistically different from the offending rates of their counterparts outside of public housing, yet public housing youth in Pittsburgh exhibit somewhat higher rates of violent offending. Kids who moved from public housing developments to low-poverty neighborhoods as part of the Moving to Opportunity program were shown to reduce their offending behavior (Ludwig et al. 2001). Nevertheless, scholars have difficulty distinguishing selection effects from causal effects when differences arise (Sampson 2008). That is, it is not clear whether crime-prone individuals are allocated to public housing (Weatherburn et al. 1999) akin to a selection effect or whether the architectural and design features of public housing projects, including physical size and the number of apartments sharing common entries, simply create more opportunities to offend for both local and nonlocal residents (Newman 1972).

On the whole, research suggests that levels of crime in public housing can be explained less by the criminality of tenants and more by the criminogenic opportunities inherent in the design and management of these developments. Public housing is a "crime attractor" such that it provides access to a "mass of prospective victims and/or erode[s] the collective efficacy of the neighborhood" (Galster et al. 2002, p. 291). Indeed, the literature has clearly established that victimization risk is significantly higher for residents of public housing relative to those who reside in private market housing and, therefore, neighborhood crime problems differentially burden public housing tenants (Dunworth and Saiger 1994; Griffiths and Tita 2009; Raphael 2001). Combined, the extreme concentration of poverty as a consequence of the spatially proximate siting of developments, the physical and social isolation of residents, and low levels of formal and informal social control are significant contributors to the elevated rates of crime and violence in housing developments.

Crime and Scattered-Site Developments

Scattered-site supportive housing, which is term typically used to describe developments with fewer than 15 units, can take the form of small apartment buildings purchased by a local Housing Authority, group homes, homeless shelters, or other small-scale subsidized housing for special needs populations such as the elderly, the infirm, substance abusers, or those suffering from mental disease or defect. Such housing has little effect on property values in the surrounding neighborhoods (Hogan, 1996). Moreover, residents in scattered-site housing report similar neighborly interactions, willingness to give and receive aid from neighbors, and embeddedness in the local community as do residents of clustered public housing (Kleit 2001). In the very few studies that have specifically examined the effects of scattered-site supportive housing on neighborhood crime trends, the authors find virtually no evidence of an increase in neighborhood crime after the opening of these small-scale subsidized developments (Galster et al. 2002, 2003; Santiago et al. 2003). The single exception is the finding of a statistically significant increase in total crime and violent crime within 500 ft of supportive housing sites in Denver, CO, but only for slightly larger developments that housed more than 53 residents (Galster et al. 2002). Therefore, the introduction of scattered-site developments into neighborhoods does not appear to increase proximate crime rates, particularly when these buildings are less dense and located in low-poverty areas with pre-construction levels of crime that approximate citywide averages.

Crime and Tenant-Based Housing

Current literature on tenant-based housing assistance suggests that relocating from densely populated high-rise public housing to voucher-subsidized private rentals has null or positive individual-level effects on children and adults, particularly among those who participated in the MTO Demonstration. For example, adolescent involvement in violent offending tends to decline after a move (Kling et al. 2005a), MTO children attend schools with higher test scores and pass rates, MTO parents report improvements in physical and mental health as well as reduction in stress, and adults perceive their new neighborhoods as safer (Goering et al. 2003).

Outcomes for HOPE VI families are more mixed. For example, Clampet-Lundquist (2010) finds that concerns over crime and perceptions of neighborhood safety change, but do not necessarily diminish, among families relocated from distressed public housing developments. More specifically, even those voucher recipients taking up residence in lower crime neighborhoods report fear as a consequence of the loss of social ties and lack of "physical and social protections" that had been available in their previous public housing projects. Indeed, residents perceive themselves to have been more physically secure from violent crime when they resided in upper-floor apartments of high-rises compared to the single-family homes in their new neighborhoods (Clampet-Lundquist 2010). Residents who personally observed offending in the previous development also tended to know the perpetrators. Consequently, the anonymity of destination neighborhoods increases residents' insecurities and perceptions of risk.

Some research illustrates that short-term gains in educational benefits for children of voucher recipients are not borne out in the long term (DeLuca and Dayton 2009) and others find few positive *or* negative effects of mobility on student achievement (Jacob 2004). Part of the explanation for these null effects is related to the characteristics of receiving neighborhoods. According to Jacob (2004, p. 251), "a large proportion of families did not take advantage of the relocation opportunity provided by public housing closings to move to a substantially different neighborhood, and even those children who did move to substantially better neighborhoods did not end up in significantly better schools."

The move to tenant-based assistance and related housing mobility programs has triggered much research on the welfare of and outcomes for affected families, but less attention is focused on how the in-migration of voucher recipients affects conditions in the receiving neighborhoods. In one of the few studies on this topic, Suresh and Vito (2007) suggest that the high assault rates characteristic of densely populated neighborhoods comprised of large-scale public housing developments in Louisville, KY, are gradually displaced into the destination neighborhoods of housing voucher recipients over time. The authors intimate that assault hotspots shift in a pattern consistent with the relocation of residents, particularly when the destination neighborhoods are low-income areas. Based on their study, however, Suresh and Vito (2007) cannot determine whether the victims and offenders involved in assaults in destination communities are actually voucher recipients. In a more recent attempt to disentangle how the in-migration of voucher recipients affects neighborhood crime rates, Van Zandt and Mhatre (2009) arrive at a different conclusion. They suggest that rather than bringing crime with them, voucher families' residential choices are constrained to higher crime neighborhoods in Dallas, TX. Consequently, any relationship between the presence of voucher recipients and neighborhood crime rates is driven by inadequate options for the use of vouchers in safer and more affluent neighborhoods.

Key Controversies

Two controversies presently dominate the literature on public housing and neighborhood crime patterns. The first is a theoretical question that situates researchers into one of two camps – those who subscribe to "types of people" explanations for aggregate levels and patterns in crime and those who subscribe to "types of places" explanations. Whether one views changes in crime rates as a function of changes in the residential population or the characteristics of place that attract offenders to the environs influences how such changes are interpreted. Second, the move to tenant-based housing assistance has spawned questions and concerns over both the consequences for destination neighborhoods and the restrictions on residential opportunities for Housing Choice voucher recipients. These respective controversies are discussed in turn below.

For those who subscribe to the types of people model, changes in crime rates that coincide with shifts in the demographic and socioeconomic composition of neighborhoods can be attributed to the criminality of newcomers. More specifically, when dangerous people move into a new area, law-abiding residents become easy targets for victimization. Rosin's (2008; but see Briggs and Drier 2008) controversial *Atlantic Monthly* article, suggesting that increases in neighborhood crime rates were associated with the movement of subsidized families (i.e., Housing Choice Voucher recipients) into the community, intimated that former public housing residents were "criminals. . .penetrating once-safe suburban neighborhoods" in Memphis, Tennessee (Van Zandt and Mhatre 2009, p. 1). Were it accurate, the speculation that those receiving tenant-based housing assistance are somehow more prone toward criminal activity than are market-rate tenants could be said to represent a "direct link" between subsidized housing and crime (Galster et al. 2002).

By contrast, "types of places" explanations for neighborhood crime trends focus on how neighborhood conditions undermine collective efforts to control public space and provide opportunities for offending. These socially disorganized and "criminogenic" environments are vulnerable to crime because there is little formal guardianship, efforts to generate social control are ineffective, and the communities lack collective efficacy. Importantly, offenders need not be local residents. In this case, scholars treat changes in levels and patterns of crime as contextual effects, in lieu of assuming that residents in high crime neighborhoods are responsible for the offenses that occur there. Here, the link between subsidized housing and neighborhood crime rates is "indirect" and public housing residents "may [simply] be more difficult for the community to enlist as. . . instruments of collective efficacy" (Galster et al. 2002, p. 293).

The second controversy presently unresolved in the literature involves the freedom with which Housing Choice voucher recipients can select destination neighborhoods. Under the logic of tenant-based assistance programs, residents ostensibly experience expanded choice in residential location such that they can choose affordable private rental units in lower poverty neighborhoods, all across the city. The Department of Housing and Urban Development requires that voucher recipients reside in rental units that are at or below HUD-determined Fair Market Rents. One of the standard performance indicators for local Housing Authorities is related to their implementation of "policies to encourage voucher holders to search outside of areas of poverty and minority concentration and to encourage owners of units in lower poverty or minority areas to rent to voucher holders" (Devine et al. 2003, p. 25). Voucher recipients' choices are restricted, however, by the preferences of private landlords to participate in the Housing Choice Voucher (HCV) program. And, unfortunately, landlord participation itself may be conditioned by the characteristics of the neighborhood and its desirability for market-rate renters and homeowners. For example, Wyly and DeFilippis (2010, p. 65) maintain that

> landlords participate only if the mix of benefits and drawbacks (stability of occupancy and rental income, vs. rules and paperwork) is better than the private rental market in the neighborhood; there has always been concern that the shift to vouchers will simply reconcentrate assisted households into a new set of bad neighborhoods.

Unlike the Moving to Opportunity Demonstration, which relocated public housing families into very specific and advantaged neighborhood contexts, the HCV program relies on a looser set of conditions, dependent on private landlords' willingness to rent to HCV holders and inclusionary zoning policies (where these county and municipal zoning ordinances exist).

On the whole, voucher recipients make short-distance moves to neighborhoods in which concentrated poverty and residential racial segregation are only slightly lower than their previous housing project (Wang and Varady 2005), although the long-term outcomes are far superior when families receive substantial housing relocation counseling and are placed in socio-economically prosperous neighborhoods (as was the case for those who participated in the Gatreaux and MTO programs; see Keels et al. 2005). Indeed, the deconcentration of poverty expected to result from tenant-based housing assistance policies that disperse low-income residents to subsidized private rental units will be undermined if voucher recipients tend to reconcentrate in new neighborhoods. In one of few studies examining this issue, spatial econometric models of HCV recipients' destination neighborhoods in New York City show

> no support for the claim that tenant-based vouchers promote poverty deconcentration. Even after accounting for racial segregation and spatial autocorrelation, vouchers have a stronger link to local poverty rates than all other types of federal low-income housing assistance – including the much-aligned category of traditional public housing (Wyly and DeFilippis 2010, p. 82).

If crime rates are already higher and socioeconomic conditions are already disadvantaged in the destination neighborhoods of HCV recipients, it would be fallacious to assume that an influx of HCV families is responsible for changes in neighborhood crime.

There is undoubtedly a complex relationship between public housing and crime that defies easy explanation. Historically, crime and violence rates tended to be exceptionally high in densely concentrated high-rise and garden-style developments, spawning NIMBY (Not In My Backyard) opposition to siting decisions. From this relatively consistent empirical relationship between crime and the location of public housing developments in American cities, however, scholars cannot conclude that public housing residents are necessarily more dangerous or inherently more likely to be involved in offending than their nonpublic housing counterparts. Instead, the physical architecture of these developments and the failure of local housing authorities to adequately maintain the units, buildings, and premises helped to undermine collective efficacy, social control, and social capital that is necessary to forestall crime and protect the public space. With the transformation in housing policy from largely project-based to predominantly tenant-based assistance, new questions about the fate of former residents' destination neighborhoods are expected to gain attention in the criminological literature.

Related Entries

- ▶ Crime Prevention through Environmental Design
- ▶ Designing Products Against Crime
- ▶ Disadvantage, Disorganization and Crime
- ▶ Housing Choice Vouchers and Escaping Neighborhood Crime
- ▶ Outdoor Serious Violence: The Role of Place
- ▶ Poverty, Inequality, and Area Differences in Crime
- ▶ Social Disorder and Physical Disorder at Places

Recommended Reading and References

Briggs XdS, Popkin SJ, Goering J (2010) Moving to opportunity: the story of an American experiment to fight ghetto poverty. Oxford University Press, Oxford

Briggs XS, Drier P (2008) Memphis murder mystery? No, just mistaken identity. Shelterforce: The Journal of

Affordable Housing and Community Building July 22, 2008, Retrieved from www.shelterforce.org/article/special/1043/

Clampet-Lundquist S (2010) ‘Everyone Had Your Back’: social ties, perceived safety, and public housing relocation. City Community 9(1):87–108

DeLuca S, Dayton E (2009) Switching social contexts: the effects of housing mobility and school choice programs on youth outcomes. Annu Rev Sociol 35:457–491

Devine DJ, Gray RW, Rubin L, Taghavi LB (2003) Housing choice voucher location patterns: implications for participants and neighborhood welfare. U.S. Department of Housing and Urban Development, Office of Policy Development and Research, Division of Program Monitoring and Research, Washington, D.C

Dunworth T, Saiger A (1994) Drugs and crime in public housing: a three-city analysis. U.S. Department of Justice, Office of Justice Programs, Washington, DC

Fagan J, Davies G (2000) Crime in public housing: two-way diffusion effects in surrounding neighborhoods. In: Goldsmith V, McGuire PG, Mollenkopf JH, Ross TA (eds) Analyzing crime patterns: frontiers of practice. Sage Publications, Thousand Oaks, pp 121–135

Galster G, Pettit K, Santiago A, Tatian P (2002) The impact of supportive housing on neighborhood crime rates. J Urban Aff 24(3):289–315

Galster GC, Tatian PA, Santiago AM, Pettit KLS, Smith RE (2003) Why not in my backyard?: Neighborhood impacts of deconcentrating assisted housing. Center for Urban Research, Rutgers University, New Brunswick

Goering J, Feins JD, Richardson TM (2003) What have we learned about housing mobility and poverty deconcentration? In: Goering J, Feins JD (eds) Choosing a better life?: evaluating the moving to opportunity social experiment. Urban Institute Press, Chicago, pp 3–36

Griffiths E, Tita G (2009) Homicide in and around public housing: is public housing a hotbed, a magnet, or a generator of violence for the surrounding community? Soc Probl 56(3):474–493

Hogan JB (1996) Scattered-site housing: characteristics and consequences (Report no. HUD-7470). U.S. Department of Housing and Urban Development, Office of Policy Development and Research, Washington D.C

Holzman HR, Hyatt RA, Dempster JM (2001) Patterns of aggravated assault in public housing: mapping the nexus of offense, place, gender and race. Violence Against Women 7:662–684

Ireland TO, Thornberry TP, Loeber R (2003) Violence among adolescents living in public housing. Criminol Public Policy 3(1):3–38

Jacob BA (2004) Public housing, housing vouchers, and student achievement: evidence from public housing demolitions in Chicago. Am Econ Rev 94(1):233–258

Katz B (2009) The origins of HOPE VI. In: Cisneros H, Engdahl L (eds) From dispair to hope: HOPE VI and the new promise of public housing in America’s cities. Brookings Institution Press, Washington D.C

Keels M, Duncan GJ, Deluca S, Mendenhall R, Rosenbaum J (2005) Fifteen years later: can residential mobility programs provide a long-term escape from neighborhood segregation, crime, and poverty? Demography 42(1):51–73

Kleit RG (2001) Neighborhood relations in suburban scattered-site and clustered public housing. J Urban Aff 23(3–4):409–430

Kling JR, Ludwig J, Katz LF (2005a) Neighborhood effects on crime for female and male youth: evidence from a randomized housing voucher experiment. Q J Econ 120(1):87–130

Kling JR, Liebman JB, Katz LF (2005b) Bullets don’t got no name: consequences of fear in the Ghetto. In: Weisner TS (ed) Discovering successful pathways in children’s development: mixed methods in the study of childhood and family life. The University of Chicago Press, Chicago, pp 243–281

Ludwig J, Duncan G, Hirschfield P (2001) Urban poverty and juvenile crime: evidence from a randomized housing-mobility experiment. Q J Econ 116(2):655–680

Massey DS, Kanaiaupuni SM (1993) Public housing and the concentration of poverty. Soc Sci Q 74(1):109–122

McNulty TL, Holloway SR (2000) Race, crime, and public housing in Atlanta: testing a conditional effect hypothesis. Soc Forces 79(2):707–729

Newman O (1972) Defensible space. Macmillan Press, New York

Popkin SJ, Gwiasda VE, Olson LM, Rosenbaum DP, Buron L (2000) The hidden war: crime and the tragedy of public housing in Chicago. Rutgers University Press, New Brunswick

Poverty & Race Research Action Council (2005) An analysis of the *Thompson v HUD* Decision. http://www.prrac.org/pdf/ThompsonAnalysis.pdf

Raphael J (2001) Public housing and domestic violence. Violence Against Women 7(6):699–706

Rosin H (2008) American murder mystery. The Atlantic Monthly 302(1):40–54

Sampson RJ (2008) Moving to inequality: neighborhood effects and experiments meet social structure. Am J Sociol 114(1):189–231

Santiago AM, Galster GC, Pettit KLS (2003) Neighbourhood crime and scattered-site public housing. Urban Stud 40(11):2147–2163

Suresh G, Vito GF (2007) The tragedy of public housing: spatial analysis of hotspots of aggravated assaults in Louisville, KY (1989–1998). Am J Crim Just 32(1–2):99–115

Van Zandt S, Mhatre P (2009) The effect of housing choice voucher households on neighborhood crime: longitudinal evidence from Dallas. Sustainable Housing Research Unit (SHRU) Working Paper 09-01. College of Architecture, Texas A & M University. www.urbanplanningblog.com/papers/HCVCrime 2008.pdf

Venkatesh SA (2000) American project: the rise and fall of a modern Ghetto. Harvard University Press, Cambridge

Wang X, Varady DP (2005) Using hot-spot analysis to study the clustering of section 8 housing voucher families. Housing Stud 20(1):29–48

Weatherburn D, Lind B, Simon K (1999) 'Hotbeds of Crime': crime and public housing in Sydney. Crime Delinq 45(2):256–272

Wilson WJ (2000) Foreword. In: Venkatesh SA (ed) American project: the rise and fall of the modern Ghetto. Harvard University Press, Cambridge

Wyly E, DeFilippis J (2010) Mapping public housing: the case of New York City. City Community 9(1):61–86

Public Opinion About Sentencing

Karen Gelb
Sentencing Advisory Council, Melbourne,
Victoria, Australia

Overview

Confidence in the courts is critical to the legitimacy and effective functioning of the criminal justice system. However, surveys measuring levels of public confidence in the system have found that public trust and confidence are at critically low levels around the world. This is especially so for the courts and sentencing.

Low levels of confidence in the courts have been of major concern for governments around the world. Within this context, a substantial body of literature has developed that attempts to understand the drivers of public confidence. There are five main themes to this body of research:

- People have little confidence in the courts.
- In the abstract, people believe that sentences are too lenient.
- People have little accurate knowledge of crime and the criminal justice system.
- The mass media is the primary source of information on crime and justice issues.
- When given more information, people become less punitive.

While there is widespread agreement in the literature that people's knowledge of crime and justice is a primary driver of their attitudes to the courts and sentencing, there is less agreement about the practical implications of this finding in terms of improving public confidence. That is, if lack of knowledge about crime and justice is closely related to lack of confidence in the courts, can public confidence be improved by increasing public understanding?

Some researchers believe that the answer to improving public confidence lies in informing and educating the public, often via bodies such as sentencing councils. Others, however, see this approach as doomed to fail in the absence of more widespread opportunities for public deliberation, debate and discussion. While informing and educating is aimed squarely at the intellectual aspect of public attitudes, it is often more emotive factors that drive attitudes. Ignoring the affective to focus on the effective thus only addresses part of the story of the relationship among attitudes, knowledge and confidence in the courts.

Introduction

Confidence in the courts is critical to the legitimacy and effective functioning of the criminal justice system. All government institutions require some level of support from the community, but confidence in the courts is arguably one of the most important for society. The public is central to the legitimacy and function of the criminal justice system: without confidence, people may be less willing to report crimes or to participate in the system as jurors or witnesses.

Surveys measuring levels of public confidence in the criminal justice system have found that public trust and confidence are at critically low levels around the world (Roberts and Hough 2005). In particular, research comparing confidence across agencies within the criminal justice system has consistently found that people have high confidence in the police but low confidence in the courts and prisons (Hough and Roberts 2004). When asked why they have little confidence in the courts, people typically cite lenient sentencing.

Some researchers have described this as a crisis of confidence in the courts. While it is difficult to determine if any given level of public confidence is objectively acceptable, there is consensus among researchers and policymakers in many countries that this lack of public confidence is a cause for some concern. The potential implications of this crisis have led to varied attempts to promote public confidence in the system as a whole, but particularly in the courts.

Background

Public opinion polls have become staple fare for governments wishing to measure community attitudes, and the area of criminal justice is no different. Along with the increased use of such polls since the 1970s has come a heightened sensitivity to the views of the public, as well as an increased role for public perceptions in the development of government policy. This is clearly seen in the rise of the public voice in criminal justice policy. That is, criminal justice systems administered by experts, independent of democratic pressures, have been replaced by systems that are more populist in style and punitive in substance. Public discourse on crime has become more emotive and emotional, with governments responding to public anxieties by seeking to give greater voice and effect to public demands.

To some extent, this more prominent public role reflects an element of penal populism: the development of penal policies that are based on their appeal to the public (and thus their ability to win votes) with little regard for their effectiveness in reducing crime or promoting justice (Roberts et al. 2003).

A penal populist approach is based upon the idea that the public is fed up with crime and with the perceived leniency of the criminal justice system. In an era of penal populism, where the vast majority of people learn about crime and the criminal justice system from the media, a combination of sensationalist reporting practices in the media and populist political responses to moral panics has resulted in widespread myths and misconceptions about sentencing in many Western countries.

Based on the perceived public concerns, penal populists take a firm “tough-on-crime” stance, using imprisonment as a central tool. Penal populism thus provides a framework within which to understand increasing imprisonment rates around the world as well as the proliferation of punitive sentencing policies such as three-strikes legislation, mandatory minimum sentences and sex offender notification laws.

Belief in a punitive public is driven primarily by the results of decades of opinion polls that show that the public believes the criminal justice system, and courts in particular, to be overly lenient. But while governments and the mass media continue to place high credence in the basic opinion poll question of whether sentencing is “too tough, about right, or too lenient” as a way to justify calls for punitive penal policy, academic researchers have repeatedly shown that public opinion on crime and justice issues, and on sentencing in particular, is far more nuanced and complex than such surveys show.

International Perspectives on Public Opinion

Regardless of the methodologies employed, research has consistently shown that public trust and confidence in the courts are especially low and are lower than public confidence in the police. While it may be unreasonable to expect levels of confidence in justice to match those of other institutions such as health care and education, especially given the different (and complex) mandate that the justice system has, the differential in levels of confidence has nonetheless been of major concern for governments around the world.

Within this context, a substantial body of literature has developed that attempts to understand the drivers of public confidence in the courts. There are five main themes to this body of research, each discussed in brief below.

People Have Little Confidence in the Courts

It is clear from the research literature that levels of confidence in the courts should not be expected to be high: one of the most consistent findings in this area is that ratings of satisfaction with, and confidence in, the courts are consistently lower than ratings for police and are objectively rather poor. The main reason for this low rating seems to be twofold: people believe that the courts are too lenient in sentencing offenders, and this perceived lenient sentencing is then held to blame for high rates of crime.

Studies that examine the correlates and causes of confidence in the courts have found few consistent results. One of the findings that does emerge consistently is that people who believe that crime is increasing are likely to be less confident in the courts. This has been linked in the literature with the role of the media in shaping people's understanding of the nature and prevalence of crime.

In their analysis of data from the 1996 British Crime Survey, Hough and Roberts (1998) found that people who thought that sentences were too lenient were more likely to believe that judges were out of touch with society and were doing a poor job. Further analysis showed that people who believed that sentences were too lenient were also those who were particularly inaccurate in their perceptions of crime. The authors concluded that public dissatisfaction with the courts and sentencing is grounded, at least in part, by a lack of knowledge of both sentencing practice and trends in crime. Subsequent sweeps of the British Crime Survey have found identical results, consistently revealing a clear relationship between attitudes to sentencing and confidence in the courts.

More recent studies from other countries have found similar results. For example, in an analysis of survey responses of 1,200 randomly selected people in Victoria, Australia, Gelb (2011) found that the strongest single predictor of confidence in the courts and sentencing was people's levels of punitiveness: the more punitive people were, the less confident they were in sentencing. As with the British Crime Survey data, perceptions of crime played a critical role, with lower confidence in the courts among people who believed that crime had been increasing. Perceptions of crime played both a direct role in predicting confidence, but also an indirect one, with perceptions of increasing crime leading to increased punitiveness, which in turn led to decreased confidence. People who relied on commercial or tabloid media were more likely to have perceived that crime had increased, demonstrating the link between confidence in the courts and the role of the media in shaping people's perceptions of crime.

In the Abstract, People Believe Sentences Are Too Lenient

When representative surveys first came into widespread use, the most common way of measuring public opinion on sentencing was to use the general question of whether sentences are "too tough, about right, or too lenient." This question, in some variant or another, has been used in opinion polls across the world for the last 40 years. And across a variety of Western countries, between 70 % and 80 % of respondents have consistently reported that sentences are too lenient.

On the basis of these survey findings alone, politicians, policy-makers, and the media have concluded that the public is substantially punitive and will therefore support increasingly punitive penal policies. It is from this conclusion that penal populism has developed.

In more recent years, however, this conclusion has been called into question. In particular, researchers have suggested that the finding of a highly punitive public is a methodological artifact – a result of the way in which public opinion has been measured, rather than an accurate representation of what people really think.

Researchers have begun to differentiate between mass public opinion and informed public judgment – between top-of-the-head, immediate public opinion and the kind of reflective, informed public judgment that can develop once people have engaged with an issue, considered it from a number of perspectives, understood the choices that each perspective leads to, and accepted the consequences of those choices. Top-of-the-head responses to simple polling questions represent mass public opinion, as opposed to informed public judgment. While public opinion tends to be volatile and have little internal consistency with other views and beliefs held by the respondent, public judgment is characterized by firmness of opinion (changing little over time) and by the degree of consistency between this view and others held by the respondent.

In order to measure informed public judgment, since the 1980s researchers have gone beyond the single question opinion poll to include additional questions that can clarify and further explain the apparent harshness of public attitudes. In particular, researchers have given respondents more detailed case studies or vignettes to consider and have provided relevant contextual information (such as the cost of imprisonment or re-offending rates) to help respondents frame their answers. This approach aims to measure informed, thoughtful public judgment, rather than the top-of-the-head mass public opinion that is derived from the abstract "too tough-too lenient" question (Green 2006).

In this way the research has attempted to address the methodological limitations of using a single abstract question to measure complex and nuanced public attitudes.

People Have Little Accurate Knowledge of Crime and the Criminal Justice System

Large-scale surveys of public opinion about crime and punishment in the United States, United Kingdom, Canada, Australia, and New Zealand have all shown that the public has very little accurate knowledge about crime and the criminal justice system. Of particular relevance to attitudes about sentencing are findings that show that people have extensive misperceptions about the nature and extent of crime and about court outcomes.

Consistent results from many of the studies in this field show that people tend to:

- Perceive crime to be constantly increasing, particularly crimes of violence, regardless of actual trends in crime
- Overestimate the proportion of recorded crime that involves violence
- Overestimate the percentage of offenders who re-offend
- Underestimate the severity of sentencing practices (e.g., the incarceration rate)

Despite this lack of knowledge, people nonetheless have strongly held opinions about crime and justice issues. In fact, representative surveys have shown that it is those who have the lowest levels of knowledge who also hold the most punitive views. For example, Doob and Roberts (1983) found that those who think that sentences are too lenient are more likely to think that crime overall is violent and to underestimate the proportion of offenders convicted of robbery and assault who are sent to prison. Studies of this kind have provided strong evidence that dissatisfaction with perceived sentencing practice is due at least in part to public misperception and misinformation. But these misperceptions are neither random nor unrelated; they reflect a systematic and cohesive view of crime and criminal justice.

The combination of underestimates of the severity of sentencing and overestimates of the severity of offending builds a grossly inaccurate picture that has serious implications for levels of public confidence in the criminal justice system. People who have the least accurate perceptions of the nature and prevalence of crime are also those who are least likely to have a positive view of sentencers. Knowledge of crime is thus intimately linked to confidence in the courts and the criminal justice system in general. Such public misconceptions have led some researchers to conclude that the answer to low levels of public confidence lies in addressing the gaps in public knowledge of the criminal justice system.

The Mass Media Is the Primary Source of Information on Crime and Justice Issues

Most people do not have direct access to first-hand information about the criminal justice system, either through personal experience or from the experience of family and friends. Instead, people tend to learn about crime and the criminal justice system through the mass media, in particular via newspapers. Given the ubiquity and popularity of the mass media (tabloid newspapers in particular), they play an integral role in the construction of both public opinion and the public "reality" of crime.

Different media provide significantly different views of the world. Broadsheet papers tend to focus on government, quoting experts, elites, and interest group representatives. Tabloid newspapers focus instead on crime victims and their families, offering dramatic and personal testimonials as counterpoint to the more professionalized discourse of the broadsheets.

Green (2006) suggests that complex public policy debates are represented by the media in increasingly constricted and emotive terms. Newspaper portrayals of crime stories do not provide a complete and accurate picture of the issue. Papers report selectively, choosing stories, and aspects of stories, with the aim of entertaining more than informing. Tabloid newspapers in particular tend to focus on unusual, dramatic and violent crime stories, in the process painting a picture of crime for the community that overestimates the prevalence of crime in general and of violent crime in particular. Analysis of British Crime Survey data has shown that tabloid readers tend to be more punitive and less knowledgeable about crime and justice than are broadsheet newspaper readers. In addition, the media provide little systematic information about the sentencing process or its underlying principles. Media emphasis on violent crime, a lack of coverage of statistical trends to place incidents in context and a focus on street crime and failures of the system all encourage a public perception that both the volume and seriousness of crime are getting worse. This perception, in turn, undermines confidence in the criminal justice system.

As people are overly influenced by single-case information, people falsely generalize that leniency characterizes the entire sentencing process. The media's focus on violent crime leads to a perception that this type of event is typical, which affects both people's knowledge of the facts about crime as well as their general levels of fear of crime. Both of these in turn have been shown to influence perceptions of leniency in sentencing.

But tabloid journalism alone cannot be responsible for the gap between the practices of the courts and the public's perceptions of sentencing. In addition to media influence, people may be responding to messages about crime and social order (either implicit or explicit) that are conveyed in tandem by law-and-order political rhetoric and the news media. That is, it is the combination of reporting practices and populist political responses to moral panics that has resulted in widespread myths and misconceptions about crime and justice in general and about sentencing in particular.

Thus, public concerns about crime typically reflect crime as depicted in the media, rather than trends in the actual crime rate. This distorted view is then used to support one particular anticrime policy approach – an expanded and enhanced punitive criminal justice system.

When Given More Information, People Become Less Punitive

There is substantial evidence that the public's lack of knowledge about crime and justice is related to the high levels of punitiveness reported as a response to a general, abstract question about sentencing. Based upon the conclusion that increasing the provision of information will decrease levels of punitiveness, many researchers have moved away from simplistic, abstract survey questions to those which provide much more information and detail to people before asking for a response. The crime vignette approach has been used in representative surveys to provide more information about the offence, the offender and the impact on the victim in order to elicit informed public judgment.

In their groundbreaking work, Doob and Roberts (1983) were the first who demonstrated the powerful effect of providing more information on respondents' attitudes. Of respondents who received a brief description of a manslaughter case (akin to the type of information provided in media accounts), 80 % rated the sentence as too lenient, while only 7 % rated the sentence as about right. In comparison, of those who received a more detailed description with information on incident and offender characteristics, only 15 % rated the sentence as too lenient and 30 % said it was about right. Fully 45 % of this group described the sentence as too harsh. Doob and Roberts concluded that, were the public to form opinions from court-based information instead of through the lens of the mass media, there would be fewer calls for harsher sentences.

To extend this analysis, the authors then turned to comparisons of newspaper accounts with court records of sentencing hearings (such as transcripts) to determine if judgments about a case differed based on the different account given. Newspaper accounts led to perceptions that sentences were too lenient, but when given more complete information, people were more content with decisions made by trial judges.

These seminal studies show that sentences described in the media are perceived by most people as being too lenient, while those described in detail in court transcripts are mostly seen as appropriate. Doob and Roberts concluded that caution should be exercised in responding to calls for harsher penalties as a fully informed public could well be quite content with the current level of severity of penalties.

Diamond and Stalans (1989) adopted a similar approach in their comparison of lay and judicial responses to case study vignettes, in which respondents were asked to impose sentences on the same four moderately severe cases in which prison was a possible, but not inevitable sentencing outcome.

Respondents were presented with detailed information about each of the four cases, including information on the nature of the offence and on the offender's background, as well as a video of the sentencing hearing. They were told the range of possible sentencing options legally available for that case and then completed a questionnaire indicating sentencing preferences.

There was no evidence in any of the four cases that judicial sentences were more lenient than the sentences of the lay respondents: judges' sentences in the study were as severe or more severe than those of the lay respondents. Diamond and Stalans concluded that the perception that judges are more lenient than the public is simply a myth.

These findings have been replicated time and again in the decades since, with a large body of research showing that more detailed information is needed in order to measure informed public judgment rather than mass public opinion.

Key Issues in the Field

There is widespread agreement in the field, based on the large body of research that has now accumulated, that a lack of knowledge about crime and the criminal justice system is a significant factor in perpetuating public misperceptions and misunderstanding – that people's knowledge of crime and justice is a primary driver of their attitudes to the courts and sentencing.

There is less agreement, however, about other factors that may underlie levels of confidence. Do demographic factors influence people's attitudes? What about experience with the criminal justice system or previous victimization? How do other criminal justice attitudes relate to people's confidence in the courts?

There is also debate about the practical implications of inaccurate knowledge about crime and justice in terms of improving public confidence. In other words, if lack of knowledge about crime and justice is closely related to lack of confidence in the courts, can public confidence be improved by increasing public understanding?

What Underlies Low Levels of Confidence in the Courts?

Studies that examine the correlates and causes of confidence in the courts have found few consistent results. One of the findings that does emerge consistently is that people who believe that crime is increasing are likely to report that sentencing is too lenient and to be less confident in the courts. This has been linked in the literature with the role of the media in shaping people's understanding of the nature and prevalence of crime. In particular, people who use commercial or tabloid media as their main source of information about crime and justice issues are more likely to have inaccurate knowledge of the nature and prevalence of crime and of the severity of sentencing outcomes. These same people are also more likely to hold punitive views toward offenders and to report lower levels of confidence in the courts.

The prominence of this constellation of factors – confidence in the courts and sentencing, punitiveness, and perceptions of crime – is found throughout the research literature, highlighting the interconnections among knowledge of crime, media use, confidence in sentencing and punitiveness.

Of the demographic factors examined in the research literature, the main consistent finding seems to be that younger people have higher levels of confidence in the courts than older people. This disparity has been found in studies from a variety of countries that measure confidence in different ways.

The role of other demographic factors is far less clear and has been the source of much study. In particular, factors such as gender, income and education have been included in many studies of the drivers of confidence in the courts but with inconsistent results.

For the most part, research on the demographic drivers of confidence has found no gender differences but has found that people with higher incomes and more education report greater confidence in the courts and indeed in other institutions as well. This finding, however, has not remained uncontested, with some studies finding no statistically significant relationship between these demographic factors and confidence, once factors such as knowledge of crime or fear of crime have been considered. The uncertainty of the role of these demographic factors may thus be a methodological issue rather than a substantive one – it may well be that demographic factors have an indirect rather than a direct effect on confidence, acting via more proximal factors such as knowledge of crime. The use of statistically sophisticated methodologies and analyses is required to address this question definitively.

Another factor that has presented inconsistent results in the literature is experience with the courts and its effect on levels of confidence. While earlier research showed that contact with the courts was related to lower levels of confidence, more recent research has resulted in a more sophisticated understanding: the nature of the experience will determine levels of confidence. That is, a person who has been treated with respect and sensitivity – regardless of whether appearing as a witness or serving as a juror – will be more confident in the justice system. It is thus the nature of the contact, rather than the fact of the contact per se, that influences ratings of confidence.

This finding has also been expressed in terms of procedural fairness: people who perceive that their treatment by the court was fair (regardless of actual outcome) are more likely to express confidence in the courts.

How Can Public Confidence in the Courts and Sentencing Be Improved?

Given the consistent and strong relationship that has been demonstrated between lack of knowledge about crime and justice and lack of confidence in the courts and sentencing, an implicit assumption has arisen in the research literature that educating the public is the solution to address low levels of confidence.

To this end, governments and courts concerned with improving public confidence have undertaken a range of strategies aimed at educating the public. Public education and information campaigns, court media liaison officers, accessible websites, large national conferences and public surveys have all contributed to efforts to promote confidence in the administration of justice by engaging more closely with the community.

One of the more formal responses to the desire to engage and inform the community has been the proliferation of sentencing councils and commissions around the world. Such bodies are being used as conduits to improve the exchange of information among governments, the courts and the public.

The Rise of the Sentencing Council

Sentencing councils that have developed over the last decade have arisen within the context of a somewhat paradoxical political and social environment. At the same time as overall crime rates have fallen in many Western countries, the judiciary has been coming under increasing pressure as the public claims a greater voice in sentencing and law-and-order auctions continue to feature in political campaigns. The status of judges and the courts more generally has been eroded over time by media polls and reports that the courts are “soft on crime” and that judges are “out of touch” with their communities.

Within this complex environment, sentencing councils have arisen as a way of engaging with the public and balancing the interests of the judiciary, the public, politicians and the media. Many of today’s councils have remits that include both providing a mechanism for incorporating community views into sentencing policy and educating and informing the public about sentencing issues. Both of these functions – providing information and allowing the community a voice in sentencing policy – build on the assumption that a better informed and more engaged public will have greater confidence in the courts and sentencing.

Some researchers, however, have questioned whether such councils can really effect any change in perceptions of either the legitimacy of the criminal justice system or the adequacy of sentencing policy and practice. In particular, it has been argued that sentencing councils cannot create more than a symbolic exercise in participation, without substantive effect. Some have argued that “public education” is a euphemism for bringing the public around to the “correct” way of thinking, rather than providing a genuine opportunity for deliberation, discussion and debate (Indermaur 2008).

This view is rather controversial. If education, information and participation are not the answers to improving public confidence, what more can be done to address the public’s lack of accurate knowledge about crime and criminal justice and thus to improve levels of confidence?

The additional component missing from the education approach is emotion. Education and information can only go so far in changing people’s perceptions of, and attitudes toward, the courts and sentencing. Education taps into the intellectual aspects of people’s attitudes. It provides people with information about the effective elements of the criminal justice system: the nature and prevalence of crime, how and why sentencing operates as it does, and the outcomes of court processes. But education cannot tap into the emotional aspects of people’s perceptions. It is often these affective components that play the more important role in shaping people’s attitudes. Indeed, research has shown the (often circular) interconnections among attitudes, beliefs, emotions and experiences.

The Tasmanian Jury Sentencing Study (Warner et al. 2011) provides an indication of this disparity between people’s effective and affective responses to crime. Comparing jurors’ sentences with those actually imposed in the real cases upon which the jurors had deliberated, the researchers showed that the majority of jurors believed that the judge had imposed an appropriate sentence in the case. However, when asked an abstract question about sentencing in general for sex offences, violent offences, drug offences, and property offences, the majority of the jurors believed that sentences were too lenient. This “perception gap” – the lack of consistency between jurors’ views of the specific offence on which they deliberated and their general attitudes toward sentencing – was most pronounced for sex offences and violent offences. Given the highly emotive nature of these types of crimes, it is arguable that this finding nicely illustrates the difference between the effective and affective components that shape people’s attitudes. Although jurors were satisfied with the sentence imposed for the case in which they had heard all the relevant information, they remained dissatisfied with sentencing in general.

As attitudes to crime and punishment are often driven by emotive rather than instrumental concerns, public education must also address the symbolic and emotional issues that punitive attitudes reflect. For example, rather than attempting to improve public confidence in and support for alternatives to prison using only arguments based purely on economic grounds, policy-makers and governments could tap the affective component of public attitudes with the idea that people can change if given the chance through treatment and rehabilitation programs.

While informing and educating is typically aimed squarely at the intellectual aspect of public attitudes, ignoring the affective to focus solely on the effective addresses only part of the story of the relationship among attitudes, knowledge and confidence in the courts.

Unanswered Questions

With the general consensus in the field that improving public knowledge is a key strategy in improving public confidence, it is perhaps somewhat surprising that the issue of the effectiveness of education and information efforts has not been tackled. There are two components to this question: can education and information genuinely change people's underlying attitudes? That is, is the provision of education and information an effective way of addressing the public's lack of confidence? And if so, is this change durable and lasting?

The research is clear that when people are provided with more information, they become less punitive in their responses to survey questions. But it remains unclear whether this change reflects a change in people's opinions – responses formed only when asked about a specific issue – or a change to people's deeper underlying attitudes, their global, enduring orientation in general.

Some researchers have attempted to explain people's punitiveness based on psychological theories of attitude formation. For example, Maruna and King (2004) identified factors that predicted support for community penalties. In their model, expressive predictors (measures of collective efficacy and trust, anxiety about youth, economic pressure and global crime salience) and core beliefs and values (measures of a belief that crime is a choice that people make rather than a product of their circumstances and a belief in people's ability to change) had a strong effect on support for community sanctions, over and above the effect of both socio-demographics and instrumental factors (measures of direct victimization, local crime salience and fear of crime). A belief in people's ability to change was the strongest predictor of support for community sanctions.

But unique and valuable as this research is, it does not address the issue of whether education and information genuinely changes people's underlying attitudes. Indeed, this is a question that remains both unexamined and unanswered in the literature.

The same holds true for the issue of whether change in responses to survey questions represents durable change. It is clear that the provision of information changes responses in the immediate context of the survey, but is this change durable over time? This issue, too, remains both unexamined and unanswered in the literature. It is likely, however, that researchers will need to draw more heavily from the field of psychology in order to address this issue, as it is within that field, not the field of criminology, where substantial work has been undertaken examining the mutability and durability of attitudes. And it is likely that measures from the psychological literature, such as the strength of people's attitudes, will need to be imported into the criminological literature in order to understand the effects of people's attitudes on their behavior (strong attitudes will have significant effects and will resist even the strongest pressure to change, while weak attitudes will have little impact on a person's thinking or actions and will be vulnerable to situational pressures). This kind of approach remains untried in criminology, leaving no evidence as to whether education and information can bring about real, enduring change.

Emerging Directions in the Field

Although the study of public opinion about sentencing has been of substantial interest to researchers for at least several decades now, the field continues to develop, with new directions emerging in both research and policy.

Emerging Directions in Research

Based on the well-documented methodological concerns with asking for top-of-the-head public opinion, in recent years some researchers have opted for a more deliberative style of discussion with respondents. The primary aim in such studies is to provide respondents with the information they might need in order to come to a reasoned, considered and thoughtful decision.

Deliberative polls are based on the kind of detailed dialogue that is critical to the generation of durable and informed preferences. By spending an extended period with respondents (such as a weekend) providing information, discussing the issues and allowing time for deliberation and debate, the deliberative poll aims to measure a kind of public opinion that is informed, nuanced and stable.

In a similar vein, but within the context of real criminal cases, is the kind of study undertaken by Warner and colleagues. Respondents in the Tasmanian Jury Sentencing Survey were actual jurors, who had been in the courtroom all the through the trial, hearing the same information the judges were hearing. With that first-hand knowledge, and with additional information on crime and justice supplied by the researchers, respondents were able to give detailed, thoughtful and nuanced responses to the survey questions. While this study did not include measures of all the expressive factors, instrumental factors, and beliefs and values that may have affected respondents' attitudes, it clearly presents a model for the future of research in the field. It is a methodologically strong and fascinating approach to the study of public opinion about the courts and may well represent a new approach in criminological research in this field.

Emerging Directions in Policy

Given the prominence of the media as people's main source of information about sentencing, one of the ways in which courts have been working to improve public confidence is by working more closely with the media to increase the flow of information. Court websites now include access to judges' sentencing remarks, allowing both the media and the public access to the full reasons behind an individual sentence. Courts are employing liaison officers who can explain sentences to the media in lay language and can create accessible summaries of complex cases. Information and education campaigns, organized by the courts themselves or by governments via the establishment of sentencing advisory councils, aim to inform and educate both the media and the community more generally.

It remains to be seen how courts, governments and bodies such as sentencing councils can address the affective as a way of improving confidence in the courts.

Related Entries

▶ Institutional Theories of Punishment

Recommended Reading and References

Diamond SS, Stalans L (1989) The myth of judicial leniency in sentencing. Behav Sci Law 7(1):73–89

Doob A, Roberts JV (1983) Sentencing: an analysis of the public's view of sentencing. Department of Justice Canada, Ottawa

Flanagan T, Longmire D (eds) (1996) Americans view crime and justice: a national public opinion survey. Sage Publications, Thousand Oaks

Gelb K (2011) Predictors of confidence: community views in Victoria. Sentencing Advisory Council, Melbourne

Green DA (2006) Public opinion versus public judgment about crime. Br J Criminol 46:131–154

Hough M, Roberts JV (1998) Attitudes to punishment: findings from the British crime survey. Home Office Research Study No. 179, Home Office Research and Statistics Directorate, London

Hough M, Roberts JV (2004) Confidence in justice: an international review. Institute of Criminal Policy Research, London

Hutton N (2005) Beyond populist punitiveness? Punishment Soc 7(3):243–258

Indermaur D (2008) Dealing the public in: challenges for a transparent and accountable sentencing policy. In: Freiberg A, Gelb K (eds) Penal populism, sentencing councils and sentencing policy. Hawkins Press, Sydney

Maruna S, King A (2004) Public opinion and community penalties. In: Bottoms A, Rex S, Robinson G (eds) Alternatives to prison: options for an insecure society. Willan Publishing, Devon

Roberts JV, Doob A (1990) News media influences on public views of sentencing. Law Human Behav 14(5):451–468

Roberts JV, Hough M (2005) Understanding public attitudes to criminal justice. Open University Press, Berkshire

Roberts JV, Stalans L (1997) Public opinion, crime and criminal justice. Westview Press, Boulder

Roberts JV, Stalans L, Indermaur D, Hough M (2003) Penal populism and public opinion: lessons from five countries. Oxford University Press, Oxford

Warner K, Davis J, Walter M, Bradfield R, Vermey R (2011) Public judgment on sentencing: final results from the Tasmanian jury sentencing study. Trends and issues in crime and criminal justice No. 407. Australian Institute of Criminology, Canberra

Public Order Crimes

▶ Moral Crimes

Public Representations of the Police

Haia Shpayer-Makov
Department of History, University of Haifa,
Haifa, Israel

Overview

Ever since the introduction of organized and systematic police forces in the Western nation-states in the early and middle decades of the nineteenth century, they generated great interest in the media. As a prominent institution which evolved in tandem with the rising weight of public opinion in the political arena, the modern police, especially in democratically inclined societies, aimed at cultivating a favorable reputation on the premise that a positive image constituted a key factor in attaining or maintaining their legitimacy, social standing, and efficiency. The interaction between the media and the police, and hence the dominant public representations of the police, changed over time and from country to another. This entry focuses on England, which, while not necessarily typical as a case study, offers an analysis of a society where the issue of police images was crucial. Although the English police did not refrain from coercive methods, whether overt or subtle, they consistently displayed a marked sensitivity to their public image, and many of their operative decisions were made in this light. Further, the English media, in its various forms, devoted exceptional attention to police activity and personnel, providing a vibrant forum for diverse and contesting views on the topic. Although media coverage contained criticism – at times vehement – of the police and its agents, and relations between journalists and the police were not without tension, current research reveals that on the whole, the portrayals of the police were supportive and thereby also enhanced the power of the state. The uncomplimentary opinion of some sectors of the public, particularly those who were the object of police heavy-handedness and control, tended to be played down in public discourse. Since a sizable portion of the population derived their perceptions from the media and not from direct contact with the police, media images had an important impact on hegemonic public attitudes.

The first section of this entry explains the focus on England. The second section traces attitudes in the major conveyors of police representations – the press and fiction – from the early nineteenth century to the First World War. At its core is the finding that despite continued criticism and varied messages in each media format, by the late nineteenth century, the media had shifted from wide criticism and rejection of the notion of a strong public police to an acceptance of its beneficial role. The third and fourth sections explore the interwar and post-Second World War periods, respectively, including, in addition to print culture, also an examination of cinema films and television, pointing to the continued largely supportive attitude in the twentieth century. Despite rising dissatisfaction with the police in the media of the turn of the twenty-first century, the English police maintained their elevated position as a national icon.

Introduction

Ever since their emergence in the early and middle decades of the nineteenth century, modern police institutions in Western nation-states have systematically attracted considerable public notice and have become recurring objects of commentary and publicity.

The fact that the evolution of professional bureaucratic policing systems coincided with the accelerated growth of literacy, mass readership, and the publishing industry intensified this phenomenon. Interest in policing was manifested in a variety of formats, including official documents, parliamentary reports, court proceedings, printed and handwritten correspondences, and public speeches, but what brought the police to the forefront of public attention were the mass media, which in some countries, particularly Britain, France, and the USA, became almost obsessed with the topic. The dramatic changes in media technologies in the twentieth century magnified this preoccupation even further. If in the nineteenth century the press in all its forms, along with published fiction and nonfiction, featured substantial amounts of material about the police, this was supplemented, from the beginning of the twentieth century, by the cinema and, after the First World War, by radio and then television, which dominated the media scene in the latter part of the century. Naturally, the extent and content of the coverage depended inter alia on the freedom of communication in each country. Wherever the media enjoyed a relative absence of interference, the more likely they were to report and convey a range of opinions on the vocation and its personnel.

The interest in policing exhibited by the media is not surprising. Narratives about transgressions, breach of the law, deviant behavior, and the responses they engendered have riveted the human imagination ever since the dawn of civilization, as exemplified in the first chapters of the Bible. As a pivotal social institution designed to detect and prevent crime and maintain order, the modern police shared this intrinsic appeal. Moreover, the mass media, which since the nineteenth century were driven by economic interests and the desire to maximize circulation, were determined to satisfy the boundless curiosity of the public in police matters.

This created an acute dependence on the police for stories, information, and insights, whether for news and entertainment or sober appraisal – a dependence that was symbiotic, since the police in turn needed the media as a channel to the public. The introduction of uniformed patrols in the public sphere made police officers ubiquitously present and highly visible. Their performance, priorities, and conduct therefore informed public perceptions on public policing. Nonetheless, whether or not people had some first-hand experience with the police, many relied on the media to fashion their opinion of the police.

Needless to say, what the police sought was positive publicity, as media attention could both benefit and harm them. In fact, few occupations so badly required favorable exposure as public policing. More and more countries in the nineteenth century witnessed the growth of civil society and democratic values. With public opinion acquiring greater influence on government policies and performance, and on the fate of the body politic, it was crucial for the sovereign state to win as much support as possible. The police, seen as the arm of government, needed to be widely perceived as impartial instruments of the law and as functioning for the good of society. This was vital especially against the background of intense opposition both to traditional policing bodies and to the establishment of new forms of law enforcement, sentiments that were pervasive in large parts of the Western world. There was a need, primarily in democratically inclined societies – typically, the French Third Republic – to demonstrate that the new police were utterly necessary and that their presence was poles apart from the dominant public fearsome image (Berlière 2007). While the police could not expect entirely uncritical backing on the part of the governed, they aimed for sustained control and gaining the confidence of the public generally.

Heightened sensitivity to the reputation of the police continued to guide the thinking of political authorities in the twentieth century even in countries where the police had managed to improve their initially uncomplimentary image. Demonstrating that the police adhered to the letter of the law was of the highest significance in parliamentary democracies, including Weimar Germany in the 1920s and Spain in the early 1930s, but a degree of consent was critical in every society. Even a totalitarian regime such as fascist Italy, which greatly relied on the coercive and arbitrary power of its police, favored good relations with the population over sheer reliance on terror (Dunnage 2007). The drive to maintain a positive police image strengthened in the liberal democracies during the second half of the twentieth century, according special importance to public approval. This did not mean that, in practice, the police always kept to the letter of the law but that throughout the period, the police and the governments they served felt constantly obliged to reinforce their standing.

The police impulse for sympathetic representation also had prosaic reasons. Their success rate, and hence their aura of proficiency and effectiveness, was markedly contingent on active public cooperation, for example, in providing evidence. Not only was it widely contended that a good reputation prompted people to come forward to aid the police (Johansen 2007) but also that a good reputation made people more tolerant of police violence and the use of other contentious means which in turn helped the police attain good results. Again, the media could contribute greatly toward the establishment of a professional image of the police both by underscoring successes and acting as sources of essential information.

The numerous points of contact and reciprocal relationships between the media and the police were manifested in a variety of forms, changing over time and diverging from country to country in light of particular national norms. As a by-product, public representations of the police, too, showed considerable variety in each national context. Space constraints dictate limiting the survey of the evolution of the police image during the nineteenth and twentieth centuries to a single case study – arguably, the most famous police force in the world – the English police. The reasons for this choice are many.

Why the Focus on England

A consciousness of the fundamental importance of good repute was deeply entrenched in the policing system in Britain. Its core structure, based on all-encompassing patrol networks, entailed constant interaction with the policed population. The underlying premise was that the success of law and order enforcement hinged on mutually respectful, or at least mutually tolerant, relations between the police and the public. While some historians stress the willingness of the new police to serve as agents of middle class control of the social and moral behavior of the lower orders, with little heed to the latter's sensibilities, others highlight the determination of senior police management to avoid public hostility and defiance in order to facilitate the constable's work (Inwood 1990). In his comparison between the ascent of police forces in London and New York in 1830–1870, Wilbur Miller quotes Commissioner Richard Mayne, one of the first commissioners of the Metropolitan Police of London, as emphasizing that "the real efficiency of the police depends upon the estimation in which it is held by the public" (Miller 1977). To promote this objective, the first instruction book of the Metropolitan Police, published in 1829 (when the force was established) and in many editions thereafter, commanded police officers to be courteous and considerate as part of a general strategy of inculcating habits of moderation, restraint, and cool-headed response. The spirit of these directives was echoed in the instruction books of the other police forces which subsequently spread throughout the country (Klein 2012).

This embedded sensitivity was a reaction, initially, to the particularly powerful and wide-ranging objections to the notion of a highly organized and systematic police – objections which preceded and accompanied the establishment of police forces in the British Isles. Different social sectors across geography, class, and political boundaries, ranging from Tory gentry to working-class radicals, were motivated by varied fears, namely, that the proposed police would undermine their own particular interests, behave as a standing army, strengthen the central state excessively, become a political tool, intervene in private life, curtail the famous British liberties, or become a financial burden on the ratepayers (Storch 1975). The source of animosity of local elites both in London and in the provinces lay in their desire to retain local control of policing. A recurrent concern equated the new policeman with the figure of the spy, associated closely with despotic governments on the continent and with unlawful and corrupt behavior. The guiding apprehension assumed that under disguise, and therefore unnoticed and unsuspected, undercover police constituted a powerful threat to the norms and habits of life dear to British subjects (Shpayer-Makov 2011).

To overcome opposition, the reformers who backed the creation of the new police molded the concept of a force that simultaneously allayed such fears while meeting the needs of the social and political elite to govern. Officers wore a uniform, but of a blue color, with top hats, to distinguish them from the military, and most were armed with a wooden truncheon only. The uniform made the policeman visible – hence easily identifiable – and approachable when help was sought by the public, thus fortifying the image of policing as a public service. Some policemen occasionally removed their uniform to perform detective tasks, but the scepter of the fearful figure of the spy had the effect of postponing the formation of a fully fledged detective branch in the Metropolitan Police for 13 years. Ultimately, this famous detective branch (popularly referred to as Scotland Yard) and the few other such branches erected subsequently, mainly in urban police forces, formed only a fraction of the entire police workforce. Additionally, apart from the Metropolitan Police of London, all other police forces were accountable not to the Home Office but to their local authorities.

Even after the institution of the police force in England gained wide acceptance in the latter part of the nineteenth century, suspicions of the potential harmful powers of the police lingered and loomed over their evolution, as in other countries. Yet, perhaps more prominently than other police forces, the English police had come to constitute an established national institution and a source of pride to its citizens, even if not unequivocally, as will be seen below. Moreover, it appears that British policy makers made special efforts to attain this consensus. So ingrained was the recognition of the weight of public opinion, that even minor police policies and ad hoc decisions were often colored by the notion of how these would be interpreted outside the police (Inwood 1990). Indeed, a focus on England in the nineteenth and twentieth centuries reveals a culture in which individuals could discuss societal issues and exchange views openly, with the media serving as primary venues for airing a variety of ideas regarding the criminal justice system. In fact, the English media evinced an exceptional and persistent preoccupation with law enforcement, a tradition which reached new heights in the latter part of the twentieth century (Allen et al. 1998). As such, the media facilitated an awareness by the police leadership and state officials of the sentiments and expectations of the public, thereby affecting the evolution of the English police importantly. Notably, research on representations of the police in England is singularly extensive. Although most studies focus on the second half of the twentieth century, quite a few deal with the history of the police image prior to the Second World War. Even if this period does not occupy entire book-length texts, it features in a host of articles and book chapters, particularly regarding the pre-1914 period, and is otherwise interwoven in any number of scholarly texts about the English police or detective fiction.

The rest of this entry will trace dominant representations of the police in the media between the inception of the first modern police force in London in 1829 and the late twentieth century, focusing on two types of cultural texts which were the principal conveyors of police images in the nineteenth century – the press and fiction, extending to include the emerging electronic media – cinema and television, in the twentieth century.

Key Issues

The Long Nineteenth Century

Against the backdrop of economic and social tensions in the 1830s and 1840s, interspersed by overt manifestations of class struggle resulting in brutal incidents of police crowd control and other constricting interventions in working-class life, many contemporary critics concluded that their worst fears regarding the inauguration of the new police had been realized. Hostility to the reformed police was so profound and widespread that their appearance in the streets of the communities which opted for the new arrangement was commonly met with such appellations as "blue devils," "blue plagues," "blue idlers," "hired mercenaries," and "unconstitutional bravoes," as reported in the press (Storch 1975). They were also called "Jenny Darbies," a misnomer for the much reviled continental gendarmes (Miller 1977). Not only did many articles disparage police actions, but their very existence and function were often described as redundant, and they themselves as "unproductive parasites" (Storch 1975). While responses were not monolithic – the press contained articles upholding the current restructuring of the old police alongside radical demands for the abolition of the new body – on the whole public debate during this period was pervaded by suspicion and distrust of the police.

In parallel with concerns about the potential or actual oppressive or intrusive nature of the police, press rhetoric criticized the reformed police as lacking the capacity to suppress crime or cope effectively with other designated duties. In as much as police manpower was recruited mainly from the laboring classes, the policeman appeared to many to be ignorant, simple minded, and limited, with his power attributable to his imposing physiology rather than his mental abilities. "Punch" mocked constables for their fondness for servant girls and cooks and their partiality for beer and food while on duty (Pulling 1964). These themes were aired time and again in ballads and songs performed in music halls. Yet, if such references indicated disrespect for the new police, they also suggested that fear was not necessarily the predominant sentiment and that even during the founding years of the modern police, some viewed constables with a touch of affability.

Predictably, the many working people who habitually found themselves at the wrong end of the policeman's truncheon and were subject to orders to "move on," close supervision, arbitrary arrest, or police violence perceived the police as unwelcome intruders in the community and agents of the political and economic elite. Those with access to the media sometimes expressed their resentment in writing, but many others resorted to physical confrontations not only while resisting arrest but also in taking the initiative and attacking police officers, sometimes unprovoked (Storch 1975). Members of the working class also demonstrated their antipathy by defying new laws and regulations enforced by the new police or simply by refusing to collaborate with them when help was required.

Before long, however, the social classes with a stake in the pacification of political radicalism and in the protection of property gravitated toward staunch support of the police institution. No doubt, the way the new police handled radical protest and public disorders in the 1830s and 1840s, especially in comparison with police on the continent, convinced the propertied classes of the utility of police functioning as being in their interests (Emsley 1991). The many journalists who served as their mouthpiece echoed their satisfaction that the police acted with effectiveness coupled with measured force (Miller 1977), implicitly suggesting that the police were their trusted protectors against the "dangerous classes" (Emsley 1983).

The process was not linear. Miller identifies the 1850s as the decade in which the image of the police improved considerably but regards the 1860s as "a crisis period in police-public relations" (Miller 1977) in light of garroting (a form of violent street robbery) panics in the early 1860s; a well-publicized demonstration in Hyde Park in 1866 which ended in riots and violence; and the rise of Fenian terrorism in the latter part of the decade. However, by the mid-1870s, even though members of the privileged classes here and there questioned certain facets of police behavior, overall, the notion that the police were indispensable for the preservation of the social order reigned supreme in these classes, as is evident in the content of the mainstream press.

Existing studies of police-public relations in the nineteenth century show that it took longer for the working classes to acquiesce in the police presence. The "rougher" elements in the working classes, who continued to be the target of police heavy-handedness during the second half of the century, still regarded the police as their enemy (Miller 1977). But the expanding "respectable" sections of the lower orders, who were victims of criminal offenses no less than their social superiors, grew accustomed to turn to the police as the legitimate representatives of the law and benefit by the services provided irrespective of class, even if they were not particularly fond of the police (Reiner 1985). That a high proportion of working-class adult males were enfranchised in the electoral reforms of 1867 and 1884, thus gaining a certain foothold in the established order, and that the messages of restraint and moderation permeated the police ranks, may have also mitigated the attitude of the respectable working classes toward the police. However, their support was tentative and not deep-rooted. The many instances of police class bias toward individuals in the streets, and their aggressive behavior in working-class demonstrations and strikes which flared up in the late nineteenth century and on the eve of the First World War, kept reviving suspicions of police partiality.

Throughout the period, the size of the press and its readership spiraled numerically, and with it public awareness of various aspects of police existence. The coverage contained a mosaic of approaches and outlooks, depending on the ideological stance of the paper. Broadly, while the Tory press became the most loyal defender of the police, liberal, and to a much greater extent radical newspapers and journals more readily adopted an offensive position vis-à-vis the police (Curtis 2001). Not surprisingly, socialist papers posited the police as the enemy of the people. However, on balance, from the early days of the police until the First World War, the press, which was not equally accessible to everyone, had undergone a metamorphosis from wide rejection of the notion of the public police to a largely consensual endorsement of it, a change which tended to downplay the nuanced, often inimical attitude in the lower social strata (Shpayer-Makov 2011).

Not that relations between the press and the police were consistently harmonious. Despite the rise in the status of journalists, and the mutual and mounting dependence between the police and the press, contacts between them were typified by reciprocal mistrust (Shpayer-Makov 2011). In varied degrees, police authorities in every community supplied newspapers and journals with the material the press needed to meet the insatiable demands of their readers for stories and reports about the criminal justice system, but they did so sparingly and often kept important information to themselves, especially when it reflected badly on the police. In such cases, journalists were forced to rely either on their own detective skills or on informal contacts within the police, particularly detectives who could provide disclosures about ongoing inquiries and who were willing to divulge inside information, sometimes for payment (Shpayer-Makov 2011). For its part, the press, which progressively saw itself as a watchdog of democratic values, monitored police and state authority and, at times, exposed unwholesome aspects of police work, which, as expected, fuelled further suspicion and even hostility between the two sides.

Notwithstanding the respective displeasure between the police and the press, and the occasional censorious account, journalists played a major role in constructing a predominantly flattering image of the English police, even if not always wittingly or actively. In addition to upholding viewpoints favorable to the police, many journalists promoted such an image in the belief that it contributed to the good of society, while undermining the police had the opposite effect (Shpayer-Makov 2011). Others refrained from judgmental commentary in order to maintain fraternal relations with the police. Moreover, the police had learnt to manipulate the mass media by providing selective material and hiding incriminating information. The outcome was partial and beautified accounts of police reality. Surely, some readers, whose number cannot be estimated, were aware of the bias implanted in the press coverage and did not adopt the perspectives expressed in it. Yet, the period was marked by rising public confidence in the press as the authoritative repository of news and hard facts in society; hence, it can safely be deduced that a great many readers accepted the favorable reports they read as mirroring authentic police life.

So beneficial did the police organization, as well as individual officers, appear that more and more columnists complained not about the police presence but about the shortage of police personnel. Admittedly, the institution of the foot patrol constituted a powerful reminder of the authority of the police and the state, but with time, the constable came to epitomize dispassionate and impartial law enforcement. Although he was often sketched as slowly trudging the streets or country lanes, these descriptions became amicable, if on occasion condescending, especially when referring to the rural constables as plods. Newspapers, including those with a conservative leaning, continued to condemn the police as inefficient when they failed to solve a case which gained headlines, such as the brutal murder of five prostitutes in the East End in late 1888 when the London police proved unable to even identify the perpetrator (popularly called Jack the Ripper) (Curtis 2001), but otherwise, the press treated the police officer as reasonably effective. The radical and liberal press vilified the police as undermining civil liberties when they dealt brutally with demonstrations and strikes, but such cases, as well as publicized instances of police corruption and the employment of spies and agents provocateurs, were increasingly posited as exceptional rather than the rule.

Significantly, as the image of the English constable was consolidated in the latter part of the nineteenth century, it became further removed from the initial adverse portrayal of the policeman as a walking panopticon, the personification of a ubiquitous and oppressive technique of surveillance. Instead, the mainstream press, which was increasingly read by the lower strata as well, was packed with stories depicting police officers sacrificing themselves when needed, for example, to save civilians' lives. The visual media, which expanded in the second half of the century, used mass-reproduced images to reinforce the verbal texts, transmitting and circulating illustrations of gallant officers, both uniformed and not, for example, in the midst of catching violent criminals. The policemen in all these images conformed to hegemonic assumptions about the ideal model of working-class masculinity (Shpayer-Makov 2012).

Time and again, the English police were compared to their counterparts on the continent and were found superior if not in efficiency than in ethical standing and respect for the law and individual rights (Shpayer-Makov 2011). The press reiterated the fact that the English police, and above all the Metropolitan Police force, became a model of imitation worldwide, inter alia, because of the bobby's deep-rooted local popularity (Johansen 2007). Further, while in Paris and Berlin the image of the policeman remained that of an agent of the state, the frame of reference in London and in the rest of the country became one of assistance to the community (Emsley 1983).

Even detectives, who initially were identified with the figure of the spy, gained acceptance. At first, journals did not distinguish between uniformed and plainclothes policemen, but before long, even though detectives constituted a small percentage of the police labor force, they drew disproportionate press attention, clearly by virtue of their more intimate connection with crime and hard-core criminals. One of the first opinion leaders to develop an inquisitive curiosity about detectives and their activities was the author and journalist Charles Dickens, who made a point of listening to their stories and accompanying them on their tours of duty, later publishing his impressions in several articles in his journal *Household Words* in the early 1850s. Partly under his influence and the growing interest shown by many other journalists thereafter, the figure of the detective assumed the role of an urban explorer probing the seamy side of life, where he commanded fear bound up with respect. If uniformed policemen attracted sympathy and even esteem, detectives, most of whom generally followed a mundane and unexciting routine, gained a romantic and heroic aura, depicted not only as adventurous and fearless but also as highly able, sharp, and enterprising. By the end of the nineteenth century, some detectives, particularly those serving in the central squad located at Scotland Yard (the headquarters of the Metropolitan Police), found themselves in the limelight and attained the status of celebrities. At the beginning of the twentieth century, with the adoption of fingerprinting as a standardized forensic practice, they gained the added luster of expert knowledge and scientific rigor.

The combined figures of the helpful and supportive community policeman and the indefatigable and clever detective operating in dangerous venues to restore order to society forged the impression, explicitly reiterated in the press, that they constituted the best police service in the world (Emsley 1991).

This was not the case in the literary world. While fictional texts portrayed police officers as moral, honest, and dedicated to their work, on the whole, these books were also more critical of the capabilities of the police and accorded them much less respect than the press. Yet, similarly to the press, though far more emphatically, literary works demonstrated greater interest in detectives than in uniformed policemen. Indeed, so compelling was the figure of the detective that, by the First World War, a whole genre had evolved around it. However, in contrast to the press, the typical fictional police detective emerged as neither highly competent nor a great achiever, and certainly not heroic. What he particularly lacked was a creative imagination and an elevated facility in reasoning. There were impressive and efficient police detectives in the growing body of books and stories featuring detectives, such as Inspector Bucket in Charles Dickens's *Bleak House* (1853) and Sergeant Cuff in Wilkie Collins's *The Moonstone* (1868). Such characters also appeared in seemingly factual accounts of crime fighters written in the first person, which flourished in the third quarter of the nineteenth century. However, with the exception of the pseudo-memoirs and certain imports, such as the translated works of Émile Gaboriau, centered on the French police officer M. Lecoq, the police detective was seldom the protagonist of the plot and not even the principal enforcer of the law.

The character who gained the admiration of contemporary authors and became the focus of numerous narratives, whether canonical or lowbrow, was the private detective, both amateur and professional, who was commonly posited as charismatic, brilliant, and effective. Sometimes he operated by himself, without the involvement of official law forces, but as the nineteenth century wore on, private and public detectives were progressively juxtaposed in the same narrative. The pervasive motif in these narratives, which contrasted starkly with the reality of criminal investigation, featured the private detective as the person who conducted the investigation and was decisive in the successful unveiling of the mystery at the root of the plot. Frequently, he/she gave the police instructions, which they followed as a matter of course, even enthusiastically. Sometimes it seems as if the mediocrity of the policeman only served to highlight the outstanding abilities of the private sleuth. This theme – the superiority of the private over the official detective – had been interwoven in stories by the American author Edgar Allan Poe, first published in England in the middle decades of the nineteenth century, and in books by English authors which appeared at that time, but with the unsurpassed popularity of Arthur Conan Doyle's Sherlock Holmes (first published in 1887), it spread like wildfire in popular literature.

Conceivably, the recurrent fictional formula implying that Holmes-like figures (rather than the official variety) were the quintessential detectives was adopted by many authors, as they viewed it as a key to success. Yet, the repetition of this theme may have also had its roots in class-based doubts of the image projected in the press that the police, originating mainly in the working class, could contend by themselves with the many dangers prevalent in the increasingly industrialized urban environment (Shpayer-Makov 2011). Indeed, most of the fictitious private detectives originated from the middle or upper classes, a lineage which apparently lent itself more readily to a heroic stature. At the same time, a subtle message was conveyed that the police, undeniably essential in modern society, should raise the social and professional quality of its manpower, as well as cooperate and consult with the private sector of the privileged classes more closely.

The Interwar Period

Despite changes in various aspects of police culture during and after the First World War, the attitude of the press and the literary world toward the police remained essentially the same as in the prewar period. The press, and not only in its radical periphery, may have felt more disposed to attack the police for misconduct and malpractice, especially in the metropolis. Newspapers embarked on campaigns exposing details of police corruption, such as the Sergeant Goddard case of 1928 (which led to a royal commission examining the police a year later) and several other scandals concerning dubious police methods, which gained headlines in the 1920s (John Carter 2010). The police handling of both fascist and left-wing demonstrations in the 1930s gave rise to critical accounts representing different political camps. The language used may have been more outspoken than before. However, seen in the aggregate, the press continued to present the police as law-abiding servants of the people, treating reports of the abuse of police power as isolated incidents. Viewed against the background of the rise of the Soviet Union and the spread of authoritarian and fascist regimes in Europe, characterized by intolerance and the use of unlawful means and the secret police to suppress dissidence, the English police shone as representing democratic values and the rule of law.

Antagonistic relationships between the police and the populations who were routinely singled out by the police for stringent control, such as slum inhabitants, working-class adolescents, street sellers, and the unemployed who lived on the margins of illegality (White 1983), was not the stuff that attracted press notice. Moreover, the masculine image of the police was not affected by the entrance of a small number of women into their uniformed ranks. Entrusted with the new tasks of directing traffic and regulating the speed of motor cars, the interwar policeman came into close contact for the first time with members of the middle and upper classes, who resented this new role and complained that he should ignore minor traffic infractions and instead spend his time catching criminals (Klein 2007). However, this did not dampen the oft-articulated conviction that the English police were the best in the world (John Carter 2010).

The reasons for the strained relations between the press and the police, and with Scotland Yard in particular, did not disappear. Notably, however, Scotland Yard manifested an increased appreciation of the importance of maintaining a good rapport with the press when in 1919 it set up a press bureau, albeit consisting of a single civil servant whose function was to issue twice-daily bulletins to journalists (Chibnall 1977). This appreciation intensified as the century wore on.

Fiction writers still found the private detective more suitable for a commendable profile than the policeman, and although a more agreeable approach to policemen in detective fiction is detectable, they were still portrayed as inferior in intellectual acumen as well as powers of observation and detection. Only infrequently was a police detective the protagonist in the narrative. The most famous author associated with this tradition in this period was Agatha Christie, who invented two private detectives – Hercule Poirot and Miss Marple – both of whom were consistently responsible for ingenious breakthroughs in investigations and for revealing the true story.

The cinematic screen, which since its pioneering days at the turn of the twentieth century featured crime films, more or less replicated printed fictional works in presenting the private detective as more accomplished than police officers. Not only were the officers not cast as heroes, but occasionally, they were depicted as corrupt, unfair, or ridiculous (Reiner 2008).

Notably, a number of real-life policemen, and not only from the highest echelons, took an active part in molding the police image in the media. Starting in the late Victorian period, several police officers, primarily retired detectives from Scotland Yard, managed to publish their memoirs, a commendable achievement given their humble origins (Shpayer-Makov 2011). This trend continued and accelerated in the interwar period (Lawrence 2003). Insulted by the belittling portraits of police officers in fictional works, they took advantage of the growing interest in their life stories as crime fighters to contest this image as unreliable and distorted. Whether consciously or not, the representational strategy adopted by these police authors succeeded in crafting an image of themselves as professional and talented policemen, possessed with a remarkable memory and cerebral acumen (Shpayer-Makov 2011). No private detective was on the scene to advance investigations and protect society. Success was theirs only and many times over.

The Post–Second World War Period

The war and its aftermath promoted greater state regulation in Britain, and with it, the expansion of the public sector, a process that was manifested inter alia in the extension of the welfare state, the democratization of the educational system, and enhanced police powers. Significantly, economic recovery in the late 1940s was accompanied by a rise in crime statistics. Although it is difficult to establish a definitive correspondence between real events and media representations, the convergence of these social and political factors undoubtedly provided the context for major shifts in the portrayal of police officers in fiction and cinema in the postwar period.

The press had long exhibited little concern for private detection, while putting faith in the police as the upholders of law and order. It also maintained a tradition of showing a certain interest in the everyday life of the constable on the beat and in the inner world of police stations, although typically, its preference was a focus on the dramatic and the sensational or at least on what was newsworthy. A new development was that fictional works, whether in print or in film, now followed a similar pattern, shifting away from placing private detection in the foreground of law enforcement and instead highlighting authentic police work. While talented private sleuths still enjoyed a lingering appeal, significantly, a novel literary sub-genre emerged in the late 1940s – the police procedural – which soon acquired popularity on both sides of the Atlantic. The title derived from a conventional plotline in which ordinary police detectives, shown as true professionals, deciphered crimes by employing real-life police procedures and methods, which were at variance with the more fanciful means used by the classic private investigator (Dove 1982). In line with the collective spirit of the age, the resolution of an investigated crime was not the product of the mind and exertions of the individual detective, as formerly, but of the collaborative efforts of a police team headed by a central figure who was invariably a policeman. In the same realistic vein, the principal investigator was not an outstanding individual or from a pedigreed background.

Until the mid-1960s, the police typically played a minor role in cinema films, as in the prewar crime fiction (Allen et al. 1998). Still, in the immediate postwar period, an invented cinematic police figure struck a special chord with contemporary audiences and left an indelible imprint on public consciousness thereafter – George Dixon from the East End of London, appearing in the film *The Blue Lamp*, released in 1950. Halfway through the film, he is killed by a criminal (played by the young Dirk Bogarde), but television, which since its inception opted for police protagonists in its crime series, resurrected him in the mid-1950s as the central figure (played by Jack Warner) in a series entitled *Dixon of Dock Green*, which ran until 1976. Dixon, depicted as an honest, kind-hearted, fair-minded, and dedicated patrol officer, respected by all in his neighborhood, fitted in perfectly with the evocation of society in postwar crime cinema "as largely based on shared values and a clear but accepted and just hierarchy of status and authority," where "criminals were normally brought to justice" (Allen et al. 1998). So idealized was this personification of the English bobby (Emsley 1991) that years later people would nostalgically refer to the period it described as the golden age of English policing.

Although Dixon continued to enjoy public adulation, changes in the police and in society at large during the 1960s and 1970s introduced new police images in the media. The police, and not private civilians, continued to be perceived as the ultimate law enforcers, but the increasing use of cars in policing resulted in a decreased role for the beat officer and a distancing of policemen from the community. This development was reflected in the acclaimed TV series *Z Cars*, shown during 1962–1965 and 1967–1978, revolving around a mobile team of policemen in the north of England.

Moreover, the general climate of police culture in film and TV no longer connoted fair play and strict adherence to the law, as in the Dixon series. Against a backdrop of student protest, growing permissiveness, the advent of a rebellious youth culture and other challenges to the established order and, more broadly, to the concept of authority, policemen were increasingly featured as callous and manipulative crime fighters who did not hesitate to bend the law and use violence yet were better able to combat hardened criminals and other grave dangers to society. Political terrorism by the IRA and the Angry Brigade in the early 1970s certainly lent force to the notion that policemen needed to be tough and, if necessary, resort to shady practices (Chibnall 1977). A number of reported scandals pointing to abuse of power within the police, while slurring the reputation of the police also reinforced the understanding that the imperfect fictional police characters in the entertainment media were true to life. Here and there, films and TV series depicted police officers as good-natured fools (Allen et al. 1998), but a more typical presentation, even if it focused on their weaknesses, still conveyed the point that they were skillful and hard workers who gave their all to serve the community and obtain just ends (Leishman and Mason 2003). Such police characters appeared in various TV series and films, the most famous of which was television's Detective Inspector Jack Regan, the protagonist in *The Sweeney* (1975–1978), which centered on Scotland Yard's elite squad. Disrespectful of hierarchy, bureaucratic rules, and genteel norms, Detective Regan was the symbol of a relentless urban fighter against threats to the community who utilized unconventional means but was incorruptible.

The late twentieth century is regarded by many scholars as a dramatic milestone period in police exposure in the media, when "the police were rarely out of the news, and very little of the police news was good news" even though police authorities were investing far more in image promotion than previously (Stead 1999; Reiner 1985). Diverse factors may account for this exposure, including controversial police conduct in various circumstances, external and internal investigations of the police which gained considerable publicity, mounting media interest in policing matters, the diminishing role of public policing due to privatization, and the use of more advanced technological means to disseminate information (Leishman and Mason 2003). Still, as Robert Reiner has pointed out, the police remained "a powerful political and cultural force" (Reiner 2008).

Conclusion and Future Research

Clearly, from the rise of the reformed police in the third decade of the nineteenth century until the close of the twentieth century, the figure of the policeman was multifaceted, with each media format recording its own version of police representations. However, viewing the media as a whole, researchers are almost unanimous in concluding that despite twists and turns and periodic sharp criticism, they painted the police in a favorable light (Reiner 2008), generally downplaying the less salubrious features of police life, such as levels of corruption and rule-breaking, embedded prejudices and recurrent failures. Admittedly, readers did not necessarily internalize the mediated messages. Media research has long affirmed that audiences are not passive recipients of the material they read or see but rather interpret and make sense of this information, largely according to how it relates to various aspects of their lives (Reiner 2008). Nonetheless, no doubt the media played a critical role in elevating the English police to a prominent and respectful place in the national psyche.

Despite the extensive body of literature on texts related to the police, whether written by or about them, there is still considerable scope for further research of such texts. Undoubtedly, each country merits an overview of its own police representations, including comparisons between the content and impact of the various popular media locally and across national boundaries, exploiting new directions and the interdisciplinary emphasis in historical research. The diverse portrayals of the English police in written texts and on the screen have drawn wide scholarly interest, particularly since the cultural turn in the 1970s, but large gaps in our knowledge remain, for example, in visual representations in print and dominant perceptions on stage. Furthermore, various themes relating to the press in the pre–Second World War period might be explored, including police-press relations, their mutual influence, the role played by journalists in this context, and the self-management of the police image. In addition, little research has been devoted to rival representations of the police in the press of the interwar period.

With the spread of academic scholarship into the area of popular culture in the 1970s, detective fiction assumed primary importance in the reappraisal of popular texts. For example, best-selling detective narratives of the Victorian and Edwardian periods, which did not enter the canon and were forgotten with the passage of time, began to be the objects of academic study, but further examination of these little-known texts is called for as a way of revealing underlying popular sentiments of the time. Similarly, further critical scrutiny of police images on the screen in the early days of cinema is merited. Such studies should focus on the cultural meanings of texts and might well utilize the more traditional approaches in historical analysis. This research would add to our understanding of social experience as well as the collective mental world of the period.

Related Entries

- ▶ British Police
- ▶ Conceptualizing of Police
- ▶ Early Modern Police and Policing
- ▶ Gendarmerie Policing
- ▶ German Police Until 1918
- ▶ History of Criminal Investigation
- ▶ Police in Nazi Germany
- ▶ Use of Social Media in Policing

Recommended Reading and References

Allen J, Livingstone S, Reiner R (1998) True lies: changing images of crime in British postwar cinema. Eur J Commun 13(1):53–75

Berlière J-M (2007) The difficult construction of a ‘Republican’ police: the experience of the French third republic. In: Blaney G (ed) Policing interwar Europe. Palgrave Macmillan, Basingstoke, pp 14–30

Broduer J-P (2010) The policing web. Oxford University Press, Oxford, Ch. 3

Chibnall S (1977) Law-and-Order news. Tavistock Publications, London

Chibnall S (1979) The metropolitan police and the news media. In: Holdaway S (ed) The British police. Edward Arnold, London, pp 135–149

Curtis LP Jr (2001) Jack the Ripper and the London Press. Yale University Press, New Haven

Dove G (1982) The police procedural. Bowling Green University Popular Press, Bowling Green

Dresner L (2007) The female investigator in literature, film, and popular culture. Mc Farland & Co., Jefferson

Dunnage J (2007) Mussolini’s policemen, 1922–43. In: Blaney G (ed) Policing interwar Europe. Palgrave Macmillan, Basingstoke, pp 112–135

Emsley C (1983) Policing and its context. Macmillan, London, pp 1750–1870

Emsley C (1991) The English Bobby: an indulgent tradition. In: Porter R (ed) Myths of the English. Polity Press, Cambridge, pp 114–135

Gilman Srebnick A, Lévy R (eds) (2005) Crime and culture. Ashgate, Aldershot

Inwood S (1990) Policing London’s morals: the metropolitan police and popular culture, 1829–1850. Lond J 15(2):129–145

Johansen A (2007) A process of civilisation? Legitimation of violent policing in Prussian and French police manuals and instructions, 1880–1914. Eur Rev Hist 14(1):49–71

John Carter W (2010) ‘The Third Degree’: press reporting, crime fiction and police powers in 1920s Britain. Twentieth Century Br Hist 21(4):464–485

Klein J (2007) Traffic, telephones and police boxes: the deterioration of beat policing in Birmingham, Liverpool and Manchester between world wars. In: Blaney G (ed) Policing interwar Europe. Palgrave Macmillan, Basingstoke, pp 216–236

Klein J (2012) Quiet and determined servants and guardians: creating ideal English police officers, 1900–45. In: Barrie D, Broomhall S (eds) A history of police and masculinities, 1700–2010. Routledge, London, pp 201–216

Lawrence P (2003) ‘Scoundrels and Scallywags, and some honest men. . .’. Memoirs and the self-image of French and English policemen, c 1870–1939. In: Godfrey BS, Emsley C, Dunstall G (eds) Comparative histories of crime. Willan Publishing, Cullompton, pp 125–144

Leishman F, Mason P (2003) Policing and the media. Willan Publishing, Cullompton
Mawby RC (1999) Visibility, transparency and police-media relations. Polic Soc 9:263–286
Mawby RC (2002) Policing images. Willan Publishing, Cullompton, Ch. 1
Miller WR (1977) Cops and Bobbies. Chicago University Press, Chicago
Pulling C (1964) Mr Punch and the police. Butterworths, London
Reiner R (1985) The politics of the police. Wheatsheaf Books, Brighton
Reiner R (2008) Policing the media. In: Newburn T (ed) Handbook of policing, 2nd edn. Willan Publishing, Cullompton, pp 313–335
Rzepka CJ, Horsley L (eds) (2010) A companion to crime fiction. Wiley-Blackwell, Oxford
Shpayer-Makov H (2011) The ascent of the detective. Oxford University Press, Oxford
Shpayer-Makov H (2012) Shedding the uniform and acquiring a new masculine image: the case of the late-Victorian and Edwardian English police detective. In: Barrie D, Broomhall S (ed) A history of police and masculinities, 1700–2010. Routledge, London, pp 141–162
Sindall R (1990) Street violence in the nineteenth century. Leicester University Press, Leicester
Stead P (1999) ‘Its not all about Nicking folks’: dramatizing the police. In: Howell DW, Morgan KO (eds) Crime, protest and police in modern British society. University of Wales Press, Cardiff, pp 207–237
Storch R (1975) The plague of blue locusts: police reform and popular resistance in Northern England, 1840–57. Int Rev Soc Hist 20(1):61–90
White J (1983) Police and people in London in the 1830s. Oral Hist 11(2):34–41

Public Transit

► Crime on Public Transport

Public Transportation

► Crime on Public Transport

Publicity Sanctions

► Naming and Shaming of Corporate Offenders

Pulling Levers Policing

Anthony A. Braga
School of Criminal Justice, Rutgers University, Newark, NJ, USA
Kennedy School of Government, Harvard University, Cambridge, MA, USA

Overview

Pulling levers focused deterrence strategies are increasingly being implemented in the United States to reduce serious violent crime committed by gangs and other criminally active groups, recurring offending by highly active individual offenders, and crime and disorder problems generated by overt street-level drug markets. These new approaches share many of the same basic elements of third-party policing, problem-oriented policing, and situational crime prevention strategies. Briefly, pulling levers focused deterrence strategies seek to change offender behavior by understanding underlying crime-producing dynamics and conditions that sustain recurring crime problems and implementing an appropriately focused blended strategy of law enforcement, community mobilization, and social service actions. Direct communications of increased enforcement risks and the availability of social service assistance to target groups and individuals are a defining characteristic of pulling levers strategies. The available evaluation evidence suggests these new approaches generate noteworthy crime reduction gains. Unfortunately, little systematic research exists to explain why these strategies seem to work in practice.

Introduction

A number of US jurisdictions have been experimenting with new problem-oriented policing frameworks, generally known as “pulling levers” focused deterrence strategies, to understand and respond to serious crime problems generated by chronically offending groups, such as gun violence among gang-involved offenders. These new initiatives generally follow a core set of activities to reduce violence: the convening of an interagency working group representing a wide range of criminal justice and social service capabilities; more or less real-time partnerships with academic researchers to develop jurisdiction-specific assessments of violence dynamics, perpetrator and victim characteristics, and related issues such as drug market characteristics and patterns of weapons use and acquisition; the design and implementation of a focused law enforcement, community mobilization, and social service/opportunity provision strategy; and the direct communication of enforcement risks and offers of help to the criminally active groups and chronic offenders.

Many of the key ideas of pulling levers focused deterrence strategies fit well within the developing discipline of third-party policing (Mazerolle and Ransley 2005). This perspective suggests that the resources of the police should be expanded to “third parties” that are believed to offer significant new resources for doing something about crime and disorder. Third-party policing asserts that the police cannot successfully deal with many problems on their own and thus that the failures of traditional policing models may be found in the limits of police powers. Using civil ordinances and civil courts or the resources of private agencies, third-party policing recognizes that much social control is exercised by institutions other than the police and that crime can be managed through agencies other than the criminal law.

The developing evaluation evidence suggests that pulling levers focused deterrence strategies generate noteworthy violence reduction benefits (Braga and Weisburd 2012). However, the varied crime prevention mechanisms that generate the observed effects are not well understood. Pulling levers strategies follow the iterative steps of problem-oriented policing (problem identification, problem analysis, development of appropriate prevention responses, and assessment; Goldstein 1990); honor core situational crime prevention ideas (Clarke 1997), such as increasing risks faced by offenders, while finding new and creative ways of deploying traditional and nontraditional law enforcement tools to do so; and are illustrative of the regulatory shift in policing that is articulated within third-party policing such that some of the “pulling lever” tools are the same interventions used in third-party policing programs that seek to engage others, either voluntarily or coercively, to do something more about crime and disorder problems than they are currently doing (Mazerolle and Ransley 2005).

Pulling levers focused deterrence strategies can be seen as a generally applicable framework for acting on the core set of situational crime prevention ideas (Braga and Kennedy 2012; Skubak Tillyer and Kennedy 2008). Mapping focused deterrence actions onto the base situational prevention ideas of increasing the effort, increasing the risks, reducing rewards, reducing provocations, and removing excuses is helpful in understanding the crime prevention mechanisms that generate the observed crime reduction results. In this entry, the emergence of pulling levers focused deterrence strategies to control gang violence and the application of these approaches to individual chronic offenders and overt drug markets is discussed within the context of third-party policing. The crime control mechanisms that seem to be associated with puller levers programs are discussed within the situational crime prevention framework.

The Emergence of Pulling Levers Focused Deterrence Strategies

Pioneered in Boston to halt youth violence, the focused deterrence framework has been applied in many American cities, in part through federally sponsored violence prevention programs such as the Strategic Alternatives to Community Safety Initiative, Project Safe Neighborhoods, and the Drug Market Initiative. In its simplest form, the approach consists of selecting a particular crime problem, such as youth homicide; convening an interagency working group of law enforcement practitioners; conducting research to identify key offenders, groups, and behavior patterns; framing a response to offenders and groups of offenders that uses a varied menu of sanctions (known as "pulling levers") to stop them from continuing their violent behavior; focusing social services and community resources on targeted offenders and groups to match law enforcement prevention efforts; focusing community normative expressions on targeted offenders and groups; and directly and repeatedly communicating with offenders to make them understand why they are receiving this special attention and what the special attention comprises (Kennedy 1997).

The Boston Gun Project was a problem-oriented policing enterprise expressly aimed at taking on a serious, large-scale crime problem – homicide victimization among young people in Boston in the 1990s. The trajectory of the Boston Gun Project, and the resulting Operation Ceasefire intervention, is by now well known and extensively documented (see, e.g., Kennedy et al. 1996; Braga et al. 2001). Briefly, a working group of law enforcement personnel, youth workers, and Harvard researchers diagnosed the youth violence problem in Boston as one of patterned, largely vendetta-like hostility among a small population of chronic offenders and particularly among those involved in 61 loose, informal, mostly neighborhood-based "gangs." These 61 gangs consisted of some 1,300 members, representing less than 1 % of the city's youth between the ages of 14 and 24. Although small in number, these gangs were responsible for more than 60 % of youth homicide in Boston.

The ability of the City of Boston to deliver a meaningful violence prevention intervention was created by convening an interagency working group of line-level personnel with decision-making power that could assemble a wide range of incentives and disincentives (Braga et al. 2002). In addition to the Harvard research team, the Ceasefire Working Group was specifically comprised of members from the Boston Police Department's Youth Violence Strike Force; US Bureau of Alcohol, Tobacco, Firearms and Explosives (ATF); US Attorney's Office (federal prosecutors), Suffolk County District Attorney's Office (local prosecutors), Massachusetts Department of Probation, Massachusetts Department of Parole, Massachusetts Department of Youth Services (juvenile corrections), Boston School Police, City of Boston youth outreach workers, and Boston TenPoint Coalition activist black clergy. As needed, Massachusetts State Police, US Drug Enforcement Administration, Youth Service Providers Network, Boston Public Schools, and other agencies were also involved in particular interventions. It was also important to place on the group a locus of responsibility for reducing violence. Prior to the creation of the Ceasefire Working Group, no one in Boston was directly responsible for developing and implementing a comprehensive strategy for reducing violence (Braga et al. 2002).

The Operation Ceasefire focused deterrence strategy was designed to prevent violence by reaching out directly to gangs, saying explicitly that violence would no longer be tolerated, and backing up that message by "pulling every lever" legally available when violence occurred (Kennedy 1997). The chronic involvement of gang members in a wide variety of offenses made them, and the gangs they formed, vulnerable to a coordinated criminal justice response. The authorities could disrupt street drug activity, focus police attention on low-level street crimes such as trespassing and public drinking, serve outstanding warrants, cultivate confidential informants for medium- and long-term investigations of gang activities, deliver strict probation and parole enforcement, seize drug proceeds and other assets, ensure stiffer plea bargains and sterner prosecutorial attention, request stronger bail terms (and enforce them), and bring potentially severe federal investigative and prosecutorial attention to gang-related drug and gun activity.

Simultaneously, youth workers, probation and parole officers, and later churches and other community groups offered gang members services and other kinds of help. These partners also delivered an explicit message that violence was unacceptable to the community and that "street" justifications for violence were mistaken. The Ceasefire Working Group delivered this message in formal meetings with gang members (known as "forums" or "call-ins"), through individual police and probation contacts with gang members, through meetings with inmates at secure juvenile facilities in the city, and through gang outreach workers. The deterrence message was not a deal with gang members to stop violence. Rather, it was a promise to gang members that violent behavior would evoke an immediate and intense response. If gangs committed other crimes but refrained from violence, the normal workings of police, prosecutors, and the rest of the criminal justice system dealt with these matters. But if gang members hurt people, the working group concentrated its enforcement actions on their gangs.

The Ceasefire "crackdowns" were not designed to eliminate gangs or stop every aspect of gang activity but to control and deter serious violence. To do this, the working group explained its actions against targeted gangs to other gangs, as in "this gang did violence, we responded with the following actions, and here is how to prevent anything similar from happening to you." The ongoing working group process regularly watched the city for outbreaks of gang violence and framed any necessary responses in accord with the Ceasefire strategy. As the strategy unfolded, the working group continued communication with gangs and gang members to convey its determination to stop violence, to explain its actions to the target population, and to maximize both voluntary compliance and the strategy's deterrent power.

A large reduction in the yearly number of Boston youth homicides followed immediately after Operation Ceasefire was implemented in mid-1996. A US Department of Justice (DOJ)-sponsored evaluation of Operation Ceasefire revealed that the intervention was associated with a 63 % decrease in the monthly number of Boston youth homicides, a 32 % decrease in the monthly number of shots-fired calls, a 25 % decrease in the monthly number of gun assaults, and, in one high-risk police district given special attention in the evaluation, a 44 % decrease in the monthly number of youth gun assault incidents (Braga et al. 2001). The evaluation also suggested that Boston's significant youth homicide reduction associated with Operation Ceasefire was distinct when compared to youth homicide trends in most major US and New England cities (Braga et al. 2001).

The assertion that pulling levers focused deterrence strategies generate noteworthy violence prevention gains has been strengthened by subsequent replications of the Boston experience. A number of cities have experimented with the pulling levers framework to control gang violence and experienced some very encouraging results. Consistent with the problem-oriented policing approach, these cities have tailored the approach to fit their violence problems and operating environments. Quasi-experimental research designs have revealed noteworthy violence prevention gains in East Los Angeles, California (Tita et al. 2004); Indianapolis, Indiana (McGarrell et al. 2006); Lowell, Massachusetts (Braga et al. 2008); and Stockton, California (Braga 2008).

The emergence of the pulling levers focused deterrence approach to controlling violent gangs and other criminally active groups can be most directly traced to the increasing use of problem-oriented policing and situational crime prevention techniques by police departments over the course of the 1990s to address violent crime problems. The devastating harms generated by the 1990s youth violence epidemic helped to push criminal justice agencies, especially police departments, toward developing innovative violence prevention strategies (Weisburd and Braga 2006). However, like third-party policing more broadly, it is important to recognize that the development of pulling levers strategies can also be placed in the context of a regulatory shift in policing that occurred during the 1990s which flowed from a changing climate of governance and regulation as well as a blurring of civil and criminal laws (Mazerolle and Ransley 2005). As will be discussed further below, the formation of interagency working groups that seek to regulate the behavior of high-risk groups through the routine application of enforcement levers mirrors the creation of an actuarial system, where the public police are one node in a network, responsible for identifying and managing risks.

Applying Pulling Levers Focused Deterrence to Individual Offenders

A variation of the Boston model was applied to individual offenders in Chicago, Illinois as part of the US Department of Justice-sponsored Project Safe Neighborhoods (PSN) initiative (Papachristos et al. 2007). Gun- and gang-involved parolees returning to selected highly dangerous Chicago neighborhoods went through "call-ins" where they were informed of their vulnerability as felons to federal firearms laws, with stiff mandatory minimum sentences; offered social services; and addressed by community members and ex-offenders. A quasi-experimental evaluation showed a neighborhood-level homicide reduction impact of 37 % (Papachristos et al. 2007). Individual-level effects were remarkable. For offenders with gun priors but no evidence of gang involvement, for example, attendance at a call-in reduced return to prison at around 5 years from release from about half to about 10 % (Fagan et al. 2008).

Applying Pulling Levers Focused Deterrence to Overt Drug Markets

There is less experience in applying the focused deterrence approach to other crime and disorder problems. In High Point, North Carolina, a focused deterrence strategy was aimed at eliminating public forms of drug dealing such as street markets and crack houses by warning dealers, buyers, and their families that enforcement is imminent (Kennedy 2009). With individual "overt" drug markets as the unit of work, the project employed a joint police-community partnership to identify individual offenders, notify them of the consequences of continued dealing, provide supportive services through a community-based resource coordinator, and convey an uncompromising community norm against drug dealing. This application of focused deterrence is generally referred to as the "Drug Market Intervention" (DMI) strategy.

The DMI seeks to shut down overt drug markets entirely (Kennedy 2009). Enforcement powers are used strategically and sparingly, employing arrest and prosecution only against violent offenders and when nonviolent offenders have resisted all efforts to get them to desist and to provide them with help. Through the use of "banked" cases, the strategy makes the promise of law enforcement sanctions against dealers extremely direct and credible, so that dealers are in no doubt concerning the consequences of offending and have good reason to change their behavior. The strategy also brings powerful informal social control to bear on dealers from immediate family and community figures. The strategy organizes and focuses services, help, and support on dealers so that those who are willing have what they need to change their lives. Each operation also includes a maintenance strategy.

A preliminary assessment of the High Point DMI found noteworthy reductions in drug and violent crime in the city's West End neighborhood (Frabutt et al. 2004). A more rigorous evaluation of the High Point DMI is currently being conducted. Quasi-experimental evaluations of a similar DMI strategy in Rockford, Illinois (Corsaro et al. 2009), and Nashville, Tennessee (Corsaro et al. 2010), found noteworthy crime prevention gains associated with the approach. Study findings suggest that the DMI strategy generates statistically significant and substantive reductions in crime, drug, and nuisance offenses in the targeted neighborhoods.

The Links Between Third-Party Policing, Situational Crime Prevention, and Pulling Levers Focused Deterrence Strategies

Third-party policing and pulling levers focused deterrence strategies are strongly linked as both approaches seek to harness the crime control and prevention work of the police to a network of third-party partners. As suggested by Braithwaite (2000), recent transformations of governance in Western democracies have resulted in a move toward networks of power and away from state sovereignty and control. These broader societal changes have influenced crime control and policing by "changing the focus from state responsibility for preventing and correcting criminal behavior to a system where crime control and prevention networks are responsible for identifying and managing risks" (Mazerolle and Ransley 2006, p. 193). In pulling levers strategies, police work within a "network of capacity" to prevent gang and group-involved violence (Moore 2002; Braga and Winship 2006). These networks are well positioned to control risky group behavior because criminal justice agencies, community groups, and social service agencies coordinate and combine their efforts in ways that magnify their separate crime reduction effects.

When implemented properly, pulling levers focused deterrence strategies combine the "top-down" approach to crime control and prevention suggested by the third-party policing perspective with the "grassroots" approach of careful analysis and response development suggested by situational crime prevention and problem-oriented policing perspectives. As documented by Mazerolle and Ransley (2006), third-party policing initiatives often, but not exclusively, occur in the context of situational crime prevention and problem-oriented policing projects. In Boston, for instance, the Ceasefire Working Group represented a system of decentralized networks of governmental and nongovernmental agencies that mobilized to prevent and control outbreaks of serious gang violence. At the same time, the Ceasefire Working Group members carefully analyzed the nature of gang disputes and the vulnerabilities of the gangs involved to a variety of sanctions and implemented a set of violence reduction actions that were customized to the targeted gangs.

Situational crime prevention and problem-oriented policing are analytic approaches that have often, either explicitly or implicitly, taken the position that traditional law enforcement is not a desirable option in dealing with recurring crime problems (Goldstein 1990; Clarke 1997). Although situational crime prevention and pulling levers focused deterrence strategies developed separately, these approaches are complements in crime prevention and control. Focused deterrence strategies involve many elements that, at face value, appear very traditional and are thus hard to see and understand in the new light in which these strategies seek to use them. By identifying the links between the situational crime prevention and focused deterrence strategies, many of these new prevention ideas are made clearer to those seeking to these approaches to deal with crime problems. This remainder of this section builds upon prior work that locates focused deterrence approaches within a situational prevention framework (Braga and Kennedy 2012; Skubak Tillyer and Kennedy 2008).

Situational crime prevention and third-party policing strategies both seek to change criminal decision making. Third-party policing initiatives exploit the coercive nature of a range of alternative civil, criminal, and regulatory rules and laws levers to persuade (or force) the decision change. Put another way, the use of the law within third-party policing interventions is a coercive way to alter offender decisions, whereas situational crime prevention interventions seem to encourage offenders to make a more rational decision by weighing the risks and benefits of committing a crime. The common ground in influencing offender decisions shared by these two perspectives is evident in the discussion below. The core situational crime prevention areas of increasing the effort, increasing risks, reducing rewards, reducing provocations, and removing excuses provide a familiar and easy-to-understand framework to summarize the varying crime prevention mechanisms at work in pulling levers focused deterrence strategies. The discussion below is also limited to decisions to commit a criminal act, such as committing a gun homicide, rather than decisions to become involved in crime or decisions to persist or desist from criminal offending (Cornish and Clarke 2003).

Increasing Risk

Deterrence theory posits that crimes can be prevented when the costs of committing the crime are perceived by the offender to outweigh the benefits (Zimring and Hawkins 1973). Most discussions of the deterrence mechanism distinguish between "general" and "special" deterrence (Cook 1980). General deterrence is the idea that the general population is dissuaded from committing crime when it sees that punishment necessarily follows the commission of a crime. Special deterrence involves punishment administered to criminals with the intent to discourage them from committing crimes in the future. Much of the literature evaluating deterrence focuses on the effect of changing certainty, swiftness, and severity of punishment associated with certain acts on the prevalence of those crimes (see, e.g., Apel and Nagin 2011). Pulling levers focused deterrence strategies honor core deterrence ideas, such as increasing risks faced by offenders, while finding new and creative ways of deploying traditional and nontraditional law enforcement tools to do so, such as directly communicating incentives and disincentives to targeted offenders (Kennedy 1997, 2009).

Increasing the risks faced by offenders to change their decisions to commit particular crimes is a central idea in both frameworks. Pulling levers strategies aimed at "gangs," for example, are designed to alter the objective sanction environment by increasing enforcement risks faced by groups of offenders and accurately and directly communicating that change to a specific population of offenders in a manner that enhances the legitimacy of the strategy. Ongoing intelligence analysis reduces the anonymity of offending groups in the commission of their crimes. Triggering events, such as committing a homicide, result in a swift and certain enforcement response by law enforcement coupled with informal sanctions by community members intended to weaken pro-offending norms. Partnering criminal justice agencies mobilize a varied set of sanctions that are tailored to the specific crime dynamics associated with the offending group and to the varied offending behaviors of the individuals that comprise the targeted group. Beyond halting crime outbreaks by groups, the pulling levers focused deterrence strategy seeks to prevent further offending by making it risky for criminal group members to encourage pro-offending norms and narratives within their social networks.

In addition, more traditional situational crime prevention techniques are often implemented as part of the core pulling levers work in practice. Extending guardianship, assisting natural surveillance, strengthening formal surveillance, reducing the anonymity of offenders, and utilizing place managers can greatly enhance the range and the quality of the varying enforcement and regulatory levers that can be pulled on offending groups and key actors in criminal networks. In High Point, North Carolina, the interagency working group conducted call-ins with landlords of drug houses to ensure that these place managers dealt with drug problems on their properties immediately or face certain and swift law enforcement and civil code violation penalties. Similarly, the Cincinnati Initiative to Reduce Violence, a multiagency community collaborative effort to apply focused deterrence principles to reduce group-based violence, used civil forfeiture techniques to close down a highly problematic bar that generated recurring serious violence.

A particularly interesting development within focused deterrence is the realization that the "guardians" so central to traditional situational crime prevention can themselves be offenders: who nonetheless can be brought to exercise powerful influences over other offenders. For instance, in Lowell, the interagency working group recognized that they could systematically prevent street violence among Asian street gangs by targeting the gambling interests of older, influential members. When a street gang was violent, the Lowell Police targeted the gambling businesses run by the older members of the gang. The enforcement activities ranged from serving a search warrant on the business that houses the illegal enterprise and making arrests to simply placing a patrol car in front of the suspected gambling location to deter gamblers from entering. The working group coupled these tactics with the delivery of a clear message, "when the gang kids associated with you act violently, we will shut down your gambling business. When violence erupts, no one makes money" (Braga et al. 2006, p. 40). These possibilities have led to the idea of deliberately mobilizing even serious offenders as "intimate handlers" of other offenders along particular dimensions of their offending portfolio.

Reducing Rewards

In many cities facing serious gun violence problems, chronic disputes among gangs and other criminally active groups generate the bulk of homicides and nonfatal shootings (Braga et al. 2002). These disputes are often personal and vendetta-like. In these settings, violence and a willingness to be violent often confer a very desirable status on particular members within group memberships and hierarchies. The rewards of violent behavior further encourage pro-offending norms and narratives within these criminal relationships and networks. Pulling levers strategies attempt to reverse this problem by associating swift and certain negative consequences with behaviors, such as committing homicides and shootings, that once enhanced the reputation of individuals within a group and the group itself. The communication strategies explicitly make a "cause and effect" link between the actions of interagency working groups and the targeted criminal behavior by the group.

Pulling levers focused deterrence strategies have extended the concept of denying the benefits of crime beyond those accrued by individual offenders to include these powerful group processes associated with repeat violent offending (Skubak Tillyer and Kennedy 2008). Nonetheless, there is considerable overlap and cross-fertilization between focused deterrence strategies and the various intervention ideas associated with particular situational prevention actions designed to reduce the rewards associated with offending. Ideas involving concealing and removing targets can be seen in the efforts of interagency working groups to use enforcement strategies, such as enhancing and monitoring probation conditions to limit their presence in high-risk places at high-risk times, and community-based actions, such as peace walks through gang turfs, to alter the presence of gang members in public places who are often the "targets" of retaliatory violence. Disrupting illegal drug markets can also produce desirable reward-reducing effects that can counter outbreaks of violence. Limiting the ability of group members to earn money when violence erupts can certainly reduce any benefits accrued to the group by violent behavior. Ideally, the violence that once won the respect of other group members would now bring anger and disassociation by one's peers.

Reducing Provocation

In pulling levers strategies, deterrent messages are framed to address the group context from which many crime problems emerge. The groups themselves can act as another internal communication vehicle for transmitting the actual sanction risk to other offenders. As Skubak Tillyer and Kennedy (2008) describe, meaningful enforcement actions and scrutiny by law enforcement agencies can leverage the rationality of group members to no longer encourage norms that provoke the outbreaks of violence. The citywide communication of the antiviolence message, coupled with meaningful examples of the consequences that will be brought to bear on groups that break the rules, can weaken or eliminate the "kill or be killed" norm as individuals recognize that their enemies will be operating under the new rules as well.

This perspective on reducing group-based provocations to commit violence developed as the Boston Gun Project unfolded. A central hypothesis within the working group was the idea that a meaningful period of substantially reduced youth violence might serve as a "firebreak" and result in a relatively long-lasting reduction in future youth violence (Kennedy et al. 1996). The idea was that youth violence in Boston had become a self-sustaining cycle among a relatively small number of youth, with objectively high levels of risk leading to nominally self-protective behavior such as gun acquisition and use, gang formation, tough "street" behavior, and the like: behavior that then became an additional input into the cycle of violence. If this cycle could be interrupted, a new equilibrium at a lower level of risk and violence might be established, perhaps without the need for continued high levels of either deterrent or facilitative intervention. The larger hope was that a successful intervention to reduce gang violence in the short term would have a disproportionate, sustainable impact in the long term.

P

Removing Excuses

Sampson et al. (1997) emphasize the capacity of a community to realize common values and regulate behavior within it through cohesive relationships and mutual trust among residents. They argue that the key factor determining whether crime will flourish is a sense of the "collective efficacy" of a community. A community with strong collective efficacy is characterized by high capacities for collective action for the public good. Pulling levers focused deterrence strategies enhance collective efficacy in communities by emphasizing the importance of engaging and enlisting community members in the strategies developed. The High Point DMI strategy, for example, drew upon collective efficacy principles by engaging family, friends, and other "influential" community members in addressing the criminal behaviors of local drug dealers (Kennedy 2009).

Community-based action in pulling levers strategies helps to remove the excuses used by offenders to explain away their responsibility for the targeted behavior. In call-ins and on the street, community members effectively invalidate the excuses for criminal behavior by challenging the norms and narratives that point to racism, poverty, injustice, and the like. In Boston, for example, black clergy challenged gang members who attempted to use these excuses by countering that poverty, racism, and injustice were not linked to their decisions to fire shots in their neighborhoods and kill other young people who have experienced the same societal ills and life difficulties. Community members also work with law enforcement and social service agencies to set basic rules for group-involved offenders such as "don't shoot guns" and, hopefully, to alert the conscience of these offenders by appealing to moral values inherent in taking the life of another, causing harm to their neighborhood, or the pain that would be experienced by their mothers if they were killed or sent to prison for a long time in a very far away location.

The social service component of focused deterrence strategies serves as an independent good and also helps to remove excuses used by offenders to explain their offending. Social service providers present an alternative to illegal behavior by offering relevant jobs and social services. The availability of these services invalidates excuses that their criminal behavior is the result of a lack of legitimate opportunities for employment, or other problems, in their neighborhood.

A recent elaboration of these ideas has been to use persons of standing to explicitly challenge the "street code" (Anderson 1999) that in many places drives much or most "street" violence. This code has identifiable, if unwritten, elements with powerful criminogenic implications: that early death is inevitable; jail and prison nothing to fear; disrespect must be met with violence; the enemy of my friend is my enemy; the failings of mainstream others justify violent offending; and the like. A long sociology of "techniques of neutralization" (Matza 1964) puts such exculpatory norms at the center of the etiology of street violence. When, for example, "original gangsters" respected by current offenders undercut such norms, it may in turn undercut felt justification for violence and other offending.

Situational crime prevention also seeks to reduce offending by making it easier for potential offenders to comply with laws and rules. Focused deterrence strategies also do so, in some very particular ways. To the extent that street norms stand against compliance, altering them for the better facilitates compliance: the law stands against violence, but the street honor code promotes it, so vitiating the honor code makes it easier to obey the law. Many potentially violent offenders do not wish to be violent but are under very real peer pressure to be so; undercutting and even reversing group and network dynamics make it easier for them to back off. Strengthening community norms against offending makes it easier to resist both internal and external pressure to offend. The original Boston project framed this "honorable exit" from violent offending as a possible consequence of changes in objective sanction risks and group dynamics.

A recent and potentially very important strain in focused deterrence thinking has focused explicitly on the gulf, often explicitly racialized, in the "norms and narratives" of law enforcement, affected communities, and offenders. These gulfs, rooted in historical and present experiences, understandings, and misunderstandings, can have powerful impacts on the ways in which these parties understand each other and on the perceived legitimacy of law enforcement, law, and informal rules and standards. If law enforcement is seen by the community and offender groups as a deliberate racial oppressor, for example, it will be difficult for community norms to support stands against drug dealing and easier for offenders to break the drug laws. If those norms and narratives are effectively addressed – which usually includes changes in actual behavior on both sides – it becomes easier for community standards to support compliance, easier for offenders to comply, and harder for them not to (Kennedy 2009; Meares 2009; Tyler 1990).

These developments draw heavily on a rich literature addressing the "legitimacy" of law enforcement in the eyes of those subject to the law (Meares 2009; Tyler 1990). Very interestingly, one of the core findings in that literature is that even offenders obey the law most of the time. This empirical fact, of considerable theoretical and practical significance, stands in contrast to situational crime prevention's establishing assumption of the "motivated offender": the idea that we can take the presence of the impulse to offend as a given and for granted. That establishing assumption is just that – an assumption explicitly deployed as a way to break with traditional criminology's pervasive and, for applied purposes not very helpful, search for the deep causal roots of individual offending. The emerging salience of legitimacy as an idea of profound practical significance could well lead to a reexamination, and refinement, of the idea of offending in situational crime prevention.

Increasing Effort

Situational crime prevention seeks to increase the effort a criminal must expend to complete a crime through a variety of specific actions such as target hardening, controlling access to facilities, screening exits, deflecting offenders, and controlling tools and weapons (Clarke 1997; Cornish and Clarke 2003). In preventing and responding to outbreaks of serious violence by groups of offenders, many of these situational actions can be used to good effect in developing a "thicker" response to the offending groups in question. Focused deterrence responses benefit from changing the environments in which gangs and drug crews congregate, commit lesser offenses together, and launch plans for retaliation. Gang members are well known to commit a wide array of crimes (known as "cafeteria-style offending"; see Klein 1995). Making it more difficult to sell drugs at specific places, commit burglaries and car breaks, and rob passersby can be a powerful deterrent response to particular groups of offenders who rely on these illicit opportunities to generate income. Disrupting the ability of groups to associate with each other in space and time, through, for instance, controlling access to and screening exit from public housing, can slow down group processes that facilitate violence and make it more difficult for rivals to locate other group members targeted for retaliatory shootings.

Focused deterrence strategies seek to create similar dynamics. Changes in group norms and in objective risks associated with particular forms of misbehavior may, for example, make it more difficult to recruit peers for particular instances of co-offending. Ethnographic research on illicit gun markets in Chicago has shown that gangs' assessment of the law enforcement responses to gun violence leads them to withhold access to firearms for younger and more impulsive members (Cook et al. 2007). DMI's goal of fundamentally disrupting overt drug markets can greatly enhance the difficulty of drug dealing: when buyers no longer routinely "cruise" once active markets, even a motivated street dealer may find it impossible to do business.

Conclusion

The evaluation evidence briefly presented in this entry suggests that pulling levers focused deterrence strategies generate noteworthy crime reduction gains. A recently completed Campbell

Collaboration systematic review identified eleven evaluations of pulling levers strategies (Braga and Weisburd 2012). The basic findings of this review were very positive; their main effect meta-analysis reported a statistically significant, medium-sized crime reduction effect associated with the reviewed programs. Nonetheless, the authors were concerned with the lack of rigorous randomized experimental evaluations of this promising approach. Only one of the reviewed interventions was tested using a randomized experimental design.

Despite their concerns over the lack of randomized experiments, Braga and Weisburd (2012) concluded that the findings of eligible focused deterrence evaluations fit well within existing research suggesting that deterrence-based strategies, if applied correctly, can reduce crime (Apel and Nagin 2011). The pulling levers approach seems to have the desirable characteristic of altering offenders' perceptions of sanction risk. Braga and Weisburd (2012) also noted, however, that pulling levers strategies contained other complementary crime control mechanisms at work that needed to be highlighted and better understood. Applying pulling levers actions to the core set of base situational prevention ideas is very helpful in unraveling the varied pathways through which these programs produce the observed crime reduction impacts.

While originally developed from a problem-oriented project, pulling levers focused deterrence approaches share some direct links with third-party policing initiatives. The successful application of pulling levers strategies to high-risk groups, individuals, or places is heavily reliant on the coordinated actions of a network of criminal justice, social service, and community-based agencies. The police are an important, but singular, node within this network. Pulling levers strategies seek to change the behavior of targeted groups and individuals by engaging a wide range of incentives and disincentives; drawing on third-party policing activities, these "levers" can include the application of criminal and civil laws and be supported through the formation of willing (and unwilling) partnerships.

Related Entries

- ► Civil Remedies
- ► Evidence-Based Policing
- ► Focused Deterrence and "Pulling Levers"
- ► Gangs and Social Networks
- ► Situational Crime Prevention
- ► Third Party Policing and School Truancy

Recommended Reading and References

Anderson E (1999) Code of the street: decency, violence, and the moral life of the inner city. Norton, New York

Apel R, Nagin D (2011) General deterrence: a review of recent evidence. In: Wilson JQ, Petersilia J (eds) Crime and public policy. Oxford University Press, New York

Braga AA (2008) Pulling levers focused deterrence strategies and the prevention of gun homicide. J Crim Justice 36:332–343

Braga AA, Kennedy DM (2012) Linking situational crime prevention and focused deterrence strategies. In: Farrell G, Tilley N (eds) The reasoning criminologist: essays in honour of Ronald V Clarke. Taylor & Francis Publishing Group, London

Braga AA, Weisburd DL (2012) The effects of focused deterrence strategies on crime: a systematic review and meta-analysis of the empirical evidence. J Res Crime Delinq 49:323–358

Braga AA, Winship C (2006) Partnership, accountability, and innovation: clarifying Boston's experience with pulling levers. In: Weisburd D, Braga A (eds) Police innovation: contrasting perspectives. Cambridge University Press, New York

Braga AA, Kennedy DM, Waring EJ, Piehl AM (2001) Problem-oriented policing, deterrence, and youth violence: an evaluation of Boston's operation ceasefire. J Res Crime Delinq 38:195–225

Braga AA, Kennedy DM, Tita G (2002) New approaches to the strategic prevention of gang and group-involved violence. In: Huff CR (ed) Gangs in America, 3rd edn. Sage, Thousand Oaks

Braga AA, McDevitt J, Pierce GL (2006) Understanding and preventing gang violence: problem analysis and response development in Lowell, Massachusetts. Police Q 9:20–46

Braga AA, Pierce GL, McDevitt J, Bond BJ, Cronin S (2008) The strategic prevention of gun violence among gang-involved offenders. Justice Q 25:132–162

Braithwaite J (2000) The new regulatory state and the transformation of criminology. Br J Criminol 40:222–238

Clarke RV (ed) (1997) Situational crime prevention: successful case studies, 2nd edn. Harrow and Heston, Albany

Cook PJ (1980) Research in criminal deterrence: laying the groundwork for the second decade. In: Morris N, Tonry M (eds) Crime and justice: an annual review of research, vol 2. University of Chicago Press, Chicago

Cook PJ, Ludwig J, Venkatesh S, Braga AA (2007) Underground gun markets. Econ J 117:558–588

Cornish D, Clarke RV (2003) Opportunities, precipitators and criminal decisions: a reply to Wortley's critique of situational crime prevention. In: Smith MJ, Cornish D (eds) Theory for practice in situational crime prevention, vol 16, Crime prevention studies. Criminal Justice Press, Monsey

Corsaro N, Brunson R, McGarrell E (2009) Problem-oriented policing and open-air drug markets: examining the Rockford pulling levers strategy. Crime Delinq doi:10.1177/0011128709345955. Accessed 14 OCT 2009

Corsaro N, Brunson R, McGarrell E (2010) An evaluation of the Nashville drug market initiative (DMI) pulling levers strategy. Michigan State University, School of Criminal Justice, East Lansing

Fagan J, Meares T, Papachristos AV, Wallace D (2008) Desistance and legitimacy: effect heterogeneity in a field experiment with high-risk offenders. Paper presented at the annual meeting of the American Society of Criminology, St. Louis, MO, Nov

Frabutt J, Gathings MJ, Hunt ED, Loggins T (2004) High point West End initiative: project description, log, and preliminary impact analysis. Center for Youth, Family, and Community Partnerships, Greensboro

Goldstein H (1990) Problem-oriented policing. Temple University Press, Philadelphia

Kennedy DM (1997) Pulling levers: chronic offenders, high-crime settings, and a theory of prevention. Valparaiso Univ Law Rev 31:449–484

Kennedy DM (2009) Drugs, race, and common ground: reflections on the high point intervention. Natl Inst Justice J 262:12–17

Kennedy DM, Piehl AM, Braga AA (1996) Youth violence in Boston: gun markets, serious youth offenders, and a use-reduction strategy. Law Contemp Probl 59:147–196

Klein M (1995) The American street gang: its nature, prevalence, and control. Oxford University Press, New York

Matza D (1964) Delinquency and drift. Wiley, New York

Mazerolle L, Ransley J (2005) Third party policing. Cambridge University Press, New York

Mazerolle L, Ransley J (2006) The case for third party policing. In: Weisburd D, Braga A (eds) Police innovation: contrasting perspectives. Cambridge University Press, New York

McGarrell E, Chermak S, Wilson J, Corsaro N (2006) Reducing homicide through a lever-pulling strategy. Justice Q 23:214–229

Meares T (2009) The legitimacy of police among young African-American men. Marquette Law Rev 92:651–666

Moore MH (2002) Creating networks of capacity: the challenge of managing society's response to youth violence. In: Katzmann G (ed) Securing our children's future: new approaches to juvenile justice and youth violence. Brookings Institution Press, Washington, DC

Papachristos AV, Meares T, Fagan J (2007) Attention felons: evaluating project safe neighborhoods in Chicago. J Empir Leg Stud 4:223–272

Sampson R, Raudenbush S, Earls F (1997) Neighborhoods and violent crime. Science 277:918–924

Skubak Tillyer M, Kennedy DM (2008) Locating focused deterrence approaches within a situational crime prevention framework. Crime Prev Community Saf 10:75–84

Tita G, Riley KJ, Ridgeway G, Grammich C, Abrahamse A, Greenwood P (2004) Reducing gun violence: results from an intervention in East Los Angeles. RAND Corporation, Santa Monica

Tyler T (1990) Why people obey the law. Yale University Press, New Haven

Weisburd D, Braga AA (eds) (2006) Police innovation: contrasting perspectives. Cambridge University Press, New York

Zimring F, Hawkins G (1973) Deterrence: the legal threat in crime control. University of Chicago Press, Chicago

Punishment

► Punishment as Rehabilitation
► Sentencing Research

Punishment and the State

Alon Harel
Faculty of Law, The Hebrew University of Jerusalem, Jerusalem, Israel

Overview

Criminal sanctions are typically inflicted by the state and by the state alone. This entry investigates (by using tools of political philosophy) the normative rationales for the exclusive control of the state over the infliction of criminal sanctions, and further, it explores whether the state has not only an exclusive right to punish but also a duty to do so.

Criminal law sanctions require the enactment of criminal law norms, the issuing of particular judgments by courts, and, last, the physical infliction of criminal sanctions. All those functions are currently controlled by the state. Yet some legal systems were almost entirely private and contemporary theorists challenge the public control over some aspects concerning the infliction of criminal punishment. Most significantly, some recent legal reforms involve privatization of central components of the criminal law system: private prisons, private enforcement mechanisms, shame penalties, etc. This entry examines whether it is a desirable (or even a necessary) feature of criminal punishment that it be inflicted by the state and, if so, why? Could the infliction of punishment be, at least in principle, privatized and, if it could, would it be desirable?

To do so, the entry differentiates among three types of justifications for the role of the state in the infliction of punishment: *instrumentalist justifications*, *normative precondition justifications*, and *state-centered justifications*. It provides examples for each type and it shows how classical theories of punishment such as those provided by Locke, Nozick, and Rousseau fall into each one of these categories. The entry also investigates whether the state has sometimes a duty rather than merely a right to punish.

Introduction

Criminal sanctions are typically inflicted by the state. The criminal law norms that set a maximal sanction or a range of permissible sanctions are enacted by the legislature; state courts determine the sanctions within the boundaries set by the legislature; and last, the state is (typically) in charge of the physical infliction of the sanction, e.g., running the prisons (maintained by public officials). Note that while typically the legislative determination, the judicial determination, and the executive infliction of the sanction are all done by the state, it is possible to differentiate these three dimensions and privatize some of them while maintaining public control of others. The central concern of this entry is to examine whether it is a desirable (or even a necessary) feature of criminal punishment that it be inflicted by the state and, if so, why? Could the infliction of punishment be, at least in principle, privatized and, if it could, would it be desirable? The entry also investigates whether the state has sometimes a duty rather than merely a right to punish.

This entry is divided into three main parts. The first Part explains why the relations between punishment and the state raise an important question of policy. This part establishes that the intimate relations between the state and punishment are neither natural nor universal. As a matter of fact, there is currently a trend towards privatizing different dimensions of the criminal process. The second Part differentiates among three types of justifications for the role of the state in the infliction of punishment: *instrumentalist justifications*, *normative precondition justifications*, and *state-centered justifications*. Last, the third Part examines whether the state has not merely an (exclusive) right to inflict criminal punishment but also a (political and/or constitutional) duty to do so and, if so, how such a duty can be justified.

Can the Infliction of Criminal Punishment Be Privatized?

Some may find the relation between criminal punishment (and the penal process as a whole) on the one hand and the state on the other hand to be natural and self-evident. Law students learn that criminal law is a branch of public law designed to promote the public interest. Public law is contrasted with private law (e.g., contract and tort law) which is designed to promote private interests (Hall 2005, pp. 241–246). Under this view held, most famously, by Blackstone, crimes “affect the whole community, considered as a community in its social aggregate capacity” (Blackstone 1769 4th book, Chap. 1). More fundamentally, the state has, as Max Weber said, monopoly of the legitimate use of physical force (Weber 1946, p. 77). Consequently, all the dimensions of the criminal law and its execution including legislation of criminal prohibitions, the determination of guilt, the size of the sanction, the enforcement system, the prosecution, and the infliction of the criminal sanctions including the maintaining of prisons are public and are governed by public law.

Yet a historical examination reveals that the public nature of these functions is “self-evident” to lawyers only because lawyers “incline to project modern practice backwards” (Langbein 1973, p. 316). Historically, it is known that for a very long time, English law relied on private prosecution (Hay 1984; Friedman 1995, pp. 475–478). As a matter of fact, “private vengeance of the person wronged by a crime” was a primary feature of the administration of criminal justice in early English law (Stephen, 245). This system has been transformed only in the nineteenth century (Kurland and Waters 1959). While the English system was based on private prosecutors, other jurisdictions relied heavily on the private infliction of criminal sanctions. The southern states of the USA had no actual penitentiary and till the second half of the nineteenth century, prisoners were leased out by the state to private entrepreneurs (Lichtenstein 2001 p. 193). Another example of private infliction of sanctions is the rule prevailing in Roman law under which the person who committed theft and was caught in the act was given up as a slave to the person against whom the theft was committed (Jolowicz and Nicholas 1972, p. 167). While many ancient legal systems recognized the difference between offenses against the state or community and offenses against the individual, the meaning of the distinction and the scope of offenses covered by each category were very different than the ones known to us now. Offenses which are now regarded exclusively as crimes were treated as private wrongs (torts) including, for instance, theft in Roman law (Maine pp. 369–371). Last, some legal systems were almost entirely private. In recent research on medieval Icelandic institutions, it was said that Icelandic institutions “might almost have been invented by a mad economist” (Friedman 1979, p. 400). In the Icelandic system that operated successfully for hundreds of years, “Killing was a civil offence resulting in a fine paid to the survivors of the victim. Laws were made by a ‘parliament,’ in which seats were a marketable commodity. Enforcement of law was entirely a private affair” (Friedman 1979, 400). Iceland is not unique in this respect. As Friedman argues, “[T]he idea that law is primarily private, that most offenses are offenses against specific individuals or families, and that punishment of the crime is primarily the business of the injured party seems to be common to many early systems of law” (Friedman 1979, p. 400; Posner 1983, pp. 119–143).

These historical observations can be described in different ways. Some may say that the Icelandic legal system did not contain *penal* sanctions (as penal sanctions are “necessarily” or “by definition” public), while others may say that the penal sanctions were privatized in Iceland. Irrespective of how these systems are described, it is evident even to contemporary legal theorists that the division between public sanctions (criminal ones) and private sanctions (torts) is not a sharp division and that the dichotomy drawn between these two systems does not reflect the legal reality (Mann 1992). More importantly, these historical observations are not merely relics of history. Contemporary theorists, in particular economists, have also favored the “privatization” of some aspects of the penal system. It was argued that a system of private enforcement of law, in which the person who caught a criminal received the fine paid by the offender, would be more efficient than the current system in which the fine is paid to the state (Becker and Stigler 1974; Friedman 1979). Further, some “private” penal practices are being reintroduced in a new form. The victim who in the past played a major role in the criminal process both as a policeman and as a prosecutor has now become again influential figure in the criminal process (Mcdonald 1975–76; Levine 2010; Cassell and Joffee 2011). Shaming penalties used extensively in medieval times have become popular again in the USA (Kahan 1996). Although not often noticed, shaming penalties privatize the infliction of punishment as the power to determine the severity of the sanction, and more generally, the power to inflict sanctions is shifted from the state to private citizens (Whitman 1998, p. 1089). Last, the privatization of prisons has triggered a bitter debate (McDonald 1994), and in at least one country, private prisons were declared by the courts to be unconstitutional. (http://elyon1.court.gov.il/files_eng/05/050/026/n39/05026050.n39.htm; Medina 2010). These historical and contemporary findings establish that there is nothing inevitable in there being public control over the penal process. The public nature of punishment is ultimately a normative decision on the part of the polity that has been gradually eroded and its desirability ought to be subjected to normative scrutiny. The rest of this entry will be devoted to examining the justifications for the involvement of the state in the infliction of punishment.

The Justification for Public Infliction of Criminal Sanctions

This section identifies three types of justifications for the central role of the state in inflicting punishment (Harel 2008; Harel and Poart 2011). Under the first type of justification – *instrumental justifications* – the agent which inflicts the sanction must be the agent most capable of inflicting the sanction. Such an agent must make accurate judgments as to the moral gravity of the offense as well as judgments as to the appropriate sanctions, be willing and capable to invest the resources necessary to punish the offender, and have the ability to act on the basis of its judgment. Further, under an instrumental justification, the characterization of who the most capable agent is (or the most successful infliction is), is independent of the identity of the entity in charge of performing it. The state is an appropriate agent to inflict the sanction simply because (and only to the extent) that it is (likely to be) deliberative, impartial, or efficient. Under a second type of justification – *normative precondition justifications* – the infliction of criminal sanctions by the state achieves goals that could in principle be fully realized through the infliction of sanctions by non-state agents. Yet, in contrast to instrumental justifications, there are normative constraints that preclude the infliction of sanctions by agents other than the state. The agency of the state is not necessary for the success of the infliction of criminal sanctions; it is however a non-contingent normative precondition for the just infliction of criminal sanctions. Under a third type of justification – *the state-centered justification* – the infliction of criminal punishment by non-state agents is conceptually impossible, as the very infliction of criminal punishment presupposes that it is the state that inflicts it. This is not mere semantics. Under the third type of justification, punishment brings about certain goods and the goods resulting from punishment hinge on the identity of the agent inflicting it. To be considered "punishment" under this view, it is not sufficient that it hurts or causes displeasure to those who have wronged (as those can result from the actions of a private agent); it must also reflect an authoritative judgment concerning the wrongfulness of an act, and such a judgment can only be made by an authoritative entity – the state. It is false therefore to say that private individuals *ought not* to punish; they simply *cannot punish* as their acts do not constitute punishment.

Instrumental justifications consist of two major steps. First, the theorist identifies the goals of punishment: deterrence/incapacitation, retribution, etc. These goals are conceptually separable from the identity of the agent inflicting the sanction. Then the theorist establishes that the state is the most likely to successfully realize or bring about these goals. To the extent that the state is more likely/capable of realizing these goals than other agents, it ought to be in charge of inflicting criminal sanctions.

The most influential instrumental justification was made by John Locke. Although ultimately, as shown below, Locke defends a non-instrumental account, Locke developed an influential instrumentalist account. More specifically, Locke believed that the state should be empowered to inflict sanctions on those who transgress the laws of nature because the state is less partial than other agents in its treatment of offenders (Locke 1960, chap. 2). Locke believed that punishment inflicted by individuals is both possible and permissible in the state of nature (Farell 1988). But he also believed that private agents inflicting sanctions are likely to be either too lenient ("Self-Love will make Men partial to themselves and their Friends") or too harsh ("Passion and

Revenge will carry them too far in punishing others"). Law and economic theorists also defend the involvement of the state on the basis of instrumental justifications. In their view, punishment should be supplied by the state because the infliction of sanctions involves a collective action problem. An individual who inflicts a sanction has to bear the costs of inflicting the sanction himself, whereas the benefits resulting from the infliction (deterrence, incapacitation) are enjoyed by everybody. It follows that individuals will underinvest in the infliction of sanctions (Mueller 1989, pp. 9–15) Other theorists also established that private agents are less likely to be accountable than public officials and that often they may act in ways that do not promote the goals of government (Verkuil 2007, chap. 1). The lack of accountability on the part of private agents is often debated in the literature (Minow 2003; Trebilcock and Lacobucci 2003). At the same time, other economists believe that the state is not always the best agent to inflict punishment. It has been argued, for instance, that states are inclined to increase sanctions above the optimal level in order to "export" their criminals elsewhere (Harel 2012, pp. 34–35).

Note that instrumentalist justifications could be based on any theory of punishment. Instrumentalists could support retributivism and argue that the state is the agent that is most likely to inflict a sanction that is proportionate to the gravity of the wrong (and is therefore "deserved" by the criminal). Deterrence theorists could argue that the state is the most likely to inflict a sanction that deters effectively or, at least, to do it efficiently. Rehabilitation theorists would maintain that the state is more likely to succeed in integrating the criminal into society etc.

Instrumentalist arguments require establishing factual claims concerning the state. Once the goals of punishment are identified, it is necessary to show that the state makes better judgments concerning the "appropriate" sanctions, is better able to inflict the sanction, is more likely to calibrate their optimal size, etc. Further, if faced with contrary evidence, an advocate of an instrumentalist argument must be willing to concede that punishment ought to be inflicted by non-state entities. There are no principled grounds why the state ought to inflict criminal sanctions. The desirability of privatization hinges therefore on contingent factors and such factors may change from time to time and from society to society.

This observation provides a basis for criticizing the instrumentalist view. Arguably, the instrumentalist fails to capture a prevalent intuition, namely, that the involvement of the state in the infliction of punishment is not a mere contingency hinging on its (alleged) ability to "get it right." Those who challenge instrumentalist views maintain that "it is not generally accepted that I have the right simply to hurt another who has done something wrong, just because he has done it, where there is no special relation between us" (Lyons 1976, p. 210). Further, "neither one's general level of virtue nor one's particular talents in the area of punishing. . . are normally taken to establish any special claim to be the one who should punish others (Simmons 1991, p. 312) Even Locke (who developed also an instrumentalist account described above) agrees with this observation and says: To justify bringing such evil [i.e., punishment] on any man two things are requisite. First that *he who does it has commission and power to do so*. Secondly, that it be directly useful for the procuring of some other good. . .Usefulness, when present, being but one of those conditions, cannot give the other, which is a commission to punish" (Locke 1823, emphasis added).

Under this challenge, one ought to separate two questions: (a) whether an agent "deserves" to be punished or ought to be punished and (b) whether a particular agent is the "appropriate agent" to inflict the sanction (Hill 1999). To justify the infliction of punishment X by an agent Y on a wrongdoer Z, it is not sufficient to show that X should be inflicted on Z and that Y is the most likely or the most capable to inflict it. Something else needs to be shown.

This powerful intuition is evident in the context of family relationship. A child may deserve punishment, or punishment may be conducive to the child's well-being or, all things considered, it may be desirable that the punishment be inflicted. The common law has recognized the parents' privilege of moderate chastisement. But, according to the common law, the only agent for whom it is permissible to inflict the sanction on the child would be a parent or, in special cases, somebody to whom the parent delegated the power, e.g., teachers (Blackstone Book I, Chap. 16). Of course such a power granted to parents could be justified purely on instrumental grounds, e.g., that parents are more likely to inflict the "right" punishment, but it seems that beyond the purely instrumental grounds, non-instrumental concerns are also at stake (Harel 2008, p. 123).

Both normative constraints theories and state-centered theories reject the instrumental view and attempt to explain why sanctions ought to be inflicted only by "appropriate" agents and appropriateness, in their view, is not merely a contingent fact. Under the normative constraints explanations, the infliction of sanctions by "inappropriate" agents may serve the purposes of punishment (whatever these purposes are). Yet there are normative constraints on the infliction of sanctions that are not grounded in instrumental concerns. Even when the punishment inflicted by an agent X is impartial, effective, and just, it could be illegitimate and wrong simply because it is inflicted by X rather than by another agent. Under the state-centered explanations, sanctions that are imposed by the "wrong" agent do not constitute punishment at all as they fail to realize the goods that punishment is designed to realize.

Locke and Nozick defend what I labeled a normative constraints theory. Under their view, individuals in the state of nature have a right to punish (designed to prevent further wrongdoing on the part of the wrongdoer herself and on the part of others) (Locke 1960, chap. 2; Nozick 1977, pp. 3–119). Further, individuals not only have a right to punish but also a right to transfer/alienate the right to punish. Last, both Locke and Nozick believe that we can attribute to individuals a decision to transfer the right to punish to the state (based on their own concern to protect their natural rights and the greater effectiveness of the state in calibrating sanctions). The authority of the state to punish is attributable not merely to the "usefulness" of the state but also to its "commission to punish" based on the voluntary (or constructive) alienation of the right to punish. Similarly Rousseau maintains that "The punishment of death inflicted on criminals may be considered from the same point of view: it is to secure himself from being the victim of assassins that a man *consents to die if he becomes an assassin*" (Rousseau 1954, p. 31, emphasis added). Punishment inflicted by private agents is therefore impermissible not merely because it is likely to be disproportionate or unjust but also because the power to punish was transferred voluntarily (or should be regarded as if it was transferred voluntarily) by those who have a natural right to punish (i.e., individuals) to the state.

Normative constraints justifications can address the concern raised earlier against instrumentalist justifications; in particular, they can explain why the power of the state to punish does not merely hinge on contingent considerations such as the fact that the state is more likely than any other agent to make better judgments concerning the appropriate size, the type of the sanctions and/or to be better able to inflict them, etc. Under the normative constraints justifications favored by Locke or Rousseau, it is not sufficient to maintain that the state "does it better." In addition, it is necessary to establish that the individual upon whom the sanction is being inflicted agreed or consented to the infliction of the sanction by the state. The nature and the conditions required for establishing such consent on the part of individuals are different, but this condition imposes restrictions on the agents who can punish.

State-centered justifications go further and argue that the power to punish is an agent-dependent power; only the state can punish, that is, realize the goals of punishment. State-inflicted sanctions are designed to realize goals or perform tasks that cannot, in principle, be performed successfully by private institutions or individuals. To develop a state-centered justification, it is necessary to develop a theory of punishment – a theory that will explain what is particularly valuable about punishment and, then, establish that the only agent capable of realizing this value is the state.

A useful analogy is the blood feud (Harel 2008, p. 121). Anthropologists have found that blood feuds can only be performed by male relatives of the deceased (Barmash 2005, p. 24; Miller 1983, pp. 162–168). In a blood feud, it is not the mere act of killing or the death of the member of the enemy clan that counts; it is rather the performance of the killing by an "appropriate agent." The agent killing the murderer in a blood feud is not perceived as a mere means to perform the allegedly just act of killing. Instead, it is the act of killing that provides an opportunity for the appropriate agent to act in order to redress the injustice. Even closer to our concerns here is the view of Joel Feinberg who believes that "punishment expresses the *judgment* of the community that what the criminal did was wrong" (Feinberg 1994, p. 76). Although Feinberg did not envision the possibility of privatizing punishment, it seems that he could maintain that the infliction of sanctions by non-state agents may be effective in many ways; it may deter wrongdoers, incapacitate criminals, satisfy retributivist concerns, etc., but private sanctions fail to punish as they do not express the judgment *of the community* that the act is wrong.

Punishment, under such a view, is a public manifestation of condemnation and disapprobation of the criminal deed. As Nozick argues: "[r]etributive punishment is an act of communicative behaviour" and, further, Nozick believed that retribution achieves two goals. The first is "to connect the criminal to the value qua value," and the second is to connect the wrongdoer to the value in a way "that value qua value has a significant effect in the criminal's life, as significant as his own flouting of correct values" (Nozick 1981, p. 370). The view is shared also by prominent criminal law theorists who believe criminal law can only be understood as a communicative practice (Duff 2001). Unlike deterrence, and perhaps other conventional goals of punishment, public condemnation is possible in the first place only if it emanates from an appropriate agent. If a wild dog attacks me while I am stealing my neighbor's property, it does not "punish" me as the dog did not make a judgment concerning the wrongfulness of my behavior. For similar reasons, if a child or a madman hits me while I am committing a wrong, the act does not constitute punishment. Mere pain or harm resulting from committing a wrong is not sufficient to constitute punishment; the pain or harm must result from a considered judgment made by an agent capable/authorized of making such a judgment.

More generally, condemnation/judgment of a wrongful act presupposes an agent who is in a privileged status to the one subjected to the condemnation, viz., one whose judgments concerning the appropriateness of behavior are worthy of attention or respect. The wrong agent fails to punish not because it fails to inflict harm – or even inflict harm resulting from a wrongful act – but because it cannot make the appropriate judgments. In the absence of such judgments, the mere infliction of "a sanction" amounts to an act of violence (Dorfman and Harel 2013, pp. 92–96).

To complete the argument, one needs to show that the state is the only agent which can make judgments of the relevant sort. Given certain assumptions, this challenge can be met. Arguably, the state is a legitimate authority whose judgments concerning the wrongfulness of the behavior count. Further, no other entity is capable of forming judgments of the right type. The judgments of the state may count either because its judgments are more likely to be correct or true (such that conforming to these judgments is more likely to lead agents to act in accordance with reason) (Raz 1986, chap. 3) or because its decisions reflect its will and its will has normative force, or for other reasons. In any case, one may draw the conclusion that the state is an agent that can effectively condemn certain types of behavior and make judgments with respect to them while other agents cannot.

A close reading of Kant's discussion of punishment in the *Metaphysics of Morals* establishes that Kant supported a state-centered justification for criminal sanctions. In a famous passage, Kant says:

> Even if a civil society were to be dissolved by the consent of all its members…, the last murderer remaining in prison would first have to be executed, so that each has done to him what his

> deeds deserve and blood guilt does not cling to the people for not having insisted upon this punishment; for otherwise the people can be regarded as collaborators in this public violation of justice. (Kant 1996, p. 106)

Kant insists in this passage that the last murderer has to be executed *before* the dispersion of society. But why is the execution of the murderer so urgent; why could it not be postponed? Perhaps Kant believes that an attempt to remedy the injustice by executing the murderer after the dispersion of society would fail to punish for this would constitute a private act of killing rather than a public act of execution and a private killing could not be done in the name of the people as a collectivity (Harel 2008, 122).

In examining the relations between punishment and the state, we did not distinguish among the legislative act (setting the maximum sentence or the range of permissible sentences for a wrong), the adjudicative act (inflicting a punishment for a particular wrong), and the execution of the sentence, e.g., maintaining a prison. It is not surprising that privatizing the execution of the punishment, e.g., privatizing prisons, seems the least controversial. This is so because it seems that privatizing the prisons does not really count as privatizing the punishment itself, but merely the manner or the technicalities of its infliction. Under this view, there is no reason to care whether prisons are run by public officials or by private contractors (as long as legislatures and courts are the ones who dictate what the punishment is).

This claim however is not self-evident. As the petition challenging the privatization of prisons in Israel indicated (Medina 2010), private prison wardens are not mere technocrats executing state decisions. They regularly make decisions that affect the welfare of inmates such as holding a prisoner in administrative isolation, carrying out an external examination of an inmate's naked body, and using reasonable force to carry out a search. These decisions require the exercise of judgment and, arguably, they ought to be made only by agents capable of exercising such judgments in the name of the state, namely, by public officials and not by private individuals, as only the exercise of judgment on the part of the former count as the judgments made by state.

To sum up, while the instrumentalist justifications maintain that the state ought to be in charge of the infliction of sanctions to the extent that it is capable of inflicting "appropriate" sanctions in an efficient manner, both normative constraints justifications and state-centered justifications believe that the state is uniquely placed to punish. State-centered justifications also maintain that the very value of punishment hinges on the identity of the agent inflicting it. It is not that it is impermissible for non-state agents *to punish*; it is rather that no other agent *can punish* and any attempt to punish on the part of such agents is bound to fail and constitute a mere act of violence.

Does the State Have a Duty to Punish

Section "The Justification for Public Infliction of Criminal Sanctions" explored arguments concerning the right of the state to punish and, in particular, whether and, if so, why this right is exclusively that of the state. However, this leaves open an additional central question concerning the role of the state, namely, whether the state has a (political or constitutional) duty (rather than merely a right) to punish (Hill 1999).

To say that only the state has a right to punish may seem too weak without acknowledging in addition the existence of a duty to punish on the part of the state. Can the state simply fail to criminalize murder or theft (or to refuse to punish those who commit these crimes) simply if it wishes to do so? Further, even if the state has such a duty, should the duty to criminalize be constitutionally entrenched or should the decision be left to the state? Intuitively, it seems that such decisions ought not to be left to the good will of the state. To, explore these questions, let me start with a real-life case – the constitutional decision of the German Constitutional Court concerning abortion.

In 1975, the German Constitutional Court declared that the law which allowed abortion on demand during the first trimester of pregnancy was unconstitutional as it violated Article 2 section 2 of the German Basic Law protecting the right to life (39 BVerFG 1 1975; Kommers 1985, p. 371; Lange 2011, p. 2033). The Constitutional Court emphasized that abortion is an act of killing that the law is obligated to condemn through the use of criminal law. Under this view, the state had not only the right but also the (constitutional) duty to criminalize abortion (Kommers 1985, 395). As a result of reunification, the German Constitutional Court had to address the matter again, and in its later 1993 decision, the Court reiterated its commitment to the view that abortion is indeed a violation of the right to life and that the state has an obligation to protect life, including the life of the fetus (88 BVerfGE 203). Further, the Court believed that protecting life requires also criminalizing abortion. It followed that abortion ought to remain criminal although the state could substitute "normative counseling" for criminal punishment as a way of fulfilling its obligations to protect fetuses.

The German Constitutional Court adhered to the view under which the state has a (both moral and constitutional) duty to criminalize abortion. To further investigate, the soundness of this claim assume that evidence is provided to the Bundestag convincing it that decriminalizing abortion would, as a matter of fact, reduce the rate of abortions in society and, consequently, more lives would be saved if abortions were decriminalized. On the basis of this evidence, the Bundestag concludes that while fetuses have a right to life, criminalizing abortions is detrimental to the protection of the right to life and, consequently, it declares that decriminalizing abortion is constitutional.

Such a conclusion would not be absurd; yet it is not self-evident either. Arguably, even if the state could better protect life if it decriminalized murder, it ought not to do it. State punishment is not merely an instrument to protect life, freedom of expression, etc. Protecting a right, so it could be argued, involves also public condemnation of violations of the right. Punishment inflicted by the state is as much about condemning the violation of the right to life as it is about the protection of life itself. Moreover, failure to criminalize abortion implies that the life of the fetus is "at the mercy" of the mother, and this seems wrong even under the assumption that criminalization does not reduce the number or frequency of abortions. Another case which may demonstrate the scope of the duty of the state is the case of *X & Y v. the Netherlands* (8 EHRR 235 26 March, 1985) in which the European Court of Human Rights decided that the state has a positive duty to charge criminally a person who allegedly committed rape even when the victim who was mentally disabled could not bring charges against him.

Further, it could be argued that the duty to criminalize is not merely a political duty on the part of the state; arguably it ought also to be a constitutional duty. If the state has no constitutional duty to criminalize, the life of the fetus, so it could be argued, is "at the mercy" of the legislature; it hinges on the inclinations of the legislature. If the legislature decides not to criminalize abortion, the fetus' life is at the mercy of the inclinations of the mother. Even if the legislature criminalizes abortion, the life of the fetus hinges on the inclinations of the legislature, namely, on its willingness to criminalize. Arguably, it is only a constitutional duty to criminalize that can remedy these defects.

The decision of the German Constitutional Court can be rationalized precisely on the grounds that protecting the fetus' right to life is not to be left either to the mercy of the pregnant woman or to the inclinations of the legislature. The legislature must not merely criminalize abortion; it must be constitutionally obliged to do so as it is not "up to the legislature" to make decisions concerning the life of the fetus.

More generally, in the absence of entrenching a constitutional duty to criminalize violations of basic rights, the mere willingness of a legislature to criminalize can be interpreted as grounded in its good will. In contrast, the constitutional duty to criminalize does not leave the decision to criminalize to the discretion of the legislature. Consequently, upon entrenching such a duty, citizens do not live "at the mercy of" the legislature.

Conclusion

Privatization may be desirable for many reasons. It has been evident to many that the desirability of privatization is grounded in instrumental considerations. Much of the debate concerning privatization focuses on instrumental arguments.

This entry does not urge anyone to ignore or underestimate instrumental considerations, and yet it challenges the assumption that instrumental considerations are all that matter. More specifically, at least in certain contexts, and within certain domains, the instrumental concerns may be too narrow to capture the moral complexity of the considerations at stake. The relations between the state, the criminal process, and criminal sanctions are grounded in principled legitimacy-based considerations touching upon foundational questions of political morality.

These observations are important not only in theory as privatization of different dimensions of the criminal law process is not (as illustrated above) a mere fantasy. Even if a solely privately inflicted scheme of sanctions is not a realistic option, there are current reforms or reform proposals to grant private individuals the power to inflict sanctions for wrongdoing. These proposals have often been initiated and discussed by economists, sociologists, and lawyers. This entry establishes that in addition to the important insights of social science, philosophers can and should contribute to this debate.

Recommended Reading and References

Barmash P (2005) Homicide in the biblical world. Cambridge University Press, Cambridge

Becker G, Stigler G (1974) Law enforcement, malfeasance and compensation of enforcers. J Legal Stud 3:1

Blackstone W (1769) Commentaries on the laws of England

Cassell P, Steven J (2011) The crime victim's expanding role in a system of public prosecution: a response to the critics of the Crime Victims' Rights Act, Northwestern University Law 105: 164

Dorfman A, Harel A (2013) Philosophy and public affairs 41:67

Duff A (2001) Punishment communication and community. Oxford University Press, Oxford

Farell D (1988) Punishment without the state. Nous 22:437

Feinberg J (1994) The expressive function of punishment. In: Duff RA, David G (eds) A reader on punishment. Oxford University Press, New York, p 76

Friedman D (1995) Making sense of English Law enforcement in the eighteenth century. University of Chicago Law School Roundtable 2:475

Friedman D (1979) Private creation and enforcement of law. J Leg Stud 8:399

Hall J (2005) General principles of criminal law, 2nd edn. Lawbook Exchange, Pyrmont

Harel A (2008) The case against privately inflicted sanctions. Leg Theory 14:113

Harel A (2012) Economic analysis of criminal law: a survey. In: Harel A, Hylton K (eds) Research handbook of the economic analysis of criminal law. Edward Elgar, Cheltenham, p 10

Harel A, Poart A (2011) Commensurability and agency. Cornell L Rev 96:749

Hay D (1984) The criminal prosecution in England and its historians. Mod L Rev 47:1

Hill T (1999) Kant on wrongdoing, desert and punishment. L Phil 18:407

Jolowicz HF, Nicholas B (1972) Historical introduction to the study of Roman law, 3rd edn. Cambridge University Press, Cambridge

Kahan D (1996) What do alternative sanctions mean. Chi L Rev 63:591

Kant I (1996) The metaphysics of morals (trans: Gregor M). Cambridge University Press, Cambridge

Kommers D (1985) Liberty and community in constitutional law: the abortion cases in comparative perspective 1985. BYU L Rev, p 371

Kurland P, Waters DWM (1959) Public prosecutions in England 1854–79: an essay in english legislative history. Duke L J 493. 1959

Langbein J (1973) The origins of public prosecution at common law. Am J Leg Hist 17:313

Lange F (2011) American liberalism and Germany's rejection of the national socialist past-the 1973 Roe v. Wade decision and the 1973 German abortion. Case Hist Perspect Ger Law J 12:2033

Levine D (2010) Public wrongs and private rights: limiting the victim's role in a system of public prosecution, NW. U. L. REV. 104 335

Lichtenstein A (2001) The private and public in penal history: a commentary on zimring and tonry. Punishment Soc 3:189

Locke J (1823) A second letter concerning toleration in the works of John Locke. Thomas Davison, London
Locke J (1960) Two treaties of government: second treatise. Laslett P (ed). Cambridge University Press, Cambridge
Lyons D (1976) Rights against humanity. Philos Rev 85:208
Maine H (1866) Ancient law. John Murray, London
Mann K (1992) Punitive legal sanctions: the middleground between criminal and civil law 101 Yale LJ 1795
McDermott D (1995) The duty to punish and legitimate government. J Polit Philos 7:147
McDonald D (1994) Public imprisonment by private means: the re-emergence of private prisons and jails in the United States, the United Kingdom, and Australia. BRIT J Criminol 34:2
McDonald D (ed) (1990) Private prisons and the public interest. Rutgers University Press, New Brunswick
Mcdonald W (1975–1976) Towards a bicentennial revolution in criminal justice: the return of the victim. Am Crim L Rev 13:649
Medina B (2010) Constitutional limits to privatization: the Israeli Supreme Court decision to invalidate prison privatization. Int J Const Law 8:690
Miller WI (1983) Choosing the avenger: some aspects of blood feud in medieval iceland. Law & Hist Rev 1:159–204
Minow M (2003) Public and private partnerships: accounting for the new religion. Harv Law Rev 116:1229
Mueller D (1989) Public choice II. Cambridge University Press, Cambridge
Nozick R (1977) Anarchy state and Utopia. Basic Books, New York
Nozick R (1981) Philosophical explanations. Harvard University Press
Posner R (1983) The economics of justice. Harvard University Press, Cambridge, MA
Raz J (1986) The morality of freedom. Oxford University Press, Oxford
Rousseau JJ (1954) The social contract (trans: Kendall W). Henry Regnery Company, Chicago
Simmons AJ (1991) Locke and the right to punish. Philos Publis Aff 20:311–349
Stephen (1883) A history of the criminal law of England. Macmillan, London
Trebilcock M, Lacobucci E (2003) Privatization and accountability. Harv L Rev 116:1422
Verkuil P (2007) Outsourcing sovereignty: why privatization of government functions threatens democracy and what can be done about it. Cambridge University Press, Cambridge
Weber M (1946) Essays in sociology. Oxford University Press, New York, Translated, Edited and with an Introduction by H.H. Gerthj and C. Wright Mills
Whitman J (1998) What is wrong with inflicting shame sanctions? R L J 107:1055

Punishment as Rehabilitation

Fergus McNeill
Scottish Centre for Crime and Justice Research,
University of Glasgow, Glasgow, UK

Synonyms

Correction; Punishment; Reform; Rehabilitation; Sentencing; Treatment

Overview

This entry explores the relationships between rehabilitation and punishment. It begins by unpacking the different meanings of rehabilitation and some of the major criticisms that have been leveled at rehabilitation. It then goes on to examine the different meanings and purposes of punishment and where and how rehabilitation fits amongst these meanings and purposes. The entry concludes by recognizing ways in which rehabilitation can have punishing effects and by suggesting when and under what conditions punishment is rehabilitative.

Introduction

The title of this entry may seem an odd one. For many people, punishment and rehabilitation are alternatives between which we must choose, rather than potential synonyms or alternative ways of referring to similar processes. Can punishment ever *be* rehabilitation? Don't rehabilitation's supporters tend to see offending as rooted in people's experience of social exclusion and injustice, to regard the trope of individual (criminal) responsibility (on which the legitimacy of punishment ultimately depends) as misconceived, and to stress the duty of the state somehow to fix the mess that produced crime? Punishment, at least for some rehabilitationists, seems little more than a fancy cloak to drape around the otherwise nakedly vengeful instincts of those that think (absurdly, these critics would argue) that there is something sensible, appropriate, and fair about piling hurt on hurt.

And yet, many practitioners of rehabilitation – even those sympathetic to these criticisms of punishment – hold on to a belief in the capacity of people to change themselves and their situations, to make different choices, to overcome their circumstances, and to author a different and better future. Even those who are deterministic when it came to the genesis of criminal behaviors and problems somehow seem to become believers in free will or the power of human agency (even under intense social and structural pressures) when it comes to the future prospects of people undergoing rehabilitation. Moreover, many rehabilitationists also have an acute sense of justice or injustice, reflected in their demands that the state honor its obligations to those whose adverse life experiences have proved "criminogenic."

Readers will already have noted a number of paradoxes and possible contradictions in play here, in terms of how we conceive of individual and political responsibility for crime and other social harms and of how we understand and deliver fairness and justice and in our approaches to the righting of wrongs. To try to unravel some of these threads and make sense of these questions requires an examination of the contested and multiple meanings of these two terms – rehabilitation and punishment – before we can begin to understand at least some of the potential interrelationships between them.

Which Rehabilitation

Both as set of concepts and as a set of practices, rehabilitation is a "fankle." Fankle is a Scots word that translates loosely as "tangle." The unraveling – or at least the teasing apart – of the distinct threads of a fankle is a tiresome and time-consuming job, but a necessary one, if only so that we can knot them together again properly at the end.

Most textbook discussions of rehabilitation begin with dictionary definitions. Raynor and Robinson's (2009: 2) excellent book, for example, tells us that the Oxford English Dictionary defines rehabilitation as "the action of restoring something to a previous (proper) condition or status." So, rehabilitation is (1) an action that (2) restores (3) for the better. Raynor and Robinson (2009: 3) also note that the OED's supplementary definition refers to the "restoration of a disabled person, a criminal etc., to some degree of normal life by appropriate training etc." This version adds the concepts of some (4) "normal" standard, returning to which requires (5) some form of third-party intervention.

Moving on to criminological uses of the term, Raynor and Robinson (2009) argue that, despite the frequency with which (offender) rehabilitation has been discussed in the literature (and, we might add, in policy and practice), it is rarely "unpacked" and critically examined. They provide the following example of a problematic criminological description of it:

> taking away the desire to offend, is the aim of reformist or rehabilitative punishment. The objective of reform or rehabilitation is to reintegrate the offender into society after a period of punishment, and to design the content of the punishment so as to achieve this. (Hudson 2003: 26)

As Raynor and Robinson (2009) note, this statement raises a number of issues. Firstly there seem to be at least two objectives in play here: "taking away the desire to offend" (i.e., somehow changing the offender) and reintegration into society (i.e., somehow changing his or her relationship with and status in society). But how are these two objectives related? Secondly, and more directly pertinent to this entry, it suggests two different relationships between rehabilitation and punishment; in one rehabilitation comes *after* punishment, and in another rehabilitation *shapes* (the nature of) punishment. There may be a third possibility, as suggested above in the introduction, where rehabilitation is cast as an *alternative* to punishment. Finally, this description elides the distinction between rehabilitation and reform; as we will see below, for others, these two ideas are distinct but related concepts.

In an effort to clarify some of these complexities, Raynor and Robinson (2009) go on to offer their own typology of perspectives on offender rehabilitation, examining the meanings and significance of correctional rehabilitation, rehabilitation and reform, reintegration and resettlement, and rehabilitation and the law.

Correctional rehabilitation, they argue, is concerned with effecting positive change in individuals. As such it is the model most commonly associated with treatment programs or other forms of offense- or offender-focused intervention. At its heart is the notion that many offenders can change for the better, given the right support. The idea of correction implies that the offender can and should be "normalized" or "resocialized" in line with commonly accepted (though rarely explicitly articulated) standards of behavior. Raynor and Robinson make the critical but often neglected point that correctional rehabilitation is a very broad church, one that allows for almost as wide a variety of theories and methodologies about how "correction" is to be achieved as there are theories about crime causation itself. That said, correctional rehabilitation, as its name suggests, tends to be preoccupied with *changing offenders* themselves and so is closely associated with theories and methods that explain crime and target intervention at the level of the individual.

Raynor and Robinson's (2009) discussion of *rehabilitation and reform* notes that some penal theorists and historians draw a distinction between twentieth-century "rehabilitation," which was concerned with individualistic (psychological) treatment programs to correct one's personality (or attitudes and behaviors), and "reform" which refers to an earlier preoccupation with offering opportunities for education and contemplation in support of the reform of one's moral character. As a form of shorthand, we might say that religion is to reform as the "psy" disciplines (psychiatry, psychology, and social work) are to rehabilitation.

The terms *reintegration and resettlement* (or in the USA "reentry") may involve or be connected with correctional rehabilitation, but they also extend beyond it; in a sense, they imply its objective. If correctional rehabilitation is the journey, reintegration is the implied destination. Raynor and Robinson (2009) suggest that rehabilitation must lead to and involve restoring the ex-offender's status as a citizen and renegotiating his or her access to its privileges and responsibilities. Whereas correctional rehabilitation dwells on the individual and the psychological, "reintegration and resettlement" signals the sociological aspects of rehabilitation; proponents of such an approach tend to have a sharper awareness of the social rather than the individual causes of crime and of the social rather than the individual issues at stake in desistance from crime. Moreover, rather than being a mode of punishment (or a method applied during punishment), this perspective tends to stress the need for rehabilitation to act as an antidote which seeks to address, compensate for, or seek to undo the adverse, usually unintended, collateral consequences of punishment itself. Rehabilitation in this vein is sometimes conceived a duty of the state that follows from its obligation to delimit punishment and to bring it to an end.

A natural (though again often neglected) corollary of the concern with reintegration and resettlement is a concern with *rehabilitation and the law*. Though we come to it last, the earliest use of the term rehabilitation (at least in legal or criminal justice contexts), according to Raynor and Robinson (2009), was in late seventeenth-century France where it referred to the destruction or undoing of a criminal conviction and to the deletion or expunging of the criminal record. A century or so later, Cesare Beccaria (1764/1963) reflected a similar meaning of the term in arguing for the use of punishment as a way of "requalifying individuals as... juridical subjects." In this sense, punishment itself was meant to be rehabilitative in settling the putative debt that offending created. Rehabilitation was thus an end of punishment in both senses of the word; it was its proper objective *and* its final conclusion. Rehabilitation was the restoration of citizenship, not in the de facto sociological sense discussed above but in the de jure legal sense.

Raynor and Robinson's (2009) taxonomy helps us a great deal with unraveling "the rehabilitation fankle," but there are one or two further strands that we need to identify before moving on to examine which meanings and versions of punishment might be consistent with which meanings and versions of rehabilitation.

Firstly, it is worth stressing that the meanings of rehabilitation are historically, culturally, and jurisdictionally conditioned. As we have already seen, the broadly correctional form of rehabilitation has been expressed in very different ways in different times and places. Two decades ago, Edgardo Rotman (1990), in a brilliant and brief introductory chapter to his book "Beyond Punishment," summarized the history of rehabilitation as being represented in four successive models. The first two of these – the penitentiary model and the treatment model – correspond loosely to our discussion of reform and rehabilitation above. The penitentiary was seen as a place of confinement where the sinner is given the opportunity to reflect soberly on their behavior and on how to reform themselves (note "reform" rather than "rehabilitate"), perhaps with divine help. But this idea was supplanted progressively by a more scientific or medical model in which rehabilitation was understood as a form of treatment which could correct some flaw in individuals, whether physical or psychological, thus remedying the problem of their behavior.

However, in the latter half of the twentieth century, this more medical or therapeutic version of rehabilitation was, according to Rotman, itself displaced to some extent by a shift in emphasis towards a model based on social learning in which behaviors were understood as learned responses that could be unlearned. In this context, rehabilitation was recast not as a sort of quasi-medical treatment for criminality but as the reeducation of the poorly socialized. Correctional rehabilitation's interventions may have remained individualized, but they changed to reflect new theories of crime causation and thus new forms of "treatment." This distinction between these two versions of correctional rehabilitation (treatment vs. social learning) is important, partly because, as we will see in the next section, it was one that was and is often ignored by rehabilitation's critics.

Criticisms of Rehabilitation

As Bottoms (1980) notes, in a typically erudite and compelling chapter that deals with the collapse of the rehabilitative ideal, one cause of its demise was that it came to be seen as being theoretically faulty. It misconstrued the causes of crime as individual when they were coming to be understood as being principally social and structural, and it misconstrued the nature of crime, failing to recognize the ways in which crime is itself socially constructed. Moreover, rehabilitative practices had begun to be exposed as being systematically discriminatory, targeting coercive interventions on the most poor and disadvantaged people in society. Rehabilitation was also coming to be seen as being inconsistent with justice itself because judgments about liberty had come to be unduly influenced by dubious and subjective professional judgments hidden from or impenetrable to the offender. Through the development of the "psy" disciplines, experts emerged with the supposed capacity to "diagnose" what was wrong with the offender, and the offender was cast as a victim of his or her lack of insight. By implication, unless and until the offender was "corrected" by the expert, he or she could not be treated as a subject. It was argued that rehabilitation faced an associated and fundamental moral problem rooted in its attempts to (psychologically) coerce people to change. Finally, at the time when Bottoms was writing, the empirical evidence seemed to suggest that, despite its scientific pretensions, rehabilitation did not seem to work.

Powerful though it is, there are flaws in this critique. Crucially, emerging evidence about "what works?" played a vital role in challenging the last point; there is now an evidence base for rehabilitation, (McNeill, Raynor and Trotter 2010). But that evidence base, in and of itself, does not address the other four criticisms which I will label the problems of *crime theory*, of *rehabilitation and injustice*, of *dubious expertise*, and of *coerced correction*. Contemporary rehabilitation theories perhaps have more in common with a social learning or social psychological than a medicalized version of correctional rehabilitation. But, even if rehabilitation is not now based on a strictly medical model, the full force of Bottoms' (1980) criticisms is only partly deflected, and in this short entry, these criticisms can be discussed only very briefly.

While contemporary correctional rehabilitation's underlying *crime theories* avoid the pathologizing traps of individualistic positivism, they still have to engage somehow with the problem that crime is (at least in part) a social construct. We do not choose to pursue all interpersonal or social harms through criminalization; the explanandum of crime theories is itself socially conditioned. That single insight has profound consequences. It affects and infuses the normative contexts of rehabilitative work (including raising difficult questions about who and what gets selected for penal "correction" and who and what does not), and it creates a series of complex methodological quandaries about how to judge rehabilitation's effectiveness (see McNeill 2012).

The problem of *rehabilitation and injustice*, or of rehabilitation's place in or association with discriminatory justice practices, remains a challenging one, though it is hardly a criticism that applies *only* to rehabilitation and not to other criminal justice practices. That said, the problem for rehabilitation is not just systematic (if unintended) biases in terms of who gets selected for "correction" of which sorts and who gets defined as "incorrigible" but also a more technical problem of the extent to which rehabilitation's resources (principally assessment tools and interventions approaches) are sufficiently sensitive to, for example, gender differences and cultural diversity. Significant though these problems are, they are problems about the proper administration of rehabilitation rather than the concept itself.

The problem of *dubious expertise* is only partly resolved by the development of a more robust evidence base for rehabilitation. While it is true that contemporary approaches, perhaps to varying degrees, tend to formally recognize the need to treat people as active subjects in their own rehabilitation (and not simply as the passive objects of expert intervention), there remains considerable force in critiques both of the professional power of the "psy" disciplines and, more specifically, of the increasing exercise of that power not in "treatment" or "therapeutic" decisions but in influencing decisions about justice itself, specifically in relation to sentencing or release decision making. These problems will be discussed below.

Of course, the enduring problem of *coerced correction* is closely related to that of dubious expertise; coercion is a problem exacerbated in systems where increasing confidence in rehabilitation's effectiveness has coincided with the development of a "toughening up" of penal policy. One apparently minor aspect of this conjunction in England and Wales was the withdrawal of any requirement for consent to accredited program conditions within community sentences. These programs therefore involve rehabilitative treatment without consent; a development that seems to have come to pass without much critique from psychologists whose ethical standards it might seem to compromise. Though it might be said that all punishment is coercive, coercion seems to be a particular problem for correctional rehabilitation, since it seeks to change the individual rather than simply to restrain, confine, or otherwise punish him or her. That suggests a particular form of intrusion into the inner world, even the identity of the subject, in respect of which coercion raises particular moral problems.

Fortunately, Bottoms (1980) also provided some suggestions about how we might best respond to these problems. Reflecting the pessimism of the times, four of his five options imply or require a shift away from rehabilitation, at least as traditionally conceived. However, he also argued that rehabilitation could be rescued through its own correction, that is, by attending more carefully to questions of consent, by committing adequate resources to make it more likely to be effective, and by conducting our rehabilitative activities in ways which are more respectful of liberty. More specifically, the intrusions that rehabilitation imposes on the offender, he suggests, should never be greater than is merited by their offending behavior, placing rehabilitative requirements within the envelope of proportionality.

There is no space here to properly review the renaissance of rehabilitation in the 1980s and 1990s (see Raynor and Robinson 2009; Robinson 2008). However, leaving aside the better known story of the advancement of the evidence base, it is worth recalling that there was also, in the 1980s, a brief flurry of writing about new normative or philosophical approaches to rehabilitation, including Rotman's (1990) work. The "new rehabilitationists" (see Lewis 2005), much as Bottoms suggested, proposed four principles to guide rights-based rehabilitation: the assertion of the duty of the State to provide for rehabilitation, the establishment of proportional limits on the intrusions opposed, the principle of maximizing choice and voluntarism in the process, and a commitment to using prison as a measure of last resort.

However, as Robinson (2008) has cogently argued, in policy and practice (in England and Wales at least), what emerged was not a more rights-based but a more profoundly utilitarian and correctional form of rehabilitation increasingly influenced by the preoccupation with public protection and risk reduction. Under this paradigm, probation officers were required to intervene with or treat the offender to reduce re-offending and to protect the public. What is critical about this shift in focus was that the "client" or intended beneficiary of rehabilitative work is no longer the offender. Rather probation is trying to change offenders to protect the law-abiding. The offender becomes less and less an active participant and more and more an object to be assessed through technologies applied by professionals and compulsorily engaged in structured programs and offender management processes as required elements of legal orders imposed, as already noted, irrespective of consent.

These developments in the relationship between correctional rehabilitation and those "offenders" – *with whom* it engages or, worse, *on whom* it operates – are a particularly important one. The shift from *subject in* to *object of* correctional rehabilitation implies a more coercive and utilitarian conception of rehabilitation: one which leaves the individual liberties and rights of the offender much more vulnerable and one which opens up new prospects of abuses of the "power to punish," justified in the putative public interest. This form of rehabilitation may still be a process of intervention aimed at some public good, but returning to Raynor and Robinson's (2009) initial definitions, the blunting both of its restorative intent and of its commitment to the offender's interests (alongside others' interests), as well as its willingness to impose (rather than renegotiate) "norms" on objectified offenders, represents major causes for concern.

Which Punishment

Of course, it is impossible to make much sense of the relationships between rehabilitation and punishment just by analyzing rehabilitation – notions of punishment must be examined too. Conceptually and practically, punishment is arguably even more "fankled" than rehabilitation. In part, as Hudson's (2003) excellent introductory text both argues and demonstrates, this is because punishment is a quintessentially interdisciplinary subject that compels and requires criminological, legal, philosophical, and sociological scrutiny, as well as raising fundamental political and practical questions.

Tonry's (2006) commanding and authoritative overview of the purposes and functions of sentencing provides a neat framework for analyzing punishment, distinguishing between sentencing's purposes or normative functions (i.e., its justifications), its primary functions (these being the proper distribution of punishment; the prevention of crime; and the communication of threat, censure, and of social norms), its ancillary functions (in contributing to the management of an efficient and effective justice system and in securing legitimacy and public confidence), and its latent functions (the ways in which it reflects self-interest, ideology, and partisanship and how and what it communicates informally). Though he does not explain his framework in these terms, we might suggest that, with notable exceptions, philosophers and jurists tend to study and discuss the normative functions and primary purposes of sentencing, criminologists tend to examine its ancillary functions, and sociologists tend to study its latent functions. Criminological and sociological analyses will be discussed briefly in the next section. For present purposes, it is perhaps most important to clarify firstly where rehabilitation fits amongst the normative justifications of punishment.

As Tonry (2006: 16) argues:

> The overriding normative function of a sentencing system in a society committed to individual liberty, procedural fairness, and limited powers of government is to assure that individuals convicted of crimes receives the sentences that, in principle, they should

At first sight, this statement seems a little vacuous or circular – the point of doing justice is to do justice – but, at least in liberal democracies, it is vital to surface this too easily taken-for-granted assumption. If punishment involves the state in imposing harms on its citizens, then the process by which this is done must be carefully bounded and governed. But of course what is in and "out of bounds" in sentencing depends to some extent on the normative principles that govern it.

Tonry (2006) goes on to expose the enduring tensions between Kantian and utilitarian approaches to punishment. The former are essentially retrospective and concerned with the delivery of retribution or "just deserts," in proportion to the severity of the offense and the blameworthiness of the offender. The latter are essentially prospective and concerned with the delivery of preventive effects; as we noted above, these are increasingly cast as the protection of the public and the reduction of offending and re-offending. These effects can be advanced through deterrence, incapacitation, or rehabilitation. Tonry (2006: 18) notes that there "are good reasons why no relatively philosophically pure sentencing system [on either model] exists or has been seriously proposed at a policy level." He goes on to illustrate the practical problems of both Kantian and utilitarian approaches and reviews a range of attempts to construct credible hybrid theories. Whatever the conceptual merits of these theories however, there is no doubt that in practice hybrid systems abound, though this is not to say that these systems are clearly articulated, coherent, or defensible. One common feature of most hybrid systems is a commitment to some form of "limiting retributivism," in which culpability sets upper limits for the severity of any punishment to be imposed. The intrusions of punishment, in other words, must never be greater than are merited by the offender's culpability (and perhaps by the gravity of the offense).

As already noted, this limiting principle surfaced in the work of the new rehabilitationists of the 1980s and 1990s, requiring the intrusions of rehabilitation to be bounded by proportionality. Though this does represent a significant and suitable response to some of the criticisms of rehabilitation discussed above, for advocates and practitioners of rehabilitation, the direct engagement with questions of justice and deserts that it entails has often been uncomfortable. Historically, at least in some jurisdictions, probation and criminal justice social workers have often considered themselves as providers and advocates of (usually rehabilitative) *alternatives to punishment*, rather than as providers and advocates of *alternative punishments*. Somehow the notion of punishing, as opposed to supporting, supervising, treating, or helping – or even challenging and confronting – has seemed inimical to the ethos, values, and traditions of probation and social work.

More recently, a different approach to the justification of punishment has perhaps offered the prospect of resolving some of these tensions and of conceiving of punishment *as* rehabilitation (Duff 2005). The penal philosopher Antony Duff – whose theory of punishment as communication has been highly influential – has argued convincingly that we can and should distinguish between "constructive punishment" and "merely punitive punishment" (Duff 2003). Constructive punishment can and does involve the intentional infliction of pains, but only insofar as this is an inevitable (and intended) consequence of "bringing offenders to face up to the effects and implications of their crimes, to rehabilitate them and to secure... reparation and reconciliation" (Duff 2003: 181). Duff (2005) is not arguing for rehabilitation as moral education; to do so would infantilize offenders as well as calling into question their culpability for their crimes. Rather than restoring or correcting a deficit *in the offender*, moral rehabilitation aims to repair a breach *in the relationships* between the offender, the victim, and the community that the offense has created. Repairing the breach is principally the offender's responsibility; the starting point is a sincere apology which represents the recognition of the wrong done and the relational breach that it has created. But in most criminal cases an oral apology – however sincere – will not be enough:

> . . .something more is required – something that will make more forceful the apology that I owe. . .; something that will show that I am taking seriously the need to avoid such wrongs in future, and to reform myself and my conduct. Two elements of such moral rehabilitation might be, first, undertaking some burdensome task to express my apologetic repentance. . .; second, taking steps to address the causes or sources of my wrongdoing, and perhaps seeking help – whether informal or formal – in dealing with them. (Duff 2005: 19)

Duff (2005) admits that he is speaking of criminal justice as it could be, or as it should be, rather than as it is. Nonetheless, he sees in some contemporary practices at least the opportunity for this kind of penal communication and for moral rehabilitation to take place:

> The burden that the offender is required to undertake, as his punishment, can be seen as constituting a formal, and forceful, apology to his victim and to the wider community. The apology has something of the quality of a public ritual rather than of a sincere expression of personal feelings, though we may hope that it will become sincere; but it serves to make clear to the offender the wrong that he has done, and for which he owes and is required to offer this apology, and to make clear to the victim our shared recognition of that wrong. Furthermore, a probation order involves, as probation officers often put it, an attempt to. . . help the offender to confront the character and implications of his crime, and to find ways of avoiding repeating it, partly by programmes that seek to address offending behavior and its causes: by undertaking such programmes the offender is also making apologetic reparation for his crime. (Duff 2005: 19)

Duff's work also helps with a second problem, already referred to above, since he recognizes that where social injustice is implicated in the genesis of offending, the infliction of punishment (even constructive punishment) by the state is rendered morally problematic. The state itself can be seen as complicit in the offending through having failed in its prior duties to the "offender," for example, as is so often the case, where the offender was once a child in the care of the state. For this reason, Duff suggests that probation officers or social workers should play a pivotal role in mediating between the offender and the wider polity, holding each one to account on behalf of the other. Here, Raynor and Robinson's (2009) discussion of the relationships between reintegration, resettlement, and rehabilitation is relevant – the rehabilitation worker's attempt to advocate for the ex-offender and to secure access (perhaps for the first time) to the full rights of citizenship acts as the counterbalance to direct work with the offender to invite and support their moral rehabilitation. This discomfiting, mediating space is one which many probation and social workers will recognize that they occupy and through which, with or without official or public support, they seek to promote social justice within criminal justice.

There are other conceptual and empirical links that could be developed here, but to which we can only briefly refer. Duff's work also resonates with the rise of reparation or paying back as an important justification of punishment (see Robinson et al. forthcoming). Intriguingly, there is some empirical evidence that "making good" is important for many of those desisting from crime (Maruna 2001). In a sense, the relevance in this body of evidence of the concept of "generativity" – referring to the human need to make some positive contribution, often to the next generation – hints at the links between paying back and paying forward, in the sense of making something good out of a damaged and damaging past. Bazemore's (1998) work on "earned redemption" examines more directly the tensions and synergies between reform and reparation, and the broader movement around restorative justice (Johnstone and Van Ness 2007) provides possible normative frameworks within which to further debate and develop these tensions and synergies.

When Rehabilitation Is Punishment

Before trying to reach some conclusions about the relationships between rehabilitation and punishment, it is necessary to turn the entry's title around and explore "when rehabilitation is punishment" or perhaps to ask more directly "does rehabilitation hurt?"

Duff's answer to the question is a qualified "yes." There is no sincere apology and therefore no real moral rehabilitation, without at least the pain of confronting and admitting the wrong and perhaps also the incidental pains of the work required to repair the breach in relationships. But beyond this principled position, there is also a growing body of empirical scholarship which shows that rehabilitation (of the therapeutic or correctional sort) also hurts in other ways, some of which may be less productive and less defensible than the pains Duff has in mind.

There is, of course, an extensive body of scholarship, in the tradition of Gresham 1958) seminal work, on the pains of imprisonment. More recent iterations of this literature point not just to the peculiar pains of life in "supermax" conditions (King 2005) but also to the pains and harms of mass incarceration itself (Haney 2005) and to the pains and harms suffered by prisoners' partners and families through "secondary prisonization" (Comfort 2007).

Perhaps more directly pertinent here are the pains of penal rehabilitation in its current risk-focused guise. Thus, for example, there is some evidence of burgeoning resentment amongst English prisoners towards what apparently seems to them to be the capricious and illegitimate exercise of "soft power" by prison psychologists involved in key decisions about prisoner progression or release (Crewe 2009). In similar vein, Lacombe's (2008) ethnographic study of prison-based sex offender programs reveals the ways in which risk-based rehabilitation invites sex offenders to contort their perceptions and presentations of self in line with the requirements of the particular program or process to which they are subject. Cox's (2012) compelling ethnographic analysis of the pains of youth imprisonment reveals a similar picture.

These pains of risk-based rehabilitation extend beyond the prison too. The traditional absence of punitive intent in probation or criminal justice social work (noted above) does not necessarily entail an absence of "penal bite," at least if studies of those subject to community supervision are to be taken seriously. For example, researchers at the RAND Corporation in the USA found that there are intermediate sanctions which surveyed prisoners equate with prison in terms of punitiveness. For some individuals, intensive forms of probation "may actually be the more dreaded penalty" (Petersilia and Deschenes 1994: 306). More recently, Durnescu (2011) has specifically explored the "pains of probation" as experienced in Romania. Alongside deprivations of time and the other practical and financial costs of compliance, and limitations on their autonomy and privacy, probationers also reported the pain of the "forced return to the offense" and the pain of a life lived "under a constant threat." The threat in question in Durnescu's (2011) study was that of breach or revocation and with it further punishment, but the works referred to in the last paragraph also point to the threat of failing to persuade a probation officer, a psychologist, or some other professional that one's "riskiness" can be and is being properly addressed and managed.

In one sense, there is little that is truly novel in this. For a prisoner of the Eastern State Penitentiary in the nineteenth century, placed in silent and solitary confinement in the hope that he or she would repent and make his or her peace with his or her Maker, the pains of penitentiary reform were doubtless profound. His or her project of "coercive soul transformation" may have been designed and delivered somewhat differently than that directed at the late-modern risk-bearing prisoner or probationer – but both are subjected to disciplinary regimes. But what may be peculiarly demanding for the late-modern penal subject is that, rather than being left to deal, before God, with his own sinfulness and redemption, he or she is compelled to *display* the malleability of his or her riskiness and to *perform* the reduction and the manageability of his or her riskiness. At least in some risk-based systems, it is the credibility of this performance which will determine progression in and release from punishment. In those circumstances, rehabilitation is both disciplinary and punishing in a particularly potent way see (Crewe 2012).

Perhaps presaging these developments, Edgardo Rotman (1994), whose work we referred to above, draws an important distinction between anthropocentric and authoritarian rehabilitation:

> The authoritarian model of rehabilitation is really only a subtler version of the old repressive model, seeking compliance by means of intimidation and coercion. Rehabilitation in this sense is essentially a technical device to mould the offender and ensure conformity to a predesigned pattern of thought and behaviour... The anthropocentric or humanistic model of rehabilitation, on the other hand, grants primacy to the actual human being rather than metaphysical fixations or ideologies, which long served to justify the oppressive intervention of the state. Client centred and basically voluntary, such rehabilitation is conceived more as a right of the citizen than as a privilege of the state. A humanistic public policy regarding crime implies the idea of human perfectibility, which at the level of rehabilitation includes not only the offenders themselves, but also the society that bred them and the institutions and persons involved in their treatment. (Rotman 1994: 292)

This distinction, and more specifically its implication that the person engaged in rehabilitation must be treated as a moral subject and not as a material object to be manipulated or adjusted in the interests of others, seems central to many of the claims that can be made for and against rehabilitation – and to the question of its relationships with punishment.

Conclusion: When Is Punishment Rehabilitation

Returning finally to Raynor and Robinson's (2009) discussion of rehabilitation reviewed at the outset of this entry, it is now possible to discern four main forms or meanings of rehabilitation. In a recent paper, McNeill (2012) has called these psychological, judicial, social, and moral rehabilitation. These are the four strands of rehabilitation that must be "de-fankled" in order to analyze their interrelationships and mutual dependencies.

Psychological rehabilitation means essentially what Raynor and Robinson (2009) refer to as correctional rehabilitation: that which seeks to somehow change or restore the offender, to develop new skills or abilities, and to address and resolve deficits or problems. A better, less loaded term, might be "personal rehabilitation," since this need not imply that any specific disciplinary perspective or set of techniques is implied in this project of personal change.

The second form of rehabilitation concerns the practical expression of Beccaria's concern with the requalification of citizens; this is *judicial rehabilitation* – which raises questions of when, how, and to what extent a criminal record and the formal stigma that it represents can ever be set aside, sealed, or surpassed. Maruna (2011) has recently argued cogently that efforts to sponsor rehabilitation and reform must address the collateral consequences of conviction – mostly notably its stigmatizing and exclusionary effects – or be doomed to fail. No amount of correctional or psychological or personal rehabilitation and no amount of supporting offenders to change themselves can be sufficient to the tasks and challenges of reintegration, resettlement, and reentry, if legal and practical barriers are left in place.

But these barriers are not just legal – they are moral and social too. A solely correctional, psychological, or personal conception of rehabilitation is inadequate to the moral and social offense that crime represents. In simple terms, doing something for or to or (better) with the offender, even something that aims at somehow changing them so as to reduce future victimization, fails to engage with other key aspects of dispensing justice. Perhaps most importantly in moral terms, psychological rehabilitation offers no moral redress per se; it operates only on the individual "offender," not on the conflict itself and not on the victim or the community. In Duff's (2005) terms, it leaves the relational breach unrepaired. Critically, reparation – and reparative work in particular – seems capable of fulfilling this function in ways in which psychological rehabilitation in and of itself cannot, perhaps principally because reparation seems better able to convey (not least visibly) that redress is being actively provided. Though, as Duff (2005) reminds us, willingly undergoing psychological rehabilitation can convey the sincerity of the offender's apology and of his or her desire to change, it is more typically a professionalized, private, and secretive business, and not one that is explicitly bound into a process of moral rehabilitation.

Reparation perhaps speaks to the insistence that moral demands have to be satisfied, and moral communication secured, before *moral rehabilitation* can be recognized (Duff 2001; 2003, 2005). In simple terms, an offender has to "pay back" or to "make good" before he or she can "trade up" to a restored social position as a citizen of good character. As Bazemore (1998) has argued, redemption needs to be earned. This is not necessarily bad news for rehabilitation; as the Scottish Prison Commission (2008, para 33) noted, "one of the best ways for offenders to pay back is by turning their lives around." But it does mean that rehabilitation theories and practices need to engage much more explicitly than hitherto with questions of justice and reparation.

Ultimately, even where psychological or personal issues are tackled, legal requalification is confirmed, and moral debts are settled, the question of "social rehabilitation" remains. In European jurisprudence, the concept of social rehabilitation entails both the restoration of the citizen's formal social status and the availability of the personal and social means to do so (Van Zyl Smit and Snacken 2009). But social rehabilitation can also be taken to imply something that is broader, deeper, and more subjective, specifically, the informal social recognition and acceptance of the reformed ex-offender. This, rather than the advancement of the "science" of correctional rehabilitation, is perhaps the ultimate problem for rehabilitation today in practice.

Ultimately, as Fig. 1 above illustrates, this entry has tried to disentangle rehabilitation and instead to show how its four forms might, in fact, constitute not a Scots "fankle," but a kind of Celtic knot; one that weaves together these four strands and reveals their interdependencies. "Constructive punishment," to use Duff's term, can be rehabilitative; it should be rehabilitative; it must be rehabilitative. But it can only work to prevent crime if it also works to deliver justice, and that requires attention to all four strands.

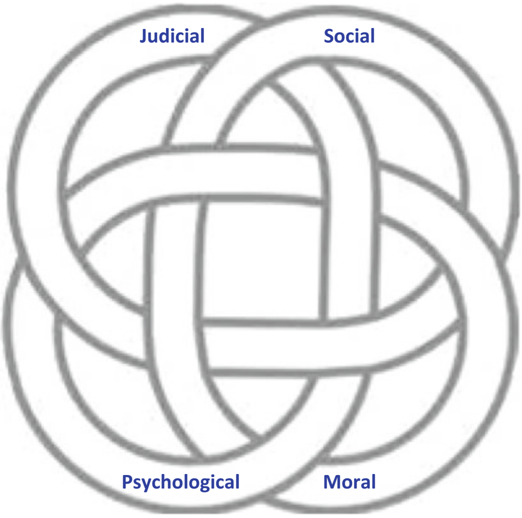

Punishment as Rehabilitation, Fig. 1 Four interdependent forms of rehabilitation

Related Entries

- ▶ History of Corrections
- ▶ Penal Justice and Social Injustice
- ▶ Penal Philosophy and Sentencing Theory
- ▶ Punishment and the State
- ▶ Rehabilitation
- ▶ Sentencing Research
- ▶ Theories of Punishment

Recommended Reading and References

Bazemore G (1998) Restorative justice and earned redemption: communities, victims, and offender reintegration. Am Behav Sci 41(6):768–813

Beccaria C (1764/1963) On crimes and punishment (trans: Pallouci H). Bobbs-Merrill, Indianapolis

Bottoms A (1980) An introduction to 'the coming crisis'. In: Bottoms A, Preston R (eds) The coming penal crisis. Scottish Academic Press, Edinburgh

Comfort M (2007) Doing time together: love and family in the shadow of the prison. University of Chicago Press, Chicago

Cox A (2012) Doing the programme or doing me? The pains of youth imprisonment. Punishment Soc 13(5):592–610

Crewe B (2009) The prisoner society: power, adaptation and social life in an English prison. Oxford University Press, Oxford

Crewe B (2012) Depth, weight, tightness: revisiting the pains of imprisonment. Punishm Soc 13(5):509–529

Duff A (2003) Probation, punishment and restorative justice: should altruism be engaged in punishment. Howard J 42(1):181–97

Duff A (2005) Punishment and rehabilitation – or rehabilitation as punishment. Crim Justice Matters 60(1):18–19

Durnescu I (2011) Pains of probation: effective practice and human rights. Int J Offender Ther Comp Criminol 55:530–545

Haney C (2005) Reforming punishment: the psychological limits to the pains of punishment. American Psychological Association, Washington, DC

Hudson B (2003) Understanding justice, 2nd edn. Open University Press, Buckingham

Johnstone G, Van Ness D (eds) (2007) The handbook of restorative justice. Willan, Cullompton

King R (2005) The effects of supermax custody. In: Liebling A, Maruna S (eds) The effects of imprisonment. Willan, Cullompton

Lacombe D (2008) Consumed with sex: the treatment of sex offenders in risk society. Br J Criminol 48(1):55–74

Lewis S (2005) Rehabilitation: headline or footnote in the new penal policy? Probat J 52:119–35

McNeill F (2012) Four forms of supervision: towards an interdisciplinary perspective. Legal Criminol Psychol 17(1):18–36

McNeill F, Raynor P, Trotter C (eds) (2010) Offender supervision: new directions in theory, research and practice. Willan, Cullompton

Maruna S (2001) Making good. American Psychological Association, Washington, DC

Maruna S (2011) Judicial rehabilitation and the 'clean bill of health' in criminal justice. Eur J Probat 3:97–117

Petersilia J, Deschenes E (1994) Perceptions of punishment: inmates and staff rank the severity of prison versus intermediate sanctions. Prison J 74(3):306–328

Raynor P (2007) Risk and need assessment in British probation: the contribution of LSI-R. Psychol Crime Law 13:121–138

Robinson G (2008) Late-modern rehabilitation: the evolution of a penal strategy. Punishm Soc 10:429–445

Robinson G, McNeill F, Maruna S (forthcoming) Punishment in society: the improbable persistence of community sanctions. In: Simon J, Sparks R (eds) The Sage handbook of punishment and society. Sage, London/New York

Rotman E (1990) Beyond punishment. A new view of the rehabilitation of criminal offenders. Greenwood Press, New York

Rotman E (1994) Beyond punishment. In: Duff A, Garland D (eds) A reader on punishment. Oxford University Press, Oxford

Scottish Prisons Commission (2008) Scotland's choice. Scottish Prisons Commission, Edinburgh

Sykes G (1958) The society of captives: a study of a maximum-security prison. Princeton University Press, Princeton

Tonry M (2006) Purposes and functions of sentencing. In: Tonry M (ed) Crime and justice: a review of research, vol 34. University of Chicago Press, Chicago

Van Zyl Smit D, Snacken S (2009) Principles of European prison law and policy. Oxford University Press, Oxford

Q

Quality of Life

▶ Discriminant Validity of Disorder and Crime

Quality-of-Life Policing

▶ Order Maintenance Policing

Quantile Regression Models to Analyze Experimental Data

Chester L. Britt
School of Criminology and Criminal Justice,
Northeastern University, Boston, MA, USA

Synonyms

Distributional analysis

Overview

The typical analysis of experimental or quasi-experimental data relies on analysis of variance (ANOVA) and ordinary least squares (OLS) regression techniques to test for differences in mean levels of an outcome or dependent variable. These models assume that cases within each treatment or control group are normally distributed and that the variance of the dependent variable is constant across groups, which then allows for a straightforward comparison of means across groups to test whether there is an effect of the treatment. In the study of crime and criminal justice, there are many applications where data are not normally distributed or do not have constant variance across groups. This is not surprising, since many of the outcome measures criminologists study – number of arrests, crime rates, bail setting, and length of prison sentence – have a high degree of positive skew. It is unlikely to find that these measures are then somehow normally distributed in experimental or quasi-experimental data, even following the use of transformations, such as the natural logarithm, which are intended to reduce the skew. All these raise concerns about the use of standard ANOVA and OLS regression models.

An ample body of research has indicated that mild to moderate violations of the assumptions of ANOVA and OLS regression models will not typically alter main findings, but authors will often insert numerous caveats about the potential volatility of their statistical results, in light of these assumption violations. Quantile regression models represent an interesting contrast to the use of ANOVA and least squares regression models, since quantile regression models are not influenced by extreme values or highly skewed distributions and analyze the full distribution of the dependent variable. Where quantile regression models may be particularly useful is in picking up treatment effects in the tails of the

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

distribution of the dependent variable. One of the frustrations commonly echoed in the conducting of experiments in criminal justice settings is the failure of treatment effects to achieve statistical significance. Quantile regression models highlight the possibility that the lack of a statistically significant effect may be limited to the mean or to a fixed range of the dependent variable. More importantly, it is possible that the treatment effect may operate differentially at the 10*th* percentile, the median (50*th* percentile), and the 90*th* percentile.

Quantile regression models represent a specific type of distributional analysis, from a much broader set of methods. In general, distributional methods offer an alternative framework for analysis that allows the researcher to assess whether the effects of independent variables on the dependent variable are constant across the distribution of the dependent variable. In roughly the last 10 to 15 years, there has been rapid growth in the availability and use of various distributional methods, particularly in research on the differential effects of background characteristics, such as education, on income and wealth (see, e.g., Buchinsky 1998; Handcock and Morris 1999). Included in the distributional methods framework are parametric quantile regression models that permit estimation of unique effects of the independent variables at any specified point in the distribution of the dependent variable. The primary purpose of this entry is to illustrate the application and interpretation of quantile regression techniques to the analysis of experimental data. As shown below, quantile regression techniques offer the potential for new findings regarding the variability of effects of treatment conditions across the distribution of the dependent variable.

Issues with Common Approaches to Analyzing Experimental Data

Standard practice for the analysis of experimental data, where there has been random assignment of cases to a control group and one or more treatment groups, has been to test for differences in group means. In the case of one control group and one treatment group, this can be accomplished with a simple *t*-test. For situations involving random assignment of cases to three or more groups, a one-way ANOVA model will serve a similar purpose: Test for differences in group means to determine whether the treatment had an effect. For those situations where the control group may have been sampled at a different location or point in time (i.e., a quasi-experimental design) or the researcher is interested in examining the effects of other case characteristics in addition to experimental conditions, OLS regression (or ANCOVA) models may be used to examine the effect of the experimental conditions while also statistically controlling for the effects of these other case characteristics on the outcome measure. Regardless of which approach is used, however, the analyst assumes that the distribution of cases within each group is normal with constant variance.

Figure 1 presents the density plots for three hypothetical groups (of 100 cases each) – one control group and two treatment groups – that meet the assumptions of an ANOVA model. Since the observations are normally distributed within each group and have approximately the same variance, the computation of mean differences using a one-way ANOVA would be straightforward.

In contrast, Fig. 2 presents the density plots for another three groups that have the same means as in Fig. 1, but the distribution of cases within each group shows a high degree of positive skew. The general shape of the plots in Fig. 2 is similar to many of the measures of interest in criminology and criminal justice, such as, self-reported delinquency and length of prison sentence. The use of ANOVA for the data as portrayed in Fig. 2 would be problematic, given the non-normal and highly skewed distribution of cases across the three groups.

As noted above, one of the ways of attempting to address non-normal distributions like those in Fig. 2 is to transform the data, say, by taking the natural logarithm of the original value. Figure 3 presents the density plots for the three groups in Fig. 2 after taking the natural logarithm. Although the transformed distributions are less skewed, they are still non-normal and show the spread (variance) of cases to be far from similar.

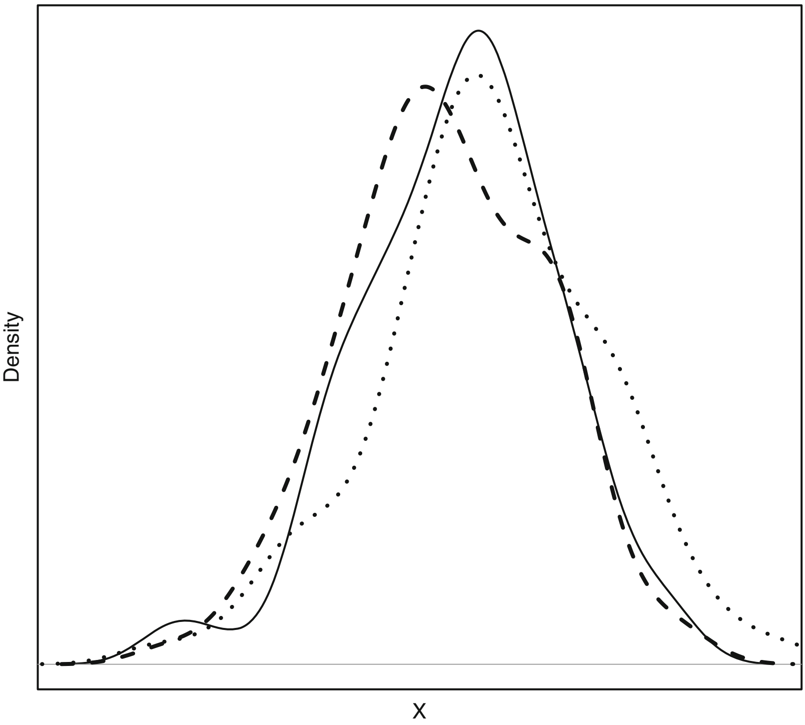

Quantile Regression Models to Analyze Experimental Data, Fig. 1 Hypothetical normal curves

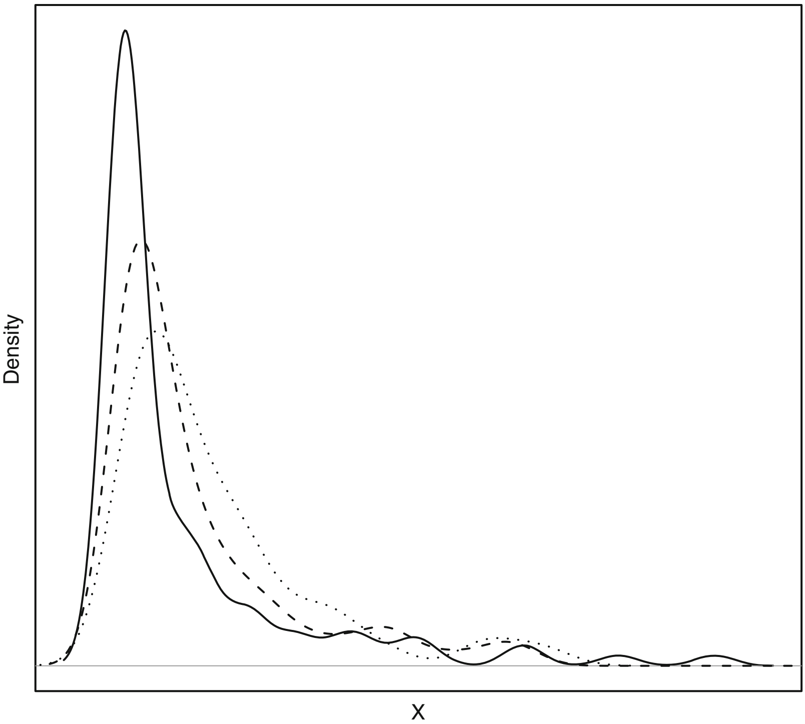

Quantile Regression Models to Analyze Experimental Data, Fig. 2 Hypothetical skewed distributions

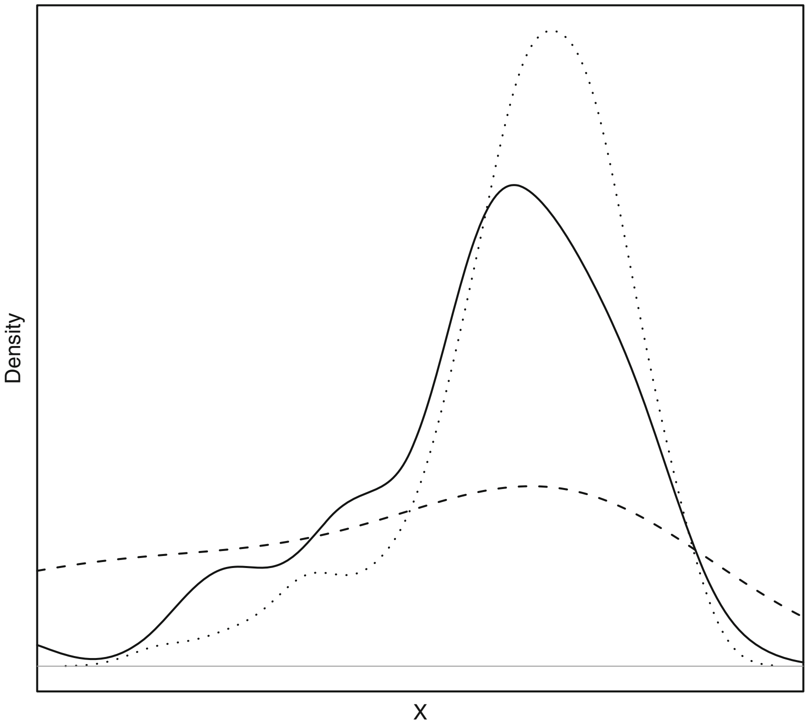

Quantile Regression Models to Analyze Experimental Data, Fig. 3 Hypothetical skewed distributions: natural log

We can see from the plots in Figs. 2 and 3 that there are important differences in the outcome measure across the three groups that would not be reflected in the test for mean differences. Consequently, when looking at the full distribution of the outcome variable, we may find important differences that are obscured by looking only at the means for each group. For example, we can see that the tails of the distributions in Figs. 2 and 3 are different, but we are unable to say how different. It is this kind of question for which quantile regression models are particularly well suited: Do the groups differ at the median? at the 25*th* percentile? at the 90*th* percentile?Are these differences statistically meaningful? How do other covariates affect these apparent differences?

Fundamentals of Quantile Regression Models

Quantile regression models were initially proposed by Koenker and Bassett (1978) as a method of robust regression that would account for a non-normal distribution of error terms and as a test for heteroskedastic error terms (see also, Gilchrist 2000; Hao and Naiman 2007; Koenker and Bassett 1982; Koenker 2005). (A quantile refers to a percentile in the distribution of a variable – a quantile of 0.25, for example, would refer to the same point in a distribution as the 25*th* percentile and the median would be represented by a quantile of 0.50.) More recently, quantile regression models have been used to study income inequality (Buchinsky 1994, 1998; Handcock and Morris 1999; Mello and Perrelli 2003; Miller 2005), earnings (Eide and Showalter 1999; Eide et al. 2002; Mora and Davila 2006; Walker 2000), and wealth (Conley and Galenson 1998; Walker 2000) distributions. In criminology, Britt (2009) has shown how this model can be used to study sentence length decisions. In cases where there may be a substantial number of zeros on the dependent variable – representing censoring of the dependent variable – quantile regression models also allow for unbiased estimates of the effects of the independent variables on the dependent variable at higher quantiles, where the influence of the zeros will be minimal (Powell 1984, 1986). For example, sentencing researchers are often confronted with offenders who are not incarcerated and are effectively a "0" on the dependent variable and then excluded from further analysis. Quantile regression techniques could be used to address this concern without relying on various sample selection models that may be more or less appropriate for the analysis of sentencing decisions (Bushway et al. 2007).

The Statistical Model

Koenker and Hallock (2001) described the estimation of the coefficients of a quantile regression model by explaining how it could be reduced to an optimization problem (for detailed descriptions, see Koenker (2005) and Koenker and Bassett (1978)). At the heart of the optimization problem is the minimization of asymmetrically weighted absolute residuals. By asymmetrically weighted residuals, they mean assigning different weights to positive and negative residuals. This results in the following minimization equation for unconditional quantiles:

$$\min_{\xi \in \Re} \sum_{i=1}^{n} \varrho_\tau (y_i - \xi),$$

where $\varrho_\tau(\cdot)$ is the absolute value function that gives the τth sample quantile, y_i is the observed value of the dependent variable, and ξ is the predicted value.

To provide a point of comparison, least squares regression can be described in similar terms as minimizing the following equation:

$$\min_{\mu \in \Re} \sum_{i=1}^{n} (y_i - \mu)^2$$

to obtain an unconditional population mean, $\mu = E(Y)$. By replacing μ with a parametric function $\mu(x,\beta)$ (McCullagh and Nelder 1989), we obtain the conditional mean of Y (i.e., E(Y | x)) by minimizing

$$\min_{\mu \in \Re} \sum_{i=1}^{n} (y_i - \mu(x, \beta))^2.$$

For quantile regression, the same type of progression is used to estimate the conditional quantile by replacing ξ with a parametric function $\xi(x, \beta)$ and specifying τ (the quantile). Thus, if we wanted to estimate the conditional median function, we would set $\tau = 0.5$ (i.e., the 50*th* percentile) and minimize the function to obtain parameter estimates for the effects of x on y at the median. To obtain other conditional quantiles, we simply replace the value of τ and minimize

$$\min_{\xi \in \Re} \sum_{i=1}^{n} \varrho_\tau (y_i - \xi(x, \beta)).$$

Interpretation of the Coefficients

In general, the coefficients from an OLS regression model may be interpreted as the expected change in the conditional mean of the dependent variable, based on a unit change in the independent variable. The coefficients from a quantile regression model may be interpreted in a similar manner, with a key difference being the coefficients from a quantile regression model represent the effect of a unit change in the independent variable at some defined point of the conditional distribution of the dependent variable (i.e., the quantile). For example, if we had set $\tau = 0.5$ – the median of the distribution of the dependent variable – then we would interpret the coefficient as the change in the conditional median of the dependent variable based on a unit change in the independent variable. Similarly, if $\tau = 0.75$, then we would interpret the coefficient as the change in the conditional 0.75 quantile (75*th* percentile), given a unit change in the independent variable.

Quantile regression coefficients permit an examination of how or whether the effects of the independent variables change at different points of the distribution of the dependent variable. When the dependent variable has a distribution that is highly skewed, such as income, earnings, self-reported crime, and sentence lengths for convicted offenders, it is possible for the independent variables to have differential effects on the dependent variable that are conditional on a designated point of the distribution of the dependent variable.

To illustrate more directly the interpretation of quantile regression coefficients, Fig. 4 presents boxplots of a measure of self-reported delinquency

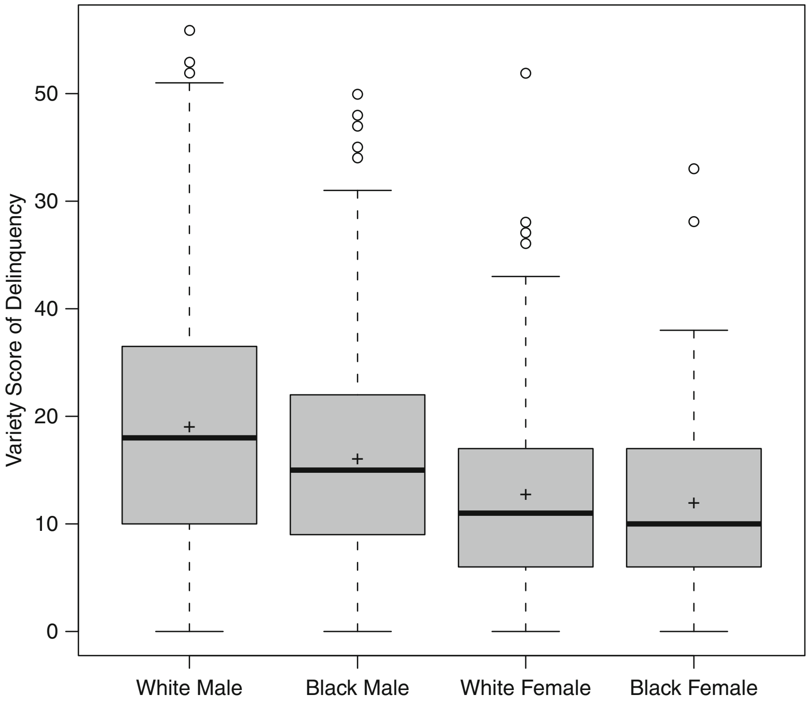

Quantile Regression Models to Analyze Experimental Data, Fig. 4 Seattle Youth Study: number of different self-reported delinquent acts by sex and race

(number of different types of delinquency) by sex and race for the sample of youth who participated in the Seattle Youth Study (Hindelang et al. 1981), which is described in more detail below. Each gray box represents the interquartile range – the range from the 25*th* percentile (lower end of the gray box) to the 75*th* percentile (upper end of the gray box). The median is represented by the horizontal line inside the gray box, while the mean for each group is represented by the + symbol. The lines extending from the gray box represent what are variably referred to as the "fence" or "whiskers" of the plot. The limits – the fence – represent 1.5 times the interquartile range. The circles appearing beyond the fence represent relatively extreme values and, in some contexts, may be treated as potential outliers. Many statistics texts contain material on the creation and interpretation of boxplots (e.g., Fox 2008). For a comprehensive treatment of boxplots, see Tukey (1977).

The interquartile range highlights some of the potential differences and similarities across the four groups, indicating a larger spread of cases for both groups of males and more compressed distributions for the two groups of females. At the 0.25 quantile (lower end of the gray box), the differences across sex start to become more apparent, where white males and black males are at 10 and 9 different delinquent acts, respectively, while both groups of females are at 6 different delinquent acts. At the median (horizontal line inside of the gray box), white males have a value of 18 different delinquent acts, while black males, white females, and black females report 15, 11, 10 different types of delinquency, respectively. The 0.75 quantile (upper end of the gray box) reveals even larger separation by sex, where the 0.75 quantile for white males is 26.5 and for black males it is 22, while the two groups of females are nearly identical (17 for white females and 16.75 for black females). The tails of the distributions are also informative. The lower end of the distribution shows virtually no difference across sex and race – the largest value at the 0.10 quantile is 5 for white males and the lowest value is 3 for white females. In contrast, at the upper end of the distribution (e.g., the 0.90 quantile), the differences are more pronounced, with values ranging from a high of 36 for white males to a low of 23.5 for black females.

The pattern of results in Fig. 4 could also be interpreted to report the relative differences at various quantiles. For example, at the 0.25 quantile, the delinquency score for white males is 1 more than for black males, and 4 more than either one of the groups of female respondents. By the median, we see that white males report 3 more acts than black males, but 7 and 8 more delinquent acts when compared to white females and black females, respectively. As we move further out on the distribution of types of delinquency, the differences become increasingly larger, where, by the 0.90 quantile, white males report 6 more acts than black males, but 9.1 more than white females and 11.5 more than black females.

Regardless of the approach used to interpret the distributional patterns in Fig. 4, the boxplots reveal interesting differences in the distribution of self-reported delinquent acts amounts by sex and race. It is also important to keep in mind that the distributions of type of delinquency by sex and race presented in Fig. 4 do not account for any other characteristics of the youths. To the extent that background characteristics (e.g., self-control, psychological stress, etc.) explain patterns of delinquency by sex and race, we would expect these differences to disappear once these characteristics had been accounted for statistically. This illustrates another way that quantile regression techniques have the potential to provide such a useful form of analysis, and allow for testing whether apparent differences in distributional patterns will disappear, once other relevant characteristics of the observations have been taken into account in a full statistical model.

Location-Shift and Location-Scale Hypotheses

The location-shift hypothesis is a test for whether a single coefficient could be used to capture the effect of an independent variable on the dependent variable across the distribution of the dependent variable. The location-shift model is equivalent to an OLS regression model where the coefficient is constant, meaning that for a fixed change in the independent variable, the conditional mean of the dependent variable is expected to change by a fixed amount. In the context of quantile regression results, if the effect of an independent variable varies around some constant value across the distribution of the dependent variable, then the effect of the variable is limited to one of shifting the location – the level – of the dependent variable.

The location-scale hypothesis is akin to testing for heteroskedasticity in the OLS model. The location-scale hypothesis reflects the possibility that as the value of the dependent variable changes in magnitude, the variance of the residuals – the spread of cases around the regression plane – will also change, sometimes increasing, sometimes decreasing. Consequently, the location-scale hypothesis tests for the possibility of a constant effect of the independent variable on the dependent variable, accounting for the increasing (decreasing) spread of cases across the distribution of the dependent variable. Koenker (2005) and Koenker and Xiao (2002a) note that the model intercept in quantile regression represents a normalized version of the residuals and to the extent the effects of the independent variable follow the same pattern as the model intercept, they represent simple shifts in location and scale of the dependent variable and residuals.

Koenker (2005) and Koenker and Xiao (2002a, b) present a test and critical values for the location-shift and the location-scale hypotheses for the full model and for the individual coefficient estimates. The critical values for the tests depend on the number of coefficients estimated, the range of quantiles, and the desired level of statistical significance; sample size has no direct bearing on the test statistic or the critical value. Practically, the calculated test statistic is compared to the critical value, and if the test statistic is greater than the critical value the null hypothesis of location shift or location scale is rejected. The rejection of the location-shift hypothesis implies that the OLS model, based on conditional means, does not provide a good representation of the effects of the independent variables on the dependent variable. Rejection of the location-scale hypothesis implies not only that the OLS model does not represent the effects of the independent variables well, but the shape of the conditional distribution of the dependent variable changes in ways not modeled by OLS, even after accounting for heteroskedasticity. In other words, there is no simple, straightforward transformation of the residuals that would result in an OLS model adequately fitting the data.

Two Applications of Quantile Regression Models

Measurement of Self-reported Delinquency

The Seattle Youth Study was an experiment conducted by Michael Hindelang, Travis Hirschi, and Joseph Weis aimed at assessing whether the method of administration of self-report delinquency items would affect the reliability and validity of the responses given by the youth who participated in the study (Hindelang et al. 1981). There were four different methods of administration of the survey that combined the form of the survey (questionnaire or interview) with conditions of administration (anonymous or non-anonymous). The 1,612 participants in the study were randomly assigned to one of the following four methods of administration: anonymous interview, anonymous questionnaire, non-anonymous interview, and non-anonymous questionnaire (for details, see Hindelang et al. 1981). Hindelang et al. (1981) performed a range of analyses to assess the reliability and validity of the self-report delinquency items and concluded that method of administration made no difference to the measurement of delinquency – each set of responses indicated a similar level of delinquency as well as common correlates of delinquency across all four methods of administration. The following discussion is intended to illustrate the application and interpretation of quantile regression techniques to experimental data and not intended to be an exhaustive reanalysis of Hindelang et al. (1981).

One of the outcome measures of delinquency created by Hindelang et al. (1981) was a variety scale that counted the number of different delinquent acts out of a maximum of 69 self-reported as having ever been committed by the youth. A boxplot of this delinquency scale by method of administration is shown in Fig. 5, revealing a high degree of positive skew in each of the four experimental conditions. A simple one-way analysis of variance shows the means across the four groups not to be significantly different ($F = 2.003$, $p > 0.10$).

Table 1 presents the results for estimating a series of quantile regression models covering the range from quantiles of 0.10 to 0.90 for the subsample of male respondents. In order to simplify the presentation of results, female respondents were excluded, since they represented only 25 % of the sample and were not assigned to either the anonymous interview or the non-anonymous questionnaire conditions. For purposes of comparison, the OLS regression results are reported in column 2, while columns 3 through 7 indicate the results for the different quantiles. In all analyses, anonymous questionnaire is used as the reference category.

The results in Table 1 show no consistent pattern of significant effects from quantile 0.10 to 0.90 – only 1 of 15 coefficients is statistically significant at $p < 0.05$. These results imply that insofar as delinquency is measured by self-report items, the method of administration has no bearing on the variety of delinquency self-reported by the respondents. An important implication of this finding – and not one highlighted by Hindelang et al. (1981) – is that the method of administration did not affect self-reported delinquency levels among either those reporting few delinquent acts or those reporting many different delinquent acts.

Philadelphia Experiment on Bail Guidelines

In the early 1980s, John Goldkamp and Michael Gottfredson conducted an experiment in the Philadelphia courts that involved judges using a set of guidelines for making bail decisions (Goldkamp and Gottfredson 1985). Prior research had indicated tremendous variability in the bails that judges set for defendants in their courtrooms. In an attempt to test whether bail decision-making could be made more consistent and equitable, Goldkamp and Gottfredson developed a set of bail guidelines organized around (1) the severity

Quantile Regression Models to Analyze Experimental Data, Fig. 5 Seattle Youth Study: number of different self-reported delinquent acts by method of administration

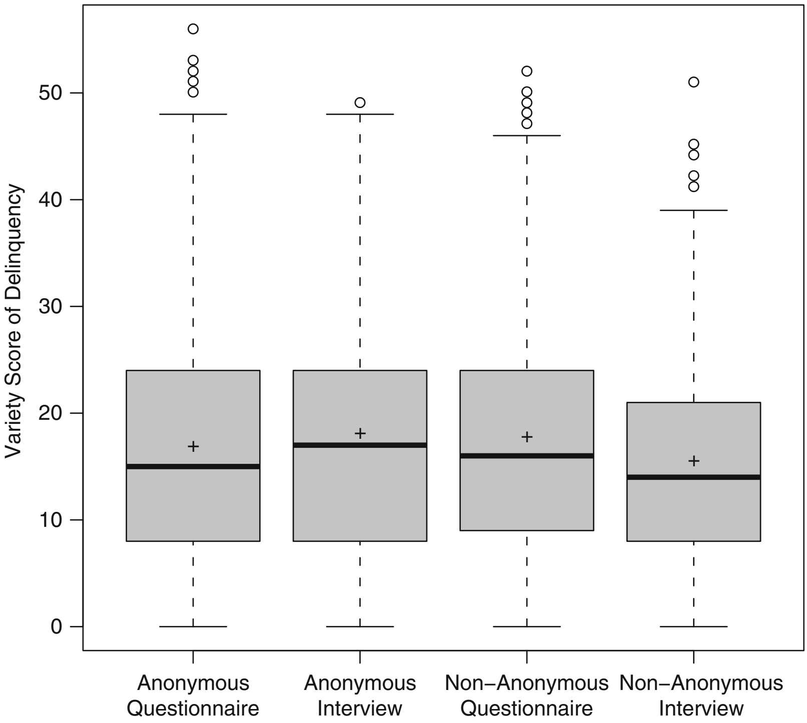

Quantile Regression Models to Analyze Experimental Data, Table 1 OLS and quantile regression results for Seattle Youth Study

		Percentile				
Variable	OLS	10th	25th	50th	75th	90th
Intercept	19.38***	5.00***	9.00***	18.00***	27.00***	36.00***
Anonymous interview	−2.11*	−1.00	−1.00	−1.00	−3.00	−4.00
Non-anonymous questionnaire	−1.62	0.00	0.00	−2.00	−3.00	−1.00
Non-anonymous interview	−1.40	1.00	1.00	−1.00	−2.00	−4.00

*p < 0.05; ***p < 0.001
N = 1,214 male respondents

of the charge offense and (2) the risk of flight or rearrest. The construction of the risk group categories and recommended bail amounts were based on an earlier pilot study in the Philadelphia courts (see Goldkamp and Gottfredson 1985 for details), so there was an empirical basis for the recommendations in the guidelines. The bail guidelines were presented in a grid format, allowing the judge to be able to cross-reference the defendant's risk group classification and current charge severity to locate a cell in the table that suggested a range of bail amounts, as well as other non-bail release options (e.g., release on recognizance (ROR)) for defendants with low risk and less severe offense charges.

Sixteen judges participated in the study: eight each were randomly assigned either to the control group (asked to make bail decisions as they had in the past) or to the experimental group (asked to follow the recommendations made in the bail guidelines grid). For all 16 judges, a quota sample of 120 total cases was selected that included an equal number of cases in each of the 6 charge severity categories (which represented a mix of felony and misdemeanor cases). The goal of the experiment was to test whether the voluntary use

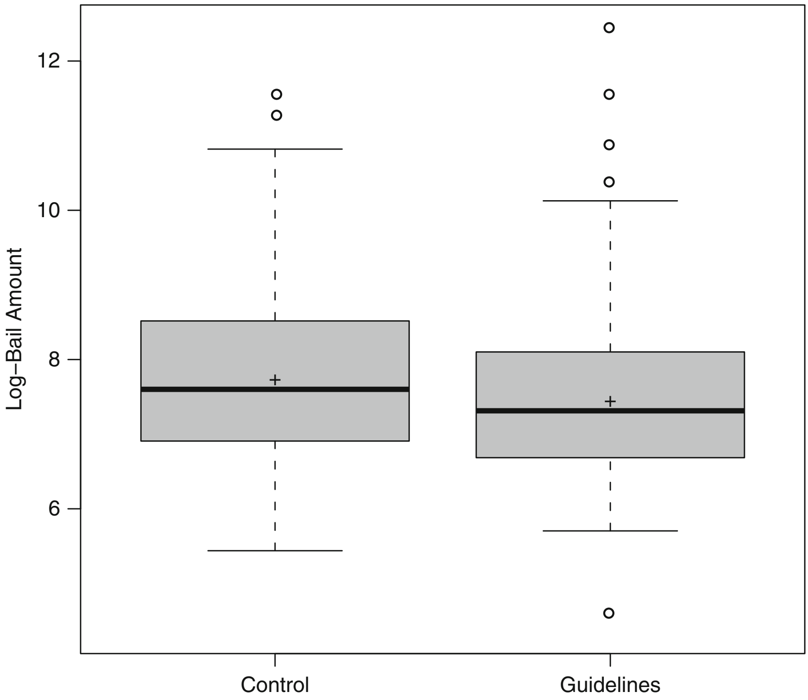

Quantile Regression Models to Analyze Experimental Data, Fig. 6 Philadelphia Bail Experiment: log-bail

of bail guidelines would result in less variability and more equitable treatment of defendants across the group of judges participating in the study. Goldkamp and Gottfredson (1985) also acknowledge that some contamination is likely to have occurred. Since all 16 judges participating in the study worked in close proximity to each other, there were no doubt conversations among the judges about the guidelines. Consequently, it is possible that those judges not assigned to the experimental group could have been sensitized to their own bail decision-making practices and unintentionally altered their bail and release decisions, thereby diluting the potential effect of the treatment.

The analyses performed by Goldkamp and Gottfredson (1985) focused primarily on a distributional comparison, where they compared median bail amounts for the control and the guidelines (experimental) judges, as well as partitioning the distributions in ways that allowed for comparisons of the percentage of cases receiving bail amounts within some specified range. Their distributional comparisons highlighted differences across the two groups of judges, but there were no tests of statistical significance for differences in the medians or other quantiles. A secondary set of analyses regressed the logarithm of bail amount on the two key criteria of the guidelines grid: severity of charge offense and risk group classification. Interestingly, while they found the overall bail amounts to differ significantly across the two groups of judges, the effects of offense severity and risk group classification on bail amount requested were nearly identical for control and guidelines judges, indicating that the relative weights judges attached to these criteria were similar, regardless of whether they used the guidelines or not. They note that this finding is not that surprising, since the guidelines were empirically grounded in the prior behavior of judges in Philadelphia.

Figure 6 presents the boxplots for log-bail amount for the control and the guidelines judges – the degree of skew in bail amount (in dollars) resulted in boxplots that were so compressed that they were not useful for discerning any differences in the distributions across experimental condition. We see that the judges who were assigned to the bail guidelines condition show

Quantile Regression Models to Analyze Experimental Data, Table 2 OLS and quantile regression results

		Percentile				
Variable	OLS	10th	25th	50th	75th	90th
Intercept	5.43	4.83	5.23	5.29	6.08	6.80
Experimental group	−0.36	*0.00*	−0.31	−0.39	−0.56	−0.58
Offense severity	0.19	0.14	0.16	0.19	0.19	0.21
Risk group	*0.07*	*0.00*	*0.02*	0.08	*0.08*	*0.03*

N = 1,007 cases with Bail >0
NOTE: Italicized coefficients are *not* statistically significant at p < .01

a more compact distribution of values in the interquartile range when compared to the control group judges. The shorter fences also indicate that the distribution of log-bail amounts is more compressed among those judges using the bail guidelines compared to the group of control judges.

Table 2 presents the results from an OLS regression (Column 2) and a series of quantile regression models (Columns 3 through 7) testing for differences in the effect of guidelines participation, offense severity, and risk group classification on log-bail amount from $\tau = 0.10$ to $\tau = 0.90$. The OLS results indicate that bail amounts are about 36 % lower in those cases presented to a guidelines judge, controlling for offense severity and risk group. The quantile regression results illustrate how the use of the guidelines affects the bail amounts requested at different points of the log-bail distribution. At the $\tau = 0.10$ quantile, there is no difference in the log-bail amounts between the two groups of judges. By the $\tau = 0.25$ quantile, bail amounts are approximately 31 % lower in the guidelines group. At the median, the OLS and quantile regression results for the effect of using the guidelines are similar (showing log-bail amounts to be about 39 % lower among guidelines judges). The primary impact of the bail guidelines can be seen in those cases at the $\tau = 0.75$ and the $\tau = 0.90$ quantiles, where bail amounts are approximately 56 % and 58 % lower among those cases heard by a guidelines judge. Substantively, these results show that as we move to the upper end of the distribution of log-bail amounts, the effect of bail guidelines increases in magnitude, resulting in increasingly larger differences in cases handled by guidelines and control judges.

There are also interesting patterns to the results for offense severity and risk group. In regard to offense severity, the OLS results indicate that each unit increase in offense severity is associated with about a 19 % increase in bail amount. The quantile regression results show that at the lower end of the distribution – the 10*th* to 25*th* percentiles – the effect of offense severity falls between 14 % and 16 %. It is at the median that the quantile regression results are comparable to the OLS results, but at the 90*th* percentile, the effect of offense severity has risen to more than 21 %. The effect of risk group is more modest – the OLS results indicate that a unit increase in risk group is associated with about a 7 % increase in bail amount. Of the quantile regression results displayed in Table 2, the only coefficient to be significantly different from zero is at the median, where each increase in risk group classification is associated with a nearly 8 % increase in bail amount.

The substance of the findings in Table 2 suggest that the bail guidelines functioned in much the same way as current practice for those cases with bail amounts below the 25*th* percentile – the least serious cases and defendants. From that point on, however, the use of the bail guidelines had an increasingly greater impact of bail amounts requested, so that the further out we go in the distributions for the guidelines and control judges, the larger the differences become, even while statistically controlling for offense severity and risk group classification.

Location-Shift and Location-Scale Hypotheses

Table 3 presents the results for the test of the location-shift and location-scale hypotheses for the effects of experimental condition, offense

Quantile Regression Models to Analyze Experimental Data, Table 3 Location-shift and location-scale hypotheses test results

Variable	Location shift	Location scale
Experimental group	3.73**	3.76**
Offense severity	1.92	2.35*
Risk group	2.53*	2.42*
Full model	7.36**	6.63**

* $p < 0.05$; ** $p < 0.01$

severity, and risk group classification on log-bail. Recall that the location-shift hypothesis is fundamentally a test for whether an OLS regression model adequately represents the effects of the independent and control variables on the outcome measure. The location-scale hypothesis tests whether fixed coefficients capture the effects of the independent and control variables after taking into account the changing spread of observations across the distribution of the outcome variable.

The critical values reported in Koenker (2005) and Koenker and Xiao (2002b) for the quantile range of $\tau = 0.15$ to $\tau = 0.85$ for the full model (3 coefficients) are 4.269 ($p < 0.05$) and 5.074 ($p < 0.01$). For the individual coefficients, the critical values are 2.048 ($p \leq 0.05$) and 2.573 ($p \leq 0.01$). Note that the range of the quantile regression process was trimmed from the results presented in Table 2. The rationale for reducing the range was the volatility of the quantile regression estimates in the tails of the log-bail distribution, which can make the estimation of the test statistics problematic (Koenker 2005). The results in Table 3 show that both the location-shift and location-scale hypotheses are rejected at $p \leq 0.01$, meaning that neither an OLS nor an OLS model adjusted for heteroskedasticity adequately model the effects of experimental condition, offense severity, and risk group across the distribution of log-bail. When we focus on individual coefficients, we find that the effects of experimental condition and risk group are not at all well modeled by an OLS coefficient (with or without heteroskedasticity adjustments). Interestingly, there are somewhat contradictory results for the effect of offense severity. The location-shift hypothesis cannot be rejected ($1.918 < 2.048$), but the location-scale hypothesis is rejected ($2.350 > 2.048$), suggesting that an OLS coefficient may represent the effect of offense severity from $\tau = 0.15$ to $\tau = 0.85$.

Summary

The primary focus of this entry has been to illustrate the analysis of experimental data with quantile regression models. Typical analyses of experimental data focus on differences in the levels of the dependent variable across the control and treatment groups. Somewhat less common, but still of interest, may be a change in the scale (spread) of the dependent variable across the control and treatment groups. In either case, the analysis will often be restricted to looking at the differences in means and variances. Quantile regression models offer the opportunity to assess whether there are differences across the control and treatment groups at any specified point in the distribution of the dependent variable. Thus, with quantile regression models, we can test for effects of a treatment (or set of treatments) at each quartile (i.e., 25*th*, 50*th*, and 75*th* percentiles), the tails of the distribution (e.g., 10*th* and 90*th* percentiles), or any other point in the distribution. The substantive importance of this type of analysis is the possibility of determining whether the effects of a treatment are constant or variable across the distribution of the outcome measure. To the extent the effect of the treatment is more or less constant, a more traditional ANOVA or linear regression analysis offers a reasonable test of the impact of the treatment. To the extent the effect of the treatment is variable, quantile regression models will be useful in highlighting those points of the distribution where the treatment is more or less effective.

The results presented in this entry illustrate well the utility of quantile regression models. In the analysis of data from the Seattle Youth Study, the quantile regression results reinforced and extended earlier findings about the effect of method of administration of self-report delinquency items on the variety of delinquency reported. Consistent with results reported by

Hindelang et al. (1981), there were no effects of method of administration on variety of self-reported delinquency. The value added by the quantile regression results was the lack of difference across the full distribution of self-reported delinquency – it did not matter if the comparison involved youth who reported a small number of different acts or a relatively large number of acts (out of a possible 69).

In regard to results from the Philadelphia Bail Experiment, the OLS results suggested that bail amounts for defendants presented to guidelines judges were about 30 % lower than those for defendants whose decisions were made by the control group judges. However, the quantile regression results illustrate that the effect of participating in the guidelines only affected cases at or above the 25*th* percentile. More important, as we move to the upper end of the log-bail distribution (i.e., the cases with larger bail amounts representing more serious offenses and/or more risky defendants), the effect of participating in the guidelines increased the difference between the two groups of defendants. The effects of offense severity and risk group classification were also shown to be variable across the distribution of log-bail, although in slightly different ways. Much like the effect of participating in the guidelines, the effect of offense severity increased in magnitude as we moved from the lower to the upper end of the distribution. In contrast, the effect of risk group was limited to the middle of the distribution – the effect of risk group was not significantly different from zero in either tail of the log-bail distribution.

Why use quantile regression models for the analysis of experimental data? As noted above, one of the key benefits to using distributional analyses, such as quantile regression techniques, is they are particularly effective at testing for variations in the effects of control and independent variables across the distribution of a dependent variable. In the context of analyzing experimental data, quantile regression models would appear to offer an interesting way of teasing out the effect of a treatment at different quantiles of the distribution of the dependent variable. For example, as shown with the analysis of data from the Philadelphia Bail Experiment, we may not expect to find dramatic differences in bail amounts at the lower end of the distribution for control and guidelines judges. In part, this is due to the manner in which the guidelines were constructed – less serious offenses and defendants who pose little risk will generally be likely to be released on their own recognizance or be required to post minimal bail. At the higher levels of the distribution of bail amounts, the range of possible bail amounts is greatly expanded, which will create more opportunities for judges to exercise discretion in their bail decisions. Since the bail guidelines were intended to reduce the range of possible bail amounts by lowering the maximum amount of bail a judge should request, the finding of an increasingly greater impact of the bail guidelines at the upper end of the bail distribution is consistent with a reduction in the range of options available to the judge.

Some of the earliest work on the development of quantile regression models was focused on dealing with censored data, where there were many zero values for the dependent variable (Powell 1984, 1986). Quantile regression models represented a way of estimating unbiased effects of the independent variables on the dependent variable, where the influence of the zeros would be minimal. Practically, the addition of the zero values for the dependent variable simply shifts the nonzero values of the distribution to the right (i.e., to the higher quantiles). The ability to test for variation in the quantile regression coefficients is unaffected by the additional zero values on the dependent variable. More recently, Koenker (2005, 2008) has started to develop the statistical theory and implement a censored quantile regression procedure that will, among other options, account for zero values on the dependent variable.

There are at least two other important benefits to the quantile regression approach described in this entry. First, by making no assumptions about the distribution of the error terms in the model, quantile regression techniques effectively deal with the non-normal distribution of errors that is so common in the analysis of data on criminal behavior and the criminal justice system. Many researchers will often transform the dependent variable (e.g., by computing a logarithm) to reduce the degree of skew or recode the variable in such a way that the fundamental distribution of the variable is changed (e.g., recoding frequency of self-reported delinquent acts from a count to something that ranges from 0 to 10 or more). Second, the ability of quantile regression models to estimate variable effects of control and independent variables across the distribution of a dependent variable suggests that this technique would be particularly valuable for testing theories that either directly or indirectly imply nonconstant relationships. Beyond the issues discussed in this entry, many theories of crime and delinquency also imply that the effect of a key theoretical variable (e.g., control balance or peers) may have effects on crime and delinquency that vary by the frequency of criminal and delinquent behavior. Similarly, in the context of criminal justice decision-making, the "liberation hypothesis" suggests that judges will exercise greater discretion in particular types of cases. Standard linear regression models will not be able to detect these kinds of effects.

The primary challenge to the application of quantile regression is the need for a sufficiently large number of cases that will allow for the computation of quantiles for the numerous combinations of the values of the independent variables. There is no magic number for the size of the sample necessary – the analyses reported above relied on just over 1,000 cases, while some of the research cited above used a few hundred observations. In light of the fact that so many researchers studying crime and criminal justice issues rely on samples (and often populations) of several thousand cases, this requirement is not likely to be prohibitive in most applications.

Related Entries

- ▶ Differences-in-Differences in Approaches
- ▶ History of Randomized Controlled Experiments in Criminal Justice
- ▶ History of the Self-Report Delinquency Surveys
- ▶ Propensity Score Matching

Recommended Reading and References

Britt CL (2009) Modeling the distribution of sentence length decisions under a guidelines system: an application of quantile regression models. J Quantitative Criminol 25:341–370

Buchinsky M (1994) Changes in the U.S. wage structure 1963–1987: application of quantile regression. Econometrica 62:405–458

Buchinsky M (1998) The dynamics of changes in the female wage distribution in the USA: a quantile regression approach. J Appl Econometrics 13:1–30

Bushway S, Johnson BD, Slocum LA et al (2007) Is the magic still there? The use of the Heckman two-step correction for selection bias in criminology. J Quantitative Criminol 23:151–178

Conley T, Galenson D (1998) Nativity and wealth in mid-nineteenth-century cities. J Economic History 58:468–493

Eide ER, Showalter MH (1999) Factors affecting the transmission of earnings across generations: a quantile regression approach. J Hum Resour 34:253–267

Eide S, Eide ER, Showalter MH, Sims DP (2002) The effects of secondary school quality on the distribution of earnings. Contemporary Economic Policy 20:160–170

Fox J (2008) Applied regression analysis and generalized linear models, 2nd edn. Sage, Los Angeles

Gilchrist WG (2000) Statistical modelling with quantile functions. Chapman and Hall, Boca Raton, FL

Goldkamp JS, Gottfredson MR (1985) Policy guidelines for bail: an experiment in court reform. Temple University Press, Philadelphia, PA

Handcock MS, Morris M (1999) Relative distribution methods in the social sciences. Springer, New York

Hao L, Naiman DQ (2007) Quantile regression. Sage, Thousand Oaks, CA

Hindelang H, Hindelang MJ, Hirschi T, Weis J (1981) Measuring delinquency. Sage, Newbury Park, CA

Koenker R (2005) Quantile regression. Cambridge University Press, New York

Koenker, R (2008) Censored quantile regression redux. http://www.econ.uiuc.edu/~roger/research/crq/crq.pdf

Koenker R, Bassett G (1978) Regression quantiles. Econometrica 46:33–50

Koenker R, Bassett G (1982) Robust tests for heteroskedasticity based on regression quantiles. Econometrica 50:43–61

Koenker R, Hallock KF (2001) Quantile regression. J Econ Perspect 15:143–156

Koenker R, Xiao Z (2002a) Inference on the quantile regression process. Econometrica 70:1583–1612

Koenker R, Xiao Z (2002b) Inference on the quantile regression process: Electronic appendix. http://www.econ.uiuc.edu/~roger/research/inference

McCullagh P, Nelder J (1989) Generalized linear models, 2nd edn. Chapman and Hall, New York

Mello M, Perrelli R (2003) Growth equations: a quantile regression exploration. Q Rev Econ Finance 43:643–667

Miller PW (2005) The role of gender among low-paid and high-paid workers. The Australian Economic Review 38:405–417
Mora MT, Davila A (2006) Hispanic ethnicity, gender, and the change in the LEP-earnings penalty in the united states during the 1990s. Social Sci Quart 87:1295–1318
Powell JL (1984) Least absolute deviation estimation for the censored regression model. J Econometrics 25:303–325
Powell JL (1986) Censored regression quantiles. J Econometrics 32:143–155
Tukey JW (1977) Exploratory data analysis. Addison-Wesley, Reading, MA
Walker TR (2000) Economic opportunity on the urban frontier: wealth and nativity in early San Francisco. Explorations Economic History 37:258–277

Quantitative Methods Used by Criminologists in the Nineteenth Century

David Maimon
Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA

Synonyms

Early quantitative methods; Historical criminology

Overview

The adoption of the "positivist approach" in the study of human behavior exposed early criminologists to a range of quantitative methods that allowed objective investigations of crime data. This entry reviews the types of data and quantitative methods used by criminologists in the nineteenth century to describe the magnitude of crime, identify offenders' unique characteristics, and search for the causes of deviance and crime. Similar to contemporary sources of "crime statistics," data on crime and criminals during this period was collected by official criminal-justice agencies and by independent scholars who observed, surveyed, and conducted scientific experiments with criminals. Due to the absence of predictive statistical tools, data analysis had mainly involved (1) classifications of crime types and construction of frequency distributions, (2) calculations of crime rates, proportions, averages, and deviations, (3) comparisons between pairs of variables, and (4) estimations of the relationships between pairs of variables using tables and graphs. The important influences of these quantitative methods on the evolution of criminological theory are discussed.

Introduction

Although the roots of criminological thinking could be traced back to the second half of the eighteenth century, it was not until the early nineteenth century, when "positivist" methods (i.e., all these methods that can predict crime and that are in harmony with the methods applied in the natural sciences (Beirne 1987)) were employed in the study of crime (Guerry 1833; Quetelet 1831) that criminology established itself as a legitimate modern social science. Drawing on the "positivist" assumption that quantitative methods should be applied in the social sciences (Quetelet 1831; Comte 1858), early criminologists adopted a range of quantitative tools (e.g., ratios and proportions, averages and deviations) in order to describe the crime problem, assess the causes of crime, and identify offenders' unique characteristics. Indeed, some scholars proposed that criminologists adopted these "positivist" methodologies in order to justify the implementation of new strategies of penalty and correction (Foucault 1979). However, most would agree that by applying quantitative methods to the study of crime, our understating of human behaviors was extended, establishing the foundation to develop crime causation theories.

This entry reviews the important quantitative methods adopted and used by criminologists during the nineteenth century. It begins with a brief overview on the emergence of the positivistic approach within the criminological discipline. It then describes the common data used by criminologists to study crime and its causes. Next, it elaborates the quantitative approaches employed by early criminologist to analyze crime and social data, using examples from early crime scholars' works. Finally, this entry concludes by discussing the unique influence of quantitative methodologies on the evolution of criminological theory. Importantly, this review encompasses articles and books that appeared between the years 1800 and 1900 and that were written in or translated into English.

Background

Positivist criminology emerged in France during the early nineteenth century as a result of the shift from an amorphous penal system (led by the ancient régime until the mid eighteenth century) to a more centralized system with a network of reform institutions (Beirne 1987). Influenced by enlightenment ideas of rationalism and humanism, and guided by the writings of prominent "classical school" philosophers (e.g., Cesare Beccaria and Jeremy Bentham), the stated objective of the new penal system was to promote moral rehabilitation among delinquent inmates (mainly alcoholics, idiots, immigrants, prostitutes, and petty and professional criminals). The penal institutions were designed to employ systematic strategies of surveillance and detention in an effort to "normalize" the conduct of the "dangerous class" (Beirne 1987). Soaring crime and increased recidivism at the beginning of the nineteenth century signaled the failure of these institutions to accomplish their task. In an effort to uncover the magnitude of crime in the country, the Ministry of Justice in France initiated in 1825 the first centralized national data collection of crime statistics. This resulted in a publication of the first annual report of national crime statistics in 1827 (Friendly 2007).

During the same time period, a positivist approach (i.e., an epistemology suggesting that natural phenomena could be empirically studied based on unbiased and objective scientific procedures) had dominated the scientific discourse within the physical sciences. Adherents of this approach suggested that unbiased scientific investigations should follow the positivist principles of observation, experiment, comparison, and historical methods (Timasheff 1967), and employ mathematical operationalization of the world. Acknowledging the unique contribution of positivist methods in the physical sciences, early social scientists (e.g., Quetelet (1831), and Comte (1858)) suggested that these methods and principles are also applicable to "social data." According to Quetelet (1831), quantitative methods could be used in the study of social data if those data are able to stand outside of the observer and are capable of being analyzed by mathematical procedures. Drawing on Quetelet's and Comte's monumental works, early social scholars developed interest in "positivist" methods and implemented them using a range of social data. This trend also influenced many crime scholars, who adopted these quantitative tools and used them to analyze crime data.

Crime Data

Similar to contemporary sources of "crime statistics" and "social data," information on crime and criminals during the early nineteenth century was collected by (1) government officials (i.e., courts and police) and by (2) individual scholars observing, surveying, and conducting experiments with criminals.

Official Crime Data. Although official efforts to collect national crime data began in 1825, the systematic collection of population-based data (i.e., social data) was initiated in the mid-fifteenth century with the London Bills of Mortality (the London Bills of Mortality were collected weekly and established the main source of mortality statistics in fifteenth-century England) (Friendly 2007). Early analyses of these mortality data demonstrated the informative value that state and public officials could gain from such "social data." As a consequence, by the mid-eighteenth century, governmental efforts to collect population-based data from its various agencies were expanding. The extensive collection of official crime statistics by the French Ministry of Justice in 1825 set the stage for the collection of data on criminal-justice agencies (e.g., police departments, courts, and prisons), as well as other population data (e.g., immigrants, prostitutes, literacy, and wealth). Beginning in the late 1830s, governmental efforts of crime data collection spread to England (Rawson 1839) and other European countries (e.g., Germany, Italy, Russia), and by the early 1850s were initiated in the USA. (Deflem 1997). While the accumulation of official crime statistics enabled early criminologists to develop an initial understanding on the distribution of crime and its magnitude, the availability of "social data" allowed these scholars to also test hypotheses regarding the social and ecological causes of crime.

Data Collected by Individual Scholars. In addition to official data collected by governmental officials, individual scholars gathered their own data on crime and delinquency directly from criminals. These scholars used a range of scientific methods – including observations, surveys, and experiments – to collect this information (Rafter 2004). Pinel, for instance, collected data on insane criminals using case studies and direct observations (Rafter 2004). Lombroso and his students collected measures of skulls, body, and morphological features of dead and alive criminals (Lombroso 1876; Lombroso and Ferrero 1893). These scholars also used experimental designs to collect data from criminals and noncriminals (Lombroso and Ferrero 1893). Finally, Mayhew (1862) and also Vidocq (Lindesmith and Levin 1937) used participant observations, testimonies by criminals, and life history documents to identify social gestures among prisoners and pickpockets. Although these methods of data collection were limited in scope, they yielded unique data sets that supported the development of biological and psychological perspectives on crime and shaped the evolution of criminological theory. The emergence of "quantitative methods" allowed early criminologists to empirically test the validity of these perspectives.

Quantitative Methods

Adolphe Quetelet was the first social scientist to apply quantitative methods in the study of crime. Drawing on his extensive background in mathematics and statistics, Quetelet believed that the same regularities and laws observed in the natural sciences could be identified in the social world. Identification of these laws, he suggested, should be based on a large amount of social data and involve the application of statistical methods and calculations (Beirne 1987; Stigler 1986). Following the publication of "*Research on the Propensity for Crime at Different Ages*" (Quetelet 1831), criminologists in France and around the globe implemented probability theory to *describe* the magnitude of crime (Guerry 1833; Rawson 1839; Levi 1880) and the work of criminal-justice agencies in their countries (Falkner 1889; Fletcher 1850; Mayhew 1862), to *identify* offenders' unique characteristics (Lombroso, 1876, 1891), and to *search* for the causes of deviance and crime (Clay 1857; Quetelet 1842). In the absence of methodologies that estimate correlations and regression coefficients (These approaches were developed around 1890 by Francis Galton and Karl Pearson (Stigler 1986)), these early crime scholars produced scientific works that: (1) classified crimes by types and presented frequency distributions; (2) calculated crime rates, proportions, averages, and deviations; (3) compared pairs of variables; (4) estimated the relationships between pairs of variables; and (5) presented quantitative graphics.

Crime Classification and Frequency Distributions. Once crime data became available to social scientists, criminologists were particularly interested in understanding the magnitude of deviance and crime. To achieve this goal, scholars like Quetelet (1831), Rawson (1839), and others (Fletcher 1841; Guerry 1833) presented annual figures on the sheer number of persons charged, convicted, and acquitted with different types of offenses. Offense type classifications were either general (i.e., distinguishing between crimes against persons and crimes against property (Quetelet 1831)) or very detailed (i.e., presenting a breakdown of all reported offenses (Rawson 1839; Fletcher 1841)). With the distinction made by official crime agencies between offenses and offenders, scholars like

Levi (1880) presented data on both the number of indictable offenses reported to the police and the number of persons apprehended for indictable offenses.

The accumulation of data on criminal-justice agencies allowed early criminologists to also describe the size, structure, and workload experienced by police stations and prisons. Levi (1880), for instance, reported the annual number of police officers in England and Wells between the years 1857 and 1878, while Fletcher (1850) provided detailed description of police wards' force size, officers' salaries, and workload in London. Prisons statistics were also popular during that time, with scholars like Mayhew (1862), Lombroso (1876), and Falkner (1889) describing the number of prisoners, their offenses, and sentence lengths in British, Irish, and American prisons, respectively.

Proportions, Rates, Averages, and Deviations. In order to allow a meaningful assessment of the magnitude of crime, early scholars adopted Laplace's method of ratio estimation (Stigler 1986) and calculated the ratios of the annual offenses to population (Guerry 1833; Levi 1880), proportions of offenders per populations (Rawson 1839), and proportions of prisoners to population (Falkner 1889). Quetelet (1831) also calculated the conviction rates of criminal offenders in France between the years 1825 and 1830, dividing the number of convicted criminals by the number of accused. Quetelet interpreted these rates as probabilities, suggesting that they indicate the chances that an average person (whom nothing more is known about) is convicted of a crime.

Next to estimating the average propensities of offenders to be convicted, Quetelet also calculated the average values of his subjects' physiques (i.e., weight, height). According to him, any aspect of a population that can be measured may produce an indicator that portrays the "average man" in that population (Quetelet 1831; Stigler 1986). This average score could then indicate either an actual physical characteristic (i.e., height, weight) or an estimated "propensity" (for instance, using crime rates to calculate the propensity to offend). Once known, this mean could be cross-tabulated with a range of demographic, social, and contextual indicators, and allow for the development of meaningful comparisons that could refine our understanding of the laws of the social world.

Comparison between Pairs of Variables. Drawing on Quetelet's (1831, 1842) influential works, numerous early crime scholars compared the distribution of crime measures across a range of social groups, time periods, and geographical regions. These comparisons included calculations of social groups' share in producing crime, and the overall proportions of offenders in the group. Results were mainly reported using tables. Levi (1880), for instance, reported that only 20 % of those committed to trial between the years 1857 and 1878 in England and Wales were women. Focusing on the differential involvement of different age groups in crime, Rawson (1839) Fletcher (1841) and Levi (1880) ordered several age categories (e.g., under 17 years, between 17 and 21 years) and compared the proportion of offenders in each category. These scholars also reported the proportion of male and female offenders across the different age categories. Finally, Fletcher (1841) calculated the "proportion of offenders" across level of instruction in the United Kingdom between the years 1836 and 1842, finding that less than 10 % of offenders were able to read and write well.

Numerous early studies also compared crime statistics across countries and various geographical regions (Levi 1880; Lombroso 1911; Rawson 1839; Fletcher 1850). Rawson (1839), for instance, compared the proportion of offenders within different age groups in England and France and found that in both countries, more than 30 % of male and female offenders were between 21 and 30 years of age. Analyzing data from England and Wales, Levi (1880) compared the average number of persons committed for trial across 51 counties and found substantial variation across counties. Finally, Falkner (1889) compared the proportion of female convicts in 27 states, finding high proportion of female convicts in North Carolina and Maryland, yet low proportions in Texas and Colorado. It should be noted that in the absence of appropriate methodologies allowing examination of significant differences (for instance, t-tests and/or one way ANOVA), early criminologists were limited in the informative value of their comparisons.

While comparisons of crime proportions and averages were common among early scholars that used official crime data, such comparisons were rare among scholars like Lombroso and his students that collected data directly from criminals (using observations and experiments). Thanks to the relatively low number of observations, these scholars used the actual values of the measures to generate comparisons between individuals. In the "Criminal Man," for instance, Lombroso (1876) compared the circumference of 66 criminals' skulls, and reported that these skulls were "abnormally small." In the third edition of this book, Lombroso also presented comparisons between the anomalies in the brains of 28 criminals as well as comparison between the anomalous physical characteristics of 219 male criminals. Indeed, Lombroso's comparisons were accepted with suspicion in the criminological field. Nevertheless, along with comparisons of crime proportions and averages, these comparisons set the stage for the emergence of the first attempts to assess relationships between pairs of variables, and provided fertile ground for the development of sociological, psychological, and biological explanations of crime.

Estimations of the Relations between Pairs of Variables. In further efforts to identify the causes of deviance and crime, several criminologists investigated the joint variation between crime indicators and social, psychological, and biological measures. In the absence of methodological tools allowing statistical estimation of the relationships between measures, these scholars used tables to organize the values of their measures and then looked for consistent variation between these measures. Rev. John Clay (1857), for instance, was interested in the relationship between the number of criminals, beer houses, and schools attendants in geographic areas. To test these relationships, he composed a table listing 40 counties in England and presented the number of criminals (per 100,000 of the population), beer houses (per 100,000 people), and attendants at school (per 10,000) in each county. Observing the joint variation of these measures, Clay determined that beer houses are associated with high levels of crime, while attendance in school is associated with low levels.

This approach allowed early scholars who represented different criminological perspectives to empirically assess the validity of their theoretical arguments. Lombroso and Ferrero (1893), for instance, compared the variation of cranial capacities across groups of normal and criminal females. In line with Lombroso's theory of the criminal man (Lombroso 1876), these authors found that the cranial capacity of female criminals and prostitutes was lower than the cranial capacities of normal females. Focusing on the social causes of deviance and crime, Durkheim (1897) compared the variation of suicide rates with the percentage of married couples across 16 Italian provinces. Consistent with the assumption that low integration increases the probability of suicide, Durkheim found that provinces with a low percentage of marriages experienced high suicide rates, while provinces with a high percentage of marriages experienced low rates. The availability of quantitative graphics further supported these early comparisons, allowing visualization of the relationship between measures.

Quantitative Graphics. By the end of the first half of the nineteenth century, all the major modern forms of data visualization (including bar-charts, histograms, time-series charts, and thematic maps) had already been invented (Beniger and Robyn 1978; Friendly 2009) and their advantage in presenting patterns, distributions, anomalies, and variations of measures was highly acknowledged. Guerry (1833) and Quetelet (1831) took active part in developing these statistical graphics. Guerry (1833), for instance, was first to produce histograms (Karl Pearson coined the term "histogram" in 1895 to describe this form of statistical graphical presentation (Stigler 1986)) by arranging ordered categories for continuous measures (i.e., age, months) and using columns of equal width to represent the frequency of each category of the data (Beniger and Robyn 1978; Friendly 2007). Generating a histogram to compare the distribution of different types of crimes over age categories, Guerry observed that indecent assault is more common during childhood and less frequent during adulthood. In his late editions of *L'uomo delinquente*, Lombroso (1889, 1896) also employed bar-graphs and histograms to compare the distributions of criminals' and noncriminals' cranial capacities, height and arm span, and parents' age. Finally, Dexter (1899) composed a histogram to compare the frequency of murder and suicide incidents over conditions of temperature. Interestingly, although utilizing bar-graphs for comparison purposes, there is no indication that early criminologists were aware of the scatterplot (first scatterplot was published in 1833 in an astronomy journal (Friendly 2007)).

In addition to bar-charts and histograms, time-series charts were commonly used by early scholars (Guerry 1833; Lombroso 1876; Du Bois 1899; Dexter 1899; Mayhew 1862). Taking advantage of increasingly accumulating annual crime data, scholars produced time-series plots to describe the fluctuation of crime over time. Guerry, for instance, presented a time-series line graph of death sentences and executions in France between the years 1825 and 1855 (Friendly 2007). Employing the unique visualization allowed by time-series diagrams, Mayhew (1862) produced a time-series chart describing the fluctuation of corn price and the ratio of criminals to population between the years 1834 and 1849 in London. Finally, Lombroso (1896) produced a graphical analysis of periodic variation of alcohol consumption in ten countries (including the USA) between the years 1870 and 1895.

The invention of the statistical shaded map by Dupin in 1826 (Friendly 2007) initiated a series of publications featuring shaded maps and describing the distribution of crime rates across geographical regions. Balbi and Guerry were first to use shaded maps for portraying crime rates across the departments of France (Friendly 2007). In their innovative work, these scholars also compared the spatial distribution of crime against persons and crime against property and education level using a set of shaded maps of France (Friendly 2007). In 1833, Guerry refined these maps (using more complete data) and added three additional shaded maps displaying the distribution of illegitimate births, donations to the poor, and suicide in France. Guerry's unique approach that compared the spatial distribution of crime data over a range of social measures had been accepted with enthusiasm in the criminological field, and was adopted by many prominent scholars (Quetelet 1842; Levi 1880; Mayhew 1862; Durkheim 1897; Lombroso 1876). Drawing on Enrico Ferri's analysis of homicide in Italy, Lombroso (1876), for instance, compared the spatial distribution of murder, parricide, and poisoning in Italy. Levi (1880) presented the spatial distribution of indictable offenses and offenses subject to summary jurisdiction between the years 1857 and 1876 across counties in England and Wells and compared it with the spatial distribution of ignorance, amount of savings, and poverty. Finally, Durkheim (1897) presented shaded maps displaying the spatial distribution of suicide rates across counties in France and compared it with the spatial distribution of wealth, family size, and alcoholism.

Summary and Conclusions

In sum, the accumulation of crime data and the adoption of quantitative methods to study crime supported early criminologists' efforts to describe the magnitude of crime, identify offenders' unique characteristics, and search for the causes of deviant behaviors. These data and methods paved the way for the development of "positivist criminology." Indeed, the common quantitative tools available during the nineteenth century were relatively limited in their ability to generate statistical inferences from a sample. Nevertheless, these initial methodologies permitted the construction of frequency distributions, calculations of crime rates, proportions, averages and deviations, and comparisons between pairs of variables. These methods were employed in a considerable amount of early criminological studies and extended our understanding of crime and delinquency.

Next to their value in describing criminals and crime, these early quantitative methods also influenced the evolution of criminological theories. Specifically, the shift from a traditional medieval view of the universe to a philosophy favoring physical facts required empirical validation of scientific explanations of the world. According to this new philosophy, the world could be characterized by mathematical and statistical principles that a rational mind could grasp. The major advantage of these early quantitative tools was their ability to operationalize the world (e.g., producing crime rates) and investigate the relationships between different constructs through scientific research. This allowed early crime scholars to offer theoretical explanation for the causes of crime, draw research hypotheses from these theories, test these hypotheses empirically, and then refine their explanations. Lombroso's (1867) consistent modification of his original thesis of the criminal man is only one example of this process. Thus, the application of quantitative methods in the study of crime shifted the focus of criminological research from narrow discrete observations to broader discussions portraying general principles on the causes of crime.

Related Entries

- ▶ History of Geographic Criminology Part I: Nineteenth Century
- ▶ History of the Self-Report Delinquency Surveys
- ▶ History of the Statistics of Crime and Criminal Justice

Recommended Reading and References

Beirne P (1987) Adolphe Quetelet and the origins of positivist criminology. Am J Sociol 92:1140–1169

Beniger JR, Robyn DL (1978) Quantitative graphics in statistics: a brief history. Am Stat 32:1–11

Clay J (1857) On the relation between crime, popular instruction, attendance on religious worship and beer house. J Stat Soc London 20:22–32

Comte A (1858) The positive philosophy of Auguste Comte (trans: Martineau H). Calvin Blanchard, New York

Deflem M (1997) Surveillance and criminal statistics: historical foundations of governmentality. Stud Law Pol Soc 17:149–184

Dexter EG (1899) The influence of weather upon crime. Pop Sci Mon 55:653–660

Du Bois WEB (1899) The Philadelphia negro: a social study. University of Pennsylvania, Philadelphia

Durkheim E (1897) Suicide: a study in sociology (trans: Spaulding JA, Simpson G). The Free Press, Glencoe

Falkner RP (1889) Prison statistics of the United States for 1888. Publ Am Stat Assoc 1:299–327

Fletcher J (1841) Progress of crime in the United Kingdom: abstracted from the criminal returns for 1842 and the prison returns for the year ended at Michaelmas, 1841. J Stat Soc London 6:218–240

Fletcher J (1850) Statistical account of the police of the metropolis. J Stat Soc London 13:221–267

Foucault M (1979) Discipline and punish: the birth of the prison (trans: Sheridan A). Vintage, New York

Friendly M (2007) A-M Guerry's moral statistics of France: challenged for multivariable spatial analysis. Stat Sci 22:368–399

Friendly M (2009) The golden age of statistical graphics. Stat Sci 23:502–535

Guerry AM (1833) Essay on the moral statistics of France (trans: Reinking VW). Edwin Mellen Press, Lewiston

Levi L (1880) A survey of indictable and summary jurisdiction offenders in England and Wales, from 1857 to 1876, in Quinquennial periods and in 1877 and 1878. J Stat Soc London 43:423–461

Lindesmith A, Levin Y (1937) The Lombrosian myth in criminology. Am J Sociol 42:653–671

Lombroso C (1891) Illustrative studies in criminal anthropology. Monist 1:177–196

Lombroso C (1876) The criminal man (trans: Gibson M, Rafter NH). Duke University Press, Durham

Lombroso C, Ferrero G (1893) Criminal women, the prostitute and the normal women (trans: Rafter NH, Gibson M). Duke University Press, Durham

Mayhew H (1862) The criminal prisons of London and scenes of prison life. Griffin, Bohan, London

Nicole R (2004) The unrepentant horse-slasher: moral insanity and the origins of criminological thought. Criminology 42:979–1008

Quetelet A (1831) Research on the propensity for crime at different ages (trans: Sylvester SF). Anderson, Cincinnati

Quetelet A (1842) A treatise on man and the development of his faculties (trans: Diamon S). Scholars' Facsimiles and Reprints

Rawson RW (1839) An inquiry into the statistics of crime in England and Wales. J Stat Soc London 2:316–344

Stigler SM (1986) The history of statistics: the measurement of uncertainty before 1900. Harvard University Press, Cambridge

Timasheff NS (1967) Sociological theory. Random House, New York

Quantitative Studies on Media and Crime

Jeff Gruenewald[1], William Parkin[2] and
Steven M. Chermak[3]
[1]Department of Sociology and Criminal Justice,
University of Arkansas, Fayetteville, AR, USA
[2]Seattle University, Seattle, WA, USA
[3]School of Criminal Justice, Michigan State
University, East Lansing, MI, USA

Synonyms

Content analysis

Overview

This entry provides an evaluation of quantitative news media studies of crime. The entry includes a summary of how scholars have approached the study of news media using quantitative content analysis over the last 20 years by identifying what type of news media have been studied, how they have been studied, sampling designs, types of analysis conducted, and how reliability issues have been addressed. In addition, gaps in the approach and specific suggestions for moving forward in this important research area are discussed.

Studying the presentation of crime and criminal justice in the news media is important for several reasons. First, most of the public has infrequent direct experiences with the criminal justice system. It is because of this limited direct experience with crime that the media have great potential to influence how the public thinks about it.

Second, the news media are critical to public debates about policy issues. The public is overwhelmed with the amount of information available on various issues and is ill-equipped to do the types of analysis that might lead to a critical understanding of these issues. What the media provide is a brief, limited discussion of the factors related to policy decisions. Policymakers recognize the significance of the media, providing reporters access to information, hearings, and individuals that can provide the context that guides these debates.

Third, although scholars have historically been interested in understanding the images of crime and criminal justice, this interest appears to have grown dramatically in the past 20 years. For example, research indicates that only 36 crime and media studies were published between 1975 and 1995 (Chermak 1998). In contrast, research conducted for this entry identified nearly 100 studies published between 1990 and 2010 to examine. There are multiple factors that have probably contributed to the increases in research in this area, including the general awareness of the important role that media play in society, the increase in number of news outlets, and changes in how the public engages the mass media. Scholars appear to have taken note of these factors, and the end result has been a significant increase in research examining media, crime, and society.

In the sections that follow, quantitative media content studies are examined. In the next section, what has been uncovered in past research is highlighted. The methodology used to evaluate the quantitative media research is then discussed. The findings are then discussed and followed by a conclusion.

Prior Research Examining Crime News

This section reviews studies that have used quantitative content analysis to examine news media portrayals of crime. First, a brief overview of common theoretical approaches used by scholars to explain apparent discrepancies between the actuality of crime and news media representations of crime is discussed. Second, a selection of key findings from relevant studies are highlighted. Third, how past research findings vary across types of media and research designs is addressed.

In the past, scholars have relied primarily on two general theoretical frameworks to account for discrepancies between news media representations of crime and the actuality of crime captured in crime statistics. The first explanation suggests that crime news content is shaped largely by the structure of the dynamic newsmaking process and the need to efficiently package and sell news to audiences (Fishman 1980). Scholars have used this structural perspective to explain why extraordinary, and especially serious, crimes are overemphasized in the news. The second explanatory model maintains that cultural typifications, often based on shared expectations of "normal crime" participants and situations, influence crime story coverage decisions. This perspective explains why racial and ethnic minorities are often portrayed differently than whites as crime participants.

Crime is an everyday occurrence, and crime stories are a consistent feature of the daily news cycle (Chermak 1994; Graber 1980; Surrette 1998). The constant stream of potential crime stories requires reporters to make daily news selection decisions. Newsmakers select which crimes fill available news space and which crimes are ignored. To work most efficiently, "crime beat" reporters develop relationships with sources that provide reliable information about local crime (Chermak 1994; Ericson et al. 1989). Reporters' close working relationships with police are especially useful because they permit access to official information about crimes from trustworthy sources. As police largely control the information released to reporters, official police sources maintain substantial influence on crime news coverage (Ericson et al. 1989).

Crime reporters also take cues from fellow newsmakers regarding which crimes stories are of most interest to audiences (Ericson et al. 1987, 1989). Understanding audience preferences is important to selling news (Chermak 1994; Sheley and Ashkins 1981). For instance, research has shown that media consumers are often most interested in "human interest stories" or stories with which audiences can sympathize. Such stories tend to simultaneously evoke sympathy for victims of crime and disdain for those that commit them. Newsmakers also consider crimes involving extraordinary or novel circumstances newsworthy and likely to catch consumers' attention. Serious crimes, usually involving acts of street violence, are also considered particularly newsworthy and lead newsmakers to adopt the mantra of "if it bleeds, it leads."

The first explanatory model suggests that organizational practices of news agencies and shared understandings of newsworthiness structure the newsmaking process and result in discrepancies between news media representations of crime and crime statistics. Moreover, it has been suggested that overemphasis on novel and serious forms of violence may lead to increased public fear of crime and perceived risk of victimization (Liska and Baccaglini 1990). Disproportionate amounts of coverage devoted to crimes against females or vulnerable groups, such as children and the elderly, create the false impression that these groups are the most likely to be targeted. This is especially true when crime stories fail to provide adequate context.

The second explanatory model suggests that news media rely on cultural understandings of "normal" crime elements and expectations of typical victims and offenders (Pritchard and Hughes 1997; Sudnow 1965). This perspective maintains that traditional indicators of newsworthiness, such as novelty and crime seriousness, may be less important for explaining why some crime participants receive disproportionate amounts of news attention. This perspective maintains that newsmakers rely on cultural scripts that outline typical victims and offender types and typical criminal situations in which they are most likely to be involved (Gilliam and Iyengar 2000). Crime scripts are rooted in gendered and racialized stereotypes of crime participants and the criminal behavior most associated with them. Negative portrayals of racial minorities as threatening predators may be linked to the dominant ideologies and worldviews of news agency personnel (Bjornstrom et al. 2010; Blalock 1967). Scholars have relied on this perspective, for instance, to explain why relatively rare forms of violence, such as female-on-male crime or white-on-black crime, receive disproportionately less news attention despite representing uncommon crime characteristics (e.g., Lundman 2003). In sum, the cultural typification perspective remains useful for explaining discrepancies between actual criminal participation of racial and ethnic minorities and news media portrayals of minorities as crime victims and offenders.

Key Findings of Past Research

Drawing from these theoretical perspectives, prior content analyses of crime news have revealed that news media representations are often misaligned with established crime patterns. Most studies have focused on print news, though others have also examined television news (Entman 1992; Entman and Rojecki 2000; Graber 1980; Sheley and Ashkins 1981). While similar newsmaking processes dictate selection decision-making for print and television news media formats, representations of crime vary across formats (Ericson et al. 1991; Sheley and Ashkins 1981). Television news content, for instance, is focused more on capturing stories visually through dramatic imagery. Therefore, stories that can be told through brief video footage are often preferred. Selling television crime news is also much more competitive than selling newspapers due to the presence of rival news organizations contending for ratings, thus making marketability more important than newsworthiness (Sheley and Ashkins 1981). While newspaper editors have more flexibility in filling news space (Chermak 1995:122), telling crime stories in video segments that may last just minutes or less can be challenging.

One of the most consistent findings has been that violent crimes are overrepresented in news media (Chermak 1994; Ericson et al. 1991; Graber 1980; Surrette 1998). In one notable study, Graber (1980:39) found that murders made up only 2 % of Chicago crime and over 26 % of all crime mentions, while the more common and less serious crimes received substantially less coverage. The few studies that have compared news media coverage across multiple research sites have found that this finding holds across cities (e.g., Graber 1980). Furthermore, it tends to be the most heinous of violent street crimes that are likely to be covered and sensationalized. Relatively rare forms of serious violence, such as murders involving multiple victims or random victims, are considered particularly newsworthy (Hall et al. 1978). On the other hand, nonviolent forms of crimes (e.g., white-collar crime) receive less attention from the news media (Chermak 1995; Graber 1980).

As for portrayals of crime participants, females generally receive more news attention as crime victims compared to males. Other victims that can be portrayed as vulnerable, such as especially young and elderly victims, are also considered “worthy victims” and receive disproportionate news attention (Chermak 1994; Graber 1980). The opposite is generally true of offenders, as male offending tends to be overrepresented by news media. Victims involved in deviant lifestyles (e.g., gangs, drugs) may also be considered less worthy victims and receive substantially less news media attention (Chermak 1994; Pritchard and Hughes 1997; Gruenewald et al. 2011). Research has found that the race of crime participants also remains important. Past studies, for instance, suggest that whites are more likely to be portrayed favorably as victims, while victims of color are more likely to be portrayed less favorably and are generally overrepresented as offenders (Bjornstrom et al. 2010; Chiricos and Escholz 2002; Dixon et al. 2003; Entman 1992; Entman and Rojecki 1994, 2000).

Different research approaches have been used by scholars to analyze crime news content. These approaches can be distinguished by if, and how, they rely on outside sources of crime data. One approach, for instance, has been to study news media crime content for a particular period of time without comparing crime representations to actual crime statistics. This type of study is concerned with content-analyzing characteristics of crimes that are covered, thus ignoring crimes not covered by news media. Studies have compared characteristics and roles of crime participants across race and gender categories (e.g., Dixon et al. 2003). This research has found that racial-ethnic minorities are more likely to appear as dangerous offenders linked to street violence than as victims (Chiricos and Escholz 2002; Dixon et al. 2003; Entman 1992; Entman and

Rojecki 1994, 2000). On the other hand, other racial minorities, such as Latinos and Asians, are often underrepresented as crime participants.

A second approach has been to compare news media coverage of crime to aggregate crime statistics in order to gauge the extent that news reflects the actuality of crime. Past research on news media coverage of crime has suggested that "...media crime reporting apparently bears little resemblance to the 'reality' of police statistics" (Sheley and Ashkins 1981:503). Official crime data sources aid in capturing general patterns of known crimes. Some scholars have relied on aggregate crime statistics, such as those from the FBI's Uniform Crime Report, and local news media coverage in a particular city to study news portrayals of crime. Others have used police statistics from a single city. Graber's (1980) seminal study comparatively examined Chicago police department crime statistics and local newspaper coverage (*Chicago Tribune*). In another important study, Chermak (1995) examined representations of crime victims in the news media for multiple cities and found that news media representations of crime are distorted in a number of ways. In particular, his research demonstrated how serious crimes and crimes against victims belonging to vulnerable groups received disproportionate news attention. Content-analyzing crime stories from multiple print and electronic news sources, Chermak (1995) compared incident, suspect, and victim crime characteristics presented in Uniform Crime Report statistics. Overall, Chermak's (1995) work revealed how these characteristics interacted to affect the newsworthiness of criminal events.

A third avenue of research has been to comparatively examine local news coverage of actual crimes. Though there have been exceptions (e.g., Sheley and Ashkins 1981), nearly all of these studies have focused on newspaper coverage of homicides. There are a number of benefits to conducting his type of content analysis (Pritchard 1985). One benefit is that it is possible to ensure that crimes are occurring within geographic locales that news organizations actually cover. Linking local coverage to local crimes guarantees apples are not being compared to oranges (Pritchard 1985:501). Another benefit to comparing local news coverage to actual crimes also helps to ensure that crimes actually occurred during the time period under consideration by content analysts. Most importantly, however, is that linking actual crime data with local news coverage allows content analysts to comparatively examine the similarities and differences between crimes that are covered with those crimes that are ignored (Pritchard 1985).

The vast majority of these "distortion analyses" have studied homicide (Gruenewald et al. 2011; Johnstone et al. 1994; Lundman 2003; Peelo et al. 2004; Pritchard 1985; Pritchard and Hughes 1997; Sorenson et al. 1998). In addition to being the most serious form of violent crime, studying homicide in this way has strategic advantages. In particular, police homicide data are likely to be reported and fully investigated. This process often creates lengthy, in-depth detectives' records that identify characteristics of offenders, victims, and situations. These characteristics then become both "search words" when seeking news media coverage of each homicide and potential indicators of newsworthiness for newsmakers to use in their news decision-making.

Findings have emerged from this growing body of research that supports both the structural and cultural explanations of crime news media distortion. Similar to findings on crime news in general, past studies have found that homicides viewed as more serious or extraordinary, such as those involving multiple victims, tend to receive more newspaper coverage (Johnstone et al. 1994). In addition, victims perceived as "vulnerable," such as young or elderly victims, are generally overrepresented (Sorenson et al. 1998). As for race, white victims are more likely to be covered than racial and ethnic minority victims (Johnstone et al. 1994; Peelo et al. 2004; Sorenson et al 1998). When black and Hispanic homicide victims receive news print coverage, then it tends to be less extensive coverage (Lundman 2003; Pritchard and Hughes 1997). Cultural typification explanations have also been utilized to explain discrepancies in news media coverage across race. While female homicide victims are also considered more vulnerable and newsworthy than male victims, recent studies have found that the victim and offender race may, in some instances, make female victim homicides less newsworthy (Lundman 2003). Other similar studies have found support for the claim that newsmakers rely on cultural scripts based on racialized and gendered stereotypes to evaluate crime story newsworthiness (Pritchard and Hughes 1997; Gruenewald et al. 2011; Lundman 2003).

Research Findings

This section discusses the basic characteristics of the sample of articles, presenting information about the frequency and types of journals that published media and crime-related content analyses, the types of media sources used in each study, the research methodologies employed, the level of statistical analyses, and finally reliability issues. Between 1990 and 2010, 95 articles were published that used content analysis of one or more media sources to study crime or the criminal justice system. Of those articles, only 25 % were published prior to 2000, while the remaining 75 % of the sample was published between 2000 and 2010. The identified articles were in 54 different journals, with *Criminal Justice & Popular Culture* publishing the most ($n = 7$), followed by *Justice Quarterly* ($n = 5$), *Deviant Behavior* ($n = 4$), and *Journalism Quarterly* ($n = 4$). The remaining 50 journals published less than four articles each over that 20 year period, 31 of which only published one article.

The topics explored in the 95 articles varied widely. Some types of crimes researched included homicide, terrorism, illicit drug use, domestic violence, identity theft, white-collar crime, carjackings, and school shootings. Other factors studied in relation to media coverage of crime included characteristics of victims and offenders such as their race/ethnicity, gender, age, and where they lived. Although the crimes and themes studied were not consistent, all used some type of media and content analysis to answer their basic research questions.

In order to identify the media and samples used in the studies, each article was coded for the media type, number of sources, sample dates, keywords used to identify appropriate content, and the subsequent sample size. These studies used a range of media types to answer their research questions. Data from newspapers were used in 77 % of studies, 28 % of studies used television news, and 7 % use magazines such as *Time* or *Newsweek*. In addition to media sources, 41 % of the articles supplemented their content analysis with non-news sources, such as police and court data, census data, survey data, and even medical records and coroner reports. Of all the articles reviewed, 85 % listed the number of sources they used for the content analysis. In 22 % of that subsample, only one media source was used, while almost a third of the articles used two or less sources. The vast majority of articles (78 %) used ten sources or less. When considering the full range of sources used, one article used 771 sources, skewing the average numbers of sources used to 21, even though the median number of sources used was 4, and the mode was 1.

Ninety-five percent of the articles specified the date range used to collect their sample. The earliest start date for materials used in the content analysis was 1892, while one study each collected data beginning in the 1940s, 1950s, and 1960s. Three studies began collecting data in the 1970s, 15 in the 1980s, 52 in the 1990s, and 16 after 2000. A little less than a third of the studies had a date range that was a year or less (31.1 %), 46.7 % took their sample from 2 years or less, and 81.1 % of the articles took their sample from a time period that was 10 years or less. The average time period from which a sample was drawn was 2,303 days (63 years) and the median was 1,096 (3 years). The minimum number of days was 3, while the maximum was 14,608 days or 40 years.

Within the identified sources, 59 % of all articles in the sample reported using some form of keywords or phrases to search for materials. Of that 59 %, 36 % did not list the keywords or phrases used. Of the remaining 64 %, the average number of keywords or phrases used was 47, with a median of 2 and a mode 1. In addition, only

28 % of all studies also used some type of random sampling to select media content, while the rest collected data on a universe of coverage. Sampling techniques used included simple random samples, stratified sampling, and constructed weeks. The units of analysis varied across study, some of which included newspaper articles, television news stories, quotes, and crime events. Almost all articles (97 %) reported their sample size. The smallest sample size was 23 and the largest was 55,000. The mean sample size was 1,771 and the median was 474. Thirty-seven percent of the studies had a sample size of 200 or less, and three quarters had a sample size of 1,000 or less. Within each sample, the number of variables used in each study ranged from 1 to 51, with an average of 11 and a median of 9. A little more than 40 % of the articles collected data on more than 10 variables. Some of the variables of interest included source type, crime type, location in paper, number of stories per crime, article or report length, and victim, offender, and incident characteristics.

Specific to the types of analyses presented in each article, the highest level of analysis for 26 % of the articles was multivariate, while 28 % analyzed data at the bivariate level, and 45 % presented information using only univariate analysis. Types of multivariate analysis included binary and multinomial logistic regression, OLS regression, path analysis, Poisson regression, negative binomial regression, and time series analysis. In addition, 19 % of the studies supplemented the empirical component of the research with a qualitative component. Importantly, only 41 % ($n = 39$) of the studies reported measuring reliability. Of that 41 %, all but one of the authors provided the actual reliability statistic. The lowest reliability score reported was 0.10, while the highest was 1.00. The average reliability score reported was 0.88 with a media score of 0.89. The types of reliability scores reported included Cohen's Kappa, Scott's Pi, intercoder reliability, coefficient of reliability, and Cronbach's alpha.

This descriptive information produces an idea of the type and frequency of articles being published that use a content analysis of media to look at one or more aspects of crime and the criminal justice system. In the next section, the implications of this data on the current and future field of crime and media research are discussed.

Discussion

After reviewing previous research that has studied media coverage of crime and criminal justice, this entry discussed preliminary findings from a study of all peer-reviewed articles published on these topics between 1990 and 2010. This study focused only on articles published about media organizations in the United States. It is important to note that books, edited volumes, and book chapters were excluded from the analysis. There were several critical findings that were discovered. First, the use of quantitative content analysis has grown dramatically. There are several explanations for this growth, including that several disciplines, including criminal justice and criminology, have increasingly relied on quantitative methodologies and access to media content is virtually uninhibited because of open source databases and newspaper retrieval services. Second, most of the studies have been published in the past 10 years. This provides evidence that the field has embraced these technologies and there are much higher expectations regarding the level of sophistication with media analysis. Third, it is interesting that most of the studies focus on a single type of medium (either newspaper or television) and articles are selected from a narrow range of sources. This is somewhat surprising because access to a broad range of sources is relatively easy because of the existence of multiple media databases. Fourth, it is also important to note that almost 75 % of the articles relied on simple statistics rather than multivariate analysis. Although quantitative content analysis has grown dramatically, there still appears to be a lack of sophistication and rigor regarding statistical analysis. Finally, less than half of the articles reported basic reliability statistics.

This is an important study that should open up dialogue about the use of different statistical techniques to study crime and justice in the news.

Future research should look more closely at the variations in coverage by discipline, topic, and media source. The vast majority of articles using multivariate statistics use what has been referred to as "media distortion analysis" – where the researcher compares an "objective reality" to a "media reality." Other studies of basic content issues, however, continue to rely on simple univariate statistics. Although these studies provide some general understanding of the coverage of key issues in the media, future research will have to consider how such statistical techniques might be more broadly used.

Media studies are critically important and engage various issues and different disciplines in interesting ways. This study provided evidence of this impact and revealed a wide variety of disciplines that have examined how news media cover crime-related issues. The good news is that scholars are looking to various types of methodologies and techniques to study these issues more closely. The downside is that there remains much room for improvement. It seems vital that future research continues to consider how different methodologies (both quantitative and qualitative) and statistical techniques can be used to further the understandings regarding images of crime and criminal justice in the news.

Related Entries

- ▶ Crime and Justice on Television
- ▶ History of the Statistics of Crime and Criminal Justice
- ▶ Use of Social Media in Policing

Recommended Reading and References

Bjornstrom RL, Kaufman RL, Peterson RD, Slater MD (2010) Race and ethnic representations of lawbreakers and victims in crime news: a national study of television coverage. Soc Probl 57:269–293

Blalock HM (1967) Toward a theory of minority-group relations. Wiley, New York

Chermak SM (1994) Body count news: how crime is presented in the news media. Justice Q 11:561–582

Chermak S (1995) Victims in the news: crime in American news media. Westview Press, Boulder

Chermak SM (1998) Predicting crime story salience: the effects of crime, victim, and defendant characteristics. J Crim Just 26(1):61–70

Chiricos T, Escholz S (2002) The racial and ethnic typification of crime and the criminal typification of race and ethnicity in local television news. J Res Crime Delinq 39(4):400–420

Dixon TL, Azocar CL, Casas M (2003) The portrayal of race and crime on television network news. J Broadcast Electron 47:498–523

Entman RM (1992) Blacks in the news: television, modern racism, and cultural change. Journal Q 69: 241–261

Entman RM, Rojecki A (1994) Representation and reality in the portrayal of backs on network television news. Journal Q 71:509–520

Entman RM, Rojecki A (2000) The black image in the white mind: media and race in America. University of Chicago Press, Chicago

Ericson RV, Baranek PM, Chan JBL (1987) Visualizing deviance: a study of news organizations. Toronto: University of Toronto

Ericson RV, Baranek PM, Chan JBL (1989) Negotiating control: a study of news sources. University of Toronto Press, Toronto

Ericson RV, Baranek PM, Chan JBL (1991) Representing order: crime, law, and justice in the news media. University of Toronto Press, Toronto

Fishman M (1980) Manufacturing the news. University of Texas Press, Austin

Garofalo J (1981) Crime and the mass media: a selective review of research. J Res Crime Delinq 18:319–350

Gilliam FD Jr, Iyengar S (2000) Prime suspects: the influence of local television news on the viewing public. Am J Polit Sci 44:560–573

Graber D (1980) Crime news and the public. Praeger, New York

Gruenewald J, Chermak SM, Pizarro J (2011) Covering victims in the news. Justice Q. Retrieved online December 1, 2011 from doi:101080/074188252011628945

Hall S, Critcher C, Jefferson T, Clarke J, Roberts B (1978) Policing the crisis: mugging, the state, and law and order. Holmes and Meier, New York

Johnstone JWC, Hawkins DF, Michener A (1994) Homicide reporting in Chicago dailies. Journal Q 71:860–872

Liska AE, Baccaglini W (1990) Feeling safe by comparison: crime in the newspapers. Soc Probl 37(3):360–374

Lundman RJ (2003) The newsworthiness and selection bias in news about murder: comparative and relative effects of novelty and race and gender typifications on newspaper coverage of homicide. Sociol Forum 18:357–386

Peelo M, Francis B, Soothill K, Pearson J, Ackerley E (2004) Newspaper reporting and the public construction of homicide. Br J Criminol 44:256–275

Pritchard D (1985) Race, homicide, and newspapers. Journal Q 62:500–507

Pritchard D, Hughes KD (1997) Patterns of deviance in crime news. J Commun 47:49–67

Sheley JF, Ashkins CD (1981) Crime, crime news, and crime views. Public Opin Quart 45:492–506

Sorenson SB, Manz JGP, Berk RA (1998) News media coverage and the epidemiology of homicide. Am J Public Health 881:1510–1514

Sudnow D (1965) Normal crimes: sociological features of the penal code in a public defender office. Soc Probl 12:255–276

Surrette R (1998) Media, crime, and criminal justice: images and realities, 2nd edn. Wadsworth, Belmont

Quasi-Experimental

► Establishing Causes of Offending in Longitudinal and Experimental Studies

R

Race

► Juvenile Justice System Responses to Minority Youth

Race and Ethnicity in Social Disorganization Theory

April Fernandes[1], Suzanna Ramirez[2] and
Robert D. Crutchfield[1]
[1]University of Washington, Seattle, WA, USA
[2]The University of Queensland, St. Lucia,
Australia

Overview

Race and ethnicity were central to the early formulations of Social Disorganization Theory, and consideration of these social categories remains significant in contemporary criminological research. In what came to be known as the Chicago School, scholars took Durkheimian conceptions of social solidarity and social disruption and created what today is known as Disorganization Theory to explain changes that were happening in the city around them. Many transitions were occurring in Chicago, as well as other Midwestern and Northeastern cities, but one of the most important of those was dramatic demographic shifts in the population as a result of migration. While earlier streams of immigrants to the United States came predominantly from Western and Northern Europe with physical features, cultural practices, and patterns of behavior that were not too dissimilar from the native population, the latter part of the nineteenth and the early twentieth centuries saw an increasing number of migrants coming from Eastern and Southern Europe. Large groups of immigrants began to populate Chicago and other American cities from countries like Italy, Greece, Poland, and Czechoslovakia. They were predominately Catholic, with languages and cultural traditions that differed dramatically from the previous immigrant waves from England, Scotland, Wales, Ireland, France, and Scandinavian countries.

During this same period, as a result of what has come to be known as the Great Migration, African-Americans moved out of the rural South to escape the oppression of Jim Crow laws and to find a better economic future into the cities of the North and West. Cities like Chicago, which needed industrial workers, were prime places for these movers to settle. Similarly, rural whites, also seeking a better economic future, moved into these cities looking for jobs.

For those early American sociologists, the changing city became their urban laboratory. They wanted to know how the changes they observed impacted social life, and a central concern was how these social forces affected crime and delinquency. Social Disorganization Theory became a critical tool for exploring and explaining the patterns that they observed. Because racial and ethnic differences were

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

central to the dynamics of this shifting urban landscape, early scholars used the important concepts of ethnic heterogeneity and ethnic succession to develop the theory used today. Both concepts will be described below, but essentially those scholars believed that when too many different groups lived in close proximity, resulting in racial and ethnic heterogeneity, the social solidarity required for an organized society could not be developed and maintained. Also, they came to believe that crime, delinquency, and other social problems were not properties of the ethnic groups that inhabited these disorganized neighborhoods, but were instead characteristics of the place. As different groups moved into that space, they would take on the levels of problems that went with the space. As the ethnic composition of neighborhoods changed, the crime rates would remain roughly the same, and there would be ethnic succession in the neighborhoods and in criminal activity.

Historical Context

Race and Ethnicity in Early Chicago School Human Ecology

Chicago School sociologists Park and Burgess ([1925]1967) provided the concentric zone hypothesis as a model of how American cities grow and adapt to change. Borrowing from the literature on plant and animal ecology, Park and Burgess's work in urban ecology was aimed at understanding both spatial relationships of social and psychological experiences and how the spatial structure of relationships within the city shaped behavior among city inhabitants. As such, this model was one of an ideal process of change rather than a static state of what cities look like (Stengel 1964).

The concentric zone hypothesis of urban growth suggested that American urban cities grew due to ecological factors, radially by concentric zones. Zone one comprises the central business district, where land is the most expensive and the most desirable particularly for commercial real estate. Zone two is the zone of transition, housing both older factories as well as cheap housing. This zone, the most undesirable area of the city, is the first point of entry for the city's newest immigrants and typically has the highest crime and poverty rates. Recent immigrants typically moved first into these areas of the city due to the availability inexpensive housing and proximity to co-ethnics and industrial employment. However, once they improved their social and economic status, immigrants, oftentimes in subsequent generations, were able to move out of this zone into more desirable areas of the city. Zones three, four, and five contained homes of blue-collar workers, middle-class residences, and upper middle-class single-family homes, respectively.

Park and Burgess ([1925]1967) argue that cities are constantly changing, growing in concentric patterns due to a process of invasion, where one population seeks to take over from another, and succession, where one population finally takes the place of another. The most striking example of this process of change is the invasion and succession of different ethnic groups into and out of city neighborhoods. Critical to the Park and Burgess model of invasion and succession of the population in urban areas was the idea that competition between inhabitants acted as the mechanism for neighborhoods to achieve equilibrium after social change, particularly the invasion of a new group into the community. Park argued that similar to the process that leads to biotic changes, cultural changes within a community occur by the same mechanism of competition. In the natural communal order of an urban area, competition for land use and economic resources determines the location of residential areas as well as the distribution of the population throughout the city. Less successful population groups will naturally inhabit the least desirable areas of the city. As new groups move in, such as new immigrant populations, competition for resources allows older immigrant groups who have become more economically stable over time to move out to more desirable areas allowing newer inhabitants to take their place.

In their classic study, The Polish Peasant in Europe and America, sociologists Thomas and Znaniecki document the transition of Polish peasants into American life, noting the struggles immigrants have in assimilating into American culture and the resulting social disorganization of the immigrant group ([1918]1996). Polish immigrants, newly arriving in American cities, brought with them the norms and morals of the old country and immersed themselves in smaller Polish communities that were unable to provide cultural and normative structures that would offer support in the face of marital conflict, disagreements between neighbors, mental health issues, and unemployment in their new country. The disorganization of the Polish peasant immigrant community is highlighted in the transition to the American way of life. New Polish immigrants struggled with economic dependency and conflict resolution with neighbors and strangers. The American economic system does not offer jobs like those the peasants were accustomed to in their home country, and therefore, immigrants found it easy to move from one job to another, unafraid of job loss, because they had developed little attachment to the positions they were able to obtain. Though the Polish community may still be intact in numbers, it now exists within the wider American community that does not recognize the function or influence of the familial or community group. Kin and kith bonds becomes broken as the church and community lose significance and power within American society, and therefore, are unable to provide support or social control when needed.

The incidence and nature of homicide illustrates the distinct nature of the social relations of Polish immigrants and the community in the American city. In the home country, stranger murders were rare since peasants had few social ties to those outside their economic and familial group, thus motives for violence rarely existed. Where social ties in American life have loosened for the Polish immigrants, contact with those outside the community is more frequent, defensive and hostile at times, and full of mistrust. The potential for a violent interaction with others can no longer be effectively subdued by kin who are now scattered, nor does the Polish community have the normative authority to intervene in the struggles of individuals.

African-Americans and Social Disorganization Theory

The African American experience in the American city, and with the process of invasion and succession, is distinct from those of other migrant groups such as the Poles discussed previously. African American migrants from Southern states inherited the most undesirable residential areas of the city. Unlike European immigrant groups, the African American experience is marked by formalized discrimination, which forced segregation of the group into the most undesirable areas of the city and restricted access to employment and education. Segregation policies and practices inhibited such essential transitions, limiting the ability of African Americans to amass savings, acquire stable jobs with good pay, and move into more desirable areas. As the “black belt” of Chicago expanded, the area became overcrowded due to the extreme density of housing from within, causing white residents and immigrants to flee to other parts of the city.

Sociologists documenting neighborhood conditions in Chicago show that African American residential areas suffered from a high rate of death and disease such as tuberculosis, mental health issues, and infant deaths (Drake and Clayton 1945). In addition, these areas were also highest in poverty and juvenile delinquency, indicative of social disorganization. Unlike African American residents, other immigrant groups who initially moved into distressed, disorganized inner-city areas were not constrained to the black belt of Chicago or other similarly disadvantaged sections of other cities.

However, these undesirable or “blighted” residential areas were invaded by warehouses and industry. The land was seen by city planners and real estate investors as more valuable for industrial establishment rather than residential areas due to its close proximity to the central business district. Therefore, investment in housing decreased because the land is more valuable for industrial establishment, while at the same time, the cost of rent was pushed up in areas already concentrated with disadvantage. Where housing investments were being made, gentrification of these areas pushes African American residents into suburban neighborhoods shared by industrial sites such as steel mills (Drake and Clayton 1945).

Ethnic Succession in Neighborhoods and Crime

Shaw and McKay's (1942) seminal work documenting the spatial patterns of male juvenile delinquency across the city of Chicago is the hallmark of Social Disorganization Theory. Shaw and McKay found that Chicago neighborhoods where male juvenile delinquency was highest were characteristically higher in poverty, residential instability, and ethnic heterogeneity. They argued that these characteristics were indicative of social disorganization which leads to some neighborhoods having higher crime rates than others. Though their work did not test this, Shaw and McKay speculated that weak social ties were the mechanism that inhibited those living in disorganized neighborhoods from combatting neighborhoods problems such as crime and delinquency.

The neighborhoods in Chicago characterized by high rates of foreign-born residents, recent immigrants, and migrants from the rural South were those with high levels of ethnic heterogeneity. Language barriers, differences in cultural and social norms, and residential instability of immigrants moving away from neighborhoods of first residence are thought to inhibit the formation of close social ties to other residents, thus decreasing the capacity for neighborhood informal social control and increasing the potential for crime.

Where Park and Burgess ([1925]1967) noted the patterns of invasion and succession for land use and industry, Shaw and McKay (1942) highlighted the patterns of ethnic and racial invasion and succession in Chicago neighborhoods and the relationship to patterns of male juvenile delinquency. Though the zone of transition experienced invasion of new immigrants and succession of older more established immigrants to more desirable areas of the city, levels of crime and delinquency remained stable over time in the same neighborhoods. This finding suggested that delinquency was tied to place rather than people or groups. Shaw and McKay argue that neighborhood crime and delinquency rates remain stable over time due to a process of cultural transmission, whereby the patterns, norms, and values that support delinquency are transmitted down through successive generations.

Shaw and McKay point towards the conflicting presence of norms and values within communities that are characterized by ethnic heterogeneity as a partial explanation for high rates of juvenile delinquency in some Chicago neighborhoods. Diversity in the cultural background of residents in the zone of transition did not foster cooperation or agreement on how to solve community problems. In addition, the constant movement of Chicago residents out from the center to more desirable neighborhoods led newer immigrants to identify with areas of residence where they planned to move rather than where they were first living. Thus, the problems that plagued their neighborhood of first residence were unlikely to be troubling to those who planned to move elsewhere.

Shaw and McKay also documented that in the African American neighborhoods of Chicago, the characteristics of social disorganization were particularly exaggerated. Like Drake and Clayton, they compared the experience of white immigrants who were able to achieve more satisfactory conditions of living and working with African Americans unable to make similar achievements due to legal restrictions on procuring residence (enforced segregation), education, and employment. As a result, African American residents of Chicago were unable to move outwardly to more desirable areas of the city, and thus, conditions of living remained in a state of concentrated disadvantage, social disorganization, and increasing crime and delinquency.

Race and Ethnicity in Contemporary Social Disorganization Theory

A central feature of Social Disorganization Theory is the argument that there is stability of crime within geographical areas. In the tradition of social disorganization, race and ethnicity are not significant as markers of cultural or behavioral proclivities for criminal offending but rather serve as indicators of pervasive social disadvantages that are affiliated with racial status. The inequality that is endemic in the educational, occupational, residential, and criminal justice contexts for people of color erects structural barriers that prevent effective social organization. European immigrants to the United States in the nineteenth and twentieth centuries faced similar restrictions, with continuing parallels to the contemporary experience of Latinos, African-Americans, and some Asian communities.

Contemporary versions of Social Disorganization Theory continue to emphasize basic principles of the theory, including informal social control and neighborhood solidarity, but scholars have added emphasis on economic disadvantage and social and cultural isolation. As was the case with their predecessors, racial and ethnic residential patterns are central to the discussion. Increased immigration and shifts in the economic and occupational market suggests that the mechanisms once understood to drive the social disorganization of neighborhoods need to consider economic and political changes and the influence on the urban environment and, consequently, the nature of neighborhood crime.

While consistent with Shaw and McKay's original formulation, Sampson and Wilson (1995) offer unique theoretical mechanisms that explore modern determinants of geographically concentrated criminal activity, emphasizing the role of neighborhood disadvantage. This work builds on Wilson's earlier research (1987) that recognizes the vital role of deindustrialization as a cause of economic disadvantage as well as pervasive residential segregation and social isolation, all causes of crime and delinquency. The decline of the urban manufacturing cities that characterized the industrial era resulted in a decline in the number of jobs available to workers with limited skills levels and low educational attainment, which disproportionately affected African-American men. The flight of middle-class whites and African-Americans from the urban core resulted in residential areas that were bereft of institutional and organizational markers of stability. Wilson suggests that these macro labor market changes contributed to the social isolation of many African-Americans who continued to live in poor urban neighborhoods, resulting in higher social disorganization and an increase in rates of violent and property crimes. Sampson and Wilson (1995) suggest that the breakdown of residential organization can be attributed to the concentrated level of economic disadvantage and the inability of residents left in the urban center to form protective relationships with neighbors. The social isolation fostered by limited residential mobility, family disruption, and high rates of poverty weakens the informal social control mechanisms that guard against the rising tide of violent and property crime.

Bursik (2000) highlights the transition from Shaw and McKay's ecological model of social disorganization to the contributions of the systemic model, which underscores the role of interactional networks that regulate behavior within neighborhoods. These networks provide the foundation for informal social control measures to control criminal behavior in the community. The breakdown of social organization is indicative of the dissolution of institutional and organizational ties to the broader community. The loss of middle-class residents, businesses, and community organizations from the urban center left the central neighborhoods with limited opportunities for public engagement and financial mobility. The exodus of residents signaled the loss of viable community structures that would aid in sustaining neighborhood cohesion. Bursik (2006) emphasizes the role of political, economic, and social network links external to the neighborhood environment in the pursuit of regulating disorderly or unacceptable behavior. The activation of these ties and their accompanying resources is predicated on the social organization of a given neighborhood. In truly disadvantaged environments, the access to these external resources is hindered by the societally disadvantaged position of its residents (Wilson 1987). Therefore, the ability to organize collectively for broader control mechanisms to combat criminal activity is limited. Predominately African-American, inner-city neighborhoods are especially affected by this process because they are more frequently characterized by hyper-disadvantage (Wilson 1987), residential segregation (Massey and Denton 1993), and as a consequence, they experience high rates of both property and violent crime (Peterson and Krivo 2010). In socially disorganized communities, social capital formation is hindered, and in communities with increased proportions of immigrant populations, the ability to effectively build and sustain these network ties is constrained by language and cultural barriers (Peterson and Krivo 2010). But recent research indicates that, contrary to the logic of early Chicago School theorists, this is not the case for all immigrant groups. Such exceptions will be discussed below. Essentially, it can be said that the social isolation experienced by many communities of color and especially some inner-city African-American neighborhoods, in tandem with increased levels of economic disadvantage, weakens the connection between neighbors, which in turn inhibits social control, allowing more crime to occur.

Sampson and Wilson (1995) highlight the substantial differences between the economic deprivations experienced by African-Americans compared to their white counterparts. They contend that an accurate comparison is not feasible given that poor whites nearly always reside in areas with less concentrated disadvantage than African-Americans, a consequence of racial residential segregation. As a result, the conditions that tear at the fabric of social organization – poverty, residential instability, and family disruption – are more pervasive and concentrated in African-American communities, making studying and replicating such circumstances in white neighborhoods an impossibility. The extent to which racism and de jure and de facto discrimination have altered the trajectories of African-Americans in the urban centers creates levels of economic, educational, occupational, and political inequity that cannot be matched in even the poorest white neighborhoods. Economically disadvantaged whites are not as isolated from nonpoor people as are their African-American counterparts. Poor whites tend to live in closer spatial proximity to socially organized middle-class communities. Therefore, disadvantaged white residents are able to benefit from the social and community resources that are most often lacking in low-income neighborhoods. The access to grocery stores, schools, community centers, and other fixtures of vibrant neighborhoods is a stark departure from the conditions in low-income African-American neighborhoods. Furthermore, these conditions, and the subsequent high crime rates that follow, remain constant over time. The toxic mixture of isolation and disadvantage ensures that the social control mechanisms will be further weakened, inhibiting the ability of African-American communities to organize effectively against crime. Consistent with Shaw and McKay's original conception, the crime conditions that exist are not racially or ethnically dependent but are rather a reflection of the structural conditions endemic to the geographical place.

New Directions

Spatial Proximity and Concentration

This social isolation is further compounded by the geographical and spatial divide between low-income communities. A major development of contemporary Social Disorganization Theory addresses the geographical concentration of disadvantage and isolation, with African-Americans often segregated in veritable islands of similarly disadvantaged others (Peterson and Krivo 2010; Sampson and Bean 2006; Sampson and Morenoff 2006). Peterson and Krivo (2010) analyze the spatial distribution of these neighborhoods, finding that low-income African-American communities are more likely to be located in a dense geographical network, bordering areas situated in similar economic straits. The spatial concentration of disadvantage heightens the levels of social isolation that residents experience, further segregating these populations from more prosocial and economically mobile segments of the population. Sampson and Morenoff (2006) suggest that this may represent a type of spatial vulnerability, where neighborhoods with low social control and collective efficacy detrimentally affect the conditions present in adjoining communities. In their analysis of Chicago neighborhoods, Sampson and Morenoff's (2006) results indicate that there was a diffusion process that spread the concentrations of high poverty rates across a densely clustered region as the national occupational landscape evolved in the late 1970s and 1980s. This geographic isolation compounds the social segregation that pervades low-income African-American communities, inhibiting the creation of social capital and the ability to establish effective informal social control. The spatial proximity of a given community in a city-wide context is affected by the influence of surrounding communities, allowing neighborhoods with strong shared value sets and personal ties to be undermined by those who do not similarly value social control but nevertheless live in close geographic proximity. These residents may be able to form strong bonds with specific fellow neighbors; however, their ability to organize effectively against crime when surrounded by socially disorganized, high-crime communities is severely limited. The permeability of adjacent geographical boundaries facilitates the spread of the detrimental effects of poverty, heightening the levels of social disorganization. Sampson and Morenoff (2006) assert that the social isolation coupled with the concentration of disadvantage promotes a vicious cycle in low-income African-American communities due to the weakness of collective efficacy. While the research focus has traditionally centered on intra-neighborhood dynamics, the inter-neighborhood research suggests that cultural and economic disadvantages are not spatially contained (Peterson and Krivo 2010; Sampson and Morenoff 2006; Morenoff 2005). Because of this diffusion, the isolation of poor African-Americans is compounded by the structural and informal barriers of the broader society. While other racial and ethnic groups face similar disadvantages, their outcomes differ from the traditional dire predictions of Social Disorganization Theory.

Contemporary Immigration and Social Disorganization Theory

Sampson and Bean (2006) reconsider Social Disorganization Theory in light of changes in immigration and mobility patterns in recent decades. Despite the original focus of the theory on ethnic heterogeneity, recent empirical work has been dominated by analyses of African-American and white neighborhoods, eschewing discussion of other racial and ethnic groups. Continuing Latino immigration has fostered a renewed interest in the dynamics that underlie ethnically heterogeneous communities and the impact on the organizational structure of neighborhoods. Extending his earlier work with Wilson (Sampson and Wilson 1995), Sampson and Bean consider the paradox that has emerged in the literature, which finds that Mexican immigrants, living in extreme economic disadvantage, experience fewer incidents of violent crime than second- and third-generation Latinos (Martinez et al. 2006; Morenoff 2005). While impoverished and historically disadvantaged African-American communities share some similar conditions with the newly immigrated Latino population, crime outcomes diverge. A number of explanations for these differences have been hypothesized, ranging from the extent of residential segregation and isolation from middle-class white communities and divergent cultural values to the severity of systemic discrimination and differential employment patterns. While African-American and Latino communities are matched on conventional predictors of violent crime, economic disadvantage, and relative social isolation, their differences on key deprivation indicators suggest immigrant status can be a protective factor. Martinez and his colleagues (2006) find that concentrated immigrant populations are not predictors of homicide at the aggregate level, which is inconsistent with the traditional pattern for African-Americans. Sampson et al. (2005) suggest that the formation of densely populated immigrant communities and the lower incidents of family disruption in Latino immigrant communities account for their lower levels of violent crime rates. While Latino immigrants face constraints on employment, residential mobility, and educational attainment, their rate of offending does not subsequently increase. Instead, researchers posit that the protective influence of their strong family and kinship ties and integration into ethnic enclaves help to ameliorate the negative effects associated with these social limitations (Peterson and Krivo 2010). The formation of ethnically homogeneous immigrant communities could similarly foster social capital among residents, producing higher levels of collective efficacy and resulting in strengthened controls against criminal activity. The formation of collective efficacy in these communities may be predicated on the shared culture, values, and language among Latino immigrants and the ability of residents to interact in spite of economic and geographic deprivation. While the original conception of Social Disorganization Theory suggested that ethnic heterogeneity constructed barriers to the formation of social capital, the evolving empirical work points to a set of unique mechanisms and processes that explain the formation of capital and collective efficacy. While African-American communities remain largely homogenous due to residential segregation, Latinos have not experienced the same levels of isolation and, therefore, often reside in more integrated neighborhoods. The research on this particular topic within the Social Disorganization Theory tradition is at a nascent stage; the continued consideration of contemporary immigrant populations holds the promise of a more nuanced understanding of the complex dynamics that underlie neighborhood crime and safety.

Conclusion

Racial and ethnic divisions have been central to the development of Social Disorganization Theory, and they remain so today. The concepts of racial and ethnic heterogeneity and ethnic succession have retained their continued empirical significance for criminologists as contributing forces to the process of social disorganization. Focusing on the process of ethnic succession diverts the onus of social disorganization from specific racial and ethnic groups to the social and structural conditions of these communities. Heterogeneity though may be more complex than the early Chicago School theorists expected. This is not to suggest that they were wrong when they hypothesized that mixing different languages, cultures, traditions, and religions together within a residential setting would inhibit the formation of social organization, but recent research suggests that a more nuanced understanding is needed. African-American communities are racially homogenous, but those marked by considerable social and economic disadvantage are more disorganized. In contrast, neighborhoods with more recent Mexican immigrants exhibit similar levels of homogeneity and disadvantage, yet local crime rates are comparatively lower. As contemporary versions of Social Disorganization Theory move forward, one of the challenges for scholars will be to evaluate and develop critical analyses of these patterns.

As is obvious, much of the literature has focused on American cities, but not all of it. For example, Sampson and Wikström (2008) used data from Stockholm to study the comparative dynamics of collective efficacy in Sweden. Other countries, for example, Australia, Canada, or South Africa, have very different racial and ethnic dynamics than the United States. That said, it is very likely that in these countries and elsewhere, if the population is marked by racial and ethnic divisions or changing immigration patterns, the influences of social organization and disorganization will have accompanying consequences for crime and criminality in these communities.

Related Entries

- ▶ Crime and the Racial Composition of Communities
- ▶ Disadvantage, Disorganization and Crime
- ▶ Early Chicago School Theory
- ▶ History of Criminological Theories: Causes of Crime
- ▶ Neighborhood Effects and Social Networks
- ▶ Poverty, Inequality, and Area Differences in Crime
- ▶ Race and Ethnicity in Social Disorganization Theory
- ▶ Social Capital and Collective Efficacy

Recommended Reading and References

Bursik RJ Jr (2000) The systemic theory of neighborhood crime rates. In: Simpson SS (ed) New advances in criminological theory. Pine Forge Press, Los Angeles

Bursik RJ Jr (2006) Rethinking the Chicago school of criminology in a new era of immigration. In: Martinez R,

Valenzuela A (eds) Immigration and crime: race, ethnicity, and violence. New York University Press, New York
Hagan J, Peterson RD (1995) Crime and inequality. Stanford University Press, Stanford
Martinez R, Valenzuela A (2006) Immigration and crime: race, ethnicity, and violence. New York University Press, New York
Massey DS, Denton NA (1993) American apartheid: segregation and the making of the underclass. Harvard University Press, Cambridge, MA
Morenoff J (2005) Racial and ethnic disparities in crime and delinquency in the United States. In: Rutter M, Tienda M (eds) Ethnicity and causal mechanisms. Cambridge University Press, New York
Park RE, Burgess EW, McKenzie RD ([1925]1967) The city. University of Chicago Press, Chicago
Peterson RD, Krivo LJ (2010) Divergent social worlds: neighborhood crime and the racial-spatial divide. Russell Sage, New York
Peterson RD, Krivo LJ, Hagan J (2006) The many colors of crime: inequalities of race, ethnicity, and crime in America. New York University Press, New York
Rutter M, Tienda M (2005) Ethnicity and causal mechanisms. Cambridge University Press, Cambridge
Sampson RJ (2011) Great American city: Chicago and the enduring neighborhood effect. The University of Chicago Press, Chicago
Sampson RJ, Bean L (2006) Cultural mechanisms and killing fields: a revised theory of community-level racial inequality. In: Peterson RD, Krivo LJ, Hagan J (eds) The many colors of crime. New York University Press, New York, pp 8–36
Sampson RJ, Morenoff J (2006) Durable inequality: spatial dynamics, social processes, and the persistence of poverty in Chicago neighborhoods. In: Bowles S, Durlauf SN, Hoff K (eds) Ethnicity and causal mechanisms. Cambridge University Press, New York
Sampson RJ, Wikström P-O (2008) The social order of violence in Chicago and Stockholm neighborhoods: a comparative inquiry. In: Shapiro I, Kalyvas S, Masoud T (eds) Order, conflict, and violence. Cambridge University Press, New York, pp 97–119
Sampson RJ, Wilson WJ (1995) A theory of race, crime, and urban inequality. In: Hagan J, Peterson RD (eds) Crime and inequality. Stanford University Press, Palo Alto
Sampson RJ, Morenoff J, Raudenbush S (2005) Social anatomy of racial and ethnic disparities in violence. Am J Public Health 95(2):224–232
Shaw CR, McKay HD (1942) Juvenile delinquency and urban areas, a study of rates of delinquents in relation to differential characteristics of local communities in American cities. The University of Chicago Press, Chicago
St Drake C, Clayton HR (1945) Black metropolis a study of Negro life in a northern city. Harcourt, Brace, New York
Stengel E (1964) Suicide and attempted suicide. Penguin, Baltimore
Thomas WI, Znaniecki F ([1918]1996) The Polish peasant in Europe and America; monograph of an immigrant group. G. Badger, Boston
Wilson WJ (1987) The truly disadvantaged: the inner city, the underclass, and public policy. University of Chicago Press, Chicago

Race and the Likelihood of Arrest

David B. Wilson[1], Tammy Rinehart Kochel[2] and Stephen D. Mastrofski[3]
[1]Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA
[2]Southern Illinois University, Carbondale, IL, USA
[3]George Mason University, Manassas, VA, USA

Overview

Opinion surveys in the United States often report that citizens believe that race influences how police officers treat the public. This entry summarizes a recent meta-analysis by Kochel et al. (2011) that examined the existing empirical evidence on the relationship between race and a police officer's decision to make an arrest. The findings support the conclusion that black suspects are more likely to be arrested than white suspects when encountering a police officer. This effect does not appear to be the result of various alternative hypotheses, such as the demeanor of the suspect or other observable legal and extralegal factors. Although the research consistently supported a racial bias hypothesis, the strength of the effect across studies did vary. Additional research is needed to better understand the factors that influence this relationship.

Background

Harvard University Professor Henry Louis Gates, Jr., was arrested for disorderly conduct in July of 2009 (Cambridge Review Committee 2010). Police arrived at his home in response to a call from a neighbor reporting a possible break-in. In fact, Professor Gates and a friend had forcibly opened a stuck door. When the police arrived, Professor Gates was already in his home. The police report indicated that in the course of questioning professor Gates, he acted inappropriately toward an officer who had requested identification from the professor and had asked him to step outside. The professor claimed that he was treated inappropriately and that it was due to his race. The officer claimed otherwise.

Professor Gates' perception that he was treated unfairly by the police is a common perception among blacks in the United States. Perceptions of racial bias on the part of the police are strongly related to race and ethnicity. A 2002 survey of the general public by Weitzer and Tuch (2005) showed that 75 % of blacks and 54 % of Hispanics believe that local police treat minorities worse than whites. This is in stark contrast to the roughly 76 % of whites who think that the police treat whites and minorities equally. Personal experiences with the police also differ substantially between racial and ethnic groups. Based on this survey, 37 % of blacks, 23 % of Hispanics, and 1 % of whites reported that they believe they were treated unfairly by the police because of their race or ethnicity. Professor Gates' perception that he was treated unfairly by the police is a common perception among blacks in the United States.

A fundamental question raised by the Gates case and the public opinion surveys is whether this perception of racially biased treatment by the police accurately reflects police behavior. Police agencies and the general public should be concerned that a high percentage of minorities believe that they are being treated unfairly by the police. This undermines police legitimacy (see [encyclopedia entry on this topic]) and reinforces racial tensions. However, an important issue is whether this perception of racially biased treatment by the police accurately reflects police behavior.

The research challenge is that in the absence of overt racist behavior, such as a racial epithet, it is difficult to know whether race played a role in an individual case, such as that of Professor Gates. To address this problem, social scientists have compared the arrest rates for white and black suspects under similar circumstances from a large number of police–citizen encounters (Although the treatment of other racial and ethnic minority groups is matter of concern, most studies focus on the comparison of whites and blacks). The credibility of such results rests on how well the research established the "under similar circumstances" comparison. The goal is to ensure that an observed difference in arrest rates between a white and black citizen is not the result of different levels of evidence, criminal behavior, or other issues that might be confounded with race. Numerous studies of this type have been conducted over the past several decades. Prior reviews of these studies, including a blue-ribbon academic panel, judged the findings to be so mixed that they have been unable to draw a definitive conclusion. The National Research Council's Committee to Review Research on Police Policy and Practices concluded that "the evidence is mixed, ranging from findings that indicate bias against racial minorities, findings of bias in favor of racial minorities, and findings of no race effect" (Skogan and Frydl 2004, pp. 122–123). Similarly, a working group of 45 social scientists formed by the American Sociological Association also was unable to draw a firm conclusion on this issue (Rosich 2007). Both of these groups relied on traditional narrative review methods and both called for additional research to better understand whether and under what circumstances police behave in a racially biased manner.

More recently, Kochel et al. (2011) published a meta-analysis of these studies. Meta-analysis is a useful tool for assessing the strength and consistency of evidence regarding a hypothesis, such as the effect of race on arrest, across multiple studies. It has also proven to be useful in areas where narrative reviews arrive at equivocal conclusions. Kochel and colleagues' meta-analysis concluded that the evidence does support the contention that the police are more likely to arrest a minority than a white suspect under similar circumstances. The methods and findings of this meta-analysis are summarized below, followed by a discussion of key issues and controversies related to this issue.

Meta-analysis of Race and Arrest

Summary of Methodology

The sample of studies included in the Kochel and colleagues' (2011) meta-analysis was comprised of observational studies of police–citizen encounters that examined the relationship between a citizen's race/ethnicity and arrest. To be included, these studies must have collected data at the level of individual suspects/encounters and measured arrest (versus a less severe alternative) and the suspect's race/ethnicity. There were several research designs that met these criteria. The two most common were field studies where researchers observed and recorded extensive data on police–citizen interactions and existing records generated by police officers, the majority of which were data routinely collected as part of police incident reports or vehicle and pedestrian stops. Also included were a survey of crime victims and a study based on data extracted from juvenile referral records.

Kochel and colleagues developed and implemented a systematic search strategy designed to identify all available studies that met the inclusion criteria for the review. This review was restricted to studies conducted within the United States, so the results cannot be generalized to other countries. To minimize publication-selection bias, both published and unpublished studies were sought and included. This process identified 40 documents that reported the results of 23 distinct research projects and 27 independent data-sets. Information was coded from these studies that reflected each study's method of data collection, sampling strategy, sample size, geographic location and years of data collection, crime type, age and sex of the samples, the data source, document type, type of statistical analyses applied, independent variables included in statistical models, and finally the statistical effect size(s) for the relationship of race on arrest.

The effect size for this meta-analysis was the odds ratio. The odds ratio is an index that compares the odds of an arrest for one group (blacks or minorities) with the odds of an arrest for another group (whites). As such, an odds ratio of 1 indicates no difference between the two groups. These odds ratios were coded so that values greater than 1 indicated bias against minorities and values less than 1 indicated bias against whites.

All of these odds ratios were based on statistical models that included additional variables that might influence police arrest discretion, such as the severity of the crime or the demeanor of the offender. As such, these odds ratios were all adjusted for observed features of the encounter, most commonly using logistic regression. A few studies relied on probit regression models or ordinary least squares models. Findings from these models were converted into odds ratios that were directly comparable to those being calculated from the logistic regression models.

Summary of Results

Kochel and colleagues coded 120 odds ratios across the 27 independent data-sets that reflected a contrast between white and black or minority suspects. When the contrast was between white and minority suspects, the minority sample was predominantly black. Thus, these results focus on the white versus black difference in likelihood of an arrest.

The multiple odds ratios from a single data-set reflected variations in the statistical models estimated on the data. This occurred both within a single publication and across publications based on a common data-set. These different models typically varied the set of independent variables included. These multiple odds ratios from a single data-set do not provide statistically independent information on the direction and magnitude of any race–arrest relationship. To address this, Kochel and colleagues restricted any given analysis to a single odds ratio per independent data-set. Thus, no analysis could include more than 27 odds ratios.

To ensure that results were not overly influenced by the specific odds ratio selected from a given data-set, Kochel and colleagues performed several analyses with different selection rules. These included calculating the average odds ratio for each data-set, selecting the smallest odds ratio for each, selecting the largest odds ratio for each, and selecting the odds ratio that met explicit criteria that they believed represent the best (most accurate) effect size for each. The selection criteria for the best effect size within a data-set gave preference to odds ratios that: (1) were based on a logistic regression model rather than an OLS or probit regression model; (2) had a standard error that was reported directly and not imputed from other information, such as sample size; (3) were based on the full sample, rather than a subset; (4) were from models where the operationalization of the dependent variable was clearly arrests versus no arrest, rather than some other less severe alternative sanction; (5) had a race variable that represented black versus white, rather than minority versus nonminority or white versus nonwhite; (6) were based on a statistical model that included demeanor as an independent variable, (7) were based on a statistical model that included officer characteristics, and (8) were based on a model with the largest number of independent (control) variables.

Table 1 presents the meta-analytic mean for these four sets of odds ratios. The meta-analytic method used was the random-effects inverse-variance weight approach. This is a conservative approach to meta-analysis and explicitly assumes meaningful variation across studies in the estimate of the race–arrest relationship.

Each of the four methods of selecting one odds ratio per data-set produced an overall mean effect size that was statistically significant. The range in the mean odds ratio across these four sets was 1.32 for the smallest odds ratio per data-set and 1.52 for the largest odds ratio per data-set. Using the average and best odds ratio within each data-set produced very similar results (1.45 and 1.38, respectively). These results suggest that minorities have a higher odds of arrest than nonminorities in a police–citizen encounter. The different selection models showed that the results were robust to which effect size is selected from each data-set.

Race and the Likelihood of Arrest, Table 1 Meta-analysis results of odds ratios for the effects of race on arrest (27 odds ratios per mean)

		95 % C.I.	
Analysis	Mean odds ratio	Lower	Upper
Average	1.45	1.28	1.64
Smallest	1.32	1.16	1.50
Largest	1.52	1.32	1.74
Best	1.38	1.24	1.53

Data from Kochel and colleagues (2011). Mean odds ratios based on a random-effects inverse-variance model estimated via method-of-moments

Figure 1 presents a forest-plot for the distribution of odds ratios for the set of effect sizes designated as the best from each data-set. Figure 1 shows that only 4 of the 27 odds ratios were in the negative direction with blacks or minorities having a lower odds of arrest. This would be expected by chance. What is of greater interest is the pattern of results across studies and this pattern supports the hypotheses that race affects the likelihood of arrest.

By conventional standards, these effects are small. To interpret the magnitude of the effect in meaningful terms, Kochel and colleagues converted the overall mean odds ratio of 1.38 (model based on the best effect size within each data-set) into the probability of arrest for minorities and blacks relative to whites. This conversion is based on an assumed arrest rate of 20 % for whites, roughly the average across all samples. Using this value as the benchmark, an odds ratio of 1.38 is equivalent to an arrest probability of 0.26 for minorities and blacks and 0.20 for whites. This is a difference that is large enough to be of practical concern.

Focusing solely on the mean ignores the variability in results across studies. This variability exceeded what would be expected from sampling error alone. Thus, there are true differences in the effect of race on the likelihood of arrest across these studies. Some studies show a stronger race effect than others. Kochel and colleagues performed a series of moderator analyses to explore this variability. These analyses examined both methodological and theoretical explanations for differences across studies.

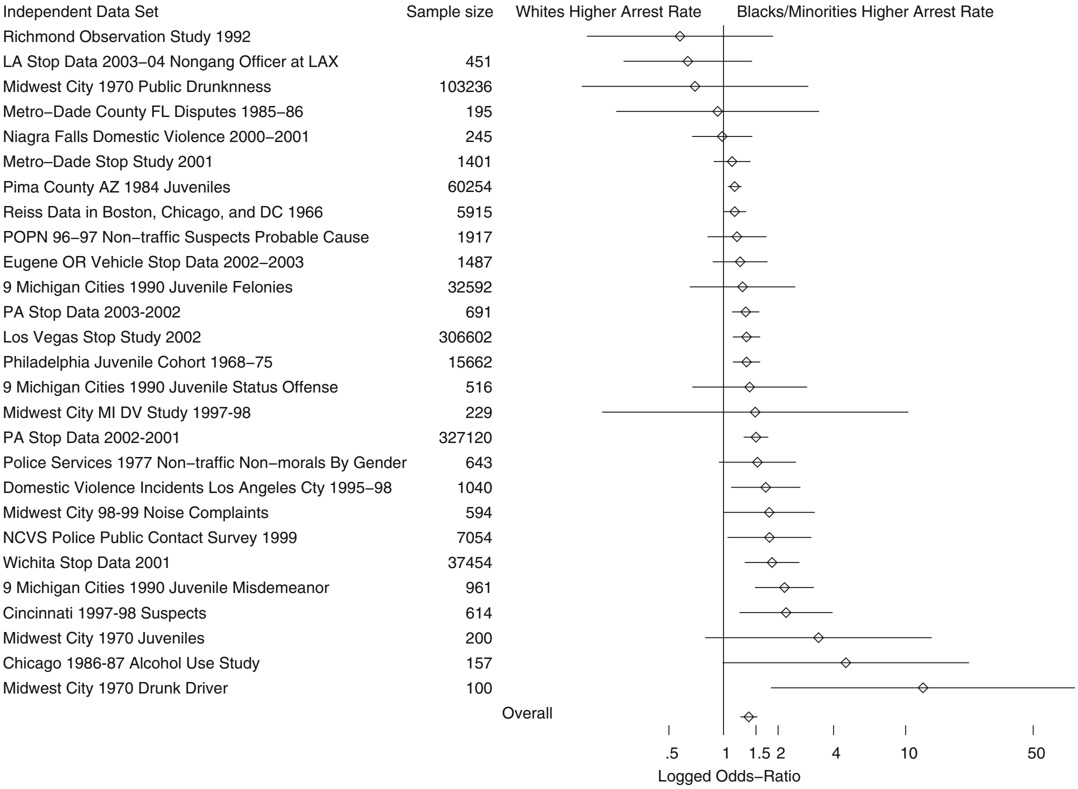

Race and the Likelihood of Arrest, Fig. 1 Forest-plot of the best effect size within each data-set (From Kochel and colleagues 2011; Reproduced with Permission from Criminology)

An important methodological consideration is how the data were collected. Three basic approaches were used across the studies reviewed. The method least susceptible to bias is the use of trained field researchers accompanying police out in the community and coding extensive information about each police–citizen encounter. However, an officer may be affected by the observer's presence, altering what they do. There is some evidence to suggest that officers are less passive and more legalistic in the presence of an observer (Mastrofski and Parks 1990; Mastrofski et al. 2010; Spano 2007). This may reduce the likelihood of finding a race–arrest relationship. This was the method used for 10 of the 27 data-set. Police officer's documentation of encounters with citizens was the data source for another 16 data-sets. This method is arguably likely to underestimate any racial bias, as officers would want to present their behavior in a positive light, given strong disincentives in most police agencies toward overt racially biased behavior. A single data-set was based on surveys of victims. As reported by Kochel and colleagues, although the victim survey produced a larger odds ratio (1.79), the officer reported and field researcher data produced highly similar means (1.39 and 1.36, respectively). The differences across these means were not statistically significant and the overall finding of a positive race–arrest relationship is clearly not being driven by a single method.

These studies also differed on the statistical methods used. Once again focusing on the best estimate per data-set, most models were estimated using logistic regression methods (23 of 27), two used probit methods, and two used ordinary least squares methods. The mean for the probit models was highest and the mean for the OLS models was lowest, although not by enough to account for a significant amount of variability in the odds ratios.

Theoretically, several variables may moderate the race–arrest relationship. In considering these variables, scholars have distinguished between disparity and discrimination. The logic is that disparate experiences by race when explained away by legal factors present during the encounter are not discriminatory (Skogan and Frydl 2004, p. 124). Therefore, differential rates of arrest for minorities and whites that dissipate when accounting for legally relevant factors indicate a disparity but not discrimination. Legally relevant factors would include the seriousness of the offense, amount of evidence against the suspect, the presence of a victim supportive of arrest, the presence of a witness on the scene, the suspect being under the influence of drugs or alcohol, or the discovery of additional criminal acts during the course of the officer–citizen interaction.

Kochel and colleagues examined the disparity versus discrimination issue by comparing odds ratios based on statistical models that either did or did not include each of the above legally relevant factors. The differences were trivial, statistically nonsignificant, and failed to explain the positive race–arrest relationship. That is, adjusting for legally relevant disparity did not attenuate the overall finding of discrimination against blacks.

It is possible that individually, these legal factors contribute little to explaining the race–arrest relationship but collectively they might. Kochel and colleagues examined this by assessing the relationship between the number of these factors included in a model and the size of the race–arrest effect using a meta-analytic regression model. The analysis showed that as more legally relevant factors were controlled, the odds ratio became smaller. However, this effect was very small and statistically nonsignificant. The race–arrest relationship remained even for those models with the highest level of statistical control for legally relevant factors.

The suspect's demeanor was an important extralegal factor that has been examined by roughly half of the studies included in the Kochel and colleagues review. Research supports the idea that a suspect's demeanor affects the arrest decision. It is important to note that some disrespectful behavior is illegal, such as physical resistance, and as such is legally relevant to an arrest decision. Other disrespectful behavior is generally not, such as using disrespectful language, and should not be considered by an officer in the arrest decision. What may appear as racial discrimination may be more appropriately attributed to suspect demeanor if demeanor and race are confounded. That is, an officer making a decision to arrest may not be reacting to a suspect's race but rather to his or her disrespectful behavior. This hypothesis is not supported by the existing data. Statistical models that adjusted for demeanor produced effect sizes roughly comparable to models that did not. Thus, controlling for the demeanor of the suspect did not attenuate the race–arrest relationship.

Key Issues and Controversies

It remains possible that unaccounted for legal and extralegal aspects of the police–citizen encounter could explain the race–arrest relationship, reducing the odds ratio to zero. This is a challenge with all observational research. It is impossible to know if all possible confounds have been accounted for. However, the most likely culprits have been addressed in the literature on this topic, thus increasing the confidence that can be placed in the conclusion that race plays a role, possibly unintentionally, in the arrest decision.

An important hypothesis in this literature has been that the observed race effect is a function of differential demeanor between white and black suspects when encountering a police officer. Research does indicate that police are affected by the demeanor of a suspect in their arrest decision. If on average, the demeanor of black suspects is more hostile than white suspects, then this would produce a differential arrest rate between the two groups. As discussed above, the evidence reviewed by Kochel and colleagues does not support this hypothesis. Other plausible confounds, such of severity of the offense, also failed to explain away the race–arrest relationship.

The robustness of the finding of racial bias to plausible alternative explanations should not obscure the variability in findings across the studies included in the Kochel and colleagues' meta-analysis. Clearly this effect varies.

The meta-analysis did not identify any explanations for this variability but the authors did theorize about possible ecological factors that might be explored in future studies (Klinger 2004). Four possibilities were discussed. The first has to do with the character of the organization, such as how actively the organization attended to unequal treatment of citizens according to their race or the degree of professionalism and bureaucracy of the police agency. Second is the context of who benefits or is served by an arrest. A police officer's decision to make an arrest may be affected as much by who benefits (e.g., the race or other characteristics of a victim) as who suffers. Third, the neighborhood context may affect an officer's discretionary judgment regarding an arrest. Police may exert their authority in more racially differentiated ways in some neighborhoods than others. And finally, differences in the race–arrest relationship may reflect the level of political power of a given minority group within a police jurisdiction. Little is known about the role of these ecological factors on police decision making.

Conclusion

In summary, the existing evidence supports the conclusion that blacks have a higher likelihood of arrest than do whites. The research does not demonstrate the causes of this racial disparity, nor does it point to a clear policy response for dealing with it. What it does establish is that where there is smoke there is indeed fire regarding racial disparity in the arrest practices of American police. This certainly shows that further efforts to delineate the legal and ethical implications of racially differentiated policing are based on a solid empirical foundation. Future empirical research should move beyond just testing for a race effect and examine what accounts for variation in this relationship and the effect of any policies designed to reduce it.

Related Entries

▶ Biased Policing
▶ Police Discretion and Its Control
▶ Police Discretion in Law Enforcement
▶ Police Discretion in Providing Services and Assistance

Recommended Reading and References

Cambridge Review Committee (2010) Missed opportunities, shared responsibilities: final report of the Cambridge Review Committee. http://www.cambridgema.gov/

Klinger DA (2004) Environment and organization: reviving a perspective on the police. Ann Am Acad Polit Soc Sci 593:119–136

Kochel TR, Wilson DB, Mastrofski SD (2011) Effect of suspect race on officers' arrest decisions. Criminology 49:473–512

Mastrofski SD, Parks RB (1990) Improving observational studies of police. Criminology 28:475–496

Mastrofski SD, Parks RB, McCluskey JD (2010) Systematic social observation in criminology. In: Piquero A, Weisburd D (eds) Handbook of quantitative criminology. Springer, New York

Rosich KJ (2007) Race, ethnicity, and the criminal justice system. American Sociological Association, Washington, DC. http://asanet.org

Skogan W, Kathleen F (eds) (2004) Fairness and effectiveness in policing: the evidence. National Research Council, Washington, DC

Spano R (2007) How does reactivity affect police behavior? Describing and quantifying the impact of reactivity as behavioral change in a large-scale observational study of police. J Crim Justice 35:453–465

Weitzer R, Tuch SA (2005) Determinants of public satisfaction with the police. Police Quart 8:279–297

Race Issues in Probation and Parole

Kathryn D. Morgan
Department of Justice Sciences, The University of Alabama at Birmingham, Birmingham, AL, USA

Overview

This entry examines the relationship between race and probation and parole decisions and the issue of racial disparity in granting probation and parole. Previous research suggests that race is an important variable at every level of processing where criminal justice authorities have the power to make discretionary decisions. This entry focuses on two levels of discretionary decision-making in the criminal justice process and how race affects decisions made at those levels. Probation is an alternative to incarceration that allows the individual to remain in the community maintaining family and community ties, and employment while avoiding the stigma of prison. Parole is granted to an individual who has served a portion of the sentence and is released back into the community to serve the remainder of the sentence under supervision. Statistics consistently show that African-American and Hispanic defendants are less likely to be granted probation and parole and more likely to be sent to prison and stay longer than Whites. While discussing the statistics, this entry covers possible reasons for this sentencing disparity. Is the disparity based on race or legal variables such as seriousness of the offense and criminal history? Finally, the entry will conclude with suggestions for future research.

Probation and Parole as Corrections Programs

Fundamentals of Probation

There has been a proliferation of programs designed to be correctional rehabilitation programs. Despite the attacks on rehabilitation and the shift away from the rehabilitative ideal in the 1970s, community-based corrections have continued to proliferate in the United States correctional system. Probation and parole are two of those programs. Many of those in correctional programs in the community are on probation and parole, causing probation and parole populations continue to increase annually. In 2008, approximately 5.1 million offenders were supervised in community corrections programs. Of that number, probationers accounted for 84 % or 4.2 million and parolees accounted for over 800,000 or 16 %. By the end of 2009, the numbers of adults supervised under probation and parole continued to increase, making it clear that probation and parole are important programs in the criminal justice system (Glaze and Bonczar 2010). Statistics for adjudicated juveniles reveal a similar pattern. Probation remains the most common disposition for juveniles who have been adjudicated as delinquent as indicated by 2007 data, showing that 56 % of these delinquent juveniles received probation.

Although probation and parole are used synonymously and viewed as the same, they are very different programs. Probation is an alternative to incarceration where the sentencing judge decides to suspend the imposition of the sentence and place the defendant on probation. The judge and the probationer sign a probation order which constitutes a contractual agreement between the court and the probationer. The probationer must obey all probation conditions to remain in the community under conditional release. The probationer is supervised under the jurisdiction of the court; therefore, all requests for amending the probation order must be approved by the court from which the case originated. If violations occur, the case may be referred back to this same court for revocation proceedings.

While parole also represents conditional release, it is only granted after an offender has served at least one-third of the sentence in prison. After serving the required period of time, the Parole Board may consider the inmate for parole release. If the Parole Board believes that the inmate is ready to return to society, he or she is released to parole supervision with rules and regulations known as conditions of parole. Unlike probation, parole is an administrative matter where the Board of Pardons and Parole has jurisdiction over the inmate. Changes to parole conditions and violations of those conditions are handled by the Parole Board.

Probation is granted by the presiding judge after a review of the presentence investigation report and a consideration of other factors including the severity of the offense; the seriousness of the prior criminal history, sentencing guidelines, the defendant's age, rehabilitation potential and attitude, and the relationship with his or her family. There has also been controversy regarding the impact of race/ethnicity, gender, and social class on the judge's sentencing decision. Research findings have been inconclusive about the extent to which these variables influence judges' decision to grant probation or incarcerate. Some studies have concluded that the strongest predictors of sentencing are severity of the offense and criminal history (Kleck 1981; Spohn and Welch 1987). Other sentencing experts have argued that these variables do affect judges' decisions in deciding to grant probation or incarcerate. Petersilia and Turner (1998) state that with the exception of drug cases, the type of attorney influences probation granting decisions. Those with a private versus public attorney are more likely to get probation instead of prison. Spohn and Welch (1987) found that in borderline cases where defendants may be sentenced to probation or incarcerated, White defendants will usually be granted probation and Black defendants will be incarcerated.

Fundamentals of Parole

Approximately 15 % of those going to prison serve their entire sentence. Over 75 % of all inmates are released through either *discretionary or mandatory release*. *Mandatory Release* which grew out of the determinate sentencing approach, occurs when the inmate has served the entire sentence minus the good time. When the inmate is sentenced to prison, a "mandatory release date" which includes good time and other credits is calculated during the classification period. When the mandatory release date arrives, the inmate must be released and cannot be held longer than the actual sentence. Mandatory release is presently used by the federal government and those states operating under a determinate sentencing structure. Recent statistics indicate that mandatory releases have surpassed discretionary release by a parole board. In early 2000, forty-one percent (41 %) of the inmates released were granted a mandatory release. *Discretionary Release* occurs when an inmate is released as a result of decisions made by parole boards or commissions. Until the mid-1970s, all states and the federal government had systems of discretionary release where the parole board determined how much the time the person spent in prison and the time of return to the community. Discretionary release is associated with the indeterminate sentencing structure where the judge sets a minimum and maximum term of incarceration and the parole board determines the release time within these limits. Discretionary parole grew out of the rehabilitation model, releasing a person at a time when he or she would benefit most from being released. Determinate sentencing abolished parole in about 20 states and the federal government.

Race Issues in Probation and Parole

One of the most important controversies in criminal justice and criminology today focuses on whether or not there is disparate treatment of racial and ethnic minorities in the criminal justice system. Blacks and Hispanics are disproportionally represented at every phase of the Criminal Justice System. One area where the disproportionality has been most acute has been in incarceration rates. According to 2009 incarceration data, over 1.6 million prisoners were incarcerated in state and federal prisons, and Black and Hispanic males had highest incarceration rates. Blacks represented almost one million of that total prison population. Black males, who comprised over 40 % of the prison population, were incarcerated at a rate of 4,749 per 100,000, six times the incarceration rate for White males. Hispanic males were incarcerated at a rate of 1,822 per 100,000 Hispanic men. White males were incarcerated at a rate of 708 per 100,000 White men. Do these numbers reflect the existence of racial bias by criminal justice authorities charged with making discretionary decisions? Several criminologists have suggested that there is racial disparity at every level of processing where criminal justice authorities have the power to make discretionary decisions. Police are more likely to arrest, verbally and physically abuse, harass, and profile racial and ethnic minorities. Criminal courts have been accused of promoting not only a racial bias but an economic bias that results in racial and ethnic minorities being denied or unable to afford bail, using court-appointed attorneys, and ultimately receiving more severe sentences. Some criminologists conclude that racial bias by decision-makers does not result in more severe sentencing outcomes for racial and ethnic minorities. Black and Hispanic defendants commit more serious offenses and have more severe criminal histories. Others suggest that even when racial disparities exist, they are not the result of systematic racial bias but the unintended consequences of well-intended crime control policies designed to reduce crime and protect society such as the War on crime and drugs. These "get tough" policies resulted in the unintended consequences of higher arrests, prosecutions, and incarcerations of Blacks who were disproportionately affected (Kleck 1981; Wilbanks 1987; Tonry 1994).

Race and Probation

At the end of 2009, there were over four million adults under probation supervision in the United States. Whites comprised 55 % of the probation population while 30 % of the population was Black defendants and 13 % were Hispanic. In some states such as Texas and Oklahoma, the percentage of Whites on probation increases to as high as 69 % and the percentage of Blacks on probation falls to 20 %. It is commonly believed that probation and presentence reports may favor White over minority defendants, causing judges to award probation to Whites more than minorities. Additionally, Whites are more likely to receive probation in jurisdictions where Blacks and Whites receive prison sentences of similar duration. Petersilia (1985) found that in California, 71 % of the Whites convicted of a felony are placed on probation compared to 67 % of Blacks and 65 % of the Hispanics convicted of a felony. In their study of sentencing in Michigan Courts, Spohn and Welch (1987) found that in "borderline" cases, in which judges could impose long probation sentences or short prison terms, Whites were more likely to get probation and African-Americans were more likely to get prison.

Drug policies associated with the "War on Drugs" dramatically increased the number of incarcerated Blacks and Hispanics. Longer sentences for drug users, mandatory sentences for possession of crack cocaine, and reclassifying misdemeanor drug possession of crack cocaine to felony possession resulted in more Blacks being arrested, prosecuted, convicted, and incarcerated for drug offenses. Approximately two-thirds of the drug users are Whites, but Blacks comprise 62 % of those incarcerated for drug offenses. Whites make up only 36 % of those in prison for drug charges (Walker et al. 2004). Blacks are often sent to prison for drug offense and serve as much time in prison for a drug offense as Whites serve for a violent offense. Clear and Dammer (2000) suggest that there is a bias that operates in the selection process for discretionary community programs such as probation that results in Blacks excluded. Black offenders are likely to be young, poor, uneducated, unskilled, and have a more serious criminal history than their White counterparts. Therefore, they are viewed as less suitable for these programs.

Wilbanks (1987) suggests that offense severity and severe criminal histories cause Black defendants to have higher high incarceration rates and lower probation sentencing rates. Black defendants, according to Wilbanks, commit more severe crimes and have more severe criminal histories.

The pattern persists with juvenile defendants; White juveniles may be granted probation while Black juveniles are sentenced more severely. A study of juvenile probation presentence reports reveals that the delinquency of Black juveniles is viewed as the result of negative attitudes and personality disorders, and the behavior of White juveniles is explained as the product of the school and family problems. As a result, White delinquents are sentenced to probation because their behavior problems can be addressed through community treatment and intervention programs. Since personality disorders are less responsive to such interventions, Black juveniles are incarcerated in order to protect the community (Mauer 2006).

The fact that fewer Blacks and Hispanics are sentenced to probation is not only a race issue, but a class issue. More Blacks and Hispanics remain in jail pending trial because they cannot afford monetary bail. Many of these same defendants use court-appointed attorneys and are tried by juries of Whites because Blacks from lower class neighborhoods are less likely to report for jury duty.

Race and Parole

At the end of 2009, there were 800,000 adults under parole supervision in the United States. Forty-one percent of the parole population was White. Blacks comprised 39 % of the parole population, and 18 % of those on parole were Hispanics. Although Black inmates make up approximately 62 % of the prison population, only 39 % of those released on parole are Black.

Findings of previous research reveal that race has an impact on parole release decisions. In a study of racial discrimination in three states, Petersilia (1985) found that Black and Hispanic defendants not only received more severe sentences than White defendants who had similar criminal records and were convicted of similar crimes but consistently served longer sentences than Whites sentenced to prison. Black inmates serve longer sentences, are less likely to receive parole, and often have additional criteria to satisfy. Carroll and Mondrick (1976) discovered racial inequities in parole decisions by comparing those decisions for Black and White inmates. They found that Black inmates not only served longer sentences before being paroled but had to show proof of participation in and completion of some institutional treatment program. This criterion was not required for White inmates being considered for parole. Results of another study found that when first time offenders are being considered for discretionary release, Black first time inmates usually serve 4 months longer than Whites (62 months vs. 58 months) (Hughes et al. 2001).

Some studies suggest that race does not significantly impact parole release decisions. Instead, legal, social, and institutional variables are the most significant predictors of parole release decisions. In a study of 958 youthful offenders, Elion and Megargee (1978) found no differences in the actual amount of time served, but a significantly higher percentage of Whites than Blacks were granted parole. Blacks who were granted parole served a smaller portion of their sentences than did Whites who were granted parole. Blacks who were denied parole served a greater portion of their sentences that Whites who were denied parole. They suggest that these differences are attributed to social, legal, and institutional variables and not race. Scott (1974) initially concluded that Black inmates serve more time and are less likely to be paroled than White inmates. However, when he controlled for the seriousness of the offense, his findings showed that Whites were less likely to be paroled than Blacks, causing him to ultimately infer that parole boards make no distinction between races when considering criteria for parole. Black and White inmates are paroled at about the same rate. In a study of the parole release process in Alabama, Morgan and Smith (2008) found that while race was not a significant predictor of parole release decisions, Black inmates served more time before being considered for parole release. While there were twice as many Black inmates as White inmates eligible for parole release and selected for a parole release hearing, fewer Black inmates were actually chosen for parole release by the parole board.

Conclusions and Future Research

Probation is a correctional alternative that allows convicted defendants to remain in the community under supervision; it is a privilege granted to defendants. Parole is granted to convicted defendants who have been sentenced to prison and served at least one-third of the original sentence. Both practices are designed to promote rehabilitation. When looking at the racial compositions for offenders supervised in these two programs, there is some concern for racial bias by judges and parole board members who are responsible for decision-making at these two levels. While Blacks make up only 13 % of the US population, Black inmates accounted for almost one million of inmates incarcerated in the nation's prisons. At the same time, they make up 30 % and 39 % of the probation and parole populations. While the question of racial bias is real, other explanations have been offered to explain this racial disparity. Some have argued that incarceration numbers are high because Blacks commit more crimes than Whites. The trend is that the Black arrest rate is disproportionate to the 13 % that they represent in the population. Recent review of arrest rates for index crime reveals that Blacks constituted 40 % of all arrests for violent crime and 30 % of the arrests for property crimes. With the exception of murder and robbery, the actual percentages of crimes committed are higher for Whites. It is also possible that police who use discretion in decisions about arrests could also exercise a racial bias. Other arguments focus on crime severity. According to these arguments, Black offenders commit more violent crimes that usually granted probation. Proponents of this argument point to the fact that UCR data indicate that Blacks comprise 40 % of all of the arrest for violent crimes and make up 50 % and 57 % of the arrest for murder and robbery. It has also been stated that Blacks have more severe criminal histories that make them poor risks for probation and more disciplinary infractions in prison that reduce the likelihood of being selected for probation and parole supervision The War on Drugs disproportionately affected Blacks by increasing incarceration rates for Blacks. In an effort to reduce drug supply, tougher sentencing policies were implemented. One such policy was a mandatory prison sentence for possession of crack cocaine, found most often in the Black community. Even though the numbers of Blacks using drugs fall far below that of Whites, Blacks are most often arrested, prosecuted, convicted, and incarcerated for nonviolent drug offenses.

The fact that Blacks and Hispanics are less likely to be granted probation and parole has implications for individuals, families, and communities. When defendants are granted probation, they remain in the community and are able to keep the jobs that they have or with the assistance of the supervising probation officer, they are able to secure employment to comply with probation conditions. When individuals are incarcerated, it takes them out of the labor force. Once released, they are viewed in a negative manner by employers; most employers are unwilling to hire ex-offenders. Failure to get and maintain suitable employment often results in individuals committing more crimes and being returned to prison. The economic impact affects not only the offender but also dependents. Incarceration affects family stability, the socialization of the children, and the psychological well-being of children. Children feel humiliation, guilt, and shame when a parent is incarcerated, often believing that they contributed to the parent's incarceration. Children may be labeled and stigmatized. Some recommendations to address this disparity include reserving prison for serious or violent offenders, expanding the use of intermediate sanctions especially for nonviolent minority offenders. There should also be a racial impact statement of every change in justice policy. These statements would prospectively assess the implications the policy shift would have on minority involvement in the justice system so that the policy can be evaluated on the basis of its consequences. Finally, it is evident that more empirical investigation needs to focus on decisions made at crucial decision points in the criminal justice system.

Related Entries

▶ Race and the Likelihood of Arrest

Recommended Reading and References

Carroll L, Mondrick M (1976) Racial bias in the decision to grant parole. Law Soc Rev 11:93–107

Clear T, Cole G (2000) American corrections. Wadsworth Publishing, Belmont

Clear T, Dammer H (2000) The offender in the community. Wadsworth Publishing, Belmont

Elion V, Megargee E (1978) Racial identity, length of incarceration, and parole decision making. J Res Crime Delinq 16:233–245

Glaze L, Bonczar T (2010) Probation and parole in the United States, 2009. U. S. Department of Justice, Bureau of Justice Statistics, Washington, DC

Hughes T, Wilson D, Beck A (2001) Trends in state parole: 1990–2000. Bureau of Justice Statistics, Washington, DC

Kleck G (1981) Racial discrimination in criminal sentencing: a critical evaluation of the evidence with additional data on the death penalty. Am Sociol Rev 46:783–805

Mauer M (2006) Race to incarcerate. The New Press, New York

Morgan K (1995) Factors influencing probation outcome. J Crim Just 22(4):341–354

Morgan K, Smith B (2005) Parole decisions revisited. J Crim Just 33(3):277–287, May–June
Morgan K, Smith B (2008) Impact of race on parole release decisions. Just Q 25(2):411–435
Petersilia P (1985a) Granting felons probation: public risks and alternatives. The Rand Corporation, Santa Monica
Petersilia J (1985b) Racial disparities in the criminal justice system: a summary. Crime Delinq 31:15–34
Petersilia J (1987) The influence of criminal justice research. RAND Corporation, Santa Monica
Petersilia J, Turner S (1998) Probation in the United States. Perspect, Am Probat Parole Assoc 30–41
Scott J (1974) The use of discretion in determining the severity of punishment for incarcerated offenders. J Crim Law Criminol 65:214–224
Spohn C, Welch S (1987) The effect of prior record in sentencing research: an examination of the assumption that any measure is adequate. Just Q 4:287–302
Tonry M (1994) Racial politics, racial disparities and the war on crime. Crime Delinq 40:475–494
Walker S, Spohn C, DeLeon M (2004) The color of justice, 2nd edn. Wadsworth, Belmont
Wilbanks W (1987) The myth of a racist criminal justice system. Brooks/Cole, Monterrey

Race, Ethnicity, and Youth Gangs

Adrienne Freng
Department of Criminal Justice, University of Wyoming, Laramie, WY, USA

Overview

It is difficult to talk about gang membership without including a discussion of race/ethnicity. The two seem invariably linked in our society, both in terms of research and our common perceptions of gang members (Esbensen and Tusinski 2007). Early gang research focused on White ethnic groups, whereas more recently, racial/ethnic minorities such as African Americans, Hispanics, and increasingly Asians and American Indians represent the groups of interest (e.g., Chin 1996; Hagedorn 1988; Major et al. 2004; Moore 1991; Vigil 2002). While the connection between race/ethnicity and gang membership remains in the spotlight in our society, the fact remains that a paucity of information regarding differences between racial/ethnic groups exists, making policy recommendations somewhat complicated.

This entry will explore the similarities and differences in gang membership between various racial/ethnic groups. As we come to understand the race/ethnicity/gang membership relationship more clearly, the intent is to develop ideas on how to become more effective at preventing gang membership among the various racial/ethnic groups. The entry begins by presenting descriptive information concerning race/ethnicity and gangs. Next, the evidence regarding shared factors for gang formation and membership between and within racial/ethnic groups is examined. Third, it explores the evidence supportive of risk factors that are "race/ethnicity specific" or unique to one or a few groups. Fourth, recommendations to inform programming regarding the connections between race/ethnicity and gang membership are provided.

Race/Ethnicity and Gangs

In our society, gang membership is often portrayed, especially by the media, as a minority issue affecting the barrios and inner cities of the United States (Esbensen and Tusinski 2007). Minority youth do, in fact, represent the majority of gang members. The National Gang Center (2010) reports that law enforcement indicates that 84 % of gang members are racial/ethnic minorities, with 49 % being Hispanic and 35 % being African American. However, increasing numbers of Whites are reporting gang membership as well (Esbensen et al. 2010; Freng and Winfree 2004; National Gang Center 2010). For example, self-report studies indicate more extensive involvement of Whites, as much as 24 % of gang members, as compared to 9 % reported by law enforcement (Esbensen et al. 2010; National Gang Center 2010).

Another recent development includes gangs becoming more multiracial (Starbuck et al. 2001). In a survey of law enforcement, agencies reported that one-third of gangs were ethnically/racially mixed (Starbuck et al. 2001). This percentage was higher in those jurisdictions that also reported a later onset of gangs. So, for example, agencies that reported gang onset as 1991–1992 as compared to 1981–1985 indicated that 55 % of gangs consisted of racial-/ethnic-mixed membership (Howell et al. 2002). As gang membership becomes more diverse in terms of race/ethnicity, it begins to call into question some of our basic assumptions regarding race/ethnicity and gang membership. So, no longer can we exclusively talk about African American, Hispanic, or Asian gangs, as more and more gangs consist of individuals from a number of racial/ethnic groups.

Sources of Information on Gang Membership

Our information on the race/ethnicity/gang membership relationship has traditionally come from three primary sources: official (law enforcement) data, self-report data, and ethnographies. Each of these data sources provides valuable information regarding gang membership but also offers different pictures concerning the involvement of racial/ethnic minorities in gangs. For example, both official and self-report data indicate that gangs are comprised primarily of racial and ethnic minorities. However, discrepancies between these data sources regarding the extent of involvement of racial/ethnic minorities arise (National Gang Center 2010). Some of this disparity could be due to the fact that the racial/ethnic makeup of a gang tends to reflect the racial/ethnic composition of the community. For example, while the National Gang Center reports that overall 9 % of gang members are White, this percentage increases to 17 % in rural counties and 14 % in smaller cities where populations as a whole tend to consist of larger percentages of Whites (National Gang Center 2010). Likewise, African American and Hispanic gang members are the most prevalent in larger urban areas (National Gang Center 2010). Thus, since the majority of data on gangs is generated from large cities, these populations are often the ones represented, giving the impression that gang membership is solely a minority issue.

The disparities in the estimates between law enforcement data and self-reports are also often tied back to the nature of each of these data sources and the various ways to define gang membership. Although defining a gang may seem straightforward, many definitions abound resulting in significant consequences in terms of informing policy and decisions about resource allocation as well as impacting a community's level of fear of crime, especially as it relates to race/ethnicity (Ball and Curry 1995; Decker and Kempf-Leonard 1991; Huff 1998; Klein 1969; Miller 1975). Many estimates and policies are based on data from law enforcement agencies; however, some gang experts believe these estimates are "probably conservative because many jurisdictions deny, often for political and image reasons, that there is a problem, especially in the early stages of youth gang development in a community" (Huff 1998, p. 1). Others, however, suggest that policies and practices of law enforcement such as conducting intelligence programs that result in lists of known gang members based on overt characteristics such as tattoos, wearing of colors, and clothing can significantly overestimate the problem. These practices can be problematic and impact data if they are not consistently purged as, contrary to popular belief, most gang members are only in the gang for about 1 year (Thornberry et al. 2003). On the other hand, although research has supported the validity and reliability of self-reports, many argue that they represent an under- or overestimation of the problem as well (Loeber et al. 1998). In other words, individuals may claim gang membership when they really might be considered peripheral to the gang, while others might not claim gang membership in an effort to keep that information from law enforcement and researchers.

Adding to the confusion are results collected by ethnographic studies (Fleisher 1998; Hagedorn 1988; Moore 1991; Vigil 2002). These typically concentrate on one particular gang comprised of a certain racial/ethnic group in a specific site. Most of these studies have been completed in large urban areas and with minority gangs, further contributing to the perception that gangs consist exclusively of racial/ethnic minorities and resulting in a tendency to separate gangs into groups such as African American gangs, Hispanic gangs, Asian gangs, and American Indian gangs. By doing this, the intricacies within groups, such as the differences between Chinese and Vietnamese gangs, as well as the multiracial nature of emerging gangs are ignored (Starbuck et al. 2001). Additionally, although ethnographies have contributed significantly to our understanding of the race/ethnicity/gang connection, they restrict the ability to generalize the findings to other groups such as Whites or to other locations such as rural and suburban areas, thus limiting our overall knowledge of how race/ethnicity impacts gang membership.

Reasons for Joining Gangs

The delinquency and gang literature is replete with various explanations as to why individuals from various racial/ethnic groups join gangs. Within this literature, the reasons for joining gangs tend to fall within two main frameworks. The first framework proposes that different etiologies or distinctive causal processes are responsible for racial/ethnic disparities in gang joining. In other words, differences in the types of factors that result in gang membership vary by group. On the other hand, the second framework suggests that the reason for more gang involvement among racial/ethnic minorities relates instead to the differential exposure of minority individuals to risk factors associated with gang membership. While there have been few efforts to empirically assess these specific orientations, the perspectives utilized to explain gang membership allude to both of these types of processes.

Early explanations outlining reasons for why individuals join gangs did not directly address the connection to race/ethnicity. Instead, they outlined conditions related to gang membership to which individuals from various minority groups were exposed, emphasizing the connection to poverty, discrimination, immigration, and urbanization and the resulting social isolation, which are still key concepts today. Theories, such as the social disorganization perspective, propose that community or environmental factors ultimately impact crime (Shaw and McKay 1942). As such, crime and gang membership is not actually related to a group of people (i.e., racial/ethnic minorities), but instead is a result of environments to which minorities are often relegated, characterized by high unemployment, elevated high school dropout rates, deteriorated housing, poverty, a significant number of single parent households, and high mobility rates, as well as industrialization, urbanization, and immigration (Klein 1995; Short and Strodtbeck 1973; Spergel 1964).

Strain and opportunity theory also points to various conditions that are experienced by racial/ethnic minorities that might result in gang membership. These theories indicate that crime is a result of lower-class frustration or strain that develops due to the inability to achieve social and financial success (Cloward and Ohlin 1960; Merton 1968). They suggest that the ability to achieve wealth and power is stratified by socioeconomic class; thus, strain occurs in those areas like disorganized slums where minorities are concentrated and conventional avenues to success are blocked. To relieve strain, people may be forced to use deviant methods to achieve their goals, such as theft, selling drugs, or joining gangs. This is mainly because while legitimate opportunities are blocked for individuals in these areas, illegitimate opportunities are plentiful. Additionally, since the availability of both legitimate and illegitimate opportunities varies across social groups, these theories also suggest that gangs emerge based on the characteristics of the community, not the individuals.

Recent work by Decker and Van Winkle (1996) suggests that threat is a key to gang development. In socially disorganized communities, where social control is at a minimum and where minority populations often live, threat rules the daily existence of individuals in the community. This threat can manifest itself in a variety of ways. One way is to perpetuate the cycle of violence. So, for example, as individuals become isolated from their communities, they become involved in violence, which further isolates them, resulting in more violence (Decker and Van Winkle 1996).

The social learning perspective proposes that individuals learn the norms, values, and behaviors associated with criminal activity from close intimate relationships with criminal peers, including other gang members (Akers 1973; Sutherland 1939). In communities, such as those with concentrated populations of racial/ethnic minorities, norms, values, and behaviors that are conducive to gang membership are prevalent and in fact have in many instances become entrenched into the community. The youth begin to learn from older individuals, thus perpetuating the attitudes and behaviors, resulting in gang behavior among multiple generations.

Other research examines the effects that concentrated disadvantage, including the changing economic structure and poverty, has had on gang development, specifically among African American groups (Hagedorn 1988; Moore 1991; Wilson 1987). During the 1980s and 1990s, jobs became scarce, and the conditions leading to the urban underclass started to breed gangs that have now become a permanent part of many urban communities. As jobs decreased and became more skilled, minority individuals had a harder time getting productive employment, thus affecting the aging out process. Minority youth no longer were able to achieve success through work because of the deindustrialization of urban areas, and thus they continued in the gang longer, making it more institutionalized within the community. Gangs then became their own social institution and a mechanism by which to "make it" in urban inner city environments (Moore 1991; Hagedorn 1988; Wilson 1987). However, while the gang fulfills the needs and provides resources for individuals in these communities in the short term, long term this involvement alienates youth from society and decreases their economic and social capital (Fleisher 1998). As a result, gang members are not socialized to enter mainstream society and pass this on to their children, thus creating an intergenerational transmission of these issues (Fleisher 1998).

More recently, Anderson (1999) considered these elements in the development of a subculture among African Americans in Philadelphia. While Anderson did not specifically address gangs in his work, his general ideas seem applicable. As with strain and opportunity theory, while not specifically addressing racial/ethnic issues, this view represents conditions that are over-representative of racial/ethnic minority populations and ties violence to structural characteristics such as poverty, unemployment, a perceived lack of quality in basic services, and discrimination. This subcultural perspective deals explicitly with gangs as a group that includes its own norms and values that are contrary to the main culture. Individuals adopt these values and create groups, such as gangs, to deal with the isolation and alienation that they experience. Anderson (1999) suggests that in disorganized minority communities, a code develops among all individuals related to toughness and respect that assists in navigating the streets. The expectation of toughness is entrenched in the behaviors of the "street" kids, while the "decent" kids utilize it to make their way in an environment characterized by constant threats (Anderson 1999). In other words, whether an individual is involved in a gang or not, they participate in and perpetuate the norms and values that are conducive to gang membership.

Explanations focusing on gang membership among African American groups tend to emphasize factors such as economics and poverty, while those examining Hispanic groups often include elements of immigration and discrimination (Vigil 2002). The multiple marginality framework (Vigil 2002) proposes that the historical background of the specific racial/ethnic groups is important in understanding how gangs develop in various communities. For example, African American and Hispanic gangs developed largely due to poverty, while Salvadoran and Vietnamese gangs arose based on migration due to civil war. These individual reasons for gang development result in exposure to various ecological, economic, social-cultural, and psychological stresses creating what Vigil (2002) refers to as multiple marginality. The more of these risks and stresses an individual experiences, the more cumulative disadvantage they face, resulting in marginalization from society. This marginalization process is set in motion by macro-historical and macrostructural factors such as immigration and discrimination, which ultimately negatively affects ecological/economic factors. As families and individuals experience these various stresses, their ability to exert social control is minimized. Gangs develop in response to this lack of social control and step into this void to provide socialization in the context of the streets. They then become a way of adapting to marginalization and provide a way to cope with the stresses of urban street life (Vigil 2002).

Risk Factors for Joining

Criminological theory and research point to a number of elements that contribute to gang membership among racial/ethnic minorities. However, much of this information still does not answer an important question: Do individuals from diverse racial and ethnic groups join for different reasons? This concern is tied intricately to programming issues as these results have led to different opinions about whether we require race/ethnic specific programming or not. Some evidence that there are different reasons does exist. Freng and Winfree (2004) examined racial/ethnic differences in reasons given for joining the gang. They found that Whites reported joining for reasons such as protection and friends in the gang, while African Americans and Hispanics tended to join more for protection and money and because peers or family were involved in the gang. Interestingly, contrary to popular belief, no group reported that they joined the gang due to force (Freng and Winfree 2004).

Recently, the risk factor approach has moved gang research in a promising direction. This perspective, developed out of the public health model, examines those factors that put an individual more at risk for gang membership or conversely protects against gang membership. The research based on this perspective does indicate some similarities between groups but also points to some differences. However, while the literature on risk factors for gang membership has become more extensive over the last decade, it still remains underdeveloped in terms of the extent to which risk factors remain constant across racial/ethnic groups (e.g., Curry and Spergel 1992; Freng and Esbensen 2007; Freng and Winfree 2004; Hill et al. 1999; Thornberry 1998; Thornberry et al. 2003).

Some researchers (Curry and Spergel 1992) propose that different factors explain gang membership among African American and Hispanic gang members, while others (Freng and Esbensen 2007) have found few differences between African American, Hispanic, and White gang members. African American gang members tend to be more influenced by social or interpersonal variables such as having family members in gangs, gang members in their classes, and friends who used drugs (Curry and Spergel 1992). Hispanic gang members, on the other hand, seem to have more intrapersonal risk factors such as educational frustration and peer and school self-esteem (Curry and Spergel 1992). For current gang members, differences have also been found (Freng and Esbensen 2007). For Whites, ecological and economic factors, such as parents with lower educational levels and increased levels of social isolation, are more important. On the other hand, for African American and Hispanic gang members, social control and street socialization elements, such as low school commitment, poor opinions or interactions with the police, and being socialized on the street, symbolize the significant risk factors. These differences do not appear for those that report ever being a gang member, however (Freng and Esbensen 2007).

Key Issues

Several weaknesses characterize the current gang research, limiting our knowledge regarding race/ethnicity and its tie to gang membership. The primary problem with most of the literature currently available is that it largely ignores race/ethnicity as an explanatory factor. In other words, cross group comparisons examining actual differences between groups remain relatively rare (Curry and Spergel 1992; Freng and

Esbensen 2007; Freng and Winfree 2004). Instead, race/ethnicity is largely used as a control factor, resulting in little information regarding actual differences or similarities. Additionally, what little research is available indicates some conflicting results on whether the same risk factors impact the various groups (Curry and Spergel 1992; Freng and Esbensen 2007; Freng and Winfree 2004). Organizations or individuals hoping to address the gang problem are thus at somewhat of a disadvantage in knowing what factors should be focused on for which groups.

Additionally, much of the gang research concentrates on specific gangs in particular urban areas that represent racially/ethnically homogenous groups (e.g., Chin 1996; Hagedorn 1988; Moore 1991; Vigil 1988). While this can present detailed information about African American gangs in Chicago or Hispanic gangs in Los Angeles, for example, it does not provide beneficial information about whether these groups are similar or different in their characteristics, attitudes, or behaviors. Furthermore, the research over the past 20 years has largely concentrated on African American and Hispanic groups, ignoring the increasing problems of gangs among Asians, American Indians, and Whites (e.g., Chin 1996; Hagedorn 1988; Moore 1991; Vigil 1988).

Finally, while the research is limited on racial/ethnic differences in joining the gang, an even greater paucity of information exists regarding racial/ethnic disparities in leaving the gang. As we begin to understand that individuals often spend a short amount of time in the gang, with the average being less than 1 year, this topic becomes even more relevant (Thornberry et al. 2003). While research points to a number of reasons and processes for why and how individuals exit the gang, examinations as to whether these reasons or processes remain consistent across racial/ethnic groups remain rare, if not nonexistent (Decker and Lauritsen 2006). Understanding the intricacies of this relationship and whether it varies for the different groups is imperative in terms of focusing our intervention strategies toward removing individuals from gangs.

Policy Issues and Recommendations

Without fully understanding the nature and extent of the problem, we are unable to thoroughly address it. The only sure way to know how best to address the issue of gang membership is to initiate more research that considers these ongoing issues, expanding beyond focusing on specific racial/ethnic gangs in specific locations, ignoring race/ethnicity as a explanatory variable, and the concentration of certain racial/ethnic gangs to the exclusion of others. This additional research will help to more completely understand the nature of gangs and assist in developing prevention efforts. While the reasons for joining should guide us in developing programming principles, surprising little information regarding differences between racial/ethnic groups actually exists, thus making developing recommendations for strong policy practices difficult. While research and theoretical explanations have focused on the importance of larger, social-structural factors (e.g., poverty, immigration, discrimination, and social isolation) which differentially impact the lives of individuals from different races/ethnicities, programs focusing on these elements would of course be much more costly and complex and thus often do not present realistic solutions. Thus, policies tend to concentrate more on individual factors, which have been assumed to be more manageable and malleable.

When considering preventing gang membership, the defining question is the following: Is race/ethnic specific programming needed? The question asks whether the general prevention programming currently utilized that targets individuals regardless of their racial/ethnic group status is sufficient or if we need more targeted programs that focus on specific factors related to gang membership for the various groups. Some evidence exists that perhaps racial/ethnic specific programming is not necessary, but that general prevention programming with some racial/ethnic group sensitive elements may provide some promising practices in effectively addressing a problem that still largely impacts minority populations (Esbensen et al. 2010). In other words, general prevention programming should be effective, in theory, if individuals from different racial/ethnic groups share similar risk factors. For example, targeting factors such as having family members in the gang, a factor tied to gang membership for African Americans, should also have an impact on gang membership for individuals regardless of their racial/ethnic background.

On the other hand, differences in both the reasons that gang members give for joining gangs and the risk factors associated with gang membership for the various groups are apparent (Curry and Spergel 1992; Freng and Esbensen 2007; Freng and Winfree 2004). This perhaps points to the need for some race-/ethnicity-specific or race-/ethnicity-sensitive programming targeting more specifically those factors tied to gang membership for each group (Esbensen et al. 2010). However, our lack of knowledge of racial/ethnic differences hinders our ability to evaluate the need for these types of programming. Unfortunately, no evidence-based programs directly aimed at addressing racial/ethnic differences in risk factors for gang membership exist. Additionally, without more extensive evaluation of programs that are currently in place, knowing the impact on gang membership among various racial/ethnic groups is difficult. Currently, little evidence of the effectiveness of general prevention programs for individuals from various racial/ethnic groups exists (Wilson et al. 2003), resulting in a lack of knowledge of whether the current programs in existence actually work to decrease gang membership among the various racial/ethnic groups. Thus, until more is known about the differences in reasons for joining the gang and risk factors for the various racial/ethnic groups, perhaps it would be sufficient to consider general prevention programming or expand upon existing promising programs to ensure that they are culturally appropriate and relevant.

Related Entries

- ▶ Crime and the Racial Composition of Communities
- ▶ Juvenile Justice System Responses to Minority Youth
- ▶ Race and Ethnicity in Social Disorganization Theory
- ▶ Risk Factors for Gang Membership

Recommended Reading and References

Akers RL (1973) Deviant behavior: a social learning approach, 1st edn. Belmont, Wadsworth

Anderson E (1999) Code of the street: decency, violence, and the moral life of the inner city. W. W. Norton, New York

Ball RA, Curry GD (1995) The logic of definition in criminology: purposes and methods for defining gangs. Criminology 33:225–345

Chin K (1996) Gang violence in Chinatown. In: Huff CR (ed) Gangs in America, 2nd edn. Sage, Thousand Oaks, pp 157–181

Cloward R, Ohlin L (1960) Delinquency and opportunity: a theory of delinquent gangs. Free Press, New York

Curry GD, Spergel IA (1992) Gang involvement and delinquency among Hispanic and African-American adolescent males. J Res Crime Delinq 29:273–291

Decker SH, Kempf-Leonard K (1991) Constructing gangs: the social definitions of youth activities. Crim Justice Pol Rev 5:271–291

Decker SH, Lauritsen J (2006) Leaving the gang. In: Egley A Jr, Maxson C, Miller J, Klein M (eds) The modern gang reader. Roxbury Publishing, Los Angeles, pp 60–70

Decker SH, Van Winkle B (1996) Life in the gang: family, friends, and violence. Cambridge University Press, New York

Esbensen F-A, Peterson D, Taylor TJ, Freng A (2010) Youth violence: understanding the role of sex, race/ethnicity, and gang membership. Temple University Press, Philadelphia

Esbensen F-A, Tusinski K (2007) Youth gangs in the print media. Crim Justice Pop Cult 14:21–38

Fleisher M (1998) Dead end kids. University of Wisconsin Press, Madison

Freng A, Esbensen F-A (2007) Race and gang affiliation: an examination of multiple marginality. Justice Q 24:600–628

Freng A, Winfree LT Jr (2004) Exploring race and ethnic differences in a sample of middle school gang members. In: Esbensen F-A, Tibbetts SG, Gaines L (eds) American youth gangs at the millennium. Waveland Press, Long Grove, pp 142–162

Hagedorn JM (1988) People and folks: gangs, crime, and the underclass in a Rustbelt city. Lakeview Press, Chicago

Hill K, Howell JC, Hawkins JD, Battin-Pearson S (1999) Childhood risk factors for adolescent gang membership: results from the Seattle social development project. J Res Crime Delinq 36:300–322

Howell JC, Egley A, Gleason DK (2002) Modern day youth gangs. Office of Juvenile Justice and Delinquency and Prevention, Washington, DC

Huff CR (1998) Comparing the criminal behavior of youth gangs and at-risk youths. National Institute of Justice Research in Brief, Washington, DC

Klein MW (1969) On group context of delinquency. Sociol Soc Res 54:63–71

Klein MW (1995) The American street gang. Oxford University Press, New York

Loeber R, Farrington DP, Waschbusch D (1998) Serious and violent juvenile offenders. In: Loeber R, Farrington DP (eds) Serious and violent juvenile offenders: risk factors and successful interventions. Sage, Thousand Oaks, pp 13–29

Major AK, Egley A Jr, Howell JC, Mendenhall B, Armstrong T (2004) Youth gangs in Indian country. Juvenile Justice Bulletin, Office of Juvenile Justice and Delinquency Prevention, Washington, DC

Merton RK (1968) Social theory and social structure. The Free Press, New York

Miller WB (1975) Violence by youth gangs and youth groups as crime problem in major American cities. U.S. Government Printing Office, Washington, DC

Moore J (1991) Going down to the barrio: homeboys and homegirls in change. Temple University Press, Philadelphia

National Youth Gang Center (2010) National youth gang survey analysis. http://www.nationalgangcenter.gov/Survey-Analysis. Retrieved 25 Jan 2010

Shaw CR, McKay HD (1942) Juvenile delinquency and urban areas. University of Chicago Press, Chicago

Short JF Jr, Strodtbeck F (1973) Why gangs fight. In: Short JF (ed) Modern criminal. Transaction Books, New Brunswick

Spergel I (1964) Racketville, Slumtown, Haulberg. University of Chicago Press, Chicago

Starbuck D, Howell JC, Linquist DJ (2001) Hybrid and other modern gangs. Juvenile Justice Bulletin. Office of Juvenile Justice and Delinquency and Prevention, Washington, DC

Sutherland EH (1939) Principles of criminology, 3rd edn. J. B. Lippincott, Philadelphia

Thornberry T (1998) Membership in youth gangs and involvement in serious and violent offending. In: Loeber R, Farrington DP (eds) Serious and violent juvenile offenders: risk factors and successful interventions. Sage, Thousand Oaks, pp 147–166

Thornberry T, Krohn MD, Lizotte AJ, Smith CA, Tobin K (2003) Gangs and delinquency in developmental perspective. Cambridge University Press, New York

Vigil JD (1988) Barrio gangs: street life and identity in southern California. University of Texas Press, Austin.

Vigil JD (2002) A rainbow of gangs: street cultures in the mega-city. University of Texas Press, Austin

Wilson SJ, Lipsey MW, Soydan H (2003) Are mainstream programs for juvenile delinquency less effective with minority youth than majority youth? A meta-analysis of outcomes research. Res Soc Work Pract 13:3–26

Wilson WJ (1987) The truly disadvantaged. University of Chicago Press, Chicago

Racial Disparities

▶ Imprisonment and Cardiac Risk

Racial Profiling

▶ Biased Policing

Racket

▶ Racketeering

Racketeering

Ernesto U. Savona and Federica Sarno
Transcrime, Joint Research Centre on Transnational Crime, Università Cattolica del Sacro Cuore di Milano and Università degli Studi di Trento, Milan, Italy

Synonyms

Extortion racketeering; Organized crime; Racket

Overview

The term "racketeering" is often used as a synonym of organized crime, although it cannot be limited to it. Indeed, it is a broader concept, encompassing a wide range of criminal activities. Such activities have to be committed on a continuative basis and as part of an ongoing criminal group which presents a certain degree of organization.

Section "Defining Racketeering: Origins and Core Elements" provides an overview of the main features of the phenomenon since its first appearance and of its development across different contexts. The various criminal activities that can be committed as part of a racketeering pattern are also examined. Section "Examples of Racketeering: The Case of the Italian Mafias" further develops the issue by providing actual examples of racketeering. The case of the Italian Mafias is analyzed, being extortion racketeering their core business.

Although extortion racketeering is commonly associated with organized crime, it can be performed in a casual or a systemic way. This depends on the market opportunities and characteristics of organized criminal groups. The prevalence of causal or systemic patterns across the European continent is examined in section "Extortion Racketeering in Europe."

Defining Racketeering: Origins and Core Elements

The term "racketeering" was originally coined by the Employers' Association of Chicago in 1927 (Witwer 2004, p. 200). It was intended to describe an emerging pattern of criminal conspiracies, aimed at extorting money from legitimate businesses and exerting a monopolistic control over certain (legal or illegal) markets. Collusive arrangements and extortion practices were not a novelty within businesses and labor unions. What the Employers' Association was more concerned about was the increasing involvement of organized crime syndicates in such practices (Witwer 2004).

The power of organized crime groups in the Unites States, in particular of the Italian Mafia, significantly increased between the end of the 1920s and the beginning of the 1930s. In that period, criminal opportunities were created by prohibition which mainly affected the vice industry (Witwer 2004; Scotti 2002). Bootlegging, as well as illegal gambling and prostitution, became profitable – and at the same time risky – activities. Criminal organizations were able to expand their influence over the illegal (and legal) sphere, for example, by extorting money from vice entrepreneurs in exchange of protection from potential assaults or external competition. The infiltration of the labor unions further increased organized crime control over legitimate businesses and its capacity of penetrating the institutional and political system (Scotti 2002). Infiltration in the legal economy may occur by using legitimate businesses as a front cover for criminal activities or by depriving them of their profits through loansharking and extortion (Kelly 1999). When extortion racketeering is established, the relationship between perpetrators (organized crime groups) and victims (businesses) can be of a predatory, parasitic, or symbiotic nature, according to the durability of the relation (short- or long term) and the degree of "complicity" between criminals and their victims. The different types of relationship will be analyzed more in detail in section "Extortion Racketeering in Europe," which specifically addresses the issue of extortion racketeering and, therefore, of the relationship between extorters and victimized businesses.

It was with the aim of curbing the infiltration of legitimate businesses by organized crime that the US Congress, in 1970, enacted the Racketeer Influenced and Corrupt Organizations Act, better known as RICO or RICO Act (Blakey 1990; Blakey and Blakey 1997; Black 1986).

RICO made it unlawful to receive and acquire income from an enterprise, as well as to conduct or participate in the conduction of the enterprise affairs, through a pattern of *racketeering activity*. "Racketeering activity" identifies "*specified felonies* committed as part of a *criminal enterprise* and as part of a *pattern*" [emphasis added] (Albanese 2011, p. 99):

- *Specified felonies*. Specified crimes which may result in a "racketeering activity" can be grouped – for merely operational purposes – into four main categories, not to be considered however as mutually exclusive: (1) violence (i.e., murder, kidnapping, arson, extortion), (2) provision of illegal goods and services (i.e., gambling, prostitution, counterfeiting, drugs), (3) corruption (i.e., bribery, extortion, embezzlement), and (4) commercial or other types of frauds (Blakey 2005). For a detailed description of the specified felonies, see the United States Code 18 U.S.C. § 1961-1.
- *Criminal enterprise*. The term enterprise is intended to include broadly "any individual,

partnership, corporation, association, or other legal entity, and any union or group of individuals associated in fact although not a legal entity" (United States Code 18 U.S.C. § 1961-4). In other words, a criminal enterprise refers to any group – or even an individual associated to it – whether legal or illegal, officially formalized or just existing, that is used as a base for committing criminal activities (Albanese 2011).

- *Pattern*. The pattern indicates two or more of the specified felonies committed within a period of 10 years (excluding any period of imprisonment of the offender) (United States Code 18 U.S.C. § 1961-5).

In order to be considered as a punishable offense according to RICO legislation, racketeering should provide that (a) two or more specified offenses are committed, (b) within a 10-year period, (c) by a person who is involved (e.g., employed, associated, related) in the activities of an *ongoing criminal enterprise*.

The innovative feature of RICO legislation was that, rather than focusing on specific offenses, it was instead centered on the ongoing criminal enterprise. What is illegal, and therefore punished, is the participation in a group that commits certain types of offenses (Albanese 2011, pp. 100–101). As such, an individual who belongs or is in a way associated with the ongoing criminal enterprise can be charged with racketeering even if the criminal offenses have been physically committed by other members (Schneider 2005; Abadinsky 2009). This provided prosecutors with an effective tool to charge, prosecute, and convict mobsters, gangsters, and other organization leaders. Before the introduction of RICO, such a result would have represented a challenging task, in that traditional criminal laws were directed at prosecuting single and specific offenses. By doing so, the heads of a criminal enterprise (i.e., an organized crime family) were hardly sentenced, since usually they were only indirectly involved in the commission of crimes (Blakey 1990).

In addition to criminalizing racketeering, the Congress also provided extended penalties and unique criminal and civil instruments to combat such an offense (Blakey 1990). The maximum criminal penalties originally established under RICO consist of a $25,000 fine and 20 years imprisonment. The civil RICO provisions also allow to seize and confiscate property and other goods assumed to be the proceeds of a racketeering activity. The purpose is to attack criminal assets and deprive criminal enterprises of their economic resources. Inclusion of the forfeiture provisions under the civil law represents one of the most significant and, at the same time, most controversial aspects of RICO (Schneider 2005). The burden of proof requested by the civil laws is lower compared to the criminal ones, so that the government may seize and confiscate assets from a criminal enterprise involved in racketeering activities without notice and "ex parte" (Schneider 2005; Albanese 2011).

Although it was originally elaborated to counter directly traditional organized crime groups – in the sense of the mafia-type organizations – and their (il) legal activities, RICO legislation cannot be considered as merely limited to organized crime. RICO provisions have demonstrated to be broader in their scope, being successfully applied even in the fight against political corruption, white-collar crime, and violent groups (Blakey 1990). Legitimate business enterprises have been charged as well (see, e.g., the court case *Sedima, S.P.R.L. v. Imrex CO.*) (Schneider 2005). RICO is intended to reach any sort of enterprise, whether legal or illegal, provided it has been involved in a pattern of racketeering activity.

The following cases provide an account of the variety of situations in which RICO has been used. In 2000, the decision of a federal judge in Los Angeles ruled that the local Police Department could be sued as a racketeering enterprise, in that some corrupt police officers had allegedly robbed, beaten, framed, and shot suspects (Albanese 2011; ABCNews 2000). In a similar way, in 2002, the governor of Louisiana was convicted under RICO, due to his involvement in various corrupt criminal schemes related to the riverboat gambling license process (Albanese 2011; United States Court of Appeals 2002). In 2008, the leader of a Colombian Drug Cartel was convicted under RICO with the charge of being at the head of a racketeering enterprise that engaged in murder, narcotics trafficking, money laundering, and bribery (Albanese 2011; Department of Justice 2008).

RICO has not been immune from criticism. An improper and overused resorting to the civil provisions of RICO has been noticed, for example, in the civil suits among legitimate commercial enterprises (see Schneider 2005, pp. 662–663).

In contexts other than the USA, as, for example, in the Eastern and some Southern European countries (especially Italy and Spain), the term racketeering could keep a more specific focus. It often refers to typical organized crime activities, such as illicit gambling and prostitution, transportation of stolen goods, loan-sharking, and, above all, extortion.

Examples of Racketeering: The Case of the Italian Mafias

As the history of Italian organized crime shows, extortion and protection represent the core businesses of the Italian Mafias (Savona 2012). For the purpose of this essay, the term "Italian Mafias" refers to the Sicilian Cosa Nostra, the Calabrian 'Ndrangheta, and the Campanian Camorra.

Gambetta (1993), in his book *The Sicilian Mafia*: *The Business of Private Protection*, defines the mafia as an "industry which produces, promotes and sells private protection." Protection is provided not only to legitimate businesses but also to criminal actors operating in the underworld.

In the upperworld, the mafia exerts a "para-state" function by taking advantage of the state's inefficiency in properly preserving legal transactions and free competition from potential abuses. In the underworld, protection is needed because criminal activities are usually carried out in a context free from any formal rule and where no sort of protection by the state is envisaged (Arlacchi 1988). In such a setting, illegal actors resort to the services offered by those agencies – namely, mafia-type organizations – which are able to regulate potential conflicts and provide protection against other cheating and competing criminals (Varese 2001; Transcrime 2009).

Protection becomes extortion when the mafia benefits from a service that it does not actually provide (Gambetta 1993). More precisely, extortion is "the obtaining of property from another, with his consent, induced by wrongful use of actual or threatened force, violence, or fear, or under color of official right" (The United States Code 18 U.S.C. § 1951(b)(2)).

When extortion is committed on a regular basis, it turns into racketeering. Extortion racketeering has been defined as "an institutionalized practice whereby tribute is collected on behalf of a criminal group that, in exchange, claims to offer (. . .) protection" (Volkov 2002, p. 1). Italian Mafias resort to it as a means of territorial sovereignty and a financial resource.

Through extortion racketeering, organized crime can first of all control the territory and the economic and political activities being implemented in it. In doing so, the mafia also consolidates its presence as part of the local culture, thus favoring the establishment of a "cultura mafiosa" (mafia culture) (Savona 2012). This means that people gradually become used to the mafia presence and therefore accept it, or even cooperate with it, in order to achieve some sort of benefits in terms of protection and advantages against competitors. In a similar way, through extortion, organized crime groups can infiltrate legitimate businesses and get hold of them when extorted entrepreneurs are no longer able to pay.

At the same time, extortion represents a source of income used by criminal organizations to finance their illegal activities and sustain the maintaining costs of the organization (e.g., payment of the affiliates, maintenance of the families of incarcerated members, military expenses) (Paoli 2003, p. 165).

As pointed out by Paoli (2003), the spread of extortion racketeering on a large scale has been driven by the fact that it is an easy crime to carry out and represents a rapid way to make money: "it does not requires a high initial investment, it carries low managing costs and, in areas where the state's protection in not regarded as adequate or reliable, it is also a low-risk operation" (Paoli 2003, p. 165). As such, extortion racketeering became one of the main resources by which the Italian Mafias achieved territorial sovereignty in Sicily, Calabria, and Campania. These are the Southern Italian regions in which the Italian Mafias have traditionally established themselves.

In its more "typical" form, extortion racketeering consists of a periodical money payment (the so-called pizzo) required by criminal organizations from legitimate businesses under the threat of violence and intimidation. In some other cases, it can take the form of either extortion in goods and services (such as the imposed removal of merchandise from a company) or extortion of the workforce (through the violent control of the local labor market) (Paoli 2003; Daniele 2009). In this regard, Monzini (1993, pp. 22 et seq.) distinguishes among different types of extortion: (See also Transcrime 2009, pp. 22-23)

1. *Extortion protection* consists in a criminal system, similar to taxation, by which regular payments are imposed using violent means. A long-lasting relationship usually develops between the extortionist and the extorted, so that a sort of legitimization of the mafia's presence and its control over the production gradually arises (Daniele 2009).
2. *Labor racketeering* is a means of violent negotiation established within single production units. The mafia extorts the workforce by controlling access to the labor market and employment.
3. *Monopolistic racketeering* is a specific market strategy enforced by the use of violence and aimed at physically eliminating competitors and creating monopolistic coalitions. Such a result is achieved by establishing a monopolistic control on the access to those resources which are crucial for a given sector.

It is interesting to note that, when extortion racketeering is carried out in mafia contexts, organized crime uses a certain degree of "professionalism" (Savona 2012). First of all, as suggested by Grasso and Varano (2002, p. 80), the mafia does not select the territory on which to establish the extortion racket. It rather prepares it. This means that in many cases the extortionists are the actors creating the conditions for the demand of protection to arise: that is, some minor crimes are committed by the same criminal organizations in order to generate a common sense of insecurity (Transcrime 2009, 22).

Furthermore, the mafia does not target its victims randomly. On the contrary, it selects them according to their vulnerabilities (Grasso and Varano 2002, p. 78; Savona 2012). First of all, the economic activity performed by the victim has to be located within the zone of influence of the mafia (Grasso and Varano 2002, p. 78). This enables Mafiosi to exert successfully their intimidating power. Secondly, the business carried out by the targeted company has to be attractive for the organized crime group (Savona 2012, p. 8), which means offering high profits at low risks. Finally, instruments and infrastructures needed for the production (e.g., warehouses, restaurants, shops, excavators, vehicles, and other machinery) must be exposed and well visible, so that they can be easily identified and targeted by the extortionists. At the same time, the entrepreneur is aware that the company's business will be closed down should those facilities be destroyed (Grasso and Varano 2002, p. 78; Savona 2012, p. 8). Once a racketeering relationship has been established, defining the type (money, goods, services, etc.) and the amount of the payment represents the next step. When imposing a protection racket, mafia groups often evaluate the economic resources of the victims and calculate the amount of money they are willing to pay (Savona 2012). Even when threatening or frightening the entrepreneur, the mafia decides the extent to which violent actions should be performed, in order to avoid that the victim reports to the police.

Extortion practices carried out by the Italian Mafias are directed not only to legitimate businesses but also to other criminals. The Neapolitan Camorra, for example, has imposed its control over the urban underworld, by establishing a comprehensive system of taxation of all the illegal traffics taking place in its territory.

Payments are required to criminals in the form of shares of the revenues arising from theft, prostitution, smuggling, gambling, and other criminal markets (Monzini 1993).

As such, extortion racketeering can be regarded, rather than as a single activity, as a hub where different spheres (legal and illegal) intersect (Savona 2012). Scientific literature has widely recognized that organized crime activities are placed along a continuum, so that it is difficult – and probably improper – to separate strictly criminal activities from non-criminal ones. The selection of the area, in which it is better to operate, depends on the risks and opportunities related to it. The Italian Mafias have engaged in both criminal and non-criminal activities since high opportunities and low risks were available in illegal businesses as well as in legal ones.

Some differences among the three mafias, especially as regards "what" is concretely done and "how," can be noted. Such a choice is in fact influenced by many factors: know-how, culture of the organization, and social and economic factors. So, for example, the Campanian Camorra is heavily involved in illegal trafficking of waste and counterfeited goods, while this is not the case for the Sicilian Mafia. In a similar way, drug dealing has been one of the main activities of the mafia in Sicily during the 1970s and the 1980s, whereas now it is one of the core businesses of the Calabrian 'Ndrangheta. Despite these differences, some similarities can be outlined. Among them, extortion and infiltration in the legitimate sphere represent features that are common to, and characteristics of, all three Italian Mafias (Savona 2012).

Extortion Racketeering in Europe

This part draws mainly on a precedent analysis (Savona and Zanella 2011, pp. 261–267).

Moving to a broader level of analysis (Europe), market opportunities and variations in the characteristics of organized criminal groups can give rise to different types of extortion racketeering: *systemic* and *casual* racketeering. Extortion racketeering is *systemic* when it is deeply rooted and extended across a territory, being a core part of the activities of organized crime groups. Extortion racketeering is *casual* when it is not extended across a territory since criminal organizations do not routinely engage in such a criminal activity.

These different types of extortion racketeering are shaped by four interrelated variables: (a) market opportunities, (b) the organizational structure of criminal groups, (c) their presence at local level, and (d) the victim/offender relationship.

In other words, on the one hand, the more organized criminal groups focus their activity on the local territory because of market opportunities, the more they develop a monopolistic position and a consequential hierarchical structure, and the more they establish parasitical and symbiotic relationships with their extortion victims, the more extortion racketeering becomes *systemic* (i.e., widespread and continuous). On the other hand, the more criminal market opportunities are open to transnational activities, the more criminal groups are organized in networks, and the more they establish predatory relationships with their extortion victims, the more extortion racketeering becomes *casual*.

The four variables and their interrelations are better described below:

(a) *Market opportunities*

Extortion is an old and simple crime committed by organized crime when the risks are low and the benefits are high. It occurs in contexts where (1) the victims do not report the crime and (2) are willing to pay the protection tax. These two conditions often arise within close-knit ethnic communities (e.g., among Italians at the beginning of the nineteenth century in New York City or among Chinese communities in the USA or in Europe). The low risks connected to this ethnic homogeneity and the consequent control of the territory are balanced by high benefits related to possible criminal market opportunities only. Extortion is systemic when other criminal alternatives are not available or cannot be adopted because of the low expertise of the group and its organization.

(b) *Organizational structure of criminal groups*

Although there is not a direct relationship between the organizational structure of criminal groups and extortion, the literature and the data show that where extortion is practiced on a large scale, and is systemic, the groups that engage in it are organized hierarchically. Where extortion is more casual, their structure is more flexible (taking the form of a network).

Due to their structure, which enables a lasting presence in a given territory, hierarchical criminal groups gain in reputation and may exercise effective threats of contingent violence against their victims. Moreover, these threats are reinforced by the fact that those threatened believe that these criminal groups "can corrupt legitimate authority or in some other way ensure that they avoid apprehension" (Reuter 1994, p. 95) and can act as an industry which produces and sells private protection (Gambetta 1993).

These elements of reputation – that is, the ability to neutralize law enforcement by means of corruption and the production and sale of protection – are closely related to the type of organized criminal group. Criminal groups with network structures "are not interested in, or capable of, exercising such a quasi-political power" (Paoli and Fijnaut 2004, p. 608). They are too small and ephemeral to be able to carry out systemic extortion racketeering. Only in a few cases they are involved in casual extortion racketeering practices. Networks may openly use violence to heighten their capacity to commit extortion, but they tend to be short lived because they lack the necessary structure and expertise.

In sum, although it is not automatic, the relationship between hierarchical structure and systemic extortion, on the one hand, and flexible structure and casual extortion, on the other, can be explained by means of the other variables shaping extortion: the local dimension of the organized crime action, its control over the territory, and the victim/offender relationship.

(c) *The presence at local level*

Why does extortion proliferate when organized crime operates at local level? And why is control of the territory so important? The explanation resides in the relationships between organized crime groups and local politicians, administrators, and businesses. The local level is the dimension where collusion with organized crime is easier and reciprocal exchange more profitable. Extortion racketeering is used to finance the criminal organization and its criminal activities and to consolidate its capacity to control local resources such as property, markets, services, and votes.

Criminal groups which exercise intense control over the local territory tend to commit systemic extortion racketeering within legal markets and in the underworld. As far as the legal markets are concerned, extortion racketeering is often viewed as the key "to infiltration and baronial domination of sections of the legitimate economy" (Bell 2000, p. 183). In regard to the underworld, it has been noted (Landesco 1968) that extortion racketeering is often used to "protect" criminal markets. By collecting extortion money from criminals, organized criminal groups establish a form of tax levying system which facilitates the establishment of monopolistic areas and creates barriers to entry that make criminal offenses less attractive.

(d) *The victim/offender relationship*

When analyzing the victim/offender relationship in extortion racketeering cases, it should be considered that "the boundaries between victim and accomplice are often (...) blurred" (Blok 2008, p. 8). These boundaries create the difference between systemic and casual extortion racketeering.

When networks are involved in extortion racketeering, they establish predatory relationships with their victims. Being unable to establish lasting relationships with them, criminal networks consequently act with the "aim or effect to destroy or bleed to death" (Passas 2002, p. 21) the victims, exacting considerable extortive payments in a short period. This is typical of casual extortion racketeering.

By contrast, hierarchical criminal groups, benefiting from their reputation and durability, may establish parasitical or symbiotic relationships with their victims. This contributes to making extortion racketeering systemic.

The relationship is "parasitical when the aim is to preserve the viability of the target, such that illegal benefits can be extorted on a more or less regular basis" (Passas 2002, p. 21). By establishing a lasting relationship with the victim, the offender "harms the host a little at a time, without killing it, or only kills it in slow motion" (Felson 2006, p. 196). In other cases, the relationship may be symbiotic in nature, so that the victim becomes a "friend" of the extorter. The victim thus gains an advantage that is not "simply that of avoiding the likely damages that would otherwise ensue, but can extend to assistance in disposing of competitors, or protection against the threat of isolated bandits, and against the risk of being cheated in the course of business transactions" (Gambetta 2000, p. 166).

Systemic and Casual Extortion Racketeering: The European Case

The link between the presence of casual or systemic extortion racketeering and the variables indicated above has been recently investigated in a study on extortion and organized crime in the 27 European Union Member States (Transcrime 2009). This study analyzed the complex variety of criminal organizations present in the 27 EU Member States, the criminal market opportunities they exploit, the differences in their organizational structures (some are hierarchical, some take the form of a network, some are permanent, some are more ephemeral) together with their differing criminal activities and control over the territory, and the various relationships they establish with their victims.

The study explained that extortion racketeering is casual in most of the European Union Member States (EU MS). The only exceptions are the Eastern EU MS (Bulgaria, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia, Slovenia) and some EU MS in the South of Europe, namely, Spain and Italy.

When the geographical locations of the 27 EU MS are divided into four clusters, the following patterns emerge (see Caneppele et al. 2009, pp. 253–254):

1. *Northern Europe* (Denmark, Estonia, Finland, Latvia, Lithuania, Sweden). Owing to the prevalence of smuggling activities in this region, extortion racketeering is casual in Denmark, Finland, and Sweden. In fact, these countries are distinguished by the presence of criminal organizations that do not exercise control over the territory because they are transnational in their smuggling activities. Proximity to Russia and the Caucasian countries and the presence of hierarchical criminal groups have generated systemic extortion racketeering in Estonia, Latvia, and Lithuania.
2. *Western Europe* (Austria, Belgium, France, Germany, Ireland, Luxembourg, the Netherlands, United Kingdom). Owing to the structure of the organized criminal groups operating in this area and to their transnational activities, extortion racketeering is casual. Most of the countries in this area suffer from extortions carried out within close-knit ethnic communities. This is the case, for example, of Austria, Belgium, France, Germany, Luxembourg, and the United Kingdom.
3. *Central/Eastern Europe* (Bulgaria, Czech Republic, Hungary, Poland, Slovakia, Slovenia, Romania). Because of its proximity to the Balkans, this region is an important transit area for criminal goods and services and in particular for smuggling and trafficking activities. However, the likely hierarchical structure of the criminal organizations operating in this area, together with their strong presence at local level, makes extortion racketeering systemic in countries such as Bulgaria, Czech Republic, Hungary, Poland, Romania, Slovenia, and Slovakia.

4. *Southern Europe* (Cyprus, Greece, Italy, Malta, Portugal, Spain). The region is highly heterogeneous. Differences outweigh similarities in the structure of organized crime groups operating in these countries, and this is also reflected in the different ways in which extortion racketeering is conducted. Extortion racketeering is casual in most of the countries belonging to this cluster.

In two countries – Italy and Spain – extortion racketeering is systemic. In Spain, it is systematic because it is carried out by terrorist groups belonging to ETA which are well structured and rooted in the territory (see Transcrime 2009, Spain country profile). As seen above, in Italy, the phenomenon is mainly linked to mafia-like organized crime, for which extortion racketeering plays a fundamental role in terms of both exercising control over the territory and financing criminal activities.

Related Entries

► History of Organized Crime in the United States
► Italian Mafia
► Organized Crime, Types of

Recommended Reading and References

Abadinsky H (2009) Organized crime, 9th edn. Wadsworth, Cengage Learning, Belmont

ABCNews (2000) Judge: LAPD can be sued as racketeering enterprise, April 20, Los Angeles

Albanese JS (2011) Transnational crime and the 21st century. Oxford University Press, New York

Arlacchi P (1988) Saggio sui mercati illegali. Rassegna italiana di sociologia XXIX(3):403–437

Bell D (2000) The end of ideology. On the exhaustion of political ideas in the fifties. Harvard University Press, Harvard

Black B (1986) Racketeer Influenced and Corrupt Organizations (RICO)—Securities and Commercial Fraud as Racketeering Crime after Sedima: What is a "Pattern of Racketeering Activity"?". Faculty Articles and Other Publications. Paper 77.

Blakey GR (1990) Foreword: debunking RICO's myriad myths. St John's Law Rev 64(4):701, Article 1

Blakey GR (2005) Commentary to definitions title 18. Crimes and criminal procedure part I. Crimes chapter 96. Racketeering influenced and corrupt organizations, South Bend, National Institute for Trial Advocacy, Notre Dame Law School. http://www.google.it/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCUQFjAA&url=http%3A%2F%2Ffreedom-school.com%2F1961_definitions.pdf&ei=T_R2UNaLCOjg4QTxzoGwAw&usg=AFQjCNFgwVxPjiNUcepaXZI5XG00gQjrqw. Accessed 11 Oct 2012

Blakey GR, Blakey JR (1997) Civil and criminal RICO: an overview of the statute and its operation. Def Couns J 64(1) pp. 36-44

Block AA (1980) East side west side. Organizing crime in New York 1930–1950. University College Cardiff Press, Cardiff

Blok A (2008) Reflections on the Sicilian mafia: peripheries and their impact on centres. In: Siegel D, Nelen A (eds) Organised crime: culture, markets and policies. Springer, Dordrecht

Caneppele S, Gosetti F, Zanella M (2009) Extortion racketeering in Europe: problem dimension and possible countermeasures. Presented at the 9th annual conference of the European society of criminology, criminology and crime policy between human rights and effective crime control, Ljubljana, Slovenia, 9–12 Sept 2009

Daniele V (2009) Organized crime and regional development. A review of the Italian case. Trends Organized Crime 12(3–4):211–234

Department of Justice (2008) Leader of Colombian drug cartel pleads guilty to racketeering charges, Oct 17, Washington. http://www.justice.gov/opa/pr/2008/October/08-crm-930.html. Accessed 11 Oct 2012

Felson M (2006) Crime and nature. Sage, London

Gambetta D (1993) The Sicilian mafia. The business of private protection. Harvard University Press, Cambridge

Gambetta D (2000) Mafia: the price of distrust. In: Gambetta D (ed) Trust: making and breaking cooperative relations. Oxford University Press, Oxford, pp 158–210

Grasso T, Varano A (2002) 'U pizzu. L'Italia del racket e dell'usura. Baldini & Castoldi, Milano

Kelly RJ (1999) The upperworld and the underworld: case studies of racketeering and business infiltrations in the United States. Springer, New York

Landesco J (1968) Organised crime in Chicago. Part III of the Illinois crime survey 1929. The University of Chicago Press, Chicago

Monzini P (1993) L'estorsione nei mercati leciti e illeciti. Liuc papers n. 1 – Serie storia, impresa e società

Paoli L (2003) Mafia brotherhoods: organized crime, Italian style. Oxford University Press, New York

Paoli L, Fijnaut C (2004) Organised crime in Europe. Concepts, patterns, control policies in the European union and beyond. Springer, Dordrecht

Passas N (2002) Cross-border crime and the interface between legal and illegal actors. In: Van Duyne P, Von Lampe K, Passas N (eds) Upperworld and underworld in cross-border crime. Wolf Legal, Nijmegen, pp 11–43

Reuter P (1994) Research on American organized crime. In: Kelly RJ, Chin K, Schatzberg R (eds) Handbook of organized crime in the United States. Greenwood Press, Santa Barbara, pp 91–120

Savona EU (2012) Italian Mafias asymmetries. In: Siegel D, van de Bunt H (eds) Traditional organized crime in the modern world. Springer, New York

Savona EU, Zanella M (2011) Extortion and organized crime. In: Natarajan M (ed) International crime and justice. Cambridge University Press, New York

Schelling TC (1967) Economics and criminal enterprise. The Public Interes 7(Spring):61–78

Schelling TC (1971) What is the business of organized crime? The J Public Law 20(1):71–84

Schneider S (2005) Racketeer influenced corrupt organizations. In: Salinger LM (ed) Encyclopedia of white-collar and corporate crime, vol 1. Sage, Thousand Oaks

Scotti B (2002) Rico vs. 416-bis: a comparison of U.S. and Italian anti-organized crime legislation, 25 Loy. L.A. Int'l & Comp. L. Rev. 143. http://digitalcommons.lmu.edu/ilr/vol25/iss1/5. Accessed 10 Oct 2012

Transcrime (2009) Study on extortion racketeering the need for an instrument to combat activities of organised crime. Final report prepared for the European Commission, Bruxelles (unpublished)

United States Court of Appeals (2002) United States of America v. Edwin Edwards, 303 F.3d 606, Fifth Circuit, Aug 23, Louisiana

Varese F (2001) The Russian mafia: private protection in a new market economy. Oxford University Press, New York

Volkov V (2002) Violent entrepreneurs. The use of force in the making of Russian capitalism. Cornell University Press, London

Witwer D (2004) The most racketeer-ridden union in America: the problem of corruption in teamsters union during the 1930s. In: Kreike E, Jordan WC (eds) Corrupt histories. University of Rochester Press, New York

Radical Criminology

► Marxist Criminology

Radical Environmentalism

► Ecoterrorism

Randomization

► Randomized Block Designs

Randomized Block Design

► Randomized Block Designs

Randomized Block Designs

Barak Ariel[1,2] and David P. Farrington[2]
[1]Faculty of Law, The Hebrew University, Jerusalem, Israel
[2]Institute of Criminology, University of Cambridge, Cambridge, UK

Synonyms

Experiments; Randomization; Randomized block design

Overview

Experimental criminologists pay close attention to baseline equality between the treatment and control groups. When the study groups are not similar to one another, this introduces biases and confusion in the interpretation of the treatment effect. Often simple random assignment procedures are used to create similar groups, where the entire sample is divided in half into treatment and control conditions. However, research shows that this simple process does not always work. By chance alone, one group can be larger and have more of a particular type of participants or type of traits. Such instances make it difficult to compare "like with like," which is one of the cornerstones of randomized controlled trials. Instead, experimentalists and statisticians advise using randomized block designs, in which cases are randomly assigned within homogeneous blocks of participants, rather than from the overall sample. The variance within the blocks is minimized, which decreases the experimental "noise" and increases the statistical power of the test. Randomized block designs can also be used to "force" size equality in order to create exactly the same group sizes, something which cannot be done with simple random assignment. Below, instances in which it is optimal to use randomized block experiments are discussed, as well as how to design them, and the obstacles to avoid when using these designs. The general type of randomized block designs is introduced, along with two special cases: permuted randomized block designs and matched pairs, which are useful in small-scale studies where group-size equality is particularly important and when unmasking of the random allocation sequence does not create a source of concern.

Introduction

In randomized controlled trials, the experimental and control groups should be as similar to one another as possible. Without this equality prior to the administration of the treatment to the experimental group and not to the control group, the advantages of the experimental approach are dramatically diminished. Therefore, researchers conducting randomized controlled trials pay particularly close attention to the way cases are assigned to the treatment conditions of the study and work hard to preserve the integrity of this process, in both the design and implementation of the study.

Such balance – or group comparability – is achieved using a process of random allocation of cases into treatment and control groups. Probability theory predicts that the equal chance of each case to be assigned to either treatment or control conditions (e.g., a 50-50 chance) creates two groups that are likely to be similar to one another – except for the intervention that "experimental participants" are exposed to but "control participants" are not. Because participants are assigned to groups by chance, the likelihood that the experimental group will turn out to be different from the control group is significantly minimized when compared with a nonrandom assignment procedure. Therefore, any baseline traits that could explain the relationship between the treatment and outcome are equally divided between the two study groups, and the potential effect of these traits is believed to be controlled for by this design. Thus, the only explanation for any difference in outcomes between the experimental and control groups ought to be attributable to the treatment effect, and nothing else.

In practice, researchers use various techniques to produce a random assignment sequence. The most popular of these techniques is simple random assignment, in which the researcher divides the entire sample into two (or more) groups, using a basic algorithm. For example, if there are 200 cases in a particular study, 100 would receive the treatment, and 100 would serve as the control participants – and that assignment to the two groups is completely random. All statistical software packages, including various online calculators, can easily produce a random assignment sequence that would enable the researcher to allocate his or her cases in such a way.

However, while this simple random assignment procedure is likely to provide an overall balance, equality is not guaranteed. The risk of creating unequal groups still exists, despite the random assignment, when the allocation is not restricted. Unbalance can become a problem in the form of unequal group sizes (e.g., 90 compared to 110 participants) or one group "holding" more participants with a particular trait (e.g., more females than males or more motivated participants in one study group than the other). These problems can be quite concerning, and a large body of literature has tried to deal with these issues over the years (see Ariel and Farrington 2010; Weisburd and Taxman 2000). For example, unbalance is usually associated with a lower statistical power of the research design (see Lipsey 1990; Weisburd and Taxman 2000), particularly in small samples. This means that the ability of the test to show there are indeed differences between the study groups *because* of the intervention is greatly diminished. It has also been shown that interpreting the results from an unbalanced trial may lead to reaching biased conclusions about the true outcome effect of the tested intervention (Torgerson and Torgerson 2003).

The Randomized Block Design

Understanding that simple random assignment cannot guarantee the necessary balance between the groups has led researchers to try and anticipate and overcome the problem, and certain measures can now be taken to address it. Some are more successful than others. For example, it was (and perhaps still is) commonly assumed that increasing the size of the sample is associated with more balance and, in turn, with greater statistical power. But, as Weisburd et al. (1993) have shown, there are instances in which there can be an inverse relationship between sample size and statistical power. Especially in heterogeneous samples, larger samples are likely to include a wider diversity of participants than smaller investigations. Because it is necessary to establish very broad eligibility requirements in order to gain a larger number of cases, large trials tend to attract a more diverse pool of participants. This increases the variability in the data as there is more "noise," which makes it difficult to detect the effect of the treatment. Therefore, the design benefits of larger trials may be offset by the implementation and management difficulties they present.

Given Weisburd et al.'s (1993) and similar research findings, along with a broader understanding of the shortcomings of simple random assignment, clinicians and statisticians have developed other research designs to more adequately handle unbalanced groups or high variability. One type which is slowly gaining recognition and is being applied more and more in experimental criminology is called the randomized block design. The concept behind randomized block designs is generally straightforward: if the researcher knows before he or she administers the treatment that certain participants may vary in significant ways, this information can be utilized to his or her advantage. The researcher can use a statistical design in which all participants are not simply randomly allocated directly into experimental and control groups. Instead, they are randomly allocated into groups *within* blocks.

The randomized block design is used in order to decrease the variance in the data (Lachin 1988). Unlike simple random assignments, where units are unrestrictedly distributed at random to either treatment or control groups (or more than two groups, as the case may be), units in a randomized block design are allocated randomly to either treatment or control *within* pre-identified blocks. The blocking process is established based on a certain criterion, which is intended to divide the sample, prior to assignment, into subgroups that are more homogeneous (Hallstrom and Davis 1988). Then, within these blocks, units are randomly assigned to either treatment or control conditions.

Variations of this design were implemented in some of the most influential studies in criminal justice research, particularly in place-based criminology, such as the Jersey City Problem-Oriented Policing Experiment at violent places (Braga et al. 1999). Recently, Weisburd et al. (2008) have used this design in an experiment on the effects of the risk-focused policing at places (RFPP) approach to preventing and reducing juvenile delinquency. Before administering the treatment, they found considerable variability in the characteristics of violent places selected for the study, which jeopardized the statistical power of the evaluation. In this case, the impact of the treatment will be more difficult to detect when there is a large variability in the assessment of study outcomes. The data were thus more adequately managed under a randomized block design. As in other hot-spot experiments (e.g., Ariel and Sherman forthcoming), the randomized block design is advantageous because it decreases the variations that characterize places, such as the drug activity, and structural and cultural characteristics. In several hotspots experiments, it was found that an examination of the distribution of arrest and call activity in the hot spots before the experiment revealed that the places fell into distinct groups – for example, high, medium, and low criminal activity blocks. Therefore, hot spots randomly allocated to experimental and control conditions *within* each of these conditions created more homogeneous groups of hot spots.

Randomized block designs can be used in person-based experiments as well. Similar issues arise in analyses of individuals in schools, prisons, and other environments, and so blocks can be created in these studies as well (see, e.g., Ttofi and Farrington 2012). For example, a large randomized controlled trial was conducted to test the hypothesis that differences in wording of letters sent to taxpayers in Israel would affect various aspects of their taxpaying behavior, such as the amount of money they were willing to report and pay to the tax authority (Ariel 2012). Nearly 17,000 taxpayers were randomly assigned to different groups, each receiving a different type of letter. However, a large sample of taxpayers introduced high variability. For example, the sample included very poor taxpayers as well as very rich taxpayers. The size of the sample also increased the likelihood that, in one of the study groups, there would be a disproportionally larger number of extremely rich taxpayers, so their effect on the results could skew the conclusion. Therefore, the sample was divided, before random assignment, into blocks of income levels. Within each of these income-level blocks, the participants were then randomly allocated into the different letter groups. Therefore, this procedure allows for measurement of not only the overall effect of letters but also of specific effects within the blocks (i.e., effect of letters on participants with different income levels).

Designing a Randomized Block Experiment

Consider the following hypothetical trial as a way to show the benefits of blocking and when this design should be considered. Imagine a trial with one experimental group consisting of 52 drug-addicted offenders, treated in an antidrug program. Another group of 48 drug-addicted offenders serve as the control group, receiving no drug treatment. The allocation to treatment and control is done randomly. The experiment is conducted in order to evaluate the merits of the program, where success is nominally defined by a decrease in drug use (on a scale of 1–8, 8 being the highest and 1 being the lowest). However, unlike other drug treatment programs, this particular treatment is very costly. Therefore, unless very high success rates are registered, the facilitators will be unlikely to recommend its implementation in the future. All eligible participants selected for this trial are known to be drug abusers before entering the program. Prior to the treatment, the drug use level was not statistically different between the two groups (both averaging about 5.85 on the said scale).

At the 6-month follow-up period, drug use was measured again for both groups. Results show that the program was successful: there was an overall 15 % reduction in the treated experimental group, when compared with the untreated control group. A visual depiction is presented in Fig. 1. At the same time, considering the costs of the program and that there was “only” a 15 % reduction, the policy implication is to not recommend further implementation of the program.

However, a closer look at the distribution of the pretest scores indicates that the sample is comprised of drug abusers of four different types of drugs (with no crossovers): marijuana, MDMA, alcohol, and heroin. Generally speaking, it could be hypothesized that drug abusers of these different drugs are not addicted in the same way (based on both pharmacological and psychobiological qualities of these substances). Therefore, the prospective success rate of the drug intervention program is also hypothesized to be different for each subgroup; it is likely that getting clean from a heroin addiction or alcohol addiction is more difficult than from marijuana or MDMA. Thus, dividing the overall sample based on the type of drug should produce more homogeneous subgroups, whose chance to get clean is different.

One primary advantage of looking at the study this “blocked” way is that even if the strength of the treatment effect within each block is similar, more precise estimates of the magnitude of

Randomized Block Designs,
Fig. 1 Hypothetical drug treatment trial: simple random assignment

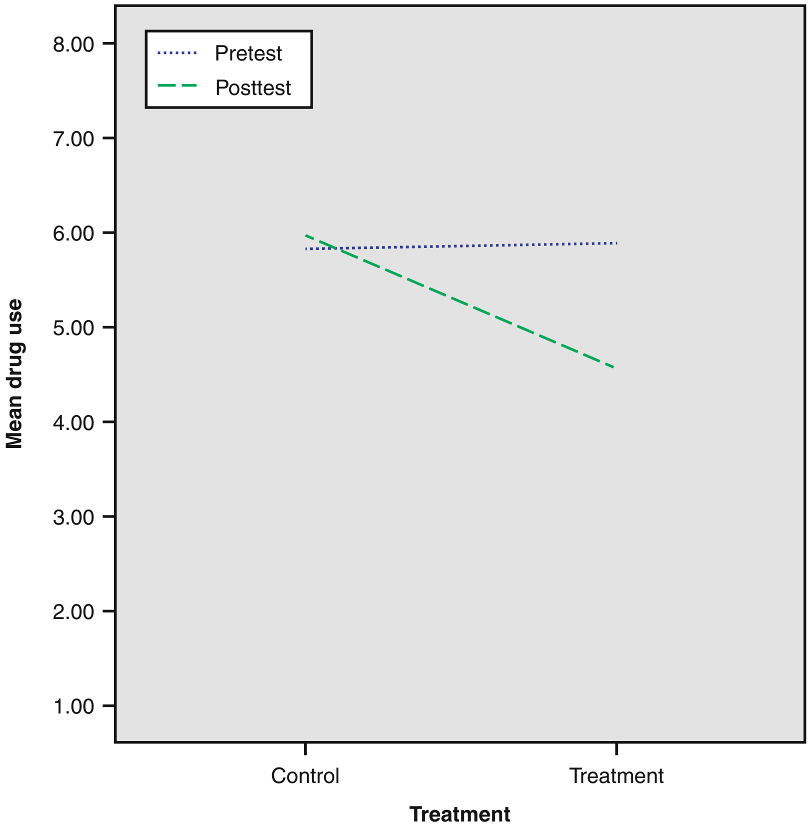

the treatment effect can be discovered because the intra-block variance is reduced. Because the "treatment estimate" is comprised of both the drug program and the variability, the treatment effect is likely to be larger within each block when the "noise" is reduced. It therefore becomes "easier" to see the effect, because the variance is reduced by the creation of homogeneous subgroups.

With this in mind, imagine another experiment, which is substantially similar to the one above: the treatment is delivered in the very same intensity and method as in simple random assignment, to a similar group of participants. However, the participants are grouped within "blocks of drug type," which creates blocks of participants who are likely to be more similar to one another, at least in terms of the type of drug to which they are addicted.

Notice that the *overall* 15 % reduction compared to the control group is still expected to be registered. This is depicted in Fig. 2. As can be seen, the overall treatment-to-control magnitude of difference is the same across blocks. However, the *intra*-block reduction is different, because there is less variance in the data. The blocking procedure shows that the Heroin Block is less receptive to the treatment (6 % decrease in drug use), as arguably expected from this subgroup, compared, for example, to the Marijuana Block or the MDMA Block (51 % and 34 % decrease, respectively). Given such findings, the policy implication could then be to continue using the program for certain types of addictions in light of the high success rate, but not for others with a lower success rate.

Planning Randomized Block Designs

When designing randomized block experiments, particular attention is drawn to three elements:

Randomized Block Designs, Fig. 2 Hypothetical drug treatment trial: randomized block design

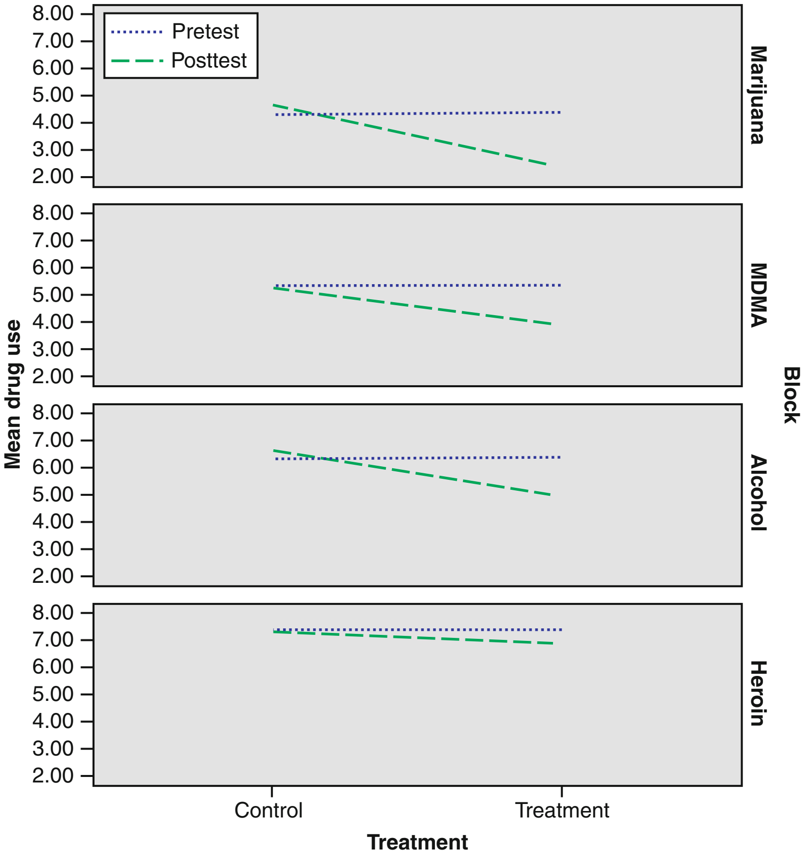

the treatment effect (and the variance associated with it), the blocking criterion (i.e., the variable which divides the sample prior to random assignment), and the interaction between the treatment factor and the blocking factor – though the latter is generally ignored in criminal justice research and therefore omitted from the discussion. There are different ways to "block" the study participants, depending on the type of study. The blocking criterion is usually based on the natural cutoff points of the outcome variable of interest, at its baseline level – for example, the number of police calls for service in hot-spot experiments (e.g., Ariel and Sherman forthcoming). The blocking criterion can be nominal or categorical as well, such as gender, ethnicity, or crime type.

The model can be extended to account for several treatments and several blocks. However, there are a limited number of interventions that one research project can study. It seems that using only one treatment and one control condition in each block is more effective, because the test statistic becomes more "stable" than with multiple conditions and blocking criteria. At the same time, it is not clear how many blocking factors should be allowed before the design becomes too "messy" – though Friedman et al. (1985, p. 69) understandingly claimed that, for studies of 100 participants, blocking using up to 3 factors is still considered manageable.

It is most important that each block should contain all treatments and that each treatment should occur an equal number of times in each block (Chow and Liu 2004, pp. 136–140). Many of the experiments on policing crime in places are constructed in this way, such as the ones reviewed in the introduction.

As in all experiments, it is desirable that the administration of the treatment in terms of potency, consistency, and procedures is identical in each block. This means that the overall experimental design is replicated as many times as there are blocks, and each block can be viewed as a disparate yet identical trial within the overall study (Rosenberger and Lachin 2002). If the administration of the treatment was different in each block, it would be rather difficult to analyze the treatment effect. Therefore, maintaining the same treatment across the study blocks is advised.

Lastly, randomized block designs were found to be particularly advantageous when there are "several hundred participants" (Friedman et al. 1985, p. 75), but recent studies suggest that randomized block designs can also be useful in much smaller studies, especially in place-based experiments, in which the unit of analysis is geographical areas rather than human participants (see review in Gill and Weisburd 2013).

Permuted Randomized Block Design

The design presented earlier reduces the variance in the data caused by certain prognostic or other general variables. However, these models are less useful to address time-related biases created by sequential assignment of units. There are times when the researcher needs to achieve near-equality in the number of units assigned in each block to different treatments, at any stage of the recruitment process, not only at the end of the study. A special type of block randomized design is required to achieve this.

First introduced by Hill (1951; see Armitage 2003), the "permuted randomized block design" (PRBD) was developed in order to deal with these conditions. The PRBD is most appropriate when there is a strong need for both periodic and end-of-the-experiment balance in the number of participants assigned to each study group in each block. This is the case, for example, in small studies, or studies with many small subgroups (Rosenberger and Lachin 2002, pp. 41–45).

Suppose that a care provider cannot wait for more victims of a similar type to surface, in order to provide them with much-needed services at the same time (for instance, group counseling). This is a small study, with only 40 participants, so it is important to have an equal assignment of cases to treatment and control conditions. Furthermore, the treatment must be delivered as soon as possible. However, the simple random assignment procedure is not optimal, because of the increased threats such as selection bias and chronological bias (Kao et al. 2007, p. 364). A PRBD model can be used instead. With 40 participants, the researcher can break the sample up into 10 blocks, with 4 participants in each block, and one treatment group (T) and one control group (C). In each block, there are exactly six possible arrangements of the treatment and control: CCTT, TTCC, CTTC, TCCT, TCTC, and CTCT. Each block is randomly allocated to one of the six arrangements. This process is then repeated 10 times (as the number of blocks), until all 40 participants are assigned to all conditions.

This procedure is relatively straightforward and provides several advantages. It "forces" group size equality by guaranteeing that there is exactly 50–50 % assignment of treatment and control participants. Furthermore, the procedure controls for time-related biases as well: using six combinations masks the next assignment sequence and therefore avoiding, at least somewhat, the risk of selection bias (but not without reservations as soon discuss). Many find it quite elegant as well, and over the years, the PRBD has become very popular in clinical trials (Abou-El-Fotouh 1976; Cochran and Cox 1957; Rosenberger and Lachin 2002, p. 154). However, in criminology, this approach is quite rare (Ariel 2009).

Matched Pairs Design

A different yet related random allocation procedure is the pairwise matching assignment (see Rossi et al. 2006). Matched pairs, in which each member of the pair is assigned to either treatment or control conditions, can also be used to achieve maximum equivalence. The researcher is able to gain more statistical power from reduced variance rather than increased sample size. In this procedure, participants are often rank-ordered based on a particular variable (e.g., the number of previous arrests), and then every pair is randomly assigned to treatment or control conditions. Thus, the top two offenders would be assigned at random, one to the control group and the other to the experimental group, and then the third and fourth participants would be assigned in a similar way, then the fifth and the sixth, and so on until the last two cases. Each pair can be viewed as a "block"; in the example above, there are 20 blocks.

Matched pairs assignments are appealing, as they produce exactly the same group sizes. However, they do pose some threats. First, any interference with the "natural" randomization of case allocation can potentially introduce bias. Particularly in sequential assignments (i.e., a trickle flow process), knowing the allocation of the first case ("treatment") immediately unmasks the assignment of the second case ("control"), and this can lead to selection bias. Second, by matching in pairs of ranked cases, one assumes that the "pairing criterion" (e.g., number of previous arrests) does not introduce a systematic bias to the procedure. Great caution should therefore be used when randomly assigning in pairs. Nevertheless, there can certainly be instances in which this design is a good fit. Hot-spot experiments can potentially benefit from this type of random allocation (e.g., Braga and Bond 2008).

A matched pairs design may also help to solve the problem of differential attrition. If one member of a matched pair drops out, the other member of the pair can be deleted as well, thus retaining the randomized design within matched pairs. For instance, in hot-spot experiments, it is conceivable that some "treatment hot spots" will have to be excluded *after* random assignment, if they are too far away from all other hot spots and difficult for the police to reach within reasonable time. In this case, the experimenter can drop the problematic "treatment hot spot," along with the paired "control hot spot," without jeopardizing the equivalence between the experimental and control groups. Nowadays, members of a pair can be matched on numerous variables by using propensity score matching (see, e.g., Apel and Sweeten 2010).

A Closer Look at the Blocking Criterion

Blocking on key variables is expected to create blocks of participants which are more similar to one another than in the generally heterogeneous random sample. In certain cases, it is immediately obvious that blocking should be utilized for the purposes of reducing experimental error resulting from variance. The advantage of blocking the data according to type of drug use, as in the hypothetical drug treatment trial presented above, is very clear. But there are other instances when the advantages of blocking are not as apparent.

Because the decision to block the data is usually based on qualitative grounds, there is always the risk that the researcher has made a poor decision. In an ordinal-level variable, for example, this means that the natural cutting points of the data might be misplaced, therefore implementing a grouping criterion that creates blocks in which the units within each block do not share common characteristics. In a study of tax evasion, for example, blocking the data according to income levels in a way that is intended to divide the sample into disparate socioeconomic backgrounds could go wrong if the researcher mistakenly categorized participants, so the blocks do not contain taxpayers with similar socioeconomic attributes. Thus, if the blocking is wrong, the block design will be disadvantageous compared with using an ordinary complete randomization design; such a procedure could actually be counterproductive and increase the error rate, because it increases the intra-block heterogeneity.

A practical approach to deal with this situation would be to ignore the blocking factor in analyses (Friedman et al. 1985). Disregarding the blocking effect is acceptable, because ignoring this factor should result in a more conservative statistical test. Analyzing the data this way will mean that the researcher sacrifices both power and precision, but not the overall integrity of the study, as long as the researcher implicitly reports this procedure.

Subgrouping Using Blocking or Subgroup Analyses

The last issue to consider in regard to these designs is whether pre-random blocking is at all necessary. Randomized block designs decrease the variance of the data and increase the precision and statistical power of the test. This is achieved by the pre-randomization blocking procedure. However, ordinary post hoc subgroup analyses, which are generally used in analysis of variance procedures (such as Tukey's HSD, Scheffe, Bonferroni, and the like), are used for exactly the same reason. These analyses allow the researcher to evaluate the treatment effect on particular groups or subgroups of participants. But in this context, they can also be used to theoretically homogenize the data according to certain key variables, much like a blocking procedure. Therefore, should data homogenization be dealt with before or after random assignment?

Subgroup analyses are customarily viewed as a natural step that comes after testing for main effects. These analyses can provide valuable information about both planned and unanticipated benefits and hazards of the intervention. At the very least, researchers use them to establish treatment benefits in subsets of participants. It makes sense to first assess the treatment by comparing all experimental units with all control units (or before–after, depending on the design) and then to account statistically for any covariates or other baseline variables. It seems that most clinicians agree with this rationale as 70 % of clinical trial reports include treatment outcome comparisons for participants subdivided by baseline characteristics at the post-randomization stage (Adams 1998; Assmann et al. 2000).

At the same time, however, as logical as subgroup analyses may be in theory, it is well established that they are often misleading. In fact, the medical community often rejects such findings (Moye and Deswal 2001), as the methods and procedures implemented are commonly misused (Assmann et al. 2000). Among some of the concerns raised against subgroup analyses, Moye and Deswal (2001) emphasize that "lack of prospective specification, inadequate sample size, inability to maintain power, and the cumulative effect of sampling error" complicate their interpretation. Some researchers go as far as saying that the most reliable estimate of the treatment effect for a particular subgroup is the overall effect rather than the observed effect in that particular group (Schulz and Grimes 2005). Therefore, using subgroup analysis instead of blocking the data before random assignment may actually be ill-advised.

Moreover, because virtually any covariate can be used to cluster units into subgroups, it allows, at least from a technical standpoint, the ability to generate multiple comparisons. This subgroup analysis is a source of concern, because it increases the probability of detecting differences simply by chance alone (type I error). It is not uncommon for researchers to be tempted to look for statistically significant, often publishable, differences between subgroups. It is therefore recommended that, without proper planning and sound rationale for conducting the analysis, subgroup analyses should not be considered as a replacement for pre-randomization blocking. Subgroup analyses should be justified on theoretical grounds a priori, in order to avoid the appearance of improper data mining. As described by Weisburd and Britt (2007, p. 320), "this is a bit like going fishing for a statistically significant result. However, sometimes one or another of the pairwise comparisons is of particular interest. Such an interest should be determined before you develop your analysis... if you do start off with a strong hypothesis for a pairwise comparison, it is acceptable to examine it, irrespective of the outcomes of the larger test. In such circumstances, it is also acceptable to use a simple two-sample t-test to examine group differences." Schulz and Grimes (2005, p. 1658) further developed this argument, by emphasizing that "seeking positive subgroup effects (data-dredging), in the absence of overall effects, could fuel much of this activity. If enough subgroups are tested, false-positive results will arise by chance alone.... Similarly, in a trial with a clear overall effect, subgroup testing can produce false-negative results due to chance and lack of power."

Thus, subgroup analyses can be misleading. The interpretation of their results should be seen as exploratory because they can suggest but not confirm a relationship in the population at large. Despite this recognition in other disciplines, most criminal justice experiments do not take multiple tests into account. It leads one to wonder how many of the findings from criminal justice experiments, in which hundreds of outcomes are reported, are really "true effects."

Notwithstanding reservations just reviewed, subgroup analyses should not be completely neglected. There are times when they can replace pre-randomization blocking designs. This, however, should be done with caution. When the researcher specifies at the beginning of the trial (i.e., in the experimental protocol before "going live" with experiment) that the efficacy of the treatment for a particular subgroup or a particular block is of particular interest and part of the research goals, then subgroup analyses can be considered. When this is the case, there are good strategies that can be used to estimate the effect size in subgroups created after random assignment (see Moye and Deswal 2001). Practically, reporting the proposed analyzes in the experimental protocol can substantially reduce the "going fishing" bias described earlier.

Conclusions

Under certain conditions, randomized block designs are more useful than simple random allocation procedures. They allow the researcher flexibility and control over the number of conditions assigned to the participants, as well as the number of blocks that are used to homogenize the data. By reducing the intra-block variance, the treatment estimates are more accurate because of the increased statistical power and precision of the test statistics.

The various statistical models designed to accommodate the different types of blocking techniques provide the researcher with accuracy that is generally superior to that which can be obtained using simple random allocation. This is particularly the case when the trial is small – several hundred or less – or when the blocking criterion is "good." These blocking procedures can deal with certain shortcomings that cannot be eliminated by randomization: increased data variance, outliers, time effects due to sequential assignment, missing data, and covariance imbalance. Blocking procedures can also create a more accurate estimate of the treatment effect – not by altering the effect of the treatment or the sample size, which are held constant, but by decreasing the variance. Therefore, given their clear advantages, these designs are gaining popularity in criminal justice research, and one should expect that they will be used more often in the future.

Related Entries

- ▶ History of Randomized Controlled Experiments in Criminal Justice
- ▶ Methodological Issues in Evaluating Police Performance
- ▶ Randomized Experiments in Criminology and Criminal Justice

Recommended Reading and References

Abou-El-Fotouh HA (1976) Relative efficiency of the randomized complete block design. Exp Agric 12:145–149

Adams K (1998) Post hoc subgroup analysis and the truth of a clinical trial. Am Heart J 136(5):753–758

Apel R, Sweeten G (2010) The impact of incarceration on employment during the transition to adulthood. Soc Probl 57:448–479

Ariel B (2009) Systematic review of baseline imbalances in randomized controlled trials in criminology. Presented at the communicating complex statistical evidence conference, University of Cambridge, UK

Ariel B (2012) Deterrence and moral persuasion effects on corporate tax compliance: findings from a randomized controlled trial. Criminology 50(1):27–69

Ariel B, Farrington D (2010) Randomised block designs. In: Weisburd D, Piquero A (eds) Handbook of quantitative criminology. Springer, New York

Armitage P (2003) Fisher, Bradford Hill, and randomization: a symposium. Int J Epidemiol 32:925–928
Assmann S, Pocock S, Enos L, Kasten L (2000) Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet 355(9209):1064–1069
Braga A, Bond B (2008) Policing crime and disorder hotspots: a randomized controlled trial. Criminology 46(3):577–607
Braga A, Weisburd D, Waring E, Mazerolle LG, Spelman W, Gajewski F (1999) Problem-oriented policing in violent crime places: a randomized controlled experiment. Criminology 37:541–580
Chow S-C, Liu J-P (2004) Design and analysis of clinical trials: concepts and methodologies. Wiley-IEEE, Taiwan
Cochran WG, Cox GM (1957) Experimental designs. Wiley, New York
Friedman LM, Furberg CD, DeMets DL (1985) Fundamentals in clinical trials, 2nd edn. PSG, Littleton
Gill CE and Weisburd D (2013) Increasing equivalence in small sample place-based experiments: taking advantage of block randomization methods. In: Welsh BC, Braga AA, Bruinsma GJN (eds) Experimental criminology: prospects for advancing science and public policy. Cambridge University Press, New York
Hallstrom A, Davis K (1988) Imbalance in treatment assignments in stratified blocked randomization. Control Clin Trials 9(4):375–382
Hill AB (1951) The clinical trial. Br Med Bull 7:278–282
Kao L, Tyson J, Blakely M, Lally K (2007) Clinical research methodology I: introduction to randomized trials. J Am Coll Surg 206(2):361–369
Lachin JM (1988) Statistical properties of simple randomization in clinical trials. Control Clin Trials 9(289):312–326
Lipsey MW (1990) Design sensitivity: statistical power for experimental research. Sage, Newbury Park
Moye L, Deswal A (2001) Trials within trials: confirmatory subgroup analyses in controlled clinical experiments. Control Clin Trials 22(6):605–619
Rosenberger W, Lachin JM (2002) Randomization in clinical trials: theory and practice. Wiley, New York
Rossi P, Lipsey M, Freeman H (2006) Evaluation: a systematic approach, 7th edn. Sage, Newbury Park
Schulz KF, Grimes D (2005) Multiplicity in randomized trials II: subgroup and interim analyses. Lancet 365(9471):1657–1661
Torgerson JD, Torgerson CJ (2003) Avoiding bias in randomized controlled trials in educational research. Br J Educ Stud 51(1):36–45
Ttofi MM, Farrington DP (2012) Bullying prevention programs: the importance of peer intervention, disciplinary methods and age variations. J Exper Criminol
Weisburd D, Britt C (2007) Statistics in criminal justice. Springer, New York
Weisburd D, Green L (1995) Policing drug hot spots: the Jersey City DMA experiment. Justice Q 12:711–736
Weisburd D, Taxman FS (2000) Developing a multicenter randomized trial in criminology: the case of HIDTA. J Quant Criminol 16(3):315–340
Weisburd D, Petrosino A, Mason G (1993) Design sensitivity in criminal justice experiments. Crime Justice 17(3):337–379
Weisburd D, Morris N, Ready J (2008) Risk-focused policing at places: an experimental evaluation. Justice Q 25(1):163–200

Randomized Control Trials

▶ Differences-in-Differences in Approaches

Randomized Controlled Trials

▶ Randomized Experiments in Criminology and Criminal Justice

Randomized Experiments in Criminology and Criminal Justice

David Weisburd[1,2], Anthony Petrosino[3] and Trevor Fronius[3]
[1]Department of Criminology, Law and Society, George Mason University, Fairfax, VA, USA
[2]Faculty of Law, The Hebrew University, Mt. Scopus, Jerusalem, Israel
[3]Learning Innovations, WestEd, Woburn, MA, USA

Synonyms

Controlled experiments; RCTS, Randomized controlled trials, Randomized field trials; True experiments

Overview

Randomized experiments are a type of experimental design that uses random assignment to gain equivalence between groups in a study. They are used most often to evaluate the impact of interventions on outcomes. This entry summarizes the rationale for a randomized experiment, provides the essential components of the design and an example from the policing literature, and briskly highlights its history and status as the “gold standard” for evaluation research. The entry concludes with a description of barriers to experiments and some of the problems in the field that threaten the design.

Background

Experimental research attempts to show a causal relationship of an independent variable (e.g., higher levels of patrol) on a dependent variable (e.g., arrest rates). In experimental research, the researcher uses various methodological approaches to control for the observed and unobserved factors that may influence the dependent variable. The major goal of experimental research is to strengthen internal validity, that is, ensuring that the observed impact is explicitly related to the changes in the independent variable (i.e., treatment condition). There are a wide variety of competing methods in experimental research used to strengthen internal validity, but randomized experiments are considered by many researchers to be the “gold standard.”

A randomized experiment is a type of design that uses random allocation of participants or other units of analysis in order to gain equivalence between subjects or units studied. Although randomized experiments may be used to test theoretical questions, they are most often used in criminal justice to evaluate the impact of interventions, programs, or policies. Randomized experiments that take place in the field, as is often the case in criminal justice, are also referred to as randomized field trials. They are also referred to as randomized control trials (RCT), true experiments, or randomized field experiments.

Although several quasi-experimental approaches are gaining popularity in the social sciences, including criminology (e.g., propensity score matching, regression discontinuity), there is consensus that randomized experiments are optimal for drawing causal inferences between independent and dependent variables. Randomized experiments, when implemented with good fidelity, permit researchers to explain observed findings as a result of the treatment or intervention, ruling out confounding factors that may explain findings.

Internal Validity

To better understand the strength of randomized experiments, a discussion of internal validity is warranted. A research design in which the impact of the intervention can be clearly distinguished from other observed factors is known as having high internal validity. If there are confounding factors involved in the impact of the intervention, then the experiment is considered as having low internal validity. Shadish et al. (2002), among others, have identified the most common threats to internal validity:

1. *Selection*: the preexisting differences between treatment and control subject or units.
2. *History*: an external event occurring at the same time of the study that may influence impact.
3. *Maturation*: changes in subjects or units between measurements of the dependent variable. These changes may be of natural evolution (e.g., aging) or due to time-specific incidences (e.g., fatigue, illness).
4. *Testing*: measurement at pretest impacts measurement at posttest.
5. *Instrumentation*: changes to the instrument or method of measurement in posttest measures.
6. *Regression to the mean*: Natural trends may cause extreme subjects or units who score extremely high or low during the pretest to score closer to the mean at posttest.
7. *Differential attrition*: the differential loss of subjects or units from the treatment group compared to the control group.
8. *Causal order*: the certainty that the intervention did in fact precede the outcome of interest (Farrington and Welsh 2005).

To further illustrate the importance of internal validity, let us suppose a researcher is interested in evaluating the impact of a youth court on a juvenile recidivism. The internal validity is considered high, if at the end of the evaluation, the researcher can show that the change in juvenile recidivism among the intervention groups is due only to the intervention (i.e., youth court) and no other confounding factors were at play. The researcher must show through either research design or analytical procedures that all confounding factors are accounted for in the measurement of outcomes. If the researcher is unable to account for other factors such as seriousness of first offense or the maturation of the study population, he or she must note that the observed effects may be due to other factors. If few threats to validity (or potential confounding factors) are not accounted for, the internal validity of the study would be considered low.

Generally speaking, a randomized experiment has the highest possible internal validity, because this approach allows the researcher to assume that other confounding causes of the outcome of interest, known and unknown, are not systematically influencing the study results. High internal validity in randomized experiments is gained through the process of randomly allocating the treatment or intervention to the experimental and control or comparison groups. Through random assignment, the researcher is not just randomizing the treatment. He or she is randomizing all other factors that may influence the outcome of the treatment. Thus, there is no systematic bias that increases the odds of one unit's assignment to the treatment group and another unit's assignment to the control or comparison group. This is not to imply that the groups are the same on every characteristic – it is very possible that differences may occur; however, these differences can be assumed to be randomly distributed and are accounted for in the probability distributions that underlie statistical tests of significance. Regardless, neither the treatment group nor the control group should have an advantage over the other on the basis of known or unknown variables. Thus, randomized experiments are the only design that allows the researcher to assume statistically unbiased effects (e.g., Boruch 1997).

The goal of most randomized experiments in criminology and criminal justice, as in other social science fields, is to disentangle the impact of the treatment or intervention from the impact of other factors on the outcomes that are to be tested. A randomized experiment allows the researcher to attribute differences between the groups from pretest to posttest as a result of the treatments or interventions that are applied. At the conclusion of the study, the researcher is able to assert, with confidence, that the difference is likely a result of the treatment and not due to other confounding factors. It is more difficult for nonrandomized studies, even a high-quality quasi-experimental design, to make this assertion. This advantage is underscored by Farrington (1983):

> The unique advantage of randomized experiments over other methods is high internal validity. There are many threats to internal validity that are eliminated in randomized experiments but are serious in non-experimental research. In particular, selection effects, owing to differences between the kinds of persons in one condition and those in another, are eliminated. (1983, p. 260)

Other Experimental Designs

Randomized experiments are distinguished from quasi-experiments and natural experiments by the random assignment of subjects or units. Indeed, these other methods have been developed as a way to approximate "true" experiments when random assignment is not possible. For example, researchers use matching, statistical methods, a comparison group, time series analyses, and other methods to create a quasi-experimental design (QED).

Researchers also take advantage of naturally occurring events to conduct natural experiments. Natural experiments often occur due to an unplanned event, which can allow for an evaluation of a particular factor or intervention. For example, when overcrowding forces the early release of certain prisoners, researchers can use this unplanned event to evaluate the impact of early release on recidivism. Although this example may show a direct short-term impact of early release on recidivism, it still suffers from potential for selection bias, as those selected for early release may be very different than prisoners remaining incarcerated.

Although QEDs and natural experiments are often necessary and can have high interval validity compared to other designs, they are generally considered less rigorous than a well-implemented randomized experiment, because they cannot control for unknown factors.

There is some debate among criminologists and across other fields in social and physical sciences, as to the type of bias that results from utilizing nonrandomized rather than randomized studies when examining program or intervention effects. Some may argue that the differences between well-designed nonrandomized studies and randomized experiments are negligible if not slightly skewed towards randomized experiments (Lipsey and Wilson 1993); however, it has been shown that in criminal justice nonrandomized experiments routinely overestimate program success when compared to randomized experiments (Weisburd et al. 2001).

Example of a Criminological Experiment

The general structure of experiments in criminology is usually similar in design regardless of the area or question of interest. Generally, experiments in criminology start with an eligibility pool, randomization, group allocation, and post-test measures relevant to the dependent variable of interest (Fig. 1).

The eligibility pool is made up of those participants or units that are eligible for the experiment. Units of analysis can be individuals or aggregated groups or other entities that are often found in clusters. For example, in an experiment that evaluates the impact of increased foot patrol on crime rates, the eligibility pool may be individual officers who "walk the beat," in a specific area, and who have a specific number of years of experience. Or the unit of analysis could be the geographic area or "beat" that will be assigned to different conditions. The eligibility pool is thus comprised of those patrol officers (or patrol "beats") that meet criteria for inclusion in the study.

Next, researchers randomly assign members from this pool of eligible participants or units to the study conditions – often a treatment group and a control or comparison group. Historically, randomization was carried out through a simple coin flip. There are now many ways to randomize subjects, but most often researchers rely on computerized statistical software to carry out randomization. Some researchers may simply use the rule of odds and evens – that is, assigning every other case to one particular group. This is often referred to as alternation and is considered quasi-random assignment as the assignment of the numbers used is not actually random. The most critical factor in randomization, however, is that each case has the same probability or likelihood of being selected for the control group as the experimental group and that the assignment is based purely on chance. Garner (1977) argued that "equal probability assignment" should replace "random assignment" when discussing such studies with practitioners as it sounds orderly and fair rather than chaotic and unfair.

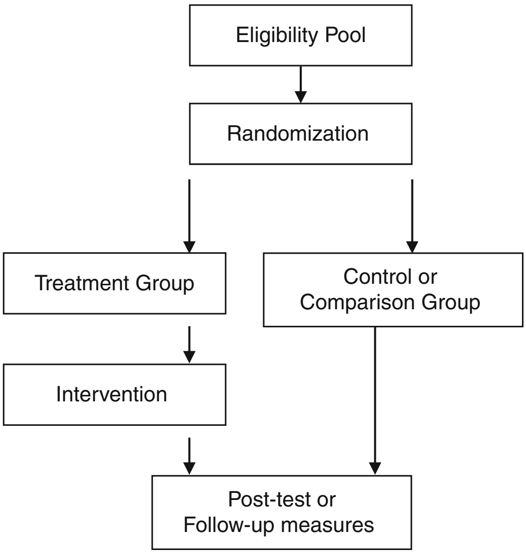

Randomized Experiments in Criminology and Criminal Justice, Fig. 1 Basic Experimental Design

In the usual criminological experiment, eligible cases are randomly assigned to one of two groups – treatment or control. Experiments in criminology may have more than two groups. But typically, an experiment is comprised of a group that receives the treatment or intervention and a control or comparison group that does not.

It is also quite common in a criminological experiment for the control group to actually receive something rather than nothing. For example, in the foot patrol example, the control group may receive treatment as usual, or the same number of foot patrol officers as typically employed.

Experimental designs can include any number of outcomes, from one to scores. If the randomization was implemented with fidelity, it should produce two equivalent groups on the pretest or baseline measures related to the outcome of interest. The researcher then conducts analyses to determine if the intervention had any impact on the posttest or follow-up measures of the outcomes of interest.

An Illustrative Example of a Place-Based Randomized Field Study: The Minneapolis Hot Spots Patrol Experiment

Randomized experiments in criminal justice, as with most fields, involve the random assignment of individuals. However, as shown in the design example above, it is possible to randomized larger units such as schools, police beats, transportation stops, or city blocks. When these larger entities are randomized, it is often referred to as cluster, block, or place-based randomized trials. The Minneapolis Hot Spots Patrol Experiment (Sherman and Weisburd 1995) provides an example of a place-based experimental design. In that study, the investigators wanted to determine in the least equivocal manner whether intensive police patrol in clusters of addresses that produced significant numbers of calls for service to police, also known as "hot spots," would reduce the levels of crime calls or observed disorder at those places.

To conduct the experiment, the investigators established a pool of 110 hot spots that were eligible for random assignment. This pool of eligible hot spots was then randomly assigned to the innovation group receiving intensive police patrol and a control group that received normal police services. In the Minneapolis study the investigators randomly assigned the hot spots within statistical blocks, or groups of cases that are alike one to another on characteristics measured before the experiment. Such block randomization is not required in randomized experiments, but as noted above is a method for further ensuring the equivalence of treatment and comparison or control groups.

The strength of the randomized experiment, if it is implemented with full integrity, is that it now permits the investigators to make a direct causal claim that the innovation was responsible for the observed results. If the researchers had simply looked at changes over time at the treated hot spots, it would have been difficult to claim that the observed changes were due to treatment rather than a natural change in levels of crime or disorder over time. Similarly, if the researchers had tried to statistically control for other potential causes in order to isolate out the effect of treatment, it could be argued that some unknown or unmeasured factor was not taken into account. In the Minneapolis study the researchers found a statistically significant difference between the treatment and control hot spots. Random allocation of subjects to treatment and control conditions allowed the authors to assume that other factors had not systematically confounded their results.

History of Randomized Experiments in Criminology and Criminal Justice

While there is evidence that randomization in experiments was used as early as 1885 (Hall 2007), most scholars credit Sir Ronald Fisher and his publication on *The Design of Experiments* (1935) as the first discussion that argues for randomization as a necessity to establish group equivalence. The medical world was the first field that used randomized experiments extensively in the field. The first randomized experiment is credited to Austin B. Hill and associates in a 1948 paper on streptomycin treatment of pulmonary tuberculosis. The work done by Hill during this time is still highly celebrated in the medical field as it paved the way for the first half century of randomized control trials (Weisburd and Petrosino 2004).

Randomized trials in criminology may have started as early as 1937 with the Cambridge-Somerville Youth Study, but the publication was not issued until 1951 (Powers and Witmer 1951) – this is likely the first criminological trial in modern times. Researchers created a youth eligibility pool from youth who were considered as "troubled kids" by their teachers and police, and matched the youth on basic characteristics. Next in the experiment, researchers randomly assigned the youth to a treatment counseling group or a no-treatment control group. The findings consistently stated no positive gains and even negative gains for the treatment group (Weisburd and Petrosino 2005).

Although Fisher and Hill generated much excitement around randomized trials in their respective fields, criminological trials were sparse in the first decade since its debut in the field. Experiments in criminology only found momentum after Campbell and Stanley (1963) published their comprehensive guide to *Experimental and Quasi-Experimental Designs for Research*. The publication provided researchers with a design template to carry out high-quality criminological research. Even then the randomized trial was touted as champion in the hierarchy of experimental design for its high degree of internal validity (Weisburd and Petrosino 2004). Randomized trials during President Lyndon Johnson's campaign – perhaps spurred on by the credibility seen throughout the medical world – were used frequently to evaluate the social endeavors designed in the "Great Society" campaign. Soon after this federal adoption, randomized trials fell out of favor with the administration as researchers would consistently report that government programs show no impact or even negative outcomes. Despite Campbell's (1969) decree that

> The United States and other modern nations should be ready for an experimental approach to social reform, an approach in which we try out new programs designed to cure specific social problems, in which we learn whether or not these programs are effective, and in which we retain, imitate, modify, or discard them on the basis of apparent effectiveness on the multiple imperfect criteria available (Campbell 1969, p. 409)

randomized experiments in criminology faded from center stage for more than a decade (Weisburd and Petrosino 2004). David Farrington's (1983) review of high-quality randomized experiments in criminal justice showed that between 1957 and 1981 there were only 41 eligible studies, of which only 35 included offending outcomes. Probably more important than the fact that only 30 % of the studies found a significant effect is the relative dearth of experiments in criminology compared to other social and physical sciences (Farrington 1983).

However, experimentation in criminology received a boost during the 1980s. First, the Minneapolis Domestic Violence Experiment (Sherman and Berk 1984), in which 330 male domestic violence suspects were assigned to the treatment (arrest on the spot) group or one of two comparison groups – mediation or separation for 8 h – was conducted. A 6-month follow-up found that arrests significantly reduced subsequent domestic violence, and the study received wide media attention and led to police departments across the USA to adopt mandatory arrest policy for misdemeanor domestic violence incidents. These results also led the US Department of Justice, under James Stewart, to support randomized experiments (Farrington and Welsh 2005; Weisburd and Petrosino 2004). Although replications in other cities were conducted with mixed results, this study is still considered a monumental factor in the promotion of randomized experiments in criminology and criminal justice (Farrington and Welsh 2005). In fact, the exposure from the Minnesota study and subsequent works led to NIJ funding more than two dozen experiments over the next several years (Weisburd and Petrosino 2004).

It was just 2 years later that Farrington et al. (1986) released their acclaimed work *Understanding and Controlling Crime*, which emphasized the unequivocal need to use randomized experiments whenever possible in criminological evaluations. The investment in experimental criminology continued in waves through the 1990's and has increased considerably into the twenty-first century.

Experiments as the Gold Standard in Criminology and Criminal Justice

There is a strong consensus that experiments are the gold standard for evaluation research in criminology and criminal justice. This is reinforced by seminal reviews that appraise the methodological quality of evaluation evidence. For example, Sherman et al. (1997) conducted a comprehensive review of evaluations across many criminal justice sectors to produce their report to congress on crime prevention entitled, "What Works, What Doesn't, What's Promising." As part of the review, Sherman and colleagues developed the Maryland Scientific Methods Scale (SMS), a set of criteria for scoring – evaluations, largely based on internal validity. The highest rating, "5," is reserved for evaluations that utilize "random assignment of program and control conditions to units" (Farrington 2003).

Another example of the rising status of experiments is the creation of organizations to promote such studies. For example, the Academy of Experimental Criminology was founded in 1998 to recognize researchers who successfully led randomized field experiments (Weisburd et al. 2007). The American Society of Criminology created a Division on Experimental Criminology (DEC).

It is also the case that the use of experiments in justice field settings is increasing. As a follow-up to earlier work, Farrington and Welsh's (2005) review showed that randomized experiments with offending outcomes that fit their review criteria grew from 35 between 1957 and 1981 to 1983 between 1982 and 2004 – an increase of over 200 %.

Barriers to Randomized Experimentation

Although randomized experiments have been widely recognized for their internal validity rigor for over 50 years, they comprise a comparatively small portion of the evaluation portfolio in criminology. For example, Petrosino (2003) examined abstracts of impact studies relevant to childhood intervention in the leading bibliographic databases for each field. The proportion of abstracts to RCTs or possible RCTs ranged from 17 % to 18 % for juvenile justice and K-12 education to 68 % in health care. And despite the increase, Farrington and Welsh (2005) were still discouraged to find as few RCTs as they did in their survey of the literature.

Why do randomized experiments encompass a small percentage in the overall catalog of research when they are widely regarded as the gold standard in research? Farrington and Welsh (2005) note that although randomized experiments are advantageous, there are "many ethical, legal, and practical difficulties that face researchers who wish to mount a randomized experiment." For one, some argue that it is legally irresponsible to provide treatment or judicial sanctions at random or to withhold necessary treatment from individuals. Boruch (1997) and others have countered this argument by noting that it is unethical to provide untested treatments and that programs are sometimes provided to persons haphazardly without evidence of their effectiveness.

Another barrier to RCTs is that randomized trials are considered too expensive. The feature of the RCT that generates the highest costs is actually not specific to randomized experiments: it is the longitudinal nature of the data collection (i.e., following treatment and control units and collecting data on various outcome measures). The Coalition for Evidence-Based Policy (2012) has argued for using large-scale administrative data for outcomes (as opposed to field data collection from surveys or interviews) to decrease costs in data collection.

Another common criticism of randomized experiments is that it emphasizes internal validity, but at the expense of external validity. External validity is the ability to generalize findings from a study to other settings. Some researchers argue that experiments set up an artificial setting with treatment and control conditions that do not approximate real world conditions.

Problems in the Field

The three major issues that threaten experimental findings of randomized experiments are as follows:

1. If the random assignment protocol was violated in substantive ways
2. If there was substantial attrition or differential loss of subjects from the study
3. If there was no clear contrast between the treatment program and control condition

Randomization integrity. Randomization is the cornerstone of true experiments. There are a number of missteps that may adversely affect the integrity of random assignments. For example, program staff may have a vested interest in manipulating the randomization assuring certain participants are funneled into certain groups. It has been shown that this type of violation occurs more frequently with practitioners than with researchers; thus, whenever possible, researchers should control the randomization and practitioners should remain blind to the group allocation. Furthermore, researchers should determine the eligibility pool with practitioners before randomization occurs, monitor assignment throughout the study, and train practitioners and data collection staff in the importance of research integrity (Petrosino 2005).

Attrition. Attrition is the loss of cases from study groups after randomization. Researchers must also be proactive in retention efforts to ensure minimal attrition throughout the study. It is very difficult, especially in the criminal justice field, to track participants for long periods of time, which is required to assess intermediate and long-term treatment impact. When participants attrite for one study group more so than another, it is called differential attrition. Differential attrition is particularly hazardous when determining confidence in the treatment impact, because often this causes a change in the composition of the groups. This is a threat to the experiment's internal validity when the treatment group differs in some important way from those who leave the study (Petrosino 2005).

Treatment-control group contrast. The intent when conducting an experiment is to test a specific treatment's impact on a given outcome variable relative to the comparison condition. To test the impact, the treatment and comparison must indeed be distinguishable from one another. On occasion, when the treatment is only a slight alteration from the comparison condition, it may only take a small change in intensity of treatment or comparison condition to create indistinguishable groups. This difference in conditions must be maintained; otherwise, the comparison and any results would be meaningless (Weisburd and Petrosino 2004).

Conclusion

Despite barriers to experiments and the hurdles researchers must jump to conduct a well-implemented RCT, experimental criminology has grown exponentially over the past several decades. There are signs that this trend will continue in criminology. First, the publicity from successful medical field trials has led to greater acceptance of this research in other fields. Second, tighter governmental accountability has created a surge in evidence-based research. The public expects the programs funded by tax dollars to be rigorously tested (Weisburd and Petrosino 2005).

Related Entries

- ▶ Cambridge-Somerville Youth Experiment
- ▶ Crime Science
- ▶ History of Randomized Controlled Experiments in Criminal Justice
- ▶ Interrupted Time Series Models
- ▶ Multisite Trials in Criminal Justice Settings
- ▶ Place-Based Randomized Trials
- ▶ Quantile Regression Models to Analyze Experimental Data
- ▶ Randomized Block Designs
- ▶ Recent Perspectives on the Regression Discontinuity Design

Recommended Reading and References

Boruch RF (1997) Randomized experiments for planning and evaluation: a practical guide. Sage, Thousand Oaks

Braga AA, Weisburd DL, Waring EJ, Mazerolle LG, Spelman W, Gajewski F (1999) Problem-oriented policing in violent crime places: a randomized controlled experiment. Criminology 37:541–580

Campbell DT (1969) Reforms as experiments. Am Psychol 24:409–429

Campbell DT, Stanley J (1963) Experimental and quasi-experimental designs for research. Houghton-Mifflin, Boston

Coalition for Evidence-based Policy (2012) Rigorous program evaluations on a budget: how low-cost randomized controlled trials are possible in many areas of social policy. Author, Washington, DC

Farrington DP (1983) Randomized experiments on crime and justice. In: Torny M, Morris N (eds) Crime and justice: an annual review of research, vol 4. University of Chicago Press, Chicago, pp 257–308

Farrington DP (2003) Methodological quality standards for evaluation research. Ann Am Acad Polit Soc Sci 587:49–68

Farrington D, Welsh B (2005) Randomized experiments in criminology: what have we learned in the last two decades? J Exp Criminol 1:9–38

Farrington DP, Ohlin LE, Wilson JQ (1986) Understanding and controlling crime. Springer, New York

Fisher RA (1935) The design of experiments. Oliver and Boyd, London

Garner J (1977) Role of congress in program evaluation – three examples in criminal justice. Prison J 57(l):3–12

Hall N (2007) R.A. Fisher and his advocacy of randomization. J Hist Biol 40(2):295–325

Lipsey MW, Wilson DB (1993) The efficacy of psychological, educational, and behavioral treatment: confirmation from meta-analysis. Am Psychol 48:1181–1209

Petrosino A (2003) Estimates of randomized field trials across six areas of childhood intervention. Ann Am Acad Polit Soc Sci 589:190–202, Special Issue on Randomized Experiments in the Social Sciences, September

Petrosino A (2005) Randomized field trials in the U.S. to evaluate justice programs: a briefing. Japan J Sociol Criminol 30:56–71

Petrosino AJ, Boruch RF, Rounding C, McDonald S, Chalmers I (2000) The Campbell Collaboration Social, Psychological, Educational and Criminological Trials Register (C2-SPECTR) to facilitate the preparation and maintenance of systematic reviews of social and educational interventions. Eval Res Educ 14(3/4):206–219

Powers E, Witmer H (1951) An experiment in the prevention of delinquency: the Cambridge-Somerville youth study. Columbia University Press, New York

Shadish WR, Cook TD, Campbell DT (2002) Experimental and quasi-experimental designs for generalized causal inference. Houghton-Mifflin, Boston

Sherman LW, Berk RA (1984) The specific deterrent effects of arrest for domestic assault. Am Sociol Rev 49(1):261–272

Sherman L, Weisburd D (1995) General deterrent effects of police patrol in crime 'hot spots': a randomized study. Justice Q 12(4):625–648

Sherman LW, Gottfredson D, MacKenzie D, Eck J, Reuter P, Bushway S (1997) Preventing crime: what works, what doesn't, what's promising. A report to the United States Congress. University of Maryland, Department of Criminology and Criminal Justice, College Park

Weisburd D, Mazzerolle L, Petrosino A (2007) The academy of experimental criminology: advancing randomized trials in crime and justice. Criminologist 32:1, 3–7

Weisburd D, Petrosino A (2004) Experiments, criminology. In: Kempf-Leonard K (ed) Encyclopedia of social measurement. Academic, San Diego

Weisburd D, Petrosino A (2005) Randomized experiments. In: Miller JM, Wright R (eds) Encyclopedia of criminology. Routledge, New York

Weisburd D, Lum C, Petrosino A (2001) Does research design affect study outcomes in criminal justice? Ann Am Acad Polit Soc Res 578:50–70

Randomized Field Trials

► Randomized Experiments in Criminology and Criminal Justice

Randomized Response Methods

Peter G.M. van der Heijden[1,2], Maarten Cruyff[1] and Ulf Böckenholt[3]
[1]Department of Methodology and Statistics, Utrecht University, Utrecht, The Netherlands
[2]University of Southampton, Southampton, UK
[3]Northwestern University, Evanston, IL, USA

Synonyms

Sensitive questions

Overview

In survey research it is often problematic to ask people sensitive questions because they may refuse to answer or they may provide a socially desirable answer that does not reveal their true status on the sensitive question. To solve this problem Warner (1965) proposed randomized response (RR). Here a chance mechanism hides why respondents say "yes" or "no" to the question being asked.

Thus far RR has been mainly used in research to estimate the prevalence of sensitive characteristics. RR can be used in, for example, self-report studies on criminal behavior (compare, e.g., the two meta-analyses in Lensvelt-Mulders et al.2005). In the Netherlands, much research has been funded by the Expert Centre for Law Enforcement of the Ministry of Justice in the area of regulatory noncompliance in areas as the taxi-driving law, under regulations for food retailers, the law of individual rent subsidy, and the law on agricultural chemicals. Elffers et al. (2003) give an overview of early research in this area. This line of research has been fine-tuned in research funded by the Dutch Ministry of Social Affairs and Employment on regulatory noncompliance among receivers of benefits (compare Lensvelt-Mulders et al. 2006, for a "best practice"). Typically, in this line of research the survey consists of RR questions on sensitive behavior and questions that may explain the psychological and background characteristics of respondents who are more likely to indulge in this behavior. Elffers et al. provide a theoretical background for the development of the psychological characteristics that they coin the table of eleven, as there are 11 explanatory areas.

It is not uncommon that researchers wrongly believe that the RR procedure has the drawback that it is not possible to relate the sensitive characteristics to explanatory variables. However, statistical tools have been developed to relate the sensitive variables to explanatory variables. Univariate RR data can be analyzed with a version of logistic regression that is adapted so that it can handle data collected by RR. Multivariate dichotomous RR data can be handled using a model from item response theory that assumes a latent variable that explains the answers on the RR variables. This entry ends with a discussion of a recent development in the analysis of multivariate RR data, namely, models that take into account that there may be respondents that do not follow the instructions of the RR design by answering no whatever the sensitive question asked.

Introduction

Is it possible to measure sensitive behavior such as noncompliance with rules and regulations that govern public life using surveys? Because it is well known that questions about compliance behavior with rules and regulations may not yield truthful responses, the randomized response (RR) method has been proposed as a survey tool to obtain more honest answers to sensitive questions (Warner 1965). In the original RR approach, respondents were provided with two statements, A and B, with statement A being the complement of statement B. For example, statement A is "I used hard drugs last year" and statement B is "I did not use hard drugs last year." A randomizing device, for instance, in the form of a pair of dice determines whether statement A or B is to be answered. The interviewer records the answer "yes" or "no" without knowing the outcome of the randomizing response device. Thus, the interviewee's privacy is protected but it is still possible to calculate the probability that the sensitive question (A and not B) is answered positively.

Meta-analyses have shown that RR methods can outperform more direct ways of asking sensitive questions (Lensvelt-Mulders et al. 2005). Importantly, the relative improvements in validity increased with the sensitivity of the topic under investigation. This entry of the encyclopedia will show recent developments of in the analysis of RR data, focusing on regression models relating the sensitive question(s) measured with RR to explanatory variables, and models that take into account that there may be respondents that do not follow the instructions of the RR design by answering no whatever the sensitive question asked.

In this manuscript the examples stem from surveys conducted for the Dutch government on noncompliance in social welfare (Lensvelt-Mulders et al. 2006). In the Netherlands Dutch employees must be insured under the Sickness Benefit Act, the Unemployment Insurance Act, the Health Insurance Act, and the Invalidity Insurance Act. Under each of these acts, an (previously) employed person is eligible for financial benefits provided certain conditions are met.

Most of the examples focus on six RR questions, four of which are health-related and the remaining two are work-related. The health questions are:

1. Have you been told by your physician about a reduction in your disability symptoms without reporting this improvement to your social welfare agency?
2. On your last spot-check by the social welfare agency, did you pretend to be in poorer health than you actually were?
3. Have you noticed personally any recovery from your disability complaints without reporting it to the social welfare agency?
4. Have you felt for some time now to be substantially stronger and healthier and able to work more hours, without reporting any improvement to the social welfare agency?

The work-related questions are:

1. In the last 12 months have you moonlighted while receiving your IIA benefits?
2. In the last 12 months have you taken on a small job alone or together with your friends that you got paid for without informing the social welfare agency?

The remainder of the entry is structured as follows. First the analysis of RR data without explanatory variables is discussed. Next, logistic regression adjusted for univariate RR data is described. After this the analysis of multivariate RR data is discussed. This section discusses how regression approaches are to be adjusted in the light of new models that take into account the fact that part of the sample may not be following the RR instructions laid out by the researcher.

Models for Randomized Response Data

Univariate RR Data, No Explanatory Variables

This section discusses the analysis of univariate RR data, where no explanatory variables are involved. The forced response design (Boruch 1971) is used as an example of an RR design.

Assume that the sensitive question asks for a "yes" or "no" answer. The forced response design is as follows. After the sensitive question is asked, the respondent throws two dice and keeps the outcome hidden from the interviewer. If the outcome is 2, 3, or 4, the respondent answers yes. If the outcome is 5, 6, 7, 8, 9, or 10, the respondent answers according to the truth. If the outcome is 11 or 12, the respondent answers no. Let Y be the latent binary RR variable that denotes the true status on the sensitive item, and let Y^* be the observed RR variable that denotes the observed answer on the randomized sensitive question, with "yes" = 1 and "no" = 2. Then

$$\begin{aligned} P(Y^* = 1) &= P(Y^* = 1|Y = 2)P(Y = 2) \\ &\quad + P(Y^* = 1|Y = 1)P(Y = 1) \\ &= 1/6 + 3/4P(Y = 1) \end{aligned} \tag{1}$$

By writing $P(Y^* = j) = c + dP(Y = 1)$, other designs can be described in the same way (compare Böckenholt and van der Heijden 2007). As a first example, in the original Warner design an individual is allocated to either statement A, with probability p, or to statement B, with probability $1 - p$, with $p \neq .5$. The sensitive question in group A is "do you belong to the sensitive population?" and in group B "do you not belong to the sensitive population?" Now $c = (1 - p)$ and $d = (2p - 1)$. A second example is the Kuk (1990) design (see van der Heijden et al. 2000, for an application). Here a (indirect) "yes" or "no" question is asked using two decks of cards with red and black cards. In the right deck the proportion of red cards is 8/10, while in the left deck it is 2/10. The respondent is asked to draw one card from each deck and keep the color of the cards hidden from the interviewer. Next, the question is asked. Instead of answering the question directly, the respondent names the color of the card he took from the related deck, that is, when the answer is "yes," the respondent names the color of the card he took from the right stack (thus, red is associated with "yes"), and when the answer is "no," he names the color of the card from the left stack (thus, black is associated with "no"). In this case we have $c = 1/5$ and $d = 3/5$.

The estimation of $P(Y = 1)$ in Eq. 1 is straightforward when the parameters are in the interior of the parameter space. If the estimate of $P(Y = 1)$ is between 0 and 1, then solving Eq. 1 provides so-called unbiased moment estimates. These moment estimates are equal to the maximum likelihood estimates (MLEs). It is however possible that the moment estimates are outside the unit interval, for example, when the observed proportion in the data is lower than 1/6 (chance level). In that case the MLEs are set to 0. See van der Hout and van der Heijden, 2002, for a detailed discussion.

Extended introductions can be found in Fox and Tracy (1986), Chaudhuri and Mukerjee (1988), and Chaudhuri (2011).

Example

In the 2002 survey 1,760 recipients of disability benefits were asked the six RR questions as presented above using the Kuk method. The respective point estimates with 95 % bootstrap confidence intervals for the four health items are .03 (.004–.052), .04 (.012–.061), .07 (.046–.098), and .13 (.102–.156). For the two work items the respective estimates are .16 (.128–.184) and .08 (.050–.103).

Univariate RR Data and Explanatory Variables

In logistic regression the dependent variable is predicted from one or more covariates. The logistic regression model for RR data is given by

$$\pi_{1i}* = c + d\frac{1}{1 + \exp(-\boldsymbol{x}_i^t \beta)} \qquad (2)$$

$\pi_{1i}*$ is the probability that individual i has score $Y = 1$, $\boldsymbol{x}_i$ the vector with the covariate scores of individual i, and $\boldsymbol{\beta}$ the corresponding vector with regression parameters.

The outcome of the RR logistic regression procedure can be interpreted in the usual way (see, e.g., Agresti 2002). The β-parameters reveal the relation between the covariates and the probability of a sensitive true status, and the significance of each parameter can be assessed using Wald tests. Chi-square tests can be used to assess the fit of the model if the observed RR data are counts that allow comparison with fitted values under the RR logistic regression model, and differences between models can be tested using likelihood ratio chi-square tests.

An early reference to logistic regression for RR data is Maddala (1983, pp. 54–56), and an elaborate discussion can be found in Scheers and Dayton (1988). Lensvelt-Mulders et al. (2006) extend the logistic regression procedure so that it can incorporate person weights that make it possible to weight a sample toward population characteristics, if known.

Example

As an illustration an example taken from Lensvelt-Mulders et al. (2006) is reported. The dependent variable is the work item “In the last 12 months have you taken on a small job alone or together with your friends that you got paid for without informing the social welfare agency?” The explanatory variable is “I think it is more beneficial for me not to follow the rules connected to my disability insurance benefit,” abbreviated as “benefit,” that is measured on a five-point scale and has a mean of 3.67 and a standard deviation of.777. Many more explanatory variables have been measured in this survey, motivated from a rational choice framework; see Lensvelt-Mulders et al. (2006) and Elffers et al. (2003) for details.

The logistic regression model logit (noncompliance) = constant + b * benefit has estimates.765 for the constant and.751 for b. In order to study the impact of these estimates, the estimated probability of noncompliance for the mean value of benefit (i.e., 3.67) and the mean plus or minus one standard deviation (i.e., 3.67 +.78 = 4.45 and 3.67 −.78 = 2.89) are compared. For the mean value of benefit, the estimated probability of noncompliance is 12 %, for 4.45 the estimated probability is 20 %, and for 2.89 the estimated probability is 7 %. This shows that benefit has a strong relation with the decision not to comply with the above disability insurance benefit regulation: the more profitable people perceive that noncompliance is, the more often they do not comply with this work regulation.

Multivariate RR Data

The data from multiple RR items can also be modeled with item response models. The simplest version of this class of models is the

Rasch model (Rasch 1980). For RR data, this model is given by

$$\pi_{1i}* = c + d\frac{1}{1+\exp(-\theta_i - \gamma_j)}, \qquad (3)$$

(see Böckenholt and van der Heijden 2004, 2007; Fox 2005). In this model, the parameter γ_j captures the sensitivity of item j, and the parameter θ_i captures the propensity of the individual i to comply with social security regulations. Since the precision in estimating individual differences in rule compliance is a function of the number of RR items per respondent, more precise measures of compliance can be obtained in multiple- than in single-item studies for equal sample sizes.

Second, mixture versions of the IRR models are developed to allow for respondents who do not follow the RR instructions. Thus, one mixture component consists of respondents who answer RR items by following the RR design and the other component consists of respondents who do not follow the RR design by saying "no" to each RR item, irrespective of the outcome of the randomizing device. These are called "self-protective" (SP) responses. By allowing for the possibility that not all respondents may follow the RR instructions, substantially higher estimates of noncompliance are found than the estimates obtained with current RR methods.

Example

Table 1 reports the goodness-of-fit statistics obtained from fitting the IRR models to the work and health items for the 2002 and 2004 surveys. Fit statistics obtained when considering the health items only are also included. Because a minimum of three items is needed to identify an IRR model without covariates, IRR models are not estimated separately for two work items. The first set of fitted models is based on the baseline RR assumptions represented by Eq. 1 and serves as a benchmark for the (mixture-) IRR models. The second part of Table 1 is obtained by fitting a Rasch version of Eq. 3 to the health item set and by fitting a multidimensional Rasch version of Eq. 3 to both item sets simultaneously.

Randomized Response Methods, Table 1 Fit statistics of RR models for work and health items

Survey year	Health items G2 (*df*)	Health and work items G2 (*df*)
1. Homogeneous compliance		
2002	124.0 (11)	282.4 (57)
2004	56.2 (11)	184.4 (57)
2. Heterogeneous compliance		
2002	39.0 (10)	100.6 (54)
2004	23.8 (10)	95.6 (54)
3. Heterogeneous compliance and SP "no" sayers		
2002	14.9 (9)	54.2 (53)
2004	10.8 (9)	63.5 (53)

The homogeneous-compliance models require the estimation of two- and six-item location parameters when applied to the four health items and the four health and two work items, respectively. None of the reported fits are satisfactory, indicating that the assumption of no individual differences does not agree with the data. This result is supported by the fit improvement obtained from the IRR models that allow for heterogenous compliance behavior without requiring item-specific discrimination parameters. With one additional parameter for the health items and three additional parameters for the bivariate covariance matrix of the health and work items, these IRR models imply major fit improvements. However, despite the better fit, these models still do not describe the data satisfactorily.

As shown by a residual analysis of the data, the main reason for the misfit is that the outcome of self-protective responses (i.e., consistent "no" responses to the items) is greatly underestimated by these models. Thus, more respondents than expected under the IRR models give exclusively "no" responses when asked questions about their compliance with the health and work regulations. A mixture component can address the problem of extra "no" responses. The models' parsimonious representation appears to be in good agreement with the 2002 and 2004 survey data as indicated by the fit statistics in Table 1.

Most importantly, the mixture-IRR models provide more accurate estimates of the noncompliance rate in the population of interest

Randomized Response Methods, Table 2 Noncompliance estimates and 95 % bootstrap confidence intervals

Domain	Item	No SP correction	SP correction
Health	1	.002 (.000, .015)	.033 (.010, .050)
	2	.014 (.000, .034)	.053 (.033, .075)
	3	.048 (.027, .070)	.083 (.055, .112)
	4	.085 (.063, .107)	.130 (.087, .159)
Work	1	.030 (.009, .052)	.074 (.050, .096)
	2	.110 (.086, .133)	.159 (.104, .190)

than RR methods that do not allow for response biases. Table 2 contains the noncompliance estimates and their 95 % bootstrap confidence intervals obtained under both the homogeneous-compliance model without a response bias correction and under a mixture model for the response bias correction. For example, for the health domain, the bias-uncorrected noncompliance percentages for the four items are estimated as 0.2 %, 1.4 %, 4.8 %, and 8.5 %, respectively. In contrast, in the model with the mixture component, the corresponding estimates are 3.3 %, 5.3 %, 8.3 %, and 13.0 %. These differences are substantial and demonstrate the value of the proposed approach for the analysis of RR data. By not taking into account possible response biases, the incidence of noncompliance is severely underestimated. It should be noted that these estimates do not include the mixture component consisting of the SP respondents. Thus, it is possible that even the response bias-corrected estimates are too low if some or all of the SP respondents are noncompliant as well.

Future Directions

Despite the fact that the respondents' privacy is protected by the RR design, it is not always perceived as such by the respondents. Because RR forces respondents to give a potentially self-incriminating answer for something they did not do, it is susceptible to self-protective responses (SP), that is, respondents answer "no" although they should have responded "yes" according to the randomizing device (see, e.g., Edgell et al. 1982). The online questionnaires used were designed in such a way that the outcome of the dice is not recorded and this was mentioned in the instructions given to the respondents. As a result, the respondents were free to give a different answer than the forced "yes" or "no" induced by the dice. Although RR performs relatively well, by eliciting more admittances of fraud than direct-questioning or computer-assisted self-interviews (Lensvelt-Mulders et al. 2005), noncompliance probabilities might still be underestimated if SP is not taken into account.

Recently, several studies have focused on the detection or estimation of SP when RR methods are used. Clark and Desharnais (1998) showed that by splitting the sample into two groups and assigning each group a different randomization probability, it is possible to detect the presence of SP responses and to measure its extent. Böckenholt and van der Heijden (2007) use a multivariate approach to estimate SP by proposing item randomized response model (Eq. 3), where a common latent sensitivity scale is assumed for a set of RR variables, and the response behavior that does not follow the RR design is approached by introducing mixture components in the IRR models with a first component consisting of respondents who answer truthfully and follow an item response model, and a second component consisting of respondents who systematically say "no" to every item in a subset of items. A similar approach is adopted by Cruyff et al. (2007) who worked out the same idea in the context of log-linear models.

The paper of Böckenholt and van der Heijden (2007) also provides insight into characteristics of respondents who have higher probabilities to use SP. Two important variables are the perceived clarity of the instructions (respondents who found the instructions to be clear are more likely to state the truth) and education (respondents with a higher education are also more likely to state the truth). It is important to try out the instruction in a pilot study, possibly using a cognitive survey lab, to see how the RR questions are perceived (see Lensvelt-Mulders and Boeije 2007). Lensvelt et al. (2006) present a best practice that may be used as a starting point.

Related Entries

- ▶ British Crime Survey
- ▶ Compliance and Corporate Crime Control
- ▶ Domestic Violence
- ▶ History of the Dutch Crime Victimization Survey(s)
- ▶ International Crime Victimization Survey
- ▶ National Victimization Surveys
- ▶ Organized Fraud
- ▶ Self-Reported Offending: Reliability and Validity
- ▶ Understanding Victimization Frequency
- ▶ US National Crime Victimization Survey

Recommended Reading and References

Agresti A (2002) Categorical data analysis. Wiley, New Jersey

Böckenholt U, van der Heijden PGM (2004) Measuring noncompliance in benefit regulations with randomized response methods for multiple items. In: Biggeri A, Dreassi E, Lagazio C, Marchi M (eds) 19th international workshop on statistical modelling, Florence, Italy, pp 106–110

Böckenholt U, van der Heijden PGM (2007) Item randomized-response models for measuring noncompliance: risk-return perceptions, social influences, and self-protective responses. Psychometrika 72:245–262

Boruch RF (1971) Assuring confidentiality of responses in social research: a note on strategies. Am Sociol 6:308–311

Chaudhuri A (2011) Randomized response and indirect questioning techniques in surveys. CRC Press, Boca Raton

Chaudhuri A, Mukerjee R (1988) Randomized response: theory and techniques. Marcel Dekker, New York

Clark SJ, Desharnais RA (1998) Honest answers to embarrassing questions: detecting cheating in the randomized response model. Psychol Methods 3:160–168

Cruyff MJLF, van den Hout A, van der Heijden PGM, Böckenholt U (2007) Log-linear randomized-response models taking self-protective response behavior into account. Socio Meth Res 36: 266–282

Cruyff MJLF, van den Hout A, van der Heijden PGM, Böckenholt U (2008) The analysis of randomized-response sum score variables. J R Stat Soc Ser B 71:21–30

Edgell SE, Himmelfarb S, Duncan KL (1982) Validity of forced response in a randomized response model. Socio Meth Res 11:89–110

Elffers H, van der Heijden P, Hezemans M (2003) Explaining regulatory noncompliance: a survey study of rule transgression for two Dutch instrumental laws, applying the randomized response method. J Quant Criminol 19:409–439

Fox J-P (2005) Randomized item response theory models. J Educ Behav Stat 30:1–24

Fox JA, Tracy PE (1986) Randomized response: a method for sensitive surveys. Sage, London

Kuk AYC (1990) Asking sensitive questions indirectly. Biometrika 77:436–438

Lensvelt-Mulders GJLM, Boeije HR (2007) Evaluating compliance with a computer assisted randomized response technique: a qualitative study into the origins of lying and cheating. Comput Hum Behav 23:591–608

Lensvelt-Mulders GJLM, Hox JJ, van der Heijden PGM, Maas CJM (2005) Meta-analysis of randomized response research: thirty-five years of validation. Socio Meth Res 33:319–348

Lensvelt-Mulders GJLM, van der Heijden PGM, Laudy O, van Gils G (2006) A validation of a computer-assisted randomized response survey to estimate the prevalence of fraud in social security. J R Stat Soc A 169:305–318

Maddala GS (1983) Limited dependent and qualitative variables in econometrics. Cambridge University Press, Cambridge

Rasch G (1980) Probabilistic models for some intelligence and attainment tests. University of Chicago Press, Chicago, Originally published 1960, Copenhagen: The Danish Institute of Educational Research

Scheers NJ, Dayton CM (1988) Covariate randomized response models. J Am Stat Assoc 83:969–974

Van den Hout A, Kooiman P (2007) Estimating the linear regression model with categorical covariates subject to randomized response. Comput Stat Data Anal 50:3311–3323

Van den Hout A, van der Heijden PGM (2002) Randomized response, statistical disclosure control and misclassification: a review. Int Stat Rev 70: 269–288

Van der Heijden PGM, van Gils G, Bouts J, Hox JJ (2000) A comparison of randomized response, computer-assisted self-interview, and face-to-face direct questioning eliciting sensitive information in the context of welfare and unemployment benefit. Socio Meth Res 28:505–537

Warner SL (1965) Randomized response: a survey technique for eliminating evasive answer bias. J Am Stat Assoc 60:63–69

Rapid Transit

▶ Crime on Public Transport

Rates of Punishment (Spatial and Temporal Variation)

▶ Institutional Theories of Punishment

Rational Choice

▶ Rational Choice, Deterrence, and Crime: Sociological Contributions

Rational Choice and Prospect Theory

Kyle J. Thomas and Thomas A. Loughran
Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA

Overview

Scholars interested in decision making under risk have traditionally relied on the standard economic assumptions of expected utility (EU) theory (von Neuman and Morgenstern 1947) to specify the process by which individual actors make decisions. Under the EU framework – which serves as the foundation for deterrence and rational choice theories of crime going back to seminal work by Becker (1968) – an individual actor faced with a risky outcome is expected to select the specific behavioral action that yields the maximized anticipated payoffs. Importantly, the utility of behaviors are weighted by their probabilities of occurrence. The domain of the utility function in this model is absolute benefits and costs, whereby individual decisions of a course of action are dependent solely on the mathematical outcome of the difference between the probability of experiencing the utility of a behavior and the probability of experiencing disutility. Over time, however, researchers in behavioral economics and psychology have become increasingly aware that expected utility theory is not an adequate descriptive model of preferences under risk. For instance, consider the famous example posed by Tversky and Kahneman (1987) of an outbreak of avian flu that is expected to result in the death of 6,000 individuals. A national government has two plans to combat the pandemic: Plan A has a 100 % chance of saving 2,000 people, while plan B has a 1/3 chance of saving 6,000 people and a 2/3 chance of saving no one. Note that the expected utility in plan A (2,000 × 1 = 2,000) and plan B (6,000 × 1/3 = 2,000) is the same, but research indicates that individuals overwhelmingly select plan A, the more certain option (Tversky and Kahneman 1987). In other words, individuals do not make decisions based exclusively on the expected outcomes, but have a tendency to overweight more certain behavioral choices. Interestingly, Tversky and Kahneman (1987) demonstrate that this preference for certainty only holds when the risky outcomes are framed as losses, but does not hold when framed as gains. In other words, if the same hypothetical pandemic is framed as a 100 % chance of saving 4,000 lives (plan A; 4,000 × 1 = 4,000) and a 2/3 chance of saving 6,000 lives (plan B; 6,000 × 2/3 = 4,000), individuals overwhelmingly select plan B – the more risky option. This suggests that preferences for risk can be situational depending on whether the individual is faced with the risk of gaining value or losing value.

The above examples are just two instances of violations of assumptions of standard expected utility theory, but many others exist. In response, several alternative models have been put forth that seek to address the shortcomings of expected utility theory. The most prominent model is *prospect theory*, developed by the cognitive psychologists Daniel Kahneman and Amos Tversky (1979). Prospect theory provides several alternative predictions that explain observed exceptions to the standard economic assumptions of expected utility theory. First, rather than choices being a function of absolute benefits and costs, value is determined by the *gains* and *losses* of an action relative to a reference point. Second, probabilities in the expected utility model are replaced by decision weights in prospect theory, which allows individuals to be more sensitive to expected losses than they are to expected gains (i.e., loss averse) and for marginal changes to be less impactful as they move away from the reference point (i.e., diminishing returns). Moreover, prospect theory posits that individuals have a tendency to underweight outcomes that are merely probable in comparison to outcomes that can be obtained with certainty. All of these predictions have received considerable empirical support, and prospect theory remains a prominent descriptive framework for the way people make decisions in the face of risk and uncertainty (Starmer 2003).

Despite its utility in understanding decision making under risk, few studies to date have applied prospect theory to the study of crime (see Loughran et al. 2011). This entry outlines the value of incorporating prospect theory to the study of crime and deviance and to rational choice and deterrence theories more specifically. The entry begins by briefly specifying the standard economic assumptions prevalent in expected utility theory. Next, it provides a brief overview of violations to these standard economic assumptions, findings that were pivotal in Kahneman and Tversky's development of prospect theory. Finally, the entry then describes the core tenets of prospect theory, and its descriptive advantages for understanding choices when faced with risky and uncertain situations, and concludes by highlighting some important questions that prospect theory raises about criminal decision making.

Fundamentals of Expected Utility Theory

Expected utility theory is a model of decision making under risk, where each behavioral action leads to a set of outcomes that can be either beneficial or costly to the actor. Importantly, the probability of each outcome is known, which enables individuals to make informed decisions when faced with risk. The core principle of expected utility theory is based on the standard economic assumption that individuals select the behavioral option that has the maximum expected utility. More specifically, the utility of each outcome is weighted by their associated known probabilities, and individuals choose the behavioral action that has the highest weighted final sum (Luce and Raiffa 1957).

They key components of expected utility theory have been incorporated into prominent criminological perspectives such as rational choice and deterrence theories (see Becker 1968; Cornish and Clarke 1986). From these theoretical perspectives, an individual is expected to offend if $p * U(\text{Benefits}) - p * U(\text{Costs}) > 0$, where $U(*)$ is a utility function that standardizes various costs and benefits from offending into comparable units, and p is that perceived probability (between 0 and 1) of experiencing an outcome. This model suggests that an individual will decide to commit a crime if the benefits from offending outweigh the potential costs. The known probabilities p are important in this model: For instance, as p increases with regard to the costs of crime, the net benefits of offending necessarily decrease. In other words, the deterrent effect of certainty on offending increases monotonically with increments of risk. What constitutes a "benefit" and "cost" of crime varies considerably across perspectives. For instance, early work in deterrence and rational choice focused almost exclusively on monetary gains and costs deriving from formal sanctions (Becker 1968), while more recent works have broadened these choice perspectives to allow benefits to include things such as prestige and social status, and costs to include things such as disappointing loved ones, moral regard, and shame (Cornish and Clarke 1986; Grasmick and Bursik 1990; Paternoster 1989). In general, the calculus of rational choice is generally not well understood, but there is strong evidence that the risk of detection, or the ostensible weight in this rational choice calculus, is associated with offender decisions (Nagin 1998).

Regardless of the outcomes that constitute benefits and costs, the process outlined in expected utility theory is thought to be a descriptive model of how individuals actually behave in risky situations. Importantly, the expected utility model makes several important implications with regard to the decision-making process. For instance, expected utility theory does not specify the functional form of the decision-making process. In this way, it is at least implied that utilities and their associated probabilities affect choices in a linear fashion, where a 10 % increase in the probability of an outcome has the same marginal effect at all points on the probability continuum (an increase in the certainty of arrest from 10 % to 20 % has the same marginal effect as an increase from 60 % to 70 %). Further, the expected utility model assumes that individuals treat benefits (gains) and costs (losses) in a similarly weighted manner. In monetary terms, a $100 gain that may be accrued by engaging in a behavior is treated the same in an individual's calculus as a $100 loss.

The core tenets of expected utility theory are prevalent throughout the field of criminology. However, research in behavioral economics has identified several situations where the predictions of expected utility theory are violated.

Violations of Expected Utility Theory and the Foundations of Prospect Theory

Kahneman and Tversky (1979) begin their seminal paper on prospect theory by presenting the results of several laboratory experiments that involve hypothetical choices. Many of their findings explicitly contradict the assumptions of expected utility theory, suggesting that individuals appear to make irrational decisions under some circumstances of risk. However, Kahneman and Tversky (1979) were able to demonstrate that the apparent irrationality has considerable consistency. Accordingly, they attempted to develop a description of decision making under risk that can explain the following observations:

1. People do not tend to make choices in terms of their net utility as predicted by expected utility theory, but instead code the outcomes of behavioral choices as *gains* or *losses* relative to a reference point. The reference point usually corresponds to one's current assets or the status quo but can also include one's expectations, and represents the zero point on one's value scale. Accordingly, the expected utility of a choice is thus evaluated as deviations from this subjective zero scale, such that "the same level of wealth... may imply abject poverty for one person and great riches for another – depending on their current assets" (Kahneman and Tversky 1979, p. 277). Put simply, objective benefits and costs of behavioral choices can have considerably varied effects across individuals, depending on one's initial asset position.
2. Because individuals encode outcomes in terms of gains and losses relative to reference points, the framing of choice options can be critical in an individual's decision. Take, for instance, the avian bird flu example presented in the introduction, where individuals are asked to select one of two programs seeking to prevent spread of the disease. We noted that when the outcomes are framed as lives saved, respondents overwhelmingly select the sure gain of 2,000 lives rather than the 1/3 chance at saving 6,000 lives. When framed as lives lost, however, individuals tend to reject the sure loss of 4,000 deaths and prefer to take the risk. This relationship can be interpreted as support for the idea that individuals are risk-averse with regard to gains and risk-seeking with regard to losses, but Starmer (2003) argues that it is just as important to understand how the consequences are interpreted. In other words, though the 4,000 lives lost is objectively the same in both situations, the framing of the scenario influences whether they are interpreted as gains or losses. Accordingly, Kahneman and Tversky go to some length to explain this in their discussion of editing and heuristics.

As will be detailed more below, prospect theory suggests that individuals edit choice prospects using a variety of heuristics, prior to evaluating their utility. One of the most important editing processes, discussed above, is the coding of outcomes as gains or losses relative to a reference point. Kahneman and Tversky go on to note that, though the reference point is typically one's current assets, "the location of the reference point, and the consequent coding of outcomes as gains and losses, can be affected by the *formulation of the offered prospects...*" (Kahneman and Tversky 1979, p. 274, emphasis added). Thus, the framing of behavioral choices can determine an individual's reference point and, in turn, whether prospects are interpreted as gains or losses. This framing effect has been demonstrated across numerous experimental studies (McNeil et al. 1982; Tversky and Karhneman 1986).

Situational characteristics can play a major factor framing the way that individuals code and evaluate prospects when making decisions. Importantly, as Levy (1992) notes, such framing effects can often be subjective, particularly when there is ambiguity with regard to the status quo. In any sense, the framing of behavioral choices can play a major role in offender decision making. For instance, framing and the use of availability heuristics may explain why individuals systematically overpredict the certainty of arrest in self-report studies of offending. Much of individuals' knowledge about the crime comes from media accounts, which can cause individuals to perceive the status quo on the certainty of arrest to be high. When faced with situations where offending is possible, however, individuals may develop a new status quo from which to analyze behavioral choices. Similarly, an individual who is socializing in a group where the perceived status quo is conformity, he/she is likely to hold conforming values as well. If that same individual is present in a group that espouses pro-delinquent values, then that individual is likely to develop a new status quo and make behavioral choices that are relative to this new reference point (Warr 2002). The point is that many of the field's most dominant explanations of crime emphasize the importance of subjective beliefs – such as perceived certainty of arrest or social reinforcement – but often treat such beliefs as static across different situations. A more nuanced understanding of the roles of subjective beliefs may be garnered by understanding how situations frame prospects and influence the decision-making processes of individuals (see Manski 2004).

3. Though Kahneman and Tversky (1979) note that individuals make behavioral choices based on changes from a reference point, they also find evidence that individuals treat gains differently from losses in two ways. First, individuals tend to be risk-averse when it comes to gains and risk-seeking when it comes to losses. A person is risk-averse if he/she prefers a certain prospect (x) to any risky prospect with expected value x. A person is risk-seeking if he/she prefers to accept risk to secure benefits or avoid losses. For instance, a person is considered risk-averse if they prefer $50 guaranteed to a 50 % chance of winning $100, while a person is risk-seeking if they prefer a 50 % chance of winning $100 to $50 guaranteed. The results of Kahneman and Tversky's experiments explicitly contradicted the descriptive model of expected utility theory with regard to the differential treatment of gains and losses. When faced with the hypothetical prospect of *gaining* money, Kahneman and Tversky (1979) found that 82 % of participants preferred the certain outcome of gaining $3,000 to an 80 % chance of gaining $4,000 and a 20 % chance of gaining nothing. Note that the expected utility is actually higher in the latter option ($3,200 vs. $3,000), but respondents are overwhelmingly risk-averse and prefer the more certain, although smaller, payoff. Conversely, when faced with the hypothetical prospect of *losing* money, individuals are substantially more risk-seeking. Kahneman and Tversky show that 92 % of participants preferred to gamble on an 80 % chance of losing $4,000 and a 20 % chance of losing nothing over a certain loss of $3,000. Again, in this scenario, individuals chose the option with the lower expected utility, which directly contradicts the assumptions of expected utility theory. This pattern has been demonstrated across a wide range of behaviors (see Edwards 1996).

4. Kahneman and Tversky (1979) also find that individuals treat losses different than gains in that (i) the utility functions of gains and losses differ such that the shape of the utility function is concave for gains but convex for losses and (ii) that the utility function is steeper in the domain of losses. Tversky and Kahneman (1992) interpret these as two general properties of decision making under risk: *diminishing sensitivity* and *loss aversion*. "Diminishing sensitivity holds that the psychological impact of a marginal change will decrease as we move further away from a reference point" (Starmer 2003, p. 128). Thus, with a reference point of zero, the difference between a gain of $10 and $20 will seem larger than the difference between a gain of $110 and $120. This trend is known in the economic literature as *diminished marginal utility*. For losses, the same principle is known as *diminished marginal disutility* and also has implications for offender decision making – i.e., the marginal change in increasing the severity of sanctions from 1 year to 2 years would have a larger impact on offending than the increase from 11 years to 12 years.

   The second property (ii) discussed with regard to the utility function of gains and losses, loss aversion, is the idea that "losses loom larger than corresponding gains" (Tversky and Kahneman 1992). This is to say that individuals place a higher value on a good that they already possess than to one that they do not. Thus, when presented with a scenario where there is a 50 % chance of winning $50 and/or a 50 % chance of losing $50, Tversky and Kahneman have demonstrated that individuals find the latter option distinctly unattractive.
5. As noted in the introduction, one finding of Kahneman and Tversky (1979) that contradicts expected utility theory is that individuals have a tendency to overweight outcomes that are certain and underweight outcomes that are merely probable, what Kahneman and Tversky label *the certainty effect*. For example, consider choosing between an 80 % chance of receiving $4,000 and a 20 % chance of gaining nothing versus a certain gain of $3,000. Notice that the expected payoff is higher in the former option (EU = $3,200), but research indicates that individuals are more prone to select the second option. Moreover, research also indicates that individuals have a tendency to treat extremely likely, but uncertain, outcomes as if they are certain, which is known as the *pseudocertainty effect* (Tversky and Kahneman 1986). The point is that individuals have a tendency to weight probabilities near 0 and 1 greater than would be anticipated in expected utility theory. Consequently, studies have also contradicted the assumptions of expected utility theory that utilities of risky outcomes are weighted linearly by their probabilities. For instance, a change in the probability of experiencing a loss from .10 to 0 is weighted more heavily than a change from .70 to .60.

Fundamentals of Prospect Theory: An Alternative Model for Choice Under Risk

It is apparent that several of the central predictions of expected utility theory are regularly violated in instances where decisions are made under risk. Prospect theory attempts to incorporate these observed violations in order to provide a more descriptively valid theory of risky decision making. In prospect theory, choice is modeled as a two-phase process: the editing phase and the evaluation phase. In the first phase, prospects are "edited" using a variety of decision heuristics. The purpose of the editing phase is to conduct a preliminary and simple analysis of the offered prospects. By using various heuristics, individuals conduct several operations that transform outcomes and their associated probabilities to allow for a simple evaluation of the prospects. Kahneman and Tversky (1979) note that there are several mental operations that occur in the editing phase for each individual prospect: *coding* which defines each prospect as either a gain or loss relative to a neutral reference point, *combination* which combines the probabilities of identical outcomes, and *segregation* which segregates the riskless components of prospects from the risky components. Moreover, Kahneman and Tversky (1979) also note that individuals compare two or more outcomes in the editing phase using the following: *cancellation* which discards the common components of different prospects, *simplification* which is rounding probabilities, and *detection of dominance* which discards outcomes that are easily dominated by other outcomes.

The editing phase is critical in the decision-making process and is one of the most distinguishing features of prospect theory. It allows individuals to prepare outcomes and their associated certainties of occurrence prior to evaluation. Note that the editing phase in prospect theory incorporates many of the observed anomalies to the standard expected utility theory. For instance, reference points become the relative point at which individuals evaluate gains and losses rather than simply assessing the absolute cost and benefits of behavior. Moreover, prospect theory allows for reference points to be determined not solely by one's current assets but also by one's expectations and by the way in which the prospects are framed. Further still, and contrary to expected utility model, the preference ordering of outcomes "need not be invariant across contexts, because the same offered prospects can be edited in different ways depending on the context in which it appears" (Kahneman and Tversky 1979, p. 275). These features of prospect theory can make complex choice situations difficult to predict, as the editing process is influenced by the norms and expectations of the individual actor (Tversky and Karhneman 1986). Because of this, Kahneman and Tversky (1979) largely focus on the second phase of decision making, but it remains noteworthy that certain individual and contextual factors can reasonably be used to predict that editing process and the determination of status quo reference points.

Once the editing phase is complete, choices among edited prospects are determined in the evaluation phase. In this phase, individuals select the maximized edited prospect as determined by a simple decision-weighted utility function that transforms probabilities into weights that impact the overall value of the prospect and evaluates the subjective value of each outcome relative to a reference point that serves as a zero point on a value scale (see Kahneman and Tversky 1979, p. 275). Hence, value is determined by the deviation of each outcome from the reference point.

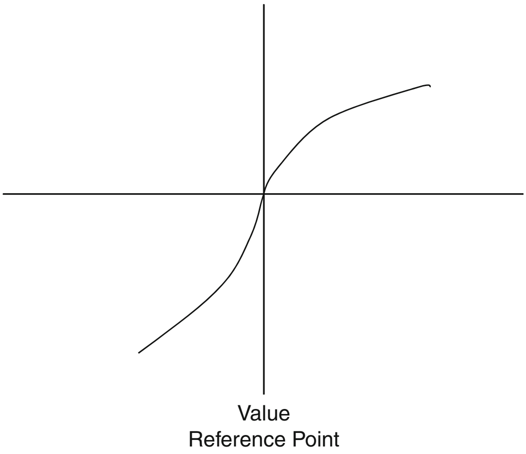

Rational Choice and Prospect Theory, Fig. 1 Value Function According to Prospect Theory (Kahneman and Tversky, 1979)

In prospect theory, outcomes are evaluated using a utility function shaped like that presented in Fig. 1. Notice that this function has several important qualities. First, the function is slightly kinked at the reference points (Starmer 2003), where $x_i = 0$. Another quality of this function is that it is concave for gains and convex for losses. These properties are important as they indicate diminishing sensitivity as described above – i.e., the psychological impact of a marginal change decreases the further one gets from the reference point. A third quality of this value function is that it is steeper in the domain of losses, which implies that individuals tend to be more loss averse. Thus, many of the properties of prospect theory's value function have been observed in experimental assessments of decision making under risk.

The decision-weighting function specified in prospect theory also possesses some important qualities. It should be noted that decisions weights are inferred from choices in a manner that is similar to probabilities in expected utility theory, but these decision weights are not probabilities and should not be interpreted as such. Instead, the "decision weights measure the impact of events on the desirability of prospects, and not merely the perceived likelihood of the events" (Kahneman and Tversky 1979, p. 280). This distinction is important because it allows the decision weight to be influenced by factors other than the probability of an outcome. For instance, ambiguity, or the uncertainty about the level of certainty one has with regard to the outcome, can also impact the decision weight. As will be noted below, this feature of the decision weight is an important quality of prospect theory and has important implications for understanding the decision to offend (Loughran et al. 2011).

The decision-weighting function has several other important features. For instance, Kahneman and Tversky (1979) propose a weighting function where "large" probabilities are underweighted and "small" probabilities are overweighted. In fact, the function is not defined for probabilities at or near 0 or 1, because "people are limited in their ability to comprehend and evaluate extreme probabilities" (Kahneman and Tversky 1979, p. 282). In particular, individuals are likely to completely ignore unusually small probabilities and exaggerate extremely high probabilities. Kahneman and Tversky (1979) note that this has another important implication: It suggests that the decision weight function is relatively shallow in the midpoints, but changes abruptly near the endpoints, which is consistent with the certainty effect described above.

In summary, prospect theory is an important alternative description of decision making under risk, because it is well suited to account for many of the observed violations of expected utility theory. The certainty effect, framing effects, loss aversion, ambiguity in uncertainty, and the finding that utility is evaluated as deviations from a reference point can all be explained by prospect theory, and this theoretical perspective is highly influential in understanding social behavior (Levy 1992).

Future Directions

The Role of Prospect Theory in Criminology

Research in behavioral economics and psychology has consistently found support for the general tenets of prospect theory. Nevertheless, the importance of prospect theory, with few exceptions (see Loughran et al. 2011), has not been integrated into criminology. Indeed, models testing offender decision making have relied almost exclusively on the descriptive model of rational choice theory, a model of decision making under risk that we know is violated in many circumstances. Given the shortcomings of expected utility theory in predicting decisions under risk, prospect theory has the potential to be highly influential in understanding offending behavior.

Loughran and his colleagues (2011, 2012b) have begun to integrate prospect theory into deterrence theory, and the results of their studies have suggested that the predictions of prospect theory fit offending data better than those derived from expected utility theory. To be sure, most studies assessing the relationship between perceived certainty of arrest and probability of offending have, at least implicitly, assumed that the relationship is linear. However, as noted above, prospect theory predicts that the relationship between certainty of risk and behavior is nonlinear, and laboratory studies have indicated that individuals only begin to weight potential losses when the certainty of experiencing an outcome is around 30 %. This was described as a "tipping effect" whereby probabilities below this were essentially underweighted or disregarded all together. This, of course, has significant implications for deterrence theory, as it challenges the functional form of the certainty-offending relationship that marginal changes in certainty affect probability of offending in a linear fashion. In support of prospect theory, Loughran and Colleagues (2012b) showed that individuals tend to weight perceived probabilities of risk non-linearly. Specifically, perceived risk only had a deterrent effect when the probability of arrest exceeded a threshold of .3 to .4, and there was a substantial increase in the marginal effect after that. Thus, Loughran et al.'s findings provided empirical support for the nonlinear probability weighting proffered by prospect theory.

In another study, Loughran et al. (2011) extended prospect theory to assess if ambiguity, or uncertainty about the true probability of detection (Camerer and Weber 1992), in the certainty of arrest was important in understanding criminal decision making. Though not a formal test of prospect theory per se, the results were consistent with framing effects. Using two different data sources, they found that ambiguity in the perceived risk of detection was a deterrent when the perceived risk was low, yet actually encouraged offending when the risk was higher.

These are the only two studies to date that have incorporated prospect theory into the study of crime. Though underdeveloped given the empirical support for prospect theory and the important role of choice in criminological theories (Akers 1990; Cornish and Clarke 1986; Paternoster 1989), this gap in the literature does provide avenues of future research that can shed considerable light on our understanding of offending behavior.

For instance, the idea that individuals make decisions based on gains and losses from a reference point raises some interesting questions regarding the decision to engage in crime and has considerable implications for criminological theory, beyond the commission of economic crimes. For example, the cost of experiencing an arrest and the gains in social status incurred when engaging in crime are typically measured as objective and absolute benefits/ costs, and researchers rarely model these costs and benefits as losses and gains relative to a reference point (i.e., one's current social status among friends). Moreover, the predictions of prospect theory also raise important questions regarding the measurement of crime and delinquency in self-report questionnaires, suggesting that the framing of questions may influence responses and that individuals may have a tendency to utilize response strategies that are highly contingent on the status quo rather than their own subjective beliefs (see Loughran et al. 2012a, forthcoming).

It may also be useful to understand how diminishing sensitivity and loss aversion affect individual decisions to offend. Many criminological theories highlight the importance of rewards to crime (Akers 1998; Cornish and Clarke 1986), but prospect theory would make distinct predictions regarding the functional form on the relationship between rewards and offending. Does the proportion of one's deviant peers affect delinquency in a linear fashion? Or would the data suggest that delinquent peers have a diminishing marginal utility when it comes to promoting deviance? Loss aversion suggests that individuals may be more influenced by the threat of losing social status and material assets than they are by the potential or gaining such things. Further, loss aversion predicts that individuals are more risk-seeking after experiencing losses in their assets. How does this affect decisions to participate in illegal markets? Are individuals more likely to initiate into or increase their involvement in markets after experiencing losses to income (both legal and illegal)? These are just two examples, but diminishing sensitivity and loss aversion have implications for a wide range of criminogenic influences.

Another avenue for future research in the role of prospect theory and crime concerns certainty in decision weighting. Research in deterrence theory has found only a weak relationship between perceived likelihood of arrest and offending (Pratt et al. 2006), leading some scholars to question whether individuals consider the threat of sanctions before deciding to offend and the rationality of offenders. But criminogenic situations are often such that individuals have almost certain gains (stealing money) but relatively less certain losses (being arrest). In this way, it may not be that individuals are irrational in their decision making or fail to consider potential costs, but rather are attracted to the more certain gains that are involved in the risk. In other words, though individuals have perceived that the cost of an arrest is relatively high, this disutility maybe underweighted because of the lack of certainty in experiencing the loss, relative to the gains.

Conclusion

Deterrence and rational choice theories continue to be two of the most prominent perspectives in criminology. Though our understanding of offender decision making has unquestionably advanced in recent years, most research assessing criminal choice is based off of the standard economic assumptions of expected utility theory that have repeatedly been shown to be violated in situations involving choice under risk. Prospect theory is an alternative model to decision making under risk that can explain many of these shortcomings (Tversky and Kahneman 1992). To date, prospect theory has not yet been widely integrated to the study of crime and delinquency. While this is unfortunate, it does provide a plethora of opportunities for scholars interested in offender decision making to apply the predictions of prospect theory – which oftentimes differ greatly from predictions derived from traditional expected utility theory – to the decision to engage in criminal and deviant behaviors and in turn further advance the field's understanding of criminal choice.

Related Entries

- ▶ Deterrence
- ▶ Rational Choice Theory
- ▶ Social Learning Theory

Recommended Reading and References

Akers RL (1990) Rational choice, deterrence and social learning theory in criminology: the path not taken. J Crim Law Criminol 81:653–676

Akers RL (1998) Social learning and social structure: a general theory of crime and deviance. Transaction Publishers, New Brunswick

Becker GS (1968) Crime and punishment: an economic approach. J Polit Econ 78:169–217

Camerer CF, Weber MW (1992) Recent developments in modeling preferences: uncertainty and ambiguity. J Risk Uncertain 5:325–370

Cornish D, Clarke RV (1986) The reasoning criminal. Springer, New York

Edwards KD (1996) Prospect theory: A literature review. International Review of Financial analysis 5:19–38.

Grasmick HG, Bursik RJ (1990) Conscience, significant others and rational choice. Law Soc Rev 34:837–864

Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. Econometrica 47:263–291

Levy JS (1992) Prospect theory and international relations: theoretical applications and analytical problems. Polit Psychol 13:283–310

Loughran TA, Paternoster R, Piquero AR, Pogarsky G (2011) On ambiguity in perceptions of risk: implications for criminal decision making and deterrence. Criminology 49:1029–1061

Loughran TA, Paternoster R, Thomas KJ (2012a) How valid are self-report offending and subjective deterrence measures? An investigation using Bayesian Truth Serum. Working manuscript

Loughran TA, Pogarsky G, Piquero AR, Paternoster R (2012b) Re-examining the functional form of the certainty effect in deterrence theory. Justice Quart 29:712–741

Luce RD, Raiffa H (1957) Games and decisions: introduction and critical survey. Wiley, New York

Manski CF (2004) Measuring expectations. Econometrica 72:1329–1376

McNeil BJ, Pauker SG, Sox HC, Tversky A (1982) On the elicitation of preferences for alternative therapies. N Engl J Med 306:1259–1262

Nagin DS (1998) Criminal deterrence research at the outset of the twenty-first century. Crime Justice Rev Res 23:1–42

Paternoster R (1989) Decisions to participate in and desist from four types of common delinquency: deterrence and the rational choice perspective. Law Soc Rev 23:7–40

Pratt TC, Cullen FT, Blevins KR, Daigle LE, Madensen TD (2006) The empirical status of deterrence theory: a meta-analysis. In: Cullen FT, Wright JP, Blevins KR (eds) Taking stock: the status of criminological theory. Transaction Publishers, New Brunswick

Starmer C (2003) Developments in nonexpected-utility theory: the hunt for a descriptive theory of choice under risk. In: Camerer CF, Loewenstein G, Rabid M (eds) Advances in behavioral economics. Princeton University Press, Princeton

Tversky A, Kahneman D (1986) Rational choice and the framing of decisions. J Business 59:S251–S278

Tversky A, Kahneman D (1987) Rational choice and the framing of decisions. In: Hargoth RM, Reder MW (eds) Rational choice and the contrast between economics and psychology. University of Chicago Press, Chicago

Tversky A, Kahneman D (1992) Advances in prospect theory: cumulative representation of uncertainty. J Risk Uncertainty 5:297–323

von Neuman J, Morgenstern O (1947) Theory of games and economic behavior. Princeton University Press, Princeton

Warr M (2002) Companions in crime: the social aspects of criminal conduct. Cambridge University Press, Cambridge

Rational Choice Theory

Bill McCarthy and Ali Chaudhary
Department of Sociology, University of
California Davis, Davis, CA, USA

Overview

Rational choice theory refers to a set of ideas about the relationship between people's preferences and the choices they make. There are several variants of rational choice theory and this essay refers to these collectively as the rational choice approach (RCA). The conceptual foundations of the RCA originate in Cesare Beccaria's1764 essay *On Crimes and Punishments* and Jeremy Bentham's 1789 work, *An Introduction to the Principles of Morals and Legislation*. One school of thought, the deterrence approach, builds on Beccaria's insights that effective punishments need to be swift and certain (Paternoster 2010). Alternative uses of the RCA focus on Bentham's formalization of the idea that the motivations for actions, criminal or otherwise, are universally grounded in individual self-interest and the desire to maximize pleasure and minimize pain. As a result, punishments require a level of rationality if they are to influence people's perceptions of the pleasures and pains associated with particular choices.

Beccaria's and Bentham's approach to crime and punishment initially had some influence on punishment and social control practices but the ideas associated with the "classical school" were superseded by two centuries of biological, psychological, and sociological explanations. Their disciplinary differences notwithstanding, these explanations of crime emphasize the uniqueness and pathological nature of criminal behavior. They argue that crime occurs because biological, psychological or social conditions motivate people to break the law. Ideas from the classical school began to remerge in a revival associated with the economist Gary Becker's (1968) *expected-utility* model of criminal decision-making, the work on "reasoning criminals" by criminologists Derek Cornish and Ronald Clarke (1986), and sociologist Jack Gibbs' writings on social control (citations for works cited but not included in the references can be found in McCarthy (2002) and Paternoster (2010)). The RCA approach to crime builds on Beccaria's and Bentham's founding principles and the centrality of self-interest for understanding behavior. The RCA uses terms such as rational or preferences that have different popular and disciplinary meanings. The following outlines the key assumptions behind the approach and explains the meaning of key terms.

The Rational Choice Approach (RCA)

According to the RCA approach:

1. People have preferences for outcomes (goods, services, states of being, etc.); preferences do not typically refer to actions or behaviors.
2. People's preferences are influenced by the expected benefits of an outcome, relative to its costs. There are several types of potential benefits (e.g., monetary, emotional, and social) and costs (e.g., opportunity, external, sunk as well as monetary, emotional, and social). The anticipated cost-benefit ratio associated with an action is an indicator of its expected utility.
3. People can order their preferences for outcomes from most to least valued. Preferences are relatively stable: they do not change during a decision, but can be modified as a result of new information.
4. People's assessments of the benefits and costs of outcomes are influenced by the information they collect. Gathering information is however, itself a cost. Thus, although people prefer to have all available information when making decisions, choices are made frequently with incomplete information. People may believe they have adequate information when they do not, they have imperfect memories, and they often miscalculate. In other words, people have *subjective expectations* about the utility they will receive from their choices.

5. Preferences are also influenced by people's orientation to time. Individuals with a positive time preference will need greater future compensation in order to forgo a present benefit, whereas those willing to forgo a current benefit for a lower level return in the future have a negative time preference. Time preferences are not fixed across all decisions but are influenced by several factors, including a person's current level of a valued outcome.
6. Preferences are further affected by attitudes toward risk and uncertainty. People do not have a preference for risk taking in itself (i.e., risk taking is not an outcome); rather, people's attitudes toward risk taking influence the utility they associated with an outcome. A risk-averse person generally refuses to accept what is calculated to be a fair gamble; those who generally have a preference for taking fair gambles, rather than a sure thing are risk-seekers; and between these extremes are people who are risk-neutral: those who are generally indifferent to accepting or refusing a fair gamble. Some rational choice theorists assume that risk disposition if relatively fixed, whereas others assume it will vary across types of decisions and situations.
7. Rational actions are those that are consistent with the above assumptions. Common shorthand is to describe such actions as being consistent with the maximization of utility. Determining a behavior's "rationality" depends on knowing, or making assumptions about a person's information, preference ordering, and approach to risk-taking, and time discounting. People's rational choices may, therefore, result in different behaviors, even when they are faced with the same situation.
8. The RCA does not preclude people from acting irrationally and people may pursue a course of action inconsistent with their preferences for a variety of reasons. Their decisions may be negatively influenced by an intense emotion or a sudden change in context. They may have limited cognitive skills that reduce their ability to use effectively the information they gather or to reflect upon previous choices, or they may be unaware of the interests that motivate them (these may be equally obscure to observers). Explanations of behavior that emphasize false consciousness, habitus, national culture, inertia, determinism (biological, psychological, or social) or similar forces suggest that these may also prompt people to make choices that are inconsistent with their preferences.
9. The RCA does not argue that people always think in ways typically associated with rationality as used in common discourse (e.g., reasoned, thoughtful, reflective), nor does it assume people undertake literal calculations. In its simplest form, the RCA refers to the consistency between people's preferences and choices. It is also a probabilistic, rather than a deterministic approach: it explains how most people make many of their decisions, without assuming that *all choices* can be explained. It does not assume that people are always conscious of their attempts to maximize their interests but simply argues that many of their actions can be understood as rational. As is the case with other accounts, the RCA simplifies the complex causal origins of behavior; however, its value lies in its parsimonious, elegant explanation that has considerable predictive power.

The RCA and Crime

In contrast to biological, psychological and sociological explanations of crime, the RCA approach assumes that crime can be understood "as if" people choose to offend by using the same principles of cost-benefit analysis they use when selecting legal behaviors. Thus, the decision to offend is influenced by people's preferences, their attitudes toward risk and time discounting, and their estimates of an illegal opportunity's availability, costs and benefits, versus a legitimate opportunity's availability, costs and potential for realizing the same or comparable returns. Or, as more commonly expressed by economists, people offend when the subjectively expected return to crime (i.e., the benefits-costs ratio) exceeds what they believe they will obtain by spending the same time and other resources to pursue legal activities (Mehlkop and Graeff 2010).

The RCA differs from many theories of crime in that it provides an account of how people's preferences affect their choices, rather than explaining the source of their preferences. Thus, it is a sharp contrast to theories that argue that crime is a result of low self-control, differential association, weak social bonds, strain, labeling, disadvantaged neighborhoods, or other social experiences or forces. Indeed, many of these explanations assume that offending is irrational and suboptimal.

Theoretically, the RCA approach shares some, but not all of the features of other explanations of criminal decision-making, such as routine activity theory, the reasoned-offender approach, and the criminal-event perspective. Although the RCA contrasts with theories that explain the origins of choice, there is a considerable conceptual overlap between the RCA and more sociological theories of offending and suggestions for theoretical integration are common. However, it is incompatible with explanations that argue that structural conditions or socialization produce character defects that make an offender's decision-making distinct from that of non-offenders. In sum, the RCA provides a fruitful approach to understanding criminal decision-making, and can be combined with explanations of the origins of preferences, and the availability of the mechanisms, or instruments, by which preferences are realized (see Dahlbäck 2003 for a more detailed explication of the RCA to crime).

Incentives and Crime

Punishment Costs

Most of the early RCA research investigates the hypothesis that all else equal, increases in punishment should decrease offending. A large number of studies have explored the deterrent effect of capital punishment (most of the research discussed here uses US data). The general finding is that executions have little deterrent effect (Levitt and Miles 2006). There is however, evidence that the costs associated with incarceration act to deter offending. Early studies have a number of shortcomings, but recent research has used more sophisticated methods and higher quality data to investigate this relationship. In their review, Levitt and Miles (2006) note that a number of recent studies find that increases in incarceration rates are associated with subsequent declines in arrest rates; however, the size of the effect is modest (i.e., a 1–10 % reduction in arrest rates for a comparable increase in prison populations), and the escalation of the cost of imprisoning people may make incarceration a less prudent deterrent.

Recent investigations that use individual-level data also document a deterrent effect for several, but not all types of offending. Research on inmates finds large deterrent effects emanating from the certainty of punishment, and smaller, generally insignificant effects from the severity of sanctions (Grogger 1991). However, punishment effects are conditional. The certainty effect is greatest for serious felony crimes and for whites, and declines in strength for non-serious offenses, blacks and Hispanics. In contrast, the severity effect is positive for whites and negative for blacks. Related research reports that inmates who successfully offended – that is who were not arrested for their crimes – are more likely to think that they could avoid being arrested in the future if they re-offended (Horney and Marshall 1992).

A number of earlier studies reported little effect of policing on crime, whereas more contemporary research that uses sophisticated methods and richer data finds that increases in the number of police have nontrivial but modest negative effects on crime (e.g., a 3–5 % reduction for a 1 % increase in the number of police; see Levitt and Miles 2006). There may however, be a ceiling effect: at some point police may have arrested most of the offenders involved in serious violent or property crimes, they may then turn to arresting offenders who commit more trivial or victim-less crimes (e.g., drug use) in part because police use crime and arrest rates to justify their employment and the need to hire more officers.

Similar to imprisonment, increases in policing may not be the most efficient way to deter crime.

Related studies focus on individual differences in perceptions about the likelihood of punishment and the consequence of this variation for crime. Many of these investigations use vignettes or hypothetical situations to measure people's willingness or intention to offend in the future. For example, a recent study of adults finds that individuals who believe that the probability of detection is low are significantly more willing to commit a tax fraud in the future (Mehlkop and Graeff 2010). Research on adolescents also finds that youth who believe that arrest is both likely and costly are less likely to have committed a theft or violent crime (Matsueda et al. 2006).

Economic Costs

The RCA approach suggests that economic costs, such as a loss of legitimate income, also influence offending. Consistent with this hypothesis, research on inmates reports a strong, negative effect of legal income on crime and a positive relationship between crime and the length of a current jobless spell (Grogger 1991; also see Uggen and Thompson 2003). As well, relatively minor criminal activity complements employment, whereas employment and serious crime are substitute activities. Thus, research that pools data on minor and more serious crime may erroneously conclude that employment does not affect criminal activity.

Other investigations also demonstrate a connection between a decrease in wages and offending. For example, research suggests that that declining wages in the USA in the 1970–1980s may have contributed considerably to youth crime increases in these years, with a 20 % fall in wages leading to a comparable increase in offending (Grogger 1991). Moreover, rising wages explain a considerable amount of the decline in offending that occurs with age.

Comparable analyses find that economic incentives and opportunity costs exert a powerful influence on offending. For example, the odds that an offender will stop offending increase with legal earnings and they decrease with illegal ones (Pezzin 1995). Furthermore, although the number of past convictions encourages desistance, the magnitude of this effect is considerably smaller than that of legal wages.

Other Costs

Sociologists have contributed further to the economic approach to crime's costs by proposing a more inclusive approach to crime's liabilities. These costs include the stigma and rejection by significant others that can accompany state sanctions; commitment to normative values; beliefs that the legal system is just and moral; and the guilt and shame that sanctions and norm violations may induce. Research that uses vignettes finds that many of these factors are better predictors of a willingness to offend in the future than is the possibility of arrest (see Nagin 1998). Moreover, studies that allow individuals to list the potential costs of crime find an even broader array of concerns. For example, when asked about the potential costs of shoplifting a sizable proportion of subjects reported including "bad karma," being banished from a store and the potential that it would be a gateway act that would lead to more serious crime (Bouffard et al. 2010).

Perceptions about social and state sanctions probably have the greatest influence on particular people: individuals who have a considerable stake in society, have internalized norms that prohibit offenses, are embedded in networks of people who appear to follow the law most of the time and whose criminal experiences are limited to a small number of common petty offenses. In other words, these perceptions may have their greatest effects on those who have few of the motivations or opportunities that encourage crime. Consistent with this claim, research finds that strong connections to normative society discourage crime. For example, in the study of tax fraud cited earlier, general support for the legal code is negatively associated with a willingness to offend (Mehlkop and Graeff 2010).

Research on homeless youth explores the deterrent effect of a different type of criminal cost: physical harm (McCarthy and Hagan 2005). Many people use violence in responding to illegal activities (e.g., victims, bystanders, police, and other offenders) and perceptions about the dangerousness of crime are negatively related to theft, drug selling and prostitution, independent of other costs, benefits and background variables. Research that asks people to describe the potential costs of crime also finds that danger is an important concern for many individuals (Bouffard et al. 2010).

The Benefits of Crime

Crime may provide a number of benefits that range from monetary returns, the excitement or thrill of the crime, to respect or status. However, critics of the RCA approach to crime typically argue that the returns to crime are so meager that they cannot be realistically viewed as incentives. Although many offenders do not profit much from their crimes, the illegal incomes of others exceed those provided by legal employment. For example, research on Washington DC drug-sellers finds that they earned monthly incomes that were more than double the median amount earned in legal jobs (Reuter et al. 1990). Other studies find Chicago crack dealers earned an average wage of $11 an hour, a substantially higher wage than the available unskilled labor jobs; moreover, although low end “foot soldiers” often made below minimum wage, several gang leaders earned between four and eleven thousand dollars a month (Levitt and Venkatesh 2000). A related study on homeless youth reports that, on average, youth who sold drugs earned three times the average legal daily wage earned from legitimate employment (McCarthy and Hagan 2001). This investigation also documents that many offenders have reasonable expectations about crime’s financial returns: offending is positively associated with the anticipation that crime will bring greater returns than legitimate employment, but is unrelated to the belief that crime will provide “lots of money.” As well, the belief that crime will provide a financial return is related to a willingness to offend in the future (Mehlkop and Graeff 2010).

Both human and social capital contribute to success in normative activities such as employment, and criminal parallels to conventional capital may influence illegal success. For example, McCarthy and Hagan (2001) find that previous experiences contribute to illegal earnings, as does specialization, a willingness to work cooperatively with others, and the number of connections with other offenders and the support these associations provide.

Other research finds that the belief that crime will provide a psychic thrill or excitement, coolness or respect, or simply good feelings are related to offending or the willingness to offend in the future. For example, Matsueda et al. (2006) find that youth who reported that offending was exciting, would enhance their “coolness,” and that both of these were likely to happen to them, reported greater involvement in theft. McCarthy and Hagan (2005) also report a positive association between offending and the perception that crime is exciting in their research on theft and drug selling among homeless youth. They also find however, that the perception does not contribute to all types of crime: the view that prostitution is exciting is unrelated to selling sex. Nonetheless, a recent meta-analysis of the results from 13 investigations that examine the returns to crime finds that the majority of studies report positive significant associations between offending and offenders’ perceptions that crime will provide valued returns (Baker and Piquero 2010).

Game Theory

The focus of RCA theory on individual preferences downplays the extent to which decisions are influenced by the choices made by others. Yet, the decision to offend may be strongly affected by the decisions made by the police, by victims, bystanders or other people involved in crime. Game theory highlights this interactional dynamics of decisions and it offers an important tool for constructing models that make explicit assumptions about people’s preferences, behavioral options and consequences, and the connections between their choices and their expectations of other’s decisions. Although game theory has most often been used to build formal mathematical models of offending, rather than guide empirical research, the propositions of game theory models (GTM) are falsifiable and therefore, subject to empirical test. The following four examples illustrate GTM contributions.

Cressman et al. (1998) use game theory to explore the dynamics between property crime victims, thieves and the police. In their model owners have two choices: they may be passive, doing nothing to guard property, or they may engage in various protective activities such as surveillance. Police also have two choices: actively pursuing offenders or passively waiting until victims contact them. Criminal opportunists choose between theft and non-theft. Owners' increased vigilance will deter theft; however, as crime decreases, the owners' incentive for choosing passivity increases, encouraging their inactivity and increasing the returns to offenders who respond to the inactivity by increasing their involvement in crime. When police actively pursue offenders they increase the likelihood that thieves will be caught, but over time, state policing makes passivity a dominant strategy for owners and theft again becomes dominant for criminal opportunists and offending increases.

Tsebelis (1990b) adds further insight into the relationships between police activity, sanctions and crime. This game begins with the following assumptions: offenders prefer to offend when the police are elsewhere and prefer to follow the law when they are present; the police prefer to enforce the law when it is violated and to not enforce it when it is not. In other words, both players' optimal choice depends on the decision of the other, and both have an incentive to change their behavior in response to the other's actions. These assumptions lead to a situation in which an increase in the severity of sanctions initially decreases crime by diminishing its expected utility. The police have less of an incentive to enforce the law as crime drops, and people respond to their disinterest by increasing their offending. This series of moves and countermoves eventually encourages a new equilibrium in which the increase in penalties has no long-term effect on criminal activity. In short, increasing the severity of penalties has the greatest effect on police behavior, lowering their monitoring and thus decreasing the certainty of arrest. These games reveal some of the processes involved in crime at the aggregate level, and may help explain why stiffer penalties and greater police enforcement may deter crime in the short run, but not over a longer period (a finding often noted in deterrence research).

Bueno de Mesquita and Cohen (1995) expand the list of decision makers in their game, adding the government. The game assumes that people's abilities to meet their preferences are influenced by the following: their level of social status, the value obtained through legitimate opportunities a government provides, the value provided by social assistance programs, the fairness of the government, the probability of apprehension and punishment for offending, and the cost of crime. People can choose to offend or to engage in socially acceptable behavior. Government can treat citizens in two ways: a fair government allows people legitimate opportunities to earn benefits that exceed the value provided by the government's social safety net; an unfair one imposes policies that shift resources from the individual to the government and limit returns.

Solving the game reveals several conclusions. First, an individual's decision to offend is strongly influenced by his level of trust in the government's expected fairness: if people are convinced that their government will treat them unfairly, punishment may have no effect on offending. Second, with level of trust held constant, the fundamental structural features of a society increase the motivation to offend for the poor, relative to those who are wealthy. Indeed, no level of trust is sufficient to reduce crime if poverty is extreme. Reducing crime among the poor requires increases in opportunities to gain social status and the rewards provided by legitimate opportunities. Third, increasing the severity of punishment will have a small impact on the decision to offend, whereas increases in the probability of apprehension have a far greater effect in deterring crime. Fourth, improving social welfare does not discourage offending, and may actually increase it if offenders receive benefits independent of their choice between legal and illegal actions.

An extension of GTM, evolutionary game theory (EGT) relaxes the assumption that people may know the benefits of behaviors when they choose them and allows for unintended rewards that can encourage people to repeat their behaviors and can encourage others to copy them. One of the most ambitious uses of EGT to study crime is Vila and Cohen's (1993) test of hypotheses derived from Cohen and Machalek's ecological theory of expropriation. Expropriation occurs when individuals (or groups) use coercion, deception or stealth to usurp material or symbolic resources from others; in short, when they steal. Cohen and Machalek argue that stealing develops as a strategy because the social organization of production (e.g., the routine patterns of activity, the availability and distribution of resources and mode of production) creates an opportunity structure that invites invasion by non-productive strategies. People adopt or copy others' stealing when they see that it is successful.

In Vila and Cohen's game, people have only two behavioral strategies: produce or exploit. They also know the costs associated with exploitation. Vila and Cohen use computer simulations to estimate models of repeated game playing over 500 generations. The simulations demonstrate that the likelihood of stealing increases when its costs are low and returns are high, and when there is little competition among thieves. It is also more likely when changes in production encourage continual innovation in illegal strategies and when these can be easily transmitted. Consistent with Cohen and Machalek's claim, Vila and Cohen's analysis suggests that expropriative crime is a normal outgrowth of routine economic, social and productive interactions.

Multilevel RCA and Crime

While many studies using a RCA to crime emphasize individual-level preferences and behaviors, recent scholarship focusing on collective processes within neighborhoods offers a multilevel model of crime. Although theories linking individual reciprocity and social structures/collectivities date back to the foundational sociological writings of Emile Durkheim and the Chicago School, contemporary studies reflect a revival in multilevel scholarship, examining the interaction between individuals and aggregate collectivities. The multilevel approach to rational choice offers new ways to analyze the connections between individual preferences and community or neighborhood characteristics. It also provides insights into how individuals and communities encourage or discourage crime. For example, Matsueda (2013) combines a RCA with ideas about neighborhood social capital and collective efficacy to understand the connection between individual utility, macro-level neighborhood phenomena, and crime. Research suggests that social capital – the potential for relationships with others to generate trust and other resources – and collective efficacy – the willingness of residents to actively engage in social control of their neighborhood – both reduce opportunities for crime.

While collective efficacy is often associated with neighborhood structure, Matsueda (2013) contends that it originates in people's attempts to maximize their individual utility. Individuals seek to maximize utility by asking for favors and by doing favors for neighbors such as exchanging information about local politics, problems with children, property or personal victimization or other neighborhood news. This reciprocal exchange fosters obligations and mutual trust; as social capital builds, neighbors' motivations to work collectively to address local problems increases. The interaction between individual-level rational behaviors of exchange, collective efficacy and neighborhood structure represent a multi-level framework for understanding how micro-level behaviors can have macro-level outcomes.

Conclusion

The RCA to crime provides a parsimonious explanation of the process by which individuals choose to offend. It assumes that people actively choose to commit crimes and that this choice is influenced by the same factors that affect the decision to behave in ways that are consistent with legal codes. Thus, both law violating and law abiding behaviors reflect the desire to maximize utility. These choices are however, influenced by attitudes toward risk and time discounting, and estimates of an illegal opportunity's availability, costs and benefits.

Our summary of the RCA suggests that many common criticisms of it are unfounded. According to some critics, the RCA adds little to existing explanations of crime; yet, the RCA grants people more agency than explanations of offending that adopt a deterministic view to explain the effects of socialization, peer associations, and other social conditions and experiences. Other detractors claim that the predictions of the RCA are inconsistent with the reality of crime. Critics charge that the RCA describes offenders who collect all relevant information, and weigh it carefully, systematically and effectively before acting. As is clear from the above summary, the RCA to offending does not make these claims.

Other critics charge that the RCA is only applicable to specific types of illegal behavior, premeditated theft for example, but not crimes of passion. RCA recognizes that an individual's choice to commit a crime may not be rational; however, there are no reasons for assuming that particular types of crime are beyond choice, or that the RCA does not apply to the decision to commit these offenses.

By far, the most important concerns with the RCA are its assumption about preferences and decision-making. As well, findings from the experimental psychology research appear inconsistent with predictions derived from the RCA. Defenders of the conventional RCA note that although a substantial proportion of experimental subjects select options that contradict the RCA, many people's choices are consistent with it. Advocates of RCA argue that combining conventional RCA with a theory of errors can correct many of the observed inconsistencies in predictions derived from the RCA, making it superior to alternatives (e.g., prospect theory and bounded rationality).

Research on the RCA to crime suggests a number of conclusions. First, people who choose to offend, as well as those who do not, have a variety of preferences. It makes no sense to assume that all offenders share the same preferences, or that one preference (e.g., the thrill of crime or its material returns) dominates in offending decisions. Preferences vary, crime provides an array of returns, and people differ in their assessments of crime's costs and benefits. Second, the decision to offend may be strategically chosen given the decisions made by others; its utility may change – from being a dominant to a nondominant strategy – depending on the strategic choices of others. Although people often assume that offenders and the police (as well as victims and other agents of social control) have unique preferences that are oppositional, independent and stable, game theory and deterrence research suggest that each group's preferences can overlap, and may change in response to the actions of the other, as does their choice of strategies.

Related Entries

- ▶ Bayesian Updating and Crime
- ▶ Certainty, Severity, and Their Deterrent Effects
- ▶ Cognitive Forensics: Human Cognition, Contextual Information, and Bias
- ▶ Deterrence
- ▶ Deterrence: Actual Versus Perceived Risk of Punishment
- ▶ Deterrence of Tax Evasion
- ▶ Econometrics of Crime
- ▶ Establishing Causes of Offending in Longitudinal and Experimental Studies
- ▶ History of Criminological Theories: Causes of Crime
- ▶ Integrating Rational Choice and Other Theories
- ▶ Offender Decision Making and Behavioral Economics
- ▶ Rational Choice and Prospect Theory
- ▶ Rational Choice, Deterrence, and Crime: Sociological Contributions

Recommended Reading and References

Baker T, Piquero AR (2010) Assessing the perceived benefits-criminal offending relationship. J Crim Just 38:981–987

Becker GS (1968) Crime and punishment: an economic approach. J Polit Econ 76:169–217

Bianco WT, Ordeshook PC, Tsebelis G (1990) Crime and punishment: are one-shot, two-person games enough? Am Polit Sci Rev 84:569–586

Bouffard JA, Exum ML, Collins PA (2010) Methodological artifacts in tests of rational choice theory. J Crim Just 38:400–409

Bueno de Mesquita B, Cohen LE (1995) Self-interest, equity and crime control: a game-theoretic analysis of criminal decisions. Criminology 33: 483–517

Clarke RV, Felson M (1993) Introduction: criminology, routine activity, and rational choice. In: Clarke RV, Felson M (eds) Routine activity and rational choice, advances in criminological theory. Transaction, New Brunswick

Cohen LE, Machalek R (1988) A general theory of expropriative crime: an evolutionary ecological approach. Am J Sociol 94:465–501

Cornish DB, Clarke RV (eds) (1986) The reasoning criminal: rational choice perspectives on offending. Springer, New York

Cox GW (1994) A note on crime and punishment. Public Choice 78:115–124

Cressman R, Morrison WG, Wen JF (1998) On the evolutionary dynamics of crime. Can J Econ 31:1102–1115

Dahlbäck O (2003) Analyzing rational crime-models and methods. Kluwer, Dordrecht

Ehrlich I (1974) Participation in illegitimate activities: an economic analysis. In: Becker GS, Landes WM (eds) Essays in the economics of crime and punishment. Columbia University Press, New York

Eide E (1994) Economics of crime: deterrence and the rational offender. Elsevier Science, Amsterdam

Gambetta D (1994) Inscrutable markets. Rational Soc 6:353–368

Gibbs J (1994) A theory about control. Westview, Boulder

Grasmick HG, Bursik RJ Jr (1990) Conscience, significant others, and rational choice: extending the deterrence model. Law Soc Rev 24:837–861

Grogger J (1991) Certainty vs. severity of punishment. Econ Inq 29:297–310

Grogger J (1998) Market wages and youth crime. J Labor Econ 16:756–791

Hoffman R (2001) Mixed strategies in the mugging game. Ration Soc 13:205–212

Horney J, Marshall IH (1992) Risk perceptions among serious offenders: the role of crime and punishment. Criminology 30:575–593

Laver M (1982) The crime game. Martin Robertson, Oxford

Levitt SD, Miles TJ (2006) Economic contributions to the understanding of crime. Annu Rev Law Soc Sci 2:147–164

Levitt SD, Venkatesh SA (2000) An economic analysis of a drug-selling gang's finances. Q J Econ 115:755–789

Matsueda RL (2013) Rational choice research in criminology: a multi-level approach. In: Wittek R, Snijders T, Nee V (eds) Handbook of rational choice social research. Stanford University Press, Palo Alto. http://www.sup.org/book.cgi?id=21110

Matsueda RL, Kreager D, Huizinga D (2006) Deterring delinquents: a rational choice model of theft and violence. Am Sociol Rev 71:95–122

McCarthy B (2002) New economics of sociological criminology. Annu Rev Sociol 28:417–442

McCarthy B, Hagan J (2001) When crime pays: capital, competence and criminal success. Soc Forces 79:1035–1059

McCarthy B, Hagan J (2005) Danger and the decision to offend. Soc Forces 83:1065–1096

Mehlkop G, Graeff P (2010) Modelling a rational choice theory of criminal action: subjective expected utilities, norms, and interactions. Ration Soc 22:189–222

Nagin DS (1998) Criminal deterrence research at the outset of the twenty-first century. Crime Justice Rev Res 23:1–42

Nagin DS, Paternoster R (1993) Enduring individual differences and rational choice theories of crime. Law Soc Rev 27:467–496

Paternoster R (2010) How much do we really know about criminal deterrence? J Crim Law Criminol 100:765–824

Pezzin L (1995) Earnings prospects, matching effects, and the decision to terminate a criminal career. J Quant Criminol 11:29–50

Piliavin I, Thornton C, Gartner R, Matsueda RL (1986) Crime, deterrence, and rational choice. Am Sociol Rev 51:101–119

Reuter P, MacCoun P, Murphy P (1990) Money from crime: a study of the economics of drug-dealing in Washington, DC. RAND, Santa Monica

Schmidt P, Witte AD (1984) An economic analysis of crime and justice: theory, methods, and applications. MIT Press, Cambridge

Tsebelis G (1990) Penalty has no impact on crime: a game theoretic analysis. Ration Soc 2:255–286

Uggen C, Thompson M (2003) The socioeconomic determinants of ill-gotten gains: within-person changes in drug use and illegal earnings. Am J Sociol 109:146–185

Vila BJ, Cohen LE (1993) Crime as strategy: testing an evolutionary ecological theory of expropriative crime. Am J Sociol 98:873–912

Viscusi WK (1986) The risks and rewards of criminal activity: a comprehensive test of criminal deterrence. J Labor Econ 4:317–340

Weissing F, Ostrom E (1991) Crime and punishment: further reflections on the counterintuitive results of mixed strategy equilibria games. J Theor Polit 3:343–350

Rational Choice, Deterrence, and Crime: Sociological Contributions

Maria S. Grigoryeva and Ross L. Matsueda
Department of Sociology, University of
Washington, Seattle, WA, USA

Synonyms

Communities; Deterrence; Incarceration; Informal social control; Perceptions of risk; Rational choice; Re-integration; Specific deterrence; Stigma

Overview

This entry reviews sociological contributions to the study of rational choice, deterrence, and crime. It reviews empirical research on the deterrence question, including macro-level studies of aggregate crime rates, micro-level studies of individual perceptions of sanction risk, and experimental studies of specific deterrence and domestic violence. It then shows the relevance of sociological research for specifying the broader context of punishment, which reveals negative externalities of mass incarceration in the USA, such as the pernicious stigmatizing effects of incarceration, the undermining of the legitimacy of the law within disadvantaged communities, and the loss of community cohesion and social capital. Such negative externalities produce criminogenic social conditions, which, in turn, undermine deterrent effects.

Rational Choice, Deterrence, and Crime

The deterrence doctrine is rooted in the writings of the classical school of criminology, and in particular, those of Caesare Beccaria and Jeremy Bentham. Influenced by the prevailing ideas of the enlightenment period, Beccaria argued that human beings freely enter into a social contract with the state, relinquishing some liberty in exchange for the protection of their individual rights. In exchange, individuals would give the state the right to punish those who violated criminal laws, consisting of a written record of the terms of the social contract. Punishments would be swift, certain, and just severe enough to offset the pleasures of crime and thereby deter the public from criminal behavior. The underlying behavioral assumption is that human beings are rational and hedonistic, and weigh the pleasures and pains associated with different behaviors. When the pleasures outweigh the pains associated with criminal behavior, individuals will violate the law. These ideas underlie the system of justice used under Anglo-Saxon law, and motivated early sociological studies of deterrence until a formal expected utility model was specified by the economist Gary Becker.

Expected Utility Models of Crime

Becker (1968) made a seminal contribution to the study of deterrence by specifying a neoclassical theory of criminal behavior. Arguing that the same economic principles explaining decisions made by firms and household members should also explain criminal behavior, Becker (1968:177) drew on von Neumann and Morgenstern's expected utility theory of risky decisions under uncertainty to specify a simple utility function for committing crimes:

$$E(U_C) = (1 - p_c)\ U(R) + p_c U(R - C) \qquad (1)$$

where $E(U_C)$ is the expected utility of crime, p_c is the probability of getting arrested and punished, $(1 - p_c)$ is the probability of getting away with crime, R is the return (both monetary and psychic) from crime, C is the cost of punishment (e.g., a fine or prison sentence), and U is a utility function translating punishments and rewards to a common metric. The expected utility model assumes that individuals have complete and transitive preference orderings for all possible decision outcomes. If we ignore noncriminal behavior or assume that the expected utility from noncrime is absorbed in C as an opportunity cost, we can specify that a crime will occur when $E(U_C) > E(U_N) = 0$, so that from equation (1), a crime will occur when the following holds:

$$U(R) > p_c U(C) \tag{2}$$

That is, when the returns to crime exceed the punishment, weighted by the probability of detection, an individual will commit a crime. The policy implication here is that by increasing the certainty and severity of punishment, the probability of crime will be reduced. Crime can also be reduced by lowering the rewards to crime – by defending public spaces through increasing surveillance, employing security guards, and using technological advances in metal detection, alarms, locks, fences, and the like (see McCarthy 2002). Historically, following Becker's (1968) work, most microeconomic research on crime has focused on the policy implications of increasing the certainty and severity of punishment.

Criminal decisions, of course, consider the utility of noncrime as well as crime, and therefore, one might specify a utility function for noncriminal activity (e.g., Bueno de Mesquita and Cohen 1995):

$$E(U_N) = p_i U(I) + (1 - p_i)\, U(W) \tag{3}$$

where I is income (returns to conventional activity), p_i is the probability of obtaining I (through having high social status, resources, or talent), and W is welfare or the social safety net for those who cannot obtain I (i.e., $p_i = 0$). Combining equations (1) and (3) yields:

$$E(U_C) = (1 - p_c)[U(R) - p_i U(I) + (1 - p_i) U(W)] + p_c U(R + W - I - C) \tag{4}$$

If follows that a crime will be committed when:

$$(1 - p_c)[U(R) - p_i U(I) + (1 - p_i) U(W)] + p_c U(R + W - I - C) > 0 \tag{5}$$

From a policy point of view, the probability of crime can be altered not only through criminal justice policies that increase the certainty and severity of punishments or that change defensible space, but also through policies that increase the utility from noncriminal activity. For example, job training, higher education, and other programs to enhance human and social capital may reduce the attractiveness of crime by increasing p_i, the probability of obtaining a desired income from legitimate activities. Returns to conventional activity include not only income but also social status and prestige, self-esteem, and happiness; policies that increase these quantities by inculcating strong commitments to conventional institutions may help reduce crime.

Limited Rationality and Crime

Criminologists, following cognitive psychologists, have questioned the assumptions of the expected utility model, arguing that criminals are unlikely to have access to full information and unlikely to conduct complex calculations necessary to maximize expected utility. For example, the utility maximization theory is commonly criticized for assuming (1) that the actor knows the probabilities of certainty and severity of crime, and (2) that criminals act on a rational calculation, rather than by impulse, either under the influence of alcohol or in responding to opportunities. In responding to this criticism of rational choice and crime, Cook (1980) drew on Herbert Simon's work on bounded rationality, and suggested that criminals may use standing decisions or rules of thumb in making decisions about crime – particularly in situations of high stress, emotion, or inebriation. Moreover, Cook argued that rational choice theories of crime need a theory of the communication of threats, and pointed to the role of the media, presence of law enforcement, personal experience, and observation of peers as important ingredients of such a theory.

Clarke and Cornish (1985) attempted to integrate rational choice with conventional criminological theories, such as social learning theory, and developed a model in which actors' pursuit of general needs and previous learning (experiences with crime, contact with police, conscience and morality, self-perception, and foresight) led to evaluation of legitimate and illegitimate solutions to produce a “readiness” to commit burglary. Readiness may lead to a decision to commit burglary, which is followed by a proactive assessment of potential middle class targets based on accessibility, police patrols, security, cover, affluence, and no one being at home. The evaluation of target accessibility is consistent with routine activities theory, which hypothesizes that criminal events result from the intersection of motivated offenders, suitable targets, and the absence of capable guardians.

Empirical Research on Rational Choice and Deterrence

Empirical research on rational choice and deterrence has made considerable methodological strides, moving from cross-sectional studies to longitudinal and panel designs, grappling with the identification problem, and considering experimental designs. Beginning with Jeremy Bentham, deterrence theory has traditionally distinguished between general and specific deterrence. General deterrence refers to an act or threat of punishment deterring the general public from committing crimes; specific deterrence refers to the special case of deterring the punished individual from committing future crimes. While these two forms of deterrence are theoretically distinct, may influence people through different mechanisms, and consequently may imply distinct punitive policies, they are often hard to tease apart empirically. Of the empirical work reviewed here, aggregate research – especially studies examining the effect of new punishment policies on rates of offending – often conflates the effects of specific and general deterrence (Durlauf and Nagin 2011). Individual-level studies can be more explicit about separating out the effects of each. We focus here on sociological and criminological studies, as econometric studies are reviewed elsewhere in this volume.

Studies Using Aggregate Data

Over the past four decades, a spate of empirical studies of deterrence by criminologists, sociologists, and economists have used aggregate data to examine whether individuals are deterred by the certainty and severity of punishment (for a review, see Nagin 1998), often employing instrumental variables to identify reciprocal effects between rates of imprisonment and rates of crime. Durlauf and Nagin (2011) provide an excellent review of empirical work on deterrent effects of incarceration, summarizing recent findings as well as identifying important issues in aggregate research on deterrence. Studies of the severity of sanctions – often operationalized as differences in the length of incarceration – have found only modest deterrent effects of severity of punishment. Studies of the deterrent effect of capital punishment – considered the most severe of all state sanctions – have historically yielded disparate results, with recent evidence that findings of deterrent effects of capital punishment were based on incorrect modeling procedures (Durlauf et al. 2012). However, there is evidence of significant deterrent effects of certainty – operationalized as greater police presence on the street, and more aggressive policing practices.

While aggregate research examining change in the crime rate has contributed much to the understanding of the deterrent properties of various policy changes, it has been criticized for several shortcomings. First, aggregate research is unable to distinguish the effect of deterrence from the effect of incapacitation on crime (Durlauf and Nagin 2011). Second, this research has not accounted for heterogeneity (especially at the state level) of policy adherence. This is problematic because systems of formal control differ not only at the state level, but also by county, city, and sometimes community. Additionally, individuals may be most aware, and concerned with, local formal punishment and detection policies rather than those at the county or state level (Apel 2012). Third, most aggregate studies do not distinguish between awareness of deterrence policy, and often assume that individuals have perfect information with which to assess risk of being caught and punished (Durlauf and Nagin 2011). Recent research indicates that there are significant gaps in policy awareness, with experienced offenders being more informed than others (Matsueda et al. 2006; Apel 2012), showing a need to treat utility maximization and possession of information as separate effects of deterrence policy on observed behavior. The next section reviews research on individuals' subjective evaluation of the severity and certainty of punishment, which has emerged as one fruitful alternative to aggregate deterrence research.

Studies Using Survey Data of Individuals

Aggregate tests of the deterrence hypothesis assume that actors know the objective certainty of arrest and imprisonment (Nagin 1998). By contrast, *subjective* expected utility models relax this assumption, replacing the single known *objective* probability with a distribution of *subjective* probabilities. Subjective utility models are still rational models because the statistical mean of the subjective probability distribution is assumed to fall on the value of the objective probability (Nagin 1998). Empirical research from a subjective expected utility framework uses survey and vignette methods to measure perceived risk of punishment directly from respondents, rather than inferring it from behavior through the method of revealed preferences.

Early empirical research by sociologists used cross-sectional data and found small deterrent effects for certainty of punishment but not for severity (e.g., Williams and Hawkins 1986). Respondents who perceive a high probability of arrest for minor offenses (like marijuana use and petty theft) report fewer acts of delinquency. Such research has been criticized for using cross-sectional data in which past delinquency is regressed on present perceived risk, resulting in the causal ordering of the variables contradicting their temporal order of measurement.

To address this criticism, sociologists have turned to two-wave panel models and found, for minor offenses, little evidence for deterrence (perceived risk had little effect on future crime) and strong evidence for an experiential effect (prior delinquency reduced future perceived risk) (see Williams and Hawkins 1986; Paternoster 1987). Piliavin et al. (1986) specify a full rational choice model of crime, including rewards to crime as well as risks, and find, for serious offenders, that rewards exert strong effects on crime, but perceived risks do not.

Recent longitudinal survey research has used more sophisticated measures of risk, better-specified models, and better statistical methods. Matsueda et al. (2006) specify two models based on rational choice. First is a Bayesian learning model of perceived risk, in which individuals begin with a baseline estimate of risk, and then update the estimate based on new information, such as personal experiences with crime and punishment or experiences of friends. Second is a rational choice model of crime, in which crime is determined by prior risk of arrest, perceived opportunity, and perceived rewards to crime, such as excitement, kicks, and being seen as cool by peers. Using longitudinal data from the Denver Youth Survey, Matsueda et al. (2006) find support for both hypotheses: perceived risk conforms to a Bayesian updating process, and delinquency is determined by perceived risk of arrest, rewards to crime, perceived opportunities, and opportunity costs. A recent study by Anwar and Loughran (2011) extends the risk updating model to high-risk offenders, finding further evidence for a Bayesian-style learning model. The study also finds that experiencing a deterrent signal about a specific type of crime results in a generalized posterior where all crimes are considered more risky.

In addition to risk updating, perceptual research has focused on other aspects of individual assessment of risk, such as interpretation of the risk signal. Loughran et al. (2011) build on Lawrence Sherman's distinction between *level* and *ambiguity* of sanction certainty, and test whether individuals are ambiguity averse – that is, interpret uncertainty as an added risk. The authors find support for ambiguity aversion in assessment of sanction certainty, but also observe ceiling effects, where extremely risky and uncertain situations will be perceived as slightly less risky than certain same-level risk. Loughran et al. (2011) additionally find that ambiguity aversion holds only for property crimes, and not face-to-face crimes, noting that property crimes may be more amenable to deterrence.

Vignette surveys examine perceived deterrent effects within a richer depiction of a scenario, which mimics actual situations of crime. Here, a crime scenario is depicted in a short written paragraph, with specific elements of the scenario – such as the presence of police or witnesses and potential monetary returns – randomly rotated across vignettes. Respondents are asked to assess the probability of getting caught or obtaining rewards from the crime and also asked their intentions of committing the crime depicted. By asking their future intentions of committing crime, vignette studies resolve the causal order problem, under the assumption that intentions are good predictors of actual behavior. A weakness of vignettes is the potential for a response effect: respondents who report high risk of arrest may be unlikely to admit to an intention to commit the crime due to social desirability effects. Vignette studies of deterrence and rational choice generally find robust effects of deterrence: certainty has a substantial effect on criminal intentions, while severity has modest effects. This holds for tax evasion, drunken driving, sexual assault, and corporate crime (see Nagin 1998, for a review).

A recent line of research in clinical psychology has focused on linking cognitive developmental differences between adults and adolescents (including middle and late adolescence) to differential decision making and criminal participation. New analyses suggest that cognitive immaturity in older adolescents and young adults, including a lesser ability to engage multiple regions of the brain to problem-solve coupled with increased dopamine activity, causes them to be more impulsive, more vulnerable to peer influence, and more likely to discount the future than adults who are in their mid-twenties and older (see Steinberg 2009). This research provides a developmental basis for differences in adolescent behavior previously noted by sociologists and criminologists, such as committing crime in groups, and discounting future punishment costs. Furthermore, these findings suggest that individuals in their late teens and early twenties may still be in a developmental stage that is not as amenable as adults to current policies of deterrence, which may explain differential rates of young adult participation in criminal activity. Such research has obvious policy implications for the distinction between juvenile and adult criminal justice.

The recent advances in perceptual deterrence have significantly added to the specification of the process by which individuals understand and respond to the policies of deterrence. However, sociological research on consequences of punishment has raised awareness of additional mechanisms that have the potential to alter individual utility functions and undermine the efficiency of deterrence. The next sections outline what is known about the ways in which the current system of punishment in the USA can affect individual utility functions and behavior.

Studies of Specific Deterrence Using Experiments on Domestic Assault

Research on specific deterrence faces the daunting task of addressing non-equivalent treatment (e.g., incarceration) and control groups (e.g., probation). Offenders who are incarcerated are likely to be more crime prone than offenders who are not incarcerated, and thus, unobserved heterogeneity is likely to bias comparisons between incarcerated and non-incarcerated offenders. A line of research that addresses this problem uses an experimental design to examine the specific deterrent effects of arrest for domestic assault on recidivism rates. Sherman and Berk (1984) conducted such an experiment by randomly instructing police to arrest domestic assault offenders versus merely counseling others. Contrasting this with a labeling effect, in which arrest may stigmatize offenders and amplify deviance, Sherman and Berk (1984) found a significant deterrent effect: arrested offenders were less likely to re-offend than non-arrested offenders. As a result, laws mandating arrest for domestic assault offenders spread across the country. Because this study was limited to one city (Minneapolis) and suffered some methodological problems – such as police departing from random assignment – the National Institute of Justice funded a set of replication studies across the USA. Generally speaking, those studies did not replicate the specific deterrent effect: overall, arrested offenders were not significantly less likely to recidivate. They did find, however, that arrest deterred a subpopulation of offenders – namely, those who had stronger stakes in conformity, such as marriage and employment (e.g., Berk et al. 1992). Thus, arrest for domestic assault may deter future assaults for those individuals who are committed to conventional society, and who consequently, have more to lose by arrest and likely incarceration. Commitment to society constitutes a key component of the broader context of punishment.

The Social Context of Deterrent Effects

Sociologists critical of using policies of deterrence as a panacea for the crime problem have emphasized the societal context in which punishment is carried out. Most research on the broader contexts of crime and punishment has not adopted a rational choice framework. Nevertheless, such research has important implications for a rational choice theory of crime, and in particular, for identifying limitations of deterrent effects. A key concept here is stigma. Research has found that small amounts of informal stigma, which is analogous to public shaming, may enhance specific deterrent effects. But widespread or permanent stigmatization of distinct groups of people can have substantial negative externalities, at times amplifying rather than deterring crime. This research has its roots in sociological labeling theory, which stresses the negative consequences of the criminal justice system, and Braithwaite's (1989) theory of reintegrative shaming, which argues that stigmatization without reintegration can amplify the crime problem. We use these ideas to organize research on stigma and crime.

Stigmatization and Crime

Goffman (1963) described stigma as an attribute that spoils an individual's identity, evokes negative stereotyping by others, and causes the individual to adopt coping strategies when confronted by "normals." Research and theory on rational choice and deterrence has shown that a specific mechanism that potentially enhances deterrent effects is informal sanctioning by family members and peer groups. That is, if the arrested or convicted offender is stigmatized by significant others – resulting in a loss of social status – the specific deterrent effect of the arrest or conviction will be strengthened (see Paternoster 1987). This effect could also enhance general deterrence as other members of society observe the stigmatizing effects of punishment. Rebellon, Piquero, Tibbets, and Piquero (2010) show that fear of arrest deters mainly through the expected shame from friends and family of the potential offender. Stigma, however, is a *relative* concept; its negative effects are dependent on its distribution in a population. Having a disfiguring scar may be enormously stigmatizing in a population whose members have no scars, but less so in a population of army veterans in which scars are frequent. Moreover, the deterrent properties of stigma may have within-individual thresholds: if an individual is severely and permanently stigmatized, any future stigmatization or shaming is rendered ineffective. In the following section, we outline how policies of deterrence may be undermined when stigmatization exceeds such population and individual-level thresholds.

Stigma and Population Thresholds

Because stigma is a relative term, its negative effects may exhibit a threshold effect or tipping point in a given group or population. When punishment is a relatively rare event, the stigmatizing effects are dramatic. As the incarceration rate of a subpopulation increases, the status of being a felon becomes more commonplace, and the effects of stigma shrink (Nagin 1998). If all members of the subpopulation receive felon status, stigma reduces to zero. Even if stigma reinforces a deterrent effect, that effect is subject to a tipping point, after which the effect declines. Thus, it is possible that mass incarceration has undermined the stigmatizing effects of having a criminal record, as more and more members of a given group – young African-American men from disadvantaged backgrounds – share the status of having a criminal record. For example, Clear (2007) and Anderson (1999) describe the spatial concentration of incarceration in poor African-American communities, where offenders cycle in and out of neighborhoods in which they resided before initial contact with the law. In such neighborhoods, incarceration may become normative, and felon status may lose its stigmatizing effect. From a rational choice perspective, the cost of incarceration has diminished in such communities, which attenuates deterrent effects.

Within-Individual Thresholds: Shaming, Stigmatization, and Reintegration

Even without reaching population-level thresholds, the effects of stigma can change from deterrent to criminogenic if stigma becomes a serious and permanent experience for an individual. The concept of permanent, self-reinforcing stigma lies at the heart of sociological labeling theories, which posit that negative labeling can at times amplify, rather than deter, crime. For example, minor forms of deviant or mischievous behavior, viewed as play by children, may be seen as bad, evil, or portending of more serious deviance by the adult community. Adults often respond by labeling the child as “evil” or “bad” and informally punishing the child. Repeated negative interactions with the community may leave the child cut off from conventional peers, stigmatized as a bad kid, and caught up with the juvenile justice system, ultimately leading the individual to adopt the stigmatizing label as permanent trait, viewing themselves as a criminal and engaging in criminal behavior. Labeling theorists emphasize the negative effects of labeling and pose the counterfactual: might the youth have been better served if their initial minor forms of delinquency had been treated as mere mischief and avoided the process of negative labeling and deviance amplification?

Braithwaite (1989) elaborated on deviance amplification with his theory of reintegrative shaming. Consistent with labeling theory, Braithwaite argues that often the punitive sanctions of the legal system amplify rather than deter crime by stigmatizing and segregating the offender from conventional society. In contrast, a system of reintegrative shaming would shame the offender – via informal disapproval by significant others (family, friends, and other community members) who are respected and trusted by the offender – followed by a program of reintegration of the offender back into conventional society. Braithwaite argued that punishment is most efficient when administered within a context of respect and love – rather than anger, retribution, and rejection – and within a context in which the offender is welcomed back into society. Consequently, for the offender, the threat of future shaming will maintain a deterrent effect, as the offender has something important – his or her renewed status – to lose.

Empirical research finds some support for ideas of restorative justice and reintegrative shaming. For example, the use of restorative justice as a model for conflict resolution and restitution may result in greater satisfaction for offender and victims compared to a traditional court-centered approach. Offenders who received restorative justice may also have lower recidivism rates than the offenders who receive court-centered justice (Braithwaite 1999). Informal shaming may be a key cost of offending and enhance deterrent effects (e.g., Rebellon et al. 2010).

Theories of labeling and reintegrative shaming provide a theoretical framework within which to review recent sociological research into the social context of punishment. The key theoretical point is that increases in the certainty and severity of punishment may produce a deterrent effect, but in the absence of policies of reintegration, may also produce massive stigmatization of a population, a negative externality that undermines deterrence. This entry emphasizes the unanticipated negative consequences of punishment, as exemplified in the recent trends of mass incarceration and severe problems of reintegration of offenders in the USA. We focus on the USA as an empirical example because massive incarceration over the last four decades has generated a large population of permanently stigmatized offenders. Problems of stigma and reintegration can become particularly acute if the same class (race, gender, and social class) of offenders is segregated from conventional society, perhaps over generations, resulting in a permanent class of individuals who have little invested in conventional activities, and consequently perceive criminal participation as relatively more rewarding. Recent research suggests this is precisely the situation in contemporary America.

Between 1972 and 2000, the US incarcerated population increased by six times. Western (2006) estimates that, among the male cohorts of 1965–1969, one in five black men had experienced prison by their early thirties. The number for white men was 3 out of 100. Moreover, he reports that nearly 60 % of black men who had dropped out of school were incarcerated by 1999. Clearly, the USA is characterized by mass incarceration, which disproportionately affects disadvantaged young black men. These trends have important consequences for examining rational choice, deterrence, and crime.

Sociological Research on the Individual Consequences of Punishment

Individuals who have been arrested, convicted, and incarcerated undergo the stigmatizing effects of the criminal justice system, including being handcuffed, appearing in court in jail uniforms, and being incarcerated away from family and friends. This is only the beginning of the stigmatizing process that hampers an offender's reintegration. Once a felon has paid his or her debt to society, he or she faces additional impediments to reentering society and refraining from future crime. Such obstacles, discussed below, suggest that the US legal system, in conjunction with other institutions, such as labor market and political institutions, operates to stigmatize and segregate the offender from conventional society, working at odds with Braithwaite's (1989) call for reintegrative shaming.

A criminal record reduces – sometimes permanently – an individual's employment and earning potential. Empirical research suggests that ex-felons are less likely to be employed, and when employed, tend to be earn lower wages than their labor market counterparts. These results, however, could merely reflect preexisting differences between individuals with and without a criminal record, such as differences in human capital. Consequently, Pager (2007) used a quasi-experimental audit study, in which matched pairs of individuals applied for jobs. She randomly assigned felon status to one of the pair, then reversed that status to obtain the counterfactual condition – what would happen if the felon statuses of the pairs were switched? – and counted the number of callbacks from employers. She found that non-felons were twice as likely as ex-felons to get a callback. That effect was greater for black ex-felons.

Furthermore, a felony conviction revokes the right to be employed in several occupations, and can be grounds for denial of such welfare programs as subsidized housing and financial aid to mothers with children (Wakefield and Uggen 2010). This exacerbates the precarious financial situation of most ex-felons. Fees and fines administered by criminal courts often create accrued debt, which undermines the solvency of economically marginalized criminals. Thus, as personal wealth and employment opportunities and earnings diminish, opportunity costs to crime decrease and illegal markets become more attractive. These persistent declines in returns to legal employment may also reduce the cost of future arrest or incarceration. At the extreme, if felony status makes legal employment virtually impossible, an ex-felon will have little to lose by re-offending.

In most states in the USA, felony status results in a long-term, and sometimes permanent, loss of voting rights (Manza and Uggen 2006). Because felon disenfranchisement disproportionately affects Democratic turnout, it is possible that recent close elections may have turned out differently had felons been allowed to vote (see Manza and Uggen 2006). Moreover, loss of voting rights undermines civic participation, an important source of integration into conventional society.

These negative consequences of punishment can have spillover effects on other commitments to conventional society. Incarceration, and the reduced economic opportunities it spawns, is associated with lower marriage rates (Western 2006) and difficulties providing for children, each of which undermines commitments to the family unit.

Without strong attachments to work, civic, and family life, ex-felons are unlikely to reintegrate into mainstream society. Strong informal ties to conventional society have been shown to dissuade individuals from criminal behavior (e.g., Sampson and Laub 1993). It follows that weak ties to the labor market, family, and civic institutions reduce the rewards from conventional activities, and in turn decrease the opportunity costs for crime. Thus, the contemporary system of punishment in the USA appears to stigmatize offenders upon release and long after their official punishment is meted out, cutting them off from participation in conventional society.

Sociological Research on the Consequences of Punishment for Communities

Incarceration, particularly on a massive scale, can also have negative externalities for communities. When punishment is repeatedly and disproportionately applied to members of specific communities, it can alienate not only the individuals punished, but the entire community, who may begin to bear the burden of disproportionate punitive targeting. In the long run, such punitive trends may cause community members to view the legal system with a jaundiced eye, ultimately undermining the legitimacy of the system. This may occur in the absence of discrimination against community members; however, when there is evidence of discrimination, such as with racial profiling by police, the problem quickly escalates.

In his ethnography of youth culture in a predominantly African-American disadvantaged inner city, Anderson (1999) documents how perceptions of hostile and racist police practices, coupled with systematic barriers to employment and frequent contact with the law, drive community members to reject conformity with the law, and adopt a view of formal systems of sanctions as unfair and illegitimate. Consistent with the threshold view of the effects of stigma, norms of informally shaming offenders not only disappear, but are replaced by the "code of the street" – a set of norms that celebrates incarceration as a right-of-passage into manhood, and accepts interpersonal violence as a reasonable strategy for obtaining personal safety and social approval. This redefines formal punishment from a social cost within a community to a social reward, undermining deterrent effects. Because the legal system and police are often the principal contact disadvantaged residents have with conventional institutions, alienation from police may lead to distrust of other institutions.

In such disadvantaged neighborhoods, with a high prevalence of incarceration, low rates of conventional economic success, and meager community resources, young people will tend to underinvest in education and careers, which lowers the opportunity costs to crime. Facing grim economic prospects, they will tend to discount the future that, by all accounts, will be bleak (Anderson 1999). This discounting, which becomes part of the code of the street, induces an attitude of "live and die in the moment," undermines delayed gratification and long-term planning, and blunts the deterrent effect of punishment.

High incarceration rates in specific communities result in large proportions of residents cycling in and out of prison, which may disrupt the local community (e.g., Clear 2007). On the one hand, removing an offender who has wreaked havoc on other residents, is isolated from others, and makes few contributions to the community may have a positive effect on the community as a whole. On the other hand, despite having committed a crime, local offenders may also have been fathers, neighbors, and friends, and thereby have been interwoven into the fabric of the local community. Their removal from the community through incarceration reduces social capital by eliminating the social ties they maintained within the community's social network. The result is diminished exchange relationships, a loss of information potential, and weakened social norms. Ironically, this reduction in social capital may be associated with reductions in informal social control, such as collective efficacy, which in turn is associated with higher rates of crime and incarceration (e.g., Sampson et al. 1997). Such informal control includes interdependencies, trust, and mechanisms of control involving community reputations, all of which are destroyed by the constant uprooting of community members. Thus, a community can get caught up in a pernicious feedback loop in which incarceration undermines social control, which increases crime and incarceration, further undermining social controls. Furthermore, an influx of returning ex-prisoners into a disadvantaged community may disrupt local legal economies (shops and restaurants), reducing opportunity costs to crime, and may increase the community's crime rate (e.g., Clear 2007). The devastating effects of mass incarceration at the community level undermine the salience of deterrence for all members, including those who would otherwise be at low risk of criminal behavior.

In sum, research on the social context of punishment and crime suggests a number of negative externalities of incarceration, particularly when carried out on a massive scale. Punitive policies in the USA may have helped deter crime, but have also produced a host of negative externalities – increasing stigmatization, hampering the reintegration of the offender, reducing the opportunity costs of crime, and undermining both social capital and informal social control of communities – all of which may have exacerbated the crime problem and reduced the efficiency of deterrence.

Conclusions and Directions for Future Research

Sociological theories of rational choice and deterrence have made strong contributions to our understanding of criminal behavior and punitive policies. Rooted in rational choice and utility maximization theories of individual behavior, aggregate studies of deterrence have been augmented by individual-level studies of decision making and cognition. In general, research suggests that deterrence has an important role in the causes of crime: based on a variety of research designs, the certainty of punishment appears to be negatively associated with crime. Research on the marginal deterrent effect of increasing the severity of punishment, including the death penalty, has been more equivocal, with effects typically modest in size or statistically indistinguishable from zero.

Increases in punishment rates, particularly mass incarceration, occur in a social context, which may lead to unanticipated consequences and negative externalities. Research in sociology and criminology suggests that a combination of mass incarceration with a failure to reintegrate offenders into the community has pronounced negative collateral consequences, including a reduction in the effectiveness of current deterrence policies. Large numbers of ex-prisoners have been stigmatized, blocked from succeeding in the labor market, prevented from civic participation, and generally isolated from conventional society, resulting in fewer opportunity costs for re-offending. The result is high rates of recidivism, and a constant cycling of predominantly disadvantaged African-American men in and out of the prison system. The removal of large numbers of men from disadvantaged communities may undermine the social organization, social capital, and informal control of those communities. Weakened community social controls combined with fewer opportunity costs for offending likely perpetuates the cycle of incarceration, stigmatization, and re-offending. As Braithwaite (1989) has argued, such a cycle will only lead to spiraling crime rates and diminished deterrent effects of stigmatized individuals; breaking the cycle requires reducing stigma, introducing community shaming, and increasing reintegration of the ex-offender into mainstream society.

Future research is needed to tease apart these countervailing processes of deterrence, stigmatization, and reintegration. Because each of these processes operates endogenously, research is needed that considers their joint nonlinear relationships. This is tricky because such processes will be fraught with feedback loops, as noted above, as well as cross-level effects: an individual decision to commit crime and risk incarceration changes the social context by providing criminal role models, reducing the certainty of arrest, and undermining social capital and informal control – all of which increases the likelihood of crime and incarceration. Research designs using experimental interventions, which provide instrumental variables for estimating treatment effects, may be useful here. Short of such costly designs, econometric methods of analyzing social interaction effects may help, although the identification problems are daunting. Another useful approach would be to estimate the models, and tease out the underidentified parameters using simulations based on game theory (see McCarthy 2002). Regardless of research design, we believe that the next generation of research on rational choice, deterrence, and crime must examine the dynamic interactions between criminal decision making and the broader social context within which decisions are embedded.

Related Entries

- ▶ Bayesian Updating and Crime
- ▶ Certainty, Severity, and Their Deterrent Effects
- ▶ Communities and the Police
- ▶ Crime Preventive Effects of Incapacitation
- ▶ Deterrence
- ▶ Deterrence: Actual versus Perceived Risk of Punishment
- ▶ Deterrent Effect of Imprisonment
- ▶ Estimating the Effects of Incapacitation
- ▶ History of Restorative Justice
- ▶ Informal Social Control
- ▶ Integrating Rational Choice and Other Theories
- ▶ Labeling: History and Concept
- ▶ Monitoring and Evaluation of Restorative Justice
- ▶ New Penology
- ▶ Rational Choice and Prospect Theory
- ▶ Rational Choice Theory
- ▶ Rational Choice Theory
- ▶ Restorative Justice and State Crime
- ▶ Shame in Criminological Theory

Recommended Reading and References

Anderson E (1999) Code of the street: decency, violence, and the moral life of the inner city. University of Chicago, Chicago

Anwar S, Loughran TA (2011) Testing a Bayesian learning theory of deterrence among serious juvenile offenders. Criminology 49:667–698

Apel R (2012) Sanctions, perceptions, and crime: implications for criminal deterrence. J Quant Crim 29(1):67–101

Becker G (1968) Crime and punishment: an economic approach. J Polit Econ 76:169–217

Berk RA, Campbell A, Klap R, Western B (1992) The deterrent effect of arrest in incidents of domestic violence: a Bayesian analysis of four field experiments. Am Sociol Rev 57:698–708

Braithwaite J (1989) Crime, shame and reintegration. Cambridge, New York

Braithwaite J (1999) Restorative justice: assessing optimistic and pessimistic accounts. In: Tonry M (ed) Crime and justice: a review of research, vol 25. University of Chicago, Chicago, pp 1–12

Bueno De Mesquita B, Cohen LE (1995) Self-interest, equity, and crime control: a game-theoretic analysis of criminal decision making. Criminology 33:483–518

Clarke R, Cornish D (1985) Modeling offenders' decisions: a framework for research and policy. Crim Justice 6:147–185

Clear TR (2007) Imprisoning communities: how mass incarceration makes disadvantaged neighborhoods worse. Oxford University Press, New York

Cook P (1980) Research in criminal deterrence: laying the groundwork for the second decade. In: Morris N, Tonry M (eds) Crime and justice: an annual review of research, vol 2. University of Chicago, Chicago, pp 211–268

Durlauf SF, Nagin DS (2011) The deterrent effect of imprisonment. In: Cook PJ, Ludwig J, McCrary J (eds) Controlling crime: strategies and tradeoffs. University of Chicago, Chicago

Durlauf SF, Fu C, Navarro S (2012) Capital punishment and deterrence: understanding disparate results. J Quant Criminol 21:1–19
Goffman E (1963) Stigma: notes on the management of spoiled identity. Prentice Hall, Englewood Cliffs
Loughran TA, Paternoster R, Piquero AR, Pogarsky G (2011) On ambiguity in perceptions of risk: implications for criminal decision-making and deterrence. Criminology 49:1029–1061
Manza J, Uggen C (2006) Locked out: felon disenfranchisement and American democracy. Oxford University Press, New York
Matsueda RL, Kreager DA, Huizinga D (2006) Deterring delinquents: a rational choice model of theft and violence. Am Sociol Rev 71:95–122
McCarthy B (2002) The new economics of sociological criminology. Annu Rev Soc 28:417–442
Nagin DS (1998) Criminal deterrence research at the outset of the twenty-first century. In: Tonry M (ed) Crime and justice: a review of research, vol 23. University of Chicago, Chicago
Pager D (2007) Marked: race, crime, and finding work in an era of mass incarceration. University of Chicago Press, Chicago
Paternoster R (1987) The deterrence effect of the perceived certainty and severity of punishment: a review of the evidence and issues. Justice Q 4:173–217
Piliavin I, Gartner R, Thornton C, Matsueda RL (1986) Crime, deterrence, and rational choice. Am Sociol Rev 51:101–119
Rebellon C, Piquero NL, Tibbetts SG, Piquero AR (2010) Anticipated shame and criminal offending. J Crim Justice 38:988–997
Sampson RJ, Laub JH (1993) Crime in the making. Harvard University Press, Cambridge, MA
Sampson RJ, Raudenbush SW, Earls F (1997) Neighborhoods and violent crime: a multilevel study of collective efficacy. Science 277:918–924
Sherman LW, Berk RA (1984) The specific deterrent effects of arrest for domestic assault. Am Sociol Rev 49:261–272
Steinberg L (2009) Adolescent development and juvenile justice. Annu Rev Clin Psychol 5:459–485
Wakefield S, Uggen C (2010) Incarceration and stratification. Annu Rev Soc 36:387–406
Western B (2006) Punishment and inequality in America. Russell Sage, New York
Williams KR, Hawkins R (1986) Perceptual research on general deterrence: a critical overview. Law Soc Rev 20:545–572

RCTS

▶ Differences-in-Differences in Approaches
▶ Randomized Experiments in Criminology and Criminal Justice

Reading Verbal and Nonverbal Behavior

▶ Evaluating Truthfulness: Interviewing and Credibility Assessment

Realism

▶ So What Criminology?

Reality Based Television Shows

▶ Cultural Criminology

Rearrest

▶ Risk Factors for Prison Recidivism

Reassurance Policing and Signal Crimes

Andrew Millie
Department of Law and Criminology, Edge Hill University, Ormskirk, Lancashire, UK

Synonyms

Neighborhood policing (UK); Signal crimes perspective

Overview

A reassurance function for policing was first considered by American psychologist Charles Bahn (1974: 338) as "feelings of safety that a citizen experiences when he knows that a police officer or patrol car is nearby." This idea was taken forward in Britain by Martin Innes and colleagues in the early 2000s through the development of a signal crimes perspective. At this time, British policing implemented a National Reassurance Policing Programme (NRPP) where local policing priorities were decided through consultation with local communities. The impact of reassurance policing has since spread and the approach has also been considered in Australia, Belgium, the Netherlands, and Sweden.

In this entry, the background to reassurance policing is considered with particular reference to the work of Charles Bahn and Martin Innes and colleagues. The development of a policy of reassurance policing in Britain is also examined. The successes and limitations of the approach are considered and three main issues identified: that reassurance needs to be a consideration for *all* policing; that increases in visible patrol need to be questioned (especially at a time of budget restraint); and that reassurance policing has the potential to be a model of democratic policing, but only if consultation is truly inclusive, for instance, including those that have been victimized and groups that have been targets of police activity such as young people, the homeless, and other minority and marginalized groups.

Fundamentals

"Reassurance policing" is an approach to policing that emphasizes the importance of the police communicating a positive image to the public, that the public is reassured that the police are doing a good job. It is closely allied to "community policing," "community-oriented policing services" (COPS), and "neighborhood policing." The first to write about the reassurance function of policing was American psychologist Charles Bahn (1974). The context was the publication of the Kansas City Preventative Patrol Experiment (Kelling et al. 1974). According to the then Kansas City Chief of Police, Joseph D. McNamara (Kelling et al. 1974: vii), the experiment showed that "routine preventive patrol in marked police cars has little value in preventing crime or making citizens feel safe." However, Bahn suggested police patrol still had an important function by increasing public awareness of police work and thereby providing visible reassurance (and hopefully making the public feel safer). He defined the reassurance function in the following terms:

> In addition to its more obvious functions, police patrol also has the function of citizen reassurance - providing feelings of safety that a citizen experiences when he knows that a police officer or patrol car is nearby. (Bahn 1974: 338)

Bahn's definition emphasized subjective feelings of safety rather than any objective measure, and the best way to provide reassurance, to make the public feel safer, was through visible and accessible police patrol:

> . . . when the man in the street asks for more police, he is really asking for the police to be on hand more frequently and more conspicuously when he is going about his daily business. (Bahn 1974: 340–341)

After Bahn, there followed a long period during which time the reassurance function was not a major focus for policing. Crime continued to rise and so controlling the level of crime was the clear priority. Rather than focus on patrol, emphases were increasingly on professionalism and crime management, a decision that appeared to be backed by the research. Not only did the Kansas City Experiment question the value of patrol work (Kelling et al. 1974), but also the Newark Foot Patrol Experiment (Kelling et al. 1981). In this study, George Kelling's team found that foot patrol did not reduce crime. However, more positively – and in support of Bahn's position – they also found that foot patrol did have an impact on fear of crime and police satisfaction (see also Clarke and Hough 1984).

Yet, a performance culture evolved in North America, Britain, and many other western nations more focused on crime control and prevention – for example, through forms of zero-tolerance or intelligence-led policing (Innes 2004). In Britain, in the late 1990s and early 2000s, the focus had become volume crime while neglecting more minor disorders and antisocial behaviors (Millie and Herrington 2005). According to FitzGerald et al. (2002: 118), "quantitative performance management was narrowing the focus of police activity in ways that might put long-term investment in police-community relations at risk."

At a time of rising crime, there were also high levels of fear of crime. However, from the mid-1990s onward, recorded crime rates fell in most western nations. This fall in crime was not thought to be matched by falls in fear of crime (Innes 2004). The meaning and measurement of fear of crime are contested and the problem could be exaggerated (Farrall and Gadd 2004). There is not a straightforward relationship between crime rates, police activity, and fear of crime; in fact, in some jurisdictions, an aggressive approach to crime by the police may have resulted in greater fear of crime among marginalized populations (Innes 2004). Also, according to Innes, an emphasis on crime management meant that much police work became behind-the-scenes analysis that was not visible to the public. An unintended consequence was an increased "social distance between police and policed" (Innes 2004: 156).

However, despite claims that the level of fear of crime was a problem, British Crime Survey measures of "worry about crime" did fall from the mid-1990s (Millie and Herrington 2004, 2005). Of greater importance than fear of crime – for a policy of reassurance policing – is public confidence in the police. In Britain, public confidence declined at the same time that recorded crime rates fell (Hough 2003). Furthermore, members of the pubic were unaware that crime rates had been falling; for instance, according to the Audit Commission (1999: 2): "Recorded crime fell by 14 per cent from 1995 and 1997, but fewer than one person in ten was aware that crime had decreased over this period." The disjuncture between falling crime, the public's ignorance of this fall (with many thinking it was rising), and falling confidence in the police became known as the "reassurance gap." Feeding into this were concerns about fear of crime. British police introduced a policy of "reassurance policing" at the start of the new millennium as a response to concern over the "reassurance gap." The answer to falling public confidence in policing was not it seemed greater managerial focus on performance, but a shift in focus to people's perceptions:

> The current over-focus on crime figures as a substitute measure for police performance will not lead to greater public confidence. The public is convinced that the UK is a high-crime society. A reassurance strategy will provide an opportunity to shift people's perception and begin to change the embedded culture of fear of crime. (ACPO 2001: para.5.3)

A focus on perceptions and feelings has been criticized. For some, the approach acquired the nicknames of "there, there policing" or "big hug policing" (Millie and Herrington 2004: 4). For others, it was little more than a public relations exercise and a re-branding of earlier attempts at community policing. According to FitzGerald et al. (2002), giving the impression that the public are safer without also tackling crime rates is hard to justify:

> If reassurance policing yields reductions in crime and disorder as well as reassurance, it is hard to see how it differs from effective policing. If it *doesn't* yield these reductions, the case for it needs scrutiny. It is difficult to justify devoting limited police resources to policing activity that serves only to give people the *impression* they are safer from crime (FitzGerald et al. 2002: 132, emphasis in original).

The UK Home Office embraced the idea of reassurance policing and commissioned a study, "to identify ways to free-up officers' time" in order to perform reassurance policing (PA Consulting 2001: v). According to the PA Consulting report, reassurance policing was defined as:

> . . . any sectoral policing activity that is visible within the community. It is the visibility of the policing effort that provides the reassurance. Hence, reassurance policing includes patrols – either directed or uncommitted, mobile or on foot – as well as visible policing of incidents. Reassurance policing is designed to address not only crime, but the fear of crime (PA Consulting 2001: 1).

PA Consulting extended the definition from Bahn's visible patrols to *any* form of policing that is visible to the public. The emphasis was on visibility as a means to address fear of crime. At the same time, a report by Her Majesty's Inspectorate of Constabulary (HMIC) investigated "the role of police visibility and accessibility in public reassurance" (Povey 2001). For the HMIC report (Povey 2001: 20) reassurance was defined as "the extent to which individuals perceive that order and security exist within their local environment." While PA Consulting focused on visibility, the HMIC report broadened this to "visibility, accessibility, and familiarity" (Povey 2001):

- "Visibility: the level, profile and impact of police resources deployed within local communities.
- Accessibility: the ease with which the public can obtain appropriate police information, access services or make contact with staff.
- Familiarity: the extent to which police personnel know, and are known by, local communities." (Povey 2001: 23–24).

Taking the different definitions together reassurance is a subjective state influenced by perceptions/sense of order and security, and by fear of crime. The job of reassurance policing is to address these subjective feelings through increased police visibility, accessibility, and familiarity. Put simply, according to Innes (2007: 135), the police are assumed to have "an important role to play not just in making people safe but in making them feel safer also."

Key to this is the semiotic quality of police-public encounters, whether this is while on patrol, when an incident is reported, through calls for service or any other mediated communication. Semiotics is an approach that attempts to understand the communicative quality or sign-value of things. Policing communicates different messages to different publics and this may be intimidation and suppression, or more positively, efficiency and reassurance. What the police communicate will be influenced by reputation, levels of trust and respect, experience of police-public encounters, and visual cues such as the uniform, police car, or even station architecture and design (Millie 2010, 2012). Also of relevance here is symbolic interaction. For instance, drawing on Goffman (1959/1990) Innes posits that, "the reassurance function of policing recognises and seeks to harness the dramaturgical power of formal social control" (Innes 2007: 133). In effect, the police officer's role is a *performance* of control and reassurance.

The Signal Crimes Perspective

In Britain, the communicative qualities of policing were the focus of the "signal crimes" perspective developed by a team from Surrey University, led by Martin Innes and Nigel Fielding (e.g., Innes and Fielding 2002). Alongside the influence of Bahn's work on police reassurance, the signal crimes perspective was informed by psychological research into risk perception, especially that of Paul Slovic (1992). According to Slovic, risk can be amplified by a combination of "psychological, social, cultural, and political factors" (1992: 124). Slovic contended that each risk-related event holds a "signal value," reflecting "the perception that the event provides new information about the likelihood of similar or more destructive future mishaps" (1992: 124). For the signal crimes perspective, this idea was transferred to criminogenic risks. According to Innes and Fielding (2002: 3.6):

> . . . public understanding of the seriousness of a risk is not defined solely by the characteristics of the event itself. Rather, it is the nature of the risk, it's semiotic properties, together with the context in which it occurs, that shapes how it is interpreted and understood.

From this perspective, it is not the seriousness of an event (however that is measured), but the signal value of an event that has the biggest effect. This will be influenced by a range of factors including, for example, perception of seriousness and impact on individual or neighborhood, frequency, time and location of event, behavioral expectations, and media coverage. Innes and Fielding went on to state that "the most important incidents in shaping popular fear and anxiety may not necessarily be only those that are traditionally defined by juridical discourses as serious crimes" (2002: 3.6). In fact, they concluded, it could be "relatively trivial" disorders, incivilities, or antisocial behaviors in one situation, but serious crimes in another that have the greatest signal value. According to Innes and Fielding (2002: 5.2):

> A signal crime/event can be defined as an incident that is disproportionately influential in terms of causing a person or persons to perceive themselves to be at risk in some sense.

The signal crime has disproportionate effect on perceptions of security, it is a warning that something is perceived to be wrong with a neighborhood or with society in general. By focusing police attention on whatever is identified as the signal crime(s) for a particular neighborhood, the hope is that there will be bigger gains in improving feelings of security and reassurance. For the police to adopt a signal crimes approach, they need to consult various publics and identify which crimes or disorders act as "signals" in particular neighborhoods. There can be strong and weak signals and not everyone will interpret the signals the same way, or the same in different spatial-temporal contexts; for instance:

> What for young people is merely part of their normal routine and lifestyle (i.e. hanging around on street corners during the evening), may be construed by older residents as signals that they should not go out at night, in order to avoid threats to their safety. (Innes and Fielding 2002: 5.10).

There is a risk that certain "usual suspects" may be unjustly perceived as signals of disorder, for instance, including young people who congregate, the street homeless, or other minority and marginalized populations. If police enforcement policy is informed by a signal crimes perspective, then the disproportionate targeting or "zero tolerance" of those labeled as signals would need to be avoided. A similar criticism has been leveled at the "broken windows" perspective of Wilson and Kelling (1982); in fact, "broken windows" and "signal crimes" have a lot in common. For the broken windows perspective, disorder is "read" as an indicator of decline, causing people to withdraw from disorderly areas leading to a decline in informal social control, thus making crime more likely. In effect, disorder is thought to cause crime to occur. However – drawing on Sampson and Raudenbush (1999) – Innes and Fielding contend that disorder does not lead to crime, but that crime and disorder are functionally equivalent: "Thus disorder is part and parcel of crime itself. Graffiti does not cause robbery, but a lack of informal social control is a cause of both" (Innes and Fielding 2002: 4.3). Alongside certain crimes, disorders, or people having a "signal value," Innes and colleagues posited that the police can act as "control signals," especially through visible patrol. This is supported by the informal social control enacted by active communities.

According to Innes (2004), "reassurance" is a style of policing that can deliver enhanced security and is not in itself an outcome. For Innes (2004), reassurance policing consists of these three parts:

- "high visibility patrols performed by officers who are known to the local public;
- the targeting of 'signal crimes' and 'signal disorders'; and
- informal social control performed by communities" (Innes 2004: 151).

The British National Reassurance Policing Project (NRPP)

From 2002 to 2003, a strategy trial of reassurance policing was run by Surrey Police and the Metropolitan Police in association with the Association of Chief Police Officers (ACPO) and the Home Office. The business case for the strategy trial had five elements:

- *Public confidence* in police choices and solutions through identifying "signal" concerns through public consultation
- *Visible control* – public reassurance by seeing visible proof that their problems are important to the police and are being controlled
- *A targeted, intelligence-led approach* focusing resources on community signals through problem solving
- *Joint action* from the police and other partner agencies
- *Dedicated resources* as much as is practicable (Millie and Herrington 2005)

A range of methods were utilized in order to identify which crimes and disorders acted as "signals" (Innes et al. 2009). In the first instance, drawing on the work of Sampson and Raudenbush (1999) and others, a visual/environmental audit was conducted of a neighborhood to identify potential "signal crimes" and "control signals." Secondly, drawing in particular on the Chicago Alternative Policing Strategy (Skogan and Hartnett 1997), local people were consulted on what were perceived to be signals. This was through the use of public perception surveys and community consultation groups. The Surrey University team also interviewed individual residents using maps to guide responses (Innes et al. 2009).

The full National Reassurance Policing Project (NRPP) was launched in 2003 and operated until 2005 in 16 sites located across eight force areas. The eight forces were Greater Manchester Police, Lancashire Police, Leicestershire Police, Metropolitan Police, Merseyside Police, Surrey Police, Thames Valley Police, and West Midlands Police. The program was evaluated by the Home Office (Tuffin et al. 2006). According to the Home Office evaluation, the NRPP aimed to achieve:

- "Reduced anti-social behaviour and improved quality of life;
- Reduced fear of crime and improved sense of safety;
- Increased public satisfaction with, and confidence in, the police; and,
- Improved social capacity" (Tuffin et al. 2006: xi).

This was an ambitious list. In fact, on the ground, reassurance policing was interpreted as having other additional objectives (Millie and Herrington 2005). For instance, the approach was regarded as a means of gathering community intelligence. Furthermore, it was thought to provide much needed structure to community policing, to lead to improvements in the local environment, and to reduce overall crime. According to Millie and Herrington (2005: 53), the central focus of reassurance policing needed to be to improve confidence in the police and to provide legitimacy for policing decisions. According to the Home Office evaluation, the aims of the NRPP were to be achieved through:

- "Targeted policing activity and problem-solving to tackle crimes and disorder which matter in neighbourhoods;
- Community involvement in the process of identifying priorities and taking action to tackle them; and
- The presence of visible, accessible and locally known authority figures in neighbourhoods, in particular police officers and police community support officers" (Tuffin et al. 2006: xii).

The involvement of communities in identifying policing priorities can be regarded as part of broader "localism" agendas. According to the Home Office evaluation, after program implementation, more people thought crime had decreased, there were positive effects on public satisfaction with the police, and positive effects on perceptions of antisocial behavior and of feelings of safety (but not on fear of crime). Furthermore, there were improvements in levels of police visibility and familiarity and in public engagement. That said, public attendance at meetings was no greater in trail sites than it was in control sites (Tuffin et al. 2006). It was suggested that public consultation should extend beyond standard public meetings to include "street briefings, door knocking and 'have a say days'" (Tuffin et al. 2006: xvi), and when extra effort was put into public engagement, this was noticed by the public. The conclusion of the evaluation was as follows:

> The research did not provide a test of the 'signal crimes' perspective, developed by Martin Innes, but does suggest that a policing approach which targets public priorities can have a positive impact both on crime and on public perceptions. . . . The limited improvements in worry and social capacity indicators suggest the need for further survey work to examine future change in the sites. (Tuffin et al. 2006: 95).

By 2004, policing policy was already moving forward with, in the Metropolitan Police, the emergence of a Safer Neighbourhoods Programme (Herrington and Millie 2006). This evolved into a national Neighbourhood Policing Programme (Innes 2005) and by 2008, Neighbourhood Policing Teams were allocated across all of England and Wales (Millie 2010). Key elements of reassurance policing survived in this new format.

In Britain, policing policy is very much tied in with national politics and in 2010, a new Conservative-led coalition government came into power. The financial crisis led to a decision to cut public sector finance – including a 20 % cut in central government funding of policing (Millie and Bullock 2012). It was not certain how reassurance/neighborhood policing would survive and in what form. That said, the coalition emphasized localism with the first democratically elected Police and Crime Commissioners in Britain introduced in November 2012. Reassurance policing's emphasis on community involvement could be attractive to the newly elected Commissioners.

Issues and Controversies

The Home Office evaluation of Reassurance Policing was broadly positive regarding the approach's effectiveness. The approach may not have had much of an impact on fear of crime or on social capacity, yet the improvements in perceptions of crime and antisocial behavior and public attitudes toward the police were not insignificant. Yet there are potential issues and controversies with the approach, which are outlined here.

Just Another Form of Community Policing

Firstly, there was always a risk that reassurance policing would be viewed as "just another project" or the same as community policing but under a different banner (Herrington and Millie 2006). While some of those involved in the NRPP thought it was the same as before, others saw it as something much more. For one officer involved in the strategy trial, the approach needed to be a "golden thread" running through all policing (Millie and Herrington 2005: 54) and not just a consideration for those involved directly in the project. There is a danger that if reassurance is restricted to a particular project, then those not involved could ruin any advances made by "loose talk" or "loose action." For instance, this could be by disproportionate targeting of "usual suspect" minority populations; or it could be the removal of police identification numbers during public protest, as has occurred in Britain (Millie 2010). If reassurance is a "golden thread" through all policing, then such activities would be questioned.

Reassurance and Visibility

A further focus for reassurance policing is improved police visibility. An original intention for reassurance policing, as outlined by Bahn (1974), was for officers to be a fixed point of reference on the street. This part of Bahn's approach was not incorporated into the NRPP; however, there is scope for improved visible (and fixed point) reassurance through the effective use of police stations and neighborhood bases in a similar fashion to the Japanese system of *koban* (Millie 2010).

Popular politics frequently call for more police officers on the beat, irrespective of their seemingly low effectiveness in reducing crime, as famously demonstrated by the Kansas City experiment. Instead, the police are there to be seen so as to improve people's *feelings* of security. According to Loader (2006: 207), there is a "self-propelling circle whereby popular demands, and the numbers of police supplied in a bid to meet them, are both endlessly ratcheted up." There are clear resource implications of highly visible policing. At a time of fiscal restraint, the deployment of large numbers of locally based officers, with a remit to be highly visible, becomes less of an option.

Reassurance Policing as Democratic Policing

A central focus for reassurance policing is that neighborhood priorities are identified by the community. There is a risk that such locally identified priorities will be in conflict with force or national policing objectives. Furthermore, there is an assumption that all neighborhoods want police intervention (some may prefer the police to go away).

Yet, on the surface, the approach is a commendable example of democratic policing. Legitimacy in policing decisions is maximized as decisions are made by "the community" rather than being dictated by police or government hierarchy. However, there is a potential problem that those consulted may not represent all the multiple communities that live in a neighborhood. Furthermore, they may constitute the "worried well," rather than a group with experience of victimization (Millie 2010). In such cases, the "signal crimes" identified will be based on perception more than actual crime and antisocial behavior (Crawford 2007). The involvement of "hard to reach" or less visible or vocal groups is notoriously difficult (e.g., Skogan and Hartnett 1997). And, as already noted, there is a danger that "usual suspects" – such as young people, street homeless, or other minority and marginalized populations – will become labeled as "signals" and disproportionately policed. According to Loader (2006), if those consulted have expectations for "total security" that cannot be met, then even they may not be reassured. Furthermore, there is a danger that those consulted may suffer consultation fatigue and the police will just assume they know what is required, as Herrington and Millie (2006: 159) have observed:

> . . . we were frequently told that the police knew what the concerns were, the public were tired of being consulted and just wanted something done. This may turn out to be the case, but effective and inclusive consultation is essential if policing is to avoid a "ready, fire, then aim" criticism that is often levelled at them when adopting new initiatives.

Conclusions

The idea of "reassurance policing" was first considered in the USA through the work of psychologist Charles Bahn (1974). However, it was not until the early 2000s in Britain that it was taken forward by Martin Innes and colleagues with the development of the signal crime perspective. A theoretically informed method was developed for delivering a reassurance approach to policing and reassurance policing became a major strand of British policing policy. The approach has also had wider international impact. For instance, reassurance policing has also been considered for policing in Australia, Belgium, the Netherlands, and Sweden.

The evaluation of the NRPP in Britain was largely positive, that reassurance policing has the potential to provide an approach that can improve confidence in the police and improve perceptions of crime and antisocial behavior. There are however obstacles that would need to be negotiated in order to maximize effectiveness:

- Firstly, reassurance would need to be integral to all policing activity, from call handling through to public order policing. Reassurance would be most effective as a consideration for all officers, rather than as a bolt-on extra and only a concern for a smaller number of "community" or "reassurance" officers.
- Secondly, the assumption that more and more visible officers are always better would need to be questioned, especially at a time of fiscal restraint. Other forms of improved visibility could be investigated, such as improved use of police stations and neighborhood bases/shop front offices.
- Thirdly the democratic potential of reassurance policing would need to be taken seriously – that consultation does not focus on the "worried well," but instead the views and experiences of those who have been victimized or have been the targets of police action are actively sought. To be truly inclusive and reassuring to all community members, consultation would need to include young people, the homeless, and other minority and marginalized groups.

Related Entries

- ▶ British Police
- ▶ Democratic Policing
- ▶ Early Modern Police and Policing
- ▶ Evidence-Based Policing
- ▶ Law of Community Policing and Public Order Policing
- ▶ Measuring Police Performance
- ▶ Order Maintenance Policing
- ▶ Organizational Change and Police Legitimacy
- ▶ Police Culture
- ▶ Police Discretion in Law Enforcement
- ▶ Police Discretion in Providing Services and Assistance
- ▶ Police Integrity
- ▶ Police Legitimacy and Police Encounters
- ▶ Police Self-Legitimacy
- ▶ Role and Function of the Police
- ▶ Theories on Policing and Communities

Recommended Reading and References

ACPO (2001) Reassurance - civility first: a proposal for police reform. Association of Chief Police Officers, London

Audit Commission (1999) Safety in numbers: promoting community safety, briefing, February 1999. Audit Commission, London

Bahn C (1974) The reassurance factor in police patrol. Criminology 12(3):338–345

Clarke RV, Hough M (1984) Crime and police effectiveness, Home Office Research Study 79. Home Office, London

Crawford A (2007) Reassurance policing: feeling is believing. In: Henry A, Smith DJ (eds) Transformations of policing. Ashgate, Aldershot

Farrall S, Gadd D (2004) Research note: the frequency of the fear of crime. Br J Criminol 44(1):127–132

FitzGerald M, Hough M, Joseph I, Qureshi T (2002) Policing for London. Willan, Cullompton

Goffman E (1959/1990) The presentation of self in everyday life. Penguin, London

Herrington V, Millie A (2006) Applying reassurance policing: is it 'business as usual'? Polic Soc 16(2):146–163

Hough M (2003) Modernisation and public opinion: some criminal justice paradoxes. Contemp Pol 9(2):143–55

Innes M (2004) Reinventing tradition? Reassurance, neighbourhood security and policing. Crim Just 4(2):151–171

Innes M (2005) Why 'soft' policing is hard: on the curious development of reassurance policing, how it became neighbourhood policing and what this signifies about the politics of police reform. J Community Appl Soc Psychol 15(3):156–169

Innes M (2007) The reassurance function. Policing: J Policy Pract 1(2):132–141

Innes M, Fielding N (2002) From community to communicative policing: signal crimes and the problem of public reassurance. Sociol Res Online 7(2). Available at: www.socresonline.org.uk

Innes M, Abbott L, Lowe T, Roberts C (2009) Seeing like a citizen: field experiments in 'community intelligence-led policing'. Police Pract Res 10(2):99–114

Kelling GL, Pate T, Dieckman D, Brown CE (1974) The Kansas city preventive patrol experiment: a summary report. The Police Foundation, Washington DC

Kelling GL, Pate A, Ferrara A, Utne M, Brown CE (1981) Newark foot patrol experiment. The Police Foundation, Washington, DC

Loader I (2006) Policing, recognition, and belonging. Ann Am Acad Pol Soc Sci 605(1):201–221

Millie A (2010) Whatever happened to reassurance policing? Policing: J Policy Pract 4(3):225–232

Millie A (2012) Police stations, architecture and public reassurance. Br J Criminol 52(6):1092–1112

Millie A, Bullock K (2012) Re-imagining policing post-austerity. Br Acad Rev 19:16–18

Millie A, Herrington V (2004) Reassurance policing in practice: views from the shop floor. Papers from the British criminology conference, vol 7. Available at: www.britsoccrim.org

Millie A, Herrington V (2005) Bridging the gap: understanding reassurance policing. Howard J Crim Just 44(1):41–56

PA Consulting (2001) Diary of a police officer, Police Research Series Paper 149. Home Office, London

Povey K (2001) Open all hours: a thematic inspection report on the role of police visibility and accessibility in public reassurance. Her Majesty's Inspectorate of Policing, London

Sampson RJ, Raudenbush SW (1999) Systematic social observation of public spaces: a new look at disorder in urban neighborhoods. Am J Sociol 105(3):603–651

Skogan W, Hartnett S (1997) Community Policing, Chicago Style. Oxford University Press, New York

Slovic P (1992) Perceptions of risk: reflections on the psychometric paradigm. In: Krimsky S, Goulding D (eds) Social theories of risk. Praeger, Westport

Tuffin R, Morris J, Poole A (2006) An evaluation of the impact of the national reassurance policing programme, Home Office Research Study 296. Home Office, London

Wilson J, Kelling G (1982) Broken windows: the police and neighborhood safety. Atlantic Monthly 249:29–38

Recall

► Front Door and Backdoor Sentencing

Recent Perspectives on the Regression Discontinuity Design

Richard A. Berk
Department of Criminology, University of Pennsylvania, Philadelphia, PA, USA

Overview

The regression discontinuity design was originally proposed in 1960 as a powerful alternative to randomized experiments. It has been little used since. Over the past decade, however, the design has been increasingly and successfully employed by economists in a variety of studies. In this entry, the fundamentals of the regression discontinuity are discussed. Recent advances are emphasized.

Introduction

The regression discontinuity design has a long and complex history. It was originally proposed by psychologists Thistlewaite and Campell in 1960 as an alternative to randomized experiments (Thistlewaite and Campbell 1960). The design became widely known with the publication of the justly famous *Experimental and Quasi-Experimental Designs for Research* by Campbell and Stanley (1963). In 1972, the design was independently rediscovered by econometrician Arthur Goldberger (1972) using a linear regression framework. A bit later, Rubin (1977) provided a formal justification for the regression discontinuity design within a more general conceptualization. That paper is still a very good read. Still later, Trochim (1984, 2001) suggested a range of extensions.

Even in these early formulations, the design was simple and powerful. But there were few applications and apparently only four published studies with significant crime and justice content (Berk and Rauma 1983; Berk and de Leeuw 1999; Chen and Shapiro 2007; Berk et al. 2010a). Over the past 15 years, a number of economists and political scientists have extended the design (Imbens and Lemieux 2008b; Imbens and Kalyanaraman 2009; Green et al. 2009) and applied it in a wide variety of settings (Imbens and Lemieux 2008a; Lee and Lemieux 2009; McCrary 2008; Butler 2009; Caughey and Sekhon 2011). An account of how and why interest in the regression discontinuity design has varied over the years can be found in recent paper by Thomas Cook (2008).

In this entry, the fundamentals of the regression discontinuity design are considered. Some recent advances are highlighted. The discussion begins with a brief introduction to the ways in which statisticians think about causal inference. Then, the classic regression discontinuity design is examined. Newer material follows.

The Fundamental Features of a Regression Discontinuity Design

In early expositions of the regression discontinuity design, there were clear links to conventional regression analysis. The name for the design is no accident. But, reliance on a regression analysis framework is formally unnecessary and can add a statistical layer that confuses matters. We turn, therefore, to a broader nonparametric perspective that at a conceptual level at least is less complicated. In so doing, we are reversing the usual order of didactic treatments. Usually, the historical chronology is employed: parametric linear regression followed by nonparametric approaches.

Imagine an evaluation of probation supervision that compared regular face-to-face meetings between probation officers and probationers with a kiosk check-in system. For the latter, probationers report to a kiosk with a touch computer screen through which they update background information (e.g., address) and enter important changes in their situation (e.g., got a job). A key outcome for the evaluation would be recidivism. Did those individuals assigned to the kiosk system do worse? At this point in the discussion, the outcome could be categorical (e.g., arrested while on probation or not) or quantitative (e.g., the number of failed urine tests).

In the absence of random assignment, agency administrators might decide how to assign probationers to one of the two supervision methods. Suppose they assigned by the risk to public safety. Only low risk probationers would be allowed to satisfy their reporting requirement using the kiosk. Low risk, in turn, would be determined by a threshold on a quantitative measure of risk. Individuals having risk scores below a certain value would be assigned to the kiosk supervision. Individuals with scores at or above that value would be assigned to the usual form of supervision. (See Berk et al. (2010a) for an illustration.)

It would seem that this design builds bias into any estimates a treatment effect. Those assigned to the kiosk already are believed to have a lower risk of recidivism that could falsely make that intervention look more effective than it really was. And this is surely true.

However, suppose the difference in recidivism between the two interventions were measured only for individuals very near and on either side of the risk threshold. For example, assume that the risk score has a probability metric from 0 to 1 with a higher score representing more risk. There is a threshold at .50. Individuals having risk values equal to or less than .50 are placed in the kiosk condition. Those with risk values greater than .50 are placed in the face-to-face condition. But – and this is a key point – the analysis is limited to individuals with risk values between .48 and .52. Given the current state of probation risk assessments, the difference in risk between a probationer with a score of .48 and a score of .52 is likely to be very small and most likely indistinguishable from noise.

In short, one has a near approximation to random assignment for individuals very close to the risk assessment threshold score. Assignment is determined by values that can be much like the product of a random number generator. Insofar as this is really true, differences on either side of the threshold in the average outcome can provide an effectively unbiased estimate of a "local" treatment effect. It is local because the estimate only formally applies to individuals with risk scores between .48 and .52, not to the full range of risk scores. Much of the recent technical work on the regression discontinuity design examines how best to obtain such estimates and what the formal properties of such estimates might be.

One potential weakness with the approach is that there can be relatively few cases close to and on either side of the threshold. With a small sample size, the treatment effect estimate may not be sufficiently precise. Another potential weakness is that an estimated treatment effect only formally applies to probationers close to the threshold. Who is to say whether the estimate applies to probationers far from the threshold? And yet, those probationers may be an important and large subset of the relevant population.

For there and other reasons, it can be important to consider treatment effect estimators that use a much large fraction of the observations and ideally all of them. This is where the links to conventional regression analysis can sometimes be exploited. We turn to those now and begin with the way in which the regression discontinuity design was initially introduced by Donald Campbell and Julian Stanley in the early 1960s.

Basic Regression Discontinuity Design: The Linear Regression Overlay

As already noted, the regression discontinuity (RD) design can have many of the assets of a randomized experiment, but can be used when random assignment is not feasible. Typically, the goal is to estimate the causal effect of an intervention such as gate money for prison inmates, anger management for troubled teens, or changes in police patrolling practices. In the simplest case, there is an experimental and comparison group with assignment fully determined by an explicit and observable rule. For example, whether community policing is introduced in a neighborhood depends on whether the crime rate is above a specific threshold. Neighborhoods falling above that threshold get community policing. Other neighborhoods get business as usual. A relatively simple and compelling analysis can follow. In some circles, the RD design is called a "quasi-experiment."

One appeal of the RD design is "political." If the assignment rule represents need, either empirically or morally, one can sometimes more easily garner the support of stakeholders. Cooperation from study subjects can also more easily follow. For example, it might seem to just make good sense to assign a policing innovation to high-crime neighborhoods. The technical complication is that neighborhoods assigned to the treatment condition will not be comparable on the average to neighborhoods assigned to the alternative. Indeed, systematic selection is explicitly built into the design. How the potential biases can be overcome in practice is a key theme in the material that follows.

To best appreciate the underlying machinery of the RD design requires some familiarity with the way causal inference has come to be formulated by most statisticians. The framework was proposed by Neyman in 1923 and later extended by Rubin (1974) and Holland (1986). These days, it is sometimes called the "Rubin Causal Model." Be clear that the Rubin Causal Model allows one to rigorously *define* a causal effect. It is a conceptual apparatus. Data play no role. It is not a causal model for how nature generated the data on hand. The "Rubin Causal Model" would perhaps be better called the "Rubin Causal Framework."

In its most simple form, there are observational units: people, neighborhood, police departments, prisons, or other entities. There is a binary intervention. Some of the units are exposed to the experimental "arm" of the study, and the other units are exposed to the comparison "arm" of the study. There is interest in estimating a causal effect. For example, one might want to learn how intensive probation supervision, compared to conventional probation supervision, affects recidivism.

Each *unit* is assumed to have two *potential* outcomes, one if exposed to the experimental condition and one if exposed to the comparison condition. These outcomes are *hypothetical* and can vary across units. Thus, a given probationer would have one response if placed under kiosk supervision and another response if placed under the usual supervision. Another probationer would also have two potential responses, which could differ from those of the first probationer. For now, assume that the outcome is quantitative and of a form suitable for conventional linear regression. To take a very different example, if the study units were census tracts and the interventions were two different methods of policing, the outcome might be the crime rate for part I crimes.

Following Imbens and Lemieux (2008b), let $Y_i(1)$ denote the potential outcome if unit i is exposed to experimental condition and $Y_i(0)$ denote the potential outcome if unit i is exposed to the comparison condition. Interest centers on a comparison between $Y_i(1)$ and $Y_i(0)$, often their difference. Thus, the difference between $Y_i(1)$ and $Y_i(0)$ can *define* a causal effect. An example is the number of failed urine tests should a given probationer be placed under the usual supervision minus the number of failed unit tests should that same probationer be placed under kiosk supervision.

In practice, however, one can never observe both $Y_i(1)$ and $Y_i(0)$. A given unit will only experience the experimental condition or the comparison condition. Let $W_i = 1$ if unit i is actually exposed to the experimental condition and $W_i = 0$ if unit i is actually exposed to the comparison condition. The *observed* outcome is then

$$Y_i = (1 - W_i) \cdot Y_i(0) + W_i \cdot Y_i(1). \qquad (1)$$

Both Y_i and W_i can be observed. For the basic RD design, one also has a single observed covariate X_i that fully determines whether unit i is exposed to the experimental condition or the comparison condition. In the simplest case, X_i is a continuous covariate with an imposed threshold. A unit that falls on one side of the threshold or right on it is assigned to the experimental condition. A unit that falls on other side of the threshold is assigned to the comparison condition (or vice versa).

X can be associated with the potential outcomes. As in our earlier illustration, X may be a measure of probation risk. A probationer who scores above some threshold is assigned to supervision as usual. If that probationer scores at or below the threshold, assignment is to kiosk supervision. The risk measure is by design associated with the potential outcome of re-offending.

There can be other observed covariates Z. In some cases, they are used to define a multivariate assignment rule. For example, probationers who are younger than 21 and who have more than two prior convictions for a violent crime may be assigned to intensive supervision. In other cases, the Z plays no role in the assignment, but is related to the potential outcomes. For now, there is no need to consider either complication.

Because by definition a causal effect, $[Y_i(1)-Y_i(0)]$, is unobservable in practice, there is a shift to the group level. (The same applies if the ratio rather than the difference is used.) Interest centers on the observed *average* response of the units exposed to the experiential condition compared to the observed *average* response of the units exposed to the comparison condition. We are seeking the average treatment effect. Given the assignment rule, however, there is the possibility that the units exposed to the experimental condition will differ systematically from the units exposed to the comparison condition. And those systematic differences can be related to the potential outcomes. There is the risk of building in selection bias.

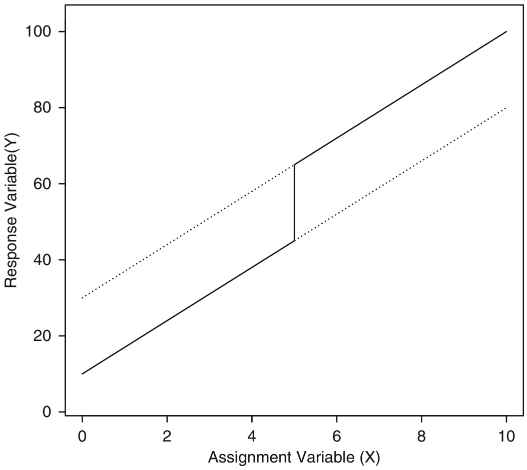

Recent Perspectives on the Regression Discontinuity Design, Fig. 1 Conventional linear case

Under certain assumptions, there is a regression-based solution. Consider Fig. 1. On the horizontal axis is the assignment variable X. On the vertical axis is a quantitative response (outcome) variable Y. Units scoring above 5 on X are exposed to the experimental condition. Units scoring at or below 5 on X are exposed to the control condition.

The higher line (dotted below the threshold of 5 and solid on or above the threshold of 5) represents the potential response for different values of the assignment variable when a unit is exposed to the experimental condition. The lower line (solid below the threshold of 5 and dotted on or above the threshold of 5) represents the potential response for different values of the assignment variable when a unit is exposed to the comparison condition. The solid parts of each response function represent the potential responses actually observed under the assignment rule. The vertical solid line at $X = 5$ represents the causal effect. In this illustration, the average value of the response increases above the threshold. One could have as easily had the average value of the response decrease or show no change at all.

Figure 1 shows the theory required for the basic RD design, and there is a lot going on. For both groups, the relationship between X and Y is positive. The larger the value of X, the larger the value of Y tends to be. If one just compared the observed mean of the group receiving the experimental condition (i.e., with $X > 5$) with the observed mean of the group receiving the comparison condition (i.e., with $X < 5$), as one might do in a randomized experiment, an estimate of the average treatment effect could be substantially biased. In this illustration, the group exposed to the experimental condition has by design larger values for X, which implies larger values of Y regardless of the intervention. Analogous results follow if the relationship between X and Y is negative.

However, note that the relationship between the assignment variable and the response variable is for both groups assumed to be linear. This is true for potential responses that are observed (i.e., the solid lines) and potential responses that are not observed (i.e., the dotted lines). The linearity is very restrictive but very convenient. Note also that the linear relationships are assumed to be parallel. This too is very restrictive but very convenient. One has a strong two-part theory about how the assignment variable is related to the response: the two response functions can differ only by a constant. That constant is the treatment effect.

If the relationships shown in Fig. 1 are a good approximation of reality, a very simple and effective analysis can follow. The difference between the vertical placement of the linear relationship for the units exposed to the experimental condition and the vertical placement of the linear relationship for the units exposed to the comparison condition can provide an estimate the average treatment effect.

More specifically, let

$$Y_i = \beta_0 + \beta_1 W_i + \beta_2 X_i + \varepsilon_i. \qquad (2)$$

Equation 2 is a conventional regression expression representing how the values of the response variable are generated. All of the observables are defined as before, and ε_i is the usual regression disturbance. When $W_i = 0$, the conditional expectation of Eq. 2 gives the linear relationship between X and Y for the group exposed to the comparison condition. When $W_i = 1$, the conditional expectation of

Eq. 2 gives the linear relationship between X and Y for the group exposed to the experimental condition. β_2 is the common slope of the two response functions. β_0 is the intercept of the response function for the group exposed to the comparison condition. β_1 is the vertical distance between the two estimated response functions, which quantifies how much the response function for the experimental group is shifted up (i.e., $\beta_1 > 0$) or down (i.e., $\beta_1 < 0$) because of the experimental condition. In Fig. 1, the length of the vertical line at $X = 5$ is equal to the value of β_1. If the experimental intervention has no effect, $\beta_1 = 0$, and the two lines collapse into one.

It is important to appreciate that Eq. 2 is *not* a causal model. Although the intent is to estimate an average treatment effect, no claims are made that the two regressors can account for all of the systematic variation in the response or that if the assignment variable is manipulated; β_2 shows how the response will be altered. Equation 2 is no more than a method to estimate the average treatment effect. It is rather like the difference between the mean of the experimental group and the mean of the control group in a randomized experiment. There is no causal model there either. But random assignment insures that on the average, all omitted predictors are unrelated to the intervention.

There is an analogous condition for the regression discontinuity design resting on a key assumption: conditional on assignment variable, $E(\varepsilon_i) = 0$. Conditional on the assignment variable, nature conducts the equivalent of a randomized experiment. This can be a credible formulation if assignment to the treatment or comparison condition is solely by the assignment variable threshold and the linear model is correct (Rubin 1977).

In practice, one can use the data on hand and ordinary least squares to estimate the parameters of Eq. 2. Consistent with the design's assumptions, if the two linear relationships are at least nearly linear and parallel, $\hat{\beta}_1$ is an effectively unbiased estimate of the average treatment effect. Likewise, a key to proper estimation is that the true selection mechanism is known and can be accurately represented by a threshold on X. The usual hypothesis test is that $\beta_1 = 0$; there is no treatment effect.

An obvious question is how one would determine in practice whether the two relationships are linear and parallel. A good way to start is to construct a scatterplot of Y against X. On each side of the threshold, one should see a truncated version of the idealized elliptical scatterplot. If there is a treatment effect, the plotted points should be shifted up or down in the region to the right of the threshold (where the units are exposed to the experimental condition). An example is discussed later.

A useful second step is to examine a scatterplot constructed from the residuals and fitted values of Eq. 2. When the residuals are plotted against the fitted values, one should see with no dramatic patterns. In particular, there should be no evidence that the assumed linear and parallel regression response functions are something else. It can also be useful to consider whether the usual assumption of a constant residual variance is credible. In short, one should examine the results from Eq. 2 as one would examine the results from any regression analysis. Graphical methods can be especially instructive and are less constrained by the untestable assumptions that test-based diagnostic methods require (Freedman 2008). An excellent discussion of a number of useful diagnostic methods can be found in the text by Cook and Weisberg (1999).

What can go wrong if proper diagnostic methods are not employed? Figure 2 provides a simple illustration. The two response functions for the potential outcomes are not parallel. The response function for the experimental group is much steeper. (The solid part of the line is observed and the dashed part of the line is unobserved.) In this particularly perverse example, the gap at the threshold between the observed outcomes for the group exposed to the experimental condition and the observed outcomes for the group exposed to the comparison condition just happens to be the same as the treatment effect when the assumption of parallel response functions is met. But this gap is not caused by a discontinuity. There is no evidence of a treatment effect.

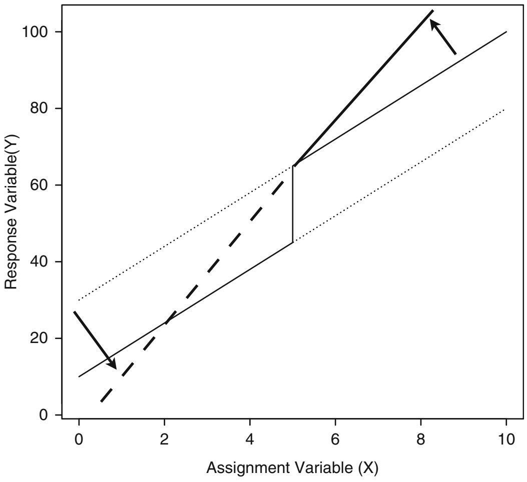

Recent Perspectives on the Regression Discontinuity Design, Fig. 2 The conventional linear case gone wrong

It is certainly possible to alter Eq. 2 allowing for linear response functions that are not parallel. Looking again at Fig. 2, however, one would have to know about the nonparallel response functions in advance. One cannot determine from the data alone whether the response functions are not parallel or whether the intervention altered the slope and intercept of the treatment group.

This can lead to an identification problem. Suppose one was prepared to bet that the two response functions were linear but not parallel. Then, β_1 combines the intercept of the response function for the experimental group in the absence of the experimental intervention with a new intercept for the experimental group should the intervention shift that group's response function. Because there are no observations for the experimental group in the absence of the experimental intervention, the confounding cannot be disentangled, and the basic regression discontinuity design fails.

Suppose that one was prepared to assume that the two response functions are linear and parallel before the intervention but that after the intervention, the slope and intercept are altered. By including the product variable ($W_i \times X_i$) as new regressor in the Eq. 2, the intervention-altered linear relationship for those exposed to the experimental condition can be identified, and the linear relationship for those exposed to the comparison alternative can be identified. One has specified an interaction effect between W_i and X_i for certain values of X_i.

But there is an important complication. The size of the gap between the response function of the experimental group and the response function of the comparison group depends on the value of X. Indeed, the treatment effect may be positive for some values of X and negative for other values of X. Although this is statistically acceptable, constructing a coherent substantive explanation can be challenging. One simplification is to focus only on the estimated average treatment effect in the immediate neighborhood of the threshold although, as already mentioned, important external validity questions can be raised. The estimate only applies to cases in the neighborhood of the threshold. Nevertheless, this is an approach to which we return shortly.

There is nothing in the RD design precluding other kinds of interaction effects that are more easily interpreted. One can subset the data by values of covariates Z_i and consider whether there are different average treatment effects. Or one can respecify Eq. 2 so that interaction effects are represented. For example, an intervention may be more effective for women than for men. One way to explore this would be to do one RD analysis for men and another RD analysis for women. Alterations in Eq. 2 would depend on the precise form the interaction effect takes. For instance, one might include a binary variable for gender and the product of the intervention variable and gender. That would allow for different intercepts and different average treatment effects for men and women.

Some of the lessons from the linear case carry over to the nonlinear case. If the two response functions are nonlinear but parallel, one can in principle estimate a shift resulting from an intervention. However, it is not necessarily clear what a change in slope means if the response functions are not linear. The nonlinear case is discussed a more later.

The Generalized Regression Discontinuity Design

One problem with the basic RD design is that the response is assumed to be quantitative. In many applications, the response is categorical or a count. An example of the former is whether a probationer reoffends. An example of the latter is the number of crimes the probationer commits.

The basic RD design is easily extended to include formulations from the generalized linear model. Logistic regression and Poisson regression are two common illustrations. Consider, for instance, a binary response.

We can proceed initially as before. Equation 3 is the same as Eq. 2 except that the quantitative response Y_i* is now completely unobserved regardless of whether a unit falls above or below the threshold. For example, Y_i* may be the proclivity of a prison inmate to engage in some form of serious misconduct, and no measures of this proclivity are available. As before, we let

$$Y_i^* = \beta_0 + \beta_1 W_i + \beta_2 X_i + \varepsilon_i. \tag{3}$$

We now hypothesize a second kind of threshold. If an inmate's proclivity toward misconduct exceeds a certain value, an act of misconduct is observed.

If that threshold is not exceeded, no act of misconduct is observed. Suppose this outcome is coded so that "1" denotes an observed act of misconduct and "0" denotes no observed act of misconduct.

Drawing on a common motivation for logistic regression (Cameron and Trivedi 2005, Section 14.4), the probability P_i of observing a "1" depends on the probability that Y_i* will exceed the misconduct proclivity threshold. If one then assumes that ε_i has a logistic distribution, Eq. 4 can follow directly:

$$\log \frac{P_i}{1 - P_i} = \gamma_0 + \gamma_1 W_i + \gamma_2 X_i, \tag{4}$$

where γ_0 through γ_2 are new regression coefficients. In other words, the systematic part of the formulation is the same as in Eq. 2, but the response is now in units of log-odds (also called "logits"). Figure 3, therefore, is analogous to Fig. 1.

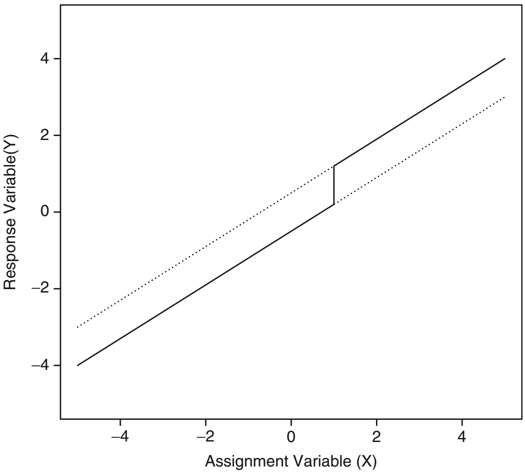

Recent Perspectives on the Regression Discontinuity Design, Fig. 3 The log-odds representation of the logistic case

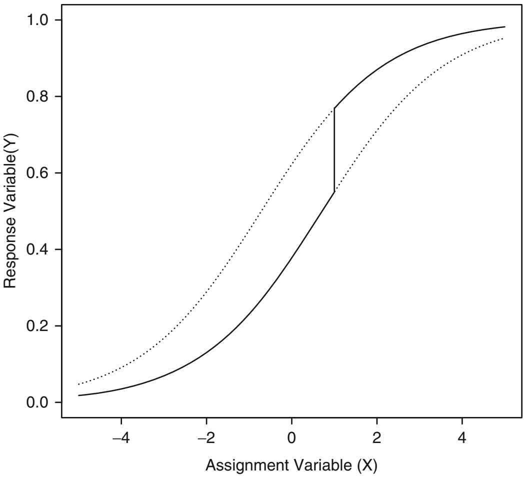

Recent Perspectives on the Regression Discontinuity Design, Fig. 4 The generalized linear case

If Eq. 4 is solved for P_i, the result is shown in Fig. 4. Note that in Fig. 3, when the response is in logit units, the two response functions are linear and parallel. But when in Fig. 4 the response is in probability units, the two response functions are neither linear nor parallel. This is not a problem as long as, consistent with Eq. 4, the estimated treatment effect is reported in logit units or as an odds multiplier. If stakeholders want the story told in probability units, however, the complications introduced by nonparallel response functions reappear. The good news is that there is no identification problem. The bad news is that the average treatment effect depends on the value of the assignment variable. In addition, any regression diagnostics should be applied to the log-odds form of the model as well. In the logit metric, there are many graphical diagnostic procedures for logistic regression. Again, a good reference is Cook and Weisberg (1999).

Berk and de Leeuw (1999) provide an instructive application. In their study, the outcome was misconduct in prison. Misconduct could include minor but common infractions such as failing to report for a work assignment or rare but serious infractions that would be felonies if committed on the outside. The assignment covariate was a risk score computed for each inmate at intake that was used to forecast prison misconduct. The intervention was prison housing at the highest level security with the alternative any other housing placement. Placement was determined by a threshold on the risk score; scoring about the threshold led to a high security placement. The empirical question was whether a high security placement reduced prison misconduct. It did, but not dramatically.

Analogous issue arises if the response variable is a count. The canonical mean function is not the log of the odds but the natural logarithm. The generalized RD design still applies and can easily be employed.

Matching Strategies

From the issues just discussed, it should be clear that the credibility of the basic and generalized RD design depends on the assumptions one makes about the two response functions linking X to Y. Both designs respond to the same fundamental problem: the observed responses of the group exposed to the treatment and the observed responses of the group exposed to the alternative reside in disjoint regions of the assignment variable. There are no values of X at which one can observe responses for both groups. Therefore, all comparisons necessarily depend on extrapolations, which, in turn, depend on the functional forms assumed.

An alternative strategy introduced earlier is to focus on the observed responses on either side of the threshold and very close to it. If this region, often called a “window,” is sufficiently narrow, the values of X within it are likely to be very similar. Because Y is a function of X, this suggests a matching strategy comparing the average observed response of the units just to the right of the threshold to the average observed response of the units just to the left of the threshold. The difference between the two averages can serve as an estimate of the average treatment effect. The same rationale applies if the response variable is categorical and one is comparing proportions. We now need to consider this approach in more detail.

An important assumption is that the small piece of the response function just to the right of the threshold and the small piece of the response function just to the left of the threshold are linear and parallel to each other. To the degree that this assumption is violated, the size of the possible bias increases. For example, if both are linear and parallel but with a positive or negative slope, the absolute value of the difference between the two averages can be larger than the absolute value of the gap size at the threshold. The hope is that if the window can be made sufficiently narrow, this bias will be negligible. Ideally, therefore, the matching strategy can provide useful estimates of average treatment effects with far less reliance on assumed functional forms over the entire length of the response functions.

Figure 5 illustrates the general idea. The two response functions are neither linear nor parallel. And these would likely be unknown in any case. If within the shaded region the two functions were linear and flat, the difference between the two averages would be the same as the treatment effect at the threshold. In this illustration, the estimated treatment effect would be a bit too large. The gap is larger than had the two functions been linear and flat.

It follows that a critical feature of the window approach is the width of the shaded area (Imbens and Kalyanaraman 2009). How wide should the shaded area be? If the shaded area is more narrow, the true causal effect often can be better

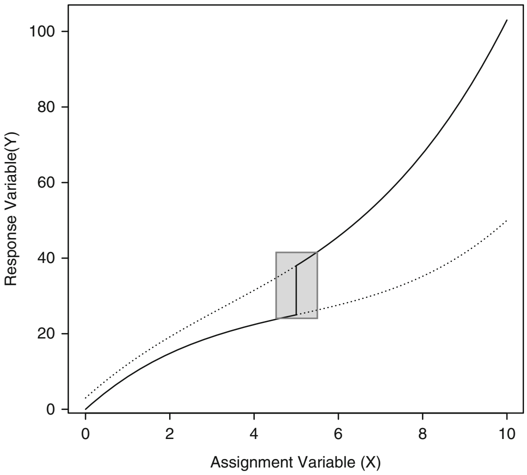

Recent Perspectives on the Regression Discontinuity Design, Fig. 5 Matching within the RD design

approximated. But there will be fewer observations with which to estimate the average treatment effect. With less information, there will be more uncertainty in the estimate. The tension is an example of the well-known bias-variance tradeoff (Hastie and Tibshirani 1990, pp. 40–42). To consider the implications of this tradeoff, we to turn to how one can estimate the average treatment effect. In this context, the terms "bandwidth" or "span" are sometimes used instead of the term "window."

Perhaps the simplest estimator of the average treatment effect within the window is the difference between the mean of the responses to the right of the threshold and mean of the responses to the left of the threshold (Imbens and Lemieux 2009, pp. 623–624). Sometimes, a better approach is to use an explicit kernel function (Hastie and Tibshirani 1990, pp. 18–19), the impact of which can be to weight observations so that those closer to the threshold are given more weight when the means are computed. Values of Y closer to the threshold will often provide more accurate information about the value of Y at the threshold.

Different weighting schemes can be used. For example, when a linear kernel is used, weights decline linearly with distance from the threshold. When a Gaussian kernel is used, the weights decline as a function of the normal distribution centered on the threshold. In practice, all of the common weighting functions that decline with distance from the threshold usually give very similar results. In fact, there is often not much difference between the estimated treatment effect using any of the common weighting functions and the estimated treatment effect using no weighting at all (i.e., all cases get the same weight).

In contrast, the size of the window can have a dramatic effect. Therefore, a criterion is needed that strikes a useful balance between the bias and the variance. A popular theoretical criterion when the response variable is quantitative is the integrated squared prediction error: the mean of the squared disparities between the fitted values $\hat{Y}_i$ and the actual values of Y_i. If a window size could be chosen to minimize the integrated squared prediction error, the sum of the variance and the bias squared will be minimized as well (Hastie et al. 2009, Section 5.5). In this sense, minimizing the integrated squared prediction error provides a good way to define an appropriate balance between the variance and the bias.

Several procedures can be used as an operational stand-in for the theoretical integrated squared prediction error. Cross validation is a useful option and with, modern computers, the calculations can be done very quickly. Other options include the generalized cross-validation statistic, AIC, BIC, and Mallows Cp. Imbens and Kalyanaraman (2009) provide a very interesting advance on this theme. In practice, each of these approaches usually will lead to about the same results, and many work well if the response variable is a count or is categorical. An excellent discussion of the underlying issues can be found in the book by Hastie and his colleagues (2009, Chapter 7).

The comparison between means within the window can be generalized. Suppose that within the window the two unknown response functions are linear, but not necessarily flat or even and parallel. Then, one can often obtain a more accurate estimate of the treatment effect by employing two linear regressions within the window, one on each side of the threshold (Imbens and Lemiux 2009, pp. 624–625). Each is a form of local linear regression. The sole regressor is the distance from the threshold. More specifically, we assume that

$$Y_{iL} = \delta_{0,L} + \delta_{1,L} \cdot (X_{iL} - T) + \varepsilon_{iL}, \qquad (5)$$

and

$$Y_{iR} = \delta_{0,R} + \delta_{1,R} \cdot (X_{iR} - T) + \varepsilon_{iR}, \qquad (6)$$

where within the window, L denotes the left side of the threshold, R denotes the right side of the threshold, T is the value of X at the threshold, X_i is the value of X for given observations, and ε_{iL} and ε_{iR} are conventional regression disturbances. Because $\delta_{1,L} \neq \delta_{1,R}$, the linear response functions do not have the same slope.

At the threshold, $(X_i{-}T) = 0$. Therefore, the average treatment effect at the threshold is $\delta_{0,R} - \delta_{0,L}$. With data for Y and X, one can estimate the values of the parameters from both equations and obtain an estimate of the treatment effect at the threshold by computing the difference between the two intercepts. That is, the estimate of the average treatment effect is $\hat{\delta}_{1,L} - \hat{\delta}_{1,R}$. The same logic can apply for the entire generalized linear model although some of the details will necessarily differ.

Although Eqs. 5 and 6 have much the same structure as Eq. 2, their purpose is rather different. The goal is to estimate the difference between the response functions at the threshold. Therefore, despite the use of regression, the matching logic still prevails. If the regression model is approximately correct, there should be gains by the integrated squared prediction error criterion. In addition, the usual statistical inference undertaken with linear regression can apply although sometimes robust standard errors are desirable (Cameron and Trivedi 2005, Section 4.4.5).

In principle, the regression estimate can be improved upon in two ways. First, one can employ a kernel weighting scheme so that when the regression coefficients are estimated, observations closer to the threshold are given greater weight. In practice, the means squared error gains are usually modest at best. Second, one can assume that $\hat{\delta}_{1,L} = \hat{\delta}_{1,R}$ and reestimate the values of $\hat{\delta}_{0,L}$ and $\hat{\delta}_{0,R}$. In other words, one proceeds as if the two linear response functions are parallel. There can be small but noticeable gains in precision if the equivalence is approximately true. Once again, regression diagnostics can be very instructive.

There can be substantial bias in the estimated average treatment effect at the threshold if regression diagnostics indicate that the response functions are not linear. One possible remedy is to replace the local linear regression with local polynomial regression (Fan and Gijbels 1996) or an even more flexible smoother such as found in the generalized additive model (Hastie and Tibshirani 1990). But, the moment one opens the door to nonparametric regression, there may be no longer a need to stay within the window, and a wide variety of tools are in play. Regression splines and regression smoothers, for instance, can be very effective (Berk 2008a, Chapter 2; Bowman et al. 2004). The task at hand can be reformulated as function estimation problem, with the amount of smoothing replacing the size of the window as a key matter for tuning. The result can be

$$Y_i = \beta_0 + \beta_1 W_i + f(X_i - T) + \varepsilon_i, \qquad (7)$$

where $f(X_i{-}T)$ is determined empirically. Equation 7 is easily extended to the generalized additive model so that binary and count response variables can be analyzed.

One interesting feature of Eq. 7 is that if the fitting procedure for $f\,(X_i{-}T)$ is made sufficiently flexible, there may be no need to include as a regressor. Should there be an important change in the response function in the neighborhood of the threshold, the nonparametric regression procedure is likely to find it. The size of the neighborhood will be determined as well. At the very least, this suggests first using a very flexible nonparametric regression procedure as an exploratory technique to help in the specification of Eq. 7. The risk is that statistical inference can be invalidated (Leeb and Potscher 2006; Berk et al. 2009) and that any findings may be the result of overfitting. Neither problem is unique to the analysis of RD designs, however.

Some Extensions Not Addressed

The RD design may be extended further, but space limitation precludes more than a very brief discussion. One easy and direct extension is to have a deterministic assignment rule constructed from more than one covariate. If there are two such covariates, for example, the threshold is a line, not a point. The various estimation procedures can be altered accordingly. Likewise it is relatively easy to have a proper RD design with more than one intervention, much in the spirit of factorial designs in true experiments.

A more complicated extension addresses the problem of compliance. Just as in randomized experiments, study subjects do not always comply with the treatment or alternative condition assigned. Then, one might be interested in trying to estimate the impact of the intervention assigned (i.e., an "intention-to-treat" analysis) or the impact of the intervention received. The former can be estimated with the procedures we have described. The latter requires more complicated and fragile procedures. A possible approach is using the treatment assigned, conditional on the assignment variable, as an instrumental variable. These and other alternatives are discussed by Imbens and Lemioux (2008b).

External Validity

The results from an RD design raise the same external validity issues as those from a randomized experiment and more. If the study subjects are probability sample from a well-defined population, generalizations to that population can be appropriate. If the study subjects are not selected by probability sampling, generalizations beyond the study subjects must rely on theory or replications. However, any estimates using the matching approach can restrict generalizations further. The relevant units are now only those that fall in the window or for some procedures, only at the threshold. For data that are a proper probability sample, generalizations can now only be to elements in the population that would fall within the window or at the threshold, respectively. When the data are not a proper probability sample, generalizations based on theory or replications must also take these restrictions into account.

In short, the basic and generalized RD designs have the same kinds of external validity constraints as randomized experiments. The external validity constraints can be far more binding if a matching estimation approach is used. The tradeoff is that the matching approach's internal validity may be stronger.

Power

To the degree that the assignment variable is correlated with the treatment indicator, there is, other things equal, a reduction in statistical power. However, other things are not always equal. A possible compensating factor is the strength of the association between the assignment variable and the response. With a larger correlation, there can be an increase in power. It is in practice difficult to determine before the data are collected how strong these two correlations will be. Consequently, conventional power analyses risk being driven very substantially by assumptions that may well prove wrong. The problems are even more serious if one attempts a post hoc power analysis after average treatment effects have been estimated. As a formal matter, post hoc power analyses are difficult to justify (Hoenig and Heisey 2001) and probably should not be taken seriously.

The safest strategy is to work with large samples whenever that is possible (e.g., samples of 500 or more). When the data come from existing datasets or administrative records, the costs of large samples may be no more than the costs of small samples.

An Illustration

We turn now to an illustration. The intent is to provide an overview of how some of the procedures described above can be used in practice.

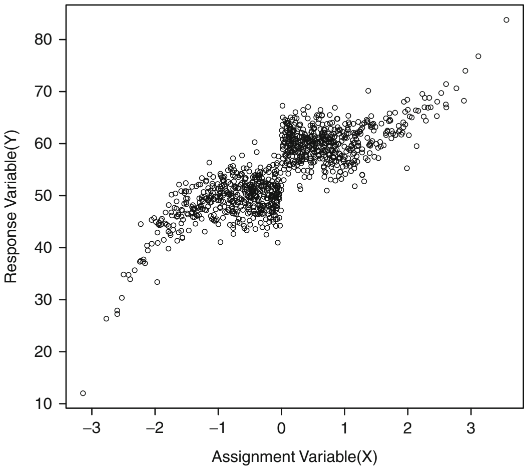

Recent Perspectives on the Regression Discontinuity Design, Fig. 6 A scatterplot for some fictitious data ($N = 1{,}000$)

Recent Perspectives on the Regression Discontinuity Design, Table 1 Estimates of the RD average treatment effect for fictitious data – true treatment effect = 10.0

Model	Estimate	SE	Fit
Correct model	10.4	.37	.83
Linear model	6.3	.37	.74
Difference in means ($N = 400$)	9.8	.31	.72
Regression splines ($k = 3$)	6.16	.37	.75
Regression splines ($k = 5$)	11.8	.43	.82
Regression splines ($k = 7$)	9.97	.51	.83

Space limitation precludes an in-depth discussion using several different datasets with varying properties.

A good way to begin the analysis of data from any RD design is to examine a scatterplot of the response variable against the assignment variable. Figure 6 is just such a plot for some fictitious data. There are 1,000 observations. The true relationship between X and Y is a fourth degree polynomial. This implies that the relationship on either side of the threshold between the response and the assignment variable is the same. Also by construction, the threshold is at 0.0, and the treatment effect equals 10.0. It is readily apparent that there is a discontinuity at the threshold. For real data, evidence of a treatment effect is usually not so easily discovered.

One would normally not know that the relationship between X and Y was a fourth degree polynomial. But it would likely be empirically apparent that the two response functions were approximately linear and flat over much the range of X, at least where there are data. Except for a few observations at the tails, two linear parallel response functions might seem to be a sufficiently good approximation for the RD analysis.

Table 1 shows some results. Its four columns contain in order the kind of estimator, the average treatment effect estimate, the standard error, and the proportion of the deviance that is accounted for by the estimator. Fit quality was evaluated using the AIC, but the adjusted proportion of deviance accounted for by the model is reported for ease of exposition.

In the first row are the results if the correct relationship between X and Y were known. This is the estimation gold standard. The estimated treatment effect is 10.4, the standard error is .37, and about 83 % of the deviance is accounted for. The estimate is about one standard error from the truth of 10.0. One would not reject the null hypothesis that the treatment effect is 10.0. (The treatment effect estimate is not exactly 10.0 because of random sampling error.)

The second row contains the results if the basic RD design is assumed and the linear model applied. The estimate of the average treatment effect is now 6.3, a reduction of about a third. The standard error has not changed (up to two decimal places), but the fit drops from .83 to .74. In this case, if after looking at the scatterplot a researcher decided that the linear model was good enough, the results would not be grievously wrong. The linear fit is not terribly far from the ideal fit: .83 versus .74 (although the ideal fit would not be known) because, where most of the data are, the linear model is about right. Whether the size of the treatment effect underestimate is large enough to matter would depend on context. In many policy settings, underestimating an average treatment effect by a third can transform a cost-effective intervention into one that is not cost-effective.

The third row shows the results for the difference between means in the region immediately on either side of threshold. The region was defined by the X between -1 and 1. Four hundred observations are contained between these two boundaries. The estimated average treatment effect is 9.8, the standard error is .31, and about 72 % of the deviance for the 400 observations is accounted for. The reduction in the standard error despite the smaller sample size results from a substantial decrease in the variance of the response variable once the more extreme values of X are trimmed. Similar improvements in fit may or may not be found for other data.

Overall, these are all very good results. A key reason is that within the window, the response functions are close to linear and flat. A close look at Fig. 6 makes this clear. Enlarging or shrinking the region by as much as 25 % does not materially change the results. Had the relationship not been approximately flat, a very different causal effect estimate could have resulted.

Rows four through six contain the results for a particular form of nonparametric regression: regression splines. In effect, the procedure seeks to fit the response function inductively (Berk 2008a, Chapter 2). A single functional form with a possible offset is assumed, just as in Eq. 7, and the entire dataset is used.

A key tuning parameter determines the amount of smoothing. Here, that tuning parameter is the number of "knots" (k). The larger the number of knots, the more flexible the fitting function and the less smooth the fitted values will be. Results when $k = 3$ suggest that the fitted values are not flexible enough. The story is about the same as for the linear model. When $k = 5$, the fit improves dramatically, but the estimated average treatment effect is a bit too large. When $k = 7$, the fit of the fourth degree polynomial is improved a bit more, and the estimated average treatment effect is almost perfect. Also, the quality of the fit is virtually the same as for the gold standard in the first row.

From the sequence of results, there are several lessons. To begin, the key to obtaining useful estimates of the average treatment effect is to first arrive at good approximation of the relationship between X and Y. To this end, no single method dominates over the variety of scatterplot patterns one is likely to find in practice. In this illustration, all of the estimates obtained were positive, some very close to the truth. In each case, one would easily reject the null hypothesis that the average treatment effect is zero.

With real data, it will often make sense to proceed in the following steps:

1. Construct and examine a scatterplot of Y against X. The goal is to obtain some initial hunches about how Y is related to X and whether there may be a discontinuity at the threshold.
2. Apply the difference in means estimator with several different window sizes. A key factor will be the number of observations within the window. If there are too few, estimated average treatment effect will be very unstable. But enlarging the window may introduce additional bias. One can use a goodness-of-fit measure such as the generalized cross-validation statistic to help determine the best window size. But new proposals have been appearing with some regularity (e.g., Imbens and Kalyanaraman 2009). The goal is to obtain an instructive initial sense of what impact the intervention may be having.
3. Apply a smoother such as lowess or smoothing splines to the data ignoring the W. Set the tuning parameters so that a very flexible fitting function is applied and then try several different values of these tuning parameters. An overlay of the fitted values on the scatterplot will help to reveal how Y is related to X and what sharp shifts up or down may be apparent for different values of X. Ideally there will be only one, and it will located at the threshold. If there are large discontinuities elsewhere, it may lead to suspicions about the one found at the threshold.
4. Apply a form of nonparametric regression with X and W as predictors. Vary the tuning parameters and use a measure like the cross-validation statistic to pick the best model. If the number of observations is relatively large, any of the common fit measures will likely

lead you to the same models. In addition, examine the usual regression diagnostics to provide additional information about model quality. It will often turn out that two or three models are about equally good. Then, the results for all three should be reported. For example, the last two estimates in Table 1 are reasonable.

Given all of the data "snooping," p-values will likely be biased downward. False power will result. It can make sense, therefore, to discount p-values by the Bonferroni method (Westfall and Young 1993, Section 2.3.1). For example, the .05 level might be represented by a probability .005. However, the problems are actually much deeper (Leeb and Potscher 2006; Berk 2008b), and it can be useful to randomly partition the data into a training dataset and a test dataset (Berk et al. 2009). All of the model building is done with the training dataset. At the end, the model is evaluated with the test dataset. A useful discussion of how to use training data and test data can be found in Haste et al. (2009, pp. 219–223). If the sample is too small to effectively partition, the best advice may be to interpret any tests judiciously and require that large and substantively sensible treatment effect estimates materialize in addition to "statistical significance." There is also some ongoing work on proper statistical inference after model selection that should be available soon (Berk et al. 2013).

Conclusions

When the RD design can be implemented properly, it has the same capacity as randomized experiments to obtain unbiased estimates of an average treatment effect. Why then has it not been more widely use? Perhaps the most important reason is that researchers have too often failed to appreciate that a wide variety of social interventions are delivered conditional on some explicit and deterministic rule. Another reason may be that because of the correlation between X and W, the RD design can sometimes deliver less precise estimates than randomized experiments with the same number of subjects. However, another factor is the relationship between the response and the assignment variable, which in some circumstances may be strong enough to effectively compensate. Moreover, when an RD design is built into the way one or more social interventions are delivered, it is often easy to obtain a very large sample at little additional expense. The bulk of the costs associated with data collection are born by the organization that is responsible for the intervention. In short, the RD design can be a very useful tool that should be far more widely exploited in crime and justice settings.

Acknowledgement This entry is an undated revision of a paper of the same name published in the *Handbook of Quantitative Criminology* (Springer, 2009) edited by Alex Piquero and David Weisberg. Work on that paper was funded by a grant from the National Science Foundation: SES-0437169, "Ensemble methods for Data Analysis in the Behavioral, Social and Economic Sciences."

Recommended Reading and References

Berk RA (2008a) Statistical learning from a regression perspective. Springer, New York

Berk RA (2008b) Forecasting methods in crime and justice. In: Hagan J, Schepple KL, Tyler TR (eds) Annual review of law and social science. Annual reviews, Palo Alto

Berk RA, de Leeuw J (1999) An evaluation of California's inmate classification system using a generalized regression discontinuity design. J Am Stat Assoc 94(448):1045–1052

Berk RA, Rauma D (1983) Capitalizing on nonrandom assignment to treatments: a regression discontinuity evaluation of a crime control program. J Am Stat Assoc 78(381):21–27

Berk RA, Barnes G, Ahlman L, Kurtz E (2010a) When second best is good enough: a comparison between a true experiment and a regression discontinuity quasi-experiment. J Exp Criminol 6(2):191–208

Berk RA, Brown L, Zhao L (2010b) Statistical inference after model selection. J Quant Criminol 26:217–236

Berk RA, Brown L, Buja A, Zhang K, Zhai L (2013) Valid post-selection inference. Working paper, Department of Statistics, University of Pennsylvania. Annals of Statistics (in press)

Bowman AW, Pope R, Ismail B (2004) Detecting discontinuities in nonparametric regression curves and surfaces. Stat Comp 16:377–390

Butler DM (2009) A regression discontinuity design analysis of the incumbency advantage and tenure in the U. S. House. Elect Stud 28(1):123–128

Cameron AC, Trivedi PK (2005) Microeconometrics: methods and applications. Cambridge University Press, Cambridge

Campbell DT, Stanley JC (1963) Experimental and quasi-experimental designs for research. Houghton Miffin, Boston

Caughey DM, Sekhon JS (2011) Elections and the regression discontinuity design: lessons from close house races, 1942–2008. Polit Anal 19(4):385–408

Chen MK, Shapiro JM (2007) Do harsher prison conditions reduce recidivism? A discontinuity-based approach. Am Law Econ Rev 9(1):1–29

Cook TD (2008) 'Waiting for Life to Arrive:' a history of the regression-discontinuity design in psychology, statistics, and economics. J Econom 142:636–654

Cook RD, Weisberg S (1999) Applied regression analysis including computing and graphics. Wiley, New York

Fan J, Gijbels I (1996) Local polynomial regression modeling and its applications. Chapman and Hall, London

Freedman DA (2008) Diagnostics cannot have much power against general alternatives. www.stat.berkeley.edu/freedman/

Goldberger AS (1972) Selection bias in evaluating treatment effects: some formal illustrations. University of Wisconsin, Madison

Green DP, Leong TY, Kern HL, Gerber AS, Larimer CW (2009) Testing the accuracy of regression discontinuity analysis using experimental benchmarks. Polit Anal 17(4):400–417, manuscript

Hastie TJ, Tibshirani R (1990) Generalized additive models. Chapman and Hall, New York

Hastie TJ, Tibshirani R, Friedman J (2009) The elements of statistical learning, 2nd edn. Springer, New York

Hoenig JM, Heisey DM (2001) The abuse of power: the pervasive fallacy of power calculation for data analysis. Am Stat 55:19–24

Holland P (1986) Statistics and causal inference. J Am Stat Assoc 8:945–960

Imbens G, Kalyanaraman K (2009) Optimal bandwidth choice for the regression discontinuity estimator. Harvard University, Department of Economics, Working paper

Imbens G, Lemieux T (2008a) Special issue editors' introduction: the regression discontinuity design – theory and applications. J Econom 142:615–635

Imbens G, Lemieux T (2008b) Regression discontinuity designs: a guide to practice. J Econom 142:611–614

Lee DS, Lemieux T (2009) Regression discontinuity designs in economics. National Bureau of Economic Research, Working paper #14723

Leeb H, Potscher BM (2006) Can one estimate the conditional distribution of post-model-selection estimators? Ann Stat 34(5):2554–2591

McCrary J (2008) Manipulation of the running variable in the regression discontinuity design: a density test. J Econom 142:698–714

Neyman J (1923) Sur Les Applications de la Thorie des Probabilits aux Experiences Agricoles: Essai des Principes. Roczniki Nauk Rol- niczych 10:1–51, In Polish

Rubin D (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol 66:688–701

Rubin D (1977) Assignment to treatment group on the basis of a covariate. J Educ Stat 2(1):1–26

Thistlewaite DL, Campbell DT (1960) Regression-discontinuity analysis: an alternative to the ex-post facto design. J Educ Psychol 51:309–317

Trochim WMK (1984) Research design for program evaluation. Sage Publications, Beverly Hills

Trochim WMK (2001) Regression discontinuity design. In: Smelser NJ, Bates PB (eds) International encyclopedia of the social and behavioral sciences, vol 19. Elsevier, Amsterdam, pp 12940–12945

Westfall PH, Young SS (1993) Resampling based multiple testing. Wiley, New York

Recidivism Versus Reoffending

▶ Risk Factors for Prison Recidivism

Recidivist Victimization

▶ Repeat Victimization

Reconstructive Mental State Evaluation

▶ Psychological Autopsy of Equivocal Deaths

Reconstructive Psychological Evaluation

▶ Psychological Autopsy of Equivocal Deaths

Recurrent Victimization

▶ Repeat Victimization

Reentry

► Drug Abuse and Alcohol Dependence Among Inmates

Reentry Courts

Christine H. Lindquist[1], Bryn Ann Herrschaft[2] and Pamela K. Lattimore[1]
[1]Crime, Violence, and Justice, RTI International, Research Triangle Park, NC, USA
[2]Department of Research, Center for Court Innovation, New York, NY, USA

Overview

Reentry courts are "specialized courts that help reduce recidivism and improve public safety through the use of judicial oversight to apply graduated sanctions and positive reinforcement, to marshal resources to support the prisoner's reintegration, and to promote positive behavior by the returning prisoners (Bureau of Justice Assistance 2010)." An estimated 26 reentry courts and 29 reentry drug courts were in operation as of December 31, 2009 (Huddleston and Marlowe 2011). These courts are intended to address the critical needs of returning prisoners – particularly in the period immediately following release – through the combination of judicial oversight and a collaborative case management process. Reentry courts take many forms, with substantial variation along key dimensions of program design, including the mechanism for jurisdictional authority, target population, whether the court is a standalone model or combined with a drug court, and the point of entry into the program. Because the reentry court model remains a relatively new innovation, very few outcome evaluations have been conducted to provide evidence supporting its effectiveness. Preliminary results from the evaluation of the Harlem Parole Reentry Court are promising; however, additional rigorous evaluation is needed.

Introduction: The Origins of Reentry Courts

Reentry courts are "specialized courts that help reduce recidivism and improve public safety through the use of judicial oversight to apply graduated sanctions and positive reinforcement, to marshal resources to support the prisoner's reintegration, and to promote positive behavior by the returning prisoners (Bureau of Justice Assistance 2010)." These courts are intended to address the critical needs of returning prisoners – particularly in the period immediately following release – through the combination of judicial oversight and a collaborative case management process. The underlying goal of reentry courts is to establish a seamless system of offender accountability and support services throughout the reentry process (Bureau of Justice Assistance 2010). For such a system to be "seamless" requires cooperation between corrections and courts, representing a new form of jurisprudence characterized by these agencies working in partnership in pursuit of the common goal of successful offender reintegration (Wilkinson et al. 2004).

Reentry courts originated as a new strategy for managing prisoner reentry, as policymakers contemplated large numbers of returning prisoners and the myriad problems associated with reintegration. These large numbers were the results of changes in incarceration rates, which, after remaining relatively stable from 1925 to 1973, began to increase steadily starting in 1973 (Travis and Petersilia 2001). Over the next 30 and more years, incarceration rates and prison populations increased dramatically, with the growth only beginning to slow in 2006 (West et al. 2010). By the end of 2009 (the most recent year for which data are available), the number of prisoners under state or federal jurisdiction was just over 1.6 million (West et al. 2010). Because nearly all prisoners are released (West et al. 2010), high incarceration rates result in large numbers of returning prisoners each year. Each year since 2005, more than 700,000 prisoners have been released from federal or state prisons. In 2009, 729,295 sentenced prisoners were released from state or federal jurisdiction (West et al. 2010). This number represented the first decline in the number of prison releases since 2000 (West et al. 2010).

The sheer volume of returning prisoners and the overwhelming obstacles faced by many during reentry have contributed to heightened interest in identifying effective strategies for managing the reentry process. It is well documented that returning prisoners have limited occupational or educational experiences to prepare them for employment, drug and alcohol addictions, mental and physical health problems, strained family relations, and limited opportunities due to the stigma of a criminal record (Petersilia 2003; Travis and Visher 2005; Lattimore and Visher 2009). Thus, it is not surprising that as many as two-thirds of released prisoners are rearrested within 3 years, and over 50 % return to prison or jail (Langan and Levin 2002).

Parole violations play a large role in reincarceration rates. The majority of released prisoners – an estimated 74 % of those released from state prisons in 2009 – are released conditionally, including release to parole, probation, supervised mandatory release, or other conditional release (West et al. 2010). However, many parolees do not complete their term successfully, although the trend toward increasingly high failure rates for parolees appears to have reversed in recent years. In 2000, only 43 % of parolees completed their term successfully or were discharged early (Glaze 2007). This percentage increased to 51 % in 2009, following a trend toward increasing success rates that began in 2006 (Glaze and Bonczar 2010). Among those returned to incarceration, the majority are reincarcerated for technical violations, with only 26 % reincarcerated for a new conviction (Glaze and Bonczar 2010). Additionally, the likelihood of successful completion is lower for parolees who have violated a previous parole term when compared with those who have not. Among state parolees discharged in 1999, only 21 % of parolees with a previous violation successfully completed parole, compared to 63.5 % of first releases (Hughes et al. 2001). Despite a shift toward more positive parole outcomes in recent years, consistently high failure rates among parolees suggest the need for alternative strategies for managing reentry.

In response to the high post-release failure rates and the overwhelming needs of returning prisoners, many reentry programs have been implemented over the past several years to facilitate returning ex-offenders' transition from prison and, ultimately, to reduce recidivism. Several high-profile efforts to promote successful prisoner reintegration by involving a variety of community stakeholders in the design and delivery of rehabilitative services have been implemented, including the Reentry Partnerships Initiative, the Serious and Violent Offender Reentry Initiative (SVORI), the Reintegration of Ex-Offenders program (RExO; formerly the Prisoner Reentry Initiative), and most recently the Second Chance Act.

Reentry courts are distinguished from other approaches to prisoner reentry because they combine rehabilitative services with the power of the courts. Reentry courts began as part of the broader national movement towards specialized "problem-solving courts" to address issues such as drug addiction, mental health problems, driving under the influence, and domestic violence. These problem-solving courts address a wide array of problems but share several elements, including treatment mandates, ongoing judicial supervision, collaboration rather than the traditional adversarial system, direct interaction between defendants and the judge, and community outreach.

Based on the success of the drug court movement, it was thought that a similar model could be applied to support prisoner reintegration. The concept of leveraging judicial authority to manage reentry for returning prisoners through reentry courts was first introduced by Jeremy Travis. Reentry courts recognize that the point of reentry is a critical juncture for intervention and that the judiciary should play a greater role in managing reentry (Travis 2000). Traditional criminal justice policy was thought to constrain the authority and reach of parole agencies and correctional institutions, requiring a new organizational strategy for effectively managing reentry (Travis 2000). Further, court-based reentry management was thought to provide an opportunity for the reentry process to begin at sentencing and continue through the release period (Travis 2000).

Since the conceptualization of this model for managing prisoner reentry, multiple jurisdictions have developed reentry courts. The following section describes the early implementation experiences, including federal efforts to support the application of the reentry court model, and program design characteristics of reentry courts.

Background Description

Federal Efforts to Promote Reentry Courts

Early demonstrations of the reentry court model were promoted by the Office of Justice Programs (OJP) through its national Reentry Court Initiative (RCI). In 2000, OJP identified nine pilot projects (including both adult and juvenile programs) to receive technical assistance for the creation of reentry courts that would provide graduated sanctions in combination with social services (OJP 1999). The goal of the RCI was to "establish a seamless system of offender accountability and support services throughout the reentry process" (OJP 1999). Several core elements were identified among the RCI pilot programs (Lindquist et al. 2004). These components include (1) *assessment and planning*, in which an interagency process was used to identify appropriate candidates and ensure needs assessment and planning prior to release; (2) *active oversight* through public judicial status hearings required regularly throughout participation and involving the reentry court team; (3) *social services*, including the identification and marshalling of necessary resources and a case management approach to ensure their appropriate provision; (4) *community accountability* through an advisory board and the use of restitution and victim involvement; (5) *graduated sanctions* in lieu of revocations, with the swift, predictable, and universal administration of a predetermined range of sanctions; and (6) *incentives for success* to recognize program milestones, ideally administered in a public forum.

Federal interest in reentry courts remained strong, and a decade after the RCI, the Bureau of Justice Assistance solicited applications to fund adult reentry courts under the Second Chance Act of 2007 (Pub. L. 110–199). Several reentry courts were funded under the Bureau of Justice Assistance's FY2010 State, Tribal, and Local Reentry Courts Program and the Bureau of Justice Assistance's FY2009 solicitation for Second Chance Act Adult Demonstration Projects. The framework for establishing a reentry court under the Second Chance Act was based on the RCI and lessons learned from other early reentry courts, with grantees required to include the six RCI components described above. An emphasis on the use of validated assessment tools, tracking of services received by participants, and the use of evidence-based practices distinguished the Bureau of Justice Assistance requirements for reentry courts from those of the RCI (Bureau of Justice Assistance 2010).

Current Estimates of Reentry Courts

The number of reentry courts currently in operation is difficult to estimate. Reentry courts have been established in the federal, state, county, and tribal systems, with some entailing a partnership between government entities. As with drug courts, some programs are subject to statewide oversight for reentry courts, including certification, compliance monitoring, training and technical assistance, and evaluation (Wolf 2011). The National Drug Court Institute reported that 26 reentry courts and 29 reentry drug courts – a common model for reentry courts, as described in more detail below – were in operation as of December 31, 2009 (Huddleston and Marlowe 2011). Including federal reentry courts and programs newly implemented as part of the Bureau of Justice Assistance Second Chance Act reentry grants would increase the estimate. However, although there has been interest in the reentry court concept since its introduction, reentry courts have not witnessed the exponential growth that has been observed for drug courts and other types of problem-solving courts. Instead, implementation of the reentry court concept has been tentative. As described below, implementation challenges – particularly jurisdictional authority and the diverse and overwhelming needs of the target population – are likely responsible for the hesitancy in embracing this model.

Program Design Characteristics

Jurisdictional Authority. The mechanism for judicial authority is a major distinguishing factor among reentry courts and represents an area of complexity for jurisdictions interested in implementing this model. Some reentry courts are administrative courts, in which programmatic authority is maintained by the executive branch. In contrast, others are criminal courts, in which the judicial branch has authority.

Administrative reentry courts are typically part of the corrections system. Under this model, participation is established as a condition of release from prison and an administrative reentry court judge presides over the court hearings. Some states use retired judges or other judicial officers to lead these status hearings (see Tauber 2008).

Reentry court programs run by the judicial branch commonly use split-sentencing jurisdiction, in which an offender is sentenced to prison for a determinate sentence and returned to the jurisdiction of the court for a probationary period after serving the prison sentence, under the active supervision of the reentry court judge. In other jurisdictions, the courts have the statutory authority to sentence offenders to prison with the possibility of early discharge for resentencing after a short term or to sentence offenders to prison to complete a rehabilitation program and then be returned to the sentencing judge after successful completion (see Tauber 2008). With reentry courts operated by the judiciary, the sentencing judge retains jurisdiction over a case during the entire sentence.

Finally, some states are experimenting with collaborative reentry court structures, in which both the courts and correctional authorities are involved in the administration of the program. Such models entail either dual jurisdiction, with the court operated by the county and parole jurisdiction held by the correctional authority, or sole jurisdiction by the correctional authority but with active participation of the reentry court (see Tauber 2008). Regardless of the source of jurisdictional authority, reentry courts are typically mandatory in nature, with eligible participants required to participate.

Target Populations. Risk variability among reentry court participants is a main deviation from the drug court model. Although some reentry courts target a particular population, including narrowly defined offense types or clients with specific treatment needs, such as mental health or addiction issues (Lindquist et al. 2004), most reentry courts do not appear to specialize with regard to client type, resulting in the need to address a variety of client risks and needs. In addition to substance abuse treatment, reentry courts commonly deliver mental health counseling, assistance with housing, transportation, education, vocational training, employment, and other support services oriented toward returning prisoners (Lindquist et al. 2004). This holistic service delivery approach implies a larger network of service providers for reentry courts than for drug courts.

In addition, whereas drug court clients are typically nonviolent, there is a growing trend among reentry courts to focus on individuals assessed at high risk (Wolf 2011). Some reentry courts enroll several "tracks" of participants, including not only newly reentering prisoners but also parole violators assigned to reentry court rather than having their supervision revoked, as well as probationers sentenced to community supervision and reentry court rather than incarceration. However, it is questionable whether serving individuals who are not returning from incarceration is consistent with the original reentry court model; some programs that primarily serve individuals diverted from incarceration are called "pre-entry courts" and seek to keep individuals out of prison (Wolf 2011).

Reentry courts have been designed for both juveniles and adults. In 2005, the National Council of Juvenile and Family Court Judges published a guide for planning, implementing, and operating a juvenile court to manage the reintegration into the community of juveniles who have undergone court-ordered, out-of-home placement.

Reentry Drug Courts. Reflecting the widespread need for substance abuse treatment among returning prisoners, a common model for reentry courts entails the combination of reentry courts with drug courts (i.e., “reentry drug courts”). As noted earlier, slightly more reentry drug courts were in operation in 2009 (n=29) than reentry courts not combined with drug courts (n=26) (Huddleston and Marlowe 2011). Although several models for reentry drug courts have been proposed, the most common model currently in operation appears to entail jail- or prison-based substance using offenders being released from incarceration directly to a drug court. (See Tauber 2009, for a discussion of potential models for diverting prison-bound nonviolent drug-involved offenders to local jails for jail-based reentry drug courts and violent drug-involved offenders to prison-based reentry drug courts.) Importantly, reentry drug courts use the drug court model in facilitating the reintegration of drug-involved offenders into the community upon their release (Huddleston and Marlowe 2011).

Reentry drug court participants receive regular judicial monitoring, intensive treatment, community supervision, drug testing, and ancillary services necessary for successful community reintegration (Tauber and Huddleston 1999). The reentry drug court model can ensure continuity of treatment and accountability for returning prisoners from incarceration through reentry. Furthermore, it can allow jurisdictions with a drug court to capitalize on this existing infrastructure.

Point of Programmatic Entry and Exit. Reentry courts were originally conceptualized so that jurisdictional oversight began at sentencing and continued through the period of release (Travis 2000). Indeed, the calls for a collaborative approach were based on the assumption that the tasks of preparing offenders for release and transitioning them back into the community would be coordinated between courts and corrections (see Wilkinson et al. 2004; Burke 2001).

Some reentry courts do entail this level of coordination and consistent involvement beginning at sentencing. For example, in some courts the sentencing judge makes decisions about where a person is incarcerated, exerts influence over the programming he or she will receive while incarcerated, and stays updated about his or her progress (Wolf 2011). However, other reentry courts do not identify or enroll participants until immediately prior to, or even after, release (Lindquist et al. 2004). The post-release phase appears to be increasingly emphasized. For example, NDCI’s current definition of reentry courts emphasizes court involvement “during the initial phases of. . .community integration” (Huddleston and Marlowe 2011). Although early identification and the ongoing involvement of the sentencing judge offer several advantages, this model may not be possible for all programs. Many courts do not have the capacity to “reach in” to prisons or jails for screening, enrollment, or monitoring of prerelease services. However, some reentry courts transfer participants to transitional housing units in the correctional facility prior to release or to secure residential facilities in the community (Wolf 2011).

The point of exit for reentry court participants, particularly the degree of overlap between reentry court participation and parole (or other post-release supervision), has not been well documented. In some programs, reentry court participation extends throughout the period of post-release supervision. In others, reentry court has a defined duration, such as 6–12 months (Lindquist et al. 2004). According to Huddleston and Marlowe (2011), reentry court graduates are often transferred to traditional parole supervision, at which point they may continue to receive case management services through the reentry court on a voluntary basis. However, very little documentation exists about the duration of reentry court participation relative to participants’ parole sentence (or other post-release supervision mechanism) and the mechanism by which graduates exit reentry court.

Reentry Court Practices

Assessment and planning has been strongly emphasized by the Bureau of Justice Assistance in its oversight of the Second Chance Act Reentry Courts. In particular, the need for using validated assessment tools to identify appropriate candidates for reentry courts, with individuals at high risk of recidivism increasingly recognized as benefitting the most from such programming, has been widely disseminated. The use of needs assessments for planning individualized services (both prior to and after release) and continually reassessing service needs has received less attention. Furthermore, the extent to which best practices in assessment and planning are implemented in the field has not been documented.

Active judicial oversight, which includes public judicial status hearings required regularly throughout participation and involving the reentry court team, has received substantial attention, and several ingredients are thought to be particularly salient (see Hamilton 2010). As with drug courts, reentry courts leverage ▶ judicial authority through formal court proceedings (Travis 2005), with the judge's role expanded beyond traditional criminal court proceeding to include involvement in reentry planning and case management in a collaborative manner. Second, ▶ procedural justice is thought to be a critical output resulting from the use of the public forum (through court hearings). Because the reentry court judge is responsible for openly and publicly explaining program requirements and the purpose of sanctions (Travis 2005), it is thought that under the reentry court model (as with other problem-solving courts), participants develop a greater perception of fairness, which, in turn, increases the likelihood of compliance (Tyler 1988). Furthermore, rather than the closed-door process typically used by traditional parole, the reentry court utilizes a public forum that is thought to contribute to a greater perception of fairness. Finally, consistent with the drug court model, ceremonies or other activities intended to recognize accomplishments are used to incentivize and mark interim compliance and successful participation culminates in a public graduation (Travis 2005).

The remaining core elements, such as graduated sanctions, incentives for success, community accountability, and social services, have not been well documented among the reentry courts currently in operation. While reentry courts do appear to use graduated sanctions and incentives for success, the manner in which they are used has not been well documented. Similarly, little is known about reentry courts' strategies for identifying and marshalling needed social services and incorporating principles of community accountability among such a diverse population.

State of the Art: The Effectiveness of Reentry Courts

Because the reentry court model remains a relatively new innovation, much of the existing research involves only case studies, process evaluations, and documentation of implementation challenges that have arisen in particular programs. Few outcome and impact evaluations have been conducted to provide evidence supporting the effectiveness of the reentry court model. The limited research available to date, most of which has substantial methodological limitations, is summarized below.

In an evaluation of Delaware's Reentry Drug Court, which was implemented in 1993 and is designed to assist offenders who reenter the community following the completion of a residential treatment program, rearrest rates for program graduates were compared to those for the general prison population released between 1981 and 1984. The reentry drug court graduates were less likely to be rearrested for a new felony crime (9 %) than general releases (28 %). Individuals who enrolled in, but did not graduate from, the program were slightly more likely to be arrested (29.1 %) than general releases (Gebelein 2003).

A similar approach was employed in an early evaluation of the Allen County, Indiana Reentry Court. The program was established in 2001 as a complement to the county's Community Transition Program, which accepts selected inmates to a prerelease probation period prior to their anticipated release date. Participation in the project is voluntary and involves 24-h supervision, judicial monitoring and sanctioning, and relevant service referrals by a reentry court team that includes a judge, case managers, treatment professionals, and parole officers. An outcome evaluation was conducted to compare recidivism rates for program participants to national averages. Overall, 3-year rearrest rates for graduates and nongraduates of the program were lower than the national average of 67.5 % (Pearson-Nelson 2009).

The federal reentry court in Oregon (the District of Oregon), which began in 2006 as a strategy to reduce recidivism among drug-involved offenders following their release from a completed term of incarceration, entails judicial supervision through monthly hearings and reentry services that address risk factors including substance abuse, employment, and housing. Participation is voluntary and graduates receive a 1-year reduction of their probation sentence following successful completion (Aubin 2009). An evaluation of the program found that reentry court participants had higher levels of supervision and service receipt than a comparison group, but lower levels of employment and more sanctions (Close et al. 2008).

These three outcome evaluations have significant limitations that make it difficult to draw conclusions about the effectiveness of the reentry court program being evaluated. The most empirically rigorous reentry court evaluation published to date focused on the Harlem Parole Reentry Court and was conducted by the Center for Court Innovation (Hamilton 2010). The Harlem Parole Reentry Court, in operation since 2001, is a collaboration between the Center for Court Innovation and the New York State Division of Criminal Justice Services and Division of Parole. The program provides judicial oversight, supervision, and services to parolees returning to Upper Manhattan in New York City for 6 months following their release from incarceration (Hamilton 2010).

The evaluation used propensity score matching to select a comparison group of parolees released to traditional supervision during the study period. Recidivism outcomes were tracked for all parolees in the sample for 3 years following their release from incarceration (Hamilton 2010). Over the 3-year follow-up period, the reentry court program was found to have a positive impact on preventing new criminal behavior, including rearrest and reconviction for new crimes. Reentry court parolees were less likely to be rearrested and were significantly less likely to be rearrested for a misdemeanor in the first year and drug offenses in the first 2 years. Reentry court parolees were also significantly less likely to be reconvicted than comparison parolees over all 3 years at risk in the community. At 3 years, reentry court participants were 10 % less likely to have been reconvicted than comparison parolees. However, reentry court parolees were found to be significantly more likely to experience revocation after 2–3 years than comparison parolees. Three years following release, reentry court parolees were 18 % more likely to have experienced a revocation, particularly revocations due to technical violations.

These results indicate that the reentry court may have a negative impact on its parolees that may be due to "supervision effects," in which lower caseloads and more intensive supervision lead to increased opportunities for the detection of (and response to) noncompliant behavior. The report suggests the exploration of graduated and alternative sanctions protocols in reentry courts to reduce or avoid supervision effects. Currently, the Center for Court Innovation is implementing a randomized control trial of the Harlem Parole Reentry Court following additional programmatic modifications that will produce an empirically rigorous evaluation of the impact of the reentry court on recidivism and other outcomes important to reintegration, including employment and substance abuse.

In addition to the ongoing experimental evaluation of the Harlem Parole Reentry Court, a multi-site evaluation of the 2010 Second Chance Act Reentry Court grantees is also underway. The evaluation is funded by the National Institute of Justice (NIJ) and led by Northwest Professional Consortium (NPC), in collaboration with Research Triangle Institute (RTI) International and the Center for Court Innovation. The study, known as NIJ's Evaluation of Second Chance Act Adult Reentry Courts (NESCAARC), includes a process evaluation, impact evaluation, and cost-benefit analysis.

In addition, the evaluation will contribute to the development of a true reentry court model by determining the extent to which the sites reveal a common framework. Results are expected to be available beginning in 2015.

Issues and Research Gaps

Awareness of the need to effectively manage prisoner reentry has been heightened in the past several decades. As a result, reentry courts are an innovative strategy for managing the complex problems associated with prisoner reentry. Yet this model remains a fledgling approach, with both implementation strategies and evaluation lagging behind other problem-solving courts.

Part of the hesitancy to embrace the reentry court model may be due to implementation challenges, which can be substantial. Issues such as establishing jurisdictional authority over the target population, developing strategies for the identification and monitoring of participants prior to release, and serving the wide ranging needs of such a diverse population may appear (or be) insurmountable for some jurisdictions. Better documentation is needed from programs that have successfully implemented a reentry court model, including documentation of the challenges that were encountered and the solutions employed to overcome these challenges. Several models of jurisdictional authority have been proposed by Judge Jeffrey Tauber, editor of Reentry Court Solutions. Yet more documentation about the application of these models or barriers associated with each is needed.

Additionally, strategies for courts (particularly programs operated by the judicial branch) to "reach in" to prisons or jails to access participants prior to release have not received as much attention. Many courts do not have the capability to engage in prerelease planning and coordination, leading to a tendency for most reentry courts to focus exclusively on post-release programming. However, an uninterrupted continuum of services, particularly substance abuse, mental health, and healthcare services, is recognized as one of the keys to reentry success following release. A necessary component of the reentry court model is the early identification of participants' needs and the provision of a seamless transition of services from facility to community. An understanding of the necessary steps to establish a continuum of services within reentry court models is necessary to provide maximum substantial benefits for program participants.

Additional attention to develop strategies for adequately meeting the needs of the target population is needed, along with strategies for generating public support for the focus on returning prisoners. It has been observed that the application of the drug court model to returning prisoners may be less successful and have less popular and political support because while drug courts concentrate on the problem of addiction and serve a relatively homogenous group of clients (nonviolent drug addicts), reentering prisoners have numerous, diverse needs and are often at high risk of violence (Maruna and LeBel 2003). Not only are reentry court participants at higher risk, they may have served substantial terms of incarceration. As a result, they may need more services than drug court participants. In addition, it has been argued that more flexibility in response to technical violations and minor offenses is required for returning prisoners, rather than a "heavy handed" approach. Encouragement, incentives, and a sense of community are thought to be particularly important.

Research on operational barriers, such as interagency cooperation, is also needed. Although not documented in the research literature, it is possible that lack of interagency cooperation may be more common with reentry courts than other problem-solving courts because partners may have different visions for the program or agreement on the target population. For example, decisions about whether to include violent or sex offenders have implications for the level of support from prosecutors, acceptance from service providers, intensity of supervision expected from probation or parole officers, and support from the public for the program. Additionally, changing attitudes and agency culture can also be difficult to bring collaborative agency staff in line with the reentry court's mission and goals. Farole (2003) reported that staff cooperation, communication, and turf issues with parole officers were a significant implementation challenge for the Harlem Parole Reentry Court in its initial stages. Without the buy-in and cooperation of all court stakeholders, including outside agency staff, the reentry court model will not achieve the desired levels of program fidelity and therefore, may fail to successfully achieve its goals.

Finally, as noted above, rigorous studies of the effectiveness and cost-effectiveness of reentry courts are needed. Multi-site evaluations and research on approaches to minimize supervision effects are particularly needed, along with the exploration of a broader set of outcomes (in addition to reincarceration) and longer-term follow-up. Importantly, such studies must employ a rigorous strategy for identifying a comparison group that is well matched to the reentry court participants on all eligibility criteria when an experimental design and random assignment are not possible. In addition, evaluations should employ an "intent to treat" approach such that all individuals assigned to the reentry court (not just successful completers) are included in the treatment group. This approach is the most rigorous test of program effectiveness because it does not rely solely on the experiences of clear program "successes."

Very little is known about the cost-effectiveness of reentry courts, and research in this area would be of substantial interest. The issue of which risk levels are targeted by reentry courts may have implications for the cost-benefit ratio of the reentry court. The growing trend is to target offenders assessed as being high risk, typically with regard to risk of violence or serious crime. However, participants in the Bureau of Justice Assistance focus group noted that greater cost savings to the criminal justice system would likely be achieved if risk were defined as risk for reoffending, which is generally higher among low-level drug offenders and the mentally ill (Wolf 2011).

Related Entries

- ▶ Drug Courts
- ▶ Mental Health Courts
- ▶ Parole
- ▶ Parole and Prisoner Reentry in the United States
- ▶ Problem-Solving Courts

Recommended Reading and References

Aubin M (2009) The district of oregon reentry court: an evidence-based model. Fed Sentencing Rep 22(1):39–43

Burke PB (2001) Collaboration for successful prisoner reentry: the role of parole and the courts. Correct Manag Q 5(3):11–22

Bureau of Justice Assistance (1997) Defining drug courts: the key components. U.S. Department of Justice, Office of Justice Programs, Washington, DC, NCJ 205621

Bureau of Justice Assistance (2010) Second chance act state, local, and tribal reentry courts FY2010 competitive grant announcement. U.S. Department of Justice, Office of Justice Programs, Washington, DC, OMB No. 1121-0329. http://www.ojp.usdoj.gov/BJA/grant/10SecondChanceCourtsSol.pdf

Close DW, Alltucker K, Aubin M et al (2008) The District of Oregon Reentry Court: Evaluation, Policy Recommendations, and Replication Strategies. United States District Court – District of Oregon. http://www.ussc.gov/Education_and_Training/Annual_National_Training_Seminar/2009/008c_Reentry_Court_Doc.pdf

Farole D (2003) The harlem parole reentry court evaluation: implementation and preliminary impacts. Center for Court Innovation, New York, http://www.courtinnovation.org/sites/default/files/harlemreentryeval.pdf

Gebelein RS (2003) Delaware's reentry drug court: a practical approach to substance abusing offenders. Paper prepared for the conference on European Perspectives on Drug Courts. http://courts.delaware.gov/Superior/pdf/Reentry_France_27Mar03.pdf

Glaze L (2007) Probation and parole in the United States, 2006. U.S. Department of Justice, Bureau of Justice Statistics, Washington, DC, NCJ 220218

Glaze L, Bonczar TP (2010) Probation and parole in the United States, 2009. U.S. Department of Justice, Bureau of Justice Statistics, Washington, DC, NCJ 231674

Hamilton Z (2010) Do reentry courts reduce recidivism? Results from the Harlem parole reentry court. Center for court innovation. http://www.courtinnovation.org/sites/default/files/Reentry_Evaluation.pdf

Huddleston C, Marlowe D (2011) Painting the current picture: a national report card on drug courts and other problem-solving court programs in the United States. National Drug Court Institute, Alexandria, http://www.ndci.org/sites/default/files/nadcp/PCP%20Report%20FINAL.PDF

Hughes TA, Wilson DJ, Beck AJ (2001) Trends in State parole, 1990-2000. U.S. Department of Justice, Bureau of Justice Statistics, Washington, DC, NCJ 184735

Langan P, Levin D (2002) Recidivism of prisoners released in 1994. US Department of Justice, Bureau of Justice Statistics, Washington, DC, NCJ 193427
Lattimore PK, Visher C (2009) The multi-site evaluation of SVORI: summary and synthesis. Research Triangle Institute, Research Triangle Park
Lindquist C, Hardison J, Lattimore P (2004) The reentry court initiative: court-based strategies for managing released prisoners. Justice Res Policy 6(1):97–118
Maruna S, LeBel T (2003) Welcome home? Examining the "reentry court" concept from a strengths-based perspective. West Criminol Rev 4(2):91–107
Office of Justice Programs (1999) Reentry courts: managing the transition from prison to community, a call for concept papers. U.S. Department of Justice, Office of Justice Programs, Washington DC
Pearson-Nelson D (2009) Allen county community corrections – reentry court program impact evaluation. School of Public and Environmental Affairs, Indiana University-Purdue University Fort Wayne, http://www.allencountycorrections.com/pdfs/reports/ReEntry%20Court%20Impact%20Evaluation%2002%2004%2009.pdf
Petersilia J (2003) When prisoners come home: parole and prisoner reentry. Oxford University Press, New York
Tauber J (2008) State prison based models. http://www.reentrycourtsolutions.com/wp-content/uploads/2009/11/STATE-PRISON-BASED-MODELS.pdf
Tauber J (2009) A state reentry/drug court proposal: a means to achieve correctional reform. http://www.reentrycourtsolutions.com/wp-content/uploads/2009/10/A-state-reentry-drug-court-proposal.pdf
Tauber J, Huddleston CW (1999) Reentry drug courts: closing the gap. National Drug Court Institute, Alexandria
Travis J (2000) But they all come back: rethinking prisoner reentry. U.S. Department of Justice, Office of Justice Programs, Washington, DC, Sentencing & Corrections: Issues for the 21st Century, No. 7
Travis J (2005) But they all come back: facing the challenges of prisoner reentry. Urban Institute Press, Washington, DC
Travis J, Petersilia J (2001) Reentry reconsidered: a new look at an old question. Crime Delinq 47:291–313
Travis J, Visher C (eds) (2005) Prisoner reentry and crime in America. Cambridge University Press, New York
Tyler T (1988) The social psychology of procedural justice. Plenum, New York
West HC, Sabol WJ, Greenman SJ (2010) Prisoners in 2009. U.S. Department of Justice, Bureau of Justice Statistics, Washington, DC, NCJ 231675
Wilkinson RA, Bucholtz GA, Siegfried GM et al. (2004) "Prison reform through offender reentry: a partnership between courts and corrections". Pace Law Review. Paper 10. http://digitalcommons.pace.edu/lawrev/10
Wolf RV (2011) Reentry courts: looking ahead: a conversation about strategies for offender reintegration. Bureau of Justice Statistics, Washington, DC, U.S. Department of Justice, Bureau of Justice Assistance

Reform

▶ Punishment as Rehabilitation

Reformation

▶ History of Corrections

Regulate Violence

▶ Responses to Domestic Violence in India

Regulatory Compliance

▶ Compliance and Corporate Crime Control

Regulatory Crime Control

▶ Legal Frameworks for Third-Party Policing

Rehabilitation

Gwen Robinson
Reader in Criminal Justice, School of Law,
University of Sheffield, Sheffield, UK

Overview

Rehabilitation is a broad but somewhat slippery concept which denotes a wide variety of interventions aimed at promoting desistance from offending, but which can also denote the symbolic restoration of an ex-offender to the status of a law-abiding citizen. This entry presents a critical overview of the ideas, theories, and practices that have been associated with the project(s) of "offender rehabilitation." This entry begins with a discussion of the meaning(s) of rehabilitation in relation to those who break the law, and in relation to broader ideas about the punishment of offenders and the imposition of criminal sanctions. It then moves into a discussion of the various rationales or justifications for offender rehabilitation and some of the key ethical issues associated with rehabilitative practices. Subsequently, this entry considers rehabilitation from an empirical perspective, addressing the thorny issue of whether rehabilitative interventions "work" and the quality of the evidence underpinning claims to effectiveness. The penultimate section analyses the "reaffirmation" of rehabilitation in light of contested ideas about developments in the social and penal spheres which have emphasized the rise of punitive and/or actuarial logics, arguing that the continuing salience of rehabilitative discourses and practices are indicative of their continuing evolution in late-modern contexts. Finally, this entry returns to the practice sphere, and considers some possible future directions for rehabilitation, with reference to some of the contemporary challenges to the "new orthodoxy" of the RNR (risk-need-responsivity) model which has come to dominate practice in many Western jurisdictions.

What Is Rehabilitation?

Rehabilitation is something of a slippery concept, which is surprisingly difficult to pin down (Raynor and Robinson 2009, Chap. 1). General dictionary definitions tend to define rehabilitation with reference to a process of "restoration" – that is, a return to a previous (proper/desirable) condition or status. This certainly makes sense in medical contexts, where rehabilitation refers to a process of regaining or recovering physical strength or cognitive functioning. And in a general sense, we can think of the rehabilitation of offenders in terms of "restoration" – that is, returning to a former, more desirable state. However, as a number of commentators have observed, there are problems with such a conceptualization. These include the fact that the social circumstances of some offenders are such that they have never enjoyed a "desirable" position within society to which they would wish to return, and the related observation that the rehabilitation of offenders often implies a process of *improvement*, by means of the acquisition of new skills, etc.

In the context of offending, rehabilitation usually implies behavioral change: the "rehabilitated offender" is generally regarded as one who has desisted from offending. And although desistance may be achieved with or without the "intervention" of professionals, programs, etc., "offender rehabilitation" commonly conjures up such images, and in much of the literature it is discussed – albeit rarely defined – in these terms. It is important to note, however, that the notion of offender rehabilitation is not confined to the behavioral realm: it also has an important symbolic dimension, although this is often neglected in the literature. Indeed, as Garland (1985) has observed, the concept of rehabilitation actually has its roots in seventeenth-century French law, where it denoted the destruction or "undoing" of a criminal conviction (see also Herzog-Evans 2011). In its fullest sense, then, offender rehabilitation implies not only behavioral change, but also the restoration of the rights of citizenship, and of reputation: or what Braithwaite (1989) has described as the "decertification of deviance."

Rehabilitation and Punishment

In the criminal courts, sentencing decisions are guided by penal philosophies which serve to justify the sanctions imposed upon offenders. Rehabilitation (or "reform") is one of three main *consequentialist* strategies, all of which seek to justify punishment or penal sanctions with reference to the desirability of their future consequences. In common with *deterrence* and *incapacitation*, rehabilitation is oriented to the goal of crime reduction, but it differs from these other consequentialist strategies in that it is usually considered in positive terms. While incapacitation involves physically restraining or removing the offender from society, and deterrence centers on emphasizing the unpleasantness of punishment, "rehabilitative punishment" seeks to reduce the likelihood of an individual reoffending by instituting changes in the offender. Cavadino and Dignan state that:

> Reform (or 'rehabilitation') is the idea that punishment can reduce the incidence of crime by taking a form which will improve the individual offender's character or behaviour and make him or her less likely to reoffend in future. (2007, pp. 41–42)

However, this image of rehabilitation as a positive process is somewhat controversial, especially when it comes to thinking about the means or methods of bringing about offender rehabilitation which have been deployed throughout history – and some in the present day. The fictitious *Ludovico Technique* featured in Anthony Burgess's (1962) novel "A Clockwork Orange" is often held up as a particularly disturbing example of what might potentially pass as a legitimate "rehabilitative" intervention. Indeed, there is nothing in the theory of rehabilitation that guarantees fair or humane treatment, albeit that its advocates have generally seen themselves as well-intentioned "good guys."

Sentencing with a view to achieving the rehabilitation of the offender does not imply a specific penal sanction. Although many of its contemporary advocates emphasize alternatives to incarceration (and indeed diversionary measures, particularly for juveniles) as the best prospect for achieving rehabilitation, early ideas about the reform of offenders centered on finding ways to render prisons rehabilitative. In late eighteenth-century England, the magistrate and penal reformer John Howard proposed that the morals of offenders might be corrected by subjecting them to a regime of "penitentiary discipline" characterized by solitude and silence, substantial doses of religious instruction, and the inculcation of a work ethic through enforced labor (Howard 1929 [1777]). Just a few years later, Jeremy Bentham put forward his infamous "Panopticon" design for a prison modeled on a factory, and his proposed regime centered upon the reformative effects of permanent surveillance and the productive labor of prisoners working in their cells for up to sixteen hours per day. Ever since then, opinion has been divided as to whether custodial institutions can and should aim to be "rehabilitative," what that might mean in practice and whether it is affordable. However, it should be noted that in some European countries (e.g., Germany), "resocialization" (a term which implies the social reintegration of ex-prisoners) continues to be an explicit objective for the implementation of sentences of imprisonment (see Van Zyl Smit and Snacken 2009). In the twentieth century, noncustodial measures (most notably probation) developed in a variety of jurisdictions with an explicitly rehabilitative remit. In some jurisdictions, at least some noncustodial options continue to have the legal status of "alternatives to punishment," but in others (such as England and Wales) contemporary supervisory penalties have evolved into sanctions in their own right.

Today, the idea that punishment or penal sanctions can deliver rehabilitation is contested. Indeed, rehabilitation has been theorized by some as a process that is conceptually divorced from punishment, such that rather than being conceived as an objective of a positive process of punishment, it has been conceived as an *antidote* to punishment (or, to be more precise, the harmful effects of the punishment process) (e.g., Cullen and Gilbert 1982; Rotman 1990). These writers argue that, while punishment may be deserved, the offender should not be unduly damaged by it, and any harm or handicap inflicted upon him or her ought to be mitigated or compensated by rehabilitative measures. According to this perspective, rehabilitation is understood as a process of undoing the harm caused by punishment, not as an objective or type of punishment or sanctioning.

Justifying Rehabilitation

For more than 200 years, rehabilitation has been advocated both as a humanitarian response to crime and as a means of bringing about the humanization (or civilization) of criminal sanctions and punishment. For example, writing in England in the late eighteenth century, John Howard developed his own ideas about the reform of offenders as an alternative to what he saw as the overly harsh and disproportionate deterrent punishments of the time, which centered upon the public rituals of hanging, whipping, and pillory as responses to the full range of offenses (Howard 1929 [1777]). More recently, writers such as Cullen and Gilbert (1982) and Rotman (1990) have continued to promote rehabilitation on similar grounds. For example, Rotman has argued that:

> The rehabilitation of criminal offenders offers the criminal justice system a unique avenue of improvement [and] has enormous potential for humanizing and civilizing social reaction against crime. Modern rehabilitative policies challenge the fantasy that the dark side of society can be forgotten and that deviants can be simply packed off to prison. They propose instead to offer inmates a sound and trustworthy opportunity to remake their lives. (Rotman 1990, p. 1)

During the twentieth century, rehabilitative punishment began increasingly to be justified with reference to positivist criminologies which emphasized the role of forces (biological, psychological, or social/environmental) beyond the control of the individual in the causation of crime. Gradually the notion of the offender as a "sinner" bearing full responsibility for his or her actions came to be displaced by a view of the offender as very much like a person suffering from an illness beyond his or her control. In the context of this emerging "medical model" of crime, it followed that offenders deserved to be helped or treated, in order to remove or ameliorate the factors responsible. It was this set of assumptions which lay behind the so-called individualized treatment model of offender rehabilitation, which dominated penal policies in the USA, the UK, and elsewhere in the middle decades of the twentieth century.

One way to justify rehabilitation, then, is with reference to positivist criminologies, and the general view of offenders as victims (more or less) of forces beyond their control. Some have, however, taken this argument further, positing that there is a case to be made for rehabilitation as a *right* of the offender, or an *obligation* of the state. Cullen and Gilbert (1982), for example, have argued that where social deprivation (e.g., poverty) or other problems which society has failed to properly address (e.g., mental illness) can be shown to have played a role in crime causation, then society arguably has an obligation to help the offender to overcome such obstacles – that is, a right to receive such help or intervention as to enable him or her to avoid offending in the future. A weaker version of the "rights-based" perspective invokes the idea of rehabilitation as an antidote to punishment, considered above, and it is most commonly linked with sentences of imprisonment. If we accept that offenders have a right not be debilitated or harmed (psychologically, socially, economically, etc.) by the process of punishment, then we can justify "rehabilitative" measures as a right of offenders. It is in this spirit that Rotman (1990) advocates the provision of educational and vocational opportunities, psychological and psychiatric treatment services and the maintenance of community and family links for prisoners. He further argues that the offender's right to rehabilitation is consistent with the full restoration of his or her civil and political rights of citizenship following release from prison. This perspective is to be found in the constitutional right to "resocialization" for (ex) prisoners which exists in some European countries (see van Zyl Smit and Snacken 2009).

Rehabilitation can also – and commonly is – justified with reference to *utilitarian* arguments. Utilitarianism is a philosophy which originates in the work of Jeremy Bentham (1748–1832) and it essentially posits that an action is ethically or morally "right" if it produces the best overall consequences, often expressed as "the greatest happiness for the greatest number." Justifying rehabilitation on utilitarian grounds therefore means emphasizing the potential benefits not just for offenders themselves, but more broadly for the communities and societies in which they live. These benefits are usually couched in terms of both maximizing the availability of useful, productive (and tax-paying) members of society, and minimizing future harm by preventing further crime and victimization (see further Raynor and Robinson 2009, Chap. 2). In contemporary discourse, these desirable outcomes are commonly described in terms of "risk management" and "public protection."

Is It Ethical? The Darker Side of Rehabilitation

Rehabilitation is generally represented and understood as a "force for good." However, it is no stranger to problems and controversies. Indeed, as Ward and Maruna (2007) have pointed out, rehabilitation has – some of the time, and to some people – been considered a "dirty word." There are, then, a number of problems and risks associated with the ideas and practices of rehabilitation, and these help to explain why it is that rehabilitation has waxed and waned in popularity across time and space.

The first risk relates specifically to *rehabilitative punishment*, and it concerns issues of justice, proportionality, and equity. Advocates of retributive justice and desert-based sentencing contend that sanctioning with a view to rehabilitating the offender is potentially problematic in that it can invite periods of punishment which are not commensurate with the seriousness of the offense committed. Indeed, this was one of the key criticisms which presaged the demise of the "individualized treatment model" in the 1960s and 1970s. The logic of the "treatment" approach was that individual offenders should be "treated" until such a time that they were considered (by an "expert") to have been "cured" of their criminal tendencies. Because the amount of time or intervention could rarely be specified in advance, rehabilitative sentencing came to be associated with sentences of indeterminate length, and the empowerment of extrajudicial professionals in decision-making. This was a particular issue in the USA, where indeterminate sentencing was more common than it was elsewhere. In the early 1970s, a report prepared by the American Friends Service Committee (1971) was instrumental in exposing the ethical flaws of the treatment model, and in particular the numerous instances in which doctors, social workers, and other "experts" were empowered to make executive decisions which directly affected people's freedom and autonomy, and which were not subject to challenge. Contemporary advocates of rehabilitation tend to accept the risks to justice posed by rehabilitative sentencing and argue that the presumed benefits of rehabilitation should not be used to justify sentences that are longer than commensurate with desert (e.g., Raynor 1997; Rotman 1990).

A second criticism of rehabilitation concerns its coercive potential and the question of whether this can ever be justified. Rotman (1990) draws a useful distinction between two models of rehabilitation – *authoritarian* and *anthropocentric* – which he sees as, respectively, authoritarian/oppressive, and liberty-centered/humanist. Rotman himself advocates only the latter model, arguing that attempts to rehabilitate offenders should be subject to the offender's consent and should invite the active participation of offenders in their own rehabilitation process. Others, however, argue that coercion is not inherently unethical; nor is it entirely avoidable, not least because rehabilitative programs are often a compulsory part of sentences, especially for those who are assessed as posing a high risk of harm to the public (e.g., sex offenders, violent offenders, and substance misusing offenders) (e.g., Carlen 1989; Day et al. 2004). Day et al. have further argued that a degree of coercion can be effective in both getting reluctant offenders into treatment programs and keeping them there.

Of course, whether we are prepared to accept a degree of coercion in efforts to rehabilitate offenders may well depend to a large extent on what it is that we are potentially requiring them to do. While most of us will be unlikely to object to the idea of forcing a reluctant pedophile to undertake a treatment program to reduce the likelihood of further sexual offending, we might be more inclined to object to the idea of compulsory surgical or chemical castration, or Burgess's violent aversion therapy, mentioned above. A third criticism of rehabilitation then concerns its *content*, and the risk of justifying abusive and/or inhumane interventions in the name of "doing good." Even a cursory knowledge of the history of attempts to rehabilitate offenders reveals that the practices for which "rehabilitative effects" have been claimed have included the physically demanding treadmill, through extended periods of solitary confinement, to psychosurgical and pharmacological interventions. Indeed, not all of these examples are confined to history: it is notable, for example, that the "chemical castration" of sexual offenders is currently enjoying something of a revival in a variety of jurisdictions.

A fourth problem with offender rehabilitation concerns the problems associated with making resources and "help" available to offenders, while such assistance may be unavailable to other members of society who have not offended but are experiencing similar or more extreme personal problems or socioeconomic disadvantages. This is a problem which is particularly pertinent in relation to the provision of resources such as employment training, education, and accommodation, which are sometimes referred to as "social rehabilitation" (Robinson and Crow 2009, Chap. 8). Where this occurs, rehabilitation is said to offend the principle of "less eligibility" formulated by Bentham in the eighteenth century. This principle states that those convicted of offenses should not enjoy conditions that are more favorable than those enjoyed (or perhaps endured) by the poorest independent laborer. In contemporary societies, this means that offenders should not obtain advantages over those in the lowest paid employment, who have not offended.

Does It Work? Evaluating Effectiveness

Despite the ethical concerns raised above, by far the most hotly debated issue with respect to rehabilitation – and in particular rehabilitative punishment – relates to its effectiveness in bringing about reductions in recidivism. However, concerns about "what (if anything) works" are a relatively recent phenomenon. Until the 1960s, rehabilitation was a dominant penal narrative in the USA and the UK, the effectiveness of rehabilitative sanctions was widely taken for granted and little research had been undertaken to establish their impact on recidivism. It was not until the 1970s that serious doubts began to emerge about their effectiveness, and these were due in large part (though by no means entirely) to the publication of a now infamous article by Robert Martinson (1974), which summarized the findings of a review of over two hundred studies of prison- and community-based rehabilitative services to offenders. The legacy of Martinson's article was the slogan "nothing works," which came to dominate mainstream penal thought for years to come and dealt a severe blow to faith in the "rehabilitative ideal" (Allen 1959; Garland 2001).

Ever since then, researchers (particularly in North America) have endeavored to demonstrate that "nothing works" was a significant overstatement, and to revive at least some optimism about the potential of penal interventions to rehabilitate offenders (e.g., Palmer 1975; Gendreau and Ross 1980). Efforts to reinstate a "what works?" agenda in the 1980s were accompanied by critiques of the old "treatment model" and the theoretical assumptions which lay behind it, alongside the development of new styles of intervention designed to confront and modify "offending behavior." The late 1980s and early 1990s saw the emergence and proliferation of "offending behavior programs" based upon a predominantly psychological (cognitive-behavioral) model of offending (e.g., Ross et al. 1988; Hollin and Palmer 2006). This period also saw the refinement of evaluative research methodologies, and the adoption (from the medical sciences) of the statistical technique of "meta-analysis." This technique enabled researchers to analyze the findings of large numbers of primary research studies and distil from them not only their impact on reoffending, but also what approaches and methods appeared to be associated with positive outcomes. Throughout the 1990s, summaries of these characteristics or "ingredients" of (apparently) effective rehabilitative interventions, derived from meta-analytic research, began to appear (e.g., Andrews et al. 1990a; McGuire and Priestley 1995) and in some jurisdictions were adopted as the basis for designing and delivering services. Canadian meta-analytic research published by Andrews and colleagues also underpinned a model of offender rehabilitation which came to be known as the "RNR" (risk-need-responsivity) model, which has subsequently had a huge influence on practice across a variety of jurisdictions and is widely regarded as the dominant approach to offender rehabilitation in the early twenty-first century (Andrews et al. 1990b).

England and Wales was one jurisdiction which saw significant investment in the development and evaluation of the so-called new rehabilitation under the auspices of its government's "Crime Reduction Programme," which sought to test the impact on reoffending of a range of programs and interventions for offenders subject to custodial and community sentences. However, the evaluative research commissioned failed to deliver clear-cut results: published studies revealed a catalogue of implementation problems; weaknesses in evaluation design; few completed reconviction studies; and results that were "mixed" and often difficult to interpret because of problems with evaluation methodologies. Far from establishing "what works," then, researchers were left pondering (once again) whether the results of their efforts were best interpreted as failures of theory, implementation, or evaluative research itself, and fearing the implications of a return to the pessimism of "nothing works" (see, e.g., Raynor and Robinson 2009, Chap. 6).

However, the picture is not an entirely bleak one. In a recent review of evidence about "what works" in offender treatment and rehabilitation, Lösel (2012) concludes that:

> various measures from the 'what works' literature are effective. Typical effect sizes range somewhere between 10 and 20 per cent less reoffending in treated than in control groups, but there are also larger or smaller effects. Desirable outcomes have been found for cognitive-behavioural programmes, therapeutic communities, social-therapeutic prisons, milieu therapy, multisystemic treatment, RJ measures, educational and vocational programmes. (Lösel 2012, p. 1006)

This conclusion is supplemented by the important observation that "what works" is rarely reducible to specific programs or interventions that will "work" in any context: there are a variety of "moderators" of program effects that include offender factors (e.g., risk level, motivation, demographic variables), treatment context (e.g., staff skills, continuity of support, institutional climate), and evaluation method (e.g., quality of evaluation design, sample size, length of follow-up period). These are all variables that can have a bearing on the findings of research. An important implication of this is that "what works?" is too crude a question: researchers instead ought to be asking what works, with whom, in what circumstances (etc.).

History has demonstrated, thus, that the apparently straightforward question – "does it work?" – is in fact an extremely complex one, which is difficult to operationalize and has failed to produce clear-cut answers. This is a perennial problem for proponents of rehabilitation who seek to persuade sentencers and policy makers of the potential of rehabilitative sanctions to cut crime, prevent victimization, and protect communities. In the absence of convincing evidence to support such claims, sentencers and policy makers are arguably going to be more inclined to resort to measures that can be shown, or are popularly believed, to impact on crime rates, such as lengthy periods of incapacitation. However, to the extent that research into the effectiveness of rehabilitative interventions continues, it is hoped that we will see more nuanced research designs that reflect some of the lessons of the past.

The Rehabilitation (R)Evolution

The so-called collapse of the rehabilitative ideal in the latter part of the twentieth century is well documented (e.g., Garland 2001). Beginning in the late 1960s and early 1970s, the dominant "treatment model" was subjected to a three-pronged critique which emphasized its theoretical and ethical flaws as well as its apparent ineffectiveness, such that both the legitimacy of rehabilitation as a primary penal objective, and faith in the possibility of rehabilitating offenders were subject to serious doubts. In the wake of rehabilitation's decline in official discourse and policy, analyses of penal transformation proliferated, and in many of these, rehabilitation was consigned to penal history. Whether proclaiming the arrival of a new "postmodern" penology or a "new punitiveness," international penal commentators tended to agree that rehabilitation was dead or at least on its last legs.

More recently, however, it has become clear that claims about the demise of rehabilitation have been overstated. Not only have penal professionals tended to retain at least some optimism about the transformative purposes of their practice, but as we have seen above, research in the "correctional" field has been relentlessly focused on developing new, more effective techniques and interventions and has far from given up the quest to discover "what works." As one of the scholars at the forefront of the "What Works" movement has explained:

> Three decades ago, it was widely believed by criminologists and policymakers that 'nothing works' to reform offenders and that 'rehabilitation is dead' as a guiding correctional philosophy. By contrast, today there is a vibrant movement to reaffirm rehabilitation and to implement programs based on the principles of effective practice. How did that happen? I contend that the saving of rehabilitation was a contingent reality that emerged due to the efforts of a small group of loosely coupled research criminologists. (Cullen 2005, p. 1)

Cullen is correct to point to the important role of the research community in the "reaffirmation" of rehabilitation, which has been evident in (for example) the state of California rebranding its Department of Corrections as the Department of Corrections and Rehabilitation under the leadership of a Republican governor, and in the UK, the Conservative-led coalition government promoting something it is calling a "Rehabilitation Revolution." However, there is arguably more to the story of rehabilitation's reaffirmation than is implied in Cullen's account: it has been argued that we should think of the recent history of rehabilitation in terms of an evolutionary process, whereby it has been successfully transformed and re-marketed in a late-modern context, such that far from going "against the grain" of other currents in penal policy, it has in fact been rendered compatible with them (Robinson 2008). Thus, for example, the "new" rehabilitation has had to adapt to social and political contexts that have become increasingly intolerant of approaches that appear to put the needs and/or interests of offenders above those of victims (actual or potential). In jurisdictions which have been subject to "populist punitiveness," proponents of rehabilitation have had to de-emphasize its humanitarian and welfarist justifications in favor of its utilitarian credentials (Garland 2001; Robinson 2008). Rehabilitation has also had to adapt and reposition itself within increasingly managerial correctional systems dominated by the discourse of "risk." In this process, it has been reframed as one of a number of possible "means" toward the "end" of risk management, as well as being increasingly rationed in line with assessments of risk which determine offenders' eligibility for "rehabilitative" sanctions or interventions.

The adaptation of rehabilitation in late-modern contexts has not however simply entailed putting a new "spin" on an old product (Robinson 2008). We have not witnessed a return to the "old" treatment model, but rather the proliferation of a new breed of "offending behavior programs" underpinned by cognitive-behavioral psychology. This "new orthodoxy" in offender rehabilitation is certainly attributable, in part, to its grounding in research and the promise of measurable reductions in reoffending on the part of program completers. But the dominance of cognitive-behavioral programs may also be attributable to their expressive and communicative qualities, and resonance with "advanced liberal" forms of governance and neoclassical criminologies which emphasize personal responsibility for wrongdoing and rely upon strategies of "responsibilization" as the dominant response to antisocial behavior (e.g., Kendall 2004). Rehabilitation has, thus, adapted and survived into the twenty-first century by transforming itself in important ways: central to its relegitimation in late-modern societies like the USA and the UK has been its successful appeal to three dominant penal narratives, which have been characterized as *utilitarian*, *managerial*, and *expressive* (Robinson 2008).

Challenging the "New Orthodoxy": Future Directions in Rehabilitation

As we have seen, the "What Works?" movement and its key protagonists have played a crucial role in the relegitimation of rehabilitation in the last two decades, and the lessons of that research have spread well beyond North America, such that the "risk-need-responsivity" (RNR) model (associated particularly with the work of Don Andrews and James Bonta) has become the "new orthodoxy" in offender rehabilitation in a host of jurisdictions. The RNR model has had a particularly significant impact on approaches to offender assessment, inspiring actuarial assessment instruments such as the Level of Service Inventory and the British Offender Assessment System, which emphasize the importance of assessing offenders' risks and (criminogenic) needs as a basis for decisions about their suitability for "rehabilitative" interventions. The RNR model has also been intimately associated with the proliferation of cognitive-behavioral offending behavior programs: indeed, the notion of "general responsivity" in the RNR formulation recommends cognitive-behavioral approaches as being the most effective in general terms.

However, both the RNR model and the specific technologies and approaches associated with it have been subject to criticism, from a variety of perspectives. Early critiques focused on the "reification" of "What Works?" (evidenced in a tendency to drop the question mark, thereby implying that the "answer" to the question had now been discovered) and the applicability of theories, assessment instruments, and/or interventions developed on the basis of research with young, white North American male offenders on different groups, such as women, ethnic minorities, and those with substantially different offending profiles. Concerns were also raised about the potentially "authoritarian" (Rotman 1990) characteristics of offending behavior programs, which were seen by some as a worrying return to the bad old days of the "treatment model," proceeding on the basis of predominantly psychological theories of crime causation, and reviving images of offenders as passive recipients of "interventions" delivered by "experts" (e.g., Kendall 2004). Several commentators argued that RNR (and particularly the assessment instruments associated with it) promulgated a *deficit model* which was overly focused on risks, needs, problems, and deficiencies and blind to their potential strengths, achievements, and "protective factors" (e.g., Ward and Maruna 2007).

Drawing on developments in what they call "positive psychology," Tony Ward and colleagues have argued for the development of "strengths-based" approaches to rehabilitation and they have proposed a "Good Lives Model" as a potential alternative to RNR. At the heart of the GLM is a positive perspective: it seeks to equip offenders with the capabilities to secure "primary human goods" (defined as valued aspects of human functioning and living) in socially acceptable and personally meaningful ways. The GLM rests on the assumption that interventions should aim to promote an individual's "goods" as well as to manage or reduce their risk. A major aim of rehabilitative work is to enable an individual to develop a life plan that involves ways of effectively securing primary human goods without harming others. The approach requires an explicit focus on conceptualizing a good life: taking account of strengths, primary goods, and relevant environments and encouraging and respecting individuals' capacities to make choices for themselves. To date, the GLM has been most extensively applied to rehabilitative work with sexual offenders, but its advocates propose that the model (and the theory on which it is based) is relevant for a wide range of offenders (Ward 2010).

Another alternative paradigm for offender rehabilitation builds on research on "naturalistic" processes of desistance – that is, what is known about how and why offenders cease offending. The "desistance paradigm" (e.g., Maruna and LeBel 2010; McNeill 2006) takes issue with the centrality of "interventions" in the RNR model and the notion that achieving desistance is simply a matter of dispensing the right program. This idea, its advocates argue, is at odds with the findings of desistance research, which points to desistance not as an event but a process, characterized by progress and setbacks; hope and despair.

In common with advocates of the GLM, proponents of the desistance paradigm argue that RNR has downplayed the role of offender motivation in the rehabilitation process, in favor of a "one-size-fits-all" model. McNeill has argued that the GLM may offer a promising start in addressing the weakness of the RNR model with respect to motivational issues. However, because both RNR and GLM are models developed by psychologists, both place too much emphasis on the individual and on psychological issues, while de-emphasizing the role of the offender's social context and the importance of developing opportunities and social capital for offenders. McNeill is also critical of the RNR model's underemphasis on the role of relationships in the offender's life (including those between the offender and the "rehabilitative" professionals with whom they engage), which also emerge from desistance research as potentially key in the desistance process.

McNeill is one of a number of commentators who have begun to draw attention to the importance of the relational dimension of, or context for, rehabilitation; that is, the idea that rehabilitation is not best conceived as something which happens to or is "done to" offenders in a social vacuum, but rather it is a process which takes place in the context of – and often seeks to build – relationships of various kinds. The idea of "relational rehabilitation" (Bazemore 1999; Robinson and Crow 2009, Chap. 9) encompasses a variety of "nontreatment" approaches that include restorative justice schemes which bring together offenders and their victims, as well as projects that utilize community members (such as mentoring schemes and "Circles of Support and Accountability" for high-risk sexual offenders). It also encompasses ideas about the potential of sentencers to act as therapeutic agents (often referred to as *Therapeutic Jurisprudence*) which have underpinned the development of specialist, problem-solving courts in the USA and elsewhere.

Finally, recent research and debate about "nontreatment" approaches, relational matters, and (particularly) the role of the law and its agents in processes of rehabilitation has prompted a revival of interest in the symbolic dimension of this concept, and the potential importance of "rehabilitation rituals" to mark the decertification of deviance (Maruna 2011). "Judicial rehabilitation" denotes all of those legal practices and tools which serve to expunge or delete the criminal record, or otherwise recognize or confirm the individual's status as a desister or ex-offender. As Maruna (2011) has noted, jurisdictions differ significantly in their willingness to adopt such mechanisms, and in the extent to which they enable ex-offenders to navigate away from stigmatic labels. In European countries such as France, which traditionally place a high cultural value on privacy, judicial rehabilitation tends to be most developed (see Herzog-Evans 2011). By contrast, in the USA there is a long history of freedom of information (including "naming and shaming" that dates back to the Salem witch trials) which at least partly explains why ex-offenders are particularly disadvantaged when it comes to the protection of their privacy and thus their ability to "live down" their convictions. However, Maruna fears a general trend toward "Americanization" in Europe and elsewhere, as is becoming evident in the practice of making criminal records more widely available, and in community notification schemes with respect to sexual offenders, which tend to prioritize "public protection" over the ex-offender's right to put their past behind them.

Related Entries

- ▶ Desistance and Supervision
- ▶ Desistance from Crime
- ▶ Good Lives Model
- ▶ History of Corrections
- ▶ Offender Change in Treatment
- ▶ Penal Philosophy and Sentencing Theory
- ▶ Punishment as Rehabilitation

Recommended Reading and References

Allen FA (1959) Criminal justice, legal values and the rehabilitative ideal. J Crim Law Criminol Police Sci 50:226–232

American Friends Service Committee (1971) Struggle for justice: a report on crime and punishment in America. Hill and Wang, New York

Andrews DA, Zinger I, Hoge RD, Bonta J, Gendreau P, Cullen FT (1990a) Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. Criminology 28(3):369–404

Andrews DA, Bonta J, Hoge RD (1990b) Classification for effective rehabilitation: rediscovering psychology. Crim Just Behav 17:19–52

Bazemore G (1999) After shaming, whither reintegration: restorative justice and relational rehabilitation. In: Bazemore G, Walgrave L (eds) Restorative juvenile justice: repairing the harm of youth crime. Criminal Justice Press, Monsey

Braithwaite J (1989) Crime, shame and reintegration. Cambridge University Press, Cambridge

Burgess A (1962) A clockwork orange. Heinemann, London

Carlen P (1989) Crime, inequality and sentencing. In: Carlen P, Cook D (eds) Paying for crime. Open University Press, Milton Keynes

Cavadino M, Dignan J (2007) The penal system: an Introduction, 4th edn. Sage, London

Cullen F (2005) The twelve people who saved rehabilitation: how the science of criminology made a difference. Criminology 43(1):1–42

Cullen FT, Gilbert KE (1982) Reaffirming rehabilitation. Anderson Publishing, Cincinnati

Day A, Tucker K, Howells K (2004) Coerced offender rehabilitation – a defensible practice? Psychol Crime Law 10(3):259–269

Garland D (1985) Punishment and welfare. Gower, Aldershot

Garland D (2001) The culture of control. Oxford University Press, Oxford

Gendreau P, Ross R (1980) Effective correctional treatment: bibliotherapy for cynics. In: Ross R, Gendreau P (eds) Effective correctional treatment. Butterworths, Toronto

Herzog-Evans M (2011) Judicial rehabilitation in France. Eur J Probat 3(1):4–19

Hollin C, Palmer E (eds) (2006) Offending behaviour programmes: development, application, and controversies. Wiley, Chichester

Howard J (1929) [1777] The state of the prisons. J.M. Dent and Sons, London

Kendall K (2004) Dangerous thinking: a critical history of correctional cognitive behaviouralism. In: Mair G (ed) What matters in probation. Willan, Cullompton

Lösel F (2012) Offender treatment and rehabilitation: what works? In: Maguire M, Morgan R, Reiner R (eds) Oxford handbook of criminology, 5th edn. Oxford University Press, Oxford

Martinson R (1974) What works? Questions and answers about prison reform. Public Interest 35:22–54

Maruna S (2011) Judicial rehabilitation and the ‘clean bill of health’ in criminal justice. Eur J Probat 3(1):97–117

Maruna S, LeBel T (2010) The desistance paradigm in correctional practice: from programmes to lives. In: McNeill F, Raynor P, Trotter C (eds) Offender supervision: new directions in theory, research and practice. Willan, Oxford

McGuire J, Priestley P (1995) Reviewing ‘what works’: past, present and future. In: McGuire J (ed) What works: reducing reoffending. Wiley, Chichester

McNeill F (2006) A desistance paradigm for offender management. Criminol Criml Just 6(1):39–62

Palmer T (1975) Martinson revisited. J Res Crime Delinq 12:133–152

Raynor P (1997) Some observations on rehabilitation and justice. Howard J Crim Just 36(3):248–262

Raynor P, Robinson G (2009) Rehabilitation, crime and justice. Palgrave Macmillan, Basingstoke

Robinson G (2008) Late-modern rehabilitation: the evolution of a penal strategy. Punishm Soc 10(4): 429–445

Robinson G, Crow I (2009) Offender rehabilitation: theory, research and practice. Sage, London

Ross RR, Fabiano EA, Elwes CD (1988) Reasoning and rehabilitation. Int J Offender Ther Comp Criminol 32(1):29–35

Rotman E (1990) Beyond punishment: a new view of the rehabilitation of offenders. Greenwood Press, Westport

van Zyl Smit D, Snacken S (2009) Principles of European prison law and policy. Oxford University Press, Oxford

Ward T (2010) The good lives model of offender rehabilitation: basic assumptions, aetiological commitments, and practice implications. In: McNeill F, Raynor P, Trotter C (eds) Offender supervision: new directions in theory, research and practice. Willan, Oxford

Ward T, Maruna S (2007) Rehabilitation. Routledge, London

Rehabilitation, Treatment

► Juvenile Diversion

Re-integration

► Rational Choice, Deterrence, and Crime: Sociological Contributions

Release

► Drug Abuse and Alcohol Dependence Among Inmates

► Front Door and Backdoor Sentencing

Reliability

▶ Identification Technologies in Policing and Proof
▶ Self-Reported Offending: Reliability and Validity

Reoccurrence of Victimization

▶ Repeat Victimization

Repeat Offending

▶ Prediction and Crime Clusters

Repeat Victimization

Graham Farrell[1] and Ken Pease[2,3]
[1]Institute for Canadian Urban Research Studies and School of Criminology, Simon Fraser University, Burnaby, BC, Canada
[2]Jill Dando Institute, University College London, London, UK
[3]Loughborough University, Loughborough, UK

Synonyms

Chronic victimization; Frequent victimization; High-frequency victimization; Multiple victimization; Near repeats; Near-repeat victimization; Polyvictimization; Prior victimization; Recidivist victimization; Recurrent victimization; Reoccurrence of victimization; Repetitive victimization; Revictimization

Overview

This entry hinges on four nonobvious propositions that, we suggest, largely encapsulate the state of the art in this field. They are that:

1. Most crime is repeat victimization against the same targets, and variation in crime across time and place is often due largely to repeat victimization.
2. One or 2 % of potential targets are the supertargets that experience about half of all crimes.
3. Preventing repeat victimization reduces crime when tactics are properly developed and implemented.
4. A broader-than-usual conception of *near-repeat victimization*, referring to crimes linked by their similarity, may prove a useful foundation for integrating theories of crime concentration.

The evidence and argument underpinning these statements is presented in what follows. References and some further reading are listed at the end, but readers are encouraged to browse the *Oxford Bibliographies Online* (*Criminology*) entry on *repeat victimization* which summarizes around 150 key studies. It documents the growth of studies from the 1970s onwards, the extent and nature of repeats and near repeats in many different countries, theoretical developments, the link to repeat offending, measurement issues, time-course and life-course studies, the link to policy and practice, and crime-specific prevention studies. The study of repeat and near-repeat victimization has become a mainstay of criminology, crime science, and police science in the last two decades, and its contribution was recently assessed as a key area of theory underpinning environmental criminology and crime analysis (Wartell 2010; Wartell and Gallagher 2012).

Definitions and Concepts

People, places, products or services, vehicle, household, business, or systems (including virtual networks) are, once victimized, more likely to be victimized again. Further victimization by the same type of crime is most likely, but polyvictimization involving two or more crime types is also far more likely than would be expected by random chance. One has only to think of hate crime for this to be recognized.

A small proportion of targets are so chronically victimized that they become supertargets. One classic study found that 1 % of people experienced 59 % of all personal crime including violence and that 2 % of households experienced 41 % of all household crime. The supertargets are the 1–2 % of persons and households, respectively, and they represented around one in eight of all persons or households who were victimized (Pease 1998).

Key features of repeats are that they tend to occur sooner rather than later (Daigle et al. 2008; Mele 2009) and are more likely to be by the same prolific offenders, with success breeding repeats. The result is a temporarily elevated risk, greater with each subsequent crime. Similar general patterns of the extent and nature of repeats have been identified for most crime types and places where informed studies have been conducted, though Sidebottom's (2011) study of repeat burglary in Malawi observes that "[t]here is little research on repeat victimization in developing countries" (p. 1).

A growth area for repeat victimization studies has been the area of *victim careers* or *life-course studies* (Gabor and Mata 2004; Buzawa et al. 2006; Averdijk 2010). Another is that of *near-repeat victimization* or *near repeats*, a term coined by Frank Morgan in a study of burglary in Perth, Western Australia (2001). In this entry, the terms repeat victimization or repeats are sometimes used as inclusive of near-repeat victimization.

Spatially, it is well established that crime hot spots are often substantially, and sometimes exclusively, repeats and near repeats. A few hard-hit stores, for example, may account for peaks in crime maps, but this can be obscured by kernel density estimation techniques. Hence, repeat research has been lauded for providing more precise predictions and insight into a real crime patterns than some traditional hot-spot studies. And what is now sometimes termed predictive policing draws on the tradition of the revictimization predictor (Farrell 1995). The significant general expansion of repeat victimization research in recent years is illustrated by a recent study of personal repeat victimization, of potentially major importance, which suggests there is a genetic aspect to the phenomenon (Daigle 2010).

The following four sections cover the propositions that began this entry. They are followed by a short conclusion.

Most Crime is Repeat Victimization Against the Same Targets, and Variation in Crime Over Time and Space is Often Due Largely to Repeat Victimization

In the United States, 75 % of personal crimes over the course of a year are repeats according to the National Crime Victimization Survey, as shown in Table 1. In England and Wales, about half of property crime and two-thirds of personal crimes are repeats (see details below). The Canadian Victimization Survey found that 2 % of the population who were repeat victims of violence experienced 60 % of all violence (Perreault et al. 2010, and Table 2 which is based the work of Nazaretian and Merolla 2013). This sounds like a lot – and it is – but these are conservative estimates. There are various methodological reasons for this. The surveys tend to exclude crimes such as child abuse and victimization of young people or nonhuman animals which have even higher repeat rates. Likewise, robberies, burglaries, and other crimes against commercial establishments (not covered in conventional victimization surveys) tend to have higher repeat rates than equivalent crimes against individuals and households. Because surveys take a time window of 6 months or a year, events within the window which are repeats of pre-window events, or which are precursors of post-window events, are repeats masquerading as one-off crimes!

Repeat rates tend to be lower for property crimes than violent crimes, but, as noted above, repeat property crime rates are disproportionately higher in hot spots and high crime areas, so the potential for focused prevention efforts is greater. Overall, however, the evidence suggests that most crimes are experienced by repeat and near-repeatedly victimized persons, places, and other targets.

Repeat Victimization, Table 1 Personal crimes in the United States

Crime type	Prevalence per 100	Incidence per 100	% Repeats
Rape/sex assault	0.4	2.3	82.6
Robbery	1.3	3.6	63.9
Assault	8.5	37.4	77.3
Personal larceny	0.5	1.2	58.3
Total	10.6	44.6	76.2

Adapted from Planty and Strom (2007, p. 199)

Repeat Victimization, Table 2 Violent crime in Canada

Crime type	Prevalence per 100	Incidence per 100	% Repeats
Sexual assault	1.4	4.2	66.1
Assault	4.0	9.7	58.9
Robbery	0.9	1.3	31.0

Adapted from Nazaretian and Merolla (2013)

Much of the crime trend in England and Wales since 1980 – both increases and decreases – is linked to change in repeats. This is shown in Fig. 1 for household and property crimes measured by the British Crime Survey and draws on the excellent work of Britton et al. (2012). In the United States, a recent study published by the Bureau of Justice Statistics suggests a similar pattern, with a lot of the crime drop being due to declines in repeat victimization (Lauritsen et al. 2012).

The bulk of variation in violence over time in England and Wales was experienced by repeat violence victims (Fig. 2). Such incidents accounted for half to two-thirds of violence, and 68 % in 1995. Repeats also played a disproportionate role in burglary trends. Repeat victims of burglary experienced around a quarter of burglaries in 1981, rising to 40 % in 1995 (Fig. 3). However, while ostensibly one-off incidents doubled between 1981 and 1995, those at repeatedly victimized households more than trebled (a 330 % increase). Between 1995 and 2010, both singles and repeats fell back to their 1981 level.

Repeated vehicle-related theft accounted for 40 % of such thefts in 1981, rising to half (49 %) by 1995. However, it accounted for less than one in three by 2010. That is, singles doubled while crimes experienced by repeat victims of car theft trebled from 1981 to 1995, after which repeats declined more rapidly than singles (Fig. 4).

The fall in repeats from 1995 was always greater than the overall fall in any particular type of crime (Fig. 5). This does not necessarily show that declining repeats drove the crime drop, but suggests explanations should certainly consider the role of repeats. Hence, aggregate crime rates seem to go up and down in large part due to variation in repeat victimization. This is consistent with crime increasing or decreasing as the same offenders victimize the same targets more or less often. Note also that these BCS statistics are quite conservative when it comes to the role of repeats because the Britton et al. (2012) analysis, brilliant as it is, stopped counting if a victim experienced more than five similar crimes! In summary, however, most crimes are repeats or nears, and they often underlie spatial, temporal, and other variations in crime.

One or 2 % of Potential Targets are Supertargets that Experience Around Half of Crime

Just as wealth is hugely, perhaps grotesquely, unevenly distributed in society, so too is victimization. Most repeat victims experience two, three, or four crimes in a year, but a few experience many more. The Gini coefficient, which is often used to measure wealth inequality, has also been applied to crime, along with the visually compelling Lorenz curves (Tseloni and Pease 2005). This section outlines some explanations for the existence of such supertargets.

Since a lot of crime is not reported to the police, and reporting rates likely diminish as the number of victimizations increases, a good tactic for police is to ask victims about their recent previous experiences with crime. By this simple means, victims, or the owners or managers of properties or other places where crime clusters,

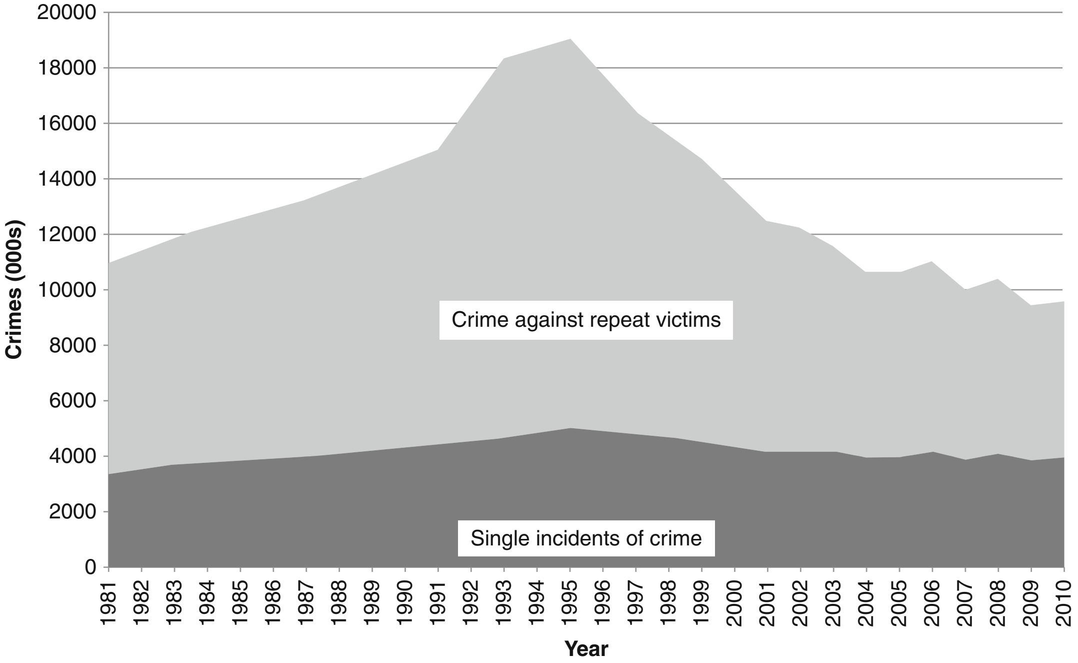

Repeat Victimization, Fig. 1 Role of repeat victimization in all crime 1981–2010

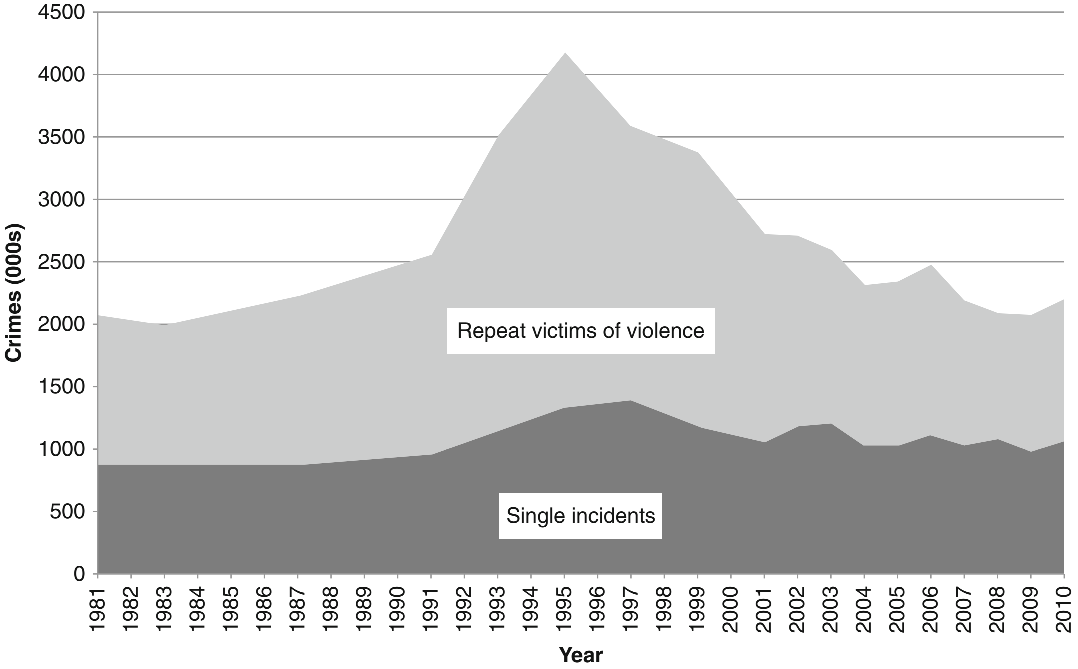

Repeat Victimization, Fig. 2 The role of repeat victimization in violence trends 1981–2010

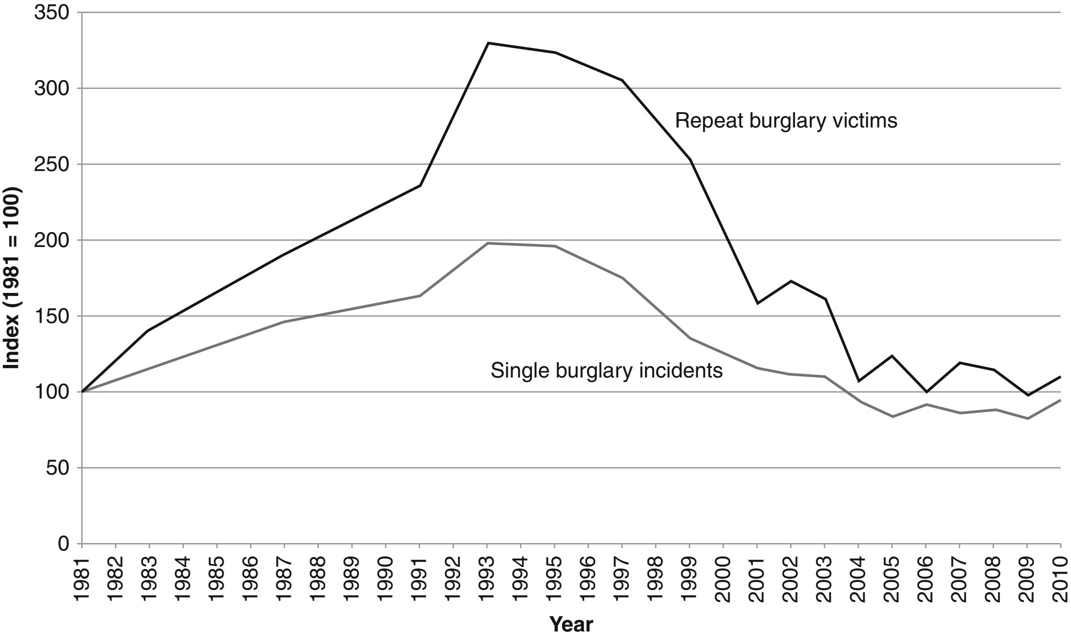

Repeat Victimization, Fig. 3 Single and repeat burglary 1981–2010

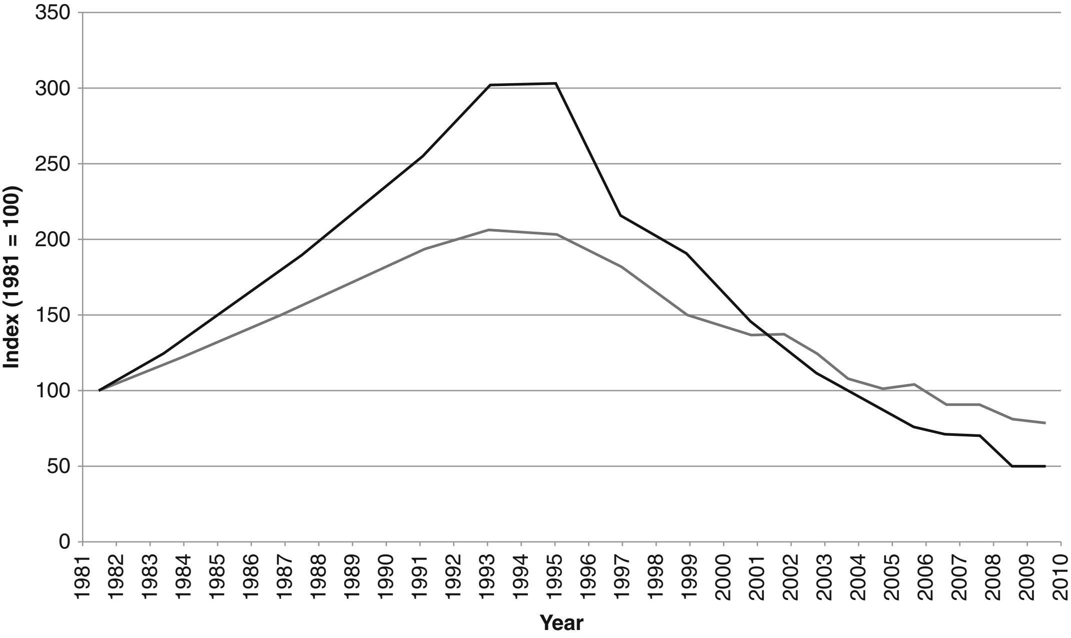

Repeat Victimization, Fig. 4 Single and repeat motor vehicle theft 1981–2010

Repeat Victimization, Fig. 5 The crime drop 1995–2010

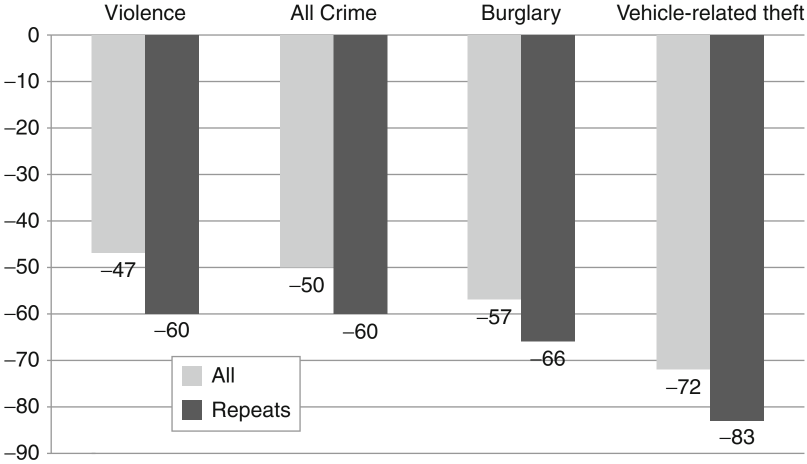

can reveal patterns of repeat crimes that will not show up in traditional analysis of police records. This is a good means of gathering information about actual crime patterns that can then inform the response that is given by the police, victim services, and others. More specifically, the amount of resources spent on a response can then be tailored to the risk, with greater risk receiving proportional response. This type of model has been termed a graded response or an Olympic response with bronze, silver, and gold responses delivered, depending on risk. Police using this approach should be prepared for an increase in reported crime, as chronic victims become persuaded that such a report has become worthwhile.

There is less skepticism about the existence of supertargets when considering crime types for which high-frequency victimization is well proven. Family violence or incest provides one example, which can produce chronic repeat victims (survivors) of domestic violence, child abuse, and elder abuse. The proximity of victim and offender, perhaps living together and in the same family with emotional or economic reasons for not reporting crime, facilitates unconstrained repetition of the crime. A range of research has established that abuse can occur on a regular and frequent basis. More broadly, those with power or responsibility sometimes exploit their position to repeatedly victimize those in their care.

Crime against businesses is another area where supertargets have been identified. Some types of store represent more attractive targets than others. These include electronic goods stores, pharmacies, convenience stores, and some supermarkets. When such stores are located in high-traffic malls, their crime count rises exponentially. The same commercial establishments are the repeat targets of many shop thieves and the same returners, for example. Particular stores and commercial premises that are perceived as lucrative easy targets are frequently hit by burglars and robbers.

Disputes between neighbors or families in a neighborhood can produce high-frequency victimization. Groups or gangs may become involved in recurring targeting of the same household. A targeted household may experience many crimes of different sorts, including burglary, vandalism of property and vehicles, theft of and from vehicles, and threats and violence of various sorts against residents. Considering such contexts, the existence of chronic victims and polyvictimization by different types of crime should not be surprising.

Stalking and harassment are largely defined by their repetitive nature. Sexual victimization can also manifest itself as different types of crime (assaults, vandalism, cyber-stalking, threats) perhaps by an ex-partner or during a cohabiting or other relationship. Hate crimes, including racial attacks, homophobic crimes, and crimes against the disabled, may result in chronic victimization. A hate crime victim may be identifiable by their appearance, which could prompt victimization by different as well as the same offenders, more so if the location of the victim's household or workplace is known. In schools, the bullying of some victims is frequent and ongoing. A victim may attract victimization from other sources if he or she acquires a reputation as a suitable target.

Some lifestyles appear related to chronic victimization. Some professions, including police, nurses, and other caring professions, are more likely to experience crime. The work can entail regular interaction with likely offenders in risky situations. One study noted that "violent series crimes that occurred in connection with a person's job were usually committed by different offenders who were generally strangers. In contrast, violent crimes committed by spouses or acquaintances were usually perpetrated by the same person who was well known to the victim" (Dodge and Balog 1987).

Repeat offenders are more likely to experience repeat victimization (Fagan and Mazerolle 2008). This may be the result of a lifestyle which includes drug dealing, gang-related violence, ongoing disputes between rivals, or frequent interaction with criminal associates. Offenders are robbed while dealing drugs, fencing stolen goods, because they carry a stash of money and in disputes with co-offenders and rivals. Since offenders are less likely to call the police, their repeat victimization occurs with relative impunity.

A supertarget may emerge due to victimization by different offenders – a *risk-heterogeneity* explanation. High-frequency shoplifting at some stores falls into this category. In many instances, however, one crime boosts the likelihood of another occurring when offenders learn it is an easy and rewarding crime to commit – an *event dependence* explanation. The extreme case is where an offender is "dedicated" to a single target with which they frequently interact and know the likely risk and reward – as may be the situation in family violence, harassment, and stalking. The balance between the two theoretical mechanisms will vary according to the circumstances. The question then becomes why supertargets are under-researched.

Why Are Supertargets Often Overlooked?

It is probably difficult for some people to believe the extent of high-frequency victimization. Or perhaps it is just easier to overlook. Chronic victims are likely those least able to voice their complaint. The NCVS had been running for around 40 years before the Bureau of Justice Statistics began to properly acknowledge the importance of high-frequency victimization in the important study by Janet Lauritsen and colleagues (2012). And even that excellent study used the median frequency of victimization as an upper ceiling on crime counts for all victims. The British Crime Survey was similarly tardy in recognition, and having the issue drawn to the attention of the relevant Ministry excited criticism rather than recognition of the problem.

It is well documented that repeat victimization is generally understated in police data. When it comes to supertargets, this is more extreme, a conservative estimate being a twentyfold undercount (Frank et al. 2012). This is due to conditional probabilities. If the likelihood of a burglary reported to police is 50 %, then the likelihood of two at the same household being reported are 25 % (i.e., .5 ×.5 =.25). It follows that the likelihood that five burglaries at a household are all reported is .5 × .5 × .5 × .5 × .5 = .03125 or 3 %. This is a simplified version, but the basic argument is robust when the initial reporting likelihood and conditional likelihood of reporting of subsequent events are varied.

For crime surveys, the situation is different. If a poly-victim reports assault, burglary, robbery, car theft, and vandalism to a survey, each offence type is typically recorded as a separate, seemingly unrelated, incident or series of incidents. The British Crime Survey does not count as repeat victims those suffering repeated crime where each offense type occurs only once. Likewise, much chronic victimization will not be brought to the attention of the police, so supertargets will be less evident in recorded crime statistics.

Unfortunately, chronic victimization is sometimes swept under the carpet. It is sometimes cut from official crime because it can be “statistically awkward” to handle (Farrell and Pease 2007; Planty and Strom 2007). Crucially, when politicians use victimization surveys to set alongside police recorded crime statistics, to see whether the trends match, allowing chronic victims to be properly counted introduces a degree of sampling variability that is statistically inconvenient. In this way, statistical convenience triumphs over proper accounting of human suffering. Victim advocates may interpret this as institutional discrimination, and the impact is not trivial: If society’s resources in response to crime were apportioned according to risk, then the 2 % of supertargets would likely receive around half.

Preventing Repeat Victimization Reduces Crime When Measures Are Properly Developed and Implemented

Crime has been reduced around 20 % on average, and by greater amounts at victimized households, via targeted prevention efforts. This is the finding of a recent systematic review and meta-analysis of 31 evaluations (Grove et al. 2012; Grove and Farrell 2012). Most of the evaluations were of efforts to prevent residential burglary of which the most successful have two key characteristics:

- Measures that are locally appropriate and prevent repeats by the same *modus operandi*
- Measures that are thoroughly implemented

This makes crime prevention sound simple. Conceptually it is, but it is not always, easy to identify which prevention measures to introduce, to implement them, and to ensure they are used appropriately.

Ensuring that victimized households receive appropriate advice and then implement measures can be difficult. Even ensuring that police officers conduct security surveys at victimized properties has proven tricky in some instances.

The most effective measures tend to be situational. These often physically block opportunities for crime. The obvious tactics include locks and bolts in the case of burglary, but it can also include a range of other measures.

It is clear that there is plenty of scope for research into the prevention of other forms of repeat crime. Efforts to prevent repeat forms of commercial crimes, identity theft, and computer crimes, for example, appear under-developed, even though these may be areas where situational measures could be relatively easily introduced.

The percentage net change in burglaries for each of the 24 projects included in the systematic review is shown as Fig. 6. The percentage change in repeat burglaries is shown in Fig. 7 for the eight projects where it could be determined.

A Broader-than-Usual Conception of *Near-Repeat Victimization*, Referring to Crimes Linked by Their Similarity, May Prove a Useful Foundation for Integrating Theories of Crime Concentration

Crimes that are linked to previous crimes but not necessarily against the same target are known as *near repeats*. Spatial nearness is easily conceived, so near-repeat burglaries are of households close to one already victimized, usually fairly soon afterwards. Similar patterns have been found for other crimes (Ratcliffe and Rengert 2008; Wells et al. 2008; Youtsin et al. 2011).

Here, *nearness* refers not only to location and timing but to other characteristics of a crime. Thefts of hot products such as a smartphones often produce near repeats defined by the identical characteristics of the product as well as the similarity of the situation and crime event. The link between repeats and near repeats is the *similarity* of the targets. Similarity is attractive to offenders as the task is easier when the method, risks, and rewards are known. Similarity (or difference) may occur in terms of characteristics of the target, or the time, the location, the offender, the modus operandi, or the crime type.

The dominant theory of hot-spot formation is predicated on the concepts of crime attractors and generators (Brantingham and Brantingham 1995; Kinney et al. 2008). A place with high frequency

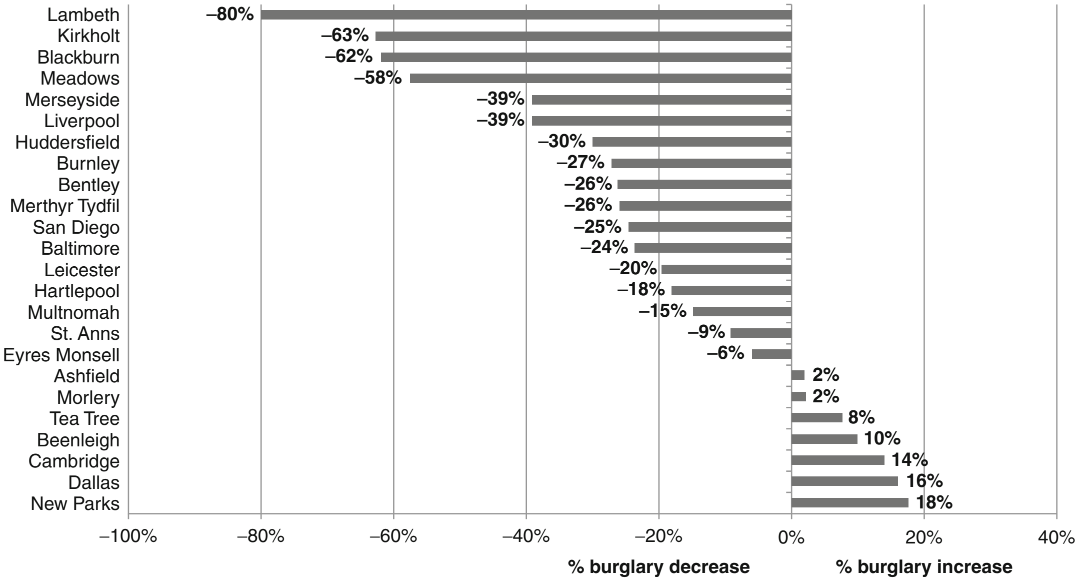

Repeat Victimization, Fig. 6 Percent change in burglaries by project

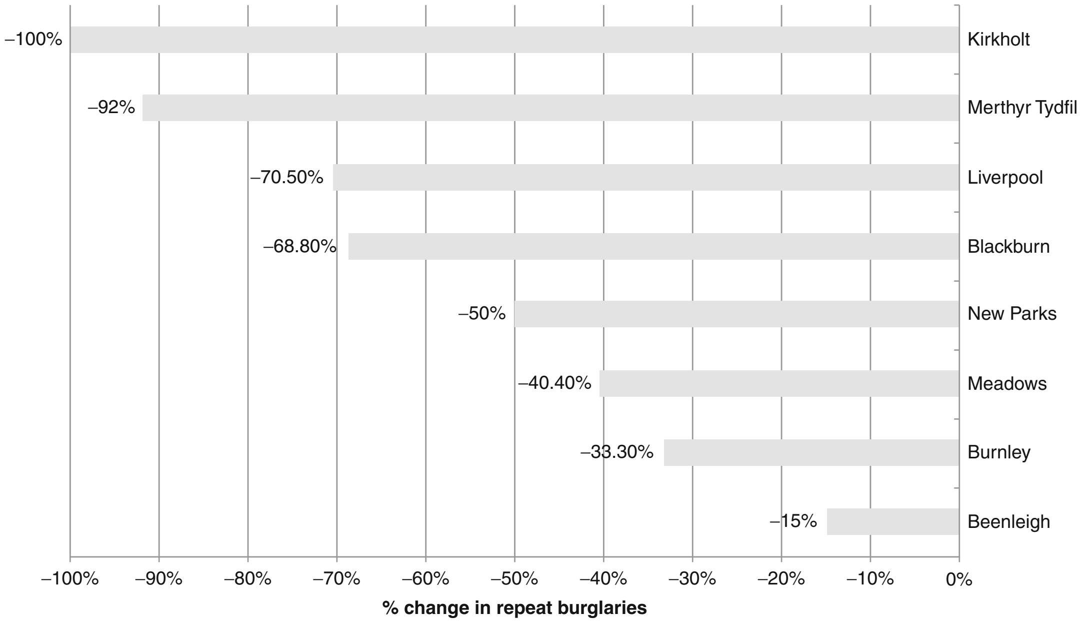

Repeat Victimization, Fig. 7 Change in repeat burglaries by project

of interaction between suitable targets and likely offenders *generates* crime. It may be a train station, a shopping mall, an entertainment district, or another activity node with a high flow of people. That is, even if per capita crime is no higher than elsewhere, a spatial concentration of crime still results. Think of a desert versus an entertainment nightspot. However, certain places then gain a reputation as locations with many suitable targets and *attract* offenders as a result.

The dominant theories of repeat victimization also embody two main components. The first is uneven risk (risk heterogeneity), such that some targets *flag* themselves as better than others. The second is greater ensuing risk (event dependence) wherein offenders learn that a target is worth revisiting, which *boosts* the likelihood they return. They have colloquially become known as flag and boost theories.

The existing theory of supertargets also largely explains why hot spots contain disproportionate repeat victimization. It can be viewed as combining a generator and a boost effect. An area may have high crime because of a generator effect wherein risk heterogeneity produces crime against particular targets. Each time a target in that area is victimized, its risk of revictimization is boosted. That risk increases with each victimization, so that a few targets progress to become supertargets. Other offenders may in turn be attracted. Incorporating near repeats into this model add to its predictive accuracy.

These key concepts – flags, generators, boosters, attractors, and near-repeat difference – may provide the foundation for a unified theory of crime concentration. This, in turn, may provide new impetus and understanding for crime prevention efforts targeting clusters of crime.

Conclusion

In the four decades since frequent returners to the emergency room of a Texan hospital were observed by Johnson et al. (1973), repeat and near-repeat victimization have emerged as a cornerstone of theory and practice in the study of crime. Repeats may often underpin, and perhaps cause, aggregate crime patterns and trends. The suggestion that the major crime drops of the last two decades may be related to reductions in repeat victimization is important when considering the precise mechanisms of change.

Yet while the importance of repeats is established, the intense concentration of crime on a few supertargets is an emerging topic with tremendous potential significance. Directing resources to prevent crime against supertargets would be particularly efficient.

Evidence from a range of studies suggests that it is possible to prevent repeat victimization and that doing so reduces crime. Situational crime prevention tactics tailored to the problem and well implemented seem to work best.

Despite a growing knowledge base, there remains much unexploited potential in repeat victimization research. This should include theory and theory testing, empirical studies, and far greater prevention effort, focused on different crime types.

Related Entries

- ▶ Multiple Victims and Super Targets
- ▶ Prediction and Crime Clusters

Recommended Reading and References

Averdijk M (2010) Individuals victimization patterns over time. Haveka, The Netherlands

Brantingham PL, Brantingham PL (1995) Criminality of place: crime generators and crime attractors. Eur J Crim Policy Res 3:5–26

Britton A, Kershaw C, Osborne S, Smith K (2012) Underlying patterns within the England and Wales crime drop. In: van Dijk J, Tseloni A, Farrell G (eds) The international crime drop: new directions in research. Palgrave, London

Buzawa E, Hoteling GT, Byrne J (2006) Understanding the impact of prior abuse and prior victimization on the decision to forego criminal justice assistance in domestic violence incidents: a life-course perspective. Brief Treat Crisis Interv 7(1):55–76

Daigle LE (2010) Risk heterogeneity and recurrent violent victimization: the role of DRD4. Biodemography Soc Biol 56:137–149

Daigle LE, Fisher BS, Cullen FT (2008) The violent and sexual victimization of college women: is repeat victimization a problem? J Interpers Violence 23(9):1296–1313

Davis RC, Taylor BG, Titus RM (1997) Victims as agents: implications for victim services and crime prevention. In: Davis RC, Lurigio AJ, Skogan WG (eds) Victims of crime, 2nd edn. Sage, Thousand Oaks

Dodge RW, Balog FD (1987) Series crimes: report of a field test. Bureau of Justice Statistics, Washington, DC

Fagan AA, Mazerolle P (2008) Repeat offending and repeat victimization: assessing similarities and differences in psychosocial risk factors. Crime Delinq 57(5):732–755

Farrell G (1995) Progress and prospects in the prevention of repeat victimisation. In Tilley N, (ed) The Handbook of Crime Prevention and Community Safety. Cullompton, Devon: Willan Publishing, pp 145–172

Farrell G, Pease K (2007) The sting in the British crime survey tail: multiple victimizations. In: Maxfield M, Hough M (eds) Surveying crime in the twenty first century, vol 22, Crime prevention studies. Willan, Cullompton, pp 33–54
Frank R, Brantingham PL, Farrell G (2012) Estimating the true rate of repeat victimization from police recorded crime data: a study of burglary in Metro Vancouver. Can J Criminol Crim Justice 54(4):481–494
Gabor T, Mata F (2004) Victimization and repeat victimization over the life-span: a predictive study and implications for policy. Int Rev Victimol 10:193–221
Grove LE, Farrell G (2011) Repeat victimization. Oxford Bibliographies Online (Criminology). Oxford University Press, New York
Grove LE, Farrell G (2012) Once bitten, twice shy: preventing repeat victimization. In: Welsh BC, Farrington DP (eds) Oxford handbook of crime prevention. Oxford University Press, New York
Grove L, Farrell G, Farrington DP, Johnson SD (2012) Preventing repeat victimization: a systematic review. Swedish National Council for Crime Prevention, Stockholm. www.bra.se
Johnson JH, Kerper HB, Hayes DD, Killenger GG (1973) The recidivist victim: a descriptive study. Criminal justice monograph, vol IV, no 1. Institute of Contemporary Corrections and the Behavioral Sciences, Sam Houston State University, Huntsville
Kinney JB, Brantingham PL, Wuschke K, Kirk MG, Brantingha PJ (2008) Crime attractors, generators and detractors: land use and urban crime opportunities. Built Environ 34(1):62–74
Lauritsen J, Gatewood Owens J, Planty M, Rand MR, Truman JL (2012) Methods for counting high frequency repeat victimizations in the national crime victimization survey. Bureau of Justice Statistics, Washington, DC
Mele M (2009) The time course of repeat intimate partner violence. J Fam Violence 24:619–624
Morgan F (2001) Repeat burglary in a Perth suburb: indictor of short- term or long-term risk? In: Farrell G, Pease K (eds) Repeat victimization. Criminal Justice Press, Monsey, pp 83–118
Nazaretian Z, Merolla DM (2013) Questioning Canadian criminal incidence rates: a re-analysis of the 2004 Canadian victimization survey. Can J Criminol Crim Justice, pp 1–23. DOI:10.3138/cjccj.2012.E18
Pease K (1998) Repeat victimisation: taking stock, Crime detection and prevention series paper 90. Home Office, London
Perreault S, Suavé J, Burns M (2010) Multiple victimization in Canada, 2004. Statistics Canada, Ottawa
Planty M, Strom KJ (2007) Understanding the role of repeat victims in the production of annual victimization rates. J Quant Criminol 23(3):179–200
Ratcliffe JH, Rengert GF (2008) Near repeat patterns in Philadelphia shootings. Secur J 21:58–76
Sidebottom A (forthcoming) Repeat burglary victimization in Malawi and the influence of housing type and area-level affluence. Secur J 1–17. doi: 10.1057/sj.2011.22
Tseloni A, Pease K (2005) Population inequality: the case of repeat victimization. Int Rev Victimol 12:75–90
Wartell J (2010) Environmental criminology: translating theory into practice. In: Presentation to the international symposium on environmental criminology and crime analysis, Brisbane, Australia, July
Wartell J, Gallagher K (2012) Translating environmental criminology theory into crime analysis practice. Policing: J Policy Pract. doi:10.1093/police/pas020
Wells W, Wu L, Ye X (2008) Patterns of near-repeat gun assaults in Houston. J Res Crime Delinq 49(2): 186–212
Youtsin TJ, Nobles MR, Ward JT, Cook CL (2011) Assessing the generalizability of the near repeat phenomenon. Crim Justice Behav 38(10): 1042–1063

Repeats

► Prediction and Crime Clusters

Repetitive Victimization

► Repeat Victimization

Reputation Sanctions

► Naming and Shaming of Corporate Offenders

Residential Burglary

Wim Bernasco
Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam,
The Netherlands

Synonyms

Break-in; Breaking and entering; Domestic burglary; Home burglary

Overview

This entry reviews what is known about residential burglary. It discusses what burglary is according to legal definitions and survey questions, it enumerates sources of knowledge about residential burglary, and it presents findings about its incidence and about which burglaries are most likely to be reported to police. The entry also describes the characteristics of the neighborhoods, individual dwellings, victims, and items that are the targets of burglary and of the offenders who commit burglaries. Further, the temporal distribution of burglary and patterns of repeat burglary victimization are discussed. The entry ends with a section on the prevention of burglary.

Introduction

A residential burglary is committed when someone enters an inhabited dwelling without permission and with the intent to steal. Although during the past two decades burglary rates have gone down worldwide, burglary is still a common offense. Most people will become a burglary victim during their lifetime. Burglary is listed as a property offense, but for most victims, the illegal entry into their homes has greater impact than the material loss they suffer. Most victims report to the police. The available knowledge about burglary is derived from the police, from victims, and from offenders. Burglars offend in disadvantaged neighborhoods with poor social control. They prefer dwellings that are unoccupied and easily accessible, and they prefer small, valuable, and disposable items (like cash, electronic gadgets, and jewelry). Most burglars are young males who steal to sustain an expensive lifestyle, often including drug use. Juvenile burglars are more likely to co-offend than adult burglars. Many burglaries are repeat burglaries of the same or a nearby address. There is some evidence that situational crime prevention measures are effective in reducing burglary.

Definitions

Residential burglary is the actual or attempted illegal entry into a dwelling with the intent to steal. Definitions vary among law enforcement jurisdictions and among victimization surveys with respect to a variety of details, including whether the entry requires force or destruction and what constitutes a dwelling.

In the United States, the Uniform Crime Reporting (UCR) program defines burglary as the unlawful entry of a structure used as a permanent dwelling to commit a felony or theft. They use three classifications in specifying burglary: forcible entry, unlawful entry where no force is used, and attempted forcible entry. In many jurisdictions, burglary also includes “distraction burglary,” where falsehood, trick, or distraction is used on an occupant to gain access to the property. In many jurisdictions, gradations of burglary are distinguished depending on whether the offender was armed, whether co-offenders were involved, whether it took place at nighttime, and whether the property was occupied when entered.

The nature of the property entered may also play a role of the definition. Some countries exclude theft from a secondary residence or from an attic or basement in multi-dwelling buildings. Some countries include theft from a car (Aebi et al. 2006). Illegal entry and theft from structures not used as a dwelling, such as schools, offices, and shops, are burglaries, but they are not domestic or residential burglaries.

Because surveys have become influential sources of knowledge about burglary, it is important not only to consider legal definitions of burglary but also to understand how burglary is defined in survey questionnaires and to note the differences between these definitions. In the International Crime Victimization Survey (ICVS), these questions are the following (van Dijk et al. 2008: Appendix 8, Q60 and Q65):

> Over the past five years, did anyone actually get into your home/residence without permission, and steal or try to steal something? I am not including here thefts from garages, sheds or lock-ups.

> Apart from this, over the past five years, do you have any evidence that someone tried to get into your home/residence unsuccessfully? For example, damage to locks, doors or windows or scratches around the lock?

The questions in the British Crime Survey (Nicholas et al. 2007, p. 75) are similar but refer to the last 12 months:

> During the last 12 months . . . has anyone GOT INTO this house/flat without permission and STOLEN or TRIED TO STEAL anything?
>
> [Apart from anything you have already mentioned], in that time did anyone GET INTO your house/flat without permission and CAUSE DAMAGE?
>
> [Apart from anything you have already mentioned], in that time have you had any evidence that someone has TRIED to get in without permission to STEAL or to CAUSE DAMAGE?

Those in the US *National Crime Victimization Survey* (*NCVS*) refer to the 6 months that precede the interview. The main burglary question in the NCVS is:

> In the last 6 months, has anyone broken in or ATTEMPTED to break into your home by forcing a door or window, pushing past someone, jimmying a lock, cutting a screen, or entering through an open door or window?
>
> Based on the answers to subsequent follow-up questions (that also cover theft from garages, sheds and hotel rooms), and in line with the UCR classification, the *NCVS* distinguishes between completed burglaries (either "forcible entry" or "unlawful entry without force") and "attempted forcible entries."

Sources of Knowledge

The available knowledge of residential burglary stems from the criminal justice system, from victims, and from offenders. Agent-based computer simulation is a new approach.

Most countries collect and publish statistics on crime recorded by the police and other law enforcement agencies. The European Sourcebook of Crime and Criminal Justice compiles and harmonizes crime and justice data, including burglary statistics, from about 40 European countries (Aebi et al. 2006). In the United States, the Uniform Crime Reports (UCR) are the major source of information at the federal level collected by law enforcement agencies.

The crime victimization survey was developed in the 1970s as an instrument to measure the "dark figure" of crime not reported to or not recorded by the police. In these surveys, a random sample of the population is asked to report their victimization experiences in the period before the survey. In the United States, the *National Crime Victimization Survey* started in 1972. A special methodological feature of the *NCVS* is that it employs a rotating panel design: households remain in the sample for, at most, seven interviews, a feature that affords longitudinal data, albeit over a relatively short period. Another distinctive feature is the use of the first interview as a "bounding interview" to minimize the telescoping effect, that is, the tendency of respondents to report that incidents are more recent than they actually are.

The International Crime Victimization Survey (ICVS), which started in 1987, is the only standardized victimization survey that includes respondents from a large number of countries. There have been five main sweeps of the ICVS (1989, 1992, 1996, 2000, 2004–2005). In the most recent sweep, 38 countries were represented, of which 30 relied on nationwide samples.

Offender accounts are the third source of information used to gain knowledge on burglary. Because the detection rate of burglary is universally low, offender accounts are not useful for estimating the size of the burglary problem. Rather, offender accounts help us understand why and how they commit burglaries. Most offender-based research on burglary has used accounts of detained offenders, but some have sought access to active offenders "out on the street" (Wright and Decker 1994).

Computer simulation, in particular agent-based modeling, is a relatively new approach to understanding phenomena related to burglary (Birks et al. 2012). In agent-based models of burglary, mathematically explicit behavioral theories are formulated by specifying behavioral rules for autonomous "agents" (e.g., burglars, victims, and law enforcement officers). The system-wide implications of their actions (e.g., the burglary rate, the amount of repeat victimization, or the level of spatial clustering of burglaries) are studied and are compared to phenomena in the real world.

Incidence, Prevalence, and Trends

Victimization surveys are generally seen as the best available source for estimating the residential burglary rate. Three related measures have been used to determine burglary victimization quantitatively. *Prevalence* is the number of burglary victims as a percentage of the population, *incidence* (or the burglary rate) is the number of burglaries per person or household, and *concentration* is the number of burglaries per victim. The distinction is important because a substantial number of burglaries have been shown to be repeat burglaries at the same address within a relatively short span of time.

Burglary is a prevalent crime worldwide. For the industrialized countries that took part in the 2004–2005 ICVS, annually there were on average 4.4 completed or attempted burglaries per 100 households. With the exception of Mexico (8.8 burglaries per 100 households), the burglary rate tends to be highest in Anglo-Saxon countries, in particular England and Wales (7.9 %), New Zealand (7.8), the United States (7.5), and Australia (6.1). Countries with a low burglary rate are Japan, Spain, and the northern European countries of Sweden, Finland, Norway, and Germany.

In most countries for which data are available, the ICVS shows that the incidence and prevalence of burglary have been decreasing since the early 1990s (van Dijk et al. 2008, pp. 66–68). The ICVS findings on the long-term trend in burglary rates are supported by data from individual countries, such as the *NCVS* in the United States and the BCS in England and Wales, which may be more reliable because they are updated annually and survey a large within-country sample.

The burglary rate (incidence) in the United States has been decreasing steadily, from 11 burglaries per 100 households in 1973 to fewer than 3 in 2005. In England and Wales, the long-term trend since 1981 is not linear. From 1982 onward, the burglary rate steadily increased until it reached a peak in the early 1990s. Thereafter, however, from 1995 to 2005, the number of burglaries dropped by 59 % (Nicholas et al. 2007, p. 74).

It has been suggested that this continuing long-term decrease in residential burglary is caused by the increased use of antiburglary devices and measures, such as alarm systems, locks and bolts, and improved lighting (van Dijk et al. 2008). The evidence for this claim is reviewed in the section on prevention.

Reporting to the Police

Not all crime victims report to the police. The seriousness of an offense is the best predictor of whether a victim reports to the police (Goudriaan 2006). In the case of burglary, victims of burglary attempts are less likely to report to the police than are victims of completed burglaries, victims of completed burglaries are more likely to report if force was used to gain entry, and victims with (larger) monetary losses are more likely to report than those who lost nothing or little. The reasons victims retrospectively give about their motivations to report or not report to the police vary from perceived moral obligations ("crimes should be reported") to cost-benefit considerations ("to recover property," "to collect insurance").

Burglary reporting rates vary across countries. No less than 90 % of completed burglaries are reported in the Netherlands, Belgium, and Scotland. In the USA, the percentage is 77. The difference between the reporting rates of completed and attempted burglaries is remarkably constant: in most countries covered, completed burglaries are reported about twice as often as attempted burglaries (van Dijk et al. 2008, p. 114).

Targets

The burglary target selection process has been described as a spatially structured hierarchical process, in which the offender first selects a geographic area that fits his or her purposes and subsequently targets a specific street and a specific property on the chosen street. Once the burglar is inside the property, the items to be stolen become the actual targets. Alternatively, in some cases, burglary is a relational offense, for example, when a former husband burglarizes the home of his ex-wife for retaliation, so that the victims rather than their possessions are the real targets. Consequently, a burglary can have multiple "targets." The present section reviews which characteristics of areas, individual premises, and victims are associated with burglary risk and what items tend to be stolen.

Areas

Urbanized areas have higher burglary rates than rural areas, and burglary rates are elevated in deprived areas, in ethnically mixed areas, and in areas with high residential turnover. The relationships between burglary risk and deprivation, ethnic heterogeneity, and residential turnover have been explained as the result of two very different mechanisms.

The first is that all three are associated with a lack of social control among residents. When social control (or social organization or collective efficacy) is lacking, residents are less likely to take notice of crimes taking place in their environment and, if they do, are less likely to intervene personally or call the police to stop it. Thus, where social control is lacking, burglars run lower risks of apprehension, and residents run higher risks of victimization.

The second reason is that most offenders themselves live in deprived, ethnically mixed, and unstable areas and usually offend within their own awareness spaces. Thus, these measures are largely synonymous to being nearby and exposed to concentrations of motivated offenders. This may explain the paradox that although offenders themselves claim that they select prosperous targets, the empirical evidence on actual targets is that burglary is concentrated in deprived areas. Apparently, offenders' preferences for prosperous targets are local preferences: they prefer the most prosperous targets within the (deprived) area they are familiar with. It has indeed been shown that visibly affluent households (e.g., detached and semidetached properties as opposed to terraced properties and flats) in the most deprived areas have the highest risks of burglary victimization.

Premises

A number of physical characteristics of residential units have been associated with burglary risk. Some of these apply to the location of the house or to the design of its immediate environment. Others apply to the structure of the unit itself. Proximity to major thoroughfares increases risk, and being located in a dead-end street or cul-de-sac decreases it, possibly because the former are more and the latter are less likely to become part of many offenders' awareness spaces. The findings are also compatible with a preference on the part of burglars for houses that allow for multiple exit routes. With respect to the units themselves, it has been shown that properties at the corner of a street block have larger burglary risks than properties in the middle of the block and that the burglary risk is elevated in detached and semidetached houses (as compared to terraced properties and flats), especially if they border playgrounds, woods, or other nonresidential area or if parts of the house are not visible from the street.

The three main property selection criteria that burglars use are surveillability, occupancy, and accessibility. Surveillability indicates whether properties can be overseen by other people. It is indicated by distance from the street, the absence of trees or hedges that block sightlines, and the presence of lighting. Occupancy is whether there are people at home. To the prospective burglar, it may be indicated by the presence of noise or light in the house, a car on the driveway, toys in the garden, or the absence of unopened mail. Accessibility indicates how easy it is to break into the property. Open doors and windows are a case in point, although they might also signal occupancy. Dogs and target-hardening devices, such as window locks and alarm systems, restrict accessibility.

Victims

The lifestyles of potential victims play a major role in their victimization risk. Because to commit their offense burglars mostly depend on the times that residents are away from home, a major predictor of burglary victimization is the proportion of time a property is unoccupied. This proportion is directly related to the frequency with which residents go out in the evening and go shopping. It is indirectly related to the composition of the households – single-person households and single-parent families have higher risks – and with age, younger households have higher risks. Renters (as opposed to homeowners) and unemployed residents also have elevated burglary risks, but the reason is unlikely to be related directly to the proportion of time their houses are unoccupied.

Although the issue is not reported extensively in the literature, some results suggest that burglary is not always a crime directed against anonymous victims but one in which victims are targeted who are known to the offender. According to the 1998 British Crime Survey, 41 % of burglary victims were able to say something about the offenders. The offender was a complete stranger to the victims in only 49 % of these cases; in other cases, the offender was casually known (17 %) or even well known (34 %). While the 41 % of cases in which the victim was able to say something about the offender are probably not representative for all burglaries, it emphasizes that burglary offenders and victims quite often know each other. In some cases, for example, those in which a former spouse burglarizes his or her partner's home, it may be the victim who is the burglary target.

Items

If the victim is not the target of a burglary, the item to be stolen is the ultimate focus of acquisitive crime. What do burglars steal? Police records and victim surveys show that the most frequently stolen items are cash, jewelry, and portable electronic gear such as cell phones, cameras, audio and video equipment, computers, game consoles, and televisions. Most of these items are "hot products" that have attractive features for thieves in general and burglars in particular (i.e., they are concealable, removable, available, valuable, enjoyable, and disposable). Cash money is most burglars' favorite item.

Offenders

Burglary is a covert crime. Contemporary knowledge of it is hampered by the fact that the detection (clearance) rates of burglary tend to be low: 12.6 % in the United States in 2006 and generally below 10 % in cities with a population above 500,000, much like those of other covert crime. As a consequence, knowledge of offenders is based on interviews with those who have been arrested or sometimes on observations of and interviews with active burglars in natural settings.

When asked about motives or precipitating factors in the decision to commit a burglary, most offenders mention financial need. Often offenders are driven by the wish to continue a lifestyle they cannot afford without offending. In particular, the continued use of expensive drugs appears to motivate burglars. The influence of others further seems to trigger the involvement in burglaries, and some mention boredom and a need for thrills. Burglaries tend to be committed by young males. Of the burglars arrested in the United States, 86 % are male and half of them are younger than 22.

Among juvenile offenders, burglary tends to be a joint activity. The major reason for this appears to be social rather than practical, as the group setting induces potential offenders for various reasons to join a risky endeavor they would not get involved in on their own. For practical reasons, such as the possibility of carrying stolen items and the usefulness of having one offender act as a lookout, one would expect co-offending to be common in burglary. Offender accounts, however, suggest that adult offenders are solitary burglars most of the time, although one might question whether these accounts are always valid, as interviewed burglars tend to be very reluctant to talk about their accomplices.

Unlike popular images of the persistent burglar, burglary specialization is uncommon. Most offenders are highly versatile, and this appears to be true for burglars as well. There is no evidence that burglars will use violence only when they are unexpectedly confronted with victims or bystanders. Many offenders who commit burglaries also commit violent offenses.

Some authors have proposed a typology of burglars. Maguire and Bennett (1982) distinguish "high-level burglars," "middle-level burglars," and "low-level burglars," a typology that is confirmed by the offense styles described by Bennett and Wright (1984, pp. 43–49), who distinguish between "planners," "searchers," and "opportunists," respectively. Opportunistic offenses occur when a burglary opportunity presents itself and is immediately, without further planning, acted upon. The decision to burglarize, the selection of the target, and the burglary itself take place with little or no time gaps in between. In the searching offense, there is a time gap between the decision to commit a burglary and the selection of a target, but not between the target selection and the act of burglary; in other words, the searcher explicitly searches for a target and attacks when it is found. A planned burglary involves time gaps both between the decision to burglarize and the target selection and between the target selection and the actual burglary; thus, there is forethought and preparation before each phase. As these findings are based on accounts of arrested offenders and because it is possible that offense styles are related to the likelihood of apprehension, it is virtually impossible to estimate how these styles are distributed in the population.

Time

The risk of residential burglary varies across the hours of the day, across the days of the week, and across the weeks of the year. Because burglars tend to avoid confrontations with residents and prefer unoccupied targets, the exact timing of a burglary is often unknown; it is typically discovered when the residents find their place burglarized on returning home. The timing of burglaries is specified in police records using a time window that reflects when residents left their home and when they returned. According to the *NCVS*, in 28 % of the burglaries in the United States in 2005, the victims or the police had no idea what time of day the burglary was committed, although daytime burglaries (between 6 a.m. and 6 p.m.) were slightly more common (53 %) than nighttime burglaries (47 %) in those cases whose timing was known. In England and Wales, 61 % of burglaries occur between 6 a.m. and 6 p.m. (Budd 1999, p. 19). But the patterns display much larger variations, with the least likely burglary hours being those that most residents spend at home.

As about 30 % of burglaries occur during the weekend, from 6 p.m. on Friday evening to 6 a.m. on Monday morning (Budd 1999, p. 19), there is little evidence for a weekly time cycle in burglaries.

Like most types of crime, burglary appears to fluctuate systematically over the seasons of the year. In the United States, burglary rates peak during the summer months and are below average in winter. Interestingly, the pattern is reversed in England and Wales, which are also in the northern hemisphere but where burglaries peak in the winter, especially December and January.

Some, if not all, of the temporal variation in the timing of burglaries is induced by the time use of potential victims or guardians: burglaries usually take place when residents or housekeepers are away from home and sometimes when they are asleep. In addition, offenders themselves have routine activities, such as school or work, that may give rise to burglary opportunities during certain times of the day.

Another part of the (daily and seasonal) temporal variation is induced by burglars' preference to offend under the cover of darkness in order to minimize their risk of being seen and recognized and arrested. As a preference for darkness is not always compatible with a preference for unoccupied homes, burglars' preference for darkness may depend on attributes of the potential targets. For example, it has been found that in daylight, burglars select targets in upmarket low-density residential areas where residents are employed, whereas in darkness, they preferred to target dwellings in deprived high-density areas.

The act of burglary itself is typically short. According to burglar accounts, the majority of burglars spend less than 10 min in the dwelling they burgle.

Repeat Victimization and Risk Communication

Repeat burglary victimization occurs when a property is burglarized more than once within a specified period (e.g., a year). Many studies have demonstrated that prior burglary victimization is associated with an elevated risk of future burglary victimization (Johnson et al. 1997). Repeat burglaries tend to occur swiftly (Polvi et al. 1991) after the initial burglary. Often, revictimization takes place within days or weeks. After a short period, the risk declines rapidly until it reaches its original level. While the data, methodologies, and outcomes differ somewhat across studies, a characteristic exponential decay in the time course of repeat burglary victimization has been confirmed in many studies.

There are two explanations for these findings (Tseloni and Pease 2003). The first is that burglary victimization simply *flags* properties with lasting attributes that attract offenders. According to this explanation, both the initial burglary and the repeated burglary reflect the elevated risk associated with stable attributes of the target. The second mechanism is that the initial victimization *boosts* the likelihood of a repeat. Under this mechanism, the initial burglary alters something about the property or the victim that increases the risk of revictimization.

It has been argued that the temporal pattern of repeat burglaries in particular often suggests the involvement of the same offender or offender group in both offenses (Polvi et al. 1991). Indeed, the boost explanation is compatible with the possibility that a repeat offense against the same premise involves the same offender who committed the initial offense and who returns to collect items not stolen during the initial burglary or that have been replaced since then. On the other hand, in the wake of a burglary, one should expect victims to be extremely vigilant and maybe to install burglary prevention devices, which should logically decrease the risk of repeat victimization. The typical time course of repeat burglary supports the boost explanation of repeat victimization. In particular, it has been viewed as tentative evidence that in a typical repeat burglary, the perpetrators are the same people who were involved in the initial event. Although the exponential decay in the time course itself is not sufficient evidence for this claim, as it may also indicate unobserved risk heterogeneity (Townsley et al. 2000), there is also evidence from interviews and offender accounts that returning to a previously targeted property is a common burglar strategy, especially among prolific offenders.

In the wake of a burglary, properties near the targeted property run heightened burglary risks as well. The phenomenon was first locally established in Australia and England. The ubiquity of such “near repeats” has been demonstrated in no fewer than ten regions around the world (Johnson et al. 2007). Involvement of the same offenders who committed the initial burglary is also likely in near repeats. Near repeats could be displaced repeats, for example, if an offender returns to a previously burglarized property but finds it well secured and subsequently targets an alternative nearby property.

Prevention

What preventive measures have been taken to reduce burglary, and what is known about their effectiveness? Burglary prevention is not typically concerned with changing the attitudes of offenders, partly because efforts to change the delinquent attitudes and behavior of burglars cannot be very specific, as offenders are quite versatile. Another reason is that the burglary detection rate is so low that only a small minority is ever arrested and eligible for rehabilitation. Increases in the burglary detection rate may decrease the number of burglaries, as increased detection will translate into burglary’s being perceived as a much more risky crime for offenders than before. Thus, although improving burglary detection cannot prevent the burglary that has been detected, it may deter the same and other offenders from committing future burglaries.

In this respect, new developments in DNA profiling are interesting, as DNA is most helpful in crimes that are the most difficult to detect. Although DNA samples currently make a relatively small contribution to all detections, they make a powerful contribution when they are available.

Most conventional preventive measures against burglary can be categorized as situational crime prevention, which include measures directed to change the immediate situation in which a burglary could potentially occur. Situational measures against burglary can be taken at various levels. First, target-hardening measures can be taken at the level of the individual property. Data on target-hardening measures of individual properties in the ICVS demonstrate that the penetration of burglar alarms and special door locks is higher in Anglo-American countries (England and Wales, Australia, United States, Canada, Scotland, Northern Ireland) than in Europe and Japan and that it is highest in England and Wales (van Dijk et al. 2008, pp. 135–139). Typically, at country level, the dissemination of target-hardening devices is more or less proportional to the burglary rate. For example, England and Wales and Australia have the highest burglary rates and the highest levels of installed target-hardening devices. The most likely explanation of this finding is that citizens respond to high burglary rates by trying to defend themselves against burglaries.

A considerable number of British householders have installed devices that strengthen the physical barriers against unlawful entry to their property, such as double locks or deadlocks (76 %), window locks (80 %), and security chains on doors (32 %). Further, many have taken measures that draw attention to illegal entries, such as outdoor (40 %) or indoor (24 %) sensor or timer lights and burglar alarm systems (29 %). These home security measures might appear to be a key defense against burglary victimization. Households with no home security measures were almost ten times more likely to have been burgled as households where there were simple security measures such as deadlocks on doors and window locks. Alarm systems only weakly differentiate between households that were victimized and those that were not. That simple security measures such as locks help to keep burglars out is also suggested by the fact that victimized households that experienced an attempted burglary were more likely to have double locks or deadlocks than those where entry was gained (Nicholas et al. 2007, p. 76). This is not the case with respect to alarm systems: an equal percentage of victims of completed burglaries and victims of attempted burglaries has an alarm system. If alarms are not effective, this might be due to their lack of specificity. In Charlotte, North Carolina, it was found that 98 % of alarm activations were false alarms (LeBeau and Vincent 1997).

The simple negative correlations between target-hardening measures and burglary risk may be spurious and may indicate that low-risk households (e.g., homeowners, high-income groups) invest more in burglary prevention than do high-risk households for reasons not directly related to the actual or perceived burglary risk (e.g., because the rich can better afford antiburglary devices or because homeowners can expect longer term benefits).

Some target-hardening measures apply to larger entities than individual premises. Alley-gating, for example, is a successful target-hardening measure that prevents unauthorized entry from the rear of properties of a block of houses. In general, the deterrent effects of individual (property-level) measures are stronger if these measures are also adopted in the wider community.

There is probably no government in the world that has invested as much in reducing burglary as England and Wales, starting with the Safer Cities project in 1988 and more specifically with the Reducing Burglary Initiative that started in 1999. As most of these initiatives are being constantly evaluated (many of them applying quasi-experimental designs with control groups and pre-intervention and post-intervention measurements), this has also generated a host of information on the effectiveness of various prevention methods.

A comprehensive evaluation of nearly 300 antiburglary schemes in England and Wales

(Ekblom et al. 1996, p. 41) found that a combination of target-hardening and community-oriented action (such as fostering Neighborhood Watch, property marking, raising burglary awareness among residents) worked best but that target hardening could also work alone. This is in line with other research on target-hardening interventions, which has concluded that whole-area target hardening can reduce local burglary rates in the short term if continuous publicity is heard or viewed by prospective offenders.

Independently of the prevention tactic chosen, publicity appears to increase the effects of prevention schemes. Apparently, this works because it informs and deters potential offenders from burgling in the area, although publicity may also sensitize potential victims and increase their vigilance. When publicity precedes the implementation, it often reduces burglary in the period before the actual intervention (this may actually undermine the proper evaluation of the effect of the intervention itself, however). It may also reduce crime in a larger geographic area than where the intervention is implemented, and it may prolong the benefits of the intervention.

In general, the largest burglary reductions are obtained by focusing on the areas or on the victims with the highest burglary risks. This explains part of the success of prevention schemes in England that were aimed at preventing repeat burglary victimization by varying means. The Kirkholt and the Huddersfield projects resulted in substantial burglary reductions through multitactic interventions, probably in part because they addressed people who had recently been victimized and thus sensitive to and aware of the emotional and material consequences.

Conclusion

While over the years there has been a lot of continuity in the literature, some important developments may be highlighted. The first is the worldwide drop in burglary. The debate on the causes of the drop is ongoing, but the drop itself cannot be mistaken.

The second development concerns the nature and comprehensiveness of information on burglary. In 1991, the *National Crime Victimization Survey* had been under way in the United States for nearly 20 years, the British Crime Survey had just started in 1982 and did not yet follow an annual cycle, and the International Crime Victimization Survey gained momentum only during the 1990s. Similar developments can be observed with respect to police-reported burglary. In the United States, the Uniform Crime Reports have been reported for decades, yet many other countries have lagged behind. The European Sourcebook is an example of a new initiative to collect and standardize information from various countries and to stimulate international comparison. Noteworthy are a number of authoritative studies based on offender accounts of burglary, either based on prison interviews (e.g., Rengert and Wasilchick 2000) or interviews with active burglars (e.g., Wright and Decker 1994).

A third development to be highlighted is the increasing salience of repeat burglary victimizations in the literature. The success of the repeat burglary prevention programs in the United Kingdom has shown that repeats can be prevented, thereby reducing overall burglary rates substantially. An emerging line of research generalizes repeat victimization by showing that risk of revictimization can be communicated to nearby dwellings.

Some issues require more research. One is the offender-victim nexus. It has been demonstrated repeatedly that offenders are disproportionately likely to be crime victims themselves; another question is whether offenders and victims in the same incident are strangers to each other. Although the issue has not been studied extensively, some findings suggest that quite often victims have an idea of who the offenders are. Ethnographic research shows that offenders quite often know who their victims are, as they tend to select them from a pool of acquaintances.

A second issue that requires more research is the effectiveness of situational crime prevention measures. Although it has been argued that surveys demonstrate strong negative correlations between target-hardening and burglary risk and although trend analyses show that increasing levels of burglary prevention devices are aligned with decreasing burglary rates, the evidence has not always been compelling, and the findings stand in strong contrast to the accounts of offenders, who generally emphasize that target-hardening measures play a minor role in the selection of dwellings.

The detection of burglary should be a third research priority. Worldwide the detection rate of burglary is low and a concern to the police. Possibilities for burglary detection offered by contemporary detection methods and tools (such as DNA matching, CCTV surveillance) have not yet been systematically investigated. The low burglary detection rate should also concern investigators who base their findings on police data or on accounts of arrested burglars. If fewer than 10 % of the burglaries are detected, is there any guarantee that those arrested are representative of the total burglar population? DNA databases may help to solve this puzzle, not by increasing the detection rate but by providing evidence of the behavioral patterns (geographic, temporal, modus operandi) of those offenders who have never been arrested but whose DNA stains have been left at multiple burglary scenes. If the behavioral patterns of these burglars resemble those of arrested offenders, it would increase confidence in present findings that are based on arrested offenders only.

Related Entries

- ▶ Crime Prevention Through Environmental Design
- ▶ Modus Operandi of Sex Offenders
- ▶ Near Repeats and Crime Forecasting
- ▶ Offender Decision Making and Behavioral Economics
- ▶ Routine Activities Approach
- ▶ Situational Crime Prevention
- ▶ US National Crime Victimization Survey

Recommended Reading and References

Aebi MF, Aromaa K, Aubusson de Cavarlay B, Barclay G, Gruszczyñska B, Von Hofer H et al (2006) European sourcebook of crime and criminal justice statistics – 2006, 3rd edn. Boom Juridische Uitgevers, Den Haag

Bennett T, Wright RT (1984) Burglars on burglary: prevention and the offender. Gower, Aldershot

Birks D, Townsley M, Stewart A (2012) Generative explanations of crime: using simulation to test criminological theory. Criminology 50:221–254

Budd T (1999) Burglary of domestic dwellings: findings from the British crime survey. Home Office, London, No. Home Office Statistical Bulletin 4/99

Ekblom P, Law H, Sutton M (1996) Safer cities and domestic burglary. Research and Statistics Directorate, Home Office, London

Goudriaan H (2006) Reporting crime: effects of social context on the decision of victims to notify the police. PhD Thesis, Leiden University

Johnson SD, Bowers KJ, Hirschfield A (1997) New insights into the spatial and temporal distribution of repeat victimization. Br J Criminol 37: 224–241

Johnson SD, Bernasco W, Bowers K, Elffers H, Ratcliffe J, Rengert G et al (2007) Space-time patterns of risk: a cross national assessment of residential burglary victimization. J Quant Criminol 23: 201–219

LeBeau JL, Vincent KL (1997) Mapping it out: repeat-address burglar alarms and burglaries. In: Weisburd D, McEwen T (eds) Crime mapping and crime prevention. Criminal Justice Press, Monsey

Maguire M, Bennett T (1982) Burglary in a dwelling. The offence, the offender and the victim. Heinemann, London

Nicholas S, Kershaw C, Walker A (eds) (2007) Crime in England and Wales 2006/07. Home Office, London

Polvi N, Looman T, Humphries C, Pease K (1991) The time-course of repeat burglary victimisation. Br J Criminol 31:411–414

Rengert GF, Wasilchick J (2000) Suburban burglary: a tale of two suburbs. Charles C. Thomas, Springfield

Townsley MT, Homel R, Chaseling J (2000) Repeat burglary victimisation: spatial and temporal patterns. Aust N Z J Criminol 33(1):37–63

Tseloni A, Pease K (2003) Repeat personal victimization. Boosts or flags? Br J Criminol 43:196–212

van Dijk JJM, van Kesteren JN, Smit P (2008) Criminal victimisation in international perspective, key findings from the 2004–2005 ICVS and EU ICS. Boom Legal Publishers, The Hague

Wright RT, Decker SH (1994) Burglars on the job: streetlife and residential break-ins. Northeastern University Press, Boston

Resistance

▶ Identification Technologies in Policing and Proof

Resistance to State Crime

Elizabeth Stanley[1] and Jude McCulloch[2]
[1]School of Social and Cultural Studies,
Victoria University of Wellington,
Wellington, New Zealand
[2]Arts Faculty, Monash University, Clayton,
Victoria, Australia

Overview

Resistance is an undeveloped aspect of state crime scholarship. Criminologists have been successful in naming state crimes and exposing the violence and harms committed by states but less so in documenting, analyzing, and theorizing the ways individuals and movements challenge, oppose, and even prevent and stop state crimes. State crime scholars are increasingly interested in defining resistance and describing the spectrum of activities that might be considered resistance. Criminologists are also interested in exploring the impact of resistance on state crime and reflecting upon what makes resistance effective in both the short and longer term, along with the factors that might encourage or discourage people to engage in resistance. State crime scholars also need to consciously consider the ways their scholarship might contribute to resistance. This entry sets out criminological contributions to the study of resistance to state crime and maps out some potential future directions for scholarship and research.

Introduction

The capacity of state officials and their contractors to kill, maim, exploit, repress, and cause widespread human suffering is unsurpassed (Green and Ward 2004). The extent and nature of state violence and harm has been successfully exposed in state crime scholarship. Crimes committed by states far outweigh those by individuals and non-state actors. State crime scholars have analyzed how acts including genocide, torture, corruption, escalating militarization, and environmental destruction have been led by the state. They have also developed understandings of how state crimes are managed at individual, societal, institutional, and structural levels; illustrating, for instance, how individual perpetrators are "made"; how bystanders fail to intervene; how state organizations fail to provide justice to victims; and how state violence is typically directed at the least powerful (see, e.g., Cohen 2001; Huggins 2010; Stanley 2009). State crime scholarship has provided significant new approaches in the way criminologists think about crimes, victims, and perpetrators.

One aspect of state crime literature that remains undeveloped is resistance – criminologists have so far paid little attention to the instances where state crime has been challenged or contested and even halted or prevented. Criminologists have tended to present a depressing catalogue of inhumanity and suffering without a corresponding analysis of the humanity and vitality of those who have resisted state violence. Many key historical moments have involved resistance to state crime: the abolition of slavery, the resistance movements of World War Two, the achievements of the civil rights movement in the United States, the implementation of the International Criminal Court, and the 2011 struggles across the Middle East provide just a few examples. Resistance to state crime is ripe for further criminological study.

This entry provides an overview of the criminological insights and debates on resistance to state crime thus far. In addition to exploring what resistance might mean in the context of state crime, it responds to a number of key questions: Who engages in resistance? What might be viewed as "successful" resistance?

And, what might be the role of criminologists in supporting or engaging in efforts to resist state crime? Finally, the entry highlights some directions that criminologists might take in future resistant responses to and analyses of state crime.

Fundamentals

While much of criminology has focused on the support of state policies and practices, even where such policies and practices lead to routine and institutionalized harms, there is a growing literature that focuses on state crime. Criminologists have debated the construction of crime, critiqued the state, and have sought to take control of the definitions attributed to harmful or injurious behaviors. In doing so, scholars have implicitly engaged in or supported resistance to state crime. State crime scholars have been adept in critically analyzing, and refusing to accept, the techniques of denial that frequently accompany acts of state criminality (Cohen 2001). Moreover, in analyzing how the identities of state crime victims are ideologically managed, so that victims are frequently defamed as terrorists or criminals, they have confronted popular conceptions that victims of state crime are deserving of harsh treatment, violation, and even elimination (Stanley 2009; McCulloch and Sentas 2006). State crime scholars have "unpacked" how state crimes come to be made acceptable and justified within complex political processes of neutralization and criminalization. In taking this route, some criminologists have asserted that numerous state actions are unfair, unjust, troubling, damaging, and in many instances crimes of grave magnitude. By their very nature, these analyses are imbued with a sense of resistance.

State crime scholars have attempted to intervene in state crimes by challenging state impunity. For instance, criminologists Hagan et al. (2005) have, through statistical and regression analyses, demonstrated how the Sudanese government directly supported the killings and rapes of Darfurians. In illustrating the racial targeting of African Darfurians by state actors, they bolstered prosecutorial evidence in characterizing these events as state-sponsored genocide. This study, like others of its kind, illustrates that criminological work can play an active role in resisting state crime.

Criminologists have, however, generally failed to consciously consider resistance either as an integral part of state crime scholarship or as a discrete topic to be pursued within state crime studies (see, however, Green and Ward 2000; Pickering 2002). While resistance can be read into state crime scholarship, it is generally embedded in another "story," such as developing a historical context, illustrating techniques of denial, or simply writing against state crime. A "more consciously 'applied' form" of analysis on resistance is needed (Friedrichs 2010, p. 5). This goal underpinned the first symposium on state crime and resistance, convened by Elizabeth Stanley in New Zealand in January 2010. This event propelled new thinking and writing about resistance – including a special edition of *Social Justice* in 2010 as well as a Routledge book edited by this entry's authors (Stanley and McCulloch 2013).

State crime scholars have grappled with defining resistance to state crime. After all, the scope of activities that might be considered as resistance is wide. Resistance is an elastic term that covers various actions. It has multiple dimensions and operates along a continuum. In relation to contesting state crime, Sharon Pickering (2002) and Rob White (2010) argue that resistance could include actions that are passive or active, hidden or overt, peaceful or violent, verbal or physical, ad hoc or strategic, a one-off event or a perpetual happening, concentrated on a limited event or led by a global focus, led by "experts" or by participatory engagement, underpinned by fun or by fear and terror, and small scale or large scale. The motives, contexts, and practices that encompass resistant acts are as varied as the state crimes that they seek to challenge.

Nonetheless, for something to be resistance rather than just a random act, there have to be some defining features. In recent work, Stanley and McCulloch (2011) establish three primary elements of resistance. These elements are intention, communication, and action. First, resistance is about opposition with social, moral, or political intent. The intention is crucial. In this respect, resistance requires some level of consciousness or human agency; it involves a choice. Mohandas Gandhi, seen as the modern father of nonviolent resistance, articulated this view in his idea that resistance was a "conscious option" for resisting injustice (see Urquhart 2011, p. 37). Other scholars, such as Henry Giroux (1983), have argued that opposition is a necessary condition of resistance but that acts of resistance must also be carried out with the aim of emancipation or transformation. The intention to change, or retain a status quo, is a key part of any resistant act.

Second, resistance is a form of communication that constitutes a message or messages. Sometimes, resistant acts can be registered as clear statements directed at dominant powers or the public. A protest march clearly aims to communicate – although, of course, underlying protest messages can sometimes be lost or subverted by state discourses of public order and control. Resistance has the most potential when it reflects the aspiration of a broad range of people. Messages are frequently designed to encourage bystanders to become participants in resistance movements. Sometimes, however, the communicatory nature of resistance may not be understood by or revealed to a wide audience. Quiet, small, or more personal acts of resistance may not be "read" by state agencies or the general public. Nevertheless, they are forms of resistance where there is an element of communication, even when the intended audience is small or intimate. For instance, during the Pinochet-led repression in Chile from 1973 to 1990, female opponents created *arpilleras*, political quilts that represented the pain, injustice, and harm of living under the regime as well as the aspiration of moving beyond repression. These quilts depicting "disappearances," torture, and other forms of violence were private, seen only by trusted family and friends. Nevertheless, they presented a clear and formidable intention to expose state crime and to delegitimize what was happening in the name of counter-terrorism. Acts of resistance can be assessed in terms of how they communicate new ideas and how they challenge dominant, state-centered discourses and generate discourse on state crime, and integral to this is how they subvert popular assumptions about perpetrators, victims, or bystanders, and raise awareness.

A third element of resistance is action towards creating, asserting, or arresting change. Resistance is bound up with the ways in which people understand and act on their capacity to make or oppose changes. These actions can be diverse, for example, they may involve the creation or consolidation of different ways of living or they may involve active or passive opposition to change or in pursuit of change. The following section elaborates upon several key issues in this emerging state crime field.

Key Issues

This section explores some of the central themes that have guided criminological writings on state crime and resistance thus far. In particular, it explores three questions: Who is involved in resistance? What might be viewed as effective resistance? And, how might criminologists resist state crime?

Who Is Involved in Resistance?

State crime scholars have highlighted the fundamental paradox that while states are major perpetrators of crime and active contributors to mass victimization at a global level, they also hold the position as the main protector of people, animals, and environments (Friedrichs 2010; White 2010). Consequently, while those engaged in activism and resistance rally against state actions, they also direct state authorities to defend and protect rights. This unique paradox can make it extremely difficult for resisters to gain truth, redress, or justice on state crime issues. The reason for this is that states will regularly prefer to preserve and protect state institutions rather than protect the people those institutions are purportedly established to serve. Criminologists such as Jamieson and McEvoy (2005) highlight the way that state perpetrators deny their involvement in state crime and engage in elaborate techniques of "othering" to hide their direct participation.

This paradox also impacts on how bystander states challenge criminal states. For instance, notwithstanding the frequent claims made by powerful states about challenging impunity, protecting human rights, and securing justice, "the international political community remains wedded to the practice of *Realpolitik* in terms of applying international criminal law" (Rothe 2010, p. 112). Challenges from bystander states – for instance, to extradite or arrest individual perpetrators – are often hindered and prevented by shifting economic, political, strategic, and military interests. Legal responses to state crimes are regularly quietly dismissed in favor of other considerations, such as developing close economic ties with perpetrating states (Stanley 2009). However, these relationships are not static or impregnable, and, at certain junctures, the constellation of interests may be such that resistance becomes possible, and (continuing the example above) some perpetrators of state crime are arrested, tried, and convicted.

In a global age, those who seek to resist state crimes must engage with multiple targets (Friedrichs 2010). The reason for this is that states are not monolithic and, in an era of globalization, state institutions will frequently operate alongside corporations, international financial institutions, the United Nations, militias, and other bodies in the committal of state-corporate crimes, eco-crimes, crimes of globalization, and finance crimes. Moreover, as Stanley Cohen's (2001) work on denial makes clear, state crimes are sustained by individuals in everyday ways. As consumers, voters, or through other social (in)action, individuals encourage, legitimize, neutralize, or become complicit in state crimes. Consider, for instance, the (media-led) public mandate given to US politicians to engage in torture following the attacks on September 11, 2001; or the usual approach of corporate shareholders not to question how their resources are spent; and popular support for punitive approaches to asylum seekers. It is often difficult to disentangle responsibilities, and criminologists have to be particularly mindful of the ways in which state crimes are operationalized across different networks of power.

Beyond identifying who or what needs to be resisted, criminologists have also begun to pay attention to who engages in resistance and how resistance strategies can cut across diverse groups. The earliest works on state criminality identified that state crime can be censured and sanctioned through a variety of different social audiences: from above, for example, through UN bodies, international courts, or domestic monitoring organizations (such as police complaints authorities); from below, through the withdrawal or erosion of public consent to state authorities (highlighted through protests against state actions); from within, by state workers exerting demands, and making changes, inside their own organizational structures; and, from without, through the work of international nongovernmental organizations (NGOs) who operate to place international pressure on criminal states. In 2009, Green and Ward elaborated upon the mechanisms and processes by which organized civil society groups engage in collaborative actions to challenge and reform criminal states. Working across different power bases at local, national, and international levels, resistant groups frequently take multipronged approaches to resist state crime. These approaches might include a combination of direct action, legal strategies, political lobbying, public relations, and media campaigns all of which are likely to overlap and feed into each other. These different approaches may be calibrated to diverse primary target audiences.

Contemporary practices of resistance transgress conventional political boundaries. The globalization of political spaces has meant that the traditional distinctions between who is a "claims maker" and who is not (at local, national, and global levels) are blurred. In addition, the rise of new technologies – such as camera-mobile phones and the Internet – has meant that state crimes can be documented and circulated across the world with relative ease and at speed. This has, in some circumstances, placed state authorities on the "back foot." Would the United States have been subject to the same level of accountability if the images of the torture and ill-treatment of prisoners at Abu Ghraib and images of hooded detainees at Guantanamo Bay had not been widely circulated? A similar question can be asked about the torture of civilians by Indonesian troops in West Papua. Likewise it has become more difficult for police to deny brutality when such events are captured on video cameras, CCTV, or mobile phones and widely circulated. Changes in technology have made it more difficult for states to claim that acts of violence have not taken place. Nevertheless, states continue to engage in denial by suggesting self-defense in the face of threats posed by criminals and terrorists. Victim-survivors are frequently recast as perpetrators despite visual evidence to the contrary.

The complexity, power, and resilience of states mean that resistance is most effective when it resonates with a community of resisters who have a shared experience and who can support each other and speak with and for each other (Scott 1990). Resisters need to mobilize a movement of concerned others. One element that celebrated resistance leaders such as Mandela, Ghandi, and Martin Luther King share was their ability to mobilize others. This point has not been lost on contemporary human rights groups, trade unions, or religious bodies who seek to develop collective strategies within and across nations in attempts to resist state crime (Stanley and McCulloch 2011). To be effective, resistance requires people to work together to achieve shared goals. In this respect, resistance is a political project in which people morally engage in practical tasks of contestation (Sivanandan 1990).

Of course, mobilizing bystanders is no easy task. Stanley Cohen (2001) has highlighted the way bystanders shut out confronting stories of state crime and distance themselves from any moral imperative to take action. This distancing occurs through various techniques. For example, bystanders can diffuse responsibility (arguing that state crimes are nothing to do with them, and because of this, they should remain uninvolved), they might struggle to identify with state crime victims (particularly if victims have been labeled as criminals or terrorists or denigrated in some other way), and they might feel powerless to help (as they cannot imagine any action that could make a substantial difference). Beyond this, people may consciously or unconsciously feel afraid to identify with or act in solidarity with state crime victims lest they too become victims. Regardless, criminologists are keenly aware that transforming bystanders into participants is crucial to any resistance strategy (Friedrichs 2010). Engaging others, within and without criminal states, is vital to challenging state crimes.

Despite the techniques of denial and the challenges of building movements of resisters, states are often faced with persistent victim-survivors who simply refuse to go away. Victim-survivors, who often have "nothing left to lose" can become "ferocious fighters and tenacious enemies" (White 2010, p. 55). This issue has been demonstrated in criminological writings on transitional justice (see Stanley 2009). In this domain, victim-survivors have been particularly strong in drawing upon international human rights standards and laws and engaging transnational campaigners and support groups to promote their demands over decades.

With fairly limited resources to push for action, victims continually make insistent claims for truth and justice. Part of this, as detailed above, can be attributed to the persistence of victim-survivors; however, it may also emerge as a consequence of the ways in which harms and injuries from state crime develop over time. Some of the crimes or harms resulting from state action may not be fully realized until many years after the initial event. For instance, the impact of depleted uranium in places like Iraq is only now becoming fully apparent in terms of the long-term health effects upon Iraqi citizens who lived on "target" sites, war veterans as well as their descendants (White 2008).

Such events highlight that resistance to state crime often requires a long-term strategy. Nonetheless, many people have shown great resilience in these endeavors. From examples like the Holocaust or the repression in Chile or the Indigenous land rights movement in Australia, it is clear that victims can achieve some "success" – decades after state crimes have occurred, victims can be acknowledged *as* victims, compensation can be paid, and prosecutions can occur (Stanley 2009). These struggles often require great levels of persistence, tenacity, and courage, and their "successful" outcomes are hard won.

What Constitutes "Successful" Resistance?

In responding to state crime, resistant acts may succeed on structural, institutional, social, and personal levels. As Stanley and McCulloch (2011) detail, successful resistance may be momentous or small; it may confront an entire regime or be limited to opposing specific state actions. It might take the form of overthrowing a criminal state, holding the state accountable for its actions, challenging the state's claim to truth, changing the socio-legal landscape to amend unjust laws or provide access to law, "answering back" to colonizing norms and values, reasserting positive values of care or dignity in society, changing the practices and values of state workers, or presenting individuals with an opportunity to take action and move beyond powerlessness.

It is impossible to quantify the multiple acts of resistance that take place all over the world on a daily basis. What is clear, however, is that "successful resistance" or "what works" is varied and depends "very much upon the immediate political struggles and social contexts" (White 2010, p. 50), although there is research to suggest that nonviolent resistance is successful more often than armed campaigns (Stephan and Chenoweth 2008). Resistance requires a recognition of state harms or injuries at a local level. Yet, in contemporary conditions, successful resistance will regularly recognize the global nature of this phenomenon and while dependent upon achieving change at a local level, may often be tied to how individual events might be cast as social problems that have a national or even international resonance. This kind of approach that goes beyond localism and generates allies across the globe can "acknowledge the intrinsic commonalities across borders and a shared moral universe among political activists" (White 2010, p. 52). Resisters do not just deal with "narrow state interests" (White 2010, p. 54) but contest fundamental problems within global structural relations of power.

Effective or successful resistance, however, is difficult to measure. What may initially be regarded as positive advances or real "victories" can turn out to be Pyrrhic, temporary, or have unintended or unforseen consequences. Resistance strategies raise "moral conundrums" (Friedrichs 2010) and may, despite altruistic intentions, adversely impact on people who have few choices and little power. State crime may also prove resilient, continuing, or reemerging in altered forms as the social, political, legal, and cultural landscapes shift (Stanley and McCulloch 2011). Indeed, some state crimes that may be viewed as having ended may continue in different forms. Slavery, for example, is regularly regarded as a historical phenomenon, but it continues unabated in contemporary forms of forced labor, debt bondage, child labor, or trafficking (Bales 2004). Contemporary slavery continues to affect millions of people; however, it often goes "unseen," as it functions as part of an advanced global capitalist system in which labor is outsourced and traded across corporate and contractor networks. In the same vein, slavery and its attendant colonial systems have also changed to fit the demands of contemporary conditions, such that the differential policing and mass incarceration of minority populations within liberal democracies such as Australia, New Zealand, the USA, and the UK have been made acceptable (Stanley and McCulloch 2011). In these instances, it is impossible to argue that substantive equality has been achieved. Instead the progressive eradication of the slave trade through the eighteenth and nineteenth centuries, and an end to formal racial discrimination, has been replaced by new forms of economic, political, and social control.

It is also apparent that what might appear to be a progressive response to resistance strategies can turn out to be merely symbolic. For instance, resisters will often draw upon the rule of law and the judiciary to bring state officials to account. Yet, using state institutions like the judiciary to oppose state crime simultaneously confirms, or shores up, the institutional basis of the state. It can reinforce the parameters of "the law" as well as enhance the legitimacy of the state; such that they present the state as a "protective body" in periods when many state actions suggest otherwise. Sometimes seemingly accommodating state responses to resistance movements or actions can be counter-productive or they can reiterate dominant norms. For example, Bibbings (2009) highlights how new, apparently liberal laws with regard to homosexuality ultimately reinforced hegemonic heterosexual norms (by continuing to advance homosexual activities as being "different" or "deviant" and by increasing the monitoring and control of sexual practices). Nonetheless, the advances in law provided an opportunity for the state to purport "to have moved beyond heteronormativity and homophobia" (Bibbings 2009, p. 42). In addition, legal advances – set against a discourse of progress and equality – made further resistance difficult to pursue as those who continued to complain were labeled as "radicals" or "extremists."

The symbolism of responses to resistant acts is, therefore, an important consideration. Progressive state responses, which might at first glance be regarded as a successful outcome to resistant acts, can just present an opportunity for the state to strengthen legitimacy and highlight that they are governing by consent. It can serve to mystify the state – in that the state (and state actors) can appear as being able to be effectively challenged or changed or "talked back to" (Coleman et al. 2009, p. 15). Where states move to accommodate, incorporate, and adjust to resistance, the solidarity of resistance movements frequently comes under challenge as those within reach different conclusions about the meaning of state actions and the need for continued resistance.

It may, however, be too narrow a frame to consider effective or successful resistance only as collective actions or those that impact discernibly on the state. Successful forms of resistance can also grow unremarked or unnoticed from a whole range of small, quiet "unsuccessful" actions carried out by unknown individuals or small numbers of people (Stanley and McCulloch 2011). This idea of the many small actions over time leading to a more just outcome in the face of state crime is captured in the title of the anthem of the Australian indigenous land rights movement: "from little things big things grow." There are circumstances in which resistance may not be seen, acknowledged, or recognized by a broader audience. It may operate instead as "offstage dissent" (Scott 1990). It may be too dangerous and repressive for individuals and groups to make open claims of victimization and demands for justice. Yet, in such circumstances, individuals and groups often continue to engage in actions that create a "hidden transcript" against the dominant force (Scott 1990). Such "transcripts" may be made up of simple acts – lighting a candle to remember someone killed by state forces or marking a stone where an atrocity occurred. While these acts may not have a broad audience and may even be done in secret, they demonstrate agency within contexts of limited power and may, in ways that are (deliberately) difficult for those not directly involved to comprehend, keep the will towards resistance alive. These struggles against power may, as Milan Kundera (1978, p. 3) maintained, be "the struggle of memory against forgetting."

There are small, as well as significant, victories to be had in resisting state crime. As a final note here, it must also be acknowledged that even if a resistance initiative is seen to have failed overall, that "does not mean that they failed to influence, mitigate or even constrain state policies and actions related to state crime" (Friedrichs 2010, p. 9). While resultant changes may often not meet the expectations of those who resist, they may still have challenged state policies and actions. In addition, the lessons learned through "unsuccessful" campaigns may form the foundation for future "success."

How Might Criminologists Resist State Crime?

From the above discussions, it is clear that state crime scholars have begun to make some progress on the elements and meanings of resistance. Criminologists have also begun to identify, and analyze, the different routes that criminologists might take in progressing a resistant approach. For while a stance on resistance might at first appear to be romantic or naïve, the costs of not challenging state crimes appear to be too high a price to pay (Friedrichs 2010).

To this end, criminologists have ascertained several approaches that might be taken by future scholars. First, state crime academics might participate in a traditional public criminology in which they engage in public education through the remit of their publications and public speaking and work with resistance movements, including with state officials. This approach dovetails with "newsmaking criminology" in which scholars expose state crime issues. This form of public criminology is seen as having limited interactions, as being relatively passive and mainstream (Kramer 2010).

Second, criminologists might engage in an organic public criminology in which they work directly with activist groups or individuals in a process of mutual education. The impetus, here, is to craft political actions and policy choices through more direct activist action. This form of criminology is more visible, thick, and active and may directly challenge public understandings of "crime" and harm (Kramer 2010). This form of criminology also connects with the "idealist critical criminology" proposed by Coleman et al. (2009) who argue that criminologists can be pragmatic and actively influence policy. It is possible to challenge the state (and be outside the ideological terrain of state) and still engage with policy. For example, criminologists can engage with organizations such as Inquest (which campaigns against deaths in custody in the UK), as well as other counter-hegemonic groups, that stand "in opposition to the state's criminal justice agenda but still remain engaged with government in consultations, lobbying and policy work" (Coleman et al. 2009, p. 16).

Within these broad approaches to the state crime issue, criminologists have a number of further concerns and issues to attend to. For instance, there are questions of whether state crime research can ever really be funded or directed by government agencies. After all, such research can be expected to support the imperatives of the state and to consolidate the continued regulation of the powerless (Walters 2009). Alongside this, there remain questions about how criminological literature may be developed to refocus attention to state crime activities. Part of this may involve a substantive reframing of criminological language to include debates on nationalism, sovereignty, universal jurisdiction, legitimacy, human rights, cosmopolitanism, global justice movements, global governance, or sustainability (Friedrichs 2010). For many criminologists, this requires new exploration of crucial ideas and concepts that currently stand outside mainstream attention.

All of these identified routes and concerns revolve around an idea of strengthening the criminological role as "claims makers." It also signifies an academic move from being invisible, passive, or inactive towards being visible and active (Kramer 2010). As academics, paid by the state, criminologists remain capable (with their scholarly privilege) to engage in resistance. Yet, such resistant approaches are not without risks or difficulties. Academics who take such stands may face isolation, attack, or dismissal; they may find that they are negatively portrayed within academic or media circles; and they may struggle to find funding or to access data (Kramer 2010; Walters 2009). In addition, within an environment in which academics face multiple constraints (such that, in many countries, academics are faced with a particularly narrow band of what constitutes "valued knowledge" – in the form of funded research, knowledge disseminated in particular journals, or by particular publishers – and employment or career progression is dependent upon being an academic producing "valued knowledge"), it may be that such academic resistance is becoming increasingly difficult to uphold or engage with (Tombs and Whyte 2003). Even so, it remains the case that western academics usually hold a comfortable position from which to resist from. This places an onus on scholars to engage; as Cohen (2001, p. 40) points out, “Intellectuals who keep silent about what they know, who ignore the crimes that matter by moral standards, are even more morally culpable when society is free and open.” The choice not to speak against state crime in circumstances of freedom and safety is to use a term borrowed from the resistance movements of World War Two, to collaborate with those crimes.

Future Directions

Given the relative dearth of literature on this topic, there are many directions that criminologists might take in the future. Here, five potential aspects of state crime scholarship are explored in relation to resistance. First, as Stanley and McCulloch (2011) note, state crime scholars must be mindful to analyze both the egregious and the mundane. That is, while it is useful and necessary to analyze genocides, such as the Holocaust, it is also important to consider state crime and resistance outside of the most notorious acts of state violence. To understand state crime, as well as resistance, criminologists need to include the study of the everyday harms and violence that are embedded in the fabric of society. Focusing on state crimes and resistance exclusively in the context of the foremost exemplars of state crime, risks overlooking the pervasive state perpetrated violence and harms that occur as normalized, hidden, or denied aspects of society, including in democratic states.

Second, criminologists might also develop analyses of resistance in terms of affirming values, or *becoming* something, rather than acting against. That is, resistant acts may be about sustaining a way of life or invoking a competing claim of universalism (such as affirming and reproducing cultural identities). This idea is illustrated in Harry Blagg’s (2008) recent work on aboriginality and justice. He highlights how difficult it is to imagine “forms of resistance that do not seek to take over the state, or reproduce it, but actively seek to create distance from state structures and recreate traditional forms of order” (Blagg 2008, p. 37). While the state may be viewed, in an Althusserian sense, as the “site and stake” of struggles, some groups want to move beyond the state. In this regard, it is crucial to understand that as a result of history there are divergent meanings attached to state institutions. In addition, some groups hold “radically incommensurate interpretations” (Blagg 2008, p. 47) of past experiences. From this perspective, to celebrate and live by another universality may be a valued form of resistance. Such ways of being move beyond the usual dichotomies of cultural/social/political norms about life, so that ways of being or ways of resisting are called into question rather than being reemphasized and bolstered.

Third, and a related point, is that criminologists may become more mindful of how resistance strategies and practices are differentially experienced across groups. State crime literature has not yet thoroughly engaged with how state crimes are structurally experienced, and the literature on state crime and how it relates to structural relations of gender, “race,” age or class remain relatively undeveloped. Future scholars could therefore be more attentive to how state crimes, and attempts to bring states into line, “will be perceived and experienced differently” (White 2010, p. 57).

Fourth, criminologists could continue to develop analyses on why people and groups become attracted to certain resistance strategies and, simultaneously, how they are also dissuaded from participating. A key question to the latter point might be the following: Why do people not resist more to state crimes? Criminologists (such as Cohen 2001; Stanley 2009; White 2010) have already demonstrated that citizens are dissuaded from resisting as a consequence of many issues, such as compassion fatigue, an acceptance of certain forms of state violence, a culture of individualism such that they are less likely to aid “others,” concerns about their own criminalization or denigration, or a sense that any strategies are doomed to failure. The dangers that resisters and their “communities” might face are worthy of further examination, particularly in terms of how states engage in “counter-resistance” techniques.

Finally, and fifth, there is significant scope to question how resistance strategies impact upon state power. That is, do resistance strategies just embed state power by emboldening states to pursue more authoritarian measures or by providing states with an opportunity to enhance their own institutions, discourses, and status? For instance, as detailed above, while there have been a range of progressive reforms around a whole host of issues, including gender and "race" relations, they have never been secure or adequate (such that essential power relations have remained unchanged). Within such circumstances, it is unclear whether resistance to state crime can be successfully undertaken by strategies that use state institutions and norms – the use of law to ban slavery and the institution of legal regimes of equal opportunity have not dismantled the economic system in which slavery and racialized punishment continue to operate, yet it has emboldened legal and state systems, at international and national levels, to present the appearance that something has been, and is being, done.

Resistance to state crime is emerging as an important aspect of state crime scholarship. Naming, describing, analyzing, and theorizing state crime has contributed to criminology by moving it beyond state-centered notions of harm and security. Harms and violence committed by individuals pale into insignificance compared to those committed by states. Incorporating the study of resistance into state crime studies and into the work of state crime scholars marks an important advance in the field. The challenge of the future is to build on the emerging insights on state crime and resistance so, as state crime scholars, we can confidently claim to have moved beyond being bystanders to participants in the struggle against state crime.

Recommended Reading and References

Bales K (2004) Disposable people: new slavery in the global economy. University of California Press, Berkeley

Bibbings L (2009) The heterostate: hegemonic heterosexuality and state power. In: Coleman R, Sim J, Tombs S, Whyte D (eds) State, power, crime. Sage, London

Blagg H (2008) Crime, aboriginality and the decolonisation of justice. Hawkins Press, Sydney

Cohen S (2001) States of denial: knowing about atrocities and suffering. Polity, Cambridge

Coleman R, Sim J, Tombs S, Whyte D (2009) Introduction: state, power, crime. In: Coleman R, Sim J, Tombs S, Whyte D (eds) State, power, crime. Sage, London

Friedrichs D (2010) On resisting state crime: conceptual and contextual issues. Soc Justice 36(3):4–27

Giroux H (1983) Theory and resistance in education: a pedagogy for the opposition. Heinemann, London

Goldsmith A (2010) Policing's new visibility. Brit J Criminol 50:914–934

Green P, Ward T (2000) State crime, human rights, and the limits of criminology. Soc Justice 27(1):101–115

Green P, Ward T (2004) State crime. Pluto, London

Green P, Ward T (2009) Violence and the state. In: Coleman R, Sim J, Tombs S, Whyte D (eds) State, power, crime. Sage, London

Grewcock M (2010) Border crimes. Federation Press, Sydney

Hagan J, Rymond-Richmond W, Parker P (2005) The criminology of genocide: the death and rape of Darfur. Criminology 43(3):525–561

Huggins M (2010) Modern institutionalized torture as state-organized crime. In: Chambliss W, Michalowski R, Kramer R (eds) State crime in the global age. Cullompton, Willan

Jamieson R, McEvoy K (2005) State crime by proxy and juridical othering. Brit J Criminol 45(4):504–527

Kramer R (2010) Resisting the bombing of civilians: challenges from a public criminology of state crime. Soc Justice 36(3):78–99

Kundera M (1978) The book of laughter and forgetting. Faber and Faber, London

McCulloch J, Sentas V (2006) The killing of jean Charles de Menezes: hyper-militarism in the Neo liberal economic free fire zone. Soc Justice 33(4):92–106

Pickering S (2002) Women, policing and resistance in Northern Ireland. Beyond the Pale, Belfast

Rothe D (2010) Exploring post-resistance to state criminality: *realpolitik* versus ideology. Soc Justice 36(3):111–121

Scott J (1990) Domination and the arts of resistance: hidden transcripts. Yale University Press, New Haven

Sivanandan A (1990) Communities of resistance: writings on black struggles for socialism. Verso, London

Stanley E (2009) Torture, truth and justice: the case of Timor-Leste. Routledge, London

Stanley E, McCulloch J (2011) State crime and resistance. International State Crime Initiative. http://statecrime.org/index.php/crime-studies-resources/gg-resistance/247

Stanley E, McCulloch J (eds) (2013) State crime and resistance. Routledge, London

Stephan M, Chenoweth E (2008) Why civil resistance works: the strategic logic of nonviolent conflict. Int Secur 33(1):7–44

Tombs S, Whyte D (2003) Scrutinizing the powerful: crime, contemporary political economy, and critical social research. In: Tombs S, Whyte D (eds) Unmasking the crimes of the powerful: scrutinizing states and corporations. Peter Lang, New York

Urquhart B (2011) Revolution without violence? vol LVIII. New York Review of Books, New York, pp 37–40, 10 March

Walters R (2009) The state, knowledge production and criminology. In: Coleman R, Sim J, Tombs S, Whyte D (eds) State, power, crime. Sage, London

White R (2008) Depleted uranium, state crime and the politics of knowing. Theor Criminol 12(1):31–54

White R (2010) Environmental victims and resistance to state crime through transnational activism. Soc Justice 36(3):46–60

Responses to Domestic Violence in India

Sesha Kethineni
Department of Criminal Justice Sciences, Illinois State University, IL, USA

Synonyms

Regulate violence

Overview

To understand the formal and informal systems of crime control, including domestic violence, in India, it is important to review the historical, social, political, and cultural perspectives that shaped these systems. The formal system of crime control relies on official agencies and laws to maintain order, deter crime, and respond to criminal activity. Informal crime control depends on families, peers, neighbors, and the communities to settle disputes, maintain peace, and promote lawful behavior. This topic covers domestic violence from historical and contemporary perspectives, specifically focusing on formal and informal methods of addressing the domestic violence in India.

Formal and Informal Systems of Addressing Domestic Violence

During medieval times, formal systems of control were enforced by various dynasties (Unnithan 2005). The kings administered justice under the Hindu religious concept of *dharma* (rules of right conduct), taking into account the circumstances under which the crime was committed. *Dharma* was supposed to sustain society, maintain social order, and ensure well-being and progress of humanity (Hayden 1984). To live in peace, proper education and training must be imparted and individuals must be taught not to transgress the freedom and liberty of others. The religious scripts also included threats of severe punishment for transgressors in their current life and the afterlife. For example, during the Mauryas Dynasty (321–185 B.C.E.) people were sentenced to punishments such as mutilation and the death penalty for minor offenses. In the second and third century C.E., Manu, an important Hindu jurist, drafted the *Dharmasastra* code, which was called *Manusmriti*. The laws of Manu were based on jurisprudence, philosophy, religion, and *varna* (caste), and described how life should be lived by lower caste, priests, kings, women, and men (Doniger and Smith 1991).

In 1500s, the Moguls conquered most of northern India. One of the most influential Mogul emperors, Shah Jahan, ruled India from 1628 to 1658 C.E. During the decline of Mogul Empire, the British East India Company arrived in India as a trading company and slowly gained control. Realizing that the East India Company was becoming a powerful virtual ruler in India, the British Crown took control of India in 1858 and ruled India until 1947. Under the British rule, although formal and informal systems of justice coexisted, the formal mechanisms were strengthened. The British used the approach of “divide and conquer” to rule: they rewarded those who supported them and punished those who sought independence or protested the British rule using the formal system of justice – police, laws, and courts. The formal system was not popular in rural areas because many people either did not understand the laws and procedures of the British or were unwilling to abide by them. Moreover, asking formal criminal justice agencies to settle disputes was not a common phenomenon, because resources such as transportation and access to lawyers were not readily available, especially for poor and disadvantaged populations. In addition, those who failed to seek reconciliation or dispute resolution, whether it was a dispute, a moral issue, or a crime, through informal justice system were shunned by the community. During this time, most of the disputes in rural communities were settled by the traditional caste *panchayats* (village councils).

The *panchayats* consisted of at least five elders, usually from Hindu upper castes, and are often referred to as *caste panchayats*. The caste hierarchy dictated the rules, format, and outcomes of the cases. The four main Hindu castes are *Brahman, Kshatriya, Vaisya*, and *Sudra*. *Brahmins* were at the top of the hierarchy and were considered arbiters of learning, teaching, and religion. The *Kshatriya*s were warriors and administrators. The *Vaisyas* were artisans and known for business ventures. The *Sudras* were farmers and peasants. The fifth class, which does not fall under the caste hierarchy, were once known as *Harijans* (untouchables) or, more recently, *Dalits* (oppressed people). The caste system continues today with many subtypes (*jatis*) in each caste.

After India gained independence from the British rule in 1947, the country moved towards a more formal system of crime control based on the British colonial model and English common law. The traditional (caste) *panchayats*, headed by village elders, were considered out of tune with the democratic processes. The Indian government created a criminal justice system with centralized control and enforcement. Although the police are organized at the state level, the high-ranking police officers (Indian Police Service) are recruited, trained, and appointed by the central (federal) government. In addition to the state-controlled police, the Indian government directed the states to establish elected *nyaya* (justice) *panchayats* as local government units in 1950 (Meschievitz and Galanter 1982).

Hayden (1984) argues that the *caste panchayat* system is different from the court-like *nyaya panchayat*. Unlike the *nyaya panchayats*, the *caste panchayats* take all considerations (e.g., the individual, place, and times) and not base decisions on predetermined rules and laws (Darrett 1978). In *caste panchayat* system, the determination of facts is not essential because the participants are already aware of what occurred. The *caste panchayat* considers known behavior and weighs its propriety (Hayden 1984), rather than labeling the act as criminal or deviant. Unlike the *caste panchayat* who applied indigenous norms, the *nyaya panchayats* relied on statutes, majority ruling, and members are selected by popular elections (Iyengar 2007). By the 1970s, these courts were considered failures due to declining caseloads and client dissatisfaction (Baxi and Galanter 1979). Scholars have asked why attempts to revive the informal justice system after the country's independence failed. Some claim it was due to deficiencies in funding and state control of legislative function, whereas others believe that the Indian elite is comfortable with the formal justice system and does not want a change (Baxi and Galanter 1979). In addition, caste rivalries, class difference, and growing egalitarianism and individualism disrupted communities, and contributed to their failure (Vincentnathan 1992). In order to promote indigenous character within the justice system and to provide an alternative to the Anglo-Saxon model of adjudication, the central government passed an amendment to the Constitution in 1992 giving the elected *panchayats* constitutional status and institutionalized them in a three-tiered system (except in states with a population of fewer than two million people): village level, block level, and district (county) level (*Panchayat* system in India 1995). In addition, because of a backlog of cases in the formal justice system, the government legislated for the establishment of *Lok Adalat* (people's courts) and Family Courts throughout India. The *Lok Adalats* are governed by Legal Services Authorities Act 39 of 1987 and are presided over by retired judges and private legal practitioners. These are court-linked mediation entities, sit on weekends, and hear a large number of gender violence cases (Relis 2011). Their main goal is to encourage parties to settle their disputes through mediation/conciliation. The Family Courts were established to handle the increasing number of marital disputes through speedy process using conciliation (Iyengar 2007).

Historical and Contemporary Views of Women

Some scholars believe that in ancient India women enjoyed equal rights along with men (Mishra 2006), whereas others hold contrasting views (Pruthi et al. 2001). During the early Vedic period (1500–500 B.C.E.), women were allowed to obtain an education. Around 500 B.C.E., following the Islamic invasion and Moghul Empire, the status of women began to change. Women and children faced religious and social restrictions. During medieval times, the practice of child marriages, *Sati* (self-immolation of women following her husband's death), and a ban on widows remarrying became part of social life in some communities. Ancient Hindu religious scriptures (i.e., the code of Manu) illustrate the secondary status of females – they have to be subservient to their fathers in childhood, to their husband after marriage, and to their sons after the death of their husband. With Muslim conquest of India, the practice of wearing a *purdah* (veil) became prevalent among Muslim women. In some parts of India, the *Devadasi* (temple women) were exploited by wealthy, upper-class men. During the colonial era, there was a cultural resurgence. Notable social reformers stood firm against all forms of social bigotry, orthodoxy, idol worship, and encouraged Western/English education. Others brought public awareness of social evils such as child marriages, polygamy, and child widows, and fought for the rights of women. These efforts resulted in widow remarriages, and prohibition of child marriages and bigamy.

Although India's independence has brought some social reforms for women and children, traditional social evils such as dowry and child marriages continue to the present day. According to the United Nations Population Fund Report, many married women in India suffer from domestic violence (Bhat 2006). "Dowry" is a cultural practice in many Indian communities. It involves giving property or valuables to the bridegroom or his family by the bride's family in consideration of marriage.

The traditional definition of domestic violence includes physical, sexual, or psychological abuse, or threats of violence inflicted by an intimate partner related to the victim through marriage, family relationship, or an acquaintanceship (Sahoo and Pradhan, n.d., p. 1). In the Indian context, the definition also includes dowry harassment, dowry death, and economic abuse. The most common causes of domestic violence include dissatisfaction with the dowry, wives refusing to have sex with their husbands, alcoholism of the spouse, the wife's inability to give birth to a male child, to minor acts such as not cooking food properly, not taking care of in-laws, or being too social with male coworkers at workplace (Kumar 2010).

Domestic Violence and Formal Justice System

Beginning in 1980s, women's organizations and other nongovernmental organizations (NGOs) made the public aware of the magnitude of the problem, the ineffectiveness of existing laws, and lack of enforcement. The prevalence of domestic violence and women's fight for equality and social justice prompted the legislators and policy makers to recognize that violence against women is not only a crime but also a violation of human rights. As a result, legislators have revised old laws and drafted new ones to criminalize offenders and provide more protection for women. These include amendments to the Indian Penal Code (IPC), the Indian Evidence Act, and the Dowry Prohibition Act, as well as changes in police procedures (Ranchhoddas and Thakore 1987).

The IPC was amended in 1983 and 1986 to specifically define crimes dealing with marital violence and abuse. Section 498A of IPC recognizes domestic violence as a cognizable (i.e., felony) offense and allows police officers to arrest the perpetrator without a warrant. Under this section, if the victim or the relatives of the victim complain to the police that the victim suffered "cruelty by husband or relatives of husband," the police have no option but to take action. The perpetrator can face imprisonment for a term in prison for a minimum of 3 years, as well as a fine. The definition of "cruelty" is broad and covers not only serious injury or bodily harm, or danger to life, limb, or physical health, but also includes mental health, harassment, and emotional torture through verbal abuse. In India, "harassment" includes such acts as coercing the wife or relatives of the wife, making an unlawful demand regarding property or valuables by the husband or his family members (generally in-laws). The law not only discourages the taking of a dowry but also the giving of a dowry. Both offenses are punishable with imprisonment for a term in prison of 6 months to 2 years, a fine, or the amount of the dowry paid or received. The law also allows women to retrieve their dowries under Section 406 of IPC (Ranchhoddas and Thakore 1987).

More important, section 304B, which was added to the IPC, deals with a new category of crime: "dowry death." This section states that if the death of a woman is caused by burns or bodily injury or occurs under unusual circumstances within 7 years of the marriage, and it is shown that prior to her death she was subjected to cruelty or harassment by her husband or his family, such a death would be considered dowry death. In such cases, the dead woman's husband or his family members would be deemed responsible for her death and shall be punishable with imprisonment for a term in prison of 7 years to "life" (generally 20 years). The inclusion of dowry death in the IPC brought about a change in the evidentiary standards in dowry death cases and a new section (113b) was added to the Indian Evidence Act. The previous section, 113(a), states that when a married woman commits suicide within 7 years of marriage and it is shown that her husband or his family members were cruel to her prior to her death, then it is presumed that her husband or his relatives aided in her suicide. The new section (113b) shifted the burden of proof to the accused, who must show why he should not be charged with the crime (Lawyers Collective Women's Rights Initiative 2010).

In addition to IPC and the Indian Evidence Act, the Dowry Prohibition Act of 1961 was amended in 1984 and 1986. These amendments made giving or receiving dowry a cognizable (felony) offense. Like dowry death cases, the burden of proof is on the person accused of taking or demanding a dowry: he must prove otherwise. A more recent piece of legislation, the Protection of Women from Domestic Violence Act (PWDVA 2005), allows civil remedies, such as a woman's right to reside in her matrimonial home or shared home. The Act recognized domestic violence as a violation of human rights. It required states to appoint regulatory authorities, such as Protection Officers (POs) and NGOs, as service providers. The Act meant that complaints of domestic violence had to be filed with the magistrates. In addition, it prohibited police officers from providing mediation in domestic violence cases (Maya 2006).

To better understand the nature and extent of violence against women, the states are required to document crimes against women under a special category. An annual report is published by the National Crime Records Bureau (NCRB). The report includes IPC offenses such as rape, kidnapping and abduction, dowry deaths, mental and physical torture, sexual molestation, sexual harassment, and trafficking of girls (NCRB 2006). In addition to IPC offenses, the NCRB also documents gender-specific offenses under Special and Local Laws (SLL). These laws include the Immoral Traffic (Prevention) Act (1956); the aforementioned Dowry Prohibition Act (1961); the Child Marriage Restraint (Amendment) Act (1979); the Indecent Representation of Women (Prohibition) Act (1986); and the Commission of *Sati* (Prevention) Act (1987).

Many state governments, by a special order, established All Women Police Stations. These police stations are staffed by women police officers (taken from all level, line staff to supervisory levels). They have the authority to handle both serious and nonserious incidents of domestic violence (e.g., family disputes). AWPS have been established in both urban and rural areas to make police stations "more approachable and less intimidating to women" (Poonacha and Pandey 1999, p. 30). It is presumed that women police officers will have a better understanding of women's needs. The police may register a cognizable (i.e., felony) case either by themselves or on the basis of a victim's complaint. These cases are registered in the First Information Report (FIR). FIR cases include dowry death, cruelty by husband or relatives, rape, intimidation and harassment of women, and demanding or receiving dowry. The noncognizable offenses include less serious offenses that are entered in the Community Service Register (CSR). CSR cases include minor dowry-related cases, drunkenness by husband, adultery, bigamy, family quarrel, seeking the custody of children, false promise of marriage, and desertion by husband. Officers try to resolve CSR cases through counseling or mediation. If the case is not resolved, the police may refer it to a family court or an appropriate court or register a criminal complaint. With the new law (PWDVA 2005), the women police officers must report the case to the magistrate and are not allowed to mediate domestic violence cases.

Formal Court Structure

The court system in India is a unified court structure in which the Supreme Court is the highest court in the country, followed by the high courts (i.e., the highest court at the state level), and other appellate and trial courts at the state level. Under the PWDVA (2005), a woman can avail herself of free legal services. The court of the chief judicial magistrate of the first class or the metropolitan magistrate (lower courts) have jurisdiction over domestic violence cases. The complainant or a protection officer (appointed under the Act) may present an application to the magistrate seeking relief. The order may include payment of compensation or damages for the injuries. A woman may also file for a separate civil suit for such compensation or damages. If the award is given in civil court, then the amount is deducted from the overall damages. The magistrate may also order the parties to receive counseling. The Protection Office must maintain a list of all service providers that offer legal aid or counseling, shelter, and medical assistance within the jurisdiction of the magistrate. A medical report of the injuries must be submitted to the police stations and the magistrate. The courts may also issue a protective order at the request of the woman, which can be modified or revoked if the circumstances of the case changes. The magistrate may also require the respondent to execute a bond with or without sureties. While passing the order, the court may direct the nearest police station to provide protection to the complainant or the person making the complaint on behalf of the victim. The protective provisions enumerated in the Act are comparable to those found in similar laws in the United States.

Those who are seeking redress through the formal justice system are often women from the middle and upper classes. The PWDVA (2005) was hailed as the most important legislation providing protection for women against physical, mental, economic abuse/violence. Some criticized the law as (1) being gender biased, (2) heavily favoring women with the potential for misuse by women, and (3) having too broad a definition of domestic violence. First, the law assumes that only women are victims and only they have rights under the law. A man who is a victim of domestic violence has no rights under the law. Second, such exclusive power will lend itself to misuse by women who want to teach their male relatives a lesson. Such trends were noticed in the case of anti-dowry law (Section 498a). A third flaw is that the definition of domestic violence includes insults and name calling as abuse, which are extremely subjective. Considering these frivolous acts, domestic violence may increase litigation and hurt the foundation of marriages. Those who opposed the law claim that the law in the present form is grossly inadequate (Deshpande 2008).

Recent Innovations in the Informal Justice System

Women in rural India, especially in western and northern states, formed the *nari adalats* (women courts). The nari adalat, an innovation by the Mahila Samakhya, is a grassroot effort by women collectives (Sanghas or groups) to address issues such as violence against women. The Mahila Samakhya is an autonomous program initiated in 1989 by the Ministry of Human Resources Development in 10 states with a primary purpose of empowering women, especially from socially and economically marginalized groups. They operate at the block and village levels and the members are trained on legal issues, discrimination of women, and women's rights. As of 2010, there were 184 nari adalats in 9 states and presided over 6,000 cases (Purushothaman 2011). Because of their popularity and their success, they are often approached by the caste panchayat members if the case involves domestic violence.

These courts consist of a jury of village women and grassroot-level workers who meet regularly to address women's issues, including domestic violence. In some villages, a group of five women, selected by the villagers, sit in a circle as part of the informal court. Many of these women have little education, but they have learned about the law and understand its implications and limitations. Although their caseload concerns women seeking justice, sometimes they hear grievances filed by men as well. Some of the cases they hear include divorce; disputes between women and their in-laws; complaints about husbands' infidelity, domestic violence, rape, dowry harassment, widows, and elderly (Iyengar 2007). They not only mediate cases and propose mutually acceptable decisions, but also serve as a monitoring body to ensure that the decision is implemented. These informal courts have been widely accepted by communities (Sharma 2000, p. 7). This unique form of informal justice is carried out quietly in many villages in states such as Gujarat "through a process of social censure, cajoling, argument and persuasion" (Sharma 2000, 7).

Some examples of cases decided by these courts follow. In one village, the *adalat* (informal court) met on the verandah of the village government office. Although space was offered inside the office, the women preferred to have the court outside so that it would be more visible to the public. An elderly woman announces the start of the sessions. The parties include a young man and a woman with a young girl sitting in a circle facing the members of the *adalat*, which consists of the elderly woman, a facilitator from a local women-empowerment program, and women from various community organizations. The young man asked for divorce because his wife left him 2 years ago to live with her parents and he does not want her back. When the *adalat* women asked why his wife left him to be with her parents, he did not respond. The facilitator states that if he wants a divorce, he must pay his wife INR (Indian Rupees) 500 (about USD$10.50) a month for 7 years, and INR8,000 (USD$170) for his daughter's education, for a total of INR50,000 (USD $1,063). He was told that he has 1 week to decide and, if he is willing to pay, he can get divorce. The second case involved dowry harassment by in-laws. The woman complains that her husband and her in-laws are insisting that her parents should buy an electric fan. Moreover, she states that she brought considerable amount of dowry into her marital home. The court told the husband, "Don't look at other people's palaces and in the process destroy your own house," and "Why don't you earn enough to buy yourself a fan?" (Sharma 2000, p. 16). The husband and his family agreed and provided a written assurance that they will not harass her any more by signing on a legal, stamped paper.

In another village, the *adalat* meets under a tree on the grounds of the government office. In a case heard at this *adalat*, a young woman complains that her husband has been unfaithful and has had affairs with her sister and sister-in-law. Her husband has thrown her out of the house and now wants a divorce. The accused replies that the women came after him. He then shouts at the adalat that it takes the side of the woman and stomps off. After a while, when tempers have cooled off, he returns and tells the court that he changed his mind because he cannot pay maintenance. The wife demands that she will return to him only if he asks for her forgiveness, which he does.

In addition to these traditional courts, recently Oxfam India, in collaboration with its partner organization, FARR (Friends Association for Rural Reconstruction) and other nonprofit organizations, conducted state-level public hearings for survivors of domestic violence in March 2011. The Oxfam India is a right-based organization that fights poverty and injustice in partnership with local NGOs. Among its five aims – the right to a sustainable livelihood, the right to basic social services, the right to life and security, the right to be heard, and the right to equality – the last one addresses gender and diversity issues. In addition, it aims to fulfill its overarching vision, "rights to life with dignity for all," by empowering poor and marginalized populations. Its recent efforts targeted seven states (Assam, Bihar, Chhattisgarh, Jharkhand, Orissa, Uttar Pradesh, and Uttarakhand) and four social groups (*Dalits*, tribals, Muslims, and women) (Oxfam India, n.d.).

The Oxfam India hearings were supported by the Department of Women and Child Development and other civil society organizations. The panel members included lawyers from the State Legal Service Authority, representatives from the State Women's Commission, and distinguished feminists. The organization heard about 50 cases of domestic violence in various locations throughout the state. The primary reason for holding these hearings was to offer a speedy resolution of cases filed under the PWDVA (2005). Moreover, these hearings offered an opportunity for women, especially from tribal villages, who were victims of informal justice systems to be heard. In some instances, an informal justice system may not be favorable to tribal women – it may even be detrimental to them. These hearings provide a voice to the survivors of domestic violence, and the panel of experts can provide a quick resolution to their disputes (Oxfam India, n.d.). An example of a case that was decided by the panel involved a woman requesting that her valuables and dowry be returned to her. Her husband died after 2 months of marriage and her in-laws forced her to leave the marital house. She wanted her valuables and dowry back so that she could start a stationary store. A lawyer from the State Legal Service Authority assisted the victim in getting her valuables and dowry back from her father-in-law. Several other women benefitted from these hearings (Das and Bandi, n.d.).

Discussion

The laws dealing with domestic violence (Section 113[a] of the Indian Evidence Act, Section 498[a] and 304[b] of the IPC, and the Dowry Prohibition Act [1961; as amended in 1986], and the PWDVA [2005]) have had limited impact on the timely disposition of cases. For example, Amnesty International reports that an analysis of court decisions in one particular district of Maharashtra showed that only 2.2 % of cases filed under 498(a) of the IPC during 1990–1996 ended in convictions (Mitra 1999). The PWDVA (2005), which was hailed as a landmark victory for the women's movement in India, faced with problems such as lack of enforcement mechanisms to implement the provisions of the Act. For example, many women who sought protection under the Act were turned away by the courts (Maya 2006) because several states failed to appoint protection officers.

A comprehensive study conducted by Relis (2011) in India covering eight states, and included interviews with victims, the accused, family members, lawyers from both sides, judges, judge-mediators, nonstates justice arbitrators (*panchayats*), court-linked mediators (*lok adalats*) in lower and high court, and nonstate women's courts (*nari adalats*), public interest lawyers, and human rights commission members, found that victims have little to no education regarding human rights. More educated women know about the concept of human rights, but still lacked information about specifics. Her study also found that less educated victims of violence preferred quasi-legal justice systems and wanted to resolve their cases through compromise so that they could return to their joint families. Victims wanted to maintain the honor of their families of origin. She explains that individuals in non-Western cultures define their identity by their membership in a group or community, not as anonymous entities. Her study also found that the quasi-legal arbitrators (e.g., women's courts) used human rights' principles in gender violence cases far more than lower courts. In order to effectively use informal justice mechanisms (e.g., *nari adalat* or women's courts), communities must have similar occupational experiences and common perspectives that bind them together (Vincentnathan 1992). Unlike the traditional *caste panchayats*, the newly emerged *nari adalats* have representatives from different communities. These *adalats* are referred to as "social justice forums largely of women, for women, and by women" (Purushothaman 2011, p. 52). These adalats have been very successful because their goal is to protect the woman's rights and facilitate a change so that she could live a violent-free live.

In order for these informal control mechanisms to work effectively, the people in the communities must have collective orientation through occupation, caste, or religion. In tribal areas of India, NGOs are working with village *panchayats* by providing training to tribal people, including women, about the various issues undertaken by the *panchayats* (Kumar, n.d.). A more effective way to deal with violence against women is to strengthen the informal system of justice mechanisms for less serious offenses, increase awareness of domestic violence, and improve the performance of formal justice systems in serious cases through timely disposal of cases, enforcement of court orders, and other forms of assistance to the victims.

Related Entries

▶ Informal Social Control
▶ Police Family Violence Services
▶ Restorative Justice and Practice
▶ Shaming in Asian Societies
▶ Surveys on Violence Against Women

Recommended Reading and References

Baxi U, Galanter M (1979) Panchayat justice: an Indian experiment in legal access. In: Cappellett M, Barth B (eds) Access to justice: emerging issues and perspectives. Sijthoff and Noordhoff, Netherlands, pp 341–386

Bhat D (2006) India's landmark domestic abuse law comes into effect. The Sunday Times, 6 Oct 2006. http://www.timesonline.co.uk/tol/news/world/asia/article614140.ece. Accessed 20 June 2011

Darrett JDM (1978) The death of a marriage law: epitaph for the Rishis. Vikas, Delhi

Das R, Bandi SS (n.d.) Oxfam India facilitates justice for survivors of domestic violence in Odisha. http://www.oxfamindia.org/content/oxfam-india-facilitates-justice-survivors-domestic-violence-odisha. Accessed 10 July 2011

Deshpande B (2008) Review of the Domestic Violence Act and other related laws. Save Family Foundation. http://www.savefamily.org/index.php?Itemid=63&option=com_content. Accessed 1 July 2011

Doniger W, Smith BK (1991) The laws of Manu. Penguin, London

Hayden RM (1984) A note on caste panchayats and government courts in India: different kinds of stages for different kinds of performances. J Legal Pluralism 22:43–52

Indian Evidence Act (1872) Universal Law Publishing, Delhi

International Center for Research on Women (ICRW) (2001) Domestic violence in India II: exploring strategies, promoting dialogue. ICRW, Washington, DC. http://www.icrw.org/docs/indiainfobulletin.pdf. Accessed 2 July 2011

Iyengar S (2007) Case study: A study of nari adalats (women's courts) and caste panchayats in Gujarat. United Nations Development Programme, Bangkok. http://regionalcentrebangkok.undp.or.th/practices/governance/a2j/docs/CaseStudy-05-India-NariAdalat.pdf. Accessed 30 July 2011

Jewkes R (2002) Intimate partner violence: causes and prevention. Lancet 359(9315):1423–1429

Kapur N (2001) Equality and the truth of Brahma. Violence Against Women 7(9):1069–1096

Koenig MA, Stephenson R, Ahmed S, Jejeebhoy SJ, Campbell J (2006) Individual and contextual determinants of domestic violence in North India. Am J Public Health 96(1):132–138

Kumar A (2010) Domestic violence in India: causes, consequences and remedies. YouthKiAwaaz, 7 Feb 2010. http://www.youthkiawaaz.com/2010/02/domestic-violence-in-india-causes-consequences-and-remedies-2/. Accessed 10 Mar 2011

Kumar MA (n.d.) Tribal women in local governance – a study of Andhra Pradesh in India. http://wgc.womensglobalconnection.org/conf06proceedings/Kumar%20A.–Tribal%20Women.pdf. Accessed 5 June 2011

Lawyers Collective Women's Rights Initiative (2010) Staying alive: fourth monitoring & evaluation report 2010 on the protection of women from Domestic Violence Act. http://www.unifem.org.in/PDF/Lawyer%20collective.pdf. Accessed 31 July 2011

Maya C (2006) India: domestic violence victims being turned away by courts. The Hindu, 20 Dec 2006. http://www.hindu.com/2006/12/20/stories/2006122004130500.htm. Accessed 13 July 2011

Meschievitz CS, Galanter M (1982) In search of nyaya panchayats: the politics of a moribund institution. In: Abel R (ed) The politics of informal justice: comparative studies. Academic, New York, pp 47–77

Mishra RC (2006) Women in India: towards gender equality. Authorspress, India

Mitra N (1999) Best practices among responses to domestic violence in Maharashtra and Madhya Pradesh. Women's Studies Unit, Tata Institute of Social Sciences. http://www.womenstudies.in/elib/dv/dv_best_practices.pdf. Accessed 1 July 2011

National Crime Records Bureau (2006) Crime in India, 2006. Ministry of Home Affairs, Government of India, New Delhi

Oxfam India (n.d.) About Oxfam India. http://www.oxfamindia.org/about. Accessed 30 June 2011

Panchayat system in India: India-local government (1995) http://www.indianchild.com/panchayats_system_india.htm. Accessed 15 Aug 2011

Panda P, Agarwal B (2005) Marital violence, human development and women's property status in India. World Dev 33(5):823–850

Poonacha V, Pandey D (1999) Responses to domestic violence in the states of Karnataka & Gujarat. Research Centre for Women's Studies, SNDT Women's University, Mumbai

Protection of Women from Domestic Violence Act (PWDVA) (2005) Professional Book Publishers, New Delhi

Pruthi RK, Devi R, Pruthi R (2001) Status and position of women: in ancient, medieval and modern India. Vedam books, New Delhi

Purushothaman S (2011) The nari adalat: A grassroots response to violence and injustice against women. Innovations towards education for empowerment, grassroots women's movement. Ministry of Human Resource and Development Department of Higher Education, Government of India, New Delhi

Ranchhoddas T, Thakore DK (1987) Indian penal code, 1860, 26th edn. Wadhwa and Company, Nagpur

Relis T (2011) Human rights and southern realities. Human Rights Quart 33:509–551

Sahoo H, Pradhan MR (n.d.) Domestic violence in India: an empirical analysis. School of Social Sciences Jawaharlal Nehru University, Centre for the Study of Regional Development, New Delhi

Sharma K (2000) Indian women pioneer informal justice courts. We-news, 30 Nov 2000. http://oldsite.womensenews.org/article.cfm/dyn/aid/357/context/archive. Accessed 11 July 2011

Unnithan NP (2005) Hindu legal traditions. In: Wright R, Miller J (eds) Encyclopedia of criminology. Routledge, London

Vincentnathan SG (1992) The social construction of order and disorder and their outcomes in two South Indian communities. J Legal Pluralism 32:65–102

Restitution as a Penal Aim

Michael M. O'Hear
Marquette University Law School,
Milwaukee, WI, USA

Overview

Across the past century, restitution has become an increasingly prominent and formalized aspect of criminal sentencing. When included in a sentence, an award of restitution requires an offender to pay compensation to her victim. Restitution is most commonly justified as a way to provide assistance to victims more quickly and less expensively than would be possible through a civil lawsuit. There are concerns, however, that restitution blurs the line between crimes and torts, and in various practical and symbolic ways may undermine the traditionally recognized objectives of the criminal-justice system.

Restitution advocates have responded in two quite different veins. Some have argued that restitution not only furthers the end of victim compensation, but also serves more conventional criminal-justice objectives, such as offender rehabilitation and atonement. Another, more radical response is that the traditional approaches to criminal punishment are fundamentally flawed, and should be supplanted by victim compensation as the system's overriding objective. The latter response has attracted much interest in the scholarly literature, but has yet to gain much support among policymakers.

As policymakers have sought to enhance victims' restitution rights, a variety of practical problems have emerged. Many of these problems seem to turn on the same basic question: To what extent should restitution rules seek to replicate in the criminal system the same damages awards that could be obtained through tort litigation in the civil system? Different jurisdictions have answered this question in quite different ways, which may reflect underlying uncertainties as to whether and how victim compensation can be reconciled with the traditional processes and purposes of criminal punishment.

Fundamentals

Definition of Restitution

Restitution refers to compensation paid by an offender to a victim. More specifically, as understood here, restitution refers to a payment ordered by a court in connection with the resolution of a criminal case. Functionally, a restitution award has many similarities to a damages award in a civil case. Indeed, the term "restitution" can also be used to refer to a type of civil remedy. Here, however, the term is used only as it relates to criminal litigation.

Restitution differs from other sorts of monetary awards in criminal cases, such as fines and assessments of court costs, inasmuch as restitution is designed to compensate a victim for an injury. The amount of the payment is determined by reference to what is necessary to repair or offset some or all of the victim's losses, and the intended beneficiary is the victim herself.

Although restitution normally involves a discrete victim or set of victims with individualized injuries, it is also possible to conceptualize an award of restitution based on a more diffuse, shared injury, such as the psychological distress experienced by all of the members of a community when a frightening crime occurs in their middle. Likewise, although restitution normally involves an award of money or other property, it is also possible to imagine a restitutionary payment in the form of services. Court-ordered community service might thus be seen as having a restitutionary character.

Formalization of Restitution as an Aspect of American Criminal Procedure

Since the late nineteenth century, restitution has become a progressively more routine and formalized aspect of criminal procedure in the United States. The rise of restitution was initially linked to the rise of probation and suspended sentences (Harland 1982). Endowed with wide discretion in setting conditions for probation, judges often found it appropriate to require probationers to pay compensation to their victims as a condition of avoiding prison. For much of the twentieth century, restitution was awarded in such cases with little or no express statutory guidance.

In the latter decades of the twentieth century, American legislatures became increasingly active in promoting restitution. This resulted from a confluence of factors. The most important was likely the appearance of a politically powerful victims' rights movement. Reflecting a growing sensitivity to the situation of crime victims, most American states adopted victim-compensation laws between the 1960s and 1980s (Harland 1982). These compensation statutes, however, were not true restitution laws, as they contemplated a nonjudicial process administered by an administrative board, from which victims could obtain compensation even if their victimizers had not been apprehended and convicted. In any event, the laws varied widely in their eligibility limitations and maximum awards Holezel (1980)). Perhaps reflecting a sense of dissatisfaction with these compensation schemes, the victims' rights movement shifted its attention to achieving various reforms in court procedures, including establishing clearer rights to restitution.

Complementing the pro-restitution advocacy of the victims' rights movement, the 1970s and 1980s saw a breakdown of the rehabilitative paradigm in criminal justice and a new emphasis on holding offenders accountable for their crimes. This emphasis on accountability, in lieu of treatment, often took the form of longer, harsher prison sentences. To some, however, accountability implied not harshness per se, but rather the performance of certain apologetic or reparative actions by the offender through which he or she might achieve a measure of atonement and reconciliation with a victim or the wider community (Garvey 1999). This view of accountability, which found effective expression in many communities through the restorative-justice movement, was quite amenable to restitution

(Umbreit et al. 2005). Of course, advocates of greater harshness might also welcome greater use of restitution, so long as restitution serves as an *additional* imposition on the offender, and not as a substitute for other forms of hard treatment.

Reflecting these various considerations, by the early 2000s, about one-third of American states and the federal government had adopted mandatory restitution laws (American Bar Association 2004). For instance, the federal Mandatory Victims Restitution Act of 1996 (MVRA) requires judges to order restitution as part of the sentence imposed for most federal offenses in which there is a victim. The law specifies, moreover, that a restitution order must cover the full amount of the victim's injuries, regardless of the offender's ability to pay. Reflecting the change from permissive to mandatory restitution, the amount of "federal criminal debt" in the United States increased from about $6 billion to more than $50 billion in the first decade after the MVRA's enactment, with about $1.8 billion in restitution collected in 2007 alone (Dickman 2009).

The growing popularity of such laws in the United States might also be understood against the backdrop of the nation's "politics of mistrust" in the latter decades of the twentieth century and related legislative efforts to curtail judicial sentencing discretion (Zimring et al. 2001). Despite increasing legislative emphasis on restitution, however, concerns remain that judges and other criminal-justice professionals have not given sufficient priority to imposing and enforcing restitutionary obligations (American Bar Association 2004).

International Developments

The United States has hardly been alone in enhancing the role of restitution in criminal procedure in recent years, or more generally in displaying greater concern for the situation of crime victims. For instance, in 1985, the United Nations General Assembly adopted a "Declaration of Basic Principles of Justice for Victims of Crime and Abuse of Power," which, among other things, encouraged restitution and recommended that governments "should review their practices, regulations and laws to consider restitution as an available sentencing option in criminal cases."

Similarly, a 2004 directive from the Council of the European Union mandates that all member states "shall ensure that their national rules provide for the existence of a scheme on compensation to victims of violent intentional crimes committed in their respective territories, which guarantees fair and appropriate compensation to victims." Indeed, even before then, most members already had systems in place for victim compensation by the state (European Commission 2001). However, despite a 1985 recommendation by the Council of Europe's Committee of Ministers that criminal courts be given the authority to order restitution, the EU has stopped short of mandating this of its members. Rather, a 2001 "Framework Decision on the Standing of Victims in Criminal Proceedings" only requires that any compensation decisions in criminal proceedings be made within reasonable time limits.

Many individual nations do, in fact, incorporate restitution into their criminal processes, although this is far from universal. In England and Wales, for instance, the Powers of Criminal Courts (Sentencing) Act of 2000 requires courts to consider ordering compensation and to provide reasons when they do not do so (Doak 2008). Germany permits restitution as a condition of probation, and also makes available a hybrid process in which victims can attach a civil tort claim to criminal proceedings (Loffelmann 2006). As a final example, France also permits victims to claim compensation as part of the criminal proceedings, and has facilitated their ability to do so by, for instance, excusing them from having to appear in court in order to make a claim (Hodgson 2002).

Key Issues and Controversies

Justification of Restitution

Since crime victims typically have civil remedies available as a means to obtain compensation from offenders (at least as a theoretical matter), the creation of an additional compensation mechanism through the criminal process seems to require some justification beyond simply a desire to shift losses from victims to offenders. The most obvious such justification would be convenience for victims. If restitution is available, then the victim may be spared the trouble and expense of initiating her own civil proceeding, and instead take advantage of what may be a speedier award through procedures that may be less formal, less costly, and more favorable in a number of other respects than conventional civil procedures. Moreover, the court system, too, may reduce its costs by folding what would have been a separate civil case into the criminal proceeding. Indeed, even some defendants might have cause to welcome the apparent efficiencies.

Yet, there may be problems with viewing restitution as simply a more efficient way to effectuate recoveries that could have been had in the civil system. Liability standards for restitution may be structured differently than tort liability standards, or may be understood differently by criminal lawyers and judges than civil lawyers and judges. Different procedural rules (e.g., with respect to rules of evidence and right to a jury trial) may produce different results in restitution litigation than in civil tort litigation. Moreover, the prosecutor and criminal judge have an extraordinary source of leverage over many criminal defendants that is generally not available in civil litigation: If the prosecutor or judge takes a real interest in ensuring a certain level of restitution, the defendant who declines to cooperate may find herself facing a threat of increased incarceration. The restitution defendant may thus feel considerably more pressure than the civil defendant to “settle” the dispute on terms favorable to the victim. (On the other side, however, if the victim relies on a prosecutor to press her claim in a restitution proceeding, the victim may receive less zealous advocacy than a civil lawyer would provide, depending on how the prosecutor prioritizes restitution among the other disparate objectives of criminal prosecution.) More generally, the incarceration-related decisions that must be made in many criminal cases may overshadow restitution decisions and lead to more careless handling of claims than would be provided in a civil lawsuit. All of these considerations suggest that restitution outcomes may often deviate from civil litigation outcomes.

To the extent that victims fare worse in restitution, the outcomes may perhaps be justified on the basis of greater speed in recovery, reduced transaction costs, and the victim’s freedom to press claims in civil court in addition to or in lieu of restitution claims. It is not immediately clear, however, on what basis *greater* awards in restitution proceedings would be justified. This might require a theory to explain why conventional civil remedies are *substantively* inadequate for some types of crime victims, as opposed merely to being slower or more costly to obtain. In any event, it appears that the defense of restitution as simply an expedited form of civil recovery may be a bit more complicated than some proponents of restitution have made out.

If one line of criticism of restitution would focus on its potential to upset the balance struck in the *civil* system between the rights and interests of injurers and the injured, another line of criticism would highlight the potential of restitution to undermine the established priorities of the *criminal* system. The imposition of large financial burdens on some offenders, especially those whose economic situation is already precarious, may impair their prospects for rehabilitation and successful reintegration into the community. To the extent that incarceration time is traded off against restitution in plea negotiations or at sentencing, it is possible that the deterrence or incapacitation aims of criminal law may be compromised. (To be sure, the same trade-offs might occur even if restitution were not a formal part of the criminal process – e.g., a judge might, as a matter of discretion, treat voluntary reparations as a mitigating factor at sentencing – but the endorsement of restitution as an express objective of the criminal process might make such trade-offs more common.) The time and resources expended in litigating and adjudicating restitution claims in some cases might mean that offenders in other cases are able to avoid punishment or receive inadequate punishment. More generally, the increasing prominence of restitution in criminal litigation might create an impression that the criminal-justice system has been co-opted by private interests, which may undermine the ability of the system to pronounce condemnation on behalf of the community in a credible fashion. The system's credibility may also suffer if victims are commonly awarded large sums in restitution that offenders are not realistically able to pay.

In light of all of these concerns, a persuasive justification for restitution might have to show an organic connection between restitution and the particular ends of the criminal-justice system; simply providing a more convenient forum for advancing the ends of the civil-justice system may not be sufficient. Moreover, having a clearer sense of how restitution is capable of contributing to the purposes of punishment might help policymakers to sort out some of the difficult design questions, such as whether ability to pay should be taken into account in making a restitution award.

Proponents have indeed argued that restitution does serve conventional criminal-justice ends. For instance, while restitution obligations may in some respects impede rehabilitation, restitution may also be seen as "an effective rehabilitative penalty because it forces the defendant to confront, in concrete terms, the harm his or her actions have caused. Such a penalty will affect the defendant differently from a traditional fine, paid to the state as an abstract and impersonal entity, and often calculated without regard to the harm the defendant has caused" (*People v. Moser*, 50 Cal. App. 4th 130 (1996)). To be sure, this view of restitution as rehabilitative seems to rest on a very different sense of the rehabilitative project than the social-service model of earlier generations; here, we seem closer in spirit to atonement-based approaches, about which more will be said below.

Likewise, in the abstract, restitution seems capable of producing helpful deterrent effects, at least in the lower-level sorts of cases in which there is little risk of lengthy incarceration (the fear of which would presumably overshadow the threat of monetary penalties). On the other hand, even in lower-level cases, it is not immediately clear that restitution offers any marginal deterrence benefits when it is simply layered on top of civil remedies and criminal fines and forfeiture; for the added threat to matter, one would seemingly need a prospective offender who has both an ability to pay the full set of potential penalties and a nuanced understanding of law and legal process – not to mention a sufficient fear of getting caught such that the person's estimation of potential financial penalties would matter.

At first blush, restitution might also be seen as advancing the ends of retributive proportionality. And, indeed, it is hard to deny the "eye for an eye" appeal of imposing on an offender a sanction that is financially equivalent to the victim's losses. Yet, the thrust of modern retributive theory has generally been to emphasize *intended* (or at least culpably risked) harm, rather than *actual* harm, as the central determinant of desert (Moore 1997). Then, too, modern retributive theory has also tended to focus on *ordinal*, rather than *cardinal*, proportionality, that is, seeking to ensure that relatively more blameworthy offenders receive relatively harsher punishments, rather than striving for some absolute moral equivalence between the crime and the sentence (Von Hirsch 1998). Restitution may thus operate in ways that are significantly out of step with some views of retributive proportionality. For instance, a person convicted of a minor, unintentional traffic violation might find herself facing a ruinous restitution award if her violation by some unhappy chance resulted in a catastrophic accident – an award that might make her total punishment exceed those of other, more blameworthy offenders who intended serious harm but did not succeed in bringing it about. On the other hand, it may be easier to imagine such hypotheticals than it is to find them in practice. Moreover, concerns may be mitigated by limiting restitution recoveries to those harms for which the offender had some minimally required level of culpability. Yet, even at that, restitution may introduce some undesirable administrative complexity if one's basic approach to punishment focuses more on the *severity* of a sentence than on the *form* that it takes; a system of fines pursuant to standard guidelines, for instance, may be a much more straightforward manner of ensuring that the severity of each offender's sentence is precisely calibrated to the relative blameworthiness of her conduct.

Atonement-oriented approaches may be even more amenable to restitution. Advocates of such approaches tend to see proportionality in relatively flexible terms (Duff 2001). The aim of punishment on this view is not proportionality per se, but rather the communication of moral condemnation of the offense in the hope that the offender will come to repent what she has done, undertake reparative actions, and achieve reconciliation with the victim and the wider community. Punishment should thus take the form of a sort of secular penance, which will serve to focus the offender's attention on the offense and provide an opportunity to make reparations. (This approach to punishment overlaps considerably with restorative justice, although they might be distinguished to the extent that restorative justice – a label applied to a diverse set of approaches to crime and social conflict – is seen as an *alternative* to community condemnation, or as merely a set of procedures to be employed without regard to the character of the outcome (Duff 2005)).

Restitution would seem a quite appropriate – perhaps even a necessary – aspect of an atonement-oriented criminal process. The presentation and adjudication of restitution claims may serve to focus the offender's attention on the consequences of her conduct and why that conduct is viewed as blameworthy by the community. Moreover, the payment of restitution affords the offender with an opportunity to take responsibility for the offense and to establish conditions that may facilitate reconciliation. Indeed, on this view, it may be easier to see why restitution is superior to fines, for fines may appear confusingly similar in form to taxes or even licensing fees, and thereby detract from a sentence's ability to communicate moral condemnation (Duff 2001). To be sure, as with other retributive approaches, it might be necessary to limit restitution to culpably caused harms to the extent that harms without culpability are not fit for condemnation.

In sum, it does seem possible to justify restitution based on various familiar purposes of punishment. How persuasive the justification is may depend both on the specifics of the restitution law or practice at issue and on one's views about which purpose or purposes of punishment ought to receive greatest emphasis. Additionally, the foregoing discussion should make clear that there may be significant practical tensions between the conventional purposes of punishment and the hope that restitution may serve as a speedier and less costly alternative to civil litigation. For instance, the strict liability and simple negligence standards that govern some civil tort claims may be difficult to reconcile with at least some versions of retributivism. Incorporating these approaches to liability into the criminal process risks undermining the project of moral condemnation that many see as a central animating principle of the criminal-justice system.

Restitution as a "New Paradigm"

The previous discussion considered how restitution might be justified as one part of a criminal-justice system that maintains the conventionally recognized aims and penal options of Western criminal law in the modern period. Restitution, however, may instead be seen more radically as an alternative, fundamentally different approach to criminal justice – one that might wholly displace important aspects of current practice, such as the use of incarceration as a standard punishment for serious crime.

Perhaps most notably in this vein, the American legal scholar Randy Barnett proposed restitution as a "new paradigm of criminal justice" in a much discussed 1977 article. At the center of his proposal lay a privatized vision of criminal justice:

> The idea of restitution is actually quite simple. It views crime as an offense by one individual against the rights of another. The victim has suffered a loss. Justice consists of the culpable offender making good the loss he has caused. It calls for a complete refocusing of our image of crime. . . . Where once we saw an offense against society, we now see an offense against an individual victim. In a way, it is a common sense view of crime. *The robber did not rob society; he robbed the victim.* His debt, therefore, is not to society; it is to the victim. (Barnett 1977, pp. 287–88)

Barnett's proposal contemplated a radically reformed criminal process. Following conviction of a crime, an offender would be sentenced to a restitution period. The amount of the award would be determined in light of civil tort principles. If the offender were not able to make immediate payment in full, then the victim would have a legal claim against the offender's future wages. If the offender is unable to gain employment or is found to be "untrustworthy," the offender would be confined to an employment project. "This would be an industrial enterprise, preferably run by a private concern, which would produce actual goods or services" (289). Here, the offender would earn wages, and would secure her release upon satisfaction of the restitution award.

Barnett identified several advantages to his reform, which he termed "pure restitution." These included that victims would receive assistance, victims would have greater incentives to report crimes and appear at trial, offenders would have a mechanism for relieving their "guilt and anxiety," and taxpayers would be spared the expense of supporting idle prison inmates.

Barnett's proposal has much in common with various restorative-justice approaches, which are also often presented as a radical alternative, and not merely a supplement, to conventional criminal justice. Like Barnett, some restorativists seek to privatize our understanding of crime and punishment (Christie 1978). Many also share Barnett's prioritization of victim compensation and his distaste for straight incarceration as a useless, possibly even counterproductive, form of punishment. Additionally, Barnett contemplates routine, direct negotiation between offenders and victims, which is a centerpiece of many restorative-justice programs. Restorativists, however, would not be so narrowly focused on financial compensation as the subject of negotiation, but might instead emphasize to an equal or even greater extent apology, dialogue for its own sake, community service or service to the victim, ceremonial shaming and reintegration, and treatment. More generally, Barnett's approach runs counter to the communitarian and moralistic tendencies of many restorativists, who might object to the extent to which Barnett would marginalize the wider community, create a two-track penal system based on ability to pay, and monetize and commodify wrongdoing. (In addition to the presumably for-profit "employment projects" that will enforce restitution obligations, Barnett envisions the development of crime insurance markets and participation by insurance companies in the criminal process.)

Thus, despite some similarities with restorative justice, Barnett's proposal is better thought of as an expression of libertarian political values, and as being related to neoclassical law and economics, the proponents of which have also urged the substitution of monetary penalties over imprisonment on efficiency grounds (Posner 1980).

In any event, Barnett's proposal has been subject to considerable criticism. It is said that Barnett fails to appreciate the condemnatory aspect of punishment, which serves an important role in reinforcing collective moral norms (Miller 1978). Likewise, Barnett's emphasis on monetary penalties fails to address the particular dignitary harms suffered by the victims of intentional crimes (as contrasted with the qualitatively different harms suffered as a result of accidental torts) (Pilon 1978). Indeed, monetary awards for serious crimes against life or bodily integrity, such as murder or rape, might be seen as degrading and morally offensive (Miller 2009). More generally, restitution "sends the wrong message": "Punishment is not something you can buy your way out of, as if you could simply purchase a license to commit a crime" (Flanders 2006).

Critics wonder, moreover, how punishment would be available in Barnett's system for such crimes as cruelty to animals, harm to public institutions, attempts, and driving recklessly or while intoxicated (Miller 1978). Similarly, consensual transactions like drug distribution and prostitution would seemingly lie beyond punishment, notwithstanding views that these acts are inherently immoral (Miller 2009).

Critics also contend that Barnett faces a dilemma in how to handle restitution for unforeseeable injuries. Miller (2009) gives the example of two thieves who steal two identical jewelry boxes. It turns out, however, that one victim's jewelry box had great sentimental value because it was a gift from her deceased mother; to the other victim, the value of the jewelry box is simply the market price. If Barnett wishes to emphasize victim compensation, then it would seem that the two thieves should be subject to quite different restitution orders (assuming that the stolen items themselves cannot be returned). Yet, from the thieves' perspective, precisely the same crime has been committed against each victim. Differential punishment would thus violate the principle that equal punishments should be imposed for equally blameworthy conduct. Similar objections might be made to pure restitution's inability to recognize distinctions between intentional and accidental harms (Miller 2009).

Finally, critics argue also argue that Barnett has failed to address numerous practical difficulties with his proposal in a convincing fashion. For instance, Miller (2009) contends that no satisfactory mechanism is available for translating emotional harms into awards of money damages, particularly in view of the way that such valuations may be influenced by the sex, race, and class of the victim and of the person doing the valuation. For his part, Hershenov (1999) focuses on the challenge posed by unskilled or disabled offenders who cannot provide labor of sufficient value to offset the costs of their confinement, let alone to compensate victims. Additionally, there are concerns that a pure restitution system would not adequately deter crime, because no "premium" is added to the punishment to offset the likelihood that a prospective criminal will avoid apprehension or conviction (Barnett 1998).

To a considerable extent, the debate over pure restitution turns on deep questions of political philosophy and punishment theory. Barnett's proposal is likely to resonate with those who are especially distrustful of the state's use of coercive power against individual citizens, who reject the enforcement of paternalistic moral norms, and who find victim compensation a more legitimate or attractive project for the courts than enhancing social solidarity. Thus, the fact that pure restitution has not had much apparent influence among policymakers may, in part, reflect the continuing strength of communitarian impulses in defining and responding to crime.

But, as a practical matter, perhaps the greatest difficulty for pure restitution has been perceptions that it is not up to the task of protecting public safety – what Barnett himself characterizes as the "most obvious objection" to his proposal (Barnett 1998, p. 216). Barnett responds, however, that existing systems of deterrence, incapacitation, and rehabilitation are dysfunctional in various ways and perhaps even counterproductive to the objective of crime control. He notes, too, that his proposal may carry greater deterrence benefits than first appears. This is because compensation must be paid not merely for direct victim injuries, but also for the costs of detection, apprehension, and prosecution of the crime, and because these financial incentives will cause victims and law-enforcement agencies to put more effort into ensuring that offenders are caught and punished. Finally, Barnett notes that pure restitution is not incompatible with robust rights of self-defense and some limited use of preventive detention by the state, both of which may also serve crime-control ends.

Greater acceptance of pure restitution in the political system may have to wait for more widespread agreement that incarceration-oriented approaches are as ineffective as Barnett claims. Even at that, it is not clear that pure restitution would be a more appealing alternative than, say, restorative justice or a revival of treatment-oriented rehabilitation.

Design and Implementation Questions

In designing and implementing restitution programs, legislatures and court personnel have confronted a multitude of difficult questions. A few are briefly described below.

First, restitution orders may be limited to injuries caused by the formal offense of conviction or may reach out to embrace harms caused by any related criminal conduct. For instance, imagine that a burglar is caught in the act by police. A search of the burglar's home then uncovers evidence that she was also responsible for nine other burglaries. The prosecutor initially charges the offender with all ten burglaries, but then agrees to drop five of the charges in exchange for the offender's guilty plea to the remaining crimes. Should the victims of the dropped charges be able to obtain restitution at the offender's sentencing? If restitution is conceptualized purely as an alternative device for obtaining what would otherwise be available through a civil judgment, then ordering compensation for all victims seems a sensible enough answer – this would likely further the aim of providing victims with compensation in the most efficient and prompt manner. On the other hand, Boldt (1986) argues that going beyond the offense of conviction undermines the integrity of the criminal adjudication process, rendering adjudication an empty formality and diminishing the ability of punishment to serve its central function of strengthening social cohesion by reinforcing shared views of moral responsibility. (To be sure, commentators have for many years recognized an erosion of the "tort-crime distinction" as a result of the multiplication of no- and low-culpability regulatory crimes (Coffee 1991). Perhaps there is no "integrity" left to preserve on the criminal side.)

Second, whether or not the offender's relevant conduct is limited to the offense of conviction, there are questions regarding what categories of harm are compensable and how close the causal connection must be between the conduct and the harm. Courts often borrow the doctrine of proximate causation from civil tort law in order to impose some minimal limitation on recoveries, which seems quite consistent with the idea of restitution as an alternative to civil recoveries. Beyond this rather imprecise limitation, different jurisdictions impose a diverse array of additional limitations (Harland 1982). For instance, some limit recoveries to economic losses, while others specifically preclude "pain and suffering." Still others provide that victims may receive the full amount that they could obtain in a civil action. Likewise, there are different approaches to the possibility of recovery by the family members of deceased victims and others who might be thought of as secondary or vicarious victims (American Bar Association 2004). All such questions again put into issue the question of how important it is that restitution liability track conventional civil liability. As against the benefits of close tracking, the more inclusive approaches to restitution liability not only risk undermining the integrity of the criminal process (as Boldt might caution), but might also entangle criminal lawyers and judges in complex and unfamiliar sorts of loss calculations that could be more capably handled by civil practitioners.

Third, legislators, prosecutors, and judges must sort out the relative priority of restitution and competing sentencing alternatives. For instance, if a defendant is unable to pay restitution, court costs, and a fine, which should be paid first? Also, to what extent should it be a priority in fashioning a plea deal or a sentence to ensure that an offender is able to maintain or obtain employment? An extended period of incarceration will make full payment of restitution highly unlikely in many cases. More subtly, certain sorts of convictions may disqualify the offender from working in some professions (e.g., a lawyer may lose her license to practice law), which may also impede the payment of restitution. Yet, selecting the disposition that maximizes the likelihood of prompt, complete restitution may frustrate other legitimate objectives of the criminal-justice system, including the objective of obtaining funds from offenders to support system operations.

Finally, there is the question of whether ability to pay should be taken into account in setting a restitution award. US jurisdictions are split on this question (American Bar Association 2004). An emphasis on replicating civil judgments would point against taking ability to pay into account. This ability-blind approach might also be seen as accountability-reinforcing by formally making the offender fully responsible for all compensable harms that she has caused. Yet, a number of arguments might be made against this approach. For instance, a large financial obligation may create economic pressures that interfere with an offender's rehabilitation. Moreover, if the offender is unable to keep up with required payments, then the court and/or other responsible agencies must bear the administrative burdens of responding to the noncompliance, which may involve readjustments to the payment schedule or the imposition of sanctions, possibly including incarceration for offenders who had been released on probation. Victims may find the process of repeated sanctions and readjustments no less frustrating than offenders, and might even prefer a more modest restitution award that is more reliably paid.

Although setting awards on the basis of ability to pay might seem to lead to unequal treatment for offenders of different economic classes who commit the same crime, it may be that the severity of a financial sanction is more properly assessed by reference to its *relative* effect on the offender's wealth or income, rather than its *absolute* monetary value. This is, of course, the premise of day-fines. Thus, from an atonement-oriented perspective, an ability-blind award that will take many years for an offender to pay off (during which time the offender will likely have to remain under supervision) may communicate an inappropriately harsh message regarding the severity of the offense and the prospects for the offender to be restored to full membership in the community. Conversely, an award that takes ability to pay into account may be more precisely calibrated to serve communicative ends.

Future Directions

To the extent that policymakers and court officials continue to respond favorably to victim demands for restitution, a multitude of practical and theoretical problems will become increasingly urgent. This section briefly surveys a few of the emerging questions.

First, a prioritization of restitution as a routine, central component of criminal sentences may increase the demand for a variety of structural reforms to the criminal-justice system. Because few offenders have substantial reserves of wealth, the satisfaction of a restitution award in cases of serious injury normally requires that an offender have a steady source of income, a portion of which may be set aside for restitution payments over an extended period of time. Yet, a criminal conviction and the nonrestitution aspects of a sentence are apt to diminish or even entirely eliminate the offender's income-earning potential. Structural reforms that might be considered so as to minimize the economic effects of a conviction include: the elimination of employment-related "collateral consequences" of convictions (e.g., categorical bars on felons working in certain trades or obtaining government-subsidized loans for higher education); recognition of stronger confidentiality rights regarding convictions (e.g., through liberalized expungement laws or "ban the box" laws that prohibit employers from inquiring about prior convictions); greater use of deferred prosecution agreements, pursuant to which criminal charges might be greatly reduced or dropped entirely if an offender satisfies certain conditions (which might include entry into a legally enforceable victim-compensation agreement); more generous provision of social services to offenders so as to enhance their income-earning potential; diminished use of incarceration; greater availability of employment-friendly "intermediate sanctions" as an alternative to straight incarceration (e.g., community-based detention centers with work-release opportunities); diminished use of pretrial detention, which may result in job loss while an offender's case is pending in court; and diminished use of pre- and post-conviction release conditions that are employment-impairing (e.g., travel and residency restrictions). Of course, structural reforms such as these would not only support restitution, but would also address growing concerns regarding prisoner reentry and recidivism.

On the other hand, such structural reforms are also objectionable on a number of grounds. The stigmatization and socioeconomic marginalization of offenders might be seen as useful or necessary from the standpoint of deterrence or retribution. It might thus be regarded as morally offensive or insulting to victims if a person who commits a serious crime is able to proceed with her career as if nothing had happened, especially if the career is in a high-status profession. Likewise, offenders might be seen as undeserving recipients of social services, especially in a society in which many poor people who have not committed crimes are unable to obtain all of the services that they would find beneficial. Moreover, public safety might be compromised if there is less offender detention or reduced public access to criminal-history information. In light of such concerns, victim compensation might be better handled through expanded social insurance, rather than through structural reforms that are intended to facilitate restitution payments. Yet, expanded social insurance would not generate the atonement-type benefits that are said to accompany restitution.

Second, much as has occurred in tort law, collectively caused or experienced harms may put pressure on traditional notions of individualized proximate causation as a limitation on restitution liability. For instance, tort causation requirements have been relaxed in some jurisdictions in cases of "toxic torts" so that an injured plaintiff may recover from any of the manufacturers of a particular type of hazardous substance, even though the plaintiff may not be able to prove which manufacturer caused her specific injury. Likewise, tort law has devised various solutions to the problem of long latency periods for toxic exposure, which may leave it uncertain at the time of a lawsuit which individuals exposed to a tortfeasor's hazardous substances will eventually fall ill as a result. Environmental crimes present obviously analogous problems for restitution law. Drug trafficking offenses may present similar challenges. For instance, should narcotics wholesalers in a given market be liable for restitution to overdose victims in that market in proportion to their volumetric share of the market? And how about child pornography? Many subjects of child pornography experience significant psychological injuries as a result of the fear that their images are being widely circulated to strangers on the Internet. Some have sought, thus far unsuccessfully, to recover under federal restitution law from those convicted of possessing their images (Jacques 2011). Their claims raise questions about whether individual participants in the child-pornography market should be required to pay restitution for their role as consumers in sustaining a market whose very existence is said to harm victims.

In such circumstances of diffuse or uncertain causation, the question of whether to make restitution available presents familiar tensions between victim-compensation objectives, on the one hand, and, on the other, the goals of avoiding complex new adjudicatory responsibilities for the criminal-justice system and maintaining the system's traditional focus on individual moral responsibility. Additionally, in cases in which there are both identified and unidentified victims (as is frequently the situation in child-pornography cases), there are questions about whether and how assets should be preserved for victims who emerge in the future. It may be fairest and most efficient in such cases for restitution to be paid to a victim-assistance organization, rather than to individual victims.

Finally, there is the question of legal representation for victims in restitution proceedings. There may be a natural tendency to regard prosecutors as advocates for victims, and some prosecutors doubtlessly welcome such a role. Other prosecutors, however, may find that such a role presents awkward conflicts of interest, for instance, in the question of whether to drop certain charges in plea-bargaining when doing so would it be harder for some victims to recover. (Consider, for instance, the example given above of the prosecutor who drops five burglary charges in order to secure a guilty plea to five others.) Still other prosecutors may feel that they lack the time and resources to litigate restitution questions on victims' behalf.

In the face of prosecutorial disinterest, a restitution claimant may engage her own counsel, of course, but many victims will lack the resources to do so. These considerations might seem to support the court appointment of lawyers for victims at public expense (at least to the extent that the cost cannot, as a practical or legal matter, be shifted to the offender as part of the restitution award). Yet, such appointments might seem inappropriate if public funding is already inadequate for the legal representation of indigent *defendants*, as seems to be the case in many jurisdictions; it seems misguided to fund a process that will likely increase the level of punishment in a system when there are reasons to doubt the reliability of the underlying guilt determinations. The difficulties may supply yet an additional reason to narrow the scope of restitution liability relative to tort, confining restitution to the sorts of direct, easily monetizable injuries so as to which victims are least likely to require legal representation in order to prove.

Related Entries

- ▶ Civil Remedies
- ▶ Fines in Europe
- ▶ History of Restorative Justice
- ▶ Penal Abolitionism
- ▶ Penal Philosophy and Sentencing Theory
- ▶ Proportionality
- ▶ Theories of Punishment
- ▶ Traditional Retributivism
- ▶ Victims and Restorative Justice
- ▶ Victims' Rights in the Criminal Justice System

Recommended Reading and References

American Bar Association (2004) Restitution for crime victims: a national strategy

Barnett R (1977) Restitution: a new paradigm of criminal justice. Ethics 87:279–301

Barnett R (1998) The structure of liberty: justice and the rule of law. Oxford University Press, New York

Benson B (1996) Restitution in theory and practice. J Libert Stud 12:75–97

Boldt R (1986) Restitution, criminal law, and the ideology of individuality. J Crim Law Criminol 77:969–1022

Christie N (1978) Conflicts as property. Br J Criminol 17:1–15

Coffee J (1991) Does "Unlawful" mean "Criminal"?: reflections on the disappearing tort/crime distinction in American law. Boston University Law Rev 71:193–246

Dickman M (2009) Should crime pay?: a critical assessment of the mandatory victims restitution act of 1996. Calif Law Rev 97:1687–1718

Doak J (2008) Victims' rights, human rights and criminal justice: reconceiving the role of third parties. Hart Publishing, Oxford

Duff RA (2001) Punishment, communication, and community. Oxford University Press, Oxford

Duff RA (2005) Guidance and guidelines. Columbia Law Rev 105:1162–1189

European Commission (2001) Green paper—compensation to crime victims

Flanders C (2006) Shame and the meanings of punishment. Clevel State Law Rev 54:609–635

Garvey S (1999) Punishment as atonement. UCLA Law Rev 46:1801–1858

Harland A (1982) Monetary remedies for the victims of crime: assessing the role of the criminal courts. UCLA Law Rev 30:52–128

Hershenov D (1999) Restitution and revenge. J Philos 96:79–94

Hodgson J (2002) Suspects, defendants and victims in the French criminal process: the context of recent reform. Int Comp Law Q 51:781–816

Holezel W (1980) A survey of 27 victim compensation programs. Judicature 63:485–496

Jacques R (2011) Amy and Vicky's cause: perils of the federal restitution framework for child pornography victims. Ga Law Rev 45:1167–1198

Loffelmann M (2006) The victim in criminal proceedings: a systematic portrayal of victim protection under German criminal procedure law

Miller F (1978) Restitution and punishment: a reply to Barnett. Ethics 88:358–360

Miller R (2009) Restitution's focus on the victim: six problems. Macalester J Philos 18:21–39

Moore M (1997) Placing blame: a general theory of the criminal law

Pilon R (1978) Criminal remedies: restitution, punishment, or both? Ethics 88:348–357

Posner R (1980) Optimal sentences for white collar criminals. Am Crim Law Rev 17:409–418

Ruback R, Shaffer J (2005) The role of victim-related factors in victim restitution: a multi-method analysis of restitution in Pennsylvania. Law Hum Behav 29:657–681

Ruback R, Cares A, Hoskins H (2008) Crime victims' perceptions of restitution: the importance of payment and understanding. Violence Vict 23:697–710

Umbreit M, Vos B, Coates R, Lightfoot E (2005) Restorative justice in the twenty-first century: a social movement full of opportunities and pitfalls. Marquette Law Rev 89:251–304

Von Hirsch A (1998) Penal theories. In: Tonry M (ed) The handbook of crime and punishment

Zimring F, Hawkins G, Kamin S (2001) Punishment and democracy: three strikes and you're out in California. Oxford University Press, Oxford

Restorative Justice

- ▶ International Responses to Victims in Criminal Justice
- ▶ Shaming in Asian Societies
- ▶ Transitional Justice

Restorative Justice and Practice

Jonathan Doak[1] and David O'Mahony[2]
[1]Durham Law School, Durham University, Durham, UK
[2]Durham University, Durham, UK

Overview

In recent times the promulgation of restorative justice has come to exert an increasingly strong influence in criminal justice policymaking. The restorative paradigm conceptualizes criminal behavior in a very different manner from orthodox models of criminal justice and aims to give "ownership" of the dispute to the primary stakeholders: the victim, the offender, and the community. There are many different forms of restorative justice practice, though they are frequently classified under four main heads: victim-offender mediation, panel-based interventions, police-led restorative cautioning, and family group conferencing. Overall, research shows that both victims and offenders are more likely to express satisfaction with restorative schemes than with court-based processes. There is also evidence to show that restorative interventions can have an impact in reducing recidivism.

Introduction

Restorative justice has developed as victim-centered approach which conceptualizes criminal behavior in a very different manner from which it has been traditionally conceived within orthodox models of criminal justice. In recent years, it has come to exert an increasingly strong influence over juvenile justice systems as policymakers have become increasingly concerned about the capacity of the traditional criminal justice system to deliver participatory processes and fair outcomes that are capable of benefiting victims, offenders, and society at large.

Proponents of restorative justice conceptualize criminal behavior in a very different way from orthodox criminal law theory. Crime is viewed first and foremost a violation of people and relationships, and the restorative process aims to repair the harm that has been caused to victims, offenders, and communities (Zehr 2005). The retributive focus of the traditional justice system and the prioritization of punishment and just deserts are thus not the primary goals of restorative justice. The normative frameworks of the traditional criminal justice system are thus viewed as being largely ineffective and even undesirable insofar as they can be counterproductive in their failure to meet the needs of those most affected by crime (Zehr 2005). By contrast, the restorative process looks to the needs of the individual parties. It seeks to repair the injuries caused, restoring relationships, repairing harm, and addressing the needs of the offender. In turn, restorative approaches are said to promote greater community protection and safety by promoting offender responsibility, restoring and reintegrating the key stakeholders, providing empathy and understanding for participants, and ultimately (hopefully) reducing the likelihood of reoffending (Braithwaite 1989; Zehr 2005).

Defining Restorative Justice

In spite of its growing popularity, restorative justice remains a contested concept, which has proved difficult to define in concise terms. One of the most widely accepted definitions is that provided by Marshall (1999 p. 5), who described it as "a process whereby all the parties with a stake in a particular offence come together to resolve collectively how to deal with the aftermath of the offence and its implications for the future." However, other commentators have argued that as an alternative to associating the with a specific archetypal process, the term should be instead thought of as encapsulating a body of core practices which aim to maximize the role of those most affected by crime: the victim, the offender, and potentially the wider community (Dignan 2005).

The potential fluidity of the concept underscores the risk that it may come to mean "all things to all people" (McCold 2000). Such a danger is highlighted by the fact that restorative justice programs worldwide vary considerably in terms of what they do and how they achieve their outcomes – let alone the level of actual "restorativeness" in terms of what is done. Practice often differs according to the situation and the manner in which programs have developed in local areas. For instance, not all schemes accept all restorative principles in equal measure, and there is also considerable divergence on the extent to which schemes may be integrated into existing criminal justice structures. As such, there is no single "prototype" format for practices that adopt the "restorative" label. While the best-known restorative schemes often operate in relation to youth or lower-level offending, there are now programs covering a diverse range of issues including corporate crime, conflicts in prisons and schools, domestic violence, serious violent crime, and political violence (see further Johnstone and Van Ness 2006).

Schemes may also differ in relation to the degree of formality or legality. Some schemes are based in statute and require that offenders be dealt with through a strict legislative (restorative) framework. In such instances, considerable resources are often invested across a range of statutory agencies, and widespread training and a change in working culture of the police, prosecutor, and courts are required. By contrast, other schemes may be practice led and "voluntary" and are much more informal. Typically, such schemes will lie on the fringes of the criminal justice system and may experience problems with resources and logistics.

Even where restorative programs become integrated within the formal criminal justice system, the conditions for referring a young person may vary considerably, with some schemes taking referrals as diversionary interventions or as a form of police caution, while others may be referred by the prosecutor or by the courts as an official form of disposal (Dünkel et al. 2011). Programs can be led by different agencies, such as the police, an independent conferencing service, or even local community mediation organizations. Programs also differ according to the level of victim involvement, with some using face-to-face meetings, while others use indirect mediation or those that rarely involve victims. Even the nature of the offence and type of offender differ, with some schemes only taking first-time offenders who have committed relatively minor offences, to others that consider a whole range of offences and individuals who have offended in the past. In essence, restorative practice is highly divergent and in considering how programs operate, one needs to be aware of their scope, range, and intentions (Doak and O'Mahony 2011).

Restorative Justice in Practice

Although restorative justice has permeate both law and criminal justice practice in around the world, there are clear differences in how these interventions are delivered and what they are designed to achieve. The following describes the main types of interventions, which include mediation, panel-based interventions, police-led restorative cautioning, and family group conferencing models.

Victim-Offender Mediation

Victim-offender mediation programs were first developed in Ontario, Canada, in the mid-1970s. They were promoted by the Christian Mennonite movement to promote personal reconciliation between victims and offenders. As such, they seek to bring together victims and offenders with a facilitator or mediator, who is usually professionally trained. Mediation aims to give victims and offenders a safe environment in which they are able to discuss the crime, its impact, and the harm it may have caused and to allow an opportunity to put right the harm caused.

Victim-offender mediation has proved very popular and is currently the most common form of restorative practice used across the United States and Canada (Dignan 2005). The technique is also widely used as a form of dispute resolution used to deal with conflicts within communities, especially where there has been a history of tension between the parties that has not been resolved by other forms of intervention. Mediation has also gained in popularity across many parts of Europe where it is currently one of the most dominant forms of restorative practice (Miers et al. 2001).

Most mediation projects are used to divert minor juvenile or adult offenders away from criminal sanctioning. Often prosecutors act as the gatekeepers to such programs and refer cases to mediation before they reach court. There is considerable variation in the types of programs available and some – particularly in cases involving young people – are based on an extension of welfare legislation, rather than being used as a penal measure.

The use of mediation is relatively a commonplace in Austria as a means of diversion for both juvenile and adult offenders. The process is designed to ensure that a much more comprehensive form of restitution is delivered which is more constructive for victims of crime than just a court-imposed sanction. The mediation process has three possible components which provide compensation for personal injury, loss, or damage; reconciliation talks which provide apologies and may help the victim; and in exceptional cases community service or payments to public welfare institutions or "symbolic restitution" (Miers et al. 2001).

Mediation services are generally delivered by organizations linked to the Probation Service. The mediation usually takes place directly with the victim, offender, and mediator in the same room. Unlike conferencing, supporters and representatives of the wider community are not generally permitted unless juveniles are involved. Occasionally indirect or "shuttle mediation" is also arranged, where the parties are reluctant to meet in person. Mediators meet regularly with members of the judiciary and representatives of the public prosecutor's office to share experiences and discuss difficult cases.

Similarly, Norway makes considerable use of mediation and the services are run directly by the central Ministry of Justice. Mediation referrals are usually made by the police or prosecutor, though they can also be ordered by courts as part of a community sentence or as a condition of a suspended sentence. The process only takes place after an admission or finding of guilt and the prosecutor usually decides whether the case is suitable for referral to mediation. Providing the parties agree on the facts of the case and consent to the process, mediation will be offered as an alternative to a formal penal sanction. The prosecutor will also take into account whether there is a personal victim, as well as considerations relating to individual deterrence. Thus, despite mediation being well developed throughout the country, it is still generally confined to less serious offences and retains its diversionary character (Kemény 2005).

The mediation process in Norway usually involves face-to-face meetings between victims and offenders. The mediation event may be brief, as is typically the case with offences against property, or prolonged, as is the case with neighbor disputes or violence (Miers et al. 2001). Services are normally provided by trained volunteers, who receive a nominal fee. Volunteers are trained through a national accreditation program, and are accountable to a coordinator based within the Ministry of Justice. In addition, the Ministry of Justice arranges annual conferences and publishes a regular journal for mediators to inform and generate exchange and good practice (Miers et al. 2001). The vast majority of mediated cases reach an agreement in Norway and nearly all of these agreements are fulfilled. The major forms of disposals include compensation, work, and reconciliation.

In the United Kingdom, there have been a number of victim-offender mediation projects; however, these have mostly developed on an ad hoc basis. Generally, they have been relatively small scale and have been local initiatives, rather than being "mainstream" projects. This has largely been the result of a lack specific legislation to formally establish them and the provision of little or no central funding, which has obviously impeded their development (Dignan 2005). Although the Home Office funded a number of pilot projects in the 1980s, funding was not continued beyond the pilot phase. Most of the schemes that operate today have been developed at a local level, through good working partnerships between criminal justice agencies such as probation, the police, and social services. However, they continue to experience considerable logistical and resourcing difficulties, concerning which agencies should fund them and where their referrals come from.

Recently, the government undertook to fund a number of these voluntary schemes and offered financial backing and an evaluation as part of its Crime Reduction Programme (Shapland et al. 2011). These schemes were established primarily to concentrate on cases involving adult offenders at different stages of the criminal justice process, from presentence through to release from prison. They are largely typical of those operating on a local level around the country, although they cover offenders convicted of serious crimes, including assaults, robberies, and burglaries.

Panel-Based Interventions

Panel-based interventions such as community reparation boards have been commonly used in the United States. While these are primarily used for adult offenders convicted of nonviolent offences, more recently their use has been extended to cover juvenile offenders. Such boards are typically composed of a small number of community representatives who hold face-to-face meeting with offenders, who have been referred by the courts. The board, with the offender, decides the sanction that should be imposed for the offence, monitor compliance, and report back to the court on its completion. The main goal of the boards is to promote community involvement and empowerment in relation to offending and to promote offender responsibility and victim reparation (Bazemore and Umbreit 2002).

In the United Kingdom, panel-based interventions are used as part of the referral order. It is a mandatory court disposal for first-time low-level juvenile offenders between the ages of 10 and 17. The referral order operates by diverting a young person away from court to a Youth Offender Panel. The panel then decides how the offence should be dealt with and what form of action is necessary. If the victim wishes, they may attend the meeting and can describe how the offence affected them. Parents are required to attend the panel meeting (if the young person is under the age of 16) and meetings are usually held in community venues.

The panel offers a discussion forum, and in theory, it gives the young person an opportunity to make reparation to the victim and community, take responsibility for their actions, and achieve reintegration into society. Although similar mechanisms in certain North American jurisdictions require the young person to consent to taking part in such a program, this is not the case with the English system of referral orders. However, the disposal is the result of a contract that must be negotiated and agreed and not simply imposed on a young person. Usually, a contract will contain specific offender-orientated interventions such as family counselling, mentoring, victim awareness sessions, and drugs or alcohol programs, which may be complemented by reparative measures, like a verbal or written apology, or an offer to repair damaged property. The amount of reparation and length of the plan have to be proportionate to the seriousness of the offence, and the young person must agree to the plan. In the event that there is no agreed outcome, the Youth Offending Panel will report back to the court, who will then reconsider how to deal with the young person.

Referral orders allow young people to play more of a role in how they are treated and they may find their plan or contract "useful," especially compared to the experience receiving a punitive sentence by the Youth Court. However, concerns have been raised about referral orders, especially where young people and parents may feel forced into agreeing plans. Some panels have experienced difficulty devising suitable plans because of a lack of local resources, and there have been ongoing difficulties of involving victims in the process (Crawford and Newburn 2003). The low level of victim participation in these panels arguably places a significant limitation on their overall restorative potential.

Police-Led Restorative Cautioning

Police-led restorative cautioning schemes were developed in the early 1990s, mostly as an alternative approach to the traditional formal police caution. The approach spread and was taken up and used in various forms in New Zealand and America, particularly in Minnesota and Pennsylvania in the mid-1990s.

The approach is largely based around Braithwaite's ideas of "reintegrative shaming" (Braithwaite 1989). In essence this approach deals with crime and its aftermath by attempting to make offenders ashamed of their behavior but in a way which promotes their reintegration into the community. It is different to the traditional police caution, which has been described by Lee (1998) as a "degrading ceremony" in which the young person, most often a first-time and minor offender, is given a "dressing-down" and good talking by a senior police officer.

The process is achieved by firstly attempting to get the young person to realize the harm caused by their actions to the victim, their family, and themselves. The focus is placed on the wrongfulness of the action or behavior, rather than the wrongfulness of the individual. The process attempts to reintegrate the young person, after they have admitted what they did was wrong, by focusing on how they can put the incident behind them, for example, by repairing the harm through such things as apology and reparation. It thereby allows the young person to move forward and reintegrate back into their community and family. The whole process is usually facilitated by a trained police officer and involves the use of a script or agenda that is followed in the conferencing process. The victim is encouraged to play a part in the process, particularly to reinforce upon the young person the impact of the offence on them, but as Dignan (2005) notes, restorative cautioning schemes have typically placed a greater emphasis on the offender and issues of crime control than on their ability to meet the needs of victims.

Despite these concerns, it has generally been found that the police are enthusiastic and committed to the restorative process. They are usually well trained and young people and their parents often place a high degree of confidence and support for the scheme. There is also evidence that restorative cautioning schemes can have other beneficial effects especially in terms of helping improve police/community relations (O'Mahony and Doak 2004; Doak and O'Mahony 2011).

Restorative Conferencing

Perhaps the best-known restorative conferencing scheme is the family group conferencing model, first developed in New Zealand. The model was devised as part of a more general initiative to address difficulties in the way young people were being treated in the criminal justice and welfare systems in New Zealand – particularly minority group offenders. The model sought to develop a more culturally sensitive approach to offending that emphasizes inclusive participation and collective decision making, bringing together young people, their families, and the broader community to determine appropriate means of redress for victims (Maxwell and Morris 1993).

The family group conferencing scheme was brought into New Zealand under a reform agenda that emphasized diverting young people away from criminal justice interventions through police cautions or informal resolutions. The legislation made conferencing the main avenue of disposal for all but the few most serious offences like murder and manslaughter. In effect, family group conferencing became the main statutory method of disposal for young offenders being prosecuted. Young people can only be prosecuted if they have been arrested and referred by the police. The courts are required to send offenders for family group conferences and they have to consider the recommendations of the conference and usually do not deal with cases until they have had a conference recommendation (Maxwell and Morris 1993).

Northern Ireland has also adopted a fully mainstreamed statutory-based restorative conferencing model for young offenders. The measures provide for two types of disposal, diversionary and court-ordered conferences. Both types of conference take place to provide a plan to the prosecutor or court on how the young person should be dealt with for their offence. For the diversionary conference, which is usually for less serious offences, the successful completion of a conference plan will avoid a court appearance and criminal conviction. Court-ordered conferences, on the other hand, are referred for conferencing by the court. An important feature of the legislation is that the courts *must* refer all young persons for youth conferences, except for offences carrying a mandatory life sentence. The court *may* refer cases that are triable by indictment only or scheduled offences under terrorism legislation. In effect, the legislation makes conferencing mandatory except for a small number of very serious offences. However, for conferencing process to take place, the young person must agree to the process and they must either admit guilt or have been found guilty in court.

The format of the youth conference itself bears much similarity to the general model used in New Zealand, though the Northern Ireland model places the victim much more at the center of the process. It normally involves a meeting, chaired by an independent and trained youth conference facilitator (employed by a Youth Conferencing Service), where the young persons will be provided with the opportunity to reflect upon their actions and offer some form of reparation to the victim. The victims, who are actively encouraged to attend, are allowed to explain to the offender how the offence affected them, in theory giving the offender an understanding of their actions and allowing the victim to separate the offender from the offence. Following a dialogue a "youth conference plan" or "action plan" will be devised which takes into consideration the offence, the needs of the victim, and the needs of the young person. The young person must give their free consent for a plan to be used. Plans can run for a period of not more than 1 year and usually involve some form of reparation or apology to the victim. Ideally the plan will include elements that address the needs of the victim, the offender, and the wider community, so as to achieve a restorative outcome (O'Mahony and Campbell 2006). Plans may include restrictions and requirements of the young person, including curfews. They may even recommend a custodial sentence.

The direct involvement of offenders and victims in conferencing and their ability to engage in dialogue contrast starkly with the conventional court process. In court, offenders are afforded a passive role – generally they do not speak other than to confirm their name, plea, and understanding of the charges – and are normally represented and spoken for by legal counsel throughout proceedings. Similarly, victims are able to actively participate in the conferencing process and many report the experience as valuable, in terms of understanding why the offence has been committed and in gaining some sort of apology and/or restitution.

The vast majority of conference plans are agreed by participants and victims and offenders usually report being happy with the content of the plans. Interestingly, plans often center on elements designed to help the young person and victim, such as reparation to the victim or attendance at programs to help the young person. The most positive aspects of conferencing for participants often relate to its ability to provide a means of moving forward for both parties, beyond simply punishing the offender. Both victims and offenders who have been through conferencing express a strong preference for it, as opposed to going to court, and they see conferencing as offering a more meaningful environment for them.

Research and Evaluation

The vast majority of research studies on restorative programs to date report high levels of satisfaction among participating victims irrespective of the seriousness of the offence or cultural or geographical variations. Satisfaction rates are particularly high where comparative research has evaluated the experiences of victims who went through a restorative process compared with those who went through conventional processes. Umbreit and Coates (1993) noted that 79 % of victims who participated in mediation schemes in the USA were satisfied with the way in which their cases had been handled, compared with only 57 % of those whose cases were dealt with in a non-restorative manner. McCold and Watchell (1998) reported statistics of 97 %

satisfaction for the restorative process compared with 81 % for the non-restorative process, and Sherman et al. (2005) reported that 70 % of victims who had participated in a restorative conference were satisfied, in comparison with 42 % whose cases where dealt with by a court. These findings are reflected in Latimer et al.'s meta-analysis (2001), where participation in a restorative scheme was found to result in higher victim satisfaction ratings when compared with comparison groups in all but one of the 13 programs examined. More recently, similar findings have been reported by Poulson (2003), who found that victims who had participated in restorative programs were half as likely to feel upset about the offence afterwards compared with those victims who attended court.

One of the more significant research evaluations to date has been the Canberra-based Reintegrative Shaming Experiments (RISE), which randomly assigned cases to a police-led conference or a court hearing and sought to compare the effectiveness of each procedure. The researchers arrived at the "inescapable" conclusion that "both victims and offenders can name many ways in which they prefer conferences to court" (Sherman et al. 1998: 165). Perceptions of fairness among victims and offenders were higher and observations reported greater participation, emotional intensity, procedural justice, apologies, forgiveness, and time and effort given to justice in conferences than in court (Sherman et al. 1998). Furthermore, conferences were said to increase offenders' respect for the law and the police, and over 70 % of contractual obligations were fulfilled by offenders (Sherman and Strang 2007). Similar positive data relating to stakeholder satisfaction have been collected in evaluations of police-led projects in the United States (see, e.g., McCold and Watchell 1998).

More mixed evaluation results have been found in relation to police-led restorative cautioning. The Thames Valley scheme evaluated by Hoyle et al. (2002) found most offenders and victims were generally satisfied and felt they had been treated fairly. But a significant minority felt they had not been adequately prepared for the process or said had been pressured into it.

Most evaluations report that victims gain a strong sense of procedural justice. Mediators/facilitators have been widely regarded as fair and objective (see, e.g., Shapland et al. 2011; Campbell et al. 2005), with victims feeling that they had adequate opportunity to express themselves (see, e.g., Trimboli 2000; Hayes and Daly 2004; Campbell et al. 2005; Shapland et al. 2011). Typically, victims place particular value on the ability to explain the effect the crime has had on them, put a face to the offender, and to have questions such as "why me?" answered (Hayes and Daly 2004; Campbell et al. 2005; Shapland et al. 2011). In relation to outcomes, satisfaction levels are similarly high (Campbell et al. 2005; Shapland et al. 2011).

Research also indicates that offenders are generally satisfied with their experience of the restorative process. In Northern Ireland, 92% of the young offenders felt the conference had helped them realize the harm they had caused and 93 % felt the conference plan was fair (Campbell et al. 2005). Similarly, 71 % of the young offenders were satisfied with their conference. In New Zealand, 84 % of offenders and 80 % of parents said that they were satisfied with the outcome of the family group conference. By contrast offender perceptions of their involvement in the process were much lower: 34 % of young people felt involved and only 9 % felt that they ultimately decided the outcome (Maxwell and Morris 1993). Results from more recent research suggest some level of improvement with around half of the sample of young people reporting that they felt involved in the conference process (Maxwell et al. 2004). Arguably, the inclusion and active participation of the offender is vital to the restorative process and the extent to which it is achieved may impact upon their perception of fairness and justice.

The perceived success or failure of restorative programs is often linked to their impact on recidivism rates. An overview of 14 different studies conducted by Nugent et al. (2003) found that those who participated in mediation had a statistically significant lower rate of recidivism. A broadly similar study conducted by the Canadian government (Latimer et al. 2001) found that restorative programs including mediation and group conferencing yielded reductions in recidivism compared with other non-restorative approaches. For programs involving group conferencing and sentencing circles, Umbreit et al. (2006) found somewhat mixed results. For the majority of the studies, the results were generally positive and showed that those who had gone through a conferencing process were less likely to reoffend than those who were traditionally processed. However, some of the studies reported differing results for different types of offenders. They note the RISE experiments from Australia showed a reduction in reoffending for violent crimes, but not a statistically significant reduction for the other categories of offences studied (mostly property offences). The research conducted by McCold and Watchell (1998) similarly found in their study of police-led conferencing that conferences had a more positive impact on recidivism rates for participants whose offences were relatively more violent.

A comprehensive meta-analysis of research looking at restorative interventions and their impact on reconviction was completed by James Bonta and his colleagues (Bonta et al. 2008). This meta-analysis was based on a wide review of differing programs that had a restorative basis and included court-imposed family group conferences and a range of community-based forums. Their analysis included studies with a comparison sample and recidivism data. Some 39 studies were used and nearly half were situated within a court setting, covering a range of restorative based programs. Overall, the researchers found the restorative programs produced a statistically significant 7 % reduction in recidivism rates over an average 17-month follow-up period (Bonta et al. 2008). In the case of violent offenders, the effect of the restorative intervention was stronger, with a reduction in recidivism of about 15 % (Bonta et al. 2008). However, the research suggests that the restorative interventions may be less effective with repeat, high-risk offenders – who by definition have a much higher rate of recidivism. These research findings are generally supported by other similar research, especially more recent studies (e.g., Shapland et al. 2011; Sherman and Strang 2007). In sum, therefore, it appears that restorative interventions have the ability to deliver lower levels of recidivism, especially compared with other criminal justice interventions. They also appear to work better with offenders who have committed more serious types of offences than petty offences.

Related Entries

- ▶ Examining the Effectiveness of Correctional Interventions
- ▶ History of Restorative Justice
- ▶ Law, Diversion and Community Sanctions in Juvenile Justice
- ▶ Monitoring and Evaluation of Restorative Justice
- ▶ Restorative Justice and State Crime
- ▶ Victims and Restorative Justice

Recommended Reading and References

Bazemore G, Umbreit M (2002) A comparison of four restorative conferencing models by the office of juvenile justice and delinquency prevention. In: Perry JG (ed) Repairing communities through restorative justice. American Correctional Association, Lanham

Bonta J, Jesseman T, Rugge T, Cormier R (2008) Restorative justice and recidivism: promises made, promises kept? In: Sullivan D, Tifft L (eds) Handbook of restorative justice. Routledge, Abingdon

Braithwaite (1989) Crime, shame and reintegration. Cambridge University Press, Cambridge

Campbell C, Devlin R, O'Mahony D, Doak J, Jackson J, McEvoy K (2005) Evaluation of the Northern Ireland youth conference service. NIO Research and Statistics Agency, Belfast

Crawford A, Newburn T (2003) Youth offending and restorative justice: implementing reform in youth justice. Willan Publishing, Cullompton

Dignan J (2005) Understanding victims and restorative justice. Open University Press, Maidenhead

Doak J, O'Mahony D (2011) Developing mediation and restorative justice for young offenders across Europe. In: Dünkel F, Grzywa J, Horsfield P, Pruin I (eds) Juvenile justice systems in Europe – current situation and reform developments, 2nd edn. Forum Verlag Godesberg, Mönchengladbach, pp 1717–1746

Dünkel F, Pruin I, Grzywa J (2011) Sanctions systems and trends in the development of sentencing practices.

In: Dünkel F, Grzywa J, Horsfield P, Pruin I (eds) Juvenile justice systems in Europe – current situation and reform developments, 2nd edn. Forum Verlag Godesberg, Mönchengladbach, pp 1649–1716

Hayes H, Daly K (2004) Conferencing and re-offending in Queensland. Aust N Z J Criminol 37:167

Hoyle C, Young R, Hill R (2002) Proceed with caution: an evaluation of the Thames valley police initiative in restorative cautioning. Joseph Rowntree Foundation, York

Johnstone G, Van Ness D (eds) (2006) Handbook of restorative justice. Willan Publishing, Cullompton

Kemény S (2006) Victim-offender mediation with juvenile offenders in Norway. In: Mestitz A, Ghetti S (eds) Victim-offender mediation with youth offenders in Europe. An overview and comparison of 15 countries. Springer, Dordrecht

Latimer J, Dowden C, Muise D (2001) The effectiveness of restorative justice practices: a meta-analysis. Department of Justice Canada, Ottawa

Lee M (1998) Youth crime and police work. MacMillan Press, London

Marshall T (1999) Restorative justice: an overview. Home Office, London

Maxwell G, Morris A (1993) Families, victims and culture: youth justice in New Zealand. Social Policy Agency and Institute of Criminology, Victoria University of Wellington, Wellington

Maxwell G, Kingi V, Robertson J, Morris A, Cunningham C (2004) Achieving effective outcomes: youth justice in New Zealand. Ministry of Social Development, Wellington

McCold P (2000) Towards a Mid-range theory of restorative criminal justice: a reply to the maximalist model. Contemp Justice Rev 3:357–414

McCold P, Watchell B (1998) Restorative policing experiment: the Bethlehem Pennsylvania police family group conferencing project. Community Service Foundation, Pipersville

Miers D, Maguire M, Goldie S, Hale C, Netten A, Uglow S, Doolin K, Hallam A (2001) An exploratory evaluation of restorative justice schemes. Home Office, London

Nugent W, Williams M, Umbreit M (2003) Participation in victim-offender mediation and the prevalence of subsequent delinquent behaviour: a meta-analysis. Res Soc Work Pract 14:408–416

O'Mahony D, Campbell C (2006) Mainstreaming restorative justice for young offenders through youth conferencing: the experience of Northern Ireland. In: Junger-tas J, Decker S (eds) International handbook of youth justice. Springer, Amsterdam

O'Mahony D, Doak J (2004) Restorative justice: is more better? Howard J 43:484–505

Poulson B (2003) A third voice: a review of empirical research on the psychological outcomes of restorative justice. Utah Law Rev 1:167–203

Shapland J, Robinson G, Sorsby A (2011) Restorative justice in practice: evaluating what works for victims and offenders. Routledge, London

Sherman L, Strang H (2007) Restorative justice: the evidence. Smith Institute, London

Sherman L, Strang H, Barnes J, Braithwaite J, Inkpen N, Teh M (1998) Experiments in restorative policing: a progress report on the Canberra reintegrative shaming experiments. Australian Institute of Criminology, Canberra

Sherman L, Strang H, Angel C, Woods D, Barnes G, Bennett S, Inkpen N (2005) Effects of face-to-face restorative justice on victims of crime in four randomized controlled trials. J Exp Criminol 1:367

Trimboli L (2000) An evaluation of the NSW youth justice conferencing scheme. NSW, Bureau of Crime Statistics & Research, Sydney

Umbreit N, Coates R (1993) Cross-site analysis of victim offender mediation in four states. Crime Delinq 39:565

Umbreit M, Coates R, Vos B (2006) Victim-offender mediation: evidence-based practice over three decades. In: Moffitt M, Bordone RC (eds) The handbook of dispute resolution. Jossey-Bass, San Francisco

Zehr H (1990) Changing lenses: a new focus for crime and justice. Herald Press, Waterloo

Restorative Justice and State Crime

Elmar G.M. Weitekamp[1] and
Stephan Parmentier[2]
[1]University of Tuebingen, Tuebingen, Germany
[2]Institute of Criminology, University of Leuven, Leuven, Belgium

Synonyms

International crimes; Political crimes; Transitional justice

Overview

Despite the continuing decline in the number of violent conflicts in recent years, the post-World War II period can be marked as one of the most violent periods in human history. The gross majority of the conflicts were intrastate conflicts involving flagrant and massive human rights violations. It is estimated that in the period 1945–1996 alone, 220 conflicts (not including international conflicts) resulted in 87 million deaths and many more millions of people stripped of their fundamental rights, property, and dignity (Balint 1996). In the final decades of the last century, various mechanisms for dealing with violent conflicts have come to light, in order to call the offenders to account and to provide compensation to the victims. The mechanisms that were put in place range from national and international tribunals or courts to nonjudicial forums like truth commissions. The emphasis on the development of more victim-oriented mechanisms is a logical continuation of the growing interest in the fate of victims in favor of the victors of conflicts. Some explain this profound shift as the progressive consolidation of the democratic model. At the same time, criminology and the understanding of crime have gone through the same paradigm shift – the social criminal as a survivor in poor communities has made room for the suffering victims.

The conflicts and crimes listed as prime examples of international crimes, that is, crimes incorporated in the Rome Statute of 1998 establishing the International Criminal Court, are as follows: (a) war crimes, (b) crimes against humanity, and (c) genocide. It should be noted that the large majority of these violent crimes are committed by states and their agents (e.g., police forces, army personnel, and intelligence services) and can therefore be considered state crimes (Rothe and Mullins 2010). However, the concept of state crimes is also wider, as it involves behavior that is not traditionally regarded as violent, such as instances of espionage or corruption.

This contribution largely equates state crimes with international crimes and leaves out the particular instances of treason, espionage, and corruption, which would merit a separate analysis. Its objectives are twofold: the first is to understand the recent developments that are gradually leading away from situations of impunity of international crimes toward situations of post-conflict justice into the face of a democratic transition; the second is to integrate the idea of restorative justice in the framework of post-conflict justice. The main argument is that the restorative justice theory, as an emerging discourse within the criminological sciences, may offer a promising approach toward the victims and perpetrators of mass violence. Moreover, it is suggested that it could contribute to broadening the object of criminology by shifting the attention from common crimes to political crimes and to deepening our theoretical understanding of the phenomenon of mass violence (Fig. 1).

The text takes the following structure. It first looks at the dominant approach to deal with international crimes, namely, retributive justice as evidenced through the criminal prosecutions of such crimes. It then shifts to a restorative justice approach and highlights its main features in the context of international crimes, by looking at truth commissions as the primary example of restorative justice. And, finally, it attempts to assess the relationship between these two very different models of justice in the case of violent conflicts and international crimes.

Retributive Justice Mechanisms and Their Shortcomings

The end of the twentieth century has witnessed the revival of international criminal justice in response to mass atrocities. The establishment of the International Criminal Court symbolizes this development. This first permanent international court will have the competency to judge individuals for having committed international crimes. It has been heralded by many as the start of a new era of justice. This era was already prepared by the establishment in 1993 and 1994 of two international ad hoc tribunals to deal with the gross human rights violations committed in the former Yugoslavia (ICTY) and in Rwanda (ICTR). Furthermore, the world has witnessed the prosecution and conviction of four Rwandans in a Belgian court, charged with complicity in genocide in Rwanda. Perpetrators of international crimes can be prosecuted outside their home country, in a "third country," on the basis of universal jurisdiction. It goes beyond the scope of this entry to describe all the existing retributive justice mechanisms that are available to post-conflict states when dealing with gross human rights violations. Instead, it focuses on the basic features of criminal prosecutions at the national and the international level.

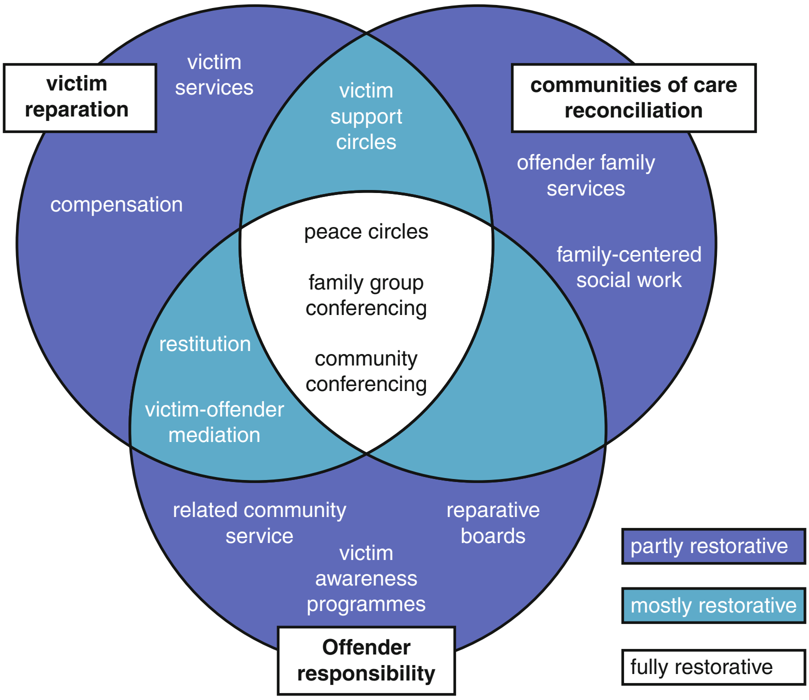
victim reparation
victim services
victim support circles
communities of care reconciliation
offender family services
compensation
peace circles
family group conferencing
community conferencing
family-centered social work
restitution
victim-offender mediation
related community service
reparative boards
victim awareness programmes
Offender responsibility
partly restorative
mostly restorative
fully restorative

Restorative Justice and State Crime, Fig. 1 Restorative justice typology (Source: McCold and Wachtel 2002)

As argued elsewhere (Parmentier et al. 2008), the classical way to deal with international crimes has been to punish them and to use the criminal justice system for this purpose. This route fits into a retributive justice approach, according to which punishment is considered a justified response to crime because those who commit offenses deserve punishment. Much more than in the case of common crimes, criminal prosecutions for international crimes may take place at various levels, national and international. From the outset it should be clear that retribution is not the monopoly of criminal prosecutions, as also civil damages to victims, or the lustration of persons having worked for a former authoritarian regime, constitute illustrations of retribution. For the sake of the argument, however, the next paragraphs discuss retributive justice in terms of criminal law and criminal justice only.

Strengths and Weaknesses of Criminal Prosecutions

Calls to bring the presumed offenders to a criminal court usually ring loud after a regime change has taken place, and the old authoritarian leaders have been replaced by a more or less democratic form of government. At that time, the political context tends to be more conducive to the necessary freedom of speech and of action. Moreover, the case of Pinochet has clearly illustrated that such calls for prosecution can be made long after the crimes have taken place. The former Chilean president and dictator was arrested in London in 1998 on charges of torturing political opponents during the early years of his dictatorship that ran between 1973 and 1990 but ultimately released by the British government on humanitarian grounds (old age and weak health).

The strong points of criminal prosecution for international crimes have been well documented. Huyse (1996) has listed two main arguments in these debates. The first is related to the reconstruction of the moral order, that is, the general idea that “justice be done” to satisfy the desire for justice of a society as a whole and of specific groups in particular. This moral argument is frequently coupled with a political one, in the sense that prosecution can also strengthen the fragile democracy, by confirming the principle of the “rule of law" and by thus providing the firm foundation on which to construct a human rights awareness and culture in the country. In such manner, criminal prosecutions serve the role of breaking through the thick walls of impunity and engage countries on the road to accountability. Closely linked to the reinforcement of the rule of law is the issue of deterrence, often voiced as a third argument. In this line of thought, to prosecute and punish perpetrators of international crimes will also serve to deter people from committing similar or other crimes in the future, both the perpetrator himself/herself but also the population at large.

On top of these two arguments that primarily pertain to the issue of the desirability of criminal prosecution, there are also aspects of legality involved. Orentlicher (2007) has argued in strong terms that there exists "a duty to prosecute in international human rights law" for serious human rights violations, founding her arguments on the various human rights treaties that contain passages to this effect and on the ensuing case law. Nevertheless, she also accepts the distinction between the "global norm" (criminal prosecution for the most responsible perpetrators) and the "local agency" of enforcing and interpreting this norm (including non-prosecution for many low-level offenders). In general legal terms, criminal prosecutions and convictions have the advantage of establishing a "legal or judicial truth," that is, an account of the acts and the facts that stands beyond doubt for the parties involved as well as for future generations. In the words of the South African Truth and Reconciliation Commission, this truth is mostly "forensic" or "factual" (TRC Report 1998), and it is of course focused on individual perpetrators and victims.

In a large number of cases, the calls for criminal prosecution are voiced by the direct victims of the violations or crimes, those who directly suffered from arrests, torture, convictions, and other forms of repression. Also their relatives, and their surviving family members, who have witnessed the crimes and have felt their consequences from close by, tend to be staunch defenders of criminal prosecutions. They usually wish to know what has happened to their beloved ones and to see the offenders be called to account. The examples of the survivors of the Holocaust or the Madres de la Plaza de Mayo in Argentina (the so-called Crazy Mothers who for 30 years demanded clarifications from the successive governments for the disappearances of their relatives during the years of the military junta between 1976 and 1983) speak for themselves. In other cases, victims conduct their activities with the support of human rights groups and committees, local but also international, such as Amnesty International, Human Rights Watch, and the Fédération Internationale des Droits de l'Homme.

It should be noted, however, that criminal prosecution is not without problems or even risks. Huyse (1996) has listed some of these risks. Contradictory as it may sound, criminal prosecutions of perpetrators may also undermine the "rule of law" of the new state. Some regimes may incur contradictions with the principle of nonretroactivity of criminal law, notably if they wish to prosecute crimes that were not prescribed or that were not punishable under the former regime. Another problem for the new state is to guarantee the independence and the impartiality of the criminal justice system. This is not always easy, because the criminal justice systems may still be populated by the same police offers and judges appointed by the former regime and still adhering to its values. The examples of Germany after World War II and South Africa after apartheid are telling in this respect. But also when the police officers and the judges are new to the system, problems may occur in that they may want to make a clear rupture with the past by adopting very repressive policies without due diligence for the rights of the accused. The reality of postwar repression in Western European countries has clearly illustrated the possible risks associated with such approach, which can lead to many controversies and tensions in the long run. And of course it is problematic to prosecute the offenders in the case of an amnesty rule that shields them off from further legal action, criminal and civil.

Another category of risks is less of a legal but more of a political nature. Many new democracies are fragile, as the old political and military elites can be resisting and actively opposing the transitional government. Prosecuting well-known offenders, or even the threat thereof, can provoke the old elites and even seduce them to seizing power again. The examples of Chile and Argentina are telling in demonstrating the power of the old elites and the caution with which the new government has proceeded. Another problem relates to the capacity of the system. In long-term autocratic regimes, only a small minority tends to possess the knowledge and the skills required to govern a country in political and economic terms. To call to account the members of this elite may lead to growing uncertainty about the future of the elites in general and may even push them to emigrate from the country, a scenario not unknown to countries from Central Europe in the early 1990s and to South Africa in the late 1990s.

Furthermore, there are problems with the logistics of criminal justice systems that are confronted with the legacy of mass violence. In some cases, the sheer numbers of offenders and potential suspects are so large that the system would become completely clogged if too many prosecutions were to take place. The example of Rwanda, where almost a decade after the genocide still over 100,000 persons were imprisoned, speaks for itself, and it has prompted the Rwandan government to look for new solutions and to revive the old conflict resolution model of Gacaca and to adjust it to dealing with the crimes of genocide (Penal Reform International 2002). Even in situations less extreme, it is unavoidable to be very selective and to only bring some perpetrators to the criminal court. Then a number of tough questions arise, namely, whether the prosecutorial agencies should aim at the heads and the planners of the crimes, who have ordered the crimes or were aware of them taking place, or whether they should target those who executed the orders and those who assisted them. And what about the so-called bystanders, who did not actively participate in committing the crimes but nevertheless witnessed them and in some cases may have benefited from the consequences?

Finally, criminal trials by their very nature are mostly focused on the offenders and on the rights of the accused and pay far less attention to the victims and to the harm inflicted upon them (Zehr 1990). This is the case with ordinary criminal trials, and it is by and large the same with international crimes and crimes of mass violence. This is not to say that victims are completely neglected but that their role in the criminal process is often reduced to that of witnesses or in other ways.

Given the many difficulties associated with criminal prosecution, many choices have to be made. Therefore, it does not come as a surprise that in practice criminal prosecution has often proved more the exception than the standard. Between the aspirations of criminal prosecutions and their realities, there exist many law and legal regulations, as well as several practical objections.

The Triptych of Criminal Prosecutions

States are traditionally well equipped to conduct criminal prosecutions against suspects of regular crimes, through the various institutions of their criminal justice systems (notably the police, the Public Prosecutor's Service, the trial courts, and the prison system or other measures for the execution of criminal sanctions). However, crimes of mass violence bring with them new challenges (Parmentier and Weitekamp 2007). One of the challenges lies in the political nature of the crimes, meaning that either the intent of the offenders or the object and context of their crimes relate to politics. Another distinctive criterion between common crimes and international crimes lies in the massive numbers of victims of the latter and sometimes in the large numbers of perpetrators as well. Therefore, criminal justice systems in many countries, particularly if poorly equipped, tend to be reluctant to engage in widespread prosecutions.

While in the past criminal prosecutions for international crimes were limited to the country where the crimes had been committed, the last two decades have witnessed two important shifts in this regard (Parmentier et al. 2008). The first shift relates to the development of so-called universal jurisdiction legislation in a number of countries. This allows third countries to prosecute and to try international crimes, without the existence of a link between the third country and the place where the crimes have been committed, the nationality of the offender, or the nationality of the victim. The main rationale lies in the fact that international crimes are considered so heinous that they not only affect the victims and the criminal justice system of the country where they took place but that they also affect humanity as such, for which reason also third countries put their criminal justice system at the disposal of the world community. Such cases of "pure" universal jurisdiction are in fact rare in today's world, the actual Spanish legislation and the former Belgian legislation (until 2003) being among the exceptions. In fact, a number of countries – mostly European – have a limited form of universal jurisdiction and require at least one specific link with the crime in order to prosecute and try it. While welcomed by some as an ethical triumph for humanity, recent events in Western Europe suggest that the efforts to establish genuine systems of universal jurisdiction are often subject to the *Realpolitik* of international relations.

The second major shift to respond to mass atrocity lies in the establishment of criminal justice mechanisms at the international level. The forerunners of this tendency were the two ad hoc tribunals, for ex-Yugoslavia (ICTY) and for Rwanda (ICTR), set up by the Security Council of the United Nations in 1993 and 1994, respectively, for a limited period of time and with a limited territorial jurisdiction. They have been followed by the establishment of a permanent International Criminal Court, entering into operation in 2002 and dealing with three main categories of international crimes, genocide, crimes against humanity, and war crimes. While the two ad hoc tribunals have a primary competence to deal with serious human rights violations, the ICC has a complementary task to prosecute and try international crimes when States Parties are "unwilling or unable" to do so, thus leaving the prime locus at the national level. It was followed by a number of mixed international-national tribunals, for Sierra Leone, East Timor, Kosovo, and Cambodia. The rapid development of international criminal justice mechanisms has been heralded by many as the final start of a new era of justice. Others, like Drumbl (2000), have argued that the practice of punishment by international tribunals is foremost an example of "legal mimicry," whereby every new institution tries to build on the experience of the previous one, despite the many different crime situations they are facing and despite the often "confusing, disparate, inconsistent, and erratic" sanctions they have applied thus far. He raises very serious questions about the fact that international criminal justice has been strongly imbued with a western conception of justice, which is liberal and legalistic, and may not be the only model to deal with these "extraordinary" crimes that involve many victims, many perpetrators, and a large group of bystanders – and all of these within a political context. Moreover, some mechanisms run the risk to be driven by economic and political motives and lack the credibility to engage in the restoration of broken relationships in the society.

Restorative Justice Mechanisms and Their Promises

While the revival of the (international) criminal justice is often regarded as the starting point of a new era for justice in the case of mass violence, the preceding section exposed some of the major problems and shortcomings of the retributive approach to post-conflict justice. Moreover, retributive justice objectives will continue to represent only one side of the "coin of justice." Given the many problems with prosecutions and retributive justice in general, recent years have arguably seen an increasing interest for other institutions and other models to deal with serious human rights violations and international crimes (Parmentier et al. 2008). The more than 40 truth commissions that have been established since the last quarter of the twentieth century are the best known examples of "second-best" solutions (Hayner 2011). Truth commissions have become strongly associated with restorative justice principles, and some like the South African one are among the few transitional justice mechanisms to be explicitly labeled as a restorative justice mechanism (Villa-Vicencio 2001). As a result there is increasing attention for the model of restorative justice and particularly for the applicability of restorative justice in situations of mass victimization and post-conflict justice (Weitekamp et al. 2006). Such approach requires "changing lenses," as aptly put by Zehr (1990) in his seminal work on restorative justice for common crimes.

In this second part, it is argued that the road to post-conflict justice requires further steps that move beyond the limits of retributive justice. By prosecuting and punishing perpetrators of mass violence, many aspects of the phenomenon of mass victimization – such as its societal, victimological, and especially its criminological relevance – are overlooked. To reach some of the key goals of post-conflict justice, namely, to prevent the reoccurrence of the violent conflict and to repair the harm that was suffered during the conflict, it is crucial to consider a number of questions that are criminological par excellence: What are the causes of mass violence? What were the conditions that made this violent conflict happen? What are the sociological and psychological reasons that could explain people's involvement in committing these atrocities? Here follows a short overview of how a restorative justice approach to post-conflict situations can offer a meaningful answer to these questions.

At this point, an important problem arises. Few concepts have been quoted more frequently in criminology during the last decade than the concept of restorative justice. At the same time, few concepts cover a more diversified meaning that this one, which is sometimes considered vague and fluid. The bottom line of restorative justice is to view crime as a violation of people and relations, thereby creating an obligation to make things right. This process should be facilitated by bringing victims and offenders together on a voluntary basis, such as in victim-offender mediation programs, and possibly also with the other stakeholders, for example, in restorative justice conferences. In these forums, room is made for dialogue and for creating an opportunity, with the help of a mediator, to restore the harm done and to reconcile the relation. For restorative justice scholars, crime is fundamentally a violation of people and interpersonal relationships, and restorative justice seeks to heal and put right the wrongs committed. In other words, restorative justice constitutes a different way of thinking about crime and our response to it. It focuses on the harm caused by crime and repairing the harm done to victims and reducing future harm by preventing crime. Therefore, need offenders assume responsibility for their actions and the harm they have caused? Restorative justice seeks redress for the victim, recompense by offenders, and reintegration of both within communities. This is achieved through cooperation by communities and the government. For Dignan and Marsh (2001), restorative justice has the following characteristics: the offenders' personal accountability to those harmed by an offense, an inclusive decision-making process incorporating the key players, and the goal of putting right the harm caused by an offense. However, looking for a meaning that is widely accepted and that can serve as the starting base for our analysis is obviously no sinecure. At least four approaches can be identified to come to terms with restorative justice: (1) looking at definitions, (2) studying criteria, (3) identifying basic principles, and (4) developing models (Fig. 2).

It is striking that developments in restorative justice are almost exclusively focused on less serious property crimes and on juvenile crime, with a limited number of isolated examples of victim-offender programs for serious interpersonal crimes (e.g., Umbreit et al. 2003). But thus far very little attention has been paid to restorative justice for crimes of a political nature, sometimes reaching the level of mass violence and mass victimization (Christie 2001). However, viewing restorative justice in such a way that it may complement retributive mechanisms in dealing with mass violence allows for an extension of the concept of governance of international crimes. Before addressing the question if and how it can be used in the case of

Restorative Justice and State Crime,
Fig. 2 Stakeholders needs (Source: McCold and Wachtel 2002)

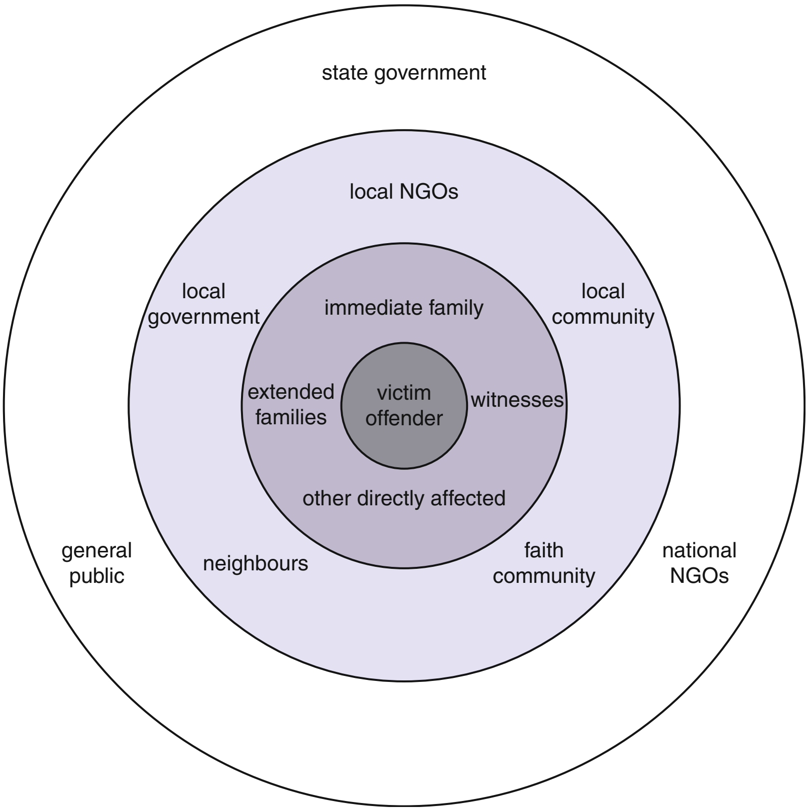

international crimes and state crimes, the meaning of restorative justice as developed in the area of criminology for common crimes has to be studied.

Definitions of Restorative Justice

Although no commonly accepted definition seems to exist, the literature makes frequent reference to two types of definitions. Marshall defines restorative justice as "a process whereby parties with a stake in a specific offence resolve collectively how to deal with the aftermath of the offence and its implications for the future" (Marshall 1996, p. 37), whereas Bazemore and Walgrave are goal-oriented as they formulate restorative justice as "every action that is primarily oriented towards doing justice by repairing the harm that has been caused by the crime" (Bazemore and Walgrave 1999, p. 48). Additional definitions abound in the broad and rapidly growing field of restorative justice.

Criteria of Restorative Justice

Definitions provided in the previous section remain somewhat general, and some would say quite vague. One way to make them more concrete is to enumerate specific criteria for restorative justice. The international community, personified by the United Nations Economic and Social Council (ECOSOC), has subscribed to the idea of restorative justice by adopting in 2002 a resolution encouraging countries to use Basic Principles on the Use of Restorative Justice Programmes in Criminal Matters to develop and implement restorative justice in their countries (www.restorativejustice.org).

The Basic Principles do not contain any formal definition of restorative justice, probably due to the fact that the text does not address the theory of restorative justice but is explicitly directed to particular programmatic expressions of restorative justice (Parmentier 2001). The text refers to "restorative justice programs," meaning a program that uses restorative processes or aims to achieve restorative outcomes, two pillars that are clearly distinguished. "Restorative processes" imply those whereby all parties "affected by a crime actively participate together in the resolution of matters arising from the crime, often with the help of a fair and impartial third party," as exemplified by mediation, conferencing, and sentencing circles. The second pillar, restorative outcomes, comprises agreements reached as the result thereof and explicitly includes restitution, community service, and other programs with an objective of reparation for victims and the community and reintegration of victims and offenders.

Rather than defining restorative justice, the Basic Principles enumerate various conditions under which restorative justice programs can and should actually work. These include, i.a., the availability of restorative justice programs at all stages of the criminal justice process, the voluntary character of restorative justice processes, the establishment of clear guidelines and standards governing the use of restorative justice programs, the importance of procedural safeguards in such processes, the issue of confidentiality, the importance of follow-up and implementation of agreements, the skills required for facilitators, and the importance of evaluation research. Through this focus on the working conditions of restorative justice programs, this text may be seen to mark the start of a more "conditional" approach to the problem of a common understanding of restorative justice.

Principles of Restorative Justice

A third way to understand the meaning of restorative justice is to identify the key principles that are underlying. By way of example, Roche in his important book (2003) has identified four such principles: (1) personalism – crime is a violation of people and their relationships rather than a violation of (criminal) law; (2) reparation – the primary goal is to repair the harm of the victim rather than to punish the perpetrator; (3) reintegration – the aim is to finally reintegrate the perpetrator into society rather than to alienate and isolate him/her from society; and (4) participation – the objective is to encourage the involvement of all direct and possibly also indirect stakeholders to deal with the crime collectively. These principles allow for a close study of the restorative character of procedures and institutions.

Models of Restorative Justice

Finally, it can be argued that restorative justice can be conceptualized by developing various models. One very useful typology was developed by McCold and Wachtel (2002) who talk about three "degrees of restorativeness," namely, "fully," "mostly," and "partly" restorative justice practices. These degrees depend on the degree of involvement of three actors, victims, perpetrators, and communities. As can be seen from the figure below, the authors consider family group conferences, peace circles, and community conferencing to be fully restorative while all other practices are only mostly or partly restorative.

The Relevance of Restorative Justice for International Crimes: Looking Through the Lens of Truth Commissions

Having sketched various approaches to restorative justice for so-called ordinary or common crimes, this contribution moves to indicating how restorative justice can be understood in the domain of international crimes, most of which are committed by states and their agencies. Rather than looking at all sorts of mechanisms that display some restorative features in dealing with international crimes, it is justified to focus on one specific mechanism, namely, truth commissions. In the following paragraphs, truth commissions will be looked upon, and it will be indicated how the principles and the models of restorative justice can be applied to them. At the same time, it should be noted that not all truth commissions share exactly the same features and therefore not all live up to the criteria of restorative justice (Villa-Vicencio 2001).

According to Hayner (2011), the following elements have to be present to speak of a truth commission: (1) it has to be set up with public (meaning state or government) support and cannot be the result of private initiatives only;

(2) its objective is not to focus on individual cases but to sketch an overall pattern of human rights violations; (3) its major focus is on the victims who are provided ample opportunities to share their experiences of the crimes committed; (4) it finishes its work by a final report that is given publicity; and (5) the report contains recommendations about how to deal with the legacy of a dark past and how to avoid similar conflicts and crimes in the future. From the above it follows that a truth commission is quintessentially a nonjudicial body, set up for a limited duration, and without the intent to establish the guilt or innocence of individual persons but more to contribute to a form of conflict settlement between parties and at the level of society.

Such commissions of inquiry are increasingly considered as a valuable and a complementary practice to civil or criminal courts because of their strong emphasis on truth, reparation, and reconciliation (Christie 2001). The restorative dimension is particularly present at two levels: first, at the institutional level of the commission, as it provides a public forum for victims and offenders to voice and share their experiences, and secondly, at the interpersonal level where individual victims and offenders can meet during or after the process of the truth commission with a view to dialogue, personal healing, or restoration on the long term. Again, not all truth commissions adhere to the same operational rules, and there it is important to conduct specific case studies to reveal in which way and to which degree truth commissions fulfill their promise of providing restorative justice (Parmentier 2001). The South African Truth and Reconciliation Commission (TRC) provides an interesting example of how such commissions may contribute to restorative justice, in this case through the work of its three separate committees (TRC 1998): first, by providing a forum to victims during the public hearings organized throughout the country as part of the Human Rights Violations Committee, which in limited cases have led to an encounter between victims and suspects or representatives of the apartheid regime; second, through actual encounters between offenders applying for amnesty to the Amnesty Committee and therefore being obliged to disclose all relevant facts, which in some instances led to a dialogue with the victims or survivors; and third, through the recommendations of the Reparation and Rehabilitation Committee that issued recommendations to the government on matters of reparation and rehabilitation.

Other examples of mechanisms with restorative justice aspects include customary mechanisms, also called community-based mechanisms, such as the Gacaca tribunals in Rwanda, set up to process the high numbers of alleged perpetrators of genocide in detention and with the objective to strike a balance between justice and reconciliation (Penal Reform International 2002).

Principles of Restorative Justice and Truth Commissions

Roche's four principles of restorative justice can easily be applied to truth commissions and particularly to the one in South Africa (Parmentier et al. 2008).

First of all, the principle of personalism, meaning that crime is a violation of people and their relationships rather than a violation of (criminal) law, finds application. This principle refers to the social dimension of "emotional involvement" that enables the conflicting parties to restore the broken relationships. Crimes of mass violence are first of all violations of people rather than violations of law. Truth commissions do bear the promise of being a "fully restorative justice process" if they stress the importance of encounters between victims and offenders and are also able to live up to these expectations (Weitekamp et al. 2006). One of the implications of such encounter is to organize victim-offender dialogues, either as part of the truth commission process or outside of its direct ambit. In that case, the key issue of "encounter" should be included as a new element in future truth commissions. To date no single truth commission has gone as far as bringing perpetrators and victims together to deal collectively with the aftermath of the violent conflict, with the possible exception of some amnesty hearings in the South African TRC. In the South African case, and presumably in other post-conflict nations as well, there is quite a lack of information on how and, more importantly, to what extent victims' expectations were effectively met through the TRC process, and very little is known about their backgrounds, expectations, and motivations. As to the offenders, it is already shown that if they are brought around the table, it might help to reverse the process of dehumanization of the perpetrator (Drakulic 2004). On a more general level, van der Merwe has illustrated how the national plan for reconciliation in post-apartheid South Africa has led to efforts to obstruct and to hamper the individual process of reconciliation for many victims (van der Merwe 2001). Finally, it should be noted that there exist many more legitimate responses to mass victimization after violent state conflict that are culturally and socially appropriate responses attempting to bring people together in some way. One of the examples could be the Gacaca tribunals in Rwanda (Penal Reform International 2002). Responses could also include a blend of mechanisms that form an integrated response to deal with the past.

The second principle of reparation, the primary goal of which is to repair the harm of the victim rather than to punish the perpetrator, is more difficult to apply straightforwardly. Reparation is becoming increasingly important to address, and even undo, the injustices of the past, and the last decade has witnessed an enormous increase of the awareness in this regard. According to the UN resolution of 2005 holding the Basic Principles and Guidelines on Reparation, reparation is to be understood as a very wide concept, including restitution of goods, financial compensation, rehabilitation through social and medical measures, symbolic measures, and guarantees of non-repetition of the alleged acts. All of these measures can be individual or collective. While the principle is now firmly established in international law (Sarkin 2004), many questions do remain (Du Plessis and Peté 2007). Although it is generally accepted that the beneficiaries of reparation are the victims, it is less clear how far to stretch this category and whether also indirect victims or society as a whole can be included. Another issue is how to enforce the right to reparation, through a general government policy or through individual administrative or judicial action? Apart from preventing conflicts to reoccur the restorative justice principle of reparation is a primary goal of transitional justice. For a long time in the debates on transitional justice, reparation was promulgated as a necessary mechanism, but only a few examples have shown a successful outcome. In the years to come, reparation is expected to rise high on the transitional justice agenda. In the context of the South African TRC, reparations were present at various levels (TRC Report 1998): the Reparation and Rehabilitation Committee awarded urgent and interim reparations to those victims who had provided written or oral testimony to the TRC and were in need of medical, psychological, or material help; it also recommended to the government to take measures of various nature, including individual reparations, symbolic reparations, and legal and administrative matters, community rehabilitation programs, and institutional reforms. It has to be said that the recommendations have not led to the expected results on the part of the government, and many victims have tried to force the government to implement these recommendations by bringing their claims to the collective level of victim support groups, direct lobbying, and even court cases. It remains to be seen what the concrete outcomes of such actions have been (Fig. 3).

The principle of reintegration, meaning that the aim is to finally reintegrate the perpetrator into society rather than to alienate and isolate him/her from society, poses another set of difficult problems. In retributive justice the focus is on the accountability of offenders, namely, calling to account those who have committed gross violations of human rights that sometimes amount to international crimes. In contrast to common understanding, accountability is far from a straightforward reality. One of the main problems relates to the type of offenders to be called to account, the heads and the planners of the violations or those who executed the orders and those who assisted them, and what to do with the "bystanders" who did not actively participate in committing the crimes but may have benefited from the consequences? On top of that, it should be clear that international crimes only constitute

Restorative Justice and State Crime, Fig. 3 Restorative transitional justice typology (Source: Parmentier et al. 2008)

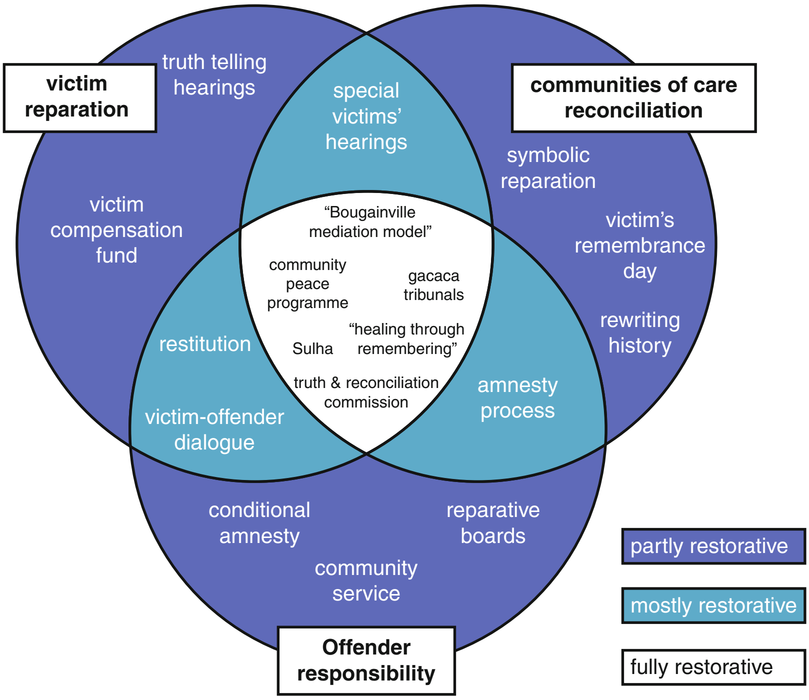
victim reparation
truth telling hearings
special victims' hearings
communities of care reconciliation
symbolic reparation
victim compensation fund
"Bougainville mediation model"
community peace programme
gacaca tribunals
victim's remembrance day
rewriting history
restitution
Sulha
"healing through remembering"
truth & reconciliation commission
amnesty process
victim-offender dialogue
conditional amnesty
reparative boards
community service
Offender responsibility
partly restorative
mostly restorative
fully restorative

the symptoms of a violent conflict. Dealing with the most serious crimes without dealing with the underlying conflict itself is merely touching the tip of the iceberg. A restorative justice approach to accountability is necessarily different, particularly when coupled with the key principle of reintegration. This principle demands from a society that it aims to hold perpetrators accountable for their wrongdoings in a supportive way, taking into account all the needed measures to work toward their reintegration into the same society. The merely punitive approach of trials leaves little room for future reintegration, and this same argument applies to states that choose not to face their violent past and decide to live in a permanent state of denial (Cohen 2001) or amnesia. A restorative justice approach to the issue of accountability is geared toward cooperation with the perpetrator, to resolve problems and to reintegrate rather than to alienate or isolate the perpetrator from his/her society. Providing a high level of support to perpetrators while not minimizing the control over the accountability process could arguably provide the most sustainable and durable response to dealing with perpetrators of serious human rights violations. Another aspect going to the heart of reintegration relates to the beneficiaries of international crimes and serious human rights violations. There is a great deal of controversy around the issue of involving the benefiting parties in the process of accountability and reparation. The experiences of the South African TRC on the (lack of) involvement of beneficiaries show the sensitive character of this matter. The amnesty hearings of the TRC focused only on individual perpetrators, and the institutional hearings with large sectors of society (e.g., business, media, and medical and legal profession) were purely voluntary. In this way, the structural evil of apartheid as a system was largely downplayed in favor of the misdeeds committed by some individuals. According to some, the beneficiaries of apartheid were let off the hook without being obliged to confront their past in a direct way (Mamdani 2000) (Fig. 4).

Restorative Justice and State Crime,
Fig. 4 Stakeholders in the case of international crimes (Source: Vanspauwen, 2003)

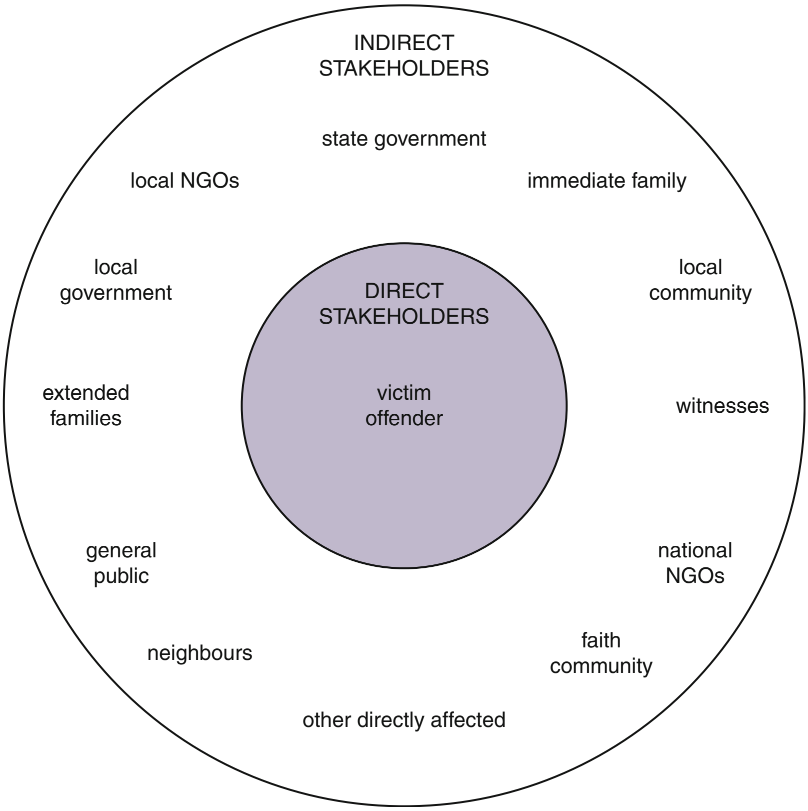
INDIRECT STAKEHOLDERS
state government
local NGOs
immediate family
local government
DIRECT STAKEHOLDERS
local community
extended families
victim
offender
witnesses
general public
national NGOs
neighbours
faith community
other directly affected

And fourthly, the principle of participation, whereby the objective is to encourage the involvement of all direct and possibly also indirect stakeholders to deal with the crime collectively, is applied. This principle refers to the dimension of empowerment and stresses that those affected by the crime need to regain their sense of autonomy. This can only be achieved when the stakeholders are actively involved in the process, and we distinguish two categories of stakeholders: those directly affected by the crime and those indirectly involved but not emotionally affected. In post-conflict situations, it is an enormous task to clarify the different roles and stakeholders that were affected either directly or indirectly and either personally or structurally by a violent past whereby gross violations of human rights were committed by the state on the one hand and by groups of the society (e.g., liberation movements, activists, minorities) on the other hand. Very few people will appear to be not affected at all, so that dealing with mass victimization in post-conflict situations will seem to be an insurmountable task to fulfill. Overall, restorative justice pays a lot of attention to victims and their needs, but a problem may arise when the group of victims is systematically narrowed. In the process of the South African TRC, for example, many victims could not be taken into account because they were not "victimized enough" according to the mandate. Only victims of gross human rights violations were defined as victims in the TRC Act. For a huge number of victims, there seems not to have been a feeling of justice at all, notably the victims of structural apartheid and the victims who suffered from the everyday apartheid policy. A lot of victims were disappointed and embittered after the proceedings of the TRC and felt that their rights and feelings had been neglected and justice was not gained. In this respect, it is also important to consider the distinction between individual and collective victims, direct and indirect victims, and see to what extent we can broaden up the definition of what victims are so that they can all have their place they deserve in the process of transition. In other words, while the process of truth seeking in the TRC has gone far beyond forensic forms of truth in individual cases and has involved narrative and social truth in a collective framework, it has still fallen short of allowing the participation of a wide range of victims of apartheid. Not all stakeholders were present in all parts of the procedures. They were mostly present in the hearings of the amnesty committee (AC), which always involved the amnesty applicants and the public and sometimes also the victims or their relatives. In the Human Rights Violations Committee hearing, the same three parties were sometimes present but sometimes not (only victims and the public). The picture is therefore mixed, moreover so because at the end of the day, only 22,000 victims in one way or another participated in the TRC's work, which in light of the many millions only constitutes a tiny fraction.

Models of Restorative Justice and Truth Commissions

Another way to assess the restorative justice aspects of truth commissions is to apply the model of restorativeness as developed by McCold and Wachtel. Using the three basic actors, perpetrators, victims, and communities, it is possible to reinvent their model into a "restorative transitional typology" to fit the purposes of international crimes and post-conflict justice (Parmentier et al. 2008). As a result, six practices could be considered as fully restorative, including truth commissions (under the condition they allow encounters between offenders and victims), community peace programs (if they allow sufficient space for victims and offenders), Gacaca tribunals (provided they reserve a proper place to the community), the mediation model applied in Bougainville (Braithwaite 2002; see infra), the practice of Sulha (the middle eastern model of doing justice), and healing through remembering programs in various parts of the world. It goes without saying that much more research is needed, particularly in the countries concerned, to find out all the details of these practices.

Toward an Integrated Approach of Retributive and Restorative Justice in Post-conflict Situations

Having sketched the strengths and weaknesses of retributive justice models in relation to international and state crimes, and having indicated the features of restorative justice and its potential to be applied for such crimes, one crucial question comes up consistently: how to understand the relationship between both models of justice in the case of international crimes and in the context of post-conflict justice?

Much has been written about this one-million-euro question, and it would go far beyond the limits of this entry to deal with all possible answers in the literature. Instead, it is focused on the theoretical work of the very influential work of Australian criminologist John Braithwaite (2002). He has consistently argued that it is always feasible (and preferable) to adopt a restorative justice approach at the start and try to uphold it as long as possible. At the same time, the level of pressure and coercion can be gradually increased, and if nothing convinces the stakeholders to enter into a restorative practice, it is possible to go back to the retributive approach. In other words, while restorative justice is the preferred model to deal with conflicts and crimes, a certain degree of coercion may be unavoidable. For Braithwaite, this theory works both in the case of common crimes (the "restorative regulation pyramid" below) as in the case of violent conflicts involving international and state crimes (the "responsive regulatory pyramid of international diplomacy" below) (Figs. 5 and 6).

The theory provides a very useful and promising framework to conceptualize the relationships between retributive justice and restorative justice. In the context of international crimes and post-conflict justice, it provides a better understanding of how both models of justice could work together and even complement

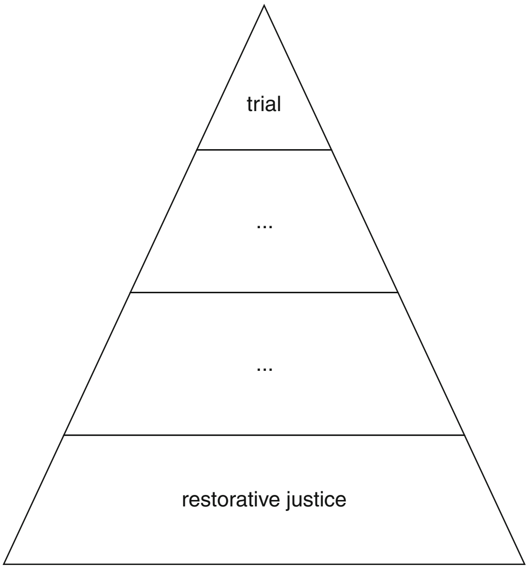

Restorative Justice and State Crime, Fig. 5 Restorative regulation pyramid (Source: Braithwaite 2002)

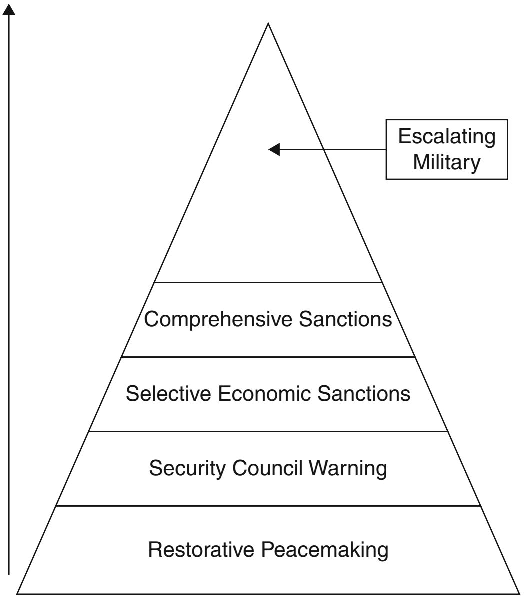

Restorative Justice and State Crime, Fig. 6 Responsive regulatory pyramid of international diplomacy (Source: Braithwaite 2002)

one another. But of course, much more research is needed, both of a theoretical and an empirical nature, to grasp the full details of such complex relationship.

Others have tried to gain further insight into this matter by developing two models of post-conflict justice. The first analyzes the building blocks of post-conflict justice and identifies seven key issues for any government to deal with in the aftermath of violent conflicts: truth, accountability, reparation, reconciliation, trauma, trust, and dialogue (Parmentier and Weitekamp 2007; Weitekamp and Parmentier 2012). The second model relates to the interplay between retributive and restorative justice instruments that can be used in national and international interventions after violent conflicts have come to an end (Weitekamp et al. 2006). The interested reader is referred to the literature contained therein.

Conclusion

In the aftermath of violent conflicts and the commission of international crimes, often by state agencies, the dominant approach is to start criminal prosecutions against the alleged perpetrators. While prosecutions take place at the national and the international level, many problems continue to abound in terms of capacity, impartiality, and the like. Moreover, little if any attention is paid to a restorative justice approach in dealing with international and state crimes. One of the reasons is the multiple understanding of what restorative justice really means and implies.

This contribution has highlighted various ways of viewing restorative justice and has applied some of these to the field of international crimes, by looking through the lens of truth commissions that have emerged in international relations over the last 30 years. It can be argued that restorative justice is a useful approach to consider in cases of international crimes. At the same time, the relationship between retributive and restorative justice is far from clear and in fact very complex. More research is needed, theoretical and empirical, to develop sound models for understanding and for intervening during and after violent conflicts.

Related Entries

- ▶ History of Restorative Justice
- ▶ Monitoring and Evaluation of Restorative Justice
- ▶ Political Crime
- ▶ Post-Conflict Traditional Justice
- ▶ Preventive Order: Civil Law
- ▶ Restorative Justice and Practice
- ▶ Transitional Justice
- ▶ Victims and Restorative Justice
- ▶ Victims of State Crime

Recommended Reading and References

Balint J (1996) Conflict, conflict victimization, and legal redress: 1945–1996. Law Contemp Probl 59:231–247

Bassiouni MC (ed) (2002) Post-conflict justice. Transnational Publishers, Ardsley

Bazemore G, Walgrave L (1999) Restorative juvenile justice: in search of fundamentals and an outline for systemic reform. Criminal Justice Press, Monsey

Braithwaite J (2002) Restorative justice and responsive regulation. Oxford University Press, Oxford

Christie N (2001) Answers to atrocities. Restorative justice in extreme situations. In: Fattah E, Parmentier S (eds) Victim policies and criminal justice on the road to restorative justice. Essays in honour of Tony Peters. Leuven University Press, Leuven

Cohen S (2001) States of denial. Knowing about atrocities and suffering. Polity Press, Cambridge

Dignan J, Marsh P (2001) Restorative justice and family group conferences in England: current state and future prospects. In: Maxwell G, Morris A (eds) Restoring justice for juveniles: conferences, mediation and circles. Hart Publications, Oxford

Drakulic S (2004) They would never hurt a fly. Abacus, London

Drumbl MA (2000) Punishment, post genocide, from guilt to shame to civis in Rwanda. New York U Law Rev 75:1221–1326

Du Plessis M, Peté S (eds) (2007) Repairing the past? International perspectives on reparations for gross human rights abuses, vol 1, Series on transitional justice. Intersentia Publishers, Antwerp/Oxford

Hayner PB (2011) Unspeakable truths. Confronting state terror and atrocity. Routledge, New York

Huyse L (1998) Justice after transition: on the choices successor elites, make in dealing with the past. In: Jongman A (ed) Contemporary genocides. PIOOM, Leiden

Mamdani M (2000) A diminished truth. In: James W, van de Vijver L (eds) After the TRC. Reflections on truth and reconciliation in South Africa. David Philip Publishers, Claremont

Marshall TF (1996) Restorative justice. An overview. Home Office: Research Development and Statistics Directorate, London

McCold P, Wachtel T (2002) Restorative justice theory validation. In: Weitekamp E, Kerner H-J (eds) Restorative justice: theoretical foundations. Willan Publishing, Devon, pp 110–142

Orentlicher D (2007) Settling Accounts' revisited: reconciling global norms with local agency. Int J Transit Justice 1:10–22

Parmentier S (2001) The South African truth and reconciliation commission. Towards restorative justice in the field of human rights. In: Fattah E, Parmentier S (eds) Victim policies and criminal justice on the road to restorative justice. Essays in honour of Tony Peters. Leuven University Press, Leuven

Parmentier S, Weitekamp EGM (2007) Political crimes and serious violations of human rights: towards a criminology of international crimes. In: Parmentier S, Weitekamp EGM (eds) Crime and human rights, Series on sociology of crime, law and deviance. Elsevier, Amsterdam, pp 109–144

Parmentier S, Vanspauwen K, Weitekamp EGM (2008) Dealing with the legacy of mass violence: changing lenses to restorative justice. In: Smeulers A, Havemann R (eds) Supranational criminology: towards a criminology of international crimes. Intersentia, Antwerp/Oxford, pp 335–356

Penal Reform International (2002) Interim report on research on Gacaca jurisdictions and its preparations (July–December 2001). Penal Reform International, London

Roche D (2003) Accountability in restorative justice. Oxford University Press, Oxford

Rothe D, Mullins C (eds) (2010) State crime: current perspectives. Rutgers University Press, Piscataway

Sarkin J (2004) Pursuing private actors for reparations for human rights abuses committed in Africa in the courts of the United States of America. In: Doxtader E, Villa-Vicencio C (eds) Repairing the irreparable. Dealing with the double-binds of making reparations for crimes in the past. David Philip, Claremont

Truth and Reconciliation Commission of South Africa (1998) Truth and reconciliation commission of South Africa report. Juta Publishers, Cape Town, 5 volumes

Umbreit MS, Bradshaw W, Coates RB (2003) Victims of severe violence in dialogue with the offender. Key principles, practices, outcomes and implications. In: Weitekamp EGM, Kerner HJ (eds) Restorative justice in context. International practice and directions. Willan Publishing, Cullompton

van der Merwe H (2001) National and community reconciliation: competing agendas in South African truth and reconciliation commission. In: Biggar N (ed) Burying the past: making peace and doing justice after civil conflict. Georgetown University Press, Washington, DC

Villa-Vicencio C (2001) Restorative justice in social context: the South African truth and reconciliation commission. In: Biggar N (ed) Burying the past: making peace and doing justice after civil conflict. Georgetown University Press, Washington, DC

Weitekamp EGM, Parmentier S (2012) On the road to reconciliation; the attempt to develop a theoretical model which applies restorative justice mechanisms in post-conflict societies. In: Plywaczewski E (ed) Current problems of the penal law and criminology. Wolters Kluwer Polska, Warszawa, pp 795–804

Weitekamp EGM, Parmentier S, Vanspauwen K, Parmentier S, Vanspauwen K, Valinas M, Gerits R (2006) How to deal with mass victimization and gross human rights violations. A restorative justice approach. In: Ewald U, Turkovic K (eds) Large-scale victimization as a potential source of terrorist activities. Importance of regaining security in post-conflict societies. Ios Press, Amsterdam, pp 217–241

Zehr H (1990) Changing lenses. A new focus for crime and justice. Herald Press, Scottdale

Restorative Justice, Victims

► Victims and Restorative Justice

Retribution

► Retributivist Theories

Retributive Punishment

► Retributivist Theories

Retributivist Theories

Christopher Bennett
Department of Philosophy, University
of Sheffield, Sheffield, UK

Synonyms

Retribution; Retributive punishment

Overview

This entry looks at recent retributivist theories that draw on denunciation and the expression of moral emotions in order to justify punishment. After setting out some of the canonical sources of the retributivist tradition, and explaining some of the most serious objections to this tradition, it looks at how these recent developments seek to overcome the objections while preserving what seems most of value in retributive ideas. The entry identifies work by P. F. Strawson and Patrick Devlin as the starting-point of these developments. In different ways, Strawson and Devlin seek to vindicate the idea that punishment should express our sense of the moral seriousness of crime as wrongdoing. Furthermore, they imply that punishment is necessary to do justice to the moral seriousness of wrongdoing. The entry considers to what extent this line of argument, if more fully developed, might offer a successful defense of retributivism. It also gives a survey of some of the major recent contributions to this field, including Jeffrie Murphy, Andrew von Hirsch, and R. A. Duff.

Introduction

Three important developments in recent thinking about the justification for punishment are: the rise of communicative theories of punishment; renewed interest in the moral emotions; and the return of retribution. This entry concerns the way the first two have informed the third. After a long period of neglect, during which it was thought to be a barbaric remnant of superstitious thought that more rational forms of society would sweep away, retribution has, over the past 40 years, returned to the fore as a rationale for punishment. There are, of course, various social and political factors that have contributed to this change. But a full explanation of the trend would surely have to refer to the fact that theorists over the past 40 years have discovered new and prima facie attractive ways of construing retributive ideas. One claim, associated with early work by Herbert Morris (1968) and Jeffrie Murphy (1973) among others, is that retributive punishment is necessary to restore the fair balance of benefits and burdens in a system of social cooperation. However, this approach has always faced the charge that it reduces all retributive action to action aimed at rectifying unfair advantage, and therefore cannot capture the sense that there is something about wrongs such as murder, rape assault, and willful neglect that call for retributive responses. A different strand of the new thinking about retributivism, on the other hand, has concentrated on the connections between retribution, moral condemnation, and the moral emotions. Wrongdoing is something that we often feel strongly about – and such feelings can, it is said, (sometimes at least) be justifiable. It is this latter strand of retributivism that is the subject of this entry.

What Is Retributivism?

Whether there is something yet living in the retributive tradition depends on whether, on due reflection, at least some of its canonical ideas continue to strike us, under a certain interpretation, as moral insight rather than illusion.

One central source for this tradition is the idea, common to Judaism, Christianity, and Islam, that there is one perfectly just God who will, at some ultimate point of judgment, reward the virtuous with perfect happiness and punish the vicious with eternal suffering. Immanuel Kant captures this thought by claiming that the highest conceivable or possible good is the "distribution of happiness in exact proportion to morality [i.e. to the goodness of one's will]" (Kant 1996, p. 135). On this view, the purpose of retributive action is to help bring about a perfectly just state of affairs. Such action may be seen, in the final analysis, as the business of the Deity. However, it is important to recognize that this is not the only way to think about retribution.

Another source of the retributive tradition is deontic or act-focused, and thinks of justice as requiring that transgression of authoritative limits on conduct must be answered by a like, or at any rate commensurate, infringement inflicted on the transgressor. For instance the claim, from the Book of Exodus, that a wrongdoer should give "life for life, eye for eye, tooth for tooth, hand for hand, foot for foot, burning for burning, wound, for wound, stripe for stripe" (often called the lex talionis) can be interpreted in this light. A similar idea can be found in Pre-Socratic Ancient Greek thought (for some discussion, see Nussbaum 1993).

Thirdly, the retributive tradition contains ideas about moral contamination or pollution. For instance, we might think of the forbidding claim made by Immanuel Kant that, were a civil society to disband, it would be necessary for the last murderer in prison to be executed before the disbandment should take place, to prevent "blood-guilt" from infecting each of the departing members (Kant 1991, p. 142). The blood-guilt presumably affects the murderer most directly. But, secondly, the contamination can be passed on to otherwise innocent parties if they do not do their duty in visiting proper justice on the murderer. Again, a similar conception is present in Ancient Greek thought, where the failure of Thebes to take action against their king Oedipus has dreadful consequences for all.

Fourthly, the retributivist can lay claim to the ideas of repentance and atonement common again to a number of religious traditions (Etzioni and Carney 1997). The idea of at-one-ment suggests a moral fracture in the offender, a separation from God or the Good, that is brought about by sin or wrongdoing. The wrongdoer needs to be made whole again, to overcome her separation from what is most important in her life (i.e., God or the Good) through sincere repentance and atonement. One prominent development of this set of ideas is the Roman Catholic sacrament of penance, in which penitents undertake something otherwise harmful or onerous in order to make amends for sin.

What each of these sources has in common is the view that it is inherently right or good that something negative should be suffered by those who do wrong. One feature commonly said to be distinctive of retributivism is that it is a backward-looking theory: It takes the importance of the wrong committed to be the central reason for punitive action. We can understand better what is meant by this if we contrast it with forward-looking approaches such as deterrence. According to a deterrent theory, the reason we punish is, not the moral nature of the crime itself, but rather the need to prevent future crimes: Punishment is necessary because (and insofar as) it produces some future good. The forward-looking view gives a clear explanation of why punishment is important, but also makes its justification derivative and conditional. Whether punishment is justified depends on whether it really is an effective means for the production of the future good we have in mind. On the backward-looking view, however, there is something about punishment that is inherently fitting for answering, or annulling, or avenging, etc., the wrong. Thus, an aspect of the retributive view is that some acts derive their importance and necessity, not from the future benefits that they bring about, but rather from their role in doing justice to what has gone before.

In addition to its backward-looking focus, another distinctive feature of retributivism is that it places some importance on harm suffered by the offender as a result of the wrongdoing. Thus, according to one prominent recent characterization, retributivism centers on “the Desert Thesis”: “that when a person has done something that is morally wrong it is morally better that he or she should suffer some loss in consequence” (Scanlon 1999, p. 274). If we put these two features of retributivism together, we come to the distinctive retributivist claim that it is morally important or necessary that wrongdoing or evil should lead to some response that involves the offender suffering in some way.

Serious Objections to Retributivism

The shape of any contemporary retributivist theory will be dictated by the need to preserve what seems most attractive in the retributive tradition, while at the same time avoiding or answering the main objections that have been raised to retributive action. We turn to look at some of these main objections now.

First of all, critics often object to retribution as a rationale on the grounds that it does not bring about any good. This is connected with its backward-looking orientation. It might be claimed that retributive action does not actually achieve anything. Indeed, the situation is worse than this, because retribution involves glorying in suffering that seems to be inflicted for its own sake. Therefore retribution can look redundant at best, and at worst, downright sadistic.

In response to this, however, the retributivist should deny that nothing is achieved by retribution (Murphy 1971). The retributivist thinks that sometimes, it is necessary to take action to do justice to a past act, independently of future benefit. Actions of gratitude, or reward, or mourning seem to share this feature. What such actions achieve is an appropriate response to that past action. Therefore, the retributivist should resist the claim that retributive action is redundant. And if the suffering of the offender is part of such appropriate response, satisfaction at suffering is not mere sadism.

Nevertheless, there is clearly an important question to be answered as to what makes the suffering of the offender the appropriate response to wrongdoing. The suspicion will be that the retributivist is influenced by ideas about moral pollution. But those ideas might be said to be part of a world view that we have long outgrown (see, e.g., Smart in Smart and Williams 1973 on guilt as a stain on the soul).

A second major line of criticism looks at the problem of free will. For retributive responses to be appropriate, it would have to be the case that perpetrators of wrongdoing acted freely. If the apparent wrongdoing was not really the wrongdoer’s fault, then they cannot deserve to suffer punishment in response. But, the objection goes: Our understanding of the basis of human action is now sufficiently advanced that we can see that no action is entirely freely chosen. Human beings are complex biological mechanisms; they act in social conditions they have not chosen, on the basis of culturally constructed understandings. Where is the room for free will in all this?

We will look in more detail at responses to this criticism below. However, it is worth saying at this point that empirical evidence about the basis of human action will not by itself refute moral responsibility. In order to assess the justice of this criticism of retributivism, we would have to look first of all at which forms of freedom are necessary for moral responsibility. This is not an empirical question, but a philosophical and moral question about the conditions under which it is fair and right to hold someone to account for what she has done. The absence of some sorts of freedom does indeed excuse or exempt from blame. But which? Only once we have answered that question can we turn to the empirical evidence and look at whether, given everything we know about human beings, it is plausible to think that we are free in the ways moral responsibility requires. Therefore, an assessment of the free will criticism depends on some understanding of what is going on in retributive action, and what forms of unfreedom make retributive action invalid.

Some other common criticisms of retributivism are worth mentioning at the moment. For instance, if we are committed to retribution, to what extent should we take into account the extent to which someone has already suffered in their life in assessing how much they presently deserve? More broadly, and leading on from this, is it not the case that the goal of apportioning happiness to personal virtue (and suffering to personal evil) requires a degree of knowledge of people's character and motivation that we routinely lack (perhaps it is no accident that the task of doing justice is often allotted to an all-knowing God). One way to respond to these problems is for the retributivist to reject a character-based concern with personal virtue and vice as a whole (the first source considered above), and argue instead for an act-focused form of retribution as a response to transgression (the second source). However, even the act-focused form of retributivism will have to address the question of how much suffering is proportionate to which crime. Once we reject the notion of literal equivalence between punishment and crime – for instance, on the grounds that it would have us raping rapists or torturing torturers – the notion of proportionality becomes more mysterious.

Another type of criticism asks whether retribution is an important, or even legitimate, state purpose. This is compatible with allowing that some form of retributivism may be morally important. But if we think of how much money is spent on criminal justice, we have to ask whether retribution gives good value for that money. If prison is necessary to bring about security of person and property, it may be money well spent. But why is it important to spend the money on something backward-looking? A related criticism is that if retributivism were the rationale for state punishment, then the state would be committed to punishing for all moral wrongdoing. So, for instance, lying, adultery, sexual perversions etc.: If any of these are moral wrongs then, the objection claims, the retributive state would punish for them. Yet this would be illiberal: It is characteristic of the value given to freedom by the liberal state that at least some moral wrongs are viewed as private and irrelevant to the legitimate concerns of law. In short, retributivism seems allied to legal moralism – the claim that law should sanction immorality as such – whereas the liberal position, it might be said, rejects this approach.

Retribution, Denunciation, and Emotion

Having looked at some of the sources of the retributive tradition, and some of its most serious objections, we can now turn to recent developments that seek to show how retributivism might yet have the resources to answer these concerns. It is central to the approach to retributivism that we will look at here that retribution is not only a rationale for punishment, but also a feature of our emotional life in a range of interpersonal relationships. Thus, according to this line of thought, there is a deep connection between retributivism in punishment and our intuitive ways of dealing with one another in everyday life. In one way, this represents a strength of retributivism over, e.g., the view that punishment is justified as deterrence or incapacitation. After all, one of the reasons why people maintain and support an institution of punishment might well be that it reflects something important about the way they think wrongdoers should be treated in nonlegal cases, for instance, when one person betrays or lies to or steals from or unjustifiably harms another. On the other hand, retributivists may be vulnerable to the criticism that they fail to recognize the extent to which legal punishment in a diverse modern society differs from blame in intimate relationships.

We can begin by identifying two interventions that helped to initiate the current revival in thinking about retributivism, condemnation, and the emotions: P. F. Strawson's "Freedom and Resentment" (1962) and Patrick Devlin's "The Enforcement of Morals" (1959). The latter initiated a debate with H.L.A. Hart (1963), one result of which was the publication of Joel Feinberg's highly influential "The Expressive Function of Punishment" (1965; see also Moberly 1968). The argument of neither, of course, has been universally accepted. Yet both operate by identifying something missing in the then-prevailing liberal-utilitarian consensus on criminal justice, with its emphasis on the efficient management of social behavior, and both suggest that the "barbaric" retributivism opposed by utilitarianism has more to be said in its favor than previously thought.

Devlin argued for retaining the criminalization of homosexuality in the UK on the basis that it was quite proper for the criminal law to target immorality or sin that arouses repugnance in "the reasonable man." Now there are many points in the Hart-Devlin debate in which the reasonable reader may find Devlin to be in the wrong. For instance, many people will find his views on homosexuality repellent, and his view that criminal law should be explicitly recognized as having a basis in Christianity is hard to reconcile with the multicultural nature of many modern democracies. Furthermore, it is often unclear whether he is arguing that criminal law should be based on what is justifiably regarded as immorality, or rather what the "common man" finds repugnant. We may share Hart's view, in *Law, Liberty and Morality*, that the criminal law ought to be shaped by the moral ideas that best survive critical reflection, rather than our gut feelings. Nevertheless, when Devlin argues that "no society can do without intolerance, indignation and disgust; they are the forces behind the moral law" – and makes it clear that he is talking about disgust "that is deeply felt and not manufactured," the presence of which "is a good indication that the bounds of toleration have been reached" – he touches upon an important aspect of the function of criminal law that is ignored if we see punishment merely as a technique for behavior control. This is a point picked up by Feinberg in his discussion of the expressive aspect of punishment. Punishment is not merely, as the utilitarian might regard it, an incentive against antisocial behavior; rather, it is a vehicle of symbolic collective condemnation, expressing the anger, indignation, or repugnance that an act arouses in the "right-thinking person." What is missing in purely forward-looking accounts of the criminal sanction is something that connects punishment with our sense of the seriousness of the wrongdoing.

Strawson's argument, on the other hand, is addressed most directly to the debate over free will and moral responsibility. As I explained above, any argument about whether we have free will depends on some conception of the kind of free will necessary for moral responsibility. Strawson gives us an account of what we are doing when we hold people morally responsible, from which he then adduces an account of the freedom we need to be morally responsible. He argues that our sense of moral responsibility and retributive justice is found in the operation of moral emotions such as resentment and indignation, which lead us to withdraw goodwill from, and inflict harm on, those to whom they are directed. Strawson defends retribution and responsibility by arguing that the tendency to have such emotions is inextricably intertwined with accountability in interpersonal relationships. The interpersonal relationships that Strawson has in mind are those grounded in mutual expectations of some level of goodwill: relationships that are based on trust rather than mere containment and accommodation. Such relationships are an important part of the value of human life, and do not seem to be a merely illusory ideal. But when we enter into such relationships, we adopt a certain practical perspective on human behavior, according to which participants are reasonably subject to certain demands for good will. Of course, this perspective on human action is not the only one that it is theoretically possible to adopt. We can also see human action, not as a response, adequate or inadequate, to legitimate expectations, but rather as the deterministic behavioral output of a number of personal, subpersonal, and environmental inputs. This is what we do when we note that human beings are biological mechanisms, that their action is the result of social forces, etc. It is also the perspective we adopt when dealing with the very young, or the insane. But human life would be barren – perhaps even impossible – if this perspective (which Strawson calls the objective attitude) were the only one we ever took up. Therefore, we should reject the abandonment of moral responsibility; there is no reason to think that our practices of accountability are in bad order.

Strawson's arguments, and the outcome of the Devlin-Hart-Feinberg debate, altered the way in which it was possible to understand the significance of retribution. We can see this if we think about how the viewpoint they developed might be used to help us address some of the criticisms of retribution we considered above. One of the main criticisms of retribution was that its backward-looking nature makes retributive action redundant, and its desire for the suffering of the offender makes it cruel. Another problem concerns free will, while another argues that apportioning virtue to happiness, even if desirable, is a pointless task for limited human beings. Yet another criticism is that retributivism is tied to outdated ideas of moral pollution. The Devlin-Strawson approach, in response, sees punishment or blame in deontic fashion as a necessary part of recognizing the significance of responsible wrongdoing in the context of relationships that are based on trust. Being subject to these retributive reactions for wrongdoing is simply a condition of being included in such relationships.

Strawson and Devlin enable the retributivist to argue that the value of retributive responses lies in the fact that they (a) recognize the seriousness of the wrongdoing, and (b) recognize the agent who did the wrong as a human being who could and should have done better, and can thus justly be held to the relevant moral demands. When a person is punished, in other words, it involves being treated harshly, but it also involves being included rather than excluded from the terms of the relationship.

This line of thinking might also allow the retributivist to turn the tables on the deterrence theorist, charging that the deterrence theorist advocates merely threatening the offender, treating her as a mere means to a social end, whereas retribution is directed at the offender as an individual, a responsible moral agent. This criticism is made most forcefully by R. A. Duff, who quotes approvingly Hegel's damning remark that, according to the deterrence theory, punishment is like raising a stick to a dog (Duff 1986, p. 180). Even if this seems too strong, a more modest point can be made. Strawson reminds us that mere deterrence is not what we are engaged in when we experience emotional correlates of punishment like resentment and indignation. This raises the question whether punishment itself could not be made more like an interpersonal interaction. Deterrence theory sees punishment as an interaction between an individual and a bureaucratic state agency motivated by an overview of social welfare. A potential strength of the retributive approach is that it envisages the criminal justice process as – ideally – something more like an interpersonal interaction.

Two Cheers for Devlin and Strawson?

In order to understand more recent developments, we turn now to evaluate how far these arguments take us. They may have convinced us that offenders should be held to account for what they have done, and that their actions deserve criticism or moral disapproval. Furthermore, we might be persuaded that condemnation of at least a certain range of wrongful actions is a legitimate state purpose. Perhaps the state has no business condemning people for adultery, or for sexually perverted behavior. But we might agree that it should condemn those who, say, violate the legal rights of others, or who rape, murder, and assault. This view of punishment provides an important alternative to the view that the function of criminal justice is simply as an effective means to protect life, liberty, and property, and has been explored by writers such as Andrew von Hirsch (1985, 1993) and R. A. Duff (1986, 2001).

However, the argument so far seems to suffer from an important gap. Serious wrongdoers may deserve condemnation, perhaps even by the state. But why, we might ask, do offenders deserve to suffer in addition to being criticized? Feinberg himself advances this criticism (Feinberg 1965). He argues that punishment is merely a conventional symbol of moral disapproval, and asks whether we could not find less harmful symbols to do that work for us. This question requires an answer, and, in response, censure theorists find themselves in two camps: those who argue that there can be no meaningful condemnation without hard treatment, and those who agree with Feinberg that the goal of censure does not by itself require hard treatment. In the latter camp lies the influential version of the censure conception of punishment put forward by von Hirsch. Von Hirsch argues that censure is an important function of punishment because it addresses the crime as a wrongful action, does justice to the victim, and conveys to the offender that some response is required to the offense (von Hirsch 1993, pp. 9–10). However, he argues that the hard treatment element of punishment is only necessary for preventive purposes: If punishment had no deterrent effect, there would be no need for hard treatment. A similar view is put forward by Matt Matravers (2000). A prominent writer in the opposed camp is Igor Primoratz (1989; see also Moberly 1968), for whom punishment is analogous to language. If we change the symbols, according to Primoratz, we change what is said or done by punishment, so the symbolism of punishment is not conventional in the sense that we can arbitrarily choose what symbols to use without change in meaning. If we aim to condemn, then we must use a vehicle that is symbolically adequate.

This raises a question about the nature of symbolic or expressive action in general and as such takes us into deep waters beyond the philosophy of criminal law and punishment. To what extent can we alter the symbolic action we use to convey, e.g., gratitude or grief or outrage while still preserving a meaningful connection to these attitudes? The retributivist – who wishes to argue for some infliction of hard treatment as the privileged symbolism for expressing outrage – requires some view on which symbols are more or less resistant to alteration. Thus, in response to the objection that we might respond to wrongdoing by turning the other cheek, or simply communicating disapproval in a calm and measured way, the retributivist will want to argue that these responses, insofar as they avoid inflicting proportionate hard treatment, fail to do justice to the seriousness of the wrongdoing involved. One way to defend such a view, which can be seen in recent work by Jeffrie Murphy and Jean Hampton (1988), is to follow Strawson and Devlin in referring to the emotions. On the basis of phenomenological investigation, we can say that the experience of being in the grip of an emotion provides a subject with a deep sense of fit between some feature of her situation and the behavior that expresses that emotion, such that that feature appears to "call for" that behavior. For instance, if I have cause for pride I might find myself "walking tall," and in that situation walking tall feels right. Or if I perceive myself as having done something shameful, then covering my face or hiding away feels right. With respect to retribution, we might point to something to do with the fact that the emotions of outrage, indignation, and resentment are expressed through aggressive action taken against the offender, such that it appears to us that the situation of wrongdoing calls for such a response. This might explain our sense that the symbolism of punishment is at least in part invariable: In order to express condemnation symbolically, we must do something that connects with the behavior we find satisfying as an expression of the emotions aroused by wrongdoing.

Morality, Retribution, and the Emotions

Given the history of hostility to the emotions in Western philosophical thought, however, this turn to the emotions to inform our thinking about punishment is likely to provoke strong criticism. Thus, a critic might ask whether the retributivist is right to give such weight to the emotions in normative theorizing – theorizing that may, after all, end up with some people (offenders) being seriously harmed. In particular, we might wonder whether, from the fact that something feels right when in the grip of an emotion, we can conclude anything about the nature of right or permissible action. This objection need not deny that there can be some value in the expression of emotion. Perhaps, in spontaneous interpersonal interactions, we cannot help expressing our emotions. And perhaps we should not always try to restrict such expression: maybe there is a value in spontaneity. But the spontaneous expression of retributive emotions leads to suffering, and the institutionalized expression of those emotions through state punishment can blight whole lives. The suspicion will be that our emotionally-driven sense of fitting response is too slender a thread on which to hang such weighty matters. Indeed, criminal justice policy, according to this challenge, may be better seen as containing rather than releasing such emotions (e.g., see Gardner 1998 on the "displacement function" of punishment).

In exploring how the retributivist might answer this criticism, we will need to say something about recent developments in the philosophy of emotion and moral epistemology. But briefly, the retributivist might attempt to argue three points. First of all, that emotions are not brute psychological states, but are what Scanlon calls "judgment-sensitive attitudes" (Scanlon 1999), attitudes in principle responsive to justificatory reasons. Secondly, the reason that emotions are judgment-sensitive attitudes is that they present a subject with an evaluatively loaded (cognitive) representation of the situation under consideration, the plausibility of which can be assessed by the subject for its coherence with her other factual and evaluative beliefs. Thirdly, because the emotions can be understood as intelligent, cognitive evaluations of our environment, and not merely automatic responses to it, we have a default epistemic entitlement to trust the way they represent our situation, in the absence, that is, of strong reason to doubt that representation. If those three points could be established, then the retributivist and the critic might agree on the following possibility: The way a situation is presented to us in the grip of an emotion could be our best way of understanding the evaluative nature of that situation and the response it demands of us. The argument with respect to retributivism would then come down to whether we have strong reason to mistrust those experiences in which it appears to us that taking wrongdoing seriously requires some type of hard treatment of the offender.

The retributivist's line of argument might start by following Nussbaum and Kahan in distinguishing a mechanical from an evaluative conception of the emotions (Nussbaum and Kahan 1996). The mechanical view sees emotions as strong sensations, immune from thought and rationality, that are triggered by certain stimuli and drive us to certain actions. One of the key developments in the philosophy of emotion in the twentieth century has been the observation that, at least in some respects, and for what might be called the "higher emotions" (Griffiths 1997), this view is too simple. The emotion of fear is not a simple sensation like the feeling of stepping into a warm bath. Fear is a state with intentional content, it is directed toward something (i.e., it is "fear of" or "fear about" or "fear that") in a way that simple sensation is not. Therefore, fear could be said to be a state that represents the world in a certain way. It is a way of seeing or understanding (according to Roberts (1988), "construing") a situation rather than just simple feeling triggered by the situation. (This is one of the main arguments against the James-Lange hypothesis that emotions are simply the effects in consciousness of bodily changes brought about by external stimuli.) Nussbaum and Kahan use this as evidence that emotion is more like a cognitive or evaluative state than the mechanical view allows (though for criticism, see, e.g., Deigh (2008)).

We could also point to the fact that it is often possible to judge emotions according to rational standards of adequacy to the facts: If I am afraid of a noise that turns out to be only the wind rattling the door-knob, then normally speaking my fear will abate; this suggests that fear is at least sometimes sensitive to reasons for that fear, that it can be justified or unjustified. Hence, the claim that emotions are "judgment-sensitive."

Any proponent of the cognitive or evaluative view of the emotions needs to recognize, however, that there is a partial truth in the mechanical view: Some emotions in particular are capable of getting out of proportion quite easily, and they can grip our attention to the extent that we find it difficult to get other things into proportion. But while the existence of irrational and disproportionate emotion is clear, this should not blind us to the fact that it is often possible to bring our emotions into line with our considered evaluative beliefs. In other words, some emotions can be justified in the sense that they embody our considered values. Furthermore, if this is allowed, it may open up the possibility that sometimes the direction of revision goes the other way round, and that it is our evaluative beliefs that have to change to accommodate the epistemic deliverances of an emotional experience that appears authentic and compelling. This is a possibility famously defended by Bernard Williams (Smart and Williams 1973; Williams 1976): Although Williams does not put it quite like this, we might say that it is the idea that justified and proportionate emotion might involve seeing the world in the right light (cf. McDowell 1985). If this is at least a possibility, then we might argue that we have a default entitlement to trust our emotions (or moral intuitions) unless we have reason to think that they cannot be made coherent with our other justified evaluative or factual beliefs (though the proper role of intuition in moral theory is the subject of great controversy: see Sinnott-Armstrong et al. (2010)). Applied to retribution, this would mean that we have a default entitlement to trust our feeling that the retributive response is fitting unless we have reason to think that this response cannot be made coherent with the demands of those other values in the validity of which we have great confidence (Moore 1988).

Nevertheless, the most we have established thus far is the possibility of justifying retribution on the basis that we might have an undefeated entitlement to trust the way our emotions present the situation of wrongdoing. In order to determine the success of this strategy for justifying retribution, we would need to investigate whether the purported validity of retributive experience in informing us of the demands of the situation has to be regarded as illusion because it conflicts with the other things we know about what the situation demands. And the reader might be forgiven for being underwhelmed by this result. For it might appear that it simply returns us to the realm of normative argument between the retributivist and her opponents about whether retributive responses are compatible or incompatible with what we know about our obligations, e.g., not to impose suffering. However, if the "retributive emotions" strategy is successful, then this reaction would be unjustified for three reasons. First of all, the success of the strategy would make it clear that the burden of justification lies with the opponent of retributivism. Retributive reactions could not be written off as mere "gut reactions." Secondly, it would provide a way of understanding ideas such as moral pollution or contamination that does not involve a commitment to unwanted metaphysical baggage. "Moral contamination" would simply be that quality that, when we are in the grip of (let us suppose) justified retributive emotion, seems to us to call for that reaction. It would not be a property of the universe with the power to cause famine, pestilence, etc. Thirdly, it would affect our view of how to conduct moral inquiry. In order to evaluate retributive reactions, what we would need to look at is, in part, whether we can understand our experience of those reactions as inauthentic and illusory, or whether it is rather our conflicting beliefs that need to change. This introduces an essentially phenomenological Gestalt-like element into moral inquiry.

The likelihood that we will conclude that retributive reactions are authentic or illusory no doubt depends on which emotions we focus on.

The relevant emotions are often taken to be those felt by third parties toward offenders, such as in Jean Hampton and Jeffrie Murphy's discussions (1988). But many readers may find that it is easier to feel comfortable with the claim that offenders should feel bad toward themselves than it is with the claim that "we" should endorse aggressive feelings toward "them." For this reason, another important line of work for retributivists is to be found in the conceptions of guilt and remorse put forward by Herbert Morris (1981), Gabriele Taylor (1985), R. A. Duff (1986), and Raimond Gaita (1991).

Emotion and Punishment

By way of conclusion, we can now distinguish three ways in which punishment might be related to the expression of the emotions.

- Firstly, state punishment might be carried out in order to displace aggressive retributive action motivated by strong feeling on the part of individual citizens.
- Secondly, punishment might be carried out in order to maintain the perception of the legitimacy of the state among the populace, a perception that might be strengthened if state action is carried out in line with the strong feelings of the citizens.

What we need to notice about these first two suggestions is that their motivation comes, not from the endorsement of the emotions involved, but rather from the desire for social stability: The policy is driven by the need to contain or accommodate strong emotion that might disrupt social relations. The justifiability or unjustifiability of those emotions is not relevant to the social need to contain them. However, the result of this is that a theory that gives this kind of role to the emotions is not distinctively retributive. For the need for punishment derives in the end from the need for social stability, not the retributive concern with doing justice.

For a distinctively retributive view, we need to consider a third possibility, which comes about if, after due reflection, we can endorse the way the situation of wrongdoing appears to us when we are in the grip of retributive emotions. If state punishment is a vehicle for treating the offender as those emotions prompt us to treat him, then we might say the following:

- Thirdly, state punishment is carried out because the way of seeing the significance of wrongdoing that is involved in the retributive emotions – namely, that it calls for hard treatment in response – is our best understanding of what the situation demands.

The retributive emotions present wrongdoing as mattering in a certain way, specifically as being impermissible or transgressive, and they present the situation as one that calls for a certain kind of response directed at the wrongdoer. The strand of retributivism in which we are interested here argues that we have reason to allow this phenomenology to inform our normative judgments, broadly endorsing our tendency to act on retributive emotions, and deploying state punishment as a vehicle for the condemnation of the offender. The reason for expressing the emotion is not a desire to release emotional pressure – as though the emotion were just a psychological force – but rather because retributive responses are our best understanding of what the situation demands.

Related Entries

▶ Civil Remedies
▶ Deterrence
▶ History of Restorative Justice
▶ Incapacitation
▶ Moral Crimes
▶ New Penology
▶ Penal Abolitionism
▶ Penal Justice and Social Injustice
▶ Penal Philosophy and Sentencing Theory
▶ Proportionality
▶ Punishment as Rehabilitation
▶ Rehabilitation
▶ Restitution as a Penal Aim
▶ Sentencing Guidelines in the United States
▶ Theories of Punishment
▶ Traditional Retributivism

Recommended Reading and References

Deigh J (2008) Emotions, values and the law. Oxford University Press, Oxford

Devlin P (1959) The enforcement of morals. Proc Brit Acad 45:129–151

Duff RA (1986) Trials and punishments. Cambridge University Press, Cambridge

Duff RA (2001) Punishment, communication and community. Oxford University Press, Oxford

Etzioni A, Carney D (eds) (1997) Repentance: a comparative perspective. Rowman and Littlefield, Lanham

Feinberg J (1965) The expressive function of punishment. Monist 49:397–423

Gaita R (1991) Good and evil: an absolute conception. Macmillan, London

Gardner J (1998) Crime: in proportion and in perspective. In: Ashworth A, Wasik M (eds) Fundamentals of sentencing theory: essays in honour of Andrew von Hirsch. Oxford University Press, Oxford

Griffiths P (1997) What the emotions really are: the problem of psychological categories. University of Chicago Press, Chicago

Hart HLA (1963) Law, liberty and morality. Stanford University Press, Stanford

Kant I (1991) Metaphysics of Morals (trans: Gregor M). Cambridge University Press, Cambridge

Kant I (1996) Critique of Practical Reason (trans: Abbott TK). Prometheus Books, Amherst

Matravers M (2000) Justice and punishment: the rationale of coercion. Oxford University Press, Oxford

McDowell J (1985) Values and secondary qualities. In: Honderich T (ed) Morality and objectivity: a tribute to J. L. Mackie. Routledge, London

Moberly W (1968) The ethics of punishment. Faber, London

Moore M (1988) The moral worth of retribution. In: Schoeman F (ed) Character, responsibility and the emotions: new essays in moral psychology. Cambridge University Press, Cambridge

Morris H (1968) Persons and punishment. Monist 52:475–501

Morris H (1981) A paternalistic theory of punishment. Am Philos Q 18:263–271

Murphy JG (1971) Three mistakes about retributivism. Analysis 31:166–169

Murphy JG (1973) Marxism and retribution. Philos Public Aff 2:217–243

Murphy JG, Hampton J (1988) Forgiveness and mercy. Cambridge University Press, Cambridge

Nussbaum M (1993) Equity and mercy. Philos Public Aff 22:83–125

Nussbaum M, Kahan D (1996) Two conceptions of emotion in criminal law. Columbia Law Rev 96:269–374

Primoratz I (1989) Punishment as language. Philosophy 64:187–205

Roberts R (1988) What an emotion is: a sketch. Philos Rev 97:183–209

Scanlon TM (1999) What we owe to each other. Belknap Press, Cambridge

Sinnott-Armstrong W, Young L, Cushman F (2010) Moral intuitions. In: Doris J (ed) The moral psychology handbook. Oxford University Press, Oxford

Smart JJC, Williams B (1973) Utilitarianism: for and against. Cambridge University Press, Cambridge

Strawson PF (1962) Freedom and Resentment. Proc Brit Acad 48:1–25

Taylor G (1985) Pride, shame and guilt. Clarendon Press, Oxford

von Hirsch A (1985) Past or future crimes: deservedness and dangerousness in the sentencing of criminals. Rutgers University Press, New Brunswick

von Hirsch A (1993) Censure and sanctions. Cambridge University Press, Cambridge

Williams B (1976) Moral luck. Proc Aristotelian Soc 50-(supplementary volume):115–135

Return to Prison

► Risk Factors for Prison Recidivism

Revictimization

► Repeat Victimization

Rhe

► Juvenile Diversion

Risk and Protective Factors

► Integrated Cognitive Antisocial Potential Theory

Risk Assessment

► Preventing of Terror at Shopping Malls

Risk Assessment, Classification, and Prediction

Edward J. Latessa and Brian Lovins
School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA

Overview

Over the years, the assessment of offenders has evolved from using one's "gut feeling" to decide who is likely to reoffend, to instruments that have focused on past behavior (static indicators), and to what are now called fourth generation assessments. These instruments combine static and dynamic (changeable) factors to more accurately predict risk of recidivism. They identify the crime-producing needs that should be targeted for change, and produce results that can be fully integrated into case plans used to manage what programs and services offenders receive as well as gauge progress (Andrews Bonta and Wormith 2006).

One of the foundations of developing effective correctional practices and programs is adoption of a validated risk assessment. Once focused primarily on the seriousness of the offense, the criminal justice system is now speaking in terms of risk and need factors. Hence, the risk level of the offender becomes the primary consideration for treatment and supervision over the offender's felony level. As departments adopt a risk, need, and responsivity (RNR) framework, the importance of assessing for risk has been elevated. In fact, many of the leading professional correctional organizations fully endorse the use of risk assessment in assisting in classification decisions and delivery of evidence-based practices (American Association of Community Justice Professionals, American Probation and Parole Association, and the International Community Corrections Association). To understand the impact risk assessment has had on corrections, this entry reviews the history of risk assessment, provides a framework for understanding the context in which assessment should be adopted, and discusses the current controversies surrounding the development of 3rd and 4th generation risk assessments. Finally, this entry provides recommendations for the field in regard to adopting risk assessment and using the results to best inform correctional practices.

What Do Risk Assessments Measure?

This seems to be a straightforward question. However, people have different concepts of "risk" and different instruments measure "risk" for different types of behavior. For some, "risk" is a concept associated with the seriousness of the crime. For example, a felon must be higher risk than a misdemeanant. Even though a felon has been convicted of a more serious offense than a misdemeanant, his or her relative risk of reoffending may have nothing to do with the seriousness of the crime. For most, "risk" refers to the probability of reoffending. A low-risk offender is one with a relatively low probability of committing a new offense (i.e., relatively prosocial people with few risk factors), while a high-risk offender has a much greater probability of future engagement in crime (i.e., more antisocial with several risk factors). While many in the criminal justice system are interested in the likelihood of reoffending, there are also assessments used in corrections that assist in determining if an offender is suicidal or likely to commit a violent act, be a victim of violence, engage in future sex offenses, engage in misconducts while incarcerated, etc. The purpose of the assessment dictates the outcome it measures. For example, a pretrial tool is likely to predict failure to appear in court and short-term behavior, while an institutional classification system will provide staff with a prediction of institutional misconducts. The focus in this entry is on the use of risk assessments to predict future criminality among offenders.

History of Risk Assessment

From the early beginning of community corrections, there has been an attempt to predict who will reoffend. Even prior to the use of actuarial tools, corrections professionals have tried to determine the likelihood that offenders would engage in future criminal behavior (Monahan 1977). Without the assistance of empirically based assessments, correctional workers have been forced to use a wide array of clinical assessments to draw on their professional experiences to make prognoses regarding offenders' likelihood of reoffending. The problem is that professional judgment alone is not very accurate. Dawes et al. (1989) found that clinical judgment was not accurate compared to statistical prediction for a range of outcomes. Ultimately, they concluded that humans take into account factors that are not predictive, weigh characteristics that are less predictive, and even when given additional training, do not outperform statistical models. Without a structured assessment process using validated and objective measures, corrections will be limited in what it can do to reduce recidivism (Latessa 2004). Adopting a risk assessment can help the criminal justice system identify the offenders who are at highest risk for recidivating, assist in enhancing public safety by ensuring that those higher-risk offenders receive appropriate supervision and interventions, help reduce bias in making decisions, and improve placement and utilization or resources.

Clinical Judgment, Structured Professional Judgment, and Actuarial Assessments

Three primary ways to assess offenders are clinical judgment, structured professional judgment, and actuarial (also called statistical) methods. As discussed in the previous section, clinical judgment is the process in which a staff person collects information and makes a decision about risk based on "gut feelings" or intuition. Clinical judgment relies on the skills, knowledge, and ability of individual staff to gather appropriate information and then to make a decision based on the totality of the information. In contrast, structured professional judgment combines clinical judgment and theoretically relevant measures into a single assessment – putting limits on the evaluators' biases while still tapping into their professional knowledge (Borum and Douglas 2003). Tools based on structured professional judgment are designed to help staff determine what they can do to disrupt future recidivism, not necessarily predict who will reoffend. The tools based on structured professional judgment are not scored per se; there are no cutoffs produced or a final composite risk score; instead specific areas in which the evaluator will rate the offender based on low-, moderate-, and high-risk levels are provided. These tools consider static and dynamic factors, protective factors, and contextual factors to assist correctional professionals to guide decisions regarding risk management (Vincent et al. 2009). One example is the Structured Assessment of Violence Risk in Youth (SAVRY), which is based on structured professional judgment along with standardized risk factors. With this tool an evaluator assesses a youth on 24 items and determines if the youth presents as low, moderate, or high risk for each indicator. Upon identifying the level of risk for each item, the assessor will make a final determination of the youth's overall level of risk. The SAVRY has demonstrated intraclass correlation coefficients of 0.81 and produced correlations with recidivism that range from 0.25 to 0.70.

A third method in conducting risk assessment is the actuarial method. Actuarial risk assessment is similar to what insurance companies use to calculate insurance rates. Actuarial instruments are based on statistical analysis of records and other information that result in the development of probability tables: if you score X you have an X chance of reoffending. The foundation of statistical risk assessments in corrections is similar to that in most actuarial sciences. For example, life insurance is cheaper for a nonsmoker in his 40s than a smoker of the same age. The reason insurance costs more for the smoker is that smokers have a risk factor that is significantly correlated with health problems – which runs a greater risk that the insurance company will have to pay out a claim before

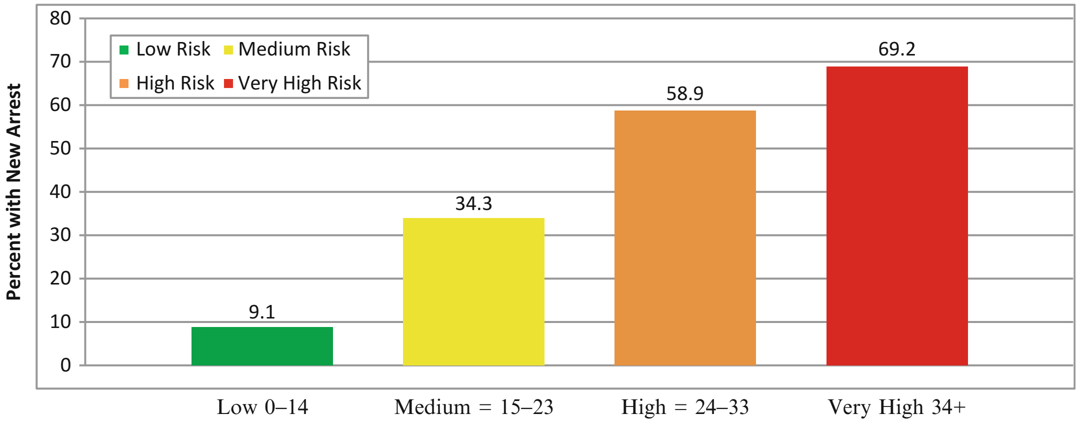

Risk Assessment, Classification, and Prediction, Fig. 1 Risk level by recidivism for community supervision sample

they recoup their money. Similarly, an offender who uses drugs and is unemployed has a higher chance of reoffending than someone who does not use drugs and has steady employment. Figure 1 shows how a distribution of offenders might be classified according to risk. Note that the probability of someone in the low-risk category reoffending is less than 10 %, moderate risk is 34 %, high risk is nearly 60 %, and very high risk is about 70 %.

Types of Assessment Tools

In addition to the methods by which assessments are designed, there are three unique types of risk assessments. Screening instruments are designed to assess a large population (usually at intake) to predict a specific outcome by sorting offenders into risk categories quickly. Screening tools are usually comprised of static items (e.g., criminal history) and can be completed by either a file review or a brief interview. They are typically used to separate low-risk from moderate- and high-risk offenders but have little utility beyond identifying levels of risk due to a limited number of dynamic items.

Beyond screening tools, comprehensive risk/need assessments are designed to gather more complete data to assist in determining levels of risk. They typically are longer to administer, require training, and can be used to identify targets for treatment. These instruments include a range of dynamic factors that can be used to reassess the offender. Furthermore, composite risk assessments can be used to develop case plans to ensure that programming targets criminogenic needs of the offender (Latessa and Lowenkamp 2005a).

Once a comprehensive risk assessment is completed, specialized assessments can be used to assess specific domains (like substance abuse) or special populations (i.e., sex offenders, mentally ill offenders, psychopaths). Specialized assessments can be completed on offenders who are flagged as risky on composite risk/needs tools. These instruments usually require additional training, while some tools require special licensures to conduct.

In many instances jurisdictions adopt all three types of assessments. A screening instrument might be used at pretrial or to assess offenders at intake to determine who needs further assessment. For those offenders who continue deeper into the system, a more comprehensive assessment tool would be used. Once assessed on a composite risk tool, specialized assessment should be prescribed on an as-needed basis. Following this approach can increase efficiency, since not all offenders will be thoroughly assessed, but those offenders who appear to pose the greatest risk to reoffend will be examined much more closely (Flores Russell Latessa and Travis 2005).

Risk, Need, and Responsivity

As agencies become more familiar with risk assessments and incorporate risk assessments across multiple decision points, it is important to understand how the results can be used to guide decisions in the field of corrections. The risk, need, and responsivity principles provide the context in which the results can be used to best serve the corrections population.

Risk Principle

The risk principle offers guidelines to correctional workers on which offenders are best served and under what conditions. Andrews Bonta and Hoge (1990) suggest that corrections should target those offenders with a higher probability of recidivism, provide the most intensive treatment to higher-risk offenders, and refrain from disrupting the lives of low-risk offenders. Although often ignored, correctional interventions that adhere to the risk principle demonstrate significantly greater reductions in recidivism (Latessa Brusman and Smith 2010). Consider a program that serves 100 low-risk offenders. Without intervention, approximately ten of the 100 offenders will reoffend. For the program to be effective, it would have to reduce recidivism for a portion of the ten offenders while not increasing recidivism for the other 90. In contrast, a program that serves 100 high-risk offenders would expect a recidivism rate of 60 %. For the program to be effective, it would have to reduce recidivism for a portion of the 60 offenders – resulting in a greater chance of reducing recidivism while reducing the chance of disrupting those that would not reoffend. The more troubling aspect to the risk principle is the fact that providing intensive interventions for low-risk offenders can actually increase failure rates (Lowenkamp and Latessa 2004). This effect is illustrated in Fig. 2, which shows the average changes in recidivism rates when we target high- and low-risk offenders.

While it makes sense that programs can have an impact on higher-risk offenders, it is not as evident why programs would increase recidivism for low-risk offenders. First, when low-risk offenders are placed in more intensive settings, they are more likely to be exposed to higher-risk offenders – increasing the likelihood of an iatrogenic effect. Considering the number of settings where lower-risk offenders are potentially mixed with higher risk – community service, treatment programs, jail, and prison – it is not surprising that lower-risk offenders are often more risky upon completion. Second, for low-risk offenders, residential programs, jails, and prisons actually disrupt what makes them low risk, causing the likelihood of recidivism to increase. For example, if you placed your low-risk child in a residential program for six months due to a substance abuse problem that child would miss out on education programming, he or she would experience family disruption, the youth's prosocial attitudes and friends would be replaced with antisocial thoughts and antisocial peers, and his or her friends' parents would likely not be interested in exposing their child to yours once released. In other words, risk would be increased, not reduced due to placement of the youth. Third, low-risk offenders are often victimized, taken advantage of, and pressured into engaging in antisocial behavior by more sophisticated offenders (Lowenkamp Latessa and Holsinger, 2006).

Grounded in the risk principle, the system should identify lower-risk offenders and provide alternative services to them without disrupting what makes them low risk. As a general rule, low-risk offenders should be excluded from intensive correctional programs. In order for the criminal justice system to adhere to the risk principle, offenders must be assessed and decisions must be made regarding the appropriate placement of lower risk offenders. If ignored, the results can be alarming. This was seen in a 2010 study conducted in Ohio that examined the effect of placing offenders in a halfway house versus regular supervision (Latessa Brusman and Smith 2010). Placing a low-risk offender in a halfway house led to an average 3 % increase in failure rates, while placing a high-risk offender in the same program resulted in an average 14 % reduction in recidivism.

Risk Assessment, Classification, and Prediction, Fig. 2 The risk principle & correctional intervention results from meta analysis (Source: Andrews and Dowden (2007))

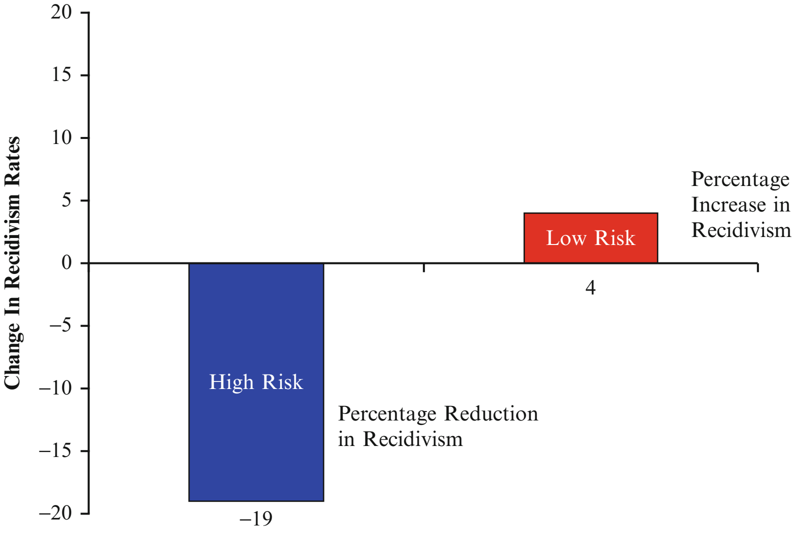

Need Principle

The need principle suggests that there are specific targets that should be addressed if the system is expecting to reduce recidivism. Research by Andrews, Bonta, Gendreau and others have identified seven dynamic areas associated with criminal conduct:

1. Antisocial/procriminal attitudes, values, and beliefs
2. Criminal peers and the lack of prosocial others
3. Antisocial temperament
4. Family
5. Education, employment, and finances
6. Limited prosocial leisure activities
7. Substance abuse

While the risk principle suggests that higher-risk offenders should be targeted for more intensive services, the need principle points to which needs should be addressed. Although most third and fourth generation risk assessments include both static and dynamic factors, the need principle is more concerned with those that can change (dynamic). Dynamic items include factors like who an offender chooses to hang around with, his or her attitudes and values towards crime, a lack of problem solving skills, substance use, and employment status. All these factors are derived from several decades of research and are correlated with recidivism. These dynamic factors are also called *criminogenic needs*: crime-producing factors that are strongly correlated with the risk of reoffending. Combining static and dynamic factors together gives the best picture of overall risk of recidivism and the most effective way to target criminogenic needs (Latessa and Lowenkamp 2005b). For example, people who are at risk for a heart attack usually have some common characteristics: over 50 years old, male, a family history of heart problems, high blood pressure, excessive weight, smoking, stress, and high cholesterol. While some of these are historical or static factors, others are dynamic. The benefit of assessing the dynamic factors is that if addressed, the high-risk patient can actually decrease his or her risk of having a heart attack. Similarly, dynamic items on a criminogenic risk assessment can provide malleable targets for risk to reoffend.

The need principle provides a roadmap to programs to ensure that appropriate criminogenic needs are targeted. Programs would benefit from targeting crime-producing needs, such as antisocial attitudes, antisocial peer associations, substance abuse, lack of problem solving and self-control skills, and other factors that are highly correlated with criminal conduct (Dowden and Andrews 1999). Identifying criminogenic needs is an important part of offender risk assessment; since it tells us “what” to focus on to reduce risk.

Responsivity Principle

The responsivity principle suggests that we should match the style and mode of service delivery based on key offender characteristics, such as culture, gender, mental health, or learning ability. This may not decrease risk, but will increase the likelihood of success in correctional interventions. One of the limitations of corrections is its ability to deliver services based on individual needs of the offender. The criminal justice system's reliance on group interventions has made it very difficult to ensure that interventions are individualized. For example, offenders who have cognitive delays may not respond well to group treatment focused on insight-oriented processes. There are often personal characteristics of an individual that should be assessed, since these factors can effect their engagement in treatment. These would include areas like mental and emotional problems, cognitive functioning, and level of motivation or readiness to change. Assessment of these areas can often improve the placement of offenders, assist with case planning, and impact the overall effectiveness of correctional treatment.

Controversies in Risk Assessment

Although the criminal justice field has moved beyond "gut feelings" as its primary form of assessment, there still remain several questions regarding the implementation of risk assessments. The first question relates to the applicability of risk assessments at the time of sentencing. From 1975 to 2002 incarceration rates increased by more than 300 % (Stemen et al. 2005). As the "get tough" movement took hold, many jurisdictions moved to determinate sentencing, stricter sentencing guidelines, and abolition of post-release supervision – resulting in massive numbers of offenders committed to state institutions. As a result, many have been calling for the implementation of risk reduction efforts at sentencing (Pew Center 2009). In fact, the Pew Center has identified the adoption of a risk assessment at sentencing as one of their top ten recommendations to control escalating prison populations. Of course, adopting a risk assessment at sentencing is not supported by all. Singer (1979) suggests that prison is not intended to address recidivism; the point of prison is to mete out a punishment commensurate with the harm to the victim. Others suggest that risk assessment should be used to develop appropriate targets for change post-sentence. In 2010, the Indiana Supreme Court upheld a decision by a trial judge to use the results of a risk assessment to inform the overall manner in which a sentence was to be served (Malenchik v State 2010).

A second question that has been raised regarding risk assessment is whether a single composite score comprised of risk and need factors should be used or should risk and needs be separated into two unique tools. Baird (2009) argues that risk assessment should be split into separate risk and needs assessments, suggesting that the risk index should be reserved for a fewer set of items that have moderate to high correlation. Limiting risk assessments to fewer, more highly correlated items will allow for greater parsimony, increase prediction, and ensure that staff can score the items accurately. The need assessment should be separately scored and based more on clinical judgment or instruments designed to determine why people commit crime – suggesting that composite risk tools cannot accurately measure individual level needs. In contrast, Bonta (2002) recommends that risk assessment should include items that are theoretically relevant, have multiple measures or items per domain, incorporate a combination of static and dynamic factors, and employ a multi-method approach to understanding risk. Risk assessments designed using Bonta's approach will typically hold more items, will measure risk using theoretically derived measures, and will focus more on the utility of the tool than parsimony.

The third question is whether third and fourth generation tools can meet acceptable levels of inter-rater reliability – that is, can different assessors come to the same decision based on the same information. Austin et al. (2003) suggests that measures of inter-rater reliability should reach a minimum of 0.70 for a risk assessment to be deemed reliable. The argument regarding third and fourth generation tools is that they typically include a significant number of dynamic items – which, in turn, are difficult to accurately rate. While Austin et al. (2003) and Baird (2009) point to third and fourth generation tools lacking reliability, there are many studies that have demonstrated that these tools can achieve acceptable levels of reliability with training. In fact, Austin found significant increases in reliability for the LSI-R when staff was provided with training on the instrument.

The fourth area in which there are questions regarding the use of fourth generation tools is in regard to using the results to develop individualized case plans. Can fourth generation risk assessments be used to identify treatment targets for individual offenders? The answer is mixed. Fourth generation tools typically identify levels of need by domain. For example, the Ohio Risk Assessment System-Community Supervision Tool (ORAS-CST) provides low, moderate, and high levels of need by each of the seven domains. These cutoffs assist staff in readily identifying areas that are riskier for offenders and help to form appropriate treatment targets. These domains are used to "flag" areas of need. Once an offender is identified as having moderate to high need in an area, the worker will want to explore each domain in greater detail to determine the individualized needs of the offender. For example, offenders that present a high risk of substance abuse may have different reasons for being high-risk substance abusers. For example, one offender may be physiologically or psychologically addicted to drugs. A second person may use drugs not because she or he is addicted, but because the offender spends significant amounts of time with peers that use drugs. A third person might use drugs to mask an underlying mental illness. All three offenders are high risk in regard to substance use, but present with uniquely different targets for treatment. Traditional drug treatment may work for the offender who is addicted, but that type of intervention may not be very effective for an offender who uses due to peer networks. In fact, sending an offender to a substance abuse group or AA/NA may not be appropriate given the treatment target is really antisocial peers. For the third group, if the underlying issue is mental health, helping an offender reduce substance use will have no impact on the symptoms of mental illness – again resulting in an intervention that has little impact on the presenting problem. To address the fact that any one need domain has a number of potential treatment targets, an offender who has a moderate to high level of need in a domain-specific area should receive additional need specific assessments to determine the individual level treatment targets.

The fifth question is in regard to the ability of risk assessments to measure change over time. Several studies have been conducted to examine the impact of a change in composite risk score on the LSI-R. Andrews et al. (2006) argues that offenders can change and that risk assessments should reflect the ability to measure these changes. Andrews and Robinson (1984) have found support for the ability of the LSI-R to predict change over time. In contrast, Baird argues that the use of third and fourth generation tools have significant limitations in measuring change over time. Moreover, he suggests that risk tools should be modified between intake and reassessment – shifting emphasis from prior history items to current behaviors.

These instruments use dynamic risk factors to measure initial risk and then reassess on these items to determine if the offender has made any significant changes in the risk he/she poses to society. In addition to targets of change, the fourth generation tools allow departments to focus their resources on those domains (broad area of need) that are moderate to high risk. Ultimately, if the criminal justice system can significantly reduce the recidivism rates for offenders, this will result in increased public safety. The fourth generation tools provide agents of change with a specific "road map" to address the needs of the offenders, manage limited resources, and protect public safety.

Obstacles to Good Practice

Here are some of the more common obstacles that exist with regard to offender assessment:

1. Offenders are assessed, but the process ignores important factors. Sometimes this is

because the tool selected is comprised mainly of static predictors, or the assessment process focuses on one or two domains (like substance abuse) to the exclusion of other important risk factors.

2. Offenders are assessed, but the process does not distinguish quantifiably determined levels (i.e., high, moderate, low). This is common with narrative assessments, and the result is often that the summaries all read the same: "offender is at risk to reoffend unless they get substance abuse treatment." This type of information tell us little about the actual risk the offender poses to reoffend or the level of need in specific areas. This is common with clinical assessment processes.
3. Even when offenders are comprehensively assessed, the results are not used – everyone gets the same treatment. As we will discuss below, adopting a risk assessment tool is only one step in the process. If the information is not going to be used then why assess?
4. Staff members often are not adequately trained in the use of the instruments, or they are only trained when the new instrument is selected. Usually, when a decision to use a new instrument is made, everyone is trained, but as time goes on and new employees are hired, little refresher training may be done, and the new staff simply learn how to use the tools by watching the older staff. The result is that reliability and validity suffer, and stakeholders lose confidence in the results of the assessment.
5. Staff resistance is one of the most persistent obstacles to overcome. Some of the common refrains that are heard by a staff include "I just need to talk to them for 5 minutes to determine their level of risk," "we don't have time to conduct an assessment," "they are all high risk," and "they all get the same treatment anyway so why assess them?" While staff resistance can be a challenge, it is not insurmountable, and as Jones Johnson Latessa and Travis (1999) found, most community correctional agencies understand the importance and value of using a valid and reliable risk assessment tool.

To avoid these and other mistakes and to derive the full value from assessments, there are some points to consider (Latessa and Lowenkamp 2005b; Latessa (2004)):

- There is no "one size fits all" assessment tool. Some domains or types of offenders will require specialized assessments, such as sex offenders or mentally disturbed individuals. In addition, the use or purpose will vary. For example, the assessment tool for making a decision about whether to grant pretrial release may be different from one for making a decision about whether to grant probation.
- Actuarial assessment is more accurate than clinical assessment, but remember, no process is perfect and there will always be false positives and false negatives – sometimes low-risk offenders reoffend, and sometimes high-risk offenders are successful.
- Assessment is usually not a "one-time" event, especially if the offender is under some form of community control. Offender risk and need factors change, so it is important to consider assessment as an ongoing process.
- Assessments help guide decisions, but they do not make them – professional discretion is part of good assessment-aided decision-making.
- While the new dynamic assessment tools can produce more useful information, they require more effort to insure reliability – they require staff training and continual monitoring of the assessment process. Fidelity and quality assurance make a difference.
- Good risk assessment serves a number of functions and helps guide decisions by providing reliable information in a systematic and objective manner. It can be the cornerstone of a more effective, efficient, and just system.
- A flexible process should be developed that expands as needed – higher-risk offenders need more assessment.
- Processes and instruments should be standardized so that everyone is speaking the same language with regard to risk assessment.
- Regardless of the assessment tool used, staff should be thoroughly trained on the rationale and use of a risk assessment tool. Proper training will ensure that the staff understand

the advantages of risk assessment and that they use the tool in an appropriate and consistent manner. The level and amount of staff "buy in" can drastically affect the level of success in implementing a risk assessment process or tool (Lowenkamp Holsinger Latessa 2004).

- Following training, agency administrators should establish quality assurance processes such as periodic audits of assessments, refresher training, or even certification of assessors.
- Assessment results should be used to develop case supervision and treatment plans and to assign offenders to programs.
- Information should be shared with service providers so that they understand the risk of the offender they are involved with as well as the criminogenic factors that need to be targeted.
- An assessment tool should be validated on the population for which it is being used. There are several widely used actuarial instruments that have been validated in numerous settings and across several subgroups (i.e., males, females, different racial, and ethnic groups). Nonetheless, agencies should still analyze assessment results based on the population for which the tool is being used.

Related Entries

▶ Actualizing Risk-Need-Responsivity
▶ Age-Crime Curve
▶ Aging Correctional Populations
▶ Behavioral Management in Probation
▶ Cambridge-Somerville Youth Experiment
▶ Career Criminals and Criminological Theory
▶ Cognitive/Information Processing Theories of Aggression and Crime
▶ Co-offending
▶ Co-offending and Offender Attributes
▶ Criminal Careers
▶ Desistance and Supervision
▶ Deterrence: Actual Versus Perceived Risk of Punishment
▶ Econometrics of Crime
▶ Effective Supervision Principles for Probation and Parole
▶ Evolutionary Theories of Criminal Behaviors
▶ Gangs and Social Networks
▶ Group-Based Trajectory Models and Developmental Change
▶ Group-Based Trajectory Models
▶ Growth Curve Models with Categorical Outcomes
▶ History of Corrections
▶ History of Probation and Parole in the United States
▶ Identification Issues in Life Course Criminology
▶ Intensive Probation and Parole
▶ Offender Change in Treatment
▶ Offender Decision Making and Behavioral Economics
▶ Onset of Offending
▶ Parole and Prisoner Reentry in the United States
▶ Pathways to Delinquency
▶ Probation and Parole Practices
▶ Probation Officer Decision-Making
▶ Psychopathy and Offending
▶ Race Issues in Probation and Parole
▶ Rational Choice Theory
▶ Risk Factors for Adolescent Sexual Offending
▶ Risk Factors for Gang Membership
▶ Risk Factors for Prison Recidivism
▶ Sex Offenders and Criminal Policy
▶ Treatment of Sex Offenders

Recommended Reading and References

Andrews DA, Bonta J (1998) The psychology of criminal conduct. Anderson, Cincinnati

Andrews DA, Dowden C (2007) The risk-need-responsivity model of assessment and human service in prevention and corrections: crime-prevention jurisprudence. Can J Criminol Crim Justice 49:439–464

Andrews DA, Robinson D (1984) The level of supervision inventory: second report. Report to Research Services. Ontario Ministry of Correctional Services, Toronto

Andrews DA, Bonta J, Hoge R (1990) Classification for effective rehabilitation: rediscovering psychology. Crim Justice Behav 17:19–52

Andrews DA, Bonta J, Wormith SJ (2006) The recent past and near future of risk and/or need assessment. Crime Delinq 52:7–27

Austin J, Coleman D, Peyton J, Johnson KD (2003) - Reliability and validity study of the LSI-R risk assessment instrument. Institute on Crime, Justice, and Corrections at the George Washington University, Washington, DC

Baird C (2009) A question of evidence: a critique of risk assessment models used in the justice system. National Council on Crime and Delinquency, Madison

Bloom HS, Singer NM (1979) Determining the cost-effectiveness of correctional programs: the case of patuxent institution. Eval Rev 3:609–628

Bonta J (2002) Offender risk assessment: guidelines for selection and use. Crim Justice Behav 29:355–379

Bonta J, Wormith SJ (2008) Risk and need assessment. In: McIvor G, Raynor P (eds) Developments in social work with offenders. Jessica Kingsley, London, pp 131–152

Borum R, Douglas K (2003) New directions in violence risk assessment. Pyschiatr Times 20(3):102–103

Dawes R, Faust D, Meehl P (1989) Clinical versus actuarial judgment. Science 31:1668–1674

Dowden C, Andrews DA (1999) What works in young offender treatment: a meta-analysis. Forum Correct Res 11:21–24

Flores AW, Russell AJ, Latessa EJ, Travis LF (2005) - Evidence of professionalism or quackery: measuring practitioner awareness of risk/need factors and effective treatment strategies. Fed Probat 69(2):9–14

Jones DA, Johnson S, Latessa EJ, Travis LF (1999) Case classification in community corrections: preliminary findings from a National Survey, Topics in community corrections. National Institute of Corrections, U.S. Department of Justice, Washington, DC

Latessa EJ (2004) Best practices of classification and assessment. J Commun Correct X11(2; Winter 2003–2004):27–30

Latessa EJ, Lowenkamp CT (2005a) What are criminogenic needs and why are they important. For the record (fourth quarter): 15–16

Latessa EJ, Lowenkamp CT (2005b) The role of offender risk assessment tools and how to select them. For the record (fourth quarter): 18–20

Latessa E, Brusman L, Smith P (2010) Follow-up evaluation of Ohio's community based correctional facility and halfway house programs. Center for Criminal Justice Research, Cincinnati

Lowenkamp CT, Latessa EJ (2004) Understanding the risk principle: how and why correctional interventions can harm low-risk offenders, Topics in community corrections. National Institute of Corrections, U.S. Department of Justice, Washington, DC, pp 3–8

Lowenkamp CT, Latessa EJ (2005) Increasing the effectiveness of correctional programming through the risk principle: identifying offenders for residential placement. Criminol Public Policy 2(4):15–24

Lowenkamp CT, Holsinger A, Latessa EJ (2004) Assessing the inter-rater agreement of the level of service inventory revised. Fed Probat 68(3):34–38

Lowenkamp CT, Latessa EJ, Holsinger A (2006) The risk principle in action: what we have learned from 13,676 offenders and 97 correctional programs. Crime Delinq 52(1):1–17

Malenchik v. State of Indiana, No. 79S02–0908–CR–365 (Ind., 2010)

Monahan J (1977) Strategies for an empirical analysis of the prediction of violence in emergency civil commitment. Law Hum Behav 1:363–371

Pew Center on the States (2009) One in 31: the long reach of American corrections. The Pew Charitable Trusts, Washington, DC

Singer R (1979) Just deserts sentencing based on equity and desert. Ballinger, Cambridge, MA

Stemen D, Rengifo A, Wilson J (2005) Of fragmentation and ferment: the impact of state sentencing policies on incarceration rates, 1975–2002. Vera Institute of Justice, New York

Vincent G, Terry A, Maney S (2009) Risk/needs tools for antisocial behavior and violence among youthful populations. In: Andrade J (ed) Handbook of violence risk assessment and treatment: new approaches for forensic mental health practitioners. Springer, New York, pp 377–423

Risk Factors for Adolescent Sexual Offending

Michael C. Seto[1] and Lesleigh Pullman[2]
[1]Integrated Forensic Program, Royal Ottawa Health Care Group, Ottawa, ON, Canada
[2]University of Ottawa, Ottawa, ON, Canada

Synonyms

Juvenile sex offenders

Overview

According to the FBI's Uniform Crime Report, 16 % of arrests for rape and 19 % of arrests for other sexual offenses are committed by youth under the age of 18 (United States Department of Justice 2001). Moreover, many adult sex offenders admit committing their first sexual offense before the age of 18. Indeed, Knight and Prentky (1993) found that half of the 1,025 adult sex offenders in their sample began sexually offending as adolescents. Because many sexual offenses are not reported, the prevalence of adolescent sexual offending is higher than official statistics indicate. Given these facts, it is not surprising that adolescent sexual offenders (ASOs) have become a major societal concern.

An etiological understanding of adolescent sexual offending is needed to refine assessment and intervention strategies. One area of research that would be particularly helpful is identifying if adolescent sexual offending can be explained by a generalist model of criminal behavior or if ASOs should be considered specialist offenders, with special factors that explain the onset and/or maintenance of their crimes. If ASOs are generalists, then existing risk and need assessment measures and interventions would likely be useful with this population. However, if ASOs are specialists, with distinct etiological factors, then different risk/need measures and interventions may be required.

ASOs have commonly been considered a distinct subtype of adolescent offender, with different developmental trajectories and criminogenic needs. Indeed, the National Adolescent Perpetrator Network, an organization of professionals involved with ASOs, concluded in an influential task force report that "…sexually abusive youth require a specialized response from the justice system which is different from other delinquent populations" (1993, p. 86). However, some empirical literature does suggest that the crimes committed by ASOs are a manifestation of general delinquency risk factors, including antisocial personality traits such as impulsivity and callousness, antisocial attitudes and beliefs, delinquent peers, early conduct problems, and substance abuse. Indeed, on average about half of ASOs would qualify for a diagnosis of conduct disorder, which by definition requires involvement in a range of antisocial behaviors (e.g., Elliott 1994).

The purpose of this entry is to summarize the theoretical literature concerning the models of adolescent sexual offending and the evidence for and against the generalist and specialist perspectives. The results of a recent meta-analysis conducted by Seto and Lalumière (2010), examining 59 studies comparing adolescent sex and nonsex offenders, will be discussed in detail to highlight where there are similarities and differences between ASOs and other adolescent offenders and to provide insight into the generalist versus specialist debate.

Key Issues/Controversies

The General Delinquency Explanation. The general delinquency explanation suggests that ASOs share etiological factors with other adolescent offenders. In a series of studies conducted by Lewis et al. (1981), the authors compared 17 ASOs and 61 other violent adolescent offenders on psychiatric, neurological, and psychoeducational variables. The authors concluded that the similarities between violent adolescent offenders and ASOs indicate these groups have similar therapeutic needs. While ASOs do warrant some specialized treatment options, general antisocial and violent tendencies must be addressed in treatment as well. Awad et al. (1984) found that the similarities between ASOs and other adolescent offenders outweighed the differences. These similarities included psychiatric history, inadequate parenting, past delinquency, violence, school misconduct, and parent-child attachment.

Recidivism studies are also consistent with a generalist model of adolescent sexual offending. Relatively few adolescent offenders commit only sexual crimes; most commit nonsexual offenses as well, though the extent of involvement in nonsexual offending varies greatly. In a meta-analysis conducted by Caldwell (2002), ASOs were six times more likely to be rearrested for nonsexual offenses than sexual offenses, and only a small proportion (10–15 %) of ASOs go on to commit sexual offenses as adults. Next, the majority of the factors that have been shown to predict recidivism among delinquents in general (criminal history, early behavior problems, antisocial personality traits, attitudes, etc.) also predict sexual recidivism in ASOs (Hanson and Morton-Bourgon 2005). Lastly, sexual offenses committed by adolescents tend to occur in an escalating pattern, starting with minor nonsexual offenses such as theft to more serious nonsexual crimes such as assaults, and eventually leading to serious sexual crimes (Elliott 1994).

Based on these findings, it is plausible to suggest that ASOs would display similar scores to that of other adolescent offenders on general delinquency risk factors. Findings of this nature would suggest that ASOs are not a distinct subtype of adolescent offender. If this were the case, current assessment tools, treatment options, and theories that are applied to adolescent offenders could be used in the study and treatment of ASOs without diminishing the reliability and validity of such options.

Specialist Explanations. Specialist explanations of ASOs suggest that these offenders have different etiological factors than other adolescent offenders. If this perspective is correct, then assessment and treatment methods developed for adolescent offenders in general will be less effective with ASOs, as they only address any common factors the two populations share. Theories describing the etiology of ASOs have focused on factors such as poor childhood attachment, heterosocial incompetence, atypical sexual experiences, psychopathology, cognitive abilities, atypical sexual interests, and sexual abuse history. Each of these factors will be briefly described below.

Poor Childhood Attachment. The ability to form mutually rewarding relationships is thought to stem from the parent-child relationship. Studies conducted with adult sex offenders have found that they are more likely than other offenders to have insecure childhood attachments that develop into insecure adult attachment styles (for a review see Marshall and Barbaree 1990). Awad et al. (1984) found that the vast majority of ASOs in their sample had been separated from their parents for an extended period of time and just over half of the youth showed serious problems in their family system, including parental abuse and neglect, witnessing violence within the home, and parental rejection. Marshall and Barbaree (1990) suggested that insecure childhood attachments increase the likelihood of sexual offending in adolescents by hindering the development of healthy emotional regulation, empathy, and social skills. Indeed, ASOs seem to be characterized by a pattern of poor social relationships. Fagan and Wexler (1988) found that ASOs were significantly more socially isolated than other violent adolescent offenders. If poor childhood attachment is a risk factor specific to ASOs, it is expected that ASOs would differ from other adolescent offenders by having more childhood attachment problems, especially insecure attachment.

Psychopathology. Studies have often found high levels of anxiety, depression, and personality problems in both adolescent and adult sex offenders. Among 22 ASOs, Galli et al. (1999) found that 82 % had a diagnosable mood disorder such as depression, 55 % had an anxiety disorder, 50 % had a substance abuse disorder, 55 % had an impulse control disorder, and 71 % had attention-deficit/hyperactivity disorder. Indeed, many participants in this small clinical study had substantial psychiatric comorbidity. The authors also reported that some participants in their study stated that their sexual impulses and sexually aggressive behavior increased when they were experiencing mood symptoms.

There are theories that postulate that sexual offending is associated with affect regulation and personality problems. For example, Hall and Hirschman's (1991) quadripartite model suggests there are four subtypes of sexual aggressors, each differing in the primacy of physiological, cognitive, affective, and personality problems. The third subtype emphasizes episodic affective dyscontrol, suggesting that ASOs are more likely to engage in sexually aggressive behavior when they are angry or sad. The fourth subtype is characterized by developmental personality problems such as neuroticism. If increased psychopathology is an etiological factor specific to ASOs, these offenders should score higher on measures of psychological problems compared to other adolescent offenders.

Cognitive Abilities. Theories that focus on cognitive deficits in ASOs suggest that offenders who have lower intellectual functioning have poorer impulse control and thus are more likely to sexually offend opportunistically against children. At the same time, youth who are of lower intellectual or cognitive functioning may be more likely to be sexually rejected by similar-aged peers and thus feel the need to turn to children or engage in sexually coercive behavior to satisfy their emotional and sexual needs. In a meta-analysis conducted by Cantor et al. (2005), the authors found that adult sex offenders scored significantly lower on measures of intelligence compared to adult nonsex offenders. However, though ASOs tended to have a lower mean intelligence score than other adolescent nonsex offenders, these results were not statistically significant. A limitation of this meta-analysis is that the authors included all available samples of sex offenders with all available samples of nonsex offenders. The samples were not matched on possible confounding variables such as how they were selected for study. Therefore, these groups may have differed in a way that influenced their intelligence scores and explains why this study did not confirm previous findings of lower intelligence in ASOs. In a direct comparison of groups recruited in the same settings using the same procedures, one would expect that ASOs would score significantly lower on measures of intellectual and cognitive functioning than other adolescent offenders.

Heterosocial Incompetence. Heterosocial incompetence refers to the lack of social skills necessary to function in a heterosexual relationship, given that most ASOs are male and sexually prefer females. Explanations that focus on this factor suggest that ASOs seek relations with younger children or use sexual coercion against peers or adults because they do not have the necessary social skills to maintain an age-appropriate relationship that satisfies their emotional and sexual needs (e.g., Marshall and Barbaree 1990). Adolescent child molesters have been found to score significantly lower than other adolescent offenders in regard to heterosocial competence (Katz 1990). Lastly, in a meta-analysis conducted by Hanson and Morton-Bourgon (2005), intimacy deficits (composed of poor social skills, negative social influences, conflicts in intimate relationships, emotional identification with children, and/or loneliness) were a significant predictor of sexual recidivism in adult and ASOs. If heterosocial incompetence is a correct explanation, one would expect that ASOs would score lower on measures of heterosocial competence compared to other adolescent offenders.

Psychosexual Development. Researchers have speculated that ASOs differ from other adolescent offenders in their psychosexual development (Marshall and Barbaree 1990). One common finding is that ASOs experience earlier and more frequent exposure to sex as children. Some studies suggest adult sex offenders are characterized by more frequent use of pornography in childhood compared to nonsex offenders, but in a meta-analysis of the adult literature, Allen et al. (2000) compared criminal sex offenders with the general population and found that the two groups did not differ in their frequency of pornography use as well as age at first exposure to pornography. However, the authors did find that criminal sex offenders were more likely than the general population to perform consensual and/or nonconsensual sexual acts following pornography viewing. If this explanation were correct, it is expected that ASOs would score higher than other adolescent offenders on measures of atypical sexual experiences such as sexually inappropriate family environments and early exposure to pornography or sex.

Atypical Sexual Interests. Atypical sexual interests usually refer to relatively strong sexual arousal to depictions of children or sexual coercion, compared to consenting sex with same-aged peers. Research has confirmed that adult sex offenders show levels of increased sexual arousal to these atypical stimuli compared to either nonsex offenders or non-offenders (see Seto 2008). Additionally, measures of these atypical sexual interests have been shown to be significant predictors of sexual recidivism in adult and ASOs (Hanson and Morton-Bourgon 2005).

Seto et al. (2000) found that as a whole, their group of 40 adolescent child molesters displayed more sexual arousal to depictions of children and coercive sex compared to young adult nonsex offenders. The difference was driven by those who had offended against boys; those who had offended against girls did not differ from the comparison group. Seto et al. (2003) found that phallometrically assessed sexual arousal to children was correlated with sexual victim history in an expected fashion, such that adolescents who had offended against boys, multiple child victims, younger child victims, or unrelated victims were more likely to show a preference for children than those who offended against only girls, a single child victim, older child victims, or related victims. Given all of these results, it is expected that ASOs will show significantly more arousal to children or coercive sex when compared to nonsex adolescent offenders.

Sexual Abuse. The most common etiological explanation of sexual offending in adults or adolescents is sexual abuse history. These theorists suggest that children and adolescents who are sexually abused are more likely to commit sexual offenses later on in life (Marshall and Barbaree 1990). In a meta-analysis conducted by Jespersen et al. (2009) comparing 1,037 adult sex offenders and 1,762 adult nonsex offenders, there was a significantly higher prevalence of sexual abuse history in adult sex offenders compared to other nonsex offenders. In fact, adult sex offenders had 3.36 times greater odds of having a sexual abuse history than their nonsex offender counterparts.

For ASOs, Johnson (1988) found that the earlier the sexually abusive behavior in ASOs begins, the more likely the youth is to have been sexually abused themselves. In the authors' sample, 72 % of the youth who began committing sexually abusive behavior before the age of 6 had been sexually abused themselves, whereas 42 % of the youth who began committing sexual abusive behavior between the ages of 7 and 11 had this history. In another study, Worling (1995) found that of the 1,268 ASOs included in his sample, 30 % had a history of sexual abuse. These explanations would suggest that ASOs should have a higher prevalence of sexual abuse history compared to other adolescent offenders. Because the explanation posits a specific link between sexual abuse history and sexual offending, any difference for sexual abuse should be greater than for physical abuse or other forms of maltreatment.

Meta-analysis. A meta-analysis is a statistical procedure that integrates the results of independent studies to produce a more reliable estimate of the actual size of the effect for the variable in question. A systematic review first identifies all relevant studies in an effort to minimize publication and other related biases. After the systematic review, the meta-analysis can be conducted to calculate a weighted statistic to give a more reliable calculation of effect size. Additionally, meta-analysis allows for the examination of moderators. Moderators are variables that affect the strength of the relationship between a predictor variable and an independent variable. Lastly, meta-analysis techniques allow for the development of more homogenous groups, which subsequently allows for the ability to compare subsets of samples found in different studies.

Seto and Lalumière (2010) examined 59 direct comparison studies, with a total sample of 3,855 male ASOs and 13,393 male adolescent nonsex offenders, on theoretically derived variables. These included offender age, criminal involvement, conduct problems, antisocial tendencies, substance abuse, childhood abuse and exposure to violence, interpersonal problems, family problems, cognitive abilities, sexuality, psychopathology, and impression management. Figures 1 and 2 summarize the results of this meta-analysis, which provided an empirical test of the generalist and specialist explanations previously discussed.

The factors suggestive of a generalist model (Fig. 1) include variables that have been identified as risk factors for general delinquency, including antisocial personality, behavior and attitudes, as well as family characteristics including family criminality, violence, and substance use. The factors suggestive of a specialist model (Fig. 2) include variables theoretically identified as special to adolescent sexual offending, including sexual abuse history, atypical sexual interests, heterosocial incompetence (poor social

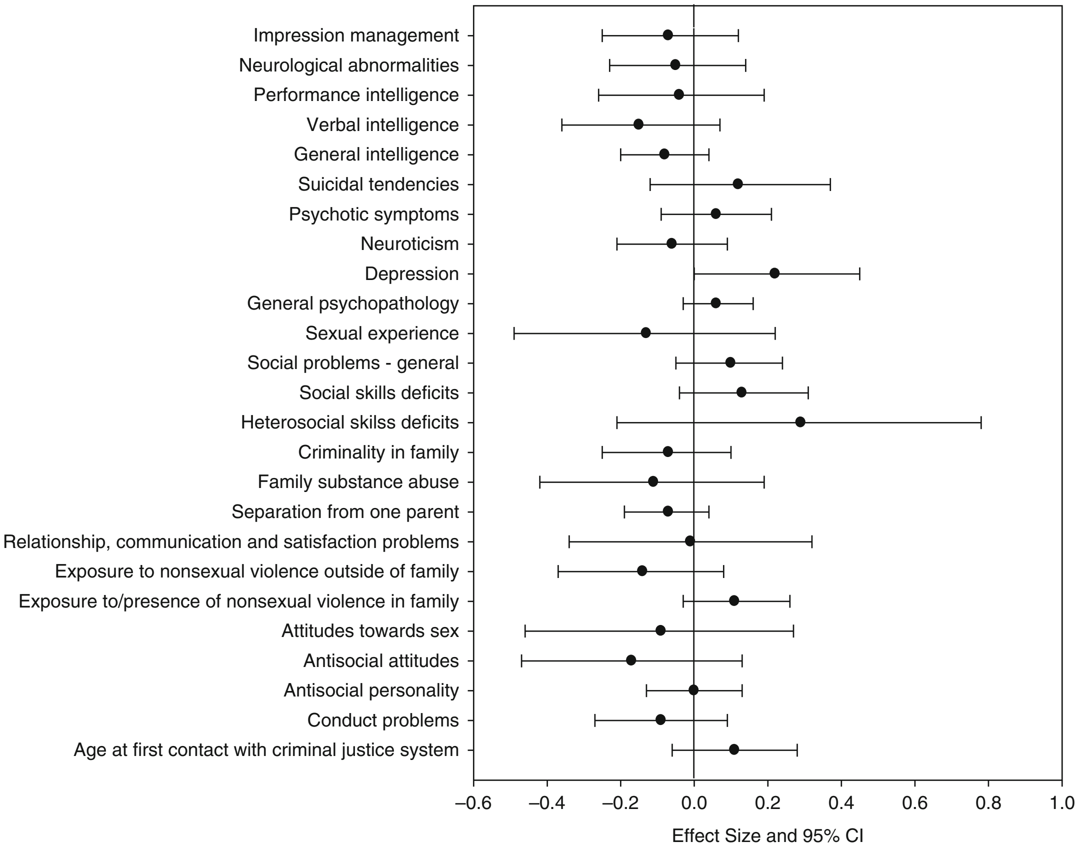

Risk Factors for Adolescent Sexual Offending, Fig. 1 Factors suggestive of a generalist model

relations), atypical sexual experiences (exposure to sex or pornography), psychopathology (anxiety and low self esteem), as well as cognitive deficits (learning problems/disabilities). The largest specialist-anticipated differences between ASOs and other adolescent offenders were found for atypical sexual interests and then sexual abuse history.

There are two main conclusions that can be drawn from the Seto and Lalumière (2010) meta-analysis. The first would be that, on average, ASOs share many of the same risk factors as other adolescent offenders, including antisocial attitudes and personality traits, which helps explain why ASOs are involved in criminal behavior. But in order to explain why a particular adolescent commits a sexual crime rather than a nonsexual crime, factors associated with psychosexual development, including sexual abuse history, exposure to sex or pornography, and atypical sexual interests, also need to be considered. The second conclusion that can be drawn from the meta-analytic results is that neither the generalist explanation nor the specialist explanations can fully explain adolescent sexual offending.

Offender Typologies. The results of Seto and Lalumière's (2010) meta-analysis need to be considered in conjunction with other research. ASOs are heterogeneous, and it is expected that there are different types of ASOs. The generalist and specialist explanations are not mutually exclusive. Both could be true: Some ASOs could be generalists, with similar risk factors and treatment needs as other adolescent offenders, whereas other ASOs could be specialists, with special factors that require other assessment or treatment methods.

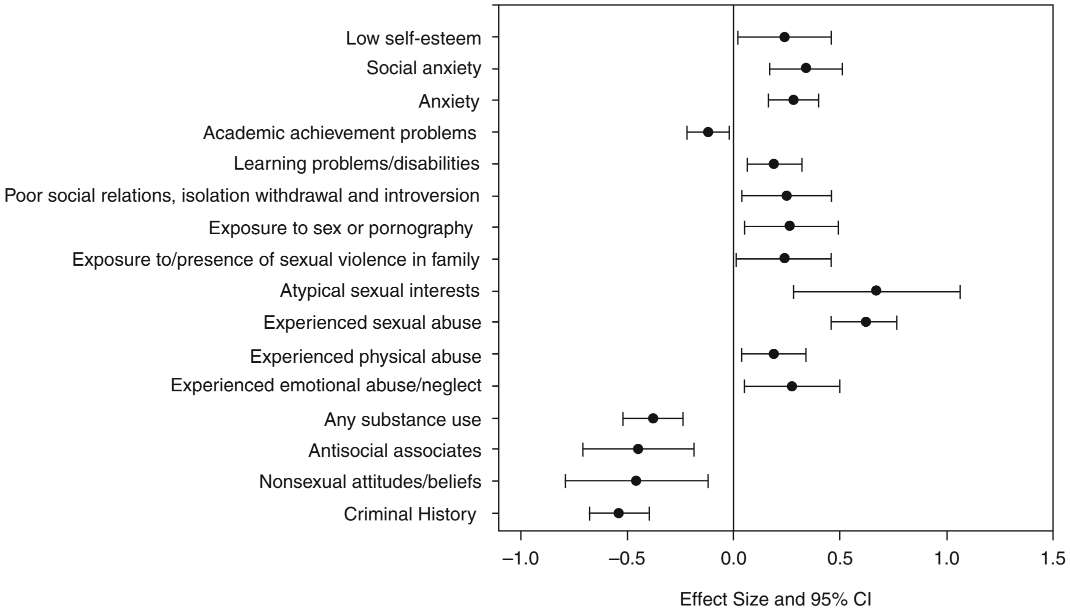

Risk Factors for Adolescent Sexual Offending, Fig. 2 Factors suggestive of a specialist model

Additional support for the specialist/generalist distinction comes from studies examining the performance of actuarial risk assessment measures. The Juvenile Sex Offender Assessment Protocol II (JSOAP-II) is a 28-item scale of theoretically derived static and dynamic risk factors for ASOs (Prentky and Righthand 2003). The items are organized into four subscales: Sexual Drive/Preoccupation, Impulsive/Antisocial Behavior, Intervention, and Community Stability/Adjustment. Rajlic and Gretton (2010) confirmed the predictive validity of the JSOAP-II using a sample of 286 male ASOs. The authors found that as a whole, the scale significantly predicted sexual as well as nonsexual recidivism. Specifically, the Sexual Drive/Preoccupation subscale predicted sexual but not nonsexual recidivism, while the Impulsive/Antisocial Behavior subscale predicted nonsexual but not sexual recidivism. The Intervention subscale and the Community Stability/Adjustment subscales predicted both sexual and nonsexual recidivism. These findings are important because it suggests that ASOs who have more risk factors in the Impulsive/Antisocial Behavior subscale (who are more likely to be generalist offenders) are more likely to reoffend nonsexually, while offenders who have higher scores in the Sexual Drive/Preoccupation subscale (who are more likely to be specialist offenders) are more likely to reoffend sexually.

The generalist/specialist distinction has been examined in ASOs as well. Butler and Seto (2002) found that 22 sex-only ASOs (adolescents who had only committed sexual offenses) and 10 sex-plus ASOs (adolescents who had committed sexual as well as nonsexual crimes) were different in many ways. Sex-plus offenders had more conduct problems as children, more antisocial attitudes and beliefs, higher risk for future delinquency, as well as a more extensive criminal history, substance abuse, and peer and family problems. All of these factors are related to general delinquency, indicating that sex-plus offenders have much in common with other adolescent nonsex offenders who were included in the study as a comparison group. These sex-plus offenders would be considered generalist ASOs. Conversely, sex-only offenders were characterized by less criminal history and conduct problems, less substance abuse, fewer peer relation problems, and fewer psychological problems.

Risk Factors for Adolescent Sexual Offending, Table 1 Overlap between typological clusters in adolescent sex offenders

Model						
	Adolescent types					
Butler and Seto (2002)	Sex-only	Sex-plus				
Chu and Thomas (2010)	Specialist	Generalist				
Oxnam and Vess (2006)		Antisocial	Normal			Inadequate
Worling (2001)		Antisocial/ impulsive		Overcontrolled/ reserved	Confident/ aggressive	Unusual/isolated

Note: Rows represent individual models while columns represent conceptual overlap between proposed types. The labels used in this table come from the original authors

The sex-only offenders, as specialists, had less in common with other adolescent offenders on these variables. Using the same generalist/specialist distinction, Chu and Thomas (2010) created a classification model for ASOs based on risk of recidivism. While there was not a significant difference in the rate of sexual recidivism between generalists and specialists (14.3 % versus 9.9 %, respectively), generalist offenders were significantly more likely to recidivate violently and nonviolently.

Oxnam and Vess (2006) used hierarchical cluster analysis to distinguish three groups of ASOs. The first group was described as "antisocial" and was characterized by youth who were prone to aggression and unpredictable. These youth lacked close bonds with others and had a propensity for substance abuse. The second group was described as "inadequate" and was characterized by severe psychopathology. This group was more likely to have a history of abuse and more likely to be negative and self-degrading. The last group of ASOs was described as being in the "normal range" and did not have extreme scores on any of the personality/risk scales.

Lastly, Worling (2001) conducted a cluster analysis of ASOs based on personality characteristics. This analysis revealed four types. The first, "antisocial/impulsive," was characterized by a propensity for rule violations that included sexual and nonsexual offenses. The second, "unusual/isolated," was characterized as emotionally disturbed and socially isolated. The third, "overcontrolled/reserved," was shy, reserved, and relatively weak at forming intimate, age-appropriate relationships. The last, "confident/aggressive," was characterized by overt aggression and narcissistic personality traits. The authors found that the "antisocial/impulsive" and the "unusual/isolated" groups were both at a higher risk to reoffend, whether violently (including sexually) or nonviolently, compared to the other two groups.

Table 1 displays the extent of overlap between these different typological clusters proposed by the above studies. While typologies may be useful in the classification and treatment of offenders, they do not provide suggestions for the etiology of the different proposed types. To develop an integrated theory of adolescent sexual offending, the origins of these different types of ASOs must be examined. Of particular interest are the developmental trajectories underlying the different types. In the following section, focus is placed on the developmental trajectories of generalist and specialist offenders.

Developmental Trajectories. Seto and Barbaree (1997) introduced a model of sexual offending in which they describe at least two developmental trajectories. The first trajectory (antisocial) is characterized by an early onset of problem behaviors and general delinquency risk factors. These offenders may show an early onset of sexual activity, have affect regulation problems, and can be manipulative and callous. The origin of this trajectory lies in the interactions between neuropsychological predispositions and criminogenic environments, consistent with Moffitt's (1993) description of the developmental trajectories of life course-persistent offenders. The second trajectory is characterized by opportunistic sexual offending. Though individuals in this trajectory may engage in nonsexual antisocial behavior as well, it is not early onset, varied, or persistent, consistent with Moffitt's (1993) description of adolescence-limited offenders.

Sexual deviance factors may be present in either trajectory; their presence increases their risk of sexual recidivism (Hanson and Morton-Bourgon 2005). Crossing the presence or absence of sexual deviance with the developmental trajectories creates four hypothetical groups of offenders: *persistent generalists*, *limited generalists*, *persistent specialists*, and *limited specialists*. *Persistent generalists* are high in delinquency risk factors, show problems early in life, and are likely to continue to nonsexually or, to a lesser extent, sexually offend in the future (they are life course persistent). Sexual offending for them would be part of a general pattern of antisocial behavior, and thus they would be more likely to engage in opportunistic sexual offenses, usually against peers/adults but sometimes against older children, particularly those showing some signs of puberty. *Limited generalists* are not expected to have sexual deviance risk factors but do have some delinquency risk factors. These offenders are expected to offend as a result of lack of sexual opportunities with peers and access to potential victims. *Persistent specialists* are high in delinquency risk factors but are also high on atypical sexual interests and thus are likely to continue to both sexually and nonsexually reoffend. They are expected to be more likely to offend against children than persistent generalists and are more likely to plan their sexual offenses. This group would pose the greatest risk of sexual recidivism. Lastly, *limited specialists* are low in delinquency risk factors and likely low to moderate in sexual deviance risk factors. Their sexual and nonsexual offending is expected to be limited to adolescence. Their sexual offenses may be attributed to their reactions to perturbations in their psychosexual development as a result of sexual abuse, early exposure to sex or pornography, social isolation, and lack of sexual opportunities with peers.

Future Directions

The results of the Seto and Lalumière (2010) meta-analysis can only indicate that both generalist and specialist explanations (but not all specialist explanations) contribute to an understanding of adolescent sexual offending. There are many unanswered questions. Of primary interest is the proportion of ASOs who can be characterized as generalists or specialists and whether these broad distinctions encompass more specific types on the basis of antisocial behavior, sexual behavior, and offense/victim characteristics. This question could be addressed by using cluster analysis or similar techniques on a large sample of ASOs with data on relevant measures. Ideally, the data would include information about developmental and historical variables to explore the developmental trajectories leading to the different empirically validated types. The different trajectories for antisocial behavior in terms of disposition, environmental interactions, age of onset, and life course are relatively understood, but the same is not true for the trajectories of sexual behavior and sexual deviance. Also of interest would be follow-up studies to determine if different ASO types respond differently to treatment or other interventions and to determine if they differ in their risk to sexually and/or nonsexually reoffend. Studies of this nature would go far in explaining the etiology of adolescent sexual offending and would have important implications for assessment and intervention practices.

Related Entries

- ▶ Establishing Causes of Offending in Longitudinal and Experimental Studies
- ▶ History of Juvenile Justice
- ▶ Juvenile Violence
- ▶ Moffitt's Developmental Taxonomy of Antisocial Behavior
- ▶ Pathways to Delinquency
- ▶ Risk Assessment, Classification, and Prediction
- ▶ Sex Offender Treatment

- ► Sexual Recidivism
- ► Specialization and Sexual Offending
- ► Specialization in Juvenile Offending
- ► Treatment of Sex Offenders

Recommended Reading and References

Allen M, D'Alessio D, Emmers-Sommer TM (2000) Reactions of criminal sexual offenders to pornography: a meta-analytic summary. In: Roloff M (ed) Communication yearbook 22. Sage, Thousand Oaks, pp 139–169

Awad GA, Saunders E, Levene J (1984) A clinical study of male adolescent sex offenders. International Journal of Offender Treatment and Comparative Criminology 28:105–116

Butler SM, Seto MC (2002) Distinguishing two types of juvenile sex offenders. J Am Acad Child Adolesc Psychiatry 41:83–90

Caldwell MF (2002) What we do not know about juvenile sexual reoffense risk. Child Maltreatment 7:291–302

Cantor JM, Blanchard R, Robichaud LK, Christensen BK (2005) Quantitative reanalysis of aggregate data on IQ in sexual offenders. Psychol Bull 131:555–568

Chu CM, Thomas SDM (2010) Adolescent sexual offenders: the relationship between typology ad recidivism. Sex J: J Res Treat 22:218–233

Elliott DS (1994) Serious violent offenders: onset, developmental course and termination. Criminology 32:1–22

Fagan J, Wexler S (1988) Explanations of sexual assault among violent delinquents. J Adolesc Res 3:363–385

Galli V, McElroy SL, Soutullo CA, Kizer D, Raute N, Keck PE et al (1999) The psychiatric diagnoses of twenty-two adolescents who have sexually molested other children. Compr Psychiatry 40:85–88

Hall GCN, Hirschman R (1991) Toward a theory of sexual aggression: a quadripartite model. J Consult Clin Psychol 59:662–669

Hanson RK, Morton-Bourgon KE (2005) The characteristics of persistent sexual offenders: a meta-analysis of recidivism studies. J Consult Clin Psychol 73:1154–1163

Jespersen AF, Lalumière ML, Seto MC (2009) Sexual abuse history among adult sex offenders and non-sex offenders: a meta-analysis. Child Abuse Neglect 33:179–192

Johnson TL (1988) Child perpetrators: children who molest other children: preliminary findings. Child Abuse Neglect 12:219–229

Katz RC (1990) Psychosocial adjustment in adolescent child molesters. Child Abuse Neglect 14:567–575

Knight RA, Prentky RA (1993) Exploring characteristics for classifying juvenile sex offenders. In: Barbaree HE, Marshall WL, Hudson SM (eds) The juvenile sex offender. Guilford Press, New York, pp 45–83

Lewis DO, Shanok SS, Pincus JH (1981) Juvenile male sexual assaulters: psychiatric, neurological, psychoeducational, and abuse factors. In: Lewis DO (ed) Vulnerabilities to delinquency. Spectrum, New York, pp 89–105

Lipsey MW, Wilson DB (2001) Practical meta-analysis: applied social research methods series, vol 49. Sage, Thousand Oaks

Marshall WL, Barbaree HE (1990) An integrated theory of the etiology of sexual offending. In: Marshall WL, Laws DR, Barbaree HE (eds) Handbook of sexual assault: issues, theories, and treatment of the offender. Plenum Press, New York, pp 257–275

Moffitt TE (1993) Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. Psychol Rev 100:674–701

National Adolescent Perpetrator Network (1993) Revised report from the national task force on juvenile sexual offending of the national adolescent perpetrator network. Juv Fam Court J 44:1–115

Oxnam P, Vess J (2006) A personality-based typology of adolescent sexual offenders using the Millon adolescent clinical inventory. N Z J Psychol 35:36–44

Prentky R, Righthand S (2003) Juvenile sex offender assessment protocol–II (J-SOAP-II) manual (NCJ Publication No. 202316). Office of Juvenile Justice and Delinquency Prevention, Washington, DC. Retrieved from http://www.csom.org/pubs/JSOAP.pdf

Rajlic G, Gretton HM (2010) An examination of two sexual recidivism risk measures in adolescent offenders: the moderating effect of offender type. Crim Justice Behav 37:1066–1085

Seto MC (2008) Understanding pedophilia and sexual offending against children: theory, assessment, and intervention. American Psychological Association, Washington, DC

Seto MC, Barbaree HE (1997) Sexual aggression as antisocial behavior: a developmental model. In: Stoff D, Breiling J, Maser JD (eds) Handbook of antisocial behavior. Wiley, New York, pp 524–533

Seto MC, Lalumière ML (2010) What is so special about male adolescent sexual offending: a review and test of explanations through meta- analysis. Psychol Bull 136:526–575

Seto MC, Lalumière ML, Blanchard R (2000) The discriminative validity of a phallometric test for pedophilic interests among adolescent sex offenders against children. Psychol Assess 12:319–327

Seto MC, Murphy WD, Page J, Ennis L (2003) Detecting anomalous sexual interests in juvenile sex offenders. J Sex Coercive Behav 989:118–130

United States Department of Justice (2001) Uniform crime report

Worling JR (1995) Sexual abuse histories of adolescent male sex offenders: differences on the basis of the age and gender or their victims. J Abnorm Psychol 104:610–613

Worling JR (2001) Personality-based typology of adolescent male sex offenders: differences in recidivism rates, victim selection characteristics, and personal victimization histories. Sex Abuse: J Res Treat 13:149–166

Risk Factors for Gang Membership

Kristy N. Matsuda
Criminology and Criminal Justice, University of Missouri-St. Louis, St. Louis, MO, USA

Synonyms

Predictors of gang membership

Overview

The question "Why do people join gangs?" has been of particular interest to those hoping to prevent gang membership. There are a number of potential methods to gain insight into the people who join gangs and the conditions or factors that lead to membership. These methods include the descriptive approach (a descriptive account of the characteristics of gang members), the direct approach (asking gang members why they joined), and the risk factor approach (identifying predictors of gang joining). Each approach is accompanied by limitations and benefits. The risk factor approach offers a number of methodological benefits and has significantly moved the gang literature forward in the most recent decades. A summary of gang risk factor studies and their theoretical contributions is offered here. While great strides have been made in the gang risk factor area, much work remains. A discussion of needed areas of research and potential hurdles is included.

Introduction

Gangs have historically been a complicated phenomenon to understand. They form, they change, some disappear, some reappear, and some show amazing stability. People join, and some stay, but many leave quickly. In short, gangs and gang members come in so many shapes and sizes it has been difficult to define with confidence, let alone identify them all and study the extent and nature of the problem. But, there is a great desire to do so. One major reason for a preoccupation with gangs and their members is the strong link between gang membership and delinquency.

Gang members commit more delinquency than nongang members. This robust finding has been replicated over time, samples, and place (both across the United States and around the world). The rate of delinquency and violence has been so noteworthy that much public attention (media, research, policy, policing) on gangs has been in an attempt to address this issue. However, the history of concentrated gang suppression efforts has shown that gangs can be quite resilient. There is, in fact, evidence to suggest that attempting to force a gang to disband may actually increase their cohesion and group bond, thus actually making them stronger entities and less likely to dissipate. Therefore, *prevention* efforts (i.e., keeping gangs from forming or keeping people from joining) have become particularly important options for this type of social problem. For prevention to be feasible, however, the impetus for gang membership must be well understood. In other words, we must ask "Why do people join gangs?" If there is a greater understanding about the reasons for gang joining, there may be better venues to combat the problem before it starts.

Background Description

Gang studies have been challenged with definition and identification issues. There has not been a universal consensus (among practitioners, policy makers, or researchers) on exactly what constitutes a gang or a gang member. Therefore, efforts to identify, classify, or study gangs have always been fraught with questions of validity, comparability, and generalizability (for a review, see Klein and Maxson 2006). While this limitation has not hindered the ability to produce quality research studies, it must always be considered when interpreting gang findings.

The history of gang research has spanned both macro- and micro-level inquiry and has included both qualitative and quantitative methodologies. The earliest gang studies were generally qualitative studies that focused on a specific area or particular gang. These primarily descriptive studies were central in providing a rich understanding of gangs and their behaviors. A number of these studies reveal the conditions under which gangs form, and some implicated reasons why gang members joined their gangs. Quantitative studies soon followed. Many of these studies cast a wider sampling net by including both gang and nongang youth. These types of studies have been critical in assessing differences between gang and nongang members, the factors led some to join and others to resist, as well as a plethora of other subjects.

The quality of the quantitative, individual-level studies has improved substantially as gang research began to adopt both the use of cross-sectional methods and the more methodologically rigorous longitudinal data collection efforts. The expansion of knowledge in the area of risk factors for gang membership has benefited tremendously from data produced by prospective, quantitative, and longitudinal studies like the Denver Youth Survey (Esbensen et al. 1993), the Rochester Youth Development Study (Thornberry et al. 2003), the Pittsburgh Youth Study (Lahey et al. 1999), the Montreal Longitudinal Experimental Study (Craig et al. 2002), and the Seattle Social Development Project (Hill et al. 1999). These studies have been used to test and validate conclusions of cross-sectional studies and to provide firm, temporal ordering for predictors of gang membership. Longitudinal research efforts are integral to the study of risk factors (see Thornberry et al. 2004 for overview). They are, however, incredibly costly, time intensive, difficult to conduct, and therefore rare. As such, we are limited to only a handful of these types of studies. It is difficult to make sweeping generalizations regarding the importance of specific factors with such few studies. Cross-sectional studies help to bolster confidence in results, but they suffer from temporal ordering problems and cannot be given as much weight. Therefore, while much progress has been made in area of gang risk factor research, conclusions are still limited and many of our conclusions are still inconclusive.

Fundamentals

There are a few ways in which to answer the question “Why do people join gangs?” The first method would be to describe the types of people that are gang members and infer that these characteristics were important determinants in why they joined a gang. This approach is often used in media or law enforcement depiction of gangs. These methods examine a group, neighborhood, city, state, or country and identify the gang members and describe them. Using this approach, results might suggest that gang members are often minority, males from highly impoverished areas. The implication for the question “Why do people join gangs?” then becomes that something about a person’s status as a minority, male, or resident of an impoverished area leads to gang membership. Generally, the descriptors available are superficial (sex, race/ethnicity, socioeconomic status, location, age) because the data being collected are from an outside observer’s determination of gang membership. In other words, this method does not include surveying every resident of a community about their gang status and then presenting descriptive information. This approach is limited to the observer’s ability to determine *gang membership* and the quality of the data that can be collected (usually) from afar. This approach is also limited in that it does not offer a good context for comparison. Even if a majority of gang members are minority, male, and residents of an impoverished area (which is debatable), it is not true that a majority of all minority, male, residents of impoverished areas are gang members.

The second way to find out why people join gangs is to ask them directly. Both qualitative and quantitative researchers have used this approach to gain insight into gang joining. This is a very direct and valuable approach. It is also limited in certain ways. First, usually, the sample captured in the direct method is more limited than the observational method due to the work intensity. Second, this approach requires finding gang members that are willing to disclose sensitive information. Self-report will always raise questions about recall and honesty. Finally, gang members can either be asked to offer their reasons for gang joining with or without prompts (i.e., answer choices). If gang members are allowed to answer without prompting the gamut of responses can be infinite and difficult to classify and generalize. If they are probed with specific responses, then an argument can be made that their responses are not as genuine. Regardless, the direct approach provides unique insight into the cognition of gang members.

Finally, there is a risk factor approach to determining who joins gangs and why. Risk factors are conditions (either individual or environmental) that increase a person's likelihood of developing some negative condition. Risk factors precede the outcome of interest and, therefore, should be measured at an earlier period than the outcome. This requires the use of at least two observational points (i.e., a longitudinal study). In addition, the risk factor approach requires a comparison group (in this case, nongang members) to identify characteristics that distinguish gang members as a special group. In order to obtain this information from gang and comparison groups, a larger sampling net must be cast. This method often relies on respondent's assessment or admittance that they are gang involved, which can assist in identifying gang members that may not be as "obvious" in the observational approach first described (but also suffers from the same limitations as the direct method described above). Usually in the study, a broad range of questions and statuses can be asked. One benefit to this method is that, unlike directly asking a respondent why they joined a gang, the predictors of gang joining need not be known to the respondent.

The Descriptive Approach

The descriptive approach uses the characteristics of gang members to understand "who joins a gang." If a status is overrepresented in a gang sample than what would be expected in the general population, one could infer that status is related to gang membership. Official data (i.e., arrest statistics, incarceration rates) is one example of data that can offer a descriptive account of gang members. These data are collected when gang members come to the attention of the justice system. These data are often limited to whatever characteristics are collected by law enforcement (i.e., mostly superficial like sex, race, age). Using official data and the descriptive approach, gang members are often described as older, male, racially/ethnically minority, and from impoverished areas.

It is important to note that the descriptive picture of gang members differs depending on the use of self-report versus official data. Conducting a similar descriptive analysis of gang members using self-report data can yield a different depiction of what a "gang member" is. Self-report data collection efforts find that the peak age of gang joining is much younger than official data would offer. Self-report also shows that though males are more likely to be gang members, females make up a more significant proportion of gang members than official data would indicate. Self-report studies have identified gang members in both high-risk communities and low-risk communities. Though official data may suggest that gang members are usually racial/ethnic minorities, self-report studies identify substantial proportions of white gang members.

From Gang Members Themselves

A number of cross-sectional and longitudinal studies have asked gang members why they joined gangs. The most common response pro-offered by gang members varies by study. Many of these studies collect these data by offering choices that gang members can select as the reasons they joined; therefore, if an answer is *not* supported by a study, it may be because it was not offered as a choice or because it was never selected by a gang member.

Many studies report that *protection* is a common reason for gang joining as suggested by members. This answer is both commonly asked by many studies and is a commonly reported reason offered by gang members. Another common reason for gang membership is that family or friends are involved in the gang. Studies that looked specifically at differences between girls' and boys' responses to gang joining found that girls were more likely to say that they joined because friends and family were in the gang than boys (though it was a common response for boys as well). Gang members also cite a desire for respect or a reputation, to feel like they are a part of their neighborhood, and to fit in better or to feel like they belong. It is also quite clear that the sheer fun and excitement of gang membership is an attractive feature for gang members.

This method has been very useful in debunking certain myths associated with gang joining. Gangs are often depicted as being organized, disciplined groups that are brought together for profit making in illicit business – in particular, drug selling. Studies show that it is actually rare for youth to report that they joined the gang for instrumental purposes like drug selling. Gang members identified by Esbensen and colleagues (1999) from a national sample of American youth, however, did often state that the general notion of "for money" was an important impetus for joining. Another commonly hypothesized cause of gang membership is that youth are forced to join gangs. A national survey of youth found that youth (boys and girls) rarely reported that they joined a gang because they were forced.

Risk Factor Approach

The risk factor approach requires both temporal ordering and a comparison group in order to truly identify the characteristics that *lead* to gang membership. As such, risk factor studies are generally far more labor intensive and costly than a cross-sectional descriptive study or even an interview study with gang members. Thus, the number of "true" risk factor studies is quite low. More cross-sectional studies (no temporal ordering) have been conducted that identify differences between gang and nongang members. These studies are useful and informative but will always suffer the lack of temporal ordering. However, longitudinal studies have generally supported the results of cross-sectional studies which lend greater confidence to the results. For the purposes of this review, the results of both longitudinal and cross-sectional studies will be included, though any result based solely on cross-sectional studies will be viewed with more caution.

Drawing from the more general literature on juvenile delinquency, gang risk factors are often presented in five domains: individual, family, school, peers, and community. Studies have shown that gang risk factors span all five domains. Therefore, no domain should be ignored. Previous work on risk factors for juvenile delinquency, more generally, shows that the importance of any of the five domains varies by social development. In other words, during some periods of the life course, certain domains may be more important than at others. It has been strongly argued that the causes of gang membership are unlikely to be set at the earliest stages of the life course. There is no evidence of an innate predisposition for gang membership. This makes intuitive sense because gang joining necessitates a gang to join – which is strongly influenced by life and environmental circumstance. This suggests that infant gang risk factor studies may not be a top priority, but starting early is still of serious consideration. Longitudinal research shows that important risk factors are present at relatively early ages. For example, using the Seattle Development Study, (Hill et al. 1999) measured risk factors at ages 10–12 that predicted gang joining at ages 13–18.

It has been repeatedly shown that risk factors for gang joining have a cumulative effect. In other words, the more risk factors present, the greater the likelihood of gang involvement. In the Seattle Study, youth with seven or more risk factors were 13 times more likely to join a gang than those with no risk factors or only one. The cumulative effect of risk factors has been shown both cross-sectionally and longitudinally. And this pattern has also been frequently demonstrated for juvenile delinquency more generally. It also appears that individuals with risk factors across multiple domains are at an increased likelihood for gang membership.

In reviewing the risk factor literature, it is important to keep measurement issues in mind. For ease of generalization and an overview, the intended spirit of measures needs to be grouped together, but the specific wording of each measure varies from study to study. And, as with any interpretation of quantitative data analysis, the determination of what constitutes "significant" is important. Bivariate relationships are the easiest types of analyses to detect any relationship between an independent and dependent variable. These types of relationships may not "hold up" in multivariate models. Does that make them irrelevant? In some cases, these same factors may be included in other multivariate models specified with other measures and be significant. Does that make them relevant? Clearly, model specification matters to the ability to detect significance. For better or for worse, this review mostly relies on the results of the most sophisticated statistical model presented by the researchers.

Individual Domain

Of all risk factor domains, the testing of the "individual factors" domain is the most expansive. Individual factors include demographic features, individual attitudes and beliefs, as well as behavioral characteristics. The ability of demographic features to predict the likelihood of gang joining has certainly depended on the sample under consideration and the type of analysis being conducted. As stated prior, males are more often thought to be gang members especially by media or law enforcement estimates, but self-report shows that a significant proportion of gang members are female. Similarly, the overrepresentation of minorities as gang members is less dramatic in self-report studies. Still, sex and race/ethnicity as significant risk factors of gang membership vary tremendously by study and the samples under investigation. Numerous school-based, self-report studies fail to find differences by sex and racial/ethnic propensity for gang membership, and other youth studies find differences. A handful of studies have examined the role of ethnic identity on an individual's propensity to join a gang. These studies consider the role of ethnic identity separately for different racial/ethnic groups under the theoretical assumption that gang membership may be related to the level of integration that youth feel within their racial or ethnic group. Studies testing ethnic identity have failed to find a significant relationship between identification within their ethnic group and gang joining.

It has been hypothesized that one factor that leads to an increased likelihood to participate in gangs are internalizing or externalizing disorders or affective states. Factors that have been examined in this category include disruptive behavior, conduct disorder, impulsivity or hyperactivity, risk seeking, low self-esteem/self-derogation, or the lack of guilt. Self-esteem is one of the most commonly tested factors and not commonly found to be associated with gang joining in either cross-sectional or longitudinal studies. Non-delinquency problem behaviors like conduct disorder, disruptive behaviors, or oppositional defiant disorder are more often available for consideration in longitudinal studies and have been consistently supported as a significant risk factor for gang membership. Impulsivity, risk seeking, and guilt have been less frequently tested (and mostly by cross-sectional studies) and have, thus far, garnered preliminary support.

Scholars have also examined whether pro-crime or pro-gang attitudes increase the likelihood of gang joining. There is some cross-sectional support that pro-gang attitudes are associated with increased gang joining, but there has not been support for the notion that youth who perceive more benefits to gang joining are more likely to join (or that youth who perceive many cons to gangs will not join). There is cross-sectional and, more limited, longitudinal support for the notion that youth with higher tolerance toward deviance, antisocial beliefs, more extensive use of techniques of neutralization for crime, or more supportive attitudes for alcohol and drugs are more likely to join gangs. Related to this, youth who hold more conventional beliefs or a stronger belief in rules have been more often shown to refrain from gang activity.

Numerous theories posit that gang membership is a response to blocked opportunities, status frustration, or other negative life events. Generally, there is consistent support that exposure to negative life events is associated with gang joining, but perceived (future) opportunities have not shown the same pattern. Experienced barriers to success or strain have been a strong basis for theories of gang membership in qualitative studies. Support for this notion in quantitative studies, both cross-sectionally and longitudinally, suggests that it may matter for male gang joining, but not for females.

Finally, gang joining may also be predicted by an individual's behaviors. As stated prior, the link between gang membership and increased rates of offending (e.g., general delinquency, violent offending and aggressive behavior, graffiti, official contact with justice system agencies, loitering/ hanging out/cruising, drug sales) has been well established. Much of this research is based on correlational studies where causal ordering could not be definitively established (i.e., which came first, the gang joining or the offending). However, cross-sectional studies that incorporate nondelinquent, problem behaviors and longitudinal studies that establish temporal ordering can be used to assess the degree to which behavior predicts gang joining. Sexual promiscuity or early sexual experiences, for example, have been found to predict gang joining. Alcohol use and drug use (both legal and illicit) and additionally, early alcohol or drug use have predicted gang joining.

Longitudinal studies have supported the notion that youth with delinquent histories may be more likely to join gangs, but the amplification effect of gang membership on the rate of delinquency (post-gang joining) has a very strong grounding. In other words, some studies show that gang members are more likely to have committed certain types of crimes before they joined gangs (e.g., violent, total delinquency), but it is also clear that the rate of offending significantly increases after youth are a part of the gang and then falls off after leaving. There has been very little support for the notion that involvement in prosocial activities decreases the likelihood of gang joining in either cross-sectional or longitudinal studies. Researchers have examined the role of religion (e.g., religious service attendance, "religiosity," or bonds to religion), employment (during school year and during the summer), and community-based activities like sports or events without noteworthy results.

Family Domain

There is some cross-sectional and longitudinal evidence that youth in families of lower socioeconomic status are more likely to become gang members. There are, however, more studies that fail to find the connection. The same can be said of research on family structure. There is inconsistent support for the role of nuclear family structure versus single parent households (or other family structures) on the likelihood of gang joining. Scholars have suggested that what is more important than the structure of the family is the quality of the parenting within the family. Risk factor studies often test the degree or quality of parental monitoring or control on youth. There is strong cross-sectional and longitudinal support for the idea that more parental monitoring decreases the likelihood of gang joining. The same has not been shown of the level of familial attachment. There is limited cross-sectional and longitudinal support that youth that are less attached to his/her parent or family are more likely to join a gang.

Many alternative family domain factors have been identified and proposed, but strong statements regarding generalization cannot yet be made. In a review of major risk factor studies, Klein and Maxson (2006) declare some results as "inconclusive" because an equal number of studies find support, fail to find support, or are inconclusive. There are also a number of factors that have not been extensively tested or exclusively tested in cross-sectional studies, and therefore, definitive conclusions would be premature. The role of familial deviance is still unclear. There is some support that the lack of parental role models, the presence of parents with prodeviant norms, or antisocial family members like siblings may increase the likelihood of gang joining. However, studies have also failed to show a link between actual familial deviance/violence and gang joining. There is preliminary evidence that familial or parental conflict, family bonding, and familial involvement and support are not as strongly associated, but these factors have had limited scrutiny. Other conceptually appealing factors like the effect of relatives in a gang have been presented as important in numerous qualitative studies (e.g., Vigil 1988), but have had limited quantitative testing.

School Domain

Most school domain risk factors are still considered "inconclusive." That is not to say that various factors have not been examined, but inconsistent results across studies and only a handful of longitudinal studies make it hard to generalize conclusions. While no variable can be argued as having "overwhelming" support, the factors with the most support at this point are educational aspirations, school achievement (as measured by grades at various times or testing), and attachment to school. These factors have been tested more often than others. The implication of these results is that the more strongly a youth is tied to school, the less likely he/she would become gang involved. There is preliminary cross-sectional and longitudinal support for the effect of negative labeling by teachers and schools on the likelihood of gang involvement. Studies have found that negative labels like "bad," "disturbed," and "disabled" are associated with gang involvement. The opposite effect was not shown. Being labeled as a "good student" did not buffer gang joining. Safety in schools has been tested less often and generally without much support.

Other factors that have been investigated in just a few studies and require more evidence before trends can be discerned are features like attachment to teachers, self-esteem about school, educational marginality, school environment, educational frustration, winning awards in school, truancy/absenteeism, participation in school sanctioned activities, and having gang members in the classroom.

Peer Domain

The peer domain seems particularly important given that gangs are a peer group. This is why temporal ordering is of particular concern to this domain of risk factors. Cross-sectional studies that report peer group characteristics at the same time as an individual reports gang membership could very well be reported on the qualities of the gang – which is interesting and informative but not predictive. Only longitudinal studies can decisively establish whether the characteristics of or commitment to certain peer group types increases the likelihood of future dealings with gangs.

While the wording of questions or scales can differ from study to study, peer domain variables have generally clustered around two types of measures: the quality of the peers present and the level of attachment to peers. In measuring the quality of the peers that are present, scholars generally attempt to capture whether youth surround themselves with delinquent and/or prosocial peers. Delinquent peers would be a risk factor, whereas the presence of many prosocial peers would indicate a potential buffer for gang joining. Cross-sectional studies often find that gang members are more likely to be surrounded by delinquent peers, but due to the lack of temporal ordering, it is not clear whether this is because their delinquent peers are, in fact, their fellow gang members. There is, however, also longitudinal support for the effect of delinquent peers to lead to gang membership. Fewer studies have looked at the buffering effect of having prosocial or *good* peers on the risk of gang joining. There is preliminary support that the presence of more prosocial peers is important though more work is certainly needed.

The second cluster of peer domain factors has been the extent of an individual's commitment to their peers. These measures have been tested less often and generally with weaker methodology, but the pattern generally follows that of the quality of peers present. Youth that are more strongly committed to their negative/delinquent peers or who have low commitment to prosocial peers are more likely to be in gangs. Youth that are more strongly committed to conventional or prosocial peers are less likely to join gangs.

Other factors that have had limited examination, but show some promise as risk factors, include the presence of or ability to resist peer pressure, the experience of being approached by a gang, and the time spent unsupervised or "hanging around." Future studies should begin to examine some of these factors more directly related to socialization with peers.

Community Domain

Despite being one of the first domains to emerge in the gang literature (mostly from qualitative studies), the community domain is one of the least often tested in the quantitative work on gang risk factors. In qualitative studies about gangs, the structural features of the community are quite prominent. Decades of research show that gangs often form in highly disadvantaged and disorganized communities that lack effective social controls. Therefore, it is known that community level factors should not be neglected as important predictors for individual-level gang membership. However, risk factor data are most commonly collected at the individual level (i.e., people are asked questions about themselves and their lives). These measures often capture an individual's perceptions of their community characteristics. These perceptions may or may not be consistent with the actual state of their neighborhood condition. Therefore, the most common method for testing community factors may not be sufficient measures of the actual phenomenon and may fail to find support even if it does exist.

Two common community measures are the level of (1) crime and (2) disorganization or disorder in a neighborhood. Studies have shown mixed results in regard to the effects of these two characteristics on the likelihood of gang joining for youth. There are longitudinal studies that show that community arrest rate and neighborhood disorganization predict later gang joining, though other studies (longitudinal and cross-sectional) do not. Other community related factors have been investigated, but not extensively. There is at least some support to suggest that perceived access to drugs in the neighborhood, attachment to a neighborhood, availability of firearms in the community, drug use in neighborhood, degree of poverty of the community, and the presence of gangs in the community should be considered.

Risk Factor Theories of Gang Membership

There are a number of theories of gang formation and gang membership. Many of the most popular theories are based on early qualitative studies on gang groups and their dynamics. Most of these theories fit under the broader umbrella of other major criminological theories. For example, many theories of gang formation have their roots in macro-crime theories like social disorganization, strain, and subculture of violence. Learning theories have also been widely cited as the mechanism to explain shared norms and their transmission among gang members.

More recently, however, the rise of longitudinal risk factor research has spurred developmental theories of gang membership. The risk factor approach assumes that there are multiple paths to the same outcome. The timing, order of risk factors, and cumulative nature play a dynamic role in the likelihood of gang joining in youth. These theories would not be possible (nor testable) without these multiple-wave, longitudinal studies that can capture stimuli at various points during development. Arguably, the most well-known theory and documented risk factor theory is Thornberry and colleagues' (2003) Interactional Theory of Gang Membership.

As the name suggests, this theory stresses that interactions between variables and bidirectional causality are important. According to this theory, it is not the case that exposure to every type of risk factor directly influences gang members. Factors like the neighborhood and family are predictive of gang membership indirectly because they loosen prosocial bonds. The decline of the bonds increases the chances for antisocial influences or values. A culmination of risk factors and disadvantage increases stress, delinquency, and general problem behaviors. Engaging in these behaviors then increases the perceived benefits of gangs (e.g., excitement, protection) and, hence, increases the odds of gang joining.

Thornberry and colleagues found support for their model using the Rochester Survey of Youth. Their theory focuses mostly on late childhood and early adolescent predictors of mid-adolescent gang membership. Since their work, other proponents of a developmental framework to explain gang membership have argued that the theory should incorporate earlier risk factors like at preschool or upon school entry (see Howell and Egley 2005) because a number of longitudinal studies (using gang and delinquency in general) have shown risk factors measured at earlier ages to be important. Regardless, risk factor research has made it possible for the understanding of gang joining and membership to be placed within a developmental context.

Future Directions

Risk Factors for Gangs vs. Other Serious Offenders

One important aspect of the gang risk factor literature that is still up for debate is the difference between gang members and nongang serious offenders. The studies reviewed here compared gang members with nongang youth. In this direct comparison, a number of significant differences can be found. The most noteworthy and consistent difference between gang and nongang members is involvement in delinquency and other problem behaviors. Gang members commit significantly more offenses than their nongang counterparts, and, as shown here, there are a number of risk factors that distinguish these two groups.

When gang members are compared to nongang, but *deviant* peer group members, their behavior is also shown to be significantly more delinquent. A number of gang scholars contend that gangs are substantively different than even nongang deviant peer groups. If this is the case, one would expect that there are unique risk factors that significantly distinguish gang members from these other serious offenders. There are a limited number of studies that have attempted to address this issue. In one study using the Denver Youth Survey (a longitudinal study of over 1,500 youth), the researchers found that while many factors distinguished gang members from the non-offending youth, few differences existed between the gang members and nongang street offenders. A separate study using a national sample of youth (n = 5,935) used the same 18 predictors on violent offenders and gang members. Of the 18 factors, 6 of them predicted both violent offending and gang membership in multivariate models (i.e., low guilt, neutralizations, delinquent peers, delinquent peer commitment, time spent with drugs and alcohol, and negative school environment). Four additional variables predicted violent offending but not gang membership (i.e., impulsivity, risk seeking, few prosocial peers, and time spent without adults). No factors independently predicted gang membership.

Scholars have struggled to predict gang membership above and beyond what we already know about delinquency more generally. Risk factors for gang members are very similar to those for violent and/or serious offenders (despite differences in level of behavior). This leads to the question of whether there is reason to study gangs as separate entities from other highly delinquent youth groups. The study using a national sample of youth also found that gang members, on average, had more cumulative risk factors than even violent offenders. This may be an important point of inquiry that requires further investigation. Perhaps, it is not specific risk factors that predict gang versus nongang serious offenders, but the sheer magnitude of cumulative risk factors that makes the difference. Gang risk factor research needs more investigation into the mechanisms or factors that lead gang members to gang groups as compared to other deviant peer groups.

Longitudinal Studies Are Needed to Move the Field Forward

True risk factor studies need to be longitudinal. Cross-sectional studies are informative, often more feasible to execute, and can be used to add support to longitudinal findings. However, they will never be able to definitively establish causal ordering and, thus, can never identify a true “risk factor” or predictor of future behavior. With newer theories like Thornberry and colleagues (2003) interactional framework for gang membership, more attention has been focused on pathway models as explanations for behavior. This makes intuitive sense. Human behavior is a very complicated thing to predict. It is highly unlikely that all factors simply work in a unidirectional fashion. The progress that the gang risk factor literature and theory building has made in recent decades has come from a handful of well-executed, longitudinal studies.

Unfortunately, there have been a limited number of these types of data collection efforts. They are costly, time-consuming, and labor intensive to conduct. The data available from these established studies have also been used quite extensively to study gang risk factors. No data collection effort can provide an infinite number of gang studies. The major longitudinal study efforts currently available also represent a limited geographic section of America, and especially with regard to gang research, diversity in location and population is an important consideration. Therefore, without more longitudinal data collection efforts, new information would not be available and gang risk factor work will not move forward. To complicate matters, tests of interactional theories (as opposed to simply direct effects of risk factors) require analysis with sufficient statistical power (i.e., lots of people to study). This requires studies with even more resources and time commitment. These types of studies are a monumental undertaking, but without more efforts, this area of work cannot make forward progress.

Related Entries

▶ Desistance from Gangs
▶ Gangs and Social Networks
▶ Interactional Theory of Delinquency

Recommended Reading and References

Craig WM, Vitaro F, Gagnon C, Tremblay RE (2002) The road to gang membership: characteristics of male gang and nongang members from ages 10 to 14. Soc Dev 11:53–67

Esbensen FA, Deschenes EP, Winfree LT Jr (1999) Differences between gang girls and gang boys: results from a multisite survey. Youth Soc 31:799–828

Esbensen FA, Huizinga D, Weiher A (1993) Gang and non-gang youth: differences in explanatory factors. J Contemp Crim Justice 9:94–116

Hill KG, Howell JC, Hawkins JD, Battin-Pearson SR (1999) Childhood risk factors for adolescent gang membership: results from the Seattle Social Development Study. J Res Crime Delinq 36:300–322

Howell JC, Egley A Jr (2005) Moving risk factors into developmental theories of gang membership. Youth Violence Juv Justice 3:334–354

Klein MW, Maxson CL (2006) Street gang patterns and policies. Oxford University Press, New York

Lahey BB, Gordon RA, Loeber R, Stouthamer-Loeber M, Farrington DP (1999) Boys who join gangs: a prospective study of predictors of first gang entry. J Abnorm Child Psychol 27:261–276

Thornberry T (1998) Membership in youth gangs and involvement in serious and violent offending. In: Loeber R, Farrington DP (eds) Serious and violent juvenile offenders: risk factors and successful interventions. Sage, Newbury Park

Thornberry TP, Huizinga D, Loeber R (2004) The causes and correlates studies: finding and policy implications. Juv Justice 9:3–19

Thornberry TP, Krohn MD, Lizotte AJ, Smith CA, Tobin K (2003) Gangs and delinquency in developmental perspective. Cambridge University Press, Cambridge

Vigil JD (1988) Barrio gangs: street life and identity in Southern California. University of Texas Press

Risk Factors for Prison Recidivism

David E. Olson
Department of Criminal Justice and
Criminology, Loyola University Chicago,
Chicago, IL, USA

Synonyms

Rearrest; Recidivism versus reoffending; Return to prison

Overview

One of the most frequently used measures to gauge the effectiveness and impact of criminal sentencing and correctional programs is the recidivism of those processed through the justice system and released from community corrections programs or prisons. Broadly defined, recidivism is the return to criminal behavior following some type of intervention by the criminal justice system; however, in practice and research, how recidivism is measured varies dramatically, and as a result, so too do rates of recidivism. Based on one definition that has been used for those released from prison – whether or not the former inmate was rearrested for a new crime within 3 years following release – the recidivism rate of a nationally representative sample of released inmates in the United States was 68 %; however, when defined as returned to prison as a result of a new conviction, the recidivism rate was 25 % (Langan and Levin 2002). In addition to recidivism rates varying depending what definition of recidivism is used, the likelihood of recidivism also varies depending on a variety of individual-level and community-level risk factors, as well as state and local criminal justice practices and policies. Over the past four decades, risk factors for prisoner recidivism have increasingly been used to guide criminal justice practice and policy, and the understanding within the field of criminal justice and criminology as to the factors associated with prisoner recidivism have increased in depth and sophistication. This knowledge has helped improve the targeting of limited criminal justice resources to those who pose the highest risk of recidivism and those who can benefit most from rehabilitative interventions that can reduce risk of recidivism. These advances in understanding of the risk factors associated with prisoner recidivism have also improved the designs of evaluations of program and policy interventions intended to reduce recidivism.

Definitions and Measures of Recidivism

The broad definition of recidivism is the return to criminal behavior following some type of intervention by the criminal justice system, such as following an individual's conviction and placement on probation or following their release from prison. Although there are a number of purposes or goals when it comes to sentencing those convicted of crimes, including individual and general deterrence, retribution/punishment, incapacitation, and rehabilitation, only two of these goals actually seek to reduce criminal behavior following the completion of a criminal sentence: individual deterrence and rehabilitation. Thus, the concept of recidivism within this context can be seen as a measure of the degree to which these interventions were effective. "High" recidivism rates can be used to draw the attention of public policy makers and practitioners to interventions that may need to change or are used to illustrate the difficulty of deterring or rehabilitating specific types of offenders. Varying levels of recidivism rates across programs or services, or among specific types of offenders, can also be used by practitioners and policy makers to invest more, or refer more offenders, to those interventions that seem to be "working" to reduce recidivism.

However, as important as recidivism is as a measure of the effectiveness of criminal justice interventions, the definition of what constitutes "recidivism" in the domains of both criminal justice researchers and practitioners varies considerably. While the conceptual definition of recidivism is the return to criminal behavior, how exactly have researchers and practitioners operationalized this concept so that it can be measured in the real world? The operational definitions and methods of measurement have ranged from researchers interviewing former prison inmates and asking them to self-disclose new criminal behavior, such as illegal drug use or theft, to state departments of corrections determining how many of the inmates they released from prison have been readmitted to prison. Because the specific measure used to gauge recidivism has such a significant impact on the level of recidivism that will be detected, it is important here to briefly describe the measures of recidivism that have been most frequently used in examining those released from prison and their relative strengths and weaknesses.

In addition, the amount of time following an inmate's release from prison during which recidivism is measured also varies considerably from study to study, ranging from relatively short periods of time at risk to much longer postprison follow-up periods. Generally, recidivism research has found that if a prison release is going to recidivate, it tends to happen relatively soon after their release and as the longer they survive without recidivating, the lower the likelihood that it will occur.

As a result of these different methods of operationalizing and measuring recidivism, recidivism rates can vary considerably and also have varying levels of reliability and validity. For example, many practitioners and policy makers would question the truthfulness (i.e., reliability) of former inmates reporting new criminal behavior to a researcher, regardless of assurances of anonymity or confidentiality that may be used to improve the reliability of the responses. In addition to the concerns regarding the reliability of self-reported information about the participation of former inmates in new criminal activity, one of the biggest limitations to the use of self-reported information as a means of gauging postprison recidivism is the enormous amount of time and resources needed to collect this information from sufficiently large samples of former inmates.

Because of the costs and concerns with the reliability of obtaining self-reported information to gauge recidivism, most researchers, agency administrators, and policy analysts examining prisoner recidivism utilize “official” records of recorded events, such as new arrests, convictions, or admissions to prison. However, while relatively easy to access and obtain, operationalizing and measuring recidivism using official records of arrests, convictions, or returns to prison also have their own limitations, and recidivism rates vary dramatically depending on which of these measures is selected. Illustrative of this is recidivism analyses performed on a nationally representative sample of prison releasees by the US Department of Justice’s Bureau of Justice Statistics which utilized four different measures of recidivism (Langan and Levin 2002). Specifically, the recidivism of a cohort of inmates released from state prisons in 1994 was examined using rates of rearrest, reconviction, return to prison for any reason (a new conviction or a technical parole violation), and return to prison as a result of a new conviction, all within 3 years of release. While somewhat dated, the results from that study are illustrative and useful for revealing the differences in recidivism rates across different definitions of recidivism. For example, when the operational definition of recidivism was rearrest for a new crime, 68 % of prison releasees recidivated, compared to a recidivism rate of 52 % when defined as being returned to prison for a new crime or a technical violation of their release conditions, 47 % when recidivism was defined as being reconvicted for a new crime, and, lastly, a recidivism rate of 25 % when defined as being resentenced to prison for a new crime (Langan and Levin 2002). Thus, depending on the operational definition of recidivism selected, the overall rates of recidivism will vary considerably, but, again, there are concerns regarding the validity of these measures. For example, some would not feel as though being convicted, or returned to prison as a result of a new conviction, is a sufficiently broad (i.e., valid) definition of recidivism, and it is likely that if it were possible to obtain reliable self-reported information regarding involvement in crime that the recidivism rate would be considerably higher. Similarly, many have pointed out that simply being arrested for a new crime does not necessarily provide a valid measure of recidivism, but merely that police had probable cause that the individual had committed a crime, and it is also obvious that not all criminal behaviors come to the attention of the police or result in arrests.

Fundamentals of Risk Factors for Prison Recidivism

While the operational definition used to measure recidivism results in different recidivism rates, regardless of the specific measures of prison recidivism used, a number of risk factors have been identified as having a consistent relationship to recidivism and have played a significant role in criminal justice policy and practice. Indeed, one of the oldest and most fundamental decision-making processes in community corrections, including the supervision of those released from prison, is the assessment of offender risk (Holsinger et al. 2001). The problem of predicting offender risk is paramount to decision making at various stages of the criminal justice system and is integral to several decisions regarding the release of those from prison, including decisions as to whether or not they should be released and, if released, the level and conditions of parole supervision. Over the past 30 years, the tools and approaches to risk assessment of prison releasees and probationers have evolved dramatically, and experts in the field of community corrections have identified three generations of risk assessment: subjective or clinical judgments, actuarial or case classification approaches, and criminogenic/dynamic evaluations (Bonta 1996). From the development of these risk assessment instruments, and extensive research on recidivism of those released from prison, a number of general conclusions regarding risk of prisoner recidivism have now become widely accepted within the field of criminal justice and criminology, while other risk factors are either evolving in terms of better understanding prisoner recidivism, and still others are the subject of considerable debate among scholars, practitioners, and policy makers. These risk factors associated with prisoner recidivism can be grouped, generally, into three main categories: (1) individual-level risk factors, (2) community-level risk factors, and (3) criminal justice policies and practices that influence the likelihood of recidivism.

Individual-Level Risk Factors of Prisoner Recidivism

The first category of risk factors for prisoner recidivism include the characteristics or traits of the individual prison releasee and have been characterized by Andrews and Bonta (1994) as falling into one of two subcategories of individual risk: individual static characteristics and individual dynamic characteristics. Static risk factors are those that cannot be changed as a result of correctional interventions and include traits such as age at first arrest, prior arrests, current age, current conviction offense, and gender. Dynamic risk factors, on the other hand, are characteristics of the offender that can be changed, either through clinical interventions or by changes in the former prison inmate's circumstances, such as antisocial thinking patterns and cognition, drug use patterns, employment status, and having antisocial or criminal peer groups and associates.

As a result of numerous individual studies and meta-analyses examining prison inmate recidivism, a number of individual-level, static characteristics have been consistently found to be associated with postprison recidivism. For example, a meta-analysis conducted by Gendreau et al. (1996) found the offender's criminal history, history of other antisocial behaviors, age, the environment within which the inmate was raised as a child (family criminality, family structure, and family-rearing practices), race, and gender were all related to the risk of recidivism. Two of the strongest individual-level, static predictors of recidivism in the Gendreau et al. (1996) study, as well across numerous individual studies examining offender populations (i.e., prison releasees and probationers), were the individual's age and the extent of their criminal history.

Generally, the older the individual when discharged from their correctional programming (i.e., prison or probation), the lower the risk of recidivism, regardless of how recidivism has been operationalized and regardless of what other risk factors are statistically controlled. While the general criminology literature has found a strong association between age and involvement in crime, or desistence from crime, even among those who are known offenders (i.e., those released from prison), this pattern also holds. Similarly, those released from prison with more extensive criminal histories, whether operationalized as prior arrests, prior convictions, prior prison sentences, or other measures of previous antisocial behavior, have a higher risk of continued involvement in crime (i.e., recidivism) independent of other risk factors or characteristics. Males released from prison also tend to have higher rates of recidivism than females, even after the influence of other characteristics have been accounted for or statistically controlled. The current conviction offense has also been found to be associated with recidivism of those released from prison, with those released from prison after having served a sentence for a property crime or a drug-law violation having a higher risk of recidivism than those who had served a sentence for a violent crime. However, the magnitude of effect for this risk factor tends to lessen after other risk factors are controlled for, such as age at release from prison, prior criminal and substance abuse history, and education level. In fact, the static, individual-level risk factors of age, criminal history, gender, and current conviction offense are even consistent predictors of prisoner recidivism in the few studies of inmate recidivism that have been performed in nonindustrialized Western nations, such as in Malta, a small republic in the Mediterranean Sea (Baumer 1997).

The Gendreau et al. (1996) meta-analysis also found the race of the offender to be related to recidivism in the small number of studies that included this as a variable, with minorities having a higher likelihood of recidivism, and some more recent studies have also found that minorities have significantly higher odds of recidivating than white offenders (DeComo 1998; Strom 2000). However, some studies suggest that in fact there is *no* statistically significant difference in the recidivism rates of minority and white juveniles (Mbuba 2005), particularly after other risk factors are taken into consideration. Thus, while some studies have found race to be predictive of recidivism for those released from prison, others have found the magnitude of the race effect on recidivism to vary or be neutral depending on what other variables are statistically accounted for in the analyses. There are also a large number of recidivism studies that have not included race as an individual-level risk factor in the analyses.

However, from the standpoint of criminal justice policy and practice, one of the biggest limitations of these static risk factors is that, by definition, they cannot be changed. Thus, while they can be used to aid in predicting the risk of recidivism, and as a result subject the individual to a higher level of supervision, they cannot always be effectively used to reduce the risk of recidivism. Therefore, it is critical to point out that there are also numerous individual-level risk factors that are considered dynamic or malleable and, if changed, can be viewed as assets or protective factors. For example, while an individual's history of drug use is predictive of their likelihood of recidivism, their current drug use patterns, their cognitive views on the appropriateness of drug use, and their completion of substance abuse treatment while in prison can reduce the likelihood of recidivism. Although the efficacy of prison-based treatment as a means of reducing the risk of recidivism was challenged in the 1970s, leading to a number of policy shifts that reduced the availability and emphasis on rehabilitation for prison inmates, there is currently agreement among most criminologists that treatment addressing dynamic criminogenic needs and utilizing evidence-based practices can reduce the risk of recidivism, which has led to a revival of rehabilitative efforts both within prisons and following release from prison.

Again, a number of meta-analyses have identified the individual-level, dynamic risk factors associated with recidivism among those released from prison, including antisocial personalities; socialization with other offenders; antisocial attitudes related to behavior, employment, and education; conflict with family; personal distress, such as anxiety and low self-esteem; social achievement, such as employment, marital status, and education; and drug use patterns (Gendreau et al. 1996). Thus, interventions that seek to alter a prison inmate's or a parolee's antisocial attitudes, teach them to more effectively handle anxiety or interpersonal conflict in noncriminal ways (i.e., without the use of drugs or violence), or improve their educational achievement or employability can change these dynamic risk factors and, as a result, reduce the risk of recidivism. Evidence regarding the impact these types of interventions can have on recidivism is also quite plentiful. For example, Wilson et al. (2000) found in a meta-analysis of corrections-based education, vocational and work programs that participants generally had lower rates of recidivism, although the existing literature is still limited in terms of fully understanding the causal relationship of these programs versus issues of selection bias regarding who participates in these types of programs. Another intervention used within prisons to addresses individual-level, dynamic risk is cognitive-behavioral therapy (CBT), which views an offender's cognitive processes related to crime as something that can be changed. "Cognitive-behavior therapy is based on the assumption that cognitive deficits and distortions characteristic of offenders are learned rather than inherent. Programs for offenders, therefore, emphasize individual accountability and attempt to teach offenders to understand the thinking processes and choices that immediately preceded their criminal behavior" (Lispy et al. 2007). In their meta-analysis of CBT programs, Lipsy et al. (2007) found that, overall, CBT interventions were effective at reducing recidivism and that some specific characteristics of programs generated greater recidivism reductions, including how well the programs were implemented (i.e., fidelity to program design), targeting of higher risk offenders, and the incorporation of specific CBT elements related to anger management and interpersonal problem solving.

Prison-based programs that seek to address issues related to an inmate's substance abuse patterns, also an individual-level dynamic risk factor, are another means by which the risk of recidivism can be reduced upon an inmate's release back to the community. One of the most effective modalities of providing prison-based substance abuse treatment to inmates is through the use of Therapeutic Communities (TCs). TCs are "residential [programs] that use a hierarchical model with treatment strategies that reflect increased levels of personal and social responsibility. Peer influence, mediated through a variety of group processes, is used to help individuals learn and assimilate social norms and develop more effective social skills" (National Institute on Drug Abuse 2002). Also, because TCs are one of the most common and widely studied drug treatment modalities for prison inmates (Lurigio 2000), their efficacy in reducing the risk of recidivism among prison releasees has been well documented and established. Mitchell et al. (2006) is one of the most recent, comprehensive, and rigorous meta-analyses published on the effectiveness of incarceration-based drug treatment. In general, most of the research on prison-based TCs, including Mitchell et al. (2006), has documented recidivism reductions for those inmates who participated, although the magnitude of the reduction varied, depending on the length of stay, the population served, the inclusion of educational and vocational programming, and the provision of access to aftercare following release from prison.

Although not generally considered among the individual-level, dynamic risk factors for recidivism, the amount of time spent in prison is certainly a factor that can be modified and altered by criminal justice practitioners and policy makers if it were found to influence recidivism. Yet despite the fact that longer prison sentences have consistently been argued as an effective crime control policy based on the belief that longer sentences would generate individual deterrence and thus reduce subsequent recidivism, much of the research that has sought to examine this question is dated, suffers from methodological limitations, and has produced inconsistent and, at times, contradictory findings. For example, a meta-analysis by Gendreau, Goggin, and Cullen (1999) found that longer prison sentences were associated with an increased likelihood of recidivism, contrary to what would be expected through the lens of individual deterrence; however, Snodgrass et al. (2011) points out that this meta-analysis was based primarily on research conducted in the 1970s, when sentence lengths among prison inmates were dramatically different than they are today. In their recent examination of the question as to whether or not length of time in prison reduces recidivism among prison releasees in the Netherlands, Snodgrass et al. (2011) found there to be no effect of time served on subsequent recidivism. More generally than just length of time served, a meta-analysis examining the relationship of harsh penalties and recidivism found that harsher punishments were related to a slight increase in recidivism, but when it came to sentence lengths, sentences under six months showed no impact on recidivism while sentences were found to increase the risk of recidivism (Smith et al. 2002). Thus, while the research literature regarding the impact of length of stay in prison on subsequent recidivism appears to suggest that it either has no effect or increases the risk of recidivism, many would also argue that the reduction of recidivism is only one goal in the sentencing of convicted offenders to prison, and thus longer sentences would be justified on the basis of retribution and incapacitation.

The Universal Nature of Individual-Level Risk Factors

It is important to note that many of the individual-level risk factors for recidivism, including both static and dynamic risks, may not be consistent or have the same magnitude of effect to all subpopulations of prison releasees. For example, since the majority of prison releasees are accounted for by males, and therefore much of what is known about risk factors for prison recidivism has been based on research predominantly of male offenders, concern regarding the generalizability of risk factors specifically to female prison releasees has been raised. Similarly, the degree to which risk factors are consistently predictive of "general" recidivism versus recidivism for specific types of crimes, such as violent offenses, also varies. For example, while under supervision or during reentry from prison, unemployment or job dissatisfaction – an individual-level dynamic risk factor – is related more to men's recidivism than to women's (Benda 2005). On the other hand, greater educational achievement significantly lowers women's risk of recidivism, but has less effect on men's recidivism (Uggen and Kruttschnitt 1998). Similarly, research has produced inconsistent findings on whether having friends who engage in criminal lifestyle – which has been identified in meta-analyses of increasing the risk of recidivism – increases the risk of recidivism for both men and women (Uggen and Kruttschnitt 1998; Benda 2005).

In addition to factors that increase or decrease the overall likelihood of recidivism, practitioners and policy makers are often concerned with recidivism for very specific types of crimes. For example, it is of greater concern from the standpoint of public safety when the recidivism involves crimes of violence against a victim than when the recidivism involves minor property crimes or drug-law violations. Similarly, for offenders sentenced to prison for sexual assault, upon their release from prison one of the biggest concerns among practitioners and policy makers is their likelihood of committing another sexual assault. However, because violent recidivism, and even more specifically sexually violent recidivism, is much less prevalent than more general recidivism, it tends to be more difficult to accurately predict. Still, a review of research on the accuracy of standardized risk assessment scales of predicting violent recidivism indicates that scales containing individual-level dynamic risk factors have higher predictive accuracy than scales that contain only individual-level static risk factors (Campbell et al. 2009).

Community-Level Risk Factors

While there is a fairly extensive body of literature and significant agreement on the individual-level risk factors that contribute to recidivism among those released from prison, as a result of the growing interest among researchers and policy makers in the area of prisoner reentry, it has increasingly been recognized that the characteristics of the community the offender is released back into may also influence the likelihood of recidivism. Moreover, these environmental conditions may modify the predictive accuracy of certain individual-level risk factors. For example, offenders who have a history of substance abuse are more likely to relapse and use illicit drugs (i.e., recidivate) in high-poverty areas than in neighborhoods with low poverty rates (Galea et al. 2003). Thus, the risk assessment field has recently recognized the importance of incorporating risk factors that relate also to the criminogenic aspects of an offender's surrounding environment in assessing offenders' risk of recidivism. The recognition of this issue by policy makers and practitioners, coupled with the application of more sophisticated statistical techniques to examine recidivism (i.e., hierarchical linear modeling), has produced a small but growing body of scientific literature that has examined the influence of community-level characteristics on prisoner recidivism.

In the limited number of studies that have been published, there appears to be some confirmation that community or neighborhood characteristics related to poverty, crime and disorder, economic opportunity, and the availability of social services does influence, independent of individual-level risk factors, the risk of recidivism among prison releasees. For example, Kubrin and Stewart (2006) incorporated Massey's (2001) Index of Concentration at the Extremes (ICE), a community-level measure of the degree of concentrated affluence relative to the concentration of poverty, in their analyses of recidivism among individuals released from prison and found that inmates released into neighborhoods with higher concentrations of affluence had a lower risk of recidivism independent of their individual-level risk factors. Similarly, Hipp et al. (2010) found that inmates released to communities with high levels of concentrated disadvantage and disorder had a higher risk of recidivism, independent of their individual-level risk factors, but communities where social service providers were not in close proximity to the released inmate also increased the risk of recidivism after controlling for the influence of other factors. A similar pattern was found by Olson et al. (2009), which suggested that inmates released back to communities where aftercare services for inmates are more readily available, and more varied, had higher rates of aftercare participation and completion, which reduced the likelihood of both parole violations and recidivism. Thus, while the Urban Institute's research on reentry across a number of states has documented that the communities many inmates are released to have high levels of social and economic disadvantage (LaVigne et al. 2004), the increased risk that this poses regarding recidivism may be mitigated if there are higher levels of social service programs available. As the research regarding the independent influence of community-level risk factors for prison recidivism increases and evolves, it will be possible to better isolate the relative influences that individual-level versus community-level risk factors play in prisoner recidivism.

The Influence of Criminal Justice Practice and Policy on Prison Recidivism

The last of the three general areas that can change the risk of recidivism has to do with changes and differences in criminal justice practice and policy. Among all of the risk factors presented, the degree to which criminal justice practice and policy impacts recidivism had more to do with how recidivism is operationalized and interpreted. For example, if return to prison within three years is the definition of recidivism, and one of the ways that a former inmate can be returned to prison is the result of violating the conditions of their release, then it is obvious that policies related to the extent and nature of post-release conditions of supervision, staffing levels, and practices of parole agents, and the policies regarding how violations of release conditions are handled, will all influence the rate at which individuals released from prison are returned to prison (i.e., "recidivism"). Generally, those released from prison that are subject to a higher frequency of supervision contacts and more intensive monitoring have higher rates of violations of release conditions being detected and, as a result, higher rates of being returned to prison. The level and intensity of postprison supervision can change over time within a specific state and also varies considerably across states.

For example, in some states, such as California, those returned to prison for parole violations accounted for 65 % of all prison admissions in 2010, while at the other end of the spectrum were the states of Florida, Virginia, North Carolina, Idaho, Ohio, and Nebraska, where parole violators accounted for less than 10 % of all prison admissions (Guerino et al. 2011). Generally, the larger the proportion of prison releasees who are subject to supervision and other conditions of release, the larger the proportion of admissions accounted for by parole violators and the higher the "recidivism" rate if measured by rates of return to prison. If only a small proportion of prison releasees are supervised and have conditions, it follows that only a small number of prison releasees can be returned to parole violations. Also, some of these changes in practice and policy can be gradual, whereas other changes can be implemented quickly and dramatically. Thus, it is important to note that policies related to whether or not those released from prison are subject to supervision, what other conditions of release are required and monitored, and how long that supervision lasts vary from state to state and, as a result, can produce different rates of "recidivism" when measured as return to prison. Further, before one can conclude that recidivism rates based on return to prison in a state have changed or are higher or lower than another state, changes or differences in postprison supervision policies and practices need to be considered and accounted for in analyses.

However, it is not only correctional or parole policy and practice that can influence and impact recidivism rates of those released from prison but those of police departments and the courts as well. For example, if individuals are released from prison during a period when police departments are cracking down or focusing on specific types of crimes or offenders, such as drug delivery offenses or parolees in specific neighborhoods, then inmates released from prison during that period of time may be more likely to be arrested than those released from prison when police officers were being deployed for other purposes or periods when police presence was lower. Thus, if the definition of recidivism is rearrest for a new crime, then obviously changes or differences in police policy and practice over time and across place can have an influence on this measure of recidivism. The same can be said for measures or recidivism that rely on reconviction as the operationalized definition of recidivism: changes in the practices and policies of prosecutor's offices regarding what types of offenses they will file charges for, take to trial, or plea bargain can potentially influence the likelihood of a prison releasee being reconvicted of a crime or a felony offense following their release. Importantly, all of these changes in policy and practice – parole supervision intensity, police department activities, and prosecutorial decision making – can impact each other and have differential impacts on the different measures of "recidivism" (i.e., rearrest, reconviction, return to prison for any reason or specifically as a result of a new conviction).

Conclusions and Directions for Future Research

The body of research and knowledge on the risk factors of prison recidivism has increased and improved dramatically over the past 40 years and is increasingly being used by criminal justice practitioners and policy makers. This knowledge of risk factors for prison recidivism is increasingly being used by clinicians working with prison inmates to better identify and target the individual-level dynamic risk factors that need to be addressed, by parole boards in deciding which inmates can be released from prison due to their lower risk of recidivism, and by policy makers as they seek to allocate resources to programs, interventions, and community-based services that can reduce recidivism among those released from prison. Still, it must be recognized that while the use of more sophisticated approaches to identifying and classifying the risk of prison recidivism has improved predictive accuracy and guided case management strategies and practices, the science is not perfect and is continually evolving and improving. When tasked with the challenge of predicting human behavior (i.e., recidivism), even when we can take into consideration individual-level static and dynamic risk factors, coupled with community-level risk factors, there are bound to be false positives and false negatives: there will be individuals who are low risk based on all of these factors that will reoffend, as well as individuals who are high risk based on their risk factors who end up not reoffending. The latter is unlikely to be a concern for criminal justice practitioners or policy makers, but the former will be of considerable concern and oftentimes results in the questioning of using actuarial risk to determine if an inmate should be released from prison and how they should be supervised. However, generally speaking, the criminal justice system can identify and utilize risk factors for recidivism much more effectively than in the past to address individual-level risks, reduce recidivism, and ultimately, improve public safety. Future research that improves the understanding in the field of the interactions of individual- and community-level risk factors, and can effectively articulate how this knowledge and information can be used within the applied setting of criminal justice practice and policy, will go a long way towards continuing to improve public safety and reduced risk of recidivism for prison releasees.

Recommended Reading and References

Andrews DA, Bonta J (1994) The psychology of criminal conduct. Anderson, Cincinnati

Baumer E (1997) Levels and predictors of recidivism: the Malta experience. Criminology 25(4):601–628

Benda B (2005) Gender differences in life-course theory of recidivism: a survival analysis. Int J Offender Ther Comp Criminol 49(3):325–342

Bonta A (1996) Risk-needs assessment and treatment. In: Harland A (ed) Choosing corrections options that work: defining the demand and evaluating the supply. Sage, Thousand Oaks, pp 4–54

Campbell MA, French S, Gendreau P (2009) The prediction of violence in adult offenders: a meta-analytical comparison of instruments and methods of assessment. Crim Justice Behav 36(6):567–590

DeComo RE (1998) Estimating the prevalence of juvenile custody by race and gender. Crime and Delinq 44(4):489–503

Galea S, Ahern J, Valhov D (2003) Contextual determinants of drug use risk behavior: a theoretic framework. J Urban Health 80(3):iii50–iii58

Gendreau P, Little T, Goggin C (1996) A meta-analysis of the predictors of adult offender recidivism: what works! Criminology 34(4):575–607

Gendreau P, Goggin C, Cullen FT (1999) The effects of prison sentences on recidivism. Public Works and Government Services Canada, Ottawa

Guerino P, Harrison PM, Sabol WJ (2011) Prisoners in 2010, appendix table 11, sentenced prisoners admitted and released from state and federal jurisdiction, by type, December 31, 2010. U.S. Department of Justice, Washington, DC

Hipp RJ, Petersilia J, Turner S (2010) Parolee recidivism in California: the effect of neighborhood context and social service agency characteristics. Criminology 48(4):947–979

Holsinger AM, Lurigio AJ, Latessa EJ (2001) Practitioners' guide to understanding the basis of assessing offender risk. Fed Probat 65:46–50

Kubrin CE, Stewart EA (2006) Predicting who reoffends: the neglected role of neighborhood context in recidivism studies. Criminology 44(1):165–197

Langan P, Levin D (2002) Recidivism of prisoners released in 1994. U.S. Department of Justice, Washington, DC

LaVigne N, Visher C, Castro J (2004) Chicago prisoners' experiences returning home. The Urban Institute, Washington, DC

Lispy MW, Landenberger NA, Wilson SJ (2007) Effects of cognitive-behavioral programs for criminal offenders. Campbell Systematic Reviews 2007:6

Lurigio AJ (2000) Drug treatment availability and effectiveness: Studies of the general and criminal justice populations. Crim Justice Behav 27(4):495–528

Massey DS (2001) The prodigal paradigm returns: Ecology comes back to sociology. In: Booth A, Crouter A (eds) Does it take a village? Community effects on children, adolescents, and families. Lawrence Erlbaum, Mahwah

Mbuba JM (2005) A refutation of racial differentials in the juvenile recidivism rate hypothesis. African J Criminol Justice Stud 1(2):52–68

Mitchell O, Wilson DB and MacKenzie DL (2006) The effectiveness of incarceration-based drug treatment on criminal behavior. Campbell Systematic Reviews 2006:11

National Institute on Drug Abuse (2002) Therapeutic community–research report series. Department of Health and Human Services, Washington, DC

Olson DE, Rozhon J, Powers M (2009) Enhancing prisoner reentry through access to prison-based and post-prison aftercare treatment: Experiences from the Illinois Sheridan Correctional Center Therapeutic Community. J Exp Criminol 5(3):299–321

Smith P, Goggin C, Gendreau P (2002) The effects of prison sentences and intermediate sanctions on recidivism: general effects and individual differences. Solicitor General Canada, Ottawa

Snodgrass GM, Blokland AAJ, Haviland A, Nieuwbeerta P, Nagin DS (2011) Does the time cause the crime: an examination of the relationship between time served and reoffending in the Netherlands? Criminology 49(4):1149–1194

Strom K (2000) Profile of state prisoners under age 18, 1985–97. Bureau of Justice Statistics, Special Report (NCJ, 193427). U.S. Department of Justice, Office of Justice Programs, National Institute of Justice, Washington, DC

Uggen C, Kruttschnitt C (1998) Crime in the breaking: Gender differences in desistance. Law Soc Rev 32(2):339–366

Wilson DB, Gallagher CA, Mackenzie DL (2000) A meta-analysis of corrections-based education, vocation, and work programs for adult offenders. J Res Crime Delinq 37:347–368

Risk Management in Policing

Carol A. Archbold
Department of Criminal Justice and Political Science, North Dakota State University, Fargo, ND, USA

Introduction

Police officers are exposed to risk every time that they put their uniforms on. This exposure can result in them being named in citizen complaints, liability claims and lawsuits. Liability claims and lawsuits can have a significant financial impact on police agencies. For example, New York City paid out $500 million in settlements and judgments as a result of police actions from 2005 to 2010 (Lui 2011). From 2000 to 2007 the city of Chicago paid out approximately $126 million as a result of litigation involving the Chicago Police Department (Midwest Human Rights 2011). Litigation costs for American police agencies have increased significantly since the late 1960s. Some police executives have recognized this increase and have started to take proactive measures to reduce officer exposure to risk. Risk management is one of the latest proactive measures being used by some police executives to manage liability.

Fundamentals

Risk management is a tool that can help identify and manage potential risks and liability problems within organizations. Organizations adopt risk management because of increasing costs associated with liability claims and litigation, to reduce the risk of harm to their clients and their employees, and to provide a higher quality of service to their clientele (Wong and Rakestraw 1991).

In the past, some people believed that the principles and practices associated with risk management were not applicable to public agencies. This perception is inaccurate as most public agencies can function better if they have the maximum amount of resources available to them. In addition, it could be argued that public organizations are held to a higher level of accountability than private organizations because they are entrusted with budgets that largely come from the community (i.e., taxpayers) (Vincent 1996). If public agencies use risk management to identify risks that often lead to organizational loss, and then make changes with the hope of reducing future loss, they will be able to spend more resources on services to the community instead of settlements for liability claims and litigation. Today, the public sector is among the fastest growing segments of the risk management profession (Vincent 1996).

There are five basic steps in the risk management process that apply to most types of organizations (Ashley and Pearson 1993; Wong and Rakestraw 1991):

1. ***Identify risks, frequency of exposure to risks, and the severity of losses resulting from exposure to risks.*** To do this, organization members can use data on the history of organizational loss including past litigation and monetary payouts resulting from liability claims. There are three general categories of assets that organizations are responsible for managing including: physical, human and financial assets (Young 2000). Vehicles, buildings, computer equipment and other technological equipment are considered **physical assets**. If this type of asset is damaged or lost, organizations face costs associated with replacing the item and also any time that is lost during the replacement of the item. **Human assets** include all employees that work within an organization. Employees are at risk for both physical harm and financial difficulties if injured while on the job. It also costs organizations money to retrain new employees when they need to replace employees that are killed or injured while on the job. **Financial assets** include any financial resources that are available to organizations that have a direct impact on how organizations are able to function. In public organizations this could include annual budgets and any external grants that can be acquired by an organization.
2. ***Explore methods to manage exposure to identified risks***. This step requires a thorough review of an organization's policies and procedures, training, and the supervision and accountability of employees.
3. ***Choose an appropriate response to manage exposure to identified risks.*** This might require making changes to policies, employee training, and/or supervision in order to reduce exposure to identified risks.
4. ***Execute the response that was chosen to manage exposure to risks***.
5. ***Evaluate the impact that the response has on exposure to identified risks.***

The risk management process is not over once the first four steps are completed; this process is ongoing because organizational risks can change over time.

The use of risk management by police agencies seems logical as most of the basic duties associated with police work expose officers to risk (Gallagher 1990b; Wennerholm 1985). Some of those duties include making arrests, pursuing fleeing suspects using patrol cars, using both lethal and non-lethal force on citizens, and serving felony warrants. There are three general benefits of using risk management as part of a police liability management strategy: (1) increase the safety of police officers and the citizens that they serve; (2) increase the quality of police service provided to the public; and (3) financial management of the costs associated with police-involved liability incidents that often result in payouts for liability claims or settlements for lawsuits (Young 2000).

The authorization of the use of force is an aspect of police work that makes it unique to all other professions. This distinctive aspect of police work also contributes greatly to police officer exposure to high levels of risk, which in turn, can lead to litigation, liability claims or citizen complaints. Improper use of lethal and non-lethal force by police officers during arrests, and improper service of due process are two incidents where damages are commonly sought and that settlements are paid out to citizens (del Carmen 1991). Liability Assessment and Awareness International (LAAW) asserts that liability related to police officer use of force can be managed by close supervision (Ashley and Pearson 1993). This group also suggests that department policies should be restrictive when identifying situations where use of force is deemed appropriate in order to keep this type of officer behavior under control. In addition, keeping standards high while reviewing all cases of alleged misuse or abuse of force will allow police executives to get a handle on this type of incident (Ashley and Pearson 1993). These suggestions made by LAAW mirror the findings of research on the use of force, which suggest that restrictive policies and enforced mandatory reporting of use of force incidents can harness this particular police officer behavior (Alpert and Fridell 1992; Pate and Fridell 1995).

Another common police activity that exposes police officers to risk is arrest (Kappeler 2006). In some cases, citizens make claims that police officers made a “false arrest.” False arrest can be defined as “the unlawful seizure and detention of a person” by a police officer (Kappeler 2006). False arrest claims consist of those cases where individuals believe that police officers did not have probable cause to stop or question them or unlawfully detained them. To avoid liability resulting from allegations of false arrest, police officers can obtain a warrant prior to making arrests (when possible). Securing a search warrant prior to arrest takes the authority of determining probable cause away from police officers and places the responsibility on judges. Because making arrests is such a common police action, it is not always possible to obtain a warrant prior to making an arrest. It would also be beneficial if police officers receive thorough training on when it is and is not appropriate to make arrests.

Another high-risk liability incident that often results in organizational loss is the operation of police vehicles. Typical liability incidents involving police vehicles include participation in high speed pursuits, the use of emergency lights and equipment when police vehicles are parked, the placement of people and vehicles during traffic stops and the investigation of traffic accidents, and the failure to use or misuse police vehicle occupant restraints (Ashley and Pearson 1993). Geoffrey Alpert and his colleagues discuss some of the ways to identify and respond to exposure to risks related to police vehicle operations in their book, *Police Pursuits: What We Know*. Alpert and his colleagues suggest that police executives scrutinize the history of their organization’s losses, review both past and pending lawsuits filed against their organization, and then survey police personnel to measure their knowledge of policies and liability issues associated with the operation of police vehicles (Alpert et al. 2000). Since the operation of police vehicles is an essential and required part of policing, the identification and management of the risks associated with vehicle operation becomes a necessary task for police executives. Similar to the research on police officer use of force, the research on police pursuits has found that clarification of pursuit policies, enforced mandatory reporting, and constant monitoring of pursuit activities will reduce the frequency of pursuits, as well as injuries and accidents that can result from this activity (Alpert and Fridell 1992).

Police agencies become exposed to risk when police services are not conducted within defined jurisdictional boundaries. Roughly 10 % of police-involved litigation is related to jurisdictional issues (Liability Assessment and Awareness International Inc. 2001). Police actions such as vehicular pursuits, surveillance or investigation of crimes require that police officers involved in those incidents understand jurisdictional boundaries.

Canine units can also be a source of liability for police agencies (American Civil Liberties Union of Southern California 1992). Some jurisdictions, such as Los Angeles, rely heavily on canine units to help them find and then capture fleeing suspects. As a result, the number of dog bites has also increased. It has been reported "that an average of nearly one person per day is bitten by Los Angeles Police Department canines" and that "more than 900 men, women, and children have been attacked and mauled by these dogs in a 3 year period (American Civil Liberties Union of Southern California 1992)." There is also evidence that the LAPD uses canine units in predominantly minority-populated neighborhoods of Los Angeles more frequently than other neighborhoods (Los Angeles County Sheriff's Department 2000). With police agencies of all sizes beginning to utilize canines in their daily operations, liability could become more of a concern in the future.

Several professional groups and national publications have recognized the potential benefits of using risk management in police organizations and have provided some guidance for the implementation of risk management. The Public Risk Management Association – PRIMA (a professional trade association of risk managers) published a police liability assessment guide that outlines the police activities that pose the greatest liability risks for police officers. PRIMA's liability guide identifies the use of firearms and other non-lethal weapons, police pursuits, defensive tactics, and hostage situations as some of the activities that pose a high level of risk to police organizations. The PRIMA guide notes that adequate training in high-risk incidents, along with department policies that clearly define procedures in high-risk incidents is critical in the control of police liability (PRIMAFILE 2000).

The Police Chief (a magazine geared toward police practitioners) has also featured articles focused on the importance of risk management as it relates to police liability. In addition to highlighting the importance of implementing risk management programs in police organizations, these articles suggest that risk management can potentially bolster the level of police professionalism (Gallagher 1990b). Several articles on the use of risk management by police agencies have been featured in *Business Insurance* magazine, *National Underwriter* magazine, and *Public Risk* magazine, which are publications geared toward risk managers and insurance professionals. These articles highlight the fact that very few police agencies actually have in-house risk management divisions, and that risk management programs can help police executives identify problematic officer behaviors that can lead to litigation and liability claims (Ceniceros 1998; Heazeltine 1986; Wojcik 1994). It is important to point out that all of these articles appearing in these publications are based on professional experiences of risk managers and other insurance professionals, and not based on empirical research.

The implementation of risk management could also provide non-monetary benefits to police agencies. Police executives need to consider the extent to which integrity violations by police officers impact individual officer careers, other police officers working within the same organization, the police organization as whole, and ultimately, the relationship between the police and the community. After non-monetary costs have been identified by risk managers, changes could be made to include information in education and training programs to teach police officers that their decisions and actions cost their departments far more than settlements or liability claim payouts.

Research on the use of risk management by police agencies in the United States can best be described as very limited. Most of the published literature on this topic has been written by professionals that have experience working with police agencies that use risk management, but is not based on empirical research (Gallagher 1990a). Only within the last decade has any discussion of risk management emerged in the academic literature. In his book, *Police Accountability: The Role of Citizen Oversight*, Samuel Walker reports that "One of the most notable failures of both police departments and other city officials has been their neglect of modern concepts of risk management and in particular their refusal to examine incidents that result in litigation and seek to correct the underlying problems (Walker 2001)." He also provides a thorough discussion on the potential benefits that can result from using risk management, early warning systems, and citizen oversight as police accountability mechanisms.

Geoffrey Alpert and his colleagues devote a chapter to risk management and police liability in their book, *Police Pursuits: What We Know* (Alpert et al. 2000). In this chapter, risk management is discussed in the context of a "plan of action" to prevent costly payouts and injuries associated with vehicular police pursuits. The authors discuss risk management in terms of police agencies taking a proactive or "front-end" approach instead of a reactive approach to the implementation of risk management. The "front-end" approach is the idea that police agencies should implement risk management practices *before* a major liability incident occurs as opposed to after (Alpert et al. 2000). This chapter provides a comprehensive overview of the principles and potential benefits of risk management, as well as suggestions for applying risk management practices to control liability related to vehicular police pursuits.

Samuel Walker and Geoffrey Alpert also wrote an article that discusses the use of early warning systems as risk management in police agencies which was included in the book, *Policing and Misconduct* that was edited by Kimberly Michelle Lersch (Walker and Alpert 2002). An early warning system is a data-based management tool which tracks problematic behavior exhibited by police officers (Walker and Alpert 2002). Early warning systems are used to identify problem officers before they do something that will result in some organizational loss (such as liability claims or lawsuits). Once problem officers are identified by the early warning system they receive some form of intervention, which can include retraining, additional training or verbal counseling by his or her supervisor. One could argue that early warning systems could be a useful tool for agencies that have adopted or plan to adopt risk management to control police liability incidents.

The chapters devoted to risk management in the three previously mentioned books are important and informative; however, these chapters did not include any evidence-based research on the use of risk management by police agencies. The first empirical study on this topic did not appear in print until 2004 (Archbold 2004). Carol Archbold conducted the first study of the use of risk management by police agencies in the United States. This study utilized interview and survey data from a national sample, and in-depth case studies of four police agencies that use risk management to control liability.

In the first stage of Archbold's study, telephone interviews were conducted with a nationwide sample of 354 local and county law enforcement agencies employing 200+ sworn employees. The interviews were used to determine if and how police agencies utilized risk management to manage their liability. A series of telephone interviews were conducted with police administrators, city attorneys, and internal affairs personnel associated with all 354 police agencies in order to get a full picture of how each agency uses risk management. The telephone interviews revealed that only 14 of the 354 (slightly less than 4 %) of the large police agencies use risk management in their police liability management efforts (Archbold 2004; Archbold 2005).

Next, each of the 14 police agencies that reported that they use risk management were faxed a survey. The purpose of the survey instrument was to learn more about how risk management is utilized, how risk management fits into the organizational structure of police agencies, and why these police agencies decided to adopt risk management. All 14 police agencies completed the survey. And finally, four of the 14 research sites were chosen for in-depth case study analyses; the research sites include Charlotte, North Carolina; Los Angeles, California; Las Vegas, Nevada and Portland, Oregon (Archbold 2004 and see also Archbold 2005). These risk management programs were visited to see first-hand how risk management is incorporated into liability management in each agency. Further details from this study can be found in its entirety in the book, *Police accountability, risk management, and legal advising* (Archbold 2004).

Darrell Ross and Madhava Bodapati conducted a study that was published in 2006 which examined the risk exposure and law enforcement liability in the state of Michigan from 1985 to 1999 (Ross and Bodapati 2006). This study used records that were collected and maintained by the Michigan Municipality Risk Management Authority (MMRMA), a group that provides insurance and risk management services to law enforcement agencies in the state of Michigan. The purpose of the study was to determine the common types of litigated cases brought against law enforcement agencies in Michigan and to also examine their trends over time. Analysis of the data revealed that incidents involving automobiles without injuries was the most frequent type of claim, followed by use of excessive force, damage or destruction of property, vehicular pursuits without injuries, and false arrest/imprisonment (Ross and Bodapati 2006). The most costly type of claims involved wrongful death/fatalities followed by attempted suicide while in custody, medical care while in custody, suicide while in custody, and automobile accidents which involved injuries (Ross and Bodapati 2006). Overall, the study found that claims filed against Michigan police agencies along with costs associated with those claims are relatively low; annually, there was an average of 752 claims for 151 law enforcement agencies. The authors concluded that "risk management services and agency practices appear to be making a difference overall (Ross and Bodapati 2006)." This study is important as it is one of the first published studies to assess the impact that risk management has on police liability.

If there are both financial and non-monetary benefits that can potentially result from using risk management, why isn't it more common in American police agencies? Risk management professionals that have helped implement risk management programs in American police organizations have identified what they believe are potential barriers to implementation. First, policing has traditionally been focused on crime fighting without any worry about financial costs (Gallagher 1990b). With an increase in litigation involving police officers in recent years, more police executives are becoming as concerned with fiscal responsibilities as they are with their crime control efforts (Ceniceros 1998). Police agencies need resources to be able to provide quality services to the public, which means that they need to utilize their resources in the most efficient way possible.

Another potential barrier is the cost associated with implementing risk management programs. Large cities are able to justify their need for hiring professional risk managers for reducing costs associated with police liability. Medium-sized and smaller cities have a harder time justifying the need for risk management as their exposure to liability is not as great as police agencies serving communities with much larger populations (Los Angeles County 2003). Some less expensive alternatives to hiring a full-time risk manager or implementing a risk management division includes the expansion of the duties for individuals that work directly with police-related liability (such as city managers or city attorneys), employing only a few risk managers to service a large area or region that encompasses several small or medium cities with less populations or more rural areas, or an "as needed" risk management committee consisting of city attorneys, police department representatives and insurance agents or risk assessors (O’Brien and Wilcox 1985).

Confusion about the best way to measure the impact of risk management within police organizations could be another reason that so few police agencies use this management tool. The monetary benefits of risk management are easy to quantify and analyze in regard to impact; however, some would argue that it is impossible to measure the liability claims and lawsuits that have been prevented as a result of risk management efforts. In addition, the impact on police professionalism and the quality of police services provided to the community are also not easy to quantify (Young 2000). Because resources are often limited in police organizations, it could be difficult to justify the need for an in-house risk manager or division if there are no outcome measures for the non-monetary benefits.

Another barrier to the implementation of risk management within police agencies is the lack of research on the use of risk management by police agencies in the United States. There are several websites on the Internet that provide information on risk management training for police agencies. There are also training seminars offered by individual risk management experts that travel across the country training police personnel (such as G. Patrick Gallagher and Gordon Graham). The problem is that there has been very little evidence-based research on the use of risk management in police agencies.

After reviewing existing literature on risk management in policing, it is clear that additional research needs to be conducted in the future. Despite the increase in costs associated with police liability in the last four decades, very little is known about how risk management might help manage those costs. Other than the two studies conducted by Archbold, and Ross and Bodapadi, there have not been any in-depth studies of the impact that risk management has on costs related to police liability. Future research should include studies similar to the Ross and Bodapadi study in other regions and municipalities across the United States. The use of risk management by medium and small police agencies should also be examined in the future. It should not be assumed that medium and small police agencies do not have organizational loss related to liability incidents. The use of early warning systems as part of risk management efforts should also be examined as there is a growing interest in the use of this particular oversight mechanism. The extent to which risk management enhances police professionalism and police officer accountability should also be studied in the future. In conclusion, we still have a lot to learn about the use and impact of risk management in American police agencies.

Related Entries

► Police and the Excessive Use of Force

Recommended Reading and References

Alpert G, Fridell L (1992) Police vehicles and firearms: instruments of deadly force. Waveland Press, Prospect Heights

Alpert GP, Kenney DJ, Dunham RG, Smith WC (2000) Police pursuits: what we know. Police Executive Research Forum, Washington, DC

American Civil Liberties Union of Southern California (1992) Analysis and recommendations: Los Angeles police department K-9 program. Los Angeles California Police Commission, Los Angeles

Archbold CA (2004) Police accountability, risk management, and legal advising. LFB Scholarly Publishing, New York

Archbold CA (2005) Managing the bottom line: risk management in policing. Policing: Int J Polic Strateg Manage 28:38–50

Ashley S, Pearson R (1993) Fundamentals of risk management. Liability Assessment & Awareness International, Inc. Available at www.laaw.com. Accessed on 12 Feb 2013

Ceniceros R (1998) Policing their own risks: formal risk management growing in law enforcement. Bus Insur 32:10–13

del Carmen R (1991) Civil liabilities in American policing: a text for law enforcement personnel. Prentice-Hall, Englewood Cliffs

Gallagher GP (1990a) The six-layered liability protection system for police. Police Chief 58:40–44

Gallagher GP (1990b) Risk management for police administrators. Police Chief 58:18–29

Heazeltine H (1986) Case study: a risk audit of law enforcement. Risk Manage 10:60–63

Kappeler VE (2006) Critical issues in police civil liability, 4th edn. Waveland Press, Prospect Heights

Liability Assessment and Awareness International Inc (2001) www.laaw.com

Los Angeles County (2003) Available at www.co.la.ca.us/bobb.html. Accessed on 12 Feb 2013

Los Angeles County Sheriff's Department (2000) 12th Semiannual report by special counsel Merrick J. Bobb and Staff. Los Angeles, California

Lui J (2011) City of New York office of the comptroller claims report – fiscal years 2009–2010. See full report at http://comptroller.nyc.gov/bureaus/bla/pdf/2011_Claims_Report.pdf

Midwest Human Rights (2011) See http://www.midwesthumanrights.org/chicago-police-misconduct-lawsuits-adding

O'Brien LE, Wilcox DE (1985) Risk management organization and administration. In: Wasserman N, Phelus DG (eds) Risk management today: a how-to guide for local government. Public Risk and Insurance Management Association and International City Management Association, Washington, DC

PRIMAFILE Online (2000) Police liability assessment guide PRIMAFILE Online, on file with author

Pate A, Fridell L (1995) Toward the uniform reporting of police use of force: results of a national survey. Crim Justice Rev 20:123–145

Ross D, Bodapati M (2006) A risk management analysis of the claims, litigation, and losses of Michigan law enforcement agencies: 1985–1999. Policing: Int J Police Strateg Manage 29:38–57

Vincent J (1996) Managing risk in public services: a review of the international literature. Int J Public Sect Manag 9:57

Walker S (2001) Police accountability: the role of citizen oversight. Wadsworth Publishing, Belmont

Walker S, Alpert G (2002) Early warning systems as risk management for police policing and misconduct. In: Lersch KM (ed) Prentice Hall Publishing, pp 219–230

Wennerholm RW (1985) Officer survival recommendations: new civil liability concerns. In: Wasserman N, Phelus DG (eds) Risk management today: a how-to guide for local government. Public Risk and Insurance Management Association and International City Management Association, Washington, DC

Wojcik J (1994) Liability lawsuits rise again for LAPD. Bus Insur 28:20–22

Wong K-SR, Rakestraw KM (1991) The A, B, C's of risk management. International City/County Management Association, Washington, DC

Young PC (2000) Risk management: a comprehensive approach. International City/County Management Association, Washington, DC

Risk Perception

▶ Fear of Crime and the Psychology of Risk

Risk Terrain Modeling (RTM)

▶ Forecasting of Shootings Using Risk Terrain Modeling

Risky Facilities

▶ Prediction and Crime Clusters

Risky Professions

▶ Prediction and Crime Clusters

Risky Routes

▶ Prediction and Crime Clusters

RNR

▶ Actualizing Risk-Need-Responsivity

RNR Model

▶ Actualizing Risk-Need-Responsivity

RNR Supervision

▶ Behavioral Management in Probation

Robbery

Philip J. Cook
Sanford School of Public Policy,
Duke University, Durham, NC, USA

Overview

Robbery is a violent crime but also a property crime since the motivation is theft. It is a fear-inspiring crime since it typically involves an unprovoked attack by strangers. In the United States, there has been a large decline in robbery rates since the peak in the early 1990s, especially in the largest cities. Some of the pronounced patterns in robbery are just what we would expect if robbers were rational actors seeking a quick score by gaining physical power or control over the victim. Most robberies are committed with lethal weapons and involve two or more perpetrators, usually young males. Gun-wielding robbers are more likely to choose stronger victims and to successfully complete the theft without much resistance. Nonetheless, gun robberies are much more lethal than robberies with knives or clubs. A variety of situational and police-manpower approaches to robbery reduction have been successful.

Introduction

Robbery is a crime of violence motivated by theft. The spectrum of robbery ranges from schoolyard shakedowns for lunch money to bank holdups with losses reaching into the millions of dollars. Property losses from robbery are, however, usually small or nil, and it is the violence that makes robbery such a serious crime.

Robbery is particularly fear inspiring because it usually involves an unprovoked attack by strangers. In that regard it differs from assault, which typically involves an altercation between acquaintances. The public concern about "crime in the streets" is to a large extent a concern about robbery, and that concern may have profound effects on the life of the city, through its influence on choices about where to live, work, shop, and go out to dinner (Cook and Ludwig 2000). James Q. Wilson and Barbara Boland note that "[i]t is mostly fear of robbery that induces many citizens to stay home at night and to avoid the streets, thereby diminishing the sense of community and increasing the freedom with which crimes may be committed on the streets" (Wilson and Boland 1976, p. 183).

This entry provides a description of trends and patterns in robbery using police statistics and statistics from crime surveys. The focus is on data for the United States, which has the best established of the national crime surveys. The survey data support a fine-grained analysis of the age, sex, race, and number of robbers and victims involved in an incident, as well as the type of weapon and the outcomes of the confrontation with respect to theft and injury. It is helpful to understand these patterns through the lens of the robbers' instrumental concerns – identifying lucrative "targets," preempting resistance, and avoiding arrest – although some violence is simply gratuitous.

Robbery prevention efforts include private protection efforts for prime commercial targets (guards, closed-circuit cameras, reduced cash on hand) and directed police patrol against street robbery. Probably more important has been the evolution toward a cashless economy, reducing the robbers' chances for a quick payoff. Reducing the overall robbery rate is not the only goal – a second public goal is harm reduction. For example, measures to reduce gun use in robbery have the potential to reduce the rate of serious injury and murder.

This entry consists of four main sections and a conclusion. The first section discusses robbery rate trends in the United States since the 1960s – increasing through 1981, down since 1991, and fluctuating in between – and relative to rates elsewhere. The second section, based on victimization dates combined for the years 2000–2005 to provide sufficiently large numbers to support generalizations, examines the demographics of robber perpetration and victimization and robbery outcomes (property loss, injury, and death). The third discusses robbers' strategic choices about whom to rob and how, with what weapon, and what differences those choices make. The fourth discusses interventions aimed at preventing robberies.

A number of conclusions are offered:

- Although data comparability issues preclude confident assertions, American robbery rates as shown in official data and victims' reports appear comparable to those of other developed countries.
- Official and victim data concur in showing a very substantial decline in robbery rates since the early 1990s.
- Most robbers are male (more than 90 % in 2000–2005), young (74 % under 30), and African American (54 %).
- Most robberies do not yield high profits for robbers (only 29 % yielded cash) but nearly half of victims are physically attacked.
- Robberies with guns are less likely to result in victim injuries than are other robberies but are much more likely to result in death.
- A variety of situational and police-manpower approaches to robbery reduction have been successful.

Data and Trends

The volume, trend, and patterns of crime can be measured by the use of two sorts of data: victimization surveys and records of crimes known to the authorities. A careful analysis of data for eight high-income nations offers insight into the difficulty in developing reliable measures (Farrington et al. 2004). As shown in Table 1, the recorded rates circa 1999 ranged from 0.5 per 1,000 (for Switzerland) up to 1.5 per 1,000 (for both the United States and England and Wales). But the survey estimates were far higher and suggest a different lineup; for example, while the survey rate for the United States was 2.5 times the recorded rate during the 1980s and 1990s, for the Netherlands the survey rate was nearly 15 times as high as the recorded rate and was a multiple of the US survey rate. Furthermore, in several of these nations, the trends in the survey measure look quite different than the trends in the recorded measure.

Of the various national crime surveys, the United States National Crime Victimization Survey is best established, generating annual estimates since 1973, and also has an important technical advantage over the others. The NCVS interviews the same households every 6 months over seven cycles, encouraging the respondents to use the previous interview as a mental baseline to define the relevant period. That device helps limit an important source of error in survey-respondent reports, namely, the tendency to report important victimization experiences that occurred outside the specified interval (6 months or a year) – a recall problem known as "telescoping." Other nations do not use this costly procedure and hence may have more telescoping and a positive bias to survey results.

The NCVS helps uncover the "dark figure of crime" that goes unreported to the police. In 2010, for example, the NCVS estimated 568,510 noncommercial robberies, compared to 367,832 robberies of *all* types tabulated through the Uniform Crime Reporting (UCR) system which compiles crimes known to the police. Only about 40 % of noncommercial robberies are reported to the police. The NCVS data are also useful in providing the statistical basis for analyzing demographic patterns of violence – both of the victims and of the perpetrators (based on respondents' reports of their impression of the age, race, sex, and number of assailants). They also provide an empirical basis for characterizing process and outcomes of criminal incidents.

Since the interview subjects are households, commercial robberies and other crimes against businesses are not included. According to 2010 UCR data, about 40 % of robberies reported to the police occur in businesses. The true percentage may be lower if businesses are more likely to report victimization than are individuals.

For the period since 1973, the NCVS data indicate that victimization rates for personal robbery were cyclical with peaks in 1974, 1981, and 1994; since 1994 there has been a sustained drop from 6.3 per 1,000 to just 2.2 per 1,000 in 2010. UCR robbery rates have roughly the same peaks and

Robbery, Table 1 Robbery rates, record, and survey data

	Time period	Growth, survey data (percent)	Growth record data (percent)	Recent recorded rate per 1,000	Ratio: survey to recorded rates
England and Wales	1981–1999	95	266	1.5	6.1
Scotland	1981–1999	61	22	1.0	3.6
Australia	1983–2000	0	300	1.2	7.9
Canada	1981–1998	2	−9	.9	9.7
United States	1981–1999	−51	−42	1.5	2.5
Netherlands	1980–1999	25	270	1.1	14.8
Sweden	1980–1998	NA	85	.7	9.0
Switzerland	1985–1999	41	48	.5	14.2

Source: Cook and Khmilevska (2005), pp. 333 and 335
NA = not available

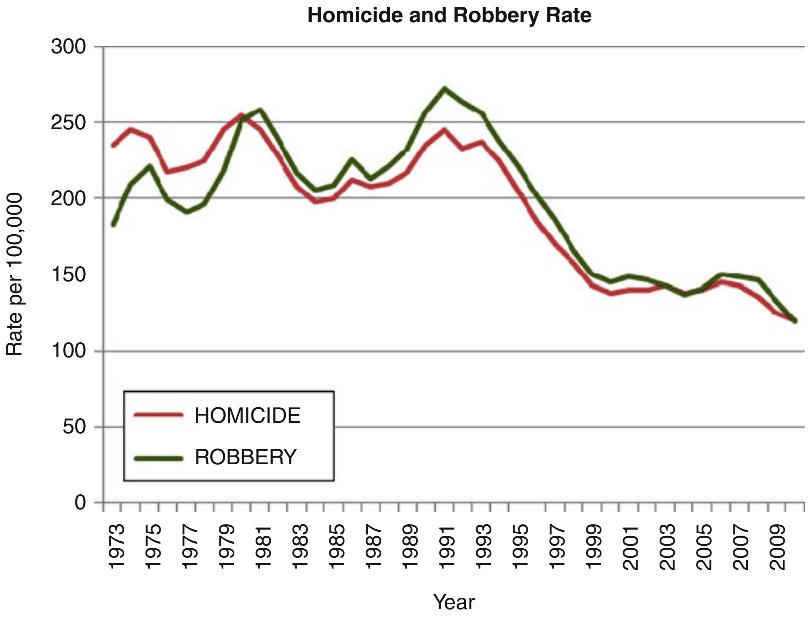

Robbery, Fig. 1 Homicide and robbery rate

valleys as the NCVS rates, and they too show a strong downward trend during the 1990s (Blumstein et al. 1991). As documented by Alfred Blumstein (2000) and shown in Fig. 1, UCR robbery rates have moved up and down for the last several decades in synch with murder rates which are the most reliably measured of all crime rates. (Note that the homicide rate has been scaled up by a factor of 25 in this diagram so that it is on a similar scale as the robbery rate.) It appears, then, that the UCR robbery trends are reliable.

The extraordinary reduction in robbery and other violent crime during the 1990s remains something of a mystery – no expert *predicted* this decline – but some credit surely goes to increases in the size of police forces and stringent sentencing policies that greatly increased the prison population (Blumstein and Wallman 2000; Cook and Laub 2002; Levitt 2004; Blumstein and Wallman 2006; Zimring 2007; Zimring 2012). The crack epidemic that began in the mid-1980s contributed to the surge and later decline of robbery: by one account, crack users turned to robbery rather than other forms of theft (such as burglary) because it was the quickest method of obtaining cash (Baumer et al. 1998). Other plausible (but

Robbery, Fig. 2 Average robbery rate by city size

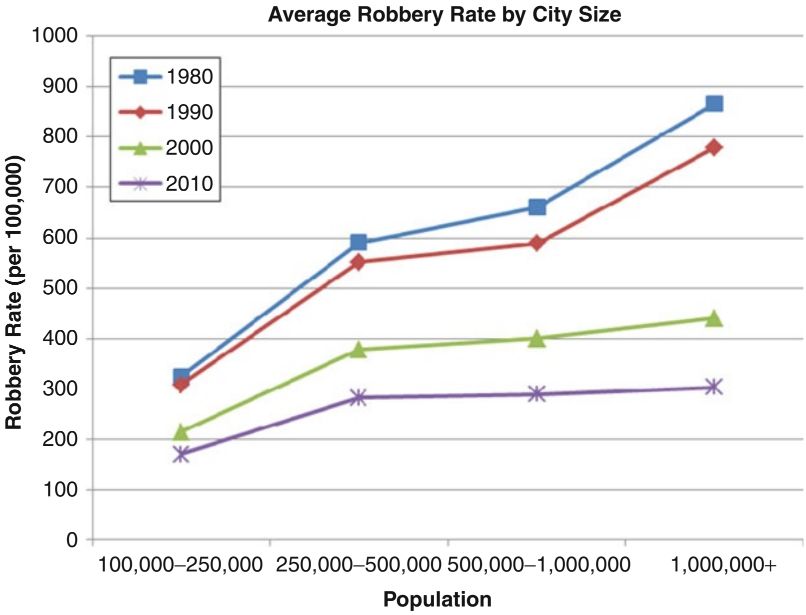

contested) explanations for the drop in robbery and other violence include the sustained economic expansion of that period, innovations in policing, the delayed effects of the legalization of abortion, and the removal of lead from gasoline.

The 1990s crime drop affected all geographic areas. The greatest percentage improvements occurred within metropolitan statistical areas (MSAs) and especially among large cities with populations over 250,000. All the 25 largest cities experienced noteworthy declines in homicide rates from their peak year (mostly in the early 1990s) to 2001, declines that ranged as high as 73 % for New York and San Diego (Levitt 2004). New York City alone had 17.5 % of all US robberies in 1990, but just 8.4 % in 2005. Remarkably, robbery, always the quintessential urban crime, was no longer correlated with city size by 2000. Figure 2 shows how the urban pattern of robbery rates changed over four decades.

Patterns of Commission, Victimization, and Violence

In what follows, robbery patterns for the United States as a whole are documented by the use of data from the NCVS on noncommercial robberies. The years 2000–2005 are combined to provide an adequate sample of incidents.

Demographics

Robberies are often committed by groups of offenders. For the period 2000–2005, 18 % of incidents involved two robbers, and 20 % involved three or more. (There has been a pronounced trend toward more solo robberies; just 50 % of robberies in 1980, by 2005 they were up to 62 %.) Such incidents are classified here according to the victim/respondent's report of the *oldest* robber's age or race. Groups of robbers almost always consist of demographically similar individuals, with the partial exception of sex: 11 % of robbery incidents involve at least one female offender, but she teamed up with a man in over one-third of those cases.

In robbery, unlike other types of violent crime, the victim and perpetrator are usually strangers (65 % of incidents) and are drawn from different demographic distributions. The majority of robbers in the United States are male, young (36 % under 21, 74 % under 30), and African American (54 %), while the victims are more representative of the US population. There is some tendency for likes to rob likes with respect to age, race, and sex (Cook 1976), but with notable "crossover" – a majority of victims of black robbers are white, and most female victims (37 % of the total) are robbed by males.

Robberies can be classified by location, and for this purpose it is best to get the full spectrum of possibilities, including commercial robberies, which requires UCR data. In 2011, 44 % of all robberies known to the police were on the street and 17 % at a residence. The remaining 39 % were at convenience stores, gas stations, banks, and other commercial locations. The nationwide drop in the robbery rate during the 1990s was concentrated on street robbery, with relative increases in residential and commercial robbery.

Outcomes

When the celebrated bank robber Willie Sutton was asked why he robbed banks, he allegedly replied "because that's where the money is." The NCVS provides data only on noncommercial robberies, committed by robbers who generally did not follow Willie Sutton's putative guidance. Combining all incident reports for 2000–2005 with at least one adult male robber, it turns out that the robbery was unsuccessful in fully one-third of the cases, either because the robber was scared off or the victim had nothing to offer. Cash was stolen in only 29 % of all cases. In the minority of cases that were successful, the median cash stolen was $90. In these and other successful robberies, the loot may include jewelry, clothing, credit cards, household items, and sometimes (in a "carjacking") a motor vehicle. But only a quarter of successful robberies resulted in theft of items worth $250 or more.

The other proximate outcome is injury to the victim. About half of victims report being physically attacked by the robber (rather than just threatened), and one-fifth require medical treatment. Some victims are seriously wounded or killed. In 2011, the FBI classified 734 murders as robbery related (6 % of all murders), which suggests as an order of magnitude that 2.5 in 1,000 robberies resulted in death that year. The true robbery-murder count is probably higher – one detailed study of Chicago robberies found the police department was conservative in using that classification (Zimring and Zuehl 1986).

Since the greatest threat from robbery is the victim's serious injury or murder, it is of considerable interest to know what distinguishes those incidents from the great majority in which the victim comes through unscathed (at least physically). One study compared robbery murders (as documented by the FBI's Supplementary Homicide Reports) to robberies, finding similar distributions with respect to prior relationship (primarily strangers) and demographics (Cook 1987, p. 366). But there was one important difference between robbery and robbery murder – the types of weapons used. About two-thirds of robbery murders are committed with guns, while less than *one*-third of robberies involve guns. Gun robberies are three times more likely to result in the death of the victim than knife robberies, and knife robberies three times more likely than robberies with other weapons (Cook 1987). While it has been much debated whether the type of weapon has a separate causal influence on injury (an "instrumentality effect"), as opposed to merely reflecting the offender's intent, a variety of evidence favors instrumentality (Zimring 1972; Kleck and McElrath 1991; Wells and Horney 2002). A regression analysis of changes in robbery-murder rates in 43 cities found a close relationship between the robbery rate and the robbery-murder rate, as if the latter were simply a probabilistic byproduct of the former. Every additional 1,000 gun robberies added 4 robbery murders to the city's total, while 1,000 nongun robberies added just one robbery murder (Cook 1987, p. 373).

Strategic Choices

Observed outcomes of injury and theft are the result of a series of choices made by the robber and his intended victim. The robber (or group of robbers) chooses the victim, the weapon, and the technique, while the victim must decide whether to cooperate or resist. Based on interviews with robbers and other evidence, it is safe to say that robbers' choices reflect a desire to preempt resistance, secure as much loot as possible, and then make a safe getaway, although actual behavior in these fraught encounters is not necessarily rational (Conklin 1972; Wright and Decker 1997; Jacobs 2000). One useful concept is the "power" of the robber to gain the victim's compliance (Cook 1976). Power depends on the weapon used by the robber; a gun, by creating a lethal threat at a distance, is most powerful, enough for a single robber to control several victims at once. Accomplices also enhance power, as does skill and experience. A solo offender who lacks a gun or accomplices is more likely to select a woman or older person as victim – a majority of gun robberies, on the other hand, are directed against relatively robust commercial targets (Cook 1980).

A multivariate analysis of NCVS data for 2000–2005 (all robberies committed by an adult male) confirms that the chances of successful theft are greatly enhanced when a gun is used or the robber has accomplices. A gun also enhances the amount of loot in those robberies that are successful, in part because gun robbers are able to choose more lucrative targets (couples, males age 25–54), who might otherwise be in a position to defend themselves. This analysis of the payoff to robbery technique can be used to compute the value of using a gun as opposed to, say, a knife. Other things equal, the likelihood of successful theft increases by 12.5 % points, and the average value of loot in successful robberies almost doubles, when compared to using a knife (the second-best weapon alternative from the robber's perspective).

Robbers use different methods for gaining compliance depending in part on the circumstances. The standard technique in a gun and knife robberies is to threaten without actually attacking (78 % and 64 % of gun and knife noncommercial cases, respectively), which is usually sufficient to gain compliance. The standard technique in unarmed robberies or robberies utilizing sticks and clubs is to attack physically (over 60 % of cases) and then attempt to take the valuables by force. The chances of injury follow the same pattern, ranging from just 11 % for gun robberies to 36 % for robberies with clubs. Of course when a gun is present and fired, the resulting injury is far more likely to be fatal, and the chances that the victim will be killed are highest for gun robberies (as documented above).

Detailed case studies have provided evidence that the violence in robbery is not necessarily instrumental. The offenders may exploit the situation by raping or otherwise tormenting the victim. In some cases of robbery murder, it would appear that the killing was purely gratuitous or even "recreational," stemming from the heat of the moment and the interplay between accomplices (Cook 1980; Zimring and Zuehl 1986). Robbery defendants who injured their victims are more likely than others to be arrested for assault following release from prison than are robbery defendants who did not use violence (Cook and Nagin 1979; Cook 1987), suggesting a taste for violence.

The strategic choice framework predicts that drug dealers and other actors in the underground economy would be especially attractive targets, both because the loot is likely to include large amounts of cash and drugs and because victims are not likely to inform the police – although they may attempt to retaliate (Jacobs 2000; Topalli et al. 2002). Crimes of this sort are unlikely to become known unless someone is killed, and indeed, it is common to have a drug connection with robbery murder (Zimring and Zuehl 1986). As observed by one group of criminologists studying drug-dealer robbery, "One of criminology's dirty little secrets is that much serious crime ... takes place beyond the reach of the criminal law because it is perpetrated against individuals who themselves are involved in lawbreaking" (Topalli et al. 2002).

Victims as well as offenders make decisions that influence the outcome. Most important is whether to comply with the robber's demands and, if not, just how to resist. In the NCVS sample of noncommercial robberies, about one-third of victims attempted to forcefully resist. Resistance is associated with a reduced chance of successful theft, but an increased chance of injury. Victims of gun robbery are less likely to resist than those confronted with other weapons. A practical question often arises of whether it is prudent to resist a robber, but unfortunately it is not possible to extract a clear answer from survey data.

Most victims act on the commonsense rule that it is foolish to resist a robber with a gun, and commercial employers often instruct their clerks that if there is a robbery, they should not attempt active resistance.

Interventions

The threat of robbery victimization has far-reaching effects on urban life. The effort to avoid this threat may influence choices about where to live, whether and where to go out at night, the use of public transportation, carrying a weapon, and so forth. A reduction in the robbery threat in a city enhances the residents' standard of living and may well contribute to the city's growth and prosperity (Cook and Ludwig 2000). There are a variety of specific measures, both private and public, that hold some promise in this regard, either in reducing the overall robbery rate or in reducing the related harms.

The private efforts to reduce robbery are to a large extent applications of "situational crime prevention" (Clarke 1983, 1995), which attempts to reduce criminal victimization of specific targets by making them less attractive to criminals. The goal is to make it more difficult to complete the crime successfully, reduce the payoff if successful, and increase the risks of arrest and conviction. For especially attractive robbery targets, most notably banks, an array of measures are typically put in place, including guards, surveillance cameras, alarms, and dye packs mixed in with the cash. (Private security against robbery and other crimes is a rapidly growing industry, and there are now as many private guards as there are sworn police officers (Cook and MacDonald 2011).) Studies suggest that convenience stores, which are frequent targets of robbery, can reduce the risk by installing bright exterior lighting and keeping at least two clerks on duty at night (Loomis et al. 2002). A variety of common robbery targets, including all-night businesses and public buses, have attempted to discourage robbery by limiting the amount of accessible cash – and advertising that fact.

One of the most controversial private responses to the threat of robbery is keeping a gun handy. The immediate goal for those who keep or carry a gun for self-protection is to enhance the capacity to resist a robbery, other assault, or criminal intrusion to home or vehicle. But the laws governing gun possession and carrying have taken on great significance in part because of the claim that widespread gun carrying would create a general deterrent to robbery and other assaults that would benefit all. With that in mind, and under pressure from gun-rights advocates, most states in the USA have loosened their restrictions on carrying concealed firearms (Vernick and Hepburn 2003), now mandating that all who meet minimum requirements have the right to a concealed-carry permit. These "shall issue" laws have been extensively evaluated with conflicting results (Lott and Mustard 1997; Black and Nagin 1998; Ludwig 1998; Lott 2000). A careful review of the evidence suggests the true effects on crime are almost certainly small and are statistically undetectable given other more powerful forces that influence crime trends (Donohue 2003).

There is less controversy about the public's stake in robbers' choice of weaponry; a gun robbery has a greatly enhanced risk of turning deadly when compared with robberies with knives and blunt objects. Gun involvement by young men is closely linked to the prevalence of gun ownership in the community, but there are a variety of regulatory and policing methods that show promise for discouraging gun carrying and criminal use (Cook and Ludwig 2006). One organizing principle for such interventions is to make guns a liability to urban youths and criminals by patrolling against illicit carrying, offering rewards for information on illegal guns, and giving priority to prosecution and sentencing of gun robberies.

More generally, expanding and focusing police resources directed at street crime is a costly but effective method of combating robbery. In England and Wales, the Street Crime Initiative provided funding for anti-robbery policing in 10 of the 43 police-force areas, with large, statistically discernible effects on robbery rates (Machin and Marie 2005). Other studies have provided additional support for the conclusion that additional police suppress crime rates (Levitt 2002; McCrary 2002; Levitt and Miles 2007). The effect sizes are large enough to make a strong case for expanded police funding (Donohue and Ludwig 2007).

Conclusions

When people worry about "crime in the streets," their first concern is robbery. While the financial losses from robbery are of little consequence overall, the fear of robbery and the actual injuries it causes have a profound effect on the quality of life in certain neighborhoods.

A quick payoff is the allure of robbery for many drug addicts and other criminals. The surest quick payoff is cash, but that is becoming increasingly rare as a means of exchange in the first world nations – except for transactions where preserving anonymity is important. For that reason robbery may focus increasingly on underground markets and as a result become less visible but perhaps more violent. Policies to reduce the scope of the underground market will thus be helpful in the effort to curtail robbery.

Related Entries

- ▶ Economic Theory of Criminal Behavior
- ▶ Offender Decision Making and Behavioral Economics
- ▶ Rational Choice Theory
- ▶ Residential Burglary
- ▶ Theories of Co-offending
- ▶ Understanding Cross-National Variation

Recommended Reading and References

Baumer E, Lauritsen JL, Rosenfeld R, Wright R (1998) The influence of crack cocaine on robbery, burglary, and homicide rates: a cross-city, longitudinal analysis. J Res Crime Delinquency 35(3):316–340

Black D, Nagin D (1998) Do "right to carry" laws reduce violent crime? J Legal Stud 27(1):209–219

Blumstein A (2000) Disaggregating the violence trends. In: Blumstein A, Wallman J (eds) The crime drop in America. Cambridge University Press, New York

Blumstein A, Cohen J, Rosenfeld R (1991) Trend and deviation in crime rates: a comparison of UCR and NCS data for burglary and robbery. Criminology 29(2):237–263

Blumstein A, Wallman J (2000) The crime drop in America. Cambridge University Press, New York

Blumstein A, Wallman J (2006) The crime drop and beyond. An Rev Law Soc Sci 2:125–146

Clarke RV (1983) Situational crime prevention: its theoretical basis and practical scope. In: Tonry M, Morris N (eds) Crime and justice: an annual review of research, vol 4. University of Chicago Press, Chicago

Clarke RV (1995) Situational crime prevention. In: Tonry M, David P (eds) Building a safer society: strategic approaches to crime prevention. University of Chicago Press, Farrington/Chicago, pp 91–150

Conklin JE (1972) Robbery and the criminal justice system. Lippincott, Philadelphia

Cook PJ (1976) A strategic choice analysis of robbery. In: Skogan W (ed) Sample surveys of the victims of crimes. Ballinger, Cambridge, MA

Cook PJ (1980) Reducing injury and death rates in robbery. Policy Anal Winter 21–45

Cook PJ (1987) Robbery violence. J Crim Law Criminol 70(2):357–376

Cook PJ, Khmilevska N (2005) Cross-national patterns in crime rates. In: Tonry M, Farrington DP (eds) Crime and punishment in western countries, 1980–1999. University of Chicago Press, Chicago

Cook PJ, Laub JH (2002) After the epidemic: recent trends in youth violence in the United States. In: Tonry M (ed) Crime and justice: a review of research, vol 29. University of Chicago Press, Chicago

Cook PJ, Ludwig J (2000) Gun violence: the real costs. Oxford University Press, New York

Cook PJ, Ludwig J (2006) The social costs of Gun ownership. J Public Econ 90(1–2):379–391

Cook PJ, MacDonald J (2011) The role of private action in controlling crime. In: Cook PJ, Ludwig J, McCrary J (eds) Controlling crime: strategies and tradeoffs. University of Chicago Press, Chicago, pp 331–366

Cook PJ, Nagin D (1979) Does the weapon matter? An evaluation of a weapons-emphasis policy in the prosecution of violent offenders. INSLAW, Washington, DC

Donohue JJ (2003) The impact of concealed-carry laws. In: Ludwig J, Cook PJ (eds) Evaluating gun policy. Brookings Institution, Washington, DC

Donohue JJ, Ludwig J (2007) More COPS, policy brief 158. The Brookings Institution, Washington, DC

Farrington DP, Langan PA, Tonry M (eds) (2004) Cross-national studies in crime and justice. U.S.

Department of Justice, Office of Justice Programs, Washington, DC

Jacobs BA (2000) Robbing drug dealers: violence beyond the Law. Aldine de Gruyter, Hawthorne

Kleck G, McElrath K (1991) The effects of weaponry on human violence. Soc Forces 69(1991):669–692

Levitt SD (2002) Using electoral cycles in police hiring to estimate the effects of police on crime: a reply. Am Econ Rev 92(4):1244–1250

Levitt SD (2004) Understanding why crime fell in the 1990s: four factors that explain the decline and six that do not. J Econ Perspect 18(1):163–190

Levitt SD, Miles TJ (2007) The empirical study of criminal punishment. In: Polinsky AM, Shavell S (eds) The handbook of law and economics, vol 1. North Holland, New York

Loomis D et al (2002) Effectiveness of safety measures recommended for prevention of workplace homicide. J Am Med Assoc 287:1011–1017

Lott JR (2000) More guns, less crime, 2nd edn. University of Chicago Press, Chicago

Lott JR, Mustard DB (1997) Crime, deterrence and right-to-carry concealed handguns. J Legal Stud 16(1):1–68

Ludwig J (1998) Concealed-gun-carrying laws and violent crime: evidence from state panel data. Int Rev Law Econ 18:239–254

Machin SJ, Marie O (2005) Crime and police resources: the street crime initiative. IZA discussion paper 1853. Institute for the Study of Labor, Bonn

McCrary J (2002) Using electoral cycles in police hiring to estimate the effect of police on crime: comment. Am Econ Rev 92(4):1236–1243

Topalli V, Wright R, Fornango R (2002) Drug dealers, robbery and retaliation. Br J Criminol 42:337–351

Vernick JS, Hepburn LM (2003) State and federal gun laws: trends for 1970–99. In: Ludwig J, Cook PJ (eds) Evaluating gun policy. Brookings Institution, Washington, DC

Wells W, Horney J (2002) Weapon effects and individual intent to do harm: influences on the escalation of violence. Criminology 40(2): 265–296

Wilson JQ, Boland B (1976) Crime. In: Gorham W, Glazer N (eds) The urban predicament. The Urban Institute, Washington, DC

Wright RT, Decker SH (1997) Armed robbers in action. Northeastern University Press, Boston

Zimring FE (1972) The medium is the message: firearm calibre as a determinant of death from assault. J Legal Stud 1:97–124

Zimring FE (2007) The great American crime decline. Oxford University Press, New York

Zimring FE (2012) The city that became safe. Oxford University Press, New York

Zimring FE, Zuehl J (1986) Victim injury and death in urban robbery: a Chicago study. J Legal Stud 15(1):1–40

Role and Function of the Police

Peter K. Manning
College of Criminal Justice, Northeastern
University, Boston, MA, USA

Overview

There is no accepted systematic theoretical paradigm within which policing is viewed. The role and function of the police are typically assumed, and a measurable facet such as crime control is defined as the scholarly interest. Those viewed as essential policing functions and how they should be performed are products of the theoretical context within which the police are viewed, their perceived political role, and the posited character of the police organization. As a result, there are alternative versions of policing and what it is good for. These might be called “practical theories of police” in that they emphasize and valorize selectively some aspects of policing while remaining wedded to the basic deployment strategies of random patrol, crime investigation and response to citizens’ calls while altering the rhetorical strategy, and some policing tactics. They are the products of fads and trends in the sense that they reflect efforts of the police and some scholars to redefine the product without modifying the structure that produces it. For this reason, it is best to treat policing in a somewhat abstract fashion, detailing its roles and functions, rather than providing overviews of research out of context. Research shows that functionally and structurally police in North America and the UK have changed very little (Maguire 2003; Weisburd and Braga 2006), and field studies suggest there is little evidence that police practices have changed (Moskos 2008; Loftus 2010). The study of the role and function of police has changed.

In this context, a review of the role and function of the police is a slightly contentious endeavor. This entry will have the following themes. The first is a necessary broadening of and consideration of systems of law within which police function and types of policing. The second theme concerns the role of police in the drama of governance, the police mandate, and their role as a functional and dramaturgical actor. The third theme is the police organization seen as a collection of roles, segments, and cultures. These are in effect various ways of seeing the role and function of the police. This outline is a kind of funnel that contracts from the beginning characterization of socio-legal types of policing systems to police systems and their role in governance in Anglo-American societies to the organization and its internal configuration. The entry concludes with some reflections.

The Rise of Police Studies

Let us consider the development of police studies in periods characterized by quite different intellectual, policy, and practical focal concerns. The foundational work on the police as an organization in the context of government was done by lawyer-administrators such as Arthur Woods, Harry Emerson Fosdick, and Bruce Smith and later by the Commissioner of the Police of the City of Chicago, O. W. Wilson (Wilson and McLaren 1963). The first modern phase can be traced forward from the work of Westley (1951, 1970) focused on the occupational culture and its support for violence; Banton's (1964) comparative ethnographic study of the police role in urban and rural context; Skolnick's work on policing on the streets as a form of justice without trial (1966); Reiss and Bordua's (1967) empirically based analysis of the police as organization, and their later paper (Bordua and Reiss 1967); Wilson's concise and masterful depiction of the dilemmas of police administration and the socio-political context of the role (1968); Bittner's formulation of the nature of the role as situated intervention (1970); Reiss's trenchant illumination of the public-driven nature of policing (1971); and Van Maanen's (1973); outline of the job as seen from the ground. Manning (1977); and Cain (1973); both influenced by Banton, complemented these works with research on English policing and the role of the officer. A second phase appeared as demands were voiced for changes in policing. This was a product of political response to the riots and protests of the late 1960s and early 1970s. Here, the demands of policy, concerns about extreme racial divisions that still exist, rioting and the police response, and the rise of reported "crime" emerged as political issues. The Law Enforcement Assistance Administration (LEAA) was created and funded research, and the metaphor of a criminal justice system was advanced in the Report of the President's crime commission. A third phase featured more intellectualized version of the same concerns arose in 1980s: a burst of more police-centric applied approaches such as community policing, "broken windows" policing, problem-solving policing, and applied managerial techniques such as crime mapping and crime-oriented meetings and analysis appeared. These quickly became amalgamated with increasingly sophisticated methodological approaches oriented exclusively to reducing officially recorded crimes and trumpeting the "experimental method" (Sherman 1992). This research featured the claim that arrests-are-deterrence and stimulated a slight but abiding countertheme of broader conception of both method (Sampson 2010) and the role and function of policing (Harcourt 2001). A fourth phase emerged arose with the drop in officially reported and recorded crime (ORC) in the early 1990s. As crime fell, the costs of policing and the number of police officers escalated in spite of this reduction that has continued until the present. A shift to a focus on stops, especially of minorities (Justice Policy Institute 2012), arrests for misdemeanors, downgrading of crimes, and drug arrests occurred in the iconic NYPD. There is considerable evidence that this was driven by management pressures (Eterno and Silverman 2012). In many respects, we are still in this fourth phase that combines a concern with crime reduction and national security post 9.11.2001. Perhaps, has it has in Europe, the economic crisis will stimulate consideration of restructuring policing, altering the conditions of work such as pensions, disability, retirement, overtime, and pay (Winsor 2012).

By way of summarizing these four modern phases, metaphorically speaking, the study of the police funded became a dance of convenience in which the music was supplied by Foundations (e.g., the Police Foundation, Heritage Foundation, Manhattan Institute), the Home Office in the UK and the Department of Justice in the USA (National Institute of Justice and the COPS agency), the police professional associations (IACP, PERF and in the UK, ACPO), and the big city police chiefs and commissioners. That is to say, the interests of the police-as-occupation, represented by their associations and leaders, the cooperative governmental agencies, and foundations, created the body of funded quantitative research captured in little technical studies that now dominate the journals. Michael Banton argued many years ago that there is a danger of developing a **sociology for the police** rather than a **sociology of the police**. The danger is quite simply that an occupation's interests, those of the police elite and supporters, are in large part but not exclusively a reflection of their political interests and, to a lesser degree, scholarly interests and theories. These two slightly overlapping streams of activity, scholarship and applied police-generated research, define a dialectic within police studies. They are comments on what the police are good for. No serious scholarship can be an illusionary reflection of the interests of the subjects of its inquiries.

Socio-Legal Systems, Types of Policing, and Defining Policing

Policing varies in historical, cultural, and social context. Even in the Anglo-American police world, it is not and never has been focused on or centrally concerned with legally defined crime and the common law that guards citizens' rights – it deals and defends politically defined order and ordering. Most research and analysis in the role of function of the police assumes a stable political order, while this is problematic even in stable democratic societies. Windows into policing are found in countries in the last century in states of civil war (Ireland, Spain, Ireland, China, Eastern Europe, especially Eastern Germany, the former USSR), occupied countries during World War II (primarily in Eastern Europe, France, the Netherlands, and Belgium), emerging states (Kosovo, Slovenia, and Croatia), police states such as Nazi Germany, Mao's China, and Stalin's Russia, and in radical transitions within and between legitimating systems of law, e.g., the current Middle East. These are anomalous systems which provide sharp contrast to democratic policing systems in stable societies.

Policing in relatively stable societies as a social form is grounded in a systems of law and cultural regulations. These include the following: (a) Islamic fundamentalist, (b) civil law-based systems (Western Europe outside Ireland and the UK, Latin America, Asia, Africa except for South Africa), (c) common-law-based systems (the remainder – North America, the UK, Ireland, etc.), and (d) quasi-legal martial law-based systems in occupied nations. There are also "mixed" systems such as those found in Macau, Hong Kong, Wales and Scotland, South Africa, Quebec, former colonies, and the State of Louisiana. All former colonial countries remain "layered" with local systems of conflict resolution operating in uneasy conjunction with formalized systems of the former colonial power. Some of these are still operating with what Weitzer (1996) calls "divided society models" of policing. These socio-legal systems in turn provide the umbrella under which the police operate in the sense that the two systems of primary concern here, the common law and the continental system, arise from deeply different conceptions of the role of the police and their connection to the state and, more specifically, the role of state security in the police mandate.

The image of policing in English language literature draws on the Peel model of policing, ignoring other sources of the role and function now apparent. While the "tradition" in Anglo-American policing studies is to cite Peel's aims and the restrained, uniformed, reactive, citizen-oriented, and crime-concerned service worker, "policing" in the Western world must include acknowledged and displayed and dramatized policing that is secretive, uses illegal means, focuses on spying, surveillance, and the use of informants in the interest of protecting the state. This form, sometimes called “high policing” (Brodeur 2010), conflates legislative, policing, and judicial powers in its actions, and polices “appearances” – i.e., as such the police watches and tracks citizens viewed as representing risks to the state and anticipates this rather than responding to actual behavior. This is an often denied yet essential feature of democratic policing (see below). As Jobard (this volume) correctly notes, this is a matter essential to the definition of all policing and its hidden yet enduring connection, even in democracies, to state security. Viewing the continental system, with explicit linking of violence and the law that the anomies of policing are most apparent, reveals the point at which the law’s limits are revealed and where violence is needed (Jobard). In this context, the connection of “politics” and “policing” assiduously blurred, denied, and obfuscated in common-law societies is made visible. The covering notion that public police are Peel-like in nature and engaged in crime prevention primarily and have little or no obligation to security functions, even post 9.11, remains. It is clear that on the one hand, the organization undertakes many tasks, as Bittner (1970) argues, in effect anything for which one “calls the cops,” while on the other hand, and that which is emphasized and dramatized as “the job” and how it should be carried out varies in the views of scholars and practitioners. Crime control is a misleading oxymoron.

Public policing has been elevated as the primary and most central of all formal sanctioning systems but bureaucratic full-time policing is a fairly recent invention and it still rests on citizen compliance, the support of complementary legitimate institutions, and the willingness of officers to serve a neutral role and to stay loyal. Policing of all kinds is tertiary to other forms of control, the primary controls such as family and peers, secondary controls such as neighborhoods, and formal associations such as schools and churches. Policing monitors and patrols order but does not create it. It often creates disorder and amplifies and sustains it when seeking to reduce chaos. It applies coercion to sustain politically defined orders. Examples of this are found easily in postriot overreactions, gratuitous violence, killings, and beatings such as happened in New Orleans in 2005. Public policing is one layer of a social control, the most visible perhaps, but, nevertheless, one of many kinds of social control.

Self-help or revenge remains the most common mode of conflict resolution, and there are modes of quasi-official policing associated with organized crime, rebellious, and resistance-oriented groups. These semi-organized systems sanction deviance often in defiance or contrary to the law but are seen as “fair” and “just” even as they are rooted in revenge rather than neutral third-party governmental actions. There are three other types of quasi-official informal policing and sanctioning organizations, i.e., those with legitimate use of force (Klockars 1985): occasional (ad hoc, hue, and cry obligations of citizens traced back to Edwardian period in England), obligatory (posses, vigilantes), and voluntary (militias, reserves, auxiliaries, the historic Texas Rangers), most of which are public, but some have been private policing, e.g., the Ulster volunteers of 1912–1914 later to be called the “b” special constables. Neighborhood associations and patrols are of this sort but are problematic in part because they have little or no accountability.

Of late, it is recognized that secondary social control occupations, those who regulate and respond to behavior informally in public and quasi-public spaces, are important semiformal control agents. These include such occupations as teachers, priests and clergy, bus and cab drivers, train and bus conductors, traffic wardens, and crossing guards.

In the last 20 years, new forms that lie between formal and informal types of policing have been officially sanctioned in the UK and other parts of Europe. Let us call these hybrid policing. This sort of policing is now included in the “police family” and denotes non-sworn officers such as PCSOs in the UK and enterprise zones in the UK and urban development zone agents in the United

States. There is confusion in regard to legitimacy of such agents because these functions can be carried out by such organizations but also by police employed privately, police employed and paid by the public, or some amalgam of these. This can also include third-party policing. Some mixed or hybrid policing, policing with combined uniforms with symbols of both private security companies and public police as in Lincolnshire, is a form of "out-sourcing." This term indicates using public funds to pay for agents to carry out functions connected the public good, e.g., issuing traffic tickets and providing order maintenance and some regulation that does not entail arrest nor authority in excess of that of the citizen. This syncretic innovation is rationalized on the grounds that it reduces costs, in part because it is budgeted differently than routine yearly personnel costs. Such policing functions have a mandate in the sense that they are licensed, legitimate, and have territorial limitations. They draw also on the mandate of public police by simulating their uniforms, interactional style, and the decoration of their vehicles and symbols.

Private Policing. While the line between public and private policing practices is blurred, private police act in the interests of those who employ them, are concerned primarily with the protection of information and property, and employ many structures and tactics to achieve this. These functions are usually connected to policing given private or semipublic areas (malls, sporting venues, etc.) In the United States, private police may carry visible weapons. They may or may not include some arrest powers in private spaces for trespassing. It is important to note also that the RCMP in Canada provides contract policing in Canada outside Quebec, Newfoundland, and Ontario that public police in the Anglo-American world work for private companies routinely and that there are many circumstances of cooperation between public and private police. This cooperative pattern is in part due to the fact that many retired senior police officers have now been employed by security firms or have begun their own. Police officers are also paid to also carry out functions that were previously exclusively public police functions such as insuring traffic safety around construction projects. In Boston these are called "details" and officers are paid by private corporations via the public payroll. The point here is that policing with private interests as the mandate only marginally, if at all, serves the public welfare, even when policing quasi-public property. The functions carried out by private police are far greater than those connected with quasi-public space (Forst and Manning 1999: 103–106). There are very large private security companies, alarm companies, and money transport companies that fall under the label of "private security" (Johnston 2003).

The public police are usually the referent of "police," but they are but one visible node in the larger network or web of policing functions. The police organization includes those who are not warranted or sworn officers. In addition, the organization can include reserve constables, cadets, and part-time officers. The organizational culture includes more than the segmented occupational culture of the police, and the percent of officers in uniform patrol in the UK has been declining for some years (Winsor 2012). The range of functions carried out by the public police is diverse and complex (Forst and Manning 1999: 27–29). Maguire (2003) lists some 28 discrete tasks carried out by patrol officers, and Winsor (2012) lists three times as many roles citing English forces. Bearing in mind that role and task are seldom closely connected in policing and that policing is always an open-ended enterprise, these are not helpful in understanding what is actually done on a day-to-day basis.

There are also publicly funded national and state armed forces, state national guards with their own investigative and intelligence units, and federal agencies with negotiated territorial obligations that also police. The ambiguity in the policing roles of the military is in part that they are constrained by the constitution with respect to operations within the continental United States. These functions are at best unclear as to accountability and scope. On the other hand, the US federal government has deemed the operating domain of their agents as unrestricted by traditional borders and costal boundaries.

There is little evidence of continuous cooperation with and across these types of policing and between formal policing agencies (Bowling and Sheptycki 2012).

A Definition of Public Policing. These considerations of types of policing and their relationship to systems of law and patterns of social control shape the nature of what is called policing in any society. So how should one define formal public policing? Consider a generic working definition of *police in Anglo-American societies* (North America, New Zealand, the United Kingdom, Australia, and Ireland), based on seeing the police as an organization rather than policing as a function. It assumes that the police are a bureaucratic organization and emphasizes the political nature of order and ordering, the new powers of surveillance and tracking that are balanced with coercion and violence. The police, because of their rather loose connection to the law and their political nature, act by exception. That is, they act in the shadow of legitimacy and the law and account or explain later only if required. In this sense, they act outside of or merely in the shadow of the law. It is an empirical question whether police practices consistently follow from these features.

> The police in Anglo-American societies are constituted of many diverse agencies, are authoritatively coordinated and legitimate organizations. They stand ready to apply force up to and including fatal force in politically defined territories. They seek to sustain politically defined order and ordering via tracking, surveillance, coercion, and arrest. As such, they require compliance to command from lower personnel and the citizens and the ability to proceed by exception.

The public police as an institution arise as a result of the development of law, differentiation, and functional specialization in a society. It is an institution that cannot survive without the consistent support of other institutions that complement its aims and practices. They are purveyors of organized coercion and the threat of coercion. The supporting institutions must complement the rationality, fairness, and procedural commitment of policing, or else it creates and sustains its own standards beyond review and remedy. This is Weber's argument Weber (1947) in respect to the collaborative nature of rationality in modern societies. The specialized functions of policing that arise as nation-states emerge and develop rest on broader societal consensus concerning fairness and justice arising from the public trust. They are a product in some sense of a balance of powers among executive, legislative, and judicial branches as set out in the original documents of a nation's establishment. One tendency, seen in crisis periods, is for the executive branch to either co-opt the police, making them directly accountable (as they still are in the Republic of Ireland and have been since 1922), or suspend *the habeas corpus* that in effect places the police under their direction. In this situation, the protections granted by legislation and by the courts are bypassed. Clearly, such considerations contain an explicit matter: the role of a "police officer" is determined by the organization's historical, cultural, and political mandate. Because the police officer in common law has original authority, the police organization is a useful fiction legally: it is constituted by the aggregated collection of officers (this is a contested legal concept and concerns the question of where accountability lies – with the officer, the Chief Constable or Chief, or the organization as a whole).

To understand the relationships between and among such a collection of types or clusters of formal policing agencies requires a label, concept, or some metaphoric work, which provides a way to think about policing organizations and that clarifies the relationships.Brodeur (2010) uses the metaphor of a "web" an assemblage of competing, interacting, organizations that function as if they were in a web. Shearing prefers the metaphor of a network with "nodes" and mentalities based on their technology and practices (Shearing 2005). Brodeur and Shearing show in their systematic treatments that "police," their role and function, cannot be restricted, except by exclusion, to the "public police." Nevertheless, the research and concern of scholars until recently has been with the public police seen within the Anglo-American Peel model. This is warranted because of their symbolic importance, their legal mandate, their control of data and exchange of this data with other agencies (Ericson and Haggerty 1997), and their huge and growing public budgets. The focus on the public urban police has meant scholars until recently have ignored for the most part the influence and importance of the range of types of nonpublic policing, private police, informal, hybrids, auxiliaries, and in public police federal police, state police, reserves, cadets, and policing in the interest of high security.

The Police in the Drama of Governance

The police organization acts, has an audience, and is a social object that people understand as real, constraining, important, and standing for something other than just a person in a blue coat and funny hat. The organization has legal status and is a legal actor. It is thus a "social unit" that has collective meaning, consequence, and symbolic significance. It performs. People respond to and animate the organization as real and doing things – fighting crime, being patriotic, issuing statements, and being an expense. It is in this sense a **social actor** with a social role in the game that is society. This is not a playful or whimsical notion; it is a way of considering the multifaceted role of the police in society. The police have many functions as an organization; it acts and reacts, defines issues, and responds to them. In many respects it is reified or seen as real when it is a rather dispersed, ineffectual entity dependent on citizen information and compliance. The police organization has a number of what might be called **governance functions** or dramas that indirectly reflect the role of the police as representatives of governmental authority. It is this representative function that is omitted in banal statements of tasks, roles, or functions, which are usually ascribed to the police officer. These individual roles and tasks, those which might be called **petite functions**, are discussed later. When dealing with "the police" as an actor, it is necessary to bear in mind at the same time that policing is carried out by individuals with a range of choice to intervene or not, when and how to intervene, and how to account for or describe why they acted or not. Even refusing to act is an action of sorts.

Here are some governance functions, ceremonial, or celebratory functions of the police, some of which are ritualized and repeated as a form of reassurance. They could be examined as propositions linking the police to the state. Let us consider these propositions in brief form. The police act as the face of civil authority displaying recognizable symbols, something that stands for something else, on public occasions – parades, funerals and weddings (of their own), and local commemorations. These stimulate collective solidarity because the police act as totems, icons of state authority for the majority. On the one hand, in their service, like the military, they symbolize volunteering for the common good and their injuries, deaths, and quests are sacrifices on behalf of us all. This is not to elevate the police per se, but to note that they play a role similar to that of the sacrificed in rituals in preliterate societies (Hubert and Mauss 1964). On the other hand, these collective occasions of ritual celebration mark and exclude those seen as outside the ambit of conventional acceptance whether they be minorities, lower class members, or immigrants. The police are conduits of governance, use, and display public resources in the name of the common good patrols, rhetorical strategies that outline their role and mandate in society, and tactics of organized action. This dramatizes their centrality to the collective dramas of control. The police are not passive actors who simply accept the legitimacy conferred upon them by tradition: they work assiduously to expand and sustain their own legitimacy as an organization, and this often conflates law as a set of mundane rules, morality as the "ought" and "should" of a society the sacred, and their miscues, malapropisms, and often banal, violent, and unbecoming actions. Their political role of the police is often subtle and less visible than their imagined media-amplified functions. They in fact symbolize their role in society by attending and protecting functions essential to the survival of democracy. These include monitoring voting, courts, judges, juries and court personnel, inaugurations, and massive events that celebrate the society to itself: American football bowl games, world series, and more mundane games also central to collective representation. They protect those whose who are "well known for their well-known-ness," those with an imagined life, "celebrities;" paid football and basketball players, movie stars, football coaches and their minions, and others temporarily thrust into the role of public icons. Why do Alabama state troopers escort the coach of the football team off and on the field? In their more mundane functions, they serve by their actions to define and perhaps obviate the nature of mundane risks and provide palliatives – crime, natural disasters, threats of terrorism, and the like. They communicate via the media their assessments and successes. They minimize, avoid, or redefine their failures. I have never heard a police officer declare publically: "We have failed miserably to contain control etc. . . . . . . .We apologize for this." They are cast as primordial eternal symbols of success. In the world of symbol work, they are active. They produce sharp and framed meanings for distant and unclear matters that lie outside normal experience – terrorism, mass murders, and impending disasters as well mundane matters of crime control. They are translators of the esoteric to the mundane. In this connection, they name, blame, and publicly constrain that which is seen as marginal, wrong, immoral, or in need of some sort of control. In this way they are redistributional agents: they distribute rewards as well as punishments. Those who are not monitored, not stopped, not questioned, or detained are rewarded with the absence of constraint, while those experiencing these are punished both formally and informally. The everyday actions of policing, those seen as profane, alter and maintain the social hierarchy in ways that are unacknowledged by the police themselves. Thus, for example, stopping those they suspect is "the job" and is disconnected from race, class, gender, and status honor. While the police may locate themselves in a mundane and unpleasant world of "dirty work" (Hughes 1971), they do act in the high and low dramas of society. That is, they seek in their everyday work to define and defend social order generally and its vicissitudes as well as dealing with everyday risks and concerns called "crime" or "disorder." High dramas, on the other hand, are those with public cachet, noticed and amplified by the media and for which accounts or explanations are called for. Low dramas are policing as usual. These dramas, via Twitter, YouTube, Facebook, and other social media, can become global-international. While they perform in little theatres, high and low local drama, they can be suddenly featured as heroes, fools, or villains in unfolding international and national dramas. The once little theatres present the events that by *media amplification* become national and international matters – serial killings and mass murders; local disasters, especially if mishandled by the police; hostage negotiations that go bad; and failed high-profile investigations. Less explosive matters – traffic stops, brutal beatings, arrests, mistakes at work – can bubble up and become unforeseen dramas of public concern. Police are expanding their stages and roles. Anglo-American police are now engaged in global policing of various kinds and in various guises (Bowling and Sheptycki 2012): officially and in informal cooperative roles with nonprofit organizations that sponsor volunteer in developing countries and in UN-based peacekeeping. There is a reflexivity about policing as they become aware of their own social capital. They reproduce by habit and practice notions of and images of social control characteristic of the police field. They frame ambiguous events as matters of police concern. They are the central players in communication dramas about themselves. They are asked for their opinions; they hold press conferences and issue press releases; they appear in court and make statements about current events. They reproduce their versions of dramas of control: who is at fault, who is to blame, who will be punished, and why. Think of public police statements about the threats represented by violent and dangerous gangs, brutal and senseless murderers, and unthinking rioters. These are symbolic packages that condense complicated social matters into known entities. The police are active actors, making up and playing roles. In this sense, they play themselves, acting out, mystifying, idealizing, and cooperating through teamwork to produce an imagery of discrete, mannered, and stylized control and service. In other words, much of their playacting is designed to magnify their own importance and is redundant in communicational terms.

The question arises: how does an organization, a handful of people, contain, control, manage, discipline, and punish a vast population in public and private spaces? They must have the cooperation and compliance of their audiences. There is a blurry, cloudy distance and unknown world "out there" that we little know directly. It cries out for some ordering, some coherence. At best, the police act to mark and sustain boundaries – horizontally and vertically – in the moral world of thoughts and feelings. They also actively act in the ecologically defined world of places, both public and private, and in personal and bodily space. They do this by carrying out patrol, stops, and arrests; by maintaining a visible presence; and by gaining access informally to private spaces – homes, buildings, corporations, and shops. They mark and patrol the edges of public space by their presence.

In summary, while the police as an organization have political interests and police officers have personal political interests, they functions as symbols, as representatives. The police organization must assiduously strive to maintain a degree of autonomy in the political web in which it operates, for this perceived neutrality is essential in a democracy. They do carry out governance functions. These functions are done in aid of the state and its power and authority, as well as sustaining the quality of life. While this list of functions is framed as "public policing," these functions can be examined in the context of the various types of policing listed above: informal, occasional, formal, hybrid, private and public, and the policing of the armed forces. Each has a mandate of some kind.

Policing: The Mandate and Acting It Out

All forms of regulation rest upon the base of self-regulation, or at best "self-help" (Black 1983), and mutual dependence, else it fails to become patterned, sustained, and acceptable. This patterning is institutionalization, the development of legitimate, named, and viable modes of accommodation to persistent anomalies, e.g., sin, death, the transmission of property and kinship, and socialization. As such, they are "going concerns" that mobilize human action and expression and customs for meeting the contingent and problematic in everyday life. In this sense, the police are a modern rationalized and rationalizing institution. Police are a legitimate, authoritative, conflict-processing system. Sociologically, and in metaphoric terms, an occupation seeks a socially accepted *license*, and if it succeeds, it strives for a *mandate* or the right to define the proper attitude and conduct toward the work, as well as even the ways of thinking and acting that are a reserved part of the practice (Hughes 1971:289). These are inclusive rights. On the other hand, it means that the occupation may well exclude and sanction those who act falsely in the name of the occupation – frauds, poseurs, or even criminals – and they will elevate that which they do that others may not. This mandate is a tacit bargain or contract between social groups and the occupation. As such, it is historically carved out and periodically defended by police and their allies. Thus, it is an expanding and contracting entity that reflects political, cultural, and economic forces. It is in this sense both shaped and shaping of public events. In part because of its association with authority, it is quasi-sacred and embedded in ritual, ceremony, and protective magic (actions and words that defend the vested interests of the occupation often counterfactually).

Far from being "transparent," policing is grounded in team secrets that are not shared with the public at large (Weber 1947:233) and in legal protection of its investigations and ongoing court cases. Certainly, policing has a mandate that varies by the public tacit acceptance or trust; it varies by race, gender, class, and age. It is best understood in this regard as compared with other occupations – physicians, plumbers, or professional sex workers. Within the ensemble of policing agencies, publicly funded policing maintains its grip on power and authority because, unlike private agencies, informal policing groups, and armed forces, they can enter a "case" into the criminal justice system for processing and potential prosecution. This is among the most powerful negative sanctions available in modern societies, regardless of the outcome of the processing.

This capacity, when combined with the on-the-ground potential for intervention in that-which-needs-responding to-else-it-will-get-worse (Bittner 1970):34, weds inextricably law, power, and authority. This formulation is the edge through which "politics" and irrationality, in the sense that it lacks long-term consideration of the law, justice, or consequence, enter the scene (Jobard, this volume). This "framing" or defining the situation is not simply a here and now situational contingency action; it is the institutionalized basis for police authority.

Policing is a role drama. The role-occupant acts, does things, and decides things, and this variation in individual performance requires what Goffman (1959):30–70 calls a "front" (props, manner, and setting-appropriate behavior) as well as devices to enhance performance: dramatic realization, mystification, misrepresentation, maintaining expressive control, all in all a kind of ritualization of actions that sustain a *definition of the situation*. Recall that each new situation is an unknown for an officer and that patrol officers subscribe to the notion that while an officer may have seen "everything," he/she believes that one must be ready for the unexpected and to cope with it and control it quickly. The parallel is succinctly stated on page one of The Presentation of Self, but one must add the additional line for police: "and manage the outcome authoritatively."

The Police as a Rational-Legal Bureaucratic Organization

The police organizations found in the United States have a number of well-known features and functions. They are diverse in size, with the modal organization small, under 15 officers; they hire and train officers using local standards and preferences, but officers are more educated and forces are slightly more diverse as time passes; they are locally funded and embedded in local politics and issues rather than national party politics; they respond to and are part of the executive branch of government and are therefore loosely tied to and accountable to the elected executive. The American system by design, a product of the revolutionary tradition and the experience of policing under the British as a colony, is "fragmented" made up of many loosely connected part or organizations. There are about 19–21,000 agencies – federal, state, and local – of which an estimated 13,500 were attached to local agencies. A 2004 census (Bureau of Justice Statistics 2007:2) revealed 12,766 local agencies, 3,067 sheriff's departments, and 1,481 special jurisdiction agencies. This category is an odd assemblage of federal agencies. In summary, the country is sprinkled with many independent agencies, with little centralized coordination of their actions, a weak and rather small federal compliment of agencies, and little of the centralization of function found in the other Anglo-American and European states.

There is no "typical" American agency, because most are quite small. Police agencies in the United States operate within a complex set of mutually contradictory laws, local regulations, and practices; their cooperation patterns are informal and their capacity to communicate with each other varies widely in both quality and quantity of channels. In contrast to the governance functions listed above, the police carry out diverse **petite functions** that are visible, local, and rather pedestrian. Petite functions range from the ridiculous and bizarre and mundane such as capturing wild beasts and sweeping up glass in the street after a road accident to dealing with hostage situations, serial killings, floods, hurricanes, homicides, rapes, and man-made disasters. They are historically frozen into a pattern of strategies of deployment developed in the eighteenth century in Dublin (The Dublin Metropolitan police were founded in 1786): random patrol and response to citizens' requests. The crime investigation function was added in the mid-nineteenth century in London.

Tactics, in part based in reduced time to respond, cheap and rapid communication systems especially the telephone and the computer, and increased education of the population are diverse but are restrained by traditional policing habits. The influence of information technologies is generally to increase speed of response, record keeping, and information retrieval, but the impact of IT on everyday functions is yet unknown. While the American police are well armed and increasingly "militaristic" in dress and dramatization of their role in the media and in their own advertisements, they have become less violent as has the public. Their dramatization or modes of organizational presentation in rhetoric and imagery are multifaceted: peaceful and child and family oriented, available, and kind as well as violent, well armed, distant, cold, and secretive. The police solicited demand, claim it is their role "you call, we haul," yet they screen and manage demand and are occasionally overloaded. They do not equally serve citizens or all parts of the city.

Changes in the Police Organization in the Last Century

The most significant changes in police organization in the post-World War II era (Reiss 1992) are changes in the composition of cities, their spread and the division of policing units into mobile beat, and improved, rapid, and inexpensive technology facilitating input and output of information. Police work, without question, is become more legalistic, demanding in respect of detail and information processing, and more subject to public scrutiny via cell phones, media, computers, and CCTV and other surveillance cameras (public and private). The police are a conservative organization, and their role and function sustains this idea. They preserve the *status quo* as defined loosely by the current legal and political realities. As such, they may find strong resistance as in the civil rights era beginning in Little Rock, Arkansas; Birmingham and Montgomery, Alabama, in the 1950s and 1960s in the South; and the bussing period in Boston in the late 1960s.

Internally, police organizations have become more specialized (the percent working in specialized named units such as "schools," "gangs," "special weapons," and investigation more generally). The percentage of employees who are non-sworn is growing and is over 25 % in large departments. This will grow and has implications for the future employment of sworn officers. With these two changes has grown an ostensive "rationalization" or developing of specific, codified, and explicit rules governing procedures relating to suspension, termination, disability, promotion, transfer, and conditions of work. These developments are countered by the traditional adherence to personalized relationships, sponsorship of protégés, networks of political oriented groups, e.g., crime control versus community orientation amongst captains and above ranks in the patrol division, resistance, and distrust of authority especially of the top command. Gouldner (1954) has suggested this kind of rule-oriented organization is a mixture of a form of formal authority called "punitive bureaucracy"– using rules to punish those who are indifferent or opposed to the management clique – and a mock bureaucracy which denies rules when convenient and works on informal compliance. These changes are accompanied by a move away from foot patrol and visible presence in neighborhoods, a diminution in face-to-face supervision – this being replaced by electronic monitoring by supervisors of mobile digital terminals, cell phones, and electronic records. While this has made policing more reliant on communications systems, the integration, cleaning and replacing, updating, and making these systems more user-friendly has lagged (Manning, this volume on police information technology). The primary capacity for surveillance, tracking, record keeping, and monitoring of citizens generally has vastly increased in the last 20 years. Finally, as Reiss (1992) has pointed out, the constraints of community-based accountability efforts, "transparency" politics, collective bargaining agreements, and contracts more generally have placed the police in a network of obligations now that were more tacit, less complex, and more easily avoided or voided than in the past. The process of rationalization reduces the power of policing as a sacred entity because its uniqueness is slowly diminishing.

Although studies of police technology suggest an increasing reliance on information technologies of various kinds and the rationalization of policing actions, i.e., aligning the ends or stated objectives of policing with the means designed to accomplish them, determining this remains difficult. The **ends** of policing are broad, undefined, and general in nature, and include maintaining order, serving the collective good or quality of life, advancing human rights, law enforcement, and representing trust. These are tacit, assumed, and implicit and police have long found them difficult to define operationally. As a result, they have resorted to indirect surrogates such as response time, arrests, officially recorded crime, clearances, and other more trivial measures such as class for service answered or school visits. In addition the primary **means** to achieve these are debated within and without the police service. Should they be accomplished by "partnerships" and "community policing"? Should they mount "crackdowns," sweeps, and targeted arrests in disadvantaged areas? Are they suited to planning and carrying out long-range problem-solving exercises? Should they do concentrated patrol or continue random wandering? In action, more traditional approaches combine the traditional structure and tactics with a crime focus, one sort of police rationality; other approaches favor community-oriented programs and partnerships, another form of rationality, while the majority of police officers have no interest in policy of any kind, see the job as a kind of entrepreneurial activity, and see their role as a rational attempt at protecting the city.

While in common sense terms, the police are an integrated and coherent rational organization; in fact, a police organization is an arena for contested rationalities associated usually with politically defined clusters or segments of officers based on these rationalities. The groups are in some tension, especially during periods of high officially recorded crime, concentrated homicides, or corruption scandals. That is, members of police departments do not agree on the goals nor the means of policing, and they are divided by segments as well as rank. At the same time, there is publicly expressed tacit agreement and consensus.

Rationality, Efficiency, and Effectiveness

As changes have been forced upon public policing, they have altered their rhetoric to simulate the language of big business. They speak of the work as "business" and supervision as "smart management" and produce annual reports, business plans, and strategic manifestos. They feature concern for "customers" and their "product." Let us consider briefly the nature of the new rational police organization now dramatized. This is an imagined organization that is based on clear objectives, long-term goals, and means geared to these goals. The resources available are tightly geared to this model of operation and closely monitored for results. Clearly, it is preferable in a capitalistic society to urge police to be rational, effective, and efficient while expecting them to deal with whatever arises with endless endurance, resources, morale, and competence. Thus, these terms are used without definition or context. Police are a patrimonial, punishment-centered bureaucracy, with loosely coupled connections between segments and units (see below), and they are a field for contested rationalities. Concepts such as efficiency and effectiveness are misleading and probably irrelevant to judging police operations.

Consider the following features that are in distinct contrast to the claims. There is no conventional market that can be used to distinguish market share, profit or loss, increased value, or net worth that can be traded or exchanged in a marketplace. Police have no customers or clients; they serve the public at large – the collective good. We cannot refuse the services or find other sellers. Taxpayers cannot opt out of paying their taxes or that part that support the police. The "customer" of the police cannot always be right if indeed right at all. Police, like other service organizations, cannot relocate to reduce their costs; they have general territorial obligations that are in effect "open-ended" and for which the demand is elastic. They are locally budgeted and monitored. They are insensitive as a result of contracts stipulating the conditions of work and pay, to variations in the market economy. They claim a mandate based on "full service" 24/7 without screening, prior prejudice, judgment, or exclusion. Their product is trust, in that they are agents of the whole designated to assess the trustworthiness of others and to act on their assessments. They, in turn, rely on public thrust and complicity to operate. This means that any attempt to screen or reduce service, to lay off officers, will be criticized. The police are expected to react to any danger to public order – man-made, natural, or artificial, e.g., nuclear power plant leakage until such time as it is resolved. This means at any given time, police hold back slack resources in anticipation of such a spiked and unanticipated demand. There is no standard number of officers per capita to hold in reserve, and there are no formulas for deployment. Overtime and extra personnel are always available in the case of an emergency. The goals are unclear and objectives contentious, and the means, e.g., arrests, only vaguely connected to the ends of public order, crime prevention, or protection of life and property. The standard practices, or "good practices," are not easily adopted and applied in El Paso and Elmira, Chillicothe, and Chicago. There is no market for police services: they have little or no competition and no standards against which to judge themselves since crime is a function of size of the city other things being equal. There is no bottom line in policing. It would appear also that to be efficient, that is, to use minimal resources to achieve an end, would require policy to guide actions – this is virtually impossible in a police organization, given the number of officers at the bottom of the organization and discretion at the bottom. Perhaps the most viable option in this regard is reducing the cost of employees by hiring civilians and increased use of technology to process data. This of course has little to do with preventing crime in any direct sense. Turning to effectiveness, it is clear no one wants a police force that enforces all or even most of the law, all or even most of the time. The sensible judgment of officers is essential in this regard. Weighing these arguments perhaps underscores the political significance of "crime control," "crime prevention," and "crime control" to police legitimacy. These statistics are matters that, given their control of reporting and arrest statistics, police can shape to their advantage. While the public agrees with the police about their concern about crime, their collective sense is in fact that the police are a public service with diverse governance and everyday functions.

The Police Organization in Action

Having sketched the external governance and petite functions of the police, it is possible to analyze the internal configuration of the police organization. This can be seen in two ways. The first is the police organization as an extension of the state's authority requiring loyalty and commitment. The second is to see it as an ensemble of roles carried out by quasi- independent actors. This role ensemble idea is taken up below.

Loyalty and Commitment. Let us take up the first. In this sense, the notion of an organization is variously a constraint on action choices. The idea of an organization assumes a degree of commitment or consistent behavior in the course of routines that are organizationally defined and attached to an income, hours, and duties. That is one definition of an occupation. Commitment is a mere job description, a way of capturing the process by routines become a role. But an organization requires involvement, attachment or even loyalty, or signs that emotional investment is being made in the organization's actions and ends, if not the means. The question of loyalty is seldom raised in police research, and when it is studied, it is revealed that police officers do like the job and, depending on age, gender, and specialization, tend to be emotionally attached to their work or "the job," but not the organization that employs them which they view with cynicism, distrust, and temerity. These attitude studies are complemented by work on the behavior of police officers in New Orleans after the Hurricane Katrina and shown in Hunt's (Hunt 2010) analysis of officers involved in a successful antiterrorist raid. The issue of loyalty has an edge in policing because the combination of distrust of the public at large and a view of the organization as unsympathetic and dangerous to their career and job interests makes for a powerful force for internal obligation and external stereotyping. It is not so much danger and displacement of aggression and hostility onto marginal others, but a combination of individualistically defined work with periodic risks. This stance toward the world when combined with a crime-fighting stereotypical self-label for overt public consumption sets the boundaries of the job and the social world of the occupation.

An Ensemble of Roles. The *police role* is embedded in uncertainty and the need to resolve it quickly. The role, with its characteristic features, arises not only from the complex of ideas that sustain the job – attitudes, values, beliefs, and material accoutrements, as well as practices – but from its craft-like nature. Like a carpenter, or plumber, officers take the situation that they define as facing them and shape it, if possible to a sensible or reasonable outcome. This result may be the preferred or even the sought after outcome, but "it is what it is" as the current saying goes. There is much that can go wrong. The management makes do with what is possible and adjusts, shapes, cuts, and shims things until they cohere or, at least, (a) do not fall apart and (b) come back to up "banana shaped," or when ". . .the shit hits the fan," (c) allows you to cover your ass in the event that something goes wrong (d) works in the here and now. Police work is done very much in the here and now. It is an intentionally conscious present-oriented doing (or not) hoping things will not get worse. If one can, one avoids paper work and further complications, violence and the related complaints, investigations, and the rest (if anticipated). That for which a record exists should be flattering and compelling to those read about it. What is acceptable to one group of citizens (or officers) will not work with another; for example, threats and exhortations may work on the street but do not suffice for middle-class people; good manners and etiquette work in the suburbs but may be dynamite and lead to ridicule and violence in disadvantaged areas.

Segments. Now consider the second way for seeing the organization: as a context for a loosely connected set of segments. This is the "motivational" outline for the politics of police organizations or how power seeking animates the internal politics of any police organization. As this perhaps suggests, there is no unified or even coherent police ▶ occupational culture in the sense that is often referred to in textbooks. It is not even useful to see it as one occupational culture with subcultures because this implies more unity in the occupation. Even this reified picture of one culture is drawn from research that reports the views of white, male urban patrol officers and their work-based oral culture. This conception excludes the meaningful experiences and views of females, minorities, and the very large civilian population that manages the infrastructure of the organization. In that sense "the police occupational culture" is a misleading gloss on complexity. Since an occupational culture is a response to the patterned uncertainties of the work, the irresolvable, recurrent difficulties of deciding that cannot be resolved factually, it has variable features depending on the prevalent uncertainties. What is usually called "police culture" is reported talk full of hyperbole and exceptions, a kind of tool kit on the one hand to resolve the fundamentally incongruous aspects of the work and a configuration of warning signs about how to keep out of trouble. This means further that much of what is talked about is exceptions, stories that are meant to express cautionary tales, ways 'round trouble, and all those things to be thought of as "on the job difficulties." It is helpful to think of the police organization as a loose confederation of sense-making *segments*, people who face similar uncertainties, express this coping verbally, and interact more frequently with those in their segment than with those outside it. Clearly, the responsibilities and functions of social segments in the organization, the patrol officer, those in supervision and management, investigators, and officers in specialized units do differ. The non-sworn segment of the organization, those called by police "civilians," provides an important contrast, but they are not discussed in detail here because there are no studies of this group. They reflect, it would appear they reflect the more stereotypical attitudes of the patrol segment with who they most interact (Manning 2008), but this is subject to more careful investigation.

A Map of Occupational Concerns

Consider Table 1. The columns across the table are the segments: lower participants, sergeants and middle management generally, top command, investigators, and specialized squads. The rows show the kinds of contingencies or uncertainties that pattern their occupational concerns. These include how officers entered the organization, their perceived audience, risks, rewards, and source of authority. There are several important points to note. Entry becomes more "political" and personalistic as one moves from left to right on the table. There is an assumption of equality of all accepted upon their entry but this changes quickly within the first few years of work (Van Maanen 1973). The perceived risks become in some sense narrower and more geared to specific skills and craft-competence. The audience of importance shifts radically when one moves from the officer and middle management segments. The audience is more public in the top command and investigator role. The authority of the street officer is personal, local, and immediate but those in supervisory and specialized units must cultivate special skills and competencies. The rewards concern change from concrete and known to more subtle questions of career and move away from immediate questions of pay, overtime, comp time, and secure retirement. These are cross-sectional rather than dynamic and set out the primary distinctions within the police world.

Now consider Table 1 in more detail. Note that the columns note the various segments of the

Role and Function of the Police, Table 1 Police occupational cultures

	Officers	Supervisors	High command	Investigators	Special units (SWAT/gangs)
Entry	• Screened • Various criteria	• Place and • An exam and • Seniority	• Political/ interview • Outside participants • Clique	• Selection • By Interview • Seniority • Brief training • Apprenticeship	• Recommendations • Request/approval • Protégés, networks
Risks	• Streets • Mistakes • Paperwork	• Officers' errors • Complaints	• Political/media • Second career	• Perjury/court • Paper lies • Boredom • Big cases	• Danger • Media • Uneven work demands
Audience	• Lateral • Peers	• Upward • Laterall	• Worthies • Clique • (Troops))	• Peers in investigative work • Media	• Other specialists • Media/public
Authority	• Original • Personal	• Streetwise • Rule-wise	• Personal/role/ position • Little actual power	• Original • Being clever • Well-spoken	• Quasi-legal
Rewards	• Overtime "details" • Comp. time • A niché	• Overtime • Promotion • Reputation with peers	• Respect (media attention) • Authority/ leadership	• Wearing plain clothes +allowance • Overtime • Participating in special squads • Flexible hours • Working "big cases"	• Peer respect • Flexible hours • Public fame

occupation (i.e., groups that interact more within that set of people than with others and partially defined by rank): patrol, supervisors, top command investigators, and special units. The rows set out the various features of the job as they see it, the lens through which the occupation is viewed: these include how they come to the position, their perceived risks and audiences, their view of the nature and source of their authority, and the rewards they associate with the work they do. Note that officers are hired by various criteria, see their risks as being "on the street" and associated paper work, are oriented to their peers, and see their authority as personal and original. Their rewards vary, and this is one bass for divisions within the segment – the nature and extent of the rewards sought – formal or informal. A niche is a kind of hideaway spot protected from transfer and demotion based on special non-rank-related skills such as working easily with computers, cars, or interpersonal relations – in the Chief's office, for example. Higher rank is a rather vague goal, not sought by most officers, so is not a key feature of the patrol officers' reward systems. Supervisors, mostly sergeants and lieutenants, gain their rank by a combination of changing factors – written exams, oral exams, and reputation. They are concerned about the risks associated with an officer's mistakes, especially those that are featured in the media, and they are "middle or marginal" people because they must balance the expectations and work styles of their squad (8–15 people for a sergeant) and those of the top command. They work with some combination of the rules, formal discipline, and informal persuasion and are generally respected. They may be rule-wise or streetwise, and their reputations as such preceded them. They may be given overtime or not or seek promotion and value the respect of their officers or squad. High or top command is constituted of the ranking officers, although the top handful in any given organization has considerably more informal power than those between them the sgts. Few studies have been done of top command and by inference one sees them as politically skilled; often rising via detective work; oriented to external audiences; seeking perhaps a second career in politics, security, or academe; and eager to cultivate power though linkages to others of like mindedness within the organization (networks of protégés, followers, and political supporters). They are sensitive to political, media, and even academic reputation and are often now highly educated (J.D., M.A., or Ph.D.). Turning now to investigators, "plain clothes" officers, they are generally chosen by reputation and connections which are then sanctified by other procedures. They are often sought out and cultivated if they have a reputation as being good at "crime work." They work as apprentices and may have been detective in a temporary position. Their risks are crime and crime-related and surround the rare case that goes to court. They always claim to be "overworked" and overloaded with cases, regardless of the actual workload, the clearance rate, or the complexity of the cases in which they are engaged. They are peer oriented and overvalue detective work considering it the heart of the job when combined with good patrol work (which can simplify their cases!) and value flair, cleverness, interviewing skills, and court aplomb. They are visibly rewarded, well paid, as well as linked into contractual pay for court whether they appear or not. Special squads vary in size, training weaponry, and rewards. Now almost all police departments have some specialized squads in addition to investigators. They are selected, cultivated, and generally "well connected" within the police for their skills and the prestige attached to carrying out violent, heroic deeds. They value and seek risk yet are aware it can be the undoing of a career in this sport of work. The work is often uneven and boring. Their audience is other specialists in their department and nationally. Because they deal with exceptions, hostage taking, riots, demonstrations, disasters, and rare events, their work is always on the borderline of legality. They are well rewarded and honored by peers, the media, and the general public. Their equipment, appearance, and especially their weapons are a notable source of public awe.

Dynamics. This scheme of segments and their features has a dynamic respect. These are vertical and horizon *cliques*, that is, informal friendships and sponsorships of like-minded officers; these links can be kinship, cohort members – those with whom you experienced the training academy – past partnerships, gender, ethnicity, or race. For example, young officers are informally selected for special squads in part on the basis of sponsorship from respected members or former members of such units, and these squads more than other segments reproduce themselves in respect to key identities – similar race, gender, and ethnicity (Hunt 2010). There are also movements in rank by promotion and demotion, changes in cohorts in given hiring patterns (numbers and composition), and leaving via disability, retirement, or termination. There are thus "cohort effects" that are visible in police organizations currently – the absence of hiring and the retirement of officers has created a thin middle segment and disproportionate number of newly hired officers serving on the street. As returning veterans are given preference in hiring, the incoming cohorts will be older, more disciplined, and experienced.

It should also be clear from Table 1 that there is no single role that police officers fill (as opposed to the rule the organization plays in the society), and no single ► occupational culture, even in the patrol division. There are in fact several segments and cohorts in competition for resources, career goals, and informal rewards. The extent to which these segments and dynamics are shaped and affected by diversity of race, gender, and sexual orientation is an open question. It would appear that several generalizations can be made: diversity exists more at the patrol officer rank than elsewhere in the organization, that women and minorities hold lower and higher ranks more than at middle-level supervisory ranks, and that changes in such ratios have moved very slowly in the 30 or so years since "diversity" has been urged. The claims about major change in police organizations as a result of recent hiring are yet to be proven.

Cohesiveness. If this is not a coherent culture, what holds this set of segments together? It would appear to be a number of features of the occupation.

- Every officer enters at the bottom, goes through the academy and field training, circulates through various units of the force, and is assigned to patrol or "fast response." All that rise to leadership share these experiences. The top command is selected by promotion within the organization and rarely is recruited from outside.
- Personalistic patterns of loyalty based on past friendships, cohorts in training, experience in special units, and the territorial units or divisions in which officers have served are highly valued. This is one source of the social glue that sustains relationships between people who work alone in widespread distances from each other carrying out quite different roles and tasks. These connections cross ranks and segments.
- Loyalty between and among ranks is patterned by but not determined by rank and command and control relationships. Yet, such personal loyalty counters rational rule enforcement and sustains what Weber (1947) calls "substantive rationality."
- Interactions on the job operate with metaphors of cohesion, e.g., "the police family" and "professional policing." These are invoked to draw out commonalities in public discourse.
- "Common sense" is used to describe the nature of the job, and it is believed to be widely shared throughout the organization. The common sense reality of policing is based upon the notion that all police officers begin at the bottom, that the craft is learned there, that shared values, practices, and norms arise from these experiences, and that "that's where the job is": at the "coal face." Insofar as the public has no access to these experiences, they are dismissed as judges of police performance. In critical incidents, the media and politicians are dismissed as ignorant and opportunistic in their criticism.
- Mutual dependencies arise from the shared uncertainties – the fear of capricious enforcement of the rules by supervisors that induces compliance amongst the lower participants and fear of their errors and mismanagement being exposed by the top command. There is also the worry of those at the top that the officers on the street might make an egregious

error that reflects on their authority and imagery. These typically involve excessive violence, such as vehicular chases ending in crashes, deaths and serious injuries, beatings, and shootings.

- Public displays of unity, participation in public celebrations such as parades, funerals, sports-based ceremonies, and as a result of acting collectively and in a unified, forceful manner when combating riots and disorder present the front stage of policing in dramaturgical terms (Goffman 1959). These are the rare occasions in which police act as a paramilitary unit with all the visible costumes, weapons, music, and audiences.
- Occupational consensus is sustained by the vague, unspecified "other" that respectable people fear and which, in turn, the police are expected to combat, manage, destroy, or erase from the publics' mind. It is the belief in the support of the respectable people that nurtures morale in spite of the "dirty work" and criticisms. This is more acute and verbalized in large cities where the valued audience is almost exclusively other "cops."
- Doing good police work, ironically, is seen as dragging some situation out of a chaos and making the best of it. There is a clear notion about what is "good police work" that is tacit and indicated by stories, and bad police work such as "bad shootings," but these are elastic notions and always put in the context "You had to be there." This might be called applying flair to complexity. "Flair" is managing and anticipating all those things which any reasonable person would do in such a situation as the person doing it saw it (not how an outsider looking at the deciding after the fact might see it).
- Powerful cliques of like-minded officers, groups including younger protégés, and their sponsors, based in part on kinship and ethnicity, shape careers. They promote and support favored officers as well as either do not protect others or actively prevent their promotions, transfers, and access to temporary postings. This is the semi-invisible tissue that binds some officers and excludes and extrudes others. Cliques are joined loosely into networks that link officers up and down the rank structure. The ambitious are typically members of the current leadership network and its protégés and a network composed of those currently out of power who may gain control with the appointment of the next Chief, Commissioner, or Chief Constable.

Other Divisions. There are matters other than already-mentioned ones of civilian versus officer status and rank and the interests of rank and rank segments (lower participants, middle management, and top command) that *divide* the organization in more subtle fashion:

- Strong lines divide the organization by gender, race, and ambition. While gender and race discrimination are well known in all organizations, the police organization is notable for its profound anti-intellectual and anti-educational milieu. This bias also includes the overtly ambitious who are often relatively well educated. Educational credentials are seen as a means to increase pay and the retirement base, rather than as a source of wisdom, judgment, skill, or knowledge. Educational credentials are deemed as irrelevant to the craft.
- Rules are the source of punishment and are used to substantiate decisions about discipline and performance. This feature, the use of rules as punishment, rather than as guidelines for choice, makes police organizations "punishment oriented bureaucracies" (Gouldner 1954).
- Cycles of rule enforcement increase the complexity of the organization in the eyes of the lower participants. They view rule enforcement as impossible to anticipate, always a possibility given any decision, capricious, and unfair.
- Rank and salary rewards are few. Ambitious officers in large departments, recognizing the personalistic and apparently arbitrary nature of "success," are often alienated and/or frustrated with the job. There are few opportunities to lateral movement to other police organizations.
- Emotional connections, losses, and achievements punctuate the police organization. Failures in high-profile cases, transfer and

symbolic demotions, unexpected promotions in which officers jump a rank or two, promotions to "inside jobs" such as the police chief's staff, forced retirements as a result of realignment of members of the powerful top command segment, and other career contingencies are all fraught with and produce emotion. Ironically, because emotional sensitivity and responsiveness other than anger are negatively sanctioned in the organization, these emotional burdens are denied and remain explicated yet powerful (Hunt 2010).

It is important, given this gloss on the occupational culture, and the role of what might be called the lower participants in the organizational hierarchy, to note several important qualifications. The patrol division is the largest segment within the organization, and it dominates the rationale for the job within the organization, i.e., the view that the organization is there in some radical real sense to "fight crime" on the streets and that this is the real job. This ideology denies the well-known facts that policing is boring, rarely involves crime, arrests, or violence, and requires scientific and legal work of a large number of "civilians" (about 26 % of large departments), investigators, managers and planners, "politicians" such as the top command, and consistent public support and finance. Nevertheless, this is the ideology shared up and down the organization and by many police researchers who carry out "experiments" designed solely to reduce officially recorded crime regardless of other (unmeasured and unanticipated) effects on the quality of life. It is in effect an organization which based on interaction is segmented, like preliterate cultures, and not a single culture, but many overlapping and competing cultures. The organization is more like a chorus than a solo performer.

Conclusion

The study of the role and function of the police is based on assumptions, most of them with an implicit interest in reform. Police as an institution reflect systems of law and culture, as well as the nature of the governance system within which they function. The police carry out both grand functions in respect to governance and more mundane or everyday petite functions. They have a flexible mandate, strategies and tactics, and function as rational-legal bureaucracies. There have been a number of changes in policing, but its core functions remain intact. Shifting attention to the inner functioning of the police organization, four major topics were addressed: the police role, loyalty, and commitment; the segmentalized nature of the organization and related dynamics; the sources of solidarity; and the remaining salient divisions and tensions within the organization.

Related Entries

- ▶ Causes of Police Legitimacy
- ▶ Comparing Police Systems Across the World
- ▶ Conceptualizing of Police
- ▶ Democratic Policing
- ▶ Police Culture
- ▶ Theories on Policing and Communities

Recommended Reading and References

Ayling J, Grabosky P, Shearing C (2008) Lengthening the arm of the law: enhancing police resources in the twenty-first century. Cambridge University Press, Cambridge

Banton M (1964) The policeman in the community. Tavistock, London

Bittner E (1970) The functions of the police in modern society: a review of background factors, current practices, and possible role models. National Institute of Mental Health, Washington, DC

Black D (1983) Crime as social control. Am Sociol Rev 48:34–45

Bordua D (1968) The police. In: Sills DL (ed) International encyclopedia of the social sciences. Macmillan, New York, Vol. 2

Bordua DJ, Reiss AJ Jr (1967) Law enforcement. In: Lazerfeld P, Sewell W, Wilensky H (eds) The uses of sociology. Basic Books, New York, pp 275–303

Bowling B, Sheptycki J (2012) Global policing. SAGE Publications, London

Brodeur JP (2010) The policing web. Oxford University Press, New York City

Bureau of Justice Statistics (2007) Census of state and local law enforcement agencies. United States department of Justice, Washington, DC

Cain ME (1973) Society and the policeman's role. Routledge, London
Ericson RV, Haggerty KD (1997) Policing the risk society. University of Toronto Press, Toronto
Eterno J, Silverman EB (2012) The crime numbers game: management by manipulation. CRC Press, Boca Raton
Forst B, Manning PK (1999) Prive policing: two views. Georgetown University Press, Washington, DC
Goffman E (1959) The presentation of self in everyday life. Doubleday, New York
Gouldner AW (1954) Patterns of industrial bureaucracy. Free Press, Glencoe
Harcourt BE (2001) Illusion of order: the false promise of broken windows policing. Harvard University Press, Cambridge, MA
Hubert H, Mauss M (1964) Sacrifice: its nature and functions. Translated W.D. Halls Foreword by E. Evans-Pritchard. University of Chicago Press, Chicago
Hughes EC (1971) The sociological eye: selected papers. Transaction Publishers, New Brunswick
Hunt JC (2010) Seven shots: an NYPD raid on a terrorist cell and its aftermath. University of Chicago Press, Chicago
Johnston L (2003) Governing security: explorations in policing and justice. Taylor & Francis, London
Justice Policy Institute (2012) Rethinking the blues: how we police in the U.S. and at what cost. Justice Policy Institute, Washington, DC
Klockars CB (1985) The idea of police. Sage, Thousand Oaks
Loader I, Walker N (2007) Civilizing security. Cambridge University Press, Cambridge, UK
Loftus B (2012) Police culture in a changing world. Oxford University Press, Oxford
Maguire ER (2003) Organizational structure in American police agencies: context, complexity, and control. State University of New York Press, Albany
Manning PK (1977) Police work: the social organization of policing. MIT Press, Cambridge, MA
Manning PK (2008) The technology of policing: crime mapping, information technology, and the rationality of crime control. New York University Press, New York
Moskos P (2008) Cop in the hood. Princeton University Press, Princeton, NJ
Ratcliffe JH (2003) Intelligence led policing. Trends in crime and criminal justice. Australian Institute of Criminology, Canberra, Australia
Reiss AJ (1971) The police and the public. Yale University Press, New Haven
Reiss AJ Jr (1992) Police organization in the 20th century. In: Tonry M, Morris N (eds) Modern policing: crime and justice, a review of research. University of Chicago, Chicago, pp 51–97, vol. 15
Reiss AJ Jr, Bordua DJ (1967) Environment and organization: a perspective on the police. In: Bordua DJ (ed) The police: six sociological essays. Wiley, New York
Sampson RJ (2010) Gold standard myths: observations on the experimental turn in quantitative criminology. J Quant Criminol 26:489–500
Shearing C (2005) Nodal security. Police Q 8:57–63
Sherman LW (1992) Attacking crime: police and crime control. In: Tonry MNM (ed) Crime and justice: a review of research. University of Chicago, Chicago, p 159, vol. 15
Skolnick JH (1966) Justice without trial: law enforcement in democratic society. Wiley, New York
Van Maanen JE (1973) Working the street; a developmental view of police behavior. In: Jacobs H (ed) Annals of criminal justice. Beverly Hills, Sage, vol. 3
Weber M (1947) The theory of economic and social organization. Oxford University Press, New York, Translated by T. A. H. a. T. Parsons
Weisburd D, Braga A (2006) "Police innovation." in Cambridge studies in criminology. Cambridge University Press, Cambridge
Weitzer R (1996) Transforming settler states. University of California Press, Berkeley, CA
Westley WA (1951) The police: a study in law, custom and morality. Ph.D. Dissertation, University of Chicago, Chicago
Westley WA (1970) Violence and the police: a sociological study of law, custom, and morality. MIT Press, Cambridge, MA
Wilson JQ (1968) Varieties of police behavior. Harvard University Press, Cambridge, MA
Wilson OW, McLaren R (1963) Police administration. McGraw-Hill, New York
Winsor T (2012) Independent review of police officer and staff remuneration and conditions. Final report submitted to the home office. Her Majesty's Printing Office, London

Role and Function of the Police Manager

Kenneth J. Peak
Department of Criminal Justice, University
of Nevada, Reno, Reno, NV, USA

Overview

If, as Shakespeare (1597) wrote in "Henry IV," it is true that "Uneasy lies the head that wears a crown," then certainly the nature of policing in general has never allowed police leaders the luxury of being able to "lie easy." The police profession, like the rest of the nation, is undergoing radical changes in leadership and supervision.

Furthermore, practices that worked for police managers of the past will no longer suffice today. The autocratic, control oriented police leadership methods of the past are changing to the empowerment and participative management styles; the paramilitary command and control operations used even a few years ago do not work well with today's officers, who respond instead to intellectual challenges, stimulation, and motivation. More specifically, past expectations and measures of the police function were quite different, often requiring that officers essentially perform like human pinballs – reactively hurrying from call to call, exiting the patrol vehicle at the scene and taking the report, and then returning to the patrol car to resume patrol (seldom if ever accomplishing anything long-term regarding crime control); furthermore, officers were typically evaluated by their police leaders in terms of numbers: arrests, traffic citations, response times, and even the numbers of miles driven during the duty shift, all of which were indicators of productivity. Officers were not encouraged to visit with citizens to obtain feedback about neighborhood crime and disorder; in fact, they were expected by their leadership, as stated by police expert Chris Braiden (1993), to "park their brains at the door of the stationhouse, and follow orders like a robot." Many of today's police executives can well remember these facets of policing that prevailed during that era, when they were young officers.

Adding to the historical challenges of police leaders are the contemporary economic woes across the USA – perhaps affected more than ever before by the financial contractions and upheavals in foreign venues – which have led to unprecedented municipal bankruptcies, police layoffs, and even a lack of job security for police personnel. Therefore, undoubtedly some of the most important and challenging positions in our society and times involve police leadership.

This essay examines the roles and functions of those leaders, including administrators, middle managers, and supervisors. Leadership is the heart and soul of any organization, and the idea of leadership has been with us for a long time. Yet leadership is still, and will always be something of an amorphous concept, dependent on the situation and the motivations and abilities of the people involved. There are theories that can guide leaders, but for some, good leadership may be a bit like the US Supreme Court's attempts at defining obscenity – summarized by Justice Stewart when he said: "I know it when I see it." This essay will thus serve to bring this concept into clearer relief.

Definitions of Key Terms

Although the terms *administration*, *manager/middle management*, and *supervisor* are often used synonymously, each is a unique concept that occasionally overlaps with the others. *Administration* encompasses both management and supervision. Administration is a process whereby a group of people are organized and directed toward achievement of the group's objective. Administration focuses on the overall organization and its mission and its relationship with other organizations and groups external to it. Administrators are often concerned with the department's direction and its policies and with ensuring that the department has the resources to fulfill its community's expectations. Police administrators generally include the chief, assistant chiefs, and high-ranking staff who support the chief in administering the department.

Middle management, which is also a part of administration, is most closely associated with the day-to-day operations of the various elements within the organization. For example, most police departments have a variety of operational units such as patrol, criminal investigation, traffic, gang enforcement, domestic violence, or community relations. Each of these units is run by someone who is most aptly described as a manager. In most cases, these managers are captains or lieutenants. These managers ensure that their units fulfill their departmental mission and work closely with other units to ensure that conflict or problems do not develop. They also attend to planning, budgeting, and human resource or personnel needs to ensure that the unit is adequately prepared to carry out its responsibilities.

Supervision involves the direction of officers and civilians in their day-to-day activities, often on a one-to-one basis. Often termed "first-line supervisors" (and typically being sergeants in a police or prison organization), these individuals ensure that subordinate officers adhere to departmental policies, complete tasks correctly and on a timely basis, and interact with the public in a professional manner. Supervisors often observe their subordinates completing assignments and sometimes take charge of situations, especially when a deployment of a large number of officers is needed. They also work closely with managers to ensure that officers' activities are consistent with the unit's mission and objectives.

This essay focuses more on the *administrative* level, but will briefly discuss the roles of middle managers and supervisors as well.

Applicable Theories

It has been said that behind every good practice is a good theory. Certainly any number of theories could be offered to inform and guide the actions of police leaders, explain how they should manage and motivate their personnel, and so on. This section discusses three applicable theories underlying police leadership: Katz's three skills of chief executive officers, the managerial grid, and situational leadership.

Early ideas about leadership tended to take a simplistic view, assuming that leadership was simply a matter of birth: the so-called "Great Man" theory of leadership, where leaders were born into leadership positions (e.g., monarchs). This definition is inadequate, however, because leadership today involves much more than one's birthright, and is a complex undertaking as well.

There is also a long-standing debate concerning whether one's ability to manage is an art or a science – that is, whether certain people are "born with" or predisposed to having the qualities necessary for being effective managers; or, conversely, whether management is an "acquired" ability or science because people can be taught to be effective managers. Many people believe management contains the elements of both, as it requires both managerial skill that is personal (art), as well as a systematized body of knowledge (science) for preparing, organizing, and directing subordinates to accomplish the goals of the organization.

Katz's Three Essential Skills

Robert Katz (1974) identified three essential skills that leaders should possess: technical, human, and conceptual. Each of these skills, when performed effectively, results in the achievement of objectives and goals, which is the primary thrust of organization and management. *Technical* skills are those a leader needs to ensure that specific tasks are performed correctly. They are based on proven knowledge, procedures, or techniques. Technical skills may involve knowledge in areas such as high-risk tactics, law, and criminal procedures. The police sergeant usually depends on training and departmental policies for technical knowledge. The areas in which leaders need technical skills include computer applications, budgeting, strategic planning, labor relations, public relations, and human resources management. They must also have knowledge of the technical skills required for the successful completion of tasks that are within his or her command. *Human* skills involve working with people and include being thoroughly familiar with what motivates employees and how to utilize group processes. Katz visualized human skills as including "the executive's ability to work effectively as a group member and to build cooperative effort within the team he leads. Katz added that the human relations skill involves tolerance of ambiguity and empathy. Tolerance of ambiguity means that the manager is able to handle problems when insufficient information precludes making a totally informed decision. Empathy is the ability to put oneself in another's place or to understand another's plight. *Conceptual* skills involve coordinating and integrating all the activities and interests of the organization toward a common objective. Katz considered such skills to include "an ability to translate knowledge into action" and emphasized that these skills can be taught to actual and prospective leaders. Thus, good leaders are not simply born but can be trained to assume their responsibilities.

All three of these skills are present in varying degrees for each level of leadership. As one moves up the hierarchy, conceptual skills become more important and technical skills less important. The common denominator for all levels of leadership, however, is *human* skills. In today's unionized and litigious environment, it is inconceivable that an administrator could neglect the human skills.

The Managerial Grid

The managerial grid, developed by Robert R. Blake and Jane S. Mouton (1962), has two dimensions: (1) concern for production and performance, and (2) concern for people (e.g., human relations, empathy). View the former, production, as being placed on a horizontal axis and the latter, concern for people, as being placed on a vertical axis, with both numbered from 1 to 9 (1 meaning low concern, to 9, indicating a high concern). The way in which a person combines these two dimensions will determine one's leadership style in terms of one of the five principal styles identified on the grid.

For example, one's management style that is located in the lower-left-hand corner of the grid would represent a minimal concern for task or service and a minimal concern for people. Conversely, one whose style is located at the lower-right-hand corner of the grid identifies a style having a primary concern for the task or output and a minimal concern for people. The upper-left-hand corner is often referred to as "country club management," with minimum effort given to output or task. The upper-right would indicate high concern for both people and production – and thus a team management approach of mutual respect and trust. In the center – a "middle-of-the-road" style – the leader has a more compromising, "be fair but firm" philosophy, providing a balance between output and people concerns.

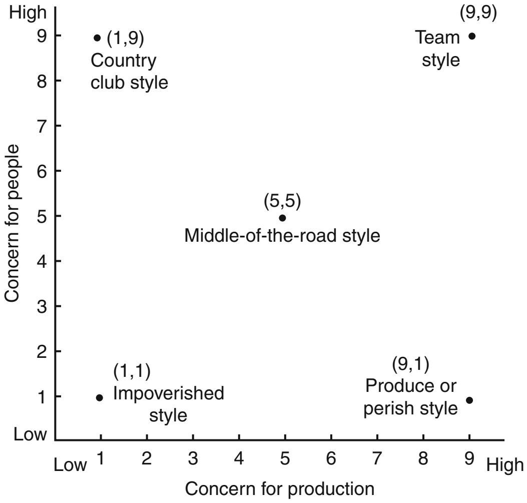

Situational Leadership

Not too different from the managerial grid is situational leadership theory, which recognizes that the workplace is a complex setting and subject to rapid changes; therefore, it is unlikely that a single leadership style would be adequate. Simply put, this theory argues that the leadership style to be used in an organization will depend on the situation – the subordinates' readiness, motivation to do their job, as well as their ability or competence to accomplish it. Therefore, as Hersey and Blanchard (1977) asserted, the leader's behavior is in relationship to followers' behavior, and the leader must evaluate subordinates' status in terms of their willingness or motivation, and ability or competence. To be effective, subordinates must be mature in their view of and approach to their job; maturity is defined by such attributes as setting high but attainable goals, taking responsibility for task accomplishment, and the education and/or experience of the individual. This ability can be easily viewed on a continuum: At one end are those subordinates who are perhaps new to the job and neither willing nor able to take responsibility for task accomplishment; at the other extreme are those who are, say, more experienced and confident, and thus both willing and able to take responsibility for task accomplishment. As the maturity level of followers grows, from one of being inexperienced to experienced and capable, this theory would maintain that the leader's style can and should change and adapt as well. For example, a police supervisor with a new subordinate who is nowhere ready to perform the demands of the job – his maturity level is low, and he is unable to do the job – would be most effective if employing an style of leadership that is more authoritarian, providing more close supervision, one-way communication supervision and telling what how, what, and when they will perform certain tasks. Conversely, an employee who is "mature," has the willingness and ability to perform the job, calls for a leadership style that is more hands-off, with the leader delegating more responsibility and allowing the subordinate more decision making authority. The onus of this style of leadership is that it is dependent on leaders to diagnose follower ability and then adjust their leadership style to the given situation. Doing so is often easier said than done.

Making a Police Organization "Great"

Another approach to successful leadership, in the form of a book that rapidly became desirable reading for many executives, was provided by Jim Collins (2001) in *Good to Great: Why Some Companies Make the Leap and Others Don't*, which sought to answer the compelling question: Can a good company become a great company and, if so, how? Collins and his team of assistants searched for companies that made a "leap to greatness," as defined by stock market performance and long-term success; they found 11 companies that met their criteria and spent more than 10 years studying what made them great. Following that, and acknowledging the growing interest in his book by non-business entities, Collins later published a monograph entitled *Good to Great and the Social Sectors* (Collins 2005) concerning how lessons concerning "greatness" could be modified to fit government agencies. Certainly, much of the success of great police organizations has to do with their leadership.

Collins coined the term *Level 5 leader* to describe the highest level of executive capabilities (Levels 1 through 4 are highly capable individual, contributing team member, competent manager, and effective leader). Level 5 executives are ambitious, but their ambition is directed first and foremost to the organization and its success, not to personal renown. Level 5 leaders, Collins stressed, are "fanatically driven, infected with an incurable need to produce results" (quoted in Wexler et al. 2007: 7).

Such leaders, Collins found, do not exhibit enormous egos; instead, they are self-effacing, quiet, reserved, even shy. Perhaps this has to do with the nature of their organizations. Unlike business executives, police leaders have to answer to the public; unions and civil service systems further inhibit their power. Therefore, Level 5 leadership in a police organization may involve a greater degree of legislative-type skills – relying heavily on persuasion, political currency, and shared interests to create the conditions for the right decisions to happen (Wexler et al. 2007).

Collins also used a bus metaphor when talking about the transformation from good to great, with leaders *first* getting the right people on the bus, and the wrong people off the bus and *then* figuring out where to drive it) (Wexler et al. 2007). Interestingly, Collins emphasized that the true challenges for a leader is not about assembling the right team who will be motivated (the right people will be self-motivated). Rather, putting them in the right seats on the bus is the critical challenge. When police executives are appointed or promoted, they inherit nearly all of their personnel, including poor performers who are unenthusiastic about the organization's vision and philosophy. Some of these people may be near retirement (about ready to "get off the bus"). Collins believes that picking the right people for the right jobs and getting the wrong people off the bus are critical (Wexler et al. 2007).

This is why performance evaluations are so critical. Unfortunately, however, many police departments still have not created evaluation tools that adequately reflect the work police do. The tendency is to measure what is easy to measure: orderliness (neatness, attendance, punctuality) and conformity to organizational rules and regulations. Until police agencies invest in valid and reliable instruments for measuring the real work of policing, it will remain very difficult to move the nonperformers out of the organization.

Perhaps the most difficult part of achieving greatness is *sustaining* that greatness. Police chiefs have notoriously short tenure in office. Therefore, in their world, some of Collins's principles may be particularly important – for example, finding Level 5 leaders who pay close attention to preparing for the next generation of leaders, giving managers authority to make key decisions, sending them to leadership academies and conferences, and encouraging them to think on their own and ask questions. This process is termed *succession planning* and is discussed below.

Police Administration Evolves: From Beginning to Present-Day

Policing Begins in the USA: The Political Era, 1840s to 1930s

Policing, as it exists today as a full-time career or occupation, began in America in 1844 in New York City, where it was deliberately placed under the control of the city government and city politicians. The American plan required that each ward in the city be a separate patrol district. The process for selecting officers was therefore different as well. City mayors chose the recruits from a list of names submitted by the aldermen and tax assessors of each ward; the mayor then submitted his choices to the city council for approval. Thus, most of the power over the police went to the ward aldermen, rather than to the police executive. Instead, the system allowed and even encouraged political patronage and rewards for friends (Johnson 1981).

Police administration during this era was thus largely a paper tiger; indeed, the New York law providing for the hiring of a chief of police gave no power to the chief in hiring the authorized 800 officers, nor to assign their duties or to fire them (Johnson 1981).

As other municipal police agencies developed in the East, three important issues confronted police administrators: whether the police should be in uniform, whether they should be armed, and whether they should use force. The wearing of a uniform was one of the basic principles of crime prevention – that police officers be visible, and that crime victims cound locate a police officer in a hurry. Furthermore, uniforms would make it difficult for officers to avoid their duties, since it would strip them of their anonymity. However, police officers themselves tended to prefer not to wear a uniform, and some even threatened to sue if compelled to don a uniform. To remedy the problem, New York City officials took advantage of the fact that their officers served four-year terms of office; when those terms expired in 1853, the city's police commissioners announced they would not rehire any officer who refused to wear a uniform. Thus in 1854 New York became the first American city with a uniformed police force.

A more serious issue was the carrying of arms. At stake was the personal safety of the officers and the citizens they served. Nearly everyone viewed an armed police force with considerable suspicion, however. Eventually, because the new police could hardly withstand attacks by armed assailants, it was agreed that an armed police force was unavoidable (Johnson 1981).

Eventually, the use of force, the third issue, would become necessary and commonplace for American officers. Indeed, the uncertainty about whether an offender was armed perpetuated the need for an officer to rely on physical prowess for survival on the streets. In New York, the police reform board was headed by Theodore Roosevelt. His appointed officers were criticized by disgruntled Tammanyites (corrupt New York City politicians) who favored the political patronage system of hiring police. Roosevelt's approach violated the American tradition of hiring local boys for local jobs.

Police administrators presided over agencies that had other problems as well. For example, from the onset of professional American policing, there was little or no lateral movement from one department to another; consequently, police departments soon became inbred, and tradition became the most important determinant of police behavior. A major teaching tool was the endless string of war stories the recruit heard, and the emphasis in most departments was on doing things as they had always been done. Innovation was frowned upon, and rookies were taught that things had to remain the same.

Police administrators of the late nineteenth century were also occupied with riots, strikes, parades, and fires. Labor disputes often meant long hours of extra duty for their officers, for which no extra pay was received. The police typically had little empathy or identification with strikers or strikebreakers; therefore the use of the baton to put down riots, known as the "baton charge," was not uncommon (Richardson 1974).

During the late nineteenth century, large cities gradually became more orderly places. The number of riots dropped. In the post–Civil War period, however, ethnic group conflict sometimes resulted in individual and group acts of violence and disorder, but daily urban life generally became more predictable and controlled. Ethnic and religious disputes were found in many police departments, however. In New York, the Irish officers controlled many hires and promotions, and there were still strong political influences at work. George Walling, an NYPD superintendent for eleven years, lamented after his retirement that he had largely been just a figurehead. Politics were played to such an extent that even non-ranking patrol officers used political backers to obtain promotions, desired assignments, and transfers (Richardson 1974).

Partly because of their initial closeness to politicians, police corruption also surfaced at this time. Corrupt officers wanted beats close to the gamblers, saloonkeepers, madams, and pimps – people who could not operate if the officers were "untouchable" or "100 % coppers" (Richardson 1974: 55–56). Political pull for corrupt officers could work for or against them; the officer who incurred the wrath of his superiors could be transferred to the outposts, where he would have no chance for financial advancement. While serving as police commissioner in New York, Theodore Roosevelt frequently made clandestine trips to the beats to check his officers. Any malingerers found in the saloons were summoned to headquarters in the morning (Richardson 1974).

The Reform Era and Attempts to End Political Patronage: 1930s to 1980

During the early twentieth century, reformers sought to reject political involvement of the police, and civil service systems were created to eliminate patronage and ward influences in hiring and firing police officers. In some cities, police chiefs did not permit their officers to live in the same beat they patrolled in order to isolate them as completely as possible from political influences. Police departments became one of the most autonomous agencies in urban government. Policing also became a matter viewed as best left to the discretion of police executives, and agencies became highly centralized, with the sole goal of controlling crime. Melding with the community was more eschewed, and any non-crime activities officers were required to perform was considered as "social work." Thus the reform era (also termed the professional era) of policing would soon be in full bloom.

Further removing policing from its earlier people-oriented, decentralized nature (and probably reducing job satisfaction in the process), the management-oriented scientific theory of administration was advocated by Frederick Taylor during the early twentieth century. Taylor first studied the work process, breaking down jobs to their basic steps and emphasizing time and motion studies, all with the goal of maximizing production. From this emphasis on production and unity of control flowed the notion that police officers were best managed by a hierarchical pyramid of control. Police leaders routinized and standardized police work; officers were to enforce laws and make arrests whenever they could, and discretion was limited as much as possible. When special problems arose, special units (for example, vice, juvenile, drugs, tactical) were created rather than assigning problems to patrol officers.

August Vollmer became a major force in the development of professional policing and its administration – but in a positive and prescient manner. In April 1905 at age 29, Vollmer became the town marshal in Berkeley, California (Vollmer 1933). Vollmer was the first chief to order his men to patrol on bicycles. He then persuaded the City Council to purchase a system of red lights, which hung at street intersection and served as an emergency notification system for police officers. He also began to question the suspects he arrested concerning their methods, and became convinced of the value of scientific knowledge in criminal investigation. Vollmer's most daring innovation came in 1908: the idea of a police school. The first formal training program for police officers in the country drew on the expertise of university professors as well as police officers (Peak 2012).

In 1916, Vollmer hired a professor of pharmacology and bacteriology to run the department's crime laboratory, and by 1917 the entire patrol force became the first to operate out of automobiles; in 1918, he began to hire college students as part-time officers and to administer a battery of tests to all applicants. Vollmer also advocated the idea that the police should function as social workers, doing more than merely arresting offenders while seeking to prevent crime (Peak 2012).

The Crime Fighter Image

The 1930s marked an important turning point in the history of police reform. O. W. Wilson emerged as the leading authority on police administration, and the crime fighter image gained popularity. Wilson, a student of Vollmer, followed in his mentor's footsteps by advocating efficiency within the police bureaucracy through scientific techniques. While chief of police in Wichita, Kansas, Wilson conducted the first systematic study of one-officer squad cars. A major contribution was the publication of Wilson's (1950) classic textbook, *Police Administration*, which set forth ideas regarding police deployment, discipline, and organizational structure. Wilson would later teach at UC Berkeley, and in 1947 he founded the first professional School of Criminology. Wilson became the principal architect of the police reform strategy.

Police as the "Thin Blue Line": Chief William H. Parker

The movement to transform the police into professional crime fighters perhaps found its staunchest champion in William H. Parker, who became the Los Angeles police chief in 1950. His greatest success, typical of the new professionalism, came in administrative reorganization. The command structure was simplified, and rigorous selection and training of personnel became a major characteristic of the LAPD – which became the model for reform across the nation (Johnson 1981). Parker conceived of the police as a thin blue line, protecting society from barbarism and Communist subversion. He viewed urban society as a jungle, needing the restraining hand of the police; only the law and law enforcement saved society from the horrors of anarchy. The police had to enforce the law without fear or favor – with few restrictions on police methods such as wiretaps and searches and seizures (Richardson 1974).

The Community Problem-Solving Era

Until the 1970s, there were few outside, scientific inquiries concerning police functions and methods, for two reasons. First was a tendency on the part of the police to resist outside scrutiny and research into their operations. Many police administrators perceived a threat to their career and to the image of the organization, as well as a concern about the legitimacy of the research itself. Second, few people in policing perceived a need to challenge traditional methods of operation, and police methods were not even open to debate.

However, the turmoil of the 1970s compelled the police to open their doors to the outside world and allow academics to study their functions. This turmoil – which centered on the raging Vietnam War, race relations, and equal opportunity – included not only riots, protests, and sit-ins, but also such brutal police actions as those that occurred in Chicago in 1968 during the Democratic National Convention; people were served a nightly television display of police using clubs and gas to control rioters. Furthermore, five national commissions were established during the 1960s and 1970s, some of which lamented the "lawlessness" of the police and their lack of professionalism

In the early 1970s it was suggested that the performance of patrol officers would improve more by using job redesign based on certain motivations. This suggestion later evolved into a concept known as "team policing," which sought to restructure police departments, improve police-community relations, enhance police officer morale, and facilitate change within the police organization. Its primary element was a decentralized neighborhood focus to the delivery of police services. Officers were to be generalists, trained to investigate crimes and basically attend to all of the problems in their area, with a team of officers being assigned to a particular neighborhood and responsible for all police services in that area – the community problem-solving era.

There were other developments for the police during the late 1970s and early 1980s. Foot patrol became more popular, and many jurisdictions even demanded it. Foot patrol was found to produce a significant increase in public satisfaction with police service, a reduction of perceived crime problems, and an increase in the perceived level of safety of the neighborhood (Police Foundation 1981).

In addition, research conducted during the 1970s suggested that *information* could help police improve their ability to deal with crime. These studies created new opportunities for the police to understand the increasing concerns of citizens' groups about disorder (e.g., gangs, prostitutes) and to work with citizens to do something about it. Police executives discovered that when they asked citizens about their priorities, citizens often provided useful information.

Simultaneously, problem-oriented policing was being tested in several cities. These studies found that patrol officers have the capacity to do problem solving successfully and can work with citizens and other agencies to solve problems when given more autonomy and training to analyze the underlying causes of problems and to find creative solutions.

Thus policing moved into the current community era, with a renewed emphasis on community collaboration for many police tasks and crime *prevention*. Administrators organize their agencies so as to be more decentralized, pushing decision making to the lower levels of the organization.

Local Police Administrators: Police Chiefs and Sheriffs

The chief of police (sometimes known as commissioner or superintendent) is generally considered to be one of the most influential and prestigious persons in local government. Indeed, people in this position often amass considerable power and influence in their jurisdiction. Mayors, city managers and administrators, members of the agency, labor organizations, citizens, special-interest groups, and the media all have differing role expectations of the chief of police that often conflict.

The mayor or city manager likely wants the chief of police to promote departmental efficiency, reduce crime, improve service, and so on. Others appreciate the chief who simply keeps morale high and citizens' complaints low. The mayor also expects the chief to communicate with city management about police-related issues and to be part of the city management team; to communicate city management's policies to police personnel; to establish agency policies, goals, and objectives and put them in writing; to develop an administrative system for managing people, equipment, and the budget in a professional and businesslike manner; to administer disciplinary action consistently and fairly when required; and to select personnel whose performance will ably and professionally promote the organization's objectives.

Members of the agency also have expectations of the chief executive: to be their advocate, supporting them when necessary and representing the agency's interests when dealing with judges and prosecutors who may be indifferent or hostile. Citizens tend to expect the chief of police to provide efficient and cost-effective police services while keeping crime and tax rates down (often an area of built-in conflict) and preventing corruption and illegal use of force. Special-interest groups expect the chief to advocate policy positions that they favor. Finally, the media expect the chief to cooperate fully with their efforts to obtain fast and complete information on crime.

Qualifications for the position of police chief vary widely, depending on the size of the agency and the region of the country. Level of education is commonly found to be an important consideration; today, many agencies require a college education plus several years of progressively responsible police management experience. Police chief executives also need several important management skills, perhaps chief among them being the ability to motivate and control personnel and to relate to the community.

Regarding their hiring, cities must often decide whether it would be better to promote someone from within the ranks or hire from outside (perhaps using the assessment center process, described below). Both approaches have advantages and disadvantages. For example, internal candidates would be better known to personnel within the agency and other agencies of criminal justice (e.g., prosecuting attorney) and possibly the governing board. Conversely, one who is brought in from the outside may have new knowledge, skills, and abilities learned and/or practiced at other city or county venues, less "baggage" than what one might accumulate in the home city or county, and so on.

The position of sheriff has a long tradition, rooted in the time of the Norman conquest of England (in 1066), and it played an important part in the early law enforcement activities of colonial America. Because of the diversity of sheriff's offices throughout the country, it is difficult to describe a typical sheriff's department; these offices run the gamut from the traditional, highly political, limited-service office to the modern, fairly nonpolitical, full-service police organization. It is possible, however, to list functions commonly associated with the sheriff's office: serving civil processes (e.g., divorce papers, liens, evictions, garnishments and attachments, extradition and transportation of prisoners); collecting certain taxes and conducting real estate sales for the county; performing routine order-maintenance duties by enforcing state statutes and county ordinances; performing traffic and criminal investigations; serving as bailiff of the courts; maintaining and operating the county jail.

Sheriffs are elected in all but two states (Rhode Island and Hawaii; note, however, that in some consolidated jurisdictions, such as Miami–Dade county, Florida, sheriffs are also appointed); thus, they tend to be aligned with a political party. As elected officials, sheriffs are important political figures and, in many rural areas, represent the most powerful political force in the county. As a result, sheriffs are far more independent than appointed municipal police chiefs, who can be removed from office by the mayors or city managers who appoint them. However, because they are elected, sheriffs receive considerable media scrutiny and are subject to state accountability processes.

Because of this electoral system, it is possible that the only qualification for the office is the ability to get votes. In some areas of the country, the sheriff's term of office is limited to one 2-year term at a time (a sheriff cannot succeed himself or herself); thus, the office has been known to be rotated between the sheriff and undersheriff. In most counties, however, the sheriff has a 4-year term of office and can be reelected.

The sheriff enjoys no tenure guarantee, although one study found that sheriffs (averaging 6.7 years in office) had longer tenure in office than chiefs of police (5.4 years). The politicization of the office of sheriff can result in high turnover rates of personnel who do not have civil service protection. The uncertainty concerning tenure is not conducive to long-range (strategic) planning. Largely as a result of the political nature of the office, sheriffs tend to be older, less likely to have been promoted through the ranks of the agency, and less likely to be college graduates and to have specialized training than police chiefs. Research has also found that sheriffs in small agencies have more difficulty with organizational problems (field activities, budget management), and that sheriffs in large agencies find dealing with local officials and planning and evaluation to be more troublesome (Hayes 2001).

What Police Executives Do: Essential Functions

Finally, to better understand what police leaders do on a more day-to-day basis, adapting Henry Mintzberg's (1975) general model of chief executive officers (CEOs) to police work is instructive. Included in the model are three major roles or categories (interpersonal, informational, decision maker) and several subcategories, which allows for grouping and delineating the police executives' occupational roles.

The Interpersonal Role

The interpersonal role includes *figurehead*, *leadership*, and *liaison* duties. As a figurehead, the CEO performs various ceremonial functions. Examples include riding in parades and attending other civic events; speaking before school and university classes and civic organizations; meeting with visiting officials and dignitaries; attending academy graduation and swearing-in ceremonies and some weddings and funerals; and visiting injured officers in the hospital. The CEO performs these duties simply because of his or her position within the organization; the duties come with being a figurehead.

The leadership function requires the CEO to motivate and coordinate workers while resolving different goals and needs within the department and the community. For example, a chief or sheriff may have to urge the governing board to enact a code or ordinance that, whether popular or not, is in the best interest of the jurisdiction; CEOs also provide leadership in such matters as bond issues (e.g., to raise money for more officers or to build new stationhouse facilities) and advise the governing body on the effects of proposed ordinances.

The role as liaison is performed when the CEO of a police organization interacts with other organizations and coordinates work flows. Here, the police chief or sheriff might meet informally each month to discuss common problems and strategies. They might also do the same with representatives of the courts, the juvenile system, and other criminal justice agencies.

The Informational Role

The informational role is composed of *monitoring/inspecting*, *dissemination*, and *spokesperson* functions. Monitoring/inspecting involve the CEO constantly looking at the workings of the department to ensure that things are operating smoothly by "roaming the ship," sitting in on staff meetings, and generally being a part of efforts that discuss agency problems and functions.

The dissemination tasks involve getting information to members of the department. This may include memorandums, special orders, general orders, and policies. The spokesperson function is related, but it is more focused on getting information to the news media. News organizations are in a competitive field in which scoops and deadlines and the public's right to know are all-important. Still, the media must understand those occasions when a criminal investigation can be seriously affected by premature or overblown coverage.

The Decision-Maker Role

As a decision maker, the police executive serves as *an entrepreneur*, *a disturbance handler*, a *resource allocator*, and a *negotiator*. As entrepreneur, the CEO must sell ideas to the governing board or the department. Ideas might include new computers or a new communications system, a policing strategy (such as community-oriented policing and problem solving), or different work methods, all of which are intended to improve the organization.

As a resource allocator, the CEO must have a clear idea of the budget and what the priorities are and must listen to citizen complaints and act accordingly. For example, ongoing complaints of motorists speeding in a specific area will result in a shifting of patrol resources to that area or neighborhood. Finally, as a negotiator, the police manager resolves employee grievances and sits as a member of the negotiating team for labor relations.

Middle Managers and Supervisors

Captains and Lieutenants

Leonhard Fuld, one of the early progressive police administration researchers, said in 1909 that the captain is one of the most important officers in the organization. Fuld believed that the position had two broad duties – policing and administration. The captain was held responsible for preserving the public peace and protecting life and property within the precinct. Fuld defined the captain's administrative duties as being of three kinds: clerical, janitorial, and supervisory (1909).

Although every ranking officer in the police department exercises some managerial skills and duties, here we are concerned with the managers to whom first-line managers report, for they generally are unit commanders. In a mid-sized or large police agency, a patrol shift or watch may be commanded by a captain. The lieutenants, reporting to the captain in the normal hierarchy, may assist the captain in running the shift, but when there is a shortage of sergeants as a result of vacations or retirements, the lieutenant may assume the duties of a first-line manager. In some respects, the lieutenant's position in some departments is a training ground for future unit commanders (the rank of captain or higher).

In a medium-sized or large police organization, captains devote a substantial amount of time coordinating their units' activities with those of other units and overseeing the operation of their units. At the same time, however, captains also perform supervisory duties (e.g., writing reports, reviewing incoming complaints and subordinates' reports, ensuring that subordinates comply with general and special orders, completing employees' performance evaluations, serving on promotional and hiring boards). Whereas a sergeant or lieutenant may be supervising individual officers, a captain is more concerned with unit activities and the overall performance of the officers under his or her command. Furthermore, as can be seen from this list, the unit commander functions to some extent like a police chief or sheriff, but on a smaller scale. The chief performs these functions for the total department, while the unit commander is concerned with only one unit (Peak et al. 2010).

Lieutenants, as the companion middle-manager to captains, perform some tasks that are purely administrative in nature but typically involve more of the day to day activities that are related to patrol operations: preparing duty rosters, conducting roll call, preparing various reports, and maintaining time sheets. At times, lieutenants may also perform duties that resemble those that are supervisory in nature, such as assisting in supervising or directing the activities of the unit, ensuring that departmental and governmental policies are followed, reviewing the work of individuals or groups in the section, responding to field calls that require an on-scene commander, and reviewing various reports. These functions include overseeing officers and sergeants to ensure that different tasks are completed.

First-Line Supervisors: The Patrol Sergeant

As indicated above, another very important level of leadership – indeed, to the majority of police, who are rank-and-file officers, the *most* important leadership level – is that of patrol sergeant. Officers are often told that it is best to rotate into different assignments before testing for sergeant to gain exposure to a variety of police functions and managers. The promotional system, then, favors well-rounded officers. As with the chief executives' hiring process, an assessment center will provide candidates with valuable training and testing experience. Other factors that might come into play as part of the promotional process include education and training, years of experience, supervisory ratings, psychological evaluations, and departmental commendations (Peak 2012).

Police administrators understand that a good patrol officer is not automatically a good supervisor. Because supervisors are promoted from within the ranks, they are often placed in charge of their friends and peers. Longstanding relationships are put under stress when a new sergeant suddenly has official authority over former equals. Leniency or preferential treatment often is expected of new sergeants by their former peers. A new supervisor therefore must go through a transitional phase to learn how to exercise command and get cooperation from subordinates.

The supervisor's role, put simply, is to get his or her subordinates to do their very best. This role involves a host of actions, including communicating, motivating, leading, team building, training, appraising, counseling, and disciplining. They are in close proximity to their subordinates, the patrol officers, and thus can determine each subordinate's strengths and weaknesses, and ensure that each officer's conduct coincides with the organization's mission, values, goals, and objectives. Supervisors must be good communicators, disseminating information to subordinates, ensuring that general and special orders from captains and lieutenants are followed, reviewing all officers' offense and traffic crash reports, listen to and assist officers with personal or professional problems, respond to officers' calls for service, keep superiors apprised of ongoing situations and problems, and be able to interpret laws and policies for subordinates.

Obtaining the Best: The Assessment Center

In order to obtain the most capable people for police chief executive positions (as well as for middle-management and even supervisory positions), the assessment center method has proved to be an efficacious means of hiring and promoting personnel. (*Note*: Sheriffs are normally elected, not hired or promoted into their position; thus, the assessment center approach is of little use for that position.)

After developing a job description which lists responsibilities and skills for the position to be filled, then each candidate's abilities and skill levels should be evaluated, using one or more of the following techniques: interviews; psychological tests; in-basket exercises; management tasks; group discussions; role-playing exercises such as simulations of interviews with subordinates, the public, and news media; fact-finding exercises; oral presentation exercises; and written communications exercises.

During each exercise, several assessors or raters analyze each candidate's performance and record some type of evaluation; when the assessment center process ends, each rater submits his or her rating information to the person making the hiring or promotion decision. Typically selected because they have held the position for which candidates are now vying, assessors must not only know the types of problems and duties incumbent in the position, but also should be keen observers of human behavior.

Although obviously much more expensive and labor-intensive to perform, the assessment center can serve to avoid many of the errors of alternative methods of selecting police leaders and is highly recommended for jurisdictions who literally wish to "test" their candidates and thus make the best choice.

Evaluating Police Administrators

How good is a municipal police chief or county sheriff? The answer to that question probably depends on whom one asks. To some, the police chief executive's arrival is the best thing that ever happened to the area; to others, he or she is the worst thing. One reason for the differing viewpoints is that people's criteria for evaluating their chief or sheriff will vary. Therefore, rating the CEO's performance in this complex job involves several potential "traps." CEOs are human; they do make mistakes, but citizen perceptions of them and their performance are often wrong. Therefore, it is difficult to assess accurately how the CEO is performing; no litmus test or simple fill-in-the-boxes exercise exists. Several broad, general guidelines can be applied, however, when evaluating these CEOs.

Inappropriate Criteria

Some criteria that are generally inappropriate for evaluating CEOs include such considerations as: *popularity in the department* (the longer the CEO's tenure, the greater the likelihood that people both inside and outside the department will have something to gripe about); d*epartmental morale* (morale is a fragile thing, and most organizations have chronic and notorious complainers; thus, the morale of a department is not necessarily the result of the CEOs leadership); *a rising crime rate* (a variety of social and economic factors influence community crime levels; therefore, to blame the chief executive for the crime rate is inappropriate if not unfair – given, however, that the chief executive should know how and when to develop programs and strategies to combat rising crime rates); *single incident or issue* (a police beating or shooting of a citizen can cause a loud hue and cry and even calls for the chief's head; however, more often a multidimensional review of the chief's total performance should be made, and the firing of the chief does not ensure that the basic problem causing the upheaval will be eliminated).

Appropriate Criteria

A police executive's evaluation should focus only on qualities, characteristics, and behavior required to perform the job. Honesty and integrity are probably the chief's most important qualities, followed by leadership effectiveness. The direction of the department, its commitment to professional and ethical standards, and its basic values emanate from the chief's leadership role. The chief's ability to motivate people without shoving them and to inspire people both inside and outside the department should also be considered.

Any assessment of the chief executive must consider leadership effectiveness. The chief's ability to be a good follower and take a subordinate position to the mayor and city manager/council is also important. It is important that the chief know where the lines are drawn, when it is time to quit leading and start following.

CEOs manage a significant amount of resources – human and financial, equipment, and physical facilities. CEOs must make timely and responsible decisions, as well as realize the impact and legal ramifications of those decisions. The chief executive must also delegate authority and decision making to others rather than taking on too much, which often results in a sluggish operation. Finally, a chief executive must be innovative and creative in thinking. Today's CEO must be willing to keep up with the challenges facing law enforcement and need to move the agency into a new era of research, experimentation and risk-taking – to become a "learning organization."

Several other areas of the chief executive's performance must also be evaluated. The first relates to departmental values: do the department's values reflect a commitment to the rights of individuals and its employees? Does the CEO manage force carefully and aggressively to ensure that the public is not subjected to unlawful or unwarranted physical force by police? Does the CEO require the agency to have clear policies, rules, and procedures to guide employee? Are those directives are constantly updated and enforced fairly and consistently? Has the CEO introduced or implemented crime strategies to make the community safer?

The labor climate of the organization is an especially difficult area for today's police executive; civil service, unions, appeal boards, the courts, and the changing profile of today's police officer have made managing labor relations a constant challenge.

Finally, the CEO's understanding of where the department stands with the community should be evaluated. No other sector of government in our society has more frequent and direct contact with the public than the police.

The Future of Police Management: Succession Planning

Soon the administration, management, and supervision of police agencies could be in a crisis stage unless measures are taken in the near future to prepare for what is coming: the current aging, turnover, and retirement of baby boomers and other generational employees. Today an essential part of every chief's job is to prepare colleagues in the organization for the next advance in their careers; indeed, today the mark of a good leader is the ability to ensure a ready supply of capable leaders for the future.

Thorough preparation of successors can help a chief establish an important legacy – one that will sustain the improvements and progress that have been made, offer opportunities for mentoring, and instill the importance of the organization's history.

Chiefs need to take a long view and look at succession planning and leadership development as a continuous process that changes the organizational culture. To provide the ongoing supply of talent needed to meet organizational needs, chiefs should use recruitment, development tools (such as job coaching, mentoring, understudy, job rotation, lateral moves, higher-level exposure, acting assignments, and instructing), and career planning (Davis and Hanson 2006).

Organizations may already have a sufficient pipeline of strong leaders – people who are competent in handling the ground-level, tactical operations but who are not trained in how to look at the big picture. While not everyone can excel at both levels of execution, efforts must definitely be made to help prospective leaders develop a broader vision – as one author put it, prepare employees to take on broader roles and "escape the silos" (Ready 2004: 93). To develop police leaders who do not see the world in zero-sum terms but instead can appreciate the bigger picture requires people to leave their comfort zones and to offer them challenging assignments in different roles (Ready 2004).

A number of excellent police promotional academies and management institutes exist for developing chief executives, middle managers, and supervisors in the kinds of desired skills described above. In addition, *agencies* can provide skill development opportunities by having those persons with leadership potential do such things as plan an event, write a training bulletin, update policies or procedures, conduct training and research, write a proposal or grant, counsel peers, become a mentor, write contingency plans, and so forth. Meanwhile, the *individual* can lay plans for the future through such activities as doing academic coursework, participating in and leading civic events, attending voluntary conferences and training sessions, reading the relevant literature, studying national and local reports, guest lecturing in college or academy classes, engaging in research, and so on (Michelson 2006).

With today's leadership theory holding that law enforcement executives should adopt a participative management style, also known as *democratic leadership*, except when emergencies arise, an autocratic management style that includes public criticism can cause resentment among subordinates rather than a sense of teamwork or a spirit of cooperation. This style may also inhibit the development of future leaders and undermine the cooperative leadership process. Leadership is learned behavior, and new leaders can be developed through properly designed leadership experiences (Hernez-Broome and Hughes 2004).

Related Entries

- ▶ British Police
- ▶ Community Policing
- ▶ Compstat
- ▶ Conceptualizing of Police
- ▶ Control of Police Misconduct
- ▶ Democratic Policing
- ▶ Managing Innovation of Policing
- ▶ Measuring Police Performance
- ▶ Police Leadership Styles
- ▶ Police Promotional Practices

Recommended Reading and References

Blake RR, Mouton JS (1962) The developing revolution in management practices. J Am Soc Train Dir 16:29–52

Collins J (2001) Good to great: why some companies make the leap. . .and others don't. HarperCollins, New York

Collins J (2005) Good to great and the social sectors: a monograph to accompany good to great. HarperCollins, New York

Davis E, Hanson E (2006) Succession planning: mentoring future leaders. Police Executive Research Forum, Washington, DC

Fuld FL (1909) Police administration. G. P. Putnam's Sons, New York

Hayes C (2001) The office of sheriff in the United States. Prison J 74:50–54

Hernez-Broome G, Hughes RL (2004) Leadership development: past, present, and future. Hum Resour Plan 27(1):24–32

Johnson DR (1981) American law enforcement history. Forum Press, St. Louis

Katz R (1974) Skills of an effective administrator. Harv Bus Rev 52(5):90–102

Michelson R (2006) Succession planning for police leadership. Police Chief 16–22

Mintzberg H (1975) The manager's job: Folklore and fact. Harv Bus Rev 53(4):49–61

Peak KJ (2012) Policing America: challenges and best practices, 7th edn. Prentice Hall, Upper Saddle River

Peak KJ, Gaines LK, Glensor RW (2010) Police supervision and management: in an era of community policing, 3rd edn. Prentice Hall, Upper Saddle River

Police Foundation (1981) The Newark foot patrol experiment. Police Foundation, Washington, DC

Ready DA (2004) How to grow leaders. Harv Bus Rev 82(12):92–100

Richardson JF (1974) Urban policing in the United States. Kennikat Press, London

Shakespeare W (1597) Henry IV, Part II, Act 3

Vollmer A (1933) Police progress in the past twenty-five years. J Crim Law Criminol 24:161–75

Wexler C, Wycoff MA, Fischer C (2007) Good to great: application of business management principles in the public sector. Office of Community Oriented Policing Services, Washington, DC, Police Executive Research Forum, 5

Wilson O (1950) Police administration. McGraw-Hill, New York

Routine Activities Approach

Jonathan Allen Kringen and Marcus Felson
School of Criminal Justice, Texas State
University-San Marcos, San Marcos, TX, USA

Overview

The routine activity approach is a theoretical perspective for describing and explaining how crime rates vary over time and space. The approach applies to variations and changes in both large and small areas, over both short and long stretches of time. The approach differs from many other criminological theories because it looks beyond individual offenders to focus on crime events and emphasizes how these events draw upon a much larger set of legal activities.

Basic Theory Described as Stated in 1979

The routine activity approach began by studying "direct-contact predatory crimes," namely, those offenses in which one person directly attacks another person or takes the property of another. Crime events normally require a physical convergence of three minimal elements: (1) a likely offender, (2) a suitable target, and (3) the absence of a capable guardian (Cohen and Felson 1979). Routine noncrime activities are central for distributing these elements across space and through time resulting in circumstances in which offenders converge with targets while guardians are absent. A guardian is not normally a police officer or security guard, but rather a citizen or employee whose proximity can discourage a crime from occurring.

While the majority of theoretical criminology focuses specifically on criminal motivation, the routine activities approach focuses on crime incidents, treating criminality as but one feature within a larger system of activities. Without denying that offender motivation is necessary, it is far from sufficient for understanding crime; accordingly, the routine activities approach focuses upon criminal events themselves and the prerequisites required for their occurrence. Variation in these prerequisites can explain changes in crime rates without requiring any changes in the structural or individual conditions that motivate offenders.

The routine activity approach argues that the absence of any of the three elements is sufficient to effectively thwart crime. Because the three elements vary as a result of the normal day to day activities of both criminal and noncriminal individuals, the effects of routine activities can be difficult to recognize. The idea of a guardian provides a straightforward example. Cohen and Felson (1979) note that because guardianship is implicit in everyday life and is often marked by the absence of violations, it is easy to overlook. Whereas police actions are widely analyzed, guardianship by ordinary individuals as they progress through their routine activities was largely neglected in sociological research on crime. However, recent research has begun to analyze the effects of guardians on crime using both experimental and quasi-experimental methods (Hollis-Peel et al. 2011). Because guardianship by ordinary individuals links together seemingly unrelated social roles and relationships to the occurrence of illegal acts, it represents a phenomena not readily explained by sociological theory. However, this phenomenon may be extremely important in understanding crime.

Massive Crime Wave and Inapplicability of Earlier Theoretical Predictions

The period between 1960 and 1975 was characterized by marked increases in reported rates of crime. Controlling for population size, robbery rates had increased by 263 %, aggravated assault by 164 %, forcible rape by 174 %, and homicide by 188 %. Nonviolent offenses had experienced similar trends with burglary up by 200 % and auto theft increasing by 150 %. These trends represented a sociological paradox, as the conditions generally assumed by sociologists to cause crime had actually improved over the same period. High school completion had risen, unemployment had fallen, median family income had grown, and the number of people living below the poverty line had diminished. Other important predictors of crime had improved as well.

The proportion of the population living in the inner city and the percent of foreign born individuals residing in high-crime areas had both declined. Although the percent black had increased slightly during that period, as did the percent of the population widowed or divorced, these changes were in no way sufficient to explain the crime rate trends. Even the influx of the Post World War II baby boom into adolescence could only account for a small part of the crime wave that actually occurred. Although cross-sectional crime analysis seemed to support conventional theories, and although such theories may be useful for explaining crime trends for particular subgroups or specific communities during certain time periods, they could not explain these general crime trends over longer periods. Incorporating microlevel assumptions about individual behavior and macrolevel ideas about how crime fits into community life that emerge from human ecology, the routine activity approach offered an alternative and more effective theoretical framework.

Human Ecology Origins

Amos Hawley (1950) propagated a theory of community life, treating communities not simply as units of territory but also as tangles of symbiotic and competitive relationships. These relationships vary as human activities are preformed over both space and time. Hawley identified three important temporal components of community structure that are especially important for the routine activity approach: tempo, rhythm, and timing. Tempo refers to the number of events occurring per unit of time, as exemplified by a crime rate. A rhythm is the regular periodicity with which an event occurs; thus, crime has a daily or hourly rhythm, a weekly rhythm, and a monthly rhythm. For example, in the course of a day, certain crimes go up in the afternoon near schools and in the evening in entertainment districts, reflecting the combined significance of time and space. Timing refers to the coordination or coincidence of different activities and is extremely important for the routine activity approach. The afternoon upsurge in crime corresponds to the daily rhythm of schooling and when school releases likely offenders and suitable targets for crime.

Hawley’s central emphasis was on sustenance activities, upon which all other activities depend. Thus, the structure of work in community life influences leisure and other activities that must draw resources from it. More generally, the structure of illegal activities depends upon the organization of everyday sustenance activities – including work, school, and family life. Accordingly, the rhythms of daily life include the daily journeys to work and school, affecting the locations of offenders and crime targets and their spatiotemporal convergences. Timing leads to questions about the interdependence between crime and noncrime activities, as well as the comovements of likely offenders and suitable crime targets in the course of a day.

Although the routine activity approach considered many temporal features of crime, it emphasized activities within each day and their impact on the specific spatiotemporal convergence of crime’s minimal elements, with offenders finding or stumbling upon targets with guardians absent. The approach supplemented Hawley’s work by including a technological impetus for change as envisioned by William F. Ogburn (1922). As a result, the routine activity approach replaced traditional cultural and social explanations of crime with an approach that brought technological change to the forefront thereby shifting crime rate analysis away from race, divorce, education, and other independent variables which could not in practice explain the crime wave during the 1960s and 1970s.

Central Social Changes Relevant to Crime

Human ecology led to the question, “Where does crime draw its sustenance?” The answer was found in the material goods produced by society and the tangible activities by which society sustained itself. A proliferation of lightweight consumer goods, high in value and low in weight, provided vast new opportunities to commit crime. The new targets for property crime emerged with the technology of plastics and transistors bringing down the weight of electronic goods. Thus, the routine activity approach looked at the increased value per pound for small television sets and the dramatic increases in things to steal. The approach looked also at the female labor force participation rate and the dispersion of activities away from family and household, these making it possible for offenders to find targets away from the security of home or to find people home alone. Intrinsic to the routine activity approach was a detailed statistical review of changes in daily life and documentation of dramatic new exposures to risk of property and violent crime victimization. The development of the victimization survey provided new data about crime events that made the routine activity approach possible, along with some related theoretical ideas developed by others.

Key Links, Developments, and Controversies

Link to Lifestyle Theory

At about the same time as the routine activity approach, similar ideas emerged in the Netherlands (Van Dijk and Steinmetz 1983) as well as Great Britain (Mayhew et al. 1976). However, the most well-known parallel work was the lifestyle theory of Hindelang et al. (1978). Lifestyle theory suggests that differential exposure to offenders results in differential rates of victimization (Fattah 1991). Therefore, understanding variance in exposure is paramount to understanding victimization. Exposure to offenders varies with structural factors commonly associated with sociological theories of crime (e.g., age, race, and neighborhood of residence). However, exposure also varies as a function of the victim's personal lifestyle. The work and leisure activities of an individual may increase exposure to potential offenders, enhancing risk of victimization. This implies that risk of victimization may be reduced by modifying such exposures.

Although routine activity theory and lifestyle theory are often viewed as identical, that is not the case. On the one hand, routine activity ideas emphasize the criminogenic effects of everyday routines, such as work, school, and family life. In contrast, lifestyle theory gives more attention to personal lifestyle choices in leisure life. The two theories are not, however, completely distinct, since the former includes lifestyles and the latter includes work. The difference is a matter of emphasis, yet it suggests that the two theories should not be treated as one and the same.

Link to Situational Prevention

Originally, the routine activity approach did not offer policy prescriptions. Instead, it simply stated that high crime rates might be endemic in a modern society given the crime opportunities it produces. When Clarke and Felson met in London in 1981, the former suggested that if more crime opportunity makes more crime, it follows that less crime opportunity makes less crime. From that point on, routine activity theory and situational prevention began to interpenetrate, as increasingly evident in subsequent publications. Thus, the routine activity approach developed policy relevance, culminating in the joint publication Opportunity Makes the Thief (Felson and Clarke 1998).

Situational crime prevention is the application of "opportunity-reducing measures that (1) are directed as highly specific forms of crime, (2) involve the management, design, or manipulation of the immediate environment in as systematic and permanent way as possible, (3) make crime more difficult or risky, or less rewarding and excusable as judged by a wide range of offenders" (Clarke 1997, p. 4). Situational prevention advocates surveillance and target hardening through the management and manipulation of the environment which equates to altering the suitability of targets and the presence of guardians. Thus, situational prevention adopts aspects of the routing activities approach.

Link to Other Approaches and the Emergence of Crime Science

The routine activity approach is increasingly linked to other approaches. Situational crime prevention, which focuses on the circumstances that give rise to specific kinds of crime, seeks to make crime less attractive to offenders by manipulating the situational factors favoring crime (Clarke 1997). Crime prevention through environmental design (CPTED) takes a similar tack by focusing on environmental conditions and the opportunities that they offer potential offenders (Zahm 2007). Victimology shifts the focus away from criminal motivation and instead focuses on victim-offender interactions and the social and personal settings where crimes take place (Fattah 2000). Environmental criminology recognizes crime as a function of law, offender motivation, and target characteristics superimposed on an unseen environmental backcloth where the rhythms of life interact with environmental constraints (Brantingham and Brantingham 1993).

While each of these approaches considers unique aspects of crime, they share the common thread of explaining crime rather than trying to explain criminal motivation. After focusing attention on crime itself, attention can be given to the problem of reducing crime rather than simply dealing with the criminal (Gottfredson and Clarke 1990). This results in an emerging area of study distinct from mainstream criminology known simply as crime science (Laycock 2003).

Incorporating Social Control

Initially, the routine activities approach simply assumed offender motivation. However, later work addressed this limitation by incorporating the idea of informal social control (Felson 1986). The idea of informal social control adopted within the routine activities perspective represents a simplification of Hirschi's (1969) control theory. While Hirschi conceived informal social control as involving four specific elements (commitments, attachments, involvements, and beliefs), the routine activities perspective considers a simplified single effect of all four. Felson (1986) described all four of Hirschi's elements using a single word: handle. Crime is reduced when society gains a handle on potential offenders. This handle on individuals is formed when society establishes a social bond with the individuals. Because individuals might lose this bond if others become upset with their behavior, society is able to exert informal social control and handle potential offenders (Felson 1986).

Handlers

Routine activities acknowledged from the beginning that the presence of a capable guardian prevented crime. However, little distinction was made between guardians who supervise motivated offenders and guardians who supervise suitable targets. Later presentations are elaborated on this distinction by distinguishing between handlers and guardians. Whereas guardians supervise targets, handlers supervise offenders. The intimate handler emerged as a handler who supervises specific offenders using personal knowledge. Intimate handlers limit an offender's ability to engage in criminal acts using informal social control. Family, community members, and peers can all act as intimate handlers.

Facilitators

While the routine activities approach has been related to situational crime prevention, Clarke (1992) argues that the routine activities approach would have greater impact on situational crime prevention if it were to incorporate the idea of a crime facilitator. A crime facilitator is anything that comprises an essential tool for the commission of a particular type of crime. Examples of crime facilitators include some items commonly associated with crime such as weapons or lock picks as well as items typically associated with noncriminal behavior such as automobiles, telephones, and computers. Further, crime facilitators can include items that act as disinhibitors such as alcohol and drugs. The distinguishing characteristic of a crime facilitator is the simple fact that the item itself is involved in the criminal act. Because crime facilitators are distributed unevenly throughout space and time, the varying availability of crime facilitators is directly related to the routine activities of individuals both criminal and noncriminal.

Place Managers and Eck's Triplets

Another contribution to routine activities theory comes from John Eck (1994). Eck incorporates the idea of a place manager into the approach. A place manager is someone who is responsible for a particular space. Unlike a capable guardian who is protecting the suitable target, a place manager protects the place itself. Because the offender and target must converge somewhere in space, an individual guarding a particular space can avert crime even though they are not specifically interested in either the target or the offender. Doormen, security guards, and building superintendents are examples of place managers who can serve to discourage crime.

For Eck, the incorporation of place managers in conjunction with intimate handlers creates a situation where criminal opportunity is strictly limited. Known as Eck's triplets, this situation occurs when a likely offender is supervised by an intimate handler, a suitable target is supervised by a capable guardian, and the time and space where the offender and the target converge are supervised by a place manager, thus drastically reducing criminal opportunity. However, the effect of the triplet can be viewed in other ways as well. Offenders can be seen as discouraged by the proximity of guardians, managers, and handlers to targets, places, and the offenders themselves. In contrast, the effect can be viewed as relying upon the social ties that guardians, managers, and handlers have with targets, places, and offenders. Either way, the result is the same. A would-be offender must lose his handler and find an unprotected target in a location free of an intrusive manager to commit a successful crime.

Supervision When Offenders Control the Area

Supervision is typically viewed as a process that discourages crime. Capable guardians, intimate handlers, and place managers all serve to limit the criminal opportunities available to would-be offenders by virtue of their familiarity with targets, offenders, and places. However, familiarity does not always reduce crime. Offenders can assume a similar role and actually supervise guardians, handlers, and managers. Offenders may either seek to prevent these individuals from interfering or may actually find suitable victims among them. The process of familiarity generally strengthens the position of whichever group has the upper hand. When adversaries of crime have a stronger position, then familiarity serves to reduce offenders' ability to commit crimes. However, when offenders have a stronger position, then familiarity serves to enhance their ability to commit crimes. Offenders may have the upper hand in areas that have lost their basic controls. For example, areas with a high ratio of active offenders to others that have large numbers of abandoned properties likely tilt the balance of power to offenders. In such areas, offenders can rely on their familiarity with individuals and locations to commit crimes and evade supervision.

Non-predatory Offenses

The routine activities approach originally considered only direct-contact predatory violations. These violations were defined as illegal acts in which an individual definitely and intentionally took or damaged another's property. The offenses were likewise limited to only those that involved direct physical contact between at least one offender and at least one person or object that the offender attempted to take or damage (Cohen and Felson 1979). Later, Felson (1987) developed a crime classification system using four general categories of crime. The categories include crimes which are exploitative (where one person takes or damages the property of another), crimes which are mutualistic (those linking multiple individuals acting in complementary roles such as gambling or prostitution), crimes which are competitive (those involving multiple parties acting in the same role against each other such as fighting), and crimes which are individualistic (isolated illegal actions such as solo drug use or suicide).

Although the original conception of the routine activities approach applied only to exploitative offenses, the basic reasoning of the perspective extends to all four types of crime. All four types require that certain minimal elements converge in space and time for a crime to take place. Whereas predatory violations require an offender, a target, and the lack of a guardian, the other types have similar requirements. Mutualistic crimes require that at least two individuals inclined toward the offense meet in the absence of others who might interfere. An example is prostitution where a prostitute must meet a client in the absence of police and spouse. Competitive offenses have a similar requirement.

Fighters must encounter each other in the absence of peacemakers. Even individualistic violations involve a convergence of events. For example, suicides must avoid meddlers.

White-Collar Offenses

Some scholars have argued that white-collar crime can be understood using the routine activities approach (Weisburd et al. 2001). White-collar offenders are offenders who have specialized access to crime targets due to their work role, profession, or organizational position. Therefore, white-collar crimes can be viewed simply as crimes of special access. Viewing white-collar crimes in this way removes the sociological stereotypes associated with the behavior and explains why white-collar crime can be largely understood through the routine activities approach.

The vast majority of white-collar crime is mundane. It is easily accomplished by unskilled individuals and results in little economic return (Wright and Cullen 2000). Further, white-collar crimes rarely involve white-collar workers. Instead, the vast majority of white-collar crimes are committed by low-level individuals seeking material gain (Felson and Boba 2010). Because the motivations for white-collar crime are remarkably similar to those for burglary, there is little evidence to suggest that white-collar crime should be viewed any differently.

The only important feature which distinguishes white-collar crime arises from the specialized access derived from an offender's work role. The offender's position can provide greater criminal opportunity because official duties can provide access to targets. All offenders need access to targets, and white-collar offenders are no different. Whereas some offenders such as robbers take advantage of overlapping activity spaces to find targets, others such as burglars may take advantage of personal ties using personal information to determine which residences to burglarize. White-collar offenders take advantage of the unique access afforded them by their employment in their search for suitable targets. Quite simply, white-collar offenders have found a particular solution to the offender's general problem of finding suitable targets.

Organized Crime

Most organized crime is much smaller in scale and organization than depicted in the media (Reuter 1983). However, the study of organized crime still suffers from a variety of distractions related to the televised image. The degree of planning and sophistication required for offender to cooperate is generally overestimated, while the interaction between criminal cooperation and legitimate activities is typically underestimated (Felson 2006). This has resulted in extant studies of organized crime focusing on social networks in attempts to understand criminal organizations. This myopic approach overvalues linkages within specific groups which limits research into the nature of criminal organizations. In contrast, the routine activities perspective suggests that the important linkages for understanding organized crime are the linkages between events rather than between individuals.

Because offenders are likely to converge in certain settings, these settings allow criminal cooperation to persist even when the individuals vary. These offender convergence settings provide a stage for engaging in a variety of illicit activities. These locations provide a place where offenders can engage in illicit transactions as well as seek out accomplices. While some settings where offenders convergence exist solely for criminal activity, others involve legitimate activities. These places provide a social life for offenders who may generate criminal plans while engaged in otherwise innocuous activities. Thus, the nature of criminal organization may owe more to the routine activities of offenders themselves than to offenders' social networks.

Cyber Offenses

Cybercrime represents a relatively new topic within the field of criminology. Although few studies have attempted to analyze cybercrime, some scholars have argued that the routine activities perspective may provide a useful theoretical model for such research (Bossler and Holt 2009). The routine activities perspective may be applicable because, like other crimes, cybercrime requires that a motivated offender and a suitable target converge in the absence of a capable guardian. However, others dispute this assertion. They argue that because virtual environments are spatially and temporally disconnected, they are remarkably different than the traditional environments in which the routine activities approach has been previously applied. The disorganization caused by individuals and websites connecting and disconnecting over short periods of time may limit the applicability of the basic constructs of the routine activities approach (Yar 2005).

While this remains an open empirical question, research into cybercrime indicates that risk of cyber victimization is related to the specific tasks an individual undertakes on the internet, not simply internet use itself (Holt and Bossler 2008). For example, malware victimization was shown to be related to computer deviance (Bossler and Holt 2009). This suggests that individuals who choose to engage in deviant cyber acts place themselves in proximity to motivated offenders. Together, this line of research supports the applicability of the routine activities approach to cybercrime as it suggests that cyber victimization may result from convergence patterns of both victims and offenders within cyberspace.

Possible Controversies in the Literature

Two controversies concerning the routine activities approach are noteworthy. The first controversy extends back to the approach's original presentation. From its inception, the routine activities approach has been criticized as not conceptually distinct from social disorganization theory. However, several differences clearly distinguish between the two perspectives. Whereas social disorganization theory emphasizes broad places, the routine activities perspective concerns much smaller geographies such as specific streets, blocks, and addresses. The temporal emphasis varies between the two as well. Social disorganization focuses on changes over decades, but the routine activities approach focuses on temporal changes from hour to hour. While social disorganization treats offenders as the key personal focus, routine activities include targets and guardians. Finally, social disorganization theory is driven by the impact of social influences. In contrast, the routine activities perspective is driven by the idea of socio-physical convergences.

The second controversy concerns the proper dependent variable for the routine activities approach. Because the approach specifies that certain activities like household and family activities entail lower risk of criminal victimization than other activities (i.e., non-household/nonfamily), victimization represents the conceptual dependent variable. However, various studies have operationalized victimization in different ways. In the original paper, Cohen and Felson considered risk per billion person hours. Later, Clarke introduced the idea of partitioning risk by time spent in particular categories of activity. This idea resulted in the idea of risk per million hours parked in each place. This concept was later extended by Lemieux (2010) who conducted a more thorough analysis of risk per 10 million person hours in each of a variety of activities. Lemieux included sleeping, shopping, commuting to and from work, commuting to and from other areas, conducting other activities at home, working, commuting to and from school, engaging in leisure away from home, and attending school. Lemieux concluded that sleeping and shopping were the least dangerous activities while leisure away from the home and attending school were the most dangerous.

Conclusions

The routine activities perspective provides an important conceptual tool in the study of crime. Originally conceived to explain the prevalence of a particular type of crime during a specific period of time, the routine activities perspective has been adopted to explain many different types of crime in a wide variety of environments. While this versatility represents one of the great strengths of routine activities, the consistency of the original insights throughout the development represents the approach's single greatest attribute. This is no doubt the result of the elegant simplicity of the perspective itself.

While the future of routine activities is unclear, the close relationship between routine activities, situational crime prevention, and CPTED will provide a variety of opportunities for continued development. Because routine activities research focuses on microlevel processes, the routine activities approach leads to specific policy recommendations. As a result, the perspective is useful to both scholars and practitioners alike.

Related Entries

- ▶ Applied Geographical Profiling
- ▶ Crime Prevention Through Environmental Design
- ▶ Crime Science
- ▶ Criminology of Place
- ▶ Effectiveness of Situational Crime Prevention
- ▶ Hot Spots and Place-Based Policing
- ▶ Lifestyle Theory
- ▶ Problem-Oriented Policing
- ▶ Situational Crime Prevention
- ▶ Theories for Situational and Environmental Crime Prevention

Recommended Reading and References

Bossler AM, Holt TJ (2009) On-line activities, guardianship, and malware infection: an examination of routine activities theory. International Journal of Cyber Criminology 3:400–420

Brantingham PL, Brantingham PJ (1993) Nodes, paths and edges: considerations on the complexity of crime and the physical environment. Journal of Environmental Psychology 13:3–28

Clarke RV (1992) Situational crime prevention: successful case studies. Harrow and Heston, Gulderland

Clarke RV (1997) Situational crime prevention: case studies. Harrow & Heston, New York

Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. American Sociological Review 44:588–608

Eck JE (1994) Drug markets and drug places: a case–control study of the spatial structure of illicit drug dealing. Doctoral dissertation, University of Maryland, College Park

Fattah EA (1991) Understanding criminal victimization. Prentice-Hall, Scarborough

Fattah EA (2000) Victimology: past, present, and future. Criminologies 33:17–46

Felson M (1986) Linking criminal choices, routine activities, informal control, and criminal outcomes. In: Cornish DB, Clarke RV (eds) The reasoning criminal: rational choice perspectives on offending. Springer, New York, pp 119–128

Felson M (2006) The ecosystem for organized crime. (HEUNI Paper No. 26: The European Institute for Crime Prevention and Control). Helsinki

Felson M, Boba R (2010) Crime and everyday life, 4th edn. Sage, Thousand Oaks

Felson M, Clarke RV (1998) Opportunity makes the thief: practical theory for crime prevention. (Police research series paper no. 98). H.M.S.O., London

Gottfredson DM, Clarke RV (eds) (1990) Policy and theory in criminal justice: contributions in honor of Leslie T. Wilkins. Avebury, Aldershot

Hawley A (1950) Human ecology: a theory of community structure. Ronald, New York

Hindelang MJ, Gottfredson MR, Garofolo J (1978) Victims of personal crime: an empirical foundation for theory of personal victimization. Ballinger, Cambridge

Hirschi T (1969) Causes of delinquency. University of California Press, Berkley

Hollis-Peel ME, Reynald DM, Bavel M, Elffers H, Welsch BC (2011) Guardianship for crime prevention: a critical review of the literature. Crime, Law, and Social Change 56:53–70

Holt TJ, Bossler AM (2008) Examining the applicability of lifestyle-routine activities theory for cybercrime victimization. Deviant Behavior 30:1–25

Laycock G (2003) Launching crime science (Department of Security and Crime Science, Jill Dando Institute of Security and Crime Science). University College, London

Lemieux A (2010) Risk of violence in major daily activities: United States, 2003–2005. (Doctoral dissertation). Rutgers, the State University of New Jersey, Newark

Mayhew P, Clarke RV, Hough M, Sturman A (1976) Crime as opportunity. (Home Office research study no. 34). H.M.S.O., London

Ogburn WF (1922) Social change with respect to culture and original nature. B. W. Huebsch, New York

Reuter P (1983) Disorganized crime: illegal markets and the mafia. MIT Press, Cambridge

Van Dijk JJM, Steinmetz CHD (1983) Victimization surveys: beyond measuring the volume of crime. Victimology 8:291–309

Weisburd D, Waring E, Chayet EF (2001) White-collar crime and criminal careers. Cambridge University Press, Cambridge

Wright JP, Cullen FT (2000) Juvenile occupational delinquency. Criminology 38:863–896

Yar M (2005) The novelty of cybercrime: an assessment in light of routine activity theory. European Journal of Criminology 2:407–427

Zahm D (2007) Using crime prevention through environmental design. (Problem-Oriented Guides for Police Problem-Solving Tools Series No. 8). U.S. Department of Justice. U.S. Government Printing Office, Washington, DC

Routine Patrol

► Preventive Patrol

Russian Mafia

James O. Finckenauer
School of Criminal Justice, Rutgers University,
Newark, NJ, USA

Overview

What is the so-called Russian mafia? Answering this question is not easy. The difficulty arises out of a number of circumstances that will shape this discussion. First, the phenomenon called Russian mafia is not and cannot be strictly Russian. A second complication is the fact that the term mafia had come to be indiscriminately used in the former Soviet Union, where it referred widely to any sort of activity or enterprise that was perceived to be outside of the law. Thirdly and most importantly, there is the most unusual story of Russia and Russians with respect to crime, organized crime, and corruption – a history that spans at least back to the seventeenth century, takes in the 70 plus years of Soviet rule, and finally focuses on the "New Russian" generation since the collapse of the Soviet Union in 1991. Each of these factors, but most particularly the last, helps explain the current situation with respect to organized crime in Russia itself, as well as the transnational organized crime that emanates from Russia with widespread global impact. Finally, following the explanation and understanding of this particular organized crime problem, there is the ultimate challenge of what to do about it.

Background

Describing and attempting to explain the popularly known "Russian mafia" is a task with a number of complications. Contrary to what might appear to be the obvious, it requires first dealing with the characterization of its being "Russian." This is because Russia and Russians have long been all-encompassing labels. As was particularly the case during the time of the Soviet Union (1918–1991), the world regarded persons from all the countries that made up the former USSR as essentially being Russians. *Russian* came to include persons not only from the other 14 Soviet republics but also from the hundreds of different ethnic groups. Although Russia is the largest and perhaps best known of the former Soviet states and although there are many ethnic Russians who have immigrated elsewhere, indeed including some who are criminals and engaged in organized crime, and not all of those currently popularly characterized as being part of the so-called Russian mafia are actually Russian. Instead, they are, for example, Armenian, Georgian, Ukrainian, Azeri, or Chechen or they are actually from Eastern Europe, such as the Albanians. The point is this phenomenon known as the Russian mafia is not and cannot be strictly Russian. This is also why law enforcement agencies in the United States more typically refer to the criminal groups in this category as Eurasian Organized Crime.

Even thornier is the concept of "mafia." Strictly speaking, mafia is a very particular kind of organized crime. It is especially associated with Sicily and southern Italy. Made members of the true mafia – inducted in formal ceremonies – are considered to be men of honor. They are bound by traditions that have been passed down for generations, such as respecting *omerta* or the vow of silence.

The mafia's traditional business has been providing protection and acting in quasi-governmental roles. The latter is a particularly important defining element. It means that mafias are most likely to be found in countries or regions that lack strong governmental structures and healthy civil societies. Such locations are usually characterized by rampant corruption, weak or absent rule of law and protections for civil liberties, and/or civil disorder. In these circumstances, a mafia or mafias take on "protective" functions and judicial functions, meting out judgments and sanctions. Although the term mafia has come to be used indiscriminately to refer to all kinds of criminal organizations, in many places, mafia is not synonymous with organized crime. Rather, it is one form of organized crime.

The term mafia was particularly indiscriminately used in the former Soviet Union. It referred widely to any sort of activity or enterprise that was being controlled or that at least appeared to be controlled. Thus, mafia could refer for example to organized criminal groups, clans that controlled politics and the economy in particular regions, corrupt government employees, or even the Communist Party itself. The noted Russian journalist, Yuri Shchekochikhin, who covered the ubiquitous mafia in the USSR extensively, said the label mafia was being applied to practically anything and everything – unions, hospitals, prostitutes, chess players, etc.

In any instance when terms are used somewhat arbitrarily, they tend to become practically meaningless. As a result of this longstanding definitional ambiguity, the subject to be dealt with here is consequently a rather amorphous one. Yet a further complication derives from the unique history and character of Russia and Russians with respect to crime, organized crime, and corruption – a history that spans at least back to the seventeenth century, the 70 plus years of Soviet rule, and then the generation since the collapse of the Soviet Union in 1991.

History and Origins

Organized crime in various forms has a 400-year history in Russia. This is a history characterized by authoritarian governments, beginning with the czars, who were then followed by the hegemony of the Communist Party and its dictatorial rulers. Indeed, authoritarianism, albeit it in a somewhat different form, extends into today under the rule of Russian President Vladimir Putin. Putin, the former head of the KGB (the Soviet secret police), calls his form of authoritarianism democracy with order.

Throughout Russian history, its rulers showed little or no regard for individual rights and liberties. Government and its laws and rules were not seen as representing and protecting the people. A consequence of this history was and is blurred distinctions between what is legal and illegal and what is public versus private. Stealing the czar's timber, for example, was not considered criminal by the Russian people, whereas one peasant stealing anything from another was. Commitment to the rule of law, a backbone of democracy and civil society, has thus always been very feeble in Russia.

Serguei Cheloukhine (2008), who has studied and written extensively about Russian organized crime, describes its beginnings as follows:

> Historically, the professionalization of criminal groups in Russia was a product of strong patriarchal traditions, hostility towards the state and an underdeveloped, largely agrarian economy. During the 17th century, corruption combined with the tyranny of state officials and landowners was widespread in Russia. This was in part a reflection of the lack of respect for the dignity of the individual – most people were regarded primarily as property and such property was not protected by a rule of law, but was due to the "benevolence" of the autocracy towards the country's nobility.

It was during these centuries prior to the Soviet era, when the czars ruled Russia, that one of the most prevalent crimes was theft, especially as indicated from the czar's properties. The thieves' rationalization was that property that belonged to the czar – to the state – did not belong to anyone, or belonged to everyone, and thus was fair game to be stolen. Although not unique to Russia, this sort of rationalization is important because it reflects the origins of the division between permissible and impermissible crimes that rose to new heights in the Soviet Union. And it continues to be very much in evidence today.

Another piece of this explanatory puzzle can be found in the reading of Russian literature. From Nikolai Gogol (*The Government Inspector*) to Ilya Ilf and Yevgeni Petrov's *The Twelve Chairs* and *The Golden Calf*, to Alexander Zinoviev (*The Yawning Heights* and *Homo Sovieticus*), and to more recent authors such as Victor Pelevin

(*Homo Zapiens*) and Gary Shteyngart (*Absurdistan*), writers have described the insidious corruption and the particular long-standing relationship between corruption and crime in Russia. A sign that this norm continues was reflected in a 2012 New York Times article that opened a discussion of the issue as follows: "CORRUPTION [sic] in Russia is so pervasive that the whole society accepts the unacceptable as normal, as the only way of survival, as the way things 'just are.'"

Crime in the USSR

Following the Russian Revolution of 1917 that resulted in the overthrow and ultimate execution of Czar Nicholas II and his family, the Bolsheviks embarked upon their ambitious agenda of creating a new society, intended to be crime-free and also to produce a new "Soviet Man" who would be hardworking, communal, patriotic, and loyal to the state. Oddly however, it was during the 1917 revolution that the Bolsheviks collaborated closely with professional criminals to overthrow first the czar and then the first short-lived revolutionary government. Thus, from the outset, there was a kind of hypocrisy about the supposed Soviet utopia. The Russian revolutionaries working hand in glove with criminals to attain the spoils of power reflects a picture of Russia today – the marriage between political power and crime. A very similar kind of criminal-political nexus – among entrepreneurs, criminal organizations, and government officials – very aptly describes the current nature of serious crime in Russia.

In most instances, organized crime arises to meet market demands for goods and services that are either illegal, regulated, or simply in short supply. In addition to goods that were illegal such as a variety of Western products, almost from the beginning, most goods and services were in chronically short supply in the USSR. This led to, or fed, the emergence of what was known as a shadow economy and to a black market. The shadow economy produced legal goods "off the books" that were sold or bartered illegally. The black market usually dealt in illegal goods, such as prohibited items from the West or drugs, for example. As indicated, it was common during the Soviet era to refer to practitioners in a variety of these activities as being mafias.

Soviet citizens used *blat* – connections and informal networks – to access the black market and shadow economy. Blat (Russian: блат) is a term which appeared in the Soviet Union to refer to the use of informal agreements, exchange of services, connections, Party contacts, or black market deals to achieve results or get ahead. While obviously not unheard of elsewhere, agreements/connections became the primary way of doing business in the USSR. *Blatnoy* came to refer to someone who got a job or got into a university using connections or sometimes bribes – again not unique to Russia, but just much more prevalent. This system was in many ways similar to the Chinese practice of *Quanxi*.

As a result of eight decades of these practices, there are at least two effects that are relevant to the current nature of Russian organized crime: It created or refined a distinct mentality about conniving to survive, and it stimulated the development of a shadowy entrepreneurial class operating outside the law. The previously mentioned writer Zinoviev's *Homo Sovieticus* was a string puller, an operator, a time server, a freeloader, and, above all, a survivor. He was also the ancestor of today's new Russians.

The offshoots of this peculiar history include, for example, the monetarization of organized crime in Russia, in which dealing in goods and services shifted to dealing in currency. This inheritance also includes the crony capitalism and widespread corruption that are so characteristic of Russia's economy and politics today.

During the 1960s, organized crime in Soviet Russia evolved into a three-tiered pyramid-like structure. The top tier was made up of high-level government and Communist Party bureaucrats – those referred to as the *apparatchiks* and the *nomenklatura*. Party officials enjoyed special privileges that were denied to the vast majority of the Russian people. They lived in special more luxurious housing; had their own vacation hideaways; consumed special foods, wines, and liquors that were otherwise unavailable; traveled widely and bought expensive clothes; and sent their sons and daughters to the best schools. Party members lived in high style off of the country's vast resources while all the while claiming to be representing the people's interests in a socialistic state of mythical equality. The hypocrisy of this lie was apparent to the population, and once again reinforced the lesson about what one really needed to do to get ahead in life.

Below the Communist Party, the middle tier of the crime pyramid was comprised of the shadow economy (underground) operators and the black marketers. They, for example, oversaw the production of the shoddy goods to fulfill the 5-year plan quotas. As long as the plan was fulfilled or better yet exceeded, the quality of the products was irrelevant – again a foundation for cynicism. These operators also procured the illegal goods and products that were then passed on up the pyramid. They did all this while siphoning off what they saw as their share.

Finally, the bottom tier consisted of the professional criminals. This group included a whole range of criminal types from petty thieves up to an including the *vory v zakone* or thieves in law. The *vory* were products of the Soviet gulag or prison system, what Cheloukhine describes as "a university of the underworld." These leaders of the underworld were (and still are in some instances) the most sophisticated of the professional criminals. They lived within a set of rules known as the thieves' law, dedicated themselves to a life of crime, and rejected involvement with the legitimate world. In many respects, although not all, they bore some resemblance to the traditional definition of mafia or at least more so than did any other form of Russian organized crime.

The Post-Soviet Era

There were great hopes for Russia after 1991. Unfortunately, many of those hopes have not materialized. Instead of enjoying the fruits of freedom and democracy and a thriving market economy, Russia has been racked by a host of social, economic, and political ills. The Russian people have been confronted with worsening living standards, poor health care, rampant inflation, corruption, alcoholism, drug addiction, and surging crime rates. Russia faces a shrinking population because of declining birth rates and decreasing life expectancy, especially among Russian men. The latter is attributable to the increased alcoholism.

The predatory nature of the Russian economy – an economy whose driving philosophy seems to be to steal as much as you can as fast as you can – is closely intertwined with official corruption and organized crime. One result is that, for the vast majority of Russians, life is much worse today than it was at the beginning of the decade of the 1990s – an outcome that would have been almost inconceivable in 1991. Another result is that the legal and criminal justice processes, and any fledgling effort toward the rule of law, have been overwhelmed by the enormity of the challenge of crime.

In the years following the collapse of the former Soviet Union, there was a rapid growth in organized economic crime. During Boris Yeltsin's presidency, both political and economic chaos reigned as Russia stumbled from a state command economy to something resembling a market economy. Predators who had cut their teeth in the shadow economy, the black market, and the criminal corruption of the Soviet era were best positioned to take advantage of the chaos – lining their pockets and those of their friends. The so-named oligarchs – the millionaires and billionaires who profited the most – were beneficiaries of the policy of privatization, that is, the selling off of the state's assets. Enormously lucrative properties were obtained for investments far below their worth largely through collusion and insider trading.

As the country moved to privatize its vast holdings, that is, sell to private owners what was previously owned by the state, organized crime in partnership with corrupt officials engaged in a wholesale rape of the privatization process, buying up valuable properties at fire sale prices via insider trading information. The economy that resulted from the various perversions of economic reform has thus not been an economy governed by market principles but is instead, as indicated, a predatory economy.

Following Yeltsin, first the Putin and then the Alexander Medvedev governments grappled with the problems of developing laws and regulations to govern private business and economic activities. But organized criminal groups were clearly able to exploit the opportunities that became available. This wave of criminal activity was a destabilizing influence as Russia struggled to achieve economic and political reform. It thwarted foreign investment, widened the gap between rich and poor, and ultimately undermined confidence in the government.

Given this context, the particular character of Russian organized crime that has arisen over the last 20 years reflects not only a high level of criminal sophistication and well-developed networks of corruption but also a broad scale of activities and influence exerted over a considerable portion of the economy. Russian organized crime is not a single monolith but rather a conglomeration of criminal groups. Some of these groups are large and some small; some are ethnically based and some are not; some specialize in particular crimes and others are more diverse. In sum, when one speaks of Russian organized crime, one is not speaking of a single entity.

The criminal activities and enterprises of these various groups can be categorized into two broad groupings. In one group are such crimes as extortion, robbery, car theft, drug trafficking, and prostitution. These kinds of crimes are the staples of organized crime in many parts of the world. But it is really the second category of crimes that most distinguishes Russia from many other countries. Between 1991 and 2008, for example, Russian organized crime gained control of a significant share of the nation's economy (estimates run to 40 % or more in certain sectors). The lack of laws, lax regulation, corruption, and violence enabled criminal groups to make substantial inroads into several lucrative economic sectors, including energy, metallurgy, construction, banking, fishery, retail trade, liquor production, and transportation. The Russian Ministry of Internal Affairs (MIA) estimated that almost 2,000 different economic enterprises were actually under the control of organized crime groups. This latter category of high-level economic crimes was much more likely to involve a class of sophisticated criminals known broadly as the *avtoritety* or authorities. This is a collection of businessmen/criminals who are drawn from the ranks of former black marketeers and shadow economy operators, government bureaucrats, and the security services, among other roots. Whereas the drug crimes, protection rackets, loan sharking, and so on were the bread and butter for the old style gangsters and the *vory v zakone*, the avtoritety favored corruption, embezzlement, financial crimes, and other forms of white-collar crime. It should be emphasized, however, that these two categories are not mutually exclusive.

Shelley has pointed out that hundreds of banks in Russia are owned or controlled by organized crime groups and that these groups use the banks to launder money, for pyramid schemes, and so on. The capital flight out of Russia, says Shelley, is one of the most damaging aspects of Russian organized crime activity.

Organized crime has been able to penetrate Russian business and state enterprises to a degree that is unheard of in most other countries of the world, even including those where organized crime is present. As already indicated, this penetration occurs in a variety of forms. For example, protection rackets are operated by what are called *kryshas* or roofs. Nearly every business in the major cities must make extortion payments for "protection" by a krysha. This protection sometimes comes in the form of acquiring unwanted partners.

Russian organized crime in Russia itself has most recently evolved into new forms that are diminishing the role of the old *vory v zakone*. Professionalization, globalization, legalization, and patrimonialism characterize these new forms. Their skill at white-collar crimes, for example, electronic crimes and crimes victimizing banking and financial institutions, gives witness to the professionalism of the new Russian criminals. At the same time, a Russian diaspora has developed – particularly in Canada, Israel, and the United States, for example. This diaspora provides a potential critical mass for the development of organized crime in those areas. The globalization is also evident in such activities as worldwide money laundering and in the trafficking of a variety of goods and persons. Through effective use of the mass media and charitable donations and through heavy giving to political campaigns and even electing their own candidates to political office, Russian organized crime has moved to legalize, in effect to legitimize itself. Finally, patrimonialism – via the kind of insider training, the issuance of preferential licenses, and the illegal banking of state funds, referred to earlier – is the other major force driving organized crime into the center of "normal" political activity. Whereas in most societies crime is an outside threat external to the society, in Russia, organized crime is not only pervasive but is very much mainstream.

One expert on Russian organized crime, Mark Galeotti, has pointed out that "Russia's internal situation is the most significant asset of Russian organized crime." The fact that law enforcement officials are corrupt and that politicians are intertwined with criminals makes Russia a safe haven for laundering ill-gotten gains through its banks but also for hiding a variety of other criminal activities and the criminals themselves. Galeotti thinks this might be changing because the political costs of being associated with criminal networks have increased because of pressures from society to reform. Also, he cites the internal "silovik wars," which bring to light rivals' illegal businesses and crimes. As a result, "the 'price tag' for political protection is increasing and those who are not willing to pay are made an example of."

Galeotti says there are now three generations of criminal groupings in Russia: The oldest group is made up of those who came out of the gulag, including the *vory v zakone*. Then there are the *avtoritety*, referred to earlier. Many of these criminal entrepreneurs have moved into the upper levels of the economic and political systems. The newest and third generation is comprised of criminal operatives who likewise have embedded themselves in the ruling political and economic apparatus. Galeotti says "these [latter] criminals come from within the elite and although they are cosmopolitan, educated in the West, and know how to operate internationally, they have opted to continue the criminal activities of their parents as an easy and quick way of making big money."

The kind of sucker mentality that is a legacy of the Russian experience obviously continues to be very much alive and well – not only in Russia but also in Russian émigré communities overseas as well. This is a mentality that extolls getting ahead by getting over. It is based upon an inherent belief that one generally moves up and makes it in the world, not by hard and honest work which is for suckers but rather by scheming and scamming and by using *blat*.

Combating Russian Organized Crime

The current environment in Russia is very conducive to the emergence and sustenance of organized crime and to the export of that crime abroad. Among the structural tools any society needs to resist and combat organized crime are an aware and active civil society, a culture of lawfulness, and a strong public disgust for crime and corruption. At least partly as a legacy of its connive to survive history, such society and culture do not now exist in Russia. Also needed, but likewise in still quite an anemic state for some of the same reasons, is an independent and incorruptible legal and judicial system.

Russia does not have effective legal tools for fighting organized crime nor does it have an effective law enforcement system. As Serio (2008) has noted:

> . . .in order to create an effective organized crime control program, five elements are needed. First, criminal statutes which allow for the prosecution of a criminal enterprise as a whole. . ..Second, there need to be laws which allow for the use of confidential informants and undercover operations. Third, the use of cooperating witness testimony in court proceedings is critical for the successful prosecution of organized crime cases. Fourth, the legal base must permit electronic surveillance and the use of wiretap evidence in court proceedings. And fifth, an effective witness protection program must be in place. Virtually none of these tools are in place in Russia in a meaningful way. . ..

There are a number of explanations for why these legal tools are unavailable or limited in Russia, ranging from holdover concerns about having too much power put into the hands of prosecutors or other law enforcement entities (where it has been abused in the past) to the fact there are a number of occupants of the Russian parliament who are or have been either directly or indirectly involved in organized crime and thus have reason to protect those interests by blocking the passage of certain laws.

"Russian Mafia" Abroad

It has been estimated by the Russian Ministry of Internal Affairs (MIA) that there are some so-named Russian criminal groups operating in upwards of 50 countries. Their greatest presence is in Europe and North America, but they have also shown up in Australia and Israel, for example. As in other historical instances of organized crime, for example, La Cosa Nostra, the most opportune locations for this Russian crime are where there is a budding Russian community that creates a critical mass for victimization. A common language, common culture, common fear of law enforcement, ignorance and distrust of local institutions, and family members back in the former USSR present an inviting target and criminal opportunity.

Russian criminals are known to have been active in the United States since the mid-1970s. Their activity at that time was, however, quite limited and localized in the main Russian-speaking community in Brighton Beach, in Brooklyn, New York. During the 1980s, the numbers of persons from the former Soviet Union allegedly involved in crime grew, and the crimes themselves became more sophisticated and organized. For example, a series of bootleg gasoline tax evasion schemes netted their perpetrators millions of dollars.

Although the types of crimes typically perpetrated require organization to carry them off, the criminal groups engaged in them do not generally possess the characteristics most associated with traditional organized crime. Instead, they have loosely connected and fluid criminal networks. These do not have continuity over time and crimes. They do not have hierarchical structures. They do not attempt to gain monopolization of the criminal market, nor do they systematically use violence and corruption to further their ends. This is not to say they are not violent, for example, but rather that their use of violence tends to be more situation specific and sporadic than systematic. Violence tends to be most associated with their extortion and protection rackets. Most importantly for this discussion, these kinds of free-floating networks do not have any of the characteristics associated with traditional notions of what a mafia is and looks like.

Growing beyond the initial Brighton Beach location, Russian (or better Eurasian) criminals have more recently been said to be active in major metropolitan areas in a number of US states, with California being a particular case. There, Eurasian Organized Crime (EOC) is principally made up of Armenians, Russians, and Ukrainians. Their criminal activities include firearms trafficking, auto theft, cargo theft, extortion, murder, prostitution, money laundering, drugs, and human smuggling and trafficking. Their real specialties, however, are health care and financial frauds. They are said to familiarize themselves with both state and federal government health care policies so as to bilk insurance companies and programs such as Medicare and Medicaid. They engage as well in a wide variety of other frauds. It is this specialization, along with such crimes as identity theft, the black market purchases of various kinds of professional licensing, and staged auto accidents, that most clearly demonstrate the continuing effects of the kind of criminal socialization that gave birth to today's criminals from the former Soviet Union. It is this that is the legacy of the hundreds of years of Russian history.

Two of the better known US prosecutions of criminals from the former Soviet Union are the cases involving Vyacheslav Ivankov, who was Russian, and Semyon Mogilevich, who is Ukrainian. Ivankov, who was nicknamed Yaponchik or Little Japanese, was convicted in 1996 of extortion and marriage fraud. He served 7 years in prison in the United States and was then deported back to Russia where he was subsequently assassinated. He is believed to have been a *vor* and the highest ranking Russian organized crime figure ever in the United States. The Mogilevich case, involving an ongoing prosecution, charges Mogilevich and others with a scheme to defraud investors in the Canadian company YBM Magnex International, Inc. The specific charges are wire fraud, mail fraud, and money laundering.

Given their history, as we see, the predominant crimes of persons from the former USSR in the United States are frauds and scams. The criminal networks they have established are diverse in the criminal opportunities they exploit, and they reflect a blend of legal and illegal activities. Their crimes reflect their descent from the old system of blat, the black market, and the shadow economy. What these criminal networks and their activities do not resemble, however, is what is commonly accepted as constituting that we know as a mafia or at least something that is mafia-like.

Conclusion

So is there a Russian mafia after all? In Russia itself, there are certainly criminal organizations that look and act like a mafia. They have indeed penetrated and embedded themselves in the governing apparatus and are running a number of security operations that entail functions normally in the domain of government. Some experts believe that there is such a nexus between the state and organized crime in Russia and that organized crime has actually become an instrument of state power. Particularly significant, they believe, is the presence of organized crime in such strategic markets as energy and metals, as well as with respect to nuclear power. Another area of both domestic and international threat is in cybercrime, where Russia has many sophisticated perpetrators that are instruments of Russian organized crime.

There are a number of other criminal groups in Russia, however, that although similarly labeled as being mafias, do not actually have mafia trappings. And interestingly, even in Russia, a large number (perhaps even a majority) of criminal groups are not actually or solely Russian. Instead, they are Chechen, Azeri, Dagestani, and Georgian, among others. Of the estimated 800–1,000 or so *vory v zakone* still believed to be active in organized crime groups, the majority of them are said by the MIA to be Georgian.

Without a doubt, organized crime emanating from the former Soviet Union is a major threat, both domestically and internationally. In this sense, it does not matter that "it" is neither Russian nor a mafia. The extent to which a variety of criminal groups under this label have infiltrated business and government in Russia threatens the possibilities for progressive reform and could threaten the stability and national security of the country. Abroad, the ease of international travel, the liberalization of emigration policies, the expansion of foreign trade, the high-tech nature of communications, and the possibilities for money laundering are all changes and vulnerabilities being exploited by criminals from the former Soviet Union. Whatever the label, more and better tools, and more and better collaboration and cooperation within the world community to combat this problem are in everyone's interests.

Related Entries

- ▶ Careers in Organized Crime
- ▶ Corruption
- ▶ Crimes of Globalization
- ▶ History of Organized Crime in the United States
- ▶ Italian Mafia
- ▶ Money Laundering
- ▶ Organized Crime and the Environment
- ▶ Organized Crime, Types of
- ▶ Organized Fraud
- ▶ Racketeering
- ▶ Social Network Analysis of Organized Criminal Groups

Recommended Reading and References

Cheloukhine S (2008) The roots of Russian organized crime: from old-fashioned professionals to the organized criminal groups of today. Crime Law Soc Change 50:353–374

Cheloukhine S, Haberfeld MR (2011) Russian organized corruption networks and their international trajectories. Springer, New York

Finckenauer J (2007) Mafia and organized crime. One world, Oxford

Finckenauer J, Voronin Y (2001) The threat of Russian organized crime. NIJ Issues in International Crime, Washington, DC

Finckenauer J, Waring EJ (1998) Russian mafia in America. Northeastern University Press, Boston

Galeotti M (2012) Transnational aspects of Russian organized crime. Chatham House, London

Gogol N (1842) The government inspector (a play). The University of Adelaide Library University of Adelaide South Australia 2012

Handelman S (1995) Comrade criminal: Russia's new mafiya. Yale University Press, New Haven

Ilf I, Petrov Y (1928 [1997]) The twelve chairs. Northwestern University Press, Evanston

Ilf I, Petrov Y (1931 [1962]) The golden calf. Muller, London

Karklins R (2005) The system made me do it: corruption in post-communist societies. M E Sharpe, Armonk

Pelevin V (2002) Homo zapiens. Viking, New York

Politkovskaya A (2004) Putin's Russia. Henry Holt, New York

Satter D (2003) Darkness at dawn: the rise of the Russian criminal state. Yale University Press, New Haven

Serio J (2008) Investigating the Russian mafia. Carolina Academic Press, Durham

Shelley L (1997) The price tag of Russia's organized crime. Transition, The World Bank, Feb 7–8

Shteyngart G (2007) Absurdistan. Random House, New York

Varese F (2001) The Russian mafia. Oxford University Press, Oxford

Volkov V (2002) Violent entrepreneurs: the use of force in the making of Russian capitalism. Cornell University Press, Ithaca

Zinoviev A (1976 [1979]) The yawning heights. Random House, New York

Zinoviev A (1982 [1986]) Homo sovieticus. Grove/Atlantic, New York

S

Safety

► Officer Safety, Health, and Wellness

Sample Selection Models

Brian Johnson
Department of Criminology and Criminal
Justice, University of Maryland, College Park,
MD, USA

Overview

Analytical issues involving sample selection are pervasive in criminological research. This entry provides an introductory overview of the sample selection problem followed by discussion of several common statistical models designed to address sample selection bias in criminological research. Examples of research involving sample selection models are discussed along with the different types of sample selection that can occur and their appropriate modeling strategies. The entry concludes with a brief overview of common criticisms and concerns surrounding sample selection models in the field and discusses future directions for development in this area.

Introduction

Social scientists have long recognized the ubiquitous threat of sample selection. Whether relying on surveys, interviews, official records, or even experiments, sample selection issues are pervasive in criminological research. Whenever information is available on a nonrandom subsample of the population of interest, or when observations are selected through a process that is not independent of the outcome of interest, sample selection bias represents a key threat to empirical validity (Bushway et al. 2007). Examples of selection bias in criminological work are wide ranging. From examinations of police discretion (Worral 2002; Lundman and Kaufman 2003; D'Alessio and Stolzenberg 2003; Kingsnorth et al. 1999) to sentencing disparity (Berk 1983; Hagan and Palloni 1986; Klepper et al. 1983), from publication bias in meta-analysis (Wilson 2010) to the effects of adolescent work on crime (Apel et al. 2008), and from tests of self-control theory (Paternoster and Brame 1998) to examinations of life-course criminology (Laub and Sampson 2003), selection bias often rears its ugly head in criminological research. Indeed, it can be difficult to identify any criminological research question that is not threatened to some degree by sample selection. This entry briefly reviews analytical issues surrounding sample selection bias, including why sample selection leads to biased and inconsistent estimates, how sample selection has been dealt with in criminological research, and how important conceptual distinctions in different types of selection can be classified and understood. The focus is on statistical models that are designed specifically for situations involving censoring or truncation of

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

the dependent variable, which are collectively referred to here as “sample selection models.”

Sample Selection in Criminological Research

The threat of selection bias is omnipresent in social research, and it is a fundamental aspect of many criminological inquiries – it can result from sundry causes, including natural human behaviors, specific research designs or implementations, or a host of other unforeseen individual, organizational, and environmental causes. Sample selection can result from instrument bias, such as when a survey instrument is designed to collect certain information from only a subset of respondents. It can be caused by sampling strategies that systematically omit certain populations, such as the homeless or incarcerated populations (Mosher et al. 2002). It can also arise through sample attrition, such as when different types of subjects are more likely to drop out of a study than others. Not only is sample selection an important source of bias, but sample selectivity itself may at times be of substantive interest to the researcher as well. If the selection mechanisms affecting the sample are fully known, they can be accounted for through sampling weights or related mechanisms, but rarely are all selection processes easily identifiable or explicitly defined in criminological research.

Various different treatments of sample selection issues have conceptualized the problem in terms of nonresponse, sample attribution, model misspecification, and/or omitted variable bias (Gondolf 2000; Maxwell et al. 2002; Berk 1983; Heckman 1979; Little and Rubin 1987). Selection bias can be understood as a sampling issue (the researcher is limited to information from a nonrandom subsample of the population of interest), a selection problem (observations are selected in such a way that they are not independent of the outcome of interest), or as a statistical problem (a correlation exists between the error term and the independent variable in the model). However it is conceived, though, the basic problem is the same – a researcher is limited to information on the dependent variable that comes from some subsample of observations that is not randomly drawn from the population of interest and therefore is not representative of the larger population to which inferences are to be made. Because statistical inferences are grounded in the principles of random sampling, generalizations based on nonrandom subsamples are likely to lead to faulty causal inferences.

The prototypical example comes from economics, where researchers are often interested in estimating wages, but wages are only observed for employed individuals (e.g., Heckman 1974; Moffitt 1999; Hagan 1990). Hagan (1990), for instance, examined this issue in the context of gender stratification in the justice system, examining income inequality among male and female lawyers. He found that both men and women who were more committed to work roles were more likely to be employed full time, and because commitment to work affected earnings, comparisons based only on full-time employees led to biased estimates of income differentials in the legal profession.

Similar examples are common in other areas of criminological research. Scholars interested in criminal punishment, for instance, only observe severity of charges for arrested offenders, sentence lengths for incarcerated offenders, and recidivism outcomes for offenders who have been released from jail or prison. Of course the same factors associated with being arrested, incarcerated or reoffending are also likely to affect initial charge severity, sentence lengths, and the likelihood of release. The problem, then, is that the process that determines which cases will be observed is fundamentally related to the outcome of interest to the researcher. This means that at any given stage of criminal case processing, the observed sample is unlikely to be representative of the larger population of all offenders from which it originated. Selective attrition through different stages of the justice system produces nonrandom samples of offenders, so estimates based on these subsamples provide biased inferences about relationships in the larger population of interest. Kingsnorth et al. (1999), for instance, investigated the role that victim characteristics and offender-victim relationships played in the severity of case processing outcomes from initial intake to final disposition in a sample of sexual assaults. As they recognized, "Unobserved factors may influence whether a case is continued to the next step," therefore "Failure to take these unobserved influences into account... leads to the problem of sample selection bias" (Kingsnorth et al. 1999, p. 284). Out of the 467 cases in their sample, only 41 % were fully prosecuted and even fewer received terms of incarceration. To address this, these authors implemented sample selection models in an attempt to control for selection bias in their analysis of sentence lengths.

Another source of potential selection bias can arise in offending or victimization surveys that have skip patterns built into them (Winship and Mare 1992). For instance, a researcher might ask survey respondents to report whether or not they have ever been arrested, and then ask additional questions of arrested respondents, such as how the arrest affected their level of satisfaction with police. Analyses based on these questions, though, would only be observed for arrested respondents, so inferences drawn from them would be subject to important selection biases. Negative attitudes towards police, for example, may affect both the probability of arrest and postarrest satisfaction levels, leading to biased inferences for the population. Early research on police satisfaction that relied on police-citizen encounters was limited in this way because it relied on data from only citizens who came in contact with the police (Reisig and Parks 2000). Estimates of satisfaction with the police from this type of selected sample will be biased if there are any differences in satisfaction between individuals who have had contact with the police and those that have not.

Even research designs that incorporate random experiments are often threatened by selection bias. Although experiments represent the quintessential tool for randomizing observations, controlling for unmeasured variables, and drawing improved causal inferences (Weisburd 2010), they are far from catholicons for sample selection. As Sampson (2010, pp. 490–491) recently argued, randomization is "the trump card that the experimentalist plays in the causal inference game" though too often scholars "focus on the benefits of randomization rather than the selection mechanisms that produce the samples in the first place." His point is that randomization can only address selection bias if the researcher begins with a random sample from the population of interest, and often this is not the case. Even when a random sample is available, though, noncompliance, information diffusion or contagion, or sample attrition, interference, and missing data problems may jeopardize causal inferences in experimental research (Berk 2005; Sampson 2010). For instance, differential nonparticipation or implementation across experimental sites can often reintroduce the types of selection bias that experimentation is designed to overcome.

In the majority of research, though, scholars are limited to nonexperimental or observational data. Spohn and Holleran (2002), for example, compared recidivism outcomes in a sample of drug-involved offenders, some of whom were sentenced to probation and others who received prison. They rightly recognized that "offenders sentenced to prison" are "qualitatively different from offenders placed on probation" and that "offenders deemed more threatening and more likely to return to a life of crime will, all else being equal, face higher odds of incarceration" (Spohn and Holleran 2002, p. 340). The problem is that offenders are not randomly assigned to prison and probation so such analyses risk comparing apples to oranges. The authors attempt to control for these differences by including "each offender's predicted probability of incarceration" in their models of recidivism (341). Although this is not explicitly a sample selection model, it follows the same general logic. Similarly, D'Alessio and Stolzenberg (2003, p. 1390) investigated the impact of race on arrest using NIBRS data for four violent crimes. To simplify the analysis, they excluded cases involving multiple offenders or multiple victims. The problem is that race may not exert the same influence in these cases as in cases with a single offender and single victim. They therefore include a "logit-based 'hazard rate' variable to account for the exclusion of crime incidents with multiple offenders and/or victims" – this is essentially a sample selection model to correct for selection bias.

Smith and Paternoster (1990) explore the role of selection bias more explicitly in their examination of the effect of juvenile placement on future offending. From the perspective of labeling theorists, formal sanctions can trigger secondary deviance, in which an individual's self-identity is transformed to accept the deviant label, leading to "deviance amplification" or increased involvement in future deviant behaviors (Lemert 1951). As these authors point out, though, if youth who are more likely to recidivate are also more likely to receive a sentence to juvenile court, then the relationship between the two could be caused by selection bias. In this case the measure of court referral would be capturing unobserved correlates of recidivism that are not included in the statistical model (e.g., predisposition to delinquency), and the result is biased and inconsistent estimates of the effect of placement on recidivism. If there are any omitted variables that affect both placement and future offending, then it may be that youth who are assigned more severe court dispositions would have been more delinquent regardless of their placement in juvenile court. Using data on juvenile offenders in Florida, Smith and Paternoster initially found a positive effect of juvenile court placement on subsequent referral; however, when they employed sample selection models to control for selection bias, this effect disappeared entirely. They interpret this as evidence that the deviance amplification thesis is largely the result of unaccounted-for selection effects in juvenile case processing.

There are numerous other examples of potential selection effects in criminological research. For instance, Brame (2000) discusses the bias that can characterize missing outcome data in randomized experiments and offers an example of this in the context of police response to domestic disputes. Western and Pettit (2005) use data from inmate surveys to impute missing wages for incarcerated offenders who would otherwise be excluded from earnings estimates. They find that the apparent decrease in the black-white wage gap in recent years is largely the product of increasing selection bias due to greater numbers of young African American males being incarcerated. Rhodes et al. (2001) provides a useful discussion of the use of instrumental variables to address selection bias in the context of nonlinear survival models, which offer an alternative approach for examining such things as treatment effects in offender recidivism studies. Paternoster and Brame (1998) examine select propositions rooted in self-control theory by investigating the association between criminal and analogous behaviors using data from the Cambridge Study. They estimate a bivariate probit model that quantifies the magnitude of the association between criminal and analogous behaviors. Similarly, Sampson et al. (2002, p. 466) in their review of neighborhood-level research suggest that "the issue of selection bias is probably the biggest challenge facing neighborhood-level research." They point out that selection bias can arise from residential mobility decisions, when individuals self-select into neighborhoods. In short, these brief examples serve to demonstrate that sample selection issues characterize a dynamic range of criminological research questions. Sample selection models offer one useful analytical approach for addressing selection bias, but before discussing these models, it is necessary to first understand the different ways that sample selection can occur.

Varieties of Sample Selection

In order to know whether or not sample selection models are appropriate, one must first know what type of selection is present. Sample selection on the dependent variable can occur in two primary ways, which directly affects the type of selection model that should be estimated. Some types of sample selection involve what is referred to as *explicit selection*, which occurs when the selection process is "explicitly" tied to the dependent variable; that is, whether or not a case is observed is a direct function of the score on the dependent variable. Put another way, Y is known only if some criterion defined in terms of Y is met. The researcher does not observe the value of Y for specific values of the dependent variable. Other types of sample selection are characterized by what is referred to as *implicit* or *incidental selection*, which occurs when the selection process results from some external, stochastic function related to the dependent variable. With incidental selection, the likelihood that an observation is excluded from the sample is determined by a separate probabilistic function, and each observation on the dependent variable has some nonzero chance of inclusion. Put another way, Y is observed only if criteria defined in terms of some other random variable is met. Implicit selection processes necessitate different types of selection models than explicit selection processes, so before the researcher can make a decision about how to best account for selection bias, this basic distinction must be understood.

Regardless of whether one is faced with explicit or implicit selection, though, the sample itself can also be characterized by an important distinction between *truncated data* and *censored data*. Truncation occurs when excluded observations provide no data on either the dependent or independent variable. Unobserved cases are unknown for all variables in the data. That is, X is observed only if Y is observed. Censoring, on the other hand, occurs when information is only missing or incomplete on the dependent variable. Information for the other variables in the data is recorded for all cases. That is, X is observed for the entire sample, regardless of whether Y is observed or not. Typically, censored data involves the clustering of observations at some threshold in the data, such that exact values above or below some cutoff are unknown and therefore set to the threshold value. With censored data we have all of the observations on the independent variables, but we don't know the "true" values on the dependent variable if they fall below or above the threshold. These basic distinctions are visually represented in Fig. 1 using the example of sentence lengths that are censored or truncated at 1 year, where each data point represents the sentence of one offender in the data.

To make the discussion more concrete, image that you want to estimate the effect of a criminal record on one's future earnings (e.g., Nagin and Waldfogel 1998). In scenario 1, you use official data that are only collected for a sample of individuals with earnings above the poverty line. This results in explicit selection because the dependent variable, income, determines which cases are observed, and it represents a truncated sample, because no information is known about either prior record or income for omitted cases. In scenario 2, income and criminal history data are collected for all offenders, but incomes below the poverty line are only recorded as "below the poverty line." Exact income values are therefore unknown for offenders who fall below this threshold. In this case, you still have explicit selection based on the dependent variable, but the sample is now censored. Data on criminal history are available for all offenders, but exact information on income is unknown for subjects below the poverty line. In scenario 3, you collect survey data in which income values are missing for a nonrandom subsample of offenders. Sample selection in this case is not an explicit function of the dependent variable in which every value above or below some cutoff value is excluded. Instead, it results from a stochastic process in which the *probability* of being unobserved is nonrandom. For instance, offenders with low incomes may be less likely to answer the survey at all, in which case you have incidental selection with truncation, or perhaps they respond to the survey but refuse to answer the income question, in which case you have incidental selection with censoring. Although different selection processes necessitate different modeling solutions, in general, similar biases result from cases involving both explicit and incidental selection processes.

These basic distinctions are further illustrated in Table 1, which uses the example of criminal sentencing data. Imagine that a researcher is interested in estimating the effect of crime severity on length of imprisonment. The first question is whether there is an explicit or implicit selection process at work. If data on the dependent variable are only collected for prisoners (i.e., offenders sentenced to at least 1 year of incarceration),

Sample Selection Models, Fig. 1 Censored and truncated data distributions

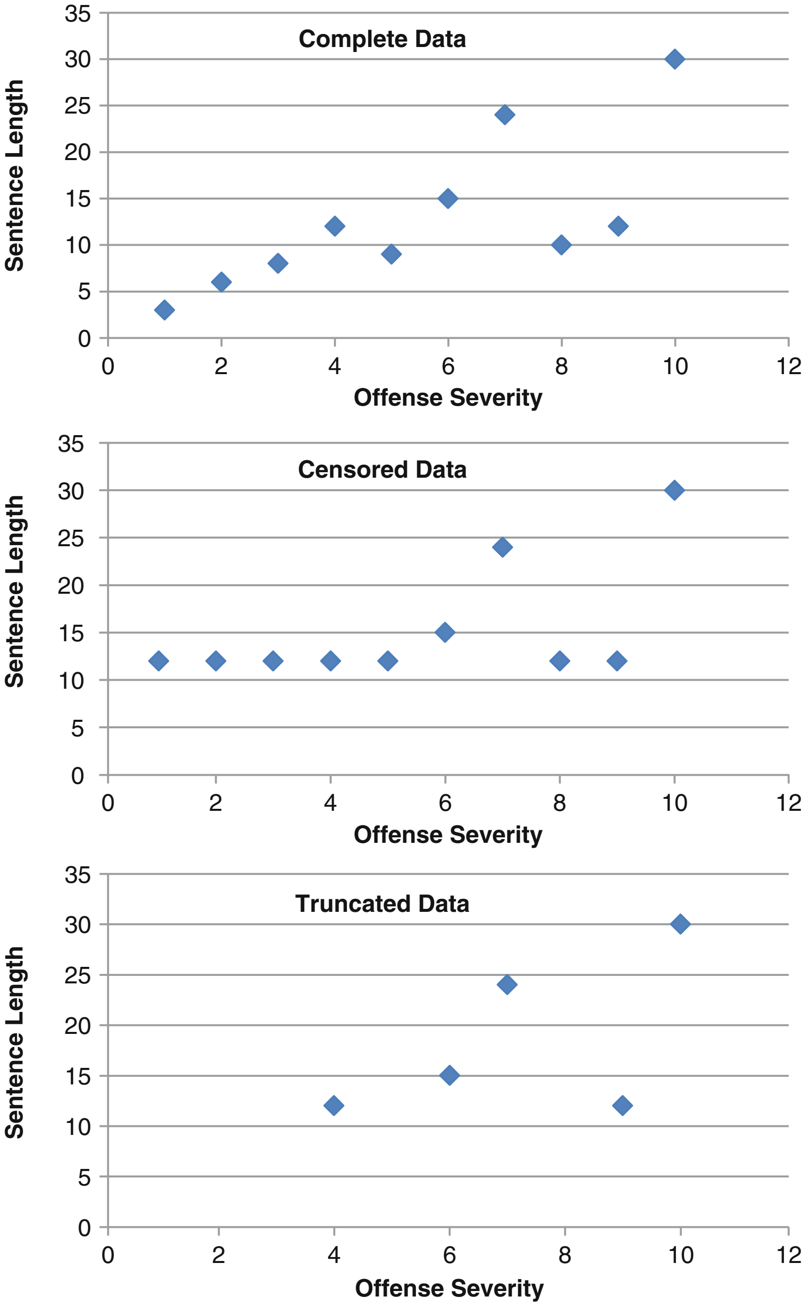

then the researcher is faced with explicit selection. Whether or not the dependent variable is observed is determined fully by the length of imprisonment. The second question then is whether the data are truncated or censored. If data on the independent variable, crime severity, is only collected for prisoners, then the data are truncated – they are missing entirely for offenders who receive less than 12 months of incarceration, as in the upper left-hand box of Table 1. If, however, crime severity data are collected on all convicted offenders, but the dependent variable is only recorded as "less than 1 year" for nonprisoners in the data, then the data are censored,

Sample Selection Models, Table 1 Typology of types of sample selection

		Type of selection	
		Explicit selection	Incidental selection
Type of sample	Truncated	Exact prison lengths are only observed for sentences of 1 year or more. No data are collected from offenders sentenced to less than 1 year of imprisonment	The probability that prison length is observed is a function of other factors related to the likelihood of imprisonment and no data are collected from non-prisoners
	Censored	Exact prison lengths are only observed for sentences of 1 year or more. Data on offense severity are collected from all offenders but sentences are only recorded as “less than 1 year” for non-prisoners	The probability that prison length is observed is a function of other factors related to the likelihood of imprisonment. Data on independent variables are collected from all offenders but prison lengths are only known for prisoners

Sample Selection Models, Table 2 Censored, truncated and sample selected models

	Y	X	Example
Censored (explicit, censored)	Y is known only if some criterion defined in terms of y is met	X is observed for the entire sample, regardless of whether Y is observed or not	Exact prison lengths are recorded only for terms of 1 year or greater. Shorter sentences are recorded as “less than 1 year”
Truncated (explicit, truncated)	Y is known only if some criterion defined in terms of y is met	X is observed only if Y is observed	Prison lengths are only recorded for terms of 1 year or greater. No data are collected for offenders sentenced to less than 1 year of incarceration
Sample selected (incidental, censored)	Y is observed only if criteria defined in terms of some other random variable (Z) are met	X and W (the determinants of whether Z = 1) are observed for the entire sample, regardless of whether Y is observed or not	Prison lengths are recorded only for offenders who are imprisoned and the probability of imprisonment is a function of other factors, such as offense severity

Adapted from Breen (1996)

as in the lower left-hand box in Table 2. Information on the independent variable would be available for all offenders, but exact sentence lengths would be unobserved for offenders sentenced to less than 1 year of incarceration. The alternative to explicit selection is implicit selection. If the likelihood that any given prison length is observed is a function of other factors related to sentencing, then implicit selection is at work. If no data are collected from certain offenders, then the sample is once again truncated as in the upper right-hand box of Table 1. If, however, prison lengths are missing as a function of other sentencing factors related to imprisonment, but offense severity is known for all offenders, then the sample is characterized by incidental selection with censoring as in the lower right-hand box of Table 1.

Figure 2, adapted from Bushway et al. (2007), provides a schematic heuristic outlining the decision-making process involved in selecting an appropriate selection model depending on the type of sample selection that is present. Detailed discussion of these models follows below, but the point here is that it is essential to understand the type of selection one is faced with before deciding how to attempt to correct for selection bias. In the face of explicit selection, the sample selection process is inherently tied to the dependent variable. When data are truncated, truncated

Sample Selection Models, Fig. 2 The process of selecting a sample selection model (Note: Heuristic adapted from Bushway et al. 2007)

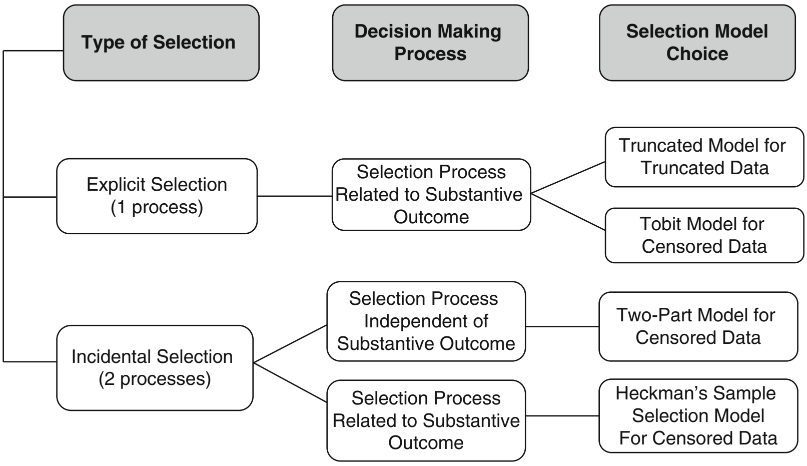
Type of Selection
Decision Making Process
Selection Model Choice
Explicit Selection (1 process)
Selection Process Related to Substantive Outcome
Truncated Model for Truncated Data
Tobit Model for Censored Data
Incidental Selection (2 processes)
Selection Process Independent of Substantive Outcome
Two-Part Model for Censored Data
Selection Process Related to Substantive Outcome
Heckman's Sample Selection Model For Censored Data

regression models can be estimated. When data are censored, Tobit regression models may be appropriate (Tobin 1958). These two approaches are similar except that the Tobit model is able to incorporate information on the independent variables of censored cases in the parameter estimates of the substantive equation, whereas this information is unavailable in truncated samples. When incidental selection is present, one must make a determination about whether or not the selection process is independent of the outcome of interest. If the selection process is unrelated to the outcome, then by definition selection bias is not a problem and two separate, independent models can be estimated, known as the two-part model (discussed below). If, however, the selection process is related to the outcome, then selection bias becomes a concern, and Heckman's two-part sample selection model may provide improved model estimates.

The astute reader will notice that regardless of which type of selection process is operative, under incidental selection, both model choices apply only to censored data. The reason for this is because with truncated data, no information is available on the characteristics of the excluded cases, so it becomes impossible to model the selection process in the presence of incidental selection involving truncated data. Theoretically, it is possible that an exogenous selection process determines which cases are omitted, but if no data is collected from the omitted cases, this process cannot be formally modeled with traditional selection models. For this reason, some treatments (e.g., Breen 1996) of selection bias collapse the four boxes in Table 1 into the three categories summarized in Table 2. Once a researcher knows what type of sample selection is occurring, then it becomes possible to choose an appropriate statistical model. For censored samples, the fundamental logic involves a two-stage modeling procedure that attempts to first capture the selection process and then correct for it in the substantive model of interest.

The Sample Selection Problem

Before turning to an elaboration of common types of statistical models that can be employed when faced with sample selection, it is instructive to first consider the limitations of the ordinary least squares regression model and how analyses conducted on selected subsamples can lead to biased statistical inferences. For illustrative purposes, the ordinary regression model is presented in Eq. 1:

$$Y = \beta X + \varepsilon \tag{1}$$

In this model, Y is a continuous outcome, such as sentence length; X is an exogenous independent variable, such as crime severity; and ε is a normally distributed error term with an

Sample Selection Models, Fig. 3 The effects of truncation on the ordinary regression line and expected values

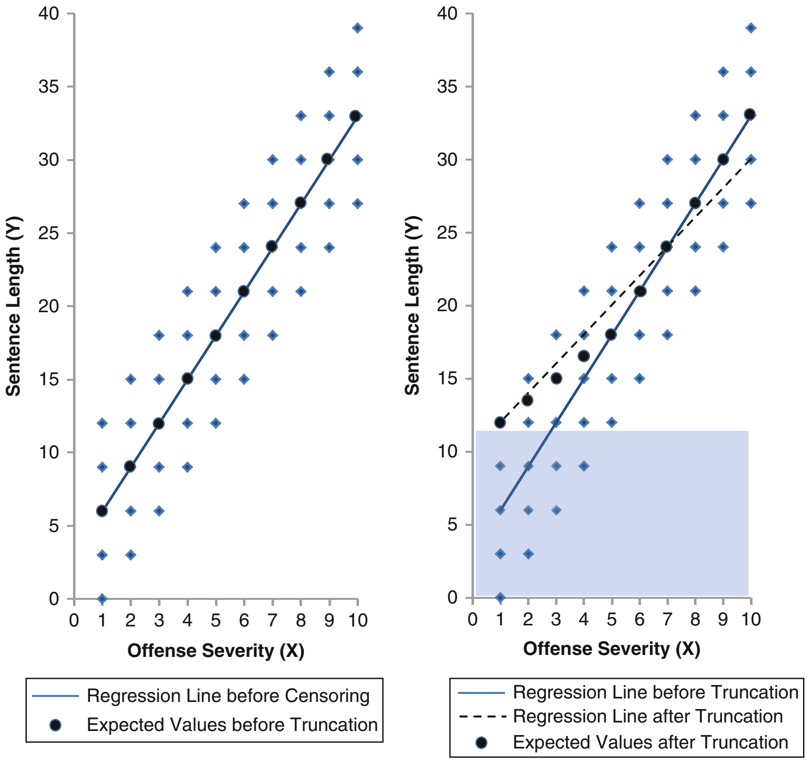

expected value of zero for each value of *X*. We assume that the functional form between *X* and *Y* is linear and additive, that the conditional distribution of *Y* is normal with uniform standard deviations, that observations are independent, and that *X* is uncorrelated with the error term ε. When these assumptions are met, the OLS estimator will provide unbiased and consistent estimates of the model parameters. Selection processes typically introduce an association between the error term and the exogenous variable in the model; in this context, then, selection bias can be understood as a violation of traditional OLS model assumptions.

Imagine that you want to examine the relationship between crime severity and sentence length in a random sample of 50 convicted offenders. The sample selection problem is that not all sentence lengths are observed. If the unobserved cases are comprised of a random subset of cases, then selection bias is not a problem. Typically, though, prison sentences are reserved for confinement sentences of 1 year or more, with shorter sentences being served in county jail. In an analysis of sentence lengths that relies on prison data, then, shorter terms of confinement (and concomitantly less serious crimes) will be systematically excluded from observation. In this example, the researcher would be faced with explicit selection in a truncated sample, though it is just as easy to conceive of related examples that involve implicit selection and/or censored samples and the biases introduced are comparable. This scenario is demonstrated in Fig. 3, which reproduces a similar example discussed in Berk (1983).

The shaded area represents sentence lengths of less than 12 months, which are omitted from the prison data. The solid line is the regression line based on all of the values if sentences of less than 12 months are included, and the dashed line is an approximation of the regression line you might expect based on only the non-truncated observations. Recall that one of the key assumptions of OLS is that the expected value for the disturbance term is zero for all values of *X*. That is, the fitted values from the model should match the means of the observed data. In the complete data shown in the left-hand panel above, the expected values fall directly on the solid regression line. For instance, when offense severity is one, the expected sentence based on the five data points would be 6 months. In the truncated sample, however, the expected values no longer conform to the original regression line, nor do they align with the adjusted line after truncation. For offense severity scores below five (where the truncation occurs), the expected values fall systematically above the original regression line. Sample selection in this case results in expected values that are shifted upward. For instance, when offense severity is one in the truncated sample, the expected sentence is shifted from 6 to 12 months. When offense severity is equal to two, the expected sentence shifts from 9 to 13.5 months and so on. The regression line no longer coincides with the expected values at each value of X, and the errors associated with the regression line will no longer have expected values of zero in this range. The regression line estimated on the truncated sample systematically falls on or above the expected values when offense severity is low, and it falls below the expected values when offense severity is high. The truncation of low values on Y therefore introduces a correlation between the exogenous variable, offense severity, and the error term in the regression model, which violates OLS assumptions and leads to bias in the estimated parameters.

Several observations are readily apparent from Fig. 3. First, the regression line for the truncated observations is different from the regression line for the complete data. Compared to the true relationship for all offenders, the truncated relationship is attenuated. The effect of selection bias in this instance is to underestimate the relationship between offense severity and sentence length. Of course other types of censoring in other scenarios may just as easily lead to overestimates of relationships in the population. Second, the expected values (the black circles) after censoring no longer conform to the expected linear relationship between X and Y. This reflects the fact that the degree of censoring is increased at lower values of the exogenous variable. The regression line, in a sense, is flattened as it nears the censoring floor. This violates another assumption of ordinary least squares, namely, that the relationship between the independent and dependent variable is linear. In this context, then, the problem of selection bias can be understood as a specification error regarding the wrong functional form of the relationship (Berk 1983).

There are several implications to the violations of these model assumptions. First, external validity is compromised. The estimated relationship in the censored sample does not represent the true relationship between offense severity and sentence length for all sentenced offenders. The regression slope is a biased estimate of the true relationship in the population. Second, internal validity is also compromised. This is seldom understood. It is common for researchers to claim an interest in only the selected subsample as a means of deflecting concerns over selection bias, but because the regression line from the truncated sample will systematically overestimate the relationship between X and Y for lower values of offense severity and underestimate it for higher values, the error term will be positively correlated with X, which will produce biased and inconsistent parameter estimates, *even if one claims an interest in only the subsample of observed cases*. In sum, the systematic exclusion of nonrandom observations from a sample is likely to inadvertently induce a relationship between observed variables and the error term, which produces biased and inconsistent parameter estimates. This unintended correlation is at the heart of the selection bias problem in diverse criminological research.

To further demonstrate, imagine an ordinary regression equation where Y is a measure of the number of future arrests, X is whether or not an offender went through some treatment program expected to lower the number of arrests, and β estimates the effect of program participation on future arrests. The least squares solution to β is provided in Eq. 2.1, and the expected value for $\hat{\beta}$ after substitution is given in Eq. 2.2:

$$\hat{\beta} = (\acute{X}X)^{-1}\acute{X}Y \tag{2.1}$$

$$E(\hat{\beta}) = \beta + E[(\acute{X}X)^{-1}\acute{X}\varepsilon] \tag{2.2}$$

If the covariance between X and ε is 0, then the second term on the right-hand side of the equation is zero and the expected value of $\hat{\beta}$ is equal to its true value β. If however X and ε are correlated, then this term does not drop out of the equation and the estimated value for β is no longer equal to its true value. This reflects the influence of selection bias on the program effect in our example. Heckman (1976, 1979) demonstrates that this type of selection problem can be understood in terms of a specification error where the second parameter in Eq. 2.2 represents an omitted variable that is not included in the ordinary regression model. This omitted variable will produce biased estimates of our program effect in the current example. By systematically excluding certain observations, one introduces the need for an additional regressor to capture the selection process. Heckman's solution is the use of a two-step estimator that first models the sample selection process and then introduces an additional regressor to capture the selection process in the outcome equation. As others have shown, the degree of bias introduced by sample selection will depend on a number of factors, including the degree and location of censoring and the strength of the correlation between the exogenous variable and the error terms in the two equations (Stolzenberg and Relles 1990, 1997; Bushway et al. 2007).

State of the Art: Statistical Models for Addressing Sample Selection

The Simple Two-Part Model

To be clear, statistical inferences based on subsets of data are not inherently problematic. If the missing observations comprise a random subset of cases from the population, OLS will continue to provide unbiased and consistent estimates with only a loss of efficiency. In the case of incidental selection with randomly missing data, for example, a simple two-part model (TPM) can be estimated (Duan et al. 1983, 1984). The researcher can model the selection process using a binary outcome model, such as a probit or logit, followed by a separate OLS model for the continuous outcome of interest, which is estimated on the selected subset of observed cases. This two-part model is presented in Eqs. 3.1 and 3.2:

$$P = \gamma Z + \delta \tag{3.1}$$

$$Y = \beta X + \varepsilon \tag{3.2}$$

In this model, P is a dummy variable equal to 1 if the subject is selected and 0 otherwise, and Y is the substantive outcome of interest. Z represents an exogenous independent variable related to the probability of being selected, and all the parameters in Eq. 3.2 are equivalent to Eq. 1. The assumption of the two-part model is that the error term in Eq. 3.2 is uncorrelated with the error term in Eq. 3.1. That is, we must assume that there are no unobserved factors that are related to both the selection process and the substantive outcome of interest. Statistically speaking, we assume that the conditional expectation of ε is zero for all values of X. Under these conditions, there is no selection bias and the simple two-part model above will provide unbiased and consistent estimates of the coefficients of interest. When these assumptions are violated, however, parameter estimates from this "uncorrected" model will be biased and inconsistent. When that is the case, a variety of statistical models can be estimated in an attempt to address the selection effect.

Tobit and Truncated Regression

As Fig. 3 suggests, the Tobit model is appropriate when a researcher is faced with explicit selection with a continuous outcome that is censored or bounded by a limiting value. In some sense, this is the simplest model for addressing sample selection because it involves a single selection process based on the dependent variable (i.e., explicit selection). This model, first proposed by Tobin (1958), accounts for the fact that all values falling above (or below) a certain threshold on the dependent value are observed as part of a continuous distribution, whereas values falling below (or above) some threshold are censored. These censored cases remain in the data and provide information on the independent, but information on the dependent variable is treated as unknown. The model assumes a normal error distribution and a linear relationship between the exogenous predictor and a latent continuous variable, which differs from the observed values because it includes the range of values that fall below the censoring threshold. The relationship between the observed and latent variables is summarized in Eq. 4:

$$\begin{aligned} Y &= Y^* \text{ if } Y^* > c \\ Y &= c \text{ if } Y^* \leq c \end{aligned} \tag{4}$$

Y represents the observed dimension of the dependent variable, Y^* the latent dimension, and c the censoring value given censoring occurs from below. In practice, the censoring threshold can take on different values under different theoretical frameworks and may involve multiple censoring values at either or both ends of the distribution. Scores for the observed values of the dependent variable, then, are equivalent to scores on the latent variable when they are above the censoring value, and when they fall at or below the censoring value, they equal the cutoff value. Conceptually, the Tobit model treats the censored values on the outcome as part of an underlying latent continuum, the values which are treated as unknown for observations that fall below the censoring value. For example, Y^* might represent the latent propensity to offend, which is only realized as actual offending, Y, for offenders who have been arrested in the data. Although numerous observations may share a value of 0 in the data, in theory they have different values on the underlying latent variable. The explanatory variables in the model, then, are linearly related to this latent dimension of the dependent variable Y^* as shown in Eq. 5, which can be rewritten in terms of the observed variable, Y, in Eq. 6:

$$Y^* = \beta X + \varepsilon \tag{5}$$

$$\begin{aligned} Y &= \beta X + \varepsilon \text{ if } Y > c \\ Y &= c \text{ if } Y \leq c \end{aligned} \tag{6}$$

The model assumes the errors ε are independent and normally distributed with a constant variance and can be estimated using maximum likelihood procedures that combine a standard probit model for the censoring of values with an ordinary linear model for the observed values (Osgood et al. 2002). The expected value of Y, then, is equal to the conditional probability of censoring times the conditional expectation of Y given it falls above the censoring threshold. The same explanatory variable is used to characterize both the probability of censoring and the linear relationship of X to Y for uncensored observations. The Tobit model assumes that the error terms from the probit and linear models are a singular normal density (i.e., that they are equal) and that the independent variables and their associated coefficients from the two models are equivalent. Heckman (1976) shows that the Tobit model can be understood as a special case of the two-step sample selection model (discussed below) in which the error terms, explanatory variables, and regression coefficients are constrained to be equal in the selection and substantive equations. Equivalency among the explanatory factors is a key assumption of the Tobit model, though alternative formulations have been proposed that relax these assumptions. For instance, Cragg (1971) offers a reformulation of the model that allows the effects of explanatory variables to vary between the selection and substantive equations, and Smith and Brame (2003) discuss the advantages of this model. Useful elaborations of Tobit regression that are relatively accessible are provided in Breen (1996), Long (1997), and Osgood et al. (2002).

The Tobit model is appropriate when there is explicit selection with censored data. As Table 1 suggests, however, explicit selection may also occur with truncated samples in which case information is missing on both the independent and dependent variables in omitted cases. In such a scenario, truncated regression models can be employed to correct model estimates (represented by the upper left-hand box in Table 1). Because implicit selection processes require information on the explanatory variables to model the stochastic selection process, there is no straightforward solution to sample selection involving truncated samples with implicit selection (represented by the upper right-hand box in Table 1). The truncated regression model is similar to the Tobit model in that the dependent variable Y is only observed if some censoring value c is exceeded. In the Tobit model, however, the selection process is captured using a probit model, whereas in a truncated model the requisite information for the probit equation is now missing. However, the log-likelihood function for the truncated regression model can be estimated using maximum likelihood. It is similar to the Tobit log-likelihood function for the uncensored observations, but the density function for the truncated normal distribution includes an additional term that captures the truncation point. In essence, the model assumes that the conditional distribution of the dependent variable is normal and that the excluded values, if observed, would also follow a normal distribution. The truncated regression model has not been used extensively in criminological research but Hausman and Wise (1977) and Breen (1996) provide useful elaborations of it.

Heckman's Two-Part Sample Selection Model

Perhaps the most widely used approach for addressing sample selection bias in criminology has been Heckman's two-part sample selection model. Figure 3 makes it clear that the model is appropriate when the researcher is faced with incidental selection in a censored sample. Unlike the Tobit model, which assumes a single process characterizes both the selection process and the outcome of interest, Heckman's sample selection model provides for a separate stochastic process that determines selection into the sample. The first step in calculating this model is to estimate a probit model for the selection process, which is then used to control for the probability of being excluded from the sample in the substantive equation of interest. In essence, the two-part model calculates a predicted probability of being excluded from the observed sample, and then it controls for that probability in the substantive equation of interest. The statistical model that summarizes Heckman's two-part solution for selection bias is represented by Eqs. 7.1 and 7.2 below:

$$P = \gamma Z + \delta \tag{7.1}$$

$$Y = \beta X + \sigma\rho\lambda(\beta X) + \varepsilon \tag{7.2}$$

Comparison of Eqs. 7.1 and 7.2 with 3.1 and 3.2 makes it clear that the essential difference between the simple uncorrected two-part model and Heckman's two-part sample selection model is that the latter includes the additional regressor $\sigma\rho\lambda(\beta X)$, where $\lambda(\beta X)$ is equal to $\varphi(\beta X)/(1-\Phi(\beta X))$, or the inverse Mills' ratio, which represents "a monotone decreasing function of the probability that an observation is selected into the sample" (Heckman 1979, p. 156). This is also sometimes referred to as the "hazard rate" because it captures the instantaneous probability of exclusion conditional on being at risk for each observation. The ρ in Eq. 3.2 represents the correlation between the two error terms, δ and ε, which is assumed to follow a bivariate normal distribution, and the σ parameter represents the effect of selection on the outcome Y. If the correlation between the two errors is zero ($\rho = 0$), then the extra parameter drops out of the equation and there is no selection bias.

To calculate Heckman's sample selection model, the probit equation is first estimated and the predicted probabilities are retained and then subsequently used to calculate the inverse Mills' ratio described above. Using the probit equation, expected values for the residuals for each individual in the data are saved, conditional on the observed variables Z, and whether or not an individual was selected ($P = 1$) or not ($P = 0$) (i.e., $E[\delta|X, P]$) (see Heckman 1978, p. 938). If the case is selected (i.e., if $P = 1$), then $E[\delta|X, P = 1]$ is equal to $\frac{\varphi(\hat{P})}{\Phi(\hat{P})}$ where φ and Φ represent the standard normal density and cumulative distribution functions, respectively. If the case is not selected, (i.e., if $P = 0$), then $E[\delta/X, P = 0]$ is equal to $\frac{-\varphi(\hat{P})}{[1-\Phi(\hat{P})]}$. In essence, the conditional residuals are used to capture information for each individual on unobservable factors that are related to the selection process. To illustrate, consider two individuals, one whose predicted value is 0.5 and one whose predicted value is −0.5. Calculating the probabilities of being selected into the sample, we get a value of 0.69 for the first individual and a value of 0.31 for the second. Now imagine that both individuals are selected, even though the first individual had a much higher predicted probability. This suggests that the individual with the lower predicted probability who was nevertheless selected into the sample was more likely to have had unobserved factors that were positively associated with selection. Among cases that are selected, then, those with lower probabilities should have larger (i.e., more positive) residuals. If we plug the hypothetical probabilities above into the expectation for $E[\delta/\mathrm{X}, P]$, we get a value of 1.14 for the case with the lower probability of selection and a value of only 0.51 for the case with the higher probability of selection. The values of the conditional residuals from the selection equation therefore serve to help capture the unobserved factors that are related to selection. In a sense, the Heckman model uses the conditional residuals from the selection equation to create an additional regressor that helps to control for unobserved heterogeneity.

This additional regressor is included in the substantive equation of interest in Eq. 3.2 to correct for selection bias. As Heckman (1979) shows, the least squares estimator will be unbiased but inefficient. Inefficiency is introduced by the censoring of the data, which produces heteroskedasticity in the error terms and leads to downwardly biased estimates of the population variance. This necessitates a correction for the standard errors in the model. According to Bushway et al. (2007), standard errors on average can be underestimated by between 10 % and 30 %. These authors identify a number of common errors that characterize prior attempts to use Heckman's model in criminological applications, such as the miscalculation of the inverse Mills' ratio and failure to properly correct for misestimated standard errors. These issues are now easily addressed by canned routines for Heckman's model that are now readily available in standard statistical packages and properly calculate lambda and correct the standard errors. In Stata, for instance, the basic two-step sample selection model can be estimated as follows: "*heckman Y X*, *twostep select(Z)*" where *Y* is the dependent variable, *X* is the vector of independent variables included in the substantive equation of interest, and Z is the vector of independent variables included in the probit selection equation. All that the researcher has to do is to specify a sample selection model and then identify which exogenous variables to include in the selection and substantive equations, respectively. In addition to the two-step procedure outlined above, the Heckman model can also be estimated using maximum likelihood. In general, the maximum likelihood estimator will be more efficient than the two-stage estimator though it may be more sensitive to violations of model assumptions (Bushway et al. 2007).

Importantly, when *Z* and *X* contain the same set of independent variables, the Heckman sample selection model is only identified by the nonlinearity introduced by the inverse Mills' ratio. Prior research demonstrates that this can lead to severe inflation of standard errors, which results from collinearity between the correction term and the other regressors in the model (Stolzenberg and Relles 1990). When the same predictors are used in the selection and substantive equations, the sigma term in Eq. 7.2 will be correlated with *X*. Therefore, whenever possible, *exclusion restrictions* should be included in the sample selection model. Exclusion restrictions are similar to instrumental variables – they are predictors that are related to the selection process but not related to the substantive outcome. For example, Albonetti (1991) included evidentiary measures in her selection equation for the likelihood of conviction, but not her substantive equation for sentence severity. Similarly, Smith and Paternoster (1990) included the caseload and number of diversionary programs in their selection model for juvenile placement but not for their substantive equation examining recidivism. The key is to identify factors that are theoretically related to the selection process but not to the outcome of interest, and whenever possible, researchers should search for and include valid exclusion restrictions when estimating this model.

The Bivariate Probit Model

Heckman's sample selection model applies to the case in which the substantive outcome of interest is a continuous variable. When the dependent variable is dichotomous, as is often the case in criminological research, then an alternative specification can be employed. The bivariate probit model involves estimation of two probit equations allowing their error terms to be correlated. This is useful when there is a binary selection process and a binary outcome in the substantive equation of interest (Heckman 1976). Whenever two dichotomous dependent variables are jointly determined, or when there is reason to believe that error terms of two binary outcomes are correlated, then the bivariate probit model can offer a useful approach for modeling this dependence. The basic model is summarized in Eqs. 8.1 and 8.2:

$$P_2 = \gamma Z + \delta \tag{8.1}$$

$$P_1 = \beta X + \varepsilon \tag{8.2}$$

If the two error terms in Eqs. 8.1 and 8.2 are independent (i.e., $Cov(\delta, \varepsilon) = 0$), then selection bias is not a problem and the two equations can be estimated separately. If, however, the two errors are related (e.g., $(\delta = \eta + u_1)$ and $(\varepsilon = \eta + u_2)$), so that part of the error (i.e., (η)) is common to both equations, then the same type of selection bias discussed above will bias model estimates. When this is the case, interest shifts to the joint probability of P_2 and P_1. The model assumes the error terms in each equation are normally distributed and that their joint distribution is bivariate normal, and as with the Heckman model above, the association between errors is represented by the correlation coefficient rho (ρ). The bivariate probit can be estimated using maximum likelihood procedures that are readily available in common statistical packages. In Stata, for instance, the bivariate probit model can be estimated as follows: "*heckprob P_1 X, select(P_2)*" where P_1 is the second probit outcome, X is the vector of independent variables included in that equation, and Z is the vector of independent variables included in the first probit equation.

The bivariate probit can be used in basic applications to allow the two outcomes to be correlated, or it can be used equivalently to the Heckman selection model to capture a binary selection process that determines inclusion in a binary substantive outcome of interest. Paternoster and Brame (1998), for instance, investigated whether or not criminal behavior was associated with analogous noncriminal behavior (Paternoster and Brame 1998) where both criminal and analogous behaviors were measured as dichotomous outcomes. They used a bivariate probit to model the joint distribution of the two outcomes in order to estimate the strength of their association. Berinsky (2004), on the other hand, utilized the bivariate probit to explicitly correct for sample selection, where the first model was whether or not a respondent answered a survey question and the second was a dichotomous measure of racial attitudes. For the bivariate probit, it is essential that valid exclusion restrictions be included. Dubin and Rivers (1990) discuss application of the binary selection model, and Sartori (2003) offers an alternative estimator that may be preferable if exclusion restrictions are unavailable.

Contemporary Controversies

A number of scholars both in and outside of criminology have discussed at length the potential strengths and weaknesses of sample selection models (Bushway et al. 2007; Winship and Mare 1992; Duan et al. 1983, 1984; Manning et al. 1987; Leung and Yu 1996; Jones 2000; Norton et al. 2008). As Breen (1996, p. 58) summarized, "it is now clear that these models need to be approached with a degree of care." Sample selection models work the best when there are strong exclusion restrictions and when distributional assumptions are clearly met. As others have shown, routinized application of sample selection models without careful circumspection can produce estimates that are no better and sometimes worse than those obtained through ordinary regression techniques obtained from simple two-part models (Duan et al. 1983, 1984). A number of potential concerns need to be considered when deciding whether or not it is advisable to employ a sample selection model.

First, sample selection models have been shown to be sensitive to violations of their distributional assumptions, including heteroskedasticity of errors and violations of normality. Whereas OLS estimates are consistent but not efficient in the presence of heteroskedasticity, estimates from sample selection models are neither consistent nor efficient (Amemiya 1985). Alternative functional forms can be specified for error distributions to address the first issue (Maddala 1983), and a semi-parametric or non-parametric methods that relax model assumptions might be used to address the latter (Heckman et al. 1999; Vella 1999; Moffitt 1999; Chay and Powell 2001). Winship and Mare (1992) discuss some of these approaches, though applications of them remain rare in criminological work. Still, the first step in deciding whether or not sample selection models are appropriate is to examine the validity of distributional assumptions regarding homoskedasticity and normality – Chesher and Irish (1987) provide a useful discussion of how this can be approached. Even if we assume that distributional assumptions are met, though, the performance of sample selection models is affected by additional factors that include the strength of the correlation (ρ) between the error terms in the selection and substantive equations, the degree of overlap in the variables included in the selection and substantive equations, and the size of the sample and degree of censoring in the data. The stronger the correlation is between errors, the larger the threat of sample selection bias and, in general, the greater need for sample selection corrections. Therefore, a researcher needs to have strong theory regarding the association between the selection process and the outcome of interest to help guide the decision to use sample selection models.

Second, it is important for researchers to carefully search for, identify, and incorporate theoretically valid exclusion restrictions into their sample selection models. A number of commentators have expressed reservations about identifying the model using only the nonlinear functional form (e.g., Berk and Ray 1982). Doing so typically leads to high variance estimates and often results in problematic levels of multicollinearity between the correction term and the other independent variables in the model. The problem tends to be exacerbated when the selection model poorly fits the data and has low levels of explained variance. If we do a poor job of capturing the selection process, then the correction for selection will have limited utility.

Third, the performance of sample selection models is also tied to the degree of censoring in the data. In samples with minimal censoring, the sample selection models may not be needed. Low degrees of censoring will introduce less bias into parameter estimates than high degrees of censoring. However, in samples with high degrees of censoring, these models can also be problematic. Stolzenberg and Relles (1990, p. 406), for instance, examined the effectiveness of sample selection corrections in small samples with severe censoring and concluded that "Heckman's method reduced the accuracy of coefficient estimates as often as it improved them." They concluded that "there is considerable evidence that the method can easily do more harm than good, and that its careless and mechanical application runs much danger" (Stolzenberg and Relles 1990, pp. 408–409). To abandon sample selection models altogether, however, would be to throw out the baby with the bathwater. In general, larger samples are preferable to smaller samples when implementing sample selection corrections, whereas the simple uncorrected two-part model is likely to be preferred when explained variance is low or when the estimate of the correlation between error terms is near zero. When sample selection models are deemed appropriate, the maximum likelihood estimators will be more efficient than Heckman's original two-step approach. In practice, the effects of the size of rho, the overlap in the predictors in the two equations, and the sample size and degree of censoring are often intertwined and can be difficult to disentangle. The researcher is therefore often faced with the unhappy decision to either estimate the simple two-part model, which will knowingly produce biased estimates, or alternatively to implement a sample selection model that can under certain conditions make estimates worse. Luckily, there are several approaches that can help assess the feasibility of implementing sample selection models to correct for selection bias.

Lueng and Yu (1996, 2000) discuss the conditions under which sample selection models can suffer from multicollinearity problems, and they discuss some practical approaches for dealing with this problem. Specifically, they recommend researchers calculate the "condition number" as a measure of collinearity, and they recommend that condition numbers over 20 indicate Heckman's correction is likely to perform worse than simple uncorrected two-part models. Other diagnostic tests are also available. For instance, Manski (1990) outlines a useful approach for bounding estimates of selection bias (see also Nagin and Manski 1999). The interval of the bounds varies inversely with the degree of censoring so lower proportions of censored cases result in narrower bounding ranges. Winship and Mare (1992) discuss his method and related approaches. More recently, Stolzenberg and Relles (1997) have outlined a useful methodology for estimating the amount of bias associated with specific parameter estimates in a model, and Bushway et al. (2007) demonstrate the utility of this approach by applying it to the context of criminal sentencing. The approach allows one to quantify the amount of bias introduced by the selection process, providing for better informed decisions regarding the necessity of using sample selection models in place of ordinary regression. Before deciding whether or not to implement sample selection models, then, it is instructive to first examine distributional assumptions, to carefully search for valid exclusion restrictions, and to conduct exploratory analyses that help assess the degree of the problem and the necessity of addressing it.

Conclusion

Sample selectivity has long been a topic of applied theoretical and methodological work in sociology and economics (Heckman 1974; Winship and Mare 1992; LaLonde 1986; Berk 1983), and in recent years it has become a popular analytical technique in criminology and criminal justice (Bushway et al. 2007). Although sample selection models can provide a very useful approach for addressing important selection biases that are pervasive in criminological research, "Infallible models for sample selection bias do not exist" (Winship and Mare 1992, p. 347) and "whatever happens to be the methodological tool du jour is too often oversold" (Berk 2005, p. 2). This suggests the need for careful consideration of the potential strengths and weaknesses of sample selection corrections in applied criminological work. As Winship and Mare (1992, p. 329) argue, "models for selection bias are only as good as their assumptions about the way that selection occurs, and estimation strategies are needed that are robust under a variety of assumptions."

Numerous extensions of these basic models are continuously being developed. Outcome measures are at times available for both the selected and unselected samples (e.g., Spohn and Holleran 2002). Moreover, it is not uncommon for certain applications to include multiple selection points across multiple outcomes of interest. Research on the criminal justice system, for instance, inherently involves multiple censoring points across successive stages of criminal case processing (Kingsnorth et al. 1999). In some applications, multiple censoring or truncation points may even be different for different subjects in a sample, and in general, the basic models discussed herein can readily accommodate such complexities without drastically increasing the difficulty of their application (Breen 1996; Maddala 1983).

The challenge for criminologists faced with sample selection then is to be diligent and circumspect, carefully formulating a theoretical model of the selection process before selecting an appropriate modeling strategy. Importantly, there are a number of different approaches that can be useful for addressing different forms of selection bias, such as experiments, statistical matching procedures, and instrumental variable approaches (e.g., Bushway and Apel 2010; Loughran and Mulvey 2010; Taxman and Rhodes 2010), and in many ways models for sample selection share considerable affinity with other quasi-experimental approaches for addressing selection bias (Lalonde 1986; Heckman and Robb 1986). Researchers should therefore investigate sample selection concerns through a variety of complementary methods whenever possible. Smith and Paternoster (1990) offer an excellent example of this type of research. They rightfully argue that "there is no one generic cure for selectivity bias and that applications of these models must make substantive sense in the context of specific applications" (Smith and Paternoster 1990, p. 1125). Similarly, Winship and Mare (1992, p. 328) conclude that "The large number of available methods and the difficulty of modeling selection indicate that researchers should be explicit about the assumptions behind their methods and should present results that derive from a variety of methods." The sample selection models discussed in this entry offer one useful approach for investigating and potentially correcting for selection bias when researchers are faced with censored or truncated data on their dependent variable of interest.

Related Entries

- ▶ Sample Selection Problems

Recommended Reading and References

Albonetti CA (1991) An integration of theories to explain judicial discretion. Soc Probl 38:247–266

Amemiya T (1985) Advanced econometrics. Harvard University Press, Cambridge, MA

Angrist J (2001) Estimation of limited dependent variable models with dummy endogenous regressors: simple strategies for empirical practice. J Bus Econ Stat 19:2–16

Apel R, Bushway S, Paternoster R, Brame R, Sweeten G (2008) Using state child labor laws to identify the causal effect of youth employment on deviant behavior and academic achievement. J Quant Criminol 24:337–362

Berinski AJ (2004) Public opinion and political participation in America. Princeton University Press, Princeton

Berk R (1983) An introduction to sample selection bias in sociological data. Am Socio Rev 48:386–398

Berk R (2005) Randomized experiments as the bronze standard. J Exp Criminol 1:417–433

Berk R, Ray S (1982) Selection bias in sociological data. Soc Sci Res 11:352–398

Brame R (2000) Investigating treatment effects in a domestic violence experiment with partially missing outcome data. J Quant Criminol 16:283–314

Breen R (1996) Regression models: censored, sample selected, or truncated data. Sage, Thousand Oaks, CA

Bushway SD, Apel RJ (2010) Instrumental variables in criminology and criminal justice. In: Piquero AR, Weisburd D (eds) Handbook of quantitative criminology. Springer, New York, pp 595–612

Bushway S, Johnson BD, Slocum LA (2007) Is the magic still there? The use of Heckman two-step correction for selection bias in criminology. J Quant Criminol 23:151–178

Chay KY, Powell JL (2001) Semiparametric censored regression models. J Econ Perspect 15:29–42

Chesher A, Irish M (1987) Residual analysis in the grouped and censored normal linear model. J Econometrics 34:33–61

Cragg J (1971) Some statistical models for limited dependent variables with applications to the demand for durable goods. Econometrica 25:241–262

D'Alessio S, Stolzenberg L (2003) Race and the probability of arrest. Soc Forces 84:1381–1398

Duan N, Manning WG, Morris CN, Newhouse JP (1983) A comparison of alternative models for the demand for medical care. J Bus Econ Stat 1:115–126

Duan N, Manning WG, Morris CN, Newhouse JP (1984) Choosing between the sample-selection model and the multi-part model. J Bus Econ Stat 2:283–289

Dubin JA, Rivers D (1990) Selection bias in linear regression, logit and probit models. In: Fox J, Long JS (eds) Modern methods of data analysis. Sage, Thousand Oaks, pp 411–443

Gondolf EW (2000) A 30-month follow-up of court-referred batterers in four cities. Int J Offender Ther Comp Criminol 44:111–128

Greene WH (1993) Econometric analysis, 2nd edn. Macmillan, New York

Hagan J (1990) The gender stratification of income inequality among lawyers. Soc Forces 68:835–855

Hagan J, Palloni A (1986) 'Club Fed' and the sentencing of white-collar offenders before and after Watergate. Criminology 24:603–622

Hagan J, Parker P (1985) White-collar crime and punishment: the class structure and legal sanctioning of security violations. Am Sociol Rev 50:302–316

Hausman J Wise DA (1977) Social experimentation, truncated distributions, and efficient estimation. Econometrica 45:919–938

Heckman J (1974) Shadow prices, market wages, and labor supply. Econometrica 42:679–694

Heckman JJ (1976) The common structure of statistical models of truncation, sample selection and limited dependent variables. Ann Econ Soc Meas 5:475–492

Heckman J (1978) Dummy endogenous variables in a simultaneous equation system. Econometrica 46:931–959

Heckman JJ (1979) Sample selection bias as a specification error. Econometrica 47:153–161

Heckman J, Robb R (1986) Altemative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes. In: Wainer H (ed) Drawing inference from self-selected samples. Springer, New York, pp 63–107

Heckman JJ, LaLonde R, Smith J (1999) The economics and econometrics of active labor market programs. In: Ashenfelter O, Card D (eds) Handbook of labor economics, vol 3A. North-Holland, Amsterdam, pp 1865–2097

Jones AM (2000) Health econometrics. In: Culyer AJ, Newhouse JP (eds) Handbook of health economics, 1st edn, vol 1, chapter 6. Elsevier, pp 265–344

Kingsnorth R, MacIntosh R, Wentworth J (1999) Sexual assault: the role of prior relationship and victim characteristics in case processing. Justice Q 16:275–302

Klepper S, Nagin D, Tierney LJ (1983) Discrimination in the criminal justice system: a critical appraisal of the literature. In: Blumstein A, Cohen J, Martin S, Tonry M (eds) Research on sentencing: the search for reform, vol I1. National Academy, Washington, DC

LaLonde RJ (1986) Evaluating the econometric evaluations of training programs with experimental data. Am Econ Rev 76:604–620

Laub H, Sampson RJ (2003) Shared beginnings, divergent lives: delinquent boys to age 70. Harvard University Press, Cambridge

Lemert EM (1945) Social pathology: systematic approaches to the study of sociopathic behavior. McGraw-Hill, New York

Leung SF, Yu S (1996) On the choice between sample selection and two-part models. J Econ 72:197–229

Leung SF, Yu S (2000) Collinearity and two-step estimation of sample selection models: problems, origins, and remedies. Comput Econ 15:173–199

Little RJA, Rubin DB (1987) Statistical analysis with missing data. J Wiley & Sons, New York

Long JS (1997) Regression models for categorical and limited dependent variables. Sage, Thousand Oaks

Loughran T, Mulvey E (2010) Estimating treatment effects: matching quantification to the question. In: Weisburd D, Piquero A (eds) Handbook of quantitative criminology. Springer, New York, pp 163–181

Lundman R, Kaufman R (2003) Driving while black: effects of race, ethnicity, and gender on citizen self-reports of traffic stops and police actions. Criminology 41:195–220

Maddala GS (1983) Limited-dependent and qualitative variables in econometrics. Cambridge University Press, Cambridge

Manning WG, Newhouse JP, Duan N, Keeler E, Benjamin B, Leiboqitz A, Marquis AS, Zwanziger J (1987) Health insurance and the demand for medical care: evidence from a randomized experiment, Publ. No. R-3476-HHS. Rand Corporation, Santa Monica

Manski CF (1990) Nonparametric bounds on treatment effect. Am Econ Rev 80:319–323

Maxwell C, Garner J, Fagan J (2002) The preventative effects of arrest on intimate partner violence. Criminol Public Policy 2:51–80

Moffitt R (1999) Econometric methods for labor market analysis. In: Ashenfelter O, Card D (eds) Handbook of labor economics, vol 3A. North-Holland, Amsterdam

Mosher CJ, Miethe TD, Hart TC (2011) The mismeasure of crime, 2nd edn. Sage, Thousand Oaks, CA

Nagin D, Manski C (1999) Bounding disagreements about treatment effects: a case study of sentencing and recidivism. Sociol Methodol 28:99–138

Nagin D, Waldfogel J (1998) The effect of conviction on income through the life cycle. Int Rev Law Econ 18:25–40

Norton EC, Dow WH, Do YK (2008) Specification tests for sample selection and two-part models. Health Serv Outcomes Res Methodol 8:201–208

Osgood DW, Finken L, McMorris B (2002) Analyzing multiple item measures of crime and deviance II: Tobit regression analysis of transformed scores. J Quant Criminol 18:859–866

Paternoster R, Brame R (1998) The structural similarity of processes generating criminal and analogous behaviors. Criminology 36:633–669

Peterson RD, Hagan J (1984) Changing conceptions of race: toward an account of anomalous findings of sentencing research. Am Sociol Rev 49:56–70

Puhani PA (2000) The Heckman correction for sample selection and its critique. J Econ Surv 14:53–68

Reisig MD, Parks RB (2000) Experience, quality of life, and neighborhood context: a hierarchical analysis of satisfaction with police. Justice Q 17:607–630

Rhodes W (1991) Federal criminal sentencing: some measurement issues with application to pre-guideline sentencing disparity. J Crim Law Criminol 81:1002

Rhodes W, Pelissier B, Gaes G, Saylor B, Camp S, Wallace S (2001) Alternative solutions to the problem of selection bias in an analysis of federal residential drug treatment programs. Eval Rev 25:331–369

Sampson RJ (2010) Gold standard myths: observations on the experimental turn in quantitative criminology. J Quant Criminol 25:489–500
Sampson RJ, Morenoff JD, Gannon-Rowley T (2002) Assessing "neighborhood effects": social processes and new directions in research. Annu Rev Sociol 28:443–78
Sartori AE (2003) An estimator for some binary-outcome selection models without exclusion restrictions. Polit Anal 11:111–138
Smith DA, Brame R (2003) Tobit models in social science research: some limitations and a more general alternative. Sociol Method Res 31:364–388
Smith DA, Paternoster R (1990) Formal processing and future delinquency: deviance amplification as selection artifact. Law Soc Rev 24:1109–1131
Spohn C, Holleran D (2002) The effect of imprisonment on recidivism rates of felony offenders: a focus on drug offenders. Criminology 40:329–358
Stolzenberg RM, Relles DA (1990) Theory testing in a world of constrained research design: the significance of Heckman's censored sampling bias correction for nonexperimental research. Sociol Method Res 18:395–415
Stolzenberg RM, Relles DA (1997) Tools for intuition about sample selection bias and its correction. Am Sociol Rev 62:494–507
Taxman FS, Rhodes A (2010) Multisite trials in criminal justice settings: trials and tribulations of field experiments. In: Piquero A, Weisburd D (eds) Handbook of quantitative criminology. Springer, NewYork, pp 519–543
Tobin J (1958) Estimation of relationships for limited dependent-variables. Econometrica 26:24–36
Ulmer JT, Johnson BD (2004) Sentencing in context: a multilevel analysis. Criminology 42:137–178
Vella F (1999) Estimating models with sample selection bias: a survey. J Hum Resour 33:127–169
Weisburd D (2010) Justifying the use of non-experimental methods and disqualifying the use of randomized controlled trials: challenging folklore in evaluation research in crime and justice. J Exp Criminol 6:209–227
Western B, Pettit B (2005) Black-white wage inequality, employment rates, and incarceration. Am J Sociol 111(2):553–578
Wilson DB (2010) Meta-analysis. In: Weisburd D, Piquero A (eds) Handbook of quantitative criminology. Springer, New York, pp 181–208
Winship C, Mare RD (1992) Models for selection bias. Annu Rev Sociol 18:327–350
Worral JL (2002) If you build it, they will come: consequences of improved citizen complaint review procedures. Crime Delinq 48:355–379
Zatz MS, Hagan J (1985) Crime, time, and punishment: an exploration of selection bias in sentencing research. J Quant Criminol 1:103–126

Sample Selection Problems

Robert Apel
School of Criminal Justice, Rutgers University,
Newark, NJ, USA

Overview

In a seminal series of contributions that would culminate in his being awarded the Nobel Prize in Economics in 2000, James Heckman proposed a deceptively simple solution to the problem posed by sample selection bias (Heckman 1976, 1979). In the classic formulation of the problem, the analyst desires to estimate a wage or earnings function from a collection of individuals sampled randomly from the population. While one wishes to generalize to the population of all wage offers, it is obvious that wages are only observed for individuals who are employed. The problem that arises is that these individuals are a self-selected (i.e., nonrandom) subset of the population, and consequently their wages cannot necessarily be used to estimate what the wages of nonworking individuals would have been, had they chosen to work. So a regression of (log) wages on a determinant such as schooling, limited to the subsample of workers, will potentially yield biased estimates of the impact of schooling on wages for the population of interest.

In criminology, a similar challenge is encountered in studies of the determinants of criminal sentencing. For example, one might be interested in studying the impact of legally irrelevant factors (e.g., gender, race) on sentencing decisions once legally relevant factors (e.g., offense seriousness, criminal history) have been netted out. However, sentence lengths are only observed for convicted offenders who are sent to prison. The determinants of sentence length among these individuals are not necessarily representative of what the determinants of sentence length would have been among non-incarcerated individuals, had they instead been incarcerated. Whereas the population of interest entails all criminal sentences rather than strictly carceral ones, the latter are a decidedly nonrandom subset of all convicted offenders. Many other examples that have criminological relevance are provided in Berk (1983), Winship and Mare (1992), Bushway et al. (2007).

Heckman (1976, 1979) formulated the problem as one of missing data giving rise to specification error. If the process that determines whether an individual is "uncensored" with respect to the outcome of interest (e.g., employed, incarcerated) is correlated with the process that determines the mean outcome, then the least squares estimator is biased and inconsistent. Heckman's insight was that exclusion of censored units from a sample will lead to bias and inconsistency, even if one is only interested in generalizing to the population of uncensored units (Berk 1983).

This chapter is devoted to an investigation of the sample selection problem, a presentation of the various solutions that have been proposed, and a discussion of the limitations of these solutions. An empirical illustration will then be provided, using data from the National Longitudinal Survey of Youth 1997 to examine the relationship between criminal behavior and legal earnings.

The Sample Selection Problem

The sample selection problem is one of estimating the relationship between two or more variables for a population, when information on the dependent variable is only observed for a nonrandom subset of a sample from that population. Consider estimating a model of earnings from employment. The population equation of interest is the following:

$$\begin{aligned} Y_i^* &= \beta_0 + \beta_1 X_{1i} + \cdots + \beta_j X_{ji} + e_i \\ &= \boldsymbol{X\beta} + e_i \end{aligned} \tag{1}$$

where the coefficients and regressors are consolidated into **Xβ** for notational convenience. Here, $\mathrm{Y_i}^*$ is a random variable representing *latent earnings*. The sample selection problem arises because $\mathrm{Y_i}^*$ is only partially observed – earnings are unobserved or censored among the subsample of individuals who are not employed. A second random variable, which can be thought of as *employment propensity*, determines whether earnings are observed or censored. This is also a latent variable, denoted $\mathrm{D_i}^*$, that is parameterized in the following manner:

$$\begin{aligned} D_i^* &= \alpha_0 + \alpha_1 Z_{1i} + \cdots + \alpha_k Z_{ki} + u_i \\ &= \boldsymbol{Z\alpha} + u_i \end{aligned} \tag{2}$$

where the coefficients and regressors are consolidated into $\boldsymbol{Z\alpha}$ for notational convenience. Concerning the residuals from these two equations, $\mathrm{e_i}$ is assumed to be distributed normal and $\mathrm{u_i}$ is assumed to be distributed standard normal. Their joint distribution is assumed to be bivariate normal:

$$\begin{pmatrix} e_i \\ u_i \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_e^2 & \sigma_{e,u} \\ \sigma_{e,u} & 1 \end{pmatrix}\right]$$

As an aside, the correlation between the unobserved determinants of the two latent response variables $\mathrm{Y_i}^*$ and $\mathrm{D_i}^*$ is provided by the equation:

$$\rho_{e,u} = \frac{\sigma_{e,u}}{\sigma_e}$$

Berk (1983) refers to Eq. 1 as the "substantive equation," and to Eq. 2 as the "selection equation." Of course, neither $\mathrm{Y_i}^*$ nor $\mathrm{D_i}^*$ are fully observed. Instead, the two latent variables have observed counterparts. Namely, latent earnings $(\mathrm{Y_i}^*)$ are only observed when employment propensity $(\mathrm{D_i}^*)$ exceeds a defined threshold:

$$Y_i = \begin{cases} 0 & \mathit{if}\ D_i^* < 0 \\ Y_i^* & \mathit{if}\ D_i^* > 0 \end{cases}$$

Note that the choice of 0 as the threshold is a convenient normalization. If we substitute the linear predictor in Eq. 2 for $\mathrm{D_i}^*$ and rearrange terms, it can be shown that latent earnings are only observed if:

$$u_i > -\mathbf{Z}\boldsymbol{\alpha}$$

Taking the expectation of the observed Y_i for the subsample with nonzero earnings, we now have the following:

$$E\left(Y_i | W_i, D_i^* > 0\right) = \boldsymbol{X\beta} + E(e_i | u_i > -\mathbf{Z}\boldsymbol{\alpha}) \quad (3)$$

The last term in Eq. 3 illustrates the source of the sample selection problem and is the key insight provided by Heckman (1976, 1979). Namely, a least squares regression model of earnings, fitted only for the subsample with nonzero earnings, can lead to omitted variables bias because of the exclusion of a potentially relevant regressor – the conditional mean of the latent earnings residual for the uncensored subsample. If this last term is 0, there is no omitted variables bias, and inferences based on the subsample with nonzero earnings are valid for the general population, including those with censored earnings. Obviously, the sample selection problem arises when the conditional mean is not 0.

Consider the implications of the above for the question of interest in this chapter – the relationship between arrest and employment earnings. It is plausible that this relationship, estimated only from the subsample of individuals with nonzero earnings, is rendered biased and inconsistent because of the fact that these individuals are not a random subset of the sample. Specifically, the unobservables which determine whether or not someone works are likely to be correlated with the observables which determine someone's mean earnings, conditional on working. This means that regressors which are truly correlated with earnings might be statistically insignificant when the model is estimated from individuals with nonzero earnings. Similarly, regressors which are truly uncorrelated with earnings might be statistically significant.

Heckman's Two-Step Solution

To overcome the sample selection problem, Heckman (1976, 1979) proposed an elegant, two-step solution – the so-called Heckman selection correction. His solution entails direct estimation of the selection term in Eq. 3 and its inclusion as a regressor in the equation of substantive interest. The first step of this procedure is to estimate a model of the probability of being uncensored, in this case, the probability of having nonzero earnings. This begins by linking employment propensity (D_i^*), which is unobserved, to a binary response variable according to the following rule:

$$D_i = \begin{cases} 0 & \mathit{if}\ D_i^* < 0 \\ 1 & \mathit{if}\ D_i^* > 0 \end{cases}$$

This binary variable is then treated as the dependent variable in a probit regression model:

$$\begin{aligned} \Phi^{-1}[Pr(D_i = 1)] &= \alpha_0 + \alpha_1 Z_{1i} + \cdots \\ &\quad + \alpha_k Z_{ki} \\ &= \mathbf{Z}\boldsymbol{\alpha} \end{aligned} \quad (4)$$

The second step of Heckman's procedure requires inserting a transformation of the linear predictor from Eq. 4 into the substantive equation, estimated from the subsample with positive earnings:

$$\begin{aligned} (Y_i | D_i = 1) \\ &= \beta_0 + \beta_1 X_{1i} + \cdots + \beta_j X_{ji} \\ &\quad + \beta_{j+1}\lambda(-\mathbf{Z}\boldsymbol{\alpha}) + e_i \end{aligned} \quad (5)$$

The new term, $\lambda(\cdot)$, is known as the "inverse Mills ratio" (IMR) of the argument:

$$\lambda(-\mathbf{Z}\boldsymbol{\alpha}) = \frac{\phi(-\mathbf{Z}\boldsymbol{\alpha})}{1 - \Phi(-\mathbf{Z}\boldsymbol{\alpha})} = \frac{\phi(\mathbf{Z}\boldsymbol{\alpha})}{\Phi(\mathbf{Z}\boldsymbol{\alpha})}$$

The numerator $\phi(\cdot)$ represents the standard normal density function (p.d.f.) evaluated at the argument, whereas the denominator $\Phi(\cdot)$ represents the standard normal distribution function (c.d.f.) evaluated at the same argument. In empirical applications, of course, the true IMR is unknown. Instead, the estimate of the linear predictor from the probit regression model in Eq. 4 is substituted into the ratio:

$$\hat{\lambda}(-\mathbf{Z\alpha}) = \frac{\phi(\hat{\alpha}_0 + \hat{\alpha}_1 Z_{1i} + \cdots + \hat{\alpha}_k Z_{ki})}{\Phi(\hat{\alpha}_0 + \hat{\alpha}_1 Z_{1i} + \cdots + \hat{\alpha}_k Z_{ki})}$$

The IMR has a number of important properties. For one, it is nonnegative and is therefore bounded by 0 and ∞. In addition, because the denominator yields the estimated probability that an individual is uncensored (i.e., has positive earnings), individuals with a small probability of being uncensored will tend to have a large λ(·). Secondarily, the parameter β_{j+1} is formally identified in Eq. 5, even when the same set of regressors is included in both the first and second steps of the procedure. This is because λ(·) is a nonlinear transformation of the parameter estimates from the first step. Finally, the significance of the estimated coefficient for λ(·) is informative about the degree to which nonrandom sample selection introduces bias into the substantive equation. To be precise, the significance of the parameter conforming to λ(·) is a test of the significance of the correlation between the unobserved determinants of Y_i^* and D_i^*. Specifically, it can be shown that:

$$\beta_{j+1} = \rho_{e,u}\sigma_e = \sigma_{e,u}$$

As a final technical note on the two-step estimator, it is important to point out that least squares regression of Eq. 5 will yield consistent estimates of the coefficients in **β**. The covariance estimator, on the other hand, is inconsistent and will produce misleading standard errors. This is partly because the residuals in the two-step model are heteroscedastic and partly because the selection term represents an estimate rather than the actual selection term. Therefore, robust estimation of the covariance matrix is typically recommended, or else bootstrapping.

Limitations of the Two-Step Procedure

The appeal of Heckman's solution to the sample selection problem derives from the fact that it is an intuitive and tractable method, requiring only estimation of a probit regression model followed by a linear regression model (although see Bushway et al. 2007, for examples of misapplication of the method in criminology). The model should not be treated as a panacea to the sample selection problem, however, for it has well-known limitations. Several of these are described below (see also Stolzenberg and Relles 1990, 1997).

One limitation of Heckman's model in many research settings is the often mechanical application of the procedure. This has been especially pronounced in sociology and criminology (see Bushway et al. 2007). This is largely a consequence of the absence of formal selection models in these disciplines, leading many researchers to conceive of the sample selection problem as a purely technical rather than theoretical issue. In economics, on the other hand, there are often well-defined theories that inform researchers about the proper specification of the selection and substantive models. Theory can provide helpful guidance about regressors which can be included in the selection model but excluded from the substantive model. Such regressors, called "instrumental variables" (IV's), can improve the empirical performance of the two-step estimator.

A second limitation of Heckman's model is that the selection term tends to be highly collinear with the regressors in the substantive equation. This is because the same regressors appear in both the selection and substantive equations. High collinearity can make the standard errors of the two-step estimator so inflated, in fact, that it loses its appeal altogether. The trade-off between consistency and efficiency is an important one. Under some circumstances, the inconsistency of the uncorrected least squares estimator is preferred to the inefficiency of the selection-corrected least squares estimator. Although the two-step model is formally identified even when the regressors in the two equations are identical, the estimator has better asymptotic properties when a researcher imposes one or more exclusion restrictions which can be justified a priori. Exclusion restrictions cannot only strengthen the theoretical grounding of the model (as mentioned in the previous paragraph) but can improve the statistical performance of the two-step estimator. As a matter of empirical practice, then, the use of instrumental variables in the selection model, along with collinearity diagnostics for the substantive model, is recommended.

A third limitation of Heckman's two-step estimator is that it is comparatively less efficient than other estimators. Heckman's model is what is known as a limited information maximum likelihood (LIML) estimator. The two-step approach had practical utility at the time that Heckman (1976, 1979) proposed the model, simply because the alternatives were computationally burdensome. With today's computing capability, however, asymptotic efficiency can be achieved by employing a full information maximum likelihood (FIML) estimator, which simultaneously models the selection and substantive equations. The FIML estimator maximizes the following log likelihood function:

$$\sum_{D_i=0} ln\ \Phi(-\mathbf{Z}\boldsymbol{\alpha}) + \sum_{D_i=1} ln\left[\frac{1}{\sigma_e}\phi\left(\frac{Y_i - \mathbf{X}\boldsymbol{\beta}}{\sigma_e}\right)\right] + \sum_{D_i=1} ln\ \Phi\left[\frac{\mathbf{Z}\boldsymbol{\alpha} + \rho_{e,u}(Y_i - \mathbf{X}\boldsymbol{\beta})/\sigma_e}{\sqrt{1-\rho_{e,u}^2}}\right]$$

An Empirical Illustration of the Two-Step Procedure

Data from the National Longitudinal Survey of Youth 1997 (NLSY97) are used to provide an empirical illustration of the two-step solution to the sample selection problem. The NLSY97 is representative of all youth in the United States born in 1980–1984. This dataset is used to explore the relationship between arrest and earnings from employment. The analysis is limited to the 4,599 males who participated in the survey, among whom 85 % ($N = 3{,}895$) contributed nonmissing data on the variables of interest. At the second interview – the interview from which the outcome variable is measured – the respondents range in age from 13 to 19, with a mean age of 15.9 years. Descriptive statistics for all measures included in the analysis are provided in Table 1.

The outcome in the substantive model is total weekly earnings from employment. This measure is taken from the second wave of the survey (1998) and represents earnings from "formal employment" or "paycheck jobs." Sample selection bias is a potential concern here, because only half of the sample (50.9 %) is employed between the first and second interviews. Among the subsample that is employed, mean earnings are 147 dollars per month. The earnings distribution among these subsamples is shown in Fig. 1.

The independent variables that are used to model employment earnings include cohort, race (white), schooling (highest grade completed), family structure (live with both biological parents, household size), residential location (central city residence, local unemployment rate), previous work experience (percentage of weeks worked, number of jobs), variety of delinquent behaviors, and total arrests. To ensure proper temporal ordering vis-à-vis earnings, these measures are taken from the first wave of the survey (1997). For this empirical exercise, the regressor of interest is arrest. The question that guides the analysis is, "What impact does arrest have on earnings?"

Three instrumental variables are included in the selection model and excluded from the substantive model. These represent variables that are thought to influence whether a respondent is employed, but are not believed to otherwise have any direct influence on earnings among those who are employed. The instrumental variables that are chosen represent provisions of the child labor law which prevails in the state in which a respondent resides as of the second interview (details on the laws are provided in Apel et al. 2008). Three such provisions impose limits on how intensively young people are allowed to work during the school year – the number of hours per week, the number of hours per weekday, and the work curfew on weeknights. For convenience, a binary indicator is constructed for each provision, denoting whether or not a respondent is allowed to work 40 h or more per week (53.6 %), 8 h or more per weekday (57.8 %), and later than 10 PM on weeknights (47.4 %). Interactions of these variables are also included in the selection model.

Sample Selection Problems, Table 1 Descriptive statistics and the selection model of employment

			Probit model of employment	
Variable	Min, max	Mean (std. dev.)	Coeff. (std. err.)	Marginal effect
Response variables (wave 2)				
Employed in a formal job	0, 1	50.9 %	–	–
Weekly earnings if employed	0.5, 800.0	147.0 (103.7)	–	–
Regressors (wave 1)				
Cohort	12, 16	14.0 (1.4)	0.275 (0.039)***	0.109
White	0, 1	52.1 %	0.350 (0.053)***	0.138
Highest grade completed	0, 12	7.6 (1.6)	0.168 (0.030)***	0.066
Live with both biological parents	0, 1	50.5 %	−0.090 (0.052)$^{+}$	−0.036
Household size	2, 16	4.5 (1.5)	−0.030 (0.016)$^{+}$	−0.012
Live in central city	0, 1	31.1 %	−0.036 (0.054)	−0.014
Local unemployment rate	1.5, 17.5	5.5 (3.0)	−0.025 (0.009)**	−0.010
Percentage of weeks employed	0.0, 100.0	12.4 (27.8)	0.012 (0.002)***	0.005
No. of different jobs	0, 7	0.4 (0.8)	0.293 (0.063)***	0.116
No. of different delinquent behaviors	0, 7	1.4 (1.7)	0.075 (0.016)***	0.030
No. of arrests	0, 33	0.3 (1.4)	−0.039 (0.018)*	−0.016
Instrumental variables (wave 2)				
(1) 40+ hours per week allowed	0, 1	53.6 %	0.223 (0.138)	0.088
(2) 8+ hours per weekday allowed	0, 1	57.8 %	0.405 (0.135)**	0.160
(3) work past 10 PM allowed	0, 1	47.4 %	1.144 (0.411)**	0.428
(1) × (2)	0, 1	49.7 %	−0.697 (0.206)***	−0.271
(1) × (3)	0, 1	41.7 %	−1.117 (0.485)*	−0.423
(2) × (3)	0, 1	45.7 %	−1.089 (0.444)*	−0.413
(1) × (2) × (3)	0, 1	40.6 %	1.378 (0.521)**	0.495

Note: $N = 3{,}895$. Estimates are unweighted. Means of binary variables are shown as percentages. Pseudo (McFadden's) R-square for the probit regression model is 0.336. The marginal effects are all evaluated at the sample means.
$^{+}p < 0.10$; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$ (two-tailed tests)

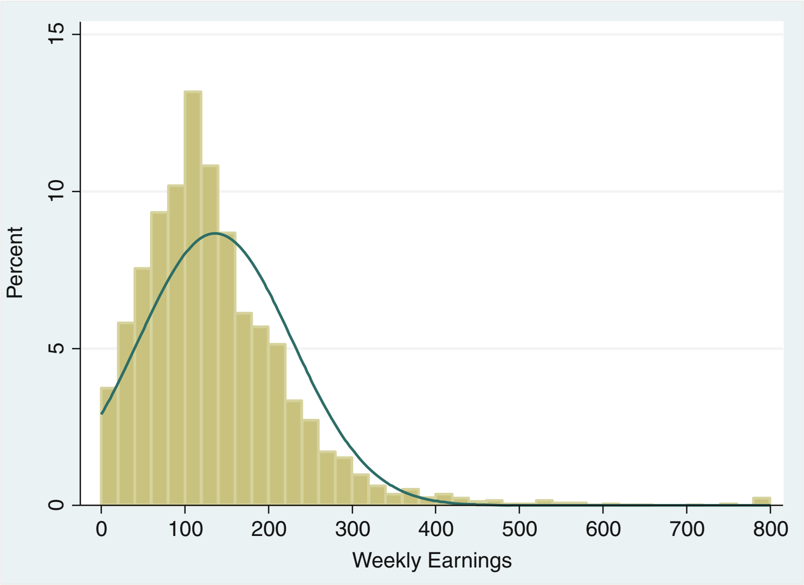

Sample Selection Problems, Fig. 1 Distribution of employment earnings

Sample Selection Problems, Table 2 Comparative estimates of the substantive model of weekly earnings

	Uncorrected models		Selection-corrected models	
Regressor	(1) Least squares regression coeff. (std. err.)	(2) Truncated regression coeff. (std. err.)	(3) Heckman's two-step coeff. (std. err.)	(4) FIML coeff. (std. err.)
Cohort	12.4 (3.3)***	27.0 (6.7)***	17.8 (4.2)***	37.7 (3.5)***
White	−10.6 (5.0)*	−20.4 (9.9)*	−6.1 (5.5)	13.0 (5.5)*
Highest grade completed	6.3 (2.8)*	12.6 (5.4)*	8.4 (3.0)**	17.0 (3.0)***
Live with both biological parents	−7.8 (4.8)	−15.0 (9.4)	−8.6 (4.8)$^{+}$	−12.3 (5.2)*
Household size	3.2 (1.6)*	5.9 (3.1)$^{+}$	2.9 (1.6)$^{+}$	0.9 (1.7)
Live in central city	−9.1 (5.2)$^{+}$	−18.0 (10.4)$^{+}$	−9.7 (5.3)$^{+}$	−9.4 (5.6)$^{+}$
Local unemployment rate	1.2 (0.8)	2.1 (1.6)	0.7 (0.9)	−1.4 (0.9)
Percentage of weeks employed	−0.6 (0.1)***	−1.2 (0.2)***	−0.5 (0.1)***	−0.1 (0.1)
No. of different jobs	17.5 (3.1)***	30.3 (5.7)***	19.1 (3.3)***	24.0 (3.8)***
No. of different delinquent behaviors	2.7 (1.3)*	5.1 (2.6)*	3.6 (1.4)*	6.8 (1.5)***
No. of arrests	−1.3 (1.4)	−2.5 (2.8)	−1.7 (1.5)	−2.8 (1.7)$^{+}$
Lambda $\lambda(-Z\alpha)$	–	–	28.4 (14.1)*	121.8 (2.7)***
Sigma σ_e	99.3	139.1	100.6	125.2
Rho $\rho_{e,u}$	–	–	0.282	0.973

Note: $N = 1{,}981$. Estimates are unweighted. The dependent variable, weekly earnings, is measured from the second interview, while all regressors are measured from the first interview
$^{+}p < 0.10$; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$ (two-tailed tests)

The first step in the analysis is to model the probability that a respondent is uncensored, that is, has nonzero earnings. This is a probit regression model, the coefficients, standard errors, and marginal effects from which are provided in the last set of columns in Table 1. By way of a brief summary, youth who are more likely to work include those who are older, are white, have more schooling, have less local unemployment, and have more prior work experience. Youth who are more heavily involved in delinquent behavior are also more likely to work, although youth who have accumulated an arrest history are less likely to work. It is important to point out that the instrumental variables are jointly significant ($\chi^2 = 36.4$; df $= 7$; $p < 0.00001$), and six out of the seven main and interaction effects are statistically significant.

In Table 2, results from a variety of specifications of the earnings function are provided. All models are estimated from the subsample of male youth with nonzero earnings, but they differ as to whether a selection correction is made. The first two columns provide coefficients from models in which no correction is made for sample selection – coefficients from the least squares regression model and the truncated regression model are provided. The coefficients from the two are of the same direction and almost identical significance. The difference is that the coefficients from least squares regression are about half the size of the coefficients from truncated regression. This is because the latter model explicitly accounts for the truncation in earnings at 0 (notice that the estimate of the model standard deviation, represented by σ_e, is 40 % larger in the truncated regression model) and is the typical starting point for analyses of this sort. As an aside, the truncated regression model represents the second part of the two-part model advocated by Cragg (1971; for criminological applications, see Smith and Brame 2003).

As for the question more immediately at hand, the number of arrests is unrelated to weekly earnings in the uncorrected models. Although the coefficient is negative, it is not statistically significant. On the other hand, more extensive delinquent involvement is positively and significantly correlated with earnings. This is consistent with other research findings that high-risk youth tend to be more extensively involved in the labor market (Apel et al. 2008). A test of an interaction between delinquency and arrest is not significant.

The last two columns provide results from models which are corrected for sample selection – coefficients from Heckman's two-step estimator and the full information maximum likelihood (FIML) estimator are provided. While the coefficient for arrest is not significant in the two-step model, it is marginally significant in the FIML model. The latter model suggests that each arrest reduces earnings by $2.80 per week. Considering that the mean wage at the second interview is $5.72 per hour, each arrest as of the first interview lowers earnings by the equivalent of about one-half hour's worth of work per week between the first and second interviews.

Notice that the selection correction entails inclusion of a new regressor in the earnings model – $\lambda(-\mathbf{Z\alpha})$, or lambda, the inverse Mills ratio. The coefficient that conforms to this regressor is $\sigma_{e,u}$, or the covariance of the residuals in the models at the first and second steps. The models also provide an estimate of the correlation between the residuals at the first and second steps – $\rho_{e,u}$, or rho. The relationship between the estimates of these quantities can be expressed as follows:

$$\hat{\rho}_{e,u} = \frac{\hat{\sigma}_{e,u}}{\hat{\sigma}_e} = \frac{\hat{\beta}_{\lambda(-Z\alpha)}}{\hat{\sigma}_e}$$

Importantly, the coefficient on lambda is meaningful. Since it is positive and statistically significant, it indicates that there is positive selection, meaning that the unobservables which make a young male more prone to work (e.g., motivation, work orientation) also tend to yield him higher mean earnings. So, selection into the nonzero earnings subsample produces higher expected earnings compared to someone drawn at random from the population, with the same set of characteristics, who chooses not to work. Therefore, models which do not perform a selection correction will produce estimates which are biased and inconsistent. Upon first glance, then, the selection-corrected models are preferred to the truncated regression model.

By the same token, there are circumstances in which the cure for the sample selection problem is oftentimes worse than the disease. Recall that high collinearity between the regressors and the selection term, lambda, can produce inefficiency, which can undermine the consistency advantage of the two-step and FIML estimators. Unfortunately, high collinearity cannot be ruled out in the present analysis, even with exclusion restrictions for the child labor law provisions. A regression of the inverse Mills ratio on all of the covariates yields a squared multiple correlation (R-square) of 0.875. And this is only marginally better than an inverse Mills ratio which omits the instrumental variables, producing an R-square of 0.907. This suggests that the selection-corrected models might be plagued by high collinearity between the regressors and the inverse Mills ratio and that the instrumental variables which are chosen might not be sufficiently strongly related to employment to justify their use as instruments. On second glance, therefore, the truncated regression model might be preferred to the selection-corrected models after all.

Conclusion

The sample selection problem which led to the formulation of Heckman's two-step estimator (Heckman 1976, 1979) is a very general problem. Casual consideration leaves one with the impression that sample selection is pervasive in social scientific analysis. Namely, any restriction that leads to the elimination of a potentially nonrandom subset of a sample that is intended to be representative of a larger population can give rise to the sample selection problem. Viewed in this light, Heckman's two-step estimator, and the growing class of models to which it belongs, is more than a peculiar method which is strictly of interest to economists. It is a tool that should be readily available in any researcher's toolbox.

To be sure, Heckman's two-step solution to the sample selection problem is not free of its detractions. It has well-known limitations, some of which have been described in this chapter. In fact, Heckman (1976, 1979) himself was cautious about his proposed model, recommending that it be used in the course of preliminary investigation rather than as the sole model. Any final determination of which set of models provides the most credible parameter estimates requires careful consideration of underlying assumptions and the consequences of violating those assumptions.

Related Entries

▶ Econometrics of Crime
▶ Parametric Sample Selection Models
▶ Sample Selection Models

Recommended Reading and References

Apel R, Bushway SD, Paternoster R, Brame R, Sweeten G (2008) Using state child labor laws to identify the causal effect of youth employment on deviant behavior and academic achievement. J Quant Criminol 24:337–362

Berk RA (1983) An introduction to sample selection bias in sociological data. Am Sociol Rev 48:386–398

Bushway S, Johnson BD, Slocum LA (2007) Is the magic still there? The use of the Heckman two-step correction for selection bias in criminology. J Quant Criminol 23:151–178

Cragg JG (1971) Some statistical models for limited dependent variables with application to the demand for durable goods. Econometrica 39:829–844

Heckman JJ (1976) The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Ann Econ Soc Meas 5:475–492

Heckman JJ (1979) Sample selection bias as a specification error. Econometrica 47:153–161

Smith DA, Brame R (2003) Tobit models in social science research: some limitations and a more general alternative. Sociol Method Res 31:364–388

Stolzenberg RM, Relles DA (1990) Theory testing in a world of constrained research design: the significance of Heckman's censored sampling bias correction for nonexperimental research. Sociol Method Res 18:395–415

Stolzenberg RM, Relles DA (1997) Tools for intuition about sample selection bias and its correction. Am Sociol Rev 62:494–507

Winship C, Mare RD (1992) Models for sample selection bias. Annu Rev Sociol 18:327–350

Sanctioning

▶ Sentencing Research

Scared Straight Programs

Anthony Petrosino[1], Carolyn Petrosino[2], Meghan Hollis-Peel[3] and Julia Lavenberg[4]
[1]Learning Innovations, WestEd,
Woburn, MA, USA
[2]Bridgewater State University, Bridgewater, MA, USA
[3]Northeastern University and University of Massachusetts, Boston, MA, USA
[4]University of Pennsylvania Medical Center, Philadelphia, Pennsylvania, USA

Synonyms

Juvenile awareness programs; Prison aversion programs

Overview

"Scared Straight" and other programs involve organized visits to prison by juvenile delinquents or children at risk for criminal behavior. Programs are designed to deter participants from future offending through firsthand observation of prison life and interaction with adult inmates.

Drawing on the findings from a recently updated systematic review, this entry describes these programs and examines the research evidence on the effects of these programs on delinquency. This entry shows that not only do these programs fail to deter crime but they actually increase it. Government officials permitting this program need to adopt rigorous evaluation to ensure that they are not causing more harm to the very citizens they pledge to protect.

Background

In the 1970s, inmates serving life sentences at a New Jersey (USA) prison began a program to "scare" or deter at-risk or delinquent children from a future life of crime. The program, known as "Scared Straight," featured as its main component an aggressive presentation by inmates to juveniles visiting the prison facility. The presentation depicted life in adult prisons, and often included exaggerated stories of rape and murder (Finckenauer 1982). A television documentary on the program aired in 1979 provided evidence that 16 of the 17 delinquents remained law abiding for 3 months after attending "Scared Straight" – a 94 % success rate (Finckenauer 1982). Other data provided in the film indicated success rates that varied between 80 % and 90 % (Finckenauer 1982). The program received considerable and favorable media attention and was soon replicated in over 30 jurisdictions nationwide, resulting in special Congressional hearings on the program and the film by the United States House Committee on Education and Labor, Subcommittee on Human Resources (Petrosino et al. 2000).

Programs such as "Scared Straight" are based on deterrence theory. Advocates and developers of these programs believe that it is possible to use these realistic depictions of life in prison and/or presentations by inmates to deter juvenile offenders (or children at risk for becoming delinquent) from further (or initial) involvement with crime. Although the harsh and sometimes vulgar presentation in the earlier New Jersey version is the most famous, inmate presentations are now sometimes designed to be more educational than confrontational but with a similar crime prevention goal (Finckenauer and Gavin 1999). Programs that feature interactive discussions between youth and inmates as speakers who describe their life experiences and the current reality of prison life have a rather long history in the United States (Michigan D.O.C. 1967). It is not surprising such programs are popular: They fit with common notions of how to prevent or reduce crime (by "getting tough"); they are very inexpensive (a Maryland program was estimated to cost less than $1 US per participant); and they provide one way for incarcerated offenders to contribute productively to society by preventing youngsters from following down the same path (Finckenauer 1982).

A randomized controlled trial of the New Jersey program published in 1982, however, reported no effect on the criminal behavior of participants in comparison with a no-treatment control group (Finckenauer 1982). In fact, Finckenauer reported that participants in the experimental program were more likely to be arrested following the program. Other randomized trials reported in the United States also questioned the effectiveness of "Scared Straight"–type programs in reducing subsequent criminality (Greater Egypt Regional Planning and Development Commission 1979; Lewis 1983). Consistent with these findings, reviewers of research on the effects of crime prevention programs have not found deterrence-oriented programs like "Scared Straight" to be effective (Lipsey 1992).

Despite the convergence of evidence from studies and reviews, "Scared Straight"–type programs remain popular and continue to be used in the United States (Finckenauer and Gavin 1999). For example, a program in Carson City, Nevada (USA) brings juvenile delinquents on a tour of an adult Nevada State Prison (Scripps 1999). One youngster claimed that the part of the tour that made the most impact on him was "All the inmates calling us for sex and fighting for our belongings" (Scripps 1999). The United Community Action Network has its own program called "Wisetalk" in which at-risk youth are locked in a jail cell for over an hour with four to five parolees. They claim that only 10 of 300 youngsters exposed to this intervention have been rearrested (United Community Action Network 2001). In 2001, a group of guards – apparently without the knowledge of administrators – strip-searched Washington, DC students during their tours of a local jail under the guise of that they were using "a sound strategy to turn around the lives of wayward kids" – claiming the prior success of "Scared Straight" (Blum and Woodlee 2001).

"Scared Straight" and other "kids visit prison" programs are not unique to the United States. For example, it is called the "day in prison" or "day in gaol" in Australia (O'Malley et al. 1993), "day visits" in the United Kingdom (Lloyd 1995), and the "Ullersmo Project" in Norway (Storvoll and Hovland 1998). Hall (1999) reports positively on a program in Germany designed to scared straight young offenders with ties to Neo-Nazi and other organized hate groups. The program has been also tried in Canada (O'Malley et al. 1993). A recent television program (*Banged* Up) aired in the United Kingdom depicting their version of the program (Blunkett 2008; Wilson and Groombridge 2010).

In 1999, "Scared Straight: 20 Years Later" was shown on United States television and claimed similar results to the 1979 film (Muhammed 1999). In this version, the film reports that 10 of the 12 juveniles attending the program have remained offense free in the 3 months follow-up (Muhammed 1999). As in the 1979 television program, no data on a control or comparison group of young people were presented. Positive reports and descriptions of Scared Straight–type programs have also been reported elsewhere (e.g., in Germany [Hall 1999], and in Florida [Rasmussen 1996]), although it is sometimes embedded as one component in a multicomponent juvenile intervention package (Trusty 1995; Rasmussen 1996).

In 2000, Petrosino and his colleagues reported on a preliminary systematic review of nine randomized field trials, drawing on the raw percentage differences in each study. They found that programs such as "Scared Straight" generally increased crime between 1 % and 28 % in the experimental group when compared to a no-treatment control group. In 2002, a formal Campbell review was published (simultaneously with the Cochrane Collaboration) – updating the 2000 work and utilizing more sophisticated meta-analytic techniques. They reported similarly negative findings for Scared Straight and juvenile awareness programs. This entry provides an update of these earlier reports, extending searches to cover literature published through 2011.

Despite the results of this review and updates, Scared Straight and similar programs continue to be promoted as a crime prevention strategy. For example, Illinois' then-Governor Rod Blagojevich signed a bill into law in 2003 that mandated the Chicago Public School system to set up a program called "Choices" (United Press International 2003). The program would identify students at risk for committing future crime and set up a program to give them "tours of state prison" to discourage any future criminal conduct (United Press International 2003). More recently, the Arts and Entertainment (A&E) station has been running a weekly series entitled, "Beyond Scared Straight." Created by the producer of the original Scared Straight program (Arnold Shapiro), the program is now the highest rated in A&E's history. The success of the television show has renewed interest in Scared Straight and similar programs as a crime prevention strategy (e.g., Dehnart 2011) but has also resulted in criticism that it ignores a long history of scientific evidence (e.g., Robinson and Slowikowski 2011). The Puerto Rico Department of Corrections and Rehabilitation also launched the "Loving Freedom" program, an intervention designed to imitate Scared Straight (Mullen 2012).

The question about whether Scared Straight and similar programs has a crime deterrent effect is best answered by examining the existing scientific evidence. Of course, prior research is no guarantee that interventions will work (or not work) in a future setting, but a reader might ask herself the following question upon reading the results of the aforementioned systematic review: Would I want a doctor to prescribe a treatment for my child that has the same track record of research results?

Methods

The findings reported here are based on the most recent update of the Campbell and Cochrane Collaboration reviews on the subject. In these reviews, studies were only included if they used random (or seemingly – quasi-random) procedures to assign participants to treatment and control groups. The control must not have received another prevention program. Furthermore, only studies that involved juveniles i.e., children 18 years of age or under) or overlapping samples of children and young adults (i.e., ages 13–21) were included. Studies could have included delinquents, pre-delinquents, or non-delinquent youths. Studies also were required to include a visit by program participants to a prison facility as a part of the intervention. Most of the included studies involved inmate presentations (Finckenauer 1982; Cook 1992), and they sometimes featured orientation sessions and prison tours. Finally, studies had to include at least one outcome of subsequent offending behavior. This outcome was measured by indices such as arrests, convictions, contacts with police, and self-reported offenses. Further information on the methodology and search methods is available in the original review.

Narrative Findings

Collectively, the nine included studies from the systematic review were conducted in eight different states of the United States, with Michigan the site for two studies (Yarborough 1979; Michigan D.O.C. 1967). No set of researchers conducted more than one experiment. The studies span the years 1967–1992. The first five studies located were unpublished and were disseminated in government documents or dissertations; the remaining four were found in academic journal or book publications. The average age of the juvenile participants in each study ranged from 15 to 17. Only the New Jersey study included girls (Finckenauer 1982). Racial composition across the nine experiments was diverse, ranging from 36 % to 84 % white. Nearly 1,000 (946) juveniles or young adults participated in the nine experimental studies. A narrative description of the included studies follows.

In an internal, unpublished government document, the Michigan Department of Corrections (1967) reported a trial testing a program that involved taking adjudicated juvenile boys on a tour of a state reformatory. Unfortunately, the report is remarkably brief. Sixty juvenile delinquent boys were randomly assigned to attend two tours of a state reformatory or to a no-treatment control group. Tours included 15 juveniles at a time. No other part of the program is described. Recidivism was measured as a petition in juvenile court for either a new offense or a violation of existing probation order. The Michigan department of Corrections found that 43 % of the experimental group re-offended, compared to only 17 % of the control group. This large negative result curiously receives little attention in the original document.

Another program at the Menard Correctional Facility (GERP&DC 1979) started in 1978 and is described as a frank and realistic portrayal of adult prison life. The researchers randomly assigned 161 youths aged 13–18 to attend the program or a no-treatment control. The participants were a mix of delinquents or children at risk of becoming delinquent. Participants were compared on their subsequent contact with police, on two personality inventories (Piers-Berne and Jesness) and surveys of parents, teachers, inmates, and young people. The outcomes are also negative in direction but not statistically significant, with 17 % of the experimental participants being recontacted by police in contrast to 12 % of the controls (GERP&DC 1979). The authors concluded that “based on all available findings one would be ill advised to recommend continuation or expansion of the juvenile prison tours. All empirical findings indicate little positive outcome, indeed, they may actually indicate negative effects” (p. 19). Researchers report no effect for the program on two attitude tests (Jesness Inventory, Piers Harris Self-Concept Scale). In contrast, interview and mail surveys of participants and their parents and teachers indicated unanimous support for the program (p. 12). Researchers also note how positive and enthusiastic inmates were about their efforts.

In the Juvenile Offenders Learn Truth (JOLT) program (Yarborough 1979), juvenile delinquents in contact with one of four Michigan county courts participated. Each juvenile spent 5 total hours in the facility. Half of this time was spent in a confrontational “rap” session. This followed a tour of the facility, during which participants were escorted to a cell and exposed to interaction with inmates (e.g., taunting). In the evaluation, 227 youngsters were randomly assigned to JOLT or to a no-treatment control. Participants were compared on a variety of crime outcomes collected from participating courts at 3- and 6-month follow-ups. This second Michigan study reported very little difference between the intervention and control group (Yarborough 1979). The average offense rate for program participants, however, was 69 % compared to 47 % for the control group. Yarborough (p. 14) concluded that, "…the inescapable conclusion was that youngsters who participated in the program, undergoing the JOLT experience, did no better than their control counterparts."

The Face-to-Face program (Vreeland 1981) included a 13-h orientation session in which the juvenile lived as an inmate followed by counseling. Participants were 15–17 years of age and on probation from Dallas County Juvenile Court; most averaged two to three offenses before the study. A total of 160 boys were randomly assigned to four conditions: prison orientation and counseling, orientation only, counseling only, or a no-treatment control group. Vreeland examined official court records and self-reported delinquency at 6 months. This evaluation also reported little effect for the intervention (Vreeland 1981). Vreeland reported that the control participants outperformed the three treatment groups on official delinquency (28 % delinquent versus 39 % for the prison orientation plus counseling, 36 % for the prison only, and 39 % for the counseling only). This more robust measure contradicts data from the self-report measures used, which suggest that all three treatment groups did better than the no-treatment controls. None of these findings reached a level of statistical significance. Viewing all the data, Vreeland concluded that there was no evidence that Face-to-Face was an effective delinquency prevention program. He finds no effect for "Face-to-Face" on several attitudinal measures, including the "Attitudes Toward Obeying Law Scale."

The New Jersey Lifers' Program (Finckenauer 1982) began in 1975 and stressed confrontation with groups of juveniles ages 11–18 who participated in a rap session. Finckenauer randomly assigned 82 juveniles, some of whom were not delinquents, to the program or to a no-treatment control group. He then followed them for 6 months in the community, using official court records to assess their behavior. Finckenauer reported that 41 % of the children and young people who attended the "Scared Straight" program in New Jersey committed new offenses, while only 11 % of the controls did, a difference that was statistically significant (Finckenauer 1982). He also reported that the program participants committed more serious offenses and that the program had no impact on nine attitude measures with the exception of a measure called "attitudes toward crime." On this measure, experimental participants did much worse than controls. We deal with Finckenauer's own concerns about randomization integrity in a sensitivity analysis, which is reported later.

The California SQUIRES Program (Lewis 1983) is supposedly the oldest such program in the United States, beginning in 1964. The SQUIRES program included male juvenile delinquents from two California counties between the ages of 14 and 18, most with multiple prior arrests. The intervention included confrontational rap sessions with rough language, guided tours of prison with personal interaction with prisoners, and a review of pictures depicting prison violence. The intervention took place 1 day per week over 3 weeks. The rap session was 3 h long, and normally included 20 youngsters at a time. In the study, 108 participants were randomly assigned to treatment or to a no-treatment control group. Lewis compared participants on seven crime outcomes at 12 months. Lewis reported that 81 % of the program participants were arrested compared to 67 % of the controls. He also found that the program did worse with seriously delinquent youths, leading him to conclude that such children and young people could not be "turned around by short-term programs such as SQUIRES…a pattern for higher risk youth suggested that the SQUIRES program may have been detrimental" (p. 222). The only deterrent effect for the program was the average length of time it took to be rearrested: 4.1 months for experimental participants and 3.3 months for controls. Data were reported on eight attitudinal measures, and Lewis reported that the program favored the experimental group on all of them, again underscoring the difficulty of achieving behavioral change even when positively affecting the attitudes of juvenile delinquents.

The Kansas Juvenile Education Program (Locke et al. 1986) was designed to educate children about the law and the consequences of violating it. The program also tried to roughly match juveniles with inmates based on personality types. Fifty-two juvenile delinquents (ages 14–19) from three Kansas counties were randomly assigned while on probation to KEP or a no-treatment control. The investigators examined official (from police and court sources) and self-report crime outcomes at 6 months. Locke and his colleagues reported little effect of the Juvenile Education Program. Both groups improved from pretest to posttest, but the investigators concluded that there were no differences between experimental and control groups on any of the crime outcomes measured. Investigators also reported no effect for the program on the Jesness and Cerkovich attitude tests.

Project Aware (Cook and Spirrison 1992) was a nonconfrontational, educational program comprising one 5 h session run by prisoners. The intervention was delivered to juveniles in groups numbering from 6 to 30. In the study, 176 juveniles (ages 12–16) under the jurisdiction of the county youth court were randomly assigned to the program or to a no-treatment control. The experimental and control groups were compared on a variety of crime outcomes retrieved from court records at 12 and 24 months. Little difference was found between experimental and control participants in the study. For example, the mean offending rate for controls at 12 months was 1.25 for control cases versus 1.32 for Project Aware participants. Both groups improved from 12 to 24 months, but the control mean offending rate was still lower than the experimental group. The investigators concluded that, “attending the treatment program had no significant effect on the frequency or severity of subsequent offenses” (p. 97). The investigators also reported on two educational measures: school attendance and dropout. Curiously, they report an effect for the program on school dropout data, but not that “. . .it is not clear how the program succeeded in reducing dropout rates. . .” (p. 97).

The only positive findings, though not statistically significant, were reported in Virginia (Orchowsky and Taylor 1981). The Insiders Program (Orchowsky and Taylor 1981) was described as an inmate-run, confrontational intervention with verbal intimidation and graphic descriptions of adult prison life. Juveniles were locked in a cell 15 at a time and told about the daily routine by a guard. They then participated in a 2-h confrontational rap session with inmates. Juvenile delinquents from three court service units in Virginia participated in the study. The investigators randomly assigned 80 juveniles ages 13–20 with two or more prior adjudications for delinquency to the Insiders program or a no-treatment control group. Orchowsky and Taylor report on a variety of crime outcome measures at 6-, 9-, and 12-month intervals. Although the difference at 6 months was not statistically significant (39 % of controls had new court intakes versus 41 % of experimental participants), they favor the experimental participants at 9 and 12 months. The investigators noted, however, that the attrition rates in their experiment were dramatic. At 9 months, 42 % of the original sample dropped out, and at 12 months, 55 % dropped out. The investigators conducted analyses that seemed to indicate that the constituted groups were still comparable on selected factors.

Most of the studies dealt with delinquent youths already in contact with the juvenile justice system. All of the experiments were simple two-group experiments except Vreeland’s evaluation of the Texas Face-to-Face program (Vreeland 1981). Only one study used quasi-random alternation techniques to assign participants (Cook 1992); the remaining studies claimed to use randomization although not all were explicit about how such assignment was conducted. Only the Texas study (Vreeland 1981) included data from

Effects of Scared Straight and other similar programs: Meta-analysis of first effect crime outcomes (Random Effects Analysis)

Study name	Statistics for each study				
	Odds ratio	Lower limit	Upper limit	Z-Value	p-Value
Finckenauer (1982)	5.454	1.650	18.022	2.781	0.005
GERPDC (1979)	1.513	0.607	3.772	0.888	0.374
Lewis (1983)	2.092	0.860	5.090	1.627	0.104
Michigan DOC (1967)	3.750	1.110	12.669	2.128	0.033
Orchowsky (1981)	1.087	0.444	2.660	0.183	0.855
Vreeland (1981)	1.476	0.569	3.832	0.801	0.423
Yarborough (1979)	1.054	0.537	2.070	0.153	0.879
	1.724	1.134	2.619	2.550	0.011

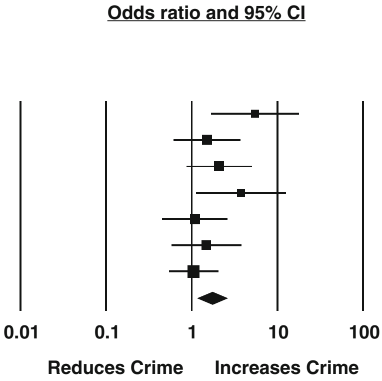

Scared Straight Programs, Fig. 1 Random effects model results (Petrosino et al. 2004)

self-report measures. In two studies (Cook 1992; Locke et al. 1986), no post-intervention offending rates were reported. Also, the follow-up periods were diverse and included measurements at 3, 6, 9, 12, and 24 months.

The results of the meta-analysis presented in the systematic review are further evidence of the harmfulness of these programs. The review reported the crime outcomes for official measures at the first-effect or first follow-up interval (and usually the only) period reported. Each analysis focused on proportion data (i.e., the proportion of each group re-offending), as the outcomes reporting means or averages is sparse and often does not include the standard deviations. Thus, because the data rely on dichotomous outcomes, analyses used the odds ratios (OR) as the measure of program effect, contrasting the odds of crime in the treatment program relative to the control.

The analysis of the data in comparison Fig. 1 from the seven studies reporting reoffending rates shows that intervention increases the crime or delinquency outcomes at the first follow-up period. The mean odds ratio across studies assuming a random effects model indicates an overall harmful effect of these programs (mean odds ratio = 1.72, 95 % confidence interval of 1.13–2.62). Thus, the intervention increases the odds of offending (Fig. 1).

These randomized trials, conducted over a 25-year period in eight different jurisdictions, provide evidence that "Scared Straight" and other "juvenile awareness" programs are not effective as a stand-alone crime prevention strategy. More importantly, they provide empirical evidence that these programs likely increase the odds that children exposed to them will commit offenses in future. Despite the variability in the type of intervention used, ranging from harsh, confrontational interactions to tours of the facility converge on the same result: an increase in criminality in the experimental group when compared to a no-treatment control. Doing nothing would have been better than exposing juveniles to the program.

Converging Evidence

Other reviews also examined the efficacy of "Scared Straight" and similar programs. A meta-analysis of juvenile prevention and treatment programs by Lipsey (1992) found a small negative effect for 11 "shock incarceration" and "Scared Straight" programs, with the experimental groups having a 7 % higher recidivism rate than controls relative to a 50 % recidivism baseline. Gendreau and his colleagues (1997) also reported a meta-analysis of "get tough" or "get smart" sanctions. These included interventions designed to deter future crime like "Scared Straight" as well as interventions designed to punish or control offenders at less cost such as intensive supervision while on probation or parole. The reviewers computed correlations of program participation and recidivism outcomes. Examining 15 experimental or quasi-experimental evaluations of Scared Straight–type programs, they also found a small negative (harmful) effect (average correlation of 0.07). Simply put, participating in the program was *associated with an increase in crime*.

Possible Controversies in the Literature

This evidence places the onus on every jurisdiction to show how their current or proposed program is different than the ones studied here. Given that, they should then put in place rigorous evaluation to ensure that no harm is caused by the intervention. Some literature indicates the program can have a positive effect on the inmate providers and that argument is sometimes used to legitimize use of the program. These arguments are undoubtedly used under the assumption that the program does no harm. In light of these findings, assertions that "Scared Straight" and similar programs ought to be used because it achieves other things raises ethical questions about potentially harming children (and others in the community who may be victimized) in order to accomplish other important, but latent, goals.

Petrosino and colleagues have received communications from different prison facilities that are using a juvenile awareness program. One argument these programs make to sustain using such programs is that the research reported here does not apply to their particular program. One recommendation here is that correctional research units, either at the facility or at a regional or national government level, collaborate with program staff to conduct a rigorous evaluation. If such units do not exist or cannot conduct their own study, they should collaborate with a local university, college, or research firm that could undertake this work to ensure that the program is working as planned and not unintentionally causing more harm than good. The findings, however, across a diversity of programs, jurisdictions, and samples (in conjunction with the converging evidence from other reviews) suggests that the fundamental concept behind Scared Straight and other juvenile awareness programs may be flawed.

Open Questions

One question that continues to arise about these findings is why "Scared Straight" and similar programs seem to lead to more crime rather than less in its participants. What is the critical mechanism? Understanding why something works or fails is of great interest to evaluators, program designers, and criminological theorists. Holley and Brewster (1996), evaluators for the Oklahoma "Speak Outs" program, wondered about the criminogenic effect of these programs when they asked:

> If one argued that a two hour visit cannot perform the miracle of deterring socially unacceptable behavior . . ., it can also be argued that it was extremely simplistic to assert that a two hour visit can perform the miracle of *causing* socially unacceptable behavior. (p. 130)

Although there were many good post hoc theories about why these programs had negative effects, the evaluations were not structured to provide the kind of mediating variables or "causal models" necessary for an empirical response to this question in a systematic review.

Another key question concerns why these programs continue to be used. The research evidence to date indicates that these programs simply do not work. Despite this evidence, these programs continue to be used. This concern is particularly problematic given the recent paucity of high-quality research studies evaluating these programs. Why do policymakers continue to implement programs that are found to be harmful?

Arnold Shapiro (cited in Dehnart 2011) criticized the studies reviewed here because none of them were reported after 1992. "Scared Straight" has evolved and is now a very different program, and two decades have passed since that last study was published. This further reinforces the need for jurisdictions using this strategy to conduct rigorous evaluation, but it is difficult to obtain funding from agencies because Scared Straight is viewed as a failed strategy for youth.

Conclusions

The research on Scared Straight–type programs cannot predict with certainty that every such program will fail or – worse yet – lead to harmful effects on juvenile participants. But, the prior evidence indicates that there is a greater probability than not that it will be harmful. Would you permit a doctor to use a medical treatment on your child with a similar track record of results?

Despite the gloomy findings reported here and elsewhere, "Scared Straight" and its derivatives continue in use, although a randomized trial has not been reported since 1992. As Finckenauer and Gavin (Finckenauer 1999) noted, when the negative results from the California SQUIRES study came out, the response was to end the evaluation – not the program. Today, the SQUIRES program continues, evaluated by the testimonials of prisoners and participants alike. Despite evidence, belief in the program's efficacy continue. Middleton and his colleagues report on the extension of this strategy in one UK town to scare ordinary schoolchildren by using former correctional officers to set up a prison-type atmosphere in the public school system (Middleton et al. 2001). In 1982, Finckenauer called this the "Panacea Phenomenon," describing how policymakers, practitioners, media reporters, and others sometimes latch onto quick, short-term, and inexpensive cures to solve difficult social problems (Finckenauer 1982). Others claim that the program by itself is of little value but could be instrumental if embedded in an overall multicomponent package of interventions delivered to youths. More recently, the success of the A&E program, Beyond Scared Straight, has increased enthusiasm for this program as a crime prevention strategy.

It may be true that Scared Straight and like programs do not work because they only convey a threat that juveniles do not think will be carried out. What about the evidence for deterrence when it is not an inmate providing a third-party threat but the juvenile system officially processes the youth? There have been a wide range of randomized trials that test for the effects of official processing in juvenile courts with some other intervention (such as diverting the kid from such processing). Is there evidence that the delivery of a threat – official system processing – deters future criminal behavior? In 2010, Petrosino, et al. examined 29 randomized trials that evaluated the effects of some diversionary alternative (services or outright release) and compared it to official processing or progression deeper into the juvenile justice system. That review, published by the Campbell Collaboration, also indicated that formal system processing or progression had no crime deterrent effect, and in some instances increased crime in contrast to diversionary alternatives. In addition, formal processing is a more expensive approach than most diversionary programs, and coupled with the crime reduction effect, could result in some savings for jurisdictions (Petrosino et al. 2010). This review indicates that the delivery of a threat (official processing) did not deter future juvenile offending, compared to doing nothing, and actually reported worse outcomes than if the youth was assigned to a diversionary program with services.

Related Entries

- ▶ Deterrence
- ▶ Examining the Effectiveness of Correctional Interventions
- ▶ Juvenile Diversion
- ▶ Juvenile Justice in the Get Tough Era

Recommended Reading and References

Blum J, Woodlee Y (2001) Trying to give kids a good scare. Washington Post, 3 June 2001, p C01

Blunkett D (2008) Scared straight? Public Pol Res 15(2):77–79

Cook DD (1990) Effects of a non-confrontational prisoner-run juvenile delinquency deterrence program. Unpublished Masters thesis, Department of Psychology, Mississippi State University

Cook DD, Spirrison CL (1992) Effects of a prisoner-operated delinquency deterrence program: Mississippi's project aware. J Offender Rehab 17:89–99

Dehnart A (2011) Beyond scared straight's real-life controversy. The Daily Beast, 23 Feb 2011.

http://www.thedailybeast.com/articles/2011/02/23/beyond-scared-straights-real-life-controversy.html

Finckenauer JO (1982) Scared straight and the panacea phenomenon. Prentice-Hall, Englewood Cliffs

Finckenauer JO, Gavin PW (1999) Scared straight: the panacea phenomenon revisited. Waveland Press, Prospect Heights

Gendreau P, Goggin C, Cullen F, Andrews D, Bonta J (1997) The effectiveness of "get tough" strategies on offender recidivism. Unpublished manuscript, University of New Brunswick, Center for Criminal Justice Studies, Saint John

Greater Egypt Regional Planning & Development Commission (1979) Menard correctional center: juvenile tours impact study. Greater Egypt Regional Planning and Development Commission, Carbondale

Hall A (1999) Jailhouse shock aims to scare youths straight. The Scotsman, 26 Oct 1999, p 12

Holley PD, Brewster D (1996) A description of speak outs in Oklahoma prisons. Okla Crim Just Res Consort J 3(Aug):123–133

Lewis RV (1983) Scared straight–California style: evaluation of the San Quentin SQUIRES program. Crim Just Behav 10(2):209–226

Lipsey MW (1992) Juvenile delinquency treatment: a meta-analytic inquiry into the variability of effects. In: Cook TC, Cooper H, Cordray DS, Hartmann H, Hedges LV, Light RL, Louis TA, Mosteller FM (eds) Meta-analysis for explanation. Russell Sage, New York, pp 83–127

Lloyd C (1995) To scare straight or educate? The British experience of day visits to prison for young people, Home Office Research Study No. 149. Home Office, London

Locke TP, Johnson GM, Kirigin-Ram K, Atwater JD, Gerrard M (1986) An evaluation of a juvenile education program in a state penitentiary. Eval Rev 10:281–298

Michigan Department of Corrections (1967) A six month follow-up of juvenile delinquents visiting the Ionia Reformatory. Research report no. 4. Michigan Department of Corrections, Lansing

Middleton J, Reeves E, Lilford R, Howie F, Hyde C, Elbourne D, Oakley A, Gough D (2001) Collaboration with the Campbell Collaboration. Brit Med J 323:1252

Muhammed L (1999) Kids and crooks revisited: some were "Scared Straight!" USA Today, 12 Apr 1999, p 4D

Mullen C (2012) Bar Association: stop putting innocent kids in jail. San Juan Daily Star (Letter to the Editor), 14 Feb 2012, p 4

O'Malley P, Coventry G, Walters R (1993) Victoria's day in prison program: an evaluation and critique. Aust N Z J Criminol 26(2):171–183

Orchowsky S, Taylor K (1981) The insiders juvenile crime prevention program. Virginia Department of Corrections, Richmond

Petrosino A, Turpin-Petrosino C, Finckenauer JO (2000) Well meaning programs can have harmful effects!: Lessons from experiments in Scared Straight and other like programs. Crime Delinq 46:354–379

Petrosino A, Turpin-Petrosino C, Buehler J (2004) "Scared Straight" and other juvenile awareness programs for preventing juvenile delinquency. Campbell Syst Rev 2004:2

Petrosino A, Turpin-Petrosino C, Guckenburg S (2010) Formal system processing of juveniles: effects on delinquency. Campbell Syst Rev 2010:1

Rasmussen DW, Yu Y (1996) An evaluation of juvenile justice innovations in Duval County, Florida. Florida State University, Tallahasse

Robinson LO, Slowikowski J (2011) Scary – and ineffective. Traumatizing at-risk kids is not the way to lead them away from crime and drugs. Baltimore Sun, 31 Jan 2011. http://articles.baltimoresun.com/2011-01-31/news/bs-ed-scared-straight-20110131_1_straight-type-programs-straight-program-youths

Scripps J (1999) Prison tour serves as a wake-up call. The Forum, 27 Oct 1999, p 1

Shapiro A (1978) Scared straight [film]. Pyramid Films, Santa Monica

Storvall AE, Hovland A (1998) The Ullersmo project: scared straight in Norway 1992–1996 (Ullersmoprosjecktet: "Scared Straight" I Norge 1992–1996). Nordisk Tiddskrift for Kriminalvidenskab 85(2):122–135

Trusty G (1995) Sheriff's office, juvenile court resources used for restarts. Louisiana Youth Care Magazine 18(5):3–5

United Community Action Network (2001) Services: Wisetalk (scared straight). www.ucan.av.org/services.htm. Accessed 26 Apr 2001

United Press International (2003) Scared straight, 21 Aug, p 1008233w3635

Vreeland AD (1981) Evaluation of face-to-face: a juvenile aversion program. Unpublished doctoral dissertation, University of Texas, Dallas

Wilson D, Groombridge N (2010) I'm making a TV programme here: reality TV's *Banged up* and public criminology. Howard J Criminol 49(1):1–17

Yarborough JC (1979) Evaluation of JOLT as a deterrence program. Michigan Department of Corrections, Lansing

Scenario Designs

M. Lyn Exum and Megan Sims
Department of Criminal Justice and Criminology, University of North Carolina at Charlotte, Charlotte, NC, USA

Overview

Suppose you drove by yourself one evening to meet some friends at a bar that is about 10 miles from

> your house. You have been drinking throughout the evening and by the time you're ready to leave, you suspect your blood alcohol level might exceed the legal limit. Suppose you have to be at work early the next morning. You can either drive home or find some other way home, but if you leave your car, you will have to return early the next morning to pick it up. (Pogarsky 2004, p. 119)

Hypothetical scenarios like the one above are commonly used to study criminal decision making. In the typical study, participants are asked to envision themselves experiencing the situation described in the scenario and then to self-report how likely they would be to respond in an illegal manner (e.g., drive while drunk). They also rate various costs and benefits that might result from engaging in the offense. Using these data, researchers attempt to reconstruct the decision to offend by modeling participants' self-reported criminal intentions as a function of the perceived costs and benefits, along with other individual and situational factors.

This entry discusses the contributions of the hypothetical scenario method (HSM) to the study of criminal decision making. The rise of the HSM within the deterrence/rational choice literature is first presented, followed by a summary of the key findings from this literature. Finally, emerging areas of research and future challenges for HSM scholars are highlighted.

The Rise of the HSM

The deterrence and rational choice perspectives are the dominant theoretical frameworks in which to study criminal decision making. Both theories assert that human beings freely chose to engage in (or abstain from) crime after assessing the consequences of the act. Both theories also contend that differences in the qualities of these consequences – such as their certainty, swiftness, and severity – will lead to different decisional outcomes. However, the two theories diverge with respect to the *types* of consequences that influence criminal decision making. In its purest form, deterrence theory focuses on the crime-inhibiting effects of formal legal sanctions, such as arrest and incarceration (Nagin 1978). In contrast, the rational choice perspective recognizes that informal sanctions as well as the benefits of crime operate in conjunction with formal sanctions to shape the decision to offend (Grasmick and Bursik 1990; Piliavin et al. 1986).

Although the philosophical underpinnings of deterrence and rational choice can be traced back to the eighteenth century, empirical research on these theories did not appear until the 1960s (Paternoster 2010). These initial tests typically compared objective measures of punishment (such as incarceration rates) to aggregate rates of crime, and in general, the results showed weak/mixed support for the notion of deterrence. Scholars subsequently argued that instead of relying on objective measures of punishment, test of criminal decision making should focus on individuals' *perceptions* of punishment, regardless of how accurate those perceptions may be. From a methodological standpoint, this shift from objective to perceptual measures meant that researchers would now need to survey individuals and have them self-report their involvement in crime as well as their perceptions of the legal consequences of their behavior. Initially, this vein of research relied heavily on cross-sectional surveys; however, given the inherent temporal ordering problems in cross-sectional data, many scholars turned to longitudinal studies to study criminal decision making.

Notwithstanding their methodological advantages, longitudinal studies of deterrence and rational choice theory are unable to capture participants' perceptions of consequences *at the exact moment* the decision to offend (or abstain) is made. Instead, these studies capture self-reported perceived consequences at Time 1 and then use these perceptions to predict participants' subsequent involvement in crime occurring prior to the Time 2 testing period (which may be as much as 1–2 years in the future). With such a large time interval, the opportunity for perceptions to change is ever present. Although evidence suggests there is some similarity in the perceived consequences at Time 1 and Time 2, one cannot be fully certain that these perceptions remained stable throughout the duration of the study interval – including the moment when participants came face to face with a real-world criminal opportunity (see Piliavin et al. 1986, pp. 115–116).

In order to better measure the perceived consequences of crime at the moment the decision to offend/abstain occurs, some researchers began focusing on whether participants *believed* they would engage in a particular criminal act, as opposed to focusing on whether participants had actually committed the act (e.g., Grasmick and Bursik 1990). As participants then contemplated their penchant toward the offense, researchers could query them on the perceived consequences of offending. While this technique seemingly allows researchers to study criminal decision making at the moment it unfolds, it does not provide participants with adequate contextual information about the potential criminal opportunity – information that participants may need in order to accurately assess the consequences of the offense. As a result, participants' perceived consequences in these studies may be unreliable (Klepper and Nagin 1989).

In response, researchers turned to the hypothetical scenario method. Through the use of structured vignettes, the HSM allows researchers to present participants with an opportunity to offend – albeit an imaginary one. The vignettes, which are usually no more than a paragraph in length, contain important contextual information about the criminal opportunity (e.g., you are drunk at a bar 10 miles from your house, it's late, you need your car early the next day). Through the use of these vignettes, researchers can present participants with virtually any type of criminal/imprudent opportunity, such as sexual assault (Bouffard 2002b), physical aggression (Exum 2002), drunk driving (Piquero and Pogarsky 2002), corporate price fixing (Piquero et al. 2005), academic cheating (Tibbetts 1999), sports doping (Strelan and Boeckmann 2006), and even police misconduct (Pogarsky and Piquero 2004).

Unlike the typical cross-sectional or longitudinal studies of decision making, the measure of offending in HSM studies is not the participants' actual involvement in crime; instead, it is the participants' self-reported *intentions* to offend. Researchers acknowledge that such self-reported intentions are not measures of real-world behavior (Bachman et al. 1992; Paternoster and Simpson 1996) nor should they be interpreted as "synonymous with actual performance" (Tibbetts and Herz 1996, p. 203). At the same time, researchers also contend that measures of criminal intentions should be strongly correlated with actual offending (Tibbetts 1999; Piquero and Tibbetts 1996) and can be interpreted as indicators of participants' predisposition or proneness to crime (Paternoster and Simpson 1996; Pogarsky 2004).

In sum, the HSM is wholly distinct in its approach to studying the decision-making processes proposed by deterrence and rational choice theories. Unlike research using objective measures of punishment, the HSM captures individuals' subjective perceptions of the consequences of crime. Unlike traditional cross-sectional studies, the HSM does not have inherent temporal ordering problems. Unlike traditional longitudinal studies, the HSM allows researchers to measure participants' perceived consequences at the moment they contemplate a criminal opportunity. Unlike studies that simply ask about participants' intentions to commit some kind of generically defined crime, the HSM provides participants with a specific narrative for the criminal opportunity that, in turn, should make participants' responses more reliable. As a result of all these features, the HSM stands as one of the most versatile and methodologically sophisticated techniques in which to study the decision to offend.

Key Findings from the HSM Literature

The Role of Perceived Punishments

Much has been written about the empirical status of deterrence theory (e.g., Nagin 1978; Paternoster 2010; Pratt et al. 2008), especially as it pertains to the certainty and severity of formal legal sanctions. In general, these reviews find that the deterrent effect of sanction certainty – though weak – is greater than the effect for sanction severity. However, these reviews generally provide a summary of the deterrence literature as a whole and therefore cannot isolate the findings from only those studies using the HSM.

Pratt et al.'s (2008) meta-analysis is the exception. It disaggregates effect size estimates for punishment certainty and severity across studies of differing research methodologies, including the HSM. Interestingly, the findings from these HSM studies largely mirror those of other, more global reviews of the deterrence literature. The certainty effect for formal legal sanctions was found to be weak (mean effect size = −0.14) but nevertheless stronger than the effect for sanction severity (mean effect size = −0.05). The threat of nonlegal sanctions, such as the loss of respect or employment, was found to exert a deterrent effect that was identical to that of the certainty of formal sanctions (mean effect size = −0.14). (Note that the severity effect for these nonlegal sanctions was not estimated in the study.) Collectively, these meta-analytic findings suggest that punishment certainty and severity play a minor role in the decision to offend.

Far less attention has been devoted to studying the impact of the swiftness of punishment. In one of the few HSM studies to examine celerity, Nagin and Pogarsky (2001) used a factorial survey to manipulate the length of time in which the punishment for drunk driving was expected to occur. Across a variety of multivariate models and after controlling for a host of factors such as the certainty and severity of punishment, celerity did not exert a significant deterrent effect on self-reported intentions to drive drunk. In a subsequent HSM study examining police misconduct, Pogarsky and Piquero (2004) manipulated the swiftness of punishment an officer might experience from failing to charge an off-duty officer with drunk driving and from performing an unauthorized background check on a new neighbor. This time, the results for celerity were found to vary by the type of misconduct. The swiftness of punishment had a significant deterrent effect on failing to report a fellow officer for drunk driving but had no significant effect on performing an unauthorized background check. Pogarsky and Piquero speculate that these differences may be partly attributed to differences in the perceived gravity of the offenses, suggesting severity and swiftness may interact to produce an overall deterrent effect. Be that as it may, the findings from these two HSM studies tentatively suggest that the swiftness of punishment – at least, by itself – has a minimal impact on criminal decision making.

The Role of Perceived Benefits

Many HSM studies, especially those with a rational choice orientation, have examined the influence of the perceived benefits of crime – with "benefits" defined to include some form of physical and/or psychological reward. Across a host of HSM studies examining a variety of offenses, this research generally shows that the potential benefits of crime are a critical component in the decision to offend. For example, HSM research finds that the benefits from crime exert a positive and significant effect on self-reported intentions to engage in computer software piracy (Higgins 2007), drunk driving (Bouffard 2007; Nagin and Paternoster 1994), physical assault (Carmichael and Piquero 2004; Exum 2002), sexual assault (Loewenstein et al. 1997; Nagin and Paternoster 1994), theft/shoplifting (Piquero and Tibbetts 1996; Tibbetts and Herz 1996), corporate offending (Paternoster and Simpson 1996; Piquero et al. 2005), plagiarism (Ogilvie and Stewart 2010), and test cheating (Sitren and Applegate 2007; Tibbetts 1999). Furthermore, the benefits of crime have been shown to be even more influential to the decision to offend than are formal and informal criminal sanctions (Nagin and Paternoster 1993).

This is not to say that perceived benefits are universally predictive of offending. In fact, some HSM studies find certain benefits of crime (and under certain conditions) to be unrelated to offending intentions and/or to contribute little predictive power in multivariate models of criminal decision making (e.g., Bouffard 2002b; Elis and Simpson 1995; Ogilvie and Stewart 2010; Paternoster and Simpson 1996; Piquero et al. 2005; Simpson and Piquero 2002). Nevertheless, and as a whole, the empirical evidence from the HSM literature overwhelmingly indicates that the perceived payoffs from crime promote the decision to offend.

The Role of Consequence Salience

Bouffard (2002b) has argued that models of criminal decision making should not only consider the certainty and severity of perceived consequences but also the salience of those consequences. Conceptually, salience is distinct from certainty, severity, and swiftness in that it captures "how much weight a given factor carried in the actual decision-making process" (p. 123; see also Bouffard 2007). In other words, salience reflects how *important* the consequence is to the potential offender when contemplating the criminal opportunity. Additionally, salience is purported to be a dynamic quality that is influenced by state-level factors such as emotional arousal. As a result, individuals in emotionally charged states may identify consequences that are highly certain and highly severe, but in the "heat of the moment," individuals simply do not view these consequences as terribly important. Similarly, under certain emotional states, the benefits of a criminal act – no matter how unlikely or how inconsequential – may nevertheless carry a lot of weight in the decision to offend.

To examine whether salience is a unique concept, Bouffard (2007) presented participants with a hypothetical drunk driving scenario and asked them to report the certainty, severity, and importance of the consequences associated with the offense. Factor analysis revealed mixed results, with the salience measures for perceived costs indeed loading on a unique factor, but the salience measures for perceived benefits loading on a factor with perceived certainty. To determine whether salience is a dynamic element of decision making, Bouffard (2002b) randomly assigned male participants to view either sexually explicit or neutral images prior to reading a sexual assault scenario. Participants then rated the certainty, severity, and salience of their perceived consequences. Here again, the findings were mixed. Sexual arousal appeared to minimize the salience of the negative consequences of sexual aggression, but it did not enhance the importance of the perceived benefits.

Arguably, the most important question regarding consequence salience is whether it influences an individual's decision to offend. The HSM research examining this question is scarce and somewhat inconsistent. For example, Bouffard (2002a) reports a series of bivariate correlations between participants' intentions to drive drunk and the salience of their perceived consequences of the act. These correlations were found to be in the expected direction, statistically significant, and weak to moderate in magnitude. However, when salience measures were correlated with intentions to engage in sexually coercion, the correlations were generally weaker with only a few statistically distinguishable from zero. Furthermore, in a series of multivariate analyses of the sexual coercion data, Bouffard (2002b) found the salience of perceived benefits – but not of the perceived costs – to impact self-reported intention scores. As a whole, the extant research on salience suggests that it is a dimension of participants' perceived consequences that appears to have some level of influence on criminal decision making. More research in this area is needed, however.

The Role of Morality

Many HSM studies have examined the role of morality on criminal decision making. In this research, the perceived immorality of the act is treated as a type of internal sanction, with high immorality ratings predicted to be inversely related with offending intentions. Note that while immorality is similar in nature to the informal sanctions of guilt and shame, it is typically regarded as a distinct construct that is generally rated in terms of its severity (i.e., "how morally wrong would it be to...") but not its certainty.

Bachman and colleagues (1992) were among the first to use the HSM to examine the influence of morality on criminal decision making. After controlling for perceptions of formal and informal sanctions, participants' ratings of immorality exerted a significant inhibitory effect on their self-reported intentions to engage in sexual assault. This inhibitory effect of morality has been found in other HSM studies as well, including those examining physical assault, drunk driving, corporate offending, shoplifting/theft, sports doping, and academic cheating (Bachman et al. 1992; Carmichael and Piquero 2004; Elis and Simpson 1995; Higgins 2007; Loewenstein et al. 1997; Nagin and Paternoster 1994; Paternoster and Simpson 1996; Piquero et al. 2005; Piquero and Tibbetts 1996; Simpson and Piquero 2002; Strelan and Boeckmann 2006; Tibbetts and Herz 1996).

Morality is sometimes viewed as the gatekeeper of the hedonic calculus. That is, before the perceived costs or benefits of a criminal act can be incorporated into the decision to offend, the act itself must first be a part of the person's repertoire of morally acceptable behavior. If it is not, then the individual is fully constrained by his morality and – as a result – any threats of punishment or enticements from benefits are irrelevant. If such a gatekeeper conceptualization is accurate, then the perceived consequences of crime should have their greatest effects among those with weak moral inhibitions and have their weakest effects among those with strong moral inhibitions (Bachman et al. 1992). Although only a few HSM studies have specifically sought to examine this interaction between morality and the role of perceived consequences, there is evidence to support the gatekeeper function of morality.

For example, Bachman et al. (1992) found that among participants who rated a sexual assault scenario to be more morally offensive, perceptions of formal punishments did not significantly alter the intention to offend. Yet, among participants who found the sexual assault to be less morally offensive, formal sanctions exerted a significant deterrent effect. Similar results have also been found among business executives and MBA students who read corporate crime scenarios (Paternoster and Simpson 1996). Although more research is needed in this area, the HSM studies that are available indicate that crime is a nonmarket area among individuals with strong moral fortitude. In other words, among the righteous, the decision to act is less utilitarian and more deontological.

The Role of Self-Control

Conceived as time-stable personality trait, self-control is purported to exert a *direct* effect on criminal intentions (Piquero and Tibbetts 1996). That is, because individuals with low self-control are impulsive risk takers, they should be less able to abstain from a criminal opportunity when it presents itself. However, self-control is also thought to exert an *indirect* effect on offending intentions by altering the individuals' perceptions of consequences of crime. That is, given that individuals with low self-control are present oriented, they should be less able to perceive the long-term negative consequences of their criminal actions, thereby resulting in greater criminal proclivities. There is some HSM research to support both of these claims.

For example, Nagin and Paternoster (1993) found evidence for a direct effect of self-control on offending intentions. After controlling for such factors as prior criminal behavior, the characteristics of the criminal opportunity, and the perceived costs/benefits of the act, a one standard deviation change in low self-control scores resulted in a 39 % increase in theft intentions, a 17 % increase in drunk driving intentions, and an 83 % increase in sexual assault intentions (see also Nagin and Paternoster 1994). In support of an indirect effect, Piquero and Tibbetts (1996) found that low self-control accentuated the perceived benefits and minimized the perceived costs associated with shoplifting and drunk driving – effects that, in turn, significantly influenced participants' intentions to offend (see also Carmichael and Piquero 2004; Higgins 2007).

However, most HSM studies that have included a measure of self-control have found either null effects or effects that are not stable across different models of decision making. For example, neither Tibbetts and Myers (1999) nor Sitren and Applegate (2007) found evidence that self-control levels directly influence college students' intentions to cheat on an exam. Similarly, Simpson and Piquero (2002) found no evidence to suggest that self-control levels predict corporate offending among business executives and MBA students. Furthermore, others have shown that the effects of self-control operate inconsistently across the sex of the participant (Tibbetts and Herz 1996), the operational definition of self-control (Piquero and Bouffard 2007), the dependent variable (Loewenstein et al. 1997; Pogarsky and Piquero 2004), and the participants' anticipated mood when completing the hypothetical act (Carmichael and Piquero 2004). Thus, while there is some support for the notion that self-control levels directly and indirectly impact criminal decision making, the overarching findings from the HSM literature raise doubt as to whether these effects are, in fact, noteworthy and robust.

The Role of Altered States of Mind

Tests of deterrence and rational choice theory have generally ignored the offender's emotional, physiological, and psychopharmacological state during the criminal decision-making process. This omission is striking given the proposed theoretical relationships between altered states of mind and cognition (see Exum 2002; Loewenstein et al. 1997). Fortunately, the HSM is easily capable of incorporating the study of altered states on decision making, and a few such studies have taken advantage of this.

For example, in their study on the effects of sexual arousal on sexually aggressive decision making, Loewenstein et al. (1997) randomly assigned male participants to view photographs of nude women or of fully clothed fashion models prior to reading a hypothetical date rape scenario. Compared to the non-aroused participants, those who were aroused endorsed greater sexually aggressive intentions. However, few differences were found across arousal conditions with respect to participants' perceived consequences of sexual aggression. In a similar experiment, Bouffard (2002b) found that self-reported arousal levels were positively related to offending intentions but inconsistently related to the perceived consequences. In a subsequent and more detailed analysis, Bouffard (2011) concluded there are in fact differential deterrent effects across arousal states, with the perceived costs of crime exerting their greatest deterrent effects among those who reported lower arousal levels. Collectively, these studies suggest that arousal increases sexually aggressive tendencies – not by altering participants' perceptions of the consequences – but by somehow moderating the underlying relationship between perceived consequences and the intention to offend.

Also using an experimental design, Exum (2002) examined the effect of alcohol intoxication and anger on participants' responses to a physical assault scenario. Participants were randomly assigned to drink either a nonalcoholic or alcoholic beverage, the latter designed to elevate blood alcohol levels to approximately 0.08 %. Half of the participants in each group were then confronted by the researcher in such a way to induce anger, whereas the remaining half was not. Participants then read a physical assault scenario and responded to a battery of rational choice questions. Results indicated that alcohol and anger interacted to increase one of two measures of aggressive intentions; however, consistent with other research (Loewenstein et al. 1997; Bouffard 2011), there was no support for the idea that the altered states of mind affected participants' perceptions of costs and benefits. At the same time, exploratory analyses revealed that the collection of consequences used to model decision making were not equally predictive of intentions across the experimental conditions. This suggests that – while the perceptions of costs and benefits were largely the same across all participants – those in the intoxication and anger conditions used these perceptions differently in their decision-making processes. These findings further suggest that altered states of mind moderate the decision to offend.

Carmichael and Piquero (2004) also examined the effect of perceived anger on physical assault. Anger was not experimentally manipulated in this study; instead, participants were asked to read a bar fight scenario and then self-report how angry they thought they would feel in that particular situation. Controlling for the perceived costs and benefits of the assault, anger scores were positively and significantly related to aggressive intentions. Furthermore, the perceived negative consequences of assault exerted significant deterrent effects among the low anger group but not in the high anger group, again suggesting that anger moderates the decision-making process. Collectively, this small body of HSM research serves as a reminder that decision making does not occur in a vacuum and that emotional, visceral, and pharmacological forces shape the decision to offend.

Conclusions from the HSM Literature

The HSM research on criminal decision making finds that the certainty of punishment (both formal and informal) has a weak deterrent effect and the severity of punishment has an even weaker effect. Although rarely examined, the swiftness of punishment also seems to carry little weight in the decision to offend. In stark contrast, the perceived benefits from crime appear to be a key element in criminal decision making. However, some criminal acts appear to be so morally objectionable to participants that they will simply refrain from the act regardless of the benefits. Similarly, dynamic factors such as one's state of mind may alter the salience of these perceived consequences as well as moderate the rational process thought to underlie criminal decision making. Trait-level characteristics such as low self-control may not only make it more difficult for individuals to self-regulate their behavior in the presence of a criminal opportunity, but they may also change the way individuals assess the risks and benefits associated with that opportunity (but note, however, that the self-control effects may not be very robust). As a result of these various state- and trait-level factors, criminal decisions that appear to be completely "irrational" to the outside observer be seen as wholly rational from within the perspective of the actor.

Emerging Areas of Research and Challenges

Given that the HSM is a relatively new technique within the study of criminal decision making, one could consider most any line of HSM research to be an "emerging" area. With that being said, below are four areas of criminological HSM research that deserve special attention. These four issues have great potential to advance (or challenge) much of what criminologists claim to know about the decision to offend.

Studying Deterrence Among Deterrables

Although there is reason to believe that individuals vary in their responsivity to the perceived consequences of crime (see Bachman et al. 1992; Carmichael and Piquero 2004; Paternoster and Simpson 1996), tests of deterrence theory typically ignore these individual susceptibilities and instead examine the properties of punishment across an aggregate sample of participants. Pogarsky (2002) challenged this approach, arguing that deterrence theory should be tested only among those who are truly deterrable, as opposed to those who are either so bound to conformity or so wedded to deviance that extralegal sanctions have no impact on their behavior. Using the HSM and participants' self-reported intentions to drive drunk, Pogarsky (2002) developed a method of identifying the subsample of "deterrables" within a group of study participants. He then examined the impact of formal sanctions among the deterrables and, for comparison, among the full sample.

Results from the full sample largely confirm the findings from past studies of deterrence theory – namely, that certainty has a greater effect than severity. However, in the analysis of only the deterrables, severity had a greater effect than certainty. Thus, the conventional wisdom that "certainty is more important that severity" appears to be challenged when studying only those people who are, in fact, swayed by sanction threats. Future HSM research should build on Pogarsky's method of disaggregating samples into deterrable and non-deterrable groups and continue to examine the role of formal and informal sanctions among those most responsive to punishments. The findings have important implications not only for criminological theory but for public policy.

The Positive Punishment Effect

Contrary to the notion that sanctions deter future behavior, some studies find a *positive* relationship between punishment and subsequent offending (e.g., Piquero and Pogarsky 2002).

Pogarsky and Piquero (2003) outlined two possible explanations for this positive punishment effect: selection and resetting. The selection explanation contends that offenders with stronger criminal propensities are more likely to come to the attention of the police (i.e., get punished) and also be more likely to recidivate. In contrast, the resetting explanation argues that offenders who experience a stroke of bad luck (e.g., get punished) are more apt to take elevated risks in the future (e.g., recidivating) because they believe their bad luck is not likely to return anytime soon. In other words, shortly after experiencing punishment, the offender's perceived certainty of sanctioning is "reset" to a nominal value.

To examine the possibility of resetting, Pogarsky and Piquero (2003) administered a hypothetical drunk driving scenario to a sample of undergraduates and asked them a series of questions about their perceptions of punishment and prior punishment experiences. Support for the resetting explanation was mixed and varied by the participant's level of risk for offending. That is, consistent with the notion of resetting, low-risk offenders who had previously been punished reported lower certainty estimates than those low-risk offenders who had not been previously punished. However, no evidence of resetting was found among high-risk offenders, with all high-risk offenders perceiving comparable sanctioning risks regardless of past punishment experiences. While more research is needed in this area to better understand the nature of positive punishment effects, future studies should also seek to integrate research on "deterrables" with that on resetting. Although it is too early to tell, it may be the case that those who are most responsive to sanction threats (the deterrables) are also the very same "low-risk" offenders who tend to reset their perceived threat levels following punishment.

Researcher-Generated Versus Self-Generated Consequences

When potential offenders encounter a criminal opportunity in the real world, they freely deduce their own set of costs and benefits to contemplate. However, in the typical HSM study, participants are asked to contemplate a predetermined, uniform set of consequences generated by researchers. Such a reliance on researcher-generated consequences inevitably forces an artificial structure on the otherwise organic decision to offend. Bouffard, Exum, and Collins (2010) examined the impact of this artificial structure by presenting participants with a hypothetical shoplifting scenario and then randomly assigning them to either (1) assess a standard battery of researcher-generated consequences or (2) assess a set of costs and benefits that the participants themselves self-generated. By comparing and contrasting the two sets of consequence ratings, Bouffard and colleagues could begin to examine the extent to which individuals' decision-making processes were being influenced by the study's methodology.

The findings show that when participants are asked to assess the benefits of crime that were identified by the researcher, they are more likely to regard the consequences as possible outcomes (Bouffard et al., 2010). Illustratively, when participants were specifically asked to consider the "thrill" derived from the hypothetical theft, 64 % reported the act would afford some degree of fun/excitement. In contrast, when participants were asked to self-identify all the "good things" that might happen to them if they committed the hypothetical act, only 10 % indicated it would be fun/exciting. Similarly, participants were more likely to see the costs of shoplifting as potential outcomes when these costs were identified by the researcher versus when they were self-generated. A noteworthy example is the cost of deviating from one's morality. Whereas 99 % of participants who received researcher-generated consequences reported that shoplifting felt immoral, only 3 % of participants who self-generated consequences reported feelings of immorality.

Recall that past research using the HSM finds moral inhibitions to be influential considerations in the decision to offend (e.g., Bachman et al. 1992; Loewenstein et al. 1997; Paternoster and Simpson 1996). The findings from Bouffard et al.'s (2010) study suggest that the morality effect reported in this literature may, in part, be a methodological artifact and largely attributed to the use of researcher-generated consequences. However, future research is needed to better understand how the hypothetical scenario methodology (in general) and the manner in which costs and benefits are presented (in particular) impact the findings from HSM studies.

The Accuracy of Self-Reported Intentions to Offend

Recall that the typical dependent variable in HSM studies is not participants' actual offending behavior but is instead their self-reported *intentions* to offend. According to the theory of reasoned action, such intentions can be viewed as valid estimates of real-world behavior because "...barring unforeseen events, a person will usually act in accordance with his or her intention" (Ajzen and Fishbein 1980, p. 5). Consistent with this notion, research generally finds the correlation between participants' intentions to act and their corresponding behavior to be positive and significant (Ajzen and Fishbein 1980; Armitage and Conner 2001). However, this correlation is largely based on the study of conventional behaviors such as voting, watching television, or going to church. Only a handful of studies have examined the intentions/behavior relationship for deviant behaviors, and while the results from these studies are promising on the surface, a closer examination reveals potential methodological and/or measurement problems that may weaken the interpretation of these intentions/behavior correlations (Exum et al. 2011; cf. Pogarsky 2004).

To further examine the predictive accuracy of self-reported intentions to offend, Exum et al. (2011) gave undergraduate students a copy of a fictitious newspaper article describing an ongoing music piracy operation on campus. Students were led to believe the article was real and had recently appeared in the local paper. The article described a graduate student who was reportedly emailing illegal copies of digital music files for free to anyone with a university email account. Also included in the article were the graduate student's name and email address. After reading the article, study participants were asked to self-report their likelihood of contacting the graduate student and requesting illegal files from him. In the following weeks, the graduate student – who was a research confederate – monitored his email account for any requests from participants. In this way, participants' self-reported intentions to offend could be validated against their own requests for illegal music files.

Despite the variability in offending intentions, including a few participants who reported a 100 % chance of emailing the graduate student, no one in the study actually requested music files from the confederate. This suggests that self-reported criminal intentions (at least, as they apply to this form of music piracy) have a low false-negative rate but a high false-positive rate. Stated differently, all participants who reported little-to-no intentions of contacting the confederate abstained from the behavior, resulting in a false-negative rate of 0 %. However, all participants who reported strong-to-definitive intentions of contacting the confederate abstained as well, resulting in a false-positive rate of 100 % (Exum et al. 2012). This suggests that when presented with a hypothetical offending scenario, participants may fall prey to a false sense of bravado and self-report strong intentions to offend, only later to cower at the real-world opportunity. Given that self-reported intentions to offend are an essential element of HSM studies of deterrence/rational choice, additional research is needed to examine the accuracy of these offending intentions (see also Pogarsky 2004). The implications of this line research have the potential to challenge most everything we know about the nature of criminal decision making.

Conclusion

More than 200 years ago, Jeremy Bentham (1970/1789) described the deliberative process in which all individuals are thought to engage as they contemplate a criminal act:

> Sum up all the values of all the *pleasures* on one side, and those of all the *pains* on the other. The balance, if it be on the side of pleasure, will give the *good* tendency of the act upon the whole with the respect to the interest of that *individual* person.... (p. 40, emphasis in the original)

Two centuries later, criminologists continue to struggle with the best way to test these ideas. The biggest impediment appears to be obtaining a measure of participants' perceived "pleasures" and "pains" at the moment participants decide to engage in or abstain from crime. The hypothetical scenario method is arguably the best technique to date for doing this, thereby making it uniquely situated to study criminal decision making at the moment it unfolds. Today, much of what we now know about the decision to offend – e.g., the negligible effect of punishment, the lure of perceived benefits, and the different calculative processes across individual- and state-level factors – has been either discovered through or confirmed by HSM research.

As scholars seek to build upon our current understanding of criminal decision making, they will inevitably rely on the HSM to some degree. The technique's versatility, its ease of administration, and its ability to accommodate experimental manipulations make it an ideal platform for designing and implementing research studies. At the same time, the HSM's greatest strength – i.e., its ability to get inside the "black box" of the mind and study decision making during the contemplation of a criminal opportunity – comes with a great weakness. The HSM can only allow researchers to model the decisions that underlie criminal *intentions* and not actual criminal *behavior*. As a result, the HSM can inform our understanding of real-world criminal decision making only to the extent that individuals' self-reported intentions are predictive of their actual criminal conduct. Until the predictive accuracy of criminal intentions is more fully understood, the study of criminal decision making will benefit most from triangulating the findings from a host of studies that use a variety of research methodologies, including the HSM.

Related Entries

- ▶ Bayesian Updating and Crime
- ▶ Deterrence
- ▶ Deterrence: Actual Versus Perceived Risk of Punishment
- ▶ Rational Choice and Prospect Theory
- ▶ Rational Choice Theory
- ▶ Rational Choice, Deterrence, and Crime: Sociological Contributions

Recommended Reading and References

Ajzen I, Fishbein M (1980) Understanding attitudes and predicting social behavior. Prentice-Hall, Englewood Cliffs

Armitage CJ, Conner M (2001) Efficacy of the theory of planned behavior: a meta-analytic review. Br J Soc Psychol 40:471–499

Bachman R, Paternoster R, Ward S (1992) The rationality of sexual offending: testing a deterrence/rational choice conception of sexual assault. Law Soc Rev 26:343–372

Bentham J (1970) An introduction to the principles of morals and legislation. Oxford University Press, New York (Original work published 1789)

Bouffard JA (2002a) Methodological and theoretical implications of using subject-generated consequences in tests of rational choice theory. Justice Q 19:747–769

Bouffard JA (2002b) The influence of emotion on rational decision making in sexual aggression. J Crim Justice 30:121–134

Bouffard JA (2007) Rational choice theory revisited: a preliminary examination of the perceived importance of consequences in offender decision-making. Int J Crime Crim Justice Law 2:49–67

Bouffard J (2011) 'In the heat of the moment': mediating versus moderating relationships between sexual arousal and perceived sanctions. J Crime Justice 34:24–44

Bouffard J, Exum ML, Collins P (2010) Methodological artifacts in tests of rational choice theory. J Crim Justice 38:400–409

Carmichael S, Piquero AR (2004) Sanctions, perceived anger, and criminal offending. J Quant Criminol 20:371–393

Elis LA, Simpson SS (1995) Informal sanction threats and corporate crime: additive versus multiplicative models. J Res Crime Delinquency 32:399–424

Exum ML (2002) The application and robustness of the rational choice perspective in the study of intoxicated and angry intentions to aggress. Criminol 40:933–966

Exum ML, Turner MG, Hartman JS (2012) Self-reported intentions to offend: all talk and no action? Am J Crim Justice. 4:523–543

Grasmick HG, Bursik RJ (1990) Conscience, significant others, and rational choice: extending the deterrence model. Law Soc Rev 24:837–861

Higgins GE (2007) Digital piracy, self-control theory, and rational choice: an examination of the role of value. Int J Cyber Criminol 1:33–55

Klepper S, Nagin D (1989) The deterrent effect of perceived certainty and severity of punishment revisited. Criminol 27:721–746

Loewenstein G, Nagin D, Paternoster R (1997) The effect of sexual arousal on expectations of sexual forcefulness. J Res Crime Delinq 34:443–473

Nagin DS (1978) General deterrence: a review of the empirical evidence. In: Blumstein A, Cohen J, Nagin D (eds) Deterrence and incapacitation: estimating the effects of criminal sanction on crime rates. National Academy Press, Washington, DC, pp 95–139

Nagin D, Paternoster R (1993) Enduring individual differences and rational choice theories of crime. Law Soc Rev 27:467–496

Nagin DS, Paternoster R (1994) Personal capital and social control: the deterrence implications of a theory of individual differences in criminal offending. Criminology 32:581–606

Nagin DS, Pogarsky G (2001) Integrating celerity, impulsivity, and extralegal sanction threats into a model of general deterrence: theory and evidence. Criminology 39:865–892

Ogilvie J, Stewart A (2010) The integration of rational choice and self-efficacy theories: a situational analysis of student misconduct. Aust New Zeal J Criminol 43:130–155

Paternoster R (2010) How much do we really know about criminal deterrence? J Crim Law Criminol 100:765–824

Paternoster R, Simpson S (1996) Sanction threats and appeals to morality: testing a rational choice model of corporate crime. Law Soc Rev 30:549–583

Piliavin I, Gartner R, Thornton C, Matsueda RL (1986) Crime, deterrence, and rational choice. Am Sociol Rev 51:101–119

Piquero AR, Bouffard JA (2007) Something old, something new: a preliminary investigation of Hirschi's redefined self-control. Justice Q 24:1–27

Piquero AR, Pogarsky G (2002) Beyond Stafford and War's reconceptualization of deterrence: personal and vicarious experiences, impulsivity, and offending behavior. J Res Crime Delinq 39:153–186

Piquero A, Tibbetts S (1996) Specifying the direct and indirect effects of low self-control and situational factors in offenders' decision-making: toward a more complete model of rational offending. Justice Q 13:481–510

Piquero NL, Exum ML, Simpson SS (2005) Integrating the desire-for-control and rational choice in a corporate crime context. Justice Q 22:252–280

Pogarsky G (2002) Identifying "deterrable" offenders: implications for research on deterrence. Justice Q 19:431–452

Pogarsky G (2004) Projected offending and contemporaneous rule-violation: implications for heterotypic continuity. Criminology 42:111–136

Pogarsky G, Piquero AR (2003) Can punishment encourage offending? Investigating the "resetting" effect. J Res Crime Delinq 40:95–120

Pogarsky G, Piquero AR (2004) Studying the reach of deterrence: can deterrence theory help explain police misconduct? J Crim Justice 32:371–386

Pratt TC, Cullen FT, Blevins KR, Daigle LE, Madensen TD (2008) The empirical status of deterrence theory: a meta-analysis. In: Cullen F, Wright J, Blevins K (eds) Taking stock: the status of criminological theory. Transaction, New Brunswick, pp 367–395

Simpson SS, Piquero NL (2002) Low self-control, organizational theory, and corporate crime. Law Soc Rev 36:509–548

Sitren AH, Applegate BK (2007) Testing the deterrent effects of personal and vicarious experience with punishment and punishment avoidance. Deviant Behav 28:29–55

Strelan P, Boeckmann RJ (2006) Why drug testing in elite sports does not work: perceptual deterrence theory and the role of personal moral beliefs. J Appl Soc Psychol 36:2909–2934

Tibbetts SG (1999) Differences between women and men regarding decisions to commit test cheating. Res High Educ 40:323–342

Tibbetts SG, Herz DC (1996) Gender differences in factors of social control and rational choice. Deviant Behav Interdiscip J 17:183–208

School Aggression

► Bullying Prevention: Assessing Existing Meta-Evaluations

School Bullying as a Risk Factor for Later Criminal Offending

Maria M. Ttofi[1], David P. Farrington[1] and Friedrich Lösel[1,2]

[1]Institute of Criminology, University of Cambridge, Cambridge, UK

[2]Institute of Psychology, University of Erlangen–Nuremberg, Erlangen, Germany

Overview

School bullying has received the attention of researchers and program planners in both developed and developing countries. It is a special category of aggressive behavior that has been addressed through numerous anti-bullying programs and, in some cases, through wider multiple component programs. Various anti-bullying agencies have highlighted the importance of intervention research for the development of safer school communities, where students can develop their full potential without being exposed to bullying and its detrimental effects. A vast number of cross-sectional studies have provided evidence of the negative impact of bullying on children's concurrent health.

This entry reports on an updated systematic review and meta-analysis that was undertaken under the aegis of the Swedish National Council for Crime Prevention and further supported by the British Academy and conducted by the current authors. Only longitudinal prospective studies were included in the review, which aimed to examine to what extent school bullying predicts later offending and violence. Significant effect sizes were found even after controlling for other major childhood risk factors. Being a bully increased the likelihood of being an offender by more than half and increased the likelihood of being violent by two thirds. These results either reflect the persistence of an underlying aggressive or antisocial tendency or a facilitating effect of school bullying on later offending and violence (or both).

The implication is that high quality bullying prevention programs (and possibly multiple component programs which also target aggression) should be promoted. They could be viewed as an early form of crime prevention. They can potentially have long-term effects by improving the future psychosocial adjustment of school bullies and reducing the associated health, welfare, education, and other costs.

Introduction

School bullying has recently become a topic of major public concern and has attracted a lot of media attention, with articles in major newspapers and magazines reporting cases of children who committed (or attempted) suicide because of their victimization at school and parents suing school authorities for their failure to protect their offspring from continued bullying victimization (e.g., Ttofi and Farrington 2012). There is, nevertheless, a number of "skeptics" who still perceive school bullying as being part of a normal developmental process, or as one of those school experiences that prepare children for the grown-up world. Scientific evidence regarding possible detrimental effects of school bullying on children's mental health and future psychosocial adjustment can only be provided through a systematic review and meta-analysis, providing an unbiased standardized effect size and defining the magnitude of the effect.

Background Research

School bullying is a special category of aggressive behavior involving repeated unprovoked acts against less powerful (emotionally or physically) individuals (Farrington 1993; Olweus 1993). Of course, schools, like other institutions, will always be a place in which the basic human motive of aggression will be demonstrated. However, school bullying should not be confused with more or less normal aggressive interactions such as rough and tumble play.

Scientific interest in the problem of bullying and its negative short-term and long-term effects emerged after the well-publicized suicides of three Norwegian boys in 1982, which were attributed to severe peer bullying (Olweus 1993). School bullying has gradually become a topic of major public concern via "bullying awareness days," national initiatives in various (European) countries (Smith and Brain 2000), and anti-bullying research networks across the world (e.g., Anti-Bullying Alliance; BRNET; International Observatory for Violence in Schools; PREVNet).

Any suggestion regarding the short-term negative impact of peer aggression and victimization seems reasonable even to the lay mind. Establishing, on the other hand, the long-term effects of school bullying and arguing that children involved in peer aggression are more likely to follow an antisocial path (compared with noninvolved students) is more challenging. Some early longitudinal studies did provide evidence of the long-term impact of school bullying and, notably, established the intergenerational transmission of school bullying. In the Cambridge Study in Delinquent Development, for example, boys who were bullies at age fourteen tended, at age thirty-two, to have children who were bullies (Farrington 1993). As another example, in his follow-up study of over 700 Stockholm boys, Olweus (1993) reported that 36 % of bullies at ages thirteen to sixteen were convicted three or more times between ages sixteen and twenty-four, compared with 10 % of the remainder.

There have been surprisingly few recently published longitudinal studies on the developmental pathways of children involved in school bullying since the seminal work of Olweus in Scandinavia and some other European examples. Two special issues in peer-reviewed journals have recently been published in an attempt to address this gap in research literature (Farrington et al. 2011; Ttofi et al. 2011a). Both issues presented new findings on the long-term negative consequences of school bullying based on major prospective longitudinal studies from around the world. Longitudinal investigators of twenty-nine studies conducted analyses for a more comprehensive British Academy funded project, which examined the long-term association of school bullying with both internalizing (such as anxiety, self-esteem, and stress) and externalizing (such as aggression, alcohol, and drug use) problems (see Farrington et al. 2012, Table 4, for a list of all contributors).

The special issue of *Criminal Behaviour and Mental Health* focused on the association between bullying perpetration at school and offending later in life. A systematic review and meta-analysis on the topic was carried out (Ttofi et al. 2011c). The special issue of the *Journal of Aggression, Conflict and Peace Research* focused on the association between bullying victimization (i.e., being bullied) and internalizing problems later in life, such as anxiety and depression. A systematic review and meta-analysis was again carried out examining the extent to which bullying victimization at school predicted depression (Ttofi et al. 2011b), showing that the probability of being depressed up to seven years later in life (M = 7.13 years; SD = 8.79) was significantly higher for victims of school bullying than for control students, i.e., children not involved in school bullying.

Building upon the above-mentioned research activities, an effort was made to update the relevant systematic reviews (Farrington et al. 2012) and to study further outcomes, such as violence (Ttofi et al. 2012). This entry presents results from the updated systematic review on the association of school bullying with offending later in life. Additional analyses are presented on the long-term link of bullying with violence.

Methods

The main objectives of the systematic review were two-fold. Firstly, to assess whether bullying at school (perpetration and victimization) was a significant risk factor predicting offending and violence later in life (unadjusted effect sizes). Secondly, to assess whether these associations were still significant after controlling for other major childhood risk factors, measured at the baseline period (adjusted effect sizes). Results on offending and violence were carefully treated in separate analyses and the outcome measures under each category generally did not overlap. However, it is possible that in some studies, outcome measures such as “police arrests” would include violence. “Offending” included outcome measures such as police or court contact, property offending, criminal convictions, property theft, vandalism, shoplifting, vehicle theft, etc. “Violence” included outcome measures such as forced sexual contact, criminal violence, physical fights, violent convictions, violent offending, weapon carrying, assault, etc.

Further analyses were conducted to investigate moderators that might explain variability in effect sizes, such as the age at which bullying was measured (Time 1), the age at which the outcome measures were taken (Time 2), the number of covariates controlled for in the adjusted effect sizes, the length of the follow-up period (measured in years), and the way in which the outcomes were measured (i.e., official data versus self-reports).

Stringent inclusion and exclusion criteria were set in advance. For example, reports were included only if they were based on prospective longitudinal data. The predictor must have been a measure of school bullying (and not other more general forms of peer aggression/victimization) and must have preceded the outcome (i.e., offending or violence). A clear measurement of offending and/or violence must have been included in the report as an outcome measure. Studies were included if participants were school-aged children in the community and exposure to bullying (perpetration and victimization) specified the school years. Published and unpublished reports of the literature were included in order to minimize the possibility of publication bias in the results.

Reports were excluded if the character of the data was qualitative in nature (e.g., qualitative data based on interviews) and did not allow calculation of an effect size. This did not apply if a qualitative method (e.g., interviews or observation studies) was used to obtain a quantitative measure. If the outcome measure (offending or violence) was part of a wider theoretical construct (e.g., externalizing or antisocial behavior), then the relevant report was again excluded. Reports based on clinic samples or incarcerated youth were also excluded.

Extensive searches were carried out and a detailed description of them can be found in the Swedish report (Farrington et al. 2012) and the most recent work focusing on violent outcomes (Ttofi et al. 2012). In total, the same searching strategies were repeated in 19 electronic databases, and the full volumes of 63 journals were searched either online or in print. In the Swedish report, readers can also find detailed tables of the key features of each report, such as the sample size, the country where the study took place, the exact confounds controlled for at the baseline period, etc.

Results

In total, 661 reports concerned with the association of school bullying with internalizing (e.g., anxiety, depression, self-esteem, etc.) and externalizing (e.g., aggressive behavior, conduct problems, offending, etc.) problems were located. All reports were screened in line with the inclusion and exclusion criteria and classified in five different categories (see Farrington et al. 2012, Table 5). Further to a detailed screening of all manuscripts, a total number of 48 reports, corresponding to 29 longitudinal studies, presented data on the long-term association of school bullying (perpetration and victimization) with offending in adolescence or young adulthood (see Farrington et al. 2012, Table 6). A total number of 51 reports from 28 longitudinal studies were included in the systematic review on the association of school bullying (perpetration and victimization) with violence in adolescence or young adulthood (see Ttofi et al. 2012, Table 1).

When different manuscripts relating to the same longitudinal study reported different effect sizes (because of differences, e.g., in the sample size or in the follow-up period that the authors have used), the combination of effect sizes across reports is not straightforward as these effect sizes are based on dependent samples. These dependencies were taken into account, as ignoring them would result in standard errors that were too small, often by a large degree. Advice from leading experts in the field was sought on this matter (Wilson 2010). Clear rules were set in advance for computing effect sizes across reports from the same longitudinal study (see Farrington et al. 2012; Ttofi et al. 2012).

Bullying Perpetration at School and Offending Later in Life

Eighteen studies provided an effect size for bullying perpetration versus offending. The summary unadjusted effect size across the 18 studies was OR = 2.64 (95 % CI: 2.17–3.20; z = 9.83) for the random-effects model. The random-effects model was used since the heterogeneity test, Q, of 84.89 was highly significant at p = 0.0001. When the three studies with only unadjusted effect sizes were excluded, the summary effect size for the remaining 15 studies – for the random-effects model – was OR = 2.54 (95 % CI: 2.05–3.14, z = 8.52). Again, there was significant variability in effect sizes across these studies (Q = 76.03, p = 0.0001).

After controlling for covariates, the adjusted summary effect size was reduced to OR = 1.89 (95 % CI: 1.60–2.23; z = 7.49) but this was still highly significant (see Farrington et al. 2012; Figs. 3 and 4). This OR indicates quite a strong relationship between bullying perpetration and later offending. For example, if a quarter of children were bullies and a quarter were offenders, this value of the OR would correspond to 34.5 % of bullies becoming offenders, compared with 21.8 % of non-bullies. Thus, being a bully increases the risk of being an offender (even after controlling for other childhood risk factors) by more than half.

For the adjusted summary effect size, various moderators were investigated to explain the heterogeneity in effect sizes across studies, which was significant (Q = 36.82, p = 0.001). These included the number of covariates controlled for at baseline (range: 1–20; M = 7.00; SD = 5.22), the age at which school bullying was measured (range: 6.23–15.54; M = 11.26; SD = 2.68), the age of participants when outcome measures were taken (range: 10.00–24.64; M = 17.10; SD = 4.91), and the length of the follow-up period, measured in years (range: 0.42–16.50; M = 5.84; SD = 4.56).

The age at which bullying was measured was positively associated with the effect size, but the regression coefficient was not statistically significant (B = 0.019, SE = 0.024, p = 0.428). The length of the follow-up period was significantly negatively associated with the effect size (B = −0.027, SE = 0.012, p = 0.018). As expected, the age of the study participants when outcome measures were taken was significantly negatively related to the effect size (B = −0.025, SE = 0.012, p = 0.039). The above two negative relationships suggest that bullying perpetration has a stronger effect in the short term. The relationship between the number of covariates controlled for and the effect size was in the expected negative direction and also significant (B = −0.027, SE = 0.013, p = 0.037). Therefore, the adjusted effect size decreased as the number of covariates controlled for increased.

Other moderators that may explain variability in effect sizes include the type of longitudinal studies (i.e., prospective versus retrospective) and the way in which the outcomes were measured (i.e., official data versus self-reports). In the Farrington et al. (2012) report, the reader can obtain information about these moderators (see their Table 6). Only three studies out of fifteen presented outcome measures based on official records for offending, making a moderator analysis inappropriate (due to uneven study numbers). Finally, only one study presented results based on a retrospective measure of bullying victimization, so any analyses on this matter would be meaningless.

If the studies included in a meta-analysis are a biased sample of all relevant studies, then the mean effect computed will reflect this bias (Borenstein et al. 2009, p. 277). It is clear from our thorough searching strategies that every precaution was taken to ensure that all eligible studies would be represented in the meta-analysis. In order to further increase the validity of the meta-analysis findings, a number of publication bias analyses were carried out.

Firstly, the Duval and Tweedie's Trim-and-Fill procedure was used. This technique displays the differences in effect sizes that could be attributable to bias by imputing effect sizes until the error distribution more closely approximates normality, offering the best *estimate* of the unbiased effect size (Borenstein et al. 2009, p. 286). No imputed effect sizes appeared on the relevant funnel plot (they would have been presented as solid black dots; see Farrington et al. 2012, Fig. 6), indicating no publication bias. The imputed summary effect size (represented by a solid black diamond) had not shifted at all.

Indeed, under the fixed effect model, the point estimate and 95 % confidence interval for the combined studies was 1.86 (95 % CI: 1.71–2.03). Using Trim-and-Fill procedure, these values remained unchanged. Under the random-effects model, the point estimate and 95 % confidence interval for the combined studies was 1.89 (95 % CI: 1.60–2.23). Using Trim-and-Fill procedure, these values were again unchanged.

Furthermore, Rosenthal's Fail-Safe N test (Rosenthal 1979) was conducted. One concern of publication bias is that some nonsignificant studies are missing from a given analysis and that these studies, if included, would nullify the observed effect. Rosenthal suggested that, rather than simply speculate about the impact of the missing studies, we compute the number of nonsignificant studies that would be required to nullify the effect. If this number is small, then there is reason for concern because some nonsignificant studies may have been never communicated to the scientific community (e.g., due to "publication bias"). However, if this number is large, one can be confident that the treatment effect, while possibly inflated by the exclusion of some studies, is nevertheless not zero.

Bullying Perpetration at School and Violence Later in Life

A total number of 15 studies were concerned with the association of bullying perpetration with aggression and violence later in life. The unadjusted summary effect size across these studies was OR = 3.09 (95 % CI: 2.35–4.07; z = 8.10). For one study, only an unadjusted effect size was available. The unadjusted effect size for the remaining 14 studies was OR = 2.97 (95 % CI: 2.25–3.92; z = 7.71; Q = 151.81, $p = 0.0001$; $I^2 = 91.44$). All individual studies yielded a significant effect size (see Ttofi et al. 2012, Fig. 1). After controlling for covariates, the adjusted summary effect size was reduced to OR = 2.04 (95 % CI: 1.69–2.45; z = 7.53) but this was still highly significant (see Ttofi et al. 2012, Fig. 2). This OR indicates quite a strong relationship between bullying perpetration and later violence. For example, if a quarter of children were bullies and a quarter were violent, this value of OR would correspond to 35.8 % of bullies becoming violent, compared with 21.4 % of non-bullies. Thus, being a bully increases the risk of being violent (even after controlling for other childhood risk factors) by two thirds.

Although all individual studies yielded an effect size supporting the link between school bullying and aggression/violence later in life, the magnitude and the significance of the effect varied across these studies. Various moderator analyses were conducted in order to explain this variability (Q = 75.801, $p = 0.0001$, $I^2 = 82.85$). These included the number of covariates controlled for at baseline (range: 2–20; M = 6.93; SD = 5.25), the age at which school bullying was measured (range: 8.00–15.54; M = 12.04; SD = 2.35), the age of participants when outcome measures were taken (range: 10.00–24.64; M = 17.65; SD = 4.83), and the length of the follow-up period, measured in years (range: 0.42–16.50; M = 5.61; SD = 4.88).

The age of participants when bullying was measured was significantly negatively correlated with the effect size (B = −0.065; SE = 0.021; $p = 0.002$), suggesting that the younger the children were when they exhibited this form of problem behavior, the more likely it was that they would be violent later in life. The age of participants when outcome measures were taken was also significantly negatively related to the effect size (B = −0.033; SE = 0.009; $p = 0.0005$). In other words, the lower the age of the participants when aggression or violence was measured, the larger the effect, possibly because this was associated with a shorter follow-up period. This is consistent with the significant negative association between the length of follow-up period and the magnitude of the effect size (B = −0.017; SE = 0.009; $p = 0.051$). As expected, the magnitude of the effect size decreased as the number of confounds controlled for increased (B = −0.013; Intercept = 0.668; SE = 0.010; $p = 0.185$), but the relevant regression coefficient was not significant.

As with the previous meta-analysis, a number of sensitivity analyses were conducted. Firstly, the Duval and Tweedie's Trim-and-Fill procedure was performed. Three imputed effect sizes appeared on the relevant funnel plot (see Ttofi et al. 2012, Fig. 3) and the imputed summary effect size (represented by a solid black diamond) had shifted slightly, suggesting a trivial overestimation of the summary effect size.

As already mentioned, the difference was very small. Under the fixed effect model, the point estimate and 95 % confidence interval for the combined studies was 1.83 (95 % CI: 1.71–1.95). Using Trim-and-Fill procedure, the imputed point estimate was 1.76 (95 % CI: 1.65–1.88). Under the random-effects model, the point estimate and 95 % confidence interval for the combined studies was 2.04 (95 % CI: 1.69–2.45). Using Trim-and-Fill procedure, the imputed point estimate was 1.77 (95 % CI: 1.45–2.16).

Finally, the Rosenthal's Fail-Safe N test was performed. This meta-analysis incorporated data from 14 studies, which yielded a z-value of 17.12216 and corresponding 2-tailed *p*-value of 0.000001. The fail-safe N is 1055. This means that one would need to locate and include 1055 "null" studies in order for the combined 2-tailed *p*-value to exceed 0.050. Put another way, 75.4 missing studies would be needed for every observed study for the effect to be nullified. It is impossible that such a large number of studies were conducted but not published or not included in our analysis.

Further Findings

Further analyses were performed to examine the association of bullying victimization with later offending before (Unadjusted OR = 1.32; 95 % CI: 1.13–1.55, z = 3.40) and after controlling for other major childhood risk factors (Adjusted OR = 1.14; 95 % CI: 0.997–1.310, z = 1.91) and relevant forest plots are shown in the Farrington et al. (2012) report (see Figs. 11 and 12). This was a very weak relationship. Moderator analyses and publication bias analyses similar to those presented in the current entry were also presented in that report.

Finally, analyses were performed to examine the association of bullying victimization with later violence before (Unadjusted OR = 1.65; 95 % CI: 1.42–1.92; z = 6.48) and after (Adjusted OR = 1.42; 95 % CI: 1.248–1.6172; z = 5.3117) controlling for covariates (see Ttofi et al. 2012, Figs. 4 and 5). Again, moderator analyses and publication bias analyses similar to those presented in the current entry are also presented in that report.

Sensitivity analyses were performed for these two sets of meta-analyses and the results showed in general no evidence of publication bias (see Farrington et al. 2012; Ttofi et al. 2012).

Possible Controversies in the Literature

The results of these systematic reviews and meta-analyses suggest that there are long-term detrimental effects of school bullying on later offending and violence. This was even the case when confounded variables that are risk factors for bullying or victimization as well as the undesirable outcomes were controlled for. Therefore, one can conclude that school bullying is an independent predictor of the later psychosocial development of perpetrators as well as of victims. It is the first time that this conclusion is not only based on a few selected primary studies and narrative reviews, but, instead, on comprehensive meta-analyses of prospective longitudinal studies that included new data from a substantial body of yet unpublished research. The findings remained robust in sensitivity analyses testing potential publication biases, of which there was no sign.

The relation of bullying perpetration with later offending and violence might reflect the persistence of an underlying disposition for antisocial behavior that has different manifestations over time (Farrington 1993; Lösel and Bliesener 2003). However, as the relation remained after controlling for other childhood risk factors, bullying perpetration may also increase the likelihood of later offending and violence.

Of course, one should acknowledge that any direct mention of causality should be carefully treated. Although most studies use bullying as the predictor of later outcomes, implying in this way a specific temporal sequence, alternative models have been suggested. Very few bullying studies have examined alternative models on whether bullying is a cause or a consequence of psychopathologic behavior (e.g., Boulton et al. 2010; Kim et al. 2006). This is not a trivial matter and it would shed more light on the temporal sequence and the causal ordering between bullying and other internalizing or externalizing behaviors. The substantial adjusted effect size for victimization versus later depression found in a previous meta-analysis (Farrington et al. 2012; Ttofi et al. 2011b), for example, suggests in a way that the frequent internalizing symptoms of victims are not only a trigger for being bullied, but a psychological consequence.

Systematic reviews on risk factors are important as they can advance theory and also help to develop effective prevention programs (Murray et al. 2009). For example, it would be interesting to examine whether victims of bullying suffer from low self-esteem or whether school bullies lack cognitive or affective empathy. Such findings, based on relevant systematic reviews, could guide future intervention initiatives, while also refining theory about the causes of bullying perpetration and victimization.

Open Questions and Future Research Directions

In the current meta-analysis, studies were included and analyzed based on "level analyses." Levels of bullying perpetration were compared with later levels of offending and violence. It would also have been interesting to complete a systematic review on "change analyses," examining whether changes in bullying from Time 1 to Time 2 are followed by changes in an outcome from Time 2 to Time 3. However, there are hardly any studies on this matter, since such analyses would require relevant data from multiple waves. Such analyses would allow, to an extent, making safer inferences about causality, although change data are subject to greater variability than level data. Systematic reviews of longitudinal studies which control for confounded variables can give some hints on whether variables are simple correlational risk factors, risk markers, or causal risk factors (Kraemer et al. 2005).

Future research should also examine possible gender-specific and ethnic-specific effects of bullying on later violent behavior and offending. Such information was hardly ever available in the current literature. To investigate and disentangle the impact of these and other variables on the relation between bullying and later outcomes, more longitudinal studies with a sound control for childhood risk factors are needed. The results of meta-regression analyses were not always as expected in the meta-analyses for the British Academy project on "Health and Criminal Outcomes of School Bullying" because of the large differences in the type of covariates researchers controlled for. However, one should note that the lack of a sufficient number of studies with consistent patterns of characteristics is a typical problem in meta-analyses (Lipsey 2003).

Future research should also examine mediators or possible causal mechanisms between school bullying and the various outcomes. The underlying mechanisms, for example, may be the reinforcement obtained by dominating others and the development of an identity as a "bully" that goes beyond the school context.

Conclusions

This is the first time that research has provided an *unbiased standardized effect size* regarding the predictive efficiency of school bullying in relation to violence and offending later in life. The significant summary effect sizes have important implications for policy and practice as they give a stronger voice to anti-bullying agencies and reestablish the moral imperative of school communities to create an appropriate violence-free school climate.

High quality bullying prevention programs should be promoted (Farrington and Ttofi 2009; Ttofi and Farrington 2011). They could be viewed as an early form of crime prevention. These programs can potentially have long-term effects by improving the future psychosocial adjustment of school bullies and reducing the associated health, welfare, education, and other costs. The effectiveness of other school-based programs for the prevention of problem behaviors has been examined through thorough systematic reviews (e.g., Wilson et al. 2001) and it is possible that such programs, or other general multicomponent programs, might have positive effects in reducing aggression and bullying behavior.

Previous research has provided strong evidence about the monetary value of saving high-risk youth (Cohen and Piquero 2009). Children involved in school bullying are undoubtedly youth at risk, with significantly higher probabilities of following an antisocial path. What remains unanswered is the identification of protective factors that interrupt the continuity from school bullying to later adverse outcomes and confer resiliency on this special category of high-risk youth (Ttofi and Farrington 2012).

Glossary

CI Confidence interval
M Mean
OR Odds ratio
SD Standard deviation

Related Entries

- ▶ Juvenile Violence
- ▶ Risk Assessment, Classification, and Prediction
- ▶ Risk factors for Adolescent Sexual Offending
- ▶ School Crime Statistics
- ▶ School Social Organization, Discipline Management, and Crime
- ▶ School Structural Characteristics and Crime
- ▶ School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis

Recommended Reading and References

Borenstein M, Hedges LV, Julian PTH, Rothstein HR (2009) Introduction to meta-analysis. Willey, Chichester

Boulton MJ, Smith PK, Cowie H (2010) Short-term longitudinal relationships between children's peer victimization/bullying experiences and self-perceptions: evidence for reciprocity. School Psychol Int 31(3):296–311

Cohen MA, Piquero AR (2009) New evidence on the monetary value of saving a high risk youth. J Quant Criminol 25:25–49

Farrington DP (1993) Understanding and preventing bullying. In: Tonry M (ed) Crime and justice, vol 17. University of Chicago Press, Chicago

Farrington DP, Ttofi MM (2009) School-based programs to reduce bullying and victimization. Campbell Systematic Reviews 6

Farrington DP, Ttofi MM, Lösel F (2011) Editorial: school bullying and later offending. Crim Behav Mental Health 21(2):77–79

Farrington DP, Lösel F, Ttofi MM, Theodorakis N (2012) School bullying, depression and offending behaviour later in life: an updated systematic review of longitudinal studies. Swedish National Council for Crime Prevention, Stockholm

Kim YS, Leventhal BL, Koh Y-J, Hubbard A, Boyce TW (2006) School bullying and youth violence: causes or consequences of psychopathologic behavior? Arch Gen Psychiat 63:1035–1041

Kraemer HC, Lowe KK, Kupfer DJ (2005) To your health: how to understand what research tells us about risks. Oxford University Press, New York

Lipsey MW (2003) Those confounded moderators in meta-analysis: good, bad, and ugly. Ann Am Acad Polit Soc Sci 587(1):69–81

Lösel F, Bliesener T (2003) Aggression und delinquenz unter jugendlichen: Untersuchungen von kognitiven und sozialen Bedingunger. Luchterhand, Hermann

Murray J, Farrington DP, Eisner MP (2009) Drawing conclusions about causes from systematic reviews of risk factors: the Cambridge quality checklists. J Exp Criminol 5(1):1–23

Olweus D (1993) Bullying at school: what we know and what we can do. Blackwell, Oxford

Rosenthal R (1979) The "File drawer problem" and tolerance for null results. Psychological Bulletin 86:638–641.

Smith PK, Brain P (2000) Bullying in schools: lessons from two decates of research. Aggressive Behavior 26:1–9

Ttofi MM, Farrington DP (2011) Effectiveness of school-based programs to reduce bullying: a systematic and meta-analytic review. J Exp Criminol 7(1):27–56

Ttofi MM, Farrington DP (2012) Risk and protective factors, longitudinal research and bullying prevention. New Direct Youth Develop 133:85–98

Ttofi MM, Farrington DP, Lösel F (2011a) Editorial: health consequences of school bullying. J Aggress Conflict Peace Res 3(2):60–62

Ttofi MM, Farrington DP, Lösel F, Loeber R (2011b) Do the victims of school bullies tend to become depressed later in life? A systematic review and meta-analysis of longitudinal studies. J Aggress Conflict Peace Res 3(2):63–73

Ttofi MM, Farrington DP, Lösel F, Loeber R (2011c) The predictive efficiency of school bullying versus later offending: a systematic/meta-analytic review of longitudinal studies. Crim Behav Mental Health 21(2):80–89

Ttofi MM, Farrington DP, Lösel F (2012) School bullying as a risk marker for aggression and violence later in life: a systematic review and meta-analysis of prospective longitudinal studies. Aggression and Violent Behavior (Accepted, in press)

Wilson D (2010) Email communication, October 25

Wilson DB, Gottfredson DC, Najaka SS (2001) School-based prevention of problem behaviors: a meta-analysis. J Quant Criminol 17(3):247–272

School Crime

► School Security Practices and Crime

School Crime Statistics

Philip J. Cook[1,2], Denise C. Gottfredson[3] and Chongmin Na[4]
[1]Sanford School of Public Policy, Duke University, Durham, NC, USA
[2]NBER, Cambridge, MA, USA
[3]Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA
[4]School of Human Science and Humanities, University of Houston-Clear Lake, Houston, TX, USA

Overview

By rights, schools should be sanctuaries against criminal victimization, but the truth is otherwise. A threatening environment is not conducive to academic success. The US federal law implementing No Child Left Behind (NCLB, the most recent national education-reform initiative in the USA) stipulates that school systems must have programs in place to reduce levels of violence. The legislation authorizing NCLB has a specific provision that "persistently dangerous" schools be identified by the states and that students attending such schools be given the option of transferring to another school. In practice, only a few dozen schools nationwide have been identified as persistently dangerous, but that fact reflects politics more than the empirical reality.

As it turns out, obtaining reliable information about crime in schools is a challenge – for researchers as well as for state and federal officials. There are several sources of data in addition to the schools' own reports, but each source is error prone. There are some rather remarkable differences among them with respect to estimated crime rates and patterns.

Based on our review of the evidence, we offer several lessons about crime in school and about the sources of statistics on crime. First, with respect to sources, we find that crime surveys with adolescent respondents give results that are remarkably sensitive to the survey method. For example, in-school victimization rates from the Youth Behavior Risk Surveillance System (YRBSS) are more than 20 times as high as comparable victimization rates estimated from the National Crime Victimization Survey (NCVS), and they produce qualitatively different patterns over time. It is always a good idea to validate estimates from one source against others, but that may be especially true with youth surveys. On the other hand, we find administrative data to be better than expected.

We are able to reach some conclusions about crime patterns that we feel are sufficiently robust to be believable. Among our conclusions are these:

- Victimization rates for 12–18-year-olds are as high in school as out (and higher for the 12–14-year-olds), despite the fact that youths spend many fewer waking hours in school.
- Victimization patterns are quite different in and out of school. Violent crimes tend to be

less serious in school (where only 1 % of homicide victimizations occur).

- Property crime victimization rates are remarkably homogeneous across race and also sex; alternative data sources are not consistent with respect to racial patterns of school violence.
- "Persistently dangerous schools" appear to be a real phenomenon and far more prevalent than are officially identified as such.

We begin this entry with a discussion of the important data sources and then go on to discuss victimization patterns for students. We then explore the issue of persistently dangerous schools using a unique data set. We conclude with an analysis of arrest patterns.

Data Sources

The primary source of the US crime data for many purposes is the FBI's Uniform Crime Reports, compiled from crimes known to the police and reported by police departments. The UCR's crime data do not provide information on the characteristics of victims and are of little help in estimating crime rates in schools. Some jurisdictions report crimes in much more detail through the National Incident-Based Reporting System (NIBRS): in this system police agencies submit a record of each known crime that includes the age, sex, and race of the victim, the location of the crime, and the characteristics of the perpetrator (when known). These data can be used to provide a detailed description of crimes involving school-aged youths, distinguishing, for example, between crimes on school grounds and elsewhere (Jacob and Lefgren 2003). There are two problems, however, with this source. First, participation rates are very low: only 20 % of police agencies, representing 16 % of the US population, were participating in NIBRS as of 2003 (http://www.ojp.usdoj.gov/bjs/nibrsstatus.htm, accessed November 19, 2008). And second, crimes committed on school property may be less likely to become known to the police than crimes occurring elsewhere.

As a result of the limitations of police data, school crime statistics are usually generated from school reports or surveys. In the School Survey on Crime and Safety (SSoCS), a sample of about 3,000 public-school principals report to the US Department of Education the number of violent incidents and thefts and indicate how many of these incidents were reported to the police. In addition there are several recurrent sample surveys: the National Crime Victimization Survey and the biannual School Crime Supplement to this Survey (sponsored by the National Center for Education Statistics, or NCES) and the Youth Risk Behavior Surveillance System (YRBSS), sponsored by the Centers for Disease Control and Prevention (CDC). The NCES compiles data from all these sources into a report called the *Indicators of School Crime and Safety* (e.g., Dinkes et al. 2007). When the estimates from these alternative sources are compared, there emerge some rather dramatic differences, leaving the investigator with the challenge of deciding where the truth lies.

Youthful Victimization in School and Out

Here we report crime victimization rates for school-aged youths, comparing, when possible, the rates at school and at other locations. We begin with murder, which is the only crime for which the statistics are reasonably accurate. Figure 1 depicts the trend in murders on school property for youths ages 5–18, compared with the overall murder count for that age group. There were about 30 school murders of youths each year from 1992–1993 to 1998–1999, a period notorious for the series of school rampage shootings that culminated with Columbine High School on April 20, 1999. During that event 12 students and a teacher were murdered and 23 students injured, before the shooters committed suicide. In the year following Columbine, the national in-school murder count dropped sharply and has remained relatively low since then. The overall murder rate for the same age group follows a similar pattern, though the decline began earlier and is less abrupt. The most important lesson from these data is that only about 1 in 100 murders of this age group occurs in school. That was true during the peak years of the early 1990s and also true a decade later. By this measure, then, school

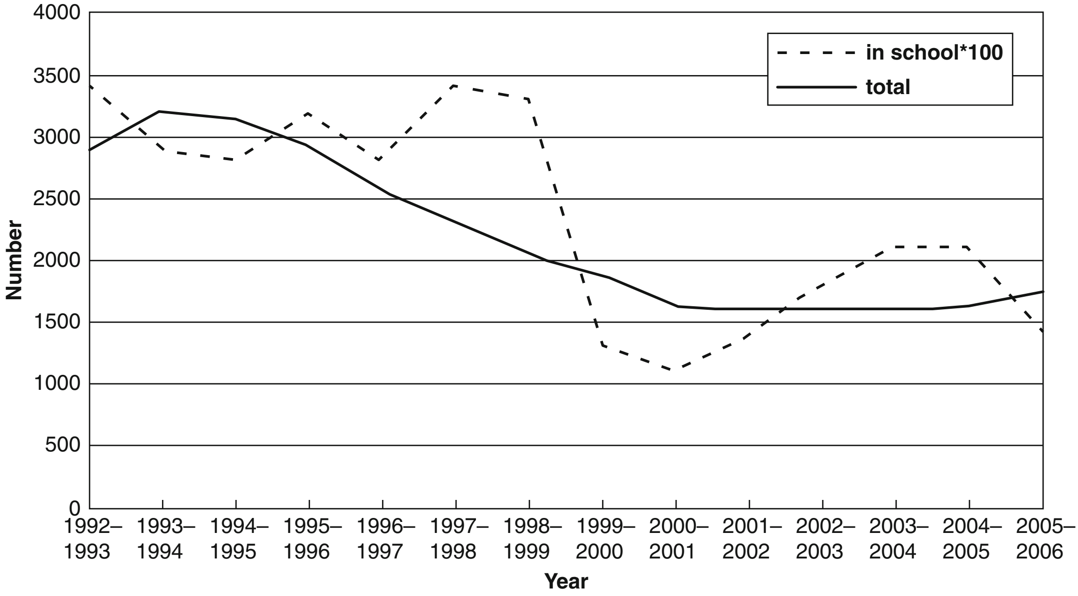

School Crime Statistics, Fig. 1 Number of homicides involving young victims, in school and out, 1992–1993 to 2005–2006. Note: "In school" includes on school property, on the way to or from regular sessions at school, and while attending or traveling to or from a school-sponsored event (Source: Data on number of homicides in school are from School-Associated Violent Deaths Surveillance Study (SAVD), tabulated in Indicators of School Crime and Safety (Dinkes et al. 2007, p. 68); and data on number of homicides total are from National Center for Injury Prevention and Control and Web-based Injury Statistics Query and Reporting System Fatal (WISQARS^{TM} Fatal), retrieved Nov 2008 from http://cdc.gov/ncipc/wisqars)

appears much safer than other locations for school-aged youths.

However, schools have a much larger share of the nonfatal crimes with school-aged victims. Figure 2 depicts the trend for victimization rates of youths aged 12–18, including both theft and violence. Youths who have completed 12 years of school are excluded from this tabulation. The rates per 1,000 follow the trend for youth homicide (as well as the national trend for criminal victimization for all age groups) – a sustained and rather dramatic reduction, so that the 2005 figures are about one-third of the peak in 1993. For our immediate purpose here, the important thing to notice is that the victimization rate in school is about the same as out of school. That parity is the net result of theft, which has higher rates at school, and violence, which for most of the period has lower rates at school (although in-school and out-of-school rates of violence converged in 2004). Note that since youths spend over 80 % of their waking hours during a calendar year out of school (Gottfredson 2001, p. 21), the parity in victimization rates implies that youths are far more likely to be victimized during an hour in school than an hour elsewhere.

For the serious violent crimes of rape, robbery, and aggravated assault, NCVS victimization rates are twice as high away from school as at school during recent years, as shown in Panel B. Since the corresponding ratio for murder is 100 to 1, we conclude that serious violent crimes committed out of school are far more likely to become murders than is true for similar crimes in school.

These NCVS results may well be misleading and should be validated. Youth survey data on crime are notoriously unreliable. In particular, crime survey results are exquisitely sensitive to the details of how the data are collected. One survey that provides an alternative to NCVS for estimating victimization rates is the Youth Risk Behavior Surveillance System (YRBSS), sponsored by the CDC. This survey yields estimates of victimization rates for serious violent crime

School Crime Statistics, Fig. 2 *Panel A*: Victimization rates at school and out for youths ages 12–18, 1992–2005: theft and violence. *Panel B*: Victimization rates at school and out for youths ages 12–18, 1992–2005, serious crimes of violence. Note: Theft includes purse snatching, pickpocketing, and all attempted and completed thefts except motor vehicle thefts. Theft does not include robbery in which threat or use of force is involved. Violence includes serious crimes of violence and simple assault. Serious crimes of violence include rape, sexual assault, robbery, and aggravated assault. "At school" includes inside the school building, on school property, or on the way to or from school (Source: National Crime Victimization Survey (NCVS), tabulated in Indicators of School Crime and Safety (Dinkes et al. 2007, p. 70))

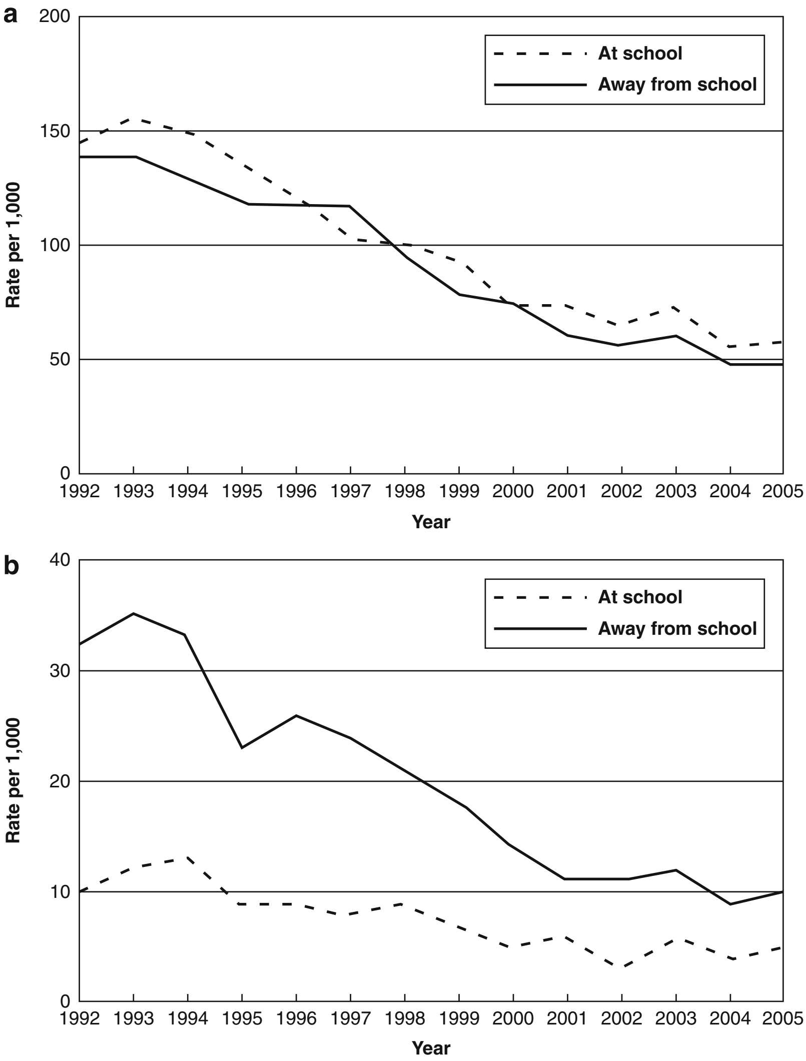

that are an order-of-magnitude higher than the NCVS rates. For example, in the 2005 YRBSS, 8 % of students in grades 9–12 reported being threatened or injured with a weapon on school property during the previous 12 months. That compares with the serious-violence victimization rate at school for 15–18-year-olds in the NCVS of 0.4 %. Thus, the YRBSS rate is 20 times as high, even though logic suggests that it should be *less*, given that the YRBSS refers to prevalence of victimization and the NCVS figure is overall incidence (so that multiple victimizations reported by the same respondent are included in computing the rate). Further, the NCVS category of "serious violence" encompasses more types of crime than the YRBSS category of "threatened or injured with a weapon."

What could account for this vast difference in results? First, the NCVS sample is interviewed every 6 months, and the previous interview serves as a bracket to help the respondent place events in time. Thus, the NCVS sample members are asked to report on events that occurred since the previous interview. The YRBSS, on the other hand, is a one-shot survey with no natural bracket on the time interval; respondents are asked to report on the previous 12 months, which creates the likelihood that some will report on serious events that occurred outside the designated period (a phenomenon known as "telescoping"). A second important difference is that all YRBSS respondents are asked the specific question about whether they were threatened or injured with a weapon on school property, whereas the only NCVS respondents who are asked about such an incident are those who first respond affirmatively to a more general screener question. Third, the NCVS questionnaire is administered to the respondent (in person or over the telephone) at home, whereas the YRBSS questionnaire is self-administered by the respondent while in school. These and other differences, none of which are relevant in a literal sense, appear to be hugely important to the respondents' answers in practice. Cook (1987) notes that the Safe Schools study (which had much in common with the YRBSS with respect to administration) estimated one million robberies in schools, compared with the estimate of 30,000 in the NCVS for the same period.

Given the disparate results from youth surveys, it is of interest to consider administrative data. The SSoCS gathers reports from public-school principals about crimes occurring during school hours. For the 2005–2006 school year, principals for middle and high schools reported a total of 928,000 violent crimes and 206,000 thefts (see Table 1). These counts are not precisely comparable to the NCVS results for 12–18-year-olds. Unlike the NCVS, the SSoCS is limited to public schools. The NCVS age range of 12–18 is roughly but not exactly comparable to the SSoCS category of "middle and high school." Nevertheless, the estimates should be close. In fact, the violence reports are higher (by half) in the SSoCS than in the NCVS for 12–18-year-olds, while the SSoCS theft reports are much lower. It is not surprising that school officials do not know about many of the thefts that occur on school property, but the fact that they are aware of more violence that shows up in the NCVS defies ready explanation.

Thus, the truth about crime in school – or even a rough approximation of the truth – is elusive. Our inclination is to believe that the SSoCS reports provide a reliable lower bound for the "true" volume of crime, understating the true total to the extent that officials are never made aware of some crimes, and may generally be inclined to underreport in order to make their schools look as safe as possible. If true, then the NCVS appears to provide a notable underestimate of the volume of violence in schools – but the difference is nothing like that suggested by the very high YRBSS results. We are inclined to believe that the NCVS data are superior to the YRBSS because the method of administration discourages exaggeration by respondents, and the bracketing provides some discipline on memory. We also note that the downward trend in NCVS rates (shown in Fig. 2) reproduces well-documented trends during that period for the entire US population and hence is credible. The YRBSS victimization rates, on the other hand, exhibit no such trend during this period, showing if anything an upward tilt since 1993. For those reasons we report additional NCVS results in what follows, even though we are willing to believe that these are also far off the mark.

School Crime Statistics, Table 1 Comparison of SSoCS and NCVS crime counts

	SSoCS crime count: public middle and high schools 2005–2006 school year	NCVS crime count in school ages 12–18 2005
Violent crimes	928,000	628,000
Theft	206,000	868,000

Sources: http://nces.ed.gov/pubs2007/2007361.pdf; http://nces.ed.gov/programs/crimeindicators/crimeindicators2007/tables/table_02_1.asp?referrer=report

Table 2 summarizes demographic patterns in victimization rates at school for youths aged 12–18. Note that these data exclude the responses of students who have already completed 12 years of schooling. They do not exclude school dropouts. The rates shown here are averaged over the three most recent years of the School Supplement of the NCVS. Note that "theft" and "violence" sum to the total – "serious violence" is included in "violence."

Theft rates are remarkably uniform across all demographic categories, averaging 41/1,000.

School Crime Statistics, Table 2 At-school victimization rates/1,000 for youths age 12–18

	Total	Theft	Violence	Serious violence
Male	73	41	32	7
Female	61	41	21	4
Ages 12–14	75	42	33	6
Ages 15–18	61	40	21	5
Urban	75	41	34	9
Suburban	67	43	24	5
Rural	58	37	22	2
White	72	45	27	4
Black	64	38	27	5
Hispanic	55	29	26	5
Other	53	33	20	2
Overall	67	41	27	6

Source: NCVS results averaged for 2001, 2003, and 2005

Violence rates are a bit lower overall and more textured, although the differences among groups are still not as large as one might expect. Males are half again as likely to be victims of violence as females, and youths 12–14 are half again as likely as older youths. Urban schools experience a higher per capita rate of violent incidents than suburban or rural schools. Most surprising is that whites, blacks, and Hispanics report virtually the same rates of violence and serious violence.

The same NCVS data provide estimates for victimization rates away from school. The patterns are not much different, with two exceptions. First, blacks report a higher rate of serious violent crimes (17/1,000) than whites and Hispanics (both at 10/1,000). Second, and perhaps most intriguing, is that the age pattern away from school is the reverse of the age pattern at school. The younger group, aged 12–14, has somewhat higher victimization rates at school than the older group, but the older group has much higher victimization rates than the younger group away from school. The results are depicted in Fig. 3. The explanation may in part be due to the fact that the older group includes a number of school dropouts who, since they are not attending school, are unlikely to be victimized on school property. Perhaps more important is that older youths have greater mobility and freedom outside of school and thus more opportunity to get into trouble.

In assessing the credibility of these results, we are inclined to believe the NCVS-based comparisons between in-school and out-of-school victimizations, since it is the same respondents and hence the same biases are likely to apply. The surprising homogeneity of theft and violence victimization between blacks, whites, and Hispanics can be checked against interschool patterns in the SSoCS. We do that below and find support for the theft results but an apparent contradiction with the results on violence.

Finally, we note the high prevalence of bullying in school. While not necessarily a crime, bullying can greatly color the school experience for some children. The NCVS School Crime Supplement found that in 2005, 28 % of youths ages 12–18 reported being bullied in school – of those, 79 % said they were bullied inside school, 28 % outside on school grounds, and 8 % on the school bus (Dinkes et al. 2007, p. 95).

Differences Among Schools

Up until this point, we have described crime patterns primarily with respect to the characteristics of the victims. From another perspective, school crime is a characteristic of the school, and there is strong evidence that school characteristics and policies influence crime victimization rates (Cook et al. 2010).

The 2005–2006 SSoCS classifies schools by grade level, enrollment size, urbanicity, and percent minority enrollment. The rate of violent incidents reported by principals is much higher for middle schools than either elementary or high schools, somewhat higher for city schools than those in suburban or rural communities, and higher in predominantly minority schools than those with less than half minority. Notably, there is little relationship between the size of the school and the violence victimization rate. The results for theft tend to be less patterned. Table 3 summarizes the results.

In one respect these patterns are at odds with NCVS victimization patterns. It appears that the relatively high rate of violence in minority schools is at odds with the NCVS finding that there is little difference in victimization rates by race. One possibility is that black respondents are

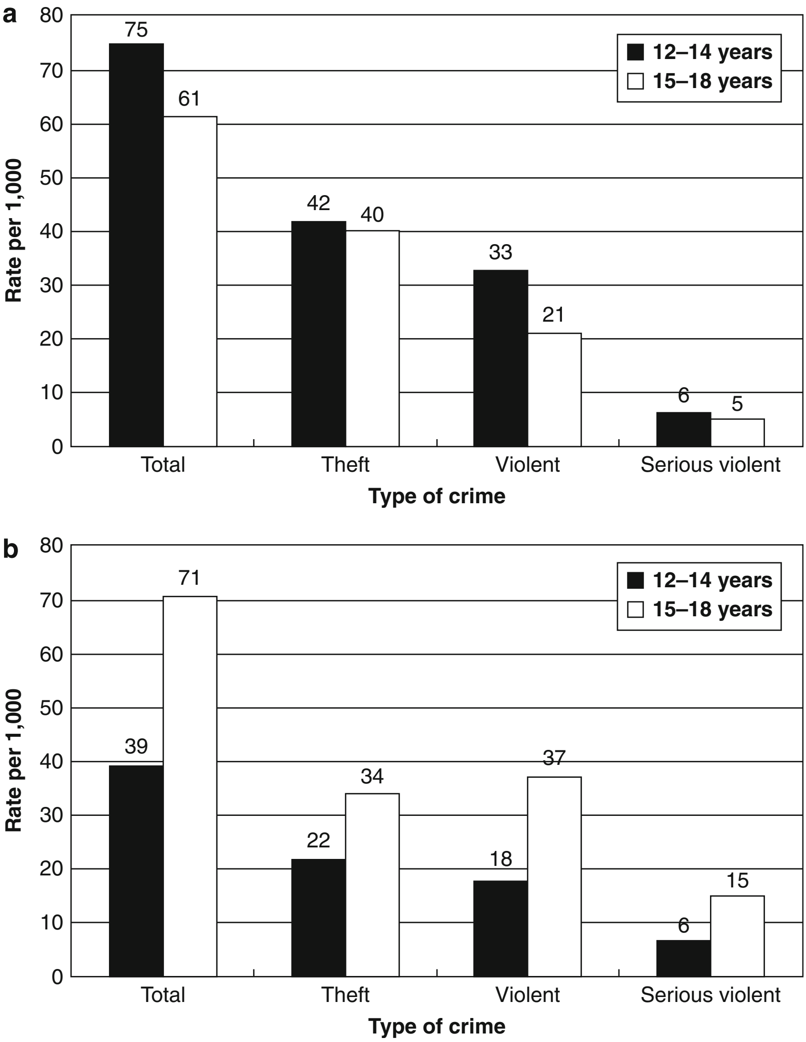

School Crime Statistics, Fig. 3 *Panel A*: Victimization rates at school for youths ages 12–14 and 15–18. *Panel B*: Victimization rates away from school for youths 12–14 and 15–18. Note: Total crimes include theft and violent crimes. Theft includes purse snatching, pickpocketing, and all attempted and completed thefts except motor vehicle thefts. Theft does not include robbery in which threat or use of force is involved. Violent crimes include serious violent crimes and simple assault. Serious violent crimes include rape, sexual assault, robbery, and aggravated assault. "At school" includes inside the school building, on school property, or on the way to or from school. NCVS results are averaged for 2001, 2003, and 2005 (Source: National Crime Victimization Survey (NCVS), tabulated in Indicators of School Crime and Safety (DeVoe et al. 2003, pp. 55–66; DeVoe et al. 2005, pp. 72–73; Dinkes et al. 2007, pp. 70–71))

less likely to report violent victimizations than white respondents in the NCVS or alternatively (as reported in Kinsler 2009) that school administrators in predominantly black schools are relatively likely to record violent crimes. The same source, SSoCS, reports information on gang-related crime. In 2005–2006, 11 % of middle schools and 16 % of high schools reported at least one crime that was gang related. Gang-related crimes were concentrated in large, urban, and predominantly minority schools (Nolle et al. 2007, Table 4). Some confirmation for these patterns comes from the NCVS School Supplement data. Students were asked about gangs in their schools. Affirmative responses were much more likely by black and Hispanic students and by students in urban areas.

Another way to illustrate the variability across schools in crime levels is by comparing the distribution of the number of crimes reported per school with a distribution that would be generated under a random process. This type of exercise is common in criminological research. In a study of more than 23,000 boys born in two birth cohorts in Philadelphia, Tracy et al. (1990), for example, discovered that a small fraction of the boys (5–6 %) committed a majority of the delinquent acts. Similarly, researchers who study neighborhood crime have observed that it, too, is clustered at certain addresses. Weisburd et al. (2004),

School Crime Statistics, Table 3 Crime rates by school characteristic

	Violence rate/1,000 students	Theft rate/1,000 students
Level		
Primary	25.2	1.6
Middle	51.6	7.8
High school	25.7	8.7
Enrollment		
<300	34.5	4.3
300–499	34.0	3.3
500–999	30.9	4.5
1,000 or more	28.6	7.2
% minority enrollment		
<5 %	26.9	4.8
5–20	22.9	5.2
20–50	28.4	5.5
50 or more	39.9	4.8

Source: Nolle et al. (2007), extracted from Table 1

School Crime Statistics, Table 4 Number of "dangerous" and "persistently dangerous" schools, 2004, 2006, and 2008

	Dangerous schools	Persistently dangerous schools[a]
2003–2004	1,547 (1.9 %)	449
2005–2006	1,542 (1.9 %)	447
2007–2008	1,624 (2.0 %)	471

Source: Original tabulations from data files from the School Survey on Crime and Safety

[a] Assumes 29 % of schools dangerous in 1 year are "persistently dangerous" (see text)

studying street segments in Seattle over a 14-year period, found that 4–5 % of the street segments account for about 50 % of incidents in each of the years examined. The distribution of crime at the school level is also concentrated, although not to the same degree.

We examined the distribution of school crime in the 2007–2008 SSoCS data using schools with enrollments of at least 1,000 students (N = 8,843), because we assumed rank orders based on crime rates would be unstable for small schools. These analyses used sample weights provided by NCES and also weighted by student enrollment. In these data, the 22.5 % of large schools with the highest *per capita* crime rates account for 50 % of all crimes reported by principals. The top 20 % of schools experience much higher crime rates than does the bottom 20 %: the victimization rates per 1,000 students for students in the most and least safe 20 % of schools are 9.8 and 111.3 for all crimes. The difference is even more striking for rates of violent crime – more than 13 times higher in the top 20 % of schools (4.9 vs. 65.7 per 1,000). These results suggest that although the average school is indeed a relatively safe place, a minority are notoriously unsafe. Figure 4 contrasts the safest 20 % of schools with the most dangerous 20 %. The more dangerous schools are much more likely to be urban schools, schools serving a high percentage of minority students, and schools serving more disadvantaged student populations.

The NCLB Act requires states to establish a mechanism for identifying such unsafe schools. Once a school has been labeled through this process as "persistently dangerous," parents must be given the option of transferring their child to a safer school. Every state has responded to this requirement by defining "persistently dangerous" and establishing a procedure for identifying such schools. The definitions differ considerably from state to state according to the specific offenses considered to be indicators of dangerousness; whether or not an official response (e.g., expulsion, arrest, conviction) to the incident is required before it is considered a dangerous offense and whether or not a dangerous offense occurring in or around the school has to have been committed by a student; the cut-point for the number or rate of offenses above which a school is to be considered dangerous; and the number of years in the dangerous status required to be considered "persistently" dangerous. Also, some states use a multistage process for identifying unsafe schools. For example, Florida's policy calls for a first stage in which a school is considered potentially dangerous if, for 3 consecutive years, it (a) has a federal Gun-Free School Act violation and (b) expels 1 % or more of a student body that is greater than 500 students, or 5

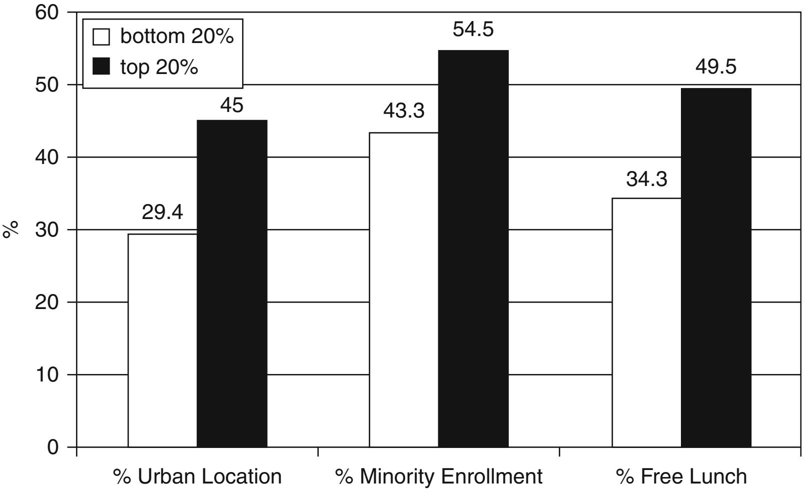

School Crime Statistics, Fig. 4 Comparison of the characteristics of the schools with the lowest and the highest 20 % of crime rate. Based on schools with enrollment of at least 1,000 students. Weighted by student enrollment and sample weight (Source: Original tabulation from data files. School Survey on Crime and Safety (SSoCS) (2007–2008))

students if the student body is 500 students or less, for homicide, battery, sexual battery, or weapons possession-related offenses. Any school so identified must, in a second stage, conduct an anonymous school-wide survey of students, parents, and school personnel. Only schools in which 51 % or more of survey respondents judge the school to be unsafe are labeled persistently dangerous (Education Commission of the States 2004).

Beginning with the 2003–2004 school year, the number of persistently dangerous schools reported by all states combined to the US Department of Education has ranged from 36 to 49, averaging 44. That low count raises the question of just how effective is the NCLB legislation for encouraging states to identify unsafe schools. We used the SSoCS survey data to identify schools that might reasonably be considered persistently dangerous. As noted above, there is no standard definition to guide this effort. Almost all states, however, include incidents involving serious violent behavior and weapons-related offenses, and many states also include incidents involving illegal drug use, possession, and distribution. Almost all states require that the rate of incidents per 100 enrolled students be between 1 and 3 before a school will be considered dangerous. Most states require that schools remain in this status for 3 years, but some require only two. We therefore classified schools in the SSoCS sample as dangerous during the 2007–2008 school year if, according to their principals' reports, the number of serious violent offenses (e.g., rape, sexual battery other than rape, physical attack or fight with a weapon, threat of physical attack with a weapon, and robbery with or without a weapon) plus the number of weapons-related incidents (those involving firearms, explosive devises, knifes, or sharp objects) exceeded 2 per 100 enrolled students. We repeated this analysis adding the number of incidents involving illegal drugs (distribution, use, or possession) to the classification of dangerous. Doing so increases the number of schools identified as dangerous. Our conclusion about the under-identification of unsafe schools using the NCLB procedure would therefore be stronger if these drug offenses were included.

Although the SSoCS survey is cross-sectional by design, it is administered to a random sample of US school every 2 years. The sample design is stratified and over-samples middle and high schools. Some schools happen to be included in multiple years just by chance. By merging the 2007–2008 survey data with data from the two prior SSoCS surveys (2003–2004 and 2005–2006), we created a longitudinal sample containing 475 schools. This longitudinal sample overrepresents secondary schools, large schools, and schools in areas that are not located in rural areas but is nevertheless helpful for examining the extent to which dangerousness persists across time.

Table 4 shows that applying our definition of dangerousness to the entire sample of schools in each of the 3 years resulted in identifying between 1.9 % and 2.0 % of the nation's schools in each survey year. The associated point estimates for the numbers of dangerous schools identified were 1,547 (2003–2004), 1,542 (2005–2006), and 1,624 (2007–2008). Using the longitudinal sample, we determined that 29 % of schools identified as dangerous in 1 year were also identified as such 2 years earlier. This exercise suggests that between 447 and 471 schools might be considered persistently dangerous in any given year. That estimate is about ten times as high as the average number of persistently dangerous schools identified to the US Department of Education each year, just 44.

Concluding Thoughts

There are a variety of sources of statistics on crime in schools, which are not entirely consistent with respect to levels, patterns, and trends. Anyone wishing to make sense of the available statistics should first become informed on the details of how the data are generated and consider the likely biases and as much as possible compare different sources.

We believe that the homicide statistics are accurate but that other police data on school crime are not to be trusted. For nonfatal crimes, we place some credence in the NCVS for students, which is a recurrent survey implemented by the US Census Bureau. What one learns from this source is that crime victimization in schools for students followed the downward trend in national crime rates during the 1990s and remains at a relatively low level since 2000. That there would be a common trend that makes sense and is one illustration of a more general result that crime in schools is closely linked to crime in the community.

Another credible result is that there is a great deal of crime in schools perpetrated by and against students. However, the estimated rates and patterns differ widely among the various sources in common use. According to the NCVS, in-school victimization rates are similar to rates experienced outside of school. But we found order-of-magnitude differences in school crime rates from NCVS and YRBSS, CDC's recurrent youth survey in which youths self-report their own victimizations. The SSoCS, based on reports from school principals, also produces counts of violent crimes that are 50 % higher than the NCVS counts. We would nominate SSoCS as the most reliable source for violent crime, but there is no basis for comparison with out-of-school crimes. We can be confident that homicide is very rare in school (relatively and absolutely). All sources confirm that a much higher percentage of minor assaults occur in schools than serious assaults. It also seems reasonable to conclude that school crime is far from randomly distributed across schools. A relatively small number of schools experience inordinately high rates of theft and violence. Rates of crime in the least safe 20 % of schools are more than ten times higher than the rates of crime in the safest 20 %. Students in these schools are likely to avoid school for fear of their own safety, teacher turnover is likely to be high, and the quality of time dedicated to learning is likely to be diminished. Unfortunately, no reliable mechanism has been established to identify which of these schools should be identified as "persistently dangerous." The consequences for being labeled "persistently dangerous" under NCLB are sufficiently punitive to create a strong incentive for principals to underreport and for state boards of education to establish policies that minimize application of that label.

What should be done to advance knowledge about school crime and to provide a reliable mechanism for identifying schools requiring assistance? The place to begin the research agenda on crime in schools is with a close look at the quality of the data in current use. Despite the many problems we identified earlier, it is too often true that users do not investigate the quality of the data or check one source against another. It would be a useful service to all users if there were a comprehensive investigation of the differences in crime rates and patterns across these data sets, together with an investigation of the sources of disagreement. One thing that is clear is that survey results with adolescent subjects are exquisitely sensitive to where and how the questions are administered.

Beyond acquiring a better understanding of the sources of discrepancy across different methods for obtaining school crime data, a serious effort is needed to develop shared understandings of what constitutes a dangerous school. Rather than asking each state to develop its own idiosyncratic definition and identification procedure, we suggest that a higher level dialogue is needed to understand the cut-point above which the school environment is trulydangerous to students and staff or seriously hinders learning. A discussion about how school safety might be monitored more systematically is needed. We suggest that rather than counting the number of certain incidents or responses to certain offenses, which are likely to be extremely rare and confounded with characteristics of school's surveillance and reporting practices, a more direct measure of school safety might come from student and staff surveys of perceptions of safety and the extent to which they avoid the school or certain areas in the school for fear of their own safety. Such survey measures have been a staple of school climate research since the 1970s and have proven to be highly reliable and stable measures of the school environment. At least one state (Florida) already uses such a strategy to identify persistently dangerous schools.

We also suggest that once a reliable mechanism is in place for identifying schools with problematic crime levels, efforts should be directed towards making the schools safer. Rigorous research (Gottfredson et al. 2002; Hahn et al. 2007; Wilson and Lipsey 2007) has identified a variety of school-based prevention and intervention practices that are likely to reduce the level of crime experienced in these schools.

Notes and Acknowledgments This entry is based on a more detailed chapter (Cook et al. 2010) on school crime control and prevention.

Related Entries

- ▶ Crime
- ▶ School Social Organization, Discipline Management, and Crime
- ▶ School Structural Characteristics and Crime
- ▶ School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis

References

Centers for Disease Control and Prevention (2008) Youth Risk Behavior Surveillance – United States, 2007 MMWR

Cook PJ (1987) Robbery violence. J Crim Law Criminology 70:357–376

Cook PJ, Gottfredson DC, Na C (2010) School crime control and prevention. In: Michael T (ed) Crime and justice: a review of research, vol. 39. University of Chicago Press, Chicago, Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1368292

DeVoe JF, Peter K, Kaufman P, Ruddy SA, Miller AK, Planty M, Snyder TD, Rand MR (2003) Indicators of school crime and safety: 2003. US Dept. of Justice, Office of Justice Programs, Bureau of Justice Statistics NCJ 210257.

DeVoe JF, Peter K, Noonan M, Snyder TD, Baum K (2005) Indicators of school crime and safety: 2005. US Dept. of Justice, Office of Justice Programs, Bureau of Justice Statistics NCJ 210697.

Dinkes R, Cataldi EF, Lin-Kelly W (2007) Indicators of school crime and safety (NCES 2008-021/NCJ 219553). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice, Washington, DC

Education Commission of the States (2004) Persistently dangerous school criteria, Updated September 2004 http://www.ecs.org/clearinghouse/52/98/5298.pdf (downloaded 25 Sept. 2009)

Gottfredson DC (2001) Schools and delinquency. Cambridge University Press, New York

Gottfredson DC, Wilson DB, Najaka SS (2002) School-based crime prevention. In: Sherman LW, Farrington DP, Brandon WC, MacKenzie DL (eds) Evidence-based crime prevention. Routledge, London

Guerino P, Hurwitz MD, Noonan ME, Kaffenberger SM (2006) Crime, violence, discipline, and safety in U.S. Public schools: findings from the school survey on crime and safety: 2003–04 (NCES 2007-302rev). U.S. Department of Education, National Center for Education Statistics, Washington, DC

Hahn RA, Fuqua-Whitley DS, Lowy J et al (2007) A review of the effectiveness of universal

school-based programs for the prevention of violence. Am J Prev Med 33:S114–S129
Hernandez N (2007) No child' data on violence skewed: each state defines 'dangerous school' (November 18, p. A01). Washington Post. 77, 67–74
Jacob BA, Lefgren L (2003) Are idle hands the devil's workshop? Incapacitation, concentration, and juvenile crime. Am Econ Rev 93:1560–1577
Jekielek S, Brown B, Marin P, Lippman L (2007) Public school practices for violence prevention and reduction: 2003–04 (NCES-2007-010). U.S. Department of Education, National Center for Education Statistics, Washington, DC
Kinsler J (2009) School discipline: a source or salve for the racial achievement gap? Unpublished manuscript, University of Rochester, New York
Nolle KL, Guerino P, Dinkes R (2007) Crime, violence, discipline, and safety in US public schools: findings from the school survey on crime and safety: 2005-6 (NCES 2007-361). U.S. Department of Education, National Center for Education Statistics, Washington, DC
Tracy PE, Wolfgang ME, Figlio RM (1990) Delinquency careers in two birth cohorts. Plenum Press, New York
Weisburd D, Bushway S, Lum C, Yang SM (2004) Trajectories of crime at places: a longitudinal study of street segments in the city of Seattle. Criminology 42(2):283–321
Whitaker C, Bastian LD (1991) Teenage victims: a national crime survey report (NCJ-128129). U.S. Department of Justice, Bureau of Justice Statistics, Washington, DC
Wilson SJ, Lipsey MW (2007) School-based interventions for aggressive and disruptive behavior: update of a meta-analysis. Am J Prev Med 33:S130–S143

School Discipline

▶ School Security Practices and Crime

School Police

▶ Police School Services

School Programs

▶ Bullying Prevention: Assessing Existing Meta-Evaluations

School Resource Officers

▶ Police School Services

School Resource Officers (SROs)

▶ Police Officers in Schools

School Security

▶ School Security Practices and Crime

School Security Practices and Crime

Allison Ann Payne and Ronald J. Eckert
Department of Sociology and Criminal Justice, Villanova University, Villanova, PA, USA

Synonyms

School crime; School discipline; School security

Overview

The crime control model of school security has led to an environment of school prisonization and student criminalization. The increasing use of restrictive security practices contrasts with both the decline in school crime and violence as well as the research suggesting that these practices may not be effective. Additionally, this intensification runs the risk of invading students' privacy and creating a negative school environment. Much research suggests that it would be better for schools to focus on other evidence-based strategies to reduce school disorder and respond to student misbehavior. If schools continue to use restrictive security practices that mirror the criminal justice system, they may, ironically, ensure that certain students are found in that system in the future.

The Prisonization of School Security

It is abundantly clear that most modern schools, particularly urban public schools, utilize a crime control model when defining and managing student behavior. This has occurred largely in response to school shootings and other visible instances of student violence (Addington 2009). More and more schools have implemented prison-like practices in an effort to improve their security (Giroux 2003), resulting in the use of many restrictive security practices and procedures intended to deter school crime and ensure the safety of students and staff. Indeed, recent statistics indicate that at least 55 % of all schools, and nearly all urban schools, now have security and surveillance programs (Devoe et al. 2005; Gottfredson and Gottfredson 2001).

Examples of restrictive security practices seen in many schools range from requiring formal dress codes (Gottfredson and Gottfredson 2001) to installing metal detectors (Brooks et al. 2000). To further detect the presence of weapons and other contraband, schools may perform regular locker searches, require students to carry clear book bags, and use drug-sniffing dogs (Brooks et al. 2000; DeVoe et al. 2005; Gottfredson and Gottfredson 2001). Visitors to campuses are generally required to sign in before entering school buildings and student identification badges are often mandated to facilitate immediate identification of rule-breakers and to deter defiance and delinquency (Brooks et al. 2000).Hallways are often supervised by school staff and administrators (Devoe et al. 2005) or even by uniformed security guards or uniformed and armed security resource officers (SROs; Giroux 2003). Additionally, many schools have chosen to install security cameras that provide constant surveillance to most areas within a school (Devoe et al. 2005). Most institutions have locked or monitored doors and gates to prevent unauthorized individuals from entering school grounds and to prohibit students from leaving campus (Devoe et al. 2005; Gottfredson and Gottfredson 2001). Schools are even investigating the use of cutting-edge technology to enhance security, such as iris recognition software, webcams, and radio frequency identification tags on ID badges (Addington 2009). Ironically, even with the increased implementation of these prison-like practices, parents and school boards have continued to call for stricter measures of control (Brooks et al. 2000) to manage the fear and anxiety surrounding school.

The intensification of school security has also led to an increase in the criminalization of students, such that schools often manage and punish student behavior in a way that is analogous to the treatment of adult criminals (Giroux 2003; Kupchik and Monahan 2006; Tredway et al. 2007). For example, the actions of students who violate school rules are often described with criminal justice language (Tredway et al. 2007): “suspects” or “repeat offenders” are subjected to “investigations,” “interrogations,” and “searches” by dogs or SROs and may then be involved in “lineups” and school “courts.” The students are then punished in ways similar to the sentences received by adult criminals; zero tolerance policies, for instance, function as the school equivalent of mandatory minimum criminal sentencing statutes. These policies have increased the use of harsh discipline, such as student exclusion through expulsion and suspension, even though these disciplinary responses have various negative impacts on both students and schools (Welch and Payne 2011). Excluded students are more likely to experience school failure, drop out of school, and engage in delinquency and drug use both in and out of school. Unfortunately, just as the use of more restrictive school security measures has increased, so has the use of these and other harsh disciplinary techniques.

Although it could be assumed that the intensification of school security is a result of increased school crime and violence, evidence suggests this is not the case (Welch and Payne 2011). In fact, these changes have occurred despite a documented decline in student delinquency and drug use, violent victimization in schools, and school-related deaths (Brooks et al. 2000; Devoe et al. 2005). Other possible explanations for this intensification include popular anxiety about high-profile instances of school violence, termed the "Columbine effect" (Addington 2009); increased school accountability for the academic performance of students (Hirschifield 2008); concerns about possible litigation in response to violent incidents on campus (Hirschifield 2008); and responses to the growing proportion of minority students (Welch and Payne 2011). While many of these explanations fall short in certain ways (Hirschfield 2008), it is likely that a combination of these factors have allowed this intensification to continue, leading schools to become less oriented around education and more like prisons focused on punishment (Giroux 2003).

Trends in School Security

The use of restrictive security practices in schools has been increasing over the past decade. Using data from the School Survey on Crime and Safety, Payne and Eckert (2011) discussed the percentage of public school principals who reported using various security measures during the years 1996, 1999, 2003, 2005, and 2007 (Table 1). The School Survey on Crime and Safety is a national questionnaire administered to public school principals by the National Center for Education Statistics. Approximately 3,500 principals from schools at all levels are asked questions on topics such as school disorder and crime, safety and discipline, and prevention programs and policies. Between the years 1996 and 2007, the imposition of school uniforms increased from 3 % to 17.5 %, drug testing of certain students increased from 4.1 % to 6.4 %, and the use of identification badges for students increased from 3.9 % to 7.6 %. Other measures show even more dramatic increases, such as the use of identification badges for faculty (25.4–58.3 %), video surveillance (19.4–55 %), and controlled access to school grounds (19.4–55 %). Finally, Table 1 shows that more and more schools controlled access to their buildings over this time, from 53 % in 1996 to 89.5 % in 2007.

School Security Practices and Crime, Table 1 Percentage of schools reporting use of security measures, School Survey on Crime and Safety

Security measure	1996	1999	2003	2005	2007
School uniforms	3.0	11.8	13.8	13.8	17.5
Drug testing of students	–	4.1	5.3	5.0	6.4
Badges or IDs for students	–	3.9	6.4	6.1	7.6
Badges or IDs for faculty	–	25.4	48.0	47.8	58.3
Video surveillance	–	19.4	36.0	42.8	55.0
Controlled access to school grounds	24.0	33.7	36.2	41.1	42.6
Controlled access to buildings	53.0	74.6	83.0	84.9	89.5
Drug sweeps	19.0	21.0	21.3	23.0	21.5
Random sweeps for contraband	–	11.8	12.8	13.1	11.4
Random metal detector checks	4.0	7.2	5.6	4.9	5.3
Pass through metal detectors daily	1.0	0.9	1.1	1.1	–
Closed during lunch	80.0	90.0	66.0	66.1	65.0
Clear book bags or banning of bags	–	5.9	6.2	6.4	6.0
Visitor sign in	96.0	96.6	98.3	97.6	98.7

Source: Payne and Eckert 2011

Similar trends are seen when Payne and Eckert (2011) analyzed student reports of security measures from the National Crime Victimization Survey: School Crime Supplement during the school years of 1999, 2001, 2003, 2005, and 2007 (Table 2). Designed by the National Center for Education Statistics and the Bureau of Justice Statistics, the School Crime Supplement is an occasional addendum to the annual NVCS that gathers specific information from individual students regarding school-related disorder and victimization on a national level. Data measuring students' perceptions of safety and crime at school are collected from approximately 6,500 12-through 18-year-old students attending public and private schools.

School Security Practices and Crime, Table 2 Percentage of students reporting use of security measures, School Crime Supplement

Security measure	1999	2001	2003	2005	2007
Security cameras	–	38.5	47.9	57.6	66.0
Security guards	54.1	63.6	69.6	67.9	68.8
Locked doors during day	38.1	48.8	52.8	54.2	60.9
Metal detectors	9.0	8.7	10.1	10.7	10.1
Badges and picture ID	–	21.2	22.5	24.7	24.3
Locker checks	53.3	53.5	53.0	52.9	53.6
Staff/adult supervision	85.4	88.3	90.6	89.8	90.0
Code of student conduct	–	95.1	95.3	95.1	95.9
Visitor sign in	87.1	90.2	91.7	92.7	94.3

Source: Payne and Eckert 2011

Of particular note is the increase of various measures between 1999 and 2001: A clear upward shift was reported in locked entrances and exit doors (38.1 % in 1999 to 48.8 % in 2001) and school security guards or law enforcement personnel (54.1 % in 1999 to 63.6 % in 2001). This trend continued through 2007, with the biggest increases in security in the areas of locked doors (38.1 % in 1999 to 60.9 % in 2007), security guards or law enforcement personnel (54.1 % in 1999 to 68.8 % in 2007), and surveillance cameras (38.5 % in 2001 to 66 % in 2007).

When schools are grouped by characteristics such as level, location, and size, differences in the type and amount of school security measures are seen (Payne and Eckert 2011). For instance, elementary schools are generally the least likely to implement restrictive security practices, followed by middle schools then high schools. One exception, however, is controlled access to buildings and school grounds, which is more often seen in elementary schools. In addition, larger schools are more likely than smaller schools to use most enhanced security measures, as are urban schools when compared with schools in towns or rural areas. Finally, schools with a greater percentage of minority students and students who receive free or reduced-price lunches are more likely to implement these practices (Payne and Eckert 2011).

The Effectiveness of Security Measures

Very little research exists that evaluates the effectiveness of school security measures. The pervasiveness of these measures combined with the lack of knowledge regarding their impact is of concern; if these practices are ineffective, they could allow schools to feel secure when they are not and may even create a dangerous environment by diverting money and resources from measures that do actually work (Addington 2009).

Several studies that have been conducted examine the effectiveness of security measures based on perceptions of specific practices rather than actual impact. Overall, school community members perceive particular measures as positive and believe that they prevent crime and disorder. For instance, both students and staff view SROs as effective; similar beliefs are held by school administrators regarding security cameras. Unfortunately, there is little evidence to support these perceptions (for a full review, see Addington 2009).

The small amount of evaluation research that has examined more objective outcomes related to the use of security measures is inconclusive. An early study of school security examined the impact of metal detectors in three program and 12 comparison high schools in New York City (Ginsberg and Loffredo 1993). Surveys were completed by 13,999 students in June 1992. Students in schools with metal detectors were less likely to carry weapons inside the school building or to and from school. However, there were no student differences in weapon carrying in other locations nor in threats or physical fights anywhere (Ginsberg and Loffredo 1993).

Mayer and Leone (1999) examined the relationships among school security, discipline management, student self-protection, and school disorder. Data were taken from 6,947 public school students in grades 7–12 who were interviewed as part of the 1995 School Crime Supplement. The structural equation model that was estimated contained four constructs: secure building, which included security measures such as guards, visitor sign in, metal detectors, and locked doors; system of law, which reflected students' perceptions of rules and rule enforcement; school disorder, which included gang presence, drug availability, and personal attack and theft; and individual self-protection, which contained student avoidance of school location and fear of attack. Results suggested that schools which used physical security measures, such as metal detectors and locked doors, experienced more crime and disorder, as did schools that used personnel-based measures, such as security guards and hall monitors. In contrast, schools that focused on communication of schools rules and consequences for rule infractions experienced less disorder.

Finally, Chen (2008) analyzed a model of school crime that included two forms of physical school security: the number of ways a school controls access to campus and the number of ways in which a school monitors student activities. Both measures were hypothesized to be negatively related to the number of crimes that had occurred in the school in the past 12 months. Other factors in the model included urbanicity, community crime, student socioeconomic status, school size, student mobility, student misbehavior, and serious disciplinary penalties such as transfers and suspensions. Data were taken from 712 secondary schools whose principals had participated in the 2000 School Survey on Crime and Safety. Bivariate correlations showed that the number of ways a school controls access to their buildings was positively correlated with the number of crimes, while the number of ways a school monitors student activities was not significantly related to crimes. When the full model was estimated using structural equation modeling, the path between school security (a latent variable containing both building access control and student activity monitoring) and number of criminal incidents was negative but nonsignificant, contrary to the study's hypothesis. These findings suggest that, while school security measures may not increase crime as reported by Mayer and Leone (1999), they do not appear to reduce crime either (Chen 2008).

Ultimately, there is little evidence supporting the effectiveness of security measures. In addition, the few studies that do exist suffer from severe limitations. First, several of the studies measure effectiveness through perceptions of school community members, rather than actual impact on school disorder (see Addington 2009). In addition, all of these studies rely on cross-sectional data with no baseline measures of school disorder, thus making it impossible to truly establish the temporal order of the implementation of security measures and the level of school disorder (Ginsberg and Loffredo 1993; Mayer and Leone 1999; Chen 2008).

Student Civil Liberties

An unintended consequence of enhanced school security that has generated much discussion is infringement on student civil liberties (Addington 2009; Berger 2003), which can be separated into two related concerns: suspicionless searches and privacy encroachments. Searches of students' persons, lockers, and belongings have been the subject of many court cases regarding the fourth Amendment, including several Supreme Court rulings. One landmark case, New Jersey v. TLO (1985), established that searches conducted by school administrators require a far lower standard of suspicion than police searches of citizens on the street, because the intrusion on students' privacy is deemed minor and is overshadowed by public health and safety concerns (Berger 2003). The rationale for this standard is the "special needs" doctrine that emerged from this case, stating that a special need exists such that schools must maintain discipline in order to have an environment conducive to learning. Therefore, it is not necessary for administrators to have probable cause before searching students and their belongings (Berger 2003), including the use of metal detectors and locker searches. Essentially, although students have the right to privacy, this right is counterbalanced by the "special need for an immediate response to behavior that threatens either the safety of schoolchildren and teachers or the educational process itself" (New Jersey v. TLO 1985, p.353). Another landmark case, Vernonia School District 47J v. Acton (1995), applied this doctrine to drug testing of student athletes; later, this was expanded to allow schools to drug test all students participating in any extracurricular activity (Board of Education v. Earls 2002). Thus, although students do not waive their fourth Amendment rights in school, the special needs doctrine offers school officials a large amount of discretion and flexibility in order to maintain a safe environment (Berger 2003).

Although these Supreme Court rulings applied the special needs doctrine to searches by school administrators for contraband in order to reduce or prevent drug use, lower court rulings have expanded the application to include searches by law enforcement personnel and security measures designed to prevent violence (Addington 2009; Berger 2003). Originally, searches allowed through New Jersey v. TLO and similar cases were limited to those conducted by school personnel rather than police officers, based on the rationale that the mission of school officials is to educate rather than collect evidence for a criminal prosecution. However, public anxiety over school crime and violence has led to a new view of student searches by police officers as acceptable under the special needs doctrine because these searches are seen as minimal, nonintrusive, and within the realm of reasonable suspicion (Berger 2003). In addition, decisions by lower appellate courts have also expanded this power to security searches designed to prevent violence, such as metal detector screenings (Addington 2009).

There is also a sense of these security measures infringing on students' privacy, even beyond the legal exploration of suspicionless searches (Addington 2009). This can be seen particularly with tactics that monitor student behavior in public areas, such as security cameras and the presence of SROS, which are also the measures that tend to be implemented most often. Concerns appear when these measures are used in ways for which they are not originally intended (Addington 2009). For example, if security cameras are originally installed for violence prevention, is it legitimate to then use them to thwart vandalism? Additionally, the use of cutting-edge technology to enhance security, such as webcams and radio frequency identification tags on ID badges, has the potential for even greater privacy infringement (Addington 2009).

The Impact on School Community

This invasion on student privacy and civil liberties may lead to another consequence of the intensification of school security: the altering of the school environment. Much research (summarized in the "School Social Organization, Discipline Management, and Crime" entry in this encyclopedia) has illustrated the importance of school climate and social organization. Schools with healthy supportive environments generally have more effective teachers who enjoy their jobs more and have more positive perceptions of the school administration (Gottfredson et al. 2005). In addition, students who attend these types of schools are more attached to teachers, more committed to school, and have stronger belief in school norms (Payne et al. 2003). This positive school community and subsequent student bonding, in turn, leads to a beneficial learning environment and less school crime and disorder (Gottfredson et al. 2005; Payne et al. 2003).

But what happens to the school community and subsequent student behavior when restrictive security practices are implemented? Restrictive measures may first lead to an oppositional relationship between students and staff, as students could protest these practices with petition drives and class boycotts (Berger 2003). Even without this type of adversarial response, aggressive security measures may negatively impact the school environment by interfering with the educational process, disrupting the learning environment, and wasting valuable class time (Berger 2003). For example, Devine (1996) described how teachers in a New York City school could not teach due to "loud noises coming from the walkie-talkies in the corridors" supervised by security staff (p. 89), while Glazer (1992) detailed how administrators in another New York City school took close to 3 h to "funnel all 3,000 students into the gym, where they are frisked with hand-held metal detectors and their book bags are probed" (p. 790).

In addition, the implementation of many of these intrusive security practices often creates a climate of fear and resentment, resulting in negative attitudes toward school (Addington 2009; Hyman and Peronne 1998). For example, (Hyman and Peronne 1998) detail how searches result in detrimental consequences for both students and staff, such as lower student morale, distrust for staff, and alienation for law enforcement authorities; other studies have shown similar student alienation and mistrust as a consequence of restrictive security measures (Noguera 1995). Metal detectors, security patrols, and lock-down drills make schools feel like prisons (Noguera 1995) and often make students more afraid rather than less (Devine 1996). As Devine (1996) describes, aggressive security measures in urban schools result in "a climate of fear that indoctrinates youth into a culture of violence and dictates that only those exhibiting a 'tough' demeanor will survive" (p.179). As fear increases, so does student resentment and hostility (Hyman and Peronne 1998), leading to lower student bonding and a negative school environment (Addington 2009). Thus, it is possible that intrusive security might have the opposite effect than intended, leading students to increasingly break school rules. Ultimately, there is a strong chance that the intensification of school security leads to students feeling as though they do not belong in their schools, which, in turn, may decrease learning and increase the likelihood of these students engaging in deviant and delinquent behavior.

The Future of School Security

Beyond a doubt, far more "evidence-based research is…needed to evaluate the costs and effectiveness of school security measures" (Berger 2003, p. 351). The lack of studies clearly shows that evaluations must be conducted to examine the impact of these practices on the levels of misbehavior and violence in schools. Further, the little research that has been done is severely limited in several ways. First, several studies examine perceptions of school community members as an outcome, rather than actual effectiveness on school crime and violence (for a full review, see Addington 2009). In addition, the few studies that do examine the impact of these measures on school disorder rely on cross-sectional data with no baseline measures of crime and deviance (Ginsberg and Loffredo 1993; Mayer and Leone 1999; Chen 2008). This makes it impossible to truly establish the temporal order of the relationship between restrictive school security and school disorder; that is, it is possible that these practices were implemented in schools that were already displaying high levels of crime, rather than leading to more crime after implementation, as suggested by Mayer and Leone (1999). This lack of research is surprising given the large amount of evidence supporting the use of other policies and programs designed to prevent student misbehavior (see "School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis" entry in this encyclopedia). Future studies should use longitudinal data and include baseline measures of crime to examine the impact of restrictive measures on school disorder.

In addition, the financial costs of these practices need to be studied. Little data exist on the cumulative cost to schools of metal detectors, SROs, security cameras, and other such measures (Addington 2009). Along with a gathering of costs, cost-benefit analyses should be conducted to evaluate whether the financial price is worth the effectiveness. Included in this consideration should be the examination of the possibility of "budgetary trade-offs" that may happen in order to pay for the enhanced security; it is possible that the cost of these measures leads to fewer resources for actual learning, such as books and staff (Addington 2009, p. 1440).

Even if they are found to be both effective and cost-effective, the use of these restrictive security measures should be balanced against the costs to student civil liberties and the school environment. Although more research is needed, studies do suggest that these measures infringe upon student civil rights (Berger 2003). As Addington (2009) discusses, it appears as though no one is protecting the students' rights, as no one with power is questioning the impact of enhanced security measures on student privacy and civil liberties. Students may want to change polices but they do not have the ability to do so, while most parents believe this security keeps their children safe and school officials are likely to give in to the demand for strong school security. With the possibility of technology, such as radio frequency identification tags, enhancing security even further, it is imperative that the impact of these measures on students' rights is considered. While student safety is clearly important, school officials should work toward developing and implementing strategies that keep student safe without infringing on their fourth Amendment rights (Berger 2003).

It is also possible that enhanced security measures create a negative learning environment, filled with fear and hostility. The few studies that have been conducted suggest that implementing restrictive measures such as metal detectors, locker checks, and security personnel may lead to an "unwelcome, almost jail-like, heavily scrutinized environment [that] may foster the violence and disorder school administrators hope to avoid" (Mayer and Leone 1999, p. 349; Chen 2008). The findings suggest, instead, that more attention should be paid to communicating school rules and norms and fairly and consistently enforcing these rules with appropriate consequences (Mayer and Leone 1999), a practice that is well supported by previous research (Berger 2003; Gottfredson et al. 2005).

This leads to a broader discussion on what schools can do to effectively reduce disorder and violence. Fortunately, a large body of research exists that establishes effective strategies. These practices are discussed in two related encyclopedia entries: "School Social Organization, Discipline Management, and Crime" and "School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis."

Conclusion

As the crime control model continues to guide security, "schools grow more like prisons than institutions of education" (Giroux 2003: 553). Reliance on this model has intensified security practices to the point of creating a picture of school prisonization and student criminalization, despite a clear decline in school crime and violence. The little research that has been conducted suggests that, although school officials appear to be pleased with enhanced security, these practices may not be effective for reducing student misbehavior. In addition, this intensification runs the strong risk of invading students' privacy and creating a negative school environment, both of which may, in fact, increase school disorder. Until evidence supporting the use of these practices is provided, schools would benefit from focusing on other evidence-based strategies to reduce school disorder and respond to student misbehavior, rather than continuing to use restrictive security practices that mirror the criminal justice system. If changes are not made, the school-to-prison pipeline may, ironically, ensure that certain students are found in that system in the future.

Notes and Acknowledgements This entry is based on a more detailed manuscript (Payne and Eckert 2011).

Related Entries

- ▶ Police Officers in Schools
- ▶ School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis
- ▶ School Social Organization, Discipline Management, and Crime

Recommended Reading and References

Addington LA (2009) Cops and cameras: public school security as a policy response to columbine. Am Behav Sci 52(10):1426–1446

Berger R (2003) The best of both worlds: school security and the disappearing fourth amendment rights of students. Crim Justice Rev 28:336–354

Brooks K, Schiraldi V, Ziedenberg J (2000) School house hype: two years later. Policy report of Justice Policy Institute and Children's Law Center, Inc

Chen G (2008) Communities, students, schools, and school crime: a confirmatory study of crime in U.S. high schools. Urban Educ 43:301–318

Devine J (1996) Maximum security: the culture of violence in inner-city schools. University of Chicago Press, Chicago

Devoe JF, Peter K, Noonan M, Snyder TD, Baum K (2005) Indicators of school crime and safety: 2005. U.S. Departments of Education and Justice, Washington

Ginsberg C, Loffredo L (1993) Violence-related attitudes and behaviors of high school students—New York city, 1992. J Stud Health 63(10):438–439

Giroux HA (2003) Racial injustice and disposable youth in the age of zero tolerance. Int J Qual Stud Educ 16:553–565

Glazer S (1992) Violence in schools. CQ Res 34:785–808

Gottfredson GD, Gottfredson DC (2001) What schools do to prevent problem behavior and promote safe environments. J Educ Psychol Consult 12:313–344

Gottfredson GD, Gottfredson DC, Payne AA, Gottfredson ND (2005) School climate predictors of school disorder: results from the national study of delinquency prevention in schools. J Res Crime Delinq 42:412–444

Hirschfield PJ (2008) Preparing for prison? The criminalization of school discipline in the USA. Theor Criminol 12:79–101

Hyman IA, Peronne DC (1998) The other side of school violence: educator policies and practices that may contribute to student misbehavior. J Sch Psychol 36:7–27

Kupchik A, Monahan T (2006) The New American School: Preparation for post-industrial discipline. British Journal of Sociology of Education 27:617–632

Mayer MJ, Leone PE (1999) A structural analysis of school violence and disruption: implications for creating safer schools. Educ Treat Child 23:333–356

Noguera PA (1995) Preventing and producing violence: A critical analysis of responses to school violence. Harvard Educational Review, 65(2), 189–212

Noguera PA (2008) The trouble with black boys...and other reflections on race, equity, and the future of public education. Jossey-Bass, San Francisco

Payne AA, Eckert RJ (2011) School security: How effective is it? Unpublished manuscript

Payne AA, Gottfredson DC, Gottfredson GD (2003) Schools as communities: the relationships among communal school organization, student bonding, and school disorder. Criminology 41:749–777

Tredway L, Brill F, Hernandez J (2007) Taking off the cape: the stories of novice urban leadership. Theory Pract 46:212–221

Welch K, Payne AA (2011) Exclusionary school punishment: the effect of racial threat on expulsion and suspension. Youth Violence Juvenile Justice 10:155–171

School Social Organization, Discipline Management, and Crime

Denise C. Gottfredson[1], Philip J. Cook[2,3] and Chongmin Na[4]

[1]Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA

[2]Sanford School of Public Policy, Duke University, Durham, NC, USA

[3]NBER, Cambridge, MA, USA

[4]School of Human Science and Humanities, University of Houston-Clear Lake, Houston, TX, USA

Overview

School violence, drug use, vandalism, gang activity, bullying, and theft are costly and interfere with academic achievement. Student misbehavior interferes with teaching and learning and is one of the primary sources of teacher turnover in our nation's schools. Gallup polls from the past 20 years show that the percentage of parents who report being concerned about the physical safety of their children while at school has ranged from 15 % to 55 %, with the highest percentages registering just after the infamous school shootings at Columbine High School in 1999. Reducing crime rates has become an increasingly high priority for America's schools.

There has also been an important trend in the official response to school crime. The response has become increasingly formal over the last 20 years, with greater recourse to arrest and the juvenile courts rather than school-based discipline. To some extent, this trend has been furthered by federal law that has imposed zero-tolerance rules for some offenses and has subsidized the hiring of uniformed officers to police the schools. The shift has been from administrative discretion to mandatory penalties and from in-school discipline to increasing use of suspension or arrest. At the same time, there has been a considerable investment in the use of surveillance cameras and metal detectors.

While the increasing formality in school response to crime has coincided with the declining crime rates, there is no clear indication of whether the new approach gets any of the credit. Indeed, the evaluation literature that we review here has very little to say about the likely effects of these changes. As so often happens, there appears to be a disconnect between policy and research.

There are alternatives to the get-tough approach with its reliance on deterrence and exclusion. We know that some schools do a much better job than others in controlling the behavior of their students. Characteristic of successful schools in this respect is that they are close-knit communities where rules of acceptable behavior are clearly communicated and consistently (if not harshly) enforced. In addition to good management practices, there is much that can be done in the classroom that has demonstrated effectiveness in improving behavior.

In this entry, we summarize evidence showing that how the school is organized and managed (in general) influences problem behavior and school safety. We then discuss two specific aspects of school organizational climate – discipline management and school culture, respectively – and how they can be manipulated to reduce crime and related problem behaviors. We conclude with recommendations to guide future evaluation research on school-based interventions. Throughout the entry, our focus is on how schools can and do influence the behavior of students while they are enrolled.

Does School Organization Matter?

How the school is organized and managed influences problem behavior and school safety. In an early national study of school disorder, Gottfredson and Gottfredson (1985) showed that even after controlling for input characteristics of students and communities in which schools were located, characteristics of schools accounted for an additional 12 % (junior high) and 18 % (senior high) of variance in teacher victimization rates. More recent national studies have replicated these findings and extended them to show that school characteristics account for a substantial amount of variance not only in teacher victimization but also in student reports of victimization and delinquency (Gottfredson et al. 2005).

Which aspects of the way schools are organized and managed influence crime and disorder? Cook et al. (2010) discuss school system decisions that influence the demographic composition of schools and the number and types of other students to whom a child is exposed. Schools and school districts have a good deal of control over the makeup of the student body. Schools can be based on neighborhood residential patterns or integrated across race and class. The grade span for elementary and middle schools can be adjusted. Truancy and dropout prevention programs can be pursued with more or less vigor, and troublesome students reassigned. Whether failing students are retained in grade or given a social promotion influences the extent of age homogeneity within classrooms. Students who are enrolled in the school can be tracked on the basis of academic potential or mixed together. These decisions influence the characteristics of other students to whom youths will be exposed. Importantly, these decisions determine the pool of youths from which highly influential peers will be selected as well as the dominant peer culture in the school.

School and school district decisions about curricular content and teaching methods are also important. These decisions determine student success in school and decisions to persist in school. As summarized in the Wilson entry elsewhere in this section, the use of specialized prevention curricula directly influences the level of problem behavior. Below, we discuss two additional characteristics of schools that influence crime and disorder: (1) policies and procedures governing discipline management that directly affect the extent to which formal sanctions are applied and the effectiveness of these sanctions and (2) aspects of the school social organization that affect the nature of interactions among teachers and students (and hence the application of social controls) and the school culture.

Discipline Management

Cook et al. (2010) summarize findings from 12 studies that looked at the association of discipline management practices with school crime. The results show remarkable consistency: when schools monitor students and control access to the campus, and when students perceive that school rules are fair and consistently enforced, schools experience lower levels of problem behavior. Inclusion of students in establishing school rules and policies for dealing with problem behaviors has also been found to be related to lower levels of problem behavior, most likely because students are apt to internalize school rules if they have helped to shape them. On the other hand, severity of sanctions is not related to a reduction in problem behaviors. These findings conform to the main findings from deterrence research that the certainty of punishment has greater deterrent effect than the severity of punishment.

Of course, there has been considerable policy attention to school disciplinary practices, especially in response to the spate of school shootings experienced in the 1980s and 1990s. Most schools employ security and surveillance strategies aimed at keeping intruders out and preventing weapons from coming into the schools. Common practices include controlled entry and identification systems, metal detectors, security personnel or volunteers who challenge intruders, or doors fitted with electromagnetic locks. Since the late 1990s, school resource officers (SROs) have also been especially popular in secondary schools as a way to prevent violence, encouraged by federal subsidies. The Payne and Eckert and Na and Gottfredson entries elsewhere in this section discuss what is known about the effectiveness of these practices.

A closely related discipline strategy is the use of zero-tolerance policies in schools – another “tough on crime” practice engendered by the epidemic of youth violence in the late 1980s and the school rampage shootings of that decade and the next. The US Congress adopted the Gun-Free Schools Act in 1994, mandating that students be suspended for 1 year if they brought a gun to school. A large majority of school districts adopted zero-tolerance policies for alcohol, tobacco, drugs, and violence. The use of suspension, especially long-term suspension, is thought to have disproportionate impact on minority and special education populations, whose behavior places them more at risk for suspension. Civil liberty advocates have argued that zero-tolerance policies rob youths of their right to a public education.

As with other security-related school polices, little high-quality evidence is available to guide decisions about which discipline management policies produce the most desirable outcomes. The issue is complex, requiring consideration of the trade-offs between in-school and out-of-school crime, the welfare of the youths who perpetrate the school-based offenses versus that of the other youths in the school, and long-term versus short-term outcomes. Clearly, removing troublemakers from school helps to maintain an environment more suitable for learning for these remaining students. But the costs of doing so for the offenders and society are not well understood. A complete analysis of the effect of zero-tolerance policies on youth crime would consider the displacement of crime from school to the community as well as the consequences for the suspended youths’ long-term criminal and academic careers. As youths lose more days of school to suspension, promotion to the next grade becomes less likely. And as youths fall further behind grade, they become much less likely to graduate, which is likely to increase subsequent crime. Clearly, although zero-tolerance policies benefit the classmates of troublesome youths, a rational discipline policy would also have to consider the broader consequences of such policies for the community.

More consistent with the research on effective crime deterrents are school discipline polices that emphasize the certainty of response to misbehavior over the severity of the response. Among the most effective school-based strategies for reducing youth violence, aggression, and problem behavior are behavioral interventions that target specific behaviors, systematically remove rewards for undesirable behavior, and apply contingent rewards for desired behavior or punishment for undesired behavior. These interventions are often applied to the high-risk youths who are most at risk for being suspended from school under zero-tolerance policies and as such could be incorporated into school routines for discipline management. Gottfredson et al. (2002) meta-analysis reported average effect size on measures of antisocial behavior and aggression of 0.34 ($p <.05$) across 12 studies of this type of behavioral intervention.

Examples of particularly effective behavioral interventions currently in use in schools are the "Good Behavior Game" (GBG; Kellam et al. 2008) and "home-based reinforcement" (Schumaker et al. 1977). The GBG is a classroom-based application of behavioral principles in which elementary school children are divided into small teams, and the teams are rewarded when the classroom behavior of the entire team meets or exceeds a preestablished standard. The GBG is played several times per week throughout the school year. The intervention was evaluated through a randomized trial involving 19 schools in Baltimore, with posttests conducted immediately following the intervention, as well as 6 and 14 years later. The results of this study indicate that participation in GBG is related to immediate reductions in aggressive behavior, rates of diagnosed antisocial personality disorder, and long-term effects (14 years later) on drug and alcohol use and smoking. Home-based reinforcement (HBR), applied to individual students displaying behavior problems, requires cooperation between teachers and parents in the management of the child's behavior. After agreeing upon specific child behaviors to be extinguished or encouraged and establishing a baseline for these behaviors, teachers systematically record data on the target behavior on a "daily report card" that goes home to the parents. The parents, who generally have access to a wider array of reinforcers and punishments than do the teachers, use the teacher's information to guide the application of rewards and punishments. As the desired behavior emerges, the frequency of reports for home is reduced, and the schedule of contingencies is relaxed. In the earliest research on HBR, application of this technique to junior high school students showed that school rule compliance, teacher satisfaction with the student, and academic performance improved as a result of participation in an HBR program (Schumaker et al. 1977). A recent review of 18 empirical studies of "school-home collaboration" interventions (Cox 2005) concluded that behavioral interventions using the daily report card strategy had the strongest effects on problem behavior. Lasting effects on crime are unknown.

These relatively simple and inexpensive behavioral interventions represent a potentially potent school-based prevention strategy that might be incorporated into routine school practice. The 1997 reauthorization of the Individuals with Disabilities Education Act (IDEA; P.L. 105–17) required functional assessment and behavioral intervention procedures to be implemented in the disciplining of students with disabilities. The evidence-based programs described here would meet these federal requirements.

Behavioral principles have also been incorporated into school-wide discipline management systems. These systems are typically designed to clarify expectations for behavior. They establish school and classroom rules, communicate these rules as well as consequences for breaking them clearly to parents and students, establish systems for tracking both youth behavior and consequences applied by the schools, and monitor the consistency of the application of consequences for misbehavior. School-wide discipline management efforts, most often implemented by a school-based team of educators, are highly consistent with the research summarized earlier suggesting that students' perceptions of school rules as fair and consistently enforced are related to reductions in problem behavior.

The meta-analysis described earlier (Gottfredson et al. 2002) also examined the effectiveness of this type of school-wide effort to improve discipline management and reported average effect size on measures of crime (0.27, $p < .05$) and alcohol and other drug use (0.24, $p < .05$). Among the studies included in the meta-analysis are two early studies of the effects of school-wide discipline management systems on problem behavior outcomes. Students in the intervention schools in the first of these efforts (Project PATHE implemented in nine Charleston, South Carolina, schools) reported less delinquent behavior and drug use and fewer punishments in school relative to the students in the comparison schools (Gottfredson 1986). A similar intervention was tested in a troubled Baltimore, Maryland, junior high school, with a special emphasis on replacing the school's reliance on out-of-school suspension with a wider array of consequences for misbehavior. This intervention, which added positive reinforcement for desired behavior to the mix of consequences routinely used, also showed positive effects on student delinquency and rebellious behavior (Gottfredson 1987a). This early research, although based on relatively small numbers of schools and lacking randomization to condition, suggested that behavioral principles could be incorporated into "normal" school disciplinary practices and that an emphasis on consistency of rule enforcement as opposed to severity of punishment provided an effective deterrent.

Contemporary approaches to discipline management incorporate behavioral principles into comprehensive systems that include school-wide discipline policies and practices as well as targeted behavioral interventions. One popular approach is School-Wide Positive Behavior Support (SWPBS). This system, adopted by over 5,600 schools throughout the United States, uses a school-team approach to apply behavioral interventions at different levels of intensity for students at different levels of need. Universal interventions focus on clarity of school and classroom rules and consistency of enforcement, and on screening for more serious behavior disorders. Group-based behavioral interventions are employed with the 5–10 % of youths who do not respond to the universal interventions. In addition, intensive, individualized behavioral interventions are employed to manage the behavior of the small segment of the population that is especially at risk. Unfortunately, the research on the effectiveness of SWPBS is not as sophisticated as it should be for such a widely disseminated program. Although dozens of studies have demonstrated that problem behavior decreases after the intervention is put in place, only one (Sprague et al. 2001) compared change in the intervention school(s) with the change that might be expected in the absence of an intervention. Even this study is not useful for isolating the effects of the behavior management strategies because it also included the introduction of a prevention curriculum along with the school-wide behavioral supports. Higher-quality research is needed to assess the effects of this promising approach on crime both in and out of school.

School Culture

School culture is potentially the most potent aspect of school climate because it involves proximal interpersonal influences on student behavior. *School culture* refers to the quality of human relationships in the school and includes both peer culture and the extent to which the organization is communally organized. All of these dimensions influence youth crime and can be successfully manipulated to reduce it.

Behavioral Norms

Cultural norms, expectations, and beliefs influence all behaviors. Recent research on school culture (summarized in Cook et al. 2010) concurs that norms and expectations for behavior, of both peers and adult, are powerful determinants of behavior, net of the individual's own beliefs.

Of course, school "inputs" are key determinants of the predominant cultural beliefs in the school. School desegregation and retention policies, as well as the grade span, of the school can influence school culture by altering the mix of students in the school. But several more programmatic attempts to alter school culture have also been studied. These programs have in common a focus on clarifying behavioral norms. That is, in contrast to the instructional programs described in the previous section that focus on teaching youths with specific social competency skills, these normative change programs focus on clarifying expectations for behavior. Some signal appropriate behavior through media campaigns or ceremonies; others involve youths in activities aimed at clarifying misperceptions about normative behavior; and still others increase exposure to prosocial models and messages.

Several studies of attempts to clarify norms for behavior have been reported. Gottfredson et al. (2002) summarized effects reported in 13 studies and concluded that such programs are effective for reducing crime, substance use, and antisocial behavior. Two of the better-known examples of programs in this category are the Bullying Prevention Program (Olweus et al. 1999) and the Safe Dates Program (Foshee et al. 1998).

Olweus's anti-bullying program includes school-wide, classroom, and individual components. School-wide components include increased adult supervision at bullying "hot spots" and school-wide discussions of bullying. Classroom components focus on developing and enforcing rules against bullying. Individual counseling is also provided to children identified as bullies and victims. A large-scale evaluation of this program in Norwegian schools demonstrated that it led to reductions in student bullying and victimization and decreases in the incidence of vandalism, fighting, and theft (Olweus et al. 1999). A very recent review of anti-bullying programs summarizing results from 59 studies conducted between 1983 and 2008 (Farrington and Ttofi 2009) confirmed that anti-bullying programs are effective for reducing bullying and student victimization and that Olweus's program is particularly effective.

The Safe Dates Program targets norms for dating violence among adolescents. The school portion of the intervention includes a theater production performed by peers; a ten-session curriculum addressing dating violence norms, gender stereotyping, and conflict management skills; and a poster contest. The community portion of the intervention includes services for adolescents experiencing abuse and training for community service providers. Foshee et al. (1998) found that intervention students reported less psychological abuse and violence against dating partners than did control students.

Based on these and other relatively rigorous evaluations, Gottfredson et al. (2002) concluded that interventions aimed at establishing norms or expectations for behavior can be effective in preventing substance use, delinquency, aggression, and other problem behaviors. It should be noted, however, that evaluations of these programs seldom provide clean tests of the proposition that culture matters, since the programs more often than not combine attempts to alter norms with other components aimed at increasing levels of supervision and enforcement (e.g., Olweus) or improving social competency skills (Foshee).

We would be remiss if we failed to mention that sometimes school-based practices seek to clarify norms for behavior backfire. One example is a peer counseling program that deliberately mixed delinquent and nondelinquent youths in counseling sessions in which youths were encouraged to share their problems. The intent was that the negative beliefs and attitudes voiced by the delinquent youths would be corrected through interaction with the nondelinquent youths. A randomized experiment testing this program as implemented in the Chicago Public Schools (Gottfredson 1987b) reported predominantly harmful effects for high school students: high school treatment youths reported significantly *more* delinquent behavior than controls. A more recent large-scale evaluation of the Reconnecting Youth program (Cho et al. 2005) also found negative effects for a group counseling program for at-risk high school students. This program sought to "reconnect" truant, underachieving high school students (and to reduce their deviance and substance use) by developing a positive peer-group culture. Students were grouped together in classes of 10–12 students for a full semester during which a trained group leader (following a standardized curriculum) attempted to develop a climate conducive to building trust. The evaluation reported only negative effects 6 months following the end of the intervention. Treatment students showed greater bonding to high-risk peers, lower bonding to school and conventional peers, lower grade point average (GPA), and higher anger than control students at the 6-month follow-up.

Communal Social Organization

A second aspect of school culture that has been studied extensively pertains to the affective bonds between students and teachers and among adults in the school. The concept of "communal social organization" (CSO) was first introduced as part of the effective schools debate in the 1980s and studied by Bryk and colleagues (Bryk and Driscoll 1988) mostly in the context of predictors of school achievement. Communally organized schools are schools characterized by high levels of social support, connectedness, common goals, and sense of shared purpose. Members of such schools are more likely to be involved and personally committed to the school. This aspect of school culture is especially important for school crime research because individual-level student affective bonds are an important predictor of delinquency, and it seems reasonable to hypothesize that schools high on CSO would produce higher levels of student bonding to school.

Research suggests that average student attachment to school and CSO more generally do inhibit student problem behaviors. The most comprehensive test of this linkage was provided by Payne et al. (2003) using data from the NSDPS. This study demonstrated that more communally organized schools experience less student delinquency and teacher victimization and that the effect of communal school organization on student delinquency is mediated by average student bonding.

This survey research dovetails nicely with an ambitious ethnographic study of school violence conducted for the National Research Council. In 2003, the Committee to Study Youth Violence in Schools of the National Research Council published its report on the circumstances surrounding several incidents involving extreme lethal violence that had occurred in the nation's schools (National Research Council 2003). The report was based on detailed case studies of six schools and communities that had experienced school shootings resulting in death. Among the committee's several insights into the factors leading to the incidents is the following:

> the sense of community between youth and adults in these schools...was lacking. In the worst example, the school allowed a school newspaper to print an article that humiliated one of the students who became a shooter. The adults involved may have been too distant from the students to prevent some social processes leading to the potential for violence or resulting in an intolerable humiliation from some potentially vulnerable youth. (p. 256)

This observation is consistent with the research on more mundane forms of school violence just summarized. It suggests that strategies that increase social bonds between students and others in their schools will reduce misbehavior by increasing informal controls. Students who care what adults in the school think about them will be less likely to act in ways that jeopardize their positive regard. More concretely, students who have close ties to the adults in the school will be more likely to report on rumors of impending attacks. But how can such bonds be built or maintained? Possibilities include organizing the school so that the typical teacher interacts with fewer students, reducing class size, and creating more "communal" social environments in which members are more tightly joined together by common goals and in which members are held in place by the support and positive regard of others in the organization. Reorganizing schools to create a smaller feel to the schooling experience is an effective strategy for increasing youths' sense of connection and that enhanced connectedness should hold criminal behavior in check.

A less drastic intervention with the same objectives is mentoring. Youth mentoring programs often target youths at risk of behavioral problems, assigning them to an adult mentor who spends time with the young person, provides support and guidance, and offers general guidance. Evaluations of such programs have been mixed, but often null or weak results can be attributed to implementation failure. As with any voluntary program, mentoring programs in practice are often not as intensive as intended (e.g., Karcher 2008). However, a recent meta-analysis of mentoring programs (Eby et al. 2007) demonstrated small but positive effects of mentoring programs on several behaviors of interest in this entry: withdrawal behaviors (e.g., school dropout, truancy – 18 studies), deviance (e.g., suspension from school, aggressive behavior, property crime – 15 studies), and substance use (7 studies). This review included a wide range of types of mentoring programs, but outcomes for youth mentoring programs were as strong on these outcomes as were the other types of mentoring programs (academic and workplace mentoring) included in the review.

One of the better-known models for adult mentoring, the Big Brothers Big Sisters program (BBBS), is a community-based program identified by BVP as a model primarily on the basis of evidence from a large-scale randomized trial that found that mentored youths were 46 % less likely than control youth to initiate drug use, 27 % less likely to initiate alcohol use, and almost one-third less likely to hit someone during the study period (Tierney et al. 1995). Community-based mentoring involves meetings between the mentor and mentee at times and places selected by the pair. Many schools now provide "school-based mentoring," (SBM), which involves meetings primarily in school during the school day. A recent evaluation of the BBBS SBM model, also involving random assignment of a large number of youths, shows that although it is not as effective as the community-based alternative, SBM does improve academic performance, reduce truancy, and reduce serious school infractions (Herrera et al. 2007) at least during the first year of mentoring. Consistent with results from smaller-scale randomized trials of SBM showing positive effects on connectedness and social support (Karcher 2008); Herrera et al. (2007) found that mentored youths reported more often than controls the presence of a nonparental adult in their life who provides social supports. At the end of the second year of the study during which minimal SBM was provided, the positive program effect on truancy was sustained but the other positive effects were not. Herrera et al. (2007) conclude that although the SBM model is promising, it needs to be strengthened to ensure longer and higher-quality mentor/mentee matches than are typically found in schools.

Discussion and Conclusions

In this brief essay, we summarized research on school discipline management policies and practices and showed that they are important determinants of school crime. Research consistently shows that in schools in which students report that the school rules are clearly stated, fair, and consistently enforced, and in schools in which students have participated in establishing mechanisms for reducing misbehavior, students are much less likely to engage in problem behaviors. We showed that evaluations of specific school-based programs that employ behavioral strategies to monitor and reinforce student behavior are effective both for controlling behavior in school and for reducing subsequent crime. Also, altering school-wide discipline management policies and practices to incorporate behavioral principles, clarify expectations for behavior, and consistently enforce rules reduces problem behavior. We discussed popular "get-tough" approaches to school discipline such as zero-tolerance policies. Although the effects of these polices on crime are not known, we argued that they might actually increase crime outside of school. There is a clear need for rigorous research on the effects of these policies.

Finally, we summarized research showing that perceptions of social norms for behavior are related as expected to problem behavior, net of individuals' personal beliefs. In schools in which the prevailing norm is to condone delinquent activities, students are more likely to do so regardless of their own personal dispositions to engage in these behaviors. But we showed that schools can intervene to change perceptions of norms and expectations for behavior and that doing so reduces delinquency, although attempts to do so sometimes backfire. We also reported on evidence suggesting that in schools in which students feel an emotional attachment to the adults in the school, their misbehavior is restrained. We discussed several strategies that might increase communal social organization and that show promise for increasing youths' sense of connection to the school. We reviewed research on school-based mentoring programs and showed that they also hold considerable promise for crime prevention. Although research documents positive effects of these programs on social relations outcomes, more work is needed to test the full potential of more potent models of school-based mentoring than have been tested to date.

Given the limitations of the evidence base, we are more confident in making recommendations about research priorities than about effective policy. Indeed, this field is burdened by a lack of timely policy research and a tendency to launch major initiatives without first (or ever!) doing a high-quality evaluation. Note in this regard the various "get-tough" policies that have been encouraged by the federal government and adopted nationwide since the 1990s, the widespread use of SROs, or the School-Wide Positive Behavior Support package that has been adopted by 5,500 schools.

We have several recommendations to guide evaluation research on interventions. The first recommendation is to actually do such research, as suggested above. An impediment to learning about the effects of many school reforms is that the reforms tend to be implemented in all schools in the affected jurisdiction at once. This hinders rigorous evaluation because it leaves no schools in which to measure what would happen in the absence of the reform. A smarter approach would be to randomly assign schools to different phase-in periods, allowing for comparison during the first few years of the schools who implement the reform early and those who will implement it in the future.

Other recommendations are to measure effects on crime and other forms of misbehavior in evaluations conducted of interventions intended to improve academic performance, to capture the most serious forms of crime in evaluations rather than only less serious misbehavior, and to assess effects of prevention practices and policies on the entire student population rather than only on the students who are targeted.

Finally, it is important to identify programs to create more cohesive, communal, personalized environments. Many approaches to creating such environments seem plausible, but no rigorous research has yet established that such changes can be accomplished and that doing so results in a reduction in crime. This appears to be the next large challenge facing research on school-based prevention.

Notes and Acknowledgments This entry is based on a more detailed chapter (Cook et al. 2010) on school crime control and prevention. Portions of this work are also included in a forthcoming chapter entitled "School-Based Crime Prevention" in the *Oxford Handbook on Crime Prevention* (Brandon C Welsh and David P. Farrington, editors).

Related Entries

- ▶ Police Officers in Schools
- ▶ School Security Practices and Crime
- ▶ School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis

Recommended Reading and References

Bryk AS, Driscoll ME (1988) The high school as community: contextual influences and consequences for students and teachers. University of Wisconsin, National Center on Effective Secondary Schools, Madison

Cho H, Hallfors DD, Sánchez V (2005) Evaluation of a high school peer group intervention for at-risk youth. J Abnorm Child Psychol 33:363–374

Cook PJ, Gottfredson DC, Na C (2010) School crime control and prevention. In: Tonry M (ed) Crime and justice: a review of research, vol 39. University of Chicago Press, Chicago, Available athttp://papers.ssrn.com/sol3/papers.cfm?abstract_id=1368292

Council NR (2003) Deadly lessons: understanding lethal school violence. National Academy Press, Washington, DC

Cox DD (2005) Evidence-based intervention using home-school collaboration. School Psychol Quart 20:473–497

Eby LT, Allen TD, Evans SC, Ng T, DuBois DL (2007) Does mentoring matter? A multidisciplinary meta-analysis comparing mentored and non-mentored individuals. J Vocat Behav 72:254–267

Farrington DP, Ttofi MM (2009) Reducing school bullying: evidence-based implications for policy. In: Michael T (ed) Crime and justice: a review of research, vol 38. University of Chicago Press, Chicago

Foshee VA, Bauman KE, Arriaga XB, Helms RW, Koch GG, Linder GF (1998) An evaluation of safe dates, an adolescent dating violence prevention program. Am J Public Health 88:45–50

Gottfredson DC (1986) An empirical test of school-based environmental and individual interventions to reduce the risk of delinquent behavior. Criminology 24:705–731

Gottfredson DC (1987a) An evaluation of an organization development approach to reducing school disorder. Eval Rev 11:739–763

Gottfredson GD (1987b) Peer group interventions to reduce the risk of delinquent behavior: a selective review and a new evaluation. Criminology 25:671–714

Gottfredson GD, Gottfredson DC (1985) Victimization in schools. Plenum Press, New York

Gottfredson DC, Wilson DB, Najaka SS (2002) School-based crime prevention. In: Sherman LW, Farrington DP, Brandon Welsh C, MacKenzie DL (eds) Evidence-based crime prevention. Routledge, London

Gottfredson GD, Gottfredson DC, Payne AA, Gottfredson NC (2005) School climate predictors of school disorder: results from a national study of delinquency prevention in schools. J Res Crime Delinq 42:412–444

Herrera C, Grossman JB, Kauh TJ, Feldman AF, McMaken J, Jucovy LZ (2007) Making a difference in schools: the big brothers big sisters school-based mentoring impact study. Public/Private Ventures, Philadelphia

Karcher MJ (2008) The study of mentoring in the learning environment (SMILE): a randomized evaluation of the effectiveness of school-based mentoring. Prev Sci 9:99–113

Kellam SG, Hendricks Brown C, Poduska J, Ialongo N, Wang W, Toyinbo P, Petras H, Ford C, Windham A, Wilcox HC (2008) Effects of a universal classroom behavior management program in first and second grades on young adult behavioral, psychiatric, and social outcomes. Drug Alcohol Depend 95:S5–S28

Olweus D, Limber S, Mihalic S (1999) Blueprints for violence prevention: bullying prevention program. Center for the Study and Prevention of Violence, Boulder

Payne AA, Gottfredson DC, Gottfredson GD (2003) Schools as communities: the relationships among communal school organization, student bonding, and school disorder. Criminology 41:749–777

Schumaker JB, Hovell MF, Sherman JA (1977) An analysis of daily report cards and parent-managed privileges in the improvement of adolescents' classroom performance. J Appl Behav Anal 10:449–464

Sprague J, Walker H, Golly A, White K, Myers D, Shannon T (2001) Translating research into effective practice: the effects of a universal staff and student intervention on key indicators of school safety and discipline. Educ Treat Child 24:495–511

Tierney JP, Grossman JB, Resch NL (1995) Making a difference: an impact study of big brothers big sisters. Public/Private Ventures, Philadelphia

School Structural Characteristics and Crime

Philip J. Cook[1,2], Denise C. Gottfredson[3] and Chongmin Na[4]
[1]Sanford School of Public Policy, Duke University, Durham, NC, USA
[2]NBER, Cambridge, MA, USA
[3]Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA
[4]School of Human Science and Humanities, University of Houston-Clear Lake, Houston, TX, USA

Overview

Other essays in this encyclopedia summarize school-based prevention strategies that have been demonstrated to be effective for reducing crime and discuss the importance for schools of having close-knit communities where rules of acceptable behavior are clearly communicated and consistently (if not harshly) enforced. Schools' capacity to implement effective prevention strategies and to establish well-ordered communities is likely to be greater in schools that are not overwhelmed by having a high proportion of the students at risk. In this essay, we discuss decisions primarily made outside of the school building that influence the demographic composition of schools and the number and types of other students to whom a child is exposed. Externally determined factors (e.g., finances, physical features of building, school size, and student/teacher ratio) also determine resources available in the school and define patterns of interaction broadly speaking. Here we will focus on school-district policies influencing the student composition of schools and school size.

Specifically, we will consider the extent to which school or school-district decisions regarding how students are organized for instruction (e.g., academic or behavioral tracking or departmentalization) further narrow the characteristics of other students to whom youths will be exposed. Importantly, these decisions determine the pool of youths from which highly influential peers will be selected as well as the dominant peer culture in the school. These structural characteristics potentially influence crime-producing mechanisms.

We find that the concentration of different types of students in a school has important implications for the amount of crime in the school: The grade levels included in the school or average age of the students in the school, the percentage male students, the social class composition, and the racial and ethnic composition of the schools are related to measures of problem behavior. Further, some policies that alter the composition of classes, grades, or schools have been shown to reduce problem behavior: Keeping sixth graders in elementary schools as opposed to moving them to middle schools reduces disciplinary infractions, and retaining students in grade increases conduct problems both for the "old-for-grade" students and for non-retained students. We find little support for the idea that the size of the school per se matters for reducing school crime.

We begin this essay with a summary of research related to school size and student demographics as they relate to school crime and disorder. We then discuss school policies and practices that are likely to alter the types of students with whom a student is likely to be exposed. We end with a discussion of implications for policy and research.

School Ecology

One category of school structural characteristics is "school ecology." Very few studies have examined the influence of several potentially important aspects of school ecology including school finances and the physical features of the school building. By contrast, many studies include a measure of school size – number of students in the school or in the grade. Our discussion will therefore focus on school size, providing a summary of the literature and some new results.

School size is thought to have a major influence on the internal organization of schools and on subsequent student outcomes. Lee et al. (1993) suggest that larger schools are likely to have increased capacity to tailor programs and services to meet the diverse needs of students in the school. The extreme example of low specialization is a one-room schoolhouse in which one teacher teaches all students all day. In small schools, the typical teacher teaches a smaller number of different students and gets to know these students well. Students in such schools may develop a greater sense of trust in the adults and be more likely to communicate potentially dangerous situations to them. Large schools are likely to be organized more bureaucratically and to involve more formalized social interactions among members of the school population. As a result, communication may be less frequent or less direct, cohesiveness may be reduced, management functions (including the management of discipline) may become less nuanced, and individuals may share less of a common experience in the school. Alienation, isolation, and disengagement may result. All of these mechanisms are plausible but speculative.

As it turns out, school size has not received much focused attention in research on schools and crime. However, many studies have included a measure of school size as a control variable when focusing on the effects of other aspects of school climate. Cook et al. (2010) summarized the associations between measures of school size and problem behavior in school-level and multilevel studies. They found nine school-level studies based on data from seven different data sources. In the studies that reported an unambiguous association between school size and a measure of problem behavior, the conclusions differed depending among other things on the measure of problem behavior used. Positive associations between school size and measures of minor misbehavior were reported for the High School and Beyond high school data and the National Education Longitudinal Study [NELS] eighth graders, but the association with more serious forms of misbehavior was not statistically significant. In another data source (Safe School Study), school size was not significantly related to student victimization but was positively related to teacher victimization. That study also showed that the effect of school size on teacher victimization is mediated by aspects of the school social organization and culture. No significant relationship with school size was found in a study of middle schools in Philadelphia.

The multilevel studies summarized provided no support for the "smaller is better" viewpoint. Fifteen different research reports based on nine different data sources were summarized. In these studies, which generally controlled for community characteristics as well as characteristics of the students who attend the school, only one data source (NELS tenth graders, as reported in Stewart 2003) produced a significant positive association between school size and a measure of problem behavior, and the measure of problem behavior used in this study was unusual because it contained mainly school responses to misbehavior (e.g., being suspended or put on probation) rather that actual youth behavior. Hoffmann and Dufur (2008) also reported on the association of school size and a broader measure of problem behaviors including substance use, arrest, and running away using the NELS tenth grade sample and found no significant association. Reports from a sample of Israeli schools containing seventh and 11th grades documented a positive association between average *class* size and student victimization, but no significant association with school size. One of the multilevel studies reported a significant negative association between school size and student victimization, but this sample included only rural schools located in New Brunswick, Canada, whose average size was 39 and 53 students, respectively, for sixth and eighth grade. A recent study by Gottfredson and DiPietro (2011) found that, net of individual-level risk factors and confounding characteristics of schools and their surrounding communities, more students per teachers and a higher number of different students taught by the average teacher were related to higher student victimization rates, but larger school size was significantly related to lower student victimization rates. Cook et al. (2010) concluded that most of the multilevel studies suggest that school size is not reliably related to student problem behavior once characteristics of the students who attend the schools are controlled.

However, the studies summarized often reported on the association between school size and problem behavior from models that may provide too conservative a test. The student characteristics that are controlled in the multilevel studies are often exactly those student characteristics that Lee et al. (1993) hypothesized to be influenced by school size (e.g., school attachment, involvement, perceived positive social climate). Also, most of the associations with school size reported in the studies are from models that partial out influences not only of the communities in which the schools are located and the average demographic characteristics of the students attending the schools but also of other school climate characteristics such as school culture and the administration/management of discipline, hypothesized to mediate the influence of school size on student outcomes. For example, Hoffmann and Dufur's (2008) study reported a negative association between school size and delinquency in the NELS data, but the equation also contained measures of "school quality," a composite measure assessing youth perceptions of their school as fair and their teachers and fellow students as caring and trustworthy. Unfortunately, many of these reports did not report the association of school size with problem behavior in models that do not control for potential effects of school size.

New results. We analyzed data from the 2007–2008 School Survey on Crime and Safety

(SSCS; Neiman and DeVoe 2009) in an attempt to establish baseline descriptive results on how school size relates to school crime. Principals were asked how many incidents of various types of crime had occurred at school during the last school year. They were asked about violent crimes (rape, sexual battery other than rape, physical attack or fight with or without a weapon, threat of physical attack with or without a weapon, and robbery with or without a weapon) and about theft and larceny. We calculated rates per 1,000 students for each school. Because school size is highly related to location and level, it is necessary to look at the association of crime rates and enrollment while controlling for these factors.

Figure 1 shows median rates per 1,000 students for theft/larceny and violent crimes, according to school principals. The figures make clear that crime rates are not systematically related to school size within level and location. However, there is some suggestion that the association between size and principal reports of crime differs according to type of crime, level, and location: In urban locations, principals in smaller elementary and middle schools report more violent crimes. This is not the case in rural/suburban schools or in urban high schools.

We conclude that school size is not generally related to principal reports of school and that whatever differences observed favor larger schools over smaller schools.

It is likely that the ratio of adults to students rather than the actual number of students in the school is related to problem behavior. Five of the studies summarized in Cook et al. (2010) looked at the association of problem behavior to student/teacher ratio. Only one of the five studies reported a significant relationship: Gottfredson and DiPietro (2011), in a multilevel study using data from the National Study of Delinquency Prevention in Schools, reported a positive association of student/teacher ratio and personal but not property crime victimization. That study reported that higher levels of social capital, as measured by student consensus about normative beliefs, partially mediated the effects of student/teacher ratio on personal victimization. We believe that a more sensitive measure of adult presence would be the ratio of all adults (rather than just teachers) to students. Many schools use parent volunteers and teacher aides in addition to teachers to help maintain order. The ratio of the total number of adults to students would reflect variability in the use of such auxiliaries. Unfortunately, no studies have reported on this association.

As far as we know, there are no intervention-based studies of how school size affects school crime. Case studies of instances in which an established school is divided in smaller units are available, but they almost never assess effects on crime, and they do not provide a clean test of the effects of changing school size because other factors (such as the curriculum, aspects of the physical space, and school finances) are always altered simultaneously.

Milieu

The SSCS data show that rates of problem behavior differ with demographic characteristics. Middle schools have higher rates of delinquency than elementary or high schools (the partial exception is substance use, which increases through high school). Schools with 50 % or more minority enrollment experience higher rates of violence than majority white schools (Nolle et al. 2007). Socioeconomic status of the student body is also associated with delinquency rates (Gottfredson et al. 2005).

More interesting from a policy perspective is the extent to which the mix of students in the school or the classroom influences the likelihood that any given student will misbehave. The mechanisms of deviant peer influence are both direct and indirect. The direct effects may arise as a result of deviant peer influence: learning and imitation, social reinforcement for deviant acts, and the creation of opportunities for deviant activities (Dishion and Dodge 2006). It may also be due to the presence of social norms that support (or at least appear to youth to support) delinquent behavior. All these mechanisms are

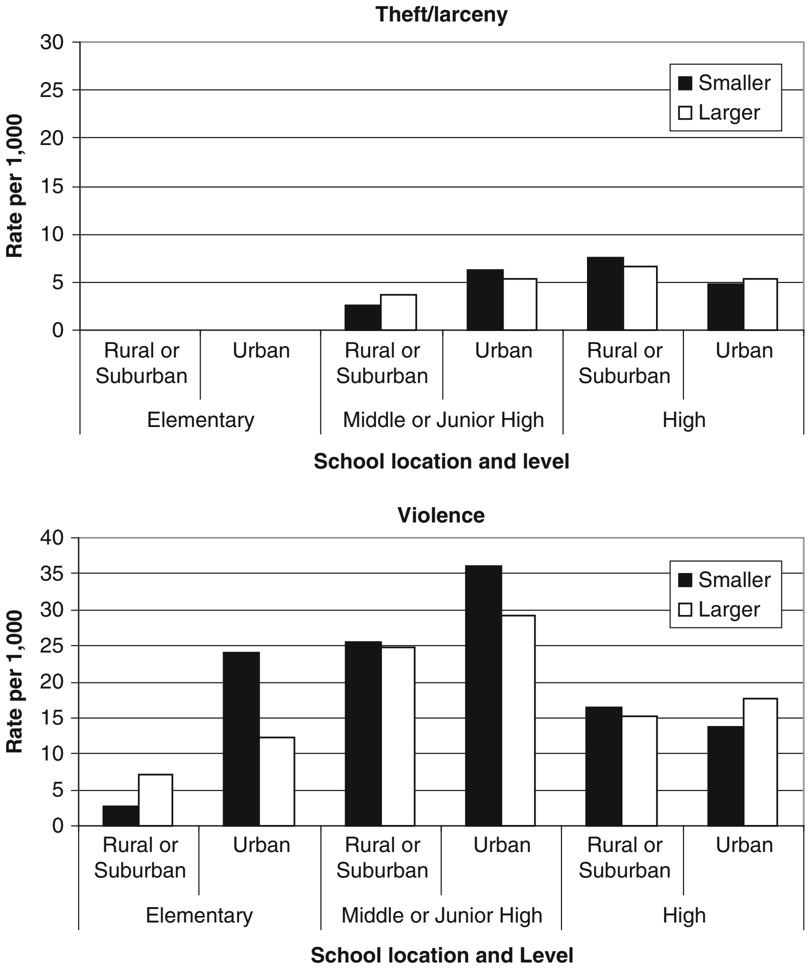

School Structural Characteristics and Crime, Fig. 1 Principal reports of crimes recorded by school, by school enrollment, location, and level – median rates per 1,000 students. Note: Size is split at the median total enrollment for each school level: 452, 585, and 885 for elementary, middle, and high, respectively. One outlier (a small alternative school for delinquent youths with a very high crime rate) is excluded. Violent crimes include rapes, sexual batteries other than rape, robberies with or without a weapon, physical attacks or fights with or without a weapon, and threats of physical attack with or without a weapon. Theft includes taking things worth over $10 without personal confrontation and includes pocket picking, stealing a purse or backpack (if left unattended or no force was used to take it from owner), theft from a building, theft from a motor vehicle or [of] motor vehicle parts or accessories, theft of a bicycle, theft from a vending machine, and all other types of thefts (Source: Original tabulation from data files. SSCS 2007–2008)

relevant for involvement with delinquency both in and out of school, including drugs and alcohol, and participation in gangs. The indirect effects may come about as a result of the dilution of authority – a teacher who can manage one or two disruptive students may lose control of the classroom when there are more than two. The same phenomenon can occur at the school level, where a high "load" of troublesome students may swamp the mechanisms of control in the corridors, cafeteria, lavatories, and grounds.

Given the real possibility of peer influence, the actual behavior of youths with a given propensity to deviant or criminal activity may well depend on whom they encounter in their classes and in the other locations in the school. A variety of policies are relevant to influencing the mix of students. At the level of the school district, the distribution of students among schools will be influenced by which grade spans are included in the middle schools, the extent to which low-performing students are held back, and whether school assignments are tied largely to place of residence or tailored to promote integration or parental choice. For a given pattern of assignments to schools, the number and characteristics of students who are actually in the building on a school day will depend on absenteeism and use of out-of-school suspension. And for a given population of students who are actually attending the school on any given day, social influence will likely be mediated by policies that influence the extent to which deviant students are concentrated, such as in-school suspension or academic tracking.

Cook et al. (2010) summarized the results from 18 multilevel studies based on 14 datasets showing associations between the milieu of the school and measures of problem behavior, controlling for individual-level demographics as well as for related characteristics of the communities from which the student body is drawn. These studies showed that the grade levels included in the school or average age of the students in the school, the percentage male of students, the social class composition, and the racial and ethnic composition of the schools are related to measures of problem behavior. These associations sometimes did not reach statistical significance, but they were nearly always in the expected direction. Here we discuss several of the strongest studies relevant to evaluating the impact of policy choices concerning grade span, grade retention, truancy prevention, racial segregation, and use of alternative schools.

1. *Grade span.* One recent study demonstrates that the grade composition in a middle school influences the rates of misbehavior of the students. A generation ago, most elementary schools included sixth grade, but now most sixth graders attend middle school. Using a quasi-experimental approach, Cook et al. (2008) compare the school records of North Carolina students whose sixth grade was located in a middle school with those whose sixth grade was in elementary school (the sample of sixth grades in middle school was trimmed to match the sixth grades in elementary school in several dimensions). While the two groups of students had similar infraction rates in fourth and fifth grades, those who moved to a middle school for sixth grade experienced a sharp increase in disciplinary infractions relative to those who stayed in elementary school. More interesting, perhaps, is that the elevated infraction rate persisted through ninth grade. A plausible interpretation of these findings is that sixth graders are at a highly impressionable age, and if placed with older adolescents, they tend to be heavily influenced by their inclination to break the rules. This example of negative peer influence is quite large and extends to all types of infractions, including violence and drug violations.
2. *Retention policies.* The age mix in a school is closely related to the grade span, but that is not the only determinant; the school district's retention policies also play a role. In response to high-stakes accountability programs (including No Child Left Behind), many school systems have ended social promotion for students who fail end-of-grade tests, thus increasing the number of old-for-grade students. Entry-level at-risk students are often held back for a year before making the transition to second grade. The effect of retention on behavior of the retained students has been extensively studied. Most studies have focused on academic outcomes: Meta-analyses of this literature conclude that the long-term effect on academic achievement is null or negative, with a greatly elevated risk of dropping out (Jimerson et al. 2006). Hence, given the robust general finding that students with academic difficulties are more prone to antisocial behavior (Nagin et al. 2003), it is not surprising that grade retention appears to increase conduct problems. One of the most sophisticated studies, using Richard Tremblay's longitudinal data on Montreal

school children, found that the effect of grade retention on classroom physical aggression (as measured by teacher reports) is conditioned by the developmental history of the child: Those showing no aggression or chronic aggression levels were not affected, whereas those whose trajectory of aggression was declining over time increased their aggression more so if retained than if not retained (Nagin et al. 2003).

There has been less attention to the contextual effect of having old-for-grade students in the classroom and school. One exception is a recent study that uses a comprehensive data set of North Carolina students. Muschkin et al. (2008) conduct a cross-sectional analysis of infraction rates by seventh graders, finding that the prevalence and incidence of infractions increase with the prevalence of retained students (students who were retained at least once in the previous 3 years) and the prevalence of old-for-grade students who were not retained during that 3-year period. These results hold after they controlled for various characteristics of the student body and the schools and the inclusion of district fixed effects. The authors also find evidence that susceptibility differs among types of students; in particular, the old-for-grade seventh graders were themselves especially susceptible to the influence of the concentration of other old-for-grade students in their school. Similar results were found when the outcome variable was the likelihood of being suspended.

3. *Truancy prevention*. The mix of students who are in the school building on any given day will be affected by absenteeism and tardiness. School attendance laws require that youths between specified ages (e.g., 7–16 in North Carolina and 5–18 in New Mexico) attend school, with possible exceptions for home schooling. This is a legal obligation for which both the child and parents are liable. In many school districts, however, these laws are widely flouted. For example, the absentee rate in DC public high schools in the 2006–2007 school year averaged 17 %.

   The rate of unexcused absence determines not only the number of students in the school building but also the behavioral propensities of those students. Chronic truants are not a representative sample of the student body but rather tend to come from dysfunctional families and be at risk for delinquency, violence, and substance abuse (McCluskey et al. 2004). It is also true that chronic truancy engenders academic problems and is associated with failure to graduate from high school and a variety of poor life outcomes, including involvement with serious crime. As a logical matter, then, programs that are effective in improving attendance rates may have several effects. First, if they get delinquent youths off the street and into school, the result may be reduced crime rates in the community. Indeed, communities concerned about the daytime crimes committed by truants have increasingly enlisted the police and the juvenile court to combat truancy. Second, effective truancy-prevention programs may come at the cost of higher crime rates within the school. And third, if at-risk youths are persuaded to attend school more faithfully, the long-term result may be to improve their chances of graduation and subsequent success.

   A number of school-based programs have been evaluated in part by their effect on school attendance. Unfortunately, there are no studies, insofar as we know, that evaluate the effect of attendance-promoting programs on school crime rates or overall (community plus school) crime rates.

4. *School desegregation*. In compliance with the 1954 Supreme Court ruling *Brown v. Board of Education*, federal courts issued a series of desegregation orders to public school districts. These orders forced a considerable increase in the extent to which African American students attended school with whites during the 1960s and 1970s. A vast literature on the effects of school segregation and desegregation has focused on academic outcomes. The results of this research offer support to the conclusion that integrated schools promote

black achievement and increase black high school graduation rates, college attendance and graduation rates, and occupational success (LaFree and Arum 2006). A persuasive quasi-experimental study of the effect of desegregation plans found that they reduced black dropout rates by 2–3 % points, with no detectable effect on whites (Guryan 2004). Given the tight link between academic success and school behavior, it is entirely plausible that the degree of segregation has a direct influence on delinquency in schools. But we are not aware of any direct evidence on the subject; segregation studies have not used school crime as an outcome variable. There have been two persuasive studies concerning the effects of segregation on crime outside of school. LaFree and Arum (2006) analyzed the incarceration rates for black males who moved to a different state following school. For any given destination state, they found that those who moved from a state with well-integrated schools had a substantially lower incarceration rate than those who moved from a state where the schools were more segregated. A more recent study (Weiner et al. 2009) utilizes a quasi-experimental approach in which the court desegregation orders serve as the experimental intervention: They report that these orders reduced black and white homicide victimization rates for 15–19-year-olds. The authors explore several mechanisms that may account for this result, including both the direct effects of changing the mix of students in the schools and indirect effects associated with police spending and relocation of some white students. In any event, since all but a handful of these homicides occurred outside of school, we are still waiting for direct evidence on crime in schools.

5. *Alternative schools.* A recent survey found that 39 % of public school districts administered at least one alternative school for students at risk of educational failure. As of October 2000, 613,000 students were enrolled in these schools, 1.3 % of nationwide total enrollment. Urban districts, large districts (those with 10,000 or more students), districts in the Southeast, districts with high minority student enrollments, and districts with high poverty concentrations were more likely than other districts to have alternative schools and programs for at-risk students. Despite the widespread use of these schools as a means of removing antisocial and violent students from the regular classrooms, there have been no systematic studies of the effects on school crime rates. The effects on the behavior of youths who are given alternative-school placements have been studied, with mixed results. Indeed, there is unlikely to be any generic answer, since effects will depend on the quality of the programming and on which students are selected. Best-practice judgments tend to rely on expert opinion rather than on evaluation studies with strong designs.

6. *Grouping within schools.* Academic tracking is nearly universal in US secondary education. The attraction of separating students into tracks that are more or less demanding academically is the belief that this is the best way to tailor coursework to the differing background, ability, and motivation of the students. Tracking tends to have the result of concentrating minorities and students from lower socioeconomic status households in certain classrooms. Given the strong association between academic success and delinquency involvement, it also has the effect of concentrating crime-prone students, setting the stage for negative peer influence.

As a device to improve academic progress, tracking has had more detractors than advocates among education specialists. The evidence base is thin: Most notably, Mosteller et al. (1996) identified only ten randomized experimental evaluations comparing the performance of students in tracked (homogeneous) and untracked (heterogeneous) classrooms. When these studies were combined, the best estimate was a zero difference in average academic performance. Still thinner is any evidence on how tracking affects misbehavior. Thus, we conclude that the possibility of deviant peer influence due to tracking is plausible but unproven.

Conclusion

School reform is typically shaped by theories of how to improve students' academic performance. But to the extent that school safety is an important goal, somewhat distinct from academic progress, the potential impacts on safety should be considered in any evaluation.

One of the most prominent reform efforts since 2001 has been the campaign funded largely by the Gates Foundation to create small high schools. While that effort was abandoned in 2008 as a result of disappointing results on academic progress, it would also be of interest to know the effect on school crime and juvenile delinquency. Given the fact, reported here, that small schools are not systematically safer than large schools (controlling for urbanicity and grade level), it appears doubtful that smaller is better in this domain.

There is very good reason to believe that the mix of students who are assembled in a school or any one classroom may influence the behavior of all. Two relevant mechanisms are deviant peer influence and "resource swamping," both implying that overall crime rates within school may increase in nonlinear fashion with the addition of deviant students to the mix (Cook and Ludwig 2006). This concern is relevant in evaluating policies regarding grade retention, truancy prevention, use of suspension and expulsion, use of alternative high schools, and even academic tracking. In each case, however, we found that relevant evaluations were lacking.

Implications for Research and Policy

Crime rates differ widely among schools. Experimental evidence suggests that the crime involvement of any given student at risk is influenced by the school that he or she attends. That fact motivates our scientific quest to find the school structural determinants of criminal activity by school-aged youths.

In another entry in this volume, we summarized findings from studies of school climate and crime and discussed numerous additional studies that reported on attempts to manipulate aspects of school climate. The school climate studies revealed sturdy associations between measures of school climate and measures of student delinquency, victimization, substance use, or other forms of problem behavior.

A starting point in accounting for interschool differences in crime is the criminal propensities of the students. Schools in which many of the students are active delinquents outside school start with a far greater challenge than those in which the students are largely law-abiding. The school crime rate of a student body with high crime propensity may be greater than the sum of the parts for two reasons. First, if the school lacks the adult resources to manage the "load" of misbehavior, then the school may become progressively more chaotic, spinning out of control. Further, delinquent and deviant youths may have a negative influence on each other and other students as well, further amplifying the problem. As noted in earlier discussion of the mechanisms involved in the production of crime and the features of schools that might influence these mechanisms, decisions that influence the demographic composition of schools are important because they determine the prevailing cultural beliefs in the school as well as the pool of youths from whom friends can be selected. In short, the crime rate in school is not just the sum of the parts but does reflect the ecological effects of the mix of students in the building.

Schools and school districts have a good deal of control over the makeup of the student body. Schools can be based on neighborhood residential patterns or integrated across race and class. The grade span for elementary and middle schools can be adjusted. Truancy and dropout prevention programs can be pursued with more or less vigor and troublesome students reassigned. Whether failing students are retained in a grade or given a social promotion influences the extent of age homogeneity within classrooms. Students who are enrolled in the school can be tracked on the basis of academic potential or mixed together and so forth. This array of policy choices has the potential to influence the "load" on teachers and other adults and the opportunity for deviant peer influence. Some of these policies have been evaluated for these ecological effects, but the evidence base is quite thin.

We also note that most of the evaluations of policies that affect the mix of students – truancy and dropout prevention, alternative schools, tracking, grade retention of failing students, and so forth – only consider the effect on the students who are targeted and fail to consider the ecological effects. But secondary effects on other students may be quite important and should be included when it is possible to implement a comprehensive study.

Finally, we noted that evidence does not support the conclusion that smaller schools are more effective for limiting problem behaviors than larger schools, but research reported in the "School Social Organization, Discipline Management, and Crime" entry of this volume suggests that conditions that make a school environment "feel" smaller and more communally organized are related to levels of problem behavior.

Notes and Acknowledgments This essay is based on a more detailed chapter (Cook et al. 2010) on school crime control and prevention.

Related Entries

- ▶ School Social Organization, Discipline Management, and Crime
- ▶ School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis

References

Carrell SE, Hoekstra ML (2008) Externalities in the classroom: how children exposed to domestic violence affect everyone's kids. National Bureau of Economic Research (NBER) working paper no.14246. Available at http://www.nber.org

Cook PJ, Gottfredson DC, Na C (2010) School crime control and prevention. In: Tonry M (ed) Crime and justice: a review of research. The University of Chicago Press, Chicago

Cook PJ, Ludwig J (2006) Assigning youths to minimize total harm. In: Dodge KA, Dishion TJ, Lansford JE (eds) Deviant peer influences in programs for youth: problems and solutions. The Guildford Press, New York, pp 122–140

Cook PJ, MacCoun R, Muschkin C, Vigdor J (2008) The negative impact of starting middle school in sixth grade. J Policy Anal Manage 27:104–121

Dishion TJ, Dodge KA (2006) Deviant peer contagion in interventions and programs: an ecological framework for understanding influence mechanisms. In: Dodge KA, Dishion TJ, Lansford JE (eds) Deviant peer influences in programs for youth: problems and solutions. The Guilford Press, New York, pp 14–43

Gottfredson DC, DiPietro SM (2011) School size, social capital, and student victimization. Sociol Educ 84(1):69–89

Gottfredson GD, Gottfredson DC, Payne AA, Gottfredson NC (2005) School climate predictors of school disorder: results from a national study of delinquency prevention in schools. J Res Crime Delinq 42:412–444

Guryan J (2004) Desegregation and black dropout rates. Am Econ Rev 94:919–943

Hoffmann JP, Dufur MJ (2008) Family and school capital effects on delinquency: substitutes or complements? Sociol Perspect 51:29–62

Jimerson SR, Pletcher SMW, Graydon K, Schnurr BL, Nickerson AB, Kundert DK (2006) Beyond grade retention and social promotion: promoting the social and academic competence of students. Psychol Sch 43:85–97

LaFree G, Arum R (2006) The impact of racially segregated schooling on the risk of adult incarceration rates among U.S. cohorts of African-Americans and whites since 1930. Criminology 44:73–103

Lee VE, Bryk AS, Smith JB (1993) The organization of effective secondary schools. In: Darling-Hammond L (ed) Review of research in education. American Educational Research Association, Washington, DC, pp 171–267

McCluskey CP, Bynum TS, Patchin JW (2004) Reducing chronic absenteeism: an assessment of an early truancy initiative. Crime Delinq 50:214–234

Mosteller F, Light RJ, Sachs JA (1996) Sustained inquiry in education: lessons from skill grouping and class size. Harv Educ Rev 66:797–842

Muschkin C, Glennie B, Beck A (2008) Effects of school peers on student behavior: age, grade retention, and disciplinary infractions in middle school. J Policy Anal Manage 27:104–121

Nagin DS, Pagani L, Tremblay RE, Vitaro F (2003) Life course turning points: the effect of grade retention on physical aggression. Dev Psychopathol 15:345

Neiman S, DeVoe JF (2009) Crime, violence, discipline, and safety in U.S. public schools: findings from the school survey on crime and safety: 2007–08 (NCES 2009-326). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC

Nolle KL, Guerino P, Dinkes R (2007) Crime, violence, discipline, and safety in US public schools: findings from the school survey on crime and safety: 2005-6 (NCES 2007-361). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC

Stewart EA (2003) School social bonds, school climate, and school misbehavior: a multilevel analysis. Justice Q 20:575–604

Weiner DA, Lutz BF, Ludwig J (2009) The effects of school desegregation on crime. National Bureau of Economic Research (NBER) working paper no.15380. Available at http://www.nber.org

School Violence Prevention

► School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis

School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis

Sandra Jo Wilson
Peabody Research Institute, Vanderbilt
University, Nashville, TN, USA

Synonyms

School violence prevention

Overview

Schools are an important location for interventions to prevent or reduce aggressive behavior. They are a setting in which much interpersonal aggression among children occurs and the only setting with almost universal access to children. There are many prevention strategies from which school administrators can and do choose, including surveillance (e.g., metal detectors, security guards), deterrence (e.g., disciplinary rules, zero tolerance policies), and psychosocial programs. Indeed, over 75 % of schools in one national sample reported using one or more of these prevention strategies to deal with behavior problems (Gottfredson et al. 2000). Other reports similarly indicate that more than three-fourths of schools offer mental health, social service, and prevention service options for students and their families (Brener et al. 2001). Among psychosocial prevention strategies, a broad array of programs can be implemented in schools. These include packaged curricula and homegrown programs for use school-wide and others that target selected children already showing behavior problems or deemed at risk for such problems. Most psychosocial prevention programs address a range of social and emotional factors assumed to cause aggressive behavior or to be instrumental in controlling it (e.g., social skills or emotional self-regulation), and it is these psychosocial programs that are the focus of this meta-analysis.

Key Issues

Various resources are available to help schools identify programs with proven effectiveness. Among these resources are the Blueprints for Violence Prevention; the Collaborative for Academic, Social, and Emotional Learning (CASEL); and the National Registry of Evidence-based Programs and Practices (NREPP) administered by the Substance Abuse and Mental Health Services Administration (SAMSHA). There is, however, little indication that the evidence-based programs promoted to schools through such sources have been widely adopted or that, when adopted, they are implemented with fidelity (Gottfredson and Gottfredson 2002).

While lists of evidence-based programs can provide useful guidance to schools about interventions likely to be effective in their settings, they are limited by their orientation to distinct program models and the relatively few studies typically available for each such program. A meta-analysis, by contrast, can encompass virtually all credible studies of such interventions and yield evidence about generic intervention approaches as well as distinct program models. Perhaps most important, it can illuminate the features that characterize the most effective programs and the kinds of students who benefit most. Since many schools already have prevention programs in place, a meta-analysis that identifies characteristics of successful prevention programs can inform schools about ways they might improve those programs or better direct them to the students for whom they are likely to be most effective. Thus, the purpose of the meta-analysis reported here is to investigate which program and student characteristics are associated with the most effective treatments.

In 2003, we published a meta-analysis on the effects of school-based psychosocial interventions for reducing aggressive and disruptive behavior aimed at identifying the characteristics of the most effective programs (Wilson et al. 2003). That meta-analysis included 172 experimental and quasi-experimental studies of intervention programs, most of which were conducted as research or demonstration projects with significant researcher involvement in program implementation. Though not necessarily representative of routine practice in schools, those programs showed significant potential for reducing aggressive and disruptive behavior, especially for students whose baseline levels of antisocial behavior were already high. Different intervention approaches appeared equally effective, but significantly larger reductions in aggressive and disruptive behavior were produced by those programs with better implementation, that is, more complete delivery of the intended intervention to the intended recipients. In 2007, we published an updated version of that meta-analysis that included 249 studies of school-based prevention programs (Wilson and Lipsey 2007). For that analysis, we separated the programs into four mutually exclusive groups, characterized by the general service format and target population of the programs. The four program groups were universal programs, pullout programs for high-risk populations, comprehensive programs, and special school programs. In addition to being distinctive in terms of service format, programs in each of the groups had a number of methodological, subject, and dosage differences that made it unwise to combine them in a single analysis. Furthermore, school decision-makers typically make choices from within a format category, rather than across them. Thus, there was little utility in lumping all the school-based prevention programs into one larger set.

Results of the 2007 meta-analysis found positive overall intervention effects on aggressive and disruptive behavior and other relevant outcomes. The most common and most effective approaches were universal programs and pullout programs targeted for high-risk children. Comprehensive programs did not show significant effects and those for special schools or classrooms were marginal. Different program approaches (e.g., behavioral, cognitive, social skills) produced largely similar effects. Effects were larger for better implemented programs and for those involving students at higher risk for aggressive behavior.

Since the publication of our earlier work, the analytic methods for handling dependent effect size estimates in meta-analysis have improved. Multiple effect size estimates from a single study sample can now be included in a meta-analysis using robust standard errors that account for the statistical dependencies present when multiple effect size estimates are used (Hedges et al. 2010). This entry provides an opportunity to take the school-based prevention programs included in the 2003 and 2007 papers and reanalyze their results using these new statistical methods.

Method

Criteria for Including Studies in the Meta-Analysis

Studies were selected for this meta-analysis based on a set of detailed criteria, summarized as follows:

1. The study was reported in English no earlier than 1950 and involved a school-based program for children attending any grade, prekindergarten through 12th grade.

2. The study assessed intervention effects on at least one outcome variable that represented either (a) aggressive or violent behavior (e.g., fighting, bullying, person crimes), (b) disruptive behavior (e.g., classroom disruption, conduct disorder, acting out), or (c) both aggressive and disruptive behavior.
3. The study used an experimental or quasi-experimental design that compared students exposed to one or more identifiable intervention conditions with one or more comparison conditions on at least one qualifying outcome variable. To qualify as an experimental or quasi-experimental design, a study was required to meet at least one of the following criteria:
   - Students or classrooms were randomly assigned to conditions.
   - Students in the intervention and comparison conditions were matched and the matching variables included a pretest for at least one qualifying outcome variable or a close proxy.
   - If students or classrooms were not randomly assigned or matched, the study reported both pretest and posttest values on at least one qualifying outcome variable or sufficient demographic information to describe the initial equivalence of the intervention and comparison groups.

Search and Retrieval of Studies

An attempt was made to identify and retrieve the entire population of published and unpublished studies that met the inclusion criteria summarized above. The primary source of studies was a comprehensive search of bibliographic databases, including PsycINFO, Dissertation Abstracts International, ERIC (Education Resources Information Center), US Government Printing Office publications, National Criminal Justice Reference Service, and MEDLINE. Second, the bibliographies of meta-analyses and literature reviews on similar topics were reviewed for eligible studies. Finally, the bibliographies of retrieved studies were themselves examined for candidate studies. Identified studies were retrieved from the library, obtained via interlibrary loan, or requested directly from the author. We obtained and screened more than 95 % of the reports identified as potentially eligible through these sources. Note that a new search was not performed for this entry and that the literature reviewed here is current through 2004.

Coding of Study Reports

Study findings were coded to represent the mean difference in aggressive behavior between experimental conditions at the posttest measurement. The effect size statistic used for these purposes was the standardized mean difference, defined as the difference between the treatment and control group means on an outcome variable divided by their pooled standard deviation (Lipsey and Wilson 2001). In addition to effect size values, information was coded for each study that described the methods and procedures, the intervention, and the student samples. Coding reliability was determined from a sample of approximately 10 % of the studies that were randomly selected and recoded by a different coder. For categorical items, intercoder agreement ranged from 73 % to 100 %. For continuous items, the intercoder correlations ranged from 0.76 to 0.99. A copy of the full coding protocol is available from the author.

General Analytic Procedures

All effect sizes were multiplied by the small sample correction factor, $1-(3/4n-9)$, where n is the total sample size for the study, and each effect size was weighted by its inverse variance in all computations (Lipsey and Wilson 2001). The inverse variance weights were computed using the subject-level sample size for each effect size. Because many of the studies used groups (e.g., classrooms, schools) as the unit of assignment to intervention and control conditions, they involved a design effect associated with the clustering of students within classrooms or schools that reduces the effective sample size. The respective study reports provided no basis for estimating those design effects or adjusting the inverse variance weights for them, so cluster adjustments were not made in the analyses reported here. This should not greatly affect the effect sizes estimates or the magnitude of their relationships to moderator variables but does assign them somewhat smaller standard error estimates and, hence, larger inverse variance weights than is technically correct. A dummy code identifying these cases was included in the analyses to reveal any differences in findings from these studies relative to those using students as the unit of assignment. The effect sizes for cluster assigned studies were not statistically different from the student-assigned studies in any analysis.

Examination of the effect size distributions for each of the four program format groups identified a small number of outliers with potential to distort the analysis. Outliers were defined as values that fell more than three interquartile ranges (IQR) above the 75th percentile or below the 25th percentile of the effect size distribution. Outliers identified using this procedure were Winsorized to the next closest value. In addition, several studies used unusually large samples. Because the inverse variance weights chiefly reflect sample size, those few studies would dominate any analysis in which they were included. Therefore, the extreme tail of the sample size distribution was Winsorized using the procedure described above, and the inverse variance weights were recomputed for those effect sizes. These adjustments allowed us to retain outliers in the analysis, but with less extreme values that would not exercise undue influence on the analysis results.

In the original publications (Wilson et al. 2003; Wilson and Lipsey 2007) only one effect size from each subject sample was used in the analyses to maintain independence of the effect size estimates. When more than one was available, the effect size from the measurement source most frequently represented across all studies (e.g., teachers' reports, self-reports) was selected. We wanted to retain informant as a variable for analysis, so we did not elect to average across effect sizes from different informants; if there was more than one effect size from the same informant or source, however, their mean value was used. However, techniques for computing robust standard errors for dependent effect sizes have since been developed (Hedges et al. 2010) and are employed in the analyses reported below. Thus, when a study reported intervention effects for more than one aggressive or disruptive behavior outcome, all effect sizes were used in the analyses. Dummy codes indicating the type of aggressive behavior measured and/or the informant or source of the outcome measure were examined for their influence on the effect sizes in the analyses.

Finally, many studies provided data sufficient for calculating mean difference effect sizes on the outcome variables at the pretest. In such cases, we adjusted the posttest effect size by subtracting the pretest effect size value. This information was included in the analyses presented below to test whether there were systematic differences between effect sizes adjusted in this way and those that were not.

A small number of studies were missing data on the method, participant, or program variables used in the final analyses; missing values were imputed using an expectation-maximization (EM) algorithm in SPSS.

Analysis of the effect sizes was conducted separately for each program format (described below) and done in several stages. We first produced the random effects mean effect size for each group of effect sizes and examined the homogeneity of the effect size distributions using the Q-statistic (Hedges and Olkin 1985). Moderator analyses were then performed to identify the characteristics of the most effective programs using weighted mixed effects multiple regression with the aggressive/disruptive behavior effect size as the dependent variable. When the number of effect sizes and studies was sufficient, three models were produced in order to examine the added contribution of different sets of variables. In the first stage of these analyses, we examined the influence of study methods on the effect sizes. Influential method variables were carried forward as control variables for the next stage of analysis, which examined the relationships between student characteristics and effect size. The third stage in these analyses involved adding important treatment characteristics to the models. Because a large number of potential moderators were available from our coding manual, moderators were selected based on the size of their weighted zero-order correlations with the effect size while being mindful of collinearity among the different moderator variables. That is, if two potential moderators were highly correlated with each other and with effect size, the one with the largest correlation with effect size was selected for the meta-regression models. Random effects analysis was used throughout but, in light of the modest number of studies in some categories and the large effect size variance, statistical significance was reported at the alpha=0.10 level as well as the conventional 0.05 level.

Results

Program Format and Treatment Modality

The literature search and coding process yielded data from 283 independent study samples. Note that many studies reported results separately for different subgroups of students (most commonly, results for boys and girls were reported separately). We treated each subgroup as a separate study sample. The 283 study samples participated in a variety of prevention and intervention programs. For purposes of analyzing their effects on student aggressive/disruptive behavior, we divided the programs into four groups according to their general service format. Programs differ across these groups on a number of methodological, participant, and intervention characteristics that made it unwise to combine them in a single analysis. The four intervention formats are as follows:

- *Universal programs*. These programs are delivered in classroom settings to all the students in the classroom; that is, the children are not selected individually for treatment but, rather, receive it simply because they are in a program classroom. However, the schools with such programs are often in low socioeconomic status and/or high crime neighborhoods, and thus, the children in these universal programs may be considered at risk by virtue of their socioeconomic background or neighborhood context.
- *Pullout programs for high-risk students*. These programs are provided to students who are specifically selected to receive treatment because of conduct problems or some risk factor (typically identified by teachers for social problems or classroom disruptiveness). Most of these programs are delivered to the selected children outside of their regular classrooms (either individually or in groups), although some are used in the regular classrooms but targeted on the selected children.
- *Special schools or classes*. These programs involve special schools or classrooms that serve as the usual educational setting for the students involved. Children are placed in these special schools or classrooms because of behavioral or academic difficulties that schools do not want to address in the context of mainstream classrooms. Included in this category are special education classrooms for behavior disordered children, alternative high schools, and schools-within-schools programs.
- *Comprehensive/multimodal programs*. These programs involve multiple distinct intervention elements (e.g., a social skills program for students and parenting skills training) and/or a mix of different intervention formats. They may also involve programs for parents or capacity building for school administrators and teachers in addition to the programming provided to the students. Within the comprehensive service format, some programs are delivered universally while others are targeted toward high-risk groups. All but one of the programs in this subcategory includes services for both students and their parents.

General Study Characteristics

Nearly 90 % of the studies were conducted in the United States with over 75 % run by researchers in psychology or education. Table 1 shows additional characteristics of the studies, broken down by program format. Among universal programs, nonrandomized designs were most common. However, cluster randomization, in which schools or classrooms were randomly assigned to treatment conditions, was utilized with some

School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis, Table 1 Study and method characteristics for each program format

	Universal		Pullout		Comprehensive		Special	
Variable	k	%	k	%	k	%	k	%
Method of assignment								
Individual randomization	6	6.7	91	71.1	5	23.8	20	44.4
Cluster randomized	32	36.0	11	8.6	10	47.6	9	20.0
Nonrandom	51	57.3	26	20.3	6	28.6	16	35.6
Year of publication								
1970s	8	9.0	36	28.1	1	4.8	8	17.8
1980s	17	19.1	33	25.8	2	9.5	18	40.0
1990s	33	37.1	48	37.5	14	66.7	13	28.9
2000s	31	34.8	11	8.6	4	19.0	6	13.3
Publication type								
Journal article	57	64.0	75	58.6	17	81.0	22	48.9
Thesis, dissertation, other	32	36.0	53	41.4	4	19.0	23	51.1
Total studies	89		128		21		45	

k = number of independent study samples

frequency for both the universal and comprehensive programs, as would be expected given how participants were typically recruited for such studies. Individual random assignment was the norm for the pullout programs and common for the special school programs, though cluster randomization and nonrandomized designs were also present for these formats. Overall, fewer than 20 % of the studies were conducted prior to 1980 and most were published in peer-reviewed journals (60 %), with the remainder reported as dissertations, theses, conference papers, and technical reports.

The student samples reflect the diversity in American schools (see Table 2). Most were comprised of a mix of boys and girls, but there were some all boy samples (19 %) and a few all girl samples (6 %). Formats targeting higher-risk children (i.e., pullout and special school programs) tended to have larger proportions of boys. Minority children were well represented with over a third of the studies having primarily minority youth; nearly 30 % of the included studies, however, did not report ethnicity information, making it difficult to examine differential program effects for different ethnic groups. Interestingly, missing information on ethnicity was more common among the special school and pullout programs, both of which tended to serve higher-risk groups of students. All school ages were included, from preschool through high school; the average age was around 9–10 for universal, pullout, and comprehensive programs and slightly older (about 13) for special school programs. A range of risk levels was also present, from generally low-risk students to those with serious behavior problems, and different risk levels were associated with the different program formats. Socioeconomic status was not widely reported, but a range of socioeconomic levels was represented among those studies for which it was reported.

Characteristics of the delivery and dosage of the programs, as well as the different program types, are shown in Table 3. Most studies were conducted as research or demonstration projects with relatively high levels of researcher involvement. Routine practice programs implemented by typical service personnel with less researcher involvement were less common. Program length varied by format, with comprehensive programs longest. Pullout and special school programs were the shortest, averaging 15 and 18 weeks, respectively. Universal programs averaged about 24 weeks. Service frequency also varied, with comprehensive and special school programs having larger proportions of daily contacts. Programs were generally delivered by teachers or the researchers themselves, though comprehensive programs tended to involve multiple types of

School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis, Table 2 Subject sample characteristics for each program format

	Universal		Pullout		Comprehensive		Special	
Variable	k	%	k	%	k	%	k	%
Gender mix								
No males	9	10.1	6	4.7	2	9.5	0	0.0
<50 % male	16	18.0	12	9.4	4	19.0	3	6.7
50–60 % male	51	57.3	30	23.4	6	28.6	12	26.7
>60 % male	3	3.4	47	36.7	6	28.6	23	51.1
All males	10	11.2	33	25.8	3	14.3	7	15.6
Age	m	range	m	range	m	range	m	range
Average age	9.2	4–16.5	10.3	4–16.0	9.1	5–16	12.7	4–16
Predominant ethnicity								
White	37	41.6	35	27.3	9	42.9	13	28.9
Black	23	25.8	30	23.4	5	23.8	11	24.4
Hispanic	10	11.2	6	4.7	3	14.3	1	2.2
Other minority	7	7.9	6	4.7	0	0.0	0	0.0
Missing	12	13.5	51	39.8	4	19.0	20	44.4
Socioeconomic status								
Mostly low SES	45	50.6	36	28.1	12	57.1	17	37.8
Middle + SES	44	49.4	39	30.5	8	38.1	15	33.3
Missing	0	0.0	53	41.4	1	4.8	13	28.9
Risk level								
General population	86	96.6	13	10.1	11	52.4	0	0.0
At-risk	0	0.0	80	62.5	8	38.1	30	66.7
Exhibiting problems	3[a]	3.3	35	27.3	2	9.5	15	33.3

[a] Three universal programs were delivered to entire classrooms, but results were presented only for the students exhibiting behavior problems

k = number of independent study samples

other delivery personnel. About 35 % of the reports mentioned some difficulties with the implementation of the program. This information, when reported, presented a great variety of relatively idiosyncratic problems, for example, attendance at sessions, dropouts from the program, turnover among delivery personnel, problems scheduling all sessions or delivering them as intended, wide variation between different program settings or providers, results from implementation fidelity measures, and the like. This necessitated use of a rather broad coding scheme in which we distinguished no problems indicated versus possible (some suggestion of difficulties but little explicit information) or definite problems with implementation.

The treatment modalities used in the four program formats varied. However, cognitive approaches and social skills training were common across all four service formats. Social cognitive strategies were cognitively oriented and focused on changing thinking patterns (e.g., hostile attributions) and developing social problem-solving skills (e.g., I Can Problem Solve; Shure and Spivack 1980). Anger management programs were also cognitively oriented but tended to focus on changing thinking patterns around anger and developing strategies for controlling angry impulses or coping with frustration (e.g., Coping Power; Lochman and Wells 2002). Social skills training focused on learning constructive behavior for interpersonal interactions, including communication skills and conflict management (Gresham and Nagle 1980). Also relatively common among the modalities were behavioral strategies that manipulated rewards and incentives (e.g., the Good Behavior Game, Dolan et al. 1993). Traditional counseling for individuals, groups, or families was also represented and tended to use a variety of therapeutic techniques (Nafpaktitis and Perlmutter 1998). The treatment coding was not mutually exclusive for any of the four program formats because many programs involved more than one

School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis, Table 3 Program characteristics for each program format

	Universal		Pullout		Comprehensive		Special	
Variable	k	%	k	%	k	%	k	%
Routine practice								
Research program	39	43.8	89	69.5	2	9.5	21	46.7
Demonstration program	38	42.7	24	18.8	18	85.7	17	37.8
Routine program	12	13.5	15	11.7	1	4.8	7	15.6
Delivery personnel								
Teacher	44	49.4	15	11.7	1	4.8	19	42.2
Researcher	17	19.1	57	44.5	1	4.8	12	26.7
Other personnel	28	31.5	48	37.5	19	90.5	14	31.1
Program length	m	sd	m	sd	m	sd	m	sd
Average weeks	24.3	32.6	15.4	17.1	54.0	34.5	17.9	12.6
Program frequency								
1× week or less	30	33.7	31	24.2	9	42.9	8	17.8
1–2× week	31	34.8	73	57.0	3	14.3	4	8.9
2–3× week	12	13.5	15	11.7	2	9.5	7	15.6
3–4× week	8	8.9	13	10.2	2	9.5	1	2.2
5× week	9	10.1	11	8.6	5	23.8	25	55.6
Implementation								
Problems = yes	32	36.0	36	28.1	12	57.1	12	26.7
Program elements	m	sd	m	sd	m	sd	m	sd
Mean # of elements	2.4	1.5	1.8	1.1	3.6	1.5	2.1	1.0
Format								
In-class (vs. out)	89	100.0	0	0.0	12	57.1	34	75.6
Program types								
Social cognitive program	58	65.1	39	30.5	7	33.3	11	24.4
Social skills training	42	47.2	28	21.9	11	52.4	11	24.4
Anger management	25	28.1	35	27.3	7	33.3	11	24.4
Behavioral program	8	9.0	31	24.2	6	28.6	14	31.1
Counseling	6	6.7	38	29.7	7	33.3	12	26.7
Academic services	2	2.2	4	3.1	6	28.6	17	37.8
Cognitive behavioral program	1	1.1	6	4.7	0	0.0	3	6.7

k = number of independent study samples

strategy. While many of the programs involved multiple program strategies, the comprehensive programs tended to have more strategies on average than the other three formats, and the selection of different components was distinctive for the comprehensive programs. The comprehensive programs tended to have multiple distinct components (e.g., a school-based cognitive component and a family-based component) and often had nonschool-based elements, while programs in the other three format categories that had multiple components tended to combine similar elements, like social cognitive and anger management components; they also tended to be solely school based.

The analyses reported below allow the 238 study samples to contribute multiple effect sizes with different characteristics, which are shown in Table 4. The aggressive behavior outcomes were grouped into six categories. Violence outcomes involved clear-cut violent behavior, including hitting. Delinquency was generally not aggressive or violent and included arrests, police contacts, and self-reported criminal behavior. Aggressive behavior outcomes were those that assessed clear physically aggressive behavior. Measures of aggressive behavior were similar to those for violence, but generally less serious. Ideally, we would have liked to examine program effects on relatively distinct groups of outcome constructs. However, very few of the measures that called themselves aggressive behavior focused solely on physically aggressive interpersonal behavior. Many included disruptiveness,

School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis, Table 4 Effect size and measurement characteristics for each program format

	Universal		Pullout		Comprehensive		Special	
	n = 175	k = 89	n = 272	k = 128	n = 72	k = 21	N = 139	K = 45
Variable	n (ES)	%	n (ES)	%	n (ES)	%	n (ES)	%
Type of dependent variable								
Violent behavior	13	7.4	11	4.0	11	15.3	12	8.6
Delinquency	17	9.7	26	9.6	11	15.3	36	25.9
Aggressive behavior	38	21.7	31	11.4	12	16.7	13	9.4
Aggression and disruptiveness mixed	53	30.3	93	34.2	26	36.1	60	43.2
Externalizing behavior	21	12.0	40	14.7	6	8.3	6	4.3
Undifferentiated problem behavior	33	18.9	71	26.1	6	8.3	12	8.6
Informant, source of outcome								
Self-report	50	28.6	34	12.5	29	40.3	41	29.5
Teacher report	68	38.9	126	46.3	15	20.8	41	29.5
Archive, record	18	10.3	43	15.8	2	2.8	33	23.7
Observation	19	10.9	31	11.4	9	12.5	11	7.9
Parent report	12	6.9	27	9.9	11	15.3	8	5.8
Other	8	4.6	11	4.0	6	8.3	5	3.6
Measurement timing								
Immediately after tx (vs. all later)	129	73.7	236	86.8	61	84.7	98	70.5
Effect size computation								
Means and sds (vs. all other)	130	74.3	199	73.2	50	69.4	105	75.5
Effect size estimation								
No estimate (vs. some)	155	88.6	227	83.5	59	81.8	117	84.2
Effect size adjustment								
Adjusted for pretest	145	82.9	232	85.3	56	77.8	86	61.9
Attrition	m	sd	m	sd	m	sd	m	sd
Average pre- to post-attrition	13 %	15 %	10 %	14 %	14 %	17 %	18 %	19 %

n = number of effect sizes, k = number of independent study samples

acting out, and other forms of behavior problems that are negative, but not necessarily aggressive. These measures were placed in the aggression and disruption mixed category. Externalizing measures generally involved disruptive and acting behaviors, but did not include physical aggression. Finally, the undifferentiated problem behavior category was comprised of those measures that included both internalizing and externalizing behaviors in the same score (e.g., the Child Behavior Checklist). Self-reports and teacher reports were the most common sources of information about aggressive and disruptive behavior, but observations, archival information, and parent reports were also used. Treatment effects were typically measured immediately after treatment, but some studies had longer follow-up periods. Effect sizes were generally computed directly from means and standard deviations, but other statistics were used in about 25 % of the cases. Direct computation from means and standard deviations or mathematically equivalent statistics was possible in over 80 % of the cases, and over 80 % of effect sizes had pretest effect sizes that allowed for adjustment. Attrition was moderate and averaged 10–18 % across the four program formats.

Overall Effects for School-Based Prevention Programs

The random effects weighted mean effect sizes and confidence intervals for the four program formats are shown in Table 5. The mean effect sizes for the universal, pullout, comprehensive, and special school programs are all positive and statistically significant, indicating that the participants in the school-based programs represented here had significantly lower levels of aggressive and disruptive behavior after the programs than students in comparison groups. The pullout

School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis, Table 5 Random effects mean effect sizes and confidence intervals for each program format

	Mean ES	*95 %* *CI*	*n*	*k*	Q_E	τ^2
Universal programs	0.16	(0.11, 0.22)	175	89	233	0.03
Pullout programs	0.37	(0.28, 0.47)	272	128	392	0.20
Comprehensive programs	0.06	(0.02, 0.11)	72	21	40	0.01
Special Schools programs	0.13	(0.03, 0.23)	139	45	91	0.07

programs evidenced the largest mean effect size of the four groups. The comprehensive programs had the smallest mean effect size at 0.06. The homogeneity statistics for the four groups of programs suggest that there is greater variability in the distributions of effect sizes than would be expected from sampling error. This variation was expected to be associated with the nature of the interventions, students, and methods used in these studies. The next step, then, is to examine some of the method, subject, and program characteristics that may be associated with the variability in treatment effects using meta-regression models.

Results for Universal Programs

There were 89 studies of universal programs in the database, all delivered in classroom settings to entire classes of students. Many studies tested program effects on more than one relevant outcome. Thus, the universal programs contributed 175 effect size estimates to the analysis. The random effects weighted mean effect size for universal programs was 0.16 ($p < 0.05$) and the distribution of effect sizes evidenced heterogeneity. The moderator analysis focused first on the relationship between study methods and the intervention effects using random effects inverse variance weights estimated via the method of moments. As mentioned above, moderator variables were selected based on their weighted zero-order correlations with effect size. This analysis is shown as Model I in Table 6.

Neither informant nor type of outcome measure was related to the effect size, so these variables were not included in the final models. Most notable is the lack of significant relationships between the study design variables and effect size. Method of assignment, pretest adjustments and computations, and attrition were not significantly related to effect size.

Our next step was to identify student characteristics that were associated with effect size while controlling for method variables. The results of this analysis are presented as Model II in the table. Only two student variables were significantly associated with effect size: gender mix and socioeconomic status. Programs with larger proportions of boys showed larger effects from universal programming than those with larger proportions of girls. Since boys tend to be more likely to exhibit aggressive, disruptive, or acting out behavior, this result is not surprising. Students with low socioeconomic status achieved significantly greater reductions in aggressive and disruptive behavior from universal programs than middle-class students ($p < 0.10$).

Model III in Table 6 shows the addition of the program characteristics to the models. Again, moderators were selected based on their correlations with effect size. In Model III, nonrandom assignment becomes active when the program variables are included, indicating that some of the variability remaining after controlling for subject and program characteristics was associated with method of assignment. Gender mix and socioeconomic status were still influential and associated with effect size. Several other variables in this analysis were also significant. Teacher and researcher delivery of programs had a significant relationship with effect size, with programs delivered by teachers or researchers having larger effects than those delivered by other personnel. In general, the other personnel were not school affiliated and included laypeople, social workers, and counselors. Well-implemented programs showed significantly larger effect sizes than those experiencing implementation problems. Dummy codes for the most common program types were also included in the model and none of those were statistically

School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis, Table 6 Random effects meta-regression models for universal programs

	Model I			*Model II*			*Model III*		
	b	*se*	*95 % CI*	*b*	*se*	*95 % CI*	*b*	*se*	*95 % CI*
Methodological characteristics									
Nonrandom assignment	0.05	0.09	(−0.13, 0.23)	0.10	0.11	(−0.11, 0.31)	0.26*	0.15	(−0.03, 0.55)
Cluster randomized	0.00	0.10	(−0.20, 0.20)	0.04	0.11	(−0.18, 0.25)	0.14	0.14	(−0.13, 0.41)
Pretest adjusted effect size	−0.11	0.08	(−0.26, 0.05)	−0.08	0.07	(−0.22, 0.06)	−0.04	0.08	(−0.19, 0.11)
ES computed with means/sds	−0.06	0.06	(−0.19, 0.06)	−0.10	0.06	(−0.22, 0.03)	−0.09	0.07	(−0.23, 0.05)
Percent attrition	−0.01	0.13	(−0.26, 0.24)	−0.10	0.13	(−0.36, 0.16)	0.08	0.15	(−0.22, 0.38)
Subject characteristics									
Gender mix				0.04**	0.02	(0.01, 0.08)	0.03*	0.02	(0.00, 0.07)
Average age				−0.02	0.01	(−0.04, 0.00)	0.00	0.01	(−0.03, 0.02)
Predominantly low SES				0.10*	0.05	(0.00, 0.20)	0.18**	0.07	(0.04, 0.32)
Program characteristics									
Teacher delivery of program							0.11*	0.06	(−0.01, 0.23)
Researcher delivery of program							0.24**	0.10	(0.04, 0.44)
Implementation problems = yes							−0.15*	0.07	(−0.29, 0.00)
Frequency of tx sessions							0.00	0.02	(−0.03, 0.04)
Social cognitive program							−0.02	0.07	(−0.15, 0.11)
Social skills training							−0.10	0.07	(−0.25, 0.04)
Anger management							0.07	0.08	(−0.08, 0.22)
Behavioral program							0.11	0.13	(−0.15, 0.37)

Mixed effects meta-regression models estimated with robust standard errors to account for dependent effect sizes estimates (n = 175; k = 89). b - unstandardized regression coefficient

*$p < 0.10$

**$p < 0.05$

significant. The different program types all produced positive effects on the outcomes and were not appreciably different from each other in effectiveness.

Results for Pullout Programs for High-Risk Students

There were 128 studies of pullout programs that targeted interventions to individually identified children. These studies contributed 272 effect sizes. The overall random effects mean effect size for these programs was 0.37 ($p < 0.05$), nearly twice as large as the average effect size for universal programs. The homogeneity test of the effect sizes showed significant variability across studies, and our analysis of the relationships between effect size and methodological and substantive characteristics of the studies proceeded much the same as for the universal programs (see Table 7). First, Model I examined the influence of each method variable on the aggressive/disruptive behavior effect sizes. Here also the study design was not associated with effect size—random assignment studies did not show appreciably smaller or larger effects than nonrandomized studies. Note that for the pullout programs, the design contrast was primarily between individual-level randomization and nonrandomization; there were only 11 cluster randomized studies. The two method variables that did show significant relationships with effect size were the

School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis, Table 7 Random effects meta-regression models for pullout programs for high-risk students

	Model I			*Model II*			*Model III*		
	b	*se*	*95 % CI*	*b*	*se*	*95 % CI*	*b*	*se*	*95 % CI*
Methodological characteristics									
Aggressive behavior outcome	−0.27	0.17	(−0.60, 0.06)	−0.22	0.15	(−0.53, 0.08)	−0.17	0.16	(−0.48, 0.14)
Problem behavior outcome	0.20**	0.09	(0.01, 0.38)	0.20*	0.09	(0.02, 0.39)	0.25**	0.10	(0.06, 0.45)
Externalizing behavior outcome	0.09	0.12	(−0.14, 0.33)	0.17	0.14	(−0.11, 0.45)	0.29**	0.14	(0.01, 0.58)
Random assignment	−0.02	0.09	(−0.21, 0.16)	−0.10	0.09	(−0.29, 0.08)	−0.06	0.10	(−0.25, 0.14)
Percent attrition	−0.93**	0.28	(−1.48, −0.39)	−1.02**	0.29	(−1.59, −0.44)	−1.07**	0.32	(−1.69, −0.44)
Subject characteristics									
Gender mix				0.01	0.04	(−0.08, 0.10)	0.02	0.05	(−0.07, 0.12)
Average age				0.01	0.02	(−0.02, 0.05)	0.01	0.02	(−0.03, 0.05)
Risk level				0.16**	0.04	(0.07, 0.24)	0.18**	0.04	(0.10, 0.27)
Program characteristics									
Researcher delivery of program							−0.21*	0.12	(−0.45, 0.04)
Other tx personnel							−0.15	0.12	(−0.39, 0.08)
Routine practice program							0.08	0.07	(−0.05, 0.21)
Implementation problems = yes							−0.36**	0.10	(−0.56, −0.16)
Frequency of tx sessions							0.01	0.02	(−0.03, 0.05)
Anger management							0.04	0.15	(−0.25, 0.33)
Social cognitive program							−0.02	0.14	(−0.29, 0.24)
Counseling							0.11	0.12	(−0.13, 0.34)
Behavioral program							0.10	0.12	(−0.13, 0.33)
Social skills training							0.24	0.15	(−0.07, 0.54)

Mixed effects meta-regression models estimated with robust standard errors to account for dependent effect sizes estimates (n = 272; k = 128). b - unstandardized regression coefficient

*$p < 0.05$

**$p < 0.1$

outcome measures of undifferentiated problem behavior and attrition. Outcome measures that included both internalizing and externalizing problem behavior tended to produce larger effect sizes than the other outcomes (all more closely associated with aggressive behavior). Attrition was associated with smaller effect sizes.

Model II in Table 7 shows the addition of the subject characteristics to the analysis. One student characteristic had a significant relationship with effect size in this model. Higher-risk subjects showed larger effect sizes than lower-risk subjects though, with the pullout programs, very few low-risk children were involved. The distinction here is mainly between indicated students who are already exhibiting behavior problems and selected students who have risk factors that may lead to later problems. Nevertheless, programs with the highest-risk students tended to have larger effects.

Model III includes the characteristics of the intervention programs. In contrast to the universal programs, researchers and other non-teacher personnel as treatment delivery agents were less effective than teachers. Programs with higher-quality implementation were associated with larger effects. None of the different program types was significantly better or worse than any other.

Results for Comprehensive or Multimodal Programs

There were only 21 studies of comprehensive programs in the database, distinguished by their multiple treatment components and formats. These programs contributed 72 effect sizes indexing treatment effects on aggressive and disruptive behavior. The average number of distinct treatment components for these programs was nearly four, whereas the universal and pullout programs typically employed one or two treatment components. The studies of comprehensive programs tended to involve larger samples of students than the other program formats and a larger proportion of cluster randomizations as well. Thus, the statistical significance of the mean effect size of 0.06 is likely overstated. Comprehensive programs were generally longer than the universal and pullout programs. The modal program covered an entire school year and almost half of the programs were longer than 1 year. In contrast, the average program lengths for universal and pullout programs were 24 and 15 weeks, respectively.

Students who participated in comprehensive programs were slightly better off than students who did not, though the effect size was small. The Q has relatively low statistical power with small numbers of studies; therefore, despite the relative homogeneity in relation to the other program formats, we ran a single meta-regression model to examine the influence of some key study characteristics. This model is shown in Table 8. Because of the limited number of studies available in this category, only a few variables were tested. None of the method characteristics had correlations with effect size above 0.20 and were not tested. We also examined student risk level and found that it was not associated with effect size. Overlapping with the risk level variable was the variable we tested in the 2007 paper which compared universally delivered programs to those targeting higher-risk groups in a pullout format. Like the risk variable, this variable was not associated with effect size. Programs with implementation problems tended to have smaller effects than those without such problems, but the relationship was not significant. However, the frequency of treatment and the number of different treatment components were both associated with effect sizes. More frequent treatments produced larger effects on aggressive and disruptive behavior, while programs with more components produced smaller effects. The comprehensive programs with larger numbers of distinct components tended to be less effective than those with fewer components.

School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis, Table 8 Random effects meta-regression model for comprehensive programs

	Model I			
	b	*se*	*95 % CI*	
Methodological characteristics				
Nonrandom assignment	−0.04	0.05	(−0.14,	0.06)
Subject characteristics				
Risk level	0.00	0.01	(−0.02,	0.03)
Program characteristics				
Implementation problems = yes	−0.04	0.03	(−0.12,	0.03)
Frequency of tx sessions	0.02*	0.01	(0.01,	0.03)
Mean # components	−0.04*	0.01	(−0.06,	−0.02)

Mixed effects meta-regression models estimated with robust standard errors to account for dependent effect sizes estimates (n = 72; k = 21). b - unstandardized regression coefficient

*$p < 0.05$

Results for Special Schools or Classes

There were 45 studies of programs delivered in special schools or classrooms. These programs generally involved an academic curriculum plus programming for social or aggressive behavior. The students typically had behavioral (and often academic) difficulties that resulted in their placement outside of mainstream classrooms. The mean aggressive/disruptive behavior effect size for these programs was 0.13 ($p < 0.05$). The Q-statistic ($Q_e = 91$) indicates that the distribution of effect sizes was heterogeneous. About 20 % of the studies of special programs assigned students to intervention and control conditions at the classroom level. As a result, there may be a design effect associated with the

School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis, Table 9 Random effects meta-regression models for the special school programs

	Model I			*Model II*			*Model III*		
	b	*se*	*95 % CI*	*b*	*se*	*95 % CI*	*b*	*se*	*95 % CI*
Methodological characteristics									
Delinquency outcome	−0.19*	0.09	(−0.37, −0.01)	−0.18*	0.08	(−0.34, −0.01)	−0.15*	0.08	(−0.31, 0.00)
Random assignment	−0.02	0.10	(−0.22, 0.18)	−0.01	0.11	(−0.23, 0.21)	0.00	0.09	(−0.18, 0.18)
Subject characteristics									
Average age				0.00	0.02	(−0.03, 0.04)			
Risk level				0.14	0.13	(−0.12, 0.40)			
Program characteristics									
Implementation problems = yes							−0.22*	0.09	(−0.40, −0.04)
Length of program							−0.01	0.00	(−0.02, 0.00)
# of components							0.14*	0.06	(0.01, 0.27)
In-class versus out of class							0.11	0.09	(−0.08, 0.30)
Anger management							0.17	0.14	(−0.13, 0.46)
Social cognitive program							−0.24	0.15	(−0.55, 0.06)
Counseling							0.05	0.08	(−0.12, 0.21)

Mixed effects meta-regression models estimated with robust standard errors to account for dependent effect sizes estimates (n = 139; k = 45). b - unstandardized regression coefficient
$^*p < 0.05$

clustering of students within classrooms that overstates the significance, though the overall effect size and the regression coefficients presented below should not be greatly affected.

The limited number of studies available for this program format necessitated that we examine fewer moderator variables in the meta-regression models (See Table 9). The two method characteristics with the highest weighted zero-order correlations with effect size were random assignment and delinquency outcomes. Studies employing individual randomization did not produce effect sizes that were smaller or larger than studies that did not use random assignment. When program effects were assessed on delinquency outcomes for the special school programs, effect sizes were significantly smaller than when any other outcome variable was used. Note that all of the other outcomes contained some form of aggressive, violent, or acting out/disruptive behavior, while the delinquency outcomes included arrests, police contacts, and self-reported crime.

In Model II, two subject characteristics were tested, average age and risk level. Neither age nor risk level were associated with effect size. These subject characteristics were dropped from Model III to allow us to test more different treatment characteristics. The length of the program did not have a significant relationship to effect size. In one form of the special school programs, students were assigned to special education classes or schools and the program was delivered entirely in the classroom setting. The other form involved students in special education classrooms who were pulled out of class for additional small group treatments. The programs delivered in classroom settings did not produce different effects than the pullout programs. Also, as in other analyses, better implemented programs showed larger effects. Three treatment modalities were tested in this model, anger management, social cognitive programs, and counseling. None were found to be significantly more effective than any of the other strategies.

Summary and Conclusions

The primary issue addressed in this entry is the effectiveness of programs for preventing or reducing such aggressive and disruptive behaviors as aggression, fighting, bullying, name-calling, intimidation, acting out, and disruptive behaviors occurring in school settings. The main finding is that, overall, the school-based programs that have been studied by researchers (and often developed and implemented by them as well) generally have positive effects for this purpose. A secondary purpose of this entry was to utilize new analytic techniques that allow for multiple dependent effect sizes from the same study to be included in an analysis (Hedges et al. 2010). Several differences from our original work are noteworthy. The mean effect size using the new techniques for universal programs was 0.16, while it was 0.21 in Wilson and Lipsey (2007). The mean effect size for the pullout programs was 0.37, quite a bit larger than the 0.29 published in Wilson and Lipsey (2007). It is well known that selecting a single effect size from each study is not ideal in meta-analysis; this reanalysis dramatizes how much difference such selections can make on results and conclusions.

The most common and most effective approaches are universal programs delivered to all the students in a classroom or school and pullout programs targeted for at-risk children who participate in the programs outside of their regular classrooms. The universal programs that were included in the analysis mainly used social cognitive approaches, anger management, and social skills training, and, like the original analysis, all strategies appeared about equally effective. In the 2007 meta-regression model for the universal programs, only SES was significant. The new analyses found that implementation quality and the type of delivery personnel were also important predictors of treatment effects.

Cognitively oriented approaches (both social cognitive programs and anger management) and social skills training were also the most frequent among the pullout programs, but many did use behavioral or counseling treatment modalities as well. In the original analysis, behavioral strategies were associated with larger treatment effects. However, differential use of the different modalities within the pullout format was not associated with differential effects in the new analysis. This suggests that it may be the pullout program format that is most important but does not rule out the possibility that the treatment modalities used with that format are especially effective ones. It is also possible that the larger effect size with the pullout programs was due to the generally higher-risk students participating in these programs. Higher-risk students generally achieve better outcomes from psychosocial programming than lower-risk students, so the larger effect size for the more targeted pullout programs may be partly due to the students with more serious behavior problems having greater room for improvement.

The mean effect sizes of 0.16 and 0.37 for the universal and pullout programs, respectively, represent a decrease in aggressive/disruptive behavior that is not only statistically significant but likely to be of practical significance to schools as well. Suppose, for example, that approximately 20 % of students are involved in some version of such behavior during a typical school year. This is a plausible assumption according to the *Indicators of School Crime and Safety*: *2005* which reports that 13 % of students age 12–18 were in a fight on school property, 12 % had been the targets of hate-related words, and 7 % had been bullied (DeVoe et al. 2004). Effect sizes of 0.16 and 0.37 represent reductions from a base rate prevalence of 20 % to about 14 % and 8 %, respectively, that is, 30–40 % reductions in aggressive and disruptive behavior. The programs of above average effectiveness, of course, produce even larger decreases.

The substantial similarity in effects across the different types of programs within the universal and pullout formats suggests that schools may choose from a range of such programs with some confidence that whatever they pick will be about as effective as any other choice. In the absence of evidence that one modality is significantly more effective at reducing aggressive and disruptive behavior than another, schools might benefit most by considering ease of implementation, compatibility with school culture, and teacher and administrator training and preferences when selecting programs and focusing on implementation quality once programs are in place. Our coding of implementation quality, albeit crude, was associated with larger effect sizes for all four treatment formats. A very high proportion of the studies in this meta-analysis, however, were research or demonstration projects in which the researchers had a relatively large direct influence on the service delivery. Schools adopting these programs without such engagement may have difficulty attaining comparable program fidelity, a concern reinforced by evidence of frequent weak implementation in actual practice (Gottfredson and Gottfredson 2002). The best choice of a universal or pullout program for a school, therefore, may be the one they are most confident they can implement well. In addition, for the universal programs, teachers and researchers as delivery personnel were generally more effective than nonschool personnel. Selecting programs that fit with teachers' interests, training, and existing curricula may help to keep implementation fidelity as high as possible.

Another significant factor that cut across the universal and targeted pullout programs was the relationship of student characteristics to program effects. Larger treatment effects were achieved with higher-risk students, even among the already higher-risk samples in the targeted pullout programs. For the universal programs, the greatest benefits appeared for students from economically disadvantaged backgrounds and for groups with larger proportions of boys. Universal programs did not specifically select students with individual risk factors or behavior problems, though many students were of low socioeconomic status and there were most likely some behavior problem students in the classrooms that received universal interventions. These findings reinforce the fact that a program cannot have large effects unless there is sufficient problem behavior, or risk for such behavior, to allow for significant improvement.

The programs in the category we called comprehensive, in contrast to the universal and pullout programs, were surprisingly ineffective. On the face of it, combinations of universal and pullout treatment elements and multiple intervention strategies would be expected to be at least as effective, if not more so, than less multifaceted programs. The small mean effect size for the comprehensive programs raises questions about their value. It should be noted, however, that there were few programs in this category and that most of these were long-term school-wide programs. It may be that this broad scope is associated with some dilution of the intensity and focus of the programs so that students have less engagement with them than with the programs in the universal and pullout categories. The comprehensive programs also suffered from implementation difficulties more frequently than the other formats (see Table 3), although implementation was not significantly associated with effect size in the reanalysis or in the original. However, the reanalysis found that the comprehensive programs with the largest number of treatment components were the least effective in this category, a finding that may be an indicator for implementation problems. This is an area that clearly warrants further study.

The most distinctive programs in our collection were those for students in atypical school settings. The mean effect size for these programs was modest, though statistically significant. As in the original analysis, implementation was important. The results also are somewhat anomalous. One of the signal characteristics of students in these settings is a relatively high level of behavior problems or risk for such problems; thus, there should be ample room for improvement. On the other hand, the special school settings in which they are placed can be expected to already have some programming in place to deal with such problems. The control conditions in these studies would thus reflect the effects of that practice-as-usual situation with less value added provided by additional programming of the sort examined in the studies included here. Alternatively, however, the add-on programs studied in these cases may have been weaker or less focused than those found in the pullout format, or the more serious behavior problems of students in these settings may be more resistant to change. Here too are issues that warrant further study.

A particular concern of our original meta-analysis (Wilson et al. 2003) was the smaller effects of routine practice programs in comparison to those of the more heavily represented research and demonstration programs. Routine practice programs are those implemented in a school on an ongoing routine basis and evaluated by a researcher with no direct role in developing or implementing the program. Research and demonstration programs are mounted by a researcher for research or demonstration purposes with the researcher often being the program developer and heavily involved in the implementation of the program, though somewhat less so for demonstration programs. In the present meta-analysis, somewhat more studies of routine programs were included, and it is reassuring that their mean effect sizes were not significantly smaller. The zero-order correlation between routine practice and effect sizes was only large enough to be included in the meta-regression models for the pullout programs, and it did not turn out to have a significant relationship with effect size in the model in which it was tested. However, only 35 of the 283 studies in this meta-analysis examined routine practice programs. This number dramatizes how little evidence exists about the actual effectiveness, in everyday real-world practice, of the kinds of school-based programs for aggressive/disruptive behavior represented in this review.

Notes and Acknowledgments The research reported in this entry was supported by grants from the National Institute of Mental Health, the National Institute of Justice, and the Centers for Disease Control and Prevention.

A bibliography of studies used in the meta-analysis is available from the first author.

Related Entries

- ▶ Juvenile Violence
- ▶ School Crime Statistics
- ▶ School Security Practices and Crime
- ▶ School Social Organization, Discipline Management, and Crime
- ▶ School Structural Characteristics and Crime

Recommended Reading and References

Brener ND, Martindale J, Weist MD (2001) Mental health and social services: results from the school health policies and program study. J Sch Heal 71:305–312

DeVoe JF, Peter K, Kaufman P et al (2004) Indicators of school crime and safety: 2004 (Report Num. NCES 2005-002/NCJ 205290) Washington, DC: U.S. Departments of Education and Justice

Dolan LJ, Kellam SG, Brown CH, Werthamer-Larsson L, Rebok GW et al (1993) The short-term impact of two classroom-based preventive interventions on aggressive and shy behaviors and poor achievement. J Appl Dev Psychol 14:317–345

Gottfredson DC, Gottfredson GD (2002) Quality of school-based prevention programs: results from a national survey. J Res Crime Delinquency 39:3–35

Gottfredson GD, Gottfredson DC, Czeh ER, Cantor D, Crosse S, Hantman I (2000) National study of delinquency prevention in schools. Final Report, Grant No. 96-MU-MU-0008. Ellicott City, MD: Gottfredson Associates, Inc. Available from:www.gottfredson.com

Gresham FM, Nagle RJ (1980) Social skills training with children: responsiveness to modeling and coaching as a function of peer orientation. J Consult Clin Psychol 48:718–729

Hedges LV, Olkin D (1985) Statistical methods for meta-analyses. Academic Press, San Diego

Hedges LV, Tipton E, Johnson MC (2010) Robust variance estimation in meta-regression with dependent effect size estimates. Res Syn Method 1:39–65

Lipsey MW, Wilson DB (2001) Practical meta-analysis. Sage, Thousand Oaks

Lochman JE, Wells KC (2002) The coping power program at the middle-school transition: universal and indicated prevention effects. Psychol Addict Behav 16(suppl 4): S40–S54

Nafpaktitis M, Perlmutter BF (1998) School-based early mental health intervention with at-risk students. Sch Psychol Rev 27:420–432

Shure MB, Spivack G (1980) Interpersonal problem solving as a mediator of behavioral adjustment in preschool and kindergarten children. J Appl Dev Psychol 1:29–44

Wilson SJ, Lipsey MW (2007) School-based interventions for aggressive and disruptive behavior: update of a meta-analysis. Am J Prev Med 33(2S):S130–S143

Wilson SJ, Lipsey MW, Derzon JH (2003) The effects of school-based intervention programs on aggressive and disruptive behavior: a meta-analysis. J Consult Clin Psychol 71:136–149

Scientific Basis of the Forensic Process

William J. Tilstone
Centre for Forensic Science, University of Strathclyde, Glasgow, Scotland, UK

Overview

The chapter reviews the contributions to the History and Philosophy of Science (HPS) that impact on the scientific basis of the forensic process. There is a close parallel between the scientific method and the forensic process, in that each consists of hypothesis creation, testing and review based on observations and rigorous challenge. The leaders in HPS whose thoughts are most relevant are Popper, for his views on the limitations of inductive reasoning, Peirce, for introducing the concept of abduction, Kuhn for his view on "normal science" and the sociology of scientific developments, and Lakatos, for formalizing Popper's views into what he termed research programs.

Having established the scientific foundation of the forensic process, the chapter turns to how it is used to convert "science" into knowledge in the form of evidence arising from an investigation and of relevance to the determination of facts by a tribunal. The fundamental types of forensic evidence, classified by purpose, are described, leading to a taxonomy of forensic evidence based on the principles of rhetoric.

Is It Science?

There is no definition of "forensic science" that addresses the question of its scientific basis. Most definitions are constructed around the concepts of "science applied to the law" and "working carefully", but these beg more questions than they answer. Is parentage testing for immigration purposes forensic? What about dope testing of Olympic athletes, or race horses? Is investigation of environmental spills forensic? Where does pre-employment drug screening fit? And "working carefully" encompasses pilots, accountants, and surgeons, but that clearly does not make them forensic.

Adopting the approach that it is something conducted by scientists in the laboratory does not help much either. For example, aspects of a homicide investigation that could be classed as "forensic" may be conducted at the scene by personnel as diverse as those who are serving police officers and whose qualification consists of successful completion of high school together with in-service training, and by toxicologists and DNA analysts with a Ph.D. degree working in a university or private laboratory.

Not too long ago, such questions would have seemed too academic and abstract for a work on forensic science, the forensic literature was dominated by case reports and papers describing new or improved technical methods. That should never have been so and is certainly less so today. The relatively recently published report by the National Academies of Science on forensic science in the US was highly critical of forensic science in general, and aspects of crime scene-related criminalistics such as tool marks and fingerprints in particular, for not being well-founded in science. The UK too has embraced basic scientific philosophy in evaluation of forensic evidence, where proposed legislation on the admissibility of expert evidence requires that a trial judge may refuse to admit expert evidence that is based on a hypothesis which has not been subjected to scrutiny (including experimental or other testing) or has failed to stand up to scrutiny.

However, the main unanswered question, other than a superficial passing mention in the NAS report, is what exactly **is** science?

This chapter addresses these issues by first of all exploring what science really is, then defining and exploring the forensic process, and thereby demonstrating how the forensic process is indeed fundamentally a scientific endeavor, with all the advantages and limitations that arise from that. The chapter ends by addressing the question "does it matter whether or not it is based in science" and explores some consequences.

Science, Technology, and the Scientific Method

The word “science” is derived from the Latin verb *scire*, to know, and its noun *scientia*, knowledge. Although the principles of the scientific method can and have been applied to other areas including psychology and the social sciences, the accepted usage today is knowledge of the natural world. Hence the generally accepted perspective of forensic science as a source of unqualified and independent information, but the history of the scientific method and the writings of the philosophers of science tell us that science is not absolute.

The ancient Chinese and early Muslim societies contributed many discoveries and inventions relevant to natural laws, but “knowledge”, including scientific knowledge, in western societies is built on the foundations provided by the ancient Greeks and Romans. As a result, up until the early sixteenth century, philosophers formulated what they regarded as laws of nature based on the universal belief that the Earth was the center of the Universe. Thus it was “known” that the Earth was at rest in the centre of the universe. This was a natural and inevitable conclusion arising from the form of deductive logic that prevailed at the time, namely that having accepted certain propositions we can increase our knowledge by using those to predict outcomes, and so new propositions come into being.

This changed dramatically when Nicolaus Copernicus (1473–1543) decided to test the accepted knowledge by carrying out observations on the heavenly bodies to see if they behaved as predicted. They did not, and the scientific method was born.

The next century reinforced the importance of inductive logic in developing our knowledge of the natural world. Francis Bacon (1561–1626) described what is still today accepted as the essence of the scientific method, namely the cycle of observations to determine facts, followed by application of inductive reasoning to formulate a hypothesis that explains the observations, followed by further experimentation to test the hypothesis, resulting in its acceptance, refinement or rejection. Bacon’s ideas were central to the subsequent three centuries of scientific enlightenment, and stimulated interest in science as an area of study in its own right.

The core of Bacon’s scientific method is the role of the hypothesis. It is the link between observation and interpretation, and one that requires the utmost rigor in its application. Contrary to the popular expression, facts do not speak for themselves. The same observations can satisfy many interpretations, and many of the advances in what can be regarded as “scientific” are concerned with the relationship between observation, hypothesis, and interpretation. A fact is something known or proven to be true and may therefore be assumed to be the starting point or end point of the cycle, but “facts” themselves may indeed not be objective, proven, immutable entities, but may be subjective perceptions or interpretations of some consequence of a true fact.

The work of the American philosopher of science Charles Sanders Peirce (Peirce 1934) is central to the understanding of the strengths and weaknesses of the science that underpins forensic science, especially in regard to current controversies regarding contextual bias. Before Peirce, scientific enquiry was either *inductive*, that is, making inferences from a particular set of phenomena in order to predict and make propositions of more general character (empiricism), or *deductive*, that is, inferring from accepted truths and using general laws to deduce outcomes (rationalism). He introduced a third option, “abduction,” using the formulation and testing of hypotheses to explore possible explanations for a surprising event or occurrence, one in which there is no existing set of phenomena. Abduction therefore gives us an alternative entry point into the scientific method, and of course in this context criminal offences are almost always surprising events.

In practical terms the hypothesis is the scientist’s best guess of the character of phenomena that he or she is going to investigate. Intuitively, the validity or strength of any guess will be proportionate to the quality and extent of the information on which it is based. Abduction is therefore counter-intuitive but provides the most fertile although least secure mode of inference. The hypothesis, being based on a "surprise event" and lacking a foundation of pre-existing phenomena, needs to have practical implications that can be rigorously tested.

Peirce himself was aware of the limitations of abduction as well as its advantages. He expressed it thus: "Deduction proves that something must be; Induction shows that something actually is operative; Abduction merely suggests that something may be."

A functional definition of forensic science is that it is concerned with identifying objects at a crime scene that can be used to provide a credible reconstruction of the event (Tilstone et al. 2012), and induction, abduction and deduction are all essential elements of the process in deriving that reconstruction. The surprise event is the crime itself and trying to explain what happened begins with the formulation of an explanatory hypothesis, followed by its testing and refinement. However, the information available as a consequence of that one event, the crime, is an insufficient basis for the conventional inductive approach to hypothesis formulation. On the other hand abduction is an imaginative, "what if," process that enjoys exploring the unexpected, captures the intuitive actions of the crime scene investigator, and can lead to formulation of one or more credible hypothesis. Peirce established abduction as a legitimate element of the scientific method, but the imaginative beginning must be balanced by rigorous, objective, scrutiny of alternate hypotheses and testing of their validity.

How then can a hypothesis be evaluated with a view to establish whether it is a matter of fact and not just a possible explanation? The cycle of the scientific method demonstrates that science is not an absolute, but a continuous process of searching for the truth, and therefore cannot prove any hypothesis to the exclusion of everything else. Karl Popper (Popper 1970) was a key figure in the debate. He held that all observation is from a point of view, and therefore any conclusion drawn from observations will be colored consciously or sub-consciously by our existing understanding. This by the way was 40–50 years before the "discovery" of contextual bias in the forensic sciences.

Popper's critical rationalism challenged the prevailing thinking in science, which held that a theory could be verified and thus acknowledged as being true on the basis of direct observations in the natural world. At its simplest, Popper's view is that no amount of confirmation proves a theory but one reliable falsification disproves it.

This point was controversial in that the prevailing theory of science preceding Popper was one that relied on empirical observation as the basis of drawing scientific conclusions. If one could make direct observations in the world which supported one's theory – then the theory was believed to be verified and thus could be acknowledged as true.

Popper used a simple example to illustrate his disagreement. Everyone in Europe and North America knew that all swans were white, but the exploration of Australasia was accompanied by the discovery of black swans, thereby challenging the accepted situation and leading to refinement of the properties that define "swan." Using this and other examples, Popper successfully argued that using observation to verify a theory does not necessarily result in secure knowledge or empirical truth. He proposed that seeking to falsify theories rather than verify them was a much more reliable way in which to produce knowledge that could be trusted.

Falsification is a simple and powerful tool for testing scientific hypotheses. Sadly it is one that seems to have been beyond the wit of the then US Supreme Court Chief Justice, at the time of the Court's decision in Daubert –v- Merrel Dow. It is itself not absolute, however, and apparent falsifications have to be subject to the same scrutiny as any other element of a hypothesis. Observations may produce information that apparently falsifies a hypothesis, but further testing may provide an explanation that maintains the integrity of the law. Such was the case with the discovery of the planet Neptune. The development of better telescopes led to the observation that the orbit of the planet Uranus was not as predicted by Newton's laws. Rather than reject Newton's laws which had stood the test of time, astronomers searched for an explanation, the most likely of which was the presence of a hitherto undiscovered planet. And so it turned out when Neptune was observed in 1846, in exactly the predicted place.

These examples – the black swans and the discovery of Neptune – show that an essential element of a good theory is its ability to predict something or even better forbid it, and this is as true for the specific application of science in forensic investigations as it is for science in general.

Mention was made above of how Popper's falsification challenged the status quo of the scientific method at the time. This leads to an alternate approach to defining science, by focusing on the behaviors of scientists rather than seeking an explanation of "science" based on the methods used or framework within which they work.

This was the approach taken by Thomas Kuhn (Kuhn 1966) who argued that scientists work within a conceptual paradigm that strongly influences the way in which they see data. Kuhn was interested in the way that science advances, and differentiated between what he called normal science, which progresses incrementally, little by little, and revolutionary science, where no change occurs until there is overwhelming evidence to overcome the prevailing paradigm which is then cast aside and replaced by the new.

In conducting his research Kuhn had to grapple with the essential question of what "science" is in order to be able to discuss how changes come about in scientific theories. He concluded that it is defined by communities of practitioners sharing properties related to the manner in which they conduct the exploration of knowledge about some aspect of the physical world. It is the existence of these communities and their established norms that make the paradigm such a strong force, inevitably resulting in conservatism.

Imre Lakatos is the last of the scientific philosophers to be considered here, but by no means the least, certainly in considering the foundation principles of forensic science. Lakatos was concerned that a strict and unthinking interpretation of Popper's falsification concept meant a hypothesis would be rejected, but its core usefulness might still be valid but lost (Lakatos 1970). Any apparent falsification should not be accepted without question, but may instead point the way to further refinement and improvement of the apparently false hypothesis.

To remedy what he saw as a problem with Popper's theory Lakatos introduced the concept of "research programmes" as a solution. A research programme is the aggregate of several theories which share a core and that progress using a common set of methodologies. In Lakatos' approach, a scientist works within a "research program" that corresponds roughly with Kuhn's "paradigm." Whereas Popper rejected the use of ad hoc hypotheses as unscientific, Lakatos accepted their place in the development of new theories, as of course did Peirce.

The main contribution to forensic science made by Lakatos is his insistence on incorporating the concept of *ceteris paribus* into the evaluation of a hypothesis. Translated as "other things being equal," *ceteris paribus* provides a context in which the formulation of a hypothesis and its acceptance, rejection, or modification depend on the state of our knowledge at the time. Its significance can be summarized as "We don't know what we don't know" and future discoveries – either new and unsuspected information like black swans or apparent contradictions to a well-established theory like Newton's law – can alter the interpretation of observations.

In summary, a single exact definition of science has eluded philosophers of science, but whether it is defined in terms of the scientific method or the behaviors of scientists, one thing is certain: science is not absolute. At any given time the apparent facts revealed by observations and their interpretation are limited by many factors: the limitations imposed by the state of technology (the quality of telescopes, the ability to travel to the opposite side of the Earth); the influence of the paradigms within which scientists work (the behavioral norms of each sect within science); the danger of seeking to confirm rather than challenge; and the impossibility of dealing with things unknown – "we don't know what we don't know" but seldom qualify (or are allowed to qualify) judgment with an admission of *ceteris paribus*.

From Science to Knowledge to Evidence

Recognizing that science is not absolute is not the same as saying science is junk, to use a term currently in vogue among the detractors of forensic science. Bacon was a champion of science as a tool for the betterment of humankind, and our lives today are shaped by it, from the trivia of the entertainment industry, to the very practical and significant advances in medicine, and the seemingly arcane activities associated with increasing our understanding of the natural world, such as the Large Hadron Collider. In the criminal justice system, the value of science is that it provides evidence that is indisputably more reliable than eye-witness testimony and that provides significant information not otherwise available to the triers of fact. For example, the Innocence Project in the United States (Innocence Project 2011) has identified eye witness misidentification as a leading factor in 75 % of the wrongful conviction cases that it has successfully pursued.

The place of forensic science within the justice system is described by Hald, in the entry on "The Philosophical Basis of the Forensic Process," and the section that follows is to some degree complementary to her paper, as it demonstrates how applying the scientific method to the investigation of crimes produces knowledge that is expressed as evidence for the legal tribunal.

An understanding of the scientific method, and to an extent, how it can produce the knowledge inherent in forensic science, permits an analysis of the purpose of the forensic process, evidence. Although dictionaries define "evidence" in a legal context as something submitted to a court to assist it in arriving at a decision, crime laboratories in the US also use "evidence" synonymously with "exhibit," namely a physical entity recovered from a crime scene. This chapter avoids that usage, preferring the sequence that an exhibit is something presented before a court, the examination of which has produced evidence that may assist the tribunal in its determinations.

Types of Forensic Evidence

The legal process distinguishes between real and testimonial evidence, which are respectively evidence resulting from the examination of exhibits and evidence resulting from the direct narrative of a witness as to what he or she has done or observed. Real evidence can be further classified according to the use to which it is put, namely inceptive, exclusionary, corroborative, or associative.

Inceptive evidence is evidence that indicates a crime has been committed. Straightforward examples include detection of accelerant traces at a fire scene and identification of cocaine in someone's possession. Less straightforward is the finding of semen in samples taken from the vagina of a woman who states that she has been raped, as the evidence speaks to intercourse and not to absence of consent – perhaps a good instance around which to discuss the application of *ceteris paribus* to forensic science. Identifying semen in intimate samples from the alleged victim, and DNA profiling to link the semen to the alleged assailant is the first and core part of the forensic science investigation of a rape. But is it possible to have rape without semen – the use of a condom, or penetration without ejaculation for example, and possible to have semen identified as being from the suspect but without rape – if the court accepts that the intercourse was consensual or that the semen was from a different act of intercourse that occurred some time before the act of rape.

Exclusionary evidence is to identity in forensic science as falsification is to hypothesis testing in science in general. There are many techniques available that speak to individual identity: fingerprints; DNA profiles; facial recognition; acquired characteristics such as tattoos, restorative dentistry, and skeletal fractures; and voice. These may be ranked in ascending order of how well they approach the critical test of knowing the identity of someone: voice, facial recognition, acquired characteristics, DNA, and fingerprints. There is a continuum from transient and variable voice, through subjective facial recognition (although computer programs are changing this), to acquired characteristics which may not be individual and unique, to DNA and fingerprints. Even the DNA and fingerprinting may fall short of what could be accepted as indisputable knowledge. The DNA profile maybe partial, and the latent print may be limited in its data content. Even if they are both complete there is considerable debate at present as to the uniqueness of any individual fingerprint, with little information on population frequencies, and in regard to even the extremely rare DNA profiles, some statisticians will argue that if there is a finite probability of something happening then sooner or later, it will (Gettinby 2012).

And of course "we don't know what we don't know" – how were the fingerprint or biological fluid rendering the DNA profile deposited? It therefore may be more scientifically sound to follow Popper and to say that if there are differences in DNA profiles or fingerprints, then can safely be concluded that this excludes a given individual as being the source, and if none are observed, then the conclusion is that the target person or object cannot be excluded as being the source.

Some of the most compelling examples of exclusionary evidence have come from the Innocence Project (2011) in the United States. As of March 2011, there have been 267 post-conviction exonerations in the US arising from the application of DNA testing, the first being in 1989. The true suspects or perpetrators were identified in 117 of the cases.

Corroborative evidence is what remains when there is no exclusion. Corroborative evidence is very similar to *associative evidence*, which seeks to establish the relationship between things – people with other people, people with places, and objects with other objects. Hairs, fibers, paint glass, soil, some examples of impression evidence, are all examples of areas of forensic science that have traditionally contributed to corroborative evidence. They are also areas that are being encountered less and less frequently because of the difficulty of attaching a weight to them, due to the absence of databases to provide reliable frequencies of the different kinds of corroborative evidence in the population.

The difference between exclusionary evidence and corroborative evidence is mirrored in another way of looking at some forensic evidence, namely identification and class characteristics. Everyday language uses identity and its derivative "identification" to mean "uniqueness" and the act of "assigning a unique identity" to something or someone. This usage does not correspond however to that found in forensic science, which distinguishes between "identification" and "individualization" thus: identification means that items share a common source or possess the same properties, and individualization means coming from a unique source.

The Latin roots of the two clarify the difference: the root of identify (identification) is *idem*, which means "the same" whereas the root of individualize is *individus* meaning "not divisible." However, the distinction is not absolute and just when "same" becomes "unique" can be unclear. For example the mechanisms used to fire a bullet and to extract the spent cartridge case from a gun impart markings on the case; imperfections left on the lands of the gun barrel during manufacture and the relationship of the lands to the grooves can be used to associate spent bullet cases and recovered projectiles with the weapon that fired the round. Some of these such as the number of lands and grooves and their twist are characteristic of all examples of a particular model of gun. Others such as the imperfections created by tools when machining the barrel will be specific to an individual gun of that model. However, wear and tear and further imperfections created by dirt in the barrel mean that the features specific to a particular gun will change in the course of time.

It may be of value to refer back to some of the classic scientific activities of the nineteenth and early twentieth centuries, where some of the advances in science were related to classifications within the natural world, that is, with taxonomy. The taxonomy of forensic science can be reduced to two levels: What? and Which or Who? The question "what is this" together with the question of "who is this" (which one of all of the higher taxonomic level) can be simplified into just one – identity. The answer to the WHAT question will be the correct identification of something as being *one of a specified group* of things and the answer to the WHO (or which) question will be the correct identification of a *unique member* within that group.

This differentiation can be illustrated by returning to Popper and his swans. If there is sufficient agreed information to define what a swan is – the highest forensic taxonomic level – then it is possible to test the hypothesis of whether or not something is a swan.

Popper's example uses one property of "swanness," color. Note that Popper does not argue that all white birds are swans, but rather that one of the properties of being a swan is to possess white plumage. If the proposition is extended to be "a large waterbird with a long flexible neck, short legs, webbed feet, a broad bill, and all-white plumage" the finding of a bird with all but one of these would have three possible outcomes: the bird would be classified as other than a swan; the bird could be classified as a new species within the Cygnus (swan) genus; or the variant would be dismissed as an anomaly.

The second taxonomic level is to identify which swan – no two will be identical (birds don't have identical twins), and to do this requires information beyond the core properties. The conclusions are contextual and depend on the question being posed. If the question is "WHAT" is this, then the object is a swan if it possesses the agreed core properties. If the question is akin to "WHO" is this, which swan, then something other than the core properties is required to confirm its individual identity. How much more is one of the major issues in forensic science.

This approach makes it apparent that the question of identity is a matter of graduation on a continuum between identification of class to unique individualization. Between these two points there are an infinite number of degrees of identification, and many opportunities for these to vary with time.

As an example, a tire - print is located in the snow at a scene in northern Denmark. A cast is made and the size and general brand of the tire determined. The tread pattern shows uneven wear and an area of damage, probably a cut. Some months later investigators identify a suspect whose car has tires of the same make, and one of which has similar wear and damage patterns. However, because of the amount of driving in the time interval and the problems of capturing fine detail in marks made in snow, there will be differences, at the same time as areas of correspondence between the tire and the print. It is therefore not always possible to conclude that the suspect car is the vehicle involved in the incident, but neither can it be concluded that it is not. This is an example of corroborative evidence that is positioned at an intermediate point in the continuum from what to who.

The above discussion leads to the conclusion that the characteristics of good forensic evidence are:

- It should be stable and not change with time.
- It should have well defined core attributes and known population statistics.
- The basis of variations in non-core attributes should be known and scientifically justifiable.

The above discusses the examination of exhibits in the context of the goal of identifying, usually as "WHAT" or "WHO." There is another framework that is widely used in practice, namely the Locard Exchange Principle and the use of comparison testing. The Locard exchange principle, summarized "as every contact leaves a trace", is the foundation of traditional associated evidence based on microscopy. The concept is simple and can be illustrated as follows: Mr. A met Miss B in a bar one evening: they knew each other slightly because Mr. A was a friend of a distant relative of hers. The conversation was amicable and Miss B had no problem with accepting Mr. A's offer to escort her home when the bar closed. The next morning Miss B's body was found in a river that ran through the park that was on the way between the bar and her home. She had been strangled and raped. This happened some years before the introduction of DNA testing and although blood grouping on the semen from her vagina showed it to be the same as Mr. A's, it was a group O which is found in 50 % of the population. Mr. A was detained on suspicion and his clothing examined microscopically along with that of Miss B. Fibers found on her coat could have come from the jacket of Mr. A and fibers found on Mr. A's underpants could have come from the skirt worn by Miss B.

This two-way exchange of fibers is exactly what is entailed by the Locard exchange principle. In considering the weight to be attached to the findings the exchange from the outer clothing is of very little significance given the known and admitted social contact in the bar. However the exchange between the skirt and the underpants is an entirely different matter.

This simple and to some extent simplistic example is taken from a real case in which the author was involved. The complexities of the Locard exchange principle and the reason why it is now not so significant in most laboratory investigations, relate to the dynamics of fiber exchange, loss, and redistribution, to the difficulty of attaching an objective meaning to the relatively few possibly transferred fibers recovered compared to the many orders of magnitude more present from the background population, and to the lack of sound databases regarding the frequencies of the types of fibers encountered.

The question was posed earlier of how much information is required in regard to the move of identity from "WHAT," or class, to "WHO," or individualization. The same issue of how much is enough applies to comparison testing. However, the fundamental aspect of individualization is that the core properties that define "WHAT" are known and if they are not present then the person or thing is not a member of the class of interest. Comparison testing does not hinge around these core properties, and professional judgment has to be employed to decide if two items could have come from the same source. This is a major shortcoming in comparative identity.

Beyond What and Who: The Science Behind the Forensic Process

The European Network of Forensic Science Institutes (ENFSI 2002) has defined the Forensic Process by means of a set of standards covering activities from the time of the initial actions at the scene to reporting the results of the investigation and testifying in court. They are:

(a) Undertaking initial actions at the scene of crime
(b) Developing a scene of crime investigation strategy
(c) Undertaking of scene of crime investigation
(d) Assessment of scene of crime findings and considering further examination
(e) Interpreting and reporting findings from the scene of crime
(f) Laboratory examination, testing and presumptive testing
(g) Interpretation of the result of examinations and tests
(h) Reporting from examinations and tests including interpretation of results

The standards involve planning, identifying and preserving potential exhibits, conducting tests, and, at (e) and (g), evaluation and interpretation of what has been done. Considering whether or not these are scientific activities, requires consideration of parts of the section discussing what science is. As discussed by Hald, even though they followed the highly deductive and anthropocentric model of advancing knowledge based on a foundation of accepted societal beliefs, the ancient Greeks and Romans developed the skills of rhetoric to the point at which logic, evidence and argument became refined and effective tools for decision-making. The integral process, due to the Greek rhetorician Hermagoras but described in Latin by St Augustine, is building the argument round the questions *quis, quid, quando, ubi, cur, quem ad modum, quibus adminiculis* (who, what, when, where, why, how, in what way and by what means/which aids). These are questions that cannot be answered "Yes" or "No" but demand that explanations based on evidence be given.

Lying at the opposite end of the time frame from the ancients, Thomas Kuhn's concept that science can be defined as something that scientists do provides a link between social behavior, facts, and rhetoric. Accepting that there is indeed an activity called "forensic science" and that it is conducted by crime scene examiners and chemists and biologists working in places called "forensic laboratories" (and the equivalents of these in the several other subgroups of forensic science described in the introduction), then what happens in forensic science can be measured against the requirements of the scientific method, and the evidence-based objectivity of Hermagoras' questions, or, simply "The 6 Ws":

- WHAT has happened
- WHERE did it take place
- HOW did it happen
- WHO was involved
- WHEN did it happen – at what time and in what sequence
- WHY did it happen

The principle underlying the maxim is that each question should elicit a factual answer. What is important is that none of these questions can be answered with a simple "yes" or "no." They therefore present the incident as a problem to be solved. Solving the problem by providing factual answers to each dimension – each question – is the purpose of investigative inquiry.

The 6Ws define a process (of inquiry) to be used within a process (the ENFSI forensic process), and can be related back to the discussion on the types of evidence that forensic science can furnish. The examples below are from scene examination, because this is the place where Peirce's surprising event begins, and abduction is a necessary part of the process. However, the same principles apply in all stages of the forensic process.

Inceptive evidence was discussed in relation to information regarding specific crimes (arson and drug possession). The first of the 6Ws – What? – is a broader exposition of inceptive evidence. Was a crime committed or not, and if so, was it planned or accidental? In many cases this may appear to be a trivial matter – a body concealed in a shallow grave, with a gunshot wound to the temple. Or a husband calls, pretty agitated. His wife has fallen from the fourth floor window. Ambulance and police are dispatched to the scene and find her lying dead on the footpath, wearing underwear, socks, pants, sweater and rubber gloves of the type normally used for household cleaning activities. She had several injuries that could have been due the fall from the window.

Applying *ceteris paribus* to these cases suggests that the first demands to be treated as murder and the second as accident. But "we don't know what we don't know," and have to challenge the hypotheses. In the apparent murder, the questions of how did the body get there, who else was involved, and why all had to be pursued and led to the discovery that it was a case of suicide and concealment of the body. The murder hypothesis was falsified by exclusionary evidence, when it was shown that the blood patterns and bullet trajectory at the scene where the shooting occurred – not where the body was found – could only have arisen from a self inflicted shot. The accident hypothesis in the other case was firstly challenged by finding blood inside of the gloves – a surprising event – that led to the discovery of widespread residues of blood in the shower in the house, providing corroborative evidence to the conclusion of murder and not accident. The woman had been bludgeoned to death by her husband.

The murder/suicide case also demonstrates one aspect of the second of the 6Ws – Where? Knowledge of where the incident took place is essential in any investigation. It establishes the boundaries within which to search for evidence and can lead to other layers of interest, for example to the primary scene in the first case above.

The 6 W question of "How did it happen" is both obvious and hidden in the murder case, obviously by a fatal gun shot but hidden that it was self-inflicted. The case of the fallen wife illustrates a different aspect of "How" since the injuries themselves could be explained by the fall, and it was the additional information from the blood inside in the glove – exclusionary or falsifying information in regard to the accident hypothesis – that led to the correct conclusion.

"Why" is the most intangible of the 6Ws. It is perhaps less valuable in recreating events than in bringing a sensible closure to a case. In the suicide/concealment case, the deceased shared a house with others who like him were illegal migrants. His friends found him dead when they returned from work and feared the consequences of reporting the death, hence the concealment of the body.

If It Is Science, Does It Matter?

The development of a sustainable description of what science is – the scientific method and its modern refinements – together with the rigorous scrutiny of the steps in the forensic process using the ancient tool of rhetoric transcribed into the 6Ws, establishes forensic science as being a well-grounded scientific discipline. The question was not properly addressed in the NAS report but set aside in favor of a discourse on methodologies and statistical interpretations. Unlike the NAS committee, the US Supreme Court did try to come to terms with "science" in Daubert –v-Merrel Dow, as did others. However, both made the mistake of assuming that "scientific" somehow meant "reliable and absolute."

This paper is about the scientific basis of forensic science. The first part shows how science is not absolute but is ever-evolving and contextual. Lakatos remains the thought leader in both areas. The limitations that *ceteris paribus* places on the formulation and evaluation of hypotheses have been described, but Lakatos was also a strong proponent that taking a broader view of Popper's ideas of falsification is a vital part of the advancement of science. He created the concept of research programs to describe how exploration of real and apparent falsifications strengthen by providing a framework for the continuous iteration of ideas and refinement of knowledge. And so it should be with forensic science, the formulation of a hypothesis and its challenge, and the understanding that falsification of the initial hypothesis is not an end but a beginning, as new testing and a new examination of the existing information leads to a better hypothesis that perhaps approaches knowledge.

However, two vital parts of the work of the forensic scientist are the reliability of the testing methods and the validity of conclusions drawn from observations and tests. These have nothing to do with science any more than the quality of the prosecution or defense in a trial have. Quality assurance in chemical, biological and physical testing is a well-worked topic (Tilstone 2012), not covered here, but are dealt with in the NAS report somewhat (but not entirely) better than it dealt with science.

A more subtle issue is that of confirmation bias. It has long been a tenet of forensic science that findings based on observation, that is, findings that are subjective, must be confirmed by an independent observer to add a degree of objectivity. The matter of objectivity has been dealt with as an element in quality assurance, by the definition of "Objective test" and a description of how to ensure objectivity in the guidance document for forensic accreditation (ILAC 2002), but recent work shows that the confirmation itself can be subjective, and systems approaches have been proposed to overcome the problem (NIST 2012). Aside from subjectivity and the breakdown (or lack of implementation) of quality assurance procedures, bias is an inevitable consequence of an inductive approach based on asking a question that can be answered "Yes" or "No" and a rigorous implementation of the 6Ws together with effective quality assurance and a more systems approach to observer-dependent testing should resolve the problem.

Related Entries

- ▶ Automated and Manual Forensic Examinations
- ▶ Causes of Wrongful Convictions
- ▶ Cognitive Forensics: Human Cognition, Contextual Information, and Bias
- ▶ Crime Science
- ▶ Criminal Investigative Analysis
- ▶ Critical Report on Forensic Science
- ▶ DNA Profiling
- ▶ DNA Technology and Police Investigations
- ▶ Evidence-Based Policing
- ▶ Eyewitness Research
- ▶ Fingerprint Identification
- ▶ Forensic Anthropology
- ▶ Forensic Environmental Evidence
- ▶ Forensic Facial Analysis
- ▶ Forensic Linguistics
- ▶ Forensic Palynology
- ▶ Forensic Science and Criminal Investigation
- ▶ Forensic Science and Miscarriages of Justice
- ▶ Forensic Science and the Paradigm of Quality
- ▶ Forensic Science Culture
- ▶ Forensic Science Effectiveness
- ▶ History of Forensic Science in Policing

- ▶ Jury Decision Making and Eyewitness Testimony
- ▶ Legal Rules, Forensic Science and Wrongful Convictions
- ▶ Philosophical Basis of the Forensic Process
- ▶ Philosophy of Forensic Identification
- ▶ Probability and Inference in Forensic Science
- ▶ Scientific Evidence Before International Criminal Tribunals
- ▶ Scientific Evidence in Criminal Adjudication
- ▶ Scientific Evidence in Criminal Prosecutions

Recommended Reading and References

Daubert –v- Merrel Dow. Daubert v. Merrell Dow Pharmaceuticals, Inc, 509 U.S. 579, 113 S.Ct. 2786, 125 L. Ed.2d 469 (1993)

ENFSI (2002) Performance based standards for forensic science practitioners. http://www.enfsi.eu/page.php?uid=108. Accessed 22 Nov 2011

Gettinby G (2012) Personal communication

ILAC (2002) Guidelines for forensic science laboratories. International Laboratory Accreditation Cooperation, Silverwater

Innocence Project (2011) http://www.innocenceproject.org/Content/Facts_on_PostConviction_DNA_Exonerations.php#. Accessed 10 Mar 2011

Kuhn TS (1966) The structure of scientific revolutions, 3rd edn. University of Chicago Press, Chicago

Lakatos I (1970) Falsification and the methodology of scientific research programs. In: Lakatos I, Musgrave A (eds) Criticism and the growth of knowledge. Cambridge University Press, Cambridge, pp 91–198

NIST (2012) Expert working group on human factors in latent print analysis. Latent print examination and human factors: improving the practice through a systems approach. U.S. Department of Commerce, National Institute of Standards and Technology, Washington, DC

Peirce CS (1934) In: Hartshorne C, Weiss P (eds) Collected papers of Charles Sanders Peirce, vol 5. Harvard University Press, Cambridge, MA

Popper KR (1970) Normal science and its dangers. In: Lakatos I, Musgrave A (eds) Criticism and the growth of knowledge. Cambridge University Press, Cambridge

Tilstone WJ (2010) Quality in the forensic science laboratory. In: Mozayani A, Noziglia C (eds) The forensic laboratory handbook procedures and practice, 2nd edn. Humana Press/Springer, New York

Tilstone WJ, Hastrup M, Hald C (2012) Fisher's techniques of crime scene investigation, First International Edition. CRC Press, Boca Raton, 2012

Scientific Evidence Before International Criminal Tribunals

Melanie Klinkner
Business School, Bournemouth University,
Bournemouth, Dorset, UK

Overview

Over the past decades, a number of international tribunals and courts have been created to help bring international human rights and humanitarian law violators to justice, thus complementing the role of domestic courts. Substantive international criminal law comprises categories of crimes such as war crimes, crimes against humanity, genocide, the crime of aggression, torture, and terrorism (Zahar and Sluiter 2008). For an accused to be found guilty of any of these crimes, the alleged perpetrator must be found criminally liable either through having materially committed the crime or through their engagement in other forms of relevant criminal behavior.

International criminal prosecutions are often exceedingly complex, as evidenced by the overambitious prosecution approach taken in the case of former Serbian President Slobodan Milošević who was charged with over 7,500 crimes committed during three wars spanning 8 years. Expert witness testimonies can be paramount in providing specialist knowledge and contextual information. In international criminal trials, as in domestic English proceedings (??x-ref Roberts), expert witnesses "are normally allowed to testify on issues about which the judges themselves, based on their personal knowledge and experience, cannot be expected to reach an opinion alone" (Schabas 2006: 480). As the extensive experience of the International Criminal Tribunal for the Former Yugoslavia (ICTY) demonstrates, scientific expertise, especially relating to mass grave evidence, has been used successfully to prosecute the categories of crimes falling under its jurisdiction, and it is fair to assume that for future international criminal proceedings, the use of expert witnesses will continue to be imperative.

There are two aspects in particular that distinguish the presentation of expert evidence before international criminal tribunals from domestic proceedings. Firstly, investigations crucially rely on state cooperation, since international criminal institutions have no enforcement agency of their own (Del Ponte 2006). A delay in investigations can mean that much of the evidence, especially physical evidence, has vanished or deteriorated in condition (McGrath 2002). Witness testimony can be lost through death, intimidation, murder, or flight from an ongoing war zone, and access to war-torn countries can be both dangerous and difficult if international and domestic support is not forthcoming. Secondly, and more relevant to the discussion here, the mixed procedural model containing civil law and common law traditions produces a unique blend with few provisions relating to evidence presentation, creating a novel – and, to some, controversial (Murphy 2008) – environment for presenting and evaluating expert evidence. The following discussion provides an overview of the rules governing expert evidence presentation at both the ICTY and the International Criminal Court, before turning to consider how scientific expertise has assisted the ICTY in its fact-finding.

Expert Evidence Provisions at the ICTY and the ICC

Informed by the development of evidentiary rules, proceedings, and jurisprudence from previous international criminal tribunals, the ICC adopts a flexible approach to the admissibility of evidence. Article 69 of the Rome Statute states that "[t]he Court shall have the authority to request the submission of all evidence that it considers necessary for the determination of the truth" (Rome Statute, Art 69.3). The court adopts the affirmative, as opposed to exclusionary, method of evidence admission, taking into account in particular the probative value of the evidence and whether it is prejudicial to fair trial requirements (Rome Statute, Art 69.4).

Likewise, the ICTY's procedural law does "not contain a detailed set of technical rules" (*Prosecutor* v *Delalić* 1998: [15]). Rule 89 of the ICTY's Rules of Procedure and Evidence (RPE) states that evidence is admissible if it is relevant, of probative value, not to the detriment of a fair trial and not otherwise excluded as being obtained through methods that would be detrimental to its reliability or might damage the integrity of the proceedings (ICTY RPE, Rule 95). While the Trial Chambers may thus be inclined to admit scientific evidence, it does not follow that expert evidence will be afforded much, if any, weight in the court's deliberations. In fact, at the ICTY it has been decided that initial admissibility rulings may be reversed at later stages in the proceedings, as and when further information relating to the validity or reliability of evidence becomes available (*Prosecutor* v *Orić* 2004).

Expert Witness Status

To the dismay of some (Derham and Derham 2010), neither the ICC nor the ICTY has specifically defined "expert witness." The ICTY Trial Chamber in *Popović* described an expert witness as someone who possesses the relevant specific knowledge, experience, or skills to help the Trial Chamber come to a better understanding and conclusion on a technical issue (*Popović* 2007a). The qualifications of an expert, summarized in the expert's curriculum vitae submitted to the court, authorize the expert – unlike an ordinary witness of fact – to state opinions, inferences, and conclusions on matters within the realm of her expertise.

At the ICC, pursuant to Regulation 44(1), the registry holds a list of experts whose qualifications have been verified and "have undertaken to uphold the interests of justice" (*Prosecutor* v *Dyilo* 2007a: [24]). Where possible, the parties are expected to rely on this list for expert instructions. To facilitate efficient trial management, joint instructions (including by victim representatives) of expert witnesses are preferred.

While qualifications are regarded as formal prerequisites for a witness to qualify as an expert, objectivity and independence are not. Rather, “the questions of objectivity, impartiality and independence become relevant to assess the weight to be accorded to that opinion evidence” (*Prosecutor* v *Popović* 2007a: [26]). Affiliation with a party, which has been commonplace in relation to ICTY investigations, does not in itself constitute grounds for disqualification. Indeed, scientific experts were routinely employed by the ICTY’s Office of the Prosecutor, and without this arrangement, little scientific expertise or original physical evidence would have been available to the ICTY. The joint instruction procedure at the ICC should serve to ameliorate perceptions of witness bias, as the opposing parties will be bound to rely on the same expert. Neither the ICTY nor the ICC have Codes of Practice for expert witnesses, but one would expect any scientific expert to be independent and to act in good faith.

Pretrial Chamber and Expert Reports

Standard practice at the ICTY is to tender and admit expert reports through Rule 94 *bis* on the testimony of expert witnesses, which provides a timetable for disclosure and other preliminaries (*Prosecutor* v *Blagojević* 2003: [20]). At this stage, the opposing party is required to indicate whether it intends to accept the expert witness statement, desires to cross-examine the expert witness, disputes her qualifications, or challenges the relevance of the witness statement. Expert evidence can be denied admissibility on three grounds. It must be excluded firstly, if it has been “obtained by methods which cast substantial doubt on its reliability” (ICTY RPE, Rule 95); secondly, if it jeopardizes the fairness of the trial (ICTY RPE, Rule 89(D)); and thirdly, evidence may be excluded pursuant to Rule 89(C) because it lacks probative value. Of these three grounds for exclusion, Rule 95 has the greatest relevance for the work of scientific experts, as it directly addresses the expert’s methods of data collection and whether, in light of the way the scientific inquiry was conducted, its results are reliable.

Provided that no objection is made by the other side, a scientific report can be admitted into evidence without hearing testimony from the expert, so long as the Trial Chamber is satisfied as to the evidence’s relevance and probative value (*Prosecutor* v *Blagojević* 2003). In *Popović*, the ICTY elaborated on the application of the general requirements of relevance and probative value to expert reports, in terms of

> (1) whether there is transparency in methods and sources used by the expert witness, including the established or assumed facts on which the expert witness relied; (2) whether the report is reliable: and (3) whether the contents of the report falls [sic] within the accepted expertise of the witness. (*Prosecutor* v *Popović* 2007a: [30])

Qualification as an expert does not automatically guarantee the admissibility of the expert’s report. The burden lies on the party tendering the evidence to convince the tribunal that it satisfies Rule 89(C).

At the ICC too, the Trial Chamber has an important role to play regarding expert witnesses. A chamber not only can instruct an expert proprio motu (ICC Regulations of the Court, Reg. 44.4.), but the chamber can also determine the subject of an expert report, number of experts in the case, the way experts are instructed, and how they are to present evidence and within what time limits (Reg. 44.5). In *Prosecutor v Jean-Pierre Bemba Gombo*, the ICC Pretrial Chamber issued detailed instructions for the disclosure of all evidence, to ensure that the defendant receives the evidence to be disclosed by the prosecutor (ICC RPE, Rules 76 and 77) and has adequate time and facilities to prepare a defense. Only evidence that is of true relevance to the case should be disclosed. In addition, the prosecutor must supply “sufficiently detailed legal analysis relating the alleged facts with the constituent elements corresponding to each crime charged” (*Prosecutor* v *Gombo* 2008: [66]). This general requirement naturally applies to scientific evidence. Pages and paragraphs of expert reports or testimonies as well as photographs, physical evidence, and maps must be analyzed in the same way, contributing to a summary table of evidence. The Trial Chamber ultimately rules on the admissibility and relevance of evidence, including scientific evidence.

Admissibility of Summary Reports and Transcript Testimony

The ICTY Rules of Procedure and Evidence contain provisions to facilitate expeditious presentation of complex scientific evidence. Rule 92 *bis* RPE authorized, for example, the presentation of summary reports by investigators in relation to mass grave excavations and examinations. These summary reports are compilations, derived from multiple sources, containing background evidence of the forensic examinations, contextualizing and reducing the complexity of the findings (e.g., Manning 2000). While summary reports can save precious trial time, they may be challenged as hearsay evidence, which is generally admissible in international criminal proceedings but may be accorded little probative value (*Prosecutor* v *Milošević* 2002: [2](i)–(ii)). That said, investigator Manning's report on physical evidence recovered from Srebrenica execution points and mass graves used in *Krstić* was subsequently also found to be "highly relevant to the case and admissible under Rule 89" in *Prosecutor* v *Blagojević* (2003: [30]).

Similarly, Rule 92 *bis* (D) authorizes the admission of trial transcripts of evidence previously given by a witness, including expert witnesses, provided the evidence does not relate to the acts and conduct of the accused. In *Blagojević* statements and transcript testimony of numerous experts relating to mass grave, investigations were admitted in this way. The Trial Chamber was satisfied that the transcript testimonies submitted under Rule 92 *bis* (D), along with the expert reports received pursuant to Rule 94 *bis*, were relevant, probative, and together provided "a complete picture of the expert evidence" (*Prosecutor* v *Blagojević* 2003: [35]).

At the ICC, Article 69.2 of the Rome Statute and Rule 68 RPE govern the use of recorded testimony, including transcripts or other documented evidence as long as the measure is not "prejudicial to or inconsistent with the rights of the accused" (Rome Statute, Art 69.2). Where the witness is not present before the Trial Chamber, both the prosecutor and the defense must have had an earlier opportunity to examine the witness. Where the witness has consented for the materials to be used and is present before the chamber, the prosecutor, the defense, and the chamber must have the opportunity to examine the witness during the proceedings if they so wish.

Expert Witness Proofing

At the ICTY, prior to giving evidence in court, experts may review their testimony with the party calling them. Lawyers and expert witnesses may discuss the relevance of the scientific evidence to eliminate any misunderstandings. Such "proofing" may assure the expert witness as to their role in court, help recollect important elements of the evidence, and assist the lawyer in using the evidence effectively. While "rehearsing, practicing, or coaching" is not permitted, it was held that "reviewing a witness' evidence prior to testimony is a permissible practice under the law of the Tribunal and, moreover, does not per se prejudice the rights of the Accused" (*Proecutor* v *Milutinović* 2006: [22]).

In contrast, the ICC's Victims and Witnesses Unit (VWU) is considered the most appropriate organ of the court to help witnesses become familiar with the experience of giving evidence. The VWU is tasked, firstly, with reminding the witness to tell the truth; secondly, to provide the witness with a copy of their original statement as a memory prompt; and thirdly, to obtain copies of any other statements made by the witness and to ensure that witnesses do not discuss their evidence with the parties. In fact, the Trial Chamber in *Dyilo* emphasized that "the ICC Statute and Rules do not expressly provide for the possibility of parties preparing witnesses for testimony, and further finds no provision in the texts to justify the practice" (*Prosecutor* v *Dyilo* 2007b: [36]). This was in marked contrast to the ICTY's more relaxed approach to witness proofing.

However, in a January 2013 decision on witness preparation, Trial Chamber V ruled differently, suggesting that the silence of the Statute does not automatically imply that witness preparation is forbidden. The judges found that witness preparation is likely to "enhance the efficiency, fairness and expeditiousness of the present trial" (*Prosecutor v Ruto and Sang* 2013: [35]).

Testing Scientific Evidence

Fairness demands that the opposing party in an adversarial proceeding must be given the opportunity to test the evidence presented to the fact finder. This is enshrined in Art 67.1(e) of the ICC's Rome Statute giving the accused the right "to examine, or have examined, the witnesses against him or her and to obtain the attendance and examination of witnesses on his or her behalf under the same conditions as witnesses against him or her." Within the general framework of victim representation at the ICC, legal representatives for victims may also question an expert witness subject to Trial Chambers' decisions on what questions may be put to the witness.

In general, testing the reliability and credibility of scientific evidence may involve some or all of the following related issues: (a) the expert's qualifications and status as an expert, (b) the scientific methods adopted, (c) norms of practice, (d) acceptance within the scientific community and validation of methods through publications and peer review, (e) whether and how the science is produced for litigation, and (f) the novelty of the scientific evidence presented (Edmond 2000). Evaluations of expert evidence in international criminal proceedings tend to focus on professional competency, methodologies, and the credibility of the findings in context (Schabas 2006: 480). Rule 140.2(b) of the ICC's RPE specifies that both the prosecution and the defense have the right to question an expert witness about the reliability of the testimony provided, the credibility of the witness, and "other relevant matters."

The following pattern is commonly observed when expert evidence is given in court. After the solemn declaration, examination in chief begins with a discussion of the expert's education and qualifications, employment record, and relevant experience, before substantial matters relating to the scientific report are queried to demonstrate the credibility of the witness and the reliability of her evidence. This is followed by cross-examination by the opposing party and, if necessary, reexamination by the party who initially called the expert. In light of the technical nature of the evidence, the accused can request the presence of her own expert during such testimony to assist with cross-examination (*Prosecutor* v *Karadžić* 2011a). Judges at the ICTY may ask questions at any stage, while ICC judges may ask questions only before or after the witness is examined by a party. The defense has the right to be the last to examine the witness (ICC RPE, Rule 140.2(c) and (d)). Challenges to the credibility of experts have been frequent before the ICTY. Experts, such as Dr Haglund in *Prosecutor* v *Popović* (2007b), have been attacked on a personal level rather than being called upon to defend the substance of their evidence (Klinkner 2009). However, if the ICC's procedural mechanism for joint instruction of expert witnesses becomes well established, one would expect there to be little strategic incentive in trying to discredit an expert whose appointment has been agreed to by all the parties.

In terms of weighing the evidence, Sluiter and Zahar observe that "[c]ross-examination, giving evidence under oath, and the direct perception of the witness's demeanour are key factors in attaching weight to live testimony" (2008: 393). Fact finders will consider whether an expert witness appears honest, independent, and impartial, and must try to establish whether the scientific examinations were conducted according to appropriate scientific standards. Qualification as a bona fide "scientific expert" does not necessarily guarantee the quality of the work conducted in any particular case.

False Testimony and Misconduct

Like all witnesses, expert witnesses are under the obligation to tell the truth. False testimony under oath may result in an investigation, preparation of an indictment for willfully giving false testimony, and, ultimately, prosecution under Rule 91 of the ICTY's RPE or Article 70 of the Rome Statute. To convict an expert witness of false testimony or an offense against the administration of justice "requires the necessary mens rea and not a mere wrongful statement" (*Prosecutor* v *Akayesu* (1998: [140])). At both institutions, conviction may entail imprisonment or fines. Misconduct can also be charged under Article 71 of the Rome Statute, potentially resulting in interdiction by the court or a fine.

Needless to add, any of these sanctions would incidentally exact a huge reputational cost for the expert's professional standing.

Weighing Expert Evidence: Scientific Evidence from Mass Graves Before the ICTY

Experience from the ICTY demonstrates that scientific evidence, especially in relation to mass grave investigations, has been mostly uncontroversial and generally accepted. First and foremost, scientific evidence recovered from mass graves provides useful corroboration for eyewitness testimony. In Srebrenica, for example, evidence derived from execution points and graves matched accounts of the events by those lucky enough to escape. In the case of Dražen Erdemović, one of the first defendants at the ICTY, who pleaded guilty to murder as a crime against humanity, it was the accused himself who led the investigations to execution and burial sites which were not previously known to the Office of the Prosecutor (*Prosecutor* v *Erdemović* 1996: [135]). Excavations that took place at the Branjevo Military Farm revealed that there were 132 male victims in the grave, 130 of whom had died from gunshot wounds, and 83 ligatures were found in the grave. Furthermore, analysis showed that the Branjevo Military Farm mass grave had been disturbed. Individuals had been removed and placed in secondary graves (Manning 2000), indicating belated attempts at concealment.

Corroboration of evidence can assist with witness selection, especially as witnesses' memories may have faded or been affected by posttraumatic stress disorder and the passage of time. It is conceivable that some witnesses will suffer from memory loss or alteration by adding to their own painful memories details they heard from fellow sufferers, which poses challenges for the investigation. Furthermore, studies on eyewitness testimony have found that "recall of details from a violent incident was significantly worse than recall of a nonviolent incident" (Clifford and Scott 1978: 356). Contrary to the common sense assumption that most people would never forget the face of an individual who had physically confronted and threatened them, large numbers of participants in one empirical study were unable to identify the person responsible for their ordeal. This research provides "robust evidence that eyewitness memory of persons encountered during events that are personally relevant, highly stressful, and realistic in nature may be subject to substantial error" (Morgan et al. 2004: 274).

Scientific evidence in general, and forensic science in particular, helps to clarify the context surrounding the crimes and contributes towards proving what crimes were committed and how they were perpetrated.

War Crimes

Grave breaches of the Geneva Conventions (ICTY Statute, Art. 2) fall within the subject matter jurisdiction of the ICTY, while Article 3 provides the tribunal with the power to prosecute "violations of the laws or customs of war," including cruel treatment, torture, and murder. Scientific evidence can assist in confirming these charges. Findings presented in *Prosecutor* v *Mrkšić* (2007) from forensic examinations of the bodies retrieved from the Ovčara mass grave in Croatia showed that 198 were male and two were female, with an age range from 16 to 72 years. The cause of death in 188 cases was attributable to single or multiple gunshot wounds. Seven individuals were believed to have died from trauma, while the cause of death is still unknown for the remaining five victims. Postmortem examinations revealed that 86 individuals had suffered from injuries prior to their death on 20/21 November 1991. In 1997, it was possible to identify 192 of the victims buried at Ovčara. With the help of forensic science, the ICTY prosecutor had little difficulty in proving the crimes that had occurred at Vukovar. The tribunal was also satisfied that the victims who were taken from the Vukovar hospital on the morning of 20 November 1991 were at that time not taking part in hostilities and therefore could not be considered legitimate military targets.

Genocide

To constitute the crime of genocide (ICTY Statute, Art. 4), the accused must have deliberately intended to destroy a protected group in whole or part (Cassese 2008). Where direct evidence of genocidal intent is absent, the requisite intent may be inferred from the factual circumstances of the crime. Perhaps the most interesting case to date where a defendant was indicted for genocide partly on the basis of scientific evidence is *Krstić*. Radislav Krstić stood accused for his actions as Deputy Commander of the Bosnian Serb Army during the Srebrenica massacre between 10 and 19 July 1995. During the trial the prosecution called six forensic experts and two ICTY investigators to give evidence in relation to the mass grave investigations. The Trial Chamber found that the forensic evidence corroborated "important aspects of the testimony of survivors from the execution sites" (*Prosecutor* v *Krstić* 2001: [71]) and was sufficiently credible and compelling to confirm the *actus reus* of genocide.

The judges concluded that "following the take-over of Srebrenica, thousands of Bosnian Muslims were summarily executed and consigned to mass graves" (*Prosecutor* v *Krstić* 2001: [73]). The investigations suggested that most of the deceased had not been killed in combat, leading the judges to infer that some 7,000 missing persons had been executed and buried in mass graves. The Trial Chamber reasoned that the disappearance of generations of men showed an intent to physically destroy Bosnian Muslims as an ethnic group. Further indication of the intent to destroy the group, as such, was provided by a "well-established pattern" (*Prosecutor* v *Krstić* 2001: [68]) of executions. Bodies were not only concealed in mass graves, but were at a later time excavated in an attempt to hide the crimes. Expert examinations of seven secondary graves found commingled and mutilated body parts rendering identification efforts, repatriation, and appropriate burials extremely difficult, causing further distress to the survivors. The fact that all located and examined gravesites associated with the Srebrenica massacre were within the Drina Corps area of responsibility contributed to the Trial Chamber's overall belief that Krstić shared the intention to commit genocide (*Prosecutor* v *Krstić* 2001). The Trial Chamber was satisfied that Krstić had participated in the joint criminal enterprise, sharing the genocidal intent to kill Bosnian Muslims, and duly convicted him of genocide. On appeal, however, this verdict was overturned as the Appeals Chamber felt that the necessary intent to commit genocide was not proven beyond reasonable doubt.

In this case, scientific evidence helped to determine that (a) a specific group was targeted; (b) the killings and burials were systematic; (c) many civilians were amongst the dead; (d) demonstrable attempts had been made to conceal the crimes; and (e) a high level of cooperation was required to undertake such executions and burials.

The *Krstić* judgments and the forensic evidence presented during that trial have since been relevant to other Srebrenica cases, especially *Blagojević*, *Popović* et al., and *Milošević*. In December 2011, three of the forensic experts that appeared in *Krstić* gave evidence before the *Karadžić* proceedings with more experts scheduled to testify in early 2012 (*Prosecutor* v *Karadžić* 2011b).

Crimes Against Humanity

Numerous defendants have been charged with crimes against humanity (ICTY Statute, Art 5), mostly in relation to attempted "ethnic cleansing" of particular regions. In the *Popović* trial, where five of the defendants stood accused of extermination as a crime against humanity, the defense was keen to clarify whether those found in mass graves had been killed legitimately in combat or whether they were identifiable as civilians whose murder would constitute a crime against humanity. An expert witness, for example, was asked whether some victims from mass graves could have died as a result of combat as opposed to execution and whether military clothing was found on the bodies (*Prosecutor* v *Popović* 2007c). According to the expert, the evidence suggested that the dead had not been killed in combat as (a) they were not wearing military clothing; (b) the deceased were of all ages, some with physical disabilities; (c) blindfolds and ligatures were found in some graves; (d) many victims had been killed from behind by a single shot to the head; and (e) there was little indication of previous injuries consistent with combatant status. While it could not be fully excluded that some had been killed in combat, the majority of dead could not be accounted for in that way (also see *Prosecutor* v *Popović* 2007d). The Trial Chamber accepted these findings and was "satisfied with the reliability of the conclusions in relation to the cause of death reached in the Prosecution expert reports" (*Prosecutor* v *Popović* 2010: [619]).

In *Prosecutor* v *Milutinović* (2007, 2009), the accused were allegedly responsible for deportation, forcible transfer, murder (as a crime against humanity and a violation of the laws or customs of war), and persecution of Kosovo Albanians. Volume two of the judgment reviews the evidence relating to the alleged crimes, relying on much of the forensic evidence gathered from investigations conducted in Kosovo during 1999. In light of these findings, the Trial Chamber concluded that over 700 bodies originally buried throughout Kosovo during the NATO bombing campaign were secretly exhumed and transported to Serbia in an attempt to conceal them from citizens of the former Yugoslavia and from the international community (*Prosecutor* v *Milutinović* 2009). These clandestine operations led the Trial Chamber to believe "that the great majority of the corpses moved were victims of crime, as opposed to combatants or people who perished during legitimate combat activities" (*Prosecutor* v *Milutinović* 2009: [1357]). Forensic science evidence thus underpinned the Trial Chamber's conclusion that some of the deceased (who included women and children) were victims of crimes against humanity.

The Future of Scientific Evidence in International Criminal Proceedings

Critics say that the amalgam of civil law and common law approaches to the rules and procedures of evidence creates an "uncertain, obscure, and unworkable body of law that does not expedite proceedings, but offers numerous possibilities for parties to submit motions for the exclusion of evidence" (Zahar and Sluiter 2008: 394). Others contend that judicial liberality results in the admission of dubious evidence which prolongs proceedings and complicates the tasks of adjudication (Murphy 2008). A flexible approach to the reception of scientific evidence does, however, mean that judges are not required to assess scientific validity when ruling on issues of admissibility.

This uncluttered approach to determining the admissibility of expert evidence at the ICC and the ICTY in no way detracts from the critical importance such evidence may play in international criminal proceedings. In such complex cases, the probative value of scientific evidence is highly contextual. Questions of scientific methodology, of reliability, and of experts' credibility, objectivity, and impartiality are addressed on a case-by-case basis through the process of testimony in court and tested through cross-examination. It is then the judges' role to weigh the evidence presented and to arbitrate between reliable and unreliable, as well as scientifically valid and invalid, evidence. The slender body of procedural rules governing expert witness testimony reflects an implicit faith that effective communication between scientific expertise and the law is possible and that traditional legal procedures will facilitate accurate fact-finding in international criminal proceedings.

Related Entries

- ► Expert Evidence and Criminal Trial Procedure
- ► Scientific Evidence in Criminal Prosecutions

Recommended Reading and References

Clifford BR, Scott J (1978) Individual and situational factors in eyewitness testimony. J Appl Psychol 63:352–359

DelPonte C (2006) Investigation and prosecution of large-scale crimes at the international level – the experience of the ICTY. J Int Crim Justice 4:539–558

Derham R, Derham N (2010) From ad hoc to hybrid – the rules and regulations governing reception of expert evidence at the International Criminal Court. Int J Evid Proof 14(1):25–56

Edmond G (2000) Judicial representations of scientific evidence. Mod Law Rev 63:216–251

International Criminal Court (2004) Regulations of the court (adopted by the judges of the court on 26 May 2004 The Hague) ICC-BD/01-01-04

International Criminal Court 'Rules of procedure and evidence' UN Doc ICC-ASP/1/3

International Criminal Tribunal for the former Yugoslavia 'Rules of procedure and evidence' UN Doc IT/32/Rev 42

Klinkner M (2009) Forensic science expertise for international criminal proceedings: an old problem; a new context; a pragmatic approach. Int J Evid Proof 13:102–129

Manning D (2000) Srebrenica investigations. Summary of forensic evidence execution points and mass graves, available at http://www.domovina.net/archive/2000/20000516_manning.pdf. Accessed 21 February 2013

McGrath R (2002) Problems of investigations into war crimes and crimes against humanity during and after ethnic conflict. In: Bassiouni MC (ed) Post-conflict justice. Transnational, Ardsley

Morgan CA et al (2004) Accuracy of eyewitness memory for persons encountered during exposure of highly intense stress. Int J Law Psychiatry 27:265–279

Murphy P (2008) Excluding justice or facilitating justice? international criminal law would benefit from rules of evidence. Int J Evid Proof 12:1–31

Prosecutor v Akayesu (Case No. ICTR-96-4-T) Judgment, 2 September 1998

Prosecutor v Bemba Gombo (Case No. ICC-01/05-01/08) Decision on the evidence disclosure system and setting a timetable for disclosure between the parties, 31 July 2008

Prosecutor v Blagojević and Jokić (Case No. IT-02-60-T) Decision on prosecution's motion for admission of expert statements, 7 November 2003

Prosecutor v. Delalić et al (Case No. IT-96-21-T) Decision on the motion of the prosecution for the admissibility of evidence, 19 January 1998

Prosecutor v. Erdemović (Case No. IT-96-22-T) Witness statement by ICTY investigator Jean-Rene Ruez, 19 November 1996

Prosecutor v Karadžić (Case No. IT-95-5/18-T) Decision on Accused's request for assistance of defence experts in the courtroom during testimony of seven prosecution expert witnesses, 27 October 2011a

Prosecutor v Karadžić (Case No. IT-95-5/18-T) Trial transcripts, 1., 2., and 8. December 2011b

*Prosecutor v. Krstić (*Case No. IT-98-33-T) Judgment, 2 August 2001

Prosecutor v Lubanga Dyilo (Case No. ICC-01/04-01/06) Decision on the procedures for instructing expert witnesses, 10 December 2007a

Prosecutor v. Lubanga Dyilo (Case No. ICC-01/04-01/06) Decision regarding the practices used to prepare and familiarise witnesses for giving testimony at trial, 30 November 2007b

Prosecutor v Milošević (Case No IT-02-54-AR73.2) Decision on admissibility of prosecution investigator's evidence, 30 September 2002

Prosecutor v Milutinović et al (Case No. IT-05-87-T) Decision on Ojdanić Motion to prohibit witness proofing, 12 December 2006

Prosecutor v Milutinović et al (Case No IT-05-87-T) Expert witness testimony by Dr Eric Baccard, 19–20 February 2007

Prosecutor v Milutinović et al (Case No IT-05-87-T) Judgment, 26 February 2009, vol 2

Prosecutor v Mrkšić et al (Case No. IT-95-13/1-T) Judgment, 27 September 2007

Proescutor v Orić (Case No. IT-03-68-T) Order concerning guidelines on evidence and the conduct of parties during trial proceedings, 21 October 2004

Prosecutor v Popović et al (Case No. IT-05-88-T) Decision on defence rule 94 bis notice regarding prosecution expert witness Richard Butler, 19 September 2007a

Prosecutor v Popović et al (Case No. IT-05-88-T) Expert witness testimony by Dr William Haglund, 15 March 2007b

Prosecutor v Popović et al (Case No. IT-05-88-T) Expert witness testimony by Dr John Clark, Transcript, 20 February 2007c

Prosecutor v Popović et al (Case No. IT-05-88-T) Expert witness testimony by Prof Richard Wright, 21 February 2007d

Prosecutor v Popović et al (Case No. IT-05-88-T) Judgment, 10 June 2010, vol 1

Prosecutor v Ruto and Sang (Case No. ICC-01/09-01/11) Decision on witness preparation, 2 January 2013

Rome Statute of the International Criminal Court (adopted 17 July 1998, as amended 16 January 2002) UN Doc A/CONF.183/9

Schabas WA (2006) The UN International Criminal Tribunals: the former Yugoslavia, Rwanda and Sierra Leone. Cambridge University Press, Cambridge

Zahar A, Sluiter G (2008) International criminal law. A critical introduction. Oxford University Press, Oxford

Scientific Evidence in Criminal Adjudication

John D. Jackson
School of Law, University of Nottingham,
Nottingham, UK

Overview

The different functions of science and law have long led to tensions between methods of investigation deployed within the sciences and the various methods of legal fact-finding that are deployed for determining legal disputes. These tensions become particularly acute in relation to scientific evidence in criminal adjudication. On the one hand, the courts have become increasingly reliant on scientific evidence; on the other hand, they cannot for the sake of their own legitimacy be seen to hand over the evaluation of such evidence entirely to scientists. The traditional "adversarial" and "inquisitorial" models of adjudication have each adopted different approaches towards the gathering and evaluation of scientific evidence which have come under critical scrutiny. The use of scientific evidence in "adversarial" systems would seem to have been particularly susceptible to miscarriages of justice while its use in "inquisitorial" systems has led to human rights violations. Against this background, it may be suggested that one way forward is to realign existing procedures in the direction of a participative model of criminal adjudication whereby *both* parties *and* adjudicators play a role in ensuring that scientific evidence is effectively prepared, presented, and tested.

Scientific Evidence in Criminal Adjudication

The growth in the use of expert or scientific evidence in legal disputes poses challenges to the traditional manner in which legal claims have been resolved. It is noteworthy how many issues of fact are increasingly coming within the domain of expert or scientific evidence (Redmayne 2001). This "creeping scientization of factual inquiry" (Damaška 1997: 143) has arisen as a result of science's ability to generate knowledge by means of sophisticated technical instruments well beyond the knowledge of laypersons. Equally significant has been the growth in the sciences of the mind which has led to the recognition of new forms of psychological evidence such as post-traumatic stress disorder, battered woman's syndrome, and recovered memory syndrome. Even the task of assessing the credibility of witnesses which has long been considered best left to the common sense of the lay judge or jury has been challenged by the growth of experts willing to testify on witness credibility and by such "scientific" techniques as polygraph evidence.

This increasing intrusion of scientific evidence into both civil and criminal adjudication can conjure up a somewhat apocalyptic vision of legal fact-finding processes being replaced altogether by scientists and experts. There are certainly instances where challenges have been made. Challenges to the traditional common law methods of eliciting testimony by means of examination and cross-examination were to be seen in the work of early twentieth century experimental psychologists (see Muensterberg 1908). Although, more recently, legal systems have refined their approaches towards certain kinds of witness testimony, such as the evidence of children and other vulnerable witnesses, behavioral science challenges have not prompted any fundamental review of legal procedures (Greer 1971). Indeed the need for testimony to be tested orally by means of cross-examination would seem to have gained rather than diminished in importance in criminal adjudication, not only in common law processes but in civil law processes as well – partly as a result of the human rights fairness requirement that everyone charged with a criminal offense has the right "to examine or have examined witnesses against him" (see Art. 14(2)(f) of the International Covenant on Civil and Political Rights, Art. 6(3)(d) of the European Convention on Human Rights).

This illustrates that scientific procedures or "scientific method" can never dictate how fact-finding should be conducted in the course of criminal adjudication, even though the same processes of inductive reasoning may be used to reach conclusions in both domains (Twining 1982; Jackson 1988). However dependent legal processes become on scientific evidence, fundamental differences between the scientific enterprise (the pursuit of truth) and the legal enterprise (the resolution of disputes through justice) mean that different methods are required to produce the desired results in each field, scientific discoveries, or legal verdicts. It is in the nature of the scientific endeavor constantly to seek better evidence, to be prepared to revise even the most entrenched claim in the face of unfavorable evidence (Haack 2004; and for the different values within science and law see Schuck 1993 and Goldberg 1994). This spirit of inquiry is incompatible with the constraints of legal inquiry where decisions have to be reached in the interest of finality (Redmayne 1997; Jackson 2004). As a former English Law Lord, Lord Scarman, once said, Justice cannot wait upon the truth; as the famous legal maxim goes, "justice delayed is justice denied."

This means that legal procedures cannot simply replicate scientific modes of inquiry. Irrespective of the accuracy of legal procedures, errors will inevitably occur and political decisions which may be highly contestable need to be reached as to how the risk of errors should be distributed. In criminal adjudication there is a consensus across the common law and civil law traditions that proof of guilt must be established beyond reasonable doubt (*in dubio pro reo*, as the principle is known in continental jurisdictions) which reflects a bias in favor of false negatives (acquittals of the guilty) over false positives (convictions of the innocent), although there are differences across the traditions as to how formidable the evidentiary barriers should be for prosecutors seeking to establish guilt (Damaška 1973). The need for finality, which can have profound consequences for those who are found guilty, correspondingly engenders expectations of fair and transparent procedures allowing defendants to contest the charges against them in an open forum and requiring that decisions are fully justified by an independent and impartial tribunal. Because of the "tragic choice" inherent in legal proceedings, whereby decisions need to be sufficiently appeal-proof even when they may be wrong in order to withstand continuing uncertainty (Nobles and Schiff 2000), the process of reaching decisions becomes a vital legitimizing tool in determining their social acceptability. Although different models of criminal adjudication have developed across legal systems expressing varying societal conceptions of adequate proof, there is a consensus across western countries that they must not only provide for accurate procedures but must also secure public acceptability through fair and open procedures.

The ambivalent attitudes that continue to be displayed towards science and technology suggest that it will be some time before "men and women in white coats" are trusted to replace lay decision-makers in the making of final decisions (Roberts 2008: 324). Scientific evidence needs to be carefully regulated and evaluated so that its probative benefits can be harnessed to assist the lay decision-makers who make final determinations in individual cases.

Adversarial and Inquisitorial Models of Adjudication

Comparative literature on criminal procedure and evidence has been dominated by a tendency to describe different criminal justice systems as either "adversarial" or "inquisitorial," with the term "accusatorial" often being used synonymously with "adversarial." Although these labels have enjoyed long currency, they have tended to be used as caricatures which often obscure rather than enlighten how different legal processes might be best understood. It is misleading, for example, to use the terms "accusatorial" and "adversarial" interchangeably to describe Anglo-American systems of criminal justice as the old accusatory forms of justice (characterized by spitting off the prosecutorial function from the final adjudication) have very little in common with contemporary "adversarial" trials, where accused persons are properly equipped through legal representation to contest the charges against them in an effective manner (Vogler 2005). Similarly, the term "inquisitorial" with its overtones of secret procedures dominated by authoritarian judges is hardly an apt description of the modern systems of justice to be seen on the European continent which replicate many of the "adversarial" features of Anglo-American trials.

Rather than use the terms as abbreviated descriptions of actual procedures, they have more helpfully been used as models to idealize opposing features to be found within the generality of Anglo-American and continental procedures. Although there is some difficulty in determining exactly which features should fit within the respective models, the essence of the contrast between adversarial and inquisitorial models of proof seems to boil down to this. Adversarial proceedings are organized around the notion of a dispute or contest between two sides – prosecution and defense – in a position of theoretical equality before a court which must determine the outcome, whereas inquisitorial procedures adopt the notion of an official and thorough inquiry driven by court officials. Thus, in the "contest" model, the prosecution prepares the case, brings the charge, and is responsible for presenting the evidence and proving the offense. The defendant either pleads guilty or attempts to rebut the charge by presenting evidence and arguments against the prosecution. The proceedings are presided over by a neutral adjudicator whose function is to see that the parties play by the rules of the contest but not to take an active part in the presentation of the evidence. In order to ensure that the parties play fair, there is a complex body of rules ensuring that relevant evidence is disclosed, particularly by the prosecution to the defense, in advance of the trial and that evidence is only admissible at the trial if it has sufficient probative value. In the "inquest" model court officials take center stage in the handling of the evidence. The prosecution may first decide the charge but officials of the court then have the responsibility for gathering, testing, and evaluating the evidence. A case file is built up containing all the documents relating to the proceedings and all the evidence that is gathered. This file can be consulted by the parties, but any role the prosecution and defense play in the proof process is incidental and subordinate to the court's primary function of finding the truth. The trial that follows the proof-gathering phase is largely based on material already in the file and this phase of the proceedings is also dominated by the court rather than the parties.

Although these models are considered to be a valuable heuristic tool for theorizing about different features of Anglo-American and continental processes of proof, more and more "real life" systems have been moving away from the models' classic features and (on some accounts) converging towards a "middle position" mixing features of both "adversarial" and "inquisitorial" traditions (Markesinis 1994: 30). For example, in the Swiss system, which is commonly perceived to fall within the inquest model, judicial authorities no longer take center stage in the handling of evidence. In Switzerland the investigating judge is now the prosecutor and, in common with a number of other continental systems, the position of the investigating judge has been abolished altogether. In the pre-trial process the prosecutor arranges confrontation hearings and is responsible for conducting them. There is no impartial third party present – just the prosecution and defense – and the defense has the chance to question the witnesses. Another novel feature in many "inquisitorial" systems is that the prosecution and defense may now reach an agreement over the outcome of the case, with the result that the court no longer plays any meaningful role in fact-finding at all (Weigend 2008). One distinguishing feature of common law systems, by contrast, is that for all the possibilities for plea bargaining, the trial is still generally the only forum where evidence can be effectively challenged by the defense. Within civil law systems, on the other hand, the defense has the opportunity and is expected to challenge evidence before trial in the various pre-trial phases, even though in many instances there may not be a judge present to oversee fairness in the process. This suggests that the contrast between criminal procedure systems is better found at the stage of the process at which "adversarial" evidence-testing takes place, rather than in any fundamental or wholesale difference between "adversarial" and "inquisitorial" proceedings.

For all this, the two models appear to retain some explanatory force if they are broadened out to include different political structures of authority and different legal cultures, for they can then be viewed as two different procedural traditions that have dominated the common law and civil law worlds, respectively, encompassing different normative expressions of how legal proceedings *should* be organized (Langer 2004; Field 2009). Jurists and practitioners within common law and civil law systems tend to identify strongly with idealized procedural models expressing different values. For example, there is a long-standing tradition of anti-inquisitorialism in American law which has acted as a kind of "negative polestar" for American criminal procedure (Sklansky 2009), while "anti-adversarialism" has tended to dominate debates surrounding many criminal justice reforms in France, such as those strengthening defense rights, the movement towards plea bargaining, and the role and function of the *juge d'instruction* (Hodgson 2005).

The Role of Scientific Evidence Within the Adversarial and Inquisitorial Procedural Systems

With the caveat that the adversarial and inquisitorial models may be more useful in reflecting beliefs about how procedural systems *should* be organized than as descriptions of how they actually *are* organized, we can begin to detail how scientific evidence is treated within each of the dominant procedural families (see generally Champod and Vuille 2011). Within the adversarial model, experts are employed by the parties in the pre-trial and trial phase of the criminal process and the court does not become involved in the appointment and use of experts except during the sentencing phase when it does make use of experts whom it appoints to assist in determining risk factors relevant to imposing penalties on convicted defendants. Before this phase, the parties employ experts to gather scientific evidence and write reports. If experts give evidence at trial, they are called as party-instructed witnesses.

The principle of party presentation by which parties call "their" own witnesses at trial extends to expert as well as lay witnesses, although no party has any exclusive property right in the evidence or any right to prevent another party from calling any witness (*Harmony Shipping Co SA* v *Saudi Europe Line Limited* [1979] 1 WLR 1380). Scientific evidence is tested in the same manner as other evidence is tested, through cross-examination at trial by counsel. Experts are treated like any other witness in this respect. However, expert witnesses are entitled to somewhat more leeway than ordinary witnesses in being allowed to state their opinion on matters within their expertise, and also to give hearsay evidence of scientific knowledge. Expert reports are often admitted as well by way of exception to the hearsay rule.

Although it is open to both sides to appoint experts, in practice defendants who are legally aided have to make a case to the legal aid authorities for scientific assistance. Even when this is forthcoming, defense experts have to overcome the structural and practical disadvantages of arriving late on the scene after prosecution experts have received items for analysis and conducted their tests and they may have to rely heavily on prosecution experts for information (Roberts 2002). Increasingly, prosecution experts are required to disclose all relevant information to the defense, while the defense is usually only obliged to disclose the scientific evidence it intends to adduce at trial.

In contrast with adversarial process, scientific experts in inquisitorial systems are in theory appointed by, and work under the supervision of, an examining judge or the court. They occupy a unique position in the criminal process, as they are neither witnesses nor judges. Increasingly, however, scientific expertise is called into the case by the police or by the prosecutor in cases where investigating judges are not involved. In reality, experts have often reached their conclusions in conjunction with the police investigation before the case reaches a judge (Nijboer 1993). The experts are usually appointed from lists or from accredited laboratories and have a status superior to that of witnesses. When they have completed their research and inquiries, experts submit a written report to the court and may be called to give oral evidence to explain their findings under questioning, which is mainly conducted by the court rather than by the parties, although counsel may be given permission to put questions directly to experts. Although the court must freely evaluate the expert evidence for itself, in practice expert opinions are afforded considerable deference.

The defense has traditionally had a limited input into the forensic process in inquisitorial systems, but the defense role has been expanding in more recent years. Defense lawyers sometimes request that particular questions are put to the expert or that certain tests are carried out (see e.g., Art 165 CPP France, Art 184 CPP Switzerland). They may comment on the expert's report once it has become part of the case file and they may request the appointment of another expert for a second opinion, although the courts can deny such petitions (see e.g., Art 167 CPP France, Art 244(4) CPP Germany). In certain systems, such as Italy (see Art 230 CPP), the defense may be able to appoint their own experts to work with the official expert but these party experts usually have inferior status to the official expert and permission may have to be sought before they are instructed (Champod and Vuille 2011: 21–22).

Miscarriages of Justice in Adversarial Systems

Although the use of scientific evidence has the potential to provide a more reliable evidential basis for verdicts than many traditional categories of evidence such as witness testimony and confessions, a litany of miscarriages of justice caused by faulty scientific evidence across a range of different forensic fields has come to light during the last 30 years, especially within the common law world. These have prompted criticism, not only of scientific evidence itself but also of systematic weaknesses in the "adversarial" model of proof. Various cognitive errors can arise when scientific evidence becomes so dependent on the parties (see generally Roberts 2002, Champod and Vuille 2011). First of all, there is the question whether it is wise to permit the parties to choose which experts to bring into the case. As one judgment over a century ago reflected: "A man may go and does sometimes to half a dozen experts . . . He takes their honest opinions, he finds three in his favour and three against him; he says to the three in his favour will you be kind enough to give evidence? And he pays the three against him their fees and leaves them alone . . . the result is that the court does not get the assistance from the experts which if they were unbiased and fairly chosen, it would have a right to expect" (*Thorn* v *Worthington Skating Ring Co*. LR 6 ChD 414, 416, 1876). Within the criminal process, the parties may not have the luxury of such a wide choice of experts, but this suggests a further problem. The parties need proper resources to be able to employ good experts which may not be available, in particular, to the defense relying on over-stretched legal aid budgets to secure expert evidence.

The traditional way in which adversarial systems have sought to protect themselves from partisan sources of evidence is by developing strict admissibility tests to screen out unreliable evidence and to give each party an opportunity to challenge the sources of evidence presented by their opponent. These approaches, however, have only had limited success in respect of expert evidence. Adversarial systems have developed a number of rules to restrict the admission of expert evidence. English law has traditionally taken a fairly deferential approach towards experts by allowing them to express expert opinions provided the subject matter of the opinion is beyond the knowledge, skill, or experience of the tribunal of fact and the expert has sufficient expertise in the field (Redmayne 2001). Although the position may now be changing, there is no established requirement that any specific threshold be met to ensure the reliability of expert evidence (Dennis 2010: 895). In practice this has meant that if an expert is accredited or has the necessary qualifications, and the subject matter of the expert's evidence is relevant to the case before the court, his testimony will be admitted (Alldridge 1999).

In 1923 in *Frye* v *US* (1923) 293 F.1013 a US federal district court appeared to impose stricter gate-keeping responsibilities on judges when it agreed that the trial court had been correct to exclude a polygraph technique because it had not achieved general acceptance in the relevant scientific community. This seemed to put a reliability threshold on the evidence by deferring to scientific consensus in the relevant field. A majority of courts in the US applied this standard and it appeared to creep into other jurisdictions as well (see, e.g., in Australia *R* v *Bonython* (1984) 38 SASR 45, *R* v *Parenzee* [2007] SASC 143). The difficulty with this approach was that it led to uneven application in practice. Courts applied varying degrees of rigor to the standard, with little indication as to which facts of the expert testimony had to be generally accepted and how widely or narrowly the relevant scientific community was to be defined (Redmayne 2001: 113; Ligertwood and Edmond 2010: 7.51–2). In *Daubert* v *Merrell Dow Pharmaceuticals, Inc.*, 509 US 579 (1993), the US Supreme Court considered that the *Frye* test had been superseded by the Federal Rules of Evidence in 1975 which made no mention of the *Frye* test but implied that a judge had to ensure that any and all scientific testimony or evidence admitted was not only relevant but reliable. *Daubert* is commonly considered to have replaced the *Frye* test in the US federal and many state jurisdictions with a reliability test involving consideration of five factors: (1) whether a theory or technique can be (and has been) tested; (2) whether the theory or technique has been subjected to peer review and publication; (3) the known or potential error rate; (4) the existence and maintenance of standards controlling a technique's operation; and (5) general acceptance in the scientific community.

The scope and application of the *Daubert* criteria have been clarified in later Supreme Court decisions. A sixth criterion was added in *General Electric Co* v *Joiner* 522 US 136 (1997) which entails examining the extent to which general findings are applicable to the instant case. In *Kumho Tire Co.* v *Carmichael*, 526 US 137 (1999) the Supreme Court broadened the applicability of these criteria to all expert evidence including that which is not strictly, or at all, "scientific." But *Daubert* remains the leading case and is now applied in federal courts in the US and has been cited with approval in certain other jurisdictions as well. In the UK, for example, the England and Wales Law Commission (2009) has recommended that *Daubert*-style criteria be introduced into the English courts.

The difficulty with requiring judges to take on an effective gate-keeping role in relation to expert or scientific evidence is that this depends for its success on judges having a sufficient understanding of the science involved. Although the test would seem to put explicit responsibilities on judges to test the validity of scientific evidence, it has been claimed that within the realm of forensic science many courts have continued to admit bite mark analysis, microscopic hair analysis, voiceprint evidence, and handwriting analysis despite the absence of any demonstrable theoretical basis, population databases, standardized methodology, or empirical data on error rates for this kind of evidence (Beecher-Monas 2007). A comprehensive report into the state of forensic science in the US concluded from reported opinions in criminal cases that, despite *Daubert*, trial judges rarely excluded or restricted expert testimony offered by prosecutors and that appellate courts routinely deny appeals contesting trial court decisions admitting scientific evidence against criminal defendants (National Research Council 2009). Judges are not helped in this endeavor by criminal defense lawyers who rarely have any background in relevant scientific expertise, whilst prosecutors have little responsibility in an adversary system to vouch for the reliability of the expertise they place before the court.

The other means of ensuring reliability of evidence in the adversarial system is to permit opposing parties to make an effective challenge to the evidence adduced. But this is only possible where the parties are properly resourced and there is sufficient disclosure of all the relevant evidence in the case. In certain notorious miscarriages of justice in England and Wales arising out of IRA terrorism in the 1970s, scientists failed to disclose important scientific information even to the prosecuting authority. The English Court of Appeal responded by requiring scientists advising the prosecuting authority to disclose material known to them which may have some bearing on the case (see *R* v *Maguire* (1992) 94 Cr App R 133). The difficulty is that, in the absence of effective oversight, investigators and forensic scientists are left to discharge these duties themselves and there may be little incentive for them to do so. Even if scientists are sufficiently objective in their approach, they may have received instructions which are incomplete and may not be alerted to all the relevant evidence in the case. When it comes to communicating their results and opinions, the requirement that they submit to examination and cross-examination by the parties means that questioning is focused on one direction or the other and expert witnesses may not be able to give as complete a picture to the court as they would like. The fact that experts are treated as witnesses also means that they are prevented from entering into open dialogue with decision-makers in the case, potentially frustrating the important educative role that experts have to explain their findings to the court (Allen and Miller 1993).

Human Rights Challenges in Inquisitorial Systems

Inquisitorial systems may be thought better able to withstand some of the cognitive weaknesses of adversarial criminal procedure. Under the inquisitorial trial model, scientific investigations are carried out under the supervision of an impartial court thereby facilitating the production of more impartial and reliable reports. Experts are also able to communicate the results of their findings to the court, whether in reports or by oral evidence, without being hampered by the artificial adversary process of examination and cross-examination. Problems of disclosure between the parties are pre-empted by having a case file made freely available to both the prosecution and the defense. Experts have full opportunity to respond to any issues raised by the parties or to call for a second opinion. Such a procedural model, however, is heavily dependent on the integrity and competence of "impartial" forensic experts. Scientific experts are perceived to be impartial because they act for the court rather than being instructed by adversarial parties, but in practice many experts work exclusively with the prosecuting authorities and this may put the defense at a considerable disadvantage. In some legal systems, such as that of the Netherlands, government forensic institutes are only permitted to investigate cases when requested to do so by the police or justice authorities (Jakobs and Sprangers 2000). If the defense wishes to commission expert evidence, it must usually pay for it and may encounter difficulties recruiting willing and able experts, particularly if it intends to challenge the findings of the experts appointed by the judicial authorities. Even if the defense is able to commission its own expertise, there is no guarantee that the court will afford the same weight to this evidence as to that produced by the official forensic scientists.

In a string of cases arising from inquisitorial systems, the European Court of Human Rights (ECtHR) has ruled that there was a violation of the right to a fair trial on the grounds that the defense had been disadvantaged under the principle of the "equality of arms" which requires that the prosecution and defense are put on an equal procedural footing. In *Bönisch* v *Austria* (1987) 9 EHRR 191 the applicant was prosecuted for preparing meat which contained an excessive concentration of benzopyrene on the basis of a report prepared by the Director of the Federal Food Control Institute who was appointed later as an expert by the court. The ECtHR considered that as it was his report that had led to the charges being brought, the Director of the Institute was more likely a witness against the accused than an independent court expert. The principle of equality of arms inherent in the concept of a fair trial required equal treatment as between the hearing of the Director and the hearing of persons who were or could be called in whatever capacity by the defense. Yet since the Director had been heard as an "expert" his statements must have carried greater weight than those of an "expert witness" called by the accused. As a formally recognized expert, the Director had enjoyed a privileged position in being allowed to attend throughout the hearings, put questions to the accused and witnesses with the leave of the court, and comment on their evidence. It followed that there had been a violation of the right to a fair trial (see also *Eggertsdottir* v *Iceland*, App. no. 31930/04, 5 July 2007; *Stoimenov* v *Former Yugoslav Republic of Macedonia*, App. no. 17995/02, 5 April 2007).

Bönisch is often contrasted with *Brandstetter* v *Austria* (1993) 15 EHRR 378, where the expert whom the court appointed was not the person who filed the report leading to the applicant's prosecution but was employed by the same Institute as the expert who had filed the report. The ECtHR held that, in this case, doubts about the neutrality of the expert could not be objectively justified and the Austrian courts had not breached the principle of equality of arms in refusing the applicant's request to appoint another expert. A court apparently does not have to appoint further experts at the request of the defense just because the opinion of the court-appointed expert supports the prosecution case. A second expert must be appointed to satisfy the requirements of the right to a fair trial only where there is some objective ground for suggesting that the expert was biased or, alternatively, some external factor suggesting that the expert's testimony may have had a distorting impact on the tribunal of fact (*GB* v *France* ECtHR 2 October 2001).

Apart from falling foul of the requirement of equality of arms, inquisitorial systems have also run up against the European principle of "adversarial procedure." This does not imply that European systems adopt fully fledged adversarial modes of proof (Jackson 2005). Instead the principle requires that the parties are given the opportunity to have knowledge and comment on the observations filed and the evidence adduced by the other party (*Brandstetter* v *Austria* (1993) 15 EHRR 378, *Rowe and Davis* v *UK* (2000) 30 EHRR 1, [60]). In *Mantovanelli* v *France* (1997) 24 EHRR 370 the ECtHR considered that the principle was violated because the applicants were not informed in advance of the dates on which an expert appointed to report on the circumstances of the death of the applicants' daughter interviewed witnesses and examined documents. The Court considered that there was no general right to be present during an expert's investigative activities. What was essential, however, was that the parties should be able to participate properly in the proceedings before the tribunal. Although the applicants could have made submissions to the court on the contents and findings of the report after having received it, the ECtHR was not convinced that the arrangement "afforded them a real opportunity to comment effectively on it." The question the expert was instructed to determine was identical with the one the court had to determine and pertained to a technical field that was not within the court's knowledge. Thus, the expert's report was likely to have a preponderant influence on the assessment of the facts by the court. In these circumstances the applicants could only have expressed their views effectively before the expert report was lodged. Since there would have been no practical difficulty in allowing them to do this, failure to provide advance notification constituted a violation of their ECHR Article 6 right to a fair trial.

Strasbourg case law has established a "two-pronged test" to ensure that criminal proceedings involving expert evidence are fairly conducted (Van Kampen 2000: 201). First of all, national courts must ensure that both parties are able to have knowledge of, and comment on, the evidence adduced by the opposing party and that they are able to question and challenge any court-appointed expert. According to one of the latest cases, *Mirilashvili* v *Russia* (2008) App no. 6293/04, 11 December 2008, [190], "if the court decides that an expert examination is needed . . ., the defence should have an opportunity to formulate questions to the experts, to challenge them and to examine them directly at the trial." This still falls short of any requirement that the defense be allowed to appoint its own expert to mount an effective challenge. If, however, objectively justified fears exist that the expert appointed by the court is not in fact acting with the presumed impartiality and neutrality, a second requirement is that national courts need to ensure that the defendant is provided with the opportunity to secure the attendance and examination of experts and/or witnesses on their behalf under the same conditions as the experts against them. This may require that the court appoints a counter-expert to ensure that the statements of the experts for the prosecution and defense are treated as probative equals.

Future Directions: A Participative Model

A recent report commissioned by the Council of Europe on the state of forensic science both globally and in the European Union has concluded that accusatorial and inquisitorial models alike do not lend themselves to the sound management of scientific evidence (Champod and Vuille 2011: 31). In those systems which give the parties prime responsibility for finding evidence, defendants who are already disadvantaged in the criminal process are penalized still further. Defense lawyers are ill-equipped to scrutinize prosecution expert reports. They may attempt to secure experts themselves but they frequently face difficulty in finding experts working in equivalent conditions to prosecution experts and even when defense experts can be found, they become involved at a comparatively late stage when critical samples have already been taken and they have to rely on information provided by the prosecution experts. The adversarial trial then requires judges and juries to assess scientific evidence in the "worst conditions imaginable" where the principle of "contradiction for contradiction's sake" holds sway (Champod and Vuille 2011: 31). Inquisitorial systems, by contrast, place too much trust in court-appointed experts who are complacently supposed to be both neutral and competent and given the misplaced status of a "judge in white" (Volk 1993: 45). The expert's work is rarely called into question and even when it is, the evidence is not assessed on its scientific merits but on the basis of exogenous factors, such as the expert's reputation, employment, or official status.

Against the failure of legal systems across the common law and civil law traditions to provide a reliable system for gathering and assessing scientific evidence, it is hardly surprising that much of the focus of recent debates has been on improving standards within forensic science itself (see, e.g., National Research Council 2009; Champod and Vuille 2011). If we extend our gaze beyond the constraints of the adversarial and inquisitorial models of proof, however, improvements might also be made to the legal process within which scientific evidence must be gathered and assessed. The principles of equality of arms and adversarial procedure developed by the ECtHR provide the basis for realigning traditional proof systems, in accordance with a new model of proof better characterized as "participative" than as "adversarial" or "inquisitorial" (Jackson 2005). The participation model is predicated on the notion that although the main actors in the criminal process – prosecutors, defense lawyers, and adjudicators – each have separate defined roles in adjudication, processes of proof should be constructed to provide all these actors with as much relevant evidence as possible, as early as possible in the process so that it can be assessed, challenged, and evaluated in a spirit of positive participation. When it comes to scientific evidence, this would mean that the parties should ensure that scientific experts are fully informed during the course of their investigations and that the experts themselves are encouraged to share and examine information between themselves. Opportunities for effective challenge and testing should be maximized. This may be better achieved by way of written exchanges before trial than in the full glare of an adversarial trial. Traditional methods of oral examination and cross-examination developed to test cognitive capacities or the good faith of ordinary witnesses are less suited to scrutinizing scientific methodology.

In common with other generalized models of adjudication, it may be that the positive participation model is most useful in expressing values that should be embedded in adjudication than as a detailed "one-size-fits-all" blueprint for conducting all cases. There are also procedural values at work that tend to detract from active participation. The defense has no obligation in Anglo-American systems to put forward any positive theory of its case and is entitled to play a passive role throughout the process. Here the defense may be more inclined to play a negative role of demolishing the prosecution case than a positive role in the generation of specific theories and hypotheses. One commentator has claimed that adversarial systems, in which two parties dance a "tango" with each other to settle for an "interpretive" truth, are ill-suited to the more interactive or collective enterprise of searching for substantive truth which resembles a "rumba" performed by coordinated troupes of dancers moving in time to a shared rhythm (Grande 2008). Where scientific evidence is involved, however, it may be argued that all the lay participants have to try to understand what the evidence means in order to assess its value. Since legal systems cannot rely on scientists themselves to validate scientific theories and techniques, lay participants have no choice but to make the necessary epistemological inquires for themselves and this suggests that, at least in exchanges between the parties and the scientists, a more positive participative exchange should be encouraged and facilitated (see Imwinkelried 2011).

Recommended Reading and References

Alldridge P (1999) Scientific expertise and comparative criminal procedure. Int J Evidence Proof 3:141–164

Allen RJ, Miller JS (1993) The common law theory of experts: deference or education? Northwestern Univ Law Rev 87:131

Beecher-Monas E (2007) Evaluating scientific evidence. Cambridge University Press, Cambridge

Champod C, Vuille J (2011) Scientific evidence in Europe – admissibility, evaluation and equality of arms. Int Commentary on Evidence 9(1):1–68

Damaška M (1973) Evidentiary barriers to conviction and two models of criminal procedure: a comparative study. 121 University of Pennsylvania Law Review 506

Damaška M (1997) Evidence law adrift. Yale University Press, New Haven

Dennis I (2010) The law of evidence, 4th edn. Sweet & Maxwell, London

Field S (2009) Fair trials and procedural tradition in Europe. Oxf J Legal Stud 29:365

Goldberg S (1994) Culture clash: law and science in America. Harvard University Press, Cambridge

Grande E (2008) Dances of criminal justice: thoughts on systemic differences and the search for truth. In: Jackson J, Langer M, Tillers P (eds) Crime, procedure and evidence in a comparative and international context: essays in honour of professor Mirjan Damaška. Hart, Oxford

Greer D (1971) Anything but the truth? The reliability of testimony in criminal trials. Brit J Criminol 11:131

Haack S (2004) Truth and justice, inquiry and advocacy, science and the law. Ratio Juris 17:15–26

Hodgson J (2005) French criminal justice. Hart, Oxford

Imwinkelried EJ (2011) The end of the era of proxies. In: Proceedings of the third international conference on evidence law and forensic science, Beijing, 8

Jackson JD (1988) Two methods of proof in criminal procedure. Modern Law Rev 51:549

Jackson J (2004) Managing uncertainty and finality: The function of the criminal trial in legal inquiry. In: Duff A, Farmer L, Marshall S, Tadros V (eds) The trial on trial volume one. Hart, Oxford

Jackson JD (2005) The effect of human rights on criminal evidentiary processes: towards convergence, divergence or realignment. Modern Law Rev 68:737

Jakobs LEMP, Sprangers WJJM (2000) A European view on forensic expertise and counter expertise. In: Nijboer JF, Sprangers WJJM (eds) Harmonisation in forensic expertise. Thela Thesis, Leiden

Langer M (2004) From legal transplants to legal translations: the globalisation of plea bargaining and the Americanization thesis in criminal processes. Harv Int Law J 45:1

Law Commission (2011) Expert evidence in criminal proceedings in England and Wales. HMSO: London. Report 325

Ligertwood A, Edmond G (2010) Australian evidence, 5th edn. LexisNexis, Chatswood

Markesinis, BS (1994) The gradual convergence. Clarendon Press, Oxford

Muensterberg H (1908) On the witness stand: essays on psychology and crime. Doubleday, New York

National Research Council (2009) Strengthening forensic science in the United States: a path forward. The National Academies Press, Washington, DC

Nijboer J (1993) The law of evidence in Dutch criminal cases in a nutshell: the role of the expert. In: Nijboer JF, Callen CR, Kwak N (eds) Forensic expertise and the law of evidence. Royal Netherlands Academy of Arts and Sciences, Amsterdam

Nobles R, Schiff D (2000) Understanding miscarriages of justice: law, the media and the inevitability of crisis. Clarendon, Oxford

Redmayne M (1997) Expert evidence and scientific disagreement. UC Davis Law Rev 30:1027–1080

Redmayne M (2001) Expert evidence and criminal justice. Clarendon, Oxford

Roberts P (2002) Science, experts and criminal justice. In: McConville M, Wilson G (eds) The handbook of the criminal justice process. Oxford University Press, Oxford

Roberts P (2008) Faces of justice adrift? Damaška's comparative method and the future of common law evidence. In: Jackson J, Langer M, Tillers P (eds) Crime, procedure and evidence in a comparative and international context: essays in honour of professor Mirjan Damaška. Hart, Oxford

Schuck PH (1993) Multi-culturalism redux: science, law and politics. Yale Law Policy Rev 11:1–46

Sklansky D (2009) Anti-inquisitorialism. Harv Law Rev 122:1634

Twining W (1982) The rationalist tradition of evidence scholarship. In: Campbell E, Waller L (eds) Well and truly tried: essays on evidence in honour of Sir Richard Eggleston. Law Book, Sydney

Van Kampen P (2000) Confronting expert evidence under the European convention. In: Nijboer JF, Sprangers WJJM (eds) Harmonisation in forensic expertise. Thela Thesis, Leiden

Vogler R (2005) A world view of criminal justice. Ashgate, Aldershot

Volk K (1993) Forensic expertise and the law of evidence in Germany (criminal cases). In: Nijboer JF, Callen CR, Kwak N (eds) Forensic expertise and the law of evidence. Royal Netherlands Academy of Arts and Sciences, Amsterdam

Weigend, T (2008) The decay of the inquisitorial ideal: plea bargaining invades German criminal procedure. In: Jackson J, Langer M, Tillers P (eds) Crime, procedure and evidence in a comparative and international context: essays in honour of professor Mirjan Damaška. Hart, Oxford

Scientific Evidence in Criminal Prosecutions

Paul Roberts[1] and William O'Brian[2]
[1]School of Law, University of Nottingham, Nottingham, UK
[2]School of Law, University of Warwick, Coventry, UK

Synonyms

Expert evidence; Expert witnesses; Forensic evidence; Forensic science evidence

Overview

There is no doubt that forensic science today makes an enormous contribution to the detection and proof of crime in modern legal systems. Scientific evidence is often useful, and sometimes vital, in proving offenses and bringing perpetrators to justice who might otherwise evade detection or conviction. Strong scientific evidence implicating the accused often induces guilty pleas without the need for a contested trial. Scientific evidence simultaneously plays an important role in excluding innocent suspects from further official inquiries or surveillance. Yet scientific evidence can also be a potent source of injustice when errors are made.

This entry explores the role of scientific evidence in criminal prosecutions, particularly with regard to common law jurisdictions such as those in the UK and the USA. It highlights challenges for criminal process actors and indicates issues of on-going controversy and concern. Prosecutors play a pivotal role in the use of scientific techniques of detection and in providing scientific evidence to the court. Common law prosecutors are instrumental in instructing expert witnesses and, in consultation with trial counsel, in adducing scientific reports and testimony in cases proceeding to trial. While the entry focuses primarily on prosecutors' professional responsibilities within an institutional legal framework informed by particular ideals of justice, the final section briefly addresses a more socio-legal issue: the impact on prosecutions of the "CSI Effect."

Scientific Evidence as Proof of Crime

The use of scientific techniques and expertise in the investigation and prosecution of crime is not a new phenomenon. The appliance of cutting-edge technology in the administration of criminal justice goes hand-in-hand with the Enlightenment and modernity's scientific revolution. Indeed, English law cases addressing issues surrounding expert evidence, such as *Buckley* v *Rice Thomas* (1554) 1 Plowden 118, CB, (which is still occasionally cited today), date back to the Renaissance. However, the twentieth century is properly regarded as the century of forensic science. This is when the early practical experimentation of the nineteenth century, immortalized by Sir Arthur Conan Doyle in the figure of Sherlock Holmes and his avowedly scientific methods of criminal detection, truly bore fruit. First pathology and fingerprinting, then blood-typing (serology) and a raft of comparison "sciences" (handwriting, ballistics, toolmarks, glass, hairs, fibers, footwear marks, dentition, facial-mapping, voice analysis, etc.), and most recently CCTV cameras, mobile phones, digital forensics, and – above all – DNA profiling have transformed criminal investigation, prosecutions, and trials around the globe.

Forensic science has become an integral part of modern criminal justice systems. Scientific evidence routinely features in high-profile serious crimes of murder, robbery, and sexual assault, but also plays an important role in relation to more mundane offending lower down the criminal calendar. Empirical research indicates that relying on DNA evidence in prosecutions of high volume crimes such as burglary yields many more identifications of perpetrators than more traditional methods of detection, including fingerprinting, at apparently reasonable cost (Roman et al. 2008: but cf. Williams and Johnson 2008: Chap. 6). It is not surprising that governments have invested heavily in the development of forensic science technologies, notably including the UK's DNA expansion program and the creation of a large National DNA Database (NDNAD) (Williams and Johnson 2008: Chap. 5) – although the Coalition Government's highly controversial recent decision to close the British Forensic Science Service (FSS) is regarded by many as a retrograde step (see House of Commons Science and Technology Committee 2011).

Acknowledging that the development and proliferation of forensic science over the course of the twentieth century has been a marvelous boon for the administration of criminal justice is not to say that it comes without any risks or drawbacks, much less to encourage complacency about it. Like any powerful tool, forensic science gets the job done efficiently and with impressive results but it is equally capable of doing appalling damage if it is not handled carefully and treated with respect. Precisely because scientific evidence has a – generally speaking, well-merited – reputation for objectivity and reliability, it is prone to be highly misleading in cases where it is either inherently unsound or put to improper uses. This is why forensic science is often implicated in miscarriages of justice. In a National Registry of Exonerations recently launched by Michigan and Northwestern law schools in the USA, "bad forensic science" features in 24 % of 891 confirmed false convictions since 1989 (Gross and Shaffer 2012: 63–65). It is often suggested that, whatever defects scientific evidence might suffer from, it has to be more reliable than other forms of evidence traditionally relied on in criminal adjudication, such as eyewitness identifications and confessions by suspects, which have often produced false convictions in the past and have been repeatedly torn to shreds by behavioral science researchers. While it might well be true that scientific evidence is comparatively more reliable than other forms of criminal proof, this is no reason not to strive to improve its performance and to make every effort to safeguard against errors, to the extent that this might be possible.

Prosecutors play a pivotal role in the use of scientific techniques of detection and in providing scientific evidence to the court. In many continental jurisdictions prosecutors direct criminal investigations, including recourse to scientific testing, and judges seek further technical advice, as required, from their own court-appointed experts. In common law jurisdictions, criminal investigations have traditionally been run by police detectives with substantial operational autonomy and comparatively little input from prosecutors. But this is changing. For example, many state prosecutors in the USA today adopt a proactive approach to participating in major criminal investigations, and even in England and Wales – where strict separation between the police and the Crown Prosecution Service (CPS) used to be an article of faith – there is far greater emphasis on interagency dialogue and cooperation between what is now, revealingly, described in official documents as "the prosecution team" (Moreno and Hughes 2008). Prosecutors are instrumental in instructing expert witnesses and, in consultation with trial counsel, in adducing scientific reports and testimony in those cases that proceed to trial.

Scientific Evidence in Adversarial Criminal Process

Reliance on scientific evidence in particular criminal prosecutions is the outcome of a deliberate, self-conscious process, comprised of a series of reasonably distinguishable, though often overlapping, temporal phases. Forensic science evidence is just as much a product of police case-building, structured by a hypothesized "theory of the case," as any other type of evidence. Research conducted for the Royal Commission on Criminal Justice identified nine key phases in the production of scientific evidence (Roberts and Willmore 1993; Roberts 1994), suggesting the following, somewhat stylized, model of "the typical case" to which real proceedings conform to a greater or lesser extent.

(1) Investigators must first of all decide to utilize scientific expertise. The initial decision is made by police and prosecutors and therefore answers to investigative, rather than strictly scientific, imperatives. This creates an observable tension between investigators' needs and expectations, and the ability of science to satisfy them. In recent times attempts have been made to educate investigators about the possibilities and limitations of scientific evidence, harnessing technological innovation to financial imperatives driving "cost-effective" policing (Lawless and Williams 2010). Police officers have been encouraged to turn to science in the investigation of "routine" volume crimes such as burglary and theft, where the potential for scientific assistance has often been overlooked in the past, but progress is uneven across the regions. It remains the case that scientific evidence is produced only when investigators think they need it, which is not necessarily when science might in fact be of most – or indeed, of any – assistance to the prosecution.

(2) Having elected to employ scientific assistance, the first task for police or prosecutors is to locate an appropriate expert. In the UK, until very recently, the FSS has been on hand to provide what was generally acknowledged to be a world-class service in a range of forensic specialisms, including DNA profiling. With the closure of the FSS, reliance must now be placed on market provision by a range of commercial suppliers, some of which – like LGC Forensics, the old Lab of the Government Chemist – were formerly state-run but subsequently privatized. Effective regulation, validation of techniques and processes, and accreditation of laboratories is vital to the integrity of this market-based system. In England and Wales, much of this responsibility has been invested in the newly created post of Forensic Science Regulator, with the assistance of the Forensic Science Advisory Council. For other types of expertise, including clinical, medical, and the more esoteric forensic sciences, investigators have to look to the hospitals, universities, research institutes, and private consultancies to locate an appropriate expert. Although these arrangements seem to

work out in the majority of cases, the process of hiring experts is surprisingly informal, and sometimes fails to produce the best evidence. It remains to be seen whether a post-FSS world of free-market provision will continue to supply an adequate range of high-quality forensic science expertise to the administration of criminal justice, with appropriate investment in quality control and research and development (Lawless 2011; Roberts 1996).

(3) Once an appropriate expert has been engaged, the next step is to supply the expert with relevant crime scene material (or other raw data) for analysis. Crime stain samples or material recovered from suspects or complainants must be identified, preserved, and transmitted to the laboratory free from contamination and protected from (further) degradation. Police and prosecutors must ensure that chain of custody is properly documented, since physical evidence is worthless unless the court can be confident about its provenance and integrity. For example, samples should be collected in "temper-evident" packaging (Lynch et al. 2008: Chap. 4). These simple administrative measures make an essential contribution to ensuring that justice is not only done, but also manifestly seen to be done.

(4) Another crucial aspect of the process of generating forensic science evidence concerns the nature of the instructions received by the expert. Scientists are inevitably influenced by the type and extent of background information provided to them by investigators, and even possibly by the form and wording of the police request for assistance. Researchers have drawn attention to the risks of unconscious biases creeping into scientific judgments through suggestive contextual information, leading forensic experts to "see," and report to prosecutors and courts, what the information they were given has led them to expect (Risinger et al. 2002). One way to respond to the risk of "contamination" by such extraneous influences would be to insist that items must be sent to the laboratory without any accompanying background information and with a request for assistance in scrupulously neutral terms. At one time the FSS considered this best practice. Yet that approach, and the language employed in its justification, betrays a fundamental misconception about the nature of forensic science. Following from the essentially applied nature of their discipline, most forensic scientists prefer to be told as much background information as possible, in order to be able to tailor their approach to the needs of the investigation. There is, most obviously, little point in a scientist wasting time and energy on matters that are not disputed in the proceedings. Background information is not so much, on this view, an external source of "contamination" or "bias," but an essential part of the scientist's data for analysis. Nonetheless, the continued risk of inappropriate, and possibly unconscious, influence suggests that – to the extent that it can reliably be identified – irrelevant and potentially prejudicial material should be filtered out of the information provided to forensic scientists, at least until they have conducted relevant tests and produced their preliminary findings.

(5) The scientist next proceeds to conduct whatever testing or examinations are judged appropriate. (6) The results of the scientific investigation are then written up into a report. Both testing and reporting reflect investigators' instructions and expectations, within the broader framework of criminal proceedings and the conventional practices of forensic science. The rules of substantive criminal law structure police investigations, which in turn influence the questions scientific experts are asked to consider, the tests they undertake, and the nature and content of the reports they write. For example, scientists asked to produce a DNA profile from blood-stained clothing in order to establish identity would not routinely comment on a self-defense theory of the case, or any other conceivable hypothesis such as provocation, mistake, or accident. But if specifically invited to do so, the expert could consider how blood spatter patterns, DNA mixtures or secondary transfer, etc., might bear on potential defense arguments and report any relevant findings or advice to prosecutors. According to one police training manual, "forensic awareness… involves the prosecution team providing the forensic scientist with all essential information in the case from the outset, keeping

them continuously updated and asking them the right questions to progress the case and get the best results" (Moreno and Hughes 2008: 25). When producing reports detailing their results and conclusions, experienced forensic scientists are conscious of the instrumental role of science in criminal proceedings, and of a report's intended audience. Investigators expect scientists to help them prove, or disprove, criminal charges against a suspect, and the style and language of expert reports is directed to that end. There are generalized pressures to work quickly and produce definite conclusions. Even if scientists were always paragons of the ideals of impartial and objective inquiry, which most of them espouse and the law demands, their work product would still remain a highly selective, constrained, stylized, and instrumentally orientated form of science. Unfortunately, forensic scientists have occasionally been seduced, or corrupted, by the institutional pressures inherent to an adversarial criminal process, allowing themselves to become partisan instruments of the prosecution rather than impartial purveyors of objective scientific facts and considered expert opinions to the courts (see Erzinçlioglu 1998; Giannelli 1997).

The production of a scientific report often effectively signals the end of particular criminal proceedings, either because the accused is induced to plead guilty in the face of compelling incriminating evidence, or because the prosecution is too weak to proceed in the absence of scientific corroboration. Scientific evidence plays an important – and possibly somewhat overlooked – role in exonerating innocent suspects from continued suspicion, in cases where DNA profiling indicates that the suspect could not, in fact, have been the perpetrator after all. In those minority of cases that do proceed to trial, the production of scientific evidence typically involves three further key stages.

(7) Defense lawyers sometimes appoint their own experts, occasionally to follow up exculpatory leads, but more often just to double-check the work already conducted by a prosecution expert. The appearance of a defense expert may present opportunities for communication, or even cooperation and the exchange of ideas, between experts on opposite sides of the adversarial divide. Forensic scientists seldom share the adversarial culture of the police and lawyers who hire them, and some scientists treat their notional adversaries as colleagues. On the occasions when scientists meet in the laboratory to review test results, defense experts have been known to persuade scientists working for the prosecution to undertake further tests or to reinterpret their results in the light of a different perspective or new information (Roberts and Willmore 1993: 56–57). Scientific investigations undertaken by the defense, in conjunction with prosecution scientists or independently, do occasionally produce significant exculpatory evidence. In the majority of cases, however, the defense examiner simply confirms the prosecution expert's data and conclusions.

(8) Expert witnesses may attend a pretrial conference with counsel. This is potentially an important meeting, at which counsel can review the expert's evidence and ensure that the expert is prepared to stand by the conclusions expressed in the expert's report. Empirical research conducted in England and Wales in the early 1990s found that pretrial conferences between experts and counsel were frequently short or nonexistent (Roberts and Willmore 1993: 57–61). This was partly due to time pressures, but also reflected a deeply held aversion on the part of some barristers to meeting any witnesses prior to trial, for fear of attracting accusations of "witness coaching." That fear seems misconceived with regard to expert witnesses, not because experts are immune to pressure and undue influence, but because counsel are unlikely to lead scientific evidence successfully if they do not themselves understand the expert's evidence. Lack of pretrial communication between experts and counsel was said on occasion to have adversely affected the presentation of scientific evidence at trial. By contrast, witness preparation is a standard feature of criminal litigation in the USA. No US prosecuting attorney would proceed to trial, in any serious matter, without first having reviewed the evidence of key prosecution witnesses, including scientific experts, in person.

Over the ensuing years, the novel idea of conducting pretrial conferences with expert witnesses has become more familiar and culturally acceptable to barristers in England and Wales, too. More significantly still, English courts have adopted a general policy of proactive judicial trial management within the framework of the Criminal Procedure Rules (CrimPR), which were introduced in 2005. The CrimPR make explicit provision for clarifying, and if at all possible resolving, scientific issues prior to trial through conferences of experts and joint reports. The Court of Appeal has repeatedly emphasized the importance of these provisions, and constantly encourages trial judges to enforce them on the parties. It has come to be widely appreciated that adversarial trials involving "battles of experts" tend to obscure genuine scientific disagreements and sometimes give the impression of scientific dispute or uncertainty where none truly exists. The prevailing philosophy among the senior judiciary is consequently to try to deal with scientific aspects of the case outside the courtroom, while still preserving the determination of genuinely contested issues for the jury at trial.

(9) Although only a very small percentage of criminal cases results in a contested trial in any adversarial jurisdiction where guilty pleas and bargains are the norm, these are disproportionately serious and important cases. Scientific evidence features in many of these contested trials, to a greater or lesser extent. Such evidence is often uncontroversial, and may be agreed – possibly as a result of successful pretrial discussions between the parties and their experts. When not disputed, the expert's report can simply be read out as documentary evidence, and is likely to be accepted at face-value by jurors (albeit that the jury is always at liberty to form its own view, and must be so directed by the trial judge in England and Wales: see e.g. *R* v *Allen* [2005] EWCA Crim 1344; *R* v *Hookway and Noakes* [2011] EWCA Crim 1989). Where the defense or, more infrequently, the prosecution wants to challenge scientific evidence, however, the expert will usually be called to court to testify in person by the party instructing the expert. There is formal provision for court-appointed experts in many common law jurisdictions. However, judges tend not to utilize court experts, in deference to the adversarial precept that the parties run their own cases at trial, with the judge adopting a relatively passive role as neutral "umpire" to ensure "fair play" and litigants' adherence to the rules of criminal procedure. (The "umpireal" model of judging in adversarial trials is an overstated and increasingly anachronistic simplification, but still serves to encapsulate a deep-seated cultural difference between "adversarial" and "inquisitorial" conceptions of the trial judge's role.)

Scientific experts testify in the courtroom like any other witness, through a series of answers to questions put by counsel. The expert witness is first taken through examination-in-chief (in the USA, direct examination) by the advocate calling the witness, and then undergoes cross-examination by counsel for the other party or parties in a multi-handed trial. If thought necessary to clear up any matter raised in cross-examination, a third phase of questioning ("reexamination" in the UK, "redirect" in the USA) may be conducted by the side calling the expert. Needless to say, this is a highly artificial way of presenting scientific evidence to the court. Its success depends in large part on the skill and scientific understanding of counsel, which cannot always be relied upon, especially if pretrial preparation has been inadequate. It might be preferable, from the point of view of communicating scientific evidence effectively, if experts could present their evidence in a more direct and less constrained fashion. As things stand, counsel may through incompetence or as a deliberate strategy distort the intended meaning of an expert's evidence, and the effect may be compounded where experts called by the prosecution and defense disagree, or appear to disagree, with each other. The evident limitations in this context of traditional oral trial procedure, whatever its efficacy or cultural significance in relation to ordinary witnesses of fact, underscore the wisdom of making more extensive use of pretrial mechanisms to fully and fairly exploit the potential of forensic science in criminal proceedings.

The Common Law Prosecutor's Role and Responsibilities

In many continental legal systems prosecutors are formally part of the magistracy. As a matter of jurisprudential theory, they are unequivocally regarded as acting in the pursuit of justice and they are supposed to be objective and nonpartisan. In common law countries, too, prosecutors are required to be "ministers of justice" who select and pursue appropriate charges against those suspected, on the basis of pertinent and reliable evidence, of having committed particularized criminal offenses. "Prosecution counsel has to exercise independent judgment throughout, with the objective not of obtaining a conviction at any cost, but of ensuring that justice is done" (Buxton 2009: 427). English prosecutors, it is said, prosecute but do not persecute. The special responsibility that prosecutors owe to justice may be rather less self-evident and unequivocal in an adversarial procedural system, in which criminal trials can sometimes take on the appearance of a gladiatorial contest. But it is no less fundamental to the legal system's claims to justice and legitimacy. There is a world of difference between a tough-minded lawyer who prosecutes firmly and fairly, and an ethically cavalier show-boater who tries to win cases and secure convictions at almost any cost.

Another important difference between civilian and common law jurisprudential theory is that, whereas many continental systems adopt some version of the "principle of legality" requiring compulsory prosecution in every serious case in which there is solid evidence of criminality, common law prosecutions are fundamentally discretionary. Of course, a suspect cannot be prosecuted in the absence of evidence capable of demonstrating his guilt. But incriminating evidence alone is not automatically enough to justify, let alone mandate, a prosecution in common law jurisdictions. Prosecution in the instant case must also be judged to be in the public interest more broadly conceived. This principle is encapsulated in the two-step test for prosecution specified by the Code for Crown Prosecutors in England and Wales (CPS 2010). The first part of the test poses the question of evidential sufficiency: is there a "reasonable prospect of conviction" of specified charges on the available evidence? This is generally understood to mean that the reviewing prosecutor must judge that conviction is more likely than not, on the balance of probabilities (ibid: para. 4.6). Otherwise, more evidence must be sought by the police; and the prosecution may ultimately have to be abandoned if better evidence is not forthcoming. Only if the evidential sufficiency test is satisfied should the Crown Prosecutor proceed to consider the second limb of the two-part Code test: is prosecution also justified in the public interest? The Code for Crown Prosecutors contains lists of public interest factors supporting or detracting from the case for prosecution, albeit that offense seriousness is always a weighty consideration and will often be dispositive in practice. Crown Prosecutors in England and Wales are under a duty of continuous review to ensure that the Code Test is satisfied in any case proceeding to trial. Since 2004, Crown Prosecutors have also been responsible for framing the initial charges in serious cases – a task formerly allocated to the police, on the traditional theory that common law prosecutions are initiated by private parties. Selection of appropriate charges likewise requires an (earlier) assessment of evidential sufficiency. Charging and evidential review are among the prosecutor's prime responsibilities in all common law jurisdictions, though detailed doctrinal arrangements and the institutional micro-dynamics of police-prosecutor relationships will naturally differ from one legal system to another.

As forensic science and other kinds of technical expertise have become increasingly prominent features of modern criminal investigations, they have also inevitably presented prosecutors with new challenges and opportunities. Some cases are effectively impossible to prosecute without scientific or medical evidence; in others, expert testimony plays a vital supporting rule. Prosecutors must grasp scientific fundamentals in order to make informed assessments of evidential sufficiency. For example, if an expert

DNA report states that the accused's DNA profile matches a crime stain partial profile with a random match probability of 1 in 30,000, the prosecutor needs to be able to make sense of this jargon in order to assess the probative value of the DNA evidence, within the evidential context of the case as a whole. Or again: if a medical expert report states that the pattern of a child's injuries is "consistent with nonaccidental trauma," does this constitute strong evidence of child abuse? Or does it simply fail to rule out abuse as one among myriad other possibilities that could be "consistent with" the injuries observed? In short, widespread use of scientific evidence in criminal prosecutions demands at least a basic level of scientific literacy from prosecutors (as well as from other relevant criminal justice professionals, including defense lawyers and judges). Ideally, prosecutors should also be equipped to spot the potential for developing lines of scientific inquiry that the police might have initially overlooked or regarded as unnecessary. This implies that prosecutors cultivate reasonable familiarity with the range of scientific specialisms and technologies potentially available to support criminal investigations, and should know how to identify and instruct suitably qualified experts, in consultation with police investigators and trial counsel where appropriate.

There are various lists of forensic practitioners produced by professional associations such as the UK Forensic Science Society, the British Psychological Society, and the Society of Expert Witnesses. National prosecution services like the CPS and state District Attorneys offices in the USA might further assist individual prosecutors by building up local intelligence on the range and quality of forensic scientific assistance available for consultation. These largely invisible, office-based processes have rarely been studied by empirical researchers, but what little we do know tends to suggest that identifying and instructing scientific experts is a rather *ad hoc* affair, heavily influenced by the motivation and personal experience of individual prosecutors. As well as failing to capitalize on opportunities to develop better evidence in the instant case, failure to share experiences more systematically may allow poorly performing experts to remain in circulation long after their consulting forensic practices should have been closed down. It is difficult to believe that this curiously unscientific approach to drawing on scientific expertise is beyond practical improvement – though it has to be said that developments over the last several decades (including the short life and ignominious demise of the UK Council for the Registration of Forensic Practitioners) provide little grounds for optimism.

The CPS in England and Wales has adopted various topic-specific policies pertaining to scientific evidence. One significant interagency document provides guidance on charging in cases involving DNA evidence (The Prosecution Team 2004), including the important principle that "[a] suspect should not ordinarily be charged solely on the basis of a match between his own profile and a DNA profile found at the scene of the crime" (ibid 11.1). This effectively introduced an informal corroboration requirement for DNA profiles in relation to suspects first identified through a speculative search of the NDNAD. It attempts to forestall the embarrassment experienced in several early DNA cases in which suspects who, viewed objectively, could not possibly have committed the crime in question were nonetheless charged purely on the basis of an adventitiously matching DNA profile. This was an object lesson in the dangerous fallacy of treating DNA evidence as though it supplied infallible proof of guilt and an invaluable reminder that, just occasionally, DNA matches can occur purely by chance (if not through contamination, analytical discrepancy, or interpretational error). CPS charging policy in England and Wales now sensibly proceeds on the assumption that, if the suspect is truly guilty, there should be more evidence linking him to the offense than an unexpected "hit" on the DNA database. Yet unjustifiably late decisions to discontinue prosecutions involving DNA matches evidently still do occur (see, most recently, Dodd and Malik 2012).

CPS training and policy specifically draw Crown Prosecutors' attention to the potential for utilizing scientific or medical evidence in a number of priority areas. In relation to domestic violence prosecutions, physical evidence of assaults backed-up by medical testimony may be vital in continuing with a prosecution if the complainant later – for whatever reason – withdraws her complaint (Dempsey 2004), by all accounts a frustratingly common occurrence. Scientific evidence can be important in building up the prosecution's case in rape and sexual assault prosecutions, partly because it can help to deflect the perception that such cases often involve "one person's word against another's," inducing the jury to acquit because it cannot be sure that the accused is guilty, despite jurors' suspicions that he probably is. At the limit, expert testimony can sometimes effectively constitute compelling evidence of guilt or innocence by itself. However, prevailing scientific opinion can and does change over time, making it risky to base a prosecution primarily on expert evidence, at least where the relevant science is rapidly evolving. This dilemma was presented starkly in England and Wales by a recent series of criminal prosecutions of mothers for murdering their children, based in part on the alleged improbability of multiple innocent cot deaths in the same family.

R v *Clark* [2003] EWCA Crim 1020 was one notorious case in which the prosecution expert testified that the probability that both of the defendant's children could have died from Sudden Infant Death Syndrome (SIDS) was 1 in 73 million. This opinion was invalid when given, since it improperly assumed that serial cot deaths are independent events when in fact they could be caused by the same underlying pathology or genetic predisposition (with the further implication that multiple SIDS might run in families). In *R* v *Cannings* [2004] 2 Cr App R 7, a case in which three of the accused's infant children had died, the Court of Appeal observed:

> Experts in many fields will acknowledge the possibility that later research may undermine the accepted wisdom of today.... That does not normally provide a basis for rejecting the expert evidence. With unexplained infant deaths, however ... in many important respects we are still at the frontiers of knowledge.... In cases like the present, if the outcome of the trial depends exclusively or almost exclusively on a serious disagreement between distinguished and reputable experts, it will often be unwise, and therefore unsafe, to proceed. (ibid. 111)

Standing alone, this statement might have been interpreted as giving a strong hint to prosecutors that they should discontinue prosecutions in cases of suspicious infant death whenever there was disputed medical evidence. In *R* v *Kai-Whitewind* [2005] 2 Cr App R 3, however, the Court of Appeal said that it would be a "startling proposition" if "whenever there is a conflict between expert witnesses the case for the prosecution must fail unless the conviction is justified by evidence independent of the expert witnesses" (ibid. 480). The Court proceeded to elucidate an essential distinction:

> In *Cannings* there was essentially no evidence beyond the inferences based on coincidence which the experts for the Crown were prepared to draw. Other reputable experts in the same specialist field took a different view about the inferences, if any, which could or should be drawn and hence the need for additional cogent evidence. With additional evidence, the jury would have been in a position to evaluate the respective arguments: without it, in cases like *Cannings*, they would not. (ibid.)

Coincidence, in other words, cannot prove murder beyond reasonable doubt, if expert opinion is divided on whether the circumstances justify an inference of foul play. But the mere fact of expert disagreement will not block a prosecution supported by other evidence.

The Court of Appeal reiterated in *R* v *Henderson* [2010] 2 Cr App R 24 that conflicts in expert evidence are in principle to be resolved by juries just like any other conflict of factual testimony. *Henderson* concerned three consolidated appeals involving allegedly "shaken babies," prosecutions which have provoked much controversy in both the UK (*R* v *Harris* [2006] 1 Cr App R 5) and the USA (Tuerkheimer 2011). Prosecution evidence in these cases often almost amounts to an uncorroborated medical diagnosis of murder.

Doctors have testified repeatedly that babies exhibiting a particular combination of symptoms (known as "the triad") were definitely injured intentionally, and even sometimes claiming that the temporal pattern of symptoms pinpointed the culprit. However, recent developments in medical research and scientific thinking have cast serious doubt on such diagnoses. In *Henderson* the Court of Appeal stated unequivocally that "the triad" is not to be treated as conclusively diagnostic of nonaccidental head injury. This is something else that Crown Prosecutors must now take into account in exercising their charging and prosecutorial discretions in cases of alleged child abuse. Current CPS policy states that prosecutions based only on the triad are unlikely to be brought or continued, in the absence of "appropriate supporting evidence (which in certain circumstances can be found in the absence of certain factors)" (CPS 2011).

Sexual assault cases based on physical examinations of the alleged victims have also posed difficulties for prosecutors. In the past, medical experts have sometimes been willing to infer intercourse from physical marks or injuries, e.g., a ruptured hymen or anal tearing. However, a recent report by the Royal College of Paediatrics and Child Health (2008) called into question the extent to which such examinations could warrant positive findings of abuse, prompting the Court of Appeal to quash a conviction of child sexual assault in *R* v *PF* [2009] EWCA Crim 1086. Medical evidence led by the prosecution at trial was also criticized for being too dogmatic by the appeal court in *R* v *Martin T* [2008] EWCA Crim 3229, but on this occasion the conviction was upheld on the strength of the complainant's testimony. In these delicate situations, involving disputed allegations of sexual abuse or assault, prosecutors must balance their duty to present as strong a case as possible today, against the possibility that changes in medical opinion might tomorrow undermine the safety of any conviction secured at trial. Recent experiences of dramatic reversals in prevailing medical opinion will predictably make prosecutors more reluctant to pursue cases built upon the shifting sands of scientific knowledge (O'Brian 2011).

These observations and illustrations prompt a more general question: to what extent should prosecutors be responsible for assessing the quality and validity of scientific evidence (as opposed to assessing its probative value on the assumption that scientific evidence is valid)? Of course, if a prosecutor knew or strongly suspected that a particular expert witness was incompetent or a charlatan it would constitute a gross breach of professional responsibility to present that person to the court as a witness of truth. Such faulty evidence could not legitimately contribute towards satisfying the evidential sufficiency test for prosecutions in England and Wales. But this is presumably – to put it no higher – a very unusual scenario. Prosecutors are not themselves trained scientists or forensic experts. When any indications to the contrary are absent, why shouldn't the prosecutor simply take at face value the validity of scientific evidence produced by ostensibly well-qualified experts? After all, scientific experts are called upon to assist in criminal proceedings precisely because they contribute knowledge and skills otherwise unavailable to the justice system. Expert evidence comprises scientific findings, interpretative judgments, and expressions of opinion which prosecutors are presumptively ill-equipped to second-guess.

Two mutually aggravating considerations belie this superficially plausible rationalization of the legal status quo. First, recent years have witnessed a growing realization that many of the traditional forensic sciences routinely employed in criminal prosecutions and trials actually lack formal, rigorous scientific validation. A report published in 2009 by the highly respected National Research Council of the National Academy of Sciences in the USA concluded that "[i]n a number of forensic science disciplines, forensic science professionals have yet to establish either the validity of their approach or the accuracy of their conclusions, and the courts have been utterly ineffective in addressing this problem" (National Research Council 2009: 1–14). Even the reliability of fingerprinting has been called into question by high-profile controversies and inquiries in the USA (see ibid: 5–7 to 5–14) and the UK (The Fingerprint Inquiry: Scotland 2011). The point is not to induce blanket skepticism about scientific evidence and the thousands of convictions based on it, which would be a counterproductive and unwarranted overreaction to these revelations. A more measured approach demands continued vigilance and scrutiny of scientific evidence's methodological and epistemic credentials, a task to which prosecutors should arguably contribute through searching pretrial evaluations of evidential sufficiency.

Less initiative might be expected of prosecutors in this regard if the validity of scientific evidence were sure to be thoroughly investigated at trial. But herein lies the second cause for concern. Common law courts have not traditionally imposed exacting admissibility requirements on scientific evidence. To the contrary, the general attitude has been that relevant scientific evidence should be admitted, leaving any (rare) challenges to validity to be resolved as questions of fact for the jury in the course of the trial. In recent decades, many common law jurisdictions have introduced somewhat more rigorous approaches to scientific validity, requiring trial judges to undertake some measure of "gatekeeping" scrutiny, at least in relation to novel techniques or applications. The best known of these tests is the admissibility standard elucidated by the US Supreme Court in *Daubert* v *Merrell Dow* 113 S Ct 2786 (1993) and subsequently adopted in many US state procedural codes. The success, or otherwise, of such interventions is still hotly debated. Commentators have noted a tendency (not necessarily restricted to US judges: Beecher-Monas 2007: 94–121) for *Daubert* to be applied quite stringently to disqualify claimants' scientific evidence in civil tort suits, while a far more relaxed and indulgent standard is applied to admit scientific evidence adduced by the prosecution in criminal proceedings.

In the meantime, there is no equivalent admissibility test for scientific evidence in England and Wales, albeit that the Law Commission (2011) has proposed legislation to introduce a locally adapted (and rather elaborate) version of the *Daubert* standard. One might argue that prosecutors' professional responsibility to inquire into the scientific validity of expert evidence they propose to adduce at trial is all the greater in the knowledge that scrutiny in the courtroom is likely to be minimal. Lay jurors can hardly be relied upon to detect unacknowledged weaknesses in prosecution scientific evidence that lawyers and judges have failed, or disdained, to notice. On the other hand, it is probably unrealistic, in terms of institutional culture and personal psychology, to expect prosecutors in adversarial proceedings to adopt an overly critical attitude towards scientific evidence which, on the face of it, supports their case and would easily satisfy the law's undemanding admissibility requirements.

Prosecution and Defense

In *R* v *Ward* (1993) 96 Cr App R 1, 51, the English Court of Appeal observed:

> [A] forensic scientist conjures up the image of a man in a white coat working in a laboratory, approaching his task with cold neutrality, and dedicated only to the pursuit of scientific truth. It is a somber thought that the reality is sometimes different. Forensic scientists may become partisan . . .. Forensic scientists employed by the government may come to see their function as helping the police. They may lose their objectivity.

Consciously partisan scientific evidence is antithetical to the expert witness's primary duty to the court. The judges in *Ward* were in no doubt that "the clear duty of government forensic scientists [is] to assist in a neutral and impartial way in criminal investigations. They must act in the cause of justice" (ibid. 52). This most basic of the expert witness's duties has been reiterated many times (e.g. *R* v *Harris*, [2006] 1 Cr App R 5; *R* v *B (T)*[2006] 2 Cr App R 3), and is now embodied in rule 33.2 of the Criminal Procedure Rules, which states unequivocally that:

> An expert must help the court . . . by giving objective, unbiased opinion on matters within his expertise. This duty overrides any obligation to the person from whom he receives instructions or by whom he is paid.

In the United States, Standard 3–3.3(a) of the ABA *Standards for Criminal Justice: the Prosecution Function* states to similar effect, but with the emphasis on the *prosecutor's* duty:

> A prosecutor who engages an expert for an opinion should respect the independence of the expert and should not seek to dictate the formation of the expert's opinion on the subject. To the extent necessary, the prosecutor should explain to the expert his or her role in the trial as an impartial expert called to aid the fact finders . . ..

Impartiality and objectivity are indubitably laudable ideals for scientific evidence and expert witness testimony. Whether it is entirely realistic to expect the parties to adversarial litigation to interpret their respective roles with a primary emphasis on achieving justice, viewed objectively and impartiality, is another question. The expert witness's abstract "duty to the court" and to justice provides useful, but very incomplete, guidance for ethical professional practice. Prosecutors and defense lawyers also owe duties to the court, but this hardly settles myriad difficult questions of legal ethics and strategic decision-making in adversarial trial proceedings.

It is accepted as axiomatic in all modern systems of criminal justice that the accused must have a fair opportunity to put his side of the case and adequate time and facilities to mount a defense. Justice is unlikely to be done, and certainly will not be seen to be done, if the accused is muzzled or unreasonably hampered in conducting his defense. However, active defense participation in the proceedings takes on additional functional and symbolic significance in adversarial systems. Adversary theory assumes that the trier of fact will best be able to discern the truth of contested events by hearing each side advance its best case at trial and adjudicating between them (Roberts and Zuckerman 2010: 46–65). There is no overriding duty on the court to discover the truth, as there would be in inquisitorial proceedings. It would consequently undermine not only the normative legitimacy, but also the epistemic efficacy of adversarial criminal trial procedure, if the trier of fact were allowed to hear only the prosecution's side of the story. In fact, the common law jury is always presented with at least two versions of events in any contested criminal trial, even if the defense "story" is reduced to a blanket denial of the allegations advanced primarily through cross-examination of the prosecution's witnesses.

Adversarialism has broad ramifications for defense participation in criminal trials. Specifically in relation to scientific evidence, the most obvious implication is that the defense should have appropriate access to high-quality forensic science assistance, to retest and verify the prosecution's scientific evidence, and to pursue further scientific inquiries of its own (Giannelli 2004; Roberts and Willmore 1993: Chap. 3). This is partly a question of providing adequate resources, which in practice means legal aid funding for work that the uninitiated might regard as wasteful reduplication. Criminal legal aid is virtually always squeezed in times of austerity. In addition, there is the related question of adequate defense access to evidential material, data, test results, and government labs and technicians (to the extent that these public goods survive free-market provision; wherein commercial confidentiality becomes a further potential barrier to access). This brings us to the more general, and perennially controversial, topic of pretrial disclosure by the prosecution to the defense.

Contrary to what simplistic models of adversarial "contests" might lead one to expect, trial by ambush is largely a thing of the past and is deprecated by modern courts and judges. Most common law procedural systems today require extensive pretrial disclosure by the prosecution, and many jurisdictions also demand more circumscribed disclosure by the defense. (The presumption of innocence preempts fully reciprocal pretrial disclosure in criminal litigation.) Moreover, late or inadequate disclosure by the prosecution has been exposed as a potent cause of miscarriages of justice, not least in relation to scientific evidence.

The extent of the prosecution's common law duty to disclose scientific evidence was clarified in *Ward*, in which the Court of Appeal stated:

> An incident of a defendant's right to a fair trial is a right to timely disclosure by the prosecution of all material matters which affect the scientific case relied on by the prosecution, that is, whether such matters strengthen or weaken the prosecution case or assist the defence case. This duty exists whether or not a specific request for disclosure of details of scientific evidence is made by the defence. Moreover, this duty is continuous: it applies not only in the pretrial period but also throughout the trial. The materiality of evidence on the scientific side of a case may sometimes be overlooked before a trial. If the significance of the evidence becomes clear during the trial there must be an immediate disclosure to the defence.

In *Ward* itself, negative test results seemingly contradicting the prosecution's case were withheld from the prosecutor, and consequently were neither disclosed to the defense nor adduced at trial, an outcome which particularly infuriated the Court of Appeal because the jury had been kept in the dark about potentially significant information:

> The consequence is that in a criminal trial involving grave charges three senior government forensic scientists deliberately withheld material experimental data on the ground that it might damage the prosecution case. Moreover [in their testimony at trial, two of them] misled the court as to the state of their knowledge about the possibility of contamination occurring from the debris of an explosion. No doubt they judged that the records of the firing cell tests would forever remain confidential. They were wrong. But the records were only disclosed about 17 years after Miss Ward's conviction and imprisonment. (ibid. 49)

Pretrial disclosure in England and Wales is now governed by the Criminal Procedure and Investigations Act 1996, as amended, and the Criminal Procedure Rules. Although the principle of full and timely pretrial disclosure by the prosecution is accepted beyond question in English criminal proceedings, disputes can still arise over its precise requirements in individual cases, especially regarding "unused material" not forming part of the prosecution's affirmative case. Defense lawyers and academic commentators continue to express skepticism about levels of compliance in practice with the prosecution's disclosure requirements (see Lord Justice Gross 2011), possibly indicating deep cultural resistance to full disclosure in an adversarial system (Quirk 2006).

If it emerges at trial that the prosecution has failed to discharge the full extent of its disclosure duties in relation to scientific evidence, the proper course will normally be for the judge to consider granting an adjournment to allow the defense time to deal with new information. The defense would not necessarily require extra time in every case: it all depends on the precise nature of the evidence and the way in which the trial is being run by both sides. It is conceivable, however, that the evidence would have to be excluded altogether, in cases where the defense position has been irreparably damaged, for example, because the opportunity to conduct further testing on perishable samples has been lost. In the event that material nondisclosure of scientific evidence comes to light after the trial has been concluded, the Court of Appeal may determine that a conviction is no longer "safe," in the terminology of the Criminal Appeal Act 1968, and must be quashed. Retrial might be a practical option in some, but by no means all, such cases. Viewed in this light, conscientious compliance with the prosecutor's pretrial disclosure duties is calculated to safeguard the reliability and legitimacy of well-founded convictions of the guilty, at the same time as ensuring the defense has a fair opportunity to present its side of the story to the jury in an adversarial criminal trial.

A "CSI Effect"?

The "CSI Effect" has generated some anxious practitioner and academic discussion, predominantly in the USA (see Cole and Dioso-Villa 2009). The basic idea is that jurors weaned on a concentrated TV diet of the *CSI: Crime Scene Investigation* franchise, *Law & Order*, *Bones*, *Silent Witness*, *Waking the Dead* and similar top-rated serials may have come to entertain hopelessly inflated expectations of scientific evidence. The worry is that jurors in thrall to these cultural fantasies might effectively "punish" the prosecution by voting for an acquittal in any case entirely lacking in scientific proof, or in which the more banal reality of expert evidence fails to live up to the glamorous fictional portrayals driving jurors' expectations. A handful of US prosecutors claims to have experienced the CSI Effect in cases in which the outcome has been, in their eyes, disappointing (e.g. Thomas 2005).

It is a perfectly plausible conjecture that fact-finding by lay jurors in criminal trials is influenced by a range of social and cultural factors, including expectations generated by crime fiction (Tyler 2006). But such cultural influences are doubtlessly diffuse and operate in concert with countless other complementary and confounding perceptions and motivations. The empirical basis for positing a genuine CSI Effect is exceedingly thin. Most academic commentators who have addressed the issue are highly skeptical, starting with the rather protean and ill-defined concept of "the CSI Effect" itself. In most of the shows in question police and forensic scientists are portrayed as civic heroes and forensic science itself invariably wins the day and puts the bad guys behind bars where they belong. One might equally posit that jurors immersed in these predictable narratives will be only too ready to accept the truth of the prosecution's case and, in particular, to defer to scientific proof of guilt whenever it is adduced (Godsey and Alao 2011).

Whatever the empirical truth of the matter, the theoretical possibility of a CSI Effect serves as a useful reminder that scientific evidence, for all its undoubted potency as a weapon in the armory of law enforcement, is neither infallible nor necessarily free from any conceivable drawbacks, downsides, or unanticipated side-effects. Part of the price of our increasing social investment in scientific proof may be eternal vigilance (not least on the part of prosecutors) to ensure that forensic science serves to promote, and not to thwart, the values and ideals of criminal justice.

Related Entries

- ▶ Behavioral Science Evidence in Criminal Trials
- ▶ Causes of Wrongful Convictions
- ▶ Criminal Defense Profession
- ▶ Critical Report on Forensic Science
- ▶ Expert Evidence and Criminal Trial Procedure
- ▶ Expert Witnesses: Role, Ethics, and Accountability
- ▶ Fingerprint Identification
- ▶ Forensic Science and Criminal Investigation
- ▶ Forensic Science and Miscarriages of Justice
- ▶ Forensic Science Culture
- ▶ Forensic Science Professions
- ▶ History of Criminal Investigation
- ▶ Identification Technologies in Policing and Proof
- ▶ Organizing and Supplying Forensic Science Services
- ▶ Prosecution and Wrongful Convictions
- ▶ Prosecutorial Discretion
- ▶ Scientific Evidence in Criminal Adjudication
- ▶ Transnational Exchange of Forensic Evidence

Recommended Reading and References

Beecher-Monas E (2007) Evaluating scientific evidence: an interdisciplinary framework for intellectual due process. Cambridge University Press, Cambridge

Buxton R (2009) The private prosecutor as a minister of justice. Crim Law Rev 427–432

Cole SA, Dioso-Villa R (2009) Investigating the "CSI effect" effect: media and litigation crisis in criminal law. Stanford Law Rev 61:1335–1373

Crown Prosecution Service (2010) The code for crown prosecutors. CPS Policy Directorate, London. www.cps.gov.uk/publications/code_for_crown_prosecutors/index.html. Accessed on 12 Feb 2013

Crown Prosecution Service (2011) Legal guidance: non accidental head injury cases (NAHI, formerly referred to as shaken baby syndrome [SBS]) – prosecution approach. CPS, London. www.cps.gov.uk/legal/l_to_o/non_accidental_head_injury_cases/. Accessed on 12 Feb 2013

Dempsey MM (2004) The use of expert witness testimony in the prosecution of domestic violence. CPS, London

Dodd V, Malik S (2012) Forensics blunder "may endanger convictions". The Guardian, 9 March:1, 5

Erzinçlioglu Z (1998) British forensic science in the dock. Nature 392:859–860 (30 April)

Giannelli PC (1997) The abuse of scientific evidence in criminal cases: the need for independent crime laboratories. Virginia J Social Policy Law 4:439–478

Giannelli PC (2004) *Ake* v *Oklahoma*: the right to expert assistance in a post-*Daubert*, post-DNA world. Cornell Law Rev 89:1305–1419

Godsey MA, Alao M (2011) She blinded me with science: wrongful convictions and the “reverse CSI-effect”. Texas Wesleyan Law Rev 17:481–498

Gross SR, Shaffer M (2012) Exonerations in the United States, 1989–2012. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2092195. Accessed on 12 Feb 2013

House of Commons Science and Technology Committee (2011) The forensic science service. Seventh Report of Session 2010–12, HC 855, TSO, London

Law Commission (2011) Expert evidence in criminal proceedings in England and Wales. Law Com No. 325. www.lawcom.gov.uk/publications/expert-evidence.htm. Accessed on 12 Feb 2013

Lawless CJ (2011) Policing markets: the contested shaping of neo-liberal forensic science. Brit J Criminol 51:671–689

Lawless C, Williams R (2010) Helping with inquiries or helping with profits? the trials and tribulations of a technology of forensic reasoning. Soc Stud Sci 40:731–755

Lord Justice Gross (2011) Review of disclosure in criminal proceedings. Judiciary of England and Wales, London

Lynch M, Cole SA, McNally R, Jordan K (2008) Truth machine: the contentious history of DNA fingerprinting. Chicago University Press, Chicago

Moreno Y, Hughes P (2008) Effective prosecution: working in partnership with the CPS. Oxford University Press, Oxford

National Research Council (2009) Strengthening forensic science in the United States: a path forward. National Academies Press, Washington, DC

O’Brian WE (2011) Fresh expert evidence in CCRC cases. King’s Law J 22:1–26

Quirk H (2006) The significance of culture in criminal procedure reform: why the revised disclosure scheme cannot work. Int J Evidence Proof 10:42–59

Risinger DM, Saks MJ, Thompson WC (2002) The *Daubert*/*Kumho* implications of observer effects in forensic science: hidden problems of expectation and suggestion. Calif Law Rev 90:1–56

Roberts P (1994) Science in the criminal process. Oxf J Legal Stud 14:469–506

Roberts P (1996) What price a free market in forensic science services? The organization and regulation of science in the criminal process. Brit J Criminol 36:37–60

Roberts P, Willmore C (1993) The role of forensic science evidence in criminal proceedings. RCCJ Research Study No 11, HMSO, London

Roberts P, Zuckerman A (2010) Criminal evidence, 2nd edn. Oxford University Press, Oxford

Roman JK, Reid S, Reid J, Chalfin A, Adams W, Knight C (2008) The DNA field experiment: cost-effectiveness analysis of the use of DNA in the investigation of high-volume crimes. Urban Institute, Washington, DC

Royal College of Paediatrics and Child Health (2008) The physical signs of child sexual abuse: an evidence-based review and guidance for best practice. RCPCH, London

The Fingerprint Inquiry: Scotland (2011) Report. Edinburgh. www.thefingerprintinquiryscotland.org.uk/inquiry/21.html. Accessed on 12 Feb 2013

The Prosecution Team (2004) Guidance on DNA charging. ACPO/CPS, London

Thomas AP (2005) The CSI effect and its real-life impact on justice: a study by the Maricopa county attorney’s office. Prosecutor 39(October):10–16

Tuerkheimer D (2011) Science-dependent prosecution and the problem of epistemic contingency: a study of shaken baby syndrome. Alabama Law Rev 62:513–569

Tyler TR (2006) Viewing CSI and the threshold of guilt: managing truth and justice in reality and fiction. Yale Law J 115:1050–1085

Williams R, Johnson P (2008) Genetic policing: the use of DNA in criminal investigations. Willan, Cullompton, Devon

Selection Bias

▶ Juvenile Justice System Responses to Minority Youth

Selective Enforcement

▶ Biased Policing

Self-Control

▶ Situational Action Theory

Self-Control Theory

▶ General Theory of Crime

Self-Report Delinquency Survey

▶ History of the Self-Report Delinquency Surveys

Self-Reported Arrests

▶ Self-Reported Offending: Reliability and Validity

Self-Reported Offending

▶ Self-Reported Offending: Reliability and Validity

Self-Reported Offending: Reliability and Validity

Darrick Jolliffe[1] and David P. Farrington[2]
[1]School of Law, University of Greenwich, London, UK
[2]Institute of Criminology, University of Cambridge, Cambridge, UK

Synonyms

Concurrent validity; Gender and race differences; Predictive validity; Reliability; Self-reported offending; Self-reported arrests

Overview

Offending is commonly measured by asking people to admit whether they have committed each of a specified number of delinquent acts such as burglary, theft, robbery, assault, and vandalism. A key question is: How reliable and valid are these admissions? Given that people may conceal, exaggerate, or forget their offenses, how accurate are self-reports as a measure of actual delinquent behavior? This entry will focus on self-reported offending (SRO) surveys of large community samples (at least several hundred), to address this question. The extent to which these reports are differentially valid by gender and race will also be explored. There is no space to review self-report surveys of prisoners or the literature on self-reported drug use.

Self-reports of offending have been used for a number of years, initially to uncover "hidden delinquency," but more recently as a complementary measure to indexes of official offending (such as arrests or convictions). Self-reports of offending are reliable, since measures administered at different times produce similar results. The validity of self-reports is usually measured by comparing them with official arrests or convictions. Individuals who self-report offenses are more likely to have official records than those who do not report offenses, and individuals with officially recorded crimes usually admit these crimes. Crucially, self-reported offenses predict future arrests and convictions for the same crimes, even among people with no official record. Overall, the research suggests that self-reports provide a valid measure of offending for white males. Validity is sometimes lower for other demographic categories (e.g., females, Asians, and African-Americans), but these findings are far from consistent. More methodological research on self-reported offending is needed, especially on the differential validity of self-reported offending (SRO) by various demographic categories.

History of the Use of SRO

For many years, people in surveys have been asked to report offenses that they have committed that have not necessarily been detected by the police, in order to obtain information about "hidden delinquency." For example, Burt (1929) in England and the Glueck and Glueck (1968) in the United States obtained this information from their interviewees and from other informants such as parents and teachers. However, the use of structured SRO questionnaires began in the 1940s. One of the most influential early studies was by Wallerstein and Wyle (1947), who distributed a 49-item mail questionnaire to nearly 1,700 adults in New York. The main finding of this research was that offending was very common: 99 % of adults admitted at least one offense, and even ministers of the church admitted an average of 8 offenses each. Reflecting on this, however, Wallerstein and Wyle (1947) concluded that acts that were technically offenses in this study were often quite trivial.

The major influence of SRO research on criminology really began in the 1950s with the research of Short and Nye (1958). Perhaps the main reason why their SRO survey was influential was because they found no relationship between socioeconomic status (SES) and self-reported delinquency. However, when they compared training school boys with high school boys, they found that the institutionalized delinquents came from much lower SES backgrounds. The implication, therefore, was that official processing was biased against low SES people – a finding that fitted well with the prevailing theorizing at the time. This result is credited with triggering the "self-report revolution," with criminologists enthusiastically embracing the self-report method, often using the same items as Short and Nye.

Gold (1970) carried out a very important SRO survey in Michigan. This study is noteworthy because of the care taken in addressing methodological issues such as random sampling, efforts to reduce attrition, and the comparison of self-reports with informant reports of offending. The major findings of this study were that three-quarters of offenses were committed with others (most commonly peers) and that getting caught by the police was followed by an increase in offending (in agreement with labeling theory). Hirschi (1969 was the first to use the self-report method to test a theory about the causes of delinquency, and this seminal and highly cited study inspired a number of similar cross-sectional surveys designed to investigate causes using self-reports.

National SRO surveys quickly followed in the United States (e.g., Williams and Gold 1972), and these provided an alternative measure of juvenile crime rates to the official arrest records. These surveys were in turn followed by the ambitious longitudinal US National Youth Survey (NYS), beginning in 1977 and continuing to the present day (Elliott et al. 1985). This survey is one of the most important sources of self-report data on criminal careers and the causes of offending, but there are numerous other influential American prospective longitudinal studies which have used the self-report methodology, such as the Causes and Correlates projects in Pittsburgh, Denver, and Rochester; the Seattle Social Development Project; and the Oregon Youth Study (for details about all these surveys, see Farrington and Welsh 2007, Chap. 2). SRO surveys have been used in many other countries, with the Pan-European International Self-Report Delinquency Study (Junger-Tas 2010) being particularly noteworthy. The most comprehensive methodological work on these surveys has been conducted in the United States and England.

History of Research on Reliability and Validity

The first attempt to review the reliability and validity of SRO surveys was in a conference report by Hardt and Bodine (1965). They noted (p. 15) that little was known about these topics because sociologists preferred substantive research to methodological examination. The first systematic assessment of SRO surveys on standard psychometric criteria such as questionnaire construction, administration procedure, objective scoring, norms for various populations, internal consistency, retest stability, and concurrent and predictive validity was completed by Farrington (1973). He published the first demonstration that an SRO survey had predictive validity. The next important methodological and substantive assessment of SRO surveys was completed by Hindelang et al. (1981), who concluded that the reliability and validity of SRO surveys was quite good in comparison with other methods and did not vary much with the method of administration (questionnaire or interview, anonymous or not). Unfortunately, this influential and highly cited book caused a great decrease in methodological SRO research, because it was now assumed that the validity of SRO surveys had been established for all time and that methodological research was not needed. From then on, the focus was on obtaining substantive results, using self-reports as the main outcome measure in criminology.

Reliability of SRO Surveys

Reliability refers to the consistency of a measure in providing similar results under similar conditions, but this concept is generally more relevant when attempting to assess attitudes or other difficult-to-observe social constructs, as opposed to the recall of actual behavior. However, there are two criteria by which the reliability of SRO surveys can be assessed: test-retest stability (the degree to which results are consistent from one questionnaire administration to the next) and internal consistency (the degree to which items on a questionnaire all measure the same underlying construct), with the former arguably being more important than the latter. This is because it would reflect poorly on SRO surveys if the same person reported a different profile of offending during a specified time period when asked on two separate occasions (low test-retest reliability). When evaluating the test-retest reliability of SRO surveys, the time lag between the two administrations is important. If this time period is too short (e.g., a few days), participants could simply be recalling what they had said previously, whereas if the time lag is too long (e.g., several weeks), participants may forget the offenses that they had committed or may report new offenses committed in the interim.

A number of studies have evaluated the test-retest reliability of SRO measures (with various time lags). Although the results varied depending on the number and types of offenses enquired about (and the methods of scoring), they suggest that these measures are at least as reliable as measures of attitudes (Thornberry and Krohn 2003). For example, Huizinga and Elliott (1986 reinterviewed a random selection of 177 participants from the NYS 4 weeks after the initial completion of the SRO survey. They found that crime index offenses (which are more serious and less frequent) had very high reliability, while less serious and more frequent offenses (e.g., public disorder offenses) had lower but still acceptable reliability. They also did not find any consistent variation in test-retest reliability over gender, race, social class, or involvement in delinquency.

Validity of SRO Surveys

The key question for SRO surveys is validity: To what extent do self-reports produce an accurate estimate of the true number of offenses committed? And how accurately do self-reports measure the prevalence, frequency, and seriousness of offending? Setting aside the less important issues of content and construct validity, the validity of self-reports is usually assessed by comparing them with some external criterion of offending. The comparison can be concurrent (measure and criterion at the same time) or predictive (measure before criterion).

The main problem centers on what to use as an accurate external criterion of offending. Unlike drug use, for example (Harrison and Hughes 1997), there are no physical traces of burglary or shoplifting in hair, blood, or urine. Some researchers have compared SRO in the usual conditions and when people are told that their lying will be detected physiologically, and generally admissions increase in these physiological conditions. There is some evidence that admissions of problematic behavior are also greater in anonymous conditions, with audio computer-assisted interviews yielding the highest response rate (Tourangeau and Smith 1996).

Notwithstanding the fact that self-reports were intended to overcome some of the perceived deficiencies of official records, SRO results are usually validated against arrests or convictions. It is possible to compare self-reports with parent, teacher, or peer reports of offending and also to compare reports of offending with direct behavioral measures of actual offending, but there have been relatively few of these types of studies (see Farrington et al. 1980). Generally, the most important validity checks are in relation to official records.

There are two general ways in which the concurrent validity of SRO has been evaluated. The first is to compare self-reported offending (e.g., reporting breaking into a house with the intention of stealing) with official records (e.g., an arrest or conviction for burglary). The second is to examine the fraction of those who are known to have an official record who self-report official offending or to compare self-reported arrests or convictions with official records of arrests or convictions.

The level of correspondence between SRO and official records is influenced by the severity and frequency of the delinquent behavior, with the expectation that more severe and more frequent behaviors would be more likely to result in an official record and more likely to be self-reported. For example, Maxfield et al. (2000) found that 59 % of those with one arrest self-reported that they had been arrested, compared with 73 % of those with 2–4 arrests and 85 % of those with five or more arrests (vs. 21 % of those with no arrests).

Concurrent Validity of SRO

Overall, the correlations between self-reports of offending and official records of arrests or convictions are generally substantial and statistically significant. However, correlations are not very good assessments of the validity of SRO surveys because most self-reported offenses will not result in an official record. The correspondence between self-reported offending and official records (both dichotomized) can be illustrated using a 2 × 2 table (Table 1).

In cell a (SR Yes, OR Yes) and cell d (SR No, OR No), the two measures of offending correspond, and this is treated as accurate responding. Cell b (SR No, OR Yes) is usually treated as concealment of offending by the individual and as such is commonly referred to as underreporting. There are of course other reasons why someone might not report an offense that was officially recorded, including forgetting and being innocent of that particular offense (e.g., because of plea bargaining). Cell c (SR Yes, OR No) represents one of the key purposes of SRO surveys, to identify offenders and offenses that have not been officially recorded. A lack of correspondence here can therefore not necessarily be attributed to a lack of validity, although this cell count might be influenced by exaggerated reporting. However, when self-reports of arrests or convictions are compared to official records, cell c might indicate overreporting (or faulty records).

Self-Reported Offending: Reliability and Validity, Table 1 Self-reports versus official records

	Self-report		
		Yes	No
Official	Yes	a	b
Record	No	c	d

When examining the validity of SRO, it might be expected that the probability of an arrest or conviction would increase with the intensity (e.g., frequency or seriousness) of self-reported delinquency. For example, Jolliffe et al. (2003) investigated the validity of self-reports by comparing these to the court referrals of over 800 boys and girls (age 11–17) in the prospective longitudinal Seattle Social Development Project (SSDP). A total of 626 youths self-reported at least one offense during this age range, and 230 of these (37 %) had a court referral. In contrast, 101 reported that they had not committed any offense, and only 16 of these (16 %) had a court referral. This comparison was statistically significant and suggested that court referrals were much more common among those who self-reported an offense than among those who denied offending. This measure of concurrent validity was highest for those who self-reported burglary (a serious offense) or drug offenses (the most frequent). A number of other studies have also demonstrated the concordance between SRO and official records.

Some researchers have questioned, however, whether self-reports of offending are equally valid across demographic categories. Generally, male/female and black/white ratios in American SRO surveys are lower than in official records. One possible reason for this is that male and black youth are more likely to conceal their offenses than females and white youths, but there are also a number of other potential explanations. Since the prevalence of offending is usually much greater in self-reports than in official records, there may be a "ceiling effect"; for example, if 50 % of white youth self-reported offending, the maximum possible black/white ratio in self-reports would be 2:1, whereas if 20 % of white youth were convicted, the maximum possible black/white ratio in convictions would be 5:1. In addition, self-reported offenses may not be comparable to (may be less serious than) official offenses.

Methods of scoring may influence race and gender ratios. Elliott and Ageton (1980) found that these ratios were much greater when the scoring allowed for large numbers of admitted crimes, rather than the maximum category being "three or more," for example. They concluded (p. 107) that race and gender differences were "more extreme at the high end of the frequency continuum, that part of the delinquency continuum where police contacts are more likely." Another possibility is that official records may reflect police and court bias against male and black youth, and more recorded male and black youth might in fact be innocent. Alternatively, male and black youth may be disproportionately arrested because of their aggressive demeanor. In a validity test, Hindelang et al. (1981) investigated how many offenses known to the police were self-reported and found that white males failed to report 20 % of serious offenses, compared with 57 % for black males, 50 % for white females, and 59 % for black females.

Similar findings on differential validity of SRO have been reported in other studies, but the pattern is far from consistent. For example, using data from the Pittsburgh Youth Study, Farrington et al. (1996) discovered that, although the self-reports of both Caucasian and African-American boys were significantly associated with their court referrals, the relationship was stronger for Caucasians. However, in the SSDP, Jolliffe et al. (2003) found that concurrent validity was lowest for Asians and females and that African-American males had the highest concurrent validity.

The comparison between SRO and official records has also been studied for racial and ethnic minorities in other countries. In a longitudinal study in New Zealand, Fergusson et al. (1993) found that children of Maori or Pacific Island descent were 2.9 times more likely to have a police record than children of European descent, but were only 1.7 times more likely to offend according to parent and self-reports. The researchers attributed these findings to ethnic biases in citizen reporting and police recording. In a national SRO survey in the Netherlands, Junger (1989) discovered that Moroccan and Turkish boys with police records were less likely to admit delinquency than indigenous Dutch boys or Surinamese boys. She pointed out that concurrent validity was not necessarily lower for ethnic minorities, because the Surinamese boys were black.

Another measure of the concurrent validity of SRO is the percentage of those with a court record for a particular offense who admit committing the same offense. For example, of those with a court referral for burglary, what percentage admit that they have committed a burglary? The association between self-reports and official records in this type of comparison is generally quite high. For example, in the SSDP, Jolliffe et al. (2003) found that, of the 246 youth referred to court, 94 % admitted committing at least one offense. The probability for specific offenses was lowest for robbery (38 %) and highest for marijuana use (100 %). Figures ranging from 54 % to 90 % have been discovered in similar studies, suggesting that most officially identified offenders self-report their particular offenses, but some differential validity by gender and race has been found. For example, in the study by Jolliffe et al. (2003), males were more likely to admit than females, and Asians were less likely to admit than either African-Americans or Caucasians. There were no differences in concurrent validity between African-Americans and Caucasians.

Concurrent Validity of Self-Reports of Arrests or Convictions

A more direct method of investigating validity is to compare self-reports of arrests or convictions with official records of arrests or convictions. For example, West and Farrington (1977) found that, at age 18, 94 % of convicted boys admitted that they had been convicted, while only 2 % of unconvicted boys claimed to have been convicted. Of offenses leading to conviction, 53 % were reported accurately, 34 % were reported but minimized, 7 % were reported but exaggerated, and 6 % were not reported. The probability of arrested or convicted offenders admitting their crimes is also high in other studies, with those who have a greater number of convictions being generally more willing to self-report offenses (e.g., Brame et al. 2004).

Again, there is some indication of differential validity by gender and race in the self-reporting of official contacts. Hindelang et al. (1981) found that 76 % of white male official delinquents reported that they had been picked up by the police, compared with 52 % of white females, 50 % of black males, and only 30 % of black females. These results fostered the widespread belief that the self-reports of black youths (especially those who were the most delinquent) were less valid because of concealment. Other studies, however, have not found race differences in reporting arrests. Maxfield et al. (2000) discovered that 75 % of convicted white youths admitted being arrested, compared with 70 % of black youths (not significantly different), and 76 % of convicted males admitted being arrested, compared to 60 % of convicted females (significantly different). More recently, in a large sample of Dutch adolescents, Van Batenburg-Eddes et al. (2012) found that 62 % of those with official police contacts self-reported a police contact (interrogation in the police station), but this was somewhat higher for older (age 14–15) adolescents compared to younger (age 12–13) ones: 65 % compared to 57 % (significantly different). They also found that indigenous Dutch (65 %) and Surinamese (70 %) boys and girls were significantly more likely than Moroccan boys and girls (46 %) to report their police contacts.

The differential validity of the self-reporting of arrests was explored by Krohn et al. (2013) using data from the Rochester Youth Development Study. Overall, 61 % of arrests were admitted. Arrested males (64 %) were more likely to admit their arrests than arrested females (46 %). Caucasian males were likely to overreport arrests, while African-American males were likely to underreport arrests. However, Krohn et al. (2013) also demonstrated that those who were arrested more frequently were least likely to report each arrest, regardless of race or gender. It could be that frequently arrested offenders had more difficulty in remembering all of their arrests.

There have been relatively few studies of how concurrent validity varies with age. However, Kirk (2006) compared age-crime curves based on official and self-reported arrests. The two curves were generally similar, but the self-report curve had a higher and earlier peak (at ages 18–19) than the official curve (at age 21). Kirk also found that the black/white prevalence ratio was higher for official arrests (2.3–1) than for self-reported arrests (1.3–1), but this was because whites overreported their arrests; 21 % self-reported an arrest compared to 13 % who had an official arrest.

Predictive Validity

Predictive validity is more impressive than concurrent validity, because being convicted may itself lead to an increase in the probability of admitting delinquent acts (Farrington 1977), perhaps because a person may assume that the researcher will know about convictions and therefore that concealment is futile. It is to be expected that current self-reported offending (of past acts) will predict future convictions, because many current self-reported offenders are also convicted and because future convictions are of course predicted by past convictions. A better test of the predictive validity of SRO surveys is to investigate the extent to which self-reports predict future convictions *among currently unconvicted people*.

There have only been four investigations (using three prospective longitudinal studies) of the predictive validity of SRO. In the Cambridge Study in Delinquent Development (a survey of over 400 London boys), Farrington (1973) showed that, for unconvicted boys, a measure of self-reported variety of offending at age 14 significantly predicted their probability of conviction in the following 3 years. Farrington (1989) later repeated this analysis for specific types of offenses. For example, among boys not convicted of burglary up to age 18, significantly more of those who self-reported burglary up to age 18 were subsequently convicted of burglary up to age 32, in comparison with the remaining boys who denied committing burglary up to age 18 (20 % as opposed to 2 %). Similar results were obtained for vehicle theft, assault, vandalism, and drug use, but not for shoplifting.

In the first test of predictive validity using American data, Farrington et al. (1996) classified the self-reported offending of Pittsburgh Youth Study boys into four categories based on seriousness of offending and compared this with their later petitions to court. For both samples, the predictive validity of self-reports was high and statistically significant (except for drug offenses in the oldest sample). Of the boys who reported the most serious forms of delinquency, 29 % were petitioned to court for Index Violence, compared to only 7 % of the boys who reported no delinquency. This difference was statistically significant, again showing that SRO had predictive validity. Farrington et al. (1996) also examined the extent to which predictive validity varied according to race, but found that predictive validity was similar for Caucasians and African-Americans. Perhaps the most important finding with regard to the validity of SRO was that there was a substantial increase in predictive validity by combining the reports of the boys with those from other sources (mothers and teachers).

Jolliffe et al. (2003) examined the predictive validity of SRO for males and females of different racial backgrounds. In the SSDP, they found that predictive validity was high and statistically significant for most offenses (except robbery and vehicle theft). It was generally higher for males than for females and highest for Caucasians and lowest for Asians (predominantly Chinese and Filipinos). The self-reports of Asian and African-American females had the lowest predictive validity.

Summary and Future Directions for Research

Self-reports of offending were initially developed to reveal the dark figure of crime, but they have evolved to be one of the most common outcome measures used in criminological research. On psychometric criteria, self-reports provide a reliable measure of offending. In addition, most research suggests that self-reports are concurrently valid, in that those who self-report offenses or arrests are more likely to have official records than those who do not self-report, and those who have been officially recorded tend to admit their offenses. Importantly, self-reports have predictive validity; among those with no official record, those who self-report offenses are more likely to be officially recorded in the future than those who do not self-report. This is good news for criminological researchers. The bad news is that the overall validity of self-reports conceals the fact that they appear to be more valid for some demographic categories than for others. Self-reports are clearly valid for white males, but they are less consistently valid for females, for Asians, and for African-Americans. Little is known about validity at older ages. That said, there are indications that other factors (e.g., number of arrests) might be more relevant to validity than demographic categories. More research is needed on the validity of self-reports of offending, and of self-reported arrests and convictions, to establish the optimal conditions for enquiring about these topics. Randomized experiments, such as that conducted by Enzmann (2013), are needed to study the effects of different methods of administration on the reliability and validity of different demographic categories. It is also important to establish to what extent differential validity might reflect biases in official processing. Since most criminological research is based on self-reports, it is very important that their reliability and validity should be measured.

Related Entries

▸ History of the Self-Report Delinquency Surveys

Recommended Reading and References

Brame R, Fagan J, Piquero AR, Schubert CA, Steinberg L (2004) Criminal careers of serious delinquents in two cities. Youth Violence Juv Justice 2:256–272

Burt C (1929) The young delinquent. Appleton, New York

Elliott DS, Ageton SS (1980) Reconciling race and class differences in self-reported and official delinquency. Am Sociol Rev 45:95–110

Elliott DS, Huizinga D, Ageton SS (1985) Explaining delinquency and drug use. Sage, Beverly Hills

Enzmann D (2013) The impact of questionnaire design on prevalence and incidence rates of self-reported delinquency: results of an experiment modifying the ISRD-2 questionnaire. J Contemp Crim Justice (in press)

Farrington DP (1973) Self-reports of deviant behavior: predictive and stable? J Crim Law Criminol 64:99–110

Farrington DP (1977) The effects of public labelling. Br J Criminol 17:112–125

Farrington DP (1989) Self-reported and official offending from adolescence to adulthood. In: Klein MW (ed) Cross-national research in self-reported crime and delinquency. Kluwer, Dordrecht, pp 399–423

Farrington DP, Welsh BC (2007) Saving children from a life of crime: early risk factors and effective interventions. Oxford University Press, Oxford, UK

Farrington DP, Knapp WS, Erickson BE, Knight BJ (1980) Words and deeds in the study of stealing. J Adolesc 3:35–49

Farrington DP, Loeber R, Stouthamer-Loeber M, Van Kammen WB, Schmidt L (1996) Self-reported delinquency and a combined delinquency seriousness scale based on boys, mothers and teachers: concurrent and predictive validity for African Americans and Caucasians. Criminology 34:493–517

Fergusson DM, Horwood LJ, Lynskey MT (1993) Ethnicity and bias in police contact statistics. Aust N Z J Criminol 26:193–206

Glueck S, Glueck ET (1968) Delinquents and non-delinquents in perspective. Harvard University Press, Cambridge, MA

Gold M (1970) Delinquent behavior in an American city. Brooks/Cole, Belmont

Hardt RH, Bodine GE (1965) Development of self-report instruments in delinquency research. Youth Development Center, Syracuse University, Syracuse

Harrison L, Hughes A (eds) (1997) The validity of self-reported drug use. US National Institute on Drug Abuse, Rockville

Hindelang MJ, Hirschi T, Weis JG (1981) Measuring delinquency. Sage, Beverly Hills

Hirschi T (1969) Causes of delinquency. University of California Press, Berkeley

Huizinga D, Elliott DS (1986) Reassessing the reliability and validity of self-report measures. J Quant Criminol 2:293–327

Jolliffe D, Farrington DP, Hawkins JD, Catalano RF, Hill KG, Kosterman R (2003) Predictive, concurrent, prospective and retrospective validity of self-reported delinquency. Crim Behav Ment Health 13:179–197

Junger M (1989) Discrepancies between police and self-report data for Dutch racial minorities. Br J Criminol 29:273–284

Junger-Tas J (2010) The significance of the international self-report delinquency study (ISRD). Eur J Crim Policy Res 16:71–87

Junger-Tas J, Marshall IH (1999) The self-report methodology in crime research: strengths and weaknesses. In: Tonry M (ed) Crime and justice: a review of research, vol. 25. University of Chicago Press, Chicago, pp 291–367

Kirk DS (2006) Examining the divergence across self-report and official data sources on inferences about the adolescent life-course of crime. J Quant Criminol 22:107–129

Krohn MD, Thornberry TP, Gibson CL, Baldwin JM (2010) The development and impact of self-report measures of crime and delinquency. J Quant Criminol 26:509–525

Krohn MD, Lizotte AJ, Phillips MD, Thornberry TP, Bell KA (2013) Explaining systematic bias in self-reported measures: factors that affect the under- and over-reporting of self-reported arrests. Justice Quart (in press)

Maxfield MG, Weiler BL, Widom CS (2000) Comparing self-reports and official records of arrests. J Quant Criminol 16:87–110

Short JF, Nye FI (1958) Extent of unrecorded juvenile delinquency: tentative conclusions. J Crim Law Criminol Police Sci 49:296–302

Thornberry TP, Krohn MD (2003) Comparison of self-report and official data for measuring crime. In: Pepper JV, Petrie CV (eds) Measurement problems in criminal justice research: workshop summary. National Academy Press, Washington, DC, pp 43–94

Tourangeau R, Smith TW (1996) Asking sensitive questions: the impact of data collection, mode, question format, and question context. Public Opin Q 60:275–304

Van Batenburg-Eddes T, Butte D, Van De Looij-Jansen P, Schiethart W, Raat H, De Waart F, Jansen W (2012) Measuring juvenile delinquency: how do self-reports compare with official police statistics? Eur J Criminol 9:23–37

Wallerstein JS, Wyle CJ (1947) Our law-abiding law-breakers. Probation 25:107–112

West DJ, Farrington DP (1977) The delinquent way of life. Heinemann, London

Williams JR, Gold M (1972) From delinquent behavior to official delinquency. Soc Probl 20:209–229

Semi-parametric Group-Based Method

▶ Group-Based Trajectory Models

Sensitive Questions

▶ Randomized Response Methods

Sentencing

▶ Front Door and Backdoor Sentencing
▶ Punishment as Rehabilitation

Sentencing as a Cultural Practice

Neil Hutton
Centre for Law, Crime & Justice, School of Law,
Strathclyde University, Glasgow, Scotland

Overview

Most of the scholarly literature on sentencing is written from a legal or philosophical perspective. Legal scholarship analyzes sentencing law. Philosophical work analyzes the normative debates about the aims of punishment in a liberal democratic society. This chapter examines sentencing as a cultural practice. Culture refers to sets of shared meanings or collective representations. To study culture is to examine the ways in which meanings are defined, enacted, mediated, communicated, and shared by a range of actors and audiences. A cultural analysis of sentencing is a study of how certain important meanings are represented. These include representations of moral boundaries, of justice, and of legitimate decision-making processes. David Garland argues that penal institutions have important cultural dimensions and consequences which shape penal policies and practices.

> Cultural categories, habits and sensibilities are embedded in and constitutive of our political and economic institutions. (Garland 2006)

A cultural understanding of sentencing seeks to understand sentencing as a collective practice which involves nonjudicial actors as well as judges. Garland distinguishes between different uses of the term “culture” in the sociology of punishment. On the one hand, “culture” describes a particular web of meanings which can be found empirically, for example, the local court culture of a particular jurisdiction which is shared and reproduced by the regular court actors. In this sense a culture is a more or less bounded set of customs, values, habits, and beliefs. On the other hand, “culture” can also be used analytically to describe ways of making meaning in a social setting, for example, how the sentencing process defines what it means to make just decisions about the allocation of punishment. In this sense cultural analysis is distinct from other forms of analysis, for example, political. A cultural analysis focuses on the creation of meaning, and a political analysis will focus on how a particular meaning becomes powerful and silences other potential meanings. This chapter focuses on the latter use of the concept; it provides a cultural analysis of sentencing rather than describing a particular sentencing culture (Yanow 1996).

There are three important ways in which a cultural approach provides a better understanding of sentencing.

First sentencing articulates the moral boundaries of society. In allocating different sorts of punishment to different sorts of offender, sentencing defines the boundaries between order and disorder, between the respectable and the disreputable, between the pure and the polluted, between good and evil, and between reason and emotion. Whatever particular forms punishment takes, it always performs this cultural task of ordering, separating the sacred from the profane (Smith 2008; Douglas 1966). Thus, sentencing helps to promote social solidarity.

Second, sentencing decision making performs and defends a particular definition of "justice." A distinctive narrative of justice underpins sentencing decision making in most common law jurisdictions, even in those US states with sentencing guidelines. This narrative both purports to describe how judges reach their sentencing decisions and also provides a normative justification for these decisions. This narrative, which is commonly known as "individualized sentencing," is described below.

Third, sentencing reproduces shared sets of meanings. It is a social practice and not just the action of an individual judge. Sentencing decision making, like other social action, is largely habitual, taken for granted, and unreflective. This does not mean that sentencers have anything less than a thoroughly professional, conscientious, and serious-minded approach to their work. But like all professionals, they work within a framework of meanings, perceptions, values, and motives that for most of the time are unquestioned. They are the taken-for-granted assumptions on which the challenging job of sentencing is based. Bourdieu calls this the "habitus" (Hutton 2006). Individualized sentencing forms the habitus for judges; it refers to the unquestioned, taken-for-granted cultural framework which defines both the way that decisions are made and also the way that these decisions are justified.

The chapter proceeds by looking at each of these three cultural tasks and in the final section analyzes the challenges to the dominant cultural approach to sentencing.

Boundaries of Moral Tolerance

From a broader Durkheimian perspective, sentencing decision making enacts deeper cultural meanings about moral boundaries which both bind society together and at the same time identify fractures and divisions (Smith 2008). Punishment expresses ideas about the sacred and the profane, about moral pollution, and about atonement and evil. As Philip Smith argues, "we can understand these basic, protean, cultural categories running through and under what appear to be more rational modern scientific instrumental or bureaucratic tendencies." Sentencing performs an "othering" function in all communities. There cannot be a community, "people like us," without people who are not like us. Sentencing therefore both includes and excludes. For Smith, following Durkheim, all societies have crime and punishment but the cultural meanings of these vary. They are invoked and put into practice in a political context and are always contested and contestible.

In allocating punishment, sentencing performs the job of defining these moral boundaries. In a broad sense, this involves invoking binary classifications such as good/evil, sacred/profane, pure/polluted, safe /dangerous, and insider/outsider. Sentencing decision making involves drawing these black-and-white distinctions, but it also involves more subtle shading which blurs the apparently sharp binary division and produces distinctions which are not as clear-cut in practice as they may appear in cultural theory.

Criminalization and the decisions of police officers and prosecutors about whether to proceed further with a reported incident patrol the boundaries between criminal and noncriminal. Sentencing is really about the shape and gradation of the negative side of this binary divide and also sometimes about the potential for an offender to shift back across to the positive side. Sentencing is about separating the good guys from the bad guys. It is also about establishing just how bad the bad guys are, about the possibility or impossibility of bad guys transforming themselves into good guys, and about what opportunities may be offered to help them to change.

In sentencing, the distinction between prison and the community is symbolically crucial. In those jurisdictions where the death penalty has been abolished, prison is the most severe sanction now available. In both England and Wales and Scotland, legislation provides that custody should only be used as a last resort where no other sanction would be appropriate Padfield (2011), Tombs (2004). The removal of the individual from the community by the State signifies both the power of the state and the subjugation of the body of the individual offender. The decision to imprison has therefore a qualitatively different significance from those other sanctions which deprive the offender of limited amounts of time, money, or association.

Aside from the custody/community tension, there are other important meanings being generated in sentencing. At the most serious end of the scale, there is a debate about how long serious offenders need to be imprisoned and the relativities both within offences (e.g., how to define different levels of seriousness of rape) and between offences (e.g., a rape and a serious assault). At the lower end, there are decisions about the boundary between fines and community sanctions (which have received less attention because of the symbolic and fiscal implications of custodial sentences) and the debates about whether fines are simply a form of economic regulation which carry little of the stigma of the other sanctions. The fine delivers pain while impacting minimally on the freedoms of movement, speech, association, and political participation that we call "liberty." The regulation of conduct virtually through automated bank transfers may be seen as a dystopian nightmare by some (Aas 2005), but it could also be the desirable freedom of a consumer society where we choose whether or not to conform and pay the price if we decide not to, literally the price of freedom (O'Malley 2009).

While it is true to say that the implementation of the criminal law performs the function of dividing conduct into acceptable and unacceptable, in practice the boundary is more accurately described as a sloping shelf, than a clearly defined wall. The boundary relates strictly speaking to actions, people can move from one side to the other. However, in practice we tend to think about criminals rather than criminal acts, and the label of criminal may persist independently of particular actions. Sentencing plays the important cultural function of defining the shape and texture of the boundary between "them and us" which turns out not to be a sharp binary distinction but a much more amorphous and liminal territory. The prison population is clearly visible "out there," but there are crowds milling around the prison walls.

In this section we have seen that punishment is about drawing distinctions between insiders and outsiders and that sentencing is the performance of this cultural process of ordering. So sentencing is not about exclusion or inclusion but about both. Sentencing can support discourses of redemption, desistance, and rehabilitation, but it also needs to support discourses of punishment, pain, and exclusion. Sentencing therefore cannot choose between rational and emotional responses to offending; it has to be able to sustain both of these approaches.

The Discourse of Individualized Sentencing

The discourse of individualized sentencing may be summarized as follows. In reaching their decision, judges take into account all of the facts and circumstances of the individual case. Each case is composed of a very large number of relevant factors and is therefore held to be unique. No two cases are exactly the same. Judges reach their decision by an "instinctive synthesis" (R v Williscroft [1975]VR292 at 300) of these myriad factors. By defining each case as unique, this approach is able to remain silent about consistency, another important feature of liberal definitions of justice. So individualized sentencing performs a particular sort of "justice" which privileges the specificities of an individual case over the demands of consistency. Judges use the discourse of "individualized sentencing" to defend a particular approach to making just decisions. "Individualized sentencing" therefore performs a particular cultural logic which both produces "justice" in sentencing and defends this definition against its potential critics, primarily those who argue that it is possible to determine similarities between cases and that it is ethically important to treat similar cases in a similar way. "Instinctive synthesis" has a transcendent quality because it is not susceptible to further rational explanation. There is thus an element of the sacred in the discourse of individualized sentencing. Intuitive synthesis presents sentencing decision making as being beyond the control of human agency, a function which can only be performed by those holding the office of judge.

Sentencing as a Social Practice

Sentencing is collective action made possible by shared cultural meanings and understandings. The actions of other criminal justice actors in the process play a part in shaping the judicial sentencing decision and are also shaped by this approach to decision making. Sentencing is a stage in the criminal justice process. Judges deal with cases which have been constructed by other actors. Each case proceeds through several processes of translation. The term "translation" (Latour 2005) is used here to make the point that what constitutes a "case" changes, as it passes through each stage of the process. Actors interpret information which is presented to them by others, act on this information in pursuit of their professional requirements, and pass this information on to the next set of actors in the next stage of the process. Cases are therefore constructed out of witness accounts, police reports, prosecutors' professional practices, reports written for the court by social workers, medical professionals, forensic scientists, psychologists, psychiatrists, and defense pleas in mitigation (Castellano 2009; McNeill et al. 2009; Sudnow 1965; Tata 2007; Yngvesson 1989). When judges pass sentence, their understanding of the case comes from reading documents and sometimes listening to evidence and argument. They have no unmediated access to the event that gave rise to the case. Thus, by the time that the case has reached the sentencing stage, it is defined in a particular way which already limits the options open to the judicial decision maker. The case is a more or less familiar narrative which prompts a more or less familiar ending.

So sentencing is not simply about judges nor is it solely about individual cases nor about individual decisions. Sentencing is a social practice and also a cultural practice.

Sentencing is the social and cultural performance of a particular definition of justice. The term "performance" is chosen deliberately to emphasize the importance of action. Justice is not something which only exists in philosophical debates or in sentencing texts. Justice has to be enacted every day in the decision-making processes of actors working in the criminal justice system. One of the most important cultural tasks of the sentencing process is to persuade audiences that sentences passed by the court are just. Sentencing is therefore about the creation of meaning performed by an interpretive community. Sentencing is a means of communicating, persuading, convincing, justifying, and providing an account of decisions which have a profound impact on individual offenders and carry powerful messages both about the kind of society we live in and the kind of society we might want to create.

Philosophical Aims of Sentencing

From a philosophical perspective, the main issue related to sentencing is the moral question of how punishment can be justified and what should be the aims of punishment (Honderich 1989). The infringement of individual liberty and/or the imposition of pain by the state requires to be both lawful and morally justifiable. Debates range about whether punishment should be backward looking and impose punishment proportionate to the seriousness of the offence ((just deserts) or whether punishment should be forward looking and seek to have an impact on reducing crime by deterring individuals or the general population, by making reparation to the community or the victim, by rehabilitating the offender, or by protecting the public. Most jurisdictions adhere to these aims at a general systemic level, but in terms of sentencing decision making, it is not clear what impact, if any, the pursuit of aims plays in selecting the type and severity of sentence in individual cases. This "cafeteria" approach (Ashworth 2010) offers great flexibility as any sentence can be fitted into one or more of these justifications. It enables sentences to be justified on a case-by-case basis. However, the flexibility of this suite of aims is also a weakness.

A sentence which appears to be fair in comparison with sentences for similar cases may not be an appropriate sentence for protecting the public or for enabling the rehabilitation of the offender. The cafeteria approach does not encourage any systematic or general policy approach to sentencing which makes it difficult to provide a more general rationale for sentencing to the public. Decisions are made on a case-by-case basis and if necessary justified post hoc with an account which can include any philosophical justification or combination of justifications. From a cultural perspective, the definition of justice is controlled by the judiciary. The mode of accountability is the public trust in the office of the judge.

Thus, although sentencing is a collective social and cultural practice, the way in which decisions are justified and defended is located at the level of the individual judicial decision maker. The following section describes the cultural framework of sentencing. It shows how individualized sentencing, seen as an approach to making and justifying decisions about justice, promotes particular values, allocates the power to define what is to count as justice to particular agencies and processes, and adopts particular rhetorical devices to persuade audiences of its propriety and legitimacy.

Sentencing as the Performance of Justice

Sentencing decisions perform justice. They define what counts as a "just" decision. These decisions also need to be justified. The discourse of individualized sentencing needs to explain how decisions are just and also needs to demonstrate that the process by which these decisions were reached was legitimate and persuade audiences of the propriety and correctness of this decision-making process. In this sense sentencing shares much in common with administrative decision making. Recent work in the cultural sociology of administrative justice provides a useful framework with which to analyze the ways in which sentencing seeks to construct just decisions and defend these decisions.

Sentencing is a legal decision in so far as decisions are made by judges, have the authority of the court, and are therefore legitimate. However, the way in which sentencing decisions are made accountable and justifiable has as much in common with administrative decision making as it does with judicial decision making. Although sentencing decisions are made by judges, the procedural form and mode of justification through which these decisions are made and defended are, in important ways, quite unlike other judicial decisions. Kagan (2010) argues that the characteristic function of an administrative decision is to get the work of society done while the characteristic function of a legal decision is to establish the legal coordinates of a situation in light of preestablished legal rules. In terms of Kagan's typology, sentencing decisions look more like administrative decisions than legal decisions.

Individualized sentencing is more about doing the job of making a "just" decision in an individual case than it is about applying rules to facts. "Rules are based on generalisations" (Kagan 2010, p. 11). A rule says that if A and B and C obtain, then the appropriate decision is X because X will advance the policy goals of the organization. In the choice of a type and amount of sentence, there are no rules which specify that if circumstances A, B, and C exist, then the judge must pass amount Y of sanction X (with the exception of mandatory life sentences in some jurisdictions). For example, there might be a well-known customary practice whereby offenders convicted of injury with a weapon which causes permanent disfigurement will receive a prison sentence, but this is a custom not a rule. The judicial decision about the appropriate type and severity of sanction is not rule governed, and thus the accounts provided to justify these decisions cannot be rule based.

Administrative decision making requires the decision maker to respond to complex and rapidly changing social circumstances which are not susceptible to the generalizations applied in fixed legal rules. Administrative decision making requires experience, expertise, and specialization to enable outcomes to be adjusted to meet the peculiarities of each individual case. Judicial sentencing decision making is based on the exercise of discretion in individual cases to deal with complex and multiple facts and circumstances. It is in this sense that sentencing more resembles administrative decision making than legal decision making. This is not to argue that sentencing is an administrative decision but rather that the style of decision making employed and its modes of accountability have more in common with the processes commonly employed to defend "just" decisions in an administrative context. "Individualized sentencing" thus serves as an important means of justifying decisions but not necessarily as an accurate account of the empirical process of making decisions.

Mashaw has argued that administrative justice refers to

> "the qualities of a decision process that provide arguments for the acceptability of its decisions". (Mashaw 1983, p. 16)

Justice, in this context, is about the decision maker being able to provide an account of the decision process that enables the public to make a judgement about whether or not the decision is a fair and reasonable decision (Tata 2002). Members of the public may disagree with the substantive decision, but still be satisfied that it was reached by a fair and reasonable process. Individualized sentencing serves a similar purpose for sentencing. A "just" sentencing decision is reached by taking into account all of the facts and circumstances of the individual case and coming to a judgement about the type and level of sanction required.

Christopher Hood has produced a cultural analysis of administrative justice which is employed here to help understand how "individualized sentencing" defines and puts into practice a particular concept of "justice" and also to identify the challenges to this discourse based on an understanding of its weaknesses (Hood 1998).

Legitimacy and Accountability

Mary Douglas famously proposed that social relations between individual actors and the institutional arrangements for managing these individuals could be classified along two variables which she called group and grid (Douglas 1982 quoted in Halliday and Scott 2010) (Fig. 1).

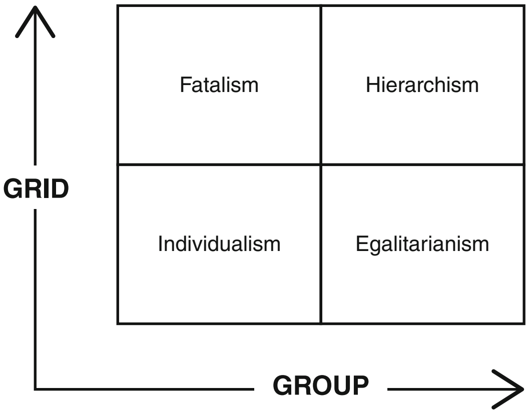

Sentencing as a Cultural Practice, Fig. 1 Grid-group analytical framework of cultural biases (From Halliday and Scott (2010))

In the context of decision-making processes, grid refers to the level of externally imposed prescriptions experienced by a decision maker. High grid means the decision makers are relatively strictly bound by rules/prescriptions over which they have little or no control, and low grid means the opposite: that decisions makers have wide-ranging discretion and are relatively unconstrained by rules. Group refers to the extent to which individuals see themselves as being incorporated into bounded units, as sharing a sense of the collective. In terms of justice, a high-grid approach believes that "justice" is most effectively delivered through a closed and controlled process and a low-grid approach implies greater openness and flexibility. A high-group approach attempts to deliver collective values, a low-group approach prioritizes individual self-interest. Group addresses questions about political legitimacy, that is about who is authorized to make decisions. Grid addresses questions about modes of accountability, about what processes make decisions transparent. So this refers to how individuals both perceive their social world and how they are able or not able to understand, challenge, and participate in decision making. It refers to the relationship between the state and the individual, a fundamental concern of political theory and law.

Examining behavior across these two dimensions gives four ideal types which can be used to exhaustively classify decision-making processes: hierarchism (high grid high group), individualism (low group high grid), egalitarianism (high group low grid), and fatalism (low group low grid). These are normative ideal types which prioritize different values in decision making. Hierarchism emphasizes expertise and skills; egalitarianism emphasises citizen participation; individualism prioritises competition; and fatalism favours stoicism and/or serendipity. As with all ideal type analysis, these are highly unlikely to exist in their pure forms in the observable world. Rather, particular decision processes will exhibit varying degrees of some or indeed all of the types, usually one being predominant.

Individualized sentencing exhibits high-group features. Authority over sentencing decision making is shared between the legislature and the judiciary. The legislature set the legal framework which judges implement. Judges are entrusted to act on behalf of the collective. In practice, the legislature leave the judiciary with wide discretion to make sentencing decisions. Individualized sentencing provides a persuasive account of how just sentences are made in order to promote the collective good.

The issue of grid is more complex. At first sight, the relative paucity of rules might suggest that sentencing is low grid, but a closer examination of the discourse of individualized sentencing suggests that it is more accurately classified as high grid. The discourse of individualized sentencing effectively restricts access to decision making to the judiciary. Individualized sentencing presents itself as the single correct way to make just sentencing decisions, an approach which can only be implemented by judges and experts by virtue of their education, knowledge, and experience and which is thus inaccessible to ordinary members of the public. Sentencing decisions are justified on a case-by-case basis. There is no set of rules which can be applied to provide a justification for the sentence chosen in each case. Judges are expected and entrusted to exercise their skill and judgement for public benefit. Citizens are not expected to participate. Individualized sentencing may not have the transparency of a grid, but it has the rigidity and stability of a grid. Individualized sentencing is thus firmly a hierarchist form of decision making.

In an individualized sentencing regime, sentencing is controlled by the judiciary. Other agents and officials play a part in the decision-making process. Defense agents have the opportunity to make a plea in mitigation. Reports provided for the court by probation or social work staff provide an opportunity for these professional groups to contribute to the decision-making process. However, ultimately the decision is made by the judge. Sentencing is "owned" by the judiciary (Tata et al. 2008). Individualized sentencing is thus substantive rather than formal, irrational rather than rational, and hierarchical rather than participatory. Hierarchical, in so far as the judge controls the process and standards for decision making and yet informal in so far as the authority of the decision, rests with the power of the decision maker and not with detailed legal rules.

Challenges to Hierarchism in Sentencing

Egalitarianism

In most late modern societies, confidence in authorities and experts is diminishing (Garland 2001). The judiciary are no exception. Judges seem to be losing public trust and may no longer be able to rely on their education, status, and tradition to sustain public confidence (Hough and Roberts 2004). The perceived monopoly of the judiciary as the "owners" of sentencing is being challenged. Kagan (2010) argues that political leaders will be comfortable with an expert judgement system so long as the electorate displays high levels of trust in the experts. However, where there is evidence that trust is in decline, leaders are likely to seek to impose a more formal and legalistic decision system with clearer accountability.

There is evidence of an increasing desire amongst citizens to have a greater input into sentencing decision making. Single-interest groups in many jurisdictions particularly those representing victims of crime and their families seek to have a greater influence on sentencing decision making. Some US states have a strong populist tradition which enables more direct citizen involvement in sentencing (California's public initiatives). A variety of sentencing institutions, usually known as councils or commissions, have been introduced in many common law jurisdictions (Hutton 2008). Some of these have the power to establish sentencing guidelines; others provide public information, education, and advice. Most of these institutions offer citizens an opportunity to influence sentencing decisions, but the discourse of individualized sentencing has proved highly resistant to change because of the powerful cultural messages that it communicates.

The history of the sentencing guidelines movement (Tonry 1996) is the story of attempts to provide a more structured approach to sentencing which allocates greater significance to the pursuit of consistency in sentencing rather than the delivery of a just sentence in each individual case. This also makes sentencing decision making more transparent and accountable and therefore more susceptible to rational debate. Individualized sentencing invites challenges about who makes sentencing and policy and what the aims for such a policy should be.

Individualism

It is hard to imagine sentencing being operated by the market, although there is increasing private sector involvement in the administration of punishment. However, it is not so hard to imagine a market approach being used to assess sentencing. In most areas of government spending such as health or education, the executive is responsible and accountable for setting budget priorities for the expenditure of public funds to manage activities in these areas. Ministers are responsible for ensuring that their officials make their decisions in a just manner, ensuring, for example, that appropriate processes have been observed and that the decisions are fair and reasonable. In common law jurisdictions where there are no comprehensive sentencing guidelines, the aggregate of sentencing decisions is, de facto, the sentencing policy for these jurisdictions. In this sense sentencing may be seen as a decision-making process which effectively allocates scarce penal resources in a particular way and provides a form of public justification for these decisions.

In an individualized sentencing approach, justice is defined with reference to an individual case and wider policy considerations about how sanctions might be allocated in a more cost-effective manner are not relevant. Decisions about justice are made by judges not policy makers although the allocation decisions of judges have significant public policy impact. Many US states are seeking to exert greater regulatory control over judicial decision making in sentencing to enable the government to exercise greater control over penal expenditure (National Conference of State Legislatures 2011). These pressures are being felt across Western jurisdictions. Politicians are keenly aware that justice clearly has a price and that for many jurisdictions, the current price is no longer affordable.

Fatalism

Fatalism is the belief that decision making is an unpredictable lottery. It can lead to cynicism and populist anger or it can lead to quietism and tolerance. There is also a strong element of fatalism in so far as citizens feel unable to influence sentencing policy and practice which is widely perceived to belong to the judiciary rather than the province of elected governments.

The strength of fatalism as a cultural bias should not be underestimated. It almost goes without saying that governments and judges need to see themselves as making positive and constructive decisions, as both being able to and having a duty to "make a difference." At the same time individuals judges, ministers, and policy makers will from time to time feel that their job is impossible. Crime will never go away and it is hard to find evidence of effective ways of dealing with crime. The media hold politicians responsible and it is not hard for them to find instances of failure. This is a cultural perspective, not an objective fact. The inevitability of crime and the apparent limitations of punishment to control crime need not lead to despair or cynicism, but rather to different ways of defining the issue and different modes of action to develop higher-group identification.

From a cultural perspective, decisions about justice in sentencing are justified by a hierarchist approach which is based on public trust in the discretionary decision making of individual judges on a case-by-case basis. This approach prioritizes particular values, most significantly those of professional expertise and experience and intuitive moral judgement. The inherent weaknesses in this approach are challenged by approaches which want to give a higher priority to different values, for example, a more rational and transparent form of accountability and a greater involvement of citizens and interest groups in decision making.

Conclusion: Why Is the Cultural Analysis of Sentencing Decision Making Useful?

Cultural analysis is useful to examine both stability and change in sentencing. In cultural terms, individualized sentencing has proved exceedingly durable and resistant to criticism. The discourse allows judges to retain control over the ways in which sentencing decisions are made and, perhaps more significantly, over how decision making is justified. The idea of fitting a sentence to the particularities of each individual case is fundamental to common sense conceptions of justice, and individualized sentencing deploys classic rhetorical devices to persuade audiences that it is the only process which can guarantee the production of justice in sentencing. To that end, justice is held to depend on scrupulous consideration of the detailed facts and circumstances of each case. The decision is made by judges and members of a respected, principled, and almost sacred profession, located above the compromised, pragmatic, world of politics. Their method is archaic, mysterious, and sacerdotal. Concerns about consistency, accountability, and rationality are presented as misguided attempts to objectify something which is inherently subjective. This representation of sentencing chimes with our common sense understandings of crime as a matter of individual responsibility, attributable to bad moral decisions made by individual offenders. It reflects the narratives of crime and punishment that form the basis of crime fiction, television crime drama, and Hollywood movies. At least part of the explanation for the resilience of individualized sentencing as an approach to sentencing lies in its insistence that "just" punishment is about the right punishment for an individual offender.

At the same time, like all justice discourses, the discourse of individualized sentencing contains the seeds of its own potential demise. As its strength is dependent on public trust in the judiciary, so it is vulnerable as this trust is perceived to be waning. Hood (1998) argues that cultural change always emerges from the perceived weaknesses of an existing state of affairs as opposed to being driven by utopian theoretical proposals premised on the possibility of starting with a blank sheet of paper. A cultural understanding tells us that any approach to decision making has both strengths and weaknesses. Change often results from the perceived failure of a process being attributed to its inherent weaknesses and solutions being sought in processes which exhibit strengths in these areas. Thus, the "problem" with individualized sentencing is its lack of transparency and limited capacity to hold judges accountable for their decisions, just as the problem with systems of sentencing guidelines is their perceived inability to take proper account of the distinctiveness of each individual case (Aas 2005).

A cultural analysis allows a better understanding of the traditional common law approach to sentencing. It shows that individualized sentencing is a political choice, not an inevitability, and that it is susceptible to challenge by approaches which offer a different version of what it means to produce justice in sentencing, for example, approaches which give a higher priority to consistency, accountability or effectiveness, and value for money. It shows that the approach to sentencing decision making in a jurisdiction arises out of political contest between politicians, administrators, judicial officers, third sector organizations, and other groups, all of which is conducted in a particular time and place and which is represented in various forms of media. It shows that there is no single objective definition of justice but rather a range of definitions over which there is continual contest.

Recommended Reading and References

Aas KF (2005) Sentencing in the age of information: from Faust to MacIntosh. Glasshouse Press, London

Ashworth A (2010) Sentencing and criminal justice. Cambridge University Press, Cambridge

Bourdieu P (1977) Outline of a theory of practice. Cambridge University Press, Cambridge

Castellano U (2009) Beyond the courtroom workgroup: case workers as the new satellite of social control. Law and Policy 31(4):429–462

Douglas M (1966) Purity and danger. Routledge and Kegan Paul, London

Douglas M (1982) Introduction to group -grid analysis. In: Douglas M (ed) Essays in the sociology of perception. Routledge and Kegan Paul, London

Garland D (2001) The culture of control. Oxford University Press, Oxford

Garland D (2006) Concepts of culture in the sociology of punishment. Theor Criminol 10(4):419–447

Halliday S, Scott C (2010) A cultural analysis of administrative justice. In: Adler M (ed) Administrative justice in context. Hart Publishing, Oxford

Honderich T (1989) Punishment: the supposed justifications. Polity Press, Cambridge

Hood C (1998) The art of the state: culture, rhetoric and public management. Oxford University Press, Oxford

Hough M, Roberts JV (2004) Public confidence in criminal justice: an international review. King's College, Institute for Criminal Policy Research, London

Hutton N (1995) Sentencing, rationality and computer technology. J Law Soc 22(4):549–570

Hutton N (2006) Sentencing as a social practice. In: Armstrong S, McAra L (eds) Perspectives on punishment: the contours of control. Oxford University Press, Oxford

Hutton N (2008) Institutional mechanisms for incorporating the public in the development of sentencing policy. In: Freiberg A, Gelb K (eds) Penal populism: sentencing councils and sentencing policy. Willan Publishing UK/Federation Press, Australia

Kagan RA (2010) The organisation of administrative justice systems: the role of mistrust. In: Adler M (ed) Administrative justice in context. Hart Publishing, Oxford

Latour B (2005) Reassembling the social: an introduction to actor network theory. Oxford University Press, Oxford

Mashaw J (1983) Bureaucratic justice: managing social security disability claims. Yale University Press, Newhaven

McNeill F, Burns N, Halliday S, Hutton N, Tata C (2009) Risk, responsibility and reconfiguration: penal adaptation and misadaptation. Punishment Soc 11(4):419–442

National Conference of State Legislatures (2011) Principles of effective state sentencing and corrections policy. NCSL, Washington, DC. www.ncsl.org

O'Malley P (2009) The currency of justice: fines and damages in consumer societies. Routledge Cavendish, London

Padfield N (2011) Time to bury the custody "threshold"? Crim Law Rev 8:593–612

Reitz K (2005) The enforceability of sentencing guidelines. Stanford Law Rev 58(1):155–173

Smith P (2008) Punishment and Culture, University of Chicago Press, Chicago.

Sudnow D (1965) Normal crimes: sociological features of the penal code in the public defenders office. Soc Probl 12(3):255–276

Tata C (2002) Accountability for the sentencing decision process : towards a new understanding. In: Hutton N, Tata C (eds) Sentencing and society. Ashgate, Aldershot, pp 399–420

Tata C (2007) Sentencing as craftwork and the binary epistemologies of the discretionary decision process. Soc Leg Stud 16(3):425–447

Tata C, Burns N, Halliday S, Hutton N, McNeill F (2008) Assisting and advising the sentencing decision process: the pursuit of 'quality' in pre-sentence reports. Br J Criminol 48(6):835

Tombs J (2004) A unique punishment? Scottish Consortium Crimed Criminal Justice. http://www.scccj.org.uk/wp-content/uploads/2011/08/A-Unique-Punishment.pdf

Tonry M (1996) Sentencing matters. Oxford University Press, Oxford

Yanow D (1996) How does a policy mean? Interpreting policy and organisational actions. Georgetown University Press, Georgetown

Yngvesson B (1989) Inventing law in local settings: rethinking popular legal culture. Yale Law J 98(8):1689–1709, 6224 words August 2012

Sentencing Commissions

Michael M. O'Hear
Marquette University Law School, Milwaukee, WI, USA

Overview

Sentencing commissions and sentencing guidelines were the preeminent institutional expressions of the sentencing reform movement that emerged in the United States in the 1970s and touched most American jurisdictions by the end of the twentieth century. This entry focuses on the history, design, and operation of American sentencing commissions. More than half of the American states have established sentencing commissions at one time or another, although several proved short-lived. At least a few commissions are believed to have played important roles in their states in advancing such ideals as uniformity, proportionality, and cost-effectiveness in sentencing policy and practice. Commissions are able to exercise influence through at least three distinct functions: collecting and analyzing sentencing-related data, advocating for or against new policy proposals in the legislature, and developing guidelines for use by sentencing judges.

The topics of sentencing commissions and sentencing guidelines are inevitably intertwined. Indeed, the core functions of many of the leading commissions have included developing and superintending guidelines. Any history or assessment of these commissions must therefore at least touch upon questions of guidelines design. However, a detailed treatment of such questions lies beyond the scope of this entry and is presented elsewhere.

Similarly beyond the scope of this entry is the history and function of sentencing commissions outside the United States, which include, for instance, commissions past or present in England, Canada, and Australia.

This entry's primary theme is the relationship between sentencing commissions and legislatures. Although sentencing commissions are predominantly legislative creations, commissions have often struggled to maintain their relevance in the face of ongoing legislative policymaking in the sentencing area, which frequently takes the form of harsh statutory responses to the "crime du jour." A secondary theme is the relationship between commissions and judges – another relationship that has sometimes proven quite challenging for the commissions to manage effectively.

Fundamentals

This section describes an early, influential articulation of the sentencing commission concept, presents case studies of two of the longest-functioning and most carefully studied commissions, and then more briefly surveys the history and design of commissions in other states.

Development of the Sentencing Commission Concept

United States District Judge Marvin E. Frankel is often credited with first proposing the idea of a sentencing commission in the early 1970s. Indeed, in his most well-known work, *Criminal Sentences*: *Law Without Order* (1973), Frankel himself characterized the sentencing commission as "the most important single suggestion in this book" (119). In particular, he saw a permanent sentencing commission as a way to mitigate systemic problems with legislative oversight of sentencing policy: "[T]he subject of sentencing is not steadily exhilarating to elected officials. There are no powerful lobbies of prisoners, jailers, or, indeed, judges, to goad and reward. Thus, accounting in good part for our plight, legislative action tends to be sporadic and impassioned, responding in haste to momentary crises, lapsing then into the accustomed state of inattention" (119).

Frankel envisioned three distinct roles for his commission. First, the commission would function as a sort of sentencing think tank, both collecting and synthesizing scholarly work in the field and designing and conducting its own studies. This function reflected Frankel's view that "the subject [of sentencing] will never be definitively 'closed,' that the process [of sentencing reform] is a continuous cycle of exploration and experimental change" (118–119). Second, putting its research and expertise to practical use, the commission would serve "as a kind of 'lobby'" with regard to sentencing issues (122). Frankel seemed to think that the commission would, in particular, represent the interests of prisoners and jailers. Indeed, he made a point of emphasizing that former or present prison inmates ought to serve as members of the commission (along with social scientists, legal professionals, business people, and artists) (120). Third, and finally, the commission would itself have some power to enact rules. Frankel did not precisely delineate the commission's jurisdiction relative to the legislature's, but suggested that the commission "could prescribe in rules of general application the factors to be considered in individual sentences, the weight assignable to any specific factor, and details of sentencing and parole procedures" (123). Thus, it seems that Frankel contemplated that the commission would promulgate what we would now call sentencing guidelines.

Minnesota's Sentencing Commission

It did not take long for Frankel's idea to achieve practical realization. In 1978, the Minnesota legislature created a sentencing commission as part of a broader package of reforms that also included the abolition of discretionary parole and the authorization of sentencing guidelines.

In broad outline, Minnesota's commission has performed each of the three roles described by Frankel. First, as to the think tank role, the commission's enabling statute directs that it "shall serve as a clearinghouse and information center for the collection, preparation, analysis and dissemination of information on state and local sentencing practices, and shall conduct ongoing research regarding Sentencing Guidelines, use of imprisonment and alternatives to imprisonment, plea bargaining, and other matters relating to the improvement of the criminal justice system" (Minn. Stat. § 244.09(6)). In furtherance of this role, the commission has gathered a "huge inventory of sentencing and corrections information," which is said to be "one of the most comprehensive and detailed databases ever assembled by any state" (Frase 2005a, b). Of particular importance has been the commission's research on pre-guidelines sentencing practices, which informed the development of the guidelines; on the corrections impact of the guidelines and proposed amendments; and on changes in sentencing practices since the implementation of the guidelines (Knapp 1987).

Second, as to the lobbyist role, the commission has participated regularly and sometimes quite successfully in legislative processes since 1978. This is not to say, however, that the commission's positions have always prevailed. For instance, a commission initiative in the 1990s to reduce the impact of plea bargaining on sentences ran into legislative resistance and proved only partially successful (Frase 2005a, b). It should also be noted that, while the commission has added an expert, well-informed voice on sentencing policy to the legislative process, it has not been quite so broadly representative a body as Frankel envisioned, at least based on its formal composition. Under the terms of its original enabling statute, the commission was to be comprised of three judges, a public defender, a prosecutor, a corrections representative, a parole board representative, and two members of the public. The commission has since been expanded to include a crime victim and a police officer. Missing, though, are social scientists, business people, artists, and (perhaps most significantly from Frankel's perspective) former or present prisoners.

Finally, as to the lawmaking role, the commission was statutorily required to promulgate sentencing guidelines for felony cases and authorized to make later amendments as necessary, subject to legislative override. The legislature initially gave the commission wide latitude in deciding which purposes and factors to emphasize in the guidelines. Notable decisions made by the commission in developing the initial guidelines, which took effect in 1980, included decisions to adopt a "prescriptive" instead of a "descriptive" approach (i.e., to seek to selectively change, rather than merely perpetuate, existing sentencing practices); to treat prison capacity as a firm constraint on guidelines severity (which implied that increasing severity levels for one offense would require corresponding severity decreases for other offenses); and to emphasize desert and offense-based considerations as primary determinants of sentence length. (Because it also gave significant weight to criminal history, the Minnesota approach was sometimes referred to as "modified desert" (Frase 1997)).

Since 1980, although the commission has retained lawmaking authority, the legislature has periodically pared back the wide policymaking discretion originally given to the commission and asserted its own lawmaking authority in the sentencing area. Often, these legislative interventions were of the type that seemed of greatest concern to Frankel: "sporadic and impassioned, responding in haste to momentary crises." For instance, in 1989, in response to a spike in violent crime, including a recent series of high-profile sexual assaults, the legislature adopted new mandatory minimum sentences for violent, sexual, and drug offenses, effectively overriding the commission's more restrained response to the crime wave (Frase 2005a, b). The legislature also amended the commission's enabling act to specify that "public safety" should be the "primary consideration" in establishing and modifying the sentencing guidelines – an implicit rebuff of the commission's emphasis on desert and prison capacity. Similarly, in 1992, after the rape and murder of two female college students, the legislature doubled the guidelines' presumptive sentences for sex offenses and imposed new mandatory minimums. In short, if an important objective of sentencing commissions is to preempt "sporadic and impassioned" penal legislation, the Minnesota commission has hardly been an unqualified success.

Yet, despite its periodic marginalization in the policymaking process, there are good reasons to regard the history of the Minnesota commission as a demonstration of the merit of Frankel's original vision. Indeed, the bare fact that the commission has survived and remained institutionally relevant for 35 years – operating in a deeply politicized field that has been repeatedly buffeted by waves of public outrage and wild swings in intellectual fashion – must itself be counted a notable success. With a half-dozen professional staffers, the commission continues to produce a detailed annual statistical report for the legislature, as well as ad hoc reports on issues of particular concern. The commission also regularly adjusts the guidelines in light of new legislation and other considerations, reports on the projected fiscal and racial impact of legislative proposals, and offers training and technical assistance for guidelines users (Minnesota Sentencing Guidelines Commission 2012). Through such activities, the commission has likely contributed to uniformity, rationality, and cost-effectiveness in Minnesota's sentencing policies and practices (Frase 2005a, b).

The commission may thus deserve some credit for Minnesota's remarkably low incarceration rate and success in avoiding the extreme prison overcrowding that has plagued so many other American states. On the other hand, Minnesota's incarceration rate was already quite low at the time the commission was formed, which suggests that preexisting aspects of the state's legal or political culture may have played an even more important role than the commission in restraining penal excess during the "get-tough" era of the 1980s and 1990s. Moreover, Minnesota's prison population has grown dramatically since 1978, even if the rate of growth has been somewhat below American norms. The commission appears to have been especially ineffective in resisting upward severity pressure in the politically charged areas of sex and drug crimes; this ineffectiveness has arguably resulted in sentences in these areas that are markedly out of step with the "just deserts" philosophy that originally animated the Minnesota guidelines (Stuart and Sykora 2011). Additionally, it should be noted that Minnesota has one of America's highest levels of racial disparity in its prison population. While these disparities seem to result largely from a complex interaction of social disadvantage, behavioral differences, and policing strategies, the commission's decisions on calculating and weighing criminal history in the guidelines may also play an important, if unintentional, role in exacerbating the disparities (Frase 2009).

United States Sentencing Commission

Although the Minnesota sentencing commission may claim the title as America's oldest, the federal commission has undoubtedly been the nation's most prominent and intensely scrutinized. Enmeshed in a political culture that seems far more polarized than Minnesota's, the federal commission has endured repeated, vociferous attacks from the both the left and the right. Perhaps even more than the Minnesota experience, the federal experience casts considerable doubt on the capacity of sentencing commissions to cure the political pathologies that Frankel associated with penal lawmaking.

As with Minnesota, proposals for a federal commission and guidelines date to the 1970s, although enabling legislation was not enacted until 1984. Once created, the federal commission faced a statutory mandate very similar to Minnesota's: gather and analyze data, make recommendations to Congress, and promulgate (and as necessary amend) sentencing guidelines (28 U.S.C. § 994). The federal commission was to be more robe-heavy than Minnesota's (three of seven voting members must be judges), but the federal statute was otherwise less prescriptive as to composition, leaving the matter to the dynamics of Presidential nomination and Senate confirmation.

One might imagine that appointment by a single president would result in a cohesive initial group of commissioners. It turned out, however, that the first commissioners were deeply divided philosophically, and one of them even publicly dissented from the guidelines the commission eventually promulgated. These divisions seem to have had important consequences for the guidelines' structure and may have contributed to perceptions that the guidelines embodied a set of unprincipled compromises (Stith and Cabranes 1998). In fairness to the first federal commission, though, it must be noted that the federal enabling statute, unlike Minnesota's, included a long list of directives regarding the content of the guidelines, many of which seemed to point in different directions (O'Hear 2006). In any event, notable decisions made by the federal commission in developing the initial guidelines, which took effect in 1987, included decisions not to endorse an overarching purpose of sentencing (in contrast to the Minnesota commission's embrace of desert); to employ the descriptive approach (except in the several important areas in which the enabling statute expressly favored harsher sentences); and to measure offense severity based not merely on the formal offense of conviction, but on a plethora of "real-offense" factors. The latter decision resulted in a set of guidelines marked by an extraordinary degree of complexity, which (along with their aggressive implementation of Congress's severity-enhancing directives) has been one of the chief sources of their unpopularity.

Since the guidelines' promulgation, the federal commission's path has paralleled that of the Minnesota commission: a strong record of data-collection and analysis has been accompanied by a more mixed record of resisting legislative interference with the integrity of the original guidelines system. With a staff of about 100, the federal commission's basic activities are similar to those of its Minnesota counterpart (publication of reports, amendment of guidelines, technical support for guidelines users, and so forth) but on a considerably larger scale.

Many commentators have effectively documented the troubled relationship between Congress and the commission, especially Congress's tendency to override the commission's policy choices in areas of particular public concern (e.g., Friedman and Supler 2008; Bowman 2004). One illustration may suffice for present purposes. In 1986, in response to a public outcry over crack cocaine and without waiting to see how the commission would deal with drugs in its pending guidelines, Congress enacted new mandatory minimum sentences for drug offenders. The law was especially tough on crack offenders: it only took 5 g of crack to trigger a 5-year minimum and 50 g for a 10-year minimum. Importantly, the corresponding weight thresholds for powder cocaine were 100 times greater, resulting in a wide disparity in the treatment of powder and crack defendants who were trafficking in similar volumes. Although not expressly required to follow suit, the commission nonetheless chose to incorporate the 100:1 ratio into its drug guidelines, thus extending the disparity across the full range of covered quantities, including those below the mandatory-minimum thresholds. Subsequent analysis by the commission, however, casts doubt on earlier assumptions that the crack form of cocaine was intrinsically more dangerous than the powder form, and also identified troubling racial disparities that arose from the 100:1 ratio (crack defendants were disproportionately black, while powder defendants were disproportionately white). Accordingly, in 1995, the commission promulgated an amendment in order to equalize the guidelines' treatment of crack and powder. Congress, however, rejected the amendment. In a series of reports over the following dozen years, the commission continued to urge Congress to soften the 100:1 ratio but to no avail. Finally, in 2007, the commission was permitted to reduce the ratio in the guidelines without congressional interference, and Congress itself then followed suit by reducing the statutory ratio to 18:1 in 2010. Although the commission eventually had some success in achieving legal change, the protracted nature of the process hardly demonstrates a high level of congressional deference to the commission's expert judgments.

In a thorough assessment of the first 15 years of sentencing under the federal guidelines, the commission claimed credit for an increase in transparency and predictability in sentencing, as well as a decrease in inter-judge disparity (United States Sentencing Commission 2004). The commission observed that it had developed "huge" databases, representing the "richest sources of information that have ever been assembled on federal crimes, federal offenders, and sentences imposed," and that its prison impact model had proven reliable in projecting prison bed and supervision needs. And those needs had become considerable; the commission noted a large increase in sentencing severity over the guidelines era. Not only were a larger percentage of federal defendants receiving prison terms, but those who were sent to prison were facing terms that were twice as long, on average, as they had been before the guidelines. Some of the increase in severity was due to congressional policy choices, such as the 1986 mandatory minimums for drug offenders, but the commission's policy choices also played a role, as in the initial decision to integrate the statutory 100:1 ratio into the guidelines. Moreover, the increase in severity was borne disproportionately by black and Hispanic defendants; the gap in average sentences between white and minority defendants grew rapidly in the guidelines era. Much of the black-white disparity was attributed to the then-prevailing 100:1 ratio.

Shortly after the 15-year report, the United States Supreme Court, in its 2005 decision in *United States v. Booker*, transformed the federal guidelines from mandatory to advisory. (Note that this decision affected only the federal guidelines, leaving intact the mandatory or presumptive character of the sentencing guidelines in Minnesota and a number of other states.) *Booker* fundamentally changed the role of the federal commission, which could no longer simply dictate policy to sentencing judges; rather, if the commission wished to maintain judicial compliance with the guidelines, it had to persuade judges that its policy choices actually merited deference. The Supreme Court made this clear in its 2007 decision in *Kimbrough v. United States*, in which the Court held that sentencing judges were no longer bound by the guidelines' 100:1 ratio. In so holding, the Court observed that the crack guidelines did not exemplify the commission's exercise of its "characteristic institutional role" because the commission relied on the statutory minimums and did not take account of "empirical data and national experience." The lesson seems to be that the commission's efficacy as a policymaking body in the future may depend in large part on its conformity to Frankel's vision of an independent, data-driven, expert agency.

Other State Sentencing Commissions

Although the Minnesota and federal sentencing commissions have been among the longest continually functioning and most carefully studied American sentencing commissions, they have hardly been alone. More than 30 other states have had sentencing commissions at one time or another (Barkow and O'Neill 2006). In some states, such as South Carolina and Alaska, the commissions were designed as temporary bodies. In other states, such as Wisconsin and Florida, the commissions were supposed to be permanent but were later abolished. In addition to Minnesota, some of the other states with well-regarded, long-established, still-functioning sentencing commissions include Kansas, North Carolina, Pennsylvania, and Virginia.

As of this writing, the National Association of Sentencing Commissions lists 22 active state sentencing commissions on its website (including a commission in the District of Columbia), although at least two of these have much broader mandates than just sentencing and at least one has recently been legislatively abolished. It is possible that states will be encouraged to develop a new wave of sentencing commissions in coming years by the American Law Institute's Model Penal Code: Sentencing project. The MPC: Sentencing "recommends to all American jurisdictions that they establish a permanent sentencing commission . . . as an essential agency of the criminal-justice system" (American Law 2007, 47).

As the Minnesota and federal case studies illustrate, sentencing commissions are often formed in connection with the abolition of parole and initially tasked with developing mandatory sentencing guidelines; in such jurisdictions, superintending the guidelines then typically becomes a core, ongoing function of the commission. A number of states, however, have not followed the Minnesota-federal model in all of these respects. In Florida and Michigan, for instance, the judiciary initially developed sentencing guidelines; commissions were only created later (Little Hoover Commission 2007). In some states, including Alaska and Tennessee, guidelines were maintained after the commission expired or was abolished. Also, in some commission states, such as Wisconsin and Virginia, the guidelines were advisory (as they now are in the federal system). In still other states, such as Louisiana and Massachusetts, commissions have operated without any guidelines at all (American Law 2007). Nor does having a commission necessarily imply the elimination of discretionary parole; both institutions existed for a time in Delaware, Pennsylvania, Virginia, and Wisconsin (Frase 2005b).

Although they vary widely in size, composition, and budget, state commissions typically have a larger and more diverse membership than the federal commission (Frase 2005b). The North Carolina commission, for instance, has 30 members, and the Virginia commission 17.

In states in which they were abolished, commissions failed for a variety of reasons. Often, the difficulties stemmed from either the resistance of judges to efforts to control their sentencing discretion or from the legislature's marginalization of the commission in making sentencing policy (Little Hoover Commission 2007).

Key Issues and Controversies

Have Sentencing Commissions Succeeded in Lessening Legislative Tendencies to Excessive Harshness?

Judge Frankel proposed the sentencing commission as a response to the problem of "sporadic and impassioned" legislative action. In this era of punitive populism, such legislative action has typically been on the harsh side, contributing to an increase in the American incarceration rate from 93 per 100,000 in 1972 to 500 per 100,000 in 2010 (Bureau of Justice Statistics 1982, 2012). Many critics contend that this ballooning incarceration rate reflects policies that "are too severe, waste lives and money, and often produce unjust results" (Tonry 2004, 3).

As Frankel's analysis suggested, a sentencing commission might bring about greater rationality and restraint in a state's criminal-justice system in at least three ways: by producing data to help policymakers better understand the costs and benefits of different policy options, by advocating on behalf of sound policies, and by creating sentencing guidelines that embody such policies. And there is some anecdotal evidence supporting the hope that commissions do sometimes succeed in reining in legislative excess (Barkow 2012; Wright 2002). Indeed, at least one multistate study finds a statistically significant, inverse relationship between the presence of a commission and growth in corrections spending (Barkow and O'Neill 2006).

Yet, it is plain that commissions are not always successful; the Minnesota and federal case studies both provide illustrations of commission marginalization. Having a commission does not ensure either a low or a stable incarceration rate. Minnesota's prison population was the nation's second-fastest growing in the first decade of the twenty-first century, while the federal system was tied for sixth-fastest (Bureau of Justice Statistics 2012). Moreover, two sentencing commission states, Alabama and Louisiana, are among the top five by incarceration rate. Commission states with low incarceration rates, such as Minnesota (notwithstanding its recent growth spurt) and Massachusetts, tend to be places that had relatively low rates even before their commissions were created; it is hard to say whether and to what extent the commissions deserve credit, as opposed to preexisting aspects of the states' political and legal culture.

It is even possible that some commissions may be counterproductive. The federal commission, for instance, is criticized for sometimes amplifying, rather than muting, ill-advised legislative policy choices, as with the 1986 drug mandatory minimums (Barkow 2012). Moreover, as Bowman argues, a complex guidelines system, like that of the federal commission, may invite, rather than discourage, legislative intervention; the federal commission, he observes, "created a mechanism that permits endless legislative tinkering in response to the crime du jour" (Bowman 2005, 250). Indeed, even a commission's data-collection and reporting activities present risks; as Wright notes, "A system that monitors sentencing practices carefully also keeps criminal justice closer to the top of the public agenda. Growth [in the prison population] may be built into this system" (Wright 2002, 90).

In the end, we cannot know for certain which commissions, if any, have truly made a long-term difference in reining in legislative excess. We cannot replay history with and without a commission, or randomly assign states to a commission or no-commission condition. States with well-financed sentencing commissions are a self-selected lot; we should not be surprised to see that these states also tend to pay greater attention to their commissions and to display greater penal restraint than other states. Whether and to what extent commissions serve to reinforce preexisting, positive tendencies in these states – although theoretically plausible and anecdotally supported – seems beyond conclusive determination.

Why Do Legislatures Create Sentencing Commissions?

In a notable series of entries, Barkow has considered the sentencing commission, not through the lens of penal policy (as it is usually considered), but through the lens of administrative law theory. One interesting question that she has posed, along with her coauthor Barkow and O'Neill (2006), is why legislatures create sentencing commissions in the first place. Although Frankel supplied good reasons to think that a commission would be in the public interest, public-choice theories suggest that legislatures are not likely to enact laws purely on the basis of public interest. In the standard explanatory model of legislative delegation of authority to an administrative agency, the legislature does so in order to avoid making a difficult decision that will inevitably offend one powerful interest group or another; legislators can then claim credit for addressing a matter of public concern without actually risking interest-group support. Sentencing, however, does not seem to fit the model: no powerful interest group opposes longer sentences, so legislatures seem free to adopt tougher sentencing laws without fear of reprisal.

Barkow and O'Neill empirically tested a number of possible explanations for commission-creation. They found statistically significant relationships between a state's likelihood of having a commission and such independent variables as a narrow partisan margin in the state, a high incarceration rate, and a high rate of corrections expenditures. In these findings, they see support for a cost-centered account of sentencing commissions: legislators support commissions in states in which corrections spending is sufficiently high so as to threaten other legislative priorities (such as keeping taxes low) and in which a close partisan divide makes it especially difficult for politicians to exercise restraint in sentencing policy and thereby risk appearing "soft on crime." In such states, delegating sentencing policy to a commission may indeed help legislators to avoid making politically risky choices.

What Actors and Activities Should Sentencing Commissions Regulate?

If, like Barkow, we think about sentencing commissions in relation to conventional government regulatory agencies, we might ask what actors and activities are regulated by sentencing commissions (Barkow 2005). Based on the example of the Minnesota commission and others that have followed in its footsteps, the answer seems to be that sentencing commissions regulate judges and, particularly, the decisions that judges make about whether and for how to long to incarcerate convicted felons. However, a sentencing commission's regulatory mission might be conceived much more broadly along both the "who" and the "what" dimensions. Indeed, if the fundamental objectives of a commission are understood to be something like achieving uniformity and proportionality in punishment or allocating criminal-justice resources in the most cost-effective fashion, then the narrowly focused approach of the Minnesota commission may be less than optimal or perhaps even counterproductive.

Consider, for instance, what may be the most perilous "third rail" for sentencing commissions: the regulation of prosecutorial plea bargaining. Plea bargaining is without question an important source of unwarranted disparities in the criminal-justice system. In jurisdictions with unregulated judicial sentencing discretion, judges could, in principle, offset some of these disparities. Depending on the overlap of statutory sentencing ranges, for example, a defendant who was able to negotiate an unusually generous reduction in charges might nonetheless ultimately receive the same sentence as other similarly situated defendants who were less fortunate in their negotiations. Sentencing guidelines, however, can impede such corrective measures, thus potentially resulting in even more prosecutor-created disparity at the same time that judicial disparity is being controlled. Indeed, the federal commission adopted "real-offense" sentencing in its guidelines precisely in order to limit the significance of prosecutors' charging and plea-bargaining decisions. Yet, federal prosecutors have proven remarkably adept at manipulating the guidelines so as to achieve desired sentencing results (Bowman and Heise 2002), which may or may not be consistent with such ideals as uniformity, proportionality, and efficiency. Indeed, in its 15-year report, the federal commission itself identified prosecutor-created disparity as an important, unresolved issue in the federal criminal-justice system. In any event, if real-offense sentencing has proven disappointing as an indirect means of regulating plea bargaining – and whatever its benefits, real-offense sentencing has undoubtedly created large complexity costs for the federal system – then more direct means might seem in order, such as the adoption of plea-bargaining guidelines. But this would put a sentencing commission at odds with a politically powerful interest group, prosecutors. Commissions have generally avoided such confrontations, and even modest commission efforts to address plea bargaining, such as the Minnesota commission's initiative in the 1990s (Frase 2005a, b have drawn strong resistance.

A thorough examination of all of the other things that commissions *might*, and in a few cases actually *do*, regulate lies beyond the scope of this entry. A few examples will serve to illustrate the range of possibilities. The North Carolina sentencing commission, for instance, has addressed misdemeanors and non-prison sentences in its guidelines (Wright 2002). Washington created a special commission to oversee juvenile sentencing in 1977, and then later transferred its responsibilities to the state's adult sentencing commission (Boerner and Lieb 2001). Hoping to better coordinate sentencing and release decisions, Chanenson (2005) has proposed a "Super Commission" that would promulgate both sentencing and parole release guidelines. A commission might also seek to regulate what sanctions are imposed for violations of the terms of probation or supervised release, as the federal commission does. Indeed, the federal commission has promulgated guidelines concerning a wide range of subjects that many state commissions have not tackled, including sentencing and plea-acceptance procedures, fines, restitution, and organizational sentencing.

Why don't commissions regulate more broadly? In part, this reflects limitations on their statutory authority. However, commissions have not always taken full advantage of what authority they do have. The Minnesota commission, for instance, could have, but did not, regulate the conditions of non-prison sentences (Frase 2005a, b). Moreover, even when statutory authority is lacking, commissions could request amendments to their enabling acts. However, as suggested by the discussion of plea bargaining above, another important constraint has doubtless been a desire by commissions and legislatures alike to avoid interagency turf battles; commissions have had enough political difficulty when their regulatory targets have been limited to sentencing judges that it is understandable why they might be less than enthusiastic about tackling prosecutors, juvenile courts, parole boards, probation offices, and so forth. And, even when it comes to regulating adult sentencing judges more fully (e.g., addressing misdemeanors and non-prison sentences), commission restraint likely reflects some combination of concerns regarding judicial resistance and resource limitations. Then, too, a commission focus on prison sentences is perfectly consistent with the Barkow-O'Neill hypothesis that commissions are formed primarily to prevent corrections costs from interfering with other legislative fiscal priorities; given the high cost of imprisonment relative to community corrections, preventing overutilization of prison seems the most direct and effective way to put a lid on state corrections spending.

How Can the Integrity of Commission Policy Choices Be Preserved From Legislative Incursions?

Although legislators may create sentencing commissions in the hope of relieving political pressures to adopt costly new sentencing laws, legislators do not seem able to keep their hands entirely off sentencing policy afterwards, even when the commission has created a comprehensive guidelines system. Legislative incursions threaten the coherence of state sentencing policy and may severely impair a commission's ability to achieve such objectives as proportionality and uniformity in punishment and stability in the size of the state prison population. Although ongoing legislative engagement with sentencing policy seems unavoidable – and, in light of democratic values, is probably even desirable at some level – a pressing question for commission design and operation is how to minimize the likelihood that the legislature will simply ride roughshod over commission policy choices; ideally, legislative engagement should be informed by commission data and expertise and should avoid large-scale disruptions to commission-created sentencing systems in the absence of some justification stronger than public outrage over the crime du jour.

Although no state seems to have found a foolproof formula, a number of devices seem potentially helpful. First, commission membership may help to strengthen the commission's relationship with the legislature or otherwise to enhance its political standing. Thus, the American Law (2007) recommends that sentencing commissions actually contain sitting members of the state legislature. Harkening back to Judge Frankel's ideas about commission membership, the ALI's model legislation also contemplates a broadly inclusive commission with representation from a range of stakeholder groups; such an approach may enhance the commission's political legitimacy. By contrast, the relatively small and judge-heavy composition of the federal commission may be one source of its political weakness.

Second, Barkow (2012) observes that reliable cost-forecasting has played a particularly important role in raising the political standing of some state commissions with their legislatures and in helping these commissions to head off costly legislative proposals with low crime-reduction benefits. This experience underscores the need for commissions to develop strong data-collection and analysis capabilities, as well as the value of statutes requiring a fiscal impact analysis of sentencing bills as part of the regular legislative process. At the same time, Wright (2002) offers an important cautionary note:

although commissions may find their greatest success by playing the role of corrections accountant, this orientation may come at the expense of promoting and contributing to public debate on some of the deeper ethical questions raised by sentencing policy.

Finally, Bowman (2005) suggests that a commission should cultivate a good relationship with judges; they are, he observes, "natural allies" in checking legislative excess in the sentencing area. He notes that some of the political weakness of the federal commission may be attributed to the adversarial relationship that developed between the commission and the judiciary in the first few years of the guidelines era. This poor relationship may seem counterintuitive, given the prevalence of judges among the commission's members. However, the judge-members were not selected by the judicial branch itself and were not necessarily reflective of mainstream judicial perspectives; the original commissioners, for instance, were tilted toward the appellate courts and had little actual sentencing experience (Stith and Cabranes 1998). By contrast, for instance, one version of the ALI proposal provides for selection of judge-members by the chief justice of the state supreme court and requires that three of five judge-members be trial-court judges (American Law Institute 2007). In any event, whether strengthened or weakened by the federal commission's membership, many judges quickly developed various negative impressions of the commission: that it was overly concerned with controlling sentencing judges and not sufficiently concerned with achieving just outcomes, that its guidelines were too harsh and too rigid, and that it was oblivious to courtroom realities and disinterested in feedback from the judges "in the trenches." This federal experience suggests that a commission wishing to maintain a good relationship with its "natural allies" should try to preserve a substantial measure of flexibility and discretion in the sentencing system, engage in regular dialogue with the trial bench, and demonstrate a willingness to modify guidelines based on judicial feedback.

How Can Sentencing Commissions Remain Relevant in an Advisory Guidelines System?

With Minnesota establishing the model, many commission jurisdictions gave their commissions the authority to promulgate and/or amend sentencing guidelines that had some level of legally binding force. In such jurisdictions, the commission has an obvious and important policymaking role (assuming the legislature does not routinely override commission policy choices). However, some commission jurisdictions chose to adopt advisory, instead of mandatory guidelines, which raises the question of whether a commission can be relevant when judges are free to ignore the commission's guidance. This question has taken on greater urgency in the wake of *United States v. Booker* (2005), which shifted the federal guidelines from mandatory to advisory, and a related line of constitutional decisions that led at least four other states to make the same transition (Pfaff 2009).

Of course, regardless of its policymaking authority, a commission may have an impact through the other two roles envisioned by Frankel, those of think-tank and lobbyist. But there are good reasons to think that a commission may also affect sentencing practices through advisory guidelines. One statistical study, for instance, found that the adoption of advisory guidelines in some states led to reduced variation in the length of sentences; although the effects were not as strong as those associated with the adoption of binding guidelines, they were nonetheless substantial (Pfaff 2006). Likewise, the *Booker* decision did not immediately render the federal guidelines irrelevant; since the Supreme Court's decision, most federal sentences have continued to be imposed within the guidelines range. Even when they are not required to follow it, some judges apparently (and understandably) appreciate and rely on expert guidance in discharging what many regard to be the most challenging responsibility in the judicial portfolio.

This is not to say, however, that advisory guidelines will inevitably succeed. Wisconsin and Louisiana, for instance, both tried and abandoned experiments with advisory guidelines

(Pfaff 2009). Moreover, since *Booker*, there have been persistent calls for Congress to reinstitute mandatory federal guidelines, prompted in part by steadily declining rates of within-guidelines sentences.

Procedural rules may enhance the relevance of advisory guidelines (and by extension of the commissions that superintend them). Following *Booker*, for instance, the Supreme Court made clear that guidelines ranges (though advisory) must still be calculated in each case; that determining the guidelines range should be the *first* step in the sentencing analysis; and that sentences within the range may be treated as presumptively reasonable on appeal. When the sentencing judge invests considerable effort in determining a guidelines range, it seems likely that the range will tend to condition the sentence imposed, even if the range is not regarded as controlling – all the more so if the range constitutes a safe harbor of sorts from the threat of appellate reversal.

Apart from such procedural rules, the commission in an advisory system may enhance the impact of its guidelines to the extent that they are persuasively grounded in sound cost-benefit research and reflect an appreciation of judicial perspectives. The commission might think of its "lobbyist" role as not merely limited to legislative policymaking decisions, but also extending to judicial sentencing decisions. This way of framing the commission's role in an advisory system underscores the importance of the commission's standing with the judiciary and the strength of the justifications it offers in support of its guidelines. In this spirit, a number of the suggestions made in the previous section regarding commission design and operation (e.g., strong representation by *trial* judges on the commission) may also be pertinent to the goal of maximizing compliance with advisory guidelines.

Future Directions

This concluding section briefly describes three important challenges that some commissions are beginning to face, and that may fundamentally alter the role of sentencing commissions in the coming years.

First, although many sentencing reformers hoped that sentencing commissions and guidelines would lessen the effects of racial discrimination in the criminal-justice system, wide racial disparities in incarceration have persisted or even grown worse in some commission jurisdictions, including Minnesota and the federal system. It does not seem that commissions have engaged in purposeful discrimination, but they have sometimes made policy choices with important racially disparate effects, such as the federal commission's original approach to crack sentencing and the Minnesota commission's treatment of criminal history. In any event, whatever the source of disparities, they seem to present an increasingly urgent threat to the legitimacy of the criminal-justice system. Commissions that concern themselves broadly with the system's fairness and effectiveness ought to pay heed. Indeed, in a few states, the legislature has prodded the commission to do so, for instance, by requiring the commission to prepare racial impact reports on sentencing bills along with a fiscal analysis (Barkow 2012). In addition to such participation in the legislative process, commissions that superintend guidelines might also do well to undertake a systematic examination of the racial impact of their own policy choices and to consider amendments so as to minimize unjustified disparities. The federal commission was a pioneer of sorts; its persistent efforts to address the 100:1 ratio were largely motivated by concerns regarding racial impact.

Second, commissions have traditionally communicated information to sentencing judges in a general way and at periodic intervals, in the form of reports, manuals, guidelines, training seminars, and so forth. In recent years, though, there has been growing interest in systems that would provide judges with real-time data at the bench or in chambers to assist with individual sentencing decisions as they arise. Such a "sentencing information system" would allow a judge to enter certain basic data about an offense and an offender, and then receive information about sentences and/or outcomes in similar cases (Miller and Wright 2005). This might include, for instance, a precisely calculated recidivism risk for an offender, or information about what probation conditions have been imposed on similar offenders in other cases. In jurisdictions in which a sentencing commission is already collecting a great deal of case-level data, the commission seems the natural agency to develop a SIS. However, putting a database created for one purpose –informing policy-level decisions by the commission and the legislature – to new uses by new users presents significant technical challenges at a minimum. The challenges would be even greater if, as some hope, different jurisdictions linked their systems and pooled their data. But the challenges may not be merely technical. For instance, one category of information that sentencing judges might like to have is cost data: what is the estimated expense to taxpayers of various sentencing options? The Missouri sentencing commission has, in fact, already attempted to make such information available (Flanders 2012). However, some critics object that cost data should not be shared with judges; this may invite a cold-blooded, cost-benefit decision that loses sight of the needs of victims or the demands of justice. Another challenge is presented by the calculation of recidivism risk. Assuming that risk forecasts may be made more accurate when data like race, sex, ethnicity, family relationships, education, income, mental health, and so forth are taken into account, difficult questions arise concerning the fairness and legality of basing sentencing decisions, even indirectly, on personal characteristics over which a defendant has little or no control, particularly when there is a history of invidious discrimination associated with those characteristics.

Finally, there is the challenge of effecting fundamental reform to an established guidelines system. Minnesota introduced the basic, widely imitated template for sentencing guidelines with the development of its two-dimensional grid more than 30 years ago. Sentencing knowledge has progressed considerably in the interim, and both public and expert beliefs seem to have shifted on a number of fundamental questions, such as the feasibility of achieving, and the desirability of attempting, offender rehabilitation in the criminal-justice system. Although commissions routinely amend their guidelines, these amendments are usually narrowly focused and rarely involve anything like a fundamental reworking of the guidelines architecture. Once a system is in place, the normal pattern seems to be that the commission becomes invested in the system's basic approach and shows little interest in reopening discussion on the big questions that had to be answered at the outset in creating the guidelines. This is all perfectly understandable, but it may become increasingly problematic as the basic models of the 1970s and 1980s fall further behind current knowledge and values. Attempts to address racial disparities and to develop sentencing information systems may exacerbate the perceived obsolescence of traditional approaches. Shifts from mandatory to advisory, and vice versa, may also necessitate structural changes in order to maximize guidelines effectiveness. Yet, fundamental change requires institutional will, political capital, and other resources that may be in short supply on many commissions. In some guidelines jurisdictions, the current systems emerged from multiyear deliberative processes and painstakingly crafted compromises among antagonistic stakeholder groups. It is hard to imagine a successful replication of such efforts. Yet, it is almost as hard to imagine that sentencing commissions and guidelines will remain viable institutions for another 30 years if they do not prove capable of periodic fundamental reform.

Related Entries

- ▶ History of Probation and Parole in the United States
- ▶ Plea Bargaining
- ▶ Proportionality
- ▶ Risk Assessment, Classification, and Prediction
- ▶ Sentencing Guidelines in the United States
- ▶ Sentencing Research

Recommended Reading and References

American Law Institute (2007) Model penal code: sentencing, tentative draft no. 1. Executive Office, American Law Institute, Philadelphia

Barkow R (2005) Administering crime. UCLA Law Rev 52:715–814

Barkow R (2012) The sentencing guidelines at the crossroads of politics and expertise. Univ Pa Law Rev 160:1599–1630

Barkow R, O'Neill K (2006) Delegating punitive power: the political economy of sentencing commission and guideline formation. Texas Law Rev 84:1973–2022

Boerner D, Lieb R (2001) Sentencing reform in the other Washington. Crime Justice 28:71–136

Bowman F (2004) Pour encourager les autres?: the curious history and distressing implications of the criminal provisions of the sarbanes-oxley act and the sentencing guidelines amendments that followed. Ohio State Crim Law J 1:373–442

Bowman F (2005) Mr. Madison meets a time machine: the political science of federal sentencing reform. Stanford Law Rev 58:235–265

Bowman F, Heise M (2002) Quiet rebellion II: an empirical analysis of declining federal drug sentences including data from the District level. Iowa Law Rev 87:477–561

Bureau of Justice Statistics (1982) Prisoner 1925–1981

Bureau of Justice Statistics (2012) Prisoners in 2010

Chanenson S (2005) The next era of sentencing reform. Emory Law J 54:377–460

Little Hoover Commission (2007) Solving California's corrections crisis: time is running out. Little Hoover Commission, Sacramento

Flanders C (2012) Cost as a sentencing factor: Missouri's experiment. Mo Law Rev 77:391–410

Frankel M (1973) Criminal sentences: law without order. Hill and Wang, New York

Frase R (1997) Sentencing principles in theory and practice. Crime Justice 22:363–433

Frase R (2005a) Sentencing guidelines in Minnesota, 1978–2003. Crime Justice 32:131–219

Frase R (2005b) State sentencing guidelines: diversity, consensus, and unresolved policy issues. Columbia Law Rev 105:1190–1232

Frase R (2009) What explains persistent racial disproportionality in Minnesota's prison and jail populations? Crime Justice 38:201–280

Friedman I, Supler K (2008) Child pornography sentencing: the road here and the road ahead. Federal Sentencing Reporter 21:83–89

Knapp K (1987) The sentencing commission's empirical research. In: von Hirsch A et al (eds) The sentencing commission and its guidelines. Northeastern University Press, Boston

Miller M, Wright R (2005) The wisdom we have lost: sentencing information and its uses. Stanford Law Rev 58:361–380

Minnesota Sentencing Guidelines Commission (2012) MSGC report to the legislature

O'Hear M (2006) The original intent of uniformity in federal sentencing. Univ Cincinnati Law Rev 74:749–818

Pfaff J (2006) The continued vitality of structured sentencing following blakely: the effectiveness of voluntary guidelines. UCLA Law Rev 54:235–307

Pfaff J (2009) The future of appellate sentencing review: booker in the States. Marquette Law Rev 93:683–715

Stith K, Cabranes J (1998) Fear of judging: sentencing guidelines in the federal courts. University of Chicago Press, Chicago

Stuart J, Sykora R (2011) Minnesota's failed experience with sentencing guidelines and the future of evidence-based sentencing. William Mitchell Law Rev 37:426–468

Tonry M (2004) Thinking about crime: sense and sensibility in American penal culture. Oxford University Press, New York

United States Sentencing Commission (2004) Fifteen years of guidelines sentencing: an assessment of how well the federal criminal justice system is achieving the goals of sentencing reform. United States Sentencing Commission, Washington, DC

von Hirsch A (1987) Structure and rationale: Minnesota's critical choices. In: Von Hirsch A et al (eds) The sentencing commission and its guidelines. Northeastern University Press, Boston

Wright R (2002) Counting the cost of sentencing in North Carolina, 1980–2000. Crime Justice 29:39–106

Sentencing Guidelines in the United States

Steven L. Chanenson
School of Law, Villanova University,
Villanova, PA, USA

Overview

Sentencing guidelines have been a central part of criminal law reform efforts in the United States since the late 1970s. Through sentencing guidelines, many American jurisdictions have attempted – with varying degrees of success – to improve their criminal justice systems by combating judicial disparity; increasing fairness, honesty, and transparency; and, in some cases, controlling costs. While a majority of states do not have them, the presence of guidelines in the federal system and a substantial percentage of high-profile states have provided sentencing guidelines with considerable visibility.

Sentencing is where the rubber meets the road in the criminal law. All of the niceties of a criminal trial, in the statistically unlikely event there was one, are over. The defendant is guilty. The system must now choose how to respond and which actor or actors should have what degree of discretion. Sentencing guidelines are a popular American tool for helping to inform and implement those choices. Depending on the method of counting, there are at least 53 full criminal justice systems in the United States, one in each state plus the federal government, the District of Columbia, and the military. Each system must decide, within the constraints of the United States Constitution (and its own Constitution if applicable), how to punish convicted offenders. Sentencing guidelines, as befitting a tool used by different sovereigns reflecting different political and legal cultures, vary quite a bit from jurisdiction to jurisdiction. There are, however, some common questions that many drafters of sentencing guidelines – whether the judiciary, the legislature, or a sentencing commission – face.

This entry will explore some of the fundamental aspects of sentencing guidelines, their structure, breadth, and challenges. This entry will not examine sentencing guidelines or their analogues outside of the United States, nor will it focus more than necessary on the sentencing commissions that often promulgate sentencing guidelines as those bodies have their own entry in this series. It will also not address capital sentencing.

This entry will, however, explore selected attributes and challenges of sentencing guidelines through the lens of some of the key sentencing guideline jurisdictions.

Fundamental Background of and Structures for Sentencing Guidelines

How Should Sentencing Systems Be Described?

Terminology describing sentencing systems is frequently confusing. A common language is thus essential. Sentencing systems are either determinate or indeterminate. "Indeterminate systems use discretionary parole release while determinate systems do not. Determinate and indeterminate sentencing schemes can take various forms. Either sort of system may be discretionary or nondiscretionary [as to sentencing]. Discretionary systems – be they determinate or indeterminate – may be guided or unguided" (Chanenson 2005, pp. 382–383). Sentencing guidelines are a popular way to provide guidance to sentencing judges, especially in discretionary, determinate sentencing systems. As noted below, sentencing guidelines can have varying degrees of force and may be best conceived of as occupying a continuum of enforceability. Common, although necessarily imprecise, terms for the relative power of the guidelines include "presumptive" and "advisory."

Why Do Some Jurisdictions Choose to Have Sentencing Guidelines?

Before the 1970s, the dominant approach to sentencing in the United States reflected a rehabilitative model and placed broad discretion in judges and parole boards. Judge Marvin Frankel and others questioned and criticized this highly discretionary approach as lawless and unfair. Academic commentators highlighted a "gross disparity in sentencing, with different sentences imposed upon similar offenders who have committed similar offenses by the same judge on different days, different judges on different days, different judges on the same day, and different judges in different jurisdictions" (Singer 1978, p. 402). US Supreme Court Justice O'Connor noted that unguided sentencing discretion "inevitably resulted in severe disparities in sentences received and served by defendants committing the same offense and having similar criminal histories. Indeed, rather than reflect legally relevant criteria, these disparities too often were correlated with constitutionally suspect variables such as race" (*Blakely v. Washington*, 542 U.S. 296, 315 (2004) (O'Connor, J., dissenting)).

Starting in the 1970s, as a result of these criticisms and other, sometimes local, concerns, various American jurisdictions reformed their approach to sentencing by implementing different forms of guidance. "One way to describe sentencing reform over the past half century is that law came to sentencing" (Miller 2004, p. 121). Some jurisdictions – most notably Minnesota and the federal government – abolished discretionary parole release and implemented a determinate system with sentencing guidelines. Other jurisdictions – most notably Pennsylvania – retained an indeterminate system but added sentencing guidelines. As the Pennsylvania Supreme Court has noted, sentencing guidelines "were promulgated in order to structure the trial court's exercise of its sentencing power and to address disparate sentencing" (*Commonwealth v. Mouzon*, 812 A.2d 617, 620 n.2 (Pa. 2002) (plurality)). Federal sentencing reform was the product of years of discussion, debate, and political compromise. Some scholars have asserted that the final product, the Sentencing Reform Act of 1984, reflected a "subtle transformation of sentencing reform legislation: conceived by liberal reformers as an anti-imprisonment and antidiscrimination measure, but finally born as part of a more conservative law-and-order crime control measure" (Stith and Koh 1993, p. 223).

Sentencing guidelines reflect the legal and policy choices of the jurisdiction. Thus, popular attitudes and opinion can play a significant role. Sentencing guidelines can spark intense debate in part because they are transparent. This law-based transparency "forces the resolution of issues of sentencing policy – and enforces the particular resolution – and thus makes the fact of resolution clear. This society is not of one mind on what values sentencing should embody and thus the resolution of these disputed issues means the values of some will prevail while those of others will not" (Boerner 1993, p. 174).

What Kinds of Structural Choices Do Legislatures Make When Starting This Process?

There are different ways to structure a sentencing system. Each jurisdiction must confront at least three common structural choices. First, should sentencing be determinate or indeterminate? Many commissions and guidelines – including in the Minnesota and federal systems – emerged as part of a package of legislation that abolishes discretionary parole release, which transforms the system from an indeterminate one to a determinate one. Other guidelines – like Virginia – began within an indeterminate structure, but the legislature transitioned to a determinate approach later for reasons similar to those motivating the Minnesota and federal systems. Finally, there are a couple of jurisdictions – most notably Pennsylvania – that continue to have both an indeterminate structure (meaning discretionary parole release) and well-developed sentencing guidelines.

Second, which actor in the system should promulgate the guidelines? In part because of their popularity and specialized expertise, permanent sentencing commissions have captured the lay imagination concerning guideline creation and maintenance, but they are not the only approach. It is possible for the legislature or the judiciary to create guidelines either themselves or through temporary bodies established for that purpose. Even when it creates a sentencing commission, the legislature often reserves for itself the power to formally create the guidelines or at least a veto power over the guidelines promulgated by the commission. A popular approach, exemplified by the federal and Pennsylvania systems, involves the sentencing commission issuing guidelines which take effect after a predetermined amount of time unless the legislature and executive affirmatively block them. This power-sharing arrangement is often described in a positive light as allowing the commission to act on its expertise with less political pressure than elected officials might feel while still respecting principles of democratic accountability.

Third, how much binding force should the guidelines have? Not all guidelines have the same amount of binding force. As exemplified by the original Minnesota and the federal approaches, powerful guidelines were initially a popular approach. Many felt that anything less binding would not rein in judicial discretion. Before a new interpretation of the Sixth Amendment to the United States Constitution emerged in the early 2000s, the federal guidelines were very binding (often called "presumptive") and judges would frequently have difficulty successfully justifying sentences that diverged from the presumptive guideline range, a process called "departing." In fact, the rigid nature of the federal guidelines – and the associated relative lack of judicial discretion – attracted severe and persistent criticism.

Terminology here can be challenging as labels like "presumptive" and "advisory" can mask many variations. Indeed, "there are an infinite number of stops between a purely advisory approach and a completely mandatory framework" (Reitz 2005, p. 157). However, a jurisdiction can influence how much power the guidelines (and, at times, functionally the prosecutor) have as opposed to the judge. One indication of a more voluntary guideline regime, exemplified by Virginia, is the absence of appellate review. The presence of appellate review generally reflects more binding guidelines, but the intensity of that review can vary widely. Releasing judge-specific sentencing information may also have an impact on guideline compliance as well as reflect a commitment to transparency (Bergstrom and Mistick 2003). It is interesting to note that Pennsylvania, a state in which judges run for election and stand for retention, releases judge-specific information while the federal government, in which judges are appointed for life, does not.

More binding or presumptive guidelines are arguably desirable because they have sufficient teeth to meaningfully limit judicial discretion. However, more advisory, and thus less binding, guidelines are arguably desirable because they provide sufficient judicial flexibility to promote fairness in a particular case. Just as there can be unwarranted disparity, there can also be unwarranted uniformity. Professor Doug Berman observed that "reformers believed that sentencing guidelines, by codifying standards which would direct judges' sentencing decisions in most but not all cases, could reduce sentencing disparities and maintain sentencing flexibility, while promoting the development of principled sentencing law and policy" (Berman 2000, p. 35). Ultimately, each jurisdiction has to find its own balance between individualization and uniformity of sentences.

How Do Sentencing Commissions Create Sentencing Guidelines?

Guideline drafters have a myriad of choices to make, and the American constitutional structure affords them substantial latitude within which to work. There are at least six decisions that can be characterized as critical:

1. What Philosophical Goals Should Animate the Guidelines?

   The legislature and/or the sentencing commission should address what philosophical goals of sentencing should be reflected in the guidelines. The standard debates about purposes of punishment play out in this process. While the rehabilitative model had been dominant, people like criminologist and law school Dean Norval Morris successfully urged that punishments should be limited by principles of just deserts. Morris argued that desert is "an essential link between crime and punishment. Punishment in excess of what is seen by that society at that time as a deserved punishment is tyranny" (Morris 1974, p. 76). Most systems have embraced Morris' idea of "limiting retributivism," pursuant to which the offender's desert defines the upper and lower limits of acceptable (not unjust) punishment and within that range other utilitarian goals can be pursued (Frase 2005, pp. 76–77).

   Although most modern American sentencing systems are now largely retributive, guideline drafters often struggle with how to articulate their views and integrate competing ideas. The United States Sentencing Commission officially sidestepped the question at the macro level when it promulgated the first set of federal guidelines in 1987. It noted that, "[a] philosophical problem arose when the Commission attempted to reconcile the differing perceptions of the purposes of criminal punishment.... Adherents [to both "just deserts" and "crime control" approaches] urged the Commission to choose between them, to accord one primacy over the other....

A clear-cut Commission decision in favor of one of these approaches would diminish the chance that the guidelines would find the widespread acceptance they need for effective implementation. As a practical matter, in most sentencing decisions both philosophies may prove consistent with the same result" (U.S. Sentencing Commission 1987, Ch 1A 1.3–1.4). Nonetheless, some scholars of federal sentencing have argued that the federal guidelines reflect a "modified just desert" theory in which "the greatest weight in determining sentences is given to matching the severity of punishment to the seriousness of the present offense," and the second most significant consideration is "the need to incapacitate for longer periods the more dangerous offenders" determined by criminal history (Hofer and Allenbaugh 2003, p. 24). This seems to be consistent with, if not a variant of, "limiting retributivism."

Putting these pieces together is not easy. Professor Frase summed it up by noting that "early guidelines reforms attempted to narrow the focus of sentencing to strongly emphasize uniformity and 'just deserts,' and to promote more rational sentencing policy. The much broader range of contemporary sentencing goals demonstrates an important underlying truth, which early guidelines reforms (and some recent proposals) seem to have overlooked: sentencing policy is very complex, requiring compromise and careful balancing of numerous, often-competing goals" (Frase 2000, pp. 435–436).

2. Should the Guidelines Be Descriptive or Prescriptive?

Sentencing guidelines can be designed primarily to reflect past judicial practices which can be described as a "descriptive" approach. This has the benefit of more likely acceptance by the judiciary as it will seem familiar in most cases, but it may not always produce the most desirable results. When the Virginia guidelines were promulgated, they enjoyed strong judicial support in part because they were largely descriptive. On the other hand, sentencing guidelines can be designed from the ground up to reflect what the sentencing commission believes is the best path even if it diverges from past judicial practice. This can be described as a "prescriptive" approach. Minnesota used a more prescriptive strategy in part to encourage incarceration for violent offenders and discourage it for property offenders. The prescriptive path may have the benefit of a more rational approach, but could encounter opposition from lawyers and judges who are accustomed to the old ways. As with so many things, the common responses can be plotted along a continuum with the bulk of guidelines avoiding the extremes.

3. Should the Guidelines Be Based More on a "Real" or a "Charge" Offense Model?

Which crime(s) should matter when sentencing a defendant? At first blush, it may seem like an odd question to ask. Should the defendant only be sentenced on the crime(s) for which he has been convicted? Does that mean that the judge must ignore compelling evidence of other criminal acts that, for whatever reason, were not the subject of the jury's guilty verdict or the defendant's guilty plea? On the other hand, should the defendant be sentenced based on every bad thing that the government alleges – and supports with proof at less than the beyond a reasonable doubt level – the defendant has ever done? Does that mean that the judge must factor in the amount of cocaine allegedly involved in a second drug charge on which the jury returned a verdict of not guilty? The restrictive approach is, at the extreme, described as "charge offense" sentencing while the inclusive approach is, at the extreme, described as "real offense" sentencing.

Charge offense sentencing runs the risk of the judge being prevented from sentencing based on everything known about the offender and the offense. It is also criticized for perhaps making it even easier for prosecutors to control the system by bringing more charges against some defendants and fewer charges against others. Real offense sentencing runs the risk of the defendant being punished on the basis of weak allegations that a jury either rejected or was never even given the chance to consider. While the extremes of both approaches may strike many people as preposterous, jurisdictions do stake out positions leaning toward one side or the other. Pennsylvania, for example, is largely a charge-based system, although judges may consider information about other behavior in exercising sentencing discretion. The federal system, in contrast, enshrines a fairly aggressive real offense scheme in its guidelines. Judges must calculate the guidelines on the basis of "relevant conduct," which encompasses acts similar to the offense of conviction. This relevant conduct must be included as long as it is proven by a preponderance of the evidence even if the government never charged those acts or the jury, employing the beyond a reasonable doubt standard, acquitted on them.

4. How, If at All, Should the Guidelines Address Multiple Convictions?

   Guideline systems struggle to deal with multiple convictions. Many people would agree that multiple offenses usually deserve greater punishments than a single offense of the same nature. But how much more? Should the increase be geometric? Is a "volume discount" appropriate for the multiple burglar? Some guidelines require consecutive sentences for certain types of offenses, such as violent crimes, but mandate concurrent sentences for other offenses, such as property crimes. Other jurisdictions, like Pennsylvania, do not speak to multiple convictions at all and thus afford the judge nearly unfettered discretion to impose sentences concurrently or consecutively. Then-Judge (now US Supreme Court Justice) Breyer criticized that discretionary model by observing that "[a] moment's thought suggests, however, that this approach leaves the prosecutor and the judge free to construct almost any sentence whatsoever" (Breyer 1988, p. 26). The federal system tries to avoid those extremes by creating a "system that treats additional counts as warranting additional punishment but in progressively diminishing amounts" (Breyer 1988, p. 27). While this federal approach may be more satisfying in theory, the practical result is a fairly byzantine process with nearly algebraic calculations that itself has been the subject of significant criticism.

5. How Should the Guidelines Deal with Mandatory Minimum Sentences?

   Despite the existence of sentencing guidelines which are designed to take a nuanced approach, many legislatures continue to create mandatory minimum provisions. Mandatory minimum sentence laws are often criticized as being crude, inconsistently applied, and in tension with the idea of sentencing guidelines. US Supreme Court Justice Breyer, who was an influential member of the original US Sentencing Commission, noted that mandatory minimum sentences "prevent the Commission from carrying out its basic, congressionally mandated task: the development, in part through research, of a rational, coherent set of punishments.... Most seriously, they skew the entire set of criminal punishments, for Congress rarely considers more than the criminal behavior directly at issue when it writes these provisions..." (Breyer 1999, p. 184). Many sentencing commissions have to decide how to respond to such conflicting legislative directives. For example, Congress created both the federal sentencing guidelines system and a series of gun and drug mandatory sentences in close succession. In fact, the mandatory sentences were passed before the new commission could even promulgate the first set of guidelines.

   There are two broad responses available to a sentencing commission. First, it could ignore the mandatory minimum sentences when crafting its guidelines. This allows the guidelines to reflect the views of the sentencing commission and simply be trumped by the mandatory minimum sentences when they apply. Doing so has the potential of creating a sentencing "cliff," which is a point where a small change in offense behavior prompts a large change in the sentence. For example, the commission may set relatively modest weight-based guidelines for a drug offense. If the amount of the drug involved reaches the level that triggers the mandatory minimum sentence, however, the defendant's sentence may increase precipitously. Sentencing cliffs are frequently criticized because of the perceived unfairness of the sharp increase in sentencing severity in response to the incremental increase in offending severity. Alternatively, the commission could accommodate the mandatory minimum scheme by integrating it into its guideline recommendations. This approach, largely followed in the federal system, avoids the "cliff" problem but requires the guidelines to be driven by the mandatory minimum sentences in a way that may diverge from what the sentencing commission would have otherwise done. The result may be greater severity for a wider swath of offenses and offenders. The Pennsylvania guidelines, providing another example of how broad labels can conceal variations, have largely declined to accommodate mandatory minimum sentences for its core recommendations. However, the guidelines have engaged with and attempted to provide alternatives to mandatory minimum sentences by providing such things as advisory sentence enhancements for conduct that might otherwise trigger mandatory minimums. It is interesting to note that Pennsylvania adopted a guideline system in part to "avoid mandatory minimum legislation that would severely restrict judicial discretion" (Kramer and Ulmer 2009, p. 15).

6. How Should the Sentencing Recommendations Be Communicated?

   The most common way to reflect the jurisdiction's sentencing choices and recommendations is through a simple two-axis grid, although there are other approaches such as narratives. This two-axis grid commonly reflects the defendant's criminal history along the horizontal axis and the severity of the crime along the vertical axis. Where those two paths intersect is the box or cell containing the guidelines' recommended sentence or sentencing range, which is often expressed in months of incarceration (*see* Fig. 1).

   Some jurisdictions, including Minnesota and Pennsylvania, have different grids for certain crimes (like sex offenses) or different ways in which crimes are committed (like with the possession or use of a deadly weapon). These separate grids are often explained as a way to reflect the particular seriousness of the targeted behavior without skewing the recommended sentences for the majority of offenses. This approach has been praised for limiting the impact of popular pushes for severity to the target offenses/offenders and has also been criticized for allowing that push for severity concerning those offenses/offenders to proceed more easily by bifurcating it from the bulk of cases.

The Mechanics of the Federal Sentencing Guidelines

Sentencing guidelines require the judge and lawyers to engage in a series of calculations that yield a particular recommendation, often as represented by a box or cell on the guideline grid. Guideline calculations vary in complexity depending on the jurisdiction. The federal system is widely viewed as the most complicated and is often criticized on that basis. Despite its distinctive federal complexities, a review of the kinds of calculations a federal judge needs to make before imposing a sentencing may afford a useful window on guideline sentencing generally.

There are 43 levels of what the federal guidelines call "offense seriousness" (US Sentencing Commission). In order to determine the ultimate offense level, the judge starts with the "base offense level" for the offense of conviction. Logically, offenses the Commission deems to be more serious are assigned a higher base offense level. The judge must then consider "specific offense characteristics," which are factors related to the offense such as the amount of money taken in a robbery or the quantity of drugs trafficked. The more the judge finds as taken or trafficked, the greater the increase in the offense level. At this point, the federal system's modified real offense feature can sweep in conduct for which there was no conviction. Next, the judge must determine whether

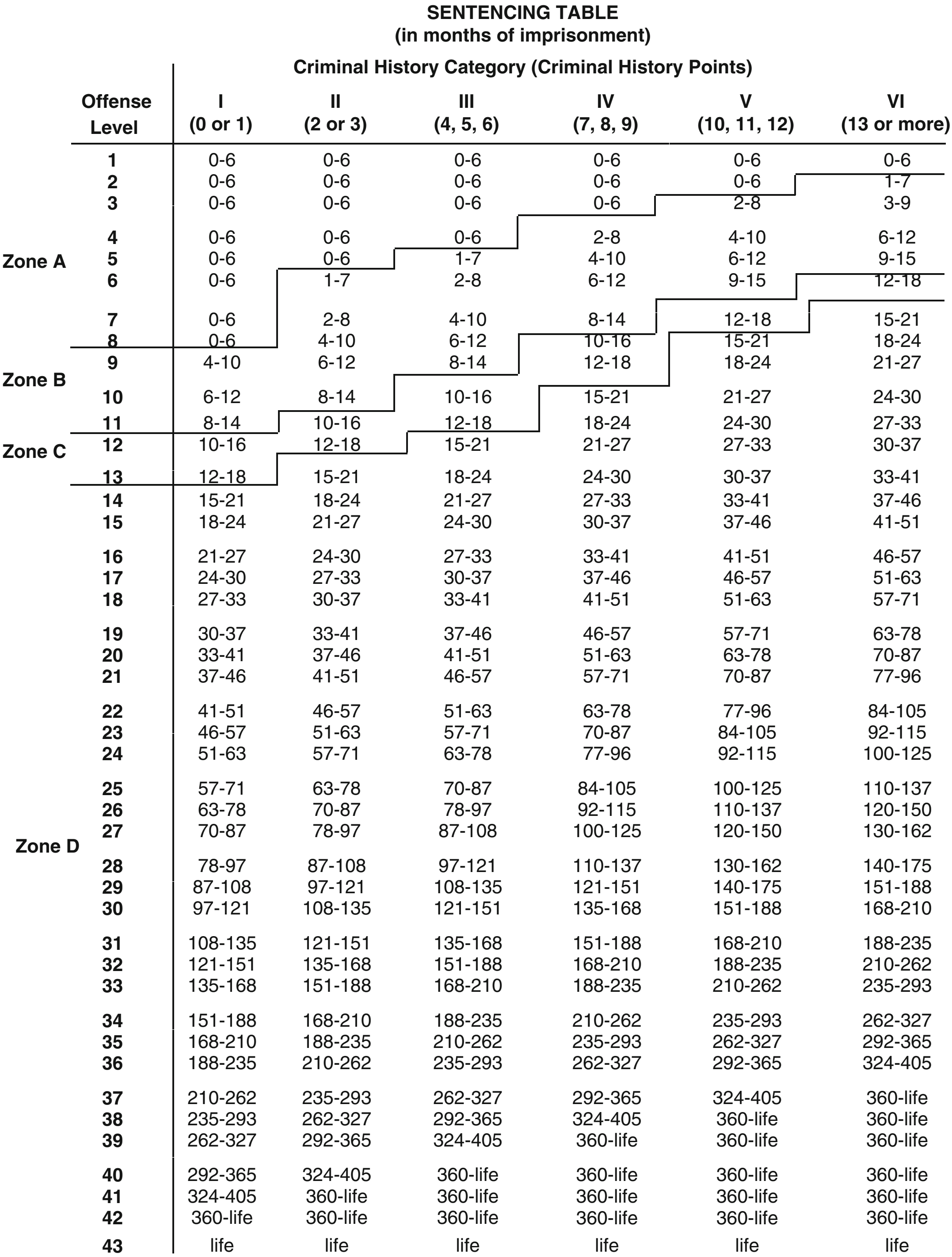

SENTENCING TABLE
(in months of imprisonment)

	Offense Level	Criminal History Category (Criminal History Points)					
		I (0 or 1)	II (2 or 3)	III (4, 5, 6)	IV (7, 8, 9)	V (10, 11, 12)	VI (13 or more)
	1	0-6	0-6	0-6	0-6	0-6	0-6
	2	0-6	0-6	0-6	0-6	0-6	1-7
	3	0-6	0-6	0-6	0-6	2-8	3-9
	4	0-6	0-6	0-6	2-8	4-10	6-12
Zone A	5	0-6	0-6	1-7	4-10	6-12	9-15
	6	0-6	1-7	2-8	6-12	9-15	12-18
	7	0-6	2-8	4-10	8-14	12-18	15-21
	8	0-6	4-10	6-12	10-16	15-21	18-24
Zone B	9	4-10	6-12	8-14	12-18	18-24	21-27
	10	6-12	8-14	10-16	15-21	21-27	24-30
	11	8-14	10-16	12-18	18-24	24-30	27-33
Zone C	12	10-16	12-18	15-21	21-27	27-33	30-37
	13	12-18	15-21	18-24	24-30	30-37	33-41
	14	15-21	18-24	21-27	27-33	33-41	37-46
	15	18-24	21-27	24-30	30-37	37-46	41-51
	16	21-27	24-30	27-33	33-41	41-51	46-57
	17	24-30	27-33	30-37	37-46	46-57	51-63
	18	27-33	30-37	33-41	41-51	51-63	57-71
	19	30-37	33-41	37-46	46-57	57-71	63-78
	20	33-41	37-46	41-51	51-63	63-78	70-87
	21	37-46	41-51	46-57	57-71	70-87	77-96
	22	41-51	46-57	51-63	63-78	77-96	84-105
	23	46-57	51-63	57-71	70-87	84-105	92-115
	24	51-63	57-71	63-78	77-96	92-115	100-125
	25	57-71	63-78	70-87	84-105	100-125	110-137
	26	63-78	70-87	78-97	92-115	110-137	120-150
Zone D	27	70-87	78-97	87-108	100-125	120-150	130-162
	28	78-97	87-108	97-121	110-137	130-162	140-175
	29	87-108	97-121	108-135	121-151	140-175	151-188
	30	97-121	108-135	121-151	135-168	151-188	168-210
	31	108-135	121-151	135-168	151-188	168-210	188-235
	32	121-151	135-168	151-188	168-210	188-235	210-262
	33	135-168	151-188	168-210	188-235	210-262	235-293
	34	151-188	168-210	188-235	210-262	235-293	262-327
	35	168-210	188-235	210-262	235-293	262-327	292-365
	36	188-235	210-262	235-293	262-327	292-365	324-405
	37	210-262	235-293	262-327	292-365	324-405	360-life
	38	235-293	262-327	292-365	324-405	360-life	360-life
	39	262-327	292-365	324-405	360-life	360-life	360-life
	40	292-365	324-405	360-life	360-life	360-life	360-life
	41	324-405	360-life	360-life	360-life	360-life	360-life
	42	360-life	360-life	360-life	360-life	360-life	360-life
	43	life	life	life	life	life	life

Sentencing Guidelines in the United States, Fig. 1 Sentencing table, United States sentencing guidelines any "adjustments" apply. Adjustments are not offense-specific and include such matters as the defendant's role in the offense (as a major or minor participant) and whether the defendant obstructed justice. Adjustments often increase a defendant's offense level, but they can reduce it as well, as in the case of a minor participant. The judge then must consider whether there are multiple counts of conviction. As noted above, the federal system employs a complicated formula to give an incremental increase for multiple offenses. Next, the judge evaluates whether the defendant has "accepted responsibility" for his offense. This often correlates with a guilty plea. If the defendant has accepted responsibility, his offense level will be reduced by either two or three offense levels. The resulting number is the defendant's offense level, which is represented on the vertical axis in the Sentencing Table (*see* Fig. 1). The horizontal axis reflects the defendant's criminal history. The judge assigns criminal history points for previous convictions of various types and whether the defendant was under judicial supervision at the time of the offense. These points translate into a "Criminal History Category." There are six Criminal History Categories. Where the offense level and the Criminal History Category intersect is the defendant's guideline range.

The US Sentencing Commission recaps the process this way. "The final offense level is determined by taking the base offense level and then adding or subtracting from it any specific offense characteristics and adjustments that apply. The point at which the final offense level and the criminal history category intersect on the Commission's sentencing table determines the defendant's sentencing guideline range" (US Sentencing Commission, p. 3). An offense level of 19 and a Criminal History Category of I, which is where a first-offender would be classified, yields a guideline range of 30–37 months. If the defendant was in Criminal History Category VI, which is the highest possibility, the guideline range would be 63–78 months.

Despite having calculated the guideline range, the judge is not ready to impose a sentence. Next, the judge must evaluate whether this is an unusual case warranting a different sentence under the guidelines. "[I]f an atypical aggravating or mitigating circumstance exists, the court may 'depart' from the guideline range. That is, the judge may sentence the offender above or below the range" (US Sentencing Commission, p. 3). For years, the availability of departures under the federal guidelines was a topic of frequent litigation and vigorous debate. Many critics argued that the appellate courts and the federal sentencing commission took an inappropriately crabbed view of these kinds of departures and that the fairness of the sentences often suffered. Yet judges did depart at times and could do so either up or down the offense level axis or left or right along the Criminal History Category axis. Since the Supreme Court of the United States made the federal guidelines "effectively advisory" (*see* below), judges must also consider whether the guidelines are appropriate in this case in light of the authorizing statute which invokes the traditional purposes of punishment as well as a parsimony provision. The judge may impose a completely different sentence if the judge believes that the guideline range does not meet the dictates of the statute, although the sentence is subject to appellate review for "unreasonableness."

The Sixth Amendment's Shifting Sands

Until approximately 2004, several guideline systems – especially the federal system – relied on facts that judges found at sentencing by a preponderance of the evidence in order to determine the presumptive sentencing range. "The top of the presumptive range was below the traditional statutory maximum for the offense of conviction. The actual sentence imposed might be higher or lower than the presumptive range, in part because of judicially found aggravating or mitigating facts" (Chanenson 2005, p. 378). In a series of decisions (including *Blakely v. Washington* and *United States v. Booker*) interpreting the Sixth Amendment to the United States Constitution, which guarantees the right to a trial by jury, the Supreme Court of the United States functionally invalidated crucial aspects of these presumptive guideline systems. The Court held these schemes unconstitutional because judges were allowed "to impose sentences higher than the presumptive guideline range based on facts found by the judge, using the preponderance of the evidence standard, instead of by the jury, using the beyond a reasonable doubt standard" (Chanenson 2005, p. 378). The key, according to the Court, was that a judge could not impose a sentence that exceeds the statutory maximum which, for these purposes, it defined as the "maximum sentence a judge may impose *solely on the basis of the facts reflected in the jury verdict or admitted by the defendant*" (*Blakely v. Washington*, 542 U.S. at 303). The Court emphasized that "the relevant 'statutory maximum' is not the maximum sentence a judge may impose after finding additional facts, but the maximum he may impose *without* any additional findings" (*Blakely v. Washington*, 542 U.S. at 303).

This could have been a mortal blow to federal sentencing guidelines, which was a heavily presumptive system and relied on significant judicial fact finding. However, the Supreme Court provided an escape hatch. These presumptive systems had two primary options. First, juries could be asked to find all of the facts beyond a reasonable doubt that judges had been finding by a preponderance of the evidence. This would then authorize the judge to impose a sentence in the same guideline range as in the past. While some jurisdictions went in that direction, many did not in part because of the factual and logistical challenges of asking juries to decide such sentencing facts as a defendant's role in the offense. This expanded role for the jury would have been particularly difficult for the federal sentencing system to adopt because it was both highly presumptive and required judges to make numerous factual determinations at sentencing, which now juries would have to make. The second option involved recasting the formerly presumptive guidelines as advisory. If the guidelines are just advisory, the judge has the ability to ignore the guidelines' advice and the power to impose a sentence up to and including the traditional statutory maximum, which is almost always higher than what the guidelines recommend. This is the path the federal system followed. However, neither the US Sentencing Commission nor the Congress made that change. Rather, it was the Supreme Court itself that recast the federal guidelines in this way. The Court construed the federal guidelines as "effectively advisory" in order to avoid the need to strike down the system as unconstitutional (*United States v. Booker*, 543 U.S. 220 (2005)). In doing so, it also broadened many people's conception of what it means for a system to be "advisory," as federal judges must still consider the guidelines and the sentences they impose are still subject to appellate review, albeit under a less stringent standard of review.

Broader Questions and Emerging Practices

Intermediate Punishments and Misdemeanors

Intermediate punishments have been the subject of discussion – and practice – for many years. Norval Morris and Michael Tonry have argued:

> We are both too lenient and too severe; too lenient with many on probation who should be subject to tighter controls in the community, and too severe with many in prison and jail who would present no serious threat to community safety if they were under control in the community. (Morris and Tonry 1990, p. 3)

It is thus unfortunate and somewhat surprising given the size and scope of community corrections programs that guidelines relating to intermediate punishment are not common. Some jurisdictions – like Pennsylvania and North Carolina – have addressed this issue to various degrees, but it remains an underdeveloped area (Frase 2000, pp. 439. –442; Frase 1999, pp. 77–78).

It is also interesting to note that relatively few guideline systems apply to both misdemeanors and felonies (Frase 2000, p. 429).

Guideline "Effectiveness"

The "effectiveness" of sentencing guidelines is often a topic of intense discussion and disagreement. Part of the disagreement flows from differing definitions of success and varying methodological views on examining the relevant issues. Overall, many believe that sentencing guidelines can be and have been effective along particular metrics, including disparity reduction, in certain circumstances (U.S. Sentencing Commission 2004, xiv–xvi, 135, 141; Frase 2000, p. 443). For example, Professor Michael Tonry has written that "[g]uidelines promulgated by commissions have altered sentencing patterns and practices, reduced sentencing disparities and gender and race effects, and shown that sentencing policies can be linked to correctional and other resources, thereby enhancing governmental accountability and protecting the public purse" (Tonry 1993, p. 713). To be sure, not every jurisdiction succeeds on every metric all the time, but if one accepts that an increase in transparency and a decrease in judicial disparity were key goals of sentencing guidelines, there are reasons to be encouraged.

Sentencing guidelines have also had some success in the realm of controlling costs. At a minimum, many sentencing commissions have become trusted sources of information for legislatures as they debate broad penal strategies. Some commissions, particularly Minnesota and North Carolina, have tried to respect financial restraints as they promulgate guidelines. More recently, Missouri started providing information about the financial consequences of particular sentencing decisions to judges in individual cases.

However, if one is focused more on controlling prosecutorial discretion (and the disparity that may flow from that discretion) sentencing guidelines have not been effective. As some predicted early on, there is a widely held view that prosecutorial power has increased after the introduction of sentencing guidelines. "Since guidelines limit the range of sentences available for a given offense, the power to drop or not drop charges is the power to select the sentence range available to the court (that is, what 'box' on the grid the case ends up in). Thus, any disparity in charging translates into disparity in sentencing" (Frase 1999, p. 77). It must be recognized that most sentencing guidelines do not speak to the prosecutorial role directly and were never intended to do so, despite the wishes of some.

Similarly, if one is focused on controlling prison populations, sentencing guidelines have not been particularly effective. The incarceration rate in the United States has risen precipitously in the last 30 years. Some jurisdictions have succeeded in curbing the growth of their prison population, but it is more than debatable whether the existence or nonexistence of guidelines is the dispositive feature. Sentencing guidelines are tools to implement broad policy in a manner that promotes systemic rationality and individual fairness. To that extent, there is reason to believe they can be – and often are – "effective." Sentencing guidelines are not inherently severe or lenient any more than fire is inherently a force for good (warmth and cooking) or evil (arson and destruction). "The experience in Minnesota and Washington, progressive states generally regarded as having the most successful guidelines systems, teach that while those systems were effective at restraining the growth of prison populations when that was the policy of those states, they were equally effective at implementing policy judgments that more punitive sentences were appropriate" (Boerner 1993, p. 176).

Evidence-Based Practices and Risk

Selected jurisdictions have started to consider how, if at all, they should incorporate evidence-based practices, including risk assessment, in their sentencing systems. While there are some echoes from the pre-guidelines era's interest in rehabilitation, the emphasis today is quite different. The focus now is on how reliable social science evidence can inform sentencing determinations – at the level of the guidelines and/or the individual sentencing judge – in a way that will reduce recidivism and thus improve public safety. Former Missouri Chief Justice Mike Wolff, who facilitated the inclusion of risk assessment information in Missouri's Sentencing Assessment Reports (its version of presentence investigation reports), has argued that, "We must acknowledge that the reason for sentencing is to punish, but if we choose the wrong punishments, we make the crime problem worse, punishing ourselves as well as those who offend" (Wolff 2008, p. 1395).

Virginia is the undisputed pioneer in integrating risk assessment into sentencing guidelines. Pursuant to a 1994 legislative directive, the Virginia Criminal Sentencing Commission explored whether an empirically based risk assessment tool could help judges divert "25 % of the lowest risk, incarceration-bound, drug and property offenders for placement in alternative (non-prison) sanctions" (Kern and Farrar-Owens 2004, p. 165). At the legislature's request, the Virginia commission later broadened the eligibility to additional low-risk offenders. If the risk instrument indicates that the qualifying offender is of sufficiently low risk, the Virginia guidelines, which are advisory and without appellate review, recommend the offender for an alternative sanction. The judge remains free to accept or reject that recommendation. A few years later, the Virginia legislature asked the commission to develop a risk assessment tool for sex offenders with the goal of identifying those with the highest risk of reoffending. In 2001, this instrument was integrated into the Virginia guidelines by increasing the upper end of the recommended range by varying amounts for higher risk offenders (Kern and Farrar-Owens 2004, p. 167). Again, the judge retains discretion and may, but need not, impose a sentence that takes advantage of the expanded sentencing range while remaining in compliance with the guidelines.

Pennsylvania is also weaving risk assessment into its sentencing guidelines. Under Pennsylvania's guidelines, the duration and intensity the recommended non-incarcerative sentence for certain drug-involved offenders has long reflected the treating of professional's judgment concerning treatment needs. Pursuant to a 2010 legislative directive, the Pennsylvania commission is now working to integrate an actuarial risk assessment instrument into the guidelines. While Pennsylvania may not follow the precise path blazed by Virginia, it is also trying to use risk assessment "to assist in the transparent – and thus accountable – decision-making by both the Commission at the policy level and judges... at the individual level" (Hyatt et al. 2011, p. 748).

There are many reasons to be cautious about risk assessment, and many policy makers and judges are quite wary, if not suspicious, of this approach. By definition, risk assessment deals with predictions and not certainties. No actuarial instrument will ever be 100 % accurate. Yet, policy makers and judges have long engaged in predictions of risk based on their clinical judgment – or gut instinct – which itself is not only imperfect, but research indicates is often less accurate than actuarial instruments (Gottfredson and Moriarty 2006). Sentencing remains deeply normative and predictions about recidivism will never – and should never – be the only consideration. Actuarial risk assessment is simply a tool that guideline systems can use to help inform judicial discretion.

Not only does risk assessment raise crucial questions concerning how to distribute punishment, but it also can prompt reflection on the severity of punishments. If a less severe sentence does not diminish – and may even improve – public safety in certain situations, was society's initial punitive judgment sound? There are no immutable answers, but the questions themselves can be important.

Concluding Observations

Sentencing guidelines have helped to refashion the landscape of American criminal justice, and they are continuing to do so. Even jurisdictions that do not yet have guidelines may be attracted by their promise of a more rational, transparent, and just approach to punishment. Indeed, the American Law Institute seems likely to recommend that all American jurisdictions adopt guidelines. Sentencing guidelines, however, remain very much a work in progress. All guidelines systems have flaws and many of them are glaring, but those same flawed systems often provide a mechanism for experimentation and, one hopes, positive refinement (Frase 2000, p. 445). Professor Michael Tonry summed it up this way:

> Like all calls for just the right amount of anything, not too much and not too little, a proposal for sentencing standards that are constraining enough

> to assure that like cases are treated alike and flexible enough to assure that different cases are treated differently is a counsel of unattainable perfection. Nonetheless, that is probably what most people would want to see in a just system of sentencing... (Tonry 1996, pp. 185–186).

Sentencing guidelines offer a real opportunity to strive for that “unattainable perfection.”

Related Entries

- ▶ Penal Philosophy and Sentencing Theory
- ▶ Sentencing Commissions
- ▶ Sentencing Research

Recommended Reading and References

Allen F (1981) The decline of the rehabilitative ideal: penal policy and social purpose. Yale University Press, New Haven

Bergstrom M, Mistick J (2003) The Pennsylvania experience: the public release of judge-specific sentencing data. Fed Sentencing Rep 16:57–61

Berman D (2000) Balanced and purposeful departures: fixing a jurisprudence that undermines the federal sentencing guidelines. Notre Dame Law Rev 76:21

Boerner D (1993) Bringing law to sentencing. Fed Sentencing Rep 6:174–177

Breyer S (1988) The federal sentencing guidelines and the key compromises upon which they rest. Hofstra Law Rev 17:1–50

Breyer S (1999) Federal sentencing guidelines revisited. Fed Sentencing Rep 11:180–186

Chanenson S (2005) The next era of sentencing reform. Emory Law J 54:377–460

Frankel M (1973) Criminal sentences: law without order. Hill and Wang, New York

Frase R (1999) Sentencing guidelines in Minnesota, other states, and the Federal Courts: a twenty-year retrospective. Fed Sentencing Rep 12:69–82

Frase R (2000) Is guided discretion sufficient? Overview of state sentencing guidelines. St Louis Univ Law J 44:425–446

Frase R (2005) Punishment purposes. Stanf Law Rev 58:67–83

Freed D (1992) Federal sentencing in the wake of the guidelines: unacceptable limits on the discretion of sentencers. Yale Law J 101:1681–1754

Gottfredson S, Moriarty L (2006) Statistical risk assessment: old problems and new applications. Crime Delinq 52:178–200

Hofer P, Allenbaugh M (2003) The reason behind the rules: finding and using the philosophy of the federal sentencing guidelines. Am Crim Law Rev 40:19–85

Hyatt J, Chanenson S, Bergstrom M (2011) Reform in motion: the promise and perils of incorporating risk assessments and cost-benefit analysis into Pennsylvania sentencing. Duq Law Rev 49:707–749

Kern R, Farrar-Owens M (2004) Sentencing guidelines with integrated offender risk assessment. Fed Sentencing Rep 16:165–169

Kramer J, Ulmer J (2009) Sentencing guidelines: lessons from Pennsylvania. Lynne Rienner, Boulder

Miller M (2004) Sentencing reform “reform” through sentencing information systems. In: Tonry M (ed) The future of imprisonment. Oxford University Press, New York

Morris N (1974) The future of imprisonment. University of Chicago Press, Chicago

Morris N, Tonry M (1990) Between prison and probation: intermediate punishments in a rational sentencing system. Oxford University Press, New York

Reitz K (2005) The enforceability of sentencing guidelines. Stanf Law Rev 58:155–173

Singer R (1978) In favor of “presumptive sentences” set by a sentencing commission. Crime Delinq 24:401–427

Stith K, Cabranes J (1998) Fear of judging: sentencing guidelines in the federal courts. University of Chicago Press, Chicago

Stith K, Koh H (1993) The politics of sentencing reform: the legislative history of the federal sentencing guidelines. Wake Forest Law Rev 28:223–290

Tonry M (1993) The success of judge Frankel’s sentencing commission. Univ Colo Law Rev 64:713–722

Tonry M (1996) Sentencing matters. Oxford University Press, New York

United States Sentencing Commission (1987) U.S. sentencing guidelines manual. http://www.ussc.gov/guidelines/1987_guidelines/Manual_PDF/1987_Guidelines_Manual_Full.pdf

United States Sentencing Commission (2004) Fifteen years of guidelines sentencing: an assessment of how well the federal criminal justice system is achievingthe goals of sentencing reform. http://www.ussc.gov/Research/Research_Projects/Miscellaneous/15_Year_Study/15_year_study_full.pdf

United States Sentencing Commission. An overview of the federal sentencing guidelines. http://www.ussc.gov/About_the_Commission/Overview_of_the_USSC/Overview_Federal_Sentencing_Guidelines.pdf

Wolff M (2008) Evidence-based judicial discretion: promoting public safety through state sentencing reform. N Y Univ Law Rev 83:1389–1419

Sentencing Research

Jeffery T. Ulmer
The Pennsylvania State University, University Park, PA, USA

Synonyms

Penalties; Punishment; Sanctioning

Overview

Research on sentencing has been an active field of inquiry for several decades. The 1980s, 1990s, and 2000s were particularly noteworthy. These decades saw advances in the quality of data, the sophistication of research methods, and improvements in theory. This entry provides a conceptual survey of research on noncapital sentencing outcomes since 2000. The entry first looks backward at the research agenda posed by reviews in the early 1980s and in 2000. Theoretical developments in the study of sentencing in the 1990s and 2000s are then discussed. An overview is next provided of recent sentencing research focused on the following: (1) court organizational and social contexts, (2) individual courtroom workgroup members, (3) disparity conditional on intersecting defendant characteristics, (4) victim characteristics, and (5) earlier case processing events and decisions.

The central focus of this entry is *noncapital sentencing decisions* as focal dependent variables. However, it is recognized that sentencing decisions are conditioned by and interdependent with a chain of earlier and later criminal justice processing decisions that are embedded in organizational and broader contexts. This entry also primarily focuses on research since 2000, discussed in the context of earlier reviews of sentencing research and theory in the 1980s and 1990s.

Sentencing Research in the 1990s: Responding to Issues and Opportunities

By far, the biggest development in the 1980s and 1990s was the availability of data collected by sentencing commissions in sentencing guideline states and the federal system. For the first time, researchers had at their disposal high-quality post-conviction, jurisdiction-wide data for very large numbers of cases. These data were disseminated by sentencing commissions with legal mandates to collect and make such data available. These data enabled progress on the following issues, raised by the important 1983 National Academy of Science assessment of sentencing research by Blumstein et al. (1983) and by Hagan and Bumiller in that report:

1. *The need for better measurement of legally relevant variables*: Sentencing commission data enabled substantially improved measures of offense severity and criminal history, which had often been measured crudely in earlier research.
2. *The need to study different sentencing outcome variables*: The Blumstein et al. (1983, p. 273) report called for "…more adequate treatments of the complexity of the dependent variable, sentence outcome."
3. *The need for larger samples from more varied jurisdictions* was particularly identified by Hagan and Bumiller (1983). Guideline data from states and the federal system provided very large samples drawn from large numbers of contextually diverse counties (for states) or district courts (with federal data).
4. *The need to investigate the effects of social contexts on sentencing and disparity* was noted by Hagan and Bumiller (1983) among others. As noted above, sentencing data from a relatively wide range of local jurisdictions enabled research on local variation in sentencing severity and disparity.

Several chapters of the Blumstein et al. (1983) report and others called for consideration of sentencing in the contexts of earlier criminal justice decisions, particularly pretrial detention, charging, and conviction. It has long been recognized that sentencing is conditioned by selection processes in earlier justice system decisions. For all their benefits noted above, guideline sentencing data are limited to convicted offenders and do not include preconviction information such as the type and number of original charges (and rarely include pretrial detention). Datasets that include preconviction data for large numbers of cases and multiple local jurisdictions have traditionally been difficult to obtain. The years since the Blumstein et al. (1983) report have seen comparatively few studies that examine sentencing in the context of earlier decisions.

Relatedly, the Blumstein et al. (1983) report called for sentencing research to examine the effects of defendant socioeconomic status variables on sentencing outcomes and to disentangle the potential influence of race and ethnicity versus socioeconomic factors in sentencing. In addition, the Blumstein et al. (1983) report and others called for studies that examine sentencing variation between criminal justice actors, especially judges and prosecutors, and the effects of decision maker characteristics in explaining variation in sentencing. But, only a handful of studies since then have examined sentencing variation between judges, and the effect of judge characteristics, and even fewer studies have examined between-prosecutor variation. Also, defendant socioeconomic data and/or data on individual court actors has generally been hard to come by in the most widely used guideline datasets. Thus, studies that disentangle socioeconomic factors from other defendant characteristics (especially race or ethnicity) have not been as common.

Reviews of the 1990s

Two comprehensive reviews (Spohn 2000; Zatz 2000) summed up research on sentencing up to 2000, primarily focusing on studies of disparity based on defendant social characteristics. Both reviews noted that research had improved in quality since the 1983 report. Both reports also highlighted research that went beyond simply assessing whether "race/ethnicity mattered" or "gender mattered" to investigating *when* and *how* such social statuses matter in sentencing – that is, investigating how the influence of race, ethnicity, and gender mutually conditioned one another and were also conditioned by other factors. Spohn's (2000) review in particular provided a quantitative summary of racial/ethnic disparity studies with an eye to conditional influences.

Regarding the incarceration decision, Spohn (2000) found that 55 % of 45 effects in 31 state court studies she reviewed, and three of seven effects in eight federal court studies, showed significant black disadvantages. More importantly, Spohn (2000, see p. 462) found that eleven studies had looked at whether race and/or ethnic effects were conditioned by other social statuses or characteristics, especially gender and/or age. All eleven found evidence of such effects. In particular, four studies found sentencing disadvantages for young black and/or Hispanic *males*, four found sentencing disadvantages for unemployed young black or Hispanic males, one found disadvantages for poor minorities, and one found disadvantages for low-education minorities.

In addition, several studies in the 1990s showed that race or ethnicity effects on sentencing were conditional on criminal history, offense type or severity, case processing factors, or victim characteristics. Eight 1990s studies found that blacks or Hispanics with more serious criminal histories were sentenced more severely. Ten studies found that race/ethnic differences were conditioned by type or severity of crimes. Three found that racial disadvantages were greatest in less serious crimes, and six found that racial disadvantages were pronounced in drug offenses. In addition, six studies found that race/ethnicity effects were conditioned by case processing factors. Two studies found that blacks were sentenced more severely if they were detained prior to trial, one found that blacks were sentenced more severely if represented by a public defender rather than a private attorney, and three found that blacks were sentenced more severely if they were convicted by plea rather than trial. Finally, one study found that blacks who victimized whites received more severe sentences than other victim-offender combinations.

These reviews also spelled out directions for further research that they hoped would be undertaken in the ensuing decade. Several of these directions called for further progress on earlier recommendations from the Blumstein et al. (1983) report. Among other directions, these reviews called for (1) investigations of how the effects of race and ethnicity on sentencing vary by social context and over time and how the effects of race and ethnicity might be conditioned by other factors, (2) more investigations of the sentencing of Latinos, (3) research that utilized better measures of defendant socioeconomic status factors and disentangling the effects of socioeconomic status from race and ethnicity, (4) research on how court social and organizational contexts shaped sentencing severity and disparity, (5) investigations of sentencing and disparity in the context of earlier criminal justice (particularly prosecutorial) decisions, and (6) qualitative research that illuminated sentencing processes.

Theoretical Perspectives on Sentencing

Hagan and Bumiller (1983) as well as other contributors to the Blumstein et al. (1983) report noted that at that time there was little theoretical development in the sentencing literature. Spohn (2000, p. 458) also noted that most pre-1990s sentencing studies were based on "an overly simplistic version of conflict theory." However, the 1990s and 2000s saw the emergence of new theoretical frameworks and the refinement of older theoretical propositions from conflict and labeling theories. These theories seem to be more complementary than competing, and some of them have mutually influenced each other. Below are brief sketches of these theoretical frameworks (readers seeking full treatments should consult the original works cited below).

Individual Cases and Actors

Celesta Albonetti articulated an *uncertainty avoidance* and *causal attribution* (1991) perspective on sentencing and court decision making. Albonetti applied insights from organizational theory to argue that sentencing suffers from operating in a context of bounded rationality in that court actors make highly consequential decisions with insufficient information, which produces uncertainty. Albonetti (1991) particularly stressed uncertainty and insufficient information regarding the recidivism risk and rehabilitative potential of offenders. Albonetti drew from attribution theory in social psychology to argue that, as a means of reducing uncertainty, decision makers fall back on attributions about reoffending risk and/or rehabilitation potential that can be linked to race and gender and other social status stereotypes. This can then result in extralegal sentencing disparity connected to these statuses.

Similarly, Farrell and Holmes (1991) presented an *interpretive theory of legal decision making*, in which they emphasized the situational role of stereotypes linked to defendant social statuses in case processing. Farrell and Holmes (1991) generated ten propositions about the conditional, situation-specific role of status-linked stereotypes for routine and nonroutine cases and defendants. A related idea is the *liberation hypothesis* which has sometimes been applied to sentencing discretion (Spohn and Cederblom 1991). The liberation hypothesis implies that as the seriousness and/or visibility of the offense or case increases, sanctioning discretion is tightened and legally relevant variables are decisive, leaving little room for extralegal influences. By contrast, in less serious/visible cases, opportunities for discretion are greater and extralegal variables can influence outcomes more than in serious cases.

Because the sentencing guidelines movement of the 1980s transformed the sentencing landscape, the 1990s saw important treatments of sentencing discretion and guidelines, which involves the management of several dilemmas: between flexible discretion and rule-bound control, between uniformity and individualization, and between centralization and decentralized localism. Joachim Savelsberg (1992) helpfully conceptualized sentencing and sentencing guidelines as an attempt to impose a regime of what founding early sociologist Max Weber called formal rationality onto a traditionally "substantively rational" process. *Substantive rationality* refers to an individualized decision process governed by criteria that are in service of ideological goals external to the law. The flexibility inherent in substantive rationality, however, permits the possibility of bias, discrimination, and unwarranted disparity.

The *focal concerns perspective* emphasizes particular kinds of substantively rational criteria at work in sentencing decisions. A recent review and statement of the focal concerns perspective in proposition form can be found in Kramer and Ulmer (2009). The focal concerns perspective argues that court actors' subjective definitions of offenders and offenses in relation to three focal concerns of punishment – blameworthiness, protection of the community, and practical constraints – determine punishment decisions. The focal concerns perspective argues that both legal and extralegal considerations affect the interpretation and prioritization of focal concerns through local substantive rationality (Kramer and Ulmer 2009). The influence of race, for example, may be conditional on defendant's gender, age, social class, legally relevant factors, and especially local contexts (see below).

Rational choice approaches have also been applied to sentencing in the times since the 1983 report (Piehl and Bushway 2007). Proponents of a rational choice approach to sentencing research would call for researchers to identify the formal and informal incentive structures of different court actors involved, make theoretical predictions about what sentencing outcomes would look like if actors acted rationally in pursuit of those incentives, and then compare these predictions to real-world data. Actors' incentives, of course, are strongly influenced by the organizational environments of prosecutors' offices, bench, and defense bar.

Court Organization and Multilayered Environments

It has long been recognized that sentencing practices varied between jurisdictions, even within states or the federal court system. Myers and Talarico (1987), Eisenstein et al. (1988), Ulmer (1997), and others drew attention to the localization of sentencing. The *court community perspective* views courts as communities based on participants' shared workplace; interdependent working relations between key sponsoring agencies such as the prosecutor's office, judges' bench, and defense bar; and the court's relation to its larger sociopolitical environment (Eisenstein et al. 1988; Ulmer 1997). These court communities are said to foster their own locally varying sentencing norms which influence sentencing as least as much as formal policies and legal structures (Ulmer and Kramer 1996).

Interest in social contexts that foster negative racial/ethnic-based stereotypes and perceptions of threat has been fostered by *racial threat theory*. This theory argues that as minority racial groups grow in size relative to whites, they are likely to develop greater power, economic resources, and political influence in the community and are better able to compete with whites for power. In the context of sentencing, racial group threat theory implies that when perceptions of minority group threat are more pronounced and when courtroom actors perceive particular racial/ethnic groups as more dangerous or morally disrespectable, such minorities may receive harsher sentences.

Finally, *organizational efficiency/maintenance models* of sentencing have long been recognized (Dixon 1995; Engen and Steen 2000). In fact, the focal concerns model would also recognize organizational efficiency as a potentially important practical constraint faced by court actors (Kramer and Ulmer 2009). These organizational efficiency model views emphasize leniency for those who plead guilty and avoid time- and resource-intensive trials as an effort by courts to keep cases moving smoothly. Relevant to rational choice theory, an organizational efficiency proposition would suggest that rewarding guilty pleas and punishing trials is an organizationally rational response to the need to move cases efficiently. Furthermore, this differential punishment would be conditioned by caseload pressure – the greater the caseload pressure, the more courts would rely on such costs and incentives.

Research Since 2000

These theoretical developments have been instrumental in five broad and interrelated areas of empirical inquiry discussed below.

Social Contexts and Sentencing

The 2000s saw a burst of studies on contextual effects and variation and how the influence of individual-level factors was conditioned by court or social context factors. Overall, the recent literature on contextual variation in sentencing shows that local variation permeates many aspects of sentencing, both under sentencing guideline jurisdictions and non-guideline jurisdictions. Studies typically find that most sentencing outcome variation exists at the individual level and is most strongly predicted by individual-level factors. However, not only does sentencing severity (and related outcomes such as guideline departures) vary between local courts and their contexts but so too sometimes do the effects of other important legally relevant and extralegal sentencing predictors.

Racial/Ethnic Population Composition

Some of this research has investigated the role of local racial/ethnic composition in conditioning racial and ethnic sentencing disparity, drawing from both racial group threat theory and the focal concerns perspective. Many multilevel sentencing studies find that the effects of race and ethnicity in sentencing decisions do indeed vary significantly across courts. However, results have been decidedly mixed regarding racial threat theory's ability to explain this variation. That is, race/ethnic effects on sentencing tend to vary across contexts, but not always in ways predicted by racial threat theory. Some studies have found that the percentage of blacks in local populations has been found to increase racial/ethnic disparities in sentencing. Other studies reveal either no support for racial threat or evidence contrary to racial threat hypotheses (see Ulmer 2012).

Local Organizational Constraints

Other research has examined how local practical constraints such as caseloads or local criminal justice resources (i.e., jail space) affect sentencing, research relevant for the organizational efficiency hypothesis, as well as the focal concerns and court communities' perspectives. Pennsylvania court caseloads were negatively related to sentencing severity (see in Kramer and Ulmer 2009). Local jail space in Pennsylvania counties affected the probability of incarceration and found that local jail space predicted the choice of county jail versus state prison (Kramer and Ulmer 2009).

Sociopolitical Influences on Sentencing Decisions

Though less common, studies have examined other sociopolitical influences on sentencing, including crime rates, political climate, neighborhood disadvantage, and local religion (see review by Ulmer 2012). Crime rates and broad political climate measures (such as percent Republican voters) have generally not been found to be strong predictors of sentencing patterns or to strongly and consistently condition individual-level predictors (an exception is Johnson 2006). Fearn (2005) and Ulmer et al. (2008) found evidence that local religious contexts may affect sentencing patterns. Fearn (2005) found that prison sentences were more likely in jurisdictions with greater proportions of evangelical Christians. Ulmer et al. (2008) found that local religious homogeneity fostered greater use of incarceration in Pennsylvania. Unlike Fearn (2005), Ulmer et al. (2008) did not find that evangelical Christian prevalence influenced sentencing but rather that local religious homogeneity interacted with percent Republican voters, in that counties that were religiously homogeneous *and* strongly Republican were most likely to incarcerate offenders.

Court Community Racial/Ethnic Composition

Farrell et al. (2009) studied the effects of federal district court community racial composition on variation in the effects of race on sentencing. They found that district US Attorney's Office's black representation was associated with significantly smaller racial disparities in incarceration, and interestingly, greater black Federal Probation Office representation was associated with greater black/white sentencing outcome disparity. In addition, studies have found that greater black representation among county attorneys attenuated local black/white sentencing disparity (see in Ulmer 2012).

Court Personnel

Earlier research had investigated the race and gender of judges on sentencing, and this was reviewed in Spohn (2000). Johnson (2006) examined court contextual and inter-judge variation in sentences in Pennsylvania. He found that black and Hispanic judges sentenced all offenders, and particularly minority offenders, more leniently than white judges. Furthermore, male judges sentenced female offenders more leniently. Studies have also examined inter-judge variation in sentencing and also found wide variation between judges in the effects of race, gender, or financial means on their sentences. It has also been found that judges considered legal and extralegal factors quite differently in their sentencing decisions (see in Ulmer 2012).

Conditional Disparity

In addition, a key theme of Spohn's (2000) and Zatz's (2000) summaries of the 1990s literature was that the influence of social status factors such as race, ethnicity, and gender was conditional and mutually contingent. Dozens of studies in the 2000s have continued to confirm this insight. Many of these studies are noted above and below in the discussions of research on court contexts, courtroom actors, victim characteristics, and earlier case processing decisions. In sum, there appears to be substantial evidence that the effects of such extralegal social statuses are conditioned by court contextual factors and provocative evidence that disparity varies across individual court community actors. Furthermore, most studies that examine the issue find young black, and to a lesser extent Hispanic, male defendants to be sentenced more severely (see reviews by Mitchell 2005; Ulmer 2012). Also, the effects of defendant social statuses may interact with case processing, offense characteristics, and criminal history.

Victim Characteristics

It should be noted that existing studies that utilize data from sentencing guideline jurisdictions actually implicitly consider victim harm, financial loss, and often victim age and vulnerability. These factors are commonly included in creating guideline offense severity rankings. Thus, the strong effects of offense severity typically found in such studies in part incorporate victim impact and vulnerability. Such research, however, does not tell us about the effects of other specific victim attributes. Curry (2010) found that Texas offenders who victimized females received longer sentences, and females who victimized males received sentences about 10 months shorter than males who victimized females. Curry's research also found that violent offenders who victimized white and Hispanic females received longer sentences, and Hispanic and black homicide offenders who victimized whites got longer sentences than other combinations. Other studies have uncovered similar findings (see in Ulmer 2012). They found no race of victim/offender combination effects, but found that males who killed females received the longest prison sentences and that females who killed males received shorter sentences than males who killed females. Overall, there is mixed evidence that victim characteristics may matter in noncapital violent crime sentencing.

Sentencing in the Context of Earlier Case Processing

A comparatively smaller set of studies has examined sentencing as related to earlier case processing events, such guilty pleas versus trial convictions, as well as charging decisions and pretrial release. This research provides some, albeit limited, insight into prosecutorial discretion as well as other joint courtroom workgroup outcomes, such as differentially punishing those convicted by trial relative to those who plead guilty.

Charging Decisions

Other researchers have investigated prosecutors' charging decisions and their impact on eventual sentencing. Studies have found that the number of indictment charges filed by federal prosecutors significantly increased federal sentence lengths, and decreased the likelihood of substantial assistance or other downward departures. Shermer and Johnson (2010) examined the likelihood of federal prosecutors reducing charges for defendants. They found that while about 12 % of federal cases in their sample involved charge reductions, race/ethnicity and gender (along with offense severity and criminal history) influenced the likelihood of those charge reductions in drug and violent offenses. These charge reductions, in turn, resulted in lower sentences for those who received them and charge reductions-mediated race/ethnic effects on sentences.

Pretrial Detention/Release

Though not as plentiful as studies of sentencing outcomes, there is a literature focusing on the consequences of these decisions for defendants' sentencing. Studies have found that defendants subject to pretrial detention were more likely to receive more severe sentences than defendants who had been released before case disposition. Other studies have Spohn examined federal offenders in three district courts and found that both legal and extralegal factors predicted pretrial status and that pretrial status significantly impacted sentencing. These found that race and gender had indirect effects on sentence severity through their effects on pretrial status.

Mandatory Minimums

Sentencing involving mandatory minimums has proved useful for understanding how earlier decisions shape and constrain sentencing choices. Bjerk (2005) found that prosecutors used their charge reduction discretion to circumvent three strikes mandatories for some defendants. He found that such circumvention of three strikes mandatory minimums was moderately less likely to occur for men, Hispanics, and to a lesser extent, blacks. Multilevel analyses of prosecutorial discretion in applying mandatory minimums among mandatory-eligible offenders sentenced for drug crimes or as "three strikes" offenders (see in Kramer and Ulmer 2009). They found that prosecutors' decisions to apply mandatory minimums were significantly affected by mode of conviction (negotiated guilty pleas greatly assisted defendants in avoiding mandatories), the type and characteristics of offenses and guideline sentence recommendations (the greater the difference between the mandatory and the applicable guideline sentence, the less likely the mandatory was applied), prior record, and gender. In addition, Hispanic males were substantially more likely to receive mandatory minimums.

Federal Substantial Assistance Departure Motions

Some studies have examined prosecutorial discretion in decisions to file motions for substantial assistance guideline departures in federal court crack and powder cocaine sentences in 2000. These found that the likelihood of substantial assistance motions increased with offense severity and criminal history and that having multiple concurrent charges reduced the likelihood of such motions for crack cocaine cases. In powder cocaine cases, black and Hispanic males had the lowest odds of receiving substantial assistance departures, while white females and males had the greatest likelihoods. The results were similar for crack cocaine cases, except that Hispanic females had the greatest odds of substantial assistance departures.

In addition, Johnson et al. (2008) examined interdistrict variations in the application of both substantial assistance and other downward departures among a variety of offense types in federal sentences from 1997 to 2000. Findings indicated considerable between-district variation in the probability of these prosecutor-initiated substantial assistance departures. This variation was explained, in part, by organizational court contexts such as caseload pressures and by environmental considerations such as the racial composition of the district.

Spohn and Fornango (2009) examined variation between federal prosecutors in three districts in their likelihood of moving for substantial assistance departures. They found substantial interprosecutor variation in the likelihood of substantial assistance departures, net of the influence of individual case and defendant characteristics. Specifically, about 24 % of the variation between prosecutors in the likelihood of substantial assistance departures was unaccounted for by case or defendant factors, and even more between-prosecutor variation was unexplained in nondrug cases.

Mode of Conviction

Finally, several studies have examined how mode of conviction affects sentencing; that is, sentencing differences between types of guilty pleas and types trials. This research exhibits some limitations in that studies typically use data on convicted offenders only, which presents issues of selection and potential missing-variable bias absent data on the likelihood of acquittal (see Bushway and Piehl 2007). Still, one overarching lesson from the recent studies below is that sentencing guidelines seem to provide a discretionary framework within which to differentially reward guilty pleas and punish trials.

Some studies have found that mode of conviction moderated the role of race/ethnicity in predicting guideline departures, with blacks and Hispanics experiencing different odds of receiving downward or upward departures, depending on their modes of conviction. Studies have looked at trial penalties for serious violent offenders and less serious offenders using hierarchical models with cases sentenced under Pennsylvania's guidelines (Kramer and Ulmer 2009). They found that defendants were substantially penalized if they were convicted by trial relative to those with negotiated or open guilty pleas. Furthermore, this jury trial penalty varied depending on the seriousness and type of offense (more severe offenses had lesser trial penalties), defendant criminal history (offenders with more substantial criminal histories actually experienced less of a trial penalty), race (blacks experienced greater trial penalties) and court contextual characteristics such as court community size (larger trial penalties in larger courts), local violent crime rates (the higher the crime rate, the greater the trial penalties), and the size of local black populations (greater trial penalties in counties with larger black populations). Studies have also found that federal trial penalties could not be fully explained by US Sentencing Guideline factors that were relevant to mode of conviction. They also found that higher district court caseload pressure was associated with greater trial penalties, while higher district trial rates were associated with lesser trial penalties (see in Ulmer 2012).

New Methodological Extensions

The past decade has seen the application of new statistical methods to sentencing research questions, and several useful quantitative modeling alternatives now confront sentencing researchers. Bushway et al. (2007) present a very useful discussion of the merits and demerits of two-part models (modeling incarceration and sentence length separately), tobit, and Heckman two-step corrections with ordinary least squares regression in addressing problems of censoring and/or selection surrounding the incarceration and sentence length decisions. Several researchers have also demonstrated the usefulness of using multinomial logistic regression to predict different types of incarceration, such as county jail versus state prison, either in an individual level (see review in Ulmer 2012). In addition, Britt (2009) proposes *quantile regression* as an interesting alternative for assessing variation in the effects of predictors of interest (legally relevant, extralegal, case processing, etc.) across the distribution of sentence length/severity. That is, the quantile regression approach allows the researcher to separate the sentence length distribution into quantiles and examine how the strength of predictors' effects varies across those quantiles. Finally, propensity score methods appear to be a promising alternative for examining sentencing disparity in a way that attempts to create balanced, comparable samples of offenders that differ only on a "treatment" or characteristic of interest (e.g., race or mode of conviction). This therefore allows the construction of useful counterfactuals, for example, "what would the sentence be if this case involved a white rather than a black defendant."

The Need for More International Research

Most sentencing research is limited to the contemporary North American – particularly the US – context. However, particularly useful overviews of the comparatively small amount of international literature on sentencing disparity that does exist can be found in Tonry and Frase (2001). Also, Johnson et al. (2010) studied 1,613 Dutch homicide offenses from 1993 to 2004 and found that homicide offenders that victimized youth under 18, elderly people, women, and Dutch (vs. foreign) victims were sentenced more severely than those victimizing other types of victims. Some other suggestive information on ethnic disparity comes from Canada and Australia (see reviews in Tonry and Frase 2001; Dawson 2006).

Comparatively little research exists on sentencing in non-Western contexts, particularly, in Asian countries, which are growing in global prominence. Lu and Kelly (2008) provide a useful summary of research on courts and sentencing in China. Two studies focus on sentencing in South Korea, a country which very recently adopted sentencing guidelines. These found that female drug offenders were sentenced more leniently than their male counterparts, though this gender difference disappeared among those with prior criminal records. Also, research has found that legal factors connected to the offense and past criminal behavior primarily determined the length of sentences for Korean marijuana and methamphetamine offenders but that males and older offenders received longer sentences. They also found that admitting guilt in court resulted in shorter sentences (see review in Ulmer 2012).

Finally, a great deal of research has been done on sentencing under sentencing guidelines, particularly, the US Sentencing Guidelines, and state guideline systems in Minnesota, Pennsylvania, Washington, and to a lesser extent Florida (see reviews by Spohn 2000; Zatz 2000; Kramer and Ulmer 2009; Ulmer 2012). More recently, research has examined the extent to which US federal sentencing and disparity changed in the wake of important Supreme Court decisions that rendered the federal guidelines advisory (see Ulmer et al. 2011). This research is mixed regarding whether racial disparity has increased under the advisory guidelines – the US Sentencing Commission's analyses argue that racial disparity increased, but independent researchers have disputed their findings, arguing that overall, the data do not support the notion that advisory guidelines have aggravated racial disparity (see Ulmer et al. 2011). Research has examined sentencing disparity, patterns, and trends within guideline systems, but relatively little research has compared differences in sentencing disparity between guideline systems (Engen 2009). Very little research has examined differences in sentencing between guideline and non-guideline jurisdictions, likely because of the lesser availability of comparable sentencing data from non-guideline systems and because of the methodological difficulties of comparing measures of key variables such as offense severity and criminal history across guideline and non-guideline contexts.

In sum, since 2000, research has made some progress on certain issues identified by the Blumstein et al. (1983) report and the Spohn (2000) and Zatz (2000) reviews. For example, in many sentencing studies published since 2000, researchers have moved beyond examining the traditional incarceration and length to look at more refined outcomes, such as distinctions between types of incarceration (e.g., probation, county jail, state prison). Also, studies have examined the imposition of mandatory minimums, which are tightly linked to prosecutorial discretion, adjudication waivers, first-time offender waivers, and special sex offender sanctions, and different types of sentencing guideline departures (see Ulmer 2012). Further, research in the 2000s generally heeded the call to include Hispanic defendants, and many, if not most, of the recent studies on US racial sentencing disparity include black, white, and Hispanic comparisons. Other directions, such as examining sentencing in the context of preconviction decisions and prosecutorial discretion, as well as research on sentencing in non-US and non-Western contexts, have seen comparatively less progress.

Related Entries

▶ Comparative Courts and Sentencing
▶ Drug Courts
▶ International Sentencing
▶ Mitigation and Aggravation at Sentencing

- ▶ New Penology
- ▶ Plea Bargaining
- ▶ Problem-Solving Courts
- ▶ Prosecutorial Discretion
- ▶ Sentencing as a Cultural Practice
- ▶ Sentencing Commissions
- ▶ Sentencing Guidelines in the United States
- ▶ Sex Offenders and Criminal Policy
- ▶ Structured Sentencing Outside the United States
- ▶ Theories of Punishment

Recommended Reading and References

Albonetti C (1991) An integration of theories to explain judicial discretion. Soc Probl 38:247–266

Bjerk D (2005) Making the crime fit the penalty: the role of prosecutorial discretion under mandatory minimum sentencing. J Law Econ 48:591–625

Blumstein A, Jacqueline C, Susan M, Michael T (eds) (1983) Research on sentencing: the search for reform. National Academy Press, Washington, DC

Britt C (2009) Modeling the distribution of sentence length decisions under a guideline system: an application of quantile regression models. J Quant Criminol 25(4):341–370

Bushway S, Brian J, Slocum LA (2007) Is the magic still there? The relevance of the Heckman two-step correction for selection bias in criminology. J Quant Criminol 23:151–178

Bushway SD, Piehl AM (2007) Social science research and the legal threat to presumptive sentencing guidelines. Criminology & Public Policy 6:461–482

Curry T (2010) The conditional effects of victim and offender ethnicity and victim gender on sentences for non-capital cases. Punishm Soc 12(4):438–462

Dawson M (2006) Intimacy and violence: exploring the role of victim-defendant relationship in criminal law. J Crim Law Criminol 96(4):1417–1449

Dixon J (1995) The organizational context of criminal sentencing. Am J Sociol 100:1157–1198

Eisenstein J, Flemming R, Nardulli P (1988) The contours of justice: communities and their courts. Little, Brown, Boston

Engen R (2009) Assessing determinate and presumptive sentencing: making research relevant. Criminol Public Policy 8:323–335

Engen R, Steen S (2000) The power to punish: discretion and sentencing reform in the war on drugs. Am J Sociol 105:1357–1395

Farrell R, Holmes M (1991) The social and cognitive structure of legal decision making. Sociol Q 32(4):529–542

Farrell A, Ward G, Rousseau D (2009) Race effects of representation among federal court workers: does black workforce representation reduce sentencing disparities? Ann Am Acad Pol Soc Sci 623:121–133

Fearn NE (2005) A multilevel analysis of community effects on criminal sentencing. Justice Quart 22:452–487

Hagan J, Bumiller K (1983) Making sense of sentencing: race and sentencing outcomes. In: Blumstein A, Cohen J, Martin S, Tonry M (eds) Research on sentencing: the search for reform. National Academy Press, Washington, DC

Johnson BD (2006) The multilevel context of criminal sentencing: integrating judge and county level influences in the study of courtroom decision making. Criminology 44:259–298

Johnson B, Ulmer J, Kramer J (2008) The social context of Guideline circumvention: the case of federal district courts. Criminology 46:711–783

Johnson BD, van Wingerden S, Nieuwbeerta P (2010) Sentencing homicide offenders in the Netherlands: offender, victim, and situational influences in criminal punishment. Criminology 48(4):981–1018

Kramer J, Ulmer J (2009) Sentencing guidelines: lessons from Pennsylvania. Lynne Rienner Publishers, Boulder

Lee M, Ulmer JT, Park MR (2011) Drug sentencing in South Korea: the influence of case processing and social status factors in an ethnically homogeneous context. J Contemp Crim Justice 27(3):378–397

Lu H, Kelly B (2008) Courts and sentencing research on contemporary China. Crime Law Soc Change 50:229–243

Mitchell O (2005) A meta-analysis of race and sentencing research: explaining the inconsistencies. J Quant Criminol 21(4):439–466

Myers M, Talarico S (1987) The social contexts of criminal sentencing. Springer, New York

Piehl AM, Bushway SD (2007) Measuring and explaining charge bargaining. J Quant Criminol 23(2):105–125

Savelsberg JJ (1992) Law that does not fit society: sentencing guidelines as a neoclassical reaction to the dilemmas of substantivized Law. Am J Sociol 97:1346–1381

Shermer LON, Johnson B (2010) Criminal prosecutions: examining prosecutorial discretion and charge reductions in U.S. Federal District Courts. Justice Quarterly 27(3):394–430

Spohn C (2000) Thirty years of sentencing reform: the quest for a racially neutral sentencing process. In: Policies, processes and decisions of the criminal justice system. Vol. 3, criminal justice 2000. U.S. Department of Justice, Washington, DC

Spohn C, Cederblom J (1991) Race and disparities in sentencing: a test of the liberation hypothesis. Justice Q 8:305–327

Spohn C, Fornango R (2009) US Attorneys and substantial assistance departures: testing for interprosecutor disparity. Criminology 47(3):813–847

Tonry M, Frase R (eds) (2001) Sentencing and sanctions in Western countries. Oxford University Press, New York

Ulmer JT (1997) Social worlds of sentencing: Court communities under sentencing guidelines. State University of New York Press, Albany

Ulmer JT (2012) Recent developments and new directions in sentencing research. Justice Q 29(1):1–40

Ulmer JT, Kramer JH (1996) Court communities under sentencing guidelines: Dilemmas of formal rationality and sentencing disparity. Criminology 34:383–408

Ulmer JT, Bader C, Gault M (2008) Do moral communities play a role in criminal sentencing? Evidence from Pennsylvania. Sociol Q 49(4):737–768

Ulmer JT, Light M, Kramer J (2011) Racial disparity in the wake of the *Booker/Fanfan* decision: an alternative analysis to the USSC's 2010 Report. Criminol Public Policy 10(4):1077–1118

Zatz M (2000) The convergence of race, ethnicity, gender, and class on court decision making: looking toward the 21st Century. In: Policies, processes and decisions of the criminal justice system. Vol. 3, criminal justice 2000. U.S. Department of Justice, Washington, DC

Sequential Lineups

▶ Eyewitness Research

Serial Victims

▶ Multiple Victims and Super Targets

Sex Offender Treatment

Jo Thakker
School of Psychology, University of Waikato,
Hamilton, New Zealand

Overview

This entry provides a brief summary of three key aspects of the treatment of sexual offenders. It begins with a description and critique of the two of the most influential treatment models, namely, the Risk Needs Responsivity (RNR) model and the Good Lives Model (GLM). It then provides a discussion of some of the relevant process issues, namely, whether treatment should be conducted in groups, and the problems of confrontation and coercion. The entry concludes with a description of the primary components of group treatment approaches and a discussion of treatment efficacy.

Treatment Models

In order to fully understand contemporary sex offender treatment, it is useful to first summarize the models which underlie treatment. Obviously, it is not possible to describe all of the relevant models herein; thus, two models which are considered to be especially influential will be discussed, namely, the Risk Needs Responsivity model (Andrews and Bonta 2003a) and the Good Lives Model (Ward and Stewart 2003).

The Risk Needs Responsivity Model

Andrews and Bonta's (2003a) Risk Needs Responsivity (RNR) approach has been important in guiding treatment across a variety of offender groups, including sexual offenders. Since its development in the late 1990s, it has played a significant role in the way that treatment has been conceptualized and delivered. As indicated by the name of the model, there are three main components. The risk aspect of the model refers to the type and degree of risk that an offender poses and suggests that the risk level should determine the intensity of treatment. Thus, low-risk offenders should receive less intensive treatment, and high-risk offenders should receive higher intensity treatment. As explained by Andrews and Bonta, intensity essentially refers to the duration of treatment.

The need aspect of the RNR model refers to the offender's specific rehabilitative requirements, which are variously termed criminogenic needs or dynamic risk factors. Essentially, these are the psychological characteristics and/or behaviors of the individual which require treatment. According to the model, treatment should target the specific characteristics that contribute to the individual's offending so that the likelihood of reoffending will be reduced. As outlined by Andrews and Bonta, some common examples of these sorts of characteristics include antisocial attitudes and beliefs, substance abuse, and relationship problems.

The third part of the RNR model is the responsivity principle which is concerned with the actual delivery of treatment. As stated by Andrews and Bonta, the responsivity principle acknowledges the significance of the therapeutic relationship and assumes that an offender's response to treatment will be dependent on this, as well as a range of other variables. Also, the model takes the view that in order for treatment to be effective, it needs to identify and manage barriers to treatment and build on the individual's unique strengths. Furthermore, the responsivity principle espouses that in order for treatment to be successful, it needs to be designed according to best practice guidelines; in other words, it is important that there is evidence that a particular therapeutic approach is effective in treating the offender's unique range of needs.

According to Ward et al. (2008, p. 180), the RNR model has "constituted a revolution in the way that criminal conduct is managed in Canada, Britain, Europe, Australia, and New Zealand." While the model is general, insofar as it has been applied to a range of offender types, it is nonetheless pertinent to this discussion as it has been widely used in the design of treatment of sexual offenders. One of the key strengths of the model is the ease with which it can be translated into treatment. It provides clear guidelines in terms of selecting appropriate treatment for offenders, and it assists in the identification of treatment targets. Also, it encourages the therapist to consider the offender's strengths and weaknesses and to design treatment accordingly. Furthermore, it emphasizes the link between theory, research, and practice and is thus consistent with the principles of the scientist-practitioner model.

However, some researchers have criticized certain aspects of the RNR model. For instance, it has been argued that the RNR model places too much emphasis on rehabilitative needs and not enough emphasis on other aspects of treatment. For example, Wilson and Yates (2009) argue that the RNR model tends to overlook the importance of therapist qualities such as empathy and respect due to its focus on the needs of the offender. However, as outlined above, in a recent explication of the RNR model, Andrews and Bonta highlighted the significance of the therapeutic relationship in treatment success. Ward argues (e.g., Ward and Stewart 2003) that the RNR model is typically implemented in an inflexible manner which fails to adequately deal with individual needs. However, this may be an unfair criticism insofar as this may reflect as inadequate implementation of the model rather than a failing of the model per se. In their critique of the RNR model, Ward and Stewart (2003) also state that the model takes an overly negative approach to therapy by focusing on risk avoidance. They suggest that the model assists offenders in understanding what to avoid, but does not assist them in establishing new pro-social behaviors.

The RNR model continues to provide the framework upon which many contemporary treatment programs for offenders are designed and administered, and this is true also for sex offender programs. Furthermore, there is now mounting empirical support for the model. According to Andrews and Bonta (2010, p. 39), "...programs that adhere to the Risk Needs Responsivity (RNR) model have been shown to reduce offender recidivism by up to 35 %."

The Good Lives Model

The Good Lives Model (GLM) by Ward and colleagues (Ward and Stewart 2003) is another model that is influential in the area of offender rehabilitation. Furthermore, the GLM has been widely discussed in relation to sexual offenders. As illustrated above, Ward and various colleagues have criticized the RNR model, and they have developed the GLM, at least in part, to respond to some of the weaknesses of the RNR model. The GLM takes a very different approach to offender rehabilitation; rather than examining the risk level of an offender, it begins by asking what purpose the offending is serving. Specifically, the question is: What gap or need is the offending filling? Ward and colleagues (e.g., Serran et al. 2007) suggest that in order to bring about long-term change in offenders' behavior, therapy should identify and respond to the underlying motivation for the behavior.

The GLM is founded on the proposition that all human activity revolves around the desire to fulfill basic needs or "goods." For instance, eating nutritious food and engaging in regular physical exercise might fulfill the basic good of maintaining physical well-being. Painting a picture or acting in a play may fulfill the basic good of creativity. When these sorts of "goods" are maintained pro-socially in an individual's life that it can be said that he or she has a "good" life. Ward also argues that the drive to attain such "goods" is inherent by virtue of one's humanity. For example, he states that "…both sexual and non-sexual offenders are naturally disposed to seek a range of primary human goods that if secured will result in greater self-fulfilment and sense of purpose" (Ward et al. 2009, p. 304).

In explicating the GLM, Ward draws a distinction between primary and secondary goods. While primary goods are essentially ends in themselves, secondary goods relate to the means of attaining primary goods. For example, as shown below, knowledge is viewed as a primary good; thus, secondary goods that might bring this about would include activities such as reading a book and engaging in educational activities. Drawing on theoretical work from a range of research areas, including anthropology, evolutionary psychology, and ethics, Ward and Stewart (2003) propose that there are at least ten primary human goods, namely, "Life (including healthy living and functioning), knowledge, excellence in play and work (including mastery experiences), excellence in agency (i.e., autonomy and self-directedness), inner peace (i.e., freedom from emotional turmoil and stress), friendship (including intimate, romantic and family relationships), community, spirituality (in the broad sense of finding meaning and purpose in life), happiness, and creativity" (p. 356).

Ward proposes that offenders, like all human beings, value these primary goods and seek to manifest them in their daily lives. However, Ward stresses that for most offenders, and indeed most non-offenders, the drive to attain these goods is not necessarily part of one's conscious awareness. Thus, the drive to attain these goods is often natural and ingrained and not typically part of a carefully thought-out plan. Ward suggests that while offenders attempt to realize the same goods as those who do not commit crimes, they go about this in a problematic way and he describes four key types of problems. These are the use of *inappropriate means* (i.e., goods are sought in ways that are inappropriate and counterproductive), a *lack of scope* (i.e., only some goods are sought), *conflict or lack of coherence* (i.e., the ways some goods are sought directly reduces the chances of others being secured), and *a lack of capacity* (i.e., individuals lack the skills, opportunities, and resources to achieve a certain good in specific ways).

With regard to sexual offenders, the use of inappropriate means would be a common example. While a male sexual offender may seek a sexual partner as part of his intention to attain the "good" of "friendship," he may seek an inappropriate sexual partner. For instance, he may engage in sexual activity with a minor rather than an adult. Furthermore, his sexual offending may be related to a lack of capacity; he may lack the social skills required to establish and maintain an appropriate sexual relationship with an adult.

As outlined by Ward, a GLM approach to offender rehabilitation requires a comprehensive assessment of the relationship between the offending and goods that the offending is being used to attain. Then, a Good Lives Plan is developed which includes setting goals for the future which will ultimately bring about the manifestation of goods in a pro-social manner. Thus, treatment involves the acknowledgement that all human beings have basic needs and the encouragement of offenders to strive to meet these in their lives but to do so in the manner that is consistent with societal norms and laws. According to the GLM, when these basic needs are met, the risk for reoffending will be greatly reduced because the offender will no longer use offending as a means of achieving his goals.

The GLM approach to offender rehabilitation is consistent with the work of Maruna (e.g., 2001) which stresses the importance of taking a positive approach in therapy. The GLM is also compatible with the strengths-based approach which has become increasingly popular within the field of mental health treatment (Wong 2006). However, as yet, there is insufficient empirical evidence that applying the GLM to offender treatment is advantageous. Also, the suggestion put forward by proponents of the GLM that the more traditional RNR model has a negative focus and is not strengths-based has been challenged. For example, it has been noted that the RNR framework does have positive and forward-looking components (New Zealand Department of Corrections 2009). Furthermore, Andrews and Bonta (2003b) take issue with Ward's proposition that a focus on criminogenic needs and risk factors is incompatible with a focus on more positive factors.

At this time, in light of current research and theory, it would seem that a combination of the RNR model and a "good lives" approach is the best way forward in the treatment of sexual offenders. While the GLM still requires empirical support, given its alignment with positive and strengths-based approaches. which have already proven their worth, it is arguably a useful model which is likely to prove efficacious.

Process Issues

Over recent years researchers have predominantly focused on the content of sex offender treatment programs rather than the process. However, with regard to the process of treatment, there are a number of important issues which require attention.

Group Versus Individual Approaches

Treatment of sexual offenders is typically conducted in groups with between eight and ten participants with either one or two facilitators (Marshall 2001). As stated by Marshall, group treatment is more economical and has been found to be more effective. Marshall points out that in group settings, clinicians are less likely to collude with offenders. Collusion is more likely with this offender group as due to society's strong aversion to sexual offending; it is especially difficult for these sorts of offenders to admit to their crimes. Sexual offenders are more likely to deny their offending or to minimize their degree of culpability, and this can pose a challenge in terms of establishing rapport in individual treatment. In contrast, in a group setting, participants are able to challenge each other which assists in maintaining the rapport between the facilitator and the offender.

As explained by Harkins and Beech (2008), effective group treatment is dependent on the presence of several qualities within the group. One that has been found to be particularly important is cohesiveness (Harkins and Beech 2008) which is essentially the positive feelings that group members have for each other and the extent to which they are able to work together. As stated by Harkins and Beech, the role of the therapist is crucial in establishing cohesiveness within the group. In particular, the therapist needs to earn the respect of the participants by displaying a range of positive characteristics, such as flexibility, warmth, and empathy. However, it is equally important that the therapist is firm and directive and is able to challenge the participant when necessary, although this should be done in a gentle and compassionate manner (Serran et al. 2003).

Many contemporary sex offender treatment programs utilize mixed groups, that is, they treat rapists and child sexual offenders together, and some researchers have questioned whether this is appropriate. For example, Harkins and Beech (2008) compared the "therapeutic climate" in mixed groups and in homogenous groups. They measured therapeutic climate with the Group Environment Scale (GES) which targets a range of phenomena, including expressiveness, cohesion, and leader support. Harkins and Beech found that there was no difference in the quality of the therapeutic climate between the group types. Furthermore, they found that both types of group appeared to have very positive therapeutic climates. Also, recidivism data examined in the study showed that the type of group that child sexual offenders participated in had no impact on the likelihood of reoffending (however, similar data was not available for rapists).

Some researchers have argued that even when treatment is delivered within a group setting, it should still be tailored to the unique needs of the individual. For example, Marx et al. (1999) suggest that facilitators should adapt their approach to suit the various personality traits and learning styles of group members. Thus, it is proposed that even within a group setting, therapists can respond to individual needs. It is important to note though that such an approach would be dependent on a thorough and accurate assessment of the group participants which would allow for the identification of individual factors that would then facilitate appropriate tailoring of the therapist's interaction with group members.

Many group-based sex offender treatment programs provide individual therapy sessions as an adjunct to group treatment. For example, the Te Piriti Sex Offender Treatment Program in New Zealand includes individual therapy sessions alongside group-based treatment. These sessions allow the facilitator to provide additional support to the offender and to discuss any problematic issues that may arise during group sessions. They also provide an opportunity for the offender to raise issues that he may not feel comfortable raising within the group. For example, often offenders feel apprehensive about the idea of disclosing the details of their offending to their fellow group members, and thus individual sessions provide an opportunity to discuss such concerns prior to disclosure and to provide appropriate guidance and support.

Confrontation

There is an ongoing debate about the merit of using outright confrontation in the treatment of sexual offenders. This sort of approach is questioned because it is incompatible with the sorts of qualities which are considered to be fundamental to effective therapy. However, it is sometimes considered to be necessary because sexual offenders often deny or minimize their offending. Marques et al. (2005) state: "…too many sex offender treatment providers appear to believe that it is necessary to be extremely confrontative when working with these clients" (p. 1098). However, as Marque and colleagues explain, such an approach is inconsistent with the principles of motivational interviewing (MI), principles which the research shows are very helpful in guiding offender treatment. Although one of the key tasks of MI is to "develop discrepancy," this should be done carefully and in conjunction with the other components, which include the expression of empathy and avoiding arguments.

Marshall and others (Marshall et al. 2003) examined the effect of therapist qualities on the efficacy of sex offender treatment and concluded that there are several therapist characteristics which are associated with treatment success. For instance, they found that emotional warmth, empathy, directiveness, and positive reinforcement were all associated with positive treatment outcomes. Arguably, these types of therapist qualities are not compatible with a confrontational approach; thus, Marshall's research raises further doubts about the usefulness of such an approach in sex offender treatment. Overall, research suggests that therapists working with sexual offenders should probably be supportively challenging. That is, they should challenge in a way that allows them to remain positive and encouraging. Supportive challenges differ from confrontational challenges in that they are typically expressed in a warm and understanding manner which displays a genuine interest in the individual and concern for his or her well-being.

Coercion

Another process issue, which is pertinent to the rehabilitation of sexual offenders, is the impact of coercion on treatment effectiveness. Although many sexual offenders are recommended by judges, parole boards, and psychologists to engage in offense-related treatment, many choose not to. A comprehensive study by Langevin (2006) found that approximately 50 % reported that they wished to undertake treatment and about 13 % successfully completed treatment. Thus, a high number of sexual offenders decline treatment and only a small proportion complete treatment. Furthermore, with regard to those who do engage in treatment, there are a variety of reasons that may underlie their decision. For example, while some offenders may believe that they need treatment and will benefit from it, others simply undertake treatment in order to gain early release from prison. Thus, the provision of consent in a prison setting may not reflect a genuine motivation to engage in therapy but may simply reflect other motivations. Also, the offender may feel coerced by others; he may believe he has no choice but to abide by the wishes of the judicial system.

As explained by Burdon and Gallagher (2002), the coercion that forms the backdrop to offender treatment is part of broader societal efforts to coerce the offender, including imprisonment, registration, and community supervision. Thus, coercion reflects society's desire to control the offender in order to prevent reoffending. But of course, in terms of treatment some degree of motivation and cooperation is necessary in order for treatment to progress, and this is recognized by most therapists who are involved in selecting treatment participants. Also, often there are constraints on program delivery which limit the number that can be run at any one time, and this in turn limits the number of places that are available. Thus, it is often necessary to choose participants that are most likely to benefit from the program, and these are often those who express a desire to undertake treatment. However, it is interesting to note that involvement in sex offender treatment tends to increase an offender's openness to the idea of engaging in treatment and can lead to success in treatment, even in individuals who are initially unenthusiastic (Burdon and Gallagher 2002).

Treatment Components

There are many different components that may be used in the treatment of sexual offenders. The components that are covered herein are those that are commonly used in group treatment approaches and which have described in detail in the research literature.

Deviant Sexual Arousal

It is not surprising that deviant sexual arousal is often targeted in treatment as it has been found to be one of the most significant risk factors for sexual reoffending. However, it has been reported that deviant arousal is now not routinely targeted in treatment as it was previously, due to the acknowledgement that problematic sexual preferences do not always play a significant role in an individual's offending (e.g., Marshall 2006). Also, the factors that lead to sexual offending are many and varied and differ across offender and offense type. For example, in rape offenses, anger and aggression may play a more significant role than problematic sexual preferences. Further, research suggests that sexual arousal to stimuli associated with rape has been found in men who have not committed rape; thus, it is not necessarily associated with deviant behavior. Thus, it may be the case that offense-related sexual preference is more likely present in child sexual offenders. Nonetheless, many sexual offender treatment programs, for those with adult and juvenile victims, continue to include a component that focuses on deviant sexual arousal (Marshall 2006).

As explained by Marshall et al. (2009), there are two general types of behavioral approaches to treating deviant sexual preferences: aversion techniques (such as covert sensitization and satiation therapy) and techniques that utilize positive reinforcement (such as masturbatory reconditioning). Covert sensitization involves pairing the deviant sexual behavior with an aversive stimulus via imaginal exposure so that the offender eventually finds the sexual behavior to be aversive. In contrast, satiation therapy involves combining the deviant stimulus with prolonged masturbation so that over time the stimulus loses its appeal. Masturbatory reconditioning involves replacing the deviant sexual stimulus (e.g., child) with a pro-social stimulus (e.g., an adult) which is then reinforced through masturbation. While these sorts of techniques have been in use for many years, there is little research available that has examined their efficacy. Also, much of the research that has been conducted has used single-case designs which of course do not allow comparison with other treatment methods.

Marshall et al. (2009) conclude that based on the available evidence, the best approach to treating deviant sexual arousal would be to use a combination of masturbatory reconditioning and satiation therapy. In this way the individual's arousal to the pro-social stimulus would be strengthened, while the arousal to the deviant stimulus would be weakened.

Emotional Regulation

Research indicates that those who commit sexual offenses tend to have difficulties regulating their emotions and may commit sexual crimes as a means of avoiding experiencing negative emotion (Mandeville-Nordon and Beech 2004). Evidence suggests that, in particular, sexual offenders often struggle to cope adaptively with stressful life events. It seems that when faced with stressful situations, sexual offenders often use maladaptive and ineffective coping strategies which increase the likelihood of reoffending (Cortoni and Marshall 2001). For example, it has been observed that rapists may experience high levels of anger and resentment in response to stress which may contribute to their offending.

Other evidence indicates that sexual offenders often use masturbation with either appropriate or inappropriate sexual fantasies as a means of escaping the experience of negative emotion (Cortoni and Marshall 2001), and such responses could increase the likelihood of sexual offending, especially, if it reinforces the idea of inappropriate sexual interaction. Thus, researchers have suggested that treatment should include a component that teaches offenders how to respond constructively to the negative emotions that may arise from stressful life events (Beech and Fisher 2002). However, to date, there is little research available that shows whether the inclusion of such components in treatment leads to a reduction in sexual reoffending. A study by Serran and colleagues (2007) found that sex offenders who had completed treatment which included a coping skills component appeared to be better equipped (than the wait-listed controls) at dealing with high-risk situations. However, the study did not include a follow-up looking at recidivism; thus, it is unclear whether this would translate into a reduction in reoffending.

As mentioned above, in terms of rape, anger is a particularly prominent negative emotion. Howitt (2006) states: “Anger control problems have to be seen as an issue with rapists for whom issues of anger are common” (p. 369). However, some researchers have questioned the use of anger management components in treatment programs for rapists due to the fact that anger does not appear to be a problem for all rapists. For example, Loza and Loza-Fanous (1999) investigated anger levels across a range of offender groups and found no difference in levels between rapists and non-rapists (or between violent and nonviolent offenders). The authors then question the usefulness of anger management in the treatment of rapists. Arguably, this study would have been more worthwhile if it had also included a non-offender population as it may be the case that offender populations have high levels of anger problems. If so, then rapists' anger levels, though similar to other offenders, may still be problematic and potentially contribute to their risk of reoffending.

Cognition

It is widely acknowledged that sexual offenders may harbor a range of offense-related attitudes and beliefs. Cortoni (2009) states: “. . .an overall predisposition to be tolerant of sexual offending is related to sexual recidivism” (p. 44). Thus, the targeting of offense-related attitudes and beliefs is a mainstay of rehabilitation approaches. One common type of belief that is endorsed by sexual offenders is a belief of “entitlement” to sex. This can include specific beliefs such as “men need more sex than women do” and “everyone is entitled to sex” (Pemberton and Wakeling 2009). Also, child sex offenders may believe that children are sexual objects and that having sex with children is good because it educates them about sex. In contrast, research has shown that those who commit rape offenses tend to objectify and sexualize women and that they hold quite specific beliefs about specific situations. For instance, rapists often endorse beliefs along the lines of "if a woman flirts, then she wants to have sex" or "if a woman accepts a free drink, then she wants to have sex."

Within the sex offender field these problematic attitudes and beliefs are typically referred to a "cognitive distortions," and they are frequently used by offenders to justify their offending. In this way, they may be used by the offender to feel comfortable with continued offending. For example, a child sex offender might justify his ongoing abuse of a child with the belief that the sexual interaction with the child is appropriate because the child is choosing to engage in the sexual activity and shows no sign of being unwilling. In association with his belief, he may also believe that the child is not being harmed in any way. In this way the abuser attempts to minimize his level of culpability, and this contributes to his decision to continue with the behavior. While sexual offenders will usually report that they had an awareness of wrongdoing, they will often use their cognitive distortions to distract themselves from this awareness.

As mentioned above, sexual offenders often find it difficult to admit to their offending due to the social ramifications. For example, they may believe that if they admit to what they have done, they will lose the support of family and friends. Also, in some cases, the sense of shame and embarrassment make it very difficult for them to talk to others about it, including those involved in correctional and rehabilitative services. It has been reported that most sexual offenders either deny that they committed the crime or at least attempt to minimize their level of responsibility (Marshall 2006). Thus, it might be assumed that denial and minimization would be important targets for treatment. In fact, many treatment programs do not allow deniers to participate because some components of treatment require the offender to admit to their offending. For example, often offenders are expected to describe their offense in detail in order to identify points at which they could have behaved differently and avoided offending.

There may be a range of other reasons that treatment programs may preclude those who deny their offending. For instance, deniers may have a negative impact on other individuals in the group; if they are not willing to be honest, then others may be less inclined to be upfront and open. Also, those who are unwilling to admit to and discuss their offending may be less likely to benefit from treatment; thus, they may take the place of someone who may have had more to gain from taking part in treatment. This issue arises because most sex offender treatment programs have limited places available. As they require significant resources to run, organizations (typically prisons) are only able to run a limited number. However, Marshall (2001, 2006) argues that excluding those who deny their sexual offending has the effect of putting the public at greater risk because those offenders are then released untreated.

When denial is addressed in treatment, there are two general approaches. One approach is to include a specific module that targets denial during the initial stages of the treatment program, while the other approach is to place deniers in a specially designed treatment program. When the former approach is taken, the issue of denial is usually discussed in an individual setting, prior to the beginning of treatment proper, so that the offender can consider and perhaps reconsider the extent to which they will be willing to discuss the offense, prior to having to communicate with fellow group members.

Intimacy and Social Skills

Evidence suggests that sexual offenders find it difficult to establish and maintain consensual intimate relationships and that subsequently they are vulnerable to experiencing social isolation and loneliness (Mandeville-Nordon and Beech 2004). Mandeville-Norden and colleague also report that individuals who commit sexual crimes tend to lack confidence and assertiveness skills; thus, it may be hypothesized that the lack of confidence and problems being assertive contribute to the intimacy difficulties. Also, research has shown that sex offenders report significant feelings of loneliness even within the context of intimate relationships. Therefore, social skills training is a mainstay of sex offender treatment programs. As explained by Marshall (2006), social skills training is usually a broad-based aspect of treatment which may include a range of components, including problem solving skills, assertiveness training, self-esteem building, and sex education.

The Extended Sex Offender Treatment Programme (ESOTP) in the United Kingdom is an example of a program that addresses social functioning. This program for high-risk offenders, which involves 140 h of treatment over 68 sessions, includes a component titled “intimacy skills.” This component includes skill development in the areas of jealousy management, conflict resolution, and giving and receiving social support. A recent survey of sex offender treatment programs in Canada reported that three quarters of their programs included components that targeted intimacy, relationships, and social skills. Overall, research shows that both intimate relationships and more general social functioning are addressed in most sex offender treatment programs.

Some research has examined the relationship between sexual offending and adult attachment styles. For instance, Lyn and Burton (2004) examined attachment styles in a “Midwestern United States” sample of sex offenders and found that the majority (85 %) were assessed as having insecure patterns of attachment. Furthermore, when Lyn and colleague compared sex offenders with non-sex offenders, they found that the former were significantly more likely to display signs of insecure attachment. In a further study Lyn and Burton (2005) explored the particular types of problematic attachment that sexual offenders tended to display, and they found that anxiety and avoidance were seen most frequently.

William Marshall is well known for his work on the role of attachment in sexual offending. He has long argued that attachment problems contribute to the development of sexual offending. Furthermore, he has examined some of the childhood difficulties which may be associated with attachment problems. For example, he has reported that sexual offenders often have problematic childhoods, characterized by disruption, neglect, and abuse (both physical and sexual) (e.g., Marshall and Marshall 2000). Also, as explained by Marshall and Marshall, while such aversive early experiences may lead to attachment problems, they may also lead to a range of more general difficulties, such as low self-esteem and limited relationship skills. Thus, attachment theory offers a psychological mechanism that may explain the intimacy and relationship problems that are seen in sexual offenders.

Victim Empathy

Many, if not most, contemporary sex offender treatment programs include a component that focuses on the development of victim empathy. Research examining empathy in sexual offenders has found that while they often lack empathy for their victims, they do not typically have more generalized empathy deficits. Thus, treatment that aims to develop sexual offenders’ empathy is usually victim-focused. For instance, some treatment programs will require offenders to listen to a voice recording of a victim talking about the impact of his or her own sexual abuse experience. Alternatively, the offenders may be required to read stories of victims’ experiences or to read accounts written by their own victims.

Empathy is a complex concept and, historically, there have been a variety of approaches to defining it. While some theorists have conceptualized it as essentially involving cognitive processes, others have construed it as the ability to recognize emotion in others and to take on the perspective of another person. More recently, it seems that many theorists have viewed it as a multifaceted concept involving thoughts, emotions, and behaviors. With regard to sexual offenders, Marshall (2006) suggests that empathy deficits are frequently associated with the presence of victim-related cognitive distortions. For example, a child sex offender may believe that his sexual activity with his victim is educational and, therefore, not harmful. Or he may believe that if the child is quiet during the sexual interaction, then the child is enjoying it; thus, the offender may misinterpret the child’s emotional response.

Given the apparent connection between empathy deficits and cognitive distortions, Marshall proposes that it may be possible to address empathy deficits simply by challenging an offender's cognitive distortions, rather than directly targeting empathy problems.

Other researchers have suggested that empathy deficits in sexual offenders may be associated with difficulties with the identification and awareness of emotion (e.g., Gannon et al. 2008). Although, this seems to contrast with the suggestion that sexual offenders' empathy deficits are specific to their offending. Arguably, if they have difficulty with their own emotional awareness, then one might expect that they would have more general problems empathizing with others. One commonly used technique for the development of emotional awareness in group treatment is encouraging participants to identify and describe the emotion that they are experiencing at that moment in treatment. This is believed to facilitate the development of an ability to put a name to one's emotional experience and to describe that experience to others.

Some treatment programs utilize victim role plays as a way of enhancing offenders' empathy for their victims. A study by Webster et al. (2005) examined the effectiveness of offense reenactments and concluded that their inclusion in treatment appeared to enhance offenders' victim empathy. The approach used in the study, which was tailored to each participant, included the development of an offense map which was then used as a guide for walking the offender through the offense while he was role-playing the victim. Any sexual or violent acts were depicted symbolically, using light touch on a nonsexual area of the body. The study included a control group who carried out "empathy deficit role play scenarios" (p. 67) but did not engage in any victim role plays. While, as mentioned above, the researchers found this approach had clinical benefits, they also found that the method seemed to be more effective in treating rapists than child sex offenders and that the differences between the two approaches were small.

It is important to point out that although, as outlined above, victim empathy is a frequently utilized component of sex offender treatment programs, research has not been able to show that it leads to a decrease in reoffending. However, demonstrating this sort of causal connection will always be a challenging task because treatment is typically multifaceted; therefore, it is difficult to know for sure which elements of treatment are bringing about change. Furthermore, given that sexual offenders present with victim-specific empathy deficits, it makes clinical sense to address this in treatment.

Substance Abuse

There is now a significant body of research that demonstrates a link between alcohol use, other drug use, and various types of criminal behavior. Further, as stated by Fridell and colleagues (2008, p. 800), "Alcohol abuse precedes or accompanies a large proportion of violent crime." Also, research has shown that problematic alcohol use is associated with a tendency to reoffend. With regard to sexual offenders, evidence suggests that a significant proportion have drug and or alcohol-related problems. For example, a study by Marshall (1996) found that 50 % of sexual offenders were intoxicated with alcohol at the time that they committed their most recent sexual offense. Also, as with general offending, drug and alcohol abuse has also been found to be associated with sexual reoffending. For instance, research suggests that alcohol abuse at least doubles the likelihood that a sexual offender will reoffend. Some researchers have observed that drug and alcohol problems are often associated with problematic lifestyle choices which lend themselves to offending behavior. Thus, drug and alcohol issues and offending may have a common origin.

While this may indeed be the case, it is important to acknowledge that intoxication with any substance has a range of consequences that can directly impact on offending. For example, alcohol can significantly diminish an individual's ability to understand and respond appropriately to a social situation. Specifically, it impairs one's judgment, thus making it more likely that an individual will respond in an antisocial manner. Furthermore, research has shown that alcohol can increase one's tendency for impulsivity, meaning that an offender will be more likely to respond to a situation quickly and without a proper appraisal of the consequences of his behavior. Also, there are well-established links between excessive alcohol use and aggression in those who have already demonstrated aggressive tendencies.

A study by Abracen et al. (2006) compared sexual and violent (nonsexual) offenders in terms of the presence of alcohol abuse, using the Michigan Alcohol Screening Test (MAST). Results showed that sexual offenders had significantly higher scores thereby indicating that they were more likely to have problematic patterns of drinking. In their discussion of the results, they suggest that myopia theory may offer an explanation of why alcohol use may contribute to sexual offending. Myopia theory suggests that when an individual is considering engaging in a risky behavior, alcohol may have the effect of decreasing their inhibition via a reduction in their information processing capacity. This is similar to effect of alcohol use on judgment which is suggested above.

Evidence suggests that there are a variety of approaches to addressing drug and alcohol problems. While some programs include a substance abuse component, others do not address it directly but direct individuals to appropriate drug and alcohol programs. Often drug and alcohol problems are construed as responsivity barriers and are thus seen as being best addressed prior to offense-specific treatment. For example, even in many prison environments, drugs are available and, therefore, it is considered that in order for an offender to make the most of offense-specific treatment, he should first address his drug use. Obviously, in community settings drug and alcohol use is more prevalent; thus, the same approach is often taken.

Treatment Efficacy

There is now a growing body of research that has examined the efficacy of sex offender treatment programs. A study a decade ago (Hanson et al. 2002) examined the outcome evaluations of 43 sex offender treatment programs with a mean follow-up period of 76 months. Hanson et al. reported a sexual recidivism rate of 12.3 % for sex offenders who had completed treatment and a 16.8 % recidivism rate for untreated sex offenders. They also found that the treated sex offenders committed significantly fewer general offenses than their untreated counterparts (27.9 % and 39.2 %, respectively). Note that the majority of treatment programs that were included in this study had a cognitive-behavioral orientation; thus, this lends support to the efficacy of cognitive-behavioral programs.

Similarly, Lösel and Schmucker (2005) carried out a meta-analysis of 69 studies of sex offender treatment efficacy that contained 80 separate comparisons of treated and untreated offenders. The authors concluded that treatment led to a mean reduction in sexual recidivism of almost 37 % (when low base rates are taken into account). The actual difference in recidivism rates between treatment groups and control groups (who were untreated or had completed another type of treatment) was 11.1 % and 17.5 %, respectively. They also found that sex offender treatment led to a significant reduction in general reoffending. In summarizing their findings, Lösel and colleague state: “The most important message is an overall positive and significant effect of sex offender treatment” (p. 135).

Another study (Seager et al. 2004) looked at recidivism rates in 109 sex offenders who had completed offense-specific cognitive-behavioral treatment and 37 who had not completed any treatment for their sexual offending. Note that of the 109, only 81 were assessed as having successfully completed the program. The recidivism rates that were examined included sexual and violent offenses. Results showed that regardless of whether individuals were considered to have been successful in completing the program, those who completed were found to have a lower recidivism rate. Specifically, 4 % of successful completers and 7 % of unsuccessful completers were reconvicted within the 2-year follow-up period. This contrasted with 18 %, 42 %, and 100 % (respectively) of those who withdrew from treatment, declined it, or were terminated from the program. However, the authors concluded that there was no evidence of a connection between offenders' change in clinical phenomena (such as empathy) and their reoffending. For example, they found that the participants risk scores on the Static 99 (an actuarial risk measure) predicted their likelihood of reoffending regardless of whether they completed. Seager and colleagues concluded that "participation in the sex offender program did not reduce recidivism rates for those who complied with treatment but merely enabled motivated offenders to concretely demonstrate their commitment to not reoffend" (p. 609).

It is unclear whether the two aforementioned meta-analyses took into consideration the various ways in which treatment completers and those who did not complete treatment might have differed. The study by Seager et al. suggests that apparent treatment effects may simply be a result of preexisting differences between those who undertake such programs and those who do not. However, it is important to note that Seager's study was comparatively small and only included a small number of offenders. Nonetheless, it does raise questions about the way in which studies of treatment efficacy are conducted and how results are interpreted.

It seems reasonable to conclude that sex offender treatment is often effective and that at the very least it does no harm. Furthermore, there is growing evidence that cognitive-behavioral approaches may be especially useful. However, further studies are needed to determine if the effects that are being found are indeed due to treatment or to the fact that particular individuals tend to undertake and complete treatment.

Summary

This entry has discussed two of the key models that underlie sex offender treatment. It has also examined some of the relevant process issues, such as whether treatment should be delivered in group settings. Further, it looked at the components that are typically found in sex offender treatment programs. The entry concluded with a brief discussion of treatment efficacy. It is hoped that this overview will provide the reader with some understanding of how the treatment of sexual offenders is ordinarily delivered and of some of the issues surrounding such treatment. Obviously, there is a need for ongoing research in many areas as many findings remain preliminary.

Related Entries

- ▶ Actualizing Risk-Need-Responsivity
- ▶ Good Lives Model
- ▶ Sexual Recidivism
- ▶ Specialization and Sexual Offending
- ▶ Treatment of Sex Offenders

Recommended Reading and References

Abracen J, Looman J, Di Fazio R, Kelly T, Stirpe T (2006) Patterns of attachment and alcohol abuse in sexual and violent non-sexual offenders. J Sex Aggress 12:19–30. doi:10.1080/13552600600722963

Andrews DA, Bonta J (2003a) The psychology of criminal conduct, 3rd edn. Anderson, Cincinnati

Andrews DA, Bonta J (2003b) A commentary on Ward and Stewart's model of human needs. Psychol Crime Law 9:215–218. doi:10.1080/10683/16031000112115

Beech AR, Fisher DD (2002) The rehabilitation of sex offenders. Aust Psychol 37:206–214. doi:10.1080/00050060210001706886

Burdon WM, Gallagher CA (2002) Coercion and sex offenders: controlling sex-offending behavior through incapacitation and treatment. Crim Justice Behav 29:87–109. doi:10.1177/0093854802029001006

Cortini F (2009) Factors associated with sexual recidivism. In: Beech AR, Craig LA, Browne KD (eds) Assessment and treatment of sex offenders: a handbook. Wiley-Blackwell, Chichester, pp 39–52

Cortoni F, Marshall WL (2001) Sex as a coping strategy and its relationship to juvenile sexual history and intimacy in sexual offenders. Sex Abus J Res Treat 13:27–43. doi:10.1023/A:1009562312658

Fridell M, Hesse M, Jaeger MM, Kühlhorn E (2008) Antisocial personality disorder as a predictor of criminal behaviour in a longitudinal study of a cohort of abusers of several classes of drugs: relation to type of substance and type of crime. Addict Behav 33:799–811. http://org/10.1016/j.addbeh.2008.01.001

Gannon TA, Collie RM, Ward T, Thakker J (2008) Rape: psychopathology, theory, and treatment. Clin Psychol Rev 28:982–1008. doi:10.1016/j.cpr.2008.02.005

Hanson RK, Gordon A, Harris AJR, Marques JK, Murphy W, Quinsey V, Seto M (2002) First report of the collaborative outcome data project on the effectiveness of psychological treatment for sexual offenders. Sex Abus J Res Treat 14:169–194. doi:10.1023/A:1014624315814

Harkins L, Beech AR (2008) Examining the impact of mixing child molesters and rapists in group-based treatment for sexual offenders. Int J Offender Ther Comp Criminol 52:31–45. doi:10.1177/0306624X07300267

Howitt D (2006) Introduction to forensic and criminal psychology. Pearson Education, Harlow

Langevin R (2006) Acceptance and completion of treatment among sex offenders. Int J Offender Ther Comp Criminol 50:402–417. doi:10.1177/0306624X06286870

Lösel F, Schmucker M (2005) The effectiveness of treatment for sexual offenders: a comprehensive meta-analysis. J Exp Criminol 1:117–146. doi:10.1007/s11292-004-6466-7

Loza W, Loza-Fanous A (1999) The fallacy of reducing rape and violent recidivism by treating anger. Int J Offender Ther Comp Criminol 43:492–502. doi:10.1177/0306624X99434007

Lyn TS, Burton DL (2004) Adult attachment and sexual offender status. Am J Orthopsychiatry 74:150–159. doi:10.1037/0002-9432.74.2.150

Lyn TS, Burton DL (2005) Attachment, anger, and anxiety of male sexual offenders. J Sex Aggress 11:127–137. doi:10.1080/13552600500063682

Mandeville-Nordon R, Beech A (2004) Community-based treatment of sex offenders. J Sex Aggress 10:193–214. doi:10.1080/1355260042000261760

Marques JK, Wiederanders M, Day DM, Nelson C, van Ommeren A (2005) Effects of a relapse prevention program on sexual recidivism+. Final results from California's sex offender treatment and evaluation project (SOTEP). Sex Abus J Res Treat 17:79–107. doi:10.1007/s11194-005-1212-x

Marshall WL (1996) Assessment, treatment, and theorizing about sex offenders: developments during the last 20 years and future directions. Crim Justice Behav 23:162–199. doi:10.1177/0093854896023001011

Marshall WL (2001) Adult sexual offenders against women. In: Holin CR (ed) The handbook of offender assessment and treatment. Wiley, Chichester, pp 333–348

Marshall WL (2006) Diagnosis and treatment of sexual offenders. In: Weiner IB, Hess AK (eds) The handbook of forensic psychology, 3rd edn. Wiley, Hoboken, pp 790–818

Marshall WL, Marshall LE (2000) The origins of sexual offending. Trauma Violence Abuse 1:250–263. doi:10.1177/1524838000001003003

Marshall WL, Serran GA, Fernandez YM, Mulloy R, Mann RE, Thornton D (2003) Therapist characteristics in the treatment of sexual offenders: tentative data on their relationship with indices of behaviour change. J Sex Aggress 9:25–30. doi:10.1080/355260031000137940

Marshall WL, O'Brien MD, Marshall LE (2009) Modifying sexual preferences. In: Beech AR, Craig LA, Browne KD (eds) Assessment and treatment of sex offenders: a handbook. Wiley, Chichester, pp 311–327

Maruna S (2001) Making good: how ex-convicts reform and rebuild their lives. American Psychological Association, Washington, DC

Marx BP, Miranda R, Meyerson LA (1999) Cognitive-behavioral treatment for rapists: can we do better? Clin Psychol Rev 19:875–894. doi:10.1016/S0272-7358

Milner RJ, Wakeling HC, Mann RE, Webster SD (in press) Clinical impact of a socio-affective functioning programme for high risk sexual offenders

New Zealand Department of Corrections (2009) What works now? A review and update of research evidence relevant to offender rehabilitation practices within the Department of Corrections. http://www.corrections.govt.nz/research

Pemberton AE, Wakeling HC (2009) Entitled to sex: attitudes of sexual offenders. J Sex Aggress 15:289–303. doi:10.1080/13552600903097212

Seager JA, Jellicoe D, Dhaliwal GK (2004) Refusers, dropouts, and completers: measuring sex offender treatment efficacy. Int J Comp Criminol 48:600–612. doi:10.1177/0306624X04263885

Serran GA, Fernandez Y, Marshall WL, Mann R (2003) Process issues in treatment: application to sexual offender programs. Prof Psychol Res Pract 34:368–374. doi:10.1037/0735-7028.34.4.368

Serran GA, Moulden H, Firestone P, Marshall WL (2007) Changes in coping following treatment for child molesters. J Interpers Violence 22:1199–1210. doi:10.1177/0886260507303733

Ward T, Stewart CA (2003) The treatment of sexual offenders: risk management and good lives. Prof Psychol Res Pract 34:353–360. doi:10.1037/0735-7028.34.4.353

Ward T, Collie RM, Bourke P (2009) Models of offender rehabilitation: the good lives model and the risk-need-responsivity model. In: Beech AR, Craig LA, Browne KD (eds) Assessment and treatment of sex offenders: a handbook. Wiley, Chichester, pp 293–310

Ward T, Gannon TA, Yates PM (2008) The treatment of offenders: current practice and new developments with an emphasis on sex offenders. Int Rev Victimol 15:179–204. doi:10.1177/026975800801500207

Webster SD, Bowers LE, Mann RE, Marshall WL (2005) Developing empathy in sexual offenders: the value of offence re-enactments. Sex Abus J Res Treat 17:63–77. doi:10.1007/s11194-005-1211-y

Wilson R, Yates P (2009) Effective interventions and the good lives model: maximising treatment gains for sexual offenders. Aggress Violent Behav 14:157–161. doi:10.1016/j.avb.2009.01.007

Wong YJ (2006) Strength-centred therapy: a social constructionist, virtues-based psychotherapy. Psychother Theory Res Pract Train 2:133–146. doi:10.1037/0033-3204.43.2.133

Yates PM, Goguen BC (2000) National sex offender treatment programme description. Correctional Service of Canada, Ottawa

Sex Offenders and Criminal Policy

Richard Tewksbury and David P. Connor
Justice Administration, University of Louisville,
Louisville, KY, USA

Overview

Over the past several years, individuals convicted of criminal acts have been subjected to increasingly severe sanctions. In particular, efforts have been made to increase the supervision of criminals and decrease offender opportunities to further perpetrate crimes. Without a doubt, many emerging laws and criminal justice policies have targeted sex offenders, especially those convicted of victimizing children. In addition to progressively long incarcerations, sex offenders potentially face civil commitment, registration and community notification, chemical castration, polygraph testing, and residency restrictions. The United States has held a leading position in passing these laws and enacting policies aimed at preventing future sex crimes.

Explanations for the renewed interest in punitive sex offender legislation may be found in the high-profile sexual attacks on children in the 1990s. Child sexual assault victims whose names saturated media headlines and generated castigating policies toward sex offenders included Jacob Wetterling, Megan Kanka, and Adam Walsh. Although not new concepts, the revival of criminal registration and civil commitment for sex offenders reflect growing public and political concerns. Technological innovations may also be responsible for intensified sex offender legislation. Chemical castration laws and polygraph testing, for example, are likely the result of expanding scientific know-how.

Background Description

Civil Commitment Statutes

In response to high-profile sex crimes, the first American sexual psychopath laws were passed in the 1930s. Community protection, as well as treatment and incapacitation of sex offenders, were the explicit goals of these laws. Often calling for the civil commitment of so-called mentally disordered sex offenders to public mental hospitals, these statutes rested on the assumption that mental health professionals were capable of identifying, confining, and treating sexual psychopaths.

Criticism surrounding sexual psychopath laws has always existed. Those offenders determined to be sexual psychopaths were often committed to mental institutions indefinitely with few procedural safeguards. These laws may have also been aimed at sex offenders other than violent recidivists and included exhibitionists, voyeurs, and homosexuals. Moreover, public disenchantment with rehabilitation generally, and lack of confidence in effective treatment methods for sexual psychopaths specifically, eventually dissuaded efforts to utilize these laws. By the late 1960s, many states started to repeal, intentionally disregard, and seldom employ these sexual psychopath laws.

Registration and Community Notification

Registration of individuals charged or convicted of various crimes has been used for decades. In 1994, the Jacob Wetterling Act put into practice the registration of sex offenders in statewide databases. Culpable for transforming sex offender registries into publicly available online domains, Megan’s Law, passed in 1996, requires state law enforcement agencies to make public information about sex offenders. Affording a sense of safety and control to the public, such laws have subsequently also been passed unanimously by many legislatures. Today, all states, as well as the federal government, have enacted and maintain publicly available, Internet-based sex offender registries.

Castration Laws

For many centuries, sex offenders were punished through physical castration. In the United States, castration was first popularly used for slaves suspected of having sex with white women. The eugenics movement (1905–1935) supported castration and sterilization of criminals and the mentally ill. Currently, voluntary chemical castration (combined with therapy) has largely replaced physical castration, as medical doctors believe that similarly effective results can be reached through treatment with medication.

Concerns over body mutilation, intrusiveness, and the lasting results of surgery have rendered physical castration of sex offenders in the United States as largely unacceptable. The widespread availability of medications as an alternative to physical castration may also explain this change in practice. Although some laws concerning sex offenders allow both physical and chemical castration as a substitute or adjunct to punishment, chemical castration seems to be the more socially acceptable solution.

Polygraph Testing

Over the last two decades, polygraph testing has been presented as yet another tool to manage, supervise, and treat sex offenders under community supervision. Numerous states require individuals convicted of two sex offenses to undergo periodic mandatory polygraph supervision. Until recently, *Frye v. United States* (1923) governed the admissibility of polygraph results in American courts. In this case, the Court of Appeals of District of Columbia ruled that when questions of fact require special knowledge, the opinions of witnesses skilled in the subject to which the questions relate are admissible in evidence.

In 1993, however, the US Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals* ruled that the Federal Rules of Evidence should provide the standard for admitting expert scientific testimony. To determine the admissibility of polygraph evidence, trial courts must now consider factors such as whether lie detector testing has been scientifically tested, largely accepted within the scientific community, and exposed to review and publication. As a result, most states currently do not consider polygraph evidence as admissible; however, if both parties stipulate, a few courts may admit polygraph results.

Appellate courts in *Himes v. Thompson* (2000) and *Kansas v. Lumley* (1999) have ruled that polygraph evidence is admissible, as well as sufficiently reliable as evidence, at probation and parole revocation hearings. Courts in *State v. Flores-Moreno* (1994) and *State v. Riles* (1997) have also held that polygraph testing limited to topics related to crimes perpetrated by sex offenders is permissible as a condition of community supervision. Moreover, the Washington Court of Appeals in *State v. Eaton* (1996) expressed that polygraph testing as a requirement for probationers and parolees was a necessary and effective way in which to monitor compliance with terms of supervision.

Residency Restrictions

Residency restriction laws have been created as a result of increasing concern about registered sex offenders living in communities. After the murder of Jessica Lunsford in Florida by a previously convicted sex offender, housing restrictions for sex offenders in the United States became more widespread. Despite punishment, treatment, and supervision, many believe that sex offenders continue to pose a serious danger to society. These housing restrictions prevent sex offenders from living near schools, parks, daycare centers, and other designated "child congregation" locations. In this way, these laws seek to limit contact between registered sex offenders and children, and subsequently reduce sex offender recidivism.

State of the Art

Civil Commitment Statutes

The premise of civil commitment statutes is that many sex offenders cannot be rehabilitated; consequently, such criminals must be incapacitated to the greatest extent possible. Under civil law, sex offenders may be committed to institutions. Modern civil commitment statutes, commonly known as sexually violent predator laws, allow governments to confine particular sex offenders to secure mental health facilities upon their release from prison or a judicial finding of incompetency to stand trial. Upon the decision of a court, sex offenders considered to have mental abnormalities or other psychological disorders that may prompt harmful sexual conduct in the future may be subjected to civil commitment. As of 2009, 20 states and the federal government have civil commitment laws aimed at sex offenders.

Following the placement of committed sex offenders to secure institutions, mental health clinicians and other professionals are assigned the responsibility of evaluating offenders at regular intervals. During periods of civil commitment, medical and psychological assessments of sex offenders are performed to assess improvements in mental status. Examinations and subsequent reports prepared for the court concerning the mental conditions of committed sex offenders are typically completed on an annual basis. After these examinations by medical and mental health officials, sex offenders civilly committed to institutions may petition for release, appear before a judge, and ask the court to determine whether their commitments continue to be necessary to protect the public. In order to be released, psychologists and the court must agree that mental abnormalities or personality disorders once exhibited by sex offenders no longer pose a threat to society.

Whereas earlier sexual psychopath laws may have centered on rehabilitation, current civil commitment laws focus on the social control and incapacitation of sex offenders. Depicting the social control function of civil commitment laws, the US Supreme Court in *Kansas v. Hendricks* (1997) upheld a sexually violent predator statute, which sent sex offenders determined to be violent predators likely to recidivate to state mental hospitals. Although Hendricks challenged the civil commitment law as a violation of due process and prohibitions against double jeopardy and (*ex post facto*) laws, the Court found such to be constitutional as the civil commitment statute was not deemed to be punishment. In 2002, the Supreme Court ruled in *Kansas v. Crane* that mental abnormalities displayed by sex offenders must differentiate committed individuals from ordinary recidivists. Further, besides showing the likelihood that offenders will commit sex crimes upon release, it must be shown that offenders have a serious inability to control their behaviors.

Registration and Community Notification

Sex offenders have consistently faced stringent sentencing laws; moreover, society has consistently looked upon them with disdain. Spotlighting society's harsh treatment of sex offenders, one of the most recent developments has been the creation of sex offender registries. Sex offender registries are utilized in every jurisdiction in the United States, and these repositories of information provide online access to a wide array of facts about convicted sex offenders and their offenses. Individuals convicted of sex crimes are typically required to provide local law enforcement and corrections authorities with name, photograph, address, birth date, Social Security number, fingerprints, offense history, date of convictions, and other information. Sex offenders must verify the accuracy of this information on a routine basis for the duration of their registration, which may range from 10 years to life.

Most arguments supporting sex offender registries emphasize public safety, particularly the protection of children. Proponents also contend that registration will permit law enforcement officials to quickly and easily ascertain the locations of sex offenders in their communities, facilitating sex crime investigations. Because Internet databases reveal the identification of sex offenders to the community, it is further maintained that sex offender registries reduce opportunities for recidivism. Despite evidence suggesting little or no effect of registration and community notification on recidivism rates, there remains general public approval concerning sex offender registries and a belief that most sex offenders will reoffend.

Community notification typically occurs in neighborhood meetings, door-to-door visits by the police, newspaper advertisements, online notices, and flyers circulated throughout a jurisdiction. Many states that use community notification have a three-tiered system based on the purported dangerousness of sex offenders that determines the degree of notification that will take place. When sex offenders are categorized as the lowest risk to public safety, notification is typically reserved for law enforcement officials only. Schools, daycares, and other neighborhood organizations are notified of the presence of sex offenders posing a medium risk to public safety. Those sex offenders considered the most dangerous, designated at high risk, will generate the most widespread notification, as the general public is notified. Some states use formal assessment tools to classify sex offender risk levels. Other jurisdictions utilize committees of clinicians and other professionals to evaluate offenders and decide upon risk level.

Four major laws are associated with contemporary sex offender registration and notification policies. The Jacob Wetterling Crimes Against Children and Sexually Violent Offender Registration Act (1994) is the first federal law that mandated statewide registration of sex offenders. This statute requires all states to establish procedures for sex offender registration. If particular jurisdictions fail to comply with the Wetterling Act, they risk the loss of crime prevention funding from the federal government. As a result of the Wetterling Act, each state has a mandatory registration law that obligates sex offenders to provide their information to law enforcement officials.

Despite the fact that Washington enacted the first community notification law in 1990, it was Megan's Law (New Jersey Stat. Ann. §§ 2C: 7-1 to 2C: 7-11, 1996) in New Jersey that created registration and community notification legislation that was ultimately replicated nationwide. Megan Kanka was sexually assaulted and murdered by a twice-convicted sex offender on parole. Amending the Jacob Wetterling Act, Megan's Law calls for the registration of sex offenders with law enforcement at various time intervals. The frequency of registration is contingent upon assumed risk levels of sex offenders. In many states, Megan's Law also requires the local police to inform communities of the presence of sex offenders moving into neighborhoods. The jurisdiction in which sex offenders reside, as well as their risk status, will determine the degree of information that is made available to the public.

In 2000, the Campus Sex Crimes Prevention Act further amended the Wetterling Act, obligating registered sex offenders studying and working at colleges and universities to provide notice of their status as sex offenders to these institutions of higher learning. College and university officials are required to inform the campus community where information regarding registered sex offenders may be obtained. In fact, many colleges and universities maintain their own distinctive online sex offender registry. At the very least, these registries provide the full names of students and university employees convicted of sex offenses. More often, campus sex offender registries include demographic information, including physical descriptions and dates of birth. Some university registries also contain the offender's home address, conviction information, description of victims, and special conditions imposed at sentencing. Further, in addition to the academic institutions with campus-based registries, many institutions of higher learning provide online links to the state sex offender registry from the webpages of campus police or the public safety department. This information is also provided to law enforcement agencies in the jurisdiction of the college or university, and it is entered into state registry records.

The abduction of Polly Klaas generated the development of three-strikes laws. Polly's murder by a repeat sex offender who was not required to register infuriated the public and generated intense media coverage. This resulted in the overwhelming public support for Proposition 184 in California. Passed in 1994, this legislation allowed a third felony conviction to result in a sentence of 25 years to life. Although those championing three-strikes laws contend that such legislation serves as a general deterrent to potential three-time offenders, critics note that there is little evidence that shows reduction in serious crime has resulted from implementation of three-strikes laws.

Accessible to the public through the Internet, a nationwide databank of registered sex offenders was created in 2005. The Dru Sjodin National Sex Offender Public Registry, as it was named in 2006, was designed by the Department of Justice to provide more efficient access to individual state sex offender registries.

The Adam Walsh Child Protection and Safety Act was signed into law in 2006. This federal law includes the Sex Offender Registration and Notification Act (SORNA) and seeks to make the archiving and monitoring of sex offenders more efficient by creating a comprehensive and national system for sex offender registration. The legislation categorizes sex offenders into three tiers of risk. Tier three sex offenders are the most serious and must verify their location with law enforcement every 3 months for the rest of their lives. Individuals defined as tier two offenders must update their whereabouts every 6 months for 25 years, while sex offenders described as tier one lawbreakers must verify their location annually for 15 years.

The Adam Walsh Act includes additional provisions for those convicted of sex crimes. Failing to register and update information with authorities are increased to felony offenses. Failure to adhere to registration guidelines under SORNA is punishable by 10 years in prison and a $250,000 fine. The federal law also eliminates the statute of limitations concerning the prosecution of child abduction and felony child sexual offense cases. It allows victims of child abuse to utilize civil proceedings to seek monetary damages from perpetrators. Further, the Adam Walsh Act sets forth mandatory minimum sentencing for particular sex crimes. For instance, the law dictates a minimum sentence of 30 years for the rape of a child and calls for increasing the minimum prison terms for offenders traveling between states with minors.

All jurisdictions must comply with the Adam Walsh Act and the provisions of SORNA or face reduced federal grant funding. In September 2009, Ohio was the first state to comply with federal sex offender provisions under SORNA. However, as of 2011, 43 states have yet to achieve compliance with the mandates of the Adam Walsh Act.

The constitutionality of sex offender registration and community notification has been challenged in two US Supreme Court cases. *Smith v. Doe* (2003) considered the Fifth Amendment double-jeopardy clause and whether registration and notification laws amount to further punishment of sex offenders. The Supreme Court found that sex offender registration and notification is regulatory and consequently is not additional punishment. Moreover, the Court ruled that individuals convicted of sex crimes prior to the establishment of registration and notification for sex offenders can still be obligated to comply with these laws. The second Supreme Court case regarding sex offender registration and community notification examined the issue of cruel and unusual punishment. The Court in *Connecticut Department of Public Safety v. John Doe* (2003) ruled that the posting of sex offender photographs online is constitutional. Registration and notification laws, therefore, are not considered to be obstacles to personal freedom.

Castration Laws

In the last two decades, medication as a means of controlling the behavior of sex offenders has gained renewed interest. Chemical castration consists of injecting Depo-Provera, the synthetic hormone medroxyprogesterone acetate, to reduce the blood serum testosterone levels in males. Taken on a regular basis, Depo-Provera is supposed to reduce sexual impulses, erections, and ejaculations. The potential side effects of Depo-Provera include migraines, nausea, weight gain, insomnia, fatigue, and loss of body hair, although most of these are thought to be rare and reversible.

Presently, many chemical castration laws call for the forced dispensing of medication to control the behavior of recidivist sex offenders. In 1997, California was the first jurisdiction to require chemical castration for repeat sex offenders (with victims under the age of 13). Under California law, repeat sex offenders are obligated to receive chemical injections prior to their release on parole. These injections persist until offenders complete their criminal sentences. Although sex offenders may refuse chemical castration, parole will be immediately denied to those deciding not to participate. Following California's example, other states including Colorado, Florida, Georgia, Louisiana, Montana, Texas, and Wisconsin have similar sex offender castration laws. However, unlike California, many state statutes allow judicial discretion in regards to implementation of castration.

Chemical castration laws regarding sex offenders generally do not require complete medical or psychiatric assessments prior to the injections of medication. Further, there is no distinction between types of sex offenders most appropriate for chemical castration in the statutes.

Polygraph Testing

Polygraph (or “lie detector”) testing with sex offenders focus on disclosure polygraphs performed after sentencing, denial and specific issues examinations, and maintenance polygraphs administered during sex offender treatment for purposes of reducing offender denial and gauging compliance with conditions of supervision. Following sentencing, the disclosure polygraph involves questions asked of sex offenders about their history of sexual deviance. Polygraph examiners identify offender deception through responses of intensified physiological arousal to relevant questions. If there are discrepancies among crime descriptions, the denial polygraph examination will likely be administered. Also, when sex offenders refuse to accept responsibility for their crimes, probation and parole officials may perform the denial examination. Specific issues polygraph examinations usually center on a specific accusation or suspicion. The maintenance polygraph detects how sex offenders manage inappropriate thoughts and measures compliance with conditions of probation or parole. In total, these applications of polygraph technology serve to manage, supervise, and treat sex offenders living in the community.

Residency Restrictions

Well over one-half of all states and numerous municipalities have sex offender residency restriction laws. Residency restriction laws often feature nebulous language to restrict registered sex offenders from living near locations described as “child congregation” areas. Such places are typically defined to include schools, parks, school playgrounds, daycare centers, bus stops, and recreational facilities. Fluctuating between 500 ft and 2,500 ft, residency restriction laws assert that specific distances must be preserved between a sex offender’s residence and various landmarks in the community.

Restrictive zoning laws prohibiting sex offenders from residing near places frequented by children are often associated with registration statutes and typically involve all registered sex offenders without regard to victims’ ages. However, other residency restriction laws concern only sex offenders under community supervision or those with child victims. For instance, the first state law addressing sex offender residency restrictions, which was passed in 1995, applied only to Florida sex offenders on probation with child victims.

In 2005, after a registered sex offender confessed to sexually assaulting and killing 9-year-old Jessica Lunsford in Florida, sex offender housing restrictions became more intensified nationwide. For example, California voters overwhelmingly passed Proposition 83 in 2006. Among other initiatives, Proposition 83 prevents sex offenders from living 2,000 ft from child congregation locations. Further, municipal sex offender housing ordinances became more prevalent. Enacted in December 2005, a Dyersville, Iowa, ordinance is among the most extreme of residency restrictions for sex offenders. The law essentially prohibits any sex offender from living anywhere in the city.

Possible Controversies in the Literature

Civil Commitment Statutes

Research highlights many disputes surrounding the civil commitment of sex offenders. Critics assert that in reality sex offenders committed to mental facilities are almost never released. Washington and Minnesota, the two states with the longest contemporary commitment programs, never discharged committed sex offenders. Others oppose civil commitment for sex offenders because such laws apparently punish individuals who have already paid their debt to society. Despite the Supreme Court decision in *Kansas v. Hendricks* (1997), Friedland (1999) contends that the civil confinement seems to hinge on the desire to continue punishment and incapacitation.

Besides apprehensions relating to sex offenders themselves, civil commitment statutes have also raised issues concerning mental health professionals. Wettstein (1992) argues that civil commitment laws regarding sex offenders inappropriately utilize experts in the mental health field. Although many sex offenders are committed due to their mental deficiencies or personality disorders, he notes that this does not necessarily mean committed sex offenders are genuinely mentally ill.

In addition, the increasing number of sex offenders and the need to commit them may exhaust government money allotted for the entire mental health field (Friedland 1999; Wettstein 1992) and may not be the most effective use of mental health resources already in short supply. If mental health professionals are trusted to confine sex offenders, it is argued that funding for treatment and other mental health services for committed sex offenders and other patients will become even scarcer. When appropriate mental health treatment is absent, some scholars note that it is possible that many mental health patients likely to be responsive to treatment may gravitate toward criminal activities.

Registration and Community Notification

Controversies over sex offender registration and community notification also exist. Those supporting sex offender registration and community notification often contend that these laws are not punishment; instead, any punitive actions resulting from registration and notification are only related to the public protection function (Brooks 1995). Further, as a result of registration and community notification, there is evidence that probation and parole officials and the community are working together to actively monitor sex offenders (Zevitz and Farkas 2000b).

And yet, it has been found that notification laws associated with sex offender registration negatively impacted probation and parole officers (Zevitz and Farkas 2000a). In particular, officers assigned to sex offender caseloads described a loss of staff, time, and financial resources as a result of a new community notification program. In addition, research on sex offender registration has shown unintended outcomes for the general community. Scholars have noted that sex offender registration and notification "can have the effect of leaving neighborhood residents frightened but feeling powerless to do anything about it" (Zevitz and Farkas 2000b, p. 405). Although registration laws were meant to assist the community by providing information about sex criminals, these requirements have also increased a sense of alarm among residents attending community notification meetings.

Registration and community notification may prevent successful reintegration into the community. Mental health professionals have expressed concern that reactions from the community resulting from registration and notification may intensify the anxiety of sex offenders, ultimately generating recidivism through poor choices (Billings and Bulges 2000).

Available literature concerning sex offender registration and community notification shows that registered sex offenders often experience numerous deleterious consequences. Specific offender aftermath associated with registration and community notification includes loss of family contact, destruction of friendships, employment difficulties, hostile confrontations, and threats of violence and personal harm. Another consequence of sex offender registration and community notification is the stigma that is associated with labeling as a registered sex offender (Tewksbury and Lees 2006, 2007). Making reintegration more challenging for sex offenders, powerful and enduring stigmas that result from public disclosure are strong obstacles to employment, education, and community activity (Uggen et al. 2004). Registered sex offenders in numerous studies have reported these marks of disgrace as common experiences Tewksbury (2004, 2005; Tewksbury and Lees 2006). The most commonly reported collateral consequences for registered sex offenders are feelings of vulnerability, stigmatization, and housing difficulties (Tewksbury 2004, 2005; Tewksbury and Lees 2006). Tewksbury and Mustaine (2009) explain that these experiences are mutually influential, as feelings of vulnerability may intensify as registered sex offenders find themselves unable to locate housing.

Castration Laws

Debates continue over the effectiveness of sex offender castration laws. Research has suggested that sex offenders treated with both medication and counseling may have improved control of their sexual behaviors (Melella et al. 1989), and sex offenders under chemical castration and participating in therapy are less likely to recidivate than those individuals discontinuing medication (Meyer et al. 1992). Supporters of castration laws regarding also contend that chemical castration does not violate the fundamental right of individuals to procreate. The effects of Depo-Provera are temporary and completely revocable upon termination of the injections (Melella et al. 1989).

However, despite Depo-Provera's only temporary obstruction to sexual activity, Spalding (1998) suggests that the ability to procreate is still infringed upon during the period of treatment, which may last for years. Further, some chemical castration laws, such as Florida's statute, permit the courts to order injections of Depo-Provera for the lifetime of the offender. In this way, sex offenders sentenced to lifetime chemical castration may have their right to procreate removed permanently. According to Meyer and Cole (1997), the long-term effects of Depo-Provera on sexual functioning are still very much in question.

Some opponents argue that judges may not be the most appropriate individuals to impose sentences of chemical castration and doubt the capabilities of medical doctors working in corrections to make impartial decisions concerning the well-being of offenders. Interestingly, Fitzgerald (1990) suggests that injections of Depo-Provera will not have any meaningful influence on many sex offenders. These sex offenders include those that deny committing the offense or refuse to acknowledge the criminal nature of their actions, those blaming their actions on environmental factors, and violent individuals motivated by nonsexual elements. Moreover, the use of medications for castration purposes is widely considered effective only when employed in combination with psychotherapy specifically designed for sex offenders (Meyer and Cole 1997). Other research warns that the administrative procedures for involuntary medication outlined in *Washington v. Harper* (1990) may not protect inmates from uninformed and arbitrary decisions by prison psychiatrists. Ryan argues that the objectivity of a committee consisting of correctional employees and the cooperative nature of the hearing afford little protection from ill-advised psychiatric diagnosis.

Polygraph Testing

There are disputes in the literature surrounding sex offender polygraph examinations. Polygraph testing has been reported to improve the management of sex offenders in the community, as it purportedly reduces denial, extracts confessions of sexual offenses, and improves treatment results (Wilcox 2000). On the other hand, Iacono and Lykken (1997) raised questions concerning the possibility of false positives and negatives, in spite of the use of control questions present in polygraph testing. They also reported no evidence for the validity of polygraph results in their review of field studies. Further, the fear of being incorrectly accused causing innocent subjects to have physiological responses that indicate guilt has been considered as a flaw of the polygraph. Proponents of polygraph testing have described accuracy rates beyond 90 %; however, critics (Iacono and Lykken 1997) contend that these studies contain fatal methodological flaws.

Residency Restrictions

Research dedicated to sex offender residency restrictions has only emerged in the past decade. The available research points to controversies in the literature. Finding that nearly half of sex offenders with child victims lived in close proximity to child congregation zones, Walker et al. (2001) studied one Arkansas county and suggested that child sex offenders may be more likely to intentionally reside near schools, parks, and daycares. And yet, when compared with nonrecidivists, research has reported that recidivating child sex offenders under community supervision did not live closer to these congregation zones (Colorado Department of Public Safety 2004). Tewksbury and Mustaine (2006) also showed that only approximately one in five registered sex offenders live in close proximity to such locations. Further, a Minnesota study found that sex offenders were more likely to seek victims in neighborhoods other than their own to avoid detection (Minnesota Department of Corrections 2003).

Research has recognized the problematic nature of residency restrictions aimed at sex offenders. Lawmakers believe that residency in close proximity to potential (child) victims influences recidivism (Sample and Kadleck 2008). However, research generally shows such restrictions have little or no effect on sex offender recidivism. And yet, laws limiting housing opportunities for sex offenders potentially exacerbate the limited choices for home placement already facing many ex-convicts. Studies concerning the limitations of where registered sex offenders may live have been widely shown to impose negative consequences on both offenders (Tewksbury 2007; Tewksbury and Lees 2006; Zandbergen and Hart 2006, 2009) and their families (Farkas and Miller 2007; Tewksbury and Levenson 2009). For instance, research confirms that residency restrictions placed on registered sex offenders significantly decreases housing opportunities perhaps to as little as 2 % of all housing stock (Zandbergen and Hart 2006, 2009). As a result of residency restrictions, registered sex offenders are also likely to be concentrated in very dense, socially disorganized communities or in rural locations with limited employment, treatment, and transportation opportunities (Minnesota Department of Corrections 2003; Tewksbury and Mustaine 2006, 2008; Zandbergen and Hart 2006, 2009).

Conclusion

Clearly there are a number of criminal justice policies that are specific to sex offenders and which establish procedures and practices that are unique for offenders convicted of sexual offenses. What may be most interesting about the ways that policies have developed and been rapidly implemented across the country are both that they tend to be in response to particular, individual cases that have generated significant public (e.g., media) attention and that these policies and their outlined practices are largely untested.

As research has started to address sex offender criminal justice policies what is quickly becoming clear is that these policies are not being shown to be very effective, they are expensive and inefficient and in many cases seem to be overreaching and including many offenders who may not need to be responded to in such ways. The future holds many questions for how and why the types of policies discussed above can and should be continued to not. However, at the present time there is little to suggest that such policies will be scaled back. Rather, it seems that criminal justice policies regarding sex offenders are only continuing to be expanded and made more harsh, despite emerging evidence suggesting that they are not necessary effective and are very inefficient.

Related Entries

- ▶ Sex Offender Treatment
- ▶ Sexual Recidivism

Recommended Reading and References

Billings J, Bulges C (2000) Comment: maine's sex offender registration and notification act. Wise or wicked? Maine Law Rev 52:175–259

Brooks AD (1995) Megan's law: the legal issues. Crim Justice Ethics 14(2):12–16

Colorado Department of Public Safety (2004) Report on safety issues raised by living arrangements for and location of sex offenders in the community. Sex Offender Management Board, Denver

Farkas MA, Miller G (2007) Reentry and reintegration: challenges faced by the families of convicted sex offenders. Fed Sentencing Rep 20:88–92

Fitzgerald EA (1990) Chemical castration: MPA treatment of the sexual offender. Am J Crim Law 18(1):1–60

Friedland S (1999) On treatment, punishment, and the civil commitment of sex offenders. Uni Colorado Law Rev 70(1):73–154

Iacono WG, Lykken DT (1997) The validity of the lie-detector: two surveys of scientific opinion. J Appl Psychol 82:426–433

Melella JT, Travin S, Cullen K (1989) Legal and ethical issues in the use of antiandrogens in treating sex offenders. Bull Am Acad Psychiatr Law 17(3):223–232

Meyer WJ, Cole C (1997) Physical and chemical castration of sex offenders: a review. J Offender Rehabil 25:1–16

Meyer WJ, Cole C, Emory E (1992) Depo-Provera treatment for sex offending behavior: an evaluation of outcome. Bull Am Acad Psychiatr Law 20(3):249–259

Minnesota Department of Corrections (2003) Level three sex offenders' residential placement issues. Author, St. Paul

Sample LL, Kadleck C (2008) Sex offender laws: legislators' accounts of the need for policy. Crim Justice Policy Rev 19(1):40–62

Spalding LH (1998) Article: Florida's 1997 chemical castration law. A return to the dark ages. Fla State Uni Law Rev 25:117–139

Tewksbury R (2004) Experiences and attitudes of registered female sex offenders. Fed Probat 68(3):30–33

Tewksbury R (2005) Collateral consequences of sex offender registration. J Contemp Crim Justice 21(1):67–81

Tewksbury R (2007) Exile at home: the unintended collateral consequences of sex offender residency restrictions. Harv Civ Rights – Civ Lib Law Rev 42:531–540

Tewksbury R, Lees MB (2006) Perceptions of sex offender registration: collateral consequences and community experiences. Sociol Spectr 26:309–334

Tewksbury R, Lees MB (2007) Perceptions of punishment: how registered sex offenders view registries. Crime Delin 53(3):380–407

Tewksbury R, Levenson JS (2009) Stress experiences of family members of registered sex offenders. Behav Sci Law 27:611–626

Tewksbury R, Mustaine EE (2006) Where to find sex offenders: an examination of residential locations and neighborhood conditions. Crim Justice Stud 19:61–75

Tewksbury R, Mustaine EE (2008) Where registered sex offenders live: community characteristics and proximity to possible victims. Vict Offender 3:86–98

Tewksbury R, Mustaine EE (2009) Stress and collateral consequences for registered sex offenders. J Public Manag Soc Policy 15:215–239

Uggen C, Manza J, Behrens A (2004) Less than the average citizen: stigma, role transition and civic reintegration of convicted felons. In: Maruna S, Immarigeon R (eds) After crime and punishment: pathways to offender reintegration. Willan Publishing, Devon

Walker JT, Golden JW, VanHouten AC (2001) The geographic link between sex offenders and potential victims: a routine activities approach. Justice Res Policy 3(2):15–33

Wettstein WM (1992) A psychiatric perspective of Washington's sexually violent predator statute. Univ Puget Sound Law Rev 15:597–633

Wilcox DT (2000) Application of the clinical polygraph examination to the assessment, treatment, and monitoring of sex offenders. J Sex Aggress 5(2):134–152

Zandbergen PA, Hart TC (2006) Reducing housing options for convicted sex offenders: investigating the impact of residency restrictions laws using GIS. Justice Res Policy 8(2):1–24

Zandbergen PA, Hart TC (2009) Availability and spatial distribution of affordable housing in Miami – Dade County and implications of residency restriction zones for registered sex offenders. American Civil Liberties Union of Florida, Miami

Zevitz R, Farkas MA (2000a) The impact of sex-offender community notification on probation/parole in Wisconsin. Int J Offender Ther Comp Criminol 44(1):8–21

Zevitz R, Farkas MA (2000b) Sex offender community notification: managing high-risk criminals or exacting further vengeance? Behav Sci Law 18:375–391

Sex Offending and Criminal Mobility

Eric Beauregard
School of Criminology, Simon Fraser University,
Burnaby, BC, Canada

Overview

For more than 60 years now, sex offenders have been the object of scrutinized attention by policy makers and the police (Barnes et al. 2009). However, it is only since 1994 that specific legislations regarding sex offenders have been passed throughout the United States to counter sex offender recidivism and alleviate community safety concerns. The Jacob Wetterling Act requires all states to track sex offenders through the use of registration systems. Such policy forces sex offenders to report to the authorities their home address. This was soon followed by Megan's Law, which requires law enforcement to notify communities when a sex offender is returned to the area (Zevitz 2006). Through the use of different means (e.g., flyers, posters, public meetings, automated phone messages), the public is informed of the identity and the residential address of the sex offender integrating into their community, allowing for an informal network of surveillance. Residence restriction laws constitute the most recent type of policy specifically targeting sex offenders. Such policies restrict sex offenders from living within close proximity (e.g., between 1,000 and 2,500 ft) of places where children are typically present (e.g., schools, parks, playgrounds, bus stops), in order to prevent them from getting access and making direct contact with potential victims (Duwe et al. 2008). What is particular to these policies is the fact that they all share a common focus – that is, *where* the offender lives. Thus, the rationale for such policies is based on the same underlying assumption – that sex offenders are geographically stable, meaning that they do not travel when committing their crimes. But is this really the case? The current paper aims to address three related questions: (1) What do we know about sex offenders' criminal mobility? (2) Is journey to crime an adequate measure of criminal mobility? (3) Are there any other ways to look at criminal mobility of sex offenders?

Criminal Mobility: The Case of the Sex Offender

Most studies interested in the criminal mobility of offenders use the journey to crime as their main measure. These studies generally show that offenders travel a longer distance to commit property offense compared to violent crimes (see for instance Brantingham and Brantingham 1981). When looking at violent crimes in particular, a total of 21 studies were identified that examined crime trips of sex offenders in North American and European cities (Beauregard et al. 2005). For all of these studies, it is interesting to note that the crime trip distance traveled by different types of sex offenders varies between no distance traveled (the offender committing the crime at home) and 40 km. On average, the distance traveled by different types of sex offenders between their home base and the crime location was a little more than 2 miles.

Interestingly, some authors have attempted to further our understanding of the criminal mobility of sex offenders by examining two types of factors influencing their journey to crime: the offender and the offense characteristics. The underlying principle of the relationship between offender characteristics and criminal mobility is that these characteristics influence the cognitive map of individuals. According to Brantingham and Brantingham (1993), cognitive maps are a representation of the awareness of space, which consists of subjective images of an individual's environment that are fundamental in determining the areas where the criminal's offense will be carried out. Furthermore, cognitive maps vary with the characteristics of individuals. For instance, age is one of the characteristics associated with criminal mobility. Most studies of sex offenders that investigate the relationship between age and distance come to the same conclusion as for other types of criminals: Younger men tend to offend nearer to home (Warren et al. 1998). According to these studies, this difference could be attributed to a greater impulsivity in the offense behavior of younger offenders, a greater access to vehicles by older offenders, or simply because of the age-related development of the cognitive map. However, recent studies have not confirmed this age-distance relationship (see Rossmo et al. 2004).

There appears to be a clear relationship between race and the criminal mobility of sex offenders. Findings show that white rapists traveled farther than nonwhite offenders (Canter and Gregory 1994; Warren et al. 1998). Even if this relationship was once again not confirmed by Rossmo et al. (2004), it is hypothesized that this finding may reflect class distinctions or cultural differences in the cognitive mapping of space. Moreover, Gabbor and Gottheil (1984) found that those with a criminal record were substantially more likely to be transient than those without one, suggesting a positive relationship between criminal career and mobility. Others have suggested that sexual fantasy is another factor related to longer sexual crime travel distance. These offenders spend long periods of time prowling for victims, and sometimes record a diary of their movements. In addition, they are willing to travel long distances to commit a crime that will reflect their fantasies (Dietz et al. 1990). Finally, a study showed that psychopaths displayed greater geographic mobility than did nonpsychopaths (Hunter 2004). The psychopath's impulsivity, short-term relationships, unstable employment, and need for stimulation/ proneness to boredom may predispose them to geographic mobility (Cooke 1998). Another explanation suggests that psychopaths frequently move locations, as their tendency to con and exploit others eventually becomes known and they are no longer able to take advantage of people in their surroundings.

As to the offense characteristics, Lebeau (1987a) focused on how the journey to rape varies as a function of the offender's approach method. Results revealed that offenders traveled the shortest mean distance to assault their victims when they illegally entered the victim's residence, suggesting that offenders travel shorter distances when using a method linked to crimes against property. In a different study, Canter and Gregory (1994) found that rapists who offend during the weekend travel farther than those who commit rape during a weekday and that rapists who attack outdoors traveled approximately 2.7 times farther to offend as those who raped indoor (e.g., in a house). Davies and Dale (1995) suggested also that rapists who target victims from a particular area (e.g., prostitutes from a red-light district), who commit sophisticated property offenses during a sexual assault, who spend large amounts of time roaming and using public transportation, and who are familiar with numerous neighborhoods (previous habitation, locations of significant people, current or past workplace locations), travel longer distances to commit their crimes. This was echoed by Warren et al. (1998) who found similar results in relation to the sophistication of the crime. Their results showed that rapists who had more extensive criminal antecedents, who used forced entry, and who burglarized the victim during the assault tended to travel farther. According to the authors, this could reflect a more generalized criminal motivation and a more experienced offender in terms of nonsexual crimes.

The research on the journey to crime of sex offenders has not only permitted to confirm that, as with other violent offenders, sex offenders do not travel long distances to commit their crime, but it has also allowed to examine the factors that could influence their criminal mobility. However, it appears that all the studies have taken for granted that journey to crime was the only and probably the most adequate measure of criminal mobility. Whether or not this is true, none of these studies have questioned the use of the journey to crime to measure their criminal mobility.

Journey to Crime: A Valid Measure of Criminal Mobility?

Previous studies on the criminal mobility of sex offenders show that their journey to crime is mostly constant, being not very far from their home base. This information is not only useful for environmental criminology but also for criminal investigations, more specifically geographic profiling techniques (Rossmo 2000). However, relying solely on the measure of journey to crime to characterize their criminal mobility can be problematic for three main reasons.

First, the studies examining the journey to crime have all used the home location of the offender as the starting point of this crime trip. Although this is congruent with theoretical models such as the routine activity theory (Cohen and Felson 1979), such a measure does not take into account a crime trip that might have originated from the offender's workplace for instance. A rapist could identify and target a victim while coming back from a friend's house. Moreover, as Bernasco (2010) found, offenders are more likely to target former residential areas to commit their crimes if they lived there for a long time. Moreover, as mentioned by Michaud and Morselli (2011), homeless offenders and many street criminals may not have a fixed home address, therefore choosing to base their criminal activities from other social activity locations such as bars or pool halls

(Rengert 1996). Again, as stated by Michaud and Morselli (2011), using the offender's home as the exclusive platform from which crimes are committed is even more surprising given the fact that most people spent around half of their days outside home (see for instance Wikström et al. 2010). Moreover, Pettiway (1995) is the only one to our knowledge who showed through interviews with crack users the true origin of the crack trip. Interestingly, he found that only 26 % of crack-purchasing trips originated from the user's place of residence. The lack of studies taking into consideration other points of origin in the journey to crime might be partly explained by the fact that most of these studies rely exclusively on police data and other anchor points of the offender might not be known.

Second, journey to crime research suggests that the whole criminal event takes place all at the same location. Although this may be the case for property crimes such as burglary, the reality is often different for crimes where the victim is mobile (Beauregard et al. 2010). In crimes such as sexual assaults, the offender may encounter a victim at a certain location, decide to attack her at another location, take her to a different location to commit the assault, and take her to another location to release her, the so-called EAMD classification used in geographic profiling (Rossmo 2000; see also Lebeau 1987b for rape). Michaud and Morselli (2011) explain that the non-consideration of multiple crime sites in journey to crime research is problematic, given that it may distort the estimation of the criminal mobility involved in the process. To illustrate their point, they provide the example of an offender who encounters a woman in a bar, takes her to a hotel, rapes and kills the woman there, to finally go back to his place with the dead body in order to dispose of her body parts in garbage bags. If the journey to crime was calculated with this case, police data would record the address where the body parts were found (offender's residence; distance-to-crime = 0 mi), thus greatly underestimating the criminal mobility that was involved in the process. Such example is not unusual in sex-related crimes. For instance, it has been shown that sexual murderers dispose of the victim's body on average 17.2 miles from the murder scene (Häkkänen et al. 2007).

Third, research on criminal mobility seems to suggest that the concept of mobility can and should only be measured in terms of distance traveled. However, the measure of distance traveled represents only one dimension of the mobility concept – that is, the movements or the mobility performances (Canzler et al. 2008). However, offenders also present other "movements" that may characterize their criminal mobility. Using the same example presented above, an offender may travel from one location to another during the same event. For instance, some sexual murderers decide to move the victim's body to a different location after the murder (Beauregard and Field 2008). Although such finding does not provide a measure of *distance* per se, it nonetheless provides some information on the mobility of the offender during the criminal event. This is not to suggest that the measure of journey to crime is not appropriate to examine criminal behavior and that this research is not important for our understanding of criminal mobility. However, this suggests that there is a need for criminology to look at complementary measures as well.

Alternative Measures of Criminal Mobility in Sex Offending

In addition to the journey to crime measure, a few alternative methods to conceptualize criminal mobility have been used in the research on sex offenders. One such approach has been to classify sex offenders according to their criminal mobility patterns. Although using different labels (see, for instance, Canter and Larkin 1993; Rossmo 1997), these typologies can be grouped under two main categories: the *geographically stable* and the *geographically mobile* offenders (Beauregard et al. 2005). This typological approach is interesting as it presents other characteristics (e.g., offender characteristics, modus operandi) associated with each mobility pattern. However, such approach presents also certain limitations. The majority of these typologies have been identified intuitively, without being tested empirically, which carries problems of validity and reliability. Moreover, the types are mainly descriptive and offer little information to explain the criminal mobility patterns of offenders (Michaud and Beauregard 2010). Also, using typologies to classify criminal mobility patterns presents the same problem as with any other typology: They assume that crime-commission processes are stable instead of being dynamic (Beauregard et al. 2007).

One way to overcome these specific limitations is to conceptualize geographic mobility as the use of multiple locations for the purpose of repetitive sexual contact with the same victim. In their study, Leclerc, Wortley, and Smallbone (2010) set out to examine whether offending differences existed between perpetrators who used multiple locations for sexual contact and those who used a single location for the entire crime-commission process. Overall, the results demonstrate that mobile offenders are more likely to isolate their victims, use violence, involve the victim in several sexual episodes, abuse the victim for over a 1-year period, and make the victim participate and perform sexual behaviors on them during sexual episodes. The authors concluded that by examining mobility of pedophiles from a location angle rather than measures of distance and direction provides a different perspective on the crime-commission process of these offenders (Leclerc et al. 2010).

Following Leclerc et al. (2010) study, Beauregard and Busina (2013) used a similar approach with serial sex offenders. They proposed that criminal mobility can be defined as the number of changes of location during the criminal event. As discussed previously, rape events present different stages – that is, encounter, attack, crime, and victim release – that may be associated with different locations. Although some sex offenders decide to commit all their action at the same location (i.e., stable offender with zero change of location), other mobile offenders may change location up to three times during the same event. As criminal mobility can be interpreted as a purposive action necessary to successfully commit a crime, the aim of their study was to predict the criminal mobility patterns exhibited in serial rape events from situational and modus operandi characteristics. The situational characteristics of the rape events and the modus operandi used in serial sex crimes might explain why some offenders need to be mobile and change location during the criminal event while others do not. Using negative binomial regression, the authors found that events which involve child or adolescent victims, committed during daytime, when the offender did not use pornography prior to crime, and where victim resistance is observed, should display more criminal mobility. Moreover, when the victim is selected, the victim is alone when approached by the offender, and the crime is characterized by sexual penetration and a lack of premeditation are exhibiting more criminal mobility. These results point toward the fact that criminal mobility is a goal-oriented action taken by serial sex offenders in order to complete successfully their crime and to avoid detection and apprehension (Beauregard and Busina 2013).

Rossmo, Lu, and Fang (2011) examined the spatial-temporal patterns of a group of reoffending parolees (many of which were sex offenders) on the Florida Department of Corrections electronic monitoring and global positioning system program. Their travel over a period of at least 8 days, including the offending day, was mapped and analyzed. This allowed analyzing the spatial activity patterns of criminals prior to, during, and after offending. At the aggregate level, results showed that the mean distance traveled 37.8 mi (min = 3.5 mi, max = 79.5 mi), covering an area of 27.2 mi^2. The average time spent traveling was 10.6 h (min = 1.0 h, max = 15.8 h), compared to an average of 12.4 h spent at home (min = 4.5 h, max = 20.9 h). On average, these offenders traveled 4.1 sites during their daily travel (min = 2.0, max = 6.5). When disaggregating these results, the findings are also very interesting, especially to further our knowledge on criminal mobility. Rossmo et al. (2011) described the spatial movements of a sex offender. This offender visited the offense location twice in the week before the crime. His movements started late in the day and went through to late in the night or early in the morning of the next day. The analysis revealed also that his routine began with a trip from his home to an activity site northwest of his residence and that he passed by the crime location for the first time during the study period, but did not stop. Moreover, their analysis showed that the actions of the offender were different when he next visited the site. On that day, the offender passed the crime location again on his way to his routine nighttime activity sites northwest of his home but he made extra turns and stops. The same travel pattern was repeated the day of the offense. As mentioned by Rossmo et al. (2011), "a better understanding on this level can assist practitioners and academics in several ways, such as providing early warning cues for offending, better knowledge of how criminals hunt for their targets, and an enhanced understanding of offender spatial behavior" (p. 39).

Finally, another alternative to the traditional journey to crime measure of criminal mobility has been suggested by Michaud and Beauregard (2010). They conceptualized the criminal mobility of offenders as having two distinct dimensions: the *criminal migration* – which is the effective distance traveled by a mobile offender between the activity spaces in which he commits crimes – and the *criminal nomadism* – which corresponds to the offender propensity to change the activity space in which he commits crimes, either by exploring new activity spaces or by traveling back to former ones. Instead of looking only at one criminal event, Michaud and Beauregard (2010) consider the entire criminal career of their sex offenders. As such, they have analyzed 3,003 criminal events (sex and non-sex crimes) committed by 461 sex offenders. Their findings showed that in terms of criminal migration, only a third of the sex offenders traveled a short distance (<150 km) during their criminal career and approximately one quarter traveled more than 1,000 km. As to the criminal nomadism, one third of the sample changed activity space in at least 70 % of their criminal career. In order to understand what could explain this criminal mobility, they examined different offender characteristics. Their findings showed that sex offenders who are psychopaths, Caucasian, educated, and specialized in sex offending present more nomadism than the sex offenders who are not.

Conclusion

Recent legislations specific to sex offenders – whether it is registration, community notification, or residence restriction – have all focused on one aspect of the crime-commission process: location. The rationale for such a focus is that by knowing where sex offenders live and by forbidding them to reside near certain location, we can prevent sexual recidivism. Although in theory, this all makes sense, all these approaches make the same mistake in assuming that criminals – more specifically sex offenders – are geographically stable and are not likely to travel when contemplating a criminal opportunity. Research shows the opposite. Sex offenders are rational individuals who may decide to travel to a different location either to avoid detection and apprehension, to complete successfully the sexual assault, or both. This suggests that even if a sex offender cannot live near a park because of a residence restriction law, nothing prevents him from leaving his residence, going to the park to encounter potential victims, and take them back to his place to sexually assault them. Studies have already shown that these legislations are ineffective to prevent sexual recidivism (e.g., Duwe et al. 2008). What has been presented here contribute further to this conclusion by showing how criminal mobility is an adaptive response to the criminal situation aiming to decrease the risks of detection and apprehension, while maximizing the chances of successfully completing the crime (e.g., obtaining sexual gratification, being able to perform sexual penetration actions).

Moreover, although journey to crime research has been helpful to the understanding of the geography of crime and to the development of investigative tools such as geographic profiling, this unique measure of criminal mobility is no longer adequate to capture the real spatial movement of offenders when committing crimes. New ways to investigate and conceptualize criminal mobility in sex offenders have been suggested which, it is hoped, will spark some interest in academics involved in the crime and place research for other types of crimes as well. The research on criminal mobility needs to be extended in order to investigate additional aspects that journey to crime has overlooked. In addition to what has been presented, two interesting concepts could further our understanding of criminal mobility. The first concept comes from Frank, Andresen, and Felson's (2012) study who found strong evidence for geodiversity (i.e., variations in the amount of area covered by various crimes depending on the variations of criminal opportunity) in cases of co-offending or co-victimization. The second concept of interest, which appears even more important to consider is directionality (Bernasco and Block 2009). Although it has been under-researched by environmental criminologists, some research has shown that directional knowledge is important for spatial decision making (Frank et al. 2012).

Related Entries

- ▶ Crime Location Choice
- ▶ Criminal Profiling
- ▶ History of Geographic Criminology Part I: Nineteenth Century
- ▶ Offender Decision Making and Behavioral Economics
- ▶ Situational Crime Prevention
- ▶ Theories for Situational and Environmental Crime Prevention

Recommended Reading and References

Barnes JC, Dukes T, Tewksbury R, De Troye TM (2009) Analyzing the impact of a statewide residence restriction law on South Carolina sex offenders. Crim Justice Pol Rev 20:21–43

Beauregard E, Field J (2008) Body disposal patterns of sexual murderers: implications for offender profiling. J Polic Crim Psychol 23:81–89

Beauregard E, Proulx J, Rossmo K (2005) Spatial patterns of sex offenders: theoretical, empirical, and practical issues. Aggress Violent Behav 10:579–603

Beauregard E, Proulx J, Rossmo K, Leclerc B, Allaire J-F (2007) Script analysis of hunting process in serial sex offenders. Crim Justice Behav 34:1069–1084

Beauregard E, Rebocho MF, Rossmo K (2010) Target selection patterns in rape. J Investig Psychol Offender Profil 7:137–152

Beauregard E, Busina I (2013) Journey "during" crime: Predicting criminal mobility patterns in rape. J Interpers Violence. OnlineFirst

Bernasco W, Block R (2009) Where offenders choose to attack: A discrete choice model of robberies in Chicago. Criminology, 47:93–130

Bernasco W (2010) A sentimental journey to crime: effects of residential history on crime location choice. Criminology 48:389–416

Brantingham PJ, Brantingham PL (1981) Environmental criminology. Waveland Press, Prospect Heights

Brantingham PL, Brantingham PJ (1993) Environment, routine and situation: toward a pattern theory of crime. In: Clarke RV, Felson M (eds) Routine activity and rational choice. Transaction, New Brunswick, pp 259–294

Canter D, Gregory A (1994) Identifying the residential location of rapists. J Forensic Sci Soc 34:169–175

Canter D, Larkin P (1993) The environmental range of serial rapists. J Environ Psychol 13:63–71

Canzler W, Kaufmann V, Kesselring S (2008) Tracing mobilities – an introduction. In: Canzler W, Kaufmann V, Kesselring S (eds) Tracing mobilities: towards a cosmopolitan perspective. Ashgate, Aldershot, pp 1–10

Cohen L, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44:588–608

Cooke DJ (1998) Psychopathy across cultures. In: Cooke DJ, Forth AE, Hare RD (eds) Psychopathy: theory, research, and implications for society. Kluwer, Dordrecht, pp 13–45

Davies A, Dale A (1995) Locating the stranger rapist (Special Interest Series: Paper 3). Police Research Group, Home Office Police Department, London

Dietz PE, Hazelwood RR, Warren J (1990) The sexually sadistic criminal and his offences. Bulletin of the American Academy of Psychiatry and Law, 18:163–178

Duwe G, Donnay W, Tewksbury R (2008) Does residential proximity matter? A geographic analysis of sex offense recidivism. Crim Justice Behav 35:484–504

Frank R, Andrrsen MA, Brantingham P (2012) Criminal directionality and the structure of urban form. J Environ Psychol 32:37–42

Frank R, Andresen MA, Felson M (2012) The geodiversity of crime: evidence from British Columbia. Appl Geogr 34:180–188

Gabor T, Gottheil E (1984) Offender characteristics and spatial mobility: an empirical study and some policy implications. Can J Criminol 26:267–281

Häkkänen H, Hurme K, Liukkonen M (2007) Distance patterns and disposal sites in rural area homicides committed in Finland. J Investig Psychol Offender Profil 4:181–197

Hunter SM (2004) Geographical mobility among male psychopaths. Unpublished Honour's thesis, Simon Fraser University, Burnaby, Canada

Lebeau JL (1987a) The journey to rape: geographic distance and the rapist's method of approaching the victim. J Polic Sci Admin 15:129–136

Lebeau JL (1987b) Patterns of stranger and serial rape offending: factors distinguishing apprehended and at large offenders. J Crim Law Criminol 78:309–326

Leclerc B, Wortley R, Smallbone S (2010) Investigating mobility patterns for repetitive sexual contact in adult child sex offending. J Crim Justice 38:648–656

Michaud P, Beauregard E (2010) Criminal nomadism: a neglected dimension of sex offender criminal mobility. Paper presented at the annual meeting of the American Society of Criminology, San Francisco

Michaud P, Morselli C (2011) Should we stay or should we go? A critical review of journey-to-crime research. Submitted for publication

Pettiway LE (1995) Copping crack: the travel behavior of crack users. Justice Quart 12:499–524

Rengert GF (1996) The geography of illegal drugs. Westview Press, Boulder

Rossmo DK (1997) Geographic profiling. In: Jackson JL, Bekerian DA (eds) Offender profiling: theory, research and practice. Wiley, Chichester, pp 159–175

Rossmo DK (2000) Geographic profiling. CRC Press, Boca Raton

Rossmo DK, Davies A, Patrick M (2004) Exploring the geo-demographic relationship between stranger rapists and their offences (Special Interest Series). Policing and Reducing Crime Unit, Home Office, London

Rossmo DK, Lu Y, Fang TB (2011) Spatial-temporal crime paths. In: Andresen MA, Kinney JB (eds) Patterns, prevention, and geometry of crime. Routledge, New York, pp 16–42

Safarik ME, Jarvis JP, Nussbaum KE (2002) Sexual homicide of elderly females: linking offender characteristics to victim and crime scene attributes. J Interpersonal Violence 17:500–525

Warren J, Reboussin R, Hazelwood RR, Cummings A, Gibbs N, Trumbetta S (1998) Crime scene and distance correlates of serial rape. J Quant Criminol 14:35–59

Wikström P-OH, Ceccato V, Hardie B, Treiber K (2010) Activity fields and the dynamics of crime. Advancing knowledge about the role of the environment in crime causation. J Quant Criminol 26:55–87

Zevitz RG (2006) Sex offender community notification: its role in recidivism and offender reintegration. Crim Justice Stud 19:193–208

Sexual Assault and Violence Against Females

► Feminist Theory in the Context of Sexual Violence

Sexual Offender Treatment (SOT)

► Treatment of Sex Offenders

Sexual Recidivism

Brian Francis and Keith Soothill
Department of Maths and Statistics, Lancaster University, Lancaster, UK

Synonyms

Sexual reconviction; Sexual reoffending

Overview

Sexual recidivism is concerned with the reoffending of sexual offenders who have already had contact with the criminal justice system. This entry emphasizes the need to carefully define what is meant by recidivism and discusses some of the pitfalls in this area. It then focuses on recidivism rates, following on with a consideration of risk factors for recidivism, both static and dynamic, before commonly used risk scores are introduced. A discussion on specialization and glances towards possible developments in this area concludes the section.

Fundamentals

This entry is concerned with sexual recidivism, that is, the "persistence" of sexual offending once an offender has been arrested or convicted and addresses the future risks posed to the public by sex offenders after conviction.

Defining Recidivism

To begin, definitional aspects of recidivism and its measurement are considered, including the outcome measure, the type of recidivism, the time horizon, and the definition of what constitutes a sexual offense.

Recidivism: Recidivism is defined as the reoffending of a known offender; however, Falshaw et al. (2003) have noted that "there is little consistency in the way this term is used and the specific behaviour it refers to." They remind the reader that Maltz, in his classic book on recidivism, pointed to the use of nine different indicators of recidivism across a total of 90 studies in the United States, i.e., absconding, arrest, incarceration, parole violation, parole suspension, parole revocation, reoffense, reconviction, and probation violation.

Focusing explicitly on sexual recidivism, Falshaw et al. (2003) distinguished between sexual reconviction, sexual reoffending, and sexual recidivism. Their view was that recidivism is primarily concerned with lapsing into previous patterns of behavior and thus encompassed other forms of potential sexual-offense-related behavior such as loitering outside a school. They defined sexual reconviction as a subsequent conviction for a sexual offense, sexual reoffending as the perpetration of another illegal sexual act (whether caught or not), and sexual recidivism as the commission of a behavior related to a sexual offense, legal or illegal, with a clear sexual motivation. A more common use of the term sexual recidivism in the literature and one which is used in this entry, however, is some kind of official contact with the criminal justice system for a sexual offense, whether it be arrest, charge, or reconviction.

There is an implicit hierarchy within these distinctions. A "sexual reconviction" is the strictest measure as, to count, one needs the endorsement of a court of law. "Sexual recidivism" is a broader measure and may include arrests or charges as well as convictions for a sexual offense. "Sexual offending" has difficult counting and legal implications as it includes illegal sexual acts, whether caught or not. It is crucial to understand the importance of such analytical distinctions as the use of different outcome measures can produce very different results.

Type of Recidivism: Many studies on the recidivism of sexual offenders use as their outcome measure any form of recidivism – a general measure of subsequent contact with the criminal justice system for *any* offense. Others are interested in whether *sexual* recidivism has taken place, whereas yet others are concerned with recidivism for *dangerous* offenses which is usually taken to be sexual or violent recidivism.

Time Horizon for Follow-Up: Some studies have a fixed time horizon or a fixed series of time gates whereby recidivism rates, say, 5 or 10 years after release can be established. Other studies use an average follow-up time – these are less valuable as their results cannot be generalized in systematic reviews.

Nature of the Sample Taken: Some recidivism studies relate to those incarcerated in prison or in a treatment facility; others involve those with community sentences as well. Other aspects are also important – are females as well as males included, and is the sample restricted to a particular age range? Specifically, are juvenile offenders included?

Definition of a Sexual Offense: Normally, legal statutes which categorize sexual offenses (e.g., rape, indecent assault, exhibitionism) are used, but these can differ from country to country and from US state to US state. Most of these differences currently relate to laws against consensual homosexual activity (e.g., Malaysia, Zimbabwe). In addition, the age of sexual consent varies from country to country, varying from 12 (e.g., parts of Mexico) to 20 (Tunisia), and in some countries sexual behavior outside marriage is illegal (e.g., Saudi Arabia, UAE). The age of consent also varies from state to state in the USA and in Australia. There is a further issue relating to long-term or historical recidivism studies in that the definition of what is a sexual offense will change over time. A specific example relates to the laws of sodomy, or "unnatural" sex, which was in effect a legal prohibition in 14 US states until 2003. This makes comparison of recidivism rates across time, across countries, and across states problematic.

There are two further issues relating to whether an offense can be classified as sexual. Some offenses may be classified as a sexual offense but may indeed have no sexual motivation. For example, some might regard "bigamy," which until recently was classified by the Home Office as a sexual offense in England and Wales, as more of a deception than a sexual offense. In contrast, other offenses captured under "theft" (e.g., stealing underwear) could be regarded as more indicative of a sexual than acquisitive motivation but are unlikely to be included. Less well recognized is the fact that most reconviction studies are based on the principal offense committed. Hence, someone convicted of both murder and rape, for example, is recorded for murder as the principal offense, while the offense of rape is masked.

Risk and Prediction: A *risk* is a chance or probability that some event (usually undesirable) will happen in the future. Thus, one can talk about the risk of a sexual offender being reconvicted for a sexual offense in the next 5 years on release from an English prison. Good risk statements should include some element of location (risks for offenders in the Netherlands may well be lower than in the UK) and some indication of a future time horizon (the next year, the next 10 years, etc.) – however, these are sometimes left implicit, leaving the reader to judge what the statement means. A *prediction* usually relates to a particular person and uses the risk measure to make a judgment about an individual. This prediction may, in turn, be acted upon and some judgment made about the individual based on the prediction. Mostly, the assessment of risk in criminology is about potential dangerousness.

The Risk of Sexual and General Recidivism

Both large numbers and a long-term follow-up are thought to be crucial in the quest for a definitive study about recidivism rates. Although systematic reviews suggest that the observed sexual recidivism rates are only 10–15 % after 5 years (Hanson and Bussière 1998), the rates continue to increase gradually with extended follow-up periods. The study by Prentky et al. (1997), using a follow-up to 25 years from a treatment center in Massachusetts, demonstrated that, if they had restricted the follow-up to less than 24 months, they would have missed up to 45 % of new charges. While even with a 5-year follow-up period, they would still miss 30 % of the charges they finally identified. In fact, they examined the cumulative failure rates for new charges for sexual offenses only as well as for new charges of any offense. They used nine time gates, broken down by charge, conviction, and imprisonment. Table 1 shows the recidivism rate for new *sexual* charges.

As Table 1 shows, the conviction rates after 2, 5, 10, and 25 years were 6 %, 11 %, 16 %, and 24 % for rapists, while the comparable figures for

Sexual Recidivism, Table 1 Cumulative sexual recidivism rates over nine time gates

Rate at given time gate									
Type of recidivism	1 year	2 years	3 years	4 years	5 years	10 years	15 years	20 years	25 years
Rapists ($N = 136$)									
Sexual reoffense charge	.09	.12	.15	.17	.19	.26	.31	.36	.39
Sexual offense conviction	.04	.06	.08	.09	.11	.16	.20	.23	.24
Sexual offense conviction with prison disposal	.04	.06	.07	.08	.10	.14	.17	.19	.19
Child molesters ($N = 115$)									
Sexual reoffense charge	.06	.10	.14	.17	.19	.30	.39	.46	.52
Sexual offense conviction	.04	.07	.10	.12	.14	.23	.31	.37	.41
Sexual offense conviction with prison disposal	.04	.07	.09	.11	.13	.21	.28	.33	.37

Source: Prentky et al. (1997)

child molesters were 7 %, 14 %, 23 %, and 41 %. The work of Prentky and his colleagues provides a very clear case in demonstrating the importance of long-term follow-ups.

More recently, the results of Prentky et al. (1997) have been validated by Cann et al. (2004), who followed up all 419 male sexual offenders discharged from prison in 1979 in England and Wales until 2000, giving a follow-up period of 21 years. The sexual reconviction rates after 2, 5, 10, and 20 years were 10 %,16 %, 20 %, and 25 %; the equivalent figures for general reconviction were 34 %, 49 %, 56 %, and 62 %. The sexual conviction rates are higher at the 2-year point than either of Prentky et al.'s subsamples, but after that point, the reconviction rates are very similar.

Risk Factors and Sexual Recidivism

As already stated, the risk of recidivism is the probability or chance that the offender, who has already offended once, will take place again in the future within some time horizon. The risk will depend on characteristics of the offender, and it is therefore important to identify a set of factors or variables which are known to be associated with the risk of recidivism. Five such factors are highlighted below.

Sexual Recidivism and Gender

The extent of female sex recidivism is still under-researched; however, differences between male and female sexual offenders in terms of recidivism have been observed. Freeman and Sandler (2008) took a matched series of 780 female and male sex offenders in New York State and demonstrated that male sex offenders were significantly more likely than female sex offenders to be rearrested for both sexual and nonsexual offenses. A later study by the same authors (Sandler and Freeman 2009) extended this work and took 1,466 female offenders who were then followed up for 5 years. The 5-year rearrest rate for these offenders was a very low 1.8 % for a sexual offense and 26.6 % for any offense. Cortoni et al. (2010) concurred with the general conclusion of low rates for female sexual recidivism and, in a meta-analysis using 10 studies, found that the rates were typically under 3 % with an average follow-up of six and a half years.

Sexual Recidivism and Age

The age variable is also theoretically very relevant, particularly in relation to developmental psychology. For instance, Hanson (2002) identifies three broad factors relevant to sexual offending – deviant sexual interests (motivation), opportunity, and low self-control – and uses these to help explain the variation in the recidivism rates of various offenses. So, he maintains that, for rapists, all three factors should decline with age – "self-control should increase in young adulthood, deviant sexual drives should decrease in late adulthood, and opportunities should gradually decline throughout the life span" (Hanson 2002). Consequently, Hanson is not surprised that most rapists are young and that their recidivism rate steadily decreases with age. In contrast, in explaining that child molesters are generally older than rapists, he points to competing factors influencing recidivism risk during early to middle adulthood with, for instance, self-control probably improving but the opportunities for child molesting may well be increasing.

Thornton (2006) has noted that "sexual offenders released at a younger age tended to be more general criminals while those released at an older age tended to be sexual specialists." But Barbaree et al. (2003), focusing more specifically on sexual recidivism, produce an extra twist. They argue that "if libido [seen as one of the important determinants of sexual aggression] decreases with aging, then it follows that sexual aggression should show similar aging effects." Certainly their results suggested that "offenders released at an older age were less likely to recommit sexual offences and that sexual recidivism decreased as a linear function of age-at-release" (Barbaree et al. 2003).

Following reviews and meta-analyses, the consensus is that there is an inverse relationship between sexual offenders' age at the time of their release from incarceration and their sexual recidivism risk (Thornton 2006; Hanson 2002; Hanson and Bussière 1998). However, Doren (2006) notes some recent challenges to this "iron law" – Doren found a series of "study-specific conclusions . . . that were often mutually exclusive." Reanalysis of existing data showed numerous potential interacting variables, such as participation in treatment, type of risk measure used, type of sexual offender, jurisdiction, and even a different measure of offender age. In other words, to explain the wide disparity of findings among the studies he reviewed, there were perhaps confounding variables in all of the reviewed research. While the age variable is definitely on the agenda, Doren concludes that "we have a lot of work to do before we can say we understand how to consider offender age in sexual recidivism assessments" (p. 156).

Sexual Recidivism of Juveniles

In law and policy, the age divide between juveniles and adults is particularly crucial. However, the activity of juvenile sex offending has only comparatively recently attracted attention and research interest. In fact, over the past decade or so, sexually abusive juveniles/adolescents have increasingly been distinguished from adult offending. Worling and Långström (2003) provide a useful and detailed review of studies indicating criminal recidivism risk with adolescents who have offended sexually. They identify, for example, that their own study on 117 sexual adolescents reported a 30 % sexual reconviction rate after a mean follow-up of 9.5 years. Such rates are higher than the Prentky et al.'s 10-year reconviction rates reported earlier in this entry.

An important counterweight to the concern that juvenile sex offenders are likely to be particularly dangerous is provided by Caldwell (2007). Caldwell compared 249 juvenile sex offenders and 1,780 nonsexual offending delinquents who were released from secured custody and contrasted the sexual charge recidivism figures for juvenile sexual offenders and nonsexual offenders and found nonsignificant differences – 6.8 % compared to 5.7 % after a 5-year follow-up. This result has important implications for policy concerns. Quite simply, most juvenile sex offenders may not be the future threat that many might have expected.

The low rates of sexual recidivism among juveniles and adolescents are certainly noteworthy. Vandiver (2006) confirms how nonsexual offenses predominate in recidivism among juvenile sex offenders. Vandiver focuses on 300 registered male sex offenders who were juveniles at the time of their initial arrest for a sex offense. The series is followed through for 3–6 years after they reached adulthood, and while more than half of the series is arrested at least once for a nonsexual offense during this adult period, only 13 (or 4 %) were rearrested for a sex offense. Similar results are portrayed in Nisbet et al. (2004) showing relatively low rates of detected adult sexual recidivism but high rates of detected nonsexual recidivism, among young men who committed sexual offenses as adolescents. Hence, one is identifying relatively stable patterns of general antisocial conduct in adolescent sex offenders, but continuing aberrant *sexual* behavior is not much in evidence. Curiously, with this study, while age at assessment was found to predict sexual recidivism, rather counterintuitively, it was an older age at assessment that predicted adult sexual offense charges. In fact, "the likelihood of being charged with sexual offences as an adult increased by 60 % with each year increase in age at assessment" (p. 230). This is an important finding as it confirms that very early sexual crime is generally unlikely to be a precursor to persistent sexual offending in adult life.

Sexual Recidivism and Type of Sexual Offending

In general, large numbers of sex offenders are needed when testing the notion that different types of sex offenders have different kinds of outcome. Earlier work such as Prentky et al. (1997) had already hinted that some types of sex offenders were more likely to commit new *sexual* crimes than others, suggesting that child molesters have a substantially higher sexual recidivism rate compared to rapists. Other authors have disagreed with this conclusion. For example, Serin et al. (2001) have indicated that recidivism rates are higher for rapists than child molesters. Hanson (2002) notes that "rapists were younger than child molesters, and the recidivism risk of rapists steadily decreased with age." Also differences in recidivism within the same kind of sex offense have been noted – for example, pedophiliac versus non-pedophiliac child molesters.

Sexual Recidivism and Prior Criminal History

The final risk factor highlighted is that of the offender's prior criminal history. There are many ways of summarizing such information, but interest has been focused on the number of prior convictions, the number of prior sexual convictions, the age of criminal career onset, and the relationship between the victim and the offender of prior sexual offending. In general, such measures have been incorporated into risk scores, and criminal history is discussed more fully below.

Risk Scores and Sexual Recidivism

A sexual recidivism risk score, put simply, identifies the risk of future recidivism of a sex offender based on a set of risk factors. For example, the risk score might classify a typical offender with a set of characteristics into one of five categories – highly unlikely to reoffend, unlikely to reoffend, equally likely to reoffend or not, likely to reoffend, and highly likely to reoffend. A prediction of recidivism, in contrast, is using that risk probability or score (and possibly other information) to make a judgment as to whether one particular person will reoffend. Less commonly, a prediction might estimate how many of a group of offenders will reoffend. Thus, a prediction might be made that a male sex offender in California released from prison who has three prior sexual convictions has a 40 % chance of reoffending with another sexual offense in the next 5 years. Based on this prediction, a decision might be made on the form of post-release supervision needed by the offender.

Actuarial and Clinical Assessment Tools

There are two methods of assessing the risk of an individual for subsequent sexual recidivism. Actuarial measures use summary measures of criminal career data on a large set of offenders together with other information about the age, gender, and circumstance of the offender to estimate a risk score. The offenders are followed up for the required period of time, and each offender is then identified as a recidivist or not. This risk score estimate is made though the application of statistical modeling methods such as logistic regression or other score-building models. The performance of the measure can be assessed by dividing the set of offenders into two – using one part of the data to build the risk score and the second part to assess how well the measure performs.

Clinical risk assessment, on the other hand, takes the judgment of professionals such as psychiatrists, probation officers, or parole board members to make a relatively informal judgment on the likelihood of an offender to reoffend. Tools such as the Violence Risk Score for Sexual Offenders can be used to guide the judgment. Although these professionals will have access to the same information on past criminal career history, they will also take into account a whole set of personal factors such as degree of remorse, demeanor, family support, stable residential status, etc., to determine risk (Milner and Campbell 1995).

Both measures would tend to be used at the start of some process – for example, in presentence reports presented to the court or in considering release from an indeterminate prison sentence. Which of these two approaches appears to give better predictions? An important study by Grove and Meehl took 136 separate studies – a mix of clinical and actuarial studies. In general, the clinical studies had a great deal of extra information available compared to the actuarial studies (Grove and Meehl 1996). Their study came to two important conclusions. Firstly, in studies which compared practitioners, there was little agreement between them. Secondly, despite using less information, the actuarial studies were either equal or superior to clinical risk assessment.

However, there have been criticisms of the actuarial approach (Quinsey et al. 1995). Firstly, it can categorize whole groups of individuals as high risk – even though it is recognized that personal circumstances will mean that some individuals are at low risk within a high-risk group. Furthermore, actuarial measures fail to take account of factors such as the intention of the offender to desist from crime. There are also other criminological arguments which need to be considered – that actuarial risk measures focus on the individual to the exclusion of other causes of crime, notably economic and social deprivation. Nevertheless, actuarial measures are perceived by many as providing a reliable estimate of risk and are increasingly used in daily criminal justice practice. For dangerous offenders, current practice relies increasingly on some combination of a structured clinical judgment and actuarial measures.

"Static" Versus "Dynamic" Factors

The main focus in predicting recidivism risk from its outset has been on historical, or "static," factors. These are factors relating to the prior criminal history of the offender such as age of first conviction; the number of previous sexual and nonsexual convictions; demographic factors, such as age and gender; historical factors in the early life of the offender such as whether both parents were present when the offender was a child; as well as victim characteristics (stranger, familial) and type of offending (child molestation, rape, etc.). Age is regarded as a "static" factor as it is taken at some fixed point in time, such as conviction or date of release from prison or treatment center. As such factors are "static," it is not possible to intervene and to change these to improve outcome.

Dynamic factors, in contrast, are subject to intervention. They can be divided into stable dynamic factors, which change slowly over time (such as job responsibility and attitudes to the opposite sex), and acute dynamic factors, which can vary day by day or hour by hour (such as day to day drinking behavior).

Building Risk Scores

Many risk scores in common use will apply statistical techniques to build a recidivism risk score. Data on a cohort of sex offenders is collected and each offender in the sample is followed up for a fixed period of time – typically 2 or 5 years – although longer follow-up periods can also be used. Typically, official data is used to assess whether the offender has recidivated by looking at arrest or court conviction records. Information on offender characteristics is also collected. These can be obtained from the criminal history of case notes of the offender. For scores involving dynamic factors, in addition, psychometric tests (e.g., for psychopathy) may need to be administered as they often contribute to the recidivism test score. Typically, logistic regression is then used to build a score and to determine the important risk factors. If follow-up times vary, then Cox regression may be used.

Once the score is built, its performance on new samples of offenders is assessed. Performance is usually measured by estimating the receiver operating characteristic curve of the score and calculating the AUC (area under the curve) which is expressed as a proportion or percentage. The AUC can be interpreted as follows: if two offenders are taken at random – one reconvicted and one not reconvicted – then the AUC gives the probability that the reconvicted offender will have a higher risk score than the unconvicted offender. Most risk scores used in assessing sexual recidivism have AUCs of around 0.70 or 70 %.

Risk Scores for Sexual Recidivism

Table 2 presents the sexual recidivism risk scores in common use, together with typical items which make up the score. As can be seen, four of the seven scores presented consist solely of static items, and only three scores attempt in addition to include dynamic factors.

Three measures are highlighted: the Risk Matrix 2000/Sexual (RM2000/S), the Rapid Risk Assessment of Sexual Offense Recidivism (RRASOR), and the Violence Risk Score for Sexual Offenders (VRS-SO).

The RM2000/S score consists of seven static items which relate to previous criminal history

Sexual Recidivism, Table 2 Some commonly used risk scores for assessing sexual recidivism for sex offenders

Measure	Full name	Number of items	Examples of items	Developers	Year
Static-99	Static-99	10 static	Conviction and sexual conviction history, relationship to victim, victim gender, lack of long-term intimate relationship, aged under 25	Hanson and Thornton	2000
RM2000/S	Risk Matrix 2000/Sexual	7 static	Conviction and sexual conviction history, age at release, relationship to victim, victim gender, marital status	Thornton, Mann, Webster, Blud, Travers, Friendship, and Erickson	2003
SORAG	Sex Offender Risk Appraisal Guide	12 static 2 dynamic	Age at offense, criminal history, failure on prior release, marital status, history of alcohol abuse, psychopathy score, phallometric test	Quinsey, Harris, Rice, and Cornier	1998, 2006
RRASOR	Rapid Risk Assessment of Sexual Offense Recidivism	4 static	Prior sexual offenses, age at release, relationship to victim, victim gender	Hanson	1997
MnSOST-R	Minnesota Sex Offender Screening Tool – Revised	17 static	Sexual and nonsexual criminal history, substance abuse, discipline history and sex offender treatment history, age at release	Epperson, Kau, and Hesselton	1998
VRS-SO	Violence Risk Score for Sex Offenders	7 static 17 dynamic	Conviction and sexual conviction history; age at release; age of first conviction; sexual deviance, criminality, treatment responsivity	Wong, Olver, Nicholaichuk, Gordon	2003
J-SOAP	Juvenile Sex Offender Assessment Protocol	23 dynamic	Sexual drive; impulsive, antisocial behavior; clinical; community adjustment	Prentky, Harris, Frizzel, and Righthand	2000

(number of previous court appearances, number of previous court appearances for sexual offenses, any conviction for a sexual offense against a male and against a stranger, any conviction for a noncontact sexual offense) together with measures of age at assessment and marital status. The test yields four risk categories – very high, high, medium, and low. The score is easy to administer and is used extensively by the UK Prison Service. AUC values of 0.75–0.77 have been reported.

RRASOR is an even briefer score and consists solely of four items and is, as its name suggests, easy to administer. The items are the number of prior sexual offenses, age of offender, gender of victim, and relationship to victim. AUC values of between 0.65 and 0.79 have been reported by various authors.

More recently, risk scores involving both static and dynamic factors have started to be introduced. The VRS-SO is one example of this development. The score includes both static items (conviction and sexual conviction history, age at release, age of first conviction) and a range of 17 dynamic items (sexual deviant lifestyle, sexual compulsivity, interpersonal aggression, and cognitive distortions are four of the items). The 17 dynamic items represent three underlying components of sexual deviancy, criminality, and treatment responsivity, together with additional items for intimacy deficits and emotional control. Theoretically, it uses "stages of change" to assess sexual-offending-related attitudes and behaviors and is administered twice, for example, pretreatment and posttreatment. The VRS-SO developers highlight its sensitivity to treatment-related change as well as to other forms of intervention. In terms of predictive validity at the time of development, AUC values ranged from 0.66 to 0.74. Beggs and Grace (2010) recently assessed the score and reported an AUC of 0.80 and also stated that the dynamic items provided additional predictive power after controlling for the static items. However, it needs to be stressed that researchers need training in the use of the instrument and the administration of the instrument is lengthy.

Researchers are divided about the utility of including dynamic measures in a risk score. The developers of VRS-SO argue that dynamic variables do not necessarily have to add to the predictive efficacy of static variables to be useful, stressing that their utility in treatment and in assessing risk change is also important.

Another area of controversy is whether age needs to be included in a risk score which includes dynamic items. The argument is that as most databases used to construct tests are cross-sectional, then age differences in the sample may represent birth cohort effects as well as age, and the distinct contribution of age cannot be determined (Harris and Rice 2007). An additional concern is that the dynamic items include measures of self-control, antisocial traits, and sexual potency, which will decline with age; hence, the inclusion of age in addition to these trait measures will overpredict sexual recidivism. However, Barbaree et al. (2009) have determined that age provides additional explanatory power over antisocial and sexual deviance measures in assessing recidivism.

In general, the choice of score depends on whether assessors have the skill, training, and time to assess offenders on the dynamic items. Scores involving dynamic items are best suited to those offenders who are incarcerated where there may be an interest in possible change, while court-based assessment is probably carried out more efficiently by using a static measure.

Recidivism and Specialization

A separate but related debate to sexual recidivism is that of specialization, which is defined as the tendency to commit the same type of offense. Specialist sexual offenders therefore have a tendency to recidivate with another sexual offense to a greater extent than the average offender.

Lussier (2005) explains that two major hypotheses have been put forward to describe the criminal activity of sexual offenders in adulthood. The first of these states that sexual offenders are specialists who tend to repeat sexual crimes. The second describes sexual offenders as generalists who are versatile in their offending. He goes on to state that the current state of knowledge provides empirical support for both the specialization and the generality hypothesis. A recent study has examined the specialization of sexual offenders both pre- and post-commitment (Harris et al. 2011). They found strong evidence of versatility but also found that those offenders who specialized prerelease were more likely than versatile offenders to specialize in sexual offending on release. They also found that child molesters were more likely to specialize than rapists or incest offenders.

The results suggest that sexual offenders are both versatile and specialist – a hypothesis originally suggested by Soothill et al. (2000). This study compared the criminal records of the 6,097 males convicted in 1973 in England and Wales of one of four offense categories – indecent assault against a female, indecent assault against a male, indecency between males, and unlawful sexual intercourse (USI) with a girl under 13 – over a 32-year observation period looking backward 10 years and forward 22 years, and very different patterns emerged for the four groups (Soothill et al. 2000). It showed that each offense group had very different criminality patterns in terms of their likelihood of being convicted of any (general) offense on another occasion – ranging from 37 % of those convicted of indecency between males to 76 % for those convicted of USI under 16. These sex offenders were shown to differ greatly in terms of general offending behavior – a *higher* proportion of those committing *heterosexual* offenses (i.e., indecent assault on female or USI with a girl under 16) tend to be convicted for violence against the person, property offenses, and criminal damage compared with the other two groups; however, a *lower* proportion of these offenders commit *sexual* offenses on other occasions. In contrast, those committing *homosexual* offenses (i.e., indecent assault on a male or indecency between males) are the mirror image. They are much less likely, compared with other sex offenders, to be convicted of violence or property offenses, while those committing indecent assault on a male are much more likely to be convicted of sexual offenses on other occasions.

At one level, therefore, sexual offenders may or may not specialize within their general criminal career, while at a more specific level offenders may or may not specialize in specific kinds of sex offending within their sexual criminal career. These two levels may act quite independently insofar that an offender could be a specialist at one level and a generalist at another. Miethe et al. (2006) have confirmed this picture by identifying, within low levels of general specialization, that 37 % of serial child molesters and 27 % of serial rapists were identified as specialists, offending with their specific types of sexual crime in the early middle and late third of their criminal careers.

Future Directions in Risk and Sexual Recidivism

This entry concludes with an indication of possible future directions in the study of sexual recidivism, focusing on new data sources, improvements in the methodology of risk score building, and better understanding of the criminal careers of sex offenders.

New Data Sources: One major difficulty with the construction of risk scores is that, in general, they are both constructed and validated on data sets containing a relatively small number of cases. Risk scores are, moreover, often built on specific samples such as sexual offenders in prison. However, the potential is there to construct longitudinal data sets through record linkage which would take complete birth cohorts of sex offenders and follow them up in terms of their conviction or arrest records for long periods of time. For example, one potential source for such data would be the population registers of the Scandinavian countries, which, when linked together, would provide information on police contacts, convictions, employment, income, and demographic variables.

New Methodologies for Assessing Recidivism: Improved data sets will allow more sophistication in the type of models which can be built to explain recidivism. Firstly, new forms of dynamic factors can be introduced. Marital status, for example, is currently included as a static variable, but having dynamic information on changes of marital status, as well as changes in job status and responsibility for children, will provide additional explanatory power which may prove useful. Modeling techniques such as the Cox discrete-time model can help to build prediction scores with time-varying covariates. Secondly, group-based trajectory modeling, which has proved its worth in identifying latent trajectories of offending frequency over the criminal history, and group membership of such classes is starting to be used as a predictor for future recidivism (Lussier et al. 2010). Thirdly, new methods of prediction from the machine-learning and data-mining disciplines (neural networks, support vector machines, random forests) will offer additional sophistication which can deal with the inherent nonlinearity of covariates on recidivism outcome.

Specialization and Recidivism: Most work on specialization and sexual offending has dealt with the whole of the criminal career. However, alternative methods are now available which propose that offenders may favor certain offense types during the short term, largely because of opportunity structures, but that because of changing situations and contexts over the life course, their offending profiles aggregate to versatility over their criminal career as a whole. This notion has not been considered specifically for sex offending. However, it seems possible that offenders may well have particular sex crime preferences in relatively narrow time periods and then transition to other kinds of behavior over time.

Reintegration into the Community: Recently, work has been carried out on how long after sentence or release general offenders become similar to non-offenders in their propensity to commit crime (e.g., Bushway et al. 2011). This idea of redemption and reintegration is important as it can help to determine policy such as how long DNA samples or old criminal records are kept for. This work needs to be extended into sexual offending, and the risk or hazard of future offending for different lengths of arrest or conviction-free periods needs to be determined and compared with non-offenders. Such results should inform the length of time sexual offenders are placed on sex offender registers.

Apart from these specific developments, it is important to realize that societal factors as well as personal and individual factors will affect sexual offending and recidivism. Indeed, it can be argued that changes in society rather than individual upbringing and attitudes have affected the long-term trends in sexual offending and changes in rates of recidivism. The focus of research into sexual recidivism in the last 20 years has focused on the individual; the next 20 years needs to focus more on society and community effects.

Related Entries

- ▶ Criminal Careers
- ▶ Risk Assessment, Classification, and Prediction
- ▶ Risk factors for Adolescent Sexual Offending
- ▶ Sex Offender Treatment
- ▶ Specialization and Sexual Offending

Recommended Reading and References

Barbaree HE, Blanchard R, Langton CM (2003) The development of sexual aggression through the life span: the effect of age on sexual arousal and recidivism among sex offenders. In: Prentky RA, Janus ES, Seto MC (eds) Sexually coercive behavior: understanding and management, vol 989. New York Academy of Sciences, New York, pp 59–71

Barbaree HE, Langton CM, Blanchard R, Cantor JM (2009) Aging versus stable enduring traits as explanatory constructs in sex offender recidivism: partitioning actuarial prediction into conceptually meaningful components. Crim Justice Behav 36(5):443–465

Beggs SM, Grace RC (2010) Assessment of dynamic risk factors: an independent validation study of the violence risk scale: sexual offender version. Sex Abuse J Res Treat 22(2):234–251

Bushway SD, Nieuwbeerta P, Blokland A (2011) The predictive value of criminal background checks: do age and criminal history affect time to redemption? Criminology 49(1):27–60

Caldwell MF (2007) Sexual offense adjudication and sexual recidivism among juvenile offenders. Sex Abuse J Res Treat 19(2):107–113

Cann J, Falshaw L, Friendship C (2004) Sexual offenders discharged from prison in England and Wales: a 21-year reconviction study. Leg Criminol Psychol 9(1):1–10

Cortoni F, Hanson RK, Coache ME (2010) The recidivism rates of female sexual offenders are low: a meta-analysis. Sex Abuse J Res Treat 22(4):387–401

Doren DM (2006) What do we know about the effect of aging on recidivism risk for sexual offenders? Sex Abuse J Res Treat 18(2):137–157

Falshaw L, Bates A, Patel V, Corbett C, Friendship C (2003) Assessing reconviction, reoffending and recidivism in a sample of UK sexual offenders. Leg Criminol Psychol 8(2):207–215

Freeman NJ, Sandler JC (2008) Female and male sex offenders – a comparison of recidivism patterns and risk factors. J Interpers Violence 23(10):1394–1413

Grove WM, Meehl PE (1996) Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: the clinical-statistical controversy. Psychol Public Policy Law 2:293–323

Hanson RK (2002) Recidivism and age – follow-up data from 4,673 sexual offenders. J Interpers Violence 17(10):1046–1062

Hanson RK, Bussière MT (1998) Predicting relapse: a meta-analysis of sexual offender recidivism studies. J Consult Clin Psychol 66(2):348–362

Harris GT, Rice ME (2007) Adjusting actuarial violence risk assessments based on aging or the passage of time. Crim Justice Behav 34:297–313

Harris DA, Knight RA, Smallbone S, Dennison S (2011) Postrelease specialization and versatility in sexual offenders referred for civil commitment. Sex Abuse J Res Treat 23(2):243–259

Lussier P (2005) The criminal activity of sexual offenders in adulthood: revisiting the specialization debate. Sex Abuse J Res Treat 17:269–292

Lussier P, Tzoumakis S, Cale J, Amirault J (2010) Criminal trajectories of adult sexual aggressors and the age effect: examining the dynamic aspect of offending in adulthood. Int Crim Justice Rev 20:147–168

Maltz MD (1984) Recidivism. Academic, Orlando

Miethe TD, Olson J, Mitchell O (2006) Specialization and persistence in the arrest histories of sex offenders: a comparative analysis of alternative measures and offense types. J Res Crime Delinq 43(3):204–229

Milner JS, Campbell JC (1995) Prediction issues for practitioners. In: Campbell JC (ed) Assessing dangerousness: violence by sexual offenders, batterers, and child abusers. Sage, Thousand Oaks

Nisbet IA, Wilson PH, Smallbone SW (2004) A prospective longitudinal study of sexual recidivism among adolescent sex offenders. Sex Abuse J Res Treat 16(3):223–234

Prentky RA, Lee AFS, Knight RA, Cerce D (1997) Recidivism rates among child molesters and rapists: a methodological analysis. Law Hum Behav 21(6):635–659

Quinsey VL, Rice ME, Harris GT (1995) Actuarial prediction of sexual recidivism. J Interpers Violence 10(1):85–105

Sandler JC, Freeman NJ (2009) Female sex offender recidivism: a large-scale empirical analysis. Sex Abuse J Res Treat 21(4):455–473

Serin RC, Mailloux DL, Malcolm PB (2001) Psychopathy, deviant sexual arousal, and recidivism among sexual offenders. J Interpers Violence 16(3):234–246

Soothill K (2010) Sex offender recidivism. Crime Justice 39(1):145–211

Soothill K, Francis B, Sanderson B, Ackerley E (2000) Sex offenders: specialists, generalists – or both? Br J Criminol 40:56–67

Thornton D (2006) Age and sexual recidivism: a variable connection. Sex Abuse J Res Treat 18(2):123–135

Vandiver DM (2006) A prospective analysis of juvenile male sex offenders: characteristics and recidivism rates as adults. J Interpers Violence 21(5):673–688

Worling JR, Langstrom N (2003) Assessment of criminal recidivism risk with adolescents who have offended sexually: a review. Trauma Violence Abuse 4(4):341–362

Sexual Reconviction

▶ Sexual Recidivism

Sexual Reoffending

▶ Sexual Recidivism

Shame in Criminological Theory

Nathan Harris
Regulatory Institutions Network, Australian
National University, Acton, ACT, Australia

Overview

The end of the 1980s marked a rediscovery of shaming in criminology. Criminological theories, and most particularly John Braithwaite's theory of reintegrative shaming, reintroduced the concept of shaming into the criminological debate. At around the same time, criminal justice systems experimented with new interventions, such as restorative justice and shaming penalties, which both draw on notions of shaming, albeit in very different ways. While these developments are principally concerned with actions that are taken by authorities, or broader communities, towards offenders, they also highlight the importance of shame as an emotion. Use of the word shaming implies that the corresponding emotion is in some way critical, but what shame is and what its benefits are for social control are less clearly articulated. The application of shaming within criminal justice is not without controversy, and the merits of shame as an emotion have provoked considerable debate. The concept of shame management and an ethical identity conception of shame have both been proposed to clarify the nature of the emotion as well as its implications for criminology.

The Rediscovery of Shaming in Criminology

The relevance of shaming to the regulation of crime has a long history. Shaming is central to accounts of social control in anthropological descriptions of Polynesian and Asian societies as well as analyses of European criminal justice practices in earlier centuries. Equally well documented is a move away from shaming practices in European-based criminal justice systems during the century and a half until the 1970s. However, in 1980s and 1990s, interest in shaming underwent something of a revival, most particularly through the publication of reintegrative shaming theory, the rise of restorative justice, and an interest in the judicial use of shaming punishments.

Reintegrative Shaming Theory

A focal point for the revival of interest in shaming was publication of John Braithwaite's (1989) book *Crime, Shame and Reintegration*. In this book it is argued that institutions of criminal justice as well as criminological theory have underestimated the importance of social disapproval. Braithwaite argues that to understand crime rates, we need to look beyond official mechanisms, such as penalties that are imposed by criminal justice systems, to the degree to which societies express disapproval of crimes. Strong social norms against criminality, which arise through community activism, are seen as critical to low crime rates because they engender a culture in which crime is unthinkable because people come to see it as abhorrent. The concept that is central to Braithwaite analysis is shaming, which he defines as "... all societal processes of expressing social disapproval which have the intention or effect of invoking remorse in the person being shamed and/or condemnation by others who become aware of the shaming" (Braithwaite 1989, p. 100). An important characteristic of this definition is that it does not limit itself to demeaning or humiliating forms of disapproval but seeks to encompass the full spectrum of ways in which disapproval might be expressed.

The fundamental distinction the theory makes is between stigmatization and reintegration. Stigmatization occurs when disapproval is directed at the person as well as at the offensive behavior, when the person is not treated with respect, when there is no ceremony to decertify the individual's deviant status, and where deviance is allowed to become a master status trait. As with labeling theories, it is predicted that stigmatization of offenders leads to greater re-offending. Being charged with a crime, found guilty of it in a court, and then sanctioned imposes a deviant identity on an individual because it ceremonially changes the position of the person within society and has important social implications, such as reduced employment opportunities. This critique of criminal justice asserts that once imposed, a deviant identity becomes a self-fulfilling prophecy and is almost irreversible: marginalization reduces the individual's access to legitimate opportunities while increasing perceptions of injustice and the attractiveness of supportive subcultures.

However, in identifying reintegrative shaming as an alternative, Braithwaite diverges from the labeling tradition by rejecting the idea that stigmatization is an inevitable product of social disapproval. Reintegration can be seen to have occurred when shaming is respectful, distinguishes between the person and their actions, concludes with forgiveness or decertification of deviance, and does not allow them to take on a negative master status trait. One context in which this often occurs is in family life and the disciplining of children, where research shows that authoritative approaches are more effective than either permissiveness or authoritarianism. Another example Braithwaite cites is Japanese society, which is both high in shaming and high in reintegrative traditions and which has a remarkably low crime rate.

In arguing for the positive effects of reintegrative shaming, Braithwaite highlights two mechanisms. One of these is that reintegrative shaming is an effective deterrent, particularly when it comes from those who the individual is close to, because it poses a threat to relationships that are valued. Yet, reintegrative shaming also transcends the rational actor model of deterrence. The second mechanism, which Braithwaite suggests is more important, is that reintegrative shaming communicates that certain behaviors are morally wrong and thus builds internalized controls or conscience. Braithwaite (1989, p. 72) argues:

> Shaming is more pregnant with symbolic content than punishment. Punishment is a denial of confidence in the morality of the offender by reducing norm compliance to a crude cost-benefit calculation; shaming can be a reaffirmation of the morality of the offender by expressing personal disappointment that the offender should do something so out of character.

Restorative Justice

Shaming has also been an important concept in the development of restorative justice programs, which exploded in popularity during the 1990s, and is now found in criminal justice, child protection, school, and prison systems in many parts of the world. Restorative justice is an alternative to the criminal justice system that redefines the goals of justice as well as the way in which it is carried out. A defining principle of restorative justice is that an offense creates an obligation for offenders to repair the harm that has been caused (Zehr 1990). Unlike the principles of traditional justice that emphasize the importance of consistent and proportional punishment, the aim of restoration focuses attention on apology, reparation, and reconciliation. While Braithwaite's concept of shaming can be applied to many different kinds of interventions, restorative justice quickly came to be seen as the principle way of implementing reintegrative shaming. The broad goals of restorative justice as well as the practices associated with it are consistent with reintegrating offenders, and there is evidence that offenders perceive it as more reintegrative.

The dynamics of restorative justice interventions, such as family group conferences, victim offender mediation, or healing circles, are rich contexts for shaming. Family group conferences, for example, involve semiformal meetings between the offender(s), people who are close to them, the victim(s), and their supporters. The focus of a conference is on finding out what happened and how the incident has affected all of the parties as well as coming to an agreement about what needs to be done to repair the harms that are identified. As a consequence, they involve communities in the kinds of conversations about the negative consequences of crime that Braithwaite argues are critical to developing individual conscience and commitment to the law. Empirical observations suggest that shame dynamics do play an important role in conferences and that well-run programs have the potential to assist in resolving these feelings (Retzinger and Scheff 1996).

Judicial Shaming

Finally, the explicit use of shaming by courts has also seen the rise of "shaming" practices that are completely contrary to the restorative approaches discussed above. Recent examples have occurred, particularly in American criminal justice, where shaming has been used in the court system as a deterrent or punishment for convicted offenders. Offenders have been ordered to complete "shame sentences" relevant to the crime they commit instead of spending time in jail. Shoplifters have been ordered to stand out the front of shops holding signs declaring that they stole, drink drivers are ordered to attach "DUI" stickers to their cars, while those convicted of soliciting sex are ordered to sweep the streets. An explicit aim of this kind of shaming is to humiliate offenders (see Kahan 1996). The significance that feelings of shame hold as a deterrent is exploited by these approaches, but the shaming they impart is explicitly stigmatizing. Another area of American criminal justice which reputedly incorporates shaming are boot camps where offenders are subject to military-style discipline but are also publicly confronted with their offense.

Concerns About Shaming

While awareness of shaming has increased, so too have concerns about the explicit use of shaming to control or respond to crime. Shaming punishments, in particular, have been seen by some as a regressive step which demeans the dignity of offenders while failing to protect basic human rights or allow for rehabilitation. Massaro (1997) argues that this "modern" kind of shaming is one that outcasts certain segments of society in a way that does not protect the individual and undermines the dignity of the whole community. In addition to arguing against the decency of this approach, she argues that the complexity of the emotion of shame is such that courts are ill-equipped to employ shaming and that the effect on offenders would be difficult to predict. Martha Nussbaum (2004) identifies five arguments in the literature against the use of shaming punishments: that they are an offense against human dignity, that they are a form of mob justice, that they are unreliable, that they don't hold the deterrent potential that they are supposed to, and that they are potentially net widening.

While it is not surprising that questions have been raised about these overt forms of humiliation, the appropriateness of shaming within more reintegrative forums such as restorative justice has also been questioned. Maxwell and Morris (2002) and others have argued that overt disapproval is not an aim of restorative practices, suggesting instead that they are oriented towards exploring the consequences that an offense has on its victims, with the aim of provoking empathy. They argue that shaming is a dangerous proposition in restorative justice because even with the best of intentions, shaming might be interpreted by offenders as stigmatizing. Shaming a young offender may exacerbate problems rather than prevent re-offending, particularly if offenses have been committed as a consequence of low self-esteem, which has occurred as a consequence of an absence of emotional support or a difficult past.

This critique of shaming is based in part on doubt as to whether shame is a positive emotion for offenders to feel. A number of scholars have argued that the more important mechanism in restorative justice is the eliciting of remorse, which occurs as a consequence of the offender coming to understand the impact that their actions had on the victims. Empathy, as understood by Maxwell and Morris, aids in this process. Shame, on the other hand, is said to be a dangerous emotion to invoke in offenders because it is a threat to the offender's sense of self-worth and is potentially destructive. These questions reflect a broader debate about the virtues of shame as an emotion, in which there is a clear division between scholars who are pessimistic about the role the emotion plays and those who are more optimistic.

Conceptions of Shame in Criminology

As just illustrated, various ideas about shaming, both positive and negative, are based on assumptions, often implicit, about the nature of the emotion that shaming invokes. This raises the following questions: what is shame, and what are its characteristics? While it is not possible to provide a neat typology because of the disparate manner in which theoretical approaches to the emotion have advanced, three broad characterizations of shame have been identified in the literature (Harris 2001). Each of these characterizations is reflected within criminological research.

Shame as a Social Threat

The first of these conceptions of shame conceives of the emotion as a response to social threat, which is precipitated by the individual's perception that they have been rejected or disapproved of in some way. This conception of shame is apparent in early anthropological perspectives which describe shame cultures as those that rely for social control on the sensitivity of individuals to negative perceptions of others, rather than through the development of conscience. This idea has been elaborated in various ways in contemporary research. While these approaches have varied in their explanations of why people are sensitive to social evaluation, they all emphasize the need to be accepted by others either because the need to have strong personal ties is a basic human motive or because there is an evolutionary need to maintain status or because shame is related to the person's perception of his or her own self-worth. An important characteristic of this conception is that it describes shame as exterior to individuals and as constraining. The individual feels shame as a result of another's decision to reject. If others do not reject in the face of the same actions, no shame is felt. Shame, or the fear of shame, is described as a powerful motivation for the individual to continually monitor and work on personal relationships and to comply with social expectations at a broader level.

As a consequence a fear of shame has been seen as a strong motivator for law-abiding behavior, and this assumption is explicitly made in early anthropological research on shame cultures. More recently, criminologists have drawn on this understanding of shame to argue that informal social sanctions represent a significant deterrent to crime (e.g., Grasmick and Bursik 1990). A number of empirical studies, which place shame within a rational choice perspective, suggest that expectations of feeling shame are associated with lower self-reported projections of offending and in some cases that the effect is comparable with, or greater than, official sanctions.

Shame as Personal Failure

A second way in which shame is described in the literature is as a response to perceptions of personal failure. This is based upon the proposition that shame occurs when an individual perceives that they have failed to live up to an ideal or standard that they uphold and that the consequence of this is the perception that the "whole" self is a failure. This proposition has been explained using a number of theoretical frameworks including psychoanalysis, attribution theory, and affect theory. Shame is contrasted with guilt because the focus of attention is the self rather than an act or omission. Unlike the social threat conception described above, perceptions of failure are not necessarily prompted by disapproval, but can occur in isolation and in relation to personal ideals.

Research that explores the impact of a disposition to feel shame as defined by personal failure has been applied to criminology. June Tangney and her colleagues, in particular, have argued that a disposition to feel shame is far less adaptive than a disposition to feel guilt because shame involves an overwhelming negative evaluation of the self that prevents individuals from responding positively. An extensive program of research shows that individuals who are shame prone are more likely to feel anger and hostility, are less likely to feel empathy for others, and are more likely to suffer from psychopathology (see Tangney and Dearing 2002). When applied to inmate populations, shame proneness is correlated with substance abuse and does not seem to play the same protective role that guilt proneness does (Tangney et al. 2011).

Shame as Ethical Threat

The third conception of shame cuts across these two literatures by incorporating the notion that shame occurs when wrongdoing is recognized by the individual and their community. Shame in this view is connected, unlike embarrassment, with serious transgression as well as the idea of fault. The individual feels shame for having intentionally committed a wrong. This is implicit in William's (1993) description of the precondition for shame being one in which a respected other, defined in ethical terms, would think badly of us. Taylor (1985) argues that shame is tied to the loss of self-respect, which defines what the individual feels is tolerable and what is not. Thus, unlike the social threat conception, shame is seen as occurring in response to the violation of internalized values, and the discomfort associated with the emotion concerns perceptions of the self. However, in emphasizing the ethical character of the emotion, this approach acknowledges the degree to which individual rely on others in forming their beliefs about what is right and wrong. Judgments about what is shameful are not externally imposed or arrived at in isolation, but they are socially negotiated.

This conception of shame is most clearly represented in criminological theory in Braithwaite's theory of reintegrative shaming. Reintegrative shaming according to Braithwaite is "conceived as a tool to allure and inveigle the citizen to attend to the moral claims of the criminal law, to coax and caress compliance, to reason and remonstrate with him over the harmfulness of his conduct" (Braithwaite 1989, p. 9). This suggests that shaming is important because of its educative value in developing or reinforcing beliefs about what is wrong. While the theory suggests that shaming can have a deterrent effect, as an informal sanction that threatens the loss of respect by valued other, this is considered secondary to its moralizing qualities. Braithwaite argues that the primary reason individuals do not commit crime is because they have commitments to shared moral norms and social institutions. Punishment is irrelevant to most people because committing serious crime is simply unthinkable to them. The socialization of children in families and schools about moral norms leads to a broad consensus about what acts should be crimes. In social contexts where there is a broad consensus that particular behavior is wrong and individuals are interdependent because of strong communities, those behaviors will be shameful. Individuals will feel ashamed for violating these values because, consistent with an ethical conception of shame, they have subscribed to these values.

Another approach that has likewise focused on shame's moral qualities is Wikström's (2004) situational action theory of crime, which argues that the emotion is a protective factor in preventing offending. Shame in this framework reflects the individual's commitment to do the right thing, which in turn influences their perception of the choices available in a given context. A number of studies have shown that juveniles who report that they would feel shame in front of others (e.g., friends) if they committed a crime also reported lower levels of delinquency.

Shame Management: The Different Faces of Shame

Evidence that feeling bad about one's actions can have both positive and negative consequences has turned attention to understanding why shame is a constructive emotion in some situations but is counterproductive in others. Why do we hope that some individuals feel shame for offenses they commit, yet also experience unease at the idea of imposing shame within criminal justice? A long tradition of research on shame emotions has explored variation in how individuals experience the emotion, and this, like more contemporary research on dispositions, has recently been drawn on to explore the notion of shame management (Ahmed et al. 2001). This theoretical perspective suggests that when confronted with feeling ashamed for their actions, individuals can manage or respond to the emotion in different ways and that this has important implication for criminal justice institutions.

Evidence of differences in shame experiences was first captured in the seminal work of psychiatrist Helen Block Lewis (1971). In her research with patients, Lewis identified three different forms of shame. The first, "acknowledged shame," involves the recognition that one feels shame and awareness of the feeling associated with it. "Overt-unidentified" shame describes the experience of feeling the negative emotion associated with shame but not recognizing it as shame and thus mislabeling it. "Bypassed" shame involves an awareness that an event may be shameful and doubt about how others see the self, but the emotion is bypassed leaving the person with "… an insoluble, plaguing dilemma of guilt thought which will not be solved" (Lewis 1971, p. 134).

One of the important findings from this work for understanding the implications of shame is that unacknowledged forms of shame are associated with feeling of anger and hostility towards others. Scheff and Retzinger (1991) extended Lewis's analysis by arguing that shame is a signal that the bond between the individual and others is threatened. When feelings of hurt associated with rejection are not acknowledged by the individual, as is the case in unacknowledged forms of shame (bypassed and overt-unidentified), then this emotion becomes redirected as anger towards the self and others. According to Scheff and Retzinger, this is the cause of humiliated fury and helps to explains not just individual anger but also conflict between nations. In both Lewis's and Scheff's accounts of shame, it is evident that when shame is not acknowledged, it manifests itself in an unhealthy reaction.

Eliza Ahmed and her colleagues (2001) have described the various manifestations of shame through the concept of shame management. This captures the notion that when confronted with a shame-inducing situation, individuals can manage the negative feelings in a variety of ways and that this is influenced by both individual characteristics and the social context. Acknowledged shame occurs when the individual accepts that they are responsible and thus acknowledges the emotion. It is argued that when shame is acknowledged, the person is more likely to make amends, feels less anger towards others, and is more likely to discharge the negative feelings. In contrast, unacknowledged shame, which Ahmed describes as displaced shame, occurs when the person does not accept that they are responsible. Failure to resolve the emotion, because of the tension between the disapproval of others and this denial of responsibility, results in shame being displaced into anger towards others.

There is growing empirical evidence that shame management predicts both bullying and criminal behaviors. Ahmed's own research in Australia and Bangladesh shows that children who are bullies are more likely to displace shame compared to children who haven't bullied, who are more likely to acknowledge shame feelings (Ahmed and Braithwaite 2006; Ahmed et al. 2001). These results have been partially supported in a study by Ttofi and Farrington (2008) which showed that displaced shame was a predictor of bullying, but did not demonstrate the expected relationship between shame acknowledgement and lower bullying. A similar result was found by Murphy and Harris (2007) in the context of white-collar crime. In this study, shame displacement predicted recidivism, but the relationship between shame acknowledgement and recidivism was mediated through a measure of remorse.

This research on shame management has significant implications for reintegrative shaming theory and has prompted a revision of the theory (Ahmed et al. 2001). While the revision does not alter the theory's prediction that reintegrative shaming reduces offending (while stigmatic shaming increases offending), it does clarify why this is the case as well as the role that shame plays. The original formulation of the theory implies that the benefit of reintegrative shaming is that it leads to greater feelings of shame in offenders. However, the implication of shame management is that reintegrative shaming reduces offending in most cases because it allows offenders to manage feeling of shame more constructively. Reintegration is more likely to result in offenders acknowledging shame, feeling remorseful for what has happened, and making amends. Stigmatization, on the other hand, is more likely to result in offenders displacing shame and feeling anger towards others. Thus it would seem, somewhat ironically, that the benefit of reintegrative shaming is that it allows offenders to resolve and diminish any shame that they feel.

The implication for criminal justice procedures that have drawn on reintegrative shaming theory, like restorative justice, is that the kinds of shaming or disapproval that are effective in those setting do not involve explicit attempts to shame offenders (Retzinger and Scheff 1996). Overt forms of disapproval, such as a focus on participants expressing their dislike of the behavior, may undermine the offender's ability to acknowledge and resolve feelings of shame. This may be exacerbated in those situations where it is culturally inappropriate to overtly disapprove of another or where offenders have already acknowledged wrongdoing. Processes like family group conferences instead focus on discussion of the consequences of the offenses for all parties. By focusing on how people have been hurt, they avoid stigmatization of the offender and allow them to express remorse and to make amends. Both of these behaviors, as well as acknowledgement and forgiveness by others, are important mechanisms for resolving shame.

An Ethical-Identity Conception of Shame

A critique of the empirical research on shame is that it fails to adequately explain the complex relationship between the individual and the social contexts in which shame occurs, either conceptualizing shame as a response to values that are extrinsic to the person (social threat conception) or having little to say about the social context at all (personal failure conception). Neither of these conceptions adequately accounts for repeated observations that shame is both intimately tied to identity and sensitive to disapproval of others.

To better explain the social context in which shame occurs, as reflected in the empirical research conducted on shame management, an ethical-identity theory of shame has been proposed, which draws on both the ethical conception of shame discussed earlier as well as insights from social psychology (Harris 2011).

The central claim of this approach is that the defining characteristic of shame is that it occurs when we experience a threat to our ethical identity. The precondition for this to occur is an awareness that we violated an ethical value that we subscribe to: we realize that we have behaved in a way that we feel is wrong or at least have significant doubts about how acceptable our behavior was. Even though shame is experienced in reference to internalized values, it is also sensitive to the opinions of others, firstly in making us aware that we might have violated a value that is important to our sense of self, but also because the opinions of others (those whose views we respect) contribute to our interpretation of our behavior.

Research on social influence and conformity suggests that a simple dichotomy between our own and others' values is too crude. A long history of research in social science demonstrates that values, attitudes, and beliefs held by individuals are influenced by others (Turner et al. 1987). Others' opinions are important because, as is illustrated by social identity theory, our values, attitudes, and beliefs, all part of our identities, are often shared with others. We expect to agree with those people whom we see as similar to ourselves, and it is disconcerting when we do not. The reason why disapproval results in shame is because it acts as a form of social validation, either reinforcing the belief that what we did was shameful or undermining the assumption that it was not.

Shame is linked to identity because violating one of our values only leads to shame if it undermines our sense of who we are. This threat to our ethical identity occurs because there is a reciprocal relationship between our values and who we think we are. Holding certain values is at the heart of personal or social identities because identities are defined in large part by sets of beliefs (e.g., "being nurturing" would be an important value if "mother" was an important identity). It follows that when we become aware that we have acted contrary to our values, our identity is called into question. The painful feelings of self-awareness, anger at ourself, and confusion that are associated with shame occur because the contradiction between our values and our behavior cannot be easily reconciled.

An important characteristic of shame is that it motivates us to resolve this dissonance between our actions and how we think of ourselves, and we are able to do so in a variety of ways. We can resolve the threat to our ethical identity by diminishing the significance of our behavior (either by concluding that there was a good excuse or reason for what we did or by making up for it in some way); we can respond by perceiving ourselves as having an alternative identity that is consistent with our behavior and in doing so rejecting any disapproval, or we may perceive ourselves as defective because we have failed to live up to our values.

It is hypothesized that the way in which we resolve this threat to ethical identity will also be influenced by the social validation we receive from others. This explains why individuals react differently to reintegrative shaming and to stigmatizing shaming. Disapproval of our behavior that is reintegrative, when it comes from those whom we respect, can reaffirm a positive ethical identity and thus encourage us to apologize and repair what we've done so as put our misdeeds within a larger positive story about ourself (Maruna 2001). Stigmatization makes this less likely because the message from others is that there is something defective about who we are. In this situation we are more likely to defend our ethical identity by deciding we had a valid reason for our behavior, we might feel very bad about ourself, or we might even decide that we have fundamentally different values to those who disapprove of us and embrace a deviant identity.

Conclusion

An important issue that has been at the center of emerging research on shame is whether it should be understood as a productive and useful emotion that allows offenders to reconsider their behavior or whether it is a dangerous and unhelpful emotion that may even promote greater offending. This question has been complicated by the various notions about what shame is and the various ways in which it has been incorporated into criminological research. However, it does seem clear that the answer is unlikely to be simply one or the other. This is unsurprising once we accept that shame is integrally connected to our belief systems, to our identities, and to our relationships with those around us. Research on shame management and an ethical-identity conception of the emotion of shame have sought to delineate the various ways in which individuals can respond to shame and have examined the consequences of these responses for both the individual and society.

Recommended Reading and References

Ahmed E, Braithwaite V (2006) Forgiveness, reconciliation, and shame: three key variables in reducing school bullying. J Soc Issue 62(2):347

Ahmed E, Harris N, Braithwaite J, Braithwaite V (2001) Shame management through reintegration, Cambridge studies in criminology. Cambridge University Press, Melbourne

Braithwaite J (1989) Crime, shame and reintegration. Cambridge University Press, Cambridge, UK

Grasmick HG, Bursik RJ (1990) Conscience, significant others, and rational choice: extending the deterrence model. Law Soc Rev 24(3):837–861

Harris N (2001) Shaming and shame: Regulating shame drink-driving. In Ahmed E, Harris N, Braithwaite J, Braithwaite V (eds) Shame management through reintegration. Cambridge University Press, Cambridge

Harris N (2011) Shame, ethical identity and justice interventions: lessons from research on the psychology of influence. In: Karstedt S, Loader I, Strang H (eds) Emotions, crime and justice. Hart Publishing, Oxford, pp 193–209

Kahan DM (1996) What do alternative sanctions mean? Univ Chicago Law Rev 63:591–653

Lewis HB (1971) Shame and guilt in neurosis. International Universities Press, New York

Maruna S (2001) Making good: how ex-convicts reform and rebuild their lives. American Psychological Association, Washington, DC

Massaro TM (1997) The meanings of shame: Implications for legal reform. Psychol Public Policy Law 3(4):645–704

Maxwell G, Morris A (2002) The role of shame, guilt, and remorse in restorative justice processes for young people. In: Weitekamp E, Kerner H-J (eds) Restorative justice: theoretical foundations. Willan Publishing, Cullompton, pp 267–284

Muphy K, Harris N (2007) Shaming, shame and recidivism: a test of reintegrative shaming theory in the white-collar crime context. Br J Criminol 47(6):900–917

Nussbaum MC (2004) Hiding from humanity: disgust, shame, and the law. Princeton University Press, Princeton

Retzinger SM, Scheff TJ (1996) Strategy for community conferences: emotions and social bonds. In: Galaway B, Hudson J (eds) Restorative justice: international perspectives. Criminal Justice Press, New York

Scheff TJ, Retzinger SM (1991) Emotions and violence: shame and rage in destructive conflicts, Lexington books series on social theory. Lexington Books/ D. C. Heath and Company, Lexington

Tangney JP, Dearing RL (2002) Shame and guilt. The Guilford Press, New York

Tangney JP, Stuewig J, Mashek D, Hastings M (2011) Assessing jail inmates' proneness to shame and guilt. Crim Justice Behav 38(7):710–734. doi:10.1177/0093854811405762

Taylor G (1985) Pride, shame and guilt: emotions of self-assessment. Oxford University Press, Oxford

Ttofi MM, Farrington DP (2008) Reintegrative shaming theory, moral emotions and bullying. Aggress Behav 34(4):352–368. doi:10.1002/ab.20257

Turner JC, Hogg MA, Oakes PJ, Reicher SD, Wetherell MS (1987) Rediscovering the social group: a self-categorization theory. Basil Blackwell, New York

Wikström P-O (2004) Crime as alternative: towards a cross-level situational action theory of crime causation. In: McCord J (ed) Beyond empiricism. Transaction Publishers, New Brunswick

Williams B (1993) Shame and necessity. University of California Press, Berkeley

Zehr H (1990) Changing lenses: a new focus for criminal justice. Herald Press, Scottsdale

Shaming in Asian Societies

Hong Lu, Wyatt Merritt and Mari Sakiyama
University of Nevada, Las Vegas, Las Vegas,
NV, USA

Synonyms

Restorative justice

Overview

Shaming plays a critical role in crime and punishment. Within the context of criminological theories, there are two different styles of shaming: stigmatizing and reintegrative. Shaming that stigmatizes is known as labeling. Through labeling theory, it has been widely recognized within criminological circles that having a negative criminal label (a stigma) contributes to the formation of a deviant self-image, which pushes offenders into criminal subcultures, facilitates the internalization of the deviant identity, and leads to secondary deviance (Becker 1973; Lemert 1951). In contrast, reintegrative shaming combines punishment with compassion and aims to rehabilitate and reintegrate individual offenders and restore broken relationships. While both types of approaches to crime and punishment have been practiced throughout human history, only relatively recently (since the 1980s) has restorative justice and reintegrative shaming caught the attention of scholars and policy makers in Western countries such as in the Unites States.

This entry first defines and describes shaming and its different forms. It then examines unique social conditions that are conducive to reintegrative shaming. Research has shown that unique social conditions, such as a high interdependency among individuals and communitarianism within a society, facilitate the use of reintegrative shaming (Braithwaite 1989). Asian countries, such as China and Japan, have been widely regarded as communitarian societies, and shaming has played an important role both in informal and formal social settings. To help illustrate the characteristics and nature of reintegrative shaming practiced in Asia, crime prevention and control programs, such as *bang-jiao* in China and *bizai shobun* in Japan, are assessed within the context of the individualism and communitarianism debate.

Shame and Reintegration

Within the context of criminology, shame refers to the disapproval of deviant acts and/or criminal wrongdoings by others and often in public. Shaming can take different forms. It can be as subtle as a frown, a slight shake of the head, or gossip. It can also be as direct as a verbal confrontation, a media broadcast, or an official pronouncement by a judge (Braithwaite 1989). Shame serves as an effective social control on two levels: (1) shame deters future criminal behavior through internally building conscience (guilt induction) and (2) shame also deters future criminal behavior by public disapproval through external referents (e.g., family members, teachers, neighbors, police officers) that set boundaries and reaffirms rules.

Developmental psychologists sometimes make the distinction between shame and guilt by depicting shame as a reaction to criticisms by others such as parents and neighbors, whereas guilt is induced internally after a wrongdoing (Braithwaite 1989). Others regard shaming and guilt induction as inextricably part of the same social process. They both imply certain moral expectations of the individual and the community.

Scholars argue that a culture that has a high set of moral expectations and reinforces these moral expectations through shaming will have a more effective social control than a culture that seeks control through strict law and punishment (Braithwaite 1989). This is because shaming implies that citizens have the responsibility to express public disapproval of behaviors that harm the community. The potency of this form of self-policing lies in its swiftness (in time) and intimacy (both in physical distance and social distance).

A new form of shaming – online shaming – takes shaming to the virtual world. Witnesses of an immoral/deviant incident, often out of frustration and outrage, may take pictures and post them online. For example, pictures of a South Korean woman who refused to clean up after her dog defecated on a train were posted online. This resulted in her being “shamed” by bloggers from all over the world. This newly emerged shaming obviously grew from modern technology but also out of the growing diversity, mobility, and fear of disorder and crime shared in modern societies. Online shaming essentially serves the function of peer surveillance, or lateral surveillance, and may help deter deviant behaviors and reaffirm moral boundaries (Skoric et al. 2010).

There are two types of shaming: reintegrative and stigmatizing. A reintegrative type of shaming has two essential components: (1) shame (public expressions of community disapproval) and (2) reintegration (efforts to reaccept the offender into the community). Community disapproval can range from a mild rebuke to a strong disapproval within a formal ceremony. Also, efforts at reintegration can range from a simple smile expressing forgiveness and love to a formal ceremony to decertify the offender as a deviant. Shaming is more potent when expressed and administered by significant others – parents, friends, and neighbors. This is because significant others tended to have a continuing and meaningful relationship with the deviants; thus, their opinions bear more importance to the deviants than would a stranger's or an institution's (i.e., the police, court officials). In addition, shaming is more effective when there is a goal of reintegration and restoration, rather than stigmatization and rejection. The best place to witness reintegrative shaming is the loving family. Within a loving family, disapproval is expressed while bonds of respect are maintained. Condemnation is made toward the deviant act, not the deviant family member. More important is the forgiveness and genuine gesture of reacceptance that leads to reintegration back into the family (Braithwaite 1989).

The stigmatizing type of shaming is disintegrative. It typically involves attaching a deviant label to the offender. Labeling is stigmatizing because it focuses on the public degradation process, the labeling of the deviant, yet with little attention on delabeling and reintegration. The deviant person, once labeled and stigmatized, is likely to become an outcast and pushed into the criminal subculture. The deviant behavior becomes their master status. A master status is a characteristic(s) that defines an individual. For example, a person could be a good athlete or a good student, but if they are caught for engaging in delinquent behavior and are labeled, this label trumps what they were characterized as being before. Consequently, it precipitates the outcast to internalize a criminal identity and reject any attempts at shaming and reintegration. This self-fulfilling prophecy happens as a response to shaming that dramatizes the offenders' evil.

Criminologists of divergent perspectives (e.g., rational choice, social learning, labeling) now recognize that most criminals do not reject criminal law outright. Instead, they rationalize their deviant actions by shifting the blame of their deviance onto others (e.g., "it's the victims' fault," "everybody else is doing it," "I did not mean it"). The ability to rationalize their deviant behavior allows the deviants to temporarily suspend their commitment to the law and insulates them from shame. In this process of rationalization, criminal subcultures provide the much-needed social support for the deviant behavior, particularly when legitimate opportunities are blocked due to negative labels. Within the world of criminal subcultures, deviant actions are rationalized as a defensible lifestyle – against the injustice and stigma often felt inflicted upon the outcasts. Deviance thus becomes a way of life. The more attempts that are made to reform the outcasts, the worse it becomes (Braithwaite 1989; Becker 1973; Lemert 1951).

Some societies are prone to reintegration, whereas other societies prefer punishment as a way of social control. In general, societies characterized as communitarian, with reciprocal relationships among its citizens, are more likely to shame reintegratively. Shaming within this context is likely to be more potent, resulting in lower crime rates, because disapproval is unlikely to be rejected. In contrast, societies characterized as individualistic, with its citizens interacting only out of necessity and convenience, are more likely to resort to punishment. Shaming in this context is less effective because stigmatization and punishment are likely to cut off the deviants' interdependencies with their mainstream social relations and push them into the criminal subcultural world.

In addition, shaming is culture specific. Similar public expressions of disapproval of a deviant act may be perceived differently. In the United States, for example, the same outward disapproval of a delinquent act in an Asian culture setting may be received differently than it would in an African American culture setting. This is based on the concept that the ideals and attitudes held by different cultures are not the same in regards to legal justice, norm violations, and the extent of the impact by external forces (Braithwaite 1989). Even within the same cultural context, different social organizations (i.e., family, neighborhood) may deliver various degrees/kinds of shaming. For example, a study on Chinese urban residents found that family members generally use shaming and reintegration highly when a family member deviated because acts of deviance reflect badly on the rest of the family and their ancestors. With the economic development since the 1980s, Chinese neighborhoods are transforming to be become more transient. The study found that residential mobility affected shaming practices in the neighborhood (making shaming less reintegrative) much more significantly than shaming practices within the family (Lu et al. 2002).

In sum, reintegrative shaming is likely to label only the deviant acts, whereas stigmatizing shaming is likely to label the actor. Reintegrative shaming is more effective at crime control than stigmatizing shaming, because reintegrative shaming minimizes the risks of pushing deviants into criminal subcultures, whereas stigmatizing shaming increases these risks. In addition, deviants who are embedded in relationships that are overwhelmingly characterized by social approval are more likely to respond positively to shame. In contrast, deviants, who are not in these types of relationships, are more likely to reject shaming and make it criminogenic rather than crime inhibiting.

Shame and Communitarianism

The effectiveness of shaming depends upon two interrelated social conditions: interdependency and communitarianism. Interdependency refers to the interrelationship among individuals. It is an individual level variable. Individuals with more interdependencies are more susceptible to shaming; thus, they are less likely to commit crime.

There is an overwhelming consensus among criminologists regarding factors critical to an individual's interdependent relationships. These factors include, but are not limited to, age, gender, marital status, education, and employment status. For example, studies have found that the most important correlate of interdependency is stage in the life cycle. Individuals are mostly likely to become a deviant between 15 and 25 years of age. The two life-defining events for deviants to go straight are getting married and having children (Braithwaite 1989).

Communitarianism is the antithesis of individualism. It is the combination of a dense network of individual interdependencies with strong cultural commitments to mutuality of obligation. There are three basic elements to communitarianism: (1) interdependencies densely enmeshed in all spheres of social life, (2) interdependencies characterized by mutual obligation and trust, and (3) interdependencies built upon group loyalty (Braithwaite 1989). Thus, while numerous interdependencies are a necessary condition for communitarianism, mere interactions among group members are not sufficient for the group/society to be characterized as communitarian.

In communitarian societies, communities are typically defined by social ties, such as in family or neighborhood, as well as other significant relationships such as fellow students and coworkers. These communities are held together by mutual dependency and the need for cooperative endeavor (typically in a strict hierarchical order based on an individual's group status and/or social roles) to function. Countries and/or tribes in many parts of the world, such as Asia and Africa, can be characterized as communitarian. Most Western societies may be identified as more individualistic rather than communitarian. As the economy progresses, societies become more urbanized, mobile, diverse, and individualized. While there may be an abundant amount of interdependent relationships in highly developed areas (i.e., Western societies), because of increasing economic activities (i.e., landlords and tenants, bank tellers and customers) and other activities accompanying the economic growths (i.e., the police and criminals), these interdependencies may be more out of necessities and convenience, opposed to mutual obligation and trust. Within this context of high individualism, shaming by communities becomes particularly challenging. This is because "As a society becomes more role-differentiated [as in a modern society], the potential for effective shaming increases in important ways, but so does the potential for stigmatization that cuts off effective shaming" (Braithwaite 1993, p. 15).

Shaming can be administered by private individuals (i.e., communitarian-based shaming), and it can also be administered by the state (i.e., state-sanctioned shaming). In most developed Western countries, shaming is handled by the state via the criminal-justice system. In contrast, communitarian based shaming is dominant in many developing nations where legal systems are not fully developed and accessible.

Studies have found that communitarian-based shaming is more effective than state-sanctioned shaming in two aspects. First, shaming delivered by people of high interdependency and within a communitarian society (i.e., family, friends) is more potent than when it is delivered by an impersonal state because of the constant contact the intimates have with the individual. Second, shaming by significant others is more likely to be expressed in ways that is reintegrative, compared to the shaming by an impersonal state. Offenders are less likely to be stigmatized and adopt an outcast master status. In this context, a communitarian society resembles an enlarged loving family, where reintegrative shaming works best.

Reintegrative Shaming and Its Effectiveness

Since its publication in 1989, the theory of reintegrative shaming has aroused great interest among scholars and policy makers. Several studies tested this theory in a variety of settings.

A study using adolescents (ages from 14 to 17) from a single urban area in a southwestern state of the United States examined the relationship between the adolescents' perceptions of their parents' sanctioning methods (reintegrative vs. stigmatizing) and their reports of predatory delinquency (Hay 2001). Findings of this study are largely consistent with Braithwaite's theory of reintegrative shaming in that the level of interdependency of the adolescent with the parents and the use of reintegration by the parent were statistically significant. Parental sanctions are a way to reinforce a close relationship within families that have strong interdependent relationships between parents and children.

Reintegrative shaming theory has been used as a valuable interpretive framework in some experimental studies. For example, a study on a drug court in a southwest state of the United States revealed that drug courts were perceived to follow more of the reintegrative shaming type of model than a traditional court that offered more of an assembly line type of service. However, drug court participants had a higher recidivism risk than nondrug court participants. A compelling explanation for the failure of drug courts is because the presumed comparability with RST in process and structure is not necessarily true. Drug courts may be more stigmatizing than general courts based on field observation data (i.e., the initial and second appearances in courts resembled more of a public degradation ceremony than reintegrative shaming). In addition, disapproval expressed by the judge may not be potent. Judges represent an impersonal state and may not be regarded by offenders as someone respectable or important. Furthermore, judges do not continue to have relationships with the drug users after seeing them in court (Miethe et al. 2000). In sum, the RST can be used to guide programs that are intended to restore offenders back into the community and can help those who manage and evaluate the programs to make them more successful.

Several studies examined family group conferences in Australia using reintegrative shaming theory. For example, a study examined basic conditions for a successful restorative/

reintegrative family conference and identified a list of key factors. First is separating the event and the offender. While the event may be defined as irresponsible, wrong, or criminal, the offender is ought to be supported and viewed as a whole person. Second, those who administer reintegrative shaming must represent all the interests involved (i.e., offender, victim, and the law) to mete out a fair and effective solution to the problem (i.e., the final goal is reintegration and restoration). Third, apology, repentance, and forgiveness must occur in order to terminate the separation of the offender and victim and to enforce reintegration rituals (Braithwaite and Mugford 1994). Another study on restorative justice conferences suggests that these types of conferences tended to have a greater and positive psychological impact on participants' view on the legitimacy of the law (Tyler et al. 2007).

Reintegrative shaming theory has also been applied to a study on nursing homes' compliance to regulatory rules in Australia. The study showed that the nursing homes that had inspection teams who had a reintegrative shaming philosophy (high disapproval and high reintegration scores) showed greater improvement in regulatory compliance, compared to inspection teams having other philosophies (i.e., tolerance [low on disapproval and high on reintegration] or stigmatizing [high on disapproval and low on reintegration]). In addition, the more interdependency between regulators and the nursing home managers, the more compliant the nursing home managers are toward the regulations (Makkai and Braithwaite 1994).

Despite the vastly different social conditions in Western and Eastern countries, scholars have also attempted to examine the validity of the claims made by reintegrative shaming theory in Asian countries. Using a sample of 1,725 adolescents (ages 11–17) in China, one study attempted to identify the general effect of reintegrative shaming theory and the predictive effect of delinquency disapproval (shaming) and forgiveness (reintegration) on delinquency involvement. The study found that parental shaming, parental forgiveness, and peer shaming had reduced the involvement of predatory offenses. However, when the interactive effect of shaming and forgiveness was introduced in the multivariate analysis, reintegrative shaming theory (represented by the interaction term of shaming and reintegration) did not appear to be a significant predictor of predatory offense involvement (Zhang and Zhang 2004).

The study in China, described above, generated somewhat mixed results about the theory. Nevertheless, it is generally believed that Asian cultures, such as in China and Japan, which are influenced by Confucianism and centered on family, frequently use shaming and guilt induction as means of informal and formal social control (Braithwaite 1989). This widespread use of shaming within the family model has been regarded as contributing to the low crime rates in these countries. In the next section, two unique crime prevention and control programs, *bang-jiao* in China and *bizai shobun* in Japan, are described in order to illustrate how the principles of reintegrative shaming work in these unique sociocultural contexts.

Reintegrative Shaming in Asia

Compared to individualism shown in Western countries, many of the countries in Asia are based on ideas of communitarianism. Due to strong Confucian influence, Southeast Asia in particular (e.g., China, Japan, Korea, Taiwan) holds high regards for social status, hierarchy, order, peace, and harmony. These values help promote conformity and submission to group interests. In contrast, the values fundamental to Western ideas of individualism – individual rights and freedom – are only placed in a relative context in these Asian countries.

As a result, a communitarian society prefers informal social control to formal social control. Collective citizen action is considered an effective strategy for maintaining peace and harmony, solving disputes, controlling crime, and reducing the fear of crime. To foster an effective informal social control, citizens are taught these social responsibilities through socialization, which is to behave appropriately according to group/social norms and rules. This takes place first and foremost within the family, then at school, in the neighborhood, and in the workplace. In addition, citizens are expected to help others observe the rules of ritual propriety and help educate and correct behaviors that deviate from the norms. This network of informal social control is largely operated on reciprocal interdependencies with mutual respect. The *bang-jiao* program widely practiced throughout urban neighborhoods in China represents such an informal social control model.

Bang-jiao and Reintegrative Shaming in China

One of the major differences between social control in China and Western countries is that the Chinese attempt to control both the behavior and minds of the people. In contrast to the Western concept of original sin, Confucianism assumes the goodness in people. Crime is thus viewed as the result of the environmental influence. To offset the negative influence of the environment and help individuals get in touch with the good of their inner selves, Confucius and his followers believed in the importance of the rule of *li* (moral code). This was because *li* was viewed as essential in fostering an internationalization of the basic moral principles, promoting voluntary compliance to the rule, and being virtuous and good. Confucius also believed in the power of education in shaping people's thoughts and behaviors. Moral awakening through thought education is considered the primary stabilizer of society (Chen 2002).

In this process of moral education, shaming is regarded as playing a critical role – a role that calls attention to the damage the deviant/criminal has done to the victim/community and a moral imperative to correct the wrongs and amend the broken relationships caused by the deviant/criminal act. To the Chinese, shaming can generate both negative (e.g., stigma) and positive (e.g., deterrence, rehabilitation) results, depending on the strategies (Chen 2002).

One strategy is early intervention. Nipping crime in the bud is common wisdom rooted in Chinese philosophy. Early intervention only requires a small dosage of shaming and small educational efforts could result in full reintegration. Even though early intervention may run the risks of intrusion into other people's lives, it is taken for granted and viewed not as meddling, but caring.

Another strategy is popular participation. Chinese social control is from the bottom up, not the top down. The Chinese prefer to handle their own crime and delinquency problems in their community rather than leaving them to the professionals. The popular participation approach ensures a swifter and more effective crime control service because those who deliver the service presumably have more intimate knowledge about the deviant person as well as the act and have more stakes in the quality of service.

The total approach is yet another strategy used in crime prevention and control. As China moves toward legalization and professionalization, greater emphasis is given to law and the legal professionals in addressing major legal issues such as crime. Nevertheless, legal professionals, such as the police and judicial officers, are expected to use all sorts of means (e.g., legal, administrative, and social) to prevent crime and reintegrate offenders. *Bang-jiao* in urban neighborhoods of China represents one such program.

Bang-jiao literally means, assisting, helping, guiding, and directing offenders, especially juvenile offenders (Zhang et al. 1996). It is a community-based program that utilizes remedial and preventative measures for controlling crime. Even though no particular group of individuals are excluded from the objects of *bang-jiao*, *bang-jiao* typically handles predelinquents, delinquents, deviants, and offenders who have committed minor offenses and are without a prior record. These groups of people are targeted, because they are the mostly likely to be "helped" (*bang*) and "educated" (*jiao*) successfully (13–28 years of age). Most *bang-jiao* programs are situated in urban neighborhoods, and they consist of parents, relatives, neighbors, teachers, coworkers, resident committee members, and/or local police who work as a team in carrying out *bang-jiao*.

Even though *bang-jiao* is community based and represents an extra legal measure to crime prevention and control, it still has some basic principles guiding its program. According to the Juvenile Delinquency Prevention Law (1991), *bang-jiao* must be conducted based on the following principles: (1) fairness and equality, (2) practical guidance, and (3) response to genuine repentance with love, emotional support, and sincere acceptance of deviants back into the community (Zhang et al. 1996).

The characteristics and process of a typical *bang-jiao* intervention can be illustrated with the following example. A teenage boy is sent to a work-study school for rehabilitation because of repeated thefts. After recommendations from his *bang-jiao* team and work-study school principals, he was transferred to a normal school. Every day he came home after dark, because he was too ashamed to be seen by his neighbors. The boy's mother finally talked to the *bang-jiao* team members. The team members visited the neighbors and found out that the neighbors wanted to communicate with the boy, not ridicule him, but did not know how. The *bang-jiao* team discussed how to overcome these barriers with the neighbors and decided to hold a community event in which they knew the boy was interested in and would participant in (community painting exhibition). They boy was invited and at the event he was highly praised for his work. This experience changed the boy, and he started being involved with community events and being a part of the community (Braithwaite 2002).

From a theoretical standpoint, reintegrative shaming facilitates this process of *bang-jiao*. Although little empirical research has examined the effectiveness of *bang-jiao* in deterring crimes and reducing recidivism rates, one study based on self-report data suggests there is a significant negative relationship based on reports by inmates living in communities with a *bang-jiao* program (Zhang et al. 1996). More specifically, the inmates who lived in neighborhoods with an active *bang-jiao* program reported a significantly lower likelihood of recidivism compared to those who lived in neighborhoods without an active *bang-jiao* program.

Even though crime rates, including juvenile delinquency rates, have surged since the economic reforms in the 1980s in China, the majority of juvenile delinquents were able to successfully have their deviant label removed and be reintegrated back into the community. This was supported with the low recidivism rate of between 8 % and 15 % annually in recent years (Chen 2002).

Bizai Shobun and Reintegrative Shaming in Japan

Similar to the Chinese view of the relationship between the individual and society, the Japanese see individuals not as an isolated entity, but as part of the network. There is a kind of social web that binds individuals into a collective. Within this largely communitarian society, punishment is not geared toward retribution, but rehabilitation and restoration.

Due to the importance of status and family to an individual, shaming has different meanings in the Japanese culture. Its potency can be different as well, particularly when compared with Western countries.

When assessing punishment, one of the criteria used in Western culture is severity of punishment (proportionality of punishment to crime), typically measured by the length of prison sentences and/or the amount of monetary fines imposed on the convicted offenders. In Japan, however, the focus is different. Contrary to isolation and punishment of the individual wrongdoer, the Japanese prefer restitution and restoration. Numerous terms, such as benevolent, paternalistic, and familiar, have been used to describe this reintegrative shaming practice unique to the Japanese. Among these different terms, an individualized decision-making model seems to best capture the Japanese crime control and prevention practice (Foote 1992).

The Japanese criminal-justice system is benevolent in that its goal is to achieve reformation and reintegration of offenders into society through lenient sanctions tailored to the offender's particular circumstances (Foote 1992). This benevolent-paternalism model emphasizes love and mutual respect so that the best interest of the offender and his/her rehabilitation is the focus of any decisions. The success of this family model depends on officers' use of widespread power of discretion. It not only permits but also expects public officials to use discretion and take a preventive approach to crime and disorder. The major factors officers consider include family circumstances, employment status, and other types of support mechanisms available to the offender and, to a lesser extent, satisfaction of the victim.

Central to the Japanese practice of restoration and reintegrative shaming are concepts of apology, repentance, and forgiveness. To apologize, the wrongdoer must first confess to the crime. The importance of confession in the Japanese criminal-justice system can hardly be overstated. Numerous scholars of the Japanese legal system concur that confession, apology, and begging for forgiveness are essential elements at virtually every stage of the criminal process in Japan. Confession and admission of guilt are not used as part of the plea bargain, but are regarded as a gesture of complete submission to authorities in Japan. Confession, accompanied by sincere apology and remorsefulness, is viewed as having both the probative and correctional value. It aids the police, the judge, and the correctional officers to better achieve the goals of holding criminal offenders accountable for their wrongdoings and reforming them into law-abiding citizens.

The ability of legal officials to exercise widespread discretionary power lies in the great faith that is bestowed on public officials. Legal officials enjoy great autonomy and widespread power sometimes beyond the legal spheres in Japan. When dealing with an offender, the classification scheme of legal (e.g., offense severity, prior criminal record) vs. extralegal (e.g., age, gender, employment) factors commonly used in Western countries may not be relevant in Japan. Instead, means of punishment/treatment will be considered favorably if they are least disruptive to the deviants'/criminals' life, fit individual circumstances, and most likely lead to the successful reintegration back into the community.

Reintegrative shaming as a punishment philosophy has been institutionalized in Japan. All three authorities in the criminal-justice system are given widespread discretion when disposing of cases. For example, the police have the authority not to report minor offenses (*bizai shobun*) (Code of /Criminal Procedure, art. 246), the prosecutors have the authority to suspend prosecution (Code of Criminal Procedure), and the courts have the authority to suspend execution of sentences (Criminal Code, art. 25).

Below the practice of *bizai shobun* is summarized to show how the police, in exercising their discretionary power, translate the philosophy of reintegrative shaming into practice.

The Japanese police enjoy widespread discretion when performing their crime prevention and law enforcement duties. For example, the Japanese local police have the authority to conduct residential surveys twice a year for the purpose of getting to know the residents, registering new comers, and preventing future crimes. The Japanese police officers may stop and question anyone in public without reasonable suspicion or probable cause. They also have the authority to interrogate a criminal suspect for a lengthy period of time without a warrant. The relative "crime-free society," combined with high citizen respect and cooperation in Japan, make the police even more powerful and free in using the tools at their disposal when dealing with criminal suspects. The Japanese police have a "quasi-judicial" function, especially when dealing with simple cases or cases involving petty offenses. One such power is discretion to drop trivial cases from further investigation and charges.

Bizai shobun is rationalized based on two principles: (1) when involving minor offenses, the police shall not put victims under undue psychological stress with an unnecessarily punitive punishment of offenders and (2) to enhance efficiency of the police work, police should focus on more serious crimes and use informal means in dealing with the less serious crimes (Arakawa 1987).

To maintain jurisdictional consistency, the police power of *bizai shobun* is granted by the prosecutor. Typically the supervising prosecutor's office will issue an order for specific types of trivial offenses to be dropped from further processing by the police. These trivial offenses commonly include assault, theft, fraud, embezzlement, and gambling. For example, in the early 1990s there were instructions for theft cases to be dropped under three conditions: (1) if the value of stolen goods did not exceed 10,000 yen ($85); (2) if the offender had a fixed residency, which excluded transients (particularly foreigners); and (3) if the offender repented (Johnson 2002).

It was estimated that up to 40 % of criminal suspects arrested by the police are released without any charges. Many of these cases were resolved through negotiation, apology, compensation, and forgiveness between the offender and the victim. Furthermore, victims, their family, and their community are critical in the police decision to employ *bizai shobun*, because they must be willing to accept the offender's apology and to forgive him/her. This is to ensure the successful reintegration of the offender back into the community.

Limitations of Reintegrative Shaming

One of the most fundamental criticisms of reintegrative shaming theory is its incomplete conceptualization of sanctions (Hay 2001). Braithwaite depicts sanctions as shaming, either reintegrative or stigmatizing, and makes the assumption that all deviance is detected and reacted to with some type of corrective action. However, deleterious reactions (no reaction) to deviance are common. Even if existing, some reactions to deviance may only intend to evoke fear of further punishment, rather than trying to evoke remorse/shame. Reintegrative shaming theory is thus regarded as incomplete, because shaming is not the only form of corrective reaction.

The concept of reintegrative shaming has the obvious strengths of maintaining continuous and interdependent relationships. However, shaming in communitarian societies (e.g., Confucian culture) may become too powerful and overwhelming, thus, having the risk of destroying robust individuation within secure social bonds. An example of the danger associated with engulfment of the individual is the high suicide rate in Japan and, to a certain degree, in China (Braithwaite 1989)

Reintegrative shaming theory is also limited in scope to personal and property offenses where there is a clear consensus regarding their moral wrongfulness (Hay 2001). In cases where societal consensus is unclear, such as white-collar crime, drug offenses, and public order offenses (e.g., prostitution, gambling), shaming may not take place (either not expressed by the public or dismissed by the wrongdoer), let alone reintegration.

Besides these conceptual challenges, reintegrative shaming theory faces difficulties of empirical validation. Some of the key concepts such as shame and reintegration are difficult to operationalize and measure with empirical data (Zhang 1995; Zhang and Zhang 2004). In addition, major moderating and mediating factors that may indirectly affect the potency of reintegrative shaming have yet to be established. For example, a study found that procedural justice (i.e., a sense of being treated fairly) affected how well offenders responded to efforts of reintegrative shaming (Tyler et al. 2007).

These issues of measurement are magnified particularly in cross-cultural settings where definitions and perceptions of shame may diverge markedly. A successful shaming program in one type of society (e.g., communitarian) may not work well in another type of society (e.g., individualistic).

Last, but not the least, shaming may run the risk of civil vigilantism. It is especially noteworthy of its potentially massive invasions of personal privacy in the online environment. An example involves the growing popularity of the "Human-Flesh Search" (*renrou sousuo*) in China where individuals' private records were dug out and posted on the website by Internet users for a variety of reasons such as personal revenge and resentment to corrupt public officials and celebrity. This counters the very intent of maintaining a civil and orderly society (Skoric et al. 2010). Within the criminal-justice system, reintegrative shaming programs may run the risk of widening the net of social control and blurring the line between moral and legal issues.

Conclusions

Reintegrative shaming, as practiced in China and Japan, place great faith on public officials – their personal morality, skills, and judgment – in making important decisions about law and order. By using extralegal means in dealing with essentially a legal matter, these criminal-justice systems runs the risk of corruption and undermining procedural transparency and the predictability and fairness of the decisions. Despite of these potential limitations, reintegrative shaming provides a plausible, alternative means for crime control and prevention. Reintegrative shaming has been demonstrated theoretically, and to a certain degree, empirically, to be more effective and efficient than the traditional means of social control.

Recommended Reading and References

Arakawa M (1987) Diversion and criminal law in Japan (ディバージョンと刑法に関する一考察 – 警察における微罪処分を中心として–). Law Polit (法と政治) 38:421–466

Becker H (1973) Outsiders. Free Press, New York

Braithwaite J (1989) Crime, shame, and reintegration. Cambridge University Press, New York

Braithwaite J (1993) Shame and modernity. Br J Criminol 33:1–18

Braithwaite J (2002) Restorative justice and responsive regulation. Oxford University Press, New York

Braithwaite J, Mugford S (1994) Conditions of successful reintegration ceremonies. Br J Criminol 34:139–171

Chen X (2002) Social control in China: applications of the labeling theory and the reintegrative shaming theory. Int J Offender Ther Comp Criminol 46:45–63

Foote DH (1992) The benevolent paternalism of Japanese criminal justice. Calif Law Rev 80:317–335

Hay C (2001) An exploratory test of Braithwaite's reintegrative shaming theory. J Res Crime Delinq 38:132–153

Johnson DH (2002) The Japanese way of justice: prosecuting crime in Japan. Oxford University Press, New York

Lemert E (1951) Social pathology. McGraw-Hill, New York

Lu H, Zhang L, Miethe TD (2002) Interdependency, communitarianism and reintegrative shaming in China. Soc Sci J 39:189–201

Makkai T, Braithwaite J (1994) Reintegrative shaming and compliance with regulatory standards. Criminology 32:361–385

Miethe T, Lu H, Reese E (2000) Reintegrative shaming and recidivism risks in drug court: explanations from some unexpected findings. Crime Delinq 46:522–541

Skoric MM, Ping J, Chua E, Liew MA, Wong KH, Jue P (2010) Online shaming in the Asian context: community empowerment or civic vigilantism? Surveill Soc 8:181–199

Tyler TR, Sherman L, Strang H, Barnes GC, Woods D (2007) Reintegrative shaming, procedural justice, and recidivism: the engagement of offenders' psychological mechanisms in the Canberra RISE drinking-and driving experiment. Law Soc Rev 41:553–586

Zhang SX (1995) Measuring shaming in an ethnic context. Br J Criminol 35:248–262

Zhang L, Zhang S (2004) Reintegrative shaming and predatory delinquency. J Res Crime Delinq 41:433–452

Zhang L, Zhou D, Messner SF, Liska AE, Krohn MD, Liu J, Lu Z (1996) Crime prevention in a communitarian society: Bang-Jiao and Tiao-Jie in the People's Republic of China. Justice Q 13:199–222

Shift Work

▶ Officer Safety, Health, and Wellness

Shock Incarceration

▶ History of Boot Camps

Shoplifting

Jason Davis
Criminal Justice, Clayton State University,
Morrow, GA, USA

Overview

Shoplifting refers to the theft of retail merchandise during hours of operation (Clarke 2003; Hayes and Cardone 2006). This form of larceny is viewed as a minor offense that does not generate significant public outrage or fear. Furthermore, it lacks the sensational appeal that necessitates extensive media coverage or immediate legislative responses (Cromwell et al. 2006; Klemke 1992). Notwithstanding these perceptions, retail theft imposes significant financial and social harms. Each year, shoplifting costs companies billions of dollars in lost revenues or merchandise. Beyond these losses, however, retail theft entails additional financial costs. Companies allocate millions of dollars each year for security personnel and measures to help deter shoplifting (Tonglet 2002; Gills et al. 1999; Cox et al. 1993). In some cases, retailers pay higher insurance premiums because they experience high rates of theft (Burrows 1998; Finklea 2011). Shoplifting can also adversely affect honest patrons. To offset financial losses, companies levy a "consumer tax" in which merchandise prices of commonly pilfered items are substantially increased (Cox et al. 1993; Schwartz and Wood 1991). Customers can also be inconvenienced by cumbersome security measures such as cable locks, key controlled glass cabinets, or other equipment that limits accessibility to merchandise (Tonglet 2002; Cox et al. 1993; Dawson 1993). Finally, state governments and local communities lose tax revenues that could be generated from stolen merchandise.

Beyond the financial tolls, shoplifting can prompt safety and domestic security problems. Unsuspecting customers or employees can be injured by fleeing or aggressive shoplifters attempting to avoid apprehension (Axelrod and Elkind 1976). A 2011 survey found that 13 % of retail crime apprehensions involved some level of physical assault or battery (National Retail Foundation 2011). When perishable goods such as infant formula and over-the-counter drugs are shoplifted, fraudulently repackaged, and sold on the open market, customers can unknowingly consume outdated or spoiled products (Finklea 2011; National Retail Foundation 2011). Finally, shoplifting by professional or organized retail theft (ORT) rings has been linked to money laundering, organized criminal syndicates, and the funding of terrorist activity (National Retail Federation 2011; Finklea 2011).

This entry provides a general overview of shoplifting. Specifically, it discusses the financial losses associated with shoplifting and how these figures likely underestimate the true costs of shoplifting. Next, a typology of shoplifters will be offered with primary attention given to amateur and professional shoplifters. A discussion of common shoplifting targets and techniques will follow as well as an examination of the theoretical explanations used to help understand why offenders engage in shoplifting. Finally, this entry will review common antishoplifting measures utilized within the retail industry and summarize their effectiveness.

Financial Estimates of Shoplifting

No other property crime including robbery, burglary, or auto theft can match the financial impact of shoplifting. Estimates have consistently shown that retailers lose anywhere between $10 and $50 billion dollars annually to shoplifting (National Association for Shoplifting Prevention 2012; National Retail Foundation 2011; Centre for Retail Research 2011; Clarke 2003). A recent study found that shoplifting cost US retailers approximately $12.1 billion (National Retail Federation 2011). In 2010, a global retail theft survey discovered that shoplifting incidents cost 1,187 retailers representing 43 countries $51.5 billion (Centre for Retail Research 2011). While the financial costs attributed to shoplifting are staggering, it is generally understood that these figures grossly underestimate the true extent of the problem (Burrows 1999; Dabney et al. 2004; Clarke 2003).

Dilemmas Associated with Shoplifting Estimates

Underreporting is a major problem associated with shoplifting data as only one-third of all thefts get reported to the police (Dabney et al. 2004; Clarke 2003). When shoplifting is officially reported, few incidents result in a formal record or prosecution (Klemke 1992; Dabney et al. 2004; Clarke 2003). Instead, law enforcement is more inclined to issue a warning or have the charges downgraded to a misdemeanor or violation offense (Dabney et al. 2004; Clarke 2003). Many retail companies are reluctant to formally report shoplifting and thus contribute to underreporting dilemmas (Burrows 1999). Part of the hesitancy may stem from legal obstacles. Stores may be unaware of shoplifting statutes, or they may be dissatisfied with the weak laws or outcomes that insufficiently punish offenders (Axelrod and Elkind 1976; Burrows 1999). In lieu of formal responses, companies more commonly seek civil restitution from the offenders. Companies may also be reticent about disclosing adverse shoplifting information (Burrows 1999). Publicity about shoplifting could alert potential offenders about a store's security vulnerabilities, alarm shareholders about potential crime problems, or signal that costly security measures are ineffective (Burrows 1999; Axelrod and Elkind 1976). Even if retailers suspect customers of shoplifting, they are apprehensive about accosting them because of a fear of litigation resulting from false arrest or imprisonment (Axelrod and Elkind 1976). Some stores are equally worried about the risk of embarrassment and loss of reputation that false arrests could yield. Collectively, there appears to be reluctance to report, classify, or respond to shoplifting as a serious felony offense.

Beyond underreporting dilemmas, sources of shoplifting information including police records, victimization or self-report surveys, and industry level data are problematic because they fail to directly measure shoplifting (Burrows 1998). For instance, shoplifting is coalesced into the larger category of larceny-theft in the Uniform Crime Report, and therefore, it is impossible to discern what percentage of thefts can be attributed to shoplifting. The National Crime Victimization Survey (NCVS) neglects shoplifting altogether since it focuses on individual victimization, not crimes committed against organizations (Burrows 1998). Store-generated data suffers similar measurement limitations. Since only a small percentage of offenders are detected at the time of the offense, store security managers or personnel must "guesstimate" what percentage of loss can be directly attributed to shoplifting (Dabney et al. 2004). Through the use of routine audits, companies measure the disparity between merchandise physically present in a particular store and financial statements regarding sold merchandise (DiLonardo 1997). Such audits determine levels of *inventory shrinkage* or shortages in inventory levels. The greater the amount of unaccounted merchandise, the more revenue lost. With the exception of employee theft, shoplifting is generally considered the second most common source of inventory loss (Hayes and Cardone 2006). In most cases, estimates maintain that roughly 33–43 % of all retail losses are caused by shoplifting (National Retail Foundation 2011; Centre for Retail Research 2011). Nevertheless, without direct observation, it is not feasible to parse out what percentage of inventory shrink is actually shoplifting related (Dabney et al. 2004). Indeed, a study relying on surveillance cameras observed more shoplifting incidents than were formally reported to law enforcement (Dabney et al. 2004). The authors concluded that shoplifting incidents from a single retail drug store approximated the total number of all larceny-theft cases reported to police in the metropolitan Atlanta area over the course of 1 year. Such findings call into question the reliability and validity of both formal and store-generated estimates regarding the prevalence and financial impact of shoplifting (Dabney et al. 2004).

Shoplifting Typology

Research has consistently identified two categories of shoplifting offenders: (a) amateur or opportunistic shoplifters and (b) professional shoplifters that generally operate as part of an organized retail crime (ORC) ring (Clarke 2003). To a lesser extent, studies have also acknowledged the existence of compulsive shoplifters known as kleptomaniacs (Blanco et al. 2008).

Amateur Shoplifters

The majority of shoplifters are considered amateurs who steal relatively inexpensive merchandise for the purpose of personal use or consumption (Cameron 1964; Cox et al. 1990;

Clarke 2003). In many cases, amateur shoplifters lack a clear motive for stealing as many offenders possess the resources to legitimately purchase merchandise or they may not have an immediate need or desire for the products they steal. Accordingly, amateur shoplifting is typically viewed as an opportunistic crime in which offenders steal goods when they perceive relatively easy opportunities to do so. Some amateurs envision shoplifting more as a challenge and thus steal to see if they can get away with it. Amateur shoplifting also tends to be impulsive and infrequent. Their acts of theft are generally unplanned, and many steal only a few times a year (Cameron 1964). Overall, amateurs are not committed to a criminal lifestyle (Cameron 1964; Clarke 2003). Formal detection or apprehension is generally enough to deter future offending (Cameron 1964).

Amateur shoplifting is not confined to a small subgroup of offenders (Cox et al. 1990). Instead, a sizeable number of people have shoplifted at some point in their life (Cameron 1964; Cromwell and Thurman 2003). According to the National Association for Shoplifting Prevention (2012), an estimated 1 out 11 people or 27 million people have shoplifted. Not surprisingly, there is not a prototypical profile of a shoplifter. Amateur shoplifters are a demographically diverse group. Various studies have concluded that men and women are equally likely to steal as are young and old and people of various ethnic backgrounds and social classes (Klemke 1992; Dabney et al. 2004; Blanco et al. 2008; Asquith and Bristow 2000; Dawson 1993; Tonglet 2002; Cromwell and Thurman 2003; Cromwell et al. 2006). While the influence of demographic characteristics has been inconclusive, behavioral clues may be stronger predictors of shoplifting. Shoppers who leave a store without purchasing an item, scan the store for security measures, or tamper with products are more likely to steal (Dabney et al. 2004).

Organized Retail Crime (ORC): Professional Shoplifting

Unlike amateurs, professional shoplifters are more inclined to steal merchandise for the purposes of reselling stolen goods for a profit (Cameron 1964; National Retail Federation 2011; Finklea 2011). This category of shoplifters tends to be more organized and skilled, steal merchandise more frequently, and target more expensive items. It is estimated that roughly one-quarter of shoplifting cases are committed by professional shoplifters (Finklea 2011). Professional thefts are generally part of a larger criminal network known as organized retail crime (ORC). At its core, ORC networks involve a relationship between *boosters* and *fences*. Boosters steal merchandise and sell these goods to fences for approximately 10–30 % of the retail value of a stolen product (Cameron 1964; Finklea 2011). For their part, fences purchase, conceal, or transport stolen merchandise. Fences use a variety of business such as small “mom and pop” community establishments, pawn shops, flea markets or swap meets, convenience stores, and online auctions to dispose stolen goods (Cromwell and Olson 2006; Finklea 2011). In some cases, the fences may represent organized crime or terrorist operations. Fences typically supply boosters with “fence sheets” that specify desired merchandise, the value or profit margin for stolen goods, and stores that commonly carry preferred items (Cromwell and Olson 2006). Some mid- or higher-level fences ostensibly operate legitimate warehouses that in reality remove security tags, store labels, repackage goods, or alter expiration dates from stolen merchandise (Finklea 2011; Cromwell and Olson 2006). The “cleaned” stolen goods are amalgamated with legitimately purchased merchandise and reintroduced on the open market. In addition to professional fences, there is the existence of part-time fences who occasionally buy stolen merchandise primarily for personal consumption. In other instances, barters exchange professional services or drugs for stolen merchandise. These nonprofessional fences do not purchase a high volume of stolen merchandise nor do they rely on fencing as their primary source of income (Cromwell and Olson 2006). The advent of online internet auction sites such as eBay has created another outlet for professional shoplifters to dispose stolen goods.

Known as *e-fencing*, this form of disposing stolen merchandise is potentially more lucrative as professional shoplifters can earn approximately 70 cents on the dollar for stolen goods (Finklea 2011).

Kleptomaniacs

While amateurs and professionals are the ubiquitous shoplifters, there is a smaller category of offenders known as kleptomaniacs who are unable to resist the temptation to steal and generally steal items they do not have a conceivable use for (Blanco et al. 2008). Historically, the term kleptomaniac was used to describe females who had an uncontrollable urge to pilfer (Hayes and Cardone 2006). More recently, kleptomania is considered a psychological disorder in which a person experiences intense feeling of anxiety prior to stealing and a sense of gratification or release after stealing (Blanco et al. 2008). It is generally believed that kleptomaniacs also suffer from other psychological disorders such as impulsivity, mood disorders, or obsessive-compulsive behaviors that compel them to steal.

Common Shoplifting Targets and Techniques

Common Shoplifting Targets

A general sample of items generally shoplifted include but are not limited to music, games, electronics, apparel, batteries, tobacco products, infant formula, and over-the-counter (OTC) medicines (Clarke 1999; Clarke 2003; Hayes and Cardone 2006; Finklea 2011). Clarke's (1999) CRAVED model focuses on six attributes that make certain targets vulnerable. In particular, items that are *concealable, removable, available, valuable, enjoyable, and disposable* are deemed *hot products*. On the whole, shoplifters commonly target merchandise that can be easily concealed and avoid stealing merchandise in large quantities or items too large to conceal for fear of attracting too much attention (Carroll and Weaver 1986). Both amateur and professional shoplifters will target widely available merchandise especially new attractive products such as cell phones, computers, video games, and apparel. While professional shoplifters typically steal merchandise that will garner greater profit margins, novices including adolescent offenders may be driven to steal merchandise that holds personal worth or cultural value. Finally, the CRAVED model predicts that shoplifters, particularly professional shoplifters, are more inclined to steal merchandise that is easy to sell or convert into cash.

Common Shoplifting Techniques

In its simplest form, shoplifting involves the illegal concealment of merchandise in one's clothing or hand. While some brazen shoplifters may simply grab goods and leave the store quickly, both amateur and professional shoplifters utilize a variety of methods to illegally pilfer retail merchandise. Research has noted that many shoplifters work in teams with some members serving as lookouts to alert the thieves about potential dangers (Hayes and Cardone 2006). Listed below are select shoplifting techniques commonly identified by researchers and practitioners (Hayes and Cardone 2006; Clarke 2003; Finklea 2011; DiLonardo 1997):

- **Booster boxes or bags**: Shoplifters use a special foil-lined dummy box or bag that has false bottoms or openings to hide stolen merchandise. The foil lining is designed to weaken electronic article surveillance (EAS) tag signals so that they do not activate or trigger the alarm upon departure from a store. In other situations, shoplifters will conceal stolen goods within a legitimately purchased boxes or bags including televisions boxes, computer boxes, sporting goods merchandise, suitcases, dog food bags, or safes.
- **Diversions**: Working in groups, one or more shoplifters will create some type of diversion (pretending to slip and fall, faking an illness or injury, dropping or breaking merchandise, knocking down a display cabinet or shelve, or triggering a fire alarm). While store employees are distracted, other shoplifters will steal merchandise.

- **Fitting room**: Shoplifters will take several items or garments into a fitting room where they will wear the stolen clothing or conceal items in a bag or purse.
- **Clothing or *crotchwalking*** This shoplifting method occurs when an offender wears large baggy clothing to hide merchandise. In some cases, shoplifters wear special clothing (i.e., coats, shirts, pants) that contains multiple hidden pockets to help conceal merchandise. Some female offenders will wear large maternity clothing to give the appearance of being pregnant while concealing merchandise.
- **Receipt manipulation**: In some instances, shoplifters will make a legitimate purchase and take the item to their vehicle. Thereafter, they will return to the store and steal the same item listed on the receipt. If stopped by employees or store detectives, the shoplifter can present the receipt as proof of a legitimate purchase. Days later, the shoplifter may return the legitimately purchased item for a full refund. In other cases, shoplifters may use discarded receipts found in the parking lot, trash bins, or store floor to steal the items listed on the receipt. The disposed receipts can also be used for refund purposes. The shoplifter may pick up the items listed on the dispensed receipt and seek a cash refund.
- **Smash and grab**: This form of theft occurs when a group of shoplifters typically steal a car or sports utility vehicle (SUV) and drive it through a storefront window. Once the vehicle has smashed through the store, shoplifters quickly grab and pack the car with as much merchandise as possible before driving away.
- **Price tag switching**: This refers to the fraudulent removal of price tags or bar codes from expensive merchandise and replacing them with cheaper price codes.

Theories of Shoplifting

Empirical studies have concluded that there is not a solitary factor that causes a person to engage in shoplifting (Cromwell et al. 2006; Tonglet 2002). Instead, shoplifting behaviors are shaped by a range of factors. Many view shopping as a crime of opportunity, while other perspectives assert that offenders view shoplifting as a trivial offense that does not cause harm. Some frameworks maintain that shoplifting is related to age-specific factors. Finally, research has argued that psychological needs or disorders may be a determinant of shoplifting behaviors.

Opportunity Perspectives

Some of the prevailing theories of shoplifting including routine activity theory (RAT) and opportunity structure for crime can be classified as theories of opportunity. Partially rooted in rational choice principles, these opportunity theories assert that would-be offenders take advantage of perceived easy opportunities to offend after careful consideration of the potential risks and rewards associated with a particular act (Clarke 1997). Opportunity theories also focus on the importance of the crime setting (Clarke 2003). In the context of shoplifting, opportunity theories are interested in understanding how the retail environment contributes to the prevalence of shoplifting. Shoplifters will seek stores with weak or ineffective security measures because they believe the benefits of shoplifting (procuring something for free) outweigh the risk of apprehension (Gills et al. 1999). Indeed, many shoplifters perceive retail pilfering as a low-risk offense that requires minimal effort, abundance of opportunities, and substantial gains (Tonglet 2002).

Routine activity theory maintains that crime occurs when a *motivated offender* comes into contact with *suitable targets* in the absence of *capable guardianship* (Cohen and Felson 1979). The convergence of these three factors in time and space represents a crime triangle. Routine activity theory is ideally suited for understanding shoplifting. Retail environments bring potential motivated offenders into contact with suitable targets including small, portable, and desirable goods that can be easily concealed. To the extent that retail environments lack adequate surveillance or security measures, motivated offenders will take advantage of the lack of guardianship and shoplift valuable targets.

Touted as an integrated theory that incorporates elements of environmental criminology, rational choice, lifestyle, and routine activity theory, Clarke's (1997) opportunity structure theory asserts that the physical environment influences opportunities for crime. Immediate situational or environmental factors dictate the supply of *targets* (objects), *victims* (vulnerable individuals), and *crime facilitators* (weapons) necessary for offending. To this end, retail environments play an important role in creating or reducing shoplifting opportunities. Through personal experiences, peer associations, or the media, would-be offenders learn the risks, effort, and rewards associated with specific stores, and they will target those with the greatest vulnerabilities or relatively easy opportunities to steal. In this context, retailers must manipulate the environment and implement target-hardening devices to decrease opportunities for offending. Research indicates that shoplifters do in fact evaluate store characteristics when deciding whether or not to steal (Carroll and Weaver 1986). In particular, experienced shoplifters assess store layouts for security devices, counter heights, or behavior of store staff in order to identify potential risks. The presence of these factors is enough to deter most expert shoplifters.

Techniques of Neutralization

Another explanation for why people shoplift is techniques of neutralization theory (Sykes and Matza 1957). This perspective asserts that most offenders hold conventional norms or attitudes, and they must rely on rationalizations to help neutralize the guilt and shame associated with offending. Potential offenders use a priori excuses that allow them to drift into criminality while temporarily suspending their moral convictions. In its original conception, Sykes and Matza (1957) identified five different techniques of neutralization that offenders use to justify criminal behavior including *denial of responsibility* (offending is beyond person's control), *denial of injury* (offending behavior is trivial and does not really harm anyone), *denial of victim* (victim deserved to have a crime committed against them), *condemnation of the condemners* (if authority figures engage in unethical behaviors with impunity, then others should be allowed to engage in similar acts), and *appeal to higher loyalties* (commitment to peers or peer pressure caused offending behavior). Recent research has identified at least seven other neutralizations including *defense of necessity* (offending is a matter of survival for basic or essential needs), *metaphor of the ledger* (crime offsets unfair conditions or can settle a grievance), *everyone else is doing it* (certain crimes are common), *denial of the necessity of the law* (laws are unfair), and *sense of entitlement* (offender should enjoy certain privileges) (Cromwell and Thurman 2003).

Studies have found that shoplifters typically use one or more justifications to help neutralize their illegal pilfering (Tonglet 2002; Cromwell and Thurman 2003; Clarke 2003). Many offenders hold proshoplifting attitudes and perceive pilfering as a relatively trivial offense that does not significantly harm the overall sustainability of profitable retailers (*denial of injury*). Some shoplifters imply that they had little control over their actions (Cromwell and Thurman 2003). This *denial of responsibility* neutralization would be consistent with shoplifters who suffer some psychological compulsion or addictive behavior to engage in theft. In other instances, offender attitudes are congruent with a *denial of victim* perspective. Many people view large organizations as wealthy and somewhat ruthless entities (Schwartz and Wood 1991; Clarke 2003; Dawson 1993). Accordingly, shoplifters are less likely to experience feeling of guilt, remorse, or shame about stealing from impersonal stores. Such antagonistic attitudes about retailers may also reflect a *metaphor of the ledger* justification in which they view shoplifting as a sense of entitlement or compensation for unfair treatment or price markups by indifferent companies (Schwartz and Wood 1991). Other shoplifters may utilize a *defense of necessity* justification in which offenders assert they had to steal in order to survive or because they did not possess the resources to make a legitimate purchase (Cox et al. 1990; Cox et al. 1993; Cromwell and Thurman 2003). A recent study identified two other neutralizations including *justification by comparison* and *postponement* (Cromwell and Thurman 2003). Offenders using the former rationalization maintain that if they were not shoplifting, they would be committing more serious offenses, while those exploiting the latter neutralization momentarily delay any thoughts about the offending behaviors. In essence, people that use postponement do not think about their crime.

Age-Specific Reasons

Motivations to shoplift can also be age specific. Studies of juvenile shoplifters have discovered that adolescents steal for thrill or novelty reasons (Cox et al. 1990; Cox et al. 1993). In addition, social desirability is a motivation for adolescent shoplifters. Juveniles steal popular merchandise including electronics or clothing in order to enhance their status among peers. In some cases, juveniles steal items that they are too young to legally purchase or too embarrassed to buy such as cigarettes, alcohol, or condoms (Cox et al. 1990). Interestingly, there is little evidence to suggest that juvenile shoplifters are directly affected by peer pressure, dares, or status attainment (Cox et al. 1993). There is also evidence that juvenile shoplifters may not be deterred by antishoplifting security measures or formal apprehension (Klemke 1992). Apprehension by parents, store personnel, or the police can actually increase the likelihood of future shoplifting for adolescent shoplifters presumably because deviant labels propel the offender, particularly female shoplifters, to commit future acts of pilfering (Klemke 1992).

Psychological Factors

Some argue that shoplifting represents coping mechanism to address stressful life events or psychological problems including depression, anxiety, bulimia, anorexia, substance addiction, financial hardship, or relationship problems (Lamontagne et al. 2000; Blanco et al. 2008; Schwartz and Wood 1991). For instance, bulimics may steal food to address eating disorders, while people who suffer from anorexia may steal clothing or beauty products to enhance their physical appearance (Schwartz and Wood 1991). Stress can also contribute to shoplifting behaviors. There is evidence that people who are experiencing stressful life events like marital problems, employment or financial problems, or health issues shoplift as a coping mechanism (Schwartz and Wood 1991). These psychological perspectives assert that shoplifting generates excitement and produces an adrenaline rush or high that temporarily subdue feelings of anger, frustration, or depression (Blanco et al. 2008). Shoplifters may also suffer some sort of psychiatric disorder generally related to pathological gambling, bipolar disease, nicotine dependency, or alcohol abuse. In support of psychological perspectives, research has noted that shoplifters are indeed more likely to seek mental health treatment compared to nonshoplifters (Blanco et al. 2008).

Common Antishoplifting Measures and Their Effectiveness

In accordance with the basic tenets of opportunity theories, many antishoplifting measures focus on manipulation of retail environment or target-hardening measures. Clarke's (1997) situational crime prevention model proposes 16 different techniques designed to make crime opportunities less attractive. The various techniques are designed to (a) increase the effort associated with offending, (b) increase the perceived risks of crime, (c) reduce the anticipated rewards of criminality, and (d) remove the excuses to offend. Some of the crime reducing-measures include target-hardening devices, formal and natural surveillance, access control, entry and exit screening, ink security tags, property identification, and warning signs. Several of these techniques are well established within the retail industry. Stores commonly use target-hardening devices like locks, safes, cabinets, mechanical cables, security tags, or benefit denial tags. Most retailers employ various forms of formal surveillance such as store security officers, employees, closed-circuit televisions (CCTVs), or dressing room attendants. Stores display dummy products or empty boxes as a means of removing the benefits of shoplifting. Finally, companies use warning signs about the prosecution of apprehended offenders or signs that indicate a store is under formal surveillance. Collectively, these security measures emphasize the importance of physical store layout, staff and security personnel, and target-hardening devices that help monitor, detect, and apprehend potential shoplifters.

Physical Store Layout

In many ways, the physical layout of a store can facilitate the commission of shoplifting (Carroll and Weaver 1986; Burrows 1998). Most store layouts openly display merchandise and encourage self-service. In addition, the size of most stores makes it difficult to effectively monitor all merchandise and customers simultaneously. Stores typically have multiple exit points, display merchandise near exits, or have high shelves that obstruct viewing angles. Therefore, retail stores provide easy access to valuable and concealable goods in the absence of effective surveillance (Clarke 2003; Tonglet 2002; Cox et al. 1993).

Accordingly, retailers employ a variety of techniques intended to manipulate the store environment, increase the level of surveillance, and reduce the opportunities to steal. For instance, stores create wide aisles to help alleviate congestion and create clear sight lines for employees to easily monitor customer behavior. Stores use convex mirrors and adequate lighting that enables staff to monitor the store from multiple angles. Merchants commonly monitor entry or exit points with employees, detectives, CCTVs, or audible alarm antennas. Smaller items are placed near checkout counters for increased visibility. Finally, stores like to maintain neatly ordered or arranged shelves so that employees can easily notice irregularities (Clarke 2003). While the presence of various security measures is necessary, retailers face the challenge of balancing a customer-friendly environment with an overly intrusive security atmosphere. The inclusion of too many cumbersome, intrusive, or highly visible crime prevention tools can make legitimate customers feel uncomfortable or alienated (Dawson 1993).

Staff and Security Personnel

Research indicates that employees and security personnel are the most important factor in reducing shoplifting (Gills et al. 1999; Clarke 2003; Hollinger and Dabney 1994; Dawson 1993). When sales associates immediately approach and greet customers, make eye contact with them, or consistently offer assistance, it sends a signal to would-be shoplifters that they have been noticed and are being monitored. Indeed, shoplifters report that they are less likely to steal if they believe they are being closely scrutinized or watched by alert employees or store detectives (Gills et al. 1999). Not surprisingly, shoplifters prefer environments in which workers are disinterested in providing quality customer service. They also prefer stores that are routinely understaffed so that employees are too busy to monitor customer behaviors. Furthermore, stores that have high turnover rates or rely heavily on part-time workers undermine effective shoplifting surveillance efforts (Hollinger and Dabney 1994). In these situations, workers are less committed to the company, lack security training, and have limited experience or awareness of shoplifting clues. To be sure, company commitment to equitable treatment of its employees and low turnovers significantly lower levels of inventory loss (Hollinger and Dabney 1994).

However, misuse or abusive employee surveillance can lead to lawsuits. Since untrained employees may misinterpret customer behaviors, they may inadvertently confront and detain innocent customers (Axelrod and Elkind 1976). In addition, employee racial biases or prejudices can affect surveillance efforts. For example, several retail companies including Eddie Bauer, Lord & Taylor, and The Children's Place have settled civil suits amid accusations that store security personnel unfairly monitored, accosted, or apprehended ethnic customers.

These high-profile cases have drawn attention to a phenomena of *consumer racial profiling* (CRP) in which employees and store detectives target shoppers, primarily African Americans, on the basis of race and ethnicity, not necessarily suspicious customer behavior (Gabbidon and Higgins 2007; Asquith and Bristow 2000). Several African Americans report that they have been victims of retail racism in terms of poor service, excessive monitoring, or being looked upon with suspicion by employees or security personnel (Gabbidon and Higgins 2007). Such practices overlook white customers who are equally as prone to shoplift as other customers.

Target-Hardening Devices: Electronic Article Surveillance (EAS) Tags

Introduced in the late 1960s, EAS tags are antishoplifting protection systems that have been widely adopted by the apparel, music, and general retail industries. In general, EAS tags are detachable devices affixed or pinned to merchandise that must be removed or deactivated by a sales associate at the time of purchase. Transmitter stands strategically placed near exit doors are designed to trigger an audible or silent alarm if a customer passes through the transmitter with an active EAS tag. Not only do EAS tags provide physical obstacles to shoplifting, they provide a psychological deterrent as potential offenders are reluctant to risk activating an alarm that alerts store security or employees (DiLonardo 1997; Dawson 1993). A multiyear study of several retail apparel stores found that stores that used EAS tags experienced significant declines in inventory shrinkage, while shortages among retailers without EAS tags actually increased by 30 % during the same time period (DiLonardo 1997). Overall, shoplifting and inventory shrinkage rates have been successfully reduced by 35–75 % after the introduction of EAS systems (DiLonardo 1997).

Other studies, however, have found that EAS tags do not deter shoplifters. Many shoplifters note that EAS tags are easy to remove with pliers or other tools (Gills et al. 1999). In some instances, shoplifters are able to use specially designed foiled bags that stunt the ability of transmitters to sound an audible alarm. More brazen shoplifters simply walk out of the store even if the alarm is sounded. The errant functioning of EAS tags can also contribute to their ineffectiveness. Stores that experience a high number of false alarms triggered by system malfunction or sales associates that fail to remove or deactivate sensor tags risk the potential of offending customers and losing their patronage (Dawson 1993). Legitimate customers are likely to feel embarrassed, annoyed, or upset by inadvertent EAS alarms. Perhaps more importantly, customers may be subject to unwarranted suspicion or scrutiny from security personnel and other customers. Beyond the potential of tarnishing a retailer's image and reputation, false alarms can levy financial costs in the form litigation and punitive damages (Dawson 1993). Another consequence of errant EAS tags is related to the fact that customers, staff, and security may, over time, become desensitized to the alarms.

Despite the best efforts of retailers, evidence suggests few antishoplifting measures including cameras, security tags, mirrors, or prosecution warning signs effectively deter the motivation of pilfers (Gills et al. 1999; DiLonardo 1997). Closed-circuit televisions (CCTVs) are considered ineffective by many shoplifters because they cannot capture quality indisputable images (Gills et al. 1999). Furthermore, shoplifters maintain that most stores have natural blind spots that make it impossible for cameras to monitor. Warning signs that imply all shoplifters will be prosecuted if detected similarly have little impact on the motivation to steal. Shoplifters typically believe the most stores will not go through the trouble of going to court or legal proceeding to recoup relatively inexpensive merchandise (Gills et al. 1999). While some shoplifters acknowledge that ink tags are more challenging to defeat, offenders will still target dye-tagged merchandise primarily because they lack an audible alarm. Once shoplifters effectively remove the merchandise from stores, they have plenty of time to carefully remove the tag (Gills et al. 1999). Even if shoplifters are deterred, the effect is temporary as most will attempt to discover new means to defeat security impediments (Gills et al. 1999). Since the chances of being detected and apprehended are relatively low and the possibility of stiff criminal sanctioning is equally low, shoplifters associate minimal risks with this type of crime and therefore are not easily deterred (Gills et al. 1999; Clarke 2003; Burrows 1988).

Related Entries

- ▶ CCTV and Crime Prevention
- ▶ Crimes of Globalization
- ▶ Effectiveness of Situational Crime Prevention
- ▶ Fencing/Receiving Stolen Goods
- ▶ Situational Crime Prevention

Recommended Reading and References

Asquith JAL, Bristow DN (2000) To catch a thief: a pedagogical study of retail shoplifting. J Educ Bus 75:271–276, May/June

Axelrod AD, Elkind T (1976) Merchants' response to shoplifting: an empirical study. Stanf Law Rev 28:589–612

Blanco C, Grant J, Petry NM, Simpson HB, Alegria A, Liu S-M, Hasin D (2008) Prevalence and correlates of shoplifting in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). Am J Psychiatry 165:905–913

Cameron MO (1964) The Booster and the snitch: department store shoplifting. Free Press, New York

Carroll J, Weaver F (1986) 'Shoplifters' perceptions of crime opportunities: a process-tracing study. In: Cornish DB, Clarke RVG (eds) The reasoning criminal: rational choice perspectives on offending. Springer, New York

Centre for Retail Research (2011) Global retail theft barometer 2011. http://www.retailresearch.org/grtb_globaltrends.php. Retrieved 29 Nov 2011

Clarke RV (1997) Situational crime prevention, 2nd edn. Harrow and Heston, Albany

Clarke RV (1999) Hot products: understanding, anticipating and reducing demand for stolen goods. Police Research Series Paper 112. Home Office, London

Clarke RV (2003) "Shoplifting." Problem-oriented guides for police problem-specific guides series, no. 11. U.S Department of Justice, Office of Community Oriented Policing Services, Washington, DC

Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44:588–608

Cox D, Cox AD, Moschis GP (1990) When consumer behavior goes bad: an investigation of adolescent shoplifting. J Consum Res 17:149–159

Cox D, Cox AD, Anderson RD, Moschis GP (1993) Social influences on adolescent shoplifting – theory, evidence, and implications for the retail industry. J Retail 69:234–246

Cromwell P, Olson JN (2006) Fencing: avenues for redistribution of stolen property. In: Cromwell P (ed) In their own words: criminals on crime, 4th edn. Roxbury, Los Angeles, pp 199–209

Cromwell P, Thurman Q (2003) The devil made me do it: use of neutralization by shoplifters. Deviant Behav 24:535–550

Cromwell P, Parker L, Mobley S (2006) The five finger discount. In: Cromwel P (ed) In their own words: criminals on crime, 4th edn. Roxbury, Los Angeles, pp 113–125

Dabney DA, Hollinger RC, Dugan L (2004) Who actually steals? A study of covertly observed shoplifters. Justice Q 21:693–728

Dawson S (1993) Consumer responses to electronic article surveillance alarms. J Retail 69:353–362

DiLonardo RL (1997) The economic benefit of electronic article surveillance. In: Clarke RV (ed) Situational crime prevention, 2nd edn. Harrow and Heston, Albany

Finklea KM (2011) Organized retail crime. Congressional Research Service. CRS Report for Congress, pp 1–25. www.crs.gov

Gabbidon SL, Higgins GE (2007) Consumer racial profiling and perceived victimization: a phone survey of Philadelphia area residents. Am J Crim Justice 32:1–11

Gills M, Bilby C, Turbin V (1999) Retail security: understanding what deters shop thieves. J Secur Adm 22:29–39

Hayes R, Cardone C (2006) Shoptheft. In: Gill M (ed) The handbook of security. Palgrave Macmillan, New York

Hollinger RC, Dabney D (1994) Reducing shrinkage in the retail store: it's not just a job for the loss prevention department. Secur J 5:2–10

Klemke LW (1992) The sociology of shoplifting: boosters and snitches today. Praeger, Westport

Lamontagne Y, Boyer R, Hétu C, Lacerte-Lamontagne C (2000) Anxiety, significant losses, depression, and irrational beliefs in first-offence shoplifters. Can J Psychiatry 45:63–66

National Association for Shoplifting Prevention (2012) Shoplifting statistics. http://www.shoplifting-prevention.org/main.asp. Retrieved 2 Apr 2012

National Retail Foundation (2011) 2011 organized retail crime survey. Washington, DC. http://nrf.com/modules.php?name=News&op=viewlive&sp_id=1132&parent_id=950&peer_rev=0&nrf_or=0. Retrieved 28 March 2012

Schwartz S, Wood HV (1991) Clinical assessment and intervention with shoplifters. Soc Work 36:234–238

Sykes G, Matza D (1957) Techniques of neutralization: a theory of delinquency. Am Sociol Rev 22:664–670

Tonglet M (2002) Consumer misbehaviour: an exploratory study of shoplifting. J Consum Behav 1:336–354

Signal Crimes Perspective

▶ Reassurance Policing and Signal Crimes

Simulation

▶ Agent-Based Assessments of Criminological Theory

Simulation as a Tool for Police Planning

Vasco Furtado
Graduate Program in Applied Informatics,
University of Fortaleza (Unifor), Fortaleza,
Ceará, Brazil

Synonyms

Crime simulation; MABS; MACS

Overview

Crime relies, directly or indirectly, upon an array of factors, ranging from the levels of concentration of wealth to the physical organization of the urban center under consideration. Modeling the highly interconnected nature of this social system has recently attracted attention in computer science. As experiments in this domain cannot be performed without high risks, because they result on loss of human lives, simulation models have been chosen as supporting tools for this process. Multiagent systems (MAS) primarily study the behavior of autonomous and organized groups of software agents with the purpose of providing solutions to complex problems that could not be achieved by each individual agent alone. Multiagent-based simulation systems have been successfully adopted because the inherent characteristics of the agents (e.g., autonomy, sociability, and pro-activity) facilitate the construction of more dynamic models, thus contrasting with conventional computer simulation approaches.

Other different approaches, dubbed in general as bio-inspired approaches, have recently been investigated in conjunction with MAS for crime modeling. Collective intelligence as that demonstrated by swarms and the evolutionary approach are two of the most prominent concepts and have been used in the design of crime models. Here, an overview of these concepts from a decision support perspective is done in order to describe how they have been applied for the development of multiagent-based crime simulation (MACS). Two particularly relevant issues in the MACS context are discussed: model calibration and model evaluation.

Fundamentals

According to Russell and Norvig (1995), an *agent* is a physical or abstract entity that can be viewed as perceiving its environment through sensors and acting on the environment through actuators. Ferber (Ferber 1999) considers a multiagent system (MAS) as comprising (i) an environment; (ii) a set of passive and active objects (agents); (iii) an assembly of relations, which link objects to each other; and (iv) operations making it possible for the agents to perceive, produce, consume, transform, and manipulate objects. One can distinguish three levels of organization of agents (Rocher and Sherif 1972): micro-social, groups, and global societies. The micro-social level concerns the interactions between agents and the various forms they relate to one another. The level of groups and of societies refers to the dynamics of intermediate structures like organization and cities, respectively. Multiagent systems have been successfully adopted in conjunction with simulation models, which are generally referred to as multiagent-based simulation (MABS) systems. According to Gilbert and Conte (1995), MABS are especially appropriate when one has to deal with interdisciplinary problem domains. Such an approach, which is bottom-up in nature, is also appropriate for the study of social and urban problems, since social or urban environments are dynamic, nonlinear, and composed of a great number of variables and entities. The main objectives behind the construction of MABS systems are the following:

- To test hypotheses related to the emergence of macro-level behavior from interactions occurring at micro-social levels
- To build theories that can contribute to a better understanding of sociological, psychological, and ethological phenomena
- To integrate partial theories coming from different disciplines (e.g., sociology, cognitive psychology, and ethology) in a common theoretical framework

A particular kind of simulation, called geosimulation, addresses an urban phenomena simulation model with a multiagent micro-social approach to simulate discrete, dynamic, and event-oriented systems (Benenson and Torrens 2004). In geosimulated models, simulated urban phenomena are considered a result of the collective dynamic interaction among animate and inanimate entities that compose the map representation. The Geographic Information System (GIS) is responsible for providing the "data ware" in geosimulations.

The study of agent self-organization and related concepts, such as emergence, is another important concept within micro-social MAS. The basic idea is that societies of agents demonstrate intelligent behavior at the collective level out of simple rules at the individual level. Moreover, these individual rules often do not explain the behavior that is attained at the collective level. Swarm intelligence is characterized (i) by strictly local communication, (ii) by the formation of emergent spatial–temporal structures, and (iii) by the agent's making stochastic decisions based on the local information available. One of the branches of swarm intelligence is ant-colony optimization (ACO), proposed by Bonabeau et al. (1999). ACO is a meta-heuristic model for solving combinatorial problems that can typically be represented as graphs. ACO gets inspiration from other areas of science, in this case, biological sciences having the special feature of adapting well to dynamic settings.

In a nutshell, ACO works by allowing agents (ants) to explore a search space, but it requires these ants to leave feedback information about locations with good solutions on the space itself. Agents are then attracted by the feedback left in the environment – the larger the amount of information (pheromone), the more attractive the agents find the position in the environment. In order to avoid early convergence to local optima, the approach assumes that the information left is volatile and impermanent; if no other activity occurs, a piece of information left in the environment "expires" or disappears over a certain period of time. Although never explored as a model for criminal behavior, ACO's characteristics appeared from the start to be an ideal fit to this purpose, as we will discuss later. Briefly, it allows for investigating variations of a learning process with or without a social factor.

Genetic algorithms (GAs) (Holland 1975) are general-purpose search and optimization algorithms that comply with the Darwinian natural selection principle and with some principles of population genetics to efficiently design (quasi-) optimal solutions to complicated computational and engineering problems. Such meta-heuristics maintain a population of chromosomes, which represent plausible solutions to the target problem and evolve over time through a process of competition and controlled variation. The more adapted an individual is to its environment (i.e., the solution is to the problem), the more likely such individual will be exploited for generating novel individuals. In order to distinguish between adapted and non-adapted individuals, a score function (known as fitness function) should be properly specified beforehand in a manner as to reflect the main restrictions imposed by the problem.

Micro-Social Level for Crime Simulation

Despite the existence of several works on MABS that investigate aggregated features of crime by means of macro-simulation, tools for police planning have essentially focused on the micro-social level. These latter follow a conceptual framework defined upon the following features: the environment into which the agent is inserted, the agents (their perceptions, productions, transformations, and objects manipulation), and the interaction between them.

The environment is a space, which generally has a volume (Ferber 1999). The agents are situated in the environment or interact with objects inserted into it. In MACS, the environment can be real or artificial. Typically, when the designer wants to reproduce the real environment, Geographic Information Systems (GIS) using digitalized maps of a geographic area for representing, for instance, streets (Groff 2008), are used. Artificial environments reproduce the main features of a geographic area via abstractions such as grids (Brantingham and Tita 2008) or graphs. Despite the claim of Elffers and Van Baal (2008) that the use of real representation in MACS is not so relevant as the quality of the model, there is a trend in this direction mainly because it facilitates evaluation by visual comparisons. Another way to model the environment is by means of cellular automata (CA) typically to represent static objects. CA is a discrete model of regular grids in which the state of a cell at time t is a function of the status of a finite number of neighborhood cells at time t-1. Examples of environments modeled as a CA can be found in Liu et al. (2005).

The related work on MACS has typically been based on the routine activity theory (RAT) (Cohen and Felson 1979), which states that in order for a criminal act to take place, three elements must coexist: a motivated offender; a suitable target, either an object or a person that can be attacked; and the absence of capable guardians in charge of preventive actions. Therefore, most of the micro-social crime models are based on these three kinds of agents: criminals, guardians, and targets.

On a micro-social level, the criminal agent tries to commit crimes or to move. Each criminal is endowed with a limited view of the environment, measured in terms of a radius following a predefined unit of measure. Criminals have one or more points of departure that we are going to call “gateways.” Such points of departure represent places where criminals are likely to start out, e.g., their residences, metro stations, and bus stops, before committing crimes (Wright and Decker 1994). Target selection is typically probabilistic, based on factors such as target vulnerability, distance between the criminal and the target, and the criminal’s experience. The decision whether to commit a crime or not is made based on the existence of guardians within the radius of the criminal’s sight.

The criminal behavior can be modeled with a learning component that exploits the agent’s experience as well as with information coming from other criminal agents. The success rate of individual agents can be computed as the ratio of the number of successful crimes to the overall number of crimes attempted in their lifetime, as in Furtado et al. (2008), or based on preferences of criminals computed from data mining and based on discrete choice theory (Xue and Brown 2006). Criminals form communities wherein hints are shared. Due to the interconnection of the communities, such hints could be relayed to other criminals in other communities, and the rate at which this happens depends directly on the topology of the network of communities.

Usually there is a set of guardians (police or not) available, each one associated with a target area. A guardian can have a route of length *n*, which is defined as a set, and each component of which is a triple composed by the target area, the interval of time the guardian remains at the target, and the daily period (patrol shift) the routes refer to. Guardians can demonstrate deterministic or stochastic behavior. A deterministic guardian will always move to the same target area and at the same pace predefined as an input parameter. Police guardians are modeled following real data, since these data are available and known by police institutions. However, finding good routes can be an important goal in order to understand the impact of police patrol on crime prevention or reduction. In Reis et al. (2006), a genetic algorithm system called GAPatrol is devoted to the specification of effective police patrol route strategies for coping with criminal activities happening in a given artificial urban environment, which, in turn, mimics a real demographic region of interest. The approach underlying GAPatrol allows for the automatic uncovering of hot spots and routes of surveillance, which, in real life, are usually discovered by hand with the help of statistical and/or specialized mapping techniques.

Notice that, while many simulation studies are aiming at understanding and analyzing the role of various assumptions about the behavior of the respective agents as a function of the ever-changing constellation of other agents and, hence, have a theoretical emphasis (cf. Birks & Elffers, Gerritsen & Klein, this volume), we like to stress here the use of simulations as a means to study the resulting pattern of offender agents and victim agents as a function of various strategies that guardian agents, especially police agents, could have. This use of simulation modeling exploits the method for investigating various possible police strategies on the simulated offence pattern and, hence, has a more applied character.

The locations to be chosen by criminals are referred to as targets, which can be differentiated with respect to their mobility. Commercial/entertainment establishment such as drugstores, banks, gas stations, lottery houses, and malls are fixed, while mobile targets are, e.g., citizens in movement. In Brantingham and Tita (2008), citizen movement is modeled according to the Levy probability distribution, while Furtado et al. (2008) and Liang et al. (2001) have concentrated their study on crimes against property, hence modeling only fixed target.

Targets have a state of vulnerability that can be either active or inactive. A vulnerable target means that it is perceivable by a criminal. Otherwise, it would not take part in the set of high-priority choices of that criminal. In this case, each target must have a probability of being vulnerable, which can follow a distribution based on past real crime data for the associated target type or based on the preference of criminals (Xue and Brown 2006). In Liu et al. (2005), a tension factor was introduced in the model by measuring the impact of crime events on human beings. After a crime in a region, the tension increases, while the vulnerability decreases. The attractiveness of a target can vary depending on cost and reward factors related to the selection of the target such as the location, income, and race composition of the area based on census data. In Furtado et al. (2008), an exponential temporal distribution is used and varies on a daytime basis. For each period and type of target, a value for a configurable parameter, λ, must be determined at the start of the simulation in order to define the pace of occurrence of crimes. For instance, in the evening, drugstore robberies may follow a distribution based on a given value for λ; whereas, in daylight periods, the crime temporal distribution might shift, achieving values four times higher for λ. At any simulation tick, at least one target is made vulnerable in accordance with the temporal distribution associated with its related type. The state of vulnerability is essential as a parameter to control the pace of crimes per type as happens in real life. However, one of the limitations of using values for input parameters from historical real-data analyses, e.g., the pace of crimes (the λ parameter), is that the simulation model will not be capable of identifying a change in the pace of crime occurrences if that change occurs during the simulation time. This problem becomes more significant as the simulation time increases.

Modeling the Interaction Between Agents

Direct interaction means that agents communicate with each other by means of message exchange and/or because they are part of a community or society. In order to be part of a society, it is imperative for an individual to establish social links with other peers. Different forms of interaction among the same individuals, even considering small groups, may take place simultaneously and may vary at different paces through time. One usual means to represent and analyze the (evolving) social structure underlying an organization of individuals is by resorting to the concept of social networks. Roughly speaking, a social network alludes to any formal, graph-based structure where individuals are represented by nodes and the social relationships that unite them are represented by links (ties) between those nodes. The topology of a social network is an important issue to be considered in the analysis thereof, as it helps to determine the network's usefulness (from the viewpoint of the individuals that participate in the network).

The social interaction and learning aspects that underlie criminal activities were investigated in Sutherland's seminal work (Sutherland 1974) in which the differential association theory was proposed. This theory advocates that interaction with others who are delinquent increases the likelihood of someone becoming and remaining a delinquent. That is, peers can play a crucial role in the development of values and beliefs favorable to law violation. In this theory, Sutherland elaborates nine postulates, out of which two are particularly relevant from a perspective of direct agent interaction:

- Criminal behavior is learnable and can be especially learned through the interactions one establishes with other persons, typically through a verbal communication process.
- The main part of the learning of criminal behavior occurs within intimate personal groups.

*A*nother important result coming from works investigating social network models within the context of criminology is that social networks are a natural way to explain the concentration of crimes per area. Crime data analyzed from different regions, and even countries, usually reflect the fact that there are huge spatial (but also temporal) variations in the crime rate between different cities and between different regions in a city. In this regard, Glaeser et al. (1996) show that less than 30 % of the spatial variation of crime (both inter- and intracity) can be explained by differences in local attributes. The remaining 70 % can be explained by social interactions, which means that the agents' decisions about crime are somewhat positively correlated. The authors also show that the impact of social relations is greater in thefts, burglaries, assaults, and robberies (i.e., crimes against property) than in homicides. It is worth mentioning the work of Bosse et al. (2007), which created a model to simulate social learning of adolescence-limited criminal behavior, and the work of Furtado et al. (2008), which designed a model in which social networks are used to model criminal communication. In the latter, the authors showed that the goal of the crime model of generating a power-law spatial distribution of crimes was correlated to the communication aspect modeled via the social network.

Indirect interaction is modeled by means of objects or variables sharing, typically represented in the environment. Other kinds of non-intentional forms of communication that are sent by diffusion or propagation into the environment, like signals, are also used. Ferber (1999) alerts to the limitation that the lack of semantics of signals can provoke. Since the signal is propagated in the environment, all the agents living there can perceive it. A cry of a citizen can be perceived by a guardian as a call for help as well as used by a criminal as a discovery of a potential prey. One important feature of these signals is that their intensity decreases with the distance from the source and with time. The concepts of tension (Liu et al. 2005) and conductivity (Dray et al. 2008) previously mentioned are examples of indirect interaction.

An example of a hybrid approach that uses direct and indirect communication is that of Furtado et al. (2008). Here, communication between the agents was proposed from the concepts of ant-based optimization augmented with a social network. In this model, criminals prefer to commit crime in locations known to be vulnerable, with high payoff, etc. In other words, their choice considers their preference and knowledge about the crime points. The link here to ACO is that, according to this approach, ants always choose their next location in the environment (the place they move toward) biased by a mechanism (the pheromone marks) of indirect interaction. Another indirect communication strategy that ACO offers is that it includes concepts intrinsically related to the notion of "collective." In ACO, ants perform their local search tasks without dictating the whole colony's behavior, which, in turn, is recognized as an emerging result coming from all these local activities. Each criminal has three possible actions: commit a crime, not commit a crime, and move to a certain location. In order to reach a decision whether to commit a crime or not, criminals make use of a probabilistic approach which is adapted from the context of ant-based swarm systems (Dorigo and Stützle 2004).

Key Issues

In MABS, the calibration of the model is a crucial step of the design process (Malleson, this volume). These models are typically characterized by the existence of a lot of parameters, which together determine the general behavior of the model. The development and the parameter setting of MABS models can be long and tedious if there is no accurate and systematic manner to explore the parameter space.

Genetic algorithm has been studied as an alternative for parameter calibration in MABS. The basic idea is to consider the tuning process as an optimization problem. The optimization function for the MABS would be the distance between the artificial model and the real system. In MACS, the calibration of certain parameters in which the complexity is high deserves special attention, e.g., the place from where each criminal starts out to commit crimes. Examples of these initial locations, here called gateways, are bus stops, metro stations, and slums. Usually there is no real data or theoretical model to help one configure those gateways in a crime simulation model. Moreover, it is a combinatorial optimization problem, i.e., the problem of assigning criminals to gateways. More formally, let $G = \{G_i, i = 1, \ldots, N_g\}$ be the set of gateways and $C = \{C_j, j =, \ldots, N_c\}$ be the set of criminals under consideration. The goal is to allocate each C_j to a G_i in a way that a quality measure F, somehow related to the aim of the simulation model, is maximized. In this allocation process, any gateway can be assigned to a criminal and all criminals must be allotted to one, and only one, gateway. Besides, more than one C_j can be assigned to a given G_i (i.e., we have not imposed any limit over the number of criminals assigned to a gateway). Since this assignment problem is combinatorial in nature, the number of feasible gateway configurations is an exponential function of the number of possible gateways.

As important as the calibration of guardians in gateways is the calibration of the number of criminals. There is no real data to help on that and even estimations are very tough to do. Also, this number can change during the simulation or remain constant. Typically in many models, the number of criminals is constant during a simulation. The absence of mechanisms that could implement variability in the number of criminals, such as arrests, can be justified when the ultimate goal is to find good police patrol routes. Considering constant the number of criminals means that crime reduction is only attained by preventing a potential criminal from acting. We could say that, by doing so, there is a preparation of the model to work in the worst of the scenarios, i.e., no way to reduce the number of potential criminals. Thus, the preventive police-planning goal is to define strategies to cover the urban space in a way that could prevent crimes from occurring; events that change the number of criminals are irrelevant.

Social sciences have struggled with this topic due to the difficulty to conduct experiments in controlled environments. MABS have emerged as a tool for social analysis in a way similar to natural controlled experiments. However, model evaluation is one of the biggest challenges of MABS and MACS in particular. A typology for validation must consider the following aspects. The first aspect regards constructing validity to account for the difference between the real world and the rendering of the simulated environment. The challenge here is to design a model representing an approximation that won't be too detailed or similar to the real world because, in this case, the model loses its pragmatic value, and rather than testing theories, it only enumerates what happens under a specific and limited set of conditions. On the other hand, the model must fit the purpose for which it has been created without variables that would bring excessive and unnecessary complexity. The second aspect refers to the internal validity of the model, also called verification. Basically, it refers here to the reliability of the software in generating a determined result from the inputs and processing function for which it was designed. Minor bugs and ill-defined implementations can be responsible for results that falsify the experiments. The third aspect refers to the external validity, meaning how reliable generalizations of the model are for populations lager than the samples. Statistical conclusion validity is another important aspect to be considered in terms of MACS evaluation. Typically, MACS are stochastic and their variation and unpredictability pose problems in the establishment of statistical validation. The identification of regularity that deviates from chance is essential in that context. This must be done across simulation runs in order to be convincing. Finally, they describe the need for models to be evaluated as to their empirical validity via comparisons with real data. In geosimulation, for instance, visual comparisons between hot spots generated from simulations are plotted on maps for comparison with real-data hot spots. Several studies conducted in different countries and different contexts have shown that crime is not uniformly distributed in space and that some victims or targets have a much greater risk of victimization than others (Pease 1998). The temporal dimension also presents a nonuniform distribution with different types of crime having different rhythms (their periodicity) as well as time (their rate of occurrence). In Johnson et al. (2007), an analysis showing how crime clusters in space and time was provided. Following victimization at one's home, those nearby experience an elevated risk of victimization, which decays as time elapses. Another strategy for validating simulation models is to find patterns in real data that indicate crime distribution instead of only relating to the exact numerical values. Doing so, it is possible to compare the results of the simulation with expected distribution of events.

Notice that the use of simulation for investigating possible police strategies is, to some extent, less hampered by the type of problems we have discussed in the present section. In such applications we simply vary the parameters governing the guardian (police) agents, in a way that they represent the strategies to be tested. Actual parameters about the police process (e.g., number of officers available at particular times, priorities) may be available from the police force for whom the simulation is run, thus limiting the variability and hence the number of simulations to be run.

Future Directions

Simulation of criminal activities in urban environments is an asset to decision-makers seeking to find preventive measures. Law enforcement authorities need to understand the behavior of criminals and their response in order to establish safety measures and policies. A conceptual framework for micro-social multiagent-based crime simulation was described involving concepts from Computer Science and AI in particular. Special attention was given to calibration and evaluation aspects, since they constitute open issues demanding further investigations and techniques.

Future investigations in the calibration field are in fact advance in terms of decision support systems, since some parameters like a police patrol route and criminal–gateway composition can shed light on non-understood aspects of preventive policing. In this context, such investigations are expected to provide satisfactory answers to questions like: How far from the optimal patrol routing strategies are those that are actually adopted by human police managers? How complex do such optimal patrolling routes need to be in terms of their total lengths and urban area coverage?

Related Entries

- ▶ Agent-based Modeling for Understanding Patterns of Crime
- ▶ Agent-based Models to Predict Crime at Places
- ▶ Crime Mapping
- ▶ Offender Decision Making and Behavioral Economics
- ▶ Predictive Policing

Recommended Reading and References

Benenson I, Torrens PM (2004) Geosimulation: object-based modeling of urban phenomena. Comput Environ Urban Syst 28(1/2):1–8

Bonabeau E, Dorigo M, Heraulaz G (1999) Swarm intelligence: from natural to artificial systems. Santa Fe Institute Studies in the Sciences of Complexity Series. Oxford Press

Bosse T, Gerritsen C, Treur J (2007) Cognitive and social simulation of criminal behavior: the intermittent explosive disorder case. AAMAS 58:367–374

Brantingham P, Brantingham P (1979) Environment, routine, and situation: toward a pattern theory of crime. In: Clark R, Felson M (eds) Routine activity and rational choice, vol 5. Transaction books, pp 259–294

Calvez B, Hutzler G (2005) Automatic tuning of agent-based models using genetic algorithms. Fourth international joint conference on autonomous agents & multi agent system, Netherlands

Cohen L, Felson M (1979) Social change and crime rate trends: a routine approach. Am Sociol Rev 44:588–608

Dorigo M, Stützle T (2004) Ant colony optimization. The MIT Press, Cambridge

Elffers H, Van Baal P (2008) Realistic spatial backcloth is not that important in agent based simulation research: an illustration from simulating perceptual deterrence. In: Liu L, Eck J (eds) Artificial crime analysis systems: using computer simulations and geographic information systems, pp 19–34

Ferber J (1999) Multi-agent systems: an introduction to distributed artificial intelligence. Addison-Wesley

Furtado V, Melo A, Coelho AL, Menezes R (2008) Simulating crime against properties using swarm intelligence and social networks. In: Liu L, Eck J (eds) Artificial crime analysis systems: using computer simulations and geographic information systems, pp 300–318

Gilbert N, Conte R (1995) Artificial societies: the computer simulation of social life. UCL Press, London

Glaeser E, Sacerdote B, Scheinkman J (1996) Crime and social interactions. Q J Econ 111(2):507–548, MIT Press

Groff E (2008) Characterizing the spatio-temporal aspects of routine activities and the geographic distribution of street robbery. In: Liu L, Eck J (eds) Artificial crime analysis systems: using computer simulations and geographic information systems, pp 226–251

Holland JH (1975) Adaptation in natural and artificial systems. University of Michigan Press

Johnson S, Bernasco W, Bowers K, Elfers H, Ratcliffe J, Rengert G, Townsley M (2007) Space-time patterns of risk: across national assessment of residential burglary victimization. J Quant Criminol 32(3):201–219

Liang J, Liu L, Eck J (2001) Simulating crimes and crime patterns using cellular automata and GIS. UCGIS student award papers 2001. Retrieved from http://www.ucgis.org/f2oppor.html

Liu L, Wang X, Eck J, Liang J (2005) Simulating crime events and crime patterns in a RA/CA Model. In: Wang F (ed) GIS and crime analysis. Idea Group, pp 197–213

Pease K (1998) Repeat victimization: taking stock. Crime detection and prevention series paper 90. Home Office, London

Reis D, Melo A, Coelho ALV, Furtado V (2006) GAPatrol: an evolutionary multiagent approach for the automatic definition of hotspots and patrol routes. In: Sichman JS, Coelho H, Oliveira S (eds) Proceedings of IBERAMIA/SBIA 2006, Lecture Notes in Artificial Intelligence (LNAI) 4140, pp 118–127

Rocher G, Sherif P (1972) A general introduction to sociology: a theoretical perspective. MacMillan Canada

Russell S, Norvig P (1995) Artificial intelligence: a modern approach. Prentice Hall Series in AI, New Jersey

Scott JP (2000) Social network analysis: a handbook. Sage

Snook B (2004) Individual differences in distance travelled by serial burglars. J Invest Psychol Offender Profiling 1:53–66

Sutherland E (1974) Principles of criminology, 4th edn. J. B. Lippincott, Philadelphia

Wright R, Decker S (1994) Choosing the target. Burglars on the job: street life and residential break-ins. Northeastern University Press, Boston

Xue Y, Brown D (2006) Spatial analysis with preference specification of latent decision makers for criminal event prediction. Decis Support Syst 41(3):560–573

Situational Action Theory

Per-Olof H. Wikström
Institute of Criminology, University
of Cambridge, Cambridge, UK

S

Synonyms

Analytical criminology; Causal processes; Crime; Deterrence; Morality; Moral norms; Self-control

Overview

This chapter introduces Situational Action Theory (SAT) and its key arguments:

Crimes are moral actions. There are many different types of actions that may constitute a crime, but what they all have in common is that they break a rule of conduct (stated in law) about what is the right or wrong thing to do (or not to do). What distinguish crime is thus not that they are a particular kind of action but that they are acts the breach rules of conduct (stated in law) and, therefore, need to be studied and explained as such.

People are the source of their actions, but the causes of their actions are situational. To explain the causes of acts of crime we need to understand the situational factors and processes that move people to break rules of conduct (stated in law). People are different and so are also the settings (environments) to which they are exposed. Acts of crime (particular acts of crime) occur when particular kinds of people (person propensities) encounter particular kinds of settings (environmental inducements) creating specific kinds of interactions that make people see certain action alternatives and make certain choices.

Social and developmental factors and processes are best studied and explained as causes of the causes. To explain the role of social and developmental factors and processes in crime causation, we need to understand how they (as causes of the causes) affect people's crime propensities and how they impact settings (environments) levels of criminogeneity and how they influence the spatiotemporal convergence of crime prone people and criminogenic settings (creating the interactions that may result in acts of crime or certain types of acts of crime).

Situational Action Theory: Background and Main Aims

Situational Action Theory (SAT) aims to explain *why* crime happens and the role of social and developmental factors and processes in crime causation (e.g., Wikström 2006, 2010; Wikström et al. 2012, pp. 3–43). It was specifically designed to overcome some recurrent problems in criminological theorizing such as the unclear definition of crime (i.e., the ambiguity about what a theory of crime aim to explain), the lack of a satisfactory action theory (i.e., the poor understanding of what it is that moves people to engage in acts of crime), and the poor integration of levels of explanation (e.g., Wikström 2010, pp. 212–216; Wikström et al. 2012, pp. 3–10). The argument is not that all criminological theories fail on all these points, but it is just that they generally fail on one or more of these points.

SAT seeks to integrate key insights from main person and environment-oriented criminological theories and research, and relevant social and behavioral science theories and research more generally, within the framework of an adequate action theory. SAT advocate an analytical criminology, a criminology that focuses on the identification and detailing of the key mechanisms involved in crime causation rather than the production of lists of statistically relevant correlates and predictors (often referred to as risk factors) of which most, at best, are likely to be only markers or symptoms and, therefore, lack any causal relevance (e.g., Wikström 2006, pp. 66–69, 2011, pp. 53–60).

Basic Assumptions About Human Nature and Social Order

Situational Action Theory is based on the key assumptions that humans are essentially *rule-guide actors* and that social order is fundamentally based on people's adherence to shared rules of conduct. People vary in their personal morals (personal rules of conduct) and settings vary in their moral norms (shared rules of conduct). A personal moral rule is held and enforced (through the process of self-control) by the actor, while a moral norm is held and enforced (through the process of deterrence) by (significant) others. Explaining human action (such as acts of crime) requires understanding how people's actions are influenced by the interplay between personal morals and perceived moral norms of the setting (environment), a problem that has been largely ignored in criminological theory.

SAT assert that humans have *agency* (powers to make things happen) but that they exercise agency (i.e., express their desires (needs) and their commitments and respond to frictions) within the constraints of rule-guided choice. The theory acknowledges that there are elements of predictability (habit) and “free will” (deliberation) in people’s choices. Explaining human action (such as acts of crime) requires an understanding of the role of agency (how it works) within rule-guided choice, a problem that has seldom (if at all) been treated in criminological theory.

Crime as Moral Actions

Criminological theories rarely specify or clearly analyze what it is they aim to explain. This is an important common omission since an explanation has to explain *something*. Without clearly defining what it is that is to be explained, it is difficult to unambiguously identify putative causes and suggest plausible causal processes that may produce the effect under study (e.g., acts of crime). A cause has to be a cause of something, and what that something is determines possible causes (and relevant causal processes).

Situational Action Theory argues that crime is best analyzed as moral actions. SAT defines morality as value-based rules of conduct about what is the right or wrong thing to do (or not to do) in a particular circumstance. *Crime* is an act that breaks a rule of conduct stated in law (rules that may be quite general or quite specific). What defines crime is thus not any particular type of action but the fact that carrying out a particular action (or refraining from carrying out an action) in a particular circumstance is regarded as breaching a rule of conduct (stated in the law).

Any particular action can, in principle, be defined as a crime, and there are variations over historic time and between places (e.g., countries) in what kinds of action are regarded as crime. Moreover, specific actions, like hitting or even shooting another person, may be considered crimes in some circumstances but not in others. The advantage of conceptualizing crime as breaches of rules of conduct (stated in law) is that it makes a general theory of crime possible by focusing on what all kinds of crime, in all places, at all times, have in common, namely, the rule-breaking. What is to be explained by a theory of crime causation is thus why people (follow and) breach rules of conduct stated in law. A theory of crime causation may be regarded as a sub-theory of a more general theory of moral action. The law is just one of many sets of rules of conduct that guide people’s action (e.g., Ehrlich [1936] 2008). The law is no different from other sets of rules of conduct; in fact, the law may be regarded as a special case of rules of conduct more generally. Explaining why people follow and break the rules of law is, in principle, no different from explaining why people follow and break rules of conduct more generally.

Analyzing crime as moral action does not imply a “moralistic” approach (no judgement whether existing laws are good or bad, just or unjust, needs to be made). However, SAT does not assume moral relativism (i.e., that all moral rules are equally likely to emerge). Common moral rules may be grounded in human nature and the problem of creating social order (see further Wikström et al. 2012, pp. 13–14). The explanation of why we have particular moral norms (and laws) and why people follow or breach moral norms (and laws) is different analytical problems. Importantly, people may breach a moral rule of conduct (stated in law) because they disagree with or care less about the rule (or their ability to exercise self-control is not strong enough to make them adhere to their own personal morals when faced with a temptation or provocation).

The Situational Model: Kinds of People in Kinds of Settings

People are different and so are the settings to which they are exposed. When a particular kind of person is exposed to a particular kind of setting, a particular situation (perception-choice process) arises that initiates and guide his or her actions in relation to the motivations he or she may experience. The elements of the situational model are the *person* (his or her relevant propensities), the *setting* (its relevant inducements), the *situation* (the perception-choice process that arises from the exposure of a particular person to a particular setting), and *action* (bodily movements such as speaking, walking, or hitting). A setting is defined as the part of the environment (objects, persons, events) that is directly accessible to the person through his or her senses (including any media present).

SAT proposes that variations between people in their *crime propensity* (i.e., the tendency to see, and choose, particular crimes as an action alternative) is essentially a question of their *law-relevant personal morality* (the extent to which their personal morality corresponds to the various rules of conduct stated in the law) and their ability to exercise self-control (which depends both on dispositional characteristics such as executive functions, and momentary influences such as intoxication and levels of stress – see Wikström and Trieber 2007).

People will vary in their propensity to see a particular kind of crime as an action alternative not only depending on their personal moral rules but also depending on their moral emotions. *Moral emotions* (shame and guilt) attached to violating a particular rule of conduct may be regarded as a measure of the strength with which a person holds that particular rule of conduct. For example, while many people may think it is wrong to steal something from another person, some may feel very strongly about this while others may not. Those who feel less strongly about stealing from others may be regarded as having a higher propensity to perceive such action as an option.

People vary in their *ability to exercise self-control* (i.e., their ability to act in accordance with their personal morality when it conflicts with a perceived moral norm of a setting). The ability to exercise self-control exerts its influence through the process of choice and is only important in the explanation of crime when a person deliberate over several potent action alternatives, of which at least one involve an act of crime. The concept of self-control in SAT is different from that of self-control in Gottfredson and Hirschi's (1990) General Theory of crime (see further, e.g., Wikström 2010, pp. 228–234). Crucially SAT makes a clear distinction between self-control as situational process (the process by which a person manage conflicting rule guidance) and people's abilities to exercise self-control.

SAT further proposes that the *criminogeneity of a setting* depends on its (perceived) moral norms (the extent to which it encourage or discourage the breaking of particular laws in relation to the opportunities a setting provides and the frictions it creates) and their level of enforcement through the process of deterrence (note that if a moral norm encourages the breaking of a particular law, a high degree of its enforcement will be criminogenic). *Deterrence* is defined as "the process by which the (perceived) enforcement of a setting's (perceived) moral norms, by creating concern or fear of consequences, succeeds in making a person adhere to the moral norms of the setting even though they conflict with his or her personal moral rules."

Key Situational Factors and the Situational Process

The cornerstone of Situational Action Theory is that people's actions ultimately is a consequence of how they perceive their action alternatives and, on that basis, makes their choices when confronted with the peculiarities of a setting. The perception-choice process is crucial for understanding a person's actions. *Perception* (the selective information we get from our senses) is what links a person to his or her environment, and *choice* (the formation of an intention to act in one way or another) is what links a person to his or her actions (see further Wikström 2006, pp. 76–84). In contrast to most choice-based theories of action, which focus on how people choose among predetermined alternatives, Situational Action Theory stresses the importance of why people perceive certain action alternatives (and not others) in the first place. Perception of action alternatives thus

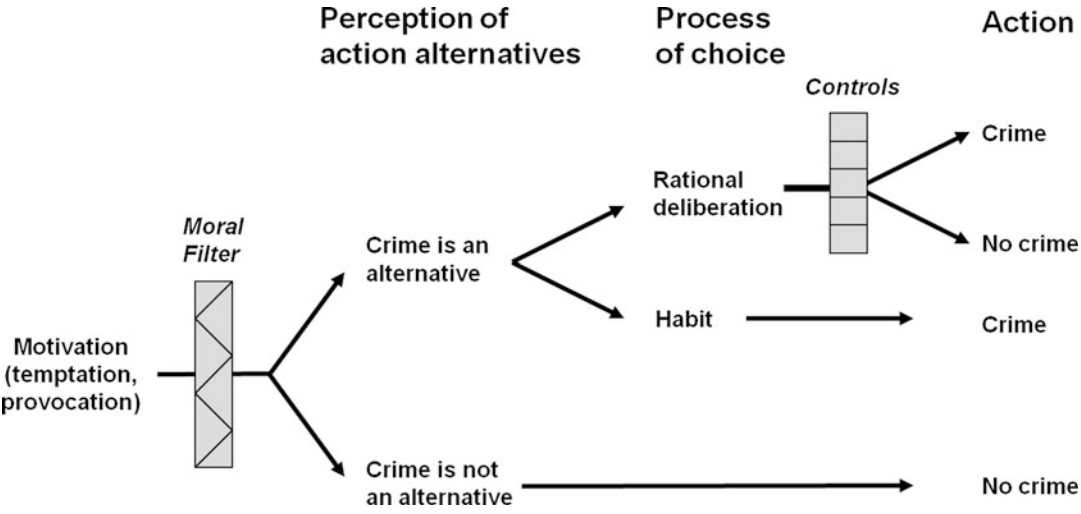

Situational Action Theory, Fig. 1 The situational process and key situational factors (applied to the explanation of crime) (Source: Wikström 2011)

plays a more fundamental role in explaining actions, such as acts of crime, than the process of choice (which is secondary to perception of action alternatives). The key stages of, and the key situational factors in, the perception-choice process, according to SAT, is illustrated in Fig. 1.

Motivation (defined as goal directed attention) is a situational factor that initiates the action process and is an outcome of the interaction between the person (preferences, commitments, sensitivities) and the setting (opportunities, frictions). According to SAT there are two main kinds of motivators; (1) temptations, which are either the outcome of the interaction between (a) a person's desires (wants, needs) and opportunities to satisfy a desire (want, need) or (b) the outcome of the interaction between a person's commitments and opportunities to fulfill a commitment (opportunities to satisfy a desire or to fulfill a commitment may be legal or illegal), and (2) provocations, which occur when a friction (an unwanted external interference) causes anger or annoyance towards the perceived source of the friction or a substitute. People vary in their sensitivities to particular kinds of frictions (as a consequence of their cognitive-emotive functioning and life-history experiences).

Motivation is a necessary but not sufficient factor in the explanation of crime. What particular action alternatives a person perceives in relation to a particular motivation (and whether that includes an act of crime) is dependent on the moral filter. The *moral filter* is defined as "the moral rule-induced selective perception of action alternatives in relation to a particular motivation" and is an outcome of a person's moral engagement with the perceived moral norms of a setting in relation to a specific motivation. If crime is not among the action alternatives a person perceives, there will be no crime and the process of choice is irrelevant as an explanation of why he or she refrained from committing an act of crime; he or she simply did not see crime as an option.

If an act of crime is among the perceived action alternatives the process of choice will determine whether or not a person will commit (or attempt) an act of crime. SAT distinguishes between two main kinds of processes of choice: habitual (automated) or deliberative processes of choice (in prolonged action sequences the action guidance may drift between habitual and deliberative influences). The important difference between the two processes is that in case of actions (such as acts of crime) committed out of habit, the actor only sees *one* potent alternative in response to a motivation (temptation or provocation) and act accordingly, while in the case of deliberation there is no predetermined response so the actor has to weigh pros and cons of the different perceived alternatives (a process that may be more or less elaborate dependent on the importance of the choice). There is plenty of evidence for the existence of a dual process of human reasoning of this kind (see, e.g., Evans and Frankish 2009). The neuropsychological basis of SAT's framework for habitual and deliberate decision making, linking these processes to areas of the brain's prefrontal cortex which have been associated with separate intuitive and cognitive functions, with particular implications for the role of self-control and emotions in action decisions, has been specifically explored by Treiber (2011).

When people act out of *habit*, they essentially react (in a stimulus-response fashion) to environmental cues, automatically applying experienced based moral rules of conduct to the peculiarities of a setting, while *deliberation* involves taking moral rules of conduct into consideration when actively choosing between perceived action alternatives. Habitual action is oriented towards the past, while deliberate action is oriented towards the future. When people act habitually, they routinely apply past experiences to guide current action (i.e., they do what they normally do in the particular circumstance without giving it much thought); when they act deliberately, they try to anticipate future consequences of perceived action alternatives and choose the best course of action. When people act out of habit (only see one causally effective alternative; although they may be loosely aware "in the back of their minds" that there are other alternatives, no such alternative is actively considered), rationality does not come into play (since there is only one potent alternative and, therefore, no weighing of pros and cons of different options) and their actions may even be irrational, that is, people may act in ways they would not consider in their best interest had they deliberated. When people deliberate they aim to be rational (to choose the best option as they see it), but, importantly, this does not necessarily involve an aim to maximize personal advantage (whether maximizing personal advantage is seen as the best option is a question of the actor's moral judgement). Habitual responses are most likely when people operate in familiar circumstances with congruent rule guidance (i.e., when there is a high level of correspondence between personal morals and perceived moral norms of the setting), while deliberate responses are most likely when people are in unfamiliar circumstances and/or there is a conflicting rule-guidance.

Only in choice processes when people deliberate do they exercise "free will" (although it is a "free will" within the constraints of perceived action alternatives) and are they subject to the influence of their ability to exercise self-control (internal control) or respond to deterrence cues (external controls). If there is no conflicting rule guidance, there is nothing to control, and, hence, in these cases (inner and outer) controls lack relevance for the action taken. A person's ability to exercise self-control is a relevant factor in crime causation when his or her personal moral rules discourage and the moral norms of the setting encourage an act of crime in response to a motivation. In this case, the extent to which a person can act in accordance with his or her personal morals (i.e., refrain from crime) is dependent on the strength of his or her ability to exercise self-control. Deterrence is a relevant factor in crime causation when a person's personal morals encourage and the moral norms of a setting discourage an act of crime in response to a motivation. In this case, whether or not a person will carry out an act of crime is dependent on the (perceived) efficacy of the enforcement of the moral norms of the setting.

SAT offer a general explanation of key situational factors and processes involved in crime causation. What may differ in *the explanation of different kinds of acts of crime* is not the process (the perception-choice process) leading up to the action, but the input to the process, that is, the content of the moral context (the action-relevant moral norms) and a person's morality (the action-relevant moral values and emotions) which drive the process. For all kinds of crime to happen (e.g., shoplifting, insider trading, rape, or roadside

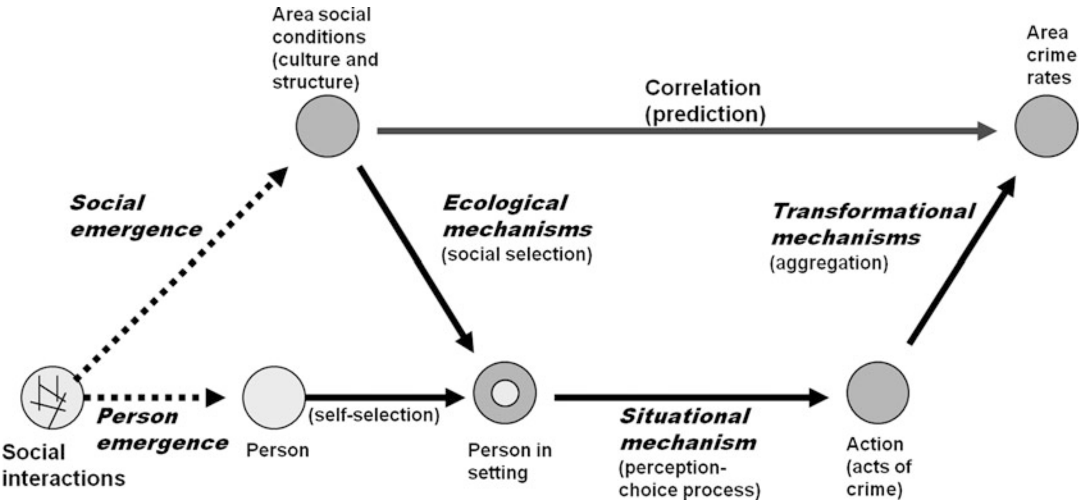
Area social conditions (culture and structure)
Correlation (prediction)
Area crime rates
Social emergence
Ecological mechanisms (social selection)
Transformational mechanisms (aggregation)
Person emergence
(self-selection)
Situational mechanism (perception-choice process)
Social interactions
Person
Person in setting
Action (acts of crime)

Situational Action Theory, Fig. 2 The social and situational models of SAT integrated (Source: Wikström 2011)

bombing), the actor has to perceive the particular action as a viable alternative and choose to carry out the act. However, the specific moral background (relevant personal morals and moral norms of the setting) which guide whether, for example, an act of shoplifting is perceived as an action alternative may differ from that which guides whether an act of rape is perceived as an action alternative.

The Social Model

Situational Action Theory insists that the causes of crime are situational and that social and developmental factors and processes in crime causation are best studied and analysed as the *causes of the causes* (of acts of crime). SAT advocates a mechanistic explanation of human action (such as acts of crime). The theory is based on four key propositions (of which the first two refer to the situational model and the subsequent two to the social model):

1. *Action is ultimately an outcome of a perception-choice process.*
2. *This perception-choice process is initiated and guided by relevant aspects of the person-environment interaction.*
3. *Processes of social and self-selection place kinds of people in kinds of settings (creating particular kinds of interactions).*
4. *What kinds of people and what kinds of environments (settings) are present in a jurisdiction is the result of historical processes of person and social emergence.*

Figure 2 illustrates how the social and situational models are integrated. The key content of the situational model has already been discussed, and the remaining part of this paper will be devoted to a brief outline of the role of processes of emergence and selection in crime causation.

The social model of SAT focuses on the role of historical processes of emergence in the creation of criminogenic environments (social emergence) and crime prone people (person emergence) and contemporaneous processes of self- and social selection that bring together crime prone people and criminogenic settings (creating the situations to which people may respond to motivators by committing acts of crime). The concept of *emergence* refers to how something becomes as it is (e.g., Bunge 2003). For example, how people acquire a certain crime propensity (personal emergence) or environments acquire a certain criminogeneity (social emergence) as an outcome of processes of social interactions. The concept of *selection* refers to the contemporaneous socio-ecological processes responsible for introducing particular kinds of people to particular kinds of settings (and thus creating the situations to which people's actions are a response).

SAT proposes that *psychosocial processes* of moral education and cognitive nurturing are of central interest in the explanation of why people develop specific and different crime propensities (i.e., tendencies to see and choose particular crimes as an action alternative) (see further Wikström et al. 2012, pp. 31–32). SAT further proposes that *socio-ecological processes* (e.g., processes of segregation and their social consequences) become of particular interest in the explanation of why particular kinds of moral contexts emerge in particular places at particular times (see further Wikström et al. 2012, pp. 2–37).

Why certain kinds of people end up in certain kinds of setting depends on processes of social and self-selection and their interaction. *Social selection* refers to the social forces (dependent on systems of formal and informal rules and differential distribution of personal and institutional resources in a particular jurisdiction) that encourage or compel, or discourage or bar, particular kinds of people from taking part in particular kinds of time and place-based activities. *Self-selection* refers to the preference-based choices people make to attend particular time and place-based activities within the constraints of the forces of social selection. What particular preferences people have developed may be seen as an outcome of their life-history experiences. Depending on the circumstances, social or self-selection can be more influential in explaining why a particular person takes part in a particular setting (see further Wikström et al. 2012, pp. 37–41).

Analyzing relevant processes of emergence and selection and how they relate help us understand (i) why people develop certain kinds of crime propensities (and why a population in a particular jurisdiction come to have a certain distribution of crime propensities and why some jurisdictions have more crime prone people than others), (ii) why certain places become criminogenic environments (and why there is a certain prevalence and kind of criminogenic environments in a jurisdiction and why some jurisdictions have more criminogenic environments than others), and (iii) why certain contemporaneous self- and social selection processes create particular kinds of crime hot spots (and why the prevalence and kind of crime hot spots vary between different jurisdictions).

Related Entries

- ▶ Criminology of Place
- ▶ Rational Choice Theory
- ▶ Routine Activities Approach
- ▶ Situational Crime Prevention
- ▶ Social Control and Self-control Through the Life Course

Recommended Reading and References

Bunge M (2003) Emergence and convergence: quantitative novelty and the unity of knowledge. University of Toronto Press, Toronto

Ehrlich E (1936) 2008. Fundamental principles of the sociology of law. New Brunswick: Transaction Publishers

Evans J, Frankish K (2009) In two minds: dual processes and beyond. Oxford University Press, Oxford

Gottfredson M, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Treiber K (2011) The neuroscientific basis of situational action theory. In: Beaver KM, Walsh A (eds) The Ashgate research companion to biosocial theories of crime. Ashgate Publishing Company, Burlington

Wikström P-O (2006) Individuals, settings and acts of crime. Situational mechanisms and the explanation of crime. In: Wikström P-O, Sampson RJ (eds) The explanation of crime: context, mechanisms and development. Cambridge University Press, Cambridge

Wikström P-O (2010) Explaining crime as moral action. In: Hitlin S, Vaysey S (eds) Handbook of the sociology of morality. Springer, New York

Wikström P-O (2011) Does everything matter? Addressing the problem of causation and explanation in the study of crime. In: McGloin J, Sullivan CJ, Kennedy LW (eds) When crime appears: the role of emergence. Routledge, London

Wikström P-O, Treiber K (2007) The role of self-control in crime causation. Beyond Gottfredson and Hirschi's general theory of crime. Eur J Criminol 4:237–264

Wikström P-O, Oberwittler D, Treiber K, Hardie B (2012) Breaking rules. The social and situational dynamics of young people's urban crime. Oxford University Press, Oxford

Situational Approaches to Terrorism

Graeme R. Newman
School of Criminal Justice, University at Albany, State University of New York, Albany, NY, USA

Overview

The situational approach to terrorism evolved from situational crime prevention, a well-established evidence-based approach to preventing and reducing crime. It argues that terrorists make choices that are limited by their perception of the opportunities afforded to them to carry out their mission. The situational approach explains *how* terrorists carry out their missions in contrast to *why* they do so. There are four pillars of terrorist opportunity: targets, weapons, availability of tools for doing attacks, and local conditions that facilitate terrorist operations. By manipulating these opportunities, interventions are directed at increasing the effort of mounting an attack, increasing the risk of carrying out an attack, reducing the rewards of attacks, reducing provocations to terrorists, and removing excuses terrorists use to justify their attacks. Because the situational approach focuses primarily on prevention, it places great emphasis on planning, which makes it a natural supplement to disaster response planning and to problem-oriented policing that seeks to solve crime and disorder problems. It is thus particularly useful to local police as a guide for developing local counterterrorism strategies.

Introduction

Situational approaches to terrorism derive from several decades of research, theory, and practice in the field of situational crime prevention. Pioneered by Ronald V. Clarke in 1988 in his seminal paper on coal gas suicides, situational crime prevention has evolved in many respects outside of mainstream criminology, which, since its inception in the nineteenth century, has been preoccupied with the root causes of crime, the search for the "born" or "made" criminal. Dominated by political science, the approach to terrorism has paralleled that of criminology: the search for the root causes of terrorism, with an overwhelming focus on its ideological, psychocultural, economic, and political causes. In contrast, the situational approach takes up where mainstream explanatory models leave off: it begins with a pinpoint focus on the specific situations in which terrorism occurs and, depending on the type of terrorism, works backwards to the causes. It does not completely eschew root causes; it views them as distant, background factors which for the most part are impermeable to counterterrorism interventions (see Fig. 1).

This conceptual distinction between situational crime prevention and other approaches in social science drives it to ask questions that others do not: how did the offender get to the point of being able to carry out his crime? What immediate factors in the social and physical environment produced a situation in which the crime was possible? Situational crime prevention asks not why but how an offender carries out his tasks. It is concerned with *motives* rather than *motivation*. It is concerned with *intervention* rather than *causation*. It takes seriously the old adage "prevention is better than cure."

Situational crime prevention is often attacked by those who hold to a contrasting view: "You can't cure an illness (disease, crime, terrorism) by treating its symptoms." The symptoms may disappear, but then reappear in some other form because the root causes have not been diagnosed. Counterterrorist experts complain, "If we protect this government building, they'll attack a shopping mall." This seems like a devastating blow to the situational crime prevention approach, but it is not as will be seen later in this entry. First, it is necessary to address the basic, intractable problem in the study of the causes and response to terrorism: its definition.

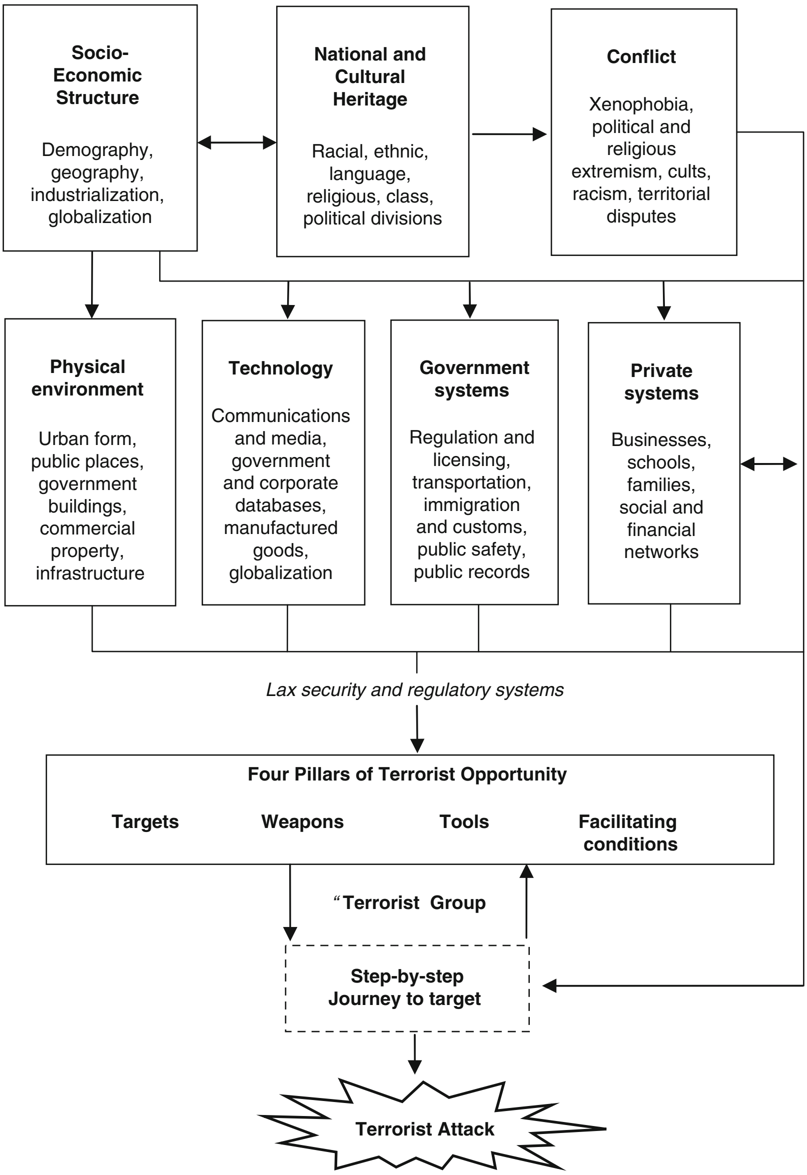
Socio-Economic Structure
Demography, geography, industrialization, globalization
National and Cultural Heritage
Racial, ethnic, language, religious, class, political divisions
Conflict
Xenophobia, political and religious extremism, cults, racism, territorial disputes
Physical environment
Urban form, public places, government buildings, commercial property, infrastructure
Technology
Communications and media, government and corporate databases, manufactured goods, globalization
Government systems
Regulation and licensing, transportation, immigration and customs, public safety, public records
Private systems
Businesses, schools, families, social and financial networks
Lax security and regulatory systems
Four Pillars of Terrorist Opportunity
Targets
Weapons
Tools
Facilitating conditions
“Terrorist Group
Step-by-step Journey to target
Terrorist Attack

Situational Approaches to Terrorism, Fig. 1 The terrorist opportunity structure (Adapted from *Outsmarting the Terrorists*, p. 8)

The Situational Definition of Terrorism

The definition of terrorism is famously elusive. The United Nations has tried for many years to agree on a definition and failed. The problem is political, legal, and conceptual. The political difficulty is reflected in the popular saying, “one man’s terrorist is another man’s freedom fighter.” The conceptual problem is tied up with the legal (which essentially binds the conceptual to the political). For example, serious political disagreements arose in the United States when the Obama administration chose to define the terrorist attacks on the World Trade Center as criminal acts and therefore subject to the regular due process of US criminal law. Opponents of that view argued that they were not criminal acts but acts of war against the USA, so that other legal procedures should follow. Others claimed that the attacks were not acts of war or criminal acts but something else – the position eventually favored by the Bush administration and resulted in setting up the prison in Guantanamo Bay, a kind of legal "nowhere land."

The situational approach finds these debates as mostly irrelevant to its enterprise. In fact, as noted earlier, situational crime prevention began with a study of suicide, an act arguably not always a crime, depending on the jurisdiction and circumstances of its commission. As it evolved, situational crime prevention developed more and more a focus on the specificity of the crime which often cuts across the traditional definitions of crime. This had important ramifications in respect to how crime was measured and recorded. Simply recording and reporting the number of burglaries, for example, in a police jurisdiction, were found to be useless for research or for prevention. It was necessary to collect very specific information about very specific types of burglaries: single-family residences or apartment blocks? Suburban or urban neighborhoods? Commercial or business districts? Adjacent to parks or woods? and so on. Generally speaking, it is the definition of the situation that is important for both research and prevention purposes, rather than the definition of the crime. So for the purposes of situational crime prevention, it does not really matter what one calls it, a crime, terrorist attack, or even an accident; it is the analysis of the situation in which these events occur that is of prime importance. As a result, Clarke and Newman in their book *Outsmarting the Terrorists* sidestep the issue by defining terrorism simply as "crime with a political motive."

The Situational Approach to Terrorism

Apart from the importance of specificity already discussed, there are three basic operating principles that play an important part in any analysis of terrorism from the situational perspective. These are opportunity, rational choice, and intervention.

1. *Opportunity*. This principle simply states that in order for a terrorist attack to be accomplished, the environmental and social conditions must be such that they offer advantage to the terrorist for carrying out the attack. For example, before there were cars, car theft was not possible. Before there were computers, software piracy was not possible. It is true that "theft" and "piracy" in the generic sense have always been possible, but it is clear that to speak of preventing or even analyzing such crimes, it makes little sense without examining their specific attributes. Furthermore, the social and physical environment may be structured in such a way that easy access to the targets of crime or terrorism is available. For example, early computers or cars were never designed in anticipation of software piracy or car theft. The mosaic of factors that make various crimes or terrorist attacks possible is called the opportunity structure as shown in Fig. 1.
2. *Rational choice*. Depending on what the terrorist wants to achieve by his attack (i.e., his motives), the terrorist will carry out his enterprise according to a rational choice from available options (i.e., part of the opportunity structure). The observer may consider what the terrorist is doing to be "irrational" such as the terrorist who kills himself in order to successfully complete an attack. This is called "limited" rationality as expounded by Cornish and Clarke in their groundbreaking work *The Reasoning Criminal*. Failure to understand the way terrorists think contributed to the shortsighted US counterterrorist policy which assumed that terrorists would not hijack an airplane and willingly kill themselves in the process – the unanticipated methodology of the 9/11 attacks.
3. *Intervention*. Situational crime prevention always begins from the point of how to disrupt the sequence of events that result from the offender's decision making. This involves generally two procedures for revealing points of weakness in the offender's or terrorist's action sequence. First, a step-by-step analysis is conducted to find out exactly the sequence of choices made by the terrorist in his journey to the target. For example, in the case of a suicide bomber, we must follow his journey from induction into the pool of potential bombers, his training, the selection of the target, and finally his journey to the target (see Fig. 2).

Situational Approaches to Terrorism, Fig. 2 Suicide bombing step-by-step (Adapted from *Outsmarting the Terrorists*, Table 5.3)

1. *Preparation*: Find safe house(s) for operations → Obtain or manufacture bomb paraphernalia → Select target or targets → Recruit bomber candidates → Specify exact location for detonation → Specify route to target → Establish group commitment → Train bombers → Prepare propaganda → reinforce individual commitment →
2. *Operation*: Intelligence gives "all clear" to target → Dispatch bomber → Bomber journeys to target →
3. *Follow up*: Claim responsibility → Broadcast propaganda → Review operation → Plan new attack → Establish supply chain for weaponry → To step 1

The journey to crime has been studied for a wide range of crimes including burglary, car theft, sex offenses, and robbery. Recent work emphasizing the step-by-step approach has adopted the concept of "scripting" from cognitive psychology which examines in detail the decision-making process and techniques used by offenders to exploit opportunities and accomplish their task. This approach has been used in analyzing credit card and check fraud, suicide bombing, Internet child pornography, identity theft, and trafficking in endangered species. The second important aspect of intervention is to conduct a careful assessment of the opportunity structure to identify weak points of security that may be exploited by criminals or terrorists. For example, particular targets may be more vulnerable, exposed, or accessible than others. Once these have been identified, the appropriate security systems may be installed. This is an important feature of situational crime prevention because it implies that there is much that can be done to protect targets even without knowing who the terrorists are or what their motivation may be.

Guided by the three principles of opportunity, rational choice, and intervention, Professors Clarke and Newman have developed a comprehensive approach to explaining terrorism which also provides a structure for organizing responses to terrorism. This helps to connect the often disjointed study of terrorism that focuses either entirely on the terrorist's characteristics or entirely on law enforcement responses to those characteristics. Instead, Clarke and Newman advocate that we "think terrorist" in order to identify a modus operandi according the opportunities available. The terrorist is placed within the opportunity structure of terrorism which is primarily composed of what they call the four pillars of terrorist opportunity: targets, weapons, tools, and facilitating conditions.

The Four Pillars of Terrorist Opportunity

1. Targets

In theory, terrorists could attack any target, as many politicians and law enforcement individuals have worried ever since 9/11. In practice, they must choose their targets carefully. And this choice will be strongly conditioned by the inherent attractiveness of those targets to terrorists, summarized as EVIL DONE:

Exposed: The Twin Towers were sitting ducks.

Vital: Electricity grids, transportation systems, and communications are vital to all communities.

Iconic: Of symbolic value to the enemy, for example, Statue of Liberty, the Pentagon, and the Twin Towers.

Legitimate: Terrorists' sympathizers cheered when the Twin Towers collapsed.

Destructible: The Murrah building in Oklahoma City and Twin Towers on 9/11.

Occupied: Kill as many people as possible.

Near: Within reach of terrorist group and close to home base.

Easy: The Murrah building was an easy target to a car bomb placed within 8 feet of its perimeter.

Of the above characteristics, the proximity of the target to the base of operations (near) is probably the most important. Research has noted that terrorists will base their operations in regions and countries where the local population is sympathetic to their cause. As noted earlier, the importance of proximity to target and base of operations has been known for some time in situational crime prevention research on the offender's journey to the crime. The reason is that the closer to base of operations, the more detailed information can be collected concerning the accessibility of the target and the route to the intended attack. Street and traffic conditions, for example, may be significant in carrying out a suicide bombing. If a target is distant, it will be much more difficult to obtain the necessary operational information. This is why all Al-Qaeda attacks against the USA, except the attacks on the World Trade Center, have been conducted against US targets overseas (embassies and military bases). These targets were generally much closer to the bases of operations of Al-Qaeda. In the case of the Al-Qaeda attacks on the US World Trade Center in 1993 and 2001, it was necessary to set up satellite bases close to the targets. Subsequent successful and failed attempts in the UK and USA have revealed that attacks inspired by Al-Qaeda have been carried out by individuals who were born of immigrant parents within the home country, though in some cases received training from terrorist camps overseas. Thus, the traditional distinction between domestic and foreign terrorism no longer holds, but the necessity of proximity of the terrorist base of operations to the target remains supreme.

2. Weapons

As with targets, terrorists favor certain weapons over others. The majority of terrorist attacks are conducted using small arms and munitions. Attacks with weapons of so-called mass destruction (biological or nuclear) have been few, and where conducted, of limited effect. The characteristics of weapons that make them appealing to terrorists are MURDEROUS:

Multipurpose	A high-powered rifle has a specific use, while explosives have a much wider application
Undetectable	Semtex is small, lightweight, and largely undetectable, ideal for penetrating layers of security
Removable	The weapons of terrorism must be portable, so they must be relatively light and reasonably small, unless transportation is available (see tools below)
Destructive	Explosive devices have a greater kill rate than guns targeted at specific individuals
Enjoyable	Terrorists are clearly attracted to their weapons. In fact, it is not just terrorists who enjoy weapons. Many ordinary people do also
Reliable	If terrorists have used a weapon, or one like it, many times before, they are likely to favor that weapon over another
Obtainable	How easy is it to get the weapon? Can it be bought or stolen easily? Or can it even be manufactured in-house?
Uncomplicated	All weapons require practice and training. A weapon that demands considerable skill, such as a free-flight armor-piercing missile launcher, will rarely be used or if used miss the target
Safe	The use of bombs as weapons is inherently more dangerous than the use of other weapons, especially when made in-house. But they remain the terrorist weapon of choice because they are so destructive

There will almost always be a trade-off between the target of choice and the weapon. If a target is particularly desirable, or difficult to reach, there may be a search for an appropriate weapon. For example, the first attack on the US World Trade Center in 1993 used a truck bomb parked under one of the towers. Though it did much damage, it failed to destroy the target which at the time was considered indestructible. The second attack used a new weapon, a passenger jet as missile, which was successful. The planning required to use this new weapon was very extensive and took years to accomplish. It also required Al-Qaeda to locate a satellite base of operations in the USA close to the targets (the relevant airports and target itself). Finally, it ignored the one attractive characteristic of weapons listed above: it was not "safe." The users, as with suicide bombers, were willing to die in order to reach the target. However, we should note that in the case of "regular suicide bombers," the explosives used are well developed and “safe” for those managing the training of the suicide bombers. The explosive vest must be well constructed so that it is detonated only when the bomber reaches the target.

Weapons of mass destruction, including biological weapons, have been rarely used by terrorists. The reason is that such weapons generally do not fit the MURDEROUS requirements of weaponry to terrorists. However, using this approach, Hassan Naqvi reviewed the main biological and chemical weapons theoretically available to terrorists and concluded that ricin toxin would be the most attractive weapon because of its exceptionally MURDEROUS qualities:

Multipurpose	Ricin can poison individuals by means of ingestion, inhalation, and subcutaneous exposure
Undetectable	Ricin is tasteless and water soluble, making it extremely concealable. Laboratory setups designed to extract ricin toxin from castor beans range from very simple to as complex as a drug lab. Any of these labs can be set up in a garage
Removable	Ricin is extremely toxic to humans in low doses. Very small amounts of poorly purified ricin are enough to harm multiple individuals. A single terrorist will be able to carry vast quantities of ricin without help
Destructive	Ricin is a thousand times more deadly than cyanide. In addition, the dissemination of ricin by means of food, water, or ventilation vectors would cause considerable panic and economic turmoil
Enjoyable	The extraction of ricin from castor beans is a fairly simple science experiment, and its clandestine use may add to the enjoyment of its production. Admittedly, it may not be quite as spectacular in its effects as are violent explosions and guns
Reliable	Ricin is an extremely hardy protein. Because it is resistant to changes in temperature, acidity, and even ultraviolet light, it retains much of its poisonous properties in environments as extreme as the stomach
Obtainable	Castor beans are available for purchase on the Internet, and laboratory equipment can easily be stolen
Uncomplicated	The extraction and dissemination of ricin is the most basic of all bioweapons. The instructions for its weaponization are easily found on the Internet
Safe	Ricin, although extremely toxic, will not harm terrorists if fairly basic precautions are taken, such as the use of masks

(*continued*)

3. Tools

Terrorist attacks cannot be accomplished without the tools of everyday life. Without such tools, it is much harder for terrorists to reach a target or use their weapons. For many of the commonest attacks, such as car or truck bombings, drive-by shootings, and targeted assassinations, terrorists are likely to need most of the following:

- Cell phones or other means of communication
- Cars or trucks to transport themselves and weapons
- Cash or (false) credit cards, bank accounts, or other means of transferring money
- Documents (false or stolen) – for example, drivers’ licenses, passports or visas, and vehicle registration documents
- Maps (and increasingly GPS), building plans, and addresses so that the target location can be pinpointed
- Television sets and monitors
- Video and still cameras for surveillance
- Internet access to collect information on street closures, traffic patterns, weather conditions, local news, and disseminate propaganda worldwide

Depending on the availability of such tools, the weaponry employed and targets reached may vary.

It may seem obvious that terrorists would need these everyday tools, but it is not obvious how they can use them without divulging at some point who they are and what they are up to. These tools are certainly widely available, but their visibility is also considerable. Using cash instead of a credit card to rent a car, for example, draws unwanted attention to the terrorist. Using a credit card exposes one’s identity – unless it is stolen. For these reasons, terrorists steal many of the everyday tools that are widely available, which of course opens them up to risk of getting caught. This is why there is a considerable, and increasing, overlap between terrorism and "traditional crime" such as money laundering, drug trafficking, and even bank robbery – a favorite way of the IRA (Irish Republican Army) to raise money. Some terrorist groups have even created entire departments that specialized in forging documents. Others have made their own money.

Despite the generality of tools needed by terrorists, some attacks do, of course, require specific tools – for example, a belt or vest to carry the suicide bomber's explosives – and it will always be important to understand what these tools might be and the supply chain that produces them.

4. Facilitating Conditions

Targets, weapons, and tools exist within physical, economic, and social environments. These environments serve to enhance their use; otherwise, they would be useless. At particular points in time or in particular regions or places, conditions may arise that facilitate terrorist ability to exploit these opportunities. These conditions make it ESEER for terrorists:

Easy: When local officials are susceptible to corruption

Safe: When ID requirements for monetary or retail transactions are inadequate

Excusable: When family members have been killed by local antiterrorist action

Enticing: When local culture/religion endorses heroic acts of violence

Rewarding: When financial support for new immigrants is available from local or foreign charities

One can readily see that the characteristics of ESEER are more or less those of failed or failing states, which is why they are the preferred abode of terrorist bases. ESEER is also exacerbated by the conditions of globalization that include the worldwide marketing of small arms whether legal or illegal, porous borders between countries making movement of terrorist operatives easier, the proliferation of nuclear technology and other toxic materials creating opportunities for terrorists to obtain or manufacture WMDs, and lax international banking practices that facilitate money laundering. The global reporting of savage or violent terrorist attacks on mass media and the Internet has also probably facilitated terrorist recruitment into various terrorist organizations. It is to the organizational aspect of terrorism that we now turn.

The Situational Approach to Terrorist Groups

With some exceptions, terrorism is a group exercise, which brings with it both advantages and disadvantages to doing terrorism. The advantages include being able to (a) mount a more complex one-off mission (the Northern bank robbery by the IRA or the 9/11 attack) and (b) sustain a continued series of attacks over a long period of time (e.g., Hezbollah in the Palestinian territories). However, the majority of terrorist groups dissolve within 1 year or less. Whether they completely desist or morph into different terrorist groups is still an open question. But this important fact suggests that there are serious problems that affect the structure and dynamics of terrorist groups.

Terrorist Group Structure

There are a number of different types of organizational structure identified among terrorist groups. Much is made of their formal or informal structures, how they arose, and their similarity to other types of organizations such as business models, specialist groups for hostage taking, or cell formation to insulate members from the overall structure of the organization. The latter protects against penetration by external agents and minimizes internal group dissention and distrust.

It is also likely that developments in communication technology may mitigate the difficulties faced in organizational structure of criminal and terrorist groups. Mobile phones in particular, especially smart phones that provide cheap access to the Internet in third world countries, have made it much easier for groups to coalesce and to overcome some of the external and internal constraints of terrorist group operations. Indeed, as has occurred in the Arab spring uprisings of 2011, many of these have been fuelled by social networking technologies. Demonstrations, sometimes turning into violent uprisings, have occurred without any formal organization at all. Social networking technologies may overcome the need to maintain a lasting formal group structure. Groups can come together – almost spontaneously for a specific event – then melt away until the next occasion. However, if routine, that is, repeated, attacks are to be sustained, obviously an enduring organization is essential, made clear as the uprisings in Libya, Egypt, and Syria unfold.

Terrorist and Criminal Groups

The situational approach views terrorist attacks that are carried out by groups or at least more than one person in much the same way as any other crime. That is, it follows the commission of the crime step-by-step; it does not presume in advance that there is a particular organized hierarchical structure. For example, the 2010 book *Situational Prevention of Organized Crime* shows that drug trafficking, contraband cigarettes, sex trafficking, mortgage fraud, and timber theft are commonly carried out by small networks of friends or family groups who live on different sides of a national border and take advantage of marketing and trafficking opportunities that are thus available.

Groups are viewed as part of the situational environment in which decisions must be made. The extent to which decision making is facilitated or constrained when it is done by groups or in group situations rather than by individuals is a difficult question to answer, though classic research on group decision making has suggested that in groups that are not well established, there is a tendency for polarization in decision making and a shifting of decision making to converge with the views of an authoritarian leader. So the group setting is yet another factor that limits or conditions terrorists' rational choice.

It is also important to recognize that in many, if not all instances, groups organized to commit traditional crimes ("organized crime") such as drug trafficking and extortion, corruption of government officials already existed before terrorist groups came along. Therefore, the existence of such groups is simply another opportunity for terrorist groups to exploit, and it is logical that they would do so. There is mounting research that demonstrates the ways in which terrorist groups develop their organizational capacities as a result of the existence of organized crime and eventually collaborate, overlap with organized crime groups, or even morph in some instances into criminal groups that do organized crime. From the situational crime prevention point of view, however, it does not matter much whether what these groups accomplish is organized crime or terrorism. All we need to know is how they make their choices, what interventions will constrain them, and how working within a group organized in a particular way facilitates or constrains the choices made along the way to accomplishing the mission.

The Situational Dynamics of Terrorist Groups

There are many sources of conflict within terrorist groups. Newman and Clarke in *Outsmarting the Terrorists* have identified the internal dynamics of terrorist groups that may constrain them as:

(a) The conflict between charismatic leadership and strict discipline. Charismatic leadership is probably the most significant attribute of terrorist groups and is necessary in order to transform terrorist groups into terrorist movements. It is needed to maintain faith in violence as a means to an end. But at the same time, strict discipline is needed to maintain operational security and efficiency. If charisma wanes (which we know always does), faith and thus discipline are undermined.

(b) The conflict between morality and hypocrisy. The moral contradiction between killing people (by terrorists) as a protest against killing people (by whatever enemy) is inherently a shaky morality, needing to be constantly fed by self-serving propaganda.

(c) Rules of entrance and exit. Some groups have no other rules than being a friend or family member. Others have strict rules for entry (applicants must prove themselves) and of exit – those who leave are obviously a security risk.

(d) Operational security. Complex operations require well-trained individuals who know each other well. But the more they know of each other, the more security will be jeopardized if one of them is captured.

The Situational Approach to Counterterrorism

The situational approach to counterterrorism is of course focused on prevention of terrorist attacks. As in crime prevention, this is its prime objective. Over the years, situational crime prevention has accumulated a range of techniques for preventing many different types of crime. These techniques, known simply as the "Twenty Five Techniques" of crime prevention, are grounded on five psychological principles that are directed at further limiting the rational choice of the offender. Applied to preventing terrorism these are:

1. Increasing the effort of mounting an attack or mission, such as protecting targets
2. Increasing the risk of carrying out an attack, such as surveillance and checkpoints
3. Reducing the rewards of attacks such as swift mitigation by first responders to reduce the injury and damage of attacks
4. Reducing provocations of terrorists such as avoiding controversial weapons
5. Removing excuses such as adopting clear rules of interrogation

If we superimpose these principles on the four pillars of terrorist opportunity, a broad array of techniques that suggest points of intervention is revealed (Fig. 3). Each of these five principles must be tailored to a specific type of terrorist attack after a step-by-step analysis has identified points of weakness.

Policy Implications

The four pillars of terrorist opportunity and the five modes of intervention that form the bases of the 25 techniques provide a framework for a systematic policy that would assess the risks of attack against various targets in various locations. After 9/11, "risk assessment" was promoted by the US Department of Homeland Security as a way of protecting against terrorist attacks. Unfortunately, because it lacked a systematic structure of how to assess comparative risks, or even what might be at risk, the major portion of resources allocated to US states by the federal government was distributed according to political largesse rather than according to a systematic comparative assessment of what was at risk. This error was in part also produced by the widespread panic by both politicians and security professionals such as Richard Clarke the then White House counterterrorism czar and reflected in the cry, "we can't protect everything." But the situational approach demonstrates that not all targets are equally attractive to terrorists. We do not have to protect everything, at least not to the same degree. By developing a systematic procedure for identifying risk, adequate planning and countermeasures can be set up both to prevent attacks and to mitigate the fallout of attacks should they occur. The manual, *Policing Terrorism*: *An Executive*'*s Guide*, by professors Newman and Clarke and published by the Center for Problem-Oriented Policing, shows how this can be done by local police departments.

However, the complaint that "we can't protect everything" reflects not only panic but also a basic misunderstanding of the situational approach to terrorism, which is that if we protect one target, the terrorist will simply shift to another target that is more easily accessible. This is called in the research literature "displacement" or the "substitution effect". And it is the major criticism of the situational approach to crime and to terrorism.

Displacement

Echoed in *The Economist* on March 8, 2008, and in other trade media, the displacement criticism argues that terrorists are so committed to their violence; if they can't attack one target, they will keep on trying until they reach another. However, we have already seen that terrorist groups, although they may be composed of committed individuals, are very likely to dissolve within a year, so the idea of commitment at least by a group to carry out an attack against anything anywhere remains highly doubtful. Furthermore, recent research in situational crime prevention by Professors Guerette and Bowers

	Targets	Tools	Weapons	Facilitating Conditions
Increase the Effort	• Identify vulnerable targets • Prioritize targets for protection • Close streets, build walls and barriers • Security training for VIPs • Control dissemination of weapons technology	• Reduce supply of cash • Design electronic products to prevent use as detonators, timers. • High tech passports, visas, driving licenses etc. • National ID cards	• Restrict weapons sales • Hold contractors liable for stolen explosives • Reduce explosives shelf-life • Bomb recognition publicity • “User unfriendly” weapons • Restrict information on weapons use	• Tighten identity and credit authentication procedures • Tighten border controls • Destroy safe houses/training camps • Disrupt recruitment (e.g. Madrasas)
Increase the Risks	• Strengthen formal and informal surveillance through CCTV, citizen vigilance, hot lines	• Technology to identify and locate cars, trucks, cell phones • Internet surveillance • RFIDs for parts on vehicles and electronic products • GIS chips in terrorist tools	• RFIDs/GIS chips to track weapons • Screen incoming cargo for weapons • Outlaw technology to circumvent screening	• “Know your customer” bank policy. • Track all financial transactions • Monitor foreign student activity • Promote ties between local police and immigrant communities • Conduct sting operations to foster climate of risk
Reduce the Rewards	• Conceal or remove targets • Bomb proof buildings/Kevlar curtains • Design guidelines to minimize injury from explosions • Swift cleanup of attack site	• Anticipate terrorist innovation in use of tools	• Avoid overreaction to attacks	• Use publicity to isolate terrorist group from community • Use publicity to portray hypocrisy, cruelty of terrorist acts • Anti-money laundering regulations
Reduce Provocations	• Unobtrusive public buildings at home and abroad • Unobtrusive dress, living accommodations for personnel abroad	• Use unobtrusive vehicles in foreign locations	• Clear and consistent rules of engagement • Minimal weapons use at protest demonstrations	• Work closely with immigrant communities and host community abroad • Clear rules for public demonstrations • Avoid provocative announcements
Remove Excuses	• Train personnel abroad in local customs, culture		• Avoid use of controversial weapons (eg phosphorous bombs) • Don’t overestimate utility of bioweapons	• Avoid maltreatment of prisoners • Clear rules for interrogation • Counter terrorist propaganda • Win hearts and minds

Situational Approaches to Terrorism, Fig. 3 Situational prevention techniques applied to terrorism (Adapted from *Outsmarting the Terrorists*, pp. 193–194) has shown clearly that displacement does *not* occur in some 70 % of cases studied where interventions were introduced for a wide range of crimes, and the same has been found for terrorism in respect to the effectiveness of airport security. Finally, should displacement be shown to occur, this is actually a demonstration that the intervention *had an effec*t on the terrorists' behavior. It means that the terrorist in displacing to a less preferred target is taking additional risks, thereby increasing the chances of being caught or thwarted in the attempt. And having to attack a target of second choice, a successful attack will produce fewer rewards to the terrorist: it may be less vital and produce less injury and damage. This leads to the final important policy implication of the situational approach to terrorism.

Attack Mitigation

If it is acknowledged that displacement *may* occur (not *will* occur), then it is possible to develop a plan of security for protection of targets that rates them according to how attractive they are to terrorists (EVIL DONE) but also the expected loss should they be attacked. This may easily be incorporated into the standard disaster response planning that occurs in most communities. Then, if a target is attacked, the fallout of that attack can be mitigated by an effective disaster response, and thus the rewards of the attack to the terrorists reduced. The importance of this mitigation should not be underrated. Certainly Al-Qaeda terrorists have understood this from the beginning. It is why they commonly include in their attacks bombs designed to disrupt the mitigating actions of first responders. It is also why they try to commit a number of attacks simultaneously. This methodology has been copied by many other terrorist groups, especially the Taliban.

The situational approach to terrorism fits nicely into the established disaster response infrastructure already established. This not only involves training of first responders, but an extensive organizational infrastructure that ensures that confusion does not occur in dealing with an attack (which occurred in the 9/11 attack) but also ensures that local communities plan systematically to protect themselves according to a rational assessment of risks. Much of this planning for protection involves regular situational crime prevention activities, well known in the field of problem-oriented policing especially partnering with businesses and local community organizations to solve crime and disorder problems.

Conclusion

In sum, the situational approach to terrorism bridges the well-known gap between the academic study of the causes of terrorism and the practical challenges of preventing and responding to terrorist attacks that are faced by policymakers, police, and security professionals. While waiting for the rest of social and political science to eradicate the *root causes* of terrorism, the situational approach provides a conceptual, practical, and evidence-based guide for eliminating the *immediate causes* of terrorist attacks.

Related Entries

- ▶ Crime Prevention Through Environmental Design
- ▶ Environmental Criminology
- ▶ Identity Theft
- ▶ Money Laundering
- ▶ Organized Crime
- ▶ Problem-Oriented Policing
- ▶ Situational Crime Prevention

Recommended Reading and References

Bullock K, Clarke RV, Tilley N (eds) (2010) Situational prevention of organized crimes. Willan/Taylor and Francis, Abingdon

Clarke RV, Mayhew PM (1988) Crime as opportunity: a note on domestic gas suicide in Britain and the Netherlands. Br J Criminol 29:35–46, 1989

Clarke RV, Newman GR (2006) Outsmarting the terrorists. Praeger Security International, Westport/London

Cornish DV, Clarke RV (eds) (1986) The reasoning criminal. Springer, New York

Freilich JD, Newman GR (eds) (2009) Reducing terrorism through situational crime prevention, vol 25, Crime prevention studies. Criminal Justice Press, New York

Guerette RT, Bowers KJ (2010) Assessing the extent of crime displacement and diffusion of benefits: a review of situational crime prevention evaluations. Criminology 48(1):1331–1368

Hsu H (2011) Unstoppable? A closer look at terrorism displacement. Doctoral dissertation, University at Albany

Jones SG, Libicki MC (2008) How terrorist groups end: lessons for countering Al Qa'ida. RAND, Santa Monica

Makarenko T (2004) The crime-terror continuum: tracing the interplay between transnational organised crime and terrorism. Global Crime 6(1):129–145

Naqvi H (2012) Situational crime prevention applied to ricin and bioterrorism. In: Freilich JD, Shoham SG (eds) Special issue on policing terrorism around the globe. Israel studies in criminology, vol 11 (in press)

Newman GR, Clarke RV (2008) Policing terrorism: an executive's guide. USDOJ, Washington, DC

Rapoport D (1988) Inside terrorist organizations. Columbia University Press, New York

Rollins J, Wyler L (2010) International terrorism and transnational crime. Congressional Research Service, Washington, DC, pp 1–56

Stohl M (2008) Networks, terrorists and criminals: the implications for community policing. Crime Law Soc Change 50(1–2):59

Situational Crime Prevention

Nick Tilley[1] and Aiden Sidebottom[2]
[1]Jill Dando Institute of Security and Crime Science, University College London, London, UK
[2]Department of Security and Crime Science, University College London, London, UK

Overview

Situational crime prevention is concerned with reducing opportunities for crime. It comprises "opportunity-reducing measures that (1) are directed at highly specific forms of crime, (2) involve the management, design and manipulation of the immediate environment in as systematic and permanent a way as possible, (3) make crime more difficult and risky, or less rewarding and excusable as judged by a wide range of offenders (Clarke 1997: 4)."

Since its inception, situational crime prevention has undergone several refinements, incorporating developments in crime prevention research and practice. Opportunity reduction measures have taken numerous forms and have been found effective in reducing a wide range of crime types. It has also encountered staunch resistance, both on theoretical grounds and on the implications deemed to arise from the implementation of situational interventions.

This entry charts the development of situational crime prevention. It begins by describing its origins, method, and key principles. Next, it discusses evidence of its effectiveness. This is followed by a review of the major criticisms of situational crime prevention and associated rebuttals. It finishes by outlining future directions for the field.

The Origins of Situational Crime Prevention

Just as there was evolution before evolutionary theory, there was situational crime prevention before situational crime prevention theory. The core tenet of situational crime prevention is that unwanted actions by others can be averted by changing the immediate circumstances of their prospective behavior. It is not necessary (and indeed it may not be possible) to alter the dispositions inclining them to act in undesirable ways.

Hedgehogs practice situational prevention. In the interests of not being attacked, they have evolved sharp spikes that increase the risk and reduce the reward to those animals that might otherwise attack and kill them. Similarly, skunks have developed smells, butterflies the appearance of large eyes, and stick insects – well – the appearance of sticks, to make them less attractive to or less easily identified by potential predators. Situational prevention can even be seen in plants: think, for example, of stinging nettles, cacti, and bramble bushes.

Humans too invented situational measures long before late twentieth-century criminologists began to articulate the relevant theory. Most of us, for example, carry around with us milled coins. These became commonplace in the seventeenth century as a way of preventing coin clipping, which was then widespread: the face value of the coins was supposed merely to reflect the value of the precious metals of which they were minted, and profit could be made by clipping off and keeping some of the metal but preserving the purchasing power of the coin. Unclipped coins thus went out of circulation as clipped coins had the same exchange value. In terms of Sir Thomas Gresham's famous aphorism: bad money drove out good.

The same fundamental attribution error may underlie the surprisingly long time it took for both evolutionary theory and situational crime prevention theory to surface. The "fundamental attribution error" refers to the human tendency to explain good and bad by the actions of agents with corresponding good or bad intentions (Ross and Nisbett 1991). Crime is bad so its causes must lie in bad intentions: in particular we should blame bad agents and the source of their wickedness. Indeed, echoes of this are still found in much criminology, which looks for the causes of crime mainly in criminal dispositions borne of bad genes, bad upbringing, or bad social conditions. In response to this, situational crime prevention theory presents revolutionary challenges to much taken-for-granted thinking.

The stimulus to the first formulations of situational crime prevention came from a crisis of confidence in conventional wisdom in the 1970s. Improvements in welfare during the postwar period and the simultaneous increases in crime cast doubt on any notion that there might be a simple causal relationship between deprivation and crime, which would be prevented by improvements in social conditions. Moreover, evaluations of efforts at offender rehabilitation, reflecting the dominant criminological opinion at the time, were not encouraging. The slogan "Nothing Works" came to be believed. Situational crime prevention emerged in the British Home Office and promised an alternative approach that focused on preventing crime events rather than addressing offenders' dispositions to commit crime. The approach was a pragmatic one. Situational crime prevention seemed to work, and policies might usefully be directed at working out how and why situations could be modified to reduce opportunities for crime.

Theoretical and Methodological Underpinnings of Situational Crime Prevention

Crime as Opportunity was published by the British Home Office in 1976 (Mayhew et al. 1976). It was followed in 1980 by a paper in the *British Journal of Criminology* with the title "Situational Crime Prevention" (Clarke 1980). These two papers articulated a fresh agenda for understanding and preventing crime.

Mayhew et al. mainly focus on physical means of reducing crime opportunities. Their starting point, however, is "the extent to which deviance may be a temporal response to the provocations, attractions, and opportunities of the immediate situation" (1976: 1). They refer to Hartshorne and May's classic study, finding that honesty in children is not a stable attribute but instead that situational changes in opportunity can lead them sometimes to cheat, steal, and lie (Hartshorne and May 1928). Mayhew et al. refer also to research showing the importance of the immediate situation on the behavior of those placed in institutions, such as borstals. The language is of "stimulus conditions," including "opportunities for action," providing "inducements for criminality," which are "modified by the perceived risks involved in committing a criminal act and – in a complex, interrelated way – the individual's past experience of the stimulus conditions and the rewards and costs involved" (Mayhew et al. 1976: 2–3). Although opportunity had been referred to in passing in previous studies, Mayhew et al.'s argument was that it warranted closer, center-stage attention.

One of the most important grounds for arguing that situational factors are worthy of consideration in their own right was the discovery that changes in the toxicity of the domestic gas supply in the UK had a substantial effect on suicide rates. Mayhew et al. cite a study which found that the 45 % drop in Birmingham's suicide rate from 122 to 67 per million between 1963 and 1969 was almost entirely attributable to reduced toxicity of the gas supply in that city (there were 87 gas-related suicides in Birmingham in 1962 but only 12 in 1970). The drop in Birmingham's suicide rate was twice the national fall. A later study by Clarke and Mayhew (1988) corroborated these findings on a national scale. It might be expected that suicides are not committed lightly, but are provoked by deep feelings of despair and, therefore, can only be prevented by looking at the underlying psychological or social source of these sentiments. Finding that the gradual switch from toxic coal gas to nontoxic natural gas largely accounted for the overall drop in suicides challenged that assumption. Having an easily available method of committing suicide, in particular one that was painless and non-disfiguring, seemed to be associated with suicidal behavior. Removing that method had a substantial effect on suicide rates. Even in relation to this deeply caused behavior, there did not appear to be displacement to alternative methods (an issue we will return to shortly). There are many ways of taking one's own life. Changing the gas supply can have no influence on the cause of suicidal dispositions. It does not make suicide impossible. It nevertheless produced a substantial effect on total numbers of suicides. If deeply motivated behavior could be prevented by a simple change in the situation, how much more readily might criminal behavior, lacking such profound emotional sources, also be prevented?

Mayhew et al. see opportunities deriving from various sources: potential offenders' age, sex, and general lifestyle; victims' status; patterns of daily activity; properties of objects (such as their abundance); physical security; and levels of supervision and surveillance (Mayhew et al. 1976: 6–7). They also spell out variables that would be important in testing opportunity theories, for example, "the greater freedom of men to be away from their homes at night, the density of residential properties in rural as against urban areas, the security of these properties, or the extent to which routes of access to them are visible to neighbours and passers-by" (Mayhew et al. 1976: 7). The main emphasis, however, was on the modifiable features of opportunities, notably "the physical security and surveillance aspects of opportunity" (Mayhew et al. 1976).

The Home Office paper goes on to discuss some examples of opportunity reduction as a means of preventing crime: the use of steering wheel locks to prevent car theft and the effects of design (front or back entry) and staffing (presence or absence of a conductor) on bus vandalism. Both studies find that situational variables were important in generating the crime patterns of interest. The introduction of steering wheel locks to all new cars in the UK, from 1971, reduced their theft but with some displacement to older cars. Areas of buses that were not easily supervised experienced far higher levels of vandalism than those open to surveillance.

Clarke (1980) adopts a more critical stance towards the then prevailing criminology than that found in the Home Office paper. He refers to the widespread "dispositional bias" which sees crime as mainly the preserve of a small minority inclined to criminality and which also continued to inform crime prevention policy and practice. He highlights the practical difficulties in changing criminal disposition or in altering the social conditions deemed to foster criminality, as well as uncertainties over the causal processes at work. He stresses, instead, the potential benefits of looking at crimes as results of consciously made, situated choices, even though those involved may not fully appreciate why they are making their decisions.

Clarke refers to three distinctive concerns of situational crime prevention theory: crime events (rather than criminality), separate categories of crime (rather than all crime), and current conditions for criminal acts (rather than dispositions). Two major types of preventive mechanism are discussed: (a) reduction of physical opportunities and (b) increase in chances of being caught, although (c) housing policies that improve chances of parental supervision of children are also mentioned. Clarke distinguishes between opportunistic offenders, whose crimes are less likely to be displaced by situational crime prevention measures and those who make their living from crime, for whom displacement is more likely. A middle group, who supplement their income from crime, produces the main puzzle, where there are plentiful alternative targets. Clarke speculates that classes of target may be protected by situational measures, for example, when much stronger steel coin boxes replaced vulnerable aluminum ones in British telephone kiosks.

While Clarke's suggestions marked a wide departure from much criminological thinking at the time, noticeable similarities are observed between the situational approach and developments in several areas of psychology. For example, Clarke's concerns with criminology's "dispositional bias" spoke directly to the ongoing person-situation debate between personality psychologists and social psychologists. The then prevailing narrative was that human behavior is best explained by more or less stable traits that make up an individual's personality. This was challenged by the pioneering work of Mischel (1968), who claimed that personality traits were poor predictors of behavior and that greater attention be paid to situational influences. Later "situationalists" could point to a series of classic experiments, notably Zimbardo's Stanford Prison Experiment and Milgram's electric shock study that demonstrated the ease with which situational forces could be manipulated in ways that trumped dispositions, often to dramatic effect. There are also hints of behaviorism in Clarke's remark that "decisions are much influenced by past experience (as) people acquire a repertoire of different responses to meet particular situations and if the circumstances are right they are likely to repeat those responses that have previously been rewarding" (Clarke 1980: 138). Likewise, Clarke acknowledges explicitly that situational crime prevention "is, perhaps, closest to a social learning theory of behaviour" (p. 139). Clarke also notes, however, that situational prevention may owe "something to the sociological model of crime proposed by the 'new criminologists'" (Clarke 1980). Moreover, towards the end of his paper, he avers that opportunity reduction "is entirely compatible with a view of criminal behaviour as predominantly rational and autonomous" (Clarke 1980: 145).

The Home Office report of 1976 and Clarke's manifesto for situational crime prevention of 1980 laid the foundations for subsequent developments. These can be considered under the headings of methodology, models of the offender, and classification of preventive techniques.

Methodology

Action research has come to be the main method used in situational crime prevention. In accordance with this, Ekblom (1988) describes the "preventive process" as one involving (a) obtaining data on a crime problem, (b) analysis and interpretation of the data, (c) devising preventive strategies, (d) implementation, and (e) evaluation, with (f) feedback via continuous monitoring of crime.

Because situational measures are applied on the basis of the analysis of specific crime problems, evaluation methods that try to determine whether a specific type of measure, say lighting upgrades or CCTV, is effective are not normally deemed appropriate. Instead action-research-based case studies attuned to the problem context have become the preferred method.

Several tools and analytic techniques to inform situational crime prevention have been developed. In particular:

- Crime pattern analysis has advanced rapidly better to specify the problems to be addressed, as well as their spatial and temporal components.
- Repeat victimization patterns have been very widely found and victimization (by place, person, or target) identified as the best predictor for future risk and hence most deserving of preventive attention.
- The phenomenon of "near repeats" has also been identified: the elevated risks faced by those near to the victimized.
- The CRAVED mnemonic (concealable, removable, available, valuable, enjoyable, and disposable) has been devised to capture the attributes of products that are most likely to be stolen.
- "Crime scripts" are used to tease out the steps involved in committing more or less complex

crimes, which can then be drawn on to work out where the most promising points of preventive intervention can be found.
- The "problem analysis triangle" organizes the critical elements that give rise to crime events.

Models of the Offender: The Rational Choice Perspective

A rational choice model has been proposed for offenders (Cornish and Clarke 1986). The simplest way of understanding how situational measures work is to assume some level of rationality on the part of offenders. A change in the situation leads prospective offenders to reassess the expected costs and benefits from engaging in a particular criminal act. If the change in the situation suggests increased rewards or decreased costs, then more of those at the margins of participation in crime will be drawn in. Likewise, if the change suggests decreased rewards or increased costs, fewer will. Little of the formal apparatus of rational choice theory, as used by economists and game theorists, has been adopted. Rather, the assumption is simply that crime choices are open to influence by changing the expected balance of advantages and disadvantages. Indeed, even if there is a very strong disposition to commit a particular crime, the presence of a police officer at the scene, for example, may be sufficient to lead the individual to decide in the event not to proceed.

The rational choice model does not suggest that each individual carefully considers all the options that are available to determine which course of action will maximize utility. Lack of information, lack of ability, and the costs involved in arduous and prolonged decision-making make that untenable. Rather, potential offenders, like everyone else, operate with "bounded rationality." This conceives of them as decision-makers who are responsive to changes in situations, but it does not embrace the unrealistic assumption that they are capable, conscious, careful, conscientious, and comprehensive calculators of costs and benefits to estimate expected utility in advance of each choice made.

Classification of Preventive Techniques

Clarke's early references to reductions in physical opportunities and increases in the chances of being caught as two categories of situational measures have since been refined and expanded. These have been generalized, respectively, to "increasing effort" and "increasing risk" for the offender. "Reducing reward" was then added, followed by "reminding of rules" (now called "removing excuses") and "reducing provocation" (Wortley 2001; Clarke and Homel 1997; Cornish and Clarke 2003). Taken together, these form a classification system of situational techniques. Each of the five major headings has a variety of subtypes, as shown in Fig. 1.

The classification of situational techniques serves four purposes. First, it provides a catalogue of the ways in which proximal conditions may affect criminal behavior. Second, it provides a repertoire of possible situational measures that can be countenanced when trying to determine what might be done in relation to a specific crime problem. Third, it provides a user-friendly teaching aide when introducing researchers and practitioners to situational crime prevention. And fourth, it highlights evidence gaps where case studies are required.

In some versions of the typology of techniques, the importance of perception is stressed in relation to risk, reward, and effort. That is, it is recognized that it is not only, or perhaps always most importantly, a question of *real* increases in effort or risk or *real* decreases in reward. For the person who might contemplate crime, the issue is that of the risk, effort, and reward that they *perceive* in relation to the criminal acts of interest. Similarly, provocation and rule recognition are about the ways in which situations are understood.

The Effectiveness of Situational Crime Prevention

A very large number of studies demonstrating the effectiveness of situational crime prevention have now been undertaken (e.g., see Clarke 1997 for a collection of successful case studies

Increase the effort	Increase the risks	Reduce the rewards	Reduce provocations	Remove the excuses
1. Target Harden • immobilizers in cars • anti-robbery screens	*6. Extend guardianship* • Routine precautions: go out in groups at night • 'Cocoon' neighborhood watch	*11. Conceal targets* • gender-neutral phone directories • off-street parking	*16. Reduce frustrationsand stress* • efficient queuing • soothing lighting/music	*21. Set rules* • rental agreements • hotel registration
2. Control access • alley-gating • entry phones	*7. Assist natural surveillance* • improved street lighting • support whistleblowers	*12. Remove targets* • removable car radios • pre-paid public phone cards	*17. Avoid disputes* • fixed cab fares • reduce crowding in pubs	*22. Post instructions* • 'No parking' • 'Private property'
3. Screen exits • tickets needed for exit • electronic merchandise tags	*8. Reduce anonymity* • taxi driver ID's • school uniforms	*13. Identify property* • property marking • vehicle licensing	*18. Reduce emotional arousal* • controls on violent porn • prohibit paedophiles working with children	*23. Alert conscience* • roadside speed display signs • 'Shoplifting is stealing'
4. Deflect offenders • street closures • separate bathrooms for women	*9. Utilise place managers* • CCTV for double-decker buses • Two clerks in liquor stores	*14. Disrupt markets* • monitor pawn brokers • licensed street vendors	*19. Neutralise peer pressure* • 'idiots drink and drive' • 'it's OK to say NO'	*24. Assistcompliance* • litterbins • public lavatories
5. Control tools/weapons • toughened beer glasses • disabling stolen cell phones	*10. Strengthen formal surveillance* • security guards • CCTV in town centres	*15. Deny benefits* • ink merchandise tags • graffiti cleaning	*20. Discourage imitation* • rapid vandalism repair • V-chips in TV's	*25. Control drugs and alcohol* • breathalysers in pubs • alcohol-free events

Situational Crime Prevention, Fig. 1 Typology of situational crime prevention techniques. **Source**: Tilley, N. (2009). *Crime Prevention*. Willan, Cullompton, Devon

and for a fuller bibliography, with over 200 examples, go to http://www.popcenter.org/library/scp/pdf/bibliography.pdf.). These cover a range of crime types and different settings. For example, in one study on residential burglary in England, lockable gates were installed to alleyways that ran behind terraced properties. The gates restricted access only to those residents in the affected properties, thereby reducing the opportunities for burglars to enter such properties from the rear, as was evidently the norm. The scheme was found to reduce residential burglary by over a third relative to the control area. Moreover, the scheme was cost-effective, recouping £1.86 for every one pound spent (Bowers et al. 2004). In another study, La Vigne (1994) describes how the implementation of a computerized telephone system in the New York City jail system led to considerable reductions both in illicit phone use by inmates and phone-related violent disputes. In other examples, the effect was immediate, as with the near disappearance of bus robberies in the USA in the 1970s following the implementation of cash drop safes and the shift to an exact-fair payment system.

Such studies do not find some specific measure or type of measure that "works" invariably – specific crime prevention panaceas are not expected. Instead, what they show is that strategies making alterations in the balance of rewards, efforts, and risks or removing provocations or adding rule reminders at the point of crime commission can be devised, which speak to specific crime problems.

The true extent of the effectiveness of situational crime prevention is likely to be higher than that gleaned from the published literature. Retailers and private businesses routinely employ situational measures to protect their assets and prevent loss or theft. Fitting merchandise with electronic tags or placing items in protective displays is commonplace. Such efforts are, however, rarely formally evaluated and hence are typically absent from the scientific literature.

Criticisms of Situational Crime Prevention

Both the theory and practice of situational crime prevention have been subject to an array of criticisms. The most common are itemized below, followed by the associated rebuttals.

Theoretical Criticisms

Situational Crime Prevention Only Displaces Crime

Arguably, the most enduring criticism of situational crime prevention concerns displacement. It is argued that situational crime prevention simply shifts crime around because it fails to address the underlying "root causes" of crime. Displacement can take several forms: by place, time, method, crime type, or offender. There are enormous measurement problems that mean that it is, in effect, impossible to rule out all forms of displacement. The best that is possible is to assess whether there is displacement where, when, and in relation to which offenses it would be most expected. On that basis, it can be concluded with some confidence that displacement rarely matches the crimes prevented by the application of situational crime prevention (Bowers and Guerette 2009). This is because (a) offenders like to offend in areas with which they are familiar and (b) crime opportunities are not uniformly available. Contrariwise, there is growing evidence that "diffusions of benefits" often occur whereby preventive effects extend beyond the operational range of the measures applied. The possible types of diffusion of benefits match those of displacement. They have been found quite widely, for example, by time, where "anticipatory benefits" include reductions in crime before the preventive measures become operative (Smith et al. 2002), and place, where the geographical range of crime reductions extend beyond the area covered by the measures (Clarke and Weisburd 1994). Proposed explanations for these patterns are that offenders tend to be unaware of the exact boundaries (temporal and spatial) of prevention measures.

Situational Crime Prevention Is Superficial and Simplistic

Crime is complex with many causes. Its explanation calls for attention to a wide range of variables relating to genetics, childhood experience, social structure, and the workings of the criminal justice system. Even if situational variables have any role to play, it is a small one that needs to be considered in the context of a wide range of other causal influences. In placing less emphasis on them, the situational approach is criticized for speaking only to the descriptive components of crime – *where* and *when* crime happens – and not to *why* crime occurs. Situational crime prevention is therefore alleged to be overly simple: lacking the theoretical depth of mainstream criminological theories and merely extending common sense.

The criticism is a weak one. Situational crime prevention draws heavily on three well-established crime event theories: routine activity, rational choice, and crime pattern. It is also underpinned by the widely accepted principle that situations yield causative influences on human behavior, in combination with predisposition, nor can it accurately be considered simple. While the superficial logic is often consistent with day-to-day observations, why the interaction between certain individuals and certain situations under contrasting conditions does or does not generate crime is in no way less complex than the question of why certain individuals do or do not enter a life of crime.

Situational Crime Prevention Is Applicable Only to Instrumental Offenses

Although situational crime prevention may be relevant to instrumental crime where some calculation of returns in relation to effort and risk might reasonably be expected and where, therefore, situational changes that alter expected utility might change an offender's decisions, for other types of crime, it is much less relevant. Many crimes, especially those involving violence, are more expressive than instrumental or are undertaken under the influence of drugs or alcohol. In these cases, those involved are unlikely to notice or adjust their behavior in response to more or less subtle situational factors.

There are clearly more examples of situational crime prevention being applied to instrumental crimes than there are expressive crimes. But generalizing this pattern to reduce situational crime prevention to an instrumental-only form of crime control is misleading. That many early situational measures relate to acquisitive crimes reflects demand not discordance. Situational crime prevention was developed by researchers based in the British Home Office, which, at the time, was primarily concerned with the high levels of residential burglary and car crime experienced in England and Wales. More generally, there is a situational component to all behaviors. The association of changes in the supply of gas in the UK with dramatic changes in suicide patterns was described earlier. There is also strong evidence to show that situational changes can lead to dramatic reductions in the probability of and harm caused by violence. Bars are a prime example. The shift from glass pint pots to shatterproof receptacles and changes to the internal configuration of bars and nightclubs have all been found to produce positive reductions in alcohol-related assaults in such settings (see Graham and Homel 2008).

Ethical and Practical Criticisms

Situational Crime Prevention Blames Victims

Offenders commit crime and so should be held fully responsible for their criminal behavior. Situational crime prevention suggests that crime problems derive from the opportunities created by those who have not committed any criminal act. Moreover, it suggests that they should accept some responsibility (and in certain cases bear some of the costs) of reducing opportunities. For many critics this is an affront, leading to accusations that situational prevention "blames the victim." Offenders choose to commit their crimes. Fault lies entirely with them. Moreover, it is the duty of the state in general, and the criminal justice system in particular, to control their behavior.

One suspects that the understandable emotions associated with victim blaming, including most obviously the unwarranted notion that some rape victims may be blamed for wearing provocative outfits, have placed greater prominence on this criticism than is deserved. Situational crime prevention recognizes that targets (broadly defined) can play a causative role in crime events. This is true of people, places, and products. Evidence suggests that certain measures are associated with reduced vulnerability to victimization (as well as its opposite). Most individuals welcome such advice and use it to take recommended precautions against crime. Most individuals would also accept that this is a reasonable expectation and that government agencies are largely impotent to protect them from crime in every situation all of the time. Moreover, there are examples where victims are open to blame and where they may be held culpable for the criminogenic effect of their actions (or inactions). For example, a shop that repeatedly experiences theft because of the way it displays goods, but passes the costs of the crimes to the state, may rightly be blamed for failing to take adequate precautions. Likewise, bar managers who blatantly supply alcohol to already intoxicated patrons may be held responsible for the violent crimes they produce.

Situational Crime Prevention "Criminalizes" Policy, Planning, and Design

The notion that the physical and social architecture of everyday life shapes crime opportunities and that it should therefore be adjusted to reduce crime, risks "criminalizing" areas of policy and practice (shaping them by a situational crime prevention agenda) to the detriment of other ends they may have. Pursuing a situational crime prevention agenda, it is argued, undermines social welfare policy, and diverts resources away from welfare provision and social crime prevention. This also applies to schools, hospitals, retailers, product designers, planners, and architects. In particular it may suggest acceptance, in the name of crime prevention, of an ugly "fortress society."

Knepper (2009) argues that situational crime prevention contributes more to social policy than the literature suggests. Many situational crime prevention case studies take place in settings that support welfare, namely, housing, hospitals, public transport, and schools. This is most clearly seen in the work of Poyner and Webb on how the design and configuration of social housing influences rates of crime (Poyner and Webb 1991). Effective situational crime prevention can thus increase the likelihood of meeting social welfare objectives.

With respect to the criminalization of products and places, it is trite but worth emphasizing that situational crime prevention amounts to more than simply target hardening. Situational interventions exhibit huge diversity, from high-tech baggage detectors to low-tech bicycle parking stands designed to facilitate improvements in locking practice. Nor are such measures ugly. To situational advocates, particularly those in the tradition of design against crime, manipulating the immediate environment in ways that reduce opportunities for crime without compromising, say, aesthetics, costs, or function holds sovereign (see Ekblom 2008). Finally, while some might consider gated communities as emblematic of a "fortress society," there are many examples of situational interventions altering the physical environment in ways that promote community togetherness by facilitating neighborly interactions. Lighting upgrades and subsequent increases in pedestrian street usage after dark is a case in point (Painter 1996).

Future Directions for Situational Crime Prevention

There is scope for some theoretical development in situational crime prevention. It is not clear that the rational choice model of the offender is necessary for it. The rational choice model emerged only after the core ideas about situational crime prevention had been formulated, at which point, as indicated earlier, various possible models of the offender were mooted. Moreover, the incorporation of provocation reduction and rule reminders suggests that agents can be responsive to situational changes that operate at the level of feelings and morals as well as at the level of utility calculation.

Displacement is a further area deserving of research. Evidence on the presence of displacement is lopsided. The assessment of spatial displacement and to a lesser extent temporal displacement predominates, while other types of displacement have been little considered, in part because of the challenging measurement problems. The empirical research that is available is also limited in what it can capture, even when technically highly sophisticated. It has long been recognized that displacement is likely to be contingent on factors such as alternative available targets and types of offender motivation. There is scope for further theoretical development to devise propositions that predict what kinds of displacement can be expected in what conditions, again delivering on the suggestions of early papers on situational prevention. The same is true for diffusion of benefits. That theoretical work then needs to be followed by empirical tests.

There is also scope to advance the application of situational crime prevention. While the range of crime types that have been targeted by research and practice has expanded enormously over the past 30 years, there remain important gaps. Studies into "unconventional" crimes, for example, organized crimes of various types (Bullock et al. 2010), terrorism (Clarke and Newman 2006), wildlife crime (Pires and Clarke 2012), and cyber crime (Newman and Clarke 2003), are available but at present are largely programmatic. Case studies evaluating measures designed to activate provocation-reduction or excuse-removal mechanisms are limited (Guerette 2009). Moreover, situational interventions have rarely been formally evaluated when applied in industrial, office, and agricultural settings (Guerette 2009). More broadly, situational crime prevention has yet to be systematically trialed in developing resource-limited countries, although interest appears to be growing (Pires and Clarke 2012). A key question relates to the challenges of carrying out the preventive process in settings where data are often absent.

The final point concerns exposure and dissemination. Despite being widely advocated and implemented in both North America and much of Europe, it is fair to say that situational crime prevention is yet to capture the imagination of the public and wield widespread influence on policymakers. There is a banality to situational crime prevention that belies its sophistication. This often means it is not credited for providing valid explanations for shifting crime patterns where traditional criminology fails. Nor is it always recognized for its role in producing major crime reductions. Take the much-storied crime drop experienced in several industrialized nations from the early to mid-1990s. Numerous hypotheses have been put forward for such declines, many receiving widespread media coverage. Yet arguably the most convincing explanation of the observed trends relates to the prevalence and quality of in-car (situational) security measures (Farrell et al. 2011). Failure to bring the successes of situational crime prevention to the attention of the public may stymie its progression and confine it to the margins of crime prevention policy and practice.

Related Entries

- ▶ Crime Mapping
- ▶ Crime Prevention Through Environmental Design
- ▶ Crime Science
- ▶ Effectiveness of Situational Crime Prevention
- ▶ Multiple Victims and Super Targets
- ▶ Problem-Oriented Policing
- ▶ Rational Choice Theory
- ▶ Routine Activities Approach
- ▶ Theories for Situational and Environmental Crime Prevention

Recommended Reading and References

Bowers KJ, Johnson SD, Hirschfield A (2004) Closing off opportunities for crime: an evaluation of alley-gating. Eur J Crim Policy Res 10(4):285–308

Bullock K, Clarke R, Tilley N (eds) (2010) Situational prevention of organised crimes. Willan, Cullompton

Clarke RV (1980) Situational crime prevention: theory and practice. Br J Criminol 20:136–147

Clarke RV (ed) (1997) Situational crime prevention: successful case studies, 2nd edn. Harrow and Heston, Guilderland

Clarke RV, Homel R (1997) A revised classification of situational crime prevention techniques. In: Lab SP (ed) Crime prevention at a crossroads. Academy of Criminal Justice Sciences and Anderson, Highland Heights/Cincinnati

Clarke RV, Mayhew P (1988) The British gas suicide story and its criminological implications. In: Tonry M, Morris N (eds) Crime and justice, vol 10. University of Chicago Press, Chicago, pp 79–116

Clarke RV, Newman GR (2006) Outsmarting the terrorists. Preager Security International, Westport

Clarke RV, Weisburd D (1994) Diffusion of crime control benefits: observations on the reverse of displacement. In: Clarke RV (ed) Crime prevention studies, vol 2. Criminal Justice Press, Monsey, pp 165–182

Cornish D, Clarke RV (eds) (1986) The reasoning criminal. Springer, New York

Cornish DB, Clarke RV (1987) Understanding crime displacement: an application of rational choice theory. Criminology 25:933–947

Cornish D, Clarke RV (2003) Opportunities, precipitators and criminal decisions: a reply to Wortley's critique of situational crime prevention. In: Smith MJ, Cornish DB (eds) Crime prevention studies: theory for practice in situational crime prevention, vol 16. Criminal Justice Press, Monsey

Ekblom P (1988) Getting the best out of crime analysis. Home Office, London. Crime prevention unit series, Paper No. 10

Ekblom P (2008) Designing products against crime. In: Wortley R, Mazerolle L (eds) Environmental criminology and crime analysis. Willan, Cullompton, pp 195–220

Farrell G, Tseloni A, Mailley J, Tilley N (2011) The crime drop and the security hypothesis. J Res Crim Delinq 48:147–175

Graham K, Homel R (2008) Raising the bar: preventing aggression in and around bars, pubs and clubs. Willan, Cullompton

Guerette RT (2009) The pull, push and expansion of situational crime prevention evaluation: an appraisal of thirty-seven years of research. In: Tilley N, Knutsson J (eds) Evaluating crime reduction initiatives, vol 24, Crime prevention studies. Criminal Justice Press, Monsey, pp 29–58

Guerette RT, Bowers K (2009) Assessing the extent of crime displacement and diffusion of benefit: a systematic review of situational crime prevention evaluations. Criminology 47(4):1331–1368

Hartshorne H, May MA (1928) Studies in deceit. Macmillan, New York

Knepper P (2009) How situational crime prevention contributes to social welfare. Liverp Law Rev 30:57–75

La Vigne NG (1994) Rational choice and inmate disputes over phone use on Rikers Island. In: Clarke RV (ed) Crime prevention studies, vol 3. Criminal Justice Press, Monsey, pp 109–125

Mayhew PM, Clarke RV, Sturman A, Hough JM (1976) Crime as opportunity. Her Majesty's Stationary Office, London. Home Office Research Study, No. 34

Mischel W (1968) Personality and assessment. Wiley, New York

Newman GR, Clarke RV (2003) Superhighway robbery: preventing e-commerce crime. Willan, London

Painter K (1996) The influence of street lighting improvements on crime, fear and pedestrian street use, after dark. Landsc Urban Plan 35(2–3):193–201

Pires S, Clarke RV (2012) Are parrots CRAVED? An analysis of parrot poaching in Mexico. J Res Crime Delinq 49:122–146

Poyner B, Webb B (1991) Crime free housing. Butterworth Architect, Oxford, UK

Ross L, Nisbett RE (1991) The person and the situation: perspectives of social psychology. McGraw-Hill, New York

Smith MJ, Clarke RV, Pease K (2002) Anticipatory benefit in crime prevention. In: Tilley N (ed) Analysis for crime prevention, vol 13, Crime prevention studies. Criminal Justice Press, Monsey

Wortley R (2001) A classification of techniques for controlling situational precipitators of crime. Secur J 14(4):63–82

Situational Crime Prevention and the Wild West

Shawna Cleary[1], Rashi K. Shukla[1] and
Marcus Felson[2]
[1]School of Criminal Justice, University of Central Oklahoma, Edmond, OK, USA
[2]School of Criminal Justice, Texas State University-San Marcos, San Marcos, TX, USA

Synonyms

Crime prevention; Natural crime prevention

Overview

Criminology now includes an increasing body of research and theory that subsumes crime prevention within a larger topic of protection and survival in nature. That body of work not only considers crime prevention as representative of a larger survival process but also takes into account human-animal interactions. Such interactions include human abuses of animals and animal populations and nature in general – including abuses that are against the law. But human-animal interactions also include humans protecting against animal incursions into their herds and home territories.

Background and Review of Literature

The situational prevention literature has recognized historical examples of human self-protection as a background topic but has increasingly brought the interactions with nature to the foreground. After reviewing a bit of that literature, the current entry shifts to a pictorial review of situational crime prevention in America's "Wild West." That term often applies to a historical period that has disappeared, but even today many Western states include sections with ranches, horses, cattle, and exposure to predatory incursions. Political struggles are found between metropolitan environmentalists who wish to reintroduce wolves to rural areas and ranchers who fear the loss of cattle and sheep to those same wolves. Migrants from city to suburban fringe and rural areas seek greater security than the areas from which they came; but they introduce new risks and fears to the areas that they enter. Thus, one person's protection becomes another person's threat. Survival from outside threats is a perennial problem in human history as in natural history.

The situational prevention of attacks also was found in the Old West. Consider this problem: ingots of silver bullion were carried by stagecoaches that were often victims of robbery by gangs of armed men on horseback. The problem was completely solved by instead producing 300 lb ingots that could not be carried on a single horse (Lingenfelter 1986). That was the end of stagecoach robbery. This example has been noted by several criminologists (Tilley and Laycock 2002; Clarke 1999; Farrell and Pease 2006).

Another example from the Old West came from pre-cowboy times. The ancient cliff-dwelling Indians in what is now Arizona and New Mexico entered their homes by climbing up ladders. At night they pulled up their ladders to prevent intrusions. This illustrates that situational prevention is far from new in behavioral terms. Moreover, Oscar Newman (1972) called our attention to territoriality and defensible space. Although situational crime prevention now includes many more ideas than that, it is still recognized that Oscar Newman placed defense of oneself, one's offspring, and one's dwelling area within a larger natural history perspective. That perspective was elaborated greatly in Felson's (2006) *Crime and Nature*, which included symbiotic and competitive ideas about offenders and the larger society.

Yet the breadth of "environmental criminology" is widened still further by new criminological papers on human-animal interactions. Recent research on parrot poaching strongly supports the idea that such crimes are highly structured in time and space and therefore subject to situational crime prevention (Pires 2012; Pires and Clarke 2011, 2012). Additional important research considers human attacks on elephants (Lemieux and Clarke 2009), and a more general review of wildlife crime is found in Pires and Moreto (2011), explaining why such offenses are subject to situational prevention. It is also clear from Eliason (2012) that wildlife crime is theoretically understood using the routine activity approach, a thread running through each of the articles reviewed in this paragraph.

Clarke (1999) offers us a framework for evaluating security by offering 25 techniques for situational crime prevention, along with many examples. His framework is widely employed to analyze crime situations and recommend solutions. These techniques are organized in five columns based on how they contribute to crime reduction. Included are (1) increasing the effort to carry out the crime; (2) increasing the risks to the offender; (3) reducing the rewards that the crime gives to the offender; (4) reducing provocations, especially for dispute-related crimes; and (5) removing excuses for carrying out the crime.

The current review considers how ranchers and others in today's Wild West use situational prevention techniques to protect against intrusions and attacks. The more important goal is to add examples and hence to widen the repertoire of ideas for studying the prevention-protection process and for understanding routine activities and situational prevention of crime. This entry shall demonstrate how the techniques of situational crime prevention have been and are currently being used in the American West, especially for nonhuman predators. Despite the urban image of rural life as idyllic, the reality is far from that. Ranchers, farmers, and other property owners often have a great deal of land and animals that must be controlled and protected from people, animals, and insects. A protective mentality is very important for their economic survival and is taken very seriously.

A Note About the Pictures

The pictures used within this entry were taken by the primary author. Many of the pictures were taken in rural Oklahoma; a few were taken at the

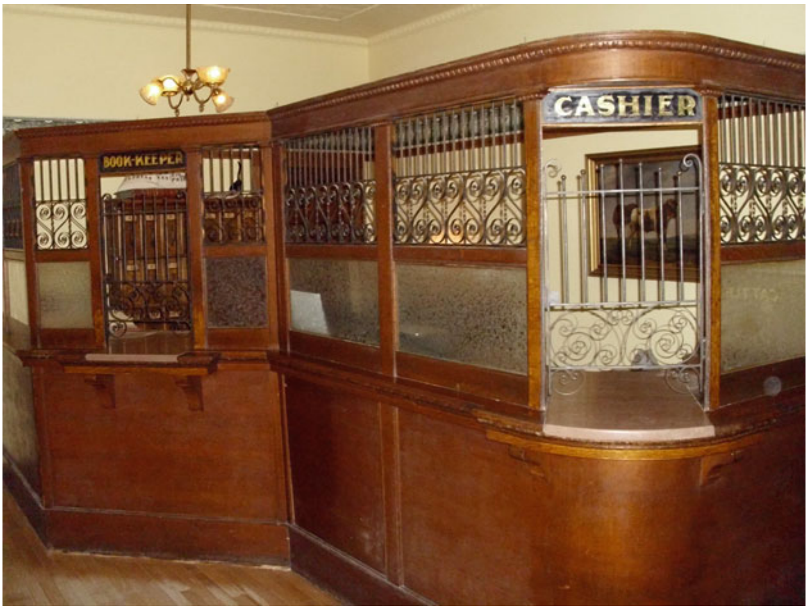

Situational Crime Prevention and the Wild West, Fig. 1 Old style bank teller window with bars

National Cowboy and Western Heritage Museum in Oklahoma City and the Woolaroc Museum in Bartlesville, Oklahoma.

Crime Prevention in the West

We begin (Fig. 1) with a bank teller window, more heavily barred than most modern banks in the United States. Although the design of the bars is attractive, their main purpose is to keep offenders out of the money area – an example of target hardening along with a kind of perimeter control. That did not prevent very aggressive robbery by multiple armed men, but it did narrow the span of vulnerability. Figure 2 shows a safe from the Cattlemen’s State Bank, designed to be too heavy for easy removal and too hard to break, given its heavy lock. Although we take these safe designs for granted today, they originated in an earlier era in response to the mobility offered by horses and a desire to respond.

One of the most important inventions in the American West was barbed wire. Without it, one can argue that the West would never have developed, since mass grazing of animals would have been impractical if they could easily wander off. The Biblical image of a shepherd watching a small herd of sheep or goats contrasts with the American process of maintaining very large herds with minimal supervision. Barbed wire fencing was developed in the late 1800s as a method of frugally enclosing animals and land while also preventing human theft. Barbed wire fence design styles range from legal and patented to illegal versions of patented designs to custom designs, resulting in thousands of versions (Liu 2009). In fact, the Cowboy and Western Heritage Museum in Oklahoma City displays 1,300 different types of barbed wire fencing. Figure 3 contains a few examples of different barbed wire styles from the late 1800s, each with its own patent. None of these styles is currently in use today, but the basic idea continues to the present in other forms. Figure 4 shows a rudimentary rural fence, while Fig. 5 shows a more professional version – each of which serves to contain cattle and signal other people; modern fencing in the West (not depicted) consists of wire woven in a small v-pattern that encloses cattle and horses without catching their hooves, is very difficult to climb, and keeps out large predators unless they can jump. Each of these fencing types applied defensible space long before Newman articulated it.

CATTLEMEN'S STATE BANK

Situational Crime Prevention and the Wild West, Fig. 2 Cattlemen's State Bank safety vault

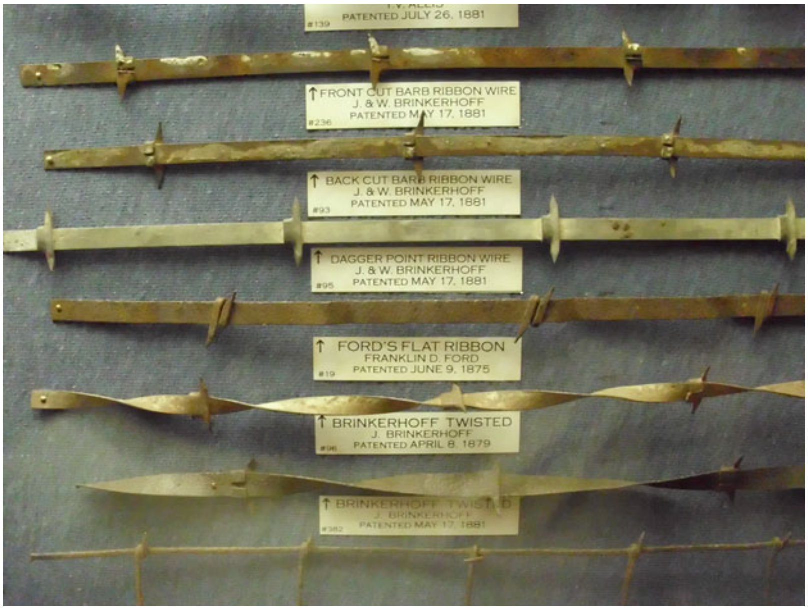
PATENTED JULY 26, 1881
#139
FRONT CUT BARB RIBBON WIRE
J. & W. BRINKERHOFF
PATENTED MAY 17, 1881
#236
BACK CUT BARB RIBBON WIRE
J. & W. BRINKERHOFF
PATENTED MAY 17, 1881
#93
DAGGER POINT RIBBON WIRE
J. & W. BRINKERHOFF
PATENTED MAY 17, 1881
#95
FORD'S FLAT RIBBON
FRANKLIN D. FORD
PATENTED JUNE 9, 1875
#19
BRINKERHOFF TWISTED
J. BRINKERHOFF
PATENTED APRIL 8, 1879
#96
BRINKERHOFF TWISTED
J. BRINKERHOFF
PATENTED MAY 17, 1881
#382

Situational Crime Prevention and the Wild West, Fig. 3 Examples of different styles of barbed wire

Not so well known outside the West is the old technique of hanging dead coyotes on a fence as a signal to live coyotes that their lives are endangered if they cross (Fig. 6).

It is important to consider in reading the current entry that predatory animals, from insects to coyotes, do more than just nibble. Attacks sometimes decimate crops or herds and other times remove the profit margin. Early ranchers worked on a much smaller scale, but today's West includes not only large-scale operations but also many small-scale ranchers who work part-time in

Situational Crime Prevention and the Wild West, Fig. 4 Rudimentary barbed wire fence in rural Oklahoma

PRIVATE PROPERTY
NO TRESPASSING

Situational Crime Prevention and the Wild West, Fig. 5 Well-constructed barb wire fence in rural Oklahoma

the city and are subject to considerable loss of income and way of life. Thus, protection of livelihood is a serious issue for them.

The use of predator against predator is an important part of their defense strategy. Owls are natural predators for farm pests, including small birds and rodents. Many rural people in the West place statues of owls in their barns to act as artificial place managers (Fig. 7). Plastic snakes and hawk statues are sometimes used for the same purpose. Scarecrows are designed to look like a farmer and are sometimes used as

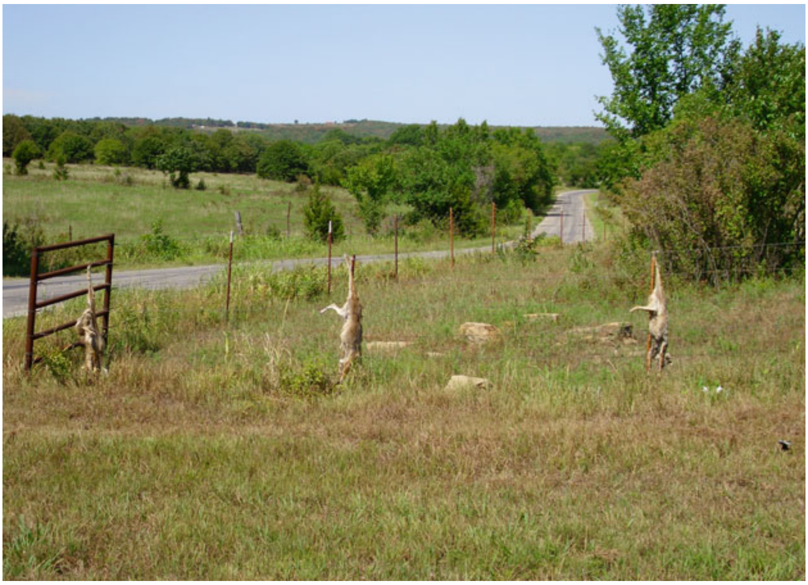

Situational Crime Prevention and the Wild West, Fig. 6 Dead coyotes hanging on fence as warning

Situational Crime Prevention and the Wild West, Fig. 7 Plastic owl serving as place manager in barn

artificial place managers to scare off birds and protect crops; many of the domestic pets in metropolitan areas become very important economically in the American West. The role of dogs was already mentioned above. Not so well known to urban people is the role of barn cats – commonly used as place managers at farms, ranches, and stables. These cats live in and around barns for the primary purpose of hunting and killing mice, rats, birds, moles, and small snakes.

Animal containment is an important rural issue for many reasons, especially for large animals. Ranchers and farmers do not want to lose their cattle or find them injured and have to

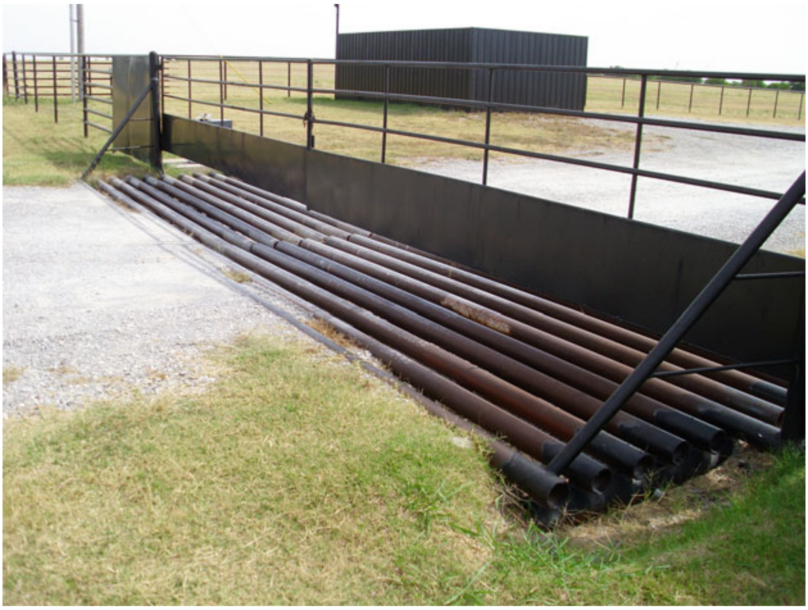

Situational Crime Prevention and the Wild West, Fig. 8 Cattle guard

Situational Crime Prevention and the Wild West, Fig. 9 Guard donkey with longhorn cattle in rural Oklahoma

euthanize them, nor do drivers want to run into stray cattle. Figure 8 illustrates a cattle guard, designed to contain large animals in pastures while allowing gates between them to remain open. Cattle guards are formed by placing steel bars close together horizontally across a shallow ditch under a gate. The bars discourage hoofed animals from exiting through the gate as their hoofs slip on the bars and get caught between them. Cows and horses will generally avoid cattle guards, although they are not foolproof. Some horses may jump over them, and both cattle and horses have been known to catch their hoofs between the bars, breaking one or more legs, leading to certain euthanization.

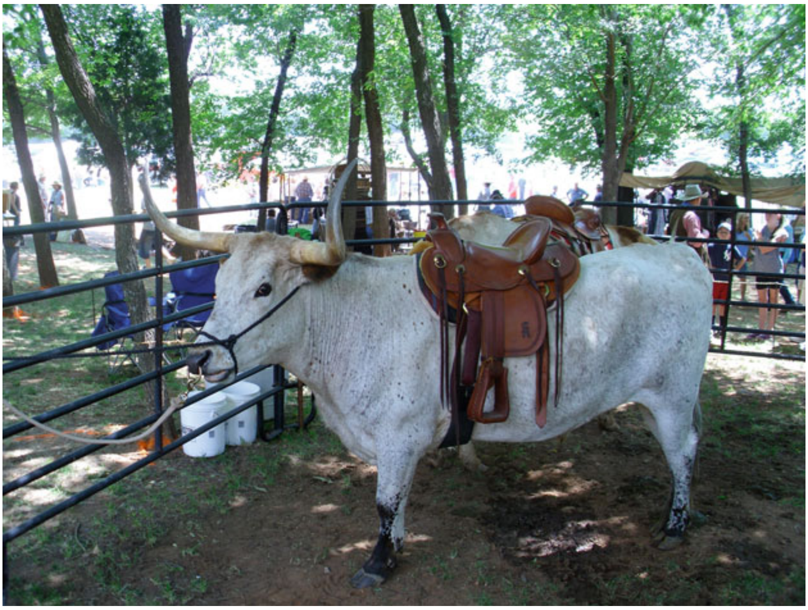

Situational Crime Prevention and the Wild West, Fig. 10 Longhorn cow wearing branded saddle

Situational Crime Prevention and the Wild West, Fig. 11 Stall divider

One of the most interesting prevention techniques is to use guard donkeys near cattle to watch fences and prevent intrusions (Fig. 9). Donkeys are good place managers because they have a dangerous kick and tend to attack animal predators. Similarly, llamas or large dogs are sometimes used for the same purpose. This gives us a new way to look at "formal surveillance," with animals functioning as security guards, requiring relatively inexpensive maintenance, and demanding no pensions or benefits.

Figure 10 illustrates branding – burning unique marks into the flesh of cattle, horses, or saddles to identify them with a particular ranch.

This fits very closely with the situational prevention method of identifying property to discourage burglary (Laycock 1991). Sometimes a brand would be altered by the thief in order to allow sale. A more modern technique is less easily thwarted: Freeze branding leaves a white brand mark on an animal's hide and is more effective than burning. Thus, we see that Ekblom's (2001) "arms race" between offenders and crime prevention techniques applies as much in the West as in metropolitan areas around the world. Modern animals and property are also identified with brass nameplates (not illustrated) that offer some protection if offenders do not have time to remove them.

Because large-scale horse operations have been commonplace in the West, managing animal interaction is important. Because horses are herd animals, it is important for them to see and communicate with each other, but physical interaction can be problematical if they begin to fight. Horses within a single farm or ranch are usually separated from one another with a variety of walkways and stall dividers (see Fig. 11) allowing them to see, hear, and smell one another, without biting or kicking.

Conclusions

Numerous types of situational crime prevention apply to the American Wild West and today's remaining Western frontier. Target hardening includes bank teller windows and heavy bank safes. Barbed wire fencing and cattle guards serve to control access. Guard donkeys and barn cats serve as place managers. Branding cattle identifies property. Stall dividers and horse barn designs serve to avoid equine disputes. Dead coyotes on fences and owl statues in barns give instructions to intruders.

We can readily see that situational prevention depends on routine security built into the design and management of daily life. This applies in the metropolis and Wild West alike, but not always exactly alike. Much more needs to be known about human-animal interaction and predation and human protection of their own animals from others. Crime is part of a much larger ecological process which we are only beginning to understand.

Related Entries

- ▶ Crime Prevention Through Environmental Design
- ▶ Designing Products Against Crime
- ▶ Informal Guardianship
- ▶ Routine Activities Approach

Recommended Reading and References

Arnold C (2001) The ancient cliff dwellers of Mesa Verde. Clarion Press, New York

Clarke RV (1999) Hot products: understanding, anticipating and reducing demand for stolen goods. Police Research Series Paper 112. Home Office, London, UK

Eckblom P (2001) Future imperfect: preparing for the crimes to come. Crim Just Matt 46(1):38–40

Eliason SL (2012) Trophy poaching, a routine activity perspective. Dev Behav 33(1):72–87

Farrell G, Pease K (2006) Criminology and security. In: Gill M (ed) The handbook of security. Palgrave Macmillan, New York, pp 179–199

Felson M (2006) Crime and nature. Sage, Thousand Oaks

Felson M, Rachel Boba Santos (2009) Crime in everyday life, 4th edn. Sage, Thousand Oaks

Jeffrey CR (1977) Crime prevention through environmental design. Sage, Thousand Oaks

Laycock G (1991) Operation identification, or the power of publicity? Secur J 2(2):67–71

Lemieux AM, Clarke RV (2009) The international ban on ivory sales and its effects on elephant poaching in Africa. Brit J Criminol 49(4):451–471

Lingenfelter RE (1986) Death valley and the Amargosa: a land of illusion. University of California Press, Berkeley

Liu JS (2009) Barbed wire: The Fence that changed the west. Mountain Press, Missoula

Newman O (1972) Defensible space, crime prevention through urban design. Macmillan, New York

Pires SF (2012) The illegal parrot trade: a literature review. Global Crime 13(3):176–190

Pires SF, Clarke RV (2011) Sequential foraging, itinerant fences and parrot poaching in Bolivia. Brit J Criminol 51:314–335

Pires SF, Clarke RV (2012) Are parrots craved? An analysis of parrot poaching in Mexico. J Res Crime Delinq 49(1):122–146

Pires SF, Moreto WD (2011) Preventing wildlife crimes: solutions that can overcome the 'tragedy of the commons. Eur J Crimin Pol Res 17(2):101–123

Tilley N, Laycock G (2002) Working out what to do: evidence-based crime reduction. Crime Reduction Research Series Paper 11, Home Office, London

Wortley R, Mazerolle L (eds) (2008) Environmental criminology and crime analysis. Routledge, New York

SNAP®

▶ Stop Now and Plan (SNAP®) Model

SNAP® Boys

▶ Stop Now and Plan (SNAP®) Model

SNAP® GC

▶ Stop Now and Plan (SNAP®) Model

SNAP® Girls

▶ Stop Now and Plan (SNAP®) Model

SNAP® ORP

▶ Stop Now and Plan (SNAP®) Model

So What Criminology?

Roger Matthews[1] and Jennifer Gibbs[2]
[1]University of Kent, Canterbury, Kent, UK
[2]West Chester University at Pennsylvania, West Chester, PA, USA

Synonyms

Administrative criminology; Critical criminology; Public criminology; Realism

Overview

Despite the continued expansion of academic criminology and the proliferation of publications, the policy relevance of criminology has noticeably decreased in recent years. Some commentators have suggested that academic criminology is becoming irrelevant socially and politically and that it has become increasingly marginalized from the process of policy making (Currie 2007). This has led to the identification of a growing body of what has become referred to as "So What?" criminology.

"So What?" criminology has a number of characteristics according to Elliott Currie (2007) including the use of impenetrable language; the adoption of highly technical research techniques which are often dauntingly quantitative, as well as a tendency to focus on trivial issues; and an orientation towards what C. Wright Mills (1959) called "abstracted empiricism." This makes research results difficult to read by anyone other than academic experts in the discipline, while the bulk of the literature remains inaccessible to both policy makers and the general public.

This body of work is produced in an academic context where there is an increased emphasis on publishing in peer-reviewed journals combined with a growing reluctance to engage in the challenging and time-consuming business of policy formation and development. Criminology is also becoming increasingly fragmented and organized into different theoretical and methodological silos, which has to some extent undermined the possibility of constructive debate in the subject producing a body of largely incompatible and partial knowledge. In contrast to these developments, realist criminology argues for the integration of theory, method, and practice and aims to move criminology in the direction of developing a more constructive and socially engaged approach to crime control. Thus, it is suggested that adopting a realist approach provides the possibility of developing an alternative to "So What?" criminology, which is both critical and useful.

Introduction

There is a growing paradox in criminology. Despite the increasing scale of the criminological enterprise involving an increasing number of students, courses, and publications, the policy relevance of criminological research has been decreasing. To make matters worse, this growing body of criminological endeavor suffers from a paucity of theory combined with the adoption of inadequate or inappropriate methodologies. This has provided the basis for the development of what has become known as "So What?" criminology.

This entry will outline some of the main features of "So What?" criminology, which include weak conceptualization, inadequate methodologies, and limited policy relevance. It will also examine some of the recent developments within criminology that have contributed to the growth of "So What?" criminology. Finally, we will offer some thoughts on the contribution that realist criminology can make to overcoming the limitations of "So What?" criminology.

The Main Features of "So What?" Criminology

The term "So What?" criminology refers to those forms of criminology that involve a low level of theorization and thin, inconsistent, or vague concepts and categories; embody a dubious or inappropriate methodology; or have little or no policy relevance (Matthews 2010). Currie (2007) has suggested that in order to overcome the deficiencies of "So What?" criminology, researchers should always aim to present their findings in a way that is easily accessible to policy makers, the general public, and practitioners. Thus:

> . . .some areas of criminological research have become so specialized, so technical and so driven by what has gone before that even many people within the discipline cannot really follow it without help. This means that the need for people who can do the work of synthesis and contextualization of research has become ever greater. (Currie 2007: 183)

Research findings and "criminological wisdoms," he suggests, should be disseminated wherever possible through public debate in order to foster a "public criminology." However, the problem with this approach is that there is little consensus in the field and that most "criminological wisdoms" are highly contested, and it is difficult to find much agreement among criminologists on any issues of significance.

Others have advocated a sort of division of labor within the subject dividing researchers into professional, policy, and critical criminologists, each developing their own forms of expertise and ideally contributing to each other's endeavors (Burawoy 2004). However, fostering such divisions within academic criminology is likely to increase the existing fragmentation of the subject and is more likely to exacerbate than resolve the growing irrelevancy of the ever-expanding criminological enterprise.

One of the most noticeable limitations of much contemporary criminology is poor conceptualization and lack of theoretical rigor. This is most evident in the way in which the central concepts like "crime" and "race" are used. Within the criminological community, there is little clarity or agreement how the concept of crime is to be conceptualized, and most criminologists are divided between those who think that it is an act and those that see it is a product of social reaction. Some use essentially legalistic definitions of crime, while others want to dispense with the concept of crime altogether and replace it with notions of "harm" or human rights (Muncie 1996; Hillyard et al. 2004). Realists have argued that "crime" is a process of action and reaction and have developed the sensitizing notion of the "square of crime" pointing out that the construction of "crime" is a *process* involving four basic elements – the offender, the victim, the state and public opinion, and the social norms (Young and Matthews 1992).

Similarly in relation to "race," rather than deconstructing this concept, all too often "race" is equated with skin color, and diverse ethnic groups are lumped together as if they formed a homogeneous group. Strangely in country like

America that has one of the most diverse ethnic populations in the world, much criminological investigation operates as if the American (criminal) population were divided into "blacks" and "whites," resulting in a form of "monochromatic criminology." In this bifurcated world, even Hispanics, which are soon to be the second largest ethnic group in America, are paid relatively little attention.

The theoretical and conceptual limitations of contemporary criminology, however, go far beyond these two examples. Administrative criminology, which emanates from governmental departments and which plays an increasingly major role in funding and conducting criminological research, is notoriously weak conceptually. Its theoretical weakness, often combined with dubious methodological approaches, tends to result in weak or equivocal conclusions and policy proposals. Often, using common sense or taken-for-granted categories of "crime," administrative and related criminologies veer towards a form of pragmatism which paradoxically is neither very practical nor makes much contribution to developing a cumulative knowledge about crime.

Administrative criminology also tends to gravitate towards empiricism or positivism and assumes that scientific investigation is based on what can be observed. Believing that the lack of theory can be compensated for by the use of "sophisticated" statistical methods, these criminologies tend to reduce the richness and complexities of crime and punishment to a "numbers game" (Young 2004). However, no amount of statistical manipulation can compensate for weak conceptualization.

The problem of method, however, goes far beyond debates about sample size, statistical techniques, and the problems of representation and generalization. Instead, criminological researchers must begin by asking a more fundamental question about how the social world can be best appropriated and understood. Criminologists operate in open systems, which are complex and messy. Social reality does not present itself in an obvious and unproblematic way. To paraphrase Marx: if the social world were intelligible at the level of appearances, there would be no need for social scientific investigation at all. Thus, the aim of investigation is to go beyond appearances and that which is directly observable and to uncover the complexity of a socially stratified and diverse reality. Thus, we need to move beyond simple description of the world and instead try to understand the causal processes involved. At the same time, we need to consider the relation between agency and structure and how the actions of particular individuals or groups are conditioned or shaped by the social structures in which they operate. The ultimate aim of criminological investigation is therefore not to produce descriptions or "facts," as if these speak for themselves, but *explanations*.

Empiricism comes in many forms. In criminology, there is a tendency to what we might call "functional empiricism" and "inverted empiricism." Functional empiricism refers to those approaches which see two observable phenomena that seem to change together and conclude that one must be influencing the other. Thus, Loic Wacquant (2009), for example, claims that there is a "functional equivalence" between the decline of black ghettos in America and the increased incarceration of African Americans and that this is a consequence of neoliberalism. However, why neoliberalism with its commitment to minimal statism would want to expand state provision in this way rather than leaving ethnic minorities to fend for themselves in the ghetto is not clear. It is also the case that in periods of neoliberalism, the number of people sent to prison has decreased as well as increased, suggesting that there is no necessary relation between neoliberalism and penal expansion (Matthews 2009b).

Inverted empiricism refers to those forms of investigation that claim, for example, that there is no causal relationship between crime rates and incarceration on the basis that these two rates do not always covary (see, e.g., Tonry 2001). However, comparing overall crime rates and imprisonment rates is problematic because measurement of these two variables changes over time and between different locations. This line of thinking by focusing on immediate observables tends to neglect a more detailed critical investigation into the attrition rates of different offenses particularly those offenses that are more likely to result in a prison sentence. Thus, naïve empiricists like Michael Tonry (2001) who claims that crime rates and imprisonment rates are unconnected because they do not correlate avoid examining the causal processes in play, and how changing patterns of crime over time as well as changes in the way in which different crimes are processed can influence the scale of imprisonment. An investigation of this type would reveal that the increase in incarceration rates in America is not so much a function of the police and judiciary getting tough on drug offenders, particularly those from ethnic minority groups, as Michael Tonry (1995) claims, but a significant increase in the number of violent offenders being processed (Bureau of Justice 2008).

The limited policy relevance of much criminology is associated with the theoretical and methodological weaknesses outlined above. In addition, many criminologists see themselves as professional academics in the way that Burawoy (2004) describes and largely divorced from policy and practice. Often, under increasing pressure to produce well-cited publications and manage increased teaching and administrative loads, a growing body of criminologists does not have the time or inclination to engage in public or political debate. Among those who are interested in engaging with policy and practice, it is often the case that because they present "facts" rather than credible explanations, their work often does not resonate with politicians or the public. In addition, findings are not presented in an accessible form.

All too often, criminologists dwell on the failures or limitations of different programs and fail to address the question of "what is to be done." Thus, in relation to major issues such as imprisonment and policing, we see endless publications listing the failures and the problems of these institutions and organizations, rather than engaging in a constructive discussion on how they might be improved or usefully reformed.

The Construction of "So What?" Criminology

There are a number of developments that have taken place within criminology over the last 20 or 30 years that have contributed to the development of "So What?" criminology in different ways. First, as mentioned above, there has been a considerable expansion in administrative criminology as a result of the increased dominance of government funding of criminological research. Reese Walters (2003) has questioned how "market-led criminology," which focuses on risk management, privatization, and cost-effectiveness, has influenced the production of criminological knowledge. Government-funded forms of administrative criminology, he suggests, are not interested in generating critical and reflexive research and have in recent years come to undermine and sideline critical criminological research. There can be little doubt that the changing nature of the academy coupled with the increasingly narrow focus of much government-sponsored research has produced a growing body of largely atheoretical research with a limited policy edge. Consequently, much "administrative criminology" often involves policy-driven evidence rather than evidence-driven policy.

Second, postmodernism has had a major influence on the social sciences in general and criminology in particular. Postmodernists have been critical of what they refer to as "grand narratives" and of the modernist belief that social progress can be achieved through the application of scientific knowledge to social problems. Postmodernists consequently present a form of relativism and generally express an anti-objectivism and antirealism. Although postmodernists often end up in producing their own meta-narratives and engage in debates which imply at least some degrees of objectivism and an independent reality about which we can debate, the advent of postmodernism has fuelled a relativism that was already present in criminology. This relativism suggests on one side that everybody's view is equally valid or that certain "standpoints" have privileged position in relation to truth.

Thus, Carol Smart (1992) in her well-referenced critique of "malestream criminology" in which she berates criminology for what she sees as its adoption of grand narratives argues instead for situated knowledge. Suspicious of what she sees as an essentially male-centered criminology, she argues that the truth claims of criminology have to be deconstructed and challenged from a feminist standpoint. Although rightly critiquing partial views of the world that claim to be universal, Smart wants to replace these one-sided views with another set of one-sided views. As John Lea (1998) has argued, replacing the views of the dominant group with those of the marginalized, victimized, and excluded women is to replace one form of fundamentalism with another. In the end, Smart muddies the waters rather than constructing a basis for the development of a feminist criminology. By advocating a form of standpoint feminism together with an antimodernism and antirealism, Smart gravitates towards relativism and foundationalism. From this perspective, it is difficult to see how she might develop a policy response that would reduce rape, domestic violence, and other crimes that victimize women. Thus, postmodernism tends towards defeatism and does little to move the field forward in developing meaningful and effective criminal justice policies.

Third, in conjunction with postmodern defeatism, there is a significant trend in academic criminology towards liberal pessimism. The implicit message of this stance as Stanley Cohen (1985) pointed out is to try to do less harm rather than more good. This brand of pessimism was popularized in the 1970s with the publication of Martinson's (1974) report on corrections – a report that was widely interpreted as claiming that "nothing works." Although Martinson in a subsequent article admitted that certain programs did work for some people under certain conditions, the pessimists embraced the "nothing works" mantra and applied it to different areas of criminology. Significantly, the claim that rehabilitation strategies were ineffective in reducing recidivism allowed a shift away from a commitment to try to rehabilitate prisoners to a policy of incapacitation and "just deserts." This in turn led to the warehousing of prisoners and the creation of what were described in the UK as "penal dustbins."

Liberals, in fact, have found it difficult to develop policies and interventions that are credible and which might resonate with the policy makers and the general public. This is partly because many are antistate or at least minimal statists, while most are anti-punishment. For the most part, their arguments are directed towards the reduction of state intervention and of limiting the use of punishment. Thus, there is a general policy of reductionism among liberal criminologists who argue for a decrease in the prison population and in many cases the use of community-based sanctions instead. However, because liberals often fail to address the issue of who should go to prison and for what purpose and for how long, the question of how far the prison population needs to be reduced can never be answered. The issue of imprisonment is, however, not just a question of numbers. Moreover, as some criminologists have pointed out, the greater use of community-based alternatives to incarceration can lead to "net widening" and the simultaneous expansion of both inclusive and exclusive strategies of control (Cohen 1985).

From the vantage point of liberal pessimists, there is a general wariness about proposing "solutions" or "alternatives." For some, there is little interest in improving prison conditions as this is only seen to relegitimize the use of incarceration. On the other hand, there is a growing skepticism about the use of community-based sanctions. At the same time, there is some suspicion concerning the use of the "welfare sanction" which is seen to deepen the level of state intervention into individual lives while extending the range of intervention to include the wider community (Cohen 1985; Garland 1981). Thus, liberal pessimists find it difficult to commit to any of these forms of regulation.

More recently, liberal pessimists have claimed that there has been a rise in "populist punitiveness" and that both politicians and the public are becoming increasingly punitive (Garland 2001;

Simon 2001; Wacquant 2009). Although these claims are based on thin theoretical and empirical foundations, there is a substantial consensus among liberal pessimists that we are witnessing a punitive turn (Matthews 2005). The main reference points for many of these liberal pessimists are the growing prison population on both sides of the Atlantic. The increase in the number of people in prison coupled with decreasing crime rates is seen as uncontroversial evidence of an increase in punitiveness. However, in relation to the American situation, the prison population leveled off at the end of the 1990s, while in a number of states, the prison population has actually decreased over the past decade. In the UK, we have seen a bifurcated response in recent years with the judiciary giving longer sentences for certain offenses like violence and murder, while giving reduced sentences for other offenses like theft and burglary (Matthews 2009a). This does not look much like the blanket surge in punitiveness that pessimistic liberals suggest. Also, in other countries such as Italy, Canada, and Germany, the incarceration rate has remained stable or declined in recent years (Meyer and O'Malley 2005). In short, these liberal pessimists add little that is useful or constructive to policy formation and in fact in relation to the debate around punitiveness have led us into a conceptual cul-de-sac. Unfortunately, doing nothing or trying to find reasons why proposed interventions are destined to fail creates a form of disengagement that allows policy makers and politicians a relatively free hand to orchestrate criminal justice policy. Criminal justice, however, is too important to allow politicians to make uniformed decisions.

Fourth, the demise of critical criminology in recent decades has contributed to the resurgence of conventional criminology with its natural scientific orientation. At its height of influence in the 1970s, critical criminology provided useful theories on crime and deviance including labeling and subcultural theories. Importantly, critical criminology encouraged critical discussions about the role of power and politics in relation to crime and justice, as well as raising questions regarding values and meanings. However, critical criminology was unable to offer a credible alternative to conventional criminology. In developing a skeptical view of convention attitudes towards crime, it tended towards a romantic and idealistic view of the criminal while downplaying the impact of crime on victims.

Eventually critical criminology was largely absorbed within the framework of mainstream criminology and became at best the "bad conscience" of conventional criminology. The demise of critical criminology was unfortunate, as it left the door open for positivism to gain ascendency within academic criminology. In recent years, however, we have witnessed the growth of cultural criminology that embodies an important critical impulse (Ferrell et al. 2008). Cultural criminology focuses on the experiences of marginalized groups, emphasizing questions of agency, motivation, emotions, and the generation of meaning. Cultural criminology, though, largely ignores the role of the state and wider issues of power. It has been criticized for acting as "zookeepers of deviance," while there is some confusion over what is meant by "culture" (O'Brien 2005). The problems are seen to limit its critical capacity, while from a realist perspective, cultural criminology like its critical criminological predecessors has a limited engagement with policy and social reform. Thus, while cultural criminology has no doubt reinvigorated critical interest among academic criminologists and in many respects breathed new life into the subject, its ability to provide a coherent alternative to mainstream criminology in its current form remains limited.

Countering "So What?" Criminology

By focusing on identifying causal connections between phenomena, realists seek to understand how and why change occurs. In contrast to postmodern defeatism and liberal pessimism, realist criminology involves a modernist perspective, promoting an evidence-based approach to social reform. Following its critical criminological predecessors, realist criminology aims to deconstruct key concepts like "crime" and "race" and to provide explanations rather than present a series of facts or decontextualized descriptions. Also drawing on the work of cultural criminologists, it aims to incorporate an appreciation of the cultural dimensions of crime control recognizing that policy formation will always be subject to cultural variations. In this way, realist criminology aims to develop a joined-up approach that incorporates sophisticated conceptualization and a robust methodology while making a contribution to policy formation.

Realists on Theory. First and foremost, realists recognize that "crime" is not simply a given nor is it an "act" or simply a product of "social reaction." Rather it is a process which involves a complex and changeable relation between offenders, victims, the state, and the public. It aims to avoid the relativism of postmodernists and the idealism of some forms of critical criminology. The aim is neither to romanticize nor demonize the offender but rather to understand the motivational and structural dynamics that create crime and victimization on one hand and which shape formal and informal responses on the other. Understanding the "square of crime" is important in order to remind ourselves that "crime" is neither a "top-down" construction imposed by the criminal justice system nor a "bottom-up" process involving certain "acts" or "behavior" or changing levels of tolerance but a complex relation between these different determinants. For realists, the development of theory is important but is not seen as an end in itself. Theory must be useful and usable.

Realists on Method. Although realists are critical of empiricism and positivism, they are also critical of those qualitative approaches that are purely descriptive. Consequently, from a realist perspective, purely descriptive ethnographies are of no greater value than the most dense and impenetrable forms of inferential statistics. Realists also reject those forms of "cookbook criminology" that claim that there is one methodological approach that is intrinsically superior to all others and should be followed on all occasions. In many respects, realists are methodological pluralists and argue that the selection of methods will be dependent on the nature of the object under study and the type of research questions that are developed (Sayer 2010).

Rather than rely on statistical correlations, the aim of realist investigation is to identify causal connections. The objective is to find out not only what works but how and why it works. It is also recognized that "what works" will vary according to context and also according to the subjects that interventions are directed towards. As Pawson and Tilley (1997) point out, it is not so much that certain programs "work" but that they are effective to the extent that they connect with the capacities of the subjects to whom they are directed.

Realists on Practice. The move away from "nothing works" (Martinson 1974) to "what works" (Sherman et al. 1997) has been a welcome break from the pessimism and impossibilism that permeate contemporary criminology. While the field today focuses on using evidence to form policy, the evidence for "what works" has been disappointing, largely due to a low level of theory and poor conceptualization combined with inadequate and inappropriate methodologies. A great deal of administrative criminology adopts an instrumentalist view of "what works" and pays scant regard for differences in contexts and in the capacities and propensities of the subjects at whom interventions are aimed. Given the complex and ever-changing contexts of social life, the difficulty is to find situations in which successful interventions can be replicated. Indeed, nothing works for everyone all the time. Thus, instead of claiming that a program or policy "works," perhaps it is more appropriate to claim that some programs work for some people under certain circumstances.

Critical realists also have a distinctive view of the nature and meaning of interventions. Interventions are not just practices but theories or hypotheses that postulate the possibility of bringing about an improved outcome. Consequently, interventions are potentially fallible particularly because they deal with complex social realities, as well as dealing with different groups of subjects and will invariably be implemented differently in different contexts.

Intervention provides tests for theories and hypotheses and potentially offers a basis for developing cumulative knowledge about what works in different contexts and for different populations.

Conclusion

Realist criminology, it has been suggested, offers an alternative to “So What?” criminology and aims to provide a coherent and constructive alternative that is able to provide an evidence-based approach to social reform. The task facing criminology is to develop a coherent and integrated approach that emphasizes the role of a theory and which is critical but grounded. Realist criminology argues for the deployment of methods of investigation that are responsive to the object under study and are designed to identify how and why measures work and ultimately to fashion interventions that are aimed at making tangible improvements in the world. In short, realist criminology is theoretical but not theoreticist, critical but not negative or impossibilist, Utopian but grounded in lived experience, methodologically flexible but rigorous, practical but not pragmatic, and policy relevant rather than policy driven.

Related Entries

- ▶ Cultural Criminology
- ▶ Left Realism
- ▶ Postmodern Criminology
- ▶ Randomized Experiments in Criminology and Criminal Justice

Recommended Reading and References

Burawoy M (2004) Public sociologies: contradictions, dilemmas and possibilities. Soc Forces 82(4):1603–1618

Bureau of Justice (2008) Adult correctional population 1980–2007. www.ojp.gov/bjs/glance/covr2.htm.

Cohen S (1985) Visions of social control. Polity, Cambridge

Currie E (2007) Against marginality: arguments for a public criminology. Theor Criminol 11(2):175–190

Ferrell J, Hayward K, Young J (2008) Cultural criminology. Sage, London

Garland D (1981) The birth of the welfare sanction. Brit J Law Soc 8:29–45

Garland D (2001) Culture of control. Oxford University Press, Oxford

Hillyard P, Pantazis C, Tombs S, Gordon D (eds) (2004) Beyond criminology: taking harm seriously. Pluto Press, London

Lea J (1998) Criminology and postmodernity. In: Walton P, Young J (eds) The new criminology revisited. Macmillan, London

Martinson R (1974) What works? Questions and answers about prison reform. Public Interest 35:22–54

Matthews R (2005) The myth of punitiveness. Theor Criminol 9(2):175–202

Matthews R (2009a) Beyond ‘so what?’ criminology: rediscovering realism. Theor Criminol 13(3):341–362

Matthews R (2009b) Doing time: an introduction to the sociology of imprisonment, 2nd edn. Palgrave/Macmillan, London

Matthews R (2010) The construction of ‘So What?’ criminology: a realist analysis. Crime Law Social Ch 54:125–140

Meyer J, O’Malley P (2005) Missing the punitive turn? Canadian criminal justice, balance and penal modernism. In: Pratt J, Brown D, Brown M, Hallsworth S, Morrison W (eds) The new punitiveness. Willan, Collumpton

Mills CW (1959) The sociological imagination. Oxford University Press, New York

Muncie J (1996) The construction and deconstruction of crime. In: Muncie J, McLaughlin E (eds) The problem of crime. Sage, London

O’Brien M (2005) What is cultural about cultural criminology? Brit J Criminol 45:599–612

Pawson R, Tilley N (1997) Realist evaluation. Sage, London

Popora D (1998) Four concepts of social structure. In: Archer M, Bhaskar R, Collier A, Lawson T, Norrie A (eds) Critical realism: essential readings. Routledge, London

Sayer A (2010) Method in social science: a realist approach, 2nd edn. Routledge, London

Sherman LW, Gottfredson D, MacKenzie D, Eck J, Reuter P, Bushway S (1997) What works, what doesn’t, what’s promising: a report to the United States congress. University of Maryland, College Park, Maryland

Simon J (2001) Entitlement to cruelty: neoliberalism and the punitive mentality in the United States. In: Stenson K, Sullivan R (eds) Crime, risk and justice. Willan, Collumpton

Smart C (1992) Feminist approaches to criminology or postmodern woman meets atavistic man. In: Morris A, Gelsthorpe L (eds) Feminist perspectives in criminology. Open University Press, Buckingham

Tonry M (1995) Malign neglect: race, crime and punishment in America. Oxford University Press, New York

Tonry M (2001) Penal reform in overcrowded times. Oxford University Press, New York

Wacquant L (2009) Punishing the poor: the neoliberal government of social insecurity. Duke University Press, Durham

Walters R (2003) New modes of governance and the commodification of criminological knowledge. Soc Legal Stud 12(1):5–26

Young J (2004) Voodoo criminology and the numbers game. In: Ferrell J, Hayward K, Morrison W, Presdee M (eds) Cultural criminology unleashed. Glasshouse Press, London

Young J, Matthews R (1992) Rethinking criminology: the realist debate. Sage, London

Social Bond

► Social Control

Social Capital and Collective Efficacy

Kevin M. Drakulich
School of Criminology and Criminal Justice,
Northeastern University, Boston, MA, USA

Overview

Criminologists have long theorized that some neighborhoods have a greater capacity than others to recognize common problems and act collectively to solve them. Though this basic notion is derived directly from social disorganization theory, collective efficacy – a recent reformulation – has brought substantial new attention to the idea. Collective efficacy borrows from the idea of social capital to rethink the ways that this capacity is rooted in the local connections among neighbors. Specifically, *social capital* refers to a resource potential facilitated by the structure of local networks, while *collective efficacy* is the ability of a group to draw on this resource to recognize common interests and achieve specific tasks related to local social control. This reformulation has also garnered attention for innovations in the operationalization and measurement of the concept. Recent research has established that collective efficacy is an important predictor of differences in the level of crime between neighborhoods and may be significant to our understanding of the disproportionate prevalence of crime in neighborhoods characterized by economic disadvantage and residential instability (Sampson et al. 1997; Morenoff et al. 2001). However, important questions remain about the way residents perceive collective efficacy, the different ways they can choose to intervene when they become aware of a crime problem, and the connections between collective efficacy and other neighborhood phenomena including cultural forces and more formal crime control efforts.

Introduction

In 1942, a now-classic volume was published which made a radical suggestion: the high crime rates found in immigrant and African-American neighborhoods were not the product of the inherent criminality of individual members of these groups – a popular explanation at the time. Instead, Shaw and McKay revealed that delinquency rates were related to the social-structural organization of the city and were highest in the "transitional zone," an area characterized by low home ownership, socioeconomic disadvantage, and other social-structural disadvantages. The enduring power of their work is derived from the carefully collected and presented evidence of this link between neighborhood structure and crime rates.

Shaw and McKay (1942) also pose an explanation for this link, suggesting that, in simple terms, there is something about some neighborhoods that allows them to organize in ways that prevent crime problems or respond to them as they occur. Notably, however, this is somewhat of an amorphous concept, and both Shaw and McKay (1942) as well as a number of subsequent researchers have struggled with two important issues. The first is how to *theoretically* conceptualize exactly what is important about the social organization of a neighborhood. The second is how to *measure* this theoretical concept systematically such that its role as a mechanism can be tested.

After more than 50 years of incremental and mixed progress on these issues, a 1997 study purported to address both issues. *Collective efficacy* attempts to capture "the differential ability of neighborhoods to realize the common values of residents and maintain effective social controls" (Sampson et al. 1997, p. 918). While earlier work relied on either administratively collected data or aggregations of individual reports of personal behaviors, collective efficacy employed individuals as reporters of neighborhood conditions. Any one of these respondents, on their own, may be biased or simply in error in their assessments. Collectively, however, and when controlling for potential individual-level sources of bias, such respondents may produce a reliable and valid measure of a collective phenomenon – or, as Warner and Rountree characterize it, a Durkheimian *social fact*: "a thing distinct from its individual manifestations" (1997, p. 520).

By situating this collective capacity for control in the resource potential of specific forms of personal and organizational networks (Sampson et al. 1999, p. 635), collective efficacy potentially represents significant progress in our understanding of the group social processes relevant to the spatial distribution of crime. This entry begins by providing background on collective efficacy as a theoretical mechanism, tracing its development through several previous propositions. Collective efficacy, however, departs from previous propositions through several theoretical and measurement innovations. The following section details these innovations and surveys research on its importance to a variety of neighborhood phenomenon. The final section raises a number of potential controversies and open questions facing future work on collective efficacy.

Background

Simply stated, social disorganization theory suggests that the structural conditions of a neighborhood, especially the level of poverty and residential instability, disrupt local social organization and thereby increase crime. The direct association between the structural conditions and the crime rate is relatively easy to observe using administrative data, such as the US Census and the Uniform Crime Reports. More difficult to observe is the social mechanism by which this structure affects crime rates. Over time, research on social disorganization has frequently revised the theoretical operationalization of what it is about the organization of the neighborhood that matters, while simultaneously experimenting with a variety of different more or less successful means of measuring these theoretical mechanisms.

Shaw and McKay (1942), in their initial proposal of the theory, suggest several theoretical mechanisms. Some components of their explanation – including value heterogeneity and the cultural transmission of delinquent traditions between generations of boys within neighborhoods – have been largely ignored or dropped in subsequent revisions to social disorganization, including in the notion of collective efficacy. One component of their explanation, however, is the direct intellectual antecedent of collective efficacy. Drawing on a classic account of Polish communities in Chicago by W.I. Thomas, Shaw and McKay (1942) suggest that an "organized" community is characterized by "the presence of social opinion with regard to problems of common interest, identical or at least consistent attitudes with reference to these problems, the ability to reach approximate unanimity on the question of how a problem should be dealt with, and the ability to carry this solution into action through harmonious co-operation" (p. 184).

In subsequent years, Kornhauser's (1978) critique of the theory led to an increasing focus on the mechanism of social control to the exclusion of the subcultural and learning mechanisms – those explanations interested in value heterogeneity or the transmission of delinquent skills and values. Simultaneously, an emphasis on social ties among neighbors rose to prominence following Kasarda and Janowitz (1974) argument for a systemic model of social organization in which the local community is "a complex system of friendship networks and formal and informal associational ties" which are in part a product of local structural conditions (p. 329). Out of these trends came a new operationalization of the mechanism at the heart of social disorganization: a disruption of the social networks among neighbors which damages the neighborhood's capacity to self-regulate (e.g., see Bursik and Grasmick 1993).

While research inspired by these ideas represented both theoretical and methodological progress, it was still hindered by basic measurement issues. In perhaps the best example of this work, Sampson and Groves (1989) operationalize the mechanism of social disorganization as a combination of sparse local friendship networks, the presence of unsupervised teenage peer groups creating a nuisance, and low levels of organizational participation. In the measure of nuisance-causing unsupervised teenage groups, Sampson and Groves are attempting to directly measure a neighborhood's capacity for informal social control. Unfortunately, by operationalizing social disorganization via an outcome (the presence of rabble-rousing teenagers) rather than the social mechanism itself (the capacity for informal social control), Sampson and Groves open themselves up to the classic criticism of social disorganization research as using delinquency "both as an example of disorganization and something caused by disorganization" (Pfohl 1985, p. 167).

Sampson and Groves (1989) find a strong relationship between their measure of the presence of trouble-causing unsupervised youth groups and a variety of measures of local crime, but their measure of local friendship networks has somewhat less success. Warner and Rountree (1997) find similarly mixed effects for the role of local social ties in relation to local crime rates: they find that measures of the prevalence of interactions between neighbors (borrowing tools, having lunch or dinner, helping neighbors with problems) are negatively related to the crime rate only in largely white neighborhoods and appear unrelated to crime in neighborhoods with larger numbers of racial or ethnic minorities. Critics of the idea of social disorganization pointed out that some high-crime communities in fact had strong *criminal* organization (recent example in Venkatesh 1997; see discussion in Sampson 2002). Critics of the systemic model of social disorganization pointed out that many poor communities had dense social networks (Gans 1962) and that in some communities, ties between neighborhoods prevented informal social control efforts when they bridged the gap between residents who were involved in crime and those who may otherwise have organized against it (Pattillo 1998).

As of the 1990s, then, research on social disorganization theory was faced with several key issues. First, the exact role of social networks in social organization remained ambiguous. Second, there appeared to be a need to distinguish organizational activities that suppress crime versus those that encourage crime. Third, research in the area continued struggling to identify direct measures of any of these theoretical concepts to test the core hypotheses.

Social Capital and Collective Efficacy

The concept "collective efficacy" attempts to resolve some of the issues in social disorganization theory by grounding it in new ideas about the dynamics of interpersonal connections within groups and the capacity these connections create for collective action. By asking residents to report on their perceptions of their neighbors' reactions to hypothetical local problems, collective efficacy attempts to capture a task-specific *potential* for collective action rooted in the social networks of a neighborhood. Collective efficacy, then, is the collective realization among neighbors that they have a common goal in the safety of their neighborhood and a collective willingness to act to achieve this goal (Sampson et al. 1997).

Collective efficacy has found success by making three key theoretical revisions to older versions of social disorganization theory and by using theses revisions to better operationalize and measure the presence or absence of this phenomenon within neighborhoods. Theoretically, Sampson et al. (1997); (see also Sampson 2002) define collective efficacy as rooted in social capital: "the resource potential of personal and organizational networks" (p. 635), as task specific, and in particular as oriented toward the universally shared goal of living in a safe neighborhood. They find empirical evidence for a role for collective efficacy by operationalizing it as potential rather than actualized organization and as an emergent property of neighborhoods best captured by using local residents as imperfect reporters of this group-level phenomenon. The following sections discuss each of these theoretical and measurement dimensions that distinguish collective efficacy from earlier versions of social disorganization.

Social Capital

Perhaps the most important theoretical development for the reformulation of social disorganization theory is Coleman's (1988) version of social capital. Coleman (1988) combines economic rational choice theory with sociological normative theory to create the notion of "social capital," with the premise that "rational or purposive action, in conjunction with particular social contexts, can account not only for the actions of individuals in particular contexts but also for the development of social organization" (p. S96).

In a simple sense, social capital suggests that certain kinds or qualities of connections among people can facilitate particular actions that may not otherwise have been possible. It can take a variety of forms which may be better or worse at facilitating different kinds of actions. One such form is a marketplace of obligations and expectations in which people do favors or help one another with the expectation of indirect reciprocation (Coleman 1988). This form of social capital provides a framework for understanding why a resident would undertake an action – for example, intervening when they see teens vandalizing public property or a neighbor's house – that has minimal direct benefits for themselves. The key is that when that resident lives in a community with an active marketplace of obligations and expectations, they can expect that their favor will be reciprocated by others at some other point when the resident currently performing the favor has a need of their own arise. Other forms of social capital include the transmission of potentially useful information through networks maintained for other purposes or the presence of norms and effective sanctions. Coleman (1988) describes, for example, a norm in which all local adults are responsible for the safety of all local children in public spaces.

These forms of social capital are facilitated by social structure. The trust required for a marketplace of obligations and expectations, for instance, is more likely to emerge in a community with relative residential stability – the presence of long-term residents who may be expected to remain in the neighborhood long enough to repay debts into the collective pool. Such a marketplace may also be more likely when residents have enough resources to be able to do favors for neighbors but not so many resources that they will never need to ask neighbors for favors (Coleman 1988). Coleman (1988) also identifies particular kinds of relational structures that produce social capital, one of which is relevant here. *Closure* in social networks refers to the presence of sufficient ties between network actors. One specific form of closure relevant to collective efficacy is intergenerational closure, in which parents know the parents of their children's friends, through which they can more effectively monitor and sanction their own children. In simpler terms, it is harder to do something without your parents' knowledge when they know all your friends as well as the parents of your friends. Thus, network closure is an aspect of social structure that facilitates certain kinds of actions within the structure.

Social capital, then, suggests that elements of local social organization provide neighborhoods with a social resource that can be drawn upon to address local problems. For instance, in highly residentially stable neighborhoods, there is more likely to be a high degree of interconnectedness among residents. Due to this interconnectedness, residents may feel a sense of trust that if they help a neighbor, another neighbor may be willing to help them when they need it. This, Coleman argues, provides a social resource potential that can be drawn upon to address local problems. Similarly, interconnections between local parents and local children may both increase the surveillance of and provide greater adult resources for local children. Thus, social capital describes a process by which structural characteristics of a neighborhood like residential stability assist in the development of particular forms of social organization, which in turn can facilitate specific kinds of local actions.

Collective Efficacy as Task-Specific Action

One key to social capital is that – like physical or human capital – it often exists as a value-neutral tool that could be appropriated to different kinds of actions. A person could use a biology degree to cure cancer or create a weapon; spray paint can be used for art or vandalism. Similarly, a neighborhood could use comparable kinds of social capital alternatively to address a rash of burglaries, to prevent African-Americans from moving into their neighborhood, or even to encourage residents to refrain from reporting crimes to or cooperating with the police.

One type of action that can result from this social resource potential is attempts to control local delinquency or crime. *Collective efficacy* is the collective realization among neighbors that they have a common goal in the safety of their neighborhood and a collective willingness and ability to act to achieve this goal (Sampson et al. 1997). Notably this conception of social organization has changed little since Shaw and McKay's (1942) description of "harmonic cooperation." The theoretical advancements since Shaw and McKay have largely come in two areas: conceptualizing this organization as task specific (organization specifically against crime will be more relevant to crime as an outcome than more general kinds of organization) and developing a better model of how such a capacity might develop, especially as a product of the structural characteristics of a community. In practice, the task-specific potential to address crime is so highly related to the presence of specific local forms of social capital that the most frequent measure of collective efficacy combines a direct measure of the capacity for informal social control with a measure of social cohesion and trust (Sampson et al. 1997; Morenoff et al. 2001).

This conceptualization helps resolve some of the questions about the role of social ties and networks from prior work. Specifically, Sampson et al. (1999) suggest that collective efficacy is a task-specific potential for action that is rooted in social capital: "the resource potential of personal and organizational networks" (p. 635). In this way, collective efficacy departs from some of the more recent versions of social disorganization by de-emphasizing the importance of close friendship networks and instead suggesting that a resource potential exists in particular forms of local social organization, including more utilitarian and less affective connections among persons. Thus, a place where neighbors trust one another enough to develop a marketplace of obligations and expectations may possess collective efficacy even when neighbors are not actively friendly with one another, while a different neighborhood with dense friendship ties may still lack it. In fact, Granovetter (1973) has described the ways in which the preponderance of dense and largely redundant ties will be less useful in exposing people to new ideas or information than ties that bridge the gaps between different groups of people.

Measurement as Potential for Action

The key to measuring this new concept is to capture the *potential* for collective action rather than measuring the presence or absence of potential outcomes of such actions, as Sampson and Groves (1989) did with their measure of perceived juvenile delinquency. The potential, in this case, is reflected in a norm: *shared expectations* for informal social control. The measurement of this kind of norm – something that exists between neighbors rather than within any one neighbor – requires the perspectives of those residents. Sampson et al. (1997) accomplish this by using questions which ask residents to speculate about how their neighbors would respond to a series of hypothetical situations involving delinquent acts by youths. Respondents are asked about the likelihood that their neighbors would take action if they encountered one of a number of instances of youth delinquency. These include children skipping school and hanging out on a street corner, children spray-painting graffiti on a local building, children disrespecting an adult, and circumstances in which a fight breaks out in front of their house. Additionally, some versions of collective efficacy have included an additional noncrime outcome: the likelihood of neighbors acting to save a local fire station threatened by budget cuts.

By asking respondents to report their perceptions of the level of collective efficacy in their neighborhood, they attempt to capture a *potential* for action which exists in the connections among persons in the neighborhood rather than within any one resident or in any simple aggregation of residents. In this sense, collective efficacy represents an emergent property of neighborhoods – shared expectations about the control of local delinquency – that theoretically would not be captured by the simple aggregation of reports of individuals' willingness to engage in such control.

Reflecting the theoretical relationship between social capital and collective efficacy, most measures of collective efficacy also include items capturing the degree of social cohesion and trust in the neighborhood. Though the two are distinct concepts theoretically and it appears plausible that collective efficacy is to some degree a product of local social capital elements such as cohesion and trust, in practice the two appear highly correlated and thus have frequently been combined to create a single scale.

In doing so, collective efficacy appears to be more successful than earlier measures of the social disorganization mechanism in predicting local levels of violent crime and even mediating some of the association between structural conditions and violence (Sampson et al. 1997; Morenoff et al. 2001). Research has also found collective efficacy to be predictive of fear of crime (Gibson et al. 2002), partner violence (Browning 2002; Wright and Benson 2011), as well as outcomes less directly related to crime including health and mental health and sexual behavior outcomes such as pro-social competency and problem behavior, birth weight, premature mortality, short-term sexual partnering, and age of sexual initiation. Collective efficacy is also suggested to be the source of a spurious disorder-crime relationship suggested by broken windows theory (Sampson and Raudenbush 1999).

Controversies, Unresolved Issues, and Open Questions

By situating the collective capacity for control in the resource potential of specific forms of personal and organizational networks, collective efficacy represents significant progress in our understanding of the group social processes relevant to the spatial distribution of crime. Operationalizing collective efficacy as an emergent task-specific potential for collective action against delinquency and crime is a noteworthy effort to resolve the difficult problem of measuring phenomena which are a property of groups, or a property of the connections among persons, rather than a property of the persons themselves. This concept represents substantial progress over earlier attempts to theorize and measure a neighborhood's capacity for social control. However, several unresolved issues remain which pose questions for future research.

Lack of Mediation and Measurement Issues

Despite evidence for an apparent role for collective efficacy in the distribution of crime over neighborhoods, the structure of a neighborhood continues to exert strong direct effects on the crime rate. Although Sampson et al. (1997) claim strong mediation effects, Morenoff et al. (2001) suggest this claim may have been overstated. Though collective efficacy appears to have a direct effect on crime, it may not be the long-sought mechanism explaining the link between social structure and crime – or it may only be one component of such a mechanism. Alternatively, this lack of mediation may be a product of measurement error related to ways in which collective efficacy has been operationalized – specifically the difficulty involved in speculating about the potential reactions of other people to hypothetical situations. This raises a key issue for both the measurement of the concept of collective efficacy and the theory itself: we still know too little about how residents make assessments of the local capacity for collective efficacy.

How Do People Intervene?

Another area for future research is to develop a better understanding of how residents choose to intervene when they become aware of crime problems. Residents may directly intervene by confronting troublemakers or they may indirectly intervene by mobilizing external resources including the police department. Residents may be fearful that intervening in local problems could result in retaliation or simply put them at greater risk for victimization. Recent work has suggested just this: that a fear of reprisals acts as an impediment to participation in traditional forms of informal social control (Carr 2003; St. Jean 2007). Instead, such residents may choose to work through the police department to exert control over local crime, actions which do not appear to require the cohesion necessary for informal efforts (Carr 2003; Warner 2007). On the other hand, neighborhoods which lack faith in the police may be wary of choosing to intervene through the police, and in some of these neighborhoods, residents may choose to attempt to address some specific crime problems through local criminal organizations like gangs (Venkatesh 1997; Pattillo 1998). Thus, while collective efficacy research has largely been interested in *whether* neighborhoods are willing to intervene, we may also want to consider *how* neighbors choose to intervene and the differential consequences for neighborhoods resulting from these choices.

Culture and Norms About Crime: Universality or Heterogeneity?

Collective efficacy assumes that communities share common values with respect to local public safety (Sampson 2002) and – following Kornhauser (1978) – has largely eschewed a role for culture in encouraging crime. Kubrin and Weitzer (2003) suggest, however, that culture may play several different roles that are relevant to the collective organization against crime; future work would do well to consider the potential links.

Early criminological work on culture emphasized the prevalence of pro-crime values in lower-class or minority communities. Often portrayed as a reaction to strain over status, respect, or access to legitimate or illegitimate economic opportunities (Cohen 1955; Cloward and Ohlin 1960), such communities were characterized as having pro-crime or crime-tolerant cultural norms. These norms emerge as a negativistic or oppositional reaction to societal norms (Cohen 1955; Anderson 1999), or as an appropriate or predictable reaction to economic circumstances (Miller 1958; Cloward and Ohlin 1960). Shaw and McKay's (1942) early work in social disorganization characterized disorganized neighborhoods as suffering from a kind of normative heterogeneity in which children will be exposed to examples of both criminal and conventional behavior. Sutherland (1947) suggests that both organization against crime and organization in favor of crime will be relevant to understanding the local crime context.

Alternatively, instead of pro-crime values mattering, it may be that communities simply differ in the strength with which they claim conventional norms (Kornhauser 1978; Warner 2003). It may be that these communities are simply more resigned or accustomed to the presence of crime – that crime has infiltrated their cognitive landscape. Finally, culture may matter in ways that have little to do with values about crime. Small (2002), for instance, suggests that cultural framing mechanisms have the ability to either incite or sustain collective organization, including organization toward the goal of addressing local crime problems. For instance, a framing of the neighborhood as in crisis or under threat may *incite* organization (Small 2002; Carr 2003), while a cultural frame defining a neighborhood as one's best opportunity for success is necessary to *sustain* continued organization (Small 2002).

Connections Between Formal and Informal Social Control

Collective efficacy is, in essence, the capacity for *informal* social control. Of course, the police also act as agents of more *formal* social control efforts, and prior work has suggested that neighborhoods differ in both the behavior of police and perceptions of the police. The capacity for informal social control, then, may also be related to local formal efforts. This idea has only just begun to receive serious attention (see Kubrin and Weitzer 2003). Recent work has found evidence for a link between perceptions of the police and informal social control efforts.

Conclusion

Collective efficacy draws on the notion of social capital to present a new theoretical and methodological conceptualization of differences between neighborhoods in their ability to recognize common problems and act collectively to solve them. Social capital is grounded in social networks – it "exists in the relations among persons" (Coleman 1988, p. S100) – and is facilitated by particular kinds of local structural conditions such as residential stability or the presence of formal organizations. In turn, collective efficacy is the ability of a community to draw on this resource to recognize a shared interest in the safety of the local area and to achieve specific tasks related to local informal social control. Recent research has provided evidence of a role for collective efficacy in explaining inter-neighborhood differences in crime rates. Important questions remain, however, about how respondents estimate their neighbor's capacity for informal social control, the different ways people choose to intervene, connections to cultural dimensions of a neighborhood, and the links between formal and informal social control efforts.

Related Entries

- ▶ Differential Social Organization
- ▶ Disadvantage, Disorganization and Crime
- ▶ Early Chicago School Theory
- ▶ Informal Social Control
- ▶ Neighborhood Effects and Social Networks
- ▶ Race and Ethnicity in Social Disorganization Theory

Recommended Reading and References

Anderson E (1999) Code of the street: decency, violence and the moral life of the inner city. W. Norton, New York

Browning CR (2002) The span of collective efficacy: extending social disorganization theory to partner violence. J Marriage Fam 64:833–850

Bursik RJ Jr, Grasmick HG (1993) Neighborhoods and crime: the dimensions of effective community control. Lexington, New York

Carr PJ (2003) Clean streets: controlling crime, maintaining order, and building community activism. New York University Press, New York

Cloward R, Ohlin L (1960) Delinquency and opportunity: a theory of delinquent gangs. Routledge, London

Cohen AK (1955) Delinquent boys: the culture of the gang. Free Press, Glencoe

Coleman JC (1988) Social capital in the creation of human capital. Am J Sociol 94:S95–S120

Gans HJ (1962) The urban villagers; group and class in the life of Italian-Americans. Free Press, Glencoe

Gibson CL, Zhao J, Lovrich NP, Gaffney MJ (2002) Social integration, individual perceptions of collective efficacy, and fear of crime in three cities. Justice Q 19:537–564

Granovetter MS (1973) The strength of weak ties. Am J Sociol 78:1360–1380

Kasarda JD, Janowitz M (1974) Community attachment in mass society. Am Sociol Rev 39:328–339

Kornhauser RR (1978) Social sources of delinquency. University of Chicago Press, Chicago

Kubrin CE, Weitzer R (2003) New directions in social disorganization theory. J Res Crime Del 40:374–402

Miller WB (1958) Lower class culture as a generating milieu of gang delinquency. J Soc Issue 14:5–19

Morenoff JD, Sampson RJ, Raudenbush SW (2001) Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. Criminology 39:517–560

Pattillo ME (1998) Sweet mothers and gangbangers: managing crime in a black middle-class neighborhood. Soc Forces 76:747–774

Pfohl SJ (1985) Images of deviance and social control. McGraw-Hill, New York

Sampson RJ (2002) Organized for what? Recasting theories of social (dis)organization. Crime Soc Disorgan 10:774–802

Sampson RJ, Groves WB (1989) Community structure and crime: testing social-disorganization theory. Am J Sociol 94:774–802

Sampson RJ, Raudenbush SW (1999) Systematic social observation of public spaces: a new look at disorder in urban neighborhoods. Am J Sociol 105:603–651

Sampson RJ, Raudenbush SW, Earls F (1997) Neighborhoods and violent crime: a multilevel study of collective efficacy. Science 277:918–924

Sampson RJ, Morenoff JD, Earls F (1999) Beyond social capital: spatial dynamics of collective efficacy for children. Am Sociol Rev 64:633–660

Shaw CR, McKay HD (1942) Juvenile delinquency in urban areas. University of Chicago Press, Chicago

Small ML (2002) Culture, cohorts, and social organization theory: understanding local participation in a Latino housing project. Am J Sociol 108:1–54

St. Jean PKB (2007) Pockets of crime: broken windows, collective efficacy, and the criminal point of view. University of Chicago Press, Chicago

Sutherland EH (1947) Principles of criminology, 4th edn. Lippincott, Philadelphia

Venkatesh SA (1997) The social organization of street gang activity in an urban ghetto. Am J Sociol 82–111

Warner BD (2003) The role of attenuated culture in social disorganization theory. Criminology 41:73–98

Warner BD (2007) Directly intervene or call the authorities? A study of forms of neighborhood social control within a social disorganization framework. Criminology 45:99–129

Warner BD, Rountree PW (1997) Local social ties in a community and crime model: questioning the systemic nature of informal social control. Soc Probl 44:520–536

Wright EM, Benson ML (2011) Clarifying the effects of neighborhood context on violence “behind closed doors”. Justice Q 28:775–798

Social Constructivism

► Transitional Justice

Social Control

Megan C. Kurlychek
School of Criminal Justice, University at Albany,
State University of New York, Albany, NY, USA

Synonyms

Social bond

Overview

Social control refers to the mechanisms through which a society is able to regulate and direct the behaviors of its members. These mechanisms take many shapes and sizes and are often classified by type and/or level. For example, social control is often divided into two types: informal and formal. Informal sanctions encompass such things as not engaging in an action so as not to disappoint one’s parents or because one believes the act to be wrong. Formal social controls involve more direct action against a person such as suspension from school or arrest and incarceration in the criminal justice system. These mechanisms also traverse various levels or domains of control ranging from the private to the public realm. The private level refers to those modes of control that are the most intimate or close to the person such as the control exhibited by family and friends and is the level most likely to contain informal measures. The next level, often referred to as the parochial or institutional level, involves those controls invoked by social institutions such as churches and schools. The public domain represents manners of control that are the most removed from the individual. This includes broader social connections between individuals, institutions, and the greater society and is often realized through formal legal sanctions. The levels of control are often complementary, interactive, and interdependent as well.

The following essay provides the reader with an introduction to the study of social control, formal theories of social control, and modern extensions and analytical techniques that have evolved to account for its multifaceted and multilevel nature. This research is grounded in two complementary sociological perspectives: social disorganization and social control. Originating in studies of the break down, or “disorganization” of communities, research in the first tradition looked to structural characteristics of a neighborhood that were associated with high crime rates. Later theorists then drew clear connections between the concepts of “disorganization” and “social control” clearly noting that the two concepts are but distinct ends of a continuum between high and low levels of social solidarity (Kornhauser 1978). When high, the community is able to exert both levels of informal and formal social control upon its members to achieve desired social outcomes. When low, on the other hand, individuals are left to their own devices.

From a social control perspective, this is not a good thing. That is, social control theorists do not begin with the question of why individuals commit crime, but rather why they do not (see, Hirschi 1969). The assumption being that left to our own devices, humans will always pursue pleasure and will do so in the easiest way possible. Since crime often offers high excitement and pleasure with less effort than traditional mechanisms (e.g., hard work), social control is then necessary to inhibit individuals from taking this path.

The following section reviews early studies of the organization, or rather the breakdown of the organization, of communities and their elaboration into formal theories of crime and delinquency. Connections between social disorganization and social control theories are described. Later sections then detail modern extensions of these theories and advances in scientific research that have informed the current state of knowledge and set forth pathways for future research.

Early Studies of Community Structure and Social Control

The 1996 publication of Hilary Clinton's *It Takes a Village: And Other Lessons Children Teach Us* brought national attention to the notion that the socialization of our nation's children is not just a family affair but involves all members of the local community. While this heightened attention to the crucial function of the community in the control and socialization of our nation's youth may seem somewhat novel, it is in fact nothing new to those who study societies and the mechanisms by which societies control their citizenry. Indeed, philosophers as far back as Aristotle noted the importance of community governance in maintaining social order and raising children. Beyond mere philosophy, a rich tradition of research and theory exists that details the important and complex relationship between community structure, interpersonal social relationships, and the exertion of social control upon a community's residents, particularly its youth.

The formal study of social control in America can be traced back to studies of the growth of American cities such as Chicago around the turn of the twentieth century. This school of thought, often referred to as "The Chicago School," approached the question of social control from the opposite direction – that is, this research examined what happens when a community is not able to exert social control upon its members. W. I. Thomas and Florian Znaniecki provide some of the earliest empirical work in this area writing on the life experiences of immigrant families trying to "make it" in the new world (1927). Of the many struggles these families faced, one of the greatest was raising their children in a community setting that was unfamiliar to them. The traditions in which they were raised no longer existed, and new traditions had yet to be established in their place. Thus these immigrant parents were ill prepared to socialize their children to this new environment. As immigrant populations tended to settle together (e.g., Polish families with Polish families, Irish families with Irish families, etc.), these problems rose above the individual or family level to become a larger group and even community phenomenon.

With the breakdown in traditional parental socialization roles, this research also found peer networks to take on a heightened influence in the lives of children. This is further demonstrated in an influential writing by Thrasher, *The Gang* (1927), that presented a chronicle of the evolution of neighborhood childhood playgroups into more formal "gangs" – some of which were criminal, some of which were not. According to this study, one primary feature of delinquent gang formation was that it filled a space or crack in social solidarity. These cracks, referred to as interstitial areas, were seen as pockets in society in which the community could not exhibit social control. These were the areas in which gangs were then free to form. As described by Thrasher, the gang provided the essential human needs of belonging and camaraderie that the family and community had failed to provide.

In addition to these qualitative accounts of life in a growing city, researchers at the University of Chicago began the development and formal testing of theories of community growth and governance. The most prominent of these theories was social disorganization theory developed by Clifford Shaw and Henry McKay. Applying a model of community growth developed by colleagues Robert Park and Ernest Burgess, Shaw and McKay were able to document patterns of arrest in the city over a 30-year time period. During this time, Chicago was experiencing rapid population growth primarily due to large influxes of immigrants from Europe as well as southern states. The city, in turn, started to experience high levels of crime. While some tended to place blame on the latest group of immigrants – be they from Poland, Ireland, Italy, etc. – Shaw and McKay conducted a scientific study of the arrest of delinquent youths and found that the criminal activity was concentrated in a small area of the city encircling the central business district which became known as the "Transition Zone." This area was in flux as neighborhoods that were once homes were swallowed up by business as the central business district (referred to in Chicago as the Loop) expanded. Housing in this area became relatively cheap as landowners stopped investing in properties they knew would soon be demolished. Because the rents in this area were cheap (and the housing in poor condition) it was often that the newest group of immigrants would first live in this area until they found jobs and gained solid footing in the "New World."

By mapping out crime data for over 30 years, Shaw and McKay were able to show that the crime remained in this area of the city regardless of the population inhabiting the area. In addition, they found that when crime did "move" to a new area it was because that area was experiencing conditions similar to those neighborhoods where crime had previously been concentrated. Thus, this study found that "social disorganization" and the high crime rates that attend it are primarily characteristics of a place and not the people who inhabit that space.

The characteristics of the place most often associated with social disorganization were high poverty, high residential mobility (people moving in and out quickly), and high racial heterogeneity. The first, poverty, was influential because there was little money to invest in improving the area. Homes would become dilapidated and community institutions such as schools, churches, and social organizations would break down. Second, because the neighborhood was "deteriorating" rather than being maintained, this is not an area where people desired to put down roots. Rather it was a stepping stone, an area where they would live briefly only until they could afford something better. Thus, once an individual or family acquired some wealth, they did not invest it in this area but rather they would move away leading to high residential mobility. Thirdly, as people came and went, especially during this time of high immigration, the area became populated by people with very different cultures, values, and even languages. This racial and ethnic heterogeneity lead to conflict and mistrust between groups as well as the inability of inhabitants to form basic social networks that help to hold a community together. Social Disorganization theory therefore posited that it was a combination of these three elements – poverty, residential mobility, and racial/ethnic heterogeneity – that lead to high levels of crime and delinquency in a given community.

Social Disorganization and Social Control: Making the Connection

While Shaw and McKay offered these basic renditions explaining why these structural factors might be of importance, it wasn't until much later that this theory was formally recognized as a version of Social Control theories and further elaborated from a pure macro-level theory to an examination of the microlevel processes that occur within a community. In a classic 1978 writing, Ruth Kornhauser examined many existing theories of crime and delinquency detailing their primary assumptions and classifying them based upon these underlying assumptions. Through this process she was able to clearly demonstrate the inherent ties between social disorganization and social control theories. According to Kornhauser, the structural factors related to disorganization of a community were indeed important predecessors of crime and delinquency *because* they lead to the disruption of the various components of both informal and formal social control.

In addition to highlighting the connections between social disorganization and social control, Kornhauser's work brought renewed attention to the social disorganization perspective by pointing out that the primary overlooked components of Shaw and McKay's work were the processes through which social disorganization at the neighborhood level is linked to individual-level delinquency. Not only are community structural characteristics important for understanding neighborhood crime levels. These ecological and social causes set the stage for a variety of undesirable consequences including weak social bonds developed between the individual and society, the inability of institutions to enforce common standards and goals, the defective socialization of children, the lack of continuous social networks, and, even the lack of the ability to take action to fix these problems. During the 1970s, 1980s, and 1990s, the search for a better understanding of the social processes by which social disorganization leads to reductions in social control became a fruitful branch of research for criminologists. The following section outlines some of the key knowledge gained during this time.

Extensions of a Theory: Systems of Social Control

John Kasarda and Morris Janowitz (1974) introduced what they coined the "systemic model" of a community control which envisioned the local community as a complex system of friendship and kinship networks. These networks were rooted primarily in family life and were the basis for the ongoing processes of socializing youth in the family and community. This research then tied these friendship and kinship networks to the social structural characteristic of residential mobility proposed by Shaw and McKay. Specifically, the authors proposed that residential mobility is a temporal variable as the development of social bonds and networks takes time. The longer one lives in a community, the more people one knows and the more opportunity one has to become involved in community activities and institutions. When people are moving in and out of an area fairly rapidly, these basic mechanisms naturally break down. This then leads to a reduction in the social networks a person is able to develop. Fewer social networks in turn diminish the ability of the family to properly socialize children and to enforce social controls, informal and formal, upon youth.

Subsequent research has further documented the importance of social ties. For example, Coleman (1988) argued that it is the resources transmitted through social ties, not the ties per se that are key to facilitating neighborhood social control. Similar to the familiar notion of financial capital (e.g., one's financial resources), these social ties and networks became known as a person's "social capital." Research in this area has examined such things as the number of family members living close by, the number of neighbors one knows by sight and by name, family composition (e.g., single-parent families), and the number of moves a family makes as potential indicators of social capital.

Although research has evolved to measure and explain social capital in various ways, it is important to remember that its primary nature is that it is something not housed within an individual but rather within the structure of the community in which the individual lives. Research by Robert Sampson and colleagues has been highly influential in the development of this concept and has resulted in the development of several terms and measures that are now generally accepted as a standard in the social control literature. The first of these is Intergenerational Closure. Intergenerational Closure reaches beyond the immediate family to measure the extent to which children are linked to adults within the community. While this includes immediate family, it also includes the parents of friends, as well as other adult figures such as teachers, coaches, church officials, or even just other concerned resident.

A second concept, reciprocal exchange, moves beyond the mere existence of such relationships and places emphasis on the ability of community members to rely on each other to exchange items (e.g., borrow a tool), information (e.g., tryouts for a sport team or new community organization), and even advice such as turning to one another for advice on child rearing. This concept encompasses more than merely knowing other community members to incorporating levels of trust between community members.

A third concept that has been highly influential in research in this area is that of "collective efficacy." According to research social capital may be a necessary, but not sufficient, aspect of neighborhood relations in the exhibition of social control. In addition to forming ties and resources in one's community, collective efficacy then refers to the willingness and ability of community members to come together and take action on a specific task. One such important task is the socialization of youth. Collective efficacy could also be visualized as a community coming together to earn funds to build a park, start a youth soccer league or improve a library. A classic example often given is the creation of a neighborhood or block watch group in which community members take turns patrolling the neighborhood and reporting suspicious people or activity in a direct effort to reduce crime. Not surprisingly, empirical evidence shows strong relationships between high levels of collective efficacy in a community and low levels of crime. In addition, some research has suggested that factors such as friendship networks and participation in community associations matter *only* through their ability to promote collective efficacy among residents (Morenoff et al. 2001).

Another aspect of social relationships brought to light in this research was social cohesion. While this particular term has several various definitions in general it refers to the sharing of common goals and values. For example, it would be hard for a community to come together to achieve a common goal (e.g., collective efficacy), if members did not agree on the goal. Just as residential mobility was shown to influence crime through its impact on social ties and networks, another social structural characteristic – racial and ethnic heterogeneity – appears to exert its influence mainly through its impact on another social interactive process. In this case, the effect is upon levels of social cohesion and the sharing of common beliefs and goals. Logically, the more different people are in a community the less likely they may be to share common goals and/or to be willing to come together and take action to achieve such goals.

Scholars note, however, that social cohesion can indeed go both ways when it comes to influencing crime and delinquency. That is, a family with strong cohesion with criminal members might then promote criminality in other members. As such, a community that accepts drug dealing, prostitution, or other types of deviance as normal can exhibit social "cohesion" in a fashion that promotes high levels of crime.

As the body of research on systemic social control continues to accumulate knowledge is gained about additional mediating processes and the reasons why they are important for the imposition of social control. For example, in addition to the factors noted above, variables such as family disruption, percent of single-parent families, sparse local friendship networks, mixed land use, population density, and the presence of unsupervised teenage peer groups all have been found to be indicators of decreased social control and, in turn, to lead to increased levels of crime and delinquency. Thus, the systemic model of social control informs us that the social structural conditions identified by Shaw and McKay as indicators of social disorganization do not directly cause crime but rather are related to high levels of crime and delinquency *because* they disrupt the more microlevel social process of human interaction that are necessary for a community to exert social control upon its residents.

Levels of Social Control

At the same time as the systemic model of social control was being developed and elaborated, complementary research was being conducted that elaborated upon the various levels of social control. Research by Albert Hunter (1985), and later Robert Bursik and Harold Grasmick (1993), detailed three distinct levels of control: private, parochial, and public. While these levels have been hinted at by the previous research, this section will more specifically identify the levels and their relationships.

The first level, private social control, is primarily informal in nature and is evoked through personal relationships. This level thus relates closely to the measures of social ties, networks, and cohesion noted above while adding in aspects of emotional attachments as well. For example, according to Travis Hirschi's original Social Bond Theory (1969), what stops an individual from committing delinquent acts are his/her bonds to society. These bonds take the form of personal attachments to people, commitment to conventional goals, involvement in conventional activities and the adoption of pro-social beliefs. Thus, simply knowing that a behavior will bring a negative response by a friend or family member or may negatively impact the ability of one to reach a goal exhibits enough "control" to stop the individual from engaging in the behavior.

The parochial level then incorporates the role of local social institutions in a community such as schools, churches, and community groups. This level of social control relates closely to the institutional controls noted in previous research and the importance of these institutions for the socialization of children. As children age such institutions take on increased importance in their lives and therefore become primary sources of socialization. The parochial level of social control may reinforce that exhibited at the private level, or in some cases, may protect the youth from further problems if the private level is not sufficiently developed. For example, a youth who is not highly attached to a parent in the home may form a strong bond with a coach or a teacher. That youth would then be less likely to break the rules than one who had no social control at either the private or parochial level. Thus, when the informal controls traditionally exhibited at the private level break down, the parochial or institutional level of social control may then serve as a second line of defense to prevent antisocial behavior.

The third level, public social control, was described by Albert Hunter (1985) as the ability of the community to access public goods and services to maintain order. This included but was not limited to police services. At this level the focus is either on the development of programs to create order and prevent crime or to catch those who are breaking the law and invoke enough formal control upon them to discourage them from continuing to break the law. At this level community financial resources and ties to greater society and social resources are of central importance.

It is easy to see the interdependent nature of these levels. If private social controls are strong, there is less need for parochial and public controls. On the other hand when private and parochial levels of control are weak, a greater use of public social control would be warranted (Black 1976). For example, an increased use of police force in an area is actually seen as a breakdown in private and parochial levels of social control. If all three levels are lacking, however, an area will be ridden with high rates of crime and violence. Research also suggests the potential for these levels to interact with one another. That is, for the effect of low social control at one level to have varying effects upon individuals based on other levels of risk or social control.

Moreover, it is important to note that this multilevel model of social control is not distinct from the systemic model but rather a further elaboration of the dynamics of social processes used to invoke social control. For example, levels of private social control in a neighborhood are determined by the ties between individuals and the strength of families. At the parochial level the strength of institutions and their ability to socialize youth is dependent on community member support both in terms of finances and participation. Research has further suggested that the social processes leading to the determination of private and parochial social controls are inherently tied to the larger structural characteristics originally identified by Shaw and McKay.

Advances in Theory, Testing, and Techniques

As highlighted by the previous discussion, social control is exhibited not only at distinct levels but through intricate interpersonal processes that are linked to the larger social institutions and structural characteristics of a community. Thus, the study of social control has become more complex in nature as have the theories that explain it.

One particular direction of theory development has been the integration of theories to address the various causes of crime and levels of social control. One example is Terence Thornberry's Interactional Theory that connects elements of social control and social learning theories as well as incorporating elements of the social structure as the interactional setting in which delinquent behavior is learned and even reinforced (Thornberry 1987). This theory also holds that different factors command primary influence at different stages of life. For example, while the family is key in early childhood through early adolescence, by mid-adolescence the school and other institutions outside of the family become more prominent with peer groups and the broad community taking dominance by late adolescence. Thus, according to this theory, the level of social control (private, parochial, public) bearing the strongest relationships to delinquency would change over the lifecourse.

John Laub and Robert Sampson (1993) also look at social control over different points in the life course and offer some of the first theories to extend this examination into adulthood noting what they call "adult social bonds." The institutions and social bonds that they incorporate as mechanisms of social control include marriage and employment as these offer new mechanisms of informal control upon the adult individual. For example, in adulthood the introduction of a spouse adds a new type of private social control while employment would expand connections to both parochial and public levels of social control.

Another fruitful theoretical avenue has been the risk-protective factor approach which seeks to identify factors at the individual, institutional, and neighborhood level that increase or decrease the possibility of deviant outcomes. Research in this area has tended to emphasize those factors that increase deviance such as previously noted correlates of delinquency (e.g., broken families, low attachment to schools). Analyses at the neighborhood level have also expanded to include such new things as number of gangs, availability of weapons, and ease of access to drugs.

Research utilizing this approach has been highly informative in gaining a better understanding of the interdependent and interactive nature of the multiple levels of social control. For example, one study asks "Do disadvantaged neighborhoods cause well-adjusted children to become adolescent delinquents?" (see Wikstrom and Loeber 2000). The answer here generally appears to be "No." However, some neighborhood enticements were found to lead some well-adjusted youth into delinquency particularly in later adolescence when influences outside of the home gain dominance. Other studies approach the question from the opposite direction asking whether "good" communities can help "not so well-adjusted" youth avoid negative outcomes (see Sampson et al. 2002; Kurlychek et al. 2011). In sum the basic consensus of these studies appears to be yes, that is, at-risk children fair better when raised in communities with high levels of parochial social control (e.g., strong social institutions and connections/cohesions among residents).

As theories of social control and crime become more complex, so do the data sets and analytical techniques needed to test them. For example, if social control spans three levels, then so must the data collected to explore it. The compilation of such data sets is time consuming and expensive thereby prohibiting a great deal of research in this area. However, several recent data collection efforts have evolved that are specifically designed to advance our understanding of the multilevel nature of social control. One such effort, the Project on Human Development in Chicago Neighborhoods (PHDCN), was specifically designed to examine how families, schools, and neighborhoods combine to affect child and adolescent outcomes. As an example of just how complicated such a data collection process can be it should be noted that this project has been ongoing for since 1994 and has received funds from a variety of entities including the National institute of Justice, the John D. and Catherine T. MacArthur Foundation, the US Department of Health, the US Department of Education, the National Institute of Mental Health the Harris Foundation, and the Turner Foundation.

In addition to the collection of data, the analyses of the data has been plagued by complexities such as the identification of compositional versus contextual effects and the spatial dependence of data as individuals are nested within schools and communities. Fortunately, recent developments in statistical techniques and packages have allowed researchers to effectively tackle many of these issues.

Perhaps the most significant advance in this area has been the use of hierarchical linear modeling (HLM) techniques to enable researchers to further disentangle the impact of individual-, institutional-, and neighborhood-level effects as well as exploring how factors at the various levels interact with one another. In addition this technique allows the researcher to identify what are known as "contextual" effects, which are effects above and beyond the mere aggregation of individual factors. For example, if being from a single-parent family increases the risk of delinquency by 5 % and one community has 10 % single-parent homes and another 50 %, the neighborhood with 50 % single-parent homes would naturally be expected to have more delinquency because more youth exhibit the 5 % greater risk of delinquency. However, a contextual effect would emerge if being in a community with more single-parent homes actually increased the risk of each youth, for example, instead of a 5 % greater risk of delinquency, it might be a 10 % greater risk of delinquency. Existent literature suggests that a pattern such as this might emerge as the percent of single-parent homes impacts other things such as the presence of adult role models, the ability of adults to supervise their children, and even the ability of parents to participate in community institutions/associations.

Examples of findings using this technique are abundant in the criminological literature. For example, research has found that the concentration of social disadvantage in neighborhoods has an independent contextual effect on adolescents' serious offending behavior even after controlling for multiple measures of individual-level risk. Studies have also found that adolescents residing in more economically disadvantaged neighborhoods were more likely to be involved in a range of serious violent behaviors after controlling for levels of individual-level risk. On the other hand, research has begun to show that youth with high levels of personal risk actually have lower chances of becoming delinquent if they reside in neighborhoods with strong social institutions and high levels of collective efficacy.

Future Directions

Although much knowledge has been gained about the role of social control, its types and levels in the causation of crime and delinquency, much is left to learn about the interactive and interdependent nature of social control. In particular, as noted above, research has just begun to show that living in an area with high levels of parochial social control can actually reduce the risk of a delinquent outcome for even high-risk youth. Research further exploring these connections holds the potential to inform crime prevention and intervention policy. For example, we might know that a youth is at high personal risk for delinquency because of family turmoil or other issues, but there may be ways to still reduce the overall probability of a criminal outcome for this youth by increasing other social controls such as attachments to school, involvement in community groups, and improving overall community levels of cohesion and collective efficacy. Or, working in the opposite direction, for youths living in high-risk communities, this research informs what might be done at the private/informal level to improve outcomes.

To answer questions such as these, there is also a need for more systematic social observation which better captures both the notion of a neighborhood and the real "feel" of living in that community. According to experts there are simply too many features of a neighborhood that are not measured readily by available public data sets (e.g., census) or cannot be systematically captured in surveys. To gain this type of information more observational and qualitative research is needed. Key researchers in this area stress the importance of collecting this information at the smallest unit possible (e.g., a block face – one side of a street within a block) (Sampson et al. 2002). This allows the researcher to then study these small geographic units, as well as providing the capacity to aggregate these units into larger geographical constructs (e.g., block groups, neighborhoods).

In conducting future research it is also important to remember that no community is an island. That is, each geographic area may be impacted by those areas surrounding it as well. As one example, when studying an adolescent it may be important to consider not only his/her home neighborhood but also the neighborhoods through which he/she travels to attend school or to which he/she travels to spend time with friends. It may also be important to understand concentrated disadvantage by not only looking at one community, but to also incorporate the resources of those communities that border it. One of the few studies in this area suggests that the consequences of being poor are worse for African-Americans in our society because not only do these individuals live in poor neighborhoods but they are much more likely than poor white individuals to live in communities embedded in areas of sustained poverty and disadvantage (Pattillo-McCoy 1999).

In summary, while we know much about social control and its importance in preventing crime, there is still much more to learn about the complex and dynamic processes through which it impacts individuals and the communities in which they live. Future research in this area holds the promise to help build stronger families, safer communities, and provide better life outcomes for even the most at-risk youth.

Recommended Reading and References

Black D (1976) The behavior of law. Emerald Group Publishing, Bingley

Bursik RJ, Grasmick HG (1993) Neighborhoods and crime: the dimensions of effective community control. Lexington Books, New York

Coleman JS (1988) Social capital in the creation of human capital. Am J Sociol 94(Suppl):S95–S120

Hirschi T (1969) Causes of delinquency. University of California Press, Berkeley

Hunter AJ (1985) Private, parochial and public school orders: the problem of crime and incivility in urban communities. In: Suttles GD, Zald MN (eds) The challenge of social control: citizenship and institution building in modern society. Albex, Norwood

Kasarda JD, Janowitz M (1974) Community attachment in mass society. Am Sociol Rev 39:328–339

Kornhauser R (1978) Social sources of delinquency. University of Chicago Press, Chicago

Kurlychek M, Krohn MD, Beidi D, Penly-Hall G, Lizotte AT (2011) Protection from risk: an exploration of when and how neighborhood-level factors can reduce violent youth outcomes. Youth Violence Juv Justice 10(1):83–106

Morenoff JD, Sampson RJ, Raudenbush SW (2001) Neighborhood inequality, collective efficacy and the spatial dynamics of urban violence. Criminology 39(3):517–559

Park RE, Burgess EW, McKenzie RD (1925) The growth of the city. University of Chicago Press, Chicago

Pattillo-McCoy M (1999) Black picket fences: privilege and peril among the black middle class. University of Chicago Press, Chicago

Rankin BH, Quane JM (2002) Social contexts and urban adolescent outcomes: the interrelated effects of neighborhoods, families and peers on African-American youth. Soc Probl 49(1):79–100

Sampson RJ, Byron Groves W (1989) Community structure and crime: testing social-disorganization theory. Am J Sociol 94(4):774–802

Sampson RJ, Laub JH (1993) Crime in the making: pathways and turning points through life. Harvard University Press, Cambridge

Sampson RJ, Morenoff JD, Earls F (1999) Beyond social capital: spatial dynamics of collective efficacy for children. Am Sociol Rev 64:633–660

Sampson RJ, Morenoff JD, Gannon-Rowley T (2002) Assessing "neighborhood effects": social processes and new directions in research. Annu Rev Sociol 28:443–478

Shaw CR, McKay HD (1942) Juvenile delinquency and urban areas. University of Chicago Press, Chicago

Taylor RB (1997) Social order and disorder of street blocks and neighborhoods: ecology, microecology, and the systemic model of social disorganization. J Res Crime Delinq 34(1):113–155

Thomas WI, Znaniecki F (1920) The Polish peasant in Europe and America. Gorham, Boston

Thornberry TP (1987) Toward and interactional theory of delinquency. Criminology 25(4):863–891
Thrasher F (1927) The gang. University of Chicago Press, Chicago
Wikstrom P-OH, Rolf L (2000) Do disadvantaged neighborhoods cause well-adjusted children to become adolescent delinquent? A study of male juvenile serious offending, individual risk and protective factors and neighborhood context. Criminology 38(4):1109–1142

Social Control and Self-control Through the Life Course

Michael L. Benson
School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA

Overview

The concept of social control has a complex and controversial history in the field of sociology. Originally, it was defined as the ability of a group to regulate itself, but the term was subsequently redefined to refer either to socialization or social repression (Janowitz 1975). Within criminology, social control typically is used in the more classical sense to refer to the mechanisms through which groups attempt to orchestrate behavior and control deviance. Self-control, on the other hand, is a relatively new concept in criminology that refers to the differential ability (or inability) of individuals to refrain from taking advantage of opportunities to satisfy their immediate desires by engaging in criminal or deviant behavior (Gottfredson and Hirschi 1990). Both social control and self-control have been proposed as important factors that influence individual involvement in crime throughout the life course (Gottfredson and Hirschi 1990; Sampson and Laub 1993). Empirical support for their importance in the etiology, longevity, and patterning of criminal behavior is substantial, but a number of issues remain to be resolved, including the origins and stability of self-control, the effects of social control on criminal trajectories at different stages of the life course, and the factors that influence a person's exposure to informal social control. This entry addresses these issues by focusing on how social control and self-control may change over the life course.

Crime and Informal Social Control

Within criminology, informal social controls are seen as emerging from "the role reciprocities and structure of interpersonal bonds linking members of a society to one another and to wider social institutions such as work, family, and school" (Sampson and Laub 1993, p. 18). For example, husbands and wives are bonded to one another through socially recognized obligations and expectations regarding proper normative behavior. These interpersonal bonds act as constraints on the participants in a role relationship, in the sense that there is pressure on participants to meet their role expectations if they wish for the relationship to continue and if they wish to be viewed favorably by others. Children may be bonded to their parents by emotional connections of love and caring, and children understand that their parents have expectations regarding how they should behave. To the extent that children care about their parents, these expectations can act as a constraining force on their behavior. Similarly, being employed requires that one meet the expectations of one's employer to show up at work, do one's job, and not bring disrepute on the work organization. Thus, informal social controls that arise out of social roles guide behavior by placing responsibilities and constraints on the individual and by directing individual action in some directions rather than others. Prosocial behavior is promoted by strong social controls, while deviancy and delinquency are promoted by the attenuation of social controls over individual conduct (Thornberry 1987).

At different stages in the life course, individuals are potentially subject to different forms of informal social control that arise of out different role relationships and participation in different social institutions (Sampson and Laub 1993; Thornberry 1987). For children, informal family and school bonds are important. Children who are strongly bonded to their parents and who care about school are less likely to be involved in delinquency than children who have difficult relations with their parents or who do not like school. As children move through the life course, the major sources of informal social control change. Parents and school are not as important for young adults as they are for children and teenagers. For young adults, employment and marriage are potential sources of informal control. Some scholars argue that variation in the strength of informal controls influences the likelihood and degree of involvement in crime and deviance at all stages of the life course (Sampson and Laub 1993; Thornberry 1987).

For young children, families and specifically parents are the most important source of informal social control. The fundamental processes that seem to be present in successful families and missing in unsuccessful ones are attachment and control. Attachment refers to an emotional connection between parent and child, in which the child has feelings of love, respect, and admiration toward the parent and the parent feels similarly toward the child. It is a reciprocal process, involving both parent and child, but it starts with the parent (Hirschi 1969). By giving love and expressive support early on, parents can foster a sense of attachment in their children. Children who develop strong attachments to their parents care about their parents' feelings and opinions. They are aware of and sensitive to the impact that their behavior can have on their parents. They understand that if they are caught doing something wrong, it will embarrass and disappoint their parents, and they do not want that to happen. Rather, they want their parents to be proud of them. This emotional connection functions as a sort of internal monitor of the child's behavior when parents are not present. Parents who work at developing a strong sense of attachment in their children when they are young are exercising informal social control and are likely to be rewarded for their efforts when their children enter adolescence in reduced levels of delinquency.

Showering children with love and affection when they are young in the hopes of developing a strong sense of attachment, however, is not enough in and of itself to prevent involvement in delinquency or deviance. Expressive support is crucial, but parents who want to keep their children out of trouble must also exercise more direct types of informal social control. Direct control involves monitoring the child's behavior, recognizing deviance when it occurs, and correcting misbehavior when it happens (Gottfredson and Hirschi 1990). Parents must pay attention to their children and be aware of what they are doing and recognize when they are doing something that is wrong or inappropriate. By monitoring their children, recognizing deviance, and correcting misbehavior, parents can foster the development of conformity in their children and help start them off on trajectories aimed away from serious delinquency (Gottfredson and Hirschi 1990; Patterson et al. 1992).

As children grow older and move into adolescence, the sources of informal social control change from families and parents to schools and peers. Children who are strongly attached to school are significantly less involved in deviance than those with weaker attachments. School attachment is a multidimensional concept that has both objective and subjective elements. Objective attachment can be seen in school performance (that it, how well one does in the classroom) and involvement in school-related activities, such as committees, sports teams, or school clubs. Subjective attachment refers to one's aspirations and expectations in regard to educational achievement and one's sense of satisfaction with and affection toward the people and activities that constitute the school system (Sampson and Laub 1993).

That people often behave differently when they are in a group as opposed to being by themselves is so well known that it is something of a sociological and psychological truism. This truism appears to apply with particular force to juveniles, as most delinquency is committed in groups (Warr 2002). Thus, peers have long been recognized as an important influence on juvenile delinquency and as a source of informal social control in adolescence in the form of peer pressure. Juveniles who have delinquent friends tend to be delinquent themselves, while their nondelinquent counterparts tend to have nondelinquent friends. Although the empirical correlation between delinquency and having delinquent friends is well established, its interpretation is a matter of dispute. One school of thought holds that delinquent friends cause or exacerbate an individual's own delinquency through modeling and social reinforcement for delinquent behavior (Akers 1998). In contrast, conforming friends model and reinforce prosocial behavior. The other school of thought holds that the correlation between delinquency and delinquent friends is spurious and results from a self-selection process in which teenagers with delinquent propensities tend to seek out other teenagers with similar propensities (Gottfredson and Hirschi 1990).

As teenagers move into young adulthood, they may become subject to new forms of informal social control. These new forms of control include employment and marriage. Individuals who are lucky enough to find good jobs or enter good marriages or both become subjugated to new sources of informal social control. One of the most important sources is marriage. Marriage entails new responsibilities and ideally a strong emotional commitment to another person (Laub et al. 1998). To the extent that a young man is emotionally attached to his wife and his children, he is likely to curb his criminal inclinations out of respect for his wife's wishes and a desire not to jeopardize his family life. Marriage may influence involvement in crime in another way by limiting a person's access to deviant peers and to opportunities to engage in deviant activities. The time that a man spends with his spouse or children at home is time that he cannot spend with criminal peers on the street. Time spent working to provide a living for one's spouse and children is also time taken away from the risky attractions of street life. Getting married reduces one's contact with friends in general and delinquent friends in particular. Reduced exposure to delinquent friends leads to reduced involvement in crime and deviance (Warr 1998).

Qualitative and quantitative evidence suggests that, for some men, becoming involved with a woman and getting married can be a route out of crime (Shover 1985). What appears to matter most is the development of high-quality marital bonds (Sampson and Laub 1993; Sampson et al. 2006). However, the effect of a good marriage on criminal offending does not occur all at once. Rather, it is a gradual process that cumulates over time. The longer time an offender invests in a good marriage, the more likely that marriage is to have a preventive effect on his criminal trajectory (Laub et al. 1998). Offenders who enter into good marriages gradually become committed to their spouses. This commitment appears to function as a source of informal social control that grows slowly over time and leads toward desistance.

Crime and Self-control

According to Gottfredson and Hirschi (1990), at the individual level, the most important causal factor in crime is a behavioral propensity called low self-control. Low self-control refers to the relative ability (or inability) of an individual to avoid taking advantage of an opportunity to satisfy his or her immediate wants and desires by engaging in some form of criminal or deviant behavior. Other monikers for this concept include self-regulation, effortful control, self-discipline, and will power. People with low self-control are thought to be impulsive, self-centered, not concerned with long-term consequences, physically active, and risk-loving. Self-control is conceived to be an individual trait that varies over individuals but remains stable for any given person over time (Gottfredson and Hirschi 1990). Thus, people who have low self-control relative to others early in life will also have low self-control relative to others later in life, while conversely people with high self-control at one point in time will tend to have high self-control at other times. People with low self-control can be distinguished from people with high self-control by their tendency to pursue "short-term gratification with little consideration for the long term consequences of their acts" and with little sympathy for the rights or feelings of others (Hirschi and Gottfredson 1987a, p. 959). Low self-control is not always manifested in criminal behavior. Depending on the situation, it may be expressed in other deviant, risky, or disreputable ways, such as reckless driving, alcohol and drug abuse, promiscuous sex, and job quitting (Hirschi and Gottfredson 1987a, b). In short, people with low self-control are impulsive, insensitive to others, and almost always interested in pursuing their own personal pleasures in the quickest way possible. They are attracted to crime, deviance, and analogous acts because they tend to provide quick rewards and easy gratification.

Where does low self-control come from? According to Gottfredson and Hirschi (1990), the answer to this question can be found in the homes and families of delinquents. The parents of delinquents are responsible for the development of low self-control. In their theory, self-control is something that parents have to instill in their children early on by engaging in specific parenting practices. If parents fail to engage in these practices, then their children will not develop a strong sense of self-control, or to put it the other way round, then the children will inevitably have low self-control. In other words, self-control is not something that people are born with or that develops naturally as they grow up. Rather, it is something that must be instilled in children by their parents. To instill self-control, parents must do four things in regard to their children. They must (1) care for their children, (2) monitor their behavior, (3) recognize their wrongdoing, and (4) discipline or correct their wrongdoing when they observe it. If parents fail to employ this suite of practices, then their children will naturally develop with low self-control, and they will do so very early in life, typically before ages 11 or 12. After that, a person's relative level of self-control is set and cannot be developed or augmented later in life.

Current Issues and Controversies

Informal Social Control and Desistance

The causal role of informal social controls in the desistance process is disputed. Some scholars who take a life course perspective believe that change is an ever present possibility and that even serious long-term offenders can and do desist at later stages in the life course (Maruna 2001; Maruna et al. 2009; Vesey et al. 2009). From this perspective, desistance is seen as a process that involves several mutually interacting factors, including turning points, informal controls, and the structuring effects of routine activities. In theory, the path to desistance begins when an individual experiences some sort of turning point, such as marriage, military service, moving to a new location, or finding a fulfilling job. Regardless what it is, this event sets in motion a series of changes in the offender's life that may over time lead to "desistance by default." Desistance by default expresses the idea that offenders do not necessarily make explicit and conscious decisions to change their lives and "go straight" as a result of some sort of epiphany or a single transformative event. Rather, their behavior and sense of identity change gradually and perhaps largely unconsciously as a result of experiencing a turning point that exposes them to new sources of informal social control and that also changes their routine activities.

Getting married, joining the military, or finding a job can serve to knife the offender off from his previous life style and expose him or her to greater informal social controls. Turning points also impose the structuring effects of conventional routine activities on offenders. Military service, marriage, and employment, all require time commitments. They require that one be at certain places at certain times. They impose duties to fulfill and obligations to keep. In short, individuals become caught up in a series of routine activities in which they spend time with conventional others doing conventional ordinary things. The time available to hang out with criminal peers and engage in deviant activities is gradually reduced until it reaches the point where an offender has for all intents and purposes desisted.

Another school of thought, however, holds that desistance is not caused by turning points and increased exposure to informal social controls. This school of thought, called the developmental perspective, holds that there are distinct developmental trajectories that people get locked into. Developmental trajectories are stable in the sense that once people enter a particular trajectory or developmental group, their future progress is more or less determined. For example, Gottfredson and Hirschi (1990) argue that criminal behavior results from an underlying behavioral propensity. At an early point in the life course, each individual's propensity toward crime and deviance is established when their level of self-control is developed. From that point on, one's level of self-control drives behavior independent of other factors and is largely resistant to change. Thus, the presence or absence of desistance in adolescence reflects only variation in this underlying propensity toward crime. Another developmental theorist, Moffitt (1993), argues that there are two distinct trajectories – life-course-persistent and adolescent limited. Individuals on the life-course-persistent trajectory are unlikely to experience turning points such as marriage, or to react badly if they do experience them, because of certain innate characteristics that are established early in life. These individuals never really desist from deviant or criminal behavior, even though the form in which their deviance is manifested may change as they age. For example, the teenage mugger becomes the adult wife-beater. Adolescent limited offenders engage in crime and delinquency for a few years during their teenage years and then desist largely for maturational reasons, not because they are exposed to new forms of informal social controls.

Developmental theorists do not deny that there is an association between exposure to informal social controls and reduced involvement in crime and deviance, but they believe that this association is spurious and not causal. Reduced involvement in crime and deviance may be associated with participation in social institutions such as marriage and work, but it is not caused by them. Rather, participation in these institutions is a matter of self-selection. It is, of course, possible and some research suggests that both social causation and self-selection are involved in the desistance process (Wright et al. 1999).

Origins of Self-control

Even though low self-control is an individual trait that would appear to bear strong similarities to other personality traits, such as impulsiveness, aggressiveness, and insensitivity, the originators of the term explicitly rejected the notion that it is biologically or genetically based (Gottfredson and Hirschi 1990). Rather, in their view, low self-control is caused solely by inadequate parenting early in a child's life. Recent advances in behavioral genetics suggest that this conclusion is either not correct or greatly oversimplified. Insofar as genetic researchers have been able to determine, it appears that all individual traits are influenced to some degree by heredity, that is, by the genes that parents pass on to their children (Rutter 2007). Height, weight, body type, eye color, hair color, and facial appearance are some of the more obvious physical traits for which it often is easy to see physical resemblances between parents and their offspring. However, in addition to these gross anatomical characteristics, children also inherit psychological characteristics and behavioral tendencies from their parents (Plomin et al. 1990). For example, aggressive parents are more likely to have aggressive children than nonaggressive parents (Huesmann et al. 1984). A growing body of evidence now suggests that genetic factors also are implicated in the development of low self-control (Beaver 2011; Wright et al. 2008a; Wright and Beaver 2005).

Genetic factors may influence the development of low self-control in several different ways. First, low self-control appears to be a heritable trait or characteristic that parents pass on to their children. Thus, parental genotype directly influences offspring genotype, including the genes related to low self-control. Second, parents with low self-control tend to create home environments that are not conducive to the development of high self-control (Moffitt 1997). Indeed, their parental management techniques may be so inadequate that they actually foster the development of low self-control (Beaver 2011). In this case, the genetic effect is indirect in that the parent's genotype is implicated in the creation of a particular home environment that then stimulates the development of low self-control in the child. Finally, children born with genetically influenced personality characteristics that make them difficult to deal with, such as lack of impulse control, aggressiveness, and resistance to parental authority, may provoke negative parenting practices that further exacerbate behavioral tendencies (O'Connor et al. 1998). This type of effect is an evocative gene environment interaction, meaning that the child's genotype in a sense creates an environment that fosters the expression of behavioral tendencies inherent in the genotype.

Stability of Self-Control

One of the most provocative and controversial aspects of low self-control theory concerns the stability of this characteristic over the life course. As noted above, the originators of the term contended that relative differences between people in self-control remain stable as people age. Note, this contention does not mean that people with low self-control never change the way they behave as they age. Rather the idea of relative stability means that even though persons with low self-control may exhibit more moderate behavior as older adults than they did as teenagers, they will still be more antisocial, self-centered, and short-sighted than their contemporaries who have adequate levels of self-control. Indeed, research does indicate that absolute levels of self-control increase with age for most people.

At this point in time, however, research on the relative stability hypothesis has produced mixed results. While it is clear that there is substantial continuity in antisocial behavior over the life course, it is also clear that the level of continuity is by no means perfect. Evidence for stability comes from a number of longitudinal studies. For example, in a study of 205 boys aged 10–16, researchers asked the mothers to recall how easy or difficult it had been to get along with their sons when they were 1–5 years old. Five years later, when the boys were 15–21 years old, those who had been characterized by their mothers as "difficult" were twice as likely to have an official record of delinquency as those rated "easy." The difficult boys also self-reported committing delinquent acts at a higher rate than the easy boys (Loeber et al. 1991). Similarly, kindergarten children who are rated by their teachers as having high levels of hyperactivity are more likely to engage in delinquency during the transition from childhood to adolescence (ages 10–13) than children who do not exhibit hyperactivity (Tremblay et al. 2003). Finally, in a study using the Dunedin, New Zealand, data, White et al. (1990) found considerable continuity in antisocial and delinquent behavior from age 3 through the early teen years. Children who scored high on measures of disobedient and aggressive behavior at age 3 were more likely to exhibit other conduct disorders later in childhood and to be arrested by the police in their early teen years than children who scored low in disobedience and aggression (White et al. 1990). The fact that some children begin to show signs of abnormal conduct so early in life suggests that some forms of antisocial behavior reflect a general temperament that may persist over time (White et al. 1990). This study provides impressive evidence for continuity in antisocial behavior because it is a prospective study of a normal population (Wright et al. 2008b).

Thus, evidence for continuity in antisocial behavior is plentiful, but nevertheless, it must be interpreted carefully (Thornberry and Krohn 2003). Conduct problems at a young age do predict delinquency in the teenage years, but the success rate of these predictions is sometimes not very good. Most antisocial preschoolers do not go on to become antisocial adults or even antisocial juveniles. In the Dunedin study, for example, there was a high false-positive rate when antisocial behavior at age 3 alone was used as a predictor of antisocial behavior at age 11. Indeed, the predictions were wrong almost 9 times out of 10. The researchers note: "Of the 209 children predicted to have antisocial outcomes at age 11, 84.7 % did not develop stable and pervasive antisocial behavior" (White et al. 1990, p. 521). Other studies have had better results in the sense that the prediction rates are better. For example, Campbell and colleagues studied children from age 3 to age 13 with assessments at ages 4, 6, and 9 in between. Of the children who were rated as hard to manage by their parents at age 3, fully 48 % met formal criteria for having externalizing disorders (attention deficit, conduct disorder, or oppositional disorder) by age 9. In contrast, only 16 % of the children in a control group displayed externalizing disorders at age 9 (Campbell 1995). In statistical terms, the difference between 48 % and 16 % is large and represents a strong effect. However, in practical terms, one could say that the study actually shows that discontinuity is more likely than continuity as less than half of the problem group exhibited problem behaviors at both times. At this point, the most accurate summary of the research would be that continuity in behavior is indeed a widespread feature of human development, but it is not a universal feature.

Although research on continuity in antisocial behavior suggests that traits such as low self-control may be stable over the life course, they do not directly assess that stability. To date, there are few studies that directly investigate stability in self-control itself over time. The few studies that are available, however, suggest that like antisocial behavior, relative levels of self-control can vary over the life course. That is, instability in individual rankings on self-control over time has been observed, suggesting that self-control may not be an immutable trait. The degree to which self-control is malleable, however, is still open to debate with some researchers finding more stability than others (Burt et al. 2006; Hay and Forrest 2006).

Future Directions

A large body of research now clearly indicates that both social control and self-control are important factors in the etiology of crime and behavior in general across the life course. Furthermore, it is becoming increasingly apparent that within the life course, self-control and social control have interactive effects on each other. For example, an individual's level of self-control may influence how he or she reacts when confronted with a turning point, such as marriage, which would in turn affect the individual's relative exposure to informal social controls. Likewise, it is possible that exposure to increased informal social controls may eventually lead to transformations in the individual that increase his or her level of self-control (Giordano et al. 2002; Laub and Sampson 2003; Sampson et al. 2006). Future research on self-control and selection effects is needed, as well as, research on the transformative power of informal social controls in regards to developing increased self-control in adulthood.

Research on the stability of self-control over the life course has largely ignored whether self-control is malleable in response to direct officially imposed rehabilitative or treatment interventions. This shortcoming is unfortunate, because, as Currie has noted, research showing continuity in antisocial behavior over time reveals only what happens when no effort is made to change how individuals develop (Currie 1998). Yet, a voluminous body of research suggests that a multitude of rehabilitative programs work in terms of reducing recidivism among offenders at all stages of the life course (MacKenzie 2006). If delinquent and criminal behaviors are an expression of low self-control, then it follows by implication that reductions in offending caused by treatment interventions may also be associated with changes, indeed increases, in self-control. Future research is needed that directly assesses whether and how measured levels of self-control can be changed in response to different forms of intervention and treatment.

Related Entries

- ► Control Theory
- ► Genes, Crime, and Antisocial Behaviors
- ► Genetic Basis to Self-Control
- ► Informal Social Control
- ► Pathways to Delinquency
- ► Risk factors for Adolescent Sexual Offending

Recommended Reading and References

Akers R (1998) Social learning and social structure: a general theory of crime and deviance. Northeastern University Press, Boston

Beaver KM (2011) The effects of genetics, the environment, and low self-control on perceived maternal and paternal socialization: results from a longitudinal sample of twins. J Quant Criminol 27:85–105

Burt CH, Simons RL, Simons LG (2006) A longitudinal test of the effects of parenting and the stability of self-control: negative evidence for the general theory of crime. Criminology 44:353–396

Campbell SB (1995) Behavior problems in preschool children: a review of recent research. J Child Psychol Psychiatry 36:113–149

Currie E (1998) Crime and punishment in America. Metropolitan Books, New York

Giordano PC, Cernkovich SA, Rudolph JL (2002) Gender, crime, and desistance: toward a theory of cognitive transformation. Am J Sociol 107:990–1064

Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Hay C, Forrest W (2006) The development of self-control: examining self-control theory's stability thesis. Criminology 44:739–774

Hirschi T (1969) Causes of delinquency. Univesity of California Press, Berkeley

Hirschi T, Gottfredson M (1987a) Causes of white-collar crime. Criminology 25(4):949–974

Hirschi T, Gottfredson M (1987b) Toward a general theory of crime. In: Buikhuisen W, Mednick S (eds), Crime & capitalism. Brill, Leiden, pp 8–26

Huesmann LR, Eron LD, Lefkowitz MM, Walder LO (1984) Stability of aggression over time and generations. Dev Psychol 20:1120–1134

Janowitz M (1975) Sociological theory and social control. Am J Sociol 81:82–108

Laub JH, Nagin DS, Sampson RJ (1998) Trajectories of change in criminal offending: Good marriages and the desistance process. Amer Sociol Rev 63(2):225–238

Laub JH, Sampson RJ (2003) Shared beginnings, divergent lives: delinquent boys to age 70. Harvard University Press, Cambridge

Loeber R, Stouthamer-Loeber M, Green SM (1991) Age at onset of problem behavior in boys and later disruptive and delinquent behaviors. Crim Behav Ment Health 1:229–246

Maruna S (2001) Making good: how ex-convicts reform and rebuild their lives. American Psychological Association, Washington, DC

Maruna S, LeBel TP, Naples M, Mitchell N (2009) Looking-glass identity transformation: pygmalion and golem in the rehabilitation process. In: Veysey BM, Christian J, Martinez DJ (eds) How offenders transform their lives. Willan Publishing, Portland, pp 30–56

MacKenzie DL (2006) What works in corrections: reducing the criminal activities of offenders and delinquents. Cambridge University Press, New York

Moffitt TE (1993) Adolesence-limited and life-course persistent antisocial behavior: a developmental taxonomy. Psychol Rev 100:674–701

Moffitt TE (1997) Adolescence-limited and life-course-persistent offending: a complementary pair of developmental theories. In: Thornberry TP (ed) Developmental theories of crime and delinquency. Transaction Publishers, New Brunswick, pp 11–54

O'Connor T, Deater-Deckard K, Fulker D, Rutter M, Plomin R (1998) Genotype-environment correlations in late childhood and early adolescence: antisocial behavioral problems and coercive parenting. Dev Psychol 34:970–981

Patterson GR, Reid JB, Dishion TJ (1992) Antisocial boys. Castalia Publishing Co., Eugene

Plomin R, Chipuer HM, Loehlin JC (1990) Behavioral genetics and personality. In: Pervin LA (eds) Handbook of personality theory and research. Guilford Press, New York, pp 225–243

Rutter M (2007) Gene–environment interdependence. Dev Sci 10:12–18

Sampson RJ, Laub JH (1993) Crime in the making: pathways and turning points through life. Harvard University Press, Cambridge

Sampson, RJ, Laub JH, Wimer C (2006) Does marriage reduce crime? a counterfactual approach to within-individual causal effects*. Criminology 44(3):465–508

Shover N (1985) Aging criminals. Sage, Beverly Hills

Thornberry TP (1987) Toward an interactional theory of delinquency. Criminology 25:863–892

Thornberry TP, Krohn MD (2003) Taking stock of delinquency: an overview of findings from contemporary longitudinal studies. Kluwer Academic/Plenum Publishers, New York

Tremblay RE, Frank V, Daniel N, Linda P, Seguin JR (2003) The Montreal longitudinal and experimental study: rediscovering the power of descriptions. In: Thornberry TP, Krohn MD (eds) Taking stock of delinquency: an overview of findings from contemporary longitudinal studies. Kluwer Academic/Plenum Publishers, New York, pp 205–254

Vesey BM, Christian J, Martinez DJ (2009) In: Veysey BM, Christian J, Martinez DJ (eds) How offenders transform their lives. Willan Pub, Cullompton; Portland

Warr M (1998) Life-course transitions and desistance from crime. Criminology 36(2):183–216

Warr M (2002) Companions in crime: the social aspects of criminal conduct. Cambridge University Press, Cambridge, UK/New York

White J, Moffitt TE, Earls F, Robins LN, Silva PA (1990) How early can we tell? preschool predictors of boys' conduct disorders and delinquency. Criminology 28:507–533

Wright JP, Beaver KM (2005) Do parents matter in creating self-control in their children? A genetically informed test of Gottfredson's and Hirschi's theory of low self-control. Criminology 43:1169–1202

Wright BRE, Caspi A, Moffitt TE, Silva PA (1999) Low self-control, social bonds, and crime: social causation, social selection, or both? Criminology 37:479–514

Wright J, Beaver K, Delisi M, Vaughn M (2008a) Evidence of negligible parenting influences on self-control, delinquent peers, and delinquency in a sample of twins. Justice Quarterly 25:544–569

Wright JP, Tibbetts SG, Daigle LE (2008b) Criminals in the making: criminality across the life course. Sage, Los Angeles

Social Control Theory of Sexual Homicide Offending

Matt DeLisi[1] and John Paul Wright[2,3]
[1]Criminal Justice Studies, Iowa State University, Ames, IA, USA
[2]School of Criminal Justice, University of Cincinnati, Cincinnati, OH, USA
[3]King Abdulaziz University, Jeddah, Saudi Arabia

Introduction

For decades, the study of offenders who commit sexually based offenses generally and who commit predatory sexual homicide specifically has come almost entirely from psychiatry, forensics, and applied criminal justice. Mainstream theoretical criminology has mostly abstained from studying predatory sexual homicide offenders for a variety of reasons that will be explored. A negative consequence of this omission is that criminological theories are often poorly suited to explain the most pathological forms of criminal violence and, by extension, poorly suited to realistically explain the most pathological offenders. This is a major limitation.

This current entry seeks to redress this by exploring the utility of self-control theory (Gottfredson and Hirschi 1990) as a conceptual vehicle to understand sexual homicide offending. Although these offenders evince lifestyles that are consistent with the theory, their instrumental, methodical approach to sexual homicide is also in many respects directly contrary to the tenets of the theory. In addition, the current authors suggest that forensic typologies that seek to characterize sexual homicide offenders as organized or disorganized (Ressler et al. 1988) can be fruitfully understood within a self-control theoretical framework.

Criminology and Sexual Homicide

Two overlapping reasons explain the presumed incompatibility of studying sexual homicide offenders within the parameters of mainstream criminological theory. The first relates to the infrequency of sexual homicide and the likelihood that offenders in traditional criminological datasets contain such offenders. If murder has a low base prevalence, then the base rate of an even more pathological crime is of course lower still. For instance, the prevalence of serial murder has been estimated to constitute less than 1 % of the total murders in society (McNamara and Morton 2004). This means that community samples and even epidemiological samples simply will not contain predatorily violent offenders such as sexual murderers. Prior studies have found that some of the most frequently used datasets in criminology have approximately zero homicide offenders in them (DeLisi 2001) and instead are comprised of less severe/violent offenders. Due in part to data accessibility issues, mainstream criminology simply declined to understand the most severe offenders in favor of explanations that are compatible with more common, normative offenders.

A second reason that criminological theory has ignored sexual homicide offenders is plain condescension toward a topic that some academics view as too sensationalistic to be taken seriously. Although this might seem like a harsh accusation, the literature is littered with examples of it. For instance, Edwin Sutherland (1950), the patriarch of American academic criminology, argued that sexual psychopath laws – which were designed to socially control instrumentally violent criminals such as sexual murderers – were based on false, questionable knowledge based in popular literature as opposed to science. The tenor of Sutherland’s differential association theory is to repudiate individual-level pathologies of the offender in favor of social learning processes. Thus, sexual homicide offenders and the nosology surrounding them (e.g., psychopathy, sexual sadism) were portrayed as antithetical to more appropriate criminological theorizing.

More recently, Ray Surette advanced that serial murderers are media icons, the purpose of which is to “generate fear, degrade social networks, increase reliance on the media, and foster social isolation and polarization” (1994, p. 147).

In a review of a book that was at best tangential to the topic of sexual homicide, Shadd Maruna (2006) opined:

> My only substantial complaint though is the embarrassing decision to include the dated, non-academic, non-theoretical chapter by Ressler, Burgess, and Douglas on serial killers in the collection. Talk about letting down the side. With one, cheap attempt to grab the attention of the 'I signed up for criminology because I like 'Silence of the Lambs' market of undergraduate readers. Cromwell almost completely undermines all of the remarkable achievements of the real criminologists pushing the boundaries of criminological research. (2006, p. 275)

Indeed, the indices of major criminological theoretical works and anthologies of criminology theory contain little to no references to sexual homicide offending suggesting that the topic is largely unrelated to mainstream criminological thought.

Fortunately, not all criminologists have been so dismissive of the most extreme and violent criminals (see, e.g., Vaughn et al. 2009). And not all criminologists have ignored potential linkages between criminological theory and homicide offending. For example, Castle and Hensley (2002) surmised that social learning theory could be a useful theoretical framework toward understanding serial murder by exploring experiences from military service and exposure to killing as potential precursors to multiple homicide offending. Social learning theory was also used to understand the prospective linkages between juvenile fire setting and subsequent serial murder (Singer and Hensley 2004). Within the literature, there is also increasing acknowledgement that criminology has for too long overlooked the study of the most violent offenders. To illustrate, DeLisi (2006) noted, "For some time, unfortunately, mainstream criminology tended to focus on more normative offenders and less extreme forms of antisocial behavior. As a result, the study of murder and other extreme forms of violent crime were important but demonstrably peripheral to the discipline. Since the general public is fascinated with more extreme forms of criminal violence, academic criminology clearly missed an opportunity to connect with public audiences, including practitioners and policy makers" (pp. 172–173). Unlike their criminologist peers, forensic psychologists, neuroscientists, historians, and journalists have focused considerable scholarly energy toward understanding the etiology of sexual homicide, and it is from these disciplines where the sexual homicide epistemology largely resides.

The current entry has an explicit aim: to explore Gottfredson and Hirschi's (1990) self-control theory – a dominant content area in criminological theory and research – and its relation to sexual homicide offending. Sexual homicide offending is defined as the predatory perpetration of murder in conjunction with some sexually oriented offense, such as rape, sodomy, or other type of sexual assault. In some ways, self-control theory provides a useful framework for understanding the lives of predatory sexual murderers particularly those who leave disorganized crime scenes. However, in other ways, the offending behavior of these offenders is characterized by extremely high level of self-control and self-regulation particularly those who leave organized crime scenes. Thus, a self-control paradox exists when studying the apex of violent criminals.

Self-Control and Violence

To date, no study has formally examined self-control theory vis-à-vis sexual homicide although a range of studies have explored the association between low self-control – characterized by Gottfredson and Hirschi as being self-centered, impulsive, poorly tempered, and action oriented as opposed to cognitively oriented, preferring simple tasks as opposed to ones requiring tenacity, and having poor gratification delay – and violence. DeLisi and Vaughn (2008) reported that self-control was a strong indicator of career criminality, a status that is disproportionately responsible for the most serious crimes in a society. Using data from a statewide sample of institutionalized delinquents, they found that those scoring 1 SD above the mean on a low self-control scale were over five times more likely to become a career criminal and disproportionately likely to commit violent crimes. The association between low self-control and violence is not limited to offender samples. Drawing on data from students, Bouffard (2010) found that self-control was inversely related to both sexual entitlement and self-reported perpetration of sexual aggression. Self-control is a potentially important moderator of criminal careers over time because persons with very low self-control are theorized to continue offending often in violent ways.

Given the nature of most violent disputes, it is unsurprising that the self-control construct would be associated with reactive violence, which frequently typifies intimate-partner violence and generalized assaults (e.g., bar fights). By definition, having a quick, volatile temper means that interpersonal conflicts are not handled internally and ignored; they are handled externally and instantaneously settled. Inherent in this statement, moreover, is the salience of impulsivity and low gratification delay to understanding violence. Consistent with Gottfredson and Hirschi's (1990) theory, violence is an easy option to quickly end the aggravating stimuli created by another person. According to the theory, those with low self-control do not handle frustration and the day-to-day adversities imposed by adult responsibilities and thus will use a quick, easy method to deal with that frustration.

To use a clearly unscientific yet accurate descriptor, it could be said that offenders characterized by self-control deficits are generally *sloppy*. They are sloppy in their school habits and attention to detail which contributes to adverse encounters with teachers and other students and generally negative academic performance outcomes (Houts et al. 2010). They are sloppy in their attention to the emotional and time commitments of interpersonal relationships, and those relationships suffer. Almost an extension of their school difficulties, they are sloppy in their dedication and commitment to work responsibilities and, predictably, experience frequent periods of unemployment. Given the syndromic nature of self-control (DeLisi 2011), it is probable that offender sloppiness is also gleaned from their murder scenes. This issue is explored within the framework of self-control theory next.

In an important work that sought to differentiate the murder scenes of sexual homicides and potentially draw inferences from the crime scenes to profile the offender, Ressler, Burgess, and Douglas (1988) advanced a typology of sexual homicide offending. On one hand was a disorganized sexual murderer whose offense was spontaneous, whose victim was known but depersonalized, and who engaged in minimal conservation during the attack. The disorganized killer's murder scene was random and sloppy, characterized by sudden violence with minimal use of restraints, and sexual acts after death. The victim's body was left in view at the death scene, and importantly, weapons and other forensic evidence were often present at the scene.

In contrast, the organized sexual murderer committed a planned killing of a targeted stranger who was personalized. There was controlled conservation during the killing, use of restraints, aggressive acts before death, and demand for submission. The entire crime scene of the organized killer reflected control. The victim's body was hidden, often transported, and there was an absence of weapons or other forensic physical evidence. Taken together, the Ressler et al. (1988) typology painted a picture of two very different perpetrators in terms of their functioning, their competence, their mastery over their murder victim, and (semantically most important) their *control* over their murder scene.

As shown in Fig. 1, the disorganized-organized sexual homicide offender typology can effectively be understood as a continuous distribution of self-control. In many respects, the disorganized offender's behavior contemporaneous to their homicide event is consistent with the conduct of the modal criminal offender. The murder is unplanned and hasty and represents a poorly contemplated, rash action to satisfy an underlying desire (recall that the victim is known to the offender). The crime scene of the disorganized sexual homicide offender is, simply, sloppy.

In contrast, the organized sexual homicide offender seemingly exudes self-control in certain respects. The crime scene evinces self-control: forensic evidence is cleaned up, and weapons

Social Control Theory of Sexual Homicide Offending,
Fig. 1 Self-control and sexual homicide organization

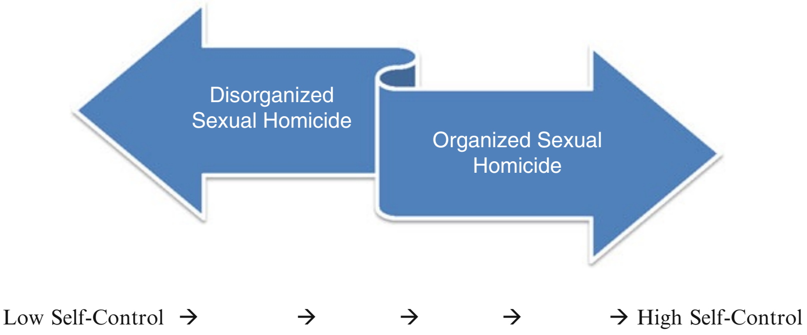

and other physical evidence are removed from the scene. The murder itself is premeditated and reflects considerable planning and control; it is a well-executed plan. Planning is not the strong suit of offenders with low self-control. All of these conditions of the organized sexual homicide offender intimate a paradoxical relationship between self-control and violence, one that could not be anticipated by self-control theory.

A sense of higher self-control has also been shown among other types of multiple homicide offenders. In their study of 34 subjects who perpetrated 27 mass murders between 1958 and 1999, Meloy et al. (2001) found that adolescent mass murderers are more often predatorily rather than affectively violent and often do not display sudden, highly emotional warning signs. In other words, their violence represents considerable control evidenced by their selection of victims.

Recasting the organized-disorganized typology as a continuous measure of self-control allows researchers to utilize a criminological theory to conceptually understand sexual murderers. Empirical research is needed to access the threshold where high levels of offender self-control are associated with extremely violent conduct. Additional research is also needed to access the linearity of this relationship. It is possible that the relationship between self-control and sexual homicide is curvilinear. It is also possible that high self-control evinced by sexual homicide offenders is comorbid with other important constructs, particularly sexual sadism and psychopathy. These correlates of sexual homicide offenders are examined next.

In his landmark *Psychopathia Sexualis*, Richard von Krafft-Ebing (1886) introduced the construct of sadism which involved a melding of sexual gratification of the offender with the infliction of pain and humiliation of the victim. In its original conceptualization, sadism was referred to as lust murder and correlated with an absence of guilty (today, known as psychopathy) and even the attempted consumption of the victim. A recent summary review of studies that examined the prevalence of sexual sadism among forensic populations produced a variety of important findings. Chief among them was the finding that the prevalence of sexual sadism was very high, about 30–40 % among offender samples. However, when the sample contained exclusively homicide offenders, the prevalence of sexual sadism was in the 70–80 % range (Krueger 2010; also see, Berner et al. 2003).

An important comparative study also shows the salience of sexual sadism to homicide offending. Yarvis (1995) interviewed and surveyed 78 murderers, 92 rapists, and ten rape-murderers to assess the prevalence of sexual sadism. None of the murderers and just 6.5 % of the rapists met the criteria for sexual sadism; meanwhile, the prevalence among rape-murderers (or, in the parlance of the current entry, sexual homicide offenders or sexual murderers) was 30 %. The criminal careers and life circumstances of sexual homicide offenders are typically severe and noteworthy for their psychopathology even compared to other serious criminals. For instance, Trojan and Salfati (2010) recently compared 137 single-victim homicide cases drawn from the Cincinnati, Ohio, Police

Department and 17 closed serial murder cases obtained from the FBI. Whereas the single homicide offenders averaged 13 arrests and eight convictions, serial murderers averaged five arrests and three convictions. Although this seems comparatively less severe, five or more arrests is a standard measure of habitual or chronic criminality (the most prolific serial killer had 25 prior arrests). The criminal histories of serial murderers also demonstrated more instrumental offenses suggesting a specific targeting of a victim that is consistent with the organized typology described earlier.

Hill et al. (2008) studied the forensic reports of 139 sexual homicide offenders in Germany who had served time in prison and 90 of whom had been released. The mean number of sexual homicide victims per offender was 3.7. Twenty years after their release, more than 23 % of offenders recommitted sexual offenses, and more than 18 % committed nonsexual violent offenses, such as armed robbery and aggravated assault. Three men or more than 3 % were subsequently convicted of attempted or completed murder. They found that the earlier that an offender committed his first sexual homicide, the more likely he was to be convicted of a sexual crime, which included extremely violent offenses such as sexual homicide, rape, sexual abuse of a child, and other sexual assault and any violent crime. Nearly one in five defendants convicted of sexual homicide has prior convictions for homicide, and more than half have other violent crimes in their criminal history (Häkkänen-Nyholm et al. 2009).

Langevin and his colleagues (Langevin 2003; Langevin et al. 1988) compared the antisocial histories of 33 sexual homicide offenders to 80 sexually aggressive offenders, 23 sexual sadists, and 611 general sex offenders. All these men had been seen for psychiatric assessments either for pretrial or as part of parole evaluation. Overall, the criminal careers and antisocial pasts of sexual homicide offenders were significantly worse than even these other severe risk groups. More than 27 % of the sexual homicide offenders had been committed to reform school, more than 21 % were gang members, over 15 % had been expelled from school for behavioral problems, and greater than 30 % had prior history of animal cruelty, vandalism, and/or fire setting.

Dietz et al. (1990) conducted a descriptive study of 30 sexually sadistic criminals. All of these men intentionally tortured their victims for sexual arousal. Their crimes often involved careful planning, the selection of strangers as victims, approaching the victim under a pretext, participation of a partner, beating victims, restraining victims and holding them captive, sexual bondage, anal rape, forced fellatio, vaginal rape, foreign object penetration, telling victims to speak particular words in a degrading manner, murder or serial killings (most often by strangulation), concealing victims' corpses, recording offenses, and keeping personal items belonging to victims. Beauregard and Proulx (2002) compared the offending process of two groups of non-serial sexual homicide offenders, 20 who were characterized or driven by anger and 16 who were motivated by sadistic pathology. Sadistic murderers were significantly more likely than non-sadistic killers to premeditate their crimes, humiliate their victims, mutilate their victims, and dismember the corpse of the victim. Sadistic murderers took longer to complete their offenses – nearly 90 % took longer than 30 min and were significantly more likely to be apprehended. They also experienced greater positive affect after the commission of their murder. Sadistic killers were also less likely to surrender to police after their crimes, less likely to admit to their crimes, less likely to admit all of the acts committed during their homicides, and less likely than other murderers to admit responsibility.

Research based on analyses of crime scene behaviors of sexual murder and rapes differentiates these forms of sexual violence. In cases of rape, offense behaviors include blindfolding and binding of victims, ripping the clothing of victims, possession and display of weapons, theft, and vaginal penetration. In cases of sexual homicide, there were more severe and depraved indicators of violence including forensic awareness on the part of the offender (e.g., cleaning up or destroying evidence at crime scene), multiple infliction of wounds, and multiple forms of penetration of the victim including objects (Salfati and Taylor 2006).

Across these studies, there is interesting evidence for considerable self-control on the part of sexual murderers. It is not known currently whether these offenders are demonstrating self-control in the Gottfredson and Hirschi (1990) sense or whether they are sadistic psychopaths whose proclivities for extreme violence are guiding their behavior. What is known, however, is that the violence of these offenders at least at the culmination of their murders does not comport with the sloppy profile of the modal criminal offender, one whose life is rife with self-control deficits.

Discussion

For the purposes of theory building, the extreme violence of sexual homicide offenders provides an opportunity to stretch the empirical veracity of constructs. At times, the behaviors of the most severe offenders cast well-known criminological constructs in a new light. For instance, cognitive functioning and intelligence are inversely associated with antisocial behavior, yet some sexual homicide offenders (most famously Ted Bundy) were highly intelligent. This does not impugn the intelligence-crime relationship but does offer opportunities to rethink theory and the putative associations between variables at different points on the offending distribution.

Gottfredson and Hirschi (1990) advanced a theoretical construct that was advertised to be *the cause* of crime, and empirically their theoretical ideas have fared well (see, DeLisi 2011). The current authors believe their theory also has utility for understanding sexual homicide offending and that the range of self-control can be superimposed on existing typologies of sexual homicide. But at another level, the pathology manifest by organized sexual murderers is contrary to the tenets of the theory. They are highly controlled. Based on prior studies of sexual homicide offenders, it is possible that what appears to be high self-control could be spurious and obscured by the effects of psychopathy and sexual sadism. Indeed, Meloy (1999) has previously characterized organized sexual murderers as compulsive (disorganized sexual murderers were referred to as catathymic in his model), which was characterized among other clinical features as having diagnoses for sexual sadism and psychopathy. Recent research has found that sadism and psychopathy are distinct factors in the psychopathology of sex offenders (Mokros et al. 2011); perhaps self-control is distinct as well.

Related Entries

- ▶ FBI Influence on State and Local Police
- ▶ Forensic Science
- ▶ Specialization and Sexual Offending

Recommended Reading and References

Beauregard E, Proulx J (2002) Profiles in the offending process of nonserial sexual murderers. Int J Offender Ther 46:386–399

Berner W, Berger P, Hill A (2003) Sexual sadism. Int J Offender Ther 47:383–395

Bouffard LA (2010) Exploring the utility of entitlement in understanding sexual aggression. J Crim Justice 38:870–879

Castle T, Hensley C (2002) Serial killers with military experience: applying learning theory to serial murder. Int J Offender Ther 46:453–465

DeLisi M (2001) Extreme career criminals. Am J Crim Justice 25:239–252

DeLisi M (2006) Book review: encyclopedia of murder and violent crime. Crim Justice Rev 31:172–173

DeLisi M (2011) Self-control theory: the Tyrannosaurus rex of criminology is poised to devour criminal justice. J Crim Justice 39:103–105

DeLisi M, Vaughn MG (2008) The Gottfredson-Hirschi critiques revisited: reconciling self-control theory, criminal careers, and career criminals. Int J Offender Ther 52:520–537

Dietz PE, Hazelwood RR, Warren J (1990) The sexually sadistic criminal and his offenses. Bull Am Acad Psychiatry Law 18:163–178

Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Häkkänen-Nyholm H, Repo-Tiihonen E, Lindberg N, Salenius S, Weizmann-Henelius G (2009) Finnish sexual homicides: offence and offender characteristics. Forensic Sci Int 188:125–130

Hill A, Habermann N, Klusmann D, Berner W, Briken P (2008) Criminal recidivism in sexual homicide perpetrators. Int J Offender Ther 52:5–20

Houts RM, Caspi A, Pianta RC, Arsenault L, Moffitt TE (2010) The challenging pupil in the classroom: the effect of the child on the teacher. Psychol Sci 21:1802–1810

Krafft-Ebing R (1886) Psychopathia sexualis. Enke, Stuttgart

Krueger RB (2010) The DSM diagnostic criteria for sexual sadism. Arch Sex Behav 39:325–345

Langevin R (2003) A study of the psychosexual characteristics of sex killers: can we identify them before it is too late? Int J Offender Ther 47:366–382

Langevin R, Ben-Aron MH, Wright P, Marchese V, Handy L (1988) The sex killer. Ann Sex Res 1:263–301

Maruna S (2006) Review of Paul Cromwell's in their own words: criminals on crime. Aust N Z J Criminol 39:274–275

McNamara JJ, Morton RJ (2004) Frequency of serial sexual homicide victimization in Virginia for a ten-year period. J Forensic Sci 49:529–533

Meloy JR (1999) The nature and dynamics of sexual homicide: an integrative review. Aggress Violent Behav 5:1–22

Meloy JR, Hempel AG, Mohandie K, Shiva AA, Gray BT (2001) Offender and offense characteristics of a nonrandom sample of adolescent mass murderers. J Am Acad Child Adolesc Psychiatry 40:719–728

Mokros A, Osterheider M, Hucker SJ, Nitschke J (2011) Psychopathy and sexual sadism. Law Hum Behav 35 (3):188–199.

Ressler R, Burgess A, Douglas JE (1988) Sexual homicide: patterns and motives. D. C. Heath, Lexington

Salfati CG, Taylor P (2006) Differentiating sexual violence: a comparison of sexual homicide and rape. Psychol Crime Law 12:107–125

Singer SD, Hensley C (2004) Applying social learning theory to childhood and adolescent firesetting: can it lead to serial murder? Int J Offender Ther 48:461–476

Surette R (1994) Predator criminals as media icons. In: Barak G (ed) Media, process, and the social construction of crime: studies in newsmaking criminology. Garland Publishing, New York, pp 131–158

Sutherland EH (1950) The sexual psychopath laws. J Crim Law Criminol 40:543–554

Trojan C, Salfati CG (2010) A multidimensional analysis of criminal specialization among single-victim and serial homicide offenders. Homicide Stud 14:107–131

Vaughn MG, DeLisi M, Beaver KM, Howard MO (2009) Multiple murder and criminal careers: a latent class analysis of multiple homicide offenders. Forensic Sci Int 183:67–73

Yarvis RM (1995) Diagnostic patterns among three violent offender types. Bull Am Acad Psychiatry Law 23:411–419

Social Disorder and Physical Disorder at Places

Sue-Ming Yang
Department of Criminology, National Chung-Cheng University, Min-Hsiung, Chia-Yi, Taiwan

Overview

Urban disorder has now become a center of public policy. Sennett (1970) viewed so-called disorder as a manifestation of the social diversity of a place where people come from different racial, ethnic, social, and economic backgrounds. Thus, disorder is considered a byproduct that comes with the modernization of urban cities. Sennett argued that disorder is actually "useful" for urban dwellers, as a disorderly environment "forces" people to get to know one another and therefore, enhances familiarity among heterogeneous population and reinforces racial integration. On the contrary, disorders are often considered as social negativities by criminologists. For example, James Q. Wilson (1975) pointed out that daily hassles such as street people, panhandlers, rowdy youths, or 'hey honey' calls trouble urban residents as much as crime does, if not more. Garofalo and Laub (1978) also argued that these urban characteristics of the modern society (disorder is one of them) are the main sources of the "fear" of residents, rather than the true fear of "crime." Wilson and Kelling (1982) further suggested that disorder has criminogenic effects with untended disorder eventually leading to crime problems. These contrasting viewpoints illustrate the subjective nature of what disorder is and its corresponding social meanings. Browsing through the studies of disorder, it is difficult to identify one single definition that everyone agrees upon. Disorder from one person's view might represent the norm to another person. In this entry, various definitions about disorder will be reviewed. Two major types of disorder, social and physical disorder, and their association with crime will be reviewed and discussed.

A city isn't just a place to live, to shop, to go out and have kids play. It's a place that implicates how one derives one's ethics, how one develops a sense of justice, how one learns to talk with and learn from people who are unlike oneself, which is how a human being becomes human.

– "The Civitas of Seeing" Richard Sennett (1989)

The Sources of Disorder

In the view of the Chicago School researchers, a city is not merely an environmental setting but is rather an organism that grows, changes, and evolves over time (Wirth 1962). Urban areas are usually characterized with diversity in the demographic composition of people who live and work within them. From this perspective, a city looks like a "mosaic of segregated people," as referred to by Robert Park (1928). In cities, different people demonstrate their cultural heritage, live in different lifestyles, yet cohabitate in the same areas. The close contact of these "differences" naturally leads to tension and conflicts among different groups and the lack of cohesion results in social disorganization. Within the disorganized environment, each social or racial group competes for resources and actively seeks to preserve their peculiar cultural forms in order to sustain their own conception of life (Park 1928).

With the expansion of a city, the influx of population brings different cultures and values into a city. For urban inhabitants, therefore, the increase in exposure to different cultures and other value systems and lifestyles tend to create uneasy feelings for individuals. Due to the unfamiliarity of other cultures, an increased level of diversity might be seen as a sign of disorder by mainstream society (Sennett 1989).

However, disorder is by no means a clearly defined concept in either research or practice. The meaning of disorder, just like all other social norms, changes along with the development of a city. In other words, the definition of disorder might vary by society and represents the social norms separating between approval and disapproval behavior (see Durkheim's Moral Education, 1961).

The definitions of social norms are always the result of competition and the transaction of different cultures and value systems. This process is critical as the social meaning of a behavior often determines how people react to the behavior. In other words, when a society defines certain behaviors as unwanted or deviant, people will start viewing them as a menace or danger to society. Conflict criminologists have long argued that crime is socially constructed and is used by the powerful groups in the society to control the disadvantaged underclass. This ambiguous nature of the social norm is even more salient when dealing with less severe social phenomena such as disorder. Similarly, the defining process of disorder is not equally influenced by all groups in the society—the powerful groups often enjoy the privilege of determining the "*norms*" (Durkheim 1961; Sennett 1970).

The Definition of Disorder

Disorder is not crime, and usually represents minor violations of social norms that do not directly harm other people. Unlike crime, there is no commonly agreeable definition to classify disorder. Skogan referred to it as a "slippery concept" (Skogan 1990, p. 4). He argued that disorders are behaviors that are not prohibited by the criminal law; or disorders in isolation constitute relatively minor offences. The ambiguity nature of disorder leaves room for different ways defining disorder.

In prior literature, various occurrences and behaviors have been identified as disorder, ranging from deteriorated buildings to teenagers hanging out on the street corner. Even the term "disorder" is not always used to describe these minor violations. Past research has referred to this particular phenomenon as "signs of crime" (Skogan and Maxfield 1981), "early signs of danger" (Stinchcombe et al. 1980), "urban unease" (Wilson 1968), "broken windows" (Wilson and Kelling 1982), "soft crime" (Reiss 1985), "public moral offenses" (Weisburd and Mazerolle 2000), "cues to danger" (Warr 1984), "incivilities" (Taylor 1999), "sign of disrespect"

(Bannister et al. 2006), “warning signals” (Innes 2004), and “disorder” (Skogan 1990). Finally, most scholars have now settled on the term of either *disorder* or *incivilities* (LaGrange et al. 1992). While disorder has been discussed and theorized prior to 1990 (see Wilson and Kelling 1982), Skogan was the first to clearly articulate disorder and its impacts on neighborhoods (Skogan 1990). In his book *Disorder and Decline*, Skogan points out that the concept of disorder is ambiguous and can represent a variety of meanings. Thus, he notes that it is important to differentiate between the friendly “active uses of the environment” (not disorder) and disorderly behavior which bothers the residents of a neighborhood, as the former may suggest a very well-connected neighborhood while the latter may represent a disoriented neighborhood.

In addition to how research defines disorder, it is also important to know how law enforcement determines what phenomena are considered disorderly. In Kelling and Sousa’s evaluation of New York City’s Quality of Life Policing effort, they used misdemeanor arrests to represent police efforts without providing a clear list of targeted disorders. In the four scenarios illustrated in the report, the enforcement was mainly targeted at public urination or public drinking (open container). Sure enough, based on the broken window thesis drafted by Wilson and Kelling, the actual practice probably included more types of “minor offense” including things like panhandlers, drunks, addicts, rowdy teenagers, prostitutes, loiterers, and the mentally disturbed. On the other hand, Thacher (2004) summarized order maintenance police practices and noted four main categories of “disorders” that were the focus of police: (1) obstructing or lying down in the subway, (2) public urination, (3) panhandling, and (4) youth or student parties (page 396).

What kind of phenomenon is considered disorderly? After reviewing the various definitions and findings from studies, Kubrin (2008) concluded the following “Definitions of disorder used by researchers and officials studying and practicing broken widows policing are not necessarily consistent with residents’ perceptions in their own communities.” (page 206). From the previous literature review, it is apparent that disorder is indeed a “slippery concept” in both research and practice.

The Disorder and Crime Association

Recently, disorder has become a new center of attention for criminologists and criminal justice practitioners. From the Broken Windows Thesis proposed by Wilson and Kelling (1982), fear of crime research (see Garofalo and Laub 1978), to broken windows policing (or called the quality of life policing in New York City), disorder has been formulated as the core element in the equation of inquiry.

Despite the popularity of the broken windows based ideas among practitioners, the association between disorder and crime has been a much debatable topic (see Taylor 1999 and Sampson and Raudenbush 2004). Wilson and Kelling (1982) asserted that disorders are the root cause of crime. They suggest that signs of disorder lead to a sense of low social control perceived by those in the community, which triggers residents’ fear. Fearful residents may further withdraw from the community, and subsequently more serious types of crimes will occur. Therefore, targeting crime would not solve the fundamental problem, but focusing on disorder intervention will result in a greater crime prevention benefit.

Similarly, Skogan believes that incivilities will eventually result in neighborhood decline because they cause a range of psychological, social-psychological and behavioral outcomes in neighborhoods. Beginning with the appearance of incivilities, the decline of a neighborhood is a gradual process (Skogan 1990, p. 65). First, signs of incivilities cause social withdrawal of residents from participating community activities. It also discourages people from cooperating with their neighbors. Second, the existence of incivilities weakens neighborhood morale. Residents become concerned about their personal safety after a number of incidents happen in their neighborhood. As a result, residents lose trust in each other. Third, from a practical standpoint, incivilities undermine the housing market. It directly affects the willingness of people to invest in the area. Due to a lack of investment, the housing market is suppressed and neighborhood decline becomes an unavoidable outcome. To test his hypothesis, Skogan (1990) collected information from 40 neighborhoods in six different cities during 1977–1983. The simple association between disorder and robbery victimization in those neighborhoods was + .80. However, when other neighborhood factors were taken into account, the association dropped to + .54, though still fairly substantial. Based on these findings, Skogan concluded that with the small number of cases at hand, it is hard to tell "*whether they have either separate 'causes' or separate 'effects' at the area level*" (1990, 73).

Keizer et al. (2008) conducted a series of field experiments on physical disorder and found that the presence of disorder, such as graffiti, trash etc. increased the likelihoods of passersby to commit minor disorderly behavior. Though this study provides some direct support of the negative effects of disorder, the study did not examine the effects of disorder on crime. Moreover, their study focuses on physical disorder only, rather than both types of disorder. The association between disorder and crime was further confirmed by Weisburd et al. (2012) longitudinal study of crime trends in Seattle, WA. Controlling for opportunity variables as well as characteristics derived from social disorganization theory at street level, physical disorder was found to be one of the strongest indicators predicting crime trends. Specifically, streets with more physical disorder incidents were much more likely to be in the chronic crime pattern instead of crime-free pattern. Also, increases in physical disorder over time were related to increased crime trends. Though the findings lend confidence to the disorder-crime association, only physical disorder measure was included in the study.

While both Wilson and Kelling and Skogan believe disorder in the neighborhoods causes crime and other social negativities, other scholars hold different opinions on the nature of their association. In a study of drug hot spots in Jersey City, Weisburd and Mazerolle (2000) found that serious crime and disorder tend to cluster together in drug hot spots. In other words, disorder and crime are positively correlated at these drug hot spots. In contemporary social disorganization theory, both Bursik and Sampson treat disorder as a consequence of the lack of social control or as a product of structural variables. Thus, disorder is only related to crime through its associations with structural variables. That is, crime and disorder should be positively correlated but the relationship is not *causal*.

Corman and Mocan (2000) examined the association between disorder and crime rates in NYC over time. However, they did not collect directly measured disorder data; rather, they used misdemeanor arrests as the proxy of disorder. They found that misdemeanor arrests in NYC from 1974–1999 were significantly, negatively related to robbery, motor vehicle theft and grand larceny after controlling for economic conditions and deterrence, but were not significantly related to the other four index crimes. Despite the empirical disagreement, within all the theoretical arguments about disorder/incivilities, most of them assume a positive association between disorder and crime. Meaning, wherever there is disorder, there tends to also have crime problem. The issue is, whether the association is correlational or causal. Also focusing on New York City, Geller (2007) examined housing conditions and the relationship to crime rates in NYC. She tested whether deteriorated housing conditions leads to crime increases as expected by the broken windows thesis. The results of her study did not support such a hypothesis. However, in the study, only the effects of physical disorder were tested. This is an important distinction as some studies have suggested that social and physical disorder may have different relationships with crime.

St. Jean (2007) conducted an ethnographic study in a high crime neighborhood in Chicago and found that disorder and crime tended to coexist in the same areas. However, he pointed out the differential association of social and physical disorder with crime. Yang (2010) in a longitudinal study of City of Seattle, WA, also found a significant correlation between disorder and crime. The direction of causation, however, was opposed to what was suggested by the broken windows thesis. The results from Granger causality tests generally showed no causal relationship between disorder and violence, and in a few places the causality appeared to run from violent crime to disorder. Additionally, Yang shared St. Jean's findings and argued that research should make a distinction between social disorder and physical disorder. They are not only qualitatively different phenomena; moreover, these two types of disorder have different associations with crime.

Separating Social Disorder and Physical Disorder

Activities under the classification of disorders, or incivilities, are not all homogeneous. Under the big umbrella of disorder, there are two commonly used subcategories: physical disorder and social disorder (Skogan 1990; Hancock 2001; Sampson and Raudenbush 1999). Later, Taylor (1999) proposed very specific operational definitions of social and physical disorder (incivilities). Social incivilities include behaviors such as public drinking, drunkenness, rowdy and unsupervised teens, sexual harassment, arguing and fighting, open prostitution, and public drug sales. While physical incivilities include things like abandoned buildings, graffiti, litter, vacant and trash-filled lots, unkempt yards and housing exteriors, abandoned cars and the conversion of houses and apartments to drug-selling locations.

From a research standpoint, social disorder and physical disorder are qualitatively different. Sampson and Raudenbush (1999) clearly pointed this out in the beginning of their article: "*[s]ocial disorder, we refer to behaviors involving strangers and considered threatening... [p]hysical disorder, we refer to the deterioration of urban landscapes* (p. 603–604)." From this statement, it is obvious that social disorder requires the involvement of human behaviors, which impose threats to others in the neighborhood. Thus, the key element defining social disorder depends on the presence of actors who perform the offensive actions. St. Jean (2007) also believes that neighborhood disorder exists in two forms: physical and social disorder. He argues that physical disorder refers to "unpleasant neighborhood characteristics" created by messy environment such as litter, trash, abandoned tires, empty lots, and broken windows. On the contrary, social disorder is a result of the "unpleasant and perhaps intimidating social interactions among people in public space." As such, social disorder includes social behaviors like panhandling, public drinking, fighting, and loitering (St. Jean 2007, p. 196).

In addition, Hancock (2001) suggested another less general way of classifying disorder by looking at its duration. Based on its temporal length, he classified disorder into two groups: the episodic disorder which occurs at no fixed time, such as public drinking, and the on-going type of disorder like abandoned buildings which are always there, unless someone takes initiative to change the condition. Putting these ideas together, social disorder is usually an episodic behavior, which only lasts for a short duration. Conversely, physical disorder represents an objective condition that might last for a long period of time unless some actions are taken to change it. The focus of the latter is on the physical conditions rather than on individuals acting within the conditions.

Many of the past studies proposed clear definitions of social and physical disorder along with characteristics included within each category (Sampson and Raudenbush 1999; Skogan and Maxfield 1981; Skogan 1990; Taylor 1999; LaGrange et al. 1992). However, when doing analysis, people often combined these two categories together into one total disorder measure – assuming that they represent the same underlying construct. For instance, Skogan and Maxfield (1981) used a general term of "sign of disorder" to represent the violations of "people's expectations about fit and proper conditions and conducts." Within this "sign of disorder" they included a wide range of circumstances like unsupervised teens hanging out on the streets, abandoned buildings, illegal drug use, and vandalism. Sampson and Raudenbush (1999), though proposed clear definitions about social and physical disorder separate, did not examine their individual effects in their test of the broken window thesis. In their analysis stage, they made a statement that "*the results are so similar for physical and social disorder, and because the two scales are highly correlated (r = .71), we combine them into a summary index of disorde*r..." (p. 626). As such, the distinction between social and physical disorder and their associations to crime were not fully examined in the study.

In a separate study, Raudenbush and Sampson (1999) tested methodological properties of social and physical disorder measures using item response modeling with hierarchically constructed data. They found that social disorder measures were recorded less often and thus, less reliable compared to physical disorder measures, due to the rarity of the social disorder observations.

Steenbeek and Hipp (2011) also followed the item response modeling approach to explore the reciprocal effects of disorder on social structures and social control under the social disorganization framework. Using data collected in the Netherlands, they argued that the correlations between social disorder and physical disorder items were too strong. Thus, they constructed a disorder measure at the neighborhood level combing all social and physical disorder items rather than separating the two constructs.

In short, despite the fact that many prior studies have acknowledged the distinction between social and physical disorder, their separate effects to crime and other social problems are still understudied empirically.

Social Disorder, Physical Disorder, and Crime

Overall, the qualitative differences between social and physical disorder are rarely addressed. Mixing the concepts of social and physical disorder hinders us from disentangling the intertwined relationship between disorder and crime. From theoretical perspectives, social disorder and physical disorder are different in many ways. First, social disorder involves actors and current actions; thus, the presence of social disorder perhaps provides a pool of potential targets, motivated offenders or both for violent offenses to occur. Physical disorder, however, does not necessarily involve actors. Most of the time, we do not see/know the actors who dump trash or break glass in the next block. We recognize physical disorder when we see it, even without seeing the creator(s) of it. As such, physical disorder provides unmistakable visual cues to the users of the space. Therefore, residents' perception of physical disorder should be more consistent than their perceptions of social disorder as the latter involves an individual's value judgment.

Compared to social disorder, physical disorder is generally considered less criminogenic for crime. An abandoned place could attract would-be offenders to engage in criminal activities. However, physical disorder itself, without the presence of potential offender/victims, should be related to crime to a lesser extent than social disorder. It is possible that there exists a sequential relationship between social disorder and physical disorder. This possibility will be examined later using the Granger causality test to see if there is a directional relationship between the two types of disorder.

The distinction between social and physical disorder also carries importance on another equally important subject—fear of crime. In addition to crime, fear is another phenomenon that has long been connected to disorder (see Garofalo and Laub 1978; Kelling and Coles 1996). Kelling and Coles (1996) argued that fear mediates the effects of disorder on crime in a community. However, the differential effects of social and physical disorder were not discussed by Kelling and Coles. Based on Sampson and Raudenbush's phrases cited earlier, it is expected that social disorder and physical disorder should have different impacts on fear. Social disorder invokes a feeling of fear to residents, while physical disorder provides a more neutral image that might still be bothersome to some residents. As such, the emotional reactions related to social and physical disorder are not necessarily the same.

Empirical studies also support that social disorder and physical disorder are different and so are their associations to crime. The empirical evidence from Taylor's (1999) observations in

Baltimore provides a good example of how different those two types of incivilities can be. In the study, physical disorder is found to be correlated with changes in rates of aggravated assault, burglary, and motor vehicle theft while social disorder is correlated only with changes in the rate of rape (all of the correlations are significant at .10 level, Taylor 1999, p. 5). St. Jean (2007) analyzed crime data and interviewed residents in poor neighborhoods in Chicago to study disorder and crime. From cross-tabulation analyses, he found that places with high collective efficacy and low social disorder have fewer crime problems. Places with low social disorder tend to have low street drug dealing. However, high physical disorder had no significant association with drug dealing, robberies and batteries. While the finding seems to support the broken windows thesis, he argued that places with high levels of social disorder do not necessarily have more crime problems. The data at best could argue that in the absence of social disorder, it is also less likely for the places to have drug dealing problems.

Nonetheless, St. Jean sought for a more thorough understanding about the underlying mechanism linking disorder and crime by conducting in-depth interviews with street criminals and other individuals who live in the crime ridden areas to get their perspectives. The presence of physical disorder alone does not attract drug dealers to the area; rather, it is the location and the activities occurring within and nearby the area that determine whether it is going to become a drug hot spot. Though the dilapidation of buildings does show that the government did not invest in the area and perhaps there is a demand for drugs in the place from people who could not find any decent jobs. The association between robberies and physical disorder is even weaker. The robbers pick places to commit crime based on accessibility and anonymity, not because of the abandoned buildings or trash lying around in the areas. As for batteries, he found that places with high levels of physical disorder tend not to have battery problems. It is reasonable because places with physical disorder tend to be occupied by abandoned lots or buildings and therefore, are less likely to have domestic disputes due to fewer people living in the area.

From the qualitative interviews, St. Jean found that social disorder has strong and significant associations with narcotic violations, robberies, and batteries. But a closer examination of the mechanism behind the association revealed the complexity in the phenomena. Places with low social disorder tend not to have drug dealings; however, places with high social disorder may or may not attract drug dealers to the areas. Moreover, the association between social disorder and narcotic violations is sometimes reciprocal. That is, people who use drugs might be involved in social disorder while they are waiting for their needs to be met. As for robberies and social disorder, the relationship is positive and significant. But further inquiry shows that the locations and actors in the places matter more than whether the areas have social disorder. Sometimes, having more people hanging around a place provides a potential pool for robbery targets. However, when the people do not seem to be vulnerable or do not possess enough monetary goods, they would not be selected as targets. The tight coupling between social disorder and batteries shows that places with more disorderly people are also more likely to have domestic conflicts. Again, St. Jean found that the association is only supported in places occupied by high social disorder and disruptive families. In places that have high levels social disorder but are occupied mainly by non-disruptive families the association diminishes. Overall, the findings from the ethnographic portion of this study were consistent with what was found in the statistical analyses. That is, physical disorder does not matter in predicting crime places. Moreover, social disorder does not matter in offenders' decisions in choosing a place as their base; what really matters to them is where the place is located and who are hanging around the place. St. Jean's study provides a comprehensive picture illustrating the relationships among crime, social, and physical disorder. Nonetheless, the analyses in the study were all based on a short time period and limited the selected poor neighborhoods in one high-crime police beat in Chicago. Therefore, it is hard to know whether the relationships remain the same over time in other places.

Finally, in another study using 16 years of data collected at the census-block-group-level in Seattle, Yang (2010) studied the long-term associations between social disorder, physical disorder, and violent crime. The dual-trajectory analysis findings suggested that a block group with a low level of disorder (for both social and physical disorder) is less likely to experience violence problems. If a block group is assigned to the low rate social disorder trajectory, then there is an 88.1 % of chance that this block group will also be found in the no (or negligible) violent trajectory. Similarly, 82 % of block groups with negligible physical disorder problems are also relatively free of violence.

However, block groups that have high disorder problems are not necessarily also plagued by violence. For block groups within the high social disorder trajectory, there is only about a 30.6 % of chance of also being classified into the high rate violent trajectory group. Also, having been assigned to the high physical disorder trajectory only corresponds to a 30 % of chance of having a high level of violence.

This means that having no disorder can be seen as a powerful protective factor for block groups in preventing future violence – as there is less than 1 % of a chance that a block group will have a high level of violence if it has negligible disorder problems. Although having high disorder can be a risk factor to increase the chance that a block group experiences violence, only about 30 % of block groups that showed a high level of disorder also showed a high level of violence. This means that by using disorder as a risk factor, we will have a 70 % chance of making an incorrect prediction of where high violence places would be located. Perhaps lack of disorder and having high-disorder are not the two sides of the same coin. Or perhaps there is not a linear relationship between disorder and violence.

Furthermore, Yang also conducted geographic analysis on the selected block groups with high levels of social disorder, physical disorder, and violence (see Fig. 1). The visual illustrations on social disorder-physical disorder dyad and violence further confirm the previous findings. First, block groups with high violence trajectory are always located within the areas with a high level of social disorder. However, places having high social disorder do not necessarily have high violence rates. In other words, to locate a high violence block group, social disorder provides more accurate information than physical disorder. This finding echoes St. Jean's study results from Chicago. Second, places that are low in both social disorder and physical disorder are also free of violence problems. When a block group is free from disorder problems, it is also more likely to be free from violence problem. On the other hand, when a block group starts to show any signs of a disorder problem, it is very unlikely that it will exhibit a low rate of violence. High physical disorder block groups, though not informative of where the high violence block groups are located, are predictive of where the *low* violence block groups will *not* be. The spatial concordance between social disorder and violence points out a possibility that there may exist a connecting mechanism that governs both disorder and violence at places. The studies of St. Jean and Yang both show that having more broken windows does not necessarily predict what the level of crime will be at a given place.

The Next Step for Disorder Research

Overall, the empirical findings and theoretical arguments point to the need for studying these two types of disorder separately. Social disorder is found to be more relevant to crime than physical disorder. Perhaps it is because people involved in disorderly behaviors are also more likely involved in violence, regardless of the environmental factors. Or, perhaps it is also possible that these people might provide visible targets for motivated perpetrators or they might be the instigators of violent behavior.

Another possible source of variation is the type of measures used in the study. One of the major obstacles on studying the linkage between disorder and crime is whether residents could actually distinguish between disorder and crime (Gau and Pratt 2008). This is crucial as the mechanism outlined in the broken windows thesis to be effective; residents need to be aware of the

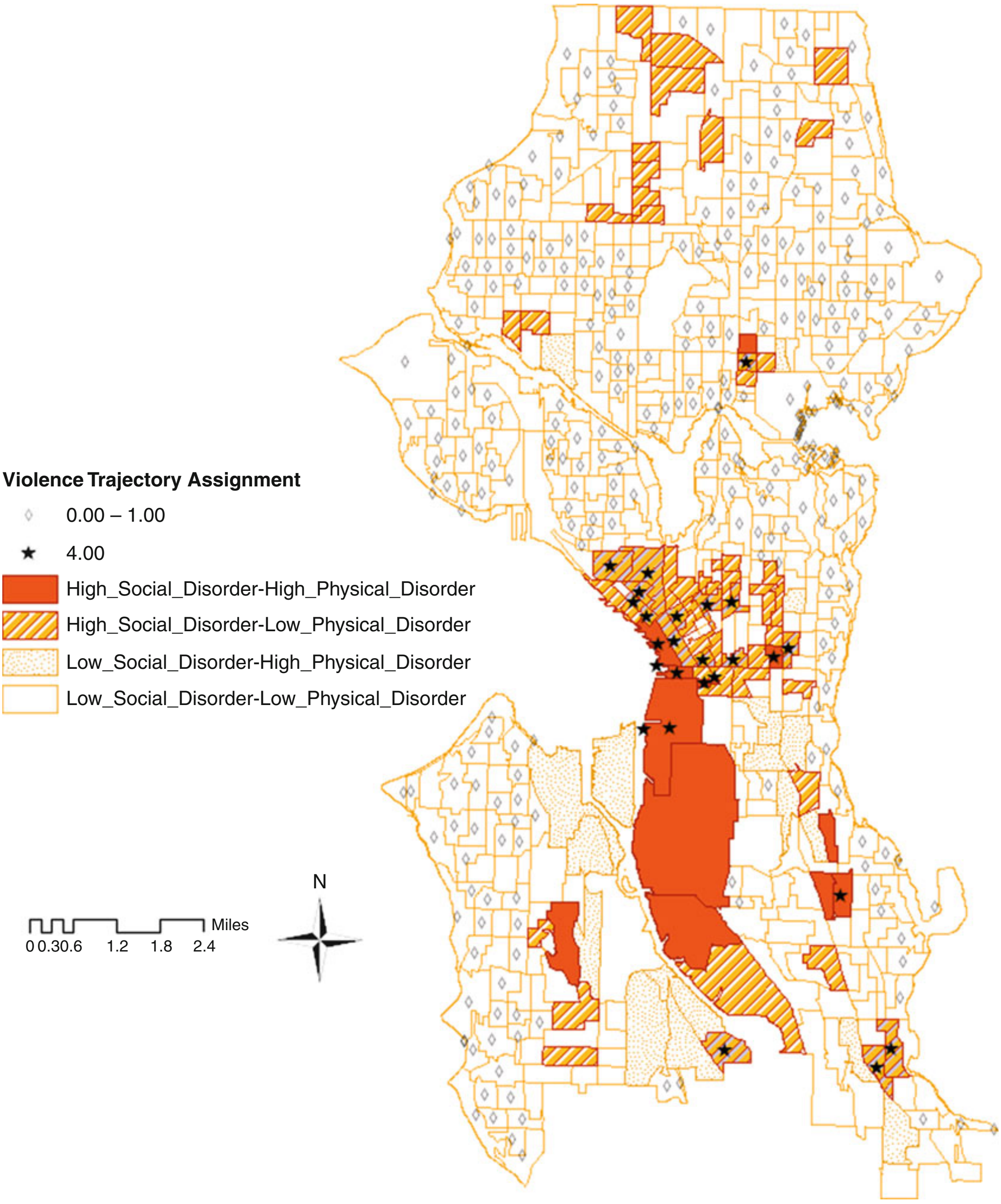
Violence Trajectory Assignment
0.00 – 1.00
4.00
High_Social_Disorder-High_Physical_Disorder
High_Social_Disorder-Low_Physical_Disorder
Low_Social_Disorder-High_Physical_Disorder
Low_Social_Disorder-Low_Physical_Disorder
Miles
0 0.3 0.6 1.2 1.8 2.4
N

Social Disorder and Physical Disorder at Places, Fig. 1 Geographic Locations of the 15 Selected Block Groups

existence of disorder. As such, some scholars started to research the impacts of perceptual disorder and the factors that could lead to perceived disorder. Clearly, observed levels of disorder play a role in this process, but perceptions are more critical. Generally, it is found that the perceptions of crime and disorder are highly related and do not have discriminant validity (see Gau and Pratt 2008; Worrall 2006). In other word, disorder and crime might be manifestation of the same underlying construct with different level of intensity. Thus, police strategies targeting disorder, as opposed to crime, might not lead to either crime or fear reduction effects as intended.

Innes (2004) also proposed a potential mechanism about how disorder affects residents' perceptions about crime and disorder. What matters most is not the nature of disorder; rather, it is the interpretation of disorder that leads to fear and negative emotions. The signs of disorder provide a framing function that helps people to give meanings to the quality of places. As such, social disorder and physical disorder might send out different signals to people. Physical disorder might reveal the lack of investment of the places by both government and residents. On the other hand, social disorder can be indicative of the type of people who hangout in the areas and the strengths of social ties in the area.

Hinkle and Yang (in progress) further point out that it is important to evaluate whether residents' perceptions of disorder matches with the observed level of disorder measured by researchers through systematic social observations. They further argue that social disorder and physical disorder might have differential conditional effects on residents' perception about the disorder problem, fear of crime, and satisfaction of the residential environment. Thus, scholars who are interested in disorder-related research need to examine social disorder and physical disorder separately. Moreover, it is also important to control the methods of measurement used to gauge levels of disorder, be it the actual level or the perception of the problem.

Disorder, unlike crime, is a common phenomenon and exists in many people's daily lives. It has also become an important issue for research in many disciplines. Recent studies have shown the importance of evaluating social disorder and physical disorder separately. These two types of disorder are not only qualitatively different, but they also relate to crime in very different ways. Moreover, the presence of social disorder and physical disorder send very different signals to passersby, residents and potential offenders and possibly lead to different perceptions about places. Future research needs to pay closer attention to the distinction between the two types of disorder and their implications for social policy.

Related Entries

▶ Broken Windows Thesis

▶ Defining Disorder

▶ Discriminant Validity of Disorder and Crime

▶ Disorder: Observational and Perceptual Measures

▶ Order Maintenance Policing

Recommended Reading and References

Bannister J, Fyfe N, Kearns A (2006) Respectable or respectful? (in) civility and the city. Urban Studies 43:919–937

Corman H, Mocan N (2000) A time-series analysis of crime, deterrence, and drug abuse in New York city. Am Econ Rev 90:584–604

Durkheim E (1961) Moral education. The Free Press, New York

Garofalo J, Laub JH (1978) The fear of crime: broadening our perspective. Victimology 3:242–253

Gau JM, Pratt TC (2008) Broken windows or window dressing? Citizens' (in) ability to tell the difference between disorder and crime. Criminol Public Policy 7:163–194

Geller A (2007) Neighborhood disorder and crime: an analysis of broken windows in New York City. Unpublished Doctoral Dissertation, New York: Columbia University

Hancock L (2001) Community, crime and disorder: safety and regeneration in neighborhoods. Palgrave, New York

Innes M (2004) Signal crimes and signal disorders: Notes on deviance as communicative action. The British Journal of Sociology, 55(3):335–355

Keizer K, Lindenberg S, Steg L (2008) The spreading of disorder. Science 12:1681–1685

Kelling GL, Coles C (1996) Fixing broken windows: restoring order and reducing crime in American cities. Free Press, New York

Kubrin CE (2008) Making order of disorder: a call for conceptual clarity. Criminol Public Policy 7:203–214

LaGrange RL, Ferrao KP, Supancic M (1992) Perceived risk and fear of crime: role of social and physical incivilities. J Res Crime Delinq 29:311–334

Park RE (1928) Forward. In: Wirth L (ed) The ghetto. The University of Chicago Press, Chicago, IL, pp vii–ix

Raudenbush S, Sampson RJ (1999) Ecometrics: toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. Sociol Methodol 29:1–41

Reiss AJ (1985) Policing a city's central district: the Oakland story. U.S. Department of Justice, National Institute of Justice, Washington, DC (March)

Sampson RJ, Raudenbush SW (1999) Systematic social observation of public spaces: a new look at disorder in urban neighborhoods. Am J Sociol 105:603–651

Sampson RJ, Raudenbush SW (2004) Seeing disorder: Neighborhood stigma and the social construction of "broken windows." Social Psychology Quarterly, 67, 319–342

Sennett R (1970) The uses of disorder: personal identity and city life. W.W. Norton and Company, Inc., New York

Sennett R (1989) The civitas of seeing. Places 5(4):82–84

Skogan WG (1990) Disorder and decline: crime and the spiral of decline in American neighborhoods. Free Press, New York

Skogan WG, Maxfield MG (1981) Coping with crime: individual and neighborhood differences. Free Press, New York

St. Jean PKB (2007) Pockets of crime: broken windows, collective efficacy, and the criminal point of view. The University of Chicago Press, Chicago, IL

Steenbeek W, Hipp JR (2011) A longitudinal test of social disorganization theory: feedback effects among cohesion, social control, and disorder. Criminology 49(3):833–871

Stinchcombe AL, Adams CA, Heimer KL, Scheppele TW, Taylor DG (1980) Crime and punishment – changing attitudes in America. Jossey-Bass, San Francisco

Taylor RB (1999) Responses to disorder: relative impacts of neighborhood structure, crime, and physical deterioration on residents and business personnel. National Institute of Justice, Washington, DC

Thacher D (2004) Order maintenance reconsidered: moving beyond strong causal reasoning. J Crim Law Criminol 94(2):381–414

Warr M (1984) Fear of victimization: why are women and the elderly more afraid? Soc Sci Quart 65:681–702

Weisburd D, Mazerolle LG (2000) Crime and disorder in drug hot spots: implications for theory and practice in policing. Police Q 3(3):331–349

Weisburd D, Groff ER, Yang S-M (2012) The criminology of place: street segments and our understanding of the crime problem. Oxford University Press, New York

Wilson JQ (1968) Varieties of police behavior. Harvard University Press, Cambridge

Wilson JQ (1975) Thinking about crime. Basic Books, New York

Wilson JQ, Kelling GL (1982) Broken windows: the police and neighborhood safety. Atlantic Monthly 211:29–38

Wirth L (1962) The ghetto, 7th edn. The University of Chicago Press, Chicago

Worrall JL (2006) The discriminant validity of perceptual incivility measures. Justice Q 23:360–383

Yang S (2010) Assessing the spatial- temporal relationship between disorder and violence. J Quant Criminol 26(1):139–163

Social Disorganization and Terrorism

Susan Fahey
Criminal Justice Program, Richard Stockton
College of New Jersey, Galloway, NJ, USA

Synonyms

Anomie; Breakdown; Collective action; Normlessness; Political action; Violent

Overview

One would expect more terrorism incidents to originate in a socially disorganized environment, that is, an environment that has experienced a rapid social change that has destroyed the regular rules and bounds on human behavior. The notion that one's surroundings may influence behavior is not a new one. In particular, the idea that the organization of the social environment, such as neighborhoods, families, and communities, may stimulate violence has been around since the origins of the social sciences. Emile Durkheim was a pioneer in describing the anticipated effects of the social environment on violent behavior.

Anomic Behavior

To understand how a socially disorganized environment influences one's behavior according to Durkheim (1930 [1951]), it is important to understand how he viewed human nature. To Durkheim, humans lacked the ability to regulate their desires and wants. In fact, humans were like other animals in that they were controlled by their needs and desires, particularly food, shelter, and the sexual drive. Unlike animals, though, humans have the power of reflection, the ability to imagine more than they have, and it is because of this ability to imagine more that they require an external force to limit their desires and needs. In normal times, the social environment provides this external limiting force; that is, one's behavior is regulated by the social rules and norms that are ordinarily in force. For example, if an individual sees that another has a very nice and desired vehicle, under most circumstances, that individual will not simply go and take that vehicle by force. Under most circumstances, the individual will refrain from taking the vehicle, because the individual is kept in the bounds of legal and moral behavior by the traditional rules of society. The fear of arrest and the disappointment expected by one's family and community as well as other social consequences keep individuals from simply taking the object of their desires.

However, Durkheim (1930 [1951]) says that if the society is experiencing temporary but rapid social changes, normal circumstances do not apply. In fact, during these times, what are called anomic or socially disorganized times, it is as if the rules of society have been stripped off and individuals are left in a world with no normal ways of doing things and no rules about wrong and right. In addition, the power of reflection, the power to imagine more, better, and different will come into play. In this world, individuals may become so disoriented by the loss of rules and controls that they may start to want everything they see. In many cases, individuals will not be able to satisfy all of their wants and desires and will become unsettled. Because Durkheim studied what led people to commit suicide, he states that such persons are more likely to commit suicide. Compared to well-regulated societies, those which experience rapid social changes that create anomie are more likely to experience increases in suicides.

Durkheim's (1930 [1951]) writings on the effects of anomie on behavior can be and have been applied to crime and terrorism. An example of a rapid social change sufficient to bring on a state of anomie in the society might be a civil war between an insurgent force and the government. In such a society, where an insurgent force is violently attempting to overthrow the government and its rights to make and enforce rules about legal and illegal conduct, the rules that dictate right and wrong in everyday life will not be in place. This constitutes anomie, a state of deregulation of behavior.

An individual living in such a state will be unsure as to what the new rules and regulations are in this different, anomic world. When the old rules were forcefully removed by the outbreak of civil war and the new rules have not been made clear or taken root yet, individuals will either need to restrain their desires and needs themselves or they will find this new anomic world without rules intolerable. The inability to tolerate this anomic world may make it more likely for some people to do behavior that is against the rules and norms of behavior that used to be in force in the society. If such individuals are sufficiently motivated, perhaps by a grievance against the government (Gurr 1976), such individuals may become involved in extreme acts, including terrorism. Anomie, according to Durkheim, is not a permanent condition. Over time, a new set of rules to define what is normal in society will develop, and the society will regain its equilibrium. This new set of rules will make living more tolerable for individuals again, and this will decrease the rate at which individuals engage in problematic behavior.

It is important to remember that this theory of human behavior is probabilistic; it suggests that the presence of such rapid social change makes it *more likely* that *some* individuals will engage in actions that are problematic. It is not a theory that predicts that all individuals will engage in problematic behavior nor is it definite that any specific individuals will be so engaged.

How Much Rapid Social Change Is Needed?

To result in anomie, a rapid social change must involve a large-scale structural challenge to the ordering and functioning of societal institutions (Piven and Cloward 1977). In other words, the rapid social change has to be serious and broad. It is also important that the rapid social change be large enough to stimulate changes in the structure and routines of everyday life. If a rapid social change occurs that does not interrupt the structure and routine of everyday life for most people, it is unlikely to change the behavior of those individuals whose routines continue uninterrupted. For example, when individuals are released from the routine of going to work every day as well as the sustenance provided by work, this changes the pattern and rhythm of their lives. If the interruption continues for a period, it is likely sufficient to produce anomie.

How Do Anomic Conditions Affect Behavior?

A socially disorganized environment likely affects behavior through its effects on the structure and routine of everyday life (Snow et al. 1998). People are more likely to take action to preserve what they know and their current routines rather than to seek something new and better. The known routines they will act to save are the things they do every day (behaviors) and the way they think about these things (cognition). The behaviors involve habitualized patterns of action. The cognition involves attitudes individuals adopt when approaching their daily routines. These attitudes chiefly involve an unquestioning, unreflective routine way of getting through the day.

These behavioral and cognitive routines of everyday life are what constitutes the *quotidian*, which is derived from the French word for daily. For example, if a rapid social change occurs that does not affect the daily routine for most people, such as the actions as well as thoughts and ideas of getting up every day to go to work, come home from work, and taking care of children and home, then it is unlikely to stimulate people to do something about the rapid social change. It is when the behavioral and cognitive routines of the everyday are disrupted that individuals are likely to be moved to act. Action is particularly likely if the routine of work is disrupted. This is because work provides both monetary sustenance that is needed for survival and dominates the waking hours.

The Great Depression is one rapid social change that severely disrupted the quotidian (Piven and Cloward 1977). In particular, many individuals lost their jobs or livelihoods. As these individuals sought work and could not find it in their communities, many left their communities in search of it elsewhere. The Great Depression vastly affected the quotidian. The key connection between a socially disorganized environment and individuals taking action is whether the anomic conditions truly affect the individuals' lives by disrupting their everyday routines.

Once the quotidian is disrupted, the present world of individuals becomes a problem that begs to be fixed. This disruption makes everyday life unbearably uncertain until such time that the old routines can be reclaimed or new routines can be adopted and adjusted to. Inherently, individuals are spurred to action by the need to get back their old routines, thus reducing the uncertainty in their lives. It should be noted, though, that there may be a threshold for the loss of the quotidian, past which action may be unlikely. This threshold may lie at the complete destruction of the individual's way of life, such as for refugees from wars or victims surviving genocidal campaigns.

How Do Social Disorganization and the Quotidian Relate to One Another?

When Durkheim and the quotidian theorists were writing, they were not trying to explain the occurrence of terrorism. However, these works have been used to explain the occurrence of nonroutine collective action (Useem 1998). Nonroutine collective actions seriously contravene social norms, particularly those against violence. Terrorist attacks constitute nonroutine collective action. The acts themselves are born of the interaction and planning of a group of individuals and involve violence against people or property and, thus, are appropriately described as nonroutine collective action.

What, then, is the process of social disorganization influencing terrorism? First, the rapid social change must occur. It must occur rapidly, and it must have effects that broadly affect the lives of many people in society. An example would be the onset of a civil war. The society must not be able to absorb the negative effects of this change. Individuals will lose their integration into the rhythm and networks of life; they will feel cut off. In addition, they will only weakly be linked to the collective identity of those in their society. These individuals are set adrift.

Second, the social change must influence the everyday lives of people. In fact, it must have substantial effects on the behavioral and cognitive routines of the everyday lives of most people. For example, if a civil war breaks out, perhaps the fighting itself has made it too dangerous for citizens to travel to work every day or perhaps the factory in which they work has been shut down due to damage from nearby bombings. Losing the routine of waking up and going to work every day as well as the thought process that goes along with such a routine may be profoundly disorienting to individuals.

Further, along with the loss of everyday routines, individuals in such societies may not know how the rules and norms of life have changed. What is acceptable and normal may have changed. Both the loss of the routine and the previously known rules and norms have effects on people. With no routine and no norms, these individuals will be poorly integrated into society. They will be included in fewer networks and activities than before the loss of the quotidian; they should also have less identification with the collective identity. The severing of networks and the collective identity and the destruction of the previous norms and rules are what classically constitutes social disorganization. This social disorganization should make it more likely that individuals will do nonroutine collective actions. They may take to the streets to protest or riot, they may commit ordinary crime, and they may engage in political violence. This political violence may be what scholars consider terrorism.

How to Define Terrorism?

The next question must be what constitutes terrorism. Definitions of terrorism abound, and the debate continues on which definition is the best. Often, the definition of terrorism chosen will reflect the main interest of the definer. For example, the United States' Federal Bureau of Investigation, which often conducts domestic terrorism investigations, has a definition of terrorism some have called narrow and legalistic (Martin 2011). This definition is likely precise and legalistic, because it is used to investigate and enforce the law. On the other end of the spectrum, academics and researchers who perform studies of terrorism often use definitions that are broader or more detailed in order to facilitate the collection of data or analysis of data.

No specific definition of terrorism is adopted here, but some guidelines might be useful. These guidelines include non-state actors using unconventional force illegally against civilian or other noncombatant targets for political motives to influence an audience (Martin 2011). Noncombatant targets are often interpreted to include military personnel when off duty. Political motives are often interpreted rather broadly and may include acts of violence that are aimed at instigating social or religious change as well as strictly political motives. In addition, groups can have multiple or overlapping motives in that they may desire political autonomy for their social or ethnic group as well as religious change. An example of this would be the group Hamas, which seeks an independent Palestinian state that would be governed by Islamic law. Further, terrorist groups seek to influence a larger audience beyond the immediate victims of their attacks. For example, the Popular Front for the Liberation of Palestine (PFLP) undertook a campaign of dramatic aerial hijackings for the express purpose of publicizing the plight of the Palestinian people to the world. They chose aerial hijackings in order to grab the world's attention and to direct it to understanding Palestinian grievances. Similarly, the group Black September is thought to have undertaken the attack on the Israeli athletes during the 1972 Olympic Games in Munich, Germany, as a way of airing their grievances to a wider audience.

Is Breakdown a Reasonable Explanation of Terrorism?

In order for the Durkheimian model of societal breakdown to be a reasonable explanation of terrorism, three things should be evident. First, there should be a rapid social change of sufficient magnitude and scope that it cuts individuals off from the social ties and bonds they had to the community before the rapid social change. Second, this loss of ties (breakdown) should stimulate discontent in individuals in that society. This discontent may stem from being cut off from society in general or from the quotidian. Third, it is important to demonstrate whether this discontent must make nonroutine collective action more likely to occur. If all of these qualifications can be met, then breakdown theory has passed a crucial test with respect to whether terrorism is more likely to occur in an environment of social disorganization.

Alternate Hypotheses

It is entirely possible that breakdown theory does not explain terrorism particularly well. What then are the alternate explanations? The major competing explanation is known as resource mobilization and is best exemplified by the work of Charles Tilly and his colleagues. Where breakdown posits that individuals who are cut off from the rules and norms of society by the loss of the quotidian will be more likely to engage in nonroutine collective action, the resource mobilization scholars argue that there is no connection between breakdown and nonroutine collective action. In fact, resource mobilization (RM) scholars do not recognize a distinction between routine and nonroutine collective action; rather, they assert that both routine and nonroutine actions spring from the same mechanisms (Tilly et al. 1975). They also contend that crime and suicide are not collective actions, but instead, they constitute social pathologies and ought not to be considered for explanation by the RM perspective.

Next, RM theorists argue that collective action is explained by dense social networks and a robust collective identity. That is, individuals who are well integrated into the society through their primary and secondary social networks are more likely to engage in collective action. Primary networks are made up of family bonds and the informal social controls provided by this type of network. Secondary social networks are informal social bonds in work, school, and church. Individuals with strong primary and secondary social networks are more likely to engage in collective action according to this perspective. That is, individuals who are well connected to the people in their lives and who identify with the society's collective identity are more likely to be involved in collective action, such as social movements or protests.

More specifically, RM theorists suggest that rather than preexisting social isolation, weak social networks, and weak collective identification, collective action grows out of preexisting social organization. Social organization can include formal and informal social bonds and resources. Individuals who are well integrated into social networks, particularly secondary social networks, are a wellspring of resources, civic-minded attitude, and connections to other people. This type of connectedness makes it *easier* rather than harder to motivate collective action. In fact, individual connections to a whole group of individuals may make bloc mobilization possible, such as mobilizing church members to protest outside of abortion clinics (Useem 1998).

Taking Sides

How, then, to decide which perspective is most useful for explaining terrorism and other forms of crime? This question can be examined using several types of studies. The first type of study would be to interview individuals who have participated in collective action. This would involve collecting their personal information, such as whether they were members of secondary groups. It would also be helpful to ask whether they feel well connected to society or cut off. It would be important to examine those who engaged in routine collective action separately from those who engaged in nonroutine collective action. For example, one set of researchers examined the characteristics of African Americans in 15 cities during the 1960s, a period of great social unrest, including both nonviolent protests and violent race riots (Miller et al. 1976). They compared those who had engaged only in nonviolent protest (routine collective action) to those who expressed attitudes supportive of rioting (nonroutine collective action). They found that those who had engaged in nonviolent protesting were more likely to be older, married, and of higher occupational and educational levels and were less likely to have been raised in "broken" homes than those individuals who had attitudes supportive of rioting. Although many of these individuals may not have participated in the riots, this study may demonstrate that those who engage in routine collective action are more likely to be well integrated in society and less likely to be socially isolated than those who engage in nonroutine collective action. This study supports the breakdown position with respect to the role of social integration and organization.

In contrast, other researchers examined the social movement against drunk driving, known as Mothers Against Drunk Driving (MADD). The movement was founded by two married mothers who had lost daughters to drunk drivers with multiple prior offenses. Interestingly, the movement seemed to take off from the beginning, with MADD's first press conference being held months after its founding on Capitol Hill with three Congressional representatives present. MADD conducts routine collective actions, including tree plantings, gatherings at the US Capitol, and political lobbying, including lobbying US President Ronald Reagan to sign the federal legislation that would effectively raise the drinking age to 21 (madd.org 2011). One study found that on average, MADD participants were socially involved, married, middle aged, college educated, and held privileged jobs (Weed 1987). Further, another study found that MADD chapters were more likely to be founded in counties with wealthier and more educated residents (McCarthy et al. 1988). Given that MADD has only been involved in routine collective actions, this movement, the types of collective actions in which they engage, and the status characteristics of those who participate in it seem to indicate that routine collective actions may require some social organization and participants who are enmeshed in social networks.

The second way of studying the connection between social disorganization and collective action is to examine a particular movement which involved both routine and nonroutine collective actions for evidence that the presence of social disorganization increased the discontent of citizens, whether this discontent made it more likely that these individuals would participate in collective action and, finally, whether those individuals who were well bonded to their families and their communities and participated in community activities before the rapid social change were more likely to engage in collective action (Useem 1980). This case study examined the antibusing movement which grew out of the Supreme Court decision that mandated the racial desegregation of public schools by busing students to racially integrate previously segregated schools. The study involved interviewing individuals who lived in parts of Boston which were likely to be affected by the integration of schools. They were directly asked about whether they had engaged in any collective action, their levels of discontent, and their ties to family and community. This case study directly examined the key questions about breakdown and preexisting community engagement on collective action.

The rapid social change of the antibusing movement was represented by the order to desegregate public schools. The antibusing movement in Boston involved both violent and nonviolent forms of collective action, including boycotts, demonstrations which involved violence, the formation of private schools to circumvent the desegregation order, and the formation of community organizations to represent the antibusing interests. This case study found that those who were closely tied to their communities and those who were members of community organizations were more likely to be discontented with the desegregation decree. That is, socially isolated individuals were less likely to be discontented. This supports the RM position. Those who were discontented, in turn, were more likely to participate in collective action against busing. This supported the breakdown position. It seems possible that those who were more connected to their communities were better informed about the order to desegregate or were more likely to be affected by the order (e.g., had school-aged children) and, thus, had more of a stake in trying to protest or circumvent the desegregation order. This type of rapid social change did not affect all community members equally; those with school-aged children would be more likely to be discontented by it. Nevertheless, this study provided support for both breakdown theory and the resource mobilization theorists.

The third way of studying this question is to examine environments that have endured the effects of rapid social change that resulted in social disorganization. For example, we could also compare the levels of nonroutine collective action in states or countries that have experienced rapid social change and social disorganization to those countries or states that have not experienced rapid social change and social disorganization. This could allow researchers to examine many different countries around the world and over long periods of time. Or alternatively, we could examine the same environment over a long time period, including periods of rapid social change as well as relatively calm periods. In this way, we could examine the effects of rapid social change on collective action while holding the country environment constant.

One study examined patterns of crime and collective action in France between 1830 and 1931 (Lodhi and Tilly 1972). For collective actions, the authors included riots, strikes, and demonstrations. For the rapid social change, they examined the effects of urbanization (i.e., increases or decreases in the proportion of people in France living in cities with 10,000 or more people). Urbanization constitutes a classic rapid social change, because it generally involves the migration of rural residents to the cities. These rural residents often moved to the cities en masse in search of work, because they were unable to sustain a living in farming or because of an increase in the availability of jobs in the cities due to industrialization. The males and occasional whole families who migrated to the cities may be predisposed to difficulties in adjusting to the new urban environment. These adjustment difficulties may have resulted in social disorganization. From the breakdown perspective, these difficulties were likely to involve the isolation of these individuals from their family and community ties, leading them to be cut off from society. This would make them more likely to engage in nonroutine collective action. From the RM perspective, urbanization itself may have no actual effects on the individuals who migrated to the cities. Rather, it is the *experience* of living in a city that makes individuals more prone to collective action. This is called urbanity, which is the proportion of people who live in cities with 10,000 or more people relative to the population of the entire country.

Thus, this research set up a direct test between breakdown and RM. The main question was whether urbanization (breakdown) or urbanity (RM) would drive either crime, collective violence, or both. During the major period of urbanization in France, the number of property crimes declined, while crimes against the person did not appear to be driven by either urbanity or urbanization. These findings most clearly contradict the breakdown perspective, which suggests that urbanization should drive both person and property crimes upward. On the other hand, urbanity seemed to predict property crime and some of the forms of collective violence. That is, the experience of living in a city was more clearly responsible for increases in property crime and some forms of collective action than the social disorganization experienced by a society which has en masse migrations of its citizens to the cities.

Another way of testing which theory better explains how collective action is produced is to study one social environment over a long period of time to assess whether crime and collective action occur during the same periods of time. The RM theorists suggest that crime is not a collective action but rather a social pathology, and thus, it should have no relationship with routine collective action. In contrast, breakdown theorists say that crime and nonroutine collective action ought to occur at similar times, because they are driven by similar processes (social disorganization). However, routine collective action and crime should not occur alongside one another because they are not driven by similar processes.

LaFree and Drass (1997) looked at this question before, during, and after the civil rights movement in the United States. Unfortunately, the study did not separate routine and nonroutine collective actions. The authors compared homicide, robbery, and burglary arrest rates for blacks and whites to collective actions. These collective actions were primarily a result of the African American civil rights movement. Thus, they included such actions as sit-in rallies, marches, boycotts, protests, demonstrations, and civil-rights-related riots.

The idea is to examine whether crime and collective action follow similar patterns over time. If they do, it can be inferred that they are being caused by the same or similar processes, and this would support the breakdown perspective, because if they are predicted by the same processes, then, it is likely that collective action does not require preexisting community ties and membership in collective groups (Lodhi and Tilly 1972). If they do not follow the same pattern over time, increasing and decreasing at similar times, then, it is unlikely that they are driven by the same processes. This type of finding supports the RM model.

Upon examination, crime and collective action did follow a similar path, increasing and decreasing together from 1955 until the early 1970s for both blacks and whites. However, collective action decreased dramatically at this point as the civil rights movement, the rapid social change, came to its conclusion. However, arrest rates stayed high and continued to follow their own path from the early 1970s until 1991. These findings do not clearly support either breakdown or RM. It is possible that if nonroutine collective actions, such as riots, and routine collective actions had been examined separately and each compared to crime that the findings would have been more clearly supportive of one side or the other.

Overall, there is no clear winner in the debate between breakdown and RM. It is likely that both theoretical explanations work to explain different segments of collective action. For example, it is likely that preexisting social organization, such as membership in community groups and having personal ties to the community, makes routine collective action more likely, such as the MADD movement against drunk driving. Individuals who have a stake in the community are probably more likely to act to defend that community using legal and nonviolent methods when they perceive a change, like in the antibusing movement. However, nonroutine collective action likely does not require such organization. Individuals who are more likely to engage in nonroutine collective action, such as collective violence or riots, appear to be more socially isolated individuals who are freed to act in response to the socially disorganized environment which surrounds them. Thus, each of these theoretical explanations appears to explain different domains of collective action.

Is Social Disorganization Felt by Citizens?

It is likely that the social disorganization created by the rapid social change is felt by those who live in the society and are affected by the changing environment. This may be felt as a change of or removal of the norms, a feeling that the old rules no longer apply (Durkheim 1930 [1951]). This may also be felt from the loss of the everyday routine (Snow et al. 1998), which may be experienced as a feeling of restlessness or rootlessness. Further, this may also be felt as impending danger or a loss of a sense of safety (Useem 1998).

In an interesting but simple test of this notion of a feeling of impending danger, Useem (1998) examined the relationship between homicide rates, domestic handgun production, and rioting activity in the United States. The time period was 1964 until 1994, a time which included the African American civil rights movement as well as race-related riots in major cities. The thinking was that handgun production would increase as the handguns previously manufactured were purchased; handguns are not usually used for hunting, unlike long guns. Handgun production rates closely tracked the homicide rate over the time period, except for those years in which there were increases in riots. During these years, handgun production rates jumped up more than would be expected given the danger presented by the homicide rate. Thus, individuals appeared to be frightened by the occurrence of these riots and purchased handguns to protect themselves even if the riot did not occur geographically close to them. This research supports the notion that individuals can feel social disorganization as a sense of impending danger.

Does the Evidence Show a Relationship Between Social Disorganization and Terrorism?

The final question is whether terrorism is more likely to occur in socially disorganized environments. One interesting research study examined this question in the context of breakdown theory (Fahey 2010). In this study, the socially disorganized environment resulted from the occurrence of political instability within the country. The instabilities included the outbreak of war, genocide, adverse regime change, which was a move toward a more autocratic government, and a combination of these events occurring within a short time period. The occurrence of these instability types can be seen as measures of rapid social change. In addition, these types of events are also likely to affect the everyday lives of individuals. If the security situation in a particular country deteriorates rapidly, as in the case of war and genocide, or the political environment has suddenly become far less open, as in the case of adverse regime change, citizens will likely be less able to continue on with the rhythm and routine of their normal everyday life, such as work, leisure, and home time. In extreme cases, they may be forced out of their homes. In addition, the political instability may create a breakdown in the informal and formal social ties that bind individuals to society; the breakdown in the ties may make it more likely that individuals living in a society which has experienced this political instability will participate in nonroutine collective action. This will be due to the idea that the informal and formal ties can no longer function to restrain the actions of these individuals. Further, the loss of the everyday routine and the loss of the controlling power of social ties may work together to increase the likelihood of nonroutine collective action.

Fahey (2010) examined whether political instability was related to the occurrence of terrorism in 147 countries from 1970 to 2005. She found that when political instability occurred within a state, increases in terrorism incidents were more likely. One of the strengths of this study is that it examined the occurrence of instability worldwide over 35 years. This helps to ensure that the results are more stable, that they do not apply only to one country or situation, and that they do not apply to only one decade or time period. Political instability does seem to constitute a rapid social change in the Durkheimian model, and it likely affects the regulatory processes in society, that is, social disorganization, as well as the everyday routines of the citizens in that society.

Conclusions

The purpose of this entry was to discuss and explain the rapid social changes that may lead to social disorganization, the two dominant theoretical hypotheses to explain the association between social disorganization and collective action, and some of the evidence that has been used to support either or both of the theoretical expectations. More evidence is needed to come to a firm conclusion on the relationship between socially disorganized environments and terrorism. However, it is likely that both theoretical explanations have some utility. Specifically, it is likely that some individuals and groups are spurred to terrorism due to the loss of prevailing norms in a newly socially disorganized environment as well as the loss of identification with the collective identity. Further, it is also likely that those who had connections to others, particularly through community groups, before the onset of the social disorganization will be in a better situation to join, recruit, and mobilize for and engage in terrorism through a terrorist group. Bloc mobilization may be particularly important for terrorist groups. What seems to be clear is that social disorganization may be one possible explanation of terrorism incidents. For terrorism that occurs after a rapid social change that produces social disorganization, it is likely that the disorganization provides at least a partial explanation.

Recommended Reading and References

Durkheim E (1930 [1951]) Suicide: a study in sociology translated by George Simpson and John A. Spaulding. The Free Press, New York

Fahey S (2010) State instability and terrorism. Doctoral dissertation, University of Maryland, College Park

Gurr TR (1976) Rogues, rebels and reformers. Sage, Thousand Oaks

Hehir A (2007) The myth of the failed state and the war on terror: a challenge to the conventional wisdom. J Int Statebuilding 1:307–332

Helman GB, Ratner S (1992) Saving failed states. Foreign Policy 89:3–20

LaFree G, Dugan L, Fahey S (2008) Global terrorism and failed states. In: Hewitt JJ, Wilkenfeld J, Gurr TR (eds) Peace and conflict. Paradigm, Boulder

LaFree G, Drass K (1997) African American collective action and crime, 1955–1991. Soc Forces 75:835–853

Lodhi AQ, Tilly C (1972) Urbanization, crime and collective violence in 19th century France. Am J Sociol 79:296–318

Madd.org (2011) 25 years of saving lives, by Laurie Davies http://www.madd.org/about-us/history/madd25thhistory.pdf. Accessed 21 July 2002

Martin G (2011) Essentials of terrorism: concepts and controversies, 2nd edn. Sage, Thousand Oaks

McCarthy JD, Wolfson M, Baker DP, Mosakowski E (1988) The founding of social movement organizations: local citizens' groups opposing drunk driving. In: Caroll GR (ed) Ecological models of organizations. Ballinger, Cambridge, MA, pp 71–84

Miller AH, Bolce LH, Halligan MR (1976) The new urban blacks. Ethnicity 3:338–367

Newman E (2007) Weak states, state failure and terrorism. Terror Polit Violenc 19:463–488

Piazza J (2007) Draining the swamp: democracy promotion, state failure, and terrorism in 19 Middle Eastern countries. Stud Conflict Terrorism 30:521–539

Piazza J (2008) Incubators of terror: do failed and failing states promote transnational terrorism? Int Stud Q 52:469–488

Piven FF, Cloward R (1977) Poor people's movements. Why they succeed, how they fail. Pantheon, New York

Snow DA, Cress DM, Downey L, Jones AW (1998) Disrupting the quotidian: reconceptualizing the relationship between breakdown and the emergence of collective action. Mobilization 3:1–20

Tikusis P (2009) On the relationship between weak states and terrorism. Behav Sci Terror Polit Aggress 1:66–78

Tilly C, Tilly L, Tilly R (1975) The rebellious century, 1830 to 1930. Harvard University Press, Cambridge, MA

Useem B (1980) Solidarity model, breakdown model and the Boston anti-busing movement. Am Sociol Rev 45:357–369

Useem B (1998) Breakdown theories of collective action. Annu Rev Sociol 24:215–238

Von Hippel K (2002) The roots of terrorism: probing the myths. Polit Quart 73:1–39

Weed FJ (1987) Grass roots activism and drunk driving issues: a survey of MADD chapters. Law Policy 31:259–278

Social Factors and Third Party Policing

Daniel Brown-Kenyon
Institute for Social Science Research, Policing and Security, The University of Queensland, St. Lucia, QLD, Australia

Overview

The Third-Party Policing model of crime control advocates for the use of partnerships between the police and other key agents of social control – for example, government, business, and community sectors – however, to date, almost no research has examined the social factors that contribute to the success or failure of these partnerships. This entry utilizes a Social Identity approach to analyze some of the key factors that affect the efficacy of these partnerships. Specifically, partnerships are broken down into three distinct phases, and various social factors are considered at each of these phases. This entry also discusses some of the organizational level variables that will be affected if these partnerships are to succeed.

Introduction

Third-Party Policing (TPP) is an approach to dealing with crime problems where police form partnerships with other public service agencies, business, nongovernment organizations, and the community to deliver more effective crime control strategies (Mazerolle and Ransley 2006). The TPP approach requires identification of "legal" mechanisms that are the domain of other parties to assist police in managing public order. Mazerolle and Ransley (2006) identify a wide range of legal levers used in a variety of different settings, such as partnerships with local councils, business, health and safety authorities, and other law enforcement agencies. The primary aim of the TPP approach is to deliver a more effective police service through a twofold approach: recruiting suitable crime control partners and making use of legal levers otherwise unavailable to the police.

This contribution aims to advance the theoretical understanding of TPP by outlining current research and theory on intergroup relations and applying this to policing partnerships. TPP advocates that partnerships are vital for the success of modern policing practices. However, to date, little to no research has been conducted within the policing arena to examine these partnerships and what drives their success and failures. Drawing on the Social Identity literature (see Brown 2000; Haslam 2004), this entry examines some of the social and organizational factors that interact with the success or failure of these partnerships. This analysis is undertaken in five sections, which are comprised of two conceptually distinct areas, focusing on the social aspects of partnerships and then outlining the organizational factors that affected partnerships. The first section presents an overview of a Social Identity approach to intergroup process. The next section separates partnerships into three core stages, including partnership seeking, partnership development, and effective partnership functioning. Next, the role of information sharing in partnership development and functioning is addressed. Finally, some of the organizational constraints within police organizations that are likely to threaten the success of a partnership approach are addressed. At the end of each section, the arguments are summarized and important relationships between the key factors and likely outcomes are discussed.

A Social Identity Approach to Intergroup Process

Social Identity theory is the preeminent theory of intergroup relations. Social Identity theory explores the social motivations and consequences of group behavior. Specifically, Social Identity theory provides a context for examining individuals' behavior within groups (Haslam 2004). According to Social Identity theory, individuals possess social representations of themselves, which are comprised of the social networks they maintain, the groups to which they belong, and the organizations for which they work (Ellemers et al. 2002). Social Identity theory argues that social identity is a critical component of self-concept, and as such, individuals seek to view their group memberships in a positive frame. This is achieved through negative comparisons between their own group and other groups. Through these comparisons, individuals are able to boost their self-esteem (Brown 2000) and show preferential treatment to those of their own groups (Brown 2000). Social Identity theory provides a meta-level context to analyze and understand a range of phenomena involved in intergroup processes.

Since the conception of Social Identity theory, there has been substantial development with a range of subtheories having been developed to explain some of the more nuanced aspects of group behavior. This entry applies a Social Identity approach to understanding the three hypothesized stages of partnerships. The first stage, termed "partnership seeking," helps to understand the internal group structure of the police organization and how these social factors impact on partnership-seeking behavior. Here, key factors of cohesion, leadership, and the norms and values within the organization are all likely to affect the willingness of individuals to seek out partnerships. In the next phase, termed "partnership formation," the core factors of diversity, distinctiveness, and identification during the formation of a new partnership are considered. In the final stage, "partnership functioning," intragroup processes and their implication on cooperation are considered. Here, group identification, procedural justice, and leadership and the impact of these factors on cooperation within a newly formed partnership are discussed.

Partnership Seeking and the Role of Social Identity Theory

This section uses Social Identity theory to examine three internal social factors that police organizations face when seeking TPP partnerships. First, the way that norms and attitudes of a police organization are likely to dictate individual officers' intentions to seek out third-party partnerships is examined. Within this, the role of self-categorization and its impact on conformity with these norms are also outlined. Second, the role of leadership within a police organization in promoting partnership seeking and development of partnerships within the context of TPP is discussed. Finally, this section explores the effect of internal cohesion and how this poses a significant barrier for the police to effectively seek out partnership.

The development of self-categorization theory is an extension of Social Identity research and is particularly useful for understanding partnership-seeking behavior (Hogg and Terry 2000). Self-categorization theory stipulates that individuals categorize their social world by the groups to which they belong, through a process of depersonalization, in which people are reduced to a prototype of the group, thereby embodying the norms, values, and behaviors that personify the group (Hogg and Terry 2000). It is through this process of self-categorization that cognitive groups are formed. Depersonalization results in cognitive perceptions of similarity between individuals and their group. These effects are emphasized for individuals of out-groups, such that when they are perceived as a prototype of the group, the individual differences between group members are reduced and the similarities are enhanced. The outcome of this process is that individuals are no longer perceived as unique, rather, they are perceived to be a representative of their group (Hogg and Terry 2000).

Police are often faced with situations in which they categorize people into groups. This process allows the police to distinguish all members of a group – for example, a child protection agency – as being conceptually the same. Furthermore, the effect of self-categorization will be more salient. That is, the officer's identity as a police officer will be more powerful, when officers are faced with the possibility of sharing aspects of their work that makes them unique, such as crime control. Evidence suggests that a critical component of group discrimination is the dimensions that the groups are compared on (Ashforth and Mael 1989). As managing law and order is a key dimension distinguishing police from other professionals, contexts requiring police to share this role with others will likely result in increased categorization of those others. This effect is likely to result in the police taking steps to demonstrate the inability of their partners to fulfill the role that they have been assigned. Empirical research supports this position, with results consistently demonstrating that self-categorization effects are most pronounced in an intergroup context (Ashforth and Mael 1989; Doosje et al. 2002). An example for policing may include a collaborative taskforce where the police share the responsibilities with a partner, such as a child protective service, who has the same power to act and responsibilities as the police. Therefore, it is possible the police would attempt to exasperate the difference between the two organizations with an emphasis on the positive aspects of police and the negative aspects of the child protection agency. This would likely result in the police demonstrating reluctance to share their responsibilities with this other agency through a distrust of their ability to deliver the required services.

Research also indicates that the degree of identification felt toward an organization is a critical factor in conformity to group norms, attitudes, and behaviors (Tyler and Blader 2001). As individuals seek to enhance both their own status within the group as well as the overall standing of the group, they are highly motivated to perform in a manner that increases their standing in both domains. Conforming to norms and attitudes is one method to achieve this. In doing so, individuals represent the prototype of the group, while demonstrating active commitment to the group. Conversely, individuals face sanctions for failing to conform. Therefore, the degree of self-categorization and meaningful value that an officer places on their position within the organization will also be predictive of their willingness to conform to the organizational rules, norms, and behaviors.

This suggests that the organizational norms and attitudes of police organizations need to foster partnership-seeking behavior if a police agency is to be successful in institutionalizing TPP. Police organizations may need to undergo a cultural change, with a strong focus on the involvement of partnerships in crime control innovation. In these situations, the self-categorization process will continue to reinforce these norms and attitudes, thereby ensuring that new police officers will also engage in this behavior.

Self-categorization theory argues that group leaders are well positioned to change the culture and develop new attitudes. According to self-categorization theory, leaders are perceived to be members of a group that are *most* prototypical (Hogg 2001), the individual(s) that embody the values, norms, and attitudes of the group (Hogg 2001). Furthermore, self-categorization theory argues that leaders have the consent of their followers to change the cultural attitudes and norms of a group. Leaders within an organization hold the unique position to install change and create a culture of active partnership formation. They hold unique positions where it is legitimate for them to deviate from the accepted cultural norms to install change. Research has demonstrated this critical role of leaders in conveying of norms (Hogg et al. 2006). Likewise, research has also consistently demonstrated the importance of leaders in fostering innovation and supporting change. Therefore, the onus is on police leaderships to create and foster norms and attitudes that will promote active partnership seeking and development among the junior ranks.

An outcome of highly salient group identity and strong leadership can lead to high levels of cohesion within groups (Hogg 1993). Cohesion is defined as the social connections that tie individuals to their fellow group members, resulting in their desire to behave and act in a manner consistent with the group (Casey-Campbell and Martens 2009). Research has extensively examined cohesive groups, with an emphasis on the cohesion-performance relationship. There is mixed evidence supporting the cohesion-performance relationship. However, there is general support for the notion that highly cohesive groups perform better on tasks, particularly in project-based team environments (Chiocchio and Essiembre 2009).

The nature of police work lends itself to high levels of internal cohesion. The police are required to rely on each other for safety and to work together to effectively achieve outcomes. Often they are required to sacrifice their own safety and wellbeing for the benefit of society. Anecdotally, police organizations are akin to military organizations, where there is a cultural of solidarity and unity. However, when considering a TPP initiative, highly cohesive police organizations pose a significant constraint on the take-up of TPP partnerships. Research indicates that highly cohesive groups are more reluctant to engage with external groups in intergroup contexts (Hardy et al. 2005). Therefore, excessively cohesive police organizations are likely to prevent officers from seeking partnerships.

It is possible that highly cohesive organizations will attempt to develop internal solutions, instead of seeking partnerships. This results in a continued drain on resources, without commiserate efficiency gains. When advocating for TPP partnerships, the role of cohesion and its effects on the success of new initiatives may need to be considered. Evidence for consideration of these effects can be seen from previous investigations into policing practices. For example, one of the key recommendations from an inquiry into police corruption in Australia was that for new recruits to be educated to a minimum of a tertiary level in an attempt to reduce the insular nature of policing (Fitzgerald 1989). This recommendation was made in an attempt to prevent the effects of a highly cohesive organization perpetuating bad practice.

This section has highlighted some of the internal social factors that police organizations face when seeking TPP partnerships. From the research outlined, the following relationships are proposed. First, the norms and attitudes of a police organization are likely to affect officer's intentions to seek out third-party partnerships. Furthermore, it is possible that this relationship will be affected by the role of self-categorization, such that those who are highly identified may demonstrate greater conformity to these norms. Secondly, it is the role of the leaders within an organization to promote active search and development of partnerships. Finally, high levels of cohesion in police organizations are likely to impede effective partnership seeking.

Partnership Formation: Two Organizations Merging and the Role of Diversity and Distinctiveness

Organizational mergers have long been the topic of interest for a range of business and academic communities. This research has largely focused on the outcomes of a newly formed organization (Terry and O'Brien 2001); however, recent research has also considered the social factors that contribute to the success or failure of partnerships (Terry et al. 2001). Policing partnerships are akin to other types of organizational mergers. They require at least two different organizations to work toward a unified goal and deliver a product – in this case public safety. Research by Terry et al. (2001) has investigated the role of social identity during the formation of a new superordinate organization after a merger. This research indicates a range of social factors that should be considered when forming a TPP partnership. Primarily, during the initial merger, the higher-status group members react more negatively toward the merger than lower-status group members (Terry et al. 2001). The implication of this finding is critical to understanding the police motivations to form a partnership. It is arguable that the police will hold the higher-status position within most new partnerships; they will typically be the organization that initiates the partnership and the organization that has the most to gain.

The inverse of the status relationship is also true. There is evidence to suggest that individuals from lower-status organizations are more willing to join higher-status organizations (Terry et al. 2001). However, some results have found that the effect of status can be moderated by legitimacy, such that, if lower-status organizations perceive their status as illegitimate, they will be less likely to join the higher-status organization (Amiot et al. 2007). When considering these findings in the context of a TPP partnership, it is clear that steps need to be taken to mitigate these effects in order for these diverse organizations to work together. This indicates that the police need to ensure that they do not portray themselves as superior to their partner organization as well as approach their partners in a collaborative and transparent manner that avoids invoking perceptions of status differences between the two groups.

Research indicates that lower-status organizations have higher staff turnover rates after merger (Amiot et al. 2007). This research also suggests that lower-status group members have less identification with new organization (Amiot et al. 2007). Together, these two findings are vital to the understanding of how partnerships function. As highlighted in the first section, the key to ensuring that individuals perform for the group, and adhere to the group norms of behavior and attitudes, is increasing the identification that they feel toward the group. Therefore, if lower-status individuals feel less identification toward the partnership, they are less likely to remain in the partnership. If the police believe that they are a higher-status organization and their key partners – such as child protection agencies – are a lower-status organization, then this is likely to result in initial reluctance from the police to form a partnership. However, should this be overcome, issues of status legitimacy, identification with the partnership, and turnover rates are all likely to be significant challenges for both parties.

When a new partnership is formed, the police and their partners come together under the banner of a new group. This partnership is formed to provide a more efficient service and reduce the resource burden from one agency alone.

However, there is likely to be historical issues of status, prestige, and potentially conflict that need to be managed. Furthermore, issues of group diversity pose significant problems to the formation of these partnerships. Diversity within organizations has long been a topic of significant interest to the academic and business community (Brewer 1995). In the context of a TPP partnership, it is important to consider diversity within the context of police and the partner agencies. These partnerships involve individuals from different groups, with different experiences/skills and different organizational cultures. Furthermore, these individuals will come from organizations and skill backgrounds, which have different underlying philosophies and ideologies about how to manage complex crime problems. The research outlines the role of different approaches to problems and the impact of these cultural ideologies on the success of group performance.

Diverse partnerships have the potential to foster conformity to organizational norms and rules, as these become the superordinate guidelines to behavior within the given context (Rink and Ellemers 2007). To enable this, organizational identities need to be highly salient and provide significant emotional value to its members. In contrast, in partnerships where there is a history of conflict and highly charged emotions, the organizational identity will be weak, thereby making the subgroup identities highly salient (Tsui et al. 1992). This results in reduced compliance to the norms and rules of the partnership and a return to the norms and values of their originating organizations (Rink and Ellemers 2007). Forming a successful TPP partnership requires that all parties are aware of organizational diversity both within and across the partnering organizations. Historical conflict between partners poses a significant constraint on the success of these partnerships. Furthermore, by identifying common goals and objectives and reinforcing the mutual commitment of the organizations, it is possible to overcome these issues.

Research in the area of optimal distinctiveness theory helps to explain the conditions under which partnership identification is achieved. Optimal distinctiveness theory (Brewer 1991) argues that we have two competing drives in social groups – one for inclusion and one for distinctiveness – and that these two drives act in opposition to each other. In the context of a new partnership, individuals will want to be included in the partnership, to feel they are a valued member and that they are able to contribute to the development and direction of the partnership. However, individuals will simultaneously want to maintain connections with their original organization. Therefore, the successful formation of a new partnership is contingent on both of these needs being met at an optimal level (Brewer 1995).

For the police to form a partnership with key partners, a significant risk arises from these organizations being identified under the banner of a large and ambiguous group, typical of “whole of government approaches,” as these groups remove the unique contributions of the individual organizations that comprise the partnership. Likewise, the assimilation of groups under the higher-status group, in this situation the police, poses the same risks. These partnerships need to develop identities that are unbiased and do not represent just one group, yet attempt to encompass the root values and goals of the partnership. Research shows that group size and overly inclusive groups are both negatively related to performance (Badea et al. 2010). According to the optimal distinctiveness literature, this is likely to result in poor functioning and lack of identification at the superordinate level (Badea et al. 2010) and likely to result in high turnover or a quick end to the partnership. Therefore, these partnerships should avoid the use of group structures that inhibit the unique contribution of each member organization.

This section highlighted the role of several key factors that inform the process and the likelihood of police forming partnerships within the context of TPP. The role of group diversity, distinctiveness, and partnership identity have all been examined and discussed. Next several relationships between these variables are examined. First, partnerships where status difference between the police and a partner organization is highly salient are likely to result in poor intergroup cooperation and weak identification with the partnership. Secondly, diversity of organizational identities in the partnerships will impact on partnership functioning. Specifically, partnerships with historical conflict and competition for resources will not function effectively as a partnership; this effect may be exasperated by differences in cultural ideology about how to approach problems. Finally, the success of a new partnership developing is contingent on the ability for individuals to retain connection with their originating organization.

Promoting Cooperation Within a Partnership: The Role of Identification, Procedurally Fair Decision Making, and Leadership

TPP partnerships require the active and continued cooperation of its members. Significant research has explored the motivations that promote cooperation within groups, with a specific focus on information sharing, extra-role behavior, and ensuring that the long-term goals of the organization are met over the temptation of short-term individual gains. Recently, Tyler (2011) examined the relationship between instrumental factors such as material rewards and role titles, and social factors such as identification and procedurally fair decision making, in fostering cooperation of employees with the organization. The results consistently demonstrated that above and beyond instrumental factors, social motivation was found to be critical in adherence to organizational rules and norms, extra-role behavior, and intragroup cooperation (Tyler 2011). Extra-role behavior is defined as undertaking work and responsibilities that are outside the individual's job description for which they do not receive direct momentary reward.

In this section two critical aspects of cooperation within partnerships are highlighted: identification with the partnership and procedural fairness of decision making within the partnership. As discussed in the earlier sections, when an individual categorizes themselves as a member of an organization, this will lead to increased adherence to the norms and attitudes of the organization (Tyler and Blader 2001). Therefore, increasing the degree of social identification with the partnership will lead to an individual feeling greater emotional significance about their connection with the group, thereby increasing their drive to enhance the group relative status. TPP partnerships require a significant commitment from both the individuals and the groups connected to them. This commitment can range from undertaking activities that are beyond the typical duties of the role they fulfill to providing extra resources to ensure the success of the partnership. By increasing the degree of identification to the partnership, there is potentially an increase in the adherence toward the goals and aims of the partnership. Furthermore, this is also likely to result in the members undertaking extra-role behavior when faced with impediments to ensure the success of the partnership (Tyler 2011).

According to group engagement model (Tyler and Blader 2003), identification with the organization is enhanced by the decision-making processes that the organization uses. Specifically, the use of procedurally fair processes to arrive at decisions increases the strength of identification that individuals feel toward their organization. Research demonstrates a link between the procedural fairness of an organization and the behavior of its employees (Blader and Tyler 2009). Procedural fairness can be described as the processes and methods that an individual or group enacts to arrive at a decision as being fair (Blader and Tyler 2003). There is consistent evidence within a range of fields that demonstrate the link between procedural fairness and cooperation, acceptance of outcomes, and extra-role behavior (Tyler and Blader 2001). Furthermore, evidence suggests that people are more concerned with the procedural fairness of a decision rather than the outcome of a decision alone. The goal of a new partnership, therefore, is to behave in a procedurally fair manner, as perceived by all members of the group. This is achieved through unbiased decision making, offering all members a voice before a decision is made, making the decision on facts rather than opinions, and using clear and transparent processes. Therefore, the relationship between procedural justice and cooperation in the group is mediated through identification with the group (Tyler and Blader 2003).

For TPP partnerships, this finding implies that it is possible to increase the degree of identification that the individual members feel toward the partnership, by utilizing procedurally fair decision-making process.

The final aspect of effective partnership functioning relates to leadership. As discussed earlier, the role of a leader can be of critical importance in the conveying of norms, attitudes, and appropriate behavior (Hogg 2001). When a leader behaves in a manner that embodies the values of the group and is seen as the most representative member of the group, then it is likely that they will be able to gain the trust and confidence of their followers (De Cremer and Van Knippenberg 2002). Likewise, there is sufficient evidence to suggest that a leader that acts in a procedurally fair manner is also likely to enhance their subordinates' willingness to accept decisions (De Cremer and Van Knippenberg 2002). Extending this, research has indicated that leaders play a vital role in the development and fostering of innovative practices (Hogg et al. 2006). Therefore, it is vital that upon formation of the partnership, a leader who will drive the strategic direction, yet encompass the requirements outlined above, is identified. As outlined in earlier sections, leaders play a vital role in the success of any venture; they provide the strategic direction (Hogg et al. 2006) and can guide the subordinates through challenging times. The success of new partnerships hinges on the effectiveness of the leader.

Three critical aspects of a partnership define its functioning. First, individuals who are part of the partnership must demonstrate high levels of identification with the new partnership, as doing so will ensure compliance and cooperation with the partnership objectives. Second, in order to increase identification, the partnership must demonstrate procedurally fair processes in its decision making, with a particular emphasis on demonstrating collegial and collaborative approach and avoiding unilateral decision-making processes. Finally, the leadership of the partnership must represent all parties within the partnership and actively engage in procedurally fair process.

Organizational Constraints of Third-Party Policing

In their analysis of TPP approaches to policing complex problems, Mazerolle and Ransley (2005) outline that police organizations face organizations' constraints that are likely to impact on the ability of the police to form partnerships. When advocating for partnerships, these factors need to be considered in conjunction with the social factors outlined above. One of the most crucial factors that require consideration relates to the ability of organizations and government to effectively share information. This is vital, as without sharing of intelligence, methods, and policies, the ability of partner organizations to coordinate the delivery of their services becomes restricted. Furthermore, it may lead to redundancy and duplication of work or frustration for lack of action from partners.

The capacity to share information between partner organizations is one of the most complex factors to be overcome to facilitate a productive working partnership. When attempting to find innovative solutions to problems, one of the most important tools that can be employed is sharing information across organizational boundaries. However, when government agencies are involved in these partnerships, the transfer of information between partners can pose significant problems. Most Western governments have laws that restrict the transfer of information (Bellman et al. 2004). These restrictions typically impose conditions that prevent the transfer of information between government agencies. These laws are designed to protect the interests and privacy of individuals within the community. However, they can also prevent the effective coordination and cooperation of core agencies. Therefore, one of the most significant structural challenges that these partnerships face is the need to overcome information sharing hurdles (Lyons 2002). It is important to consider what information needs to be shared, how the information will be used, and who, under relevant laws, is able to hold that information. Addressing these considerations early in the formation of a partnership is critical to the success of a partnership.

Mazerolle and Ransley (2005) identify a number of organizational factors that either facilitate or constrain the adoption of Third-Party Policing. One of the most significant challenges that police agencies face in forming partnerships is the centralization of decision making. They argue that the concentration of decision-making power with the executive is not conducive to the formation of partnerships. The crux of their argument is that officers at all levels of the organization need to have the freedom to identify problems and innovate solutions, and doing so can only occur effectively when there has been a decentralization of decision making (Mazerolle and Ransley 2006). Furthermore, decentralization of decision making offers unique advantages to geographically dispersed police agencies. For example, issues that affect police forces in large metropolitan areas are likely to be significantly different from those that affect police officers in rural or remote areas. Therefore, officers need to be able to identify local solutions to local problems. This argument is not unique to TPP advocates. Problem-oriented policing, arguably one of the most influential developments in policing practice, has advocated for support from the senior executive for officers to implement localized strategies to complex crime problems (Goldstein 2003).

Finally, Mazerolle and Ransley (2006) outline that role specialization has a significant impact on the ability of officers to form partnerships, stating that more specialized roles are conducive to the formation of specific partnerships than roles that are more general. This argument follows the notion that officers and personnel that fulfill highly specialized roles within an organization will have greater insight into the needs from the partners and the ability of the partners to engage in a cooperative approach. Extending this argument, the seniority of officers will also have an impact on their ability to develop and formalize partnerships, with senior officers having the ability to implement formal cross-organizational partnerships and junior officers seeking ad hoc episodic partnerships. Therefore, identification of the right personnel within the police organization and gaining the support of the senior executive are factors that may well contribute to the success of potential partnerships.

Conclusion

In an era of social change in the way that the police conduct their duties, there is inherent value in a TPP approach to crime control. Through the use of effective partnerships, the police are able to provide a more efficient and effective service that has the potential to develop more holistic outcomes for the targets. This entry examines the partnership aspect of TPP using Social Identity theory to explore the intergroup processes that affect the ability of the police to seek out and develop crime control partnerships. Three critical phases comprise the process of partnership building within TPP: partnership seeking, partnership formation, and partnership effectiveness.

In the partnership-seeking phase, research outlined the role of the internal social structure of the police organization, with an emphasis on the norms and attitudes, to ensure that they are focusing on the ultimate goal of the organization. In conjunction with this, the role of internal cohesion within a police force was considered, with a particular emphasis focusing on what impact this has on the ability and desire of the organization to seek out partnerships. Research also examined the role of leadership in shaping the future direction of police agencies, with an eye toward fostering partnership-seeking behavior. In the formation phase of TPP partnerships, the research outlined indicates that through the consideration of the diversity of ideological backgrounds and organizational culture, it is possible to improve working relationships. Extending this, allowing individuals to join and have meaningful input into the future directions of a partnership while still allowing them to maintain a strong connection with the organization where they originated is another avenue to success. In the final stage, partnership functioning, the research outlined above argues that it is critical to help the individual form a sense of identity with the partnership in order to foster cooperation. It is possible to achieve this with procedurally fair decision-making process, with a strong emphasis on the use of unbiased, fair, and neutral process. Engaging and developing a strong and consistent leader is also important in fostering a strong sense of identity with the new partnership.

While this entry has largely focused on the importance of social factors in the formation and functioning of a partnership, there are also instrumental factors that contribute to the success and failure of partnerships. Information sharing is a significant challenge to the way that partnerships can function, with the inability of some information to be shared between partners potentially leading to duplication of work and frustration with partners. Likewise, the decision-making ability of individuals within an organization represents challenges for partnerships. In situations where all decisions need to be passed through the senior executive, this has the potential to limit the ability of officers to seek and contribute to partnerships. Finally, officer specialization is likely to impact on the quality and type of partnership, with more specialized officers having greater insight in the needs of their own organization in respect to specific issues but also insight into how partners function their ability to contribute to a coordinated effort.

The issue of partnership formation is not one that is solely the responsibility of the police. All key stakeholders, including government, policy makers, communities, and businesses have a role in overcoming the challenges faced in addressing complex crime. The police face many challenges in attempting to develop their practices in a new era of policing; therefore, drawing attention to these challenges provides opportunities for further research and input from policy makers and practitioners alike. If the police and society more generally are going to deliver a more effective service, then it is incumbent on the community to work collaboratively to develop solutions. As we learn more about how and why crime is committed, it is no longer feasible to assume that insular police organizations will be able to recruit and retain the knowledge, skills, and resources required to combat future criminal behavior.

Related Entries

- ▶ Civil Remedies
- ▶ Communities and the Police
- ▶ Communities That Care
- ▶ Community Policing
- ▶ Legal Frameworks for Third-Party Policing
- ▶ Procedural Justice and Cooperation
- ▶ Third Party Policing and School Truancy

Recommended Reading and References

Abrams D, Hogg MA (2004) Metatheory: lessons from social identity research. Pers Soc Psychol Rev 8(2):98–106. doi:10.1207/s15327957pspr0802_2

Amiot CE, Terry DJ, Callan VJ (2007) Status, equity and social identification during an intergroup merger: a longitudinal study. Br J Soc Psychol 46(3):557–577. doi:10.1348/014466606x146015

Ashforth BE, Mael F (1989) Social identity theory and the organization. Acad Manage Rev 14(1):20–39

Badea C, Jetten J, Czukor G, Askevis-Leherpeux F (2010) The bases of identification: when optimal distinctiveness needs face social identity threat. Br J Soc Psychol 49(1):21–41. doi:10.1348/000712608x397665

Bass BM (1997) Does the transactional–transformational leadership paradigm transcend organizational and national boundaries? Am Psychol 52(2):130–139. doi:10.1037/0003-066x.52.2.130

Bellman S, Johnson EJ, Kobrin SJ, Lohse GL (2004) International differences in information privacy concerns: a global survey of consumers. Inf Soc 20(5):313–324. doi:10.1080/01972240490507956

Blader SL, Tyler TR (2003) A four-component model of procedural justice: defining the meaning of a “Fair” process. Pers Soc Psychol Bull 29(6):747–758. doi:10.1177/0146167203029006007

Blader SL, Tyler TR (2009) Testing and extending the group engagement model: linkages between social identity, procedural justice, economic outcomes, and extrarole behavior. J Appl Psychol 94(2):445–464. doi:10.1037/a0013935

Brewer MB (1991) The social self: on being the same and different at the same time. Pers Soc Psychol Bull 17(5):475–482. doi:10.1177/0146167291175001

Brewer MB (1995) Managing diversity: the role of social identities. American Psychological Association, Washington, DC

Brewer MB (1996) When contact is not enough: social identity and intergroup cooperation. Int J Intercult Relat 20 (3–4):291–303. doi:10.1016/0147-1767(96)00020-x

Brown R (2000) Social identity theory: past achievements, current problems and future challenges. Eur J Soc Psychol 30(6):745–778. doi:10.1002/1099-0992(200011/12)30:6<745::aid-ejsp24>3.0.co;2-o

Casey-Campbell M, Martens ML (2009) Sticking it all together: a critical assessment of the group

cohesion–performance literature. Int J Manag Rev 11(2):223–246. doi:10.1111/j.1468-2370.2008.00239.x
Chiocchio FO, Essiembre H (2009) Cohesion and performance. Small Group Res 40(4):382–420. doi:10.1177/1046496409335103
De Cremer D, Leonardelli GJ (2003) Cooperation in social dilemmas and the need to belong: the moderating effect of group size. Group Dyn Theory Res Pract 7(2):168–174. doi:10.1037/1089-2699.7.2.168
De Cremer D, Van Knippenberg D (2002) How do leaders promote cooperation? The effects of charisma and procedural fairness. J Appl Psychol 87(5):858–866. doi:10.1037/0021-9010.87.5.858
Doosje B, Spears R, Ellemers N (2002) Social identity as both cause and effect: the development of group identification in response to anticipated and actual changes in the intergroup status hierarchy. Br J Soc Psychol 41(1):57–76. doi:10.1348/014466602165054
Ellemers N, Spears R, Doosje B (2002) Self and social identity. Annu Rev Psychol 53(1):161–186. doi:10.1146/annurev.psych.53.100901.135228
Fitzgerald GE (1989) Report of a commission of inquiry pursuant to orders in council. Queensland Government, Brisbane
Goldstein H (2003) On further developing problem-oriented policing: the most critical need, the major impediments, and a proposal. Crime Prev Stud 15:13–47
Hardy J, Eys MA, Carron AV (2005) Exploring the potential disadvantages of high cohesion in sports teams. Small Group Res 36(2):166–187. doi:10.1177/1046496404266715
Haslam SA (2004) Psychology in organizations: the social identity perspective. Sage, London
Hogg MA (1993) Group cohesiveness: a critical review and some new directions. Eur Rev Soc Psychol 4(1):85–111. doi:10.1080/14792779343000031
Hogg MA (2001) A social identity theory of leadership. Pers Soc Psychol Rev 5(3):184–200. doi:10.1207/s15327957pspr0503_1
Hogg MA, Terry DJ (2000) Social identity and self-categorization processes in organizational contexts. Acad Manage Rev 25(1):121–140
Hogg MA, Fielding KS, Johnson D, Masser B, Russell E, Svensson A (2006) Demographic category membership and leadership in small groups: a social identity analysis. Leadersh Q 17(4):335–350. doi:10.1016/j.leaqua.2006.04.007
Lyons W (2002) Partnerships, information and public safety: community policing in a time of terror. Polic Int J Police Strateg Manag 25(3):530–542
Mazerolle L, Ransley J (2005) Third Party Policing. Cambridge University Press, New York
Mazerolle L, Ransley J (2006) Third party policing. Cambridge University Press, Cambridge, UK
Rink F, Ellemers N (2007) Diversity as a basis for shared organizational identity: the norm congruity principle. Br J Manag 18:S17–S27. doi:10.1111/j.1467-8551.2007.00523.x
Tajfel H, Turner JC (1986) The social identity theory of intergroup behavior. In: Worchel S, Austin WG (eds) Psychology of intergroup relations, 2nd edn. Nelson-Hall, Chicago, pp 7–24
Terry DJ (2000) Inter-group relations and organizational mergers. In: Hogg MA, Terry DJ (eds) Social identity process in organizational contexts. Taylor & Francis, New York
Terry DJ, O'Brien AT (2001) Status, legitimacy, and ingroup bias in the context of an organizational merger. Group Process Intergroup Relat 4(3):271–289. doi:10.1177/1368430201004003007
Terry DJ, Carey CJ, Callan VJ (2001) Employee adjustment to an organizational merger: an intergroup perspective. Pers Soc Psychol Bull 27(3):267–280. doi:10.1177/0146167201273001
Tsui AS, Egan TD, Iii CAOR (1992) Being different: relational demography and organizational attachment. Adm Sci Q 37(4):549–579
Turner JC, Hogg MA, Oakes P, Reicher SD, Wetherell MS (1987) Rediscovering the social group: a self-categorization theory. Blackwell, Oxford, UK/Basil
Tyler TR (1989) The psychology of procedural justice: a test of the group-value model. J Pers Soc Psychol 57(5):830–838. doi:10.1037/0022-3514.57.5.830
Tyler TR, Blader SL (2001) Identity and cooperative behavior in groups. Group Process Intergroup Relat 4(3):207–226. doi:10.1177/1368430201004003003
Tyler TR, Blader SL (2003) The group engagement model: procedural justice, social identity, and cooperative behavior. Pers Soc Psychol Rev 7(4):349–361. doi:10.1207/s15327957pspr0704_07
Tyler T (2011) Why people cooperate: The role of social motivations. Princeton University Press. Princeton

Social Harms

► International Responses to Victims in Criminal Justice

Social Learning Theory

Kyle J. Thomas and Sally S. Simpson
Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA

Synonyms

Behavioral learning; Differential association-reinforcement theory; Learning theory

Overview

In their differential association-reinforcement theory, Burgess and Akers (1966) reformulated, and restated, Edwin Sutherland's principles of differential association (1947) to include the learning principles of modern behavioral learning. Sutherland's theory had been roundly criticized on two main grounds: (1) its failure to make explicit the precise mechanisms of learning and (2) difficulties operationalizing and empirically testing the main ideas of the theory leading some to assert that the theory was unfalsifiable (Nettler 1984). Akers (1973, 1998) subsequently revised differential association-reinforcement theory into a social learning theory of crime and deviance that included four key constructs: differential association, definitions, differential reinforcements, and imitation. In the most general terms, Akers' social learning theory asserts that both conforming and deviant behavior are learned in the same way, that one's likelihood of engaging in antisocial behavior is influenced by his/her prior and anticipated consequences of behavior, or that the probability of engaging in deviant behavior is contingent on the prior and anticipated future reinforcements and punishments one has experience regarding deviance.

Social learning theory remains a dominant perspective in criminology, and the empirical literature testing this theory is arguably one of the largest and most persuasive in the field (see Warr 2002). This entry provides a general overview of the assumptions, constructs, and propositions that are central to social learning theory, summarizes the empirical literature testing this perspective, addresses some of the key criticisms, and concludes with new directions for the theory.

Background

In 1965, C. Ray Jeffrey recommended that Sutherland's differential association theory be replaced with a single statement drawing on the principles of modern behavioral learning. Specifically, Jeffrey (1965, p. 295) noted that criminal behavior, like all other human behavior, is governed by the principles of differential reinforcement such that it "occurs in an environment in which in the past the actor has been reinforced behaving in this manner, and the aversive consequences attached to the behavior have been of such a nature that they do not control or prevent the response." The importance of Jeffrey's recommendation cannot be overstated, as he correctly pointed out that criminologists from the learning paradigm had neglected to incorporate the significant advancements in behavioral psychology into the understanding of crime since Sutherland's original contributions.

Around the same time, Burgess and Akers were in the process of integrating these modern learning principles into the differential association perspective. They believed that all of the explanatory variables and processes in differential association theory could be subsumed under a more general theory of human behavior that is based on behavioral learning principles. Burgess and Akers accomplished this by revising each of Sutherland's nine statements to form seven new propositions that invoke language from behavioral psychology and modern learning. For instance, the first and eighth propositions in Sutherland's differential association theory which read "Criminal behavior is learned" (p. 6) and "The process of learning criminal behavior by association with criminal and anti-criminal patterns involves all of the mechanisms that are involved in any other learning" (p. 7), respectively, were combined into a single proposition that read "Criminal behavior is learned according to the principles of operant conditioning" (Burgess and Akers 1966, p. 132). Further, the key proposition of Sutherland's theory which read "A person becomes delinquent because of an excess of definitions favorable to violation of law over definitions unfavorable to violations of law" was similarly revised in a manner consistent with behavioral learning, such that "Criminal behavior is a function of norms which are discriminative for criminal behavior, the learning of which takes place when such behavior is more highly reinforced than non-criminal behavior" (p. 143).

The initial formulation of social learning theory in criminology was heavily influenced by the behavioral learning paradigm developing in psychology (Bandura 1969). In this seminal work, Burgess and Akers discussed the importance of things like discriminative stimuli (signals that elicit a particular behavioral response), saturation (when the potency of a reinforcer is maximized and no longer affects behavior), and extinction (the reduction or loss of a behavioral response when reinforcement is withheld) when describing the learning process. The complexity inherent in this perspective, however, led many criminologists to question the utility and testability of this learning perspective. These concerns, along with the movement towards a variable-based criminology (Hirschi 1969), led Akers to revise social learning theory into a more tenable set of constructs and propositions.

Akers' Revised Social Learning Theory

Social learning theory is developed to explain not just the acquisition of criminal behavior but also the maintenance and modification of criminal behavior (Akers 1998). Put simply, the process of social learning offers not just explanations on the motivation to initially engage in criminal behavior but also explanations on how certain variables work to maintain criminal behavior throughout time and how these variables also can maintain and promote conformity over time. Akers (1998, p. 50) argues that "The basic assumption in social learning theory is that the same learning process. . . produces both conformity and deviant behavior. The differences lie in the direction. . .. [of] the balance of influences on behavior." Thus, individuals are at an increased likelihood of engaging in deviant behavior when they differentially associate with others who commit criminal behavior themselves, who define deviance as desirable, and who provide (or have provided) greater reinforcement than punishment for that deviance. Conversely, individuals are more likely to be prosocial to the extent that they differentially associate with individuals who engage in and espouse norms conducive to conformity.

When developing his revised social learning theory, Akers asserts that the social learning process can be principally captured by four distinct, yet interdependent, constructs: differential association, definitions, differential reinforcement, and imitation. The subsequent sections of this entry will elaborate on these four constructs and discuss how they influence the probability of engaging in criminal and deviant behavior.

Differential Association

Social learning theory begins with the assumption that all learning takes place in a context of structural interactions with others. The most important social context in which mechanisms of learning operate is through the differential associations individuals have. This construct draws heavily on Sutherland's theory, which emphasizes the importance of "intimate personal groups" in the learning process.

Akers asserts that differential association can influence delinquency both through behavioral/interactional dimensions and normative dimensions. The former refers simply to the direct association and interactions individuals have with others and whether the behavioral patterns that these associations expose to an individual are criminal or conforming. In other words, it refers to the behavioral nature of one's associates (e.g., deviant vs. non-deviant). The latter, however, are the different patterns of norms and values that an individual is exposed to through these associations. Sutherland's discussion of differential association was primarily concerned about the transmission of definitions that occurred in these differential interactions, but Akers essentially allows the differential association construct to incorporate a much broader range of deviant influences present when interacting with others, including the transmission of deviant definitions, providing reactions that are favorable to delinquent conduct that alters the perceived rewards/costs of crime and exposing individuals to deviant models from which to imitate.

Like Sutherland, Akers' social learning theory does not view all interactions as being equally influential in affecting human behavior. Instead, Akers retains the four modalities discussed by Sutherland, asserting that associations, and their subsequent influence on deviant behavior, vary in terms of frequency, duration, priority, and intensity. Frequency is how often an individual interacts with a group or individual. Duration concerns the length and relative amount of time spent in different associations. Priority refers to when in time the association began, in other words, which associations came earliest in time. Finally, intensity has to do with the "significance, saliency, or importance of the association to the individual" (Akers 1998, p. 64). Drawing on these four modalities, Akers (like Sutherland) argues that associations will have a greater impact on shaping an individual's behavior if they occur more frequently, if they have been interacting for longer periods of time, if they began earlier in the individuals life, and if they are more meaningful relationships to the individual.

This construct of differential association subsumes many of the major elements of Sutherland's differential association theory. Sutherland's recognition that the interactions that individuals have when associating with others is central to the learning process is acknowledged by Akers, and he continues to stress the importance that different modalities have on influencing how these associations shape behavior. In fact, given Akers' assertion that most learning is shaped by social reinforcement contingencies – rather than nonsocial ones (e.g., intrinsic rewards/ punishments) – differential association is a necessary condition in the social learning perspective.

Definitions

Sutherland's emphasis on the transmission and acquisition of definitions favorable to delinquency when explaining deviant behavior made it difficult to distinguish that construct from differential associations. However, Akers explicitly separates these two constructs in his social learning theory. Despite this, Akers' characterization of definitions is nearly identical to that described by Sutherland. Definitions can generally be thought of as the normative beliefs, attitudes, or meanings that individuals attach to any given behavior. These can include the motives, drives, rationalizations, justifications, and attitudes that define the circumstances in which the commission of an act is right or wrong, desirable or undesirable, justified or unjustified (Akers 1998, p. 77). In social learning theory, definitions can be both general and specific. General definitions refer to the broad religious and moral beliefs and norms one holds that favor either criminal or conforming behavior. Specific definitions refer to specific acts and situations in which behavior is, at minimum, justified. The distinction between general and specific definitions is of theoretical importance. Individuals may generally hold the belief that "breaking the law is wrong" but at the same time believe that "smoking marijuana is OK." Similarly, individuals may hold the general definition that "thou shall not steal" while simultaneous believing "it is morally justifiable to steal when hungry."

Definitions in Akers' theory do not *require* deviant behavior, as some scholars have wrongly interpreted in Sutherland's theory (Kornhauser 1978). Akers (1998) argues that *some* definitions favorable to delinquency can be so strongly held that they require delinquency. More common, however, is that individuals hold definitions that rationalize or justify delinquency in certain situations. In this way, definitions in Akers' theory are similar to the belief construct in Hirschi's (1969) social control theory, in that the definitions themselves do not directly provide motivation for delinquency but rather are conventional beliefs that are weakly held and can allow for other motivating factors (e.g., differential reinforcement) to have important and direct effects on antisocial behavior. That is to say, their moral reservations against delinquency are weak enough that it can be bypassed in some situations and thus allow for delinquent behavior when the right circumstances or reinforcement contingencies are present (Warr 2002).

Differential Reinforcement

Arguably the construct of greatest importance in Akers' social learning theory is differential reinforcement. Though Sutherland stated that criminal behavior is learned like all forms of human behavior, he spent little time discussing the learning process. Drawing on research in psychology, Akers argues that the process of learning antisocial behavior is best captured through the construct of differential reinforcement. "Differential reinforcement refers to the balance of anticipated or actual rewards and punishments that follow or are consequences of behavior" (Akers 1998, p. 67). If an individual believes that they are likely to be rewarded for engaging in deviance and/or punished for deciding upon conforming behavior, then that individual is likely to engage in deviance. Moreover, when faced with a similar situation in the future, that individual is likely to draw on their prior experiences and view that deviant behavioral choice as more reinforcing and the conforming option more punishing. This ultimately leads to what is arguably the most important proposition in Akers' theory: "Whether individuals will refrain from or initiate, continue committing, or desist from criminal and deviant acts depends on the relative frequency, amount, and probability of past, present, and anticipated rewards and punishments perceived to be attached to the behavior" (Akers 1998, p. 66).

From the social learning perspective, human behavior is governed by both reinforcements and punishments, whereby the former *increases* the probability that a behavior occurs and the latter *decreases* it. Both reinforcements and punishments can be either positive or negative, which references the fact that some consequences influence behavior by adding stimuli (positive) while others influence it by removing stimuli (negative). Accordingly, positive reinforcements increase the likelihood of engaging in or repeating deviance by providing a positive outcome or reaction to it (e.g., friends' approval, money, or intrinsic feelings of pleasure). Negative reinforcements increase the likelihood of deviance by removing an aversive or unpleasant feeling or event (e.g., eliminating the ridicule of classmates). Meanwhile, positive punishments decrease the likelihood of criminal behavior by providing a painful or unpleasant consequence associated with a behavior (e.g., an arrest), and negative punishments affect behavior by removing pleasant consequences or rewards (e.g., parental respect). These reinforcements and punishments influence the likelihood of engaging in deviance by altering an individual's perceived anticipated consequences when faced with similar situations in the future, which in turn affect one's decision-making calculus at the time of the offense (Akers 1990, 1998).

Just as with differential associations, reinforcement and punishment contingence are not all weighted equally but are influenced by the amount, frequency, and probability of their occurrence. The greater the amount or value placed on the reinforcement contingencies, the more frequently it is reinforced, and the greater the likelihood of experiencing the reinforcement all increase the likelihood of engaging in deviant behavior. Thus, for instance, if an individual places great value on the relationship she shares with her friend who frequently rewards deviant behavior, and she is certain that her friend will reward her for stealing, that individual will have a high probability of engaging in deviant behavior. However, the reinforcements and punishments that are provided by strangers for whom the individual has no personal relationship with will likely have little effect both on the present and on future perceptions of rewards and punishments.

Though the examples provided above have focused largely on social reinforcements and punishments (i.e., peers and parents), Wood et al. (1997) note that reinforcement contingencies can be nonsocial. For instance, the physiological effects of drugs and alcohol can be rewarding to some, which in turn increases the likelihood of repeated use. Moreover, some individuals may view deviant behavior itself as intrinsically rewarding. This would occur if the behavior itself, regardless of social reactions, is viewed as "fun" or "thrilling." Wood et al. provide support for this premise: positive intrinsic sensations from crime are significantly related to the maintenance of criminal behavior. Interestingly, Akers vests little in the idea that nonsocial factors are inherently reinforcing. For instance, he argues that individuals learn that the effects of drugs are enjoyable or unpleasant *through social reinforcement*. Similarly, Akers also argues that some individuals are prone to be risk-seeking but that they learn to become risk-seeking, and learn to see delinquency as "fun," through social interactions with others. In other words, individuals may experience physiological and intrinsic rewards from drug use and crime but primarily only do so if they are predisposed by a *social* learning history to do so.

Akers believes that the primary source of learning occurs through social reinforcements and punishments. When individuals value the relationships they share with others, the prior, anticipated, and actual reactions that they attach to deviant behavior are likely to be highly influential in affecting the likelihood that an individual offends. Importantly, this is not the same as having peers pressuring individuals to offend, but rather this process refers to the influence that these friends have by altering the perceived consequences of different behavioral choices. Social learning theory recognizes that humans are social beings who are concerned with the reactions and opinions of others and that our perceptions on what the anticipated consequences of behavior entail are necessarily a function of prior experiences in similar situations. Even tangible and material items, such as money and wealth, are valued by individuals because of the social rewards and prestige that they provide. Put simply, Akers argues that tangible and material rewards from crime derive their power of influence because they are symbolic of social rewards (i.e., prestige and status) and that their value (and thus their influence on behavior) is of little importance when the social rewards are disassociated from them.

In sum, the inclusion of differential reinforcement into the learning paradigm of criminology filled the important gap as to *how* criminal behavior is learned (i.e., a focus on the "reception" and not the "transmission" process). Though this contribution by Akers is arguably the most important aspect in social learning theory, it also remains one of the most misunderstood. The idea that individuals are affected by reinforcement contingencies does not mean that humans are passively conditioned by their environment, simply reacting to stimuli like Pavlovian dogs. On the contrary, social learning theory views that individuals are mostly actively engaged and attuned to the anticipated consequences of behavioral actions (Akers 1990), which presupposes that individuals employ some cognition before engaging in delinquency. If an individual determines that the relative balance of the consequences of deviance provides greater reinforcements than punishments, then that individual is at an increased likelihood of engaging in deviant behavior. Moreover, social learning theory argues that the actual consequence attached to the behavior will affect their future likelihood of offending: if an individual engages in antisocial behavior and they are rewarded for it, they are at an increased likelihood of repeating the action when faced with a similar situation in the future. Conversely, if individuals engage in deviance but experience a relatively greater level of negative consequences, they are likely to avoid deviant behavior in the future. Simply put, social learning theory argues that an individual's prior experiences influence how they interpret anticipated consequences in the future and that these anticipated consequences are influential when deciding on a behavioral action choice (Akers 1998).

Imitation

Imitation refers to the modeling of behavior after similar behavior is observed in others. The classic example of imitation derives from Bandura's (1961) Bobo doll experiments, where the behavior of children was observed after watching a model aggress on an inflatable doll. As is evident in these experiments, not all behavior that is observed is imitated by others. Of particular importance is the observed consequence of the behavior: if one observes an individual engage in a behavior and that individual is subsequently rewarded for the behavior, it is more likely that that behavior will be modeled. This is known as *vicarious reinforcement*. Conversely, if one observes behavior that is subsequently punished, it is less likely that the behavior will be modeled – what is known as *vicarious punishment*.

Another factor that influences whether an observed behavior is modeled is the characteristic of the model. In this way, imitation shares some of the modalities as the other constructs in social learning theory. The most salient models are those persons "with whom one is in direct contact in primary groups" (Akers 1998, pp. 76–77). This most obviously means that individuals are more likely to imitate the behavior of close, personal associates than that of strangers. But this statement also implies that individuals are more likely to model behavior observed in direct interaction than behavior observed through media forums. Akers (1998) has argued that technological and social changes have increased the relative importance of modeling behavior on television, movies, and video games. More specifically, given that the relative time spent watching television and movies and playing video games has increased over the last several decades, the modalities of learning have also increased (frequency, duration, priority, and intensity), making media models more salient.

Intuitively, the imitation of others' behavior is most influential in the acquisition and performance of novel behavior but plays little role in the maintenance of deviant behavior. The continuation of the behavior is more affected by direct reinforcements and punishments. For example, an individual may learn initially that violence is rewarded when one's status is challenged, but whether they continue to behave in that matter will be contingent on their own experiences with such behavior. However, imitation may still play a role in the acquisition of criminal skills that can make crime easier and less risky.

Though these constructs are treated and discussed independently, there is considerable interdependency among them. Indeed, social learning is itself a process that incorporates all of these constructs in a complex and recursive manner. Initial deviance is influenced by the balance of definitions one has been socialized to hold, by the modeling of others' behavior, and/or by the anticipated rewards and punishments that derive from a criminal event. Immediately after this event, the actual consequences of the action play an important role – if the individual experiences some reward, then he/she is likely to believe there will be positive consequences in the future. Further, if one does experience reinforcement, this will likely alter the definitions they hold and, in turn, could potentially alter the individual's criminal and non-criminal associates (i.e., self-selection). Indeed, social learning theory does not claim that associations are random but predicts that individuals do self-select into peer networks. However, social learning theory does predict that more often than not delinquent associations precede delinquency.

Research on Social Learning Theory

Few theories have been tested empirically to the extent that social learning theory has. These tests most often assess the relationship and direction between variables from the social learning perspective – differential association, definitions, differential reinforcement, and imitation – and some deviance outcome measure (i.e., general delinquency, drug use, violence). The overwhelming majority of these studies have found strong support for social learning predictions, with very few studies finding negative support for the theory.

One of the earliest and most seminal tests of the theory was conducted by Akers and his colleagues in 1979. Using substance use as an outcome, the results found that differential association (i.e., delinquent friends), definitions (how right/wrong substance use is), and differential reinforcement (how parents and friends would react to respondent drug use) were all strongly and statistically related to substance use, providing support for the social learning predictions. Imitation, however, was only weakly and inconsistently related to substance use, though Akers et al. (1979) note that this finding is not unexpected, as imitation is only predicted to have strong effects on the initiation of delinquent behavior. Though the research findings were strongly supportive of the theory, the clear operationalization of the theory into testable measures and predictions within the article is an equally important contribution. Since this work, many empirical studies have assessed the empirical validity of social learning, with most supportive of the perspective (Pratt et al. 2010).

Another common manner in which research has inferred the empirical validity of social learning theory is through an assessment of the relationship between delinquent peers and offending. As would be predicted by social learning theory, the relationship between peers and delinquency is one of the strongest and most consistent findings in all criminological research (Warr 2002). More frequent, longer-term, and closer associations with peers who engage in delinquent behavior is consistently related to the delinquency or respondents, while associating with prosocial peers is consistently related to conformity. This strong relationship has generally held using various samples and methodological specifications. Ultimately, these findings have led some scholars to conclude that delinquent peers are one of, if not, the most important causes of adolescent delinquency (Warr 2002) and provide considerable support for both the differential association and social learning perspectives.

Perhaps the most telling support for social learning theory comes in two forms: its relative superiority when faced with theoretical competition and its ability to explain known correlates of offending. Regarding the former, many empirical works in the field of criminology have sought to pit variables from one theoretical perspective against those from others. Most generally, this involves putting learning variables in statistical models and comparing the size of the effects against variables from control and strain theories. When this is done, the variables from learning theory tend to explain more of the variance in delinquency than variables from other perspectives (Warr 2002). Learning variables also do an impressive job explaining some of the well-known correlates of offending. For instance, Warr (1993, 1998) has demonstrated that associating with delinquent peers can explain both the age and marriage effect. Similarly, Mears et al. (1998) have shown that learning theory also explains a substantial portion of both race and gender differences in crime, respectively.

The empirical support for social learning theory has most recently been summarized in a meta-analysis conducted by Pratt and colleagues (2010). This study reviewed the published extant tests of the perspective to provide a synthesized estimate of the effect size of the social learning theory constructs. In reviewing these studies, Pratt et al. conclude that the empirical support for social learning theory is stronger than many other criminological perspectives and is at least as comparable as the most strongly supported theories (see Pratt and Cullen 2000). Interestingly, however, the different constructs of social learning theory appear to have varying levels of explanatory power. The constructs most strongly related to deviance are differential associations and definitions, the two constructs already captured in Sutherland's theory. Differential reinforcements and imitation are only modestly related to delinquency. If indeed differential reinforcement is the most important construct in the social learning perspective, these findings raise some questions about Akers' perspective and provide researchers with an important avenue to further assess the interdependence of the four constructs of social learning theory.

An extension of social learning occurred in 1998 when Akers (1998) suggested that an integrated theory (social structure and social learning) could explain structural variations in crime rates across communities. Although this premise has yet to receive much empirical attention, the basic argument is that social structure is linked to individual behavior – and in turn, crime rates – through its effects on social learning variables, namely, differential association, differential reinforcement, definitions, and imitation. Structural characteristics provide the contexts within which social learning variables operate: it causes individuals to differentially associate with delinquents and exposes individuals to a greater ratio of deviant reinforcement contingencies and to definitions favorable to delinquency, as well as a greater number of delinquent models to imitate.

Four dimensions make up the SSSL perspective. *Differential social organization* refers to the distribution of structural correlates of crime (age, race, gender, SES) in the community or society that affect rates of delinquency. *Differential location in the social structure* refers to one's own social and demographic location in the social structure that affects social learning variables (e.g., younger individuals are more likely to be exposed to other younger individuals who, in turn, are more crime prone). *Theoretically defined criminogenic aspects of social structure* draws on anomie, social disorganization, and conflict theories which, again, Akers argues affects one's exposure to criminal associates, definitions, models, and reinforcements that can promote deviant behavior. Finally, *differential social location in primary, secondary, and reference groups* refers to the more immediate personal networks that filter the larger social environment to either promote or discourage criminal behavior. In sum, the general premise of SSSL is that social learning variables fully mediate social structural characteristics correlated with high rates of crime. To date, however, this perspective has not fully been tested by researchers, and the ability of social learning theory to explain differences in crime rates remains in question.

Criticisms

Pratt and colleagues' (2010) finding that differential reinforcement and imitation had relatively smaller effects in predicting delinquency when compared to differential association and definitions can be seen as challenging to Akers' contributions to the learning paradigm, but several other criticisms also plague the theory. Most notable, several scholars have challenged the supposed causal relationship between peers and delinquency that is hypothesized by learning theorists. The argument is that the strong correlation between delinquent peers and delinquency is not evidence that differential associations cause delinquency but instead reflects the tendency for "birds of a feather to flock together" (Gottfredson and Hirschi 1990). This "selection" argument has more recently been championed most fervently by scholars from the control perspective who assert that both delinquency and the association with delinquent peers are both outcomes of similar causal factor (i.e., weak attachments to society or low self-control). The criticism is that delinquent associations are not random, but instead individuals self-select into friendship networks based on delinquent tendencies. Thus, proponents of the selection perspective argue that the relationship between peers and delinquency is spurious, as delinquent behavior is thought to actually precede delinquent peer associations.

Though often seen as one of the greatest challenges to social learning theory, this criticism falls short for both theoretical and empirical reasons. First, social learning theory does not predict that peer associations are randomly constructed but instead explicitly acknowledges that individuals self-select into peer groups based on their prior learning. Unlike control theories, however, social learning theory predicts that once this selection does occur, these associates are likely to have a continued influence on delinquent behavior through differential reinforcement and other learning mechanisms. The empirical research is supportive of the learning predictions: even when statistically controlling for competing variables that are thought to render the peer-delinquency relationship spurious (Gottfredson and Hirschi 1990), the association between peers and delinquency remains (Matsueda and Anderson 1998). Associations with delinquent peers most often precede delinquent behavior (Elliott and Menard 1996), which is supportive of the predictions of social learning theory. Still, this selection versus socialization debate is far from settled (see Weerman 2011). Future research using experimental and quasi-experimental designs will likely shed more light on this issue.

Another criticism of social learning theory concerns the types of deviant behaviors in which the theory is able to explain. Many tests of social learning theory have assessed whether Akers' theory can explain variation in substance use and other forms of minor delinquency. Much less research has examined how well social learning theory explains more serious forms of delinquent behavior, and the research that does finds that learning theory does a poorer job explaining these more serious forms of delinquency. The reason behind this remains relatively unexplored, but one possibility is that individuals who engage in more serious forms of violence have such high levels of criminal propensity (i.e., high in impulsivity) that they do not require learning mechanisms to engage in delinquent behavior. In any case, the finding that social learning has more trouble explaining more serious forms of delinquency is problematic for the theory, as Akers' perspective is designed to explain all forms of crime and deviance.

Finally, following Kornhauser's critique (1978, p. 34) of differential association/social learning theory assumptions that "man has no nature, socialization is perfectly successful, and cultural variability is unlimited," social learning theory has been interpreted by some as a theoretical perspective unable to explain individual differences and only applicable to group differences in crime where subcultural definitions adhere to deviance (Akers 1996, p. 230). A corollary is that the theory focuses entirely on the development of novel criminal behavior that, once learned, *requires* deviance. In essence, critics see little to differentiate differential association and social learning from cultural deviance theories. Akers (1996), in an exchange with Travis Hirschi on this subject (Hirschi 1996), suggests that these criticisms stem from ambiguities associate with and misrepresentations of Sutherland's statement of the differential association – problems that Akers (1996, p. 238) believes have been rectified in his specification of social learning principles by allowing for variations in both deviant definitions held by individuals and the social reinforcement they receive. "These deviations can develop within the same normative or cultural system. They do not require the existence or participation in an organized deviant subculture."

New Directions

Despite these criticisms, social learning theory remains one of the preeminent theories of crime, and the considerable empirical support for this perspective seems to justify this prominence. Nevertheless, there are some important questions deriving from social learning theory that have been left unexplored.

Saturation Effects and Discriminative Stimuli. In describing a social learning theory of crime and deviance, Akers (1998; see also Burgess and Akers 1966) invokes language from psychology that describes how learning differs across time and situations. Most notably are the concepts of saturation and discriminative stimuli. Saturation occurs when the potency of a reinforcer has been maximized, and that reinforcer no longer has an effect on behavior. This has clear implications for the structure of peer networks: some individuals may no longer be influenced by associating with delinquent associates, as their prior learning and already existing associates may have already hit a saturation point. Put simply, the relationship between delinquent associates and delinquency may be nonlinear, increasing at a decreasing rate. Zimmerman and Vasquez (2011) have assessed this saturation hypothesis at a neighborhood level. Their results provide support for the saturation effect: delinquent peers have a strong effect for individuals residing in non-disadvantaged neighborhoods but a diminishing effect in disadvantaged ones. Still, there is considerable research left unexplored on the saturation effect at the more proximal friendship network level and over the life course. How does the inclusion of a delinquent peer change the delinquency of an individual in a relatively nondelinquent network versus a relatively highly delinquent network? Further, do peers have the same effect over the life course? Or do individuals become saturated from delinquent peers leading to a declining peer effect as one ages?

The discriminative stimuli effect has been relatively neglected in criminological research (however, see McGloin et al. 2011), despite the direct implications it has for theory. Discriminant stimuli refer to stimuli that influence how an individual will act in a given situation. Discriminant stimuli are most important when they signal which situations and circumstances will yield reinforcement. To be clear, it is highly unlikely that individuals view that their friends will reward violence, theft, or other forms of deviance in all situations. Rather, it is more common that an individual's delinquent friends will reward it in some situations and under some circumstances. Discriminative stimuli serve as the cues as to when these situations and circumstances occur. The same logic applies in understanding one's definitions favorable/unfavorable to delinquency. As noted above, most delinquents do not hold the universal values that "deviance is right" but rather hold definitions that justify delinquency in some situations (Akers 1998). Unfortunately, most measures tapping into peer reactions and individual definitions use global statements on the rightness and wrongness of antisocial behavior. Considerable theoretical insight can be gained by beginning to incorporate this discriminative stimuli logic into learning measures and assessing what situations and circumstances allow individuals to justify deviant behavior. The overall point is that there remains important elements from the more general behavioral learning paradigm that have not been explored in criminological research that can shed some important light on the understanding of crime.

Differential Susceptibility: *How and When Learning Variables Matters*. Just as learning variables may not have ubiquitous effects across all situations, there may also be variability among individuals in this regard. For generations the primary question surrounding social learning theory has been the central question of causality: do peers and other associates (e.g., family) hold causal role in the facilitation of delinquency? As noted above, the overwhelming evidence supports the predictions from social learning theory. At this point, it seems worthwhile to move beyond simple assessments of *if* learning variables influence delinquency and begin to assess if they matter differently for different people.

Take, for instance, the relationship between delinquent peers and delinquency. Akers suggests that delinquent peers influence delinquent behavior by providing positive reactions to such behavior. He goes on to specifically acknowledge that the influence of such reactions is necessarily dependent on the ability of individuals to consider the longer-term consequences of behavior; in other words, it requires individuals to reflectively consider how their peers would react before deciding to offend (Akers 1990). If this is the case, then some individuals may not be susceptible to the risk of associating with delinquent peers. In particular, individuals high in impulsivity, who are characterized by their tendency to make decisions based almost entirely on situational stimuli, may not need or consider the reactions of their peers when deciding to offend. Instead, these individuals will be most susceptible to the situational risk of unstructured socializing with peers (Osgood et al. 1996). This hypothesis was supported by Thomas and McGloin (2013) using three different large-scale data sets: delinquent peers have their strongest effects on individuals low in impulsivity but only weak (and nonsignificant) effects on individuals high in impulsivity. Further, the situational risk of unstructured socializing with peers has its strongest effects on this high in impulsivity, but a weak (and nonsignificant) effect on this low in impulsivity. Thus, some individuals may be particularly susceptible to the reactions of their peers, while others may not need such reactions to motivate them to deviate. The larger point is that researchers should begin to think more critically about the mechanisms by which learning variables facilitate delinquency. These more nuanced considerations can shed light on important considerations as to how and when peers and other associates facilitate delinquent behavior. Do all individuals rely on delinquent associates to deviate? How and when do individuals consider their own definitions before deciding to offend? Indeed, the question as to if social learning variables are related to crime and deviance appears to have been answered, and researchers should begin to assess whether these variables matter differently for different people.

Objective versus Perceptual Measures. An area of growing interest among criminologists deals with the most appropriate way to measure variables of theoretical interest, and social learning theory has been particularly in the crosshairs of this discussion. Several scholars have expressed concerns regarding the adequacy of traditional perceptual measures of delinquent peers, whereby respondents are asked to report on the behavior of their friends. Most critically, critics have suggested that perceptual measures have had a tendency to overestimate the delinquent peer effect, because individuals are not accurately reporting the behavior of their friends, but instead are merely projecting their own behavior to them (Gottfredson and Hirschi 1990). The limited research that has been conducted is supportive of this claim: when objective measures of peer deviance are used, the relationship between peers and delinquency is reduced, though still substantively large. These findings have led many scholars of crime to encourage the use of objective measures over the use of perceptual ones while outright rejecting the utility of the latter measures because they reflect an individual's misperceptions of their friends' behavior and not the reality of it.

It is interesting that this movement has occurred during the same time period in which rational choice and deterrence theorists have advocated the use of perceptual measures over objective ones when assessing criminal decision making. After decades of empirical work challenged the core propositions of rational choice theory when using objective measures, scholars began to realize that sanction threats can only influence behavior if they are perceived by the individual. The crux of their argument was that these choice theories are perceptual in nature and, thus, require scholars to assess one's perceptions of their beliefs of risk rather than utilizing the objective reality of that risk. The positive research using these perceptual measures has led to some resurgence in the rational choice perspective (Nagin 1998). These positive findings has led Nagin (1998) to argue that scholars interested in the choice perspective should move beyond assessing if risk perceptions influence behavior and focus on how individuals actually formulate their perceptions of risk, in other words, what information (or "signals") go into the development of subjective risk perceptions.

Interestingly, social learning theory is also an inherently perceptual theory of crime. The core proposition of the theory is that the perceived reinforcements and punishments associated with behavior influence the present and future likelihood of offending. Given this strong emphasis on perceived reinforcements and punishments, and the robust relationship between these perceptions and offending, one future direction for social learning scholars is to heed Nagin's (1998) call and end the discussion on the superiority of objective measures over perceptual ones and begin to consider what factors go into one's perceptions of social reinforcements and punishments. To be clear, a recent study conducted by Young et al. (2011) found that individuals have a tendency to misperceive their own friends' delinquent behavior. Of course, this finding could be used as more ammunition for the objective-measure-superiority debate, but it also can be used to ask more interesting questions regarding individual perceptions of reinforcements: what causes individuals to misperceive their friends' behavior? Or more generally, how do individuals formulate their subjective perceptions of social reinforcements? The answer to these questions has considerably more theoretical utility and can shed important light on the learning process, in general.

Conclusion

Social learning theory begins with the assumption that criminal behavior, like all human behavior, is learned in the interaction with others. This theory went beyond Sutherland's perspective by more clearly specifying what this learning process entailed. Simply, social learning theory asserts that the probability an individual commits deviant acts depends on the relative frequency, amount, and probability of past, present, and anticipated rewards and punishments perceived to be attached to the behavior. The empirical literature supportive of the theory is one of the most impressive in the field of criminology. Nevertheless, considerable avenues of ongoing and future research can lead to more theoretical clarity.

Related Entries

- ▶ Behavioral Management in Probation
- ▶ Cognitive/Information Processing Theories of Aggression and Crime
- ▶ Co-offending and Offender Decision-Making
- ▶ Cultural Criminology
- ▶ Deterrence
- ▶ Differential Association Theory
- ▶ Informal Social Control
- ▶ Punishment as Rehabilitation
- ▶ Rational Choice Theory
- ▶ Sex Offender Treatment

Recommended Reading and References

Akers RL (1973) Deviant behavior: a social learning approach. Wadsworth, Belmont

Akers RL (1990) Rational choice, deterrence, and social learning theory: the path not taken. J Crim Law Criminol 81:653–676

Akers RL (1996) Is differential association/Social learning cultural deviance theory? Criminology 34:229–247

Akers RL (1998) Social learning and social structure: a general theory of crime and deviance. Northeastern University Press, Boston

Akers RL, Krohn MD, Kaduce LL, Radosevich M (1979) Social learning and deviant behavior: a specific test of a general theory. Am Sociol Rev 44:635–655

Bandura A (1969) Principles of behavior modification. Holt Rinehart and Winston, New York

Bandura A, Ross D, Ross SA (1961) Transmission of aggression through imitation of aggressive models. J Abnorm Soc Psychol 63:575–582

Burgess RL, Akers RL (1966) A differential association-reinforcement theory of criminal behavior. Soc Probl 14:128–147

Elliott DS, Menard S (1996) Delinquent friends and delinquent behavior: temporal and developmental patterns. In: Hawkins JD (ed) Delinquency and crime: current theories. Cambridge University Press, New York, pp 28–67

Gottfredson M, Hirschi T (1990) A general theory of crime. Stanford University Press, Palo Alto

Hirschi T (1969) Causes of delinquency. University of California Press, Berkeley

Hirschi T (1996) Theory without ideas: reply to Akers. Criminology 34:249–256

Jeffrey CR (1965) Criminal behavior and learning theory. J Crim Law Criminol Police Sci 56:294–300

Kornhauser R (1978) Social sources of delinquency. University of Chicago Press, Chicago

Matsueda RL, Anderson K (1998) The dynamics of delinquent peers and delinquent behavior. Criminology 36:269–308

McGloin JM, Schreck CJ, Stewart EA, Ousey GC (2011) Predicting the violent offender: the discriminant validity of the subculture of violence. Criminology 49:767–794

Mears DP, Ploeger M, Warr M (1998) Explaining the gender gap in delinquency: peer influence and moral evaluations of behavior. J Res Crime Delinq 35:251–266

Nagin DS (1998) Criminal deterrence research at the outset of the twenty-first century. Crime and justice: a review of research, vol 23. University of Chicago Press, Chicago, IL

Nettler G (1984) Explaining crime, 2nd edn. McGraw Hill, Boston

Osgood DW, Wilson JK, O'Malley PM, Bachman JG, Johnson LD (1996) Routine activities and individual deviant behavior. Am Sociol Rev 61:635–655

Pratt TC, Cullen FT (2000) The empirical status of Gottfredson and Hirschi's general theory of crime: A meta-analysis. Criminology 38:931–964

Pratt TC, Cullen FT, Sellers CS, Winfree LT, Madensen T, Daigle LE, Fearn N (2010) The empirical status of social learning theory: a meta-analysis. Justice Q 27:765–802

Sutherland EH (1947) Principles of criminology, 4th edn. Lippincott, Philadelphia

Thomas KJ, McGloin JM (2013) A dual-systems approach to understanding differential susceptibility to different mechanisms of peer influence. Criminology (Forthcoming)

Warr M (1993) Age, peers, and delinquency. Criminology 31:17–40

Warr M (1998) Life-course transitions and desistance from crime. Criminology 36:183–216

Warr M (2002) Companions in crime: the social aspect of criminal conduct. Cambridge University Press, Cambridge

Weerman FM (2011) Delinquent peers in context A longitudinal network analysis of selection and influence effects. Criminology 49:253–286

Wood PB, Gove WR, Wilson JA, Cochran JK (1997) Nonsocial reinforcement and habitual criminal conduct: an extension of learning theory. Criminology 35:335–366

Young JTN, Barnes JC, Meldrum RC, Weerman FM (2011) Assessing and explaining misperceptions of peer delinquency. Criminology 49(2):599–630

Zimmerman GM, Vasquez BE (2011) Decomposing the peer effect on adolescent substance use: Mediation, nonlinearity, and differential nonlinearity. Criminology 49:1235–1273

Social Network Analysis and the Measurement of Neighborhoods

Rick Grannis
Department of Sociology, University of California, Irvine, Irvine, CA, USA

Synonyms

Neighborhood effects; Neighborhoods; Neighboring; Social networks

Overview

Neighborhoods shape a variety of outcomes for children, families, and residents in general, influencing behavior, attitudes, and values as well. While some neighborhoods foster continuous patterns of criminal activity, others develop collective efficacy, the shared understanding that their constituent members have social capital resources which they are mutually able and willing to use to achieve collective outcomes. Neighborhoods cluster outcomes, some of which cannot be accounted for in terms of the characteristics of the individuals or households currently residing in them; they prove to be real communities with enduring characteristic patterns that survive the replacement of their constituent members. A useful neighborhood definition would be one that helped us better understand these neighborhood communities and their effects. These neighborhood communities are not only geographically meaningful but geographically identifiable as well because the networks of interactions among neighboring residents which produce them, which translate neighboring interactions into neighborhood communities and their effects, are constrained by predictable urban geographic substrates. New research has proposed behaviorally oriented definitions of neighborhoods, defining them in terms of their potential for interaction among residents. Defining neighborhoods in this way provides a lens to focus more closely on neighborhoods as effect-generating communities emerging from the networked interactions of their constituent residents.

Fundamentals/Key Issues

Why Neighborhoods Matter in the Twenty-First Century: The Continual Emergence of Neighborhood Effects

Neighborhoods: Social Capital, Collective Efficacy, and Crime There appears to be a continually increasing interest in the role of neighborhoods in shaping a variety of outcomes for children, families, and neighborhood residents in general (for an overview, see Brooks-Gunn et al. 1997a, b). These "effects" have included a vast array of phenomena ranging from child and adolescent development (e.g., abuse and maltreatment, school completion and achievement, drug use, deviant peer affiliation, delinquency and gangs, adolescent sexual activity and pregnancy and childbearing, parenting behaviors) to concentrated disadvantage and its many corollaries (economic attainment and labor market success, crime and violence, physical disorder, the perpetuation of racism, to name just a few). An overwhelming conclusion reached by all of these studies of neighborhood effects is that neighborhoods influence our behavior, attitudes, and values. They shape the types of people we will become and expose us to or shield us from early hazards that would seriously restrict the opportunities available to us later in life. After our homes, and in conjunction with them, they are where we first learn whether the world is safe and cooperative or inchoate and menacing.

Not all neighborhoods are alike, however. Some neighborhoods are characterized by high levels of effective community. They clearly offer social capital to their residents, a social organization which facilitates and coordinates cooperative action for mutual benefit, which allows residents to deal with daily life, seize opportunities, reduce uncertainties, and achieve ends that would not otherwise have been possible. This social organization is a resource which is not individually attainable because social capital is not a characteristic of individuals; it is a supra-individual property of social structure and it seems to be particularly well grounded in neighborhood communities. These sources of social capital tied to the neighborhood community context are analytically distinct from and as consequential as the more proximate family processes and relationships occurring in the home. Some neighborhoods develop a further layer, mutual trust and shared norms, values, and expectations, beyond the resource potential of neighbor networks, which allows them to utilize these networks to achieve desired outcomes. Collective efficacy occurs when members of a collectivity, with social capital resources, believe they are mutually able and willing to use these resources to achieve an intended outcome (Morenoff et al. 2001). The distinction is a subtle, but important, one. A neighborhood may have social capital resources available for its constituent members to utilize but they may not trust the willingness or ability of their fellow residents to use these networked resources for the collective good or they may not even be certain that they agree as to what the collective good might be.

From a less positive perspective, neighborhoods show remarkable continuities in patterns of criminal activity as well. For decades, criminological research in the ecological tradition has confirmed the concentration of interpersonal violence in certain neighborhoods, especially those characterized by poverty, the racial segregation of minority groups, and the concentration of single-parent families. Even in neighborhoods with less socioeconomic or racial isolation, crime rates persist despite the demographic replacement of neighborhood populations (Brantingham and Brantingham 1993). In addition, neighborhoods determine not only one's exposure to crime and violence but also a host of less tangible deleterious factors which contribute to the development of an urban underclass, signs of social disorder which lead residents to perceive their neighbors as threats rather than as sources of support or assistance (Massey and Denton 1993).

Neighborhoods: Geography and (Potential) Effects These neighborhood effects, both the community enhancing and the community degrading effects, necessarily involve a geographic context. Thus, to analyze and understand them, they necessarily require a geographic equivalent of a neighborhood and a geographic definition of one. Researchers have utilized a variety of such definitions. In fact, they have used so many that Galster (2001, p. 2111) argued that "urban social scientists have treated 'neighborhood' in much the same way as courts of law have treated pornography: a term that is hard to define precisely, but everyone knows it when they see it." Apparently, however, researchers often don't know it when they see it. The modifiable area unit problem (MAUP) is a statistical bias affecting areal unit summary values (e.g., totals, rates, proportions) when they prove "arbitrary, modifiable, and subject to the whims and fancies of whoever is doing, or did, the aggregating" (Openshaw 1984, p. 3). Miller's (1999) survey suggests that, when the spatial units used to study a phenomenon are not clearly defined by theory, any conclusions derived about the studied phenomenon may be hopelessly prejudiced by the arbitrary, at least from a theoretical perspective, choice of spatial unit.

While many statistical techniques and error modeling approaches have been used to try and counteract, reduce, or remove the effects of MAUP, Miller perceives that the ultimate solution has to involve a more behavioral-oriented definition of neighborhood. One needs better intuitions about the general nature of neighborhoods, not better statistical methods. The very existence of the modifiable areal unit problem (MAUP) evidences that theory has taken a back seat. Researchers who have developed methods for creating optimal analytic units with respect to predefined objective functions note correctly that MAUP would be irrelevant if neighborhood equivalents were chosen for theoretical reasons rather than administrative convenience (Alvanides et al. 2001).

Despite this, however, when a geographic definition of neighborhood is required for the purpose of quantitative analysis, "most social scientists and virtually all studies of neighborhoods . . . rely on geographic boundaries defined by the Census Bureau or other administrative agencies. . . [which] offer imperfect operational definitions of neighborhoods for research and policy" (Sampson et al. 2002, p.445). Administratively defined units such as census tracts and block groups do not directly measure, nor were they designed to measure, the potential for interaction among resident members, the primary process hypothesized to produce neighborhood communities and their effects. For example, the Census Bureau appears to theoretically focus on segregation type effects, defining tracts as "relatively homogeneous units with respect to population characteristics, economic status, and living conditions." In most cases, however, the sheer ubiquity of data gathered by the Census Bureau or other administrative agencies (e.g., school districts, police districts) proves to be an overwhelming temptation for most researchers. Theory succumbs to the preponderance of data. The very existence of the modifiable area unit problem (MAUP), however, evidences that census geography is not measuring what researchers are studying.

Undoubtedly, neighborhood effects involve a geographic context. Neighborhood effects, however, are not produced by neighborhood geography. Neither are neighborhood effects, at least not a lot of them, merely a by-product or spurious confound of the geographic co-location of residents with particular demographic characteristics or psychological profiles. Neighborhood effects hypothesize that there exists a thing, a social entity, a neighborhood community, which has effects. Neighborhood effects are the product of these neighborhood communities. Neighborhood communities and their effects emerge from neighboring interactions among their constituent members. Neighborhood communities are geographically constrained because the interactions which produce them are geographically constrained. Neighborhood communities are both geographically identifiable and have effects which persist through the replacement of their residents because the networks of interactions which produce them, which translate neighboring interactions into neighborhood communities, are constrained by predictable urban geographic substrates.

When we think about geography having an effect on community, it is because we believe that something about physical space affects something about how individuals interact within that space. Neighborhoods are more than colored boxes on a map or sets of geo-referenced variables for use in a geographic information system (GIS). A focus on maps, especially maps based on census or administrative geography, emphasizes those aspects of neighborhoods and their residents which can be effectively displayed or referenced to administratively defined polygons and ignores those which cannot. To understand the social-interactional aspect of neighborhoods, we may not necessarily have to think outside the box, but we do have to think about what's inside of it, residents potentially interacting with each other as neighbors.

Neighboring and Geographic Neighborhoods While it may seem obvious, it is worth highlighting that, at its most fundamental level, neighboring is a proximity-dependent relation. When we say that someone is our neighbor, we are making a statement about them being proximal to us. Neighbors must, by definition, live close to each other; but what constitutes the geographic proximity or availability that defines neighboring?

Many studies have called attention to the strong role of extremely short distances in neighborly contacts. At least from an individual household's perspective, the distances associated with neighboring are often effectively measured in feet and yards (Festinger et al. 1950). Residential propinquity's influence on social interaction is typically limited to those who live within a few households away. What is most important is who lives next door, not who lives in the same census tract; who lives a few houses away, not who lives a few blocks away. Neighbors have also been defined to be people who live within walking distance (Grannis 2009). Walking distance, of course, varies by person, being much greater for some than others due to their age and physical fitness and even changing seasonally in some areas. No matter how far walking distance is, however, for any particular person, the probability that someone will be identified as a neighbor declines rapidly with increased distance from one's home.

Because neighboring is so close at hand, it depends upon very subtle geographic features. Besides focusing on the number of houses or yards separating two households, a natural division, in both cognition and behavior, occurs at the face block. The face block includes all of the dwellings that front on the same street and are situated between only two cross streets (an exception would be cul-de-sacs which are face blocks delimited by only one cross street). The face block has been found to be an important sociospatial unit (Suttles 1972). At one level, the face block includes virtually all neighbors who live either next door to each other or directly across the street from each other and most of those who live within a few house lengths; therefore, it could be viewed as simply a reflection of the more general effects of proximity. However, studies have shown that residents have more interaction with those on the same face block than they do with residents beyond an intersection, even if they were spatially closer to the latter group (Greenbaum 1982).

Not all face blocks, however, are oriented towards the pedestrian nature of neighboring. Some front on large arterial streets devoted to providing access for travelers while others front on smaller streets more devoted to local living space. Studies of peoples' perceptions of street life reveal that residents clearly perceive the difference between heavy traffic face blocks continuously filled with strangers which are used solely as thoroughfares and corridors between the local neighborhood and the outside world on the one hand and light traffic face blocks which form the basis of lively, close-knit communities where everyone knows each other and residents consider the boundaries between house and street space to be quite permeable on the other hand (Appleyard and Lintell 1986). A tertiary face block has been specifically designed and maintained by governing authorities to promote local and pedestrian traffic. The tertiary face block is a more or less "natural" unit of face-to-face neighborly interaction (Suttles 1972). Tertiary face blocks are both oriented towards pedestrian travel and local residents, rather than outsiders who arrive by automobile or mass transit. Thus, tertiary face blocks are the types of face blocks most likely to give rise to social interactions (Rabin 1987). Tertiary face blocks provide a meeting place for neighbors (de Jong 1986). People use them for a host of activities including walking pets, riding bicycles, and chatting with neighbors. Shared tertiary face blocks provide a "permeable boundary" between households' private spaces. Tertiary face-block neighbors are "used for easy sociability and assistance when quick physical accessibility is an important consideration."

Face blocks terminate at intersections. An alternative way of thinking about this, of course, is that intersections connect face blocks with each other. The important question then becomes: Do they also connect neighbors and do neighbor networks terminate at intersections or do they bridge them to form larger structures? Intersections form a different metric than face blocks for measuring functional distance. Just as different types of streets differentially induce or fail to induce neighborly relations, different types of intersections may induce or inhibit neighborly relations from bridging them. Fortunately, the intersections of streets can also be operationally defined in a convenient and meaningful way. When streets of different classifications intersect, planners consider the intersection to be of the higher classification. For example, if a larger street intersects a tertiary street, planners would consider the intersection to be part of the larger street, but not part of the tertiary street. This makes intuitive sense because the inhibiting effects of the larger street will dominate. The nature of intersections either facilitates or impedes pedestrian-based neighborly interaction; therefore, an intersection is a tertiary intersection if all of the face blocks, however many, contiguous with it are tertiary face blocks. A non-tertiary intersection, in contrast, is an intersection such that at least one of the face blocks contiguous with it is a non-tertiary face block.

Studies have shown that tertiary intersections combined with tertiary face blocks can serve as a "pedestrian circulation system." Sidewalks provide access between residence and parks, churches, and neighborhood shops. Neighborly relations bridging face blocks occur "through the routes people take in meeting an average day's basic needs and desires. The newsstand where one buys the Sunday paper, the store one runs to for a quart of milk, and the streets one travels on to visit a friend" (Anderson 1992, p. 46). People come to envision their neighborhoods as networks of paths and channels along which they move (Lynch 1971). Tertiary intersections guide this "natural movement" (Hillier 1996) within a city.

Potential Neighbor Networks and Actual Neighbor Networks Individual neighbor networks evolve into neighborhood networks through the process of concatenation. Residents have relations with their neighbors who interact with other neighbors, and so on. These neighborly relations concatenate and consolidate neighbor-to-neighbor-to-neighbor. There are several important corollaries of this fact. First, the resultant network is as far reaching as its most extensive ramification. Relations concatenate to form a network typically larger, both relationally and geographically, than any individual's relations. Thus, relatively micro-level relations can result in a macro-level structure. Second, the resultant network is as fragile as its weakest link. Anything which can cause a relation to not form, no matter how trivial, breaks the network. In contrast to the first corollary, micro-level fragilities can destroy a macro-level structure. Third, the characteristics of the resultant network are not readily predictable from the characteristics of the local networks which concatenate to form it. Only as sets of individual networks concatenate do the characteristics of this aggregated network emerge. Finally, because a neighboring relation cannot exist unless residents are geographically available to each other, the network of potential neighbors cannot transcend the network of geographic availability; it is logically impossible. While individuals' lifestyles and habits may prevent them from having contacts and interactions with those who are geographically available to them, they cannot cause them to have contacts with those who are unavailable.

To study efficacious neighborhood communities emerging from neighbor networks, therefore, we need a definition of a neighborhood community whose importance is derived from the potential for neighbor networks to concatenate within it. Grannis (2009) defined this geographic availability in terms of shared walking arenas which mediate, guide, and constrain potential neighborly encounters. Building on this, the concatenated network of overlapping neighborly contacts can be no larger than the concatenated network of walking arenas; conversely, the network of potential neighborly relations, based on concatenated interactions, is a subset of the concatenation of these walking arenas. Grannis argued that tertiary block faces effectively proxy walking arenas in urban areas and thus the maximal concatenation of contiguous tertiary block faces, of walking arenas, represents the maximal consolidation of individual residents' potential to access each other.

Such a neighborhood equivalent would signify internal access. All residents within it would have a potential for neighborly relations using walking arenas. While it is unlikely that all, or even any, residents would traverse the entirety of this neighborhood equivalent, this internal contiguity would allow residents to interact with their neighbors down the street who interact with other neighbors further down the street, and so on throughout the network. Such a neighborhood equivalent would also signify constraint. To the extent to which the potential for neighboring relations depends upon walking arenas, it would constrain their concatenation. Based on these criteria, Grannis (2009) defined two types of neighborhood equivalents, t-communities and islands, one connecting tertiary block faces using only tertiary intersections and the other connecting tertiary block faces using all intersections. A t-community is a maximal contiguous network of tertiary block faces and tertiary intersections, while an island is a maximal contiguous network of tertiary face blocks and any intersections. The label "t-community" is short for "tertiary street communities" indicating that these neighborhood equivalents are connected by tertiary streets (Grannis 2009). A pair of neighborhoods would be geographically unreachable if they could only reach each other by crossing water; Grannis (2009) analogously label these discontinuous networks of tertiary block faces as "islands" since it is impossible for households in two different islands to access each other using tertiary block faces, even if they use non-tertiary intersections. Since islands are maximal networks of tertiary block faces and any intersections while t-communities are maximal networks of tertiary block faces and tertiary intersections, t-communities are necessarily subsets (although not necessarily proper subsets) of islands.

Selection vs. Influence These t-communities and islands, these walking arenas, interact with two different social forces to create neighborhood communities and their effects. First, neighborhoods are more than just neighbors residing nearby each other. They are vital entities, or at least they have the potential to be. Even when it appears static, neighborhood social life generally is not; it is a stable equilibrium reached amidst the strife of social flows. The vibrant, living part of neighborhoods consists of the flow and exchange, both spoken and silently modeled, of norms, values, identities, symbols, ideas, affect, sentiment, and other social and cultural goods and resources among neighbors along the conduits provided by neighbor networks. This flow pressures neighbors towards conformity.

This certainly happens verbally, through the exchange of personal information, life histories, and stories, as well as through establishing and enforcing rules for neighborhood children; however, it happens even more nonverbally. Community members model and enforce "appropriate behaviors" in their daily interactions with each other, especially with children (Coleman 1990). In addition to modeling appropriate behaviors, neighbors may use rewards to encourage normative behaviors or sanctions to discourage behaviors not compliant with social or personal norms. The degree of these sanctions or rewards varies greatly with the nature of the society; for example, in communities that emphasize social control and social cohesiveness, sanctions may be enforced by direct social pressure for conformity (Hogan and Kitagawa 1985). As norms, values, ideas, and other social goods and resources traverse and commingle along neighbor networks, they have the potential to engender a sense of community and identity, social capital, mutual trust, social control, collective efficacy, and many other important facets of neighborhood life social researchers interest themselves in. The collectively efficacious community network which emerges, or which fails to emerge, however, is embedded in the network of neighborly interactions which is embedded in the network of potential neighborly interactions which is embedded in the network of geographic availability.

Second, locational choice and homophily may certainly account for some of the effects of neighborhoods but numerous studies have shown that neighborhoods with similar population demographics, in terms of race, socioeconomic status, family structure, and a host of other characteristics, often yield different outcomes for their constituent members. Market explanations, while intuitively appealing, have failed to account for the richness and complexity of effects correlating to neighborhoods.

Locational-based neighborhood effects such as residential differentiation and segregation correspond to the influence-based neighborhood effects such as social capital and collective efficacy because in choosing to move away from dissimilar households, residents are implicitly choosing to segregate their networks of potential neighborly interactions as well. Since contact is a necessary prerequisite for interaction, if households settle in such a way that their immediate neighbors are similar to themselves, then they have settled in such a way as to not have neighborly interactions with those different from themselves. Neighborhood communities result from both the concatenation of homophilous locational choices and the flow and exchange of norms, values, and beliefs among neighbors. Their correspondence is not additive, as in a regression model, but rather sequential. Relocation, which is responsible for residential differentiation and segregation, determines geographic availability and the potential for neighborly interactions and thus, of necessity, the actualized neighborly interactions which influence works upon to create social capital, collective efficacy, and other important neighborhood effects.

Children and Their Families Neighborhoods are especially important for households with children because children are much less mobile, and thus more geographically dependent, than adults. Children and their playful interactions depend upon proximity much more than adults and their interactions do. Since children cannot drive and have little, if any, voice in relocation decisions, they are forced to share lives with neighboring children even more than are their parents. For children, the street in front of their home is "the mediator between the wider community and the private world of the family" (Appleyard 1981, p. 4). This is where children first learn about the world. They often play games in the middle of these streets, use them to walk pets and to ride bicycles, and the majority of their recreational activity occurs there (Brower 1977). Sidewalks provide access between residence and schools and parks. As a result, the relationships children form will primarily depend upon the opportunities to interact provided by walking arenas immediately surrounding them (Appleyard and Lintell 1986). Especially for young children, neighboring children are the most likely to become their playmates (Hillier 1996). Thus, the networks of relationships they form will be much more dependent upon the network of geographic availability. Unlike children, adults have many venues for social relationships beyond their neighborhood including work and voluntary activities. School-age children may have some of these to the extent their parents allow. Preschool children, however, have few, if any, of these alternative venues for social opportunities. Their lives are tightly bound by geography.

"The micro-ecology of pedestrian streets bears directly on patterns of interaction that involve children and families. Parents are generally concerned with demarcating territory outside of which their children should not wander unaccompanied by an adult, to ensure that their children stay in areas that are safe for play and conducive to adult monitoring. To the extent that these limited spaces of children's daily activities usually do not cross major thoroughfares, defining tertiary communities may provide a foundation for constructing neighborhood indicators of child well-being and social processes more generally " (Sampson et al. 2002).

Not only are your neighbors' children predisposed to become your children's friends, but they also determine the character of your children's playmates (Cochran 1994) and the kinds of role models they emulate (Massey and Mullen 1984). The flow of norms and values discussed above acts not only on children but their families as well. "For example, when parents know the parents of their children's friends, they have the potential to observe the child's actions in different circumstances, talk to each other about the child, compare notes, and establish norms. Such intergenerational closure of local networks provides the child with social capital of a collective nature" (Sampson 2001: 9). As a result, households with children are far more influenced by the norms and values of surrounding households with children than households in general are influenced by the norms and values of their surrounding neighbors.

Neighboring parents may become intimately involved in the socialization of each other's children. Neighbors rear children side by side and together have the potential to co-create a safe and value-laden environment. Parents monitor their own children as well as those of their neighbors (Sampson et al. 1999). Some neighborhoods expect that residents share values and are willing and able to intervene on the behalf of children. They expect that residents will actively engage themselves in the support and social control of children (Bandura 1997) and that the community will work together to successfully support and control children. Parents get to know the parents and families of their children's friends; they observe children's actions, both their own and their neighbors', in a variety of circumstances; they talk with other parents about their children; and they establish norms (Coleman 1990). Such structural and normative adult-child closure gives children social support, provides parents with information, and facilitates control (Sandefur and Laumann 1998). The choice to live in a neighborhood is to some extent a choice to rear children together with one's neighbors. Ultimately, a community of parents may develop around the community of children, mirroring it. People whose children play together form friendship relations based in part on that fact (Abu-Gazzeh 1999; Grannis 2009). While it is the children who are immobile and thus confined to neighborhoods that are most immediately impacted by neighborhoods, children's geographic dependence encumbers their parents as well.

The Menu of Neighborhood Equivalents Altogether, for both children and adults, through influence and selection, neighborhoods cluster outcomes which cannot be accounted for in terms of the characteristics of the individuals or households currently residing in them. It is as if neighborhoods have personalities, enduring characteristic patterns that survive the replacement of their constituent members. Neighborhood communities are geographically constrained because the interactions which produce them are geographically constrained. Neighborhood communities are both geographically identifiable and have effects which persist through the replacement of their residents because the networks of interactions which produce them, which translate neighboring interactions into neighborhood communities, are constrained by predictable urban geographic substrates. To study efficacious neighborhood communities emerging from neighbor networks, therefore, we need a definition of a neighborhood community whose importance is derived from the potential for neighbor networks to concatenate within it.

Grannis (2009) defined this geographic availability in terms of shared walking arenas which mediate, guide, and constrain encounters. What do neighborhoods such as t-communities and islands, defined and measured by their potential for interactions, offer us that traditional neighborhood equivalents do not? They provide us with a lens to focus more closely on neighborhoods as communities emerging from the interactions of their constituent residents. They use an entirely different metric than census geography, one based precisely on the potential for community generating neighborly relations. In contrast, administrative geography often focuses on neighborhoods as statistical abstractions, perhaps reflecting segregation but agnostic to any potential for community, for interaction, for neighboring. While both types of neighborhood equivalents have their uses, researchers need to use care as to which one they choose and perhaps use both to disentangle different mechanisms that are at work in neighborhood communities, one mechanism provided by the concatenation of neighboring relations into neighbor networks and another provided by service areas, such as those offered by schools or marketplaces or police which unite residents around similar needs and opportunities. T-communities precisely measure the first. Neighborhood equivalents, defined solely by their boundaries, measure the second to the extent their boundaries coincide with the service areas. A careful use of both t-communities and neighborhood equivalents defined by their boundaries could tease apart the different mechanisms at work (Grannis 2009).

Related Entries

- ▶ Neighborhood Effects and Social Networks
- ▶ Social Network Analysis of Urban Street Gangs
- ▶ Spatial Models and Network Analysis

Recommended Reading and References

Abu-Gazzeh TM (1999) Housing layout, social interaction, and the place of contact in Abu-Nuseir, Jordan. J Environ Psychol 19:41–73

Alvanides A, Openshaw S, Macgill J (2001) Zone design as a spatial analysis tool. In: Tate NJ, Atkinson PM (eds) Modeling scale in geographical information science. Wiley, Chichester, Chapter 8

Anderson E (1992) Streetwise: race, class, and change in an urban community. University of Chicago Press, Chicago

Appleyard D, Lintell M (1986) The environmental quality of city streets: the residents' viewpoint. In: de Boer E (ed) Transport sociology: social aspects of transport planning. Pergamon, Oxford, pp 93–120

Appleyard D (1981) Livable streets. University of California, Berkeley

Bandura A (1997) Self-efficacy: the exercise of control. Freeman, New York

Brantingham PL, Brantingham PJ (1993) Nodes, paths, and edges: considerations on the complexity of crime and the physical environment. J Environ Psychol 13:3–28

Brooks-Gunn J, Duncan GJ, Aber JL (eds) (1997a) Neighborhood poverty, vol I, Context and consequences for children. Russell Sage, New York

Brooks-Gunn J, Duncan GJ, Aber JL (eds) (1997b) Neighborhood poverty, vol II, Policy implications in studying neighborhoods. Russell Sage, New York

Brower SN (1977) Streetfronts and backyards: two ways of looking at neighborhood open spaces. Department of Planning, Baltimore

Cochran S (1994) Understanding and enhancing neighbourhood sense of community. J Plann Literat 9(1):92–99

Coleman JS (1990) Foundations of social theory. Harvard University Press, Cambridge

de Jong R (1986) The recapture of the street? In: de Boer E (ed) Transport sociology: social aspects of transport planning. Pergamon, Oxford, pp 77–91

Festinger L, Schachter S, Back K (1950) Social Pressures in informal groups: a study of human factors in housing. Harper and Row, New York

Galster GC (2001) On the nature of neighborhood. Urban Stud 38(12):2111–2124

Grannis R (2009) From the ground up: how the layered stages of neighbor networks translate geography into neighborhood effects. Princeton University Press, Princeton

Greenbaum S (1982) Bridging ties at the neighborhood level. Soc Networks 4:367–384

Hillier B (1996) Space is the machine. Cambridge University Press, Cambridge

Hogan DP, Kitagawa EM (1985) The impact of social status, family structure, and neighborhood on the fertility of black adolescents. Am J Sociol 90:825–852

Lynch K (1971). The image of the city. MIT Press, Cambridge

Massey DS, Mullen B (1984) Processes of hispanic and black spatial assimilation. Am J Sociol 89:836–873

Massey DS, Denton NA (1993) American apartheid: segregation and the making of the underclass. Harvard University, Cambridge

Miller H (1999) Potential contributions of spatial analysis to geographic information systems for transportation. Geogr Anal 31:373–399

Morenoff JD, Sampson RJ, Raudenbush SW (2001) Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. Criminology 39(3):517–560

Openshaw S (1984) The modifiable areal unit problem. Geo Books, Norwich

Rabin Y (1987) The roots of segregation in the eighties: the role of local government actions. In: Tobin GA (ed) Divided neighborhoods: changing patterns of racial segregation. Sage, Newbury Park, pp 208–226

Sampson, RJ, Morenoff JD, Earls F (1999) "Beyond Social Capital: Spatial Dynamics of Collective Efficacy for Children". Am Sociol Rev 64:633–660

Sampson RJ (2001) How do communities undergird or undermine human development? Relevant contexts and social mechanisms. In: Booth A, Crouter N (eds) Does it take a village? Community effects on children, adolescents, and families. L Erlbaum, Mahwah, pp 3–30

Sampson RJ, Morenoff JD, Gannon-Rowley T (2002) Assessing "neighborhood effects": Social processes and new directions in research. Ann Rev Sociol 28:443–478

Sandefur RL, Laumann EO (1998) A paradigm for social capital. Rational Soc 10:481–501

Suttles G (1972) The social construction of communities. University of Chicago Press, Chicago

Social Network Analysis of Organized Criminal Groups

Francesco Calderoni
Transcrime, Università Cattolica del Sacro Cuore, Milan, Italy

Overview

This entry discusses the state of the art in the application of network analysis methods to the study of organized criminal groups. It provides an overview of the development of the field from both the academic and law enforcement perspective and discusses the current approaches in the literature. Particular attention is given to methodological issues (data sources, type of network analysis) and to the discussion of the limitations of the application of social network analysis to criminal organizations. In the light of the current state of the art, the entry also discusses the recurrent claim that network analysis may help the law enforcement agencies to more efficiently disrupt criminal networks. Finally, it attempts to identify the future trends in the application of network methods in the study of organized crime, while suggesting some promising paths from both a research and policy perspective.

Introduction

The use of social network analysis (SNA) in criminology is a relatively recent trend compared to other social sciences. Among the multiple possible applications of SNA in criminology, it may be hard to identify trends and specific fields. Carrington (2011) has identified three main topic areas where network methods have found significant application in criminology: The first is the analysis of the influence of personal networks on delinquency and crime which, according to the author, is "the most common use of social network analysis in criminology" (Carrington 2011, 236). The second area is the influence of neighborhood networks on crime in the neighborhood. The third area is the organization of criminal groups and activities. In turn, the latter may be framed into three fields: The first concerns the application of SNA methods to terrorist networks; the second deals with street gangs, youth gangs, and delinquent groups (Carrington 2011, 244–246); and the last, discussed in this entry, concerns organized criminal groups.

The application of SNA methods to organized crime has a background which is based on the history of both research and law enforcement.

From the research side, the idea that organized crime may be better understood as a network rather than a hierarchical and structured organization is not new in criminological literature. The reaction against the alien conspiracy approach, which suggested to analyze organized crime as a bureaucratic organization, organized along a formal hierarchy and with detailed rules for its functioning, soon led to the alternative hermeneutic perspectives leaning towards more flexible and informal mechanisms. For example, Albini argued that organized crime is "a system of loosely structured relationships," mainly based on patron-client relations (1971, 288); the works of Ianni suggested that mafia-type organizations should be better understood as social systems based on shared social cultural and ethnical relations. He explicitly mentioned network analysis, as "an anthropological tool that is used to chart social interactions" (Ianni 1973, 4). Ianni analyzed a number of Puerto Rican and African-American criminal groups as networks, although the application was quite far from current use. Overall, although the concepts of network and network analysis were recurrently evoked to describe the functioning and structure of organized crime, there were very limited empirical applications using network analysis methods (Ianni 1973).

From the law enforcement side, since the mid-1970s there has been a growing attention on the processing and analysis of intelligence data in organized crime. In this context, link analysis (visual representation of the structure of a criminal group performed through manual or computer-assisted drawings) was increasingly applied by law enforcement agencies and a private industry quickly developed offering methodologies and training courses (Lupsha 1983, 63). Link analysis allowed to "establish the relationships that exist among individuals and organizations from bits and pieces of available evidence" (Harper and Harris 1975, 158). It became increasingly popular thanks to the development of intelligence software. The step from link analysis to SNA was relatively short.

Academic interests on organized crime networks gradually met with law enforcement network analysis attempts. Not surprisingly, Ianni and Reuss-Ianni, who had refuted the bureaucratic approach to organized crime for the social system approach, contributed to a volume on criminal intelligence analysis suggesting it be applied to organized crime (Ianni and Reuss-Ianni 1990). Indeed, from the very beginning, this idea was linked with its operational exploitation for law enforcement, at a time when the focus on strategic intelligence and analysis of organized crime was very strong.

Indeed, the first contributions in this field, dating back to the first half of the 1980s, already focused on the opportunities offered by network analysis in the strategic analysis and enforcement of criminal organizations (Davis 1981; Lupsha

1980, 1983). In his 1980 contribution Peter Lupsha suggested that effective enforcement of organized crime required a move towards strategic analysis of groups and operations. He suggested some “first steps that departments might engage in to move from a tactical to a strategic analysis perspective” (1980, 37). Among such steps, he included network analysis as an “essential and necessary step” (1980, 38). One year later, FBI Special Agent Roger Davis provided a first hypothetical example of the application of network methods to a criminal organization, using some basic network concepts, such as density and centrality, to a fictional criminal organization (Davis 1981, 18). One of the first empirical network analyses of a criminal organization was carried out by Lupsha (1983). The study focused on “The New Purple Gang,” a group trafficking in heroin and cocaine active in the New York City metropolitan area in the second half of the 1970s and composed of Italian-Americans with connections to La Cosa Nostra. The author maintained that his study was an application of network analysis and actually developed matrixes and graphs. Rather, the study merged broad concepts familiar to network analysis, such as reciprocity and ego network structure, with other approaches and data. Interestingly, some of the most common network analysis concepts, such as centrality, were barely mentioned. As a result, Lupsha’s study provided a first exploration of network analysis of organized crime, demonstrating how multiple methods could help to extract relevant information from police data.

In the following years, some scholarly contributions advocated the application of network methods to criminal organizations (among the first contributions Ianni and Reuss-Ianni 1990; Sparrow 1991). However, most of these nowadays “classic” contributions, albeit with some exceptions (e.g., Lupsha 1983), did not engage in empirical analysis of criminal groups, possibly due to the limited availability of datasets and software. Interestingly, Mastrobuoni and Patacchini (2010) and Papachristos and Smith (2011) have recently demonstrated that old law enforcement databases and archives may provide valuable opportunities for the application of SNA to organized crime. Not surprisingly, only 10 years ago, Coles still complained about the “failure by criminologist to adopt Social Network Analysis techniques and concepts in the investigation of criminal networks, particularly of organized crime”.

Only in the last decade the use of network analysis in the study of organized crime has seen significant developments. Since the beginning of the 2000s, interest in this specific field has significantly increased and contributed to opening new research directions in the study of criminal organizations. Indeed, the number of articles and papers on criminal, illicit, or, more generally, “dark” networks has grown only since the 2000s. Significantly, a special issue of a specialized academic journal was dedicated to SNA and organized crime (*Trends in Organized Crime*, Volume 12, Number 2/June 2009). Studies on criminal organizations are now an important sector in this increasing trend. The following sections of this entry attempt to review the current state of the art of the application of SNA to organized crime groups and to discuss present issues and future trends.

Criminal Organizations and Markets

In a very rough categorization, most current applications of SNA to organized crime have adopted either a “micro” or a “macro” approach.

Studies within the “micro” approach have normally focused on one organization or network with a relatively low number of nodes. One of the first examples was Natarajan’s study of a cocaine trafficking organization (2000). The study provided a detailed analysis of the structure and functioning of a criminal organization, through a mix of multiple methods (contact analysis, task analysis, status analysis, and network analysis). As to SNA, Natarajan applied some network concepts (centrality and density) and demonstrated that a systematic quantitative analysis of law enforcement and judicial sources could complement more traditional research methods in the study of higher-level drug trafficking.

In the following years, further empirical works have contributed to analyzing criminal organizations using network analysis. Most of them focused on drug trafficking (Morselli 2005, 2009; Calderoni 2012; Malm et al. 2011; Bright et al. 2012), but some studies analyzed car-rigging organizations (Morselli and Roy 2008), extortion/money-laundering activities (Varese 2006), or the structure or evolution of a prominent criminal within a New York mafia family (Morselli 2005).

As to drug trafficking, the application of SNA has provided unprecedented insight in the structure of trafficking organizations. In most drug-trafficking networks, a subset of nodes concentrated the majority of criminal contacts and communications. This emphasized that drug trafficking was mostly conducted by a subset of highly connected and active individuals who were able to bring different people and resources together. In general, these subjects were skilled criminal brokers tasked with the management of drug-smuggling operations. Conversely, the majority of nodes involved in drug-trafficking groups appeared to be marginal players, involved only in one specific operation or task. On the one side, the results revealed a specific organizational structure of drug trafficking. This was a rather horizontal structure, ideal for a flexible and fast-changing environment, where criminal activities and illicit partnerships depended more on criminal opportunities than on any internal organizational arrangement. On the other side, the mentioned structure did not correspond to an extremely elaborated hierarchy where different organizational layers/ranks could be observed and where bosses delegated to underbosses or capos who in turn command soldiers. Signals of structure and hierarchy did not emerge from the network analysis, but rather from complementary methods highlighting the informal division of tasks and differences in status among the participants. Overall, leaders/brokers were actually in contact with most of the other individuals in the criminal networks and maintained direct control over the criminal activities. Hardly any traces of delegation and rigid organizational layers were found, and this is probably due to the highest risks that may be faced due to information dispersion, lack of trust, and difficulties in controlling subordinates.

This picture was substantiated by evidence coming from different studies on drug networks from a variety of countries and typologies of criminal organizations, such as the United States, Canada, Spain, Australia, the UK, and Italy. These findings were consistent with, and provided further empirical evidence to, previous drug-trafficking research.

Studies within the “macro” approach have developed in very recent years, possibly because they were based on larger and more complex databases. They have generally analyzed larger networks, focusing on specific national or regional criminal markets or offender categories. This research has aimed at analyzing specific aspects of criminal networks and population. Malm and Bichler studied a network of 2,197 individuals derived from law enforcement intelligence data in Canadian Pacific Region (Malm et al. 2009; Malm and Bichler 2011; Malm et al. 2011). They used multiple networks (co-offending, kinship, formal organization, and legitimate association) and this allowed to focus on the interplay of different types of connections (e.g., family relation or partnership in legal business) within a large network of individuals involved in criminal organizations (Malm et al. 2009). In another study, they focused on a subset of 1,696 individuals involved in drug trafficking and analyzed the network characteristics of specific market niches (i.e., specific tasks within the drug-trafficking chain) (Malm and Bichler 2011). They have also explored the extent of co-offending patterns among criminal groups of different (ethnic or nonethnic) origin (Malm et al. 2011). Heber (2009) studied a sample of 127 serious drug offenders (principal offenders) in Sweden and further enlarged her network including the principal offenders’ co-offenders and eventually the co-offenders of the latter. As a result, her network included several thousands of offenders allowing an analysis of their distribution across the Swedish territory and of the structure of drug-trafficking activities. Some further contributions focused on American organized crime, namely, in Al Capone's Chicago (Papachristos and Smith 2011) and La Cosa Nostra in 1960 (Mastrobuoni and Patacchini 2010).

Overall, network studies have pointed out that the majority of drug-trafficking (and other criminal) groups are flexible and rapidly changing organizations, hardly fitting the typical pyramidal picture which is frequently suggested by the media and sometimes by law enforcement. Network analysis has contributed critically to the production of empirical evidence in this field.

Methodologies

Compared to the most advanced application of SNA in other social sciences, the methods currently applied in most criminological contributions may appear somewhat simpler. This may be due to the relatively recent development of network approaches in this specific field. In turn, this may imply a limited experience of most criminologists in the application of SNA but also, and more critically, a limited scientific acceptance of such methods within the discipline. Not surprisingly, in the past years most criminologists applying SNA have experienced difficulties in publishing in mainstream peer-reviewed journals. Another reason for the relative lag in the application of more complex network methods to organized crime may inherently relate to the this specific field. Indeed, available data on criminal organizations and criminal markets rarely allow to apply more complex analyses such as dynamic network analysis (DNA), multiplexity (multiple types of relations, e.g., co-offending, phone communications, and kinship), and exponential random graph models which would allow to study the evolution of organized crime networks in time and structure. Notwithstanding these difficulties, some studies have applied ERGMs and multiplex relations (Malm et al. 2009) and one recent paper announced further developments in the field of organized crime (Papachristos and Smith 2011).

Concerning data sources, the majority of studies relied on secondary data. Most micro studies used judicial documents, intelligence reports, or investigation files. Macro studies have used intelligence databases collected by law enforcement agencies (Malm and Bichler 2011; Malm et al. 2009; Heber 2009) or databases created through archival analyses (Mastrobuoni and Patacchini 2010; Papachristos and Smith 2011).

As for the types of SNA, most of the literature on organized crime groups has worked with complete (or whole or full) networks rather than the ego networks, differently from other main areas identified by Carrington (2011). Indeed, the complete network approach is naturally suited to the analysis of criminal organizations or criminal markets. Nevertheless, some studies used ego network approaches for studying the career of a prominent La Cosa Nostra member (Morselli 2005) or the impact of drug dealers' personal networks on their survival and success (Bouchard and Ouellet 2011) and the co-offending patterns among different typologies of criminal enterprises (Malm et al. 2011).

Within the full network perspective, the overwhelming majority of the reviewed contributions have focused on one-mode networks with individuals as nodes. Exceptionally, some contributions used two-mode networks with individuals and specific positions in the drug-trafficking chain (Bouchard et al. 2010), while another study, focusing on the world drug-trafficking market, elaborated a full network with world countries as nodes (Boivin 2011).

As already pointed out by Carrington (2011, 244), most network analyses of organized criminal groups are not particularly concerned with the testing of specific criminological theories. Nevertheless, most studies are concerned with the verification of hypothesis and concepts suggested by previous literature on organized crime (Natarajan 2000, 2006; Varese 2006; Morselli 2009; Malm et al. 2011). The existing literature shows a wide variety of approaches and different methods as to the network measurements and concepts. The most popular concepts are density and centralization, the analysis of subgroups (e.g., factions, cut-points, core-periphery), and the centrality of individuals (e.g., degree, betweenness, flow betweenness, eigenvector, closeness, clustering coefficient). Some contributions have merged network analysis with other methodologies such as script analysis (Morselli and Roy 2008), the content analysis of transcripts (Natarajan 2000; Varese 2006; Natarajan 2006; Calderoni 2012; Campana 2011), or spatial analysis (Malm et al. 2008).

Finally, a number of contributions discussed the theoretical, methodological, and practical repercussions of the application of SNA to organized crime groups (e.g., Morselli 2009). These works signal the ongoing meditation about the implications of network analysis in the study of organized crime.

Problems and Issues

One of the main problems encountered in studies applying social network analysis to organized crime is data validity and reliability. The different data sources used in the literature are all exposed to possible biases. Indeed, most of them are secondary sources originating from judicial cases, law enforcement, or intelligence databases. Inevitably, the analysis based on such sources will reflect the perception of law enforcement. Interestingly, however, most of the abovementioned problems seem to be inherent to organized crime research in general. The biases affecting network analysis are often likely to occur also when using alternative methodologies. Indeed, scholars using network methods frequently reported difficulties in controlling for law enforcement biases, and they discussed the possible limitations of their analyses in detail (Morselli 2009, 44–50; Bouchard and Ouellet 2011, 83–85; Varese 2006, 45–47; Calderoni 2012; Malm et al. 2009, 70–71). A recent contribution further explored how the selection of different judicial sources affects the extracted networks (Berlusconi 2013). In general, authors appeared aware of the limitations of their studies. This has not spared criticism and skepticism towards the uncritical use of network methods in the study of organized crime. Undeniably, some contributions have exploited the growing interest in network analysis and have inserted it in the titles, abstract, or keywords, but without actually using any network analysis concept. This may suggest that the increasing use of network analysis may have generated some side effects, not excluding the unnecessary abuse of network-related terminology and techniques, just because this has become "fashionable." Besides such collateral effects, critics have rarely pointed out what alternative methods could replace the advantages provided by SNA. In fact, an increasing number of scholars are using network methods and concepts and applying them to the study of criminal organizations in different countries. This general trend suggests that SNA may actually bring significant added value to this field.

Concerning the risk of biases related to the law enforcement perception of crime, this appears to be a major issue particularly for studies adopting a macro approach. When trying to analyze regional or national markets, these studies most often reflect law enforcement knowledge of markets and criminal networks. The validity and reliability of the sources relate to the effectiveness of law enforcement and to the quality of the databases. Inevitably, however, any law enforcement source is inherently incomplete. Individuals and activities not (yet) uncovered or investigated by law enforcement will be missing (Sparrow 1991, 268; Bouchard and Ouellet 2011, 84; Malm et al. 2009, 70–71).

Also studies adopting a micro approach, focusing on specific groups or networks, may be affected by the mentioned problems of validity and reliability. However, these problems may have a lower impact than in the case of macro studies. Focusing on particularly long and detailed investigations is likely to reduce the risk of missing data in a criminal network. When law enforcement has been intercepting and monitoring a criminal group for months, if not years, the chances for some skilled criminal to avoid detection are likely to be low (Morselli 2009; Calderoni 2012). Nevertheless, the risks that law enforcement may have misunderstood or misjudged the relevance of specific nodes remain and so also the possibility of a willing omission by the police (e.g., this may be the case of informants or undercover agents which have not yet been disclosed).

Interestingly, recent studies have demonstrated that network properties and measures are strong even if tested for missing data (Morselli 2009, 48; Malm et al. 2009, 70–71). Although these findings are encouraging as to the validity and reliability of network methods applied to organized crime, there is a need of further research in this direction.

Social Network Analysis as a Tool for the Enforcement of Criminal Organizations

As already discussed, one of the main drivers of the development of network studies on criminal organizations has always been its possible application in law enforcement. Scholars have claimed the operational benefits of network analysis for intelligence and investigations for years. Given the increasing interest in network analysis of organized crime, a number of law enforcement agencies have started to apply SNA in their criminal investigations, and some contributions from law enforcement analysts acknowledged the use of network methods in the enforcement of criminal organizations. However, operational results so far appeared quite weak and many law enforcement agencies showed some skepticism towards the application of network methods in organized crime cases. Although this may be related to some difficulties in cooperation between law enforcement and academics in general (an issue which is not limited to the application of network analysis and which goes largely beyond the scope of this entry), discussion with law enforcement analysts (from Germany, the United Kingdom, the Netherlands) who have actually experimented network analysis in their job provides more detailed information. According to their experience, the current applications of network analysis are not likely to provide critical advantages to law enforcement agencies in their everyday investigations. Indeed, in long-lasting investigations, the police normally acquires a detailed knowledge of the case. Each individual is monitored, her/his background scanned, and phone and e-mails are frequently intercepted. The application of SNA can hardly provide any additional knowledge to the officers and prosecutors working daily on the case. The idea that SNA can effectively identify the most important actors and show the police the most efficient way to disrupt a criminal network does not take into account that the police normally has a far better insight on the case than any researcher. More often than not, scholars identify the most important individuals, the best targets for network disruption, and the "usual suspects" that law enforcement had already identified.

This may explain the skepticism so far shown by some law enforcement professionals (some of which are increasingly being trained in SNA methods) when claims are made that SNA can help the police to identify the best targets. Network analysis can assist a researcher in reconstructing a network and in making sense of the structure and functioning of a criminal organization. While these analyses have provided unprecedented academic insight on the structure and functioning of criminal organizations, the risk is that current applications may have limited interest for law enforcers.

The above considerations should not lead to hasty conclusions about the futility of SNA and the enforcement of organized crime. Rather, they suggest that the exchange of experiences between academics and law enforcement should be improved. Notwithstanding the operational limits of current analyses, SNA concepts and methods have gained the interest of many law enforcement agencies around the world. As with most new techniques, it will take time to have it accepted and applied. The future trends discussed in the next section suggest that the use of SNA in the enforcement of criminal organizations may become a reality.

Future Issues and Trends

As already mentioned, the application of SNA to organized criminal group is a very recent phenomenon. Therefore, it is likely that further developments will come in the next years. Some trends may be reasonably anticipated.

The number of publications will increase. The rate of publications applying SNA to criminal organizations has been significantly growing, and this trend should continue, if not increase, in the coming years. Hopefully, this may improve the overall quality and standards in this specific field. It will also prevent and eventually remove the inappropriate exploitation of the word network analysis in contributions which do not actually make any use of SNA methods and concepts. Eventually, this may provide SNA applications wider academic acceptance in this specific field and reduce skepticism about the network methods.

Further, the increase in the number of published works will likely provide more empirical studies of specific organized crime groups (micro approach) and/or markets (macro approach). At the moment, the promising results produced by the first papers need further support or verification with other groups and countries. The replication of methodologies and analyses will play an important part until a critical mass of criminal networks and datasets will be analyzed and available.

The methodology will improve. As already mentioned above, so far most SNA studies on organized crime have applied relatively basic measures and concepts. Although this is not necessarily a problem, since also very basic analyses have provided valuable results, the forthcoming years will likely bring new, more complex network methods to be applied also in the study of organized crime groups. Inevitably, the main challenge in this direction will be data availability and, ultimately, the improvement of data collection procedures by law enforcement. The availability of new and better data sources will likely allow the application of advanced SNA techniques such as dynamic network analysis, multiplexity, and ERGMs. These will provide new avenues for analyzing the structure of complex networks and their evolution through time.

In addition to more complex network techniques, future studies will likely mix network methods with other methods, both quantitative and qualitative. Network analysis can be complemented with other quantitative techniques such as spatial analysis and hotspot analysis. Furthermore, the inherently quantitative approach implied in SNA may provide the best results when balanced by qualitative analyses. Studies within the micro approach may take advantage of a detailed analysis of the case studies, including interviews with law enforcement to verify possible biases, interviews with network participants, or content analysis of judicial sources and transcripts (Campana 2011). As anticipated by Morselli, "content analysis is indeed the next step toward enriching the various analyses conducted throughout this book" (Morselli 2009, 164).

SNA and the enforcement of organized crime groups. Notwithstanding the issues highlighted in the previous section, future improvements of network applications to organized crime groups may provide interesting developments for the operational/enforcement point of view. In particular, SNA may be applied not only within investigations, but across different investigations. This contribution may have a geographical and a chronological dimension.

From the geographical perspective, criminal organizations are frequently operating in different areas which may not always be within the jurisdiction of the same law enforcement agency. This is particularly true for large criminal organizations such as mafia-type associations which may commit crimes in different countries. Inevitably such extended criminal activity is likely to involve a number of different police forces and prosecutors' offices, with increasing problems in coordination and information sharing. Different cases and files are going to be opened. It may happen that a marginal subject in the investigation led by District A shows up to be a high-ranking criminal in the investigation conducted by District B. The application of SNA methods across investigations may provide critical information which would not otherwise be available to the single investigators/prosecutors. Through network analysis it may be possible to merge investigative data from a number of cases and provide an overall picture which may be otherwise too complex or difficult to achieve. The literature has already provided some examples of SNA studies merging data from different investigations (Morselli 2009, Chaps. 8 and 9). Results demonstrated that network structure and positioning may significantly change when different cases are merged.

From the chronological perspective, some criminal investigations are capable to survive prosecution and convictions. Frequently, family ties and kinship provide important connections among the members, and this allows such criminal groups to last over years. Mafia-type organizations (e.g., the five families in New York) may last for years, notwithstanding constant law enforcement monitoring and repeated prosecutions. In these cases, it may be important to achieve a very good knowledge of the history of a criminal group. Surely, prosecutors and policemen achieve excellent knowledge of the most dangerous groups in their jurisdictions, but in some cases this important knowledge asset may be dispersed due to reorganization, retirement, transfer, or resignation. The application of network analysis may provide important strategic information about the development and structure of a single criminal organization through time. Furthermore, SNA may allow to assess the changes and reactions of criminal organizations as a reaction to law enforcement. In this perspective, Morselli and Petit showed that systematic drug seizures by law enforcement had a significant impact on a criminal organization, although the group was to continue its operations for months (Morselli and Petit 2007). Further studies may analyze the impact of arrests (either random or targeted) on criminal organizations, applying dynamic network analysis techniques to organized crime groups. Ultimately, this may allow law enforcement agencies to estimate the impact of different strategies and investigations on the evolution of the structure of large and continuing criminal organizations. SNA methods may represent a valuable quantitative benchmark in this evaluation.

In conclusion, organized crime is one of the last criminological areas where SNA methods have been applied. If Carrington has argued that "the use of social network analysis in criminology is in its infancy," this is even more relevant for the use in the study of organized criminal groups. In the last 10 years, important steps have been made. The progress are encouraging and it is likely that, through the trial and error which is one of the basic methods of science, SNA will play an important part in the study of organized crime in the next years.

Related Entries

- ▶ Gangs and Social Networks
- ▶ Neighborhood Effects and Social Networks
- ▶ Network Analysis in Criminology
- ▶ Social Network Analysis and the Measurement of Neighborhoods
- ▶ Social Network Analysis of Urban Street Gangs
- ▶ Spatial Models and Network Analysis

Recommended Reading and References

Albini JL (1971) The American mafia; Genesis of a legend. Sociology series. Appleton, New York

Berlusconi G (2013) Do all the pieces matter? Assessing the reliability of law enforcement data sources for the network analysis of wire taps. Global Crime 14(1): 61–81.

Boivin R (2011) Drug trafficking networks in the world economy. In: Montréal. http://www.erdr.org/textes/boivin.pdf

Bouchard M, Ouellet F (2011) Is small beautiful? The link between risks and size in illegal drug markets. Glob Crime 12(1):70–86

Bouchard M, Descormiers K, Gallupe O, Wang W (2010) Panel: analyzing illegal drug supply chains as 2-mode social networks. Paper presented at the crime and social institutions, 62nd annual meeting of the American society of criminology, San Francisco, 17 November

Bright DA, Hughes CE, Chalmers J (2012) Illuminating dark networks: a social network analysis of an Australian drug trafficking syndicate. Crime Law Soc Change 57(2):151–176. doi:10.1007/s10611-011-9336-z

Calderoni F (2012) The structue of drug traffcking mafias: The Ndrangheta and Cocaine. Crime Law Soc Change 58(3):321–349.

Campana P (2011) Eavesdropping on the mob: the functional diversification of mafia activities across territories. Eur J Criminol 8(3):213–228

Carrington PJ (2011) Crime and social network analysis. In: Scott JP, Carrington PJ (eds) The SAGE handbook of social network analysis. SAGE Publications, London, pp 236–255

Davis RH (1981) Social network analysis: an aid in conspiracy investigations. FBI Law Enforce Bull 50(12):11–19

Harper WR, Harris DH (1975) The application of link analysis to police intelligence. Hum Factors 17(2):157–164

Heber A (2009) The networks of drug offenders. Trends Organ Crime 12(1):1–20

Ianni FAJ (1973) Ethnic succession in organized crime: summary report. U.S. Government Printing Office, Washington

Ianni FAJ, Reuss-Ianni E (1990) Network analysis. In: Andrews PP, Peterson MB (eds) Criminal intelligence analysis. Palmer Enterprises, Loomis, pp 67–84

Lupsha P (1980) Steps toward a strategic analysis of organized crime. The Police Chief 47:36–38

Lupsha P (1983) Networks Vs. networking: analysis of an organized crime group. In: Waldo GP (ed) Career criminals. Sage, Beverly Hills

Malm AE, Bichler G (2011) Networks of collaborating criminals: assessing the structural vulnerability of drug markets. J Res Crime Delinq 48:271–297

Malm AE, Kinney JB, Pollard NR (2008) Social network and distance correlates of criminal associates involved in illicit drug production. Secur J 21(1–2):77–94

Malm AE, Bichler G, Van De Walle S (2009) Comparing the ties that bind criminal networks: is blood thicker than water? Secur J 23(1):52–74

Malm AE, Bichler G, Nash R (2011) Co-offending between criminal enterprise groups. Glob Crime 12(2):112–128. doi:10.1080/17440572.2011.567832

Mastrobuoni G, Patacchini E (2010) Understanding organized crime networks: evidence based on federal bureau of narcotics secret files on American mafia. Carlo Alberto notebooks 152. Collegio Carlo Alberto, Moncalieri

Morselli C (2005) Contacts, opportunities, and criminal enterprise. University of Toronto Press, Toronto

Morselli C (2009) Inside criminal networks. Studies of organized crime. Springer, New York

Morselli C, Petit K (2007) Law-enforcement disruption of a drug importation network. Glob Crime 8(2):109–130

Morselli C, Roy J (2008) Brokerage qualifications in ringing operations. Criminology 46(1):71–98

Natarajan M (2000) Understanding the structure of a drug trafficking organization: a conversational analysis. In: Mangai N, Mike H (eds) Illegal drug markets: from research to prevention policy, vol 11, Crime prevention studies. Criminal Justice Press/Willow Tree Press, Monsey, pp 273–298

Natarajan M (2006) Understanding the structure of a large heroin distribution network: a quantitative analysis of qualitative data. J Quant Criminol 22(2):171–192

Papachristos AV, Smith CM (2011) The small world of Al Capone: the embedded nature of criminal and legitimate social networks. In: Third annual illicit networks workshop, Montréal. http://www.erdr.org/textes/papachristos_smith.pdf

Sparrow MK (1991) Network vulnerabilities and strategic intelligence in law enforcement. Int J Intell Counter Intel 5(3):255–274

Varese F (2006) The structure of a criminal network examined: the Russian mafia in Rome. Oxford Legal Studies Research Paper No. 21/2006

Social Network Analysis of Urban Street Gangs

Jenny C. Piquette[1], Chris M. Smith[2] and Andrew V. Papachristos[3]
[1]Department of Sociology, The Ohio State University, Columbus, USA
[2]Department of Sociology, University of Massachusetts Amherst, Amherst, USA
[3]Department of Sociology, Yale University, New Haven, CT, USA

Overview

This entry synthesizes the current state of network-related research on street gangs around the themes of network structure, network action, and network location. At their core, street gangs are social networks created by the coming together, socializing, and interacting of individuals in particular times and in particular places. The employment of social network analysis has the potential to examine patterns of interaction among gang members and gangs, illuminate structural variation across gangs, and measure the influence of gang networks on individual action. This entry includes an overview of social network analysis, suggestions and directions for future gang network research, a discussion of limitations, and a vision for how a social network approach to gangs might inform theory, research, and practice.

Introduction

Over the past decade, the field of criminology has increasingly employed both the theoretical and methodological tools of social network analysis (McGloin and Kirk 2010; Papachristos 2011). Although criminology lags behind other disciplines in this network turn, criminologists have applied social network analysis to investigate organized crime syndicates, narcotics trafficking, terrorist organizations, white-collar conspiracies, and adolescent delinquency. But perhaps one of the areas of criminological inquiry most in need of a network-related approach is the study of street gangs.

Street gangs are, if anything, social networks created by the coming together, socializing, and interacting of individuals in particular times and in particular places. Though scholars still heavily debate the definition of what constitutes a gang, virtually all gangs have some element of a group and it is this groupness that is at the heart of this entry. Street gangs are groups of individuals engaging in activities that are bolstered by group processes, culture, and structures. Social network analysis studies street gangs *empirically* to examine precise patterns of interaction among individuals and collectivities (see Papachristos 2006). *Theoretically* gangs operate as unique sociological entities with various structural and interactional manifestations. Research gathers, compares, and analyzes dozens of network properties at the individual, group, and system level, which brings fresh perspectives to old debates. Contemporary social network approaches have the potential to push past criminology's definitional bulwarks and typologies and reveal how real-life behavior of individuals and groups might inform theory, research, and practice.

This entry synthesizes the current state of network-related research on street gangs. Rather than simply plod forward in the standard literature review format of "this article says this" and "that article says that," the aim will be to review research in a way that provides a road map for future lines of inquiry. Specifically, the focus will be on measuring and theorizing street gangs from multiple levels of analysis in order to shed light on a broader range of empirical gang behaviors, structures, and processes. To wit, this entry is structured around three themes: (1) the structure of gang networks, (2) the actions of gang networks, and (3) the location of gang networks. In the following sections particular attention will be paid to the group nature of gangs as well as the locations and actions of individuals in gangs. This entry concludes with a brief discussion of the challenges to and limitations of social network analysis and gang research. To begin, the next section provides a brief discussion of social network analysis more generally and proposes its application to street gangs.

Social Network Analysis and Applications to Street Gangs

Social network analysis (SNA) refers to the study of social relationships among sets of actors and, more importantly, the analysis of how the patterns of relationships affect actors' behaviors (Wasserman and Faust 1994). Unlike traditional regression analyses in which effects are net of other variables (Emirbayer 1997), SNA assumes *interdependence* rather than *independence*. Indeed, it is such interdependence that is of analytic interest. Most network studies attempt to (a) analyze and describe the observed structure of relationships among a set of actors and/or (b) investigate how sets of relationships affect actors' behaviors, actions, opinions, and attitudes. Empirical research has demonstrated the importance of networks on a variety of phenomena such as getting a job, income attainment, political decision making, the diffusion of ideas, and even political revolutions.

Formally, social network analysis relies on graph theory: mathematical models that capture the pairwise relationships among a specific class of objects (Wasserman and Faust 1994). In graph theoretical terms, a network consists of a set of vertices (V) that represent a bounded set of actors or units and a set of edges (E) that define the relationship among the vertices. Relationships are commonly represented in a square matrix $X = x_{ij}$ where x_{ij} is the strength of relationship between unit i and unit j (Wasserman and Faust 1994). Edges can be binary indicating the presence of a relationship. Alternatively, edges can be valued measuring the intensity of a relationship, the

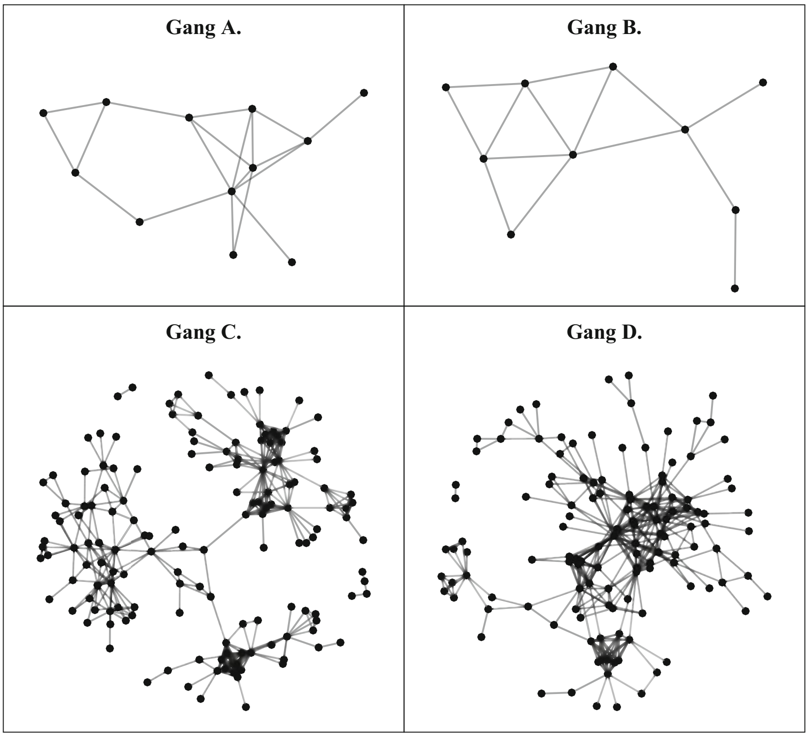

Social Network Analysis of Urban Street Gangs, Fig. 1 Four African American gang networks from Chicago's west side

level of interaction, or any other theoretically sound quantity of a relationship. Ties can also have direction, such that one node sends a tie and another receives it. Sociograms or graphs display networks visually in which nodes depict the vertices and lines depict the edges (see, e.g., Fig. 1).

Conceptually, gangs are clearly social networks. Gangs consist of actors who are tied to each other through group participation and identity, as well as neighborhood, familial, and peer relationships. Furthermore, gangs collectively act in particular moments and places, whether simply hanging out or engaging in gang warfare. Of particular interest for researchers is a gang's ability to activate ties in order to accomplish certain tasks and the ensuing consequences of those tasks. Figure 1 illustrates this point by mapping the network structures of four African American gangs in Chicago using police observational data. Each node represents a unique individual and each tie represents a non-arrest observation made by the police – sometimes called "field observations" or "field contacts" (see Papachristos et al. 2012a). Each tie is a police observation of two individuals hanging out on a street corner, which is perhaps one of the most fundamental gang activities. By aggregating such information over time and within multiple locations, a network structure emerges representing the patterns of direct and indirect associations among gang members as well as their non-gang companions.

All four of the gangs in Fig. 1 are from the west side of Chicago and are part of larger gang nations that date back to the 1960s. By virtually any survey or qualitative account, all four of these gangs should have the following: a hierarchical structure, formalized sets of rules and regulations, a division of labor, and intricate drug-dealing portfolios. Yet, the patterns of association shown in Fig. 1 reveal that some of the networks deviate from such accounts. Gangs A and B, for example, are smaller in size and less dense than gangs C and D. Moreover, the network structures of gangs A and B look more like ordinary friendship networks and are decidedly *not* hierarchical. In gang B, triangles of individuals overlap to make up the larger networks; this pattern suggests that the network is comprised of small cliques with overlapping membership as opposed to some formalized division of labor. In contrast, gangs C and D have a starburst type of structure with dense pockets of individuals linked together by one or two ties. These gangs are more centralized, a trait that is especially evident in gang C with a single individual joining three different parts of the graph.

The analysis of gang structures such as those in Fig. 1 can augment more traditional analyses of gangs by comparing structural traits of networks across time and space. A whole range of network measures and statistics – such as density, clustering, and degree distribution (Wasserman and Faust 1994) – can be compared and analyzed across samples of gangs to detect points of variation and similarity. Perhaps more importantly, SNA can examine the structural properties that distinguish a gang from other types of social groups.

A networked approach to gangs requires two related and nested units of analysis: (1) gangs as collectives and (2) gangs as a collection of individual gang members. The questions gang researchers ask rely on the unit of analysis. At the group level, researchers examine organizational structures, gang interactions within larger social spaces, and intragroup relations affecting larger social processes. Individual-level hypotheses center on relations between gang members and gang influence on individual behavior. Research agendas might include (1) how friendship networks within a gang mediate or facilitate violence, (2) the types of relationships from which gang leadership or influence derive, (3) patterns of social status and power among gang members, or (4) the ways in which gang members integrate into lager noncriminal networks (see Papachristos 2006). The remainder of this entry grapples with some of these hypotheses, which are organized around the themes of network structure, network action, and network location.

Network Structures of Street Gangs

Though not specified in formal network language, researchers' conceptualization of street gangs has relied on network concepts and ideas for decades. At the most basic level, gangs can be conceptualized as groups of delinquent individuals who socialize with each other. Rather than a random collection of strangers who happen to share social and geographic space, gangs have a collective identity, a culture, and sense of togetherness. As such, criminologists strive to understand group formation, cultural precepts, or collective processes. Prior research has provided tremendous insight into street gangs' culture, experiences, and understandings of their group, but has fared less well in explaining the structural variation and scope of street gangs. SNA, however, provides theoretical and methodological tools to evaluate gangs' organizational structure and participation in those structures.

Organizational Structure

Descriptions of the gangs as organizations range from informal friendships and associations (Fleisher 1999) to formalized corporate bureaucracies (Venkatesh and Levitt 2000). Multiple typologies have emerged attempting to classify gangs according to various organizational or group characteristics. For example, research has debated whether leadership and influence within a gang are vertical (bureaucratic) or horizontal (influence based) (Sanchez-Jankowski 1991). Other typologies differentiate gangs based on certain organizational characteristics such as membership size, meeting regularity, formalized rules and regulations, and gang activities (Bouchard and Spindler 2010; Pyrooz et al. 2012). Gangs are also frequently differentiated by the kinds of criminal behaviors in which they engage – i.e., violent gangs, income generating gangs, and delinquent gangs (Starbuck et al. 2001). Thus far, the majority of the research on gang structure has been largely descriptive and, therefore, limited from making topological comparisons.

SNA can advance this line of research by moving away from descriptive organizational characteristics to formal measures of group structures. This transition allows for testing organizational types, examining structural changes over time, as well as the ability to look at patterns of behavior, influence, and interaction. For example, network research on formal organizations suggests that informal relationships are often more significant in organizational behavior than formal hierarchical positions. Asking whether or not a gang has regular meetings overlooks the informal associations that may be at the heart of gang life. SNA has challenged conventional understandings of gang membership by showing how ties among individuals extend beyond the in-group and has also revealed more intricate relations within the larger gang structure (McGloin 2005). Frequently gang members are not friends, do not know one other, or have animosity toward one another. From a networked perspective, one builds the organizational structure from the ground up by tracing lines of influence and interaction among members.

A network approach to gangs has the potential to illuminate structural variation, model organizational changes over time, and examine organizational threat or stability. Figure 1, for example, presents structures representing actual lines of interaction among members, as opposed to a theoretical structure that may or may not fit the actual structure of the group. Formal analysis of networks sheds light on current typologies and develops new understandings of gang structures. One potential avenue of inquiry would compare the patterns of interaction within specified groups to group structures as reported by law enforcement or gang members. In many non-gang settings, often the informal networks of friendship have a stronger influence on behavior (see Haynie 2001). Another avenue of inquiry would compare structures at the group level in order to examine if particular network forms are more or less conducive to certain behaviors such as violence or drug dealing.

Structural Position

Not all gang members are alike. For decades, gang research has made important distinctions between the levels at which members participate in gang life and gang culture. Terms such as wannabes, hanger-ons, and associates typically refer to those individuals only causally involved in a gang, whereas terms such as hard core, down, or lifer are reserved for those more committed to gang life (Klein and Maxson 2006). Scholars often make the distinction between core and periphery gang members where the former refers to those who are active in the gang and the latter refers to members not central to the gang structure. The basic presupposition is that those who are more central to the gang structure are participating more fully in gang activities and those who are on the outskirts participate less.

Scholars also recognize the naïveté of such a core-periphery distinction. Gang participation is more amorphous than most typologies allow. Although popular media often depict gangs as secret societies with blood rituals and initiations, the reality is that gang membership is fluid (Fleisher 1999). Furthermore, membership is often unofficial; it is just as likely that membership is based on familial relations than a formal initiation. Some fully indoctrinated members may actually exert very little influence, while others who are not formal members may exert tremendous influence. Small cliques form within the group, which can at times divide the group into subdivisions, each with its own level of participation (Fleisher 1999; McGloin 2005; Papachristos 2006).

SNA can move research on gang participation forward by considering how positions within networks may or may not relate to levels of participation or vice versa. By empirically returning to basic gang networks (such as those in Fig. 1), analysts can use various network measures to find structural similarities or differences within and across groups. For example, one promising idea might be to analyze a sample of gang structures and test for structural equivalence in order to identify unique positions or roles in gangs. Another potential line of inquiry might locate sources of influence within gang structures – i.e., locating individuals, such as brokers or gatekeepers, who might be significant not because of formal titles or positions but because of their influence on member behavior.

Network Action of Street Gangs

Like other types of social groups, gangs can pursue collaborative goals, form alliances, and engage in collective action. Perhaps one of the most important areas of research on SNA and gangs investigates how gang network structures affect collective and individual action.

Gang Landscapes

SNA has the capacity to survey *gang landscapes*: the general patterns of relationships among a population or system of gangs (Kennedy et al. 1997). Social mapping of the collective relationships within a gang or across gangs offers insight into the organizational and territorial niches in which gangs exist. A gang landscape can map the alliances and rivalries between groups and individuals across a geographic space. Gang landscapes have the potential to situate specific gang activities and exchanges (such as violence and drug dealing) in the context of a total networked system. In other words a structural analysis illuminates gang actions and behaviors. For example, Descormiers and Morselli (2011) analyze how gang-level attributes (such as gang size, age of members, and ethnicity) and ecological factors (such as gang turf and underground economies) predict maps of gang alliances and conflicts in Montreal. In a slightly different approach, Papachristos (2009) maps the networks of violent exchanges between Chicago gangs over time and finds that prior relationships and the position of a gang in conflict networks are strong predictors of retaliation. This latter example demonstrates how patterns of the gang landscape provide insight into a social process of retaliation.

Group Processes

Gangs' group-level actions include protection, socializing members, and supplying resources. These behaviors imply a number of social processes that are negotiated within the gang. Prior research has revealed the potential mechanisms within gangs that contribute to their crime and delinquency – e.g., loyalty, status, and protection (e.g., Short and Strodtbeck 1963). Protection is of particular importance because survey research has consistently found that protection is one of the most commonly cited reasons for joining youth gangs (Klein and Maxson 2006). In the quest for mutual protection, gangs face a basic collective action problem: they must motivate members to actually come to each other's mutual aid in the face of a threat.

Network structures can facilitate or impede acts of mutual protection. In particular, the idea of *cohesion*, often measured as the density of ties within a network (Wasserman and Faust 1994), offers a fruitful direction for empirical analysis. A second approach to examining protection is by calculating the network density of gangs and correlating this with acts of violence – especially retaliation. For example, Papachristos (2009) found that gangs are more likely to reciprocate an act of violence when their actions are visible to other gangs. In other words, the collective processes involved in retaliation are more likely to occur when retaliation becomes a signal to other groups that a gang is capable of vengeance and protection. In addition to reinforcing the *esprit de corps* of the group, Papachristos (2009) finds such acts of retaliation can actually flame the spread of violence and create enduring networks of gang conflict that persist over time.

Research on gangs' group processes and the ensuing consequences has been virtually untouched. The launch pad for such inquiry should focus on how specific hypothesized processes might manifest themselves structurally or on how structures affect the hypothesized processes. Just as understanding the sociodemographic characteristics of neighborhoods allow one to understand how social processes ecologically unfold, understanding the networks of street gangs better unpacks the social context in which group processes unfold.

Group Influence on Individual Action

One of the fundamental questions in gang research is the extent to which gangs (as groups) actually facilitate individuals' delinquent or criminal behavior. This is in contrast to the selection effect perspective, which assumes like-minded and previously delinquent individuals come together to form a group (Bendixen et al. 2006; Thornberry et al. 2003). Mounting evidence from longitudinal studies supports a facilitation effect: individual delinquency is amplified in the gang context even when controlling for factors such as individual self-control, prior delinquency, and sociodemographic characteristics (Krohn et al. 2011; Thornberry et al. 2003). Individual delinquency, even among those with higher baseline delinquency, increases during periods of gang membership. However, what is still unclear is exactly how the gang facilitates crime and delinquency for the individual.

A growing body of network research shows that the structure of adolescent friendship networks is associated with self-reported delinquency, which can range from statutory crimes like underage drinking to more serious crimes like theft (Haynie 2001, 2002; McGloin 2009). In particular, this research finds considerable evidence for peer influence on delinquency: one's own delinquency is related to the attitudes, behaviors, and delinquency of one's peers. At the same time, only a handful of studies look at co-offending networks involved in serious crimes. Given the observed facilitation effect of gangs and the simple fact that gang associations appear more group-based than friendship networks, it is reasonable to hypothesize that the network effect of gangs on individual action should be larger than effects of friendship networks or co-offending networks. Furthermore, certain network configurations are more apt than others to facilitate imitation of behavior, to provide opportunities for deviance, and to exert stronger influence on its members. As such, we hypothesize that (a) the gang context amplifies group-level processes that foster delinquency and (b) gangs exhibit additional processes not found in delinquent peer groups.

Two recent studies by Papachristos and colleagues provide evidence for the group facilitation effect of gangs. In a study of gang members and their associates in Boston's Cape Verdean community, Papachristos et al. (2012a) examine a social network of more than 700 individuals, approximately 30 % of whom are gang members. Their study finds that individuals in the network are more likely to be gunshot victims when (a) they are gang members, (b) their immediate social network is saturated with gang members, and (c) their immediate social networks contain other gunshot victims. In other words, network exposure to gang members and gunshot victims increases one's own risk of being a victim. The second study surveys 150 active gun offenders in Chicago. Papachristos et al. (2012b) find that the network structure of individuals affects both their perceptions of the law and their subsequent law-violating behavior. In particular, individuals are more cynical of the law and subsequently more likely to engage in deviant behavior when their social networks are saturated with gang members. These two studies offer only a small glimpse into the ways network structure might influence individual actions and/or individual-level consequences. Future research should consider a wider range of group types (such as comparing gang members vs. non-gang members or members of other types of delinquent groups), different types of group processes (imitation, peer influence, and homophily), as well as various types of individual-level outcomes (beliefs, decisions, and behaviors).

Violence Prevention

From a social problem perspective, one of the most critical gang activities requiring attention is the incredibly high rate of lethal and nonlethal violence. A small but growing body of research has taken gang landscape mapping quite seriously as both a means of gang violence prevention and a research area. As part of a crime reduction initiative, Kennedy et al. (1997) mapped out gang rivalries and alliances in Boston. Violence intervention and prevention efforts relied on these network maps to direct attention toward gangs that were actively involved in shootings and target resources scarce to those gangs. This targeted group-based intervention strategy (of which gang network mapping was a key part) produced significant reductions in youth homicide and nonfatal shootings (Braga et al. 2001). Similar approaches evaluate the efficacy of group-based and outreach worker strategies in other US cities (Skogan et al. 2009; Tita et al. 2003).

Network Locations of Street Gangs

Gangs are decidedly local phenomena with strong connections to the character and culture of urban neighborhoods. In fact, part of sociology's continued fascination with gangs stems from their relationship to neighborhood social organization. Gangs are seen as de facto social institutions or collectivities that meet the basic needs of community and youth in otherwise disadvantaged communities. More than that, however, the neighborhood becomes a defining element of gang identity and group membership (Garot 2007; Grannis 2009). This association is so strong that many gang names incorporate their neighborhood of origins: the 35th Street Kings, Columbia Point Dogs, Westville Saints, and so on. Indeed, often the simple question "Where you from?" is a shorthand way for identifying gang members (Garot 2007). In a way, the gang itself becomes a reflection of many of the conflicts, cultures, and idiosyncrasies of the larger neighborhood.

Prior research has shown that space matters in shaping gang behaviors in at least two ways. First, and foremost, gang violence is commonly associated with the defense of gang turf or territory. Given the aforementioned attachment of gangs to urban space, threats to turf are considered threats to home, members' safety, and the identity of the group. Violence stemming from the protection of turf becomes as much a symbolic measure signaling the solidarity and reputation of the gang and community as well as an instrumental gesture in protecting a piece of land over which a gang claims ownership. Secondly and perhaps related to such matters, previous research also suggests that gang violence is spatially contagious; gang violence in a given neighborhood affects levels of violence in adjacent neighborhoods. In a sense, the group and dynamic nature of gang violence spills over into adjacent communities. Though such spatial effects are associated with crime and violence more generally, a recent study of gangs in Los Angeles suggests that the spatial effects of gang violence extend beyond immediately adjacent geographic areas to affect more geographically distant neighborhoods (Tita and Radil 2011; Tita et al. 2003).

To date, with few exceptions, such spatial analyses of gangs rely mainly on regression models of aggregate crime measures – i.e., the extent to which crime rates in a focal neighborhood are associated with rates in surrounding communities. Yet, crime rates themselves do not move across gang turf or other geopolitical boundaries: gangs and gang members do. Nearly all empirical studies on the spatial nature of crime theoretically acknowledge that the social networks of criminal activity – such as drug dealing, prostitution, and gangs – drive the observed contagion of crime across neighborhood boundaries (e.g., Cohen and Tita 1999; Morenoff et al. 2001). Yet only a handful of studies actually incorporate social networks and geographic space into the same analysis, and even fewer studies have applied both to the study of gangs.

The inherent networked nature of gangs and their strong links to urban space propose an excellent area of study for understanding the convergence of social and geographic space. This line of inquiry would be relevant not only to gang research but also to social science more broadly. In fact a recent issue of *Social Networks* (Adams et al. 2012) was dedicated entirely to the convergence between spatial and network thinking. Statistical models of geographic and network autocorrelation share many similarities and, in their most basic principles, have a similar functional form. The gang, both theoretically and empirically, affords an important instance in which to understand the convergence of both geographic and social space. For example, combining social networks of gang relationships with geographic models of diffusion could shed light on whether the clustering of gang violence in particular neighborhoods is related to ecological conditions, group processes (such as reciprocity), or some combination therein. Likewise, analyses such as these might advance our understanding of whether institutional memory captured in social networks might trump geographic barriers. Alternatively, spatial network research could also provide valuable insight into the ways that group processes and geography may impede or promote behavior – the relevance of such findings extends well beyond the world of the street gang.

Concluding Remarks

To review, SNA affords many unique opportunities to consider the group nature of contemporary street gangs. By considering gangs as networks of individuals or by mapping systems of gangs, researchers can more accurately measure structures and processes that make the gang a unique sociological entity. This entry has outlined some promising avenues for inquiry and, no doubt, many more exist. In advancing this network agenda for gang research, researchers should keep two things in mind. First, at the core of gang research is the inherent "groupness" of the gang. The application of SNA to gang research should proceed in a way that exploits new methodologies and data to understand this aspect of the gang. Second, the potential influence between SNA and gang research is bidirectional. This means that the application of SNA to gangs should not simply be a "cut-and-paste" of existing methodologies and ideas. Rather, the diligent application of SNA to gang research should provide insight into a range of small group behavior relevant to social science more generally.

Though enthusiastic about the direction of SNA and gang research, this entry recognizes that the methodology is not without challenges or limitations, two of which are mentioned here. First and foremost, boundary specification issues haunt all SNA research (Laumann et al. 1989) and pose even greater limitations for theoretical and empirical considerations of gangs. Boundary specification refers to the problem in SNA of identifying where the network begins and where it ends. Boundary specification can be driven by data considerations (who is in a dataset) or by theoretical considerations (who is part of the group). As discussed above, SNA research on gangs has already demonstrated that boundaries of gangs – defined by gang membership – are rather amorphous (Fleisher 1999; McGloin 2005). This suggests that finding the "true" group may prove elusive. More precisely, the shape of the network will be contingent upon where analysts draw the boundary of the group – whether, for instance, to include only "gang members" or to also include their associates. Figure 1, for instance, stopped at 2° of separation from known gang members (i.e., those individuals who were friends of friends of known gang members). In short, SNA provides a new way to discuss group membership and structure, but ultimately boundary specification becomes immensely important for theory and empirical research.

Second, future research should explore the limits of existing data as well as consider how to improve future data collection efforts. SNA requires relational data: a data source that contains some reference to the relationship among units. In some instances, this means revisiting old data sources with a new theoretical lens. For example, the study of co-offending networks (Schaefer 2012) and the study of gang homicide networks (Papachristos 2009) relied on official police data common in criminological research. But it required these authors to understand a relational aspect of the data not previously considered: how events link individuals and groups. The network literature refers to this as two-mode data (see Wasserman and Faust 1994). Similarly, ethnographic data often provides extremely rich information on the social relationships of its subjects. Fleisher (1999), for example, coded his field notes for specific instances of social ties to help uncover the patterns of the group – a time-tested anthropological technique for understanding group membership. Papachristos (2006) has even gone back and coded published ethnographies to a similar end. Moving forward, gang researchers should include network-related questions in future survey collection efforts. Indeed, much of the growth of SNA in criminology can be attributed to the network focus of a handful of surveys of school youth. Subsequent surveys of gang members or law enforcement could advance this area of research considerably by adding even a single network question to their survey instruments.

Related Entries

- ▶ Co-Offending
- ▶ Gangs and Social Networks

- ▶ Neighborhood Effects and Social Networks
- ▶ Network Analysis in Criminology
- ▶ Race, Ethnicity, and Youth Gangs
- ▶ Risk Factors for Gang Membership
- ▶ Social Network Analysis of Organized Criminal Groups

Recommended Reading and References

Adams J, Faust K, Lovasi GS (eds) (2012) Capturing context: integrating spatial and social network analysis [Special issue]. Social Networks 34(1):1–158

Bendixen M, Endresen IM, Olweus D (2006) Joining and leaving gangs: selection and facilitation effects on self-reported antisocial behavior in early adolescence. Eur J Criminol 3(1):85–114

Bouchard M, Spindler A (2010) Groups, gangs, and delinquency: does organization matter? J Crim Justice 38(5):921–933

Braga AA, Kennedy DM, Piehl EJ, Morrison A (2001) Problem-oriented policing: an evaluation of Boston's operation ceasefire. J Res Crime Delinq 38(3):195–225

Cohen J, Tita G (1999) Diffusion in homicide: exploring a general method for detecting spatial diffusion. J Quant Criminol 15(4):451–493

Descormiers K, Morselli C (2011) Alliances, conflicts, and contradictions in Montreal's street gang landscape. Int Crim Justice Rev 21(3):297–314

Emirbayer M (1997) Manifesto for a relational sociology. Am J Sociol 103(2):281–317

Fleisher M (1999) Dead end kids: gang girls and the boys they know. University of Wisconsin Press, Madison

Garot R (2007) 'Where you from!': gang identity as performance. J Contemp Ethnogr 36(1):50–84

Grannis R (2009) From the ground up: translating geography into community through neighbor networks. Princeton University Press, Princeton

Haynie D (2001) Delinquent peers revisited: does network structure matter? Am J Sociol 106(4):1013–1057

Haynie D (2002) Friendship networks and delinquency: the relative nature of peer delinquency. J Quant Criminol 18(2):99–134

Kennedy DM, Braga A, Piehl A (1997) The (un)known universe: mapping gangs and gang violence in Boston. In: Weisburd D, McEwen T (eds) Crime mapping and crime prevention. Criminal Justice Press, New York, pp 219–262

Klein MW, Maxson CL (2006) Street gang patterns and policies. Oxford University Press, New York

Krohn M, Ward J, Thornberry TP, Lizotte A, Chu R (2011) The cascading effects of adolescent gang involvement across the life course. Criminology 49(4):991–1028

Laumann EO, Marsden PV, Prensky D (1989) The boundary specification problem. In: Freeman L, White D, Romney A (eds) Research methods in social network analysis. George Mason University Press, Fairfax, pp 61–87

McGloin JM (2005) Policy and intervention considerations of network analysis of street gangs. Criminol Pub Policy 4(3):607–635

McGloin JM (2009) Delinquency balance: revisiting peer influence. Criminology 47(2):437–477

McGloin JM, Kirk D (2010) Social network analysis. In: Piquero A, Weisburd D (eds) Handbook of quantitative criminology. Springer, New York, pp 209–224

Morenoff JD, Sampson RJ, Raudenbush SW (2001) Neighborhood inequality, collective efficacy, and the spatial dynamics of homicide. Criminology 39(3):517–560

Papachristos AV (2006) Social network analysis and gang research: theory and method. In: Short JF, Hughes LA (eds) Studying youth gangs. AltaMira Press, Lanham, pp 99–116

Papachristos AV (2009) Murder by structure: dominance relations and the social structure of gang homicide. Am J Sociol 115(1):74–128

Papachristos AV (2011) The coming of a networked criminology? Using social network analysis in the study of crime and deviance. In: MacDonald J (ed) Measuring crime and criminality: advances in criminological theory, vol 17. Transaction Publishers, New Brunswick, pp 101–140

Papachristos AV, Braga AA, Hureau DM (2012a) Social networks and the risk of gunshot injury. J Urban Health 89(6):992–1003

Papachristos AV, Meares TL, Fagan J (2012b) Why do criminals obey the law? The influence of legitimacy and social networks on active gun offenders. J Crim Law Criminol 102(2):101–144

Pyrooz DC, Fox AM, Katz CM, Decker SH (2012) Gang organization, offending, and victimization: a cross-national analysis. In: Esbensen FA, Maxson CL (eds) Youth gangs in international perspective: tales from the Eurogang program of research. Springer, New York, pp 85–105

Sanchez-Jankowski MS (1991) Islands in the street: gangs and American Urban Society. University of California Press, Berkley

Schaefer DR (2012) Youth co-offending networks: an investigation of social and spatial effects. Social Networks 34(1):141–149

Short JF, Strodtbeck FL (1963) The responses of gang leaders to status threats: an observation on group process and delinquent behavior. Am J Sociol 68(5):571–579

Skogan WG, Hartnett SM, Bump N, Dubois J, Hollon R, Morris D (2009) Evaluation of ceasefire-Chicago. http://www.ipr.northwestern.edu/publications/ceasefire_papers/mainreport.pdf. Retrieved 8 Mar 2012

Starbuck D, Howell J, Linquist D (2001) Hybrid and other modern gangs. U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention, Washington, DC

Thornberry TP, Lizotte AJ, Krohn MD, Smith CA, Porter PK (2003) Causes and consequences of delinquency: findings from the rochester youth development study. In: Thornberry TP, Krohn MD (eds) Taking stock of delinquency: an overview of findings from contemporary longitudinal studies. Kluwer/Plenum, New York, pp 11–46

Tita G, Radil S (2011) Spatializing the social networks of gangs to explore patterns of violence. J Quant Criminol 27(4):521–545

Tita G, Jack Riley K, Ridgeway G, Grammich C, Abrahamse AF, Greenwood PW (2003) Reducing gun violence: results from an intervention in East Los Angeles. RAND Corporation, Santa Monica

Venkatesh S, Levitt SD (2000) An economic analysis of a drug-selling gang's finances. Q J Econ 115(3):755–789

Wasserman S, Faust K (1994) Social network analysis: methods and applications. Cambridge University Press, New York

Social Networks

▶ Network Analysis in Criminology

▶ Social Network Analysis and the Measurement of Neighborhoods

Societies

▶ History of Police Unions

Socio-spatial Criminology

▶ Geography of Crime and Disorder

Spatial Approaches to Crime

▶ Inferential Crime Mapping

Spatial Influence

▶ Forecasting of Shootings Using Risk Terrain Modeling

Spatial Models and Network Analysis

Steven M. Radil
Department of Geography, Ball State University,
Muncie, IN, USA

Overview

Several issues in criminology have led to the widespread use of spatial econometric statistical models. This entry outlines the principles and concepts behind the use of spatial models in criminology, including a description of the most commonly used models in spatial criminology. Next is a discussion of a critical but often overlooked element of spatial model specification, the construction of the spatial weights matrix, which should be theory dependent. The entry concludes with a discussion of the limitations of distance- or proximity-based approaches to constructing the spatial weights matrix, including the use of social network analysis to overcome these limitations.

Foundations of Spatial Modeling in Criminology

The recognition of geography as a factor in the explanation of a multitude of social phenomena has been an increasingly notable component of quantitative criminology. Research concerned with basic and applied questions pertaining to crime, criminal justice, and policy now typically incorporate geographic or spatial elements into analyses that utilize quantitative methodologies. An important reason for the adoption of spatial perspectives for quantitative criminology has been a growing recognition of the importance of context to human action. Incorporating context into quantitative models of human behavior is an ongoing focus of the subfield of spatial criminology.

The analysis of spatial phenomena in criminology has been made possible in recent years by the ongoing development of statistical techniques that attempt to deal with some of the unique problems of spatial data, especially spatial dependence. Dependence, or the tendency of characteristics of a given location to correlate with those of nearby locations, is a foundational issue in quantitative geographical analysis (e.g., Anselin 1988). When present in spatially organized data, spatial dependence can result in the spatial autocorrelation of regression residuals, which simply means that the values of variables sampled at nearby locations are not independent from each other (such values may be either positively or negatively correlated). The presence of spatial autocorrelation may lead to biased and inconsistent regression model parameter estimates and increase the risk of a type I error (falsely rejecting the null hypothesis). Accordingly, a critical step in statistical model specification when using spatial data is to assess the presence of spatial autocorrelation.

The statistical models developed to deal with spatial dependence, commonly called spatial econometric models, are now increasingly of interest to researchers from diverse disciplines investigating a wide range of issues, including crime and criminal justice. Many criminologists utilizing quantitative techniques now view spatial econometric regression models as an improvement non-spatial techniques as a result of the growing recognition that dependencies in the spatial structure of research data limit the reliability of inferences from conventional regression models (Anselin 1988). While there are now a variety of methods to address spatially auto correlated data in regression models, the standard workhorse in spatial regression is what are commonly referred to as simultaneous autoregressive (SAR) models.

SAR models can take three different basic forms (see Anselin 1988, 2002). The first SAR model assumes that the autoregressive process occurs only in the dependent, or response, variable. This is called the "spatial lag" model, and it introduces an additional covariate to the standard terms for the independent or predictor variables and the errors used in an ordinary least squares (OLS) regression (the additional variable is referred to as a "spatial lag" variable which is a weighted average of values for the dependent variable in areas defined as "neighbors"). Drawing on the form of the familiar OLS regression model and following Anselin (1988), the spatial lag model may be presented as

$$Y = pWy + X\beta + \varepsilon,$$

where Y is the dependent variable of interest, ρ is the auto regression parameter, W is the spatial weights matrix, X is the independent variable, and ε is the error term.

The second SAR model assumes that the autoregressive process occurs only in the error term. In this case, the usual OLS regression model is complemented by representing the spatial structure in the spatially dependent error term. The error model may be presented as

$$Y = X\beta + \varepsilon,\ \varepsilon = \lambda W\varepsilon + \mu,$$

where λ is the auto regression parameter and ε is the error term composed of a spatially auto correlated component ($W\,\varepsilon$) and a stochastic component (μ) with the rest as in the spatial lag model.

The third SAR model can contain both a spatial lag term for the response variable and a spatial error term, but is not commonly used. Other SAR model possibilities include lagging predictor variables instead of response variables. In this case, another term must also appear in the model for the auto regression parameters (γ) of the spatially lagged predictors (WX). This model takes the form

$$Y = X\beta + WX\lambda + \varepsilon.$$

Combining the response lag and predictor lag terms in a single model is also possible (sometimes referred to as a "mixed" model).

As Anselin (1988) observes, spatial dependence has much to do with notions of relative location between units in potentially different kinds of space, such as a conceptual social space. Accordingly, SAR models share a number of common features with network autocorrelation models. Substantively, spatial and network approaches have been used to explore similar questions pertaining to influence and contagion effects, albeit among different units of observations (see Marsden and Friedkin 1993 for examples). In such cases proximity or connectedness is assumed to facilitate the direct flow of information or influence across units. Individuals or organizations are also more likely to be influenced by the actions, behaviors, or beliefs of others that are proximate on different dimensions, including geographical and social space.

The Spatial Weights Matrix (W)

Spatial econometric regression models demand a careful consideration of both the theoretical and empirical spatial structure of the data in question (e.g., Florax and Rey 1995). However, the advent of new spatial econometric software packages has lowered the traditional technical barriers to the use of these techniques while simultaneously making it easy for users to choose among a few predefined spatial structures. New users may therefore be unfamiliar with the importance of the formal spatial structure to analytic outcomes and less likely to carefully consider their choice (Leenders 2002). The small literature new users may draw upon to understand this issue remains quite technical in nature, even when proclaiming the opposite intent (e.g., Griffith 1996). Taken as a whole, this state of affairs is less than ideal and unlikely to encourage careful thinking about space. However, the need to formalize the empirical spatial structure of the data for modeling is also an opportunity to reflect on the theoretical interaction of the geography and social processes being studied. In this manner, spatial analyses of human behavior and outcomes are at the core of an emerging "spatially integrated social science" identified by Goodchild et al. (2000), and an opening for the investigation of how human behavior and space is mutually constituted.

As typically argued in many geographic literatures and increasingly in quantitative criminology, spatial perspectives are important for both theoretical and practical reasons. Theoretically, spatial perspectives are of interest because of the long-standing interest in the social production of space in geography. Following Lefebvre's (1991) understanding of space as both social process and social product, the spatial structures of a given phenomenon are commonly investigated as underlying cause, constructed outcome, or both. For example, the perception of rivalries over territorial control between a set of gangs and the actual construction of bounded turf are simultaneously a social and a geographic phenomenon. The competitions are inseparable from the geography and vice versa. Hence modeling a social process such as gang rivalry requires a consideration of the modeling the social construction of space (Radil et al. 2010).

On the other hand, consideration of geography is a practical methodological matter. Data with a geographical component has important implications for statistical analyses; if processes that are affected by the underlying spatial structure in a study area are not accounted for, inferences will be inaccurate and estimates of the effects of independent variables may be biased (Anselin 1988). Perhaps the most important reason for the interest in quantitative spatial methods is the most straightforward: nearly all social science data is spatially organized, and ignoring this structural element is increasingly seen as untenable. To accommodate these issues, statistical models have been developed that attempt to deal with issues of spatial dependence. Conventionally, this is done through either introducing an additional covariate (the "spatial lag" variable discussed in the previous section) or by specifying a spatial stochastic process for the error term.

These models, now seen as both viable and important for social science research, have been discussed and exemplified in depth (see Anselin 2002 for an overview), but an essential element of these models remains largely ignored in the literature, despite the major theoretical and methodological implications. Both the lag and error models attempt to estimate regression parameters in the presence of presumably interdependent variables (Anselin 1988; Leenders 2002). This estimation process requires the analyst to define the form and limits of the interdependence and formalize the influence one location has on another. In practice, this is accomplished by identifying the connectivity between the units of the study area through a $n \times n$ matrix. The matrix is usually described in the literature as a "spatial weight" or "spatial connectivity" matrix and referred to in the preceding SAR models as "*W*."

This *W*, or matrix of locations, formalizes a priori assumptions about potential interactions between different locations, defining some locations as influential upon a given location and ruling others out. A simpler way of describing this is that *W* identifies, in some cases, who is a neighbor and who is not or with whom an actor interacts. However, the construction of *W* is more than just an empirical choice about neighbors. It is a theoretical decision regarding the spatiality of the social processes being discussed and one that has implications for the statistical estimates generated.

As *W* is supposed to represent a formal model of connections between geographic locations, how one translates theories about influence and its mechanisms across space into a formal mathematical construct is an important step. Put another way, at its core, *W* is really a theoretical geography of interaction. However, as a practical matter, the spatial analytic geography literature focuses on modeling interaction through a distance-based logic that typically takes one of two forms: contiguity or distance (Cliff and Ord 1981; Griffith 1996). Both of these spatial themes have been mobilized for constructing *W*s for various kinds of measures of spatial autocorrelation dating back to Cliff and Ord's seminal work (1981). Contiguity, or the physical connections between locations, is emphasized in issues that focus on areal spaces, especially ecological studies that use aggregated data and attempt to evaluate the relationships between neighborhood composition and crime outcomes. Distance between locations of interest also remains an important concept for criminology, in particular for journey-to-crime research.

Theorizing *W* with Social Network Analysis

Whether contiguity- or distance-based, the geography of influence is typically imagined and implemented in SAR models as a kind of gradient that uniformly diminishes with increasing distance, also known as the concept of distance decay (see Stephenson 1980). Classic examples of distance-decay thinking in social science are Boulding's (1962) "loss of strength gradient" which argued that military power has a direct inverse relationship with distance and Tobler's (1970) so-called "first law" of geography which states that everything is related to everything else, but near things are more related than are far things.

At a basic level, distance decay is the foundational concept behind the implementation of *W* in most spatial economic models, and the development and dissemination of spatial analytic software allow users to easily create *W*s from spatially organized data using the classic spatial forms of areal contiguity or point-to-point distance. While such software often guides researchers through the practical steps needed to create a matrix of spatial interaction or influence grounded in distance decay, there is no drop-down menu to offer guidance as to how best to capture the specific geography, or spatiality, of the social processes being analyzed. For any given research topic, are immediately contiguous areal neighbors enough, or should more distant neighbors also be included? If distance matters, at what distance does influence begin to diminish? Does influence diminish over distance in the same way for all social actors or behaviors? More to the point, why and in what way does distance "matter" in the operation of the social process under investigation? Addressing these sorts of questions remains the key challenges for a theoretically informed spatial analysis in criminology.

Somewhat surprisingly, theoretical discussions about the nature of *W* and practical discussions of how different specification choices may affect regression results have also been underemphasized in most spatial analytic literature. Despite these noteworthy efforts, Leenders (2002: 44) is correct in his assessment that "the effort devoted by researchers to the appropriate choice of *W* pales in comparison to the efforts devoted to the development of statistical and mathematical procedures." The net effect of this lack of attention is that theoretical conceptions about the role space plays in producing empirical patterns in a given dataset are often afterthoughts and little effort is given to understanding how sensitive model results may be to different specifications of *W*. Hence, the vision of a "spatially integrated social science" (Goodchild et al. 2000) remains unfulfilled, because when space is included in the analysis of social processes, it is often added in a default form without consideration of the geographic expression of the processes in question.

Criminologists have begun to explore such questions, often turning explicitly to the concept of social networks and the associated field of social network analysis. For example, social networks within neighborhoods are often implicated as being important mechanisms in the production and maintenance of safe, low-crime neighborhoods (e.g., Sampson et al. 1997). Such approaches argue that individual-level social bonds among local residents facilitate the formation of informal social control and the creation of shared goals and trust that regulate and censure local activities. A general acceptance of the importance of networks of local social relationships for understanding rates and patterns of neighborhood crime has prompted researchers to look into the kinds of social networks operating in communities in an effort to gather information regarding social ties among local residents as well as peer associations with delinquent others.

Beyond examining the degree to which individual-level ties influence crime patterns within a community, researchers are also beginning to explore the importance of institutional or organizational ties that can bridge communities. Sampson (2004: 158) argues for reconceptualizing neighborhoods "as nodes in a larger network of spatial relations" in order to account for the various ties that can link residents across space. Although Sampson (2004) does not specifically suggest what kinds of local institutions or organizations one should consider as important to explaining crime, in general terms he refutes the notion that neighborhoods are analytically independent and argues that ecological models of crime need to consider the different ways in which the observable outcomes in one neighborhood are partly the product of social actions and activities that can stretch beyond local communities (Morenoff et al. 2001; Sampson 2004).

Drawing on such arguments about the importance of both localized and no localized social networks, an emerging framework in quantitative criminology argues for a "deductive approach" to incorporating space built on the understanding that social networks are inherently geographic, existing both within and across localities and connecting communities separated by distance (e.g., Radil et al. 2010; Tita and Radil 2010). This analytic framework allows influence to take place not just between geographically proximate neighbors (as with conventional distance-decay-based spatial regression models) but also between locations that are connected directly by social networks. By carefully considering and allowing for processes that extend beyond (or perhaps preclude) spatially adjacent areas, one can ensure that the spatial weights matrix adequately captures the realities of the mechanisms of influence. Just as Morenoff (2003: 997) argues that spatial analysis "expands the neighborhood-effects paradigm by considering not only the local neighborhood but also the wider spatial context within which that neighborhood is embedded," a careful consideration of the spatial dimensions of social influence is under way as part of an attempt to facilitate the inclusion of the "wider *social* context of neighborhoods" into quantitative models of crime.

An early example of this emerging framework was found in the work of Mears and Bhati (2006), which links adjacent as well as nonadjacent areas to one another if the residents are economically and demographically similar by constructing *W* based on "similarity" that links socially similar areas together regardless of spatial proximity. Most recently, this framework has been advanced by research into the importance of both social and spatial linkages among gangs embedded within various types of local networks (Tita and Greenbaum 2009; Radil et al. 2010; Tita and Radil 2010). Rather than inferring connections through social similarity, the framework first demonstrated by Tita and Greenbaum (2009) accounts for influence by simultaneously considering linkages through geographic proximity between neighborhoods as well as through specific social ties that connect places. This approach carefully identifies direct social connections between neighborhoods based on rivalries between urban street gangs that are not geographically proximate, while also preserving the underlying spatial structure of the entire study area. Tita and Radil (2010) examined the utility of such an approach by comparing spatial regression model outcomes of gang violence using three differently constructed *W*s based on geographic adjacency, social connectivity, and a blend of both. Tita and Radil (2010) argued that violence committed by, and against, gang members in a socially and geographically distinct area of Los Angeles is largely a function of a *social* process that spans the local geography in such a way that violence in noncontiguous areas impacts levels of violence in a focal neighborhood and that blended *W* that incorporated both distance-based spatial effects and connections formed by specific social relations across neighborhoods offered the most meaningful results to the explanation of gang violence, both theoretically and statistically.

These studies reflect a growing interest and recognition in quantitative criminology of the need to not simply incorporate space and spatial effects in the study of crime but to do so in a way that moves beyond the uncritical notions of uniform distance decay that underpin SAR models. Specifications of the spatial weights matrix, or *W*, that rely solely on spatial contiguity or distance to define the spatial reach of the various social processes posited to be responsible for clustering force researchers to assume that all such processes decay rapidly and uniformly over geographic distance and therefore matter only among spatially adjacent neighbors. Furthermore, even when multiple social processes are considered, the conventional modeling approach has been to specify a single *W* rather than specify different kinds of connections between places for different social processes. In addition to making it impossible to parse the impact of one process from that of another, this is an a theoretical approach to understanding why and how space matters (Leenders 2002). By carefully considering the socio-spatial dimensions of the criminal phenomena of interest and of the actors behind them, the work of Mears and Bhati (2006) and (Tita and Radil 2010) demonstrate new possibilities by creating *W*s that explicitly capture the social and geographic dimensions of spatial influence through incorporating social networks. Further, such approaches allow theories of spatial influence to guide the construction of *W*, a meaningful advance given Leenders' (2002) concern with the lack of careful consideration of the underlying social processes of influence exhibited by researchers in their construction of weights matrices.

This is not to say that all the challenges of incorporating space into studies of crime have been met. For example, the developing arguments about the importance of microscale units of analysis in criminology (Hipp 2007) complicate the need to model interaction as the likelihood for spillover or other unmeasurable spatial effects increases as the areal scale of the unit of analysis decreases. Additional challenges include the need to incorporate dynamism and change over time in social networks and space and the need to consider overlapping networks (Ettlinger 2003) or multiple spatialities (Leitner et al. 2008) when theorizing the interactions between people and space. Nonetheless, the value of theoretically grounded notions of the geographic complexities of the mechanisms of influence in terms of positing various theories and mechanisms responsible for observable patterns of crime cannot be overstated. The current efforts underway in quantitative criminology have demonstrated the value of employing such an approach which should encourage future research into how social networks are involved in the social production of space and how the methods of social network analysis may be further utilized to understand if and how the structure of social networks have implications for material geographies of crime.

Recommended Reading and References

Anselin L (1988) Spatial econometrics: methods and models. Kluwer, Boston

Anselin L (1998) Exploratory spatial data analysis in a geocomputational environment. In: Longley P, Brooks S, McDonnell R, Macmillan B (eds) Geocomputation, a primer. Wiley, New York

Anselin L (2002) Under the hood: issues in the specification and interpretation of spatial regression models. Agric Econ 27(3):247–267

Anselin L, Florax R, Rey S (2004) Advances in spatial econometrics: methodology, tools and applications. Springer, Berlin

Boulding K (1962) Conflict and defense. Harper, New York

Cliff A, Ord J (1981) Spatial processes, models, and applications. Pion, London

Ettlinger N (2003) Cultural economic geography and a relational and microspace approach to trusts, rationalities, networks, and change in collaborative workplaces. J Econ Geogr 3(2):145–171

Florax R, Rey S (1995) The impacts of misspecified spatial interaction in linear regression models. In: Anselin L, Florax R (eds) New directions in spatial econometrics. Springer, Berlin, pp 111–135

Goodchild M, Anselin L, Appelbaum R, Harthorn B (2000) Toward spatially integrated social science. Int Reg Sci Rev 23:139–159

Griffith D (1996) Some guidelines for specifying the geographic weights matrix contained in the spatial statistical models. In: Arlinghaus S (ed) Practical handbook of spatial statistics. CRC Press, Boca Raton, pp 65–82

Hipp J (2007) Block, tract, and levels of aggregation: neighborhood structure and crime and disorder as a case in point. Am Sociol Rev 72:659–680

Leenders R (2002) Modeling social influence through network autocorrelation: constructing the weight matrix. Soc Netw 24:21–47

Lefebvre H (1991) The production of space (trans: Nicholson-Smith D). Blackwell, Oxford/Basil

Leitner H, Sheppard E, Sziarto K (2008) The spatialities of contentious politics. Trans Inst Br Geogr 33(2):157–172

Marsden P, Friedkin N (1993) Network studies of social influence. Soc Methods Res 22:125–149

Mears D, Bhati A (2006) No community is an island: the effects of resource deprivation on urban violence in spatially and socially proximate communities. Criminology 44(3):509–548

Morenoff J (2003) Neighborhood mechanisms and the spatial dynamics of birth weight. Am J Sociol 108:976–1017

Morenoff J, Sampson R, Raudenbush S (2001) Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. Criminology 39:517–560

Radil S, Flint C, Tita G (2010) Spatializing social networks: using social network analysis to investigate geographies of gang rivalry, territoriality, and violence in Los Angeles. Ann Assoc Am Geogr 100(2):307–326

Sampson R (2004) Networks and neighborhoods: the implications of connectivity for thinking about crime in the modern city. In: McCarthy H, Miller P, Skidmore P (eds) Network logic: who governs in an interconnected world? Demos, London, pp 157–166

Sampson R, Raudenbush S, Earls F (1997) Neighborhoods and violent crime: a multilevel study of collective efficacy. Science 277:918–924

Stephenson L (1980) Central graphic analysis of crime. In: Georges-Abeyie D, Harries K (eds) Crime: a spatial perspective. Columbia University Press, New York, pp 146–155

Tita G, Greenbaum R (2009) Crime, neighborhoods and units of analysis: putting space in its place. In: Weisburd D, Bernasco W, Bruinsma G (eds) Putting crime in its place: units of analysis in spatial crime research. Springer, New York, pp 145–170

Tita G, Greenbaum R (2011) Spatializing the social networks of gangs to explore patterns of violence. J Quant Criminol 27(4):521–545

Tita G, Radil S (2010) Making space for theory: the challenges of theorizing space and place for spatial analysis in criminology. J Quant Criminol 26(4):467–479

Tobler W (1970) A computer movie simulating urban growth in the Detroit region. Econ Geogr 462:234–240

Weisburd D, Bernasco W, Bruinsma G (eds) (2009) Putting crime in its place: units of analysis in spatial crime research. Springer, New York

Spatial Perspectives on Illegal Drug Markets

Lusine Tarkhanyan
Department of Security and Crime Science,
University College London, London, UK

Synonyms

Drug corners; Drug dealing places; Illicit retail marketplaces; Open drug market; Street drug market

Overview

Illegal drug markets have been the subject of considerable research in environmental criminology. Two main types of study have been conducted: evaluations of law enforcement strategies designed to police illegal drug markets and empirical studies which focus on the social and environmental context of places where drug distribution and consumption tends to occur. By addressing the dynamic relationship between drug trading, market procedures, environment, and police interventions, both approaches share the common aim of informing more effective crime prediction and control strategies. Since the main purposes of illicit drug markets are to both provide and secure the supply chain of illicit commodities, understanding *how* and *where* buyers and dealers position themselves to engage into transactions is critical to crime prevention.

This entry addresses what is normally the last and lowest stage in the illegal drug supply chain – the drug market. Here, the term *drug market* refers to urban location(s) where illicit commodities are illegally traded. This entry considers the extent to which illicit drug markets depend on the specific geography of places and on their amenities. The discussion will highlight how the areas surrounding these places affect their suitability as drug markets and what makes them attractive from an economic perspective. One proposition that will be discussed is that since the demand for drugs displays a nonrandom spatial pattern across the urban environment, crime prevention strategies might usefully focus on targeting and altering specific locations that are conducive to drug crime, rather than or as well as, focusing on particular individuals or groups of people.

This entry is organized as follows: first, the theoretical context and a dilemma regarding the location of drug markets will be discussed. Next, a marketing perspective will be discussed that has been proposed to explain the reasoning behind locational choices. The subsequent sections will explore the relational links between drug markets, drug-related crimes, and law enforcement actions. And, finally, further questions and potential research areas will be highlighted.

Theoretical Rationale

An urban-dweller’s routine consists of a number of activities distributed across the town or city, and the practice of committing crime is no exception. Environmental criminology identifies three factors which influence the probability of a crime event occurring: the “when,” the “where,” and the “how” (Brantingham and Brantingham 1995). The “where” refers to the spatial location of a crime. Researchers suggest (Cohen and Felson 1979) that patterns of criminal activity are influenced by both the physical environment and a criminal’s perception of what constitutes a good or bad opportunity for committing crime. Furthermore, it is proposed (Brantingham and Brantingham 1984; Cohen and Felson 1979) that the way in which a criminal moves around during their daily routine provides them with the knowledge about their surrounding environment and the opportunities it offers for criminal activity. Therefore, according to this perspective, crimes should be more highly concentrated within criminals’ awareness spaces, and these are likely to be defined by the constellation of places they move around within their noncriminal daily routines (Cohen and Felson 1979; Brantingham and Brantingham 1984).

This rationale has been applied by scholars (Eck 1995; Rengert et al. 2005) to try to explain how both drug buyers and dealers identify places that are potentially suitable for transactions of the kind they involve themselves in. Moreover, it is proposed that rational decision making on the part of both participants – buyer and dealer – in the transaction influences the geographical distribution of drug dealing locations (Eck 1995). Depending on the method of transaction, drug markets can be very large and dispersed, or small and more concentrated (Eck 1995). Scholars (Rengert et al. 2005; McCord and Ratcliffe 2007) have identified that profitable drug markets are often situated in or near socially disorganized neighborhoods, where there is a lack of social resistance to the market’s existence and from which there may be a ready supply of people willing to purchase drugs. However, the rationale for choosing these neighborhoods remains disputed: it is still unknown whether the locations are chosen initially due to their accessibility to a large number of drug users and their presence undermines the neighborhoods’ social organization (Eck 1994) or whether the initial lack of social resistance provides optimal grounds for establishing drug markets (McCord and Ratcliffe 2007).

Additionally, it has been noticed that the success of trading locations is associated with types of land uses that are conducive to, or generate, other types of crime. For example, drug dealing is likely to happen close to facilities which inherently and routinely generate a large flow of people. These are mainly open public spaces, retail, entertainment facilities, and transport interchanges that are associated with low levels of adequate guardianship or place management (Eck and Wartell 1996). In their analysis, Rengert and colleagues defined two types of built environment facilities that may be associated with the locations of drug markets. First, there are those which indirectly increase the profits from drug sales, because they facilitate nonresidents' access to an area. An example of this would be transport interchanges, which provide easy access to the drug markets (Brantingham and Brantingham 1995). Second, there are those which generate opportunities for drug transactions because they are used routinely by potential drug buyers, for example, areas near to homeless shelters or pawnshops, where potential buyers can readily convert stolen goods to cash (Anderson 1999).

The Main Dilemma

In the same way that legitimate businesses do not select their location at random, drug market placement may reflect rational decision making (Eck 1995; Rengert et al. 2005): the main aim of both activities is to attract customers and supply products to them in order to make a profit. From a strategic perspective, when deciding where to locate their stores, retailers focus on finding an accessible spatial location which will attract many potential customers. However, their locational choices are also constrained by different planning regulations and environmental impact assessments, which are required by authorities to allow the placement of the shop.

Rengert (1996) suggests that drug dealers may follow a similar logic and try to identify potential profitable sites. As Rengert states, "the quality of the sale's location is directly related to the quantity of profit for the illegal drug sales" (Rengert 1996). However, in contrast to legitimate trading, in the case of illegal markets, offenders have an additional goal: staying safe and unnoticed so as to avoid arrest (Reuter and MacCoun 1993). Thus, drug dealing locational choices are also constrained by the presence of legitimate and capable guardians who discourage criminal opportunities (Eck 1994). These include security guards, home and shop owners, and generally people who manage such places. In their study of drug and disorder problems, Mazerolle et al. (1998) found that at the level of street blocks, the place managers who engaged in crime prevention activities played an important role in guarding places from drug problems. Moreover, they found that place managers who engage their neighbors from the same street block in crime prevention efforts are more effective than individual efforts.

Eck (1995) has termed this specific aspect of drug markets the accessibility vs. security dilemma: "how to exchange illicit goods or service when the exchange process is very risky" (p. 71). Illicit drug markets typically face the conflict between needing to be accessible to many customers, including complete strangers, and avoiding the vulnerabilities associated with drug sales. Due to the illicit nature of the market, both customers and dealers are usually in a vulnerable position, since they run the risk of both legal intervention and being cheated or robbed by their counterpart during the course of a transaction. Of course, there is no means of securing the transaction through law enforcement or of resolving such conflict of interests through legal channels (Reuter and MacCoun 1993). Thus, violence is a very common means of regulating and resolving disagreements (Goldstein 1985), especially in street-based sales, where exchanges take place outdoors, often between anonymous participants. Both buyers and dealers will thus be motivated to limit their accessibility so as to reduce risk, seeking locations that they personally consider to be safest. Such places may be enclosed and familiar locations. For example, indoor sales may be made from fixed locations such as drug houses or deliveries made to indoor locations specified by the customers (Curtis and Wendel; 2000). However, in both of these scenarios, one of the participants will always be at a greater risk than the other, since they will not be familiar with the location. In some cases, unfamiliarity may lead to them avoiding a location altogether.

Eck proposes two models of drug markets which can overcome the access-security dilemma, given the constant risk of police presence. The first is the *social network-based transactions model*, in which security is provided through a network of trusted people. The second is the *routine activity model*, in which both participants use their legitimate daily activities to search for places which are potentially appropriate for engaging in drug deals. In the first model, there is no attachment to a specific location: through a social network both parties can arrange a mutually accessible location, potentially based on their routine. This type of market tends to be closed in nature, and the market itself may be dispersed over a large geographical area, leading to a low spatial concentration of drug dealing incidents (Eck 1995). Although a network-based exchange offers security, it also limits the number of participants in the transaction, which in turn reduces the potential for profit.

The second model proposed assumes that the spatial-temporal patterns of both participants' legitimate daily activities determine where their drug transactions occur. These markets tend to be spatially clustered and to focus on locations which are familiar to both participants, thereby reducing their perception of potential risks. These markets also need specific operating conditions: when surrounded, for instance, by a large number of legitimate activities and a constant flow of foot traffic, it is easier to blend into the crowd and search for potential customers. This type of market is often an open street market with a high frequency of transactions between anonymous participants. The market permits equal access to all participants and is located near places with mixed land uses and a high concentration of activities, such as shopping centers, high streets, and transport interchanges. Eck proposes that place managers from the legitimate sphere who control and manage these locations have a particularly important role in impeding this type of market.

It should be noted that not all locations which are next to shopping malls or transport facilities, for example, will form drug markets. According to this model, places which are more likely to have drug markets will be attractive from a retail perspective, offering a balance between the spatial distribution of those who demand the product and the distance they would be required to travel to the market. This concept will be discussed further in the following section.

State of the Art: Locational Choices and Spatial Economics

As with many crime types, open drug markets tend to concentrate geospatially: from all the available urban locations, there are very few which are well suited to illegal drug trading. Weisburd et al. (2004) found that a small number of street junctions in Jersey City (4.4 %) accounted for almost half of the drug sales arrests in the city. This demonstrates considerable spatial concentration of markets. According to Kleiman (1991 cited in Taniguchi et al. 2009) the reason for spatial clustering is that a large number of dealers operate in a single location, gaining security from arrest by spreading the risk of apprehension across all dealers. The same logic applies to customers who would prefer to be in a crowd than alone.

Rengert et al. (2005) propose that the exact location of open drug markets is likely to depend on the convergence of conditions that are most suitable for both participants involved in transactions and spatial patterns in demand. They suggest the model of "agglomeration economics" operates, as is the case with legal trade. That is, after a location becomes known as a site for specific goods, more customers will visit the area in the search of that good. At a certain point, the number of buyers that visit the area will be sufficient to support further suppliers of the product(s) and so more retailers locate there. In Wilmington, USA, researchers discovered that the likelihood of large drug markets being located near to each other is quite high. They say "there are best places to sell illegal drugs because these are places where demand is focused spatially" (Rengert 1996). That is, in order to stay profitable, a drug market's location should be attractive enough to a sufficient number of drug users. The attractiveness of the place is partially determined by the surrounding facilities, but above all by how far a buyer is prepared to travel to make a purchase (Pettiway 1995).

In legal trading, markets for goods which are highly valuable but purchased infrequently are usually found in very accessible urban locations, which may attract many potential customers from remote locations. In contrast, local markets tend to supply items, which consumers will wish to purchase frequently but will be prepared to travel only short distances to get them. In their study, Rengert and colleagues (et al. 2005) discuss concepts from the economics literature – *threshold population* and *range* – to frame a discussion on how the demand for drugs and the distance required to reach a market likely affect the location and stability of drug markets. Threshold is defined as a minimum number of customers required for a market to stay profitable. The concept of range concerns the distance that a buyer is prepared to travel to purchase a good. Pettiway (1995) suggests that the spatial range of a market that caters for vehicular movement will be larger than one that caters for pedestrian movement. Simply put, if the demand for a drug market is situated within the physical catchment area (an area to which they do or will travel) of many potential drug buyers, that market will remain stable or may even grow.

According to this rationale, dealers will try to establish markets at locations that reduce the total distance that customers will have to travel to reach them. In order to determine the demand for drug markets, US researchers have used methods of sociodemographic profiling to identify those types of neighborhoods with the characteristics that are associated with increased risk of drug use. They propose that if a drug dealer wants to sell drugs to a local community, he must first identify possible users. If the local demand is not enough to sustain a drug market, the dealer must consider factors that would attract potential customers into the neighborhood. For example, the market should be situated in close proximity to transport facilities, which are used routinely by many potential drug addicts, or it should be accessible to modes of private transport.

Furthermore, since drug markets are established along routes which are used on a daily basis by many potential drug buyers, the location and retail characteristics of these markets can vary considerably. Depending on where users and dealers live relative to the market, drug markets are classified as *neighborhood, open regional, semi-open regional, and closed regional markets* (Reuter and MacCoun 1993; Eck 1994). Neighborhood markets are described as places in which both customers and dealers are from the vicinity: they might be neighbors or know each other. Open regional markets are usually established close to places which are routinely used by a large number of nonresidents, such as shopping malls. These markets are large enough to support several competing drug dealers and can secure a high frequency of transactions. Semi-open regional markets are typically established when both participants aim to lessen the risks involved in open street markets. They do not live in the same area and they only interact if they know each other or have been referred by a third person. Closed regional markets' locations are determined by both participants of the transaction and are distributed over a wide area. The clientele of such markets are established through a network of friends and other trusted people. This type of market is the norm in wholesale drug dealing.

Alongside spatial clustering, drug markets are typically located in close vicinity to certain facilities: shopping centers (Eck 1995), high schools (Roncek and LoBosco 1983), bars (Roncek and Maier 1991), cash stores and pawnshops (Anderson 1999), transport links, train stations and highways (Eck 1994), and vacant homes (Rengert et al. 2005). Rengert and colleagues (2005) found that there are also facilities that discourage the establishment of drug markets in an area, such as police and fire stations or courts and federal buildings. Given that accessible drug dealing locations should offer good retail potential, drug marketplaces can be further classified according to the level of pedestrian and traffic accessibility, which may bring potential customers to the area. It can be suggested that, depending on a market's geographical positioning in the city, the level of accessibility will vary – from locally to regionally accessible markets. For example, Eck (1994) found that in San Diego, outdoor drug markets formed at locations about two blocks away from major transportation arteries, suggesting that they were regionally accessible markets. Importantly, this suggests that although offenders aim to sell drugs from accessible locations, they do not tend to do so on the major roads (presumably as a way of reducing risk). That is, for regionally accessible markets, operating in close proximity to major roads may offer an acceptable balance of custom and safety. In comparison, in Philadelphia, Rengert and colleagues (2005) found a high concentration of drug markets located in the suburbs, located away from major roads, suggesting that these marketplaces are oriented to local rather than regional demand.

State of the Art: Drug-Related Crime

The drug-crime relationship is usually examined through the theoretical perspective of a "tripartite framework" advanced by Goldstein (1985), which categorizes drug-related crime into three groups. First, there is "psychopharmacological" crime in which drug use affects aggression, thereby increasing the chance of criminal behavior. Second, there is "economic-compulsive" crime, in which crime is committed to finance drug use. Third, there is "systemic" crime, in which criminal behavior occurs in order to secure the drug supply chain, protect profits, control locations, compete for customers, or resolve conflicts of interests. A fourth category has also been proposed (Blumstein 1995 – cited in MacCoun et al. 2003; Curtis and Wendel 2000; Rengert 1996), which focuses on the negative impact of drug markets on local communities, mainly in terms of how drug dealers' behavioral codes and manners impact on community members who are not directly related to the market, for instance, by increasing gun carrying for the purpose of self-defense or conflict resolution.

Despite extensive research on the drug-crime nexus, little is known about how drug-related crime and drug markets relate to each other geographically. Based on crime pattern theory, it is suggested that drug markets can be both crime generators and crime attractors. The former may occur due to the inherent criminality of drug markets – with those who buy or sell drugs engaging in other types of crime while they are in an area, even though they may not have travel to such locations with the intention of so doing. Even though they do not travel to such locations to commit other forms of crime, this will nevertheless affect such crime rates in the area. Apart from potential drug buyers, drug markets can *attract* other offenders who purposely seek out criminal opportunities. In their research, Weisburd et al. (2004) found that drug-related crime clusters spatially around drug markets. Rengert and Wasilchick (1989) state that property offenses committed by drug users are spatially concentrated in close proximity to drug markets. Moreover, ethnographic research suggests that violence against dealers is not uncommon since they possess drugs and money.

Others (Eck 1994; Reuter and MacCoun 1993) suggest that depending on the retail nature of the market (street based, indoor from fixed locations, or delivery based), levels of violence and criminal behavior differ considerably. The indoor and delivery-based methods of transaction are safer than outdoor sales, since they enjoy an established and protected territorial boundary and/or security is provided by limiting sales to a network of trusted people.

The nature of a market – *local, export, public, and import* – is thought to affect levels of violence. According to Reuter and MacCoun (1993), local markets have the least violence, given that there is an established network of local social ties and all market participants interact on an ongoing basis within recognized territorial boundaries. Export markets also enjoy low levels of violence, since local dealers are interested in discouraging violence in order to attract more customers from outside the local neighborhood. Public markets feature a high level of anonymous participants operating in public places and hence lack clear territorial boundaries. Due to the resultant competition and weak social ties between participants, which in turn promote potential mistrust and disagreement both between dealers and buyers and among dealers, these markets have considerable potential for violence. Import markets may feature a similar risk of violence, since neither dealers nor customers belong to the given neighborhood and the process of establishing territorial boundaries may cause dissatisfaction among local residents and tensions between dealers. As a result, conflict may arise between dealers and locals as well as between competing dealers.

"Systemic" violence, which occurs in drug markets as a means of controlling a location and competing for customers, has been investigated in relation to drug dealing gangs, who usually operate from the so-called street corners (Taniguchi et al. 2009). These street corners are financially more beneficial than other street segments due to the possibility of targeting a greater volume of passersby (Eck 1994). Consequently, gangs compete to defend and control the corners as a matter of prestige and economic gain. Scholars (Taniguchi et al. 2010) have found that there is more violence when multiple gangs operate from a single corner than when a single gang controls it. In the former scenario, violence between gangs arises from their competing interests in finding clients and establishing territorial boundaries. In comparison, when a single gang occupies a strategically positioned corner, it has an economic interest in keeping the area safe from violence in order to attract more customers (Cohen and Tita 1999; Goldstein 1985; Levitt and Venkatesh 2000 – quoted in Taniguchi et al. 2010).

Several authors (Eck and Maguire 2000; MacCoun et al. 2003) have noticed that drug markets are highly responsive and adaptive to law enforcement. They have suggested that violent behavior linked to competition may be caused by intensive law enforcement actions. The link between police activities and drug markets will be discussed in detail in the following section.

State of the Art: Drug Markets and Law Enforcement

Apart from arresting drug dealers, one of the police's main goals in targeting drug markets is to understand how the operational structure of these markets can be disrupted. Since drug markets cluster in relatively few locations, which also feature specific land uses which support them, crime prevention researchers (Taniguchi et al. 2010) suggest targeting specific locations instead of drug dealers. By applying the categorization of drug markets according to where dealers and buyers live, as discussed above, scholars (Reuter and MacCoun 1993) have identified the markets which are expected to be the most and least vulnerable to law enforcement activity. Local markets are, for example, the most resilient to police activities, because both participants in the transaction are from the same neighborhood with strong social ties. Consequently, if the market's functioning is disrupted, the participants' shared knowledge of the neighborhood allows the easy reestablishment of the market elsewhere. In contrast, public drug markets can be easily disrupted, because there will be relatively fewer locations which are well known to both dealers and buyers and which are suitable for drug transactions. Targeting these locations may be sufficient to keep drug dealers and buyers apart.

However, it has been claimed that disrupting the operations of drug markets will not solve the problem, since they are highly adaptive to law enforcement and might reopen nearby or simply start trading at different times. This amounts to claims of spatial or temporal displacement (Repetto 1974). Although displacement is a common criticism of geographically focused crime prevention efforts, this concern has not been substantiated by empirical research. Furthermore, a number of studies (Clarke and Weisburd 1994; Weisburd et al. 2004) have found the reverse effect, namely, a diffusion of crime control benefits, whereby there is a reduction in the crime level in locations surrounding the area where police implemented crime prevention strategies. There are several explanations for this effect. From the perspective of agglomeration economics, "the immediate spatial diffusion is more likely than immediate spatial displacement" (Taniguchi et al. 2010). For example, it has been suggested that closing down the most profitable drug dealing location will make dealing in the neighboring locations less rather than more profitable. Dispersing the cluster of competing drug dealers will lead to a series of smaller marketplaces, each with fewer participants and drug transactions. According to this perspective, a forced change in spatial positioning can be sufficient in disrupting a drug market due to its detrimental effect on economic conditions.

Ethnographic research suggests (Weisburd et al. 2004) that after preventative interventions, drug dealers prefer to adapt their behavior and stay in the targeted area rather than move location. Both participants may implement less open modes of transaction, for instance, arranging the drug transaction time and place through mobile devices.

Weisburd and colleagues (2004) have studied the effects of geographically focused or hot-spot crime prevention strategies. Despite the widely held belief discussed above that hot-spot policing causes displacement, the researchers found no evidence on one of these. They strongly support the use of focused market targeting in police interventions since dealers resist moving away from a chosen area, as discussed above. They justify this support by emphasizing that dealers are part of the social and business community and when they relocate they may encounter violence caused by competition with established dealers in the new area. Moreover, existing networks of customers are not easily displaced to another location, since their limited geographical familiarity makes it harder to locate their dealer there. This can interrupt the supply-demand chain and hence affect the profitability of drug dealing at a particular place. This is one reason why focused hot-spot prevention may be beneficial to an area since one consequence is that reestablishing the market is likely to require effort on the part of the dealer.

Overall, researchers (Rengert 1996; Weisburd et al. 2004) are in favor of hot-spot policing, believing it to produce strong crime prevention results. Since the change in economic and physical conditions makes the location less attractive for a dealer, they have strongly suggested paying more attention to the surrounding conditions which affect the marketplace's economic attractiveness. However, caution should be taken in generalizing these results as several crackdown studies have found evidence that a diffusion of benefits to adjacent areas may be accompanied by partial displacement of the drug scene to indoor locations, other neighborhoods, or nearby metropolitan areas.

Future Research Directions

Overall, it can be noted that there is a particular environmental balance between physical and social characteristics which facilitate the optimal grounds for establishing a drug marketplace. As has been described above, the retail chain of drug supply operates differently depending on both the spatial structure of the area and the participants' profile. Moreover, drug markets are more likely to be located in areas with low levels of social organization. It is hard, however, to establish whether the choice of location is determined by the level of social organization or whether the social organization is undermined by the presence of a drug market. Whatever the case, the ensuing lack of legal guardianship provides, from the dealers' perspective, valuable protection from police enforcement.

For the future research directions, it may be worth looking at geography of separate crack or heroin markets alone and on crime concentration nearby. There is an opinion that crack markets tend to be more violent than markets for other drugs, but it is not supported with any empirical evidences. Additionally, it is useful to know whether markets that offer particular types of drug that is more expensive and difficult to source are harder to stamp out for good and are more liable to displacement.

Although many researchers recognize the relationship between legitimate movement and drug markets, it is still unknown how much and what type of movement is required for the establishment of drug markets. Locational choices may, for example, be influenced by whether the market is frequented by pedestrian or vehicular traffic. A drug market which, for example, relies on vehicular traffic may need to be located in close vicinity to main roads with easy and quick access *to* and *from* the marketplace. In the case of pedestrian traffic, the market will probably be located near places where there are a sufficient number of legitimate activities which can conceal the drug transaction process. Presumably, there should be multiple escape routes, lookout points, and workarounds (Eck 1995). Since street networks influence the way people navigate towns and cities (Brantingham and Brantingham 1995; Hillier and Hanson 1984), and the street networks vary from place to another, caution should be taken when generalizing findings from one location to another. This is particularly true when comparing findings from studies conducted in automobile-based societies with regular grid-like street networks, such as America, with those from pedestrian-based societies with street networks which have grown organically, such as the UK. To be more explicit, all drug markets are not equal, and those factors that influence one type of market may be different to those that influence others. Understanding the factors that affect different types of markets would be a useful next step, and consideration of how the morphology of the street network might shape or support opportunities for drug crime would seem to be a logical starting point.

Another priority research area is the drug-crime connection. Rengert and Wasilchick (1989) discovered that offender's search behavior is oriented towards the location of drug marketplace. They showed that drug-dependent property offenders search for potential houses to burgle on their way to purchase a drug. Despite the fact that many scholars have pointed to a significant relationship between drug use and crime, little is known about how drug-related crime patterns are associated geographically with drug markets: while those who use drugs may also commit crime, how are their spatial targeting decisions influenced by the location and type of drug markets? From a methodological perspective, it is necessary to apply appropriate crime data modeling techniques to examine the geography of criminal activities taking place in different locations and explore how potential drug users are positioned and move about within the city. Recent developments in the use of geographical information systems (GIS) allow the input, storage, merging, and analysis of geographical, statistical, and other types of datasets within a single file format. The high operational quality of these platforms enables the integration of many data formats and the construction of very large and detailed models. This means that it is possible, for example, to map drug dealing events in conjunction with sociodemographic statistics, data on land use, or crime statistics. This information can be aggregated according to different spatial units, such as the administrative boundary, ward unit, and building block. Remarkably, most of the research has focused on the patterns of drug dealing only at the area or grid-based unit of analysis. With few exceptions (e.g., Friedrich et al. 2009), drug markets have not been examined at the street segment level. Thus, in future work, it will be useful to conduct analyses using street segments as the unit of analysis. The advantage of detailing at such a fine scale of analysis allows more accurate data modeling – after all criminals do not navigate areas, they move along the streets. Moreover, from the detection and prevention perspective, it is quite useful to know the type of street segments that have a high probability of being or becoming drug dealing places and require more police attention.

Recent advances in computational methods allow very detailed modeling and analysis of data which should facilitate this sort of approach. For example, a group of researchers from Japan used a GIS platform to develop the Spatial Analysis along Networks (SANET) software. This provides a set of statistical tools that facilitate the analysis of point data with reference to the street network. The network statistics used represent a modification of numerous traditional spatial but also graph theoretical techniques. For instance, the software enables the analysis of the density of crime incidents across street segments, the shortest path between points, and the enumeration of the fraction of clustered crime incidents and so on.

The morphology of street network per se is another factor that can be informative in understanding and analyzing crime incidents within a city. Space Syntax theory (Hillier and Hanson 1984) and research discuss how the spatial properties of the street network influence the distribution of pedestrian movement patterns. Space Syntax has developed analytical tools, which using computer modeling (Depthmap software; Turner 2001) provide graphic (and quantitative) representations of the probabilities of pedestrian and traffic movement along the street segments. Namely, it estimates the accessibility levels for every street segment in relation both to its immediate surroundings (local scale) and to the whole city network (global scale). The quantifiable differences in the level of movement accessibility between different locations enable testing statistically why particular locations are prone to crime.

In conclusion, it should be emphasized that illegal drug markets are difficult and multidimensional problems involving many social and spatial factors. Mechanisms of illicit retail may vary from country to country, and so the theoretical frameworks described in this entry require further testing to determine the extent to which they do and do not apply in other cultural contexts. Methodological and computational advances offer the opportunity to do so and will likely provide new insight into patterns of drug crime.

Related Entries

- ▶ Crime Location Choice
- ▶ History of Geographic Criminology Part I: Nineteenth Century
- ▶ Rational Choice Theory
- ▶ Routine Activities Approach
- ▶ Spatial Models and Network Analysis

Recommended Reading and References

Anderson E (1999) Code of the street: decency, violence, and the moral life of the inner city. W.W. Norton, New York

Brantingham PL, Brantingham PJ (1984) Burglar mobility and crime prevention planning. In: Clarke R, Hope T (eds) Coping with burglary. Kluwer-Nijhoff, Boston

Brantingham PL, Brantingham PJ (1993) Environment, routine, and situation: toward a pattern theory of crime. In: Clarke RV, Felson M (eds) Routine activity and rational choice. Transaction Publishers, New Brunswick, pp 259–294

Brantingham PL, Brantingham PJ (1995) Criminality of place: crime generators and crime attractors. Eur J Crim Policy Res 3:5–26

Clarke RV, Weisburd D (1994) Diffusion of crime control benefits: observations on the reverse of displacement. Crime Prev Stud 2:165–184

Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44:588–608

Cohen J, Tita GE (1999) Diffusion in homicide: exploring a general method for detecting spatial diffusion processes. J Quant Criminol 15:451–493

Curtis R, Wendel T (2000) Toward the development of a typology of illegal drug markets. In: Hough M, Natarajan M (eds) Illegal drug markets: from research to policy. Criminal Justice Press, Monsey, pp 121–152

Eck J (1994) Drug markets and drug places: a Case–control study of the spatial structure of illicit drug dealing. Dissertation, University of Maryland, Ann Arbor, University Microfilms International

Eck J (1995) A general model of the geography of illicit retail market places. In: Eck J, Weisburd D (eds) Crime and place. Criminal Justice Press, Monsey

Eck JE, Maguire ER (2000) Have changes in policing reduced violent crime? An assessment of the evidence. In: Blumstein A, Wallman J (eds) The crime drop in America. Cambridge University Press, Cambridge, pp 207–265

Eck J, Wartell J (1996) Reducing crime and drug dealing by improving place management: a randomized experiment: report to the San Diego Police Department. Crime Control Institute, Washington, DC

Friedrich E, Hillier B, Chiaradia A (2009) Anti-social behaviour and urban configuration using space syntax to understand spatial patterns of socio-environmental disorder. In: Koch D, Marcus L, Steen J (eds) (Proceedings) 7th international space syntax symposium, Royal Institute of Technology (KTH), Stockholm

Goldstein PJ (1985) The drugs/violence nexus: a tripartite conceptual framework. J Drug Issues 15:493–506

Hillier B, Hanson J (1984) The social logic of space. Cambridge University Press, Cambridge
Hillier B, Penn A, Hanson J, Grajewski T, Xu J (1993) Natural movement or, configuration and attraction in urban pedestrian movement. Environ Planning (B): Planning Design 20(1):29–66
Levitt SD, Venkatesh SA (2000) An economic analysis of a drug-selling gang's finances. The Quarterly Journal of Economics 115:755–89
MacCoun E, Kilmer B, Reuter P (2003) Research on drugs-crime linkages: the next generation. Toward a drugs and crime research agenda for the 21st century. Retrieved September 2003 from: http://www.ojp.usdoj.gov/nij/drugscrime/194616.htm
Mazerolle L, Kadleck C, Roehl J (1998) Controlling drug and disorder problems: the role of place managers. Criminology 36(2):371–402
McCord E, Ratcliffe J (2007) A micro-spatial analysis of the demographic and criminogenic environment of drug markets in Philadelphia. Aust N Z J Criminol 40:43–63
Okabe A and Okunuki K and SANET Team SANET a spatial analysis along networks (Ver.4.0). Tokyo, Japan
Pettiway LE (1995) Copping crack: the travel behaviour of crack users. Justice Q 12(3):499–524
Rengert G (1996) The geography of illegal drugs. West View Press, Boulder, Colorado
Rengert G, Wasilchik (1989) Space, time and crime: ethnographic 87 drug sales, burglary, and neighborhoods insights into residential burglary. Final Report, National Institute of Justice, Grant # 88-IJ-CX-0013.
Rengert G, Ratcliffe J, Chakravorty S (2005) Policing illegal drug markets: geographic approaches to crime reduction. Criminal Justice Press, Monsey
Repetto TA (1974) Residential crime. Ballinger, Cambridge
Reuter P, MacCoun R (1993) Street drug markets in inner-city neighbourhoods. Rand Corporation, Santa Minica
Roncek D, LoBosco A (1983) The effect of high schools on crime in their neighborhoods. Soc Sci Quart 64:598–613
Roncek D, Maier P (1991) Bars, blocks, and crime revisited: linking the theory of routine activities to the empiricism of 'hot spots'. Criminology 29:725–754
Taniguchi TA, Rengert GF, McCord ES (2009) Where size matters: agglomeration economies of illegal drug markets in Philadelphia. Justice Q 26:670–694
Taniguchi TA, Ratcliffe J, Taylor R (2010) Gang set space, drug markets, and crime around drug corners in Camden. J Res Crime Delinq 48(2011):327–363
Turner A (2001) Depthmap: a program to perform visibility graph analysis. In: Proceedings 3rd international symposium on space syntax, Georgia Institute of Technology, Atlanta, Georgia, pp 31.1–31.9
Weisburd D, Wyckoff J, Ready J, Eck J, Hinkle C, Gajewski J (2004) Does crime just move around the corner? A controlled study of spatial displacement and diffusion of crime control benefits. Criminology 44:549–592

Specialization and Sexual Offending

Danielle Arlanda Harris
Justice Studies, San Jose State University,
San Jose, CA, USA

Overview

The field of research into sexual aggression and offending is largely based upon the assumption that sexual offenders are inherently different from generic or nonsexual criminals. This position has permeated through more than just research and is now evident in the policies and practices that treat and manage sexual offenders and seek to prevent sexual abuse. This assumption of offense specialization proposes that sexual offenders commit sexual offenses persistently and exclusively (or at least predominantly) throughout their lives (Simon 1997; Zimring et al. 2007). It is often further implied that these offenders have a detectable degree of sexually deviant arousal and commit sexual offenses in a planned and repetitive way (Simon 2000; Zimring et al. 2007). Meanwhile, almost no other type of offender is viewed like this. Rather, they are considered to be criminally versatile, showing no apparent predilection for one crime over another.

This entry provides an overview of recent international literature regarding the nature and extent of offense specialization and versatility in the criminal careers of sexual offenders. A discussion of the many methodological considerations that arise from this research will be presented, followed by a brief overview of relevant theoretical perspectives. Finally, comments are provided regarding the implications for further research in criminology and criminal justice as well as prevention efforts aimed toward addressing sexual offenders.

Introduction

There are wide-reaching practical implications that stem from the view that sexual offenders are either specialist or versatile in their offending. The sexual offender "profile" has become increasingly visible in recent years. This has likely been encouraged by a small number of horrific incidents that received extraordinary attention by the international media. As a consequence, individuals who are identified as sexual offenders are now subjected to more discretionary decision making in the criminal justice system than almost any other type of offender (Simon 1997). This has led to more frequent and aggressive prosecution of sexual offenders as well as the passage of various special legislative initiatives aimed exclusively at sexual offenders (Simon 1997). Judgments about an individual's dangerousness and their perceived risk of reoffense are predicated upon the view that sexual offenders are "special cases" requiring or necessitating specific assessment, treatment, therapist certification, polygraphy protocols, probation models, protective custody, and so on.

If offenders do specialize in the types of crime they commit, then it makes sense for policy to be tailored toward specific individuals who have been determined likely to participate in a particular crime (Mazerolle et al. 2000). In this way, crime-specific incarceration policies would be influential in reducing the incidence of specific crime types. If knowledge of prior offenses could actually predict subsequent criminal events, then the identification of specialized offenders would certainly be useful in criminal justice decision making (Farrington et al. 1988).

If, on the other hand, offenders are predominantly versatile, crime-specific policies would have less impact. If versatility is the standard offending tendency, a reevaluation of many legislative initiatives and policies already in place might be necessary. One interpretation of this state of affairs would suggest that such initiatives, "fuelled by the assumption of specialization" (Simon 1997, p. 2), have evolved to cater to a sexual offender population that does not actually exist.

Competing Perspectives

The Specialist Perspective: Offense specialization is a central component of the criminal career paradigm and refers to an individual's tendency to engage in the same crime or type of crime on successive arrests or offending occasions (Blumstein et al. 1986). Specialist offenders are thought to become proficient at a certain crime, which they subsequently commit almost exclusively (Harris 2008). The consequence of this definition is that it is thought to be possible to predict any particular arrest in someone's criminal career based on their most recent prior arrest (Britt 1994).

Compelling evidence of versatility first emerged in the 1920s (Harris 2008) which left the notion of specialization largely marginalized for decades. Despite this, books continue to be written, research projects funded, and theories constructed, all with an assumption that offenders specialize in their offending (Gottfredson and Hirschi 1990). The strong belief in specialist offenders – shoplifters, drug offenders, and white-collar criminals – conveniently allows "us" to separate ourselves from "them." Nowhere is this trend more obvious than in cases of rapists or pedophiles.

The Versatile Perspective: The opposing position of offense versatility proposes a common construct of deviance that views all crimes as "acts of force or fraud undertaken in pursuit of self-interest" (Gottfredson and Hirschi 1990, p. 15). Gottfredson and Hirschi (1990) claim that all crimes are alike and argue that any distinction between them (such as trivial or serious, or victim or victimless) is irrelevant and misleading. It would follow then that distinguishing between sexual and nonsexual offending (or, e.g., between sexual offenses against adults and sexual offenses against children) is an equally futile endeavor.

Specialization and Versatility in Sexual Offending: Theoretical Context

Although offense specialization is probably best understood empirically, it is necessary to situate offending patterns within a meaningful theoretical context. Theory building in sexual offending research has largely occurred beyond the boundaries of traditional criminology (Simon 2000). Meanwhile, criminological theories tend not to consider sexual offenders on their own. Some have observed that these developments have created an almost philosophical apartheid between conventional criminology and sexual offending theory (Soothill et al. 2000). There are, however, appropriate explanations that attend to both of these observations. The following discussion will focus explicitly on those theories of sexual offending that speak to specialization or versatility.

Sexual offenses are usually understood within the context of deviant sexual arousal and paraphilia. But a more generic model would consider a sexual offense within the context of general criminality. Here, a sexual crime would be seen as an expression of general antisocial behavior that likely occurs rarely and in a haphazard fashion within an offender's criminal career. Some researchers have concluded that generic models provide sufficient explanation of sexual offending, while others hold that sexual offenders warrant more specific theoretical consideration.

The propensity to offend can certainly be understood by either general or specialist theories. In particular, sexual offending can be understood as either an example of general crime or as a special case requiring the presence of sexual deviance. The Integrated Theory of Sexual Offending (Marshall and Barbaree 1990) and the Conditioning Theory of Sexual Offending (Laws and Marshall 1990), respectively, are proposed as two such examples.

The Integrated Theory of Sexual Offending. Marshall and Barbaree's (1990) Integrated Theory of Sexual Offending focuses on opportunity and self-control and can be seen as a sexual offending-specific version of Gottfredson and Hirschi's (1990) General Theory published in the same year. Marshall and Barbaree (1990) combined developmental elements with situational and environmental factors to create an integrated explanation of sexual offending (Ward 2002). The authors proposed that men are biologically equipped to learn to express their sexuality aggressively. They contend that "the acquisition of attitudes and behaviors during childhood sets the stage for the developing male to respond to the sudden onset of strong desires characteristic of pubescence with a prosocial or an antisocial mental set" (Marshall and Barbaree 1990, p. 260). Similar to Gottfredson and Hirschi (1990), they illustrate the importance of parental attachment and argue that sexual offending is more likely in individuals whose childhoods are characterized by poor parenting or poor socialization (Ward 2002). Provided these preexisting variables are present, situational elements have the ultimate impact on sexual offending. Stress, intoxication, anger, presence of a potential victim, and the belief that detection can be avoided are the most important situational variables (Marshall and Barbaree 1990; Ward 2002).

Marshall and Barbaree's (1990) contribution does not address the issue of specialization versus versatility directly. For them, sexual offending remains independent of other behaviors and therefore warrants a unique explanation. They recognize that sexual offenders have the capacity to be criminally versatile and do engage in an array of antisocial behaviors. This represents a significant variation from other theories constructed to explain sexual offending which have essentially ignored this aspect of criminality. Ward et al. (2006) suggest that Marshall and Barbaree altered the theoretical landscape (within sexual abuse research) by simply acknowledging the component of nonsexual criminality.

Like the critics of Gottfredson and Hirschi (1990), researchers have also argued that the Integrated Theory is too general (Ward 2002). Their main criticism is that it does not account for important distinctions between types of sexual offenders (i.e., distinguishing, e.g., between rapists and child molesters) (Ward 2002). Drawing on typological traditions, Ward recommends that theorists should "develop more specific rather than general explanations" (p. 217). Evidently, influenced by the work of Gottfredson and Hirschi, the criminological tradition has moved in the opposite direction.

Conditioning Theory of Sexual Offending. Also published in 1990, Laws and Marshall's Conditioning Theory provides a coherent explanation of specialization in sexual offending by emphasizing learning experiences and motivation. Like other well-regarded theories of sexual abuse, Laws and Marshall (1990) do not preclude a sexual offender from being versatile. Instead, they simply do not make any reference to a sexual offender's nonsexual offending. In this way, they seek to explain only how an individual with certain experiences will come to offend sexually.

Conditioning Theory provides a theoretical basis for the deviant sexual preference hypothesis (Ward et al. 2006). Although it has a specific sexual offender focus, the theory argues that sexual deviations are learned responses to possibly accidental experiences with sexually deviant behavior (Ward et al. 2006). That is, they could evolve in anyone.

Deviant arousal patterns are said to develop through the process of classical conditioning and the "elaborated use of deviant fantasy in masturbation" (Laws and Marshall 1990, p. 226). This is likely a gradual process occurring during masturbation to a memory which need not have been sexually stimulating at the time of the initial experience (Ward et al. 2006). Classical conditioning describes the unconscious and "repetitious or traumatic pairing of sexuality and some negative experience [which] produces some type of intensive emotional response that distorts subsequent sexual gratification" (Schwartz and Cellini 1996, p. 14). This explains the unconscious compulsion to reenact one's own abuse to gain mastery over the experience. This has been supported empirically in samples of sexual offenders, particularly for juveniles.

A Theory of Specialization and *Versatility*. At this point, two compelling empirical realities must be considered and reconciled. First, it is known that not all sexual offenders present with deviant sexual preferences or deviant sexual arousal (Simon 2000; Weinrott and Saylor 1991). Second, it is known that many people who commit sexual offenses also engage in a considerable amount of nonsexual criminal behavior in addition to their sexual offending (Lussier 2005; Simon 2000; Weinrott and Saylor 1991). What is missing from these perspectives is a sensible explanation of two realities: sexual offending by men without deviant sexual interests and nonsexual offending committed by men who have explicitly been designated "sexual offenders."

Harris et al. (2009a) concluded that the components of Gottfredson and Hirschi's (1990) theory were appropriate for almost all (89 %) of the 506 sexual offenders in their sample. They indeed revealed the analogous behaviors that were predicted by the General Theory of Crime. At the same time, it still appeared that a small group of highly specialized offenders warranted the more sexually specific explanations such as those provided above (Marshall and Barbaree 1990; Laws and Marshall 1990). That is, for the remaining 11 % of the sample it was relevant to focus on such factors as sexual preoccupation; emotional congruence with children; and having a male, unrelated, or stranger victim. Essentially, the authors suggested that some aspects of sexual offending could be understood akin to the way Moffitt (1993) explained delinquency.

Borrowing Moffitt's (1993) framework amounts to a "Dual Taxonomy of Sexual Offending" where two distinct trajectories of offending are proposed. Thus, the observation of sexual offenses is understood to conceal two qualitatively distinct types of offending behavior: sexual offenses that are committed rarely or sporadically as part of the broader range of criminal activity in which an individual engages and sexual offenses that are committed by highly specialized offenders who seldom, if ever, engage in any other rule-breaking behavior.

The first category of offenders were referred to as "versatile" and made up the vast majority of offenders. They were characterized by general indicators of low self-control including substance abuse, adolescent antisocial behavior, elementary school problems, and unemployment (Harris et al. 2009a). They were more criminally versatile, engaged in a broader range of different criminal behaviors, and were more likely to offend upon release. This is in line with the results of a previous study (Meloy 2005) which concluded that the majority of sexual offenders could be described as situational or opportunistic, with versatile criminal histories.

The second category was "highly specialized" and represented 11 % of offenders, comparable in size to Moffitt's life-course-persistents or Wolfgang's chronic offenders. Contrary to their versatile counterparts, this group consisted mostly of child molesters and was characterized by sexual preoccupation, emotional congruence with children, and a greater likelihood of abusing known, related, and male victims (Harris et al. 2009a). Highly specialized sexual offenders were also statistically significantly more likely than versatile offenders to begin their offending in adulthood. This group best resembled the expectations of the theoretical perspectives that expect specialization (Laws and Marshall 1990; Marshall and Barbaree 1990).

It is definitely conceivable that more groups could be identified, particularly given the sample used in the study (Harris et al. 2009a) which is arguably biased toward sexual offenders in the direction of specialization. The group that is most notably absent in that study are offenders who have *not* been convicted of a sexual offense. Although the studies reviewed for this entry have demonstrated considerable similarities between convicted sexual offenders and the theoretical expectations of more generic nonsexual offenders identified in the literature, the opportunity certainly remains to assess this similarity empirically by including a control group of nonsexual offenders. This represents a worthwhile area of future inquiry.

Methodological Considerations in Specialization Research

Extant research on offense specialization is confounded by methodological inconsistencies which has an impact on current understanding of specialization. The most common concerns include the operational definitions of specialization and inconsistencies in the number and type of crime category classifications that are used (McGloin et al. 2007). Within the field of research on sexual offending, the added question of offender classification also impedes one's ability to compare results across studies..

Definition of Specialization. Many different techniques exist to define and measure specialization. Although each one has limitations when used alone, a combination of measures can be used together to paint a clearer picture of specialization. The simplest measure, the Specialization Threshold (ST), is a crude cutoff that detects specialization if an arbitrary proportion of one's arrests is for a particular offense or category of offenses (Cohen 1986; Harris et al. 2009b). Various thresholds have been used in prior research but Cohen's (1986) definition is used most often and declares specialization if 50 % of a person's arrests are for a particular type of crime. As the most simplistic measurement of specialization, the ST's main challenge is its inability to capture change or crime switching over time. The Forward Specialization Coefficient (FSC) (Farrington et al. 1988) and Diversity Index (DI) attend to that limitation by taking into account transitions between offenses (Sullivan et al. 2006). These models generate elegant coefficients that can be compared across studies. Although the FSC relies on consecutive transitions only, the DI can be calculated on nonadjacent transitions as well (McGloin et al. 2007). For some researchers, the DI has been preferred over the FSC for its more intuitive application and interpretation (Harris 2008).

Number of Crime Categories. It is difficult to compare results across studies because researchers seldom apply the same or similar crime classification schemes. Although there is no consensus on which to use (Sullivan et al. 2006), the convention in criminology has been three or four offending categories: personal/violence, property, drugs, and other (where "drugs" and "other" are sometimes combined) (Mazerolle et al. 2000). Mazerolle and colleagues showed that the outcomes of using three or four categories are substantively similar and opted for the parsimonious model of combining "drugs" and "other." So far, the emerging research on sexual offending has further integrated "personal" offenses into sexual violence and nonsexual violence.

Use of Official Data. The limitations of studying crime using official statistics have been discussed elsewhere (Farrington et al. 1988; Rice et al. 2006). In addition to the actual nature of their behavior and whether they got caught, an individual's criminal record can also be influenced by a victim's willingness to report an offense, local law enforcement practices, and sentencing policies (Harris 2008). Further, in the case of sexual offending, official statistics tend to underestimate the "severity of sexually motivated violent offenses" (Rice et al. 2006, p. 525). Procedures such as plea bargaining often mean that the actual charges brought against a person no longer represent the sexual nature of the crime (e.g., rape being pled down to assault). Despite these disadvantages, official statistics remain the most common source of data for studies of specialization, but their subsequent results should be considered with these limitations in mind. A combination of official records and self-reported data is recommended.

Number of Previous Criminal Events. The concept of specialization implies the presence of another important element of the criminal career paradigm: persistence. Because it is meaningless to discuss specialization in one or two offenses, specialization research is often restricted to frequent and persistent offenders or "career criminals." Of course, analyzing only those offenders with more than five arrests, for example, greatly reduces one's sample of offenders. In addition to reduced statistical power, this introduces a possible sample bias by including only sufficiently persistent offenders. Although some researchers have limited their analysis to participants with at least five offenses or nine offenses, the general convention is to include participants with two or more (Farrington et al. 1988; Harris 2008; Harris et al. 2009b; Sullivan et al. 2006) or three or more separate arrests or sentencing occasions.

Use of Arrest or Conviction. Specialization studies either use arrests, convictions, or sentencing occasions as the primary unit of analysis. Because each occasion might include multiple counts or charges, the convention is usually to record only the most serious charge at each event (Farrington et al. 1988). Using charges will hide arrests that failed to result in a charge and using sentencing occasions will hide convictions that failed to result in a sentence. The seriousness of most sexual offenses (compared with almost all other crime) means that recording only the most serious offense would conceal other co-occurring behaviors, thus inflating specialization. Although the opposite practice (of including every individual charge) has the complementary limitation of inflating versatility, given that the dark figure of sexual crime is so substantial, this method is considered the most appropriate for samples of sexual offenders (Harris 2008).

Classifying Sexual Offenders. Specialization is usually taken to mean the likelihood of committing the same offense on a subsequent occasion. Blumstein et al. (1986) notion of "offense clusters" broadened this definition so that rather than only detecting specialization in burglars who subsequently commit burglary, for example, (Britt 1994), an offense cluster of theft could be used to encompass such crimes as burglary, larceny, shoplifting, or fraud (Britt 1994). This approach is important for the present discussion because it is argued that some sexual offenders specialize *within* sexual offense categories (Soothill et al. 2000).

Consider a male child molester who is subsequently convicted for rape of a woman. Blumstein et al. (1986) would consider him a specialist sexual offender. A clinician, however, would consider the same individual quite differently from an exclusive child molester (who, e.g., might demonstrate the more typical characteristics of emotional congruence with children and difficulty in maintaining adult relationships). To Blumstein et al., he would be a "sexual offender," but to a clinician, although his offenses might all be sexual, he apparently shows no predilection for certain types of victims and might be considered a more challenging candidate for treatment. A mixed offending pattern (that includes both adult and child or male and female victims) might also indicate that he is potentially more likely to reoffend upon release, because his sexual crimes are unpredictable and his victim preferences are unclear. These characteristics are also those that would render an individual at higher risk of reoffending (according to many risk assessment tools) than his more specialized counterparts.

Empirical Evidence of Specialization and Versatility in Sexual Offenders

Almost a quarter century of research on offense specialization is summarized by the findings of one of the first articles on the topic: there is a "small but significant degree of specialization in offending superimposed on a great deal of versatility" (Farrington et al. 1988, p. 461). Regardless of method or sample, there is a clear indication of versatility among offenders within the criminological literature. A modest amount of specialization is detected in some samples, and this is most notably observed in persistent offenders who use drugs.

Despite clinical assumptions to the contrary, research that has focused specifically on sexual offenders over the last 10 years has drawn similar conclusions: many sexual offenders do not restrict their criminal activities to sexual offenses (Lussier 2005; Miethe et al. 2006; Simon 2000; Smallbone and Wortley 2004; Soothill et al. 2000; Weinrott and Saylor 1991; Zimring et al. 2007). Further, notwithstanding possible differences in reporting, arrest, and conviction rates for sexual and nonsexual offenses, adult sexual offenders are about twice as likely to be convicted for nonsexual offenses as they are for sexual offenses, both before and after being convicted of a sexual offense (Smallbone and Wortley 2004). The following section details the findings of a selection of specific studies in this area.

As early as 1991, Weinrott and Saylor (1991) concluded that specialization in sexual crimes was relatively rare. The 99 convicted sexual offenders in their study collectively self-reported nearly 20,000 nonsexual crimes in just the 12 months prior to their incarceration. The majority of these offenses were minor and mostly included public intoxication and petty theft.

Smallbone et al. (2003) found substantial offending versatility in the official histories of the 88 incarcerated adult male sexual offenders in their sample. The participants in their study included 33 rapists, 29 extrafamilial child molesters, and 26 incest offenders. Almost two thirds (60 %) of their total sample and 88 % of the offenders who recidivated reported having prior nonsexual criminal convictions.

Smallbone and Wortley (2004) later assessed the extent of persistence and versatility in the self-reported offending behaviors of 207 child molesters. Their sample was broken down further by victim relationship with 98 participants having committed incest exclusively, 72 having extrafamilial victims, and 37 having both intrafamilial and extrafamilial victims. Results from an extensive self-report questionnaire revealed considerable versatility with 69 % of the sample having been convicted of at least one nonsexual offense and 80 % having first been convicted for a nonsexual crime.

In an ambitious examination of almost 10,000 sexual offenders and some 24,000 nonsexual offenders released from prisons in 15 US states in 1994, Miethe et al. (2006) also found low levels of persistence and specialization among sexual offenders compared to nonsexual offenders. Sexual offenders had fewer arrests of all kinds than any other offender group except murderers. Although 70 % of the sexual offender sample had been arrested only once, some 95 % of those who had been arrested more than once had nonsexual offenses in their arrest histories.

Harris et al. (2009b) explored offense specialization in the officially recorded criminal histories of 506 sexual offenders referred for civil commitment in Massachusetts. They examined the sample by offense type (comparing rapists, child molesters, incest offenders, and offenders with mix-aged victims) and by referral status (comparing offenders who were observed and released with those who were observed and committed for treatment). Contrary to their hypotheses, their results indicated that treated participants were no more likely to specialize in sexual offenses than observed participants. When compared by offense type, as expected, child molesters were substantially more likely than rapists to specialize in sexual offenses. Child molesters were also more likely to specialize in sexual offenses against children (than rapists were to specialize in sexual offenses against adults). As a whole, the sample was also more likely to specialize in sexual offenses than in the other offending categories of nonsexual violence, property, and other offenses.

Later, Harris et al. (2011) examined the post-release offending records of the same sample of 506 observed and committed offenders for a period of up to 10 years. Again, they found that the tendency to be criminally versatile remained the most compelling offending pattern across the sample. Few differences were detected between treated and observed participants, and almost no differences were found between rapists and child molesters with respect to recidivism or to post-release specialization. Rapists were more likely than child molesters to reoffend at all and to reoffend violently. There were no differences between groups on their likelihood of committing sexual offenses after release.

Interestingly, specialization prior to incarceration was not indicative of specialization in sexual offending after release. Of the 53 participants who specialized after release, approximately half had specialized criminal histories (n = 27) and half did not (n = 26) (Harris et al. 2011). The authors concluded that the predictive power of a sexually specialized criminal history was limited.

Harris and Knight (forthcoming) elaborated upon the findings from the previous two studies by using an addition of almost 300 more cases and attempting to establish the existence of an identifiable group of highly specialized sexual offenders that may have been masked by the liberal definition of specialization (50 %) used in prior works. A small group of 166 highly specialized sexual offenders existed out of a larger group of 748 offenders and was made up mostly of child molesters. This was consistent with both existing empirical evidence and the findings from Harris et al. (2009b) which concluded that child molesters were the most likely offender type to specialize in sexual offending.

A deeper analysis of the sample's responses to various items from risk assessment tools demonstrated that the majority of the sample shared many of the general criminal attributes proposed by conventional criminology including the presence of analogous behaviors such as alcohol and drug use, problems in elementary school and adolescent behavior, and involvement in a range of different types of crimes. By comparison, the small group of highly specialized offenders was characterized instead by the variables identified by more specific theories of sexual offending. This study provided added empirical support for the Dual Taxonomy of Sexual Offending described earlier.

Specialization Within *Sexual Offenses.* As mentioned earlier, sexual abuse researchers distinguish between offenders who have sexually assaulted adults (rapists) and offenders who have sexually abused children (child molesters) (Meloy 2005; Simon 1997). Further distinctions are also made between familial and nonfamilial child molesters. A review of recent studies that have addressed the question of specialization *within* sexual offenses is provided here.

Rapists are repeatedly found to resemble violent nonsexual offenders most closely (Harris 2008). They are thought to be predominantly versatile in their offending and the commission of rape is seen to be part of a broader propensity to act in a generally antisocial manner (Lussier et al. 2005; Smallbone et al. 2003). Each of the 37 convicted rapists in Weinrott and Saylor's (1991) sample admitted to at least one nonsexual offense. Further, over half of the sample committed burglary and almost half committed robbery (Weinrott and Saylor 1991). The mean number of different offenses (out of a possible list of 22) that was self-reported by each rapist was 10.5 (Weinrott and Saylor 1991). Further, a third of convicted rapists admitted to having had sexual contact with a child (Weinrott and Saylor 1991).

Lussier et al. (2005) found that when compared with child molesters, rapists engaged in a larger variety of all crimes, had a higher frequency of property and violent crimes, and had also committed a "significantly greater variety of violent crimes than of sexual crimes" (p. 180). Addressing specialization using recidivism statistics, Langan and Levin (2002) found that 3 years after incarceration, drug offenders (41 %), larcenists (34 %), burglars (23 %), nonsexually violent offenders (23 %), defrauders (19 %), and robbers (13 %) were all much more likely than rapists (2 %) to be rearrested for the same type of crime for which they were originally convicted.

Child molesters, by comparison, are generally considered much less versatile than rapists (Harris 2008; Harris et al. 2009b). In fact, if a specialist, persistent offender exists, the stereotypical extrafamilial child molester is likely the most convincing example. Some evidence of crime switching has been detected in child molesters. For example, 12 % of the convicted child molesters in Weinrott and Saylor's (1991) sample admitted to attempting forced sex with an adult female and 34 % had abused children inside the family as well as outside the home. Lussier et al. (2005) also concluded that child molesters committed a variety of different sexual crimes. A quarter of Weinrott and Saylor's (1991) sample reported substance abuse and 20 % reported assault, theft, burglary, possession of stolen goods, and drug-related offenses in the year prior to incarceration.

Finally, incest offenders tend to most closely resemble non-offenders. When compared to rapists and child molesters, they display more social competence, having maintained steady employment and a stable marriage or marital-type relationship. Incest offenders are often thought to present with the lowest risk of reoffending, having engaged in a discrete and isolated event. But this idea has been challenged by evidence of versatility among samples of intrafamilial offenders (Harris 2008). Although a small sample, half of the 29 incest offenders in Smallbone et al. (2003) study "had at least one previous conviction for a nonsexual offense" (p. 57). All of Weinrott and Saylor's (1991) also small sample of 18 incest offenders self-reported at least one nonsexual offense. Other studies also find that incest offenders have reported committing nonsexual violent crimes and sexual crimes outside the home (Harris 2008). For example, half of Weinrott and Saylor's (1991) sample of incest offenders admitted to abusing extrafamilial children.

A Lingering Methodological Concern

In the methodological concerns raised above, the "offense cluster" of sexual offending was seen as irrelevant for sexual offenders. The research presented here, however, indicates that a substantial number of sexual offenders commit sexual offenses *outside* their assumed "specialty." The observation that identified rapists offended against children and identified child molesters against adults poses a unique problem for specialization studies that has not yet been evaluated properly. Much is made in the theoretical and treatment literature about the empirically significant and clinically valuable differences between rapists and child molesters.

The response to this conundrum is not a simple one because the distinctions between offender types are clearly valuable in practice. But, within the context of research, the considerable amount of crime switching results in the label of "rapist" becoming meaningless for a man who later admits to having child victims. Likewise, calling someone a "child molester" is similarly misleading if he has also offended against adults.

Policy Implications

Men convicted of sexual crimes (particularly against children) are now subject to specific, expansive, discretionary, and controversial legislation that has impacted almost every stage of the criminal justice system (Meloy 2005). Policies include community registration and notification, residency restrictions, mandatory specialized treatment, and civil commitment. Taken together, the studies reviewed above support existing criminological perspectives regarding versatility and challenge the assumption of specialization. By concluding that sexual offenders are largely indistinguishable from general criminals, these findings cast considerable doubt upon the value of these policies.

The focus of many sexual offender laws is to address the most chronic and persistent offenders. Evidently, specialized, persistent, and predatory sexual offenders do exist but they are an identifiable minority. So, the laws and regulations aimed specifically at them amount to a substantial financial commitment toward policies that addresses the rarest of circumstances. Further, it is now well established that children are much more likely to be abused by those who are known or related to them. Thus, a law that notifies a community of an offenders' name and address or that restricts offenders from living within a certain distance of a school or park is fundamentally flawed. These laws have the unfortunate consequence of increasing homelessness for an already disenfranchised population and reducing employment opportunities, thus elevating the likelihood of reoffending. The considerable burden that these laws place on police, parole, and probation officers should also be taken into account.

Concluding Remarks

Research now indicates that many sexual offenders neither restrict their criminal behavior to sexual offenses nor their sexual offending behavior to specific sexual offense subtypes (Soothill et al. 2000; Weinrott and Saylor 1991). Even accounting for the bias inherent in clinical samples of sexual offenders, or the difference between official statistics and self-reported data, the ultimate message is the same: a broad trend of versatility is maintained across the studies. When compared by offender classification, child molesters are consistently found to be more likely (and rapists less likely) to specialize in sexual offenses. Although incest is predominantly seen as a distinct behavior, existing studies of these men indicate that they also offend outside the home.

Various methodological considerations were examined. While it is clear that no approach is without limitations, researchers should manage these obstacles by using multiple measures of offending, multiple definitions of specialization, and mixed methods of analysis. The construct of offense specialization versus versatility should be considered in terms of the extent both to which sexual offenders commit nonsexual offenses and to which they commit sexual offenses outside their presumed sexual offense subtype.

Broader prevention efforts clearly need to do much more than rely on the selective policing and incapacitation of identified sexual offenders because there is now sufficient evidence to indicate that sexual offenders are generally criminally versatile. Individual offending tendencies are likely dynamic and could be influenced by a range of factors that remain underexplored. How these offending patterns emerge and perhaps change over time requires further examination. Such inquiries could in turn inform policy makers, criminal justice practitioners, and politicians alike with empirically valid evidence-based alternatives from which decisions regarding community safety could be made.

Related Entries

- ▶ Criminal Careers
- ▶ Measuring Crime Specializations and Concentrations
- ▶ Offense Specialization: Key Theories and Methods

Recommended Reading and References

Blumstein A, Cohen J, Roth J, Visher C (1986) Criminal careers and career criminals. National Academy Press, Washington, DC

Britt C (1994) Participation and frequency. In: Hirschi T, Gottfredson M (eds) The generality of deviance. Transaction, New Brunswick

Cohen J (1986) Research on criminal careers: individual frequency rates and offense seriousness. In: Blumstein A, Cohen J, Roth J, Visher C (eds) Criminal careers and career criminals. National Academy Press, Washington, DC

Farrington D, Snyder H, Finnegan T (1988) Specialization in juvenile court careers. Criminology 26(3):461–487

Gottfredson M, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Harris DA (2008) Offense specialization and versatility in a sample of male sexual offenders referred for civil commitment. Unpublished PhD dissertation, Griffith University

Harris DA, Mazerolle P, Knight RA (2009a) Understanding male sexual offending: a comparison of general and specialist theories. Crim Justice Behav 36(10):1051–1069

Harris DA, Smallbone S, Dennison S, Knight RA (2009b) Specialization and versatility in sexual offenders referred for civil commitment. J Crim Just 37:37–44

Harris DA, Knight RA, Dennison S, Smallbone S (2011) Post release specialization and versatility in sexual offenders referred for civil commitment. Sex Abuse 23:243–259

Langan P, Levin D (2002) Recidivism of prisoners released in 1994. US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics, Washington, DC

Laws D, Marshall W (1990) A conditioning theory of the etiology and maintenance of deviant sexual preference and behaviour. In: Marshall W, Laws D, Barbaree H (eds) Handbook of sexual assault: issues, theories and treatment of the offender. Plenum Press, New York

Lussier P (2005) The criminal activity of sexual offenders in adulthood: revisiting the specialization debate. Sex Abuse: J Res Treatment 17(3):269–292

Lussier P, LeBlanc M, Proulx J (2005) The generality of criminal behaviour: a confirmatory factor analysis of the criminal activity of sex offenders in adulthood. J Crim Just 33:177–189

Marshall W, Barbaree H (1990) An integrated theory of the etiology of sexual offending. In: Marshall W, Laws D, Barbaree H (eds) Handbook of sexual assault: issues, theories and treatment of the offender. Plenum, New York

Mazerolle P, Brame R, Paternoster R, Piquero A, Dean C (2000) Onset age, persistence, and offending versatility: comparisons across gender. Criminol 38(4):1143–1172

McGloin J, Sullivan C, Piquero A, Pratt C (2007) Local life circumstances and offending specialization/versatility: comparing opportunity and propensity models. J Res Crime Delinq 44(3):321–346

Meloy M (2005) The sex offender next door: an analysis of recidivism, risk factors, and deterrence of sex offenders on probation. Crim Justice Policy Rev 16(2):211–236

Miethe T, Olson J, Mitchell O (2006) Specialization and persistence in the arrest histories of sex offenders: a comparative analysis of alternative measures and offense types. J Res Crime Delinq 43(3):204–229

Moffitt T (1993) Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. Psychol Rev 100:674–701

Rice M, Harris G, Lang C, Cormier C (2006) Violent sex offenses: how are they best measured from official records? Law Hum Behav 30:525–541

Schwartz B, Cellini R (1996) The sex offender: new insights, treatment innovations and legal developments. Civic Research Institute, Kingston

Simon L (1997) The myth of sex offender specialization: an empirical analysis. N Engl J Crim Civ Confin 23:387–403

Simon L (2000) An examination of the assumptions of specialization, mental disorder, and dangerousness in sex offenders. Behav Sci Law 18:275–308

Smallbone S, Wheaton J, Hourigan D (2003) Trait empathy and criminal versatility in sexual offenders. Sex Abuse: J Res Treatment 15(1):49–60

Smallbone S, Wortley R (2004) Onset, persistence, and versatility of offending among adult males convicted of sexual offenses against children. Sex Abuse: J Res Treat 16(4):285–298

Smallbone S, Marshall W, Wortley R (2008) Preventing child sexual abuse: evidence, policy and practice. Willan, Cullompton

Soothill K, Francis B, Sanderson S, Ackerley E (2000) Sex offenders: specialists, generalists – or both? Brit J Criminol 40:56–67

Sullivan C, McGloin J, Pratt T, Piquero A (2006) Rethinking the "norm" of offender generality: investigating specialization in the short-term. Criminology 44(1):199–233

Ward T (2002) Marshall and Barbaree's integrated theory of child sexual abuse: a critique. Psychol Crime Law 8:209–228

Ward T, Polaschek D, Beech A (2006) Theories of sexual offending. Wiley, West Sussex

Weinrott M, Saylor M (1991) Self-report of crimes committed by sex offenders. J Interpers Violence 6:286–300

Zimring F, Piquero A, Jennings W (2007) Sexual delinquency in Racine: does early sex offending predict later sex offending in youth and young adulthood? Criminol Public Policy 6(3):507–534

Specialization in Juvenile Offending

Jesse Cale
Key Centre for Ethics, Law, Justice and
Governance, Griffith University, Brisbane, QLD,
Australia

Synonyms

Crime repetition by youth; Patterns in youthful offending

Overview

Whether and to what extent juvenile offenders specialize in committing certain types of crimes during their delinquent careers has substantial implications for theory and policy. The question of offender specialization arguably first emerged with the advent of the positivist school in criminology in the late nineteenth century when Caesare Lombroso sought to identify criminal types. With the introduction of specialized juvenile justice in laws in the early twentieth century and their fundamental reforms in many countries in the second half of the twentieth century and into the initial decade of the twenty-first century, there have been considerable theorizing, research, and debate about specialization in youthful offending. Political debates in countries such as Australia, Canada, England-Wales, and the United States concerning the reform of youth justice laws have been influenced, in part, by the considerable research on specialization youthful offending. The main theoretical debate has been focused on the attempts to specify potential distinguishing characteristics of certain types of violent juvenile delinquents such as serious sexual offenders, murderers, and gang members, compared to less serious violent offenders, property offenders, and drug users/traffickers, for example. Many diverse labels have been applied to young offenders, such as serial fire-setter/rapist/murderer, career criminal, life-course-persistent/adolescence-limited offender, chronic/nonchronic offender, and adolescent psychopath/conduct disordered, that typically include attempts to specify types and patterns of crime. Oftentimes these labels or categories have been introduced into the media, political, and legislative circles, and debates emerge about whether to incarcerate, punish, and/or rehabilitate juvenile delinquents depending on what "type" of offender they are.

The essence of the theoretical debate is whether juvenile offenders can be classified into an "offender type" based on the qualitative nature, and/or quantity of their offending, versus whether there is more simply a single type of juvenile offender who engages in diverse and multiple antisocial and criminal behaviors. Not surprisingly, a considerable empirical research theme over the past half a century has focused on the nature and extent to which juvenile offenders demonstrate specialization in offending. The related theoretical debate has centered on considering that if specialization is not evident, then a single or general theory of juvenile delinquency could be sufficient to explain this phenomenon. In turn, criminal justice responses, therefore, could be directed toward a central set of causes or risk/protective factors to prevent juvenile delinquency from beginning as well as reoccurring. In contrast, if research demonstrates specialization or a diversity of types of juvenile delinquent offending patterns, then typologies require far more complex and unique theoretical explanations along with specialized juvenile justice responses, including treatment and intervention strategies, depending on the type of juvenile delinquent or young offender. To better understand how this debate on specialization has evolved over the years, it is necessary to describe the key theoretical controversies historically.

Key Issues and Controversies: Theoretical Debates

Typologies of adult and youth offenders gained popularity in the 1960s with the assertion that offenders could be classified into conceptually valid types based on fundamental differences in the crimes they committed. The most basic and obvious crime-centered typologies generally consisted of nominal categories of specialized criminal behavior consisting of property versus violent offender types. More types emerged as well including white-collar criminals, organized crime, street gangs, and political terrorists, for example. It was then theorized that these criminal types were explained by different sets of causal factors.

By the 1980s, there was a wide range of theories to explain ever-increasing types of offenders. In turn, typologies and related theories generated intense debates concerning basic definitions of deviance, delinquency, and criminality that are essential to assessing youthful offending specialization. For example, the assertion that delinquency was overwhelmingly associated with youth from low-income groups along with those belonging to certain ethnic/racial minorities was challenged by self-report studies of delinquency (e.g., Elliott and Ageton 1980). These studies demonstrated that minor delinquency was common for youth from all income and ethnic groups. A second criticism of initial crime-centered typologies was the reliance on offense records whereby the presence of a particular type of offense (e.g., sexual assault) determined a youth's "specialization" in offending. Critically, such a snapshot-based examination of offending patterns did not consider variations in offending over time. The longitudinal cohort studies of the 1970s and subsequent major cross-sectional studies provided a more valid basis for establishing typologies and specialization in youthful offending because they took into consideration changes in the frequency of offense types over the entire course of adolescence. Finally, a third key criticism involved arbitrary and income/ethnic group biases of juvenile justice laws regarding charges and convictions up to the 1970s. For example, juvenile justice laws from the early 1900s to the early 1970s virtually criminalized many childhood and adolescent antisocial and deviant behaviors. Typologies ranged from the enormous diversity of trivial antisocial behaviors such as status offenses (e.g., swearing and drinking alcohol) to the most violent crimes (e.g., rape and murder). However, because juvenile justice laws governed the ages between 8 and 21 depending on the country and within-country jurisdiction (e.g., state, province, region), it was abundantly clear that attempts to assess the nature and extent of specialization in juvenile offending had to be based on more sophisticated approaches.

What emerged from both longitudinal cohort and major cross-sectional research was that simple typologies based on the presence of certain offenses in a juvenile's history were inadequate for assessing specialization for several reasons. For example, status, property and violent crimes committed by juveniles varied substantially in not only degree of seriousness but frequency. To assess specialization, theoretically driven and complex equation-based indices often consisting of ratios of different types of crime such as property compared to violent crimes became central to investigating specialization in youthful offending. Seriousness of types of crimes, for example, based on the extent or amount of damage in dollars involved with property crimes, and the nature and extent of injuries for violent crimes also were incorporated into indices determining offender types. Another dimension was the variation in the extent of switching to different types of crime in certain developmental periods, such as childhood, adolescence, and adulthood. In other words, early crime-centered typologies were unable to address rapidly emerging evidence that delinquent careers were characterized by significant heterogeneity involving different patterns in the type, seriousness, and frequency of offenses committed over childhood and adolescence.

Several of the key concepts from the criminal career paradigm, particularly the concept of the career criminal, evoked intense theoretical and empirical debate. Klein (Klein and Malcolm 1979; Klein 1984) maintained that an abundance of empirical evidence demonstrated that juvenile offending was predominately unpatterned and reflected the notion of "cafeteria-style" offending. Subsequently, Gottfredson and Hirschi (1990) introduced their General Theory of Crime based on Hirschi's (1969) original research and his Bonding Theory of social control. This theory asserts that specialization does not exist and instead delinquents engage in a wide range of delinquent behaviors, stemming broken bonds to family and other societal institutions such as the church, school, and community. The absence of strong prosocial bonds further explains the time-stable personality trait of low self-control, which in turn is caused by lax and/or inconsistent parenting practices. Low self-control is established in childhood and remains stable throughout life. It is manifested by the pursuit of easy and instant gratifications inherent in deviant behaviors such as smoking, excessive drinking, short-term and multiple sexual conquests, recklessness, and crime, when opportunities are present. In effect, the General Theory of Crime predicts that juvenile offending is best understood as involving substantial offense versatility.

By the 1990s the developmental theoretical perspective of crime introduced more complex predictions about patterns in juvenile offending such as specialization and versatility. This perspective identified risk and protective factors for several forms of antisocial behavior throughout the life course beginning with early childhood and into early adulthood. Loeber (1990) asserted that children who exhibited recklessness, problems with authority, aggression, and coercive/manipulative behaviors were more likely to follow a more violent pattern in adolescence but also engage in diverse crimes. In contrast, children who were more withdrawn and antisocial were more likely to engage in covert patterns and drug use as adolescents.

The most prominent developmental theory relevant to specialization involved Moffitt's (1993) developmental taxonomy. She predicted that more sustained aggressive and violent life-course patterns were based on the age of onset of offending. Moffitt proposed two distinct types of offenders in the population: life-course-persistent' (LCP) offenders and "adolescence-limited" (AL) offenders. LCP offenders are characterized by the interaction, and accumulation, of individual deficits and environmental adversities that begin early in childhood and cascade over subsequent developmental periods. In effect, these individuals begin offending early in childhood/adolescence, engage in a wide range of deviant and criminal acts including bullying and other violent behavior, and persist into adulthood. In adolescence, therefore, this developmental group was predicted to exhibit frequent, versatile, and serious offending including violence. In contrast, AL offenders are predominately characterized by risk factors that emerge in adolescence. The primary risks involve status frustration driven by the expanding maturity gap between biological and social maturities (i.e., the observation that while adolescents experience physical or biological maturity at earlier ages when compared to the early twentieth century, their social maturity as evidenced by the age they assume adult social roles such as husband, breadwinner, father is delayed), as well as exposure to deviant or delinquent peers and role models. AL offending was predicted to be more specialized than LCP offending, in the context of offenses reflecting the effect of the maturity gap such as vandalism, public order, substance abuse, and status offenses.

As with all classic debates in criminology, measurement of key constructs are often central to resolving competing hypotheses, in this instance, about specialization versus versatility in youthful offending.

The Measurement of Specialization in Youthful Offending

Historically, methodological and measurement innovations have had to take into consideration the inherent limitations associated with data sources and especially official juvenile justice information recording. Most importantly, there is a consensus that official data significantly underrepresents the total amount of crimes committed, and even more for official juvenile offending data. During the past 30 years, measures based on informal and formal diversion programs have been introduced in many Western countries. Often, informal diversions from formal justice proceedings have not been recorded. In addition, there has been concern over the variations in the offense recording and charging process across jurisdictions even within countries. When only the most serious offense is proceeded with in terms of offense recording, trial, and/or sentencing, both the range and seriousness of actual offenses are under-recorded. This affects the accuracy of officially recorded crime records. Another limitation has been the concern with variations in the number of police charges that are recorded when police practices include deliberately over charging in order to enhance their ability to gain information about other crimes and to an advantage in plea bargaining to gain a confession of guilt. As well, police and prosecutors utilize their ability to increase or decrease the seriousness of the charges since criminal codes include wide discretion concerning the degree within a criminal offense category. For example, a violent assault resulting in death ranges from manslaughter to first-degree murder. All of these concerns raise questions over the validity of measures when using, for example, charges versus convictions in assessing specialization, especially across different juvenile justice jurisdictions included in comparative research designs. Similar validity concerns have been a source of debate with self-report methodologies utilized to assess specialization. Most importantly and especially when youth are involved, memory recall is often problematic in terms of overstating offenses, and the timelines in which they occurred in their past. Nonetheless, despite these limitations, most research on specialization in youthful offending has used a combination of methodologies to identify typologies based on a variety of specific measurement criteria. One advantage with official data is that a more complete delineation of potentially different offending trajectories over time is possible because this data can record patterns across childhood, adolescence, and adulthood. Similarly, complex cohort designs typically include multiple data recording approaches (e.g., including both official and self-report data) to measure the types, frequency, and temporal patterns of offending.

The seminal Philadelphia Birth Cohort Study (Wolfgang et al. 1972) was the most widely recognized initial methodological innovation in juvenile offender specialization research. Wolfgang and colleagues examined crime type patterns from offense to offense as a series of transitions using transition matrices. This method assessed the probability of juvenile offenders having received charges for similar crimes across adjacent offending events. Transition matrices identified whether specialization in certain crime types was evident by evaluating whether the probability of being charged with the same offense consecutively was greater in value than the probability of being charged with a different offense. Importantly, these techniques were also adopted to assess specialization using major cross-sectional research that included the complete court histories of charges and convictions for juvenile offenders (e.g., Farrington et al. 1988).

Related to the transition matrix methodology was the development of statistical tests in the 1980s to measure whether the probability of observed transitions to similar offense types was greater than could be expected on the basis of chance alone. Coefficients representing the level of specialization for specific types of offending in a sample became a standard measure. For example, Bursik (1980) introduced the Adjusted Standardized Residual (ASR) that determined whether the ratio between observed and expected transition values was statistically significant. Farrington et al. (1988) further introduced the Forward Specialization Coefficient (FSC) to control for the influence of infrequent cases that would distort the basic observed to expected ratio. The FSC represents a single value summary of specialization for different offense types based on the sequence of similar offenses in transition matrices. Fisher and Ross (2006) more recently asserted that the FSC is the most widely utilized measure of specialization to date because of its theoretical sophistication and predictive validity.

Despite widespread use, the transition matrix approach generally, and the FSC specifically, has several limitations (e.g., see Britt 1996; Fisher and Ross 2006; LeBlanc and Frechette 1989; Nieuwbeerta et al. 2011; Sullivan et al. 2009). Most importantly, the FSC describes specialization aggregately for a group of delinquents. This aggregate measure is theoretically appropriate for the explanation of the shifting from different types of crime within groups. However, it does not measure shifts in crime for specific individuals. In effect, the aggregate nature of the FSC does not produce information about the individual extent of specialization.

Within-individual change was central to the concept of criminal careers in the 1980s (Blumstein et al. 1986) and the individual-level examination of specialization in juvenile offenders. From this perspective, the measurement focus is on the proportions of similar offense types in individual criminal histories. Here, specialization was defined by the absence of diversity of different types of crime in an individual's offense history. Regarding a specific measurement, for example, Wikstrom (1987) calculated the total proportion of crimes belonging to a single category of offenses in the histories of offenders, which allowed for the utilization of different percentage thresholds for determining specialization. Wikstrom (1987) asserted that if, regardless of the type of offense, two-thirds or more were of one type, then the offender's criminal offending history is considered specialized. However, this measurement methodology is not applicable to occasional or limited offending. In a Swedish sample of delinquents, Wikstrom (1987) identified approximately one-third of juvenile offenders as specialists. Theoretically, different thresholds can be asserted as establishing specialization. In addition, this threshold measurement methodology raises validity concerns depending on the number of types of crimes available for analysis. For example, certain studies can and have calculated a threshold based on three categories of crime, such as property, violent, and status offenses. Others have used even more specific crime categories, making cross-study comparisons problematic because the two-thirds threshold used for three crime types does not appear an appropriate threshold for a larger number of categories. In short, determining specialization based on percentage threshold approaches is somewhat arbitrary and ultimately does not easily capture the nature of offense specialization.

More recently, the extent of specialization in an offender's history has been calculated using the diversity index which also focuses on the proportion of different offenses committed by individuals but also provides a measure of the probability that any two observed offenses drawn randomly from a criminal history belong to dissimilar offense types. In effect, by calculating the proportions of offenses across a specified number of crime categories, this approach provides an individual-level measure of specialization based on the complete offense patterns, rather than the sequence of offenses. This measure calculates the overall amount of similar/dissimilar offenses and allows for the comparison of specialization patterns for both individuals and groups. Again, however, when studies used different numbers of crime categories to calculate the amount of diversity, cross-study comparisons become problematic. In addition, some scholars have asserted that the diversity index does not provide precise information about the type of offense specialization, such as the frequency of different offense types (Sullivan et al. 2009).

Several recent studies have applied new and innovative approaches to measure specialization in juvenile offending (and more generally). These approaches typically involve complex multivariate techniques that have the ability to simultaneously capture the nature and extent of specialization, something that past techniques typically accomplished independently (i.e., the FSC compared to the diversity index). One such approach is Latent Class Analysis (LCA). While not designed specifically for the study of offense specialization, LCA is useful for identifying homogeneous subgroups of offenders that cluster around specific types of offenses within a sample. LCA has been used in the classification of offenders' criminal career trajectories that are based on the varying frequency of total offending over time (Francis et al. 2004). In effect, this technique has the potential to distinguish groups according to both the frequency and type of offenses committed.

Research Findings

The original empirical studies using transition matrices were characterized by mixed findings because of differences in sample composition, data sources, and several methodologies. Nonetheless, most of these studies identified considerable versatility in juvenile offending behavior, which therefore, provided initial support for the General Theory of Crime and the prediction of "cafeteria-style" offending patterns of juvenile offenders (e.g., Klein and Malcolm 1979; Klein 1984). However, several of the initial studies indicated some support for specialization for certain types of minor offending, such as theft (Wolfgang et al. 1972). Subsequent studies in the 1980s found evidence of specialization, again, for less serious crimes such as property and status offenses. For example, Farrington et al. (1988) demonstrated that the most specialized offenses involved status offenses (e.g., running away, liquor, incorrigibility, curfew, and truancy) and property offenses (e.g., burglary, vehicle theft) as well as drug use. As well, there was some indication of a modest tendency toward crime switching to more serious forms of offending, particularly from status to property and property to violent offenses (e.g., LeBlanc and Frechette 1989; Smith et al. 1984). The same pattern of mixed support for specialization was evident in other studies. For example, females were more likely to specialize in status offenses compared to males, who were comparatively more likely to specialize in serious offending (e.g., Kempf 1986, Farrington et al. 1988). Yet still, other studies did not report such differences. Even regarding ethnicity/racial groups, versatility was more prevalent than specialization despite some studies that indicated black youth were slightly more likely to specialize in more serious property offenses than white youth (e.g., Lattimore et al. 1994). Again, several studies did not report this pattern.

By the late 1990s, studies based on developmental theories' predictions concerning specialization in juvenile offending revealed even more complex findings. Mazerolle et al. (2000) examined Moffitt's typology and found AL offenders were more likely to specialize in nonserious offenses compared to LCP offenders. The latter also were more violent (see also Piquero et al. 1999). Subsequent developmental theory-based studies expanded into the adulthood stages revealing that offending trajectories over the life course confirmed the adolescence period had the least amount of specialization with specialization typically emerging later in adulthood. Most importantly, both onset and desistance occurred more often in the adulthood stages than originally predicted. In effect, as offenders aged into adulthood, their offending became more patterned but less frequent.

In sum, three main empirical themes regarding specialization in youthful offending were evident. First, adolescence generally involved the least amount of offense specialization compared to later periods. Second, a small but significant amount of specialization occurred during adolescence. This variation however, typically consisted of the less serious end of the offending spectrum, such as status offenses and theft. However, while this small but significant amount of specialization occurred, versatility was far more prevalent. Third, a small minority of juvenile offenders was responsible for much of the versatile offending committed in adolescence and, most importantly, for more serious violent offending.

Juvenile Justice Policy Implications Related to the Specialization Debate

The original juvenile justice laws reflected early crime-specific theories and crime-centered typologies that presupposed specialization in juvenile delinquency and, in particular, status offenses. Given the apparent extent of nonserious and more specialized offending, welfare/rehabilitation models of juvenile justice were dominant and focused on the family-based causes of delinquency through rehabilitation interventions in juvenile courts. In effect, status offenses were a key target for policy intervention with youth, and in many Western jurisdictions, formal status offender programs proliferated and were based on the hypothesis that youth status offenders would escalate to more serious forms of offending if formal intervention and treatment was not pursued. In other words, one rationale for these programs was that early and formal intervention/treatment was warranted if "status offenders" represented a distinct group of youth who specialized in nonserious misbehavior to prevent them from becoming involved in delinquency and more serious offending. Proponents of this view, therefore, subscribed to the belief that youthful offending was (a) characterized by some degree of specialization and (b) escalated to more serious forms of delinquency in the absence of intervention.

By the mid-1970s, however, the welfare model-based juvenile justice laws and assumptions about specialization in youthful offending were challenged theoretically, empirically, and politically. One reform theme was that diversion of youth outside of the formal juvenile justice system was necessary to avoid the hypothesized delinquent effects of labeling. In other words, the common view was that minor versatile offending such as status offenses and vandalism escalated into more diverse and serious offending types when juvenile justice systems labeled minor offending youth as delinquents. In effect, specialization was conceptualized more in qualitative terms such as status offenders versus serious criminal, property, and violent offenders. However, beyond the theoretical debates, research on specialization in juvenile offending in the 1970s and 1980s did not support the assertion that, in the absence of intervention, the majority of youth who specialized in status offenses escalated to more serious offending types.

The reform debates in several jurisdictions, especially Canada and the United States, focused on specialization themes involving the identification of certain types of offender categories including career criminals, chronic offenders, serious and violent offenders, and sexual offenders. In effect, the political and research debates did not just center on more contemporary themes and measures of specialization but rather on more broad conceptualizations focused on distinguishing types of minor offenders, types of property offenders, types of violent offenders, and combinations of offender types. Youth justice laws were and are still based on this conceptualization of juvenile delinquent or young offender specialization. In many jurisdictions, different sets of procedural processes are utilized depending on the above specialized categories as well as versatile types. For example, as stated above, minor property offenders and minor violent offenders in certain jurisdictions in Australia, New Zealand, Canada, the United Kingdom, and the United States are completely diverted away from juvenile justice court proceedings. In contrast, in Canada, the United States, and most other jurisdictions, the most serious violent juvenile offenders, such as murderers, are either processed in adult criminal courts or subject to adult length custodial sentences. In addition, patterns of offending, especially predominately minor violent offenses, automatically subject such offenders to more punitive sentences. The key assumption informing many of these sentencing options is that the minor violent offending specialization pattern typically indicates the potential for such youthful offenders to escalate into specialized violent and chronic offending.

The mixed results regarding the existence of specialization in youthful offending belie the above assumptions that there is a strong correlation between chronic minor juvenile offending and long-term chronic violent offending. While there is more support for the relationship between serious chronic violent juvenile offending and long-term violent offending trajectories, only an extremely small proportion of all juvenile offenders have been identified within this specialized type. The concern is that juvenile justice laws have become over focused on this rare offender category to the detriment of programming for the overwhelming number of less serious juvenile offenders. In other words, juvenile justice reforms have been influenced by the debate in research concerning specialization, arguably, in a progressive way in some countries (e.g., Australia, New Zealand, Canada, the United Kingdom) and in a regressive manner in several US states, for example, those with life sentences with no parole and formally capital punishment for the most serious juvenile offenders. Even more recently in Canada, a conservative federal government introduced punitive reforms to the national youth justice law based on the assertion that specialized violent and specialized property offenders require more punitive sentences.

Conclusion

There is a consensus that versatility typifies most juvenile delinquents or young offenders. There also is considerable evidence that for a small proportion of juvenile offenders, certain types of offenses distinguish them from other offense patterns. For example, sexual offenses, while not constituting the largest proportion of a young offender's offending profile, nonetheless, distinguishes such an offender from another young offender who has no sexual offenses in their offending profile. In effect, the conceptualization of offender specialization remains open to different operational definitions. The key to the different conceptualizations is the theory or policy rationales. Major theoretical insights, for example, have emerged from developmental theories of offending that have asserted certain types of specialization as discussed above. The related theoretical debates and empirical research has propelled further theorizing and research that has several benefits. First, multiple measures and indices involving increased methodological sophistication have emerged and enhanced the understanding of juvenile offending trajectories. Second, the specialization research has revealed far more complex and diverse patterns of offending than were originally hypothesized. Third, it is vitally important that sophisticated research designs continue to be employed in assessing specialization because the ongoing juvenile justice policy debates concerning how to react to different types of offenders are best informed by valid and empirically supported categories of offending as opposed to impressions and emotion regarding what constitutes serious youthful offending.

Related Entries

- ▶ Crime Specialization, Progression, and Sequencing
- ▶ Juvenile Violence
- ▶ Measuring Crime Specializations and Concentrations
- ▶ Offense Specialization: Key Theories and Methods

Recommended Reading and References

Blumstein A, Cohen J, Roth JA, Visher CA (1986) Criminal careers and 'career criminals', vol 1. National Academy Press, Washington, DC

Britt CL (1996) The measurement of specialization and escalation in the criminal career: an alternative modeling strategy. J Quant Criminol 12:193–222

Bursik RJ (1980) The dynamics of specialization in juvenile offences. Soc Forces 58:851–864

Cloward R, Ohlin L (1960) Delinquency and opportunity: a theory of delinquent gangs. Free Press, New York

Colvin M, Pauly J (1983) A critique of criminology: toward an integrated structural-Marxist theory of delinquency production. Am J Sociol 89:513–551

Elliott DS, Ageton SS (1980) Reconsidering race and class differences in self-reported and official estimates of delinquency. Am Sociol Rev 45:95–110

Farrington DP, Snyder HN, Finnegan TA (1988) Specialization in juvenile court careers. Criminology 26:461–485

Fisher G, Ross S (2006) Beggarman or thief: methodological issues in offender specialization research. Aust N Z J Criminol 39:151–170

Francis B, Soothill K, Fligelstone R (2004) Identifying patterns and pathways of offending behaviour: a new approach to typologies of crime. Eur J Criminol 1:47–87

Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Hirschi T (1969) Causes of delinquency. University of California Press, Berkley

Kempf K (1986) Offense specialization: does it exist? In: Cornish D, Clarke R (eds) The reasoning criminal: rational choice perspectives on offending. Springer, New York

Klein MW (1979) Deinstitutionalization and diversion of juvenile offenders: a litany of impediments. In: Crime and justice, vol 1. University of Chicago Press, Chicago, pp 145–201

Klein MW (1984) Offense specialization and versatility among juveniles. Brit J Criminol 24:185–194

Lattimore PK, Visher CA, Linster RL (1994) Specialisation in juvenile careers: Markov results for a California cohort. J Quant Criminol 10:291–316

Le Blanc M, Loeber R (1998) Developmental criminology updated. In: Crime and justice, vol 23. University of Chicago Press, Chicago, pp 115–198

LeBlanc M, Frechette M (1989) Male criminal activity from childhood through youth. Springer, New York

Loeber R (1990) Developmental and risk factors of juvenile antisocial behavior and delinquency. Clin Psychol Rev 10:1–41

Mazerolle P, Brame R, Paternoster R, Piquero A, Dean CW (2000) Onset age, persistence, and offending versatility: comparisons across gender. Criminology 38:1143–1172

Moffitt TE (1993) Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. Psychol Rev 100:399–420

Nieuwbeerta P, Blokland AAJ, Piquero AR, Sweeten G (2011) A life-course analysis of offence specialization across age: introducing a new method for studying individual specialization over the life course. Crime Delinq 57:3–28

Piquero AR, Paternoster R, Mazerolle P, Brame R, Dean CW (1999) Onset age and offence specialization. J Res Crime Delinq 36:275–299

Smith DR, Smith WR, Noma E (1984) Delinquent career lines: a conceptual link between theory and juvenile offences. Sociol Quart 25:155–172

Sullivan CJ, McGloin JM, Ray JV, Caudy MS (2009) Detecting specialization in offending: comparing analytic approaches. J Quant Criminol 5:419–441

Wikstrom P-OH (1987) Patterns of crime in a birth cohort: age, sex, and class differences (Project metropolitan: a longitudinal study of a Stockholm cohort, no. 24). University of Stockholm, Department of Sociology, Stockholm

Wolfgang ME, Figlio RM, Sellin T (1972) Delinquency in a birth cohort. University of Chicago Press, Chicago

Specialty Courts

▶ Problem-Solving Courts

Specific Deterrence

▶ Rational Choice, Deterrence, and Crime: Sociological Contributions

Stalking

Troy McEwan[1] and Michele Pathé[2]
[1]Centre for Forensic Behavioural Science, Monash University and the Victorian Institute of Forensic Mental Health, Melbourne, Australia
[2]Queensland Forensic Mental Health Service and Griffith University, Brisbane, Australia

Synonyms

Harassment, Threat assessment

Overview

Stalking affects hundreds of thousands of people around the world every year, and criminal justice and mental health professionals are often required to manage or treat stalkers or their victims. This entry provides an overview of the phenomenon of stalking, its legal status, explanatory models, and victimology and impact. A detailed description of best-practice assessment procedures for perpetrators and victims and treatment/management considerations for both groups is also provided.

Fundamentals

What Is Stalking?

Stalking is a problem behavior characterized by repeated and unwanted intrusions inflicted by one person upon another in a manner that causes reasonable fear and/or distress. Intrusions typically come in the form of unwanted communications, such as those via telephone (including SMS), electronic media (e-mail, social networking websites), letters, or graffiti. In many stalking episodes, intrusions will also take the form of unwanted contacts, including loitering near the victim, following, maintaining surveillance, and making approaches. In addition to these core stalking behaviors, a range of other forms of harassment are often present, including property damage, theft, malicious use of the Internet (e.g., creating defamatory websites), sending or leaving unsolicited materials (ranging from gifts to items intended to frighten or implicitly threaten, such as dead animals), inveigling others to harass the victim, vexatious complaints, uttered threats, and assault.

The defining features of stalking behavior are that it is *repeated*, is *unwanted*, and *causes significant distress*. Even within this broad definition, stalking has been shown to occur in two basic patterns (Purcell et al. 2004). The first is a short-lived burst of self-limiting intrusions that occur over the course of a day or two and usually involve unwanted approaches by a stranger or distant acquaintance. The second involves both unwanted communications and contacts, and frequently other associated behaviors, tends to be perpetrated by an ex-intimate or acquaintance, and persists for weeks, if not months. The latter type tends to have a far greater psychological and social impact on the victim and is more frequently associated with threats and physical assault, so is more likely to attract criminal justice or mental health attention. In recent years, stalking behaviors have also been observed in a purely virtual environment, a phenomenon labelled "cyberstalking." Those who engage in cyberstalking use the mechanisms of the Internet to harass their target(s), but do not engage in the aforementioned behaviors commonly seen in "off-line" stalking. Studies of this behavior are few and bedevilled by problems with definition and methodology. The best investigation to date concluded that, while Internet-based harassment is common, pure cyberstalking is actually rare and most victims who experience stalking via the Internet are also targets of off-line stalking behavior (Sheridan and Grant 2007).

Legal Definitions

Stalking is a relatively new crime in most jurisdictions; the Californian law of 1990 was the first legislation specifically intended to prevent the behavior. By the end of the 1990s, stalking laws had been introduced in most English-language jurisdictions, with many countries in mainland Europe following suit throughout the first decade of the 2000s. The first wave of anti-stalking laws (1990–1999) usually conceptualized stalking as a precursor to assault, and so the purpose of the law was to prevent physical attack. More recently, stalking has been conceptualized as harmful in and of itself, and the introduction of the second wave of laws (those implemented since 2000) has tended to focus less on violence, although this is still a concern. Anti-stalking legislation is relatively uncommon outside of Australasia, Europe, and North America. Where present, it is usually included in domestic violence laws and so is typically poorly defined and largely restricted to female victims of ex-intimate partners.

Depending on jurisdiction, anti-stalking legislation typically requires three elements to prove the offence. The behavior or conduct element is usually defined as a "course of conduct," or conduct engaged in on more than one occasion (intentionally broad so as to allow for early intervention). There is no requirement that the individual instances of behavior be unlawful (e.g., repeated nonthreatening telephone calls would satisfy a course of conduct, although none would be illegal in isolation). Some jurisdictions leave the behavior element at this point without further defining the nature of the acts that constitute harassment or stalking, while others go on to define the types of conduct that are prohibited. The second necessary element is perpetrator intent. Most anti-stalking laws require either that the perpetrator intended to harass or cause harm to the victim or that they ought to have known that their behavior could have had that effect (akin to being reckless as to the effect of their behavior). The final element, which is present in many North American and European jurisdictions, although not in Australasia, is that the victim experiences the behavior as intimidating and feels fear or apprehension as a result. Two levels of proof are often required: a subjective standard (this victim felt fear) and an objective standard (a reasonable person would have felt fear). In many jurisdictions, stalking is divided into misdemeanor or felony offences, depending on the nature of the behavior engaged in. Furthermore, many anti-stalking laws provide for the civil option of a protection order for the victim, breaches of which are a criminal offence. Stalking laws have provoked considerable discussion, which is summarized by Kapley and Cooke (2007) and Mullen and colleagues (2009a).

Prevalence

Estimates of the prevalence of stalking vary according to the definition, sampling, method of enquiry, and willingness of participants to respond and answer candidly. Nonetheless, estimates are remarkably consistent across location and time, with lifetime victimization rates of 17–30 % reported for women and 4–12 % for men (Purcell et al. 2002; Smith et al. 2011; Stieger et al. 2008). Reported prevalence rates in the USA tend to be lower than other countries (7 % for women and 2 % for men) as the national survey included only those victims who experienced the stalking as dangerous or life-threatening (Basile et al. 2006). Victims are largely female (70–80 %) and stalkers are predominantly male (80–85 %), with 20–25 % involving same-gender stalking, usually male-male.

Explaining Stalking Behavior

Stalking, like any complex human behavior, is a product of a range of psychological, social, and cultural influences. People stalk out of anger, hope, lust, ignorance, or a combination of these factors. The influence of psychopathology, discussed later, is also relevant in many cases of stalking, either as a direct or indirect causal factor. To date there are no comprehensive theories to explain the variety of stalking behavior that is observed. Some single factor theories have been described, most commonly disrupted attachment (Meloy 2007). This theory posits that stalking is a consequence of insecure and, specifically, preoccupied attachment. Research has supported aspects of this theory, finding that stalkers do indeed often present with insecure adult attachment, although there is little evidence for a specifically preoccupied style. Proponents of stalking as an expression of disrupted attachment tend to focus on those seeking relationships and do not address other possible trajectories into stalking behavior. Moreover, the mechanisms by which insecure attachment produces stalking in some contexts but not in others, even within the same individual, have not been adequately explained. Nonetheless, there does seem to be some evidence for chronically disturbed attachment style in some stalkers.

Where attachment theory examines distal dispositional characteristics that may lead to increased proclivity to stalk, Spitzberg and Cupach’s (2007) application of relational goal pursuit theory to relationship-based stalking suggests more proximally relevant factors. In this theory, the goal of possessing a particular relationship becomes entwined with the perpetrator’s value system and is integral to their sense of self-worth and life happiness. In this context, pursuit of the goal (a relationship with the victim) takes on disproportionate significance. The failure to achieve the goal in spite of persistent attempts results in rumination and consequent negative emotional states, which the perpetrator attempts to escape by further efforts to achieve the desired goal. This is perhaps the most well-explicated theory of stalking behavior, although it also does not account for the variety of trajectories seen in stalking behavior, and to date it remains largely untested. A significant gap in the stalking literature remains, awaiting a truly comprehensive multifactor theory that adequately explains this complex behavior.

Classification Systems

In the absence of such a theory, many classification systems have been developed in an attempt to reduce the heterogeneity of stalkers into manageable categories that can assist decision-making. These taxonomies vary depending on the needs of the authors and audience. Those developed by clinicians emphasize the goal of treatment and management, while those for law enforcement tend to be used more as shorthand for management and referral options. In both realms, classification schemes tend to focus on three variables: mental disorder, prior relationship, and motivation. The earliest classification schemes focussed heavily on psychopathology, reflecting the fact that stalking behavior was seen as a manifestation of emotional and psychological disturbance. By the mid-1990s, prior relationship and affective state were also considered useful ways of differentiating stalkers. More recently, Mohandie and colleagues (2006) proposed a scheme based on the nature of the prior relationship and the context in which the stalking occurred. This scheme in particular may be useful for law enforcement personnel as it is relatively easy to apply based on minimal information. Many typologies have been proposed, and a useful review is provided by Pinals (2007).

A Multiaxial Classification Scheme

One of the most commonly used typologies, endorsed by the Group for the Advancement of Psychiatry Committee on Psychiatry and the Law, is that of Mullen et al. (2009). This typology, based originally on a sample of 145 Australian stalkers, took the unique step of proposing classification on multiple axes to determine type. The first axis is the nature of the prior relationship between stalker and victim; the second, the initial motivation for and context in which contact with the victim occurred; and the third, the presence and nature of any psychiatric diagnoses. Using these three axes, five stalker types were identified, which are described below. More lengthy descriptions can be found in Mullen et al. (2009).

The Rejected: These stalkers begin to stalk following the dissolution of an intimate relationship, however brief, in an attempt either to reclaim the relationship or to exact revenge against their former partner for leaving. In many cases, both of these motives are present at different times. Not typically affected by severe mental illness, these stalkers tend to present with antisocial, narcissistic, borderline, or paranoid personality traits, depression, and/or substance use disorders. In a minority of cases, dependent and obsessive-compulsive personality traits are evident.

The Resentful: These stalkers target an acquaintance or stranger, but invariably someone who they perceive has mistreated them, either in their own right or because they are representative of some organization that has provoked the stalker's ire. The stalking usually begins with the aim of righting the perceived wrong and over time becomes the only way to assuage feelings of resentment and regain their lost sense of power and control. These stalkers often suffer from paranoid or narcissistic personality disorders and occasionally present with paranoid delusions incorporating the stalking victim(s).

The Incompetent Suitor: These individuals target strangers or acquaintances with the initial intention to pursue a friendship or date. The stalking is an inept and often self-centered attempt to achieve that outcome. Over time it may take on an angry or aggressive tone as the stalker's wishes are thwarted, but this group typically does not stalk for an extended period, instead transferring their interest elsewhere. While many stalkers who fall into this category do not experience any form of mental disorder, this behavior does emerge in the context of impaired social skills and interpersonal deficits, and intellectual disability or developmental disorders are not infrequent. Also found in this group are narcissistic individuals who are assertive, overbearing, and simply can't imagine why any woman would refuse the opportunity to date them.

The Intimacy Seeker: These stalkers target acquaintances or strangers in an expression of their love and seeking love in return. Usually the stalker persists in an affectionate and amorous way, but very occasionally they can become angry and vengeful if they realize that the longed-for relationship is under threat. Often driven by a severe mental illness, these individuals may believe that they already have a mutually loving relationship with the victim, even in the absence of any true connection. Conversely, they may be intensely infatuated and believe that the idealized relationship is possible if they persist. Overvalued ideas of intimacy are almost ubiquitous in this group, arising either out of a severely disturbed borderline personality or at the more extreme end, a delusional disorder or schizophrenia with erotomanic delusions.

The Predatory: These men most often target strangers with the goal of achieving some form of sexual gratification, usually deviant. They use stalking as part of a wider repertoire of sexually inappropriate behaviors, and their stalking is usually characterized by surreptitious loitering, following, spying, and, in many cases, sexual violence. The predominant diagnoses in this group are paraphilias, depression, and substance misuse.

Psychopathology

Stalking is a set of behaviors, not a diagnosis, but studies suggest that at least 80 % of community referred stalkers have some form of psychiatric condition (McEwan et al. 2009). Personality disorders, schizophrenia, and major mood disorders are the primary presenting problems (McEwan et al. 2009; Rosenfeld 2003). In their retrospective file review of a forensically derived sample of stalkers, Mohandie and colleagues (2006) found 46 % had a probable diagnosis of mental disorder. Referral bias limits the generalizability of these findings, in that forensic samples are weighted towards more serious offending and severe mental illness.

There is a higher prevalence of major mental illness among those who stalk public figures (James et al. 2009). Over 80 % of individuals who made inappropriate communications or approaches to the British Royal Family were found to have a major mental disorder. In a study of individuals who pursued and ultimately physically attacked a Western European politician, over half were mentally disordered, most of whom were psychotic.

Suicidal ideation is also common among stalkers. Using retrospective file review, Mohandie and colleagues (2006) identified suicidality in 25 % of 763 North American stalkers, while McEwan and colleagues (2009) observed that one in ten of their Australian sample reported suicidal ideation. In a 3-year follow-up, McEwan and colleagues (2010) found that 3 of 138 stalkers (2.2 %) had committed suicide. Though small, these numbers represent substantially higher rates, indeed six times greater, than would be expected among psychiatric patients or community-based offenders. In all three cases, the stalker had a history of suicidal ideation or intent. This finding is of particular concern because suicidal stalkers pose an enhanced risk of violence towards their victim and third parties. Stalking risks will be discussed further in a later section.

Particular psychopathology is found more commonly in certain types of stalker, as noted in the multiaxial classification scheme presented earlier. For instance, personality disorders or traits predominate in Rejected stalkers though morbid jealousy of delusional intensity is also encountered in this group. Intimacy seekers are far more likely to have serious psychopathology, in the form of psychotic disorders, and female stalkers have higher rates of psychosis than their male counterparts (Meloy and Boyd 2003). As elaborated later, the risk of serious violence in the stalking situation is greater when the stalker has a nonpsychotic as opposed to a psychotic condition.

Earlier assumptions that stalkers have above average intellectual functioning have been called into question by a prospective study of stalkers referred to a clinic specializing in their assessment and treatment (MacKenzie et al. 2010). Intelligence testing found a mean intelligence quotient (IQ) in the average range (91.59 [SD 16.2]), with a verbal IQ significantly lower than performance IQ. These findings have important implications for psychological intervention programs for stalkers, which must reflect their cognitive abilities.

Victims and the Impact of Stalking

Stalking is an event that has a perpetrator and a victim. The victim is pivotal to stalking because stalking is, in effect, a victim-defined crime. Behaviors that are inappropriate and gauche are transformed into behaviors which are harmful and illegal by the apprehension that is evoked in the victim. While the line between the inappropriate and the fear inducing is in some respects influenced by the tolerance and sensitivities of the victim, most stalking that comes to the attention of the criminal justice system could be expected to provoke fear or disquiet in the majority of people.

Stalking has been viewed as a product of failed relationships and a form of domestic violence perpetrated by men against women. Earlier in the evolution of our understanding of stalking, the concept generalized from behavior directed at the famous to that of males who battered their partners. While lending attention to a long neglected issue, the concept of stalking is problematic when used to describe behaviors that occur in the context of an ongoing relationship. To begin with, the focus on stalking as a form of domestic violence has predominantly been on women as victims and men as abusers, while those who stalk former intimates can be female and victims can also be male. More importantly, while behaviors akin to stalking undoubtedly emerge among cohabiting partners, especially in relationships characterized by jealousy, this usually comprises behaviors that are aimed at information gathering (e.g., monitoring, surveillance) or coercive physical or verbal abuse and threats. A distinction should be made between such conduct and stalking. The aim of stalking is to make one’s presence felt *where it would not otherwise* *exist*. Where there is an ongoing relationship, however conflicted and potentially damaging, the behavior should be classified as a form of terrorizing and controlling intimate partner violence, not as stalking. Confusing the two is semantically and practically problematic as approaches to assessing and managing stalkers will be ineffective in situations where the stalker and victim continue to cohabitate.

A number of classifications have been proposed for stalking victims, all of them based upon the preexisting relationship, if any, between victim and stalker. The following is a relational classification of stalking victims, derived from the typology proposed by Mullen and colleagues (2009):

1. *Former sexual intimate*. The most common victim profile is a woman who previously shared an intimate relationship with the stalker, usually male, though women can be perpetrators and stalking can also arise in the context of same-gender relationships. This category includes only those cases in which the relationship has been explicitly terminated.

   Acquaintances, who can be:

2. *Family/friend*. Rejected stalking may follow the breakdown of a close friendship or estrangement from a family member. Occasionally, Resentful stalking can emerge in relation to some perceived injustice (e.g., dispute over a family estate).
3. *Professional relationship*. As below, those in the health, law enforcement, and teaching professions are especially at risk.
4. *Casual contact*. Intimacy seekers, Incompetent suitors, and Predatory stalkers may all target their victim in the context of a casual social encounter. Neighbor stalking usually falls into this category, most often involving a Resentful stalker incited by disputes over noise or a fence line.
5. *Workplace contact*. Where the victim meets their stalker in a workplace context. The stalker may be in a subordinate or superior position, or in some cases a client or customer.

   Those without any real connection to the victim:

6. *Stranger*. Some victims have no prior knowledge of their stalker. The stranger may identify themselves, but some remain anonymous.
7. *The famous*. As noted below, prominent figures such as politicians and media personalities attract a range of stalker types, particularly Intimacy seekers and Resentful stalkers. Some celebrities suffer the attentions of multiple stalkers.

To this can be added secondary victims, who are impacted by stalking as a consequence of their relationship to the primary victim (e.g., the victim's partner or other family members, co-workers, or neighbors). Stalking tends to have a "ripple effect" on the victim's social network. In some instances, stalkers pose a greater risk to secondary victims than their primary target.

It has long been observed that stalking can have damaging, if not devastating, consequences and it is the reason, ultimately, that stalkers have attracted the interest of mental health professionals. Stalking impacts upon the victim, the third parties, the stalker themselves, and the community as a whole. Such impacts have received systematic attention over the past decade. The initial studies were mainly descriptive and highlighted the psychological, occupational, and social impacts of these behaviors (Pathé and Mullen 1997). Subsequent large epidemiological studies (e.g., Purcell et al. 2002; Purcell et al. 2005) found that stalking was prevalent in Western nations and also confirmed the higher rates of psychiatric morbidity among victims of stalking.

Victims of stalking commonly report a profound sense of violation and loss of control over their lives. Many experience a pervasive mistrust of others that can persist long after the stalking has ceased. In their epidemiological study that compared matched controls, victims of brief harassment, and victims of protracted stalking (>2 weeks), Purcell and colleagues (2002, 2005) found that the rates of general psychiatric morbidity were significantly higher among those who reported a history of protracted stalking (36 % vs 19 % for controls and 22 % for brief harassment). Those subjected to severe stalking were also more likely to report recent suicidal ideation (10 %) and clinically significant levels of posttraumatic symptomatology (three times more likely than the briefly harassed group).

More recent studies of stalking victims have considered groups who are particularly vulnerable to stalkers. These include health professionals (Whyte et al. 2011) and public figures (Mullen et al. 2009), but there is also anecdotal evidence of higher rates of stalking victimization in other professions such as teaching and law enforcement. Studies to date suggest that the lifetime risk for a health professional being stalked by a client or patient is 20 % (Purcell et al. 2005), and those working in the mental health field, by virtue of the patients they encounter, may be more susceptible to such behaviors. One of the commonest motivations for stalking health practitioners is morbid infatuation, where a lonely, disordered patient misconstrues sympathy and attention as romantic love. These patients-turned-stalkers can be categorized as Intimacy seekers. Resentment is another common motivation, stemming from some supposed injury or dereliction. These stalkers fall into the Resentful category. A smaller group is pursuing reconciliation or revenge following the termination of a (usually long-term) therapeutic relationship (Rejected stalkers). The impact of stalking on health professionals can be significant, with over 70 % of affected psychologists in one study (Purcell et al. 2005) modifying aspects of their personal and professional lives, some of whom relocated their practice and/or their private address. Almost a third contemplated leaving their profession altogether, this figure rising to nearly 50 % among those whose stalker made malicious complaints to their professional registration board.

Assessment Guidelines

Assessments of stalkers or victims often take place in a stressful context. For the stalker, this may be as part of a presentence or parole report, or in a mandated treatment situation. For victims, help is often sought when the pressures of the stalking become overwhelming and they feel that they can no longer cope on their own. Stalkers tend to deny, minimize, rationalize, and justify their behavior. Victims often minimize the experience of stalking and overemphasize their responsibility for the harassment. In both scenarios, obtaining collateral information about the stalking episode is essential. Where available, police or court documents summarizing the behavior are invariably useful, as are accounts by those associated with the victim or stalker. When possible, it is usually best to avoid having contact with both the victim and stalker. Victims may experience attempts to obtain collateral information as simply another unwanted intrusion by someone acting on behalf of the stalker, while for stalkers, having a professional make contact with the victim can be understood as implicit support for an ongoing “relationship” or as an opportunity to obtain new information about the victim.

Assessing a Stalker

It is usually possible to form hypotheses about the motivation for stalking reasonably quickly. Information about the prior relationship between victim and stalker and about the apparent initial motivation can be gleaned from collateral sources or identified very early in an assessment. Often the best way of determining motivation is to simply ask the stalker what drove them to this behavior or what they originally wanted to achieve. A central aspect of the assessment is conducting a functional analysis of the stalking. This involves examining the common situational, emotional, and behavioral antecedents to and consequences of each intrusion. An analysis of this kind can provide useful information upon which to base hypotheses about the cognitive and emotional states associated with offending and the skill deficits that might lead them to stalk rather than use a more adaptive strategy to achieve their goals.

For mental health clinicians, particularly those with forensic expertise, the assessment will be like most others, with slightly more emphasis on the individual’s experience of other, similar interpersonal situations. For example, for a Rejected stalker, additional questions about prior relationships and breakups would be relevant, and for a Resentful stalker, enquiry into how they have handled previous injustices or disputes would be useful. Where sexually deviant behavior is present, a full sexual behavior assessment is appropriate, as for sex offenders. If the clinician is not trained in assessment and diagnosis of psychopathology, referral to an appropriately qualified expert is strongly recommended.

Assessing a Victim

Some stalking victims present for mental health intervention after the stalking has ceased. They are commonly seeking help to deal with post-trauma symptoms and reconstitute their life. An appraisal of the risk of recurrence of the stalking may also be appropriate. Many more stalking victims seek assistance when the stalking is ongoing, and the priority in these instances is to ensure all reasonable steps are taken to protect the victim's safety and privacy. Some may be engaging in behaviors which place them and other parties at risk (e.g., retaliating against their stalker), necessitating urgent guidance. Around 10 % will have suicidal ideation requiring assertive intervention.

As a consequence of their experiences, many stalking victims are distrustful, and some will have previously encountered negative attitudes from the "helping" professions. In engaging stalking victims, it is essential that the assessment is conducted by a professional familiar with stalking issues and the available interventions. It is important at the outset that stalking victims are reassured that the stalking is not their fault and that they are not alone in their ordeal.

Stalking victims require a comprehensive assessment, including a full psychosocial history that encompasses past victimization and a detailed account of the stalking. This includes the nature of the victim's prior relationship with the stalker, a description of the onset of the stalking, and the frequency and nature of the intrusions, as well as any strategies employed to end the stalking. How effective were these measures, and has anything exacerbated the behavior? Have they involved the police, and if not, why not? Are there any current protective injunctions, or is the stalker subject to any other legal directives? Is the stalker known or suspected to be mentally ill, and if so, are they known to the mental health system? Some victims, especially those stalked by strangers, may have limited information, but victims should be encouraged to bring as much documentation as possible to the assessment (including any evidence of the stalking in the form of letters, e-mail, phone records, photographs, court orders, or witness statements).

Stalking victims should be questioned about suicidal inclinations, any homicidal intent towards the stalker, and any measures they may have taken to defend themselves (e.g., keeping weapons in the house). One should endeavor to gauge the victim's current level of support, others affected by the stalking, the victim's employment situation, and mental health status. Are they suffering current anxiety or depressive symptoms, and how have these been managed? Maladaptive coping mechanisms such as substance abuse or gambling are not uncommon and may require specific interventions.

Risk Assessment

In any stalking situation, a primary concern for the victim is the likelihood of further damaging behavior. Often their worries focus on the likelihood of physical violence, the likelihood that the stalker will persist, or if they have stopped, that they will start again. Therefore, for those charged with managing stalking situations, risk assessment is a primary focus. For those working with stalkers, an additional concern is the potential for psychological and social harm to the perpetrator and what this means for other areas of risk.

While the stalking risk assessment literature continues to lack strong retrospective or prospective studies of representative samples, a sizeable body of published work now exists, and relevant information can also be gleaned from the much larger body of research into violence and offending risk assessment more generally. Two structured professional judgement tools for assessing stalking have been published and are undergoing validation trials – the *Stalking Risk Profile* (*SRP*) (MacKenzie et al. 2009) and the *Guidelines for Stalking Assessment and Management* (*SAM*) (Kropp et al. 2008). Where these tools are not available, clinicians and law enforcement personnel must rely on clinical experience and integrating the literature themselves.

Risk of Persistence or Recurrence Stalkers that persist for at least 2 weeks are at increased risk of continuing for months. Those with a prior relationship with the victim tend to stalk for longer, and as a rule, women tend to be more persistent than men. By far the strongest predictor of extreme persistence is psychosis, and these stalkers are unlikely to desist without mental health intervention. Recurrent stalking appears to be the domain of the personality disordered and those who share the victim's children, property, or locale, although a lack of research makes firm conclusions difficult (Rosenfeld 2003).

Risk of Violence As a group, ex-intimate partners present the greatest risk of violence, with up to half in this category using physical violence against their victim or a related third party (McEwan et al. 2007). Identified risk factors for violence for all stalkers include:

- *Demographic variables*. It has been repeatedly shown that the gender of the stalker or the victim has no impact on the prevalence of either threats or assault, and same-gender stalking victims are at equal risk of violence (Strand and McEwan 2011).
- *Stalking behaviors*. Stalkers who follow, loiter near, accost the victim, or enter the victim's home are at increased risk of violence. Escalations in the intrusiveness or intensity of the stalking have not been subject to research, but should be treated as periods of increased risk.
- *Threats*. Threats are common in stalking situations and have been shown to be a risk factor for violence, particularly for ex-intimate partner violence (McEwan et al. 2007).
- *Prior violence and criminal behavior*. The evidence is inconsistent; however, on balance, it seems that prior violence in particular is a predictor of stalking-related violence (McEwan et al. 2007).
- *Mental disorder*. Psychotic stalkers have consistently been found to be less violent than nonpsychotic stalkers; however, the presence of other risk factors should not be discounted because the stalker is psychotic (McEwan et al. 2009). The relationship between personality disorder and violence in stalkers is unclear, although findings from the wider risk assessment literature indicate that antisocial personality disorder in particular is related to violence.
- *Substance misuse*. Substance misuse is associated with stalking violence, likely having a compounding effect on other risk factors.

Formulation and Treatment Planning for Stalkers

Clinical formulation links assessment information with a systemic treatment and management plan. A good formulation will explain the mechanisms underlying the stalking and direct interventions to stop it. For the clinician assessing a stalker, the central focus of formulation is the *stalking behavior* itself rather than the presence of psychopathology or other matters. A detailed explanation of how to formulate stalking behavior is beyond the scope of this entry, but some guidelines are available in McEwan et al. (2011).

Treatment and Management

Stalkers

Management of stalkers often involves the use of the criminal justice system, which is described further below. Presented here are options for interventions designed to address the underlying causes of the behavior in conjunction with such environmental management strategies. The information provided in this section assumes the availability of an appropriately trained and supported mental health practitioner and a client that is willing and/or mandated to attend treatment. Whenever possible, treatment of stalkers should be undertaken by a clinician experienced in working with offenders, who has training in structured risk assessment, familiarity with the stalking research literature, and an understanding of local stalking laws. Those who work with stalkers (or their victims) should also be cognizant of the risks of such work and should take appropriate steps to ensure their own safety and security (Pathé 2002).

There is limited information available about the efficacy of offence-specific psychological interventions for stalkers, with Rosenfeld and colleagues' (2007) trial of dialectical behavior therapy the only published study. Treatment strategies described by these and other authors (e.g., MacKenzie and James 2011) take a cognitive-behavioral perspective and focus on helping the stalker to develop skills that they can apply in situations where they have previously used stalking (e.g., emotional regulation, problem solving, and communication skills are all targets of psychological intervention). Which particular skill deficits are most relevant to an individual stalker, and the prioritization of one skill over another, should be evident following a comprehensive formulation as outlined above. In addition to skill development, another central aspect of treatment is helping the stalker recognize potential triggers for stalking behavior, identify high-risk thinking and behavioral patterns, and implement alternative strategies.

As with any offender treatment, attention must also be paid to the client's readiness to desist from their behavior. Treatment readiness can be defined as having appropriate motivation, capacity, and ability to engage in treatment so as to maximize the likelihood of therapeutic change. Where motivation to engage is low, time should be spent helping the client identify personal goals that would be assisted by avoiding stalking in the future, encouraging the view that they are able to change their behavior, and reducing emotional distress that might be provoked by thinking or talking about their offence.

Maximizing treatment readiness may require other issues to be dealt with prior to beginning any offence-specific interventions. Perhaps the most common factors that interfere with a stalker's capacity and ability to engage in offence-related treatment are active symptoms of mental illness and substance misuse. As discussed previously, these may also be risk factors for further stalking behavior, and so appropriate treatment can both reduce risk and increase the client's opportunity to engage in other offence-related interventions. Mullen et al. (2009) provide a detailed description of the management of mental disorders in stalkers. In essence the appropriate treatments are no different to those offered to non-stalkers, although consideration of risk may lead to increased emphasis on the safety of the client and others and so greater impetus for involuntary treatment. Similarly, where substance misuse is present, interventions offered to individuals who do not stalk are usually appropriate, although with due consideration of the need to protect the stalking victim while treatment for substance misuse is undertaken.

Victims

Strategies to Combat Stalking These approaches need to take account of the individual circumstances, including the prior relationship between victim and stalker; the presence of shared children, property, or workplace; the methods of harassment; and the assessed risk to the victim and/or other parties. In most situations, the victim should be encouraged to inform relevant people of the stalking, so they are alert to the problem and better equipped to protect the victim and themselves. It is important to advise victims at the outset to avoid all contact with their stalker. Attempts to appease the stalker through "one last meeting," or to explain, yet again, why the relationship ended, are, in effect, perpetuating the "relationship" and reinforcing the stalker's efforts.

Using Anti-stalking Laws As stalking is a criminal offence in many jurisdictions, informing the police may not only offer a resolution but also open an official paper trail documenting the stalker's course of conduct. To maximize the likelihood of police action, victims should be encouraged to keep a clear chronological record of harassing intrusions and any tangible evidence of the stalking. While there are few surveys of police responses to stalking complaints, those that exist show that in the majority of cases, stalkers are charged with offences other than stalking (e.g., trespass), even where specific anti-stalking legislation exists. Anecdotal evidence suggests that police often recommend that the victim obtain a protective injunction, enabling them to prosecute a breach; however, this strategy may carry some risks (see below). Storey and Hart's (2011) investigation of 32 cases showed that legal intervention is one of the more effective tactics to stop stalking. This includes everything from a formal police warning to a conviction. Informal warnings from police were shown to be ineffective in this study.

Protective Injunctions Also termed restraining, non-molestation, apprehended violence, protection, no contact, or intervention orders, their use against stalkers remains a contentious issue. These orders should not be used indiscriminately with stalkers as they have the potential to inflame the situation, provide a false sense of security, and exacerbate the victim's distress. Protective injunctions are more likely to be effective if they are properly policed, and the perpetrator is an Incompetent suitor or a Rejected stalker who has less invested in the relationship (e.g., a brief relationship with no children) and with a nonviolent, law-abiding background. These orders are less likely to succeed with Intimacy Seekers, especially those with erotomanic delusions, and Rejected stalkers who have a strong investment in the relationship (long-standing, shared children and assets) and/or an antisocial personality disorder, morbid jealousy, or a history of domestic violence.

Privacy, Safety, and Security There are a range of measures that can enhance the victim's security. These are beyond the scope of this entry but are detailed in Pathé (2002). Victims and potential victims such as health practitioners should try to ensure that as far as possible all personal information is removed from the public domain. This includes delisting their telephone and fax numbers, obtaining a post office box address, and removing their name from the electoral roll and property titles. Stalking victims should be cautious in their use of social networking websites such as Facebook and Twitter and take appropriate steps to protect their privacy.

Clinical Management of Stalking Victims Cognitive-behavioral approaches have been recommended with stalking victims. As a consequence of their stalking experiences, a victim's previously held beliefs in their strength and resilience are replaced by feelings of extreme vulnerability, and their previous assumptions about the reasonable and predictable nature of the world are shattered. Cognitive-behavioral therapies aim to correct the unhelpful assumptions that underlie their anxiety and depression and the avoidance behaviors that frequently emerge in this context. Priority should always be given to the client's safety, and if the stalking is ongoing, some avoidance behaviors may be entirely appropriate. For stalking victims presenting with chronic or complex posttraumatic stress disorder, dialectical behavior therapy (DBT) may prove efficacious (Mullen et al. 2009b).

Although there is a dearth of randomized clinical trials for pharmacological agents in the treatment of stress-related syndromes, those most frequently used with victims of stalking, as an adjunct to psychological approaches, are the selective serotonin reuptake inhibitor (SSRI) antidepressants. These agents have demonstrated efficacy for stress-related symptoms, panic, and depression. Care should be exercised in prescribing benzodiazepines for anxiety, as the symptoms, and thus the need for treatment, may be protracted, increasing the risk of drug dependence. Furthermore, the benzodiazepine agents such as Valium may diminish alertness and self-control, which can exacerbate the victim's feelings of vulnerability.

Other Therapies Wherever possible, the victim's partner and other relevant family members and supports should be involved in their management and the development of a safety plan. They frequently share in the victim's distress and may have been more directly impacted than the primary victim. They often seek information to try to make sense of the stalking, and educating significant others can alleviate pressure on the victim, by stemming criticism and unhelpful advice. Some may require referral for separate counselling.

Group therapies and support organizations can provide some validation for stalking victims and assist in reducing the victim's sense of isolation and self-blame. They are a useful source of information, support, and advocacy. Stalking victim groups need to be mindful of security at the meeting venue and the confidentiality of their members.

Conclusions

Stalking is a complex and damaging behavior that has commanded the attention of behavioral scientists, mental health professionals, the legal system, and the general public. Stalkers vary according to their motivations and psychopathology, and this determines their level of risk in multiple domains, and their management. While theories of stalking are not well understood, this problem behavior is amenable to psychological and environmental interventions. The mental health and criminal justice systems have a key role to play in bringing stalking episodes and victim suffering to an end.

Related Entries

► Police Family Violence Services
► Risk Assessment, Classification, and Prediction

Recommended Reading and References

Basile KC, Swahn MH, Chen J, Saltzman LE (2006) Stalking in the United States: recent national prevalence estimates. Am J Prev Med 31:172–175

James DV, Mullen P, Pathé M, Meloy J, Preston L, Darnley B et al (2009) Stalkers and harassers of royalty: the role of mental illness and motivation. Psychol Med J Res Psychiatry Allied Sci 39:1479–1490

Kapley DJ, Cooke JR (2007) Trends in antistalking legislation. In: Pinals DA (ed) Stalking. Psychiatric perspectives and practical approaches. Oxford University Press, New York, pp 141–163

Kropp RP, Hart SD, Lyon DR (2008) Guidelines for stalking assessment and management (SAM). ProActive Resolutions Inc, Vancouver

MacKenzie RD, James DV (2011) Management and treatment of stalkers: problems, options, and solutions. Behav Sci Law 29:220–239

MacKenzie RD, McEwan TE, Pathé MT, James DV, Ogloff JRP, Mullen PE (2009) Stalking risk profile: guidelines for assessing and managing stalkers. StalkInc & Centre for Forensic Behavioural Science, Monash University, Melbourne

MacKenzie R, James D, McEwan T, Mullen P, Ogloff J (2010) Stalkers and intelligence: implications for treatment. J Forensic Psychiatry Psychol 21:852–872

McEwan TE, Mullen PE, Purcell R (2007) Identifying risk factors in stalking: a review of current research. Int J Law Psychiatry 30:1–9

McEwan TE, Mullen PE, MacKenzie R, Ogloff JRP (2009) Violence in stalking situations. Psychol Med J Res Psychiatry Allied Sci 39:1469–1478

McEwan TE, Mullen PE, MacKenzie R (2010) Suicide amongst stalkers. J Forensic Psychiatry Psychol 21:514–520

McEwan TE, Pathé M, Ogloff JRP (2011) Advances in stalking risk assessment. Behav Sci Law 29:180–201

Meloy JR (2007) Stalking: the state of the science. Crim Behav Ment Heal 17:1–7 [Editorial]

Meloy JR, Boyd C (2003) Female stalkers and their victims. J Am Acad Psychiatry Law 31:211–219

Meloy S, Hoffman (2008) Stalking, threatening and attacking public figures. Oxford University Press, New York

Mohandie K, Meloy JR, McGowan MG, Williams J (2006) The RECON typology of stalking: Reliability and validity based on a large sample of North American stalkers. J Forensic Sci 51:147–155

Mullen PE, James DV, Meloy J, Pathé MT, Farnham FR, Preston L et al (2009a) The fixated and the pursuit of public figures. J Forensic Psychiatry Psychol 20:33–47

Mullen PE, Pathé M, Purcell R (2009b) Stalkers and their victims, 2nd edn. Cambridge University Press, New York

Pathé M (2002) Surviving stalking. Cambridge University Press, Cambridge

Pathé M, Mullen PE (1997) The impact of stalkers on their victims. Br J Psychiatry 170:12–17

Pinals DA (2007) Stalking: classification and typology. In: Pinals DA (ed) Stalking. Psychiatric perspectives and practical approaches. Oxford University Press, New York, pp 27–60

Purcell R, Pathé M, Mullen PE (2002) The prevalence and nature of stalking in the Australian community. Aust N Z J Psychiatry 36:114–120

Purcell R, Pathé M, Mullen PE (2004) Editorial: when do repeated intrusions become stalking? J Forensic Psychiatry Psychol 15:571–583 [Editorial]

Purcell R, Pathé M, Mullen PE (2005) Association between stalking victimisation and psychiatric morbidity in a random community sample. Br J Psychiatry 187:416–420

Rosenfeld B (2003) Recidivism in stalking and obsessional harassment. Law Hum Behav 27:251–265

Rosenfeld B, Galietta M, Ivanoff A, Garcia-Mansilla A, Martinez R, Fava J et al (2007) Dialectical behavior

therapy for the treatment of stalking offenders. Int J Forensic Ment Health 6:95–103
Sheridan LP, Grant T (2007) Is cyberstalking different? Psychol Crime Law 13:627–640
Smith KE, Coleman K, Eder S, Hall P (2011) Homicides, firearms offences and intimate violence 2009/10. Home Office, London
Spitzberg BH, Cupach WR (2007) The state of the art of stalking: taking stock of the emerging literature. Aggress Violent Behav 12:64–86
Stieger S, Burger C, Schild A (2008) Lifetime prevalence and impact of stalking: epidemiological data from Eastern Austria. Eur J Psychiatry 22:235–241
Storey J, Hart S (2011) How do police respond to stalking? An examination of the risk management strategies and tactics used in a specialized anti-stalking law enforcement unit. J Police Crim Psychol 26:128–142
Strand S, McEwan TE (2012) Violence among female stalkers. Psychol Med 42:545–556
Whyte S, Penny C, Christopherson S, Reiss D, Petch E (2011) The stalking of psychiatrists. Int J Forensic Ment Health 10:254–260.

State Crime

Dawn L. Rothe
Department of Sociology and Criminal Justice,
Old Dominion University, Norfolk, VA, USA

Overview

State crimes have been one of the foremost social problems of the past 100 years, with wide-reaching costs and levels of victimization. During the course of the twentieth century, the state crimes of Turkey, Nazi Germany, Stalinist Russia, Pol Pot and the Khmer Rouge, and Maoist China were especially large-scale, dramatic examples. More current examples include the continued possession of nuclear weapons by some states in violation of the Nuclear Nonproliferation Treaty; the US wars of aggression in Iraq and Afghanistan; genocides and crimes against humanity that include the Serbian attack on Bosnian and Croatia, Rwanda, Yugoslavia, and Sudan's actions against Darfur; and the Democratic Republic of Uganda and the Democratic Republic of Congo's wars and stealing of natural resources by states and corporations alike. Other charges of state crime occurring around the world as of 2011 include the political, economic, and militarized oppression involving Egypt, Israel, the United States, China, Russia, Chechnya, and North Vietnam all of which have resulted in the victimization of tens of millions that lost their lives, were rendered homeless, imprisoned, and psychologically and physically damaged through the illegal or socially harmful actions of governments.

Despite the gravity, costs, and extensiveness of crimes committed by states, the field of state crime remains understudied relative to conventional street crimes in the field of criminology and criminal justice. Nonetheless, over the last two decades, considerable theoretical, conceptual, and empirical progress has been made by criminologists to better specify the nature, extent, distribution, causal variables, and issues associated with state violence. However, there remain two primary areas of debate within the field of state crime: the standard by which to define state crime and the issue of controlling state criminality for practitioners, politicians, citizens, as well scholars of state crime. This entry provides an overview of the history of state crime and subsequent costs, followed by a discussion of the contentious issues that remain within the field. The entry concludes with a potential philosophical change that could prove to be the most fruitful means in which to constrain and control state criminality.

The Fundamentals

History and Overview of the Field of State Crime

The criminological study of state crime can be traced back to Edwin Sutherland (1939), who called attention to a then-neglected form of crime, namely, the crimes of respectable people in the context of a legitimate occupation and of corporations. His extension of the concept of crime, beyond its conventional parameters, provided an important foundation built upon by several later scholars of white-collar crime. However, one can argue that William Chambliss'

1988 American Society of Criminology Presidential address on state-organized crime provided the more direct and immediate inspiration for more systematic attention to crimes of the state. Exploring crimes such as piracy and smuggling, Chambliss showed how states can be crucial in the organization and support of activities that violate their own laws and international laws when doing so fulfills their broader political and economic objectives. A number of criminologists, particularly critical criminologists, quickly adopted the concept, broadening and enriching the field. Their early work focused not only on crimes tacitly supported or organized by a sovereign polity but also on actions committed on behalf of states themselves. However, the early research on state criminality was plagued by definitional issues and generated much debate regarding whether the individual or the state (organization) was culpable for acts deemed state crime and what standards should be used to define state criminality. These two contested areas cut to the core of the field of criminology in general; thus, it was not surprising that this debate influenced the early development of the field of state crime and, in some cases, continues today and will be discussed further in the following sections.

Regardless of the remaining contentious issues associated with standards and definitions, since the onset of criminological inquiry concerning state criminality began the field has grown exponentially. For example, literature has been produced on state crime with topics ranging from the US invasion of Iraq, the illegal use of and threatened use of nuclear weapons, the ongoing genocide in Darfur, crimes against humanity in Uganda, the treatment of illegal aliens, the US role in and lack of response to Hurricane Katrina, to the many cases of state-corporate crime such as the Challenger, Imperial Foods, ValuJet cases, and more recently cases involving Halliburton, BP, and Abu Ghraib. Additionally, there are now two comprehensive texts on state crime, Penny Green and Tony Ward's (2004) *State Crime* and Dawn L. Rothe's (2009) *State Criminality: The Crime of all Crimes*, and six edited anthologies on state crime, including *Resistance to State Crime*, edited by Elizabeth Stanley and Jude McCulloch (2012); *State Crime in the Global Age* edited by William Chambliss, Raymond Michalowski, and Ronald Kramer (2010); *State Crime, Current Perspectives*, edited by Dawn L. Rothe and Christopher W. Mullins (2010); *State-Corporate Crime: Wrongdoing at the Intersection of Business and Government* edited by Raymond Michalowski and Ronald Kramer (2006); and *Varieties of State Crime and its Control*, edited Jeffrey Ian Ross (2000). Additionally, John Hagan's (2010) *Who Are the Criminals? The Politics of Crime Policy from the Age of Roosevelt to the Age of Reagan* examines both state policies towards street crime as well as acts of state crime within and beyond the borders of the United States. Other cognate areas of state crime include political crimes, political white-collar crimes, environmental crimes, finance crimes, and the recently added crimes of globalization (see entries in this volume).

Costs of State Crime

Committing the most harmful of crimes are entities and individuals acting on the behalf of, or in the name of, the state. The types of costs include physical costs such as death, dismemberment, torture, ill health, and other costs associated with interpersonal violence directly and indirectly; environmental costs that include natural resources; economic costs; and psychological costs that are associated with a host of victimization and perpetrator mental and emotional health issues. For example, crimes of the state also can lead to the destruction of infrastructures, resulting in additional devastating harms (e.g., the 2002 sinking of the Senegalese ferry Le Joola with a loss of more than 1,000 lives) and environmental destruction with lasting effects on generations of citizens (e.g., the 1986 disaster at Chernobyl). Perhaps most importantly, state-committed crimes rip asunder the social trust between a state and its citizenry as well as trust between states (e.g., invasion and occupation of sovereign states, including most recently Iraq). When states violate such trust, domestically or internationally, they threaten the security of the global order, peace, and their own legitimacy. Given the above, it is impossible to estimate an overall cost associated with state crime, yet, we do know it far outweighs that of street crime.

Key Issues and Controversies

History of Contention: Definitions and Standards

Dating back to Sutherland, the notion of an organization being criminally liable had consistently been met with resistance by criminologists until the late 1970s through the mid-1980. During this period some criminologists began to incorporate ideas from organizational sociologists' research as they argued that social scientists needed to move beyond focusing on the individuals who make up an organization and to recognize that the aggregate whole functions as an entity. As such, it was suggested that organizations, as social actors, can and should be the primary focus of analysis in state and corporate crime.

Others strongly objected to the notion of a state, as a social actor, in an analysis of state or corporate crime. Furthermore, within criminology, the idea of a state being criminally liable was met with significant resistance. There were those that denied state criminality was possible. However, within the international legal arena, the notion of a state as an actor that could be held accountable was already well underway as the concept of a state, as an entity possessing individual rights and subject to criminal liability, emerged back in the mid-1900s. Likewise, several legal scholars observed that there was a connection between individual criminal responsibility and state criminal responsibility under international law.

On the other hand, scholars who supported the idea of state criminality were divided upon the standards to be used to define such acts as criminal between those who favored a legalistic frame and others who favored a broader frame ranging from social harms to human rights. The tensions and debates over defining state crime, however, reflect a broader debate within criminology itself. Utilization of state-produced legal codes has long been the stated and unstated norm. Such a reliance on state-produced definitions has caused tensions within street crime and white-collar crime studies, with critically orientated criminologists rejecting state-produced definitions. The political nature of law production has long been the main rationale for this rejection given that one cannot separate the nature of the political process that guides legislatures (and legislators) from the legislation produced. Further, states have an inherent drive to fulfill their own self-interest and not define harmful and problematic behavior as criminal (especially their own). Consequentially, within the field of state crime in particular, alternative formulations have been advanced, including the standard of basic human rights precepts, socially analogous harms, and social harm framework, to the perceptions of the state's citizens, each not without their own criticisms, strengths, and weaknesses.

Standards

The substance of a debate about definitions goes beyond merely critiquing the source or the substance of a given definition as alternative formulations were put forth. For example, in 1970, Herman and Julia Schwendinger suggested using a humanistic approach that would draw from objectively identifiable harms to humans and violations of human rights as the core definition of crime. Others have advocated that crimes are any socially injurious actions, regardless of the actor in question. Still some scholars have advocated that state crime should be defined by a social audience that recognizes the act as deviant. More recently, some criminologists have called for the abandonment of the concept of crime entirely in favor of Zemiology, the study of harm, thus, a social harm standard. In general, two positions on standards to be used to classify state actions as criminal remain within the broader definitional debate among scholars of state crime: crime as a social harm definition and a legalist approach.

The legalist approach includes a state's own domestic law as well as the broader umbrella of international public law (customary law, treaties, charters, and the newly emerged criminal law). This framework includes other approaches and standards such as human rights and social and economic harms. Furthermore, international criminal law covers individuals as well as states, thus resolving any enduring reservations of the state as actors versus individuals. Additionally, the legalists' use of extant statute identifies an external reference point, while other approaches are said to use a more amorphous and relativistic definitional rubric. Those that accept the use of a legalistic standard suggest that it adds legitimacy to the field's definition. Legalists argue that if a critique of state crime studies is that they are not truly scientific but rather politically inspired diatribes, establishing the illegality of such actions under a legal code is a fitting response to such critiques. Nonetheless, this approach has been criticized as another example of continuing to use law as a "tool of the state" to control the very entities that create it.

The social harm approach begins with the realization that crime has no ontological reality and it is subjectively defined by states within the context of broader issues of power and political and economic interests. In the case of crimes of the powerful, harmful activities are rarely defined as criminal. Many acts and behaviors that cause serious harm are not part of the domestic or international criminal law, thus being omitted by those using a stricter rubric such as the legalist standard. Those advocating this framework suggest that harm be defined as physical, financial and economical, psychological and emotional, sexual, and cultural. This includes the observable forms of harm, but also those seemingly more "natural" including death and illness caused by starvation, untreated treatable illnesses due to lack of medical care, those who lose their traditions and communities due to economic displacement and relocation, and a host of other ways millions of humans suffer and experience harm as a result of state policies and actions. Most importantly, this approach focuses on the origins of the harm rather than merely the actors or states involved or the act itself. This is not to say the other approaches do not do this in their theoretical and analytical assessments; however, those working from a social harm perspective begin with this focus.

There are pros and cons to using either standard to determine what is to be considered a state crime. For example, the use of international law provides clarity and precision in the definitional processes. One need not negotiate the problematic aspects of defining "harm" per se. One of the major issues with using the legalist framework is that both state-produced law and international law are the result of a questionably legitimate political process. Additionally, given the objective nature of law as defined, there is the criticism that using it as a frame to define behavior as criminal serves to maintain particular power relations at both the state and international level. In fact, international law often fails to be created or is created in a nonjusticiable way, to express politicization of its construction. While states not only have the ability to define crime, they are more often than not powerful enough to resist definitions of crime that label their own behavior criminal. The above issues are also related to a separate but equally important problem: that of enforcement.

On the other hand, the social harm framework allows for various forms of negligence and actions that are harmful yet do not meet any extant legal definition of crime to be examined and decried. This perspective allows criminologists to develop their own determination of their subject matter rather than rely on predetermined laws and human rights to dictate the field of inquiry or make the claim that a particular state behavior is "wrong," thus removing the external political influences. The social harm perspective then avoids the formal institutionalized problematic way in which crime is defined and potential additional levels of harm that could be generated by controlling crime in general through a formal system of response. Nonetheless, using a social harm frame has its own weaknesses. While this framework might add a conceptual and ideological purity to definitional processes, it takes an already broad subject matter and casts the net wider to what could be perceived as nearly the entirety of individual and institutional behavior within contemporary societies. Additionally, as with any other standard, the question becomes who then defines what is or is not harm?

Given the strengths and weaknesses of the above standards, the field of state crime now incorporates both positions. For example, while a legalist may well stick to the standard of international public law, there are times when such a standard is not acceptable. Consider the research on state crime involving, explicitly and implicitly, international financial institutions (IFIs) or crimes of globalization. Here, the intersection of these institutions not only results in immense harm, but the harms are not covered through the application of international public law, save for human rights violations in some cases. Although the policies and implicit and explicit actions of these IFIs can result in or facilitate state criminality, they are not covered under international public law, thus requiring a broader standard for defining such actions as "criminal." On the other hand, the use of a social harm standard can be complementary to the legalist approach when such harms include human rights violations. In both of the examples noted, the intermixing of these standards provides researchers of state criminality with the objective foundation as argued by the legalists within the limitations of the legal perspective. The value of both positions should not be recognized as the field of state crime research continues to grow.

The Quagmire of Controlling State Criminality

Resisting or controlling state criminality is bound to some degree with the definitional quagmire. Those that draw from the legalist standard promote formal and informal accountability mechanisms from social condemnation to prosecution. On the other hand, scholars working within the social harm framework generally see resistance and controls not in terms of accountability but in terms of policy responses. There are additional forms of harm that result from relying on a crime control industry. According to those who embrace a social harm standard, it is far more important to address not just the harms committed but also the underlying structural conditions that facilitated them. Accordingly, they suggest that this requires debates about policy and resources rather than handing the problem over to another arm of the broader power structure that, more often than not, further facilitates social harms: a criminal justice system. It requires using something beyond "the masters' tools" to confront, constrain, and control state crime. This can include movements from below, including social movements as has been witnessed in 2011 in Egypt, Tunisia, Libya, Bahrain, Yemen, and other Middle Eastern countries and with the Occupy Wall Street Movement.

Other scholars of state crime view international law as fundamental to the potential control of crimes of the state and mechanisms of accountability. These mechanisms include an international court such as the International Criminal Court and other states' domestic criminal justice systems. Additionally, impunity for crimes of the elite has long been an issue. Granting them near immunity for their actions in favor of extensive definitional and policy-related debates would suggest a reinforcement of impunity. Further, controlling state crime and holding heads of state or other high-ranking government officials accountable have long been met with contradictions and controversy from the perspective of state crime scholars as well as juristic practitioners. Consider that current geopolitical and international legal structures offer no threat of consequences. International law can be violated without threat of prosecution simply because there is no empowered institution to do so, and when it does occur, it is riddled with issues of selectivity. This reality has led many criminologists to deduce that the legalist standard and subsequent domestic and international laws are meaningless and not a valid framework for identifying or controlling state criminality. Indeed, for the legalist standard, the greatest obstacle is to actually control state criminality in some meaningful way. This includes ending impunity in the face of *realpolitik* and power differentials at the international level.

Discussions of resistance to complex forms of state criminality range from the formal to the informal, from the individual to the local to the state level, and to the international level. At the individual level, there are cases of dissenters and whistleblowers, as examples. At the local, organizational, and state levels, there are NGOs, media outlets, social movements, and political and civil groups that attempt to change existing conditions and expose state criminality. At the international level, there are political pressures from other states and international intergovernmental organizations and international laws that serve as deterrents to state crimes or after-the-fact accountability. Within each of these categories, resistance fluctuates and occurs at various points of time in relation to specific state criminality and the resources of organizations. It is through all of these mechanisms that awareness of the harms produced and illegalities of states' actions can come to be recognized and generate appropriate responses.

Future Directions

The term *cosmopolitan* is an ideology expressing that all of humanity belongs to a single moral community: an ideology for "global citizenship." Such an ideology is not new. For centuries philosophers have speculated on the conditions for achieving enduring peace between nations, one of which includes the notion of cosmopolitism or universalism. Consider Kant's thoughts in his 1795 essay *Perpetual Peace*, where he suggests *ius cosmopoliticum* (cosmopolitan law/right) be a guiding principle to protect people from war – grounded in the belief of universal hospitality. In more contemporary times, the idea of cosmopolitism reemerged in both formal and informal political realms that included the development of a United Nations, the global citizen, and an international community. While there remains a distinct contradiction between this ideology and the current state of international affairs, one vision for the future would be a reality of the ideal type where global citizens become the primary consideration in foreign policy and state actions. There is no magic answer or quick policy fix for state crimes: the worst crimes. It is an ongoing problem that requires diligence, commitment, and a change in ideology, praxis, and relations.

Related Entries

- ▶ Crimes of Globalization
- ▶ Crimes of the Powerful
- ▶ International Responses to Victims in Criminal Justice
- ▶ Resistance to State Crime
- ▶ State-Corporate Crime

Recommended Reading and References

Chambliss W (1990) State organized crime. Criminology 27(2):183–208

Chambliss W, Michalowski R, Kramer R (eds) (2010) State crime in a globalized age. Willin Press, Devon

Green P, Ward T (2004) State crime: governments, violence and corruption. London: Pluto Press

Hagan J (2010) Who are the criminals? The politics of crime policy from the age of Roosevelt to the age of Reagan. Princeton University Press, Princeton

Michalowski R, Kramer RC (eds) (2006) State-corporate crime: wrongdoing at the intersection of business and government. Rutgers University Press, Piscataway

Ross JI (ed) (2000) Varieties of state crime and its control. Criminal Justice Press, Mosney

Rothe DL (2009) State criminality: the crime of all crimes. Lexington/Roman and Littlefield, Manheim

Rothe DL, Mullins CW (eds) (2010) State crime, current perspectives. Rutgers University Press, Piscataway

Stanley E, McCulloch J (2012) Resistance to state crime. Routledge Press, London

Sutherland E (1939) White collar criminality. Presidential address to the American society of sociology. Reprinted 1940. American Sociological Review 5:1–12

State-Corporate Crime

Elizabeth A. Bradshaw[1] and Ronald C. Kramer[2]
[1]Department of Sociology, Anthropology and Social Work, Central Michigan University, Mt. Pleasant, MI, USA
[2]Department of Sociology, Western Michigan University, Kalamazoo, MI, USA

Overview

State-corporate crime has been defined as the illegal or socially injurious actions that occur when one or more institutions of political governance pursue a goal in direct cooperation with one or more institutions of economic production and distribution (Kramer et al. 2002). Originating from a series of papers and articles produced by Ronald Kramer and Raymond Michalowski in the early 1990s (see Michalowski and Kramer 2006), the concept of state-corporate crime unites two parallel streams of criminological literature to draw attention to the mutually dependent interorganizational relationships between governments and private corporations. Scholarship on state-corporate crime seeks to breach the conceptual wall between economic crimes and political crimes in order to create a new lens through which criminologists can examine the ways illegal acts and social harms often emerge from intersections of economic and political power. While the concept of state-corporate crime could be applied to illegal or other socially injurious actions in societies ranging from private production systems to centrally planned political economies, most of the early research focused on state-corporate crimes within the private production system of US capitalism. State-corporate crimes within a global capitalist economy involve the active participation of two or more organizations, at least one of which is in the civil sector and one of which is in the state sector.

Antecedents

The concept and theory of state-corporate crime evolved within the broader tradition of white-collar crime research in the field of criminology, and it introduces a third form of organizational crime to examine in addition to the separate focus on either corporate or state crime. The roots of the concept can be traced to Edwin Sutherland's analysis of corporate offenses in his classic work, *White-Collar Crime* (1949). Over time, two divergent approaches to white-collar crime research emerged. One approach embraced the occupational dimension of white-collar crime, while the other stressed the importance of focusing on organizational offenders. Occupational crimes consist of offenses committed by individuals for their own personal gain in an occupational setting, while corporate crimes are illegal or socially injurious acts committed by corporate officials acting on behalf of the corporation for the benefit of the corporation (Clinard and Quinney 1973). The organizational turn in white-collar crime research retained the importance Sutherland placed on studying crimes of the powerful and paved the way for an organizational analysis of corporate crime (Clinard and Yeager 1980; Kramer 1982).

Although the broader study of white-collar crime also includes offenses committed by government officials, the concept of state crime more explicitly draws attention to the illegal or harmful actions committed by the state as an organizational entity. Highlighting the important relationship between politics and the economy, state crime is often motivated by the need for capital accumulation by modern nation states. Defined as acts prohibited by criminal law and committed by state officials in the pursuit of their jobs as representatives of the state, "state-organized crime" includes acts such as piracy, smuggling, and illegal spying on citizens (Chambliss 1989). As with corporate offenses, deviant behavior by state actors can take the form of both crimes of commission and of omission (Kauzlarich et al. 2003).

To sum up, after the organizational turn in the field of white-collar crime, two nearly independent bodies of research had developed. Theory and research in the area of corporate crime had concentrated primarily on organizational deviance within private business corporations. Paralleling that work but seldom intersecting with it, others had examined crimes by governments. Despite its ubiquity, the structural relations between corporate and governmental organizations had been relatively peripheral to the study of organizational crime. Kramer and Michalowski suggested that many forms of organizational deviance are generated at the interstices of corporations and government and they introduced the term state-corporate crime to denote these particular forms of organizational deviance.

In addition, rather than limiting inquiry to acts in violation of criminal law, white-collar and organizational crime studies broadened the scope of criminological concern to include violations of other forms of law, such as regulatory law for the study of corporate crime and international law for the analysis of state crime. With regard to crimes of the state, since governments select which behaviors to criminalize, socially harmful actions committed by the state are often exempt from any legal scrutiny. For this reason, some criminologists argue that it is necessary to move beyond the confines of law in favor of approaches that focus on violations of human rights or other forms of social harm (Schwendinger and Schwendinger 1970; Hillyard et al. 2004). The study of state-corporate crime fully supports this expansion of criminological inquiry to encompass violations of regulatory and international law, as well as other legal, but socially injurious behavior, as legitimate phenomena for analysis.

Development

Initially, four case studies (Kramer 1992; Kauzlarich and Kramer 1993; Aulette and Michalowski 1993; Matthews and Kauzlarich 2000) established the foundation for understanding two distinct forms of deviant state-corporate interactions: state-initiated corporate crime and state-facilitated corporate crime. State-initiated corporate crime occurs when corporate entities undertake organizational deviance at the behest or direction of, or with the tacit approval of, government institutions. Illustrating this concept, the case studies of the *Challenger* explosion and the US production of nuclear weapons both draw attention to explicit role of the state in initiating deviant interorganizational actions. Alternatively, state-facilitated corporate crime occurs when government institutions of social control fail to restrain deviant business actions, either due to direct collusion between corporations and government or because they adhere to common goals that would be thwarted by forceful regulation. Both the investigation of the fire at the Imperial Food Products chicken processing plant and the crash of *ValuJet* 592 highlight the failure of the state to effectively constrain corporate criminality.

The original development of the concept of state-corporate crime stems from a case study of the deviant state and corporate intersections that resulted in 1986 *Challenger* space shuttle explosion. At first glance, the explosion of the *Challenger* appeared to simply be an accident. However, Kramer's (1992) examination demonstrated how the interactions between a government agency, the National Aeronautics Space Administration (NASA) and Morton Thiokol, Inc. (MTI), a private business corporation, led to risky decisions and unsafe actions that resulted in the death of six astronauts and one school teacher. While the technical cause of the disaster was the failure of the O-ring seal on the solid rocket booster, larger structural and organizational forces increased the likelihood that a dangerous outcome would occur. Budgetary compromises, political pressure to launch, and what (Vaughn 1996) later identified as the "normalization" of deviant practices within the organizational culture at NASA led to a catastrophic explosion.

Also focusing on the historical, political, and contextual factors contributing to deviant state and corporate interactions is Kauzlarich and Kramer's (1993) study of the environmental damage caused by the US nuclear weapons manufacturing industry. For more than 50 years, the Department of Energy (DOE) and the Atomic Energy Commission contracted with private multinational corporations such as Westinghouse, DuPont, General Electric, and Martin Marietta to produce nuclear weapons. While DOE owned the equipment and oversaw the production of nuclear weapons and materials around the country, the corporations were responsible for daily manufacturing operations. Producing nuclear weapons results in enormous amounts of radioactive and nonradioactive hazardous waste. This waste was improperly disposed of causing irreversible environmental damage. For example, two of the most environmentally harmful nuclear weapon facilities, the Hanford facility and the Savannah River plant which both produce plutonium and tritium, have released billions of gallons of liquid waste contaminating the local air, soil, groundwater, rivers, and watersheds. Encouraged by Cold War cultural beliefs and lacking interorganizational oversight, the nuclear weapons industry placed production of defense materials above the environmental consequences of warhead production (Kauzlarich and Kramer, 1993).

Another foundational case study of state-corporate crime by Aulette and Michalowski (1993) details the 1991 fire at the Imperial Food Products chicken processing plant in Hamlet, North Carolina, that resulted in the deaths of 25 workers and injured an additional 56. Although the technical cause of the fire was a rupture in the hydraulic line near the deep fryer that sent a wave of fire throughout the plant, it was discovered that Imperial had deliberately locked the fire doors to prevent employee theft, thereby denying them access to a safe exit. Beyond the actions of Imperial, a complex pattern of regulatory failure was revealed. Facilitated by a long history of privileging business interests over labor, North Carolina's neglect to fund the state's Occupational Safety and Health Program severely weakened regulatory oversight designed to protect workers. In contrast to the examinations of the *Challenger* explosion and the US manufacturing of nuclear weapons that demonstrate the direct role of the state in the commission of corporate wrongdoing, the Hamlet fire study identified a different type of relationship in which the state indirectly creates the conditions for corporate crime to occur (Aulette and Michalowski 1993).

Matthews and Kauzlarich's (2000) examination of the crash of *ValuJet* Flight 592 in the Florida Everglades on May 11, 1996, helps to further define the role of the state in facilitating corporate crime. While the explicit cause of the crash that killed all 105 passengers and five crewmembers was the explosion of oxygen generators in a cargo compartment that resulted in fire, government investigations also identified the failure of both ValuJet and SabreTech (an airline maintenance company) to comply with numerous regulations as important factors. Dually tasked with the conflicting mandates of regulating the safety of the airline industry while simultaneously promoting it, the Federal Aviation Administration (FAA) refused to implement safeguards and guidelines that could have protected passengers in favor of the economic interests of the airline industry. By ignoring two specific recommendations by the National Transportation Safety Board (NTSB) to place smoke detectors in cargo holds (the exact area the fire started in *ValuJet* 592), as well as to reclassify cargo holds to prevent the spreading of fire to the rest of the plane, the FAA indirectly sets the stage for the crash to occur (Matthews and Kauzlarich 2000).

A Theory of State-Corporate Crime

State-corporate crime has three useful characteristics as a sensitizing concept. First, it refutes the notion that organizational deviance is a discreet act by illuminating the relationships between social institutions. Second, by embracing the relational character of the state, the concept of state-corporate crime demonstrates how the horizontal interactions between political and economic institutions contain the potential for illegal and social injurious actions to occur (Wonders and Solop 1993). Finally, adopting a relational approach to the state not only allows for a consideration of horizontal interactions but of the vertical relationships between different levels of organizational action: political-economic, organizational, and interactional.

Corresponding to each level of social action are three theoretical approaches to the study of corporate crime: differential association theory, organizational theory, and political economy. Developed by Edwin Sutherland, differential association theory is a social psychological theory that seeks to establish the processes by which individuals learn deviant behavior. While differential association has been criticized for failing to consider the institutional context, the organizational perspective helps to link the external political-economic environment with the work-related thoughts and actions of individuals occupying structural positions in the social organization of work. Motivated by pressure to achieve organizational goals in a competitive setting, there exists an inherent inducement for the organization to participate in crime (Kramer 1982). The perspective of political economy demonstrates the implicit criminogenic potential of pressures for profit maximization within capitalist markets. By drawing attention to the structural relations between political and economic institutions, the political-economic perspective recognizes that many crimes of corporations and of governments are ultimately crimes of capital arising from the ownership or management of the accumulation process (Michalowski and Kramer 2006).

Independent of one another, each of these three theoretical approaches is inadequate to explain organizational deviance. In an effort to overcome these shortcomings, state-corporate crime scholars propose an integrated theoretical model of organizational crime that combines the three levels of analysis with three catalysts for action: motivation, opportunity, and control (Michalowski and Kramer 2006). The first catalyst for action concerns goal attainment. As the emphasis on goal attainment by political-economic institutions, organizations, and individuals increases, corporations and state agencies become more susceptible to engaging in organizational deviance. The second catalyst for action – opportunity – assumes that organizational deviance is more likely where legitimate means are scarce relative to goals. Finally, the third catalyst for action examines the presence or absence of social control at all three levels of analysis. Organizations subjected to a high operationality of social controls are more likely to cultivate organizational cultures that favor compliance with laws and regulations, and those organizations that are not subject to such controls are more likely to develop cultures of resistance. By investigating the linkages between levels of analysis and catalysts for action, a more nuanced understanding of state-corporate crime can potentially be developed.

Recent Research Trends

Beyond its early incarnations, the concept and theory of state-corporate crime has been increasingly used to understand diverse forms of organizational deviance engaged in by state and corporate actors. In addition to the early empirical and theoretical research on state-corporate crime, *State-Corporate Crime: Wrongdoing at the Intersection of Business and Government* (Michalowski and Kramer 2006) includes case studies on topics such as globalization, women and state-corporate crime, corporate facilitation during the Holocaust in Nazi Germany, the National Highway Transportation Safety Administration (NHTSA), Ford Motor Company and the Bridgestone-Firestone tread separation case, the Exxon Valdez oil spill, Enron era economics and economic democracy, violations of the treaty rights of indigenous peoples, the invasion of Iraq, as well as the role of Halliburton in Iraq. In addition to these early case studies, more recent research on state-corporate crime has emerged, particularly concerning international and environmental issues.

International Issues

A number of criminologists have paid increasing attention to the changing nature of state power in the global neoliberal economy and have begun to focus on the international arena. One theme among numerous recent case studies of state-corporate crime concerns the organizational deviance that has occurred due to the US invasion of Iraq. A second prominent theme is the relationship between private military firms and governments and the implications this has for state-corporate deviance. Moreover, other case studies on the state and corporate interactions surrounding international conflict over scarce natural resources, as well as the antecedents of the most recent global financial meltdown, have also been undertaken.

Kramer and Michalowski (2005) argue that the 2003 invasion and occupation of Iraq were illegal under international law and can therefore be classified as an instance of state crime. Laying the foundation for international relations is the United Nations (UN) charter, which at its core prohibits the use of aggressive force between nations with only two exceptions: self-defense and humanitarian interventions. Initially, the

Bush administration sought to implicate Iraq in the September 11th terrorist attacks in an effort justify war as self-defense. After being forced to concede that there was in fact no connection between Iraq and 9/11, the Bush administration argued for preemptory war due to allegations that Iraq possessed weapons of mass destruction, thereby constituting a direct threat requiring defensive military action. As a final strategy, the Bush administration attempted to convince the UN Security Council that it was necessary to go to war in order to protect Iraqi citizens from human rights abuses by Saddam Hussein. Despite all the efforts to legitimate war in Iraq, none met the legal requirements of international law and were unable to receive necessary approval from the UN Security Council. In the end, the March 2003 decision to go to war in Iraq was undertaken in clear violation of international law and paved the way for continued state-corporate crimes as a result (Kramer and Michalowski 2005).

Whyte (2007) explores how the overarching principle behind the US invasion and occupation of Iraq was the creation of a new rule of law based on the opening up of the economy to privatization by Western, and particularly US, corporations in breach of international law. Facilitating this transition, the Coalition Provisional Authority (CPA) was given the power to rebuild the Iraqi economy over a 14-month period. Guided by neoliberal economic principles, the CPA spent over $20 billion in Iraqi oil revenue to restructure the economy, much of which was disbursed to US corporations with scant oversight or documentation. Through 100 legally binding administrative orders and decrees, the CPA set the foundations of Iraq's new economy, justice system, and political structure based on the idea of "trickle down" economics. The Development Fund of Iraq (DFI) provided the CPA with immediate money from Iraqi oil revenues to be dispersed for the humanitarian needs of the people. Since the DFI funds were disbursed in bundles of cash, CPA transactions did not leave a paper trail and contracts frequently went undocumented entirely. Moreover, the Bush administration in conjunction with the CPA took steps to thwart audits and investigations into the dispersal of funds. A US government appointed auditor has since established that an unknown sum of the DFI funds have disappeared as a result of bribery, overcharging, embezzlement, product substitution, bid rigging, and false claims. These forms of state-corporate corruption under the CPA have been essential in achieving neocolonial dominance in occupied Iraq (Whyte 2007).

In the course of the occupation and invasion of Iraq, privatized security has been deployed on a scale never before seen. Rather than abdicating its authority to private military firms (PMFs), Welch (2009) contends that governments work in direct cooperation with them, creating a situation of fragmented power. Although it has enjoyed close ties with the Bush administration, even recruiting former officials from the CIA and the Pentagon, Blackwater has received criticism for its lethal actions in Iraq. For example, in May 2007, Blackwater employees opened fire on the streets of Baghdad twice in 2 days, including a standoff with Iraqi security forces. Another incident, labeled "Baghdad's Bloody Sunday," occurred on September 16, 2005, when Blackwater guards shot and killed between 8 and 20 innocent civilians and wounded dozens of others in Nisour Square. A documented repeat offender, Blackwater has also been investigated for at least six other episodes of excessive force. Despite this pattern of criminality, the US government has not only neglected to prosecute Blackwater employees and other contractors but has extended immunity from wrongdoing altogether with respect to the reconstruction of the Iraqi economy. In this manner, the decoupling of police and government forces enables private contractors to escape accountability for war crimes and human right abuses (Welch 2009).

Part of a broader trend of privatizing military functions, private military companies (PMCs) today have become a legitimate industry involved in a wide range of activities including protecting governmental and nongovernmental organizations during humanitarian missions, in addition to protecting corporate interests such as the extraction of oil and mining. Whyte (2003) argues that the unfolding of the PMC market can be understood as transference of law from international prohibition treaties and national criminal law to civil contracts as the principal means of legal regulation. Far from a reduction of state sovereignty in the era of neoliberal globalization, private military markets are dependent on the consent and support of governments for their livelihood. Moreover, governments have also come to recognize the benefits to be gained by forging a state-corporate alliance in the private military market. Accepting the proliferation of PMCs as not only inevitable but also desirable, the United Kingdom has embraced a "soft-touch" regulatory mix that strengthens state-corporate military relations in a number of ways. First, such an approach enhances British industrial competitiveness in weapons manufacturing and other industries. Second, encouraging the growth of PMCs also allows governments to monitor and influence remote territories while remaining ostensibly neutral, thereby conducting foreign policy by proxy. Finally, the difficulty of prosecuting PMCs under international law and the reluctance of the UK government to support the UN convention on mercenaries create a law-accountability gap. By expanding the PMC market, the opportunity structure for state-corporate crime is increased as states and corporations are able to engage in high risk or politically sensitive conflicts while evading accountability for their actions. Absent any new criminal legal controls, state-corporate crime in the PMC market is only likely to accelerate (Whyte 2003).

Providing a revised theoretical framework of state-corporate criminality, Rothe and Ross (2010) analyze how anomie and social disorganization, resulting in a lack of regulation, are significant factors in explaining the criminal propensity of private military companies (PMCs). Private forces (such as Bechtel, Blackwater, CACI International, DynCorp, Halliburton and subsidiary Kellogg, Brown, and Root, Logo Logistics, and Titan) provide a wide range of services including direct tactical military assistance, military consulting (strategic advisory and training), and logistic, intelligence, and maintenance services to armed forces. Within each of these sectors, however, there are variations in the types of crimes committed including murder, fraud, and war profiteering. PMCs are not held to the same rules of engagement as the military and have an unclear legal status that is undefined by international law. Lacking internal and external constraints at every level, PMCs operate in anomic (lawless or normless) conditions that cultivate criminogenic behavior. Social disorganization occurs when communities are no longer able to create and enforce informal mechanisms of social control. As they operate in war ravaged areas, PMCs function in disorganized environments with uncertain mandates, high rates of employee turnover, and little social support. These factors converge to produce an environment in which regulation and social control breaks down and PMCs are free to engage in deviant behavior without consequence (Rothe and Ross 2010).

In the face of recent US actions that violate international law such as the war on terror, the war in Afghanistan, and the war in Iraq, a large segment of the citizenry nevertheless shows support for these policies. Part of the ideology of military aggression, Klein and Lavery (2011) argue, is perceived in-group victimization. A corporately owned political and cultural institution, the mainstream media functions as a tool by which politicians emphasize and validate the national experience of collective victimization. In an attempt to legitimate their illegal actions, elites and organizations use the media to amplify their claims of victimization and thereby neutralize criticism. At least in part, fear of being a terror victim seems to be related to the content of news media. Moreover, there also seems to be support for the notion that terrorism victimization in the media can contribute to public acceptance of state-corporate criminality (Klein and Lavery 2011).

Confronted with widespread political and economic violence for much of the past century, the experience of the Democratic Republic of Congo (DRC) cannot solely be explained by internal factors experienced in the transition from colonial to postcolonial rule. Rather, Mullins and Rothe (2008) argue that neighboring states, transnational corporations, and international organizations all play a significant role in the continuing violence within the DRC. Seeking control of large concentrations of valuable mineral fields including gold, diamonds, silver, copper, and coltan, transnational corporations have been more than willing to negotiate access with warlords and militias controlling the hinterlands despite widespread human rights abuses. While demand from European markets is an important factor spurring participation by transnational corporations and their Western trading partners, states and international organizations also play a crucial role in facilitating the sale of illegally misappropriated resources. Capitalizing on the chaos surrounding the Second Congolese War in 1995 by enacting patterns of illegal resource appropriations particularly for gold and diamonds, Uganda acknowledged to the World Bank (WB) that production levels might reflect exports usurped from the DRC. The WB and the International Monetary Fund praised Uganda and Rwanda for their increases in production, despite the illegal means by which it was achieved. While it is largely recognized that there is no indigenous production of gold in Uganda, Metalor Technologies, a Swiss refinery, has been responsible for purchasing the Congolese gold from Ugandan sources. Although it had been demonstrated that over 100 companies from over a dozen countries are benefiting from illegal mineral exploitation, there has been reluctance on behalf of governments to intervene. The institutional structure of the larger international community therefore plays an important role in perpetuating the political and economic violence in countries engulfed by genocidal civil war such as the DRC (Mullins and Rothe 2008).

As an investigator for a 2002 Royal Commission examining fraud within the Dutch construction industry, Van Den Heuvel (2005) helped reveal that rather than an isolated incident of collusion, the entire sector had engaged in illegal practices including fraud, undercutting the market, unjustified subsidies, monopolization resulting in higher prices, and bribery of politicians and public servants. At the core, the Commission identified the industry-wide pandemic to be due to multiple forms of collusion or a secret agreement for a fraudulent purpose. One form of collusion occurred between contractors themselves, such as illegal price fixing. A second type involved collusion between authorities and contractors, such as favoring particular contractors over others. Finally, a third type of collusion identified by the Commission occurred at the individual level in the bribing of public servants, for example, providing generous gifts and favors. The Commission concluded that these forms of collusion within the Dutch construction industry were so interconnected that they constituted a culture that placed contactors above the law and in control of authorities. Because of the pervasive collusion between contractors and authorities, it seems necessary to enact stricter rules governing state-corporate interorganizational relationships (Van Den Heuvel 2005).

Liederbach (2010) contends that while the state established the framework for the 2008 mortgage default crisis, corporate malfeasance was at its core. From the 1930s through the 1970s, there were stringent government regulations on mortgages which restricted lending practices such as issuing variable rate loans, capped interest rates, and prohibited partial monthly payments (known as negatively amortized mortgages). However, statutory changes and deregulation by Congress during the early 1980s removed these restrictions. Coasting on the economic success of the 1990s, an effort to increase national homeownership led by the US Department of Housing forced the government-sponsored entities, Fannie Mae and Freddie Mac, to engage in riskier subprime mortgage markets since the targeted borrowers could not qualify for prime loans. Setting the conditions for the crisis to thrive, during the early 2000s, the Federal Reserve expanded the availability of credit by setting and maintaining surprisingly low interest rates and decreasing the amount of money banks were required to keep in reserve. All of these factors culminated in an economic meltdown that resulted in millions of home foreclosures and an unprecedented $750 billion taxpayer-funded bailout to rescue distressed financial institutions (Liederbach 2010).

Environmental Issues

Another emerging trend within the research on state-corporate crime is case studies examining the environmental harm resulting from organizational deviance. Of increasing significance are the state and corporate interactions contributing to the critical problem of global warming and climate change.

Smandych and Kueneman (2010) contend that Canadian officials and oil corporations are guilty of manipulating environmental regulations as well as the practice of "greenwashing:" disinformation provided by an organization attempting to present an environmentally friendly public image, in this case surrounding the Alberta tar sands project. Criticized for its concentrations of toxic waste and a major source of greenhouse gas emissions, the aggressive state-corporate development of the Alberta tar sands project began in 1990s. Effectively surrendering sovereignty over the nation's resources, the 1989 Canada-USA Free Trade agreement obliges Canada to share its oil and gas resources with the USA. The two means of extracting the oil from a mixture primarily comprised of sand are strip mining and underground "in situ" mining, both of which are environmentally disastrous leading to the devastation of pristine boreal forests, contamination of the Athabasca watershed, and depletion of natural resources. In this regard, the USA accrues the benefits while Canada suffers pollution of its land, air, and water. Working hand in hand with the oil industry, specifically US-based multinational oil corporations, the conservative governments of the Province of Alberta and the federal government of Canada have suppressed attempts to define the harm caused by the tar sands as criminal activity (Samandych and Kueneman 2010).

Lynch et al. (2011) applied the concept of state-corporate crime to global warming by examining the politicization of global warming under the G.W. Bush administration. Although industry leaders are often selected for governmental positions, appointing corporate leaders from the oil and mining industries to crucial environmental policy positions was a dominant trend in the Bush administration that strengthened state-corporate ties and deterred action on climate change. Despite the overwhelming consensus of climate change scientists, the Bush administration colluded with corporations in the fossil fuel industries in an effort to discredit and suppress science on the human causes of global warming. White House officials deliberately sought to undermine science on global warming in a number of ways including blocking publications, editing government reports, altering federal policy, and pressuring climate change scientists to delete references to global warming and climate change in government-sponsored research. Furthermore, the corporate strategy for undermining information on global warming was to fund and create front organizations to disseminate misinformation about climate change to the public. In this manner, the policies of the Bush administration privileged the interests of the fossil fuel industry over the interests of the public (Lynch et al. 2011).

Four forms of state-corporate crime shape the social and environmental harms caused by global warming. Kramer and Michalowski (2012) argue that state and corporate actors produce these harms by (1) denying that global warming is caused by the actions of humans (anthropogenic), (2) thwarting attempts to reduce greenhouse emissions, (3) excluding from the political arena ecologically just adaptations to climate change, and (4) responding to the social conflicts that arise as a result of climate change with militarism and violence. While anthropogenic global warming is the result of over 200 years of industrialization and fossil fuel consumption, it is possible to identify state-corporate relationships that caused knowable and predictable harm and that could have been avoided. Therefore, the failure of state institutions to mitigate or reduce carbon emissions in the private and public sectors should be understood as a state-corporate crime of omission. More than just a failure to act, however, the orchestrated denial of climate change despite overwhelming scientific evidence to the contrary constitutes a state-corporate crime of commission. Designed to cast doubt on the evidence for anthropogenic global warming, the global warming denial counter-movement has been directed, organized, and funded by corporations and conservative think tanks. States and corporations have also refused to seriously consider socially just adaptation policies despite the increasing number of social conflicts resulting from climate change. This too, it is argued, constitutes a state-corporate crime of omission (Kramer and Michalowski 2012).

Conclusion

The study of political economy has been around for a long time, and the idea that corporations and states often act together in ways that have serious social consequences is not new. However, the concept and theory of state-corporate crime filled a gap in the evolving study of organizational crime within criminology. By examining social harms that result from the interaction of political and economic organizations, scholarship on state-corporate crime has made an important contribution to the field. The concept has generated a substantial body of research, and some argue, "the approach developed by state-corporate crime scholars is a significant advance toward developing a powerful integrated theoretical model" (Green and Ward 2004, p. 51). Judged by the growing number of criminologists who find the concept state-corporate crime useful in their work, research and theory in this area have the potential to make additional contributions to the study of organizational offending, particularly with regard to international and environmental crimes.

Related Entries

- ▶ Corporate Crime Control
- ▶ Corporate Crime Decision-Making
- ▶ Crimes of Globalization
- ▶ Crimes of the Powerful
- ▶ Green Criminology
- ▶ Nuclear Weapons and State Crime
- ▶ Political Crime
- ▶ Resistance to State Crime
- ▶ State Crime
- ▶ Victims of State Crime

Recommended Reading and References

Aulette J, Michalowski R (1993) Fire in Hamlet: a case study of state-corporate crime. In: Tunnell K (ed) Political crime in contemporary America: a critical approach. Garland, New York, pp 171–206

Chambliss W (1989) State-organized crime- the American society of criminology, 1988 presidential address. Criminology 27:183–208

Clinard M, Quinney R (1973) Criminal behavior systems: a typology, 2nd edn. Holt, Rinehart and Winston, New York

Clinard M, Yeager P (1980) Corporate crime. The Free Press, New York

Friedrichs D (2010) Trusted criminals: white collar crime in contemporary society, 4th edn. Wadsworth, Belmont

Green P, Ward T (2004) State crime: governments, violence and corruption. Pluto Press, London

Hillyard P, Pantazis C, Tombs S, Gordon D (eds) (2004) Beyond criminology: taking harm seriously. Pluto Press, London

Kauzlarich D, Kramer R (1993) State-corporate crime in the US nuclear weapons production complex. J Hum Justice 5:4–28

Kauzlarich D, Mullins C, Matthews R (2003) A complicity continuum of state crime. Contemp Justice Rev 6:241–254

Klein J, Lavery C (2011) Legitimating war by victimization: state-corporate crime and public opinion. Crime Law Soc Chang 56:301–317

Kramer R (1982) Corporate crime: an organizational perspective. In: Wickman PW, Daily TB (eds) White collar and economic crime. Lexington Books, Lexington, pp 75–94

Kramer R (1992) The space shuttle challenger explosion: a case study of state-corporate rime. In: Schlegel K, Weisburd D (eds) White collar crime reconsidered. Northwestern University Press, Boston, pp 214–243

Kramer R, Michalowski R (2005) War, aggression and state crime: a criminological analysis of the invasion and occupation of Iraq. Br J Criminol 45:446–469

Kramer R, Michalowski R (2012) Is global warming a state-corporate crime? In: White R (ed) Climate change from a criminological perspective. Springer, New York

Kramer R, Michalowski R, Kauzlarich D (2002) The origins and development of the concept and theory of state-corporate crime. Crime Delinq 48:263–282

Liederbach J (2010) Pass the trash: the mortgage default crisis as state-corporate crime. Sociol Crime Law Devianc 15:17–41

Lynch M, Burns R, Stretesky P (2011) Global warming and state-corporate crime: the politicization of global warming under the bush administration. Crime Law Soc Chang 54:213–239

Matthews R, Kauzlarich D (2000) The crash of ValuJet flight 592: a case study in state-corporate crime. Sociol Focus 33:281–298

Michalowski R, Kramer R (eds) (2006) State-corporate crime: wrongdoing at the intersection of business & government. Rutledge University Press, New Brunswick

Mullins C, Rothe D (2008) Gold, diamonds and blood: international state-corporate crime in the Democratic Republic of the Congo. Contemp Justice Rev 11:81–99

Rothe D, Ross JI (2010) Private military contractors, crime, and the terrain of unaccountability. Justice Q 27:593–617

Samandych R, Kueneman R (2010) The Canadian-Alberta tar sands: a case study of state-corporate environmental crime. In: White R (ed) Global environmental harm: criminological perspectives. Willan Publishing, Portland, pp 87–109

Schwendinger H, Schwendinger J (1970) Defenders of order or guardians of human rights? Issues Criminol 5:123–157

Sutherland E (1949) White collar crime. Dryden Press, New York

Van Den Heuvel G (2005) The parliamentary enquiry on fraud in the Dutch construction industry: collusion as concept between corruption and state-corporate crime. Crime Law Soc Chang 44:133–151

Vaughn D (1996) The challenger launch decision: risky technology, culture, and deviance at NASA. University of Chicago Press, Chicago

Welch M (2009) Fragmented power and state-corporate killings: a critique of blackwater in Iraq. Crime Law Soc Chang 51:351–364

Whyte D (2003) Lethal regulation: state-corporate crime and the United Kingdom government's new mercenaries. J Law Soc 30:575–600

Whyte D (2007) The crimes of neo-liberal rule in occupied Iraq. Br J Criminology 47(2):177–195

Wonders N, Solop F (1993) Understanding the emergence of law and public policy: toward a relational model of the state. In: Chambliss W, Zatz M (eds) Making law. Indiana University Press, Bloomington, pp 204–225

STDs

► Drug Abuse and Alcohol Dependence Among Inmates

Stigma

► Rational Choice, Deterrence, and Crime: Sociological Contributions

Sting Operations

► Law of Undercover Policing

Stolen Auto

► Motor Vehicle Theft

Stolen Vehicle

► Motor Vehicle Theft

Stop Now and Plan (SNAP®) Model

Leena Augimeri[1], Margaret Walsh[2], Kathy Levene[2], Karen Sewell[2] and Erin Rajca[2]
[1]Centre for Children Commiting Offenses, Child Development Institute, Toronto, ON, Canada
[2]Child Development Institute, Toronto, ON, Canada

Synonyms

Centre for Children Committing Offences (CCCO); Earlscourt Child and Family Centre (ECFC); Earlscourt Girls Connection (EGC); SNAP®; SNAP® Boys; SNAP® GC; SNAP® Girls; SNAP® ORP; Under 12 Outreach Project (ORP)

Overview

This entry presents a comprehensive summary of the SNAP® (Stop Now and Plan) model including its related programs which are part of a crime prevention strategy for at-risk children under the age of 12. At Child Development Institute (CDI), a multiservice, not-for-profit children's mental health organization, the mandate for two key evidence-based clinical programs, SNAP® Boys and SNAP® Girls, and the Centre for Children Committing Offences (CCCO) promotes effective services for these "forgotten" children who have or are at-risk of having police and/or child welfare contact for their disruptive behavior. The comprehensive strategy includes (1) police-community referral protocols; (2) gender-specific risk assessment using Early Assessment Risk Lists (EARLs); and (3) gender-specific interventions, the evidenced-based SNAP® programs (for further details, see Koegl et al. 2008). The overarching goal of SNAP® is to keep high-risk children in school and out of trouble. SNAP® has been found to reduce aggression, delinquency, and antisocial behavior; increase social competency; prevent further and future delinquency; improve academic success by decreasing behavioral issues at school; engage high-risk children and their families in service; increase effective parent management skills; and connect children and parents to community-based resources. Authors discuss key aspects of the SNAP® model including SNAP® principles, theoretical foundation, model framework, and research summary of findings to date.

Fundamentals of SNAP®

Over 27 years, CDI has developed its expertise in responding to children with disruptive behavior problems and their families. In 1985, with the decriminalization of children under the age of 12 in Canada, CDI (with support from the provincial Ministry of Children and Youth Services in Ontario, Canada) developed SNAP® programs in response to the need of mental health services for this age group of high-risk children and families. The SNAP® model is based on a comprehensive framework (Fig. 1, SNAP® Model Framework) for effectively teaching children with serious behavior problems emotion regulation, self-control, and problem-solving skills. Parents also learn SNAP® skills as well as researched cognitive-behavioral parenting techniques. Children and families learn how to stop and think in order to find solutions that "make their problems smaller, not bigger." The

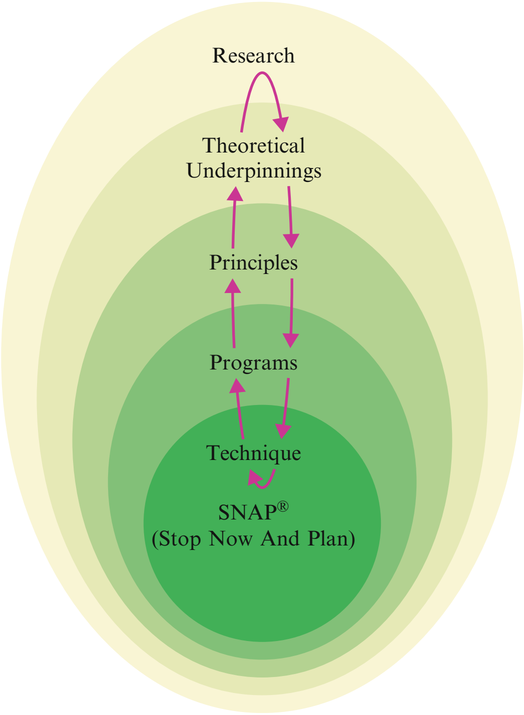

Stop Now and Plan (SNAP®) Model, Fig. 1 SNAP® model framework

SNAP® Model Framework depicts the interconnectedness and relationship of the theoretical underpinnings, principles, programs, and technique and how research plays a role in each of these key areas. The SNAP® programs have been developed with the *technique* as the cornerstone of the program components, have been continuously informed by theory and research, and are delivered through adherence to the SNAP® principles. This dynamic model provides feedback loops which allow for fluidity between the elements to influence and inform ongoing development. The evidence-based SNAP® programs have been adapted to different populations and settings: SNAP® Boys, SNAP® Girls, SNAP® Schools, SNAP® for Children with Asperger Syndrome, SNAP® for Aboriginal Communities, SNAP® Youth Leadership, and SNAP® for Youth in Custody.

Technique

There is robust evidence that early childhood interventions focused on enhancing self-control are likely to bring greater return on investments (Moffitt et al. 2011). Further support for this approach is identified in other investigations which highlight that children tend to be good candidates for learning self-control strategies, especially before the age of 10 (Piquero et al. 2010). The SNAP® technique is a cognitive-behavioral emotion regulation, self-control, and problem-solving strategy intended to help children *stop and think before they act* and come up with socially appropriate plans to address their problems – helping to control impulsivity, challenge cognitive distortions, and think about the consequences of their behavior. A key aspect of the SNAP® technique is to help children identify their bodies' physiological responses (*body cues*), thoughts, feelings, and triggers (things that make them feel angry/sad/worried) and help them to make the connection between their body cues, feelings, thoughts, and what they can do to effectively regulate arousal levels and help their bodies calm down (*Stop*), so that they can come up with an effective *Plan*. As discussed in the article, *Rolling Out SNAP® – An Evidence-Based Intervention: A Summary Of Implementation, Evaluation, and Research* (Augimeri et al. 2011a), SNAP® was first developed in the former Earlscourt Child and Family Centre's day treatment classroom for children with behavioral problems in the late 1970s, and the technique underlies the entire foundation of the SNAP® Model Framework (see Fig. 1). This was then formalized with the creation and publication of program manuals (Earlscourt Child and Family Centre 1990a, b, 2001a, b, 2002; Levene 1998) and trademarked in 1998.

As noted in the SNAP® program manuals, there are a number of steps to the SNAP® technique that have been mapped onto the image of a stoplight – red light (*Stop*), yellow light (*Now and*), and green light (*Plan*). These steps are used to teach children to regulate their emotions by helping them to calm down (e.g., by taking deep breaths and/or counting to ten) (*Stop*); replace "hard thoughts with cool thoughts" (coping statements, cognitive restructuring) to help them remain calm (e.g., "this is hard but I can do this") (*Now and*); and generate effective solutions which meet these three criteria: 1. make their problems smaller instead of bigger; 2. make them feel like a winner; and 3. not hurt anyone, anything or themselves (*Plan*).

Programs

The introduction of the first SNAP® program (SNAP® Under 12 Outreach Project, now known as SNAP® Boys) in Toronto in 1985 was designed specifically to address the gap in services when the age of criminal responsibility in Canada was raised from seven to 12 under the *Young Offenders Act* (YOA) in 1984. Prior to the YOA, the *Juvenile Delinquents Act* (JDA) enacted in 1908 prosecuted children as young as 7 years of age. The YOA placed these children under the responsibility of child welfare legislation versus criminal justice; this remained when Canada replaced the YOA with the Youth Criminal Justice Act (YCJA) in 2003.

SNAP® Boys was launched in partnership with Toronto Police Service through provincial funding from the former Ministry of Community and Social Services (today, Ministry of Children and Youth Services). The mandate of SNAP® Boys is to serve children under the age of 12 who are engaging in antisocial behaviors who do not legally fall under the purview of the YCJA. SNAP® Boys is noted as the most fully developed, longest sustained, empirically based multicomponent intervention specifically for "pre-offender" youth under the age of 12 (Howell 2001, 2003). Its sister program, the SNAP® Girls Connection (now known as SNAP® Girls), began in 1996 and is the first reported *gender-specific* intervention for girls under the age of 12 with disruptive behavior problems. Both programs are fully manualized and are in various stages of replication worldwide.

Presenting issues of the children admitted into the SNAP® programs typically include stealing, lying, mischief, vandalism, aggression, assault, bullying, and truancy. A significant number of these children also experience academic difficulties and comorbid mental health symptomology, such as depressive and anxious behaviors or ADHD (Pepler et al. 2004; Walsh et al. 2002).

Children admitted into the SNAP® clinical programs (SNAP® Boys and SNAP® Girls) have had police contact for their own misbehavior and/or have a score within the clinical range on standardized measures assessing externalizing behavioral issues (aggression, conduct, rule breaking). Primary referral sources include schools, police, child welfare, parents, and other mental health and medical professionals (see Fig. 2).

As noted in the SNAP® Logic Model (Fig. 2), assessment is informed by an ecosystemic approach (also a SNAP® *principle*, see Table 1) that takes into account interventions targeting the child, the family, the school, and the community. The Early Assessment Risk List (EARL-20B for boys or EARL-21G for girls), a structured clinical risk/need assessment device for use with aggressive and delinquent children, is also completed to provide a comprehensive framework for evaluating risk factors known to influence a child's propensity to engage in future antisocial behavior. Informed by the ecosystemic assessment, the risk assessment takes into account multi-informant perspectives (child, parent, teacher, and clinician), identifies the unique treatment needs of children and their families, and assists clinicians with treatment planning in order to mitigate these risks.

The SNAP® clinical programs (SNAP® Boys and SNAP® Girls) offer multifaceted services including core and adjunct components which are available to children and families based on their level of risk and need. In the SNAP® programs, components are goal oriented, skill focused, and developmentally responsive. Integrated into each component are key skill acquisition training techniques (i.e., role-play, modeling, self-talk) and generalization activities (home practice assignments) to transfer learning of the SNAP® technique and SNAP® parenting skills from the clinical environment to real-life settings. Following the ecosystemic and EARL assessments, a treatment plan is tailored to the child and family's strengths, risks, and needs. Children and families typically begin with completing the core components. Core components include:

SNAP® Children's Group – a gender-specific manualized core component that focuses on teaching children emotion regulation, self-control, and problem-solving skills with a special emphasis on challenging cognitive distortions/thinking errors. Examples of topics covered include introduction to SNAP®, peer pressure, dealing with anger, and bullying. Children participate in a 13-week SNAP® group, occurring once a week for 1.5 h.

SNAP® Parent Group – a manualized core component that runs concurrent with the SNAP® children's groups. Parents learn emotion regulation, self-control, and problem-solving skills and effective child management strategies with a special emphasis on challenging cognitive distortions/thinking errors, reducing isolation, and enhancing parent-child relationships. Examples of topics covered include effective communication, positive reinforcement, limit setting and consequences, and family problem solving. Three of the group sessions are joint parent-child sessions where parents and children practice skills together.

Girls Growing Up Healthy (GGUH) – a manualized core component unique to SNAP® Girls, this mother-daughter group focuses on enhancing relationship capacity, healthy relationships, and physical and sexual health. Topics covered include preparing for puberty, deconstructing stereotypes, and planning for the future. Caregivers and daughters meet for 8 weeks, once a week for 2 h.

The following adjunct components are offered based on the child's and family's identified goals, strengths, risks, and continued treatment needs:

Individual Counseling/Mentoring/Community Connections – provides children with individualized support with a SNAP® worker to reinforce and enhance skills learned in the SNAP® children's group and address treatment targets (i.e., social skills, coping ability, cognitive restructuring) and goal attainment. Children can also be matched with volunteers who connect them with structured recreational activities in their community.

Stop Now and Plan Parenting (SNAPP): individualized family counseling – based on our SNAPP Manual, is offered to families who are

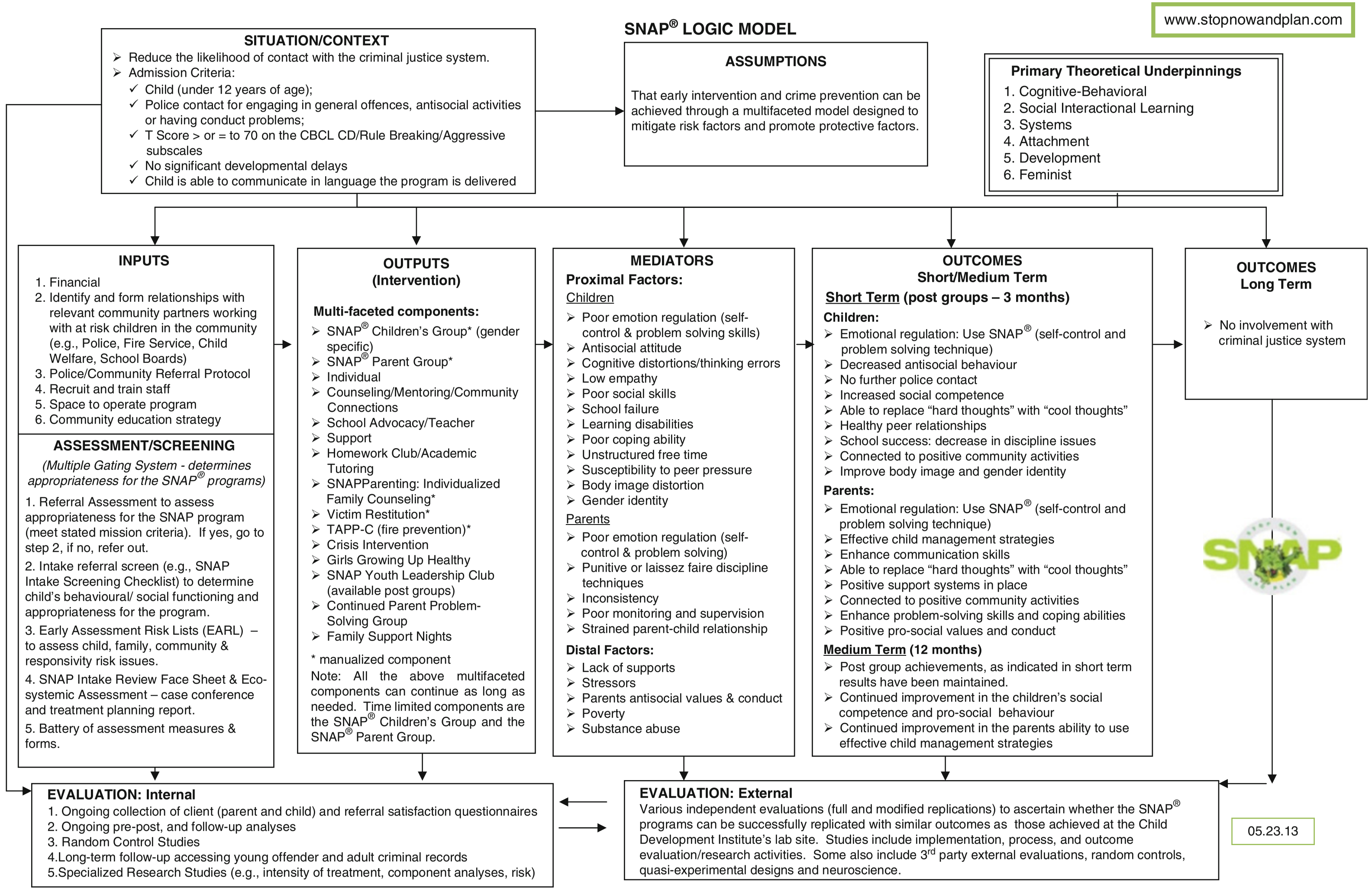

www.stopnowandplan.com

SNAP® LOGIC MODEL

SITUATION/CONTEXT

- Reduce the likelihood of contact with the criminal justice system.
- Admission Criteria:
  - ✓ Child (under 12 years of age);
  - ✓ Police contact for engaging in general offences, antisocial activities or having conduct problems;
  - ✓ T Score > or = to 70 on the CBCL CD/Rule Breaking/Aggressive subscales
  - ✓ No significant developmental delays
  - ✓ Child is able to communicate in language the program is delivered

ASSUMPTIONS

That early intervention and crime prevention can be achieved through a multifaceted model designed to mitigate risk factors and promote protective factors.

Primary Theoretical Underpinnings

1. Cognitive-Behavioral
2. Social Interactional Learning
3. Systems
4. Attachment
5. Development
6. Feminist

INPUTS

1. Financial
2. Identify and form relationships with relevant community partners working with at risk children in the community (e.g., Police, Fire Service, Child Welfare, School Boards)
3. Police/Community Referral Protocol
4. Recruit and train staff
5. Space to operate program
6. Community education strategy

ASSESSMENT/SCREENING

(Multiple Gating System - determines appropriateness for the SNAP® programs)

1. Referral Assessment to assess appropriateness for the SNAP program (meet stated mission criteria). If yes, go to step 2, if no, refer out.

2. Intake referral screen (e.g., SNAP Intake Screening Checklist) to determine child's behavioural/ social functioning and appropriateness for the program.

3. Early Assessment Risk Lists (EARL) – to assess child, family, community & responsivity risk issues.

4. SNAP Intake Review Face Sheet & Eco-systemic Assessment – case conference and treatment planning report.

5. Battery of assessment measures & forms.

OUTPUTS (Intervention)

Multi-faceted components:

- SNAP® Children's Group* (gender specific)
- SNAP® Parent Group*
- Individual
- Counseling/Mentoring/Community Connections
- School Advocacy/Teacher
- Support
- Homework Club/Academic Tutoring
- SNAPParenting: Individualized Family Counseling*
- Victim Restitution*
- TAPP-C (fire prevention)*
- Crisis Intervention
- Girls Growing Up Healthy
- SNAP Youth Leadership Club (available post groups)
- Continued Parent Problem-Solving Group
- Family Support Nights

* manualized component

Note: All the above multifaceted components can continue as long as needed. Time limited components are the SNAP® Children's Group and the SNAP® Parent Group.

MEDIATORS

Proximal Factors:

Children

- Poor emotion regulation (self-control & problem solving skills)
- Antisocial attitude
- Cognitive distortions/thinking errors
- Low empathy
- Poor social skills
- School failure
- Learning disabilities
- Poor coping ability
- Unstructured free time
- Susceptibility to peer pressure
- Body image distortion
- Gender identity

Parents

- Poor emotion regulation (self-control & problem solving)
- Punitive or laissez faire discipline techniques
- Inconsistency
- Poor monitoring and supervision
- Strained parent-child relationship

Distal Factors:

- Lack of supports
- Stressors
- Parents antisocial values & conduct
- Poverty
- Substance abuse

OUTCOMES Short/Medium Term

Short Term (post groups – 3 months)

Children:

- Emotional regulation: Use SNAP® (self-control and problem solving technique)
- Decreased antisocial behaviour
- No further police contact
- Increased social competence
- Able to replace "hard thoughts" with "cool thoughts"
- Healthy peer relationships
- School success: decrease in discipline issues
- Connected to positive community activities
- Improve body image and gender identity

Parents:

- Emotional regulation: Use SNAP® (self-control and problem solving technique)
- Effective child management strategies
- Enhance communication skills
- Able to replace "hard thoughts" with "cool thoughts"
- Positive support systems in place
- Connected to positive community activities
- Enhance problem-solving skills and coping abilities
- Positive pro-social values and conduct

Medium Term (12 months)

- Post group achievements, as indicated in short term results have been maintained.
- Continued improvement in the children's social competence and pro-social behaviour
- Continued improvement in the parents ability to use effective child management strategies

OUTCOMES Long Term

- No involvement with criminal justice system

EVALUATION: Internal

1. Ongoing collection of client (parent and child) and referral satisfaction questionnaires
2. Ongoing pre-post, and follow-up analyses
3. Random Control Studies
4. Long-term follow-up accessing young offender and adult criminal records
5. Specialized Research Studies (e.g., intensity of treatment, component analyses, risk)

EVALUATION: External

Various independent evaluations (full and modified replications) to ascertain whether the SNAP® programs can be successfully replicated with similar outcomes as those achieved at the Child Development Institute's lab site. Studies include implementation, process, and outcome evaluation/research activities. Some also include 3rd party external evaluations, random controls, quasi-experimental designs and neuroscience.

05.23.13

Stop Now and Plan (SNAP®) Model, Fig. 2 SNAP® logic model

Stop Now and Plan (SNAP®) Model, Table 1 SNAP® principles and indicators

Principle	Indicator
Scientist-practitioner	An interactive science-practice paradigm is sustained by regular cross discipline workgroups to support the high-quality evolution of SNAP® program development, outcomes, and research
Client centered	Services are informed by client personal and cultural narratives and goals, ensuring client responsiveness through growth-oriented engagement and alliances
Gender sensitive	Specific gendered factors which account for differential development, learning styles, and trajectories of boys and girls with conduct problems are integrated throughout SNAP® programs
Ecosystemic	Each client's ecological system (individual, home, school, community) is assessed to identify and inform strengths, needs, and risk and protective factors and to match them with appropriate service components and treatment intensity
Strength and skill based	Specific, consistent use of positive reinforcement, as part of our evidence-based, cognitive-behavioral practice, promotes and strengthens individual capabilities and the acquisition of primary SNAP® skills: emotion regulation, self-control, and problem solving
Continuing services	Continuing needs and commitment to service are regularly and jointly assessed to support and ensure high-risk families, children, and youth are engaged in services
Collaborative	Effective collaborations with appropriate child- and family-oriented services are conducted to ensure service coordination and family support system development during and after SNAP® services
Community responsive	SNAP® programs are adapted to diverse, cultural, and socioeconomic factors that characterize communities in order to be responsive to social determinants of child and family mental health
Accountable service excellence	Combination of high-quality staff development activities that include consistent supervision, training, integrity, and the attainment of accountable standards assessed through a series of well-developed research, evaluation, fidelity, and quality assurance activities fosters overall service excellence

unable to attend the SNAPP parent group, who need additional assistance/practice with parenting skills, or who need parenting support to address barriers to skill acquisition (i.e., mental health, parent cognitive restructuring, attachment).

School advocacy/teacher support – ensures that children receive the best possible education meeting their individual behavioral and learning needs. Teachers of identified clients are contacted at the start of the program to introduce the program skills and offer behavior management support if needed. Parents are supported in advocating for their children within the school system environment.

Crisis intervention – a service available to assist parents and children involved in the SNAP® program in dealing with challenging situations as they arise and/or referral to appropriate crisis services.

TAPP-C (The Arson Prevention Program for Children) – offered to children with fire interest or fire setting as a presenting problem. It involves a fire interest assessment and recommendations, a home safety visit, and education regarding fire by Toronto Fire Services.

Victim restitution – activities that encourage children to apologize to their victim, redress behaviors, and begin to learn how to take responsibility for their actions.

Homework Club/academic tutoring – provides remedial sessions for children functioning below grade level. Weekly 1 hour tutoring sessions with teachers or specially trained volunteers are held in the child's home or community.

SNAP® Youth Leadership Club – a component offered in both the boys' and girls' programs for youth who have completed the core components of the SNAP® program but continue to be high risk. Staff provide group, individual, and family work to prepare at-risk youth for self-sufficiency, increase motivation for school involvement and success, improve their workforce career trajectories, and reduce their involvement with the law.

Parent problem-solving group – a 9-week group component for parents who have completed the SNAP® parent group. The focus is on enhancement and refinement of family problem-solving skills in relation to ongoing issues the families are experiencing (i.e., keeping rewarding effective, difficulty with consequencing, school issues, media awareness) while providing continued support.

Long-term connections/continued care – families may continue to be involved in all components of SNAP® as long as there is a need and interest. In addition to previously listed components, this may also include activities such as participation as a peer or parent mentor.

Principles

Nine *principles* with specific indicators have been identified to describe the approach to service delivery and guide SNAP® programming. In addition, the *principles* ensure service and clinical excellence when organizations replicate SNAP® in their communities (sites). These are also used to assist in measuring implementation adherence, fidelity, and integrity of the SNAP® prevention and intervention programs (see Table 1).

SNAP® Theoretical Underpinnings

From the very beginning, the SNAP® model was built on well-known theoretical approaches showing promise in the early 1980s. These included social skills training, cognitive problem solving, self-control and anger management strategies, cognitive self-instruction, family management skills training, and parent training. As noted by Augimeri et al. (2011a), the model continued to evolve as SNAP® scientists and practitioners consulted with the Oregon Social Learning Center (OSLC) in Eugene, Oregon. These consultations helped to strengthen the SNAP® parenting component by adopting aspects of OSLC's Social Interactional Family Therapy's (now known as Social Interactional Learning) approach to working with families (Patterson et al. 2010). The SNAP® model programs have evolved to reflect the contributions of six core treatment theories, including Systems, Social Interactional Learning, Cognitive-Behavioral, Attachment, Feminist and Developmental Theories (see Fig. 2). It is important to note that these theories are not viewed as stand-alone entities, but as interactive in their contributions to the foundation and ongoing development of the SNAP® model programs.

The SNAP® Logic Model (see Fig. 2) also illustrates that the primary targeted outcomes include improved overall child and family functioning with an emphasis on emotion regulation, self-control, and problem solving for both parent/caregiver and child. As Strayhorn (2002) indicates, "self-control difficulties are of central importance for many psychiatric disorders. . .[it] is also a crucial, and often missing, ingredient for success in most treatment programs" (p. 7).

Research

Research on the SNAP® programs has been an integral part of the model's ongoing development since its inception and continues on an ongoing basis. Scientists and clinicians work within a collaborative process (SNAP® *principle: scientist-practitioner*) to inform and update the theoretical approaches of the model, ongoing evaluation, and program development. Rigorous internal and external evaluations of SNAP® programs (e.g., process and outcome evaluation, quasi-experimental designs, random control trials, long-term follow-up – criminal record searches, cost-benefit analyses, third-party external evaluations, and neuroscience) have consistently demonstrated positive treatment effects over time: children improve significantly more than children receiving an attention-only group, delayed treatment, or an alternative treatment with notable effect sizes (moderate to large); treatment gains are maintained at 6, 12, and 18 months; parents report less stress in their interactions with their children and increased confidence in managing their children's behavior; children report improved quality of interaction with parents, less yelling, and more limit setting; children report more positive attitudes and less anxiety and demonstrate more pro-social skills with teachers, peers, and family members. Longitudinal research analysis showed that

91.8 % of the boys and 95 % of the girls had no history of criminal offences by age 14 and approximately 68 % of the children have not had a criminal record by age 19 (Pepler et al. 2010; Augimeri et al. 2007, 2011b). Brain imaging studies conducted by the Hospital for Sick Children in Toronto and the University of Toronto showed that children who responded positively to SNAP® treatment manifest changes in brain systems responsible for cognitive control and self-regulation, and a number of SNAP® families showed an ability to "repair" after engaging in a difficult parent-child interaction (see Granic et al. 2007; Lewis et al. 2008; Woltering et al. 2011).

Designations

As a result of these promising research findings, SNAP® has achieved the highest levels of recognition from independent reviewers who rate evidence-based programs. In 2012, the US Department of Justice, Office of Justice Programs, designated SNAP® as an "effective" crime prevention model (see http://www.crimesolutions.gov/ProgramDetails.aspx?ID=231). In 2011, Public Health Agency of Canada designated SNAP® as a Canadian best practice under their Preventing Violence Stream – Canadian Best Practice Portal (see http://cbpp-pcpe.phac-aspc.gc.ca/intervention/707/view-eng.html; http://cbpp-pcpe.phac-aspc.gc.ca/intervention/706/view-eng.html. In an excerpt outlining effective options for young children with conduct problems, Cipriani (2009) highlighted the SNAP® model as the "best example" of effective early intervention strategies and discussed the program and its successes at length. In 2008, the Canadian *National Crime Prevention Centre* designated SNAP® as a "Model Program" (see www.publicsafety.gc.ca/res/cp/res/2008-pcpp-eng.aspx), in 2012 it was designated as an "Effective Program" by the US Department of Justice's OJJDP (see www.ojjdp.gov/mpg/mpgProgramDetails.aspx?ID=699) because of its robust treatment outcomes, and in 2006 it was given the highest effectiveness designation (Level 1) by the United States' Whitehouse program *Helping America's Youth*, now titled *FindYouthInfo.gov;* (see www.findyouthinfo.gov/ProgramDetails.aspx?pid=699). At the beginning of 2000, a study group on very young offenders led by Drs. David Farrington and Rolf Loeber found the SNAP® program was the "most fully developed intervention to date for child delinquents" (Howell 2001, p. 312).

Current Issues/Controversies

SNAP® Model Replication and Implementation

For a discussion of the program's replication standards and principles, licensing agreements, accreditation, and fidelity frameworks ensuring successful replication of the SNAP® model, see Augimeri et al. 2011b. In this discussion, having a team of dedicated scientists and practitioners was emphasized in order to create an effective, well-established, and recognized program, and it stressed that host organizations need commitment, support, and resources when incubating an evidence-based model program within a community setting.

Five important criteria were identified for the successful implementation and replication of SNAP®: (1) *Adherence* to the model is critical, (2) *restraint from making modifications* is essential, (3) *training and ongoing consultation* is mandatory, (4) *ongoing fidelity/integrity audits* are necessary to ensure the highest possible efficacy, and (5) *selecting the right staff* is paramount to program success.

SNAP® licensing agreements are established with the Child Development Institute. CDI's experience in disseminating SNAP® is that with adequate training and support, this model can be successfully replicated and implemented with strong fidelity in a variety of settings. SNAP® fits in the classroom, in the clinician's office, and at home. The program can be situated in a variety of diverse community settings and real-life community conditions. Currently, there are a number of successful SNAP® implementations in Canada, the United States, and Europe.

SNAP® Fidelity and Integrity Framework

For successful implementation and replication of evidence-based programs such as SNAP®, training and ongoing consultation activities between the site and its replicators are considered paramount (Augimeri et al. 2011b). As the SNAP® model incorporates a complex therapeutic approach, it requires strict documentation of the services being delivered and records of any integrity activities conducted to ensure successful replication and outcomes. As a result, there is a need to identify all the intricate elements of the various treatment components within the SNAP® model (e.g., SNAP® core groups, individual counseling/mentoring, family counseling) in order to effectively monitor if the delivery of these key elements is done correctly and skillfully, when and where necessary. The SNAP® principle, *accountable service excellence*, highlights the requirement of fidelity practices that include case file audits, consultations, adherence to group manuals, and consistency of facilitation skills. It is essential that the integral pieces of SNAP® related to long-term positive outcomes (e.g., decreased criminal activity) are delivered with appropriate timing, skill, and adherence. Ultimately, SNAP® researchers and facilitators are concerned with delivering an effective program that adequately meets all objectives that were predefined with the clients.

Key Aspects to Ensure a Successful International Implementation

Even though successful SNAP® sites have been established worldwide, we continue to recognize that there are many obstacles to successful implementation. As noted earlier, the implementation of an evidence-based model can be challenging on its own. This is especially true when it is being adopted in another country or culture. As SNAP® implementations continue to reach communities worldwide, the onus is on SNAP® developers to explore creative methods for ensuring successful replications. There are several important factors (e.g., language, culture, travel) that may need to be considered when replicating a “foreign” intervention, even though its core strategies are proven to have universal applicability (e.g., cognitive-behavioral therapy).

Future Directions

A sixth implementation criterion that would greatly contribute to not only the successful implementation of the program, but most importantly, its sustainability, would be the *adoption of community teams for children under 12* (Augimeri et al. 2001a; Goldberg et al. 1999). *Community teams* would be comprised of representatives from child welfare, school personnel, and criminal justice systems such as the police, health, and children’s mental health. The *community team* would be responsible for setting up police-community referral protocols, managing and maintaining a centralized referral intake line, conducting comprehensive risk and needs assessments/screenings, and connecting children and families to the appropriate gender-sensitive services to at-risk children and their families. The creation of a government-supported *National Advisory Working Group for Children and Youth Involved in Offending Behaviour* would act as a knowledge-based resource center (similar to CDI’s CCCO) dedicated to knowledge transfer activities that would support the dissemination of current research and services tailored to the needs of high-risk children and their families. This national working group would act as an external, unbiased body responsible for monitoring fidelity of implementation and replication, thus ensuring integrity and accountability from those engaged in assessment and/or services.

Conclusion

After numerous decades of working with young children in conflict with the law, their families, and communities, CDI and SNAP® continue to support and advocate on behalf of these “forgotten children.” CDI and SNAP® researchers and clinicians remain committed to keeping such children out of the youth justice and adult criminal systems. These vulnerable children deserve our utmost attention and help to develop to their fullest potential.

Recommended Reading and References

Augimeri LK, Koegl C (2012) Raising the bar: transforming knowledge to practice for children in conflict with the law. In: Loeber R, Welsh BC (eds), The future of criminology. Oxford University Press, New York, pp 204–210

Augimeri LK, Koegl CJ, Goldberg K (2001a) Children under 12 years who commit offenses: Canadian legal and treatment approaches. In: Loeber R, Farrington DP (eds) Child delinquents: development, intervention, and service needs. Sage Publications, Inc., Thousand Oaks, pp 405–414

Augimeri LK, Koegl CJ, Webster, Levene K (2001b) Early assessment risk list for boys: EARL-20B, version 2. Earlscourt Child and Family Centre, Toronto

Augimeri LK, Farrington DP, Koegl CJ, Day DM (2007) The under 12 outreach project: effects of a community based program for children with conduct problems. J Child Fam Stud 16:799–807. doi:10.1007/s10826-006-9126-x, Published Online, January 10, 2007

Augimeri LK, Enebrink P, Walsh M, Jiang D (2010) Gender-specific childhood risk assessment tools: early assessment risk lists for boys (EARL-20B) and girls (EARL-21G). In: Otto RK, Douglas KS (eds) Handbook of violence risk assessment. Routledge, Taylor & Francis, Oxford, pp 43–62

Augimeri LK, Walsh MM, Liddon AD, Dassinger CR (2011a) From risk identification to risk management: a comprehensive strategy for young children engaged in antisocial behavior. In: Springer DW, Roberts A (eds) Juvenile justice and delinquency. Jones & Bartlett, United States, pp 117–140

Augimeri LK, Walsh M, Slater N (2011b) Rolling out SNAP$^{®}$ an evidence-based intervention: a summary of implementation, evaluation and research. Intern J Child Youth Fam Stud 2.1:162–184

Augimeri L, Walsh M, Woods S, Jiang D (2012) Risk assessment and clinical risk management for young antisocial children: the forgotten group. Univ Psychol 11(4):1147–1156

Cipriani D (2009) Children's rights and the minimum age of criminal responsibility: a global perspective. Ashgate Publishing, Farnham

Earlscourt Child and Family Centre (2002) SNAP™ girls group manual: the girls club. Earlscourt Child and Family Centre, Toronto

Earlscourt Child and Family Centre (1990b, 2001b) SNAP™ parent group manual. Earlscourt Child and Family Centre, Toronto

Earlscourt Child and Family Centre (1990a, 2001a) SNAP™ children's group manual. Earlscourt Child and Family Centre, Toronto

Goldberg K, Augimeri LK, Koegl CJ, Webster CD (1999) Canadian children under 12 committing offences: legal and treatment approaches. Earlscourt Child and Family Centre, Toronto

Granic I, O'Hara A, Pepler D, Lewis M (2007) A dynamic system analysis of parent–child changes associated with successful "real-world" interventions for aggressive children. J Abnorm Child Psychol 35(5):845–857. doi:10.1007/s10802-007-9133-4, Printed On-Line

Howell JC (2001) Juvenile justice programs and strategies. In: Loeber R, Farrington DP (eds) Child delinquents: development, interventions, and service needs. Sage Publication, Thousand Oaks, pp 305–322

Howell JC (2003) Preventing & reducing juvenile delinquency: a comprehensive framework. Sage, Thousand Oaks

Koegl CJ, Augimeri LK, Ferrante P, Walsh M, Slater N (2008) A Canadian programme for child delinquents. In: Loeber R, Slot NW, van der Laan P, Hoeve M (eds) Tomorrow's criminals: the development of child delinquency and effective interventions. Ashgate, Aldershot, pp 285–300

Levene K (1998) SNAPP stop-now-and-plan-parenting: parenting children with behaviour problems. Earlscourt Child and Family Centre, Toronto

Levene KS, Augimeri LK, Pepler DJ, Walsh MM, Koegl CJ, Webster CD (2001) Early assessment risk list for girls: EARL-21G, version 1, consultation edition. Child Development Institute, Toronto

Lewis MD, Granic I, Lamm C, Zelazo PD, Stieben J, Todd RM, Moadab I, Pepler D (2008) Changes in the neural bases of emotion regulation associate with clinical improvement in children with behaviour problems. Dev Psychopathol 20:913–939

Loeber R, Farrington DP (2001) Child delinquents: development, intervention, and service needs. Sage, Thousand Oaks

Loeber R, Slot NW, van der Laan P, Hoeve M (2008) Tomorrow's criminals: the development of child delinquency and effective interventions. Ashgate Press, Burlington

Moffitt TE, Arseneault L, Belsky D, Dickson N, Hancox RJ, Harrington H, Houts R, Poulton R, Roberts BW, Ross S, Sears MR, Thomson WM, Caspi A (2011) A gradient of childhood self-control predicts health, wealth, and public safety. Proc Natl Acad Sci (PNAS) 108(7):2693–2698

Patterson GR, Forgatch MS, DeGarmo DS (2010) Cascading effects following intervention. Dev Psychopathol 22:949–970

Pepler DJ, Walsh MM, Levene KS (2004) Interventions for aggressive girls: tailoring and measuring the fit. In: Moretti MM, Odgers CL, Jackson MA (eds) Girls and aggression: contributing factors and intervening principles. Kluwer Academic/Plenum, New York, pp 41–56, Perspectives in Law and Psychology Series, Volume 19

Pepler DJ, Walsh MM, Yuile A, Levene K, Vaughan A, Webber J (2010) Bridging the gender gap: interventions with aggressive girls and their parents. Prev Sci 11:229–238

Piquero AR, Jennings WG, Farrington DP (2010) Self-control interventions for children under age 10 for

improving self-control and delinquency and problem behaviors. Campbell Systematic Reviews 2010 2. doi:10.4073/csr.2010.2
Strayhorn JM (2002) Self-control: theory and research. J Am Acad Child Adolesc Psychiatry 41(1):7–16
Walsh MM, Augimeri LK (2009) SNAP® fidelity rating index. Child Development Institute, Toronto
Walsh M, Pepler D, Levene K (2002) A model intervention for girls with disruptive behavior problems: the earlscourt girls connection. Can J Couns 36:297–311
Woltering S, Granic I, Lamm C, Lewis MD (2011) Changes in the neural correlates of emotion regulation with treatment. Biol Psychiatry 70(9):873–879

Strategies of Policing Terrorism

Simon Perry
Institute of Criminology, Faculty of Law, The Hebrew University, Jerusalem, Israel

Synonyms

Counterterrorism strategies; Internal security strategies

Overview

In the last decade terrorism has become a prominent topic in many areas, including criminology. Since terrorism affects almost every aspect of life, it has become a central priority for the police in Western democracies. American police and even countries that had been prepared to fight terrorism prior to 9/11, such as Great Britain and Germany, began after that date to review their readiness and rethink the role of police in counterterrorism (Weisburd et al. 2009a; Bayley and Weisburd 2009; International Association of Chiefs of Police 2005; Bamford 2004). Howard (2004) argues that police departments should start thinking of themselves as proactive valuable assets in deterring, defeating, and recovering from terrorist attacks. Law enforcement, intelligence, and security agencies are expected to team up, join forces, and work together with other organizations to uncover terrorist networks, foil terrorist attacks, respond to suspicious situations, and serve as first responders (Weisburd et al. 2009a). Yet there is a lack of evidence-based models for this new role of policing terrorism. A Campbell Collaboration (Lum et al. 2006) systematic review of strategies to combat terrorism could only identify seven studies that met minimal methodological requirements. None of these seven studies examined a police intervention. In fact, to date there are only a few descriptions of possible models for strategic and tactical activities of policing terror. Therefore, little is known about what the best antiterrorism strategies and tactics are. Furthermore little is known about how models can be systematically measured and assessed for their effectiveness (Weisburd et al. 2009a).

This entry will attempt to summarize recent developments in the field of counterterrorism. It will introduce and portray principal strategies, tactics, and practices that are presented in the literature and/or that have been widely adopted by practitioners in policing terrorism.

Why Police Bear Primary Responsibility for the Terrorism Threat in Democratic Countries?

Before discussing what these counterterrorism strategies, tactics, and practices are, four principal themes that naturally surface will be discussed: (a) the characteristics of the terrorism threat phenomenon, (b) the complexity of developing and evaluating a counterterrorism model, (c) the tension between preserving democratic principles and counterterrorism, and (d) why the police (according to Bayley and Weisburd 2009) bear primary responsibility for preserving public security in most countries.

The Terrorism Threat

In order to address the threat and develop an effective response to terrorism, one should first define this criminological phenomenon and its goals (Weisburd et al. 2009b). According to

Ganor (2009), even after the world recognized the scale of the terrorist threat after 9/11, there is still no agreement on a definition of terrorism. After studying various definitions of terrorism (Hoffman 2006; Ganor 2009; Martin 2009; Hasisi et al. 2009) in the context of seeking to develop an effective response to terrorism, one can conclude that the phenomenon of terrorism includes the following basic elements: (a) the creation of *fear and anxiety*; (b) doing so through the use, or threat of use, of *violence*; (c) non-state actors; (d) the goal of weakening the fabric and the *resilience of society*; (e) inflicting *random casualties* among the *general population*; (f) increasing the *frequency of attacks*; (g) targeting the *weakest and most vulnerable* members of society; (h) taking advantage of the *media*; and (i) all in order to pressure decision makers to *accept terrorist demands*.

The Complexity of Developing Strategic Policing-Terrorism Models

The lack of evidence-based models in the policing-terrorism field, as well as a lack of systematically evaluated strategic and tactical activities, is a result of two main reasons. First and foremost, law enforcement intelligence and security agencies are very reluctant to cooperate with such research, as they fear the possible ramifications including the exposure and compromise of counterterrorism methods, tools, sources, and tactics. Secondly, this type of research encounters difficulty in the measurement of success and the determination of cause and effect. These difficulties stem from the presence of other variables in the tested model, the effect of which is to make creating a control situation extremely difficult.

Preserving Democratic Principles and Human Rights in Counterterrorism

In counterterrorism the democratic liberal values of the society limit the capability of the state to make use of the full potential means and technological advantages that the state has. In its struggle against terrorism, the democratic state is obliged to select measures and utilize capabilities that will cause minimum damage to human rights. Innes (2006) argues that collecting and using intelligence is part of the "dirty work" of a democracy. Usage of superior capabilities harming those who have no connection to terrorism, as well as harming fundamental moral principles that are essential for the democratic state, would constitute a victory for the terrorist organization. And such misuse could alienate parts of society, playing into the terrorists' hands (Ganor 2009).

Bayley and Weisburd (2009) argue that legitimacy is the foundation of successful policing, whether related to terrorism or to crime in general. Even though protection from terrorism is a moral cause, it can lead to the violation of human rights, causing officers to approach individuals as suspects rather than as individuals who deserve service. Losing police legitimacy endangers the advantages of public cooperation that are greatly needed in the war on terror.

This is especially significant in minority communities that are linked ethnically or nationally to terrorist groups. For example, in the Israeli context, this applies to familial and national ties of Israeli Arabs to Palestinians in surrounding countries and in the Palestinian-controlled territories (Hasisi et al. 2009).

Why Should the Police Lead the Response to Terrorism?

There are five main reasons why the police force has been chosen to lead the response to terrorism in most countries:

1. Since *terrorists are criminals, and terrorism is a type of crime* in all essential characteristics (Clarke and Newman 2006), it is appropriate that, in democracies, the police force, which is best qualified to deal with criminals and crime, should lead the response to terrorism. **The criminal-justice model** for counterterrorism features terrorism as a crime and portrays the terrorist as violent criminal who should be arrested and punished according to the rule of law by the police and the criminal-justice system (Greene and Herzog 2009). According to Perliger, Hasisi, and Pedahzur (2009), most of the literature discussants on

democracy and counterterrorism believe that **the criminal-justice model** permits dealing with terrorism without seriously undermining the legal and moral foundations of the democratic system. Therefore, it is seen to be better for democratic countries to leave counterterrorism in the hands of the police which operate in the civilian arena. By contrast, **the war model** (Greene and Herzog 2009) features terrorism as an act of war that disputes and threatens the well-being of the state and the political system. Therefore, in this model, it is believed that the terrorist and the terrorist organization should be eliminated by the use of intelligence and military forces. However, employment of **the war model** paradigm would lead to situations in which military forces conduct combat warfare within their own territory, constituting a severe undermining of human rights and morality of the democratic state and its legal system.

2. *Many of the methods and means necessary to fight terrorism are used routinely by the police* in their daily operations. These include investigation and information and evidence collection; forensics (identification of weapons, explosives, victims, etc.); police-operated call centers which are first responders; and police liaison with the private sector, including the issuance of licensing to businesses. Moreover, many activities associated with the handling of terror attacks, whether before, during, or after such an attack, are an integral part of routine police activity (traffic control, managing crime scenes, and maintaining or restoring public order).
3. *The connections between terrorists and other criminals give the police an advantage.* Criminals facilitate terrorism with many necessary means such as weapons and explosives; documentation; vehicles; collecting, transferring, and laundering money; information; communications; technology; and even operational subcontracts for specific missions. Terrorist organizations, in order to finance their activities, have used classic organized-crime illegal activities such as money counterfeiting and the smuggling of drugs, counterfeit goods, and taxable merchandise like cigarettes. This connection gives the police, who have a familiarity with the criminal world, an edge in counterterrorism.
4. *Police are responsive to irregularities in the environments in which they operate routinely.* During their everyday policing of crime and disorder, the police can be on the lookout for suspicious indicators connected to terrorism activities in their communities (Innes 2006). This puts the police in an exceptional position to collect information (that remote intelligence agents generally cannot obtain, since they are not connected to the community) and develop critical counterterrorism intelligence that can be disseminated in real time to other law enforcement agencies (Connors and Pellegrini 2005). Perliger, Hasisi, and Pedahzur (2009) argues that it is better to leave the struggle against terrorism in the hands of the police as part of a more comprehensive model – the criminal-justice model – which treats terrorism as violent criminal action in the civilian arena.
5. *Legal authority to perform policing procedures*. In many democracies only the police have legal authority to perform policing procedures that are required in counterterrorism. The military, for example, in the USA, Canada, UK, and Israel, is not legally empowered to engage in ordinary police functions within the borders of the country. This situation is unlike many European countries (e.g., France, Spain, and Italy) where the military play an active policing role (Weisburd et al. 2009a).

Defusing Terrorist *Motivations* Versus Preventing Terrorist *Opportunities*

Based on the “rational choice” theory (Clarke and Cornish 2001), crime is mainly a result of a ***process*** in which the individual evaluates ***opportunities*** weighing the anticipated benefits against the expected costs of the behavior, in the context of a certain time and place. Terrorism, like any other criminal behavior, is an outcome of two necessary conditions: ***motivation*** to commit the terrorist behavior and ***opportunity*** to carry out a terror attack at a particular occasion (Clarke and Newman 2006). Therefore, any counterterrorism strategy should defuse *terrorist motivations* or prevent *terrorist opportunities* or do both.

The strategy that tries to prevent the terrorist from attacking by defusing the *terrorist motivation* is known as the **conciliatory model**. According to this model, terrorism is a political problem, and it should be given a political solution that will address the root cause of terrorism by the action of policymakers, brokers, and diplomats (Greene and Herzog 2009). However, Clarke and Newman (2006) argue that counterterrorism "must not rely on changing the heart and minds of terrorists. The motivation for terrorism results from long-term social, cultural and psychological pressures, which are difficult to alter" (Clarke and Newman 2006:11).

According to Weisburd and Waring (2001), the efficient approach for prevention of crime involves decreasing the ***opportunities*** (the core of the situational prevention approach presented by Felson and Clarke 1998) present in certain situations which encourage commission of a crime. Clarke and Newman (2006) argue also that in counterterrorism it is easier to reduce ***opportunity*** than to reduce terrorist ***motivation***. This is especially true in the case of suicide bombers that are very hard to deter. Therefore, the efficient approach to prevent terrorism is to adopt strategies that attempt to reduce the opportunities to commit terror attacks. In any case, they claim, easy opportunities encourage terrorists to attack.

In short, defusing terrorist motivations is mainly not a law enforcement intelligence operational task and seems to be less efficient in preventing terrorism than the approach of reducing opportunities. **Therefore, most counterterrorism strategies concentrate on attempting to reduce the *opportunities* rather than reducing *motivation*.**

Reducing opportunities could be achieved by the ***defensive model*** that aims to protect potential targets and victims from attacks, through "target hardening" (Clarke and Newman 2006), or by ***proactively preventing operational capabilities*** of terrorist organizations and individuals that pose the physical and psychological threat. According to Clarke and Newman (2006 p. 9), there are four "pillars of terrorism opportunity": targets, weapons, tools, and facilitating conditions.

The Responsive Approach and the Proactive Approach

Law enforcement practices have two main models of operation: the responsive approach and the proactive approach. In the responsive approach, after a crime or specifically a terrorist act has occurred, law enforcement responds by gathering intelligence, identifying perpetrators, collecting evidence, and arresting suspects. This responsive approach calls for defensive methods that could perhaps foil an attack that has been launched, or at least minimize the damage. Therefore, in counterterrorism we also refer to it as a responsive defensive approach. This approach better protects vulnerable places and effectively restores order in the event of a terrorist attack (Weisburd et al. 2009b).

The proactive model attempts to foil the crime **before** it is launched. This approach has been used for decades by law enforcement against organized and other forms of crime, especially drug trafficking. In this approach law enforcement intelligence identifies potential criminals/terrorists, gathers intelligence and evidence, makes the arrest, and foils the crimes/attacks before they can occur. The main goal of this approach, which can be classified as "high policing," utilizes covert intelligence gathering, surveillance, and foiling operational tools (Bayley and Weisburd 2009; Brodeur and Dupeyron 1993).

Terrorism seeks to create fear and anxiety, weaken the fabric and the resilience of society, and destabilize a social order, all by increasing the frequency of attacks and inflicting mass casualties. The objective of law enforcement is therefore to reduce the operational capabilities of terrorist organizations through counterterrorism measures which can be characterized as proactive, preventive, and offensive in nature. This goal is a crucial part of every democratic country's duty to protect its citizens (Ganor 2009).

The proactive method is intended to deter, disrupt, and prevent terrorist activities (Weisburd et al. 2009b; Hasisi et al. 2009). This counterterrorism strategy damages terrorist organizations physically, psychologically, and economically and hurts their operational infrastructures. It pushes them into a defensive mode, where they spend a great deal of time and resources concealing their activity, which thus limits their effectiveness and reduces their ability to carry out attacks. According to Hasisi et al. (2009), this proactive strategy is becoming more complex as a result of the ever-changing nature of terrorism. In the contemporary age of globalization, terrorist groups tend to be stateless, making them a more elusive target that readily exploits open societies and advanced communication technologies, while evading surveillance by traditional security agencies.

The ***responsive approach***, which is defensive in nature, and the ***proactive approach***, which is preventive and offensive in nature, complement rather than contradict one other.

Since, unfortunately, it is unrealistic to expect to entirely prevent all terrorist attacks, a good responsive approach is important to have in place alongside an efficient proactive approach. This combined approach both apprehends terrorists and damages terrorist organizations. By combining both approaches, counterterrorism forces can prevent or at least significantly minimize terrorist ***opportunities*** to attack before the attacks have been launched, ***responsively*** foil the attacks once they have been launched, and effectively minimize the extent of the casualties and damage, as well as restore order and evacuate the scene during and after the attack. Such a comprehensive method of counterterrorism makes it possible for the general public to maintain their everyday routines and preserves their morale and sense of personal security.

In the United States the view of local law enforcement officers as primarily "first responders" is slowly changing, and they are becoming viewed also as "first preventers" of terrorism. In New York City, the NYPD adopted this concept of "prevention" that has been used so successfully against crime, and adapted it for the war on terror (Bratton and Kelling 2006). Connors and Pellegrini (2005) claim that local officials and police in the United States must be prepared to take the lead in the war on terror and not to wait for direction from federal agencies hundreds of miles away, if they want to prevent or recover from any future terrorist attacks.

According to Innes (2006), in the United Kingdom as early as the 1990s, a new intelligence-led policing improved police agencies' effectiveness by proactively identifying problems and then targeting individuals who were most likely to cause harm to the general public. This change was aimed at pursuing a more proactive mode of working based on the principles of risk management. Given that the police have limited resources and developing effective live informants ("HUMINT") is difficult, a network of community intelligence contacts was established to provide fairly effective surveillance over potentially dangerous groups and individuals. Expert community engagement units were set up to develop and maintain this "community intelligence feed" that would serve both anti-criminal and anti-terrorist purposes.

The Israel National Police (INP) has a long history of experience struggling with terrorism and is considered highly efficient and professional in its approach to counterterrorism (Weisburd et al. 2009b). The Israeli model for policing terrorism applies this principal strategy of combining the ***proactive offensive*** and ***responsive defensive*** approaches. It is implemented in three circles of activity: (a) ***sources of terrorism circle*** – proactive early prevention, interdiction, and treatment of the sources of terrorism in order to foil terrorist attacks before they are launched; (b) ***the attack route*** – response activities once the attack has been launched, to foil terrorist attacks before they reach the target; and (c) ***the terrorist targets*** – defending and "hardening" potential targets before any attack and response activities at the scene during and after an attack.

In the Israeli antiterrorism model, there is a great deal of emphasis on the proactive model approach, since it enables capturing terrorists, decreases the frequency and severity of attacks, and results in fewer casualties. A proactive model is preferable since it has a greater counterterrorism effect, preventing terrorists from achieving their goals and making it possible for the general public to maintain their everyday routines.

Proactive Offensive Early Interception Counterterrorism

As discussed above, the offensive-proactive approach is the most effective counterterrorism strategy and is the key for preventing attacks before they are launched. This strategy identifies and responds to terrorist threats before they actualize, by developing intelligence, establishing operational capabilities, and uprooting terrorists and their infrastructure (Weisburd et al. 2009b). The proactive approach is based on three crucial competencies: producing ***quality intelligence***, building ***operational capability***, and creating a ***hostile environment for terrorists*'** ***operations***.

Effective counterterror thwarting operations rely on the ability to gather reliable information about the terrorists' capabilities, intentions, and specific plans. Quality intelligence takes away from the terrorist the element of surprise, which is a critical advantage for a terrorist attack. Intelligence allows security forces to efficiently protect in advance targets that may be selected by terrorists. Furthermore, intelligence puts the element of surprise in the hands of law enforcement.

Once a terrorist attack has already been set in motion, security forces utilize intelligence both for proactive offensive thwarting operations and for responsive defense measures (Perliger et al. 2009).

The Intelligence Process

Producing intelligence is a circular flow process which constantly aims to portray an accurate picture about the threats. The intelligence analyst defines (based on the initial information) the *information gaps*, which is the information he needs but does not have in order to portray a reliable *intelligence picture*. The *intelligence picture* is composed of all the information about subjects that are connected significantly to the different threats, which are called *topics of interest*, and all the information about individuals who are connected significantly and compose the threat; they are called *targets*.

In order to fill in the *intelligence gaps*, an *intelligence collection and coverage plan* is prepared. The plan includes three levels of coverage: the abovementioned *topic of interest* and *targets*, as well as the *territorial coverage*, which is the entire information about a specific area (neighborhood or town) including who is doing what, when, and where, and who knows about it. At the first stage, potential sources of information are identified in order to cover all three levels. The next stage is the preparation of the *recruitment plan*, which is the methods of recruiting and running live and technical sources. The various sources utilized for coverage are as follows: live informants ("HUMINT") and echnical sources (called "SIGINT") such as wiretaps, surveillance activities, investigations and debriefings, archives, and databases; public open information such as the news media and the Internet; and fellow agencies from the same country or international. The role of the analysts is to prepare and present an *integrative intelligence picture*. Their job is an unending process, since they are the "nerve center" of information: they direct the collection and thwarting of intelligence according to the requirements of the command guidelines, they assist in the preparation of an operational perception, they support the development of "HUMINT" sources, they acquire and deploy intelligence capabilities and tools, and they administer the Intelligence Database.

Unlike military intelligence, which tends to rely on advanced technological monitoring and surveillance capabilities, HUMINT sources are considered more effective for penetrating terrorist groups. Police intelligence generally deals with criminal organizations which are small active groups from within the civilian setting and whose structure and character resemble terrorist groups (Perliger et al. 2009).

Intelligence Sharing and International Cooperation

In many cases the terrorism threat is both national and international. Therefore, cooperation and quality intelligence sharing are essential for law enforcement and intelligence organizations to prevent attacks. Kelling and Bratton (2006) emphasize the importance of collecting and sharing intelligence in the fight against terrorism.

The mutual interrelation between terrorist organizations and the classic hardcore criminal organizations enables the police to extend the well-established national and international cooperation on fighting organized crime to the foiling of terrorism.

Special Operations Capabilities

The mere collection of counterterrorism intelligence is not sufficient, since intelligence without the ability to reach a target or prevent an attack is of little value. These tasks are performed by special operations units which are trained to enter locations where terrorist actions are being planned and prepared. The goal is to foil the attacks before they are launched and to threaten the terrorists' own sense of security, keeping them busy and on the run. Such units should be able to conduct undercover operations in which they reach their target and arrange an arrest without being detected. The police should have specialized elite counterterrorism units, trained to handle very specific terrorist situations, such as releasing hostages and carrying out special operations using small disciplined teams highly trained in commando style military operations.

Bratton and Kelling (2006) argue that the police need special training and a new mindset so they can proactively assault terrorism. Police departments from all over the world are already exchanging information with other countries as the war against terror is gradually becoming a global threat. The government of Israel, for example, has welcomed police forces from all over the United States for training and exchange visits.

Proactively Changing Terrorists' Operating Environment

Bratton and Kelling (2006) explain that the theory of "broken windows" is used to constantly create a hostile environment to potential criminals, building up the uncomfortable sentiment that they are the ones who are in danger. In the same way police officers should create a hostile environment, mainly through the potential terrorist support structures, in which terrorists will feel uncomfortable. Constantly changing the terrorists' operating environment by establishing "bottleneck passage" creates a hostile environment while taking away the element of surprise from the terrorist. Bottleneck passage, such as obstacles and barriers, compels the terrorist to take counteractions, involving other coconspirators by using communication channels. These communication channels leave "intelligence footprints," increasing the opportunities for intelligence collection, by both human and signal technical means.

Defensive Response

The ***defensive response*** to attacks (that were not foiled by the offensive-proactive activity) should be dealt with in two circles of activities: (a) response activities once the attack has been launched and it is on its way to the target and (b) response activities before, during, and after an attack has occurred on a terrorist targeted site.

Response Activities Foiling the Attack on Its Way to the Target

Once an attack is ongoing and the terrorist is on the route to the target, there are a series of possible tactics to delay the terrorist's movement, including setting up road blocks, creating traffic jams, and closing specific public facilities or streets. There are two main goals for creating such obstacles. The first objective is to slow down the terrorist as much as possible, thus delaying the attack. This gives the police more time to bring special operations units to engage with the terrorist on route and at the same time to organize a better defense at potential terrorist targets. The second objective of the obstacles is (as mentioned above) to compel the terrorist to establish communication channels that increase the opportunities for "HUMINT" and "SIGINT" intelligence collection. The enhanced intelligence and better police deployment increases the possibility for interdiction before the attack takes place.

Once there is a specific threat that terrorists have got through the security and are out of control, the police should consider making an announcement though the media, informing the public to stay away from crowded places.

Defensive Activities at a Terrorist's Potential Targeted Site Before an Attack

The number of potential target is endless, and police resources are limited. Therefore, the police need to conduct an efficient defensive effort, using risk assessment in order to build an effective protection plan. Vulnerability and risk analysis must then drive operational responses in order to create a truly effective policing apparatus against terrorism (Connors and Pellegrini 2005). The plan needs to prioritize the allocation of defensive tools in order to harden vulnerable potential targets. According to Davis et al. (2004) after 9/11 three-fourths of the police departments in the United States conducted risk assessment compared to only one-fourth that performed such analysis before 9/11. Private security firms are the police's main partner in "target hardening". In this partnership the police should train, provide necessary information, and supervise so that the private security firms can better protect their clients' facilities. In Israel, all malls, shopping areas, restaurants, hospitals, office buildings, or any other public facilities have private security guards who check customers entering the facility and conduct other security-related activities. The police, prior to approving the business licenses for a facility's operation, review the business facilities and their security procedures. Additionally the police perform periodic security checks, in which facilities that do not hold to the security standards set by the police may be forced to shut down through issuance of a court order (Weisburd et al. 2009a).

Realistically, not all targets can be rendered attack proof; thus, the protection plan should aim to minimize the damage in case terrorists are able to attack. To reduce the number of victims and amount of damage in terrorist attacks that were not foiled, the police work with the private security agencies to prevent terrorists from accessing public facilities. The underlying principle here is that an explosion that occurs within a closed environment (especially if crowded) will result in a much more destructive outcome than if the same explosion took place outdoors, hopefully away from the large crowd.

An important element of target hardening is educating the public and improving routine security preparedness. The police should play a central role in educating the public from an early age to be aware of indications of possible terrorism events and to provide information by reporting their suspicions to the police. In Israel, for example, police visit elementary schools, teaching the children to be attentive towards suspicious people and objects and to report to grown-ups or, if possible, to a policeman. Despite the fact that most security calls to the police are false alarms, the police treat each and every call as if it were an actual explosive device or other security threat. It is important to do so because of the potential damage from every terrorist attack and to demonstrate that the police are responsive, so that the public will continue calling (Weisburd et al. 2009b). Such responsiveness increases public confidence in the police, when in the field of counterterrorism, they are viewed by the public as responding efficiently.

These approaches are also deployed in the United Kingdom, where the police invest a great deal of attention in developing public awareness of suspicious behavior (suspicious short-term tenants, suspicious people who have bought or rented a car, etc.). The police in the United Kingdom develop suitable reporting mechanisms, alongside creating working relationships with the business community to protect businesses from potential threats and to provide guidance on appropriate security measures (Howard 2004).

Defensive Activities on a Terrorist Targeted Site During and After an Attack

The police, being the first responders, are naturally expected to act in response to a terrorist attack in such a way as to effectively manage the crisis. The Israel National Police (INP), which has tremendous experience managing many scenes of terror attacks over the years in Israel, has developed effective practices and protocols. Through the whole process, the highest territorial police commander who is on the scene is responsible for all activities that take place throughout the procedure, until he has been released by his commander or when the scene is cleared. The overall responsibility is never divided and it is always clear. During the entire process, all organizations that are involved are subordinated to the police commander, including the medics, the firefighters, and even the employees of the local city council who later clean up the area. The first mission is to secure the scene from secondary explosive devices or additional terrorists, to prevent a tactic that has been used by terrorists to hit the first responders – the policemen and the medics. Therefore, the first on the scene are the bomb-squad technicians, who close down and search the inner scene. At this stage, along with the bomb-squad technicians, only the medics who treat and evacuate the most critically injured are allowed on the scene. Only after the inner scene has been secured are the rest of injured treated and evacuated. Forensic units follow to identify the dead and – with a special unit – evacuate the bodies; collect intelligence and try to identify the type of explosive device or weapons used, for the purpose of linking it to a specific terrorist organization or laboratory; and collect evidence with the criminal investigators. At the same time traffic and patrol officers set up road blocks, closing down the outer ring and directing traffic, clearing the way for the ambulances, searching for accomplices who might have assisted the attacker and are trying to get away, and controlling crowds. Intelligence units collect information that may assist in identifying the source of the explosives and the people responsible

As mentioned, in order to defeat the goals of terrorism, the main object of the police in counterterrorism is to preserve and strengthen the resilience of the population, enabling the public to continue with their daily life routine. It is expected that a rapid clean-up of the terrorist scene will minimize the psychological effect of the attack. Therefore, timing is critical, both for psychological reasons and also for forensic reasons – to assure the collection of evidence before the scene is contaminated. In the Israeli model, all of these activities of clearing terrorist scenes are expected to be accomplished in no more than 4 h. Another important element in preserving public resilience and reducing distress and fear is keeping the important communication channels open to the public thru the media, in which the territorial commander (or deputy or spokesperson) reports calmly and informatively during and after the attack event to news agencies. The ongoing briefing prevents damaging rumors, giving the public the safe feeling that things are under control by providing such information as the description of the event, the areas or roads that have been shut down, and alternative routes (Weisburd et al. 2009a).

Organizational Elements in Policing Terrorism

A Clear Division of Authority and Responsibility Alongside Vital Cooperation and Partnership

Some countries have a highly centralized police organization at the national level, with a clearly distinguished purpose and responsibility related to crime, terrorism, and public order. However, even in such a setting, the police force has partners in the counterterrorism effort. Furthermore, any police force is composed of different units, so that harmonized cooperation doesn't always come naturally. In the chaotic reality of a terrorist attack, it is fundamental to have an unambiguous division of authority and responsibility that clearly directs who is the one person in charge at a certain time and place and who bears responsibility and accountability. Vagueness about who is running the show leads to confusion, lack of necessary decision making, and failure that will end with unnecessary victims and destruction.

For example, Israel has only one national police force (Israel National Police, INP), and it is highly hierarchical and centralized, with unique centralized operational combat special units. Since intelligence gathering is also centralized, it makes the processes of detecting, identifying, deterring, and thwarting potential terrorist attacks more efficient (Greene and Herzog 2009). There are close and ongoing formal and informal contacts between the police and the Internal Security Agency (ISA) (Weisburd et al. 2009a). This intimate cooperation between the INP, which has overall responsibility for internal security, and the ISA, which is the main source of counterterrorism intelligence, was not a trivial matter to achieve. It required tremendous effort to build trusting relationships at all operational and command levels. Such a relationship makes it possible in a matter of minutes to turn critical ISA information into an INP foiling operation. More broadly, the close connection between the INP and the ISA facilitates the exchange of intelligence while enabling consistency in both offensive and defensive counterterrorist operations (Hasisi et al. 2009).

Countries that do not have a centralized police system have to synchronize their counterterrorism activity on the national level as well as the local level. For example, in the United States before 9/11, there was a lack of intelligence coordination that has been strongly criticized (Weisburd et al. 2009b). Currently local, state, and federal law enforcement agencies exchange information on initiatives, as well as creating centralized special officers and units, including special response teams.

According to Bratton and Kelling (2006), intelligence-led policing is making its mark in the USA, and major efforts are being made to restructure police capabilities for an increasingly proactive intelligence gathering and analysis apparatus. Instead of relying on the federal government for intelligence, many state and local departments are now creating their own systems. Among other actions, they are assembling databases and sharing information. This very important development requires highly sophisticated coordination, especially in such a large country as the USA that has around 17,000 police organizations on the federal, state, and local levels. These organizations are adopting a set of national strategic guidelines that clearly define the division of authority and responsibility, setting vital cooperation and partnership procedures.

Recruitment, Training, Weapons, Equipment, and Relevant Exercises

Since counterterrorism is a major responsibility of the police, police officers need to be trained and equipped to confront the terrorism threat. Kelling and Bratton (2006) argue that counterterrorism has to be woven into the working protocols of every police department, so that it becomes part of the everyday thinking of the officers who are on the street.

All officers, including those whose main role within the police is *not* counterterrorism, should undergo basic counterterrorism training. The training should prepare officers for an unexpected encounter with a terrorist incident. It should focus on acquiring first response skills such as isolating the site of a terrorist attack and effectively handling the personal weapon. Fast response teams that would defuse terrorist situations and bring them to an end as soon as possible should be put together and trained. The goal of such teams is to be prepared for fast intervention to minimize the damage, and to control and contain the scene until the attack is defused or the special counterterrorism unit takes over. Response time is crucial in containing a terrorist attack; therefore, the fast response teams should have sufficient training (such as urban warfare), relevant exercises, equipment, and suitable transportation, such as motorcycles (Weisburd et al. 2009a).

Shortcomings of Policing Terrorism

The leading role of police in counterterrorism elevates new problems and dilemmas for the police in democratic countries. On the one hand, counterterrorism is a natural extension of the classic police duties, as previously mentioned. On the other hand, it is clear that terrorism may call for new responsibilities that emphasize “high policing,” which is characterized by its focus on strategic issues at a macro level, rather than local crime and disorder problems (Bayley and Weisburd 2009; Weisburd et al. 2009b). High policing emphasizes controlling rather than servicing the public, a standpoint very different from the community policing ideas that have reinforced community-police relationships. Counterterrorism functions of the police are likely to clash with the goals of creating closer police-community relations, especially with minorities, since it is difficult to be “officer friendly” and at the same time collect intelligence on suspects that are part of the community or related to it (Weisburd et al. 2009b).

Summary

Counterterrorism is geared to making it possible for the general public to maintain their everyday routines, thus preserving their morale, feelings of personal security, and resilience. This entry has introduced, described, and explained the following ***principal strategies***, ***tactics***, ***and practices*** that are presented in the literature and that are adopted by practitioners in policing terrorism:

1. Most of the counterterrorism strategies concentrate on attempting to reduce the ***opportunities*** for terrorist attack, since it is easier than to reduce terrorist ***motivation***.
2. Efficient counterterrorism combines ***proactive offensive*** and ***responsive defensive*** approaches in order to reduce terrorist ***opportunities***.
3. ***Proactive offensive*** counterterrorism identifies and responds to terrorist threats before the terrorists have an ***opportunity*** to attack.
4. The first crucial component of the ***proactive offensive*** approach is ***quality intelligence*** that is an outcome of a circular flow process. This process aims to constantly portray an accurate picture about the threats, turning information into timely action-operational intelligence.
5. The second crucial component of the ***proactive offense*** is establishing ***operational capabilities*** that have accessibility to uproot terrorists and their infrastructure, pushing them into a defensive and ineffective mode by keeping them busy and on the run, thus foiling attacks before they are launched.
6. The third crucial component of the ***proactive offense*** is constantly changing terrorists’ operating environment by establishing bottleneck passage such as obstacles and barriers, causing an intelligence footprint. Bottleneck passage compels the terrorist to take counteractions, involving other coconspirators and using communication channels. These communication channels boost the opportunities for intelligence collection, both human and signal technical.
7. Since in many cases the terrorism threat is both national and international, cooperation and quality intelligence sharing are essential for law enforcement and intelligence organizations to prevent attacks.
8. The ***responsive defensive*** model enables the general public to maintain their everyday routines, strengthening their resilience and preserving their morale and feelings of personal security. The model includes “target hardening,” foiling launched attacks, and effectively restoring order during, and efficiently evacuating the scene rapidly after, the attack.

9. In order to foil launched attacks, there are a series of possible tactics to delay the terrorist's movement, compelling the terrorist to establish communication channels. This allows enhanced intelligence gathering, potentially allowing special operations units to engage with the terrorist on route, as well as better police defensive deployment at potential targets, increasing the possibility for interdiction before the attack takes place.
10. In order to build an effective protection and "target hardening" plan, a vulnerability and risk analysis must be part of the police apparatus. Private security firms are the police force's main partner, to whom the police should provide necessary information and supervision so that the private security firms can better protect their clients' facilities and minimize the damage in case a terrorist was able to attack.
11. The police should educate the public from an early age to be aware and to report any possible threat, which should always be treated by the police as if it was a genuine attack.
12. It is fundamental to have clear division of authority and responsibility that clearly directs who is the one person in charge at a certain time and place in the event of an attack and who bears responsibility and accountability. Vagueness leads to confusion, lack of necessary decision making, and failure that will end with unnecessary victims and destruction.
13. At an attack scene the first mission is to secure the scene from secondary explosive devices and to treat and evacuate only the critically injured. After the scene has been secured, the rest of the injured are treated and evacuated. To minimize the psychological effect of the attack, the forensic units follow, to effectively and rapidly identify and evacuate the dead, and collect evidence and intelligence, allowing a quick clean-up of the scene.
14. The police commander needs to keep the important communication channels open to the public thru the media, reporting calmly and informatively during and after the attack in order to reduce distress and fear.
15. Countries that do not have a centralized police system need to synchronize their counterterrorism activity on the national and local levels, by adopting a set of national strategic guidelines that clearly defines the division of authority and responsibility, setting vital cooperation and partnership procedures.
16. All police officers should undergo basic counterterrorism training that should prepare officers for an unexpected encounter with a terrorist incident. Response time is crucial in containing a terrorist attack; therefore, the fast response teams should have sufficient training, relevant exercises, equipment, and suitable transportation.

Now that the policing-terrorism models have been described, the key question that remains is "How effective are these ***principal strategies***, ***tactics***, ***and best practices*** of 'policing terror'?" As Weisburd, Jonathan, and Perry (2009a) argue, it is clearly time for scholars to begin to evaluate the effectiveness of police responses to terrorism.

Related Entries

▶ Biometrics and Border Control Policing
▶ Democratic Policing
▶ Ecoterrorism
▶ Effectiveness of Situational Crime Prevention
▶ History of Criminal Investigation
▶ Homegrown Terrorism in the United States
▶ Incarcerating Terrorists
▶ Intelligence-Led Policing
▶ Methodological Issues in Evaluating Police Performance
▶ Policing Terrorism and Legitimacy
▶ Preventing of Terror at Shopping Malls
▶ Rational Choice Theory
▶ Risk Assessment, Classification, and Prediction
▶ Routine Activities Approach
▶ Situational Approaches to Terrorism

▶ Social Disorganization and Terrorism
▶ Terrorist Organizations
▶ Theories for Situational and Environmental Crime Prevention
▶ Use of Social Media in Policing

Recommended Reading and References

Bamford BWC (2004) The United Kingdom's war against terrorism. Terrorism Polit Violence 16:737–756

Bayley D (2006) Changing the guard: developing democratic police abroad. Oxford University Press, New York

Bayley D, Weisburd D (2009) Cops and spooks: the role of police in counterterrorism. In: Weisburd D, Feucht TE, Hakimi I, Mock LF, Perry S (eds) To protect and to serve: policing in an age of terrorism. Springer, New York, pp 81–99

Bratton W, Kelling G (2006) Policing terrorism. Civic Bull 43:1–10

Brodeur J, Dupeyron N (1993) Democracy and secrecy: the French intelligence community. In: Brodeur J, Gill P, Tollborg D (eds) Democracy, law and society. Ashgate, Aldershot, pp 19–23

Clarke R, Cornish B (2001) Rational choice. In: Paternoster R, Bachman R (eds) Explaining criminals and crime: essays in contemporary criminological theory. Roxbury, LA, pp 23–42

Clarke R, Newman G (2006) Outsmarting the terrorists. Praeger Security International, Westport, CT

Connors TP, Pellegrini G (2005) Hard won lessons: policing terrorism in the United States. The Manhattan Institute, New York, Safe Cities Project

Davis LM, Riley JK, Ridgeway G, Pace J, Cotton SK, Steinberg PS, Damphousse K, Smith BL (2004) When terrorism hits home: how prepared are state and local law enforcement? Rand Corporation, Santa Monica

Felson M, Clarke R (1998) Opportunity makes the thief: practical theory for crime prevention. Police Research Series. Paper 98

Freilich J, Newman G (2009) Introduction. In: Freilich J, Newman G (eds) Reducing terrorism through situational crime prevention. U.S. Criminal Justice Press, Monsey, pp 1–7

Ganor B (2009) Trends in modern international terrorism. In: Weisburd D, Feucht TE, Hakimi I, Mock LF, Perry S (eds) To protect and to serve: policing in the years of terrorism, and beyond. Springer, New York, pp 11–42

Greene JR, Herzog S (2009) The implications of terrorism on the formal and social organization of policing: some concerns and opportunities. In: Weisburd D, Feucht TE, Hakimi I, Mock LF, Perry S (eds) To protect and to serve: policing in the years of terrorism, and beyond. Springer, New York, pp 81–99

Hasisi B (2006) Protest policing in deeply divided society: Israeli Arab protest under fire. Paper presented at the annual meeting of the American Society of Criminology, Toronto, Oct 2006

Hasisi B, Alpert GP, Flynn D (2009) The impacts of policing terrorism on society: lessons from Israel and the U.S. In: Weisburd D, Feucht TE, Hakimi I, Felson Mock L, Perry S (eds) To protect and to serve: policing in the years of terrorism, and beyond. Springer, New York, pp 177–202

Hoffman B (2006) Inside terrorism. Columbia University Press, New York

Howard P (ed) (2004) Hard won lessons: how police fight terrorism in the United Kingdom. Manhattan Institute, New York

Innes M (2006) Policing uncertainty: countering terror through community intelligence and democratic policing. Ann Am Acad Polit Soc Sci 605:222–241

International Association of Chiefs of Police (2005) Post 9-11 policing: the crime-control-homeland security paradigm—taking command of new realities. International Association of Chiefs of Police, Alexandria

Kelling GL, Bratton WJ (2006) Policing terrorism. Manhattan Institute for Policy Research, New York, Civic Bulletin 43

Lum C, Kennedy LW, Sherley A (2006) Are counterterrorism strategies effective? The results of the Campbell systematic review on counter-terrorism evaluation research. J Exp Criminol 2:489–516

Martin G (2009) Understanding terrorism challenges, perspectives, and issues. Sage, California

Newman G (2009) Reducing terrorist opportunities: a framework for foreign policy. In: Freilich J, Newman G (eds) Reducing terrorism through situational crime prevention. Criminal Justice Press, USA, pp 33–59

Newman G, Clarke R (2008) Policing terrorism: an executive's guide. U.S. Department of Justice, Washington, DC, Office of Community Oriented Policing Service (COPS)

Perliger A, Hasisi B, Pedahzur A (2009) Policing terrorism in Israel. Crim Justice Behav 36:1279–1304

Sherman L, Gottfredson D, MacKenzie D, Eck J, Reuter P, Bushway S (1997) Preventing crime: what works, what doesn't, and what's promising. University of Maryland, College Park

Weisburd D, Waring E, Chayet E (2001) White-collar crime and criminal careers. Cambridge University Press, Cambridge

Weisburd D, Jonathan T, Perry S (2009a) The Israeli model for policing terrorism: goals, strategies, and open questions. Crim Justice Behav 36:1259

Weisburd D, Feucht T, Hakimi I, Mock L, Perry S (2009) Introduction. In: To protect and to serve: policing in the years of terrorism, and beyond. Springer, New York, pp 1–9

Weisburd D, Hasisi B, Jonathan T, Aviv G (2010) Terrorism threats and police performance: a study of Israeli communities. Br J Criminol 50:725–747

Street Drug Market

▶ Spatial Perspectives on Illegal Drug Markets

Stress

▶ Officer Safety, Health, and Wellness

Structured Sentencing Outside the United States

Julian V. Roberts
Faculty of Law, Centre for Criminology,
University of Oxford, Oxford, UK

Overview

How much discretion should judges have at sentencing? If they have too much, disparity of outcome will be the inevitable consequence. On the other hand, removing their discretion entirely will result in injustice. The solution to this problem lies in the concept of sentencing guidelines which in theory at least provide direction to courts but without unduly restricting their discretion. The US-style grid-based guidelines are only one form of sentencing guidelines; other models are possible. This entry reviews the experience in jurisdictions other than the USA. Several countries have attempted to structure judicial discretion at sentencing in different ways. These other forms of guidance represent alternative approaches to promoting consistency at sentencing. After noting some developments in some other countries, this entry focuses on the experience in England and Wales which is the only common law jurisdiction outside the United States to have developed and adopted a comprehensive sentencing guidelines scheme.

Introduction

When most judges or other criminal justice professionals think about sentencing guidelines, the systems found across the United States come readily to mind. Sentencing guidelines have been evolving in that country for over 30 years now, at both the state and federal levels. Most states have a formal sentencing guidelines scheme to assist judges at sentencing. The best known guidelines model involves a two-dimensional sentencing grid – much like a mileage chart which shows the distance between two cities. Under a sentencing grid, the two dimensions are crime seriousness and criminal history. In order to determine the sentence that should be imposed, a court selects the appropriate level of seriousness and the appropriate criminal history category. Where the crime seriousness row and the criminal history column intersect, there is a grid cell containing a relatively narrow range of sentence length. Sentencing grids of this kind are found in a number of states including Minnesota (e.g., Minnesota Sentencing Guidelines Commission 2010a; Frase 2005, 2009).

If the guidelines are presumptively binding (some states use advisory guidelines, and the US federal guidelines are now considered advisory following several judgments from the US Supreme Court), the court must impose a sentence within the range found in the guidelines grid. A court wishing to impose a more or less severe sanction than that which is prescribed by the guidelines must first find "substantial and compelling" reasons to justify what then becomes a "departure" sentence, namely, one outside the official guideline ranges. Departure rates vary across the US systems. Statistics from Minnesota reveal that in 2009, 25 % of all felony offenders received a sentence different from that prescribed by the guidelines, while a further 14 % of all sentences of custody were outside the guidelines sentence length limits (Minnesota Sentencing Guidelines Commission 2010b, p. 26). The total departure rate for 2009 was therefore 38 %. Departure rates in other states are lower. In Pennsylvania, only 10 % of sentences imposed in 2008 fell outside the guidelines; over the period 1985–2008, the departure rate averaged 11 % (Pennsylvania Commission on Sentencing 2009, p. 46). In Oregon, less than one-fifth of sentences were outside the guidelines (Oregon Criminal Sentencing Commission 2003, p. 13).

Structuring judicial discretion at sentencing is one of the most significant challenges for a legislature. If they prescribe specific sentences – such as mandatory terms of custody – courts are prevented from doing justice by reflecting the individual circumstances of specific offenders. For example, legislating a mandatory sentence of imprisonment for all convictions of robbery means that offenders of different levels of culpability will receive the same sentence – a clear injustice. On the other hand, if legislatures leave the courts to regulate themselves at sentencing, outcomes may be too variable, leading to sentencing disparity. In short, it is a fine balance between offering too much and too little structure.

Sentencing Reform in Other Jurisdictions: Guidance by Words Alone

Sentencing guidelines do not have to be numerical in nature, providing a specific range of sentence for each crime. A number of Scandinavian countries have developed what may be termed "guidance by words" (see also Ashworth (2009) for discussion of techniques to reduce disparity through increased guidance). This approach to structured sentencing involves the legislature placing relatively detailed guidance in a sentencing law. For example, the Swedish Penal Code identifies proportionality as the primary rationale for sentencing and requires courts to assess the seriousness of the crime in order to determine sentence. A number of mitigating and aggravating factors are also specified in the Swedish sentencing law in order to guide judges in the determination of sentence. Finally, the law also contains guidance for courts with respect to the choices they should make between different sentencing options (for further information, see von Hirsch and Jareborg 2009). The advantage of the "guidance by words" approach is that it leaves courts with considerable flexibility to determine an appropriate and proportionate sentence. On the other hand, this may result in much greater disparity than would be the case in a jurisdiction such as Minnesota where judges have to follow detailed and prescriptive sentencing guidelines.

Sentencing Structures: Commissions and Councils

How would a country go about creating a sentencing guidelines scheme? The first step in any move towards structuring judicial discretion involves the creation of an independent authority to developing and issuing sentencing guidelines. All US guidelines schemes emerge from a sentencing commission, such as the Minnesota Sentencing Guidelines Commission or the US Sentencing Commission at the federal level. In other countries, these bodies are usually called Sentencing Councils, and there is significant variation in their structures and functions. The Sentencing Council of England and Wales is headed by the Lord Chief Justice and is tasked with devising and disseminating guidelines as well as a range of other functions (see Roberts 2011, 2012). On the other hand, sentencing councils in Australia such as the Sentencing Advisory Council in New South Wales are, as the name implies, advisory in nature. These councils do not issue sentencing guidelines per se but rather provide advice and conduct research upon a wide range of sentencing matters.

All sentencing councils are involved in public legal education of one kind or another. This may mean publishing reports to help the public understand the sentencing process better, or it may mean releasing comprehensive sentencing statistics. For example, some guidelines authorities publish periodic Sentencing Bulletins which summarize sentencing trends for specific offenses (see http://sentencingcouncil.vic.gov.au/page/about-us/council). The public typically rely on news media accounts of sentencing decisions, and these generally focus on unusual or exceptionally lenient sentences – those which are newsworthy in some respect. It is important therefore for a guidelines authority to dispel public misperceptions of sentencing.

Progress towards developing sentencing guidelines around the world has been fitful and slow. The Law Commission of New Zealand developed a comprehensive and principled set of guidelines, but the legislature in that jurisdiction has yet to proclaim the necessary legislation to permit implementation (see Young and Browning 2008). The New Zealand scheme involved a comprehensive guideline for each offense; the guideline contained categories of crime seriousness, each with an associated range of sentence. A sentencing court would match the case appearing for sentencing to the guideline category using information in the guideline. The system was more flexible than the US-based schemes. The New Zealand proposed guidelines also included "generic" advice – guidelines which apply to more than a single offense. For example, the guidelines provide guidance on considering the impact of the crime upon the victim and also the way in which courts should approach the sentencing of multiple crimes on the same occasion.

Other jurisdictions – including Scotland, Western Australia, Northern Ireland, and Israel – have explored the use of guidelines for sentencers but so far have not actually adopted a formal scheme. The Sentencing Commission for Scotland recommended creation of an Advisory Panel on Sentencing to assist the introduction of sentencing guidelines, but this has yet to become operational (see Hutton and Tata 2010). Following recommendations from a Sentencing Working Group (2010), Northern Ireland held a consultation on the possible options for a form of sentencing guidelines (Criminal Policy Unit 2010).

In January 2012, the Israeli Parliament (Knesset) approved a sentencing law. This law adopted parts of a Bill which provide for "guidance by words" but without establishing the guidelines authority which would have been empowered to develop and issue guidelines scheme involving "starting point sentences" (see Gazal-Ayal and Kannai 2010). Under the new legislation, courts are required to devise their own proportionate sentence range for the case being sentenced and to provide reasons if they impose a sentence outside this range. South Korea has also launched a guidelines scheme (see Park 2009). Several jurisdictions (including New South Wales and the state of Victoria in Australia) have created advisory bodies which disseminate information about sentencing but which do not actually disseminate guidelines (see Abadee (2008) and more generally, Freiberg and Gelb 2008).

Countries Without Guidelines

Finally, some countries – Canada, South Africa, Ireland, and India, for example – have resisted all appeals for greater structure at sentencing (e.g., Terblanche 2003; Doob 2011; Roberts et al. 2011). Although scholars and practitioners in those countries have long advocated creation of some kind of guidelines scheme, legislatures in these countries have so far rejected calls to introduce sentencing guidelines. The consequence is that judges in these jurisdictions continue to impose sentence much as they have for decades, with the only guidance coming from the appellate courts. This approach to sentencing may be termed "judicial self-regulation" (see Ashworth 2009). The limitation of this approach is that higher courts hear only a small proportion of cases on appeal, which means that the opportunities for guidance are limited. When a court of appeal does hear a sentence appeal, it does not always give general guidance.

Guidelines Structures

If appropriately constructed, and not subject to political interference, sentencing guidelines remain the best hope for constraining prison populations and achieving principled sentencing (see Stemen and Rengifo (2011); von Hirsch et al. (2009)), but the question remains: what form of guidelines is appropriate for any given jurisdiction? The guidelines movement remains strong across the USA, but despite its high profile, the model employed in states such as Minnesota has not proven a popular penal export. Canada was the first country to reject this approach to structured sentencing.

In 1984, the Canadian government created a term-limited Sentencing Commission which visited several American states (including Minnesota and Pennsylvania) and concluded that two-dimensional grids held no promise for sentencing in Canada (see Canadian Sentencing Commission 1987). Canada's flirtation with guidelines ended in the 1990s when its Parliament approved a modest sentencing reform package and declined to create a permanent sentencing commission or to create guidelines of an even advisory nature (see Roberts and Cole 1999; Doob 2011). A generation later the Sentencing Commission Working Group in England and Wales visited the home of numerical guidelines and drew the same conclusion (Sentencing Commission Working Group 2008). Western Australia considered adopting a two-dimensional sentencing grid in 1999 but also ultimately abandoned the idea.

The proliferation of two-dimensional sentencing grids across the USA since the 1970s may paradoxically have undermined the appeal of all presumptively binding guidelines. Sentencing guidelines of any kind are often regarded by judges as harbingers of grids and as being antithetical to sentencing as a "human process" (see Hogarth 1971). Calls for the introduction of any kind of sentencing guidelines system are perceived as an attempt to move towards the ultimate goal of a grid. Indeed, opposition in Canada (and England and Wales) to sentencing guideline schemes of all stripes was fueled by predictions that any move towards structuring judicial discretion would culminate in the imposition of a rigid two-dimensional grid. In England and Wales, despite considerable judicial and professional resistance to the concept of guidance derived from a source other than the Court of Appeal, guidelines have slowly emerged over the past decade. Definitive guidelines now exist for most high-frequency offenses.

Sentencing Guidelines in England and Wales

Distinguishing Guideline Schemes in the USA and England

The English guidelines offer a different but equally viable model to follow, and there are lessons for other jurisdictions seeking to impose greater structure at sentencing without adopting a US-style two-dimensional grid. Guidelines have been developing slowly in England and Wales since creation of the Sentencing Advisory Panel in 1998. In addition, the Court of Appeal has been issuing guideline judgments periodically over the years. Courts in England and Wales have therefore a tradition of guidance on which to draw. However, with the creation of the Sentencing Council in 2010 (see Roberts 2011), the guidelines entered a new area, and the new Council began issuing guidelines more frequently and in a new format.

How do the English sentencing guidelines differ from the US-based schemes? The principal difference is that guidelines in England and Wales promote consistency by requiring sentencers to proceed through a series of steps. This represents a different approach to promoting consistency in sentencing. The effectiveness of the English approach has not been evaluated, as the Council responsible for issuing the guidelines has yet to publish the data which will shed light on any changes in consistency of sentencing outcomes. The second principal difference between the English and the US guidelines is that Sentencing Commissions across the USA devise and issue a complete package of guidelines encompassing all offenses – usually, as noted, within a single guidelines grid. The English Sentencing Council issues guidelines sequentially for each offense category (such as all assault offenses, or drugs crimes). This means that it takes longer for guidelines to be available for all offenses. (All English guidelines are available at http://sentencingcouncil.judiciary.gov.uk/).

Step-by-Step Sentencing Methodology

How do the English guidelines work? The English guidelines structure contains a series of

Offence: Assault occasioning actual bodily harm

Step 1: Use the factors provided in the guideline which comprise the principal elements of the offence to determine the category which is appropriate:

- Category 1: Greater harm *and* high culpability;
- Category 2: Greater harm *and* lower culpability *or* lesser harm *and* higher culpability;
- Category 3: Lesser harm *and* lower culpability.

Step 2: Use the *starting point* sentence from the appropriate *offence category* to generate a provisional sentence within the *category range*. The guideline contains a list of additional aggravating and mitigating factors which should result in upward or downward adjustment from the *starting point* sentence.

Offence Category	Starting Point Sentence	Category Sentence Range
1.	18 months custody	12 months – 3 years' custody
2.	6 months custody	Community order – 12 months custody
3.	Community order	A fine – community order

Step 3: Consider if any reduction in the provisional sentence should be made to reflect assistance offered or provided to the prosecution.

Step 4: Consider the level of reduction appropriate to reflect a guilty plea.

Step 5: Consider whether the offender meets dangerousness criteria necessary for imposition of an indeterminate or extended sentence.

Step 6: If sentencing for more than one offence apply the totality principle to ensure that the total sentence is just and proportionate to the total offending behaviour.

Step 7: Consider whether to make a compensation order and/or other orders.

Step 8: Give reasons for and explain the effect of the sentence on the offender.

Step 9: Consider whether to give credit for time on remand or bail.

(For further information, see Sentencing Council (2012); Roberts and Rafferty (2011)).

Structured Sentencing Outside the United States, Fig. 1 Example of sentencing guidelines format in England and Wales

nine steps, of which the first two are the most critical. (See Fig. 1 for a summary of the steps). The idea is that if all courts follow the same step-by-step procedure, sentencing decisions across courts should be more consistent.

For the purposes of illustration, let us examine the guideline for *Assault occasioning actual bodily harm* (see Sentencing Council of England and Wales 2011). As with many offenses for which a definitive guideline has been issued, ABH has been stratified into three categories of seriousness. The guideline provides a separate range of sentence and starting point sentence for each category. Step 1 of the guidelines methodology requires the sentencing court to match the case appearing for sentencing to one of the three categories of seriousness. The three categories have been created to reflect gradations in harm and culpability, with the most serious category (1) requiring greater harm and enhanced culpability. Category 2 is appropriate if either greater harm or higher culpability is present, while Category 3 involves lesser harm and a lower level of culpability.

Determining the Offense Category

Step 1 of the guideline identifies an *exhaustive* list of sentencing factors which should be used to determine which of the three categories of seriousness is most appropriate for the particular offender appearing for sentencing. These factors constitute what the guideline describes as the "principal factual elements" of the offense. Their importance is reflected in the fact that determination of the category range is the step which will have greatest influence on severity of sentence. For example, if the court chooses the lowest category, the most severe sentence it should impose is a community-based (noncustodial) sentence. In contrast, the next category up has a maximum sentence of 51 weeks imprisonment. Having determined the relevant category range, a court should use the corresponding starting point sentence to move towards a sentence which will then be shaped by the remaining steps in the guideline.

Step Two: Shaping the Provisional Sentence

Step 2 requires a court to "fine-tune" the sentence by considering circumstances which provide "the context of the offense and the offender." The guidelines provide a non-exhaustive list of sentencing factors for courts to consider at this step of the methodology. The aggravating factors include committing the offense while on bail or license, while under the influence of drugs, or with abuse of trust. The guideline factors which reduce seriousness (and which result in less severe sentence) include an absence of prior convictions and the fact that the crime was an isolated incident. A diverse collection of factors is cited as relevant to personal mitigation, including remorse, the fact that the offender was a sole or primary carer and "good character and/or exemplary conduct."

Other Steps Towards the Final Sentence

Leaving Step 2 leads a court to the remaining seven steps of the guidelines methodology, which may be briefly summarized. Step 3 directs courts to reduce sentence in cases where the offender has provided or offered to provide assistance to the prosecution or police. Step 4 involves the reduction for a guilty plea. Offenders who plead guilty receive a reduction of up to one-third off their custodial sentence, depending upon how early they enter their plea. The remaining steps to be followed relate to other sentencing provisions such as the consideration of any time that the offender spent in remand while awaiting trial. (This time is taken off any custodial sentence ultimately imposed). The guideline steps also require the court to give reasons for the sentence imposed and to explain the effect of the sentence for the benefit of the offender. Finally, courts give reasons for the sentence and, in particular, if they have imposed a sentence which falls outside the guidelines range.

Legal Status of English Guidelines

As noted, in many US states, the sentencing guidelines are binding on courts: a judge may depart from the guideline range if he or she finds "substantial and compelling" reasons why the guideline sentence is inappropriate. The English guidelines are less restrictive, although they are formally binding on courts. When sentencing an offender in England and Wales, a sentencing court "must follow" any relevant sentencing guidelines – unless it would be contrary to the interests of justice to do so (see discussion in Ashworth 2010). However, courts may impose a sentence within a wide range and still remain compliant with the guidelines. In this sense, although they are required to follow the guidelines, courts retain considerable freedom to sentence offenders: first, they have considerable discretion within the guidelines, and second, they may depart from the guidelines if they believe it is necessary in the interests of justice.

Conclusion

What have we learned about the experience with structuring sentencers' discretion outside the United States? A number of lessons can be drawn. First, the Scandinavian model suggests that numerical guidelines are not necessarily the only model to follow. It is possible to offer guidance to courts without prescribing specific sentencing ranges in terms of numbers of months or years. Second, the English guidelines demonstrate that there is a middle ground lying between the relatively tight sentencing guidelines found across the United States and the looser systems of "guidance by words" found in countries like Sweden and Finland (see von Hirsch et al. 2009, Chapter 6). Along with the New Zealand proposals, the English guidelines offer a system which is numerical, prescriptive, and yet quite flexible in application. These two systems offer a plausible improvement upon the highly discretionary sentencing arrangements found in countries like Canada, South Africa, and India. Third, judicial acceptance of greater structure (and reduced discretion) is far more likely when the judiciary is heavily implicated in the development and evolution of the guidelines. The statutory bodies responsible for the guidelines in England and Wales have generally been dominated by the judiciary. The Canadian Sentencing Commission proposals failed, in part because judges perceived the guideline scheme to be a bureaucratic scheme created by academics.

Fourth, there may be an advantage to the gradual evolution of the guidelines. The English guidelines have been criticized for being slow to develop – with guidelines for specific offenses issued periodically over the years rather than in one step as was the case in the United States or the proposals advanced in New Zealand. There was no "big bang" to the English experience in the sense of a comprehensive guidelines package covering all offenses. In retrospect, this potential weakness of the guidelines may paradoxically have ensured their survival and development. They have evolved incrementally – from their modest origins in 1999 (see Ashworth and Wasik 2010) through to the much more comprehensive and detailed scheme of 2011. Judges who are traditionally resistant to any attempts to curb their discretion may be more likely to accept guidance when it comes in this format.

The ultimate question to pose, however, is the following: Are the guidelines proposed or implemented in other countries better or worse, more or less effective than those developed across the United States? Unfortunately the absence of truly comparative research makes it impossible to resolve the issue one way or another. In addition, the non-US-based guidelines such as those in England have yet to be evaluated. For example, we do not yet know how much the English guidelines have improved consistency in that country. At the very least, however, the experience in that country demonstrates that it is possible to introduce detailed and prescriptive sentencing guidelines even in a common law jurisdiction which, in the 1980s and 1990s, was committed to the traditional model of privileging judicial discretion (see Ashworth 2010).

Related Entries

▶ Mitigation and Aggravation at Sentencing
▶ Sentencing as a Cultural Practice

Recommended Reading and References

Abadee A (2008) The New South Wales sentencing council. In: Freiberg A, Gelb K (eds) Penal populism, sentencing councils and sentencing policy. Willan, Cullompton

Ashworth A (2009) Techniques for reducing sentence disparity. In: von Hirsch A, Ashworth A, Roberts JV (eds) Principled sentencing. Readings on theory and policy, 3rd edn. Hart Publishing, Oxford

Ashworth A (2010) Sentencing and criminal justice, 5th edn. Cambridge University Press, Cambridge

Ashworth A, Wasik M (2010) Ten years of the sentencing advisory panel. In: Sentencing advisory panel annual report. Available at: www.sentencing.council.org

Canadian Sentencing Commission (1987) Sentencing reform: a Canadian approach. Supply and Services Canada, Ottawa

Criminal Policy Unit (2010) Consultation on a sentencing guidelines mechanism. Northern Ireland Department of Justice, Belfast

Doob AN (2011) The unfinished work of the Canadian Sentencing Commission. Can J Criminol Crim Justice 55:279–298

Frase R (2005) Sentencing guidelines in Minnesota, 1978–2003. In: Tonry M (ed) Crime and justice. University of Chicago Press, Chicago

Frase R (2009) Sentencing policy under the Minnesota sentencing guidelines. In: von Hirsch A, Ashworth A, Roberts JV (eds) Principled sentencing. Readings on theory and policy, 3rd edn. Hart Publishing, Oxford

Freiberg A, Gelb K (eds) (2008) Penal populism, sentencing councils and sentencing policy. Willan, Cullompton

Gazal-Ayal O, Kannai R (2010) Determination of starting sentences in Israel – system and application. Fed Sentencing Report 22:232–242

Hogarth J (1971) Sentencing as a human process. University of Toronto Press, Toronto
Hutton N, Tata C (2010) A sentencing exception? Changing sentencing policy in Scotland. Fed Sentencing Report 22:272–278
Minnesota Sentencing Guidelines Commission (2010a) Minnesota sentencing guidelines and commentary. Revised, Aug 2010. MSGC, St. Paul
Minnesota Sentencing Guidelines Commission (2010b) Sentencing practices. Annual summary statistics for felony offenders sentenced in 2009. MSGC, St. Paul
Oregon Criminal Sentencing Commission (2003) Sentencing practices, 2001. State of Oregon, Criminal Sentencing Commission, Salem
Pennsylvania Commission on Sentencing (2009) Sentencing in Pennsylvania. Annual report 2008. Available at: http://pcs.la.psu.edu/
Park H (2009) The basic features of the First Korean sentencing guidelines. Korean Ministry of Justice, Seoul
Roberts JV (2011) Sentencing guidelines and judicial discretion: evolution of the duty of courts to comply in England and Wales. Br J Criminol 51:997–1013
Roberts JV (2012) Structured sentencing: lessons from England and Wales for common law jurisdictions. Punishment and society. Int J Penol 14(3):267–288
Roberts JV, Azmeh U, Tripathi K (2011) Structured sentencing in England and Wales: lessons for India? Natl Law School India Rev 23:27–46
Roberts JV, Cole D (eds) (1999) Making sense of sentencing. University of Toronto Press, Toronto
Roberts JV, Rafferty A (2011) Structured sentencing in England and Wales: exploring the new guideline format. Crim Law Rev 9:680–689
Sentencing Commission Working Group (2008) Sentencing guidelines in England and Wales: an evolutionary approach. SCWG, London
Sentencing Council of England and Wales (2011) Assault. Definitive guideline. Available at: http://sentencingcouncil.judiciary.gov.uk/
Sentencing Working Group of Northern Ireland (2010) Monitoring and developing sentencing guidance in Northern Ireland. Northern Ireland Department of Justice, Belfast
Stemen D, Rengifo A (2011) Policies of imprisonment: the impact of structured sentencing and determinate sentencing on state incarceration rates, 1978–2004. Justice Q 28:174–201
Terblanche S (2003) Sentencing guidelines for South Africa: lessons from elsewhere. S Afr Law J 120:858–882
von Hirsch A, Ashworth A, Roberts JV (eds) (2009) Principled sentencing. Readings on theory and policy, 3rd edn. Hart Publishing, Oxford
von Hirsch A, Jareborg N (2009) The Swedish sentencing law. In: von Hirsch A, Ashworth A, Roberts JV (eds) Principled sentencing. Readings on theory and policy, 3rd edn. Hart Publishing, Oxford
Young W, Browning C (2008) New Zealand's sentencing council. Crim Law Rev, April 2008, pp 287–298

Subsidized Housing

▶ Public Housing and Crime Patterns

Suicide

▶ Officer Safety, Health, and Wellness

Suicide and Prisons

Gregory F. Kellermeyer[1] and Hal S. Wortzel[2]
[1]Denver Health Medical Center and Division of Forensic Psychiatry, Department of Psychiatry, University of Colorado Denver School of Medicine, Denver, CO, USA
[2]Division of Forensic Psychiatry, Department of Psychiatry, University of Colorado Denver School of Medicine and VISN–19 MIRECC, Denver Veterans Affairs Medical Center, Denver, CO, USA

Overview

An increasing body of literature reflects the heightened risk for suicide faced by justice-involved persons, whether they are in jails, in prisons, or in the process of transitioning from a correctional facility back into the community. Suicide and suicide prevention have garnered increased attention, efforts, and resources within correctional facilities. Despite this increased awareness of suicide and heightened efforts to prevent suicide in correctional settings, suicide rates among incarcerated persons in the United States have continued to exceed that of the general population (albeit less so than in previous decades). Pertinent case law clearly creates potential liability for both correctional and health-care staff in the wake of a death by suicide. This entry will review applicable case law; explore statistical trends, risk factors, and characteristics of suicides in correctional settings;

review methods for prevention of suicide supported by the literature; discuss select special populations of incarcerated individuals at risk; and explore the suicide risk surrounding the transition from correctional settings back into the community. This entry will validate the importance of the issue of suicide in correctional settings and highlight areas where further research and progress are needed.

Case Law History

The potential for legal liability resulting from suicide in correctional settings originates in two seminal cases: *Estelle v. Gamble* (1976) and *Bowring V. Godwin* (1976). In the *Estelle* case, the plaintiff filed suit claiming that the medical director and two correctional officers of a state correctional department violated the Eighth Amendment (cruel and unusual punishment) by not providing appropriate medical care for the inmate's back pain. Upon appeal, the US Supreme Court established the legal standard of "deliberate indifference" and by using this standard held that the inmate's medical care did not violate the 8th Amendment. Deliberate indifference requires more than mere negligence. Negligence typically involves the failure to exercise the care toward others which a reasonable or prudent person would do in similar circumstances, with the resulting harm generally being accidental. A reckless culpable mental state typically involves gross negligence, meaning that the individual could foresee the potential for harm, but was heedless of the consequences. The deliberate indifference standard is more akin to recklessness, necessitating that in order to be deemed liable, prison staff must have the subjective knowledge of risk for serious harm and then disregard that risk.

In the *Bowring* case, an inmate in a state correctional facility claimed Eighth Amendment (cruel and unusual punishment) and Fourteenth Amendment (due process) violations in lower court due to being denied parole, in part as a result of a psychological opinion that he would not be a successful parolee. The plaintiff argued that he was not provided the psychiatric or psychological diagnosis or treatment needed to address the psychological concerns of the parole board. The Fourth Circuit Court of Appeals held that psychiatric illnesses are medical in nature, thereby extending the treatment obligations incumbent upon correctional facilities to mental illness. The combination of these two seminal cases sets the stage for subsequent decisions finding liability in failure to prevent inmate suicide.

Concerning suicide specifically, circuit-level court findings in *Partridge v. Two Unknown Police Officers* (1986) and *Colburn v. Upper Darby Township* (1991) establish inmate suicide as a potential source of liability for correctional facilities (Scott 2010). The *Partridge* Fifth Circuit court determined that psychiatric illnesses potentially resulting in deaths by suicide constitute a "serious medical need." In *Colburn v. Upper Darby Township* (1991), the Third Circuit court established that correctional facilities have an obligation to train staff members to identify potentially suicidal inmates and to intervene in an effort to prevent suicide deaths. The court determined that the standard for correctional staff's competency in identifying potentially suicidal inmates would be comparable to that of a layperson's and that correctional staff would be held to a standard of "reckless indifference" when legally evaluating efforts to intervene and prevent an inmate death by suicide. The Third Circuit court confirmed the lower court's analysis, indicating that a plaintiff must prove not only that the deceased inmate had a "particular vulnerability to suicide" but that custody staff knew or should have known about the vulnerability and that the custody staff acted with "reckless indifference" in relation to the individual's particular vulnerability. Existing case law clearly sets a bar for correctional facilities' role in suicide prevention, though sets that bar relatively low.

Suicide Rates in Correctional Settings

Historically, suicide rates have been exceedingly high within correctional settings relative to the general population. In 1983, the suicide rate for jails was 129 per 100,000 inmates, and suicide accounted for 56 % of all jail deaths (Noonan 2010). Fortunately, suicide rates in both jails and state prisons have decreased since the 1980s. From 2000 to 2007, the suicide rate within jails decreased by approximately 25 %, from 48 per 100,000 to 36 per 100,000. Despite this relative decline in suicide rates, suicide remained the single leading cause of preventable death in jails. When jail suicide deaths rates from 2000 to 2007 are adjusted for demographic factors such as age, race, and sex, the rates still remain high relative to those observed in the general population in the United States (14 per 100,000). Overall, the adjusted rate for suicide in jail is 59 per 100,000 for males and 27 per 100,000 for females. Comparable rates of suicide in the US general population are 21 per 100,000 for males and 5 per 100,000 for females. It is worth noting that suicide rates within jails may vary substantially depending on the size of (and presumably resources available within) the correctional facility. Large jails had a lower suicide rate (27 per 100,000) relative to small jails (167 per 100,000), with small jails being defined as those housing 50 or fewer inmates. Noonan theorizes that a lack of suicide prevention measures, including appropriate mental health services and more inmate turnover in smaller jails, may account for this difference (2010). State prisons have also demonstrated a fortunate downward trend in suicide rates over the past few decades, though they still exceed that of the general population. The rate of suicide was 34 per 100,000 in 1980 (Mumola 2005) and decreased to a rate of 16 per 100,000 for the years 2000–2007. This rate has held fairly stable since the 1990s (Noonan 2010). Federal prison suicide rates are reported at 10 to 17 per 100,000 (White et al. 2002).

A poignant criticism has been raised regarding the method employed in measuring the rate of suicide in jails. These reported rates are based upon the total national average daily inmate population rather than the number of admissions to the specific incarceration facilities (Hayes 2010). If the rates were calculated based upon the number of admissions rather than the average daily population, calculated suicide rates would be lower. Calculating suicide rates based on the total number of admissions would arguably be more accurate because this method takes into account the high turnover of jails and thus more fully captures the at-risk population. Larger jails release approximately 46 % of inmates within the first 2 days of confinement and release 80 % of their population within an inmate's first month of incarceration. Calculations based on admissions would thus more fully take into account all the unique individuals incarcerated. Others argue that community suicide rates are not calculated on total "admissions" to the population, and moreover, the majority of admissions are brief incarcerations and thus should not be included in the statistical calculation. But, as will be discussed below, many suicides occur very early after incarceration, illustrating the potential importance of capturing all admissions when calculating rates.

Incarceration within correctional facilities does seem to carry increased risk for suicide. Why has the rate of suicide been high in correctional settings, and furthermore, why has the rate decreased since the early 1990s? Various investigations and expert opinions have proposed several possible explanations.

Prevalence of Mental Illness in Correctional Settings

One significant risk factor for suicide, applicable to individuals across settings (whether they are in jails, prisons, or the general community), is a diagnosis of mental illness. In one study of individuals who died by suicide in jails between 2005 and 2006, 38 % suffered from mental illness, 20 % had a history of being prescribed psychiatric medications, and 34 % had a history of self-directed violence, including suicide attempts and nonsuicidal self-directed violence (Hayes 2010). Patterson reported that 73 % of the California prison inmates who died by suicide in a 6-year period had a history of mental health treatment and 62 % had a history of suicidal thoughts or behavior (2008). Fazel conducted a systematic review of suicides in prison and found that history of attempted suicide, current psychiatric diagnosis, and current prescription for psychotropic medication were risk factors for suicide in prison (2008).

Rates of suicide in correctional settings are almost certainly related to the prevalence of mental illness. For jails and prisons, the number of admitted individuals suffering mental illness has increased over the last half of the twentieth century. Since the 1960s, the number of long-term care mental health institutions (such as state mental health hospitals) has decreased under the deinstitutionalization movement (Lamb and Weinberger 2005). Deinstitutionalization evolved from legal and ethical concepts favoring treatment of mentally ill individuals in the least restrictive environment possible. Under this theory, individuals would be released from highly restrictive psychiatric facilities and would instead be treated and supported in community-based mental health centers. Unfortunately, many experts agree that the deinstitutionalization movement has largely failed due to the inadequate creation and funding of needed community mental health centers, such that many severely, chronically mentally ill individuals are not receiving the treatment and support needed to sustain tenure in the community (Scott 2010). In the absence of such support, many of these mentally ill persons go on to engage in behaviors that eventually lead to incarceration. Hence, the failed deinstitutionalization movement has, over time, resulted in a shift in location for the severely mentally ill, from mental health facilities to correctional settings. In 1955, there were 559,000 psychiatric hospital beds for a total United States population of 165 million, compared to 59,403 psychiatric hospital beds for a total United States population of 273 million in 2000 (Lamb and Weinberger 2005). The notion that individuals suffering from a mental illness have shifted institutional settings, from mental health hospitals to correctional facilities, has been termed the criminalization of the mentally ill (Scott 2010). Other proffered explanations for this shift include lack of resources for community mental health care, peace officers assuming the role of triaging individuals to hospitals versus jail, and more restrictive involuntary psychiatric hospitalization laws (Lamb and Weinberger 2005).

Simultaneous to the deinstitutionalization of the mentally ill, the United States began its incarceration binge which has resulted in an increased rate and prevalence of incarceration. The total number of inmates in the United States in 1980 was 501,886, compared to over 2.2 million in 2009 (Glaze 2010). As a result, the total number of mentally ill incarcerated individuals increased with the incarceration binge. A study by Bradley-Engen supports this additional explanation (2010). The study reported the change in the total number of individuals admitted to Washington State prisons from 1998 to 2006 and the relative proportion of inmates diagnosed with serious mental illness and co-occurring substance use at time of admission. The results of the study suggest that the increase in the diagnosis of serious mental illnesses and co-occurring substance use disorders is related to the increase in total admissions to prisons during the study time period and is not due to an increase in the percentage of admissions involving serious mental illness. In other words, this study suggests that at least some of the increase of the total number of incarcerated mentally ill is a result of the overall higher number of individuals being incarcerated. Collectively, it seems that an unfortunate combination of factors, chiefly involving the failed deinstitutionalization movement and a criminal justice system featuring a strong predilection toward incarceration, has resulted in growing populations of mentally ill persons behind bars in the United States.

Estimates vary for prevalence rates of mental illness in jails and prisons depending on the method of the study. The Bureau of Justice Statistics reported that 56 % of state prisoners, 45 % of federal prisoners, and 64 % of jail inmates suffered from some type of mental health problem (James et al. 2006). Bradley-Engen reported that past studies suggest that the prevalence of serious mental illness in prisons ranges from 16 % to 24 %, compared to 5–7 % in the general community (2010). A recent study suggests that the rate of serious mental illness in jail is 14.5 % for men and 31.0 % for women (Steadman et al. 2009).

Risk Factors for Suicide Pertinent to Correctional Settings

Besides housing an influx of many individuals that might have otherwise been institutionalized in mental health facilities during an earlier era, the correctional setting tends to contain populations with high rates of other important risk factors for suicide. Many of the risk factors associated with inmates' suicide deaths match risk factors associated with suicide deaths in the community: male gender, Caucasian race, substance abuse, psychiatric treatment history, and prior suicide attempts (Anasseril 2006; Patterson and Hughes 2008; Hayes 2010). Fazel reported that being a male and being Caucasian were risk factors for suicide in prison; he also found that being African American was inversely related to suicide (2008). Noonan reported that the suicide rate for African Americans was lower compared to Hispanic and Caucasian inmates (2010). However, it is vital to appreciate that demographic risk factors of this kind are of very limited utility in determining risk at the single inmate level, keeping in mind that favorable demographic factors identified via population-based research do not keep individuals safe; inmates of any race or gender may of course present at heightened risk for suicide.

Complicating risk assessment in correctional settings is the fact that so many of the identified risk factors for suicide occur at a very high prevalence, higher than the prevalence in the general community. For example, inmates in correctional settings are predominately male (Glaze 2010). The male gender is dramatically overrepresented in correctional settings, is overrepresented in suicide deaths in the general community, and also represents approximately 93 % of jail suicides (based on figures from 2005 to 2006) (Hayes 2010). In federal prisons, despite having a 7 % female population, from 1993 to 1997, 100 % of suicides were completed by males (White et al. 2002). Noonan reports that white males in jail have a rate of suicide 5.7 times greater than white males in the general population (2010).

Substance use disorders also have a high prevalence in correctional settings. As a conservative estimate, 50 % of incarcerated individuals have substance use disorders, with jail inmates having a higher prevalence rate than prison inmates (Scott 2010). Hayes found that of the individuals that died by suicide in jail between 2005 and 2006, 47 % had a history of substance abuse; 20 % were intoxicated at the time of the suicide (2010). In New York state prisons between 1993 and 2001, 95 % of inmates who died by suicide had a history of substance abuse (Way et al. 2005).

Many incarcerated individuals have experienced violent trauma during their lifetime (Wolff et al. 2010). As cited by Wolff, at least 50 % of incarcerated women have reported victimization in their lifetime. Additional profile characteristics for individuals that committed suicide in prisons are low education level and limited social supports in the community (White et al. 2002; Anasseril 2006). The frequency of trauma, victimization, and limited support among correctional populations seems particularly relevant to suicide risk when one considers the interpersonal-psychological theory of suicide risk offered by Thomas Joiner (2005). Under this model, the convergence of acquired ability (relating to habituation to pain and suffering along with perceived burdensomeness and/or a sense of failed belongingness results in a heightened risk for death by suicide). The above-referenced research indicates that inmates are frequently themselves victims of trauma, potentially contributing to habituation to pain and suffering, and thus acquired ability to engage in suicidal behaviors. Additionally, limited psychosocial support and the impact incarceration has on families likely engender feelings of burdensomeness and failed belonging. It would seem that the convergence described by Joiner is probably a common occurrence among inmates and may in part explain the observed elevation in suicide rates.

The high rates of violent perpetration and aggression among correctional populations are a consideration when thinking about suicide risk and suicide rates. Violent offenders have been found to face increased suicide rates across a number of studies, with the Bureau of Justice statistics reporting a suicide rate nearly three times that of nonviolent offenders in jail, and violent offenders in prison being more than twice as likely to die by suicide. Hayes reported that 43 % of inmates that died of suicide had been arrested on personal and/or violent charges, including murder, negligent manslaughter, armed robbery, rape, sexual assault, indecent assault, child abuse, domestic violence, assault, battery, aggravated assault, and kidnapping (2010). Noonan also found that violent offenders had the highest suicide rate, with public-order offenders having the second highest rate of suicide, compared to individuals with property and drug offenses (2010). Inmates incarcerated for the alleged offenses of homicide, kidnapping, and rape had the highest rate of suicide (Noonan 2010). Notably, aggression features prominently in a theoretical model for suicide risk offered by Kerr et al. (2007). Under that model, aggression may lead to suicide attempts via various routes: aggression may lead to depression and then to suicidal ideation; aggressive acts may yield negative consequences leading to worsening depression and suicidal ideation; or aggression may more directly be turned inwardly resulting in self-directed violence and suicide attempts.

The incarcerated individual's coping styles, changing emotional states, or emotional traits have also been associated with suicide risk. Individuals that have a maladaptive style of interacting with the environment and coping with stress have increased risk for suicide and self-directed violence. Specifically, rigidity and dichotomous thinking, poor problem-solving ability, avoidant behavior, and poor positive reinterpretation are personality styles that lend themselves to suicidal behavior (Polluck and Williams 2006). Emotional states or traits involving anxiety, depression, aggression/agitation, hostility, fear, and impulsivity are characteristic of individuals who have suicidal behavior (Polluck and Williams 2006; Anasseril 2006; Way et al. 2005). An individual's self-report of feeling hopeless is arguably a highly significant risk factor for suicide in most if not all populations, including incarcerated populations (Anasseril 2006). Recent "behavior changes" of incarcerated individuals from baseline observations also has been noted as a characteristic profile of individuals who are at risk for suicide in prison (Way et al. 2005). For example, prior to some suicide deaths, staff members report that the inmate had not been acting in his or her usual manner.

Psychosocial stressors unique to incarceration and the criminal justice process may also contribute to increased suicide risk. For example, many jail and prison suicides occur during the initial incarceration phase or after transfer to a new facility (Anasseril 2006; Hayes 2010). After arrest and initial incarceration in jail, an individual has gone from living in a free society to living in isolation, or possibly in a cell under 23 h lockdown where freedoms are severely curtailed and most aspects of daily life kept under strict control (Patterson and Hughes 2008; Hayes 2010). Many recent detainees are unclear of the specific consequences of their legal charges or have not spoken with an attorney, are intoxicated or experiencing withdrawal from drugs or alcohol, and/or have not yet communicated with their primary supports in the community. Some detainees are experiencing guilt or shame concerning their charge and/or have just experienced a personal/family crisis that is related to their charge (Hayes 2010). Although the above-described stressors are challenging for all inmates, individuals from higher social economic backgrounds incarcerated for the first time seem to be particularly susceptible to such stressors and at higher risk for suicide (Anasseril 2006).

For longer stays in jail or prisons, stressors from both within the correctional setting or crises occurring outside the correctional setting can contribute to suicide risk. Specifically, death of a loved one, relationship problems/losses, financial losses, new charges or setbacks in court, gang problems, victimization, bullying, recent experience or risk of sexual assault, and recent institutional disciplinary action are often antecedents of inmate deaths by suicide (Way et al. 2005; Anasseril 2006). A life sentence is considered a risk factor for suicide in prisons (Fazel et al. 2008). Death row inmates have the highest suicide risk (Anasseril 2006). Although counterintuitive to some, marriage was found to be a risk factor for suicide in prisons and during periods of longer incarceration in jails (Fazel et al. 2008; Hayes 2010). It is interesting to note that jail and prison suicides were not associated with certain holidays but evenly distributed through the year (White et al. 2002; Anasseril 2006; Hayes 2010).

Despite curtails in freedom and relatively limited access to means (compared to life in the community), the correctional environment does offer sufficient opportunity for inmates that are having suicidal thoughts to engage in lethal acts of self-directed violence (Hayes 2010). A one-person cell, time spent alone in a cell, and access to lethal means (such as anchor points for hanging) provide the necessary means for an inmate to die by suicide (Patterson and Hughes 2008; Hayes 2010). Fazel found that single-cell occupation is a risk factor for suicide (2008). The most common method of suicide in both jails and prisons is hanging, featuring in over 80 % of suicides (Patterson and Hughes 2008; Hayes 2010). Other methods employed include overdose, self-inflicted laceration resulting in exsanguination, self-strangulation, swallowing objects, electrocution, poisoning, self-immolation, self-inflicted gunshot wound, and intentional fall (Felthous 2011). All these techniques are more easily accomplished when an inmate is isolated.

Suicide Prevention Programming

The primary method to mitigate suicide risk and implement protective factors is to create an effective comprehensive suicide prevention program in individual correctional facilities (Hayes 2010). Effective suicide prevention programs have been credited with reduction of suicides in correctional settings. National organizations and credentialing bodies such as the American Psychiatric Association, the National Commission on Correctional Health Care, and the American Correctional Association all recommend that correctional facilities have suicide prevention programs and articulate specific standards for these programs. Hayes recommends the following eight essential components for a comprehensive suicide prevention program: staff training, intake screening/assessment, communication, housing, levels of observation, intervention, reporting, and follow-up and morbidity-mortality review (Hayes 2010). A brief summary highlighting the major aspects of Hayes' recommendations for a comprehensive suicide prevention program follows.

Hayes states that all custody, mental health, and medical staff should initially have 8 h of training and 2 h of training annually (Hayes 2010). Training custody staff is crucial due to the large amount of time they spend with inmates, resulting in their significant role in identifying potentially suicidal inmates. Hayes asserts that staff training should include, at a minimum, "reasons why correctional environments are conducive to suicidal behavior, staff attitudes about suicide, potential predisposing factors to suicide, high-risk suicide periods, warning signs and symptoms, identification of suicide risk despite the denial of risk, liability issues, critical incident stress debriefing, [review of] recent suicides and/or serious suicide attempts within the facility or agency, and components of the facility's or agency's suicide prevention policy" (Hayes 2010).

Hayes recommends that individuals receive screening and assessment for suicide risk upon intake to a correctional facility (Hayes 2010). Hayes asserts that screening should include questions concerning suicidal thoughts/plans (including such thoughts during past incarcerations), past suicidal attempts, history of mental health treatment, recent stressors and losses, suicide behavior by family or close friends, and reports from arresting officers concerning suicide risk (Hayes 2010). As in other settings, suicide risk assessment is a process, not an event; suicide screening does not end after intake, but continues throughout the individual's incarceration. Inmates' self-reporting should be considered in concert with behavioral observations and other collateral information in determining the level of suicide risk.

Communication between arresting officer, correctional staff, mental health staff, and medical staff is essential to reduce suicide risk (Hayes 2010). Protocols for appropriate suicide precautions should be in place to enable staff to take appropriate action if an inmate is considered to be at risk for suicide in order to keep the inmate safe. Hayes suggests that custody staff should not be underestimated in their importance in identifying potentially suicidal inmates and communicating this concern to mental health staff (2010). In additions, barriers need to be reduced in order for inmates to obtain mental health care.

Appropriate housing of inmates with significant suicide risk is an important component of a proper suicide prevention program (Hayes 2010). Inmates with clinically determined differing levels of suicide risk should be housed in the least restrictive environment needed to provide the appropriate level of safety, such as medical/mental health units or general population housing where staff can readily interact with them. Only individuals with the highest acuity of suicide risk should be housed in a locked-down environment, and only when the appropriate suicide precautions are provided. Although segregation may be the safest option for individuals at imminent suicide risk, isolation may also be detrimental to the inmate's emotional well-being and seldom addresses the factors driving the crisis. Careful consideration regarding the risks and benefits of segregation is warranted. Cells used for individuals at acute risk for suicide should be free of methods which might be employed to act upon suicidal thoughts, such as anchors or hanging devices in the cell or on the inmate. Level of observation of a suicidal inmate in a particular housing setting depends on the particulars of the individual's suicide risk (Hayes 2010). Observation should be, at a minimum, staggered in 15-min intervals and, at least for inmates judged to be at imminent risk, should consist of continuous in-person observation. Cameras in cells can help monitor inmates at acute risk for suicide but do not replace the need for continuous in-person observation. Inmates flagged for suicide precautions should be closely followed for evaluation and treatment by mental health staff.

All staff with regular contact with inmates should be provided with first aid, cardiopulmonary resuscitation, and automated external defibrillator training to be prepared to intervene after a suicide attempt (Hayes 2010). Staff should be prepared to quickly triage the emergency situation, contact appropriate medical staff, and initiate the appropriate above treatment(s). Staff should not assume that the individual is deceased. If a serious suicide attempt or completed suicide has occurred, family and appropriate authorities should be notified immediately. Morbidity-mortality reviews and psychological autopsies are recommended. Within 24–72 h, Hayes recommends that staff involved in the incident be offered critical incident stress debriefing by a trained professional (Hayes 2010).

Correctional facility culture and environment seem to be related to suicide risk. Liebling argues that inmates that are already vulnerable to suicide and self-directed violence face even higher suicide risk if the institution creates feelings of lack of safety, lack of respect/fairness, alienation, and frustration (2006). In fact, she argues that institutions that provide individual support create a protective factor. She recommends that institutions focus on improving both correctional custody and healthcare staff culture, specifically avoiding environments that use excessive authority, have a lack of personalization of the individual inmate, and avoid addressing or listening to inmate complaints. An international study analyzed possible interventions or factors that reduced the incidence of suicide in prisons. "Purposeful activity" (as defined as classes or activities for inmates, such as education, substance abuse treatment, work, and family visits) was significantly associated with lower suicide rates (Leese et al. 2006). In a study conducted in one jail, protective factors included reasons for living and social connectedness.

Special Populations

Special populations mandate specific considerations when evaluating the risk for suicide and self-directed violence within correctional settings. A few examples follow.

Adolescents

In the United States general population, suicide is one of the leading causes of death for adolescents. Suicide is the fourth leading cause of death of individuals between the ages of 10 and 14, while suicide is the third leading cause of death for teenagers between the ages of 15 and 24. The Centers for Disease Control and Prevention reported that the suicide rate in 2007 for adolescents between the ages of 15 and 19 is 6.8 per 100,000. According to Roberts, a leading cause of death in juvenile correctional settings is suicide, with incarcerated juveniles having a risk for suicide approximately four times greater than juveniles in the general population (2006). As with incarcerated adults, confined juveniles have a high rate of mental illness compared to the general population. Hayes cited multiple studies which indicate prevalence rates of mental illness in the confined juvenile population to be well above 50 %, with many of the juveniles having multiple psychiatric diagnoses, including substance use disorders, conduct disorder, and history of childhood abuse (2009). Penn reported that 12.4 % of 289 adolescents recently admitted to a New England correctional facility had a previous history of attempting suicide (2003).

Characteristics of juveniles who have died by suicide while in confinement are similar to characteristics of adult victims of suicide in correctional settings. A review by Hayes of 79 juvenile suicides between 1995 and 1999 documented many characteristics similar to adult suicide victims in correctional settings, as well as some demographics unique to this population (2009). For example, juvenile suicide victims were 68 % white and 79.7 % male. Also similar to incarcerated adults, 73.4 % had a history of substance abuse, 74.3 % had a history of mental illness, and 69.6 % had history of prior suicidal behavior. The method of suicide, similar for adult suicide deaths, was hanging by 98.7 %. As with adult correctional facilities, the physical plant characteristics provided opportunities for suicide behavior with 50 % being in room confinement and 74.7 % housed in single-occupancy rooms. In addition, the data provided specifics for juvenile suicide victims, including a mean age of 15.7 with 70 % between ages of 15 and 17, 69.6 % confined for nonviolent offenses, and 78.5 % with history of previous offenses. Approximately 40 % of juvenile suicide victims confined in detention centers had died by suicide during the first 72 h of confinement. Other confinement settings including training school/secure facility, reception/diagnostic center, and residential treatment center had more diverse time frames of juvenile suicide deaths. Hayes argued that although adolescent suicide has gained national attention, suicide of juveniles in confinement has gained much less consideration. He recommended more research into investigating the possible precipitating factors of juvenile suicide in confinement (Hayes 2009).

Women

The total percentage of incarcerated women in the United States has increased by 14 % from 1990 to 2009 (Glaze 2010). Steadman estimated that the rate of serious mental illness for female jail inmates is 31 % (2009). The Bureau of Justice Statistics reported that the rate of suicide for female inmates in both jails and prisons between 2000 and 2002 (32 per 100,000 and 10 per 100,000, respectively) is lower than that for males (Mumola 2005). Recent rates using data from 2000 to 2007 reflect a decrease in the rate of suicide deaths for female jail inmates (28 per 100,000) (Noonan 2010). However, the suicide rate still remains five times higher than the general population when these rates are demographically adjusted (Noonan 2010). Noonan also reports that incarcerated Hispanic females have a rate of suicide 10.5 times greater than the rate of suicide for Hispanic women in the general population (2010). Incarceration does appear to substantially increase the risk for suicide among women.

Research on incarcerated female suicides has been sparse. Clements-Nolle reported that one risk factor for suicide attempts in prison is childhood trauma (2009). Marzano found that in English and Welsh prisons, there was a higher prevalence of current depression, two or more psychiatric diagnoses, history of inpatient psychiatric hospitalization, and history of suicide attempt in women with near-lethal self-directed violence compared to a control group (2010). Fazel hypothesized that substance use disorder and loss of contact with dependent children may contribute to incarcerated female suicide risk (2009).

Veterans

Another special population is veterans. Estimates suggest that nearly 10 % of incarcerated persons are veterans, although such estimates were derived from data that probably fails to fully reflect the impact of current conflicts in the Middle East. Clinical experience among those familiar with the population of returning veterans suggests that such numbers may be on the rise (Mumola 2000; Noonan 2004). The Veterans Administration now officially recognizes that veterans are arrested for a variety of offenses, some of which may be related to extended periods of battle readiness and combat exposure during multiple deployments and to maladaptive coping with the return to civilian life. The VA's Uniform Mental Health Services Package now calls not only for assistance for veterans reentering the community from state and federal prisons but also for outreach efforts to veterans who are interfacing with jails, courts, and law enforcement and for education to these agencies regarding mental health problems relevant to veteran populations, such as posttraumatic stress disorder (PTSD) and traumatic brain injury (TBI). Unfortunately, relatively little is known about the subpopulation of incarcerated veterans, including their suicide rates. It has been argued that suicide risk factors faced by the general veteran population may interact with risk factors specific to incarceration or justice involvement, potentially placing the incarcerated veteran at especially high risk for suicide. Additional research is needed to better characterize this potentially high-risk population to determine rates and opportunities for intervention (Wortzel et al. 2009).

Risk After Release

The suicide risk associated with incarceration unfortunately does not end upon release. Rather, research suggests that the time of release represents yet another difficult transition period that carries increased risk for mortality in general and increased risk for suicide specifically. A study by Pratt examined suicide following prison incarceration in England and Wales (2006). Compared to the rate of suicide of 18 per 100,000 in the general population, the rate for recently released inmates was 348 per 100,000, with 21 % of the suicides occurring within 28 days of release. Suicide risk was also found to be high for individuals after release from New South Wales, Australia, prisons (Kariminia et al. 2007). The suicide rate for individuals within 2 weeks of release from prison was found to be 3.87 times higher when compared with the suicide rate for individuals 6 months after release. Binswanger conducted a retrospective cohort study of inmates released from the Washington State Department of Corrections (DOC) and reported an adjusted risk of death among former inmates 3.5 times higher than that faced by other state residents (2007). More specifically, former inmates had a 3.4 relative risk for suicide, a 10.4 relative risk for homicide, and a 12.2 relative risk for accidental overdose. Meaningful efforts to mitigate suicide risk among recently incarcerated individuals will necessarily involve linkage to appropriate resources in the community.

Conclusions

The ongoing high risk of suicide death in correctional settings compared to the general population has resulted in more attention by the academic community, the courts, and correctional settings. This attention has led to a better understanding of the risks that exist in these settings and pragmatic solutions to decrease suicide risks. More recent studies have suggested that the rate of suicide in correctional settings has decreased over time. However, the decreased rates continue to remain higher than the general population rates, suggesting that efforts need to continue to find solutions to further reduce correctional suicide rates. The academic examination of suicide death and self-directed violence in correctional settings involving adolescents, women, and veterans is relatively new and is an area requiring further study. The high suicide risk/rate of recently released inmates from correctional settings is an area that is ripe for further study. Future studies could help provide direction in formulating a plan to target and reduce the high suicide risk/rate in this subpopulation. Finally, this entry demonstrates the responsibility that correctional institutions have in further developing policies and procedures that aim to reduce the risks of inmate suicide.

Related Entries

- ▶ Drug Abuse and Alcohol Dependence Among Inmates
- ▶ History of Substance Abuse Treatment
- ▶ Mental Health Courts
- ▶ Mortality After Release from Prison
- ▶ Surveys on Violence Against Women
- ▶ Treating Mentally Ill Offenders

Recommended Reading and References

Anasseril DE (2006) Preventing suicide in prison: a collaborative responsibility of administrative, custodial, and clinical staff. J Am Acad Psychiatry Law 34(2):165–175

Anasseril DE, Fleming J (2006) Suicides in a state correctional system, 1992–2002: a review. J Correct Health Care 12(1):24–35

Binswanger IA, Stern MF, Deyo RA, Heagerty PJ, Cheadle A, Elmore JG et al (2007) Release from prison – a high risk of death for former inmates. N Engl J Med 356(2):157–165. doi:10.1056/NEJMsa064115 [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]

Bradley-Engen MS, Cuddeback GS, Gayman MD, Morrissey JP, Mancuso D (2010) Trends in state prison admission of offenders with serious mental illness. Psychiatr Serv 61(12):1263–1265. doi:10.1176/appi.ps.61.12.1263 [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]

Clements-Nolle K, Wolden M, Bargmann-Losche J (2009) Childhood trauma and risk for past and future suicide attempts among women in prison. Womens Health Issues 19(3):185–192. doi:10.1016/j.whi.2009.02.002 [Research Support, Non-U.S. Gov't]

Fazel S, Benning R (2009) Suicides in female prisoners in England and Wales, 1978–2004. Br J Psychiatry 194(2):183–184. doi:10.1192/bjp.bp.107.046490 [Research Support, Non-U.S. Gov't]

Fazel S, Cartwright J, Norman-Nott A, Hawton K (2008) Suicide in prisoners: a systematic review of risk factors. J Clin Psychiatry 69(11):1721–1731 [Meta-Analysis Review]

Felthous AR (2011) Suicide behind bars: trends, inconsistencies, and practical implications. J Forensic Sci 56(6):1541–1555. doi:10.1111/j.1556-4029.2011.01858

Glaze LE (2010) Correctional population in the United States, 2009. U.S. Department of Justice Statistics

Hayes LM (2009) Juvenile suicide in confinement – findings from the first national survey. Suicide Life Threat Behav 39(4):353–363. doi:10.1521/suli.2009.39.4.353 [Multicenter Study Research Support, U.S. Gov't, Non-P.H.S.]

Hayes LM, National Institute of Corrections (U.S.), National Center on Institutions and Alternatives (U.S.) (2010) National study of jail suicide: 20 years later. U.S. Dept. of Justice, National Institute of Corrections, Washington, DC

James DJ, Glaze LE, United States Bureau of Justice Statistics (2006) Mental health problems of prison and jail inmates. U.S. Dept. of Justice, Office of Justice Programs, Bureau of Justice Statistics, Washington, DC

Joiner TE (2005) Why people die by suicide. Harvard University Press, Cambridge

Kariminia A, Law MG, Butler TG, Levy MH, Corben SP, Kaldor JM et al (2007) Suicide risk among recently released prisoners in New South Wales, Australia. Med J Aust 187(7):387–390 [Research Support, Non-U.S. Gov't]

Kerr DC, Washburn JJ, Feingold A, Kramer AC, Ivey AZ, King CA (2007) Sequelae of aggression in acutely suicidal adolescents. [Research Support, N.I.H., Extramural]

Lamb HR, Weinberger LE (2005) The shift of psychiatric inpatient care from hospitals to jails and prisons. J Am Acad Psychiatry Law 33(4):529–534

Leese M, Thomas S, Snow L (2006) An ecological study of factors associated with rates of self-inflicted death in prisons in England and Wales. Int J Law Psychiatry 29(5):355–360. doi:10.1016/j.ijlp.2005.10.004

Liebling A (2006) The role of the prison environment in prison suicide and prisoner distress. In: Dear GE (ed) Preventing suicide and other self-harm in prison. Edith Cowan University, Australia

Marzano L, Fazel S, Rivlin A, Hawton K (2010) Psychiatric disorders in women prisoners who have engaged in near-lethal self-harm: case-control study. Br J Psychiatry 197(3):219–226. doi:10.1192/bjp.bp.109.075424 [Research Support, Non-U.S. Gov't]

Mumola C (2000) Bureau of Justice Statistics Special Report, veterans in prison or jail. (178888). Retrieved from http://www.ojp.usdoj.gov/bjs/pub/pdf/vpj.pdf

Mumola CJ, United States Bureau of Justice Statistics (2005) Suicide and homicide in state prisons and local jails. U.S. Dept. of Justice, Office of Justice Programs, Washington, DC

Noonan M (2004). Veterans in state and federal prison. (http://www.ojp.usdoj.gov/bjs/abstract/vsfp04.htm)

Noonan M, United States Bureau of Justice Statistics (2010) Mortality in local jails, 2000–2007. U.S. Dept. of Justice, Office of Justice Programs, Bureau of Justice Statistics, Washington, DC

Patterson R, Hughes F (2008) Review of completed suicide in the california department of corrections and rehabilitation, 1994–2004. Psychiatr Serv 59(6):676–682

Penn JV, Esposito CL, Schaeffer LE, Fritz GK, Spirito A (2003) Suicide attempts and self-mutilative behavior in a juvenile correctional facility. J Am Acad Child Adolesc Psychiatry 42(7):762–769. doi:10.1097/01.CHI.0000046869.56865.46 [Research Support, Non-U.S. Gov't]

Polluck LR, Williams MG (2006) Psychological perspectives on intentional self-harm. In: Dear GE (ed) Preventing suicide and other self-harm in prison. Edith Cowan University, Australia

Pratt D, Piper M, Appleby L, Webb R, Shaw J (2006) Suicide in recently released prisoners: a population-based cohort study. Lancet 368(9530):119–123. doi:10.1016/S0140-6736(06)69002-8 [Research Support, Non-U.S. Gov't]

Roberts AR, Bender K (2006) Juvenile offender suicide: prevalence, risk factors, assessment, and crisis intervention protocols. Int J Emerg Ment Health 8(4):255–265

Scott CL (2010) Legal issues regarding the provision of care in a correctional setting. In: Scott CL (ed) Handbook of correctional mental health (2nd edn). American Psychiatric Publishing, Inc., Washington, DC, London

Steadman HJ, Osher FC, Robbins PC, Case B, Samuels S (2009) Prevalence of serious mental illness among jail inmates. Psychiatr Serv 60(6):761–765. doi:10.1176/appi.ps.60.6.761[Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]

Way BB, Miraglia R, Sawyer DA, Beer R, Eddy J (2005) Factors related to suicide in New York state prisons. Int J Law Psychiatry 28(3):207–221. doi:10.1016/j.ijlp.2004.09.003

White TW, Schimmel DJ, Frickey R (2002) A comprehensive analysis of suicide in federal prisons: a fifteen-year review. J Correct Health Care 9(3):321–345

Wolff P, Nancy L, Shi J, MS (2010) Trauma and incarcerated persons. In: Scott CL (ed) Handbook of correctional mental health (2nd edn.). American Psychatric Publishing, Inc., Washington, DC, London

Wortzel HS, Binswanger IA, Anderson CA, Adler LE (2009) Suicide among incarcerated veterans. J Am Acad Psychiatry Law 37(1):82–91 [Comment Review]

Supervision

▶ Desistance and Supervision

Surveillance Technology and Policing

Gavin J.D. Smith
Senior Lecturer in Sociology, The Australian National University, Canberra, ACT, Australia

Overview

Policing is in a state of transformation. Traditional paradigms founded on reactive and responsive principles are being superseded by contemporary risk-based models, proactive and preemptive approaches that encompass an array of stakeholders and sophisticated technologies. The "new policing" is underpinned and driven by intelligence-led doctrines, performance indicator metrics, and desires to better identify crime patterns. Although many relational factors account for these changes in policing ideology and praxis, the integration of – and reliance on - surveillance technologies and associated "visibility" rationalities has been particularly pivotal. Yet the widespread availability of surveillance technologies is simultaneously diversifying policing practices and responsibilities. Civil/commercial organizations and citizen/consumers are, for example, increasingly participating in the inspection and assessment of behavior, albeit for often differing purposes. The perpetual interweaving of visibility-inducing systems into the cultural fabric of everyday living and concomitant heterogeneity of actors who now engage in regulatory activities has generated two major effects: (a) policing opportunities have increased; and (b) supervisory webs have become progressively intricate and multi-dimensional in semblance. "Policing" is now a pluralized social practice that extends far beyond "the police." Indeed, the "policing creep" emerging from intensifying visibilities generates for criminologists important theoretical, empirical and ethical questions, some of which inform the substantive focus and thematic arrangement of this entry.

In order to arrive at an understanding of the surveillance technology-policing nexus, this entry commences with an overview of the emerging techniques of visibility now infusing the social orders of privileged global provinces. This will enable us to better grasp the assortment of observational devices being routinely deployed, the heterogeneous nature of operational procedures, and the differing populations now engaged in the policing of subjects. It will also permit us to comprehend some emergent social effects. A case study analyzing closed-circuit television (CCTV hereafter) camera system growth in the global north will then show empirically how surveillance technologies can modify and transform existing policing structures. This section draws further attention to the evolving labor force now performing, in this case, paid policing duties.

This entry, therefore, is not a story about "surveillance and the police" per se. Rather, it is an account describing how the widespread distribution of surveillance technologies ruptures previous modes of social control and instigates novel policing configurations. Criminologists are only beginning to comprehend the social implications attendant on these changes.

Surveillance Technologies and Rationalities

The rapid and extensive diffusion of surveillance technologies into the architecture of contemporary living has, in recent decades, been a particularly salient development in social governance. Everyday actions are targeted by an expanding arsenal of monitoring devices. Indeed, it has become well-nigh impossible to complete commonplace tasks in privileged wordly regions without encountering a diverse set of data capturing "probes." CCTV camera systems, airport body scanners, DNA swabs, X-ray scanners, and Web browsing cookies, for example, perpetually extract personal information "particles" from the data trails that accompany our daily activities. Tracking and transmission technologies such as radio-frequency identification (RFID) microchips and Global Positioning Systems (GPS) are now embedded in all manner of mundane objects: credit cards, cell phones, and smart passports being obvious illustrations.

Despite embodying differences in terms of technical components and physical dimensions, there are general similarities between and among surveillance technologies on an operational front. Most function to procure, transport, and then ultimately deposit data emissions derived from embodied acts – that is, from walking down a street, making an online purchase, and sending a SMS text message – into a complex cybernetic matrix commonly known as "the database." These vast informatic chambers decontextualize behaviors both by systematically removing them from their circumstantial origin and by rendering behavioral expressions into a series of discrete data flows. A person's consumption profile, for instance, can be inferred from analyzing her or his credit card statements. Buying habits, in this example, are deemed to reveal social status and thus customer "worth." Rather than take into consideration the infinite causal variables influencing the decision to purchase a product, these systems focus attention instead upon tangible outcomes and seek to aggregate data from multiple sources so that "tendencies" can be ascertained. Similar to how social research data is arranged, personal information is recurrently coded, collated, mined, or "recontextualized" by a set of surveillance workers – be they human (e.g., a CCTV operator) or nonhuman (e.g., a computer code) – in accordance with an assortment of organizational imperatives. This allows for historical activities to be logged, social relations to be mapped, actions to be compared, and causal patterns to be identified. Based on such information, risk/return evaluations of specific phenomena can be undertaken and decisions reached, for example, whether or not to invest capital in a specific enterprise, whether or not to trust a job applicant, and whether or not to proffer credit to a prospective customer. Direct human mediation in data analysis and processing is declining as systems of surveillance develop in capacity and sophistication. Computerized software is increasingly favored to autonomously dispose the information collected.

Surveillance technologies are a rationalized means to enhance visibility, to sharpen decision-making, and to more effectively manage resources, helping their overseers determine risk propensities, investment potentials, trustworthiness, and suspiciousness. They expose and scrutinize discrete social processes (such as disease intensity, voting patterns, or market relations) and help identify probable correlations between and among variables. They permit interest groups to accumulate intimate knowledge about specific phenomena (e.g., population health in specific regions) and to fashion corresponding interventions from an informed perspective (e.g., the proposed location of a health clinic). They provide a calculative lens through which all manner of truth claims can be assessed and adjudicated (e.g., an asylum seeker application or a credit request). As such, surveillance operates at the very heart of organizational protocol, specifically as it (a) enables social complexity in all its manifestations to be captured and (b) facilitates complexity governance and uncertainty reduction. Indeed the visibility – that is, the capacity to see and be seen – produced by systems of surveillance is fundamentally transforming the conditions of contemporary social life and how we relate to one another. Desires to know, order, govern, and entertain are exposing people and processes in much finer detail, and this situation is both diversifying and intensifying formal and informal policing arrangements.

Surveillance technologies reveal previously concealed facets of life and circulate mediated personal intimacies to an audience that increasingly is pluralized, remote and anonymous. An example of this process can be discerned from examining Web 2.0 engagement. Social media sites like Facebook proactively encourage users both to self-disclose/self-promote and to peruse the exteriorized musings ("newsfeeds") of fellow participants. User curiosity and voluntarism is exploited and manipulated in strategic ways, system programmers systematically designing and assembling an array of infrastructural mechanisms to sustain interest levels – and thus asset sustainability and value. Incentivizing architectures ensure that acts of revelation are celebrated and rewarded and that tracking the lives of others is experienced as an addictive and pleasurable pastime. The pellucid nature of these online registries permits emotive posts and graphic images to be accessed and deciphered by an army of unseen inspectors (e.g., peers, employers, marketers, and law enforcement agencies), the biographical information deposited revealing vital clues about a particular person's ideological values, consumer preferences, and psychological disposition. This knowledge can then be utilized by the various spectators to construct character profiles and to rationalize interventions, specifically procedures aimed at modulating and engineering conduct in some desired way (Trottier 2012).

We can perceive from this illustration the contemporary heterogeneity of policing, in terms of contributors and purposes. The operational reach and remit of policing continues to broaden as surveillance technologies extend purviews deeper into the backstage regions of personal life. Indeed, it is helpful to conceive people as "cyborg" entities, comprised of physical biomatter and informatic particles: the latter material being willingly emitted – and coercively extracted – from our embodied actions in the course of everyday living. The point to note is that personal data has become a highly valued unit of currency for problem analysis and decision-making, its textual content shaping significantly how particular individuals and social groups are perceived and subsequently treated by differing organizational systems. Rather than trusting a person's narrative or status claim at face value, interlocutors increasingly prefer to discern legitimacy and authenticity from the information encased within a set of corresponding archival records or what I term, "surveillance texts." Although often derived from certified or valid documents – for example, medical consultation notes, criminal records, bank statements, and tax returns – employers, law enforcers, and insurance actuaries also obtain knowledge fragments from unofficial and aspatial communicational mediums like Twitter and Facebook, and from personal email correspondence. Each source comprises discrete "event descriptors" that, when collated, depict previous conduct (e.g., consumption preferences) and circumstances (e.g., health conditions). It is believed that analysis of multiple surveillance texts will reveal (a) a detailed picture of a person's disposition and (b) a more accurate representation of her or his social positioning and future trajectory. This permits the adjudicating authority to effectively bypass or short-circuit what is perceived as the contrived nature of self-presentations, unveiling instead preconscious forms of cognition (Andrejevic 2010). In a world increasingly defined by information glut and uncertainty (a situation ironically accentuated by the intensification of visibility), surveillance systems act as critical arbitration tools, a means for truth establishment and narrative verification, a means for individuating and stratifying populations, and a means for regulating access to resources.

Surveillance systems mostly operate in hegemonic ways, functioning to maintain the interests of a privileged elite and thus a particular fiscal status quo. Yet the gradual diffusion and sedimentation of surveillance should not be understood as some ideological conspiracy that has been engineered by a state-corporate cartel. In contrast, a surveillance system or device intrinsically is neither benign nor malign, the sociocultural relations in which the apparatus is entangled effectively shaping its everyday orientation (see Smith 2013). Indeed, state authorities and administrators depend upon an assortment of surveillance machineries and practices to ensure the deliverance of democratic commitments and ideals: that voting rights are upheld, that law and order is maintained, that taxation thresholds are accurately set, and that designated populations are granted access to welfare services and benefits. Surveillance technologies, for example, are regularly utilized to assess and manage a nation's health. Disease epidemiologists use biostatistics to identify and gauge immanent biosecurity threats, while medical practitioners use hi-tech screening equipment to better envision patients' bodies so that malignant cells can be located and tissue injuries reliably evaluated. Law enforcers employ a range of surveillance techniques to investigate suspects, to acquire evidence and to direct operations. Moreover, surveillance is exploited strategically as both a thematic currency and a textual commodity by an assortment of cultural producers, be they formal commercial organizations or civic communities. People are regularly exposed to surveillance symbolism and output in the course of consuming a range of media broadcasts. Surveillance texts often depict graphic content the witnessing of which can be both shocking and entertaining. As a distinctive subject-matter, surveillance is artfully appropriated by groups possessing a diversity of political agendas, artistic aesthetics and economic motivations. The ensuing struggle over surveillance representations in popular culture encourages audiences to approach surveillance technologies with an attitudinal cynicism and ambivalence stimuli and representations. Surveillance is also deployed to increase the speed and simplicity of transactions between individuals and organizations and to enhance the flexibility, convenience, and efficiency of these encounters. In addition, surveillance is fast becoming a medium through which powerful groups can be made more accountable for their actions. Footage captured on a G-20 protestor's personal video camera, for instance, revealed a civilian being unlawfully assaulted by a police officer (see http://www.guardian.co.uk/uk/2009/apr/07/ian-tomlinson-g20-death-video, accessed 18/5/2012).

Despite notable exceptions – where surveillance technologies are harnessed to serve the interests of vulnerable individuals and periphery collectives – data-hungry social and commercial institutions are the overwhelming winners of surveillance ubiquity. This is because capturing and analysis technologies afford these agencies statistical knowledge on social trends and attitudes, allowing evidence-based responses – in terms of policy or commodity design – to be instrumentally engineered and actualized. Surveillance infrastructures and circuitries greatly abet elite groups, to the extent that they help to legitimize their domination, accentuate their expertise, foster their definitional capacities, and protect their capital, interests and resources.

Indeed, a core operational objective is the differential classification and division of groups in accordance with their perceived position in the social hierarchy, their projected risk level, and their assigned market value (Gandy 1993; 2009). Thus, in an economic order defined by mass consumption and in a political order characterized by omnipresent security concerns, population distrust, and risk-sensitive policies, it is people residing in disadvantaged locales without the means to consume, or those whose ethnicity, piousness, skin color, and occupation contravene a dominant (normative) security discourse, who are recurrently subject to disproportionate levels of scrutiny (Norris and Armstrong 1999; Lyon 2003). Procurement of a stigmatizing or discrediting label can mean that an individual or collective is denied access to certain spaces and/or refused basic rights, resources and privileges (Monahan 2008; Zureik et al. 2011). Thus, because surveillance systems embody the prejudicial essentialism of their controllers, they tend to produce exclusionary outcomes and perpetrate social injustices.

The Diversification of Policing Responsibility

A major outcome of neoliberal policy has been an aspiration by elite groups to instill a "responsibility ethic" and "cosmopolitan sensibility" within the psyches of the population: an expectation that citizens play an active role in managing risk. This expectation primarily relates to personal health and safety practices but also to deportment in the economic and ecological spheres. In the law and order field, adoption of a "responsibilization" rationality (see O'Malley 1996) by state authorities has profoundly transformed the meaning of "criminality" and "citizenship." What historically had been a matter for formalized policing agencies has now become a community issue demanding multiagency interventions; policymakers eager to enlist in quotidian supervision practices both civil society actors and the private sector. Criminality itself has been discursively reconstituted as a set of hierarchical thresholds, ranging from lower level civil transgressions to higher order terrorist/biosecurity risks. This stratification – and concomitant governmental desire to mobilize as "crime managers" various groups – has produced a distinctive set of stakeholders (e.g., citizens, proprietors, and council workers) who are now equipped with specialized powers – and ascribed specific obligations – for addressing particular expressions of disobedience (Rose 2000; Garland 2001).

The emergence of civil "on-the-spot" penalty notifications, laws persecuting parents for child truancy, community "wardens," neighborhood watch programs, and dispersal/antisocial behavior orders are notable empirical instances of a general trend to prescribe codes of etiquette and to assign any violation – or manifestation of deviance – with a particular regulatory custodian (e.g., a parent, a schoolteacher, a social worker, or a transport marshal). Indeed, the diversification of actors now responsible for everyday surveillance and social ordering marks a distinct paradigm transition in policing ideology and application, from a traditional model which was largely "reactive and localized/centralized," that is, post facto and exclusive, to one which is now "proactive and dispersed/pluralized," that is, anticipatory and inclusive (McLaughlin 2007).

Such changes in policing praxis correlate with the neoliberal state's remit to buttress declining security budgets/provision and to "govern at a distance" (Miller and Rose 1990) with a corresponding engagement (or exploitation) of community groups and corporate agencies in the collective "fight against crime" (Garland 2001). They also accord with an allied desire to sponsor the implementation of hi-tech surveillance infrastructures which operate according to intelligence-led axioms and preemptive principles (see Elmer and Opel 2008). The Texas Virtual Border Watch program is an apposite illustration (http://blueservo.com/vcw.php, accessed 30/5/2012). This online initiative effectively transforms citizens into custodial guardians of the United States–Mexico border. It encourages participants, through financial incentives, to identify and report suspicious mobilities and to adopt a nationalistic sensibility: a cognitive disposition that is hostile to the context prompting a decision to enter illegally a foreign sovereign territory. An uncompassionate and uncompromising mentality orientated to informing and supervising is, in this process, deemed to be a desirable civil ethic (Koskela 2011). These policy transformations (e.g., enlisting communities in addressing lower order criminality) simultaneously grant formal policing organizations special license to concentrate finite resources on tackling serious and organized crime and an assortment of high-level national security threats. Moreover, it is in this transitional climate that policing authorities have increasingly adopted surveillance technologies as ways of managing shrinking quotas and the complexities associated with transnational law enforcement in the information age.

The rolling out of large-scale socio-technical assemblages such as DNA databanks and CCTV camera networks demonstrates how responsibilities for managing crime and disorder are transferring from a dependence on theological precepts, authoritarian physicality, and human informants to a faith in scientific expertise and technical solutions. The fetishization or overprivileging of scientific discourse and technical systems by governmental actors, while partly operational, is in large part symbolic and in response to a perceived "crisis" in traditional policing approaches, brought on by austerity measures, rising crime fears, and high profile institutional failings/operational incompetence (McLaughlin 2008). Utilizing sophisticated technological infrastructures and computer-powered calculative metrics permits policing agencies to identify criminogenic patterns and to anticipate prospective security threats (Ericson and Haggerty 1997). It is now possible to collate previously discrete data fragments and create "criminality constellations." The Metropolitan Police's crime map (http://maps.met.police.uk/, accessed 20/5/2012) is an apt example. Based on the quantitative aggregation of crime reports and statistics, policing authorities are able to spatially and temporally envision felony rates and correlations within the urban metropolis. This ensures that any proposed crime prevention initiative or awareness-raising campaign has an empirical foundation and is measurable. It also means that resources can be rationally deployed in strategic and targeted ways and decision-making protocol as it relates to criminal investigations can be externally validated and legitimized. It additionally permits bureaucratic officials and police managers to set performance targets and impose evaluative frameworks.

Evidently, these applications have multiple users and serve numerous purposes. Insurance companies, for example, are keen to exploit crime data to calibrate premiums, privileged citizens to distinguish (un)desirable locations for settlement, and marketers to direct advertising toward amenable communities (and away from others). The employment of actuarial techniques by formalized policing entities means that, based on available indicators, persons and groups acquire a risk rating (Feeley and Simon 1994). This computer-generated calculation effectively comes to determine the type and intensity of scrutiny that is then actioned. Gradual erosion of the "innocent until proven guilty" legal criterion has been an outcome of these developments and wider anti-social behavior and anti-terror policies. Increasingly, policing logic has more unity with a "guilty until proven innocent" aphorism, as shown in recent UK legislation endowing extraordinary preventive and regulatory powers to police forces. These laws grant police officers special license to "randomly" stop and search individuals, to disperse groups of young people, to preemptively detain terror "suspects" without charge, and to demand that protestors register their personal details before attending demonstrations. They also permit law enforcers to acquire sensitive personal data derived from the activities of law-abiding persons (http://www.bbc.co.uk/news/technology-18102793, accessed 20/5/2012) and to ascribe "suspect status" to an individual on the basis of her or his communicative position within a criminalized informatic network (http://news.bbc.co.uk/2/hi/technology/8245056.stm, accessed 24/5/2012). Similarly, landmark statutes such as the USA Patriot Act (2001) bestow exceptional powers to national security agencies in terms of intelligence gathering and information sharing protocol, and facilitate the introduction of ever more intrusive surveillance measures. These bills tend to be passed with limited public consultation or debate in moments of intense emotionality, for example, in the immediate aftermath of a terrorist act or in advance of mega events like the Olympic Games or a G20 convention, when anger and fear – rather than critical rationalism and reflexivity – orientate the collective consciousness and the political will.

Surveillance Technologies and Surveillance Work

The previous sections described how and why governmental desires and reality-capturing devices have extended policing practices into the circuitries of everyday living. CCTV is an insightful example of how policing praxis has co-evolved with technological supplementation, specifically in terms of the emerging complexity of social relations and operational remits that now constitute such regulatory systems. There are additional reasons for considering CCTV. Focusing on the exponential growth of camera networks in the global north illuminates the complex politics and rationalities underpinning surveillance diffusion. Moreover, considering novel policing modalities such as CCTV focuses attention on the unique type of labor – or “surveillance work” – that the camera operators routinely practice. These “order custodians” are subsidized to remotely encounter, capture, and “manage” mediated scenes of criminality and disorder, transgressive spectacles to which such workers are repeatedly exposed but over which they have no direct control. Reflecting on how the introduction of CCTV cameras transforms policing practices in both intentional (official) and unintentional (informal) ways provides clues for envisaging the likely impact of additional surveillance technologies on conventional policing protocol.

CCTV provision has dramatically increased in recent years. This is particularly true of the UK, a world leader in terms of its camera coverage. While a complex politics, at local, regional, and national levels, surrounds the augmentation process (see Fussey 2007; Norris and Armstrong 1999), the 1980s economic recession was a decisive factor in the state’s decision to legislatively promote and financially sponsor a series of installation programs. Rising crime rates associated with growing unemployment, a declining welfare system and resource inequities, and a corresponding amplification of the issue by media reportage impelled a public perception that policing agencies were failing in their bid to control criminality. Successfully redressing this situation became a central political issue, Home Office officials desperate to introduce a progressive “silver bullet” solution that would enhance the policing service and “treat” in an effective manner the physical symptoms of structural disparities (note similarities between this scenario and the one depicted in science fiction movie, *Robocop* (1987)). The intelligence battles of the Cold War period meant that an array of advanced militarized surveillance technologies were available for civil market appropriation. CCTV was one such innovation (Norris and Armstrong 1999).

Preliminary evaluations of pilot projects in several jurisdictions showed that public space camera networks could reduce particular acts of disobedience by around 40%. Although this research was methodologically problematic and predominantly administered by self-interested practitioners, the extraordinary statistics persuaded ruling technocrats that CCTV camera networks were worthy of significant financial investment (Gill and Spriggs 2005). Cities were encouraged to compete with one another for allocated monies totalling millions of pounds, and camera systems rapidly became akin to a “fifth utility” service (Graham 2001).

The favouring of CCTV systems by urban authorities was also the outcome of neoliberal “urban renaissance” policies which advocated restructuring of the metropolis around commercial imperatives and resiliency ideals (Atkinson and Helms 2007; Coaffee et al. 2009). This principally involved spatial revitalization, market liberalization, and the assembly of vibrant, leisure economies. Manipulating consumer perception became a central directive of the project, and CCTV cameras were deployed to help promote an impression of safety that would facilitate increased consumption and a "feel good factor".

A further factor fostering CCTV's wider cultural appeal was the global circulation of images showing Merseyside toddler, Jamie Bulger, being abducted by two ten-year-old boys in a Bootle shopping mall (Norris and Armstrong 1999). Although the cameras did nothing to prevent the child's violent death, the moral outrage and anxiety associated with this act as symbolized by the grainy pictures was, for different reasons, exploited by an array of state, media, and industry actors. Notwithstanding an assortment of competing rationales and rationalities, each agency had a vested interest in sponsoring CCTV deployment. The government had an interest in endorsing system efficacy to accrue public support and thus political capital/leverage. The police had an interest in supporting CCTV implementation as cameras could bolster conviction rates, sharpen operational activities and enhance investigative practices. The media had an interest in acquiring "authentic" camera footage to both disturb and amuse a scopophilic audience, that is, an audience obsessed with the spectacle. The security industry had an interest in promoting CCTV's effectiveness to generate increased demand for the technologies and for training programs.

The symbolism embodied in these initiatives is worth reflecting upon. The cameras act as materialized signifiers of sovereign authority and symbolize a profound distrust of the public by ruling groups. They cultivate conformity to the consumer order and, in the process, reify dominant definitions of appropriate urban conduct. They become the instruments through which discrediting labels can be ascribed to and enacted upon particular stigmatized groups (Norris and Armstrong 1999). They operate as architectural barriers for rationally thinking criminals to circumnavigate, obstacles hindering the ease with which an act of criminality can be perpetrated in a particular locale. Camera networks are aesthetic additions to the urban landscape and represent a city's prosperity and progressiveness (Graham 2001). They become part of the political discourse on risk and security, and function as discursive horizons on which an array of propagandistic vistas can be fashioned. Such partisan conceptions tend to overemphasize the technology's disciplinary influence and to fetishize its role in crime reduction and public safety. CCTV camera initiatives are also a tokenistic reminder, at least in the public imaginary, that the state is actively engaged in addressing both the fear and actuality of criminality.

The above analysis demonstrates that surveillance technologies are much more than mere technical-material structures. It shows, in contrast, that they are also social enterprises. Although surveillance assemblages emerge as a result of political economic rationalities and comprise technical components, their everyday operational vitality and verve is entirely dependent on a complex chain of social actors and relations. Considering the role of workers in the daily production of surveillance is, therefore, an important part of the story. This focus permits us to discern the exploitative dimensions of policing reality, the artful practices comprising the operational lifeworld and the changing configuration of policing praxis.

Surveillance workers tend to occupy a privileged symbolic position, wielding discretionary interpretive powers and an anonymous and asymmetrical gaze. Recent research on the topic of surveillance work(ers) has illustrated (a) the extent to which such laborers impose their own definitional criteria upon that which is observed and (b) the consequences for those who are involuntarily exposed. Norris and Armstrong's (1999) ethnographic study of several CCTV control rooms in the UK is an illustrative example. The two scholars sought to establish the interpretive frameworks or "working rules" – effectively, subjective schemata – operators drew upon to determine camera positioning. The researchers show that suspicion is not an innate or discernible behavioral quality but is rather the outcome of social construction processes. It was racist, sexist, fascist, and classist ideologies that largely determined where and at whom cameras were pointed, the operators predominantly associating criminality with young males, minority groups, and underprivileged populations, specifically those not explicitly engaged in legitimized forms of consumption. These findings emphasize the categorical or labeling power surveillance workers exercise both in the arrangement of social reality and in concomitant decision-making protocol.

Another way of comprehending the social dimensions of surveillance work has been developed by Smith (see 2007, 2009, 2013). He found camera operators to be active consumers of social reality, attaching anecdotal stories to the mediated events encountered. The surveillance workers' musings comprised a mix of imagination and memory, and this creative interchange between the fictional and the real helped foster a set of ritualized occupational customs that made the trying work of watching a more pedagogically meaningful and pleasurable activity. It is easy to imagine and perceive "the gaze" projected by surveillance workers as inevitably embodying disciplinary properties and to assume watching practices as invariably empowering, rewarding and seductive. Yet surveillance work is ambiguous. It comprises important affective and exploitative components. Cameras operators, for example, are required to search for and identify events threatening a desired normative order and, as such, recurrently confront graphic scenes depicting social disorder and human suffering. They must dispassionately film people attempting suicide, injecting drugs, committing robbery, and being viciously assaulted. As a result of the labor that they practice and the traumatic spectacles to which they are habitually exposed, camera operators come to perceive the social world around them as chaotic, precarious, and dangerous. Being routinely subjected to mediated scenes of distress distils in such laborers a distinctive set of pathological beliefs about human conduct and relations, such values reorientating their everyday engagements with the world (Smith 2013).

A sequence of coping strategies are drawn upon by camera operators to manage their general estrangement both from the commodity that they are trying to produce (i.e., social order and its inherent fragility) and from the manufacturing process more generally (i.e., system technologies and occupational conditions). Cameras, for example, are used informally by operators in an attempt to escape, albeit temporarily, from the stresses and tensions associated with witnessing behavioral transgressions and social disarray. Nature becomes an attractive focal point. Its structural stability and intrinsic grace, as represented in dew-covered spider webs, cumulus cloudscapes, and majestic sunrises, helps operators reestablish a sense of trust and coherency in a world that so often appears fractured and anarchic. Thus, the same technologies fuelling the operators' general malaise with, and alienation from, social reality, also facilitate transcendental pursuits, in this case, identification of the natural order's aesthetical mechanics, internal consistency and harmonic patterns. Camera operators also partake in individualized and collective forms of emotional management. The application of black humor, bullying, and sensory detachment – and exploiting camera optics for voyeuristic ends, for instance – allows potentially pathological (and health degenerative) feelings to be channeled and mitigated (and to be projected onto external others).

Surveillance workers are situated within and stratified according to wider occupational hierarchies. They are increasingly embroiled in struggles over camera positioning/control and continually receive quasi indictments requiring them to publicly defend their productions in law courts (Smith 2009). Moreover, research by Goold (2004) and by Newburn and Hayman (2002) has shown how the introduction of CCTV into the public sphere and the custody suite has increased the visibility of policing practices and placed novel accountability pressures on law enforcement officers – and generated corresponding occupational dilemmas for camera operators. Despite the fact that CCTV networks are supposed to support the police in their pursuit of occupational goals, it is often the case that such systems challenge traditional structures in the police's occupational culture, particularly their monopoly on the legitimate use of violence. The routine exposure of policing practices has been further facilitated by the mass diffusion of portable capturing devices and by global circulations of information, especially as facilitated by Internet platforms such as Facebook and YouTube.

Future Directions

This entry has considered how emerging cultures of visibility transform the concept and practice of policing. In particular, the proliferation of surveillance devices into organizational circuitry and cultural canals has been shown to diversify the nature and meaning of policing. Institutions are becoming as much "the watched" as "the watchers." The same can be said of individuals. People embrace surveillance platforms to derive knowledge about social phenomena, to speed up their worldly transactions, and to be entertained. As such, persons increasingly find themselves experiencing surveillance regimes as both exposed "subjects" and contributory "agents." It would seem that, for many, scopophilia has become a seductive practice and voluntary disclosure a desirable state of being.

A plurality of conditions account for these transformations, and more research need examine the political economies, rationalities, and imaginaries underpinning the design and construction of surveillance technologies and attendant product marketing. Research on how different social groups understand, negotiate, and partake in surveillance activities is also required. Further criminological research, therefore, must excavate the nuances associated with surveillance "work," specifically how the introduction of computerized algorithms to autonomously determine patterns of causality in social reality – and to dispose specific interventions – demands a critical rethinking of how contemporary policing is increasingly performed. Today's world is defined by the visibility of particular social flows and the corresponding invisibility of general surveillance processes. The latent nature of computer processing and the overvaluing of technology means that power and domination are becoming harder processes to empirically locate, discern, follow, and – importantly – hold to account. If criminology is to remain analytically relevant to the rapidly evolving structures of social regulation, it must innovate methodologically and epistemologically, particularly appropriating knowledge and tools from quantum physics, mathematics and computing sciences. Policing now operates through informatic circuits, and it is to a better understanding of these cybernetic territories - as distinctive "fields of power" - that criminologists increasingly need focus their energies.

Related Entries

- ▶ Biometrics and Border Control Policing
- ▶ CCTV and Crime Prevention
- ▶ Crime Mapping
- ▶ Crime Science
- ▶ Criminal Profiling
- ▶ DNA Profiling
- ▶ DNA Technology and Police Investigations
- ▶ Electronic Monitoring
- ▶ Fingerprint Identification
- ▶ Forensic Science and Criminal Investigation
- ▶ History of Technology in Policing
- ▶ Identification Technologies in Policing and Proof
- ▶ Information Technology and Police Work
- ▶ Innovation and Crime Prevention
- ▶ Intelligence-Led Policing
- ▶ Prediction and Crime Clusters
- ▶ Predictive Policing
- ▶ Risk Assessment, Classification, and Prediction
- ▶ Telemetric Policing
- ▶ Use of Social Media in Policing
- ▶ Video Technology and Police Interrogation
- ▶ Visualizing Data: A Brief History

Recommended Reading and References

Andrejevic M (2010) Thin-sliced thoughts and theory's ends. Media Tropes 2(2):45–64

Atkinson R, Helms G (eds) (2007) Securing an urban renaissance: Crime, community and British urban policy. The Policy Press, Bristol

Coaffee J, Murakami-Wood D, Rogers P (2009) The everyday resilience of the city: how cities respond to terrorism and disaster. Palgrave Macmillan, London

Elmer G, Opel A (2008) Preempting dissent: the politics of an inevitable future. Arbeiter Ring Press, Winnipeg

Ericson RV, Haggerty KD (1997) Policing the risk society. University of Toronto Press, Toronto

Feeley M, Simon J (1994) Actuarial justice: power/knowledge in contemporary criminal justice. In: Nelkin D (ed) The future of criminology. Sage, London

Fussey P (2007) An interrupted transmission? Processes of CCTV implementation and the impact of human agency. Surveill Soc 4(3):229–256

Gandy O (1993) The Panoptic sort: A political economy of personal information. Westview, Boulder

Gandy O (2009) Coming to terms with chance: Engaging rational discrimination and cumulative disadvantage. Ashgate, Farnham

Garland D (2001) The culture of control: Crime and social order in contemporary society. Oxford University Press, Oxford

Gill, M. and A. Spriggs (2005) 'Assessing the impact of CCTV', Home Office Research Study 292, London: Home Office Research, Development and Statistics Directorate

Goold B (2004) CCTV and policing: public area surveillance and police practices in Britain. Oxford University Press, Oxford

Graham S (2001) CCTV: the stealthy emergence of a fifth utility? Plan Theory Prac 3(2):237–41

Koskela H (2011) Hijackers and humble servants: individuals as camwitnesses in contemporary controlwork. Theor Criminol 15(3):269–282

Lyon D (ed) (2003) Surveillance as social sorting: privacy, risk and digital discrimination. Routledge, London

McLaughlin E (2007) The new policing. Sage, London

McLaughlin E (2008) Hitting the panic button: policing/'mugging'/media/crisis. Crime, Media, Cult 4(1):145–154

Miller P, Rose N (1990) Governing economic life. Econ Soc 19(1):1–31

Monahan T (2008) Editorial: surveillance and inequality. Surveill Soc 5(3):217–226

Newburn T, Hayman S (2002) Policing, surveillance and social control: CCTV and police monitoring of suspects. Willan Publishing, Cullompton

Norris C, Armstrong G (1999) The maximum surveillance society: the rise of CCTV. Berg, Oxford

O'Malley P (1996) Risk and responsibility. In: Barry A, Osborne T, Rose N (eds) Foucault and political reason. UCL Press, London

Rose N (2000) Government and control. Br J Criminol 40(2):321–339

Smith GJD (2007) Exploring relations between watchers and watched in control(led) systems: strategies and tactics. Surveill Soc 4(4):280–313

Smith GJD (2009) Empowered watchers or disempowered workers? The ambiguities of power within technologies of security. In: Aas KF, Gundhus HO, Lomell HM (eds) Technologies of insecurity: the surveillance of everyday life. Routledge–Cavendish, London

Smith GJD (2013) Opening the black box: surveillance in everyday life. Routledge, London

Trottier D (2012) Social media as surveillance: rethinking visibility in a converging world. Ashgate Publishing, Farnham

Zureik E, Lyon D, Abu-Laban Y (eds) (2011) Surveillance and control in Israel/Palestine: population, territory and power. Taylor–Francis, Abingdon

Surveys on Violence Against Women

Kristiina Kangaspunta[1] and
Ineke Haen Marshall[2]
[1]United Nations Office on Drugs and Crime (UNODC), Vienna, Austria
[2]Department of Sociology and Anthropology and School of Criminology and Criminal Justice, Northeastern University, Boston, MA, USA

Synonyms

Domestic violence; Gender-based violence

Overview

The general consensus to measure violence against women is that population-based standalone surveys are the instruments of choice for collecting statistics on the victimization of women while recognizing that a well-designed module within a general or other purpose survey would be an appropriate tool as well. Several countries have measured violence against women using different types of survey methods. This entry examines different ways of measuring violence against women giving examples of national and multi-country surveys.

Introduction

The international community has been concerned with violence against women (VAW) or gender--based violence for a number of years. The first United Nations resolution on "Abuses against women and children" was adopted in 1982 calling upon the Member States to "take immediate and energetic steps to combat these social evils." The Convention on the Elimination of All Forms of Discrimination against Women was adopted by the United Nations General Assembly in 1979. The Committee on the Elimination of Discrimination against Women (CEDAW) monitoring the national implementation of the Convention defines gender-based violence as violence that is directed against a woman because she is a woman or that affects women disproportionately. It includes acts that inflict physical, mental, or sexual harm or suffering, threats of such acts, coercion, and other deprivations of liberty.

In 1993, the United Nations General Assembly adopted the Declaration on the Elimination of Violence against Women which also introduced the first internationally agreed upon definition on violence against women:

> For the purposes of this Declaration, the term "violence against women" means any act of gender-based violence that results in, or is likely to result in, physical, sexual or psychological harm or suffering to women, including threats of such acts, coercion or arbitrary deprivation of liberty, whether occurring in public or in private life. (United Nations 1993)

Soon after the adoption of the Declaration in 1993, the lack of reliable data on gender-based violence was recognized. The United Nations Beijing Platform of Action in 1995 recommended the collection of prevalence data to better understand different forms of violence against women. The recommendation has been repeated several times in different international forums.

The hidden nature of gender-based violence makes it difficult to measure its prevalence. Police records reflect no more than the tip of the iceberg. Victimization surveys have been seen as key to reveal the experiences of the victims of violence against women which could not be captured in the official records. The first large scale victimization surveys focusing on gender-based violence were carried out in the Netherlands and Canada. The prevalence rates for sexual violence and other violence were generally seen as shocking and caused disbelief. Soon after, other countries started to carry out surveys with similar results. Most violence against women surveys have been done at the national level. Comparative research involving several countries has proven to be problematic because of differences in cultural and legal definitions and data collection methods. However, there are now a growing number of comparative studies on different aspects of violence against women even though the number of countries participating in these surveys has been rather limited.

Measuring Violence Against Women

Indicators of Violence

Much of the research on violence against women has focused on domestic violence and intimate partner violence (IPV), but gender-based violence is a much wider issue. The United Nations, true to its commitment to human rights, has played an important role in broadening the concept. Gender-based violence can take many forms, and it can take place within or outside the family; it can occur in the workplace, the doctor's office, in state institutions, on the street, or other public places. The victims may be children, adults, or the aged, and the offenders can be total strangers or those that are known to the victim. Violence can take the form of harassment, emotional and psychological violence, economic abuse or exploitation, sexual and physical assault, or murder. Even though many forms of violence against women are criminalized, in many parts of the world women still experience violence outside the reach of the law. It is virtually impossible to capture this very broad range of harmful practices by any standardized measurement process (Kangaspunta and Marshall 2012).

Gender-based violence is a universal phenomenon which takes different forms in different cultural, social, and economic contexts. Our knowledge about the different forms of violence against women is spread very unevenly across the globe, the most developed countries often having the most sophisticated and elaborate survey data and record-keeping systems. This situation is likely to improve in the future, since important progress is being made on the development of global statistical indicators of violence against women.

In 2006, the United Nations adopted a resolution on "Intensification of efforts to eliminate all forms of violence against women" requesting the Member States to ensure systematic collection and analysis of data. Also, the Statistical Commission was requested to develop a set of possible indicators on violence against women in order to assist States in assessing the scope, prevalence, and incidence of violence against women.

Following the request of the resolution, the United Nations Statistical Commission adopted a list of indicators on violence against women which are based primarily on two criteria, that is, on the availability of data at the national level and the seriousness of the violence itself. The below list is suggested to be a starting point for initiating further work on identifying the most appropriate measurements:

1. Total and age-specific rate of women subjected to physical violence in the last 12 months by severity of violence, relationship to the perpetrator(s), and frequency
2. Total and age-specific rate of women subjected to physical violence during lifetime by severity of violence, relationship to the perpetrator(s), and frequency
3. Total and age-specific rate of women subjected to sexual violence in the last 12 months by relationship to the perpetrator (s) and frequency
4. Total and age-specific rate of women subjected to sexual violence during lifetime by relationship to the perpetrator(s) and frequency
5. Total and age-specific rate of women subjected to sexual or physical violence by current or former intimate partner in the last 12 months by frequency
6. Total and age-specific rate of women subjected to sexual or physical violence by current or former intimate partner during lifetime by frequency

In addition, harmful practices including female genital mutilation/cutting and early marriage were considered for inclusion in the indicators list. Also, a number of other manifestations of violence against women were identified that need to be assessed as possible topics for measurement, such as psychological and economic violence, stalking, physical and sexual violence in childhood, forced marriage, discrimination and economic violence at work, trafficking in women, impact of incidence of sexual violence against women on sexually transmitted diseases and HIV/AIDS, assessing risk factors, assessing the extent to which women recognize the suffered violence as a crime, and the percentage of hidden violence unreported to the authorities, or, indeed, even within the community.

It should be pointed out that currently only impressionistic and qualitative information is available for many countries. It is often collected by non-governmental organizations or, for those countries which lack accessible academic research on the topic, provided by the news media. In addition, in some countries we can glean information about gender-based violence through a review article by a local scholar (e.g., Simister et al. 2010; Wasileski and Miller 2010).

One of the most difficult tasks in measuring violence against women is to operationalize the definition of violence against women. During the1990s, there was a considerable controversy over the concepts and terms used in different studies. "Conflict tactics" was used to describe how couples try to settle their differences. Some other studies used rather "violence" and "force" to describe the incidences of violence against women. Also, the question on whether the act or the impact of the act should be measured created prolonged discussion (Walby and Myhill 2001).

Currently, it is widely accepted that surveys should use specific behavioral measures of a range of types of violent acts rather than general terms such as violence or assault. This can at least partly prevent different interpretations of violence among the respondents who might have individual and culture-bound understandings of what constitutes violence. For example, the International Violence against Women Survey used a list of concrete violent acts in its questionnaire including acts like throwing, hitting, pushing, grabbing, twisting arm, pulling hair, slapping, kicking, biting, strangling, suffocating, burning, scalding, and using a knife or gun. Also different acts of sexual violence were mentioned as well as the impact of violence and the steps victims take to obtain help (Johnson et al. 2008).

Measuring Prevalence

Official, administrative records, in particular police and court statistics, have as said limited capacity for measuring the prevalence and incidence of violence against women. Lack of victim willingness to report the incident to the police and reluctance by the police and court officials to record such reports in an adequate manner have historically been cited as impediments to using official data to measure violence and crime. Police statistics may be seen as measures of the criminal justice system's response to crime, rather than as measures of the crime itself. Police statistics about violent crimes against women such as sexual assault and rape, physical assault, and partner violence can be seen as indicators of social sensitivity and changing norms with regard to acceptable levels of interpersonal violence (see, for example, Mucchielli 2010). A number of countries now have specific legislation to record intimate partner violence or the more inclusive concept of domestic violence. However, these statistics primarily measure the social response of victims, bystanders, service providers, and the police to victimization (Kangaspunta and Marshall 2012).

The general consensus to measure gender-based violence is that population-based stand-alone surveys are the instruments of choice for collecting statistics on violence against women while recognizing that a well-designed module within a general or other purpose survey would be an appropriate tool as well. Several countries have measured violence against women using different types of survey methods. Targeted surveys usually cover all forms of violence against women including physical violence, sexual violence, intimate partner violence, and threat of violence. In some cases, also harassment and stalking have been studied.

Some countries use a specific module on partner violence which is added to general victimization surveys or to health surveys to assess gender-based violence. Also, a self-completion module about personal physical and sexual victimization has been added to the general victimization surveys. In some countries, only some forms of violence against women have been studied, usually addressing partner violence and sexual violence.

Surveys have also been targeted in different ways in different countries. In some countries, only women were included in the survey; in other countries also men could participate. Also age of the respondents differ: in most surveys the minimum age has been 18, while the maximum age has varied. The UN Statistical Commission recommends the surveys to be started from age 15.

The mode of interviewing has varied in different countries. Violence against women surveys have been conducted by interviewing respondents face-to-face or over the phone; also mailed questionnaires, self-completion on a computer, or combined methods have been used. While some research argues that the use of different methods might have an impact on the data collection, some others have not found much evidence for this. Face-to-face interviewing can build up more rapport between the interviewer and the respondent making it easier to speak about victimization experiences. On the other hand, the anonymity of telephone-based interviews and self-completed interviews may increase the likelihood to reveal sensitive information. In any case, the enquiry mode may have an impact on the response rate. In general, mailed questionnaires tend to have the lowest response rate compare to the other methods (Walby and Myhill 2001). However, the impact of the enquiry method might be different in different countries depending on cultural and other factors. For example, the first violence against women survey in Finland was conducted using a mailed questionnaire with a relatively high response rate of 70 % (Heiskanen and Piispa 1998, p. 8).

Violence against women data drawn from general victimization surveys should be interpreted with caution for many reasons. General surveys might not be able to include specific-enough questions on experiences of violence, and they may lack sensitive question wording. Also lack of selection of and training for interviewers needed to sensitively conduct interviews or the use of male interviewers may have an impact on the willingness to disclose painful memories of violence. Lack of efforts to facilitate victims to speak about their experiences such as ensuring privacy as well as the broad scope of crime victimization surveys not allowing adequately to address sensitive issues might prevent women to disclose their victimization experiences (Johnson et al. 2008, p. 12).

Also, data collection methods based on face-to-face or telephone interviews might lower the reporting of these crimes. When comparing different methods used in the British Crime Survey, it was noticed that prevalence rates for domestic violence from the self-completion method are around five times higher than rates obtained from face-to-face interviews (Flatley et al. 2010, p. 58).

The International Crime Victim Survey (ICVS) which has been carried out since 1989 includes questions on sexual violence which could be relevant in measuring violence against women. However, there are only small changes in the average sexual assault against women 1-year prevalence rate which varies between 0.6 % and 1 % in different sweeps. Measuring violence against women based on ICVS is not easy, and the figures should be interpreted with caution as mentioned above. Relatively low numbers of female respondents as well as low figures on sexual violence experiences make it difficult to ascertain statistically significant differences. Some findings also suggest that rates of sexual offences are less stable over the years than those of other types of crime. This may indicate that media campaigns or other measures can temporarily raise awareness about sexual violence which in turn increases the reporting of these incidents (van Dijk et al. 2007, p. 77). Even so, national victimization surveys (with larger samples) when used with due caution, may be a useful source for estimates of the rate of victimization of women.

National Violence Against Women Surveys

Targeted Stand-Alone Surveys

Targeted stand-alone surveys have been carried out in many countries. Examples of the first such surveys can be found in Australia, Finland, and the USA, besides Canada and the Netherlands. The first Women’s Safety Survey was carried out in Australia in 1996 to provide information on women’s safety at home and in the community and to depict the nature and extent of violence against women in Australia (Mc Lennan 1996). The survey was repeated in 2005 as part of the Personal Safety Survey in Australia (ABS 2005) updating the information of the first survey concerning women’s experiences of violence (Phillips and Park 2006). In Australia both male and female violence against women was studied, and the focus was on the experiences during the previous 12 months or during the relationship. In 1996, around 7 % of the respondents had experienced violence during a 12-month period; the corresponding figure in 2005 was around 6 % (ABS 2005).

In Finland, the first national survey on men’s violence against women and its consequences was carried out in 1997 (Heiskanen and Piispa 1998). Also in Finland, the survey was repeated, and the second report on Violence against Women in Finland was published in 2006 presenting information about the prevalence, patterns, and trends of violence committed by men against women in 2005. Also, figures on fear of violence and how victims have sought and received help from different agencies were presented. The results showed that in 1997, 40 % of women aged 18–74 had at least once been exposed to men’s violence or threats since the age of 15. In comparison, the percentage was 44 in 2005. In general, there were only small changes between 1997 and 2005 in women’s violence experiences, the most significant being the decrease of intimate partner violence and the increase of work-related violence (Piispa et al. 2006).

The National Violence Against Women (NVAW) Survey was carried out in 1995 and 1996 in the USA, and it sampled both women's and men's experiences with violent victimization. Respondents to the survey were asked about physical assault, forcible rape, and stalking they experienced at any time in their life by any type of perpetrator. Like the other surveys already mentioned, the US Survey showed that violence against women was widespread: nearly 60 % of surveyed women said they were physically assaulted either as a child or as an adult. The survey also showed that many American women were raped at an early age. Out of all women surveyed who said they had been the victim of a completed or attempted rape at some time in their life (18 %), 54 % were younger than 18 years when they experienced their first attempted or completed rape. The survey showed that violence against women in the USA was primarily intimate partner violence. Twenty-two percent of all surveyed women and 64 % of those who reported being raped, physically assaulted, and/or stalked since age 18 were victimized by a current or former husband, cohabiting partner, boyfriend, or date. Only 7 % of all men and 16 % of victimized men had similar experience of intimate partner violence (Tjaden and Thoennes 2000, pp. iii–iv).

NVAW was a onetime survey which has not been compared with any information collected after the survey. However, the Bureau of Justice Statistics' National Crime Victimization Survey (NCVS) has included a special module to their general victimization survey from 1993 to 2008. The findings cover nonfatal and fatal violent crimes committed against females including nonfatal intimate partner violence (IPV), fatal IPV, rape and sexual assault, and stalking. The victimization survey includes women and girls age 12 or older (Catalano 2009).

One of the first violence against women surveys was carried out in Ghana in 1997 where 3,047 women and men aged between 15 and 72 years were interviewed throughout the country. The high number of languages made the survey particularly challenging in Ghana since the country has more than 27 local languages. All the terms used in the study were dually translated into eight local languages representing the main ethnic groups. Another challenge was the access to the people especially women, because in many cases women need permission from husbands/household heads for interviews. Households also needed to confer with chiefs and elders (Ardayfio-Schandorf 2005).

The violent acts identified in the study included wife beating, rape, defilement, widowhood rites, forced marriages, and female circumcision. Around 23 % of the female respondents reported that they have been beaten by their husbands or boyfriends. Six percent of the female respondents stated that they had been defiled; 78 % of the perpetrators were either close relations, acquaintances, or family friends. Eight percent of women said they had been raped; 59 % of the respondents said they never reported these cases to anybody. On forced marriage, 22 % of the married females stated that their parents decided for them, and 30 % of the females reported that their parents and other closer relatives chose their partners for them. In addition, 12 % of the female respondents indicated that they have been circumcised. Respondents were also asked about widowhood rites. Thirty-one percent of widowed respondents said they were being asked to marry their dead husband's brother. Other widowhood rites which the respondents mentioned include shaving of hair, ritual bath, being confined to a room for days, and wearing of rope around the neck (Ardayfio-Schandorf 2005).

The latest targeted stand-alone surveys include the Nationwide Survey on Domestic Violence Against Women in Armenia, 2008–2009 and the National Research on Domestic Violence Against Women in Georgia in 2009, as part of the UNFPA project on Combating Gender Based Violence in the South Caucasus which was launched in April of 2008 (UNFPA 2008). In the Netherlands large scale surveys of violence against women have been conducted in

1988, 1997, and 2010 (Van der Veen and Bogaerts 2010).

General Victim Surveys with Special Module

Canada was the first country to carry out a series of large scale-targeted survey to measure violent victimization of women aged 18 and above. The survey found that 51 % of women had experienced violence at some point of life out of which 39 % was sexual violence, 34 % physical violence, and 29 % intimate partner violence (Statistics Canada 1993). The survey established a baseline to measure violence against women in Canada, and it was used to include a module on spousal violence in the Statistics Canada's General Social Survey (GSS) on victimization which is carried out every 5 years which collects information of Canadians aged 15 years and older; a spousal violence module has been used in 1999, 2004, and 2009 surveys. Sexual assault on the other hand is covered as one of the eight crime types addressed routinely in the general survey (Johnson 2006).

In Denmark, the survey information on violence against women is collected through health-related surveys. The 2000 and 2005 national health interview surveys included a self-administered questionnaire on violence against women including questions on specific forms of violence (Helweg-Larsen and Frederiksen 2008).

For France, there are two different types of surveys available. First, there is the general victimization survey in the Paris region conducted every 2 years between 2001 and 2009. It is not a national population-based survey, but it does cover a large population (representative sample of 10,500 households). This survey includes questions about sexual violence, sexual assault by an intimate, insults, and threats. Estimates of two-year victimization rates are possible. Since 2006, a more focused survey "Cadre de vie et securite" (CVS) is conducted annually by the Observatoire national de la delinquance (OND). This national survey has a representative sample of about 17,000 households. This survey uses a confidential self-completion module on personal violence asking respondents between ages 18 and 75 about annual victimization by physical or sexual violence, either by members of their household or by non-family members (INHESJ 2010).

Since 2004/2005, the yearly British Crime Survey (BSC) has included a self-completion module asking respondents aged between 16 and 59 about their experiences of domestic abuse, sexual assault, and stalking (Smith et al. 2011, p. 69). With this method the sensitive nature of the victimization related to violence against women is taken into consideration, and reporting of these crimes is evidenced to be higher (Flatley et al. 2010, pp. 56, 58).

Comparing Victimization

Multi-Country Violence Against Women Surveys

As shown above, because of several differences in national surveys, they cannot be used to compare the level of victimization across different countries. For that reason, international surveys were developed where victimization data were collected in several countries using the same methodology in order to allow comparison between different countries. However, these surveys also have challenges particularly with the translation of questionnaires, interpretation of concepts, structural differences of societies, and other cultural issues.

The International Violence Against Women Survey (IVAWS) was specifically designed to target violence by males against women, assessing the level of victimization of women in a number of countries worldwide. The IVAWS relies largely on the methodology of the International Crime Victim Survey (ICVS). Between 2003 and 2005, standardized surveys were carried out in Australia, Costa Rica, Czech Republic, Denmark, Greece, Hong Kong, Italy, Mozambique, the Philippines, Poland, and Switzerland. The main findings of the IVAWS are:

- Between 35 % and 60 % of women in the surveyed countries have experienced violence by a man during their lifetime.
- Between 22 % and 40 % have experienced intimate partner violence during their lifetime.

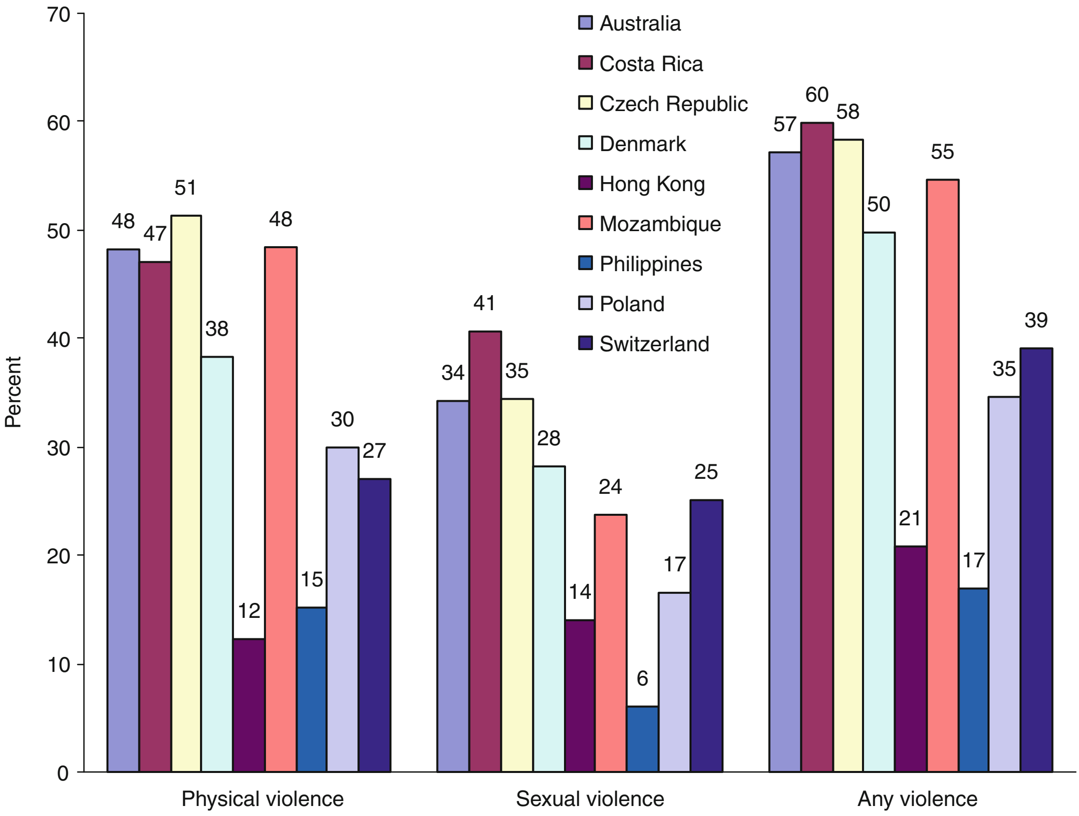

Surveys on Violence Against Women, Fig. 1 Percentage of lifetime physical, sexual, or any violence among adult women (Source: Johnson et al. 2008, 39)

- Less than one third of women reported their experience of violence to the police; women are more likely to report stranger violence than intimate partner violence.
- About one fourth of victimized women did not talk to anyone about their experiences (Johnson et al. 2008) (Fig. 1).

The WHO Multi-country Study on Women's Health and Domestic Violence against Women collected data from over 24,000 women between 2000 and 2003 in the following ten countries: Bangladesh, Brazil, Ethiopia, Japan, Peru, Namibia, Samoa, Serbia and Montenegro, Thailand, and the United Republic of Tanzania. The study estimated the prevalence of physical, sexual, and emotional violence against women, with particular emphasis on violence by intimate partners. The study consisted of standardized population-based household surveys which were carried out in different locations. In five countries (Bangladesh, Brazil, Peru, Thailand, and the United Republic of Tanzania), surveys were conducted in the capital or a large city and one province or region, usually with urban and rural populations. One rural setting was used in Ethiopia, and a single large city was used in Japan, Namibia, and Serbia and Montenegro. In Samoa, the whole country was sampled. The results showed that among women aged 15–49 years:

- Between 15 % (Japan) and 70 % (Ethiopia and Peru) of women reported physical and/or sexual violence by an intimate partner.
- Between 0.3 % and 11.5 % of women reported experiencing sexual violence by a non-partner.
- The first sexual experience for many women was reported as forced – 24 % in rural Peru,

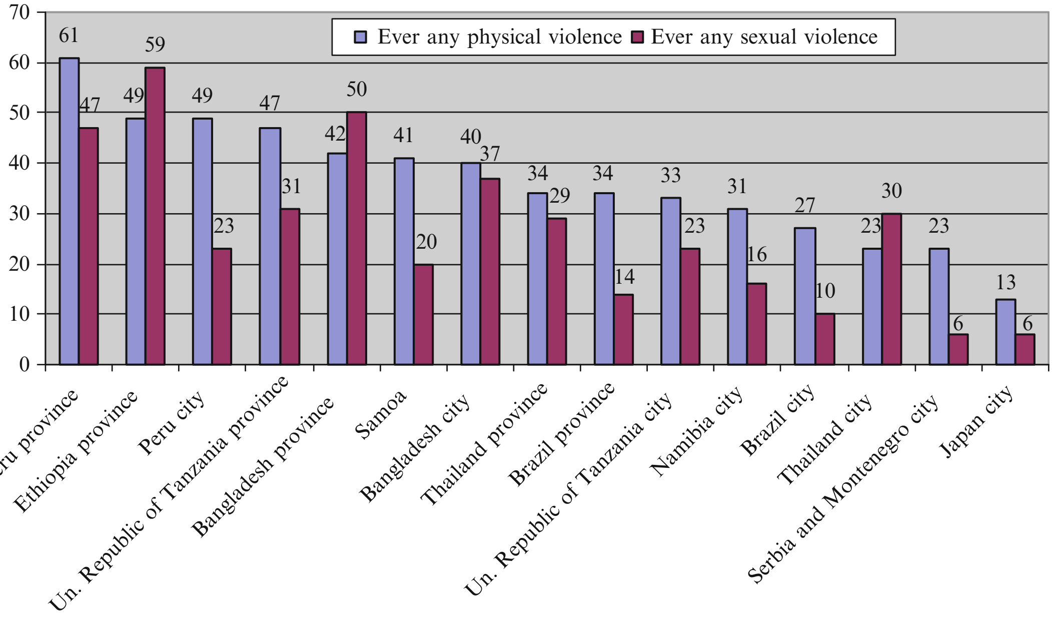

Surveys on Violence Against Women, Fig. 2 Percentage of lifetime physical and sexual violence by an intimate partner among ever-partnered women (Source: WHO 2005, 29)

28 % in Tanzania, 30 % in rural Bangladesh, and 40 % in South Africa (WHO 2005) (Fig. 2).

Demographic and Health Surveys (DHS) are nationally representative household surveys that provide data for a wide range of topics in the areas of population, health, and nutrition. They also include a module on domestic violence assessing the prevalence of intimate partner violence (IPV) among married or cohabiting women. A comparative analysis was carried out based on data collected between 2001 and 2006 in ten developing countries including Bangladesh, Bolivia, Cambodia, Colombia, the Dominican Republic, Egypt, Haiti, India, Kenya, Malawi, Moldova, Nicaragua, Peru, Rwanda, Zambia, and Zimbabwe. The survey analysis showed that there was wide variation across countries in the prevalence of physical or sexual violence experienced by women and perpetrated by their partners: from 75 % in Bangladesh to 16 % in the Dominican Republic. The highest reported rates of physical violence were in Bangladesh (71 %), Bolivia (52 %), and Zambia (45 %), and the lowest reported rates were in Haiti (12 %) and the Dominican Republic (15 %). The highest rates of sexual violence were reported in Bangladesh (26 %), Kenya (15 %), and Bolivia (14 %), whereas the lowest rates were reported in Moldova (3 %), the Dominican Republic (5 %), and Zambia (6 %) (Hindin et al. 2008) (Fig. 3).

Ongoing Demographic and Health Surveys (DHS) can be found in many parts of the world; the latest ones were finalized in 2011 in Angola, Ethiopia, Liberia, Madagascar, Nepal, and Rwanda (Measure DHS 2012).

A standardized survey is also being carried out in all 27 European Union countries including the candidate country Croatia on women's well-being and safety in Europe. The survey will involve standardized face-to-face interviews with about 1,500 women in each country. The interviews will cover women's experiences of violence including physical, sexual, and psychological violence, harassment, and stalking by current and former partners and non-partners. The survey will also look at violence experiences in childhood in order to create a comprehensive

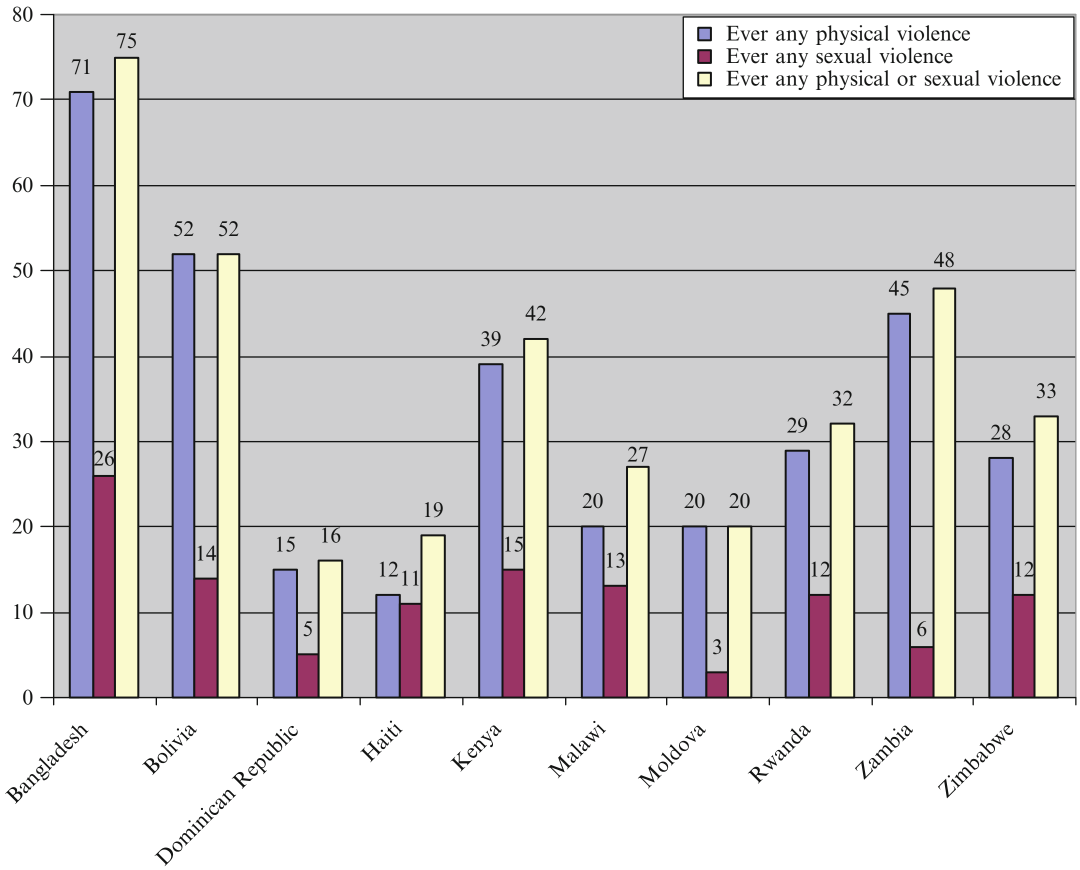

Surveys on Violence Against Women, Fig. 3 Percentage of lifetime physical and sexual violence by intimate partners among currently partnered women age 20–44 (Source: Hindin et al. 2008, 18)

picture of women's experiences of violence during their lifetime (FRA 2012).

Trend Analysis

When trends in violence against women are measured, the comparison is not done between countries but within countries based on information from different time periods. For many reasons, the level of violence against women differs significantly between countries, but that is not the focus of a trend analysis. Rather, the interest is in seeing if the data suggest that the level of violence is increasing, decreasing, or remaining about the same in countries across the globe. An important methodological advantage of comparing trends across countries (rather than levels at one point in time) is that this tends to minimize the comparability measurement problems which hamper international comparisons of levels of violence (Marshall and Summers 2012).

A proper trend analysis is possible only if there is sufficient survey information on violence against women covering a prolonged time period. In a recent article Kangaspunta and Marshall (2012) found only seven countries with (repeated) victim surveys related to violence against women. They were Australia, Canada, Denmark, Finland, France, UK and the USA. There are also other countries which have conducted repeated violence against women surveys; however, they do not necessarily allow for a trend analysis. The case of Belgium presented below clearly shows some of the challenges of trend analysis.

Surveys on Violence Against Women, Table 1 Trend of different forms violence against women in Australia, Canada, Denmark, Finland, France, UK, and USA

	Australia	Canada	Denmark	Finland	France	UK	USA
Survey data source, violence against women	Targeted surveys 1996 and 2005	Combination of targeted survey, 1993 and general social surveys including a module on spousal violence 1999, 2004, and 2009	National health interview survey including a self-administered questionnaire on violence against women 2000 and 2005	Targeted surveys 1998 and 2005	Local victimization survey Paris region every 2 years 2001–2009 and general victimization survey including a self-completion module on personal violence since 2007	General victimization survey including a self-completion module on violence against women yearly since 2004/2005	Targeted survey 1996 and general victimization surveys 1993, 1996, 1999, 2002, 2005, and 2008
Survey-based trends in VAW							
General violence against women	–	NA	0	+	0 or +	NA	–
Partner violence	–	–	–	–	0	–	–
Sexual violence	–	–0	++	+	+	0	–
Trends in sexual assault and domestic violence – based on agency records							
Sexual assault	++	–	–	++	+ –	– +	
Domestic violence/ IPV			+	+	+	+	
Trends in homicide – based on agency records							
Female victim homicide	–	_	–	0	–	–	–
Female partner homicide	+	_	NA	+	–	+	–

Source: Kangaspunta and Marshall 2012

In Belgium, three large-scale studies were conducted on the prevalence of gender-based violence in 1988, 1998, and 2009. The first study in 1988 analyzed incidences of violence against women; the second study in 1998 was extended to men. The 2009 survey studied the occurrence, forms, and severity of the physical, sexual, and emotional abuse to which women and men are exposed. These studies illustrate some of the difficulties performing trend analyses. While the 1998 study was exclusively concerned with violent experiences over the course of the respondents' whole life, the 2009 survey included also the experiences during the past 12 months. In the 1998 survey, a violent act was defined in a detailed manner listing 17 acts of physical and 24 acts of sexual violence; in the 2009 study violence was measured in a more synthetic manner, by grouping physical acts into three questions and one general question on sexual abuse.

In the Belgian surveys of 1988 and 1998, the age group included in the sample ranged from 20 to 50 years; the 2008 survey studied individuals between the ages of 18 and 75. In the case of the 2009 study, only abuse experienced in adult life (after the age of 18) was taken into account, whereas the figures for 1998 concern abuse experienced over the respondent's lifetime. Also, the 1988 survey was conducted using face-to-face interviews, while the more recent surveys were carried out either over the phone or online. In the report presenting the findings of the 2009 results, the authors conclude that – given the differences in methodologies – the surveys do not adequately support trend analysis (Pieters et al. 2010).

In the USA, the National Violence Against Women (NVAW) Survey was a onetime effort which has not been compared with any information collected after the survey. The US violence against women trend analysis has been conducted by using data from the Bureau of Justice Statistics' National Crime Victimization Survey (NCVS), 1993–2008 covering nonfatal and fatal violent crimes committed against females including nonfatal intimate partner violence (IPV), fatal IPV, rape and sexual assault, and stalking. The victimization survey included women and girls age 12 or older (Kangaspunta and Marshall 2012).

The results of the trend analysis of violence against women in the most developed western countries suggest a drop in partner violence against women as measured by the standard surveys (Table 1). On the other hand, (non-partner) sexual violence appears to be increasing. This drop in partner violence is accompanied by an apparent increase in the level of female victimization reported to the police and other agencies. As Table 1 shows, homicides with a female victim were shown to be decreasing (possibly reflecting a general decline in homicides in general) (Kangaspunta and Marshall 2012).

Conclusions

The need to have sound and up-to-date information on violence against women is well recognized both at the national and international level. The United Nations as well as many other international and regional organizations has stressed the need to collect reliable and comparable data on gender-based violence. Many countries have responded to this call since 1993 when the first highly publicized violence against women survey was carried out in Canada. The first surveys were carried out in highly developed western countries only; however, soon after also countries in other regions followed with their own surveys. The latest efforts to conduct such surveys are covering more and more countries in all geographical areas. Also several new initiatives to collect comparative data have recently been launched.

Often, violence against women surveys have focused on intimate partner violence which raises particular concerns in many countries. Also other forms of violence have been studied including sexual violence and violence in public places. In general, all violence against women surveys have greatly increased our knowledge on the victimization of women, showing that high proportions of women in all parts of the world suffer from violence and abuse.

While some women face the dangers at home, most women need to be aware of the risks outside the private space as well. Women's everyday life contains numerous situations where harassment, abuse, and violence could take place such as during doctors' appointments, at the workplace, at school and other educational institutions, during leisure time – for example when participating in sports or just when having a walk in a park. Some of the new studies, particularly the survey on women's well-being and safety in Europe, aim at highlighting these latter forms of violence. The final objective of violence against women surveys should be to generate positive changes so that women all over the world could enjoy everyday life without fear of violence of any sort.

Related Entries

- ▶ British Crime Survey
- ▶ Domestic Violence
- ▶ Feminist Theory in the Context of Sexual Violence
- ▶ Gendered Theory and Gendered Practice
- ▶ International Crime Victimization Survey
- ▶ National Victimization Surveys
- ▶ Police Family Violence Services
- ▶ Responses to Domestic Violence in India
- ▶ Understanding Cross-National Variation
- ▶ Understanding Victimization Frequency
- ▶ US National Crime Victimization Survey
- ▶ Victimization, Gender, and the Criminal Justice System

Recommended Reading and References

ABS (2005) Personal safety survey Australia. Australian Bureau of Statistics, Canberra

Ardayfio-Schandorf, Elizabeth (2005) Violence against women: the Ghanaian case. Expert presentation at the expert group meeting organized by: UN division for the advancement of women in collaboration with: Economic Commission for Europe (ECE) and World Health Organization (WHO). "Violence against women: a statistical overview, challenges and gaps in data collection and methodology and approaches for overcoming them", Geneva, Switzerland, 11–14 April, 2005. http://sgdatabase.unwomen.org/uploads/Violence against women in Ghana – paper by Elizabeth Ardayfio-Schandorf.pdf

Catalano S, Smith E, Snyder H, Rand M (2006) Intimate partner violence in the United States. Bureau of Justice Statistics

Flatley J, Kershaw C, Smith K, Chaplin R, Moon D (eds) (2010) Crime in England and Wales 2009/10. Home Office Statistical Bulletin 12/10. London, Home Office. http://www.homeoffice.gov.uk/rds/pdfs10/hosb1210.pdf

FRA (2012) FRA survey on women's well-being and safety in Europe. European Union Agency for Fundamental Rights. http://fra.europa.eu/fraWebsite/research/projects/proj_eu_survey_vaw_en.htm

Heiskanen M, Piispa M (1998) Usko, toivo, hakkaus. Kyselytutkimus miesten naisille tekemästä väkivallasta. Tilastokeskus. Tasa-arvoasiain neuvottelukunta. Oikeus 1998:12. Sukupuolten tasa-arvo. SVT. Helsinki

Heiskanen M (2006) Miesten Naisille Tekemän Väkivallan Kehitys ja Kokonaiskuva. In Piispa, M., Heiskanen, M., Kääriäinen, J. and Sirén, R. (2006) Naisiin Kohdistunut Väkivalta 2005. Oikeuspoliittisen Tutkimuslaitoksen Julkaisuja 225; Yhdistyneiden Kansakuntien yhteydessä toimiva Euroopan Kriminaalipolitiikan Instituutti (HEUNI) Publication Series No. 51: Helsinki

Helweg-Larsen K, Frederiksen ML (2008) Men's violence against women. Extent, characteristics and the measures against violence – 2007. English summary. Minister for Gender Equality, National Institute of Public Health

Hindin MJ, Sunita K, Ansara DL (2008) Intimate partner violence among couplesin 10 DHS countries: predictors and health outcomes. Macro International, Calverton, DHS Analytical Studies No. 18

INHESJ (2010) La criminalité en France. Institute National des Hautes Etudes sur la Sécurité et de la Justice and Observatoire National de la Délinquance et des Réponses Pénales (ONDRP)

Johnson H (2006) Measuring violence against women. Statistical trends 2006. Statistics Canada 2006. Catalogue no. 85–570-XIE

Johnson H, Ollus N, Nevala S (2008) Violence against women: an international perspective. Springer, HEUNI, New York

Kangaspunta K, Marshall,HI (2012) Trends in violence against women: some good news and some bad news. In: van Dijk J, Tseloni A, Farrell G (eds) The international crime drop: new directions in research. Palgrave, Basingstoke

Marshall HI, Summers D (2012) Contemporary differences in rates and trends of homicide among European nations. In: Liem MCA, Pridemore WA (eds) Handbook of European homicide research: patterns, explanations and country studies 2012, pp 39–69

Measure DHS (2012) Demographic and health surveys. http://www.measuredhs.com/What-We-Do/survey-search.cfm?pgtype=main&SrvyTp=year

Mc Lennan W (1996) Women's safety Australia. Australian Bureau of Statistics, Canberra
Mucchielli L (2010) Are we living in a more violent society? A socio-historical analysis of interpersonal violence in France, 1970s-present. Br J Criminol 50:808–829
Piispa M, Heiskanen M, Kääriäinen J, Sirén R (2006) Naisiin kohdistunut väkivalta 2005. Oikeuspoliittisen Tutkimuslaitoksen Julkaisuja 225; Yhdistyneiden Kansakuntien yhteydessä toimiva Euroopan Kriminaalipolitiikan Instituutti (HEUNI), Helsinki, Publication Series No. 51
Phillips J, Park M (2006) Measuring domestic violence and sexual assault against women: a review of the literature and statistics. E-Brief: http://www.aph.gov.au/library/intguide/sp/ViolenceAgainstWomen.htm
Pieters J, Italiano P, Offermans A-M, Hellemans S (2010) Emotional, physical and sexual abuse – the experiences of women and men. Institute for the equality of women and men, Brussels
Simister J, Mehta PS (2010) Gender-based violence in India: long-term trends. J Interpers Violence 25(9):1594–1611, http://jiv.sagepub.com/content/25/9/1594
Statistics Canada (1993) Violence against women – survey highlights and questionnaire package. Statistics Canada, Canada
Tjaden P, Thoennes N (2000) Full report of the prevalence, incidence, and consequences of violence against women, findings from the national violence against women survey, national institute of justice, office of justice programs, U.S. Department of Justice, and the Centers for Disease Control and Prevention, Washington, DC
UNFPA (2008) Combating gender based violence in the South Caucasus. http://www.genderbasedviolence.net/content/show/4/about-us.html
United Nations (1993) Declaration on the elimination of violence against women. U.N. Document A/48/49, New York
Van der Veen HCJ, Boogaerts S (2010) Huiselijk Geweld. WODC, Den Haag, Nederland
Van Dijk JJM, van Kesteren JN, Smit P (2007) Criminal victimization in international perspective, key findings from the 2004–2005 ICVS and EU ICS. Boom Legal Publishers, The Hague
Van der Veen HCJ, Bogaerts S (2010) Huiselijk geweld in Nederland: overkoepelend synthese-rapport van het vangst-hervangst-, slachtoffer- en daderonderzoek 2007–2010. Den Haag, WODC
Walby S, Myhill A (2001) New survey methodologies in researching violence against women. Br J Criminol 41(3):502–522
Wasileski G, Miller S (2010) Review essay: the elephants in the room: ethnicity and violence against women in post-communist Slovakia. Violence Against Women 16:99–125, http://vaw.sagepub.com/content/16/1/99
WHO (2005) Multi-country study on women's health and domestic violence against women: summary report of initial results on prevalence, health outcomes and women's responses. World Health Organization, Geneva

Suspended Sentence

► Probation and Community Sanctions

Synthesizing Biological and Social Theorizing

Jon Heidt[1] and Gail S. Anderson[2]
[1]School of Criminology and Criminal Justice, University of the Fraser Valley, Abbotsford, BC, Canada
[2]School of Criminology, Simon Fraser University, Burnaby, BC, Canada

Overview

The synthesis of biological and social environmental theories is a relatively recent phenomenon in criminology. While the early criminologists focused heavily on biological factors related to criminality, criminology eventually came to be dominated by sociological explanations of criminal behavior in the 1930s, and this continued for several decades. However, starting in the 1970s, criminologists began to reexamine biological findings relevant to understanding criminal behavior.

In modern criminology, there are several prominent attempts to synthesize biological and social environmental theorizing. In particular, these integrated theories in the biosocial and developmental research programs use a combination of biological, psychological, social, and environmental factors to explain criminality. These new theories have been successful fusing existing explanations of criminal behavior and explaining interactions between key variables from biology, psychology, and sociology.

The Reemergence of Biology in Criminology

For many years, biological explanations of criminal behavior were excluded from mainstream criminology. This was largely a product of the domination of sociological and environmental explanations in the study of crime from the 1930s to the 1970s (Ellis and Walsh 2003). However, during the late 1970s, some criminologists began to acknowledge the importance of biology and individual differences in the formation of criminality. The emergence of the biosocial perspective offered a way to synthesize existing social environmental explanations with important biological findings relevant to understanding criminal behavior (Ellis 1990).

Recent attempts to fuse biological and social environmental theories can be found in developmental-life course theories of criminal behavior (e.g., see Farrington 1992; Moffitt 1993) and biosocial criminology (e.g., see Ellis 2005; Robinson and Beaver 2009). The notion of gene-environment interaction is important to these theories. Genes determine the different traits that are passed from parents to offspring. In some cases, these traits lead to the formation of an impulsive or aggressive personality creating a predisposition towards antisocial and/or criminal behavior (Walsh 2009). If the environmental conditions are right, an individual with an impulsive or aggressive personality may become involved in crime.

Many of these recent contributions also rely upon a risk factor approach to theorizing (Bernard and Snipes 1996). The goal of the risk factor approach is to identify important variables from different levels of explanation and explain how they interact with one another (Robinson and Beaver 2009). Before reviewing the key contemporary biosocial theories of criminal behavior, it is important to be familiar with the origin of the biosocial perspective in criminology.

Origin of the Biosocial Perspective in Criminology

In the late nineteenth century, biological and multiple factor explanations of criminality were quite popular and could be found in the work of many well-known positivist criminologists including Lombroso, Ferri, Garofalo, and Hooton. However, starting in the early twentieth century, these explanations slowly started to lose favor and began to be displaced by theories that emphasized sociological and environmental explanations. Most prominent were theories offered by the Chicago School (e.g., social disorganization and differential association theory) and Merton’s strain theory.

Modern fusions of biological and social environmental explanations can be traced back to Eysenck’s (1964) theory of the criminal personality. He argued that criminality resulted from the interplay between neurophysiological and environmental conditions and that behavior was learned through classical (or Pavlovian) conditioning. More specifically, individual differences in neurochemistry may cause some people to condition less effectively which leads to the formation of various personality types some of which predispose people to criminality. He identified two personality traits, extraversion and neuroticism, as particularly important to proper conditioning. To summarize, Eysenck (1964) suggested that people with high levels of extraversion and neuroticism would have difficulties learning from punishment and other aversive stimuli, making them the most likely to engage in crime.

Eysenck (1964) claimed that extraverted individuals are especially prone to criminality because their behavior consists of attempts to stimulate their underactive arousal systems. The inactivity can be traced to a small bundle of neurons in the central part of the brain known as the ascending reticular activating system (Eysenck 1964). This system is responsible for keeping the cerebral cortex stimulated and alert. Extraverts are thought to have a pronounced insensitivity in this area of the brain. Consequently, they tend to seek out stimulation in their environment by engaging in risky, thrilling, and impulsive behaviors which often equate to various forms of criminal and/or antisocial behavior.

Neuroticism refers to how one reacts to stress. People with high levels of neuroticism are very sensitive to stress and have difficulties recovering from stressful situations. People with low levels of neuroticism tend to be calm and emotionally stable even when stressors are present and recover quickly after experiencing stressful events. According to Eysenck (1964), high levels of neuroticism are connected to high baseline levels of activity in the autonomic nervous and limbic systems in the brain. These systems are responsible for triggering one's fight or flight response in emergencies and other dangerous situations. Neurotics are thought to be more prone to criminality because the emotional instability created by the high activity in these areas can cause overreactions or inappropriate reactions to stressful situations which may result in criminal and/or antisocial behavior.

C. R. Jeffery (1978) was another early proponent of the biosocial perspective in criminology. He noted that for nearly a half century, criminologists had ignored the findings of the biological sciences that were an important part of a full explanation of criminal behavior. While he was clearly a pioneer in the field of biosocial criminology, Jeffery failed to produce a specific theory of criminality that synthesized biological and social environmental factors. However, he did suggest the application of systems theory to the field of criminology (Jeffery 1990). The systems approach would become a key component of Robinson and Beaver's (2009) integrated systems theory of antisocial behavior which will be discussed later in this entry.

Another early example of an attempt to synthesize biological and social environmental explanations can be found in Wilson and Herrnstein's (1985) *Crime and Human Nature*. Wilson and Herrnstein argued that certain biological predispositions (e.g., aggression, impulsivity, and low IQ) affect how people respond to punishment and, consequently, their propensity to commit crime. In this theory, biological predispositions are thought to affect the operant (instrumental) learning process; this is in contrast to Eysenck's (1964) theory which stressed classical (or Pavlovian) conditioning. In other words, one's susceptibility to deterrence is affected by individual differences in biology. According to Wilson and Herrnstein (1985), behavioral problems begin with parents who fail to discipline their children properly and continue with the criminal justice system which, in their view, had become overly lenient on criminals during the 1960s and 1970s.

It is clear that the work of both Eysenck (1964) and Wilson and Herrnstein (1985) are attempts to combine biological factors (i.e., predispositions) and social environmental factors (i.e., learning and conditioning) to explain criminality. One major shortcoming of these theories was a failure to seriously consider social factors involved in criminality outside of the family and psychological learning processes. In addition, these theories focused exclusively on explaining more serious crimes involving violence like murder and robbery while ignoring corporate and white-collar crime and other crimes common in the middle and upper classes (e.g., driving under the influence, tax evasion, and minor theft). It is interesting to note that early twin studies conducted by Mednick and his colleagues (1977) found stronger connections between heredity and petty crimes than between heredity and violent crimes. Finally, both theories seem to be geared towards explaining chronic and persistent forms of offending rather than sporadic or infrequent criminal behavior.

Contemporary theorists have attempted to address these issues by broadening the scope of their theories and by specifying interactions between key variables in their theories. These more recent theories will be reviewed in the following section.

Contemporary Syntheses

Farrington's (1992) integrated cognitive antisocial potential theory is one example of a recent attempt to fuse biological and social environmental explanations of criminal behavior. His theory also incorporates a developmental-life course approach meaning that it attempts to account for within-individual change as well as between-individual differences in offending. The key construct in this theory is the notion of antisocial potential (AP) which assesses one's potential to commit antisocial acts. In addition, AP serves as a vehicle for integrating existing biological and sociological explanations of criminality.

Two types of AP are identified: long term and short term. Long-term AP refers to between-individual differences in individuals that are relatively stable over time. Factors that contribute to long-term AP include low cortical arousal, impulsivity, high levels of strain, weak attachments, and exposure to antisocial models (Farrington 1992). This is a clear attempt to combine variables derived from key biological findings with prominent sociological theories of criminality (e.g., strain, social control, and social learning). Short-term AP focuses upon within-individual variation in criminality. Factors for this dimension of AP are situational (e.g., anger, boredom, and intoxication) and are drawn out of the rational choice, routine activities, and opportunity theories (Farrington 1992).

Like Wilson and Herrnstein (1985), Farrington's (1992) theory addresses street crime committed by lower-class males, so the scope of the theory is quite limited. In other words, like its predecessors, this theory fails to account for middle-class and elite forms of crime and "normative" crime committed by adolescents. The main advancement here is the use of a developmental-life course approach and the focus on within-individual change in criminality. This emphasis allows the theory to account for changes in offending patterns over time.

Another theory that incorporates a developmental-life course approach is Moffitt's (1993) developmental taxonomy. She posits the existence of two types of offenders: life-course-persistent (LCP) and adolescence-limited (AL) offenders. According to Moffitt (1993), LCPs offend frequently throughout their life and account for approximately 5–10 % of all offenders. LCPs are thought to suffer from neuropsychological deficits that affect verbal (i.e., speech, reading, and writing) and executive (i.e., impulse control and attention span) functioning. These deficits also cause problematic social interactions, especially with parents, school officials, and other authority figures. Family interactions are particularly problematic since, in many cases, parents tend to have similar neuropsychological deficits making interaction all the more difficult. In addition, these deficits predispose the individual to engage in risk-taking and impulsive behavior which often results in crime and other forms of antisocial behavior. Finally, these neuropsychological problems impede the learning of prosocial alternatives to crime, and eventually LCPs become ensnared by the consequences of crime (e.g., negative labeling and a reduction in legitimate opportunities that often lead to desistance from criminal activity). This portion of the theory is a clear attempt to explain how individual differences in biology (i.e., neuropsychological deficits) interact with various environmental factors (i.e., family and school interactions and negative labeling) to produce chronic offending patterns (Moffitt 1993).

Individual differences are thought to be much less important in explaining AL offending patterns; the key factors in AL offending are primarily social and environmental. AL offenders compose the vast majority of offenders (about 90–95 %) and commit crime during adolescent years but eventually desist by their early 20s. Moffitt (1993) argues that this type of crime is normative and is not the result of special traits or biological factors. Instead, AL offending occurs through a process of social mimicry of LCP behavior.

According to Moffitt (1993), modern society has created a "maturity gap" for teenagers in which they lack true roles. This also means that they are prohibited from engaging in adult activities like working, driving, voting, and living on their own until their late teens and early 20s. This role vacuum causes teens to search for other ways to feel mature and independent. For some of these adolescents, maturity and independence are defined as engaging in adult behaviors including drinking alcohol and having sex. Eventually, some of these individuals observe the LCPs getting rewards for their delinquent behavior (e.g., premarital sex, underage drinking, drug use, and possessions obtained through theft). For the ALs, these rewards symbolize freedom and autonomy. In order to obtain feelings of independence, the ALs begin to mimic the delinquent behavior of the LCPs.

AL offending has very little to do with individual differences per se; however, biological factors are still taken into account. Moffitt (1993) claims that societal changes have effectively lengthened adolescence by limiting or delaying roles that require responsibility and maturity. The factor of maturity and its role in development is the biological component in this part of the theory, while the notion of social mimicry and societal changes affecting developmental patterns of youth represents the social environmental component.

Ellis's (2005) evolutionary neuroandrogenic theory attempts to explain why sex is the strongest correlate of crime. To do this, Ellis (2005) suggests that high levels of testosterone give rise to competitive/victimizing behavior. Competitive/victimizing behavior is conceptualized along a continuum. On one end of the continuum are "crude" or criminal forms of behavior, which refer to attempts to injure people or deprive them of their property; in most cases, these acts have been criminalized by modern governments. Notably, white-collar or elite forms of crime are also addressed here as well. On the other end of the continuum are "sophisticated" or commercial forms of behavior, which include acts that "make no profits on the sale of goods or services, although those who administer and maintain the organization under which they operate usually receive much higher wages than do those who provide most of the day-to-day labour" (Ellis 2005: 288). According to this theory, the most serious criminality will be concentrated among adolescent and young adult males of low social status. Victimless forms of crime are left unaddressed by this theory.

This theory consists of two interrelated propositions. The evolutionary proposition suggests that aggressive and acquisitive criminal behavior evolved as an aspect of human reproduction. In short, females are thought to have mating preferences for higher status males that are capable of providing resources. These preferences have been observed in other mammalian species and are thought to hold true in humans as well. Consequently, the theory predicts that rape will always be more common among males who are unable to be stable providers.

The neuroandrogenic proposition identifies three important aspects of brain functioning that play a role in criminal behavior. First, high perinatal testosterone levels can alter brain functioning in ways that promote competitive/victimizing behavior. This is because higher levels of testosterone can affect brain development and may lead to suboptimal arousal levels, may make one prone to seizures, and may cause a rightward shift in neocortical functioning. These are all correlates that appear in other research on serious criminality (see, e.g., Eysenck 1964). Second, one's ability to learn or IQ can serve to inhibit criminal behavior. Third, executive functioning or planning ability is thought to inhibit "crude" or criminal forms of behavior. After adolescence, those with higher levels of learning and planning ability will move quickly from "crude" to "sophisticated" forms of behavior (Ellis 2005).

Because on it focuses upon testosterone and brain development, Ellis's (2005) evolutionary neuroandrogenic theory may appear to be a purely biological theory of criminality. However, the evolutionary proposition considers the role of social environmental factors relating to sex and social status. In contrast to some of the other formulations presented here, mainstream theories from the control, strain, and differential association/cultural deviance trajectories in criminology do not play a role in Ellis's (2005) evolutionary neuroandrogenic theory.

Robinson and Beaver's (2009) integrated systems theory is a wide-ranging and holistic attempt to combine biological and social environmental explanations into a unified theory. The approach used is composed of two key influences. First, the theory itself is derived from the evolutionary-ecological paradigm proposed by Vila (1994). Vila claimed that ecological, micro-, and macrolevel factors form the basis for the formation of personality (referred to as "strategic style") which determines the likelihood of criminal behavior.

Second, Robinson and Beaver's theory (2009) incorporates a systems theory perspective into the evolutionary-ecological paradigm. This is an interdisciplinary biosocial approach, meaning that it seeks to synthesize findings from a variety of disciplines. Systems theory posits six levels of analysis: cell, organ, organism, group, organization, and community/society. Each of these levels represents an individual system, and all of these systems interact with other systems above and below them. This is essentially a risk factor approach similar to the one offered by Bernard and Snipes (1996) and used by the developmental-life course theorists (Farrington 1992; Moffitt 1993).

The biosocial basis of Robinson and Beaver's (2009) formulation is made clear in the first proposition of the theory:

> All behaviours are the result of gene-environment interaction. Genes do not cause behavior; they predispose individuals to react to environmental stimuli in certain ways, meaning some will be more likely to behave in an antisocial manner. Genes are linked to numerous factors relevant for antisocial behavior, including but not limited to personality, drug use and abuse, IQ, violence, and mental illness. (367-368)

The rest of the propositions explain how various criminogenic factors contribute to antisocial and criminal behavior. A variety of factors and explanations from different disciplines are discussed and grouped into the various levels of analysis proposed by systems theory. Biological and psychological factors are considered to be part of the cell, organ, and organism levels. The cellular level of analysis refers to how genes may impact behavior and provide certain predispositions for antisocial behaviors. For example, genes determine IQ, neurotransmitter, and hormone levels all of which are connected to the brain. The main emphasis of the organ level is the brain since it plays a prominent role in behavior. The organism level is focused on the formation of personality which is considered to be a result of gene-environment interactions. This level also accounts for factors relating to diet and nutrition, drug consumption, and mental illness (Robinson and Beaver 2009).

The social and environmental factors are placed into the group, organization/community, and societal levels. The group level deals with small group interactions and incorporates explanations from sociological and social-psychological criminology like the social learning and social control theories. Community variables are derived from sociological theories like social disorganization, routine activities, and lifestyle theory which focus on how certain areas and neighborhoods produce high levels of crime. This level of explanation also considers the impact of organizational level variables proposed by deterrence and labeling theories (i.e., related to the criminal justice system and societal reaction) might have on criminal activity. Finally, societal level explanations are drawn from macrosociological explanations of crime and incorporate variables from strain, anomie, subcultural, and culture conflict theories.

Robinson and Beaver (2009) also attempt to specify how the levels and variables within them interact with each other. Other theorists (Eysenck 1964; Farrington 1992; Moffitt 1993) have identified some of these interactions, especially those at the biological and psychological levels, but none have done so with the level of rigor applied here. For example, Robinson and Beaver (2009) explain that abnormal neurotransmitter, enzyme, and hormone levels impact the formation of the brain and personality, which can later lead to problems in social interaction with parents, friends, peers, and teachers. However, they identify other variables like destructive labeling, economic stress, diet, and pollution that can also affect brain development and may influence subsequent behavior. In addition, larger socioeconomic factors (e.g., tax and corporate policies) may contribute to social disorganization in certain communities. Levels of social disorganization can influence incidences of family conflict, which can also affect biosocial development. Further, social disorganization often gives rise to unsupervised peer groups which play role in the formation of criminal and antisocial behavior (Robinson and Beaver 2009).

Robinson and Beaver's (2009) theory does an effective job of organizing the important findings from a number of disciplines. The attempts made to specify the relationships between the different variables are also useful; however, further research will be needed to truly understand, identify, and refine all of these relationships.

Problems in previous theories (Farrington 1992; Moffitt 1993) are also avoided because the scope has been broadened, allowing this theory to more seriously consider how white-collar and corporate crime may be explained.

Conclusion

The biosocial and developmental approaches have proven to be extremely useful as a starting point for synthesizing biological and social environmental explanations of criminal behavior. However, there are still a number of lingering controversies and contentious debates that should be discussed. First, early biosocial theories (Eysenck 1964; Wilson and Herrnstein 1985) generally downplayed the role of social and environmental factors in criminality. Much of this was related to an implicit emphasis on identifying serious and violent chronic offenders rather than explaining more common types of crime committed by youth or less serious offenders. Again, it is important to bear in mind that early studies of twins found biological connections to petty, and not violent, crimes (Mednick and Christianson 1977). This disjunction between theory and research indicates that there is still an interesting puzzle for criminologists to solve.

Later theories (Farrington 1992; Moffitt 1993; Ellis 2005; Robinson and Beaver 2009) attempted to address this problem by more directly incorporating social factors into the mix of variables. However, biological factors and individual differences still play a central role in these formulations. Further, many of these theories fail to account for white-collar and corporate forms of crime, so their scopes are somewhat narrow. Notable exceptions here include Ellis's (2005) evolutionary neuroandrogenic theory and Robinson and Beaver's (2009) integrated systems theory.

Second, some commentators have suggested that there is currently an overreliance on the risk factor approach in criminology (Moffitt and Caspi 2006; Wikstrom 2008). Specifically, these critics charge that simply assembling lists of key risk factors does not advance the state of criminological theory. The underlying issue here is confusion between correlation and causation; in other words, many risk factors are based on correlates of crime rather than causal factors. In order to break free of this problem, theorists have attempted to specify interactions between different variables and more specific causal processes that exist within theories (Robinson and Beaver 2009). Unfortunately, many of these relationships and causal mechanisms remain somewhat unclear and unspecified.

Despite these problems, biosocial and developmental-life course theories have done an admirable job of advancing criminological theories past the point of purely sociological explanations of criminal behavior. Recent attempts to integrate biological and social environmental explanations have clarified the relationships that variables have with one another and have identified some interactions that may exist between these variables. However, more unified theories will require more research that clearly specifies all of the potential relationships the various biological and social environmental factors may have with each other.

Related Entries

- ▶ Career Criminals and Criminological Theory
- ▶ Cognitive/Information Processing Theories of Aggression and Crime
- ▶ Evolutionary Theories of Criminal Behaviors
- ▶ Eysenck's Personality Theory
- ▶ History of Criminological Theories: Causes of Crime
- ▶ Hormones, the Brain, and Criminality
- ▶ ICAP Theory
- ▶ Neurology and Neurochemistry of Crime
- ▶ Risk factors for Adolescent Sexual Offending

Recommended Reading and References

Beaver KM, Walsh A (2010) Biosocial theories of crime. Ashgate Publishing, Burlington

Bernard TJ, Snipes JB (1996) Theoretical integration in criminology. Crime Justice 20:301–348

Ellis L (1990) Introduction: the nature of the biosocial perspective. In: Ellis L, Hoffman H (eds) Crime in biological, social and moral contexts. Praeger, New York

Ellis L (2005) A theory explaining the biological correlates of criminality. Eur J Criminol 2(3):287–315

Ellis L, Hoffmann H (1990) Crime in biological, social and moral contexts. Praeger, New York

Eysenck H (1964) Crime and personality. Methuen, London

Eysenck H, Gudjonsson GH (1989) The causes and cures of criminality. Plenum, New York

Farrington D (1992) Explaining the beginning, progress, and ending of antisocial behavior from birth to adulthood. In: McCord J (ed) Facts, frameworks, and forecasts: advances in criminology theory, vol 3. Transaction, New Brunswick

Fishbein D (2001) Biobehavioral perspectives in criminology. Wadsworth Publishing, Belmont

Jeffery CR (1978) Biology and crime. Sage, Beverly Hills

Jeffery CR (1990) Criminology: an interdisciplinary approach. Prentice Hall, Englewood Cliffs

Mednick SA, Christiansen, KO (1977) Biosocial bases of criminal behaviour. New York: Gardiner

Moffitt T (1993) Adolescence-limited and life-course-persistent antisocial behaviour: a developmental taxonomy. Psychol Rev 100:674–701

Moffitt T, Caspi A (2006) Evidence from behavioral genetics for environmental contributions to antisocial conduct. In: Wikstrom P-O, Sampson RJ (eds) The explanation of crime: contexts, mechanisms, and development. Cambridge University Press, Cambridge

Rafter N (2008) The criminal brain: understanding biological theories of crime. New York University Press, New York

Robinson M, Beaver K (2009) Why crime? An interdisciplinary approach to explaining criminal behavior. Pearson/Prentice Hall, Upper Saddle River

Vila BJ (1994) A general paradigm for understanding criminal behaviour: extending evolutionary ecological theory. Criminology 32:501–549

Walsh A (2009) Biology and criminology: the biosocial synthesis. Routledge, New York

Walsh A (2011) Social class and crime: a biosocial approach. Routledge, New York

Walsh A, Beaver K (2009) Biosocial criminology: new directions in theory and research. Routledge, New York

Walsh A, Ellis L (2003) Biosocial criminology: challenging environmentalism's supremacy. Nova Science, New York

Wikstrom P-O (2008) In search of causes and explanations of crime. In: King RD, Wincup E (eds) Doing research on crime and justice. Oxford Press, Oxford

Wilson JQ, Herrnstein R (1985) Crime and human nature. Simon and Schuster, New York

Systematic Social Observation

John D. McCluskey[1], Roger B. Parks[2] and Stephen D. Mastrofski[3]
[1]Department of Criminal Justice, Rochester Institute of Technology, Rochester, NY, USA
[2]Indiana University Bloomington, Bloomington, IN, USA
[3]George Mason University, Manassas, VA, USA

Overview

Systematic social observation (SSO) is the direct observation of social phenomena in their natural settings. It is often a group enterprise with many researchers using a systematized protocol to gather quantified data. Its application in criminology has been sparse compared to official data or survey methods, but it offers unique measurement capacity that can prove valuable for many issues in criminology.

Below, SSO is first defined, and then the method is described along with issues of validity and reliability. Next, a review of the application of SSO to criminological questions offers some insight into the variety of settings in which the approach has been used, and finally this entry concludes with speculation regarding the future configuration of SSO and its role in contributing to an understanding of important questions in criminology.

Definition of Systematic Social Observation

Many individuals engage in social observation, for example, standing on a street corner and people watching or observing audience members at a public event. Viewing persons in natural social settings (or even the setting itself, e.g.,, assessing the quality of neighborhood housing in a location one wishes to move) is an act of observation. What distinguishes SSO is the *systematic* application of rules and protocols that structure the act of observation. Those explicit rules determine what constitutes relevant information for the observer and how that information is to be structured (coded) for subsequent analysis. Explicit rules for observation and coding make meaningful comparisons of events observed by multiple observers and at different times and places (Reiss 1971). One might say that SSO is structured watching with a purpose. SSO is a powerful tool for the study of human behaviors. Its power derives from the focus and structure it brings to observation of human behavior in its natural settings.

History

In the late 1920s and early 1930s, scholars studying early childhood social development practiced systematic social observation, using methods "... designed ... to ensure consistent recordings of the same events by different observers ..." (Arrington 1943, p. 83). Systematic social observation came to criminology at the hand of Albert J. Reiss, Jr., who encouraged social scientists to shed some "nonsensical" views about the limits and benefits of different forms of observing social phenomena (Reiss 1968, 1971). Reiss objected to the notion that direct observation of social phenomena in their natural setting was work for solo researchers using qualitative methods, while survey research was suitable as a group enterprise with many researchers using a systematized protocol to gather quantified data. Reiss argued that both direct social observation and survey research were in fact forms of observation that must confront the same set of challenges to producing interpretable information, that both were amenable to either solo or group practice, and that both could be used effectively for discovery or validation of propositions about social phenomena. Beyond these insights, Reiss's important contribution to criminology in this area was the development and practice of the techniques of SSO. Reiss demonstrated how SSO could be used to answer important questions about what influences police-citizen interactions, with implications for theories about police-citizen relationships and for public policies concerning justice, race relations, and crime control.

Suitability of SSO

What makes SSO especially valuable to researchers gathering data directly in the natural setting are precision of the observations and the independence of the observer from that being observed (Reiss 1971, p. 4). For example, some classic qualitative field research pioneered researcher access to the police occupation, but the necessarily selective samples of these solo researchers appear to have overstated the uniformity of police practice (Skogan and Frydl 2004, p. 27). SSO researchers have observed considerable variation in the way police use their authority, and some have shown the high degree of variability that may be found with the same officer over time. Precision is also accomplished through the sequencing of events and the detailing of context, matters that may not be well documented by official records or accurately recalled by participants when interviewed – for example, how police encounters with the public escalate into rebellion or violence (Sykes and Brent 1983). Sometimes SSO precision derives from the application of complex standards or expectations to the practices of persons with obligations to perform in particular ways. For example, the extent to which legal actors conform to constitutional standards or a professional standard can be assessed. Further, SSO can be used to determine the extent to which justice officials comply with the preferences of citizens they encounter or whether citizens comply with the preferences of justice officials in everyday situations (e.g., Mastrofski et al. 1996).

SSO is especially desirable when the question demands detailed knowledge of situations, conditions, or processes that are not otherwise well illuminated or where there is reason to question the validity of knowledge based on other forms of data collection. SSO may also be useful in studying people who might find it difficult to provide an objective or accurate account of what the researcher wishes to know (such as their behavior and the context of that behavior in highly emotional situations). Where there are strong temptations to omit, distort, or fabricate certain socially undesirable features, such as illegal, deviant, or otherwise embarrassing situations, SSO offers an independent account. This is, for example, a limitation of survey-based citizen self-reports of encounters with police to deal with a problem caused by the survey respondent, and especially problematic if there is systematic variation in the degree of error across important subgroups within the sample, for example, according to race.

While much of the SSO research has focused at the level of individual persons as decision-makers, the 1980s saw the beginning of studies that use an ecological unit, such as the neighborhood block face, as the unit of SSO analysis. Noting that neighborhood residents find it difficult to offer accurate descriptions of their neighborhood's social and physical environment, Raudenbush and Sampson (1999) highlighted the value of an "ecometric" approach that uses SSO in conjunction with neighborhood survey research to more fruitfully characterize the state of neighborhood physical and social structure.

SSO may be especially well suited to situations and events where all of the relevant actors and events pertinent to the phenomenon of interest can be observed from start to finish in a limited, well-defined time period. For example, the police decision on how to deal with a traffic violator is clearly bounded in time and place. To the extent that the decision is heavily influenced by the context of the immediate situation (e.g., the offense, the evidence, the driver's demeanor), the decision on how to treat the traffic offender can be captured by SSO and has been especially useful in understanding informal sanctions, which are not completely captured in official records. In general, SSO lends itself to observing phenomena that occur either with high frequency, such as drivers' noncompliance with speed limits on public highways, or at predictable times and places, such as criminal trials, scheduled meetings between probation officers and offenders, or even field tests of prison security systems. Events that occur less frequently, such as acts of social disorder in public places, may require considerably more observation time to obtain reliable estimates (Raudenbush and Sampson 1999), or they may be so infrequent and unpredictable as to make SSO simply impractical, such as the police use of lethal force or the life course of criminality in a sample of individuals. One of the most frequent uses of SSO has been investigating how criminal justice workers operate in the context of role expectations generated by their organization or profession. Primarily focused on police, SSO research in criminology has been very concerned with how officers negotiate the tension between the formal (legal, bureaucratic, and professional) standards set for them and those that issue from the occupational culture. SSO could also be applied to role conformance in the context of informal or illegitimate organizations, such as gangs.

SSO is often used in conjunction with other forms of observation. Some studies have used SSO to measure the extent to which treatment conditions in randomized trials have been maintained (Sherman and Weisburd 1995, p. 685). SSO data have been linked to other forms of data collection on research subjects, such as census data (on neighborhoods), survey interviews of police officers, and follow-up interviews with citizens who were observed in encounters with police. And sometimes SSO is used to supply data not otherwise available, such as objective measures of the physical and social disorder in urban neighborhoods (Raudenbush and Sampson 1999).

Perhaps the most frequent reason that criminologists have turned to SSO is their dissatisfaction with the data they could obtain by other means, such as official records and survey research (Buckle and Farrington 1984, p. 63). Self-report and victim surveys have a number of biases and limitations, but since these are different from those inherent in SSO, it can provide an alternative perspective on specific phenomenon relative to those approaches (Parks 1984).

Unit of Analysis and Sampling

Planning the selection of what is to be observed is an essential element for SSO. Like survey interviewing, SSO requires a careful focusing of what is to be observed and makes it possible to estimate parameters and evaluate error. The first step requires establishing the unit of analysis. Given that much SSO focuses on social interactions, there are three distinct approaches (McCall 1984, pp. 268–269). One uses a time period as the unit of analysis, observing what happens within each discrete time segment, such as what behaviors police officers display in a 15 minute segment of time or the level of social disorder on a street segment during a one-hour period. Another uses a behavior or act as the unit, tracking the sequencing of different behaviors over time, such as the behavioral transactions between officers and citizens who engage each other. The third approach is to socially construct an event as a unit of observation, such as a face-to-face encounter between a police officer and citizen (Reiss 1971) or a public meeting between police and members of a neighborhood organization (Skogan 2006).

Once the unit of analysis is decided, the researcher must consider the sampling frame. The same principles of sampling apply to SSO as any other data collection method, such as survey research. The researcher must consider where and when the units of interest may be found and determine an efficient method of capturing a representative sample. An example of a straightforward sampling strategy is an SSO study of shoplifting that randomly selected shoppers entering a store, systematically varying the location of observers among entrances to the store (Buckle and Farrington 1984).

SSO researchers often use more complex sampling strategies focusing on geographic space. They have sampled police beats and specific days and times within them, oversampling places and times where higher levels of police-citizen encounters were expected. Some researchers rely upon the observed subjects making their own choices as to where observers conduct their observations. This makes sense when the object of study is a specific research subject, but when the object of study is the geographic entity itself, an independent sampling plan is required. For example, a study of public order in a park required researchers to conduct hourly park patrols to observe and record activities of persons by location within the park (Knutsson 1997). Some researchers have used a smaller geographic unit than a police beat or park. Several studies use the face block to apply SSO to the measurement of disorder on public streets, defined in terms of traces of physical and social disorder (trash, graffiti, loitering, public intoxication) (Sampson and Raudenbush 1999). In one study, observers drove through street segments videotaping what was viewable from the vehicle (ibid). Others have performed live observation at the “epicenter” of each block face (best location to observe the most activity), randomly selecting short periods of time for observation from that location and recording them on check sheets (Weisburd et al. 2006). But observers could focus on single addresses, as might be done if one were interested in observing the extent of different kinds of desired and undesired social activity at crime hot spots. Even smaller spatial units have served as the sampling frame. A study of the relationship between crowding and aggression in nightclubs selected high traffic areas within the establishment (10 m^2) to observe levels of patron aggression for 30-min time periods (Macintyre and Homel 1997).

While much of the extant SSO research must develop time- or area-based sampling frames that capture unpredictable or unscheduled events, some SSO studies have focused on scheduled events, such as the delivery of therapeutic community programs in correction institutions or the previously mentioned police-community neighborhood meetings. Sampling of regularly scheduled events is common in research on educational practices and physician behavior, a practice easily replicated for certain aspects of the legal process of interest to criminologists.

Sometimes the practicalities of conducting successful field observation make the research vulnerable to sample biases. In cases where consent of those to be observed must be secured, a clear bias is introduced when those who refuse to be observed differ in their behaviors from those who are willing to be observed. The practical requirements of observation can introduce bias as well. For example, the observation of disorder on Chicago block faces required light that was sufficient for observation only between 7 am and 7 pm (Sampson and Raudenbush 1999), meaning that researchers were unable to measure many forms of disorder that mostly occur in the darkness. It would also be challenging to observe many aspects of law enforcement inquiry and exchanges in the investigative and prosecutorial processes, because much of the effort is not limited to face-to-face encounters, but rather occurs through telephone and computer, modes of communication that may necessitate very different sampling frames and observational methods. Particularly challenging are studies that require a sampling of *cases* rather than individual decision-makers, inasmuch as it is difficult to track and observe the behavior of many different persons who may be involved in making decisions about a case.

Instrumentation

Principles that apply to other forms of research also apply to the creation of instruments for structuring and recording SSO. Sometimes the instrument takes the form of a tally sheet or log for recording the frequency at which phenomena were observed, such as counting disorderly elements at block faces (Sampson and Raudenbush 1999) – or the timing and duration of events, such as police presence in a hot spot (Sherman and Weisburd 1995). Often the instrument takes the form of a questionnaire that is directed to the observer. A study of police use of force, for example, might ask observers to code a series of close-ended questions about the citizens involved in an incident (their personal characteristics, their appearance, their behavior), the behavior of police (how much and types of force used), and other features of the situation (location of the event, its visibility, the presence of bystanders).

SSO instruments have the desired effect of focusing observers' attention on items selected for observation. Field researchers have demonstrated a substantial capacity to recall the relevant features of long sequences of these events, given the repetitive use of the protocols. Nonetheless, greater complexity in the coding system heightens the risk of error. The accuracy of such recall is undoubtedly variable, but researchers have not assessed most of the correlates of recall accuracy (e.g., observer characteristics, instrument characteristics, and the observational setting).

Recording Observations and Use of Technology

Two issues arise in recording of phenomena observed through SSO: (a) whether it is contemporaneous with the observation or later and (b) whether technological recording devices are employed. Resolving these issues requires choosing the highest priority and what must suffer as a consequence. The more contemporaneous the recording of an observation, the less the vulnerability to recall error and various forms of bias, but in many cases, the act of recording may increase the reactivity of the observed parties to the process of being observed, as, for example, when observers posing as shoppers follow actual shoppers to observe whether they are shoplifting (Buckle and Farrington 1984, 1994).

Employing technological aids is usually intended to increase the accuracy or detail of observation from that which would be otherwise available. Handheld electronic recording devices have been used in observing police-public interactions and in observing the social and physical environment of neighborhoods. Audiotaping of calls for service to police telephone operators has been used to gather data on police workload. Videotaping neighborhood block faces from a slow-moving motor vehicle has been used to observe neighborhood disorder. Use of handheld personal digital devices allows contemporaneous observation and recording of brief, frequent events and is most practical when the number of aspects to be observed per event is small in number, which minimizes the interference of recording events occurring in close succession with observing them (McCall 1984, p. 272).

The major advantage of initially recording events electronically and then encoding those records later for analysis is not only the elimination of recall problems but also that more detailed and accurate observations may be made, and the testing of interobserver reliability is facilitated. Further, field researchers may in some instances feel safer when they make the initial recording of their observations from the security of a moving vehicle or when events are recorded with a remote and unobtrusive device (time-lapse photography of a street corner). Nonetheless, there are a number of drawbacks that have concerned other researchers. Video recordings cannot exactly replicate what all of the available senses would communicate to an observer who was there "in the moment." The use of drive-by photography can be expensive; it raises ethical concerns because denizens of the neighborhood find them intrusive and perhaps anxiety producing; it may raise legal and human subject protection issues if alleged criminal acts are recorded; and they may cause significant reactivity among those being observed. In other instances, however, the pervasiveness of already-present surveillance technology that records observations in readily shared (digital) formats (closed circuit television in public and mass-private settings or in-car police video cameras to record traffic stops) may afford researchers a relatively unobtrusive source of data that does not encounter these problems. Dabney, Hollinger, and Dugan (2004), for example, used augmented video surveillance to study shoplifters. The researchers followed a sample of drugstore customers at a single location equipped with high-resolution video surveillance cameras. They were able to determine which customers engaged in shoplifting and coded data on customers' personal characteristics, as well as behavior. Increasingly, much technology-based surveillance derives its unobtrusiveness, not from its being unknown to subjects, but that it is taken for granted (Shrum et al. 2005, p. 11). And with the advent of nonlinear editing packages for digital video, the data itself (traditionally analyzed in quantitative or text format) can be readily manipulated and analyzed as images (Shrum et al. 2005, p. 5).

Error, Reliability, and Validity

SSO data are subject to the same range of threats that befall other methods. Error can be introduced by the observer, and issues of reliability and validity of the method must be addressed. Observers can introduce error intentionally (cheating) or unintentionally (bias or reactivity). Cheating is rarely reported in SSO, although its frequency is unknown. It seems likely that most instances of SSO cheating go undetected. Shirking, a more subtle form of cheating, may occur if observers attempt to reduce their workload by failing to record events that would require writing extensive narratives and structured coding. There has been no direct systematic assessment of the extent and impact of this form of shirking in SSO, but one researcher did examine the effects of time on the job (a proxy for burnout) on researcher productivity but found that productivity was not significantly related to time on the job lower (Spano 2005, pp. 606–608).

Sources of unintended biases in SSO are the mindset and prejudices that observers bring to the field or develop on the job. These may affect what they observe and how they interpret it. The research exploring these issues for SSO does not offer clear and consistent findings. Reiss (1968, 1971b, pp. 17–18) found that an observer's professional background (law student, sociology student, or police officer) did have consequences for some types of information, but not others. A later study attempted to determine whether a statistical relationship between observed police orientation to community policing and officer success in securing citizen compliance could be attributed to observers' own views on community policing (Mastrofski et al. 1996, p. 295). A clear association was not found between the observers' attitudes and the effects that their analysis produced.

Some types of observation judgment are undoubtedly more vulnerable to personal bias than others. Some research, for example, required field observers to judge whether police officers applied excessive force against citizens. But a more effective strategy may be to bifurcate the process into (a) recording narrative accounts of what happened (without asking the field observer to make a judgment about excessive force) and (b) having separate, specially trained experts review these accounts and make an independent judgment about whether they constitute a violation of some standard, legal, or otherwise.

No responsible researcher can casually dismiss the risks of reactivity in SSO, but a number suggest that it is dependent on the context of the observational setting and the nature of the relationship between observer and observed. In SSO of police it has been argued that reactivity to the observer can be reduced by the observer downplaying any evaluative role and emphasizing one's naivety as a "learner" (Mastrofski et al. 1998; Reiss 1968). Yet even this approach may generate more "teaching" or show-off activity in police subjects. Observed officers who engage in such teaching have been reported making contact with citizens to illustrate elements of police work to the observer (Mastrofski and Parks 1990, p. 487). And some types of observers (e.g., females in the presence of male police officers) may produce more of this effect than others (Spano 2007, p. 461).

One of the distinct advantages of SSO over solo field research is that it facilitates testing and improvement of the *reliability* of observations. Early on, much attention was given to the use of multiple observers and estimating their inter-rater reliability. Where many researchers can independently observe the same phenomenon by having multiple observers on scene or by using video recordings, the testing of inter-rater reliability is accomplished by measuring the extent of agreement among the pool of observers for the same set of events. Sometimes disparate independent observations of the same event are resolved by a process of discussion and negotiation. Where multiple independent observations of the same event are not possible, and that is often the case in situations where having more than one observer would be too disruptive, observers might be tested by using their detailed narrative descriptions to determine (a) if they are properly classifying phenomena according to the protocol and (b) the extent of agreement among persons who use those narratives to make classifications. For example, this has been done for characterizing a wide range of police and citizen behaviors in predicting citizen compliance with police requests (McCluskey 2003, pp. 60–74).

Recently, SSO researchers have broadened their reliability concerns to incorporate measurement accuracy and stability. Raudenbush and Sampson (1999) apply psychometrics to the development of "ecometrics" to better understand the error properties of SSO data gathered from observing physical and social disorder in urban neighborhoods. They adapt three psychometric analytic strategies: item response modeling, generalizability theory, and factor analysis to illuminate the error structure of their observational data and to make judgments about the best ways to limit different sources of error in future observational studies. For example, they find that physical disorder can be more reliably measured at lower levels of aggregation than social disorder, due to the much lower frequency of the latter in their observations (Raudenbush and Sampson 1999, p. 30).

While textbooks often note that field studies are less vulnerable to validity problems than surveys because the method places the observer "there" while events are unfolding, they are of course only valid insofar as they produce data that measure what the researcher intends for them to measure. This suggests that issues of validity are related to the protocols, rules, and frameworks that guide the observation. For example, the validity of observations is weakened when observers make inferences about motives or psychological conditions, which cannot be directly observed.

SSO Contributions to Criminology and Future Opportunities

SSO data have made major contributions in two areas of criminological research: the behavior of rank-and-file police officers and the measurement of disorder in urban neighborhoods. SSO has dominated the empirical research on the discretionary choices of police officers: making stops and arrests, issuing citations, using force, assisting citizens, and displaying procedural justice (Skogan and Frydl 2004:ch. 4). One of SSO's special contributions has been the scope of explanatory elements made available for the researchers' models. These include many details of not only the officer's behavior but also the context in which it occurs (nature of the participants, their behavior, the location, and the neighborhood). SSO has also been instrumental in detailing the nature of the *process* of police-citizen interaction, illuminating the interactive quality of temporally ordered micro-transactions or stages that may occur in even a relatively short police-citizen face-to-face encounter (Sykes and Brent 1983). And SSO has also allowed researchers to observe elements of organizational control and community influence on the work of police officers. For example, researchers can learn more about the influence of police supervisors on subordinates' practices (Engel 2000), the dynamics of police-community interaction and their consequences when police and neighborhood residents deal with each other at community problem-solving meetings (Skogan 2006).

A second area where SSO research has concentrated is the examination of neighborhood physical and social disorder. It has been used to test the impact of police interventions in hot spots, showing that police interventions in these "micro-places" not only reduce crime and disorder, they also diffuse those benefits to nearby areas (Sherman and Weisburd 1995; Weisburd et al. 2006). The largest project in this area has focused on Chicago neighborhoods and has produced a number of insights relevant to the testing and development of theories of the role of neighborhood disorder in causing crime in urban neighborhoods (Raudenbush and Sampson 1999). Using SSO-based measures of "objective" disorder described earlier in this entry, researchers have examined the sources and consequences of public disorder. The research has demonstrated the importance of "collective efficacy" in predicting lower crime rates and observed disorder, controlling for structural characteristics of the neighborhood (Sampson and Raudenbush 1999). Collective efficacy also predicted lower levels of crime, controlling for observed disorder and the reciprocal effects of violence. The researchers found that the relationship between public disorder and crime is spurious, with the exception of robbery, which is contrary to the expectations of the well-known "broken windows" theory of neighborhood decline.

In general, SSO has afforded precision that has in many cases shown the phenomena of interest to be more complex than other forms of data collection had indicated. For example, SSO researchers have found rich variation among police officers in their patterns of discretionary choice and even noted the instability of those patterns for individual officers over time. And the independence of SSO observers from the phenomenon of interest has provided a means to understand the contributing factors to the social construction of phenomena, such as the contributions of a neighborhood's racial profile in assessing its level of disorder (Sampson and Raudenbush 2004).

There are many opportunities to expand the use of SSO to increase knowledge and understanding. Largely untapped is the observation of crime and disorder, especially at the microlevel, where observers have the opportunity to make detailed observations of offenders in the act. SSO studies of shoplifting and aggressive or disorderly behavior in bars and clubs show that this is feasible where observers can easily blend into the environment. Where that is not possible, access to unobtrusive surveillance technologies appears to offer opportunities for detailed observation that reduce reactivity concerns. It is highly likely that criminologists will take advantage of the ubiquity of electronic surveillance to capture events that would otherwise be costly to observe. For example, the growing sophistication of surveillance and identification technology may make it possible to use facial identification software to gather data for a network analysis of persons who frequent hot spots. This includes not only the growing use of video recording devices by government and private sector organizations but the now ready availability of miniaturized recording devices to the general public (through cell phone recording devices).

In searching for efficient ways to use SSO, criminologists will likely capitalize on the growing body of evidence about the predictability of crime and disorder occurring in small geographic spaces. Because much "street" crime is so highly concentrated in a relatively small portion of addresses or face blocks, the location of observers or observational devices can very efficiently generate lots of information on what occurs, especially in public areas. In addition, given heightened levels of obtrusive surveillance in public places, SSO should prove an excellent way to understand how security and surveillance operate, why certain methods are effective, and the collateral impacts of various methods of monitoring and control designed to increase public safety.

Another venue for SSO to be used fruitfully is in experimental studies. SSO can be used to measure key aspects of the process that presumably operate to link treatments to outcomes. For example, if the physical redesign of bars and serving practices of bartenders are intended to reduce violence in those establishments, do patrons in fact alter their patterns of behavior in the ways that are expected to produce less violence (Graham et al. 2004)?

Conclusion

Four decades ago Albert Reiss showed criminologists the utility of systematic social observation, but it remains a method used infrequently. This is due in no small part to two things. First, criminologists are rarely exposed to training and opportunities to do SSO during their course of study. Second, those who know a little about it may often expect that it requires more time and resources than they have available. This may indeed be the case. Many projects could be taken on a smaller scale with a narrower scope of questions than the better-known, large SSO projects, especially if technological advances are leveraged for both recording and coding purposes. Some researchers may decline to use SSO because of reactivity concerns, but the available evidence suggests that these problems are often manageable and may be no more severe in any event than found with other data gathering methods.

Increased use of SSO will undoubtedly attract and stimulate greater scrutiny of its limitations, as well as its advantages. The error properties of most SSO data sets have been underexplored, and more attention here is needed. Expanding the use of SSO and more comprehensively assessing its strengths and limits could be fruitfully combined into a more comprehensive assessment of other methods of gathering data on crime and justice phenomena.

SSO deserves the consideration of researchers because of its many advantages. It offers enhanced prospects of validity for the study of crime and justice phenomena, and it increases confidence in reliability because of the researcher's direct access to the phenomenon of interest and greater control and transparency of data encoding. It affords greater precision in capturing details of the phenomenon and its context, such as the sequencing of what happens before, during, and after those events. In many cases it may be the least problematic method for acquiring information. Criminology, which has strong roots in the traditions and methodologies of sociological research, remains heavily reliant on the use of sample surveys and official records (McCall 1984, p. 277; Reiss 1971). But as the field matures and diversifies intellectually, more of its researchers may with justification be inclined to make systematic social observation the method of first, not last, resort.

Related Entries

- ▶ Defining Disorder
- ▶ Hot Spots and Place-Based Policing
- ▶ Measuring Police Performance
- ▶ Methodological Issues in Evaluating Police Performance
- ▶ Police Discretion in Law Enforcement
- ▶ Police Discretion in Providing Services and Assistance
- ▶ Police Legitimacy and Police Encounters

Recommended Reading and References

Arrington RE (1943) Time sampling in studies of social behavior: a critical review of techniques and results with research suggestions. Psychol Bull 40(2):81–124

Buckle A, Farrington DP (1984) An observational study of shoplifting. Br J Criminol 24(1):63–73

Buckle A, Farrington DP (1994) Measuring shoplifting by systematic observation: a replication study. Psychol crime law 1(2):133–141

Dabney DA, Hollinger RC, Dugan L (2004) Who actually steals? A study of covertly observed shoplifters. Justice Q 21(4):693–728

Engel RS (2000) The effects of supervisory styles on patrol officer behavior. Police Q 3(3):262–293

Graham K, Osgood D, Zibrowski E, Purcell J, Gliksman L, Leonard K, Pernanen K, Saltz R, Toomey T (2004) The effect of the safer bars programme on physical aggression in bars: results of a randomized controlled trial. Drug Alcohol Rev 23:31–41

Knutsson J (1997) Restoring public order in a city park. In: Homel R (ed) Policing for prevention: reducing crime, public intoxication and injury. Crim Justice, Monsey, pp 133–151

Macintyre S, Homel R (1997) Danger on the dance floor: a study of interior design, crowding and aggression in nightclubs. In: Homel R (ed) Policing for prevention: reducing crime, public intoxication and injury. Crim Justice, Monsey, pp 91–113

Mastrofski S, Snipes JB, Supina A (1996) Compliance on demand: the public's response to specific police requests. J Res Crime Delinq 33:269–305

Mastrofski SD, Parks RB (1990) Improving observational studies of police. Criminology 28(3):475–496

Mastrofski SD, Parks RB, McCluskey JD (2010) Systematic social observation in criminology. In: Piquero A, Weisburd D (eds) Handbook of quantitative criminology. Springer, New York, pp 225–247

Mastrofski SD, Parks RB, Reiss AJ Jr, Worden RE, DeJong C, Snipes JB, Terrill W (1998) Systematic observation of public police: applying field research methods to policy issues. National Institute of Justice, Washington, DC, NCJ 172859

McCall GJ (1984) Systematic field observation. Annu Rev Sociol 10:263–282

McCluskey JD (2003) Police requests for compliance: coercive and procedurally just tactics. LFB Scholarly Publishing LLC, New York

Parks RB (1984) Comparing citizen and observer perceptions of police-citizen encounters. In: Whitaker GP (ed) Understanding police agency performance. U.S. Government Printing office, Washington, DC, pp 121–135

Raudenbush SW, Sampson RJ (1999) Toward a science of assessing ecological settings, with applications to the systematic social observation of neighborhoods. Sociol Methodol 29:1–41

Reiss AJ Jr (1968) Stuff and nonsense about social surveys and observation. In: Becker HS, Geer B, Riesman D, Weiss RS (eds) Institutions and the person: papers presented to C. Everett Hughes. Aldine Publishing, Chicago, pp 351–367

Reiss AJ Jr (1971) Systematic observation of natural social phenomena. In: Costner HL (ed) Sociological methodology. Jossey-Bass, San Francisco, pp 3–33

Sampson RJ, Raudenbush SW (1999) Systematic social observation of public spaces: A new look at disorder in urban neighborhoods. Am J Sociol 3:603–651, November

Sampson RJ, Raudenbush SW (2004) Seeing disorder: neighborhood stigma and the social construction of "broken windows. Soc Psychol Q 67(4):319–342

Sherman LW, Weisburd D (1995) General deterrence effects of police patrol in crime "hot spots": a randomized, controlled evaluation. Justice Q 12(4):625–648

Shrum W, Duque R, Brown T (2005) Digital video as research practice: methodology for the millennium. J Res Pract 1(1):1–19, http://jrp.icaap.org/index.php/jrp/article/view/6/11

Skogan WG (2006) Police and community in Chicago: a tale of three cities. New York, Oxford

Skogan W, Frydl K (eds) (2004) Fairness and effectiveness in policing: the evidence. National Academies, Washington, DC

Spano R (2005) Potential sources of observer bias in police observational data. Soc Sci Res 34:591–617

Spano R (2007) How does reactivity affect police behavior? Describing and quantifying the impact of reactivity as behavioral change in a large-scale observational study of police. J Crim Justice 35:453–465

Sykes RE, Brent EE (1983) Policing: a social behaviorist perspective. Rutgers University Press, New Brunswick

Weisburd D, Wyckoff LA, Ready J, Eck JE, Hinkle JC, Gajewski F (2006) Does crime just move around the corner? A controlled study of spatial displacement and diffusion of crime control benefits. Criminology 44(3):549–592

Worden RE, McLean SJ (Forthcoming) Systematic social observation of the police. In: Reisig MD, Kane RJ (eds) Oxford handbook on police and policing. Oxford University Press, Oxford

T

TB

► Drug Abuse and Alcohol Dependence Among Inmates

Technology

► Biometrics and Border Control Policing

Telemetric Policing

Pat O’Malley
Sydney Law School, The University of Sydney,
NSW, Australia

Overview

In the past 50 years, policing has undergone a major change as a result of the development of new technologies of telemetric surveillance. The new surveillant technologies, in theory at least, are perfectly accurate, tireless, produce photographic evidence, cost little to operate after installed, and may generate revenue through fines. Telemetric policing technologies move one step further, linking surveillance with electronic identification of offenses through built-in algorithms. In other words, they *perform* a police function by registering offenses and infractions directly. The information revolution has enormously enhanced the impact of this by embedding telemetric policing into a fully digital “simulated” justice wherein all phases of justice from detection to sentencing to expiation of sanction can be performed online. Telemetric policing with respect to traffic has been credited with reducing offending, rates of injury and death, and costs of policing. However, many of the advantages attributed to it have been challenged by conflicting data, and the instrumentation itself has proven fallible. Resistance has also focused on telemetric policing as merely revenue raising, on human rights grounds, fears of “Big Brother,” and demands for more personal forms of justice. Nevertheless, it continues to expand: new technologies allow automated intervention in vehicle control, and applications of the telemetric model may be extended to many minor public order infractions.

Telemetric Policing Distinguished

“Telemetrics” refers to measurement at a distance, and telemetric policing is thus an assemblage of actors, instruments, technologies, and so on that operates through the remote measurement of offending. It involves the electronic apparatuses of surveillance now familiar parts of everyday life: CCTV, traffic “safety cameras,” bar code readers, magnetic strips, RFIDs (radio frequency identification devices) and so on. Not all of these measure offending,

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

or at least, not most of the time. CCTV, for example, usually operates as a "dumb" technology recording actions in its field of view but doing nothing with the images, which are accessed (if at all) when policing agents seek information. Tracking the movements of the London underground bombers via CCTV turned out to be an important means for identifying offenders. But this process relied upon live police agents scanning the hours of recordings. Even if aided by facial recognition software – important because it may improve the speed, reliability, and costs of police identification – the assemblage remains passive, merely recording specific presences or absences. Whether an offense has occurred remains a question for the interpretation and intervention of live policing agents. Such forms of policing technology have rightly been the focus of a considerable volume of research; they merge criminology with some of the most exciting new areas of social theory, and have given rise to an enormous volume of political attention and concern (Norris and Wilson 2006).

Because much of the development is quite recent, and perhaps because there is no clear cutoff point from what could be referred to as such surveillant policing, police telemetry has been subject to rather less attention. Yet when algorithms are built into the electronic platform of surveillance to register problematic identifiers (e.g., license numbers, passport chips) or patterns of action, a major step toward telemetric policing has been taken, into "automated surveillance" or "algorithmic surveillance" (Norris et al. 1999). The use of CCTV to identify patterns of movement associated with potential suicides on the subway, or processing of a variety of electronic identifiers on passports are two examples. With the former, the action identifies a risky pattern of action, alerting live agents and calling for intervention (Graham and Wood 2003: 236). It remains, however, merely a televisual record and live agents must both interpret and intervene. With the example of the passport, things become more complex. It may not simply be the identity of the passport holder that trips a risk switch, but a variety of other data linked to that identity through various data banks, algorithms, and so on. Amoore (2011), for example, outlines the ways in which algorithms modeled on those used in derivatives markets throw up alarms based on correlations between such factors as the previous patterns of travel, the traveler's destination, or whether the airline ticket was paid for in cash. But while these are cutting edge surveillant technologies – venturing into a regime of estimations and suggestive possibilities – they still do no more than mobilize agents to investigate further.

Telemetric policing is achieved where one more step is taken, and the offense itself is identified through informatic calculation: by a machine deploying an algorithm, that does not necessitate the intervention of live agents. It is, literally, *a policing machine* (O'Malley 2010b). Policing agents may then be mobilized, for example, to effect apprehension of the offender, but identification of the offense has been carried out by the machine. One example that has received attention is electronic tagging of those under a restraining order forbidding access to or exit from certain spaces. Through satellite telemetry, any breach of the order – for example approaching too close to a schoolyard or a former spouse's residence – can be registered automatically (Jones 2001). The offense may still have to be confirmed by later investigation. For example, if the subject fled the house because it was burning down. But as far as the policing process goes, the offense *prima facie* has been registered and apprehension and arrest normally will follow.

Yet while such forms of active surveillance are within the territory of telemetric policing, a still more familiar assemblage takes things one step further by dispensing with apprehension by police, at least in the vast majority of cases. The diverse systems of policing traffic through "safety cameras" – speed cameras, red-light cameras, point-to-point cameras, and so on register an offense through telemetry. They also issue a penalty notice in the form of a reverse-onus money sanction (and/or issue license demerit points and license cancelation). In the vast majority of cases – generally well in excess of 90 % – the policing process is thus entirely operated through information technology, without the intervention of live agents. Where fines are paid electronically or by check through the mail, as is normal, no live agents have been involved even to the point of sanctioning: this is simulated justice (O'Malley 2010b).

An understanding of the nature, significance, and potential of telemetric policing is thus best effected through analysis of these traffic assemblages. Far from "trivial," these operations of telemetric policing very likely constitute the bulk of citizens' experience as subjects of police interventions. They result in the imposition of significant fines and/or license cancelation, and in the case of dangerous driving charges (or nonpayment of fines), quite possibly involve imprisonment.

The Rise of Telemetric Policing

In 1935, a new innovation hit the streets of downtown St Louis, Missouri. Little more than a coin-driven alarm clock on a pole, it contained within it the key elements of telemetric policing. The parking meter itself was the solution to a series of challenges. As downtown parking became an acute problem in the USA, it became necessary to optimize the circulation of traffic, shoppers, and workers. Parking time – limits were established early on, but these soon proved to present significant policing problems. Sworn officers were deployed to ensure that drivers did not overstay the time limits, primarily by chalking tires and returning to book those who infringed. This was expensive, absorbing a considerable amount of police time, but was made more so by its openness to dispute. Drivers claimed that tire marks were from some previous occasion and disputed the officers' time calculations. Defense lawyers exploited every loophole and ambiguity. Police were taken off the beat to testify in court, adding considerably to the cost of policing, while vulnerability to contestation in prosecutions made for many other inefficiencies in policing (Fogelson 2001).

While parking meters did not sweep all these problems away, they had a major effect. It was far easier for infringements to be detected and established, police time was saved and court disputes over timelines and tire marks dropped away. The invention spread rapidly, and often the "objective" nature of the parking meter made it possible to displace expensive sworn police with cheaper "parking officers." What had been an alarming demand for ever-growing volumes of police time was significantly curbed as policing had in key respects become automated. Telemetric justice of the twenty-first century can be seen simply as an electronic extension of this mechanical innovation, for it displaced police through the process of automatically registering an offense or infringement – albeit it that this still required policing intervention to record and process the matter.

Apart from the technology, a second aspect of this development was that citizens were paying on the spot for being policed. Fees and fines made this machine-governed process a form of self-funding regulation. The machines did not merely displace expensive police time; they produced revenue that could be used to fund more machines and other costs of enforcement. In principle therefore, it illustrates a key potential of telemetric policing: it is capable of almost infinite expansion. In pursuit of further economies and efficiencies in policing, a new generation of meters today takes timed and dated digital photos of the registration number of any vehicle overstaying a limit, and transmits this to a central computer that issues a fine. Even parking officers are displaced by virtual police: this is the maturation of telemetric policing. Indeed, a huge volume of policing already has become virtual in this way, transforming the policing not merely of parking but – in the name of public safety – of many "moving offenses" such as speeding, crossing double white lines, and running red lights (Fogelson 2001).

At the turn of the twentieth century speeding, dangerous driving and similar offenses presented problems similar to those giving birth to parking meters. The principal technology for assessing speeding was either the stopwatch, the subjective judgment of the police office, or the speedometer reading of a pursuit police cruiser. All were time consuming, and all were seen to require sworn officers – in good measure because all were easily and often challenged in court. Gaining convictions was difficult especially because motoring then was an elite activity. High status motorists faced courts often unwilling to proceed against such "reputable" citizens. As well, speeding was widely regarded as an illegitimate offense – a "tax on progress" and an excuse for revenue raising (Simon 1998). If charges of dangerous driving existed, then what could the justification be for merely speeding? Motorists' organizations mounted further pressure, and motorists of the time could afford the best in legal representation. Failed prosecutions were frequent. So draining was this of police time, and so difficult to achieve success, that police avoided pressing charges: indeed the British Home office had to issue instructions that speeding had to be enforced (Plowden 1971).

The rise of the automobile as a prime feature of modern life exacerbated problems by vastly increasing the volume of offending. And if courts became more sympathetic to police, nevertheless police prosecutions still relied on primitive technologies of detection vulnerable to court challenge. From an early date, American states attempted to deal with this by establishing traffic courts with greatly simplified procedures. But speeding retained its status as a "merely technical" offense in the eyes of many. The great change in this respect – and vital to the rise of telemetric policing – was the rise during the 1970s of risk management as a widespread technique for government. During this period, alongside drink driving, speeding changed its character. Now increasingly it appeared not as a merely "technical" offense, but as an "objectively" demonstrable risk factor for traffic casualties (O'Malley 2010a). As the road toll emerged as a high-profile problem for governments at all levels, they mounted major campaigns against speeding and drink driving. This move was matched both with increased levels of policing and the application of new technologies. Most significantly at this time this meant mobile speed cameras in police cruisers, and within a few years, fixed speed cameras and more recently point-to-point cameras. The development of fixed speed cameras was a key step in the genealogy of telemetric policing: policing that was *completely performed* (not merely assisted) by remote electronic sensing devices.

It is important to recognize that telemetric policing was shaped by risk-based governance in two ways. First, by directly linking these offenses – especially speeding – to public safety by "objective" scientific means, a new strength was given to the morality of traffic law. Public and personal safety came to be "the" issue in a new and demonstrable fashion. Second, this moral jurisprudence was transformed by risk's creation of a *quantitative morality*. Analysis of speeding and red-light violations in terms of risk calculation created a precisely measurable offense. As speed increased so too did risk and thus the moral wrongness of the offense. And with this went increases in the penalty. The new telemetric machines could not only simply detect offending, but also calculate seriousness of the offense. The next step, logically, is for them to calculate the penalty also.

Risk techniques and informatic technologies are thus vital to telematic policing. But no understanding of the phenomenon, its implications and resistances to it, is possible unless the crucial role of a third element – money penalties – is recognized. Because the penalty predominantly appears as "only money" rather than imprisonment (sidestepping liberal shibboleths of liberty and freedom), this has permitted a considerable streamlining of legal procedure (O'Malley 2010a). Even before the emergence of telemetric policing, the sheer volume of traffic offending led to a series of innovations aimed at speeding up the processing of cases, culminating in the "on the spot fine" or "infringement notice" in the 1960s. Their most salient features include a fixed monetary penalty and a reverse-onus provision (Fox 1996). These innovations created the conditions of existence for telemetric policing to become part of "simulated justice" – that is, justice completely transacted in the virtual environment. Once fixed penalties apply, then a computer can receive the information transmitted by a camera – identifying vehicle and its precise speed – and calculate the penalty. The computer can then issue the infringement notice and email or mail it to the presumed offender. (Usually this is the registered owner of the vehicle. But even where the owner claims another to have been driving, this involves only a minor bureaucratic formality.) It is here that other key features of money come into play. Money is anonymous and can be digitized. Because one lot of money cannot be told from another, courts have had to allow that fines may be paid by someone other than the offender. This makes possible payment remotely, by check. In turn, because money can be digitized, it is a small step to allow payment through electronic banking. At this point, justice itself becomes virtual: electronic and informatic. But because, nevertheless, it retains its reality – imposing real sanctions on real people – the real and the virtual are identical: hence "simulated" justice (O'Malley 2010b).

In turn, the rise of simulated justice means that telemetric policing substantially displaces police from court appearance in relation to traffic offending. Apart from the reverse-onus assumption in infringement notices, incentives for offenders not to go to court are manifold. Fines paid promptly and without challenge are generally "rewarded" by a substantially discounted fine and often without a conviction being recorded. Money, trauma, and time are saved by transacting the whole matter remotely, and simulated justice virtually secures the public anonymity of the offender (Fox 1996). Again, little of this happened by accident: it was precisely the pressure of mounting police work and backlogs in processing cases that forced through this mode of justice to deal with what is, in effect, the principal modern form of mass disobedience to the law. Further, once set in motion, the process of growth becomes self-sustaining. As telemetric policing, operating 24/7 and with unwavering attention, can detect and process unprecedented numbers of offenses and infringements, so there is little choice for administrators but to expand the realm of simulated justice more or less in lockstep. Conversely, as simulated justice efficiently processes cases and generates revenue to fund itself, so telemetric policing encounters no obstacles to expansion arising from organizational problems of logistics. A world of infinite regulation opens up.

The Impact and Effectiveness of Telemetric Policing

Governments and their agencies routinely promote telemetric policing in the traffic domain. Typical is the New South Wales Road Transport Authority (RTA 2012) in Australia that argues that "mobile speed cameras have been introduced because they are recognised internationally as a best practice road safety countermeasure to reduce speeding, leading to a reduction of crashes. The introduction of mobile speed programs in Queensland and Victoria has reduced casualty rates in those states by at least 25 %" (RTA 2012). With respect to fixed safety cameras – which may be speed cameras, red-light cameras, or both – the claim is made that "the use of cameras to enforce speeding has proven road safety benefits." It points out that where cameras have been installed, there has been a 70 % reduction in speeding, resulting in a 90 % decline in fatalities and a 23 % reduction in injuries. These are very high estimates. In an international study, Thomas (2008) found that fixed cameras reduced the number of personal injury accidents by between 20 % and 25 %, figures that correspond to the findings of Gains' (2005) analysis of UK camera sites, while Ercke and his colleagues' (2009) international meta-analysis indicated a 35 % reduction in personal injury crashes. If these claims are taken at face value, then as noted, they are a powerful justification for rolling out even more widespread and sophisticated telemetric policing. However, there is no shortage of contradictory evidence. The most frequent challenges involve indications that data are far more varied and unclear than is admitted to be the case by police and road safety authorities.

One general criticism is that such figures ignore the effects of regression to the mean. In other words, cameras are installed in high accident areas, but accidents in such areas will tend to revert to the mean level even without any action being taken. Figures showing a reduction in accidents after installation rarely allow for this, and consequently exaggerate impacts (Mountain et al. 2005). Likewise, it is frequently argued that speed cameras displace traffic to other routes, and thus, their effectiveness in one site needs to be set against possible rises in accident frequencies elsewhere. This is argued to explain why, in many jurisdictions, traffic cameras are effective in specific locations but overall rates of accidents and casualties do not decline. Indeed, in some settings, speed cameras appear to have *increased* accident rates. One study reported that in Britain, 28,000 road accidents have been triggered by speed cameras as drivers slowed down ahead of them and then sped up once passed (Cavenagh 2008). The introduction of speed cameras and new speed limits in Australia's Northern Territory was associated with a 36 % increase in road deaths, and in Cumbria, England, they were reported to have risen by 40 % after the introduction of speed cameras (Cavenagh 2008). A UK Ministry of Transport (2008) study showed a 55 % increase in injury accidents when speed cameras are used on highway work zones, and a 31 % increase when used on freeways where there were no road works. Likewise British government figures "showed for 32 speed camera sites there were an average of 48 more accidents involving death or serious injury" over the previous 12 months, while at 38 red-light camera sites, there had been an average of 62 more accidents (Williams 2005: 5).

In Britain, such evidence led to decisions by local councils in 2010 to scrap the widespread use of safety cameras because of doubts as to their effectiveness. Prior to this, in 2005, the UK government had blocked the installation of nearly 500 new speed cameras on the same grounds. The 38 traffic safety partnerships' nationwide (involving local government and police forces) were ordered not to use cameras in any new sites (Wilson 2008). It is worthy of note that the Association of Chief Police Officers condemned this ban saying that it could cost lives because dangerous roads were being left unprotected by cameras (Webster 2011). As in Australia and other jurisdictions, police in England and Wales generally are heavily supportive of such telemetric policing. But this argument seems largely to have been ignored in the face of contradictory evidence.

As the British were winding back speed camera installation around 2010, other objections emerged in that country. It was reported that the Department for Transport was reviewing the rules on deploying cameras after concerns that partnerships had failed to consider alternatives such as improving junctions or erecting warnings. Even the Parliamentary Advisory Council for Transport Safety had admitted that while it supported speed cameras, "in some cases partnerships may have chosen to install a camera when an engineering solution may have been better" (Ministry of Transport 2008). Findings by Mountain et al. (2005), for example, indicate that engineering solutions such as speed humps have twice the impact on lowering accident rates (44 %) than do speed cameras (22 %). It has been argued strongly by opponents that cameras represent a considerably cheaper and easier option than engineering – even if less effective – where concerns had arisen about accident rates. Indeed, in 2010, the UK Road Safety Minister stated that the government would cut funding because local governments had relied too heavily on safety cameras for too long, and the focus needed to be on other safety measures.

Red-light cameras have come in for an almost equal amount of criticism. Overall, research reports indicate that their benefits are mixed, with a tendency for right angle crashes (which are more often fatal and injurious) to decrease, but for tail-end accidents to increase significantly – largely attributed to drivers braking suddenly at camera-controlled intersections. However, some studies, including a 7-year study of red-light cameras in the District of Columbia, conclude there is no evidence of reductions in accidents or injuries, and there was evidence of significant increases in certain locales. As a result of such mixed findings, several US states that had installed red-light cameras abandoned the program, including Virginia and Hawaii, while an array of others banned their future use (Texas House of Representatives 2006).

Finally, claims to the success of speed cameras have been undermined by challenges on various technical grounds. Concerns have arisen around demonstration that individual speed cameras exaggerate recordings, resulting in unjustified prosecutions. In 2011, Alberta canceled around 100,000 speeding tickets issued in the previous 14 months due to concerns with the accuracy of equipment (Edmonton 2011). The switching-off of speed cameras on Melbourne's Ring Road system in 2004 was the direct consequence of an admission by the government that possibly thousands of drivers had been fined as a result of faulty cameras.

By challenging the scientific standing of the jurisprudence of safety in key respects, these difficulties represent a major source of vulnerability of telemetric policing: its apparent objectivity and justification in safety. Perhaps precisely because telemetric traffic policing's moral ground is that of safety, and because its techniques for producing safety are scientific technologies, in no other domain of justice does the practice of policing appear so vulnerable to empirical critiques of ineffectiveness.

Public Resistance

As may be seen by reference to the history of speeding, without the justification of harm-minimization, telemetric policing and simulated justice may appear simply to tax motorists, and thus come to be regarded either as an unjust tax or as an implicitly illegitimate source of state revenue masquerading under a guise of security. Necessarily, this leads to a consideration of telemetric policing's seemingly essential nexus with money sanctions. There could be a reasonable expectation that money sanctions would lessen resistance to telemetric policing, certainly compared with imprisonment. This would render them just another annoying but relatively trivial (for middle class offenders) cost of life in modern society. However, in this respect, fines have proved to be a vulnerable strategy. In the USA, in particular, campaigns around this issue were prominent among factors leading to cameras being turned off in Texas as early as the 1980s, and later in Arizona, Illinois, and Alaska (Texas House of Representative 2006: 6–8). In 2011, the New South Wales government banned further rollout of speed cameras and removed 38 out of 141 cameras after public complaints, because they were shown not to have contributed to safety and thus appeared merely as cash cows. Two of these cameras alone had generated in excess of $2 million in the previous year (*The Sydney Morning Herald*, 28 July 2011, p. 1).

Equally problematic has been the widespread use of arrangements with private security companies to install and operate safety cameras under contract. Commonly, governments would license companies to install and maintain cameras, placed according to official guidelines. In exchange, the company kept a certain proportion of the revenues generated. Unsurprisingly, this model generated very high levels of opposition, on the grounds that it was a direct inducement to maximize fines rather than produce road safety. In many jurisdictions, this practice has been curtailed and been replaced by funding private telemetric policing through fees paid by the relevant government. Even so, public opposition to private contractors operating speed camera systems remains high, and this was a factor in the curtailments to camera operations in Arizona during 2010.

It may be imagined that opposition to telemetric policing could rarely get coordinated: for offenders are widely dispersed geographically and are comparatively un-homogeneous – especially when compared with wealthy early motorists. All the more so, because their offending is not normally a matter of public record and visibility, being sheltered from the public gaze by the online environment of telemetric policing and simulated justice.

While rendering the subjects of telemetric policing publicly invisible, information technologies, especially the internet, unexpectedly have proven an important resource for opponents. Innumerable online discussion forums, blogs, online media publications, websites, and so on have sprung up, focusing on the issue of new traffic-tracking technologies. Such mobilizations are an example of a new form of political coordination that has emerged in a society seemingly fragmented and loosening its ties with identity politics. Separated from large-scale political movements, they form political groupings around single issues – such as telemetric policing – that, as Agamben (1993: 86) expresses it "form a community without affirming an identity." Many of the objections to, and critiques of, speed and red-light cameras were generated, or contributed to, by online politics (O'Malley 2013). As Introna and Gibbons (2009: 238–239) argue, "a virtual network may be a powerful actor through the enactment of *information politics*." They suggest that such information networks connect with traditional media and thus traditional politics becoming "an important element of resisting state surveillance practices."

As could be predicted, and as O'Malley (2013) and Wells and Wills (2009) found, key themes of popular resistance were focused on the problems with the accuracy of the telemetry apparatus. Linked with this, many drivers and online opponents raised questions about the appropriateness of rigid and universal criteria of speed limits. Echoing arguments raised in a theoretical context by Wynne (1996), opponents argued that risk was very much shaped by local road and weather conditions, local knowledge of the road, and the skills of the driver. In effect, a counter-expertise was mobilized which drew upon local knowledge and contributed to critical assessments of the effectiveness and accuracy of the telemetric policing apparatus. Thus, while it may be easy to dismiss such resistance as merely based on anecdotal and "subjective" evidence, the findings reviewed earlier in this entry indicate that there are effective and valid grounds for "local" opposition to universal expertise. As Wynne argues, even the categories of "lay" or "subjective," when applied to such resistant views of risk, may be inherently problematic. Simply because universal knowledge always requires adjustment to specific contexts, he argues that the "local" is no less "expert" just because not based on questionable "universal" and "objective" data.

Apart from an information politics, forms of guerrilla action have emerged. Some activists began destroying cameras. In the UK, in Essex alone, six cameras were set on fire (with each unit costing an estimated GBP 24,000). In one instance, websites created a social bandit out of a figure named "Captain Gatso" (GATSOMETER is the name of a brand of speed camera) who grandly announced a struggle "against an unjust form of taxation" and threatened "increased operations across the country" (Khan 2003). A group called MAD (Motorists Against Detection) which reported itself as having a hard core of 200 members claimed responsibility for having destroyed 1,000 cameras since 2000. The group claimed to use internet chat forums, encrypted email, and pay-as-you-go phones to keep in touch and plan campaigns. In practice, this group may have involved only handful of isolated individuals, although news media estimated that some 700 cameras had been destroyed across the nation with particularly active cells in Wales and London. The importance of such accounts is not so much the reality of the threats they posed. It is, rather, the symbolic role they appeared to occupy in an active online opposition movement. This was mobilized primarily by the perception that money fines were an index of government agencies using the jurisprudence of risk as a cover for oppressive imposition of questionable universal standards and for thinly disguised revenue raising.

Even so, it could be imagined, telemetric policing would offer considerable attractions because it is objective and impersonal, a technical policing apparatus free from accusations of bias that frequently dog physical policing. As well, the procedures are streamlined and anonymous. However, Wells (2008) found that while her sample of speeding drivers objected to what they regarded as an unfair system of justice for all the reasons outlined above, in addition, they objected strenuously to the denial of a "voice." That is, the fact that the automated nature of policing and justice effectively denied them the opportunity to put their side of things, to offer issues of mitigation, criticisms of the placement of cameras, denials that speeding in a specific site was dangerous and so on. Regardless of whether they believed in the opportunity to have their say would make any difference to the outcome, the loss of voice was seen to undermine the legitimacy of this form of justice.

Perhaps, more than any single issue, it might have been predicted that privacy, and also related human rights issues, would have been matters receiving most attention in public opposition and related politics. In the USA, such concerns have been raised as contributory concerns influencing several states (Arizona most recently) to wind back or even shut down camera systems. In part, more specific debates in the USA have been curtailed by argument stemming back to the case of Katz v. United States 389 *US* 347 (1967) that the Constitution does not protect the privacy of those who are engaged in breaking the law, and that there is no privacy interest in what is routinely and regularly displayed in public. Much the same argument is regularly aired in relation to cases before the European Court of Human Rights that found the Human Rights Act 1998 was not being violated where the owner of a vehicle was required to identify the driver in relation to a speeding offense.

Nevertheless, what is perhaps surprising is that issues of privacy and human rights comparatively rarely seem to be pushed by online and other opponents – unlike the issues of accuracy and revenue raising. Wells and Wills (2009), for example, found that in the array of issues objected to by drivers caught by automated speed cameras, privacy, civil rights, and "Big Brother" matters were not major concerns. This is also reflected in O'Malley's (2013) findings in relation to analysis of internet sites dedicated to opposing safety cameras.

Futures of Telemetric Policing

The uneven but marked success of campaigns against telemetric policing has focused in many ways on what could be seen as its perceived strengths: objectivity and impersonality, low cost to state budgets, the low profile of money sanctions, and the discourse of safety. This is more than simply an irony. It suggests that, apart from the matter of inaccurate recording of offenses, these are not simply technological glitches that are likely to be resolved. Rather they go to the heart of the features that make telemetric policing attractive to governments and police. It should therefore not be assumed that telemetric policing will simply advance inexorably. Additionally, new oppositions will spring up as the technology progresses. Two examples are significant.

First is Intelligent Speed Adaptation (ISA), where telemetric policing is enabled to intervene directly in offender behavior by taking over operation of vehicle speed and ignition controls. Such technology has existed for some years and is deployed by private vehicle fleets where warnings are transmitted to an offending vehicle and if ignored, the accelerator and/or brakes may be operated remotely. Of course, to allow for emergencies, these can temporarily be overridden by the driver – in much the same way that cruise control works – but in the event this override proves unjustified, a more serious offense may be registered. As well as speeding matters, erratic driving, red-light transgression, crossing double white lines, and so on are all capable of being governed in this fashion. In addition, it is already the case that stolen vehicles can be located and immobilized, a technology publicly available but not yet deployed widely by police. But in principle, all vehicles could be immobilized remotely for all manner of policing purposes. Such interventions, however, are unlikely to escape strong public reaction, and as the level of intervention escalates in this fashion, it would seem reasonable to expect that opposition would likewise become more concerted and widespread.

Second is the use of facial recognition technologies. While, once more, the technology is now well developed, many police forces have as yet abstained from introducing it in telemetric contexts – preferring to rely on automatic number plate recognition (ANPR). In some instances, this self-imposed restriction has been to avoid an anticipated level of public outcry (O'Malley 2013). Here, privacy issues that so far have been rather muted may come more to the forefront. Certainly, the forces promoting such advances in technology are considerable. Not least are concerns that drivers accruing demerit points against their licenses may – policed by ANPR alone – be able to persuade (or pay) others to state that they were driving the detected vehicle. Thereby, dangerous drivers are not removed from the roads. But privacy commissions (NSW 2005) already have noted that the records of telemetric policing are unlikely to be restricted to the issues of traffic control to which they appear dedicated, and are likely to be fed into or linked with many other data bases tracking citizens in everyday life.

The outcome of these kinds of development is difficult to predict. Certainly the suspension and closure of many safety camera schemes should caution against simple assumptions that policing and security concerns will always trump citizen concern and resistance. Even so, some of the more concerning developments may be arising in areas where resistance may be more muted because they touch far less on the lives of the ostensibly respectable "average" citizen. As mentioned, most obvious among these has been the development of electronic tracking devices for offenders released into the community and/or who are subject to restrictions on their movement as part of other sanctions and orders. Yet other developments create new possibilities and new concerns.

The rise of telemetric policing was closely linked with the development of streamlined forms of procedure culminating in the infringement or penalty notice. While well in excess of 90 % of these notices are associated with traffic issues, penalty notices have been extended to govern many kinds of civil and criminal transgression. Lansdell et al. (2012), for example, record 120 agencies with the power to issue notices with respect to 2,000 kinds of infringement in the Australian state of Victoria alone. They also note the widening use of such notices with respect to minor public order offenses, and elsewhere in Australia, it has been noted approvingly that such developments are extremely cost-effective for police (NSWLRC 2010: 145). Of course, the ongoing spread of penalty notices does not imply the necessary growth and extension of telemetric policing per se. However, the further development and implementation of, for example, face-recognition technologies in CCTV begins to open out new directions in telemetric policing in relation to public order offenses. While identification of rioters in recent UK civil disturbances has been mediated by live agents aided by facial recognition software, it is not a huge step for more minor offending to be handled in much the same telemetric way as traffic infractions. This should not be thought of as fanciful. The inducements to the wrongdoer offered by a penalty notice scheme have been seen with traffic issues to be a powerful factor in reducing challenges in court settings, and this assemblage already operates in other aspects of "street justice" (Slapper 2010). Resistance to telemetric policing may thus be overtaken by inertia, lack of resources, and representation. Likewise, while alleged inaccuracies in identification have been effective in some traffic instances, this has been uneven and the advance of telemetric traffic policing has generally extended over the past decade: the same could well be true for facial recognition in telemetry. Telemetric policing of moving traffic appeared fanciful 30 years ago, yet within 10 years, it had become normal in many jurisdictions. As "criminal" law and civil law become increasingly blurred around questions of low-level public disorder (Crawford 2009; Lansdell et al. 2012), and as the intake of increasing numbers of public disobedience infringements consequently expands – as happened with traffic – it would not be at all surprising to see innovations in this telemetric direction within the next few years.

Related Entries

▶ Electronic Monitoring
▶ History of Technology in Policing
▶ Surveillance Technology and Policing

Recommended Reading and References

Agamben G (1993) The coming community. Notes on politics. University of Minnesota Press, Minneapolis, p 86

Amoore L (2011) Data derivatives. On the emergence of a security risk calculus for our times. Theor Cult Soc 28:24–43

Cavenagh R (2008) Northern territory road toll soars as speed cameras take their toll. 20 Sept 2008. www.roadsense.com.au.latest.html

Crawford A (2009) Governing through anti-social behaviour. Regulatory challenges to criminal justice. Brit J Criminol 49:810–831

Deleuze G (1995) Negotiations 1972–1990. Columbia University Press, New York

Edmonton City Government (2011) http://www.edmonton.ca/city_government/news/intersection-safety-camera-spe.aspx

Ercke A, Goldenbleld C, Vaa T (2009) Speeding, drink-driving and seatbelt-wearing: results from a meta-analysis. European Commission, Brussels

Fogelson R (2001) Downtown, its rise and fall 1880–1950. Yale University Press, London

Fox R (1996) Criminal justice on the spot. Infringement penalties in Victoria. Australian Institute of Criminology, Canberra

Gains A (2005) The national safety camera programme: four-year evaluation report. PA Consulting Group, London

Goold B, Neyland D (eds) (2009) New directions in surveillance and privacy. Willan, Devon

Graham S, Wood D (2003) Digitizing surveillance. Categorization, space, inequality. Crit Soc Polic 23:227–248

Introna L, Gibbons A (2009) Networks and resistance: investigating online advocacy networks as a modality for resisting state surveillance. Surveill Soc 6:233–258

Jones R (2001) Digital rule. Punishment, control and technology. Punishm Soci 2:5–22

Khan S (2003) Saboteurs take out 700 speed cameras. The Observer, 7 Sept 2003

Lansdell G, Eriksson A, Saunders B, Brown M (2012) Infringement systems in Australia. A precarious blurring of civil and criminal sanctions. Altern Law J 37:41–45

Ministry of Transport (2008) Safety performance of traffic management. Transport Research Laboratories, London

Mountain L, Hirst W, Maher M (2005) Are speed cameras more effective than other speed management measures? The impact of speed management schemes on 30 mph roads. Accid Anal Prev 37:742–754

Norris C, Wilson D (eds) (2006) Surveillance, crime and social control. Ashgate, Aldershot

Norris C, Moran J, Armstrong G (1999) Algorithmic surveillance: the future of automated visual surveillance. In: Norris C (ed) Surveillance, closed circuit television and social control. Ashgate, Aldershot, pp 255–275

NSW Parliament (2008) Privacy: the current situation. Research report 2008. http://www.parliament.nsw.gov.au/prod/law/qala.nsf/0/CA25708400173F6CA25704

NSWLRC (New South Wales Law Reform Commission) (2010) Penalty notices. Department of Justice and Attorney General, Sydney

O'Malley P (2010a) Fines, risks and damages. Curr Iss Crim Just 21:265–382

O'Malley P (2010b) Simulated justice. Risk, money and telemetric policing. Brit J Criminol 50:795–807

O'Malley P (2013) Mass preventive justice. In: Ashworth A, Zedner L (eds) Preventive justice. Hart Publishing, Oxford

Plowden W (1971) The motor car and politics 1986–1970. The Bodley Head, London

RTA (Road Transport Authority of NSW) (2008) Mobile speed cameras at http://www.rta.nsw.gov.au/roadsafety/speedandspeedcameras/avespeedsafetycameras/index.html

Simon J (1998) Driving governmentality, automobile accidents. Insurance and the challenge to social order in the inter-war years 1919–1941. Conn Insur Law J 4:522–588

Slapper G (2010) Pay-as-you-go street justice. J Crim Law 74:1–3

Thomas L (2008) Safety effects of automated speed enforcement programs. Critical review of international literature. Transport Research Record 2078. Transportation Research Board, National Academy of Sciences, Washington, DC

Tozer J Town that scrapped 'motorist tax' speed cameras sees no increase in accidents. www.dailymail.co.uk Accessed 24 Apr 2010

Webster B (2011) Speed camera U-turn as 500 sites rejected. http://www.timesonline.co.uk. Accessed 5 July 2011

Wells H (2008) The techno-fix versus the fair cop: procedural (in)justice and automated speed limit enforcement. Brit J Criminol 48:798–817

Wells H, Wills D (2009) Individualism and identity. Resistance to speed cameras in the UK. Surveill Soc 6:259–274

Williams D (2005) Speed cameras may cause accidents. Evening standard, 24 June 2005, p 6

Wynne B (1996) May the sheep safely graze? A reflexive view of the expert-lay knowledge divide. In: Lash S, Wynne B (eds) Risk, environment and modernity. Sage, New York, pp 44–83

Termination

▶ Desistance from Crime

Terrorism

▶ Preventing of Terror at Shopping Malls

Terrorism Inside the USA

▶ Homegrown Terrorism in the United States

Terrorist Groups

▶ Terrorist Organizations

Terrorist Organizations

Erin Miller
University of Maryland/START Consortium,
College Park, MD, USA

Synonyms

Terrorist groups; Terrorist perpetrator groups; Terrorist perpetrator organizations

Overview

Unlike many types of violence, terrorist attacks are frequently conceived of, planned, and executed within the framework of a group or organization rather than by isolated individuals. As a result, a key element of understanding terrorism involves understanding the groups and organizations that engage in terrorism. This entry explores the fundamentals of terrorist organizations in a historical and global context, focusing on three general areas: definitions, data, and dynamics. First, what constitutes a terrorist organization and what implications do definitional complexities have for analysis and policy? Second, what is the current state of data and scholarly analysis on terrorist organizations? Finally, how do terrorist organizations vary on key dimensions such as formation, typologies, attack patterns, organizational structure, and evolution?

Defining Terrorist Organizations

Insofar as terrorist organizations are simply organizations that engage in terrorism, the practice of comprehensively defining terrorist organizations inherits all of the complexity that goes along with defining terrorism and then some. The elusive definition of terrorism is discussed at length elsewhere (Schmid and Jongman 1988; Schmid 2004), with the conclusion that there are many definitions used and though they share common themes, none is universally accepted. Beyond this challenging inconsistency, defining terrorist organizations is further complicated by the fact that organizations are typically multifaceted, engaging in a wide variety of legal and illegal activities. Classifying organizations as terrorist organizations requires consideration of how many acts of terrorism a group must commit before this becomes a key identifying characteristic. Further, how great of a threat must the organization pose, and to whom? Finally, groups of individuals that engage in terrorism range from very small, clandestine, informally related clusters of people, to broad networks united by “leaderless resistance,” to formally established, even hierarchical organizations that are more explicit about their existence and their objectives. This structural dimension also impacts the problematization of terrorist groups. One man’s terrorist organization may be another man’s insurgent group, or criminal gang, or even political party.

In light of these issues, identifying terrorist organizations is ultimately in the eye of the beholder. A number of international entities and nations formally designate terrorist organizations for the purpose of law enforcement and sanctioning. Many of these designations originated or expanded following the unprecedented terrorist attacks carried out by al Qa’ida in the United States on September 11, 2001. The most widely adopted list is that established by the United Nations Security Council (UNSC). The UNSC (resolutions 1267 (1999) and 1989 (2011)) maintains a list of “entities and other groups and undertakings associated with al Qa’ida,” which member nations are bound to sanction by (1) freezing the entity’s funds, assets, and resources; (2) preventing transit to or through their territory; and (3) preventing the supply, sale, or transfer of arms or related material. The UNSC’s list includes 68 entities, both terrorist organizations linked to al Qa’ida and organizations that provide financial and material support to al Qa’ida and related groups.

In addition to upholding the sanctions established by the United Nations, the European Union maintains sanctions against additional groups involved in terrorist activity identified through a process governed by the Common Foreign and Security Policy (CFSP). The Council Common Position 2001/931/CFSP indicates that the function of the EU's list of terrorist groups is twofold. It combats the financing of terrorist activity insofar as the European Community will freeze the assets and resources of listed groups, and it facilitates cooperation in criminal justice matters by stating that the EU and its member states will "afford each other the widest possible assistance" with respect to police and judicial enquiries regarding the entities on the list. Until recently the list included a distinction between groups that are subject to both provisions and groups that are subject to only the latter provision regarding cooperation among states. In fact, the initial list of terrorist groups identified by the EU in 2001 included only two whose funding and financial assets were targeted: Hamas-Izz al-Din al-Qassem (the terrorist wing of Hamas) and Palestinian Islamic Jihad (PIJ). The remaining groups, for which member states pledged assistance in criminal justice matters, included the Continuity Irish Republican Army (CIRA), Basque Fatherland and Liberty (ETA), the Anti-facist Resistance Group First of October (GRAPO), the Loyalist Volunteer Force (LVF), the Orange Volunteers (OV), the Real Irish Republican Army (RIRA), the Red Hand Defenders, (RHD), the Revolutionary Nuclei, the Revolutionary Organization 17 November, the Revolutionary Popular Struggle (ELA), and the Ulster Defense Association/Ulster Freedom Fighters (UDA/UFF).

The EU list is reviewed and updated regularly, however, and as of January 2011, it has changed considerably. It now includes 26 groups that are all fully covered by the provisions of Common Position 2001/931/SCFP. Furthermore, the organizations on the list are far more geographically diverse than the originally designated list, including groups not only from Europe and the Middle East but also Latin America (e.g., the revolutionary Armed Forces of Colombia (FARC), Shining Path), Asia (e.g., Aum Shinrikyo, the Liberation Tigers of Tamil Eelam (LTTE)), and Africa (e.g., Gama'a al-Islamiyya).

Other international organizations have not engaged in explicitly designating terrorist organizations to the extent the European Union has. The Association of Southeast Asian Nations (ASEAN) has released declarations affirming its commitment to combating terrorism in 2001 and 2003 (2012). They pledge to uphold UN anti-terrorism resolutions and, among other things, enhance the sharing of information on terrorist organizations and their movement and funding but stop short of formally designating particular organizations for investigation. In 2009, however, ASEAN established a "Comprehensive Plan of Action on Counter Terrorism" that includes the following provision:

> 10.2 Introduce a system for the designation/proscription of terrorists/terrorist groups, without prejudice to domestic law and in accordance with international standards, and share such information with other ASEAN Member States in order that they may take the appropriate action, including, inter alia, monitoring and deterring terrorist movement, freezing assets/property, and preventing recruitment. (2012: 72)

Likewise, the African Union has resolved to strengthen its capacity and that of its member states to address the problem of terrorism (2004). Despite the fact that the AU is actively involved in combating the al Qa'ida-linked militant group al Shabaab in Somalia, thus far it has only officially designated one terrorist organization, the Lord's Resistance Army (LRA) based in Uganda. This designation took place in 2011, according to a November 2011 *Reuters* report.

Individual states have unique protocols for designating terrorist organizations as well and typically focus on those that are specific threats to the interests of the country or its allies. For example, the United States Department of State (2012a) describes its practice of "listing" Foreign Terrorist Organizations (FTOs) as follows:

1. It must be a foreign organization.
2. The organization must engage in terrorist activity or terrorism or retain the capability and intent to engage in terrorist activity or terrorism.
3. The organization's terrorist activity or terrorism must threaten the security of US nationals or the national security (national defense, foreign relations, or the economic interests) of the United States.

There are several legal implications of a US Department of State FTO designation. It is illegal for a person subject to the jurisdiction of the United States to provide material support or resources to the FTO. Also, members or representatives of the group who are not US citizens are not allowed to travel to the United States. Finally, US financial institutions are required to freeze any funds known to belong to designated FTOs. The Department of State notes several informal sanctions that come with an FTO designation, namely, the potential deterrent impact on donations and financing for FTOs and the stigmatization of the organizations, signaling to the public and to other nations the United States' view of them as a threat to national security. As of September 12, 2012, the United States designates 51 organizations as FTOs.

A number of other countries follow similar protocols for identifying and sanctioning terrorist organizations, with similar implications of sanctioning. For example, the United Kingdom's Home Office maintains a list of proscribed terrorist groups that, as of July 2012, includes 45 "international terrorist organizations" as well as 14 proscribed Northern Irish groups (Home Office 2012). Likewise, the Australian government (2012) officially identifies 17 organizations that have been designated either by a court in the context of a prosecution for a terrorist organization offense or "listed" by the government under the Criminal Code Regulations. Under Canada's Anti-Terrorism Act, the Governor in Council has established a list of entities that have participated in or facilitated terrorist activity. This list, which is reviewed every 2 years, currently includes 44 entities. India's Ministry of Home Affairs maintains a list of 34 banned organizations declared as terrorist organizations, in addition to the al Qa'ida and Taliban-related organizations designated by the United Nations Security Council pursuant to resolutions 1267 and 1989.

Listing and delisting terrorist organizations can be contentious and influenced by many factors, often including political considerations. For example, the Mujahedin-e Khalq (MEK) is an anti-government Iranian group currently exiled in Iraq that has carried out numerous violent attacks since the 1970s. The MEK has recently been successful in its controversial campaign to be delisted by a number of states. A 2012 background report on the organization published by the Council on Foreign Relations indicates that the MEK engaged a number of advocates, including many paid high-profile officials, to lobby on its behalf claiming that it is no longer a terrorist threat. As a result, the MEK has been delisted by the European Union, the United Kingdom, and most recently the United States. The United States, which initially listed the MEK in 1997, cited the group's "public renunciation of violence, the absence of confirmed acts of terrorism… for more than a decade, and their cooperation in the peaceful closure of Camp Ashraf, their historic paramilitary base" in its decision to remove them from the FTO list in 2012 (Department of State 2012c).

Likewise, there is often great deal of debate about which groups should be added to sanctions lists. For example, in 2012 media sources including the *Wall Street Journal* and the *Jerusalem Post* reported that representatives from several governments, including the United States, Israel, the United Kingdom, and the Netherlands, have applied pressure to the European Union regarding the absence of the Lebanon-based group Hezbollah from the EU's list of designated terrorist organizations. The EU, citing a lack of consensus, diplomatic concerns, and lack of tangible evidence of Hezbollah's involvement in acts of terrorism, has long resisted this pressure. A more recently emerging group, Boko Haram, is reportedly responsible for the deaths of over 1,000 people in Nigeria since 2009, leading US authorities to consider designating it as an FTO. However, the *International Business Times* reports that despite ongoing violence against Nigerian targets, Nigerian officials are concerned that doing so would have unintended negative consequences for travel and trade between the United States and Nigeria.

Given the many special considerations that produce such disparate and incomplete lists of officially defined terrorist organizations, scholars interested in systematic analysis of terrorist organizations may seek a more comprehensive solution to establishing this problem space. One alternative is to define terrorist organizations empirically, based on an objective definition of terrorism that is exhaustively, rather than selectively applied. Analysts have increasingly turned to terrorism incident databases, which are discussed in more detail below, to establish a comprehensive record of terrorist organizations insofar as they are simply those organizations identified as perpetrators of terrorist attacks. This approach is not without limitations. For example, its all-inclusiveness provides a rather poor initial assessment of a group's potential threat outside of the basic question of how many attacks it has executed and the severity of those attacks. Activity related to the preparation, planning, funding, or material support for violent attacks is rarely accounted for in such sources. However, this post hoc approach certainly avoids the inherent biases that exist in governmental strategies to define terrorist organizations.

Data on Terrorist Organizations

Despite the typically clandestine nature of terrorist organizations, an impressive amount of information about them exists in unclassified literature, though the level of detail is not uniform for all organizations. Due to terrorists' reliance on the media to disseminate their message, primary media sources have proven to be fruitful sources of information on the activities of terrorist organizations. In addition, court documents pertaining to the prosecution of terrorism often include information about groups and organizations. In certain cases, researchers have been very successful at obtaining valuable information through archival documents or interviews with members or former members of terrorist organizations.

There are several general types of data or information on terrorist organizations. These include in-depth case studies, narrative profiles compiled into reference volumes or online portals, structured data on terrorist organizations, and structured data on terrorist activity. Each of these contributes a unique and important perspective on terrorist organizations, and together they provide a more complete understanding of the nature and dynamics of organizations that engage in terrorism.

Case studies offer in-depth analysis of the motivations, structure, and functions of terrorist organizations, covering the entire spectrum of their violent and nonviolent activity such as recruitment, planning, strategy, financing, activism, communication, and criminal activity. They typically take a broad view of the organization, placing it in a broader historical and political context. Case studies typically focus on a particular group or conflict, such as Wright's (2006) chronicle of the history of al Qa'ida, Wickham-Crowley's (1992) comparative study of guerrilla groups in Latin America, or Sayigh's (1997) exhaustive account of the Palestinian National Movement. Each of these covers approximately half a century of events in great detail, drawing on archival records, published documents, and personal accounts to illustrate the inner workings of terrorist organizations. Others compile case studies on contrasting organizations, typically with the aim of identifying common patterns of activity. Examples of this include Crenshaw's (1995) edited volume, *Terrorism in Context*, which includes chapters on terrorism and terrorist organizations in Europe, Russia, Italy, West Germany, Argentina, Peru, Ireland, India, Spain, Algeria, Israel and Palestine, and Iran. Ross (2006) seeks to illuminate the causes and consequences of political terrorism using case studies of three key organizations, al Fatah, the Revolutionary Armed Forces of Colombia (FARC), and the Provisional Irish Republican Army (PIRA), as well as a review of terrorism and various terrorist groups active in the United States.

These are just a few examples of the use of case studies to compare, contrast, and better understand the organizational nature of terrorist activity. There are certainly many other studies that synthesize a great deal of information about particular organizations, both large and small. The studies mentioned above tend to aim at providing exhaustive accounts of the organization's existence, but others may isolate a particular dimension of interest. For example, a compilation of case studies written by a team of researchers at the RAND Corporation (Jackson et al. 2005) analyzes five terrorist groups: Aum Shinrikyo, the Radical Environmentalist Movement, Hezbollah, Jemaah Islamiyah, and the Provisional Irish Republican Army, focusing specifically on exploring their aptitude for organizational learning and adaptation. Others use the case study method to investigate how terrorist organizations decline (Alterman 1999; Crenshaw 1991; Cronin 2009; Jones and Libicki 2009; Ross and Gurr 1989). While the richness of case studies has obvious benefits for understanding how terrorist organizations operate, they do sacrifice breadth for depth. Case studies generally focus on a small number of high-profile groups that are not necessarily representative of the global landscape of terrorist organizations. In fact, more inclusive sources of information on terrorist organizations tend to indicate that these high-profile groups, while responsible for a great deal of terrorist violence, are actually quite atypical.

Compilations of narrative profiles of terrorist organizations provide a far more comprehensive, though less detailed, source of information. Several of the official listings of terrorist organization discussed above include narrative descriptions of the groups in question that serve as justification for their designation as terrorist organizations (e.g., United Nations 2012; US Department of State 2012b). There are a number of reference texts that are far more inclusive, however, such as the "World Directory of Terrorist and Other Organizations Associated with Guerrilla Warfare, Political Violence, and Protest" compiled by Albert Jongman in collaboration with Alex Schmid (Schmid and Jongman 1988). This directory is based on information from a variety of journals, newspapers, and magazines and includes profiles of approximately 3,000 groups, parties, movements, and organizations worldwide that have engaged in violence, including state agencies that have engaged in gross human rights violations.

Similarly comprehensive reference resources on terrorist organizations include Janke's (1983) *Guerrilla and Terrorist Organizations: A World Directory and Bibliography*, Jessup's (1998) *An Encyclopedic Dictionary of Conflict and Conflict* Resolution 1945–1996, Hill's (2002) *Extremist Groups: An International Compilation of Terrorist Organizations, Violent Political Groups, and Issue-Oriented Militant Movements*, Kushner's (2003) *Encyclopedia of Terrorism*, and Atkins' (2004) *Encyclopedia of Modern Worldwide Extremists and Extremist Groups*.

More recently, several online portals including information on terrorist organizations have been developed. Perhaps the first and most well known of these, the Terrorism Knowledge Base (TKB), was developed and maintained by the Memorial Institute for the Prevention of Terrorism (MIPT). The TKB included narrative profiles and basic structured information on aliases, bases (countries) of operation, date formed, group size, ideological classification, sources of funding, key leaders, related groups, and official designations for over 800 organizations. Although the TKB was discontinued in 2008, these profiles were transferred to the National Consortium for the Study of Terrorism and Responses to Terrorism (START) where they are published online in their original form under the name *Terrorist Organization Profiles (TOPs)* (2008). Similar resources have since been developed, including the Violent Extremism Knowledge Base (VKB) at the Institute for the Study of Violent Groups (ISVG) and Crenshaw's Mapping Militant Organizations project at Stanford University (Crenshaw 2012; ISVG 2012). The VKB contains profiles of over 150 groups and organizations around the world. The Mapping Militant Organizations project, though currently limited to Global al Qa'ida, Iraq, Italy, North Caucasus, the Philippines, Somalia, and Pakistan, includes both general profiles and in-depth timelines documenting the complex evolution of dozens of organizations in these contexts.

Structured, quantitative datasets on terrorist organizations are extremely resource intensive to collect and therefore are quite rare. Due to the large number of terrorist organizations active worldwide, along with their clandestine and continually evolving nature, it is difficult to systematically collect reliable, comprehensive data on these groups. Certain group-level datasets do exist, though they are somewhat limited in scope either by time or place. For example, the Big Allied and Dangerous (BAAD) Lethality database collected by Asal, Rethemeyer, and Anderson contains data on organizational ideology, location, size, structure, funding, and relationships to other groups for those organizations that were active worldwide between 1998 and 2005 (2009). Likewise, the Profiles of Perpetrators of Terrorism in the United

States (PPT-US) database collected by Miller and Smarick (2011) contains over 100 variables regarding the organization's history, structure, and violent and nonviolent activity for 142 groups that have carried out or attempted to carry out terrorist attacks in the United States between 1970 and 2010.

The final piece of the data puzzle for understanding terrorist organizations is incident-level data. In order to fully understand the dynamics of terrorist organizations, one must consider their patterns of activity, including the location, timing, tactics, and lethality of their attacks. With the increased availability of incident-level datasets on terrorist attacks, analysts are able to aggregate data on the perpetrators of these attacks to first identify the universe of perpetrator groups and then conduct analysis on any of these dimensions of activity. Several such global datasets are currently available, including the RAND Database of Worldwide Terrorism Incidents (RDWTI) (RAND Corporation 2012) and the International Terrorism: Attributes of Terrorist Events (ITERATE) dataset (Mickolus 1982). The ITERATE dataset, distributed by Vinyard Software, now includes structured data on international terrorist attacks from 1968 through 2007, and the RDWTI includes data on international terrorist attacks from 1972 to 1997 and both international and domestic attacks from 1998 to 2008.

For a complete account of an organization's terrorist activity, it is important to consider both international and domestic attacks. The Global Terrorism Database (GTD) is maintained by the National Consortium for the Study of Terrorism and Responses to Terrorism (START 2012). The GTD, which includes over 104,000 domestic and international terrorist attacks that occurred worldwide from 1970 to 2011, identifies names of over 2,000 organizations as perpetrators of the attacks.

Dynamics of Terrorist Organizations

Ultimately, it is extremely challenging to establish a "ground truth" universe of terrorist organizations, given many divergent interests that influence official designations and various methodologies for collecting data on these groups. It is critical that researchers make appropriate inferences with this in mind. Nonetheless, drawing on the variety of sources described above, analysts have built a considerable body of research on the dynamics of terrorist organizations, how they form, what drives them, how they evolve, and how they ultimately end. The common theme in much of the literature is that terrorist organizations are extremely diverse with respect to many of these dimensions.

How Terrorist Organizations Begin

Martha Crenshaw (2001) contrasts two theoretical paradigms of terrorism, one *instrumental* and the other *organizational*, both of which have important implications regarding how terrorist organizations form and begin to engage in violence. An instrumental approach is entirely goal based, suggesting that everything an organization does, including choosing to engage in violence, is decided because it is in the best interests of achieving the group's stated goals. The organizational framework suggests that the stated ideology or goal of the group is actually secondary to its true objective: simply continuing to exist, even as political purpose evolves over time. In this scenario, insofar as the potential for grievance is universal, the formation of potentially violent organizations becomes critical to understanding and combating terrorism.

A number of factors influence the formation of terrorist organizations. Crenshaw suggests that the necessary components include skilled, committed leaders who are able to attract and incentivize other members, a demand for the organization among an actual or potential constituency, available mobilizable resources, and a salient purpose marked by ideas that legitimize violence. Likewise, Oots (1989) explores the importance of not only entrepreneurial leadership, recruitment of members, and political and financial support but also a group's ability to form coalitions with other groups, internal and external competition, and internal cohesiveness. The salience of these factors with respect to terrorist organizations suggests that in many ways they are not unlike other types of organizations, particularly other political organizations.

Research on the experiences of group members indicates that the incentives to join a terrorist group often have little to do with the group's objectives (Crenshaw 2001; Sageman 2004). Crenshaw highlights the importance of factors like personal daring, social rebellion, action, and feelings of belonging, acceptance, solidarity, and contribution. Furthermore, Sageman observes that the key to understanding modern Salafi jihad terror networks is not found in conventional wisdom about "top-down" recruiting efforts and brainwashing. Instead, it is frequently the case that adopting an alliance with al Qa'ida and forming a terrorist cell is about social relationships involving friendship, kinship, discipleship, and worship. That is, terrorist cells are formed among personal acquaintances within social circles as a "bottom-up" phenomenon.

Typology

Analysts frequently organize terrorist groups into typologies to provide a framework for understanding the groups and their activities. These typologies are most frequently based on the groups' ideology or goals (Ganor 2008; Merari 1978; Schmid and Jongman 1988). The motivational underpinnings of terrorist groups may be political, religious, social, economic, philosophical, practical, or a complex, shifting combination of many of these. Although ideology is frequently the basis for classification of terrorist groups, it is not the only basis and it has certain limitations. For example, in some cases, the group's ideology is held universally, and in other cases, it may be largely immaterial to the violent activity of members. Likewise, it can be challenging to articulate a meaningful classification scheme when the organizations themselves are often complex.

For example, the *dominant ideology* variable in the Profiles of Perpetrators of Terrorism (PPT) database is a fairly typical ideological classification of perpetrator groups based on the following five categories: Extreme Right-Wing, Extreme Left-Wing, Religious, Ethnonationalist/Separatist, and Single Issue. Variations on this scheme may disaggregate these categories further; however, even at this fairly high level of abstraction, certain groups fall into more than one category, compromising the mutual exclusivity of the typology. Often these complexities contribute to rifts within groups that lead to the development of factions or splinter groups. For example, the Moro National Liberation Front (MNLF) is a Muslim group based in the Philippines that sought for decades to establish independence for the Moro people. Because the Moro people are Muslim, the MNLF is frequently classified as Religious/Nationalist-Separatist, despite the fact that their ideology is generally secular and somewhat left leaning, guided principally by the goal of Moro nationalism (ISVG 2012; Tan 2007). In contrast, however, MNLF splinter groups such as the Moro Islamic Liberation Front (MILF) and Abu Sayyaf Group (ASG) have adopted a far more "overtly religious" radical Islamist ideology, establishing links with the global jihadist movement and al Qa'ida (Tan 2007, pp. 13, 34).

Fitting groups like this into a typology based on ideology can be challenging and, as Merari (1978) suggests, not particularly useful with respect to forecasting a group's strategy or tactics or the efficacy of counterterrorism measures. Instead, Merari recommends a two-dimensional typology that classifies terrorist groups based on (a) whether they fight against a foreign or domestic government, organization, or population ("xenofighters" vs. "homofighters") and (b) whether their base of operations is domestic or abroad relative to their constituency. Merari's typology of terrorist organizations focuses particularly on their operational goals and how they relate to group behavior. For example, he suggests that xenofighter terrorist groups tend to adopt more indiscriminate tactics and are more likely to harm innocents or use weapons of mass destruction because they do not rely on the target population for support like homofighter groups do. Likewise, in Merari's typology, homofighter groups are more likely to use targeted tactics such as assassinations or kidnappings of key figures.

Ganor (2008) articulates an exhaustive list of typological dimensions that extend beyond an organization's motivations. He suggests that violent groups vary meaningfully with respect to targets of violence (military vs. civilian), size, extent of public support, characteristics of the decision-making process, degree of independence or subordination with respect to sponsor states, financial strength, arena of operations (rural vs. urban), geographic range, control of territory, demands, seniority, place in organizational development, activities, and structure. Ganor recommends concentrating in particular on the two elements that are most likely to impact an organization's activity: degree of motivation and degree of operational capability. Although any number of these dimensions might individually or in combination provide insight regarding an organization's capabilities and threat, it is often difficult to comprehensively and systematically classify groups because reliable information on these characteristics is typically scarce and the characteristics themselves are often dynamic. Consequently, the application of typology-based strategies for understanding terrorist organizations is poorly developed.

Attack Patterns

Although it is difficult to obtain systematically collected data on many of the more clandestine attributes of terrorist organizations, information about their attack patterns is more readily available from incident-level databases like the Global Terrorism Database (GTD). The most widely recognized terrorist organizations are long lasting and either highly active, highly lethal, or both; however, a comprehensive analysis of the perpetrator groups identified in the GTD indicates that while some groups endure, the vast majority (74 %) last less than one year and are associated with only a few attacks (Dugan 2012). Not surprisingly, given this high proportion of short-lived groups, Dugan's analysis also indicates that 26 % of terrorist organizations identified in the GTD are responsible for over 93 % of all attacks for which a perpetrator was identified.

The attack patterns of terrorist organizations are quite diverse and multifaceted. The Liberation Tigers of Tamil Eelam (LTTE) in Sri Lanka and the Irish Republican Army (IRA) in Northern Ireland are both nationalist/separatist organizations among the most active and longest active groups between 1970 and 2007, yet the LTTE is nearly twelve times as lethal on average than the IRA. Conversely, al Qa'ida is responsible for the most lethal terrorist attacks in history on September 11, 2001, but is not among the most active terrorist organizations, though several al Qa'ida-linked groups are, such as al Qa'ida in Iraq (AQI).

The most recent version of the GTD indicates that between 1970 and 2011 there were 91 groups that carried out suicide attacks. Although the Taliban and the LTTE are the most frequent perpetrators of suicide attacks, this tactic represents only 9.8 % and 6.5 % of their attacks, respectively. Among groups that have carried out more than ten attacks, several adopt suicide tactics in more than one third of their attacks, including the Haqqani Network (71.8 % of attacks), al Qa'ida in Iraq (40 % of attacks), and al Qa'ida (36.8 % of attacks).

The GTD also indicates that perpetrator organizations attack a wide variety of targets. Although some of the typologies discussed above seek to classify terrorist organizations based on target selection, terrorist organizations vary considerably with respect to both targeting practices and diversity of target selection. For example, among perpetrator groups that have carried out at more than ten attacks between 1970 and 2011, only one organization attacked exclusively military targets, the 16 January Organization for the Liberation of Tripoli which attacked 24 Syrian military targets in Tripoli, Lebanon throughout 1989. A far more prevalent pattern of attacks involves a wide variety of target types including military, police, government, business, and private citizens and property. The GTD includes 21 different types of targets and over half of the perpetrator groups that have carried out more than ten attacks have attacked more than seven different types of targets.

How Terrorist Organizations Evolve

Just as the activity of terrorist organizations is highly variable, the organizations themselves are not static but rather dynamic and often subject to extensive evolution over time. Careful examination of the perpetrator attribution data in the Global

Terrorism Database reveals that terrorist organizations do not exist in a vacuum, but instead they frequently splinter, create factions, merge with other groups, or form coalitions or umbrella groups. They often change names, key personnel, base of operations, or even ideology. Terrorist organizations may form strategic alliances or partnerships with each other or rivalries that lead to further violence. The evolution of terrorist groups and the often scarce information about their organizational development have important implications for analysis. Naming conventions alone, and the extent to which they reflect meaningful shifts in group identity, can pose critical challenges for the analysis of terrorist organizations. Consider, for example, the following narrative about Devrimci Halk Kurtulus Partisi/Cephesi (DHKP/C), the Revolutionary People's Liberation Party/Front, from the Terrorist Organization Profiles (TOPs):

> The DHKP/C (Revolutionary People's Liberation Party/Front) is a Marxist, anti-Western splinter group of the Turkish terrorist group Devrimci Sol (Dev Sol). Dev Sol originated as a splinter group of Devrimci Yol (Dev Yol) which was itself a splinter group of the Turkish People's Liberation Party-Front (THKP/C). The THKP/C was an offshoot of the broader Revolutionary Youth movement (Dev Genc) within Turkey.
>
> In the early 1990s, infighting within Dev Sol resulted in the emergence of two factions. Dursun Karatas, who founded Dev Sol by combining splintered factions of Turkish radical leftist groups in 1978, changed the group's name to DHKP/C in 1994. Bedri Yagan, also a founding member of Dev Sol, broke from the Karatas faction and created a new faction, THKP/C (not to be confused with the original THKP/C). Confusingly, the Yagan faction of DHKP/C is still often referred to as Dev Sol.

The evolution of DHKP/C described here represents only a small segment of the broader Turkish Left and is not particularly atypical. Many organizations with different names in some way represent the same broader entity and often organizations with the same name are actually different entities. Furthermore, individual membership to a group is rarely a discrete status and various groups with similar interests share members. These qualities frequently make it difficult to determine where one terrorist organization ends and another begins.

The Mapping Militant Organizations project led by Martha Crenshaw at Stanford University is an extremely useful resource that chronicles the family trees of terrorist groups that are part of the global al Qa'ida network, as well as those that have been active in specific countries and regions. It includes a web-based application that graphically displays relationships between key organizations, including alliances, affiliations, mergers, rivalries, splits, and umbrella organizations. The "maps" illustrate clearly the complex reality of terrorist group dynamics and allow analysts to understand how interrelated these organizations are.

How Terrorist Organizations End

For many years research on terrorist organizations concentrated primarily on how they form and their attack patterns; however, recently scholars have devoted more attention to the question of how terrorist organizations decline and ultimately end. Many of these studies focus primarily on articulating and evaluating typologies of reasons for decline among terrorist organizations (Alterman 1999; Crenshaw 1991; Cronin 2009; Ross and Gurr 1989). In addition, several scholars have used statistical analysis of terrorist organizations and terrorist attacks to help understand patterns of decline (Blomberg et al. 2010; Jones and Libicki 2009; Miller 2012).

The proposed typologies include a range of factors, such as success or achievement of goals, perhaps coinciding with a transition to the legitimate political process; loss of popular support; and repression by the state. Various frameworks share in common the contention that organizational desistance from terrorism results from both internal influences (e.g., organizational breakdown) and external influences (e.g., military intervention) and likely results from several factors operating simultaneously. Scholars also observe that an effective strategy for countering one terrorist organization might have unintended negative consequences for countering another terrorist organization, depending on various factors. For example, killing or capturing a key leader of a group may lead to its decline or its resurgence, depending on the group's organizational structure.

Statistical analysis of how terrorist organizations decline and end has been growing increasingly sophisticated and accountable to the complexities of terrorist organizations. The first large-N quantitative analysis of how terrorist groups end classified 648 groups as either active or inactive and then described the reason for decline among those groups that had ended (Jones and Libicki 2009). Although groundbreaking, this strategy oversimplifies both the status of terrorist organizations and the factors that impact decline. Subsequent efforts have improved upon the analysis by using statistical tools like duration modeling which accounts for the entire length of an organization's life span (Blomberg et al. 2010) and group-based trajectory analysis which accounts for the organization's level of activity across its life span (Miller 2012).

General findings from this body of research suggest that the effectiveness of counterterrorism policies remains largely untested with respect to the goal of bringing about organizational decline. Reasons for decline appear to be varied and politicization and suppression are equally common. The cause and rate of decline may be influenced by key characteristics of the group, such as ideology, size, organizational structure, or attack patterns. It is important to note that these findings are not yet robust enough to fully account for the complex evolution of organizations within a broader movement as discussed above. Rather than signifying the death of a terrorist organization, cessation of activity under a particular group name may indicate organizational splintering, merging, or simply a change of name. Future research on the decline of terrorist organizations must take into account a realistic appreciation for the complex and vague boundaries between groups.

Conclusion

The fact that terrorist attacks are frequently planned and executed by groups and organizations presents important challenges to understanding both terrorism and counterterrorism. The task of identifying terrorist organizations is subject to numerous influences and interests, leading various sources of data to produce widely divergent accounts of the prevalence and activity of terrorist organizations. A number of useful data sources exist, however, including in-depth case studies on particular groups, narrative profiles or structured data on a wider range of groups, or incident-level data that can be aggregated to provide details of organizations' activity. Each of these has strengths and limitations but contributes an important perspective to the landscape of knowledge about terrorist organizations. From these sources of information, we know that terrorist organizations are extremely diverse with respect to many dimensions such as lethality, tactics, targeting practices, ideology, and organizational evolution. Existing research reveals complex relationships between terrorist organizations that analysts must take into account in order to fully understand patterns of terrorist activity.

Related Entries

- ▶ Co-offending
- ▶ Counterterrorism
- ▶ Counterterrorism and the Media
- ▶ Incarcerating Terrorists
- ▶ Strategies of Policing Terrorism

Recommended Reading and References

African Union (2004) Protocol to the OAU convention on the prevention and combating of terrorism. Retrieved from: http://www.au.int/en/content/protocol-oau-convention-prevention-and-combating-terrorism

Alterman JB (1999) How terrorism ends (Special report no. 48). United States Institute of Peace, Washington, DC

Asal VR, Rethemeyer K, Anderson I (2009) Big allied and dangerous (BAAD) database 1 - lethality data, 1998–2005. Retrieved from: http://hdl.handle.net/1902.1/16062UNF:5:2Z77QCNImKUu2OVS6hqccw == Project on Violent Conflict [Distributor] V3 [Version]

Association of Southeast Asian Nations (ASEAN) (2012) A compilation of ASEAN declarations, joint declarations, and statements on combating transnational crime and terrorism. Retrieved from: http://www.asean.org/resources/archives

Atkins SE (2004) Encyclopedia of modern worldwide extremists and extremist groups. Greenwood Press, Westport

Australian Government (2012) Australian national security: listing of terrorist organizations. Retrieved from: http://www.nationalsecurity.gov.au/agd/www/nationalsecurity.nsf/AllDocs/95FB057CA3DECF30CA256FAB001F7FBD

Blomberg SB, Engel RC, Sawyer R (2010) On the duration and sustainability of transnational terrorist organizations. J Conflict Resol 54:303–330

Crenshaw M (1991) How terrorism declines. Terrorism Polit Violence 3:69–87

Crenshaw M (1995) Terrorism in context. The Penn State University Press, University Park

Crenshaw M (2001) Theories of terrorism: instrumental and organizational approaches. In: Rapoport DC (ed) Inside terrorist organizations. Frank Cass, London, pp 13–31

Crenshaw M (2012) Mapping militant organizations. Stanford University. Retrieved from: http://www.stanford.edu/group/mappingmilitants

Cronin AK (2009) How terrorism ends: understanding the decline and demise of terrorist campaigns. Princeton University Press, Princeton

Dugan L (2012) The making of the Global Terrorism Database and its applicability to studying the lifecycles of terrorist organizations. In: Gadd D, Karstedt S, Messner S (eds) The Sage handbook of criminological research methods. Sage, Los Angeles, pp 175–195

European Union: Europa (2001) Council common position of 27 December 2001 on the application of specific measures to combat terrorism (2001/931/CFSP). Retrieved from: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32001E0931:EN:NOT

Ganor B (2008) Terrorist organization typologies and the probability of a boomerang effect. Stud Conflict Terrorism 31:269–283

Hill SD (2002) Extremist groups: an international compilation of terrorist organizations, violent political groups, and issue-oriented militant movements. Office for International Criminal Justice, Huntsville

Home Office (2012) Proscribed terror groups or organisations. Retrieved from: http://www.homeoffice.gov.uk/publications/counter-terrorism/proscribed-terror-groups/

Institute for the Study of Violent Groups (2012) Violent extremism knowledge base (VKB). Retrieved from: http://vkb.isvg.org/

Jackson BA, Baker JC, Chalk P, Cragin K, Parachini JV, Trujillo HR (2005) Aptitude for destruction, volume 2: Case studies of organizational learning in five terrorist groups. RAND Corporation, Santa Monica

Janke P (1983) Guerrilla and terrorist organizations: a world directory and bibliography. Macmillan, New York

Jessup JE (1998) An encyclopedic dictionary of conflict and conflict resolution. Greenwood Press, Westport

Jones SG, Libicki MC (2009) How terrorist groups end: lessons for countering al Qa'ida. RAND Corporation, Santa Monica

Kushner H W (2003) Encyclopedia of terrorism. Sage, Thousand Oaks

Merari A (1978) A classification of terrorist groups. Terrorism, 1: 331–346.

Mickolus EF (1982) International terrorism: attributes of terrorist events, 1968–1977 (ITERATE 2). Inter-University Consortium for Political and Social Research, Ann Arbor

Miller E (2012) Patterns of onset and decline among terrorist organizations. J Quant Criminol 28:77–101

Miller E, Smarick K (2011) Profiles of perpetrators of terrorism in the United States (PPT-US). Retrieved from: http://hdl.handle.net/1902.1/17702 UNF:5:rMan9QGKaQXo4bRl8luCzQ== National Consortium for the Study of Terrorism and Responses to Terrorism [Distributor] V3 [Version]

Ministry of Home Affairs, Government of India (2012) Banned organisations: list of organisations declared as terrorist organisations under the Unlawful Activities (Prevention) Act, 1967. Retrieved from: http://www.mha.nic.in/uniquepage.asp?Id_Pk=292

National Consortium for the Study of Terrorism and Responses to Terrorism (START) (2008) Terrorist Organization Profiles (TOPs). Retrieved from: http://www.start.umd.edu/start/data_collections/tops/

National Consortium for the Study of Terrorism and Responses to Terrorism (START) (2012) Global Terrorism Database (GTD). Retrieved from: http://start.umd.edu/gtd/

Oots KL (1989) Organizational perspectives on the formation and disintegration of terrorist groups. Terrorism 12:139–152

Public Safety Canada (2012) Listed entities. Retrieved from: http://www.publicsafety.gc.ca/prg/ns/le/index-eng.aspx

RAND Corporation (2012) RAND Database of Worldwide Terrorism Incidents (RDWTI). Retrieved from: http://www.rand.org/nsrd/projects/terrorism-incidents.html

Ross JI (2006) Political terrorism: an interdisciplinary approach. Peter Lang, New York

Ross JI, Gurr TR (1989) Why terrorism subsides: a comparative study of Canada and the United States. Comparative Polit 21:405–426

Sageman M (2004) Understanding terror networks. University of Pennsylvania, Philadelphia

Sayigh Y (1997) Armed struggle and the search for state. Oxford University Press, New York

Schmid A (2004) Terrorism: the definitional problem. Case Western Reserve J Int Law 36:375–419

Schmid AP, Jongman AJ (1988) Political terrorism: a new guide to actors, authors, concepts, databases, theories, and literature. Transaction Publishers, New Brunswick

Tan ATH (2007) Terrorism and insurgency in Southeast Asia. Edward Elgar, Cheltenham

United Nations Security Council (2012) Narrative summaries of reasons for listing: Entities and other groups and undertakings associated with Al-Qaida.

Retrieved from: http://www.un.org/sc/committees/1267/entities_other_groups_undertakings_associated_with_Al-Qaida.shtml

United States Department of State, Bureau of Counterterrorism (2012a) Foreign terrorist organizations. Retrieved from: http://www.state.gov/j/ct/rls/other/des/123085.htm

United States Department of State, Bureau of Counterterrorism (2012b) Country Reports on terrorism. Retrieved from: http://www.state.gov/j/ct/rls/crt/

United States Department of State, Office of the Spokesperson (2012c) Delisting of the Mujahedin-e Khalq. Retrieved from: http://www.state.gov/r/pa/prs/ps/2012/09/198443.htm

Wickham-Crowley T (1992) Guerrillas and revolution in Latin America: a comparative study of insurgents and regimes since 1956. Princeton University Press, Princeton

Wright L (2006) The looming tower: al Qa'ida and the road to 9/11. Vintage Books, New York

Terrorist Perpetrator Groups

▶ Terrorist Organizations

Terrorist Perpetrator Organizations

▶ Terrorist Organizations

THB

▶ Trafficking in Human Beings

Theft at Construction Sites

Rachel Boba Santos[1] and Roberto Santos[2]
[1]School of Criminology and Criminal Justice, Florida Atlantic University, Boca Raton, FL, USA
[2]Professional Standards Division, Port St. Lucie, Police Department, Port St. Lucie, FL, USA

Synonyms

Burglary at construction sites; Construction site theft and burglary

Overview

Theft at construction sites is recognized as a concern by the United States as well as many other countries around the world. Estimates from the United States indicate that between one and four billion dollars worth of materials, tools, as well as large and small equipment are stolen every year. This type of crime occurs at a number of different places including individual house construction sites, subdivisions of homes, larger multifamily sites, and commercial sites; however, there are a number of factors that contribute to opportunities for these crimes that result in their concentration at construction sites and the areas around them. This entry reviews the theoretical foundation and current research literature that seeks to explain why crime occurs at construction sites, the factors that contribute to these crimes, and ends with a discussion of the next steps in understanding this specialized crime type.

Introduction

Some crimes have the potential to occur throughout much of the environment. For others, opportunities are less ubiquitous. Construction site theft is recognized as a concern by the United States as well as many other places around the world such as Canada, Australia, Europe, and Japan (Berg and Hinze 2005; Lambertson 2005). Estimates from the United States indicate that between one and four billion dollars worth of materials, tools, as well as large and small equipment are stolen every year (Berg and Hinze 2005; Lambertson 2005). The wide range of these estimates is attributed to the lack of reporting to both police and insurance by builders and contractors (Lambertson 2005).

In addition to property losses, there are indirect costs that are often not quantified but are important as they show the overall impact of construction site theft. They include, among others, job delays, downtime for operators, higher insurance premiums, and possible cancellation of insurance (Berg and Hinze 2005). The impact of these losses is often not borne by the construction companies themselves, but the costs are passed on to the home buyer with an average of a 1–2 % increase in the cost of a new home (Patton and Oleck 2005).

This type of crime is specific to the nature of the place where it occurs (i.e., the construction site) and may not be present in a community on a continual basis (i.e., may be seasonal or appear every few years when there is a construction boom). However, because it can be a significant problem and impacts nearly every community at some point, it is important to understand the nature of the crime and particularly how it concentrates by place. Thus, this entry reviews the crime and place theoretical foundation and the current research and practice of construction site theft to identify the specific factors that facilitate this specialized crime.

State of the Art: Crime Concentration at Construction Sites

There are a number of reasons why construction sites are vulnerable to theft. Opportunity theories offer a lens through which to consider how the ecology of an environment affects the likelihood of crime occurrence. They shift the attention from seeking to understand the individual criminal motivation of offenders to understanding how the convergence in time and space of an already motivated offender with a suitable target absence of a capable guardian (Cohen and Felson 1979) might affect the probability of a crime taking place.

In terms of construction site theft, *guardians* are people who have the ability to protect movable property of buildings by watching over them, such as construction workers, residents living nearby, and business or other people in the area. *Handlers* are people who know potential offenders and are in positions that allow them to monitor and/or control potential offenders' actions. Depending on who offenders of construction crime are, the people might be construction site supervisors (for offenders who work on site) to parents (of juvenile offenders living near the construction site), among others. Place *managers* are those who are responsible for the construction sites, such as construction site supervisors or the owners of construction companies. These people set rules and employ practices that can encourage or discourage crime, so that the rules, how they are enforced, and the physical environment of the place all affect the crime opportunities. An important aspect of the effect of guardians, handlers, and managers is that they have the ability to *both* discourage or reduce opportunities as well as encourage or increase the opportunities for crime.

Consequently, the nature of the construction site environment is an important factor in understanding patterns of construction theft because often it is something about the place, not the offenders or the victims that best explain crime (Eck et al. 2007). However, even though some practical research has found that there is not direct repeat victimization at construction site locations (Boba and Santos 2007), we know more generally from research that crime is not randomly distributed across the geographic landscape, but rather, it is concentrated in space and also by type of place (Sherman et al. 1989).

The following sections break down specific factors that may contribute to theft at construction sites and how they are generated in terms of the nature of the construction site environment, the influence of place management practices of construction companies, the timing of crimes committed at these types of places, the nature of construction site theft offenders, as well as repeat victimization at and goods stolen taken from construction sites. The factors are derived from theoretical components discussed above as well as from research results which, unfortunately, are somewhat limited. However, the factors indicate that opportunity and the nature of the place plays a significant role in construction site theft.

Nature of the Place

A characteristic of construction sites that make them an environment different than most places where crime occurs is that they are in transition. Clarke and Eck (2005) define open or transitional environments as "areas without consistent or regular designated uses" (Step 15). They note, for example, that these types of places differ from parks that are used for a consistent purpose (i.e., recreation) by many different people even if they have not been specifically designated for such purpose. More specifically, construction sites exist only temporarily as environments where construction occurs and subsequently become residences, businesses, or other types of environments within a reasonably short amount of time. This is a key factor when considering the explanation for crime at these places.

Because of their transitional nature, construction sites are places of interest and curiosity. Those living or working near construction sites may want to examine what type of residence or business is being built near where they spend time. Others simply pass by occasionally or regularly look to see what is being built, and may even stop and walk through a site. Many of those interested likely visit the sites multiple times to see the "work in progress" and what new changes have been made since their last visit. Eventually, this curiosity may become tempting for some which may after repeated visits provoke those individuals to trespass and take property that has been left unprotected.

For example, an individual living near a residential construction site is curious about the site and begins by simply walking through the site when no one is around to see what type of house is being built, the layout, design, etc. During this first visit, the individual may notice several unused boxes of ceramic floor tile on the site. As the construction continues, the same individual visits again to check on its progress and changes to the home, and notices the same floor tile has not been used, but the floors are nearly complete. In a final visit after several days of thinking about it, the individual goes to the site one last time in the evening when no one is around with a vehicle and loads several, not all, of the unused boxes of tile and takes them home. In this case, the curiosity leads to the identification of an opportunity which in the end was too tempting for the individual to pass up, especially since the opportunity was nearby, the site was semi-private, there was minimal guardianship and no clear ownership, and the property itself was easy to take.

Another characteristic of the construction site environment is that they are places with a large number of transient employees; that is, people who conduct work for a short period of time on the site and then move on. If security is lacking and management indifferent, the temptation for these employees to take items that are improperly secured or unaccountable may be too much to resist (Fennelly 1996). For example, a construction worker who freelances for different companies and is in need of building supplies for a new job and/or newer tools and equipment may decided to take the property from a site he is working on for a day or two.

For all places, the physical environment impacts the amount of opportunities for crime (Felson and Clarke 1998). Studies of residential burglars find that burglars consider occupancy cues (e.g., presence of cars, visible residents, and noise/voices), surveillability cues (e.g., can neighbors see them), and accessibility cues (e.g., how well is the site protected with doors, fences, and locks) (Cromwell et al. 2002; White 1990). Nearly all the cues that would prevent an offender from committing a crime are lacking at construction sites (e.g., lack of neighbors, inability to lock buildings).

Additionally, research has shown that crime is more concentrated in areas with physical disorder, such as abandoned buildings, broken streetlights, junk-filled and unmowed vacant lots, litter, graffiti, and the visible consequences of vandalism (Skogan 1990) because it indicates to potential offenders that residents have little attachment to their neighborhood, symbolizing to offenders a lack of capable place guardianship. Construction sites are locations that, by the mere nature of their existence, are in physical disarray. With so many different types of work being conducted on a site (e.g., building construction, electrical wiring, plumbing, finishing) by different contractors for different periods of time, construction materials (e.g., lumber, dry wall, pipes) are often lying around because they are delivered and sit on the site for a time before installation. Even after work has been done, there can be a significant amount of scraps and other materials that appear as trash but could be, in fact, valuable.

Place Management

Another key characteristic of construction site environments is the fact that various contractors, sub-contractors, and even sub-sub-contractors often come and go from the site and have minimal concern for the appearance and well being of the site. As Newman (1972) asserts, defensible space is a place that allows inhabitants or users of the space to become key agents in ensuring its security. On a construction site, there is not always a clear feeling of ownership; that is, the construction company is not the true owner of the land or what is being built, but is temporarily overseeing it during the construction process. Therefore, it may also not be clear who is responsible for protecting the site and its materials from theft. For example, there are cases in which the builders have required homeowners to purchase insurance to cover any theft or damage that occurs during the home construction period, before they even own the home (Boba and Santos 2007).

Because construction sites are transitional, the different stages of construction, such as clearing the lot, laying a foundation, installing the roof, installing doors and windows making the building lockable, create different stages of opportunity for theft. Different materials are necessary and vulnerable at different times. In addition, the speed of construction can affect the amount of time the construction site is at risk. That is, the less time the construction takes, the less opportunity for theft since the property is vulnerable for shorter periods of time.

Additionally, construction sites are not yet occupied by residents or employees, so they lack the guardianship an inhabited home or office building would provide. Some sites may be isolated from view by being set back from the road, on large lots, or next to nonresidential land (e.g., parks, waterways, wooded areas), which reduces the chance that neighbors or passersby will see or hear a thief. Finally, construction sites may not have fencing or other mechanisms to protect the location from trespassers. Property left on the site is often left unprotected lying on the ground, in open garages, or in a partially constructed, unsecurable buildings. Some property, such as air conditioning units, can be vulnerable even when it is installed because it is located outside the building, virtually unprotected.

Productivity is a key component of the construction business (Lambertson 2005), and the management practices that maximize productivity may also maximize opportunities for crime. For example, delivering appliances or other materials several days or weeks before installation ensures that there is no delay once installers are there; however, this practice increases the time and opportunity for theft, especially in a building or a site that is not secured. Another example is a lax policy of tracking tools. Builders and contractors save time by not checking them in and out daily, but lack of control and oversight may give employees the perception the builders do not care if they are taken and decrease their perceptions of being caught (Berg and Hinze 2005). Finally, many builders see theft as a cost of business (i.e., trust luck and swallow any losses) (Clarke and Goldstein 2002) and cover their losses by simply increasing the price of a home or incorporating it into the price of a construction bid. Thus, there is little incentive for improving their practices.

Timing of Construction Site Theft

Unlike residential and commercial burglaries, there is little research that indicates when construction site theft predominantly occurs. Opportunities for theft at construction sites indicate that thefts could occur any time of day because of the lack guardianship and the large number of workers that frequent a site. In addition, because there are typically no witnesses to construction site theft, the exact time of occurrence is difficult to determine. However, one study found from interviewing several CEO’s and managers of construction companies that they believed the majority of thefts occurred during the day in the late afternoon hours after the site manager had visited the construction site the last time for the day (Boba and Santos 2007).

The stage of building when a construction site theft occurs may be more important than the time of day it occurs. For example, logic tells us that the most vulnerable time for theft at a single-family home construction site may be between when the house is securable (i.e., has windows and doors that are lockable) and when the home is handed over to the resident since a large number of desirable types of property (e.g., washers, dryers, refrigerators, heating and cooling units, etc.) are in the home. Construction also can vary by area of the country, by national and local economics, and by weather which may cause crime at these locations to also vary.

Nature of Construction Site Theft Offenders

Research on burglars in general indicates that they engage in crime because of the desire for money, the ease of the crime, and influence of others (Piquero and Rengert 1999). They do not tend to think about the consequences of their actions, or they believe there is little to no chance of getting caught. They often know their victims, who may include casual acquaintances, neighborhood residents, people for whom they have provided a service, or friends or relatives of close friends. Although very little research has been done on construction site thieves, researchers group them into three categories in order to understand the motivations and opportunities they exploit (Boba and Santos 2007, 2008; Clarke and Goldstein 2002).

Amateur opportunists are offenders who live near or travel near construction sites, see movable property on the site, and take it based on an immediate assessment of rewards and risks. They do not necessarily plan their crimes and tend to seize the opportunity at hand. However, they may also take property after seeing it unprotected for a long period of time since they may drive by or live near the construction site. These offenders might be more likely to take smaller, generic materials and tools at times when workers are not present. The example previously mentioned of the neighbor taking the box of tile is an example of this type of offender.

Professional thieves are offenders who make a living at theft and selling high dollar stolen goods. They spend time planning their crimes and tend to have intricate knowledge of the environment and areas where the crimes are committed because the reward is very high. For example, Boise law enforcement agencies arrested two men and recovered $750,000 worth of material and equipment stolen over 7 months, including tractors, trailers, generators, and power units (Anderson 1999). In another case, a Detroit man had rented lockers to store a large amount of stolen property and had nearly 200 thieves working for him, even filling orders (Bouffard 2004). Unfortunately, there is no research on the nature of professional construction thieves to know how they differ from professional thieves of other crimes or why and how they commit crimes in this environment.

Insiders are a third type of offender who are or work for builders and contractors or rival companies. They are called “insiders” either because they have knowledge of a specific builder’s construction practices and access to keys, tools, and materials, or they have more general knowledge of construction practices, such as how to disassemble an air conditioning unit and handler and what stage of building appliances or other desirable items are typically delivered and installed. Research shows that generally, a high percentage of employee thefts begin with opportunities that are regularly presented to them, which is similar to the example given about the amateur opportunist (Fennelly 1996). The inside knowledge and the benefits of acquiring materials and tools for free make these crimes very rewarding for these types of offenders. The type of offender that falls into this category of “insiders” appears in some research to be the most prevalent of the three. In Boba and Santos (2007), interviews with builders reveal that they believe that employee or “insider” theft was a significant problem. The same study found that the nature of the crimes, the type of property taken also indicated the prevalence of insider offenders.

Repeat Victimization and Goods Stolen

Repeat victimization is an important consideration because it indicates systematic patterns of opportunity that exist for a particular type of crime and identifying when repeat victimization is occurring assists in prioritizing crime prevention efforts (Farrell and Pease 1993). For example, one of the biggest predictors of residential burglary is whether the home has been burglarized in the past (Farrell and Pease 1993).

However, repeat victimization for construction site theft does not seem to be as obvious as it is for other types of crime because construction sites are transitional places. First, victimization may be difficult to determine since the place starts as a construction site but eventually becomes a residence, business, or other type of place which means the systematic opportunities that are the basis for identifying repeat victimization also change. Furthermore, the change of the place from one type of environment to another makes the crimes themselves hard to aggregate and count since they will be coded as different crimes (e.g., construction site theft, residential burglary, commercial burglary).

In examining repeat victimization of construction site theft, researchers have found it is more fruitful to focus on aggregating crime by individual builders or type of property rather than by individual place (Boba and Santos 2007). For example, a comprehensive study in Port St. Lucie, FL found that only 12 of 254 individual sites were burglarized twice within 1 year but that 20 % of the builders were named victims for nearly 70 % of the burglaries (Boba and Santos 2007). This implies that repeat victimization is explained by the general practices and policies of a particular builder, such as systematic delivery lag times, use of fencing, and storage of materials, rather than the specific opportunities present at a specific construction site. Repeat victimization might also be explained considering virtual repeats (Pease 1998), which are places that are virtually identical, although not necessarily close, to one another, because they share some of the same physical characteristics as well as place management practices. Examples are electronics retail stores with similar store layouts and materials used or residential homes that are designed and built by the same builder.

In terms of goods stolen, property taken in construction site theft is rarely recovered (Berg and Hinze 2005). Property types that are stolen include building materials such as drywall, plywood, and studs; appliances, such as refrigerators, ovens, air conditioners; construction equipment and tools, such as drills, saws, generators, cement mixers; pool equipment; internal cosmetic materials such as carpet, toilets, sinks, ceramic tile, and other flooring; as well as other property (Anderson 1999; Berg and Hinze 2005; Boba and Santos 2007; Clarke and Goldstein 2002). Yet, the type of goods that are taken may indicate the motivations of the offender. Amateur opportunists may take generic building materials they need to repair or use in their own homes, such as drywall, lumber, or ladders. Professional thieves may take property they know they can sell in an unregulated second-hand market (e.g., the Internet), such as appliances ceramic tiles, faucets, toilets, doors, and windows (Bradley 2005). Insiders may be more likely to take materials, tools, and small equipment they have access to on a regular basis or items that take some skill or effort to remove.

Related to the discussion of “insiders,” the high cost of construction materials induces some people to steal materials from construction sites in order to reduce their own building costs. In recent years, there has been a dramatic increase in the price of construction materials. For example, from January 2003 to May 2004, the price index of 11 key construction materials more than doubled (Bradley 2005; Bouffard 2004; Scarcella 2005) and the price of precious metals, such as copper, fluctuates as well. Increases such as these create incentive for contractors as well as individuals to take materials from construction sites to save money themselves or trade the materials for cash (Sidebottom et al. 2011).

Controversies and Key Issues

There are a number of considerations that make it difficult to conclude generally about the nature of thefts at construction sites. First, patterns of victimization are difficult to determine because of data coding. Because construction sites are transitional and become commercial, residential or other types of places, crimes that occur at different stages of development are coded differently, thus are difficult to aggregate. Even something as simple as searching all crimes at a particular address may be difficult for construction sites as many sites may not have addresses when the crimes occur or their addresses may change once construction is complete. Second, in general, thieves are rarely caught (FBI 2011) and even though rates for construction theft are not recorded nationally, those taking from construction sites seem to be caught even less often (Boba and Santos 2007) making it difficult to understand who the offenders are and why they are committing these crimes. Third, the business of construction varies significantly by area of the country, by national and local economics, and by weather; therefore, construction site theft might also vary by these factors.

In addition to these concerns, there are a number of questions that still need to be answered in more detail to further the understanding of theft at construction sites. They include, but are not limited to:

- How do levels of crime vary by type of construction location (e.g., single family residential, multi-family residential, commercial, retail, public roadways)?
- How do levels of crime compare to other types of public sites (e.g. parks, universities or hospitals)?
- Are the "professional" thieves specialized and focus only on construction sites or are they "professional" offenders for other types of crime as well?
- How do current crime recording practices affect the evidence that we have about this crime and how can they be improved?
- Are construction sites that cluster spatially and are of a specific type (e.g., bars, homes, industrial) at a higher risk for crime?
- What is the rate of "virtual" repeat victimization of theft at construction sites?

Conclusion

Bringing together the previous discussion of the nature of construction site theft and the opportunities for this crime to occur, several key conclusions can be deduced from the theory and research results. The first is that the mere nature of a construction site and the use of the area and the area surrounding the construction site is typically in disarray, is in transition, and appears to have no owner. All of these characteristics are relatively unique to crime place and are significant contributors to increased opportunities for any type of crime at construction sites. In addition, construction sites themselves are typically concentrated spatially, especially when they are managed by the same construction company, thus it follows that the crime would also be spatially concentrated.

Secondly, place management practices are likely the most important factors in understanding and subsequently preventing theft at construction sites. Theory indicates and research shows that practices for securing property and equipment, managing employees and subcontractors, and providing guardianship for sites (e.g., security) have a significant impact on which sites are more vulnerable than others. Typically, these place management practices vary by builder/construction company rather than by individual construction site, which is supported by research that shows that repeat victimization tends to occur, not by individual address or area, but by individual builder.

Finally, there are several types of offenders that find construction sites appealing based on the fact that the construction sites are places of interest and employment (e.g., amateur opportunists and insiders) or are targeted because of the availability of high dollar items (e.g., professional thieves). The most prevalent type of offender seems to be the insider because they have the most knowledge of management practices and activity space, the most direct motivation for committing these crimes, and, therefore, have the most opportunities to commit the crimes.

Related Entries

▶ Broken Windows Thesis
▶ Crime Prevention Through Environmental Design
▶ Designing Products Against Crime
▶ Fencing/Receiving Stolen Goods
▶ History of Criminological Theories: Causes of Crime
▶ Rational Choice, Deterrence, and Crime: Sociological Contributions
▶ Repeat Victimization

- ▶ Residential Burglary
- ▶ Routine Activities Approach
- ▶ Situational Crime Prevention

Recommended Reading and References

Anderson J (1999) Battling construction theft: contractors lose millions in work site rip-offs: but Clark county detectives and an industry group are pursuing thieves. *The Columbian*. (Vancouver, Washington), 27 June, E1

Berg R, Hinze J (2005) Theft and vandalism on construction sites. J Constr Eng Manag 7(131):826–833

Boba R, Santos RG (2006) Burglary at single family house construction sites, Problem-oriented policing guide series. Office of Community Oriented Policing Services, Washington, DC

Boba R, Santos RG (2007) Single-family home construction site theft: a crime prevention case study. Int J Constr Educ Res 3(3):217–236

Boba R, Santos RG (2008) A review of the research, practice, and evaluation of construction site theft occurrence and prevention: directions for future research. Secur J 21(4):246–263

Bouffard K (2004) Construction site thieves profit from building boom: burglars grab all tools and supplies not nailed down. *The Detroit News*, 7 May

Bradley D (2005) Construction-site thefts escalate. *The Seattle Times*, 13 March

Clarke RV, Eck J (2005) Crime analysis for problem solvers: in 60 small steps. Office of Community Oriented Policing Services, Washington, DC

Clarke RV, Goldstein H (2002) Reducing theft at construction sites: lessons from a problem-oriented project. In: Tilley N (ed) Analysis for crime prevention. Criminal Justice, Monsey

Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44:588–608

Cromwell P, Olson J, Avary D (2002) Decision strategies of residential burglars. In: Cromwell P (ed) Their own words: Criminals on crime: an anthology. Roxbury, Los Angeles

Eck J, Clarke RV, Guerette RT (2007) Risky facilities: Crime concentration in homogeneous sets of establishments and facilities. In: Farrell G, Bowers KJ, Johnson SD, Townsley M (eds) Imagination for crime prevention: essays in honour of Ken Pease. Willan, Portland, pp 225–264

Farrell G, Pease K (1993) Once bitten, twice bitten: repeat victimisation and its implications for crime prevention. *Crime prevention unit series*, Paper No. 46. Home Office Police Research Group, London

Federal Bureau of Investigation (FBI) (2011) Crime in the United States 2009. U.S. Department of Justice, Washington, DC

Felson M, Clarke RV (1998) Opportunity makes the thief. *Police research series*, Paper No. 98. British Home Office Research Publications, London

Fennelly L (ed) (1996) Handbook of loss prevention and crime prevention, 3rd edn. Butterworth-Heinemann, Boston

Lambertson G (2005) Law enforcement focuses on construction site theft. Retrieved 29 June 2011 from www.cegltd.com/story.asp?story=5408 & headline = Law

Newman O (1972) Defensible space: Crime prevention through urban design. Macmillan, New York

Patton N, Oleck C (2005) Even the kitchen sinks are swiped: metro construction site thefts on the rise. *Detroit Free Press*. 14 March

Pease K (1998) Repeat victimization: taking stock. *Crime detection and prevention series*, Paper No. 90. Home Office, London

Piquero A, Rengert G (1999) Studying deterrence with active residential burglars. Justice Q 16:451–471

Scarcella M (2005) Work site bandits cash in on growth: in Manatee, authorities use helicopters and infrared cameras to try to stop the theft of building materials. *Sarasota Herald-Tribune*, 7 April, A1

Sherman L, Gartin PR, Buerger ME (1989) Hot spots of predatory crime: routine activities and the criminology of place. Criminology 27(1):27–56

Sidebottom A, Belur J, Bowers K, Tompson L, Johnson S (2011) Theft in price-volatile markets: on the relationship between copper price and copper theft. J Res Crime Delinq 48(3):396–418

Skogan WG (1990) Disorder and decline: crime and the spiral of decay in American neighborhoods. Free Press, New York

White G (1990) Neighborhood permeability and burglary rates. Justice Q 7(1):58–67

Theoretical Integration

- ▶ Integrating Rational Choice and Other Theories

Theories for Situational and Environmental Crime Prevention

Richard Wortley and Nick Tilley
Jill Dando Institute of Security and
Crime Science, University College London,
London, UK

Overview

The situational and environmental perspective in criminology refers to a collection of theories and approaches that focus on the role that immediate circumstances play in producing criminal acts. This contrasts with criminology's more usual preoccupation with what creates offenders' criminality: depending on the disciplinary orientation, the disposition to commit crime is accounted for by genetic makeup, developmental experiences, or current social, cultural, and economic conditions. Correspondingly, the usual solution to any crime problem is deemed to require altering these long-term and systemic causes of criminality, for example, by investing in early intervention with at-risk families, increasing educational opportunities, removing social and economic disadvantage, or providing treatments for known offenders.

Situational and environmental perspectives, on the other hand, focus instead on the dynamics surrounding crime events. Attention shifts from distal causes of criminality to proximal factors that enable and encourage specific crimes. Individuals highly disposed to commit crime do not do so all of the time. Moreover, many of those lacking any particular criminal disposition will occasionally commit crimes and account for a high proportion of all crimes. Even those individuals determined to commit crime can only do so when there is opportunity, while some situations can induce ordinarily law-abiding individuals to commit crimes that would not otherwise have been committed. Whether crimes occur in any given location or at any given time depends upon the conduciveness of the conditions there. The prevention of crime, therefore, is conceptualized in terms of altering the criminogenic aspects of the immediate situations most liable to produce crime.

This entry presents the key theories underpinning situational and environmental crime prevention. These are of two broad sorts. One set addresses the situational basis of criminal behavior. The other examines the patterning of crime in time and space according to the distribution of criminogenic environments.

Situational Models of the Offender

While situational and environmental perspectives are concerned with crime events rather than with criminality, in order to understand the role played by the immediate environment in crime, we must begin with a coherent model of the offender that explains how his or her behavior is influenced by momentary contexts. That all behavior occurs as a function of a person-situation interaction has become a commonplace principle of much social science (Mischel 1968). Behavior can be highly variable from one situation to the next; how an individual acts is shaped by where he or she is and with whom he or she is interacting. The mechanics of the person-situation interaction have been explained by numerous theories. As for the application of the person-situation interaction to crime, two main approaches have been taken. The first approach is to view offenders as rational, utility-maximizing actors whose choices are informed by available criminal opportunities and how they perceive them; the second approach is to view the offender as a respondent to situational pressures and provocations, which sometimes positively encourage crime.

Rational Choice

In 1967 Ron Clarke published a research paper on absconding from a residential school for juvenile delinquents (Clarke 1967). He noted that the best predictors of absconding were aspects of the environment – hours of daylight, features of the school's regime, and the distance home – and not any individual factors associated with the absconders. These and similar findings ultimately led Clarke to conclude that opportunity was a central and neglected cause of crime (Mayhew et al. 1976). In order to explain how opportunity leads to crime, Clarke and Cornish (1985) reviewed a wide range of research from sociology, criminology, economics, and cognitive psychology examining the variability of criminal behavior across situations. From the array of perspectives reviewed, they settled on a rational choice model as their preferred way to account for the mechanisms through which opportunity leads to deviant behavior.

The basic premise of the rational choice perspective on crime is that individuals make cost-benefit assessments of their situations in deciding whether or not it is in their interests to offend. Criminogenic environments are those that offer ample opportunities for prospective offenders: where they can expect the benefits from committing a crime to outweigh the costs of doing so. Situational crime prevention involves altering such criminogenic environments in ways that will lead prospective offenders to conclude that the expected costs of offending outweigh expected benefits. Note that while the label "rational choice theory" is frequently used in the literature, Cornish and Clarke have generally eschewed the term "theory," preferring instead "model" or "perspective" (see Cornish and Clarke 2008). They present rational choice as a heuristic that provides a framework for practitioners seeking to implement situational crime prevention. For this purpose it's a "good enough theory."

Underpinning the rational choice perspective is a view of offenders, like all others, as goal-driven, rational agents (Cornish and Clarke 2008). People commit crime for the purpose of deriving some benefit. These benefits are not restricted to the obvious economic ones that accompany property crimes. Offenders can also commit crimes because doing so is exciting, because it gives them a sense of power, because their friends will approve, because they receive sexual gratification, and so on. The costs and benefits attached to crime vary from one crime to another and also vary within crime types. The reasons for committing rape are very different from the reasons for committing car theft. Likewise, the goals of a professional car thief are different from the goals of a joyrider. Furthermore, the goals pursued by offenders – money, sexual gratification, and excitement – are universal motivators, even if the means by which they pursue these goals are illegal. When viewed from the offender's perspective, the decision to commit crime is rational and represents the most effective strategy for achieving his or her goals at that moment.

Critics have challenged the presumption of offender rationality that underpins the rational choice perspective, arguing that many of the choices made by offenders do not result in benefits and may be ultimately self-defeating (e.g., leading to imprisonment). A related criticism of the rational choice perspective is that it cannot explain so-called expressive crimes that appear to be spontaneous, associated with psychological disorder or committed when the offender is in a state of emotional arousal. Cornish and Clarke acknowledged that offender rationality is not perfect. Borrowing from the work of Simon (1957), they describe offender rationality as "bounded." The concept of bounded rationality recognizes that there are limits on the offender's capacity to make rational decisions. Rationality may be limited by cognitive biases, lack of information, time pressures, emotional arousal, drugs or alcohol, individual values, and a range of other factors. The utility of an anticipated outcome, therefore, is not necessarily consciously or accurately worked out when someone commits a given crime. Even crimes carried out in the heat of the moment, however, or which are the product of psychological disorder, are committed for a purpose. Offender decisions may be suboptimal, but they are "satisficing" – satisfactory and sufficient to meet the offender's immediate needs (Cornish and Clarke 2008).

Although at first sight rational choice approaches may seem to imply that offenders systematically weigh up the options each time they are about to act, few rational choice proponents believe this to be the case. As people become practiced in a behavior, the thought processes surrounding that behavior become automatic and take place below the level of conscious awareness. A common example is driving. When people first learn to drive, they focus on every aspect of the task; with practice they can drive "without thinking" as if on autopilot. (Every driver will have had the experience of reaching their destination and realizing that they have little or no recollection of the journey just completed.) So too offenders learn with practice to commit crimes without going through a laborious decision-making exercise, but proceed on the basis of "standing decisions" (Clarke and Cornish 1985). Experienced burglars, for example, make much more rapid and efficient judgements about suitable burglary targets than do inexperienced burglars (Garcia-Retamero and Dhami 2009). They may even use terms such as "instinctive," "automatic," "routine." and "second nature" when they describe how they go about committing their crimes (Nee and Meenaghan 2006).

As the above discussion of standing decisions suggests, many criminal events are not confined to single points in time but involve a sequence of stages and decisions and may extend over a prolonged period of time (Cornish and Clarke 1985, 2008). Throughout the crime commission process, the offender must make many choices as the offense unfolds. Cornish (1994) developed the concept of crime scripts to help elaborate the stages that comprise the criminal event. Crime scripts are "step-by-step accounts of the procedures used by offenders to commit particular crime" (Cornish 1994, p. 31). They are templates of the required elements of an offense and may include associated activities that occur both before and after the commission of the crime. By way of illustration, to commit a burglary a burglar will typically (1) research likely targets and assemble the necessary tools, (2) drive to the target suburb, (3) drive around the suburb, (4) scan houses for a likely target, (5) approach a likely target to check for occupancy, (6) break in and search for goods, (7) select and steal the goods, (8) load up the goods and departs, (9) avoid attracting attention, (10) leave the target suburb, and (11) store or sell the goods (Cornish and Clarke 2008). Breaking the criminal event down in this way helps identify "pinch points" for intervention. For example, prevention of burglary may include improving household security (step 4), introducing neighborhood watch (step 9), and clamping down on pawnshops that accept stolen goods (step 11).

Crime Precipitators

As Clarke and Cornish (1985) recognized when they first proposed the rational choice perspective, there are many different ways to account for the role situations play in behavior. Decision-making models such as rational choice assume that the prospective offender enters the crime scene already motivated to offend; the emphasis in these models is on the role played by the actual and perceived consequences of criminal behavior. Other explanations of the person-situation interaction, however, examine aspects of immediate environment that create or intensify the offender's motivation to offend; the emphasis in these approaches is on the antecedents of behavior. More than just enabling crime to occur (i.e., through providing opportunities), some environmental conditions may actively precipitate criminal events. The various psychological mechanisms that have been proposed in the literature to account for the precipitating effects of situations on behavior can be collected into four categories – prompts, pressures, permissions, and provocations (Wortley 2001, 2008).

Prompts refer to environment features that bring to the surface thoughts, feelings, and desires that may be lying dormant. Prompts jog people's memories, stimulate them, tempt them, evoke moods, and provide examples for people to follow. The primary theoretical underpinning for prompts is learning theory, which holds that for behavior to be produced on any given occasion, it needs to be evoked by an appropriate environmental stimulus. Hence, Pavlov's dogs did not salivate randomly but only when a bell was rung. Alternatively, prompts may be explained by the concept of priming – the capacity of environmental stimuli to facilitate retrieval of information stored in memory – described in cognitive psychology. Applying the principle of prompting to crime, the immediate environment may trigger the impulse to engage in criminal acts. For example, it has been shown that increased levels of aggression follow exposure to a range of violence-related stimuli including guns, knives, Ku Klux Klan clothing, aggressive verbalizations, vengeance-themed bumper stickers, aggressive films, and boxing films (Carlson et al. 1990).

Pressures refer to the effects that the presence of others has on an individual's behavior and are examined in social psychology. Human beings are social animals, and a great deal of social behavior is influenced by the demands and expectations of other members of the species. Social pressures include the tendency for individuals to conform to group expectations, to obey the instructions of authority figures, to submerge their identities in a crowd, and to try to please significant others. Under certain conditions, these social influences can allow good people to do bad things, a phenomenon that Zimbardo (2007) has been labelled the Lucifer Effect. Zimbardo, famous for the Stanford prison experiment (Haney et al. 1973), applied these social influences to explain the abuses of power that occurred at Abu Ghraib prison in Iraq in 2003, where US soldiers guarding the prison punched, sexually abused, urinated on, threatened, and tortured detainees. Zimbardo argued that the behavior of the soldiers could not be understood in terms of individual pathology, but as a consequence of the social dynamics of the prison that promoted a mob mentality.

Permissions refer to the neutralizations that offenders invoke to excuse their crimes (Bandura 1977; Sykes and Matza 1957). Offenders may deny the essential wrongness of their actions ("It is nothing compared with the things politicians get away with"), minimize their personal accountability ("I lost control and couldn't help myself"), deny causing any harm ("The shop was insured"), and minimize the worthiness of the victim ("She was just a whore"). Neutralizations can be facilitated by situational conditions that help obscure the relationship between the offender's behavior and its negative outcome. For example, Bandura (1977) argued that the division of labor within organizations facilitates corruption by allowing individuals to hide behind a collective responsibility ("I was just doing my job"). One of the common defenses of Nazi prisoners at the Nuremberg trials was that while they might have played a role in the deportation of Jews to the concentration camps, they were not personally responsible for any deaths.

Provocations refer to aspects of the immediate environment that induce emotional arousal and so facilitate aggression. Aversive arousal-inducing experiences include being thwarted, constrained, frustrated, insulted, threatened, annoyed, overwhelmed, crowded, and discomforted. Physiological arousal prepares the individual for a flight-or-fight response as means of managing or adapting to noxious conditions and events. In a series of studies Homel and colleagues (Homel and Clark 1994; Homel et al. 1997; Macintyre and Homel 1997) examined the relationship between violence in licensed premises (bars and nightclubs) and aggravating features of those premises. Violence was associated with a range of environmental irritants including overcrowding, the amount of cigarette smoke, poor lighting, lack of ventilation, aggressive security staff, and poor floor plans that promoted jostling as patrons tried to buy drinks.

The notion of precipitation is not proposed as a rival to rational choice but as a complementary process. Both precipitation and rational choice may play a part of every crime, but the relative importance of each may vary from crime to crime and from individual to individual. To capture the respective contributions of precipitation and rational choice, Cornish and Clarke (2003) put forward an offender typology comprising three types – antisocial predators, mundane offenders, and provoked offenders.

Antisocial predators are stereotypical, calculating criminals. They actively seek out or create criminal opportunities and enter the crime scene with preexisting motivation to commit the crime. They utilize situational data to weigh up the relative costs and benefits of criminal action, and as such their behaviors may be largely accounted for by rational choice. Mundane offenders vacillate in their criminal commitment and are opportunistic in their offending. They have generalized impulse-control weaknesses, committing their crimes with minimal planning and forethought. Like predatory offenders, they seek to derive benefits from their crimes. However, they are easily tempted into crime and may use neutralizations to help them ease any lingering feelings of guilt they may be experiencing. Provoked offenders are reacting to a particular set of environmental circumstances – situational frustrations, irritations, social pressures, and the like – that induce them to commit crimes they would not have otherwise committed. They may have conventional value systems and lead otherwise law-abiding lives. Their involvement in crime may represent an aberration and would not have occurred if it were not for the precipitating events.

As with all typologies, the categories should not be viewed as hard and fast. The offender types may be more appropriately regarded as *offending* types, and an individual offender may commit crimes across all three types. In particular, antisocial predators are versatile and prolific offenders and are likely to also commit opportunistic and provoked crimes as the situation arises (Wortley 2012).

Crime Patterns

It is a fundamental principle of the situational and environmental perspective that crimes do not occur randomly. Because crime is dependent upon criminogenic situations, it is patterned in time and space according to the distribution of those criminogenic situations. Nightclub violence, for example, will invariably be concentrated in a relatively small number of premises and will peak on certain days of the week (usually the weekend) and at certain times of the day (usually in the early morning as patrons leave) (Graham and Homel 2008). Crime hotspots are the logical places to direct policing and crime prevention resources. Thus, the set of theories that is examined in this section is concerned with explaining spatial and temporal crime patterns that follow the assumption that situations are crucial to the production of criminal acts. Two approaches are examined – routine activities and crime pattern theory.

Routine Activities Approach

At the heart of Cohen and Felson's (1979) routine activity approach lies the most parsimonious account possible of the crime event. Direct contact predatory crime events require the coincidence in time and space of three minimal elements: "(1) motivated offenders, (2) suitable targets, and (3) the absence of capable guardians against a violation" (p. 589). For example, a residential burglary occurs when a motivated burglar encounters an unguarded premise that contains desirable products. If any one of the three elements is absent – a motivated offender does not pass by the house, the property is guarded, or it contains nothing of value – then burglary will not occur. The routine activities account makes no assumptions about the psychological history or attributes of the offender beyond the fact that he or she is motivated to commit crime. In fact, in later writings, Felson (e.g., 2008) substituted "likely offender" for "motivated offender."

For such a deceptively simple idea, the routine activities approach has important implications for the distribution of crime. Cohen and Felson argued that the convergence of likely offenders, suitable targets, and unguarded locations is dictated by the natural rhythms of everyday life that produce predictable crime patterns. They set about demonstrating their thesis by examining crime rates after World War II. They noted an apparent paradox in the data – crime rates after the War rose substantially at the same time that economic conditions improved dramatically. This finding defies conventional criminological theory that predicts that crime increases with poverty. Cohen and Felson argued that higher crime rates could be explained by the changes in the routine activities of the population that accompanied economic prosperity. As economic conditions improved, the three necessary conditions for crime were often brought into alignment. Consider the case of burglary. Growing affluence and technological advances meant that there were more compact, valuable personal possessions available to steal in the average home (suitable targets). At the same time, as women increasingly took up work outside of the home, more houses were left unattended during the day (absence of capable guardians). These changes altered the behavior of likely offenders and resulted in significant shifts in burglary patterns. Between 1960 and 1975 rates of commercial burglary in the United States almost halved, while the rates for domestic burglary doubled as offenders chased the more lucrative pickings in suburban houses. Furthermore, burglary went from being a nighttime to a daytime activity, as offenders targeted the growing numbers of properties that became temporarily empty when women took paid employment outside the home.

There are many examples of crime patterns that can be explained by the respective routine activities of victims and offenders. Young males experience high rates of physical victimization because they have risky lifestyles that involve visiting bars in dangerous districts, staying out late at night, and mixing with other alcohol-affected youths (Jensen and Brownfield 1986). Juvenile violence peaks in the hour or so around the close of school in the afternoon when large numbers of young people – potential offenders and victims – spill into the streets (Snyder et al. 1996). Burglary is more likely to occur when the weather is fine and people are likely to go on excursions away from home (Hipp et al. 2004). As the previously discussed negative association between poverty and burglary demonstrates, an appealing feature of the routine activities approach is its capacity to explain apparently counterintuitive crime patterns. In a further example of this, Felson (2002) cites research showing that young people who have part-time jobs unexpectedly commit *more* crime than those who do not have jobs. As Felson explains, a job allows a young person to fund a relatively risky lifestyle (in terms of crime opportunities) that might include taking illegal drugs, drinking alcohol, visiting shopping malls, and driving his or her own car.

While Cohen and Felson were concerned with the impact of macro-level social forces on patterns of criminal behavior, subsequent theorizing has explored the compatibility of the routine activities approach with the rational choice perspective and situational crime prevention (Clarke and Felson 1993). It was realized that the three necessary elements for crime – an offender, a target, and an absent guardian – provided a framework for analyzing the dynamics of individual crime events and for determining points of intervention for crime prevention. The familiar crime triangle ("hot" offenders, "hot" locations, "hot" targets) used in operational policing has affinities with the routine activities approach (Clarke and Eck 2003). All prevention efforts involve neutralizing one or more of these three elements. Some crimes occur because of the presence of motivated offenders and so intervention involves strategies to divert them from likely trouble spots. Other crimes are the result of poor management of facilities and so require interventions that strengthen guardianship in those locations. And other crimes occur because of the easy accessibility of vulnerable targets and so intervention requires target-hardening strategies.

Crime Pattern Theory

Crime pattern theory was developed by Brantingham and Brantingham (1991, 2008) to account for nonuniformity and nonrandomness in the geographical distribution of criminal events. As with the routine activity approach, crime pattern theory assumes that crime is associated with the everyday movements of offenders and victims in time and space. However, while the routine activity approach emphasizes the supply, distribution, and movement of suitable targets, likely offenders, and capable guardians for explaining crime patterns across space and time, crime pattern theory is concerned specifically with the role played by the urban landscape. According to crime pattern theory, crime is associated with the distribution of key facilities and activities in a community that shape the offender's daily movements and consequent familiarity with his or her environment.

Crime pattern theory explores the role in crime of three key urban elements – nodes, paths, and edges – that individuals encounter and become familiar with in their daily movements. Nodes are locations that an individual regularly visits in the course of his or her routine activities. They include places where a person lives, works or goes to school, shops, and goes to for entertainment and recreation. Paths are the usual routes the person takes to move between nodes. An edge is the boundary between areas of land use, which individuals encounter at the fringe of their nodes and as they traverse paths between their nodes. The most common edge is the dividing line that separates one neighborhood from another.

Together nodes, paths, and edges form an individual's awareness space. Awareness spaces are the parts of the urban environment that the individual knows well and in which he or she feels comfortable. Most people, for example, have a set way that they travel from home to work or school, and they become very familiar with the environment along that path. They may, however, know little about the environment just one or two streets outside of their usual route and feel disoriented if they stay too far from their awareness space.

Crime pattern theory predicts that crime will occur in locations in which crime opportunities intersect with an offender's awareness space. Crime opportunities cluster around certain nodes referred to as crime generators and crime attractors. Crime generators are locations that attract large numbers of people for legitimate, noncriminal activities. They include sports stadiums, shopping malls, and transport hubs. The presence of so many people provides many criminal opportunities (e.g., pickpocketing) for offenders who may happen to be mixed in with the general crowd of patrons. Crime attractors are districts or premises that attract potential offenders for the specific purpose of committing crime. They include certain bars, red light districts, and non-secured parking lots. Offenders frequent these areas to sell or obtain drugs, to fence stolen goods, to pimp for prostitutes, and so on. Crime attractors are particularly likely to be located near freeway exits and public transport hubs, which provide convenient access and escape. Crime generators and attractors are likely hotspots for crime. Edges, too, are often high-crime locations, where strangers come together and offenders can commit crime with little fear that they will be recognized. Edges can also be sites for conflict between those identifying with areas on either side and for graffiti to mark territorial claims (Brantingham and Brantingham 1995).

Crime pattern theory can be used to help explain specific movement patterns of offenders. One such pattern is the so-called journey to crime. A prediction that follows from crime pattern theory is that offenders generally will not travel far from their key personal nodes – their home, work, or school – in order to commit crime. They will prefer to access criminal opportunities that occur within their awareness space, but at a distance that reduces the chance that they will be recognized. This prediction may at first seem counterintuitive – it might be expected that offenders would travel away from their usual haunts in order to search out the most desirable targets and to put a safe distance between themselves and the crime. However, the prediction is also consistent both rational choice perspective and routine activities approach. Going out of one's way to offend requires more effort than committing crimes in locations that are encountered in the course of one's daily routine. Research examining the travel habits of offenders confirms that the frequency of offending decays as a function of the distance the offender is from home. Snook (2004) examined data from 41 serial burglars. He found that the median distance travelled to commit a burglary was 1.7 km and that 85 % of burglaries were committed within five kilometers from the burglar's home. The rate of decay was rapid; 33 % of burglary sites were within one kilometer from home, 25 % between one and two kilometers, and 15 % between 2 and 3 km. Offending not only occurs close to home but also along familiar pathways. Rengert and Wasilchick (1985) examined the direction of burglary sites in relation to the burglar's home. They found that over 75 % of burglary sites were within a 45° arc around the path between the burglar's home and place of work, while 57 % were in a 45° arc around the path between the burglar's home and place of recreation.

Conclusion

A common criticism of the situational and environmental perspective in criminology is that it is atheoretical, naive, and ignores the "root causes" of crime. Prevention efforts based on the situational and environmental perspective are, in turn, often characterized as amounting to little more than a "locks and bolts" approach to reducing crime. The purpose of this entry has been to provide an overview of the theories that underpin situational and environmental prevention. The intention has been to show that human behavior – and hence criminal behavior – is by its very nature situational. All crime involves a complex interplay between the disposition of the offender and the situation he or she is in at the time of the offense. The situational and environmental perspective does not ignore root causes of crime; the immediate environment is a root cause of crime.

The models of offender behavior and the accounts of crime patterns described in this entry provide the bases for situational and environmental crime prevention. There are a number of different approaches taken to prevention, but all are based on the basic premise that altering some aspect of the immediate crime environment has the potential to disrupt crime. Crime Prevention Through Environmental Design (CPTED) is one of the earliest formal models of situational/environmental crime prevention. Its focus is on the role that architecture and urban design play in crime and crime reduction. Situational crime prevention (SCP) presents a microanalytic approach to crime reduction. Its focus is on identifying and altering the specific elements of the crime scene that permit or encourage specific crimes to occur. Design against crime (DAC) examines the implications of the situational and environmental perspective for the design of products and systems. It explores the premise that some products encourage crime and that these criminogenic features may be designed out at the production stage. Crime science is the newest model in the situational and environmental perspective. Crime scientists argue that crime prevention initiatives need to be based around the scientific method. They also argue that the crime prevention task should be viewed in truly multidisciplinary terms and take in contributions of "hard" scientists – mathematicians, engineers, computer scientists, and so on – as well as social scientists.

Related Entries

- ▶ Crime Mapping
- ▶ Crime Prevention Through Environmental Design
- ▶ Crime Science
- ▶ Designing Products Against Crime
- ▶ Effectiveness of Situational Crime Prevention
- ▶ Hot Spots and Place-Based Policing
- ▶ Rational Choice Theory
- ▶ Repeat Victimization
- ▶ Routine Activities Approach
- ▶ Situational Crime Prevention

Recommended Reading and References

Bandura A (1977) Social learning theory. Prentice Hall, Englewood Cliff

Brantingham PJ, Brantingham PL (1991) Environmental criminology, 2nd edn. Waveland Press, New York

Brantingham P, Brantingham P (1995) Criminality of place: crime generators and crime attractors. Eur J Crim Policy Res 3(3):2–26

Brantingham PJ, Brantingham PL (2008) Crime pattern theory. In: Wortley R, Mazerolle L (eds) Environmental criminology and crime analyses. Willan, Cullompton

Carlson M, Marcus-Newhall A, Miller N (1990) Effects of situational aggressive cues: a quantitative review. J Pers Soc Psychol 58:622–633

Clarke RVG (1967) Seasonal and other environmental aspects of abscondings by approved school boys. Br J Criminol 7:195–206

Clarke RV, Cornish DB (1985) Modeling offenders' decisions: a framework for research and policy. Crime and Justice 6:147–185

Clarke RV, Eck J (2003) Become a problem-solving crime analyst. Jill Dando Institute of Crime Science, London

Clarke RV, Felson M (eds) (1993) Introduction: criminology, routine activity, and rational choice. In: Clarke RV, Felson M (eds) Routine activity and rational choice. Advances in criminological theory, Vol. 5. Transaction Publishers, New Brunswick

Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44:588–608

Cornish D (1994) The procedural analysis of offending and its relevance for situational prevention. In: Clarke R (ed) Crime prevention studies. Criminal Justice Press, Monsey

Cornish DB, Clarke RV (2003) Opportunities, precipitators and criminal dispositions: a reply to Wortley's critique of situational crime prevention. In: Smith MJ, Cornish DB (eds) Theory and practice in situation crime prevention, vol 16, Crime prevention studies. Criminal Justice Press, Monsey

Cornish DB, Clarke RV (2008) The rational choice perspective. In: Wortley R, Mazerolle L (eds) Environmental criminology and crime analyses. Willan, Cullompton

Felson M (2002) Crime in everyday life, 3rd edn. Sage, Thousand Oaks

Felson M (2008) Routine activity approach. In: Wortley R, Mazerolle L (eds) Environmental criminology and crime analyses. Willan, Cullompton

Garcia-Retamero R, Dhami MK (2009) Take-the-best in expert-novice decision strategies for residential burglary. Psychon Bull Rev 16:163–169

Graham K, Homel R (2008) Raising the bar: preventing aggression in and around bars, pubs and clubs. Cullompton, Willan

Haney C, Banks C, Zimbardo P (1973) Interpersonal dynamics in a simulated prison. Int J Criminol Penol 1:69–97

Hipp JR, Bauer DJ, Curran PJ, Bollen KA (2004) Crimes of opportunity or crimes of emotion: testing two explanations of seasonal change in crime. Soc Forces 82:1333–1372

Homel R, Clark J (1994) The prediction and prevention of violence in pubs and clubs. In: Clarke RV (ed) Crime prevention studies, vol 3. Criminal Justice Press, Monsey

Homel R, Hauritz M, Wortley R, McIlwain G, Carvolth R (1997) Preventing alcohol-related crime through community action: the Surfers Paradise Safety Action Project. In: Homel R (ed) Policing for prevention, vol 7, Crime prevention studies. Criminal justice Press, Monsey

Jensen JF, Brownfield D (1986) Gender, lifestyles, and victimization: beyond routine activity. Violence Vict 1:85–99

Macintyre S, Homel R (1997) Danger on the dance floor: a study of interior design, crowding and aggression in nightclubs. In: Homel R (ed) Policing for prevention: reducing crime, public intoxication and injury. Criminal Justice Press, Monsey

Mayhew P, Clarke RV, Sturman A, Hough J (1976) Crime as opportunity, vol 34, Home office research study. HMSO, London

Mischel W (1968) Personality and assessment. Wiley, New York

Nee C, Meenaghan A (2006) Expert decision-making in burglars. Br J Criminol 46:935–949

Rengert G, Wasilchick J (1985) Suburban burglary: a time and place for everything. Charles Thomas, Springfield

Simon HA (1957) Models of man. Wiley, New York

Snook B (2004) Individual differences in distance travelled by serial burglars. J Investig Psychol Offender Profiling 1:53–66

Snyder HN, Sickmund M, Poe-Yamagata E (1996) Juvenile offenders and victims: 1996 update on violence. US Department of Justice, Washington, DC

Sykes GM, Matza D (1957) Techniques of neutralization: a theory of delinquency. Am Sociol Rev 22:664–670

Wortley R (2001) A classification of techniques for controlling situational precipitators of crime. Secur J 14:63–82

Wortley R (2008) Situational precipitators of crime. In: Wortley R, Mazerolle L (eds) Environmental criminology and crime analysis. Willan, Cullompton

Wortley R (2012) Exploring the person-situation interaction in situational crime prevention. In: Tilley N, Farrell G (eds) The reasoning criminologist: essays in honour of ronald V. Clarke. Routledge, London

Zimbardo P (2007) The lucifer effect. Random House, New York

Theories of Co-offending

Frank M. Weerman
Netherlands Institute for the Study of Crime and Law Enforcement (NCSR), Amsterdam,
The Netherlands

Overview

This entry is focused on theories and theorizing about co-offending. This includes explaining why people co-offend in the first place, how co-offenders find each other and initiate their crimes, and how co-offending may have consequences for offenders and offending rates. The entry starts by outlining three general approaches in the field. After that, a small number of "Theories of co-offending" are addressed in more detail—elaborated theoretical publications about understanding the causes of co-offending or certain aspects of it. Finally, a number of other theoretical issues are addressed, which have received quite some attention in the field. Following each of these issues, a short outline is provided for the remaining questions that are still in need of further theoretical elaboration.

Introduction

One of the basic findings of criminology is that a substantial part of crime is committed in the company of others—"co-offending." Early twentieth-century scholars already signaled that co-offending is abundant, particularly among juveniles. Since then, a body of empirical research has emerged in which considerable knowledge has been gathered about the regularities of co-offending and the characteristics of co-offenders (see the other entries in this area).

Co-offending is not only an interesting subject for empirical study, it also poses many interesting and profound theoretical questions. Since breaking the law does not necessarily require others, the sheer phenomenon of co-offending itself is meaningful and in need of explanation.

Not everyone is willing and capable to break the law, and it is important to understand how multiple offenders find and choose each other and communicate to make co-offending possible. And since co-offenders interact with each other, offending in company may have important consequences for the modus operandi of the offense and the criminal careers of those who are involved. Co-offending implies that the "antisocial" activity of offending must have many social aspects. It may teach us many things about the relationships and social structures that govern the collective behavior of offenders (Weerman 2003).

Although theorizing about co-offending has received much less attention as empirical study of the subject and much less attention as theorizing about crime and delinquency in general, a few contributions in the field explicitly aimed to theorize the causes and consequences of co-offending (Felson 2003; McCarthy et al. 1998; Tremblay 1993; Weerman 2003). More theorizing about co-offending, explicit or implicit (scattered notions and hypotheses), can be found in several empirically oriented studies (e.g., Carrington 2009; McGloin et al. 2008, McGloin and Stickle 2011; Morselli et al. 2006; Reiss and Farrington 1991; Van Mastrigt and Farrington 2011; Warr 1996). It is possible to distinguish a few general approaches or perspectives and a couple of more formally elaborated "theories of co-offending." However, there are also remaining theoretical questions about co-offending that are in need of further elaboration.

Three General Perspectives

Sometimes explicit, but more often implicit, basic assumptions are made about the nature of co-offending, and scholars build their arguments upon them. Some regard co-offending as living proof of the "group character" of crime and the importance of peers; some scholars think of co-offending as a merely coincidental convergence of offenders at the same place and time; and others view co-offending as an instrument to optimize the execution of an offense. These three diverging views coincide with important general perspectives in criminology (social learning theories, control theories, and the routine activities or rational choice approach), and as such they may be regarded as competing perspectives. They may also be seen as points of view that highlight different aspects of co-offending.

Co-offending as the Result of Peer or Group Influence

Many scholars writing about co-offending see it as the necessary result of peer influence or group dynamics. For example, Shaw and McKay (1931) regarded delinquency as mainly group behavior and co-offending as one of the ways in which offending was transmitted among young people. Delinquents learned their techniques from older, more experienced offenders. Scholars who revived interest in co-offending in the 1970s also regarded it as the result of group dynamics. Some of them argued that the abundance of co-offending is an indication that "delinquency is still group behavior" (Erickson and Jensen 1977). More recently, Warr (2002) argued that co-offending, together with the association between delinquency and delinquent peers, provides evidence of the social nature of criminal behavior.

If one would reconstruct the basic arguments of this perspective, the causal path would run from being part of a group and/or having delinquent peers to being exposed to peer influences that lead someone to join others in offending. Peer influence may consist of social learning mechanisms and the acquisition of delinquent definitions and group processes like fear of ridicule and the urge to show loyalty and obtain status during adolescence (see Warr 2002). People co-offend because the group in which they became delinquent also leads them to join in criminal activities within the group.

Co-offending as the Result of Social Selection

Some scholars explicitly denied that peer influence is a major cause of delinquent behavior (e.g., Hirschi 1969). According to them, other causes, mainly within the individual and the family (e.g., a weak bond of society or lack of self-control), are responsible for the onset of delinquency. Having delinquent peers is a byproduct: criminal or delinquent groups are formed because offenders find each other in the same circumstances or select each other as friends ("birds of a feather stick together"). Their argument also accounts for co-offending: this is the result of offenders who are in each other's company because of social selection. Most explicit about this is Kornhauser (1978), who argued that co-offending is nothing special, since adolescents do most of their activities in groups. More recently, Stolzenberg and D'Alessio (2008) posited that co-offending among juveniles is "theoretically inconsequential." They argue that youths tend to be more gregarious than adults (having a tendency to spend time with others), and therefore also tend to commit their crimes more often in groups as adults and more often than alone.

The basic arguments of this perspective would be that co-offending occurs as byproduct when offenders associate together and spend a lot of time in groups. For reasons other than peer influence, offenders are prepared to join criminal activities, and they will co-offend when they are coincidentally together and an offense opportunity emerges.

Co-offending as an Instrumental or Rational Choice

Finally, several scholars analyze co-offending from an instrumental or rational angle. According to the rational choice perspective in criminology, decisions of offenders may be regarded as if they were rational calculations of the expected rewards and costs of certain actions. The decision to co-offend is also analyzed as a rational or instrumental decision. Researchers who have studied specific crimes like burglary and robbery often adhere to this approach, when they describe co-offending as a way of making the offense easier and more rewarding (e.g., Walsh 1986). Also a number of scholars who explicitly theorized about co-offending have reasoned from an instrumental view on co-offending. For example, Tremblay (1993) assumes that co-offending is required for many types of offending, including organized crime, and many motivated offenders need to find suitable co-offenders. According to him, this quest can be regarded as the outcome of individual decision making depending on situational circumstances (see the next section). Likewise, McCarthy et al. (1998) analyzed co-offending as a form of cooperation between in essence selfish parties to optimize the perpetration of the offense. However, they went a step further by looking at collective forms of rationality and ways in which co-offenders deal with the risks of being betrayed.

The basic arguments of this perspective would be that the decision to co-offend is more or less rational, and that it involves weighing advantages of co-offending against the potential costs. Co-offending can ease the execution of the offense and increase the expected catch, but it also implies that profits have to be shared and that co-offenders may cheat or talk to the police. Co-offending is chosen when it is expected to be easier and more rewarding than solo-offending.

Explicit Theoretical Contributions to the Literature About Co-offending

A few publications can be regarded as more or less explicit theories of co-offending, focused on explaining the phenomenon itself, particular aspects of it, and/or variation in co-offending among offenders and situations. These publications will be addressed in more detail below.

Understanding How Co-offenders Find Each Other (Tremblay, Felson)

One of the first scholars who explicitly theorized about co-offending was Tremblay (1993). In a paper titled (Searching for Suitable Co-offenders), he theorized about the criteria for suitable co-offenders and how they may be found. This paper employed a broader conception of co-offending than usually followed in the literature: encompassing not only the actual companions during the crime but also accomplices and acquaintances that are needed before and after the offense—advisers, fences, and sellers and buyers of illegal goods.

According to Tremblay, many acts of crime become possible only when an offender can find suitable co-offenders. This is more complex as it may appear: an offender needs to get in contact with someone that can be trusted and is able to provide the needed services, which can be very specialized. The availability of co-offenders may be increased by certain social circumstances: neighborhoods with relatively more offenders, periods of high unemployment, and time spent in a prison. However, even if there are high concentrations of potential co-offenders, choices need to be made on who is actually a partner in a particular offense or a series of crimes.

Tremblay assumes that there are two main criteria to decide which co-offenders are suitable and how many: trustworthiness and usefulness. Trustworthiness is most benefited by a small number of co-offenders with who one has a strong relation or bond (e.g., family or friends). However, because this is often not possible, offenders often adopt a basic attitude of distrust towards potential co-offenders. The criterion of usefulness is most benefited by a large social network of friends and vague acquaintances. These relatively weak bonds may enhance contacts with others who have skills and knowledge that the offender needs.

The two basic criteria imply that offenders would be best served by two kinds of social networks at the same time: a cohesive network of strong ties to enhance trust and a large and relatively loose network of weak ties. In practice, this is usually difficult to achieve and Tremblay supposes that most offenders choose one of three strategies to deal with the problem of finding suitable co-offenders. In the "exclusive strategy," offenders choose one type of network and concentrate on establishing strong bonds with trustworthy others or getting a large social network with many potentially useful co-offenders. In the "mixed strategy," offenders try to develop a few strong relationships as well as many weak ties. Finally, in the "avoidance strategy," offenders keep out of the problems by committing only solo-offenses.

More recently, Felson continued theorizing about the problem of finding suitable co-offenders in a paper titled "The process of co-offending" (Felson 2003). Felson concentrates on actual types of co-offending in which at least two offenders are directly and simultaneously cooperating. Drawing on the routine activities approach, combined with a social learning perspective on co-offending, Felson develops a theory that emphasizes the importance of meeting places for offenders, or as he names it "convergence settings." According to him, these places strongly facilitate co-offenders to find each other and strongly enhance offending rates.

A basic assumption in Felson's theory is that "access to accomplices is inherently criminogenic" (Felson 2003, p. 151). Several arguments are provided for this stance. First of all, when potential offenders get together there is an increased chance that they will persuade each other to break the law in the near future. Here, Felson clearly adheres to a peer influence perspective on co-offending. Second, offenders may learn a lot of practical things from each other. They may provide each other with information about potential targets and their locations, about the best way to approach them, and dispose of stolen goods. And third, access to accomplices is needed when an offender wants actual assistance and help in carrying out an offense. By getting into contact with other potential offenders, one may find some necessary extra manpower or someone with specialized skills or knowledge.

Having stressed the importance of getting access to co-offenders, Felson argues that finding them is less easy as it may appear. Drawing on routine activities theories, Felson formulates three requirements. First of all, co-offenders need to get together at the same place and time. Second, they must be able to interact with each other without outside interference. And third, they need to have substantial time available to socialize with each other. With regard to the latter, Felson emphasizes that screening the availability and suitability of potential co-offenders is a slow and informal process: offenders need to size each other up and bond together by talking, drinking, or other informal activities. These three requirements imply that "convergence settings" need to be stable and predictable sources of potential co-offenders. In practice, a small number of places apply: places like street segments, bars, houses, and neighborhood facilities, often only during particular hours. This relative rarity implies that they are important determinants of crime in a certain area, making them potentially interesting targets for intervention.

Understanding How the Risks of Co-offending Are Dealt With (McCarthy et al.)

Co-offending involves risks and costs. In a paper titled "Uncertainty, cooperation, and crime: understanding the decision to co-offend," McCarthy et al. (1998) ask themselves how it is possible that offenders cooperate, despite their conflicting interests and basic uncertainty about the real intentions of the other. Offenders, they argue, can never be sure whether co-offenders stick to their agreements and refrain from cheating or talking to the police. Co-offending has resemblance to a "prisoner's dilemma," in which the least preferred option for co-offenders would be the situation in which they would cooperate and the other would cheat. To explain how co-offending is possible in this situation, McCarthy et al. draw on sociological literature about social dilemmas, situations in which individual and collective goals are contradicting. This literature suggests that people are capable of reasoning with "collective rationality." Individuals can take other people's wishes and needs into account when the end result is also in their own interest. In the case of co-offending, collective rationality dictates that it is in each other's interest not to talk to the police and share the profits of the crime, to ensure future collaboration. Adding to this, trust is needed to be confident that co-offenders also take collective rationality into account in their decision making.

A complicating factor is that not all offenders have the luxury to contemplate who is trustworthy and who is not. McCarthy supposes that offenders who are low in status and live in adverse circumstances (e.g., the street youths that were studied by the authors) will be more inclined to join others than others who are less dependent upon others.

Understanding Who Is Co-offending and Why: Co-offending as Social Exchange (Weerman)

An attempt to integrate many of the then-existing perspectives on co-offending was undertaken by Weerman (2003) in an article titled "Co-offending as social exchange: explaining characteristics of co-offending." This paper aimed to formulate a comprehensive theory about why people co-offend and about variations and regularities in co-offending. The theory is based on the social exchange perspective from sociology and psychology. The basic idea of the theory is that co-offending can be analyzed as an event where goods are exchanged between offenders. The "exchange goods" in co-offending are not only material or instrumental, like a share in the gains of the offense (the catch or a payment), but also immaterial, such as social approval and acceptance. According to Weerman, people co-offend because it is a way of obtaining rewards through the exchange of goods that cannot be obtained by solo-offending. Because the exchange is not limited to the material "catch" of the offense, co-offending can occur even when it is not necessary to carry out a specific crime or it can be done in ways that are not the most efficient to complete the offense.

Six types of "co-offending exchange goods" are distinguished in the theory: services, payment, "catch," appreciation, acceptance, and information. The category of "services" can be considered as basic: these are the co-offenders' activities that make the offense possible (actual help during the offense, providing devices or resources, being a lookout or just being present during the offense). The category of "catch" consists of the share in the money or goods from the offense, often divided among the co-offenders. The category of "payment" consists of all the material rewards (money or goods) that a co-offender gets from another as payment for his service. The category of "appreciation" is immaterial: the social rewards for the actions of the co-offender (approval or praise). The category of "acceptance" is also immaterial: the social rewards for the person of the co-offender (ranging from toleration in a group to affection, status, and respect). The category of "information" is also immaterial: this refers to knowledge obtained through co-offending (practical knowledge about techniques or contacts or knowledge about the capacities and trustworthiness of co-offenders).

Based on these "exchange goods," the theory distinguishes different forms of co-offending. In instrumental forms of co-offending, services are exchanged with a share in the catch or with payments. In strategic forms of co-offending, the goal is to obtain important information apart from the catch (e.g., to test someone or to learn a special technique). Material and immaterial rewards are both involved in quasi-instrumental forms of co-offending, when co-offenders try to impress each other with their skills or with their bravado, or when participating in co-offending is required to be a member of a group. In expressive forms of co-offending, only social rewards are involved (e.g., youth groups committing acts of vandalism). Further, co-offending may be equal or unequal, depending on the exchange goods that everyone is receiving.

The theory explains variation among offenders and offenses by referring to three conditions that make co-offending possible. The first condition is that the offender is willing to co-offend. This depends on whether someone perceives co-offending as "profitable enough" and offering material and immaterial exchange goods that exceed the potential costs (sharing the gains, investing time and energy, the possibility of being caught). The theory also assumes that offenders who have mainly immaterial needs, for example, young adolescents who offend to be accepted and appreciated, are in many cases rewarded by co-offending. Offenders with mainly material needs, like professional or addicted criminals, are only rewarded when they get a lot of catch or payment out of co-offending or when co-offending is the only way in which they can obtain the goods they want.

The second condition is that one or more potential co-offenders are directly available or easy to contact. It is assumed that having a large criminal network and being part of a delinquent group facilitates access to potential co-offenders. Further, regular meeting places and times are expected to ease access to potential co-offenders (cf. Felson 2003). The third condition is that offenders need to convince others that they are worthy and useful to co-offend with. This is only the case when the offender has resources that make offending with him "profitable enough" for others. This attractiveness is mainly influenced by the criminal capacities and possibilities of the offender to offer valuable material and immaterial rewards. Someone can be attractive by his or her physical capital (tools like guns and crowbars), human capital (knowledge, smartness, and criminal insight), and social capital (knowing other useful people).

Other Theoretical Issues and Pending Questions About Co-offending

This section elaborates some important theoretical issues about co-offending that received considerable attention in the literature, though not in the form of systematic theoretical papers. Current notions and hypotheses about these issues are described, and pending questions and gaps in our knowledge are signaled.

Understanding the Process Leading to Co-offending: Recruitment, Instigation, and Decision Making

An interesting theoretical issue is how co-offending is initiated and whether particular offenders serve as recruiters or initiators. This issue has been the subject of several empirical studies which are based on various theoretical assumptions (see the entry about co-offending and group crime for empirical results). Reiss (1988) and Reiss and Farrington (1991) suspected that within the population of offenders, there is a small number of high-rate offenders who act as recruiters. These recruiters would initiate many offenses with less or non-experienced co-offenders. They theorized that these recruiters may get many others involved in their crimes who would otherwise commit no or much less offenses. Recently, Van Mastrigt and Farrington (2011) further investigated the existence of recruiters. They suggested that the developmental taxonomy theory of Moffitt (1993) offers an additional angle to understand recruitment. This theory would predict that recruiters are usually life-course persistent offenders who differ strongly in their characteristics from joiners, presumably adolescent-limited offenders.

In a seminal paper titled “Organization and instigation in criminal groups,” Warr (1996) proposed that the organization of co-offending is much more dynamic and situational as the term “recruiter” suggested; instead co-offenders may act as “instigator” or “joiner” depending on the situation. Based on his empirical results, he supposed that instigators of co-offending do not have certain stable personality characteristics; instead who initiates a crime largely depends on the situation and the group composition among co-offenders. Those who appear to be the oldest among the present adolescents would have the highest probability of becoming an instigator in that situation, because older offenders generally have the highest status in a group.

Recent contributions to the literature suggest that it is also possible that co-offending emerges in other manners, not because one of the present offenders takes an explicit initiative but because subtle processes occur in which offenders become mutually aware of each other’s willingness to commit a crime in a certain situation. Based on a qualitative study among convicted robbers, Hochstetler (2001) distinguishes three ways in which co-offenders may arrive to their decision to commit an offense together. One way is that co-offenders signal to each other with implicit verbal and nonverbal expressions that they recognize a crime opportunity and are willing to act. This is a decision that is slowly built and sometimes only reached in the last moment before the actual offense. A second way is that co-offenders know from each other that they are willing to offend and only need a gesture or a few words to act. A third way is that co-offenders have developed a common identity around a certain type of offense and do not need any communication at all. They commit these offenses with no need for discussion.

A potential next step in theorizing the initiation of co-offending would be to integrate these different opinions and formulate hypotheses about when co-offenders initiate crimes as sole instigators or recruiters and when they do so in concert with other offenders. It might be helpful to distinguish different roles within groups of (potential) co-offenders, each with their own group status and motives to participate in the offense. A further step would be to theorize how co-offending evolves after different ways of initiating it. Social psychological literature about group processes and decision making in small groups may be helpful in this regard (see the entry about co-offending and group crime).

Understanding Co-offending Over the Life Course

Another issue that has received much attention in the co-offending literature is the relation between age and co-offending (as offenders become older, they tend to co-offend less often or with less accomplices on average). Much of the literature on this issue is covered by the entries of Van Mastrigt and Carrington. Within these studies, several explanations have been put forward to explain the relationship. Reiss and Farrington (1991) offer two competing hypotheses, which are in fact based on the group influence perspective and the selection perspective on co-offending. One possibility, they suggest, is that most co-offenders are predominantly breaking the law because of group processes; they usually end their criminal career at the end of adolescence, while those who offend alone continue. The other possibility, Reiss and Farrington suggest, is that some co-offenders change their modus operandi as they grow older, partly because they get more experience, but also partly because the availability of co-offenders decrease after adolescence. This would be the logical result of the age-crime curve, which implies that there are relatively many potential co-offenders during adolescence, but increasingly less when offenders grow older. Adding to this, Carrington (2009) suggests that the relation between age and co-offending partly stems from the general tendency of young people to spend time together and do things in groups. He argues that delinquent behavior is just another type of group activity among many others during adolescence (see also Stolzenberg and D’Alessio 2008).

Empirical studies on the relation between age and co-offending offer contradictory findings (see the entry about co-offending and offender attributes). This has led some scholars in the field to theorize that co-offending over the life course depends on the type of offense, or that there are actually different types of co-offending trajectories over the life course, with some categories of offenders who do not change their level of solo- or co-offending and other categories with offenders who switch from co-offending to solo-offending (Carrington 2009; McGloin et al. 2008). This is interesting since it suggests that co-offending over the life course is a heterogeneous and dynamic phenomenon: it may differ among offenders, offenses, and situations. Carrington (2009) formulates a number of hypotheses about the relation between co-offending and criminal careers, based on the distinction that Reiss (1988) made between low- and high-rate offenders and the distinction that Moffitt (1993) made between life-course persistent and adolescent-limited offenders. Low-rate offenders would usually start their career with co-offending and may shift towards solo-offending when they get more experienced; high-rate offenders develop their delinquency independent from other offenders and mix solo- and co-offending from the start.

The logical next step in theorizing about the link between age and co-offending would be to understand further why there are different trajectories of co-offending over time, how these trajectories are connected to different motives for co-offending, and, more generally, how co-offending careers are linked to the etiology of delinquency in general.

Understanding the Consequences: When and Why Does Co-offending Lead to (More) Delinquent Behavior

Many contributions about co-offending stress the fact that it represents a large share of the total amount of crime, in particular if one considers the fact that one act of co-offending involves multiple people (see, e.g., Andresen and Felson 2010). Several studies also suggest that co-offending may have consequences for future offending, in the sense that it may lead to increased levels of (violent) delinquency. An important theoretical issue, then, is to get insight in the consequences of co-offending: does it lead to more crime in the future, and if so why and when?

To start with, opinions diverge whether co-offending is related to increasing levels of crime. According to those who adhere to a selection perspective on co-offending, the phenomenon mainly represents the group character of many activities during adolescence (Kornhauser 1978; Stolzenberg and D'Alessio 2008). The real causes of delinquency are found somewhere else, and this would imply that co-offending would not substantially increase involvement in crime.

Most scholars, however, believe that co-offending is associated with increased levels of offending. Several theoretical arguments have been offered to substantiate this expectation. First of all, co-offending may actually mark the onset of delinquency for many adolescents. Scholars who assign a substantial role to "recruiters" (Reiss 1988; Reiss and Farrington 1991; Van Mastrigt and Farrington 2011) or "mentors" (Morselli et al. 2006) believe that initiators of co-offending often lead inexperienced or novice offenders into (more) offending. Second, many adolescents start their first offense in situations of "pluralistic ignorance," when young people believe that rule breaking is expected from the others in a group, while in reality this does not need to be the case (see Warr 2002).

Many have also provided arguments why co-offending may enhance involvement in crime. An important mechanism that is offered is social learning. Conway and McCord (2002) suggested that offenders learn particular types of crime through co-offending. Young offenders who would commit their offenses in the company of violent offenders, for example, would imitate this behavior more often in their own future criminal career (see also the entry about co-offending and group crime). Other group mechanisms may be responsible as well. McGloin and Piquero (2009) suggest that offending with relatively many accomplishes facilitate collective behavior and the diffusion of responsibility, which increases the chances of violent behavior in the future.

Another argument why co-offending may lead to increasing and more serious involvement in crime is that it offers access to knowledge, skills, and contacts that make future offending possible. Felson (2003) argues that this is a reason why co-offending not only leads to more future co-offending but also to more future lone offending. Further co-offending increases because accomplices learn how to work together and how to get access to future accomplices. Lone offending is also facilitated when skills and knowledge are obtained through co-offending. Kleemans and Van de Bunt (1999) argued that co-offending often leads to a "social snowball effect." Offenders who become involved in an ongoing criminal operation use the knowledge and resources that they get from co-offending to set up new and separate illegal businesses that attract other co-offenders.

Future theorizing may focus on the question which categories of offenders are influenced and stimulated most by their involvement in co-offending. It seems logical to connect this question to the issue of recruitment and instigation, because this suggests that some co-offenders may initiate law breaking and convince others to join them. Another fruitful route may be to use life course and developmental criminology. Co-offending may facilitate involvement in crime at particular moments during a criminal career, and life events may counteract these effects or make offenders more susceptible to them.

Understanding Failure and Success in Co-offending

Another interesting theoretical issue relates to explaining the results of co-offending: when is it rewarding or not and which offense and offender characteristics contribute to the failure or success of co-offending?

In the past, several scholars suggested that co-offending results in higher apprehension risks for those who participate in it, in what has become known as "the group hazard hypothesis" (e.g., Hindelang 1976). They supposed that the presence of two or more offenders increases the chance that one of them makes a mistake, is caught by the police, or when caught, gives in to the pressure to talk about the others. A similar tendency has been suggested in studies focused on offenders of burglary and robbery (e.g., Cromwell et al. 1991). These studies also suggest that burglars and robbers may take more risks when they are in groups. Further, co-offenders may betray each other or cheat by taking an uneven share in the catch (McCarthy et al. 1998).

Whether co-offending exceedingly results in apprehension or not is still an open empirical question. Some studies have suggested that this is the case, but others did not find any difference (see Feyerherm 1980). More importantly, insight is needed on the conditions and circumstances that increase or decrease the risks of co-offending. Several scholars have suggested that social network characteristics of co-offenders and potential co-offenders may facilitate or counteract detection. There is a trade-off between success and risk in criminal networks: having a lot of ties gives you a lot of opportunities, but also adds to the risks of betrayal and apprehension (Morselli 2009). Further, it is suggested that offenders with a central position in the network (having many contacts with others) are the most vulnerable, not only to harmful actions of co-offenders but also to interventions by law enforcement agencies, because they are the "most visible players" (Morselli 2009). Offenders with more strategic but less visible positions are expected to be less vulnerable, because they are often not recognized as the "big man."

The other side of co-offending (positive from the perspective of the law breaker) is that it may lead to higher gains and more success. Many scholars assume that this is usually the case. Tremblay (1993) supposed that co-offending is required for more sophisticated and organized types of offending, Felson (2003) suggests that co-offending leads to the combination of specialized skills and knowledge that make certain offenses within reach of the individual offenders, and Morselli (2009) argues that co-offending leads to "pooling of resources."

Several authors have theorized that it may be particularly interesting to co-offend with more experienced offenders to get "on-the-job training" and learn the profession. Shaw and McKay (1931) already emphasized the importance of learning from experienced others in the development of criminal careers. Recently, Morselli et al. (2006) argued that having a "mentor" may add directly to criminal success in the sense that it may increase the financial gains that are obtained from offending. Having a mentor may lead to "criminal maturation": being committed to a criminal lifestyle, carefully planning offenses and taking effort to decrease the risks. Mentors are also pointing others towards their apprentices as being a "criminal talent", which may offer access to other people and resources.

More generally, many scholars who adhere to a social network perspective assume that access to many potential co-offenders is a key to a successful criminal career. Morselli (2009) argues that the acquirement of new criminal contacts is a leading force behind the criminal careers of major players in organized crime. Success in co-offending may also be acquired by the use of existing social bonds and structures. In this regard, Kleemans and Van de Bunt (1999) emphasized the importance of strong social ties in understanding success in organized crime. Co-offenders who know each other as friends or, even better, as family may have common interests and have access to each other's resources. Further, they may have built trust in each other, because they have a history together and very probably will meet each other in the future again.

An interesting question is which types of criminal contacts are most beneficial to success in co-offending. Recently, several scholars emphasize that strategic positions in criminal networks offer important advantages in this regard. Offenders who connect relatively distinct parts of the network can act as a "broker" and bring different parties in contact with each other. They have unique access to information and opportunities, and with this position they have the potential to provide important services to others and to take initiatives to co-offend themselves (Morselli 2009). More generally, having relatively few redundant contacts (people that do not know each other) is hypothesized to be associated with criminal success. It implies that someone's co-offenders provide a high variety in knowledge and skills, which may lead to more offending versatility and opportunities for criminal gains (McGloin and Piquero 2010).

A next step in theorizing success and failure of co-offending would be to combine the literatures about co-offending and social networks with notions about the risks of co-offending and the "group hazard hypothesis." Further, current thinking about this issue seems to be divided between scholars interested in adolescent group crime and researchers on sophisticated and organized forms of offending. It would be highly interesting to understand when co-offending remains limited to minor and simple forms of illegal behavior, when co-offending enables more serious and sophisticated types of crime, and how simple forms of co-offending may evolve towards serious and successful illegal enterprises.

Final Remarks

After decades of relative silence, interest in co-offending has emerged again since the early 1970s and seems to have increased recently. The last two decades have resulted in several systematic attempts to formulate theories of co-offending or certain aspects of it, and apart from that, an increasing amount of less systematic and explicit theorizing is provided in various empirical studies on co-offending, juvenile delinquency, and organized crime. Nevertheless, in comparison to the large amount of theories on crime and criminal behavior on a general level, theorizing about co-offending is still scarce. What is lacking most are systematic attempts to combine various notions and insights we already have.

Apart from further theorizing, a lot of work is needed in empirically testing the existing theories and hypotheses. Several studies are available yet in which competing perspectives about certain aspects of co-offending are analyzed (e.g., Reiss and Farrington 1991) and in which competing perspectives on co-offending are tested indirectly, for example, by distinguishing peer influence from co-offending (McGloin and Stickle 2011). Yet lacking are explicit attempts to empirically test theoretical notions about fundamental processes behind co-offending directly.

It is important to note that the study of co-offending is not only interesting in its own regard. It may also provide crucial information about the nature of delinquent behavior itself. The three general views on co-offending outlined earlier in this entry (which we can summarize as the group, selection, and instrumental perspective) represent major perspectives in criminology. In this regard, empirical knowledge and further insights in co-offending may offer crucial information about the validity of etiological theories like social control, self-control, and social learning theory and of the usefulness of more situational approaches like routine activities and rational choice theories.

Finally, theoretical insight into co-offending may be useful to improve criminal justice efforts to prevent and intervene in crime (see also the entry about co-offending and offender attributes). As many authors have argued, the impact of co-offending on the total amount of crime may be very large, and preventing offenders to cooperate with each other may offer a substantial contribution to crime reduction (see Andresen and Felson 2010; Felson 2003; Van Mastrigt and Farrington 2009).

Related Entries

- ► Co-offending
- ► Co-offending and Offender Attributes
- ► Gangs and Social Networks

Recommended Reading and References

Andresen MA, Felson M (2010) The impact of co-offending. Br J Criminol 50:66–81

Carrington PJ (2009) Co-offending and the development of the delinquent career. Criminology 47:301–335

Conway KP, McCord J (2002) A longitudinal examination of the relation between co-offending with violent accomplices and violent crime. Aggress Behav 28:97–108

Cromwell PF, Olson JN, Avery DW (1991) Breaking and entering: an ethnographic analysis of burglary. Sage, Newbury Park

Erickson ML, Jensen GF (1977) Delinquency is still group behavior! Towards revitalizing the group premise in the sociology of deviance. J Crim Law Criminol 68:262–273

Felson M (2003) The process of co-offending. Crime Prev Stud 16:149–167

Feyerherm W (1980) The group hazard hypothesis: a reexamination. J Crime Delinq 17:58–68

Hindelang MJ (1976) With a little help from their friends: group participation in reported delinquent behaviour. Br J Criminol 11:109–125

Hirschi T (1969) Causes of delinquency. University of California Press, Berkeley

Hochstetler A (2001) Opportunities and decisions: interactional dynamics in robbery and burglary groups. Criminology 39:737–763

Kleemans ER, Van de Bunt HG (1999) The social embeddedness of organized crime. Transl Organized Crime 5:19–36

Kornhauser R (1978) Social sources of delinquency. an appraisal of analytic models. University of Chicago Press, Chicago

McCarthy B, Hagan J, Cohen LE (1998) Uncertainty, cooperation, and crime: understanding the decision to co-offend. Soc Forces 77:155–176

McGloin JM, Piquero A (2009) 'I wasn't alone': collective behaviour and violent delinquency. Aust N Z J Criminol 42:336–353

McGloin JM, Piquero AR (2010) On the relationship between co-offending network redundancy and offending versatility. J Res Crime Delinq 47:63–90

McGloin JM, Sullivan CJ, Piquero AR, Bacon S (2008) Investigating the stability of co-offending and co-offenders among a sample of youthful offenders. Criminology 46:155–188

McGloin JM, Stickle WP (2011) Influence or convenience? Disentangling peer influence and co-offending for chronic offenders. J Res Crime Delinq 48:419–447

Moffitt TE (1993) Adolescense-limited and life-course persistent antisocial behavior: a developmental taxonomy. Psychological Review 100:674–701

Morselli C, Tremblay P, McCarthy B (2006) Mentors and criminal achievement. Criminology 44:17–43

Morselli C (2009) Inside criminal networks. Springer, New York

Reiss AJ, Farrington DF (1991) Advancing knowledge about co-offending: results from a prospective longitudinal survey of London males. J Crim Law Criminol 82:360–395

Reiss AJ (1988) Co-offending and criminal careers. In: Morris N, Tonry M (eds) Crime and Justice, 10. University of Chicago Press, Chicago

Shaw CR, McKay HD (1931) Report on the causes of crime, vol II. Government Printing Office, Washington

Stolzenberg L, D'Alessio SJ (2008) Co-offending and the age-crime curve. J Res Crime Delinq 45:65–86

Tremblay P (1993) Searching for suitable co-offenders. In: Clarke RV, Felson M (eds) Routine activity and rational choice (Advances in criminological theory 5). Transaction Books, New Brunswick, pp 17–36

Van Mastrigt SB, Farrington DP (2009) Co-offending, age, gender and crime type: implications for criminal justice policy. Br J Criminol 49:552–573

Van Mastrigt SB, Farrington DP (2011) Prevalence and characteristics of co-offending recruiters. Justice Q 28:325–359

Walsh D (1986) Heavy business: commercial burglary and robbery. Routledge & Kegan Paul, London

Warr M (1996) Organization and instigation in delinquent groups. Criminology 34:11–37

Warr M (2002) Companions in crime: the social aspects of criminal conduct. Cambridge University Press, Cambridge

Weerman FM (2003) Co-offending as social exchange: explaining characteristics of co-offending. Br J Criminol 43:398–416

Theories of Punishment

Anthony Ellis
Department of Philosophy, Virginia Commonwealth University, Richmond, VA, USA

Overview

Taking punishment to be the infliction of suffering for an offense, we may say that, traditionally, theories about the justification of punishment have been divided into retributivist and non-retributivist, with further sub-categories in each category. In this entry, we shall take retributive theories to hold that the connection between the aim of punishment and the method used to achieve this aim is conceptual, not causal.

A simple retributive theory holds that the justification of punishment is simply that offenders *deserve* to be punished – for no further reason than that they have done wrong. A different theory is that offenders take an unfair advantage over non-offenders and punishment restores the just balance of benefits and burdens. A third theory also emphasizes the desert of the offender, but holds that what offenders deserve is the opportunity to reform: punishment provides the occasion for this. A fourth type of theory holds that the point of punishment is allow the community to express its condemnation of the offender, or of the offense. Such condemnation might be intrinsically good; or it may be that it is required for the offender to admit wrongdoing and be reconciled with the community.

Non-retributive justifications of punishment usually, though not invariably, hold that much the most important point of punishment is the reduction of crime, and probably the most common way in which it is thought to do this is by deterring potential offenders, sufficiently so as to justify its costs. Quite differently, some forms of punishment also make it impossible, or at least very difficult, for offenders to re-offend (prison and execution being the most obvious examples). A third non-retributive theory, paralleling the retributive version of the reform theory, is that punishment reforms offenders, or at least gives the opportunity for them to be rehabilitated, but, unlike the retributive version, this theory holds that punishment is simply one method among other ways of achieving this aim, a method to be discarded if it is not found to reduce the level of crime. And there is also a non-retributive version of the expressive theory, which also holds that punishment helps to reduce crime, perhaps by educating the population about the wrongness, or illegality, of offenses, or by allowing the vicarious desire for revenge an official outlet, or inducing shame in the offender.

Preliminaries

Punishment we may take to be the infliction of suffering for an offense. We may thus distinguish punishment from restitution and compensation. One is punished for an offense: One is not required to make restitution or to compensate for *an offense* but for the damage that one has caused. (When courts order “punitive damages,” punishment and compensation are being run together.) We may also distinguish punishment from medical treatment. Medical treatment is not administered *for* an offense, though it may be administered *because of* an offense.

The aims of punishment may be many and various: They might include, for instance, propitiating the gods, or trying to make it rain. But this entry will deal with the main aims that have been put forward as justifications for the modern state in punishing its citizens.

So we are concerned here with the classifications of punishment according to what its aims are thought to be. There are, of course, other theories of punishment, such as theories about its historical origins, or its evolutionary purpose, but these theories are not our concern.

But speaking of “aims” too is ambiguous. The word could refer to its sociological function, for instance, but again this is not our concern. We are concerned with the aims that might be thought to *justify* punishment, and to justify it *morally*, not merely *explain* it. The sociological claim that punishment serves some function such as promoting solidarity, for instance, is not typically offered as a moral justification for it, though it may explain its origin or persistence.

Punishment requires a moral justification because when we punish someone we do to them something which, in most other contexts, it would be wrong to do: Punishment involves the infliction of suffering. The suffering might be relatively slight (a small fine) or it might involve serious deprivation of liberty or even of life. The theories we shall be concerned with aim to offer a moral justification for this. There are a few theorists who think that punishment is not in fact justifiable; (Boonin 2008) takes this view, suggesting that restitution should take the place of punishment. Others have suggested that restitution, instead of replacing punishment, should be thought of as its main aim. But most people think that punishment, in its normal understanding, is justifiable by one or more of the traditional theories.

Justifying punishment might be thought to be either, or both, of two different things. First, the focus might be on finding a justification for a particular act of punishing someone. Such an act, it might be thought, could be carried out, properly called “punishment,” and justified, even if there were no penal institutions of any sort. Second, the focus could be on the justification of a penal institution or system itself: Why is it permissible for a state, or a community, to set up a system in which people are actually punished if they contravene some set of norms? And a theory might try to answer both of these questions.

Theories

“Retributivist” Theories

A traditional way of classifying theories of punishment divides them into *retributive* and *non-retributive* theories. It is questionable how precisely theorists have used this classification: Retribution has been defined in lots of ways that are clearly not logically equivalent, and it will be useful to stipulate a definition that captures something like the basic thought. By a “retributive theory” will be meant a theory of punishment which holds that the connection between the aim of punishment and the method used to achieve this aim is conceptual, not causal. All other theories will be called “non-retributive.” So a theory that holds that the aim of punishment is simply to make offenders suffer for their offenses will be retributive: On such a theory, it is conceptually necessary for achieving this aim that one can make offenders suffer for their offenses. By contrast, the deterrence theory is a non-retributive theory: It holds that the aim of punishment is the prevention of crime, but the connection between punishment and the prevention of crime is purely contingent and causal.

This division lines up fairly closely with some other characterizations, such as that retributive theories are “backward looking” (since the justification for punishment lies in what was done in the past), whereas other theories are “forward looking.” But it is not an obvious necessary truth that future actions cannot merit retribution – New (1992) suggests that future acts may legitimately be the object of punishment and of retribution. The common use of the word "consequentialist" to denote non-retributive theories is, however, misleading, because it does not distinguish between causal consequences and logical consequences. The perhaps more common use of the word "utilitarian" to denote non-retributive theories is even more misleading, since most such theories are not actually utilitarian in the proper sense of the word. But there is, of course, no possibility of exactly aligning a precise theory with a range of theories which is not precise.

In any case, "internalist" and "retributivist" will be used interchangeably here, and likewise "externalist" and "non-retributivist".

In popular thought, the idea of retribution typically suggests that a punishment should consist of "eye for eye, tooth for a tooth…", sometimes referred to as the *Lex Talionis* (see, e.g., *Exodus* 21:22ff., *Leviticus* 24:17ff., *Deuteronomy* 19:16ff.). This is not necessarily a retributivist theory, since it does not actually mention an aim of punishment. But in any case, it is inapplicable to most offenses (loitering, speeding on the highway): The phrase is also sometimes used to justify punishments that clearly do not follow its rule, such as cutting off the hands of thieves. And despite the, perhaps, poetic justice in such amputations, they are usually justified by reference to their deterrent effect. Few serious thinkers now regard the *Lex Talionis*, as just described, as a viable theory of punishment, though some defend it at least as an ideal to which punishment should aspire even if it never fully realizes that aspiration (Reiman 1985).

Simple Desert

A more sophisticated attempt to justify punishment on retributive grounds we may call the *Simple Desert Theory*, which simply holds that offenders deserve to suffer *in proportion to* their wrongdoing: Their wrongdoing is some function of the harm that they caused and their culpability in causing it (Moore 1993; Moore 1997; Moore 2009). In principle, perhaps, their deserving to suffer gives anyone the right or the duty to punish them because if they remain unpunished the "universe is morally out of joint." Typically, however, the state takes this role by setting up institutions that are thought to be the most efficient and just way of seeing to it that offenders get the suffering they deserve; but however elaborate the penal system may be, the individual punishment does not derive its justification from the system; the reverse is the case. There is no further explanation of why wrongdoers deserve to suffer: This is simply a basic, foundational moral truth.

Some people think that a punishment and an offense are strictly incommensurable, so that no sense can be attached to the idea that offenders deserve punishment "in proportion" to their offense, so that assigning punishments in this way can be, at best, expressing the arbitrary will of the community. For an attempt to overcome this problem, see Davis (1983). Others find it hard to accept that desert here is a basic notion with nothing further to explain it. Others think that no one *deserves* to be made to suffer (Braithwaite and Pettit 1990, p. 175; Parfit 2010, p. 455). And many think that the notion of simple desert is merely vicarious revenge dressed in the language of justice.

Just Distribution

But this is not the only way that the notion of moral desert can be the basis for a theory of punishment. One might agree that offenders deserve to suffer, but hold that there is a further, more basic, reason for this. That is the position of what might be called the *Just Distribution Theory* and it has been held by many theorists (for two out of a great many proponents, see Morris 1976 and von Hirsch 1976). Here, the general idea is that society requires its members to make certain sacrifices in return for the goods that society makes possible; the members must, e.g., curb their desires to harm other members of society, and this represents a cost to them. The offender, in offending, in not curbing such desires, does not carry his fair share of this burden, and so has gained an unfair advantage over the other members of his society. Fairness requires that this advantage should be taken away, and the suffering involved in a proportionate punishment achieves this. (This theory requires, of course, a functioning society but not, in principle, any set of penal institutions; but, again, questions of efficiency and justice might justify the state in taking over this function).

This theory need not suffer from the problem of incommensurability. It will depend on the precise account of the burden that was not carried in an offense. Merely described as not curbing one's desires might suggest that some form of imprisonment might match this, but not, say, a small fine, and small fines are the majority of punishments administered, even in the USA. There has been considerable dispute among adherents as to how to specify the nature of the wrongdoing so as to overcome this question. We may add that the theory, originally a least, was based upon the so-called Principle of Fair Play. There has been considerable discussion about this principle itself and its applicability to punishment.

Reform Theory

Jean Hampton has argued that the ultimate aim of punishment is the prevention of crime (Hampton 1984, p. 211), but that offenders, like anyone else, can deserve only *good* (in this life, at least), not unproductive suffering. So punishment must be a good to the offender. So long as that is true, then it seems as if the ultimate aim of punishment can only be through reform of the offender. She was also impressed by the thought that punishment, like all human actions, must respect a demanding conception of autonomy and thus not coerce the offender's reform. So punishment must make real to the offender the wrong that he did by making him suffer for it, but it leaves him to decide whether to accept the lesson or not.

But this is not an easy thesis to sustain. There is, as we shall see, a non-retributivist reform theory. However, the claim that punishment reforms offenders, offered as an empirical claim, has little evidence to support it. (The classic work was R. Martinson et al. 1974. The conclusion, based on recidivism, itself part of a larger study using other measures, did not conclude that "Nothing Works," as is often said, but only that nothing had been shown to work.) And indeed Hampton claims that it is a conceptual claim. This seems to rest upon the further claims that the wrongness of offenses consists in their causing suffering (which is surely not true for all offenses) and that offenders do not fully understand the suffering that they cause (which seems doubtful in many cases) and that making them suffer is the only way to get them to appreciate this (which also seems doubtful).

Expressive Theories

In later work, Hampton moved toward an *Expressivist*, or *Denunciatory* theory (Hampton 1992; Murphy and Hampton 1988). An offense, she claims, demeans the victim, does not show him the respect that he deserves, a respect which, as rational beings, we all share equally. So a moral statement has been made, a message sent, that the wrongdoer is superior to the victim, and this has to be effectively countered. Official pronouncements that the offender is not superior to the victim will not do, because the offense still stands and "counts as evidence that he is" superior (Murphy and Hampton, 128). Punishment is all that can do this: In its mastery of the offender, it reestablishes the worth of the victim. And only punishment can do this, because "*any... method*, so long as it was still a method for *defeating* the wrongdoer, *would still count as punishment*" (Murphy and Hampton 1988, p. 125f.) or "retribution" (Hampton 1992, p. 16).

This theory has many moving parts. For instance, that an offense "sends a message" can be true in only the most trivial sense, since few offenses have that even as a subsidiary aim; and even if such a message is sent, it is not clear that it will necessarily be heard, or believed: and the claim that any method for defeating a wrongdoer would count as punishment does not seem clear: when an unjust invader has been expelled from the invaded country, it has been mastered but not punished.

Antony Duff forges a closer connection between the prevention of crime and the expressive function of punishment. While the ultimate aim of punishment is the prevention of crime (Duff 2003, p. 112), this aim must be achieved by repentance and reform on the part of the offender, and reconciliation between the offender and the community (Duff 2003, 107f.; see also Duff 1986). Duff's theory is complex, but at its heart lies the notion of penance. Like Hampton, Duff founds his theory on the Kantian requirement of respect for persons, i.e., respect for rational, moral autonomy (Duff 1986, p. 6). His theory also has built into it a "liberal communitarianism": The legal system should express the community's moral consensus as to what the law should be. He thinks that these requirements preclude a deterrence theory of punishment because deterrence, instead of appealing to the community's sense of morality that underlies the law, is simply a system of threats which create reasons for obedience and play upon the potential offender's fear. The real justification must lie in the goal of reforming the offender and reconciling him with the community. But how is making the offender suffer supposed to achieve this goal? Duff's answer seems to be through the idea of penance. The offender's trial and punishment are an expression of the community's condemnation of the offender's conduct with the hope that this will bring him to repent of his wrongdoing and condemn himself. Duff thinks that punishment is crucial here because punishment is a form of penance which will induce repentance in the offender (Duff 1986, p. 245; cf. Duff 2003, p. 106). Duff candidly admits in his early work that punishment, as we know it, rarely achieves this end and that moral disaster would follow if we were to give it up. In his later work, he favors such things as victim-offender mediation schemes, probation orders, and community service orders. These are not alternatives to punishment, he says: They, like an imposed penance, "more less fit" the standard definition of punishment and should be understood as punishment "even if this requires us to modify conventional understandings of both mediation and punishment" (Duff 2003, p. 97). (Such programs are in fact typically regarded as forms of punishment, since they involve, at a minimum some sort of deprivation of liberty).

There is much to be discussed here, but, as with Hampton, the main problem is the claim, to which Duff steadfastly holds, that it is a conceptual truth that only punishment – the infliction of suffering – can bring about repentance and reform in the offender. The empirical claim that it *can* do this is obviously beyond question: that it does so in a significant number of cases seems doubtful. The claim that it is a conceptual truth and that only punishment can do this seems yet more doubtful.

However, the general idea that a justifying aim of punishment is that it expresses denunciation of the offender or the offense has been very influential.

Andrew von Hirsch, for instance, argues that punishment has two aims, neither of which justifies punishment alone, but both of which, taken together, do so. One aim is the prevention of crime. But the mere prevention of crimes does not, by itself, justify punishment: This would leave out a crucial moral dimension. That dimension is supplied by the fact, as claimed by von Hirsch, that punishing an offender necessarily expresses moral disapproval. And the overt expression of disapproval of those we think have done wrong calls for no justification: It is, conceptually, part of what it is to think that someone has done wrong. To express denunciation *by causing harm*, however, does need justification, and that is presented by the claim that punishment also reduces crime.

Victim Compensation Theory

Last, we should perhaps also mention what are called "victim compensation theories." The idea here is that victims of a crime deserve not just material compensation, but the satisfaction of seeing those who offended against them brought to justice: If being brought to justice involves being punished (as opposed to merely being convicted), then this aim cannot be achieved without the offender being punished. It is therefore an internalist theory. But, as with many theories, there is in externalist version too: If the ultimate aim is to reduce the likelihood of victims using "Lynch law" to satisfy their desire for justice, then it will be an externalist theory (for there is no conceptual guarantee that making offenders suffer is either necessary or sufficient for diminishing the incentives to "Lynch law").

Justice is an impartial idea, and it is not clear that victims have any greater right to see *justice* being done than does anyone else; so the suspicion is strong that we are really talking here about the desire for vicarious revenge.

Non-retributivist Theories

Despite the current popularity of retributivist theories (among philosophers at any rate), it seems likely that the majority of post-enlightenment penal theorists have held that the connection between punishment and its aims is causal and contingent rather than conceptual. Widely different such aims have been ascribed to punishment, but it seems likely that the dominant two are now deterrence and incapacitation.

Deterrence

It is difficult to find explicit theories of punishment before the fourth century BC, though *Joshua*, 17.12 seems to refer to a deterrent aim for punishment; but it is hard to believe that at least part of the aim of the gruesome punishments mentioned in, e.g., the Code of Hammurabi (C17 BCE) was not deterrence. Plato (though in the mouth of Socrates) explicitly accepts deterrent punishment for "incurable" offenders (*Gorgias*, 525b; see also *Protagoras* 323c–324a). Locke also accepted the deterrence theory (Locke 1690, §§ 8, 12; this is what Locke means by "restraint"), as did Beccaria (1764, Chap. 12) and Bentham 1789, Chap. XIII, fn. a.).

In its simplest form, deterrence assumes that potential offenders are roughly rational, economic agents who will weigh the likely costs and benefits of their actions before acting. They will take into account the costs and benefits of all of the actions that are available to them and take that course with the most favorable cost-benefit ratio. The aim of punishment is to make the cost-benefit ratio of prohibited actions less eligible by increasing one or both of the likelihood of conviction and the severity of the punishment if convicted.

It seems clear that most people do not adhere strictly to this simple model. And it is clear that there are many other deterrents to crime than the fear of punishment (the fear of retaliation from victims, socialized aversion to crime, fear of the loss of valued relationships, the possible shame attached to criminality, etc.). So the question for the deterrent theorist is whether the costs of the penal system add sufficiently to other disincentives to crime to justify itself. Answering this question, however, is very difficult, precisely because it is difficult to disentangle the motives that drive people to offend, or to avoid offending. That is all the more so because, it is generally thought, the fact of punishment *itself*, contributes to the sort of informal sanctions just mentioned. For a brief account of the methods used in deterrence research, see (Zimring and Hawkins 1973; Nagin 2000).

Despite these difficulties, most penologists agree that punishment, in tandem with these other factors, and when separated from them, does make some contribution to deterring crime. The classic econometric work was done by Becker (1968). For a more general survey, see also, e.g., Nagin 1998; and Nagin 2000 ("I am persuaded that the collective actions of the criminal justice system exert a substantial deterrent effect", at p. 346). For a more skeptical view about "deterrence-optimizing analysis," see Robinson and Darley 2003 – who do, however, accept that punishment actually does have some deterrent value ("There seems little doubt that having a criminal justice system that punishes violators. . .has the general effect of influencing the conduct of potential offenders. This we concede: Having a punishment system does deter"). For a (perhaps) contrary view, see Meares, Katyal and Kahan (2004) who think the evidence for the deterrent effect of punishment is "speculative and inconclusive" (p. 1186), but it is not clear that they think it has no deterrent effect (see, e.g., p. 1196).

Deterrence is traditionally divided into "special" deterrence (punishment's effect on the person punished) and "general deterrence" (the threat of punishment on the population at large). Strictly speaking, however, this is a confusion. Punishment cannot itself act as a deterrent: Only the *threat* of punishment can do that. That is a conceptual point. And punishing an offender is not usually, in itself, a *threat* of punishment; in very peculiar circumstances, one could perhaps punish someone precisely as a threat of further punishment, but that is not how punishment works in penal systems. All that punishment itself can do is to make the person punished a more deterrable character, by making real to him the nature of the punishment. But making someone a more deterrable character is not itself deterring him: It is a way of *reforming* him. All that can deter him is a further *threat*. So there is no such thing as "special deterrence:" If the person who is punished is thereby made into a more deterrable character, he is then simply a member of the general population to be deterred by "general deterrence." With sufficient publicity, actual punishments may have a similar effect on the general public too, by making the threat of punishment more credible, but that is not "general deterrence" either. The effect on potential offenders of the *threat* of punishment is all that is properly called deterrence.

To the extent that general deterrence works, it works by threats. One might then wonder why it is necessary actually to carry out the threat if the threat fails to deter. The usual answer is that if the threat were not typically carried out, then it would lose its credibility and so become ineffective. So the point of actually punishing offenders, as opposed merely to threatening potential offenders, is to make the threat credible. And this has added to a worry about deterrence that we have already encountered: In actually punishing offenders we are using them as a means to an end, using their suffering to mold the behavior of others. And this does not seem, in normal circumstances, to be morally acceptable behavior.

Incapacitation

The idea here is simple: Punishment, either by execution or incarceration, actually prevents offenders from committing more crimes. But this cannot be a justification for punishment in general, because most punishments, fines for instance, have no significant incapacitative effect. Revocation of driving licenses may be thought to be incapacitative, but they do not prevent anyone from driving, only from driving legally.

The use of prisons, until fairly recently, was not widely thought of as a method of keeping offenders out of circulation. They were mainly used to hold suspected offenders until trial, or to coerce them into paying debts. But transportation, begun in Britain in the reign of Elizabeth I as an extension of hard labor, must surely have also been intended to remove offenders from society. And it exceeds credibility that monarchs and others did not use prisons simply to keep troublemakers out of harm's way (The Tower of London was built specifically to house the king's enemies, not to punish them).

As a justification for imprisonment generally, the idea has simple charm, but is not without problems (see Zimring and Hawkins 1995). For one thing, crime is often committed by groups, and if not all the group is imprisoned then others may take their place in the group. Again, high-ranking offenders can often carry out their criminal purposes from behind bars. And confining offenders is highly expensive, and increasingly so as they get older: There must be some question as to whether the money required for this could not be used in other ways to reduce crime, and to do so without leaving offenders with a stigma makes a continued life of crime difficult to avoid.

Enhanced prison sentences for recidivist offenders are typical in most jurisdictions. But here the problem has been the accurate prediction of which of these offenders would in fact offend again, and the results have not been encouraging. One suggestion has been to try to identify a few offenders who are particularly "dangerous" and confine them to long, perhaps indeterminate, sentences simply as a matter of protection of the public (see Floud and Young 1981). A somewhat less subtle strategy has been pursued in many US states and by the Federal government: to sentence those who have committed a third felony to very long sentences, usually without the possibility of parole. This, of course, typically provides a perverse incentive to those who are committing their third felony not to mitigate the offense.

And there is a moral worry in all of this: Offenders are being confined not for what they have done, but for what it seems likely that they would do if set free. Even if our predictions were accurate, some might still find this morally troubling.

Reform

In its very broadest sense, "reform" may be regarded as any sort of treatment that makes the offender less likely to offend: At the limit, this would, of course, include capital punishment.

More usually, the aim has been simply to change the offender's desires so that he loses the desire to offend. More narrowly still, the aim has been to change the offender's desires while, in some way, not reducing his autonomy. More extendedly, the aim has been to *rehabilitate* the offender so that he can resume a useful life in society.

The reform theory has a long history, dating back at least to Plato, and including Aristotle (*Nicomachean Ethics*, Bk. II, Chap. 1). Jeremy Bentham also thought that reform was one of the main aims of punishment, though less important than deterrence (Bentham Chap. XIII, fn. A). As a matter of policy, reform became the main aim of punishment, imprisonment in particular, during the first half of the nineteenth century, an aim which led to the redesign of prisons so that prisoners were effectively isolated from each other and under constant supervision. Jeremy Bentham's Panopticon became the model for such prisons and many prisons were designed according to its model. But, despite remaining one of the official aims of most penal systems, it has had a mixed history since then, and since the 1970s the lack of any evidence of success has led largely to its demise.

As we have already seen, it is unclear how punishment can reform offenders to any significant degree. Plato thought that wrongdoing was, roughly, a matter of ignorance, and if that were true it would go some way toward answering the question – but, as we have already seen, not very far. In modern times, the reform movement has often been linked with the idea that criminality is some sort of disease, or mental abnormality, but, unless the claim is to be a mere tautology, this can be shown to be true of only a very small proportion of offenders. Starting in the nineteenth century, prisons in particular came to be seen not so much as the reforming vehicle as the *place* where the reform could be carried out, and thus gave rise to the idea of prisons as places where offenders could be rehabilitated. In the last few decades, such programs as probation or community service have been tried with certain sorts of offenders (Ashworth 2005, Chap. 9). There is some dispute about whether such programs are punishments, or alternatives to punishment, but there is in any case little evidence of their efficacy.

Expressive Theories

Just as internalist reform theories have an externalist counterpart, so to do internalist expressive theories. These theories claim that, as a contingent matter, punishment or penal systems reduce crime by sending some sort of message. Most popularly, it has been held that they do so by expressing *denunciation* or condemnation of offenses. This may be thought to reduce crime by educating citizens as to what is acceptable behavior. Or it may be thought to siphon off feelings of outrage which otherwise may lead to social disorder. Or it may be thought that the actual experience of punishment expresses condemnation to the offender, thus inducing shame and the greater likelihood that he will not offend again.

All of these claims are empirical and very difficult to confirm. It seems likely that some such effect operates entwined with such other effects as punishment may have, but it is difficult to be more precise than that.

Assurance Theory

The assurance theory holds that the point of punishment is to assure law-abiding citizens that other citizens will obey the law too (Lacey 1988, 182f.). If they did not have this assurance, it is thought, they would be less likely to obey the law themselves – perhaps, e.g., because they would see the legal system as operating unfairly, or because they would think it foolish to obey laws that others did not obey too, or perhaps because, without that assurance, they would be more likely to take the law into their own hands.

There is no doubt some truth in this. However, this theory is parasitic on one or more of the theories mentioned above: Assurance will be given to law-abiding citizens only if punishment reduces the level of crime, and it can only do that if one of the theories mentioned above (or more than one) is actually correct.

Concluding Remarks

Some of the theories outlined above are easily combined: There is no problem about holding, for instance, the traditional version of the deterrence theory along with the incapacitation theory. Other combinations will produce simple incoherence: One could not coherently hold that the only appropriate response to an offense is punishment and, at the same time, hold that, if punishment does not reduce crime then we should turn to other methods (but see Duff 1986, p. 296ff.). Yet other combinations will not be incoherent but will seem to embody a moral tension, as might be the case if one derived the justification of punishment from a utilitarian desire to reduce crime *and* the demand that we give offenders what they deserve. Whatever the views of theorists may be, however, it seems likely that most people would agree that punishment is in fact justified, and that the justification will consist of more than one of the theories outlined above. But an account of a suitable combination is beyond the scope of this entry.

Recommended Reading and References

Aristotle (1934) Nicomachean ethics (trans: Rackham H). Harvard University Press, Cambridge MA

Ashworth A (2005) Sentencing and criminal justice, 4th edn. Cambridge University Press, Cambridge

Beccaria C (1764) In: Richard B (ed) On crimes and punishments. Cambridge University Press, Cambridge

Becker GS (1968) Crime and punishment: an economic approach. J Political Econ 76:169–217

Bentham J (1789) In: Burns JH, Hart HLA (eds) Introduction to the principles of morals and legislation. Methuen, London

Boonin D (2008) The problem of punishment. Cambridge University Press, Cambridge

Braithwaite J, Pettit P (1990) Not just deserts. Clarendon, Oxford

Davis M (1983) How to make the punishment fit the crime. Ethics xciii:726–752

Duff RA (1986) Trials and punishments. Cambridge University Press, Cambridge

Duff RA (2003) Punishment, communication and community. Oxford University, Oxford

Floud J, Young W (1981) Dangerousness and criminal justice. Heinemann, London

Hampton J (1984) The moral education theory of punishment. Philos Public Aff xiii:208–238, at p 211

Hampton J (1992) An expressive theory of punishment. In: Cragg W (ed) Retributivism and its critics. Franz Steiner, Stuttgart, 1992

Lacey N (1988) State punishment. Political principles and community values. Routledge, London

Locke J (1690) Second treatise of civil government (ed with Intro: Laslett P). Cambridge University Press, Cambridge

Martinson R et al (1974) What works? Questions and answers about prison reform. Public Interest 22:22–54

Meares TL, Katyal N, Kahan DM (2004) Updating the study of punishment. Stanford Law Rev 56:1171–1210

Moore M (1993) Justifying retributivism. Israel Law Rev xxvii:15–49

Moore M (1997) Placing blame: a general theory of the criminal law. Clarendon, Oxford

Moore M (2009) A tale of two theories. Crim Justice Ethics xxviii:27–48

Morris H (1976) On guilt and innocence. Essays in legal philosophy and moral psychology. University of California Press, Berkeley

Murphy J (1979) Retribution, justice, and therapy. Reidel, Dordrecht

Murphy JG, Hampton J (1988) Forgiveness and mercy. Cambridge University Press, Cambridge

Nagin D (1998) Criminal deterrence research at the outset of the twenty-first century. Crime Justice 23:1–42

Nagin DS (2000) Deterrence and incapacitation. In: Tonry MS (ed) The handbook of crime and punishment. Oxford University Press, Oxford, pp 345–368

New C (1992) Time and Punishment. Analysis lii:35–40

Parfit F (2010) On What Matters, vol 1

Plato Gorgias (1925) In Lysis, Symposium, Gorgias (ed. and trans. by Lamb WRM). Harvard University Press, Cambridge MA

Plato Protagoras (1924) In Laches, Protagoras, Meno, Euthydemus. (ed. and trans. by Lamb WRM). Harvard University Press, Cambridge MA

Reiman J (1985) Justice, civilization and the death penalty: answering van den Haag. Philos Public Aff xiv:115–148

Robinson PH, Darley JM (2003) The role of deterrence in the formulation of criminal Law rules: at its worst when doing its best. Georgetown Law J 91:950–1002

von Hirsch A (1976) Doing justice. Hill and Wang, New York

Zimring FE, Hawkins G (1973) Deterrence: the legal threat in crime control. Chicago University Press, Chicago

Zimring FE, Hawkins G (1995) Incapacitation. Oxford University Press, Oxford

Theories on Policing and Communities

Christian Mouhanna
Centre de Recherches Sociologiques sur le Droit et les Institutions Pénales (CESDIP), CNRS, Guyancourt, France

Overview

In most countries with an elaborate theory on policing, any discussion on topics such as "the police" and "the community" is bound to raise the issue of community policing. Although somewhat eclipsed in the early 2000s, the concept remained quite popular among policing organizations, many of which had built upon it in the 1980s–90s to reshape their outreach strategies. The actual meaning of the word "community" across the various countries, however, turns out to be rather loose and inconsistent: While in the USA or the UK, in particular, it is understood spontaneously as a rather obvious piece of common knowledge (Lewis 1969) with a natural impact on policing actions, in other countries such as France, for instance, it has a clearly negative aspect, linked to communitarianism as a perceived threat to the nation state. Beyond local discrepancies, a closer examination of the concept of community from an academic perspective raises complex epistemological issues that reach to the very foundations of sociology and which this entry will not be able to cover. While refraining from delving too deep into intellectual discussions, however, this entry will examine how the emergence of the concept influenced the perception of crime in societies and the subsequent responses by police forces. Indeed, policing actions are literally shaped by how the police perceive the people they are serving and/or controlling. Hence, a given community may be recognized by policemen either as an entity they belong to, or as an intruder in their own community, to be watched or even fought.

After an examination of the concepts and interpretations that tie together communities, social processes, and the modes of regulation of these processes, the link between community and policing will be explored via a literature review. Interestingly, the invention – or reinvention – of community policing in the 1970s–80s owes a lot to scholars – Anglo-Saxon scholars in particular – whose research on police efficiency and impact on people had led them to offer an alternative model to traditional policing. In many countries, especially those that never experienced a close-knit relationship of police forces with the people, such research elicited strong opposition from field professionals. Many a policeman found it hard to accept being challenged, sometimes painfully, from outside their profession. Resistance and reversals were frequent reactions. Why a model that has so often been advocated as ideal proved so difficult to deploy will have to be analyzed and explained. Whatever tensions were triggered by the implementation of community policing may have been underestimated by scholars, even though some studies did eventually tackle the issue later on. As for opponents to this policy, their perspective will be interesting to explore as well.

The Concept of Community, Cities, and Security

In a human and social science perspective, the concept of community was first and foremost meant to describe some form of primarily rural social organization, as described by Tönnies (1887) and the early Max Weber (1948). This matters to us insofar as, in countries where rural and urban policing are organized separately, community policing practices are sometimes nothing more than successful country police endeavors transposed to urban policing. Tönnies (1887) distinguished the community (*Gemeinschaft*) from the society (*Gesellschaft*). To him, the community was a form of organization peculiar to traditional rural life, where people are bonded by family, neighborhood, and religious and custom ties. Community (*Gemeinschaft*) members understand each other and form a closely knit group, organizing their lives together. The *Gesellschaft*, on the other hand, is a by-product of modernity characterized by truly diverse ways of thinking. Far from sharing common values, individuals compete, hence the rise of individualism. Such an outlook generates a "naturally" pessimistic view of modern societies and how to regulate them. As far as this entry's field of study – namely security – is concerned, policing such societies obviously seems more complex and difficult than policing communities, where self-regulation does prevail somehow.

The vanishing agrarian community – understood as a community of interests uniting farm workers and landowners – was a theme that also appeared in Weber's work, albeit in a slightly different sense. Rural exodus was compounded by a refusal to further community ties. Weber's communities were not fossilized (Weber 1971): although rooted in feelings and emotions – *Vergemeinschaftung* –, they also involved a process of rational consultation among individual members – *Vergesellschaftung*. Weber's contribution to the understanding of the agricultural world notwithstanding, his analyses are useful to us for at least three reasons. First because he took research themes that hitherto pertained specifically to rural world communities and applied them to urban communities, thus probably influencing scholars from the School of Chicago. Secondly because Weber is the one who connected and matched social organization matters and legitimation mode issues (Weber ibid.): Modern societies are regulated by law, and this legal, rational legitimacy forms the foundation of the State's authority. And thirdly, within the framework thus designed, the State holds the monopoly of legitimate physical coercion on its territory. By contrast, as can be seen, such lines of reasoning highlight another issue, namely, the tensions that may appear between a given political community, legally endowed with this monopoly, and other types of communities which, although embedded in this larger political community, nevertheless challenge its legitimacy.

Building up on this body of research, and confronting it to extensive field work, the School of Chicago of the 1930s–40s inscribed this concern for the community in its treatment of crime, adding what at the time was a novel dimension: preventing crime and looking for its roots. Scholars working on the Chicago area Joliet prison – Clifford Shaw in particular (Shaw 1930; Shaw and Mc Kay 1942) – were instrumental in the inception of these groundbreaking theories. Following up on the works of Robert E. Park (1952), they identified as the main source of juvenile delinquency the changes undergone at the time by primary institutions – the family, the church, the social group – because of urbanization. Outside the rural context, the socializing power of these institutions, whereby they traditionally induced young people to adopt conventional – or at least noncriminal – attitudes, was being upset by the characteristics of urban life. Scholars found that city life allowed a manifold increase in social exchanges that led to segmented, superficial, and utilitarian relationships, with social ties becoming too numerous, heterogeneous, and instable (Wirth 1938). Faced with the opportunity of making varied acquaintances outside the groups "covered" by primary institutions, young people were less supervised and had more occasions to go astray. Crime and delinquency were thus viewed as a by-product of urban life, as opposed to individual instances of misbehavior. Massive immigration only contributed to reinforce this disruption pattern.

To address such "drifts," the socialization of young people in urban areas was entrusted to schools and police forces – secondary institutions as per the above-mentioned scholars' classification, who considered such institutions as generally unable to tackle the problem. The main option to address this desocialization process, they thought, would have been to restore some social control by helping primary institutions adjust and strengthen themselves. One of the key ideas is that communities should solve their problems on their own, possibly with some help from support groups. Some of these scholars envisioned a radical shift in the role of police forces, who should not restrict themselves to law enforcement, but actively partake in the reinforcement of the capacity of primary groups to socially integrate young people. Social control would thus become the main weapon in the struggle against deviance.

This primarily "environmental" view of crime – and its implications on policing – attracted early and numerous criticisms, including those voiced by Robert Merton or William F. Whyte. The former (Merton 1938) considered that deviance stems from a discrepancy between whatever goals society offers to its members and the means it affords them to reach these goals. For instance, barring adequate educational or social networking resources, underprivileged youths willing to "strike it rich" – one of the major values in modern societies – will be led to seek alternative means of achieving their goal, i.e., they will resort to deviant behavior patterns. In such a framework, deviant behavior patterns are not ascribed to discrepancies between the norms of the various groups embedded in a given society and the ensuing imbalances. Quite the opposite, actually, all members of society share common values and goals, but follow different paths to reach them.

Then, W. F. Whyte showed that while traditional rules inherited from peasant societies do tend to decline when confronted to modern society, new rules and groups emerge as well (Whyte 1943). Whatever anomie scholars may have found is largely attributable to their looking for a model of society that would replicate theirs, thus diminishing their ability to identify rules that do not fit prior knowledge. It should be noted that, in Whyte's Cornerville, local police were integrated to the local society, through corruption and gift practices in particular. According to Whyte, policing can be conceived of in two different ways. The middle class considers that police forces should fearlessly enforce the law and refrain from any preferential treatment, while Cornerville people and the police themselves think that the latter should be trusted by locals so that they can solve problems without arresting anybody. Both perspectives are largely antagonistic: Extreme law-abidance makes it impossible for policemen to act as conciliators, while excessive familiarity stands in the way of their repressive role.

The Impact of Victimization

In the USA, the late 1960s gave new impetus to discussions on the relationships between the police and the community, thanks to a new approach that, instead of focusing on wrongdoers only, started to take victims into account. The idea was to explore the causes of "crime fear" and measure the potential link between such a fear and actual victimization: in other words, to distinguish facts – direct or indirect (i.e., through personal acquaintances) confrontation with crime – from the perceived level of fear of persons who never experienced such a confrontation themselves. So-called victimization surveys, by confronting concerns and actual life experiences, were supposed to help test how relevant this link between fear and actual crime was. More importantly perhaps, they offered an opportunity to check the relevance of existing measures, i.e., to test police statistics. Surveys conducted directly with the population allowed scholars to spot potential discrepancies between statistics provided by policing organizations and their own, later dubbed the "dark figures of crime." This eventually led to questioning the propensity of victims to lodge complaints, their faith in policing actions, or the actual willingness of the police to register complaints. Such interrogations on fear reflect – more or less directly – questionings on the actual role of police forces, how much citizens trust them, and their ability to penetrate the various communities.

Victimization surveys revealed a conspicuous lack of congruence between actual crime and the feeling of insecurity, or more precisely between the number of registered criminal acts – thefts, burglaries, assaults, abductions, murders – and crime fear levels (Ennis 1967; Reiss 1967). This lack of direct causality triggered many interpretations, from the concept of vulnerability (Stinchcombe 1978), allegedly explaining highest fear levels among women and senior citizens, to the linking of crime fear and fear of foreigners – foreign to the neighborhood, or foreign in the sense of not belonging to the same ethnic community –, a frequent staple of research at the time (Skogan 1976). John Conklin established a link between crime fear and the community (Conklin 1975). He opposed Durkheim's idea that crime contributed to bonding the community together by reinforcing social ties between "honest" people (Durkheim 1998). In fact, his field studies rather tended to show that crime produced insecurity, defiance, and a negative view of the community. It seemed to him that crime reduced interactions between neighbors, since fear and suspicion tend to estrange people rather than unite them, at which point social control on deviant behaviors becomes impossible. Faced with crime, people react individually, with fear and withdrawal, rather than in the collective fashion Durkheim envisioned. To Conklin, individual responses to delinquency tend to sever social links, a vicious circle that reinforces fear – including the fear of one's own neighbors – and, as a backlash, fosters criminal behavior patterns, which are encouraged by the resulting lack of social control. In a similar vein, James Q. Wilson highlighted the fact that crime impedes the building of sustainable communities (Wilson 1975). Following up on Tönnies' conception of individualism, he insisted that crime tends to destroy the formal and informal links that bond people to their neighbors, which "atomizes society" and turns all of us into increasingly individualistic schemers. Thus, a shared response to crime would be the only option granting us an opportunity to build or rebuild the community.

Policing is also impacted in this approach: Subjection to crime, when individualized, reduces public support to law and order enforcement institutions. Impacted people are reluctant to press charges, preferring to criticize police activity, or the lack thereof, feeling that the authorities are unable to protect them or prevent crime. Conklin listed several potential collective responses to crime, which included "work[ing] for a political candidate who promises to restore law and order", "call[ing] meetings of community residents", "band[ing] together in a civilian police patrol [. . .]". Residents of the areas most exposed to threats or fear should simultaneously be involved in a reassessment of police work and priorities. Similarly, the fear of foreigners should be tackled by countering the general hostility of the police toward them.

Community and Policing

In the wake of these and similar analyses, the late 1980s saw the emergence – first in the USA, then in Europe – of various concepts and methods aiming both to restore this community bond and profoundly renovate the role of police forces. While in France, any estrangement of the *Police Nationale* from the citizenry was largely attributable to the primacy of law and order in the policing agenda, municipal police forces in the USA mainly had to answer charges of corruption and nepotism. Still, in the 1980s, although their organizational structures and guiding principles differed significantly, both forces conjured up a rather similar imagery of brutality, racism, and social conservatism (Skolnick 1966; Westley 1970; Ericson 1982). Both also reflected a classical phenomenon, pervasive in policing literature, namely, the freedom of action enjoyed by street policemen, who may cherry-pick whatever cases they are interested in. Such discretionary behavior can be held in check by community control (Mastrofski 2004).

In the USA, the rebuilding of communities was thus also dependent on improving policing as a service to the public through far-reaching amendments of its practices, and transforming the relationships between the police and the people they were confronted with. Still, community policing remains hard to define in general terms – after all, adjustment to the local context is supposed to be one of its salient features in the first place. To simplify, one might say that it is more a process and a state of mind than a fixed operating model. *"In a definitional sense, community policing is not something one can easily characterize. It involves reforming decision-making processes and creating new cultures within police departments; it is not a packet of specific tactical plans. It is an organizational strategy that redefines the goals of policing, but leaves the means of achieving them to practioners in the field. It is a process rather than a product."* (Skogan and Hartnett 1997). Lowering the feeling of insecurity is just as much – possibly even more – a goal as lowering the crime rate or actual criminality, but it should be done with the community in mind, i.e., the locals should be involved. Community policing should contribute to improving communication between the police and the people, with the former working to achieve goals fixed by the latter. This general goal can translate into multiple practices, from satisfaction surveys to local meetings, neighborhood watches, activities for the youth, bicycle or foot patrols, etc. The basic idea is to decentralize and scale down policing to increase proximity with the field. Skolnick and Bayley identified four elements that make community policing the future of policing (Skolnick and Bayley 1986): police-community reciprocity, areal decentralization of command, reorientation of patrol and civilianization. Capitalizing on their privileged relationship with the citizenry, policemen are expected to “coproduce” security.

On a related note, Herman Goldstein (1990) noticed that police forces mostly focused on their means – organizational structure, staffing, equipment – as opposed to their objectives, considering that improving or increasing the former must result in an improved quality of service. Goldstein offered an alternative in the concept of “problem-oriented policing,” an operating mode that would put a stronger emphasis on meeting the public’s needs. He particularly insisted on the necessity for the police to respond not just to crime, but also to other problems raised by people, always with a nongeneric, non-stereotyped, customized response. This was meant to deal with minor disturbances, such as noise, which trigger repeated incidents and police interventions, but find no real solution. Besides, problems should not be addressed through legal channels only. Other resources, including partnerships, are available. Generally speaking, it can be said that problem-oriented policing explicitly raises the issue of how to assess police work and its efficiency. Beyond corporatist demands, policemen are encouraged to question their own methods as well as the impact of their action on identified problems.

On this basis, many local police forces in the USA implemented their own, more or less preventive, repressive, wide-scale, or focused on one specific department versions of community policing and problem-oriented policing, eliciting a rather keen interest from the management and active volunteer participation, but with mixed interest at best from many rank-and-file policemen. In Europe too, the problem-solving approach and the idea of closing the gap between the police and the people enjoyed a real, if uneven success, depending on the country. More often than not, in France or the UK, for instance, this rapprochement was made to look like a “back to basics” approach, with the management building up on the image of an idealized policing past – Robert Peel’s bobbies, the Paris *îlotiers* – to convince policemen to establish better relationships with the people.

This drive toward community-oriented policing, however, never took the shape of a unanimous, universal wave. Starting in the 1980s, law and order rhetoric and policies gradually took over in western democracies, so much so that even such an ideal of the iconic English bobby was weakened (Reiner 2000).

Zero Tolerance: A Backlash?

To a certain extent, the so-called New-York model and its corollary, “zero tolerance,” are usually considered to be the antithesis of community policing. To many commentators, only one facet of this model has ever mattered: increased repression whereby each and every illegal act is penalized – the judicial system (or other institutions) even starting to take “acts of incivility” into account. It appears, however, both to promoters of this strategy and to many commentators that this course of action is not entirely at odds with community policing, theoretically at least (Silverman 1999). The New-York Police Department, although sometimes reluctant to admit it, was inspired by a peculiar reading of James Q. Wilson and Georges L. Kelling’s now famous broken windows theory (Wilson and Kelling 1982). While the theory is unanimously recognized for explaining with utmost clarity the absolute necessity of bringing a swift response to any urban phenomenon that could be construed as suggesting that law enforcement authorities are giving up the fight, many beg to differ as per the “coercive” reading of it. Some, including G. Kelling (Kelling and Coles 1996), deny that they ever automatically associated the broken windows theory and the aggressive policing strategy known as zero tolerance. Others, possibly encouraged by further publications by Wilson himself, do however see a direct link between both, and either praise or criticize it. They insist on the fact that the “soft policing” interpretation of this theory neglects the fundamental function of police forces, namely, coercion, while inevitably advocating the existence of unrestrained, as opposed to legally constrained, police forces (Terrill and Mastrofski 2004). The underlying idea, quite popular among police officers, is that community policing stands in the way of their "real" job, which consists in fighting crime. The zero tolerance discourse is well received insofar as it focuses back on this priority.

In practice, the New-York police were seen implementing a policing strategy aimed at certain criminal activities – narcotics trafficking, illegal gun possession, car theft, corruption – while also focusing on street crime and "quality-of-life" policing – coercive police practices targeting street disorder, including prostitution, public alcohol consumption, public urination, littering, panhandling, or even "squeegeeing." Heckling youths were targeted as well. Against such behaviors, reconquering the streets involved arresting, controlling, warning people, and resorting to force whenever necessary. Tolerance zero strategies flourished in the Western world, next to community policing. The popular success of discourses and practices based on this concept – not just with policemen, but also with part of the general population – begs the question of what a community is and how useful the concept can be in the modern, even postmodern, individualistic world where even its definition is problematic. Is it dealing with ethnic, religious, generational, or just territory-based communities? In which case, how can a common policing strategy be designed to meet the needs of all citizens, although their interests diverge – the youth vs. the elderly, those who enjoy an animated life vs. those who long for a quiet environment, shopkeepers vs. people living on the street, etc.? More often than not, community policing may dissimulate some favoritism toward specific social groups – the well-off, the most integrated, those who "most resemble the police" – at the expense of others. As compared to zero tolerance, community policing, instead of fostering a police force able to restore communities and fight anomie, may prove to be nothing more than an ersatz of evolution, some sort of concealment and circumlocution (Klockars 1988). In other words, announcing a shift toward community policing may not be enough to alter the fundamental practices and philosophy of the police. In particular, a zero tolerance policy may turn out to be nothing more than a pretext to increase police pressure on those who are considered uncivil or unruly – beggars, prostitutes, youths occupying public spaces.

"Who does the community police force work for?" is a question that can also be framed in ethnic or racial terms. As pointed out by Skogan (2006) about Chicago, one of the areas where community policing has been most efficiently implemented, it was perceived quite differently in each the "three Chicagos" identified by him: White, African-American, and Latino. In Europe, several countries have faced racial riots that highlighted the urgency of questioning the adequacy of policing with regard to racial issues. In this respect, a comparison of the French riots of 2005 and the English riots of the years 1980–2005 provides a case in point (Waddington et al. 2009). Few policing systems are more dissimilar indeed: Ever since 2002, the French model has been characterized by the rejection of any notion of proximity policing (Monjardet 1996), whereas historically, English police forces are known to have consistently championed community policing. Still, in both countries, young people from ethnic minorities and underprivileged neighborhoods are specifically "targeted" by the police (Bowling et al. 2003). Such groups are always subjected to more surveillance and more zero tolerance policies than others. Besides, it is quite clear that the general attitude of the police toward them has been a key factor in both outbreaks of riots. The triggering factor has often been the arrest and subsequent death of a youth at the hands of the police. Integrating youths from these very minorities in the police forces is an option that has been implemented in some English cities, with disappointing results, especially as far as discriminations with regard to identity checks go. France, however, is adamant that such a differential treatment conflicts with a "republican" tradition holding that no discrimination can be made among French citizens based on ethnic criteria. Theoretically, all nationals have equal legal rights, regardless of their origins. Hence, no community is ever recognized as such. Yet, in practice, several studies did reveal the existence of ongoing discriminatory practices.

The emergence of stringent anti-terrorist policies in the wake of 9/11 and the developing struggle against illegal immigration in most western democracies only accentuated this drive toward the "targeting" of ethnic or religious minorities. At this point, not only do the professional habits of the police incite them to focus their action on minorities and youths from alleged rough areas, but the struggle against illegal immigration officially encourages this. As a matter of fact, illegal immigrants are primarily sought among visible minorities, which reinforces police pressure on these and makes them acutely aware of not quite belonging to the national community. The conscious or unconscious amalgamation of matters such as illegal immigration, terrorism, and incivilities concurs to the gradual estrangement of these ethnic or religious communities. Again, the USA, with their wide range of practices, gives us some food for thought. While many cities responded favorably to the federal policy initiated in 2002 to reinforce the struggle against immigration, many others, among the largest – including San Francisco, Los Angeles, Detroit, Chicago, Houston, Seattle, or Washington – flatly declined to engage in this. Refusing to allow the police to inquire into the immigration status of persons arrested for minor offenses, they started designating themselves as "sanctuary cities" for immigrants (Skogan 2009). The arguments put forward are key to this discussion. When the local police is deeply involved in community policing programs, any mass action against illegal aliens that might jeopardize progress made in terms of cooperation with immigrants will simply be ruled out. Information from these communities can only be accessed on the basis of a trust-based relationship, which in turn rests on a benevolent attitude from the police. The bottom line is that illegal immigrants are an ideal prey for criminals, insofar as they do not have the option of lodging complaints, for fear of being convicted themselves. Besides, because they entered the country illegally, they can only gain illegal access to whatever services, housing, jobs, or usury they might need. They become potential recruits for criminal organizations.

In that case, the presence of a lenient, understanding police force is an asset in the fight against this multiform criminality. From this point of view – and not unusually, as far as policing studies are concerned – establishing privileged contacts that might yield useful, legally exploitable information entails building a mutually beneficial relationship: Using their discretionary, police officers may turn a blind eye on some petty offenses in order to gain access to such information. Which raises yet another issue, one that obviously impacts not just illegal immigrants and cooperating offenders, but the citizenry as well, namely, the increasing judicialization of policing activities, which restrains the discretionary leeway of police officers.

Scrupulous law-abidance is indeed a growing trend among police officers, both because of increased surveillance from police control institutions – magistrates, internal control departments, ombudsmen, human rights NGOs – and because of the higher education officers have received in legal matters, which has improved dramatically. This can only be positive, since police abuse should considerably decrease as a result. There is, however, a downside in that this judicialization induces the police to reject all flexibility in order to abide by the law, which is detrimental not just to their efficiency, but also to their negotiation margins. They are less inclined to adjust their behavior to the members of the various communities they are facing, and refuse to engage in such negotiations as mentioned above, trading leniency for information. Given increasingly restricting modes of management, this legal burden is hampering all attempts at developing any kind of proximity police. While problem-solving policing requires bottom-up operational processes, police management is increasingly top-down orientated.

Avatars of Community Policing

The coexistence of such different logics as community policing and law and order policies has bred hybrid practices that make it difficult to unambiguously interpret their induced effects. A case in point here is intelligence-led policing, a more recent concept that is quite illustrative of the ambiguities produced by the various interpretations of community policing and its confrontation to other logics. Intelligence-led policing was developed in the 1990s when authorities in Great-Britain and Australia were striving for efficiency in both policing and cost-management (Ratcliffe 2002). Even prior to 9/11, it was recognized that the traditional policing model needed to be reformed because it was unable to cope with the major changes induced by globalization. The idea was to turn the police into “knowledge workers” trained in dealing with the risk society (Ericson and Haggerty 1997). This could only be achieved by constantly maintaining high levels of awareness to the environment and the changes it was going through. Next, with the help of analysts organized in support teams, they had to be able to build up an understanding of this changing environment. Finally, they were expected to transmit their findings to decision-makers. The targeted goal was increased efficiency: Analysis is expected to help define clear-cut priorities, not letting the police spread themselves too thin and operate too routinely on day-to-day habits. Efficiency and budgetary requirements are two closely intertwined factors here.

From a theoretical perspective, intelligence-led policing is perfectly compatible with both community policing and problem-solving policing. All three approaches share the common purpose of trying to bypass mundane bureaucratic police procedures and adjust the priorities imposed upon the organization to the ongoing changes of the environment. In practice, intelligence-led policing has helped remotivate a number of policemen who had grown disenchanted of community policing. Indeed, given the lack of enthusiasm elicited by community policing – particularly in the general context of developing terrorist threats and focus on zero tolerance – entrusting a proximity police force with intelligence assignments certainly contributed to enhancing its professional legitimacy. Undoubtedly, promoting rank-and-file community policemen, who occupy the lowest echelons in the informal hierarchy of policing, to the status of intelligence and analysis agents could only be perceived as gratifying.

The discourse on intelligence-led policing, however, is certainly not devoid of contradictions, either intrinsic or relative to the principles of community policing. Its very implementation is fraught with practical difficulties, to begin with. But what is primarily interesting here are the tensions generated by the confrontation of beat policemen, who spontaneously adjust to their environment, with external analysts. This is a classic dilemma of policing, and one which was already covered here: should the policing agenda be set by public demand or by uninvolved analysts who do not in any way relate to the contingencies of police-public interactions? Whether such people belong to the traditional hierarchical organization, as in the classic militarized model, or are mere “technocrats” outside that hierarchy makes no difference from this point of view. Secondly, performance evaluation raises more questions than it can answer, being based on criteria that are all too prone to excesses and manipulations. Policing productivity criteria have lots of pernicious consequences, including in particular the targeting of “small fry,” minor offenders that generate high productivity statistics, instead of the “big fish,” who require more work for a seemingly inferior output.

Finally, to conclude on community policing, it should be noted that recent developments may be taking us toward absolutely novel forms of community policing. In countries with highly centralized, nation-wide policing processes, such as France, local forces are increasingly tasked with proximity assignments, leaving “real policing” to the *Police Nationale*. More generally, communities tend to rely on themselves to assume a regulation role that traditional police forces do not, or not properly, fulfill. In upscale neighborhoods, this leads to the creation of gated communities. In poorer areas, either self-defense groups get organized to protect the community, or criminal groups see to the security of those they consider part of their own community, thus effectively supplanting the police, who are alien to it. Indeed, the concept of “community” has a risky side to it, and can lead to self-imprisonment and withdrawal.

Related Entries

- ▶ Community Policing
- ▶ Democratic Policing
- ▶ Problem-Oriented Policing

Recommended Reading and References

Bowling B, Phillips C, Parmar A (2003) Policing ethnic minority communities. In: Newburn T (ed) Handbook of policing. Willan Publishing, Devon, UK, pp 528–555

Conklin JF (1975) The impact of crime. MacMillan, New York, 1975

Durkheim E (1998[1893]) De la division du travail social. Quadrige-PUF, Paris

Ennis PH (1967) Criminal victimization in the United States: a report of a national survey. US GPO, Washington, DC

Ericson RV (1982) Reproducing order: a study of police patrol work. University of Toronto Press, Toronto

Ericson RV, Haggerty KD (1997) Policing the risk society. Clarendon Press, Oxford

Goldstein H (1990) Problem-oriented policing. McGraw-Hill, New York

Kelling G, Coles C (1996) Fixing broken windows: restoring order and restoring crime in our communities. The Free Press, New York

Klockars C (1988) The rhetoric of community policing. In: Greene JR, Mastrofski SV (eds) Community policing- rhetoric or reality? Praeger, New York

Lewis DK (1969) Convention: a philosophical study. Harvard University Press, Harvard

Mastrofski SD (2004) Controlling street-level police discretion. Annals Am Acad Political Social Sci 593:100–118

Merton RK (1938) Social structure and anomie. Am Socio Rev 3–1938:672–682

Monjardet D (1996) Ce que fait la police, sociologie de la force publique. La Découverte, Paris

Park RE (1929) Human communities. The Free Press, New York

Ratcliffe JH (2002) Intelligence-led policing and the problem of turning rhetoric into practice. Policing Soc 12:53–66

Reiner R (2000) Crime and control in Britain. Sociology 34(1):71–94

Reiss AJ (1967) Studies in crime and law enforcement in major metropolitan areas, Field Survey III, Presidential Commission on Law Enforcement. US GPO, Washington, DC

Shaw CR (1930) The Jack-Roller: a delinquent boy's own story, Chicago. University of Chicago Press, Chicago

Shaw CR, Mc Kay HD (1942) Juvenile delinquency and urban areas. University of Chicago Press, Chicago

Silverman EB (1999) NYPD battles crime: innovative strategies in policing. Northeastern University Press, Boston, [Epilogue, 2001]

Skogan WG (1976) Sample surveys of the victims of crime. Ballinger Publishing, Cambridge, MA

Skogan WG (2006) Police and community in Chicago, A tale of three cities. Oxford University Press, Oxford

Skogan WG (2009) Policing immigrant communities in the United States. In: W Mc Donald (ed) "Immigration, crime and justice", Sociology of crime, law and deviance, vol.13. Emerald Publishing, Bingley, UK

Skogan WG, Hartnett SM (1997) Community policing, Chicago style. Oxford University Press, New York

Skolnick JH (1966) Justice without trial: law enforcement in democratic society. Wiley, New York

Skolnick JH, Bayley DH (1986) The new blue line: police innovation in six American cities. Free Press, New York

Stinchcombe AL (1978) Theoretical methods in social history. Academic Press, New York

Terrill W, Mastrofski SD (2004) Working the street: does community policing matter ? In: Skogan WS (ed) Community policing (Can it work?). Wadworth/ Thomson, Belmont CA

Tönnies F (1887) Gemeinschaft und Gesellschaft, Leipzig. Fues's Verlag, Leipzig

Waddington D, Jobard F, King M (2009) Rioting in the UK and France, A comparative analysis. Willan Publishing, Cullompton, Devon, UK

Weber M (1948) Capitalism and rural society in Germany. In: From Max Weber. Essays in Sociology, translated, edited and with an Introduction by H.H. Gerth and C. Wright Mills. New York, p. 363–385 {Routledge, 1991}

Weber M (1971[1922]) Economie et société. Plon, Paris

Westley WA (1970) Violence and the police, A sociological study of law, custom and morality. MIT Press, Cambridge

Whyte WF (1943) Street corner society, The social structure of an Italian slum. The University of Chicago Press, Chicago

Wilson JQ (1975) Thinking about crime. Basic Books, New York

Wilson JQ, Kelling GL (1982) Broken windows: the police and neighborhood safety. *The Atlantic Monthly*

Wirth L (1938) Urbanism as a way of life. Am J Sociol 44:1–4

Therapeutic Courts

▶ Problem-Solving Courts

Therapeutic Jurisprudence

Barbara A. Babb[1] and David B. Wexler[2,3]
[1]Center for Families, Children and the Courts, University of Baltimore School of Law, Baltimore, MD, USA
[2]International Network on Therapeutic Jurisprudence, University of Puerto Rico School of Law, San Juan, Puerto Rico
[3]University of Arizona, Tucson, AZ, USA

Synonyms

TJ

Overview

Therapeutic jurisprudence, developed in the late 1980s, is a field of inquiry. It is a lens through which to examine the effects of substantive laws, legal rules, legal procedures, and the behavior of legal actors, including judges, lawyers, court personnel, and service providers, on the psychological and emotional well-being of justice system participants, including the legal actors themselves. Therapeutic jurisprudence is a perspective or framework, and its use suggests the need to conduct empirical research to determine whether outcomes resulting from the application of substantive laws, legal rules, and legal procedures and from the behavior of legal actors have therapeutic (helpful) or antitherapeutic (harmful) consequences, both intended and unintended. In addition, therapeutic jurisprudence involves a reform agenda, as it urges that findings from the behavioral and social sciences be used to transform laws, rules, procedures, and the behavior of legal actors in a manner that promotes well-being. This interdisciplinary focus enables therapeutic jurisprudence scholarship and practice to encompass a broad array of subject areas.

From its beginnings as a concept developed in relation to mental health law, therapeutic jurisprudence now enjoys wide and international application in almost every area of the law, including criminal law, family and juvenile law, contract law, tort law, and health law, to name a few. The task of therapeutic jurisprudence in each area of the law is to ensure that therapeutic outcomes can be maximized, while still honoring other justice system values, such as due process. A comparative law approach now is advancing which allows therapeutic practices and approaches from one country's legal system to be assessed in relation to another country, and therapeutic jurisprudence scholarship now exists in many languages.

Although the use of therapeutic jurisprudence principles is best known in the context of problem-solving or solution-focused courts, therapeutic jurisprudence is by no means so limited. The first essay explicating the perspective predates the establishment of problem-solving courts. In addition, from its beginning, principles of therapeutic jurisprudence have been used in general judicial contexts, particularly in criminal cases (Wexler 2005).

Defining Therapeutic Jurisprudence

David Wexler became interested in the notion of law as therapy and of therapy through law as a result of his work in the area of law and mental health. In the summer of 1987, he wrote a paper in preparation for a National Institute of Mental Health workshop. In that paper, which he presented at the workshop in October 1987, he referred to this perspective as "juridical psychotherapy." That term, however, did not survive the meeting, and Wexler thereafter began to use "therapeutic jurisprudence" to define this approach. Wexler's first anthology on the subject was published in 1990 (Wexler 1990).

The late Bruce Winick shared Wexler's interest in mental health law and in the notion of law as therapy. Both were interested in civil commitment, the insanity defense, incompetency to stand trial, and the fact that the mental health system often functioned in a manner that impeded people's recovery (Wexler 2010b [hereinafter Application to Criminal Justice]). The two men became collaborators and codevelopers of therapeutic jurisprudence doctrine and, in 1991, jointly published a book on the subject (Wexler and Winick 1991). For the next two decades, until Winick's untimely death in 2010, these two scholars often worked together. With a growing group of other professionals, they brought psychological insights into the development of legal thought, scholarship, practice, and reform.

Now, after Winick's passing, his legacy continues to be felt. Wexler and a large number of colleagues internationally and across a number of disciplines are developing energetically in the therapeutic jurisprudence field.

Since its origins as an academic approach to mental health law, therapeutic jurisprudence has spread to almost every other area of the law, including criminal law, family and juvenile law, tort law, health law, and housing law, among others. The breadth of its application is evident in a 1,000-page anthology published by Wexler and Winick in 1996, *Law in a Therapeutic Key: Developments in Therapeutic Jurisprudence*. In addition, multitudes of international scholars and practitioners are engaged in the study and practice of therapeutic jurisprudence, as evidenced by postings on and membership in the International Network on Therapeutic Jurisprudence, a therapeutic jurisprudence listserv, a Facebook page, and several international conferences to date on the subject. Many law school course offerings worldwide focus on therapeutic jurisprudence, as referenced in *TJ Across the Law School Curriculum* (Wexler 2012). One United States law school has created a family law and family justice system center, the University of Baltimore School of Law Center for Families, Children and the Courts, whose work is grounded in therapeutic jurisprudence.

Some scholars and practitioners have criticized the use of the word "therapeutic" as "too vague" or "too medical sounding." Winick has suggested that anything related to psychological functioning is therapeutic (Johnston 2012). Others have commented that "jurisprudence" is inappropriate, as they have searched for a comprehensive theory, which they find lacking. Wexler and Winick themselves have encouraged wide scholarly inquiry and debate (Johnston 2012). Nonetheless, the term therapeutic jurisprudence has survived and is thriving.

Therapeutic jurisprudence is a field of inquiry – a research agenda, so to speak. It has evolved from a focus on mental health law to a mental health approach to the law in general. It has contributed to a deeper understanding of the law. Therapeutic jurisprudence is the study of the role of law as a therapeutic agent by serving as a lens that focuses on the law's impact on an individual's emotional and psychological well-being. Therapeutic jurisprudence looks at the law as a social force that can produce therapeutic (helpful) or antitherapeutic (harmful) consequences. These consequences flow from substantive law, legal rules, and legal procedures (the "legal landscape") and from the behavior (the "practices and techniques") of legal actors, including lawyers, judges, court personnel, and others working within a legal context. Returning to Wexler's original idea of law as therapy, therapeutic jurisprudence sees the law itself as the therapist or healing agent. In the same manner as iatrogenic or harmful consequences exist in medicine, law has the potential to produce psychological harm, which Wexler has referred to as "law-related psychological dysfunction" or "juridical psychopathology" (Wexler 1992). Therapeutic jurisprudence instructs that we seek to maximize the therapeutic consequences of the law and its intervention and to minimize its antitherapeutic consequences.

As a field of inquiry, the task of therapeutic jurisprudence is to identify relationships between legal arrangements and therapeutic outcomes. In contrast to the traditional legal analogical process, therapeutic jurisprudence research involves applying a body of social science literature to a body of law and determining their interactions. The ultimate goal of the inquiry is to examine these pairings empirically, and the research has the potential to be interdisciplinary by involving law, philosophy, psychiatry, psychology, social work, criminal justice, public health, and other fields. This empirical understanding then can serve as a basis for policy and law reform (Wexler 1992). In fact, bringing relevant social science research into the legal arena is an important undertaking for therapeutic jurisprudence (Wexler 2001). According to

Wexler, therapeutic jurisprudence is an optimistic and creative perspective, as it tries to identify promising practices from the social sciences and to import those practices into the law. Therapeutic jurisprudence aims to produce tangible, positive change; to promote the well-being of all legal actors; and to improve the justice system so that it is more relevant and helpful for participants and their communities (Goldberg 2005). Therapeutic jurisprudence seeks to determine whether the law can be made, applied, or practiced in a more therapeutic way (Wexler 2010b [Application to Criminal Justice]).

One important question when approaching legal problems and issues from this perspective is whether the goals of therapeutic jurisprudence always should prevail. According to Wexler and Winick, therapeutic consequences should not trump other considerations, such as due process, nor is the approach paternalistic or coercive, as it emphasizes the importance of self-determination and autonomy. Abiding by the notion of therapeutic jurisprudence, if all other judicial and legal issues are equal, the law ought to be restructured to accomplish therapeutic outcomes. The difficulty arises when determining what normative values should take priority. Therapeutic jurisprudence does not answer this (Wexler 1992). It initiates the question and then sharpens and focuses the debate (Wexler 2010b [Application to Criminal Justice]).

Therapeutic jurisprudence has the added value of identifying some of the more subtle, unintended consequences of substantive law and legal rules (Wexler 2010b [Application to Criminal Justice]). It also offers a platform from which to ask and raise questions that otherwise might go unaddressed (Wexler 2010b [Application to Criminal Justice]).

Therapeutic Jurisprudence and Problem-Solving Courts

In August 2000, both the Conference of Chief Justices and the Conference of State Court Administrators endorsed the concept of problem-solving courts and the application of therapeutic jurisprudence by these courts (Wexler 2001). The resolution adopted by both groups focused on methods founded in therapeutic jurisprudence, including the following: the integration of treatment services with legal case processing, ongoing judicial intervention during the case, close monitoring of and responsiveness to a party's behavior, multidisciplinary involvement in a case, and collaboration with community and government organizations (Wexler 2001). The therapeutic purpose of these courts was found in their intent to promote positive behavioral change on the part of the participants (Wexler and King 2010).

Problem-solving courts, such as drug treatment courts, domestic violence courts, prostitution courts, mental health courts, and reentry courts, to name a few, exemplify therapeutic jurisprudence as applied or therapeutic jurisprudence in action. These courts, which now operate internationally, create reciprocal accountability among the judge, the parties, and the services providers that differs markedly from the traditional adversarial roles (Dorf and Fagan 2003). Problem-solving courts seek to identify and address on an individual basis a legal problem's underlying issues or causes of the criminal behavior, such as substance abuse, family violence, and mental illness, thereby promoting positive behavioral, psychological, and emotional change in court participants – all therapeutic outcomes. By focusing on the reasons for offending and by consciously attending to them, a problem-solving approach strives to decrease recidivism and to impede the revolving door of the criminal justice system.

The first scholars to connect therapeutic jurisprudence to problem-solving courts were Peggy Hora and William Schma, along with John Rosenthal, who posited that therapeutic jurisprudence formed the explicit framework for drug treatment courts, which began to operate in 1989 (Hora et al. 1999). Building on their work, most, if not all, problem-solving courts now routinely are associated, implicitly or explicitly, with therapeutic jurisprudence.

While not abandoning the place of therapeutic jurisprudence, Michael King has suggested that problem-solving courts be recast as solution-focused courts (King 2011). King argues that this solution-focused approach relies less on the court's ability to solve a problem than on an individual's self-efficacy and own ability to initiate and sustain positive behavioral change (King 2011). The judge functions as a facilitator of change rather than as the problem-solver, thereby empowering the participant (King 2011). King suggests that this approach also should guide lawyers and other professionals involved in problem-solving courts. He states that the focus on self-determination and intrinsic motivation is more consistent with therapeutic jurisprudence than is the potentially more coercive, paternalistic approach of a problem-solving court (King 2011). Therapeutic jurisprudence acknowledges that the individual must confront and solve her own problems (Wexler and King 2010). King also cautions, however, that therapeutic jurisprudence does not place the focus on the individual above all other justice system values, including the integrity of the justice system (King 2011).

Adopting a therapeutic, problem-solving, or solution-focused approach to legal problems does not mean that judges and lawyers function as therapists or social workers. Instead, it requires that legal actors consciously consider the problems that may have precipitated the criminal behavior and how to effectively address those concerns. Legal actors need to recognize their potential to function as change agents and to acknowledge the impact that their behavior can have on the participants (Goldberg 2005). Through a non-adversarial, team-based approach to an individual's legal problems, the focus of problem-solving or solution-focused courts is on offender compliance with treatment and rehabilitation (Goldberg 2005). To be truly meaningful and therapeutic in these settings, all legal actors' interactions with participants must be characterized by empathy, respect, active listening, a positive focus, non-coercion, non-paternalism, and clarity (Goldberg 2005). In support of these notions, Goldberg provides wonderful examples of behaviors under each of these characteristics in the document she edited for the Canadian National Judicial Institute (Goldberg 2005).

Many problem-solving or solution-focused courts require signed contracts with the offender that outline goals and conditions, along with appropriate rewards and sanctions. In order to maximize the therapeutic outcomes from these contracts, they should adhere to the following guidelines: involve the offender, identify high-risk situations, require the offender to take responsibility for his actions, set specific goals, enumerate specific rewards and sanctions, encourage the participation of family and community members, treat the offender with dignity and respect, and schedule regular review hearings with judicial supervision (Goldberg 2005). Again, Goldberg offers examples of how to accomplish each of these considerations (Goldberg 2005).

The integration of therapeutic jurisprudence into the design and operation of problem-solving or solution-focused courts is paramount. It reminds us that all legal actors must operate from and embody an ethic of care, recognizing that the law truly is a helping and healing profession (Wexler and Winick 2003).

Therapeutic Jurisprudence and Related Doctrines

In the study, practice, development, and reform of the law and legal processes, therapeutic jurisprudence has become partnered with many other doctrines, which has served to enhance the power and relevance of the perspective.

One of the first such pairings was the collaboration between therapeutic jurisprudence and preventive law. The development of a preventive law approach preceded that of therapeutic jurisprudence by many years. Preventive law encouraged lawyers to try to identify potential legal problems early and to address those problems before they developed into true legal issues. It advocated that lawyers anticipate and attend to "legal soft spots," including both "trouble spots," or attempts to avoid legal problems, and "opportunity spots," or areas upon which to focus in order to achieve desired outcomes (Wexler 2010a [hereinafter From Theory to Practice]). Applied to the practice setting, preventive law emphasized the importance of "periodic legal checkups," analogous to regular medical checkups, and the use of the "rewind" technique, or replaying a situation back and determining what could have been done differently to avoid any legal problems encountered (Wexler 1999). Wexler and Winick joined Dennis Stolle, who first suggested the therapeutic jurisprudence/preventive law partnership, to create the alliance (Stolle 1996). According to them, preventive law offered law office practices, including client counseling approaches, and a framework within which to apply the law therapeutically. Therapeutic jurisprudence expanded preventive law's focus on "legal soft spots" to "psycholegal soft spots and strategies," suggesting and justifying that attorneys, judges, and other legal actors consciously address the psychological and emotional issues often accompanying both civil and criminal legal problems (Wexler 2010a [From Theory to Practice]). Therapeutic jurisprudence also enhanced the notion and actions of lawyers as counselors (Wexler 1999).

Another powerful partnership occurred in the family law context, where therapeutic jurisprudence was joined with Urie Bronfenbrenner's research paradigm from the social sciences called the ecology of human development (Babb 1997). This ecological approach to family legal issues and to family justice system reform offered a method to promote consideration of the interaction among individuals, institutions, and the social environment; to assist with a more complete identification of problems; and to contribute to the development of more comprehensive, effective solutions. The ultimate aim of this approach was identified as strengthening the connections among these interactions, institutions, and influences to improve families', children's, and communities' functioning. In order to reform the family justice system, Barbara Babb advocated applying a therapeutic and ecological framework to restructure family law decision-making forums into unified family courts (Babb 1998). She suggested that these courts, if designed and implemented according to her blueprint, could improve families' and children's lives by identifying and addressing holistically all the related legal and nonlegal issues contributing to the underlying family legal problems (Babb 1998).

In a similar vein, Susan Brooks and Robert Madden linked therapeutic jurisprudence with a social science doctrine often applied in the social work context and known as family systems theory (Brooks and Madden 2010). The authors examined the use of a holistic approach to the practice of law through the application of this conceptual framework, similar to the ecological approach described above, and designated by them as "relationship-centered lawyering" (Brooks and Madden 2010). Their paradigm was not limited to the family law context, however, but was applicable to most areas of the law.

Procedural justice, or the focus on legal proceedings and processes themselves rather than exclusively on decisions or findings, also has complemented therapeutic jurisprudence. As Susan Goldberg has noted, for example, therapeutic outcomes for individuals depend largely on their sense of feeling fairly treated, of being accorded respect, of being able to understand the proceedings, and of being heard during the legal process (Goldberg 2005). Goldberg also emphasizes that people who are satisfied with legal proceedings are more willing to comply with courts' decisions and, in turn, have greater trust and confidence in the justice system (Goldberg 2005). As Wexler and Michael King suggest, in the criminal law context, procedural justice must be coupled with some principles and techniques of therapeutic jurisprudence in order to decrease recidivism (Wexler and King 2010).

In the criminal law area, restorative justice often is paired with therapeutic jurisprudence. Restorative justice aims to hold offenders accountable and, at the same time, attempts to address the needs of crime victims and the community at large, as the offense is characterized as harm done to both (Goldberg 2005). Generally, the offender, the victim, and the community engage in an active dialogue to identify a resolution that allows offenders to apologize, make amends, pay restitution, and give back to the community (Goldberg 2005). The aim of this approach, then, is conflict resolution that promotes a healing or therapeutic outcome on the part of all participants, including rehabilitation of the offender and strengthening of the community (Goldberg 2005).

Wexler and King also note that in the criminal law area, therapeutic jurisprudence has adopted a "tripartite" framework (Wexler and King 2010). Under this framework, they urge attention to (1) the pertinent legal landscape or a jurisdiction's substantive law, rules, and procedures; (2) the treatments and services available to defendants; and (3) the possible practices and techniques that the array of legal actors (judges, lawyers, court personnel, therapists) can employ (Wexler and King 2010). Each of these three areas may be either therapeutic jurisprudence-friendly or therapeutic jurisprudence-unfriendly, with the aim being to enhance their affinity to therapeutic jurisprudence (Wexler and King 2010).

Susan Daicoff has included therapeutic jurisprudence as one of the vectors of her legal framework, the comprehensive law movement (Daicoff 2000). This approach to the law is intended to be more holistic and humane.

Critiques of Therapeutic Jurisprudence

Over the years, there have been a number of critiques of therapeutic jurisprudence, well-summarized by Australian barrister and academic Ian Freckelton (2008). Some commentators have questioned whether therapeutic jurisprudence is truly a new approach or whether it is just "old wine in a new bottle" (Freckelton 2008). Similarly, others have argued that the concept of "therapeutic" is problematic and is not distinct enough to differentiate itself from current mental disability law (Arrigo 2004; Petrila 1993; Slobogin). Additional critiques are that therapeutic jurisprudence is covertly paternalistic through involuntary or mandated treatment and that therapeutic jurisprudence attempts to balance many values that are at odds with one another, giving it unclear applicability to judicial decision-making (Slobogin 1995).

Bruce Arrigo argues that therapeutic jurisprudence is too conservative, in that it views the public as homogenous, without recognizing individual and group differences (Freckelton 2008). Samuel Brakel has characterized therapeutic jurisprudence as redundant and not helpful to understand the interface between law and human behavior (Brakel 2007). Brakel also has argued that therapeutic jurisprudence intrudes upon civil liberties by resisting mandated inpatient treatment; therapeutic jurisprudence places some offenders at a disadvantage, as some mental illnesses might be resolved through treatment during brief hospitalization (Brakel 2007).

Freckelton responds to each of these critiques. Generally, however, he notes that many are based on a perspective that "therapeutic jurisprudence has purported to be what it does not claim to be – a coherent body of scholarship with a unified focus that proffers coherent and straightforward answers to complex issues in law and practice" (Freckelton 2008, p. 591). Instead, proponents such as Winick and Wexler merely have claimed that therapeutic jurisprudence can provide a new perspective on the law and that a legal system where therapeutic consequences factor into law and policy when appropriate is preferable to a system in which they do not.

Perhaps the strongest frontal attack on therapeutic jurisprudence and lawyering in the criminal context comes from Mae Quinn. Quinn argues that therapeutic jurisprudence principles should not extend to criminal defense practice – essentially making defense attorneys "rehabilitative change agents" – because doing so undermines the core traditional value of zealous representation. Quinn points to a lack of data showing that traditional criminal defense lawyering has antitherapeutic consequences as evidence that good defense lawyers already effectively take client needs into account in plea-bargaining and defending their clients.

Wexler has responded to Quinn that a defense lawyer who takes therapeutic jurisprudence into account would still exercise due diligence in investigating all possible offenses but would incorporate therapeutic jurisprudence principles as an "add-on." Wexler views therapeutic jurisprudence lawyers as holistic, using psychologically sensitive practice techniques. If, as Quinn maintains, this is simply good lawyering, therapeutic jurisprudence may add a conceptual scheme to examine these practices and to consider other areas of application (Wexler 2008).

As noted above, therapeutic jurisprudence has been criticized for going too far or doing too much. But lawyer/commentator/journalist Mark Satin has criticized therapeutic jurisprudence for not going far enough. Satin argues that therapeutic jurisprudence should be viewed as the basis for an alternative approach to justice, rather than as a lens relative to the existing legal system (Satin 2008).

Future Directions for Therapeutic Jurisprudence and Conclusion

A focus on therapeutic jurisprudence encourages creative thinking in all areas of the law, including practicing, judging, administration, teaching, policymaking, scholarship, and reform, among others. As Wexler indicates, we are called upon to search for promising developments and practices in the behavioral and social sciences, including psychology, psychiatry, criminology, and social work, even if these practices appear to have nothing to do with the law (Wexler 2010b [Application to Criminal Justice]). The task is to determine if any of these promising developments can be introduced into some aspect of or undertaking involving the law (Wexler 2010b [Application to Criminal Justice]).

In the general criminal law context, therapeutic jurisprudence principles have been applied to the granting of probation – especially to the process of crafting probation conditions with which the probationer is likely to comply (Wexler 2008).

Therapeutic jurisprudence also is used in the correctional system, such as in encouraging confined persons to adopt a perspective consistent with a law-abiding future (Wexler 2006) and to the context of parole release (Wexler 2011a [Retooling Reintegration]). A recent article on cognitive interviewing has applied the therapeutic jurisprudence perspective to police investigations (Fisher and Geiselman 2010).

A daily opportunity that has the potential to incorporate therapeutic jurisprudence is at judges' sentencing of criminal defendants who are convicted of some offense. At sentencing, by the way the judge interacts with the defendant, the judge has the opportunity to assist the defendant to confront the wrongdoing and to begin to change behavior (Goldberg 2005). This may include making some type of amends to both the victim and the community. In other words, judges can adopt a problem-solving approach to sentencing based upon restorative justice principles.

Judges can perform another important function relative to therapeutic jurisprudence. Once they have grasped the importance of therapeutic jurisprudence to the justice process and to the work of the courts, judges can draft or assist with the drafting of justice system mission statements that explicitly account for therapeutic jurisprudence (Wexler and King 2010). An example of one such mission statement exists for the Family Divisions of the Maryland judiciary, which has adopted and codified a therapeutic, holistic, ecological approach to family law decision-making (Babb and Kuhn 2003).

One promising practice Wexler details is a "reentry moot court" for incarcerated individuals who are about to participate in the parole process (Wexler 2010b [Application to Criminal Justice]). Through this reentry moot court, the prisoner would participate in a rehearsal of his parole board appearance before a group of incarcerated peers and at least one trained facilitator. This group would assist the prisoner to identify important issues regarding his release and reentry. In addition, the reentry moot court could be useful to the other prisoners as a means to assist them to prepare for their own parole hearings (Wexler 2010b [Application to Criminal Justice]). Practices that could be included in the reentry moot court include restorative circles, modified restorative circles, and relapse prevention programs discussed below.

Hawaii currently offers two "circle" programs for incarcerated individuals. In the restorative circle, the incarcerated person plans for reentry by meeting with loved ones and prison representatives. Together, they create a detailed reentry plan to assist with the inmate's return to society (Wexler 2011a [hereinafter Retooling Reintegration]).

By involving loved ones, this model is a pure restorative justice model. In the modified restorative circle, loved ones are absent, and trained facilitators work with inmates to detail their reentry plans (Wexler 2011a [Retooling Reintegration]). Both programs have had therapeutic effects on inmates' attitudes, including enhancing their feelings of optimism, hope, and readiness to return to society (Wexler 2011b).

Relapse prevention planning, or RPP, is a practice that examines an individual's past behavior in an effort to determine how that individual can move forward in life without further harmful results or effects (Wexler 2011a [Retooling Reintegration]). These programs appear to achieve positive results, such that policymakers are becoming more receptive to incorporating them into the prison setting (Wexler 2011a [Retooling Reintegration]).

A new type of specialized problem-solving court that shows great promise is a child support collection court operating in Colorado. It is a criminal nonsupport docket where individuals appear who are in contempt of court for failure to pay child support. Instead of being called the nonsupport docket, it is called simply "Problem-Solving Court" (Griego 2011). The court team has built a network of services to connect court participants, most of whom are fathers, with substance abuse treatment, free mediation services, vocational rehabilitation, workforce development, GED programs, and fatherhood programs (Griego 2011). Since the court began operating, the number of participants sentenced to jail has decreased dramatically, and the amount of back child support paid has increased substantially.

In an effort to capture all of the innovative practices incorporating therapeutic jurisprudence that judges and attorneys are employing with increasing frequency, Wexler hopes to develop a body of therapeutic jurisprudence case law (Wexler 2010a [From Theory to Practice]). Through this process, therapeutic jurisprudence cases could be continuously gathered for distribution, discussion, and revision or rewind, including suggested improvements for future practice. In addition, it would be possible to create a new body of "practical interdisciplinary scholarship," once there is a method to collect these approaches (Wexler 2010a [From Theory to Practice]).

Wexler also suggests creating continuing education programs on therapeutic jurisprudence for lawyers, judges, and mental health professionals. These programs could include time for input by the group regarding promising therapeutic practices and techniques (Wexler 2010a [From Theory to Practice]).

While law school clinical programs have begun to include the study and application of therapeutic jurisprudence into both their seminar components and into their practice of law, there exists a rich opportunity for even more programs to incorporate this perspective (Wexler 2010a [From Theory to Practice]). Law school clinical programs also represent another opportunity for law students and law faculty to write about their experiences, thereby increasing the available literature about therapeutic jurisprudence as applied (Wexler 2010a [From Theory to Practice]).

Within the law school setting, classes other than clinical programs can continue to include examination of the therapeutic jurisprudence perspective (Wexler 2010a [From Theory to Practice]). As noted earlier, a bibliography currently exists that lists readings on therapeutic jurisprudence for classes across the broad law school curriculum. To maintain and enhance the momentum to include therapeutic jurisprudence within the law school context, legal academics must continue to produce therapeutic jurisprudence scholarship and to speak about therapeutic jurisprudence at professional conferences.

In fact, one therapeutic jurisprudence scholar has suggested that legal scholarship around-therapeutic jurisprudence ought to differ from traditional legal scholarship. He calls for this new scholarship to be shorter, more direct, and more readable, including discussions of law reform issues (Wexler 2010a [From Theory to Practice]).

Michael King suggests that there is a dearth of research about judging in problem-solving courts (King 2011). He believes that there is an urgent need for this research and that the research should focus on the nature and process of behavioral change in individuals. According to King, problem-solving courts are too focused on the practice of lawyering and judging. Instead, King argues that these courts ought to be designed based upon a more comprehensive understanding about how change occurs within individuals (King 2011).

With a similar focus on judging, Wexler suggests that judges who are aware of therapeutic jurisprudence should begin to document appropriate situations or circumstances for its application (Wexler 1999). This documentation then can be shared with other judges. Since judges create the legal culture in their courtrooms, it stands to reason that if judges begin to care about therapeutic approaches to the law, lawyers appearing before them also are likely to attune to this perspective (Wexler 1999). Both judges and lawyers should be aware of what Professor Winick has called the "therapeutic moment" (Wexler 1999).

Wexler also suggests incorporating practices from drug treatment courts, such as graduation ceremonies, judicial praise, and family and friend participation, into ordinary juvenile and criminal cases (Wexler 2001). While these activities acknowledge a participant's progress, they also may contribute to maintaining the participant on a path of positive behavior change (Wexler 2001). The judge has the potential to be a tremendous influence on participants due to her stature and authority, and these activities offer a means for the court to offer true support to participants.

Therapeutic jurisprudence is now a well-established field of inquiry with a large international following and an extremely broad scholarship component. In addition, judges, lawyers, court personnel, and services providers all are adopting and applying therapeutic approaches to their roles and behavior. The most noteworthy example is the effusive and continuing proliferation of problem-solving courts. The potential for therapeutic jurisprudence to continue to enhance and improve justice systems internationally is unlimited.

Related Entries

- ▶ Community Courts
- ▶ Drug Courts
- ▶ Mental Health Courts
- ▶ Problem-Solving Courts
- ▶ Reentry Courts

Recommended Reading and References

Arrigo B (2004) The ethics of therapeutic jurisprudence: a critical and theoretical enquiry of law, psychology and crime. Psychiatry Psychol Law 11:23–43

Babb B (1997) An interdisciplinary approach to family law jurisprudence: application of an ecological and therapeutic perspective. Indiana Law J 72:775

Babb B (1998) Fashioning an interdisciplinary framework for court reform in family law: a blueprint to construct a unified family court. South Calif Law Rev 71:469

Babb B, Kuhn J (2003) Maryland's family divisions performance standard 5.1: a therapeutic, holistic, ecological approach to family law decision making. In: Winick B, Wexler D (eds) Judging in a therapeutic key: therapeutic jurisprudence and the courts. Carolina Academic Press, Durham

Brakel S (2007) Searching for the therapy in therapeutic jurisprudence. N Engl J Crim Civil Confin 33:455–500

Brooks S, Madden R (eds) (2010) Relationship-centered lawyering: social science theory for transforming legal practice. Carolina Academic Press, Durham

Daicoff S (2000) The role of therapeutic jurisprudence within the comprehensive law movement. In: Stolle D, Wexler D, Winick B (eds) Practicing therapeutic jurisprudence: law as a helping profession. Carolina Academic Press, Durham

Dorf M, Fagan J (2003) Problem-solving courts: from innovation to institutionalization. Am Crim Law Rev 40:1501

Fisher R, Edward Geiselman R (2010) The cognitive interview method of conducting police interviews: eliciting extensive information and promoting therapeutic jurisprudence. Int J Law Psychiatry 33:321

Freckelton I (2008) Therapeutic jurisprudence misunderstood and misrepresented: the price and risks of influence. Thomas Jefferson Law Rev 30:575

Goldberg S (2005) Judging for the 21st century: a problem-solving approach. National Judicial Institute, Ottawa

Griego T (2011) Court lends support to dads. *The Denver Post*, September 4, B-01.

Hora P, Schma W, Rosenthal J (1999) Therapeutic jurisprudence and the drug treatment court movement: revolutionizing the criminal justice System's response to drug abuse and crime in America. Notre Dame Law Rev 74:439

John P (1993) Paternalism and the unrealized promise of essays in therapeutic jurisprudence. New York Law School J Hum Rights 10:877

Johnston EL (2012) Theorizing mental health courts. Washington University Law Rev 89:519–579

King M (2011) Should problem-solving courts be solution-focused courts? Monash University Research Paper, Victoria

Satin M (2008) Healing first! Time for the U.S. justice system to get less mechanistic and more compassionate. *Radical Middle*

Slobogin C (1995) Therapeutic jurisprudence: five dilemmas to ponder. Psychol Public Pol Law 1:193

Stolle D (1996) Professional responsibility in elder law: a synthesis of preventive law and therapeutic jurisprudence. Behav Sci Law 14:459

Stolle D, Wexler D, Winick B (eds) (2000) Practicing therapeutic jurisprudence: law as a helping profession. Carolina Academic Press, Durham

Wexler D (1990) Therapeutic jurisprudence: the law as a therapeutic agent. Carolina Academic Press, Durham

Wexler D (1992) Putting mental health into mental health law. Law Hum Behav 16(1):27

Wexler D (1999) The development of therapeutic jurisprudence from theory to practice. Rev Jur Univ Puerto Rico 68:691

Wexler D (2001) Robes and rehabilitation: how judges can help offenders “make good.”. Court Rev 38:18, Spring

Wexler D (2005) A tripartite framework for incorporating therapeutic jurisprudence in criminal law education, research, and practice. Fla Coast Law Rev 7:95

Wexler D (2006) Therapeutic jurisprudence and readiness for rehabilitation. Fla Coast Law Rev 8:111

Wexler D (2007) Therapeutic jurisprudence and readiness for rehabilitation. University of Arizona, Arizona Legal Studies, Tucson

Wexler D (ed) (2008) Rehabilitating lawyers: principles of therapeutic jurisprudence for criminal law practice. Carolina Academic Press, Durham

Wexler D (2010a) From theory to practice and back again in therapeutic jurisprudence: now comes the hard part. University of Arizona, Arizona Legal Studies, Tucson

Wexler D (2010b) Therapeutic jurisprudence and its application to criminal justice research and development. University of Arizona, Arizona Legal Studies, Tucson

Wexler D (2011a) Retooling reintegration: a reentry moot court. University of Arizona, Arizona Legal Studies, Tucson

Wexler D (2011b) Therapeutic jurisprudence and family-friendly criminal law practice. University of Arizona, Arizona Legal Studies, Tucson

Wexler D (2012) Therapeutic jurisprudence across the law school curriculum. Phoenix L. Rev. S. 791–810

Wexler D, King M (2010) Promoting societal and juridical receptivity to rehabilitation: the role of therapeutic jurisprudence. University of Arizona, Arizona Legal Studies, Tucson

Wexler D, Winick B (eds) (1991) Essays in therapeutic jurisprudence. Carolina Academic Press, Durham

Wexler D, Winick B (eds) (1996) Law in a therapeutic key: developments in therapeutic jurisprudence. Carolina Academic Press, Durham

Wexler D, Winick B (2003) Putting therapeutic jurisprudence to work. Am Bar Assoc J 89:54

Winick B, Wexler D (eds) (2003) Judging in a therapeutic key: therapeutic jurisprudence and the courts. Carolina Academic Press, Durham

Third Party Policing

► Legal Frameworks for Third-Party Policing

Third Party Policing and School Truancy

Faye Nitschke[1], Lorraine Mazerolle[2] and Sarah Bennett[1]

[1]Australian Research Council Centre of Excellence in Policing and Security (ARC), Institute for Social Science Research (ISSR), The University of Queensland, St. Lucia, QLD, Australia

[2]Institute for Social Science Research, The University of Queensland, St. Lucia, QLD, Australia

Synonyms

TPP

Overview

Third-party policing is an approach to crime prevention and control involving the police partnering with organizations or individuals to prevent or reduce crime problems. It relies on available civil, criminal, or regulatory rules and laws (legal levers) that enable third parties to take responsibility, at least in part, for control of crime. This entry examines the complexities of the legislative and regulatory framework around the problem of school truancy to illustrate the range of opportunities for using legal levers within a third-party policing context.

Introduction

Third-party policing is a strategy where police collaborate with organizations or nonoffending persons to encourage them to take part responsibility for preventing crime or reducing crime problems (Buerger and Mazerolle 1998). In a third-party policing approach, the police aim to improve crime control, where previous efforts were ineffective, by creating third-party partnerships to enforce the law. These third-party partnerships may be cooperative and involve community members, parents, inspectors, and regulators to convince third parties to take on more crime prevention or control responsibility. Importantly, third-party policing relies on available civil, criminal, and regulatory rules and laws (referred to as legal levers) that enable third parties to take crime-control responsibility.

Third-party policing is an international and growing phenomenon in policing. Mazerolle and Ransley (2005) show that about 50 % of problem-oriented policing (POP) projects utilize at least one third-party policing tactic and occur outside the Scanning, Analysis, Response, Assessment (SARA) policing approach to problem-solving. Problem-oriented policing provides a management infrastructure (see Goldstein 1990) and a stepwise approach to solving a crime problem (Eck and Spelman 1987) within which third-party policing may be one response utilized by police to solve a crime problem (Goldstein (2003). Situational crime prevention provides police with ideas and techniques for reducing crime opportunities (Clarke 1992). In fact, third-party policing can be integrated into any of the 25 techniques of crime prevention proposed by Cornish and Clarke (2003).

When third-party policing is implemented as a crime-control mechanism, the structure of the available legal levers dictates how potential crime-control problems are solved. While third-party policing is merely one of many categories of problem-solving responses or situational techniques, the use of coercive power, from the available legal levers, differentiates the third-party policing approach from other policing strategies (see Buerger and Mazerolle 1998). In third-party policing, the police partnerships with the chosen third parties focus on encouraging or coercing the third party to utilize the legal lever available to them to respond to the relevant crime-control issue (Mazerolle and Ransley 2005).

This entry explores the legal context that underpins third-party policing. Third-party policing occurs within a legal framework that establishes the authority for police to partner with or coerce third parties, the contexts in which they can do this, and the types of interventions this may produce. Arguably, it is this legal basis of third-party policing that both defines it as a unique strategy and distinguishes it from other policing interventions, most notably problem-oriented policing.

The most common legal levers used in third-party policing include local, state, and federal statutes (including municipal ordinances and town bylaws); health and safety codes; uniform building standards; liquor licensing; and a wide variety of statutory programs that use injunctive remedies (e.g., domestic violence restraining orders, gang injunctions, drug nuisance abatement). Importantly, the legal lever may not be aimed at crime control but may instead regulate another form of activity that, if managed, could have a crime-control impact. In fact many legal levers used in third-party policing approaches are not sourced from criminal law. For example, many of the legal levers that can be applied in a third-party policing intervention for truancy come from education law but operate to control crime by improving student attendance and reducing the likelihood that students become either a victim or perpetrator of crime.

This entry examines the complexities of the legislative and regulatory framework around the issue of truancy to illustrate the breadth of opportunities for using legal levers within a third-party policing context. It begins with an overview of the general understanding of the problem of truancy, followed by an overview of a wide range of legal, civil, and informal levers and sources of authority available to support a range of possible interventions in truancy across a range of jurisdictions. Next, the challenges for police in initiating and implementing a third-party policing approach to truancy are discussed, and finally, some conclusions are drawn.

Truancy and Third-Party Policing

Truancy is generally defined as unexplained and persistent absenteeism from school. Research has shown a relationship between truancy and numerous negative developmental outcomes for a young person including school dropout (Alexander et al. 1997), unemployment and social service dependency (Attwood and Croll 2006), and teen pregnancy (Spencer 2009). Research shows truancy is symptomatic and causally related to domestic violence, general family dysfunction (Baker et al. 2001), substance use and abuse (Henry and Huizinga 2007), low socioeconomic status (Baker et al. 2001), and bullying or school violence (Reid 2005). Importantly, truancy is also directly related to, among other antisocial behaviors, antisocial peer influence and drug and alcohol use (Henry 2007). This makes effective intervention for truancy critical for the young person, their family, and the community, as addressing truancy may have a crime-control function. However, the variation in underlying causes of truancy makes systematic intervention for truancy difficult for government agencies.

A third-party policing approach to truancy provides the requisite flexibility for a systematic intervention to effectively address truancy. This flexibility is created by the multitude of differentially focused legal levers available to authorities to use in deterring truancy and encouraging school attendance. While each legal lever offers police an opportunity to use a third-party policing intervention approach in response to truancy, the number of legal levers can make it difficult for police to identify which legal levers will be appropriate and effective for different truancy cases. It can also be difficult for police to select an appropriate third-party partner to work cooperatively in a third-party policing approach to address truancy.

Broadly speaking, legal levers fall within two overarching categories: those that seek to *deter truancy* (taking the “stick” approach) and those that seek to *encourage or normalize attendance* (taking the “carrot” approach). Two kinds of third-party policing interventions for truancy emerge based on this classification of available legal levers. The first is a formal third-party policing approach that aims to deter students from truancy by using coercive or punitive legal levers (deterrence-focused legal levers) in organized partnerships with other agencies as third-party partners. The second is an informal third-party policing approach, with police using legal levers that focus on engagement in school (engagement-focused legal levers) to partner directly with parents as the third-party partners. The next section discusses these two approaches to truancy by exploring the available legal levers in various jurisdictions and explaining their anticipated impact on students’ school attendance behavior.

Deterrence-Focused Legal Levers

Deterrence-focused legal levers, those that aim to stop truanting behavior, generally use punishment or the threat of punishment to induce compliance with a legal mandate. For truancy, the legal mandate is consistent across jurisdictions and requires children attend school on all school days (e.g., Education (General Provisions) Act (Qld) s176(1)). Deterrence-focused legal levers authorize, and thus legitimize, a government agency or department to deal with truancy using some form of punitive sanction. To be properly utilized in a third-party policing intervention, deterrence-focused legal levers generally require police to develop a more formal partnership with the agency authorized by the relevant legal lever to take action in relation to truancy. In this way, the identification of a useful deterrence-focused legal lever will lead police to a natural third party whom they can partner with in order to engage in a third-party policing intervention for truancy. As deterrence-focused legal levers typically involve the third party administering punishment, schools and other government organizations, rather than nongovernment organizations, are authorized to act by deterrence-focused legal levers. Schools and other government agencies have access to less punitive deterrence-focused legal levers, such as notifying parents of their duty to ensure their child attends school, and more punitive legal levers, such as revocation of welfare payments to coerce attendance at school.

School-Initiated Legal Levers

In many jurisdictions schools are deputized by law, principally by legislation and regulations, to administer punitive deterrence-focused legal levers. This makes schools an ideal third party for police to partner with in a third-party policing intervention approach to truancy. Additionally, school staff have regular contact with families and consequently knowledge of the contributing factors to an individual student's truancy. In combination with their legal authority to deal with truancy, by way of deterrence-focused legal levers, this means schools can be a powerful cooperative partner for police in a third-party policing approach to truancy. Some of the commonly available school-initiated legal levers and their deterrent value on truancy behavior will now be discussed.

Mandatory Attendance Recording Most countries, including Australia and the United Kingdom, require schools to keep accurate attendance records (Education (General Provisions) Regulation 2006 (Qld) ss18, 20–21; Education Act 1990 s24; School Education Act 1999 (WA) s28; Education Act 1994 (Tas) s6(3); UK Department for Children, Schools and Families, 2008). Some legislation requires that the school track attendance throughout the day, contact parents promptly in the event of an unexplained absence, and confirm explanations for unauthorized absences that are provided by students (Education (General Provisions) Regulation 2006 (Qld) s20(2)). Schools can also be obliged to report an absence where there is reasonable cause to suspect that nonattendance is associated with a risk of harm to the child (e.g., Education (General Provisions) Regulation 2006 (Qld) s20(6)). In New South Wales, it is an offense for a school principal to fail to keep accurate attendance records (Education Act 1990 (NSW) s24(5)). While record keeping may not appear to have a deterrent impact on children and parents, it has two important deterrent functions. First, regular role calls and following up on absences make families aware that attendance is taken seriously and may trigger voluntary reengagement in school before truancy becomes habitual. Second, accurate records of student nonattendance are an integral component in initiating more punitive legal levers.

Notice of Parental Legal Obligations and Consequences In many jurisdictions, schools are required to inform parents of the obligation for children to attend school regularly if a child begins to truant from school. In Queensland (Australia), schools must notify parents in writing of their duty to ensure that their child attends school and meet with parents to discuss how continued truancy can be resolved (Education (General Provisions) Act 2006 (Qld) s178(2)–(3)). Other Australian States and Territories, New Zealand, and the United Kingdom also require schools to warn parents of the requirement that children attend school. Notice to parents often warns that continued truancy will trigger more punitive sanctions, like prosecution or control orders (e.g., Education Act 1990 (NSW) s22C, Education Act 2011 (NT) s23B). Notification procedures deter truancy by raising awareness of the punitive consequences that may arise if truancy is not resolved. Like attendance recording, notification procedures may trigger a voluntary return to school.

Fines Fines are used to punish and deter individuals who enable or engage in truancy in many jurisdictions, including all Australian States and Territories, parts of Canada and the United States (US), the United Kingdom (UK), New Zealand (NZ), and Hong Kong. Fines are principally administered in response to two kinds of truancy-related offenses: adult employers who employ children during school hours (e.g., Education Act 1990 (Ontario) s30(3), Education Act 1995 (Nova Scotia) s115, Education Act 1997 (New Brunswick) s17) and parents who fail to ensure their child attends school (e.g., Education Act 1990 (Ontario) s30(1), Education (General Provisions) Act (Qld) s176(1), Education Act 1990 (NSW) s23(1)). Generally, fine amounts vary based on the target of the fine. Smaller fines apply for first-time offenses, with larger fines for subsequent or repeat offenses. Fines deter truancy by imposing a punitive financial consequence for enabling truancy and place pressure on adults to enforce mandatory school attendance by ensuring children attend school.

Contracts to Formalize Behavioral Undertakings In some instances, parties who are responsible for improving a child's attendance record are required to formalize undertakings to improve attendance in a written agreement or contract. Some contracts are forged between parents and government agencies to arrange support for a return to school (e.g., Education Act 1990 (NSW) s22B, Anti-social Behaviour Act 2003 (UK) s19). Other contracts are made with any responsible adult in the student's life, including grandparents and community elders (Education Act 2012 (NT) s23B). The different forms of contracts in different jurisdictions implicate different family members in remedying the child's truancy and offer a diverse range of partners for a third-party policing intervention.

Contractual undertakings may require that the parents, or responsible adults, engage in counseling or other intervention for their personal problems that interfere with their supervision of the truanting child (e.g., Youth Justice Act 2011 (NT) s140E). In some jurisdictions, including the Northern Territory (Australia) and the United Kingdom, contracts do not give rise to enforceable legal obligations if breached. These contracts are less coercive, and consequently less of a deterrent, because no punishment arises for failing to adhere to the contract. However, these contracts still have a deterrent function for truancy by consolidating the steps that need to be taken to remedy the truancy in one document. They may also trigger a voluntary compliance with the mandatory attendance requirement by creating a moral obligation on parents and families to comply. Some contracts, like those in New South Wales (Australia) (e.g., Education Act 1990 (NSW) s22B(4)), are admissible as evidence in truancy-related court proceedings. Breach of these contracts may be part of a court's decision to impose a more coercive legal lever. This means admissible contracts have a stronger deterrent impact on truancy because they may initiate harsher penalties for truancy.

Government Agency-Initiated Legal Levers

Less frequently, other government agencies are deputized by legislation or a regulation to use deterrence-focused legal levers to reduce truancy. Other government departments may have access to expertise or services that could assist in the implementation of a third-party policing approach to truancy, positioning them as another powerful third-party partner for police. The legal levers available to these agencies generally focus on removing appealing alternatives to attending school, such as driver's license revocation for truanting minors, or punishment to deter further truancy, such as welfare payment restriction schemes.

Investigation Powers Police officers or appointed "attendance," "truancy," and "liaison" officers are in some instances provided with statutory powers to conduct investigations into suspected instances of truancy to curb nonattendance. These powers may include the ability to question school-aged children seen in public during school hours, to return a school-aged child to school or to the student's home, or take other appropriate action in relation to the child (Education Act 1990 (NSW) s122; Education Act 1972 (SA) ss79, 80A; School Education Act 1999 (WA) s36; Education Act (NT) s23C; Education and Training Reform Act 2006 (Vic) s2.1.10; Education Act 1989 (NZ) s31). Some state police services also conduct "street sweeps" of popular locations for truanting such as train stations, shopping malls, and parks. Investigative powers work in several ways to deter truancy. Investigative actions remind parents and children that attendance is important and that missing school is considered by schools and education departments to be a serious act. Investigations of truancy may also trigger either willing compliance with the school attendance mandate or the application of more punitive legal levers like prosecution.

Welfare Payment Suspensions Currently, the Australian government is trialing suspension of family welfare payments as a truancy response in selected communities in Queensland and the Northern Territory (in the SEAM trial). Social security benefits, pensions, defense force income supplements, and armed forces service pensions can be suspended under the scheme if a child in the care of the payment recipient is not attending school (Australian Social Security (Administration) Act 1999 (Cth) s124D). Payment suspensions can only be initiated after the payment recipient has been issued a warning notice and has been given 28 days to remedy the child's truancy (Australian Social Security (Administration) Act 1999 (Cth) s124K). If truancy is not remedied within the warning period, payments can be suspended for up to 13 weeks. If truancy persists after an initial suspension, payments can be suspended again or ultimately canceled (Australian Social Security (Administration) Act 1999 (Cth) s124M). Welfare payment suspension potentially has a powerful capacity as a deterrence-focused legal lever. The financial consequences of a welfare payment suspension far exceed any fine based in education law in most jurisdictions.

Motor Vehicle License Revocations Some parts of Canada and the United States permit the removal of habitual truants' driver's licenses (Reimer and Dimock 2005). For example, the driver's license suspension regimes in California enable suspension for up to a year (Vehicle Code (California) s13202.7). This type of legal lever can only be used on older students who have a driver's license, making it an ineffective legal lever for younger students. However, license revocation has a strong deterrent capacity as it limits students' ability to engage in an alternative activity that is more desirable than attending school. As the decision to revoke a license is administrative rather than judicial, it can be made more quickly and responsively to habitual truancy than those legal levers that rely on a judicial decision from the courts.

Parent-Targeted Control Orders Some jurisdictions have legislative regimes that permit courts to make control orders that force parents to ensure their child attends school. In New South Wales (Australia), the Children's Court can make a compulsory schooling order over the parent to force the truanting child's attendance at school. These orders can include directives to agencies to provide support to the family so that the order is upheld (Education Act 1990 (NSW) s22D(7)(b)). Parenting orders are also used in the UK and can be issued by the court when a parent refuses to be a party to a parenting contract in relation to persistent truancy on the part of their child (UK Department for Children, Schools and Families, 2008). In New South Wales, the penalty for breaching a control order, by failing to ensure that the truanting child attends school, is four times the penalty for breaching the general parental duty to ensure that children attend school (Education Act 1990 (NSW) s22D(9)(a)). This strengthens the deterrent impact of control orders and makes control orders an excellent legal lever for a third-party policing intervention.

Truanting students can also be prosecuted for nonattendance at school in some jurisdictions, including New South Wales and the Northern Territory (Australia) (Education Act 1990 (NSW) s22D(9)(b), Education Act 2012 (NT) s23B(12)). Control orders may have a large deterrent impact on truancy because unlike other deterrence-focused legal levers, the child can be directly punished for nonattendance.

Engagement-Focused Legal Levers

Engagement-focused legal levers, those that seek to reengage the student in school, typically provide resources and support to encourage attendance. These engagement-focused legal levers, which are chiefly sourced in policy, are comparatively "soft" to deterrence-focused legal levers. Unlike deterrence-focused legal levers, which generally require a legally legitimate third-party agency to operate the lever, by using engagement-focused legal levers, police can partner directly with parents. As parental condemnation of truancy may have a significant influence on a student's behavior (Reimer and Dimock 2005), parents can be an influential third party for police to partner with in a third-party policing approach to truancy. While this kind of third-party policing intervention may appear simple to implement, one challenge of using engagement-focused legal levers is finding a nonthreatening way for police to contact and develop a dialogue with parents. The use of a family group conference to implement a deterrent and engagement third-party policing approach between police, parents, and support agencies is currently being trialed in Australia (Mazerolle et al. 2012).

Awareness Campaigns

The broadest of the engagement-focused legal levers explored in this entry are attendance awareness campaigns initiated by governments. In Australia these campaigns include "Every Day Counts" (Queensland), "It's Not OK to be Away" (Victoria), "Every Child Every Day" (Northern Territory), and "All Starts At School" (Western Australia). Typically these campaigns employ sports stars, celebrities, and local or past student mentors as the face of the campaign to promote the importance of regular and consistent school attendance (e.g., Phillips 2011). They aim to change existing normative attitudes toward school attendance (Queensland Department of Education and Training 2010). Awareness campaigns, by informing parents of the problems associated with truancy, may make parents willing to partner with police in a third-party policing intervention for their child's truancy. This makes awareness campaigns clearly engagement focused.

Early Parental Notification of Unexplained Absences

Another simple engagement-focused legal lever is prompt notification to parents of unexplained absences at school. Notification legal levers facilitate a strong relationship between schools and parents by encouraging schools to engage with parents on a regular basis (Kronholz 2011). Legal levers include immediately text-messaging parents if a child is not at school, liaison officers completing regular home checkups with parents of truanting students (Queensland Department of Education and Training 2010), reporting school and statewide attendance levels in the school newsletter to promote awareness of comparative attendance discrepancies (Victorian Government 2006), and conducting family counseling sessions for at-risk students (Reimer and Dimock 2005). Prompt notification of absences may initiate voluntary reengagement with school before truancy becomes habitual.

Programs for Improving Attendance

Some jurisdictions have programs that offer students incentives for regular attendance. Incentive initiatives include mentoring programs, such as the "Big Brother" program in the United States (Reimer and Dimock 2005), and gifts including money, vouchers, special meals, and excursions. Some programs randomize the award of any incentives to avoid strategic school attendance by habitual truants.

A number of schools use incentives and support to ensure students arrive on time. Breakfast clubs have been used to encourage early arrival at school (Chilcott 2009). Daily pickup services (Queensland Department of Education and Training 2010), free bus passes, and wake-up calls (Kronholz 2011) have also been used to facilitate a timely arrival at school. Difficulty with transportation to school has been identified as a cause of truancy (Reid 2005) so these legal levers engage students in school by resolving an underlying cause of truancy.

To make reintegration into the classroom after truancy easier, some programs seek to engage students in school by providing academic support for classes missed. These programs include homework clubs, tutoring, or makeup classes (Victorian Government 2006). These engagement-focused legal levers should assist return to school where academic problems contributed to the student's decision to truant.

Flexible Schooling Arrangements

Flexible arrangements authorize schools to develop alternative methods of attending for students who face difficulties attending standard school hours (e.g., School Education Act 1999 (WA) s24). Flexible arrangements can be used, for example, for older students who have caring responsibilities or young children at home. Under the legislation in Queensland (Australia), arrangements are developed by the parent and can be authorized by the relevant teacher (Education (General Provisions) Act (Qld) ss182, 183). Flexible arrangements encourage engagement with school by removing pragmatic barriers to regular school attendance experienced by some students.

Community Partners and Group Action

Although group action by community partners generally is intended to deter students from truanting, frequently this kind of third-party policing intervention also engages the community to proactively mobilize against truancy. Group action in response to truancy generally involves community members taking action to reduce the availability of activities that students may engage in instead of attending school. An excellent example of group action is the “No School, No Service” program in the Northern Territory, in which business owners agree to refuse to serve school-aged children during school hours, providing a strong disincentive to truant. In these group interventions, the community, unlike a parent, is not legally obligated to play a role in the enforcement of the legal mandate that children attend school. Group action stems from a decision to stop enabling truancy, even though the community is not legally obligated to be part of the enforcement of the mandatory school attendance requirement.

Initiation and Implementation of a Third-Party Policing Intervention

The relative effectiveness of a third-party policing approach to truancy is shaped by how the approach is initiated and implemented. Specifically, the choice of the focus of the third-party policing approach in initiating third-party policing partnerships and the role of the police within an established third-party policing partnership can have significant impact on whether the third-party policing approach to truancy is effective. These initiation and implementation issues are discussed in this section.

Initiating Third-Party Partnerships

In a sense, the success of a third-party policing approach to truancy is dependent on the willingness of police, the third party, and the target to comply with the legal lever used in the intervention. The burden is on police to initiate the partnership process and make the decision about the focus of the third-party policing intervention. These choices can, at first glance, appear difficult in the face of the multitude of legal levers available to address truancy. For each truancy case, the police will be faced with numerous engagement- and deterrence-focused legal levers that they could use in a third-party policing intervention. In a way, the core decision to be made for a successful third-party policing approach is whether to use deterrence- or engagement-focused legal levers. By virtue of the way the law functions in many jurisdictions, the legal lever identifies to the police the target of a third-party policing approach. Third-party policing interventions for truancy, based on the available legal levers discussed in this entry, are mostly targeted at either parent or child.

If a deterrence-focused third-party policing approach is chosen, in most cases the decisions about third-party partners and intervention targets will flow naturally from the choice of legal lever. As discussed, deterrence-focused legal levers, sourced most frequently from legislation and regulations, will typically legitimize a third-party agency, who naturally becomes the third-party partner in a third-party policing approach that uses that legal lever. Deterrence-focused legal levers of a punitive nature will also always identify a target for the third-party policing intervention, which in the case of truancy may not always be the child. This significantly reduces the complexity of implementing a third-party policing approach for police. For example, if the police elect to use a deterrent approach in response to truancy, they may use a parental prosecution legal lever. In most jurisdictions, the legal lever for parental prosecution for the nonattendance of a child will authorize the school or education department in that jurisdiction to engage in the prosecution decision-making process. In making the decision to use a deterrent approach and use the parental prosecution legal lever, the police have already identified a target for the third-party policing intervention, the parent of the truanting child, and a third-party partner, the school or education department.

A similar relationship between legal levers, third-party partners, and targets exists in relation to engagement-focused legal levers. Although engagement-focused legal levers tend to be sourced in policy and thus do not identify a natural third party for police to partner with, most engagement-focused levers have a clear target. In most cases, the target of an engagement-focused legal lever is the child. This is the case with truancy awareness campaigns, incentives, and programs that support attendance, flexible schooling arrangements, and typically community action. In focusing on the child, the parent or caregiver can become a natural choice of third-party partner for police to use a proactive engagement-focused lever to facilitate their child's regular attendance at school.

Once the partnership has been established, police play a critical role in implementing the third-party policing approach to truancy. Importantly, police will often need to provide the third-party collaborator with encouragement and support, to ensure that the third party initiates the chosen legal lever to cause change in the behavior of the intervention target. This is especially true when a third-party policing approach to truancy is used, as legal levers that respond to truancy are often infrequently used. For this reason, it is important that the police remain an active party to the third-party policing partnership to implement the third-party policing approach to impact truancy.

Maintaining Third-Party Partnerships

Timely and consistent activation of the chosen legal lever is critical to the success of a third-party policing approach to truancy. Many engagement- and deterrence-focused legal levers are infrequently used to respond to truancy. Within a third-party policing approach, activation of the chosen legal lever is the focus of the intervention. Ultimately, an infrequently, or unapplied, legal lever will not change truancy behavior. Thus police need to address any third-party reticence to utilize the chosen legal lever within the third-party policing partnership if the intervention is to be effective.

Deterrence-focused legal levers, and in particular prosecution-based legal levers, seem to be particularly infrequently used in relation to truancy. Determining the number of parent- and child-focused prosecutions that take place is difficult because in many jurisdictions the prosecutions occur in lower courts, many of which do not keep record databases of their decisions. Australian states seem reluctant to address truancy by parental prosecution. News reports indicate in April 2009 seven parents were "facing" prosecution in Queensland, with an eighth being referred to the police for action (Chilcott 2009). In South Australia, 24 cases were brought between 1984 and November 2010, only six of which were successful (Keller 2011). Just six parents were prosecuted in Tasmania between 2005 and 2009 (Dickson and Hutchinson 2010). In Western Australia and the ACT, truancy prosecutions were commenced for the first time in February 2010 and 2009 respectively (Dickson and Hutchinson 2010). It is reported that Northern Territory and Victorian truancy laws have never been enforced (Dickson and Hutchinson 2010). In contrast, New South Wales initiated 310 prosecutions in 2010 against parents or children for persistent school nonattendance (McDougall 2011).

The United Kingdom regularly prosecutes parents for failing to ensure their children attend school (Morris 2009). In contrast, enforcement is anecdotally lax in the United States (Kronholz 2011), and in New Zealand there were two prosecutions recorded in 2011 (Fisher 2011) and only 17 between 2004 and 2007 (Rich 2007).

Another related and important issue is how prosecution is initiated. The number of notification steps that must be taken prior to a prosecution, fine, welfare payment suspension, or control order in most jurisdictions precludes immediate use of a deterrence-focused legal lever in response to truancy. In fact, the policy guidelines of some jurisdictions, including Queensland, Tasmania, and New Zealand, specifically identify prosecution as a last resort measure (Chilcott 2009; Dickson and Hutchinson 2010; New Zealand Ministry of Education 2010). Although appropriate consideration before using deterrence-focused legal levers afforded by longer escalation processes is important, delays in the application of a legal lever after truancy occurs may weaken the deterrent power of the legal lever. This is recognized in some jurisdictions, like the United Kingdom, where local authorities are able to override escalation prerequisites to fast track prosecutions in cases where this is considered appropriate (UK Department for Children, Schools and Families, 2008).

The negative institutional attitude toward prosecution and other punitive deterrent legal levers reflects uncertainty in the assumptions that underpin the use of punishment in response to truancy. More specifically, there seems to be a belief that punishing parents for truancy unfairly assumes parental failure causes truancy. Some also object to truancy relevant fines because they disadvantage families with the smallest disposable income.

In a third-party policing approach, possible negative attitudes toward using legal levers, particularly those legal levers that are deterrence focused, can be counteracted by police encouraging and supporting the third party to use the legal levers to address the truancy problems. In this way, a third-party policing approach could make the most effective use of preexisting legal levers to reduce truancy and consequently the prevalence of antisocial behavior.

Concluding Comments

The legal context within each jurisdiction informs the way a third-party policing intervention for truancy is initiated and implemented. The differentially focused legal levers across jurisdictions provide flexibility to address the root causes of truancy, streamline the decision-making process for police determining how to implement a third-party policing approach, and assist police in defining their role within a third-party policing partnership to achieve effective intervention for truancy. Importantly, the various legal levers available offer police a plethora of opportunities to engage in a third-party policing intervention for truancy, which may have a crime prevention impact for the community. Third-party policing offers police the opportunity to address crime prevention effectively by utilizing preexisting legal levers and third-party partnerships in a resource responsible manner.

Related Entries

- ▶ Civil Remedies
- ▶ Cost-Benefit Analysis
- ▶ Family Engagement Strategies to Reduce Crime
- ▶ Juvenile Justice in the Get Tough Era
- ▶ Legal Frameworks for Third-Party Policing
- ▶ Pulling Levers Policing

Recommended Reading and References

Alexander KL, Entwisle DR, Horsey CS (1997) From first grade forward: early foundations of high school drop-out. Sociol Educ 70:87–107

Attwood G, Croll P (2006) Truancy in secondary school pupils: prevalence, trajectories and pupil perspectives. Res Pap Educ 21:467–484

Baker ML, Sigmon JN, Nugent ME (2001) Truancy reduction: keeping students in school. Juv Justice Bull September:1–15

Buerger ME, Mazerolle L (1998) Third party policing: a theoretical analysis of an emerging trend. Justice Q 15:301–328

Chilcott T (2009) Truancy parents set to be prosecuted. Courier Mail. Retrieved from http://www.couriermail.com.au/news/parents-face-truancy-action/story-e6freon6-1225699228733

Clarke RV (1992) Situational crime prevention: successful case studies. Harrow and Heston, New York

Cornish D, Clark R (2003) Opportunities, precipitators and criminal decisions: a reply to Wortley's critique of situational crime prevention. In: Smith M, Cornish D (eds) Theory for practice in situational crime prevention, vol 16, Crime prevention studies. Criminal Justice Press, Monsey, pp 41–96

Dickson E, Hutchinson T (2010) Truancy and the law in Australia: the Queensland example. Int J Law Educ 15:87

Eck JE, Spelman W (1987) Problem solving: problem-oriented policing in Newport News. Police Executive Research Forum and National Institute of Justice, Washington, DC

Fisher A (2011) More prosecutions urged to halt truancy. The Dominion Post. Retrieved from http://www.stuff.co.nz/dominion-post/news/5161027/More-prosecutions-urged-to-halt-truancy

Goldstein H (1990) Problem oriented policing. McGrawHill, New York

Goldstein H (2003) On further developing problem oriented policing: the most critical need, the major impediments and a proposa. In: Knutsson J (ed) Mainstreaming problem oriented policing, vol 15, Crime prevention studies. Criminal Justice Press, Monsey, pp 13–48

Henry KL (2007) Who's skipping school: characteristics of truants in 8th and 10th grade. J Sch Heal 77:29–35

Henry KL, Huizinga DH (2007) Truancy's effect on the onset of drug use among urban adolescents placed at risk. J Adolesc Heal 40:9–17

Keller C (2011) Parents face court on truancy crackdown *The Advertiser*. Retrieved from http://www.adelaidenow.com.au/news/in-depth/parents-face-court-on-truancy-crackdown/story-e6frebvu-1225959253748

Kronholz J (2011) Truants: the challenge of keeping kids in school. Ed Next 11:32–38

McDougall B (2011) Targeting truants, teens face court. Daily Telegraph. Retrieved from www.dailytelegraph.com.au/targetting-truants-teens-face-court/story-fn6b3v4f-1226085427823

Mazerolle L, Ransley J (2005) Third party policing. Cambridge studies in criminology series. Cambridge University Press, Cambridge

Mazerolle L, Bennett S, Antrobus E, Eggins E, Nitschke F (2012) Project ABILITY information pack. Australian Research Council Centre of Excellence in Policing and Security, University of Queensland, Brisbane

Morris E (2009) To tackle truancy, we need consistency. The Guardian. Retrieved from http://www.guardian.co.uk/education/2009/feb/24/truancy

New Zealand Ministry of Education (2010) Improving attendance: case management of truancy and the prosecution process. Retrieved from http://www.minedu.govt.nz/~/media/MinEdu/Files/Boards/Support/ImprovingAttendance2010.pdf

Northern Territory Government (2012) Every child, every day: the strategy 2010–2012. Retrieved from http://www.det.nt.gov.au/__data/assets/pdf_file/0012/15141/EveryChildEveryDayStrategy.pdf

Phillips Y (2011) Fresh new face to anti-truancy campaign. Perth Now. Retrieved from http://www.perthnow.com.au/news/western-australia/fresh-new-face-to-anti-truancy-campaign/story-e6frg14u-1226060492934

Queensland Department of Education and Training (2010) Every day counts policy. Retrieved from http://education.qld.gov.au/everydaycounts/

Reid K (2005) The causes, views, traits of school absenteeism and truancy: an analytical review. Res Educ 74:59–82

Reimer M, Dimock K (2005) Truancy prevention in action: best practices and model truancy programs. National Dropout Prevention Centre Network, Clemson

Rich K (2007) Truancy prosecution rates pathetic. Scoop NZ. Retrieved from http://www.scoop.co.nz/stories/PA0703/S00289.htm

Spencer AM (2009) School attendance patterns, unmet educational needs, and truancy: a chronological perspective. Remedial Spec Ed 30:309–319

UK Department for Children, Schools and Families (2008) Ensuring children's right to education: guidance on the legal measures available to secure regular school attendance. Retrieved from https://www.education.gov.uk/publications/eOrderingDownload/DFES-00529-2007.pdf

Victorian Government (2006) It's not okay to be away: attendance strategies and sample policies. Retrieved from http://www.education.vic.gov.au/healthwellbeing/wellbeing/attendance/initiative.htm

Third-Party Policing

Michael E. Buerger
Bowling Green State University, Bowling Green, OH, USA

Overview

The original articulation of third-party policing (Buerger and Mazerolle 1998) focused on coercive tactics police use against nominally respectable citizens to impel them to join anticrime efforts. The underlying premise was that the semiformal and informal social control exercised by "place guardians" (Spelman and Eck 1989; Eck 1997; Buerger 1993) was greater overall than that of police intervention. More personal and more consistent over time, it could be more immediate in its consequences (such as denial of entry or eviction for cause). If tapped and coordinated with an overt police effort, it had greater potential for stabilizing place-related behaviors.

Mazerolle and Ransley (2005) extended the concept further, outlining the benefits of formal coordination between police and governmental agencies with specific mandates (oversight over pharmacy sales or alcohol beverage control). Joint endeavors to prevent abuse of prescription drugs (and over-the-counter medicines containing precursor chemicals) were differentiated by specific legislative charges in three Australian states, creating a natural experiment highlighting different levels of authority, commitment, and engagement with the goals of the supposedly joint project. Those considerations also attend local efforts and can be even more pronounced when third-party policing is attempted in nonregulatory environments.

The most salient limit of third-party policing is that it constitutes a *coerced volunteerism*, a response attained by implied threats when appeals for truly voluntary actions fail. The police power to overcome resistance through physical force in criminal matters (Bittner 1970) does not extend to third-party policing, which occurs almost exclusively in the realm of civil law. The proscribed actions police seek to prevent are not those of the recruited citizen but the conduct of their customers, neighbors, or even passersby. Coerced volunteerism imposes an enforcement-like role with no perceptible foundation in law or capacity. Initial resistance should be anticipated in such cases: indeed, one of the fundamental demands of third-party policing is that of finding ways to overcome such resistance.

Coercion is most effective against those who are already vulnerable by dint of playing fast and loose with regulations that define their basic responsibilities, overseen and enforced by nominal allies of the police. Implied coercion is less able to compel individuals in compliance with standard regulations, and those who are accustomed to gaming the existing system. It is most tenuous when the targets are themselves sovereign entities or parties who have already forged alliances with nominal regulators, politicians, and other sovereigns who can intercede on their behalf.

Coerced volunteerism invariably requires actions and expenses outside the normal scope of the third party's life or business. Whether a one-time expense such as the construction of a barrier or a continuing commitment to regulate the conduct of patrons, an obligation is imposed that disrupts the routine activities of the enlisted individual or organization. Some requested efforts may be perceived as injurious to business interests, even dangerous to the participants. Unlike volunteer efforts like Neighborhood Watch, third-party policing does not mobilize any inherent desire or willingness to "do something" about a mutually agreed-upon noxious condition; it is far more likely to be a second-tier enlistment for a police-defined operation, after the softer approaches have been exhausted. Unless the benefits of the requested actions are recognized, resentment and resistance must be anticipated.

Moreover, "recognition of benefits" is not unidimensional. In the best case, the third party either understands intuitively the link between the police request and an improved quality of life or experiences the benefits soon after implementing the requested actions. Much more likely is the probability that the only "benefit" perceived is to be relieved of the annoying presence of the police officers repeating their requests. At the lower end of the scale, the third party may interpret the "benefit" of cooperation to a *quid pro quo* for later favors or leniency from the police.

The limits and boundaries for third-party policing initiatives lie in two distinct domains: among the potential third-party partners of the police and within the police organization itself. Both domains are also vulnerable to environmental influences affecting one or both. Periods of economic constriction not only place severe financial constraints on landowners and businesses but also can redirect police resources personnel from efforts viewed as peripheral. Unless clear benefits can be shown, both agency and political overseers will demand that resources be directed back into "meat and potatoes" channels: traditionally, that has meant staffing patrol to answer calls in the old reactive policing mode.

Several impediments – notably “innovation fatigue” and “limited subscription” – are found in different manifestations in both camps. Though linked in some arenas, they are conceptually distinct in important ways. Nevertheless, the different aspects of individual third parties and organizational third parties are presented together here. Greater differences exist between external organizational resistance and internal police agency resistance, and those are treated separately. Regardless of the class, resistance tends to manifest itself in the distinct forms outlined below.

Innovation Fatigue. Institutional history in police departments (and others) often includes a compilation of highly touted but short-term efforts to “do things better.” Even successful innovations often fail to achieve the critical level of acceptance needed for adoption as a regular practice. While not unique to police organizations, the kiss-of-death pronouncement that “We tried that once; it didn’t work” is recognized by police researchers as the most visible legacy of promoting innovation in policing. (The police see this differently: researchers “arrive with a great idea they’ve dreamed up in the ivory tower, make a mess for a year, and then disappear for 3 years before they come back to tell us what we did wrong.” That pronouncement embodies the perception that externally imposed innovation almost always fails to capture important realities of the practical level.)

A comparable condition may affect community members and partner organizations. Considerable publicity accompanies the onset of any new program, conveying to multiple constituencies both the dimensions of the project and its needs. People invest in the new idea, make the necessary arrangements to participate (disrupting or sacrificing other parts of their routine to do so), and expect to see results. Often those results do not materialize, and those who are most affected by such failure are the community members who stood to benefit from a successful project.

Some programs simply do not achieve the anticipated results, either because of theory failure or because of implementation failure. Some cannot garner the critical threshold of resources to be effective. Others simply fail to materialize for reasons that are never made clear. The cumulative effect is the modern-day equivalent of Aesop’s fable of the boy who cried “Wolf!” Each new iteration of change becomes less and less compelling, until finally people are unwilling to budge. History tells them that nothing good will emerge, no matter what they do.

Uneven Subscription: Whether in organizations or by individual citizens, uneven subscription can occur in one or more distinct manifestations: overt resistance, covert disruption (“monkey-wrenching”) and lip service masking “this too shall pass” passivity, token compliance, and good-faith but witless attempts to participate. Innovation fatigue certainly lends itself to “this too shall pass” passivity, but native skepticism can produce the same condition even when there is no prior history. If there is a perceived gap between the known resources and the stated goals, or distrust of the promoters, *inter alia*, a “wait-and-see” attitude may be prominent even if there is no history of prior innovation collapse.

Wide variation of participation across third-party efforts highlights one of the most prominent limits of the approach: it is inherently time-intensive for police officers. High-end efforts like Oakland’s SMART team, with focused political support and sufficient resources at all levels, are relatively rare. When resistance and reluctance are the norm, officers working with third parties must recontact them frequently to insure that plans or instructions are understood and promised actions taken. Periodic contacts are equally important to learn of legitimate new complications and to debunk new illegitimate excuses for noncompliance.

Overt Resistance. Managers and others in position of authority in agencies approached to become “partners” in a police initiative can say no. This is the most visible hurdle facing officers seeking a third-party solution to problems. Outright refusal is usually couched civilly in terms of one or more of four major objections: (1) lack of centrality of the police request to the business’s or individual’s primary mandate, (2) lack of legal authority, (3) lack of resources, and (4) competing demands from external sovereigns with greater call on the organization's resources.

"Overt resistance" does not necessarily involve confrontation with the police: a common form of resistance is simply to ignore police requests to meet about a problem, and patterns of ownership insulate some place managers by distance. Several RECAP properties were owned by corporations with headquarters on the east coast: leasing was handled by agents in Minneapolis and rent checks mailed to addresses out-of-state. One landlord ignored RECAP attempts to speak with him until the case officer drove to the landlord's residence (in a third-tier suburb outside the Twin Cities) to leave a handwritten note. The landlord's comment that "if he came all the way out here, I figured he must be serious" demonstrated the "ignore them and they will go away" legacy of innovation fatigue. There are other forms of resistance to be aware of, however, hidden under various guises of compliance.

Lip Service. In the face of continued police pressure, the easiest form of noncompliance is verbal acquiescence without substantive action. If the police can be stalled for long enough, they will either give up or be redirected to other pursuits. Follow-up contacts are vital and tangible intermediate actions and goals desirable wherever they can be articulated. Lip service is difficult to maintain in the face of continued reminders that promised work has not been done.

Token Compliance. A step above lip service is token compliance: "face time" attendance at meetings without commitment to action or designation of liaison personnel without real powers or even direction; in short, providing the most minimal level of participation or resources to make a marginally credible claim of cooperation. Token compliance is another stalling technique, but it also constitutes a visible demonstration of good faith: that demonstration is often more for political sovereigns than for the police, though it serves both in the initial stages. Where lip service "says something," token compliance "does something." To overcome ineffective results, police must make a case both for the ineffectiveness of the token effort and the actual capacity of the organization or individual to do more.

Lip service and token compliance illustrate one of the most stringent limits to coerced volunteerism: the "coercion" often is not and cannot be real. The "inherent coercion" of police presence is anchored in the criminal law and has little purchase in the civil arena without a specific regulatory mandate.

Covert Disruption. Third parties may take actions to undermine the entire thrust of the police initiatives. Asked to appoint a manager for an apartment building, one landlord designated a low-functioning tenant as a manager: the individual was incapable of controlling the behaviors of the other residents and limited himself to doing minor physical repairs. A problematic landlord was at first thought to be cooperative: he quickly made the physical alterations suggested and directed his on-site manager to implement the other provisions of the officers' plan for reducing calls to the property. Several months later, the police learned that he had countermanded those orders as soon as the meeting had ended and fired the manager for actually attempting to implement changes.

One of the more notorious landlords reacted to the news that one tenant in his RECAP building was a medium-level drug dealer by evicting the individual. Several weeks later, a precinct raid at another address owned by the same landlord revealed that he had "evicted" the tenant by moving him into another building, in effect transferring the associated problems to a new neighborhood.

Drug dealers were a point of contention that became emblematic of the cognitive gaps in problem definition that sometimes affect third-party policing efforts. To the police and the surrounding neighborhood, problems caused by a resident drug dealer are all too obvious: the influx of unsavory characters and associated rise of property crimes; unpleasant confrontations with marginal individuals and those under the influence of drugs; detritus including feces and urine, needles, and used condoms; constant noisy traffic; and related incivilities.

To absentee landlords, however, drug dealers were almost ideal tenants: they paid the rent on time, in cash; they made no demands on the landlord; and they did not complain about minor deficiencies in the physical plant. Insulated by distance from the toxic by-products of the illicit activity, absentee landlords were ignorant of the problems, and not disposed to oust a cash-paying tenant who had caused the landlord no problems. The problem at the time was accentuated by a surplus of available rental units in Minneapolis. The natural response of the landlords was to categorize the ancillary problems as "police responsibilities" because the police requests flew in the face of the basic purpose of renting property (making money with minimal expenditure of capital and effort).

Good-Faith Incompetence. At a large residential complex, a suggestion that security be employed led to the hiring of a local private service whose employees were gang-connected and essentially used their legitimate employment as a cover for increasing drug sales. It was never certain that they had been hired to suborn the police request in that fashion; they were lowest bidders for the job, and the situation was chalked up to a form of good-faith incompetence.

Similar results were encountered in a variety of settings, including a large homeless shelter that hired a security guard to control behaviors within the facility. One night, several male shelter residents overwhelmed the guard, pinned him to the floor, and ostentatiously cut the buttons off his shirt, one by one, before releasing him unharmed. The security guard quit that night.

A number of caretakers encountered similar problems at multifamily dwellings on the RECAP list. The common link was that they – like the security guard at the shelter – were unprepared to deal with the troublemakers in their facility. The place managers were capable enough for the routine duties of the position but were vastly overmatched by the residents who were hardened by street life and their criminal associates.

Client and Partner-Centered Limits

"Ask, and ye shall receive" is a religious precept. In the secular and often profane world of policing, "to ask" often means to be told to "[bleep] off." Compliance is not the first resort of persons who are asked to take on additional obligations, in order to insure a return that perhaps is perceived as dubious, and in any case is defined in terms of benefit to persons other than the individual so entreated. As a result, forays into third-party policing are constrained by a number of factors centered in the perceptions of the would-be partners (such as the differing perspectives on the drug-dealing tenants, above), many of which are operational counterparts of the techniques of negation articulated by Sykes and Matza (1957).

Corporate and Agency Resistance

Resistance from external organizations can be anticipated in four different guises: (1) the request lies outside the organizational mandate as interpreted by the organizational hierarchy, (2) the lack of legal authority to take on the actions requested, (3) the request cannot be fulfilled because of budgetary restrictions, and (4) the organization is already overloaded with the demands of their own mandate. The rationales sometimes overlap, and they often are legitimate: shrinking budgets are a primary reason that fulfilling organizations' own mandates is difficult, as American police departments are discovering in the "new normal" of the economy.

Organizations: Outside the Ordinary Mandate. While there may be occasions when this is a proper remonstrance, in many cases the argument is really that the request lies outside the routine *practices* of the organization. The analog is to Sykes and Matza's concept of "denial of responsibility." There is a deeper issue, however: the police are seeking an intervention by an agency that possesses defined statutory powers, in order to coerce compliance with police requests in areas that are outside the partners' specific mandates, and for which no coercive authority exists to enable direct police action. While the regulatory authorities do have a mandate to intervene in one or more of the intermediate conditions, there is no corresponding requirement that they do so to enable other actions for other agencies: in most western countries, regulatory functions exist in their own silos.

The challenge of third-party policing is to forge credible links between the separate enforcement rationales. While that might imply a veiled threat that exposure of the conditions will make the regulatory agency look bad in the public eye, such a premise should not be overstated, and certainly cannot be overused. Agencies (including the police) have well-developed defense mechanisms to defend their standard practices against complaints that come from all quarters. The most powerful is that of competing or superseding mandates.

In Jersey City, a grant-funded program to forge a joint Police Department-School Department venture to address absenteeism and in-school misbehavior foundered on issues related to specific legal mandates. It was the position of the School Board that privacy concerns did not permit the sharing of student information with the police, and repeated entreaties to bridge the gap failed. When RECAP tried to persuade the Minneapolis Housing Authority not to house young alcoholics and addicts in senior citizen high rises, the authority invoked the provisions of their charter, comparable to the soon-to-be-enacted Americans with Disabilities Act (ADA).

In Minneapolis, a powerful nominal ally – the City Prosecutor's office – refused to cooperate with RECAP's attempts to curb high-volume shoplifting. Standard practice was to *nolle prosse* misdemeanor shoplifting cases because the volume alone would overwhelm the office's capacity to prosecute cases (a legitimate "lack of adequate resources" refusal). Analysis suggested that most shoplifters were one- or two-time offenders who desisted after they were caught the first time; real damages were inflicted by the serial shoplifters and organized shoplifting gangs. The organized groups were more likely to be from outside the area; when apprehended, they had enough stolen goods to support felony theft charges. RECAP targeted local prolific shoplifters, proposing a plan to monitor all shoplifting arrests (and releases) to aggregate serial offenders' thefts from a series of low, no-prosecution levels to a total amount that constituted a felony, whereupon they would be presented for prosecution.

Though the city attorney's office still balked, invoking its sovereign status vis-à-vis the police, the unit took the issue into the political arena. Their proposal had the backing of the chief, raising the specter of a public spat with adverse political consequences, and so overt resistance gave way: prosecutors grudgingly accepted the RECAP compromise plan. What followed, however, was an institutional process of covert disruption: without notifying RECAP, the city attorney's office moved the monetary bar for prosecution higher and higher, to *nolle prosse* the lower-threshold aggregations that RECAP forwarded. In token compliance with the request, the office had other means to insure that normal practice was not disrupted by RECAP's third-party gambit.

RECAP also attempted to enlist a comparable effort on the part of major retailers whose stores were frequent targets of thieves. Retailers readily agreed to participate in the plan: from their perspective, it was little different than the existing state of affairs, with some promise of a better curb on a small group of local offenders. In the process of the discussion, however, it became evident that the corporate security apparatus – individually and collectively – had a considerably more sophisticated grasp on the situation than did the police. A logical extension of their efforts to track the shoplifting gangs across multiple jurisdictions created an extensive database of apprehended shoplifters. It could have saved the Minneapolis police a considerable amount of effort that ultimately went into recreating the local version. Naturally, the corporations were reluctant to share it with the police.

The polite distance of the corporate entities from anything beyond token compliance with RECAP's plan illustrated another set of limitations when working with private sector entities: overlapping layers of constraint under which the private sector operates. The need to curtail losses to theft certainly fell within the corporate security mandate, but so did protecting company assets from other threats: corporate liability extended well beyond shoplifting. National and regional corporations certainly possessed adequate resources, but a much broader set of threats quelled more extensive cooperation. Logical negative consequences (never confirmed) included exposure to individual or class-action lawsuits for sharing information with the police, possible negative publicity related to accusations of behavior that we now term "racial profiling" (see, e.g., Meeks 2000), and a potential for loss of market share based on rumors. Negative returns on investment conceivably could include the neutralization of security techniques (if revealed publicly, as part of a lawsuit or because of "loose talk" by police officials) and an influx of even more serious shoplifting.

Organizations: Budget Constraints. Organizational budgets are fixed, particularly in the regulatory and social services. Like the police, the budgets tend to be devoted in large part to personnel. The police request often is for a temporary reassignment of employee time, malleable, and already covered by the budget. However, some community policing initiatives involving large-scale activities go beyond a simple redirection of time, and third-party requests may be comparable in their scope.

Overcoming the budget defense often involves building political coalitions that alter, if only temporarily, the priorities of agencies sought as partners. One-time interventions or projects (such as the cleanup of an abandoned lot to establish a neighborhood playground) are easier to mount than long-term, resource-intensive systematic projects to rehabilitate neighborhoods. One-time projects tend to be quick political capital, with *Before* and *After* photos; bringing about long-term changes (improving slipshod management at problem bars, shutting down thinly veiled storefront facades for the drug trade) are fundamentally different problems, inherently more difficult, with fewer opportunities for participants and political sovereigns to claim victory publicly.

On the rare occasion when nonprofit organizations are the focus of third-party efforts, budget will be both a defense and a potential lever. A privately held residential facility serving clients in rehabilitation or under duress was identified as an attractive nuisance: a number of its clients committed crimes both inside the facility and in the neighborhood and contributed substantially to neighborhood incivilities. The proprietor pled a lack of resources, attempting to turn the RECAP officer into an advocate in the proprietor's quest to obtain a higher level of certification that would bring in more money. After a protracted effort against the stalling tactics above, the tables were turned: revocation of existing certifications became the lever for establishing substantive changes. That decision was complicated by the fact that the facility served a large number of needy clientele who were not causing problems and were in need of shelter and other assistance provided there. Such moral dilemmas will attend third-party efforts as well as traditional policing.

Budget decisions made at remote corporate levels are less amenable to police intercession. Many corner gas-and-grocery convenience were beset by large numbers of shoplifting calls. Corporate policy forbade employees from detaining shoplifters, on the theory that the costs of a single successful lawsuit or worker's compensation claim would far outweigh the losses to shoplifting. One franchise owner revealed that while shoplifting losses could run about a thousand dollars in a bad month, the store turned a monthly profit between 50 and 60,000 dollars a month even after shoplifting losses were factored in. Police assertions that shoplifting (free beer, cigarettes, and munchies) was an attractive nuisance with a negative impact on the areas simply had no traction against this economic reality. RECAP officers went so far as to travel to the out-of-state headquarters of one of the chain companies to plead their case, but to no avail.

Leverage: Mobilizing City Organizations

In Minneapolis, building inspections were conducted in routine manner by multiple agencies: building codes, health and sanitation, and others. Advance notices mere mailed to property owners before inspections were to be conducted, allowing owners to make any needed repairs or cosmetic changes, masking deeper structural problems that would resurface eventually. Nominal compliance served an important goal of inspectorial agencies by maximizing their effectiveness to superficial review and minimizing the resources needed for follow-up and enforcement. It also helped to minimize pressure from political sovereigns, who were third-party defense mechanisms for property owners seeking to divert police attention from their own lack of responsibility.

RECAP needed a series of spot inspections outside the normal two-step of nominal compliance. To break the accepted routine of scheduled inspections, the inspection agency needed substantial evidence that such a change was needed. A high volume of calls to the police (the defining condition for selection as a RECAP address) was a behavioral problem, not a structural one. In some cases, the deadlock was broken by documentation: RECAP officers became informal inspectors, documenting the visible conditions of the building during the period when inspections were not anticipated. In others, the time-honored tradition of trading favors – standing by when a district inspector examined a particularly problematic address where physical harm was possible, for instance – would bring an inspector to a property unannounced. In a handful of cases, the documentation of problems at the property would go directly to the ward's Council Member, a preemptive move giving the police version primacy over the subsequent self-serving description by the property owner. It also undercut any agency invocation of past "successful" routine inspections as evidence that no intervention was necessary.

There are definite limits to police officers' ability to tap other city agencies (or employees) for favors outside the normal course of business, but the ramifications of doing so can be enormous. The sudden appearance of a formerly predictable regulatory agent, and the penalties she or he levies subsequent to the surprise inspection, is the civil equivalent of "a public hanging." Both specific and general deterrence (compliance with requests the first time the police ask, in this case) are desired results.

The various workarounds to routine must be used sparingly and cannot be squandered on trivial cases. The best initial case to go forward with will have extensive documentation, including support from neighbors, police records, and photographs where appropriate (RECAP used video surveillance on one particularly nettlesome property with a huge criminogenic impact on the surrounding neighborhood). High-profile targets also improve the media impact of the police efforts, though they can also be a liability if things go sour.

Once it is known that the police have that "power," the word will spread quickly among those accustomed to gaming the system: there is a new set of rules we have to deal with. In Minneapolis, there was a two-tiered division of landlords: individuals who owned one or two properties in an attempt to improve their financial standards, and well-financed owners of multiple properties. The majority in both categories were entrepreneurs who toed the line and maintained good properties; the handful of multiple-site owners who were gaming the system owned a surprising number of RECAP problem properties, and it was to that group that the "public hanging" efforts were directed. Smaller independent owners were more likely to be overwhelmed by unanticipated problems and welcomed police attention and the assistance it brought.

Individual Resistance

Denial of responsibility, denial of injury, and condemnation of the condemners (Sykes and Matza 1957) are the most commonly proffered justifications for refusal by individual citizens. Financial concerns analogous to organizational budgets are frequently offered, sometimes combined with spurious appeals to "higher loyalty" (usually the well-being of the individual's family): like organizational budgets, some of the claims are real, while others are spurious and self-serving.

Individuals: Denial of Responsibility. Unlike the criminals who were the focus of Sykes and Matza's original thesis, third-party individuals who rebuff police attempts to enlist them are not rationalizing any criminal behavior on their part. They are rationalizing a refusal to take on additional burdens, the compelled volunteerism articulated above. This variation of "denial of responsibility" exists upon a continuum, however: at one end are individuals who find themselves caught in the changing social conditions that produce the problems besetting the neighborhood. They are not a part of "the problem" through any actions of their own, and attempts to cast them as part of the problem purely on the basis of not being "part of the solution" (a bumper-sticker platitude at the ground level) will leave a sour taste.

At the other end are slumlords and do-nothing place managers whose *de minimus* discharge of ordinary responsibilities constitutes a complex set of visual and behavioral cues that allow, abet, or even encourage the criminal or disruptive acts that were reported to the police. Sometimes these conditions were due to sloth on the place managers' part, but sometimes they were deliberate: a number of bartenders in Minneapolis were found to have profited from allowing illegal drug sales and prostitution to occur in their premises. Instinctively, police understand that such individuals *are* part of the problem.

That assertion carries the onus of proof, however, so police must document their claim in several ways. The first is to provide evidence to the individual, in hopes that the person will acquiesce to avoid further consequences. Failing that, the case must be made to partner agencies that the individual is thumbing a nose at the regulatory agencies by superficial compliance with routine inspections. Additionally, officers seeking to employ third-party policing solutions may have to justify the effort to their superiors. Often, those justifications must take a form that allows the command staff to justify the efforts to political sovereigns, to whom the proximate targets of the third-party efforts have appealed to "get the cops off my back."

In this context, a variant form of *denial of injury* comes into play, sometimes abetted by a variant of *denial of victim*. Several Minneapolis landlords, and several bar owners, portrayed the police problems at their properties as an inevitable spillover effect of neighborhood demographics: "those people." In some cases, the third-party targets attempted to portray themselves as advocates for the troublemakers: "where else are they going to go?" It was a thin rationale and easily countered by the rejoinder that "help" does not mean allowing the same destructive habits that brought the troublemakers to their present state. Nevertheless, it is a rationale that might find sympathy elsewhere, a triangulation of support that might make the third-party efforts more difficult: police are not the only agents capable of developing coalitions. Because it consists of "soft" coercion applied to nominal respectables, third-party policing carries the onus of anticipating challenges and countermeasures, and taking steps to neutralize them before they are invoked.

Individuals: Budget Limitations. Budget limitations may be very real for the individual landlord or owner of a business establishment; the counterargument that investments made in the police efforts will work to improve the property's bottom line is always speculative at first. Small-time owners (with one or two properties) have fewer resources, and often lesser skill sets for managing, than do the more established and well-capitalized owners of multiple properties.

RECAP's exemplar of this condition was an individual who had just taken back a property when a contracted buyer defaulted, only to find that the well-managed property had been run into the ground: the physical plant had been allowed to deteriorate, and the middle-class tenants had left when more problematic individuals were let into the property by fly-by-night speculators. When the RECAP officer contacted him and laid out the list of police calls to the property, the owner responded with "I'll work on that if you'll work on these" and presented a stack of repair orders from various inspectors' offices that amounted to a huge amount of money. In that case, third-party policing took on a different form, with the RECAP officer interceding with the other regulatory bodies, buying time for physical repairs in exchange for the owner's reestablishing the former rules of conduct. Both sides of the effort were successful.

Another variation of the budget defense – sometimes a bluff in Minneapolis's renters' market, but also exercised – was to threaten to "walk away from" a property. . . essentially abandoning it and allowing it to become derelict. Few landlords actually walked away, but several did sell the property to another landlord, often buying a new property from a third owner in turn. A retrospective analysis of real estate transactions revealed a well-established pattern of such "trading sales." One of most prolific problem-causing properties changed hands five times in the course of the year-long RECAP experiment: each time a new owner took over a property, previous gains and concessions were lost, and the RECAP officer had to begin anew.

Police Organization Limits

Nominally, police agencies are hierarchical organizations that rest heavily on principles of command and control. In reality, those principles are loosely observed, and in some agencies are more honored in the breach than the observance. Leadership in subsidiary parts of agencies may cleave to a nominal subscription to agency values, but add beliefs and inclinations that are independent of it. Individuals' skill sets may or may not be amenable to new duties or approaches. As a result, attempts to gain leverage from third-party policing may falter on one or more of the same deficiencies that are found in nominal partners.

Overt resistance can be particularly sharp in police agencies. When problem-solving officers attempt to enlist the help of others in the organization, it is usually because their particular skills are needed: the work of undercover drug officers, gang unit members with special expertise, and the like. As such, the "lack of centrality" and "lack of authority" defenses available to outside agencies are not viable. Lack of resources and competing demands from sovereigns (within or outside the agency) are generally the first reason given for refusal: "I'd like to help you, but…" Given that the demands for specialized units usually outstrip their ability to meet them, these initial refusals are not unreasonable.

Another form of overt resistance is the darker side of autonomy: police officers take pride in their ability to work through their own problems and to do so on their own terms. Backup requests are routinely honored, especially at the peer level, but when line officers must ask for help from superior officers in other units, that equation changes. Jersey City attempted to evoke a problem-oriented response to street prostitution in two areas but was rebuffed by the precinct command staff. Treating the grant funds as an overtime windfall, the patrol units chose to ignore the problem-oriented plan in favor of traditional suppressive approaches that displaced the street trade into neighborhoods previously free of it. (A simultaneous gambit asked the Jersey City courts to levy jail time instead of fines and credit for "time served" while awaiting arraignment, escalating sentences after multiple convictions; the idea was to deprive pimps of the income from the prostitutes. Like RECAP's shoplifting initiative, the effort was dismissed out of hand: the judiciary claimed a greater need for bed space to house violent offenders.)

Police agencies generally have a silo approach to specialty work: if a special unit has been created, then that work is somebody else's responsibility. That ethos is slowly changing in many places but is real enough to be considered a limitation on third-party policing initiatives, whatever their origin. Overcoming silo mentality requires a skill set and an approach similar to dealing with external agencies: some plausible benefit, or a win-win overlap of separate responsibilities, must create a foundation rationale for internal cooperation.

Going up the chain of command to realign the priorities of nominally equal peer units is risky, though if it is done skillfully enough, the tit-for-tat repercussions that come from "stepping on toes" may be avoided. If persistent requests come from resident groups or other civilian partners, some of the onus might be lifted from the requesting officers. Deferring the decision upward to higher command levels, by presenting various strategies with and without the available units, might be a workaround. In many cases, however, the processes are so heavily linked to personalities at all levels that it is impossible to prescribe a successful path to enlisting organizational assistance.

Covert disruption is less likely within the agency context than elsewhere, because the consequences of disruption attach to the agency and because memories are long. While there are some documented episodes of disruption in the history of policing experiments (beginning with the Kansas City Preventive Patrol Experiment), they tend to be expressions of disdain at interference by outsiders, not internal tampering. More serious disruptions tend to work in the opposite direction: "cooking the books" to make ineffective anticrime initiatives look better for political reasons.

Lip Service requires much less effort than covert disruption and is a more likely result in police agencies. The demands of external sovereigns, the changing priorities within the agency (or the precinct, or other unit), and the demands of the 9-1-1 system all become plausible reasons why promised assistance does not materialize. One RECAP officer attempted to enlist the assistance of peers he had worked with in patrol, asking them to make arrests when called to domestic violence incidents at a pair of properties where domestics were the primary driver of police numbers. Though they readily agreed, no arrests were forthcoming, even though an analysis of the active call information indicated a sound foundation of probable cause in multiple interests.

Token compliance is a close cousin to lip service within agencies, but as with outside partners, it involves a "did something" component that falls well short of the assistance needed. A day of directed patrol in a problematic neighborhood where 2 weeks or more are needed is one such manifestation. A couple of undercover drug buys where a sustained effort is needed is another. In all cases, a resort to competing demands for short resources is the explanatory rationale for levels of participation that are less than desired.

Good-Faith Incompetence. All departments of any size have individuals who meet minimal levels of competence but are unable or unwilling to reach more desirable levels of performance (the subtle Minneapolis appellation was "the sick, the lame, and the lazy"). Seconding an underperforming or "problem" member to the third-party request is a way of answering two distinct needs: the home unit is no longer responsible for the individual, and the unit can claim to have responded positively to the request. (While "good faith" may be a stretch in such cases, the response can be distinguished from covert disruption because no particular ill effects are intended.)

Like many policing initiatives, third-party policing is also vulnerable to a systemic problem in policing: lack of stability and continuity. Though no employee is irreplaceable, some are more attuned than others to the nuances of third-party work (like any other specialty). Local knowledge and trust are important to patrol officers, who tend to be regarded as interchangeable cogs in the machine. The "here today, gone tomorrow" transfers that are the norm for patrol can be fatal to third-party efforts. Knowledge and trust are vital to being able to sustain third-party policing, but when police agencies fall short, it inadvertently adds to innovation fatigue. Because building coalitions factors into many third-party and community-based initiatives, it is important to sustain third-party effort with long-term planning for promotions, transfers, or retirements.

Summary

The basic premises of third-party policing are simple: the police can be more effective by enlisting the efforts of partners with effective, consistent influence over the lives of the offenders who are the ultimate targets of police efforts ("policing *by* third parties"). Those potential partners who are reluctant to participate can be brought into conformity by civil pressures unrelated to the criminal law, if necessary ("the policing *of* third parties"). Excuses are illegitimate, and effective third-party policing is simply a matter of finding the right tool and establishing leverage.

In reality, those premises sometimes founder upon harder realities. Some reasons for nonparticipation are legitimate, most often those related to available resources: attempting to push past them will create ill will at least and quite possibly may create active resistance sufficient to undermine an otherwise legitimate effort. One of the greatest unknowns of third-party policing efforts is whether or not the narrow interest of a particular initiative will run afoul of competing priorities of more sovereign entities, as was the case with the RECAP shoplifting initiative and the Jersey City prostitution abatement project.

The right tools are not always available, particularly when political relationships lie behind the problem definition. Some political decisions will be legitimate choices to forego one opportunity (the third-party initiative) for another, larger political goal. Others may be the result of hard or soft corruption, but in any case will not yield to police insistence. "The wisdom to know the difference" is one of the mantras applicable to third-party policing efforts.

The "we tried that once; it didn't work" pathology of organizational thinking can be deadly. Overcoming it requires a nonconfrontational investment in analyzing both prior and current efforts. Special operations teams routinely conduct after-action critiques to identify areas needing improvement, but SWAT and SRT units have a cachet within police culture that community policing still strives to achieve. The concept of meaningful review of other efforts has made only modest inroads in other areas of policing and needs to expand to improve third-party policing.

Against those limits, third-party policing requires a continuity and a sense of long-term strategy quite distinct from the call-driven siloing of patrol work. In some instances, success requires the building of coalitions to overcome obstacles, and there is much to be learned in that respect from the community organizing movement. The existence of coalitions of affected parties lifts the onus of "being the bad guy" from the police, and both legitimizes and broadens the pressure on third-party targets. A clear understanding that "coerced volunteerism" works best when the coercion is minimal, and the benefits of cooperation are perceived early in the process, should be a fundamental tenet of third-party policing.

Related Entries

- ▶ Broken Windows Thesis
- ▶ Civil Remedies
- ▶ Cost-Benefit Analysis
- ▶ Crime Location Choice
- ▶ Crime Prevention Through Environmental Design
- ▶ Criminology of Place
- ▶ Deterrence
- ▶ Effectiveness of Situational Crime Prevention
- ▶ Evidence-Based Policing
- ▶ Focused Deterrence and "Pulling Levers"
- ▶ Hot Spots and Place-Based Policing
- ▶ Innovation and Crime Prevention
- ▶ Law of Crime Concentrations at Places
- ▶ Legal Frameworks for Third-Party Policing
- ▶ Problem-Oriented Policing
- ▶ Pulling Levers Policing

Recommended Reading and References

Buerger ME (1993) Convincing the recalcitrant: reexamining the Minneapolis RECAP experiment. Unpunished dissertation, University Microfilms, Ann Arbor

Buerger ME, Mazerolle LG (1998) Third-party policing: a theoretical analysis of an emerging trend. Justice Q 15(2):302–327

Eck JE (1997) Preventing crime in places. In: Sherman LW, Gottfredson D, MacKenzie D, Eck J, Reuter P, Bushway S (eds) Preventing crime: what works, what doesn't, what's promising – A report to the Attorney General of the United States. United States Department of Justice, Office of Justice Programs, Washington, DC

Eck JE, Spelman W (1987) Problem solving: problem-oriented policing in Newport News police executive research forum. National Institute of Justice, Washington, DC

Green LA (1996) Policing places with drug problems, vol 2, Drugs health and social policy series. Sage, Thousand Oaks

Mazerolle L, Ransley J (2005) Third party policing. Cambridge University Press, Cambridge

Meeks K (2000) Driving while black: highways, shopping malls, taxicabs, sidewalks: what to do if you are a victim of racial profiling. Broadway Books, New York

Sherman LW, Buerger ME, Gartin PR et al (1989) Repeat call address policing: the Minneapolis RECAP experiment. Final report to the National Institute of Justice. Crime Control Institute, Washington, DC

Spelman W, Eck JE (1989) Sitting ducks ravenous wolves and helping hands: new approaches to urban policing. Public Affairs Comment Lyndon B Johnson School of Government, University of Texas, Austin, Winter

Sykes GM, Matza D (1957) Techniques of neutralization: a theory of delinquency. Am Sociol Rev 22(6):664–670

Third-Party Policing (TPP)

▶ Family Engagement Strategies to Reduce Crime

Time Series

▶ Longitudinal Crime Trends at Places

Time Series Intervention Analysis

▶ Interrupted Time Series Models

Time Series Quasi-Experiment

▶ Interrupted Time Series Models

TIP

▶ Trafficking in Human Beings

TJ

▶ Therapeutic Jurisprudence

Tobacco Abuse in Correctional Settings

Karen L. Cropsey[1], Gloria D. Eldridge[2] and Brendan C. Clark[1]
[1]Department of Psychiatry and Behavioral Neurobiology, University of Alabama at Birmingham, Birmingham, AL, USA
[2]Center for Behavioral Health Research and Services, University of Alaska at Anchorage, Anchorage, AK, USA

Overview

The United States incarcerates the largest percentage of its population of any country in the world, with over eight million adults and 650,000 youth under some form of local, state, or federal supervision, including prisons, probation, or parole (Taxman et al. 2007). The number of individuals under criminal justice supervision in the community has quadrupled over the past 25 years from 1.12 million in 1980 to 4.9 million in 2010 and comprises the largest segment of the criminal justice population (Glaze and Bonczar 2011). The increase in the criminal justice population has been attributed in part to the war on drugs and a political climate that favors increasingly punitive sentencing. Although a common misconception is that incarcerated people are removed from the community indefinitely, as many as nine million individuals are released from prisons and jails and return to their communities each year (Beck 2006). Consequently, policies surrounding the availability and use of tobacco products and the availability of resources for tobacco cessation in correctional settings have a profound impact on efforts to reduce the morbidity and mortality associated with tobacco use in the United States.

This entry discusses the prevalence of smoking and trends in tobacco use in correctional settings, common comorbidities such as substance abuse and dependence and psychiatric illness associated with tobacco use among incarcerated people, the health consequences of tobacco use among people in the criminal justice system, policies surrounding tobacco use in correctional settings, interventions for smoking cessations with criminal justice populations, and implications for the public health impact of tobacco use in the United States.

Fundamentals

Prevalence and Trends in Smoking

Despite over 40 years of public health messaging as well as population- and individual-based interventions, smoking remains the leading cause of death and disability in the United States. While smoking rates have declined to about 19 % among the general population, smoking remains disproportionately high for individuals in the criminal justice system (Cropsey et al. 2010).

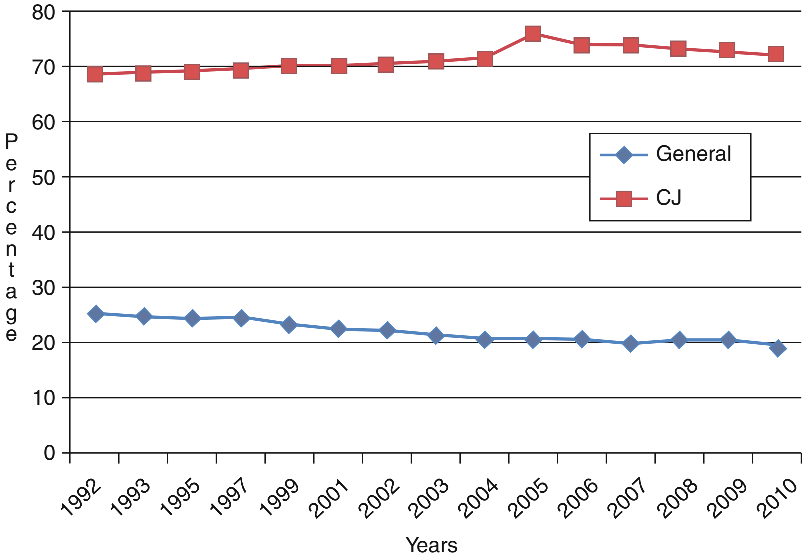

Tobacco Abuse in Correctional Settings, Fig. 1 Prevalence of smoking for criminal justice (*CJ*) and general population (1992–2010). Note: CJ data was available for the following years in the United States: 1992, 2004, 2005, 2006, and 2010; missing points were estimated. General population smoking estimates were taken from CDC, 2011, "Trends in Current Cigarette Smoking Among High School Students and Adults, United States, 1965–2010"

Figure 1 illustrates the prevalence of smoking from 1990 to 2011 for the general population and correctional populations. While individuals under criminal justice supervision represent approximately 3.6 % of the general population, they represent over 12 % of current smokers (Centers for Disease Control 2011). The high rates of continued smoking among individuals in the criminal justice system suggest that public health messages and interventions have been largely ineffective or have not penetrated to this population.

Smoking prevalence among adults in the general population is about 19.3 % (Centers for Disease Control 2011). Among incarcerated men, smoking prevalence is 70–80 % and has remained steady over the past 20 years (Cropsey et al. 2010). Similarly, a recent survey of individuals under community corrections supervision indicated that 72.3 % were current smokers (Cropsey et al. 2010). Smoking rates among female arrestees range from 42 % to 91 % – two to four times greater than among women in the general population (Nijhawan et al. 2010). Finally, approximately 70 % of adolescents in the juvenile justice system have tried smoking and half are daily smokers (Cropsey et al. 2008b). This compares to about 44.7 % of adolescents not in the juvenile justice system who have tried smoking and 6.4 % who reported smoking on 20 or more days of the month (Centers for Disease Control 2012).

Smoking is entrenched in the culture of prisons and jails and has a place in prison life that differs from that in the "free world." With smoking rates three to four times higher among criminal justice populations than among individuals in the general population, smoking is normative in correctional settings and lacks the stigma attached to smoking in the "free world" (Cropsey et al. 2008a). In correctional settings, smoking connotes peer acceptance and creates a shared social and emotional connection. In addition, tobacco products are an important part of the informal economy in correctional settings and are traded and bartered for other commodities or services.

Smoking Comorbidities

Mental Illness

In the general population, individuals with histories of psychiatric disorders have higher rates of lifetime and current smoking and are less likely to quit smoking. Rates of smoking are substantially higher among individuals diagnosed with depression, schizophrenia, and anxiety disorders. Active psychiatric disorders are risk factors for onset of daily smoking and progression to nicotine dependence. Nicotine-dependent individuals with comorbid psychiatric disorders make up less than 10 % of the US population but consume over a third of cigarettes smoked in the United States (Grant et al. 2004). In addition, the burden of smoking-related disease is significant in smokers with chronic mental illness. For example, among individuals with chronic mental illness, the prevalence of chronic obstructive pulmonary disease (COPD) is over four times greater than the estimated COPD prevalence for individuals of a similar age in the general population.

Psychiatric disorders are prevalent in correctional populations, which may contribute to the high prevalence of smoking in criminal justice population. The prevalence of psychiatric disorders in correctional populations ranges from 7.6 % to 36.5 %, with 14.7 % of incarcerated individuals diagnosed with a psychiatric disorder in the past year (Glaze and James 2006). Additionally, one in four incarcerated people reports a learning impairment, which may compound the morbidity from other psychiatric diagnoses. The high prevalence of psychiatric disorders in correctional populations has been linked to the deinstitutionalization of mentally ill persons in the United States and the lack of access to mental health services in the community. Suicide was once the leading cause of death among jail inmates. Over the past 20 years, suicide has decreased drastically in the incarcerated population; however, the suicide rate in jails remains at 47 per 100,000 (Mumola 2005) compared to 12 per 100,000 in the general population (Heron 2012). Outside the correctional setting, incarcerated individuals are exposed to factors that contribute to the development of, and relapse to, mental illness, including sexual assault, childhood abuse and neglect, and abusive relationships. In addition, incarceration may exacerbate mental health problems, with violence, isolation from family and friends, and overcrowding being likely contributors to the high rates of mental illness in correctional populations. Thus, the rates of smoking and the high prevalence of psychiatric disorders in correctional populations suggest that a large proportion of incarcerated individuals may have difficulty quitting smoking without intensive interventions and may suffer long-term and serious health consequences from smoking.

Substance Abuse

In the general population as well as the incarcerated population, smoking is closely linked to the use of alcohol and other substances. In the general population, individuals with histories of dependence on alcohol and other substances smoke more heavily, are more nicotine dependent, and are less likely to quit successfully without intensive interventions. For example, 70–90 % of individuals who abuse alcohol and other drugs are cigarette smokers (Bowman and Walsh 2003), and tobacco use is associated with initiation and increasing levels of consumption of alcohol and other drugs.

Population studies show a clear, bidirectional relationship between nicotine dependence and dependence on alcohol and other drugs. Nicotine dependence is most prevalent among individuals with a current substance use disorder. Alcohol-dependent smokers are more physically dependent on nicotine, and heavy drinkers smoke more cigarettes, smoke sooner after awakening (an indicator of nicotine dependence), and inhale more smoke from each cigarette. In addition, individuals who are dependent on other substances smoke higher nicotine content cigarettes. Smokers with comorbid abuse of other substances are less likely to maintain long-term tobacco cessation and require more intensive behavioral and pharmacological interventions. Finally, alcohol and tobacco have a synergistic impact on health; health risks for individuals who abuse alcohol *and* smoke are greater than health risks for individuals who abuse either substance alone. In fact, among treated alcoholics, more deaths are attributable to smoking than to alcohol.

The “war on drugs” has been credited with the sharp increase in drug-related offenses among individuals in the correctional population. Most incarcerated people enter prison or jail with a history of abuse and dependence on alcohol and other substances. For example, about 56 % of inmates reported using drugs in the month prior to their offense, while about a third reported using drugs during the time of their crime (Karberg and Mumola 2006). Overall, about half of individuals in prison meet DSM-IV criteria for a substance use disorder. Given the overwhelming use of criminal sanctions for drug use behavior in the United States, as well as the association between drug use and smoking, it is not surprising that so many prisoners have high rates of smoking.

Medical Problems

The increase in incarceration over the past 30 years has resulted in increased correctional costs, from $20 billion in 1982 to over $74 billion in 2007, more than a threefold increase in annual expenditures (Bureau of Justice Statistics 2012). Healthcare costs have contributed substantially to this increase and are expected to double over the next 10 years, adding further financial burden to correctional systems already facing economic hardships (Lamb-Mechanik and Nelson 2000). Smoking-related diseases contribute substantially to this increase in correctional medical costs. Even though 88 % of individuals in the correctional system are younger than 50, 43 % of incarcerated males and 57 % of incarcerated females report at least one medical problem (Maruschak 2008). The most common medical problems are directly related to or exacerbated by smoking, including diabetes and heart, circulatory, respiratory, kidney, and liver problems.

Correctional populations are especially vulnerable to the negative health consequences of smoking. Many smokers increase smoking during incarceration, and many incarcerated people report a greater need to smoke due to the stress, boredom, and deprivation of prison life. Because of restricted availability and the expense of purchasing tobacco products from prison or jail commissaries or on the black market, many smokers switch to unfiltered hand-rolled cigarettes, which are higher in tar and nicotine. Recent studies have demonstrated that smoking roll-your-own cigarettes was associated with greater smoke exposure compared to factory-manufactured cigarettes. Thus, smoking in prison may present a greater health risk than smoking outside prison.

Racial Disparities

Over the past decade, smoking prevalence rates for African-Americans have come to equal those of Caucasians (21.3 % vs. 22.1 %; Centers for Disease Control 2011). This increase in smoking for African-Americans is further complicated by the fact that they tend to have higher rates of smoking-related diseases and mortality from smoking and more difficulty with smoking cessation. The differences in cessation rates have been documented even when controlling for potential moderators in both incarcerated (Cropsey et al. 2009) and nonincarcerated populations. This increase in smoking-related consequences and difficulties with cessation may be due to African-American's preference for smoking mentholated cigarettes which results in higher CO concentrations. African-Americans are overrepresented in the criminal justice system such that while they comprise only about 30 % of the general population, approximately 60 % of African-American men are incarcerated for at least 1 year before age 35 (Pettit and Western 2004). The overrepresentation of African-Americans in the criminal justice system in conjunction with their increased difficulty with cessation and greater health risks highlights both the importance and difficulty of cessation in this population.

Prison Policies

Tobacco policies have changed profoundly in US correctional facilities in the past several decades, with a consistent trend toward increasingly restrictive policies toward tobacco availability and use (Kauffman et al. 2008). Historically, tobacco use was an accepted part of life in US penal institutions, and state and federal correctional institutions routinely provided free tobacco. Although some correctional systems provided tobacco products to make incarceration more humane, cigarettes also played a role in maintaining order within facilities.

In the mid-1980s, about half of state correctional facilities provided free tobacco. By the early 1990s, tobacco products were sold in about 88 % of prison commissaries, 26 % of prison systems provided tobacco at no cost for indigent inmates, and 8 % of prison systems provided no cost tobacco products to all inmates (Vaughn and del Carmen 1993). Some correctional systems phased out distribution of free tobacco as incarcerated people gained the opportunity for paid employment and had money to purchase tobacco products from prison commissaries.

The initial impetus for changing tobacco policies was cost containment, including the cost of providing tobacco products, maintaining buildings and facilities, and treating smoking-related medical conditions. The trend toward restricting tobacco gained momentum because of concerns about exposure to environmental tobacco smoke (ETS). ETS is a particular health risk in correctional facilities because of the high prevalence of smoking, overcrowding and poor ventilation, inadequate facilities, and unrestricted indoor smoking or poor enforcement of indoor smoking policies. Smoke concentrations in correctional facilities are greater than in the homes of smokers. Finally, the desire to limit amenities for incarcerated people provided additional impetus for changes in tobacco policies. In 1986, only 5 % of US prisons had smoke-free living areas; by 2007, 96 % had smoke-free living areas, 27 % banned indoor smoking, and 60 % banned smoking completely (Kauffman et al. 2008).

Tobacco policies in correctional settings range from policies restricting smoking to policies prohibiting possession of tobacco products or tobacco use anywhere on facility grounds. A complicating factor in correctional tobacco policies is that the correctional facility is the incarcerated individual's home and there is some debate whether incarcerated people have the right to smoke as long as nonsmokers, other inmates, and staff members are protected from exposure to ETS. Complete tobacco bans force incarcerated people to quit tobacco use involuntarily or continue to use tobacco products illicitly (Eldridge and Cropsey 2009).

Although there is relatively little research on the impact of tobacco bans and restrictions in correctional settings, the research that is available indicates that the majority of people continue to smoke after the imposition of tobacco restrictions. Partial tobacco bans (i.e., incarcerated individuals are allowed to possess tobacco products but smoking is restricted in indoor areas) may be more difficult to enforce than complete tobacco bans. Organizational and environmental factors in correctional settings often mitigate against consistent enforcement of tobacco bans and the imposition of disciplinary sanctions for tobacco possession or use. Although administrators and others are often concerned about the possibility of violence and disciplinary infractions during the transition to smoking restrictions, there is little evidence that this occurs.

Despite the fact that tobacco use tends to persist after the imposition of smoking bans and restrictions, there is evidence of a reduction in the amount of tobacco consumed, which has implications for health and air quality in correctional settings. There is evidence of an improvement in air quality and a reduction in ETS and suspended particulates after the imposition of tobacco restrictions.

Despite these positive changes as a consequence of tobacco bans and restrictions, one negative change is a gradual reduction in the availability of tobacco cessation programs and materials. For example, tobacco cessation programs were offered in 86 % of correctional facilities with indoor tobacco restrictions compared to 39 % of facilities with total tobacco bans. Similarly, tobacco cessation aids (e.g., nicotine replacement) were available in 65 % of facilities with tobacco restrictions but only 35 % of facilities with total bans. This raises the question about whether the goal of tobacco restrictions in correctional settings is to suppress tobacco use during incarceration or to accomplish lifelong tobacco cessation. The trend toward restricting tobacco use while decreasing availability of tobacco cessation programs and materials is likely to result in temporary suppression of tobacco use during incarceration rather than long-term tobacco cessation during incarceration and after release (Kauffman et al. 2008). Although data are sparse, almost all individuals released from smoke-free correctional facilities resume smoking within 6 months of release (>98 %), with only 37.3 % remaining abstinent from smoking at the end of the first day post-release (Lincoln et al. 2009).

Interventions

Very few smoking cessation studies have been conducted with correctional populations, and none have been conducted with individuals in community corrections. Initially, two pilot studies examined smoking interventions with prisoners, and both used small samples and focused exclusively on male inmates. More recently, a large randomized controlled trial of smoking cessation with female prisoners who were provided NRT and 10 weeks of group smoking cessation counseling showed cessation rates comparable to community smoking cessation interventions. The literature suggests that spontaneous quit attempts among female prisoners are rare. However, this study demonstrates that female prisoners are interested in smoking cessation that it is feasible to provide smoking cessation in correctional environments, and that incarcerated smokers quit smoking successfully when provided with community standard smoking cessation interventions. However, no studies have examined providing integrated smoking cessation treatment to community corrections populations, despite the high need for such interventions and the public health relevance of reaching this low-income population. Given that smoking is now concentrated in individuals with low incomes who may have substance use and mental illness problems, smoking cessation interventions must be able to reach these individuals who are at risk and who do not have access to standard smoking cessation interventions in the community. Individuals in community corrections represent such a population of high-risk smokers who are poor and do not have access to community-level smoking cessation interventions.

Recommendations for the Field

Correctional settings house individuals with high rates of smoking, smoking-related medical conditions, histories of drug and alcohol abuse and dependence, and psychiatric illnesses that present challenges to long-term smoking cessation. Incarceration (and criminal justice supervision in the community) provides an opportunity to address tobacco use in a population that may otherwise have less access to tobacco cessation programs and materials. In 2002, the National Commission on Correctional Health Care in its report on the health status of soon-to-be-released inmates recommended that incarcerated people have access to smoke-free environments and smoking cessation materials and programs. However, smoking among criminal justice populations, particularly individuals under community corrections supervision, is virtually ignored in the research literature, despite the enormous human, health, and economic costs to correctional settings and to the community after release (Cropsey et al. 2004).

The current state of tobacco policies (i.e., tobacco bans and restrictions and availability of tobacco cessation programs and materials) in correctional settings highlights the gap between temporary suppression of smoking during incarceration and long-term smoking cessation. Although data are sparse, most individuals released from correctional facilities relapse to smoking soon after release. Prisons and jails provide an opportunity for achieving lifelong smoking cessation for disadvantaged men and women, with the potential for an enduring positive impact on health and economic security. However, as correctional settings impose tobacco bans and restrictions while diverting resources away from smoking cessation programs and materials, that opportunity is being lost.

Recommended Reading and References

Bailey BA, Daugherty RA (2007) Intimate partner violence during pregnancy: Incidence and associated health behaviors in a rural population. Matern Child Health J 11:495–503

Beck AJ (2006) The importance of successful reentry to jail population growth. Presented at the Urban Institute's jail reentry roundtable, 27 June 2006. www.urban.org/projects/reentry-roundtable/upload/beck.PPT

Bowman J, Walsh R (2003) Smoking intervention within alcohol and other drug treatment services: a selective review with suggestions for practical management. Drug Alcohol Rev 22:73–82. doi:10.1080/0959523021000059857

Bureau of Justice Statistics (2012) Direct expenditures by justice function, 1982–2007. Extracted on 24 July 2012. http://bjs.ojp.usdoj.gov/content/glance/tables/exptyptab.cfm

Centers for Disease Control and Prevention (2011) Vital signs: current cigarette smoking among adults aged ≥18 years – United States, 2005–2010. Morb Mortal Wkly Rep 60(33):1207–1212. Accessed 24 Jan 24 2012

Centers for Disease Control and Prevention (2012) Trends in the prevalence of tobacco use national YRBS: 1991–2011. Extracted on 19 July 2012. http://www.cdc.gov/healthyyouth/yrbs/pdf/us_tobacco_trend_yrbs.pdf

Cropsey KL, Eldridge GD, Ladner TL (2004) Smoking among female prisoners: an ignored public health epidemic. Addict Behav 29:425–431

Cropsey KL, Eldridge GD, Weaver MF, Villalobos GC, Stitzer ML, Best AM (2008a) Smoking cessation intervention for female prisoners: addressing an urgent public health need. Am J Public Health 98:1894–1901

Cropsey KL, Jones-Whaley S, Jackson DO, Hale GJ (2010) Smoking characteristics of community corrections clients. Nicotine Tob Res 12:53–58

Cropsey KL, Linker JA, Waite DE (2008b) An analysis of racial and sex differences for smoking among adolescents in a juvenile correctional center. Drug Alcohol Depend 92:156–163. doi:org/10.1016/j.drugalcdep. 2007.07.018

Cropsey KL, Weaver MF, Eldridge GD, Villalobos GC, Best AM, Stitzer ML (2009) Differential success rates in racial groups: results of a clinical trial of smoking cessation among female prisoners. Nicotine Tob Res 11:690–697

Eldridge GD, Cropsey KL (2009) Smoking bans and restrictions in U.S. prisons and jails: consequences for incarcerated women. Am J Prev Med 37(2s): S179–S180

Glaze LE, Bonczar TP (2011) Probation and parole in the United States, 2010. U.S. Department of Justice, Bureau of Justice Statistics, NCJ 236019

Glaze LE, James DJ (2006) Mental health problems of prison and jail inmates. Bureau of Justice Statistics, NCH 213600

Grant BF, Hasin DS, Chou P, Stinson FS, Dawson DA (2004) Nicotine dependence and psychiatric disorders in the United States: results from the national epidemiologic survey on alcohol and related conditions. Arch Gen Psychiatry 61:1107–1115. doi:10.1001/archpsyc.61.11.1107

Hammett TM, Harmon MP, Rhodes W (1997/2002) The burden of infectious disease among inmates of and releases from US correctional facilities. Am J Public Health 92:1789–1794

Harrison PM, Beck AJ (2003) Prisoners in 2002. U.S. Department of Justice, Bureau of Justice Statistics, NCJ 200248

Heron M (2012) Deaths: leading causes for 2008. Natl Vital Stat Rep 60:9–10. http://www.cdc.gov/nchs/data/nvsr/nvsr60/nvsr60_06.pdf. Accessed 25 Feb 2013

Himelhoch S, Lehman A, Kreyenbuhl J, Daumit G, Brown C, Dixon L (2004) Prevalence of chronic obstructive pulmonary disease among those with serious mental illness. Am J Psychiatry 161:2317–2319

Karberg JC, Mumola C (2006) Drug use and dependence, state and federal prisoners, 2004. Bureau of Justice Statistics, NCJ213530

Kauffman RM, Ferketich AK, Wewers ME (2008) Tobacco policy in American prisons, 2007. Tob Control 17:357–360

Lamb-Mechanik D, Nelson J (2000) Prison health care survey: an analysis of factors influencing per capita costs. National Institute of Corrections. http://nicic.org/pub/2000/0159.99.pdf. Accessed 9 Sep 2003

Lincoln T, Tuthill RW, Roberts CA, Kennedy S, Hammett TM, Langmore-Avila E, Conklin TJ (2009) Resumption of smoking after release from a tobacco-free correctional facility. J Correct Health Care 15:190–196

Maruschak LM (2008) Medical problems of prisoners. U.S. Department of Justice, Bureau of Justice Statistics, NCJ 221740

Mumola CJ (2005) Suicide and homicide in state prisons and local jails. U.S. Department of Justice, Bureau of Justice Statistics, NCJ 210036

Nijhawan AE, Salloway R, Nunn AS, Poshkus M, Clarke JG (2010) Preventive healthcare for underserved women: results of a prison survey. J Womens Health 19:17–22

Pettit B, Western B (2004) Mass imprisonment and the life course: race and class inequality in U.S. Incarceration. Am Sociol Rev 69:151–169

Taxman FS, Young DW, Wiersema B, Rhodes A, Mitchell S (2007) The national criminal justice treatment practices survey: multilevel survey methods and procedures. J Subst Abuse Treat 32:225–238. doi:10.1016/j.jsat.2007.01.002

Vaughn MS, del Carmen RV (1993) Smoking in prisons: a national survey of correctional administrators in the United States. Crime Delinq 39:225–230

TPP

► Third Party Policing and School Truancy

Traditional Retributivism

Alfonso Donoso
Instituto de Ciencia Política, Pontificia
Universidad Católica de Chile, Santiago
Metropolitan Region, Chile

Overview

This entry offers an account of some of the canonical ideas and authors of traditional retributivism. For present purposes, retributivism is understood as a type of justification of criminal punishment. One of the most salient characteristics of this justification is that it focuses on a past event, prior to the infliction of hard treatment upon the offender. Retributivism is thus defined as a backward-looking justification of punishment. Rather than emphasizing the likely consequences that the imposition of penal sanctions may bring about, traditional retributivism insists that the core of justified punishment is the intrinsic wrongness of the act committed by the agent, which makes the offender deserve the hard treatment entailed by this practice. This entry unpacks and explains these different retributivist claims by offering an overview of the *lex talionis*, Kant's retributivism, and Hegel's retributivism, three of the most fundamental sources of retributivism (for an account of contemporary retributivism, see Christopher Bennett's entry in this volume).

Introduction

Punishment is a perplexing practice. On the one hand, it imposes burdens upon individuals that under most circumstances we would consider inadmissible. The punished may see her freedom of movement radically hindered, her property expropriated, and even her life terminated. On the other hand, punishment is a pervasive practice the state undertakes on a daily basis. Indeed, the state claims punishment is part of its legitimate functions and, therefore, a practice to which it is entitled.

How can a practice like that be justified? Or, more fundamentally, is there anything that could possibly justify it? Abolitionists argue that punishment cannot be justified and, thus, should be abolished. However, *pace* abolitionists, there is a long tradition that has devoted time and effort to develop an argument for the justification of punishment. One such type of justification is consequentialism. Another is retributivism. Consequentialists hold that punishment is a justified practice if and only if it brings about consequences whose benefits outweigh the costs it causes and these benefits are greater than (or as good as) those obtainable by nonpunitive means. By contrast, retributivists hold that offenders deserve to be punished for their wrongdoings. Thus, for retributivism it is not the case that punishment is a justified practice only if it produces a benefit such as deterrence or rehabilitation and produces such a benefit better than alternatives. The retributivist insists that those ends cannot justify the infliction of hard-treatment. Instead, rather than focusing on future consequences, retributivists focus on a past act – the criminal offense – thereby offering a backward-looking justification of punishment. For the retributivist, those who committed a wrong are guilty of that wrong, and the guilty deserve to be punished. Punishment is, the retributivist claims, the intrinsically adequate response to wrongdoers.

Keeping these features of retributivism in mind, let us now consider three traditional sources of retributivist principles and ideas. In the course of presenting these different sources, emphasis is put on the questions of how to determine adequate types and amounts of punishments, the justificatory connection between punishment and crime, and the place of punishment in the broader context of moral and political theory.

Lex Talionis

Some of the oldest expressions of retributivism can be found in various passages in the Hebrew Bible. In "Leviticus," for example, it is said that "if a man cause a blemish in his neighbour; as he hath done, so shall it be done to him; breach for breach, eye for eye, tooth for tooth: as he hath caused a blemish in a man, so shall it be done to him again. And he that killeth a beast, he shall restore it: and he that killeth a man, he shall be put to death (24:19–21). And Exodus states: "if *any* mischief follow, then thou shalt give life for life, eye for eye, tooth for tooth, hand for hand, foot for foot, burning for burning, wound for wound, stripe for stripe" (21: 23–25). In ordinary parlance, this is what has come to be known as the "eye for an eye" principle or, in its Latin formulation, the *lex talionis*, which is aptly translated as the law that imposes a punishment that resembles in kind and degree the wrong committed (for another, older source of the *lex talionis* see the Hammurabi's Code of Laws (c. 1780 BCE)). For the purposes of this section, this principle's most relevant feature is the idea of "likeness" between present punishment and past wrongdoing. This feature shall lead us to consider in some detail both the plausibility and nature of the *lex talionis*.

To contemporary eyes, the *lex talionis* seems to be a very unlikely source for justified criminal punishment. The most obvious objection against this principle emerges when we think of the type of punishment that should be inflicted upon offenders for very serious wrongdoings. For example, should we rape the rapist? Or, should we torture the torturer? For some people, this makes a conclusive point against the revival of the *lex talionis* in current criminal practices (see Bedau 1978; Lacey 1988). However, this response is too quick. First of all, defenders of the "eye for an eye" principle may argue that although the *lex talionis* may not work for any type of crime, it does help us to explain the type of punishment that should be inflicted on some offenders for some crimes in certain circumstances. An offense at point is murder. In cases of intentionally killing of a person without justification or excuse, the *lex talionis* shows straightforwardly the punishment that is to be meted out to the murderer. Furthermore, proponents of the *lex talionis* may argue that this law has a much more sophisticated form than may first appear and that this qualified understanding of the principle should make us see the *lex talionis* as an adequate principle for a theory of punishment. The rest of this section shall consider in some further detail this more nuanced depiction of the *lex talionis*.

We should begin by asking what kind of principle is the *lex talionis* and in what sense it is connected to retributivism. This is a relevant question because it would be a mistake to understand the *lex talionis* as a version of a retributivist justification of punishment. The *lex talionis* does not tell us why someone may be morally justified in inflicting punishment upon an individual for her offense. Instead, it advances an answer to the question of what ought to count as the adequate punishment to be imposed upon an offender. This means that the *lex talionis* is logically posterior to the justification of punishment. In turn, this opens the possibility that the *lex talionis* is a principle that is compatible with different types of justification of punishment, including consequentialist justifications (see Waldron 1992).

The possibility that the *lex talionis* may be part of a consequentialist justification of punishment forces us to specify the sense in which this principle is connected with retributivism. The clearest sense in which both are linked is that both focus on a past event. We have already seen that retributivism offers a backward-looking justification of punishment. Similarly, the *lex talionis* answers the question of the type of punishment that is to be visited on an offender in the present by looking at the act performed by the offender in the past. This gesture to a foregoing act provides us with an answer (at least part of it) to the question of how we should punish an offender. In short, both retributivism and the *lex talionis* contemplate a time prior to the present to unravel relevant questions about punishment. This is what connects the "eye for an eye" principle with retributivism.

This connection to the past brings us to the core of the possibility of a more nuanced, sophisticated, and thus plausible account of the *lex talionis*. Such an account depends largely on how we interpret the equivalence that ought to exist between a past offense and present punishment. The bad name that the *lex talionis* may have derives from interpreting this equivalence strictly (i.e., raping the rapist and torturing the torturer). Indeed, strict equivalence makes punishment look more like sheer cruelty than like a justifiable practice grounded in a sound moral theory. Thus, the moral plausibility of the *lex talionis* depends on the existence of a more charitable interpretation of the connection between a past offense and its present punishment. The kernel of this more nuanced account of the *lex talionis* is to deny that we should punish offenders with an action identical to the offense committed. Instead, an adequate account of this principle is that an offender should be visited with the same relevant type of act as the one that constituted the offense. There are different and independent reasons for dismissing a strict interpretation of the *lex talionis*, but perhaps the two most evident are, first, that it is not possible (perhaps, not even logically possible) to inflict upon an offender at a present time an action identical to the past offense and, second, that a practice like that is vengeful, cruel, and inhumane.

If the strict interpretation is successfully dismissed, what can be offered as a plausible account of the *lex talionis*? Rather than asserting the *identity* between a past offense and the punishment that ought to be visited upon the offender, a plausible interpretation of this principle should affirm the *similarity* between these two events. But this is insufficient: what are the features of a past offense that are to be matched in present punishment? Should punishment bring about a physical and/or psychological harm similar to the one the past offense caused the victim? Should it be performed with a violence similar to that used by the offender? Should punishment be inflicted the same day of the week in which the offense was committed? All these are features that would make present punishment similar to a past offense but, surely, they are not what a defender of a morally grounded account of the *lex talionis* will advance in support of the principle. Instead, the *lex talionis* is considered in its best light when it is understood as a principle that, first, makes us focus on what constitutes the wrong of a past offense and, second, makes us choose the adequate type of punishment in light of that wrong. Thereby, the *lex talionis* tells us that only by establishing the wrongness entailed by the past offense can we have an answer to the question of the punishment that must be inflicted on the offender. Of course, which is that wrong and what are the relevant features that must constitute present punishment for a past wrong are things on which we will disagree. That is inescapable. However, inescapable disagreement does not imply that no arguments can be advanced in this matter.

Kant's Retributivism

Immanuel Kant's views on punishment and its justification are generally considered as paradigmatic of retributivist reasoning. Although some commentators have challenged this interpretation, or, at least, challenged that Kant's account is *purely* retributivist, this entry will be mainly concerned with the most widespread view on Kant's account of punishment. Thus, irrespective of the plausibility of a non-retributivist account of Kant's philosophy of punishment, Kant is here interpreted as an exemplar theorist of traditional retributivism (for non-retributivist interpretations of Kant, see Byrd 1989; Cummiskey 1990; Tunick 1996; Brooks 2003. For an account that challenges the consistency and coherence of Kant's views on punishment on grounds that they advance both a deterrence and retributivist system of punishment, see Murphy 1987.

The most developed account of Kant's retributivism is found in "The Doctrine of Right," the first part of his late work *The Metaphysics of Morals* (Henceforth, Kant 1991 followed by the pagination of the Prussian Academy of Sciences edition of Kant's writings and the page number of the English edition. For Kant's views on punishment, see also *Lectures on Ethics* [27: 286, 547, 552f]; *Critique of Pure Reason* [A317/B373]; and *Critique of Practical Reason* [A65-A67]). In "The Doctrine of Right," Kant emphatically tells us that criminal punishment.

> can never be inflicted merely as a means to promote some other good for the criminal himself or for the civil society. It must always be inflicted upon him only *because he has committed a crime*. For a man can never be treated merely as a means to the purposes of another or be put among the objects of rights to things [. . .]. He must previously have been found *punishable* before any thought can be given to drawing from his punishment something of use for himself or his fellow citizens. The principle of punishment is a categorical imperative, and woe to him who crawls through the windings of eudaemonism in order to discover something to release the criminal from punishment or even reduces its amount by the advantages it promises [. . .]. (Kant 1991, 6:331–332, pp. 140–141)

Kant is clear in insisting that an offender must be punished *only because* he has performed an action deemed criminal by the civil society. It is only because the offender has made himself worthy of punishment that punishment must be visited upon him. To do otherwise is to act impermissibly. To punish an offender O for action A because under the circumstances this would increase or maximize society's feeling of security by deterring O or other potential offenders from A-ing (or from performing other actions deemed criminal) is illegitimate and wrong. Similarly, to punish O because punishment would reform O, so after punishment O would be less likely to A again, is also an expression of illegitimate punishment. For Kant, only the guilty must be punished, and the guilty must be punished *because* it is deserved. To do otherwise is to use the offender as a mere means to an end, a violation of the Kantian moral law.

However, despite this strong retributivist approach, Kant notes that consequentialist considerations can also enter into the justificatory scheme. The point is that they cannot figure as having a primary or exclusive role in the justification of punishment. Indeed, by opening the door to consequentialist features in the justification of punishment, Kant advances an attractive and plausible account of retributivism that may well anticipate contemporary mixed theories of punishment (for Kant as a mixed-theorist see Byrd 1989. For mixed theories in general, see Hart 1959; von Hirsch 1976, 1993). Thus, Kant's account presents a version of retributivism that is more complex than some critics have been willing to recognize (e.g., Dolinko 1992). However, and to be sure, independently of the good and desirable consequences that may follow the imposition of punishment upon an offender, for Kant it is the intrinsic evil of the act performed by the offender that makes her deserve punishment.

This conclusion, however, can only be a part of a complete answer to the question of the justification of punishment. As Kant himself realizes, we also need to know "what kind and what amount of punishment is it that public justice makes its principle and measure?" (Kant 1991, 6:332, p. 141). To answer this question, Kant appeals to something close to the *lex talionis* conceived in terms of a "principle of equality." According to this principle, "whatever undeserved evil you inflict upon another within the people, that you inflict upon yourself. If you insult him, you insult yourself; if you steal from him, you steal from yourself; if you strike him, you strike yourself; if you kill him, you kill yourself" (ibid.). In Kant's own words, "the law of retribution (*ius talionis*) [. . .] can specify definitely the quality of and the quantity of punishment" (ibid).

How would Kant reply to the difficulties posed in the previous section of this entry in relation to the "eye for an eye" principle? Kant recognizes that although there are times in which the *lex talionis* principle cannot be strictly followed, it remains as a valid way to specify what the wrongdoer deserves: the spirit of the principle is upheld insofar as the punishment inflicted "is quite similar" to the offense performed (Kant 1991, 6:332, p. 141. See also 6:363, pp. 168f). It is worth noting that Kant's proposal on this issue seems to expand the types of punishment that may be rightly imposed upon an offender. He does not reduce the practice of state punishment to widespread types of penal retribution such as expropriation, restrictions of freedom of movement, or the termination of life. Indeed, he goes as far as to consider a variety of punishments, such as public apology or hand kissing (presumably to demonstrate the equal status as citizens of individuals from different social classes). Moreover, for cases in which the strict exercise of the *lex talionis* may seem offensive, cruel, or inhumane, Kant authorizes alternative forms of punishment (although to our eyes these alternatives may still be offensive, cruel, or inhumane). Thus, for rape and pederasty he proposes castration, while for bestiality he suggests "expulsion from civil society, since the criminal has made himself unworthy of human society" (Kant 1991, 6:363, p. 169). There is, however, one class of offense that cannot be responded to by proxy, namely, murder. "Here there is no substitute that will satisfy justice" (Kant 1991, 6:333, p. 142). In this case, death is the only legitimate response to be visited upon the offender since this sentence is proportional to the criminal's "inner viciousness" (ibid.).

Beyond the reasonable and serious doubts these types of punishments may bring about, Kant's conception of the *lex talionis* is successful in advancing an important consideration not to be overlooked when establishing both types and quantum of legitimate punishment. Kant's nuanced account of the principle of retribution brings to the fore the duty we have to uphold a principle of humanity in retribution, the violation of which contravenes not only human dignity but also the very concept of punitive justice (see Kant 1991, 6:363, p. 168f). Thus, Kant sees the *lex talionis* as the adequate principle of retribution, the appropriate way to determine how to punish offenders in a way required by the respect owed to all human agents. The satisfaction of this principle of respect is a core aspect of Kant's retributivism.

Besides this and other virtues, Kant's theory of punishment faces a series of difficulties that have been largely noticed by commentators (e.g., Murphy 1987; Merle 2000; Hill 2003). These problems are either inherent to retributivism, and thus, shared by all or most versions of it, or characteristic of Kant's moral and political theory from which his retributivist theory is built. Because many of the former difficulties are discussed in different places of this and Bennett's entry on retributivism, this section concludes with some brief comments on the latter type of difficulties.

By now it should be clear that desert has a central role in Kant's retributivism: it strongly denies the legitimacy of punishment inflicted, primarily, for reasons other than desert. However, given Kant's emphatic distinction between actions – external and the proper subject of justice – and motives – internal and the proper subject of virtue – it is difficult to provide desert with a coherent place within the theory, at least in the way it is presented in the "Doctrine of Right." To some extent, the difficulty derives from the lack of a sufficient account of intentional actions that make sense of the relevant internal factors of agents that Kant's retributivism must take into account to justify punishment.

This gap in the theory is closely linked to another feature missing in Kant's account. Kant affirms that offenders ought to be punished because their wrongdoings show inner wickedness. At the same time, Kant makes clear that the law is there to deal only with external actions – that is, it "does not include the incentive of duty in the law and so admits an incentive other than the Idea of a duty itself" (Kant 1991, 6:219, p. 46). From a liberal point of view, precluding the state and its coercive apparatus from meddling in the inner sphere of individuals is something to welcome. However, this insight is difficult to reconcile with Kant's concern with the internal wickedness of offenders that may make punishment justifiable. These are not irremediable difficulties, but they do cast some doubts about the internal coherence of Kant's theory of punishment.

Hegel's Retributivism

In the Anglo-American tradition of legal and political philosophy, G.W.F. Hegel's retributivism has received much less attention than Kant's (for exceptions, see Wood 1990, Chap. 6; Tunick 1992; Brooks 2012). This is regrettable, as Hegel's account of crime and punishment – mainly developed in his *Elements of Philosophy of Right* – provides enlightening insights into the nature of retributivism in particular and the institutions of criminal justice in general. It is worth noticing that, despite this dearth of attention, Hegel's theory of punishment has still managed to influence contemporary accounts of criminal punishment and penal justice (e.g., Hampton 1984; Morris 1968; Brudner 2009).

As happens with Kant's theory, there is room for debate on the extent to which Hegel does advance a pure retributivist account. However, whether he coherently introduces non-retributivist elements in his theory is something that this entry will not consider in further detail. Rather, our concern is confined to the retributivist aspects of Hegel's views in relation to punishment. For a recent treatment of Hegel as a mixed-theorist, see Brooks 2012.

Before considering the sense in which Hegel is a retributivist, a general contextual preliminary is in order. In the *Elements of the Philosophy of Right* (henceforth Hegel 1991, followed by the paragraph and page number in the English edition), Hegel is concerned "with the concept of right and its actualization" (Hegel 1991, §1, p. 25). Hegel's project is thus devoted to the manifestation in the social world of the concept of right, which manifests itself from different but interrelated positions that Hegel calls the "Abstract Right," "Morality," and "Ethical Life." The *Elements*, in short, aim to account for the dynamic and connection of these different appearances of right. Hegel's discussion on punishment appears mainly in "Abstract Right" where persons are seen as abstract agents, stripped from their particularities, and capable of manifesting their freedom and self-determination in external things. A consequence of this conception of the individual is that, to treat someone as a person, we are obliged to act in ways that recognize her ability to conduct her life as a free agent in the world. As we will see shortly, this has important consequences for Hegel's retributivism. For a useful synthetic exposition of these general preliminaries, see Dubber 1994.

The essence of Hegel's retributivism emerges with some clarity when he articulates the relationship that exists between crime and punishment. Hegel states in his characteristic style that the "determination of the concept [of punishment], however, is precisely that necessary connection [which dictates] that crime, as the will which is null and void in itself, accordingly contains within itself its own nullification, and this appears in the form of punishment" (Hegel 1991, §101, p. 128). In this passage, Hegel is drawing our attention to the structure of punishment and how it is essentially linked to the criminal act. For Hegel, crime implies punishment and punishment implies crime. This is not to suggest the obviously mistaken point that any time a crime is committed a punishment is meted out. Neither does it advance the view (consistent with Hegel's position) that who is to be punished and to what extent depends on the nature of the criminal act. Hegel is rather pointing to a logical, and thus necessary, connection between these two parts of a single totality: "punishment is merely the manifestation of the crime, i.e., it is one half which is necessarily presupposed by the other" (Hegel 1991, §101 addition (H), p. 129).

Given this argument, it seems natural to elaborate further on the notion of crimes in order to ground the justification of punishment. Indeed, Hegel dedicates various paragraphs of "Abstract Right" to elucidate the nature of crimes. He distinguishes three varieties of wrongdoing: (1) civil wrongs, in which a particular will judges by itself, mistakenly, what is right about a particular case. These judgments are wrongs because they are mistaken judgments, but they do not thereby negate or intentionally disrespect the system of right as such (Hegel 1991, §§84–6, pp. 117–118). (2) Deception, by contrast to civil wrongs, involves the intentional violation of the law. However, in committing this wrong, I am simultaneously recognizing the system of right: I deceive someone by pretending to be acting according to the law (Hegel 1991, §§87–9, pp. 118–119). (3) Crimes are wrongs that negate right. They are neither unintentional, like civil wrongs, nor do they try to deceive others, like deception. Rather, crimes are an attempt to debunk right. But crimes go even further in their negative judgment. When a criminal act is committed, it is not only right as such that is negated (the offender has acted according to a maxim that contradicts right itself), but also the victim's status as a person (her external freedom has been violated), and the offender's right (he can then be treated according to the maxim generated by his action, namely, that one should violate other people's external freedom).

With this account of crimes in place, the logical connection between punishment and crime should become clearer. For Hegel, "Punishment is the negation of the negation" (Hegel 1991, §97 addition (H), p. 123), which is to say that punishment cancels the negations of right involved in the criminal act. To make further sense of this idea we must notice that for Hegel right is absolute, that is, right is something that cannot possibly be canceled. Crimes are thus "null and void," because offenders could never *really* negate right by means of their criminal actions. Punishment simply exposes such a vacuity and reaffirms right as right. Understood in this manner, punishment "is the reconciliation of right with itself"; "this reconciliation applies to the *law*, which restores and thereby *actualizes itself as valid* through the cancellation of the crime" (Hegel 1991, §220, p. 252). Punishment is just "in and for itself" (Hegel 1991, §99, p. 125). (For clarification on Hegel's use of cancelation (*Aufheben*), see Hegel 1969, p. 107. For a critique, see Honderich 2006, pp. 41–45).

The previous analysis allows us to see how punishment reaffirms right as right (punishment reaffirms the validity of the system of law) as well as the right of the victim of crime (punishment denies the universal validity of the right violations entailed in the criminal act). But Hegel also contends that the reaffirmation of right via punishment does justice to the offender. This is the case because punishment is "a right for the criminal himself" (Hegel 1991, §100, p. 126). This is surprising since, as we have seen, punishment involves treating the offender in ways that we normally consider inadmissible. How then could punishment be an offender's right? A characteristic aspect of Hegel's theory becomes salient at this point. In "Abstract Right," persons are defined in terms of their capacity for freedom and rationality, and by punishing offenders – who also happen to be persons (!) – we honor them as rational beings and recognize them as subjects of right. Justice is done, and the right as right is asserted, first, when punishment is imposed, and second, when it is imposed not as the result of a private but of a universal will (Hegel 1991, §104, p. 131). To consider punishment otherwise, as a mere threat or an act of revenge imposed by a private will, would mean to treat "a human being like a dog instead of respecting his honour and freedom" (Hegel 1991, §99, pp. 125–126. See also §102). The criminal "is denied this honour if the concept and criterion of his punishment are not derived from his own act; and he is also denied it if he is regarded simply as a harmful animal which must be rendered harmless, or punished with a view to deterring or reforming him" (Hegel 1991, §100, p. 126). This shows Hegel's retributivist credentials in full light. By being punished, the wrongdoer is treated as he deserves, a rational human being capable of acting on universalizable maxims – in the case of crimes, the maxim of permitting the violation of another person's external freedom. Punishment simply applies the universalized maxim of his offense to the wrongdoer himself.

If all this is sound – namely, that punishment is an integral part of the rational manifestation of right and, thus, it is in itself an act of justice that reaffirms the right that the offender attempted to cancel – then, how and how much should we punish the offender? This is a particularly pertinent question in this context. If punishment is also a right of the offender, as Hegel believes, he must surely have something to say about the amount and types of punishment that the state may legitimately inflict upon the offender. Unfortunately, Hegel says very little about this. Hegel seems aware of the difficulties involved in many interpretations of the *lex talionis*, and thus he proposes what he calls an "equality of value" principle between the past offense and present punishment (Hegel 1991, §101, pp. 127ff). However, this looks like a very unspecified, if not obscure, way to establish how and how much we ought to punish offenders. In addition to some form of ordinal equality, according to which the more serious the offense the harsher the punishment, it is unclear how Hegel means the principle of "equality of value" to be understood. It is not until "Ethical Life," the third part of the *Elements*, that Hegel elaborates a bit further on these matters. In this later stage of the *Elements*, the individual is not depicted as a formal agent, as in "Abstract Right," but as a particular person (Hegel 1991, §182, p. 220). Since for Hegel philosophy is concerned only with general and abstract principles, whereas the precise determination of these principles – one of which concerns the quantum of punishment – is a matter that depends on history and other contingencies (Hegel 1991, §3, pp. 28ff), it is no surprise that he left considerations about the distribution and proportionality of punishment to this later, less abstract, stage of the *Elements*. In short, Hegel believes that how and how much we should punish offenders is a matter to be resolved within the specific society in which punishment is performed. "A penal code [and with it, the definition of crimes and the harshness of punishments] is therefore primarily a product of its time and of the current condition of civil society" (Hegel 1991, §218, p. 251).

Hegel seems to be right. A universal code that establishes, irrespective of the circumstances of societies, conclusive measures and types of punishments for specific offenses is far-fetched. Hence, it seems sensible to adopt something like Hegel's contextual proposal. However, his reflections on the importance of history and circumstances regarding proportionality and types of punishment can only work as the beginning of an adequate answer. Indeed, as it stands, there is nothing in his proposal that may impede it from being hijacked by undesirable consequentialist considerations on the question of proportionality. For example, in a time different from ours, a rational and decent consequentialist judge could conclude, in consistency with Hegel's view, that given his society's current circumstances, not only the stability but also the very survival of society demands torturing terrorists in the public square. He could defend this conclusion by expressing that this is what "equality of value" requires in a time like his. To be sure, Hegel's account may have the conceptual tools to advance an answer that would avoid undesirable outcomes such as this, however, as it stands, it is unclear how this solution would proceed.

Concluding Remarks

In 1939, J.D. Mabbot stated: "the retributive view is the only moral theory except perhaps psychological hedonism which has been definitely destroyed by criticism" (Mabbot 1939, p. 152). Indeed, up to the mid-twentieth century, the established view on punishment was some version of consequentialism, whereas retributivism was dismissed as something akin to superstition. The panorama is now very different. Since the last years of the 1960s, retributivism has come to life again and became an influential form of justification of punishment.

In essence, current retributivism maintains many of the main claims found in traditional retributivist ideas. Some of those claims have been explored here, namely, present punishment can be inflicted on an offender only because he deserves it due to his past offense; wrongdoers must not be used as mere means for achieving a social end; the amount of punishment that can legitimately be inflicted upon an offender depends on some feature of the wrong committed. These are retributivist considerations that, it is fair to say, are well-established principles in most plausible theories of punishment, and rightly so. They adequately capture widespread intuitions about punishment and the criminal justice system in general.

However, the virtues of traditional retributivism should not make us forget the difficulties and gaps we have encountered in the process of exposing this theory. Before finishing, it is worth recapping some of them.

We have seen that traditional retributivism works well as a type of justification of punishment. However, as Hegel himself realizes, this justification cannot be considered, nor successfully grounded, in isolation from wider state institutions and social circumstances. In particular, a comprehensive account of punishment and criminal justice must provide a sufficiently well-defined account of moral and political theory. This is, *pace* Hegel's suggestions, something that traditional retributivism does not advance in an adequate detail or with precision. (For a contemporary attempt to integrate the justification of punishment with broader political and moral theory, see Matravers 2000).

Retributivism in general should also make additional efforts to avoid the charge of advancing an account of punishment that is utterly insensitive to consequences. Critics have asked, with reason, what is the point of punishment – an expensive and morally problematic practice – if it does not bring about any further good. When detached from any positive consequences, is not punishment a mere expression of revenge or uncivilized practices? Without having to deny the core premise of this objection, traditional retributivism may respond that this criticism is ill grounded for two different reasons. First, in the course of this entry it has been mentioned that retributivism is capable of including non-retributivist considerations as part of the justification of punishment. Thus understood, it is not true that traditional retributivism disregards consequences all the way, but rather that it does not locate them as the first and main element in the justification of punishment. Second, traditional retributivists seem to believe that punishment does produce a good, indeed, a significant one. This is especially clear in the case of Hegel: punishment reaffirms right, which in turn is at the basis of the existence of the civil society. Thus understood, punishment is necessary for the maintenance of the polity. Such is the good, not a trivial one, to which punishment contributes (notice that although in this sense Hegel's account is teleological, his justification of punishment is still strongly non-consequentialist). Be that as it may, it is far from clear how these considerations would suffice or be convincing enough. A plausible justification of punishment must not only take consequences into account, but it also needs to emphasize the relevant type of consequences and locate them in the right place within the theory.

Finally, although part of the appeal of traditional retributivism derives from its insistence on the connection that must exist between the present punishment and the past offense, the details of this aspect of retributivism are left very much underdeveloped. The *lex talionis*, in its most plausible formulation, only indicates the need of preserving the relevant connection between punishment and crimes, but it does not tell us much about how to establish this connection. In turn, although affirming the obligation we have to respect the dignity of the offender in sentencing, Kant offers little insight into this issue. In the case of Hegel, things are not very different. He only tells us that this is a matter of detail to be considered in the social and political circumstances in which punishment is effectuated. As they stand, none of these sensible suggestions are very helpful. They work as general guidelines, but do not provide sufficient grounds to direct our practices.

Nothing of this is to say that we should dismiss traditional retributivism. Indeed, it is unclear whether any alternative justification of punishment would do much better than retributivism in these and other issues. Moreover, the difficulties noticed here are not necessarily insurmountable, and retributivism may well develop (and indeed has developed) answers to many of these difficult and problematic issues.

Related Entries

- ▶ Deterrence
- ▶ Proportionality
- ▶ Theories of Punishment

Recommended Reading and References

Bedau HA (1978) Retribution and the theory of punishment. J Philos 75:601–620

Brooks T (2003) Kant's theory of punishment. Utilitas 15(2):206–224

Brooks T (ed) (2012) Hegel's philosophy of right. Wiley-Blackwell, Oxford

Brudner A (2009) Punishment and freedom: a liberal theory of penal justice. Oxford University Press, Oxford

Byrd BS (1989) Kant's theory of punishment: deterrence in its threat, retribution in its execution. Law Philos 8(2):151–200

Cottingham J (1979) Varieties of retribution. Philos Q 29:238–246

Cummiskey D (1990) Kantian consequentialism. Ethics 100(3):586–615

Dolinko D (1992) Three mistakes of retributivism. Univ Calif Law Rev 39:1623–1657

Dubber M (1994) Rediscovering Hegel's theory of crime and punishment. Mich Law Rev 92(6): 1577–1621

Duff RA (2001) Punishment, communication, and community. Oxford University Press, Oxford

Hampton J (1984) The moral education theory of punishment. Philos Public Aff 13(3):208–238

Hart HLA (1959) Prolegomenon to the principles of punishment. Proc Aristot Soc 60:1–28

Hegel GWF (1969) Science of logic (trans: Miller AV). Allen and Unwin, London

Hegel GWF (1991) Elements of the philosophy of right (ed: Wood A). Cambridge University Press, Cambridge

Honderich T (2006) Punishment. The supposed justifications revisited. Pluto Press, London

Hill TE (2003) Treating criminals as ends in themselves. Jahrbuch für Recht und Ethik 11:17–36

Kant I (1991) The metaphysics of morals (trans: Gregor M). Cambridge University Press, Cambridge

Kant I (1997) Lectures on ethics (eds: Heath P, Schneewind J). Cambridge University Press, Cambridge

Kant I (1998) Critique of pure reason (eds: Guyer P, Wood A). Cambridge University Press, Cambridge

Kant I (2001) Critique of practical reason (ed: Gregor M). Cambridge University Press, Cambridge

Lacey N (1988) State punishment. Routledge, London

Mabbot JD (1939) Punishment. Mind 48(190):152–167

Matravers M (2000) Justice and punishment. The rationale of coercion. Oxford University Press, Oxford

Merle JC (2000) A Kantian critique of Kant's theory of punishment. Law Philos 19(3):311–338
Morris H (1968) Persons and punishment. Monist 52:475–501
Murphy JG (1971) Three mistakes about retributivism. Analysis 31(5):166–169
Murphy JG (1987) Does Kant have a theory of punishment? Columbia Law Rev 87(3):509–532
Steinberger P (1988) Logic and politics: Hegel's philosophy of right. Yale University Press, New Haven
Tunick M (1992) Hegel's political philosophy: interpreting the practice of legal punishment. Princeton University Press, Princeton
Tunick M (1996) Is Kant a retributivist? Hist Polit Thought 17(1):60–78
von Hirsch A (1976) Doing justice: the choice of punishments. Hill & Wang, New York
von Hirsch A (1993) Censure and sanctions. Oxford University Press, Oxford
Waldron J (1992) Lex talionis. Ariz Law Rev 34:25–51
Wood A (1990) Hegel's ethical thought. Cambridge University Press, Cambridge

Traffic Enforcement and Control

Joseph A. Schafer
Department of Criminology and Criminal Justice, Southern Illinois University, Carbondale, IL, USA

Overview

Police agencies engage in a variety of traffic services. These efforts are designed to encourage motorists to obey traffic laws and regulations through a variety of methods. It is hoped that such compliance will maximize traffic efficiency and obedience of the law while minimizing traffic disruptions and accidents. The traffic services provided by state, county, municipal, and special jurisdiction law enforcement agencies are frequently associated with a range of organizational objectives. Though enforcement of traffic laws is a major objective, traffic services are also provided with the intention of minimizing injuries and property damage due to traffic accidents, determining facts when accidents and collisions have occurred (both to determine fault and to provide data for accident prevention efforts), assisting accident victims, obtaining compliance with traffic laws, and maximizing efficient use of congested roadways (Sheehe 1971). Traffic services include efforts to enforce traffic laws, regulate and direct the flow of traffic, minimize traffic accidents (both volume and severity), and efficiently respond to accidents when they do occur.

After providing generalized patrol services, traffic enforcement is one of the major ancillary tasks undertaken by local police agencies (Walker 1992). Citizens report that traffic stops are the most common situation in which they have contact with the police (Durose et al. 2007). Data suggest that nearly 10 % of citizen calls to the police relate with a traffic problem of some sort (most typically reporting traffic accidents but also to complain about traffic violations or traffic flow problems) (Scott 1981). Other studies have found that nearly one-quarter of patrol officer time is spent on traffic enforcement activities (Greene and Klockars 1991). Traffic services and enforcement serve a crucial secondary purpose for police officers and organizations. For officers, traffic enforcement may have less to do with the enforcement of traffic laws and more to do with providing an officer legal grounds to look for more serious violations (Schafer and Mastrofski 2005). Though traffic services may not seem glamorous and are often downplayed in media portrayals of policing, they represent a range of tasks that are central to policing, both in terms of their frequency and importance.

Fundamentals

Agencies vary considerably in the extent to which traffic services are a specialized or generalized function (Walker 1992). Specialized traffic services assign traffic enforcement, traffic regulation, and/or traffic accident investigation to specific officers who are usually freed from other policing duties, such as responding to emergency calls for service. This situation is increasingly common as the size of an agency increases and specialization of all types begins to be more evident within the organization's structure.

Specialization of any policing task is presumed to increase the efficient use of personnel resources and to free general patrol officers from time-consuming tasks not directly related with crime control. The presence of a specialized traffic service function does not mean general patrol officers will not engage in any traffic-related activities; rather, having a specialized function would be expected to reduce the amount of time general patrol officers spend on these activities, at least during the times when specialized officers are working.

Generalized traffic services responsibilities occur when all officers are expected to routinely engage in traffic enforcement, traffic regulation, and accident investigation. This is more likely to be noted in medium and small agencies, where specialization is less common or the extent of specialization is modest. For example, in a 30-officer department, 1 or 2 officers working the day shift might be given responsibility for traffic services. During other times of the day, general patrol officers would be expected to provide those services. There may be additional variation in how agencies assign traffic service duties, even where a specialized unit is present (i.e., a traffic unit might handle traffic control and traffic accident investigation, while all officers are expected to engage in routine traffic enforcement). In agencies that place a heavy emphasis on traffic enforcement, such as state police and highway patrol, traffic services may be a generalized responsibility even when the agency is quite large.

Though contemporary evidence is sparse, some research suggests agencies produce a higher volume of traffic enforcement activity when traffic is a specialized function (Gardiner 1969). A variety of reasons could explain this situation. In larger agencies, general patrol officers may not have the time to engage in traffic services because of a high volume of calls for service. General patrol officers may also lack some resources in support of traffic enforcement; for example, radar units might not be installed in vehicles used by general patrol officers and/or not all officers may be certified (if required by the state) in the use of radar equipment. There might also be an absence of a culture of traffic services in a given jurisdiction; traffic services might be viewed as a "soft" assignment, something an officer does when she/he wants to work the day shift or escape the "grind" of routine patrol.

Traffic Enforcement

When most citizens think about police efforts centering on traffic services, it is likely that traffic enforcement is the first activity that comes to mind. Traffic enforcement and regulation efforts represent a curious convergence of circumstances in the understanding of police organizations and discretionary choices of police officers (Lundman 1979). Organizations have various incentives to regulating the volume of enforcement efforts by personnel, including the desire to ensure public safety, to control crime, and to respond to public demands for action on select offenses (i.e., drunk driving) (Mastrofski et al. 1987). Within those circumstances, however, traffic services remain a highly discretionary activity for officers and, at times, a low visibility activity (Brown 1981). Though officers may have to meet certain overall expectations for their production and provision of various traffic services, they often enjoy considerable latitude (sometimes granted by state law) to decide when to invoke the law. Situations where officers show such leniency are rarely captured in existing police data systems and are rarely known to anyone aside from the officer and the motorist.

Traffic enforcement activities can be the source of tension, danger, and inconsistencies for the police and the public. Citizens may resent being stopped and subject to extra attention by the police, even if a citation is not issued. It should be noted, however, that the overwhelming majority of citizens stopped by the police report that they believe the police had legitimate reason for initiating that contact (Durose et al. 2007). Officers killed or injured on the job are often involved in traffic service activities. Around 13 % of police fatalities and nearly 10 % of assaults on the police occur while officers are engaged in traffic stops (Lichtenberg and Smith 2001). For decades researchers have noted inconsistencies in the traffic enforcement actions of agencies. James Q. Wilson first noted this variation in his classic text *Varieties of Police Behavior* (1968), observing that agencies of similar size and jurisdiction often vary considerably in the rates with which they engage in traffic stops and, when stops are made, take formal enforcement action by issuing a citation. Some agencies demonstrate low levels of enforcement and/or respond to traffic violations with verbal warnings, while other agencies are quite actively involved in traffic enforcement, including the issuance of a high rate of citations (Gardiner 1969; Mastrofski et al. 1987; Wilson 1968). This may contribute to citizen frustration, as it is not always clear when and where a traffic offense will invoke a formal police response.

Explaining Variation in Traffic Enforcement

Variation in the frequency of engaging in traffic stops and issuing traffic citations can be explained by a number of potential factors. In large part, such variation is likely a reflection of differing department objectives and community expectations. Though rarely discussed, in many communities traffic enforcement often contributes appreciably to city and county coffers (Radelet and Carter 1994). When a traffic citation is issued and the motorist is subsequently ordered to pay a fine, the issuing agency is not the direct recipient of those funds. However, the overriding governmental unit (the city or county) often does receive a proportion of the fine paid by the motorist. Particularly in difficult economic times, governmental units can become dependent on a certain level of resources. This might not pressure an officer to issue a citation on any given traffic stop, but it might result in implicit pressure on an agency to maintain or even increase the volume of citations issued over time. Some agencies might try to limit local opposition to heavy traffic enforcement by targeting efforts in areas where out-of-town drivers are more likely to be caught committing a violation (i.e., encouraging officers to "run radar" on a state highway).

High levels of traffic stops and/or traffic citations can also be a reflection of myriad noneconomic pressures. For example, a community may be calling upon the police to increase traffic enforcement, particularly for select offenses (i.e., speeding) or in specific areas (i.e., near schools or commercial area). Top agency administrators may request officers to modify their behavior in specific ways, such as requesting that officers increase the number of traffic stops or show more restraint in issuing traffic citations. Calls for increased traffic stops or enforcement may reflect a belief that such actions will enhance public safety, though evidence demonstrating a clear ability of the police to achieve that outcome is mixed (see below). Leaders may also formally or informally ask officers to increase traffic enforcement because traffic stops are one way agencies can document that officers are working. Agencies will be reluctant to refer to these expectations as "quotas," but it is not uncommon for agencies to establish "productivity standards" (or similar expectations) based in part on the frequency of traffic stops and/or citations for each officer. Given the limit direct supervision most officers receive, a traffic-related productivity standard is one way supervisors and agency leaders can ensure that officers are being vigilant in engaging in work-related activities while on routine patrol (Walker 1992). Agencies cannot easily assign officers to make a given traffic stop (as an agency might assign an officer to handle a specific complaint or take a specific report), but expecting officers to document a select number of traffic stops per shift or issue a certain number of citations or warnings per month is thought to demonstrate a minimal level of vigilance on the part of each officer. Such expectations are predicated on the assumption that officers can routinely identify legal cause to stop motorists for violations or other pretense.

The volume and nature of traffic enforcement behavior by officers has been noted to vary not only across agencies, as discussed earlier, but also across officers. Within the cultural and organizational expectations regarding traffic enforcement, officers may opt to merely meet standards or choose to exceed organizational expectations (Schafer and Mastrofski 2005). This reflects an established reality in policing research that demonstrates that officers vary in when, where, and how they use discretion in performing their work-related duties (Brown 1981). In the context of traffic enforcement, some officers will have a strong inclination to engage in a high rate of traffic stops and issue a large number of citations. This vigor may be due to the belief that such activities will either ensure greater control in the community or because of a desire to remain active during unassigned duty time. Other officers will engage in these activities with enough frequency and vigor to satisfy the expectations of their supervisors and organization, but will not go substantially beyond that threshold. Still other officers may engage in a high rate of traffic stops while issuing few citations. For these officers, traffic stops have instrumental value; they allow officers to initiate contact with citizens to investigate suspicious or unusual circumstances with the ultimate objective of detecting criminal conduct. Though organizations and communities can exert some influence, individual officers maintain some latitude to decide how they prioritize traffic enforcement efforts in the performance of their duties.

Nonenforcement Traffic Services

In addition to enforcing the law, police officers and organizations engage in other traffic services aimed at regulating the flow of traffic and responding to traffic accidents. Nonenforcement traffic services can also generate tension and dangers for the police. Directing traffic is a common policing activity, both in special circumstances (i.e., at a special event or during the aftermath of a traffic accident) and as a routine assignment (i.e., directing traffic near a school or at busy urban intersections). Motorists may be frustrated with the perceived inflexibility of officers or delayed movement of traffic. A driver may feel ill will toward the police because they are not allowed to take their preferred route. Officers may become frustrated while working long assignments in full uniform, with full exposure to the elements. Citizens may perceive the police are exacerbating traffic problems, and officers may perceive shortcomings in the cognitive and driving abilities of the driving public, though neither of these outcomes will be universally present.

In addition to enforcing motor vehicle laws and regulating the flow of traffic, an important traffic service function performed by the police is the investigation of accidents. Traffic crashes are a leading cause of death (causing more annual deaths than crime) and generate a tremendous economic impact. As a matter of public safety, officers are tasked with responding to traffic accidents in order to facilitate medical care, regulate the flow of traffic while a matter is resolved (to minimize the likelihood of a secondary accident occurring), coordinate the removal of disabled vehicles, and assist the involved parties. Officers will generally question the motorists and any witnesses, examine the vehicles involved, note the position of vehicles on the roadway, and in some cases take photographs, collect physical evidence, or otherwise examine and record the existing evidence (i.e., skid marks). These investigative efforts are intended to support officers in determining how an accident occurred, what took place, and, where relevant, how to assign blame for the accident. The latter point becomes relevant if it is determined that enforcement actions are needed but also to support decisions of responsibility relevant to insurance and other financial and legal claims that may arise as a result of an accident.

Data generated by police accident investigations and reports can also be used to support traffic safety enhancement efforts. On a local level, agencies use data to identify high-risk intersections. This information might shape traffic regulation decisions within a jurisdiction, such as changing the pattern of a traffic light or installing some form of traffic regulation technology. Local data might further be used as a rationale for increasing enforcement efforts in a particular area. On a large scale, traffic data might be used to inform traffic laws or policies. Knowing who is often involved in accidents, when those accidents occur, and the nature of those accidents can all provide support for changes. For example, a state might opt to lower speed limits on certain road, restrict select driver behavior (i.e., texting while driving), or modify who is allowed to operate certain types of vehicles (i.e., allowing additional testing to authorize someone to operate a motorcycle). Though large volumes of data amassed by street-level officers often are used to justify these policy changes, it is not always clear those changes achieve the intended outcomes (typically a reduction in the frequency and severity of traffic accidents).

The Effectiveness of Traffic Services

The efficacy of traffic services and enforcement is subject to some debate. The relationship between police enforcement behaviors and actual reductions in crime, traffic violations, and traffic accidents has not been conclusively established. Steven Rock (1989) evaluated a traffic enforcement program that was implemented to reduce traffic accidents in Batavia, Illinois. Rock's analysis suggested that accident rates did decline in Batavia, though perhaps not for the expected reason. He attributed at least some of the decline to improved accident investigations conducted by Batavia officers, which in turn helped generate data that allowed the city to redesign intersections where a high rate of traffic accidents occurred. Thus, the efforts of the Batavia police did not directly influence the behavior and actions of motorists. Instead, the effect was indirect; the police influenced traffic engineering and roadway design, which influenced driver behavior and resulted in fewer accidents.

It remains unclear how police should direct efforts to reduce traffic violations and associated problems (i.e., the volume and severity of accidents) (Scott 2001). Though the limited available research has tended to focus on reducing traffic issues through the use of enforcement, other mechanisms might be presumed to achieve the same outcome. Citizens can be educated using signage, media alerts, and the distribution of literature. Potential violators might be deterred through visible police presence and/or a heightened volume of traffic stops, even if enforcement (i.e., traffic citations) is not an outcome of those efforts. Roadways might also be engineered to improve the flow of traffic through high-accident areas and/or by building design elements (i.e., "traffic calming" features such as narrowing roadways, intentionally creating curving roadways, or creating speed humps) into roadways to make violations less likely. Tactics deemed to be less effective in reducing traffic violations and accidents include decreasing speed limits, increasing fines and penalties, and reengineering vehicles (i.e., the use of speed governors) (Scott 2001).

Traffic enforcement is a common element in police efforts to reduce crime and violence (McGarrell et al. 2002). In such instances, traffic stops are not used with the explicit intension of reducing traffic violations. Instead, traffic stops are used to provide lawful pretext for officers to stop vehicles, either in general or when the vehicle is deemed to be "suspicious." Officers are likely focused more on identifying motorists, running want/warrant checks on vehicles and their occupants, and obtaining lawful authority to search vehicles and their occupants. Such authority might be obtained by a citizen granting consent, when officers determine they have lawful authority to conduct a *Terry* search, or when other legal circumstances are present (i.e., contraband observed in plain view, incident to an arrest, or incident to a vehicle being towed). When the police engage in this type of traffic enforcement, the objective is usually to reduce crime and disorder, not to influence or regulate driving behavior. The traffic stop becomes the legal means to achieve some other end objective.

Only a few quality studies have attempted to assess whether aggressive police enforcement of traffic laws can achieve a meaningful reduction in crime rates. Early studies of this phenomenon supported that areas with higher levels of traffic enforcement had lower rates of robberies (Sampson and Cohen 1988; Wilson and Boland 1978). In contrast, Weiss and Freels (1996) found that increasing traffic enforcement in high-crime areas did not influence robberies or auto thefts nor did it increase the frequency of felony arrests, in Dayton, Ohio. Their analysis also could not support that higher enforcement resulted in more arrests for drug crimes, weapons crimes, or DUI.

Despite the common presumption that increasing the vigor of traffic enforcement will engender ill will toward the police, some studies have found the opposite to be true. This situation may be a function of the traffic enforcement methods used by the police. If the police seek to influence the behavior of the driving public, they might do so by increasing the issuance of traffic citations. If the police seek to influence crime and disorder, they might increase the frequency and investigative rigor of traffic stops, without actually issuing more citations. In such circumstances, traffic stops are not the end objective but rather a means to that objective. Research has found citizen support for more aggressive traffic enforcement actions by the police when those actions are undertaken in the name of community safety and order. This finding is tempered to some extent by the reaffirmation that while aggressive strategies might enhance citizen perceptions of the police, they might achieve that outcome without actually influencing crime (Chermak et al. 2001; Weiss and Freels 1996). In other words, while citizens might support aggressive traffic enforcement efforts initiated in the name of reducing crime, those efforts have not conclusively been shown to influence crime rates.

Key Controversies

Traffic services have been a contentious aspect of police-community relations, since police agencies first assumed responsibility for investigating accidents, enforcing motor vehicle laws, and regulating the flow and movement of traffic. Citizens presumably recognize the need for, and value of, traffic enforcement and regulation by the police. Citizens may not, however, appreciate being the specific recipient of such police attention and intervention (Lundman 1979). A motorist might question why an officer is “hassling” them for failing to completely stop at a stop sign, instead of addressing “real crime” within the community. Despite the presumption that aggressive traffic enforcement efforts will engender citizen ill will toward the police, studies have not always found that to be the case. An aggressive enforcement strategy used by the Indianapolis Police Department was implemented as a means of targeting select high-crime neighborhoods. Citizens in these neighborhoods were found to be supportive of the police department’s efforts (Chermak et al. 2001), though it is worth noting that officers were instructed not to simply conduct more traffic stops but to engage in more targeted enforcement efforts and to conduct more thorough contacts with motorists they stopped.

Corruption

A number of investigations into police corruption have involved issues related with traffic services (see The Knapp Commission Report on Police Corruption 1972). This problem has included officers accepting bribes to not issue citations, police personnel being paid to “fix” traffic citations and parking tickets, police soliciting payment for providing favorable traffic-related services, and companies paying police officials for favorable consideration (i.e., tow truck companies paying to be called more often when citizens needed motor vehicle assistance). Across the country, agencies of all sizes and types have allegedly been involved in such scandals. The problem appears isolated to very small numbers of officers and non-sworn personnel but demonstrates the financial motivations citizens may have to avoid legal sanctions or secure more favorable treatment with traffic-related matters.

Racial Profiling

Traffic enforcement behavior reached a new level of scrutiny in the mid-1990s with the emergence of concerns over racial profiling (also referred to as “Driving While Black” and “Driving While Brown,” among other labels). Though the practice of racial profiling is often seen to have deep roots in the dynamics of American criminal justice, this behavior achieved a more tacit level in the 1980s when the Drug Enforcement Administration (DEA) began to train local law enforcement personnel under the Operation Pipeline program (Buerger and Farrell 2002). Operation Pipeline was predicated on the belief that drug traffickers brought their goods into major port cities using various methods of bulk transportation. From these ports, drugs were distributed throughout the country using couriers who normally used cars to transport their deliveries. The DEA developed a “profile” of the behaviors and other characteristics of drug couriers, including that drivers were often members of minority groups. Though not an explicit endorsement of using race as a factor in deciding whom to stop, Operation Pipeline is often considered to have legitimized what some view as a long-standing practice of incorporating race into police enforcement decisions and practices.

Racial profiling received considerable attention beginning in the 1990s as a number of court cases and media exposés highlighted disparities in the characteristics of citizens stopped and searched by various departments (Weitzer and Tuch 2006). Prominent among these anecdotes was *State of New Jersey v. Pedro Soto*, et al. (1996), a court decision that incorporated statistical analysis of behavior exhibited by officers patrolling the New Jersey Turnpike. In particular, a rolling survey of turnpike users and violators found that African American motorists accounted for approximately 15 % of drivers on the turnpike, but accounted for 46 % of violators stopped by police on that roadway. In the first decade of the new millennium, dozens of states created laws regarding racial profiling, and thousands of agencies began collecting data regarding traffic stops in their jurisdictions. Some of the latter efforts were voluntary, while many others were compulsory under state law.

The net result of these laws and data collection efforts remains largely unclear. A wide variety of reports and studies were published in the 2000s seeking to examine whether and how driver characteristics influenced police discretion in traffic stops. Many of these studies were based on data that suffered from a range of limitations and often data were from a single agency and/or a relatively short period of time, making conclusions tenuous. The majority of studies lacked data of sufficient validity to help determine whether minority motorists were more likely to be stopped when compared with nonminority motorists. A few studies suggested that young, male, and minority drivers were more likely to commit more serious traffic violations, which could explain at least some disparities in patterns of who the police stop for traffic offenses (Lundman and Kowalski 2009). The studies generally did suggest minority drivers (especially those who were young and male) were more likely to be searched when stopped by the police, though many datasets did not allow researchers to clearly control for the reason a search was conducted. It was often noted that minority drivers were less likely to be found in possession of contraband when searched, though many studies did not include data on the specific type and quantity of illicit materials and goods. The net result is that racial profiling clearly remains a matter of concern to the public, politicians, and police officials, but the available data provide a very incomplete picture of the extent to which this problem exists and, if so, how profiling actually manifests itself across the range of choices officers make during traffic stops.

Technology-Assisted Traffic Enforcement

The introduction of technology-assisted traffic enforcement efforts has also become a controversial practice. Communities have used video-capture technology and other applications to detect violations and automate the issuance of citations for offenses such as running red lights and stop signs, as well as speeding. Often these systems operate in a "passive" manner without the real-time presence of an officer and the use of a traffic stop. The camera/radar system detects and records the violation, along with the identity of the offending vehicle. The registered owner(s) of the vehicle often receives a citation by mail with some documentation of the offenses in question. In many communities, the expense of installing, maintaining, and operating these systems has been assumed by a private entity. A jurisdiction allows a private corporation to place enforcement devices in the community; that corporation receives a portion of the revenue generated by operating the device. The privatized nature of enforcement and the high revenues automated systems can generate have made the use of technology-assisted devices quite controversial. At times these systems are seen being more focused on generating revenue and profits than actually improving public safety.

Future Directions

Traffic services remain an integral part of modern policing. Traffic regulation and enforcement consume a large proportion of uniformed policing resources and attention. Traffic stops are the most common venue in which the police and public interact; they also provide important pretext to justify police-citizen contacts when authorities are primarily interested in identifying and intervening in more serious criminal matters. These situations are not likely to change in the immediate future.

Recommended Reading and References

Brown MK (1981) Working the street: police discretion and the dilemmas of reform. Russell Sage Foundation, New York

Buerger ME, Farrell A (2002) The evidence of racial profiling: interpreting documented and unofficial sources. Police Q 5:272–305

Chermak S, McGarrell EF, Weiss A (2001) Citizens' perceptions of aggressive traffic enforcement strategies. Justice Q 18:365–391

Durose MR, Smith EL, Langan PA (2007) Contact between the police and the public, 2005. US Department of Justice, Washington, DC

Gardiner J (1969) Traffic and the police: variations in law enforcement policy. Harvard University Press, Cambridge, MA

Greene JR, Klockars CB (1991) What police do. In: Klockars CB, Mastrofski S (eds) Thinking about police: contemporary readings. McGraw-Hill, New York, pp 273–285

Lichtenberg I, Smith A (2001) How dangerous are routine police-citizen traffic stops? J Crim Justice 29:419–428

Lundman RJ (1979) Organizational norms and police discretion: an observational study of police work with traffic law violators. Criminology 17:159–171

Lundman RJ, Kowalski BR (2009) Speeding while black? Assessing the generalizability of Lange et al.'s (2001, 2005) New Jersey Turnpike speeding survey findings. Justice Q 26:504–527

Mastrofski SD, Ritti RR, Hoffmaster D (1987) Organizational determinants of police discretion: the case of drinking-driving. J Crim Justice 15:387–402

McGarrell EF, Cherkmak S, Weiss A (2002) Reducing gun violence: evaluation of the Indianapolis Police Department's directed patrol project. US Department of Justice, Washington, DC

Radelet LA, Carter DL (1994) The police and the community, 5th edn. Macmillan, New York

Rock SM (1989) Impact of the 1980–1983 Selected Traffic Enforcement Program on traffic accidents in Batavia, Illinois. Eval Rev 13:409–419

Sampson RJ, Cohen J (1988) Deterrent effects of the police on crime: replication and theoretical extension. Law Soc Rev 22:163–189

Schafer JA, Mastrofski SD (2005) Police leniency in traffic enforcement encounters: exploratory findings from observations and interviews. J Crim Justice 33:225–238

Scott EJ (1981) Calls for service: citizen demand and initial police response. Government Printing Office, Washington, DC

Scott MS (2001) Speeding in residential areas. Problem-oriented guides for police series no. 3. US Department of Justice, Washington, DC

Sheehe GH (1971) Police traffic supervision. In: Eastman GD, Eastman EM (eds) Municipal police administration. International City Management Association, Washington, DC

State of New Jersey v. Pedro Soto, et al. (1996) 324 NJ Super. 66, 734 A.2d 350

The Knapp Commission Report on Police Corruption (1972) Commission report with summary and principle issued recommendations. George Braziller, New York

Walker S (1992) The police in America: an introduction, 2nd edn. McGraw-Hill, New York

Weiss A, Freels S (1996) The effects of aggressive policing: the Dayton traffic enforcement experiment. Am J Police 15(3):45–64

Weitzer R, Tuch SA (2006) Race and policing in America: conflict and reform. Cambridge University Press, Cambridge

Wilson JQ (1968) Varieties of police behavior. Harvard University Press, Cambridge, MA

Wilson JQ, Boland B (1978) The effect of the police on crime. Law Soc Rev 12:367–390

Traffic Police

Quentin Deluermoz
Quentin Deluermoz, Université Paris 13,
Sorbonne Paris Cité, CRESC/Pleiade,
Paris, France

Overview

Despite its seemingly obvious denotation, the phrase "traffic police" is far from self-evident. One may see it in the charts describing the structure of national, local or municipal police forces, but the term may as well be never written in the records of organizations where, functionally, a traffic police does exist. In its broadest extension, the phrase denotes the group responsible, within the police organization of a state, for the management of traffic and the safety of individuals – be they drivers, passengers or pedestrians – during their travel and transportation. In turn, "traffic policing" refers to the activity associated with this body of officers. It includes the definition of problems by higher ranking officials or experts, the implementation of technical solutions and the enforcement of the law, up to the procedures for actions taken on the field. The importance of traffic police and traffic policing is evident, given the place of mobility as a major element in contemporary societies: traffic, the structural environment facilitating it, and the formal and informal rules regulating it have a profound impact upon our understanding, use and organization of space. These modes of action thus have become routine aspects of any space, so much so that they are the first phenomena identified by a foreign traveler.

It is then a paradox that the body of scholarly work focusing on this branch of policing as such is very thin. Studies on the matter are diverse in perspective and discipline (history, sociology, political science, criminology, law, psychology), which renders all efforts toward a synthetic work less easy, but also makes it all the more needed and interesting.

A Constructivist History of Traffic Police

It is commonly claimed that the problem of traffic regulation is as old as traffic itself. The existence of a body of officials specifically appointed to the control of this activity, however, is not that old. For historians studying the matter, a number of assumptions must be questioned. "traffic police," as defined above, emerged during the seventeenth and eighteenth centuries in western capital cities. In Paris, for instance, it was linked to two phenomena: on one hand, an increase in trade and exchanges, the idea that a city can be seen as an organism with a healthy circulatory system, and the speeding of the urban pace (going from walk to full gallop); on the other hand, the development of "police" forces in the modern acception of the term, leading to more officers paid by the state, and the cultivation of autonomous knowledge and practices. *Commissaires* were initially responsible for the enforcement of traffic regulations and the protection of freedom of movement, then a brigade of the newly visible police created in 1829 was appointed to this duty. The same effort is also at play when London's Metropolitan Police was created that same year. The role of this police and its officers became increasingly important throughout the following century, as traffic became more and more significant, and the road and railroad systems were being developed. The police force took various forms throughout the country, and local regulations vary, leading to a "patchwork of recommendations" (Roche 2008), both precise and varied.

The situation changed in the late nineteenth century, for several reasons: the (slow) establishment of automobile as the dominant form of transportation, leading to changes in the perception of speed; the fear of accidents (numerous early on, and rapidly becoming more frequent); the rising intolerance for traffic congestion, earlier seen as a necessary evil. Changes in political and police response were also numerous: the number and visibility of traffic agents rose, more ambitions laws were drafted (*Code de la route* in France in 1921–1922, Road Traffic act in the United-Kingdom in the 1930s). Urban landscape and road network underwent transformations that reflected the growing use of both personal and collective motor vehicles. To the ongoing rationalization of streets and roadways (sidewalks, definition of street width and pavement material) were then added roundabouts, one-way streets, and traffic lights – the very first of which was installed in Cleveland, Ohio, in 1914, the first Parisian light coming in 1923. In a sense, a new traffic regime was substituted to the old one, with agents of traffic policing taking part to the change. Other transformations took place after 1945, as automobiles became available to the masses, urban population increased, land use planning gained importance, technical innovations appeared, and "traffic safety" is redefined.

This brief historical overview could be written for other locations, following the particulars of each. Three conclusions can be drawn from these works: first, traffic police is not, as it is often noted, a simple adaptation of the police to a linear development of modes of transportation. Rather, one can observe configurations of social, political, technical, legal or cultural issues, within which arise specific problems and solutions. Secondly, each configuration leaves, when being replaced, its imprint on the next – especially visible in the development of roadways, or in the institutional structures owing to police structure. Such layering can explain the complexity of current situations. Lastly, the definition of traffic police is far from self-evident, and is inseparable from a group of phenomena and external influences playing decisive roles in issues of roadworks and transportation: elected officials, decision-makers, road-users, engineers... Traffic police is but one element in a larger network, and from it draws its specificity.

Within Police Organization: Between Specialization and Globalization

The place of traffic police within police organizations gained importance in the 1960s and 1970s, when road transportation increased drastically, and developing countries started to catch up. In several industrialized countries, traffic police personnel make up about 10–15% of the police force (Carnis et al. 2004), overall a non negligeable branch of the force.

As can be seen in the rare studies of the organizational aspect of traffic police, it still remains difficult to propose a fixed definition of it. This is chiefly because of the broad spectrum of tasks it is expected to perform, and prevents one single specialization from being pointed out. Those tasks include the safety of individuals on streets and highways, control of vehicle speed, radar surveillance, accident investigations, schoolchildren education, prevention of traffic congestion, direction of traffic, verification of proof of insurance, help to distressed road-users, investigation of auto theft, policing of individuals, surveillance of national security and drug trafficking.

A recurring question, linked to this aspect, is that of the relation of traffic police to other, sometimes more visible or legitimate branches of the police force (such as criminal police). Furthermore, the particulars vary from one corps to another. The very concept of “traffic policing” has its own history, connected to American police officer turned researcher Franklin Kreml (1907–1998). Among the first scholar to identify as a Traffic Safety researcher, Kreml created in 1936 the Northwestern University Traffic Institute (NUTI) and the Traffic Safety Division of the International Association of Chiefs of Police (IACP). NUTI defines, and circulates throughout the country, knowledge and training adapted to traffic issues. This could explain the stricter definition of the term in the English-speaking world.

Traffic branches are also diversely integrated within their institutional environment. They may take the form of an autonomous corps devoted to road policing (like the California Highway Patrol, one of the most accomplished forms of traffic police), be a branch only of the whole organization (like in New Zealand), or be an important activity without a specific service (like in France, where the *gendarmerie* plays a major part). Along the same lines, this police raises with particular sharpness the issue of its relation with other actors of civil society (victims associations, coalitions of contractors or insurers) who participate to some degree in the production of order of the road. Obvious in the United States, this phenomenon can also be observed in Ghana where the private sector is organized into a Transport Union defining the conduct codes and paying wages to guards overseeing drivers. Lastly, this police also bring up with intensity the old debate on the relationship between local and national levels of action, on allocation of road networks, modes of police administration, or inside the kind of State’s organization (federal, centralized, strong or weak). These forms of policing thus appear as an indicator of the complex and layered administrative and political structures marking the action of policing organizations. This raises the issue of the forms of specialization and definition of the traffic policemen’s professional identity. Research on this domain is only in draft form, but suggest that this identity would depend on the existence of a well identified service, the need for training, the wearing of a specific uniform, or the institutional recognition of the specificity of road surveillance.

Paradoxically, while deeply impacted by local and national contexts, this sector is singularly open to the international standardization of methods and techniques. It is far from a new turn of event, as can be exemplified by the internationalization of road signage (McShane 1999): inaugurated in the United States in the mid 1910s and the 1920s, then standardized within various organizations (League of Nations, Institute of Traffic Engineers), traffic lights and street signs are adopted by most industrialized countries after 1945. The exchange and large scale adaptation of signage is organized by more or less formal political, industrial or police networks, with global activities (such as tourism) as an incentive. This also concerns developing countries and former colonies who were confronted to a rise in traffic accidents following the rise, after World War II, of the automobile, and who sought outside help. The Egyptian ministry of interior, for instance, followed in the 1970s the advice of the Overseas Unit of the Transport and Road Research Laboratory (TRRL) to set up the best adapted tools for road policing. The works on this transnational dimension still remain limited in number, and connected to a diffusionist model or to the mere acknowledgment of socio-economic inequalities: matters of adaptation, potential drawbacks, or issues of sovereignty are rarely studied.

The best example of the phenomenon is probably the standardization of training, and the ubiquity of traffic radars and traffic lights, which are systematically integrated even in the less developed countries and those where traffic policing is nearly non-existent. This technical aspect of traffic policing is probably the most globalized. Far from excluding each other, globalization and integration into social and administrative layers specific to a nation rather go hand in hand. As these researches point out, the study of traffic police then seem all the more interesting, for both the field of police studies and that of the sociology of organizations.

Managing the “Public Problem” of Road Safety

The importance of traffic police also comes from its role in now very sensitive issues, that are all at the same time human, economic and symbolical issues. Road accidents are one of the chief causes of mortality in a number of countries, especially among the younger population – any given year, they are the cause for 1.3 million death and between 20 and 50 million injuries. Traffic accidents also reflect inequalities in development: in 1978, the number of death per 100,000 inhabitants was 238 in Nigeria, 60 in Pakistan, and below 40 in Korea, and below 10 in Western Europe, North America and Japan. This pattern is still current, judging from the figures of death per billion of vehicle kilometer – in 2011, it was 8.5 in the United States and Canada, 39.8 in Romania, 310 in the United Arab Emirates, and 106,6 in Yemen.

Yet these data have to be in relation to the context of the social representations that give them meaning in each country, and that organize the public issue of road safety, as was noted by a well-known study on driving and alcohol intoxication (Gusfield 1981). The emergence of the problem of deadly accidents in late twentieth-century Europe is at first tied to the arrival of motor vehicles than to a real hike in the number of killed – which comes at a later date. Studies of the social representations of automobile traffic and its regulations show the evolution in the points of concern, made obvious in the type of figures chosen, or by the shifting focus of the press: to the insistence on human losses is added the economic toll and the lamentation over the time lost in traffic, before the recurring theme of “road violence” come back with strength. The media coverage also tends to criminalize irresponsible road behavior. It is significant that most industrialized countries have recently substituted the notion of “traffic lawlessness” to that of “road safety.” These two aspects – number of victims and perception of the issue – go hand in hand: the problems caused by the increased number of motor vehicles after 1945 also reflect mutations within the society, and the transformation of the questions the society adresses to itself. Automotive traffic, as noted by Joseph Gusfield, is thus marked by the social and cultural contexts of the spaces it modifies.

This aspect is in direct relation to the activity and structure of the traffic police: first, it presses for a response to traffic problems, and requires solutions adapted to the type of concerns it raises. Second, traffic policing institutions are also responsible for the production of symbols and discourse. Those may concern legitimation – the choice of insignia, symbols, uniforms, promotion campaigns for the force, for modes of intervention or modes of transportation – but may also be the very instruments of police actions, as can be seen in road traffic safety campaigns striving to leave its mark on the public and target at-risk road users. This symbolic aspect concerns all visible police branches, but the social and global magnitude of these issues give it a particular importance in traffic policing. These efforts draw attention to the relation between traffic policing and its social environment, which takes nothing away from the reality of the issues.

Man or Machine? Questioning Police Technology

The weight given to solutions to traffic issues, for the sake of economic continuity and the safeguarding of human lives, is easily understood. This aspect of police activity has been the primary focus of several research studies, which have evaluated the efficiency of policing, and tried to provide police intervention with more better performing tools.

The means of public action are varied: the first is the legal aspect – voting of bills, definition of violations and crimes, choice of penalties. The second mean is enforcement, and as such is directly tied to policing: increase in the number of agents, compilation and teaching of adapted policing knowledge, choice of modes and location of intervention (possible mobility, on foot, motorbikes or in cars). Intervention can be of punitive or preventative nature: one important aspect of police action resides in education, as can be seen in the constant work of police officers within the school system. Such educational focus proved a major aspect of the effort to establish the sharing of the road system and the statute of the pedestrian, in the early twentieth century, and is still ongoing.

The fundamental principles of road traffic safety, expressed as early as 1923 by Director of the Kansas City Safety Council Julian Harvey, and never contradicted since then, can be summed up in what is called the 3 "E" rule: Education, Enforcement, Engineering. The latter is important too, and takes on various forms. A number of psychology studies, for example, through the factorial analysis of answers to questionnaires, focus on the identification of social groups that are the most likely to commit infractions, on determining external, involuntary causes (state of the vehicles and roads, presence of driving schools) and the emotional motivations (aggressiveness in young males, irritability caused by the daily experience of traffic congestion). The goal of these studies is to better target police response. Research conducted in two Chinese cities (Beijing and Chengde) concludes for example that potential countermeasures should be sought among young women and thirty-somethings (Xie and Parker 2002). Cognitive psychology, for its part, is more focused on the determination of causes to loss of attentiveness, and on the stimulation of the ability to focus on the road. Lastly, engineers – administrative or urban engineering, transport or computer systems specialists – propose innovations aiming to facilitate vehicle flows and reduce the number of accidents through changes to the external environment – car structure, pavement material, traffic information displays, computerization of data. Undoubtedly one of the most dynamic field today, traffic telematics and the automation of the control/penalty chain lessen the need for police presence. Where police agents stand in this vast social, political and technical arena is a delicate matter. Police often appears as a liaison within a very varied group of modes of intervention, which is understood as typical of new forms of management and governing, raising issues of negotiation, acceptation and reject.

Parallel to these studies concerned with public action, others develop a more critical reflection on the means themselves, and their evolution.

A number of them consider the ever-progressing behavior regulation technologies as what Michel Foucault called “biopower”: they reveal the insertion of always more subtle power relations within bodies and behaviors. This police activity and its extensions are described as the milieu *par excellence* where these standardizations become accepted, as they are performed on a daily basis and justified by the goal of reducing the loss of human human lives. Other researchers show that an organizational study of police systems raises the question of the efficiency of such programs. Technological and administrative tools are in fact incorporated to the “music” of the organization (Manning 2008) and act according to its criteria and definitions of issues, which are sociological constructs. Following a reasoning at times more philosophical, several authors question the relation between man and machine as it is in place in this form of policing, and the dehumanization of the road simultaneously in process. Cars and roadways are seen as the space for a specific social experience, an experimental field for future evolutions, which in this respect must be studied. The major question is that of the acceptability of these technical and automatized artifacts. While it is an effect of public policies, it is also dependent on habits developed with broader social mutations. Rather than offering an opportunity for the denunciation of the control of the body, the disappearance of the police body and physical interactions, replaced by remote, computerized surveillance, would then invite the questioning of the creation of new forms of social relations, navigating between internalization, adjustments, and resistance.

What Efficiency? Reducing Accidents, Normative Incorporation, Social and Cultural Variations

The last question can be worded as follows: what is the efficiency of this multiform action? Answers are contradictory, and are open to a debate that is not defined as of yet. Criminologists and sociologists have long shown that police action is generally of little influence on the evolution of criminality. Yet, traffic police may be the exception to the rule: it is one of the only field where this impact is established and measurable. It can be seen in the evolution of the figures: the number of death per 100,000 inhabitants within the OECD went from 23.4 in 1970 to 11.4 in 1999, while in some countries such as Israel, the rate of deadly accidents has been divided by ten in the same period. This trend has been confirmed in recent years, as can be seen in the United States, Hungary or Ireland, where the number of death has followed a stronger declined between 2008 and 2011 than it had during the 15 preceding years. Overall, when a variety of road safety measures are massively implemented in coordination, a rise in the number of drivers can be observed, as well as a decline in the number of accidents. This observation, however, is disputed: some researchers think that the decline is not due to police action or law enforcement, but rather to technical progress in vehicles and road environments, as well as the social acceptance of the norms, the real condition to the efficiency of these measures. Other studies focus more on some countries, some but not all of them developing countries, or on a selection of regions within countries, where the trend cannot be observed.

In any case, traffic police action is often considered a field of choice for observation, because of the intense impact of normative pressure, of the repetition of contacts, of the will to make mobility easier, and of the internalization of driving laws and norms. The analysis of the action of traffic police agents (most often at an intersection) is a staple of sociological studies: on the changes in behavior when drivers are faced with an officer (E. Goffman), on the internalization of social order into non-conscious habitus, through the “natural” respect for driving rules (P. Bourdieu), or on the revealed cognitive styles, through a relation to the rule that would be specific to differentiated societies (H. Garfinkel).

This shared observation of an interaction that is continuous or temporary, physical or technical, and that isn’t without its effect son behaviors, is challenged by other forms of studies. Some of them focus on the persistence of inequalities, based on rank, race or gender, and are visible even in traffic policing. Officers in some US states may, consciously or not, tend to be more zealous with younger drivers, as well as females, and Black and Hispanic drivers. The study of "Driving while Black" has led to a number of scholarly works and to a debate on how to observe and analyze the phenomenon. Other researchers, on the contrary, highlight the varying cultural interpretations of these forms of social control. In Indonesia, the law on the compulsory helmet for motorbike users, for example, seems to have little effect on the seriousness of injuries. The frequency and gravity of accidents is not linked to the use of alcohol, as is the case in numerous other countries (religious rules are strictly observed and alcohol plays a very minor part in traffic accidents); rather, accidents are in part due to the very formal interpretation of the law, on the part of drivers and agents. They wear a helmet, and control that a helmet is worn, but pay no attention to how it may prevent injury, as no protection in case of accident is attached to the wearing of helmets. So helmets are worn, but without efficiency. In another area, on the Polynesian island of Niue, 20 % of non-intentional deaths are caused by road accidents, amounting to a cost of 1–2 % of the GDP. An anthropological study shows that this rate is tied to the nature of vehicles, the nature of the road environment, and forms traffic policing, but also to cultural factors such as low sensitivity to risks and a culture of road warriors among the youth, reflecting older Polynesian representations. Social, institutional and cultural imperatives thus constantly intervene in the possible imposition of a norm by the authorities, posing them dreadful problems. Traffic police may at times be in opposition to these imperatives, and sometimes heavily influenced by them: traffic police seems to be at the crossroad of diverging normative domains, making its action and the usual interpretation of its efficiency less mechanical. Future research could study this further.

Traffic police appears as an important, fragmented and stimulating research subject. The state of the research on the topic is tied to the diversity of the forms and the names it takes, and to the relative invisibility of its action, integrated to daily life or gradually disappearing as new forms of control take over. However, it may also be due to the fact that the various academic fields involved are separated by debates over their epistemological foundations, their methods and their conclusions (transport engineering, cognitive psychology, macro-economy, urban history, sociology of organizations. . .) Above all, studies are separated by their intended destinations: there is a rift between research targeting public action and aiming to improve policing, and more reflexive or critical studies aiming to define the stakes and issues of traffic police and policing. The bridging of that rift, while often claimed to be accomplished, is still fragile (Monjardet 1996). Yet, the study of this somewhat neglected branch of policing is of great importance, when mobility in all its forms takes a major place in the experience of daily life as well as in social organization. A lot can be gained through well regulated interdisciplinary research focusing more directly on the aspects mentioned here. In this human and organizational work on mobilities and individuals, are at stake the major challenges to the modes of functioning of law keeping and enforcing institutions, and more broadly the changes in the relation to the law, to the norm, to the freedom move, and to safety within contemporary societies.

Related Entries

- ▶ Cognitive Forensics: Human Cognition, Contextual Information, and Bias
- ▶ Comparative Legal Cultures
- ▶ Crime Prevention Through Environmental Design
- ▶ Eyewitness Evidence: Historical Background
- ▶ Labelling, Deviance, and Media
- ▶ Organizational Change and Police Legitimacy
- ▶ Police Discretion in Law Enforcement
- ▶ Police Legitimacy and Police Encounters
- ▶ Social Control
- ▶ Surveillance Technology and Policing
- ▶ Traffic Enforcement and Control

Recommended Reading and References

Accident analysis and prevention

Berlière J-M (1996) Le monde des polices en France. Complexe, Bruxelles

Campos PHF, Lagares R (2002) The traffic and driving social representations. In: Acts of the 5th international conference on social representations, Montreal, pp 454–472

Carnis L, Hamelin F, Spenlehauer V (2004) Les polices de la route: une approche comparée (Etats-Unis, Australie et Nouvelle-Zélande), etude commandité à l'Institut National de Recherche sur les Transports et leur Sécurité (INRETS) par l'Institut des Hautes Etudes de Sécurité Intérieure (IHESI), Convention de recherche n°02/1023, rapport final

Castro C (ed) (2009) Human factors of visual and cognitive performance in driving. CRC Press, Boca Raton

Conley J, Tigarmac Laren A (2008) Cars troubles, critical studies of automobility and auto mobility. Ashgate, Farnham

Delaney A, Ward H, Cameron M (2005) The history and development of speed camera use. Monash University, Accident Research Centre, report no. 242, September

Deluermoz Q (2010) The prefecture de police's car service and the Parisian traffic: politics, reglementations and policing of a public problem (1892–1921). In: 8th European social science history conference, Ghent, Belgium

Emsley C (1993) Mother, what did policemen do when there weren't any motors? The Law, the police and the regulation of motor traffic in England, 1900–1939. Hist J 36:357–381

Engel R, Calnon J (2004) Examining the influence of drivers' characteristics during traffic stops with police: results from a national survey. Justice Q 21:49–89

Finney M (1997) Scale economies and police department consolidation: evidence from Los Angeles. Contemp Econ Pol xv:121–127

Gusfield J (1981) The culture of public problems: drinking-driving and the symbolic order. The University of Chicago Press, Chicago

Hamelin F, Spenlehauer V (2005) Traffic police and policing: a comparative and socio-historical perspective. Paper proposed for 2005 Australian road safety research policing education conference

IRTAD (2011) Road safety annual report. http://internationaltransportforum.org/irtadpublic/pdf/11IrtadReport.pdf

Kratcoski P, Das D (2002) Traffic policing: an international perspective. Policing 25(3):619–630

Lawton R, Parker D, Stradling SG (1997) Predicting road traffic accidents: the role of social deviance and violations. Br J Psychol 88:249–262

Lundman R, Kaufman R (2003) Driving while black: effects of race, ethnicity, and gender on citizen self-reports of traffic stops and police actions. Criminology 41(February):195–220

Manning P (2008) The technology of policing: crime mapping, information technology, and the rationality of crime control. New York University Press, New York

McShane C (1999) The origins and globalization of traffic control signals. J Urban Hist 25:379–404

Monjardet D (1996) Ce que fait la police. Sociologie de la force publique. La Découverte, Paris

Monkonnen E (ed) (1992) Policing and crime control, part 3. Saur, Munich/London/New York/Paris

Roche D (2008) La culture equestre de l'occident, XVIe – XIXe siècle. L'ombre du cheval. Fayard, Paris

Rothe JP (1994) Beyond traffic safety. Transaction Publishers, New Brunswick

Schmucki B (2001) Der traum vom verkehrsfluß. Städtische verkehrsplanung seit 1945 im Deutsch-Deutschen vergleich. Campus Verlag, Frankfurt

Xie C, Parker D (2002) A social psychological approach to driving violations in two Chinese cities. Transp Res F Psychol Behav 5:4

Trafficking in Human Beings

Monika Smit[1] and Peter van der Laan[2]

[1]Administration of Justice, Legislations, and International and Aliens Affairs (RWI), Research and Documentation Centre (WODC) of the Dutch Ministry of Security and Justice

[2]Netherlands Institute for the Study of Crime and Law Enforcement (NSCR) and Faculty of Law, VU University Amsterdam, Amsterdam, The Netherlands

Synonyms

Human trafficking; THB; TIP; Trafficking in persons

Overview

Trafficking in human beings (THB) concerns the exploitation of vulnerable individuals for the purpose of financial profit. Victims of trafficking, women but also men and children, fall prey to different forms of exploitation, such as forced labor or services, slavery, servitude, or the removal of organs. However, the most frequently reported type of trafficking is the one that results in the exploitation of women in the sex industry.

This contribution focuses particularly on this type of trafficking.

The concept of THB came into use in the early twentieth century in connection with white slavery, a term then referring to forced or fraudulent recruitment of women and girls for prostitution. Currently THB refers to a broader phenomenon that can be approached from different perspectives, for example, as a public order problem, a migration issue, a criminal justice issue, a violation of human rights, or a combination thereof.

Usually the trafficking process is distinguished in three phases: the recruitment of victims, their transportation, and the eventual exploitation. These phases are reflected in the most widely shared definition of THB, the definition in article 3 of the United Nations Protocol to prevent, suppress, and punish trafficking in persons, especially women and children:

> Trafficking in persons shall mean the recruitment, transportation, transfer, harboring or receipt of persons, by means of the threat or use of force or other forms of coercion, of abduction, of fraud, of deception, of the abuse of power or of a position of vulnerability or of the giving or receiving of payments or benefits to achieve the consent of a person having control over another person, for the purpose of exploitation. Exploitation shall include, at a minimum, the exploitation of the prostitution of others or other forms of sexual exploitation, forced labor or services, slavery or practices similar to slavery, servitude or the removal of organs.

This definition, although widely shared, has also been heavily criticized. It answered some of the difficult questions that were raised. It became, for example, clear that not only women and girls but also boys and even men can be trafficked and that THB is not restricted to the purpose of forced prostitution. However, because the definition in the protocol describes a process involving several actions and outcomes which themselves are not clearly defined, there is still dispute about what falls under its umbrella.

Furthermore, in practice, it is difficult to distinguish THB from the smuggling of persons. In fact, assisted illegal immigration may precede exploitation, as some smuggled illegal immigrants end up as victims of trafficking later on. Lured with bogus offers of employment, these migrants fate only becomes clear when they are gradually forced to do other work than they had agreed upon or under very different circumstances than expected. The most obvious difference between the two phenomena is that human *smuggling* primarily relates to illegal immigration and the violation of immigration laws, whereas human *trafficking*, which often but not necessarily involves border crossing, implies exploitation, and the violation of an individuals' human rights.

Until the endorsement of the UN Trafficking Protocol, the response to THB was exclusively human rights based. Human rights law nowadays remains important to ensure the protection of the rights of victims of trafficking, but a different perspective was necessary in order to realize the current level of global attention and resources to combat contemporary exploitation (Gallagher 2010). Since the shift in perspective and the protocol, several regional treaties and treaty-like instruments on trafficking have been developed, dealing with issues like criminalization, international cooperation, and the protection of victims. In addition, 80 countries have implemented a comprehensive anti-trafficking act, 64 countries have provisions in their penal codes criminalizing all forms of trafficking, 22 have provisions criminalizing some forms of THB, and only 2 countries have no such laws. Furthermore, 58 countries have a National Plan of Action (NPA) to deal with THB, and 12 have NPAs that target human right violations related to THB (Protection Project 2012). Attention to the phenomenon of THB has grown significantly, and numerous activities to combat THB and deal with its victims have been implemented but little is known about which interventions are the most effective in preventing human trafficking, protecting victims, and prosecuting traffickers.

This contribution presents information about traffickers and their modus operandi, a description of some of the many policies and interventions dealing with THB. Furthermore, limitations of the amply available research on this topic are discussed, followed by the findings of a systematic review that underline these limitations, and a conclusion.

Traffickers and Their Modus Operandi

Individual Characteristics of Perpetrators

Information about the background of individual traffickers is mainly based on national statistics concerning those who have been prosecuted and convicted. Most are males in their mid-20s or mid-30s, but women, who are traditionally regarded as victims, are also involved in recruitment, organizing the trafficking process, and controlling victims. The percentage of women involved differs per nationality of convicted perpetrators. Some perpetrators recruit or exploit fellow countrymen. Already present migrant networks can help newcomers with their start in the new country, by assisting in finding shelter or work, but newcomers also are an easy prey for malicious fellow countrymen.

Trafficking and Organized Crime

Trafficking is often described as a crime perpetrated by ruthless criminal networks, and the above-mentioned "trafficking protocol" supplements the UN Convention against Transnational Organized Crime (2000). Article 2a of this Convention defines an organized crime group as *a structured group of three or more persons existing for a period of time and acting in concert with the aim of committing one or more serious crimes or offences established in accordance with this Covention, in order to obtain, directly or indirectly, a financial or other material benefit*. A "structured group" is *a group that is not randomly formed for the immediate commission of an offense and that does not need to have formally defined roles for its members, continuity of its membership or a developed structure* (article 2c).

According to the US trafficking in persons (TIP) report 2012, organized crime and complex networks indeed contribute significantly to the trafficking of persons. The TIP report is a yearly comprehensive report by the US Department of State on efforts made by governments worldwide to combat THB.

Human trafficking networks are likely to be organized in small groups, which operate both independently and in cooperation with other crime groups. The interaction between groups is often connected to the provision of specific services such as the recruitment of victims in particular countries or areas or their transport to locations where (forced) prostitution takes place. Among other roles individuals take on are those of investors, debt collectors, and money launderers.

However, not only organized crime groups commit THB. It is known, for example, that within the European Union, the removal and relaxation of internal border controls provided opportunities for individual criminals and less sophisticated, smaller, or mid-level groups to operate across borders, and there are also individual perpetrators enriching themselves by exploiting other humans. The actual share of organized crime in THB worldwide is unknown.

In her study on THB in the Netherlands, Van Dijk (2002) distinguished three main forms of criminal cooperation: solo-offenders, self-supporting criminal groups, and criminal macro-networks. Solo-offenders force one or more individuals into prostitution. Self-supporting criminal groups control all phases in the trafficking process, from recruitment to prostitution. They do not have established contacts with other offenders or criminal groups involved in human trafficking. The main suspects are often in charge of brothels or sex clubs, relying on personal contacts to recruit and transport victims. Criminal macro-networks include solo-offenders as well as criminal groups, clustered by geographical proximity, family ties, friendships, commercial circuits, or similarity in criminal activities. The characteristics of organized trafficking groups are related to the perpetrators' nationality/ethnicity and the region in which they operate.

Shelley characterized organized trafficking groups in five different ideal typical "business models," linked to specific geographical areas and modus operandi. These are (1) the post-Soviet *national resource model* (women recruited in source countries are trafficked for prostitution), (2) the Chinese and Chinese-Tai *trade and development model* (smuggling and trafficking men for labor exploitation and to a lesser extent women for prostitution), (3) the US-Mexican trade *supermarket model* (smuggling large numbers of men and women, the women trafficked for forced prostitution), (4) the Balkanese violent *entrepreneur model* (women trafficked for prostitution by middlemen are controlled from bases in the Balkan), and (5) the *traditional slavery with modern technology model* (trafficking young women out of Nigeria and West Africa for prostitution, forcing compliance by using contracts and voodoo practices) (Aronowitz 2009).

Even THB cases in which organized crime and foreign criminal groups are involved often have a local component, sometimes referred to as "glocalization." Criminals have to trust each other in order to be able to do business. Friends, relatives, or others with whom they have had good experiences are eligible as "partners in crime."

The Recruitment of Victims

Traffickers recruit their victims mostly in deprived, disadvantaged, or poorly integrated sectors of society, offering them employment abroad. Unemployment and low earnings are among the strongest push factors for victims. The Internet provides anonymous opportunities to recruit victims. Most migrate willingly, and many are aware of the fact that they will work in prostitution. Other victims have been promised jobs in the entertainment industry or as, for example, nannies or domestic servants. Outright abduction or selling into prostitution is rather the exception than the rule (Andrees 2008). This can also be concluded from data in the International Organization for Migration (IOM) Counter Trafficking Database. In 9.646 cases (who received assistance from 1999 to 2006), 60 % had been recruited via personal contacts (in 25 % of these cases by friends, relatives, or partners), and only nine victims had been kidnapped.

An often used recruitment strategy is the so-called loverboy method: after manipulating victims by faking love, they are gradually coerced or threatened into prostitution. Not all recruitments of foreign victims take place in countries of origin. Some victims are recruited in the country of destination, after having arrived there as undocumented migrants. They are vulnerable to exploitation, as they risk alien detention and expulsion.

Means of Control

Once the exploitation has started, and victims are working in street prostitution, behind windows, in clubs, private homes, massage parlors, or as escorts, they have to be persuaded to keep "delivering." The principle means of control are debt bondage, dislocation, and mobility (rotation of victims between different locations, sometimes across borders), as well as close monitoring, intimidation, and violence. Recently, however, perpetrators have started to use more subtle methods of control and share a small part of the profits with the victims, thus discouraging them from complaining to the authorities and making it more difficult to prove coercion. In cross-border trafficking, fuelling the victims' fear of local authorities and the prospect of expulsion is a common control method.

Some Policies and Interventions in the Fight Against THB

Many different policies and interventions have been implemented to deal with THB. These partly depend on the perspectives with which THB is approached. Most are not specifically or not exclusively aimed at trafficking by organized crime groups.

Policies Dealing with Prostitution and the Sex Industry

Policies against human trafficking for sexual exploitation are often directly linked to policies dealing with prostitution and the sex industry. Among these are interventions to address the demand for commercial sexual services, including a variety of campaigns and educational programs to change men's attitude and behavior towards buying sexual services, and so-called John Schools to educate buyers in order to prevent them from soliciting again. Shaming tactics, such as making the names of the males involved public, are also mentioned. According to this report, most men believe that criminal penalties would have a deterrent effect only if they believed they were at risk of being caught. However, in most countries this risk is rather minimal.

In many countries, prostitution is tolerated to some extent, though different forms of prostitution and exploitation are regulated or combated in different ways. In Sweden, however, prostitution is regarded as a form of violence, committed by men against women. In 1999, the purchase of sexual services became punishable by law, while the sale of such services was decriminalized. Since then street prostitution has been halved, while at the same time, it increased in Denmark and Norway. Internet prostitution increased in all three Scandinavian countries but on a larger scale in the neighboring countries, while the proportion of Swedish men reporting that they have purchased sexual services has decreased. However, some of these conclusions are based on estimates, and the effects of the law on *trafficking*, which is not the same as prostitution, remain unclear.

The Dutch authorities took a different approach. In 2000, after a policy of tolerance for many years, the general ban on brothels was lifted. Under certain conditions, making money out of voluntary prostitution by adult prostitutes was no longer prohibited, and brothels were legal if they complied with certain licensing conditions. The intention was to simultaneously act forcefully on prostitution of minors and illegal aliens, as well as involuntary prostitution. An extensive evaluation showed that this approach of location-bound prostitution produced some positive effects: business owners tend to comply with licensing conditions and the number of prostitutes without legal documents decreased. However, the position of prostitutes did not improve much, and pimps are still a common phenomenon. Furthermore, law enforcement investigations show that human trafficking still thrived behind the legal façade of regularization. The main reason is that the policy focused on business owners, whereas the exploitation of women is more often carried out by pimps.

In Finland, yet another position was taken: the law prohibits the buying of sex from a trafficking victim or from a subject of procurement. But will a client be able to know who is a victim, or who is being procured? And if so, will he act upon it? Some clients are willing to report on involuntary prostitution but most men who buy sex are aware of trafficking and exploitation, but this does not affect their behavior. A quarter of 103 London men who buy sex in the UK report having encountered a woman in the sex industry whom they believe was forced, but only five reported this to the police (Farley et al. 2009).

Policies Against Illegal Migration

Another way to combat human trafficking is to treat it as a form of illegal migration and increase border controls in destination and transit countries. Examples are exercising due diligence in regulating entry of foreigners, in order to identify individuals at risk, checking transit documents, and providing protection to suspected victims of trafficking in transit. However, this is easier said than done as much THB for sexual exploitation takes place over short distances within countries, and if borders are crossed, this frequently involves future victims who bear legal travel documents.

Cooperation in the Investigation and Prosecution of Traffickers

Next to national policies regarding the attitude towards the sex industry and border protection, there are policies directed at the investigation of trafficking cases and the conviction of perpetrators. Cooperation is an important characteristic of this type of policies. It concerns cooperation between different actors within regions and municipalities as well as cross borders. In the Netherlands, for example, several investigative bodies cooperate with local authorities, victim assistance, and health care in a combined criminal and administrative approach, which also contains preventive measures. In order to gather enough evidence in cases in which victims do not cooperate in the investigation and in which it is difficult to conclude a case because it is one's word (the victim's) against another's (the suspect's), the Dutch police developed a proactive investigation technique. The method implies the piling of all established facts, from testimonies and telephone tapping to observed injuries in order to establish a body of evidence without using victim statements.

The fight against cross-border trafficking cases that involve foreign suspects requires cooperation with criminal investigators in other countries. Such cooperation takes place ad hoc as well as in the form of well-planned actions, in Europe called Joint Investigation Teams (JITs). The work in such teams is not always easy because of differences among countries in laws, jurisdiction, and priorities. Eurojust can help in solving jurisdiction problems that occur in the cooperation in the fight against cross-border criminality, among which trafficking cases.

Awareness Raising Campaigns, Training Material, Manuals, and Toolkits

Finally, the many campaigns and materials that have been developed for the various target groups who may fall prey to traffickers as well as for those who may come across victims of trafficking are worth mentioning. Among these are awareness raising campaigns to warn potential victims and manuals for professionals. Awareness raising campaigns include radio and television advertisements, educational school programs, and posters at airports informing incoming passengers of hotline numbers they could call should they become victims of THB. To strengthen the law enforcement response to THB, training materials or modules have been developed. The UN, for example, developed an “anti-human-trafficking manual for criminal justice practitioners,” and organizations like the International Labour Organization (ILO) and the International Trade Unions Confederation have developed training material about the way to handle indications of trafficking for Labor Inspectorates, unions, and employees (see, e.g., the manuals on the ILO site www.ilo.org and the UNODC online “Toolkit to combat trafficking in persons”). Whenever a victim has been identified, protection and assistance have to be offered. Several of the above-mentioned manuals and toolkits offer handles in this matter.

Limitations in Research on THB

Scarcity of Data

Because of the definition problems mentioned in the introduction, the criminal nature of THB, and the fact that victims rarely come forward with what has happened to them, accurate data on THB are scarce. Most often victims are discovered during police investigations of some sort of criminal activity, for example, during raids on locations where prostitution takes place.

Victims refrain from reporting to the police, some even after having been discovered, because they do not consider themselves victims but migrants whose journey went wrong, because they are embarrassed, they fear expulsion, or because they either fear their perpetrator or are emotionally attached to him. Notwithstanding the lack of accurate data, many figures are presented to map the scale of the phenomenon. Goodey (2008) describes how “guesstimates” are used for that purpose. Between the estimates, for which the computing methodology is rarely provided, and the number of identified victims exist huge differences.

Much has been written about initiatives to deal with trafficking for the purpose of sexual exploitation, but knowledge on “what works” is particularly limited. Although thorough evaluation is important, as evaluative knowledge on trafficking can be used to develop adequate prevention techniques or policies, this is scarce. The United States Government Accountability Office (GAO), for example, reviewed documents of 23 US government-funded anti-trafficking projects in Indonesia, Thailand, and Mexico. It revealed that 21 of the 23 projects included one or more monitoring elements, but only ten stated how performance was measured. The majority lacked a logic framework of monitoring that linked activities to goals, indicators, and targets (GAO 2007).

Limited Scope of and Methodological Fallacies in Trafficking Research

Most publications on THB are primarily descriptive, for example, describing the hardship of or social assistance to victims. Others mainly aim at collecting numbers to define the scale of the phenomenon. Few, if any, interventions are accompanied by evaluation research to assess their impact and cost-effectiveness and not much systematic data collections exist. There has been relatively little *independent* research – research carried out by others than the ones who developed or delivered the counter trafficking initiative – to evaluate and assess the effectiveness of counter trafficking policies, programs, and interventions. This can be problematic since organizations evaluating their own initiative may have explicit or implicit political or practical agendas that influence their conclusions, for instance, by selecting or disposing of certain information.

Another frequently encountered restriction in the research on THB is the lack of large samples. This is often due to the limited access to victims and perpetrators. Using a variety of methods on small samples can produce illuminating results but often preference is given to a single method such as either interviews or questionnaires. Furthermore, much of what is known comes from just one source, such as governmental statistics, data that reflect the legal framework and enforcement strategies in a country rather than the actual scope of the problem, or victims who have been or are being assisted. As a result, (reliable) knowledge on effectiveness of interventions in this area is scarce. In the meanwhile, anti-THB interventions can cause collateral damage, for example, by putting restrictions on women's migration or travel.

The Need to Know More

Because of the severity of the crime, the impact on its victims – some are severely traumatized by what they went through – and the possible collateral damage of anti-THB measures, it is important to gain more insight into the working and effectiveness of anti-trafficking strategies and interventions. This was the reason to undertake a review to assess the presently available evidence on the effects of interventions that aim to prevent and suppress this type of crime (Van der Laan et al. 2010).

A Systematic Review

Objectives of the Review, Research Questions, and Method

The review aimed at contributing to a more evidence-based approach in the prevention and suppression of cross-border THB for the purpose of sexual exploitation. The questions to be answered were as follows: Which anti-THB strategies were evaluated in a way that is rigorous enough to determine their effect, and what are the outcomes of these (quasi-)experimental studies? Only studies focusing on cross-border trafficking for the purpose of sexual exploitation were reviewed. The focus was on cross-border trafficking because internal trafficking was not (yet) or just recently recognized in many countries. Furthermore, the review was limited to trafficking for prostitution or sexual exploitation, in any form there is. Trafficking for forced labor or services, slavery, servitude, or the removal of organs was kept out of consideration.

All eligible evaluation studies in the abovementioned field were categorized according to the Maryland Scientific Methods Scale (SMS), in order to assess the strength of the evidence (Sherman et al. 2002). The studies dealt with initiatives targeting individual perpetrators and criminal groups as well as vulnerable individuals and professionals in the field of anti-THB. Studies exclusively regarding victim assistance, the reintegration of victims, or arrangements for their legal status were excluded. Eligible studies in nine European languages, either published or unpublished, from the year 2000 up to the end of June 2009 were reviewed. For the broad search strategy, including the databases that have been searched and the key words used, see www.campbellcollaboration.org.

Main Results

The search resulted in 144 potentially eligible studies. These concerned interventions of a primarily preventive nature, suppressive interventions (mostly regarding legislation and prosecution) or were of a combined preventive/suppressive nature. In 48 studies no specific initiative was described. Most studies were in the field of sexual exploitation or concerned more than one type of exploitation. The target group varied and consisted of both adults and children, focusing on women and children, but studies on programs that were especially developed for minors were found as well. After a more extensive examination, not one study was found to be applicable to be included in the review since they could not be qualified as (quasi)-experimental study.

Four studies qualified in most respects (they concerned interventions developed for prevention or suppression of cross-border THB for the purpose of sexual exploitation), except for using pre- and post-measures and a control group (corresponding with level 1 or 2 on the SMS). In the following these studies will be described briefly in order to give an idea of the kind of interventions being evaluated and the limitations in the evaluation designs used.

A Closer Look at Four Studies

The first study (Association for Emancipation, Solidarity and Equality of Women 2006) concerned a qualitative evaluation of the economic and social stabilization program for potential victims of trafficking in the border regions of Former Yugoslavic Republic Macedonia. The main goal of this program was empowering these potential victims – vulnerable groups, especially women – by providing employment opportunities through skills training. Basic business training courses were organized aiming to develop knowledge on how to start and manage a small business and how to generate income. The study assessed the impact of the project on the beneficiaries, their families, and local stakeholders as far as socioeconomic position, quality of life, and vulnerability towards human trafficking are concerned. The study consisted of a posttest measure only (using questionnaires and interviews), there was no control or comparison group involved, and no empirical outcomes were reported.

The second study (Balanon and Barrameda 2007) evaluated the impact of the Young Men's Camp in the Philippines, an intervention addressing the demand for sexual services. Its goal was to change the sexual attitudes and practices of boys and young men. The study contained pre- and posttests to assess changes in their knowledge and perceptions. Next to these assessments, there were follow-up activities, consisting of 1-day fora and workshops. It remained unknown whether the intervention focused only on cross-border trafficking.

The third study (Centre for Research on Environment Health and Population Activities 2004) included an evaluation of an awareness raising program in rural Nepal. The primary target group of the program was adolescent girls, and their peers, parents, and the community were the secondary focus group. This intervention educated the girls and aimed to prevent them from becoming victims of trafficking. The effects of the program were measured using interviews and surveys. The study contained a pre- and posttest measure without a comparison or control group. Outcome measures were the increase of awareness and knowledge on trafficking.

The fourth study (Hashash 2007) evaluated an Israelian prevention program aiming at raising awareness among the wider public and creating attitudinal and institutional change among decision makers and law enforcement personnel. Education on trafficking was achieved by distributing information, providing training, publicizing reports, using media advocacy, and giving lectures. The program intended to influence public policy and opinion. The researchers conducted a pre- and posttest using interviews and questionnaires. No control or comparison group was used.

Even though these studies were better designed compared to most of the studies in this field, they do not allow for firm conclusions regarding the effectiveness of the interventions concerned.

Conclusion and Discussion

Trafficking in human beings is a crime conducted by individuals as well as by smaller and larger organized crime groups. They recruit vulnerable victims – women, men, and children – mostly in deprived regions or sectors of society, in order to exploit them in the sex industry or in other economic areas. The perpetrators use different and more or less violent means of control, ranging from close monitoring and threats to outright violence.

The attention to THB has risen during the last two decades.

Most countries now have laws in place making trafficking punishable and protecting its victims. Many have National Action Plans to combat THB, and numerous policies and interventions to deal with THB have been developed. Depending on the perspective from which THB is approached, these policies and interventions range from addressing the demand for commercial sexual services, border protection, or improving the investigation of trafficking cases and the prosecution of its perpetrators. In addition, there are the awareness raising campaigns and manuals, toolkits, and training materials that have been developed for potential victims and for parties who may come across them.

For various reasons there is not much reliable data about THB. Still, some progress has been made. An increasing number of countries have trafficking data available, also thanks to the pressure to provide national trafficking data, for example, by the US government for its yearly TIP report. In addition, there have been several initiatives to collect data across countries in a uniform way, and guidelines for the collection of data have been published.

Still lacking is evaluation research on anti-trafficking initiatives that is methodologically rigorous enough to be able to reach substantive conclusions about the effectiveness of the present anti-trafficking initiatives. Many go without pre-test measures and/or a control group. Improvements in research design are seriously needed. This does not have to be that difficult since individuals who do not participate in an intervention or who are on a waiting list can, for example, be treated as control group. Researchers, funding institutions, and policy makers should demand higher methodological standards of evaluation studies in order to gain knowledge on outcomes. This would enable them to better succeed in preventing and suppressing trafficking in human beings for the purpose of sexual exploitation. Partnerships can be sought in the area of politics and funding agencies.

Debates about the approach of THB, especially where it concerns THB for sexual exploitation, are currently often heated and clouded by generalizations and myths. Better research might also help to make these debates more evidence-based and more fruitful.

Related Entries

▶ Forced Migration and Human Rights
▶ Human Smuggling
▶ Prostitution

Recommended Reading and References

Andrees B (2008) Forced labour and trafficking in Europe: how people are trapped in, live through and come out. ILO working paper 57

Aronowitz A (2009) Human trafficking, human misery. The global trade in human beings. Prager, Westport/London

Association for Emancipation, Solidarity and Equality of Women (2006) Qualitative evaluation of the economic and social stabilization programme for potential victims of trafficking in the border regions of FYR Macedonia – ESS. CEB, IOM & the Norwegian government

Balanon FG, Barrameda TV (2007) Coalition against trafficking in women – Asia Pacific 2004 to 2006. Evaluation of programs. CATW, Philippines, www.fafo.no/pub/rapp/20040/20040.pdf

Centre for Research on Environment Health and Population Activities (2004) The anti-trafficking programme in rural Nepal: assessment of change in awareness and communication among adolescent girls, peers and parents in Baglug District. Centre for Development and Population Activities DC, U.S. Department of Health and Human Services, Gairidhare, Kathmandu

Farley M, Bindel J, Golding JM (2009) Men who buy sex. Who they buy and what they know. Eaves/Prostitution Research and Education, London/San Francisco

Gallagher AT (2010) The international law on human trafficking. Cambridge University Press, Cambridge

Goodey J (2008) Human trafficking: sketchy data and policy responses. Criminol Crim Justice 8:421–442

Hashash Y (2007) Evaluation report. Trafficking project. Hotline for migrant workers and Isha, L'Isha

Sherman LW, Farrington DP, Welsh BC, MacKenzie DL (eds) (2002) Evidence based crime prevention. Routledge, London

The Protection Project (2012) The protection project review of the trafficking in persons report. Johns Hopkins University, Washington, DC

United States Department of State (2012) Trafficking in persons report. Department of State, Washington, DC

United States Government Accountability Office (2007) Human trafficking: monitoring and evaluation of international projects are limited, but experts suggest improvements. Government Accountability Office, Washington, DC

Van der Laan PH, Smit M, Busschers I, Aarten P (2010) Cross-border trafficking in human beings: prevention and intervention strategies for reducing sexual exploitation. http://www.campbellcollaboration.org

Van Dijk E (2002) Mensenhandel in Nederland 1997–2000 (Trafficking in human beings in the Netherlands 1997–2000). KLPD, Zoetermeer

Trafficking in Persons

▶ Trafficking in Human Beings

Trajectories

▶ Longitudinal Crime Trends at Places
▶ Onset of Offending

Transactional Theory

▶ Interactional Theory of Delinquency

Transitional

▶ Transitional Justice

Transitional Justice

Carl S. Stauffer
Center for Justice and Peacebuilding, Eastern Mennonite University, Harrisonburg, VA, USA

Synonyms

Communal context; Deviance theory; Indigenous; Restorative justice; Social constructivism; Transitional

Overview

The exercise of defining criminal deviance entails a complex and mystifying process often fraught with tension and paradox. The locus questions of *who* defines deviation and *how* it is defined remain the key "flash-points" in this debate. The crux of the matter is enmeshed in the dissatisfaction of the current criminal justice system which has maintained a highly positivist (clearly defined, scientific) approach to the notion of deviance. This positivist inclination is deeply rooted in rational choice and/or criminogenic theories (psychopathic explanations) and has adopted certain tenants of social control theory which rotate around three major "pillars" of rationalization for the institutions of criminal justice: deterrence, rehabilitation, and incapacitation (Kaplan 1973).

With the rise of transitional justice applications in post-war contexts (where gross human rights violations, mass atrocities, and genocide have occurred), these tensions around defining deviance and the location of justice have become even more pronounced. An alternative agenda is required in order to reframe the on-going debate around criminal deviance and its implications for the administration of justice.

An alternative agenda is emerging in the transitional justice field through the current revival of restorative, indigenous justice practices that are "filling in the gaps" of what constitutes satisfying human justice. Simply put, numerous forms of ancient indigenous justice that are restorative in nature are providing new lenses for reconfiguring what justice means and how it is implemented (Cobban 2007; Pouligny et al. 2007). There are three critical observations regarding this renewal of indigenous justice systems:

1. The Transitional Justice field has unwittingly provided the scaffolding of language and the platform of dialogue for indigenous justice forms to surface in order to meet the desperate need for justice to be felt as transformative by local communities on-the-ground who otherwise are often left out of the national and international transitional justice discourse and accompanying mechanisms.

2. These ancient forms of indigenous justice are offering coherent frameworks that engage ethical values and processes of facilitation that are necessary in order to critique and create new future hybrid models of justice.
3. There are numerous paradoxes and negative externalities resulting from the attempts to concretely define deviance in the current Western legal system. Indigenous justice applications are showcasing the restorative corrective "DNA" to assist in redefining the meaning of human deviance away from a stagnant code of legislated behaviors toward a nested paradigm of human interaction that is negotiated through relationship ties, and the norms and mores of community networks.

Fundamentals

What Is Transitional Justice?

The survivors of war, mass atrocities, and genocide across the Globe are crying out for justice. The question is: what does justice require? How is justice satisfied? What does justice *feel* like? Many say justice meets violence with violence – a revenge justice characterized by a visceral form of "blood rites" or a need to honor the memory of the dead who have been immortalized in a premature death of injustice (Ignatieff 1998). In its "civilized" expression this punitive justice is meted out in the form of state-sanctioned (legal) revenge in response to individual or communal (illegal) violence. This form of "retributive justice" is the dominant model of the western legal system which is based on the foundational notion that punishment (usually in the form of incapacitation or prison) results in individual change (rehabilitation) and collective restraint (deterrence). Some have relinquished the attainment of justice to a universal power beyond themselves – a Divine command, often requiring a form of "sacred violence" (Girard 1972). This is prevalent today in the political application of ancient legal codes such as the Mosaic or Sharia Law believed to be direct revelations from God and therefore sacred forms of justice. Others see justice as a human construct – a Natural Law. This concept emphasizes the rational law of "cause and effect" and that every action results in a ripple effect of consequential responses. The familiar adage of "you reap what you sow" contains this idea well. Social contract theorists such as Hobbs (1651), Locke (1689), and Rousseau (1762) are some of the best known proponents of this form of justice. Still others understand justice as social equity (freedom of choice), rights-based egalitarianism (fairness), and the search for the maximization of the "common good" – a kind of utilitarian or distributive justice (Rawls 1971; Sen 2009). Elements of all these conceptions of justice permeate the guiding principles of what is now termed *Transitional Justice* – a burgeoning field of systematic justice that aims to reconstruct the human and material capital of societies that have experienced mass atrocities (Hamber 2009).

Transitional justice involves a range of approaches that nation-states employ to address past human rights violations including trials and prosecution, truth commissions, lustration or vetting, reparations, reintegration of ex-combatants and war-affected populations, memorialization, and various institutional measures of reform and accountability (good governance, security sector, human rights, gender equity, and anticorruption structures). The said overarching aim of transitional justice is to end the culture of impunity and establish the rule of law in a context of democratic governance. To this end, the transitional justice field has identified eight broad objectives:

- Truth-telling at both the macro and micro levels
- Giving voice to victims – public platforms that "bears witness" to atrocities
- Ensuring accountability – perpetrator acknowledgement and responsibility
- Providing victim and survivor restitution and reparations
- Promulgating individual and collective healing and reconciliation
- Bolstering good governance structures that are sustainable
- Advocating for institutional reforms for the prevention of future violence
- Reinforcing public participation and dialogical democracy

What Is Restorative Justice?

In its best form, Transitional Justice should be *restorative* in nature. The field of Restorative Justice as an academic discipline and as a global practice movement is relatively young (early 1970s) with its contemporary origins rooted in experiments in alternative criminal diversion in Canada (Peachey 1989) and Indiana (Zehr 1990). However, in many ways it is simply a re-enactment of past practices of justice not too far distant in many ancient cultural traditions, but eclipsed by the current Western justice system (Braithwaite 2002). Restorative Justice is concerned with right relationships and the revitalization of community and collective harmony after a breach of violence. It is a justice that demands *accountability* (recognizing the harm and taking responsibility), making clear that no act that destroys human dignity goes unnoticed. However, along with this accountability it creates an avenue for *reconstruction* (equalizing power and addressing future intentions through restitution, reparations, and reconciliation).

Instead of a justice system that is obsessed with apportioning blame and shame, and administering punishment and isolation, restorative justice seeks to heal the harms of victims, rehabilitate offenders, and reintegrate both of them into community networks of support as a safeguard of justice. It is simple, but never easy. A comprehensive definition of restorative justice is:

> a broad term which encompasses a growing social movement to institutionalize peaceful approaches to harm, problem-solving and violations of legal and human rights. Rather than privileging the law, professionals and the state, restorative resolutions engage those who are harmed, wrongdoers and their affected communities in search of solutions that promote repair, reconciliation and the rebuilding of relationships. Restorative justice seeks to build partnerships to reestablish mutual responsibility for constructive responses to wrongdoing within communities. Restorative approaches seek a balanced approach to the needs of victim, wrongdoer and community through processes that preserve the safety and dignity of all. (Center for Restorative Justice, Suffolk University 2011)

Definitions and Theories of Deviance

In general, deviance is understood to be those actions or behaviors that stand in defiance to legislated legal boundaries (e.g., crimes) and/or culturally enforced social norms and mores (expected societal traditions and customs). Leading sociological thinking has extended the meaning of deviance to include any "*thought*" *or* "*feeling*" not just actions or behaviors (Douglas and Waksler 1982); deviance as contextually based in "*a particular society at a particular time*" (Dinitz et al. 1975); and deviance as having to do with "*a disapproved direction and of sufficient degree to exceed the tolerance limit of the community*" (Clinard and Meier 2008).

By in large, the *rationalist-utilitarian* conception of deviance remains dominant in the current Western legal system. This school of thought promotes a clearly defined set of "objective" legal codes that philosophically and practically delineate between deviant and nondeviant behaviors. Although to a lesser degree than the former, the current legal system has also given ample attention to the *structural-functionalist* approach to deviation which understands deviance as a necessary function in society that need only be managed effectively in order to maintain social equilibrium. The concept of "necessary function" entails the idea that through the existence of social deviance social service institutions and professional employment is created, social identity is formed, and a comparative standard of what is acceptable and unacceptable behavior is shared and reinforced for the "common good" (socialization). *Conflict theorists*, on the other hand, see the clear definition of deviance as an instrument of modern day institutions of domination (including the prison industrial complex) which are primarily concerned with maintaining the status quo in order to solidify power and harness the means for repressive social control – especially as it relates to class struggle, labor rights, and marginalized groups in society (Foucault 1977). Finally, there is a rising postmodernist critique (in the vein of *micro- or symbolic-interactionsim*) of the current Western legal system which would question the essence of the idea of deviance to the point of obscurity claiming that the very conceptualization and language of deviance as an illegal behavior is irrelevant and incomprehensible without a particular context to interpret them in. Thus, the two critical intersecting factors of analysis that lie at the heart of the debate around deviance theory are: Issues of *agency* (does the cause and responsibility of deviant behavior lie at the individual or collective level?) and issues of structural *purpose* (does deviance serve a functional or dysfunctional purpose in society?).

One of the best examples of how theories of deviance impact actual criminal offenders is found in what is termed *labeling theory* (Becker 1963). While there are many forms of labeling theory, when it comes to the present criminal justice system one could argue that the current penal codes serve as an official, state-sanctioned labeling process – often imposed for a lifetime (criminal offender stigmatization), without consideration of the particularities of context, void of the historical and relational complexities surrounding a crime, and bound to a mind-set and a timeframe of the present. This focus on the "present" labeling of a criminal offender does not allow for a process of reinterpretation (renarrating) of the past or for the opportunity to address future intentions such as exploring the possibilities of personal and relational transformation and/or creating new opportunities for healing and reconstruction. Winslade and Monk (2000) in their work on narrative mediation refer to this process of dealing with the past and the future as "deconstructing a conflict saturated story (labeling) and reconstructing a new story." In summary, when the criminal justice system attaches a label to certain codified "deviant" activity it does so in a contextual vacuum, it opens the door for the labeled person to became a societal "scapegoat" and it denies the offender the opportunity to take responsibility, show remorse, and offer some form of redress for the wrong they have committed. In short, transformative change is denied.

Comparative Study of Three African Transitional Justice Processes

Effective transitional justice processes call forth intersectoral and multidimensional interventions; approaches that can coordinate communication and collaborative action across the vertical and horizontal planes of human organization. In this section, a brief comparative study of post-violence justice efforts in South Africa, Rwanda, and Sierra Leone will be highlighted with the intersections of the formal legal system, transitional, restorative, and indigenous justice applications, and their impact on the understandings of deviance in mind. In its broadest sense, transitional justice as a process has the overarching mandate to (re)define the meaning of, and to deal effectively with, deviant violence enacted at a mass systems-level of society.

The Truth and Reconciliation Commission (TRC) in South Africa

The South African Truth and Reconciliation Commission (TRC) represents an example of a state-sanctioned (official) structure set up to legally define and come to terms with deviance as it was expressed through the Apartheid system and the ensuing violent struggle for liberation. For this post-war reconstruction context, the criminal courts were the only other form of institutional justice that South Africans had at their disposal. While the TRC moved away from blanket amnesty (the primary experience of Latin/South America) and embraced *conditional* impunity, it fell short on a number of other critical measures. First, it was perceived to be *perpetrator-biased* as opposed to victim-centered in part because of the amount of time and attention given to the legal aspects (rights and procedural undertakings) of the amnesty applicants. For example, victim-offender interactions were seriously hampered by the legal constraint that restricted perpetrators from discussing the past outside the parameters of the amnesty hearings themselves. The perpetrator was liable in a court of law for any confessions or admissions of guilt expressed outside the amnesty hearings. Hence, in this case, the law actually stood as a barrier to accomplishing the essential outcomes of relationship-building and reconciliation, forms of deviant resolution. Second, the TRC functioned from a *top-down approach* which was successful in opening up a robust debate on reconciliation at a national level, but it failed to translate that reconciliation experience in practical application at the local community-level context. Across South Africa there was no formal interface between the TRC and other traditional, indigenous practices of justice, healing, and reconciliation, and as such many citizens are living with the ambivalence of daily encounters with past, unresolved deviance at a neighborhood level. Third, the TRC ended up making a once-off payment of money to its victims, thereby, failing to engage perpetrators and communities in creative, meaningful efforts at *compensation and reparations* that would restore dignity, build community, and bring some form of vindication to a violent, repressive past.

The South African TRC along with many other transitional mechanisms carried the heavy responsibility of dismantling Apartheid as a governmental system that had defined deviance in an illegitimate way (the black majority was seen as deviant because of the color of their skin, their unwillingness to support the Apartheid enterprise, and their declared armed struggle against minority white rule). The African National Congress (ANC) government, voted into power in the first all-inclusive elections in April 1994, had the onerous responsibility to transform this racially discriminatory definition of deviance in a legitimate manner that would have impact across all sociopolitical spheres and societal sectors. This dilemma presented a quandary to the ANC who had politically declared themselves as a nonracial, democratic government and yet from an ideological perspective, the surest way to legitimize their freedom struggle was to declare their revolutionary violence as justified and therefore out of reach of the definitions of deviance.

However, the TRC using its legal standing as an autonomous structure (not aligned or controlled by any political party) determined to address the deviant violence of the past and not to favor any particular political affiliation in the process. The TRC significantly accomplished this by equally applying the international standard of deviant violence: gross human rights violations (GHRVs) to all armed political groups in South Africa. The five categories of GHRVs covered by the TRC were murder, attempted murder, torture, kidnapping/abduction, and disappearance. All other Apartheid government violations were outside the purview of the TRC. Every fighting faction no matter of what political or ideological persuasion was required to make legal application for amnesty with the TRC. This egalitarian approach to past deviant violence was a relief to the foot soldiers of the Apartheid security forces, but for the Liberationist groupings (freedom fighters) there was outrage and protest as they understood themselves to be fighting a just and upright cause regardless of the violent means utilized. However, what this insistence on equal treatment of all GHRVs accomplished was to send a strong signal to all aspiring governments, military formations, and the citizenry of South Africa that GHRVs would not be structurally or criminally tolerated in the future.

As was earlier indicated, the TRC did not grant “blanket amnesty” and instead instituted a conditional amnesty clause requiring all amnesty applicants to make full disclosure, establish political motivation, prove they were a member of a bona fide political organization, show that their activities were proportional to their goals, and ensure that their case(s) fell within the timeframe (1960–1993) and GHRV categories of the TRC. Upon the receipt of each amnesty application, statements were then formally investigated in order to correlate and substantiate the truthfulness of the evidence presented.

Particular issues emerged out of the context of the TRC which revealed the nuanced complexity of working within the specificity of legislated definitions of deviance. First, in the process of negotiating the TRC legislation the white-factions of the Apartheid system insisted that the TRC amnesty hearings be conducted in-camera thereby keeping the deviant violence of the Apartheid state hidden; covered-up except for only the designated few. The ANC flatly refused this demand on the claim that the granting of amnesty was already a significant political compromise. As a result, all amnesty hearings were held in public purview thereby opening up the space for the narrative discourse of the secret deviant political violence to be exposed and rearticulated in the shared memory of the national psyche.

Second, as result of these public hearings which were broadcast across the nation for the 2-year life-span of the TRC, social sanctions against the perpetrators were surfaced in the public domain. Many amnesty applicants stepped out of their positions in public office and were forced to search for employment in the private sector. Many changed identities, moved locations, and in essence went into hiding to start life over anew. This process of lustration and vetting sent a strong message to the present and future governing "powers" that deviant violence of this nature would not be tolerated in those leaders who have been entrusted with the stewardship of the "common good" of the country.

Third, as the amnesty hearings began to unfold increasing numbers of offenders began to claim Post Traumatic Stress Syndrome (PTSS) as the motivating rationale for their deviant behaviors. This trend, while substantiating a "legal argument" for impunity, caused considerable alarm in the victims' communities and among the psychological fraternity. From the professional psychological perspective, while the symptoms of PTSS are often found in common between both victim and offender in a traumatic encounter of violence, to manipulate the PTSS diagnosis as a legal justification for committing acts of violent deviance on a consistent and protracted basis over time (such as performing acts of torture on political prisoners for decades) was not scientifically acceptable. In this situation, other explanations for this form of systematic, repeated violent deviant behavior were required. One possible explanation could be the notion of anomie which refers to an individual lack of the internal sense of guiding social norms (Durkheim 1897). Other explanations have their origins in the research on various antisocial personality and psychopathic disorders that revolve around the searing of conscience, lack of empathic motivations, and the suppression of social "anchors" of right and wrong behavior appropriate to healthy relationships with others (Hare 1999; Checkley 2011).

Gacaca Courts in Rwanda

Building on the South African TRC experience, Rwanda took their transitional process a step further and engaged in a state-sanctioned hybrid model of justice which entailed incorporating a community-based approach called *Gacaca Courts* (*GCs*) which were mandated to run parallel (complimentary and supplementary) to the formal legal process of an International Tribunal that had already been launched in Arusha, Tanzania. The GCs represented an innovative justice project birthed in the midst of a prison over-crowding crisis in Rwanda following the 1994 genocide whereby 120,000 prisoners accused of crimes of genocide were being held. Based on a traditional custom of dealing with a range of community disputes, GCs involved the community electing civic leaders/traditional elders who functioned as the third-party judges in each case. These nine community arbitrators were tasked with gathering as much information as possible about the genocide activity in their local village. They then brought together the survivors, accused offenders, family support networks, and the community at large (at least 100 persons needed to be in attendance for a hearing to go forward). Truth-telling was core to the process with the use of witnesses to corroborate the findings. Opportunity for offender admission of guilt, confession, and apology were emphasized throughout the process with the occasional possibility of survivor-offender mediation and reconciliation resulting from the procedure. Forgiveness, reduced prison sentence, compensation, or punitive discipline was then decided on (official sentencing) by the community judges. With the closure of the GC process in April of this year (2011), the benefits of this process are believed to be myriad and the world is watching closely to see the long-term, potential success of this effort at building a sustainable climate of justice, healing, and reconciliation in Rwanda (Wolters 2005).

Some of the most stinging critique of the Rwandan experience of transitional justice has *not* been in relation to the internal-structure or effectiveness of the GCs hybrid justice approach, but instead the political *interference* of governmental ideology and policy that swirls around the national definition of "genocide" and as such the conception of deviant acts of violence. Vigilant against any language or action that smacked of "genocidal ideology," the current government in Rwanda outlawed the use of the ethnic designations of "Hutu" and "Tutsi," and to this day keeps a tight surveillance on any organizations that appears to only work with or favor one ethnic group to the neglect of the other. While this is quite understandable in lieu of the horrific nature of ethnic genocide that has transpired in that nation's history, it has had the unintended consequence of "silencing" a rigorous and honest debate about ethnicity and genocide deviance as experienced in that country (Tiemessen 2004).

On top of this, in an eagerness to exonerate itself from the violence of 1994, and suppress all hints of genocide ideology in the country, the ruling party has embarked on a rigorous campaign to carefully distinguish between the language of genocide (organized violence intent on eliminating an entire ethnic group) and all other forms of violence (mass killings or massacres). On the surface this linguistic differentiation seems straight forward; however, for the narrative discourse of deviance being discussed in this entry it has dangerous implications. First, it has allowed the ruling Rwandese Patriotic Front (RPF) government to excuse itself from implication in partaking in any form of deviant violence, claiming that whatever massacres or mass killings that may have transpired under its watch were strictly a consequence of war and in defense against the Genocide. Second, this clear definition of language infers that only Hutu people could have had genocide intentions and therefore they carry the responsibility for acts of past deviant violence. Third, as the GCs were set up to deal with the aftermath of the 1994 genocide specifically and not the historical violence in Rwanda generally, by implication the GCs appeared to carry a bias in favor of Tutsi and against Hutu (Mamdani 2001). Thus, with such a distinct political narrative of "genocidal ideology" as the only deviant form of violence and "massacres" and "mass killings" as justified forms of violent self-defense in the face of genocide, the conceptions of deviance have become partial and arbitrary for some communities in Rwanda. With over 100 years of genocidal history as a backdrop, a "thicker analysis" of the roots of genocide in Rwanda would need to take into consideration deviance as a product of "mimesis" especially as it relates to the deviance theories of social learning (Bandura 1977) and social disorganization (Kurbin and Weitzer 2003). Social learning theory emphasizes three critical phases: retention (observing deviant behaviors), reproduction (imitating deviant behaviors), and motivation (reinforcing deviant behaviors). Social disorganization theory speaks to the fundamental desires or needs for recognition, domination, and security that drive the performance of deviant actions especially as it relates to the phenomena of social stratification and enemy formation.

The pivotal learning here is that any process of transitional justice (whether in amnesty or prosecution) must hold an impartial interpretation of the deviant nature of violence to all persons or organizations involved in perpetrating gross human rights violations (GHRVs) regardless of the political-ideological persuasion, or whether or not one claims the "moral high ground" in the implementation of a militant struggle for freedom. While the process of the GCs allowed for a community-based micro-level complexity in the definition and management of criminal deviance to occur, at the macro-national level this contextualization of deviance was inhibited due to the government's interference in the definition of "genocidal ideology" and its labeling effect to one ethnic portion of the population. Without a demarcation against violence as deviation across the board, there will remain a mind-set of victors (nondeviant winners) and vanquished (deviant losers) which will only lend itself to continued revenge cycles in the future. Plus, the transitional justice movement runs the risk of sending the signal that certain violence is justified and permissible if it fails to hold all perpetrators of violence in equitable standing (Lemarchand 1994).

Fambul Tok in Sierra Leone

After a brutal terrorizing 11-year civil war, Sierra Leone negotiated a peace settlement in which all militia and rebel movements (including the Revolutionary United Front – RUF) were granted blanket amnesty. In response to this unconditional amnesty, the International Criminal Court (ICC) seized the moment and set up a Special Court to apply prosecution measures against the highest ranking leaders responsible for crimes against humanity perpetrated during the civil war. At a national level, a Truth and Reconciliation Commission (TRC) was instituted in a bid to promote healing and restoration across the country. The TRC embarked on a truth-telling exercise seasoned with a collective historical rewrite and the spice of public "confession-apology-forgiveness" transactions. Unfortunately, under the weight of a voluminous final report, limited funding, and ineffectual mobilization of the international and national commissioners, the TRC struggled to find innovative ways to inculcate the ownership of justice and reconciliation at a community grassroots level.

However, one of the most promising civil society responses to this dearth of community instituted justice and healing processes is the Fambul Tok (in the local Kriol language this is literally translated, "Family Talk"). Fambul Tok was launched in early 2008 by a Sierra Leonean human rights organization, the *Forum of Conscience* with long-term support from *Catalyst of Peace*, a foundation based in the United States. Fambul Tok is touted as a community-driven effort that claims no formal state sanctioning or governmental political interference in its pursuit of reconciliatory justice.

Fambul Tok boasts the following vision: "*Fostering sustainable peace in Sierra Leone through reviving our communities' traditions and values of confession, forgiveness and reconciliation*" (Fambul Tok, Annual Report 2008). The approach itself is based on traditional practices of resolving conflicts within the confines and safety of community-family networks. Planned, organized, and implemented within the local village context, the encounters integrate innovative measures of dialogical and relational justice in what is termed "truth-telling bonfires" and various traditional cultural cleansing ceremonies. These interactions are facilitated by local leaders and designated elders who provide the wisdom, moral structure, and social capital for the interface. Following these events, the momentum for reparations and reconstruction is capitalized on through practical activities of educational radio peace clubs, cross-faction football (soccer) games, and intercommunal farming projects.

Fambul Tok cases range from accusations of chicken theft during the war, serious assault, rape, and murder. The accused may be anyone from traditional community leaders, child soldiers, self-appointed militia fighters, and former army officers. For example, it was commonly known that some army officers played the role of soldier by day and rebel by night – they were euphemistically called "Solbels" by the villagers. This kind of layered violence is best understood through the social deviance theory of differential association (Sutherland 2006) and subcultural theory (Cohen 1955; Matza 1964). Differential association promulgates the idea that people will choose criminality if the balance of benefits for law-breaking outweigh the benefits of law-abiding. Subcultural theory grows out of delinquency research with urban gangs and suggests that societal subcultures form around values and attitudes that are conducive to crime and violence and as a consequence deviance then becomes the bonding agent for meeting the recognition and identity needs of individuals and groups.

Initially, 161 Fambul Tok ceremonies were planned at a chiefdom level around the country; however, the significance of this approach has spawned a great demand for this process at all levels of the society and as such there are plans to see thousands of these kinds of ceremonies conducted across the nation. With the Sierra Leonean TRC completed and the closure of the International Criminal Courts coming to an end, it is predicted that the process of Fambul Tok will play a leading role in securing a satisfying form of justice, healing, reconciliation, and ultimately durable peace in Sierra Leone for the future.

The significance of this transformative approach is described as follows:

> Fambul Tok . . . is a face-to-face community owned program that brings together perpetrators and victims of the violence in Sierra Leone's 11-year civil war through ceremonies rooted in the local traditions of the villages that were affected. It provides Sierra Leonean citizens with an opportunity to come to terms with what happened during the war, to dialogue, to experience healing, and to chart a new path forward – together. (Fambul Tok, Annual Report 2008)

A paramount learning from this case scenario is that Fambul Tok shows the potential of community-initiated programs to deliver a satisfying form of restorative justice across a nation. Despite government inabilities to translate the reconciliation experience from the top echelons to the ground, Fambul Tok illustrates the spark of innovation and the creative genius of civil society and local communities when allowed to dream and act out a better future for themselves. A cursory comparison of African customary law (indigenous justice) and restorative justice practice reveals a great deal of overlap:

- Both have the objective of reconciliation and restoration of peace and harmony in the community.
- Both share the function to serve the goal of restoring human dignity to all parties involved in the wrong-doing.
- Both promote a normative system which stresses an individual's duties, obligations, and responsibilities and not only their legal human rights.
- Both consider all offenses as human and personal wrongs against another person(s).
- Both employ procedures that are simple, cost-effective, and informal, yet powerful.
- Both encourage community participation and ownership of the process of justice; thus, those who have offended are more likely to be held accountable, to accept responsibility, apologize, and offer reparations/restitution for their wrong-doing.

Indeed, the very ideations surrounding the meaning and application of justice change in the process of reconceptualizing deviancy within the ambit of established community norms. When the significance of deviancy and the process of managing the consequences that result from deviant behaviors are placed within the worldview of broken relationships and breached harmony within community then justice becomes a socially constructed reality. Deviance and justice become contested subject matters that are negotiated through public participation, sustained dialogue, and deliberative spaces of narrative reiteration. In other words, an understanding of deviance that is embedded in human capital, social contracts, and community regulation is open to a whole new set of interpretations as opposed to the current justice system that defines deviancy in a particularly rigid, essentialist, and universal manner.

- In the first place, the definition of what qualifies as deviancy is much more flexible, nuanced, and complex in a community-based justice process than in our current justice system.
- Secondly, in a community-based justice process the understanding of deviancy becomes much more culturally appropriate and contextually situated and therefore can be more effectively sanctioned.
- Thirdly, the process of managing the consequences of deviancy is much more transitory (less legally binding and permanent) and less stigmatizing (it moves away from the current focus on labeling of offenders). Rupert Ross (1996) a Canadian magistrate writing out of the First Nations indigenous tradition where "sentencing circles" are used as a form of justice contrasts the current justice system (The Crown) as symbolic of an offender "degradation ceremony" whereas the "sentencing circles" are an example of an "reintegration ceremony."

Future Directions

Considering the novelty of this discussion on deviance embedded within a community-sanctioned process, there are numerous divergent routes that could be considered for future research:

1. Currently, there is very little debate or dialogue occurring between the legal fraternities and those working with indigenous justice processes. This is due in part to the fact that these two groups do not share a common language with which to talk about the meaning of justice and its applications. Further research is needed that will build "bridging" language between jurisprudence and the working values that guide indigenous justice processes so that a constructive conversation can be enabled between the two fields.
2. How to measure the "satisfaction" of indigenous justice processes has not been clearly defined or articulated. Research that could contribute to the development of standard frameworks for monitoring and evaluation of indigenous justice systems is of critical importance in order to give these new forms of justice the wide-spread attention they need (Van der Merwe et al. 2009). A starting point could be to borrow from research efforts in the restorative justice field such as work done by Skelton and Batley (2006) who propose four critical sets of interactions that should be used to assess the practice of restorative justice: *Encounter, Amends, Inclusion, and Reintegration.*
3. Other possible research topics to pursue relate to *recidivism* and decrease in deviant criminal behavior (longitudinal research on the impact of indigenous justice on recidivism rates is important), *scale* (can these indigenous justice systems be scaled-up to deal with the deviance of mass atrocities in a sustainable manner?), *structure* (what forms of hybrid western and indigenous models could be developed to more effectively contain deviance?), and *application* (What are the "intersectionalities" of micro and macro understandings of deviance and how can they be coordinated to efficiently administer justice?).

Conclusion

This entry has explored various applications of the definition and resolution of criminal deviance within the frameworks of transitional, restorative, and indigenous justice contexts. The contrasts and integrations of state- and community-sanctioned justice systems as well as hybrid models have been described and critiqued. Out of the various African transitional justice experiences discussed above, one could deduce that there are a number of critical elements to the process of identifying and transforming deviance that should not be overlooked. First, the definition and resolution of deviance is best served from within its own indigenous tradition by an "*insider-impartial*" (trusted "in-context" leaders) that can facilitate all the stakeholders to the conflict in a balanced manner. Second, deviance is best understood to be a breach of community harmony, well-being, and order and therefore to repair this corporate fissure the victims, offenders, extended families, and community networks must all be involved; it is a communal problem. Third, in dealing with deviance culturally appropriate platforms for truth-telling, confession, apology, forgiveness, and reconciliation should be explored throughout the process. Fourth, in dealing with deviance material forms of reparations, restitution, and compensation are expected as signs of peaceful goodwill and as indications of human responsibility and obligation to make right the wrong. Fifth, in dealing with deviation symbolic gestures of ritual healing, cleansing, and resolution are an important means of reintegration of offenders, release of the victim-survivors, and psychosocial closure for all who have been affected by the violence.

Related Entries

- ▶ Post-Conflict Traditional Justice
- ▶ Restorative Justice and State Crime

Recommended Reading and References

Bandura A (1977) Social learning theory. General Learning Press, New York

Becker H (1963) Outsiders. Free Press, New York

Braithwaite J (2002) Restorative justice and responsive regulation. Oxford University Press, New York, p 7

Checkley H (2011) The mask of sanity: an attempt to reinterpret the so-called psychopathic personality. Literary Licensing, LLC, Whitefish

Clinard M, Meier R (2008) Sociology of deviant behavior. Wadsworth Cengage Learning, Belmont, p 28

Cobban H (2007) Amnesty after Atrocity – healing nations after genocide and war crimes. Paradigm Publishers, Boulder

Cohen A (1955) Delinquent boys: the culture of the gang. Free Press, Glencoe

Dinitz S, Dynes R, Clarke A (1975) Deviance: studies in definition, management, and treatment. Oxford University Press, New York, p 4

Douglas J, Waksler F (1982) The sociology of deviance: an introduction. Little, Brown, Boston, p 10

Durkheim E (1897) Suicide – a study in sociology. The Free Press, New York

Foucault M (1977) Discipline and punish – the birth of the prison. Random House, New York

Forum of Conscience and Catalyst of Peace (2008) Fambul Tok – community healing in Sierra Leone. First Annual Report, pp 3–4

Girard R (1972) Violence and the sacred (trans: Gregory P). The John Hopkins University Press, Baltimore/London

Hamber B (2009) Transforming societies after political violence. Springer, London/New York

Hare R (1999) Without conscience: the disturbing world of the psychopaths among us. Guilford, New York

Ignatieff M (1998) The warrior's honour: ethnic war and the modern conscience. Viking, published by the Penguin Group, Toronto

Kaplan J (1973) Criminal justice: introductory cases and materials. The Foundation Press, Mineola, New York, pp 15–29

Kurbin C, Weitzer R (2003) New directions in social disorganization theory. J Res Crime Delinquency 40:374–402

Lemarchand R (1994) Burundi – ethnic conflict and genocide. Woodrow Wilson Center Press and Cambridge University Press, Cambridge, pp 160–177

Mamdani M (2001) When victims become killers – colonialism, nativism, and the genocide in Rwanda. James Curry, Oxford, pp 264–282

Matza, D. (1964) *Delinquency and Drift*. Transaction Press

Peachey DE (1989) The Kitchner experiment. In: Wright M, Galaway B (eds) Mediation and criminal justice: victims, offenders and community. Sage, London, pp 14–26

Pouligny B, Chesterman S, Schnabel A (2007) After mass crime – rebuilding states and communities. United Nations University Press, Tokyo

Rawls J (1971) A theory of justice. Harvard University Press, Cambridge, MA

Ross R (1996) Returning to the teachings – exploring aboriginal justice. Penguin, Toronto

Sen A (2009) The idea of justice. The Belknap Press of Harvard University Press, Cambridge, MA

Skelton A, Batley M (2006) Charting progress, mapping the future: restorative justice in South Africa. Restorative Justice Centre and Institute for Security Studies, Pretoria

Sutherland E (2006) The Michael-Adler report: searching for the soul of criminology. Criminology. http://www.ccjs.umd.edu/Faculty/userfiles/27/laub2006.pdf. Accessed 8 Mar 2011

Tiemessen A (2004) After Arusha: Gacaca justice in post-genocide Rwanda. Afr Stud Quart 8(1). http://web.africa.ufl.edu/asq/v8/v8i1a4.htm. Accessed 19 Sept 2009

Van der Merwe H, Baxter V, Chapman A (eds) (2009) Assessing the impact of transitional justice – challenges for empirical research. USIP Press, Washington, DC

Winslade J, Monk G (2000) Narrative mediation – a new approach to conflict resolution. Jossey-Bass, San Francisco

Wolters S (2005) The Gacaca process. Afr Security Rev 14(3). http://www.iss.co.zapubs/ASR/14No3/AWWolters.htm. Accessed 19 Sept 2009

Zehr H (1990) Changing lenses. Herald Press, Scottdale

Transitional Work Programs for Ex-prisoners

Robert Apel
School of Criminal Justice, Rutgers University,
Newark, NJ, USA

Overview

Employment has the potential to be a major point of intervention in an offender's criminal career, particularly those who have recently been released from prison. Consequently, employment-based reentry programming has obvious appeal as a policy lever intended to slow the "revolving door" of prison. President George W. Bush (a Republican) and President Barack Obama (a Democrat) have both advocated for federal grants, under the Second Chance Act, to fund prisoner reentry initiatives with employment training provisions. This is the clearest illustration imaginable that resolving the employment challenges faced by ex-prisoners is a bipartisan issue.

From a theoretical standpoint, employment provides a means of strengthening prosocial and law-abiding behavior. Rational choice theory suggests that work experience is an important source of human capital that imposes "opportunity costs" on criminal activity. Social control theory posits that high-quality jobs are a potent source of "social capital" that can override natural temptations to violate the law. Routine activities theory proposes that working fosters changes in individuals' "unstructured socializing" outside of the workplace in ways that channel them into better supervised settings with law-abiding companions. Strain theory proposes that lack of success in the labor market can motivate individuals to "innovate" through criminal behavior as a way to acquire valued goals or to alleviate negative feelings. Learning theory suggests that the workplace provides a forum in which individuals are exposed to a variety of "models," "definitions," and "reinforcements" which are conducive to law-abiding behavior.

From an empirical perspective, it is not a stretch to say that literally dozens of studies demonstrate that employment, job stability, and work quality are inversely correlated with criminal offending. While this research cannot be considered in detail in this entry, the kinds of jobs which are most strongly correlated with criminal behavior tend to be ones in which individuals earn higher wages, work for longer duration, report higher satisfaction, and work in more prestigious occupations. Empirical research thus confirms the theoretical expectation that improvement in work prospects provides one avenue for offenders to give up criminal behavior.

The labor market, and crime prevention interventions based on it, is therefore viewed as a prominent institution in the desistance process. This entry is concerned with a specific kind of employment-based reentry program known as transitional work. It begins with a look at the underlying rationale and historical background of this class of employment programs. It then provides a review of contemporary transitional work programs, highlighting recent experimental evidence of their (in)effectiveness. Because many of these transitional work programs have shown limited potential, a number of plausible explanations will be considered for their poor record of increasing employment and reducing recidivism. This entry closes with suggestions for the future design and evaluation of transitional work programs.

Rationale of Transitional Work

The rationale of employment-based reentry programming is straightforward and goes something like this: incarcerated offenders have poor work prospects, which often predate their incarceration; imprisonment substantially worsens their work prospects, in part because of stigma; and these worsened work prospects exacerbate criminal recidivism. Each of these three claims will be examined more closely.

First, individuals with a criminal record serious enough to result in incarceration have well-documented employment problems. This is due, in part, to the fact that they tend to be unskilled and poorly educated – qualities that make them unattractive to potential employers. Nationally, just 35 % of jail and prison inmates have at least a high school diploma, compared to 82 % of the general population (Harlow 2003). And the employment problems experienced by such offenders tend to be long standing. For example, Apel and Sweeten (2010) demonstrate that young people who experience their first incarceration spell exhibit unstable work histories well before they receive their first criminal conviction, compared to young people who are also convicted but are not incarcerated. Namely, they are modestly less likely to have been employed prior to conviction (60 % vs. 67 %) and work fewer weeks when they are employed (29 vs. 33 weeks). Yet a more alarming finding is that to-be-incarcerated youth show weaker attachment to legal work, as indicated by their higher probability of labor force nonparticipation prior to the conviction that led to their confinement (76 % vs. 69 %).

Second, individuals with a prison record experience deterioration in their work prospects as a consequence of their incarceration. For one, a prison record stigmatizes individuals in the marketplace. Pager (2003), for example, reports 50 % lower callback rates among job applicants who report a prison sentence on their application. Many ex-prisoners invariably discover this from their own job search experience, reporting overwhelmingly that they feel their criminal record has hindered their ability to find a job (Visher et al. 2008). Besides the stigma that ex-prisoners face in the labor market, they also accumulate substantial human capital deficiencies as a result of their incarceration, as time spent in prison is time not spent acquiring work experience. And incarceration apparently weakens ex-prisoners' already tenuous attachment to legal work, by promoting illegal income earning and labor market dropout (Apel and Sweeten 2010).

Third, studies of the determinants of criminal recidivism indicate that employment reduces the probability of, and increases the length of time to, re-offending among recently released prisoners. For example, work by Uggen (1999; Uggen and Thompson 2003) demonstrates that ex-prisoners who spend more time employed, work in higher-quality jobs, and have higher legal earnings have a significantly lower probability of criminal behavior and illegal earnings. The obvious problem is that only 35–50 % of ex-prisoners are employed in the year following their return to the community. For example, in the Urban Institute's Returning Home Study, less than half (45 %) of ex-prisoners were currently employed by the eighth month following release (Visher et al. 2008).

In principle, then, sustained efforts to improve the employment prospects of ex-prisoners have the potential to lower their risk of criminal behavior and at the same time give them the capacity to earn a livable wage that lessens the attraction of the illegal market. Furthermore, considering that hundreds of thousands of such individuals return to the community each year from state and federal prisons, successful employment interventions have the potential to benefit society at large by increasing tax revenue, improving public safety, and lessening the burden of correctional spending. One viable strategy to overcome the difficulties experienced by ex-prisoners in the labor market is transitional work.

Background of Transitional Work Programs

The US Department of Labor defines transitional work in the following manner:

> A "transitional job" is generally defined as temporary, paid work experience intended to improve participants' employability, earnings, and opportunities for advancement, and to promote self-sufficiency and long-term success in the unsubsidized labor market. (Quoted from a solicitation for grant applications under the Enhanced Transitional Jobs Demonstration, posted by the Employment and Training Administration. Accessed online at http://www.doleta.gov/grants/pdf/SGA-DFA-PY-10-11.pdf on January 9, 2012)

Transitional work programs are rooted in the *supported work model* of employment services delivery. There are four key features of these programs (Bloom 2006, 2010):

1. *Temporary subsidized work* in a minimum wage job for a nonprofit or government agency typically in a small "work crew" that is intended to strengthen peer support for employment. The workplace introduces graduated stress that increases the amount of responsibility expected from the trainee, who undergoes routine performance evaluations by the worksite supervisor.
2. *Work support services* that facilitate on-site employment and industry- or firm-specific skills acquisition including such things as transportation assistance, job coaching, and occupational skills training.
3. *Skills development* that emphasizes training in "soft skills" and general "work readiness" and can include a variety of wrap-around services related to literacy educational certification, life skills, housing assistance, legal assistance, parenting skills, and substance abuse intervention.
4. *Unsubsidized job placement*, which often includes an assessment of core competencies as well as resume development and job search assistance. Incentives might be made available to promote job stability and retention.

These programs thus have a dual focus on employment and employability. In other words, the "give 'em a job" aspect of subsidized work is supplemented by efforts to strengthen the work orientation and work readiness of ex-prisoners. The most influential precursor of modern transitional work programs was the National Supported Work Demonstration of the 1970s. This program was targeted at hard-to-employ populations – welfare recipients, youth dropouts, former drug addicts, and ex-prisoners. The program was carried out from 1975 to 1980 and evaluated in ten cities across the United States. The total cost of the program and evaluation was $82.4 million, with contributions from a variety of public, private, and philanthropic agencies (MDRC 1980).

The population of interest to this entry was the group of participants 18 years of age or older who had been imprisoned in the 6 months prior to enrollment in the program and were currently unemployed with no more than 3 months of work in the preceding 6 months ($N \approx 1{,}500$). Participants who were assigned to the supported work intervention were assigned to supervised work groups in such activities as building maintenance and construction. They were guaranteed 12 months of subsidized employment with a starting wage at or just above the prevailing minimum wage (with provisions for bonuses and merit increases).

The ex-prisoners in the study were a highly disadvantaged group (see MDRC 1980). They were mostly male (94 %) and African American (84 %), with a 27 % high school completion rate and just over 10 years of schooling, on average. Half of the sample had either never worked (11 %) or had not worked for more than 2 years (39 %). Almost one-third reported daily recent use of heroin (31 %). The sample was also highly criminally involved, with a mean of more than nine arrests, three convictions, and 195 weeks (3.7 years) of accumulated incarceration time.

Concerning the success of the program at increasing employment and reducing recidivism among the ex-prisoners, the influential evaluation by MDRC (1980) revealed only limited impact. At 3 months post-baseline, the employment rate for the control group was 38 % (with a mean 37 h and $139 during months 1–3), compared to 94 % in the experimental group (with a mean 144 h and $447 during months 1–3). Yet by 12 months post-baseline, the employment differential narrowed to 39 % in the control group (with a mean 53 h and $221 during months 10–12) and 51 % in the experimental group (with a mean 74 h and $277 during months 10–12). However, the differentials disappeared entirely after the 12th month. Thus, the large and statistically significant employment differential observed early in the study period was attributable entirely to the unsubsidized work offer provided to the experimental group. But the program had no lasting impact on employment.

The findings with respect to recidivism were even less favorable. Nine months after enrollment in the study, 34 % of the control group had been arrested, a figure that increased to 46 % by 18 months. Among the experimental group, the respective figures were 32 % and 47 %. Therefore, even during the early part of the follow-up period, when the experimental group was benefiting substantially from the subsidized work offer, the program had no apparent impact on criminal activity.

There were other noteworthy findings concerning the ex-prisoners in MDRC's (1980) evaluation. First, there was a high rate of dropout from the program, as the ex-prisoners spent an average of only 6 months in the program, far less than the other high-risk groups targeted for study. A large portion of the premature dropout was attributable to being terminated from the program job, suggesting that ex-prisoners are a particularly difficult group to manage in the workplace. Second, a significant arrest differential in favor of the experimental group was found after 36 months. Because only the earliest cohorts to enroll in the study were observed for this length of time, it suggests that early enrollees might have benefited from subsidized work throughout the study period. Third, there was some suggestion from follow-up analysis that ex-prisoners over 25 years of age experienced lasting reductions in recidivism. This led some to suggest that older ex-prisoners may be more highly motivated to take advantage of opportunities to change (Bushway and Reuter 2004). However, this interesting finding was not observed among the other targeted groups in the study, rendering any interpretation of it inconclusive.

(In)Effectiveness of Modern Transitional Work Programs

An example of a modern transitional work program is the Center for Employment Opportunities (CEO), based in New York City. The CEO program has all four characteristics of the supported work model. (The ensuing description relies heavily on material provided on CEO's website [www.ceoworks.org] as well as a program description published by the National Institute of Justice [Finn 1998].) The program begins with 4 days of "life skills education" in order to establish work readiness. This includes instructional workshops related to resume writing, job search, interviewing, and workplace behavior. Following completion of the pre-employment course, participants spend 1 day with an employment specialist, who assesses the participant's interests and develops an employment plan.

At the start of the second week, participants begin employment in a subsidized job in a crew of about 5 workers. The work involves performing maintenance, janitorial, and grounds keeping services to city or state agencies and educational institutions. Workplace supervision is provided by a CEO staff member, who provides coaching in work performance as well as enforcement of rules of conduct (e.g., punctuality, dress code, workplace behavior). Participants are paid at the end of each workday. Four days each week are spent at the workplace, while the fifth day is spent in the CEO office with the employment specialist, who provides additional coaching. When participants are deemed work ready, they are matched with a job developer, who lines up job interviews in the private sector for positions which match the skills and interests of each individual.

For a period of 6 months after participants acquire an unsubsidized job, a CEO retention specialist continues to provide work-related counseling, career planning, and job search assistance. As an incentive for stable employment in a full-time job, and for up to 1 year, monthly bonuses are made available to participants who present their pay stubs. Throughout the CEO program period, moreover, specialized programs offer training in parenting skills.

An evaluation of 2-year CEO program impacts was conducted by Redcross et al. (2009). Almost 1,000 participants were randomly assigned to either receive the full complement of CEO program services or job search assistance only. The participants were largely male (93 %), African American (64 %), or Hispanic (31 %) and had either no educational certification (42 %) or only a GED certificate (43 %). While 81 % of the sample had ever been employed, just 61 % had worked for the same employer for 6 consecutive months, and only 24 % had been employed in the year prior to baseline in a job covered by the state unemployment insurance (UI) system. The average participant had over seven prior arrests (over four felony arrests), almost seven prior convictions, and had accumulated 60 months in state prison.

Examination of the filtering of the experimental group through the stages of the CEO program illustrates the challenge of retaining high-risk individuals in a transitional work program. Only 79 % completed the pre-employment course, 72 % worked in a subsidized job (for a mean 8 weeks), 57 % met with a job developer, and 30 % were successfully placed in an unsubsidized job, either through CEO or on their own initiative.

With respect to CEO program impacts on employment, the experimental group was far more likely to work in a UI-covered job (81 %) than the control group (57 %) in year one, as expected. However, the program had no impact on the subsequent likelihood of employment, as half of both the experimental and control groups worked in a UI-covered job during year two. Although not statistically significant, the experimental group did have higher UI earnings in year two than the control group ($5,167 vs. $4,670), for a difference of almost $500 in annual earnings. Apparently, the initial employment differential was attributable entirely to the subsidized job offer, as CEO had no impact on unsubsidized employment, even though monetary incentives were provided.

With respect to CEO program impacts on recidivism, the difference in 2-year arrest likelihood was not statistically significant, with 38 % of the experimental group and 42 % of the control group experiencing an arrest, a relative difference of 10 %. The experimental group also had a lower risk of conviction (31 % vs. 38 %) as well as a lower risk of incarceration (49 % vs. 55 %), differentials that were both statistically significant. The rate of incarceration for a new crime was also lower, although not significantly so, among the experimental group (14 % vs. 17 %). Interestingly, CEO program impacts on arrest and conviction were concentrated in the second year of the follow-up. For example, while there was no difference in arrest in year one for the experimental group (22 % vs. 23 %), there was a significant difference in year two (23 % vs. 27 %).

The perplexing finding from the CEO evaluation is that the employment differential was observed in year one, during the subsidized work phase, while the recidivism differential was observed in year two, during the unsubsidized work phase. Therefore, the impact of the program on recidivism was due to something other than the program's impact on unsubsidized employment, contrary to the theoretical model underlying transitional work programs. The authors of the evaluation speculated that the extensive wrap-around services (e.g., support and guidance provided by CEO staff), and not employment per se, was responsible for the recidivism reduction (Redcross et al. 2009). Irrespective of the underlying mechanism, though, the impact of CEO on recidivism was apparently long lasting.

As it turns out, the recidivism reductions observed in the CEO evaluation are the exception rather than the rule. Yet even the results from CEO are difficult to interpret. A recent evaluation of the Transitional Jobs Reentry Demonstration yielded no program impacts on either unsubsidized employment or recidivism (Redcross et al. 2010). A meta-analysis of employment-based reentry programs performed by Visher et al. (2005) concluded unambiguously that such programs were ineffective in increasing employment or reducing recidivism. The CEO experience aside, therefore, transitional work programs for ex-prisoners inspire little optimism about their utility as a policy option.

Why Transitional Work Does Not Work

Transitional work programs tend to produce large in-program effects on employment and earnings, but these gains tend to be lost during the post-program period. In fact, a year after the start of the program, transitional work participants are often indistinguishable from nonparticipants in their labor market experiences. In both cases, employment prospects remain dim. Nor is there a recidivism reduction, during either the in-program period or the post-program period. What could possibly account for this poor track record? There are a number of possibilities, only a few of which will be considered in detail here.

First, compliance problems tend to plague field experiments of the sort evaluated in transitional work programs. Individuals assigned to a control condition can seek non-program assistance or training on their own, while individuals assigned to an experimental condition can refuse to show up for training or subsidized employment. For example, among the ex-prisoners in the National Supported Work Demonstration, 33 % were fired from their program job, and another 20 % were terminated for other negative reasons such as reinstitutionalization (MDRC 1980). Noncompliance is problematic for evaluation of the "treatment effect" of transitional work. The problem arises from the fact that the causal estimand is the "intention to treat" (ITT), that is, the causal impact of randomized offers to participate in a work program. When there is substantial noncompliance with the program conditions, the ITT can be attenuated – often substantially so – relative to the "average treatment effect on the treated" (ATT), or the causal impact of actual participation in a work program. Therefore, the failure of transitional work programs may be more apparent than real, because a substantial portion of participants are either terminated from or voluntarily leave the program.

Second, ex-prisoners face inequality in the job search process. Recall from Pager's (2003) audit study that ex-prisoners had a substantially lower callback rate from prospective employers. Employer reluctance to hire individuals with a criminal record has also been confirmed by surveys of employers themselves. For example, Holzer et al. (2004) found that 62 % of the establishments they surveyed reported that they would "probably not" or "definitely not" have hired someone with a criminal record to fill their most recent job opening. Employer willingness to hire offenders (38 %) for low-skill positions was markedly lower than their willingness to hire welfare recipients (92 %), applicants with only a GED (96 %), applicants who had been unemployed for a year or more (83 %), and applicants with a spotty work record (59 %). Therefore, the failure of transitional work programs to improve the employment prospects of ex-prisoners could be attributable to demand-side preferences that severely penalize individuals with a prison record.

Third and finally, the time is probably right to consider the possibility that transitional work is based on a model of service delivery (and a theory of behavior change) that is outdated and outmoded, at least in the context of prisoner reentry. Visher et al. (2005), for example, lament that most of the programs in their meta-analysis are quite old. Bloom (2010) observes that transitional work programs are based to a large extent on a welfare-to-work model. While transitional work does have a good track record of improving the employment prospects of former welfare recipients, ex-prisoners represent a population with deficits that make them much harder to serve than welfare recipients. It might be that transitional work, as currently conceived, is simply not a good prison-to-work model. Furthermore, since the 1970s, when the basic features of modern transitional work programs were put into place, a much better understanding of "what works" in correctional rehabilitation has been acquired. Indeed, integration of modern correctional principles might inspire the design of future transitional work programs.

Conclusion

The theoretical rationale of transitional work programs for ex-prisoners is unassailable. Since ex-prisoners have long-standing work difficulties that are worsened by imprisonment, focused efforts to improve their employment and employability should produce lasting individual and societal benefits. Regrettably, transitional work programs have not been shown to uniformly increase employment and reduce recidivism. Indeed, evidence of effectiveness is the exception rather than the rule. And exceptional programs (e.g., CEO) tend to produce perplexing results, namely, recidivism reductions in the absence of better unsubsidized work prospects. A brutally honest assessment of the evidence therefore compels this author to conclude that "transitional work does not work," at least not in the context of prisoner reentry and not as currently designed. Claims to the contrary are Pollyannaish.

While this conclusion might be needlessly pessimistic, there are still good reasons to believe that transitional work programs can be salvaged. Collaboration between criminal justice officials, service providers, and academic researchers can help transform good intentions into good practice, with greater fidelity to the expanding knowledge base about "what works" in the design of correctional interventions. Successful efforts are likely to involve a combination of strategies, including integration of prison- and community-based services, experimentation with a variety of financial incentives for ex-prisoners as well as employers, and systematic efforts to address substance abuse and mental health problems.

The fact that more than one-half million individuals leave the nation's prisons each year lends urgency to a renewal of innovative employment-based strategies designed to ease their reentry into the community. Transitional work programs should be central to such efforts, and steps taken to understand why these programs are ineffective, and what can be done to make them effective, will yield dividends for future reentry policy. Such efforts promise to improve the lives of ex-prisoners, reduce their burden on the criminal justice system, and protect society at large.

Related Entries

- ▶ Desistance from Crime
- ▶ Effective Community Supervision Programming

- ▶ Examining the Effectiveness of Correctional Interventions
- ▶ History of Randomized Controlled Experiments in Criminal Justice
- ▶ Randomized Experiments in Criminology and Criminal Justice
- ▶ Re-entry

Recommended Reading and References

Apel R, Sweeten G (2010) The impact of incarceration on employment during the transition to adulthood. Soc Probl 57:448–479

Bloom D (2006) Employment-focused programs for ex-prisoners: what have we learned, what are we learning, and where should we go from here? Paper presented at the National Poverty Center, Gerald R. Ford School of Public Policy, University of Michigan, Ann Arbor

Bloom D (2010) Transitional jobs: background, program models, and evaluation evidence. Unpublished manuscript. MDRC, New York. mdrc.org/publications/553/full.pdf. Retrieved online on 13 Jan 2012

Bushway SD, Reuter P (2004). Labor markets and crime. In: James QW, Joan P (eds) Crime: public policies for crime control. Oakland, CA: Institute for Contemporary Studies

Finn P (1998) Successful job placement for ex-offenders: the center for employment opportunities. Program Focus No. 168102. National Institute of Justice, Washington, DC

Harlow CW (2003) Education and correctional populations. Bureau of Justice Statistics Special Report No. NCJ 195670. US Department of Justice, Washington, DC

Holzer HJ, Raphael S, Stoll MA (2004) Will employers hire ex-offenders? Employer perceptions, background checks, and their determinants. In: Patillo-McCoy M, Weiman D, Western B (eds) Imprisoning America: the social effects of mass incarceration. Russell Sage, New York

Manpower Demonstration Research Corporation [MDRC] (1980) Summary and findings of the national supported work demonstration. Ballinger, Cambridge

Pager D (2003) The mark of a criminal record. Am J Sociol 108:937–975

Redcross C, Bloom D, Azurdia G, Zweig J, Pindus N (2009) Transitional jobs for ex-prisoners: implementation, two-year impacts, and costs of the center for employment opportunities (CEO) prisoner reentry program. MDRC, New York

Redcross C, Bloom D, Jacobs E, Manno M, Muller-Ravett S, Seefeldt K, Yahner J, Young AA Jr, Zweig J (2010) Work after prison: one-year findings from the transitional jobs reentry demonstration. MDRC, New York

Uggen C (1999) Ex-offenders and the conformist alternative: a job quality model of work and crime. Soc Probl 46:127–151

Uggen C, Thompson M (2003) The socioeconomic determinant of ill-gotten gains: within-person changes in drug use and illegal earnings. Am J Sociol 109:146–185

Visher CA, Winterfield L, Coggeshall MB (2005) Ex-offender employment programs and recidivism: a meta-analysis. J Exp Criminol 1:295–315

Visher C, Debus S, Yahner J (2008) Employment after prison: a longitudinal study of releasees in three states. Justice policy center research brief. Urban Institute, Washington, DC

Transnational Corporate Bribery

Nicholas Lord[1] and Alan Doig[2]
[1]School of Social Sciences, Cardiff University, Cardiff, Wales, UK
[2]International Development Department, The University of Birmingham, Edgbaston, Birmingham, UK

Synonyms

Transnational corporate corruption

Overview

Large-scale investigations involving multinational corporations (MNCs) such as Siemens and BAE Systems – the former involved a system of slush funds used to pay bribes to win overseas contracts, while BAE was investigated by the UK Serious Fraud Office for alleged commissions and hospitality payments to Saudi officials involved in a major arms procurement – demonstrate how large commercial enterprises may be the subject of allegations of bribing public and private officials to further or maintain their business interests.

Some would accept that as the cost of development or the unavoidable interdependence between licit and illicit commerce in "gray" markets, others would argue that corruption and bribery have devastating consequences, in particular for developing countries where much corporate corruption is directed. Increasingly, the argument that corruption is very harmful is having a significant impact on the way international business is seen and has shaped legal frameworks and enforcement practices, particularly in terms of international conventions and business-focused initiatives. These, however, not only raise questions about definitions of what is corruption but also of the causes of corporate corruption. These questions are also central to an understanding of why and where corporate corruption occurs and whether control strategies would work and, if not, why not.

Addressing Transnational Corporate Corruption

What Is Corporate Corruption? Definitions and Disciplines

Bribery is only one offense within the wider term of corruption, as academics and commentators try and encapsulate into a single term that conduct and behavior which reflects private or other partisan interests over official duties in both public and private sectors. Defining "corruption," has long been the subject of conceptual and definitional debate, in part because of differing country and cultural perceptions and in part because of the wish to seek as near as possible universal definition that allows transnational applicability. The terms bribery and corruption are often used synonymously. Bribery is considered the main aspect of corruption, which is often interpreted more widely in terms of conduct that may be in breach of external criminal or regulatory requirements or deviates from the formal internal expectations, governing holding of an appointment in public or legal entities.

It is possible to locate such conduct and its control in the criminological literature within traditional conceptual debates around "corporate" and "white-collar crime" – the ground-breaking work of Edwin Sutherland (1883–1950) that pioneered criminological research into this area. Sutherland's investigation of business or corporate crime was based on studying the ways in which "crime" was learned by "differential association." His focus was very much on corporate wrongdoing in its widest sense (such as restraint of trade, unfair labor practices, and stock manipulation). His themes have been developed, expanded, and adapted by academics subsequently to include computer-related crime, management fraud, and corruption, with an emphasis on motivations, deterrence, criminality, offenders, and recidivism. Other criminologists study corporate corruption within the context of state-corporate crime, including state-initiated corporate crime and state-facilitated corporate crime. These argue that governments either (a) allow corporations to act deviantly on behalf of the former to achieve the former's objectives or (b) more generally work to minimize any regulation of the latter's deviant activities.

Whichever the intradisciplinary perspective within criminology, however, the specific topic of corporate corruption has not been successfully defined and conceptualized. Empirical *criminological* research into corporate corruption, with the exception of a few notable studies (see, e.g., Clinard 1990), is sparse, with most literature either analyzing the broader concepts of corruption and bribery found within other academic disciplines, such as political science, sociology, economics, and law, or taking a broader management handbook approach (see, e.g., Brytting et al. 2011).

Within the field of political science, Joseph Nye offered an early definition of corruption, seeing it as "behaviour which deviates from the formal duties of a public role because of private-regarding (personal, close family, private clique) pecuniary or status gain; or violates rules against the exercise of certain types of private-regarding influence" (1967, p. 419). The definition addresses the key distinction between private and official roles which has continued to be of significance and results in a primary focus on public officials, as can be seen in more recent practitioner definitions:

– **World Bank** (www1.worldbank.org/publicsector/anticorrupt/corruptn) – the term corruption covers a broad range of human actions. To understand its effect on an economy or a political system, it helps to unbundle the term by identifying specific types of activities or transactions that might fall within it. In considering its strategy, the Bank sought a usable definition of corruption and then developed a taxonomy of the different forms corruption could take consistent with that definition. We settled on a straightforward definition – the abuse of public office for private gain. Public office is abused for private gain when an official accepts, solicits, or extorts a bribe. It is also abused when private agents actively offer bribes to circumvent public policies and processes for competitive advantage and profit. Public office can also be abused for personal benefit even if no bribery occurs, through patronage and nepotism, the theft of state assets, or the diversion of state revenues. This definition is both simple and sufficiently broad to cover most of the corruption that the Bank encounters, and it is widely used in the literature.

– **Transparency International** (**TI**) (www.transparency.org/news_room/faq/corruption_faq) – corruption is operationally defined as the abuse of entrusted power for private gain. TI further differentiates between "according to rule" corruption and "against the rule" corruption. Facilitation payments, where a bribe is paid to receive preferential treatment for something that the bribe receiver is required to do by law, constitute the former. The latter, on the other hand, is a bribe paid to obtain services the bribe receiver is prohibited from providing.
– **INTERPOL Expert Group on Corruption** (www.interpol.int/Public/Corruption/IGEC) – corruption is any course of action or failure to act by individuals or organizations, public or private, in violation of law or trust for profit or gain.

A more sociological definition that focuses on the social relationships involved in bribery and corruption and develops the legal aspect is offered by Deflem, who defines corruption as a "colonisation of social relations in which two or more actors undertake an exchange relation by way of a successful transfer of the steering media of money or power, thereby sidestepping the legally prescribed procedure to regulate the relation" (1995, p. 243). This second definition highlights the key role of social interaction as well as developing a more normative aspect in relation to socially constructed laws. It also allows private sector-private sector corruption to be explored in a more balanced way.

In a preface to an edited volume of articles on definitions, however, Robert Williams noted: "the first volume in this collection, *Explaining Corruption*, is more conceptual then empirical. But despite the giant lava flow from the "corruption eruption", the literature on this aspect of corruption still generates more heat than light. Many contributions appear ideological in character and corruption remains a highly contested concept. We still lack firmly grounded theories of corruption and the shortage of analytically informed empirical inquiries continues. Too many participants in the contemporary corruption debate content themselves with resurrecting tired clichés about the topic or with slaying long-dead dragons" (Williams 2000, p. xi).

The difficulties of a definition with universal applicability moves toward subjective interpretation and country-specific relevance, where what may be seen as criminal in one context or country may be unethical in another or acceptable in a third (gift giving is one area where these issues are often debated). For such reasons, other organizations have chosen not to seek a general definition but imply a definition by content. Thus, the OECD argues that

> The OECD, the Council of Europe and the UN Conventions do not define "corruption". Instead they establish **the offences for a range of corrupt behaviour**. Hence, the OECD Convention establishes the offence of bribery of foreign public officials, while the Council of Europe Convention establishes offences such as trading in influence, and bribing domestic and foreign public officials. In addition to these types of conduct, the mandatory provisions of the UN Convention also include embezzlement, misappropriation or other diversion of property by a public official and obstruction of justice. The conventions therefore define international standards on the criminalisation of corruption by prescribing specific offences, rather than through a generic definition or offence of corruption. (OECD 2008, p. 22. **Bold** in original)

While specific, legally derived bribery definitions encompass the private sector, the wider definitions, as well as the emphasis on direct private gain, may be less clear (as is the active and passive distinctions and the specific transaction of a bribe for a specific advantage). In particular, taking the definition of bribery in the conduct of the enterprise's business in Transparency International's Business Principles for Countering Bribery, "the offer or receipt of any gift, loan, fee, reward or other advantage to or from any person," may be intended to be an inducement to do something which is dishonest or illegal, but it is not necessarily a breach of trust when the company itself decides to use bribes to secure contracts. Indeed, it should be noted that some countries do not criminalize bribery of the private sector, while others do not prioritize corruption in the private sector that does not involve the public sector. Such a decision may also not be related to a specific contract but to creating a climate of favoritism toward the company. An easier approach may be to view definitions through the OECD's definition of what is not acceptable:

> No offer, promise or give undue pecuniary or other advantage to public officials or the employees of business partners. Likewise, enterprises should not request, agree to or accept undue pecuniary or other advantage from public officials or the employees of business partners. Enterprises should not use third parties such as agents and other intermediaries, consultants, representatives, distributors, consortia, contractors and suppliers and joint venture partners for channelling undue pecuniary or other advantages to public officials, or to employees of their business partners or to their relatives or business associates. (OECD 2011, p. 45)

Given such a breadth of offenses and an increasingly subjective interpretation of corruption, a number of commentators distinguish between what might be considered the core corruption offense of bribery and other offenses (some of which, such as the inclusion of embezzlement or money laundering in the UN Convention against Corruption 2005 (UNCAC), would be termed and indeed prosecuted as fraud, economic crime, or financial crime rather than corruption). Certainly some countries and commentators would attach "corruption" to conduct that, while it may be considered self-regarding or in contravention of their official roles and responsibilities (such as the direction of state resources for election purposes), may be criminalized under general legislation (misappropriation) or may be considered unacceptable or unethical but not necessarily illegal depending on context, circumstance, or country (e.g., the traffic of public officials on resignation or retirement to private appointments in areas or activities relating to their previous employment; see, e.g., David-Barrett 2011).

Definitions of bribery direct us toward the key actors, relations, processes, objects, and employment settings that exist during the processes of what is essentially a transaction. Distinctions between "active" (i.e., giving, offering) and "passive" (i.e., soliciting, accepting) bribery, or otherwise thought of as "supply" and/or "demand" side bribery, are also made. So defined, bribery poses multiple complexities for international and state agencies aiming to legislate for, control, and regulate corruption, but these agencies and actors are largely concerned with illegal acts. Wider "abuse of public office" or violation of "trust" definitions fall into a gray area of bribery and corruption but, along with issues of non-contract-specific payments (such as the US Foreign Corrupt Practices Act 1977's reference to bribes to obtain or retain business, the use of facilitation payments or the UK Bribery Act 2010's reference to excessive hospitality), are also requiring moral definitions or subjective assessments to become more narrowly defined and restrictive legal interpretations.

Corporate Bribery and Corruption: Sources and Causes

One cause of the lack of systematic attention to corporate bribery and corruption is the absence of valid data on its prevalence and practices on which to base definitions and explanations. There is anecdotal information on both public-private and private-private sector bribery and corruption, although material on the latter is even thinner than that on the former. The latter, for example, is discussed in material available from consultancies (see, e.g., Control Risks 2009) and occasional academic research. Material on the former within a specific country is more extensive in that it relates to bribery of public officials; data may be obtained from a wide range of academic and practitioner institutions although most of the emphasis has been on the public sector than on the corporate sector.

One consequence has been that in attempting to explain the causes of corporate corruption, as with other corporate and white-collar crimes, the lack of empirical research in this area, particularly that which engages with the corporate perspective, makes any concrete analysis difficult and leaves researchers seeking to view corporate conduct through traditional explanations of crime. These have sought to explain crime through classicist or neoclassical models (e.g., rational choice theories), individual and sociological positivist theories (e.g., anomie, strain, learning and biological theories, social disorganization, subcultures), and later social constructionist notions (e.g., labelling perspective). Such theories largely neglected white-collar and corporate crimes, instead focusing on crimes occurring within lower socioeconomic areas by so-called conventional offenders. Corruption, however, could be explained through various sociological/criminological, cultural, political-economic, and structural approaches which resonate at the individual, organizational, and societal levels. Table 1 (below) illustrates these three dimensions although the examples here are not comprehensive and there may be significant interplay and overlap in any given act of corruption.

A number of these attempts to explain the causes of corruption have adopted rational, economic-based models tied in with the level of government intervention. A key example of this is Klitgaard's (1998, p. 4) corruption formula – "corruption equals monopoly *plus* discretion *minus* accountability" – which argues that, whether the activity is public, private, or nonprofit and irrespective of where it occurs, corruption will tend to be found when an organization or person has monopoly power over a good or service, has the discretion to decide who will receive it and how much that person will get, and is not accountable. Furthermore, "corruption is a crime of calculation, not passion" (1998, p. 4).

Transnational Corporate Bribery, Table 1 Explanations of corruption

Level	Causality	Examples/contexts
Individual	Economic rationality, pathology	Individual greed or individuals who think they can "get away with it"
Organizational	Complex corporate structures, devolved decision making, implied consent, organizational goals, ambiguous accountability	Direct – pervasive and organized, for example, company policy to use inappropriate hospitality or bribery to facilitate business
		Indirect – pervasive but unorganized, for example, managerial/cultural pressure to meet targets and profit margins
Social structural	Demands of/ opportunities within (export) markets, criminogenic political economy (i.e., promotion of economic power and profits)	Common and expected business practice to bribe in some overseas jurisdictions – "everyone does it"

Rational choice theory as applied by Shover and Hochstetler to white-collar crime would also argue that "all criminal decision making represents a single field of study and that much of what has been learned in studies of street criminals and their decision making almost certainly is paralleled in white-collar crime decision making" (2006, p. 3).

This ties in with Gottfredson and Hirschi's (1990) "general" theory of crime that also adopts this model of rationality, suggesting human nature is motivated by self-interest and a pursuit of pleasure over pain, therefore echoing the earlier classicist notions of Cesare Beccaria and Jeremy Bentham. They suggest the official crime data they analyzed showed few differences between "conventional" and "white-collar" offenders in relation to variables such as age, social class, and gender and that short-term gratification is a key feature of conventional and white-collar crime. Individuals' lack of self-control along with the absence of attachments to others and sufficient controls leads individuals to offend. This approach has been greatly criticized, most notably by Steffensmeier (1989), on the grounds that their sample of white-collar criminals was biased toward lower-class offenders, and by Slapper and Tombs (1999). The latter point out that explanations of crime in relation to human nature do not account for sociological, environmental, cultural, and social structural influences on behavior. In addition, such control theories neglect that organizational crime involves the pursuit of organizational goals ahead of individual gratification (Friedrichs 2007, p. 205).

Economic and individual rationality, then, should be considered only one perspective in understanding the causes of corruption as explanations of the causes of corruption can be found within legal, economic, social, organizational, and cultural reasoning. For example, other explanations of white-collar and corporate crime focus on pathologies at the individual level whereby individual characteristics and personality traits differentiate offenders and non-offenders. It has, for example, been acknowledged that corporations often put such wrongdoing down to "rotten apples in the barrel" albeit such explanations are seen as inappropriate and have little evidence to support any connection between white-collar crime and individual pathologies (see Croall 2001, pp. 83–85). Certain characteristics such as a tendency to take risks, recklessness, ambitiousness and drive, and egocentricity and a hunger for power have been associated with this type of offending, and as Friedrichs notes, "some white collar crime does seem to be difficult to explain without reference to personality and character" (2007, p. 203).

The work of Matza and Sykes (1961) that focused on "techniques of neutralisation" is one theory that has been applied somewhat successfully to corporate crimes where offenders justify or neutralize their behavior. Justifications include "denial of responsibility," "denial of injury," "denial of the victim," "condemnation of the condemners," and "appeal to higher loyalties." Such culturally based justifications can demonstrate corporate attitudes toward certain wrongful activities. For example, Chibnall and Saunders (1977), who examined business corruption and bribery in England in a different era, showed that business individuals, despite being fully aware of their illegal actions, were able to justify them by referring to the pervasiveness of such practices in business. As Clinard and Yeager note, "[a] variety of justifications are available to those executives who are confronted with doubt or guilt about illegal or unethical behaviour; these justifications allow them to neutralize the negative connotations of their behaviour" (1980, p. 67). Indeed, as Croall notes, "[s]ome illegal activities are regarded as 'normal' business practice and others are readily justifiable as being for the 'good of' the organisation" (2001, p. 94).

The Corporate Organization

Much of the organizational deviance literature looks at managers and executives who may not only act deviantly on behalf of the organization but also against it. In looking at the private sector, Punch discusses a number of contextual possibilities that may trigger corporate deviance. These may be structural – markets (either competition or anti-competition factors); size of organizations (which dilutes both control and responsibility); achieving organizational goals; opportunity; total loyalty to the company; corporate culture, and personal, depersonalization in deference to the corporate ideologies and practices; corporate-focused rationalization; business as war; risk-taking; dominant personalities; necessary dirty hands as part of corporate competition; pressure; and rewards, which lead to the development of a "corporate mind." The mind is attuned to the "signals" from the corporate environment (what they are expected to do, what others are also doing, or what it takes to succeed) that are set by the "fundamental concerns of senior management. . .centred on corporate survival, continuity, power, reputation/face, and profits. . ." (Punch 1996, p. 239).

The question might be asked – deviant in relation to domestic legal frameworks, to legal frameworks in export markets, or to conduct considered acceptable in domestic markets and replicated in export markets? Certainly recent work of criminologists on organizations, culture, and deviance would, as in the case of Clinard and Yeager (1980), argue a shift away from theories of deviance and crime applicable to individuals to a focus on the corporation as a complex organization. By doing so, they sought to adopt organizational theory in their understanding of corporate crime. They state:

> [t]he immensity, the diffusion of responsibility, and the hierarchical structure of large corporations foster conditions conducive to organizational deviance. In addition, the nature of corporate goals may promote marginal and illegal behaviour, as may the characteristics and the social climate of the industries within which the firms operate (1980, p. 43).

Organizational understandings of corporate crime, and therefore corporate corruption, take into consideration individual, organizational, and sociological factors which can influence corporate offending, and Schrager and Short succinctly explain this relationship when they state that

> . . .preoccupation with individuals can lead us to underestimate the pressures within society and organisational structures which impel those individuals to commit illegal acts. . .recognising that structural forces influence the commission of these offences does not negate the importance of interaction between individuals and these forces, nor does it deny that individuals are involved in the commission of illegal organisational acts. It serves to emphasise organisational as opposed to individual etiological factors, and calls for a macro sociological rather than an individual level of explanation (1977, p. 410; see also Punch 1996).

Further certain intradiscipline perspectives argue that such factors cannot be viewed in isolation from the wider political and economic context. Thus, the state-corporate approach "directs attention toward the way in which deviant organisational outcomes are not discrete acts, but rather the outcome of relationships between different social institutions. Second, by focussing on the relational character of the state, the concept of state-corporate crime foregrounds the ways in which horizontal relationships between economic and political institutions contain powerful potential for the production of socially injurious actions. This relational approach, we suggest, provides a more nuanced understanding of the processes leading to deviant organisational outcomes than approaches that treat either businesses or governments as closed systems. Third, the relational character of state-corporate crime also directs us to consider the vertical relationships between different levels of organisational action: the individual, the institutional, and the political-economic" (Kramer and Michalowski 2006, p. 22).

A more extreme analysis within this perspective would propose the organization as reflecting macrolevel influences, which ties in with social structural theories explaining corporate criminality as well as with theories from Marxist scholars who argue that capitalist economic structure and ideology itself is criminogenic. For example, Durkheim's "anomie" and later strain theories have been applied to white-collar criminals in relation to "boundless aspirations" and "greed" that affect all individuals (see Croall 2001, p. 91). Furthermore, in terms of the economic structure, Slapper and Tombs (1999, p. 141) argue that corporate offending is endemic to capitalism. This creates "amoral" corporations that are focused on profit above business ethics, although ethical business activities may be used as a means of creating profit. On the other hand, other criminologists would argue that it is not:

> convincing to assume that all businesses act as "amoral calculators" and would choose to offend but for the availability of serious sanctions. . .The desires to continue in business and to maintain self-respect and the good will of fellow businessmen, go a long way to explaining reluctance to seize opportunities for a once-only windfall. . .Trading competitors (as well as organized consumer groups, unions, and others) can serve as a control on illicit behaviour for their own reasons. Law-abidingness can often be definitely in the competitive interests of companies. Marxist theory has no need to assume that all business crime will be tolerated. Many forms of business misbehaviour made into crimes may reflect changing forms of capitalism or inter-class conflict. At any given period, some corporate crimes, such as anti-trust offences, will not be in the interest of capitalism as a whole, so it is important to distinguish what is in the interests of capitalism from what suits particular capitalists. Even if the latter may succeed in blocking legislation or effective enforcement, at least in the short term, this does not prove that it is capitalism as such which requires the continuation of specific forms of misbehaviour (Nelken 2007, pp. 747, 748).

As Slapper and Tombs note, however, while it is difficult to pinpoint why a company at a point in time may opt for corrupt business dealings or to understand why "normal" rather than pathological organizations should act this way, societal context or culture is crucial: "a capitalist economy constantly pushing people with targets to hit, promotions to seek and emotions to avoid, recessions to try and survive, and so on, can thus be seen as a society likely to engender corporate crime" (Slapper and Tombs 1999, pp. 161–162).

On the other hand, given Nelken's comments, assumptions about likelihood do not address the *absence* of corporate corruption in some companies and/or in some countries. Part lies with the increasing extraterritorial reach of developed countries' law enforcement and justice bodies; part lies with the promotion of international instruments to foster cooperation, asset recovery, and so on. Part lies in the increasing influence in some countries of the corporate governance agenda which sees both corporate value and profitability in promoting corporate social responsibility, environmental protection, fair trade, and organizational standards. In part, it may also be the consequence of other (often related) factors, such as government's interests, the specific markets in which the companies are working, and the often asymmetrical and inefficient nature of market competition which are likely to impact more on countries exporting internationally than on those with limited exposure to more volatile and corruption-prone contexts.

Countries, Contexts, and Consequences

It is clear that a number of theoretical perspectives and ideas can be relevant to explanations of corporate corruption. Also clear is the complexity between individual, organizational, social structural influences, as well as the legal, economic, and cultural frameworks within which transnational corporations operate. Thus, Slapper and Tombs suggest corporate offending can only be understood in the context of wider political economy and that any theoretical analysis of corporate crime should include various social structures in relation to capitalism and inequality. They suggest that

> [k]ey aspects of any such analysis would integrate understandings of the general state of national and the international economy, the nature of markets, industries, and the particular products or services with which particular corporations are involved, dominant ideologies and social values, formal political priorities and the nature of regulation, particular corporate structures, the balances of power within these, the distribution of opportunities within and beyond these, and corporate cultures and socialisation into these (1999, p. 160).

This takes us to corporate bribery and corruption in terms of the international dimension, where the volume of academic and practitioner material is much more extensive. Two points are worth making. First, as discussed below, the anticorruption drivers in terms of corporate bribery and corruption derive from international organizations as part of their development agenda. As a consequence, developed countries associated with that agenda have been those who have increasingly required domestic business to adhere to domestic legislation not to engage in bribery. On the other hand, businesses from emerging economies are also heavily involved in lesser developing countries but whose governments are less committed to enforcement of anticorruption measures (e.g., those of China, India, and Russia have been identified (see Transparency International's 2008 Bribe Payers Index)). Second, reflecting on Slapper and Tombs' comments, how business behaves in its domestic market and different international markets may depend significantly on country, context, and product, as well as the propensity of the industry or sector itself to dishonest practices in pursuit of markets and profitability.

Thus, the World Health Organization (WHO) has reported in 2009 that "the medicines chain refers to the steps required for the creation, regulation, management and consumption of pharmaceuticals. Corruption in the pharmaceutical sector occurs throughout all stages of the medicines chain, from research and development to dispensing and promotion. . .countries with weak governance within the medicines chain are more susceptible to being exploited by corruption. These countries lack: appropriate legislation or regulation of medicines; enforcement mechanisms for laws, regulations and administrative procedures; conflict of interest management" (WHO 2009). Similarly, the arms trade is known for not only the interdependence between companies and governments, whether through lobbying, revolving appointments, party funding, and export subsidy, or a focus on export-led production, often involving exports to those countries, as one of the few inquiries into international arms trafficking noted, where "the age-old pressures of human nature, greed, the thirst for power and finally unbridled corruption" (Blom-Cooper 1990, p. 131) continues to adversely affect "open societies, the rule of law and national security" (SIPRI 2011, p. 5).

At the same time, countries with poor governance frameworks often provide permissive environments for domestic and international companies, as well as for the behavior of those domestic companies when they too operate in the international markets. Thus, networks and payments in China have provided a means for domestic and international business "to negotiate the unstable and complicated world of China in the early market reform era." Nevertheless, the traditional *guanxi* practice which "allowed actors to construct social networks and build relationships of trust, which encouraged investment and facilitated the rise of small businesses,. . .also encouraged a conducive environment for rent-seeking, bribery, collusion and exploitation" (Smart and Hsu 2007, p. 186). When the United Nations undertook its major investigation into the oil-for-food scandal, Chapter III of the report provided a detailed analysis of the individual roles of 23 companies that participated in the payment of kickbacks on humanitarian contracts. The inquiries not only showed a refusal by companies (including European as well as Russian and Chinese companies) to cooperate with the inquiry but an unwillingness to acknowledge that the payments were either known to senior management or were, if known, illegal (see Independent Inquiry Committee 2005).

The apparent continuing absence of business to act responsibly in such markets has also been the subject of a number of international organizations and NGOs, covering healthcare, mineral resources, and oil. Such reports often emphasize the consequences. Thus, in relation to natural resources in one country, the 2009 Human Rights Watch report on forestry in Indonesia estimated that the Indonesian government lost $2 billion in 2009 due to illegal logging, corruption, and mismanagement. The total includes forest taxes and royalties never collected on illegally harvested timber, shortfalls due to massive unacknowledged subsidies to the forestry industry (including basing taxes on artificially low market price and exchange rates), and losses from tax evasion by exporters practicing a scam known as "transfer pricing." It went on to note that forestry sector corruption has widespread spillover effects on governance and human rights:

> The individuals responsible for the losses are rarely held accountable by law enforcement and a judiciary deeply corrupted by illegal logging interests, undermining respect for human rights. In addition, the ability of citizens to hold the government accountable is curtailed by a lack of access to public information. And the opportunity costs of the lost revenue are huge: funds desperately needed for essential services that could help Indonesia meet its human rights obligations in areas such as health care, go instead to line the pockets of timber company executives and corrupt officials (2009, p. 39).

Indeed, attempts to reduce carbon emissions through deforestation by international and regional aid agencies have already been identified as likely to cause further corruption place in countries where corruption is pervasive through the illicit purchase of land and carbon rights by developers and logging companies (see UNDP 2010).

Control Context: Legislation, Conventions, and Initiatives

It has been this level of continuing the failure of business (and on occasion the governments of their home countries) to act "responsibly" at an international level which has led to pressure to criminalize such conduct. A number of countries, primarily in the developed world, have taken steps to regulate corruption in their domestic markets; a lesser number also have regulatory agencies to address corporate standards. Nevertheless, in terms of international corruption, multilateral and bilateral agencies have not only sought to manage corporate misuse of their own funds but also, under continuing pressure from international NGOs, such as Transparency International, Global Witness, and so on, have sought to institute more effective control environments. By doing so at international level, however, the intention has been to address domestic legislation in a way that addresses not only bribery abroad but also the question of domestic jurisdictions being responsible for such extraterritorial conduct.

The international measures include the Organisation for Economic Cooperation and Development's (OECD) Anti-Bribery Convention (the OECD 1997 Convention on Combating Bribery of Foreign Public Officials in International Business Transactions), the UN Conventions (including the United Nations' 2003 Convention against Corruption [UNCAC] and, in part, the 2000 UN Convention against Transnational Organized Crime [UNTOC]), and numerous regional and EU-level Conventions (including the 1999 Council of Europe Criminal and Civil Law Conventions on Corruption and the 1995 EU Convention on the protection of the communities' financial interests and the fight against corruption, First and Second Protocol; and the 1997 EU Convention on the fight against corruption involving officials of the European Communities or officials of Member States 1997).

These provide anticorruption frameworks within which to tackle these crimes (although public discourse often refers to corporate "wrongdoing," "misbehavior," and so on, rather than "crime"). Furthermore, monitoring and evaluation of national anticorruption policies takes place in the form of UNCAC peer reviews, the Group of States against Corruption (GRECO), and the OECD's Working Group on Bribery, among others, which highlight inadequacies and subsequently recommend legislative, institutional, practical, and structural improvements.

The three international conventions that significantly influence the control of corporate corruption at the national level are the OECD Convention, UNCAC, and UNTOC:

The OECD Anti-Bribery Convention

The OECD Convention was the first and remains the only legally binding instrument focusing on the *supply side* of bribery. The Convention deals with what is termed "active bribery" in contrast to "passive bribery." This means the focus is on the offense committed by the person who promises or gives a bribe. That said, the Convention does not use the term "active bribery" in order to avoid nontechnical readers mistakenly viewing the briber as taking the initiative and the recipient as a passive victim (OECD Convention 1997, p. 14, paragraph 1): it is often the case that the recipient will have induced or pressured the briber and thus be more "active." The Convention seeks a "functional equivalence" among the measures taken by the parties to sanction bribery of foreign public officials and therefore does not require uniformity or changes in fundamental principles of a party's legal systems (e.g., there is corporate criminal liability in the UK but not Germany). The Convention is highly focused and targeted toward those countries that account for the majority of global exports and foreign investment and introduced the required specific criminalization of the bribery of a foreign public official in international business transactions. It provides a broad and inclusive definition of foreign public official. Senior politicians through managers at state-owned organizations to local police officers would all fall within this definition.

UNCAC

The UNCAC is the first global legally binding instrument in the fight against corruption. It requires the States Parties to implement numerous and detailed anticorruption measures impacting upon their laws, institutions, and practices. The purpose is to aid prevention, detection, and sanctioning of corrupt practices and encourage the cooperation of States Parties. The Convention requires States Parties to establish a range of offenses associated with corruption and attaches particular importance to prevention and the strengthening of international cooperation to combat corruption. It also includes "innovative and far-reaching" provisions on asset recovery and technical assistance and implementation. UNCAC follows an almost identical definition of the term "foreign public official" to that of the OECD Convention but includes also any person holding an "executive" office of a foreign country.

UNTOC

Although UNTOC is the primary international tool for tackling organized crime, it also acknowledges the significant role of corruption in transnational organized crime and therefore requires corruption to be addressed as part of combating organized crime. It focuses only on domestic bribery. UNTOC covers public sector corruption. Of most significance, the Convention requires States Parties to criminalize corruption, with a focus in particular on public sector corruption and bribery of public officials. It further requires effective measures to detect and punish such corruption and significantly requires public authorities to be provided with adequate independence in order to deter the exertion of inappropriate influence of their actions.

Signatories to all such Conventions are required to translate the mandatory components into domestic legislation. All three mention money laundering (although UNCAC makes asset recovery a fundamental principle) and argue for international cooperation in investigating and sanctioning offenders registered in their jurisdiction.

There also a range of corporate-focused initiatives to support corporate anticorruption implementation. Some provide guidance to business, such as the *OECD Guidelines for Multinational Enterprises* intended to provide voluntary principles and standards for responsible business conduct consistent with applicable laws and the TI *Business Principles for Countering Bribery*, or specific guidance, such as ICC *Guidelines on Agents, Intermediaries and Other Third Parties*. Others are more specific and applied. The Global Compact (GC) is a UN initiative to encourage corporate responsibility among private sector companies by asking them to sign up to a set of core values in the areas of human rights, labor standards, the environment, and anticorruption. The World Economic Forum has set up Partnering Against Corruption Initiative (PACI). This engages with chief executive signatories of industry-leading companies in the implementation of a zero-tolerance policy toward corruption and bribery. Other initiatives are directed at specific industry sectors, such as the Global Infrastructure Anti-Corruption Centre, the Medicines Transparency Alliance, and Publish What You Pay; the two major initiatives are the EITI and Collective Action in which the first seeks publication by extractive companies of their payments to host governments and the second provides an anticorruption resource for business by focusing on fighting corruption jointly with other companies and stakeholders, including government.

Control Context: Impact and Implications

Much of the international approaches rely on advocacy and influence; international law and international judicial institutions effectively have no jurisdiction in relation *trans*national and *multi*-jurisdictional corporate crimes. Sovereign states, despite agreeing to ensure their domestic legislation allows for extra-jurisdictional policing, face significant difficulties. Anticorruption authorities, limited by procedural, evidential, legal, and financial obstacles, as well as interstate cooperation, are thought to be easily outflanked by corporations that bribe as part of international business transactions. Enforcement frameworks at the national level largely reflect "command and control" regulatory approaches that exercise influence through enforcing standards supported with criminal sanctioning – albeit corporations rarely face criminal liability, instead negotiating civil settlements with prosecutors and enforcement agencies.

Historically, however, there have been numerous debates between regulators and between academics as to the most appropriate approach to effective regulation. On the one hand, there are those who believe that compliance to the law requires a punitive policing approach incorporating the use of severe sanctions and criminal prosecution. On the other hand, there are those who advocate an approach more in line with the persuasion of businesses to comply with the law. These polemical approaches have been referred to as models of “deterrence” as opposed to models of “compliance” (Reiss 1984) although some analysts have argued it is a question of “when to punish; when to persuade” (Ayres and Braithwaite 1992, p. 21). It is this latter perspective that largely reflects contemporary thinking on regulation (see, e.g., Black and Baldwin 2010). Enforcement practices alone, however, are insufficient, with more preventive mechanisms required to ensure behavior change within corporations and business. It is here where initiatives such as GC and PACI (see above) can influence the broader regulatory landscape.

The final question is whether or not such initiatives – whether focused on criminalization and sanctions or on voluntary agreements – are effective in managing or reducing corporate bribery and corruption. A number of more mainstream criminologists would argue that the undifferentiated assumptions that see all business is potentially criminogenic all of the time “as an inevitable consequence of capitalism” are too sweeping a generalization. They suggest that “it is well to bear in mind some reservations about the idea that capitalism as such is criminogenic. The argument appears to predict too much crime and makes it difficult to explain the relative stability of economic trade within and between nations, given the large number of economic transactions, the many opportunities for committing business crimes, the large gains to be made, and the relative unlikelihood of punishment” (Nelken 2007, p. 747).

Certainly, policing of corporate offenses both domestically and internationally varies significantly, depending on country, context, and purpose – whether the role of the Elf Aquitaine oil company in party finance in France or the 15–20 % of the Iraq Relief and Reconstruction Fund ($3–5 billion) “wasted” by US companies according to the Special Inspector General for Iraq Reconstruction (Tarnoff 2009, p. 20) – and how far governments balance their various interests, such as trade, influence, and oil, with commitment to international standards. In terms of the OECD’s own figures on criminal or administrative sanctions, three countries – Germany, Italy, and the USA – were responsible for nearly two-thirds of the cases (232 from a total of 310 between 1999 and 2010 with 38 countries responding).

This raises questions about the geopolitics of integrity because a pragmatic approach by developed countries rules out large areas of the world, including a number of Caspian, Caucasian, and Middle East countries where, as the *Economist* pointed out, “oil and ethics mix about as well as oil and water” (10 May 2003), those countries whose corruption is overlooked in the continuing pursuit of greater political objectives (whether arms sales, the war on drugs or regional spheres of influence), those countries too big to influence and too powerful to want to listen (such as Russia), or those who are written off as failed states with no strategic or geopolitical value.

Of course, some would argue that the overall developmental agenda that includes the current anticorruption approach is itself shaped by multilateral and key bilateral donors’ perception of the wider developmental trajectory of and objectives for the donor-aided countries and influenced through the uses and conditions attached to their aid and support to achieve this. In terms of both, the preferred direction is toward the “standard liberal democratic practices and norms – representative and responsible government, the rule of law, and the absence of corruption. But it also appears to privilege a neo-liberal faith in the superiority of market economies and in the importance of introducing market mechanisms into the public sector” (see Bevir 2004 p. 2) which benefit developed economies’ businesses as much as, if not more so, developing economies’ businesses. Whether that would resolve the issue of corporate corruption, certainly in export markets, is another question.

Conclusion

This entry set out to discuss corporate corruption in the context of bribery of public and private officials. Corruption and bribery are broad concepts that incorporate a diverse array of activities, actors, and social relations. Endeavors to conceptualize their meaning and extent, their perpetrators and victims, their regulation and control, or even whether they constitute crime, deviance, immoral behavior, or common business practice have been faced with significant theoretical, conceptual, and methodological challenges. The lack of empirical data adds further to these complex phenomena, as do variations in understandings of the problem within different social contexts and jurisdictions.

Distinctions between “active” and “passive” bribery and “public” and “commercial bribery” at the “international” or “domestic” level create a useful analytical framework for understanding how corporations can use corrupt practices in the context of commerce. But the organization of such complex and serious crimes creates significant challenges for control agencies limited by their jurisdictional boundaries and their traditional enforcement approaches. International conventions have attempted to harmonize the problem through their legal frameworks, using content definitions, rather than analytical definitions, that can be operationalized at the national level. Principles of “functional equivalence” that place emphasis on “goals” rather than “means” ensures the fundamental beliefs of nation states’ enforcement systems remain. How sovereigns deal with the problem of corporate corruption can therefore vary greatly. Sanctioning alone, however, may be insufficient, as long-term behavior change within businesses and corporations that use bribery requires a broader regulatory approach incorporating nonenforcement mechanisms such as self-regulation and anticorruption initiatives. Measuring the impact of such prevention-focused approaches, however, is a difficult task in itself.

Related Entries

- ▶ Corporate Crime Decision-Making
- ▶ Corporate Crimes and the Business Cycle
- ▶ Corporate Liability
- ▶ Crimes of Globalization
- ▶ Deterrence of Tax Evasion
- ▶ Naming and Shaming of Corporate Offenders
- ▶ State-Corporate Crime
- ▶ Victims of Corporate Crime
- ▶ Women and White-Collar Crime

Recommended Reading and References

Ayres I, Braithwaite J (1992) Responsive regulation: transcending the deregulation debate. Oxford University Press, New York

Bevir M (2004) Democratic governance, working paper 2004: 5. Institute of Governmental Studies, University of California, Berkeley

Black J, Baldwin R (2010) Really responsive risk-based regulation. Law Policy 32(2):181–213

Blom-Cooper L (1990) Guns for Antigua: report of the commission of inquiry into the circumstances surrounding the shipment of arms from Israel to Antigua and transhipment on 24 April 1989 en route to Colombia. Duckworth, London

Brytting T, Minogue R, Morino V (2011) The anatomy of fraud and corruption. Gower, Farnham

Chibnall S, Saunders P (1977) Worlds apart: notes on the social reality of corruption. Br J Sociol 28:138–153

Clinard MB (1990) Corporate corruption: the abuse of power. Praeger Publishers, New York

Clinard MB, Yeager PC (1980) Corporate crime: the first comprehensive account of illegal practices among America’s top corporations. Collier Macmillan Publishers, London

Control Risks (2009) Business, corruption and economic crime in central and south-east Europe. Control Risks, London

Croall H (2001) Understanding white-collar crime. Open University Press, Buckingham

David-Barrett L (2011) Cabs for hire? Fixing the revolving door between government and business. Transparency International UK, London

Deflem M (1995) Corruption, law, and justice: a conceptual clarification. J Crim Justice 23(3):243–258

Friedrichs DO (2007) Trusted criminals: white collar crime in contemporary society, 3rd edn. Thomson Wadsworth, Belmont

Gottfredson M, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Human Rights Watch (2009) “Wild Money”. The human rights consequences of illegal logging and corruption in Indonesia’s forestry sector. http://www.hrw.org/sites/default/files/reports/indonesia1209webwcover.pdf. Accessed 30 Jan 2013

Independent Inquiry Committee into the United Nations Oil-for-Food Programme (2005) Report on programme manipulation: chapter three – humanitarian goods transactions and illicit payments. http://www.iic-offp.org. Accessed 8 Aug 2011

Klitgaard R (1998) ‘International cooperation against corruption’ finance and development. Available at: http://www.imf.org/external/pubs/ft/fandd/1998/03/pdf/klitgaar.pdf. Accessed 10 Jan 2011

Kramer RC, Michalowski RJ (2006) The original formulation. In: Michalowksi RJ, Kramer RC (eds) State-corporate crime. Rutgers University Press, New Brunswick

Matza D, Sykes G (1961) Juvenile delinquency and subterranean values. Am Sociol Rev 26(5):712–719

Nelken D (2007) White collar and corporate srime. In: Maguire M, Morgan R, Reiner R (eds) (2009) The Oxford handbook of criminology. Oxford University Press, Oxford

Nye JS (1967) Corruption and political development: a cost-benefit analysis. Am Polit Sci Rev 61(2):417–427

OECD (1997) Convention On Combating Bribery Of Foreign Public Officials In International Business Transactions. Available at: http://www.oecd.org/dataoecd/4/18/38028044.pdf

OECD (2008) Corruption. A glossary of international standards in criminal law. OECD, Paris, Accessed at www.oecd.org/dataoecd/59/38/41194428.pdf

OECD (2011) OECD guidelines for multinational enterprises: recommendations for responsible business conduct in a global context. OECD, Paris

Punch M (1996) Dirty business: exploring corporate misconduct. Sage, London

Reiss A (1984) Consequences of compliance and deterrence models of law enforcement for the exercise of police discretion. Law Contemp Probl 47(4):83–122

Schrager LS, Short JF (1977) Towards a sociology of organisational crime. Soc Probl 25:407–419

Shover N, Hochstetler A (2006) Choosing white-collar crime. Cambridge University Press, Cambridge

SIPRI (Stockholm International Peace Research Institute) (2011) Armaments, disarmament and international security. Oxford University Press, Oxford

Slapper G, Tombs S (1999) Corporate crime. Longman, London

Smart A, Hsu CL (2007) Corruption or social capital? Tact and the performance of Guanxi in market socialist China. In: Nuijten M, Anders G (eds) Corruption and the secret of law. Ashgate, Farnham

Steffensmeier D (1989) On the causes of "white collar" crime: an assessment of Hirschi and Gottfredson's claims. Criminology 27(2):345–358

Tarnoff C (2009) Iraq: reconstruction assistance. Congressional Research Service, Washington

UNDP (2010) Staying on track: tackling corruption risks in climate change. UNDP, New York

Williams R (2000) Introduction. In: Williams R (ed) Explaining corruption. Edward Elgar, Cheltenham

World Health Organisation (2009) Medicines: corruption and pharmaceuticals. Fact sheet N°335. Accessed at www.who.int/mediacentre/factsheets/fs335

Transnational Corporate Corruption

▶ Transnational Corporate Bribery

Transnational Exchange of Forensic Evidence

Carole McCartney
School of Law, University of Leeds, Leeds, UK

Synonyms

International evidence exchange; Transnational policing

Overview

Policing and judicial cooperation across international borders is now an expectation. Strategies to combat terrorism, organized and serious crime, as well as domestic crime problems, increasingly incorporate the exchange of forensic information and intelligence between national institutions and agencies. While networks of policing cooperation are not novel, their powers and capabilities have been growing in recent years and bilateral and multi-lateral agreements to exchange data between countries proliferate apace. With increasing demand and capacity for the extra-territorial exchange of evidence under mutual legal assistance, as well as direct and more informal exchanges of data between law enforcement professionals there are increasing numbers of criminal investigations where evidence may have been collected, examined, or interpreted across national borders. Since the Prüm Treaty of 2005, automated exchange of DNA profiles, dactyloscopic data (fingerprints) and vehicle registration data has become mandatory across the EU. These institutional developments highlight broader concerns raised by the exchange of scientific evidence for use in criminal proceedings in a different jurisdiction from that where the evidence originated. As well as presenting technological challenges of effective and secure information exchange, mutual judicial assistance in evidence-gathering poses questions of principle demanding urgent consideration, given that international exchanges of forensic information and intelligence are set to increase exponentially.

Institutional Context: Mutual Judicial Assistance and EU Criminal Policy

The modern escalation in the mobility of goods and people across international borders has not been limited to those legally permitted to travel. While governments encourage goods and individuals to cross borders for economic gain, this comes at the cost of needing to protect national borders against the illegal movement of goods and people. The benefits to be accrued from crossing borders have been capitalized upon by organized criminals, terrorists, and human traffickers. In addition, many ordinary criminals have gained an advantage over law enforcement agencies, avoiding detection by committing offences in different legal jurisdictions. In response, States have likewise had to work across borders in order to combat these illegal activities. As such, international judicial cooperation has become not merely desirable, but essential in those countries in the European Union that share borders. In 1985, the Schengen Treaty paved the way for the creation of a "borderless" Europe and facilitated (some would say necessitated) transnational exchanges of information on European citizens and visitors.

Policing and judicial cooperation across countries is not a new phenomenon. Older informal networks became formalized with the creation of Interpol in 1923. Within (what is now) the EU, the anti-terrorism focused group "TREVI" was created in 1975. The Maastricht Treaty of 1992 subsequently formalized policing and judicial cooperation within the "Third Pillar" of the EU's constitutional framework relating to "Justice and Home Affairs." Europol was then created in 1995, becoming operational in 1999. The Amsterdam Treaty of 1999 created an "Area of Freedom, Security and Justice," strengthening Europol and continuing the trajectory of expanding legal and judicial cooperation across the EU. The Treaty of Amsterdam endorsed: "The principle of mutual recognition which... should become the cornerstone of judicial cooperation... in criminal matters within the Union." This was boosted by the Hague Programme of 2004 which propounded the "Principle of Availability" of law enforcement information. This principle, adopted in November 2004 (Brussels European Council, 4/5 November 2004, Presidency Conclusions), anticipated that: "throughout the Union, a law enforcement officer in one Member State who needs information in order to perform his duties can obtain this from another Member State and that the law enforcement agency in the other Member State which holds this information will make it available for the stated purpose, taking into account the requirements of ongoing investigations in that State." This aspiration was realized across the EU in 2008.

In 2006, the "Swedish Initiative" marked a significant step towards implementing this Principle of Availability, with the adoption of Council Framework Decision 2006/960/JHA on simplifying the exchange of information and intelligence between EU Member States' law enforcement authorities. Adopting a broad and inclusive approach to information exchange, 49 different types of relevant information were identified, including DNA and Fingerprints (CRIMORG 7, 5815/2/05), the underlying principle being that national parties should apply the same criteria to international exchange across the EU as they do to information-sharing between criminal justice professionals at the national level. Transnational exchanges of information could take place through any of the existing channels of mutual judicial cooperation, including Sirene Bureaux (utilizing the Schengen Information System, SIS), Interpol National Contact Bureaux (NCBs), Europol National Units (ENUs) or Europol Liaison Officers and Bilateral Liaison Officers.

Despite these organizational developments and political exhortations to increase criminal justice cooperation, most day-to-day policing remains a national matter. This truism necessitates that policing efforts to detect criminals whose activities cross borders must primarily focus on sharing information and intelligence. Mutual assistance treaties have made provision for information exchange between law enforcement agencies and judicial bodies across the EU Member States. One of the first treaties was the 1959 European Convention on Mutual Assistance in Criminal Matters (20 April 1959, ETS No 30), which came into force on 12 June 1962, and provided for the transfer of information on previous convictions. This treaty set out the rules for the enforcement of letters rogatory issued by a "requesting" State, to procure or communicate evidence for the purposes of criminal proceedings being undertaken in that State from another State. Requirements were specified for requests for mutual assistance, including who was permitted to make a request, the language to be used and under what circumstances mutual assistance could be refused.

The 1999 Amsterdam Treaty and Tampere Programme for policy development and implementation laid down more precise provisions for cooperation between EU police authorities. These instruments boosted inter-governmental co-operation within the EU on policing matters, while endorsing "mutual recognition" as a cornerstone of European integration. In 2001 a detailed programme of measures was set out in the European Convention on Mutual Assistance in Criminal Matters (Official Journal of the European Union 2001, C326), which introduced a suite of practical tools for enhancing cross-border cooperation, starting with the European Arrest Warrant, later adding the European Evidence Warrant, and most recently the European Investigation Order (Mitsilegas 2009). Co-operative efforts have not stopped there. Nijboer (2009, p. 23) predicts that "the next development will be the EU members" mutual recognition, not merely of court decisions, but also of investigative results such as forensic experts' reports'. The Lisbon Treaty, which came into force in 2009, took the final constitutional step of abandoning the "three pillar" structure introduced by the Maastricht Treaty, placing criminal justice policy on an equal footing with other core business at the heart of the EU's activities.

Transnational exchanges of forensic science data and intelligence between EU member states take place within this evolved institutional framework. Resort to forensic science by national intelligence services and by EU and other international security agencies is increasingly frequent. Rapidly evolving strategies to tackle transnational crime foster the exchange of information across borders as well as the linking of databases of different provenance (Lewis 2007). Indeed, international utilization and exchange of forensic intelligence has become a firm expectation of the sector, matched by burgeoning expansion of the mechanisms needed to achieve it. While there have long existed informal professional networks between law enforcement authorities facilitating cooperation across international borders (Loader 2002), modern developments in ICT have produced increased automation of information-sharing and paved the way to multiple-country access to large databases.

An important distinction needs to be drawn at this point in the exposition between the sharing of raw forensic data and exchanges of scientific reports incorporating evaluative expert opinions. Often the two go hand-in-hand, but data and its interpretation do not always coincide and they might not be contributed by the same individuals or institutions. It is far more likely that exchanges of forensic data could be standardized and automated, whereas sharing forensic expert opinions across national borders and legal jurisdictions is a significantly more complex and challenging endeavor. The following discussion will consider each type of information-sharing in turn, drawing out its further implications for principles of mutual recognition and mutual judicial assistance in European criminal proceedings.

Forensic Science Data-Sharing

Direct law enforcement co-operation, or cross-border evidence exchange during investigations and pre-trial processes concerning known crimes or individuals, has been increasing for many years, gaining impetus from the terrorist attacks of 9/11 (Loof 2011). Interpol initially drove the technological phase of international cooperation involving DNA on a global scale. Resolution No. 8 of the Cairo General Assembly in 1998 articulated an aspiration to harmonize DNA profiling and quality assurance systems internationally and recommended that its member countries should create national DNA databases. The Interpol DNA Gateway, created in 2002, enabled its member countries to search unsolved crime-scene profiles and search against DNA profiles from other donor countries using the Interpol Standard Set of Loci (ISSOL). In July 2006, it went online via Interpol's I-24/7 global police communications system. The DNA Gateway does not hold any nominal (demographic) data but acts simply as a conduit between countries who can follow-up any "hits" independently. The DNA Gateway is available for use by all 190 Interpol countries. It can only be used by police forces, however, and this excludes from the system countries (e.g., Belgium and the Netherlands) in which DNA data are classified as judicial information: that is, the information is collected, managed and used by magistrates or prosecutors rather than police agencies. One consequence of these domestic institutional arrangements is that data can be shared internationally only through judicially controlled mechanisms. For this and other reasons, there has consequently been a low submission rate, possibly also attributable to practitioner scepticism that submission will reap much investigative reward (Johnson and Williams 2007, p. 108).

In 1997, the EU adopted Council Resolution 97/C193/02 on the exchange of DNA, recommending the establishment of national DNA databases and the standardization of DNA technology to facilitate information exchange between EU member states. In 2001, Council Resolution 2001/C 187/01 further recommended the limitation of DNA analysis to "non-coding" regions as well as the ongoing standardization of reporting and recording, and the creation in each member state of single national points of contact. This programme was designed to promote harmonization of DNA analysis and the procedures necessary for information exchange and data-sharing. Informal exchange of DNA "in the absence of a supranational authority to govern" was sporadic at this formative stage (Johnson and Williams 2007).

The Prüm Treaty, concerned with increasing the exchange of EU law enforcement data in order to combat terrorism, illegal immigration, and cross-border crime, was signed in 2005. The original signatories were Austria, Germany, the Netherlands, Luxembourg, Belgium, France and Spain, who were subsequently joined by six others. The Treaty was transposed into EU law in 2007 (Decision (2008/615/JHA; 2008/616/JHA), at which point all 27 member states committed themselves to working within the treaty obligations by August 2011. The adoption of the Treaty into the EU *acquis* (legal framework) created, for the first time, a requirement that member states establish an infrastructure to permit the exchange of DNA, fingerprint, and vehicle registration data. This implies the creation of national databases as a necessary precondition to data-sharing, which some member states still do not have. By January 2012, twelve EU countries were actually exchanging DNA information (with ten exchanging fingerprints and vehicle registration data). The Prüm Treaty thus represents a partial step towards the realization of the Hague Programme's aspiration of free-flowing law enforcement information between EU States under the "Principle of Availability." The ultimate objective is "a virtual EU-wide database for law enforcement that makes an entry loaded onto any forensic database in any EU country potentially accessible to every police force in the EU" (Human Genetics Commission 2009, para 5.10).

The Prüm Treaty stipulates that automated searches by other member states are conducted on a "hit"/"no hit" basis. Only after a "hit" has occurred will there be an exchange of personal (demographic) details and case information via existing mutual legal assistance channels. The second step of the exchange process then falls back on national legislation, affording discretion to each country to decide whether or not to exchange demographic data. To satisfy the stipulated standard for information-sharing, DNA profiles must contain at least six of the seven European Standard Set of Loci (ESSL) for known persons or a minimum of six loci (of any type) for crime scene stains. Mixed profiles are excluded. So also, for speculative searching, are profiles derived from crime stains that have already been "matched" to a particular person. A country may make some profiles unavailable for searching (e.g., profiles of criminal justice personnel compiled for elimination purposes). Declaration of a "match" requires at least six fully matching loci (one mismatch is permitted, to accommodate "near" matches; and one base-pair difference is permitted to allow for microvariants that could be artefacts of the profiling process). A report following a Prüm search will include details of: the responding country; the type of "match" (full or near); the characteristics of both matching DNA profiles; the crime stain or reference sample; and the sample code to facilitate follow-up requests for further details via mutual legal assistance mechanisms.

Introduction of the Prüm automatic information exchange system has been credited with notable operational successes. After less than 2 months of a trial period of information-sharing between Austria and Germany, for instance, the German authorities had obtained 1,510 "hits" for DNA profiles from unsolved cases against data held by the Austrian authorities, leading to 710 detections (including, 41 hits in homicide cases, 885 hits in theft investigations, and 85 hits relating to robbery or extortion) (House of Lords European Union Committee 2007, pp. 15–16). Advocates, no less than critics, of the Prüm system recognize that these impressive immediate gains partly represent a one-off windfall. When a country first initiates data exchange, large amounts of information about earlier crimes will become available to the prosecuting authorities for the first time (House of Lords European Union Committee 2007, p. 16). It remains to be seen how effective EU data-sharing will be in terms of producing new "hits" leading to detections and prosecutions on a more sustainable, on-going basis.

The biggest practical limitation of the Prüm information-sharing framework is geographical, being restricted to EU member states (plus a few additional signatory countries such as Norway and Switzerland). There were soon calls for exchanges of DNA data, in particular, to be extended internationally. The UK's presidency of the G8 in 2005 inaugurated a project to identify the requirements for greater international exchange and propose technical solutions. The "DNA Search Request Network" (DNA SRN) was duly established, utilising Interpol's I-24/7 secure email network to send search requests to other G8 countries. Pilot testing began in 2007 with search requests sent to national DNA Bureaux from the USA, Canada and the UK. Japan has since joined the original trio and all four countries have commenced sharing crime scene DNA profiles. The nature and extent of these exchanges has not been made public, however.

Sharing Expert Opinions

Forensic data might consist of nothing more than raw scientific test results. Such information is readily amenable to automatic recording and transmission. For these data to become "evidence," however, a further vital step is required, that of "interpretation." Interpretative conclusions may be offered by qualified scientists and other experienced forensic practitioners, but they are also routinely contributed by non-scientific personnel, including police and prosecutors, in many countries. Within the EU, such data fall under the umbrella term "intelligence." They are transmitted via a range of policing communication channels, including Sirene Bureaux (within the Schengen framework), Europol National Units and Europol Liaison Officers, Interpol National Centre Bureaux, and bilateral mutual assistance arrangements. Such channels are open to the transmission of almost any type of law enforcement data (Gill 2008, p. 143). It is anticipated that the flow of forensic science intelligence will increase within the Schengen system once "SIS II" is brought online, with this express objective.

International policing networks, such as Europol and Interpol, are increasingly focusing efforts on the production and sharing of "forensic intelligence." It is intended that "the accurate, timely and useful product of logically processing forensic case data" (Ribaux et al. 2003) will inform policing tactics, operations and strategy. In 2002, Europol's mandate was extended to cover all serious forms of international crime (including crimes against persons, financial crime and cyber-crime) perpetrated by organised criminals, or where more than one EU Member State is directly affected. Europol Liaison Officers (ELOs), based in The Hague and representing their respective national law enforcement agencies, form bureaux to exchange a range of law enforcement information and intelligence, either on a bilateral or multi-lateral basis. They also share information and intelligence with other competent national authorities, such as customs officials, border control agencies, etc. Each Liaison Bureau contains a Europol National Unit, which acts as a gateway to supply Europol with information and intelligence for its databases. In addition, Eurojust, established in 2002, is tasked with fighting serious crime in the EU by stimulating judicial cooperation and the coordination of investigations and prosecutions which involve more than one Member State. Eurojust aims to improve cooperation by facilitating international mutual legal assistance in order to make cross-border crime investigations more effective. There are also cooperation agreements in place with Norway, Iceland, the USA, Switzerland and Croatia. The organization collates forensic intelligence as part of its mandate as a centre of terrorism-related expertise.

Beyond raw data exchange, it is the further step of interpreting forensic data and producing “intelligence” within the context of a particular investigation that perhaps poses the greatest challenge to effective cross-border information-sharing. Suppose, for example, that DNA profiling produces a database “hit” indicating an association with a crime scene stain from another country. This “hit” then requires further interpretation and investigation, and may prompt further forensic processing. These steps will determine whether and to what extent the declared “match” between two profiles is relevant and sufficiently reliable to inform the course of criminal investigations and, perhaps, to be used later as evidence in criminal proceedings. The standards for this interpretation, who conducts the interpretation, and the amount of further information required will all have an impact upon the “strength” of the evidence and the reliance that can reasonably be placed upon it in criminal investigations or prosecutions. However, there are currently no standardized protocols or mechanisms of international quality assurance governing forensic scientific interpretation or the formulation of expert opinions. Furthermore, there are huge variances across the globe, and even within the EU, in who can offer expert opinions in courts and the stringency of the tests that expert opinion must undergo before being admissible at trial. The degree of critical scrutiny afforded to admissible expert evidence is also highly variable across legal jurisdictions. This degree of variability naturally generates concerns about the quality of scientific evidence being proffered by experts and relied on by courts in different countries. In the context of mutual judicial assistance and intelligence-sharing, it raises the spectre of flawed scientific opinion or misinterpreted forensic data spreading miscarriages of justice across borders.

Key Issues/Controversies

Forensic Data Exchange

Information exchange under the Prüm Treaty typically *precedes* the institution of formal criminal proceedings, since the national authorities will usually have evidence of suspected criminality but, at this stage, no identified suspect. Investigators hope that intelligence sharing will produce an identifiable suspect, e.g., through a matching DNA profile or vehicle registration records. Since information shared through mutual judicial assistance is therefore often pivotal to the progress and outcome of an investigation, ensuring its integrity and reliability is vital to the proper administration of criminal justice. The need for such guarantees is all the greater in relation to evidence widely considered inherently reliable (such as fingerprint and DNA evidence), as domestic courts (and juries) may be in the habit of taking its reliability on trust.

There are rudimentary quality controls in place. For example, there must be compliance with ISO 17025 or equivalent standards for all laboratories processing DNA profiles. No personal data is transferred between member states, via mutual legal assistance channels, until a "hit" has been declared. There must be a commitment to the European Data Protection Directive (or equivalent national law) and also an estimation of the number of adventitious matches expected after a search (Van der Beek 2008). This "two-step" approach is intended to provide an opportunity to assess domestic legislation and gauge whether the exchange falls within legal parameters and whether data protection measures are in place prior to the exchange of personal data. It also provides an opportunity for a "near match" to undergo further testing by the country reporting the "hit" prior to the recipient country relying upon the "hit" to commence investigations.

Widespread distribution of automated database searching could potentially generate candidate match overload. Such problems were anticipated by the European Network of Forensic Science Institutes (ENFSI) DNA Working Group in 2009 (ENFSI 2009). The Working Group concluded that while the current ESS loci are capable of handling occasional exchanges of DNA-profiles between countries, the much higher volume of profile-sharing and searching made possible by the Prüm Treaty means that the chance of adventitious or false matches may significantly increase. An adventitious match occurs when the DNA profiles of two individuals (who are not identical twins) genuinely match (theoretically a rare, but not impossible, occurrence with the latest, highly discriminating, DNA profiling systems). A "false match" is a match declared in error, where closer examination shows the result to be invalid (e.g., owing to laboratory contamination or interpretational errors). Investigators will often be able to eliminate adventitious or false matches when it becomes apparent that the individual concerned had nothing to do with the crime. However, those wrongly suspected may have suffered considerable intrusion into their private lives before the confusion was resolved. The reverse situation – a false elimination – may be almost equally damaging for the administration of justice. With increasing horizontal integration of criminal justice data, international liaison officers could be overwhelmed by information and in major investigations may experience considerable pressure to confirm "near matches" as reliable leads. As Schneider (2009, p. 1) explains: "When massive exchanges of DNA profiles are undertaken. . . the seven ESS loci will not be sufficient because the chance of adventitious matches will no longer be negligible." Both Germany and the Netherlands have reported having to undertake extensive re-testing when matches are identified, to exclude adventitious and false matches. The Netherlands now refuses to release demographic data for "near matches" without the assurance of confirmatory testing (Van der Beek 2008).

The Interpol International Gateway was a relatively low-cost initiative for information searching. The political momentum driving Prüm clearly envisages much greater investment to facilitate multiple database searching and the generation of new data, as well more effective exchange of existing data. Germany claims that it cost less than €1 million to set up its national Prüm data exchange system. However, the UK government estimate a start-up figure of some £31 million (including £2.5 million first-year operating expenditure) (House of Lords European Union Committee 2007, pp. 23–24). This does not include the additional costs of tracing suspects, their arrest, detention and deportation, following "hits" identifying foreign nationals or fugitives residing overseas. In 2009–2010 UK authorities effected the arrest, detention and extradition of 1,032 people under the European Arrest Warrant framework, prompting some commentators to question why UK taxpayers should foot the £25 million annual bill for sending individuals abroad for trial The UK police process 20 times as many arrest warrants as they submit (Slack 2011).

Even greater obstacles confront fingerprint data exchange, since there is no accepted international standard for declaring valid matches. Transmission of raw data and the preparation or presentation of information for analysis, theoretically at least, presents fewer technical problems. However, use of the best quality images is crucial for fingerprint comparison. An international standard for minimum pixel quality would be desirable, for example, but little progress has yet been made on devising any international standards for transmission of fingerprint data, beyond the specification of an agreed file format. Meanwhile, proprietary search and image capture systems are being developed with little regard for interoperability.

Responding to threats of "match overload" and excessive strain on examiners required to analyze thousands of prints, the EU has placed limits on the number of searches that can be conducted of fingerprint databases (Document 5860/5/10 REV 5 JAI 92 CRIMORG 16 ENFOPOL 29). Dror and Mnookin (2010) argue that the scale and speed of automated database searching inevitably increases the risk of erroneous identifications by multiplying the number of candidate matches based on incidental similarities, some of which may be artifacts "both of the relative similarity of the patterns being compared and of the human cognitive architecture involved in pattern matching" (Dror and Mnookin 2010, p. 56). This parallels the problem of DNA "near matches," and may similarly require significant investment of resources to re-check and eliminate specious "hits." Reliance must be placed on national authorities to be diligent in undertaking the necessary re-testing to confirm initial results, before treating shared fingerprint "matches" as a reliable basis on which to initiate criminal prosecutions.

Laboratories developing imperceptible latent prints from crime scenes must satisfy the rudimentary quality controls required for accreditation under ISO 17025. However, police fingerprint bureaux are exempted, placing Automatic Fingerprint Identification Systems (AFIS) outside this quality framework. The UK Association of Chief Police Officers (ACPO) has declared its aspiration to ensure that all fingerprint bureaux are accredited to ISO 17025 by 2020. In the meantime, fingerprinting, like DNA profiling, is widely perceived as reliable and routine. This may foster the impression that regulatory frameworks are equally robust and unproblematic, reinforcing implicit faith in these types of evidence. However, the regulation and governance of both fingerprint and DNA data in England and Wales, and internationally, have attracted critical scrutiny in recent years (see e.g., Nuffield Council on Bioethics (2007); Human Genetics Commission (2009); House of Commons Home Affairs Committee (2010); McCartney et al. (2010)). More effective regulation of the forensic sciences in general has been demanded for many years. Regulatory initiatives in England and Wales were recently expanded EU-wide under the Council Framework Decision 2009/905/JHA on the accreditation of forensic service providers carrying out laboratory activities (EU Council 2009a).

Exchanging Expert Opinions

Scientific data are virtually meaningless unless interpreted within the context of other information. As we have seen, this interpretation can be undertaken by forensic science practitioners, police or prosecuting authorities, often in a different country from where forensic data were originally generated. It is therefore important to ensure the integrity of systems of accreditation of forensic experts in various countries, and to set appropriate international standards for interpreting forensic data. While forensic laboratories may be accredited to ISO 17025, this standard does not guarantee the quality of individuals' case-work nor does it govern activities outside the laboratory (e.g., crime scene investigations). Accreditation of individual experts remains a major unresolved issue, which has been addressed in a variety of ways, and with varying success, by different countries. The Council for the Registration of Forensic Practitioners (CRFP) was established in the UK in 1999 to accredit individual experts for forensic work, but this encountered various practical problems and was disbanded in 2007, largely on economic grounds. Subsequently, the Netherlands Register of Court Experts (NRGD) was set up in 2009 to provide Dutch courts with a register of scientific experts whose professional quality can be relied upon (see http://english.nrgd.nl/home/). However, these are exceptional efforts in international terms. Few other countries have introduced similar mechanisms or systems to gauge the professional competence of scientific experts. There is consequently significant variation in the use of forensic expertise in criminal proceedings, much depending on individual trial judges' discretion.

Yet judicial evaluations are often largely based upon an expert's formal professional qualifications or training, a poor substitute for a thorough examination of the competence of individual experts and of the reliability of their methodology, data and conclusions. Rarer still are sanctions for an incompetent expert or one who provides flawed testimony in court. In the absence of international standards for interpreting forensic data, it is almost impossible to harmonise the format or substance of expert reports.

Future Directions

The Human Genetics Commission (2009, p. 84) noted the growing pressures for sharing information:

> [T]here are . . . substantial pressures to create the conditions for the horizontal integration of criminal justice databases, and the sharing of information between national jurisdictions. Whereas the exercise of law enforcement powers of European Union (EU) member states are confined within national borders, for the transfer of information relevant to law enforcement these borders are increasingly dissolving. We identify three developments in the 'third pillar' of EU policy (police and judicial co-operation in criminal matters) that have contributed to the cross-border linking of forensic bioinformation: the establishment of a system of co-operation based on a central information system making use of electronic databases; the creation of a presumption and then an obligation in favour of data sharing; and the commitment of the UK and other member states to align with such a system.

The UK Government previously explained how the Prüm Treaty would "speed up and improve the quality and quantity of information exchanged" about DNA, fingerprints and vehicle registration in order to identify and bring to justice terrorists and criminals. Ministers stated that the Government had negotiated hard "to get an outcome that enables [the UK] to sign up to Prüm and get all the benefits in terms of fighting cross-border crime and counter terrorism where so much depends on good data exchange and intelligence led policing" (House of Lords European Union Committee 2007, Ev.1-2 and Q8). In reply, parliamentarians and ethicists emphasized that "privacy-related issues concerning the use and transfer of DNA and other data for inter-jurisdictional criminal matters must be considered and agreed in parallel with arrangements for availability, exchange and linkage" (Nuffield Council on Bioethics 2007, p. 105). The UK has yet to announce when it expect to be taking part in information-sharing under the Prüm Treaty.

For the time being, mutual judicial cooperation and data exchange in relation to DNA and fingerprints remains on a fairly modest scale, compared with efforts to detect criminals travelling with stolen or lost travel documents. At the end of 2011 the Interpol Stolen and Lost Travel Documents (SLTD) database held 31 million records, accessed by 161 countries (see http://www.interpol.int/@en/INTERPOL-expertise/Databases). The database can be made interoperable with national border control data systems. In 2008 law enforcement and border control officers carried out 25 million international searches of the database. This resulted in the identification of more than 5,000 individuals travelling on fraudulent documents (McCartney et al. 2010, p. 44).

In 2011 the Polish EU presidency called for a "Vision for European Forensic Science 2020." Building on increasing police cooperation and in recognition of the demands of maintaining and developing the EU as an area of "freedom, security and justice," the EU is working to create a "European Forensic Science Area" by 2020. This policy programme aims to "ensure the even-handed, consistent and efficient administration of justice and the security of citizens," by accrediting forensic institutes and laboratories and conducting international proficiency testing, establishing minimum competence criteria for forensic practitioners and introducing best practice manuals for laboratory case-work. There will also be minimum quality standards for examining crime scenes and managing the production of scientific evidence, from the crime scene to the courtroom. Duplication of effort in different countries should be minimized through mutual recognition of testing and processing of forensic data, and the sharing of forensic databases.

These objectives are laudable and, if implemented successfully, would go some way towards overcoming current obstacles to effective forensic data exchange and intelligence sharing. As things stand, enormous differences in national modes of data production and intelligence generation, coupled with markedly discrepant regulatory regimes, conduces to an extremely complex environment in which to attempt to foster mutual judicial cooperation (Gill 2008, p. 132). These complexities are overlaid by major differences in national legal traditions and judicial systems, which are far from being comprehensively harmonized (even supposing this to be desirable). For example, in almost all of the initiatives, mechanisms and organizations which seek to build collaborative links and facilitate forensic data and intelligence sharing, the emphasis has been squarely on policing and prosecutorial cooperation. There is almost a complete lack of attention to defence rights and interests. Procedural safeguards for defendants in criminal proceedings and equivalent access to evidence for the defence have largely been ignored in the drive to re-arm investigators and prosecutors in the fight against cross-border crime. These deficits are likely to become more salient in the coming months and years, as the provision of adequate procedural rights for the defence rights rises up the EU criminal policy agenda (see, in particular, EU Council 2009b).

Conclusions

All forms of human cooperation and collaboration require mutual trust. The EU's 2009 Stockholm Programme, which promised to deliver "An Open and Secure Europe Serving and Protecting Citizens," stressed that mutual trust was "the basis for efficient cooperation in this area. Ensuring trust and finding new ways to increase reliance on, and mutual understanding between, the different legal systems in the Member States will thus be one of the main challenges for the future" (EU Council 2010, para.1.2.1). However, significant obstacles to effective cross-border policing cooperation have yet to be overcome. Immediate practical difficulties such as language differences, time limits and technical legal constraints (e.g., the doctrine of double criminality in extradition law) seem to preoccupy policymakers, at the expense of more fundamental considerations of proportionality, necessity, and public acceptability. The Stockholm Programme did acknowledge that "[i]n order to improve cooperation based on mutual recognition, some matters of principle should also be resolved" (EU Council 2010, para 3.1.1.). Yet "principled" considerations have largely been absent from much subsequent debate.

Pressing issues demand resolution prior to further expansion of international exchanges of forensic data, information, and intelligence. The existing fragmentary regulatory framework must be developed to set international quality standards to ensure the competency of individual forensic practitioners and the scientific validity and reliability of their results. Agreed protocols for data interpretation and report writing need to be adopted, and issues of training and resources addressed. At a more fundamental level of principle, urgent consideration must be given to implementing procedural safeguards and equivalent powers for the defence, to ensure that transnational data and intelligence sharing does not undermine the accused's fair trial rights. Most "cooperation" instruments are silent on these matters. Robust rules need to be in place to ensure that evidence exchanged across borders is relevant, reliable and not obtained through illegal or unfair means. The exchange of evidence between national authorities should not provide an excuse to drive down procedural standards to the lowest common denominator, nor to enable prosecutors to gain access to evidence that would have been legally unavailable to them in their own jurisdictions. To the contrary, mechanisms of mutual judicial assistance should be used to establish and promote best international practice in evidence gathering, production and presentation.

Perhaps most important is the oversight of these transnational flows of law enforcement information. The process lacks transparency, and consequently citizens' ability to challenge exchanges of their personal data is almost nonexistent. Concerns raised by den Boer (2002) that policing networks sit outside political frameworks of democratic governance and accountability remain apposite. Simply ensuring that the system "works" for the law enforcement professionals involved is inadequate for maintaining public confidence. Transparency and accountability are key to the long-term acceptability of expanded networks of data and intelligence sharing across national borders and mutual recognition of judicial evidence. As Professor Spencer explained to a House of Lords Select Committee: "What is done in trans-border cases has to be acceptable to public opinion, not just prosecutors and people who work the system. If there are dysfunctions in the criminal justice systems of some other Member States. . . they are not likely to be sorted out just by people getting to know each other. . .Unless these matters are addressed, public opinion will not accept the too ready functioning of cross-border criminal justice" (House of Lords European Union Committee 2012, p. 11).

Related Entries

- ▶ Behavioral Science Evidence in Criminal Trials
- ▶ Comparative Legal Cultures
- ▶ DNA Profiling
- ▶ Expert Evidence and Criminal Trial Procedure
- ▶ Expert Witnesses: Role, Ethics, and Accountability
- ▶ Forensic Science and Criminal Inquiry
- ▶ Forensic Science and Miscarriages of Justice
- ▶ Forensic Science and the Paradigm of Quality
- ▶ Scientific Evidence Before International Criminal Tribunals
- ▶ Scientific Evidence in Criminal Adjudication
- ▶ Scientific Evidence in Criminal Prosecutions

Recommended Reading and References

Den Boer M (2002) Towards an accountability regime for an emerging European policing governance. Polic Soc 12(4):275–289

Dror IE, Mnookin JL (2010) The use of technology in human expert domains: challenges and risks arising from the use of automated fingerprint identification systems in forensic science. Law Probab Risk 9(1):47–67

ENSFI (2009) DNA database management review. http://www.dnaresource.com/documents/ENFSIdocumentonDNAdatabasemanagement20092.pdf

EU Council (2009a) Accreditation of forensic service providers carrying out laboratory activities. Council framework decision (2009/905/JHA) OJ EN L 322/1

EU Council (2009b) Roadmap for strengthening procedural rights of suspected or accused persons in criminal proceedings, Maastricht. (2009/C 295/01), OJ C 295, 4 Dec 2009

EU Council (2010) The Stockholm programme (2010/C 115/1) OJ C 115, 4 May 2010

Gill R (2008) Study of the obstacles to cooperation and information-sharing between forensic science laboratories and other relevant bodies of different member states and between the latter and counterparts in third countries (JLS/D1/2007/025)

House of Commons Home Affairs Committee (2010) The National DNA database: eighth report of session 2009–2010, vol I report. The Stationery Office Limited, London

House of Lords European Union Committee (2007) 18th report of session 2006–2007. Prüm: an effective weapon against terrorism and crime? HL paper 90. The Stationery Office Limited, London

House of Lords European Union Committee (2012) 30th report of session 2010–2012. The European Union's policy on criminal procedure, HL paper 288. The Stationary Office Limited, London

Human Genetics Commission (2009) Nothing to hide, nothing to fear. Department of Health, London

Johnson P, Williams R (2007) Internationalizing new technologies of crime control: forensic DNA databasing and datasharing in the European Union. Polic Soc 17(2):103–118

Joubert C, Bevers H (2000) International exchange of police data and forensic expertise. In: Nijboer JF, Sprangers WJJM (eds) Harmonisation in forensic expertise: an inquiry into the desirability of and opportunities for international standards. Thelma Thesis, Amsterdam

Law Society (2009) Safeguarding the use of expert evidence in the European Union (JLS/2006/AGIS/058). http://www.ecba.org/extdocserv/projects/expertwit/SafeguardingExpertEvidenceEUProjectRep2009.pdf

Lewis C (2007) International structures and trans-national crime. In: Newburn T, Wiliamson T, Wright A (eds) Handbook of criminal investigation. Willan, Cullompton, pp 175–198

Loader I (2002) Policing, securitization and democratization in Europe. Crim Justice 2(2):125–153

Loof R (2011) Obtaining, adducing and contesting evidence from abroad: a defence perspective on cross-border evidence. Crim Law Rev 1(1):40–57

McCartney C, Williams R, Wilson T (2010) The future of forensic bioinformation, The Nuffield Foundation. http://www.nuffieldfoundation.org/future-forensic-bioinformation

Mitsilegas V (2009) The third wave of third pillar law: which direction for EU criminal justice? Eur Law Rev 34(4):523–560

Nijboer JF (2009) Current issues in evidence and procedure – comparative comments from a continental perspective. Int Comment Evidence 6(2). doi:10.2202/1554-4567.1095

Nuffield Council on Bioethics (2007) The forensic use of bioinformation: ethical issues, Cambridge. http://www.nuffieldbioethics.org/bioinformation

Ribaux O, Girod A, Walsh SJ, Margot P, Mizrahi S, Clivaz V (2003) Forensic intelligence and crime analysis. Law Probab Risk 2(1):47–60

Schneider P (2009) Expansion of the European standard set of DNA database loci – the current situation, Profiles in DNA, MarchSlack J (2011) Police hit by flood of EU arrest warrants as officers process 20 times the number of foreign extradition requests as they make. Daily Mail, 1 Feb 2011. http://www.dailymail.co.uk/news/article-1352337/British-police-hit-20-times-number-EU-arrest-warrants-make.html#ixzz26k7B02V9

Slack J (2011) Police hit by flood of EU arrest warrants as officers process 20 times the number of foreign extradition requests as they make. Retrieved 1 February 2011 from http://www.dailymail.co.uk/news/article-1352337/British-police-hit-20-times-number-EU-arrest-warrants-make.html#ixzz2T94Jt4qO

Van der Beek C (2008) Exchange of DNA profiles by the Treaty of Prüm. Presented at DNA data exchange in Europe conference, Maastricht, 5–6 June 2008. http://www.dna-conference.eu/ppt/Van%20der%20Beek.pdf

Transnational Policing

► Transnational Exchange of Forensic Evidence

Treating Mentally Ill Offenders

William B. Lawson
Department of Psychiatry and Behavioral Sciences, Howard University Health Sciences, Washington, DC, USA

Overview

Prisons and jails now include a substantial number of individuals who are mentally ill or substance abusers. This entry reviews how offenders with mental health and substance abuse disorders are handled for their behavioral health disorders. More importantly, the overrepresentation of minorities in the correctional system presents challenges to diagnoses and treatments offered, given the disparities in health care in the community. The importance of medications, and the continuation after release from prison or jail, is emphasized as a necessary tool to safely manage and assist individuals with behavioral health disorders. The failure to provide adequate diagnoses and treatment in confinement facilities, as well as in the community, has implications for efforts to reduce the demand for incarceration.

Introduction

The United States has shown a substantial increase in its correctional population, with much of it due to the “war on drugs” increasing the incarceration of the mentally ill and substance abusers. This special population often requires unavailable treatment. Moreover, minority and poverty-stricken populations make up the majority of the correctional population and face the same disparities in access to care as these individuals see in the noncorrectional population. Diagnostic concerns, lack of access to psychotherapy, and pharmacotherapy are quite prevalent. Nevertheless, treatment availability can improve the correctional environment and has positive consequences for recidivism and the larger society. Moreover, substance abuse treatment can reduce the spread of diseases such as AIDS. Approaches such as drug courts and integrated programs have helped to improve access to treatment and maintain adherence to care in the community.

Prisons and jails now include a substantial number of individuals who are mentally ill or substance abusers. Estimates of the inmate population that is severely mentally ill range from 15 % to 30 % (Fazel and Seewald 2012). Fifty percent to 65 % of state and federal prisoners and jail inmates have a diagnosis of mental illness (James and Glaze 2006). Jails and prisons are now the largest provider of mental health services in the country, serving more mentally ill than any public mental health hospital or clinic. Half of the correctional population meets standard diagnostic criteria for drug dependence or abuse (Mumola and Karberg 2006).

Unfortunately, individuals in the correctional system are often underdiagnosed and undertreated. Once incarcerated, inmates have little chance of accessing mental health services with fewer than 17 % actually receiving services. Moreover, individuals are under-referred for mental health services and correctional systems lack the professional staff to address mental disorders, such as providing evidence-based psychotherapy (Foulks 2004). The culture of corrections further complicates the problem since the need for security and control supersedes the need for treatment. Nevertheless, the need for treatment services is well documented. Consequently, many correctional settings offer treatments of various types or at least make an attempt to manage those mentally ill individuals or substance abusers who are clearly behaviorally disturbed or suffering from intoxication or withdrawal.

Treating Mentally Ill Offenders, Table 1 Comparison of the correctional and mental health systems

Correctional system	Mental health system	Consequence
Focus is on security	Focus is on treatment	Treatment needs often have to be sacrificed for security
Abnormal behavior considered to be manipulative	Abnormal behavior considered to be an illness	Dealing with willful disobedience is a priority over care
Management of overt behavior considered important	Diagnosis considered relevant	Diagnosis or the need for diagnosis ignored or minimized
Hierarchy relationships are maintained for order	Seeking therapeutic alliance	Lack of trust and need for security overrides therapeutic considerations
Medication needed for control	Medication needed for treatment	Therapeutic benefits have less priority than the need for order

Barriers to Treatment and Services

The correctional system as an institution has a different perspective on the role of treatment, as shown in Table 1.

The correctional system is a punitive system with a focus on security. In such a setting, a therapeutic alliance, while possible, is sacrificed to the need for order; trust is secondary. Interventions necessarily are focused on control rather than therapeutic improvement and treatment needs are sacrificed for security.

There are also key staffing issues. Jails and prisons face multiple challenges in providing inmates with adequate care because of several reasons. One such challenge is the lack of knowledge of treatment histories or medications. Second, jail and prison officials do not have access to medical records. Finally, there is an inability to communicate with providers because of confidentiality restrictions.

Demographics are also important. The incarcerated are disproportionately poor and minority. African-Americans are disproportionately represented in the criminal justice system – 10 % of African-American men between ages 18 and 34 were in prison in 2005 (three times the rate of Hispanic men and seven times that of White men). African-Americans and Latinos make up over half the prison population, which is a far greater representation than in the larger society. For example, African-American juveniles with conduct problems are sent to youth detention while their White counterparts receive treatment. In addition, African-Americans arrested for possession are more likely to be incarcerated than their White counterparts (Iguchi et al. 2005).

Disparities in services for the poor and minorities in the general population are well documented (Kessler et al. 2005; Primm and Lawson 2010). African-Americans with depression are only half as likely to get treatment services even when they are appropriately diagnosed. Incarceration can contribute to these disparities as inmates are less likely to get treatment, or have difficulty linking to services after release due to structural issues related to incarceration (Springer et al. 2012). Outside of the correctional system, less than a quarter of African-Americans receive evidence-based treatment for most mental disorders in any setting.

Consequences of Lack of Treatment

The need for treatment is also related to the fact that most incarcerated individuals are ultimately released. The failure to treat mental illness and substance abuse contributes to recidivism, trapping individuals in an endless cycle of mental illness and incarceration. This cycle appears particularly prominent among African-Americans (Primm and Lawson 2010). African-Americans are more likely to be underserved in correctional settings (Hartwell 2001) and due to their overrepresentation, their release greatly affects the African-American community that often is one with limited resources and high rates of poverty.

Moreover, prisons have become vectors for chronic diseases such as AIDS, Hepatitis C, and treatment-resistant TB (Springer et al. 2012). Recently released inmates have been found to be infected with HIV and continue to be involved in risky behavior such as high-risk sex or drug abuse. These disorders can be spread to the general population through high-risk behavior if the individuals are not treated for both the underlying mental illness or substance abuse and the chronic illness.

Mental disorders can be associated with violence when untreated. Yet when treated, individuals with mental disorders are no more violent than the general population. Substance abuse can also be associated with continued criminal behavior. Treatment of substance abuse reduces the risk of such behavior in the community and reduces crimes rates. Lack of access to treatment therefore affects both the incarcerated and general community.

The First Step: Accurate Diagnosis

Misdiagnosis and underdiagnosis is common in correctional settings. Moreover, the poor and ethnic minorities are probably more likely to be underdiagnosed or misdiagnosed, given the findings from the non-correctional community. Outside of the correctional system, clinicians tend to grossly overdiagnose ethnic minorities with relatively uncommon disorders such as schizophrenia at the expense of much more familiar ones, such as mood and anxiety disorders (Strakowski et al. 2003). This underrecognition of some mental disorders and excessive misdiagnosis of others contributes to treatment disparities among African-Americans and Latinos. Such disparities also contribute to the failure of recognizing these disorders in the prison system. Mentally ill, underrepresented minorities are often not recognized as having a disorder and are instead arrested, which further contributes to the increase in mental illness within the prison population (Foulks 2004).

Anxiety Disorders

Anxiety disorders are less likely to be observed in the community for African-Americans diagnosed by clinicians without the use of assessment instruments, which are related to misdiagnosis and underrecognition of symptoms. However, anxiety disorders such as generalized anxiety disorder or posttraumatic stress disorder (PTSD) are among the most common mental disorders and may show a higher prevalence among those in the correctional system because of the correctional setting itself. Prisoners can and do exploit each other, and the problem of violence and victimization within the correctional system is well documented. Mentally ill prisoners are far more likely to be victimized and exploited (Crisanti and Frueh 2011).

Many individuals enter the correctional setting with preexisting lifestyle factors that are similar in the inner city. Individuals are often exposed to violence if not victimized themselves, and PTSD is widespread in inner cities where minorities and the poor are over-represented (Alim et al. 2006). PTSD is both underdiagnosed or misdiagnosed, as those with the disorder tend to be hypervigilant, and show autonomic hyperactivity and emotional numbness, which could be interpreted as psychotic or antisocial (Lawson 2009). African-Americans are generally at risk of being perceived as guilty or troublemakers, and those with suspicious behavior, such as the hypervigilance and emotional numbness associated with PTSD, are often perceived as being criminals (Hicks 2004). The result is delayed or no treatment.

Depression

Depression is among the most common disorders in the correctional system (Baillargeon et al. 2000). Although depression is often treated in the correctional system, there is a large unmet need as suicide rates are higher in the prison system than in the community (Rabe 2012). Failure to recognize depression may arise from a lack of awareness of African-American and Latino culture (Primm and Lawson 2010). Moreover, low income individuals and many ethnic minorities, especially males, do not report sadness and often will focus on the vegetative or physical signs of depression, including sleep or eating disturbance, and somatic complaints. The problem is further enhanced by the failure of many inmates to recognize depression. Gang members may show their depression by "acting out" and commit antisocial or violent acts.

Suicide is an important consequence of depression. Suicide and suicide attempts have been recognized as a worldwide problem in correctional systems. It is clearly more common than in the general community, perhaps due to the high rates of victimization as well as unrecognized psychopathology (Crisanti and Frueh 2011). Early recognition and treatment of depression is one essential approach to reducing the suicide risk.

In the past, suicide was thought to be rare among ethnic minorities. However, recent studies have shown young African-American males do not differ in rates from their White counterparts, while Native Americans have among the highest rates in young men (Joe et al. 2006).

Bipolar Disorder

Misdiagnosis of bipolar disorder is common, and these individuals may instead be diagnosed for years with schizophrenia and/or depression (Lawson 2005). African-Americans are much more likely to be under- or misdiagnosed and often receive a diagnosis of schizophrenia instead of bipolar disorder. It is also under-recognized in the correctional system with a third of inmates having unrecognized bipolar disorder (Kemp et al. 2008).

Recognizing and treating bipolar disorder is extremely important in the correctional system. The symptoms of the disorder involve impulsivity, pleasurable high-risk behavior, and irritability. Impulsive violent behavior, drug use, overspending, sexual assault, gambling, and disregard for the rights of others are all common in a manic episode and make these individuals especially vulnerable to arrest (Lawson 2005). The increased underdiagnosis among African-Americans makes the risk for arrest far greater than the likelihood for treatment. Secondly, bipolar disorder has been found to be the most single important factor in recidivism in jails (Quanbeck et al. 2005).

Schizophrenia

Individuals too impaired to make it in society may result to criminal activity to survive, resulting in the severely mentally ill ending up incarcerated. Many are regarded as criminals or psychopaths, and often the first time, a mental illness is diagnosed in jail or prison. Symptoms, such as impulsivity, poor judgment, callousness, a blunted affect, command hallucinations, and delusions, may be interpreted as antisocial or criminal especially for minorities. Once in the correctional system, individuals with schizophrenia are at great risk for being victimized. Upon release, transition to the outside world is especially difficult, leading to readmissions and recidivism especially when the diagnosis is not made. In addition, African-Americans are at increased risk of being perceived as hostile, further reducing the likelihood of treatment (Lawson 2007).

Preventing Misdiagnosis

Prison officials must be made aware of the risks associated with a misdiagnosis. A missed diagnosis increases the inmates' risk of exploitation and suicide, and the correctional staff's ability to prevent risks to themselves and other prisoners. Secondly, cultural awareness is important. Stereotypes about the rates of illness based on race must be discarded. An appreciation of cultural issues can increase diagnostic accuracy (Lawson 2005, 2007).

Language issues must be addressed as well. Many Latinos speak English as a second language. Outside of the prison setting, patients who speak English as a second language are more likely to be misdiagnosed as having schizophrenia when they have depression and may report different symptoms in their primary language. Limited English proficient individuals may be misdiagnosed, ignored, victimized, lack adherence to treatment, or may not receive appropriate and necessary treatment in the correctional system (Foulks 2004).

The use of screening tools can also be effective. Screening tools used with clinical judgment can lead to a greater likelihood of a valid diagnosis. Screening has been found to be effective in primary care settings which have successfully used various instruments to identify mental health illness (Gilbod et al. 2002). These instruments can be very brief and self-administered and often have good validity and reliability, although not all have been standardized in ethnic minority populations.

Non-pharmacological Management

Psychotherapy can be effective in the correctional setting. Some contend that the practice of psychotherapy in prison is an unethical endeavor due to a lack of trust or confidentiality between the therapist and client and the possible threat of coercion. However, if these issues are addressed directly, a therapeutic relationship can be established. In fact, when therapy was provided, mental health providers offered a full range of services with goals that include mental illness recovery, emotions management, institutional functioning, reentry, risk-need, and personal growth, with the main focus being mental health recovery (Bewley and Morgan 2011). Moreover, many systems offer special areas of treatment that are key in the correctional system including: (1) management of self-mutilation, posttraumatic reactions (e.g., prison rape survivors), terminal illness, and crisis intervention, (2) maintenance of psychotic and major affective disorders, (3) outpatient therapy on adjustment issues and elective psychotherapy, and (4) treatment Programs for sex offender therapy and chemical dependency.

Moreover, various therapeutic approaches have been found to be effective. Meditation-based programs including Transcendental Meditation, mindfulness-based stress reduction, and 10-day Vipassana retreats have been found to enhance psychological well-being, decrease substance use, and decrease recidivism (Himelstein 2011).

Dialectical behavioral therapy has been used effectively as well. Various cognitive behavioral therapy approaches and interpersonal therapy have been shown to be effective in depression and anxiety disorders. Psychosocial interventions such as cognitive behavioral therapy (CBT), family interventions, and education programs have been found to be effective in maximum security settings (Walker 2004). Modified therapeutic communities modified for the correctional setting have been shown to be effective.

Pharmacotherapy

Despite the strong evidence for the effectiveness of psychosocial interventions, medication treatment is the primary mode of treatment within the correctional system (Thornburg 1995). Part of the reason may be the absence of personnel trained in psychosocial interventions. Another reason may be due to ethnic minorities who outside of the correctional system are thought to be poor psychotherapy candidates (Primm and Lawson 2010), leading to a decrease in the utilization of different types of psychotherapy.

Pharmacotherapy

One of the most prescribed medication groups are benzodiazepines which are safe, effective, and have less abuse potential than their older counterparts for anxiety and sedation. In the correctional system, benzodiazepines are used to address behavioral problems and a wide range of psychiatric disorders (Lerat et al. 2010). However, their overuse could lead to problems in the correctional system. The medications tend to contribute to disinhibition and tolerance is quickly acquired, making their use for many behavioral and psychiatric problematic. In an older study, violent or aggressive incidents occurred significantly more frequently in inmates on psychotropic medication than inmates who were not on psychotropic drugs. Of these, antianxiety agents (diazepam in 81 % of the cases) appeared to be the most implicated, with 3.6 times as many acts of aggression occurring when inmates were on these drugs (Workman and Cunningham 1975). Apparently the benefits were outweighed by the disinhibiting effects on behavior.

While these agents are less likely to be abused than most street drugs and older sedative/hypnotics, overuse and dependence often occur. To maintain control and order, hypnotics and sedatives such as benzodiazepines are often prescribed and their use is widespread among prisoners. Both illegal and medical drug use can occur in prison. Moreover, in this population in the jail/prison setting, drug-seeking behavior is frequently encountered. Prescribers usually take into account the potential for abuse before prescribing pain medications, benzodiazepines, and other hypnotic/sedative drugs, and they will reduce their use with guidelines and education. Nevertheless while agents such as SSRI are more likely prescribed by psychiatrists in the community to treat anxiety disorders, use of these agents to treat disorders in prisons remains widespread.

Some prescribed agents may not be abused in the community, but they are frequently abused in prison. Practitioners appear less likely to consider medications such as atypical antipsychotics and antidepressants as having the potential for abuse or dependence. Quetiapine (Seroquel®), an atypical antipsychotic, and Bupropion (Wellbutrin®), an SSRI antidepressant, have been identified as possible drugs of abuse within the correctional setting (Del Paggio 2005). Abuse was indicated by: (1) clinical presentations atypical at best; (2) doses that would be considered subtherapeutic for control of the reported "psychotic" symptoms; (3) refusal to consider other treatment options for their symptoms; (4) asking specifically for Seroquel® and Wellbutrin®, and (5) clinical presentations frequently suggested malingering of symptoms. Educational programs and close monitoring is often effective to combat this abuse.

As noted above, depression is common and often treated with antidepressants (Baillargeon et al. 2000). Antidepressants are effective in the correctional setting when the diagnosis is appropriately made, as shown by their increased use relative to other agents such as antipsychotics despite their cost and relatively strong adherence (Baillargeon et al. 2000). Although African-Americans are more likely to receive antidepressants in the correctional system, they are still less likely to receive them compared to Caucasians and more likely to receive them compared to Latinos.

Depression is strongly associated with suicide which is unfortunately not uncommon in the correctional system. As such, effective treatment of depression should have a positive effect on suicide rates within the correctional system. In fact, a marked decrease of suicide rates occurred after the introduction of SSRIs, and a significant association was seen between SSRIs, other new-generation antidepressants, and lower suicide rates, while TCAs are associated with higher suicide rates (Gibbons et al. 2005).

Antipsychotics have been widely used because of the incarceration of the severely ill and the need to treat aggressive behavior. The comparatively high acquisition costs of the newer antipsychotic medications have caused the mental health community to look closely at their potential benefits. Initiation of the atypical olanzapine therapy was associated with favorable clinical outcomes with the incarcerated mentally ill, particularly in patients with a formal diagnosis of thought disorder. The decline in total resource utilization was not statistically significant, although it may be of practical importance (Del Paggio et al. 2002).

Mood stabilizers have been useful as well because they are effective in treating both bipolar disorder, which disproportionately affects recidivism and aggression. Medication adherence is important in preventing re-incarceration and homelessness for those patients incarcerated with bipolar disorder and treated with mood stabilizers (Copeland et al. 2009). Lithium has been found to reduce aggression in a variety of settings including corrections (Müller-Oerlinghausen and Lewitzka 2010). Divalproex is effective in reducing impulsivity and mood liability in bipolar and non-bipolar patients. However, the impulsive/aggressive subgroup was the only non-bipolar subgroup in which DVX yielded clinical benefit (Kamath et al. 2008). The author found that divalproex was effective in reducing disruptive behavior in both groups (Lawson and Nanos 2008).

Ethnopsychopharmacology

Ethnic differences have been reported for the metabolism of many psychotropic medications. Given the large ethnic minority populations in the correctional system, such findings are extremely relevant. African-Americans may require less medication than Caucasians due to differential pharmacological response (Bradford 2002). For instance, African-Americans may require lower doses of antipsychotic and antidepressant medications than Whites. Cytochrome P 450 enzymes metabolize over 90 % of drugs in clinical use including antipsychotic and antidepressant medications, and individuals with relatively inactive CYP2D6 alleles (which account for 25 % of metabolism of commonly used drugs) tend to have higher plasma levels of antipsychotics and antidepressants (Bradford 2002). However, 50 % of people of African ancestry have reduced functioning or nonfunctioning alleles, which lead to slower metabolism of older antipsychotics or tricyclic antidepressants and higher plasma levels. Chronically higher plasma levels may be associated with an increased risk of side effects, such as intolerance of the medication, which in turn can lead to poorer adherence to treatment. African-Americans are also more likely to develop tardive dyskinesia when receiving first-generation antipsychotics (Lawson 2007). A noted disparity in the provision of treatment is that African-Americans are more likely to be prescribed antipsychotics and are less likely to be prescribed newer antipsychotics or antidepressants (Lawson 2007). As mentioned above, African-Americans are also less likely to be offered different types of psychotherapy.

Many clinicians are unaware of potential racial/ethnic differences in pharmacological responses. For example, many anticipate that African-Americans will respond to medication similarly as other racial/ethnic groups do since considerable genetic similarities exist across such groups. Many also feel that African-Americans require more medication based on the misconception that African-American males are more hostile (Lawson 2002). However, the evidence suggests otherwise. Since the older medications are more likely to be metabolized by the CYP2D6 enzymes, African-Americans may be at greater risk of experiencing side effects when given the same dose as Whites because the drugs are metabolized more slowly. Moreover, since African Americans are more likely to be perceived of as being hostile or having schizophrenia, they may end up in receiving even higher doses and ultimately less likely to adhere to treatment after release.

Substance Abuse Pharmacotherapy

Substance abuse treatment has become an important intervention for reducing recidivism among inmates, especially since substance abuse is a major factor in the growth of US prison populations. However, pharmacotherapy has become an effective treatment in addressing drug abuse. An estimated 12–15 % of inmates have histories of heroin addiction (Mumola and Karberg 2006) and not surprisingly much of the research has involved opiate pharmacotherapy. The full agonist methadone and related agents have been found to be effective in addressing drug use problems in the correctional setting and post-release. Methadone treatment begun in prison and continued in the community upon release extended the time parolees remained in treatment, reduced further drug use, and produced a threefold reduction in criminal activity (Kimlock et al. 2009). A review of current studies showed that prerelease opiate maintenance therapy (OMT) was significantly associated with increased treatment entry and retention after release if arrangements existed to continue treatment. For other outcomes, associations with prerelease OMT were weaker but included postrelease reductions in heroin use. Evidence regarding crime and re-incarceration was equivocal, but as noted above, some studies found positive results.

Buprenorphine is a long-acting partial agonist that acts on the same receptors as heroin and morphine, relieving drug cravings without producing the same intense "high" or dangerous side effects. Congress passed the Drug Addiction Treatment Act, permitting qualified physicians to prescribe Schedule III–V narcotic medications for the treatment of opioid addiction. This legislation created a major paradigm shift by allowing access to opiate treatment in a medical setting rather than limiting it to federally approved opioid treatment programs. Buprenorphine has advantages over methadone including less associated stigma; fewer regulations, which permit its use outside opioid treatment programs; and lower risk of overdose. Its combination with naloxone reduces the likelihood of intravenous abuse of the medication. It can be a valuable tool in addressing the problem of adherence since in 85–90 % of such persons relapsing to opioid use within 1 year after release, regardless of duration of incarceration (Kinlock et al. 2002). Overall, Buprenorphine has been found to be effective during incarceration and for post-release (Sullivan et al. 2006).

Corrections and Public Health

Since the first reports of HIV/AIDS occurring in US prison populations were published more than 25 years ago, epidemiologic studies have consistently documented a high prevalence of HIV infection and AIDS among prison inmates. The most recent estimates show that the prevalence of HIV infection in the US prison population is more than three times higher than that of the general population (Spaulding et al. 2009). Moreover, one-sixth of all infected individuals in the USA pass through the criminal justice system annually (Spaulding et al. 2009). Consequently, interventions are urgently needed for individuals with HIV disease who are transitioning back to the community. Moreover, 20 % of HIV-infected inmates released from prison were re-incarcerated within 3 years of their release date, creating a cycle of reinfection (Balergeon et al. 2010). Buprenorphine was associated with reduced relapse to opioid use and sustained viral suppression among released HIV-infected prisoners meeting criteria for opioid dependence (Springer et al. 2012) which helps combat this cycle of reinfection.

The impact of HIV infection and the cycle of reinfection on the African-American community is especially important. African-Americans have both higher HIV infection rates and have a greater representation in the correction system (El-Sadr et al. 2010). African-Americans have less access to services upon release (Primm and Lawson 2010).

Mental disorders also contribute to the spread of AIDS. Depression has been linked to disease progression in HIV, risky behaviors, and medication non-adherence. Moreover, treatment of depression leads to better adherence of HAART medications (Walkup et al. 2008). Unfortunately it is well documented that ethnic minorities have limited access to treatments for affective disorders which may further exacerbate the risk of the spread of AIDS in the African-American community (Primm and Lawson 2010).

Need for Integrated Treatment

Treatment in the correctional system must address the complex problem of substance use comorbidity. Because drugs of abuse affect similar brain circuits or receptor mechanisms proposed for mental disorders, they can cause drug abusers to experience one or more symptoms of mental illness (Lawson et al. 2011). When substance abusers present with similar symptoms of mental disorders, an underlying mental disorder may be missed and a drug abuser may be incarcerated rather than provided treatment. The result could be misdiagnosis and inappropriate treatment. Additionally, individuals with overt, mild, or even subclinical mental disorders may abuse drugs as a form of self-medication (Lawson et al. 2011). As a result, the mental disorder is considered yet another symptom of substance abuse, resulting in a delay in or absence of treatment for the mental disorder, especially in the correctional settings. Substance abuse adds to the burden by exacerbating mental illness and leading to a poorer outcome. In correctional systems, co-occurring disorders greatly exacerbate the rate of recidivism (Baillargeon et al. 2010). And, those with co-occurring substance abuse disorders are more likely to be held in jail longer than other inmates charged with similar crimes (McNiel et al. 2005). The combination of mental illness and drug abuse contributes to a revolving door of incarceration.

HIV means that many of these inmates may have three disorders, a triple whammy. It has already been noted that drug abuse worsens the outcome of HIV. Untreated mental illness is also associated with poor outcome for HIV. Treatment for comorbidity is inherently more difficult and often requires specialized services. Yet, such treatment is important and perhaps necessary to improve outcome and protect the community.

Ensuring Treatment

Mental health and drug courts are a valuable and effective source for early detection, referral, and navigation of services (Gallagher et al. 2011). They offer fresh insight into the potentially beneficial and detrimental effects of legal decisions and view one of the roles of law as that of a healing agent. At present, several states have instituted mental health courts based on these concepts; incorporating previous drug court experiences (Gallagher et al. 2011) in their development. Their goal is to avoid the criminalization of the mentally ill and their recidivism through the creation of special programs.

Mental health courts promote a non-adversarial atmosphere in which participants interact directly with the judge and in which praise and encouragement are issued far more often than sanctions. One program reported a significant drop in psychiatric hospitalization days for mentally ill participants and a decrease in positive drug and alcohol tests. Another mental health court program in Nevada reported successfully keeping mentally ill offenders out of the correctional system while concomitantly improving their mental condition. Mental health courts have been found to lower posttreatment arrest rates and days of incarceration (Steadman et al. 2010). While mental health and drug courts are not in and of themselves treatment and depend heavily on existing treatment resources, they are an important first step to providing services to improve adherence and linkage to post-release services.

Conclusion

The changing inmate population in the correctional system makes treatment essential to prevent recidivism and protect the community. We now have treatments that are effective in addressing mental disorders, substance abuse, and general medical problems such as AIDS. Most importantly, these treatments are shown to be effective in the correctional system and show strong benefits post release. Unfortunately, the issues that limit treatment in the community at large also impact the correctional community. As in the community, a large minority population, problems with disease recognition, need for integrated treatments, and barriers to treatment adherence exist in the correctional setting. Moreover, the culture of the system and financial challenges limit access to evidence-based treatment. Yet the need for ongoing and integrated treatment cannot be more urgent. For many with mental illnesses, substance abuse problems, or medical issues, the correctional system provides treatment opportunities for the first time in their lives. When they are released, it is essential that the cycle of high-risk behavior and recidivism is broken, both for the individual and the larger community. Educating the correctional system administrators and the general public on the benefits of evidence-based treatment is more important than ever. Recognition of mental disorders and substance abuse early in the process is crucial. Integrated service approaches combined with mental health/drug courts may produce the effective outcomes necessary for changing the life trajectories of former inmates. Dual diagnosis programs have now been introduced in many settings often with excellent results. However, many of those with comprehensive services in the correctional system discontinue them when released, which is why programs that promote adherence and linkage to services are essential (Foulks 2004). Providing early detection, personnel education, alternatives to incarceration, evidence-based treatment in corrections, and strong aftercare will be economically feasible and may help to reduce the disparities in health outcomes seen in ethnic minorities.

Recommended Reading and References

Alim TN, Graves E, Mellman TA, Aigbogun N, Gray E, Lawson W, Charney DS (2006) Trauma exposure, posttraumatic stress disorder and depression in an African-American primary care population. J Natl Med Assoc 98:1630–1636

Baillargeon J, Contreras S, Grady JJ, Black SA, Murray O (2000) Compliance with antidepressant medication among prison inmates with depressive disorders. Psychiatr Serv 51(11):1444–1446

Baillargeon J, Penn JV, Knight K, Harzke AJ, Baillargeon G, Becker EA (2010a) Risk of reincarceration among prisoners with co-occurring severe mental illness and substance use disorders. Adm Policy Ment Health 37:367–374

Baillargeon J, Giordano TP, Harzke AJ, Spaulding AC, Wu ZH, Grady JJ, Baillargeon G, Paar DP (2010b) Predictors of reincarceration and disease progression among released HIV-infected inmates. AIDS Patient Care STDS 24(6):389–394

Bewley MT, Morgan RD (2011) A national survey of mental health services available to offenders with mental illness: who is doing what? Law Hum Behav 35(5):351–363

Bradford LD (2002) CYP2D6 allele frequency in European Caucasians, Asians, Africans and their descendants. Pharmacogenomics 3:229–243

Copeland LA, Miller AL, Welsh DE, McCarthy JF, Zeber JE, Kilbourne AM (2009) Clinical and demographic factors associated with homelessness and incarceration among VA patients with bipolar disorder. Am J Public Health 99(5):871–877

Crisanti AS, Frueh BC (2011) Risk of trauma exposure among persons with mental illness in jails and prisons: what do we really know? Curr Opin Psychiatry 24:431–435

Del Paggio D, Finley PR, Cavano JM (2002) Clinical and economic outcomes associated with olanzapine for the treatment of psychotic symptoms in a county mental health population. Clin Ther 24(5):803–817

El-Sadr WM, Mayer KH, Hodder SL (2010) AIDS in America – forgotten but not gone. N Engl J Med 362(11):967–970

Fazel S, Seewald K (2012) Severe mental illness in 33,588 prisoners worldwide: systematic review and meta-regression analysis. Br J Psychiatry 200(5):364–373

Foulks EF (2004) Commentary: racial bias in diagnosis and medication of mentally ill minorities in prisons and communities. J Am Acad Psychiatry Law 32:34–35

Gibbons RD, Hur K, Bhaumik DK, Mann JJ (2005) The relationship between antidepressant medication use and rate of suicide. Arch Gen Psychiatry 62:165–172

Himelstein S (2011) Meditation research: the state of the art in correctional settings. Int J Offender Ther Comp Criminol 55(4):646–661

Iguchi MY, Bell J, Ramchand RN, Fain T (2005) How criminal system racial disparities may translate into health disparities. J Health Care Poor Underserved 16:48–56

James DJ, Glaze LE (2006) Mental health problems of prison and jail inmates, Bureau of Justice Statistics special report. Department of Justice, Washington, DC, NCJ Publication No. 213600

Joe S, Baser RE, Breeden G, Neighbors HW, Jackson JS (2006) Prevalence of and risk factors for lifetime suicide attempts among blacks in the United States. J Am Med Assoc 296:2112–2123

Kamath J, Temporini HD, Quarti S, Zhang W, Pagano KL, Demartinis N, Trestman RL (2008) Psychiatric use and utility of divalproex sodium in Connecticut prisons. Int J Offender Ther Comp Criminol 52(3):358–370

Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE (2005) Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 62:593–602

Kinlock TW, Battjes RJ, Schwartz RP (2002) A novel opioid maintenance program for prisoners: preliminary findings. J Subst Abuse Treat 22:141–147

Lawson WB (2005) Bipolar disorder in African Americans in perspectives. In: Georgiopoulos AM, Rosenbaum JM (eds) Cross-cultural psychiatry. Lippincott Williams and Wilkins, Philadelphia, pp 135–142

Lawson WB (2007) Schizophrenia in African Americans. In: Mueser KT, Jeste DV (eds) Clinical handbook of schizophrenia. Guilford, New York, pp 616–623

Lawson WB (2008) Identifying interethnic variations in psychotropic response in African-

Lewis DO, Shanok SS, Cohen RJ, Kligfeld M, Frisone G (1980) Race bias in the diagnosis and disposition of violent adolescents. Am J Psychiatry 137:1211–1216

Müller-Oerlinghausen B, Lewitzka U (2010) Lithium reduces pathological aggression and suicidality: a mini-review. Neuropsychobiology 62(1):43–44

Mumola C, Karberg J (2006) Drug use and dependence, state and federal prisoners. U.S. Department of Justice, Office of Justice Programs, Washington, DC

Primm AB, Lawson WB (2010) Disparities among ethnic groups: African Americans. In: Ruiz P, Primm A (eds) Disparities in psychiatric care: clinical and cross-cultural perspectives. Wolters Kluver/Lippincott Williams & Wilkins, Baltimore, pp 19–29

Spaulding AC, Seals RM, Page MJ, Brzozowski AK, Rhodes W et al (2009) HIV/AIDS among inmates of and releases from US correctional facilities, 2006: declining share of epidemic but persistent public health opportunity. PLoS One 4:e7558

Springer SA, Qiu J, Saber-Tehrani AS, Altice FL (2012) Retention on buprenorphine is associated with high levels of maximal viral suppression among HIV-infected opioid dependent released prisoners. PLoS One 7(5):e38335. doi:10.1371

Steadman HJ, Redlich A, Callahan L, Robbins PC, Vesselinov R (2010) Effect of mental health courts on arrests and jail days: a multisite study. Arch Gen Psychiatry 68:167–172

Strakowski SM, Keck PE Jr, Arnold LM, Collins J, Wilson RM, Fleck DE, Corey KB, Amicone N, Adebimpe VR (2003) Ethnicity and diagnosis in patients with affective disorders. J Clin Psychiatry 64:747–754

Sullivan LE, Barry D, Moore BA, Chawarski MC, Tetrault JM, Pantalon MV, O'Connor PG, Schottenfeld RS, Fiellin DA (2006) A trial of integrated buprenorphine/naloxone and HIV clinical care. Clin Infect Dis 43(4):S184–S190

Walker H (2004) Using psychosocial interventions within a high-security hospital. Nurs Times 100(31):36–39

Walkup W, Wei W, Sambamoorthi U, Crystal S (2008) Antidepressant treatment and adherence to combination antiretroviral therapy among patients with AIDS and diagnosed depression. Psychiatr Q 79(1):43–53

Workman D, Cunningham D (1975) Effect of psychotropic drugs on aggression in a prison setting. Can Fam Physician 21(11):63–66

Treatment

► Punishment as Rehabilitation

Treatment Matching

► Actualizing Risk-Need-Responsivity

Treatment of Sex Offenders

Friedrich Lösel[1,2] and Martin Schmucker[2]
[1]Institute of Criminology, University of Cambridge, Cambridge, UK
[2]Institute of Psychology, University of Erlangen–Nuremberg, Erlangen, Germany

Synonyms

Sexual offender treatment (SOT)

Overview

This entry gives a brief overview on the effects of sexual offender treatment drawing on evidence from a number of meta-analyses and systematic reviews. Overall, this research reveals a positive effect on sexual reoffending. The syntheses suggest a mean effect of approximately $d = 0.20 \pm 0.10$. Because of the low base rate of sexual reoffending, this equals about 25 % less recidivism of treated offenders as compared to untreated groups. However, the studies and findings are extremely heterogeneous and outcomes vary depending on many program, setting, offender, and methodological factors.

Unfortunately, the quality of the majority of evaluation studies on sex offender treatment is rather weak. However, the overall design quality is not consistently related to effect sizes, and the same applies to various other methodological and content characteristics. Cognitive-behavioral treatment, relapse prevention, and programs that adhere to the Risk-Need-Responsivity Model show relatively consistent positive effects, although even such interventions sometimes fail to demonstrate positive effects. Hormonal medication for subgroups of sexual offenders is promising, too, but it needs more methodologically sound evaluations on this topic. By trend, programs in ambulatory settings, with some individualization and proper implementation, and for offenders at higher risk of reoffending, show above-average effects.

The small number of methodologically sound studies with specific characteristics and confounded moderator variables makes it difficult to draw robust conclusions. Despite practical problems in undertaking randomized experiments in this area, more high-quality evaluations are highly needed.

Introduction

In many countries sexual offending is a topic of particularly high concern in the general public and the media. Accordingly, governments implemented two main strategies to deal with sex offenders. On the one hand, they put into effect tougher sentences and security measures for this offender population. On the other hand, they invested in sexual offender treatment programs to reduce reoffending. Although treatment of sexual offenders became common in many Western countries, there is much controversy about its effectiveness (Marshall & Marshall 2010; Rice & Harris 2003). A consensus on "what works" in sexual offender treatment is complicated by various issues. First, sexual offending is a very heterogeneous category that contains, for example, various forms of child molesting, rape, exhibitionism, distribution, and consumption of child pornography in the internet. Second, there are very different types of sexual offenders such as those with or without a deviant sexual preference (paraphilia), an antisocial personality, an opportunistic orientation, and neuropsychological deficits. Third, although much research on risk factors for reoffending and structured assessment instruments has been carried out, the evidence on the causal mechanisms is less clear. Fourth, treatment approaches are very heterogeneous ranging from cognitive-behavioral programs and relapse prevention over psychodynamic approaches and multi-systemic therapy to hormonal treatment by medication or surgical castration, and some of these categories embrace rather different approaches in themselves (McGrath et al. 2010). Fifth, sound treatment evaluation of routine practice is difficult because in many jurisdictions serious sexual offenders cannot simply be left untreated in control groups and sexual reoffending requires rather long follow-up periods.

Accordingly, compared to general or violent offender treatment, methodologically sound evaluations of programs for sexual offenders are less frequent, particularly outside North America (Lipsey and Cullen 2007; Lösel 2012). However, since the 1990s, controlled evaluations have increased and a number of systematic reviews and meta-analyses could be carried out. Table 1 summarizes important features and findings of 12 reviews and meta-analyses that (a) mainly addressed adult sexual offenders, (b) used sexual reoffending as outcome measure (e.g., arrest, conviction, incarceration, or self-reported offenses), and (c) calculated effect sizes (ES) or at least provided systematic data that can be used for effect size estimations. Concentrating on meta-analyses reduces problems of selective file drawing and subjective evaluation. A focus on effect size instead of mere statistical significance provides more adequate information because many sex offender treatment studies contain rather small samples and thus lack statistical power. The remaining entry will provide a brief overview of the main findings and the controversies that arise from those research integrations. Although there is overlap between these syntheses, they vary substantially with regard to the included primary studies, categories for coding, methods of data analysis, calculation and integration of effect sizes, and investigation of outcome moderators.

Some meta-analyses concentrated on psychotherapeutic/psychosocial interventions only. Others also included hormonal treatment by medroxyprogesterone-acetate or cyproterone-acetate to reduce male sexual hormones as well as surgical castration. In the general category of psychotherapeutic/psychosocial interventions, the individual treatment programs vary considerably and contain interventions such as behavior modification, cognitive-behavioral treatment (CBT), relapse prevention, psychodynamic approaches, social therapy, or multi-systemic therapy. Some of

Treatment of Sex Offenders, Table 1 Content and selected results of meta-analyses on the treatment of sexual offenders

Meta-analysis	Program types	*k*	Total *n*	Design quality[a]	Recidivism TG/CG (%)[b]	ES (*d*)[c]	ES of different program types[d]
Furby et al. (1999)	P, C	11	3,977	2–5 (2)	8.2/17.1	0.34	C > P
Hall (1995)	P, H, C	12	1,313	2–5 (2)	18.9/26.8	0.24	C/H and CBT/RP > BM
Alexander (1999)	P	79	10,988	1–5 (1)	13.8/18.0	0.12	CBT/RP > other P
Grossman et al. (1999)	P, H	12	1,283	2–5 (2)	17.6/27.4	0.28	H > CBT/RP > other P
Polizzi et al. (1999)	P	12	5,542	2–4 (2)	13.9/14.5	0.16	CBT/RP > other P
Gallagher et al. (2000)	P, H, C	23	NA	2–5 (3)	12.0/22.0	0.47	C > H > CBT/RP > other P
Lösel (2000)	P, H, C	20	1,048	1–5 (1)	19.3/22.3	0.08	C/H > P
Aos et al. (2001)	P, H, C	14	3,342	3–5 (3)	NA	0.19	CBT/RP > other P
Hanson et al. (2002)	P	38	8,164	2–5 (2)	12.3/16.8	0.12	Current P (often CBT) > noncurrent
Lösel and Schmucker (2005)[e]	P, H, C	74	22,181	2–5 (2)	11.1/17.5	0.29	C > H > BM/CBT/RP > other P
Hanson et al. (2009)	P	23	6,746	3–5 (3)	10.9/19.2	0.23	P adhering to RNR > other P
Schmucker and Lösel (2013)	P	27	8,347	3–5 (3)	9.0/12.4	0.20	MST and CBT > other P

P, psychotherapeutic/psychosocial interventions (overarching category that includes CBT, BM, RP, MST, and other P); H, hormonal medication; C, surgical castration; CBT, cognitive-behavioral treatment; RP, relapse prevention; BM, behavior modification; MST, multi-systemic therapy; RNR, risk-need-responsivity model; *k*, number of studies/comparisons; *n*, total number of offenders in this review (TG + CG); TG, treatment group; CG, control/comparison group; ES, effect size; *d*, Cohen's coefficient

[a] Rating on the Maryland Scale of Methodological Rigor (1 = No CG, 2 = Non-equivalent CG, 3 = Indicators of equivalent CG, 4 = Matching procedures, 5 = Randomized Control Trial)

[b] In meta-analyses that included primary studies on Level 1 of the Maryland Scale, the reported data refer to the studies with CG only

[c] ESs were transformed to Cohen's *d* for comparability when another effect size measure was reported in the original meta-analysis. In reviews that did not report ESs, we used the available information provided to estimate *d*- and the *n*-weighted overall ES

[d] The differences between program types are not always significant, but show a systematic tendency

[e] It includes analyses of Schmucker and Lösel (2008) that addressed the same set of primary studies

these intervention types share similarities. For example, the contents of CBT (e.g., reducing deviant sexual attitudes, improving self control, enhancing social skills, promoting perspective taking, or coping with stressors) overlap with those of relapse-prevention programs that focus more strongly on risk situations and fantasies of offenders with a deviant sexual orientation (paraphilia). And multi-systemic therapy for young sexual offenders and social therapy/therapeutic communities make particular use of the social context of the clients but also incorporate elements from CBT or psychodynamic approaches. Hormonal treatment is primarily used for those offender subgroups who are mainly motivated by sexual drive (and not by dominance, etc.), but the medication is often accompanied by CBT or other psychosocial interventions.

Table 1 also presents the methodological quality of the included primary studies using a slightly modified version of the Maryland Scale of Methodological Rigor (MSMR; Farrington et al. 2002) which has five levels: *Level 1*: Only treatment group (TG) data, no control or comparison group (CG); *Level 2*: Nonequivalence of TG and CG, differences on relevant variables effecting recidivism are reported or are to be expected (e.g., treatment dropouts or nonstarters); *Level 3*: Incidental assignment but equivalent CG, sound statistical control, or empirical data on equivalence in relevant variables; *Level 4*: Matching procedures, systematic strategy to attain equivalence of TG and CG; and *Level 5*: Random assignment of subjects to TG and CG, no selective dropout.

Most meta-analyses contain a majority of studies that are rather weak methodologically (Levels 1 and 2). This is a serious problem because the results from such studies can be distorted. However, while high internal validity is a precondition for any causal inference and generalization, an exclusive reliance on randomized controlled trials (RCTs) would lead to a very narrow database. For example, a Cochrane review of RCTs on sex offender treatment ended up with only three studies on very different modes of treatment (White et al. 2009). Thus, hardly any conclusion could be drawn except stating the lack of RCTs. A review of Brooks-Gordon et al. (2006) contained nine RCTs, but the majority did not use reoffending as outcome measures. Most recently Schmucker and Lösel (2013) found six RCTs on sexual offender treatment with offending as outcome. However, the findings of these studies were very different and did not reveal a clear picture that could guide practice and policy making. It is necessary to conduct more RCTs in the evaluation of sexual offender treatment. But one has to be aware of the practical problems in implementing *and* maintaining an RCT in this research area. One should also take into account that methodological quality is a multidimensional category and sound quasi-experiments can provide useful information (Lösel 2007). Against this background it seems helpful not yet to ignore findings from suboptimal studies on sexual offender treatment.

In total, the meta-analyses indicate a substantial increase of evaluations in the 1990s and 2000s. The large number of studies and high *n* in Alexander (1999) is due to the inclusion of methodologically weak studies without comparison groups. The increase in evaluation research is also accompanied by better design quality. As the most recent review of Schmucker and Lösel (2013) shows, nearly half of the studies at Level 3 or above of the MSMR have been published in the 2000s. Perhaps, the public concern about sexual offending led to more interest in sound evidence and research funding in this field. However, one must bear in mind that even the recent evaluation reports do not necessarily address the most recent programs. Sexual reoffending has a relatively low base rate and often only occurs after rather long time periods. Therefore, many evaluations of sexual offender treatment have rather long follow-up periods. Whereas in the evaluation of general offender treatment a follow-up of 1–2 years is frequent, the average follow-up in most meta-analyses reported in Table 1 is clearly over 4 years, ranging to more than 20 years for individual primary studies. This is a positive feature of the respective primary research, but it also has the implication that completed evaluations often address programs that have been revised in the meantime.

In all meta-analyses, except for Lösel (2000) which integrated German studies, the majority of included evaluations came from North America. This may be partially a consequence of the focus on English-language reports by native English-speaking authors. However, similar to other fields of correctional treatment, the dominance of the United States and Canada also reflects a stronger evidence orientation and research activity in the Anglo-American World. It is good news that some meta-analyses on sexual offender treatment also found sound evaluation research in other parts of the world. For example, in the most recent review, approximately one third of the eligible studies stem from outside North America. Despite this positive tendency there are still too few studies from different countries to draw conclusions about the impact of the cultural or legal framing conditions on the effects of sexual offender treatment.

Overall Effect

All meta-analyses suggest a desirable mean effect of sexual offender treatment. The average rates of sexual recidivism in treated offenders varied between 8.2 % (Furby et al. 1989) and 19.3 % (Lösel 2000). In contrast, in the comparison groups, they varied between 12.4 % (Schmucker and Lösel 2013) and 27.4 % (Grossman et al. 1999). In the various meta-analyses the recidivism rates for the TG were approximately 1–10 percentage points lower than the rates for CG. Although, such differences may look small at first glance, one must take the relatively low sexual recidivism rates in the CG into account. For example, Hanson et al. (2002) reported an average sexual recidivism rate of 12.3 % for TG and 16.8 % for CG. This difference of 4.5 percentage points translates into a reduction of 27 % as compared to untreated offenders. The absolute difference between 9.0 % (TG) and 12.4 % (CG) in the meta-analysis of Schmucker and Lösel (2013) was even smaller than in Hanson et al. (2002); however, the reduction in percent was the same (27 %). Of course, one must take into account undetected or unreported recidivism so that the "real" reoffending rates may be higher than those presented in Table 1. However, this is a general criminological problem that even may have decreased a bit in sexual offending because in many countries victims are more encouraged to report to the police or aid organizations.

Computing an overall ES from the various reviews is problematic because of the overlap between the reviews, different sample sizes, methods of analysis, and other issues. However, scanning the reported overall effect sizes across the different meta-analyses reveals that most of them are in the range of $d = 0.10$–0.30. Therefore $d = 0.20$ plus/minus 0.10 (approximately $r \sim 0.10$; odds ratio ~ 1.5) seems to be a realistic estimate of the effect that can be expected on average. This ES range of sexual offender treatment is similar to the ESs reported in the much more comprehensive research on general offender treatment (e.g., Lösel 2012). It is noteworthy that sexual offenders do not only reoffend with sex offenses. In fact, the meta-analyses show that the recidivism rates of sexual offenders were higher for nonsexual offenses. For example, Hanson et al. (2009) reported 31.8 % (TG) versus 48.3 % (CG) and Schmucker and Lösel (2013) reported 24.0 % (TG) versus 32.4 % (CG). This indicates that sexual offender treatment has a broader effect as well.

Treatment Characteristics as Moderators

Mode of Treatment. Nearly all meta-analyses show that cognitive-behavioral treatment (CBT) and relapse-prevention approaches (RP) exhibit larger effects than unspecific psychosocial or psychodynamic interventions. This result is in accordance with general offender treatment (Lipsey & Cullen 2007; Lösel 2012). But although CBT is shown to be effective in principle, this program type does not guarantee reduced sexual reoffending. In well-controlled evaluations the effects of similar CBT programs ranged from strong positive over nearly zero to even negative effects (Schmucker and Lösel 2013; see also individual studies in Hanson et al. 2009).

Beyond the evaluation of specific psychotherapeutic "schools," the literature on "what works" in offender rehabilitation developed broader program characteristics such as the Risk-Need-Responsivity Model (Andrews et al. 2011). Similar to the findings from general offender treatment, Hanson et al. (2009) demonstrated that programs that fulfill all three principles are most successful for sexual offenders, while treatment that does not meet any of the three criteria even shows a slightly negative effect.

Those meta-analyses that also include hormonal therapy via antiandrogen medication and/or surgical castration suggest relatively large effects for this kind of treatment. However, hormonal interventions are only indicated for a subgroup of offenders who are primarily motivated by sexual drive. Surgical castration is now rarely used because of legal and ethical reasons (Weinberger et al. 2005). The respective evaluation studies are of older date and they do not contain equivalent untreated groups. The respective CGs seem to be a negative selection of offenders not motivated for this kind of treatment or rejected by the decision committees. None of the studies reaches Level 3 of the MSMR (Lösel and Schmucker 2005). Although the designs of hormonal medication evaluations are somewhat stronger, they also fail to meet the Level 3 eligibility criterion of Schmucker and Lösel (2013). Furthermore, hormonal therapy has the highest refusal and dropout rates (e.g., Hall 1995). In these cases, those offenders who remained in treatment may have been a highly motivated selection. More sound evaluations of hormonal medication are necessary. One must also bear in mind that hormonal medication is rarely applied as a stand-alone measure but is normally combined with some form of psychosocial intervention. In that sense, these primary studies evaluated the effects of medication added to psychotherapeutic measures. Combinations of CBT and RP approaches with pharmacological treatment are promising for some groups of sex offender, and such approaches would be in accordance with common practice in other fields such as the treatment of depression.

Treatment Setting. Most meta-analyses report a tendency for larger effects in community than in custodial programs. However, some syntheses did not observe this difference (e.g., Alexander 1999; Hanson et al. 2002; Hanson et al. 2009). In principle, one would expect better effects of interventions in the community because treatment progress can be directly tested in risk situations, whereas the transfer of what has been learnt in custody is a key issue in resettlement and relapse prevention. However, the offender populations in both settings are not the same. No study compared treatment in institutional and ambulatory settings for the same type of sexual offenders. As more high-risk cases are to be expected in custody, the outcome differences between settings may be confounded with other features. When institutional treatment is further differentiated into prisons and forensic hospitals, some data suggest that the latter goes along with larger ES (Lösel and Schmucker 2005; Schmucker and Lösel 2013). In contrast, Alexander (1999) found larger ESs for prison compared with hospital settings, but this synthesis contains many methodologically weak studies.

Other Treatment Factors. In addition to treatment mode and setting, only a few other features of program delivery have been addressed in some meta-analyses. For example, Schmucker and Lösel (2013) found that programs with a group format were less effective than programs that included individual sessions or were fully individualized. Although there are plausible arguments for treating sexual offenders in groups (Ware et al. 2009), this finding suggests that some differentiation according to specific offender needs is valuable (see also Marshall 2009).

Surprisingly little is known about the impact of the quality of program delivery. Only few primary studies report data on this issue so that it could not be addressed in the meta-analyses. Research in other areas of correctional treatment has shown that quality of delivery and program integrity are highly relevant for effectiveness (Goggin and Gendreau 2006). However, this issue needs to be investigated more systematically for sexual offender treatment, too. This also applies to the impact of therapist characteristics and the therapeutic relationship. Although research on general psychotherapy has shown that such features are as important as the treatment content, there is only little research on this issue in programs for sexual and other offenders.

Offender Characteristics as Moderators

Offense or Offender Types. The meta-analyses report only few differences in ES for different types of sexual offenses or offenders. For example, Gallagher et al. (2000) and Hall (1995) found no significant differences between various offense types. Alexander (1999) revealed larger ES for non-incest child molesters with male victims and the smallest ES for rapists. Lösel and Schmucker (2005) reported significant effects for all offense categories except intra-familial child molesting. The unsatisfactory state of systematic knowledge about this issue is due to a lack of detailed diagnostic information in the primary evaluation studies.

Research on sex offender recidivism has shown that specific offender and offense characteristics such as use of violence, type of victims, general psychopathology, and paraphilia correlate with reoffending (Hanson and Morton-Bourgon 2005). However, most treatment studies differentiate only according to the index offense between child molesters, rapists, and exhibitionists or even use the summary characteristic of sex offender. Some narrative reviews suggest better success for CBT of child molesters and exhibitionists (e.g., Marshall et al. 1991), but the evidence is rather scarce and as yet there are no well-replicated results on differential indication from the treatment evaluations.

Risk Level. There is a clearer picture of findings with regard to offenders' risk of sexual reoffending. Two recent meta-analyses showed that ESs are larger for groups at higher than lower risk (Hanson et al. 2009; Schmucker and Lösel 2013). A simple explanation for this finding is that the base rate for sexual reoffending in low-risk groups is so low that even successful programs cannot further reduce them. In part, this finding may also be confounded with the age of offenders.

Age. According to the age-crime curve, young offenders have a higher risk of reoffending. Therefore, young age is included in evidence-based actuarial risk assessment instruments for sexual offenders. Although the present entry focuses on the treatment of adult offenders, some of the meta-analyses also included studies on juveniles. Lösel and Schmucker (2005) and Schmucker and Lösel (2013) found a tendency of larger ES in studies that addressed adolescents only. In contrast, Hanson et al. (2009) reported an opposite trend of larger ES in adult samples. Due to the small number of primary studies that specifically addressed adolescent sex offenders, the respective differences in all three meta-analyses did not reach statistical significance. However, two meta-analyses on the treatment of juvenile sexual offenders support the assumption of larger effects in these target groups: Walker et al. (2004) found a mean effect of $d = 0.54$ and Reitzel and Carbonell (2006) a mean of 0.47. With the exception of Gallagher et al. (2000), both ESs are clearly above those in the meta-analyses of mainly adult offenders (see Table 1).

Treatment Motivation. The motivation to engage in treatment and avoid reoffending is seen as a key factor for success. However, in contrast to classical psychotherapy treatment, motivation is no longer seen as an all-or-none category and a more or less static characteristic of the offender, but as a dynamic concept of interaction between the offender and his social context (McMurran 2002). Therefore, similar to interventions for drug-addicted patients, the treatment of sexual offenders is not only a voluntary offer. Recent findings support this view. Whereas Lösel and Schmucker (2005) found a somewhat (but not significantly) larger effect in voluntary than in mandatory programs, Schmucker and Lösel (2013) revealed nearly identical ESs for both approaches to treatment participation.

Methodological Features

Although practice is primarily interested in program and offender factors that moderate treatment effects, characteristics of the evaluation method are also very important. In the meta-analysis of Lösel and Schmucker (2005), methodological characteristics of the primary studies even explained the largest amount of ES variation. However, in the field of sexual offender treatment, it is difficult to disentangle the independent influence of a factor because the subsamples with a specific characteristic are small and there are often confounded variables.

Design Quality. Table 1 suggests that the methodological quality varied substantially both within and between meta-analyses. In general it is still not satisfactory. The findings on the relation between design quality and the effect of sexual offender treatment are mixed. The two meta-analyses with a large portion of uncontrolled outcome studies showed the smallest effects (Alexander, 1999; Lösel, 2000). The Gallagher et al. (2000) synthesis of studies with relatively sound designs reported the largest mean effect. However, the latter meta-analysis also contained studies that used not only official recidivism as outcome data but also self-reports, family reports, or other information. As these studies revealed relatively large effects, ES differences between meta-analyses do not necessarily depend on a single dimension of methodological rigor. The other meta-analyses on relatively sound studies revealed ESs that are lower. All three reviews found a mean ES at around 0.20 (Aos et al. 2001; Hanson et al. 2009; Schmucker and Lösel 2013).

There is also no clear picture on this issue *within* the various meta-analyses. Some syntheses observed a tendency of lower ES in studies with weaker design (e.g., Gallagher et al. 2000; Polizzi et al. 1999). Others found an opposite tendency (Hanson et al. 2002, 2009) or no systematic relation between design quality and ES (Hall 1995; Lösel and Schmucker 2005; Schmucker and Lösel 2013). The same was the case with regard to RCTs versus other designs. In reviews on general offender treatment, there is also no clear relation between overall design quality and ES (Lipsey and Cullen 2007; Lösel 2012). This finding deviates from studies on a broader range of criminological interventions where rigorous designs often go along with lower ES (Weisburd et al. 2001). However, multivariate analyses of the impact of methodological quality on the ES in sex offender treatment are hindered by lacking information and the small number of RCTs. The relation between ES and design quality may even depend on the use of a fixed versus random effects model of ES integration (e.g., Hanson et al. 2009). In some meta-analyses, results depended on whether a single study with extreme effects was included or not (e.g., Gallagher et al. 2000; Lösel 2000).

Sample Size. The problem of disentangling moderators is also relevant with regard to sample size. Similar to findings on general offender treatment (Lipsey and Cullen 2007), some meta-analyses reported the largest effect in very small samples. This may be due to better program implementation in small studies, in particular when the evaluation was carried out in a demonstration project and the authors evaluated an own program. However, the impact of sample size could also reflect a publication bias because smaller studies require a larger ES to get a significant result. Accordingly, Gallagher et al. (2000) and Hanson et al. (2009) reported larger effects in published studies, however, the latter authors for the fixed effects model only. In contrast Hanson et al. (2002) found larger effects in unpublished studies and Lösel and Schmucker (2005) even observed a somewhat greater moderator effect of sample size in unpublished studies.

Length of Follow-Up. Although the rate of recidivism increases with the length of follow-up, the ES in treatment evaluations seems not to decrease generally over time. For example, Lösel and Schmucker (2005) and Schmucker and Lösel (2013) found no significant correlation between ES and follow-up time. Gallagher et al. (2000) and Hall (1995) even reported that the length of follow-up related positively to ES. This contradicts results on general offender treatment (Lipsey and Cullen 2007) but may have been due to outliers with very long follow-up periods and very high ES. Very long periods of follow-up might also cause problems because data may have been erased from official registers for some (inactive) offenders.

Other Methodological Moderators. There are many other potential moderators of ES. However, they were not systematically addressed across the various meta-analyses. For example, Gallagher et al. (2000) and Lösel and Schmucker (2005) found that *official recidivism* showed lower ESs than other outcome measures (i.e., self-reports or composite measures). Perhaps, official data are less sensitive to change but such effects are difficult to demonstrate due to confounding with other variables. *Attrition rates* also seem to be relevant for ES. High attrition rates may indicate a low quality of program implementation and also lead to lower effects when the dropouts are counted as part of the TG in an intent-to-treat analysis (Lösel and Schmucker 2005).

A particularly important issue is the transparency of the report on the evaluation. Although the concept of *descriptive validity* (Lösel and Köferl 1989) is not a characteristic of the evaluation process itself, it shows substantial correlations between $r = 0.24$ and 0.50 with ES (Lösel and Schmucker 2005; Schmucker and Lösel 2013). It can be assumed that a detailed and transparent documentation of the treatment concept, implementation, outcome measurement, and statistical analysis indicates a good overall quality of the respective evaluation. Therefore, it is strongly recommended that program evaluations pay more attention to this issue.

Conclusions

This brief review of meta-analyses on the effects of sex offender treatment revealed an overall positive effect on sexual reoffending and various characteristics of the most promising interventions.

However, the heterogeneity of study outcomes accompanied by a lack of detail in the documentation of programs, offender populations, setting variables, and other features as well as often weak study designs leaves many questions open. A further increase of high-quality primary studies and the systematic integration of their results are crucial to answer open questions and help turn suggestive findings into confirmed knowledge. Such studies should also focus on more homogeneous and well-assessed subgroups of sexual offenders and provide sound information on the treatment process. This would lead to a more solid and differentiated evidence base about what works for whom, under what conditions, in what contexts, and with regard to what outcomes. As sexual offending is a topic of much public concern, such evaluations are particularly important to replace stereotypes and opinions in policy making by sound evidence.

Related Entries

- ▶ Behavioral Management in Probation
- ▶ Child Pornography
- ▶ Criminal Career of Sex Offenders
- ▶ Effective Community Supervision Programming
- ▶ Effective Supervision Principles for Probation and Parole
- ▶ Empathy and Offending
- ▶ Estimating the Effects of Incapacitation
- ▶ Evidence-Based Policy in Crime and Justice
- ▶ Examining the Effectiveness of Correctional Interventions
- ▶ Good Lives Model
- ▶ History of Corrections
- ▶ History of Randomized Controlled Experiments in Criminal Justice
- ▶ Intervention and Prevention in China
- ▶ Offender Change in Treatment
- ▶ Pornography
- ▶ Propensity Score Matching
- ▶ Randomized Experiments in Criminology and Criminal Justice
- ▶ Risk Factors for Adolescent Sexual Offending
- ▶ Scared Straight Programs
- ▶ Sex Offenders and Criminal Policy
- ▶ Sexual Recidivism
- ▶ Specialization and Sexual Offending
- ▶ Treating Mentally Ill Offenders
- ▶ Victim-Focused Prevention of Child Sexual Abuse

Recommended Reading and References

Alexander M (1999) Sexual offender treatment efficacy revisited. Sex Abus J Res Treat 11:101–116

Andrews DA, Bonta J, Wormith S (2011) The risk-need-responsivity (RNR) model: does adding the Good Lives Model contribute to effective crime prevention? Crim Jus Behav 38:735–755

Aos S, Phipps P, Barnoski R, Lieb R (2001) The comparative costs and benefits of programs to reduce crime. Version 4.0 (No. 01-05-1201). Washington State Institute for Public Policy, Washington

Brooks-Gordon B, Bilby C, Wells H (2006) A systematic review of psychological interventions for sexual offenders I: randomised control trials. J Forensic Psychiatr Psychol 17:442–466

Farrington DP, Gottfredson DC, Sherman LW, Welsh BC (2002) The Maryland scientific methods scale. In: Sherman LW, Farrington DP, Welsh BC, MacKenzie DL (eds) Evidence-based crime prevention. Routledge, London, pp 13–21

Furby L, Weinrott MR, Blackshaw L (1989) Sex offender recidivism: a review. Psychol Bull 105:3–30

Gallagher CA, Wilson DB, MacKenzie DL (2000) A meta-analysis of the effectiveness of sex offender treatment programs. Paper retrieved from http://www.wam.umd.edu/~wilsondb/papers/sexoffender.pdf

Goggin C, Gendreau P (2006) The implementation and maintenance of quality services in offender rehabilitation programmes. In: Hollin C, Palmer E (eds) Offending behaviour programmes. Wiley, Chichester

Grossman LS, Martis B, Fichtner CG (1999) Are sex offenders treatable? A research overview. Psychiatr Serv 50:349–361

Hall GCN (1995) Sexual offender recidivism revisited: a meta-analysis of recent treatment studies. J Consult Clin Psychol 63:802–809

Hanson RK, Gordon A, Harris AJR, Marques JK, Murphy WD, Quinsey VL, Seto MC (2002) First report of the collaborative outcome data project on the effectiveness of psychological treatment for sex offenders. Sex Abus J Res Treat 14:169–194

Hanson RK, Morton-Bourgon KE (2005) The characteristics of persistent sexual offenders: a meta-analysis of recidivism studies. J Consult Clin Psychol 73:1154–1163

Hanson RK, Bourgon G, Helmus L, Hodgson S (2009) The principles of effective correctional treatment also apply to sexual offenders: a meta-analysis. Crim Justice Behav 36:865–891

Lipsey MW, Cullen FT (2007) The effectiveness of correctional rehabilitation: a review of systematic reviews. Ann Rev Law Soc Sci 3:297–320

Lösel F, Köferl P (1989) Evaluation research on correctional treatment in West Germany: a meta-analysis. In: Wegener H, Lösel F Haisch J (eds) Criminal behavior and the justice system. Springer, New York pp 334–355

Lösel F (2000) The efficacy of sexual offender treatment: a brief review of German and international evaluations. In: van Koppen PJ, Roos N (eds) Rationality, information and progress in law and psychology. Metajuridica, Maastricht, pp 145–170

Lösel F, Schmucker M (2005) The effectiveness of treatment for sexual offenders: a comprehensive meta-analysis. J Exper Criminol 1:117–146

Lösel F (2007) Doing evaluation in criminology: balancing scientific and practical demands. In: King RD, Wincup E (eds) Doing research on crime and justice, 2nd edn. Oxford University Press, Oxford pp 141–170

Lösel F (2012) Offender treatment and rehabilitation: what works? In: Maguire M, Morgan R, Reiner R (eds) The Oxford handbook of criminology, 5th edn. Oxford University Press, Oxford, pp 986–1016

Marshall WL (2009) Manualization: a blessing or a curse? J Sex Aggress 15:109–120

Marshall WL, Marshall LE (2010) Can treatment be effective with sexual offenders or does it do harm? A response to Hanson (2010) and Rice (2010). Sex Offender Treat 5(2), online

Marshall WL, Jones R, Ward T, Johnston P, Barbaree HE (1991) Treatment outcome with sex offenders. Clin Psychol Rev 11:465–485

McGrath RJ, Cumming GF, Burchard BL, Zeoli S, Ellerby L (2010) Current practices and emerging trends in sexual abuse management: the Safer Society 2009 North American Survey. The Safer Society Press, Brandon

McMurran M (2002) Motivation to change: selection criterion or treatment need? In: McMurran M (ed) Motivating offenders to change. Wiley, Chichester, pp 3–13

Polizzi DM, MacKenzie DL, Hickman LJ (1999) What works in adult sex offender treatment? A review of prison- and non-prison-based treatment programs. J Offend Ther Comp Criminol 43: 357–374

Reitzel LR, Carbonell JL (2006) The effectiveness of sexual offender treatment for juveniles as measured by recidivism: a meta-analysis. Sex Abus J Res Treat 18:401–421

Schmucker M, Lösel F (2008) Does sexual offender treatment work? A systematic review of outcome evaluations. Psicothema 20:10–19

Schmucker M, Lösel F (2013) Evaluating the effects of sexual offender treatment: an international meta-analysis of sound quality studies. Paper submitted for publication

Walker DF, McGovern SK, Poey EL, Otis KE (2004) Treatment effectiveness for male adolescent sexual offenders: a meta-analysis and review. J Child Sex Abus 13:281–293

Ware J, Mann RE, Wakeling HC (2009) Group versus individual treatment: what is the best modality for treating sexual offenders? Sex Abus Aust NZ 1:70–79

Weinberger LE, Sreenivasan S, Garrick T, Osran H (2005) The impact of surgical castration on sexual recidivism risk among sexually violent predatory offenders. J Am Acad Psychiatr Law 33:16–36

Weisburd D, Lum CM, Petrosino A (2001) Does research design affect study outcomes in criminal justice? Ann Am Acad Pol Soc Sci 578:50–70

White P, Bradley C, Ferriter M, Hatzipetrou L (2009) Managements for people with disorders of sexual preference and for convicted sexual offenders [Cochrane Review]. The Cochrane Library (4). Oxford: Update Software

Trends

► Juvenile Violence

Triads and Tongs

T. Wing Lo and Sharon Ingrid Kwok
City University of Hong Kong, Kowloon,
Hong Kong

Overview

The ancestor of Chinese triad society is Hung Mun, which was originally a loyal and righteous society. It was a secret society whose members, bound by oaths of blood brotherhood, were pledged to restore the ancient Empire, the Chinese Ming Dynasty, to the throne by overthrowing the foreign conqueror, the Munchow Qing Dynasty. Being a massive and highly coordinated political and military organization, it had a clear structure of commanders, officers, rank and file, and followers. Its members were all brave heroes rather than groups of criminals or vandals. Since China became a Republic in 1911, such loyal and patriotic nature had gradually diminished, leading to the disintegration of Hung Mun into dozens of separate triad societies engaging in criminal activities. The political turmoil in China between the two world wars, including the Japanese invasion and the rise of

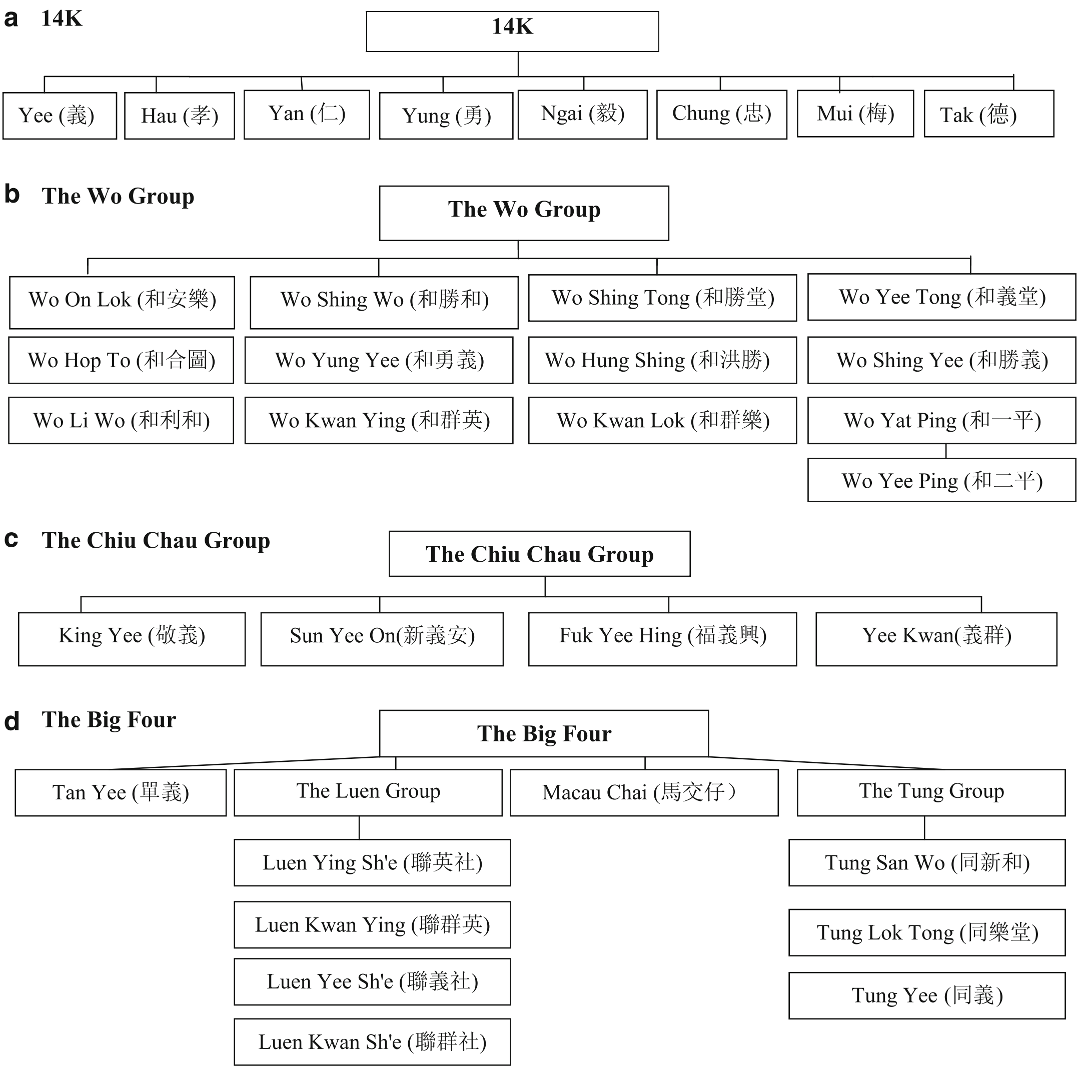
a 14K
14K
Yee (義)
Hau (孝)
Yan (仁)
Yung (勇)
Ngai (毅)
Chung (忠)
Mui (梅)
Tak (德)
b The Wo Group
The Wo Group
Wo On Lok (和安樂)
Wo Shing Wo (和勝和)
Wo Shing Tong (和勝堂)
Wo Yee Tong (和義堂)
Wo Hop To (和合圖)
Wo Yung Yee (和勇義)
Wo Hung Shing (和洪勝)
Wo Shing Yee (和勝義)
Wo Li Wo (和利和)
Wo Kwan Ying (和群英)
Wo Kwan Lok (和群樂)
Wo Yat Ping (和一平)
Wo Yee Ping (和二平)
c The Chiu Chau Group
The Chiu Chau Group
King Yee (敬義)
Sun Yee On(新義安)
Fuk Yee Hing (福義興)
Yee Kwan(義群)
d The Big Four
The Big Four
Tan Yee (單義)
The Luen Group
Macau Chai (馬交仔）
The Tung Group
Luen Ying Sh'e (聯英社)
Luen Kwan Ying (聯群英)
Luen Yee Sh'e (聯義社)
Luen Kwan Sh'e (聯群社)
Tung San Wo (同新和)
Tung Lok Tong (同樂堂)
Tung Yee (同義)

Triads and Tongs, Fig. 1 The four largest triad consortiums in Hong Kong

the Communist Party, forced the triads to retreat into Hong Kong. In 1960, the then Police Commissioner of Hong Kong made such a remark, "Of the approximately 3,000,000 inhabitants of Hong Kong it is estimated that about one in six is a triad member" (Morgan 1960:ix). Hong Kong is undoubtedly the capital of triad societies, which serves as the best venue for the study of triad structure, subculture, and transformation. Today, triads are criminal organizations acting independently of each other and use triad names and rituals for their own benefit. Although they are not as powerful as before, the problem still exists and the heroic figures presented by triad members in movies and cartoons have won the heart of working class youth (Lo 2012a).

Fundamentals

Organizational Structure

The triads in Hong Kong were originally affiliated to four large triad consortiums (see Fig. 1). Take the Wo group, one of the largest and oldest triad consortiums, as an example. It was formed by many triad societies, such as Wo

Triads and Tongs, Fig. 2 Structure of a traditional triad society by rank

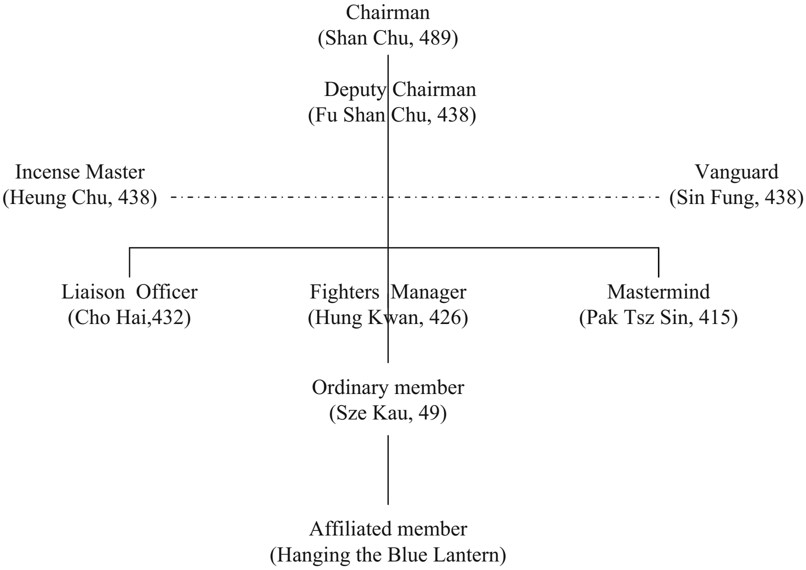

Shing Wo, Wo Shing Yee, and Wo On Lok. The 14K was formed by such branches as Yee, Hau, Yan, and Yung (Cheung 1987). Their internal structure was characterized by a parent and branch relationship. Each triad society was controlled by an administrative headquarters whose authority extended to the branches and sub-branches formed from them. As one of the functions of the headquarters was arbitration when interbranch or inter-society warfare occurred, the office bearers appointed to the headquarters were usually senior officers of various branches whose standing could command a measure of respect and influences onto other members.

Traditionally, triad activities were facilitated by an established organizational structure (Fig. 2). The chairman of a triad society was known as Shan Chu (secret triad code 489), who was assisted by the Deputy Shan Chu (438) and had the authority in making final decisions in all matters relating to the triad. Assisting him were two officers of equal standing, Heung Chu (Incense Master, 438) and Sin Fung (Vanguard, 438), who had seniority in ranking and age. Their duties were to hold initiation and promotion ceremonies as well as recruitment and expansion of the triad (Morgan 1960). On the middle-level were three officers: Hung Kwan (426), Pak Tsz Sin (415), and Cho Hai (432). Hung Kwan in Chinese means “red pole,” signifying a weapon used in fighting. This officer is well-trained in Chinese martial art or tough and brave in fighting. He is the leader of fighters during triad warfare and also responsible for inflicting punishment on traitors. Pak Tsz Sin in Chinese means “white paper fan,” usually used by ancient Chinese intellectuals. This officer is the mastermind of the triad and normally appointed from among the most intellectual and educated members. His role is an adviser, thinker and planner. Cho Hai in Chinese means “straw sandal” which was frequently used by messengers and travelers of ancient time. Being the liaison officer of the triad, he becomes the intermediary between the headquarters and its branches. At the bottom of the hierarchy is Sze Kau (49) or ordinary members (Morgan 1960). Under the Sze Kau, there is a group of affiliated members called Hanging the Blue Lantern. The former has undergone a formal initiation ceremony in order to become a formal triad member (Sze Kau), while the Blue Lantern has not. But both of them are now used to identify generally triad members of the lowest rank.

In postwar decades, triad societies had undergone organizational transformation. The headquarters had no control over branches and procedures on promotion and recruitment had not been adhered to. Incidents of internal conflict always emerged. As time passed, the four large

Triads and Tongs, Fig. 3 Structure of a modern triad society (Chu 2000: 28)

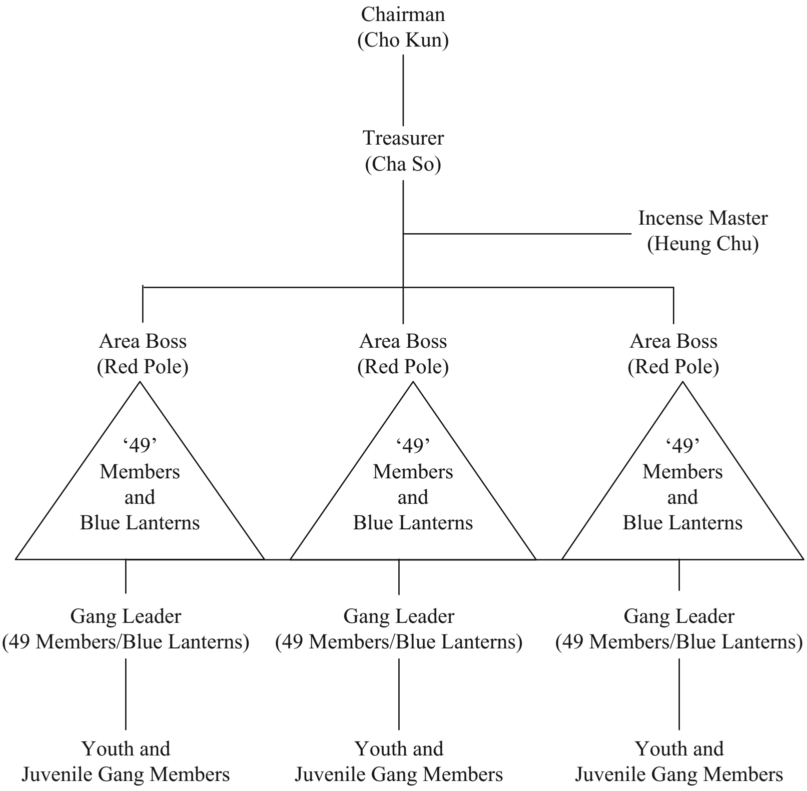
Chairman
(Cho Kun)
Treasurer
(Cha So)
Incense Master
(Heung Chu)
Area Boss
(Red Pole)
Area Boss
(Red Pole)
Area Boss
(Red Pole)
'49' Members and Blue Lanterns
'49' Members and Blue Lanterns
'49' Members and Blue Lanterns
Gang Leader
(49 Members/Blue Lanterns)
Gang Leader
(49 Members/Blue Lanterns)
Gang Leader
(49 Members/Blue Lanterns)
Youth and
Juvenile Gang Members
Youth and
Juvenile Gang Members
Youth and
Juvenile Gang Members

triad consortiums outlined in Fig. 1 dissolved. The triad societies were highly autonomous and used their own triad titles in the underworld. Some large triads have their own central committee to take care of their sub-branches and staff promotion. Some of the triads became inactive or faded out (e.g., Wo Yat Ping, Wo Yee Ping, Wo Kwan Ying), while some new criminal gangs emerged. A study in 2002 (Lo and Kwok 2012) confirmed that the following triads existed in Hong Kong:

- 14K – Hau, Tak, and others
- The Wo Group – Wo On Lok (Shui Fong), Wo Hop To, Wo Shing Wo, Wo Shing Yee, Wo Yee Tong, and Wo Kwan Lok
- The Chiu Chau Group – Sun Yee On, Fuk Yee Hing, and Hwok Lo.
- The Big Four – Tan Yee, Luen Kung Lok, Luen Ying Sh'e, Chuen Yat Chi, Lo Tung, Li Kwan, and Lo Wing
- Gangs emerged in the 1990s – Big Circle Gang and Hunan Gang

Following the disorganization of traditional triad structure, some triads adopted a more flexible hierarchical structure. Some ranks such as Shan Chu, Sin Fung, and Cho Hai became inactive. Pak Tsz Sin has also lost its popularity (Chu 2000). Yu (1998) interviewed triad officers and members in jail and found that 14K had already discarded the rank of Cho Hai. The ranks in existence today are Hung Kwan, Sze Kau, and Hanging the Blue Lantern (Chu 2000). These three ranks are still widely used in modern triads. The headquarters is turned into a central committee and much simplified, composing of Cho Kun (chairman), Cha So (treasurer), and probably Heung Chu (Incense Master), usually elected among triad officers (Fig. 3). Taking the leadership role in the triad, this central committee has the authority to control promotion, enforce discipline, and settle disputes, but it often does not assign illegal activities to its members. In fact, the leaders in the central committee do not get any bonus from the activities that members engage in. Rather, the leaders only obtain money from occasions such as Chinese New Year, initiation ceremonies, and promotions, in the form of “red packets” (lucky money) (Chu 2000). However, the reputation of being a Cho Kun of an active triad in Hong Kong would help solicit both licit and illicit business opportunities in mainland China and overseas countries.

Another part of a triad is the mid-level area-focused sub-branches led by Hung Kwans (Red Pole) and followed by Sze Kau members and Blue Lanterns (Fig. 3). The gangs are organized by area and occasionally if the area is large enough, one or more gangs may occupy the same area but under different area bosses. They engage in criminal activities on the street level and their influences are extended to local youth gangs by means of a “Dai Lo–Lan Tsai” (Big Brother–Follower) relationship (Lo 2012a). This line represents a type of fictive kinship between them. The Dai Lo is normally a Hung Kwan or Sze Kau, who commands a triad located in the area. Under this relationship, a Lan Tsai becomes a Blue Lantern and can seek protection from his Dai Lo. He can claim the name of Dai Lo and the triad he belongs to during a conflict with another gang. On the other hand, the Dai Lo has the right to order the Lan Tsai to perform any tasks for him. The Blue Lantern (affiliated member) can become a Sze Kau (formal member) if he has accumulated sufficient criminal experience and attended an initiation ceremony. Through this “Dai Lo–Lan Tsai” relationship, a process of *triadization* (Lo 2012a) occurs whereby the youth gang members assimilate triad subculture and learn crime techniques from their seniors.

In addition to the above organizational structure, different triads have their own uniqueness. For instance, Sun Yee On, the most cohesive triad in Hong Kong, adopts a hereditary system in leadership succession and is managed by the “Heung” family. The 14K, the largest triad in terms of the number of members, is a loosely organized society, consisting of different street gangs under the control of their own area bosses who cooperate with one another on an ad hoc basis even though they share the same triad name. Wo On Lok, one of the powerful triads in the Wo Group, is still managed by a central committee, which controls promotions, supervises internal discipline, and settles internal and external disputes. It is led by a Cho Kun (chairman) and Cha So (treasurer), who are elected at biennial meetings. The chairman and committee officers are usually the most influential or wealthiest office bearers of the 426 rank (Chu 2000).

Rituals and Subcultural Values

Triad subcultural values include loyalty to the gang, eye for an eye, secrecy, and sworn brotherhood (Booth 1990; Morgan 1960). Triad members are required to regard themselves as one family and blood brothers; they are expected not to do harm to their brothers even under untoward circumstances and to sacrifice themselves for their group (Chin 1990; Lo 1984). There are rules, rituals, oaths, codes of conduct, and chains of command, as well as rigid control over the behavior of triad members, who have to follow triad norms that reflect sworn brotherhood (Lo 1993):

1. Don’t “cal dai” (過底, don’t join other triads).
2. Don’t be a “yee ng tsai” (二五仔, no squealing against the triad).
3. Don’t “po wong” (報皇, don’t report anything to the police).
4. Don’t “ngau dai” (淆底, don’t be scared in committing crime).
5. Don’t “tuk pui chek” (篤背脊, don’t spread disadvantageous information against other triad brothers behind their back).
6. Do obey Dai Lo.
7. Do “bong tall” (幫拖, help triad brothers in fighting).
8. Do “chow kei” (籌旗, give financial help to other triad brothers in times of trouble).
9. Do “fuk cheuk” (覆桌, take revenge if the triad was attacked by rivals).

The essence of sworn brotherhood is also reflected in the triad initiation ceremony. Chu (2000: 32) highlighted 18 steps commonly used in the ceremony: “(1) tapping the new recruit’s left shoulder, (2) dashing the joss stick, (3) covering with the yellow gauze quilt, (4) passing through the Heaven and Earth ring, (5) passing through the fiery pit, (6) passing the two plank bridge, (7) eating the five seasonal fruit, (8) drinking the three river water, (9) smashing the bowl, (10) chopping off the chicken's head, (11) pricking the middle finger of the left hand, (12) drinking the Read Flower wine, (13) eight-step worshipping in front of the altar, (14) old and new members bowing to each other, (15) washing the face of new recruits, (16) teaching triad hand signs, (17) bowing to the officer-bearers, and (18) giving lucky money to protectors." All these rituals carry special meanings. The first eight steps highlight the process of entering into the Hung Mun and worshipping its ancestors, signifying the new recruits are going to join the triad concerned like their ancestors. Steps 9 and 10 emphasize that traitors will be punished with death. Steps 11–14 indicate blood brotherhood in the triad, signifying that after the members drink the wine that mix with the blood dripped from their fingers, they become blood brothers. Steps 15 and 16 signify that the new members will start a new life and learn the triad ways of communication. In the last two steps, 17 and 18, the new members thank their Dai Lo for recruiting and protecting them to round up the initiation ceremony.

In recent decades, there have been radical changes in the ideology and aims from those of traditional triads. Rituals and initiation ceremonies have been greatly simplified or even abandoned in order to avoid police targeting (Bolton et al. 1996). There has been a decrease in the use of hand signs and poems as secret methods of communication (Morgan 1960). The traditional triad principles of brotherhood and loyalty have more or less disappeared or modified, depending largely on particular Dai Lo's interpretations of them. Members' loyalty and righteousness have begun to diminish. The triads are no longer cohesive as the senior triad officers fail to control the middle and lower ranks, and conflicts between branches of the same triad often occur (Chu 2005). Moreover, triad members can cooperate with members of other triads to pursue their own financial goals if opportunities arise, which was previously regarded as a taboo.

There are several reasons for the gradual disorganization. The long chain of command and hierarchical structure increases the risk of exposure of triad activities. As triads are moving towards entrepreneurship, members from different triads, or triad members and ordinary businessmen, could work together to achieve the same moneymaking ends, while triad brotherhood is a less-valued subcultural element nowadays. Benefits and profits become a drive for bonding criminal partners instead of emphasis on brotherhood, which make triad rituals and values less appealing. In addition, the rituals of triads could be used as evidence in the prosecution of triad-related offenses (Kwok and Lo 2013). Consequently, the rituals do not help in smoothing operation but pose a risk to the triad. In view of the government's increased control of triad activities and fast changing social environment, triads have to be flexible in their operational strategies and responsive to market demands and social conditions (Broadhurst and Lee 2009).

Triad Territory and Crime

Territoriality is a key element of triad organized crime. Even today, different triad societies have their own turf or monopolize certain economic sectors, without interfering with one another. They extort or provide protection services to shops, restaurants, hawkers, construction sites, wholesale and retail markets, bars, karaoke, and nightclubs which are located inside their territories (Broadhurst and Lee 2009). Clashes and violence occur when rival triads cross the line and step into their territory (Lo and Kwok 2012). Triads also operate different kinds of lucrative illicit trades, including but not limited to prostitution, pornography, soccer and other forms of gambling, human, cigarette and fuel smuggling, selling of counterfeit products, drug dealing and trafficking, and loan sharking (Broadhurst and Lee 2009; Curtis et al. 2002).

In addition to conventional criminal activities, some triads also engage in other forms of legitimate business, such as monopolizing the home decoration services in some housing estates, film and entertainment industry, car valet services, waste disposal industry, mahjong parlors, and nonfranchised minibus routes (Broadhurst and Lee 2009). Since the 1990s, triads have been found involving in insider trading and financial crime. Lo (2010) reported such a crime committed by the powerful Sun Yee On triad in collaboration with a Hong Kong listed company. The Sun Yee On leader and his associates used the privileged business information they "manufactured" in mainland China to drive the stock market crazy so as to obtain huge financial gain through insider trading and market manipulation.

International Perspectives

Mainlandization and Patriotic Triads

Since the 1990s, Hong Kong has experienced a process of mainlandization of its political, economic, and legal systems (Lo 2012b). Triads have also gone through a similar process of mainlandization as a result of China's economic growth and rising demand for limited goods and services. Mainlandization is defined here as a process of making Hong Kong triads more reliant on mainland China for financial gain through social networking with Chinese officials, enterprises, and criminal syndicates and taking advantage of the rising legitimate and illegitimate business opportunities. In view of increasing licit and illicit opportunities in the mainland (the pull factor), triads have gradually shifted their moneymaking focus on China. The traditional triad rituals are now replaced by a business approach to crime. Chu (2000, 2005) observed that triads have increasingly invested in legitimate businesses and teamed up with entrepreneurs to monopolize the newly developed mainland market. In the illegal side, triads have also forged cooperative relationships with mainland criminal groups, trying to capitalize in the booming underworld (Broadhurst and Lee 2009). They provide their Chinese counterparts with crime techniques and intelligence, in return for cheap labor, sex workers and dangerous drugs to be consumed in Hong Kong.

In 1992, Tao Siju, the Minister of Public Security of China, remarked that some triad members were patriots. Tao even suggested that Deng Xiaoping had made a similar remark back in the 1980s because the triads protected him during his visit to the USA in 1979 (Arsovska and Craig 2006: note 127; Liu 2001). In 1993, Ye Xuanping, the vice-chairman of the National Committee of the Chinese People's Political Consultative Conference, was seen officiating in a mainland film production center owned by the Sun Yee On triad, with two of its leaders, Jimmy and Charles Heung, standing beside Ye (see Picture 1 in Lo 2010). The National Committee is one of China's highest political bodies, and the presence of its vice-chairman in such an event confirmed the "united front" tactic frequently used by the Chinese Communist Party to incorporate opinion deviants into the establishment.

The Chinese national leaders applied the united front tactic on Sun Yee On to serve three political functions (Lo 2010). First, it was to close the Yellow Bird Action in which over 150 prodemocracy leaders and student activists were smuggled out of China with the assistance provided by Hong Kong triads after the Tiananmen crackdown on democratic movement in Beijing in 1989. Second, it was to uphold Hong Kong's law and order in the run-up to 1997 when the people doubted whether "one country, two systems" would work and whether the rule of law would be upheld after China's takeover of Hong Kong in 1997. The Chinese government was also concerned with the threats of the triads to Hong Kong's social order, and thus the united front tactic was applied to recruit them to its side. Third, it was to prevent the infiltration of Taiwanese secret societies (e.g., the Bamboo United) into Hong Kong because of Taiwan's model of *heijin zhengzhi* (or black-gold politics), where criminal gangs (black) join hands with businessmen (gold) to penetrate into the political arena and obtain power through local elections (Chin 2003). If the Chinese leaders aimed to maintain political control in Hong Kong after 1997, they could not allow prodemocratic forces in Hong Kong to join hands with any Taiwanese forces to take part in the elections.

In order to control Hong Kong's political arena and social stability, China had to ensure that the most powerful triads were friends not enemies. Consequently, the triad leaders were included into the establishment by the label of patriotic triads and simultaneously performed some law and order functions for the communist state in exchange for business opportunities. Thus, triads played the roles of both angel and devil (Lo 2010). On the one hand, they served national interest and complemented established political and social structures. On the other hand, triad leaders used this label as a passport to expand its business network with mainland enterprises and criminal groups to further their interests, posing serious threats to the legal order.

Tong

Traditionally, a tong (堂) was a place for Chinese people with the same ethnicity, trade or industry to assemble for social interaction and recreational activities, settle interpersonal disputes, and gain mutual assistance and support from each other. In every village of China, there was a "che tong" (祠堂) which serves the above functions. In Hong Kong, "tong hau" (堂口) is constantly used to describe triad societies, such that "which "tong hau" one belongs to" is used to ask "which triad one is affiliated with." In fact, some triads have the word "tong" in their name, such as Wo Shing Tong, Wo Yee Tong, Tung Yee Tong, and Tung Lok Tong. Chu (2000:19) contended that "numerous triad societies were set up simply because their original members were victimized by triads. In about 1910 the Tung Lok Tong triad society was established by coolies of the government hospitals in the Western District in order to protect the coolies against extortion by members of the Wo group....The Wo Yee Tong triad society was founded in 1940 in Kennedy Town area of Hong Kong Island. It originally consisted of hawkers and stall-holders who banded together in defense against exploitation by triads."

Tongs also exist in Chinatowns of the UK. For instance, the "Zhi Gong Tong Chinese Association (UK)" was found in Liverpool in July 2012, with the following introduction printed on a street board at the entrance of Chinatown: "The Origin of the Association had a strong link with Dr Sun Yat-sen, often referred to as the "father of modern China," it supported the overthrow of the Qing Dynasty in 1911. The association has branches across the continents. The association in the UK was set up to unite overseas Chinese and provide support to each other. It is one of the oldest Chinese community organizations in Liverpool. The association organizes transport for their members to the Everton and Anfield cemeteries twice a year to pay respect to Chinese ancestors, including those who had no family in the UK." The above introduction indicates that tongs were patriotic, welfare and mutual-help organizations that were active in Europe as early as the beginning of the twentieth century.

In the USA, tongs appeared in the Chinatown of San Francisco as early as the 1850s. The US Chinatowns had the characteristics of transitional zone, such as poverty and heterogeneous cultural background among residents. The local governments failed to provide effective social services to assist Chinese immigrants. Due to the lack of formal social services and resources within the towns, the tongs became the only self-help organizations that provided assistance and welfare, such as job referral, housing, networking and entertainment activities, to this group of socially excluded Chinese. They also acted as mediators in individual and group conflicts within the Chinese communities, as well as liaison between local government agencies and Chinese citizens in the constituencies (Chin 1990). Over the years, the power of tongs grew rapidly due to the constant social exclusion of Chinese immigrants and unrestricted recruitment of members. In recent decades, the tongs have been transformed from social service providers and self-defense organizations to illegal service providers. Their victims are mainly Chinese immigrants who have language and cultural barriers, thus being alienated from the mainstream society (Chin 1990, 1996).

Within each tong, there are many factions. Although there is a headquarters above these factions, it has no actual power over them. Most tongs have an elected president, but he only serves as a puppet controlled by the powerful officers and members in the headquarters, which consists of the president, vice-president, secretary, treasurer and auditor, and several public relations officers (Chin 1990). Ordinary members have to pay fees regularly, but they are not involved in any daily affairs and have no decision making power.

It has been argued that Chinese street gangs are affiliated with tongs. They provide protection services for tongs' gambling business (Chin 1990, 1996; Chin et al. 1998; Curtis et al. 2002; Kelly et al. 1993). However, the relationship between tongs and gangs are complicated and indistinguishable. For example, tongs' elders are involved in resolving conflicts between cliques within the gangs. Being the mentors of gang members, they transmit triad subculture and values to them and provide financial support and recreational venues. Their relationship is intertwined as some gang leaders are officially recruited into tongs, thus having dual membership. Some also introduce their followers into tongs. As a result, the elders of tongs can control gang members indirectly through the gang leaders, who serve as a broker between the tong elders and the street soldiers (Chin 1990). In addition, tongs also support street gangs through other means. For example, the tongs' youth activity coordinators serve as the middlemen between tongs and gangs, sponsoring the sports activities of youth members of the tongs, including gang members. Tongs also provide job opportunities for the youth in their gambling houses, and offer financial compensation to gang members or relocate them to other cities when they are in trouble.

Despite such close relationships, mistrust and misunderstanding between the two parties are rampant. Many tongs use the gangs as a tool to commit crime so that they can be insulated from the criminal activities, but they do not truly regard the gangs as part of the organization. On the other hand, some gangs do not want to be controlled by tongs. They do not feel compelled to follow the command of tongs and want to be independent (Chin 1990). Therefore, the relationship between tongs and street gangs is never static and monolithic.

Tongs differ from triads. Triads are illegal organizations, but tongs are legal entities although they consist of both law abiding and law breaking members and provide both licit and illicit services to the Chinese communities. With their original goals of self-protection and mutual assistance, tongs were not established exclusively for criminal activities. As such, law enforcement agencies tend to chase after individual criminals of the tongs rather than the tongs. Therefore, even if the criminals are arrested, the tongs will still survive and recruit new blood to replace the caught members (Zhang 2012).

Triads, Tongs, and Transnational Organized Crime

Before China resumed the sovereignty of Hong Kong in 1997, there was serious concern that the communist takeover would lead to a mass exodus of Hong Kong triads to the West (Kenney and Finckenauer 1995). However, Lo (1995) argued that due to the increase of both licit and illicit business opportunities in mainland China, the possibility of such exodus was slim; rather, it was the internationalization of triad activities that should be of major concern to the western nations. Reports from Southeast Asia, North America, and the UK revealed that triad organized crimes, such as drug trafficking and human smuggling, were on the rise in these jurisdictions (Curtis et al. 2002). Since some individual triad members migrated overseas in fear of the communist rule, they become Hong Kong triads' overseas trustful partners. They possess the necessary skills, experiences and triad values. Therefore, criminal partners at the two ends of the international servicing channel collaborate with each other, sharing the same goal, subculture, and norms.

Zhang and Chin (2003) argue that the traditional structure of triads facilitate them to commit crime in local territories but hinder their involvement in transnational criminal activities. Despite the increasing number of reports indicating triads' or tongs' involvement in transnational organized crime, such as human smuggling and narcotics trafficking (Curtis et al. 2003), the findings remain inconclusive, and the role of triads and tongs in transnational crime is still unclear. There is no credible evidence to prove their collective involvement in narcotic trafficking

(Chin 1996). Empirical research found the involvement of triads and tongs in smuggling is only restricted to individual members and operated in a disorganized basis (Zhang 2007). It was observed that many heroin trafficking and smuggling cases were conducted in dynamic form involving multi-criminal groups, legitimate business, and other social sectors, rather than monopolized by a single triad or tong (Chin and Zhang 2007; Zhang and Chin 2003; Zhang 2007).

Zhang and Chin (2003)'s structural deficiency perspective shed light on the reason why triads and tongs, being territorial-based criminal organizations, fail to expand their operations into the arena of transnational organized crime. Their hierarchical structure, preservation of continuity and group identity, restricted membership, monopolization in activities, and preferring long-term and low-risk benefit from illicit operations do not fit with the structural requirements of transnational organized crime (Zhang 2012). Both smuggling and trafficking markets are fluid and multifaceted due to the extensive movement involved in illicit goods and services. The markets of transnational crime are fragmented, risky, and uncertain because the clientele are scattered over different regions, and the crime is exposed to law enforcement agencies of different nations. Therefore, transnational organized crimes are best operated by loosely organized ad hoc groups rather than hierarchical organizations like the triads and tongs, because small-sized groups could be more adaptable and responsive to the unstable and hostile environment. The risky smuggling and trafficking environment requires limited vertical hierarchy with high level of division of labor, instead of bulky, hierarchical structure with long reporting line.

Moreover, due to the liberalization of the smuggling and narcotic trafficking market, the illicit market is open to anyone with resources, social capital, and access to business opportunities. Consequently, it is difficult for one particular group to dominate the market and the long-term stable benefit is not guaranteed. Therefore, it may not fit with the organizational goal of triads and tongs. In addition, flexibility in recruitment is important for transnational crime operation, as the clientele is limited and the opportunities are not constantly available. The restricted membership system and identity of triads and tongs reduce flexibility in developing new network and grasping new opportunities, making them difficult to compete with freelance participants who now share the illicit transnational market (Zhang and Chin 2003).

Future Directions

Given the rapid economic development and rising interest of overseas investors, triads are taking a stake in the Chinese market. Mainlandization is the characteristic of triad crime in the new century. A possible analogy of mainlandization is a "vacuum cleaner." Once the hover starts, the current is so strong that everything will be sucked into it. Triads are pulled into mainland China because of the lucrative profit and low risk. Economic prosperity in the mainland provides ample opportunities for the triads to generate profit, whereas the risk is low due to the power monopoly and corruptibility of judicial, municipal and police officials, the absence of the rule of law, a weak civic society and censored mass media, and the existence of *guanxi* and protective umbrella where criminal syndicates can bribe officials in exchange for protection in their activities (Zhang and Chin 2008). Such economic and political climate encourages triads to leave their local territories to embrace the "patriotic triad" claim in the mainland (Lo 2010; Lo and Kwok 2012).

The unknowns ahead, however, are whether the triads can secure their position in the mainland. Does mainlandization take place simply in the form of crime displacement? Can triads transplant their illegal operations from Hong Kong to the mainland? Can Hong Kong triads colonize the underworld of the mainland? Will transplantation and colonization (Varese 2006) elicit resistance from mainland criminal syndicates against Hong Kong triads? On the mainland, the Chinese government has been aware of the potential risk of triad displacement or transplantation to the mainland. In order to control this trend, the government amended its Criminal Law in 2002 and Article 294 stipulates that triad members outside

China trying to recruit members or develop a triad branch in China will be sentenced to 3–10 years of imprisonment. The government has paid close scrutiny to the development of "triad-nature" crime and activities, but how effective the law, procuratorate, and public security organs are in controlling triads in the mainland is a major concern. All of the aforementioned questions need to be further addressed and require more research.

Related Entries

- ▶ Careers in Organized Crime
- ▶ Gangs and Social Networks
- ▶ History of Organized Crime in the United States
- ▶ Organized Crime and the Environment
- ▶ Organized Crime, Types of
- ▶ Russian Mafia
- ▶ Social Network Analysis of Organized Criminal Groups
- ▶ Social Network Analysis of Urban Street Gangs

Recommended Reading and References

Arsovska J, Craig M (2006) Honourable behaviour and the conceptualisation of violence in ethnic-based organized crime groups: an examination of the Albanian Kanun and the code of the Chinese triads. Global Crime 7(2):214–246

Bolton K, Hutton C, Ip P (1996) The speech-act offence: claiming and professing membership of a triad society in Hong Kong. Lang Commun 16:263–290

Booth M (1990) The triads: the Chinese criminal fraternity. Grafton Books, London

Broadhurst R, Lee KW (2009) The transformation of triad 'dark societies' in Hong Kong: the impact of law enforcement, socio-economic and political change. Secur Chall 5(4):1–38

Cheung S (1987) The activities of triad societies in Hong Kong. Cosmo Books, Hong Kong

Chin K (1990) Chinese subculture and criminality: non-traditional crime group in America. Greenwood, Westport

Chin K (1996) Chinatown gangs: extortion, enterprise, and ethnicity. Oxford University Press, New York

Chin K (2003) Heijin: organized crime, business, and politics in Taiwan. M.E. Sharpe, NY

Chin K, Zhang SX (2007) The Chinese connection: cross-border drug trafficking between Myanmar and China. Final report submitted to the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice for Grant #2004-IJ-CX-0023. Rutgers University, Newark

Chin K, Zhang SX, Kelly R (1998) Transnational Chinese organized crime activities. Transnatl Organ Crime 4(3/4):127–154

Chu YK (2000) The triads as business. Routledge, London

Chu YK (2005) Hong Kong triads after 1997. Trends Organ Crime 8(3):5–12

Curtis EG, Elan SL, Hudson RA, Kollars NA (2002) Transnational activities of Chinese crime organizations. Trends Organ Crime 7(3):19–59

Curtis GE, Elan S, Hudson R, Kollars N (2003) Transnational activities of Chinese crime organizations. Federal Research Division, Library of Congress, Washington, DC. http://www.loc.gov/rr/frd/pdf-files/ChineseOrgCrime.pdf. Accessed 30 Sept 2012

Dombrink J, Song JHL (1996) Hong Kong after 1997: transnational organized crime in a shrinking world. J Contemp Crim Justice 12(4):329–339

Kelly R, Chin K, Fagan J (1993) Structure, activity, and control of Chinese gangs: law enforcement perspectives. J Contemp Crim Justice 9(3):221–239

Kenney DJ, Finckenauer JO (1995) Organized crime in America. Wadsworth, San Francisco

Kwok SI, Lo TW (2013) Anti-triad legislations in Hong Kong: issues, problems and development. Trends Organ Crime 16:74–94

Liu B (2001) The Hong Kong triad societies before and after the 1997 change-over. Net e-Publishing, Hong Kong

Lo TW (1984) Gang dynamics. Caritas, Hong Kong

Lo TW (1993) Neutralization of group control in youth gangs. Groupwork 6(1):51–63

Lo TW (1995) Patriotic triads and the 1997 exodus, paper presented in the conference on Hong Kong and its Pearl River Delta Hinterland: links to China, links to the world. University of British Columbia, Vancouver, 4–6 May 1995

Lo TW (2010) Beyond social capital: triad organized crime in Hong Kong and China. Br J Criminol 50:851–872

Lo TW (2012a) Triadization of youth gangs in Hong Kong. Br J Criminol 52:556–576

Lo TW (2012b) Resistance to the mainlandization of criminal justice practices: a barrier to the development of restorative justice in Hong Kong. Int J Offender Ther Comp Criminol 56(4):627–645

Lo TW, Kwok SI (2012) Traditional organized crime in the modern world: how triad societies respond to socioeconomic change. In: Siegel D, van de Bunt H (eds) Traditional organized crime in the modern world: responses to Socioeconomic change, vol 11, Studies of organized crime. Springer, New York, pp 67–89

Morgan WP (1960) Triad societies in Hong Kong. Government Printer, Hong Kong

Varese F (2006) How Mafia's migrate: the case of the Ndrangheta in northern Italy. Law Soc 40(2):411–444

Yu KF (1998) The structure and subculture of triad societies in Hong Kong. City University of Hong Kong, Hong Kong

Zhang S (2007) Smuggling and trafficking in human beings: all roads lead to America. Praeger/Greenwood, Westport

Zhang S (2012) China tongs in America: continuity and opportunities. In: Siegel D, van de Bunt H (eds) Traditional organized crime in the modern world: responses to socioeconomic change, vol 11, Studies of organized crime. Springer, New York, pp 109–128

Zhang S, Chin K (2003) The declining significance of triad societies in transnational illegal activities: a structural deficiency perspective. Br J Criminol 43:469–488

Zhang S, Chin K (2008) Snakeheads, mules, and protective umbrellas: a review of current research on Chinese organized crime. Crime Law Soc Change 50:177–195

True Experiments

► Randomized Experiments in Criminology and Criminal Justice

Tuberculosis

► Drug Abuse and Alcohol Dependence Among Inmates

Under 12 Outreach Project (ORP)

► Stop Now and Plan (SNAP®) Model

Understanding Cross-National Variation

Steven F. Messner[1] and Gregory M. Zimmerman[2]
[1]Department of Sociology, Arts, and Sciences, University at Albany, State University of New York, Albany, NY, USA
[2]School of Criminology and Criminal Justice, Northeastern University, Boston, MA, USA

Overview

Cross-national inquiry can make several distinctive contributions to the understanding of criminal violence. One, it permits an assessment of variability in levels and patterns of violent crime. In his classic discussion of the "normality of crime," Durkheim (1964:66) argued that crime is "closely connected with the conditions of all social life," leading him to conclude that a society without crime is inconceivable. The same may very well hold true for violent crime in contemporary nations. Coercive action has intrinsic utility for obtaining compliance from others, and thus a certain degree of violence may be an inevitable feature of complex forms of social organization. Nevertheless, cross-national research elucidates the range of possibilities for minimizing levels of criminal violence.

Two, cross-national inquiry is indispensable for establishing the generality of theories of violence and the need for any scope conditions (Kohn 1987). As Lilly et al. (1989:11) observe, "social context plays a critical role in nourishing certain ways of theorizing about crime." The substantive content of theories of criminal violence, therefore, reflects to a large degree what "makes sense" to criminologists given the specific environments with which they are familiar. Therefore, the causal processes depicted in criminological theories may in fact be limited in scope to a restricted set of sociocultural situations. The application of theories of violence beyond the confines of a single nation provides an opportunity to distinguish between causal processes that are truly general and those that vary by the larger national context.

Three, cross-national inquiry permits an assessment of the role of macrosocial factors in the explanation of criminal violence. Criminological research, especially in the United States, is most commonly conducted at the individual level of analysis. This kind of research is directed towards answering the question: "what is it about individuals that explains their behavior" (Short 1985:53). While such inquiry is without question valuable in criminology, it is also

This essay incorporates and extends material published in Messner (2003) with the permission of the publisher.

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

important to consider how basic features of social organization such as demographic structures, institutional arrangements, and broad cultural orientations help explain violent crime. These kinds of variables are by their very nature properties of large-scale aggregates rather than of individual persons. Moreover, to discover the causal significance of such macrolevel properties, it is necessary to observe variation in violence across aggregate units.

The present essay is restricted to research dealing with nation-states. As Ember and Ember (1995) explained, two general types of comparative research can be distinguished: cross-national and cross-cultural. Cross-national research compares politically organized entities, whereas cross-cultural research is directed towards "societies" in the larger anthropological sense, that is, populations that occupy contiguous territories and share a common language not normally understood by outsiders. In addition, this essay focuses primarily on research that is analytical rather than descriptive and that employs quantitative rather than qualitative techniques. For example, comparative research on genocide, which has adopted primarily qualitative techniques (Straus 2006), is not discussed. The bulk of research on cross-national variation in criminal violence is oriented to testing theoretically derived hypotheses. In recent years, quantitative, theory-testing studies almost always employ multivariate statistics.

Theory

Cross-national research on criminal violence has been informed by a variety of theoretical perspectives. Specific arguments are often complex and multifaceted, which makes it difficult to summarize the theoretical approaches concisely. Nevertheless, previous reviews of the literature have offered typologies of theories that delineate core themes and that overlap to a considerable degree (Neuman and Berger 1988; LaFree 1999; Neapolitan 1997). Building upon these earlier reviews, the various theoretical perspectives can be organized into two general categories reflecting the underlying strategy for theorizing.

One strategy can be labeled macro-sociological theorizing. Theories of this type begin with arguments about properties of large-scale social systems and their interrelationships. These system-level properties are then related to criminal violence (and crime more generally) by means of intervening mechanisms suggested by conventional theories of crime. An alternative strategy begins with theoretical explanations developed to account for crime within nations, at the individual-, group-, or neighborhood-level. The specified causal processes are then linked with national characteristics, and the explanations of crime are essentially extrapolated to apply to national crime rates. For convenience, this approach can be referred to as "extrapolative" theorizing about cross-national variation in violent crime.

Macro-Sociological Theories

One of the most influential macro-sociological theories in cross-national research on crime is modernization theory, which draws liberally on the Durkheimian tradition. Although several prominent proponents of this approach can be identified in the literature (see Eisner 1995, for a review), Louise Shelley (1981) is usually credited with developing the most thorough formulation of the thesis. Modernization theory posits an evolutionary model under which social structures are fundamentally altered, and levels and patterns of criminal activity change, as nations modernize.

The precise nature of the consequences of modernization for crime varies depending on the stage of national development and the offense under consideration. At early stages of urban and industrial growth, traditional social structures are undermined in the growing cities. The concomitant social disorganization, anomie, and weak social control promote increases in property crimes. At the same time the newly arriving migrants from the countryside bring with them traditions of violence associated with rural life, which leads to increases in violent crime. At later stages of development, patterns of crime change. Property crimes continue to rise, becoming the most prominent type of criminal activity. In contrast, the growth in criminal violence subsides as rural migrants become adjusted to urban life. Increasingly, when criminal violence occurs, it does so in the context of the commission of property crimes. In sum, modernization theory posits that changes in levels of criminal violence (and the preponderance of criminal violence relative to property crimes) can be understood with reference to macrosocial processes accompanying societal development.

A second influential argument relevant to the effects of large-scale social change and criminal violence draws upon Elias' (1978) thesis of the civilizing process. According to this view, societal development entails important changes at both the micro- and macrolevel (Eisner 1995; Heiland and Shelley 1991). At the microlevel, personality structures are transformed. Customs and manners are refined, the "threshold of repugnance" against acts of violence is lowered, and action is increasingly guided by internalized self-control. At the macrolevel, the institutionalization of capitalist economies renders interpersonal violence largely dysfunctional because such violence undermines the mutual trust upon which markets are based. In addition, the modern nation-state begins to monopolize power, thereby creating a relatively stable framework for social interaction (Neapolitan 1997:71). The clear implication of the civilization thesis is that levels of criminal violence should decrease as nations become more "civilized."

A third macro-sociological approach to the explanation of cross-national variation in violent crime is dependency/world systems theory (Barak 2001). This perspective is concerned primarily with crime in lesser developed nations. Different variants of the theory have been proposed, but the core claim is that criminogenic conditions within less developed nations arise from the subordinate and dependent position of these nations within the larger world economy. Economic dependency, reflected in conditions such as transnational corporate penetration and trade imbalances, promotes distorted economic and political development. Extreme poverty, inequality, and political oppression characterize subjugated nations. In turn, high levels of criminal violence can be understood as a response to the "frustration and misdirected anger created by deprivation and demoralizing living conditions" (Neapolitan 1997:76). In sum, the distinctive claim associated with dependency theory is that the national conditions that serve as the proximate causes of violent crime must be understood within the context of an inherently exploitive international division of labor.

Finally, recent scholarship has directed attention to processes of capitalist globalization and the emergence of postmodern societies. A common theme in much of this literature is that the expansion of markets and the accompanying triumph of capitalism have unleashed a host of criminogenic forces. These include the spread of rampant individualism (Currie 1991), anomic cultural orientations (Messner and Rosenfeld 2000), the decline of social reforms that previously mitigated the more deleterious consequences of capitalist society (Teeple 1995), and an expansion in the opportunities for organized criminal activities at a transnational level, activities which are often accompanied by violence (Castells 1997).

Extrapolative Theoretical Arguments

Much of the cross-national research on violence does not involve the kind of macro-sociological theorizing just described. Instead, researchers often begin with premises about processes described in conventional criminological theories, and then national characteristics are identified that are suggestive of the operation of these processes. In practice, virtually all of the major sociological theories of crime have been applied in the cross-national literature on criminal violence. This has resulted in a diverse array of hypotheses encompassing a wide range of independent variables. In her research on homicide rates across 18 advanced nations, Gartner (1990) proposed a useful framework for organizing the hypothesized determinants of national crime rates. She observed that explanations of violent crime emphasize some combination of motivations, controls, and/or opportunities for victimization. Applying the extrapolative model of theorizing, national characteristics can be expected to affect levels of criminal violence to the extent that they raise motivations, lower controls, and/or increase opportunities (Gartner 1990:94).

Gartner (1990) proposed that such national characteristics can be grouped into four general types: the material context, the integrative context, the demographic context, and the cultural context. The material context refers to conditions that entail economic deprivation and stress. Variables commonly used to represent the material context include levels of poverty and income inequality and the extensiveness of governmental efforts to alleviate economic stress through welfare policies. The integrative context refers to the strength and depth of social ties that serve to control violent behavior. Common indicators of the integrative context include measures of family structure and racial/ethnic heterogeneity. The demographic context refers to features of population structure. Age distributions, population size and growth, and the sex ratio are major components of the demographic context. The last context – the cultural context – has received the least amount of attention in the research literature because of the difficulty of measuring "culture." In a few studies, researchers have called attention to the potentially important role of general value orientations (Karstedt 1999; Messner 1982; Smith and Kwong Wong 1989), religious values (Groves et al. 1985), and values supportive of legitimate violence (Gartner 1990).

Empirical Evidence

Cross-national research on crime was very limited prior to (around) 1970, but studies have accumulated steadily since that time. The vast bulk of the contemporary work deals with homicide. Researchers generally agree that the comparability of definitions and the quality of official data are greater for homicide than for other offenses. Most of the quantitative studies of homicide share a basic analytic design. The effects on homicide rates of theoretically strategic national characteristics are assessed in multiple regression equations.

Comprehensive reviews of cross-national empirical research on homicide are available in the literature (Neuman and Berger 1988; LaFree and Kick 1986; LaFree 1999; Messner 2003; Neapolitan 1997). Rather than detailing the results of all studies represented in these previous analyses, the following discussion focuses on general patterns and substantive conclusions.

The Material/Economic Context

Most cross-national studies of homicide include some measure of economic development. Common indicators are Gross National (or Domestic) Product per capita, the prevalence of modern media or communications (telephones, radios, newspapers), energy consumption, the sectoral distribution of employment (e.g., agricultural vs. industrial), and composite indexes of such indicators. The results reported in the research are not entirely consistent, but the general picture is one of null or weak negative effects on homicide rates for measures of economic development.

Numerous studies have also focused on the effects of a related concept: economic deprivation. In these analyses, researchers typically distinguish between "absolute deprivation" and "relative deprivation." Indicators of absolute deprivation sometimes include the same measures as those used to represent economic development (e.g., GNP/capita). Infant mortality and unemployment rates have also been examined as indicators of absolute deprivation. Indicators of absolute economic deprivation generally fail to exhibit consistent associations with national homicide rates in multivariate models, with the exception of measures of infant mortality (Pridemore 2011).

There is considerable support, however, for the hypothesis that homicide rates vary with the degree of "relative deprivation" as reflected in levels of income inequality. Reviews of cross-national studies relating measures of income inequality and poverty to homicide (e.g., LaFree 1999; Messner 2003) show that, with few exceptions (e.g., Neumayer 2003; Pridemore 2008, 2011), income inequality exhibits significant, positive effects on homicide rates. This relationship is observed across different measures of income inequality (e.g., the Gini coefficient, quintile shares) and different homicide data sources (e.g., Interpol's *International Crime Statistics*; the World Health Organization's *World Health Statistics*; the United Nation's *World Crime Surveys*). The presence of the positive effect is so pervasive that scholars have called "the scale of income differences between rich and poor… the most well-established environmental determinant of levels of violence" (Wilkinson 2004:1).

Most of these studies are based on cross-sectional assessments of the effect of income inequality on homicide, but a few longitudinal analyses have appeared in the literature. Sophisticated statistical procedures have been applied, such as fixed-effects pooled time series techniques (e.g., Jacobs and Richardson 2008; see also Messner et al. 2010). The evidence generally indicates that the widely observed cross-sectional relationship between income inequality and homicide rates is replicated in longitudinal analyses in multivariate analyses with statistical controls for other relevant predictors.

Research has also investigated the possibility that the relationship between income inequality and violent crime depends on other socioeconomic features of nation-states. Pratt and Godsey (2003) discovered an interaction effect between income inequality and social support: as social support increases, the positive effect of income inequality decreases. In addition, Savolainen (2000:1021) found a negative interaction effect between economic inequality and the strength of the welfare state: "nations that protect their citizens from the vicissitudes of market forces appear to be immune to the homicidal effects of economic inequality."

In a similar vein, a few studies have also examined the direct effects of governmental efforts to provide a safety net via the provision of various social security and welfare benefits. Although the evidence is sparse, the general pattern of findings is consistent. More generous and expansive social welfare policies are associated with lower homicide rates. These effects have been observed for child homicides as well as total homicides and in both cross-sectional and longitudinal analyses (Fiala and LaFree 1988; Gartner 1990; Messner and Rosenfeld 1997).

The Demographic Context

Most multivariate studies of cross-national variation in homicide rates include measures of population structure and distribution. Researchers drawing on modernization theory commonly examine the effects of urbanism/urbanization, often measured as the relative size of the population residing in urban areas. The results of these studies are mixed. The bulk of studies report nonsignificant relationships in multivariate models, although several analyses find modest but significant negative relationships.

For the most part, other demographic factors have not emerged as robust predictors. Population size and density typically exhibit weak or null effects. Similarly, measures of age structure (e.g., the proportion of the population in "high-risk" age brackets) and sex ratio are generally non-related to homicide rates. These results are surprising (even with the potential lack of variation in age structure and sex ratio across aggregate units) given the strong associations between criminal violence and age and sex at the individual level. The one demographic factor that has emerged as a significant predictor with a fair degree of regularity is population growth. Several studies report that rapid population growth is associated with high homicide rates (see Messner 2003).

There is also scant but suggestive evidence that demographic factors may interact with other national characteristics to affect homicide rates. One study (Krahn et al. 1986) reported an interaction effect between population density and income inequality. The positive impact of income inequality on homicide rates is higher in the more densely populated nations. Another study (Pampel and Gartner 1995) found that demographic structure interacts with the degree of "collectivism" in society. Increases in the relative size of the young population tend to elevate homicide rates under conditions of weak collectivism.

The Integrative and Cultural Contexts

As noted above, various criminological theories emphasize the importance of social integration and cultural values in the explanation of levels of criminal violence. One common way of operationalizing social integration is with measures of racial/ethnic, linguistic, and/or religious heterogeneity. Such heterogeneity is generally interpreted as a potential source of value conflict and as an impediment to social cohesion. While some studies have found the expected positive effects of indicators of heterogeneity on homicide rates, null effects have often been observed in the literature. The general consensus is that conventional indicators of heterogeneity are at best only weakly related to homicide rates. Similar to some of the results noted above for demographic variables, however, there is some evidence to suggest that heterogeneity may interact with economic inequality to affect homicide rates. Specifically, economic disadvantage that is rooted in group characteristics has been related to high homicide rates.

A few cross-national studies have examined the relationship between divorce rates and homicide. Divorce is often viewed as both a consequence and a cause of weak integration. Once again, the evidence is mixed, with some studies reporting appreciable positive effects and others reporting null effects.

With respect to the potential impact of the cultural context, the research is quite limited. A few studies have reported positive associations between battle deaths and homicide rates and between the use of capital punishment and homicide rates (Archer and Gartner 1984; Gartner 1990; Landau and Pfeffermann 1988). These findings are consistent with the claim that cultural support for the legitimate use of violence is likely to "spill over" and increase the use of illegitimate forms of violence. Similarly, a limited body of evidence relates gun availability to lethal violence (Killias 1992; Walker 1999). Although gun availability is sometimes interpreted as an "opportunity" factor, it has also been interpreted as an indicator of cultural values conducive to violence.

Researchers have also considered the impact of individualistic values on levels of violence. Competing theoretical arguments have been offered about whether individualism is likely to promote or inhibit criminal violence (Currie 1991; Karstedt 1999). The difficulties associated with the measurement of values preclude definitive conclusions. Nevertheless, there is scant but suggestive evidence pointing to the importance of values of "moral individualism." This value orientation emphasizes a universalistic respect for human rights and independence from strong collective ties (especially kinship and ethnic-group ties). In one of the few efforts to use direct measures of value orientations in cross-national research on criminal violence, Karstedt (1999) reported that a measure of individualism exhibits significant negative effects on homicide rates in multivariate models.

Nonlethal Criminal Violence and "Unofficial" Data Sources

As noted earlier, the majority of quantitative cross-national research on criminal violence has examined homicide rates. Indeed, some have argued that cross-national research should be restricted to homicide because differences in legal definitions and official recording practices render the official data for nonlethal offenses suspect (Neapolitan 1997:38). An important methodological development over the past decades has been the application of survey techniques to overcome limitations associated with official data on nonlethal forms of criminal violence. The most prominent and ambitious of these efforts have been multinational self-report studies of delinquent behavior among youths (Junger-Tas et al. 1994; Gatti et al. 2011) and victimization surveys (van Dijk et al. 2008).

Another advance in this area is the use of multilevel analysis to simultaneously examine individual, structural, and cross-level interaction effects on homicide across nations. For example, Stein (2011) found significant direct effects of individual-level and structural-level measures of opportunity on violent victimization in a multilevel analysis of 45 western industrialized and nonindustrialized countries (see also Stein 2010).

Furthermore, victimization surveys have been used to study the cross-national covariates of intimate partner violence against women. The theoretical and empirical evidence suggests that three factors may decrease national rates of intimate partner violence against women: the empowerment of women through socioeconomic equality, a cultural context (i.e., religion, institutions) that cultivates independence among females, and national globalization (see Ellsberg and Heise 2005; García-Moreno et al. 2005; Johnson et al. 2008; Kaya and Cook 2010).

Given the myriad of methodological difficulties associated with cross-national research on criminal violence, it is not surprising that the literature yields a somewhat cloudy picture. A few generalizations, however, appear warranted. From a purely descriptive standpoint, the evidence demonstrates remarkable cross-national variation in levels of criminal violence, especially in the case of lethal violence. Homicide is a major social problem and a significant cause of mortality in some nations, whereas it is a very rare event in others. Clearly, highly variable levels of violence are compatible with complex forms of social organization.

With respect to the more prominent macrosociological theories reviewed above, the failure to observe appreciable negative effects of standard indicators of economic development on national levels of lethal violence runs counter to simple versions of modernization theory. In particular, the evidence raises questions about the interconnectedness of different dimensions of modernization and about the utility of strict evolutionary models. The consequences of economic development for levels of criminal violence evidently depend to a considerable degree on the larger historical and cultural context of a nation (Shelley 1986). The finding of weak effects of developmental measures in the cross-national literature also suggests that the theory of a civilizing process offers a better explanation of changes in criminal violence associated with the transformation of agrarian, rural communities into industrial and postindustrial nation-states than it does of variation in violence across such nation-states. Considering dependency theories, the finding of positive effects of income inequality on homicide rates is consistent with an underlying premise of the theory. However, the distinctive prediction of this approach is that deprivation within a nation reflects a dependent position in the international division of labor and that this deprivation statistically interprets the effect of dependency on crime. Few quantitative studies actually consider these claims.

Despite ambiguities about the validity of dependency theory, the research does offer clear evidence indicating the importance of the material context. Not only does income inequality emerge as a significant predictor of cross-national variation in homicides rates but indicators of governmental efforts to protect the citizenry from economic hardship also yield reasonably consistent effects. In contrast, standard demographic factors are less important than might be expected in accounting for cross-national variation in criminal violence. This suggests that simply generalizing from individual-level evidence on violence (e.g., age and gender patterns) to the national level through a simple aggregation process is potentially misleading. Evidently, "contextual" effects can modify in an important ways the purely "compositional" effects associated with relatively large numbers of persons with various demographic statuses. Finally, as noted above, only limited evidence is available concerning the role of social integration and culture. However, when indicators of the integrative and cultural context are statistically significant, they are usually in accord with the predictions of criminological theory (Gartner 1990).

Future Directions

The cross-national research on criminal violence has without question advanced markedly over the course of recent decades. Researchers have become increasingly sensitive to issues of data quality, advanced techniques of causal modeling have been applied in multivariate analyses, and efforts to systematize the literature through comprehensive reviews have appeared. Nevertheless, most scholars in the field concede that knowledge about the reasons for cross-national variation in criminal violence is still at a relatively primitive stage.

The compilation of accurate, reliable, and comparable data on criminal violence continues to be a vexing problem. The efforts of the United Nations to standardize data collection and to enhance participation in its crime survey are encouraging. Nevertheless, the collection of international crime statistics presupposes well-developed criminal justice systems at the national level. Some nations simply lack the reporting infrastructure to provide the data requested by multinational organizations.

Victimization and self-report surveys are largely independent of and circumvent official governmental agencies. Yet these surveys are often expensive to conduct, particularly when samples are nationally representative. Thus, to a considerable extent, advancement in cross-national research on criminal violence will continue to be dependent on progress in the development of national criminal justice systems.

There is also a dilemma surrounding priorities for data collection. Researchers desire highly detailed information from large samples of nations. Yet the two objectives of detail and sample size are in some respects incompatible. Given the limited data-collection infrastructures in many nations (noted above), research on large samples will undoubtedly be limited to broad categories of offenses, such as aggregated rates of selected forms of criminal violence.

Cross-national research on criminal violence is also hindered by limitations in the measurement of strategic independent variables. Most studies employ a fairly standard set of predictors regardless of the theoretical objectives of the research. As Neapolitan (1997:88–89) observed, this has led to troublesome ambiguities in the literature. Diverse concepts are often operationalized by the same measure, and the same measure may be interpreted as reflecting different concepts. This practice is understandable – researchers tend to select measures of national characteristics that are readily available in accessible publications, such as those of multinational organizations. The priorities of these organizations and participating governments, however, do not necessarily coincide with those in the scholarly community. In particular, measures of cultural orientations are not well represented in standard data sources. A key challenge for researchers is to initiate efforts to expand the scope of nationwide data collection to encompass indicators of variables relevant to criminological theories beyond the economic and demographic data that have traditionally been compiled under governmental auspices.

Another important methodological challenge for cross-national researchers is to design, implement, and assess interventions. While a body of literature based on correlational designs has accumulated steadily, comparable efforts to conduct field experiments and intervention assessments have yet to emerge on a significant scale (Weiner and Ruback 1995:180–181; see also Eisner et al. 2012). Such research has particular significance for the development of social policies to deal with the problem of violence.

There is also a pressing need for theoretical development in the field. For the most part, researchers have drawn upon existing criminological theories to identify potentially important independent variables, but the nature of the relationships between these variables and levels of violence has not been stated with much precision. For example, with few exceptions, researchers implicitly assume simple additive, linear relationships (see Gartner 1995), or "impose" nonlinearities for methodological reasons. Theoretical rationales for nonlinear relationships warrant greater theoretical attention. Similarly, as noted, several studies have detected interaction effects of various national characteristics on levels of violence. These effects are typically observed in exploratory analyses and are "under theorized." An important task for the future is to incorporate interactions more systematically into theorizing about criminal violence.

More generally, researchers should devote greater efforts to explain the mechanisms that link independent variables and national rates of violence. At best, researchers provide a theoretical rationale for anticipating an association between some national characteristics and rates of criminal violence, but alternative interpretations for the same predictions are usually possible. The precise nature of the hypothetical processes needs to be specified so that distinctive implications of different theoretical arguments can be derived.

A final task worthy of concerted attention in the future is the incorporation of diffusion processes into theories of cross-national violence. There has been a growing interest in spatial analyses of violent crime in the "within-nation" literature over recent years, and sophisticated spatial econometric techniques have been developed to identify diffusion or contagion processes in multivariate models. Moreover, evidence indicates pronounced differences in levels of violence across major regions of the world, which is consistent with diffusion processes. Further theorizing about the possible diffusion of violence is particularly important given the growing recognition that issues of crime and criminal justice increasingly transcend national borders.

Related Entries

- ▶ Anomie and Crime
- ▶ Comparative Courts and Sentencing
- ▶ Comparative Legal Cultures
- ▶ Comparative Political Economy of Punishment
- ▶ Comparing Police Systems Across the World
- ▶ Crimes of Globalization
- ▶ Cross-National Performance in Policing
- ▶ Development of International Criminal Law and Tribunals
- ▶ Effective Community Supervision Programming
- ▶ Extent of Imprisonment: Global View
- ▶ Historical and Comparative Perspectives on Incarceration
- ▶ History of Geographic Criminology Part I: Nineteenth Century
- ▶ History of Geographic Criminology Part II: Twentieth Century
- ▶ Human Smuggling
- ▶ Institutional Anomie Theory
- ▶ Institutional Theories of Punishment
- ▶ International Human Rights Standards and Community Sanctions
- ▶ International Policy Transfer
- ▶ International Responses to Victims in Criminal Justice
- ▶ International Sentencing
- ▶ Policing in Developing Democracies
- ▶ Privatization of Policing in an International Context
- ▶ Transnational Corporate Bribery
- ▶ Transnational Exchange of Forensic Evidence

Recommended Reading and References

Archer D, Gartner R (1984) Violence and crime in cross-national perspective. Yale University Press, New Haven

Barak G (2001) Crime and crime control in an age of globlization: a theoretical discussion. Crit Criminol 10:57–72

Castells M (1997) The power of identity. Blackwell, Oxford, UK

Currie E (1991) Crime in the market society: from bad to worse in the ineties. Dissent (Spring):254–259

Durkheim E (1964 [1895]) The rules of sociological method. Free Press, New York

Eisner M (1995) The effects of economic structures and phases of development on crime. Criminol Res 32:17–43

Eisner M, Malti T, Ribeaud D (2012) Large-scale criminological field experiments. In: Gadd D, Karstedt S, Messner SF (eds) The Sage handbook of criminological research methods. Sage, London, pp 410–424

Elias N (1978) The civilizing process: the history of manners (trans: Jephcott E). Urizen Books, New York

Ellsberg MC, Heise L (2005) Researching violence against women: a practical guide for researchers and activists. World Health Organization, Washington, DC

Ember CR, Ember M (1995) Issues in cross-cultural studies of interpersonal violence. In: Barry Ruback R, Weiner NA (eds) Interpersonal violent behaviors: social and cultural aspects. Springer, New York, pp 25–42

Fiala R, LaFree G (1988) Cross-national determinants of child homicide. Am Sociol Rev 53:432–445

García-Moreno G-M, Jansen HAFM, Watts C, Ellsberg MC, Heise L (2005) WHO multi-country study on women's health and domestic violence against women: initial results on prevalence, health outcomes and women's responses. World Health Organization, Geneva

Gartner R (1990) The victims of homicide: a temporal and cross-national comparison. Am Sociol Rev 55:92–106

Gartner R (1995) Methodological issues in cross-cultural large-survey research on violence. In: Barry Ruback R, Neil Alan W (eds) Interpersonal violent behaviors: social and cultural aspects. Springer, New York, pp 7–24

Gatti U, Haymoz S, Schadee HMA (2011) Deviant youth groups in 30 countries: results from the second international self-report delinquency study. Int Crim Justice Rev 21:208–224

Groves WB, McCleary R, Newman GR (1985) Religion, modernization, and world crime. Comp Soc Res 8:59–78

Heiland H-G, Shelley L (1991) Civilization, modernization and the development of crime and control. In: Heiland H-G, Shelley LI, Katoh K (eds), Crime and control in comparative perspective. Walter de Gruyter, New York, pp 1–19

Jacobs D, Richardson AM (2008) Economic inequality and homicide in the developed nations from 1975 to 1995. Homicide Stud 12:28–45

Johnson J, Ollus N, Nevala S (2008) Violence against women: an international perspective. Springer, New York

Junger-Tas J, Terlouw G-J, Kein M (eds) (1994) Delinquent behavior of young people in the western world – first results of an international self-report delinquency study. Kugler Publications, Amsterdam

Karstedt S (1999) The moral strength of weak ties: a cross-cultural analysis of individualism and violence. In: Fifty-first annual meeting of the American society of criminology, Toronto, 17–20 Nov 1999

Kaya Y, Cook KJ (2010) A cross-national analysis of physical intimate partner violence against women. Int J Comp Sociol 51:423–444

Killias M (1992) Gun ownership, suicide, and homicide: an international perspective. In: Alvazzi del Frate A, Zvekic U, van Dijk J (eds) Understanding crime: experiences of crime and crime control. UNICRI, Rome, pp 289–302

Kohn ML (1987) Cross-national research as an analytic strategy – American Sociological Association, 1987 presidential address. Am Sociol Rev 52(December):713–731

Krahn H, Hartnagel TF, Gartrell JW (1986) Income inequality and homicide rates: cross-national data and criminological theories. Criminology 24:269–295.

LaFree G (1999) A summary and review of cross-national comparative studies of homicide. In: Dwayne Smith M, Zahn MA (eds) Homicide: a sourcebook of social research. Sage, Thousand Oaks, pp 125–145

LaFree G, Kick EL (1986) Cross-national effects of developmental, distributional, and demographic variables on crime: a review and analysis. Int Ann Criminol 24:213–235

Landau SF, Pfeffermann D (1988) A time series analysis of violent crime and its relation to prolonged states of warfare: the Israeli case. Criminology 26:489–504

Lilly JR, Cullen FT, Ball RA (1989) Criminological theory: context and consequences. Sage, Newbury Park

Messner SF (1982) Societal development, social equality, and homicide: a cross-national test of a Durkheimian model. Social Forces 61:225–240

Messner SF (2000) Market dominance, crime, and globalization. In: Karstedt S, Bussman K-D (eds) Social dynamics of crime and control: new theories for a world in transition. Hart Publishing, Oxford

Messner SF (2003) Understanding cross-national variation in criminal violence. In: Heitmeyer W, Hagan J (eds) International handbook of violence research. Kluwer, Dordrecht, pp 701–716

Messner SF, Rosenfeld R (1997) Political restraint of the market and levels of criminal homicide: a cross-national application of institutional-anomie theory. Soc Forces 75:1393–1416

Messner SF, Raffalovich LE, Sutton GM (2010) Poverty, infant mortality, and homicide rates in cross-national perspective: assessments of criterion and construct validity. Criminology 48:801–830

Neapolitan JL (1997) Cross-national crime: a research review and sourcebook. Greenwood Press, Westport

Neuman WL, Berger RJ (1988) Competing perspectives on cross-national crime: an evaluation of theory and evidence. Sociol Q 29:281–313

Neumayer E (2003) Good policy can lower violent crime: evidence from fixed effects estimation in a cross-national panel of homicide rates, 1980–97. J Peace Res 40:619–640

Pampel FC, Gartner R (1995) Age structure, socio-political institutions, and national homicide rates. Eur Sociol Rev 11:243–260

Pratt TC, Godsey TW (2003) Social support, inequality, and homicide: a cross-national test of an integrated theoretical model. Criminology 41:611–643

Pridemore WA (2008) A methodological addition to the cross-national empirical literature on social structure and homicide: a first test of the poverty-homicide thesis. Criminology 46:133–154

Pridemore WA (2011) Poverty matters: a reassessment of the inequality-homicide relationship in cross-national studies. Br J Criminol 51:739–772

Savolainen J (2000) Inequality, welfare state, and homicide: further support for the institutional anomie theory. Criminology 38:1021–1042

Shelley LI (1981) Crime and modernization: the impact of industrialization and urbanization on crime. Southern Illinois University Press, Carbondale

Shelley LI (1986) Crime and modernization reexamined. Annales Internationales de Criminologie 24:7–21

Short JF Jr (1985) The level of explanation problem in criminology. In: Meier RM (ed) Theoretical methods in criminology. Sage, Beverly Hills, pp 51–72

Smith E, Kwong Wong S (1989) Durkheim, individualism and homicide rates re-examined. Sociol Spectr 9:269–283

Stein RE (2010) The utility of country structure: a cross-national multilevel analysis of property and violent victimization. Int Crim Justice Rev 20:35–55

Stein RE (2011) The contextual variation of routine activities: a comparitive analysis of assault victimization. Int J Hum Soc Sci 1:11–24

Straus S (2006) Beyond Rwanda: new directions in comparative genocide research. Invited Presentation, Department of Political Science, Yale University

Teeple G (1995) Globalization and the decline of social reform. Humanities Press, Atlantic Highlands

van Dijk JJM, Kesteren JN, Smit P (2008) Criminal victimisation in international perspective: key findings from the 2004-2005 ICVS and EU ICS. Boom Legal Publishers, The Hague

Walker J (1999) Firearm abuse and regulation. In: Newman G (ed) Global report on crime and justice. Oxford University Press, New York, pp 151–169

Weiner NA, Barry Ruback R (1995) Inquiry through a comparative lens: unraveling the social and cultural aspects of interpersonal violent behaviors. In: Barry Ruback R, Weiner NA (eds) Interpersonal violent behaviors: social and cultural aspects. Springer, New York, pp 171–182

Wilkinson R (2004) Why is violence more common where inequality is greater? Ann N Y Acad Sci 1036:1–12

Understanding Homicide and Aggravated Assault

Richard Rosenfeld
University of Missouri-St. Louis, St. Louis,
MO, USA

Overview

This entry considers the cultural meanings, empirical patterns, theoretical explanations, and social responses connected with homicide and aggravated assault. The paper addresses:

- Homicide and aggravated assault trends in the United States over the past several decades
- Cross-national comparisons of homicide and assault
- Characteristics of victims, perpetrators, and incidents
- The relationship between victims and perpetrators
- Theories of violence
- Policies to reduce violent crime

The main empirical patterns in homicide and serious assault are easily summarized, less easily explained. Rates of homicide and aggravated assault have exhibited substantial volatility in the United States, rising to peak levels by the early 1990s and then falling over the next decade to 30-year lows. Cross-national comparisons reveal strikingly higher rates of homicide, but not assault, in the United States than other developed nations. In all nations where they have been recorded, rates of homicide and serious assault are higher among men and the young than women and older persons. In the United States, the prevalence of homicide and serious assault is greater among African-Americans and Hispanics than non-Hispanic whites. In all groups, firearm use greatly increases the risk that a violent encounter results in the victim's death. Large fractions of victims are acquainted with the perpetrator in homicide and aggravated assault, sometimes intimately. Women are more likely than men to be killed or injured in attacks by an intimate partner. Rates of intimate partner violence have fallen in recent years, but the decrease differs by sex and lethality. Greater reductions in assault victimization have occurred among women than men, and homicide victimization has decreased more among men than women.

Although homicide and serious assault have declined in the United States, the reasons why remain something of a mystery. Mass imprisonment is one contributor to the crime decline, but its other social consequences are troubling, and even its crime-reduction benefits may dissipate over time. The same sense of political purpose, widespread mobilization, and forceful leadership that ushered in the era of mass imprisonment will be required to bring about needed reforms under the banner of "prisoner reentry," the reintegration of released prisoners into the social and political mainstream. A surprising degree of political support for prisoner reentry currently exists, but that is likely to last only until violent crime rates begin to rise again.

Introduction

Homicide and aggravated assault are forms of violent crime. A “crime” is a violation of criminal law, but what is “violence”? That turns out to be a difficult question. Should the definition of violence be limited to acts causing physical harm? What about threats of harm? Can mere words or symbols constitute violence? Is defacing religious property or burning a cross on someone’s lawn a violent act?

Social scientists do not agree on a single or unified definition of violence. Psychologists generally prefer broad definitions that include behaviors producing emotional harm. The Committee on Family Violence of the National Institute of Mental Health, for example, offers a definition that not only includes acts threatening or inflicting physical harm but those resulting in “restraint of normal activities or freedom, and denial of access to resources” (quoted in Crowell and Burgess 1996, p. 10). By this definition, prisons and poverty could be considered forms of violence.

Criminologists tend to favor narrower definitions of violence focusing on physical harm or threats. Many, but not all, criminologists accept the definition provided by an influential National Research Council study of violence as “behaviors by individuals that intentionally threaten, attempt, or inflict physical harm on others” (Reiss and Roth 1993, p. 2). This definition includes a diverse assortment of behaviors, including homicide, assault, rape, torture, capital punishment, and boxing. But it excludes many acts that are encompassed by other, equally reasonable, definitions. How one chooses to define violence, as we shall see, prefigures the types of behavior that are counted as violence, the levels of violence observed across place and time, the theories that “make sense” of violent behavior, and the social response to violence.

Regardless of the definition, not all violence is criminal violence. Violence is made a crime through the *criminalization process*, which entails the selective application of the criminal law to violent acts. Criminal law, and law itself, are not universal. They have not always existed, and they do not impose uniform standards or sanctions on behavior, across societies or in the same society over time. It is a criminological truism that acts legally prohibited at one time or place may not be at others. A common example is prohibition of the manufacture, sale, and distribution of alcohol in the United States during the 1920s. But this maxim applies to violent behavior and not just so-called victimless crimes. Until late in the twentieth century in most US states, for example, a man could not be criminally charged with raping his wife; now all states have laws against marital rape (see Russell 1990).

Patterns of Criminal Homicide and Assault

The foregoing discussion implies that how one chooses to define violence and the cultural meanings attached to particular instances of violent behavior will affect how violence is measured, the levels of violence observed across places at any given time, and trends in violence over time. What is true for violent behavior generally is also true for homicide and aggravated assault. Homicide is the willful killing of one human being by another. Some homicides are lawful, such as those committed in warfare or as legal punishment for crime, and some are criminal. Aggravated assault is an attack by one or more persons on another with the purpose of inflicting serious bodily injury, including threats and attempts, or an assault committed with a dangerous weapon – an “aggravating” condition (Federal Bureau of Investigation 2006). As with homicide, the behavior itself does not distinguish a criminal from noncriminal assault. The sport of boxing is a legally regulated form of assault.

In the United States, three sources of data can be used to track levels and trends in criminal

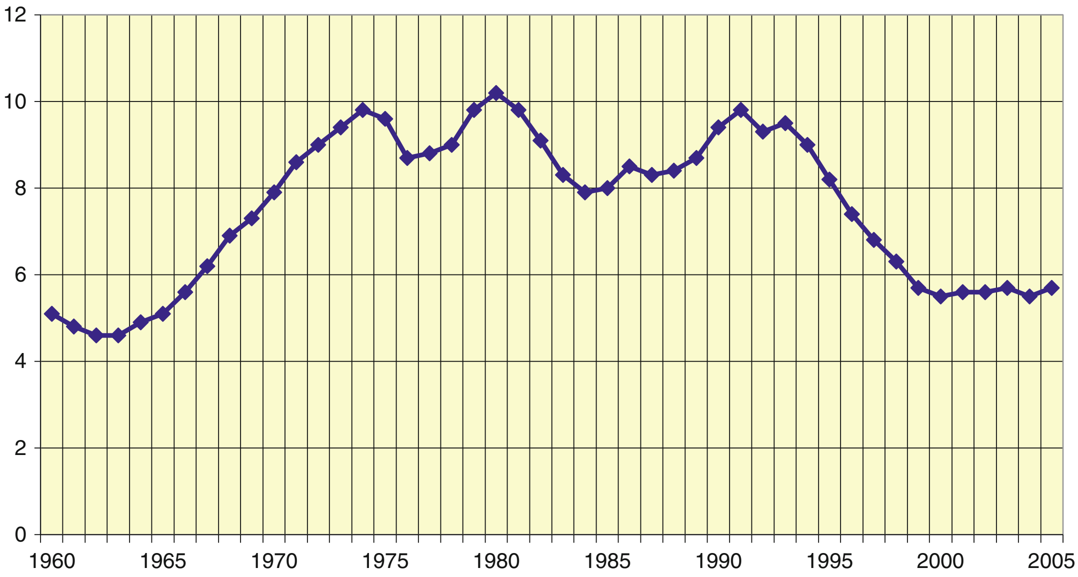

Understanding Homicide and Aggravated Assault, Fig. 1 Homicides per 100,000 population in the United States, 1960–2005. Source: Uniform Crime Reports

homicide and assault: the FBI's Uniform Crime Reports (UCR), the Bureau of Justice Statistics' National Crime Victimization Survey (NCVS), and the National Center for Injury Prevention and Control's (NCIPC) mortality and intentional injury data series. The UCR data are based on offenses reported to and recorded by police departments. The NCVS data are from representative surveys of the population age 12 and over. The NCIPC data are from hospital and death records. The three data sources complement one another in several ways. For example, the NCVS picks up crimes that are not reported to the police, and the UCR and NCIPC count homicides, which are not included in victim surveys. The NCIPC mortality series includes the nearly 3,000 homicides resulting from the 2001 terrorist attacks, whereas the FBI omitted these victims from its 2001 homicide statistics. Finally, the three series categorize rapes and other sexual assaults separately from other assaults.

Recent Trends in the United States

Criminal homicide and aggravated assault rates have exhibited marked fluctuation over the past several decades in the United States. UCR homicide rates between 1960 and 2005 are displayed in Fig. 1. Homicide rates were relatively stationary at 4–5 homicides per 100,000 population during the 1950s and early 1960s and then rose substantially over the next several years, peaking at just over 10 per 100,000 in 1980. They cycled between 8 and 10 per 100,000 until the early 1990s when they began a nearly decadelong drop, what one analyst has called the “Great American Crime Decline” (Zimring 2007). By 2000, US homicide rates had returned to levels not seen since the mid-1960s.

Aggravated assaults are measured in both the UCR and the NCVS, using virtually identical definitions. But the two series show very different patterns of change since 1973, the first year of the NCVS (see Fig. 2). Although both UCR and NCVS aggravated assault rates have fallen since the mid-1990s, the UCR series rose almost continuously before then, while the NCVS series decreased from 1973 until the early 1990s, increased for a few years, and then began to fall again.

Some of the divergence between the two series could reflect a growing willingness by victims to report crimes to the police. All of the increase

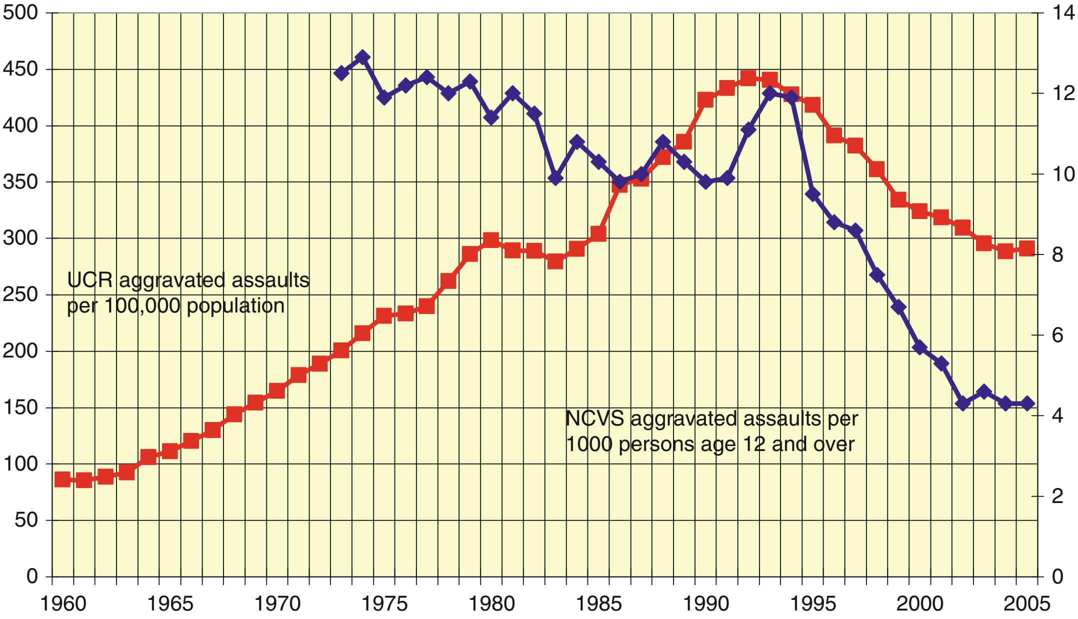

Understanding Homicide and Aggravated Assault, Fig. 2 UCR and NCVS aggravated assault rates. Source: Uniform Crime Reports, National Crime Victimization Survey. Note: NCVS rates prior to 1993 adjusted to reflect 1992 NCVS redesign

in reporting to the police, however, occurred well after the NCVS and UCR aggravated assault trends began to converge in the 1990s and, therefore, cannot explain the divergent trends of the earlier years (see Catalano 2006a).

The contrasting trends prior to the 1990s more likely reflect better police recording of crimes with the advent of computerized records and 911 emergency call systems, and police “upgrading” of many assaults as aggravated that in the past had been considered less serious crimes. The latter interpretation is supported by evidence showing much stronger correspondence between the UCR and NCVS time trends for assaults committed with firearms than those without a firearm. The police are unlikely to record firearm incidents as simple assaults, whereas they have much greater discretion in classifying incidents without firearms or very serious injury. Beginning in the 1970s, the women’s movement and domestic violence advocates began to press police and prosecutors to treat assaults more seriously, especially those with female victims, and those pressures contributed to the increase in aggravated assaults reflected in the UCR data series. Because the NCVS is based on standardized screening questions to determine criminal victimization and lacks the police recording “filter” of the UCR, the NCVS aggravated assault series is much less likely to be affected by such changes (see Rosenfeld 2007).

Changes in police recording of assaults provide an excellent illustration of how the cultural meanings of violence are redefined by social movements and affect the definition and measurement of violent crime. Assaults that the police, reflecting commonly held views, once had classified as misdemeanor offenses, or may not have recorded as crimes at all, were upgraded in seriousness with the increasing social status of victims. The rights and prerogatives of crime victims have grown during the past 30 years, with victims of so-called domestic violence leading the way. Mandatory and pro-arrest policies have reversed the legal protection once granted to violence occurring in the home. But, as we shall see, the effects of such policies on the incidence of domestic violence are more difficult to discern.

Cross-National Comparisons

Comparative assessments of changes in homicide or other crime rates over time are important because they can reveal limitations in explanations of crime trends that are restricted to a single nation, as is true of several leading explanations of the recent violent crime decline in the United States (Zimring 2007). Like those in the United States, the homicide rates of many developed nations rose during the 1970s and 1980s and have fallen since then, but there are plenty of exceptions. As an illustration, Figure 3 displays the homicide rates for Japan, Canada, and the United States between 1960 and 2000. Japan's homicide rate dropped steadily over the 40-year period. By contrast, the Canadian trend mirrors the increases and declines in the United States, as shown by the best-fitting curves superimposed on the two nations' homicide time series. Explanations of recent US crime trends that emphasize the importance of prison expansion and urban crack markets leave open the question of why Canada should have experienced much the same trends, without comparable changes in imprisonment or drug markets (Zimring 2007).

Explanations of cross-national differences in homicide also must contend with the substantial difference in homicide *levels* between the United States and other developed nations, regardless of the time period. Even after dropping by 40 % during the 1990s, the US homicide rate greatly exceeds those of other developed nations. Figure 4 displays the homicide rates of 14 highly industrialized, democratic nations in 2000. With six homicides per 100,000 population, the United States is the clear outlier in the group. The US homicide rate is three times those of Australia and Canada, its nearest competitors, and over six times the average rate of the remaining 11 nations.

Some analysts attribute the large difference in homicide rates between the United States and other developed nations to the widespread availability of firearms in the United States (Zimring and Hawkins 1997). Others point to the fact that the rate of homicides committed without guns is also greater in the United States than other

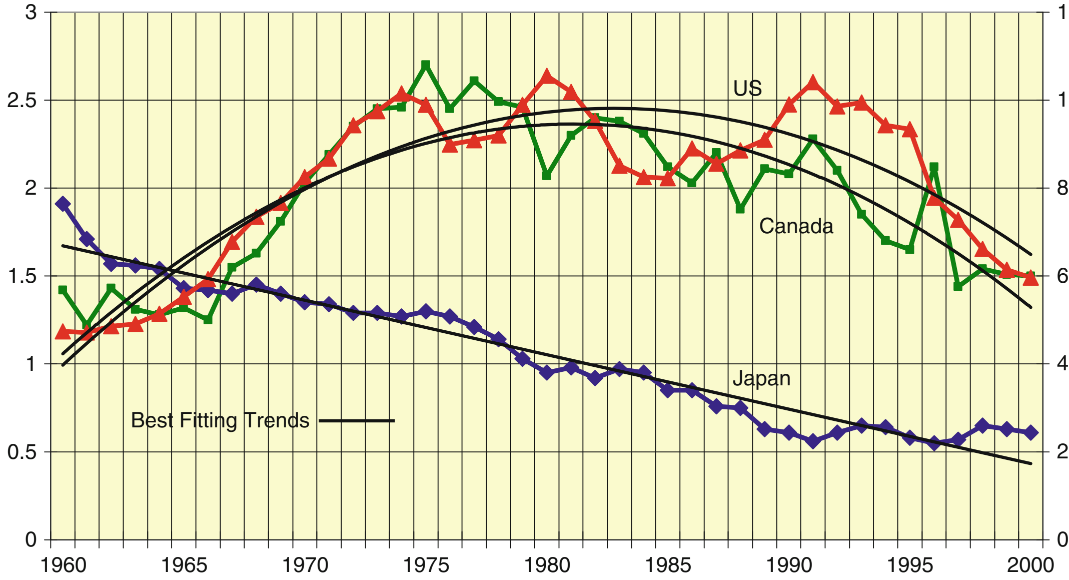

Understanding Homicide and Aggravated Assault, Fig. 3 Homicides per 100,000 population in Japan, Canada, and the United States, 1960–2000. Source: World Health Organization, compiled by Gary LaFree

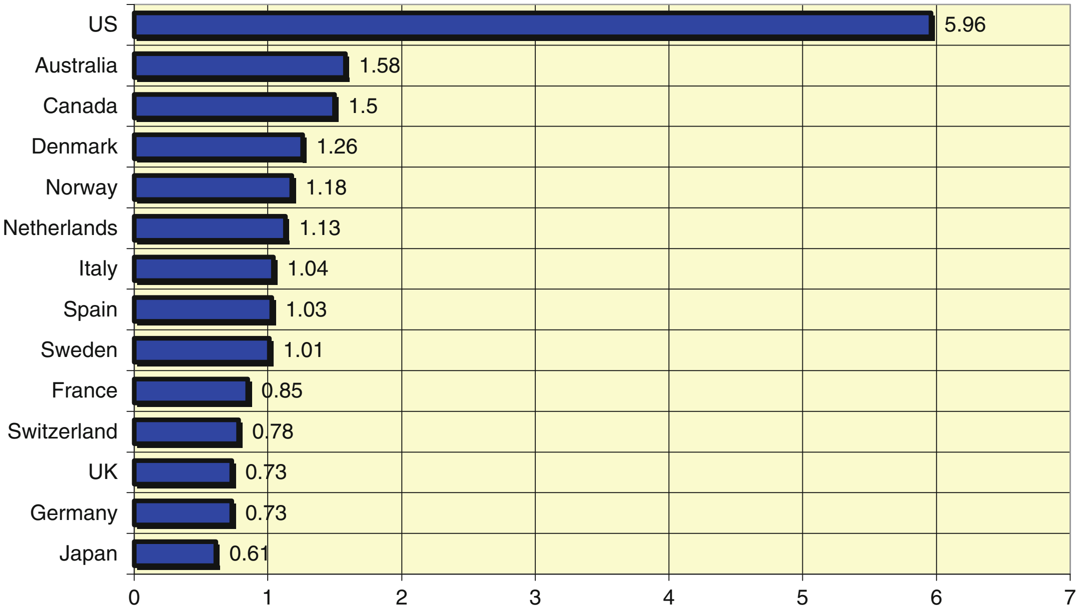

Understanding Homicide and Aggravated Assault, Fig. 4 Homicides per 100,000 population in 14 nations, 2000. Source: World Health Organization, compiled by Gary LaFree

developed nations and call attention to a history of racial oppression, economic inequality, and lack of protection from market forces, in addition to the availability of guns, as reasons for the exceptionally high levels of criminal violence in the United States (Beeghley 2003; Messner and Rosenfeld 2007).

But are rates of "criminal violence" greater in the United States than other developed nations, or is American exceptionalism limited to homicide, as Zimring and Hawkins (1997) maintain? The prevalence of nonlethal assault does not appear to be greater in the United States than other developed countries. Figure 5 displays the percentage of respondents to the International Crime Survey (ICS) in 10 nations who reported that they were victims of one or more assaults in 1999. The United States is far from the top of the list, ranking just below Denmark and Sweden and well below Australia and the United Kingdom. This finding favors causal accounts able to explain why the United States dominates international comparisons of homicide but not other forms of criminal violence.

Characteristics of Victims and Firearm Use

Nearly everywhere they have been assessed, rates of aggravated assault and homicide are higher among males and the young than females and older persons. In the United States, serious interpersonal violence is also more common among blacks and Hispanics than non-Hispanic whites. Table 1 presents 3-year average rates of aggravated assault and homicide in the USA by sex, race, ethnicity (Hispanic and Non-Hispanic), and age. The data represent rates of victimization; offending rates exhibit the same demographic patterns, except that offenders tend to be a few years younger than their victims.

Looking first at rates of aggravated assault, we see that they are 78 % greater among males and 58 % greater among blacks than females and whites, respectively. Aggravated assault is somewhat more common among Hispanics than non-Hispanics, but this difference is smaller than the sex and race differences. Age-specific rates of aggravated assault exhibit a distinctive,

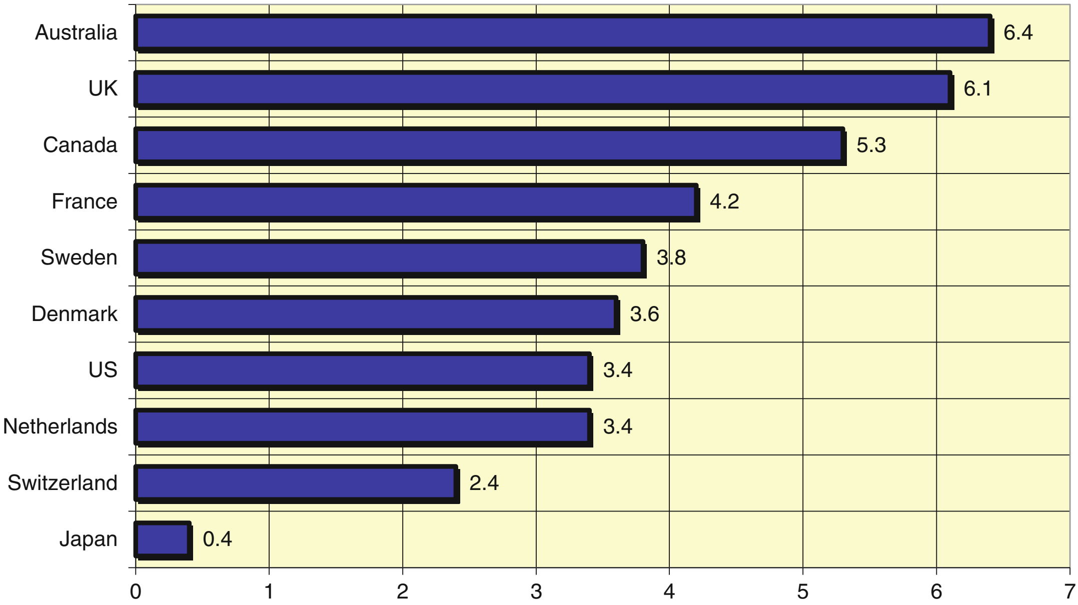

Understanding Homicide and Aggravated Assault, Fig. 5 Percentage of respondents reporting assault victimization, 1999. Source: International Crime Victimization Survey

curvilinear pattern, rising from early to late adolescence and falling steadily thereafter. With a victimization rate only one-tenth of the average rate of 440 per 100,000, persons over the age of 65 are particularly unlikely to be victims of serious assault. This "age-crime curve" characterizes many common criminal offenses.

Firearms are used in about one-quarter of aggravated assaults victims report to the NCVS. Recall that use of a dangerous weapon is one of the "aggravating" conditions that distinguish aggravated from simple assaults. The other condition is serious bodily injury. It turns out that aggravated assaults involving firearms are much less likely to produce injury than those committed with other weapons or none at all. But, if the presence of a firearm reduces the chance of nonfatal injury, it increases the chance that the victim will be killed, thereby converting an aggravated assault into a homicide (Kleck and McElrath 1991).

The homicide rate for persons 12 and older in the United States is 6.8 per 100,000, meaning that the average American over the age of 11 has a one-in-14,706 chance of becoming a homicide victim. The sex, race, ethnic, and age patterns in homicide are roughly similar to those in aggravated assault, although the group differences are larger. Males are about four times more likely than females to be the victim of a homicide. The homicide rate among blacks is more than six times that among whites, and the risk of homicide is about 60 % higher among Hispanics than others. Like aggravated assaults, homicide rates rise during adolescence, but peak a few years later in the early 20s before falling through the remainder of the life course, as shown in Fig. 6. Finally, in contrast to aggravated assault, about two-thirds of homicides are committed with a firearm, so that the firearm homicide rate in the United States is more than twice that of homicides committed without firearms.

The third column in Table 1 presents the "lethality ratio," or the number of aggravated assault victims for every homicide victim, by characteristics of the victim and whether or not the victim was attacked or threatened by an offender with a firearm.

The lethality ratio is a rough approximation of the degree to which death results from the serious violence a group experiences. A full accounting

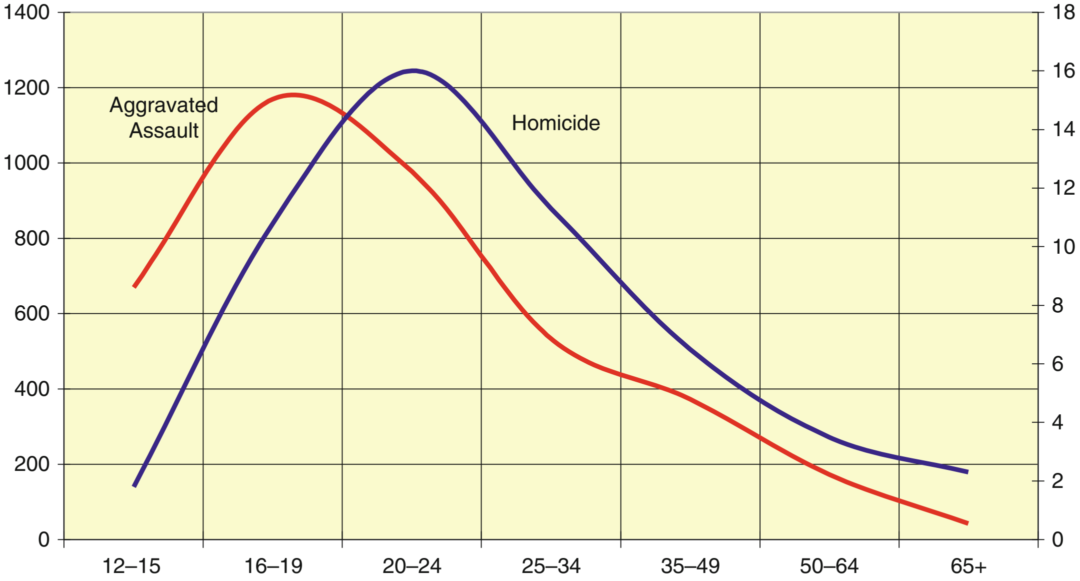

Understanding Homicide and Aggravated Assault, Fig. 6 Age-specific aggravated assault and homicide victimizations per 100,000 population, 2002–2004. Source: National Crime Victimization Survey, National Center for Injury Prevention and Control

of the lethal outcomes of violent crime would have to include other "precipitating" offenses, such as robbery and rape, in addition to aggravated assault, but the group differences shown in the table would not be altered substantially.

For the population over the age of 11, the average ratio is 65 aggravated assaults for every homicide. Some groups – teenagers, whites, and women – have ratios that are much higher than average, meaning that the serious violence they experience is less likely to result in death than is true for groups with lower-than-average ratios, such as blacks and, interestingly, senior citizens. Persons age 65 and over have a lethality ratio of 19 aggravated assaults for each homicide, the lowest ratio shown. It is tempting to attribute this result to the frailty of older persons, which would transform more of their serious assaults into homicides, but their comparatively low homicide rate of 2.3 per 100,000 suggests that is not the full story. Rather, older persons are much less likely to experience interpersonal violence, serious or not, than other age groups. Even their level of simple assault (assaults without a dangerous weapon or serious injury) is far lower than that of younger persons. But when they are attacked, they are more likely to be killed.

The group with the largest lethality ratio is near the other end of the age spectrum. Adolescents between 12 and 15 years old experience 372 aggravated assaults for each homicide. Teenagers, who are said to have replaced fists and knives with guns when they fight, remain far less likely to die in violent encounters than their parents or grandparents.

Finally, attacks involving firearms produce far more deaths per incident than those committed with a knife or club or no weapon. There is no question that firearms are lethal weapons, and many analysts argue that decreases in the availability of firearms would result in fewer homicides. But critics of this view maintain that strict controls on firearm availability would deprive victims of needed protection and even increase violent crime rates, if offenders assume their victims are unarmed. A comprehensive assessment by the National Research Council finds fault with the methods and data used in studies of the "more guns, less crime" hypothesis, regardless of their results (Wellford et al. 2004). The relationship between firearms and violence

Understanding Homicide and Aggravated Assault, Table 1 Characteristics of aggravated assault and homicide victims age 12 and older and incidents in the United States, 2002–2004

	Assaults[a]	Homicides[a]	Lethality ratio[b]
Sex			
Male	563	11.1	50.7
Female	317	2.8	113.2
Race			
White	410	4.1	100
Black	647	26.2	24.7
Ethnicity[c]			
Hispanic	463	10.1	45.87
Non-Hisp.	437	6.3	69.4
Age			
12–15	670	1.8	372.2
16–19	1,170	10.8	108.3
20–24	977	16	61.1
25–34	533	11.3	47.2
35–49	373	6.5	57.4
50–64	173	3.5	49.4
65+	43	2.3	18.7
Weapon			
Firearm	106	4.8	22.1
Non-Fire.	334	2.1	159
Total	440	6.8	64.7

Source: National Crime Victimization Survey, National Center for Injury Prevention and Control

[a]Rate per 100,000 population

[b]Number of aggravated assaults for every homicide

[c]Hispanics may be of any race

remains a contentious research and public issue in the United States, which is not surprising in a nation with persisting traditions of popular justice and self-protection, high homicide rates, and the largest private arsenal in the world.

Victim-Offender Relationship

Victims of homicide and serious assault are often acquainted with the offender, sometimes intimately. According to the NCVS, about 47 % of aggravated assault victims in 2004 knew their attacker. It is more difficult to determine the proportion of homicide victims who knew their killer, both because the victim is unable to say and because an offender is arrested in only about 6-in-10 homicides committed in the United States (Federal Bureau of Investigation 2006). The FBI's Supplementary Homicide Reports (SHR), a detailed compilation of homicide incidents, victims, and known offenders, show that 43 % of homicide victims in 2004 were acquainted with the offender, 13 % were strangers, and the victim-offender relationship could not be determined in the remaining 44 % of cases. Women and younger victims are more likely to know their attackers than men and older victims. The same patterns emerge in violence among family members and other intimates. About 16 % of aggravated assault victims and 12 % of homicide victims in 2004 were attacked by a family member, including spouses or ex-spouses. So-called intimate partner violence has received widespread attention by researchers, victims' advocates, and policymakers in the United States. It is worth taking a closer look at patterns in intimate violence.

Intimate Partner Violence

Two facts stand out in patterns of intimate partner violence: (1) Women are more likely than men to be victims of serious intimate partner violence. (2) Levels of intimate partner violence have declined markedly over time. Both of these patterns are revealed in Figs. 7 and 8, which show US trends in nonfatal and fatal intimate partner violence, respectively, by sex of the victim.

Both figures show higher levels of victimization among female than male intimate partners and decreases in victimization over time, but the rate of decline differs for men and women by type of victimization. Continuous data from the NCVS on nonfatal intimate partner violence are available only since 1993. They show a greater decline in the nonfatal victimization of women than men. By contrast, the drop in male victims of intimate homicide over the past 30 years is steeper than the drop in female victims, which only began in the early 1990s. These contrasting declines are, at first glance, puzzling.

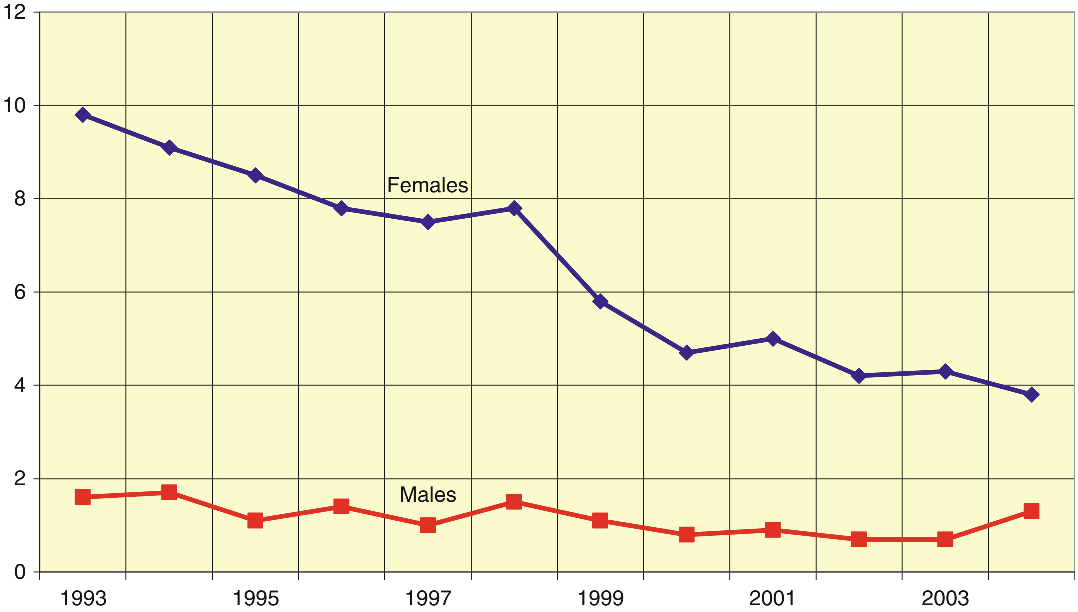

Understanding Homicide and Aggravated Assault, Fig. 7 Victims of nonfatal intimate partner violence per 1,000 population by sex of victim, 1993–2004. Source: Catalano (2006b)

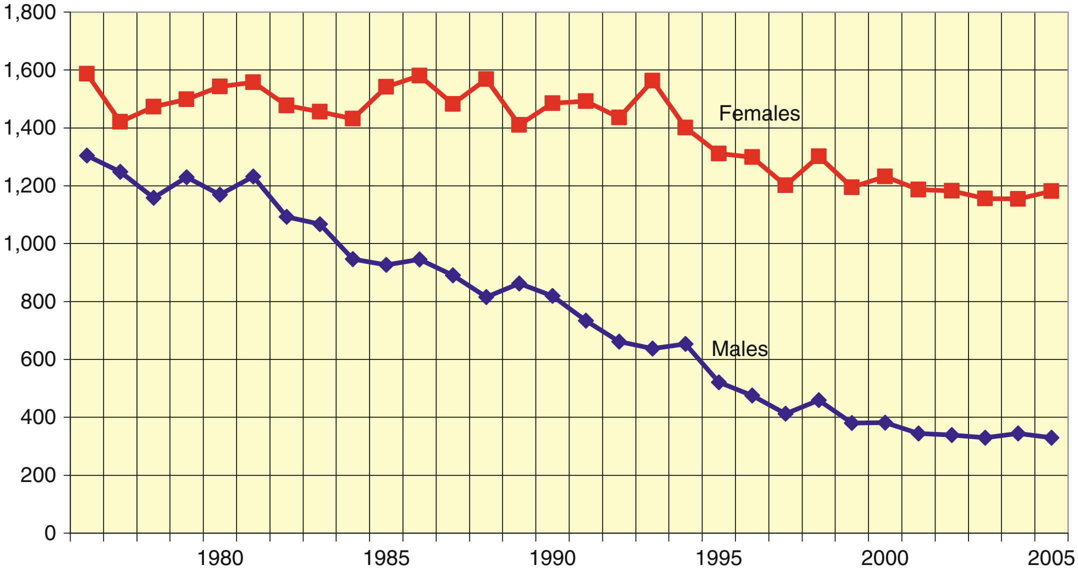

Understanding Homicide and Aggravated Assault, Fig. 8 Intimate partner homicides by sex of victim, 1976–2005. Source: Catalano (2006b)

What accounts for the sex differences in nonfatal and fatal victimization trends?

The conditions responsible for reducing violent crime generally during the 1990s – including a booming economy, more police, and a dramatic expansion in imprisonment – might be expected to have reduced violence against women (Levitt 2004; Rosenfeld 2004). But that makes the sharper drop in male than female intimate homicide victimization all the more puzzling. A key to the puzzle lies in the differing circumstances of male and female intimate partner homicide.

When women kill their male partners, the homicide is much more likely to be "victim precipitated" than when men kill their female partners (Rosenfeld 1997; Wolfgang 1975 [1958]). A homicide is victim precipitated when the victim initiates the violent conflict that ends in his or her death. The person who starts a fight is not always the one left standing when it ends. When women kill their male partners, they typically do so in response to an attack or threat. The same is not true of men who kill their intimate partners (Rosenfeld 1997). Women are more apt than men to kill as a last resort, when all else fails, to stop the violence directed at them, sometimes over a long period of time. Women, then, should be more willing than men to take advantage of nonviolent alternatives to homicide that effectively reduce their exposure to their partner's violence.

Dugan et al. (1999) used that reasoning to explain what otherwise is a very curious result in their study of the impact of hotlines, shelters, legal advocacy, and other domestic violence services and resources on trends in intimate partner killings: intimate killings of men, not women, dropped more rapidly in cities where such resources were more prevalent. It is a fair bet that the domestic violence prevention advocates who demanded more services and greater legal protection for battered and abused women and their children did not intend that their struggle would result in saving batterers' lives. But if we view domestic violence resources as providing nonviolent alternatives to killing a violent partner, it is not surprising in light of the sex difference in the circumstances of intimate partner homicide that they would have a greater impact on the rate at which women kill their male partners than the rate at which men kill their female partners.

In a second study, Dugan et al. (1999) found that the homicide reductions associated with domestic violence services and legal reforms were uneven across racial groups and by marital status: the availability of domestic violence resources produced greater reductions among whites and married couples than among blacks and unmarried intimate partners. These results are consistent with evaluations of so-called mandatory and pro-arrest policies for domestic violence, which have found stronger deterrence effects of arrest on subsequent violence for married and employed men than for the unmarried and unemployed (Sherman 1992). Domestic violence policy interventions explain only part of the drop in intimate partner homicide; rising divorce rates also contributed to the decline as more women exited relationships with violent or abusive partners (Dugan et al. 1999). In fact, an important rationale for liberalizing divorce laws in Europe during the nineteenth and twentieth centuries was to reduce the number of domestic homicides (Gillis 1996). The institution of divorce was the first and arguably remains the most effective program to prevent domestic violence.

Theories of Homicide and Assault

Homicide and assault are studied from a variety of theoretical perspectives, and no single theory commands consensus among students of interpersonal violence, even when the focus is limited to criminological perspectives. It is useful to categorize theories of violence according to their level of analysis and whether they apply to all forms of violence, legal as well as criminal, or are restricted to nonnormative or criminal violence. Microlevel theories direct attention to characteristics of individuals or their immediate social context (e.g., family and peer influences) to explain individual differences in violent behavior. Macrolevel theories explain the variation in rates of violence across groups, populations, and societies. Both micro and macro criminological theories of violence are limited primarily to explaining socially disapproved violence, although two perspectives, Felson's (2004) interactionist theory and Black's (1998) theory of social control, encompass normative violence as well.

While conceptually and analytically distinct, specific micro and macro perspectives share underlying assumptions about the individual and social processes implicated in crime. The three major microlevel perspectives in criminology – social learning theory, social control theory, and strain theory – share behavioral assumptions with three of the major macrolevel perspectives on crime – cultural deviance theory, social disorganization theory, and anomie theory, respectively (Messner and Rosenfeld 2007). Social learning theory posits that criminal behavior is learned under the same social conditions that shape learning of all kinds. Its macrolevel counterpart, so-called cultural deviance perspectives, explains variations in violent crime across groups and communities on the basis of cultural or subcultural values, beliefs, and norms that promote or tolerate violent conduct. Social control theory explains criminal behavior as resulting from weakened or absent social bonds that, when strong, discourage crime and delinquency. Its macrolevel analogue, social disorganization theory, specifies the social conditions that weaken social bonds, including economic deprivation, residential instability, and population heterogeneity. Finally, strain theories posit that criminal behavior is a function of failure to attain desired goals, aversive stimuli, and negative emotions. At the macrolevel, anomie theories hold that high levels of crime characterize communities and societies in which cultural goals extol the virtues of economic success for everyone, access to the legitimate means to attain cultural goals are unequally distributed in the population, and the social controls and supports of basic institutions are weakened.

The three pairs of criminological theories make very different assumptions about how criminality is facilitated by personal and social characteristics. Cultural-learning theories propose that criminal propensities are generated in interaction with others whose values and beliefs conflict with prevailing legal standards. In sharp contrast, social control theories view criminal behavior as needing no special explanation; what must be explained is not why some people steal or behave violently, but why the rest of us do not. People conform to law, control theorists argue, because not doing so is costly. To the degree nonconformity threatens valued social relationships and investments in rewarding legitimate activities, people will conform. In addition, time and energy devoted to legitimate pursuits limit the time and energy available for criminal activity. Contemporary strain and anomie theories incorporate elements of social learning and control theories but also specify the personal and social processes that weaken social controls and encourage the adoption of deviant or criminal attitudes and beliefs (Messner and Rosenfeld 2007).

As mentioned, none of these perspectives holds a commanding position among criminologists. Researchers tend to mix and match arguments from multiple perspectives, often oblivious to the contrasting assumptions they make about how and why persons violate legal standards (Kornhauser 1978). In recent years, rival theories have emerged that seek to transcend the limitations of established theories or recast them around new insights. Routine activity theory (Felson 2002) downplays the significance the older perspectives attach to criminal motivations and emphasizes the opportunities to commit crime that emerge when offenders, however motivated, come into contact with victims in situations that facilitate theft or attack (e.g., an unguarded cash register, a lone individual on a darkened or empty street). Self-control theory (Gottfredson and Hirschi 1990) transcends social control theories by locating the source of deviant behavior in early childhood socialization. Poor parenting produces persons incapable of forming deep attachments to others or controlling their impulses and who will commit crimes whenever opportunity permits. Life course theories also emphasize the importance of early childhood in facilitating or retarding delinquent and criminal behavior in adolescence and adulthood but reject self-control theory's deterministic assumption that faulty socialization spells trouble throughout the life course (Laub and Sampson 2003; Sampson and Laub 1993). Age-graded "turning points," such as marriage, military service, or a good job can promote conformity and discourage criminality.

None of these established or emerging criminological perspectives applies specifically to criminal violence; they are intended to explain criminal behavior generally: violent offending, property offending, and public order offending, such as possessing illegal drugs. At the same time, the explanatory scope of all of them, unsurprisingly, is restricted to *criminal* violence. That is a significant limitation. None of the criminological theories can explain why some forms of violent conduct are socially condemned while others are not. Implicitly or by design, all take the prevailing legal order as given and therefore are fundamentally culture bound.

Not all theories of violence are so restricted in their application. Black's (1998) theory of social control views violence as one among several means of conflict management. Whether disputants engage in violent "self-help" rather than avoidance, toleration, mediation, arbitration, or other ways of managing conflict depends on how closely they are related to one another, the degree of economic inequality or cultural difference between them and relevant third parties, and their integration into broader collectivities. Black's theory is intended to explain dispute-related violence of all kinds, at all times, everywhere.

Felson's (2004) interactionist theory explains both dispute-related and predatory violence. Whereas persons engage in dispute-related (sometimes termed "moralistic") violence in response to a perceived wrong, predatory violence is committed without real or imagined provocation by the victim. When a parent spanks a child for misbehaving, she is engaging in moralistic violence. When a robber demands his victim's money at gunpoint, he is engaging in predatory violence. Rejecting the traditional duality of "expressive" versus "instrumental" violence, Felson argues that all violence is intended to achieve some goal, whether to exact compliance, achieve justice, or secure a desired social identity. His theory applies to both criminal and noncriminal violence, in any and all societies and historical periods.

Many forms of violence are roundly condemned and, in principle, can be explained with reference to the various problems and pathologies highlighted in traditional criminological perspectives. But the commonalities underlying normative and nonnormative violence are obscured by such a restrictive focus on the problematic origins and manifestations of violence. Science craves generality (Black 1995). All else equal, theories of violence capable of explaining the executioner's behavior, and not merely that of the condemned, are better.

Violence Reduction and Mass Incarceration

Homicide and aggravated assault rates in the United States have declined markedly over the past two decades. Homicide fell by 43 % between 1991 and 2005. UCR aggravated assault rates dropped by 33 % over the same period. As measured in the NCVS, aggravated assault rates fell by 57 %. Declines of this magnitude in death and serious injury must be counted as good news. But valued effects are not always brought about by equally appealing causes. That is the case with the relationship between recent crime reductions and mass imprisonment in the United States.

Over a quarter of a century ago, the United States embarked on an unparalleled program of prison expansion. For most of the twentieth century, US imprisonment rates were on par with those prevailing in Europe and the rest of the developed world. Since the mid-1970s, the US imprisonment rate has risen fivefold. No other nation incarcerates its citizens at the now customary levels in the United States.

The rise of mass incarceration in the United States resulted from deliberate policy choices and was not the unavoidable outcome of increasing crime rates. One state after another, with strong inducements and pressures from the federal government, instituted mandatory-minimum sentences, sentence enhancements, and truth-in-sentencing provisions to insure that more convicted felons would be sent to prison and would serve longer sentences (Messner and Rosenfeld 2007, pp. 115–118). Drug offenders were a prime target of the sentencing reforms. Presumably African-Americans were not, but changes in drug laws and other get-tough policies, whether intended or not, have disproportionately affected the life chances of African-Americans (Tonry 1996). Fully 10 % of African-American men between the ages of 25 and 29 resided in a state or federal prison in 2002, compared to 2 % of Hispanic men and 1 % of white men of the same age (Messner and Rosenfeld 2007, p. 5). It is fair to assume that few policymakers intended or foresaw a quintupling of the prison population as a consequence of the mandatory-minimum sentencing and truth-in-sentencing laws passed over the last 30 years, but once the dramatic escalation in incarceration – and its racially disparate effects – became evident, if policymakers were inclined to limit future imprisonment growth, presumably they would have done so by now (Tonry 1996).

Although debate remains about the size and even the direction of the effect, accumulating research evidence shows that the massive expansion in imprisonment has reduced crime rates in the United States (Levitt 1996, 2002, 2004; Marvell and Moody 1994; but see DeFina and Arvanites 2002, Tonry 2004). It would be astounding if mass imprisonment at the level used in the United States had no impact on crime rates. But the effects are complex. They are likely to differ by age, with the strongest crime-reduction effects among offenders in their early 30s, who have the longest residual criminal careers, and the weakest effects among juveniles, who are at low risk for incarceration in state and federal prisons. And recent research has shown that the crime-reduction effects of incarceration diminish with increases in the scale of imprisonment (Liedka et al. 2006).

Mass imprisonment is a stopgap and costly crime-control policy. Most persons released from prison are rearrested for new crimes and half return to prison within 3 years (Langan and Levin 2002). The costs of the huge correctional complex are borne primarily by state and local governments, and they have risen at roughly double the rate of state and local outlays for education, health care, and policing over the past 30 years (Hughes 2006). Each additional dollar spent on prisons, jails, and community corrections is money that states and local areas cannot devote to other pressing needs, including improvements in education and early childhood interventions that hold some promise for reducing *criminality* and not just containing crime rates by incapacitating offenders (Greenwood et al. 1998).

Add to the economic costs of mass imprisonment the political and social costs in the form of a growing disenfranchised population of ex-prisoners, further destabilization of the communities hardest hit by the continuous “churning” of young men moving in and out of prison, the high recidivism rates of ex-prisoners, and the racially disparate impact of prison expansion – and policy alternatives that promise to limit *both* crime rates and prison expansion begin to have growing appeal to policymakers at all levels of government, regardless of political affiliation. A “prisoner reentry” movement has attracted the attention of the federal government and state legislatures across the country. Advocates of prisoner reentry point to the limited employment, housing, health, and social services available to persons returning home from prison and the spotty and inconsistent supervision of many ex-prisoners in the community as primary reasons for their high recidivism rates (see Petersilia 2003; Travis 2005). To protect the communities to which they are released and staunch the flow of offenders back to prison, reentry advocates call for sharply lower parole caseloads and enhanced supportive services for ex-prisoners and their families.

Whether current reentry programs reduce recidivism rates remains uncertain, but if they are found to be effective and are widely implemented, they could well limit the overall growth in imprisonment, for the simple reason that most persons entering prison in recent years have been there at least once before (Blumstein and Beck 2005). The essential challenge to prisoner reentry programs is finding the right mix of incentives, supports, and controls that will reduce the chances that persons who have already been in and out of prison, and then are reincarcerated, will not return to prison yet again when they are released.

Effective crime-control policy requires more than addressing the needs of released prisoners. The best policy is multifaceted and also should include promising approaches to the prevention and policing of violent crime and in-prison rehabilitation programs of proven effectiveness. There is no shortage of such initiatives (http://www.crimesolutions.gov/about_ojp.aspx), but combining them in a coherent policy set and providing the resources necessary to bring them to scale poses enormous political challenges.

One is tempted to say the challenge is insurmountable, but recent history indicates it is not. The United States showed the world it is possible to effect startling changes in crime-control policy when the necessary sense of urgency and political will are mobilized for action. In a nation renowned for its political decentralization and fragmentation, a coherent and self-reinforcing policy of mass imprisonment took hold almost instantaneously across the 50 states with the federal government, no matter which party was in power, leading the way. Limiting the costs of mass imprisonment and instituting effective alternatives are no less possible. The challenge is to mobilize the necessary political resources and instruments before crime rates begin to rise again.

Acknowledgments This entry is adapted from Rosenfeld, Richard. 2009. Homicide and serious assaults. In Michael Tonry (ed) *Oxford handbook on crime and public policy*. New York: Oxford University Press, pp. 25–50.

Related Entries

- ▶ Juvenile Violence
- ▶ Lawful Killings
- ▶ Youth Homicide in the United States

Recommended Reading and References

Beeghley L (2003) Homicide: a sociological explanation. Roman & Littlefield, Lanhan

Black D (1995) The epistemology of pure sociology. Law Soc Inquiry 20:829–870

Black D (1998) The social structure of right and wrong, Revisedth edn. Academic, San Diego

Blumstein A, Beck AJ (2005) Reentry as a transient state between liberty and recommitment. In: Travis J, Visher C (eds) Prisoner reentry and crime in America. Cambridge University Press, New York

Catalano SM (2006a) Criminal victimization, 2005. US Department of Justice, Washington, DC

Catalano SM (2006b) Intimate partner violence in the United States. US Department of Justice, Washington, DC, http://www.ojp.usdoj.gov/bjs/intimate/ipv.htm

Crowell NA, Burgess AW (eds) (1996) Understanding violence against women. National Academy Press, Washington, DC

DeFina RH, Arvanites TM (2002) The weak effect of imprisonment on crime: 1971–1992. Soc Sci Quart 83:635–653

Dugan L, Nagin D, Rosenfeld R (1999) Explaining the decline in intimate partner homicide: the effects of changing domesticity, Women's status, and domestic violence resources. Homicide Stud 3:187–214

Dugan L, Nagin D, Rosenfeld R (2003) Exposure reduction or retaliation? the effects of domestic violence resources on intimate partner homicide. Law Soc Rev 37:169–198

Federal Bureau of Investigation (2006) Crime in the United States 2005. US Department of Justice, Washington, DC, http://www.fbi.gov/ucr/05cius/

Felson MK (2002) Crime and everyday life, 3rd edn. Sage, Thousand Oaks

Felson RB (2004) Predatory and dispute-related violence: a social interactionist approach in routine activity and rational choice. Adv Criminol Theory 5:103–125

Gillis AR (1996) So long as they both shall live: marital dissolution and the decline of domestic homicide in France, 1852–1909. Am J Sociol 101: 1273–1305

Gottfredson M, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Greenwood PW, Karyn Model C, Rydell P, Chiesa J (1998) Diverting children from a life of crime. RAND Corporation, Santa Monica

Hughes KA (2006) Justice expenditure and employment in the united states, 2003. U.S. Department of Justice, Washington, DC

Kleck G, McElrath K (1991) The effects of weaponry on human violence. Soc Forces 69:669–692

Kornhauser RR (1978) Social sources of delinquency: an appraisal of analytic models. University of Chicago Press, Chicago

Langan PA, Levin DJ (2002) Recidivism of prisoners released in 1994. U.S. Department of Justice, Washington, DC

Laub JH, Sampson RJ (2003) Shared beginnings, divergent lives: delinquent boys to Age 70. Harvard University Press, Cambridge
Levitt SD (1996) The effect of prison population size on crime rates: evidence from prison overcrowding litigation. Quart J Economics 111:319–352
Levitt SD (2002) Deterrence. In: Wilson JQ, Petersilia J (eds) Crime: public policies for crime control. ICS Press, Oakland
Levitt SD (2004) Understanding why crime fell in the 1990s: four factors that explain the decline and six that do not. J Econ Perspect 18:163–190
Liedka RV, Piehl AM, Useem B (2006) The crime-control effect of incarceration: does scale matter? Criminol Public Policy 5:245–276
Marvell TB, Moody CE (1994) Prison population and crime reduction. J Quant Criminol 10:109–139
Messner SF, Rosenfeld R (2007) Crime and the American dream, 4th edn. Wadsworth, Belmont
Petersilia J (2003) When prisoners come home: parole and prisoner reentry. Oxford University Press, New York
Reiss AJ Jr, Roth JA (eds) (1993) Understanding and preventing violence. National Academy Press, Washington, DC
Rosenfeld R (1997) Changing relationships between men and women: a note on the decline in intimate partner homicide. Homicide Stud 1:72–83
Rosenfeld R (2004) The case of the unsolved crime decline. Sci Am 290:68–77
Rosenfeld R (2007) Explaining the divergence between UCR and NCVS aggravated assault trends. In: Lynch JP, Addington LA (eds) Understanding crime statistics: revisiting the divergence of the NCVS and the UCR. Cambridge University Press, New York, pp 251–268
Russell DEH (1990) Rape in marriage. Indiana University Press, Bloomington
Sampson RJ, Laub JH (1993) Crime in the making: pathways and turning points through life. Harvard University Press, Cambridge
Sherman LW (1992) Policing domestic violence: experiments and dilemmas. Free Press, New York
Tonry M (1996) Malign neglect: race, crime, and punishment in America. Oxford University Press, New York
Tonry M (2004) Thinking about crime: sense and sensibility in American penal culture. Oxford University Press, New York
Travis J (2005) But they all come back: facing the challenges of prisoner reentry. Urban Institute Press, Washington, DC
Wellford CF, Pepper JV, Petrie CV (eds) (2004) Firearms and violence: a critical review. National Academies Press, Washington, DC
Wolfgang ME (1975/1958) Patterns in criminal homicide. Patterson Smith, Montclair
Zimring FE (2007) The great American crime decline. Oxford University Press, New York
Zimring FE, Hawkins G (1997) Crime is not the problem: lethal violence in America. Oxford University Press, New York

Understanding Victimization Frequency

Andromachi Tseloni
School of Social Sciences, Nottingham Trent University, Nottingham, UK

Synonyms

Victimization incidence; Victimization rate

Overview

In the past two decades, the central place of repeat and multiple crime incidents against the same victims in boosting crime rates, predicting vulnerability to victimization and, with that knowledge, assisting crime prevention, has been recognized. This entry overviews the accumulated evidence on measuring and predicting victimization frequency and hints at possible crime prevention implications. It concludes with a list for future improvements in measuring and thus understanding repeat victimization.

Context

The field of criminology has traditionally focused on the offenders, the origins of offending behavior, and how to contain or ultimately eliminate it. The Chicago School of the inter-world war period broke away from that tradition by examining the societal (demographic, social, and economic) conditions that were associated with crime concentration. In doing so, researchers, such as Shaw and McKay (1942), introduced context to the study of crime and examined the conditions that enabled offending propensities to be formed and acted out. Since then, considerable research has been conducted in this vein and valuable lessons have been learned while Chicago still yields some groundbreaking analyses of the contextual behavioral parameters of crime concentration (Sampson 2006)!

The development of criminology via the study of offending in relation to area crime concentration has been encouraged and facilitated by the type of information that has been available to researchers and practitioners alike. Police records and data from other criminal justice agencies have existed roughly since the formation of these institutions (readers are invited to explore the rich archive held by the Galleries of Justice located in the old Sheriff of Nottingham mansion, court, prison, and "serious" offenders' execution site of the writer's home city should they visit it). Recorded crime data however hold only basic, if any, information on the victims of crime, such as name, sex, and age. As police alongside other state/public institutions' accountability came into question, it became clear that recorded crime gives a partial picture of how much crime is committed. Population-based criminal victimization surveys were thus launched in the 1970s as a means to estimate the "dark" figure of crime and implicitly or explicitly evaluate the quality and efficiency of the police service. Victims of crime, who at that time were conveniently distinguished from offenders, came, therefore, to the forefront of criminological enquiry to offer information about the criminal incidents they have experienced, the circumstances of their occurrence, and any consequences to those involved.

The National Crime Survey was launched in 1973 in the USA as a tool to "measure" crime in society. After two decades, it was revamped as the National Crime Victims Survey (NCVS) with improved information about victim and crime situational characteristics. In 1981, the British Crime Survey (BCS) was launched in the UK to offer data to supplement that on recorded crime (with regular sweeps in England and Wales) and provide appropriate measurements for testing victimization theory and crime-related national policy initiatives. National crime surveys have also been introduced in Canada, Australia/New Zealand, and a number of European countries. In 1989, the first regular comparative survey was launched, the International Victimization Survey, which followed the same sampling design and administered the same questionnaire cross-nationally until recently (2005). Since 2000, it has been known as the International Criminal Victimization Survey (ICVS). For many countries and some parts of the world, the ICVS is the only crime survey data held. The above are the main data sources upon which this entry draws. For a history of the development of victimization surveys, please, see the entry by Jan van Dijk and Pat Mayhew herein.

The pioneering victimization research that originated from the forerunner to the NCVS and resulted in the well-known lifestyle theory examined the association between the likelihood of experiencing crime and each socioeconomic attribute at hand (Hindelang et al. 1978). For instance, whether victims belonged disproportionally to specific population subgroups was assessed in an effort to identify "victim profiles" in a similar manner to that by which offender characteristics had been and continue to be drawn. The victimization theories of lifestyle and routine activities are well documented in the literature and textbooks (Felson 2002). The developments in statistical modeling research of the last two decades have shown that lifestyle and routine activity effects alone do not predict victimization and that context, that is, the social disorganization parameters put forward by Shaw and McKay (1942), is equally important (for instance, Kennedy and Forde 1990; Osborn and Tseloni 1998). Finally, the appearance in the new millennium of studies based on hierarchical modeling established that individual characteristics and lifestyles are not only additional factors to but also conditioned by context (Rountree et al. 1994; Tseloni 2000) as well as victimization history (Tseloni and Pease 2004). Therefore, a micro- and macro-level victimization theory seems to be emerging that recognizes the context-specific importance of socioeconomic attributes, routine activities, and previous crime experiences. This work will not revisit the original theories but dwell on the established evidence that the world, victimization *not* excepted, is a multi-dimensional place.

The Pervasiveness of Victimization: Definitional Issues

This part focuses on the different ways in which victimization and victims can be defined. This is central in understanding, predicting, and thus preventing criminal victimization.

The likelihood of becoming victim of crime, also known as victimization risk or prevalence, gives the probability that a crime target experiences at least one crime incident during a specific period. The population of eligible crime targets varies according to the crime type in question. For instance, personal crimes may affect any individual in a society, property crimes are against households, or car theft refers to the number of car-owning households. The specific period that offers a snapshot of crime problems in a society depends on the sampling design and the nature of the employed survey. It can be, for instance, the quarter, the calendar year (January to December), the financial year (April to March), or a lifetime.

Victimization risk has been widely used and still appears as the main variable of interest in empirical research. However, it confounds a wide spectrum of victim experiences, from unlucky chance encounters to dangerous daily lives. Therefore, examining the all-encompassing probability of being victimized one or more times does not tell much about the mechanisms of criminal victimization and, therefore, how to prevent it. Indeed, each victim type from the typology set out below represents quite different crime experiences affecting the perceived seriousness of that experience (for identical crime types) and requires different prioritization as for a minority of victims, the "continuity" of crime experiences affects their mental health and/or diminishes their quality of life:

- Single victimization refers to one incident during the reference period.
- Multiple victimization refers to two or more different crime types in the same period.
- Repeat victimization exists when victims have experienced the same crime type more than once.
- Series victimization is defined as a large (what constitutes a large enough count is still unresolved between crime survey methodologists) recurrent number of very similar crime incidents that occur under similar circumstances and possibly by the same perpetrators. Finally,
- Composite victimization has recently been suggested to characterize incidents during which more than one crime type occurred, but according to survey methodology practice, they are classified under the most serious one (Tseloni et al. 2010).

Therefore examining the crime count, namely, the entire distribution of incidents including zero crimes, that is, non-victims, offers an alternative way to examine victimization. Such an approach does not mask some of the aforementioned distinctions. In particular, the crime count clearly distinguishes single and repeats, while series, depending on the study's research design, may be included in the upper truncation point of the distribution (Osborn and Tseloni 1998). This is not perfect but much preferred than lumping all victims together in a "one or more" truncation of self-reported victimization. In addition, when aggregate crime categories are investigated, such as personal or property crime, victimization by two or more crimes may include both multiples and repeats. The variable that summarizes the mean number of crimes per individual or household rather than the victim/non-victim dichotomy is the so-called victimization incidence.

The question of whether repeat, multiple, or series victims differ significantly from single victims has not been thoroughly addressed at the time of writing. The only study that tested whether repeat victims have different attributes or live in different contexts than single victims suggested that the former have simply higher crime exposure but no dissimilar risk factors than the latter (Osborn et al. 1996). In the absence of evidence suggesting otherwise, therefore, the full count of crime incidents should be examined (Tseloni et al. 2002).

Rarity and Concentration of Victimization

Crime is a relatively rare event. But for some people, it is not as rare as it would be

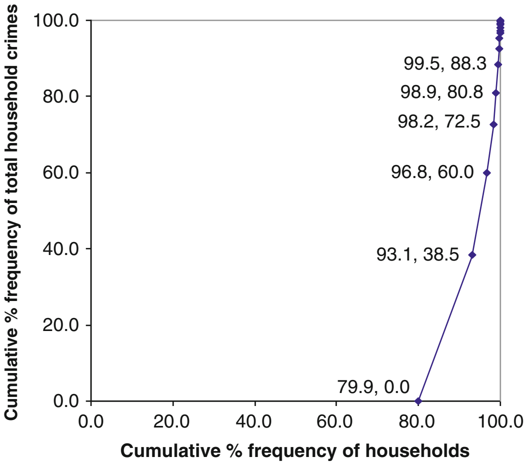

Understanding Victimization Frequency, Fig. 1 Lorenz curve for total household crimes over population (Adapted from Tseloni and Pease 2005)

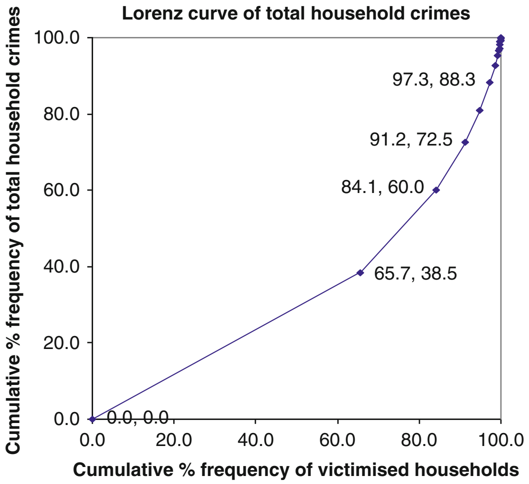

Understanding Victimization Frequency, Fig. 2 Lorenz curve for total household crimes over victims (Adapted from Tseloni and Pease 2005)

if victimization was a matter of chance. Counterintuitively perhaps, high crime areas have a lower proportion of victims than predicted if crime was random. The difference between victimization rate and crime rate is made up by more crimes per victim, that is, repeat victimization (Trickett et al. 1992). In particular, for more than five crimes for every hundred people, the number of victims is overestimated if we assume that crimes are random events and crime concentration, that is, the number of crimes per victim, is underestimated (Osborn and Tseloni 1998).

Figure 1 shows the Lorenz curve of the distribution of property crimes in England and Wales according to the 2000 British Crime Survey (BCS). The Lorenz curve is traditionally employed in public economics to describe how a society's wealth is distributed across the population, and it is related to the Gini coefficient of income inequality. The closer the Lorenz curve is to the low right-hand corner of its square outline, the higher the inequality. In the extreme, one person holds all wealth or in this case experiences all crime.

Figure 1 shows that 80 % of households have not experienced any crime over the 1-year risk period used in the BCS. This is good news for the majority, but bad news for the minority. Tracking the Lorenz curve upward, it is seen that the 6.9 % of the most victimized households experience 61.5 % of crime in the sample. Further, the 3.2 % of such households experience 40 % of property crimes, and so on and so forth, until nearly 12 % of crimes affect the 0.5 % most heavily victimized households. This indicates that the crimes reported by the 20 % victimized households are allocated unequally. Figure 2 demonstrates this more clearly.

For instance, the least victimized 65.7 % of victims experienced 38.5 % of crimes (see Fig. 2). By contrast, the most victimized 2.7 % of victims experienced 11.7 % of crimes. Therefore, crime is affecting both the population and victims unequally. Assuming therefore an appropriate statistical model for highly skewed counts, the chances of two crimes (or any other given number) can be predicted (Osborn and Tseloni 1998).

As mentioned, a large portion of empirical victimization research looks at risk rather than the number of crimes per person. It thereby ignores repeat victimization which is an important component of high crime rates. This seems to be confirmed further by the significant reductions of repeat rather than single victimization rates (Thorpe 2007) during the last 15 or 20 years of considerable crime falls.

Measuring Repeat Victimization and Crime Concentration

Repeat victimization did not feature in a consistent manner in criminological research prior to the work by Ken Pease and colleagues in the 1990s (Forrester et al. 1990). With the exception of few still inspiring studies, such as Reiss (1980), a number of issues with regard to counting and analyzing crime incidents hindered the allocation of crimes over victims. For instance, the discontinuous nature of police statistics, which, as mentioned in the beginning, is offender- or at best location-centered, does not allow a straightforward victim-based aggregation. In addition, the infrequency of crimes overall, that is, in the general population (see Fig. 1), which becomes even more skewed when individual crime categories are analyzed, such as larceny or theft from the person, has led quantitative population-based victimization research to focus on the proportion of victims. Therefore, new ways to look into crime experiences were introduced (Farrell and Pease 1993).

Repeat victimization can be measured in a number of ways depending on the objectives of the analysis. The following fictional crime distribution of Table 1 demonstrates how. It assumes that 100 people were surveyed and a minimum of 0 and a maximum of 4 crimes were reported, the latter thankfully by just one respondent. The majority of hypothetical respondents (52) had no crime experience which gives a victimization risk of 48 %. But 18 respondents were repeat victims; therefore, the percentage of repeat victimization is 100 × 18/48, that is, 37.5 %. The total number of crimes is 75 which gives crime incidence of 75 % or 0.75 crimes per respondent. But crime concentration is 75/48, that is, 1.56 crimes per victim. Crime concentration should be distinguished from crime incidence which refers to the mean number of crimes in the entire population not just victims. Finally, 45 crimes are the result of repeat victimization; therefore, the percentage of repeat crimes is 100 × 45/75 which gives 60 %.

Ideally, therefore, examining the entire (skewed) distribution of crimes via compound Poisson statistical models that are appropriate for counts, such as the distribution given by the first two columns of Table 1, addresses both the binary victim/non-victim outcome that was the focus of early research and repeat victimization.

Understanding Victimization Frequency, Table 1 Fictional crime distribution

Number of crimes	Frequency	Total number of crimes
0	52	0
1	30	30
2	10	20
3	7	21
4	1	4
Total number of respondents	100	75
Total number of victims	48	

Table 2 shows the extent of repeat victimization for each crime type based on the 2000 BCS. "Concentration" measures the mean number of crimes per victim. The second column of "% Repeat crimes" shows the percentage of total crimes that affected the same victims. The last column, "% Repeat victims," measures the percentage of victims who suffered at least two incidents by the respective crime type. For instance, threats and assaults are the most recurring individual crime types. Sixty-four percent of threats were against victims already threatened. Thirty-six percent of threat victims were threatened twice or more within a year. Evidently aggregate categories, such as total household or personal crime, have more repetition because a number of different offense-type counts and the boundaries between multiple and repeats are unclear.

Table 2 suggests that all crime types tend to be suffered disproportionately by the same targets. This is not limited to England and Wales. It is well established that repeats exist across crime types and cross-nationally with roughly similar victim characteristics (Farrell and Bouloukos 2000).

Why Is Crime Concentrated?

Having established that repeat victimization and crime concentration exist, the question calls for

Understanding Victimization Frequency, Table 2 The extent of repeat victimization based on the 2000 British Crime Survey for England and Wales (Source: Tseloni and Pease 2005)

Crime type	Concentration	% repeat crimes	% repeat victims
Total household crime	1.71	61.5	34.3
Burglary	1.32	37.9	17.9
Theft from dwelling	1.46	50.0	26.9
Total personal crime	1.55	53.4	27.8
Assault	1.68	61.2	34.7
Theft and robbery	1.29	31.2	11.0
Threat	1.75	63.7	36.4
Total vehicle crime	1.51	53.7	30.1
Theft from vehicle	1.35	42.7	22.5
Theft of vehicle	1.13	19.1	9.7

an answer. Why are individuals or households likely to experience more than one crime, especially when there are so many more non-victims in the population? Three explanations, which were conveniently borrowed from the literature on statistical analyses of accidents, have been suggested (Tseloni 1995). They are given below:

- Population heterogeneity (apparent contagion) – The simple fact that people are and behave differently and encounter dissimilar situations because they live and spend time in diverse places and/or settings. According to this explanation, all individuals have constant but unequal victimization risks. In the next section, the profile of individuals and settings which are vulnerable to crime and crime repetition is outlined.
- Event dependence (true contagion) – The chances of a second, third, and so on and so forth crime depend upon the outcome of the previous one. All individuals have the same initial risks, but these change after each victimization event. For instance, during his/her first attempt, an offender acquires more knowledge about the target. It is only natural that this new experience will inform and increase the chances of subsequent victimization of the same target.
- Spells – It is the time periods during which an unusually large number of crimes occur. For instance, burglars may operate in an area for a period of time and then move elsewhere perhaps being locked up or reformed. At the time of writing this, there is a spell of gold-related burglaries due to its increased price, for instance.

Traditional victimization theory and the majority of academic empirical research focus on the first explanation. Indeed, population heterogeneity has guided the prediction of the victim/non-victim dichotomy (please see the considerable work by Professors Lauritsen, Miethe, McDowall, and colleagues). The frequency of crime experience and victimization history is however ignored with few exceptions (for instance, Lauritsen and Davis-Quinet 1995), and as mentioned, it may not be a priority in terms of crime prevention as it is repeat crimes that contribute to high crime rates (Pease 1998). The second explanation, event dependence, has guided crime prevention by the Home Office in England and Wales since the 1990s and elsewhere (Pease 1998). The third explanation, spells, has hardly been explored to date (Bowers and Johnson 2005).

Predicting Victimization Frequency

In an oversimplified statement, which, as such, is open to legitimate criticism, victimization frequency depends on "who you are, where you live, and past crime experiences." To be more precise, it can be predicted from the following:

- Individual risk and protective factors and their interactions
- The conditioning of the risk and protective factors according to context and
- The conditioning of the risk and protective factors according to victimization history

The various combinations of the above elements, given reliable victimization survey data and good statistical model fit, can tell how many times crime is expected to occur. The relative importance of each individual and contextual attribute as well as victimization history depends on the specific crime type examined. It has long been established that the mechanisms of victimization are better captured when focusing on narrower crime categories rather than aggregate crime types (Kennedy and Forde 1990). After modeling burglary incidence in England and Wales for almost two decades, the current author can safely say that, although the major risk and protective factors remain unchanged over time, their relative importance may change. Therefore, crime prevention practitioners and other readers should take the order of the independent effects given herein with a pinch of salt.

Independent Effects and Their Interactions

This section overviews the strongest (in terms of size of statistically significant effect) risk and protective factors associated with crime-specific victimization frequency. Perhaps not surprisingly, these effects are consistent both across periods and cross-nationally (Osborn and Tseloni 1998; Tseloni 2000; 2006). The following bullet points give the eight most important crime-specific factors. Eight is an arbitrary cutoff point (the full lists of risk and protective factors are available in the original studies). These are risk factors (indicated via R) which increase the mean number of crimes or protective (indicated via P) that reduce them. More than one attribute within the same bullet point implies that the effect sizes are effectively identical although the signs may be opposite, that is, presenting an (R) and a (P) along the same line.

The crime types overviewed here are total household and total personal crime (contrary to the writer's own wisdom given in the previous section), burglary, household theft, and criminal damage.

The frequency of household crime can be predicted from the following:

- 3+ Cars (R)
- Lone parent (R)/prior assault (R)
- Inner city (R)/prior burglary (R)
- Social renting (R)
- Number of cars per household in the area (P)/terraced or town house (R)
- Percentage of 5–15 years old in the area (R)/area population density (R)/urban (R)
- Prior car theft (R)
- Over £30,000 household income (R)

The well-established opposite effect of individual and contextual affluence on victimization is demonstrated here. Area affluence, which is indicated by the average number of cars per household, is a protective factor against household crimes, while household affluence, given by the number of cars in the household and high income (Over £30,000), is a clear risk factor (Tseloni et al. 2002). Therefore, it seems that from a policy perspective, the relative well-off households in poor areas require more household crime protection than their neighbors which may contradict intuitive justice sentiments. In this vein, the relatively modest households living in affluent areas may consider themselves safe from being targeted.

The second fundamental lesson to be retained is that victimization history, even by a different crime type, as in the prior assault effect on aggregate household mean crimes above, is of high importance. Therefore, event dependence effects via the use of panel data, such as the NCVS, or questionnaire items about crime experiences that happened prior to cross-sectional surveys' reference period are equally worth capturing as socioeconomic and contextual attributes (population heterogeneity).

Household crimes are disaggregated into burglary, household theft, and criminal damage – motor-vehicle crime is ignored here as it can happen away from the victims' residence. The following set of bullet points shows the main factors associated with each crime type.

The frequency of burglary including attempts can be predicted from the following:

- 3+ Cars (R)/lone parent (R)
- Social renting (R)/prior burglary (R)
- Prior assault (R)
- Number of cars per household in the area (P)
- Prior car theft (R)/flat 2nd floor or above (P)
- 1–2 years in the area (R)

- Inner city (R)
- Percentage of 5–15 years old in the area (R)

The best predictor for frequent burglaries is three or more cars in the household; the second and the third most important risk factors are again related to event dependence: been burgled and assaulted in the past (prior to the study period). To demonstrate the earlier point about crime definition- and period-specific effects, in analyses based on more recent BCS data, car ownership loses its top place in the set of associated factors to inner city/urban residence. In addition, when isolating burglary with entry from attempted burglary, a different order of the same set of risk factors appears (Tseloni 2011).

The frequency of household thefts can be predicted from the following:

- 3+ Cars (R)/lone parent (R)
- Prior assault (R)
- Social renting (R)
- Ethnic minority (P)
- Prior burglary (R)
- Terraced (R)
- Prior car theft (R)
- Nonmanual household reference person, henceforth HRP (R)

The following are most associated with the frequency of criminal damage:

- 3+ Cars (R)
- Prior assault (R)
- Percentage of 5–15 years old in the area (R)
- 1 or 2 cars (R)
- Terraced (R)
- Social renting (R)
- Number of cars per household in the area (P)/3 + adults in the household (P)
- Percentage of Indian-subcontinent household representative person (politically correct term for "head of household") in the area (P)

As seen, for instance, the availability of young people in an area is the third most important factor for criminal damage but does not feature in the top eight factors for household theft and makes it last for burglary.

The risk and protective factors of personal crime are the following:

- Divorced (R)
- Single (R)
- Ethnic minority (P)
- 3+ Adults in the household (R)
- Social renting (R)/inner city (P)
- Private renting (R)/children (R)/flat (R)
- Area population density (R)
- Terraced (R)/over £30,000 household income (P)

Three things should be noted here. First, the ninth most important risk factor, which did not make it to the list, is area deprivation. Second, individuals' affluence, ethnic minority, and living in inner city are protective factors for personal crimes. Third, gender is not within the eight (or nine) most associated factors. All three do not only appear in incidence analyses but replicate victimization risk results. Indeed, some studies have not evidenced any significant difference in personal victimization between men and women (Rountree et al. 1994), especially within households of high vulnerability (Tseloni 2000). If anything, females have higher chances of experiencing personal theft than males (Miethe and Meier 1990). Ethnic minorities are significantly less victimized by personal crimes than Whites overall (Tseloni 2000; Lauritsen 2001) and, in particular, when spending at least a night out per week, in full-time employment or study (Miethe et al. 1987) and within ethnically heterogenous communities (Rountree et al. 1994). More research is under way at the time of writing this both in the USA and the UK, but the traditional lifestyle theory propositions that males and/or ethnic minority individuals are more exposed and therefore victimized than others (Hindelang et al. 1978) seem to be counterindicated by results of contextualized individual effects.

The frequency of victimization increases if households or individuals possess multiple risk factors. It diminishes if they possess risk alongside protective factors. In addition, some combinations of attributes reinforce victimization over and above the simple sum of the respective individual effects. For instance, in general, men are more frequently "threatened" than women. But divorced women experience triple the number of threats as divorced men (Tseloni 1995). Similarly, widowed people, who are in general the

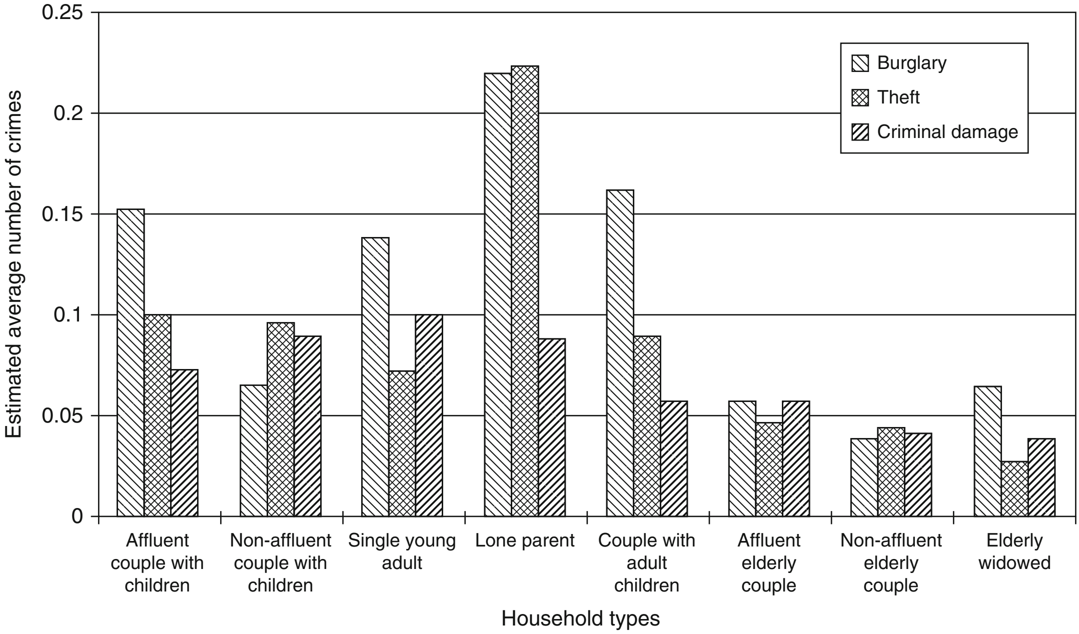

Understanding Victimization Frequency, Fig. 3 Estimated mean rates of property crimes for fictitious households in an average area of England and Wales based on the 1992 BCS (Source: Tseloni et al. 2002)

least victimized by personal crime marital status group, become disproportionally vulnerable in densely populated areas (Tseloni 2010). Therefore, context is as important as individual characteristics and lifestyle. The following section illustrates this point and the one after next explores it in more detail.

So What? How Practitioners Might Make Use of Predicted Population Heterogeneity

Figure 3 shows how estimated crime rates from empirical statistical models can provide information about the relative vulnerability of different households. Having this knowledge, group-specific crime prevention can be developed. For instance, a nonaffluent family experiences considerably more thefts than burglaries. Therefore, if this result is replicated, this group would benefit more from awareness about the risks of offering access into their house than from technological target hardening via burglary prevention devices. By contrast, a couple with adult children which probably possesses more valuables and gadgets would benefit from target hardening.

Figure 4 offers the same information as Fig. 3 but for risk rather than average victimization frequency. It is evident that notable differences exist between predicted risk and incidence. For instance, lone parents are at greater risk of burglary than theft but the mean number of respective events is similar. Single adults experience more incidents of criminal damage than theft but similar risks. This demonstrates that a different story might be told when risk and incidence are examined. By the same token, if interest centered upon victims of three or more repeats, a different order of specific-crime-type frequency per household composition might have appeared. Therefore, modeling (and predicting) the entire distribution of crimes offers flexibility in the choice of a cutoff point for analysis and intervention.

Figure 5 shows how the predicted household crime incidence over household types is influenced by area of residence. Even low vulnerability groups, such as a nonaffluent elderly couple, may experience more crimes in deprived inner city areas than the most vulnerable population subgroup, lone parents, in affluent areas.

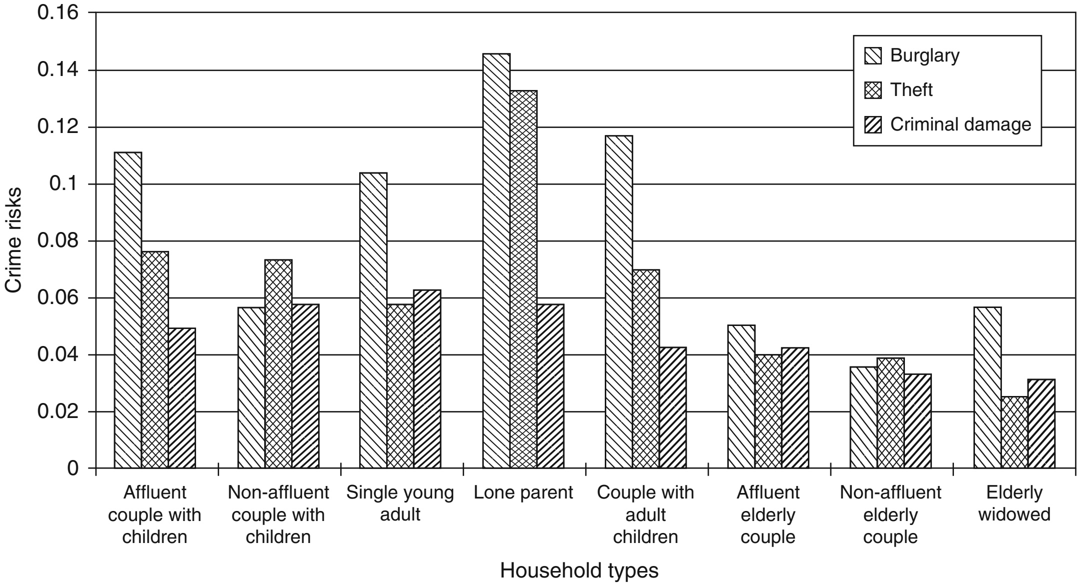

Understanding Victimization Frequency, Fig. 4 Estimated risks in property crime categories for hypothetical households in an average area (Source: Tseloni et al. 2002)

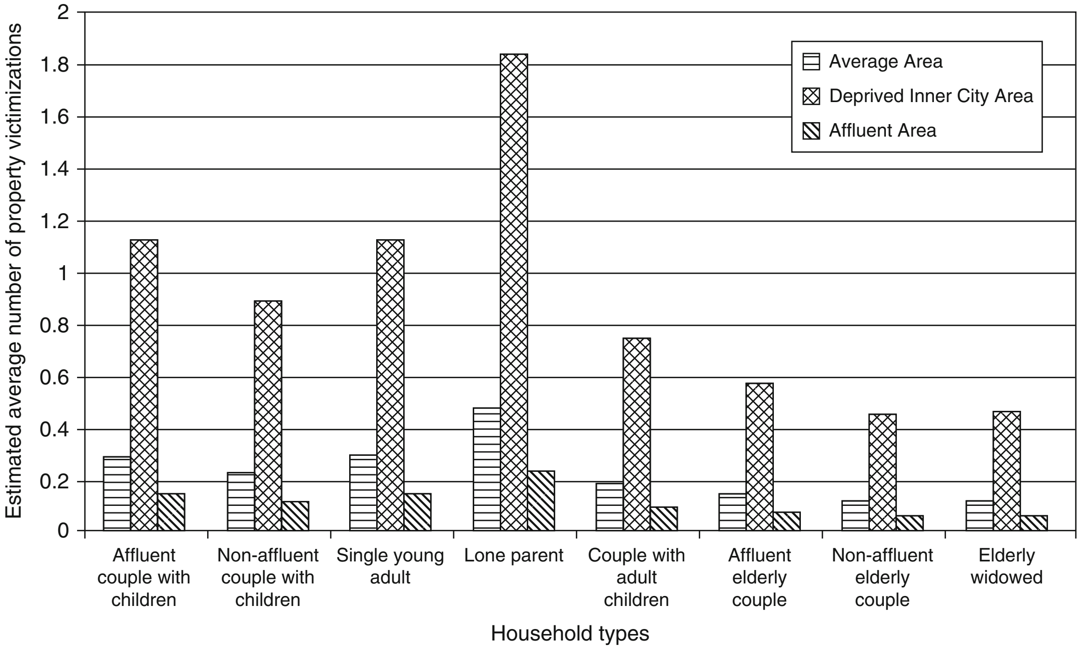

Understanding Victimization Frequency, Fig. 5 Estimated mean rates of aggregate property crime victimization for hypothetical households across selected areas of England and Wales (Source: Tseloni et al. 2002)

Therefore, context has a direct effect on victimization frequency. The next section discusses how it may also interact with and condition the individual and routine activity effects.

Conditioning of Individual Factors According to Context and Victimization History

The interaction between individual and contextual characteristics, the so-called cross-cluster effect (Rountree et al. 1994; Tseloni 2006), addresses the following question: Why is person A more vulnerable than person B in a given situation but not always?

There are two types of cross-cluster interactions: explained, such as the area deprivation interaction with widowhood mentioned in the earlier subsection "Independent Effects and Their Interactions," and unexplained or random. In the latter case, the importance of a risk or protective factor alters by context. It may jump up or down few places in the series of bullet point lists given earlier (subsection "Independent Effects and Their Interactions") from one area to another. To illustrate this, please, consider the classic negative age effect on victimization. Figure 6 shows the average effect for England and Wales in 1999.

Age is indicated as the horizontal axis of Fig. 6. Household crimes (indicated in the vertical axis) increase for HRPs between the ages of 16 (the lower bound of the BCS sample at the time) and about 34 years old and then drop sharply as HRPs grow older. This curve however portrays the average of a wide range of age effects, such as seen in Fig. 7. Households with young HRPs experience about six times more household crimes in the "worse" area than in the best area. As they grow older, the number of household crimes they may experience converges regardless of where they live. As a result, ageing reduces victimization incidence more sharply in the "worst" area compared to the "good" ones.

Another example is the lone parent risk factor: In "good" areas, lone parents are expected to experience roughly the same low number of household crimes as others, but in the "worst" area, they may experience roughly ten household crimes per year (Tseloni and Pease 2010). Similar conditioning of individual effects and lifestyle has been evidenced in victimization risk analyses especially, as mentioned in subsection "Independent Effects and Their Interactions," with regard to ethnicity (Lauritsen 2001; Rountree et al. 1994).

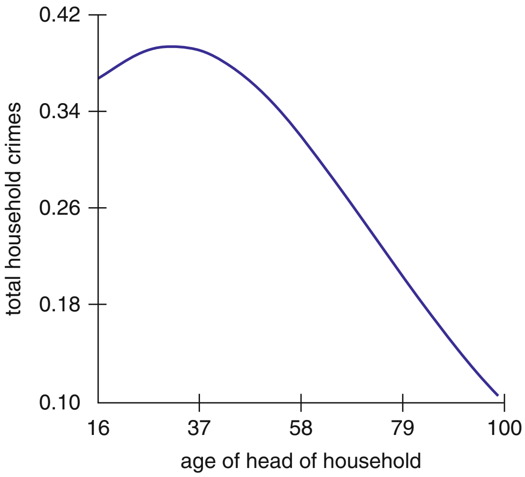

Understanding Victimization Frequency, Fig. 6 Total property crimes over age of household reference persons (Source: Tseloni and Pease 2010)

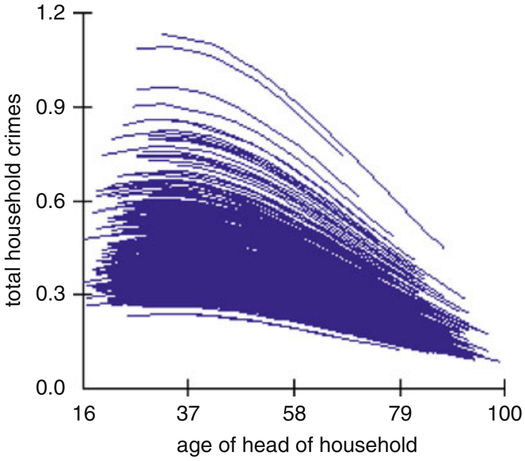

Understanding Victimization Frequency, Fig. 7 Total property crimes over age of "household reference person" for households clustered within English and Welsh areas (Source: Tseloni and Pease 2010)

The last issue in predicting crime concentration is to examine whether the effect of victimization history on current crime experiences depends on sociodemographic characteristics and lifestyle of individuals. To some extent, and especially during adjacent periods, when victims' attributes and routine activities remain the same, event dependence is an indication of unmeasured population heterogeneity, namely, of subtle differences between people or households. The effects of recent victimization history in the short run are exacerbated by outgoing lifestyle and vary widely according to area of residence (Tseloni and Pease 2004).

Future Agenda

This essay discussed the acquisition of knowledge during the last two decades or so on victimization frequency and its prediction. In particular, crime risk and concentration can be examined simultaneously via crime count models and predicted from individual and contextual factors, victimization history, their interactions, and contextual conditioning.

That prediction will most definitely improve if population crime surveys:

- Incorporate a longitudinal element with data on the same households (rather than addresses as currently in the NCVS) that span longer than the 1-year reference period.
- Collect detailed information about each incident that occurred per victim.
- Locate as precisely as possible incidents on a timeline.

Given that crime concentration is predictable, it can also be prevented.

Related Entries

- ▶ Bullying
- ▶ Crime and the Racial Composition of Communities
- ▶ Crime Prevention Through Environmental Design
- ▶ Disadvantage, Disorganization and Crime
- ▶ Domestic Violence
- ▶ Effectiveness of Situational Crime Prevention
- ▶ Hate Crime
- ▶ International Crime Victimization Survey
- ▶ Juvenile Victimization
- ▶ Lifestyle Theory
- ▶ Motor Vehicle Theft
- ▶ Multiple Victims and Super Targets
- ▶ National Victimization Surveys
- ▶ NCVS
- ▶ Near Repeats and Crime Forecasting
- ▶ Neighborhood Effects and Social Networks
- ▶ Poverty, Inequality, and Area Differences in Crime
- ▶ Prediction and Crime Clusters
- ▶ Residential Burglary
- ▶ Social Capital and Collective Efficacy
- ▶ Social Network Analysis and the Measurement of Neighborhoods
- ▶ Stalking
- ▶ Surveys on Violence Against Women

Recommended Reading and References

Bowers KJ, Johnson SD (2005) Domestic burglary repeats and space-time clusters: the dimensions of risk. Eur J Criminol 2:67–92

Farrell G, Bouloukos AC (2000) A cross-national comparative analysis of rates of repeat victimization. Crime Prev Stud 12:5–25

Farrell G, Pease K (1993) Once bitten, twice bitten: repeat victimization and its implications for crime prevention. Crime prevention unit paper 46. Home Office, London

Felson M (2002) Crime and everyday life. Sage, Thousand Oaks

Forrester D, Frenz S, O'Connell M, Pease K (1990) The Kirkholt burglary prevention project: phase II. Crime prevention unit paper 23. Home Office, London

Hindelang MJ, Gottfredson MR, Garofalo J (1978) Victims of personal crime: an empirical foundation for a theory of personal victimization. Ballinger, Cambridge, MA

Kennedy LW, Forde DR (1990) Routine activities and crime: an analysis of victimization in Canada. Criminology 28:137–152

Lauritsen J (2001) The social ecology of violent victimization: individual and contextual effects in the NCVS. J Quant Criminol 17:3–32

Lauritsen JL, Davis-Quinet KF (1995) Repeat victimization among adolescents and young adults. J Quant Criminol 11:143–166

Miethe T, Meier R (1990) Opportunity, choice, and criminal victimization. J Res Crime Delinq 27:243–266

Miethe TD, Stafford MC, Long SJ (1987) Social differentiation in criminal victimization: a test of routine activities/lifestyle theories. Am Sociol Rev 52:184–194

Osborn DR, Tseloni A (1998) The distribution of household property crimes. J Quant Criminol 14:307–330

Osborn DR, Ellingworth D, Hope T, Trickett A (1996) Are repeatedly victimised households different? J Quant Criminol 12:223–245

Pease K (1998) Repeat victimization: taking stock. Crime detection and prevention series paper No. 90. Home Office, London

Reiss AJ (1980) Victim proneness in repeat victimization by type of crime. In: Fienberg S, Reiss AJ (eds) Indicators of crime and criminal justice quantitative studies. Department of Justice, Washington, DC, pp 41–53

Rountree PW, Land KC, Miethe TD (1994) Macro–micro integration in the study of victimization: a hierarchical logistic model analysis across Seattle neighborhoods. Criminology 32:387–414

Sampson RJ (2006) Collective efficacy theory: lessons learnt and directions for future inquiry. In: Cullen FT, Wright JP, Blevins L (eds) Taking stock: the status of criminological theory, advances in criminological theory, vol 15. Transaction Publishers, New Brunswick, New Jersey, pp 149–167

Shaw CR, McKay MD (1942) Juvenile delinquency and urban areas. Chicago University Press, Chicago

Thorpe K (2007) Multiple and repeat victimization. In: Jansson K, Budd S, Lovbakke J, Moley S, Thorpe K (eds) Attitudes, perceptions and risks of crime: supplementary volume 1 to crime in England and Wales 2006/7, Home Office Statistical Bulletin 19/07. Home Office, London, pp 81–98

Trickett A, Osborn DR, Seymour J, Pease K (1992) What is different about high crime areas? Br J Criminol 32:81–89

Tseloni A (1995) The modelling of threat incidence: evidence from the British Crime Survey. In: Dobash RE, Dobash RP, Noaks L (eds) Gender and crime. University of Wales Press, Cardiff, pp 269–294

Tseloni A (2000) Personal criminal victimization in the U. S.: fixed and random effects of individual and household characteristics. J Quant Criminol 16:415–442

Tseloni A (2006) Multilevel modeling of the number of property crimes: household and area effects. J R Stat Soc A Stat Soc 169(Part 2):205–233

Tseloni A (2010) ‘Individual and contextual risk factors of personal victimisation in England and Wales.’ Interdisciplinary workshop on ‘Urban social dynamics: inequalities, segregation and criminality’, EHESS, Paris. Available online http://cams.ehess.fr/docannexe.php?id=1011. Accessed 16 Nov 2010

Tseloni A (2011) ‘Household burglary victimization and protection measures: who can afford security against burglary and in what context does it matter?’ Crime Surveys Users Meeting, Royal Statistical Society, London. Available online http://www.ccsr.ac.uk/esds/events/2011-12-13/index.html. Accessed 13 Dec 2011

Tseloni A, Pease K (2004) Repeat personal victimization: random effects, event dependence and unexplained heterogeneity. Br J Criminol 44:931–945

Tseloni A, Pease K (2005) Population inequality: the case of repeat victimization. Int Rev Victimol 12:75–90

Tseloni A, Pease K (2010) Property crimes and repeat victimization: a fresh look. In: Shoham S, Knepper P, Kett M (eds) International handbook of victimology. Taylor and Francis, Boca Raton, pp 127–149

Tseloni A, Osborn DR, Trickett A, Pease K (2002) Modelling property crime using the British Crime Survey: what have we learned? Br J Criminol 42:89–108

Tseloni A, Ntzoufras I, Nicolaou A, Pease K (2010) Concentration of personal and household crimes in England and Wales. Eur J Appl Math Spec Issue Math Models Crim 21:325–348

Unintended Effects of Imprisonment

Anja J. E. Dirkzwager[1], Marieke van de Rakt[2], Robert Apel[3] and Paul Nieuwbeerta[4,5]

[1]Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, The Netherlands

[2]Avans University of Applied Sciences, Hertogenbosch, The Netherlands

[3]School of Criminal Justice, Rutgers University, Newark, NJ, USA

[4]Department of Criminology and Criminal Law, Leiden University, Leiden, The Netherlands

[5]Utrecht University, Utrecht, The Netherlands

Overview

A growing international literature has been attentive to the collateral consequences of net-widening imprisonment policies, observing that the social costs of steadily rising prison admissions in an era of rapidly declining crime rates might very well outweigh the crime-control benefits. The potential irony of mass imprisonment is that, to the extent it has unintended adverse effects on life outcomes that are correlated with criminal offending, large-scale growth in the incarceration rate may actually exacerbate the crime problem over the long run by stigmatizing an ever larger class of individuals.

One unintended collateral consequence faced by ex-inmates might be felt acutely in the institution of marriage. Specifically, a prison record might erode one's prospects for family formation by delaying or permanently disrupting the transition to marriage. Imprisonment might also disrupt intact marriages by increasing the risk of divorce. In light of compelling research demonstrating an inverse relationship between marriage and criminal behavior, such difficulties could inadvertently sustain an individual's criminal career. Marriage has a prominent place in criminological theory and research as one institution that has the potential to genuinely foster desistance from a criminal career. Mass imprisonment policies in the United States and elsewhere, therefore, pose a potential threat of increased crime if they impede the ability of ex-prisoners to reintegrate into society by stigmatizing them and limiting their chances in the marriage market.

A second unintended collateral consequence of imprisonment that is central in criminological literature pertains the effects of incarceration on physical and mental health. Criminological literature suggests that former inmates are relatively more vulnerable for mental disorders (e.g., depression and suicide) and chronic and infectious diseases such as HIV, tuberculosis, and hepatitis. Moreover, there can be expected to be an ultimate collateral health outcome of incarceration, that is, premature death. Scholars and policymakers are relatively uninformed about the causal effects of incarceration on premature death after release from prison. However, several studies revealed relatively high mortality within prisons and during life after prison.

A third unintended collateral consequence of imprisonment pertains the effects of parental imprisonment on their children's behavior and well-being. There is increasing concern that imprisonment may have far-reaching undesirable consequences for families of prisoners and the wider community. However, these implications are understudied.

This entry provides an overview of the theoretical background and the empirical earlier studies on three unintended collateral consequences of imprisonment: (a) marriage and divorce, (b) mortality, and (c) criminal behavior of children of (ex-)prisoners.

Unintended Effects of Imprisonment I: Marriage and Divorce

Theoretical Background

A number of plausible theoretical explanations can be marshaled to interpret the inverse correlation between incarceration and marriage among unmarried offenders or the positive correlation between incarceration and divorce among married offenders (see for a review Apel et al. 2010). Four possibilities are discussed below, although it is cautioned that the review is not intended to be exhaustive.

First, one mechanism by which incarceration would conceivably erode marriage prospects is through the stigma of imprisonment and the "signal" that such a label transmits to potential spouses. From the perspective of signaling theory and its sociological counterpart – labeling theory – a prison record conveys (however, imperfectly) information about a person that entails more than just his or her future risk of criminal behavior. For example, it might communicate something about one's prospects for success in the labor market, about one's ability to provide for a family, about the company that one chooses to keep, about a potential spouse's risk of becoming a victim of domestic violence, and so on. In other words, a label of "ex-inmate" provides important information that is used by prospective spouses to assess the labeled individual's reputation, integrity, and future prospects – criminal and non-criminal alike. If potential spouses are responsive to cues about one's "marriage potential," it is conceivable that a prison record conveys, quite simply, that one is not marriage material. Incarceration would thus lead to exclusion from the pool of eligible marriage partners because one is viewed as a less viable and attractive partner.

Second, stigmatization might operate in a second, related way by constraining the social networks of ex-inmates and thereby limiting the opportunity to meet potential dating or marriage partners. While crime-prone individuals might run in social circles and frequent geographic locations dominated by like-minded others (through a process of self-selection or social homophily), incarceration and the formal labeling process might exacerbate this tendency through a process of subculture formation among similarly stigmatized individuals. To the extent that policies of mass imprisonment create a society of outcasts through processes of social exclusion, incarceration might result in growing isolation from conventional peer contexts and growing integration into criminal peer networks. This "criminal embeddedness," or social embeddedness in crime, would provide fewer opportunities than before to meet marriageable partners.

Third, for married offenders, incarceration imposes a period of time-out from the marriage. Because incarceration often removes offenders from their communities of residence, married inmates are physically isolated from their spouses and families, which can lead to disruption in the quality of existing marital ties (Lopoo and Western 2005). The stress of separation could lead to withdrawal and dissolution as alternative sources of emotional support are sought by the "surviving" spouse. Additionally, the loss of economic support by the imprisoned individual might motivate a spouse to separate or divorce in the interest of finding a new partner who can help stabilize the financial situation in the household, particularly if children are present (see Phillips et al. 2006).

Fourth, an entirely different view on the incarceration–marriage relationship flows from the possibility that individuals sort themselves into certain institutional settings on the basis of a differential tendency to consider the long-term consequences of their actions, what Gottfredson and Hirschi refer to as self-control. Because individuals with low self-control seek immediate gratification of their desires with minimal effort or long-term planning, they are less likely to be married or, if married, will tend to have difficulty maintaining a stable union: "People who lack self-control tend to dislike settings that require discipline, supervision, or other constraints on their behavior" (Gottfredson and Hirschi 1990, p. 157). Additionally, individuals with low self-control are theorized to be self-centered, insensitive, and nonverbal, qualities that would also likely interfere with union formation and stability. Simply put, non-marriage, an unstable marriage, and crime are all manifestations of the versatility of individuals with low self-control.

Earlier Research

Four datasets provide the only quantitative evidence pertaining to the effect of imprisonment on the likelihood of marriage. For example, data from the Glueck delinquents support the notion that juvenile incarceration disrupts marital unions in adulthood. Sampson and Laub (1993) found that men sentenced to a reformatory as juveniles were more likely to be divorced in young adulthood (ages 17–25) and middle adulthood (ages 25–32). In their model of the propensity to marry (in order to estimate the effect of marriage on crime), Sampson et al. (2006) found that the length of juvenile incarceration was inversely correlated with the probability of marriage throughout adulthood and into the seventh decade of life. Thus, early imprisonment appears to erode a man's long-term marriage prospects. Evidence from the Glueck men also suggests that incarceration impedes marriage formation in adulthood, both contemporaneously and in the short term. For example, Sampson et al. (2006) found that the length of incarceration in the previous year (as well as cumulative incarceration length) was inversely correlated with the probability of being married in the current year.

A pair of studies utilized the Fragile Families and Child Well-Being Study (FFCWS) to estimate the effect of incarceration on family formation among a high-risk sample of parents of newborn children. Western and McLanahan (2000) found that the cumulative incarceration status of the father (at the first follow-up) was strongly related to whether or not the family was intact, that is, whether the father resided with the child's mother in either a cohabiting or marital situation. Specifically, an incarceration history (as reported by the father) reduced the odds of living together by 49 % (odds ratio = 0.51). Using the same dataset, Western et al. (2004) reported that the father's incarceration history (as reported by either the father or mother) was inversely associated with the likelihood of cohabitation and marriage relative to nonresidence (see also Western 2006).

Several recent studies using the National Longitudinal Survey of Youth 1979 (NLSY79) have investigated the relationship between incarceration and marriage. Western (2006) observed that 40 % of males with a prison record remained unmarried until age 40 compared to just over 10 % of never-incarcerated males. Lopoo and Western (2005) reported that male offenders were significantly and substantially less likely to get married during periods in which they were interviewed in an institution. In their discrete-time event history model of first marriage using the 1979–2000 interviews, they estimated that the odds of the transition to marriage were 78 % lower (odds ratio = 0.22) for incarcerated individuals compared to their non-incarcerated counterparts. Similar results were reported from the 1983 to 2000 interviews by Huebner (2005); for race-specific analyses, see Huebner 2007) in a fixed-effects model of marriage, in which she estimated that the odds of being married were 39 % lower for currently incarcerated individuals (odds ratio = 0.61). One notable exception was the study by Raphael (2007) using the 1979–1996 interviews, who found that current incarceration had no impact on the probability of never being married once individual effects were controlled.

With respect to the long-term (versus contemporaneous) relationship between incarceration and marriage in the NLSY79, Huebner (2005, 2007) reported that a history of incarceration as an adult – from the 1983 survey onward, when all respondents were at least 18 years of age – was inversely associated with the probability of marriage but that youthful incarceration (prior to the 1983 interview) was not correlated with marriage. Raphael (2007) noted that having ever been incarcerated prior to the current interview was associated with about a 14-point increase in the probability of having never been married. Lopoo and Western (2005), on the other hand, observed no such effect of having ever been incarcerated on transitions to marriage and divorce. The preponderance of evidence from the NLSY79 suggests, therefore, that incarceration is inversely correlated with marriage, although when measured contemporaneously, this is conceivably an incapacitation effect of prison on marriage likelihood. The findings about the long-term or accumulative effect of incarceration on marital transitions are more equivocal.

Research on the likelihood of divorce is rarer still, but the results are quite consistent. Lopoo and Western (2005) and Western (2006) found that a transition to divorce was significantly more likely to occur contemporaneous with periods of incarceration in the NLSY79. In particular, there was an astonishing twofold increase in the odds of divorce among individuals who were incarcerated relative to their non-incarcerated peers (odds ratio = 2.99). In the FFCWS, Western (2006) similarly found that incarceration significantly increased the risk of separation for men in marital or cohabitational unions, and in fact it was among the strongest predictors in his model (odds ratio = 2.23). Thus, incarceration appears to have a major disruptive effect on preexisting unions.

Finally, Apel et al. (2010) use a long-term study of a conviction cohort in the Netherlands to ascertain the effect that first-time imprisonment has on the likelihood of marriage and divorce. The analysis uses data from the Criminal Career and Life-Course Study (CCLS), a large-scale longitudinal study. The results suggest that the effect of imprisonment on the likelihood of marriage (among unmarried offenders) is largely a selection artifact, although there is very weak evidence for a short-lived impact that does not persist past the first year post-release. This is interpreted as a residual incapacitation effect. On the other hand, the results strongly suggest that the experience of incarceration leads to a substantially higher divorce risk among offenders who are married when they enter prison.

Unintended Effects of Imprisonment II: Mortality

Theoretical Background

Although some criminological theories have suggested that offenders are more likely than non-offenders to suffer early morbidity and mortality, knowledge on the exact nature of the relationship between incarceration and mortality remains scarce. A number of criminological and health theories were reviewed to learn more about the possible effects of imprisonment on mortality. A literature review shows that, based on different theories, at least four conflicting hypotheses can be formulated about the expected effects of imprisonment on mortality (see Dirkzwager et al. 2012).

First, it is possible that imprisonment has no effect on mortality at all. In that case, differences in mortality rates between ex-prisoners and others result from the fact that other common factors influence both imprisonment and mortality. For instance, individual factors such as low intelligence and high impulsivity can cause criminal behavior, as well as poor health and mortality. In their general theory of crime, Gottfredson and Hirschi (1990) emphasize the importance of the causal agent of low self-control in explaining crime. Compared with their high self-control counterparts, individuals with low self-control are more likely to engage in all sorts of criminal, reckless, and risk-taking behavior throughout their life course. Empirical studies have provided evidence for this theory and showed that personality factors such as high impulsivity, low self-control, and a high need for sensation seeking were related to criminal behavior as well as to an irregular and unhealthy lifestyle with excessive drinking, drug use, risky sexual behavior, and risk-taking in traffic.

Another factor that may influence both criminal behavior (including imprisonment) and mortality is socioeconomic status. It is well established that individuals of low socioeconomic status are overrepresented in prison populations. Throughout history, socioeconomic status (even without delinquency) has been related to health, with persons higher in the social hierarchy having a better health than those lower in the hierarchy. Additionally, research consistently demonstrated a relationship between low socioeconomic status and mortality risk from all causes, as well as specific causes of death.

Second, it can be hypothesized that imprisonment has a direct positive effect on health and consequently decreases the risk of premature mortality. It is, for instance, possible that imprisonment decreases mortality risk by improving prisoners' health. Life in prison can have some healthy elements. For instance, in some countries, prisoners receive three meals a day, have a right to exercise, have free access to medical and mental health care that must meet the same quality requirements as health care outside prison, and have fewer possibilities to use drug and alcohol while imprisoned. Such health-improving elements of imprisonment may be particularly beneficial for individuals who were already living in problematic circumstances before imprisonment, including poverty, addictions, and homelessness.

Imprisonment can also be hypothesized to change prisoners' skills, knowledge, and values. Some theories such as learning theories, and economic and social control theories, argue that investment in human capital (i.e., an individual's skill level, knowledge, and experiences) will increase a person's (market) value, will increasingly embed a person into conventional society, and will increase the costs associated with further criminal behavior. This idea is consistent with the rehabilitation goal of imprisonment. By offering training (such as education, job training, and social skill training), prisoners are expected to acquire new skills that may increase the possibilities of a conventional life and decrease the chance of future criminal behavior. Thus, when the rehabilitation of prisoners is successful, former prisoners are equipped with more constructive skills that may enhance a healthier lifestyle with a decreased risk of mortality.

Third, it is possible that imprisonment has a direct negative effect on health and consequently increases mortality risk. A prison experience can be seen as a major stressful event in prisoners' lives, which may negatively affect their mental and physical health, and thereby increase their risk of mortality. Stress theories argue that major life events and exposure to prolonged stress require major behavioral adaptation. Exposure to (chronic) stress will result in an increased awareness of the body and a burdening of certain physiological and bodily systems (e.g., cardiovascular, immune, and hormone systems). On the long term, if an individual is unable to successfully adapt to stress, this can predispose an individual to disease. Major life events have been related to a number of negative health outcomes, such as psychiatric problems, morbidity, and mortality.

Moreover, in some countries, the conditions of confinement are harsh, including a lack of food and medical services, poor quality of care, limited possibilities for self-care, aggressive incidents, and overcrowding. Such conditions can negatively affect prisoners' health on the long term and may increase the risk of premature death after release from prison. Furthermore, relatively high rates of chronic and infectious diseases are observed within prison facilities. Therefore, prisoners may be disproportionately exposed to such diseases that can negatively affect their further health and increase their risk of premature death. Finally, it has been argued that prisons can be "schools of crimes" and breeding grounds for further crime. In line with this belief, a prison experience will increase the chances of future criminal and risky behavior and associated risks of experiencing violent encounters in life.

Fourth, imprisonment may have an indirect negative effect on health and therefore increase the risk of premature death. Research has shown that imprisonment can be a turning point, changing the lives of those involved in a negative way. Imprisonment can, for instance, result in long-lasting problems on the labor market, a loss of income, and homelessness. Additionally, incarceration removes the prisoners from their spouses, families, friends, and neighbors. As a consequence, former prisoners can experience difficulties in family formation and dissolution and a decrease in social networks and social capital. Former prisoners often have to deal with stigmatizing reactions that may also negatively affect the degree of social support. All these factors, that is, a low socioeconomic status, unmarried civil state, and decreased social support, are associated with increased mortality risks in themselves.

In sum, criminological and health theories do not provide clear and unequivocal answers but confront us with alternative hypotheses about the way imprisonment can result in either higher or lower risks of premature mortality.

Earlier Research

The limited knowledge on the effects of incarceration on premature death is, to a large extent, due to limitations in the research design of the existing studies. Determining an effect of incarceration on mortality is difficult because a selective group of offenders will enter prison. A review of the literature (see Dirkzwager et al. (2012) for more information) shows that to date only 24 studies worldwide have compared postprison mortality rates of ex-prisoners with the mortality rates of some kind of comparison group. The studies investigated mortality rates among former inmates in different parts of the world, such as the United States, the United Kingdom, Australia, and Europe. There was a substantial difference in the length of the postprison follow-up periods.

In most of these studies (i.e., 22 of the 24 studies), the mortality rates of former prisoners were compared with rates in the general population (age and gender adjusted). This is problematic, however, because there are well-known preexisting differences between prisoners and the general population (e.g., in criminal history, educational level, ethnicity, social class, mental health). Therefore, such studies may be seriously biased in their estimates of the effects of incarceration on mortality (i.e., probably overestimating prison effects). Subsequently, such studies are problematic in drawing causal conclusions regarding the effects of imprisonment on postprison mortality.

However, even the two studies (Fleming et al. 1992; Sattar 2001, 2003) using more appropriate control groups (i.e., offenders receiving noncustodial sentences instead of the general population) did not take selection effects adequately into account. Obviously, the assignment of offenders to custodial or noncustodial sentences is not random. Judges will be more likely to imprison offenders who have committed serious crimes, who have more prior convictions, and who have a high risk of recidivism. The idea that more serious and frequent offenders encounter more difficulties in their lives, including health and mortality problems, is well established.

As mentioned, to identify studies on the effects of imprisonment on postprison mortality, Dirkzwager et al. (2012) performed an extensive literature search. Based on this search, they identified 24 studies that examined mortality rates among former prisoners, relative to a comparison group. Of the 24 studies, 22 compared mortality rates of former prisoners with rates in the general population, adjusted for age and gender (sometimes also adjusted for race). The results consistently show that compared with gender- and age-adjusted persons from the general population, former inmates were more likely to die from both natural and unnatural causes of death. Former inmates appear particularly at risk of dying during the first few weeks after release from prison. Mortality during the first weeks following release of prison was particularly associated with death by drug overdose. This suggests that, soon after their release, addicted prisoners used drugs again and died due to overdose (perhaps by accident, having not used drugs during imprisonment). Other prevalent causes of unnatural deaths among former prisoners were accidents, suicide, and homicide. Leading causes of natural deaths among former prisoners were cardiovascular diseases, cancer, liver diseases, and infections.

Comparing mortality rates of former prisoners with gender- and age-adjusted persons from the general population can be informative. However, this method can provide only limited information regarding the relationship between imprisonment and the increased risk of premature death. The group of former prisoners and the general population differ on more characteristics than age and gender alone. Prison populations are, for instance, overrepresented by ethnic minorities, people from lower social classes, and people with poor mental and physical health (Sattar 2001); such preexisting differences can also cause differences in mortality rates.

A more adequate approach is to compare former prisoners with a more similar control group, such as offenders who were sentenced to a noncustodial sentence. Only three studies have investigated mortality rates among former prisoners and offenders receiving noncustodial sentences. First, in England and Wales, Sattar (2001, 2003) compared the nature and risk of death of prisoners, with offenders serving community sentences, former prisoners being supervised in the community, and the general population. Standardized mortality rates showed that, compared with the general population, community offenders and former prisoners were both more likely to die. Mortality rates of former prisoners were either somewhat lower (in 1996) or similar (in 1997) to mortality rates of offenders serving community sentences (Sattar 2003). Although these overall mortality rates may seem similar, the study also showed that ex-prisoners were more likely to die within the first weeks after release from prison. These results suggest that the phase immediately after release is particularly risky for ex-prisoners in terms of accidents or incidents with drugs, alcohol, or suicide (Sattar 2003).

Second, in Australia, Fleming et al. (1992) investigated people who died while they were serving a noncustodial correction order during the years 1987 and 1988. The sample consisted of offenders on parole (i.e., ex-prisoners) and offenders serving other community correction orders, such as probation and community service orders. The mortality rate of ex-prisoners was 15.1 per 1,000 orders compared with 10 per 1,000 orders for offenders on probation and 2.2 per 1,000 orders for offenders serving community service orders. The results of their study suggest that former prisoners serving parole orders are at increased risk of premature death compared with people serving other forms of noncustodial sentences. It should be noted, however, that the group of parolees in their study was older than those serving other types of custodial orders, and age is also associated with mortality risks.

Third, in the Netherlands, Dirkzwager et al. (2012) examined the effects of first-time imprisonment on postprison mortality. Data are used from a longitudinal study examining criminal behavior and mortality over a 25-year period in a representative group of 2,297 Dutch offenders who had their criminal case adjudicated in 1977. Of these offenders, 597 were imprisoned for the first time in their lives in 1977. The remaining 1,700 offenders got a noncustodial sentence. Ex-prisoners' mortality rates and causes of death are compared with those in the general population and those in a matched control group of non-imprisoned offenders. Propensity score matching is used to minimize selection bias. Odds ratios with 95 % confidence intervals are used to examine whether mortality among the ex-prisoners differs significantly from the general population or from the non-imprisoned controls. About 18 % of the imprisoned offenders died over the 25-year follow-up period. Compared with the general population (age and gender adjusted), ex-prisoners are three times as likely to die during the 25-year follow-up (odds ratio [OR] = 3.21). Compared with a more appropriate control group of non-imprisoned offenders (matched on age, gender, and propensity score), ex-prisoners are no longer significantly more likely to die (OR = 1.40). In line with the available research, the Dutch results demonstrate significantly elevated mortality rates among former prisoners when compared with age- and gender-adjusted individuals from the general Dutch population. In the present study, ex-prisoners were three times more likely to die during the 25-year follow-up period than persons from the general population. The difference in mortality rates between former prisoners and non-imprisoned offenders of similar age and gender was substantially smaller and even became nonsignificant when former prisoners were compared with non-imprisoned offenders who were similar regarding age, gender, and propensity score. The results of the present study emphasize the importance of constructing appropriate comparison groups when examining the effects of imprisonment on postprison mortality.

Unintended Effects of Imprisonment III: Criminal Behavior of Children of Prisoners

Theoretical Background

Reviews by Murray and Farrington (2008a) and Van de Rakt et al. (2012) distinguish four theories that explain how paternal imprisonment might cause an increase in child criminal behavior. First, trauma theories suggest that the parent–child separation caused by the imprisonment might be harmful for children (for a full discussion of this perspective, see Murray and Murray 2010). Attachment theory and social bonding theory both acknowledge the importance of children's attachment to their parents. The separation caused by imprisonment usually is unexpected, and opportunities for contact are often very restricted during their imprisonment. Following trauma theories, disrupted attachment to the parent following parental imprisonment might cause an increase in children's delinquent behavior. Parental imprisonment might have stronger effects on children if it is experienced early in childhood, because parent–child bonding in early childhood is particularly important for child development, and separation can be more disruptive when young children have fewer cognitive skills to process the event. Longer separations might predict worse outcomes for children.

Second, modeling and social learning theories suggest that parental imprisonment might cause an increase in child criminal behavior because children become more aware of parental criminality and imitate their parents' behavior. Differential association theory (Sutherland et al. 1992) proposes that criminal behavior is learned in the same way as normal (accepted) behavior is learned. Thus, learning of criminal behavior primarily takes place in intimate personal groups, such as the family. Not only are the techniques to commit crime learned, but motives, values, and attitudes toward crime are also learned. Association with criminal parents might be especially influential in determining children's criminal behavior. Parental imprisonment could make children more aware of their parent's criminality and encourage the idea that committing crimes is normal and desirable. Such learning and imitation processes might be especially strong after early childhood, during adolescence, when children are more aware of the meaning of their parents behavior and imprisonment. At the same time, removal of a highly antisocial parent from the household, because of imprisonment, might mitigate these learning processes and, in some cases, actually reduce the probability that children develop criminal behavior.

Third, strain theories suggest that the loss of economic and social capital due to parental imprisonment causes children to commit more crimes. Parents cannot contribute to family income while in prison, and telephone calls and prison visits can add considerably to family expenses. In the long term, imprisonment might lead to unemployment and fewer labor market opportunities for ex-prisoners. Also, children are more likely to have unstable care arrangements when one of their parents is in prison. Therefore, the quality of parental care and supervision may be reduced by parental imprisonment. Because of reduced income, social capital, and supervision, children with imprisoned parents may have fewer chances in school and in the labor market. This may increase their risk for delinquency compared to children of non-imprisoned parents.

Finally, labeling theories suggest that parental imprisonment might cause children to experience stigma, bullying, and teasing, which increases their criminal behavior. Children might become more reluctant to go to school and to socialize with other children. As a result, children with imprisoned parents might perform worse at school and in the labor market than their peers. There might also be official bias against children of prisoners, making them more likely than other children to be monitored by the police and to be convicted. Labeling mechanisms will be most likely to influence children when parental imprisonment endures for a long period of time and when it is experienced in adolescence. When children are somewhat older, they understand the meaning of parental imprisonment better. In these settings, stigmatization and bullying will be more powerful.

Earlier Research

A critical problem for research is to separate out the effects of parental imprisonment on children from the influence of other childhood risk factors. Large-scale longitudinal studies show that children of prisoners are much more likely than their peers to be exposed to other risk factors, such as low IQ, high daring, poor school attainment, poor parenting practices, parental criminality, and low family socioeconomic status. It is particularly important to disentangle the effects of parental imprisonment from the effects of parental criminality, because prisoners tend to be highly criminal, and there is strong evidence that crime runs in families.

To accurately estimate the long-term risks for children after parental imprisonment and the effects on children's criminal careers, studies need to use representative samples, suitable control groups, and long-term follow-ups. Five main studies have done this to date (for an extensive review of prior research, see Murray and Farrington 2008a; Van de Rakt et al. 2012).

In the first study, Huebner and Gustafson (2007) compared rates of adult offending behavior between 31 children whose mothers had been imprisoned and 1,666 children whose mothers had not been imprisoned in the National Longitudinal Survey of Youth (NLSY). The NLSY is a prospective longitudinal survey of males and females in the United States, who were aged 14–22 in 1979. Of children with imprisoned mothers, 26 % were convicted as an adult, compared with 10 % of controls. Controlling for background variables (including child, maternal, paternal, family, and peer risk factors), maternal imprisonment still significantly predicted adult convictions. These results are consistent with the idea that maternal imprisonment has a causal effect on children.

In the second study, Murray and Farrington (2005) investigated the effects of parental imprisonment on children in the Cambridge Study in Delinquent Development (CSDD). The CSDD is a prospective longitudinal study of 411 boys, born in 1953 and living in a working-class area of South London. Outcomes were compared between 23 boys who were separated because of parental imprisonment (between birth and age 10) and four control groups: (1) boys with no history of parental imprisonment or parent–child separation (up to age 10), (2) boys separated because of hospitalization or death, (3) boys separated for other reasons, and (4) boys whose parents were imprisoned only before the boy's birth. Parental imprisonment during childhood was a strong predictor of antisocial-delinquent behavior through the life course. For example, of boys separated because of parental imprisonment, 65 % were convicted themselves between ages 19 and 32, compared with 21 % of boys with no history of parental imprisonment or separation. Effects of parental imprisonment remained even after controlling for other childhood risk factors in the study (including parental criminality), suggesting that parental imprisonment might have a causal effect on children. Parental imprisonment also strongly predicted boys' mental health problems, educational failure, drug use, and unemployment in the CSDD.

In the third study, Murray et al. (2007) compared rates of adult offending behavior of 283 children whose parents were imprisoned (from the children's births until they were age 19) and 14,589 children without imprisoned parents in Project Metropolitan. Project Metropolitan is a prospective longitudinal survey of children born in 1953 and living in Stockholm, Sweden (Janson 2000). Parental imprisonment was strongly predictive of offspring criminal behavior between ages 19 and 30. Of prisoners' children, 25 % offended as adults, compared with 12 % of controls. However, when account was taken of background criminality of parents (using regression analyses and comparing children exposed to parental imprisonment in childhood with children whose parents were imprisoned only before the child's birth), there were no additional effects of parental imprisonment on children. This suggested that parental imprisonment did not cause offspring offending in Sweden.

In the fourth study, Kinner et al. (2007) compared 137 children whose mothers' partners had ever been imprisoned with 2,262 children whose mothers' partners had not been imprisoned in the Mater University Study of Pregnancy. This is a prospective longitudinal survey of 8,458 women who were pregnant in Australia in 1981 and their children. At age 14 years, children whose mothers' partners had been imprisoned were significantly more likely to have externalizing problems than their peers. However, after controlling for other parental and family risk factors, there was no effect of the imprisonment on child externalizing problems suggesting that the original association was spurious.

In the fifth study, Van de Rakt et al. (2012) investigated the effects of fathers' imprisonment on criminal convictions of their children (aged 18–30). Unique official data of the Criminal Career and Life-Course Study (CCLS) are used on a nationally representative sample of Dutch men convicted in 1977. Growth curve analysis is used to establish the influence of paternal imprisonment on the development of criminal careers of children. Special attention is paid to the timing and the duration of the imprisonment. The authors demonstrate an association between fathers' imprisonment and child convictions, especially when fathers are imprisoned when the child is between 0 and 12 years old. When fathers' criminal history is controlled for, the influence of paternal imprisonment becomes much weaker, although it remains significant. The dose–response relationship between the length of the father's imprisonment and children's convictions disappears after controlling for other variables. In the Netherlands, effects of paternal imprisonment on children are very weak and similar to the effects found in another study in Sweden.

In summary, five longitudinal studies show that parental imprisonment is strongly associated with child antisocial and criminal behavior. However, the evidence on causal effects is mixed: two studies suggest that there are causal effects, and three studies suggest that the relationship is weak or even spurious. It is possible that these differences are accounted for by differences in the social and penal contexts where the studies took place.

Recommended Reading and References

Apel B, Blokland AAJ, Nieuwbeerta P, van Schellen M (2010) The impact of imprisonment on marriage and divorce: a risk set matching approach. J Quant Criminol 26:269–300

Dirkzwager AJE, Nieuwbeerta P, Blokland AAJ (2012) Effects of first-time imprisonment on post-prison mortality: a 25-year follow-up study with a matched control group. J Res Crime Delinq 49:383–419

Farrington DP, Coid JW, Murray J (2009) Family factors in the intergenerational transmission of offending. Crim Behav Ment Heal 19:109–124

Fleming J, McDonald D, Biles D (1992) Deaths in non-custodial corrections, Australia and New Zealand, 1987 and 1988. In: Biles D, McDonald D (eds) Deaths in custody, Australia, 1980–1989, research paper no. 12. Aust Inst Criminol, Canberra, pp 239–276

Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Huebner BM (2005) The effect of incarceration on marriage and work over the life course. Justice Q 22:281–303

Huebner BM (2007) Racial and ethnic differences in the likelihood of marriage: the effect of incarceration. Justice Q 24:156–183

Huebner BM, Gustafson R (2007) The effect of maternal incarceration on adult offspring involvement in the criminal justice system. J Crim Justice 35:283–296

Janson C-G (2000) Project metropolitan. In: Janson C-G (ed.) Seven swedish longitudinal studies in the behavioral sciences. Forskningsradsnamuden, Stockholm, Sweden, pp 140–171

Kinner SA, Alati R, Najman JM, Williams GM (2007) Do paternal arrest and imprisonment lead to child behaviour problems and substance use? A longitudinal analysis. J Child Psychol Psychiatry 48:1148–1156

Laub JH, Vaillant GE (2000) Delinquency and mortality: a 50-year follow-up study of 1,000 delinquent and nondelinquent boys. Am J Psychiatry 157:96–102

Lopoo LM, Western B (2005) Incarceration and the formation and stability of marital unions. J Marriage Fam 67:721–734

Murray J, Farrington DP (2005) Parental imprisonment: effects on boys' antisocial behaviour and delinquency through the life-course. J Child Psychol Psychiatry 46:1269–1278

Murray J, Farrington DP (2008a) Effects of parental imprisonment on children. In: Tonry M (ed) Crime and justice: a review of research, vol 37. University of Chicago Press, Chicago, pp 133–206

Murray J, Farrington DP (2008b) Parental imprisonment: long-lasting effects on boys' internalizing problems through the life-course. Dev Psychopathol 20:273–290

Murray J, Murray L (2010) Parental incarceration, attachment and child psychopathology. Attach Hum Dev 12:289–309

Murray J, Janson C-G, Farrington DP (2007) Crime in adult offspring of prisoners: a cross-national comparison of two longitudinal samples. Crim Justice Behav 34:133–149

Nieuwbeerta P, Piquero AR (2008) Mortality rates and causes of death of convicted Dutch criminals 25 years later. J Res Crime Delinq 45:256–286

Phillips SD, Erkanli A, Keeler GP, Jane Costello E, Angold A (2006) Disentangling the risks: parent criminal justice involvement and children's exposure to family risks. CPP 5:677–702

Raphael S (2007) Early incarceration spells and the transition to adulthood. In: Danziger S, Rouse CE (eds) The price of independence: the economics of early adulthood. Russell Sage Foundation, New York, pp 278–306

Sampson RJ, Laub JH (1993) Crime in the making: pathways and turning points through life. Harvard University Press, Cambridge

Sampson RJ, Laub JH, Wimer C (2006) Does marriage reduce crime? A counterfactual approach to within individual causal effects. Criminology 44:465–508

Sattar G (2001) Rates and causes of death among prisoners and offenders under community supervision. Home office research study 231. Home Office, London

Sattar G (2003) The death of offenders in England and Wales. Crisis: J Crisis Interv Suicide Prev 24:17–23

Sutherland EH, Cressey DR, Luckenbill DF (1992) Principles of criminology. General Hall, New York

van de Rakt M, Murray J, Nieuwbeerta P (2012) The long term effects of parental imprisonment on criminal trajectories of children. J Res Crime Delinq 49:81–108

Western B (2006) Punishment and inequality in America. Russell Sage Foundation, New York

Western B, McLanahan S (2000) Fathers behind bars: the impact of incarceration on family formation. In: Fox GL, Benson ML (eds) Families, crime and criminal justice: contemporary perspectives in family research, vol 2. Elsevier Science, New York, pp 309–324

Western B, Lopoo LM, McLanahan S (2004) Incarceration and the bonds among parents in fragile families. In: Patillo M, Weiman DF, Western B (eds) Imprisoning America: the social effects of mass incarceration. Russell Sage Foundation, New York, pp 21–45

US National Crime Victimization Survey

Lynn A. Addington[1] and Callie Marie Rennison[2]
[1]Department of Justice, Law, and Society, American University, Washington, DC, USA
[2]School of Public Affairs, University of Colorado, Denver, Denver, CO, USA

Synonyms

National Crime Survey; NCVS

Overview

The importance of the National Crime Victimization Survey (NCVS) and its predecessor the National Crime Survey (NCS) to the study of crime and development of crime statistics cannot be emphasized enough. For over 40 years, the NCS-NCVS has been one of only two ongoing national measures of crime in the United States that provide annual level and change estimates. The NCS-NCVS not only has played a groundbreaking role in shaping US crime statistics, but also international ones. This entry discusses the rich history of the NCS-NCVS with a focus on its methodological development. This entry also describes the survey's trends and key findings over time and the continued influence of the NCS-NCVS on crime measurement.

Introduction

The importance of the National Crime Victimization Survey (NCVS) and its predecessor the National Crime Survey (NCS) on the study of crime and development of crime statistics cannot be emphasized enough. The NCS-NCVS is one of only two ongoing national measures of crime in the United States that provide annual level and change estimates. In addition, the NCS-NCVS played a groundbreaking role in shaping US crime statistics. Before the NCS, no national crime statistics had been based on self-reported survey data from the general public. The other national source is from the Uniform Crime Reporting Program (UCR), which measures crime using official police records. To provide an understanding of the NCS-NCVS, this entry examines the history and development of the survey with a focus on its methodology. This entry then reviews key findings that have been obtained from the survey as well as a discussion of its effect on the study of crime and measurement of victimization.

History of the NCS-NCVS

In 1965, President Lyndon Johnson appointed two Presidential Commissions with charges that included examining the causes and characteristics of rising crime in urban areas in the United States and recommending policies and programs to reduce this apparent growing crime problem (Cantor and Lynch 2000). The Commissions immediately confronted the problem of a significant lack of information from which to accomplish their charges. The Commission strongly criticized the quality of US crime data at the time and described the situation as one where "the United States is today, in the era of the high speed computer, trying to keep track of crime and criminals with a system that was no less than adequate in the days of the horse and buggy" (President's Commission on Law Enforcement and Administration of Justice 1967, p. 123).

At the time the Commissions were instituted, the only source of national crime data available was from the Uniform Crime Reporting Program (UCR). The Federal Bureau of Investigation collected UCR data from state and local law enforcement agencies. UCR data had four fundamental limitations for studying crime and for the Commissions' purposes. First, as an accounting of crimes known to police, UCR data by design could not identify the nature and extent of offenses that victims did not report. A second limitation was that UCR data were several steps removed from the crime incident and likely reflected law enforcement activity and agency priorities rather than criminal behavior and trends in crime. Sellin (1931) identified this problem shortly after the UCR was implemented. Third, police data could be subject to manipulation and misrepresentation through the classification and counting rules needed to establish uniformity in the UCR data (Beattie 1941). While this problem did not appear to be widespread, well-publicized cases highlighted this concern and brought all UCR data under question (Seidman and Couzens 1974). Finally, UCR data at the time were aggregate counts of crime and as such failed to provide details about the criminal incidents. Without any incident-level details, these data could not identify important information about criminal events nor the victims and offenders involved (Dodge and Turner 1981).

The Commissions reached the conclusion that the United States needed a national survey of crime victimization (President's Commission on Law Enforcement and Administration of Justice 1967). The genesis for using victim surveys to measure crime originated from the UCR data limitations as well as an environment receptive to a self-reported survey-based data collection. This new attitude toward survey research arose due to a convergence of factors that included increased attention on crime as a social problem and overall advances in the field of survey methodology (see Cantor and Lynch 2000, for a discussion). Victimization surveys offered an alternative source of crime data that provided greater incident details than the existing aggregate-based UCR as well as a system independent from, and not influenced by, police practices and policies.

Victimization surveys in theory addressed the data limitation identified by the Commissions. The unknown issue was whether victimization surveys would work in practice, both in terms of the cooperation of respondents in reporting victimization experiences, the reliability and validity of their responses, and the ability to produce annual crime estimates. No other country had relied on self-reported survey data to produce national crime estimates (Hough et al. 2007; UNODC 2010). To test this method, the Presidential Commissions supported three sets of pilot victimization surveys: one household-based sample in Washington, DC; one group of city based household samples; and one national household sample. All of these pilots produced two important findings: confirmation that victim surveys could provide reliable crime estimates and identification that a significant amount of crime that occurred was never reported to the police (Biderman 1967; Cantor and Lynch 2000). These pilot surveys also assisted in identifying methodological issues such as ways to facilitate recall, temporal placement, and collecting crimes against the household (Cantor and Lynch 2000).

Given these findings, the Commission recommended a routine collection of national victimization data on an ongoing basis as well as the infrastructure to collect this data in the form of a national criminal justice statistics center. In 1968, the Law Enforcement Assistance Administration (LEAA) was created within the US Department of Justice and charged with implementing a national victimization survey. The LEAA was the predecessor to today's Bureau of Justice Statistics (BJS), which now oversees the crime survey as well as a full portfolio of other criminal justice statistics. The Bureau of the Census was selected to design and implement the survey on LEAA's behalf.

Based on recommendations from the Commissions and findings from methodological studies, the NCS was created and first implemented in 1972. Originally the NCS was a system of four interrelated victimization surveys: one national household sample (the Crime Panel), a city-level household sample (the Central Cities Survey), and two Commercial Victimization Surveys (one at the national level and one at the city level). The Commercial Victimization Surveys initially included a sample of 2,000 commercial establishments in 26 large cities, and a sample of 15,000 businesses across the nation. The Central Cities Survey included sample from 26 large cities, and the Crime Panel was a national probability sample of 72,000 households. Almost immediately, budgetary and methodological issues beleaguered this design and prompted the removal of three of the surveys. After 1976, only the Crime Panel remained and is what is referred as the NCS. This entry uses NCS to refer to the Crime Panel's national sample of households.

The original goals for the NCS concerned providing an indicator of the crime problem independent from police data and the UCR and an ongoing measure of victimization risk (Rennison and Rand 2007). The NCS also served other purposes that included shifting the focus of the criminal justice system to the victim and away from the offender and providing a measure to assess changes in police reporting (Rennison and Rand 2007). As a result of this close tie to monitoring the UCR and police activities, the crimes covered by the NCS were closely linked with the UCR crimes. The NCS crimes included rape, robbery, aggravated and simple assault, personal larceny (pickpocketing and purse snatching), burglary, motor vehicle theft, and property theft.

Methodological Development of the NCS-NCVS

Original NCS Methodology

The NCS originally was based on a sample of approximately 72,000 housing units and group quarters in the United States. Given this sampling frame, individuals who served as crewmembers of merchant vessels, military personnel housed in military barracks, and institutionalized persons (including but not limited to prisoners) were excluded from the survey. This sample of households was developed based on a rotating, stratified, multistage cluster panel design to take advantage of efficiency and cost-saving qualities of this approach as well as to comport with the Census field staff assignments for its Current Population Survey (Hubble 1995; Lehnen and Reiss 1978).

Each NCS household remained in-sample for three and a half years and was interviewed at 6 month intervals. Sampled households entry into the survey was staggered with an introduction of new panels of households being rotated in and out of the overall sample during each month. The rotation was designed such that one-sixth of the sample was interviewed for the first time, one-sixth for the second time, one-sixth for the third time, and so forth, each month (Lehnen and Reiss 1978). As a result, the overall sample changed on a monthly basis. Because the NCS was a survey of households, individuals or families were not followed if they moved. Thus the composition of households could change due to events such as relocation, marriage, divorce, incarceration, children aging into the sample, or death. In the most extreme case, entire families moved out of housing units and were replaced by new families during the 3 and a half year interview period.

At each 6 month interview period (or enumeration), all persons age 14 or older residing in the household were interviewed either in person or over the phone using a paper and pencil interviewing mode. Household residents ages 12 and 13 were included in the NCS but only were interviewed by proxy. Each household had a single respondent designated as the household respondent. This individual not only answered questions regarding any personal victimization experienced in the 6 month retrospective period, but also questions regarding victimizations against the household, which included burglary, motor vehicle theft and other theft. The household respondent was required to be a household member at least 18 years old who was deemed likely to give accurate answers regarding the household questions. This person was always the first person interviewed in the household during each enumeration. Household respondents could vary within each household across enumerations. All other respondents were questioned regarding their own personal victimization experiences.

The NCS utilized three primary survey instruments: the Control Card, the Basic Screen Questionnaire, and the Crime Incident Report. The Control Card served as the administrative record to gather basic information about the sampled unit including a roster of all persons residing in the household, a record of visits, telephone calls, interviews, and non-interview reasons. The household respondent provided the names, ages, genders, marital statuses, education levels, and relationships of all persons in the household. This information was updated at each interview and included a brief description of all victimizations previously reported by all members of the household. With this information, the Control Card served as a quick reference during later interviews to minimize double-counting of victimizations already reported to the survey during a previous interview.

The NCS collected its victimization data using an innovative two-step process to promote recall. The first step concerned screening for any victimization. Here the Basic Screen Questionnaire ascertained whether any crimes had been committed against the household or individual household members during the previous 6 months. The structure of the NCS screener that was implemented departed from the pilot tests that recommended a short cue approach to facilitate recall (Cantor and Lynch 2000). One reason for this departure was to facilitate matching reported victimization incidents with UCR defined crimes. While the legal jargon of the UCR was not used in the screener, the questions did have a direct correspondence with UCR crimes. For example, screening questions related to assault victimizations asked: "Did anyone beat you up, attack you or hit you with something, such as a rock or bottle?" and "Were you knifed, shot at, or attacked with some other weapon by anyone at all?" If a respondent reported a victimization incident, a separate set of questions would gather details of the event for later determination as to whether a crime (and what type of crime) occurred.

The Crime Incident Report (CIR) collected detailed information about the reported victimizations identified during the screening stage. These details included incident characteristics such as the time, date, and place of the crime; information about the offender including any relationship to the victim; consequences to the victim including injury and loss of or damage to property; and whether the incident was reported to police. The detailed information collected on the CIR was used to determine whether the incident reported was within the scope of the NCS and to classify the type of crime. The NCS used an algorithm based on responses to several questions to classify the crime type.

Counting crimes for the NCS's annual estimation purposes required addressing additional issues in the case of respondents who experienced more than one crime in a particular incident (such as a rape and a robbery) or multiple incidents during the 6 month recall period. For respondents who experienced more than one crime during an incident, the NCS created a hierarchy of crime, where only the most serious crime would be counted. The NCS hierarchy from most to least serious was as follows: rape, robbery, aggravated assault, simple assault, personal larceny (purse snatching and pickpocketing), burglary, motor vehicle theft, and property theft. In the example where a victim was robbed and raped in a single incident, only the rape would be counted for annual estimation purposes. It is important to keep in mind this hierarchy is only for BJS's annual estimation purposes. Public use NCS-NCVS files allow researchers to discern all crimes that occur within a victimization incident (see, e.g., Addington and Rennison 2008).

A second difficulty in crime counting concerns measuring victims who experience on-going or repeated violence. To address the difficulties that these respondents had in remembering needed details of each incident and to minimize the burden of completing a large number of CIRs, the NCS utilized a special protocol for such "series victimizations" Series victimizations referred to three or more victimizations, similar in nature, in which a respondent unable to remember the details in the previous 6 month reference period. In these instances, details of only the most recent incident in the series were recorded in the CIR along with the number of times the person was victimized.

Changes to the NCS Methodology Over Time

The NCS underwent a significant redesign effort (see Cantor and Lynch 2000, for an in-depth discussion). The redesign efforts began almost as soon as the survey went into the field (Cantor and Lynch 2000). The NCS was a significant event for both those interested in survey methodology as well as social statistics. One area of criticism came from academics who challenged the significant financial expenditure on a data system that provided limited analysis options and few variables that could be used for theory testing (Cantor and Lynch 2000). Critiques also came from the law enforcement community who questioned this new survey crime data collection method based on victim self-reports and the associated criticisms of police-based crime data (Cantor and Lynch 2000). At one point, the NCS had attracted enough criticism that the US Department of Justice considered discontinuing the survey until it could be redesigned (Lynch 1990). While the survey was never suspended, it was the subject of a re-evaluation by National Academy of Sciences and a 5 year program of research, instrument development, and redesign planning (Lynch 1990).

This redesign effort focused on developing a better screening instrument to promote respondent recall of victimization incidents and revising the incident report to collect additional details about the victimization event (Cantor and Lynch 2000). Other redesign work concerned research devoted to recall periods, respondent fatigue and other methodological issues (Cantor and Lynch 2000). Research and development of these efforts began in the 1980s and the fully redesigned survey was introduced into the field in 1992.

Before full implementation of the redesigned NCS, certain changes were implemented that addressed methodological issues as well as minimized the cost of the survey. These so-called "near term" changes were limited to those that were not likely to affect the NCS's estimates of annual crime rates for trend purposes also referred to as non-rate affecting. Examples of these changes concerned reducing proxy interviews, changes to the CIR, changes to telephone interviewing, and a name change. In 1986, the survey reduced the number of proxy interviews by directly interviewing persons age 12 or older (vs. 14 or older as it had been originally), and additional questions concerning incident details were added to the CIR. In 1987, centralized computer-assisted telephone interviewing (CATI) started being phased in alongside continued paper and pencil mode (both in-person and over the telephone). Over the years, the number of in-person interviews was reduced to the point where only the first interview is conducted in person. All other interviews are conducted over the telephone unless this mode is not acceptable to the respondent or not feasible due to circumstances such as a household moving into sample. Finally in 1991 prior to the launch of the fully redesigned NCS, the survey changed its name to the National Crime Victimization Survey (NCVS). For the remainder of this entry, the acronyms "NCS" refers to the pre-redesign survey, "NCVS" refers to the post-redesign survey, and "NCS-NCVS" refers to both surveys.

The completely redesigned NCVS went into the field in 1992. The redesigned survey kept the same three instrument format (control card, screener and incident report). Otherwise the survey had been completely overhauled, most noticeably in terms of the screener. The post-redesign screener included a series of short cues to prompt recall. In the NCS, each screener question corresponded to a particular crime, such as assault. In the NCVS, screener questions instead focused on aspects of the victimization such as where it occurred, what the respondent was doing at the time, and who committed the offense. The premise was to facilitate recall and to use various memory triggers to do so. An example of the screening question that focuses on who the offender is the following: "People often don't think of incidents committed by someone they know. (Other than any incidents already mentioned,) did you have something stolen from you OR were you attacked or threatened by – (a) Someone at work or school – (b) A neighbor or friend – (c) A relative or family member – (d) Any other person you've met or known?"

The redesigned screening process also added identification of new crimes (vandalism and sexual assault) and specific prompts were added to better collect data on certain crimes that had been included in the NCS (rape and domestic violence). Rape and domestic violence were more clearly screened by including questions related to sexually based offenses and offenses committed by people known to the respondent. In addition, the redesigned NCVS took advantage of the use of supplements or additional sets of questions beyond the Control Card, Screener and CIR. These supplements were used to collect information on additional crimes beyond those included for annual crime estimation purposes. Supplement topics included workplace violence, school crime, stalking, and identity theft. Other changes that occurred with the redesign concerned repeated (or series) crimes were counted and certain methodological changes (Rennison and Rand 2007).

The 1992 redesign was implemented over an 18-month period using a split-sample design data in which half the sample was collected using the old NCS methodology, and half was collected using the redesigned NCVS methodology. Analyses indicated that the redesigned survey provided a better measure of crime and as expected that the redesign had a greater effect on certain crime types than others. In particular, the redesigned survey and its use of short cues in the screener uncovered more rape and domestic violence and the redesigned survey collected a higher percentage of crime that was not reported to the police than the NCS.

The NCVS has experienced additional changes since its redesign. Until recently, many of these changes centered on efforts to reduce costs as the NCVS has been flat funded for most of its existence since the redesign. The most significant of these changes has been the dramatic cut in sample size (Rennison and Rand 2007). Concurrent with the flat funding and sample cuts, external pressures demanded that the NCVS do more in terms of data collection and speed in releasing data. Questions were added concerning hate crimes, crimes against the disabled, computer crimes and identity theft. The NCVS was also under pressure to produce its annual estimates more quickly (Rennison and Rand 2007).

The deep cuts to the NCVS program generated an unsustainable situation and a request by BJS for the National Academies' Committee on National Statistics (CNSTAT) for a panel to review the NCVS and "consider alternative options for conducting the [NCVS]" (National Research Council 2008, p. 23). The CNSTAT Panel's review found the NCVS was in need of more resources and restoration. Specifically "[a]s currently configured and funded, the NCVS is not achieving and cannot achieve BJS's legislatively mandated goal to 'collect and analyze data that will serve as a continuous and comparable national social indication of the prevalence, incidence, rates, extent, distribution, and attributes of crime . . .'" (National Research Council 2008, p. 78). The CNSTAT Panel offered many recommendations to restore and modernize the NCVS as well as an overall call to encourage appropriate funding of the survey. In response to this report, BJS began conducting a redesign effort to address the recommendations. These efforts have included restoring the sample size to increase precision and reliability, as well as efforts to generate sub-national victimization estimates (BJS, n.d.).

Limitations of the Victimization Survey Methodology

Victimization surveys provide a perspective of crime independent from official police statistics. While this approach offers valuable information for understanding victimization, these surveys in general and with the NCS-NCVS in particular are not without limitations. These limitations can be categorized as ones based on the sampling frame, sampling error, and non-sampling error (see Maltz 1998 for a detailed discussion of these issues).

Certain limitations arise due to the selection of the population to be included in the survey sample. For the NCS-NCVS, this methodological decision does exclude certain crime victims due to selections based on age and residential requirements as well as it being a non-commercial survey. Currently one must be age 12 or older to be eligible for the NCS-NCVS. In addition, only households in the United States are sampled for purposes of the NCS-NCVS. Therefore homeless persons, those living in military barracks, incarcerated individuals, persons visiting from out of the country and other institutionalized persons are excluded from the survey. The NCS-NCVS only covers non-commercial crimes since businesses are not included in the population sampled. As such, while a household burglary is counted, a burglary of a business is not.

Like all surveys, the NCS-NCVS are subject to sampling error. Depending upon the crime and population of victims considered, this error can be large (Maltz 1998; Rand and Rennison 2002).

The NCS-NCVS is also subject to nonsampling errors including those related to inaccurate respondent recall. Respondents may forget criminal events, or be in error about characteristics of a recalled event including when it happened, who was involved and what transpired. In some instances victims may refuse to share details of criminal events especially if the offender is a relative or intimate partner.

Key Findings from NCS-NCVS Data Over the Years

The NCS-NCVS is a rich resource of information on crime trends and incident details. In terms of long-term trends, NCS-NCVS data show that non-fatal violence most recently peaked in the early 1990s, and has since decreased dramatically to rates unseen for decades (Bastian 1995; Ringel 1997; Truman 2011). In addition to identifying overall long-term trends, the NCS-NCVS data are uniquely suited to provide incident level details over time for US crime rates. When these details are examined, similar trends are observed across different types of victims, offenders and crimes (Bastian 1995; Ringel 1997; Truman 2011).

One unique aspect of the NCS-NCVS in comparison to the UCR is the collection of data regarding patterns of reporting victimizations to police. The pre-NCS pilots highlighted the large proportion of crimes that were not reported to police and prompted Biderman and Reiss's groundbreaking 1967 article that identified this "dark figure of crime." The underreporting of crime continues to this day. The NCVS data indicate that a higher percentage of violent crimes come to the attention of police than property crimes (50 % vs. 40 %) (Catalano and Rand 2007; Rand 2009, 2008; Truman 2011; Truman and Rand 2010). Within violent crime, the percentage of victimizations reported to the police varies. For example approximately 57 % of robberies and 55 % of aggravated assaults are reported to police (Hart and Rennison 2003). Historically, rape and sexual assault is the violent crime least likely to come to the attention of the police as only about one-third are reported (Hart and Rennison 2003). Finally, less than half (about 44 %) of simple assault tend to be reported to the police (Hart and Rennison 2003). In contrast, while motor vehicle thefts are the least common type of property crime, it is the individual crime most consistently reported to the police (Catalano and Rand 2007; Hart and Rennison 2003; Rand 2009, 2008; Truman 2011; Truman and Rand 2010).

Data from the NCS and NCVS also provide the basis for current understandings about victimization, particularly the relationship between victimization and particular individual characteristics. With the exception of rape and sexual assault, males are more likely to be victimized than females (Catalano and Rand 2007; Rand 2009, 2008; Truman 2011; Truman and Rand 2010). Those with less education, a lower income, and who are younger are at a higher risk for all forms of victimization than the well-educated, those with higher incomes and older individuals (Catalano and Rand 2007; Rand 2009, 2008; Truman 2011; Truman and Rand 2010). Most non-fatal violence is not committed with a weapon, but when it is, a firearm is most often used; and similarly most victims are not injured as a result of non-fatal violence that is reported to the NCS-NCVS (Rand 2009, 2008; Rennison 1999; Truman 2011; Truman and Rand 2010). The NCS-NCVS also put victimization into context and challenged popular beliefs. Stranger violence against females is less common than is violence by a known offender (Rand 2009, 2008; Rennison 1999; Truman 2011; Truman and Rand 2010). The bulk of non-fatal violence is intra-racial and occurs close to the victim's home (Truman and Rand 2010; Rand 2009, 2008).

In addition to the core NCS-NCVS questions, the NCS-NCVS supplements have allowed unique insights with regard to particular types of victimization including information on characteristics of workplace violence, types of identity theft, and trends in school violence. The workplace violence supplement shows that law enforcement officers, security guards and bartenders as those who experience the highest rates of workplace violence (Harrell 2011). Data from the NCVS's identity theft supplement shows that from 2005 to 2010 credit card theft was the most common type of identity theft occurring in a household (Langton and Planty 2010). With regard to school violence, regular fielding of the NCVS's School Crime

Supplement shows that school violence has been declining at a steady rate since 1992 (Robers et al. 2012).

Effect of NCS-NCVS on Crime Data Collection

The NCS-NCVS have been influential on the measurement of crime and collection of crime statistics. One of the most immediate effects of the NCS was not on a survey data collection, but on the UCR and its use of police records to collect crime data. The incident details provided by the NCS highlighted the limitations with the UCR's aggregate-level system. With the available details from the NCS, researchers and policymakers began to demand the production of police data that provided incident-level information needed to explain the crime problem (Poggio et al. 1985). This situation prompted the creation of the National Incident-Based Reporting System (NIBRS) as a new method for collecting UCR data. In addition to expanding the number of crimes reported to the UCR, NIBRS captures incident-level characteristics not included in the summary reporting system.

Given the groundbreaking nature of the NCS-NCVS as demonstrating the feasibility of survey data as a resource for generating national crime statistics, the NCS-NCVS also have influenced the development of omnibus crime data collections abroad. In particular, the methodological experiences gained from the NCS-NCVS informed victimization surveys in other countries. For example, both the idea and the model for the British Crime Survey came from the NCS (Hough et al. 2007). Since the NCS began in 1972, the number of crime surveys has flourished. According to one inventory sponsored by the United Nations Economic Commission for Europe and the United Nations Office on Drugs and Crime, there were approximately 62 country-wide victimization surveys as of 2006. Approximately two-thirds of these surveys were purely victimization surveys (e.g., Scottish Crime Survey, British Crime Survey, Northern Ireland), while the remainder were victimization components of more general surveys (e.g., Australia and Canada). The NCS also influenced cross-national studies of crime. Since 1989, the International Crime Victimization Survey (ICVS) has provided for a comparative approach to understanding crime and victimization without the limitations associated with official records by gathering victimization data for several types of crime from participating nations. Since its inception, 78 countries have participated in the ICVS at least once, and more than 300,000 people have been interviewed.

Conclusion

The NCS-NCVS fundamentally changed how crime is measured and shaped how crime data are collected. The history of the survey highlights the wealth of research that has been devoted to the initial development of the NCS and its redesigned successor. The surveys have shaped knowledge about crime and the experience of victims as well as provided important insights regarding trends in crime over time and details about particular aspects of victimization. The future of the NCVS is one of continued evolution both in response to new knowledge about crime and a continued demand for more detailed data. Current recommendations for the NCVS focus on considering including data collection on emerging crime types such as identity theft as well as expanding the use of supplements to collect detailed information on a rotating basis (NRC 2008). The NCVS must balance these demands for expansion and change with ongoing concerns particularly its historic primary goal of generating annual level and change crime estimates.

Recommended Reading and References

Addington LA (2009) Studying the crime problem with NIBRS data: current uses and future trends. In: Krohn MD, Lizotte AJ, Hall GP (eds) Handbook on crime and deviance. Springer Science + Business Media, New York

Addington LA, Rennison CM (2008) Rape co-occurrence: do additional crimes affect victim reporting and police clearance of rape? J Quant Criminol 24(2):205–226

Bastian L (1995) Criminal victimization, 1993. U.S. Government Printing Office, Washington, DC
Beattie RH (1941) The sources of criminal statistics. Ann Am Acad Polit Soc Sci 217:19–28
Biderman AD (1967) Surveys of population samples for estimating crime incidence. Ann Am Acad Polit Soc Sci 374:16–33
Biderman AD, Reiss AJ Jr (1967) On exploring the 'dark figure' of crime. Ann Am Acad Polit Soc Sci 374:1–15
BJS (n.d.) NCVS redesign. http://bjs.ojp.usdoj.gov/index.cfm?ty=tp&tid=91 Accessed 4 Feb 2012
Cantor D, Lynch JP (2000) Self-report surveys as measures of crime and criminal victimization. In: Duffee D (ed) Measurement and analysis of crime and justice. National Institute of Justice, Washington, DC, pp 85–138
Catalano S, Rand M (2007) Criminal victimization, 2006. United States Department of Justice, Washington, DC
Dodge RW, Turner A (1981) Methodological foundations for establishing a national survey of victimization. In: Lehnen RG, Skogan WG (eds) The national crime survey: working papers, volume I: current and historical perspectives (NCJ-75374). U.S. Department of Justice, Washington, DC
Harrell E (2011) Workplace violence, 1993–2009. U.S. Government Printing Office, Washington, DC
Hart TC, Rennison CM (2003) Reporting crime to the police, 1992–2000. United States Government Printing Office, Washington, DC
Hough M, Maxfield M, Morris B, Simmons J (2007) The British crime survey over 25 years: progress, problems and prospects. Crime Preven Stud 22:7–31
Hubble DL (1995) The national crime victimization survey redesign: new questionnaire and procedures development and phase-in methodology. http://www.amstat.org/sections/srms/proceedings/papers/1995_009.pdf. Accessed 31 March 2012
Langton L, Planty M (2010) Victims of identity theft, 2008. U.S. Department of Justice, Washington, DC
Lehnen RG, Reiss AJ (1978) Response effects in the national crime survey. Victimology 3:110–160
Lynch JP (1990) The current and future national crime survey. In: MacKenzie DL, Baunach PJ, Roberg RR (eds) Measuring crime: large-scale, long-range efforts. SUNY Press, Albany
Maltz M (1998) Displaying violent crime trends using estimates from the national crime victimization survey. U.S. Department of Justice, Washington, DC
National Research Council (2008) Surveying victims. National Academies Press, Washington, DC
Poggio EC, Kennedy SD, Chaiken JM, Carlson KE (1985) Blueprint for the future of the uniform crime reporting program: final report of the UCR study. U.S. Department of Justice, Washington, DC
President's Commission on Law Enforcement and Administration of Justice (1967) Task force report: crime and its impact – an assessment. U.S. Government Printing Office, Washington, DC
Rand M (2008) Criminal victimization, 2007. U.S. Department of Justice, Washington, DC
Rand M (2009) Criminal victimization, 2008. U.S. Department of Justice, Washington, DC
Rand MR, Rennison CM (2002) True crime stories? Accounting for differences in our national crime indicators. Chance 15(1):47–51
Rennison CM (1999) Criminal victimization, 1998: changes 1997–98 with trends 1993–98. U.S. Department of Justice, Washington, DC
Rennison CM, Rand M (2007) Introduction to the national crime victimization survey. In: Lynch JP, Addington LA (eds) Understanding crime statistics: revisiting the divergence of the NCVS and the UCR. Cambridge University Press, Cambridge
Ringel C (1997) Criminal victimization 1996, changes 1995–96 with trends 1993–96. U.S. Department of Justice, Washington, DC
Robers S, Zhang J, Truman J, Snyder T (2012) Indicators of school crime, 2010. U.S. Department of Justice, Washington, DC
Seidman D, Couzens M (1974) Getting the crime rate down: political pressure and crime reporting. Law Soc Rev 8:457–493
Sellin T (1931) The basis of a crime index. J Am Inst Crim Law Criminol 22:52–64
Truman J (2011) Criminal victimization, 2010. U.S. Department of Justice, Washington, DC
Truman JL, Rand MR (2010) Criminal victimization, 2009. U.S. Department of Justice, Washington, DC
UNODC (2010) Manual on victimization surveys. Jointly promoted by the United Nations Economic Commission for Europe (UNECE) Statistical Division and the United Nations Office on Drug and Crime (UNODC). http://www.unece.org/fileadmin/DAM/publications/oes/STATS_Manual_Victimization_surveys_E.pdf. Accessed 27 Mar 2012

US Responses to Terrorists

Brent L. Smith[1], Christopher Shields[2] and Kelly R. Damphousse[3]
[1]Department of Sociology and Criminal Justice, University of Arkansas, Fayetteville, AR, USA
[2]Terrorism Research Center in Fulbright College, University of Arkansas, Fayetteville, AR, USA
[3]University of Oklahoma, Norman, OK, USA

Overview

America's response to terrorism has changed dramatically over the past 30 years. Changes have included everything from the way in which terrorism is portrayed politically to the manner in which terrorists are investigated, prosecuted, and punished. Modifications to governmental interventions also have affected the manner in which terrorists plan and conduct terrorist activities as well as the manner in which they defend themselves in court. Although it would be difficult to address all of these changes in a single paper, the purpose of this current entry is to provide an overview of the most significant events that evoked changes in the manner in which terrorists are portrayed, pursued, and prosecuted as well as the way in which terrorists and their defenders have responded to federal prosecutorial efforts. The specific ways in which federal agencies respond to terrorism, however, are rooted in much larger political and social issues.

American Terrorism and Governmental Response in Historical Context

For whatever reason, the United States has avoided the concept of "political crime" and "political criminality." Some have contended that this is due to ideological considerations – that America has tried to portray itself as a nation characterized by consensus and, as a result, America is simply immune to violent political conflict (Ingraham and Tokoro 1969; Turk 1982). For many years, terrorism and political crime were associated almost exclusively with the extreme left and Marxist revolutionaries. The Cold War, communism, Fidel Castro's support of revolutionaries in the United States, and the civil rights and student antiwar movements of the 1960s and 1970s shaped America's views on those who sought political change through violence and terrorism. Terrorists came to be described as exclusively young, Marxist, urban, educated revolutionaries who were bent on destroying capitalism (Russell and Miller 1977; Smith and Morgan 1994). The potential threat posed by these revolutionaries was not lost on the FBI. Efforts to suppress the activities of the Black Panthers and other leftist groups through the FBI's now infamous counterintelligence program (COINTELPRO) of the late 1960s and early 1970s are well documented (Poveda 1990).

Abuses within these programs and the impact of Watergate in the mid-1970s severely tarnished the public's image of the FBI. Confidence in the agency was shaken, and Congress and the American people demanded change. In the wake of the post-Watergate investigation, two significant events occurred that helped shape the definition of American terrorism and how the federal government responds to it. First, in April 1976, new FBI investigative guidelines were implemented under the guidance of then Attorney General Edward Levi. The guidelines identified the standards by which internal security investigations could be initiated and the length of time they could last (Hearings 1978). As evidence of this change, the number of domestic security investigations dropped from more than 20,000 in 1973 to less than 200 in 1976 (Elliff 1979). Second, in August 1976, the FBI dismantled its domestic intelligence units, moving investigations of domestic terrorism from its Intelligence Division to the General Investigative Division (Kelley and Davis 1987; Poveda 1990). This move limited the types of investigative techniques that could be used in terrorism cases to the standards used for traditional crimes.

The impact of these changes was two-fold. First, it reaffirmed that terrorism in the United States was to be viewed as "conventional crime" and that terrorists were to be treated as such. Second, it marked the end of an era in FBI history in which the agency focused on domestic intelligence gathering. The next generation of FBI agents generally was trained to avoid data collection on American citizens. The FBI adopted a much more "reactive" stance, investigating terrorist groups only when a "criminal predicate" could be established. The more proactive "intelligence gathering" days of the pre-Watergate and COINTELPRO era were replaced with a desire to restore the FBI image and characterized by a reluctance to invade the "personal privacy" of domestic extremist groups. Domestic terrorist groups that survived the FBI's COINTELPRO operated almost with impunity during the late

1970s and early 1980s. During this time, the number of terrorism incidents in the United States reached its zenith according to the FBI's annual reports on *Terrorism in the United States*. (It should be noted that the title of the FBI's annual reports has changed over the years. Initially, they were called *FBI Analysis of Terrorist Incidents in the United States*. In 1984, the title was changed to *FBI Analysis of Terrorist Incidents and Terrorist Related Activities in the United States*. The 1986 report reverted to the pre-1984 title. In 1987, the FBI adopted the title *Terrorism in the United States* followed by the year of the report.) Leftist terror continued unabated, Puerto Rican separatists targeted both the island and the US mainland, and far-right terrorism emerged partly as a response to gains made by African Americans as part of the civil rights movement. But like Watergate and COINTELPRO, which triggered the changes described above, new events would swing the pendulum to the other direction.

The Impact of President Reagan on Counterterrorism Policy

With the election of Ronald Reagan in 1980, a substantially different trend in federal response to terrorism began to emerge. A staunch conservative President Reagan saw leftist and Puerto Rican separatist groups as significant threats to American security. The FBI, however, was reluctant to reengage in domestic intelligence. Congressional criticism of the FBI subsequently mounted. A seminal event would trigger renewed FBI vigilance. In October 1981 in Nyack, New York, an armored truck was robbed in an incident that left two police officers dead. The robbery involved members of the long-forgotten Weather Underground and Black Liberation Army, which had also been providing assistance to the Puerto Rican separatist group, the Armed Forces of National Liberation (FALN). Later that year, members of another holdover leftist group from the mid-1970s, the United Freedom Front, killed a New Jersey state trooper. The leftist threat suddenly appeared to be quite real.

These and other events resulted in revision of the Levi guidelines for FBI investigations. The new guidelines, issued by Attorney General William French Smith in 1983, set the standard for FBI investigations of domestic terrorism groups for the next 20 years. Although minor modifications were made by subsequent attorney generals, the theme and spirit of the "Smith Guidelines" remained intact until modifications after the September 11 attacks in 2001. After the Smith guidelines were issued in the spring of 1983, the FBI implemented a series of counterterrorism initiatives, including the creation of counterterrorism task forces around the country to combat specific regional and local threats. The success of these efforts was remarkable. Leftist groups like the Weather Underground and its radical East Coast offshoot, the May 19th Communist Organization, the United Freedom Front, and the Puerto Rican groups, the Macheteros and the FALN, were decimated by arrest, indictment, and convictions in federal courts. The extreme right met a similar fate. Leading members of the Order, the Sheriff's Posse Comitatus, the White Patriot Party, and the Aryan Nations, among others, were indicted and convicted in a series of dramatic trials in the mid to late 1980s.

The Smith Guidelines were quite explicit – terrorism investigations were to be "concerned with the investigation of *entire enterprises, rather than individual participants* [emphasis added]." Furthermore, terrorism investigations could not be opened unless "circumstances indicate that *two or more persons* [emphasis added] are engaged in an enterprise for the purpose of furthering political or social goals... that involve force or violence and a violation of the criminal laws of the United States" (Office of the Attorney General 1983). Although the Smith Guidelines provided an opportunity for the FBI to reestablish itself in collecting domestic intelligence, the new policy was clearly intended to prevent a return to the abuses of the past. Furthermore, the FBI continued to self-impose restrictions on when terrorism investigations could be opened. Although the new guidelines did not mandate the necessity of a "criminal predicate," the term came to be the standard descriptor used by agents when discussing the threshold for opening a terrorism enterprise investigation.

For domestic terrorists, the Smith Guidelines had a significant impact in four major ways. First, a "focus on groups rather than individuals" would later lead terrorist groups in the USA to implement "leaderless resistance" or lone wolf strategies in an effort to avoid terrorism investigations. Second, the guidelines made it clear that dismantling a terrorist organization, primarily by decimating its leadership, was more important than convicting the actual perpetrators of specific terrorist incidents. This approach had a significant impact on negotiated pleas as a result of plea offers to cadre members in exchange for governmental cooperation against key leaders. Third, the FBI continued its "reactive" policy begun under Levi, as evidenced by a reluctance to get involved in domestic intelligence gathering. The number of active terrorism investigations generally ranged from 8 to 12 in any given year (although the actual number and names of groups investigated is classified) from 1983 to the end of the century. Finally, the FBI and Department of Justice continued to maintain the position that terrorists should be perceived, portrayed, tried, and punished as "conventional" criminals (Smith and Damphousse 1996).

Complicating this scenario was a new emphasis on international terrorism under Reagan's administration. As a result of the bombing of the US Marine Corps barracks in Beirut in 1983 and the bombing of the LaBelle Disco in Germany, which targeted American service members, Reagan pushed for greater authority and jurisdiction in dealing with terrorism overseas. With passage of the Comprehensive Crime Control Act of 1984 and the Omnibus Diplomatic Security and Antiterrorism Act of 1986, the federal courts now assumed jurisdiction over terrorists who committed acts of terrorism against US citizens or property overseas. These persons were typically investigated under a different set of FBI investigative guidelines – the *Attorney General Guidelines for FBI Foreign Intelligence Collection and Foreign Counterintelligence Investigations* (FBI 1998). While classified, it is clear that investigations involving international terrorists provided greater latitude to investigators than investigations under the domestic guidelines. Essentially, a "dual system" was emerging in which domestic terrorists were portrayed as "common criminals," while international terrorists were much more likely to be portrayed and described as "terrorists" (Smith 1994). This line of reasoning emerged in the 1990s and was in place when international threats to the US homeland began to become a reality.

America Awakens to the Threat of Terrorism

Despite having a long history of domestic terrorism, prior to the 1990s, most Americans viewed terrorism as something that happened in Europe and the Middle East. We typically viewed ourselves as immune from the leftist-anarchist terror associated with the Red Brigades, Black September, and the Baader-Meinhof Gang. The first World Trade Center bombing in 1993 and the Murrah Federal Building bombing in 1995 awakened the American public to the catastrophic damage that terrorists can inflict on a nation. Although the latter event precipitated significant focus on the extreme right and the so-called militia movement in the late 1990s, the first WTC attack was a seminal event for federal officials. Although Islamic extremists had targeted Americans overseas since the Iranian Revolution in 1979, the US mainland had remained relatively secure. That perception changed for federal officials in 1993. Subsequent plots uncovered over the years that followed relating to the bombing of New York City landmarks further emphasized that the threat of international terrorism on American soil was no longer merely a "threat."

Congress responded by passing the Antiterrorism and Effective Death Penalty Act of 1996. One of the most notable provisions of the Act involved the designation of "foreign terrorist organizations" (FTOs) and penalties for providing material support to any group on this official list. For scholars studying those indicted for terrorism-related activities in the United

States, the law created a new, and fairly large, group of persons indicted in federal courts. Over the next decade (1996–2006), over 325 persons would be indicted for providing material support to foreign terrorist groups. Emphasis on identifying these perpetrators blurred the previously rather simple distinctions between "domestic" and "international" terrorism. In the past, almost all domestic terrorists had been American citizens, while the overwhelming majority of international terrorists had been noncitizens. By the turn of the century, identifying who was being investigated under which set of Attorney General's (AG) Guidelines was becoming "muddied." After the attacks of September 11, 2001, it became even more difficult to distinguish.

Prior to 9/11, most domestic terrorists were portrayed in court as common criminals. In most cases, every effort was made to avoid raising the specter of a "political trial." Common practices used by both defense and prosecutors included filing motions "in limine" to avoid mention of the word "terrorist," the name of the terrorist group, or the ideology to which the group adhered; avoiding seditious conspiracy charges; and using "presumed" or "strict liability" counts that did not require mention of the terrorists' intent in order to obtain a conviction. Prosecutors had learned over the past 20 years that discussion of "motive" in terrorism trials could be extremely hazardous to conviction rates.

In contrast, international terrorists were usually portrayed quite differently. In many cases, their trials were explicitly politicized, and the defendants were frequently called "terrorists," "revolutionaries," and other terms that evoked a threat to national security. The successful conviction of the first WTC bombers and their associates in 1995 represented the first successful prosecution of terrorists on seditious conspiracy charges in over a decade.

Policy Changes After September 11

After September 11, the strategies used to combat terrorism in the United States underwent dramatic change. In broad terms, American antiterrorism policy is now focused on terrorist organizations, affiliated networks, and state sponsors in an effort to identify potential terrorist threats and *proactively prevent* future attacks. In a seminal statement released after September 11, Attorney General John Ashcroft explained that the policy of the United States Government changed from prosecuting terror-related crimes that had already occurred to thwarting attacks before they could happen (Ashcroft 2002).

Prior to this time, the FBI had remained primarily a "reactive" agency – in which "preventive" intelligence gathering played second fiddle to the more conservative standard associated with a "criminal predicate." After September 11, 2001, Congressional and public complaints regarding why the FBI did not have an intelligence apparatus in place to prevent such an attack were commonplace, and the nation moved quickly to encourage intelligence gathering and intelligence sharing.

With these new proactive goals in mind, the US Congress and Executive Branch agencies changed a number of policies to provide new tools for combating terrorism. The creation of fusion centers, the National Counterterrorism Center, and passage of new, less restrictive investigative guidelines by Attorney General Ashcroft reflected this paradigm shift. Some policies expanded legal authority to engage international terrorism that occurs away from American soil (Perl 2003). Other policies expanded legal authority to intercept, investigate, and prosecute domestic terrorists.

The research presented here focuses on changes in domestic policies that affected the prosecution of terrorists. For example, the Antiterrorism and Effective Death Penalty Act was revisited, and the USA PATRIOT Act of 2001 (Patriot Act) was created to extend and strengthen US antiterrorism policy. In addition, the Executive Branch changed Department of Justice policy on how the FBI and US attorneys would handle the investigation and prosecution of terror suspects, namely, through the aforementioned Ashcroft guidelines.

Pre-9/11 AG investigative guidelines (Smith) had required FBI field offices to refer potential terrorism investigations involving two or more persons to the director or assistant director of the FBI; they, and only they, could authorize a "terrorism enterprise" investigation. Once the director authorized such an investigation, he had to report that fact to the Office of Intelligence Policy and Review. The prior guidelines also required the director or another top official to monitor the progress of the investigation at 180 day intervals. Section (B)(4)(a) of the Ashcroft Guidelines loosened those standards by allowing agents in the field to authorize a terrorism investigation for a period of up to 1 year (unlike the Smith Guidelines, the Ashcroft Guidelines allowed the Special Agent in Charge to renew the investigation without additional authorization from FBI HQ). While a field office was required, within 1 year, to report to FBI HQ any terrorism investigations it initiated and provide reports, permission to open an investigation was no longer necessary. The research that follows examines how these changes impacted government interventions in, and outcomes of, federal terrorism cases.

Terrorism on Trial

Defendants investigated as suspected terrorists under the Attorney Generals' Guidelines and later charged with crimes in federal court have been the focus of empirical research for more than 20 years, in part, because they behave differently than similarly situated federal defendants and they have been processed differently by the federal criminal justice system. Each phase, from investigation strategies and charging decisions, to prosecution and defense strategies and their relationship to negotiated pleas and trial outcomes, to conviction rates and sentencing outcomes, deserves separate treatment.

Investigating and Indicting Suspected Terrorists

In the early 1980s, the government's response focused on infiltrating terrorist cells, either by undercover government agents or by turning rank and file members of suspected terrorist organizations into confidential informants, and gathering evidence of planned operations and attacks for the purpose of charging group leaders with conspiracy crimes centered on those plans. The impetus on indicting leaders and putting them behind bars was driven by the policy goal of tearing the organizations apart from the top down. By targeting group leaders, the government hoped to disrupt the groups and help assuage the threat of their radical agendas.

US Responses to Terrorists, Table 1 Proportion of cases with confidential informant

Era	Proportion	N	Std. deviation
Pre-9/11	0.59	133	0.493
Post-9/11	0.14	86	0.349

$t(102.2) = 7.430$, $p < 0.001$ 20 system missing pre-9/11, 40 system missing post-9/11

US Responses to Terrorists, Table 2 Proportion of cases with undercover agents

Era	Proportion	N	Std. deviation
Pre-9/11	0.29	130	0.457
Post-9/11	0.03	86	0.185

$t(185.9) = 4.96$, $p < 0.001$ 23 system missing pre-9/11, 40 system missing post-9/11

This policy was time-consuming, requiring time to infiltrate and build relationships with group members, but prior to September 11, the FBI and other federal agencies had successfully infiltrated groups to the extent that 29 % of all cases filed in federal court involved an undercover agent (see Table 1). Federal agencies were even more successful in convincing group members to turn states evidence. Prior to September 11, prosecutors were using evidence provided by confidential informants in 59 % of all federal terrorism cases (see Table 2).

The post-September 11 policy shift ushered in new strategies directing federal investigators and prosecutors to arrest and prosecute suspected terrorists more proactively. In an effort to disrupt potential threats, the policy goal of infiltrating groups and building conspiracy cases gave way to prosecuting suspected terrorists for any crime as soon as there was sufficient evidence to do so. This shift had a profound and predictable impact on the type of investigations the government conducted, as well as the proportion of cases involving successful infiltration. As Tables 1 and 2 suggest, after September 11, the government made use of confidential informants in just 14 % of cases filed against suspected terrorists and undercover agents in only 3 % of those cases. It should be noted that Smith et al. (2011) found no significant drop in the amount or quality of evidence provided by confidential informants after September 11, though the average number of confidential informants per case did decrease.

Fewer informants and undercover agents to provide evidence resulted in prosecutors moving forward with a smaller proportion of cases in the post-September 11 era that were based on evidence of complex conspiracies. In turn, prosecutors relied on a greater variety of statutes to convict suspected terrorists (Shields 2008; see also, Chesney 2007; and Bradley-Engen et al. 2011). Indeed, the number of cases involving planned acts of terrorism dropped substantially after September 11. Shields (2008) and Jackson (2011) determined that the proportion of cases linked to planned or completed terrorism events declined from 84.2 % to 29.9 % after the policy change.

Not only did the government shift the types of cases it pursued, it pursued more of them. From 1980 to 2001, the government prosecuted approximately 500 individuals after they were referred by the FBI following a domestic security/terrorism investigation. According to Smith et al. (2011) and Shields (2008), over 500 individuals were similarly referred and prosecuted in the first 4 years following September 11, 2001, and that number increased to nearly 1,000 by the end of 2009 (see also, Greenberg 2011). In addition, the average number of defendants per case decreased after September 11, dropping from ten to four defendants, and the average number of counts per indictment decreased from 13.8 to 7.9 (Shields 2008; Smith et al. 2011). The average severity of charges filed against terrorist defendants also decreased (see, Shields 2008; Jackson 2011; and Smith et al. 2011). Post-September 11 policy changes also impacted the prosecutorial and defense strategies used once suspected terrorists were indicted.

Prosecuting and Defending Suspected Terrorists

Among the more intriguing aspects of prosecutions involving suspected terrorists are the strategies used by prosecutors to convict them and the strategies used by defense attorneys to counter federal prosecutors. Smith (1994) observed that in some cases federal prosecutors explicitly politicize the ideological beliefs and political motives of a suspected terrorist defendant. Prosecutors using this strategy frequently seek conspiracy charges that openly tie the predicate criminal act to the defendant's group affiliation or ideological beliefs. This, in turn, is typically done in concert with labeling the defendant as a member of an extremist group or the follower of an extremist ideology in charging documents, motions, and in open court.

Federal prosecutors used this strategy in the majority (64 %) of the cases they pursued before September 11 (Shields 2008; Smith et al. 2011). Prosecutors also employed a second, subtler strategy that implied a link between the defendant and terrorism. Prosecutors using this strategy seek counts charging defendants with any number of crimes, but unlike the previous strategy, prosecutors make no attempt to link the defendant's alleged crimes to terrorism in the indictment. Rather, prosecutors drop "hints" during open court implying the defendant belonged to an extremist group or held extremist views. Shields (2008) reported the use of these hints at all phases of the criminal prosecution, from bail determinations to sentencing hearings. Prior to September 11, prosecutors used this strategy about 25 % of the time.

The third strategy, the conventional approach, involved treating suspected terrorist defendants as traditional offenders, and in doing so, prosecutors avoid any reference to the defendant's group affiliation or belief system. Smith (1994) observed cases where federal prosecutors filed motions in limine to prevent defendants from using their trials as a soapbox to draw attention to their cause. While the conventional strategy was only used in one out of ten cases before September 11, prosecutors relied on it in half of all cases afterward (see Table 3). There has been

US Responses to Terrorists, Table 3 Crosstab prosecution strategy before and after 9/11

	Conventional criminality	Implicit politicality	Explicit politicality	Total
Pre-9/11	57	132	340	529
	10.7 %	24.9 %	64.0 %	100.0 %
Post 9/11	117	37	79	233
	50.2 %	15.9 %	33.9 %	100.0 %
Total	174	169	419	762
	22.8 %	22.2 %	55.0 %	100.0 %

$X^2 = 135.970$, df $= 2$, $p < 0.0001$ 21 system missing (fugitive, awaiting trial, etc.)

US Responses to Terrorists, Table 4 Crosstab defense strategy before and after 9/11

	Defense strategy			
Case type	Political persecution	Disassociate	Conventional	Total
Pre-9/11	105	153	208	466
	22.5 %	32.8 %	44.6 %	100.0 %
Post 9/11	29	20	159	208
	13.9 %	9.6 %	76.4 %	100.0 %
Total	134	173	367	674
	19.9 %	25.7 %	54.5 %	100.0 %

$X^2 = 62.258$, df $= 2$, $p < 0.0001$, 109 system missing

a marked and corresponding decrease in the use of both politicized strategies (explicit 33.9 %; implicit 15.9 %) after September 11.

Shields (2008) and Smith et al. (2011) have suggested that the increased use of conventional prosecution strategies may be a result of the aforementioned proactive policy to prosecute suspected terrorists earlier leading to a decrease in the number of confidential informants and the evidence they would provide. Whether that is the case, defense attorneys have responded by using more traditional defense strategies and, as will be discussed below, defendants are pleading guilty at a higher rate.

In response to politicized prosecution strategies, defense attorneys developed a number of strategies calculated to counter prosecutors. Defense attorneys who presented evidence that their clients were the focus of political persecution because of their beliefs dropped from 22.5 % before September 11 to 13.9 % afterward (see Table 4). Similarly, defense attorneys who tried to disassociate their clients from a terrorist group or ideology decreased from 32.8 % to 9.6 % after September 11. Probably due to prosecutors increased reliance on conventional prosecution strategies, defense attorneys turned to conventional defense methods in greater numbers, from 44.6 % to 76.4 %, after September 11.

Case Outcomes in Terrorism Trials

Differences in the investigation and indictment of suspected terrorists, as well as the strategies used to prosecute and defend them, resulted in significant differences in case outcomes. While the conviction rates for indicted defendants increased slightly from 77.1 % in the pre-September 11 era to 78.1 % in the post-September 11 era, the means by which those defendants were convicted differed dramatically (see Table 5).

The biggest change occurred in the proportion of defendants who entered guilty pleas versus those who went to trial between the eras. While less than one-half (43.2 %) of defendants indicted prior to September 11 pleaded guilty, more than two-thirds (66.5 %) did so after September 11. Mistrials and dismissals increased after September 11 from 14.7 % to 20.0 %, and jury convictions decreased from 33.9 % to 11.5 %.

It is possible that the "early prosecution" mandate demanded by Attorney General Ashcroft impeded the government's ability to infiltrate extremists groups with agents, and likewise, it may have limited the amount of time government agents had to develop relationships with potential informants who were associated with group members. In preliminary studies, Shields (2008), Jackson (2011), and Smith et al. (2011) do suggest a significant shift in the way the government pursued and prosecuted terrorists after September 11.

The policy shift may have been responsible for limiting the amount of evidence available to prosecutors. Shields (2008) suggested that less evidence probably caused a shift in the type of cases prosecutors pursued and the type of prosecution strategies they employed resulting in an increase in negotiated pleas. It is also possible that these lower levels of evidence might be the cause of the noted increase in the number of case

US Responses to Terrorists, Table 5 Disposition of defendants before and after 9/11

Era	Dismiss/mistrial	Acquittal at trial	Jury conviction	Plead guilty	Total
Pre-9/11	73	40	168	214	495
	14.7 %	8.1 %	33.9 %	43.2 %	100.0 %
Post-9/11	48	3	27	155	233
	20.6 %	1.3 %	11.6 %	66.5 %	100.0 %
Total	121	43	195	369	728
	16.6 %	5.9 %	26.8 %	50.7 %	100.0 %

$X^2 = 62.148$, df $= 3$, $p < 0.0001$, 55 system missing (fugitive, transferred, etc.)

dismissals. Significant changes in sentencing occurred as well.

Sentence Outcomes for Convicted Terrorists

While conviction rates increased after September and, ostensibly, defendants were charged with slightly less severe lead offenses, Smith et al. (2011) noted an average drop of 138 months in average prison sentences after September 11, from 203 to 65 months. While the gap narrowed to 120 months when immigration and financial fraud cases were excluded from the post-September 11 cases, there remained a rather dramatic and statistically significant difference between the eras. These findings are not unique (see also, *Terrorism Trial Report Card*, Center for Law and Security 2010).

It is important to note that it is somewhat inappropriate to simply use "sentence in months" as an indicator of changes in sentence length over the past 30 years due to changes in federal sentencing procedures. Changes in federal sentencing policies render direct comparisons tenuous at best. One goal of the federal sentencing guidelines was to reduce disparity among "similarly situated" defendants. Smith (1994) and Smith and Damphousse (1996) addressed this issue and found that individuals involved in terrorism cases received sentences that were, on average, about three and one-half times longer than others convicted of the same lead offenses. A later examination of post-sentencing guidelines in terrorism cases found that this disparity had dropped considerably but that it remained significant (Smith and Damphousse 1998).

The magnitude of the reduction in sentencing disparity in the post-September 11 era suggests that a number of other factors may be at work. First, there was a small but significant decrease in count severity in lead offenses in post-September 11 cases. This likely had a small but significant impact on the average sentence length. Second, one would expect shorter prison sentences for defendants who plead guilty as compared to those who are convicted at trial. As noted above, prosecutors and defendants agreed to plead guilty in two-thirds of all cases filed after September 11, an increase of 23 % from before September 11. Third, Jackson (2011) determined that prosecutors were more likely to accept guilty pleas on fewer counts within an indictment in the post-September 11 era, reducing the opportunity for consecutive sentences. Jackson found the percentage of unconvicted counts per indictment before September 11 was roughly 37 % while the percentage of unconvicted counts per indictment after September 11 was about 82 %. In essence, a chain of events is probably responsible for the decline in sentence length. As the FBI has moved to a "proactive" rather than "reactive" stance regarding terrorism after September 11, they have begun to intervene earlier in investigations. This frequently means evidentiary strength is not as great in these cases (Jackson 2011), which results in greater willingness among prosecutors to negotiate a plea agreement.

Related Entries

- ▶ Ecoterrorism
- ▶ Homegrown Terrorism in the United States
- ▶ Incarcerating Terrorists
- ▶ Plea Bargaining

- ▶ Preventing of Terror at Shopping Malls
- ▶ Prosecutorial Discretion
- ▶ Situational Approaches to Terrorism
- ▶ Social Disorganization and Terrorism
- ▶ Strategies of Policing Terrorism
- ▶ Terrorist Organizations

Recommended Reading and References

Ashcroft J (2002) Ashcroft fact sheet on new FBI investigative guidelines. Politechbot.com, September 21, 2002

Bradley-Engen MS, Engen RL, Shields C, Damphousse KR Smith BL (2011) The time penalty: examining the relationship between time to conviction and trial vs. plea disparities in sentencing. Justice Q, pp 1–29. doi:10.1080/07418825.2011.624112

Chesney R (2007) Federal prosecution of terrorism-related cases: conviction and sentencing data in light of the “data reliability” and “soft sentence” critiques. Available at http://ssrn.com/abstract_id=1005478

Elliff J (1979) The reform of FBI intelligence operations. Princeton University Press, Princeton

Federal Bureau of Investigation (1998) Terrorism in the United States. U.S. Government Printing Office, Washington, DC

Greenberg K (Ed) (2011) Terrorism Trial Report Card: September 11, 2001 - September 11, 2011. Retrieved from Center for Law and Security, New York School of Law: http://www.lawandsecurity.org/Portals/o/Documents/TTRC%20Ten%20Year.

Hearings before the subcommittee on administrative practice and procedure of the committee of the judiciary on FBI statutory charter (1978) 95th Cong, 2d sess, June–Sept. U.S. Government Printing Office, Washington, DC, pp 98–99

Ingraham B, Tokoro K (1969) Political crime in the United States and Japan: a comparative study. Issues Criminol 4:145–171

Jackson S (2011) Measuring intervention success in countering terrorism, thesis (defended May 2011 University of Arkansas). Unpublished manuscript

Kelley CM, Davis JK (1987) Kelley: the story of an FBI director. Andrews, McMeel, and Parker Publishing, Kansas City

Office of the Attorney General (1983) Attorney general’s guidelines on domestic security/terrorism investigations. Office of the Attorney General, Washington, DC

Perl R (2003) Terrorism, the future, and American foreign policy. Issue brief for congress, order code IB95112, Foreign Affairs, Trade and Research Division

Poveda J (1990) Lawlessness and reform: the FBI in transition. Brooks/Cole, Pacific Grove

Russell C, Miller B (1977) Profile of a terrorist. Terror Int J 1:17–27

Shields CA (2008) An analysis of prosecutorial and defense strategies in federal terrorism trials from 1980 to 2004 dissertation (defended December 2008), UMI Number: 3341235

Shields CA (2012) American terrorism trials: prosecutorial and defense strategies. LFB Scholarly Publishing LLC (print release September 2012)

Smith B (1994) Terrorism in America: pipe bombs and pipe dreams. State University of New York Press, Albany

Smith B, Damphousse K (1998) Terrorism, politics, and punishment: a test of structural-contextual theory and the “liberation hypothesis”. Criminology 36(1):67–92

Smith B, Morgan K (1994) Terrorists right and left: empirical issues in profiling American terrorists. Stud Confl Terror 17:39–57

Smith B, Shields CA, Damphousse K (2011) Patterns of intervention in federal terrorism cases Year 2 Progress Report, June 2011, for DHS project via the START Center of Excellence, University of Maryland: Creation and Analysis of an Integrated U.S. Security Database (IUSSD), BAA number 08-01

Smith L, Damphousse K (1996) Punishing political offenders: The effect of political motive on federal sentencing decisions. Criminology 34(3):86

Stern J (1999) The ultimate terrorists. Harvard University Press, Cambridge, MA, p 11

Transactional Record Access Clearinghouse (2003) Criminal terrorism enforcement since the 9/11/01 attacks. A TRAC Special Report, 8 Dec found at http://trac.syr.edu/tracreports/terrorism/report031208.html

Turk A (1982) Political criminality. Sage, Beverley Hills

United States Department of Defense, Office of Joint Chiefs of Staff Joint Publication 1-02: Department of defense dictionary of military and associated terms. United States Department of Defense, Washington, DC, 12 Apr 2001 – As amended through 5 June 2003, p 531

Uniting and strengthening America by providing appropriate tools required to intercept and obstruct terrorism act, (P.L. 107-56) enacted, Oct 2001

Use of Social Media in Policing

Sean P. Varano and Ryan Sarasin
School of Justice Studies, Roger Williams University, Bristol, RI, USA

Overview

Social media has in many ways altered the landscape of communication and information sharing in recent years. Nearly 750 million people across the globe are estimated to have accessed *Facebook* by August 2011, or nearly one-tenth of the world's population (Facebook 2012). Its impact has been felt across the social landscape, from basic interpersonal communications to shaping the short-term and long-term socio-political future of some countries. It has been argued, for example, that it played a central role in the toppling of tyrannical governments during the "Arab Spring" of 2011 (Howard and Hussain 2011). With the breadth and depth of this impact in mind, law enforcement agencies have begun to adopt social media strategies in an effort to expand the scope of their communication strategies. It is estimated that by September 2011, approximately 88 % of law enforcement agencies across the United States had adopted some form of social media strategy (International Association of Chiefs of Police 2011). The purpose of this entry is to provide an overview of the use of social media in law enforcement. After providing a definition and brief history of social media, the authors provide a conceptual framework for understanding the role of social media, like other mechanisms for communication and information sharing, within a policing environment. The entry then provides an overview of the benefits and potential liabilities associated with the integration of social media into organization communication strategies. Finally, the entry offers a set of policy recommendations that can assist police organizations considering implementing social media strategies.

Background

The City of Los Angeles was rocked by a series of fires that occurred between December 29, 2011 and January 2, 2012 when more than 50 separate instances of arson indiscriminately occurred. By December 31, it was clear that there was a serial arsonist on the loose wreaking havoc in the greater Los Angeles area. While motor vehicles were the primary targets, several fires quickly spread to nearby buildings, which only intensified the mounting problems. The damage was extensive in many situations as evidenced by one small cluster of fires in the West Hollywood area that caused more than $350,000 worth of damage to vehicles and property (CNN 2012).

Fire officials scrambled across the widely dispersed city in an effort to get a handle on the expanding threat. But with multiple fires erupting within short periods of time, the public safety capacity of Los Angeles was pushed to its limits. Public safety officials quickly turned to social media as an outlet to both collect quickly emerging information from the public about a rash of fires, but also to communicate back to the public about the evolving situation, particularly when a suspect was eventually identified. Almost immediately it was recognized that social media played a critical role in a quick resolution to this public safety crisis. It greatly enhanced the ability of public safety officials to disseminate information to the public and media alike quickly, but also because it provided a capacity to access segments of the community that might otherwise not frequent more conventional news outlets. In the end, a suspect was identified and apprehended by January 2, 2012. The police and traditional media have universally recognized the value of social media in helping to bring this situation to a timely end.

The 2011–2012 New Year's fires represent an important milestone for the Los Angeles Police Department; the fires represented one of the first times that social media has been used by the LAPD in an effort to facilitate bi-directional communication with the public about an emerging public safety threat. This example shows the promise that social media presents to public safety officials trying to do more with less. Yet, while the potential advantages to implementing a robust social media capacity are apparent to many, some police departments across the United States have been reluctant to accept this new form of communication. While the reasons for this reluctance might be many, it is clear that at least some of this apprehension is due to basic lack of familiarity and expertise.

Social media is a complex concept that generally refers to a broad range of non-traditional forms of electronic communication. It is not completely clear what exactly the term "social

Use of Social Media in Policing, Table 1 Various forms of social media

Forms of social media	Illustrative platforms
Blogs	Blogspot, Live Journal, BlogCatalog, Wordpress
Forum	Yahoo! Answers, Epinions
Media sharing	Flickr, YouTube, Justin.tv, Ustream, Scribd
Microblogging	Twitter, foursquare, Google buzz
Social networking	Facebook, Myspace, LinkedIn, Orkut, PatientsLikeMe
Social News	Digg, Reddit
Social Bookmarking	Del.icio.us, StumbleUpon, Diigo
Wikis	Wikipedia, Scholarpedia, ganfyd, AskDrWiki

Source: Tang and Liu (2010, p. 1).

media" encompasses. Social media ranges widely in form and purpose. According to Kaplan and Haenlein (2010), social media is "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content" (p. 61).

As of 2010, forms of social media included blogs and microblogs, discussion forums, media sharing websites, social networks, social news, social bookmarking, and Wikis (see Table 1). Although these forms of social media differ, a common feature that differentiates them from the conventional web and traditional media is that they are consumers of content. Compared to traditional media (TV, radio, movies, and newspapers), in which only a small number of people decide which information should be produced and how it is distributed, a user of social media can be both a consumer and a producer. With millions of users active on the various social media sites, everyone can be a media outlet. Thus, social media allows for the production of timely news and leads to volumes of user-generated content.

Kaplan and Haenlein (2010) take on a slightly different approach when defining social media by developing a more robust classification scheme. According to the authors, there are six different types of social media. The first type of social media is collaborative projects, which allow for the joint and simultaneous creation of content by end-users (e.g. Wikipedia). The second type of social media is blogs and micro-blogs, which are particular websites that normally display date-stamped entries in reverse chronological order (e.g. Twitter). Third, content communities refer to the sharing of media content between users (e.g. YouTube). The fourth type is social networking sites, which allow users to connect by creating profiles, enabling friends to have access to those profiles, and sending emails and instant messages between each other (e.g. Facebook). The fifth type is virtual game worlds, which refers to platforms that portray a three dimensional environment whereby users can create and appear in the form of personalized avatars and interact with each other (e.g. World of Warcraft). Finally, virtual social worlds are similar to virtual game worlds, except there are no rules restricting the range of possible interactions, such as the Second Life application (Kaplan and Haenlein 2010). These technologies allow users to simply create content on the Internet and share it with others. Thus, social media is the infrastructure that enables users to become publishers of content that is interesting to them and others.

Social media allows for the production of timely news and information. For example, during the London terrorist attack in 2005, some witnesses blogged about their experience in order to provide first-hand reports of the incident. Similarly, witnesses provided updates on Twitter during the bloody clash that resulted from the Iranian presidential election in 2009. In the event of an emergency or disaster, people have demonstrated to turn to social media platforms. This has prompted the city of Arcadia, California to introduce social media to their Emergency Operations Center (LeVeque 2011). In addition, social media enables collaborative writing to create high-quality bodies of work that was not previously possible. For example, over the last decade, Wikipedia has grown into one of the largest reference web sites, with more than 85,000 active contributors working on more than 14,000,000 articles and attracting more than 65 million visitors monthly. Furthermore, one of social media's unique characteristics is in its user interaction. The participation of users on social media websites is what makes them successful, advancing eight social media sites to the top 20 websites in the United States (Tang and Liu 2010).

Evolution of Social Media

Because the term social media is not exactly clear, examining the history also varies from different sources. However, several articles agree that an early social media service was launched in 1997. According to Boyd and Ellison (2007), the first recognizable social network site, Sixdegrees.com, was launched in 1997, which allowed users to create profiles, list their friends, and, starting in 1998, view the "friends" list. In 1999, Pyra Labs' Blogger was introduced, an early blog-publishing service that allows blogs with time-stamped entries (Paquet 2002). Another social media website created in 1999 included LiveJournal, a virtual community in which users can keep a blog. The next series of social networking sites began in 2001 when Ryze.com was created to allow people to leverage their business networks (Boyd and Ellison 2007).

According to Boyd and Ellison (2007), three key social networking sites that shaped the business, cultural, and research landscape included Friendster (2002), Myspace (2003), and Facebook (2004). Friendster was created in order to help friends-of-friends meet, unlike meeting strangers on dating websites. However, due to a substantial increase in users, the site encountered technical difficulties, leading to a reduction in the number of users. Social networking sites really hit the mainstream from 2003 onward. Most sites used a profile-centric form, similar to Frendster. Myspace was created in 2003 in order to compete with other social networking sites like Friendster and Xanga. Myspace differentiated itself from other social media sites because it allowed users to personalize their pages. A significant number of teenagers began joining Myspace in 2004, and eventually evolved into three distinct populations: musicians/artists, teenagers, and the post-college urban social crowd. Facebook was originally designed to support distinct college networks only. Although it started as a Harvard only site, it eventually expanded to include other colleges, high school students, professionals inside corporate networks, and, ultimately, everyone. The latest key social media site that has substantially increased in popularity over the last few years is Twitter. Twitter, which began in 2006, is a microblogging site, limiting posts to 140 characters or less. This restriction allows real-time posts, also known as "tweets," to be made using short message service (SMS) technology, which is the basis for text messaging on cell phones and other mobile devices. Although these social media sites have been created and evolved to seek broad audiences, several sites also target business people, such as LinkedIn, Visible Path, and Xing. A recent study argues that LinkedIn is a fantastic tool for law enforcement professionals (Stevens 2011). Table 1 illustrates the evolution of major social networking sites. Because the history of social media encompasses a huge phenomenon, this list is unavoidably incomplete.

Explosion of Social Media Use

In recent years, there has also been a considerable expansion in the levels of use of social media. According to Kaplan and Haenlein (2010), 75 % of Internet users used "social media" in the second quarter of 2008 by becoming members of social networks, reading blogs, or writing reviews on shopping websites, compared to 56 % in 2007. This growth is not only attributable to teenagers, but also populations well beyond their forties and fifties. According to Chou et al. (2009), participation in social networking sites more than quadrupled from 2005 to 2009. Boyd and Ellison (2007) claim that although there is not any exact data regarding how many people use social media websites, marketing research suggests that these sites are growing exponentially worldwide. As a result, many corporations now invest time and money into creating, purchasing, and promoting social networking sites. Boyd and Ellison (2007) also argue that the rise of social networking sites suggest a shift in the organization of online communities. Social networking sites are now mainly organized around people, not interests, and are structured around personal networks. Therefore, social media is an extremely active and evolving domain that now accounts for a significant proportion of Internet traffic.

A recent poll by the Pew Research Center indicates that approximately 66 % of online adults use social media, with the vast majority reporting the use social media to "stay connected" to family and friends (Smith 2011). Although the extent of connectivity via social media sites varies across countries, data indicate that approximately 30 % of adults in most industrialized countries use social media to some degree (Pew Research Center 2011). Facebook, one of the leading social media companies, has experienced an exponential increase in the number of registered users in recent years. The number of registered "active" users, for example, increased from 12 million in December 2006, to 100 million in August 2008, to 500 million in July 2010, and finally to over 750 million "active" users by August 2011 (Facebook 2012). This figure, which represents nearly 10 % of the *world's* population, is nearly impossible to grasp. The magnitude and speed of this growth defies comprehension and has placed its founder, Mark Zuckerburg, as one of the wealthiest people in the world. . .at age 27 (Forbes 2012). The use of social media has affected the social landscape in many ways, and has added terms such as "friending," "defriending," and "tweets" into the daily lexicon of many.

The Legacy of Technology in Policing: Setting the Stage for Social Media

It should come as no surprise that social media has developed into an issue that police departments across the world have sought to understand and harness. Policing is in many ways a technology driven profession that has historically been quick to adopt burgeoning forms of technology, particularly as it relates to communication and information sharing. The motorized vehicle, radio communication systems, and computer systems have been integrated into policing in highly visible and instrumental ways. The tendency toward the adoption of technology is functional and symbolic as well as practical and ceremonial. From the outside, it appears that police have enjoyed and exploited major technological advancements of the modern era (Brady 1997). As Varano et al. (2007, p. 263) argued, "technology [in policing] holds the symbolic potential for providing a degree of respect as it implies access to resources, commitment to innovation, organizational leadership, and a degree of sophistication." That many police organizations have adopted social media as a way of communicating to the public, then, should be of little surprise.

Social media represents one of the next biggest waves of technological advancement in law enforcement. Like other governmental sector agencies, many police departments around the world have been quick to adopt the use of social media to communicate with the public. A recent survey by the International Association of Chiefs of Police (2011) found that 88 % of 800 surveyed agencies reported some use of social media, with Facebook, Twitter, and YouTube among the most common (p. 4). These strategies, in many ways, lie at heart of more innovative policing strategies such as community-policing or problem-oriented policing, both of which hold police-citizen communication/interaction as a core philosophy. The most common reasons cited by police for the adoption of social media include community outreach, information dissemination, emergency/disaster notification, and investigations (International Association of Chiefs of Police 2011, p. 9). Not only do social media provide a greater opportunity for bi-directional communication with the public in general, it is even more so for the younger generation. Young people across the socioeconomic divide use social media to stay connected with family and friends, as a primary source of information about the world around them, and even as a primary source of emotional support (Ellison et al. 2007; Greenhow and Robelia 2009). There is a growing sense that while exploitation of social media might be optional today, it will not remain so for very long as these new forms of communication expand in their dominance of the social landscape and discourse.

Factors Underlying the Adoption of Social Media Practices

From a conceptual point of view, the trend toward adoption and expansion of social media in policing is part of a larger trend centered on greater communication with the public. Community policing, a policing model that stresses community engagement and proactive problem-solving, has pushed police departments and communities alike to reorient themselves in ways that view one another more as allies and less as adversaries. Community policing represents an important paradigm shift that moves police out of the role of the crime fighting professional who is solely responsible for establishing organizational priorities. Instead, the community is viewed as an active participant in organizational processes. The community has a role in defining organizational priorities, but just as important, the police are held accountable to the public they police. Community policy sees the community as the *customer* whose needs are worthy of consideration. This is part of a lager "customer service" trend in the public sector.

Another important dimension of community policing has been the notion of *police accountability* to the public. That is, community policing is based on the premise that the community not only has a direct role in helping to shape organizational priorities, but that the police have a responsibility for sharing information with the public. It is important to remember that information sharing is *not* something that comes natural in policing environments. Information is generally something to be tightly controlled and it is disseminated on a "need to know" basis. Police are traditionally not in the habit of sharing information about individual cases or overall trends with the public. This, however, is not unique to the relationship between the police and the public. Most police departments are reluctant to share information with other police departments and even other units (e.g., vice, narcotics, or gang units) within the same police department. It is also not uncommon for officers in the same unit within the same department to be reluctant to share information with each other. The need to tightly control the internal and external flow of information is viewed as so important that many departments employ a "Public Information Officer" to specifically handle requests for information. Thus, information sharing is something that is neither natural nor routine in a policing environment.

From a conceptual point of view, information sharing can be viewed not only as a service to the public, but also as image control whereby police agencies have adopted deliberative strategies intended to create positive public presences. Driven both by the desire to share information and to craft an image of a progressive department, many agencies have some sort of presence on the World Wide Web. Social media then is likely part of this larger trend toward information sharing and community outreach.

While there is little doubt that the movement toward community policing partially explains the expansion of social media, it is not the exclusive factor. The movement toward a "virtual presence" on the web, including the adoption of social media, has also been influenced by resource constraints that have plagued policing and other public sector agencies, particularly in the past two decades. In policing, calls for service (e.g., 911 calls) are one of the primary sources of workload. Large call volumes to emergency 911 centers and shrinking personnel resources due to layoffs have put pressure on police departments to develop innovative ways to reduce workload, but to do so in a way that compromises neither public safety nor community relations. Police departments across the United States have experimented with alterative reporting strategies that have moved non-emergency calls out of the 911 system. The goal of these efforts is to free up officer patrol time by limiting their involvement in non-emergency events that do not need a police response. Moreover, the intent is to also enhance customer satisfaction by creating a reporting strategy that is more convenient. At the end of the day, alternative reporting systems are premised on leveraging existing resources in a way that maximizes return and enhances customer service.

Communities experimented with a variety of non-emergency reporting strategies including 311 systems and phone-based police reports. What few evaluations are published indicate that, by and large, alternative systems can be effective alternatives to the traditional 911 system (Mazerolle et al. 2003). The explosion of the internet, however, created a new and relatively inexpensive crime reporting strategy – online crime reporting. By the 1990s police departments across the United States were experimenting with the use of online crime reporting technologies as a means for permitting the reporting of crimes in a way most convenient to the victims. Since many departments were already developing an online presence as part of the community outreach efforts identified above, this seemed to be a natural extension of these efforts. While the actual prevalence of online reporting is unclear, the Executive Director of the COPS Office, Bernard Melekian, has referred to it as the "wave of the future" (Cisneros 2009).

It has been the movement toward community engagement and the impact of financial constraints requiring police to do more with less, phenomenon that occurred almost simultaneously, that set the stage for the adoption of social media. The ever-expanding capacity of internet, the diffusion of computer technology across most segments of society, and rapid expansion of evolution of social media all occurred within a short 10–15 year period. The quickly evolving technical infrastructure provided the means for police, like other public and private sector agencies, to develop a virtual presence and cautiously explore the dynamics of information sharing.

The Use of Social Media in Policing

As social media continues to expand in its dominance of the social landscape, organizations in any profession that dismiss social media will find themselves lagging behind. Utilizing social media has several benefits and can be extremely invaluable in a contemporary policing environment. To keep up with the quickly evolving landscape, many police departments have experimented with different strategies. There remains, however, wide variation in the extent to which social media has been adopted within the policing community. Below is an overview of how social media has been integrated into policing.

Recruitment

Police recruitment in the twenty-first century can be especially challenging. Agencies are faced with staffing challenges associated with budget-related attrition, retiring wave associated with the baby-boom generation, and certain "generational trends" that move qualified candidates to other professions (Wilson and Castaneda 2011). In order to be competitive employers, law enforcement agencies need to be resourceful in their approach for recruiting applicants and seek new tools to accomplish this task. Social media offers law enforcement this new set of tools with an ability to attract, engage, and inform potential applicants.

There are several advantages of using social media platforms for recruitment volume. With the vast number of users who utilize social media sites on a daily basis, it presents a great opportunity to connect with both passive and active job seekers. Also, unlike commercials and paid investments, social media sites are inexpensive. For example, there is no cost to create an account, but personal time is necessary to maintain and update content on the sites. In addition, social media allows agencies to answer any recruitment questions applicants might have on a personal level and in a publicly accessible fashion (International Association of Chiefs of Police 2011).

Law enforcement agencies such as the Houston Police Department (Texas) and the Vancouver Police Department (British Columbia) currently use social media platforms including blogs, social networks, and multimedia-sharing sites to give potential applicants a unique and distinct view of the police profession. In 2008, a Houston Police Department senior officer launched a successful blog in order to provide insight and offer a resource to jobseekers, applicants, and new recruits. The blog consists of stories, information, videos, and pictures, with daily weekday posts. The recruitment blog substantially expanded, starting with approximately 15 hits per day to now over 1,500 site views daily. The Houston Police officer argues that it offers the ability to share insight with potential applicants that they cannot get anywhere else (International Association of Chiefs of Police n.d.-b). The Kentucky State Police has also turned to various social media platforms to attract new recruits. The agency has launched recruiting videos on its YouTube Channel, while already having presence on Facebook, Twitter, Flickr, and Blogger.com (Gazaway 2011).

Social media tools such as the methods used by the Houston Police Department and Kentucky State Police have offered law enforcement agencies the ability to connect with the general public and answer any questions about a future career in law enforcement. Blogs have given agencies a method to communicate more regularly and more personally than a conventional department website offers. For example, some topics discussed include a day in the life of a new recruit, advice or comments from the police chief, and miscellaneous blog posts from other department personnel. Law enforcement agencies have used Facebook to communicate with both potential and current applicants. Facebook posts allow applicants to have specific conversations that can be viewed by other applicants. In addition, departments have created a personal YouTube channel in order to post recruitment and other career-related videos.

Investigations

Social media is transforming the way in which law enforcement agencies conduct investigations and should be embraced by the department as a whole. Law enforcement agencies use social media for investigative purposes when seeking evidence or information about missing or wanted persons, gang participation, as well as Internet crimes such as cyberstalking or cyberbullying. For example, in Franklin, Indiana, police work with prosecutors and use social network sites as evidence in cases ranging from underage drinking to child custody cases. Information found online helps supplement other evidence already gathered by the police (National Law Enforcement Policy Center 2010). In one case, police and prosecutors convicted an individual of operating a vehicle under the influence of a controlled substance. The individual had smoked marijuana on the day he crashed his vehicle and killed another individual inside. On his Facebook page, the individual referenced his marijuana use and also posted photos of a bong (Michalos 2010). This example illustrates what many law enforcement agencies are doing throughout the country in which police and prosecutors are using information found on social media sites to gather and enhance the evidence against an individual (National Law Enforcement Policy Center 2010).

The Utica, New York Police Department (UPD) has been using Facebook since 2010 and began using Twitter and YouTube in 2011. During a period of 4 months, UPD made 11 arrests directly from information posted on the UPD social media sites, including a bank robbery and numerous grand larcenies. In some cases, people have turned themselves in, possibly out of fear or embarrassment. In other cases, people who recognize the pictures and videos contact the police department with information. The Department believes that social media empowers the community to get involved in the crime fighting process (International Association of Chiefs of Police, n.d.-a). This leads to enhanced relationships with community members.

Community Outreach and Citizen Engagement

Law enforcement agencies have begun to use social media tools to enhance community policing initiatives. Police departments throughout the country, for example, are posting crime prevention tips through various social media platforms, providing opportunities to report crime online, posting crime maps and other data, and disseminating alert information to the community (National Law Enforcement Policy Center 2010). Agencies throughout the country are also using social media platforms to reach out to their community members in order to foster better relationships and connections. For example, Baltimore Police Department uses social media sites such as Facebook, Twitter, and Nixle to enhance their relationship with the community and increase knowledge and safety throughout the city. The Baltimore Police Department uses social media sites as "an extension of the local news media because the media can't cover everything that happens and involves the department" (National Law Enforcement Policy Center 2010, p. 2). The Richmond (Virginia) Police Department (RPD) also uses social media sites to strengthen community relations. RPD posts information on their sites such as its "officer of the month" (Wagley 2011). The Colorado Springs Police Department (CSPD) uses social media sites and includes several features on its pages including: From the Chief, Quote of the Week, and Officer of the Month. The department argues that there is real value in sharing information with the public and illustrating how the different processes and procedures work within the police department. CSPD has received positive feedback from the community (International Association of Chiefs of Police n.d.).

In terms of citizen engagement, the Boca Raton (Florida) Police Department has clearly structured and defined goals regarding social media outreach. The department's social media and communications strategies prioritize effectively engaging the community. Thus, instead of simply disseminating information to the public, the department's social networking sites allow for two-way communication, which allow for heightened levels of transparency. The two-way communication also offers a mechanism for the police department to gather information from community members (Alexander 2011).

Background Investigations

Social media has also been used to screen potential police officers during the application phase. Much to the consternation of some, several law enforcement agencies request that applicants sign waivers that allow investigators to have access to their various social media sites in the search of disqualifying information. Some agencies even require candidates to provide private passwords, Internet pseudonyms, text messages, and email logs as part of the expanding cybervetting process for law enforcement jobs (Rose et al. 2010). According to a report on law enforcement social media use, more than a third of 728 law enforcement agencies check applicants' social media sites during background checks. A potential recruit in Malden, Massachusetts was disqualified after background investigators found evidence of a past threat of suicide. Another potential candidate in Middletown, New Jersey was disqualified for posting racy photographs of himself with scantily clad women (Johnson 2010). Although it is important for law enforcement agencies to use cybervetting procedures during the background investigation process for potential applicants, they must be diligent to keep in mind privacy concerns to ensure fair and just hiring practices (National Law Enforcement Policy Center 2010)

The use of social media for background investigations is so common that many potential recruits are aggressively warned to delete any social media accounts they have maintained sooner than later. The prevailing theory is that departments who fail to properly screen/monitor potential and current employees are open to negative media attention. Nearly one-third of 800 agencies responding to the IACP survey indicated their agency had experienced negative media attention from social media use by on-duty or off-duty employees (International Association of Chiefs of Police 2011, p. 11).

Global Social Media Impact

The use of social media is not limited to the United States. According to the International Telecommunication Union (2012), there are now over one billion users of social media worldwide. Police agencies in various countries have began to embrace the benefits of using social media. The Metropolitan Police of London, for example, have turned to Flickr in an effort to identify individuals involved in serious crimes. After the 2011 riots on London streets, local police posted a set of photos on Flickr showing people they believed to be involved in the riot. Police were asking the public to identify any individuals from the photographs captured from CCTV surveillance cameras in areas where stores were looted. On the Flickr page, police posted that their intent was to "bring to justice those who have committed violent and criminal acts" (Wortham 2011).

German police will soon use a social media platform to assist in criminal investigations. Police in the German state of Lower Saxony are preparing to use their networks of Facebook "friends" to find missing persons and find suspected criminals. The decision to use social media comes from their successful implementation of a pilot program in the capital of Lower Saxony. This pilot program helped solve six criminal investigations and two missing persons cases. Photographs of suspects were posted on Facebook asking for public assistance. Two cases were solved just hours after the information was uploaded to the site (Baghdjian 2012).

Police agencies across the globe are also using social media platforms for community outreach and citizen engagement. The Toronto Police Service (Canada) is highly regarded as one of the most forward-thinking law enforcement agency users of social media platforms, especially for community engagement. Service members can continually be seen tweeting and posting to Facebook and YouTube (Stevens 2011). Several police agencies in Australia have also begun to utilize the community outreach strategy. For example, the South Australian Police (SAPOL) has embraced multiple social media platforms in hopes of better connecting with communities at a local level. SAPOL launched Facebook and Twitter accounts as well as a YouTube Channel, in an effort to increase community support. SAPOL also shares information on such things as travel hazards and CCTV images of wanted offenders (Fedorowytsch 2012). In addition, the New South Wales Police Force's eyewatch program brings community members and police together to communicate and solve problems using Facebook. The Facebook page includes crime prevention messages, alerts, updates, and other important community information. Since this social media platform has been implemented, there has been a 20 % increase in the information flow to the crime stoppers tip line (Maxwell 2012).

Police Facebook sites are also popular in New Zealand. One local police department's Facebook page had more than 5,000 views within 3 months of implementation and is responsible for solving several burglaries. The police agency uses the Facebook page for leads, while the public uses it for information. A community constable in the police agency argues that it's about bringing policing into the twenty-first century, using new tools available to get messages across to the community (Hueber 2012).

Opportunities and Challenges

There is little doubt that social media presents tremendous opportunities for the policing community. In a time of an unprecedented tightening of resources and ever-increasing demands on the policing community, organizations continue to struggle with the realities of doing more with less. In fact, the concept of "doing more with less" seems to be the prevailing mantra of most organizations today, and like other complex organizations, police are not immune from these demands. In a time of tough economic times that project to remain constant for the near future, police departments around the globe are strained with making the tough choices between what they *can actually do* in lieu of budgetary constraints. The opportunities are real and tangible.

In many ways, social media and related technologies offer to help law enforcement agencies grasp the realities those more contemporary philosophies of policing have to offer. Policing paradigms like community policing and problem-oriented policing have promised a new era in police community relations. These paradigms have stressed the centrality of the community to the policing function, the need to share information with the public, the value of connecting with the community on *their* terms, and a need to do these things within the natural constraints of available resources. While laudable, these goals are often hampered by a larger policing culture that does not necessarily support or recognize the value of community involvement in what can otherwise be considered “police business.”

Social media, however, offers the opportunity to facilitate these goals in important and unique ways, especially as it relates to information sharing. It affords the ability to fundamentally change the ground rules of police-community relationships in ways that increase transparency. “To truly develop a police culture that can exchange data, information, and intelligence” (National Law Enforcement Policy Center 2010, p. 5). Social media provides a relatively inexpensive fix for a broad array of information exchanges. It provides an opportunity for rich information exchange that is done on the terms of the public, but which is also not too taxing on either the police or the community. It offers the opportunity for meaningful information exchange and “dialogue.” It expands the “reach” of police departments into segments of the community they may otherwise have difficulty accessing. It is also remarkably efficient in doing so in a way that minimizes actual resource investment.

Yet social media offers more than just information sharing. It offers bi-directional communication that may actually increase *public safety*. The Los Angeles fires during the 2011–2012 New Year season are a noteworthy case in point. As described earlier in this entry, this situation represented a rapidly expanding threat that pressed the limits of the public safety apparatus. The sheer magnitude of the unfolding events made it nearly impossible for serious investigations to be initiated during this early period. Police and fire officials quickly began to take notice that social media sites such as *Facebook* and *Twitter* were inundated with “chatter” about fires. Users were quickly “reporting” new fires, sometimes before 911 was even notified. Videos and pictures of freshly set blazes were appearing on these sites on a near instantaneous basis. *Facebook* and *Twitter* sites such as “Arson Watch L.A.” were used quickly and became venues for individuals interested in following the situation with the arsons to track the latest news, and have the most up-to-date information on what was occurring. By December 31st, public safety officials had begun to fully exploit this technology to both gather information about events, but also to communicate the most current information such as descriptions of suspects, particularly to segments of the community who might otherwise not frequent more conventional media outlets. Social media would eventually emerge as one of the primary ways Los Angeles police were able to communicate directly with the residents of affected neighborhoods. During similar situations in the past, police would have to organize last-minute community meetings in an effort to disseminate information to residents. As one police official noted, “By the time we’d be done with that, we’d be behind the curve, at best, 24 h.” Now, he argued, “we can turn out information in minutes” (Stevens and Winton 2012). By January 2nd, police had identified and apprehended a suspect after aggressively disseminating footage from a surveillance video depicting a person of interest leaving the scene of a crime. Social media played a vital role in helping police and fire officials manage a dynamic public safety threat.

The advantages of a favorable orientation toward social media are obvious, but less obvious are the potential complications presented by social media. Police organizations are inherently risk-adverse institutions intent on communicating to both internal and external constituencies a sense of professionalism. It is an occupation rife with potential complications associated with violations of standard operating procedures. The potential costs associated with scandal are a perennial management concern, and the first hint of scandal often brings a search for internal “suspects” on whom the blame can be laid. The action and inaction of police organizations can often be viewed through this “risk-averse” lens.

It should come as no surprise that many police departments around the country have been trepid about integrating social media into their broader community outreach initiatives. Numerous police departments, for example, have been faced with the very real challenges associated with the discovery of unsavory behavior on the part of sworn employees prominently displayed on social media sites. Examples of problematic behavior have included provocative pictures, apparent participation in illegal behavior, unsavory text-based comments, or other behavior that is considered unbecoming of an officer. One BBC News (2011) news story reported that 150 police officers in the UK were recently disciplined for inappropriate use of *Facebook*, at least 1 of which was fired. Infractions included harassing former paramours and the posting of racist comments. In one example, officers were disciplined after bragging of beating up members of the public during recent protests.

Concerns about the misuse of social media are evidenced by the extensive attention given to the "legal aspects" of social media use in IACP's document titled *Social Media, Concepts and Issues Paper* (National Law Enforcement Policy Center 2010). This document, for example, devotes nearly two-thirds of the entire document to potential legal issues. Concerns raised include important issues such as first amendment rights of employees, the inappropriate dissemination of sensitive information via social media, and the impacts of personal use of social media on the employability of current and future employees. Agencies, IACP warn, have a *duty* to educate and inform their employees, both civilian and sworn, about departmental "proper use" policies.

Recommendations for Moving Forward

The very notion that nearly 10 % of the world's population were active users of Facebook in just a few short years since the product was first developed signifies an unimaginable trajectory of growth. Social media is here to stay and its dominance as a mode of communication is growing by the day. It seems reasonable to assume that new and emerging forms of social media will be the dominant form of social interaction in the near future. With that in mind, police organizations are encouraged to exploit the advantages that social media has to offer, but to do so in a way that mitigates the potential negative tradeoffs. The following are a list of recommendations leaders should consider when planning for the adoption of social media:

1. *Clearly define your purposes*: Agency leaders are encouraged to fully explore their rationale for the adoption of social media practices. Social media is not an all or nothing strategy. It is a set of technologies that evolve over time to fit the changing needs of the department. Different forms of social media have particular advantages/disadvantages and primary forms of social media should be driven by a clear understanding of its purpose. A chaotic approach that is neither focused nor strategic has the potential to not only be ineffective but actually be counterproductive if it communicates a sense of unprofessionalism. Agencies are encouraged to start small but smart.
2. *Commit*: Regardless of the purposes social media plays in a police organization, commit to doing it well. Like static websites that rarely are updated with current content, the success of any social media strategy depends heavily on the extent to which agencies commit to making it part of operational strategies. The most successful social media strategies are those which help create a cultural of interaction and information sharing. Purposeful and consistent utilization of social media technology will help institutionalize its larger role in public safety.
3. *Appreciate the expectations social media creates on the part of the public*: Police departments often justify their forays into social media based on the value of transparency and open communication. Such statements along with the accompanying technology may also create an *expectation* of transparency and open communication on the part of the public. Before an agency fully commits to the use of social media, agency leaders need to fully appreciate how such an effort may also create expectations for more information sharing. Organizational representatives need to create reasonable expectations about the circumstances of information sharing.
4. *Devotion of necessary resources*: Agencies are encouraged to devote a reasonable set of resources to social media strategies in terms of

both money and personnel. While social media is attractive to many agencies because it is seemingly inexpensive, this does not suggest there are no costs. Like any attempts at communication, social media strategies need to be viewed as an important component of community outreach and interaction.

Recommended Reading and References

Alexander D (2011) Using technology to take community policing to the next level. Police Chief 78:64–65

Baghdjian A (2012) German police use Facebook pictures to nab crooks. http://www.reuters.com/article/2012/02/07/us-germany-facebook-idUSTRE8161LG20120207. Accessed 12 Sept 2012

Boyd DM, Ellison NB (2007) Social network sites: definition, history, and scholarship. J Comput Mediat Commun 13(1):210–230

Brady T (1997) The evolution of police technology (No NCJ 163601). U.S. Department of Justice, National Institute of Justice, and Office of Community Oriented Policing Services, Washington, DC

BBC News (2011) 150 officers warned over Facebook posts. http://www.bbc.co.uk/news/uk-16363158. Accessed 18 Jan 2012

Chou W-YS, Hunt YM, Beckjord EB, Moser RP, Hesse BW (2009) Social media use in the United States: implications for health communication. J Med Internet Res 11(4)

Cisneros T (2009, April 1) Santa Ana launches online crime-reporting tool. http://www.ocregister.com/news/reports-168503-police-online.html

CNN (2012) Official: Los Angeles arson suspect under investigation in Germany. http://articles.cnn.com/2012-01-04/us/us_california-arson_1_los-angeles-arson-fires-hollywood-apartment?_s=PM:US

Ellison NB, Steinfield C, Lampe C (2007) The benefits of facebook "friends:" social capital and college students' use of online social network sites. J Comput-Mediat Commun 12(4):1143–1168

Facebook (2012) Timeline statistics. http://www.facebook.com/press/info.php?statistics. Accessed 12 Jan 2012

Fedorowytsch T (2012) SA policing making Facebook connection. http://www.abc.net.au/news/2012-06-28/police-facebook-pages-connect/4098074?section=sa. Accessed 18 Sept 2012

Forbes Magazine (2012) Forbes 400 richest Americans. http://www.forbes.com/profile/mark-zuckerberg/. Accessed 15 July 2012

Gazaway C (2011) KSP turning to social media for recruits. http://www.wave3.com/story/15176887/ksp-turning-to-social-media-for-recruits. Accessed 29 Jan 2012

Greenhow C, Robelia B (2009) Old communication, new literacies: social network sites as social learning resources. J Comput-Mediat Commun 14(4):1130–1161

Howard PN, Hussain MM (2011) The role of digital media. J Democr 22(3):35–48

Hueber A (2012) Fighting crime with Facebook. http://www.theaucklander.co.nz/news/fighting-crime-facebook/1289252/. Accessed 1 Sept 2012

International Association of Chiefs of Police (2011) 2011 IACP Social media survey. International Association of Chiefs of Police, Alexandria

International Association of Chiefs of Police (n.d.-a) Case Study: Colorado Springs, Colorado, Police Department – community education, community engagement. http://www.iacpsocialmedia.org/Resources/CaseStudy.aspx?termid=9&cmsid=4734. Accessed 1 Sept 2012

International Association of Chiefs of Police (n.d.-b) Houston, Texas, Police Department – Blogging for recruitment and beyond. http://www.iacpsocialmedia.org/Resources/CaseStudy.aspx?termid=9&cmsid=4734. Accessed 12 Jan 2012

Johnson K (2010) Police recruits screen for digit dirt on Facebook, etc. http://www.usatoday.com/tech/news/2010-11-12-1Afacebookcops12_ST_N.htm. Accessed 3 Jan 2012

Kaplan AM, Haenlein M (2010) Users of the world, unite! The challenges and opportunities of social media. Bus Horiz 53(1):59–68

LeVeque T (2011) Introducing social media to an emergency operations center. http://lawscommunications.com/category/tom-le-veque. Accessed 21 Nov 2011

Manning PK (1996) Information technology in the police context: the "sailor" phone. Inform Syst Res 7(1):52–62

Maxwell JJ (2012) A new platform for community engagement and neighbourhood watch in the 21st century. http://www.abc.net.au/news/2012-06-28/police-facebook-pages-connect/4098074?section=sa. Accessed 1 Sept 2012

Mazerolle L, Rogan D, Frank J, Famega C, Eck JE (2003) Managing citizen calls to the police: an assessment of non-emergency call systems. United States Department of Justice, Washington, DC

Michalos S (2010) Social media to solve crimes. Journal Gazette. http://www.journalgazette.net/article/20100620/NEWS07/306209898/0/FRONTPAGE

National Law Enforcement Policy Center (2010) Social media: concepts and issues paper. International Association of Chiefs of Police, Alexandria

Paquet S (2002) Personal knowledge publishing and its uses in research. Seb's Open Res Oct:1–15

Pew Research Center (2011) Global digital communication: texting, social networking popular worldwide. Pew Research Center, Washington, DC

Rose A, Timm H, Pogson C, Gonzalez J, Appel E, Kolb N (2010) Developing a cybervetting strategy for law enforcement. International Association of Chiefs of Police, Alexandria

Smith A (2011) Why Americans use social media. Pew Research Center, Washington, DC

Stevens L (2011) Social media quick tip: 3 reasons why LEOs should be on linkedin.http://www.lawofficer.com/article/technology-and-communications/social-media-quick-tip-3-reaso

Stevens L (2011) The Toronto Police Service launches social media program. http://connectedcops.net/2011/07/25/the-toronto-police-service-launches-social-media-program/. Accessed 12 Sept 2012

Stevens M, Winton R (2012) Social media sites are crucial in arson probe. Los Angeles Times. http://articles.latimes.com/2012/jan/02/local/la-me-social-media-arson-20120103

Varano SP, Cancino JM, Glass J, Enriquez R (2007) Police information systems. In: Schafer JA (ed) Policing 2020. Federal Bureau of Investigations, Washington, DC

Wagley J (2011) Police embrace social media. Secur Manage 55(11):36–37

Wilson JM, Castaneda LW (2011) Most wanted: law enforcement agencies pursue elusive, qualified recruits. RAND Rev 35(1):25–29

Wortham J (2011) London police use Flickr to identify looters. http://bits.blogs.nytimes.com/2011/08/09/london-police-use-flickr-to-identify-looters/. Accessed 12 Sept 2012

Victim Input at Sentencing

Edna Erez[1] and Julian V. Roberts[2]
[1]Department of Criminology, Law, and Justice, University of Illinois at Chicago, Chicago, IL, USA
[2]Faculty of Law, Centre for Criminology, University of Oxford, Oxford, UK

Overview

This encyclopedia entry surveys victim rights at sentencing in adversarial justice systems and explores the continuing debate surrounding victims' claim to have a voice in the sentencing of "their" offender. Although victims claim input at various stages of the criminal process – from bail to parole – the sentencing process is of most significance to victims. At the same time, victim input at sentencing generates the most opposition among scholars and practitioners, making the topic of continuing significance to criminal justice professionals as well as victims and their advocates. The essay reviews the arguments for and against victim input at sentencing and the empirical research which has explored the effects of victim-related reforms in the area of sentencing.

Introduction

Within the adversarial model of justice, a criminal trial is construed as a conflict between two theoretically equal adversaries – the state and the defendant – played out before an impartial adjudicator – the judge. According to this model, the victim is a passive participant who may be called to testify and respond to questions in direct and cross-examination, if necessary. In jurisdictions which employ an adversarial legal system, the crime victim serves as the principal witness for the prosecution and, having served this function, has no further role to play in criminal proceedings.

Over the past three decades, the role of the victim has been transformed from passive witness to active participant, with rights for input into proceedings. While still lacking full "standing" as a party to the proceedings, victims in adversarial systems such as those in USA, Canada, England, or Australia are consulted, informed, and provided input into proceedings more than ever before. To some, this evolution represents a threat to core values of the adversarial model. Victims' rights advocates, on the other hand, view the new privileges and powers of the victim as evidence of progress toward the goal of full participation in the criminal process.

Currently, most countries with adversarial justice systems provide victims with the right to receive information about the status of the case in which they are involved (in the USA, referred to as victim notification laws), the right to apply for and receive financial and psychological assistance, and input rights throughout the criminal process (such as in bail, sentencing, and parole hearings). Although most of these rights and

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

benefits have been generally accepted, the right to provide input into sentencing decisions has proved controversial and continues to be the subject of heated debate. Controversies revolve around the extent and forms of victim input and its impact on the principles and outcomes of criminal justice proceedings. Underlying this debate is the continuous struggle for the "ownership of the conflict": many prosecutors and other court workgroup members still consider victims as outsiders, or at best symbolic participants, while victims and their advocates, particularly in serious crimes, view themselves as interested parties and thus demand input into court proceedings (Englebrecht 2011). Some victims' advocates have even proposed providing crime victims with "Parallel Justice" – a system that mirrors in its scope and importance the rights accorded to offenders (see Herman 2011). These debates have also emerged at the international level, where the international community pursues justice in special courts established to address mass victimization involved in crimes against humanity, genocide, and war crimes (Zappala 2010).

In the USA, victims' input rights have also assumed a constitutional dimension. The *Victims' Right Amendment* is a proposed constitutional amendment that would establish various participatory rights for crime victims nationwide, with remedies in cases of noncompliance. These rights include the right to speak during the course of legal proceedings, including pretrial release, plea bargains, sentencing, and parole. To date, despite campaigns for its passage and joint resolutions presented in congress, the proposed Victims' Rights Amendment has failed to gain the necessary number of states in support. The principal argument against its passage has been that victim participation and input can be accomplished by enforcing existing laws found in state constitutional amendments or statutes (Beloof et al. 2010).

The Emergence of Victim Input Rights

The role of victims in a criminal prosecution has changed drastically over the centuries in common law countries – from a system in which victims were expected to deal with their offenders directly, through a system in which the monarch assumed the duty of imposing punishment, to the present system in which the state prosecutes a defendant on behalf of the surrogate victim who is relegated to a role of (at best) lead witness. For most victims, even their role as witness never materializes. Because of the high case attrition in the criminal process, most cases do not result in arrest, much less prosecution. Even when charges are laid, many are subsequently stayed or withdrawn by the prosecutor acting in the public interest or because there is no reasonable prospect of a conviction. Further, in the overwhelming majority of cases, the offender enters a guilty plea, often following negotiations with the prosecutor. The victim's principal role in most cases is as a backup – a threat waiting in the wings if plea bargaining breaks down. If the defendant pleads guilty, there is no need for a trial, and the victim has no opportunity to testify. Unlike continental legal systems, which provide victims with party rights, the adversarial system accords victims no formal standing in the prosecution of their offenders and no input rights in the disposition of "their" offender.

The demand by crime victims to have a voice in the criminal justice system has been recognized by many national committees established to study victims in the criminal justice process. The President's Task Force on Victims of Crime (1982) was a particularly important undertaking which explored victims' experiences and revealed the extent of their secondary victimization as a result of their treatment by the criminal justice system. The international community has also recognized the need to integrate victims into the criminal justice process in the 1985 United Nations Seventh Congress on the Prevention of Crime and Treatment of the Offender, which adopted a declaration that required that victims be allowed to present their views and concerns at appropriate stages of the criminal justice process (UN Declaration of Basic Principles of Justice for Victims of Crime and Abuse of Power).

Following these demands and pressures to integrate victims in proceedings, many countries with adversarial legal systems initiated legal reforms that aimed at providing victims with input rights. In the United States, following the recommendation of the President's Task Force on Victims of Crime (1982) that the Sixth Amendment be amended to guarantee victims a right to be present and to be heard at all critical stages of a judicial proceeding, most states adopted laws that address victims' right to be present in proceedings by excluding victims from sequestration requirements or otherwise accommodating their wish to participate. Many states have enacted victims' bills of rights which vary in their scope from mandating that criminal justice officials show respect toward victims, to establishing a victim's right to be present and heard, to allowing victims to sit at the prosecutor's table during trial. In several states, victims' rights legislation is by specific statute. A substantial number of states have adopted constitutional amendments to give victims' rights greater permanence and visibility.

Victim Impact Statements in Major Adversarial Legal systems

In the USA, the states and the federal government allow for victim participation in sentencing hearings. The mechanism for providing victim input in the USA and many other countries is the *victim impact statement* (VIS) or *victim personal statement* as it is known in England and Wales. The VIS is commonly administered by either victim assistance team within the prosecutor office or probation departments within the court. Using the VIS victims can provide details on the physical, psychological, social, and economic harm they and their significant others have suffered as a result of the crime. In some states victims may make specific sentence recommendations in their victim impact statements. Elsewhere, in other common law jurisdictions such as England and Wales and Canada and New Zealand, victims are instructed not to make sentence recommendations. In these countries the purpose of the statement is to allow the victim to describe the harm created by the crime and not to influence the court in any particular direction with respect to the sentence that should be imposed.

The victim impact statement can be presented in writing and by video and some states allow for victim allocution, according to which victims can also present their VIS orally during sentencing. States also provide for victim input in plea bargaining and at parole hearings. For example, crime victims in Arizona have the right to be consulted with respect to any potential plea agreement. Furthermore, prosecutors are required to demonstrate that they have complied with victims' rights legislation. While some victims are hesitant or unwilling to partake in the arrest and prosecution of their violators (e.g., battered women who fear retaliation, are economically or emotionally dependent on their abusers, or are concerned with issues of privacy (e.g., Dichter et al. 2011)), most crime victims, particularly of serious offenses, welcome the opportunity to provide input into sentencing. The suffering endured by parties other than the direct victim, whether it consists of physical, psychological, or financial lasting impact, is known as "ancillary harm" (Roberts and Manikis 2010), and there is growing evidence that courts recognize such harm in sentencing (Cassell and Erez 2011). Most American jurisdiction do not provide for cross-examination of victims about their VIS.

It should be noted, however, that when input rights are provided, most US jurisdictions construe them as courtesies to be extended or withheld at the discretion of the police or prosecutor, without guaranteed remedies for nonenforcement (Tobolowsky et al. 2010), although the trend is to recognize the importance of providing legally enforceable rights with remedies, particularly at the federal level (*Crime Victims' Rights Act, 2004*), and a few states (e.g., Arizona) provide victims with remedies when their rights for input are breached (Beloof et al. 2010).

Other countries with adversarial legal systems have also adopted victims' right for input into sentencing, although not to the degree found in the United States. Australian states provide victims with input rights that are comparable to the rights accorded American victims. The purpose of the VIS in Australia is assisting the court in determining an appropriate sentence. VIS can be made by the primary victim in all of the Australian jurisdictions, and there are also provisions allowing for others to make VISs on behalf of themselves or, in certain circumstances, on behalf of the primary victim. Some jurisdictions, such as South Australia, allow a VIS to include recommendations as to the quantum of sentence, while others, such as Western Australia, expressly disallow the inclusion of sentencing recommendations. In all Australian jurisdictions, it is possible for a victim to orally present their VIS to the court. In Canada, victims have a statutory right to submit a victim impact statement at sentencing and to deliver the statement orally if they so desire. In addition, there is an obligation on courts to inquire of the prosecutor whether the victim has been apprised of her right to submit a statement. If no VIS is present, the prosecutor or the victim may ask for a remand of the case in order for the victim to be able to submit a statement. These provisions go further toward meeting the needs of crime victims than those found in many other countries.

A number of countries have adopted "Victims' Charters," which create expectations that the system will conform to certain standards without actually conferring legal rights upon crime victims. For example, in 2001, England and Wales implemented a victim impact scheme, known as the victim personal statement (VPS). This is a statement expressing the physical, emotional, and financial consequences of the incident. The policy was supposed to be a unique policy, crafted with the US practices as the backdrop and was intended to differ from them in several ways: obtaining the victim's statement early in the process (during the police report, with a right to update it when necessary), presenting this information as a written statement (as opposed to an oral statement), and notably, specifying that victims were not to comment on what sentencing options they prefer for the accused. Finally, the importance of victim participatory rights is apparent from Article 68 of the Rome Statute of the International Criminal Court (ICC) which recognizes the security interests and participatory rights of victims and witnesses and allows victims to provide input into sentencing decisions.

Debates Surrounding Victim Impact Statements at Sentencing

Arguments for Victim Input

Advocates of victims' rights to participate in the criminal justice process have advanced various moral, penological, and practical arguments in support of victim input into sentencing. The right for input recognizes victims' wishes to be treated as a party to the proceeding and with dignity. It reminds judges, juries, and prosecutors that behind the "state," there is an individual with an interest in how the case is resolved. Some argue that sentencing outcomes will be more proportionate if victims provide accurate information about the impact of the crime (Cassell 2009). Without hearing from the crime victim, a judge bases his or her sentence on the description of the crime provided by the prosecution – an indirect account of the crime. By deposing a victim impact statement, the victim provides the court with a more comprehensive picture of the consequences of the crime. Victim input may also promote psychological healing by helping victims who express their harm to recover from the emotional trauma associated with their victimization and court experiences. Finally, there may also be benefits in terms of the offender. Hearing from the crime victim may sensitize the offender to the consequences of his actions in a way that is not possible if the victim does not provide input. This sensitization in turn may promote the offender's rehabilitation and strengthen his intention to refrain from further offending.

Arguments Against Victim Input

The objections to victim input into sentencing range from assumptions that vengeful justice will result to predictions that the system will grind to a halt as a result of the additional time needed to process cases if victims provide input (Erez 1994). Critics argue that a victim's input into sentencing is "irrelevant to any legitimate sentencing factor, lacks probative value in a system of public prosecution, and is likely to be highly prejudicial" (Hellerstein 1989, p. 429). Permission to deliver a VIS in person – exercising victim allocution right – has been regarded as particularly objectionable, as an oral delivery in a very serious crime may be emotionally moving for the judge or jury; this may increase sentence severity or sentencing disparity (Bandes 2009). Critics argued that similar cases might end up being disposed of differently depending on whether a VIS is available and the persuasiveness of the victim. Legal professionals and scholars also argued that victim input violates the fundamental principles of the adversarial legal system, which, as previously noted, do not recognize the victim as a party to proceedings. Including victims would transform the trial between the state and the defendant into a tripartite court proceeding (state – victim – offender). Such practices, it was argued, belong only in continental legal systems with adhesive prosecution or *partie civile* procedures, or to restorative justice schemes, but not in adversarial legal systems (see Ashworth 2002).

Research on Victim Input into Sentencing

Research provides answers to some of the issues raised in the debate about victim input into sentencing (see Roberts (2009) for a review of research findings). Findings from this research sustain the following conclusions: (a) levels of victim input vary across jurisdictions and are influenced by personal reasons as well as by jurisdictions' communal and organizational contexts; (b) victim input does not generally delay the criminal justice system by increasing the time taken to arrive at an adjudication; (c) victim input does not necessarily result in harsher punishment of offenders because courts are able to use victim impact statements appropriately; (d) victim input has the potential to increase victim satisfaction with the judicial system, although victim exposure through participation to the working of the criminal justice may also lead to disappointment; and (e) the implementation of victim input laws is still problematic, and many victims do not exercise their rights or benefit from these reforms.

Levels of VIS Submission

Research has revealed various impediments to the implementation of VIS programs (the term "VIS" is used as generic to describe VIS or VPS or other forms of victim input into sentencing). Many victim-related reforms such as the VIS in practice never reach victims. Victims are sometimes unaware of their rights to provide input, elect not to exercise them, or are confused about the nature and purpose of VIS. When VIS laws were first introduced, victims either did not know what VIS is or claimed that they did not fill out such statements when, in fact, they had done so. Being questioned by a seemingly endless array of people may have caused victims confusion about the purpose of particular interviews. More recently, as jurisdictions have improved their notification schemes and increased outreach and attention to victims' welfare, victims have become more of aware of their rights (Erez et al. 2011b).

One observation applicable to victim impact schemes in all common law jurisdictions is that only a minority of all victims submit a VIS. In England and Wales, a national survey of crime victims found that less than half of all victims interviewed recalled having been informed of the opportunity to submit a statement. Of those who had been informed, only approximately half actually submitted a statement (Roberts and Manikis 2011). Similarly, judges in surveys conducted in Canada reported seeing a VIS on average in 11 % of cases proceeding for sentencing (Roberts 2009).

The decision to submit a VIS is largely determined by personal considerations, but communal and social factors also play a role. Contextual factors such as the availability of victim resources, the outcomes victims receive, the procedures used to determine those outcomes, the manner in which victims were treated, and the information available to victims have all been shown to influence levels of participation (Haynes 2011). The social context in which the victimization occurred and was processed likewise affected participation. Higher levels of victim participation have been documented in urban areas, where anonymity often helps victims to avoid the discomfort or embarrassment associated with involvement with police and the courts (Haynes 2011).

Reasons for Submitting a Victim Statement at Sentencing

The reasons why victim choose to submit VIS or VPS vary across victimization types and jurisdictions. In the survey of crime victims in South Australia, Erez et al. (1994) report that the main reason that victims cited for providing a VIS was to ensure that justice was done (cited by more than two thirds of the respondents), with other reasons ranging from a wish to convey the harm to the offender, to considering the harm in sentencing. Research in other counties has found that victims are more likely to want the court to understand the effect of crime than to influence the sentence imposed. Other reasons cited by victims for submitting a statement include communicating the impact of the crime to the offender and in order to discharge a civic duty. Some studies have found that VIS also provides a platform for victims to convey their concerns, feelings, or views about the crime to the community at large or to authorities who were negligent in protecting victims or preventing the victimization (Cassell and Erez 2011). In some instances victims also wished to correct misconceptions the media created about them or the reasons for their victimization. Lastly, victims commonly do not consider their input, sentencing outcomes, or even capital penalty imposed on an offender as leading to “closure” (Bandes 2009) but rather describe participation in justice as allowing three interrelated benefits: empowerment, validation, and “moving on” (Erez et al. 2011). Clearly, there are many reasons why victims participate or submit an impact statement.

The Effect of Victim Input on Criminal Justice Administration

Researchers have evaluated the effects of victim input at sentencing in many ways, through the analysis of criminal justice statistics as well as surveys of criminal justice professionals such as prosecutors and judges. Research in jurisdictions that allow victim participation indicates that including victims in the criminal justice process does not cause delays or additional expense and that very few court officials believe that victim input creates or exacerbates problems or slows down the proceedings (Hillenbrand and Smith 1989). Court officials (prosecutors, defense attorneys) even in very busy jurisdictions generally view victim input positively.

Perceptions of the Judiciary

The views of judges are particularly relevant to the debate, as they are best placed to know whether victim impact statements contain relevant or extraneous information. One of the criticisms of victim impact statements is that they add little or nothing to the court record and may actually detract from the process by introducing irrelevant assertions about the offender. Surveys of the judiciary demonstrate that most judges see a benefit in receiving crime impact information directly from the victim by means of a victim impact statement. Although much of the information in VIS should already be reflected in evidence adduced at trial or prosecutorial submissions at sentencing, at times VIS may provide additional information useful to the judge who is determining sentence.

One high-profile example of judges’ appreciation of the VIS can be found in the case of Bernard Madoff who ran a “Ponzi” fraud scheme involving the lost of many billions of dollars, devastating the lives of large numbers of victims. In his sentencing remarks, the judge referred to the content of many VIS and listed what they revealed as one of the factors in his decision for a commensurate sentence (Cassell and Erez 2011). Surveys of judges in Canada and Australia found that most judges acknowledge that VIS contains information that is relevant to the purposes of sentencing (Erez and Rogers 1999; Roberts 2009). In short, the most important constituency at this stage of the criminal process – sentencing judges – perceives considerable utility in VIS. This is particularly true for crimes in which the impact on the victim was disproportionate or unusual or involved a large number of victims, violence, sex offenses, or crimes in which property was stolen, damaged, or misappropriated by deception.

Accuracy and Meaning of VIS

Although defense lawyers occasionally express concern that victim statements are inaccurate or express victim vindictiveness, research indicates that, in fact, victim statements seldom include inflammatory or prejudicial material (Roberts 2009). If this occurs, the prosecution can edit the statement to prevent the offending phrases from coming to the attention of the court. Another unfounded criticism of the VIS is that victims often exaggerate the level of harm that they sustained. This appears to happen only rarely, and when it does, judges and prosecutors report that these exaggerations involve financial matters, not emotional or mental suffering (Erez and Rogers 1999). In general, victims express feelings that others who have not been in their shoes or have prior experience with listening to victims may not consider valid or believable (Erez and Rogers 1999). Judges have also rejected the claim that the VIS serves as a platform for victims' vindictiveness or that similar statements by multiple victims suggest "mob vengeance" (Cassell and Erez 2011).

The Impact of Victim Input on Sentencing Practices

A major criticism of VIS is that its submission, and particularly its content, will tip the scales against the defendant, resulting in harsher sentencing patterns. Research largely refutes assumptions that victim input results in harsher sentences for defendants. Studies suggest that sentences are determined predominantly on the basis of legal considerations such as the seriousness of the offense and the offender's criminal record. As a practical matter, VIS will have limited relevance in the US federal system or in jurisdictions that employ a determinate sentencing scheme. Judges and prosecutors in one study (Erez and Rogers 1999) reported that VIS results in harsher sentences in some cases (e.g., when the intended harm was particularly serious or the crime was especially heinous) and in less severe sentences in other cases (e.g., when no harm occurred or when the harm was much less than would be expected). This may in part explain why aggregate studies (e.g., Erez and Tontodonato 1990) do not find any effect of victims' participation on sentencing trends. With respect to other jurisdictions such as Canada and England and Wales, the trends seem consistent; sentencing patterns did not get harsher after the introduction of victim impact schemes (Roberts 2009).

The Effect of VIS Submission on Victims' Welfare and Satisfaction with Justice

Findings on the effect of providing input into sentencing or participation in proceedings are inconsistent with respect to the issue of victim welfare, distress, and satisfaction with justice. At best, research has shown modest impact for certain types of victims or victimization. The evidence, however, may simply reflect a problematic implementation of the law, variations of the impact on different type of victims, or the use of different research methods and measures to study the impact. One study found that filing VIS usually results in increased satisfaction with the outcome (Erez and Tontodonato 1992), and studies of domestic violence victims (Cattaneo and Goodman 2010) have found that victim participation generally increases victims' satisfaction. Studies that examined victims in general, even with random assignment to various treatments, have found that VIS had no effect on victims' satisfaction with the criminal justice process or its outcome (Davis and Smith 1994).

Similarly, studies of the VIS programs in Canada and in Australia have revealed that victims who submit an impact statement are not necessarily more satisfied with the outcome or with the criminal justice system. The Australian study found that lack of impact of the input scheme may be related to the problematic implementation of the law; many victims did not realize that the purpose of the interview they had was to gather input that would be provided to the judge. In some cases, filing VIS heightens victims' expectations that they will influence the outcome. When that does not happen, victims may be less satisfied than those who do not submit a statement. In contrast, a comparative study of victims in the continental criminal justice systems (which allow victims a party status and significant input into the proceedings) suggests that victims who participated as subsidiary prosecutors or acted as private prosecutors were more satisfied than victims who did not participate (Erez and Bienkowska 1993). These differences suggest that in jurisdictions in which victims have more input into proceedings, levels of satisfaction are higher.

The most compelling evidence to support the position that victims see a benefit in submitting a statement comes from responses to the question "Would you submit a statement again if you were victimized?" This question has been posed to crime victims in several countries including the USA, Canada, and the United Kingdom, and the result is consistent: most victims state that they would submit a statement in the event of future victimization (see Roberts and Manikis 2011).

Research has identified three major factors that increase victims' overall satisfaction with the justice system and reduce their trauma. They include (a) procedural justice concerns such as whether the victim had the opportunity to be heard and whether he or she was treated with respect and informed of key developments in their cases (see Wemmers 1996), (b) the final decision of the court (e.g., whether the victims received financial compensation), and (c) whether there was an admission of guilt or request for forgiveness from the perpetrator. These variables were found to be more predictive of victim satisfaction than the severity of punishment imposed. This research suggests that victims' interests or concerns relative to proceedings are not simply to generate a severe sentence but pertain to the court addressing a broad range of issues that are within its purview, issues that a well-implemented VIS program can facilitate. Recent research has confirmed that victim overall participation in justice is influenced not only by the outcomes victims received but also the procedures used to determine those outcomes, the manner in which victims were treated, and the information available to victims about their rights and benefits in the cases they were involved (Haynes 2011).

Finally, research with victims of violent crime in several countries has revealed that victims appreciate judicial recognition of the harm they sustained. Judicial acknowledgment may be expressed in a direct statement if the victim is present in court at sentencing, or it may be articulated in the reasons for sentence (see Cassell and Erez 2011). Judges appear aware of the importance of this validation of the harm: a survey of the judiciary found that most reported acknowledging victim harm directly by addressing the victim in court or indirectly by citing victim impact in their reasons for sentencing (Roberts 2009).

To summarize, the substantial research record on victim impact statements demonstrates that when input schemes are properly administered, most victims benefit from the experience. Unfortunately, VIS regimes around the world are only partially successful in reaching all crime victims. One consequence of this is that only a minority of all victims benefit from participating in the sentencing process.

Conclusion

Victim input into sentencing is now recognized as an important component of criminal justice proceedings in adversarial legal systems around the world. Reports from practitioners indicate that few administrative problems, serious defense challenges, longer trials, or higher incarceration rates result from victim input into sentencing. However, because victim input challenges traditions and established patterns within the criminal courts, the implementation of this right still encounters resistance. Legislative reforms typically lack remedies for noncompliance. Victims' input at times depends on such fortuitous factors as whether a victim encounters criminal justice personnel who support victims' rights and inform him or her what they are or the levels of resources that a jurisdiction allocates to victims (see Haynes 2011). At present then, the victim input schemes at sentencing which employ victim impact statements as the vehicle for victim input have achieved partial success. Although only a minority of victims depose impact statements, most who do report the experience as being a positive one. Moreover, the benefits to victims do not seem to disadvantage offenders. As noted, there is a compelling argument that hearing the victim describe the impact of the crime may be beneficial to offenders.

The road to incorporating a victim voice in adversarial proceedings has not been smooth (see U.S. Department of Justice 1998) nor has it overcome all the barriers to implementation. Yet victim input into sentencing is no longer questioned and is considered an integral part of sentencing in both adversarial and continental legal systems.

Recommended Reading and References

Ashworth A (2002) Restorative rights and restorative justice. Br J Criminol 42:578–595

Bandes S (2009) Victims, "closure", and the sociology of emotion. Law Contemp Probl 72(2):1–26

Beloof D, Cassell P, Twist S (2010) Victims in criminal procedure. Carolina Academic Press, Durham

Cassell PG (2009) In defense of victim impact statements. Ohio State J Crim Law 6:611–648

Cassell PG, Erez E (2011) Victim impact statements and ancillary harm: the American perspective. Can Crim Law Rev 15(2):150–204

Cattaneo LB, Goodman LA (2010) Through the lens of therapeutic jurisprudence: the relationship between empowerment in the court system and well-being for intimate partner violence victims. J Interpers Violence 25(3):481–502

Davis R, Smith B (1994) Victim impact statements and victim satisfaction: an unfulfilled promise? J Crim Justice 22:1–12

Dichter ME, Cerulli C, Kothari CL, Barg FK, Rhodes KV (2011) Engaging with criminal prosecution: the victim's perspective. Women Crim Justice 21(1):21–37

Englebrecht CM (2011) The struggle for "Ownership of Conflict": an exploration of victim participation and voice in the criminal justice system. Crim Justice Rev 38(2):129–151

Erez E, Bienkowska E (1993) "Victim Participation in Proceedings and Satisfaction with Justice in the Continental Legal Systems: The Case of Poland." J Crim Justice 21:47–60

Erez E (1994) Victim participation in sentencing: and the debate goes on. Int Rev Victimol 3:17–32

Erez E, Roeger L, Morgan F (1994) Victim Impact Statements in South Australia: An Evaluation. Adelaide: Office of Crime Statistics, South Australian Attorney-General's Department

Erez E, Rogers L (1999) Victim impact statements and sentencing outcomes and processes. Br J Criminol 39:216–239

Erez E, Tontodonato P (1990) The effect of victim participation in sentencing on sentence outcome. Criminology 28:451–474

Erez E, Tontodonato P (1992) Victim participation in sentencing and satisfaction with justice. Justice Q 9:393–415

Erez E, Kilchling M, Wemmers J (eds) (2011a) Therapeutic jurisprudence and victim participation in justice: international perspectives. Carolina Academic Press, Durham

Erez E, Ibarra PR, Downs DM (2011b) Victim participation reforms in the United States and victim welfare: a therapeutic jurisdiction perspective. In: Erez E, Kilchling M, Wemmers J (eds) Victim participation in proceedings and therapeutic jurisprudence. Carolina Academic Press, South Carolina, pp 15–39

Haynes SH (2011) The effects of victim-related contextual factors on the criminal justice system. Crime Delinq 57(2):298–328

Hellerstein D (1989) Victim impact statement: reform or reprisal? Am Crim Law Rev 27:391–430

Herman S (2011) Parallel justice for victims of crime. National Center for Victims of Crime, Washington, DC

Hillenbrand S, Smith B (1989) Victim rights legislation: an assessment of its impact on criminal justice practitioners and victims. American Bar Association, Washington, DC

President's Task Force on Victims of Crime (1982) Final report. Government Printing Office, Washington, DC

Roberts JV (2009) Listening to the crime victim: evaluating victim input at sentencing and parole. In: Tonry M (ed) Crime and justice. University of Chicago Press, Chicago

Roberts JV, Manikis M (2010) Victim impact statements at sentencing: exploring the relevance of ancillary harm. Can Crim Law Rev 15(1):1–29
Roberts JV, Manikis M (2011) Victim personal statements at sentencing: a review of the empirical research. Office of the Commissioner for Victims and Witnesses of England and Wales, London
Strang H (2002) Repair or revenge: victims and restorative justice. Clarendon, Oxford
Tobolowsky PM et al (2010) Crime victim rights and remedies, 2nd edn. Carolina Academic Press, Durham
U.S. Department of Justice (1998) New directions from the field: victims' rights and services for the 21st century. U.S. Department of Justice, Washington, DC
Walklate S (2007) Handbook on victims and victimology. Willan, Cullompton
Wemmers J (1996) Victims in the criminal justice system. Kugler, Amsterdam
Zappala S (2010) The rights of victims v. the rights of the accused. J Int Crim Justice 8(1):137–164

Victim, Witness, Suspect Interviewing

► Evaluating Truthfulness: Interviewing and Credibility Assessment

Victim-Focused Prevention of Child Sexual Abuse

Yi-Ning Chiu and Benoit Leclerc
School of Criminology and Criminal Justice,
Griffith University, Brisbane, QLD, Australia

Overview

Reppucci and Haugaard (1989) have argued that before designing prevention programs to prevent sexual victimizations, we should first know what actually happens in sexual offenses. In other words, prevention should be based on empirical studies focusing on a detailed analysis of the actual offense not anecdotal accounts. To our knowledge, a very limited number of studies have been completed on victim resistance and its impact on the offender-victim interchange and related outcomes in child sexual abuse. A clearer understanding of the patterns in offender-victim interchanges may assist in informing and improving prevention programs. This entry begins with an overview of child sexual abuse, leading to current knowledge regarding offender-victim interchange and victim resistance. Second, it reviews victim-focused prevention programs and its effectiveness in reducing the number of sexual abuse incidents. Lastly, it makes a number of recommendations as to potential future directions.

Fundamentals

Child Sexual Abuse: An Overview

In child sexual abuse (CSA), the victim is most often in a trusting relationship with the offender with whom they share an emotional attachment. Approximately 95 % of CSA are committed by offenders already known to the victim, with 47 % who were related to or lived with the victim (Smallbone and Wortley 2000). Most incidents of sexual abuse occur in a familial or acquaintance relationship where the offender has established a previous nonsexual relationship with the child, often in authority and care-taking roles (Smallbone and Wortley 2000). The "grooming process" in CSA can be subtle, involving a graduation of giving attention, touching, and sexual talk to more explicit sexual behaviors, but is rarely associated with overt force and violence (e.g., Leclerc et al. 2011a; Smallbone and Wortley 2000). Offenders often target children who appear to have low confidence and self-esteem, who are more vulnerable to emotional exploitation (Elliott et al. 1995). CSA is a highly underreported form of crime, where studies have found as little as 3 % of incidents were reported to the police (Finkelhor and Dziuba-Leatherman 1994). Explanations such as reluctance to see the offender face serious police investigations, fear of stigmatization and retribution, and fear of blame or disbelief from non-offending adults have been speculated (Finkelhor and Ormrod 2001).

The Victim in Focus: Offender-Victim Interchange

Differences have been found in CSA regarding female and male children. Offender sexual behaviors such as touching and fondling are predominant for girls, whereas masturbatory, oral, and anal abuse is more likely for boys (Ketring and Feinaur 1999). Girls are more likely to be abused by familial offenders (in particular, by a stepfather), whereas boys are more likely to be abused by nonfamilial offenders (by a family friend) and also experience threats and force in conjunction with the abuse (Gold et al. 1998). These differences are indicative that offender methods and crime-commission processes vary for boys and girls. Due to the complex nature of offender-victim relationships, the offender-victim interaction is of particular importance in understanding the processes and outcomes in CSA.

Although recent reconceptualization of CSA from a criminological standpoint saw progress toward victim incorporation in crime analyses, only a few studies have examined offender-victim interaction in CSA with a deeper focus. An example is taking into account victim characteristics (e.g., Leclerc et al. 2009) and including the temporal order of events (e.g., Leclerc et al. 2011a). To date, research has suggested that the degree of intrusiveness of offenders' sexual behaviors in CSA is contingent on the victim performing sexual behaviors on the offender, which is in turn dependent on the offender's *modus operandi* (*MO*) or strategies used to gain the victim's cooperation (Leclerc and Tremblay 2007). Furthermore, in examining what factors might lead to a more intrusive sexual outcome, Leclerc et al. (2009) found that modus operandi as well as victim characteristics of gender and age had an important effect on crime event outcomes. Their findings suggest that offenders who get their victim to perform sexual behaviors on them (or victim participation) need to first secure the victim's trust through manipulation (including seduction, money and gifts, playing with the victim, and intoxicating the victim with alcohol or drugs). In fact offenders who used manipulation strategies were nearly six times more likely to make their victim perform sexual behaviors on them in sexual episodes; however, they were also less likely to perform penetration for female victims. In their study, Leclerc et al. (2009) suggested that offenders who used manipulation were more likely to perceive their victims as partners and therefore stopped short of performing penetration when met with resistance. Manipulation was found to decrease the risk of penetration for female victims, but not male victims, and it was assumed that female victims are more reluctant to experience intrusive sexual behaviors than male victims. This interpretation was supported by another pattern found in the study – an interaction effect was discovered between victim characteristics and sexual outcomes. Offenders were more likely to make the victim perform sexual behaviors on them in sexual activity as the victim became older when abusing a male victim. The opposite effect was apparent for female victims; older female victims were less likely to perform sexual behaviors in sexual episodes but were at greater risk of experiencing penetration.

A broader way to examine CSA is by mapping the complete sequence and events of the crime, such as through the use of crime scripts. Crime scripts evolved as an extension of the event model in rational choice perspective, where Cornish extended the analysis of offender decision-making from target selection to the whole process of the crime. A crime script (Cornish 1994) is a step-by-step account of the procedures and decisions used by offenders during crime commission. It breaks the crime down into steps or scenes, which enables potential for situational prevention at each step of the crime, and acts as a framework for breaking down the crime-commission process. The first empirical CSA crime script was developed by Leclerc et al. (2011a), which drew on the script model provided by Cornish (1994). It outlined the sequence of behaviors in CSA through a comprehensive framework. This script involved several stages including (1) entry to setting, setting where victim is first encountered; (2) instrumental initiation, gaining trust strategies; (3) continuation, proceeding to crime location strategies; (4) location selection, location for sexual contact; (5) instrumental actualization, isolation; (6) completion, gaining cooperation strategies; (7) outcomes, amount of time spent in sexual activity, victim participation, and offender sexual behaviors; and (8) post-condition, avoiding disclosure strategies. By breaking down the sequence of actions into a script, it allowed the authors to devise potential situational intervention strategies at different stages of the script, such as parent training sessions on modus operandi and context of abuse and information sessions for teachers on how to detect intrafamilial abuse and limit long-term and repetitive access to children. Offender-victim interaction was incorporated in the script by examining crime event outcomes; however, in this study victim resistance was not accounted for.

Victim Resistance

Victim resistance has been examined in a range of crimes. It has been argued that the interaction between victims and offenders largely determines the severity and level of injury during the crime (Block 1981). Studies that incorporated violence-resistance sequence analysis found that related injuries were attributed to the offender's physical attack rather than the subsequent result of victim resistance (Ullman and Knight 1991, 1992). This indicates that apparent inconsistencies in research may be explained by factoring in sequential relations of offender attack, victim resistance strategies, and assault outcomes. Although it is unclear whether victim resistance decreases the likelihood of injury, results generally indicate that victim resistance is successful and at worst inconsequential for different types of crime including robbery, burglary, sexual offenses against women, assault, and personal larceny (Tark and Kleck 2004).

However, there exists a discrepancy in CSA. The issue of victim resistance is more complicated in CSA, as victims represent a vulnerable group with less ability to physically resist against an adult offender. Finkelhor et al. (1995a, b) examined the efficacy of child's resistance and determined that children who resisted were more likely to suffer injuries and did not experience lower levels of victimization. A few earlier studies in victim resistance asked the offenders regarding what self-protection strategies children should use to avoid abuse (e.g., Elliott et al. 1995). Suggestions made included saying "no" to offenders, telling someone about the abuse, trying to get away, and learning about the proper touching of private parts. In more recent research, the efficacy of resistance strategies in CSA was analyzed based on empirical evidence in crime events. Of the different forms of resistance, it was discovered that the most successful resistance strategies included saying "no" and being assertive and the less effective resistance strategies included trying to get away, fighting back, and yelling for help (Leclerc et al. 2011b). In Leclerc et al.'s (2011b) sample of 94 adult offenders, a total of 74.3 % of offenders reported that their victim was able to avoid sexual contact at some point by telling them they do not want to participate in such activities, and 56.4 % by saying "no." Fighting back (11.8 %) and yelling for help (4 %) were the least successful strategies to avoid sexual contact.

Leclerc, Wortley, and Smallbone's (2010) work supported this finding and added another dimension into their analysis by examining the circumstances where children were more likely to successfully resist abuse. Factors such as pre-offense situation, offender modus operandi, and victim characteristics were examined in association with victim resistance in CSA. In relation to victim characteristics, it was found that younger girls were more likely to employ non-forceful verbal resistance than older girls and also use a greater number of strategies. Additionally, the older the child, the less likely they were to use non-forceful verbal resistance. As younger girls were more likely to use self-protection strategies, they were able to prevent episodes of abuse than older girls. This is in line with Asdigian and Finkelhor's (1995) study, which found that younger victims were more likely to use passive forms of resistance and a greater number of resistance strategies; however, older children were more likely to use active forms of resistance during the assault. Leclerc et al. (2010) found that victim resistance was also affected by offender modus operandi. Examination of offender modus operandi and victim resistance at a multivariate level showed that only violence was significantly linked to victim resistance. Offender violence was related to all forms of resistance, where the use of violence increased the chance of victim resistance by nearly five times when compared to strategies without violence (e.g., desensitization, gifts/privileges). Assuming that violence preceded resistance, it was interpreted that victims had no other way to avoid abuse but to resist in this context.

Victim-Focused Prevention Programs

Crime prevention of CSA can take many forms, including situational crime prevention, developmental crime prevention, community approaches, and criminal justice interventions. Current approaches to CSA prevention are dominated by formal interventions that take effect only after the crime has been committed, such as investigations, selective incapacitation, and offender rehabilitation (Smallbone et al. 2008). Against this backdrop, child protection programs have emerged as a non-criminal-justice response and alternative prevention to CSA incidents. Victim-focused prevention largely revolves around personal safety programs that seek to equip children with self-protective skills, including recognizing and avoiding sexually abusive situations. CSA prevention programs operate on several levels. The overall aim is to prevent potential occurrences and reoccurrences of child maltreatment, as well as reduce the likelihood of intergenerational transmission, where victims of CSA become perpetrators of the next generation. A composite approach to prevention programs was developed from the public health model of disease prevention, based on the level of severity: child maltreatment programs are commonly categorized as primary, secondary, or tertiary interventions (Tomison and Poole 2000). An example table of program classifications is adapted from Holzer, Bromfield, Richardson, and Higgins (2006), as seen in Table 1.

Victim-Focused Prevention of Child Sexual Abuse, Table 1 Types of child maltreatment prevention programs (Holzer et al. 2006)

Focus	Intervention level		
	Primary	*Secondary*	*Tertiary*
Child	Personal safety programs	Assertiveness training for “at risk” children	Therapeutic programs for abused children
Parents/ family	Home visiting programs	Parent education programs	Child protection service referrals
Community	General media awareness campaigns	Targeted media campaigns in “at risk” communities	Intensive community interventions

Primary interventions are strategies aimed toward targeting communities and building public resources. This includes personal safety programs taught in schools for children and parent education such as home visiting programs designed to increase knowledge of child development, assist in parenting skills, and normalize the challenges and difficulties faced in parenting. Secondary interventions target families with higher risk of child abuse, involving interventions such as home visitation, parent education, and skills training. This involves early screening to detect children who are potentially at risk due to the presence of one or more risk factors for child maltreatment, in order to prioritize early intervention. Tertiary interventions involve families in which child abuse has already occurred and is therefore seen as a reactive approach. Interventions such as statutory child protection and child welfare services seek to prevent the reoccurrence of abuse and mitigate the negative impacts of trauma and maltreatment. In summary, prevention programs differ in their focus, whether it is pre-intervention or post-intervention, or vary in their target audience. Education can be specified for parents, for children, for the family as a unit, for the community, or for the offender/potential offenders. Greater emphasis has been placed on primary and secondary interventions following reports indicating greater difficulties in treating abusive parents at the tertiary level, as such abuse may have become a fixed pattern in parent-child interactions (Geeraert et al. 2004).

Victim-Focused Prevention of Child Sexual Abuse, Table 2 Core concepts in the curricula of school-based prevention programs (adapted from MacIntyre and Carr 2000)

Core concept	Description
1. *Body ownership*	The child's body belongs to herself/himself and has a right to control access to her or his body
2. *Touch*	There is a distinction between "good" touch and "bad" touch, where a good touch can be permitted and bad or confusing touches should be refused
3. *Saying "no"*	The child has the right and should practice the skill of saying "no"
4. *Escape*	The child should be aware of the importance of escaping potential offenders and practice the skills necessary for escape
5. *Secrecy*	The child should be able to talk about any touch he or she is asked to keep secret
6. *Intuition*	The child should trust his or her feelings when a situation does not feel right
7. *Support systems*	The child should identify adults that he or she can turn to for help in case of disclosure and seek help from another adult if the first does not listen or believe the disclosure
8. *Blame*	The child is not to blame if he or she is abused or victimized

Recognition of the importance of proactive strategies has increased among government bodies, community alliances, and nongovernment organizations. Victim-focused safety programs adopted across schools are included in this category.

MacIntyre and Carr (2000) summarized the core concepts generally taught in school-based prevention programs, including body ownership, touch, saying "no," escape, secrecy, intuition, support systems, and blame (see Table 2). Body ownership and saying "no" try to teach the child that they have a right to refuse other people access to his or her body. This also teaches children to be assertive and that the child's body belongs to himself/herself. Touch and intuition methods rely on the child's ability to detect what feels confusing or wrong, including the concepts of "good touch, bad touch." Escape includes teaching methods and skills to escape potential offenders. Secrecy and support systems encourage disclosure from the child by assuring children they can talk about anything (even what they were told to keep a secret) and identify a trusted network of adults to whom they can do so. Programs also try to reduce victim blame by assuring children that they are not responsible and are not to blame if he or she is abused.

Are Self-Protection Programs Effective in Reducing Child Sexual Abuse Incidents?

Since its early development, programs such as the Protective Behaviors (PB) program from the 1970s have been subject to several criticisms. For example, early programs were criticized for not incorporating children's developmental and learning needs in the concepts of the program (Finkelhor and Strapko 1992). Also, it was pointed out that the child's ability to clearly recognize and safely respond to unsafe feelings was ineffective, as adults are generally viewed as protectors (Briggs 1991). This is particularly relevant in CSA where 95 % of incidents are perpetrated by adults with prior established relationships with the victim (Smallbone and Wortley 2000). For young children, the attachment toward the offender combined with lack of sexual knowledge means the abuse might not be recognized by the child. This limits the effectiveness of understanding what constitutes as a "bad touch" and increases the risk of victim shame and self-blame.

Outcomes of child safety programs are generally determined by evaluations during a follow-up period. There are three main types of evaluations (Richardson and Tomison 2004): *process evaluations* examine how a program has been implemented or practiced; *impact evaluations* investigate whether a program has achieved its operational goals, such as increased knowledge, skill gain, or increased disclosure. Finally, *outcome evaluations* determine whether the program has resulted in reductions of subsequent incidence of CSA. The latter has rarely been examined by studies, and to date only one study by Gibson and Leitenberg (2000) reported positive results: a survey of 825 female undergraduates revealed that 8 % of students experienced subsequent child sexual victimization after attending protection programs, compared to 14 % of students who had not participated in a program. This result is a contrast to Finkelhor et al.'s (1995a, b) study, which examined the efficacy of self-protection in several types of victimization, including assaults by peers, family members, gangs, kidnappings, and sexual offenses. First, Finkelhor et al. (1995a) completed telephone interviews with a nationally representative sample of 2,000 children who participated in school-based prevention programs. Children who took part in prevention programs were more likely to use self-protection strategies, to disclose abuse when it does occur, and to perceive themselves as having been more effective in avoiding or minimizing the harm of sexual victimization and were less prone to self-blame. It was also found, however, that children involved even in the most comprehensive personal safety programs were more likely to suffer injuries in coping with sexual assaults. According to the authors, this finding was perhaps related to children's more aggressive resistance. In a follow-up study, Finkelhor et al. (1995b) then recontacted previous participants for a second interview, with the average delay being 15 months. Children involved even in the most comprehensive personal safety programs did not experience lower levels of completed victimizations.

However, in actual CSA situations using recent offender self-report data, some strategies were found to be successful in avoiding sexual abuse, such as saying "no" (Leclerc et al. 2010). Although the sample used in the study was limited to children who were ultimately abused at some point, the research indicates that some forms of resistance may be effective in avoiding episodes of sexual contact in real-life situations. In summary, while there is general evidence supporting process and impact evaluations of child safety programs (MacIntyre and Carr 2000), evidence regarding outcome evaluations is lacking and inconsistent. Whether programs are successful in reducing the incidence of subsequent sexual victimization is still unclear (Smallbone et al. 2008).

Where to from Here?

A better understanding of offender-victim interchanges has the potential to contribute to child sexual abuse prevention and inform prevention programs. Below is a list of general recommendations to guide future studies.

Sequence of Behaviors (Offender-Victim/ Offender-Outcome)

Although it has been demonstrated that offender-victim interchanges can impact the outcome, the time sequence of behaviors has not been incorporated. For example, offender response to victim reaction and associated outcomes has yet to be studied. This sequence of behaviors can be valuable in understanding why some resistance strategies are more effective than others. An excellent example of this kind of study is Luckenbill's (1977) work in criminal homicide. In the examination of 70 cases, he found that victims did not play a passive role in the event. Murder was found to be the outcome of a dynamic interchange between the offender, victim, and bystanders. The actions of the actors were shaped in part by each other. Alternatively in sexual offending, when incorporating sequence analysis to sexual offenses against women, it was found that victim injury generally did not precede offender attack (Ullman and Knight 1991, 1992). Therefore, victim resistance is often adopted in response to offender violence. By factoring in sequential relations of offender attack, victim resistance, and assault outcomes, it was further discovered that resistance rarely precedes injury, where 85 % of women studied responded with physical force only in reaction to the offender's initiated violence (Ullman and Knight 1992), and that most forms of resistance are not significantly associated with higher rates of subsequent victim injury (Kleck and Sayles 1990). Similarly, by incorporating sequence analysis to CSA, it will provide a better framework to understand the offense process and interpret the associated outcomes.

Scripts Based on Victims

Although Leclerc et al. (2011a) developed an empirical script model in CSA, the sequence of events can be examined in greater detail using scripts or script tracks. For situational purposes, it is important to be crime-specific. Other offense characteristics such as victim characteristics, victim resistance, encounter settings, or offense location could be incorporated. Studies have already shown the victim gender and age have an impact on the crime (e.g., Leclerc et al. 2009, 2010). For example, what is the crime script for sexual offenses against younger boys compared to older boys? What is the crime script for sexual offenses in public places compared to private locations? Are there differences in offender modus operandi or victim reaction? By operating at a higher level of specificity, crime scripts may offer further insight into prevention strategies. This is valuable for prevention programs. If programs can be tailored to the victim's unique needs and circumstances, based on higher likelihood of the situations they may encounter, more effective and informed advice can be given. This is congruent with Asdigian and Finkelhor's (1995) recommendation, which does not encourage a one-size-fits-all approach to CSA prevention.

What Works, What Does Not, and What Is Promising?

Further impacts and links regarding offense characteristics and victim resistance can be investigated in greater detail and with larger sample sizes. Other factors such as duration of the abuse, escalation pattern, victim gender, guardianship, offender-victim relationship, and settings can be explored. In particular, analyzing the crime process for girls and boys has potential to inform self-protection programs. Examining the efficacy of victim resistance in child sexual abuse empirically is also important. Although some studies have found evidence that some self-protection strategies may be more effective than others [i.e., saying "no" and saying they do not want to participate (Leclerc et al. 2011b)], whether these strategies work for boys and girls of different ages and whether the effectiveness varies in different contexts can be explored. This evidence gives practitioners clearer indications of what works and what does not and what has the potential to be developed. An example to draw from may be in the area of sexual offenses against women, as victim resistance in this area has been more thoroughly explored in comparison. For instance, several types of victim response have been central to women's rape avoidance studies, including forceful physical resistance, non-forceful physical resistance, forceful verbal resistance, and non-forceful verbal resistance. Forceful physical resistance has been consistently related to successful rape avoidance and found to be directly related to less severe sexual abuse without increase of physical injury (Ullman and Knight 1991). The method of resistance also affected the outcome, where forceful fighting and screaming was related to decreased severity of sexual abuse. Non-forceful verbal resistance such as pleading, begging, and reasoning, however, were related to greater severity of sexual abuse (Ullman and Knight 1991). Better understanding of victim resistance strategies in CSA, its efficacy and associated outcomes, and the contexts in which it occurs has important implications for prevention programs.

Conclusion

In conclusion, despite existing knowledge in the field, there are still many avenues to be investigated in CSA. Currently the effectiveness of child protection programs in reducing subsequent victimization has not yet been validated. The design of early victim-focused prevention programs was subject to various criticisms. Some concepts adopted in programs, such as "bad touch" or "unsafe feelings," are ineffective for children (Briggs 1991) and based on common assumptions rather than empirical research. This is in part due to a lack of understanding of the processes that occurs in child sex offenses. Existing literature in CSA indicates that victim characteristics (such as gender and age) and offender modus operandi are important elements that can influence abuse outcomes in terms of severity and level of injury (Leclerc et al. 2009). These factors also impact victim resistance (Leclerc et al. 2010). However, there is still much to be explored and understood regarding the child sexual abuse crime event. The complexity of CSA means that without a more detailed context it can be hard to determine the complete behavioral sequences and interactions in the crime. This is an important area to investigate, as the understanding of what happens during an offense is essential in driving the design of prevention programs (Reppucci and Haugaard 1989; Smallbone et al. 2008). Building further knowledge of patterns in offender-victim interchange may help to interpret crime event outcomes and identify intervention strategies for prevention programs. The recommendations mentioned above provide a guideline for future research to inform practitioners; the pursuit of such knowledge can improve practical intervention and programs aimed at reducing the risk of child sexual abuse.

Recommended Reading and References

Asdigian N, Finkelhor D (1995) What works for children in resisting assaults? J Interpers Violence 10:402–418

Block R (1981) Victim-offender dynamics in violent crime. J Crim Law Criminol 72:743–761

Briggs F (1991) Child protection programs: can they protect young children? Early Child Dev Care 67:61–72

Cornish DB (1994) The procedural analysis of offending and its relevance for situational prevention. In: Clarke RV (ed) Crime prevention studies, vol 3. Criminal Justice Press, Monsey

Elliot M, Browne K, Kilcoyne J (1995) Child sexual abuse prevention: what offenders tell us. Child Abuse Negl 19:579–594

Finkelhor D, Dziuba-Leatherman J (1994) Children as victims of violence: a national survey. Pediatrics 94:413–420

Finkelhor D, Ormrod RK (2001) Factors in the underreporting of crimes against juveniles. Child Maltreat 6:219–229

Finkelhor D, Strapko N (1992) Sexual abuse prevention education: a review of evaluation studies. In: Willis DJ, Hoder EW, Rosenberg M (eds) Child abuse prevention. Wiley, New York

Finkelhor D, Asdigian NL, Dziuba-Leatherman J (1995a) The effectiveness of victimization prevention instruction: an evaluation of children's responses to actual threats and assaults. Child Abuse Negl 19:141–153

Finkelhor D, Asdigian NL, Dziuba-Leatherman J (1995b) The effectiveness of victimization prevention programs for children: a follow-up. Am J Public Health 85:1684–1689

Geeraert L, Noortgate WVD, Grietens H, Onghena P (2004) The effects of early prevention programs for families with young children at risk for physical child abuse and neglect: a meta-analysis. Child Maltreat 9:277–291

Gibson LE, Leitenberg H (2000) Child sexual abuse prevention programs: do they decrease the occurrence of child sexual abuse? Child Abuse Negl 24:1115–1125

Gold SN, Elhai JD, Lucenko BA, Swingle JM, Hughes DM (1998) Abuse characteristics among childhood sexual abuse survivors in therapy: a gender comparison. Child Abuse Negl 22:1005–1012

Holzer PJ, Bromfield LM, Richardson N, Higgins DJ (2006) Child abuse prevention: what works? The effectiveness of parent education programs for preventing child maltreatment. Research brief no. 1, from http://www.aifs.gov.au/nch/pubs/researchbrief/rbl.html

Ketring SA, Feinaur LL (1999) Perpetrator-victim relationship: long-term effects of sexual abuse for men and women. Am J Fam Ther 27:109–120

Kleck G, Sayles S (1990) Rape and resistance. Soc Probl 37:149–162

Leclerc B, Tremblay P (2007) Strategic behavior in adolescent sexual offenses against children: linking modus operandi to sexual behaviors. Sex Abuse J Res Treat 19:23–41

Leclerc B, Proulx J, Lussier P, Allaire J-F (2009) Offender-victim interaction and crime event outcomes: modus operandi and victim effects on the risk of intrusive sexual offenses against children. Criminology 47:595–618

Leclerc B, Wortley R, Smallbone S (2010) An exploratory study of victim resistance in child sexual abuse: offender modus operandi and victim characteristics. Sex Abuse J Res Treat 22:25–41

Leclerc B, Wortley R, Smallbone S (2011a) Getting into the script of child sex offenders and mapping out situational prevention measures. J Res Crime Delinq 48:209–237

Leclerc B, Wortley R, Smallbone S (2011b) Victim resistance in child sexual abuse: a look into the efficacy of self-protection strategies based on the offender's experience. J Interpers Violence 26:1868–1883

Luckenbill DF (1977) Criminal homicide as a situated transaction. Soc Probl 25:176–186

MacIntyre D, Carr A (2000) Prevention of child sexual abuse: implications of programme evaluation research. Child Abuse Rev 9:183–199

Reppucci ND, Haugaard JJ (1989) Prevention of child sexual abuse: myth or reality. Am Psychol 44:1266–1275

Richardson N, Tomison AM (2004) Evaluating child abuse prevention programs. Child Abuse Prevention Resource Sheet 5, from http://wwww.aifs.gov.au/nch/sheets/rs5.html

Smallbone S, Wortley R (2000) Child sexual abuse in Queensland: offender characteristics and modus operandi (full report). Queensland Crime Commission, Brisbane

Smallbone SW, Marshall W, Wortley R (2008) Preventing child sexual abuse: evidence, policy and practice. Willan, Cullompton

Tark J, Kleck G (2004) Resisting crime: the effects of victim action on the outcomes of crimes. Criminology 42:861–909

Tomison AM, Poole L (2000) Preventing child sexual abuse and neglect. Findings from an Australian audit of prevention programs. http://www.aifs.gov.au/nch/pubs/auditreport.html

Ullman SE, Knight RA (1991) A multivariate model for predicting rape and physical injury outcomes during sexual assaults. J Consult Clin Psychol 59:724–731

Ullman SE, Knight RA (1992) Fighting back: women's resistance to rape. J Interpers Violence 7:31–43

Victimhood

▶ Victimization, Gender, and the Criminal Justice System

Victimization Incidence

▶ Understanding Victimization Frequency

Victimization Rate

▶ Understanding Victimization Frequency

Victimization, Gender, and the Criminal Justice System

Sandra Walklate
Department of Sociology, Social Policy
and Criminology, University of Liverpool,
Liverpool, UK

Synonyms

Crime and justice; Men; Victimhood; Women

Overview

Since the development of the criminal victimization survey in the late 1960s, appreciation of the impact of crime has risen up national and international criminal justice agendas. Its more recent refinement as a tool, alongside the input of feminist work, has contributed to a rising awareness of the gendered nature of that impact. This is especially the case in respect of sexual violence. However, as awareness of this impact has increased so has awareness of the shortcomings of criminal justice systems to appreciate and respond appropriately to this kind of criminal victimization. In this chapter some of the tensions that exist between what is known about the gendered nature of sexual violence and what is done about it will be considered. By way of conclusion areas for further work and investigation will be suggested.

What Does Gender Mean?

It is beyond doubt that gender as a word is now increasingly used not only in everyday language and but also routinely in policies and practices. The word gender is frequently used in monitoring and other questionnaires in which people are routinely asked: what gender are you? However, in asking this question, those same documents follow this up by providing the respondent with the option of the choice of being either male or female: categories that denote sex. These practices illustrate, more profoundly than in any other way, both the embrace of gender as a word and the deep confusion between sex and gender that this embrace exemplifies. In many ways gender has become the word that has, in a rather unthinking way, replaced sex. In a chapter of this kind then it is important to appreciate what this author takes, not only the words sex and gender to mean, but also what they refer to as concepts.

Put simply, sex is a biological given. A person is either male or female. Whilst there might be some confusion about the nature of this biological given for a small percentage of people, being male or being female is a *biologically* ascribed attribute. Gender refers to the *socially* achieved qualities of being male or female, that is, being a man or a woman. More specifically the concept of gender alludes to socially constructed understandings of masculinity and femininity. So, whilst everyone is biologically either male or female, sociologically, being male or being female does not necessarily equate with being masculine or being feminine. These latter attributes refer to the social expectations associated with biological characteristics and they are expectations that may vary over time, space, and place. Moreover, it is important to remember that the term "gender," by implication, refers to both men and women.

Why should being clear on the distinction to be made between these terms matter for understandings of gender, victimization and the criminal justice system? In part the answer to this question is important because of the observations already made concerning the substitution of the word sex for gender. The mere substitution of one word for another in policy and practice might at first glance suggest that, the questions that gendered understandings of the world pose, have been addressed. However, such an understanding would be superficial, since gendered approaches demand changes at the level of conceptual understanding as well as practices so informed, rather than merely a change in words.

One way of making sense of how the conflation between sex and gender has come about is to situate this within the broader policy drive for gender mainstreaming. In particular, it is through the conduit of gender mainstreaming that sexual violence has become a part of the policy agenda on a global scale, from the United Nations to the European Community, to national governments. Moreover it is this drive that sets a framework for policy development for their respective member states and/or organizations. Indeed Fraser (2009) posits that as a result of these activities, gender equality is now part of the social mainstream but has yet to be realized in practice. Whilst Fraser (2009) is addressing a wide range of equality issues (unequal pay and so on), the gap between theory and practice that she comments on is a very pertinent one for the criminal justice system. It is this gap between theory and practice, that forces a consideration of the implications for making sense of the ambivalent relationship between what is *known* about the relationship between gender and victimization and what it is that is *done* in the light of this knowledge.

By implication the question of gender challenges not only intellectual understandings of victimization but also what is to be done on the basis of those understandings. In addressing this challenge this chapter will do three things. First it will briefly reflect upon the domain assumptions associated with the concept of victim and victimology: the theoretical context. Second, it will map these concepts against the domain assumptions of the criminal justice system in which one particular kind of victimization that poses considerable challenges to that system will be considered, men and women who present themselves as victims of rape: the policy context. Finally it will reflect upon the extent to which the experiences of the rape complainant are matched by the criminal justice response: the gap between theory and policy.

The Theoretical Context: What Does the Term 'Victim' Mean?

When the word "victim" is gendered, as in French for example, being la victime, it is feminine. The links between this word and being feminine implies that the passivity and powerlessness associated with being a victim are also feminine qualities. This link is problematic for those working within the feminist movement who prefer to use the term "survivor" to try to capture women's resistance to victimization. At the same time the tensions between being a victim or being a survivor are also problematic for others interested in criminal victimization since this either/or distinction fails to capture the *processes* of victimization. In other words it is possible that an individual at different points in time in relation to different events could be an active victim, as passive victim, an active survivor, a passive survivor, or at a point on a whole range of experiences in between. The label "victim," a consequence, seems quite sterile when considered in this way.

There is, however, another problem associated with the word "victim" that is derived from appreciating the process whereby an individual becomes identified as a victim. This is connected with Christie's (1986) concept of the "ideal victim." In other words, there are certain assumptions attached to the label "victim" that mean not everyone actually acquires the label of victim. For Christie the "ideal victim" is the Little Red Riding Hood fairy story victim; a young, innocent female out doing good deeds, who is attacked by an unknown stranger. Indeed this "ideal victim" fits all the common sense stereotypes of the "legitimate" victim of rape, as will become clear in the example to be developed in this chapter. It can result in some people being "deserving" victims, that is acquire the label of victim very readily and easily, and others as "undeserving" victims and may never acquire the label of victim at all. This distinction between deserving and undeserving victims and how it impacts upon experiences of the criminal justice process is one of the issues that we shall be concerned with. This distinction suggests a hierarchy of victimization experiences with some people more likely to be recognized and embrace the label of victim. Such a hierarchy alludes to a third problem intrinsically associated with the term victim: it is also a highly gendered. As shall be seen men and women find themselves in potentially quite different places in this hierarchy even though they may have experienced the same kind of criminal victimization. It will be informative, then, to reflect upon how this concept has become gendered and what kind of impact this gendering has. But first, a few words on the interconnected domain assumptions of victimology.

The origins of victimology are located in the work of Von Hentig and Mendelsohn, amongst others. Like many other like-minded intellectuals of the late 1940s and the 1950s they were perplexed by the events that had happened in Germany during that time. That perplexity led Von Hentig and Mendelsohn to think about the dynamics of victimization, though being lawyers how they understood those dynamics was very much at the level of the individual and very much informed by their legal training. Both of them were concerned to develop ways of thinking about the victim that would enable the victim to be differentiated from the non-victim. In other words both were clearly suggesting that there is a normal person who, when the victim is measured against them, the victim falls short. Each developed victim typologies. Von Hentig's typology worked with a notion of "victim proneness." He argued that there were some people, by virtue of their socio-economic characteristics who were much more likely to be victims (in this case of crime) than other people. These he identified as being women, children, the elderly, the mentally subnormal etc. If this categorization is thought about from a gendered perspective it suggests that Von Hentig thought the normal person against whom the victim was to be measured was the white, heterosexual male. Von Hentig's work has been very influential in the development of victimological thinking and is perhaps most keenly identified in the concept of "lifestyle" that has informed much criminal victimization survey work.

Mendelsohn adopted a more legalistic framework in developing his typology. His underlying concept was the notion of "victim culpability." Using this concept he developed a sixfold typology from the victim who could be shown to be completely innocent, to the victim who started as a perpetrator and during the course of an incident ends as the victim. Arguably his typology is guided by what might be considered a reasonable or rational way of making sense of any particular incident given the nature of the law. Again gendered thinking suggests that his understanding of reasonable equates with that which the white, heterosexual male would consider reasonable. This is especially demonstrated in the later work that was generated from Mendelsohn's ideas, in which victim culpability became translated into "victim precipitation." The furore that greeted the publication of Amir's (1971) work, whose analysis of police data on recorded rapes that suggested 19 % of these were "victim precipitated" stands as a case in point. That furore, driven by the feminist movement, railed against the way in which this work perpetuated myths about rape. However, as shall be demonstrated, such concepts, and the myths that they serve, still have considerable cultural and professional resonance despite their questionable acceptability intellectually.

From these beginnings the concepts of lifestyle and victim precipitation have formed the core of much traditional victimological work and illustrates the influence of positivism on victimology. This kind of victimology reflects an emphasis on measuring differences, seeing those differences as being somehow abnormal, and looking for explanations of those differences that lie outside of individual choice. In all of this, the development of the concept of lifestyle has been of central importance in perpetuating certain assumptions about criminal victimization and how to measure it so we shall consider its relevance to the discussion here in a little more detail.

The concept of lifestyle was developed initially and related to criminal victimization by Hindelang et al. (1978). This book presents a way of thinking about personal victimization in which lifestyle refers to "routine daily activities, both vocational activities (work, school, keeping house etc.) and leisure activities" (ibid: 241) and in which individuals are constrained by role expectations and structural characteristics that are reflections of their demographic positions (class, age, sex, ethnicity etc.). The authors argue that there is a direct link between an individual's routine daily activity and their exposure to high-risk victimization situations from which personal victimization occurs. It is important to note the implicit acceptance of the view that the chances of personal victimization vary in relation to the amount of time spent in public places. This assumption does two things simultaneously. First it reflects a very male view of what constitutes a problematic arena for people: the public. As a result it hides the world of the private (all that goes on "behind closed doors") as being an arena in which personal victimization might occur. In other words this assumption hides the world of women, children, and some men. Second, in focusing on what goes on in public it implicitly accepts the view that this is the context in which legitimate criminal victimization might occur. Thus the victim out doing good deeds who is attacked in public by a stranger stands as an "ideal victim" (Christie 1986) provided that the victim was out doing good deeds and did nothing to "invite" the attack.

Centering lifestyle reflects an interesting twist in the domain assumptions of victimology. No longer is the white, male exempt from victim status, as in the work of Von Hentig and Mendelsohn. The concept of lifestyle centers the importance of public space and places men in the foreground of potential victimhood with women and children in the background, in the world of the private. However, despite the potential of this assumptional twist, the early criminal victimization work that emanated from this development did little to build on its potential for problematizing men and masculinity. Indeed, it was not until the early 1990s, arguably as a consequence of the input of radical feminism, that the conundrums that this work generated (for example, the empirical finding that whilst men were most at risk from criminal victimization they expressed the least concern about it) were subjected to closer (gendered) scrutiny.

What feminist work did was, not only to offer a different conceptual apparatus, informed by the concept of patriarchy, with which to understand women's lives in general, it applied that apparatus to the crimes that affected women's lives. Those crime that occurred in private, behind closed doors; mainly rape and "domestic" violence. From this viewpoint the key variable in understanding women's experiences including their experiences of criminal victimization is, therefore, men's power over women. Put simply radical feminism clearly focuses attention on all those criminal victimizations that go on in private between people who are for the most part known to each other. This included rape, sexual assault, and domestic violence. The seminal work of Russell (1990) marked a key turning point in placing these acts on the criminological and victimological agenda, aided by the conceptual development of a ‘continuum of sexual violence’ introduced by Kelly (1988). This work rendered men, and their relationship with masculinity, as centrally problematic to understandings of the nature of women’s victimization. In so doing it made men as victims of such violence invisible (see also Graham 2006). Conceptually this in itself was problematic since it did not resonate with the relatively small volume of work that suggested men too could be victims of such violence and that it occurred in the community as well as in prison.

In one intervention on this invisibility, Goodey (1997) used the phrase “big boys don’t cry” to help explain how it was that young men struggled with being victimized. Put simply it contradicted their understanding of themselves as men. The epidemiological work of Coxell et al. (1999) on non-consensual sex between men, the pioneering work of Lees (1999) on male victims’ experiences of the criminal justice system, the in-depth interviewing work with male victim of rape reported by Allen (2002), all add both quantitative and qualitative depth to our understandings of men’s differential experiences of violence, victimization, and their responses to it. More recently further detail has been added to this picture in appreciating how blame works for male and female victims of rape (Davies et al. 2006), the continued efficacy of the stranger rape stereotype as mediated by gender (Anderson 2007), the comparative role of shame in reporting behaviors (Weiss 2010), and the persistent inability of the law to recognize the gendered nature of the act of rape i.e. that it includes men as victims (Graham 2006). The work referenced here affords just a snapshot of the ongoing intellectual engagement with rape as both a male dominated act and a male experienced process of victimization.

To summarize, whilst much victimological work leaves the impression that victims are not likely to be male, this is clearly empirically not the case. Much male violence is committed against other men, and in recent years there has been an increasing awareness that much of that male violence against other men is also sexual violence. So quite clearly men can be victims even if they have been relatively hidden from the victimological gaze. The empirical fact of their (sexual) victimization, however, has not only been relatively invisible in conceptual understandings of victimization, but also in criminal justice responses to such victimization. The next section considers the extent to which, despite this intellectual context, gendered understandings have permeated or failed to permeate policy responses to sexual victimization of both men and women.

The Policy Context: Implementation Failure?

Charting the nature and extent of sexual violence and developing appropriate policy responses to this problem is fraught with difficulties. Nevertheless is it the case that much effort over the last 20 years has been focused on improving the response of all the criminal justice agencies to the rape victim across many different jurisdictions. These improvements have ranged from the development of specialized police units designed to respond to rape complainants, to the establishment of specialized investigative procedures, specialist health responses, to changes in the law. In addition legislation designed to curtail the use of questioning about the past sexual history or character of complainants has been introduced both in the UK, Australia, New Zealand, and United States of America. Nevertheless the conviction rate for rape cases in England and Wales remains low at 5.6 %, a finding replicated across Europe and elsewhere (Daly and Bouhours 2009). Her Majesty’s Crown Prosecution Service Inspectorate Thematic Report published in 2007 attempted to offer some understanding of why the conviction rate remains so low in England and Wales despite the evidence that the reporting of complaints of rape has increased. Its findings reveal that, of 573 recorded complaints of rape (and it must be noted that all these cases involved female complainants), in 491 (85.7 %) of the cases the suspect and the victim knew each other, in 102 of these cases (20.8 %) and the victim withdrew the charges, leaving 398 cases. Of these the suspect was charged in 160 cases, the main reason for not charging was insufficient evidence (229 cases). Of 160 cases in which the suspect was charged, 75 case files were examined further that revealed that out of these 75 cases there were 39 convictions, 20 as a result of a guilty plea, 19 on the basis of a jury decision.

This snapshot of what happens to cases in England and Wales is illustrative of problems experienced across a range of jurisdictions. It points to not only the problem of attrition but also the points at which cases of rape pose problems for the criminal justice system. This data points to at least three of these critical moments; the point at which the incident is reported (this involves the police); the point at which a case for charging a suspect is being made (this involves the Crown Prosecution Service), and the point of conviction (this involves the court and the jury). Indeed the Stern Review (2010: 8), of how rape is handled in the UK, concluded that despite these problems the:

> . . .policies are the right ones and we have only a few changes to recommend. In some areas the policies are applied consistently and with commitment by all involved. The policies are not the problem. The failures are in the implementation.

Is this analysis correct? The following summary of the Warboys case (Brown and Walklate 2011: 2–3), provides some insight into why this analysis might not be the appropriate one.

Warboys was the driver of a London black cab. He was first identified as a suspect in cases of sexual assault reported to the police in July 2007 but it was not until February 2008 that he was charged with a large number of such offences dating back to October 2006. He was found guilty and sentenced in April 2009. Some 80 victims contacted the police after Warboys' arrest, many of whom had not previously reported their assault to the police. The Independent Police Complaints Commission's (IPPC) enquiry into this case noted that not only there were failures to comply with Standard Operating Procedures for the investigation of rape but also systemic failures to identify and link offences. The IPCC concluded that there had been a poor initial investigative response, a failure to trust victims, failures in front line supervision, lack of cross checking systems to link similar offences, and they found that the detective sergeant missed crucial investigative opportunities. Such a catalogue of failings is not peculiar to the UK. Jordan (2008) documents a similar set of failings during the investigation of a case in New Zealand that revealed majors flaws within policing practices and the wider criminal justice system.

As has already been suggested, embedded here are patterns of a bigger picture relating to sexual violence, like for example, under reporting, attrition of cases as they drop out at key stages of the criminal justice process, disbelieving of complaints, and giving men's explanations for what happened greater credence. For example, in England and Wales the British Crime Survey reveals that a significant number of victims still tell no one about their assaults with only 11 % making a report to the police (Povey et al. 2009). Indeed in Brown et al. (2010a), two-fifths (40 %) told no one about their experience and while men and women were equally likely to tell someone, who they told varied somewhat. Of those who did tell, the most common confidantes were friends, relatives or neighbors (44 %). The way in which women's (and men's) experiences of rape and sexual assault are silenced, within and outwith the criminal justice system, is subjected to critical analysis by Jordan (2011). This is an analysis that needs to be considered against criminal victimization survey data. For example, Povey et al. (2009) also report that nearly 1 in 4 women (23.3 %) and 1 in 33 men (3 %) men had experienced some form of sexual assault (including attempts) since the age of 16. For rape (including attempts) the prevalence was nearly 1 in 20 women (4.6 %) and 1 in 200 men (0.5 %) since the age of 16. Taken together, findings like these speak volumes as to the nature and extent of such violence and the wider (lack of) recognition of it.

As the thematic report referenced above indicates much work remains to be done in supporting complainants of rape through the criminal justice process. However this need for support needs to be put into a wider context as Jordan's (2011) analysis suggests. Complaints of sexual assault face challenges in being heard and responded to in a wider cultural context. For example, an Amnesty International sponsored poll (ICM 2005) revealed that 26 % of those asked thought that a woman was partially or totally responsible for being raped if she was wearing sexy or revealing clothing, and more than one in five held the same view of a woman who had had many sexual partners, with 30 % "blaming the woman" if she was drunk. These findings echo those of Brown et al. (2010b) and highlight the kinds of attitudes that are exploited in the criminal court during the course of a trial.

Women's experiences of the rape trial arguably are still governed by the need to be beyond reproach or blame in every respect especially in the witness box. The process of cross-examination in the witness box is, more often than not, designed to render a woman's evidence less than believable on the grounds of her womanhood rather than the quality of her evidence. It is well established that trials involving sexual offences requires counsel to test the credibility of complainants. This, of course, is a strategy not confined to sexual offence trials though it is argued that the impact of such questioning is qualitatively more traumatic in these cases than in those involving other kinds of offences. Moreover, Burman (2009: 394) reports that, despite legislative changes made in Scotland intended to improve women's experiences of the witness box:

> The 2002 Act has clearly not gone in the direction intended; it has resulted in neither a decrease in the use of sexual history and character evidence overall, nor a tighter focus on the type of questioning that is allowed. Rather, it has resulted in the introduction of much more sexual history and character evidence than under the 1995 Act. The proportion of trials with applications to introduce prohibited evidence have increased markedly, questioning on sexual history and character is now sought by both the prosecution and the defence, the numbers of multiple applications have doubled and the "belt and braces" approach adopted by the defence means that the questioning or evidence sought in written applications is now far more detailed and extensive than that sought in verbal applications made during the trial under the 1995 Act.

Whilst less well documented, the finding reviewed above can also be translated to men's experiences of the criminal justice process as victims of sexual assault. Rumney (2008: 4) reports that "While some men do report a positive response, some police officers and other criminal justice professionals appear to attach to gay men or those they perceive as gay highly questionable assumptions regarding credibility, trauma and truthfulness." Of course, assumptions in respect of homosexuality, and indeed, experiences of homophobia have the potential to permeate men's experiences of reporting sexual assault despite the empirical data pointing to the fact that the majority of sexual assaults are committed by heterosexual men on other men both within and outwith the context of prison. As Lees (1997: 106) has commented:

> One of the most damaging insults to be thrown at a man is to call him a woman, a bitch or a cunt. The act of coercive buggery can be seen as a means of taking away manhood, of emasculating other men and thereby enhancing one's own power.

Experiences at court, then, can not only challenge a woman's testimony based on her qualities of womanhood, they can also challenge a man's based on his qualities of manhood. In each case these qualities are measured against normal forms of femininity and normal forms of masculinity. These are forms that are permeated with assumptions around what constitutes normal heterosexuality. Thus, in facing questioning at court, the issue of penetration, various legislative interventions in different jurisdictions notwithstanding, still remains crucial to a rape trial: who did what to whom with what implement. The focus on the penis characterizes that which Smart (1989) has called the celebration of phallocentrism. A celebration that tells us much that is implicit in the law with respect to gender and why the rape trial proceeds in the way that it does.

The celebration of the penis privileges a male centered view of both female and male sexuality. Women do not know what they want; they have to be persuaded. So a woman whose dress or behavior challenges this assumption is problematic. Men cannot be blamed for acting on deep-seated sexual needs: in the face of temptation these are uncontrollable. As a result the few who are convicted must be the psychologically deranged "other men" in whose name, men as a group, dare not speak. Moreover it is this manhood that proves equally problematic for the male complainant of rape. As Allen's (2002) typology of men's reactions to having been the victim of rape, clearly indicates the necessity of understanding these through the lens of what it is to be a man. Thus presumptions surrounding appropriate manhood (masculinity) render the male complainant of rape hugely problematic for the criminal justice system in general and the criminal trial in particular, given the implicit celebration of (heterosexual) phallocentrism commented on above. Naffine (1997: 119) suggests:

> It is not only women who suffer from the culture of the strong man, though it is women who are consistently dispossessed by it..... There are a range of other masculinities which are simultaneously implied and then cast out. They are by implication, rendered unnatural and undesirable. Thus are the men who depart from the masculine ideal rendered silent.

Thus sexual assault and rape are framed by deep-seated understandings of who the law is for and what can be achieved in the recourse to law. This frame inexorably regresses to very conservative understandings about who can do what to whom with what kind of implement. Such conservatism inherently buys into ideal typical notions of victimhood that are differently informed for women and for men, but in each case can result in a silencing of their respective voices. The question remains: is this resistance to gendered understandings of sexual violence, implementation failure, as many policy and practitioners would claim, or reflection of a deeper problem?

Current Issues and Controversies: Implementation Failure or Conceptual Failure?

The foregoing discussion offers an overview of the extent to which addressing sexual violence, rape in particular, has been responded to as a social problem. This overview clearly implies that whilst there has been a good deal of policy activity directed towards addressing sexual violence against women, there has been less attention paid to the problem of sexual violence against men. Moreover, despite the desire to improve policy and professional responses to women as complainants of sexual violence, whilst much is promised, much also remains to be delivered, leaving responses to men as complainants arguably further in the shade. This overview also implies a deeply held believe that the criminal justice systems could "do better"; that is, could implement agreed policies much more effectively. A good deal of the debate around how a better response might be achieved has focused on the question of attrition and, what attrition actually means, which is neither simple nor straightforward. Yet it is also important to note that criminal justice system policies and practices does not act within a vacuum. As the foregoing discussion has also intimated, there are inherently conservative images embedded in the law around sexuality and difference and there are similarly conservative views articulated in public opinion around the factors that contribute to sexual violence. Moreover such views are held about men and women as victims and perpetrators (Anderson 2007). In the face of such conservative forces, implementation failure seems not only a likely outcome but also, arguably, an unfair charge.

Mooney (2007: 269) comments that the values whereby men's violence to women is sustained in the face of public imperatives otherwise "exist throughout the width and breadth of popular culture." For example, consider the vicarious pleasure gained by some young males in witnessing violence on a "good night out" (Winlow and Hall 2006). Thus violence becomes part of everyday life; as is the case of a woman living with violence who judges that the risk of poverty and homelessness are worse than the violence she knows she will be subjected to. This is the ordinary violence of everyday life. In fading out the voices of those who know this, as Jordan (2011) suggests, the ordinariness of violence is rendered absent. So too is the ordinariness of the (sexual) violence that men commit against each other both within and outwith the prison. So to answer, Mooney (2007) it is with the ordinary nature violence in which it can be both public anathema and a private common place all at the same time. It is within these tensions that the problems of the "implementation gap" are to be found.

Conclusions and Future Research

In a recent commentary on her own conceptual intervention on sexual violence Kelly (2011: xxiii) suggests that "neither recognition nor redistribution provide an adequate framework for understanding the persistence of gender inequality, VAW [violence against women] and the intersections involved" and that the apparent policy failures that have ensued from over 20 years of campaigning, "need new theoretical framings, which place the continuum of violence at the core of gender inequality." In other words, what we have witnessed, and what has been articulated in this chapter, is not implementation failure, but conceptual failure. This does not necessarily mean that current policies should be dropped but does mean that greater recognition that gender refers to men and women is required. By implication, more work is needed in considering sexual violence from a gendered perspective. Work that does not assume what might be known about women's experiences also translates to men's, but considers how a gendered understanding might transform their respective experiences of the criminal justice system as both women and men as victims, offenders and professionals working within the criminal justice system.

Related Entries

- ▶ Domestic Violence
- ▶ Feminist Criminological Theory
- ▶ Feminist Theory in the Context of Sexual Violence
- ▶ Gendered Theory and Gendered Practice
- ▶ Gendering Traditional Theories of Crime
- ▶ Police Family Violence Services
- ▶ Surveys on Violence Against Women
- ▶ Women of Color and Crime

Recommended Reading and References

Allen S (2002) Male victims of rape: responses to a perceived threat to masculinity. In: Hoyle C, Young R (eds) New visions of crime victims. Hart Publishing, Oxford

Amir M (1971) Patterns of forcible rape. University of Chicago Press, Chicago

Anderson I (2007) What is a typical rape? Effects of victim and participant gender in female and male rape perception. Br J Soc Psychol 44:245–265

Brown J, Walklate S (2011) Introduction. In: Brown J, Walklate S (eds) Handbook on sexual violence. Routledge, London

Brown J, Horvath M, Kelly L, Westmarland N (2010a) Has anything changed? Results of a comparative study (1977–2010) on opinions towards rape. Government Equalities Office, London

Brown J, Horvath M, Kelly L, Westmarland N (2010b) Connections and disconnections: assessing evidence, knowledge and practice in responses to rape. Government Equalities Office, London

Burman M (2009) Evidencing sexual assault. Probation J 56(4):379–398

Christie N (1986) The ideal victim. In: Fattah EA (ed) From crime policy to victim policy. Macmillan, London

Coxell A, King M, Mezey G, Gordon D (1999) Lifetime, prevalence, characteristics and associated problems of non-consensual sex in men: cross sectional survey. Br Med J 318(7187):846–850

Daly K, Bouhours B (2009) Rape and attrition in the legal process: a comparative analysis of five countries. Forthcoming in crime and justice: an annual review of research 39. Downloaded 26 Oct 2010

Davies M, Pollard P, Archer J (2006) Effects of perpetrator gender and victim sexuality on blame toward male victims of sexual assault. J Soc Psychol 146(3): 275–291, British

Fraser N (2009) Feminism, capitalism and the cunning of history. New Left Rev 56:97–117

Goodey J (1997) Boys don't cry: masculinities, fear of crime and fearlessness. Br J Criminol 37(3):401–418

Graham R (2006) Male rape and the careful construction of the male victim. Soc Leg Stud 15(2):187–208

Hindelang MJ, Gottfredson MR, Garofalo J (1978) Victims of crime: an empirical foundation for a theory of personal victimisation. Ballinger, Cambridge, MA

HMCPSI (2007) Without consent. HMCPSI, London

Jordan J (2008) Serial survivors: women's narratives of surviving rape. The Federation Press, Sydney
Jordan J (2011) Silencing rape, silencing women. In: Brown J, Walklate S (eds) Handbook on sexual violence. Routledge, London
Kelly L (1988) Surviving sexual violence. Polity, Oxford
Kelly L (2011) Standing the test of time? Reflections on the concept of the continuum of sexual violence. In: Brown J, Walklate S (eds) Handbook on sexual violence. Routledge, London
Lees S (1997) Ruling passions. Open University Press, Buckingham
Lees S (1999) Carnal knowledge. The Women's Press, London
Mooney J (2007) Shadow values, shadow figures: real violence. Crit Criminol 15:159–170
Naffine N (1997) Feminism and criminology. Temple University Press, Philadelphia, Polity
Povey D, Coleman K, Kaiza P, Roe S (2009) Homicides, firearm offences and intimate violence. supplementary volume 2 to crime in England and Wales 2007/08. Home Office Statistical Bulletin. London: Home Office
Rumney P (2008) Policing male rape and sexual assault –. J Crim Law 72(1):67–78
Russell D (1990) Rape in marriage. Collier, New York
Smart C (1989) Feminism and the power of law. Routledge, London
Stern V (2010) The stern review: Government Equalities Office, London
Weiss KG (2010) Too ashamed to report: deconstructing the shame of sexual victimization. Feminist Criminol 5(3):286–310
Winlow S, Hall Winlow S, Hall S (2006) Violent night: urban leisure and contemporary culture. Berg, London

Victimology

▶ International Responses to Victims in Criminal Justice

Victims and Restorative Justice

Ross London
Berkeley College, New York, NY, USA

Synonyms

Restorative justice, victims

Overview

A major theme expressed in many of the chapters of this volume is the dissatisfaction that crime victims have long felt concerning the policies and practices of the traditional criminal justice system. After centuries of neglectful and, all too often, abusive treatment in criminal courts, crime victims in countries throughout the world within recent decades have demanded fundamental reforms to address their needs and concerns. What many victims' advocates envisioned was a criminal justice system not simply fixated on the goals of assessing blame and punishing offenders, but one which offered crime victims the opportunity to meaningfully participate, to have their voices heard, and to repair the emotional and material harm they have suffered.

With the emergence of the restorative justice movement in the 1980s, it appeared that crime victims had gained an important ally in their quest for reform. Restorative justice, after all, was seeking many of the reforms crime victims had sought: greater involvement in the criminal justice process, compensation for losses and a path to healing the emotional wounds of crime. One could scarcely have imagined a more natural alliance of interests, and yet, despite its early promise, the relationship between the victim's advocacy movement and restorative justice has been and continues to be tenuous at best. (Amstutz 2004, p. 85). In the most extensive survey of victims' advocacy groups conducted to date, widespread dissatisfaction was uncovered concerning restorative justice programs in the United States, including concerns about injustice, disrespect, exclusion, lack of empathy, and irrelevance of the programs (Mika et al. 2002). A number of respondents believed that restorative justice was created primarily for the benefit of offenders in search of reducing punishment and avoiding responsibility and that the needs of victims were merely an afterthought. Some respondents reportedly felt they were manipulated into participating by restorative justice "in order to promote and rationalize its agenda" and argued that the characterization of crime as mere interpersonal conflict "trivializes the nature of deep harms and the character of their relationship to the offender" (2004, p. 34). Despite the promises of healing, "little victim relief" was actually attained, and crime victims asked: "What is the compelling proof that it works?" Without proof of the effectiveness of its programs, the report maintained, "there will continue to be resistance to their blanket implementation and reluctance in the victim community to embrace them" (2004, pp. 33–34).

What went wrong? Can this relationship be mended? What are the prospects for future relations as both the victims' movement and the restorative justice movement continue to evolve? In this chapter, these questions will be explored, beginning with a brief historical overview of both the victims' movement and the restorative justice movement, followed by a critical assessment of the benefits to victims claimed by restorative justice advocates, and an analysis of several key policy issues that have impeded the growth and public acceptance of restorative justice. In the final section, we shall consider how a more expansive conception of restorative justice that has emerged in recent years may make restorative justice available to a wider variety of crime victims and become of greater importance in their advocacy for increased public support.

Fundamentals

The Restorative Justice Movement

Restorative justice arose in the 1980s as a radical alternative to a conventional criminal justice system that many had regarded as dysfunctional, morally deficient, and neglectful of the legitimate needs of victims, offenders and communities. Not only was the criminal justice system seen by many critics as ineffectual, unable to stem the epidemic of drug use, and, through its "get tough" tactics, the cause of further alienation and rebellion in the inner cities, the system was increasingly deemed unresponsive to the needs of crime victims and their communities. Social critics from the Left renewed their attacks on the criminal justice system as a means of social control exercised by a repressive state, and proposed decentralized, informal alternative means of resolving conflicts. Prisoners' advocacy organizations that stressed the need to reintegrate, rather than permanently exclude, offenders obtained support from Christian Mennonites who advocated healing, repair and reconciliation in place of retributory justice. The 1980s also saw a rise in interest in community-based problem solving and indigenous justice. During this period, the "just deserts" theory, initially embraced by legislators as a principled solution to the problem of sentencing disparities that resulted from unbridled discretion, came under strenuous attack for being rigid, obsessed with punishment, retrospective, and more concerned with uniformity rather than conflict resolution. (Bazemore and Walgrave 1999, p. 46).

Out of this amalgam of critiques of the existing criminal justice system grew an interest in a new conception of justice based on healing injuries rather than the assignment of blame and punishment for legal transgressions. These "restorative justice" theorists sought an informal, non-punitive alternative to a system in which many victims felt they were treated as little more than a "piece of evidence" and who were "twice victimized, first by the offender and then by an uncaring criminal justice system that doesn't have time for them" (Umbreit et al. 1994, p. 196.)

Because restorative justice arose from a diversity of sources, there has never been an "official version", or even a commonly agreed-upon definition. What has been achieved has been not so much a set of principles as a set of shared values:

- Viewing crime as a source of harm, not simply as a transgression of law, and, consequently, specifying the mission of the criminal justice system as the repair of harm instead of only the determination of guilt and imposition of punishment;
- Providing compensation for the victim and achieving social reintegration of the offender; and

- Involving the victim, the offender, and relevant community members to the fullest extent possible in voluntary dialogue in order to devise restorative solutions.

The Victims Movement

Beginning in obscurity in the 1960s as a scattering of neglected and abused "outsiders" to the criminal justice system, victims' advocacy groups, within a few decades, became a powerful influence on criminal justice policy and practice throughout the world.

Their demands that the voices of crime victims be heard in court, that the full extent of their suffering be acknowledged, and that their expectations of appropriate justice be considered by the courts in sentencing fell upon the receptive ears of law enforcement officials and politicians in the 1970s who seized upon the incipient victims' rights movement as a means to secure the cooperation and political support of crime victims. Crime victims' advocacy achieved a number of significant criminal justice reforms including the establishment of victims' assistance units within prosecution offices that provided court information, protective services, and assistance to victims as well as procedural changes that permitted victims to be heard in court at sentencing and parole hearings. Along with these and other procedural reforms, support for victims' welfare reform also developed. Overturning centuries of legal impediments to obtaining fair compensation for their material losses within the traditional criminal justice system, victims' advocates achieved the remarkable goal of establishing court ordered restitution as a standard component of criminal prosecutions in court systems throughout the world. As the full extent of victims' losses was brought to light both in court and through the news media, it soon became apparent that the prospect of payment for these losses solely through the financial resources of criminal defendants was woefully inadequate. Therefore, advocates of victims' welfare looked to the state for compensation, whether or not the offender could pay restitution, and indeed, whether or not the offender was even brought to justice. Many of these programs, therefore, operated independently of the criminal justice system altogether. Despite these gains, however, there were persistent concerns from victims' advocates that state compensation programs were either flawed or insufficient, and riddled with exceptions, exclusions and limitations. (Dignan 2005, pp. 54–55).

Benefits to Victims Offered by Restorative Justice

While restorative justice advocacy has not been significantly involved in assisting victims groups to obtain increased government support and compensation, it has offered an intriguing new way to fulfill victims' emotional and psychological needs in 100s of innovative programs implemented throughout the world during the last 3 decades.

Restorative justice, indeed, may very well be the first major criminal justice reform that has focused on repairing both the material and psychological harm suffered by crime victims. Besides their physical need for safety and the economic need for restitution of their material losses, crime victims endure emotional and psychological losses as a consequence of the crime. Emotional reactions reported by victims include anger, anxiety, depression, overall physical distress, resentment, and hostility. Psychological harms to victims include a sense of isolation, loss of faith, shock, enmity, loss of control, self-blame, denigration, and fear (Lurigio and Resick 1990, pp. 55–57). Restorative justice advocates have primarily stressed three vehicles for emotional and psychological recovery: apology, receipt of restitution and victim participation.

The central mechanism for emotional and psychological healing cited by virtually all restorative theorists is apology (Strang 2002, pp. 18–23). The demonstration of respect inherent in apology is thought to counteract the message of disrespect that often underlies the feeling of injustice, thus lessening the victim's demands for punishment. Because crime devalues the self-worth and denigrates the status of the victim, proper balance is restored by (1) the offering of an apology by the offender and (2) the communication of condemnation and reproach by society. Through expressive acts, rather than through physical punishment, it is believed that the victim is vindicated and the psychological harm of the crime is repaired.

Additionally, the requirement of restitution is not only a means of material compensation, but also a means of accountability–a symbolic gesture of penance and a corroboration of the sincerity of apology. “Restitution,” it has been contended, “serves both the emotional and practical needs of the victim and the idea of social justice–as opposed to punishment, which instead satisfies the rather archaic requirements of retribution” (Sessar 1999, pp. 287–288). Finally, restorative justice theorists have contended that the disempowerment, loss of control, and sense of isolation experienced by many victims can be ameliorated by greater involvement and more respectful treatment in the criminal justice system itself.

Empirical Testing of Restorative Justice Claims

Overall, the results of victim surveys from dozens of programs throughout the world have been very positive, indicating a high degree of satisfaction on the part of both victims and offenders (Umbreit et al. 2003, pp. 25–28). Victims involved in restorative justice sessions have reported higher levels of satisfaction than their counterparts in the traditional criminal justice system concerning their experience of fairness, receipt of restitution and reduction in fear and anxiety. In his meta-analysis of seven studies screened for statistical comparison, Poulson found significant positive results from victims including their satisfaction with the handling of the case, their opportunity to tell his or her story, their belief that the offender was held accountable, whether apology was offered, and whether the victim was still upset about the crime (Poulson 2003). These studies also indicate that the material and emotional benefits to victims were achieved via a non-punitive restorative alternative while decreasing, or, at least without increasing, the recidivism rate of the offenders who participated.

Not all findings are uniformly positive, however. In her survey of Australian programs, Kathleen Daly found that, contrary to the goal of eliciting empathy and remorse from offenders, most offenders were fixated on negotiating a better sentence and reported that the victim’s story had “little or no effect on them” (Daly 2002, p. 70). Correspondingly, only 27 % of the victims who attended these sessions regarded the offender’s apologies as sincere. Indeed, Maxwell and Morris’ 1997 study found that 25 % of victims who participated in a restorative justice family conference program actually felt worse after attending due to, among other reasons, the perceived insincerity of the offender’s apologies. Even in the limited context of victim satisfaction, researchers have found that victims were the least satisfied of the participants at restorative justice conferences, and that their negative feelings were linked to dissatisfaction with the leniency of the outcome (Maxwell and Morris 1994, p. 119).

Compounding the problem of assessing the success of restorative justice programs are methodological issues including low response rates to questionnaires, problems in identifying and standardizing key variables, the inability of research designs to randomly assign sentencing variables, and difficulties in measuring victims’ emotional and psychological reactions. Studies, for example, that have focused on victim satisfaction have not assessed the victims’ emotional and psychological recovery, and even when attempted, the goal of measuring satisfaction is singularly elusive: “It is virtually impossible to disentangle the multitude of variables which together determine how satisfied people are with any social service” (Johnstone 2011, p. 20).

Moreover, the limitation of restorative justice programs to those victims willing to personally meet with their offenders restricts the scope of the conclusions that can be drawn from the victim satisfaction studies conducted to date. In contrast to the claim that victim participation is, in itself, a source of personal empowerment and a means of repairing the emotional harm of crime, the fact remains that most victims, after being offered the opportunity to engage in dialogue with offenders, have failed or refused to do so. The largely positive results of these surveys, therefore, were very likely obtained from a small, self-selected group of victims who were amenable to the process to begin with and saw a positive benefit in face-to-face encounters with offenders.

Notwithstanding these qualifications and limitations, there is little reason to doubt the largely positive effect of restorative justice practices on the emotional well-being of those crime victims *who choose to participate in restorative justice programs*. What remains highly controversial, however, is the claim that the benefits to the material, psychological and emotional well-being of crime victims achieved though restorative justice programs result from its informal, non-punitive approach to criminal justice in contrast to conventional prosecution of crimes. Unfortunately, empirical researching into this critically important issue has been impeded by an underlying selection bias toward relatively minor offenses that are eligible for diversion away from formal prosecution. These "divertible" cases are necessarily of a different character than those that law enforcement officials refuse to divert. The criteria for diversion typically includes the stipulation that diversion away from prosecution will not jeopardize the safety of the victim or the larger community, nor will it offend community norms of justice. Even if it were possible to randomly assign cases to either restorative justice or conventional prosecution, the most one could conclude would be that, for *cases eligible for diversion*, restorative justice conferencing achieves better outcomes for the victim, the offender and society.

But no such claim could be made about the other, more serious cases, that are not eligible for diversion. As to these serious cases – those that are of greatest concern to victims and to society-the many claimed advantages of restorative justice lack firm empirical support, especially the highly controversial claims regarding the use of punishment and the role of the state in administering criminal justice.

Key Controversies

The Relevance of Punishment to the Emotional Well-Being of Victims

The fact that victims respond positively to apology, greater involvement when desired and receipt of restitution is neither surprising nor controversial. On the other hand, the contention that healing the emotional wounds of crime victims can be better achieved without punishing the person responsible for their suffering is undoubtedly the most controversial claim advanced by restorative justice advocates, and vividly contrasts with the retributory approach to justice held by many, if not most, members of the public, especially for serious crime.

The non-punitive approach of restorative justice has been primarily advanced not on moralistic or religious grounds, but rather on the strength of the claim that it offers victims a better path to healing their emotional wounds than the path offered by retributive justice. The prospect of healing is put forward not merely as an incidental benefit to victims. It is, rather, the primary inducement to their involvement in restorative justice programs and is inextricably linked with the concept of restoration itself. "Restorative justice," it is claimed "is about healing (restoration) rather than hurting. Responding to the hurt of crime with the hurt of punishment is rejected..." (Braithwaite and Strang 2001, p. 1.) If crime is about harm suffered by victims, and if the goal is to repair that harm, merely punishing offenders, restorative justice advocates have claimed, "does not help the injuries caused by crime and simply creates new injury–now both the victim and the offender are injured" (Van Ness and Strong 1997, p. 38). They have asked "How can still more suffering undo the pain of crime?" (Kaptein 2004, p. 83) and have claimed that the "personal needs of those offended are not met by revenge, but by addressing the feelings behind the [victims'] anger" (McCold 2000, p. 361). What all these observations have in common is the claim that, whatever benefits punishment may have in maintaining social order, it is not necessary for the restoration of a victim's well-being and, in fact, may be counterproductive.

Unfortunately, these claims are far from proven by research conducted to date. Despite 2 decades of implementation in programs throughout the world and scores of studies on various aspects of restorative justice programs, no study has shown that the emotional recovery of crime victims can be improved by removing the option of punishment (London 2011, p. 99). Perhaps the best evidence to date for a positive effect of a non-punitive restorative justice approach upon the emotional recovery of crime victims can be found in Strang's study of the Reintegrative Shaming Experiments (RISE) Project in Canberra, Australia (Strang 2002). This study had the methodological advantage of random assignment of cases between conventional court processing and restorative justice conferencing, but the limitation of the research to cases deemed eligible for diversion restricts the scope of its conclusions. Punishment of the offender may indeed have been irrelevant to the emotional well-being of these victims but, then again, in these cases, the need to punish the offender was judged by law-enforcement officials to be so minimal as to warrant diversion from prosecution. Additionally, unless similar procedures are compared- with and without the option of punishment- one cannot discern whether any resulting emotional benefits to victims arose solely from their opportunity for greater involvement, their better treatment by criminal justice staff and, perhaps most saliently, their receipt of an apology, rather than the presence or absence of punishment as an available sanction.

In reviewing the literature, one researcher concluded that most theories concerning the therapeutic effects of punishment on the emotional losses of victims are based on anecdotal evidence and still wait valid empirical testing (Shuman and Smith 2000, p. 103). Indeed, as Parsons and Bergin noted in their 2010 review of the psychological outcomes of restorative justice programs: "there has been no research to date contrasting the mental health impact of restorative justice with that of conventional trials" (Parson and Bergen 2010, p. 186). Remarkably, therefore, the central claim of restorative justice concerning its power to heal the emotional wounds suffered by victims without punishing their victimizers-undoubtedly its most controversial claim as well as its primary attraction to victims- remains unproven by empirical research.

The eligibility of a case to diversion is important to consider when assessing the victims' safety concerns as well. Although restorative justice advocates have cited the greater ability of restorative justice programs to satisfy the victims concerns about his or her safety following the proceedings, it is doubtful whether victims of serious crime would be similarly assured of their safety if their victimizers faced no prospect of punishment for re-offending. Many victims, indeed, would not want "to trade off the protection of the current system for an opportunity to meet with the offender" (Achilles, p. 70). Similarly, even if recidivism rates are unaffected- or lowered- by diverting petty offenses away from prosecution toward non-punitive alternatives, it is unlikely that re-offending by violent criminals would be unaffected or diminished by removing the threat of incarceration. On the contrary, Levrant and her colleagues concluded, "a fundamental weakness of restorative justice is its failure to provide a plausible blueprint for how to control crime" (Levrant et al. p. 22).

Restorative Justice and State Prosecution

A key feature of restorative justice advocacy has been its criticism of the state's control over criminal proceedings. "There is no denying" one author claimed, "the fundamental incompatibility between the state system of doing justice and the principles of restorative justice" (Boyes-Watson 2004, p. 215). By using the state as the offended party in criminal cases, it has been argued, the needs of the actual victims are virtually ignored. Restorative justice has been proposed as an alternative that stands apart from any state-sponsored system of criminal justice, returning the conflict to its rightful owners. Yet justice freed from control by state authorities has thus far failed to engage the support of many victims' advocate for a number of reasons.

First, the rejection of state prosecution in favor of personal conflict resolution would make unavailable to participants the power of the state to compel compliance with the law adequate to ensure their personal safety, and would make unavailable the resources of the state necessary to compensate victims for their losses when the personal resources of the offender are insufficient. Second, personalized justice in the form of victim-offender conferencing is not subject to any standards of procedural fairness, equality of treatment or safeguards against biased or arbitrary decision-making. Unlike criminal justice professionals, participants at these sessions are untrained in the law, are subject to no standards of performance, and are accountable to no reviewing authority. Therefore, in contrast to the advantages claimed for "deprofessionalized justice," there are legitimate concerns that restorative justice conferencing can deliver a higher quality of justice than that available under the conventional system (Ashworth 1993).

Third, even with respect to the purely private concerns of the victim, victim – offender conferencing, offered as a substitute for state prosecutions, may be an inherently inadequate means for victims to satisfy their demand for public vindication. Rather than accusing the offender of a criminal offense against the victim, the goal of the restorative justice mediator is "the creation of a safe or even sacred space to foster dialogue", and for this to happen, the mediator must have an "attitude of unconditional positive regard and connectedness with all parties, while remaining impartial (e.g. not taking sides)" (Umbreit et al. 2003, p. 17). There is no dominant "narrative" imposed on the event as in state prosecutions of a crime, but rather an exchange of viewpoints. The admonition to mutuality thus may burden the victim with the expectation of empathizing with the offender, acknowledging the validity of the offender's needs as well as her own, and providing support for the offender's rehabilitation (Achilles 2004, pp. 69–70). Finally, in those cases in which the crime was made possible by a power imbalance between the parties, the same power in imbalance that gave rise to the crime are likely to persist during the conference. The danger is that the very vulnerability of the victim that was exploited in the crime will also be exploited at the conference, thereby undermining the validity of the victims' moral assessment of the crime and the appropriateness of the sanction he or she seeks. The victim, lacking the support of the state as the complaining party, is now expected to stand up for herself and demand justice.

To sum up, the claims advanced by restorative justice advocates concerning its unique power to heal the emotional wounds of crime and its even more controversial claims rejecting retributory justice in favor of non-punitive personalized dialogue between victim and offender are best substantiated in minor matters that may be safely diverted away from prosecution without impairing the public interest in safety and justice. This limitation to diversionary programs for petty crimes by youthful offenders falls very short of the sweeping claims that restorative justice represents an entirely new paradigm that transforms the way we administer criminal justice. In fact, if restorative justice continues to function merely as a diversionary strategy, it would not be significantly different from other types of diversionary strategies employed by the conventional criminal justice system, including face-to-face conflict resolution, counseling and non-punitive means of holding offenders accountable, and therefore will continue to be of limited assistance to crime victims.

Future Directions

Despite its present isolation as an exotic alternative to conventional criminal justice practice, a more expansive view of restorative justice has emerged in recent years that would make restorative justice applicable to a greater diversity of victims and a wider variety of cases- from the least to the most serious.

In contrast to the early proponents of restorative justice who conceived of it as a radically different alternative to criminal justice, a growing number of restorative justice scholars and advocates have proposed a "maximalist" view of restorative justice which, uncoupled from its former anti-penal and anti-statist positions, "does not require abandoning deterrence, retribution or the adversarial process" (Bibas and Bierschbach 2004, p. 112). Rather than creating an either/or choice between restoration and other essential objectives, the maximalist version of restorative justice would attempt to achieve restorative goals for crime victims without forfeiting the advantages offered by state resources and mechanisms of enforcement. Bazemore and Walgrave's definition of restorative justice stresses its unique goal rather than its opposition to conventional practices: "restorative justice is every action that is primarily oriented towards doing justice by repairing the harm that is caused by crime" (1999, p. 48).

Despite a diversity of maximalist positions, what is shared is an interest in paring down restorative justice to its essential mission - *repairing the harm of crime*–that is capable of application in diverse settings. These applications might include:

- Diversionary programs
  A maximalist version of restorative justice would continue to offer a diversionary option for appropriate cases, but unlike the purist model, using this diversion strategy would not entail rejecting state resources or coercive power. On the contrary, agreements reached by the victim and the offender would be fully enforceable by the state.
- Applications with victim involvement
  For more serious cases that are not eligible for diversion, opportunities may be presented to victims who wish to meet the offender and participate in fashioning a restorative solution, including sentencing recommendations within a range of penalties prescribed by law. The involvement of victims in sentencing recommendations would not be an alternative to judicial sentencing authority, but rather would be an alternative to sentencing recommendations by prosecutors and defense attorneys in the context of plea bargains.
- Applications without victim involvement
  Because the maximalist model of restorative justice is based on the attainment of a new criminal justice goal rather than any particular procedure, it may be applied as a guide to sentencing in a multitude of different settings, including cases in which the parties fail to agree to participate in restorative justice conferencing and in "victimless crimes" and crimes affecting the population in aggregate such as tax evasion and environmental offenses, in which victim-offender encounters are clearly inapplicable.

The advantages of an expanded view of restorative justice which utilizes rather than rejects the coercive power of the state may be seen in many current post – conviction restorative justice programs that take place within prisons. In these programs, victims who wish to meet the offender to express their feelings, obtain answers to their questions and be offered an apology need not renounce their demand and for justice as a precondition for participation since the offender has already been convicted and sentenced. In one such program analyzed by Susan Miller, "the victims . . . felt strongly that the offender needed to be held accountable by the criminal justice system. They did not wish to sidestep that aspect of justice and punishment but rather, the victims wanted the opportunity to meet with the offenders after the conclusion of the criminal justice process . . ." (Miller 2011, p. 12). Miller concluded that this post-conviction format for restorative justice offers victims the opportunity for healing and empowerment beyond those available either through conventional criminal justice or diversionary programs: "Since the process begins after guilt is established and punishment impose, victims. . . perhaps for the first time since the abuse and violence began – felt empowered by the process" (2011, p. 198).

Example of Application: Domestic Violence Cases

An example of how a more expansive approach to restorative justice can serve the needs of a wider variety of victims may be seen in domestic violence cases. Restorative justice, since its inception, has had an uneasy relationship with advocates of victims of domestic violence. The approach of peacemaking, reconciliation, and conflict resolution that is offered as an alternative to prosecution has been seen by feminist critics as a regression to the time when domestic violence was considered a private matter between partners, best dealt with by counseling and cooling-off periods. The idea of reprivatizing domestic violence by the use of restorative justice as an alternative to the state's punitive powers has been viewed as a form of "cheap justice" (Coker 1999, p. 85) that, by diverting these cases from prosecution de-criminalizes violence and devalues the seriousness of the offense. Other feminist advocates, however, have cautioned that insisting upon arrest and prosecution is no panacea for victims of domestic violence, given the low participation rates of victims, low conviction rates of offenders and poor treatment of victims often associated with conventional prosecution. The use of restorative justice in cases of domestic violence, it is therefore argued, is at least an alternative to an adversarial system that still isn't working for victims of intimate violence (Curtis-Fawley and Daly 2005).

But must there be a choice between the "hard option" of criminal prosecution and the "soft option" of restorative justice diversion? A number of feminist advocates have recognized that this dualistic choice is unnecessarily self-limiting. As Barbara Hudson recognized nearly a decade ago, it is not restorative justice conferencing per se, but it is rather the diversionary nature of privatized, non-punitive restorative justice conferencing that renders it inappropriate for cases of gendered violence (Hudson 2002). Curtis-Fawley and Daly have similarly argued that restorative justice programs should include, but should not be limited to nonpunitive diversions. Instead of setting itself apart as a radical alternative to prosecution, they propose that restorative justice should be regarded as a flexible concept, capable of being integrated into the existing criminal justice system (2005, pp. 632–633).

The kind of integration of criminal justice and restorative justice envisioned by these feminist advocates is precisely the goal of a maximal version of restorative justice. Rather than confining itself to diversions from state persecution, the coercive power of the state would be available to enforce agreements designed to safeguard victims' safety and to mandate compliance with the law. Instead of an either/or choice between prosecution or diversion, a maximalist version of restorative justice would accommodate a number of procedural options, both formal and informal, that are deemed appropriate to the individual case, backed up by the coercive power of the state to compel compliance.

Extending Aid to Victim Beyond Criminal Justice Reforms

Finally, an expansive view of restorative justice opens opportunities for engagement in advocacy on behalf of victims for increased public support. Victims' advocates have long noted that victims' needs exceed the capabilities of the conventional criminal justice system. Given the extent of victims' losses including medical, psychological and rehabilitation expenses, property damage and loss of income, few offenders can be expected to fully compensate their victims from their own resources. If the conventional system cannot adequately address the needs of victims, victim advocate Susan Herman argues, restorative justice, despite its "commendable effort to humanize the justice system" (Herman 2004, p. 78) cannot expand upon these capabilities if it is dependent on the personal resources of those who participate in restorative justice programs. Clearly, the "purist" version of restorative justice offers no rationale for increased public spending since, on the basis of its foundational claim that crime is a personal, not a public harm, it limits redress to measures undertaken by the offender individually to repair the damage he or she has caused.

On the other hand, a maximalist version of restorative justice may very well offer a constructive role in such advocacy because it is unencumbered by an anti-state ideology which, in regarding crime as a personal harm, ignores the public dimension of crime. The simplicity of the maximalist approach gives it its flexibility to extend beyond the confines of criminal justice case administration to reach even those victims whose offenders are unknown or unprosecuted. According to this view, if a person is harmed by crime, the work of restorative justice is to repair that harm in any way possible, utilizing both the resources of the state and the resources of the individual offender. An expanded view of restorative justice, therefore, can play a positive role in advocacy for an expanded view of victims' rights.

Limitations

The evolution of restorative justice from a diversionary strategy to the mainstream of criminal justice practice, made possible by acknowledging the public dimension of crime, paradoxically contains within itself an important limitation; for once the public dimension of crime is acknowledged, it is no longer sufficient for restorative justice to be responsive only to the interests of the victim and the offender. Their interests, in fact, may very well conflict with the public interest in safety, justice and equality of treatment (Ashworth 1993).

This is a topic that cannot adequately be pursued within the confines of this chapter. Suffice it to say, however, that, even as restorative justice continues to evolve and continues to outgrow its original conception as a criminal justice diversion and thereby becomes more responsive to the needs of crime victims, those personal needs of victims must, in the end, remain subordinate to the needs of the public. The challenge to future legislation will be to permit victims the widest latitude in pursuing their personal interests without thereby impairing the public good.

Conclusion

Restorative justice began as a vision of a better way to do criminal justice and, in 100s of programs throughout the world, it has proven to be just that. It has helped victims feel more satisfied with the process and more secure in their personal safety. It has increased offenders' compliance with restitution orders without adversely affecting recidivism rates. Yet it has not nearly reached its potential for transforming the conventional criminal justice system because it has failed to articulate a theory or set of policies applicable to the kinds of cases that most concern us: serious crime and adult offenders. Whether or not restorative justice can effectively deal with serious cases and adult offenders represents its primary challenge today because, unless it can do so, it shall remain on the sidelines of criminal justice practice, "doomed to irrelevance and marginality" (Dignan 2002, p. 179).

The growing trend of restorative justice toward a maximalist version of its mission, however, disassociated from its early insistence on an anti-penal and anti- statist ideology, holds the promise of restorative justice becoming a significant influence on criminal justice policy for a full range of cases- both divertible and non-divertible. Assuming this evolution from insularity to universality continues, the long-hoped-for alliance between restorative justice and the victims' movement may finally come to fruition.

Recommended Reading and References

Achilles Mary (2004) Can restorative justice live up to its promise to victims? in In: Zehr H, Toews B (eds) Critical issues in restorative justice, Criminal Justice Press, Monsey, NY, pp 65–73

Amstutz L (2004) What is the relationship between the victim service organizations restorative justice? In: Zehr H, Toews B (eds) Critical issues in restorative justice. Criminal Justice Press, Monsey, pp 85–94

Ashworth A (1993) Some doubts about restorative justice. Crim Law Forum 4(2):277–299

Bazemore G, Walgrave L (1999) Restorative juvenile justice: in search of fundamentals and an outline for systemic reform. In: Bazemore G, Walgrave L (eds) Restorative juvenile justice: repairing the harm of youth crime. Criminal Justice Press, Monsey, pp 45–74

Bibas S, Bierschbach R (2004) Integrating remorse and apology into criminal procedure. Yale Law J 114:85–148

Boyes-Watson C (2004) What are the implications of the growing state involvement in restorative justice? In: Zehr H, Toews B (eds) Critical issues in restorative justice. Criminal Justice Press, Monsey, pp 215–226

Braithwaite J, Strang H (2000) Connecting philosophy and practice in restorative justice. In: Strang H, Braithwaite J (eds) Restorative justice, from philosophy to practice. Dartmouth, Hants

Coker D (1999) Enhancing autonomy for battered women: lessons from Navaho peacemaking. UCLA Law Rev 47:1–111

Curtis-Fawley S, Daly K (2005) Gendered violence and restorative justice: the views of victim advocates. Violence Against Women 11(5):603–638

Daly K (2002) Restorative justice: the real story. Punishm Soc 4:55–79

Dignan J (2005) Understanding victims and restorative justice. Open University Press, Berkshire

Dignan J (2002) Restorative justice and the law: the case for an integrated, systematic approach. In: Walgrave L (ed) Restorative justice and the law. Willan Publishing, Cullompton, pp 168–190

Herman S (2004) Is restorative justice possible without a parallel system for victims? In: Zehr H, Toews B (eds) Critical issues in restorative justice. Criminal Justice Press, Monsey, pp 75–84

Hudson B (2002) Restorative jusrice and gendered violence: diversion or effecrive justice? Br J Criminol 42(3):616–634

Johnstone G (2011) Restorative justice: ideas, values debates, 2nd edn. Willan Publishing, Cullompton, Devon

Kaptein H (2004) Against the pain of punishment: retribution as reparation through penal servitude. In: Kaptein H, Malsch M (eds) Crime, victims and justice. Ashgate, Aldershot, pp 80–111

Levrant S, Cullen F, Fulton B, Wozniak J (1999) Reconsidering restorative justice: the corruption of benevolence revisited. Crime Delinq 54:3–27

London R (2011) Crime, punishment and restorative justice: from the margins to the mainstream. FirstForum Press, Boulder

Lurigio A, Resick P (1990) Healing the psychological wounds of criminal victimization. In: Lurigio AJ, Skogan W, Davis R (eds) Victims of crime: problems, policy and programs. Sage, Newberry, pp 50–68

Maxwell G, Morris A (1994) The new zealand model of family group conferences. In: Alder C, Wundersitz J (eds) Family conferencing and juvenile justice: the way forward or misplaced optimism? Australian Institute of Criminology, Canberra, pp 15–43

McCold P (2000) Toward a holistic vision of restorative juvenile justice: a reply to the maximalist model. Contemp Justice Rev 3:357–414

Mika H, Achilles M, Halbert E, Amstutz L, Zehr H (2004) Listening to victims: a critique of restorative justice policy and practice in the United States. Fed Probat 68:32–38

Miller S (2011) After the crime: the power of restorative justice dialogues between victims and violent offenders. New York University Press, New York

Parson J, Bergin T (2010) The impact of criminal justice involvement on victims' mental health. J Trauma Stress 23(2):182–188

Poulson Barton (2003) A third voice: a review of empirical research on the psychological outcomes of restorative justice. Utah Law Rev 1:167–203

Sessar K (1999) Punitive attitudes of the public: myth and reality. In: Bazemore G, Walgrave L (eds) Restorative juvenile justice: repairing the harm of youth crime. Criminal Justice Press, Monsey, pp 287–304

Shuman D, Smith A (2000) Justice and the prosecution of old crimes. American Psychological Association, Washington, DC

Strang H (2002) Repair or revenge: victims and restorative justice. Clarendon, Oxford

Umbreit M, Coates R, Kalanj B (1994) Victim meets offender: the impact of restorative justice and mediation. Criminal Justice Press, Monsey, pp 287–304

Umbreit M, Vos B, Coates R, Brown K (2003) Facing violence: the path of restorative justice. Criminal Justice Press, Monsey

Van Ness D, Strong K (1997) Restoring justice, 1st edn. Anderson, Cincinnati

Wright M, Galaway B (1989) Mediation and criminal justice: victims, offenders and community. Sage, Newbury Park

Zehr H (1990) Changing lenses: a new focus for crime and justice. Herald Press, Scottsdale

Victims and the ICC

► Victims and the International Criminal Court

Victims and the International Criminal Court

Ilaria Bottigliero
International Development Law Organization (IDLO), Rome, Italy

Synonyms

Victims and the ICC

Overview

This entry reviews the treatment of victims in the International Criminal Court, including the victims' participation and reparation regime; the establishment of a Trust Fund for Victims, the institution of an office of Public Counsel for Victims, state cooperation procedures, and relevant jurisprudence.

Fundamentals

The treatment of victims in the International Criminal Court (ICC) represents a substantial innovation in the field of victims' redress under international law. It marks a significant departure from traditional interstate approaches to individual claims for breach of an international legal obligation based on classical principles of state responsibility, towards a more comprehensive regime in which individual victims have a right under international law to seek and obtain reparations directly from individuals found responsible for genocide, war crimes, crimes against humanity, or the crime of aggression. Unlike the Nuremberg and Tokyo International Military Tribunals or the International Criminal Tribunals for the Former Yugoslavia (ICTY) and for Rwanda (ICTR), the ICC Statute allows victims ample opportunity to participate in the Court's proceedings and to receive some form of redress directly from the Court by way of judicial award. The ICC legal regime ensures that the right of victims to redress is no longer left exclusively in the hands of domestic authorities.

The ICC Statute's provisions on victims figure centrally within the ICC's general architecture, linking to other relevant provisions, such as those relating to state cooperation as enshrined in Part 9 of the Statute. This approach anchors the victim's rights to redress solidly in the ICC's norms and operation and improves the coherence and comprehensiveness of the victims' participation and reparation regime beyond that of the ICTY or ICTR.

Key Issues

The Rome Statute's victims' regime rests on five main pillars:

1. Participation in the proceedings
2. Establishment of reparation procedures
3. Creation of a Trust Fund
4. Institution of an Office of Public Counsel for Victims
5. State cooperation procedures

Participation in the Proceedings

The ICC allows victims some degree of direct participation in the Court's proceedings, in line with modern victimology theories which recognize the importance of victims' participation in the retributive process as a fundamental component of the healing process, especially where serious human rights violations are involved. In this sense, the ICC helps to crystallize one of the emerging, fundamental normative components of the victims' right to redress, namely, the right to be involved in the justice process, irrespective as to whether this process is to be administered by a judicial body, such as the ICC, or by quasi-judicial bodies such as truth and reconciliation commissions, enquiry commissions, human rights bodies, or similar entities.

Based on these premises, Article 68(3) on "Protection of the victims and witnesses and their participation in the proceedings" provides that:

> Where the personal interests of the victims are affected, the Court shall permit their views and concerns to be presented and considered at stages of the proceedings determined to be appropriate by the Court [. . .]. Such views and concerns may be presented by the legal representatives of the victims where the Court considers it appropriate [. . .].

This provision basically allows victims to be heard at certain stages of the proceedings, from the beginning of an investigation to the sentencing stage, thus conferring upon them a measure of locus standi. In particular, the ICC regime provides victims with the faculties discussed below.

Victims Can Provide Information to the Prosecutor for the Purpose of Initiating an Investigation

The ICC Statute does not provide victims with an independent power to institute proceedings. Unlike the European Court of Human Rights or the Inter-American Court of Human Rights, where individual victims can petition the Court either directly or, in the Inter-American system, through the Commission, the drafters of the Rome Statute decided against conferring individual victims full capacity to petition the Court. As such, victims have only "indirect" access to the Court through the action of the Prosecutor. At the very initial stages of the judicial process, victims are among those who can provide the Prosecutor with information which may trigger the initiation of a proprio motu investigation into a crime falling within the jurisdiction of the Court under Article 15. At this stage, the Prosecutor has an obligation to "analyze the seriousness of the information received," and, in this context, he or she may "seek additional information" and "receive written or oral testimony at the seat of the Court." Should the Prosecutor decide that the information received is not adequate to satisfy the "reasonable basis" requirement to proceed with an investigation, victims are allowed to submit further information on the same situation "in light of new facts or evidence" (Article 15(6)).

The limitation of not having an independent power to initiate proceedings can prevent victims from accessing the Court, should the Prosecutor fail to give proper consideration to the information provided by victims. In practice, the Prosecutor has ample discretion to decide whether or not to initiate an investigation or to proceed to a prosecution upon investigation.

Under Article 53(1)(c), should the Prosecutor have "substantial reasons to believe that an investigation would not serve the interests of justice," he or she may determine that there is no reasonable basis to proceed to an investigation. Similarly, the Prosecutor can decide whether or not to proceed with a prosecution even following an investigation, as provided in Article 53(2)(c). Although this determination must take into account elements such as "the gravity of the crime and the interests of victims" (Article 53), the Prosecutor's ample discretionary powers to decide on whether a particular prosecution should go ahead or not may limit the possibility of victims seeking and obtaining redress through the ICC.

The Rome Statute, however, provides a series of checks and balances to prevent abuse of power on the Prosecutor's part. For example, should the Prosecutor decide not to proceed with an investigation or with a prosecution following an investigation, the Pre-Trial Chamber retains the power to review and confirm the decision, either on its own motion or at the request of the referring party.

Should the Prosecutor decide that there is a "reasonable basis" to proceed with an investigation, he or she must inform the victims or their legal representatives of the decision to proceed, unless the Prosecutor, at his or her discretion, decides that this "would pose a danger to the integrity of the investigation or the life or wellbeing of victims and witnesses" (Rule 50).

Also, in terms of limitations for the victims, one needs to consider that, unlike the Security Council's authority to refer to the Court situations of genocide, war crimes, or crimes against humanity taking place anywhere in the world, even where the country in question is not a party to the Rome Statute, the Prosecutor's capacity to investigate a particular situation out of his or her own motion must comply with the general conditions of territoriality or nationality. In other words, the Prosecutor can only initiate proprio motu investigations into situations that either have occurred in the territory of a State Party to the ICC or that have involved accused who are nationals of a State Party to the ICC.

Victims Can Make Representations to the Pre-Trial Chamber

Full participation of victims at the pre-trial stage is an essential prerequisite both for the preparation of an eventual trial and for the purposes of reparation. Once the Prosecutor has decided to proceed with an investigation of a situation under the jurisdiction of the Court, the matter goes to the Pre-Trial Chamber for a judicial review of the decision. At this stage, victims can make representations to the Pre-Trial Chamber (Article 15(3)), and the Chamber may decide to hold a hearing to collect additional information.

Victims are the only ones who can make representations at the pre-trial stage, since neither States Parties nor individual suspects can challenge the admissibility of the Prosecutor's request for authorization of an investigation. During the pre-trial stage victims can also trigger the power of the Pre-Trial Chamber to take protective measures for the purpose of forfeiture, which is often an essential step in ensuring proper financial coverage of an eventual reparation award.

Also, in case the Prosecutor seeks a ruling on a question of jurisdiction or on the admissibility of a case, victims or their representatives can make representation to the Court to submit their observations (Article 19(3) in combination with Rule 59).

Victims Can Make Opening and Closing Statements and Participate in Hearings

The Chamber has wide discretion to determine ways in which victims can participate in the proceedings. Once victims have filed a written application to "present their views and concerns" (Article 68(1)), it is up to the Chamber to decide whether or not to accept the application and, in case it does, to specify appropriate ways for victims to participate in the proceedings, including through the making of opening and closing statements (Rule 89).

However, the ICC Statute indicates at least one way in which victims may participate in the proceedings, and this is by way of participating in a hearing and posing questions to witnesses, experts, or the accused either in writing or orally, subject to the Chamber's prior approval. In the case of a hearing for reparations, the victims' legal representative must not be subject to any restriction on the questioning of witnesses, experts, or the accused prior to the approval of the Chamber (Rule 91). Yet, both the Statute and the Rules clarify that victims' participation in the proceedings will be allowed only if it is "not prejudicial to or inconsistent with the rights of the accused and a fair and impartial trial" (Article 68(3)) or if it does not hinder the "need for a fair, impartial, and expeditious trial" (Rule 91). It is therefore up to the Chamber to be responsive to the requirements of the Statute on victims' access to the proceedings and to ensure proper ways for victims to participate in the justice process.

Establishment of Reparation Procedures

The ICC's powers in relation to the victims' right to seek and obtain reparation for damages represent a major advance. The Court's capacity to award reparations directly to the victims was among the more controversial issues debated both prior to and during the Rome Diplomatic Conference. Delegations expressed concern that a direct reparation system would be difficult to manage and would not be compatible with the Court's main responsibilities to prosecute and punish the offender. However, delegations finally agreed on a rather innovative formula providing the Court with the power to determine the scope and extent of damages, losses, and injuries suffered by victims once an individual is found guilty of one of the crimes under the jurisdiction of the ICC. According to Article 75:

> The Court shall establish principles relating to reparations to, or in respect of, victims, including restitution, compensation and rehabilitation. On this basis, in its decision the Court may, either upon request or on its own motion in exceptional circumstances, determine the scope and extent of any damage, loss and injury to, or in respect of, victims and will state the principles on which it is acting.

Article 75(1) adopts non-exhaustive language but also lists specific reparation elements such as restitution, compensation, and rehabilitation, thus resolving what had been a lengthy debate over redress terminology.

Although the Court can determine damages, losses, or injuries either upon request or proprio motu, the reparation procedure is designed to be accessible mainly at the request of victims. The Court therefore acts on its own initiative only "in exceptional circumstances." A victim's request for reparation must be made in writing and filed with the Registrar and must contain certain specific information such as the "identity and address of the claimant"; a "description of the injury, loss, or harm"; the location and date of the incident; the identity of the person the victim believes to be responsible for the violation, if possible; a description of the objects for which restitution is claimed; a claim for monetary compensation, rehabilitation, and other forms of remedy; and any other relevant supporting documentation.

In general, the drafters of the Rome Statute conceived the reparation regime and relevant proceedings to function as a widely publicized process. Under Rule 96, the Registrar has an obligation to give publicity "as widely as possible and by all possible means" to the reparation proceedings before the Court "to other victims, interested persons and interested states." To this end, the Registrar may seek the cooperation of relevant States Parties and the assistance of NGOs.

Following the determination of the scope and extent of damages, losses, and injuries suffered by the victims, the Court can expedite payment in one of two ways. It could:

– Make an order directly against the convicted person "specifying appropriate reparations to, or in respect of, victims, including restitution, compensation, and rehabilitation"
– Where appropriate, order that "the award for reparations be made through the Trust Fund" established under Article 79 of the Statute.

The Court has the authority to exercise its full discretion in the choice of the reparation awards. Reparations may be awarded in the form of restitution of property, compensation, or rehabilitation but also in forms other than those specifically listed in the Rome Statute, such as orders for apologies, satisfaction, or guarantees of non-repetition. Although the Rome Statute does not contemplate expressly the possibility of awarding punitive damages in addition to compensatory damages, this option might fall within the Court's discretion. The Court must give consideration "to the extent of the damage caused, in particular the harm caused to the victims and their families" in determining sentence under Article 78. Conversely, wherever appropriate, the Court must also consider efforts by the convicted person to compensate the victims as a mitigating circumstance (Rule 145).

Creation of a Trust Fund

The ICC has the authority to channel a reparation award against a convicted person through the Trust Fund, which is established under Article 79 "by decision of the Assembly of States Parties for the benefit of victims of crimes within the jurisdiction of the Court and of the families of such victims." Broadly speaking, the Fund can be used almost as an intermediary for the transfer of awards, for example, in the case of mass claims. The Statute identifies two sources of income for the Trust Fund as follows:

– Reparation awards against a convicted person ordered by the Court to be made through the Trust Fund (Article 75(2))
– Money or other property collected through fines or forfeiture under Article 77, ordered by the Court to be transferred to the Trust Fund for the benefit of the victims and their families (Article 79(2))

The choice of the language in Articles 75 and 79 reflects the different scope of application of the two provisions with respect to the Trust Fund. Article 75(2) establishes that "the Court may order a reparation award against a convicted person to be made *through* the Trust Fund," which suggests that the Trust Fund is bound to use the proceeds of a specific reparation award exclusively for the benefit of the victims with respect to the particular case before the Court, rather than to use it for other cases or situations. In this regard, the Trust Fund functions more as an intermediary for the transfer of individual awards. Rule 98(2) indicates accordingly that the Court could order an award for reparation against a convicted person to be deposited with the Trust Fund only in cases where, at the time of the order, "it is impossible or impracticable to make individual awards directly to each victim." This particular award will remain separate from other resources of the Trust Fund and will be distributed to each victim "as soon as possible." However, Rule 98(3) also specifies that "where the number of victims and the scope, forms and modalities of reparations makes a collective award more appropriate," the Court may order such award to be made through the Trust Fund. Indeed, collective awards resulting from class actions appear to be the most likely form of reparation to be distributed by the ICC, as compared to narrowly targeted and administratively onerous individual awards.

In contrast, Article 79(2) indicates broadly that the Court can order "the transfer of money or other property collected through fines or forfeiture to the Trust Fund." This suggests that the Trust Fund is not bound by the Statute with respect to the use of money or property collected from fines or forfeiture, as long as the funds are used "for the benefit of victims of crimes within the jurisdiction of the Court, and of the families of such victims" as per Article 79(1). Rule 85(a) defines "victims" as "natural persons who have suffered harm as a result of the commission of any crime within the jurisdiction of the Court," but also:

> Organizations or institutions that have sustained direct harm to any of their property which is dedicated to religion, education, art or science or charitable purposes, and to their historic monuments, hospitals and other places and objects for humanitarian purposes.

Accordingly, the Trust Fund has the power to recognize and approve intergovernmental, international, or national organizations as beneficiaries of reparation awards directed through the Fund. The ICC Statute establishes that the Court and the Victims' Trust Fund have different budgets and financing schemes. Because the Trust Fund is not funded directly from the budget of the Court, voluntary contributions to the Trust Fund take on added importance.

A crucial question for the operation and effectiveness of the Trust Fund is precisely who can be considered to qualify as a victim for the Fund's purposes. On this point, the Statute's only indication is that the Trust Fund is established for "victims of crimes within the jurisdiction of the court" (Article 79(1)). Some have interpreted the provision of Article 79(1) as conferring "victim's status" for the purpose of the Trust Fund exclusively on family members and victims of violations under consideration in ICC proceedings, thus aligning the Trust Fund's scope of application *ratione personae* to that of the Victims and Witnesses Unit originally provided for in the Statute. Others have suggested that the Trust Fund's scope of action should cover also victims of violations who do not fall within the ambit of the Court's proceedings. This latter approach is based on the argument that Article 79 creates the Trust Fund for the benefit of victims of "crimes within the jurisdiction of the Court" in general, without specifically mentioning their participation in the proceedings. Based on this approach, the Trust Fund's resources could be accessed by any victim of crimes falling within the Court's jurisdiction, regardless of the victim's participation in the proceedings, as a witness or in another capacity.

Institution of an Office of Public Counsel for Victims

The Office of Public Counsel for Victims (OPCV) replaces the Victims and Witnesses Unit, originally established under the ICC Statute to function as a permanent statutory organ of the ICC which is financed by regular budget funding. The OPCV is intended to provide "legal research and advice to victims and their legal representatives at all stages of the proceedings," as well as legal representation to victims. Placed under the Registrar for administrative purposes, the Office comprises Counsel, Legal Officers, and Administrative Assistants. It is designed to function completely independently from the Registrar and the organs of the Court, and in this connection, all members of the Office are obliged to follow a "Code of Professional Conduct" for counselors of traumatized or threatened victims.

State Cooperation Procedures

To work effectively for the benefit of victims, the ICC reparation regime requires the full cooperation of States Parties, non-States Parties, and all others concerned. State parties are under the obligation to cooperate fully with the Court in relation to investigations or prosecutions and to ensure that reparation orders are fully enforced. To this end, the Rome Statute embodies a number of provisions aimed at supporting and implementing the victims' participation and reparation regime. Cooperation provisions come under the general umbrella of Articles 86, 87, and 88 in Part 9 of the Rome Statute. Article 86 provides for a general obligation for states to "cooperate fully with the Court in its investigation and prosecution of crimes within the jurisdiction of the Court." Article 87 grants the Court the power to request state cooperation and sets out some basic procedures to seek assistance, including a provision specifically dedicated to the assistance of victims, potential witnesses, and their families. Finally, under Article 88, States Parties have an obligation to ensure that "there are procedures available under their domestic law for all forms of cooperation" specified under Part 9. State cooperation is required from the early stages of the proceedings, particularly where the Pre-Trial Chamber, acting under Article 58, issues a warrant of arrest or a summons to appear before it and consequently decides to take "protective measures" for the purpose of forfeiture, in particular for the ultimate benefit of victims' under Article 57(3)(e). In its preliminary work, the Pre-Trial Chamber can request the cooperation of states under Article 93, which could include the identification, tracing, and freezing of the defendant's assets; the seizure of proceeds, property, and assets and instrumentalities of crimes; and the execution of searches and seizures.

The Court can also request cooperation at subsequent stages in the enforcement of ICC reparation orders. State cooperation is essential, particularly during early phases of investigation, for example, with regard to identification, tracing, and freezing of assets which could later become the object of reparation orders. Prompt State action at the request of the Court could help to prevent the alleged offender from relocating his or her assets to countries willing to shield the suspect from reparation orders.

Significantly, in its first reparations decision in the Lubanga Case, the Court ruled that Part 9 of the Statute did not prevent it from awarding forms of reparation other than those expressly mentioned in the Statute. In particular, the Court could award reparations as follows:

> By establishing or assisting campaigns that are designed to improve the position of victims; by issuing certificates that acknowledge the harm particular individuals experienced; setting up outreach and promotional programmes that inform victims as to the outcome of the trial; and educational campaigns that aim at reducing the stigmatisation and marginalisation of the victims of the present crimes (ICC-01/04-01/06-2904 at para. 239).

Jurisprudence

In the ICC's first substantial decision on the situation in the Democratic Republic of the Congo (*Decision on the applications for participation in the proceedings of VPRS1, 2, 3, 4, 5, and 6*, "Lubanga Case" ICC-01/04-01/06, 17 January 2006), the ICC Judges addressed some of the Statute's ambiguities concerning the definition and role of victims in the proceedings.

In the *Lubanga Case*, six individuals requested to participate as victims at the investigative stage of the proceedings. This prompted the Pre-Trial Chamber to consider "whether the Statute, the Rules of Procedure and Evidence ... and the Regulations of the Court accord victims the right to participate in the proceedings at the stage of investigation of a situation and, if so, what form such participation should take." In its decision, the Court also considered the question of whether the applicants met the criteria for being considered victims within the meaning of Rule 85.

The Prosecutor contended that victims could not participate in the proceedings at the investigative stage because "there are, strictly speaking, no proceedings within the meaning of Article 68 of the Statute during the investigation phase" since "from a terminological point of view, the word 'proceedings' does not encompass the investigation of a situation." The Prosecutor considered victims' participation as "inappropriate" and further argued that, in any case, the applicants had failed to show that their personal interests were affected at the investigation stage.

In response to the Prosecutor's challenge, the Pre-Trial Chamber observed that the Court had a general obligation to "take appropriate measures to protect the safety, physical and psychological well-being, dignity, and privacy of victims and witnesses," including at the investigation stage and that, in fact, there was no "explicit exclusion" of the investigation stage from the scope of application of Article 68(3) on the question of victim's participation. On this basis, the Chamber concluded that Article 68 of the Statute gave victims a general right of access to the Court, including at the stage of investigation of a specific situation." The Court noted that the victims' right of participation at the investigation stage was in fact consistent with the "object and purpose of the victims' participation regime of the ICC" and, more generally, with "the growing emphasis placed on the role of victims by the international body of human rights law and by international humanitarian law." The Chamber went on to say that victims had the right, under the Statute, to express an "independent voice and role" in ICC proceedings and noted that the victims' right to participate in the proceedings was part of a more general right to participate "in the fight against impunity," thus adopting a broad approach. During the course of the Lubanga Trial, 129 victims participated in the Court's proceedings with submissions, requests to introduce evidence, and more directly, by questioning witnesses. Three victims themselves testified as witnesses.

The approach that the Pre-Trial Chamber adopted to victims' participation in the *Lubanga Case* has come under a certain degree of criticism, especially as concerns the distinction between a victim in the context of a "situation" and a victim in the context of a "case." The difference is that, under the ICC Statute, a situation is a set of background circumstances in which crimes may have been committed, and which the Prosecutor is investigating, for example, the Darfur situation or the situation in the Central African Republic. A case, on the other hand, involves allegations, investigation, and prosecution of a particular individual or individuals who have been accused of crimes within the jurisdiction of the Court. In *Lubanga*, the Trial Chamber decided that whoever had a "personal interest" in the investigation of a situation could participate in the investigative phase of the proceedings. Potentially, this opened the door to a very large number of victims, which raises the question as to whether the Court's having to handle perhaps many hundreds or even thousands of people participating at an early investigative stage could hinder or even compromise the Prosecutor's investigative efforts which require a methodical, organized, and systematic approach.

Once the Prosecutor decides to move on from investigating a "situation" to the laying of charges and preparation of a formal indictment, the Court will have to narrow its recognition of individuals as "victims" to only those persons directly relevant to the issue of criminal guilt. Only those individuals with a clear nexus to the crime ("personal interest linked to the charges," to be assessed by the Court) would be in a position to offer testimony relevant to guilt or innocence and, ultimately, to apply for reparations. A risk is that where the Prosecutor fails to make a strong enough case, the Pre-Trial Chamber can decline to authorize the Prosecutor to proceed further, which could then leave out victims related to the particular crimes in question.

The potential shortcomings of this approach for the victims have been partially addressed by the Court's first reparation order of 7 August 2012, issued following the Court's decision to sentence Thomas Lubanga Dyilo to 14 years imprisonment for war crimes and crimes against humanity in the DRC, for having conscripted and used child soldiers in his rebel army. As regards causation, the Court observed that:

> Neither the Statute nor the Rules define the precise requirements of the causal link between the crime and the relevant harm for the purposes of reparations. Moreover, there is no settled view in international law on the approach to be taken to causation. Reparations should not be limited to "direct" harm or the "immediate effects" of the crimes of enlisting and conscripting children under the age of 15 and using them to participate actively in the hostilities, but instead the Court should apply the standard of "proximate cause." (ICC-01/04-01/06-2904 at paras 249–250)

On the same point, the Court also clarified that:

> Reparations may be granted to direct and indirect victims, including the family members of direct victims [. . .]; anyone who attempted to prevent the commission of one or more of the crimes under consideration; and those who suffered personal harm as a result of these offences, regardless of whether they participated in the trial proceedings. (ICC-01/04-01/06-2904 at para 194)

In its order, the Court noted that "reparations can be directed at particular individuals, as well as contributing more broadly to the communities that were affected" and that reparations must:

> Relieve the suffering caused by [the] offences; afford justice to the victims by alleviating the consequences of the wrongful acts; deter future violations; and contribute to the effective reintegration of former child soldiers. Reparations can assist

> in promoting reconciliation between the convicted person, the victims of the crimes and the affected communities. (ICC-01/04-01/06-2904 at para 179)

The Court also opted for a "collective approach" to reparations to ensure that reparation awards reach both victims who have applied for reparations and those who have not been identified, specifying that individual and collective reparations are not mutually exclusive and that they may be awarded concurrently. In terms of implementation, the Court ordered that the processing of claims and the subsequent awards should be channeled through the Trust Fund for Victims. Eighty-five victims have launched applications for reparations, and it is estimated that there are more than 8,000 victims who were directly affected by the crimes prosecuted in the *Lubanga Case*.

Related Entries

▶ Civil Remedies
▶ Comparative Courts and Sentencing
▶ Development of International Criminal Law and Tribunals
▶ Genocide
▶ Hate Crime
▶ Human Rights Violations in Criminal Court
▶ International Responses to Victims in Criminal Justice
▶ Multiple Victims and Super Targets
▶ Rehabilitation
▶ Restitution as a Penal Aim
▶ Victims and Restorative Justice
▶ Victims' Rights in the Criminal Justice System

Recommended Reading and References

Bottigliero I (2004) Redress for victims of crimes under international law. Martinus Nijhoff, The Hague

International Criminal Court, http://www.icc-cpi.int

Schabas W (2004) An introduction to the international criminal court, 2nd edn. Cambridge University Press, Cambridge

Sunga LS (1997) The emerging system of international criminal law: developments in codification and implementation. Kluwer Law International, The Hague

Trust Fund for Victims, http://www.trustfundforvictims.org/

Victims and the Police

Rob I. Mawby[1,2] and J. Monckton-Smith[3]

[1]University of South Wales, Newport, South Wales, UK

[2]Department of Natural Sciences, University of Gloucestershire, Cheltenham, UK

[3]University of Gloucestershire, Cheltenham, UK

Overview

The way the police respond to reported crime is a delicate balancing act. While it is almost universally accepted, and a key component of police culture, that they should prioritize apprehending the offender, complainants and victims often have broader expectations of what services the police should be offering them, and a successful detection and prosecution may not be their priority.

This entry considers police support for crime victims in different countries. However, because variations in the ways in which the police handle victims of different offense types are arguably just as notable, four contrasting offense types are considered: burglary, where the victim is generally considered "deserving" but where detection rates are low; homicide cases, where indirect victims may also be perceived to be suspects; and domestic violence and rape, both examples of where the police have traditionally been criticized for victim blaming.

Police Support for Victims

It is important that the police provide a satisfactory service for victims for a number of reasons. For example, the police are highly dependent upon the public, and especially victims, for bringing crime to their attention and providing leads on the offender. Victims who feel that the police do not provide an appropriate response may fail to report crimes or even take unsanctioned vigilante action. Moreover, despite the emergence of victim assistance programs in many countries, the police service is still the main agency with which victims have contact. As a result, police response to victims may be their most significant post-crime experience (Mawby 2007). Most important though, victims deserve to be treated with courtesy and respect.

However, traditionally, police forces in a range of countries have provided a very poor service for victims, with the result that they have often exacerbated the effects of the crime, promoting secondary victimization rather than support. This is particularly the case in countries where the police system is (or was) organized on paramilitary lines, as in colonial and communist societies and in many parts of continental Europe (see chapter by Mawby). However, cross-national studies, including the international crime victims survey (ICVS), indicate that victims' criticisms of the police are reflected across a broad range of industrial societies, including the USA, England and Wales, and even Japan, and that criticisms cover lack of interpersonal skills, poor response, failure to take the incident seriously, and not keeping the victim informed of further developments, as well as failure to apprehend the offenders or recover stolen property (van Dijk et al. 2008; van Kesteren et al. 2001; Zvekic 1998).

Given the extent of victims' criticisms, it is scarcely surprising that policies have been implemented in a number of countries aimed at improving police services. Approaches adopted by police services in the USA include specialist police training in victim-oriented work (Rosenbaum 1987) and enhanced police support (Skogan and Wycoff 1987). In the European Union, legally binding minimum standards of victim treatment came into force in 2003 (van Dijk et al. 2008). These included the requirement that victims be treated with consideration and respect by the police and kept informed about investigation and prosecution decisions. Additionally, some countries have introduced more specific requirements and services. For example, in the Zaarstrad area of the Netherlands, Winkel (1989) described an initiative that incorporated both specialist training for the police and additional support for victims from the police. Also in the Netherlands, a series of guidelines for police and prosecutors were issued, requiring the police to treat victims sympathetically, provide all the relevant information, and, where necessary, refer them to other agencies, and victims now have a right to cite the guidelines should they subsequently take legal action against the police (Wemmers and Zeilstra 1991). A similar approach has been adopted in England and Wales, with the 1990 and 1996 Victims' Charters spelling out expectations of how the police should respond to victims and the 2005 Code of Practice detailing the police's obligations (Mawby 2007).

The introduction of international and national standards suggests a greater victim orientation of the police in many countries. However, evaluations of different initiatives indicate a resistance to change. A traditional emphasis upon action, excitement, and "real" policework means that the police often accord a lower priority to work with victims. In fact, despite legislative and policy changes, there is considerable evidence that victims' satisfaction with the way the police have responded has fallen. For example, the 2005 ICVS shows a general decline in victim satisfaction, with the most notable fall in England and Wales, the USA, the Netherlands, Canada, and Sweden (van Dijk et al. 2008). National surveys sometimes paint a different picture. In England and Wales, for example, the 2009/2010 British Crime Survey (BCS) reported a rise since 2008 in victims' satisfaction, with 37 % very satisfied and 32 % fairly satisfied with the police in 2010 (Flatley et al. 2010). Nevertheless, the same survey reiterates earlier findings that victims are more critical of the police than are other citizens.

The way the police respond to victims of different types of offense also varies markedly. Local or national priorities, plus different perceptions of the precise impact of different offenses, mean that agency interpretations of need – and, perhaps more importantly, desert – vary. In some countries (such as England and Wales), the needs of burglary victims have been prioritized, although this is changing (Mawby 2001). In others (like the USA), there has been a greater focus on victims of violent and/or sexual crimes (OVC 2001), partly reflecting feminist influence on early initiatives and partly the high homicide and rape rates. A rather different distinction in the types of victim receiving specialist police support is the recent development in tourist areas of programs dedicated to helping tourist victims. While in some countries, such as the Irish Republic and the Netherlands, such services are provided out with police departments, in others, including the USA, South Africa, and India, specialist Tourist Police units exist.

Because of the stark differences between the ways in which the police handle victims of different offense types, the following sections cover four contrasting examples. The first focuses on burglary, where the victim is generally considered "deserving" but where detection rates are low. Then we address police response in homicide cases, where the police often have to balance the need to treat surviving relatives as both indirect victims and suspects. Finally, we consider domestic violence and rape, both examples of where victim blaming has traditionally featured on the agenda.

Police and Burglary Victims

The circumstances of police intervention in cases of domestic burglary are quite distinct. Firstly, in most cases, the police are called in to a "cold" crime, that is, some time after the incident occurred. Secondly, and as a result of this, detection rates are low, in England and Wales currently about 13 %. Thirdly, in most cases, the status of the complainant as an "innocent victim" is not disputed, although in some cases, the police may suspect the complainant of an insurance scam or similar and in others they may blame the victim for not protecting their property adequately.

Despite low detection rates, victims tend to be relatively positive about police handling of their complaint. In the UK, for example, they tend to be more positive than victims of violent crime, where detection rates are higher and less critical of the police than victims of vehicle-related crime (Mawby 2001). However, the most recent ICVS suggests that burglary victims' criticisms were directly related to the specific circumstances of the offense. Thus, 30 % of critical victims complained at the time it took the police to respond, an indication of the fact that the police give low response priority to "cold" crimes, while the low detection rate is reflected in criticisms that the police failed to apprehend the offender (cited in 58 % of cases) nor recover stolen goods (49 %).

Research by Mawby and his colleagues assessed victims' feelings about the way the police dealt with them (see, e.g., Mawby 2001), comparing cities in two Western European countries (England and Germany) with three former Warsaw pact countries (Poland, Hungary, and the former Czechoslovakia). They addressed three general questions vis-a-vis police services: How do the police handle complaints of burglary, how far do burglary victims from East and West Europe share similar perceptions of the police, and to what extent are their assessments of the police influenced by policing traditions, their perceptions of the crime problem, and by the services provided by the police today?

In general, victims from Poland expressed considerably more criticism than did those from the other four countries, including Hungary and Czechoslovakia. Differences were pronounced when those who expressed any dissatisfaction were asked what it was about the police response that they disliked, using a similar pre-coded checklist to that of the ICVS and BCS. Overall victims were most likely to cite lack of feedback, the fact that the crime was undetected, property not being recovered, and criticism that the police did not do enough or were not interested. However, while lack of feedback was one of the most common criticisms in all five countries, in other respects, the emphasis was different. Most notably, whereas in England victims were also most likely to criticize the police because they "did not do enough" or were "not interested," in Eastern Europe in general, and Poland in particular, the most common complaints were that the crime was not detected and that property was not recovered. This parallels Zvekic's (1996) analysis of the second ICVS and may relate to the fact that in these countries, far less victims were insured.

Arguably then, insurance compensation may mean that criticisms of police reactions are more muted than they otherwise would be.

Victims were also asked how sympathetic they felt the police were when dealing with victims of (a) burglaries, (b) disasters like fires and floods, and (c) rape and sexual assault. Although responses varied with each situation, overall respondents from Poland and Hungary were about as likely as those from England to see the police as sympathetic, and it was in Germany, with its militaristic policing tradition, that victims were least likely consider the police sympathetic. Victims also felt that the police would be most sympathetic in the case of a burglary, less so for sex crimes and in disaster situations. This may indicate that victims felt the police had been particularly victim-oriented vis-a-vis their burglary. On the other hand, it may reflect a real difference in police response to the three situations, a conclusion that is in line with the ICVS data discussed above and may reflect perceptions of the "innocent victim."

Victims were then asked whether, over the last few years, "the police have got better or worse at handling the victims of crime." Burglary victims from Poland, Czechoslovakia, and (especially) Hungary were likely to perceive an improvement rather than a deterioration in this respect. Victims from England were least positive, although in many cases, lack of police resources was cited as the main reason for a decline in police service. It is also possible that as specialist victim services have expanded in recent years in countries such as the UK, so the police have abdicated their responsibilities and instead referred victims elsewhere. In this respect, van Dijk, van Kesteren, and Smit (2008) note that levels of satisfaction with police response have dropped most in countries like the USA, Canada, the United Kingdom, Sweden, and the Netherlands where specialized victim assistance agencies have become well established.

International differences in burglary victims' satisfaction with the police are at least partly due to the nature of police organization and priorities and the expectations citizens hold of their police. However, differences between countries are also, at least partly, due to police practices that may be crime specific. For example, in Poland, where victim dissatisfaction levels were considerable, it was routine for the police attending a burglary to remove door locks for "forensic examination," a practice that actually did nothing to improve detection rates but unequivocally added to victims' grievances!

Police and (Indirect) Victims of Homicide

Homicide is a crime judged very much in context, not only in the public perception but also police perceptions, which can have serious repercussions for victims (Monckton-Smith 2010, 2012). For example, in policing terms, a so-called stranger murder will be approached very differently in the amount of resources and expertise it receives to a domestic homicide. Those homicides where the offender is immediately identifiable, and the victim is adult and an intimate partner of the killer, may well be processed by less experienced officers. In such cases, the police priority is not to handle the victims with skill, but to produce a file of evidence for court, which may impact on the victim experience. Also, as noted above, victims will be assessed as more or less deserving, and this can impact on the way they are treated. The primary victims are often perceived as more deserving the more intimate distance there is from the killer (Monckton-Smith 2010, 2012), and this has been mirrored in approaches to victims of domestic abuse where the outcome is not fatal (Dawson 2003). In this sense, the reputation of the primary victim can be attacked, especially in court, which causes further trauma for families who cannot defend their loved one and a more negative perception of their dealings with criminal justice agencies (Casey 2012; Monckton-Smith 2012). There are many aspects to a homicide which will make the relationship between secondary victims and the police more problematic and may lead to dissatisfaction. Despite the comparatively high clear-up rate for homicide (around 80 % in the UK), victim experience of their dealings with criminal justice agencies is significantly negative on the whole.

Because in a homicide the primary victim is dead, police contact will be with surviving relatives and friends (indirect victims) who are suffering from a severe and exaggerated grieving process (King 2004). Research has shown that in cases of homicide, it is the way that these victims are treated by the police that has the most significant impact on their overall level of satisfaction (King 2004; Kilpatrick et al. 1990) and that more careful consideration of the way indirect victims are treated can have a major impact on the extent to which they experience psychological trauma (Kilpatrick et al. 1990, p. 1). It is even suggested that those indirect victims who have no involvement with criminal justice agencies at all actually cope with trauma better (Casey 2011).

An added difficulty is that the families of homicide victims are more likely to have sustained interactions with the police, resulting from quite complex and diverse issues. Due to the seriousness of the crime, investigations and the court process can be protracted; there is much at stake for the accused and a high likelihood of a trial which is stressful for victims. Where offenders plead guilty, victims may have no involvement in the court process which leaves them feeling frustrated (Casey 2011; Kilpatrick et al. 1990). Unsolved cases can be investigated for lengthy periods of time, sometimes years, and then re-investigated if new evidence is discovered or forensic techniques are developed. Even successful convictions may become the subject of an appeal and further extend family interactions with the criminal justice process. Seemingly simple procedures like the release of the body for funeral can also take a long time while the inquest is completed, and the need for a postmortem, or very often, more than one postmortem, is carried out. This in itself has been a subject of criticism from victims who question the need for second postmortems and lengthy inquests (Casey 2011). In the UK, the former Victims Commissioner Louise Casey campaigned to have a process whereby all bodies are released to families within 1 month to try and make the process more efficient and easier for relatives.

In cases where no killer is identified, the process is especially long, and family suffering is even more exaggerated, with their need for justice frustrated by what might be perceived as police failings in the investigation, and anger may be directed at them (King 2004), creating a very complex and sensitive situation. Kutnjak Ivkovic (2008) found that the more contact that individuals have with the police, in general terms, the more dissatisfaction there is. So there is much scope for victims to feel dissatisfied with their treatment by the police in cases of homicide.

It is fair to say that family and friends of victims have been sidelined by the criminal justice process, having few rights and no legal status in the court or police process. Many complain that the offender has more rights than they do and is given a more robust voice in proceedings (Monckton-Smith 2012; Casey 2012; Kilpatrick et al. 1990), and this is often most noticeable in those cases where families have been so frustrated by failings or weaknesses in the police response that they campaign for change.

It is the high-profile enquiries and reinvestigations resulting from homicide cases, which are the clearest demonstration of victim dissatisfaction. There is little research which documents the homicide victim's experiences of the police in a generic sense but an abundance of individual examples. A prime example from the UK is the Macpherson (1999) enquiry into the racist murder of Stephen Lawrence in 1993 where the police treatment of the family was heavily criticized; the Bedfordshire police report into the Cambridgeshire police investigation into the murder of Claire Oldfield-Hampson, again criticized the way the families were dealt with (Cambridgeshire Police 2011). Similarly in Belgium, the police handling of the investigation into the murders committed by Marc Dutroux rocked public faith in the Belgian police and led to a complete restructuring of policing (Robbers 2008). The murders of Ronald Goldman and Nicole Brown Simpson in Los Angeles, USA, allegedly committed by OJ Simpson, brought significant criticism of the police; the murder of journalist Alfredo Villatoro saw Honduran Police Chief General Ricardo Ramirez el Cid removed from position, shortly after the removal of his predecessor, Jose Luis Munoz, for failings by police in the murder investigation of two university students (Fox 2012). There is a reported nose dive in confidence in the Nigerian police partly as a result of their failings in high-profile homicide cases (Adisa 2012). In some jurisdictions, it is corruption that is cited as failing victims; in others, it is arrogance and incompetence of the police. But across the world, it is homicide cases that have the most potential, it seems, to bring criticism to police services and to instigate change.

These high-profile criticisms into police handling of homicide cases have probably led to many of the changes which have sought to improve the victim's experience. The relationship between the police and victims in any jurisdiction will be affected by wider relationships with the state which will also impact on police performance and attitude (Kutnjak Ivkovic 2008). The biggest criticism from victims' families is police attitude, more so than prosecution outcomes and sentencing decisions. The key criticisms are that the defendant was treated better than victims (Casey 2011; Kilpatrick et al. 1990) with many families expressing the view that there was no aspect of the criminal justice response with which they were satisfied.

Varying levels of satisfaction have been found in the respect given to secondary victims, with some citing police responses as supportive, but lack of consideration was also cited as a key problem in other cases (Kilpatrick et al. 1990). Similarly in the UK, in one study, 76 % considered the police fairly or very supportive (Casey 2011), but this was often different where no offender was identified. The biggest issue here and in other cases was the lack of information given to families, and this has been replicated across all studies. King (2004) reports that some of the major complaints from indirect victims include being shown a lack of compassion, being given incomplete information, a lack of accessibility, ignorance of the grieving process by police, and dishonest answers to questions. A Russian study found that in more general terms, victim interactions with police were marred by rudeness of officers and a lack of trust (Kutnjak Ivkovic 2008).

Victims have long complained that they are not listened to, and in the UK, the USA, and Canada, among other countries, there are now Domestic Homicide Reviews/Fatality Reviews which are inclusive of indirect victims, giving them more room for a voice in the process, but these would only occur after an initial investigation. It has been suggested that more categories of homicide could benefit from such a process (Casey 2011). The introduction of Victim Impact Statements again addresses the need for the victim to be heard.

It is clear from the studies across the world that the way victims are treated is key to their satisfaction with the process, and dignity and respect are the main characteristics seen as lacking. It appears to be a recurring theme that sensitive treatment is what is required but also what is lacking.

Police and Victims of Intimate Partner Abuse

Intimate partner abuse (IPA) or, more broadly, domestic abuse and violence is receiving unprecedented attention worldwide in an attempt to reduce its prevalence and respond to victim's needs (Monckton-Smith 2012). It is a problem of global proportions. Despite the emergence of the notion of gender symmetry in domestic abuse which has precipitated much more sympathetic police responses to male victims, it is the case that the predominant victim group is female. They are killed at a far higher rate and receive more sustained, prolonged, and serious abuse than men (Stark 2007; Websdale 1999). It is also difficult in policing terms because even police departments with more liberal attitudes and proactive response policies are not really geared to respond to the most common forms of abuse against women, which may not be characterized by violence, but by what Evan Stark (2007) calls coercive control. However, coercive control is significantly associated with serious and fatal abuse.

There are now at least 45 countries with specific domestic violence legislation (Unifem 2010), but the most intractable problems are universally acknowledged to be in deeply held prejudicial beliefs about male and female gender roles and gendered behavioral characteristics, which can permeate the official response and undermine the spirit and practice of legislation, as well as direct the behavior and practices of individuals and communities (Monckton-Smith 2012, p. 25). It is well documented that victims of IPA have received poor attention from the police in the past with accusations that they were disbelieving, disinterested, and keen to minimize the harm (Stark 2007; Websdale 1999; Dobash and Dobash 2002). However, despite the changes implemented to improve the response to victims, Dobash and Dobash capture the criminal justice and political zeitgeist in claiming there has been great change and no change at all (2002, p. 21).

Underreporting of domestic abuse is a significant issue with many women not wanting police intervention for fear of antagonizing the assailant or as the result of cultural beliefs which construct wife abuse as a private matter (Hoyle and Sanders 2000; So-Kum Tang 2003; Sun et al. 2011). Many victims seem to desire police intervention to end the immediate danger or to calm the assailant down, but no more (Hoyle and Sanders 2000). The purpose of involving police appears to be an attempt to end the ongoing violence, rather than instigate the power of the law to achieve a conviction. Hoyle and Sanders (2000) found that police intervention rarely has this effect, and abuse and violence often continued. If the victim is seeking to remove herself from danger in calling the police, then she will need a particular response which acknowledges this and has the capacity to manage her safety. Such is the institutional frustration with women who either refused to cooperate with arrest and prosecution; that in the UK and USA, pro-arrest or mandatory arrest policies were implemented for domestic violence calls. The follow-up studies had mixed results suggesting that those men in the higher socioeconomic groups were most likely to be deterred from future violence, having more to lose (Schmidt and Sherman 1993). It is also the case, however, that while an immediate arrest may reduce the threat in the very short term, women without adequate resources to use that time window when the abuser is incarcerated will be in danger of repeat violence on his return and in the long term.

Unfortunately, in many cases where women do report IPA, there is a general dissatisfaction with police handling of the incident and the subsequent investigation, if indeed there is one (Stephens and Sinden 2000). Victims across the world complain that police officers are disinterested, hostile, or frustrated with victims for their reluctance to follow complaints through or permanently leave the abuser. It was found in one American study of trainee law enforcement personnel that their attitudes to women and IPA were significantly more censuring of victims than students in social work (Meyer 2011), suggesting that those entering law enforcement are holding discriminatory attitudes prior to any institutional influence. Studies in China show that those police officers who hold traditional attitudes which see women as “wives” and “home builders” rather than equal citizens are far more likely to be disbelieving and disinterested (Sun et al. 2011). Stephens and Sinden (2000) found that there were four key police demeanors noted by domestic abuse victims which caused dissatisfaction: minimizing the situation, disbelieving the victim, “we don’t care,” and “macho cop.” Similar to the findings in the treatment of homicide victims, domestic abuse victims were most satisfied with police contact when they were shown care and respect with listening and believing the most important characteristics (Stephens and Sinden 2000).

Meyer (2011) reports a study in the USA which shows that over 70 % of victims in their sample reported dissatisfaction with the police response, largely characterized by disinterest, but also that taking the report was a waste of time and effort. Research in China also shows that traditional attitudes to female gender roles are implicated in poor police response to IPA and a tendency to blame the victim (So-Kum Tang 2003).

Even though China has specific legislation outlawing IPA, authorities still do not always consider IPA a matter for the police, sometimes refusing to even record the initial complaint (Sun et al. 2011). Respondents in this research reported they would be more likely to seek help from friends and family (Sun et al. 2011). Reports from Sweden show better victim satisfaction with IPA intervention, largely attributed to their social welfare approach to the problem which gives women more access to resources and state support which helps them manage their safety (Sun et al. 2011). A Canadian study found that not all victim groups have benefitted even where there are developments in victim treatment by the police. This study found that Aboriginal women experience violence at disturbingly high rates with their mortality rate from violence being three times higher than for non-Aboriginal women (Dylan et al. 2008). The police response was said to be marked by disrespect, dismissal and professional failure, and the revictimization of women by the system a common feature (Dylan et al. 2008).

Victims of IPA almost universally express similar problems with police responses which inhibit them from reporting and are related to cultural ideas of male and female gender roles which place women in more danger after reporting. There is a prevailing belief in institutional response to IPA that legislation to further conviction should be used and is helpful to the victim, but this fails to take account of victim needs which are largely about safety planning.

Police and Victims of Sexual Assault

Sexual assault is another form of offending which is receiving worldwide attention. As a result of human rights legislation and more inclusive and nondiscriminatory policies, in Europe and Westernized countries, the victimization of men is also receiving far more attention and practical support. Historically, similar to victims of IPA, victims of sexual assault have received poor service from the police, and this has been articulated in the strongest terms by victims and victim's advocates. It was high-profile feminist campaigning that raised awareness of its prevalence and the damage it causes leading to widespread policy changes with regard to the treatment of victims (Monckton-Smith 2012). Sexual assault is a diverse offending category which encompasses a diverse victimology. The most common forms of sexual assault, however, are those perpetrated against women or children by male family members or acquaintances. Similar to IPA, this form of assault was largely hidden within the family. Pedophilia was made visible by feminist campaigning in the 1970s (Kitzinger 2004) and has become more openly reviled as time has moved on, with pedophiles perhaps being the most hated of offender groups. Despite this, child victims are still not always believed, especially where they are close to the offender (Kitzinger 2004). Similarly, rape is widely considered one of the worst of violent offenses, and it is perhaps this construction of the crime, as particularly evil, that protects offenders from conviction. It is difficult, after all, for a seemingly ordinary and plausible man to look like a "rapist" in this construction of the offense and offender (Monckton-Smith 2010).

There is still a belief that a "real rape" occurs in a public place, committed by a stranger to the victim, and involves aggravating violence and use of a weapon, a perception that is described as more akin to an aggravated rape assault (Kelly et al. 2005). This perception, however, is setting the standard for what "rape" is and reflects what Kelly et al. (2005) refer to as the "rape template." This template reflects a very skewed idea of what a rapist or a rape victim looks like and even the victims themselves do not always identify their assault as a rape. The British Crime Survey found, for example, that "less than half of women who experienced an assault that met the legal definition of rape defined it as such themselves" (Kelly et al. 2005, p. 33).

Attrition in cases of rape is still extremely high and has been the subject of sustained research interest (Kelly et al. 2005). In the UK, the Fawcett Society claims that a rape is reported to the police every 34 min, but only 6.5 % will attract a conviction. Vera Baird reports that of those rapes where the alleged offender is charged by the

CPS, 59 % will receive a conviction, suggesting that the problems are at the very start of a victim's involvement with the criminal justice system. This is not a problem peculiar to the UK; it is a worldwide issue. Kelly and Regan (2001) found similar low conviction rates across Europe. For example, of the 468 rapes reported in Finland in 1997, only 46 resulted in a conviction; of the 1,962 reported in Sweden, only 115 achieved a conviction; and of the 424 reported in Norway, only 35 achieved a conviction. In the USA and Australia, similar problems are documented, suggesting a conviction rate of just 12 % (Kelly et al. 2005). South Africa has one of the highest rates of sexual violence against women in the world, but it is estimated that only around 16 % of those reported attracted a conviction (cited in Monckton-Smith 2010, p. 4). The attrition rates are a stark reminder of the problems victims have with the police. As noted, most cases are lost in the very earliest stages, that is, where the victim either makes a decision to approach the police or not, and most are choosing not, and then in the initial investigation and early contact with officers. It has been found, similar to the experiences of indirect victims of homicide, that those women who have no contact with police and the criminal justice system are more less likely to suffer trauma (Sleath and Bull 2012).

The UK is one jurisdiction where it could be said that improvements for rape victims have been made, yet victims still report significant problems with the police response. The case of taxi driver John Worboys, convicted of 19 counts of rape in the UK in 2009, is a recent example of the prejudice victims of sexual assault can suffer from police. An IPCC (Independent Police Complaints Commission) enquiry into this case found police to have given poor service to victims. The accusation, which was supported, was that victims were not believed and some were allegedly openly scorned. These findings show little improvement in police response from the time of the now infamous Thames Valley fly on the wall documentary *Police* (broadcast 18.01.82 BFI 2005) which revealed the institutional prejudice against certain victims of rape. MP Fiona MacTaggart said of the Worboy case:

> I was depressed by the Independent Police Complaints Commission report into Warboys, because it showed that still there persists, even among specialist police officers, insufficient sensitivity to the experience of the victim of rape and an insufficient determination to treat rape as a major violent crime that must be investigated. (Hansard 2010, p. 180)

It has been consistently found that police and prosecutors overestimate the amount of false allegations and that victims are treated with skepticism (Kelly et al. 2005) and that there is a relationship between belief in the rape myths/template and a failure to recognize sexual assault. Some of the rape myths include women lie about rape, only certain types of women get raped, women encourage rape in their manner of dress, rape is sexual, you can recognize a rapist by the way he looks, women fantasize about rape, rape is impulsive and spontaneous, rape is about passion, and a person who has been raped will be hysterical (Monckton-Smith 2010).

A culture of blame is held largely responsible for the poor response to rape victims by the police, and it has been found that even officers who receive specialist training show no decrease in acceptance of rape myths and levels of victim blaming. However, in the USA, it was found that officers experienced in rape investigation were less accepting of rape myths (Sleath and Bull 2012).

Research demonstrates that as rape is so traumatic, a sensitive response by police can be easily appreciated, whereas negative attitudes can have a devastating effect (Sleath and Bull 2012, p. 650). Sleath and Bull (2012) report, in accordance with other studies, that victims of stranger rape are considered less blameworthy than victims who knew their attacker; similarly, victims dressed in revealing clothes or who were intoxicated were assessed as more blameworthy. These kinds of attitudes are correlated with victim dissatisfaction with police responses to the rape; they deter women from reporting and lead to secondary victimization when they do report (Monckton-Smith 2010).

Summary and Discussion

Ideally, police response to crime victims should encompass a service approach alongside a concern to do everything possible to clear up the crime and provide justice for victims. Certainly, as the example of burglary illustrates, victims are concerned about detection and the return of stolen property. But victims also expect the police to respond sympathetically, to treat them as people rather than crime numbers, to provide help or advice where necessary, and to keep them informed of any progress regarding "their" crime. It is against these criteria, in addition to detection rates, that police services for crime victims need to be measured. Moreover, the balance between detection and service is not a zero sum equation. Police who respond appropriately to victims of domestic violence may prevent reoffending. Police who are sympathetic and supportive to all victims of crime may gain better cooperation from victims and consequently elicit evidence to identify a suspect and/or gain a conviction.

In considering police response to crime victims in different parts of the world, it is evident that there are some differences between victims of different types of crime. Of the examples detailed here, burglary victims, while somewhat critical of the police, are generally more positive about the services they receive, suggesting that part of the problem lies in the police being judgmental and stereotyping some victims, or victims of certain offense types, as more or less innocent (in terms of provocation) and more or less deserving (of any support). There are also international differences where traditional police practices in some countries may exacerbate the problems faced by victims, and new initiatives elsewhere may improve, even if only marginally, victims' experiences. An example of the former is police practices during burglary investigations in Poland. An example of the latter is police practices around domestic violence in the UK, which are reputed to be pioneering in understanding and responding to the problems (Choi 2009).

Nevertheless, the depressing finding is that despite policy changes, the police are regularly found wanting in their treatment of victims of crime. The findings that victims who are in contact with the police frequently feel worse affected by the whole crime experience and are often more critical of the police in general than those who do not report their crime are sad indictments of police service. The reasons for this are, perhaps, twofold. In part, police failure to respond appropriately to victims may be a consequence of a police culture where excitement and action are prioritized over support and care. In part, and not unrelated, it might reflect the types of citizen who become police officers and the features that attract them to a police career. While legislation and policy initiatives aimed at improving police procedure are to be welcomed, it is only by changing this police mind-set that the service provided to victims of crime can be significantly improved.

Related Entries

- ▶ Domestic Violence
- ▶ Feminist Theory in the Context of Sexual Violence
- ▶ International Responses to Victims in Criminal Justice
- ▶ Police Family Violence Services
- ▶ Residential Burglary
- ▶ Victim-Focused Prevention of Child Sexual Abuse
- ▶ Victims' Rights in the Criminal Justice System

Recommended Reading and References

ABS (1996) Women's safety Australia. Catalogue no 4128.0. Australian Bureau of Statistics, Canberra

Adisa T (2012) Nigeria police: threat to national security? Nigerian Tribune, Friday 27 January

BFI (2005) British Film Institute screenonline. Accessed at www.screenonline.org.uk/tv/id/464502

Cambridgeshire Police (2011) Enquiry into the conduct of the investigation into the death of claire oldfield Hampson. Accessed at https://www.cambs.police.uk/about/foi/disclosure/PUB0999-2011.pdf

Casey L (2012) Review into the needs of families bereaved by homicide. Accessed at http://www.justice.gov.uk/downloads/news/press-releases/victims-com/review-needs-of-families-bereaved-by-homicide.pdf

Choi K (2009) Tackling domestic violence in the UK: Persistent problems. Int Area Stud Rev 12:17–43

Dawson M (2003) The cost of ‘lost’ intimacy: the effect of relationship state on criminal justice decision making. Br J Criminol 43:689–709

Department of Justice (2007) US Department of Justice National Crime victimisation survey 2007. Accessed at http://www.ojp.usdoj.gov/bjs/pub/pdf/cv07.pdf

Dobash RE, Dobash RP (2002) Women, violence and social change, 4th edn. Routledge, London

Dylan A, Regehr C, Alaggia R (2008) And justice for all? Aboriginal victims of sexual violence. Violence Against Women 14:678–696

Fox E (2012) Honduran police chief removed amid scandal over journalist murder. In-Sight. Tuesday 22 May 2012. Accessed at http://www.insightcrime.org/insight-latest-news/item/2659-honduran-police-chief-removed-amid-scandal-over-journalist-murder

Hansard (2010) House of commons Hansard debates. Westminster Hall violence against women 21 January 2010 Column 146WH – 180WH. Accessed at http://www.publications.parliament.uk/pa/cm200910/cmhansrd/cm100121/halltext/100121h0001.htm#10012127000001

Hoyle C, Sanders A (2000) Police response to domestic violence. From victim choice to victim empowerment. Br J Criminol 40:14–36

Kelly L, Regan L (2001) Rape: the forgotten issue. Child and woman abuse studies project JA1 1999/DAP/161/WP

Kelly L, Lovett J, Regan L (2005) A gap or a chasm? Attrition in reported rape cases, vol 293, Research study. Home Office, London

Kilpatrick DG, Amick A, Resnick HS (1990) The impact of homicide on surviving family members. US Department of Justice. Accessed at https://docs.google.com/viewer?a=v&q=cache:nDK_olBL4cUJ:www.ncjrs.gov/pdffiles1/Digitization/130823NCJRS.pdf+&hl=en&gl=uk&pid=bl&srcid=ADGEESjGmQgrDimlOCU5Puuag10qTH1aOxm9h7tSNCUlcCeCo6rDsk4AVNT9xqwEGxm2gUahCHsuP0Wh-487_tJqnBbDI-UWtDB-TA6HUtbz1UyS77R4yFiVGUIujXKyg_DJpWlHnQGz&sig=AHIEtbR9OBcLAftJiqzqZpnhAXygsPQb4w

King K (2004) It hurts so bad: comparing grieving patterns of the families of murder victims with those of families of death row inmates. Crim Justice Policy Rev 15:193–211

Kitzinger J (2004) Framing abuse. Pluto Press, London

Kutnjak Ivkovic S (2008) A comparative study of public support for the police. International Criminal Justice Review 18:406–434

Macpherson Report (1999) The Stephen Lawrence Inquiry. Report of an Inquiry by Sir William Macpherson of Cluny. Accessed at www.archive.official-documents.co.uk/document/cm42/4262/4262.htm

Mawby RI (ed) (1999) Policing across the world: issues for the twenty-first Century. UCL Press, London

Mawby R (2001) Burglary. Willan Publishing, Cullompton

Mawby RI (2007) Public sector services and the victim of crime. In: Walklate S (ed) The handbook of victims and victimology. Willan Publishing, Cullompton, pp 209–239

Meyer S (2011) Seeking help for intimate partner violence: victims’ experiences when approaching the criminal justice system for IPV-related support and protection in an Australian Jurisdiction. Fem Criminol 6:268–290

Monckton-Smith J (2010) Relating rape and murder: narratives of sex, death and gender. Palgrave Macmillan, Hampshire

Monckton-Smith J (2012) Murder, gender and the media: narratives of dangerous love. Palgrave Macmillan, Hampshire

OVC (2001) Law enforcement (New Directions from the Field, bulletin 3). Accessed at www.ojp.usdoj.gov/ovc/new/directions/bulletins/welcome.html

Robbers M (2008) Blinded by science: the social construction of reality in forensic television shows and its effect on criminal jury trials. Crim Justice Policy Rev 19:84–102

Rosenbaum DP (1987) Coping with victimization: the effects of police intervention on victims’ psychological readjustment. Crime Delinq 33:502–519

SAPS (2005) South African Police Service Crime Information Analysis Centre: Rape statistics. Accessed at http://www.saps.gov.za/statistics/reports/crimestats/2005/_pdf/crimes/rape.pdf. Accessed 23 June 2009

Schmidt JD, Sherman LW (1993) Does arrest deter domestic violence? Am Behav Sciist 36:601–609

Skogan WG, Wycoff MA (1987) Some unexpected effects of a police service for victims. Crime Delinq 33:490–501

Sleath E, Bull R (2012) Comparing rape victim and perpetrator blaming in a police officer sample: differences between police officers with and without special training. Crim Justice Behav 39:646–665

So-Kum Tang C (2003) Factors influencing responsibility attribution to wife abuse: a study of Chinese police officers. Crim Justice Behav 30:584–601

Stark E (2007) Coercive control. The entrapment of women in everyday life. Oxford University Press, Oxford

Stephens BJ, Sinden PG (2000) Victims’ voices: domestic assault victims’ perceptions of police demeanour. J Interpers Violence 15:534–547

Sun I, Su M, Wu Y (2011) Attitudes toward police response to domestic violence: A comparison of Chinese and American College students. J Interpers Violence 26:3289–3315

Unifem (2010) Facts and figures on VAW. United Nations development fund for women. Accessed at http://tinyurl.com/q42ws. Accessed 21 Apr 2012

van Dijk J, van Kesteren J, Smit P (2008) Criminal victimisation in international perspective: key findings from the 2004-2005 ICVS and EU ICS. Boom Legal Publishers, The Hague. Accessed at http://rechten.uvt.nl/icvs/pdffiles/ICVS2004_05.pdf

van Kesteren J, Mayhew P, Nieuwbeerta P (2001) Criminal victimisation in seventeen industrialised countries: key findings from the 2000 international crime victims survey. Accessed at www.minjust.nl:8080/b_organ/wodc/reports/ ob187i.htm

Varelas N, Foley L (1998) Blacks' and whites' perceptions of interracial and intraracial date rape. J Soc Psychol 138:392–400

Websdale N (1999) Understanding domestic homicide. Northeastern University Press, Boston

Wemmers JM, Zeilstra MI (1991) Victims services in the Netherlands, vol 3, Dutch penal law and policy. Ministry of Justice, The Hague

Winkel FW (1989) Responses to criminal victimization: evaluating the impact of a police assistance programme and some social psychological characteristics. Police Stud 12(2):59–72

Zveric U (1996) Policing and attitudes towards police in countries in transition. In: Pagon M (ed) Policing in Central and Eastern Europe: comparing firsthand knowledge with experience from the West. College of Police and Security Studies, Ljubljana, pp 45–59

Victims of Corporate Crime

Hazel Croall
School of Business and Society, Glasgow Caledonian University, Glasgow, UK

Introduction

While the much heralded "rediscovery of the victim" in the latter part of the twentieth century exposed the previously hidden victimization associated with violence against women and minority ethnic groups, it did not extend its gaze to the victims of corporate crime (Croall 2007, 2010; Whyte 2007). Yet this form of crime has a huge impact, encompassing for example, the mass deaths, injuries, illnesses and long term health dangers associated with corporate neglect of safety, health and environmental considerations and the enormous financial losses suffered by individual victims, the general public and, very often, the least affluent.

This contribution will explore the factors underlying this neglect including the extent to which corporate crime victimization fails to "fit" the constructions of crime and victimization prevalent in criminology and criminal justice. It will look at issues of research before outlining the potential scope of victimization and providing selected, contemporary, examples. It will outline the relationship between victimization and structural inequalities and at the relatively limited role of criminal justice before considering alternative strategies within and beyond criminal justice.

Constructing Corporate Crime Victimization

The relative invisibility of corporate victims is strongly related to the widely acknowledged ambiguous criminal status of corporate, along with white collar, crime and to the very different nature of victimization. In very general terms, crime and criminal victimization are constructed around *individual* offenders and victims, an immediate *event* involving *intentional direct* physical and economic harm, often, if not always, accompanied by *interpersonal* contact. Corporate offences very often lack individual "bleeding" victims and involve an *impersonal* company with *no intent* to harm specific offenders. There may be no "event," rather a series of offences building up over time (Tombs and Whyte 2010), and, as Sutherland (1949), in his classic work on white collar crime recognized, many, such as consumer crime, have a "rippling effect," causing a small loss to a large number of victims, but yielding large profits. Moreover, many offences are not widely perceived as "crime" or those harmed as "crime victims," raising important issues of what activities to include.

The absence of intent, in legal terms, (*mens rea*), and an individual, guilty offender, gave rise, during the era of public welfare legislation in the nineteenth century, to the development of what is known as regulatory law. In general terms this involves the enforcement, by a vast range of agencies, of regulations covering a very wide range of corporate activities including compliance with financial, health and safety, environmental and consumer legislation. Agencies have the power to criminally prosecute, although such prosecutions are rare and penalties often perceived as lenient (Croall 2004).

This effectively creates a distinction between "real" and "quasi" crimes, which, to critical criminologists, reflects the way in which corporations have been able to influence the criminalization of their harmful acts.

From Sutherland onwards debates have raged over whether criminologists should include these activities, along with other harmful acts not subject to criminal penalties, as "crime." Some critical criminologists, arguing against an acceptance of state definitions of crime suggest using an alternative concept of "harm" (Hillyard and Tombs 2004), while others advocate using the concept of "deviance" rather than crime, which incorporates notions of a reaction by a respected audience and the imposition of sanctions, be they criminal, administrative or less formal forms of censure and reputational damage (Green and Ward 2000). It could be argued therefore that it is reasonable to include, in the category of corporate crime, corporate activities, such as financial fraud, which are unambiguously "real crime," "regulatory" crime as outlined above, activities which are "illegal but not criminal," such as some restrictive trade practices or misleading advertising, and those which are widely recognized as "lawful but awful" (Passas 2005), such as tax avoidance or the continued production of unhealthy food and other consumer products.

Like crime, the construction of criminal victimization is also based on conventional and state definitions of crime, with critical victimology stressing the importance of looking at the "victims we cannot see" (Walklate 2007), including the victims of gendered and racist violence as well as those of corporate crime, state crime and human rights abuses. Green criminologists have argued that "human centered" constructions of victimization neglect harms against non human species and of the planet itself (White 2008), and others challenge an "occidentalism" which downplays the exploitation of groups in developing countries (Hoyle and Zedner 2007). As these activities so often involve corporations they are highly relevant. Following these arguments, this Chapter will adopt a critical inclusive approach.

Research

The nature of corporate victimization also makes it difficult to "count" and directly compare with victimization from conventional crime (Croall 2007; Whyte 2007). Many victims do not regard themselves as "crime" victims and indeed are very often unaware of any harm as they cannot, for example, test for themselves the contents of food, the quality of consumer goods or the degree to which air is polluted. As outlined above, many victims suffer only a very small loss, and others blame themselves for being taken in by fraudsters or other persuasive sales practices. Responsibility for a so called "accident" may only become evident after lengthy investigations, and the long term effect of harmful substances such as asbestos or food additives only slowly recognized, and in some cases, "redefined" as criminal. When victims do choose to report problems to relevant agencies, they are not reported as "crime" but more likely as an "incident" or "complaint." The majority of crime or victimization surveys exclude corporate crime, even where victims would be aware of harm. Researchers aiming to "count" different forms of offences often therefore have to deconstruct a range of different statistical sources, such as the records of enforcement agencies, to estimate the amount which might be attributable to management (Tombs 2000).

Nevertheless, a variety of sources can be useful. Although limited to offences of which victims are aware, there have been academic studies of offences such as consumer frauds (Titus 2001; Shichor et al. 2000), and, in the US, the National White Collar Crime Center carries out a survey every five years on victimization from a range of white collar crimes including mortgage fraud, home and auto repairs, price misrepresentation and losses due to false stockbroker information, the latest in 2010 (National White Collar Crime Center 2010). Other information is found beyond traditional criminological sources. As will be seen below, some surveys, such as, in the UK, the Labor

Force Survey or those carried out by consumer organizations or Government Departments such as the Office of Fair Trading (OFT) are equivalent to victim surveys, although not restricted to "criminal" complaints. A wide range of organizations also carry out surveys on specific issues and investigative journalism provides another invaluable source by exposing aspects of victimization. Much of this is easier, via the Internet, to gather together than hitherto. While relatively unsystematic and difficult to directly compare with "conventional" crime, research tends to confirm the general assumption that victimization from corporate crime is widespread and likely to exceed other equivalent forms of crime. A victim survey which did include questions on health and safety at work, consumer fraud and offences by landlords found higher levels than for other forms of crime (Pearce 1992), and the above mentioned 2010 National Public Survey on White Collar Crime found that 24 % of households and 17 % of individuals reported at least one experience of victimization from the crimes included within the previous year (National White Collar Crime Center 2010). Other examples are outlined below.

Dimensions of Corporate Crime Victimization

While this chapter cannot provide extensive detail about the vast impact of corporate victimization (Croall 2007, 2010; Whyte 2007), the following, albeit selective, examples illustrate some of its main forms and suggest its potential impact.

Corporate Financial Frauds

Corporate financial fraud has a massive impact. It affects millions directly, through loss of investments and savings, the "misselling" of pensions, mortgage endowment policies and other financial products such as payment protection insurance. Indirectly it affects us all, through for example, taxes not paid or higher prices as a result of restrictive trade practices. For example,

- The collapse of the *Enron* corporation as a result of widespread frauds led to thousands of jobs and millions of dollars of retirement savings being lost and its wider impact included higher interest rates; a reduction in capital for investment; associated physical and mental distress and an erosion of trust in major institutions (Friedrichs 2004).
- *Taxation* involves a very fine, and constantly shifting, line between legal tax avoidance and illegal *tax evasion* which, while often seen as "victimless," impacts on state expenditure. While notoriously difficult to "count," it has been estimated that in 2005 tax fraud in the UK, *excluding* income tax fraud, involved losses of nearly £2 billion to businesses; £2.75 billion in charity, consumer, investment and pension frauds and £6.434 billion in frauds on public bodies such as the National Health Service (Levi and Burrows 2008).

To this could be added the massive effects, including job losses, unemployment, insecurity, lower pensions, lower interest rates, loss of savings and recession of the global financial crisis, which, while leading to few prosecutions, has been likened to a "bank robbery," was accompanied by considerable illegality and led to considerable reputational damage being suffered by banks and other financial institutions (McGurrin and Friedrichs 2010). While often seen as primarily economic, corporate fraud also has severe emotional impacts, particularly where loss of financial security or feelings of self blame are involved.

Safety in the Workplace

It has long been recognized that corporate crime "kills," readily illustrated in the numbers of deaths, illnesses and injuries associated with the workplace, so often caused by failures to comply with, often routine, safety regulations. These may involve mass or individual deaths and injuries in one off events, often misleadingly described as "accidents" or the long term impact of working with unsafe substances such as asbestos or chemicals. Thus:

- The mass deaths attributable to neglect of corporate safety regulations include the 500 deaths associated with the faulty design of the Ford Pinto car in the United States; the 26 Canadian miners killed in 1992 in the Westray mine in Nova Scotia; the spate of transport "disasters" in the UK, which included the 192 drowned in the sinking of the ferry *The Herald of Free Enterprise* near Zeebrugge in 1987 and a series of fatal rail crashes (Whyte 2007) and the 167 workers killed in the world's worst off-shore oil disaster on the Piper Alpha Oil rig in Scotland in1988 (Ross and Croall 2010). While these tragedies did not all lead to prosecutions, subsequent inquiries revealed a familiar story of managerial incompetence and disregard of regulations.

- *Individual deaths* in the workplace, across a range of industries and occupations, have been estimated in the UK to far exceed deaths from homicide (Tombs 2010).
- *Occupationally caused diseases* include those associated with asbestos – which, in Britain alone, account for around 4,000 deaths per annum, expected to peak between 2011 and 2015 (Tombs 2010).
- *Injuries from work* affect around 30,000 UK workers and 15,000 members of the public with Labor Force surveys, which can be likened to victim surveys, estimating that as many as 1,000 per 100,000 workers were injured in 2006–2007, a figure exceeding that of "real crime" (Tombs 2010).
- The 2006/2007 Labor Force survey also recorded that 2.2 million respondents suffered from an *illness* which they associated with work.

As well as the direct effect on victims and their families, these offences can have spillover effects on the wider community – especially where they lead to places of employment being closed down.

Crimes Against Consumers

A vast number of consumer goods and services have been associated with fraud, misrepresentation and health and safety issues (Croall 2009). Individual consumers can be poisoned by food, killed or made very ill by pharmaceutical products and be threatened by unsafe products. They can be directly defrauded by or misled by misrepresentations about the quantity, quality or origins of pre packaged goods and misleading advertisements. More indirectly their long term health is threatened by, often untested, chemicals and additives in foods, cosmetic, cleaning and other products, and by the continued production and sale of unhealthy processed foods which yield greater profits than their healthy, unprocessed equivalents (Croall 2012). For example:

- Pharmaceutical products such as the drug Thalidomide and the Dalkon Shield contraceptive caused mass deaths worldwide.
- Unsafe consumer products can be imported from countries with lower safety standards. It has been estimated that unsafe goods, particularly toys and fireworks, entering the UK through one port alone, cause or contribute to 95,000 injuries, 100 fires and 3 deaths per annum (Croall 2010).
- Food poisoning outbreaks, generally associated with lack of attention to hygiene regulations, include outbreaks of E.coli 0157 such as that in Lanarkshire Scotland where 21 pensioners died and in Walkerton, Ontario, which killed 7 (Croall 2010). Many more become ill and other food borne diseases such as campylobacter or salmonella are also linked to poor hygiene – the chocolate firm Cadbury's was prosecuted in 2007 following illnesses to babies and children caused by Salmonella Montevideo (Tombs and Whyte 2010).
- Food adulteration can also kill, seen most recently in the deaths of 6 and illnesses of 300,000 Chinese babies after consuming milk contaminated with melamine (Croall 2012).
- *Consumer fraud* is widespread with frauds in relation to car servicing and repair having been estimated to cost British consumers around £4 billion each year (Croall 2009). A 2008 survey, again equivalent to a victim survey, carried out by Britain's Office of Fair Trading (OFT), of consumer "detriment," which involves instances where customers suffer "partly or wholly as a result of the organization accidentally or deliberately

treating the customer unfairly" estimated its cost to consumers at £6.6 billion annually with around one third reporting at least one problem (Croall 2010). While many losses are small, (only 4 % involved sums of over £1,000), it should be borne in mind that surveys like this can only cover harms of which consumers are aware and perceive to be "unfair." The 2010 National White Collar Crime Center's survey found that 28.1 % of US respondents reported experiences of price misrepresentation, and 22.3 % reported cases of unnecessary repairs to motor vehicles (National White Collar Crime Center 2010). This leaves out a wide range of misleading marketing practices, such as deceptive packaging and misleading descriptions, including, for example descriptions of food as "healthy," which are widespread but not widely perceived as deceptive.

Environmental Crime

Environmental crime also has an enormous impact. Individuals can be killed and made ill by pollution and a host of toxic emissions, which also affect communities through loss of amenities, in some cases threaten the survival of habitats and entire species and endanger the future of the planet (White 2008). Food production can also involve the inhumane treatment of animals as seen in recurrent exposees of the conditions in intensive dairy or chicken farms (Croall 2012). Intellectual property, knowledge and natural and human resources in developing countries can be plundered by international corporations in what is known as bio-piracy (South 2007) and other forms of exploitation which underlines the global impact of victimization. For example:

- The world's worst industrial "accident" took place in the Indian town of Bhopal in 1984, where an explosion of poisonous gas led to a death toll of over 20,000 and at least 200,000 injuries and illnesses.
- In 2005, in the wake of the tsunami, hundreds of barrels containing radioactive waste disposed of by European companies were found on the shores of Somalia, causing infections, skin diseases and as yet un-measurable cancers (White 2008).
- Globally it is estimated that up to 800,000 premature deaths can be attributed to ambient air pollution, and UK Health Department estimates attribute at least 24,000 premature deaths each year to air pollution. In the North West of England air pollution and the Sellafield nuclear power plant have been linked to high levels of cancer (Tombs and Whyte 2010).

These selective examples illustrate several features of corporate victimization, although it is important to recognize its heterogeneity. Despite often being assumed to be "merely" economic, and to have a diffuse and indirect effect, many forms of corporate crime, which can be described as "corporate violence" (Tombs 2010), have an immediate and direct physical impact on victims, some fatal, some causing long term illness or debilitating injuries. Its direct economic effects on individual victims can also be devastating, involving loss of savings, employment and financial security. Like other forms of violent crime, it can have a considerable emotional impact ranging from the need to adjust to life changing losses or, particularly in cases of fraud, feeling that trust has been violated and blaming oneself. Like other groups of crime victims, corporate crime victims can be constructed as "non deserving" or having "asked for it" as in cases where investments have been willingly made, or consumers have chosen to buy cheap but unsafe goods, cosmetics, unhealthy junk food or environmentally damaging products (Croall 2010). This reflects the doctrine of "*caveat emptor*," let the buyer beware, a long accepted principle in consumer law.

While some offences may appear to have relatively trivial effects they may, like minor consumer frauds, occasion irritation and frustration and time taken to complain or seek compensation. In addition, while individual losses may be small, they can have a cumulative effect, which can lead to a loss of trust in business, and all citizens can be repeat and multiple victims – many suffer, for example, from pollution *and* the "misselling" of financial and consumer products, and it is very difficult to avoid harmful chemicals in household products or foods.

Other forms of corporate crime, sometimes regarded as "victimless," have a diffuse and not inconsiderable effect on tertiary victims such as the general public. Tax fraud is the clearest example of this but also significant are the public health costs of pollution and obesity and other problems created by the mass production and marketing of unhealthy junk foods with widely recognized dangers. These are akin to other corporate harms in that, deaths from obesity and diet related problems are "avoidable," with food corporations consistently and fiercely resisting attempts at regulation (Croall 2012). Victimization also has serious global implications. Global corporate crime (Croall 2005a) very often involves corporations taking advantage of "regulatory voids" to locate production in countries with less stringent safety or employment regulations and utilize tax havens. Many forms of forms of production exploit poorly paid, migrant or child labor – to a point where the term "new slavery" is widely used (Lawrence 2008).

Who Are the Victims?

While it has been argued that we are *all* victims of corporate crime, and victims are often described in aggregate terms such as "consumers," workers, or citizens, it is important to ask *which* of these groups are more likely to be at risk from *whic*h forms of corporate crime (Croall 2009). The depiction of one off events as "accidents" masks the greater vulnerability of, for example, workers, to deaths and illnesses caused by inattention to safety regulations and the emphasis on the diffuse nature of victimization masks the greater risks posed to those who have less ability to protect themselves. The following section also illustrates the significance of interrelated inequalities such as gender, race and age.

Global inequalities are well illustrated in examples in which corporations, as outlined above, legally exploit regulatory voids in relation to taxation, employment, safety, product quality or environmental regulations (Croall 2005a). The Bhopal plant would not have complied with regulations in the United States, Thalidomide and the Dalkon Shield were widely sold in the third world after their dangers were known in the West and toxic and e-waste is "dumped" by richer countries in poorer ones, and its impact is most severely felt by the poorest within these countries (White 2008). Major food corporations, most of whom are located in tax havens, extract much needed revenue from developing nations and have also been linked with the "new slavery" associated with poor conditions and low wages in South America (Lawrence 2008).

Within countries many offences prey more particularly on the poor although the affluent, by having more to invest, are more at risk from investment frauds. Low income groups, as well as losing potential benefits through tax evasion, have also been identified by Britain's National Consumer Council, along with those living in a deprived area and those with "skills difficulties," as vulnerable consumers (Croall 2009) and are also said to have less access to healthier food, with some living in so called "food deserts" (Croall 2012). Lower level employees, with some exceptions, are more likely to be seriously and fatally injured at work, with industries relying more heavily on casual and non unionized labor such as the building industry and agriculture having higher amounts of injuries and deaths (Tombs 2010). In the UK, some of the worst cases of chemical pollution have a disproportionate effect on lower class communities (Tombs and Whyte 2010).

A combination of *class and race* have been identified as significant in victimization from environmental crime with, in the US for example, more hazardous chemical plants being located nearer lower class communities and communities of color, black communities being most at risk from health hazards, and chemical accidents occurring more frequently near African-American and low income census tracts (Lynch and Stretesky 2001). Across the globe, migrant workers are particularly vulnerable to exploitation, seen dramatically in, for example, the food industry with "sweatshop" conditions and appalling living conditions being noted in England and Spain as well as in developing countries (Lawrence 2008).

G*ender* is a significant factor although many analyses of corporate crime have been "gender blind" (Snider1996). "Little old ladies" have been seen as archetypal victims of fraud; historically were more vulnerable to investment frauds and their assumed technical and financial incompetence makes them more likely targets of, for example, investment frauds and car sales and service frauds (Croall 2009, 2010). Goods and services marketed specifically for women, often playing on images of femininity, such as cosmetics and household products, contain allergenic and potentially carcinogenic "chemical cocktails," diets, which rarely work, are widely marketed on dubious evidence and botox and silicone breast implants have widely reported dangers (Croall 2009). Men are more at risk from other activities – particularly from workplace fatalities in so called "man's work" (Whyte 2007).

Age inequalities are variously related to victimization. Following the life cycle for example:

- The *very young, babies and children* are particularly vulnerable to offences involving illness such as food poisoning and products containing chemicals such as the Chinese melamine in milk case (Croall 2012), and cheap toys, often manufactured abroad, are amongst the unsafe goods mentioned above. Some food additives, particularly food colorings, have been associated with hyperactivity in children (Lawrence 2008). Globally, child labor is exploited in manufacturing to drive down production costs.
- *Teenagers* constitute a major consumer market making them vulnerable to a host of dubious marketing practices – mobile phone companies have, for example, been widely criticized for misleading indications about phone contracts and fashion goods and clothes, music, and DVD's are subject to faking (Croall 2009). Teenagers also enjoy "risky" and dangerous life styles, rendering them vulnerable to safety offences in, for example, night clubs or adventure centers.
- *Older people* are often seen as particularly vulnerable. Physically, they suffer more from some forms of food poisoning, are more vulnerable to physical and financial abuse in care homes and a variety of, often dangerous or ineffective, pharmaceutical and other products, such as "anti wrinkle cream" which target this market (Croall 2009). Pension frauds play on fears of economic insecurity and dependency in old age. Nonetheless, age is strongly related to class and the vulnerability of the elderly can be overstated – more affluent elderly people may indeed have sufficient knowledge and experience to avoid some obvious frauds (Titus 2001).

These, interrelated, inequalities in victimization reflect wider structural inequalities which form the context for production, marketing and other corporate strategies. Goods and services, for example, are targeted at specific income groups and reflect assumptions about masculinity, femininity, and, often commercially encouraged, consumption patterns. This is seen in, for example, women's greater vulnerability to unsafe household and "beauty" products or children and teenagers' vulnerability to offences involving mobile phones or toys. Children become brand aware at a very early age and parents may be pressured to purchase cheap goods or unhealthy junk foods (Croall 2009). The "race for the bottom" in terms of production costs underlies the exploitation of the least powerful groups whether in Western societies or in developing countries who have little power to resist. While the more affluent are victims, they are protected to some extent by their cultural and social capital which means that they are more aware of the dangers of cheap products, unhealthy food, more able to resist unsafe working conditions or avoid polluting industries being located nearby (Croall 2007).

Issues for Criminal Justice and Beyond

The move of the victim to "centre stage" of criminal justice did not include corporate crime, whose victims do not feature in victim compensation schemes or services to victims in court. This is largely due to the different constructions of crime and victimization outlined above and victims' lack of awareness or failure to report many offences. Many corporate offenders are not part of criminal justice proceedings as so many are subject to the "quasi criminal" proceedings of regulatory law where prosecutions are rare and other harms are not subject to criminal proceedings. Inevitably therefore the criminal justice process plays a limited role. At the same time however, when it is used, the extent to which it meets victims' interests can be questioned and alternatives suggested. Moreover, corporate crime victimization raises issues of social as well as criminal justice, and victims' interests are represented by groups such as those involved in consumer, trade union and environmental issues – for whom criminal law and justice are only one aspect of broader policy concerns.

This is illustrated by the range of groups representing victims' interests. Some perform functions equivalent to victim support groups for conventional crimes, providing emotional and practical support for victims and expert advice about how to deal with inquests, investigations and seeking compensation. Many of these emerged in the wake of major tragedies such as the sinking of the Herald of Free Enterprise, which saw a growth in relevant legal and other expertise. It was seen as important to recognize the need for victims to explain "what happened" and to attribute responsibility (Wells 1993). As well as individual cases, some of these groups, such as the former Centre for Corporate Accountability and Trade Union based organizations also saw victims' interests as being served by campaigning for legal reforms, in particular for changes to the criminal law to remove obstacles to the conviction of companies and individual directors for corporate manslaughter or homicide (Croall 2007; Tombs 2010). Broader based groups such as consumer or environmental groups mount a variety of campaigns which research and expose harms, albeit that they are more concerned with strengthening relevant regulations to prevent offences than with individual cases or discussions of *criminal* law and justice.

Some of these groups have been particularly critical of the scarcity of prosecutions under regulatory law, both for individual cases, where victims may express a desire to see "justice" being done by having those responsible publicly tried and punished, and the deterrent value of a prosecution could prevent further victimization. Against this, it can be argued that compliance strategies, based on persuading and negotiating with organizations, are more "cost effective" and preventative (Croall 2004).

Sentences are also widely perceived as too lenient and as not reflecting elements of victimization such as the harm done. The main sentence for many corporate offences is monetary, reflecting the absence of a guilty individual to imprison or rehabilitate (Croall 2005b). Fines are often constrained by their potential spillover effects of higher prices and potential job losses and the impact on the wider community should the company have to close. Regulatory law, in the absence of intent to harm a specific victim, tends to focus on the breach or neglect of regulations rather than on the "harm done" and, even where deaths have occurred, which are an aggravating factor, sentences can be low. In one Scottish case, known as Stockline, where a gas explosion, subsequently attributed to persistent maintenance failures, killed 9, the two companies concerned were fined £200,000, publicly criticized as amounting to a "mere" £44,000 per life (*Edinburgh Evening News*, 2007). The fine was limited by the need to keep the company in business (Ross and Croall 2010). Higher fines are envisaged under the 2007 UK Corporate Manslaughter and Corporate Homicide Act, which covers death and serious injury and which recognizes the need for heavier sentences where serious harm has been done, but prosecutions under this Act are likely to be rare.

While larger fines may express greater disapproval and go some way to meeting victims' concerns, they do little directly for victims and there have been recurrent arguments that reparative and restorative principles can be appropriate for corporate offenders particularly in view of a corporation's ability to pay and the expertise which it may have at its disposal (Braithwaite and Geis 1982; Croall 2005a). More use could be made of individual compensation orders and where offences impact on communities or the general public, Community Sentences could require offending corporations to "pay back" through, for example, the provision of leisure amenities. Other suggestions include requiring companies to fund research into disaster prevention or other relevant issues (Croall 2005b). Restorative justice has the potential to work particularly well with companies, and Braithwaite (1995) reports a case in which senior executives of an insurance company, convicted for selling deceptive policies to Aboriginal communities, were required to visit the communities and negotiate settlements. Provisions for these kinds of options have been made in parts of Australia for consumer and environmental offences (Law Reform Commission, New South Wales 2003). In the UK, orders containing restorative elements and payments to victims were recommended by the McRory (2006) review, adopted in the Regulatory Enforcement and Sanctions Act of 2008, although initially only as administrative penalties as a desire to avoid criminal prosecution was clearly stated (Ross and Croall 2010). To date however there has been little research on the use and impact of these options although it can be argued that they have a considerable potential.

The role of criminal justice is however inevitably limited, and some forms of private justice also have serious shortcomings. While individual victims can pursue private redress, many, as seen above, are unaware of any harm or of their right to seek compensation and, as so many victims are amongst the least affluent, many do not have sufficient resources to pursue cases against corporations who can utilize their superior resources to dispute liability. This means that victims' interests in preventing future victimization may best be met outside the formal arenas of justice, in the many wider campaigns around social and public policies. These include for example, the role played by Trade Union, environmental or consumer groups in highlighting and conducting research which exposes corporate harms and campaigns for fuller and more comprehensible information to assist consumers, such as those advocated by the UK consumer's organization *Which*? While worthwhile, these are also inevitably limited particularly in view of the structural aspects of victimization which limit the extent to which different groups, however well informed, are able to exercise choices to, for example, not work in unsafe factories, or avoid purchasing cheap food or consumer goods. Ultimately therefore corporate harms are linked to wider issues of social justice.

Concluding Comments

It can be seen from the above therefore, that, when the victimological gaze is extended to corporate crime, the many victims whom we do not see include some of the least advantaged within Western societies and across the globe. Indeed victimization from corporate crime most probably vastly exceeds that from conventional crime. The role of criminology, criminal law and justice has inevitably been limited by a number of factors associated with the failure of corporate crime and victimization to "fit" conventional constructions of victimization. Nonetheless, some improvements could be suggested.

It is often argued, for example, that the different nature of victimization means that corporate crime is not appropriate for the widely used crime survey method of "counting" victimization. This is indeed the case for many offences, but, as seen above, research equivalent to victim surveys has been carried out in respect of those crimes of which victims are aware, and victim surveys could easily be adapted to include, for example, some offences which would perform the valuable function of exposing their extent and impact and comparing this to conventional crime. The exclusion of these harms is therefore as much ideological as it is related to technical difficulties of measurement, as crime surveys are designed around perceived governmental priorities and inevitably include only those activities which governments wish to know about – thus restricting their scope to conventionally defined crime and anti-social behavior.

Conventional constructions are also reflected in criminal law and criminal justice and, as seen above, the ambivalent criminal status of many corporate crimes raises complex issues about the extent to which the interests of corporate victims can best be met by criminal law and criminalization. While few dispute that there are limitations, and that compliance strategies can be cost effective, it can also be argued that the continued existence of so many corporate offences indicates the failure of current strategies to act as a deterrent. Moreover, the criminal law remains the sanction which best expresses public disapproval of any activities (Croall 2004) and continued corporate resistance to further regulations and criminalization also indicates that, for some at least, a fear of the stigma of the criminal sanction and reputational damage remain important.

Criminal law could also be considerably strengthened. While invoking victims' interests has been negatively associated with calls for more punitive treatment of conventional offenders, in respect of corporate crime, some tougher and more victim centered sanctions could be viewed positively. As outlined above for example, victims and their representatives stress the need to see the "guilty" publicly prosecuted and sentences could better reflect the harm done – all of which may be preventative by providing a greater deterrent. More innovative use of restorative and reparative measures, whether criminal or administrative, could go some way to including corporations in concepts such as "community payback."

The likelihood of such measures being introduced is however reduced by the current preference of many Governments for neo liberal deregulatory policies which, in a bid to reduce "red tape" and public expenditure, are likely to reduce resources available to regulatory agencies and reduce, rather than expand, the scope of criminal law. This underlines the importance of social as well as criminal justice in relation to corporate harms and researching, exposing and campaigning to prevent these harms is the province of a wide range of social movements for whom exposing victimization, calling for more stringent regulations, be they criminal or otherwise, is a vital aspect of their work.

Related Entries

- ▶ Corporate Crime Control
- ▶ Corporate Liability
- ▶ Fraud Victimization
- ▶ Victims and Restorative Justice
- ▶ Victims of State Crime

Recommended Reading and References

Braithwaite J (1995) Corporate crime and republican criminological praxis. In: Pearce F, Snider L (eds) Corporate crime: contemporary debates. University of Toronto Press, Toronto

Braithwaite J, Geis G (1982) On theory and action for corporate crime control. Crime Delinq 28(2):292–314

Croall H (2004) Combatting financial crime: regulatory versus crime control approaches. J Financ Crime 11(1):45

Croall H (2005a) Transnational white collar crime. In: Sheptycki J, Wardak A (eds) Transnational and comparative criminology. Glasshouse, London, pp 227–246

Croall H (2005b) Penalties for corporate homicide, annex to Scottish executive expert group on corporate homicide. http://www.scotland.gov.uk/Publications/2005/11/14133559/36003. Accessed 12 June 2013

Croall H (2007) Victims of white collar and corporate crime. In: Davies P, Francis P, Greer C (eds) Victims, crime and society. Sage, London

Croall H (2009) White collar crime, consumers and victimization. Crime Law Soc Change 51:127–146

Croall H (2010) Economic crime and victimology: a critical appraisal. Int J Victimol 8(2). http://www.jidv.com/njidv/index.php/archives/150-jidv-23/425-economic-crime-and-victimology-a-critical-appraisal

Croall H (2012) Food crime: a green criminology perspective. In: South N, Brisman A (eds) Handbook of green criminology. Routledge, London (forthcoming)

Edinburgh Evening News (2007) Stockline blast firms fined £44k a life. http://www.scotsman.com/news/stockline-blast-firms-fined-163-44k-a-life-1-1329928. Accessed 28 Aug 2007

Friedrichs DO (2004) Enron et al.: paradigmatic white collar crime cases for the new century. Crit Criminol 12(2):113–132

Green P, Ward T (2000) State crime, human rights, and the limits of criminology. Soc Justice 27(1):101

Hillyard P, Tombs S (2004) Beyond criminology? In: Hillyard P, Pantazis C, Tombs S, Gordon D (eds) Beyond criminology: taking harm seriously. Pluto Press, London

Hoyle C, Zedner L (2007) Victims, victimization and criminal justice. In: Maguire M, Morgan R, Reiner R (eds) The Oxford handbook of criminology, 4th edn. Oxford University Press, Oxford, pp 461–494

Law Reform Commission, New South Wales (2003) Sentencing corporate offenders, law reform commission of New South Wales Report 102. www.lawlink.nsw.gov.au/lrc.nsf/pages/r102. Accessed 12 June 2013

Lawrence F (2008) Eat your heart out: why the food business is bad for the planet and your health. Penguin, London

Levi M, Burrows J (2008) Measuring the impact of fraud in the UK: a conceptual and empirical journey. Br J Criminol 48(3):293–318

Lynch MJ, Stretesky P (2001) Toxic crimes: examining corporate victimization of the general public employing medical and epidemiological evidence. Crit Criminol 10(3):153–172

MacRory R (2006) Making sanctions effective final report. HM Treasury, HMSO, London

McGurrin D, Friedrichs DO (2010) Victims of economic crime- on a grand scale. Int J Victimol 8(2). http://www.jidv.com/njidv/images/pdf/JIDV23/4_jidv23_word_pdf_friedrichs.pdf

National White Collar Crime Center (2010) National public survey on white collar crime national white collar crime center. http://www.nw3c.org/Home#. Accessed 12 June 2013

Passas N (2005) Lawful but awful: 'legal corporate crimes'. J Socio Econ 34:771–786

Pearce F (1992) The contribution of "left realism" to the study of commercial crime. In: McLean B, Lowman J (eds) Realist criminology: crime control and policing in the 1990's. University of Toronto Press, Toronto, pp 313–335

Ross J, Croall H (2010) Corporate crime in Scotland. In: Croall H, Mooney G, Munro M (eds) Criminal justice in Scotland. Willan, Cullompton

Shichor E, Sechrest D, Doocy J (2000) Victims of investment fraud. In: Pontell H, Shicho D (eds) Contemporary issues in crime and criminal justice: essays in honor of Gilbert Geis'. Prentice Hall, Englewood Cliffs

Snider L (1996) Directions for social change and political action. In: Szockyj E, Fox JG (eds) Corporate victimisation of women. Northeastern University Press, Boston

South N (2007) The 'corporate colonization of nature': bio-prospecting, bio-piracy and the development of green criminology. In: Beirne P, South N (eds) Issues in green criminology. Willan, Cullompton, pp 230–247

Sutherland E (1949) White collar crime. Holt/Rinehart and Winston, New York

Titus RM (2001) Personal fraud and its victims. In: Shover N, Wright JP (eds) Crimes of privilege: readings in white-collar crime. Oxford University Press, New York, pp 57–66

Tombs S (2000) Official statistics and hidden crimes: researching health and safety crimes. In: Jupp V, Davies P, Francis P (eds) Doing criminological research. Sage, London, pp 64–81

Tombs S (2010) Corporate violence and harm. In: Brookman F, Maguire M, Pierpoint H, Bennett T (eds) Handbook of crime. Willan, Cullompton, pp 884–903

Tombs S, Whyte D (2010) Reflections upon the limits of a concept: "victims" and corporate crime. Int J Victimol 8(2). http://www.jidv.com/njidv/index.php/archives/150-jidv-23/425-economic-crime-and-victimology-a-critical-appraisal

Walklate S (2007) Imagining the victim of crime. McGraw Hill/Open University Press, Maidenhead

Wells C (1993) Negotiating tragedy: law and disasters. Sweet and Maxwell, London

White R (2008) Crimes against nature: environmental criminology and ecological justice. Willan, Cullompton/Portland

Whyte D (2007) Victims of corporate crime. In: Walklate S (ed) Handbook of victims and victimology. Willan, Cullompton

Victims of State Crime

David Kauzlarich
Department of Sociology and Criminal Justice Studies, Southern Illinois University Edwardsville, Edwardsville, IL, USA

Overview

This entry reviews the definitions, measurement, extent, and nature of victims of state crimes as well as relevant justice system policies and resistance to victimization by the state. Current theorizing and research on victims of state crime are discussed as are public and criminal justice policy initiatives. Propositions on the victimology of state crime as well as current research findings suggest that the study of state crime victimization requires an appreciation of creative and innovative social and criminal justice policy changes that attend to the diverse needs of those victimized by the state.

As with definitions of state crime itself, there are many competing ways to conceptualize those victimized by the state. There is obviously a vast array of harm and victimization which result from state crime, as well as multiple types of actors, agents, and organizations which may be directly or indirectly involved in this form of organizational crime. The crimes can be against individuals, property, communities, groups, and states and may cause a range of psychological, physical, community, and emotional trauma. Despite differing ways to classify victims of state crime, most of the scholarly literature on this subject within criminology fit under the following definition: those individuals or groups of individuals who have experienced economic, cultural, or physical harm, pain, exclusion, or exploitation because of tacit or explicit state actions or policies which violate law or generally defined human rights (Kauzlarich et al. 2001). The range of victims of the state is very broad, and can include, but is not limited to, targets of genocide; ethnic cleansing; human and political rights violations; war crimes; natural disasters caused by state action or inaction; those discriminated against by any state agency, especially within justice systems; and, more broadly, those who suffer the worst pain and harm because of larger structural state systems that cause or correlate with unnecessary poverty, illness, and other troubles. Each form of victimization can be broken down into different types of injury. For example, Turkovic (2002, p. 206) has compiled an inventory of the harms experienced by victims of war. These include those killed, unemployed, wounded, forced to conscript, tortured, missing, confined, psychologically distressed (e.g., insomnia, depression), and sexually abused as direct personal harms, along with property victimization. To this we could add the destruction of schools, hospitals, water and food supplies, and so on.

There is some debate in criminology over how to define victims of state crime, especially as it relates to the implication of larger state policies, both implicit and explicit. No one would debate that citizens who are indiscriminately killed or raped in the course of a military invasion are victims of a state. Likewise, there would be little debate on the criminality of the genocide committed some years ago in Rwanda or Cambodia. However, differences do exist on whether states that allow homelessness, suffering because of treatable medical illness, hunger, and economic marginalization are really state crimes or simply social problems. On the one hand, those who favor legalistic definitions of state crimes are less likely to view capitalist or exclusionary activities themselves as criminal, while radical scholars are more likely to speak of victims of the economy in the same way others speak of victims of human rights violations. Further, extensive research has been conducted by Faust and Carlson (2011), Faust and Kauzlarich (2008), and Green (2009) on environmental and natural disasters that are linked to harmful actions or inactions by the state. Interestingly, there may be widespread disagreement among non-scholars in the use of the term victim as applied to natural disasters and indeed among victims themselves. This is very rare in traditional street crime victimization save for so-called public order crimes like prostitution, gambling, and drug use.

On an epistemological level, victimology in general must take into consideration the sources and values upon which it assumes candidacy for classification as a victim. A key controversy is about whether victims and their constituency, both of a state and non-state nature, should also define their experiences as victim-like or whether other standards should be used that may identify injustices and those harmed with or without those subject agreement. Any attempt at a universal definition of state crime victimization is questionable because of the subjective and objective elements of social life in general, let alone injury to a person or group. Using legal standards is not objective because laws and political processes are interest-laden, while more obviously subjective interpretations by people who may be candidates for classification as state crime victims may not accept the label of victim. Consider what Strobl (2010) offers as a constructionist concept of victims: (1) an identifiable event, (2) negative evaluation by the victim, (3) an uncontrollable event, (4) attributable to a personal or social offender, and (5) a violation of a socially shared norm. The second and last points are especially questionable given that the concepts of social norms and cognizance are subject to hegemonic effects in which the attitudes and beliefs of those in power are packaged to be understood as acceptable by the masses, whose interests are not served by elite definitions of social reality. For example, the recent global recession has links to state crimes of omission in terms of the underregulation of financial industries, yet many citizens who have lost jobs or their homes might be unlikely to view such negligence as criminal, or themselves as victims of crime. This debate is similar to what is seen in social problems research in sociology, wherein some scholars believe that in order for a social problem to exist, it should be recognized as such by a majority of people in the citizenry. From a radical approach, such a definition, as that with definitions of state crime victimization that depend on public cognizance, is unacceptable because of the ideological, material, and political control those in power have over those who are less powerful.

Core Propositions

Six propositions on the victimology of state crime were developed by Kauzlarich et al. (2001) and have guided several critical analyses of the phenomenon (see, e.g., Westervelt and Cook's use of the propositions):

1. Victims of state crime tend to be among the least socially powerful actors.
2. Victimizers generally fail to recognize and understand the nature, extent, and harmfulness of institutional policies. If suffering and harm are acknowledged, it is often neutralized within the context of a sense of "entitlement."
3. Victims of state crime are often blamed for their suffering.
4. Victims of state crime must generally rely on its victimizer, an associated institution, or civil social movements for redress.
5. Victims of state crime are easy targets for repeated victimization by the same organization or institution.
6. The unethical, immoral, or illegal state policies and practices, while committed by individuals and groups of individuals, are manifestations of the attempt to achieve organizational, bureaucratic, or institutional goals.

Indeed, unlike most victims of traditional street crime, those victimized by the state are likely to have even less power to address the injury. Victims of war crimes and genocide, for example, are often in conditions of chaos and most everyday or normative social processes are absent. In these conditions, it is not feasible for victims to contact policing or regulatory bodies for help or redress. Further, victims of state crime are usually harmed by a relatively large and organized institution or group, not by lone offenders as is the case with much traditional street crime. Depending on the organization of the state doing the offending, state crime victims may also have little to no way of finding justice. Citizens of states who are victimized by their own government will often not have any opportunity for redress in weak states, whereas in stronger states they may be able to work through a different arm of the government for a remedy. For example, Chinese citizens have very little chance of finding remedy for human rights violations through their domestic justice system, whereas in the United States, there are examples of victims of nuclear weapon experimentation who have, albeit over decades, received a measure of compensation for the harms. Others harmed by states may have to seek international help in addressing state crimes committed against them, such as systems developed under the International Criminal Court (ICC) or ad hoc international criminal tribunals. Some state crime victims may also be less likely to trace their harm and suffering to the particular offending organization. Unlike a traditional street crime such as robbery or rape where the victim is directly attacked and immediately experiences some or all elements of victimization, the suffering from state crimes and the identification of the offender, like with many white-collar crimes, may come years or decades later. For example, illegal government pollution of the environment or harmful chemicals and agents to which military personnel are exposed may cause health problems long after the initial exposure. Further, layers of bureaucracy and secrecy may hinder the pursuit of the key offenders. It is important to keep in mind, however, that some state crimes are exactly like traditional street crimes but differ only by the context. Sexual and physical assault, theft, robbery, and several other crimes do occur in times of war, genocide, and ethnic cleansing and are committed by state affiliated groups as a way to exercise power and control over a population.

It is not uncommon that revictimization, or compounded victimization, occurs in all kinds of crime situations, but with state crime it can be more profound. Consider only the emotional and social injury caused by illegal military aggression, genocide, and war. Whole families and communities may be killed or injured, and such demolition compounds the initial criminal events, possibly for generations. Rebuilding infrastructure, whether for public health, schooling, roads, or water sanitation, forces resources to be placed into fundamental human needs building rather than improving what once existed. Victims, then, are considerably vulnerable and often public sympathy, if there was any in the first place, relaxes if concerted organizational efforts are not sustained. In the United States, we have seen this take place in myriad circumstances, especially for the victims of Hurricane Katrina, who some even blamed for their loss of property and community. In the United States invasion of Iraq, some several hundred thousand people died by some accounts, and many of these appear to be forgotten in the West, if they were even recognized as victims in the first place. Drawing from studies of genocide and war in Bosnia, Ajdukovic (2006) outlines a host of repairs and initiatives that should follow such conflicts including provision of mental and physical health services, securing social support for families and communities, promotion of nondiscrimination in social, political, legal, and cultural contexts, training in conflict resolution, encouraging tolerance and diversity, and a range of other small- and large-scale social changes.

Measurement

Victims may or may not be recognized as such by formal legal, regulatory, or human rights commissions. With crimes of the state, this would necessarily include control mechanisms at the domestic and international levels. Consider, for instance, the victims of the Rwandan genocide at the state level, where the recognition is of Tutsis as victims. Yet, many Hutu were also victimized. Likewise, international institutions of control remain selective who they define and label as a victim. This has serious ramifications for victim recourse as well as victim healing and accurate accounts of facts and subsequent history of their victimization. Additionally, it must not be overlooked that in many cases the victim can be rightly labeled as victim/offender. Here again, looking at Rwanda, testimony of various Hutu highlight how they view themselves as perpetrator but victims due to the fact that they were faced with being killed or killing Tutsi. In many cases, the processes of labeling or lack thereof can result in new forms of victimization and/or revictimization. The complexities and multiple layers of seeing oneself as a victim, accepting such a label, being given a victims label by others, exclusion or inclusion as a victim in mass atrocity settings facilitates another issue: How can we measure and know the totality of state crime victimization (Rothe and Kauzlarich forthcoming a)?

A key issue associated with the study of state criminality is to know the actual numbers of victims (Bijeveld 2007). As we know with street crime, there is a dark figure involved. Yet, criminologists attempt to get beyond this and obtain estimates that are more reflective by using, though imperfect, multiple venues available (self-report surveys, the United States' National Crime Victimization Survey, and the British Crime Survey are well known examples). However, with international crimes, this dark figure is a "doubly dark figure" (Bijeveld 2007, p. 4). In part, this can be due to a state's unwillingness to disclose the information for multiple reasons, victims' desires to remain silent, lack of survivors, lack of pre-conflict census data, lack of post-conflict citizenry data, significant population displacements, and a score of other variables. As such, criminologists must attempt to use multiple methods to make the doubly dark figure of the crimes, at best, a dark figure of the crimes. The important task of counting the precise number of victims remains a laborious task, yet not only is it important but it also speaks to the difficulty of using quantitative methods. For example, the death toll in Darfur has been estimated to be between 60,000 and 160,000. However, the Coalition for International Justice reports estimates near 400,000. The number of victims from the genocide in Rwanda is estimated to be between 500,000 and 1,000,000. The mortality rate of Pakistanis, due to the conflict in 1971, varies by a threefold variation, between one and three million, and estimates of the death toll in Congo between 1964 and 1965 varies tenfold. During the twentieth century, it has been suggested that 170 million people have been killed in “conflicts of a non-international character, internal conflicts and tyrannical regime victimization” (Bassiouni 1996, p. 2). Since the beginning of the twenty-first century there have been hundreds of thousands more killed, maimed, tortured, displaced, and/or raped. For those who use a more expansive definition of state criminality, the harms are even more insurmountable when we consider those generated by states’ omitting to alleviate specific conditions, or respond to natural disasters, cases of institutionalized racism, ethonocism, classism, and a host of other injuries. As you can see, figures for victims display enormous variance.

There are various forms that can be used to obtain statistics, no different than those used for traditional street crime, namely, victim surveys. These have been used for estimating mortality. Yet, in situations of these types of crime, significant portions of the population may have fled the country or may reside in refugee centers abroad, making it more difficult to account for the double-dark figures; or, if entire families were killed, mortality would be underestimated. An additional barrier to obtaining exact numbers or using victimization surveys is the inability to often go to the regions affected if the conflict is ongoing. Security issues can affect access to areas and thus the representativeness of the total numbers of victims. Further, some victims may not be willing to open up and share their experiences; this is particularly the case with victims of genocidal rape. Additionally, many victims may not know or perceive themselves as victims of state crime. Others may self-identify as such but be denied the status through efforts to ensure state legitimacy and to cover state criminality. Regretfully, the true numbers of victims of state crime will remain a “doubly-doubly-doubly” dark figure that is really beyond what most of us could comprehend (Rothe and Kauzlarich forthcoming).

Justice

Popular thinking about victims often dubiously assumes that victim healing can be accomplished simply through the formal processing of the offender(s). The logic here is that if the perpetrator is caught, tried, convicted, and punished, the victim will be able to somehow regain what she or he has lost. This common-sense belief, however, has been shown to be overly optimistic and quite naïve (Lobwein 2006). While it is true that some victims report feeling better about, for example, honoring their dead family members by telling their story on the stand, just as many appear to find formal criminal justice procedures alienating, frustrating, and without much healing power (Stover 2005). This is an exceptionally common finding in traditional street crime victimological studies.

Most truth commissions and international courts cannot regularly deliver outcomes or structure their proceedings in an expressly victim-centered manner. With truth commissions, it is often uncertain before and sometimes after a commission completes its investigation whether or not individual offenders will be named, and further, whether trials may be conducted on the basis of the information collected in a truth commission proceeding. The record shows that of the several dozen truth commission convened to date, only four have named the perpetrators. In some instances, complete and wholesale amnesty was given to offenders, as in the case in El Salvador, while the Argentinean truth commission investigation eventually led to criminal trials (Hayner 2001, 2006). In the case of South Africa, the Truth and Reconciliation Commission was allowed to offer amnesty or a case by case basis, although the names of offenders were a matter of public record. Indeed, Hamber et al. (2000) study shows that almost all of the South African victims they interviewed felt that granting amnesty to offender is unfair or simply "wrong." Victims of the Lord's Resistance Army in northern Uganda echo this sentiment in a sense as almost half of those surveyed indicated that, among other things such as restorative justice processes, perpetrators should be held accountable for their crimes and punished (Pham et al. 2005). It must also be understood that truth commissions are principally fact finding units, not a victim services center. Historically, truth commissions tend to be carried in places trying to recover from war, massive oppression, and economic and political crisis. The lack of material resources along with the primary business of fact gathering often lead to dissatisfaction and anger.

There are better alternatives for victims than these bureaucratic justice procedures that rely on principles of restorative rather than retributive justice. Some justice systems use a combination of these justice philosophies, while others are strictly one or the other. An interesting hybrid is evidenced by the relatively new International Criminal Court (ICC) which has an unprecedented system for helping victims. While the Court is still very traditional in terms of adversarial justice processes, the promising Victims Trust Fund has done much more than simply focus on punishing offenders. The trust fund has primarily helped victims of state crimes in Congo and Uganda, and has provided resources for psychological counseling, physical rehabilitation, education, various health services, and material support for education and agriculture. The Trust Fund is especially concerned with the following groups: Victims of sexual and gender-based violence, widows/widowers, former child soldiers and abducted youth, orphans and other vulnerable children, and victims suffering from physical and mental trauma (Trust Fund for Victims 2011). Interestingly, a sitting judge on the ICC has recently written a fascinating essay on her concerns with the manner in which the ICC can truly give victims voice, provide reparations, and facilitate healing (Van den Wyngaert 2012). Her main concerns are that the ICC is a bureaucratic, formal legal-adversarial system, and not a fully victim-centered process, and that because the Trust Fund is based mostly on individual and state donations, one should expect less than ideal results for victims. Alternatively, a fully restorative justice approach to harm and healing would center on three major stakeholders, the victim, the offender, and the community. The goals would be for offenders to recognize, apologize, and atone for their crimes, for victims to be free to voice their pain and suffering and be given resources to heal, and for the community to rebuild itself after crime or violence.

Resisting state crime and victimization has become a major point of interest for criminologists. Stanley and McCulloch (2013) have produced an anthology that considers a whole range of techniques and approaches that may be used to object to state crimes such as human rights violations in the Middle East, West Papua, Sri Lanka, the United States, the United Kingdom and other areas focusing on civil, popular, state, and international forms of resistance. The main question for scholars is what actions may be taken to reduce the harms caused by states, whether in official criminal or civil law, international regulation, social movement activities, social media outlets, or structural changes in the essential functions of states. As with any scholarly examination, the crime and the victim are obviously intertwined, but because of the likelihood for state crime to directly involve politically legitimate entities as offenders and their subjects or enemies as victims, the fundamental relationship between institutional power and individuals or groups of individuals with considerably less power reveals two major dynamics. First, resisting some types of state crimes, such as discrimination against the rights of gays and lesbians to marry may be facilitated by a combination of legal challenges, publicity, and social movement activities. Likewise with incidences of police or correctional officer brutality, states with greater freedoms may be challenged quite directly through politics and the media. In these types of situations, challenges to victimization and repeat victimizations may ultimately sway social opinion, the legal or political systems, and the populace to take notice of state crimes and effectively challenge them through both civil and state avenues. For states with less freedom of the press and assembly, resisting victimization is considerably more difficult as any resistance itself may be criminalized. Numerous examples of resistance to state crime victimization in the form of human rights violations in China, Iraq, Cuba, and many other areas of the globe illustrate how without substantial social movement activity, as found in some revolutions in the Arab Spring, resistance is weak and further victimization is more likely. Second, state crime victimization can occur as a result of higher level international affairs issues, which is not at all how most traditional crimes and victimizations take place. War always results in innocent killing and injury, and both soldiers and civilians are victimized. Most actors caught in the crossfire are unable to resist victimization for obvious reasons, and international aid agencies such as the Red Cross or Doctors Without Borders become primary aides in helping address injury. As recently witnessed in the face of gross human rights violation in Syria, without substantial political agreement among major powers within the United Nations, especially within the Security Council, little can be done to relieve the suffering from war.

As the state crime literature in criminology has advanced over the last two decades, the study of the victims of this form of crime has lagged behind, but has been bolstered in the last few years with more interest in the subject matter. Enduring issues in this area of inquiry are the definition of state crime itself, conceptualizations of victims, measuring and recording victimization, resisting victimization, and justice processes, whether formal or informal, to address state harms both prior to and after the victimization occurs.

Related Entries

- ▶ Crimes of the Powerful
- ▶ Repeat Victimization
- ▶ Resistance to State Crime
- ▶ Restorative Justice and State Crime
- ▶ State-Corporate Crime
- ▶ State Crime
- ▶ Victim Input at Sentencing
- ▶ Victims and Restorative Justice
- ▶ Victims and the International Criminal Court

Recommended Reading and References

Ajdukovic (2006) Barriers to social reconstruction of communities in the aftermath of organized violence. In: Ewald U, Turkovic K (eds) Large-scale victimization as a potential source of terrorist activities: importance of regaining security in post-conflict societies. IOS Press, Amsterdam, pp 269–277

Bassiouni C (1996) Searching for peace and achieving justice: the need for accountability. Law Contemp Probl 59(4):9–28

Bijeveld C (2007) So many missing pieces: some thoughts on the methodology of the empirical study of gross human rights violations. Paper presented at the expert meetings on supranational criminology. Maasstricht

Ewald U, Turkovic K (eds) (2006) Large-Scale Victimisation as a Potential Source of Terrorist Activities: Importance of Regaining Security in Post-Conflict Societies. U. Ewald and K. Turkovic

Faust K, Carlson S (2011) Devastation in the aftermath of Hurricane Katrina as a state crime: social audience reactions crime. Crime Law Soc Change 55(1):33–51

Faust K, Kauzlarich D (2008) Hurricane Katrina victimization as a state crime of omission. Crit Crim 16(2):85–103

Green P (2009) Women and natural disasters. In: Cain M, Howe A (eds) Women, crime and social harm. Hart, Oxford, pp 161–177

Hamber B, Nageng D, O'Malley G (2000) Telling it like it is…: understanding the truth and reconciliation commission from the perspective of survivors. Psychol Soc 26:18–42

Hayner PB (2001) Unspeakable truths: confronting state terror and atrocity. Routledge, New York

Hayner PB (2006) Truth commissions: a schematic overview. Int Rev Red Cross 862:295–310

Kauzlarich D, Matthews RA, Miller W (2001) Toward a victimology of state crime. Crit Criminol Int J 10(3):173–194

Lobwein W (2006) Experiences of the victims and witnesses section at the I.C.T.Y. In: Ewald U, Turkovic K (eds) Large-scale victimization as a potential source of terrorist activities: importance of regaining security in post-conflict societies. Ios Press, Amsterdam

Pham PN, Vinck P, Wierda M, Stover E, di Giovanni A (2005) Forgotten voices: a population-based survey of attitudes about peace and justice in northern Uganda. Human Rights Center, Berkeley

Rothe D, Kauzlarich D (forthcoming a) Toward a Victimology of State Crime. Routledge. London. forthcoming.

Rothe D, Kauzlarich D (forthcoming b) Introduction In: Rothe D, Kauzlarich D (eds) Toward a victimology of state crime. Routledge, New York

Rothe D, Mullins C (2011) State crime: current perspectives. Rutgers University Press, New Brunswick

Stanley E, McCulloch J (2013) State crime and resistance. Routledge, New York

Stover E (2005) The witnesses: war crimes and the promise of justice in The Hague. University of Pennsylvania Press, Philadelphia

Strobl R (2010) Becoming a victim. In: GioraShoham S, Knepper P, Kett M (eds) International handbook of victimology. CRC Press, Taylor & Francis, London, pp 9–25

Turkovic K (2002) Overview of the Victimological Data Related to War in Croatia. Eur. J. Crime Crim. L. & Crim Just.10(202):202–215

Trust Fund for Victims (2011) Programme progress report, summer. International Criminal Court, The Hague

Van den Wyngaert, Christine (2012) Victims before international criminal courts: some views and concerns of an ICC trial judge. Case W Res J Int Law 44:475–496

Westervelt SD, Cook KJ (2010) Framing innocents: the wrongly convicted as victims of state harm. Crime Law Soc Change 53(3):259–275

Victims' Rights in the Criminal Justice System

Jonathan Doak
Durham Law School, Durham University,
Durham, UK

Overview

The concept of victims' rights has come to bear considerable influence on the formation of criminal justice policy on both the domestic and international platforms. While commentators have long accepted the desirability of rights for accused persons, it is only in more recent times that a discourse concerning the rights of victims has emerged. Prior to exploring the evolution and extent of such rights, it is worth noting at the outset that the tendency of policymakers and politicians to adopt language couched in the terminology of *victims* rights' often departs from the notion of a *legal* right, in the sense that it can be enforced through the justice system. In other words, while certain benefits or dispensations may be framed as rights on paper, they frequently lack any enforcement mechanism and may not be considered binding on courts or other public authorities.

History and Development

There are diverse and varied accounts which document the ascendancy of victims' rights within the criminal justice system (see generally Kearon and Godfrey, 2007). Many such accounts begin with reference to Christie's (1977) critique, in which he argues that the State effectively *stole* the conflict from its primary protagonists and resulted in the evolution of contemporary criminal justice system being normatively and structurally built around a contest between the state and the accused. For many centuries, victims were "underestimated, ignored and undervalued" (Walklate 2007: 11). It was not until the latter part of the twentieth century that the interests of the victim began to impact upon the shape of criminal justice reform in any meaningful way.

A growing awareness of victimization and its consequences emerged in the years following the Second World War. The postwar era was marked by a "shifting understanding of accountability and citizenship" (Walklate 2001: 204), which was evidenced in the shift towards welfarism, concern for vulnerable members of society, and the emergence of civil society. Although criminal justice policy was not a widely contested political issue during this time, it was within this climate of emerging social rights that a loose association of groups and individuals became involved in campaigning for victim-specific issues (Mawby and Walklate 1994). The feminist movement, in particular, played a key role in highlighting aspects of hidden victimization stemming from domestic violence, rape, and sexual assault across many countries, including the USA, Australia, and the United Kingdom (Dignan 2005; Wolhuter et al. 2009). Increased attention to the plight of children also showed that child abuse was much more prevalent than had been once thought and was a form of hidden victimization. In the United Kingdom, for example, several high profile cases came to light concerning the abuse of children in the care of social services, or where social services had failed to intervene (Mawby and Walklate 1994).

The 1980s and 1990s witnessed a dramatic growth in the number of specific interest groups in many different countries, including the USA, the United Kingdom, Australia, and New Zealand. These organizations included those campaigning for the registration of sex offenders, incest survivor groups, relatives of murdered and missing children, relatives of victims of drunk driving, and those concerned with combating racism, homophobia, and discrimination generally. While such groups were generally unconnected and pursued their own specific agendas, the net effect of their efforts was to highlight the plight of weaker and more vulnerable members of society on many different levels under existing legal and political frameworks (Mawby and Walklate 1994).

Notwithstanding this fragmentation, the victims' movement was becoming gradually more coherent. Despite their differing agendas and policy objectives, Karmen (1992: 159) argues that, from an international perspective, the movement reinforced a fivefold basic critique of the criminal justice system:

> (1) That criminal justice personnel – the police, prosecutors, defence attorneys, judges, probation officers, parole boards... were systematically overlooking or neglecting the legitimate needs of crime victims until they all began their campaign; (2) that there was a prevailing tendency on the part of the public as well as agency officials to unfairly blame victims for facilitating or even provoking crimes; (3) that explicit standards of fair treatment were required to protect the interests of complainants and prosecution witnesses... (4) that people who suffered injuries... ought to receive reimbursement... (5) that the best way to make sure that victims could pursue their personal goals and protect their own best interests was by granting them formal rights within the criminal justice system.

The victims' lobby in the USA has been traditionally perceived as being much more proactive and rights focused than its European equivalents, and the movement grew rapidly during the latter part of the 1970s into the 1980s. The National Organization for Victim Assistance (NOVA) was founded in 1975 and has since been involved in a campaign of national advocacy, championing the victim's cause and exerting pressure on federal and state authorities. Although many European countries also witnessed the formation of more consolidated national organizations, these tended to focus on meeting the needs of individual victims rather than leading political campaigns to enhance their rights. In the United Kingdom, local associations came together to form Victim Support in 1979, and the organization rapidly expanded its operations during the 1980s and 1990s and secured a twofold increase in Home Office funding (Dignan 2005). The organization now performs the role of a quasi-public body in championing victims' rights and supporting victims in their capacities as witnesses in court. Similar victim assistance organization evolved in Germany (1976), Ireland (1983), and the Netherlands (1985) (see generally Hall 2010).

As crime victims became more visible, western governments became increasingly concerned with being seen to address their needs as well as the wider problems stemming from the perceived fear of crime. As Geis (1990: 260) describes, "[t]he plight of the crime victim is dramatic and determinable. Their relief is feasible. It has strong political, social and personal appeal. Any of us, at any time, could become a crime victim." In Garland's (2001) view, such politicization stems from the wider collapse of *penal welfarism* and the subsequent need for governments to redefine the role of the criminal justice system. Concerns have, however, been expressed in some quarters that governments throughout the western world are increasingly willing to use the crime victim as a front so that they are perceived as doing something about crime (Garland, *ibid*.). In this sense, official rhetoric may be used as a mask to introduce policies that are primarily aimed at securing the interests of the state as opposed to victims (Jackson 2003). Such a trend has arguably resulted in governments embarking upon piecemeal and hurried approaches to statutory reform in the name of the victim, in order to maximize the political appeal of certain policies and, arguably, to increase conviction rates. Koffman (1996), for example, has warned that piecemeal measures couched in the language of victims' rights make it appear that the government is addressing concerns over apparently rising crime while avoiding having to incur the major outlay of expenditure that would be required for more far-reaching reviews of the criminal justice system or major programs of reform.

The 1990s and 2000s witnessed a drift away from a needs-based or welfare-orientated approach to victims' issues, and a rights-based discourse dominated the policy agenda in the following years (Wolhuter et al. 2009). Many jurisdictions began to produce minimum standards for victims in the form of charters, codes of conduct, or statements of rights. In the United Kingdom, the first Victims' Charter was published in 1990. The Charter consisted of a set of aspiring principles, while the second edition of the Charter, published in 1996, committed the government to benchmarking existing practices through a set of standards which victims could legitimately expect in relation to criminal justice services. Specific standards were laid down in relation to the role of the police, Victim Support, compensation, prosecution, and conviction/ release of an offender. However, its provisions remained unenforceable through the courts, and the implementation of these standards was largely dependent upon the discretion of individual criminal justice agencies (Koffman 1996). Similar charters were launched in New Zealand (1987), Australia (1993), and Ireland (1999).

In the USA, many states amended their constitutions to give effect to victims' rights, while others opted to entrench rights through specific legislation. At the federal level, Congress passed the Victim and Witness Protection Act 1982, which sought to guarantee fair treatment for victims and witnesses in criminal courts, and the Victims of Crime Act 1984 established the Crime Victims Fund, comprising federal fines and bond forfeitures to support state victim compensation and local victim assistance programs. Since 1980, all 50 states have adopted over 1,000 pieces of legislation enshrining victims' rights within their legal systems, including Bills of Rights (Doak 2008).

The beginning of the twenty-first century heralded further advances for victims, with governments seemingly more willing to introduce legislative reforms and revise charters in order to give effect to more radical rights. In the USA, the Crime Victims' Rights Act 2004 (forming part of the Justice for All Act) conferred victims with eight specific rights, these being the right to protection from the accused, the right to notification, the right not to be excluded from proceedings, the right to speak at criminal justice proceedings, the right to consult with the prosecuting attorney, the right to restitution, the right to proceedings free from unreasonable delay, and the right to be treated with fairness and respect for the victims' dignity and privacy. In contrast to legislative attempts elsewhere, the rights contained in the legislation are justiciable in the sense that victims are given legal standing to petition the federal (though not the state) courts in the event that any of the above rights have been breached.

In the United Kingdom, the Victims' Charter was replaced by a statutory code of conduct pursuant to the Domestic Violence, Crime and Victims Act 2004. The new code of practice conferred victims with statutory rights to support, protection, information, and advice in relation to all the major criminal justice agencies. If victims feel the code has not been adhered to by any criminal justice agency, they can file a complaint in the first instance, with the respective agency, and then, via their Member of Parliament, with the Parliamentary Ombudsman. In addition, the legislation has created the office of an independent commissioner for victims to promote the interests of victims and witnesses. Exercising an oversight role, he or she is also responsible for reviewing the code on a regular basis and encouraging good practice. However, the limited extent of the powers conferred on the commissioner underlines concerns that the code fails to provide for any effective enforcement mechanism. Section 34 of the Act provides that a failure by any agency to abide by a duty contained in the code "does not of itself make [any person] liable to criminal or civil proceedings." There is thus no freestanding right for victims to enforce the code through the courts: the judiciary is generally unable to give it legal effect. Manikis (2011) criticized the complaints process on the grounds that it was overly complex and largely inaccessible for many victims. Hall (2010) reported that the lack of effective enforcement mechanisms is a common limitation found in most charters, with the US Crime Victims' Rights Act constituting the most robust system of protection.

The International Context

The ascent of victims' rights within domestic criminal justice systems has been mirrored – and indeed partly driven – by developments on the international platform. Williams and Goodman (2007) suggest that this is particularly true of European jurisdictions, which took their cues from international benchmarks as opposed to Anglo-American countries, where the victims' movement had been more *proactive* in nature.

The first major development came in 1985, when the United Nations issued the Declaration of Basic Principles for Justice for Victims of Crime and Abuse of Power, which defined crime prevention as a victims' rights issue, and enshrined victims' rights to access to justice, fair treatment, information, assistance, and access to informal dispute resolution methods. Since then, a range of UN instruments, including the Basic Principles for the Treatment of Prisoners, the UN Convention against Transnational Organized Crime and the Standard Minimum Rules for Non-custodial Measures, and the Vienna Declaration on Crime and Justice, have required that the rights of victims to be taken into account in various ways.

On a regional level, crime victims have also been subject to a range of protections. In Europe, both the Council of Europe and the European Union have also been involved in standard setting. In 1983, the former organization was responsible for formulating the European Convention on the compensation of victims of violent crimes which lays down minimum standards for the provision of state compensation to victims of crime. Signatory states to the Convention are required to ensure that victims suffering serious bodily injury or impairment of health receive compensation for at least the loss of earnings, medical and funeral expenses, and the loss of maintenance to any dependents. The Convention was followed in 1985 by the adoption of Recommendation 85(11) on the position of the victim in the framework of criminal law and procedure, which contained guidelines aimed at protecting victims of crime and safeguarding their interest at each stage of the criminal procedure as well as calling on signatories to examine the possible advantages of mediation and conciliation. Since then, the Council of Europe has published many more recommendations, setting down a range of standards in relation to the provision of support, protection, and assistance for victims and witnesses (see especially Rec 87(21), calling on member states to adopt a range of measures to develop assistance and support programs for victims, and Rec 06(08), which stipulates that states should assist victims in obtaining compensation from the offender through both criminal and civil proceedings).

The European Union has also been active in standard setting. In 1999 the commission issued a communication to the European Parliament entitled *Crime Victims in the European Union: Reflections on Standards and Action*. This document contained 17 proposals grouped under five main headings: prevention of victimization, assistance to victims, standing of victims in the criminal procedure, compensation issues, and general issues (information, language, training) and called on all member states to implement fair and effective legislation in these areas. Following its adoption by the European Parliament, in March 2001, the Justice and Home Affairs Council adopted the *Framework Decision on the Standing of Victims in Criminal Proceedings*.

The Framework Decision is particularly significant because, in contrast to the various declarations, recommendations, bodies of principles, and other soft law pronouncements of international bodies, the Framework Decision is legally binding and, as such, is directly applicable in all member states of the European Union. More recently, the commission has proposed that the Framework Decision be replaced by a new directive. The *Directive establishing minimum standards on the rights, support and protection of victims of crime* replicates much of the original wording of the Framework Decision. Given that the objectives of the Framework Decision were not wholly realized, concerns have been expressed that the new Directive will similarly fail to meet its own objectives and will not really add anything new in terms of concrete measures to protect victims (Doak and Taylor 2012).

The Substance of Victims' Rights

Victims' rights are commonly classified as falling within two broad groups. First, victims may possess *service* or *social* rights, which encapsulate those specific entitlements which the state makes available to victims in the aftermath of a crime. Secondly, victims may be afforded *procedural* rights, which entail some form of participation within criminal procedure itself. In addition to this conventional taxonomy, it can be noted that victims are also entitled to the same *human* rights as everyone else in society, which may (depending on jurisdiction) be entrenched through a constitution or legislation act and enforced through the courts.

Service Rights

The first group of rights to which victims may be entitled has been variously labeled *service* or *social* rights. These cover matters such as access to practical support and counseling and the provision of information about the case as it passes through the criminal justice system. The past three decades have witnessed a growing body of evidence which suggests that the lack of information is one of the greatest sources of dissatisfaction for victims and witnesses (Wolhuter et al. 2009). The European Framework Decision sought to address this deficit through stipulating that victims must be given access to any information relevant to the protection of their interests, which should include the types of support and services or organizations available for victims, places and formalities for reporting an offense as well as the ensuing procedures, conditions for obtaining protection, conditions for access to legal or other advice and aid, requirements for receiving compensation, and arrangements available for nonresidents. However, a particular difficulty in providing for such a right in practice is that criminal justice agencies tend to have discrete responsibilities at different stages of the criminal process. This means that information is not readily shared between the different agencies, and it is not always clear which agency bears the burden of sharing which part of the information with the victim (Dignan 2005).

The European Framework Decision also commits member states to ensuring that “victims are treated with due respect for the dignity of the individual during proceedings” and that “victims who are particularly vulnerable can benefit from specific treatment best suited to their circumstances.” There is, however, considerable leeway as to how terms such as *due respect*, *dignity*, and *treatment best suited* are interpreted, let alone realized in practice (Doak 2008).

While such service rights may prove often expensive or burdensome from an administrative perspective, they are not generally contentious since they only touch upon the passive or indirect role of the victim. Two further major service rights are more radical in nature and have been the subject of significant critique, these being the provision of compensation for criminal injuries and the use of special measures to support vulnerable victims and witnesses while testifying.

Compensation

Many international instruments now recognize that victims of crime ought to be able to exercise a right to receive compensation in the aftermath of violent crime that has caused death or bodily injury. Examples include the European Convention on the Compensation of Victims of Violent Crimes, Council of Europe Recommendation 06(8), and Article 9 of the EU Framework Decision. However, provision for criminal injuries compensation was one of the first service rights acquired by victims many years before such needs were recognized by international benchmarks.

New Zealand was the first to develop such a scheme in 1963, followed by the United Kingdom the following year. Schemes were created in California and New York in 1965 and 1966, respectively, while New South Wales became the first Australian jurisdiction to launch a program in 1968. Schemes are now widespread throughout Europe, Australasia, and North America, with a number of developing countries including the Philippines and Colombia also operating similar programs.

While schemes vary considerably in terms of scope and reach, they usually provide compensation to victims of violent crime, either using a tariff-based approach or assessing each case on its individual merits. Some such schemes have come under criticism for failing to guarantee access as a *right* to all victims; payments are usually made on an ex gratia basis and are only available to those who meet strict eligibility criteria. While compensation is paid by the state, it can be noted that criminal courts in many jurisdictions have the power to order offenders to make compensation payments to victims directly for any loss or damage resulting from the offense.

One of the criticisms that is most frequently levied at state compensation schemes is the fact they appear to distinguish between *deserving* and *undeserving* victims. Most schemes, including that in the UK, only pay out in those cases where the victim is free from blame, thereby promoting the dangerously inaccurate concept of the *ideal* victim and precluding *deviant* victims from accessing any recompense (Miers 2000). Only those victims who strike a chord with society's view of the *innocent* victim are deemed to deserve state support; *bad* victims are cast aside by the system (Freckleton 2004). The *ideal* victim will be happy to cooperate fully with the criminal justice system and the police. Failure to report an offense promptly, failure to make a statement, failure to participate in an identity parade, and failure to attend court have all been cited as grounds for refusing compensation (Spalek 2006).

Despite the widespread proliferation of state compensation schemes, many jurisdictions will seek to prioritize mechanisms that purport to achieve reparation directly from the offender. Such mechanisms, however, are prone to two major shortcomings. First, as noted above, they can only be used where the offender is actually identified. Secondly, most offenders do not have the means to make substantial compensation payments to victims. In the United Kingdom, there are two main avenues for victims to achieve reparation directly from the offender: these are through a legal action in tort through the civil courts or through a compensation order issued by a criminal court.

Since 1973, courts in England and Wales have been empowered to order the offender to pay compensation as part of a sentence for any personal injury, loss, or damage resulting from the offense. Originally introduced as an ancillary penalty, from 1982, courts were given the power to award compensation orders as penalties in their own right. While it is still the norm for them to be awarded in conjunction with other penalties, the court may make a compensation order either instead of, in or addition to, any other penal sanction. Similarly, the German *adhesion* procedure, distinct from subsidiary prosecution described above, also makes it possible for civil damages to be claimed within the criminal action.

Special Measures

The past two decades have witnessed considerable efforts to reduce secondary victimization in the courtroom, which had become widely documented in adversarial jurisdictions (see generally Ellison 2001). Common measures adopted to alleviate the stress of testifying include the use of screens, televised links, and prerecorded video evidence. Jurisdictions differ in terms of which witnesses may benefit from such measures. Although many schemes were originally introduced with the specific aim of protecting child witnesses, some have been extended to include, inter alia, complainants in rape and sexual assault cases, witnesses in fear of intimidation or reprisal, elderly and disabled witnesses, and witnesses suffering from psychiatric or developmental disorders.

Vulnerable witnesses in inquisitorial jurisdictions have generally not had to contend with the same degree of secondary victimization as their counterparts in adversarial trial systems (Ellison 2001). However, screens and video links are becoming more commonplace in continental systems, with trial judges generally being able to exercise a broad discretion as to what general steps might be taken to accommodate or protect vulnerable witnesses (Doak 2008). Other common measures that may be adopted include the use of written evidence taken in advance of the trial, holding some or all of the trial *in camera*, or imposing restrictions on media reporting (Doak, *ibid*). Some countries also make special provision for complainants to be questioned in the absence of the defendant. For example, Spain provides that the offender can be expelled from the court if he behaves in a way that upsets the victim by behaving in an inappropriate manner (Brienen and Hoegen 2000: 875), and the Dutch courts may make provision for a threatened witness to testify anonymously, subject to strict conditions (Ellison 2001).

Procedural Rights

Procedural rights entail some form of participation by the victim in criminal procedure. As such, they raise much deeper questions concerning how far the victim's interest ought to be accommodated in what is often perceived to be a contest between the State and the accused. International standards have tended to shy away from laying down explicit requirements in terms of stipulating that victim participation ought to be enshrined as a generic benchmark, and the concept of participation does not feature at all in any of the Council of Europe's recommendations. Those instruments that touch on the issue tend to do so in a relatively vague or non-prescriptive manner. For example, Principle 6(b) of the UN Declaration on Victims states that the judicial process should allow "the views and concerns of victims to be presented and considered at appropriate stages of the proceedings where their personal interests are affected, without prejudice to the accused." Likewise, Article 3 of the EU Framework Decision grants victims a "right to be heard and supply evidence." Neither instrument, however, offers specific details as to how such rights are to be realized in practice or as to what stage(s) of the process they are to be applied.

Yet, most jurisdictions from both the common and civil law traditions do make provision for some form of participatory rights. The common-law system of England and Wales, for example, still permits victims to commence and conduct private prosecutions. This procedural right is mostly alien to civil law systems, although for various reasons, it is rarely resorted to in practice (not least because of costs and access to evidence). If victims feel disenfranchised by a decision not to prosecute, it may instead be possible for them to challenge the decisions of public prosecutors by means of review. This means of challenge is found in many continental European systems. Review will either lie with a superior authority, as in Denmark, Luxembourg, and Norway, or a court, as in Germany, Portugal, and the Netherlands (Brienen and Hoegen 2000).

Participation in the Trial Process

The ability of victims to actively participate in the trial process is generally limited, and international standards do not lay down any requirement that victims ought to be considered parties to proceedings. However, it can be noted that Article 68 of the Rome Statute confers victims with a general right to participate in proceedings at the International Criminal Court. Victims may choose their legal representatives, who have the right to present their views and make submissions to the court when their interests are likely to be affected. Such views and submissions may be made at all stages of the court proceedings with only the limitation that it would not be prejudicial or inconsistent with the rights of the accused.

By contrast, victims have little opportunity to participate within domestic criminal trials in common-law jurisdictions. Victims in England and Wales have no right to be present at the trial, no right to legal representation, no right to question witnesses, and no right to present evidence. Other adversarial systems have relaxed their rules in this regard. In the USA, a number of individual states have already adopted amendments to their constitutions and have created a special office of "victim advocate" to enforce provisions contained in their constitutional amendments. For example, *a* victims' service advocate was added to the New Mexico Office of the Attorney General in 1999 to provide assistance to victims of violent crime and their families. Some states make provision for the victim to rely on the presence of a legal representative at court. Washington State provides that victims of violence and sex crimes have the right to have an advocate present at any prosecutorial or defense interviews with the victim and at any subsequent judicial proceedings. Similarly, Illinois confers a constitutional right on victims to have the presence in court of "an advocate or other support person of the victim's choice" (Section 8(1), Illinois Bill of Rights). In these cases, however, the right is confined to the mere presence of a legal representative as opposed to any actual involvement.

Other US states now take a more radical approach and permit victims to hire their own attorneys to represent them at various parts of the criminal process, and certain states go further and permit victims' counsel to intervene in rape and sexual assault trials. Wisconsin, West Virginia, and New Hampshire allow complainants' representatives to make representations when questions governing the admissibility of sexual history evidence are being considered by the court. South Carolina Statute 16–3–1510 is even broader, insofar it permits representations from a victims' advocate in any type of case where the defendant alleges improper or illegal conduct on the part of the victim as part of his or her defense.

The inquisitorial jurisdictions of continental Europe have been historically more relaxed about concept of victim participation in the trial, with a number of jurisdictions permitting the victim to assist the prosecutor as a subsidiary counsel. Germany, Austria, Malta, Norway, Sweden, and various eastern European countries operate some form of *subsidiary prosecutor* schemes, which allow victims an active role in both pretrial decision-making and within trials themselves. The procedure will generally allow them to submit evidence, comment on representations made by the prosecution and defense, and express their opinions on key decisions taken. In Germany, victims of certain serious offenses or the relatives of a murder victim may act as a subsidiary prosecutor (*Nebenkläger*). By declaring his solidarity with the prosecution, the victim derives certain active participatory rights. These include a right to be present at all stages of the process, to put additional questions to witnesses, to provide additional evidence, and to make a statement or to present a claim for compensation. In this sense, the victim's lawyer can be an important ally to the public prosecutor, who nonetheless retains the burden of preparing and presenting the prosecution.

An alternative model which allows for the exercise of victims' procedural rights is commonly referred to as the *adhesion* or *partie civile* procedure. Participation of the victim as an independent civil party bears some similarity to the subsidiary prosecution model, although it has a distinct advantage in that it acknowledges the victim's status as a separate party to the trial. The procedure is widely utilized in France and Belgium, where the victim must formally demonstrate his or her intention of becoming a party to the proceedings by initiating an independent action before the *juge d'instruction* (*constitution de partie civile*) at any stage in the proceedings.

Victim Participation in Sentencing

Participation within sentencing procedure is more commonplace than participation within the trial *proper*. Many criminal justice systems now make provision for victim impact statements to be admitted as part of the sentencing procedure. Two common justifications are offered. First, such statements are said to have a functional purpose through giving the court a better picture of the physical, emotional, or financial harm suffered by the victim as the result of an offense. Alternatively (or in addition), victim impact statements are commonly said to provide a communicative or expressive function, insofar as they provide the victim with a means to exercise voice within sentencing hearings (Roberts and Erez 2004).

Schemes vary somewhat in their nature, and legislation often fails to stipulate precisely how victim impact statements ought to be taken into account by the court or the degree of weight that ought to be attached to them (Roberts and Manikis 2010). Such matters are generally left to the judiciary in common-law settings (Roberts and Manikis, *ibid*.). The content of victim statements may also differ substantially. For the most part, statements are restricted to explaining the impact of the offense on the victim, though some schemes permit victims to lay down specific penal demands. Different methods are also used to tender the statement. These may be appended in written form to the judge's papers (as in England and Wales) or may form part of a presentence report compiled by probation or social services. Alternatively, statements may be tendered orally in court by the prosecutor, victim, or victim's lawyer (see generally Booth and Carrington 2007).

Research has revealed that, in general, victim statements only have a minimal impact upon sentencers. In their interviews with judges in Scotland, Chalmers et al. (2007) found that many found it difficult to isolate the effect of the victim impact statement from the vast array of other factors that were built into the sentencing equation, though the researchers do note a number of incidents where judges were minded to "consider a sentence of a different nature from the one they were initially minded to impose" (at 376). Similar findings have been reported elsewhere (see Roberts and Erez 2004).

Substantive Rights

In addition to the service and procedural rights discussed above, crime victims – as citizens – are also entitled to rely on substantive human rights provisions which may be triggered in the aftermath of an offense. Human rights law has developed rapidly in recent times. Although for many years primarily associated with victims of state crime or abuse of power, it is now becoming frequently recognized that its provisions may also be significant for victims of *ordinary* or *non-state* crime (Doak 2008).

For example, various international treaties require states to put in place legislative frameworks together with practical measures to protect individuals from serious forms of crime that threaten their life or physical integrity. Article 2 of the Convention Against Torture provides that "each State Party shall take effective legislative, administrative, judicial or other measures to prevent acts of torture in any territory under its jurisdiction," and Article 25 of the United Nations Convention against Transnational Organized Crime states that "[e] ach State Party shall take appropriate measures within its means to provide assistance and protection to victims of offenses covered by this Convention." Preventative obligations are particularly commonplace in circumstances where particularly vulnerable parties are perceived to be at heightened risk. Children, those at risk from domestic violence and those at risk from intimidation or repeat victimization, have all been subject to some form of special recognition within international instruments.

Under the European Convention, very specific demands are placed upon signatory states. These include a duty to put in place preventative measures to ensure that non-state parties do not breach the human rights of others, and the

European Court has made clear that states must not only refrain from taking life but must also take steps to protect life against threats from third parties (see e.g., *Osman v United Kingdom* (1998) 29 EHRR 245 and *Edwards* v *United Kingdom* (2002) 35 EHRR 19). It has also been held that the State has a duty to take positive steps to protect the rights of victims under Article 3 (prohibition against torture and inhuman or degrading treatment) by ensuring that they are afforded adequate protection under the criminal law (see *A* v *United Kingdom* (1999) 27 EHRR 611; *MC* v *Bulgaria* (2005) 40 EHRR 20). Similar approaches have been adopted by the Inter-American Court of Human Rights (see, e.g., *Velasquez Rodriguez* v *Honduras* (1989) 28 ILM 291; *Jessica Gonzalez v. United States*) (Case 1490–05, Report No. 52/07, 2007). Where a witness is in fear of intimidation or reprisal, the European Court has also accepted that evidence may be given anonymously in exceptional circumstances, providing sufficient counterbalancing measures are in place (*Doorson* v *Netherlands* (1996) 22 EHRR 330). Articles 3 (prohibition against torture and inhuman or degrading treatment) and 8 (right to privacy and family life) may also be relevant to victims if called on to testify at court. It has been suggested by Doak (2008) that the treatment experienced by victims of rape and sexual assault under cross-examination in adversarial courtrooms could raise potential issues under both of these provisions.

Controversies: Victims' Rights and Defendants' Rights

The idea that the rights of victims and offenders are strictly oppositional is reflected by frequent resort to the metaphor of *balance* in political rhetoric and official publications (Hall 2009; Jackson 2003). In the UK, a number of commentators have drawn attention to the fact that a number of policy initiatives, launched in the name of *victims*, have done very little to assist them while actively encroaching upon the rights of the accused (Jackson 2003, 2004). Some commentators have been keen to stress that it does not necessarily follow that the interests of victims and offenders will always conflict (Hall 2009, 2010; Klug 2004; Walklate 2007); Doak (2008) cites the provision of information, support at court, and the provision of good facilities as examples of recent improvement to victims' services which carry little or no impact on the rights of the accused. Indeed, there are a number of scenarios in which the victim and the defendant will share mutual concerns, such as the desire for a prompt and efficient trial process and to be provided with information about procedure.

However, concerns about recent initiatives designed to increase the procedural rights of victims have been expressed (Ashworth 2000; Coen 2006), and both domestic and international case law are replete with conflicts between the rights of victims and offenders. For example, a victim may well object to a decision not to prosecute, to release a suspect on bail, to pursue a particular line of questioning in court, or to impose a community sentence. John Spencer (1994) has argued that the fair trial standards laid down in the Convention are not absolute, and a fair trial does not mean a trial which is free from all possible detriment or disadvantage to the accused. A similar view was expressed by the Strasbourg Court in *Doorson v Netherlands* (*supra*, see para 70).

A number of commentators have argued conceptualizing victims' rights as human rights provides a framework whereby competing rights can be assessed (Doak 2008; Klug 2004; Wolhuter et al. 2009). There are relatively few attempts, however, to devise very specific frameworks where this can be done. Sanders et al. (2010) propose a quasi-utilitarian *freedom model*, which basically states that where competing sets of rights or values are in conflict, those that safeguard the freedom of others to the greatest extent should prevail. Other models of note include Cavadino and Dignan's (1997) *integrated restorative justice* model (Cavadino and Dignan 1997), Doug Beloof's *victim participation* model (Beloof 1999), and Kent Roach's *non-punitive victims' rights* model (Roach 1999). Others have suggested utilizing theoretical models to resolve rights-based conflicts is not necessarily a fruitful exercise and that such conflicts can realistically only be on a case-by-case basis (Doak 2008).

Related Entries

- ▶ Bullying
- ▶ Victim Input at Sentencing
- ▶ Victimization, Gender, and the Criminal Justice System
- ▶ Victims and Restorative Justice
- ▶ Victims and the International Criminal Court
- ▶ Victims of State Crime

Recommended Reading and References

Ashworth A (2000) Victims' rights, defendants' rights and criminal procedure. In: Crawford A, Goodey J (eds) Integrating a victim perspective within criminal justice. Ashgate, Aldershot

Beloof D (1999) The third model of criminal process: the victim impact model. Utah Law Rev 2:289–330

Booth T, Carrington K (2007) A comparative analysis of victim policies across the Anglo speaking world. In: Walklate S (ed) Handbook of victims and victimology. Willan, Cullompton

Brienen M, Hoegen H (2000) Victims of crime in 22 European criminal justice systems: the implementation of recommendation (85) 11 of the council of Europe on the position of the victim in the framework of criminal law and procedure. Wolf Legal Productions, Niemegen

Cavadino M, Dignan J (1997) Reparation, retribution and rights. Int Rev Victimol 4:233–253

Christie N (1977) Conflicts as property. Br J Criminol 17:1–15

Coen R (2006) The rise of the victim – a path to punitiveness? Ir Crim Law J 16:10–14

Dignan J (2005) Understanding victims and restorative justice. McGraw-Hill/Open University Press, Maidenhead

Doak J (2008) Victims rights, human rights and criminal justice: reconceiving the role of third parties. Hart, Oxford

Doak J, Taylor L (2012) EU tools for the protection of victims of serious and organised crime. In: Ruggieri S (ed) The protection of fundamental rights in criminal proceedings. Springer, Amsterdam

Ellison L (2001) The adversarial process and the vulnerable witness. Hart, Oxford

Erez E, Roeger L (1995) The effect of victim impact statements on sentencing patterns and outcomes: the Australian experience. J Crim Justice 23:363–375

Freckleton I (2004) Compensation for crime victims. In: Kaptein H, Malsch M (eds) Crime, victims and justice: essays on principles and practice. Ashgate, Aldershot

Garland D (2001) The culture of control: crime and social order in contemporary society. Oxford University Press, Oxford

Geis G (1990) Crime victims – practices and prospects. In: Lurigio A, Skogan WG, Davis RC (eds) Victims of crime: problems, policies, programs. Sage, Newbury Park

Hall M (2009) Victims of crime: policy and practice in criminal justice. Willan, Cullompton

Hall M (2010) Victims and policymaking: a comparative perspective. Willan, Cullompton

Jackson JD (2003) Justice for all: putting victims at the heart of criminal justice? J Law Soc 30:309–326

Karmen A (1992) Who's against victims' rights? St John's J Leg Comment 8:157–175

Kearon T, Godrey B (2007) Setting the scene: a question of history. In: Walklate S (ed) Handbook of victims and victimology. Willan, Cullompton

Klug F (2004) Human rights and victims. In: Cape E (ed) Reconcilable rights? Analysing the tension between victims and defendants. Legal Action Group, London

Koffman L (1996) Crime surveys and victims of crime. University of Wales Press, Cardiff

Manikis M (2011) Navigating through an obstacle course: the complaints mechanism for victims of crime in England and Wales. Criminol Crim Justice 12:149–173

Mawby R, Walklate S (1994) Critical victimology: international perspectives. Sage, Thousand Oaks

Miers D (2000) Taking the law into their own hands: victims as offenders. In: Crawford A, Goodey J (eds) Integrating a victim perspective within criminal justice. Ashgate, Aldershot

Roach K (1999) Due process and victims' rights: the new law and politics of criminal justice. University of Toronto Press, Toronto

Roberts J, Erez E (2004) Communication in sentencing: exploring the expressive and the impact model of victim impact statements. Int Rev Victimol 10:223–244

Roberts JV, Manikis M (2010) Victim impact statements at sentencing: exploring the relevance of ancillary harm. Can Crim Law Rev 15:1–29

Sanders A, Young R (2007) Criminal justice. Oxford University Press, Oxford

Spalek B (2006) Crime victims: theory, policy and practice. Palgrave-Macmillan, Basingstoke

Spencer J (2004) Criminal procedure: the rights of the victim versus the rights of the defendant. In: Cape E (ed) Reconcilable rights? Analysing the tension between victims and defendants. Legal Action Group, London

Walklate S (2001) The victims' lobby. In: Ryan M, Savage S, Wall D (eds) Policy networks in criminal justice. Palgrave Macmillan, London

Walklate S (2007) Perspectives on the victim and victimisation. In: Walklate S (ed) Handbook of victims and victimology. Willan, Cullompton

Williams B, Goodman H (2007) The role of the voluntary sector. In: Walklate S (ed) Handbook of victims and victimology. Willan, Cullompton

Wolhuter L, Olley N, Denham D (2009) Victimology: victimisation and victims' rights. Routledge-Cavendish, London

Video Surveillance

► CCTV and Crime Prevention

Video Technology and Police Interrogation

David Dixon
Faculty of Law, University of New South Wales,
Sydney, NSW, Australia

Overview

The audio-visual recording of police interviews with suspects provides a good example of several important issues in the relationships between technology and policing. Beyond matters of technological development and deployment, this entry notes the social and political reception by police of new technology; the relevance of the social meanings of technology as a potential solution to policing problems; problems which are raised by over-reliance on technological solutions; and problems in the social reading of images produced by audio-visual technology.

Key Issues

Technological Development and Deployment

The recording of questioning has a long history before introduction of the electronic recording which is the focus of this entry. In the thirteenth century, technologies (record-keeping procedures, synopses, archives, and indices) were developed by those responsible for the Inquisition along with bureaucracies to store and retrieve records of procedures, hearings, admissions, and confessions (Murphy 2012: 40–43). Interest in using electronic recording in police interviews tracked the development of accessible, reliable, and simple recording equipment. Serious interest in the topic emerged in the 1960s and field trials and informal local use occurred in a number of places (Baldwin 1992; Geller 1992). However, widespread systematic use emerged in response to concerns about policing practices in the 1980s.

In a number of jurisdictions, courts and legislatures have required the use of electronic recording, notably in response to recognition of miscarriages of justice. Notably, audio-recording was one of the products of the reform of criminal procedure in England and Wales resulting from the Confait Inquiry and the Philips Royal Commission (Dixon 1997). The Police and Criminal Evidence Act 1984 required police manually to record contemporaneous notes from the time of the legislation's implementation in 1986 until audio-recording equipment and procedures could be made available. In Australia, the various federal, state, and territory policing agencies introduced general audio-visual recording in the 1990s: a High Court decision restricting the admission of confessions and admissions which were not electronically recorded hurried this process along (Dixon 1997: Chap. 5).

The availability of reliable, economical audiotaping made a crucial difference in consideration of this issue. Judicial willingness to allow police the benefit of the doubt in admitting contested confessions waned. There was, it could be argued, no need for any such doubt if electronic recording facilities could provide an indisputable record of what was said. It then found legislative standing in a series of federal and state reforms of criminal procedure. There were similar developments in New Zealand and Canada. In the USA, there has been widespread adoption of electronic recording across the vast archipelago of policing agencies. Illinois, Maine, and the District of

Columbia legislated to require electronic recording of custodial interrogation in homicide investigations. In addition, taping is required in Alaska and Minnesota as a result of court rulings (Sullivan 2010). Such recording is often limited to a rehearsed confession rather than the whole interview. (Problems with this approach are addressed below.) Other jurisdictions are drawing on these experiences as reforms of criminal process are considered: for example, Japanese authorities are currently looking with interest at the audio-visual recording experience in Australia.

Rapid developments in technology facilitated the spread of recording. Notably, the replacement of bulky, relatively fragile VHS tapes with digital recording has made electronic recording much more manageable. (The basic responsibility to store recorded VHS tapes created great difficulty in some stations.) Jurisdictions have differed in their choice of audio or audio-visual recording. England and Wales rely primarily on audio-recording, despite recurrent sporadic interest in audio-visual recording and various field trials. As well as cost and equipment reliability, there has been concern about the interpretation of visual images (an issue to be addressed below). By contrast, audio-visual recording has been the norm from the start in Australia.

Parochialism in development has been a common theme. In decentralized jurisdictions such as Australia and the USA, there has been considerable inefficiency in development as neighboring agencies go through similar development processes independently. Lack of knowledge and policy sharing, conflicts and jealousies, and the career benefits to those seen to innovate rather than follow neighbors have been common. Unfortunately, there have been similar trends at international level. Notably, in the USA, academics and criminal justice professionals continue to produce numerous discussions of the possibility of audio-visual recording as if its introduction would be novel: there is little recognition that several Australian jurisdictions have been using audio-visual recording since the early 1990s, not just in field trials, research experiments, or selected cases, but routinely for questioning about all indictable offenses. Simple ignorance is not the issue: some such commentators have attended conferences at which Australian practice has been discussed. An American colleague acerbically explained this as a product of the belief that any innovation from outside the USA was either second-rate, socialist, or, most likely, both.

In Anglo-American discussions of audio-visual recording, technology provided a simple solution to one of the common concerns about recording police questioning, that the record would be subject to interference and alteration. VHS and audiocassette tapes were inscribed with a timeline, making any editing clearly detectable. Even more simply, recording machines were developed which simultaneously recorded several tapes, one of which was given to the suspect on completion of interview. Similar security procedures apply to digital recording. No doubt, interference with a recording may exist as a technological possibility: however, this has not been a practical concern in jurisdictions which have implemented general recording regimes. Digital recording provides an additional level of security. In sophisticated, connected digital systems, it is increasingly difficult to delete a record permanently and comprehensively. One of the less-noted scandals in the Guantanamo story was the destruction of tapes recording the torture and coercive interrogation of detainees. Such interference with justice processes should be more difficult today.

Emerging technology should solve one current problem, resulting from the involvement of people in the process. In England and Wales, significant problems have been reported in the accuracy of synopses and transcriptions of interviews (Baldwin 1993; Gudjonsson 2003: 86, 114). In the state of New South Wales, Australia, police rely on transcripts produced by an external contractor. This introduces a new realm of potential problems, with concerns over the security and reliability of such contractors. As the accuracy and reliability of voice recognition technology improves, the need for human transcription diminishes.

The Problem Which Technology Is to Address

Police interrogation has long been a source of problems and controversy in criminal justice around the world. Police malpractice has ranged from unintentional inducement of false confessions, to fabrication of confessions ("verballing"), to torture. Sometimes of equal concern to the authorities have been allegations of abuse which have been false or unverifiable, but which cause delay in the justice process and harm the reputation of police. A series of connected responses developed in English-speaking countries. In the 1960s, the US Supreme Court interpreted the Constitution to require protection of suspects' rights, notably through access to legal counsel (*Miranda v Arizona* 384 US 436, 1966). In the 1980s, detailed statutory regulation of custodial interrogation was introduced in England and Wales via the Police and Criminal Evidence Act 1984. In Australia, the High Court developed the law of evidentiary admissibility, and Commonwealth and state statutory regulation was introduced (Dixon 1997: Chap.5).

In the wake of these responses, official concern about interrogation and confessions waned, partly because it was thought that judicial and statutory responses had been adequate, partly because of the shift in public concern from due process to crime control, partly because other issues became more fashionable for policy-makers, grant-funding agencies, and academic researchers. However, such concern has been sharply revived in this century as false confessions emerged as a significant source of the miscarriages of justice which have been disclosed by the use of DNA analysis. This is particularly the case in the USA, where the limits of judicial supervision even in capital cases have been exposed by the acknowledgment of a mass of miscarriages of justice through use of DNA. False or coerced confessions have been a significant contributor to the wrongful convictions which have become a national scandal (Gross 2008; Justice Project: 2007). Similarly in England, some of the contentious disputes over alleged miscarriages of justice stemming from false confessions have been resolved by DNA analysis.

There has been a common theme in many responses to these controversies: police interrogation should be audio-visually recorded. Indeed, electronic recording is frequently presented as a solution to the ills of custodial interrogation. Interest in such recording is not new: there had been calls for its use from the time that recording equipment was widely available. These inquiries did not just identify the problem: they also usually reported that a solution was available in the form of electronic recording. In 1987, McClintock and Healey observed that in "Australia over the last 20 years, virtually every report which has touched on the law relating to criminal investigation" had recommended the tape recording of interviews and that the "same concerns have been echoed in nearly every common law jurisdiction across in the world" (1987: 7). However, the contemporary calls for audio-visual recording are more widespread, united and urgent than before. They draw on the widespread familiarity with electronic recording as a means of social control via CCTV (Newburn and Hayman 2002).

Notably, calling for the use of electronic recording has become a standard component of proposed programs to avoid miscarriages of justice in Canada (FPT Heads of Prosecutions Committee 2004; PPSC 2011: Chap.6) and the USA (Drizin and Reich 2004; Leo 2008). A prominent example was provided in 2003 by the state of Illinois. In response to concern about the execution of people who had been wrongfully convicted, Illinois required police to electronically record interviews with murder suspects. The reform was designed "to restore the integrity of the criminal justice system." (Governor Rod Blagojevich, quoted "Ill. Law 1st to order taping murder confessions" *USA Today*, 18 July 2001, 3A.) All too often, electronic recording is put forward as a panacea. There is little consideration of how it will deal with the problem: it is taken for granted that it will.

Police Attitudes to Electronic Recording

In the 1980s, many police officers expressed outright opposition to mandatory recording schemes, stressing the practical problems of audio or audio-visual recording: it was claimed that electronic recording would be expensive, mechanically unreliable, difficult to arrange in remote locations and open to manipulation by suspects pretending they had been assaulted by investigators during audio-recording or playing to the camera in audio-visual recording. (Some objections were dealt with the way electronic recording was introduced: notably, the incorporation of video recording met the concern that suspects would pretend they were being assaulted.) These instrumental police objections to electronic recording are only part of the story. They were often vehicles for expressing opposition to proposals seen as a slight on their integrity and an inappropriate interference in their business. Specifically, the interrogation room has been the heart of private police territory and interrogation practice seen as a confined, confidential police concern. Having outsiders intrude both symbolically and instrumentally into the interrogation room touched raw nerves among many police.

However, such opposition is usually short-lived. Experience in several jurisdictions suggests that most fears about electronic recording of interviews with suspects were exaggerated or misplaced. Police reluctance to adopt electronic recording inevitably raised suspicion that they had something to hide. It also sat uncomfortably with their enthusiasm for other technological developments, such as wiretapping. As appears to be typical, police resistance to taping faded away as officers were reassured about its detrimental effect, or simply came to accept electronic recording as part of normal, everyday practice. Police also came to appreciate the benefit of electronic recording: they could rebut accusations of malpractice and (when using audio-visual) present to courts images of the defendant as they were when arrested, which often contrast unfavorably with that of the polite, well-dressed person in the dock. Police officers and prosecutors routinely were enthusiastic about the court being able to see the contrast between the neatly dressed, polite defendant in the dock and the scruffy, abusive suspect shown in the audio-visual record. The outcome is that police and prosecutors are much more favorable towards electronic recording than defense lawyers. More worryingly, police have come to rely on images for other purposes.

Reading Images and Detecting Deception

In Australia, there was concern that using visual images in evidence could be problematic, considering whether "a record of things such as tattoos, speech, mannerism, dress, demeanor and language" might be prejudicial to some defendants. This was balanced by advantages to suspects, such as showing the pressure applied by police or the drug-affected condition of the suspect (CLRD 1986: 15). However, the CLRD did not foresee that the problem would be not responses to objectively identifiable matter such as dress and tattoos, but subjective interpretations of behavior – the reading of "body language" in order to draw inferences and, particularly, to detect deception. This meant that a potential problem of using video was (and continues to be) underestimated.

In the USA, an industry has grown providing training to police in detecting deception. The major interrogation trainers include deception detection as a key part of interrogation training. The availability of videotaped records of interview encourages such training, but also allows prosecutors, judges, and juries to participate in deception detecting. In Australia, some judges accepted unscientific claims about detection with remarkable ease. Judges and prosecutors are more likely than police to believe that demeanor is an indicator of veracity (Dixon 2007). Such judicial enthusiasm led to pressure for improvement in audio-visual technology picture quality, and, most significantly, the introduction of technology providing an image alternating between a general picture of the interview room and those present to a close-up of the suspect.

Alternating images have both advantages and disadvantages. The most obvious advantage is that for the first time the viewer can see a large, clear image of the suspect's face during the interview. After years of (at times frustrating) attempts to make out how the suspect looks (Are his/her eyes closed? Is he/she falling asleep during some questions? How serious an injury is that mark on the forehead which is a blur from the distance? Is she visibly alcohol affected?), it is good to be offered such a large clear image of his or her face. The size and clarity of this image of the suspect greatly reduces what previously may have remained in the realm of guess work.

Among the disadvantages are that other persons present in the interview are only seen briefly, if at all. The camera records the whole interview table only for some 20 s every 3 min, before reverting to the face of the suspect. For most of the time, the interviewers are not on screen. If audio-visual technology is to be used as a mechanism of supervision and accountability of interviewing officers, something is lost by focusing on the suspect. There is a relatively simple technological solution to this aspect of the problem. Replacing audio-visual recorders with units including two cameras which could produce split image or "picture in picture" images would allow simultaneous recording and presentation of both the suspect's face and the room as a whole. However, this would not deal with the problem of misinterpretations of images.

This is not the place for a review of the extensive psychological literature on this topic. It is sufficient for present purposes to point out that the research evidence clearly establishes that, whatever a highly trained psychologist may be able to do in detecting deception, a judge (or indeed prosecutor, jury or police officer) cannot do so accurately, and that standard interview training does not increase the capacity to correctly identify deception (Vrij 2008). The widespread dissemination of schlock psychology through magazine articles or brief professional education courses is a matter of real concern, indicating the need for a vigorous program of appropriate education and training for criminal justice professionals, including judges and prosecutors, in any jurisdiction considering the use of video to record interviews with suspects.

The value of good psychological research in this area is shown by studies by Lassiter et al. (1986, 1992) which show how changes in the angle and direction of video-recording significantly influence understandings of interaction and assessments of police and suspects. This reinforces the crucial lesson that technology is not neutral: its social and psychological impacts are as important as its technical details.

The Context of Recorded Interrogation

In the USA, electronic recording has set a significant challenge to criminal investigation practice. The dominant approach to police interrogation, the Reid Technique, emphasizes the psychological benefits to police of isolating suspects. Employees of Reid & Associates explain that interrogation involves psychological pressure on the suspect via "pretense and duplicity. It is therefore understandable that investigators were reluctant to publicly disclose these techniques to criminals who may benefit greatly from this knowledge and thus fortify resistance to telling the truth" (Buckley and Jayne 2005: 1). Perhaps a more substantial concern should be that judges, juries, and the general public who saw the whole interrogation process might disapprove of investigators who "may use vulgar language, may have physical contact with the suspects, may look at the suspect in an intimidating fashion, may make reference to incriminating, sometimes fictitious evidence against the suspect" (Buckley and Jayne 2005: 42) and other types of psychological coercion which the Reid Technique involves. If investigators fear criticism if their techniques were to be seen by courts, an obvious corollary is that they should not use them. If the introduction of comprehensive audio-visual recording leads to US police following their colleagues in the UK and Australia in abandoning interrogation techniques which have been shown to produce unreliable confessions, this would be a beneficial effect of the technology (Dixon 2010).

By contrast, in Australia, when audio-visual recording was introduced, it was presented as a technology which would deal with the problems – real and alleged – of police questioning by allowing outsiders to see what happened in the police interview room. It was taken for granted that the full interview, not just a rehearsed confession would be recorded. Through this clear "objective" window, the court should be able to see how a defendant came to confess.

As McConville comments, “What seems to be on offer, for judges, lawyers and juries, is the chance to have the past replayed, enabling the viewer to look on as reality is being constructed. It is that promise which invests the video with such persuasive character” (1992a: 548). If, however, the recorded interaction is the product of earlier, unrecorded questioning, then video’s promise may be illusory. The danger that video may give a false gloss of authenticity is real: simply by virtue being on camera, the interview seems more compelling and tends to be accepted as legitimate.

Anyone who feels complacent about police questioning practices in an age of audio-visual recording would do well to read McConville’s disturbing account (1992a) of how some English police officers evaded controls on the questioning of suspects and were able to present audio-visually recorded accounts which gave no indication of the unrecorded misconduct in “interviews” which preceded them. While these were not officially recorded, they were captured by cameras and microphones installed in the station as part of a documentary project by a television company. The police force was experimenting with the use of audio-visual recording. In England and Wales, a system based on audio-taping was introduced in the later 1980s: there has been spasmodic interest in videotaping (Newburn et al. 2004). Apparently voluntary confessions given in bland interviews were shown to have been produced by deals, threats, and inducements. The official record of these interviews gave a misleading account of what occurred in a way that would have been convincing had not an unofficial record been available.

McConville argues that the misrepresentation effected by incomplete recording jeopardizes suspects, and that, far from protecting suspects’ rights, electronic recording undermines them. He warns that “where the police make threats or inducements or strike deals with suspects in private which then lead to a confession in the formal interrogation, the position of a complaining suspect will be weakened rather than strengthened by the supporting videotaped record of the confession because of its apparent ability to capture reality” (1992b: 962). Responding to similar earlier concerns in Australia, the Criminal Law Review Division had warned of the danger that electronic recording might be used to record rehearsed interviews, and consequently recommended that all questioning should be recorded (CLRD 1986:17, 41–42).

Much of this preparatory interviewing is innocuous. Far from the dramatic myth, many police investigations and interviews are mundane and not contentious. An assumption that unrecorded questioning routinely involves attempts to coerce or persuade suspects into confessing distracts attention from a less dramatic reality. Typically, officers talk to suspects to find out how they will respond to formal questioning and to prepare themselves for a formal interview. From this perspective, much informal interviewing is part of the process of planning and preparing for the formal, recorded interview. At its simplest, it tells an officer how much work he or she is going to have to do for the interview to be successful. These comments seek to be realistic about the nature of everyday police investigations without being complacent. It is recognized that, as McConville’s examples showed in England, an audio-visual record may present a completely misleading picture from which unreliable, unfair, and inappropriate tactics used during preparatory interrogation are obscured.

Equally, even if such tactics are not employed, a suspect or defendant could allege that they were. The result could be to revive the costly and damaging cycle of allegation and denial which electronic recording was intended to kill off. Dixon’s (2007) research reported some confessions which may have been obtained entirely properly, but the recording of which raises rather than dispels doubts. In one, a suspect confessed on a tape to a long series of armed robberies, including several for which he had apparently not been a suspect. On tape were merely the bland confessions, with no indication of the circumstances of their production. The investigating officers were at pains to record the suspect’s statement that he had not been offered any inducement. However, incidental references to access to legal advice and entry to a witness protection program indicated that these were important factors in the production of his confession. The potential for conjecture, and for lengthy legal dispute, about the reliability of such confessions is evident.

Competing claims have been made about whether malpractice preceding formal interviews can be detected from observing tapes. On one hand, McConville asserts that "it is not possible to tell from the video recording whether suspects have been the subject of improper pressure" (McConville 1992b: 962). Those of us who are skeptical about the ability of police interviewers to detect deception should be modest in their own claims that they can identify deception by police officers (Dixon 2007). Baldwin is slightly more optimistic, suggesting that "a recording is valuable in offering some insight into what has happened when a suspect is questioned and in providing a means by which an assessment might be made of whether a suspect has been bullied or primed beforehand" (Baldwin 1993: 328). Identifying pre-recording police misconduct was not straightforward, even with the availability of a visually recorded police interview. The audio-visual record may be valuable in indicating how a suspect has been treated earlier, but it is by no means conclusive (Dixon 2007).

Baldwin also suggests that "the techniques of discourse analysis have already been used in the courts in challenges to various forms of confession evidence, and there is no reason why they could not be used to good effect to expose indications of earlier conversations from the transcripts of formal interviews" (1992: 1096). Discourse analysis will, of course, only be a resort available to a small minority of suspects. For most of those who confess and plead guilty, the prospect of obtaining linguistic experts to analyze their interview will be remote.

It would be naive to think that the need to produce an audio-visual record exerts no influence on officers' behavior, or that suspects can be coerced into agreeing to anything. Even officers who are prepared to coerce suspects have to be confident that any mistreatment of, pressure on, or deal struck with a suspect is going to be effective enough to ensure that there is no embarrassing outburst when the interview is recorded. Equally, it would be naive to suggest that it is only police who prefer some interaction to be unrecorded. Some suspects may be prepared to speak informally, but not to cooperate during the recorded interview. Some detectives suggest that many experienced criminals will talk to the interviewer informally, but will not cooperate when what they are saying is recorded (Dixon 2007). Perhaps more significantly, suspects may well be reluctant to talk on the record about other people's involvement in offenses, or indeed their own involvement in offenses other than that for which they were arrested.

Concentrating on coerced confessions which have been obtained before the audio-visual recorder is activated would divert our attention from more mundane but significant issues. It would be valuable to pay closer attention to cases in which such interviewing is openly acknowledged. For example, from a psychological perspective, the compliant and responsive role allocated to the suspect in procedures for adopting previous questions and answers may have significant effects. Dixon reports one extreme instance in which a suspect was asked no less than 96 "Do you agree . . . ?" questions in 15 min, all of which were answered "Yes." The repetition of questions in this form is highly conducive to compliance. Psychological and linguistic analysis may demonstrate the subtle reconstruction of statements in these processes. Almost inevitably, a DYA question will contain the officer's paraphrase of the original exchange even if an attempt at contemporaneous note-taking is made: a suspect's cursory signature confirms the semblance of quoted speech.

This point must not be overstated. It is certainly better that a suspect is asked to "adopt" on tape a confession or admission which has been made away from recording facilities than that police are permitted to give evidence of unrecorded confessions and admissions. This was vividly demonstrated by a case in which the High Court of Australia unfortunately declared admissible police evidence of an unrecorded, incriminatory comment made by a suspect in police station car park soon after the conclusion of a recorded interview in which he had denied the offense. (*Kelly* [2004] HCA 12. By contrast with this narrow legalism, the High Court adopted a purposive approach in *Nicholls and Coates* [2005] HCA 1.) This approach simply invites process corruption.

Responding to the Problem of Unrecorded Questioning

It has been suggested here that unrecorded, preparatory interviewing is usually a matter of routine. Nonetheless, it threatens the integrity of the system. An audio-visual recording shows that a suspect made confessions or admissions, not how he or she came to do so. As noted above, establishing the reliability of a confession is harder than merely proving that it was in fact made. Given what we know from many jurisdictions about deliberate misconduct and inadvertent influence by interviewers and about apparently irrational responses by suspects, there is no room for complacency. It is necessary to have as much questioning recorded as possible. The problems of recording field interrogations mean that questioning should be conducted in police stations wherever feasible. This should not put unrealistic demands on police. Claims that spontaneous outbursts make recording impracticable echo arguments from the 1980s that electronic recording would be impossible. The response now should be same as then: of course, exceptions must be allowed, but these must be in defined circumstances and/or subject to rigorous scrutiny.

So long as much interviewing is conducted before the recorder is activated, there will be room for controversy about what happened. Such controversy includes doubt about the reliability of recorded confessions. The potential benefits of electronic recording are dissipated if it is used to record rehearsed material. It should be stressed that, from all the evidence available, the costs and problems to police of comprehensive recording are minimal. If a police officer feels uncomfortable about using an interviewing technique on tape, then that technique may well not produce reliable results. The costs are minimal, but they exist. As explained above, there are going to be occasions when recording is impossible or inappropriate – for example, when a suspect insists that he/she will not name an accomplice while being recorded. This is not an unusual dilemma in policing: the objective is the minimization of problems, not some problem-free utopia. Police should record all questioning of suspects conducted within police stations and should only interview suspects in police stations (except in cases of exceptional need which fall within specified categories). Wherever possible, suspects should be asked to repeat unrecorded confessions on tape. If they are not asked to do so (or their refusal to do so is not recorded), there should be grave suspicion about such confessions.

Dealing with partial recording requires a more general remedy. What is needed is a renewed commitment to the legal regulation of policing by the development of rules, policies, and standards (Dixon 1997: Chap. 7). This does not mean more rules, a message which would find favor with no-one. It means having better rules which become "working rules," (i.e., part of the cultural and other norms which guide everyday working practice) rather than "inhibitory rules" (which are effective only if there is an immediate prospect of their enforcement) or "presentational rules" (whose main purpose is to placate a public audience).

Such rules must be made with legislative authority, not left to the police to produce. None the less, police should be directly involved in the production of the rules in this as in other areas (Dixon 1997: Chap. 7). While courts will play an important role in interpreting and enforcing rules, they cannot be expected to take the leading role in regulating policing. In their different ways, the experiences of the USA and Australia demonstrate that judicial control is inadequate because it depends on the vagaries of case law, which does not allow for detailed prospective regulation.

Legal regulation should establish positions from which a variety of pressures are put on the investigatory practice of police officers. Audio-visual recording is just one of those potential pressures. Others include proficient, well-resourced legal advisers (and, for vulnerable suspects, social workers trained to take the role of appropriate adult); rules of evidence in the hands of judges and magistrates who are prepared to be active in the control of policing; and senior officers who are prepared to supervise in order to ensure that investigators work within the rules and use approved techniques for questioning suspects. None of these is a panacea or a silver bullet. While it is as foolish to think of them as such, it is equally foolish to reject one or the other on the ungrounded assumption that it will not change police practices. Progress may be possible through the combination of various (admittedly flawed) mechanisms, of which audio-visual recording is one.

Conclusion

This assessment of electronic recording of police questioning of suspects raises much broader and more complex issue of regulating police practice. Audio-visual recording is not enough by itself: it must be used as a tool in a general regime of regulation. The recorded interview is just one stage in a suspect's detention. Its reliability and propriety depend substantially on legal regulation of the context in which interviewing takes place. Such problems can only be tackled by much more rigorous regulation of investigative practices and, in particular, by requiring that (with the caveats noted above) all interviews should be electronically recorded in full. There are obvious incentives for officers to question suspects before a formal recorded session. If electronic recording is to have a significant role in controlling police interviewing and ensuring the reliability of confessions by providing more than confirmation of what a suspect said in a rehearsed interview, then effective legal and supervisory regulation of investigative practices is necessary. Audio-visual recording offers significant benefits to criminal justice, but is no panacea. It can even be counterproductive if it instills a false confidence that all is well. Indeed, there is an urgent need to acknowledge and deal with other problems in the regulation of custodial interrogation.

Related Entries

- ► Criminal Investigative Analysis
- ► False Confessions and Police Interrogation
- ► Forensic Linguistics
- ► Investigative Psychology
- ► Law of Police Interrogation

Recommended Reading and References

Baldwin J (1992) Videotaping police interviews with suspects – an evaluation, vol 1, Police research series paper. Home Office Police Department, London

Baldwin J (1993) Police interview techniques: establishing truth or proof? Brit J Criminol 33:325–352

Buckley DM, Jayne BC (2005) Electronic recording of interrogations. John E Reid & Associates, Chicago

CLRD (1986) A proposed system of electronically recording police interviews with suspected persons. Criminal Law Review Division, Sydney

Dixon D (1997) Law in policing: legal regulation and police practices. Clarendon, Oxford

Dixon D (2006) A window into the interviewing process? The audio-visual recording of police interrogation in NSW Australia. Policing Soc 16:328–348

Dixon D (2007) Interrogating Images: Audio-visually Recorded Police Questioning of Suspects. Institute of Criminology, Sydney

Dixon D (2010) Questioning suspects: a comparative perspective. J Contemp Crim Just 26:426–440

Dixon D, Travis G (2007) Interrogating images: audio-visually recorded police questioning of suspects. Institute of Criminology, Sydney

Drizin SA, Reich MJ (2004) Heeding the lessons of history: the need for mandatory recoding of police interrogations to accurately assess the reliability and voluntariness of confessions. Drake Law Rev 52:619–646

FTP Heads of Prosecutions Committee (2004) Report on the Prevention of Miscarriages of Justice. Ottawa: Ministry of Justice, Canada

Geller WA (1992) Police videotaping of suspect interrogations and confessions: a preliminary examination of issues and practices. National Institute of Justice, Washington, DC

Gross SR (2008) Convicting the innocent. Ann Rev Law Soc Sci 4:173–192

Gudjonsson G (2003) The psychology of interrogations and confessions. Wiley, Chichester

Intelligence Science Board (2006) Educing information: interrogation science and art: foundations for the future. National Defense Intelligence College, Washington, DC

Justice Project (2007) Electronic recording of custodial interrogations: a policy review. The Justice Project, Washington, DC

Kassin SM, Drizin SA, Grisso T, Gudjonsson GH, Leo RA, Redlich AD (2009) Police-induced confessions. Law Hum Behav 34:3–38
Lassiter GD, Meissner CA, Ware LJ, Lindberg MJ, Ratcliff JJ (2010) Videotaping custodial interrogations. In: Lassiter GD, Meissner CA (eds) pp 143–60
Lassiter GD, Meissner CA (eds) (2010) Police interrogations and false confessions. American Psychological Association, Washington, DC
Lassiter GD, Slaw RD, Briggs MA, Scanlan CR, Irvine AA (1986) Videotaped confessions: the impact of camera point of view on judgments of coercion. J Appl Soc Psychol 16:268–276
Lassiter GD, Slaw RD, Briggs MA, Scanlan CR (1992) The potential for bias in videotaped confessions. J Appl Soc Psychol 22(23):1838–1851
Leo R (2008) Police interrogation and criminal justice. Harvard University Press, Cambridge
McClintock I, Healey A (1987) Getting it taped: recording police interviews. In: Zdenkowski G, Ronalds C, Richardson M (eds) The criminal injustice system: volume 2. Pluto, Sydney, pp 5–41
McConville M (1992a) Videotaping interrogations: police behaviour on and off camera. Crim Law Rev 532–548
McConville M (1992b) Video taping interrogations. New Law J 11 May:960–962
Murphy C (2012) God's jury: the inquisition and the making of the modern world. Houghton Mifflin Harcourt, Boston
Newburn T, Hayman S (2002) Policing, surveillance and social control: CCTV and police monitoring of suspects. Willan, Cullompton
Newburn T et al (2004) The introduction of visual recording of police interviews with suspects. Unpublished, LSE/Kent University
PPSC (Public Prosecution Service of Canada) (2011) The path to justice: preventing wrongful convictions. Report of the federal/territorial/provincial subcommittee on the prevention of miscarriages of justice, PPSC, Ottawa
Sullivan TP (2010) The wisdom of custodial recording. In: Lassiter GD, Meissner CA (eds). pp 127–142
Vrij A (2008) Detecting lies and deceit. Wiley, Chichester

Violent

▶ Social Disorganization and Terrorism

Violent Pornography

▶ Pornography

Virtual Prison

▶ Electronic Monitoring

Virtual Repeats

▶ Prediction and Crime Clusters

Visualizing Data: A Brief History

Michael D. Maltz
Department of Sociology, Criminal Justice Research Center, The Ohio State University, Columbus, OH, USA
Emeritus, Department of Criminal Justice and Department of Information and Decision Sciences, University of Illinois at Chicago, Chicago, IL, USA

Overview

This essay describes the benefits and uses of data visualization. It provides an overview of its history, from the pioneering work of William Playfair to the more recent Web-based activity.

Introduction

Among the many attributes of the criminal justice system, one that stands out is the sheer quantity of data it generates. Since Western governments do not (generally) have secret arrests, trials, or imprisonments, almost every transaction is recorded in some database, most of which are computerized – and large. And these databases, although subject to restrictions on privacy and confidentiality, are (generally) open to public – and researcher – scrutiny and analysis.

Analyzing these data sets can be very useful, to look for patterns in offender and offending characteristics and to determine the effect of different policies on outcomes. As the criminal justice system focuses more and more on "evidence-based" policies, developing practical methods of analyzing these large data sets becomes more important. One very useful set of analytic tools for this purpose falls under the heading of data visualization techniques.

One problem with using these techniques is lack of training in their use. Rare is the criminology/criminal justice curriculum that has courses in data visualization. [One possible exception in North America is the criminology program at Simon Fraser University].

Consequently, data visualization has yet to be implemented to any great degree in criminology and criminal justice. This is the case despite the fact that criminal justice agencies provide a great deal of data suitable for graphical analysis. In fact, so much data exists in this area that it often makes it difficult to apply sample-based statistical methods: every test is statistically significant!

This essay first describes the pros and cons of using data visualization techniques. Then some examples of their use are presented. Following this, the early (pre-1950) work in this area is discussed, which leans heavily on works by Wainer (1997, 2005, 2009), and on an early history of graphical representation of data (Funkhouser 1937).

This is followed by some recent examples of visualized data, showing how it can be used to obtain a better understanding of its attributes. Of particular benefit is its use in checking data to insure its correctness and integrity, described next. Its early history is then covered, followed by a description of more recent work in data visualization. Following the conclusion, the reference section includes a partial listing of websites that contain interesting examples of visualized data. (Please note that websites are not necessarily permanent, so some links may cease to exist after a time).

Disadvantages and Advantages of Data Visualization

As previously mentioned, there is a noticeable lack of training in visualization techniques compared to statistical analysis techniques. This is in part due to the lack of a standard set of procedures for visualizing data, because the characteristics of the data should dictate the visualization technique. This means that the analyst needs to have an idea of how to present the data, and how to put that presentation into practice.

That often requires that the analyst first do some experimenting by hand, determining how best to portray the data. Once an idea presents itself, the analyst often has to do some computer programming, sometimes simple, other times more complex. This is probably the main reason that data visualization has not been used or taught to any great degree: normally one can't just import the data into a statistical package (e.g., the four Ss of social science – SAS, SPSS, STATA, SYSTAT), click on a few menu choices, and generate a beautiful and insightful plot. The analyst needs to consider the nature of the data, figure out how best to portray it visually, and write a sequence of commands to manipulate the data to achieve that result. That is, data visualization is more of an art than a (codified) science.

In addition, the lack of standard techniques means that the visual analyses can't be compared directly with other studies; each analysis is *sui generis*. This means that there is no particular number or figure of merit ("statistical significance") that provides a – supposedly – objective measure of the result of the analysis (Maltz 1994). Thus, it may work against the gradual accumulation of studies that then give weight to a particular understanding of a pattern in the data.

There are substantial advantages as well. Unlike other methodologies used to analyze data, data visualization requires no assumptions about the characteristics of the data: they need not be linear, normal, independent, homoscedastic, or have (or be assumed to have) any specific statistical characteristics. That is, one can consider the data without any preconceived notions or requirements. Data visualization permits the data to tell its own story. [After this essay was in final draft, an article was published

(Barbe and Horace 2012, 1) that underscored this advantage. In analyzing automobile searches conducted by the Maryland State Police, they showed that the "complex dynamics" in the data "preclude the use of standard statistical inference techniques," and instead "show that simple graphics are well-suited to capturing the nuances of the data."]

It may take the analyst some effort to determine how to best represent the data, but on the plus side of the ledger, and as will be seen in the following examples, the data becomes easy to interpret. This is in contrast to manipulating the data using a statistical package; in that case it is easy to run the program and get results, but it may mask some major anomalies in the data. While using a standard statistical method may be a good approach when testing specific hypotheses, there is an implicit assumption that the data fits the method and the method fits the data. These assumptions are often made without actually inspecting the data, especially when dealing with very large data sets.

This is where visualization comes into play. It is always useful to plot some of the data's variables, looking (with one's eyes) for errors and inconsistencies that too often creep into data sets, while at the same time discovering the unique characteristics of the data. One needn't be constrained by assumptions about the data that may facilitate analysis but are rarely true – rare is the behavioral variable that has a normal (or lognormal, or Poisson) distribution nor are there any behavioral variables that are linearly related to each other. Despite this, most statistical procedures assume that relationships among the variables are linear and additive, like the recipe for a cake – take two parts poor education, one part impulsivity, and three parts poverty and you cook up a delinquent.

Instead, data visualization permits the analyst to discover for oneself what relationships exist within the data. There are no quick-and-dirty (or cookie-cutter) techniques that can be applied; data visualization is more of an art than a specific set of procedures, and each analyst may have different ways of depicting data so as to bring out its qualities.

Examples

One of the best examples of visualizing data is found at http://www.ted.com/talks/view/id/92. Since it is a video, it cannot be put between the pages of a book, but it is well worth viewing. In it Hans Rosling animates a number of time series to tell a story of how infant mortality and other health-related characteristics of countries have changed over the past 50 years.

Another example, not as impressive or important, of how baseball "closers" (pitchers who pitch the last few innings to "save" games) compare, is depicted in Fig. 1. [A brief explanation for those not familiar with America's (so-called) pastime: the pitcher is the person on a baseball team who exerts the most control over the number of points scored by the opposing team. Toward the end of a game in which a team's point advantage is three or fewer, the team's current pitcher is often replaced by a "closer," a pitcher who, it is hoped, prevents the opposing team from adding a sufficient number of points to win, thereby closing out the opposing team and "saving" the team's win. Figure 1 compiles the cumulative success record of these closers over time]. Note how the figure shows:

- who first achieved a particular milestone (X00 saves),
- when the pitchers were active,
- their career length,
- how injury affected one pitcher's (Trevor Hoffman's) trajectory, and
- how fast they accumulated their saves.

Websites and books that describe this data visualization technique (and others) are included at the end of this essay. This graphic was created using R, a free graphics-oriented statistical program, one used by many analysts for visualizing data. A large number of specialized programs written in R are available on the Web, as are complete instructions of how to program in R. However, a graph like this can also be generated by Excel and other spreadsheet programs. The horizontal axis is the date (apparently only dates during the baseball season are included) and the vertical axis is cumulative number of saves.

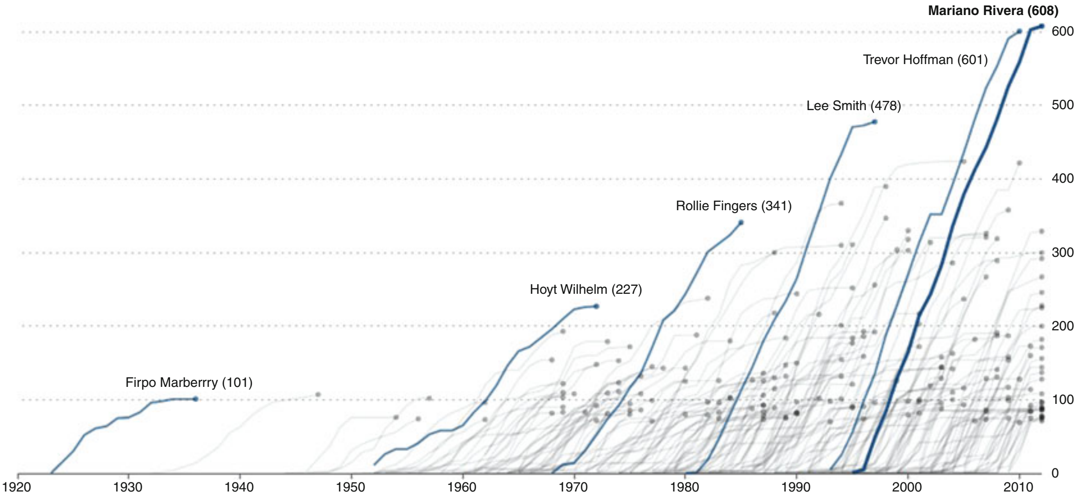

Visualizing Data: A Brief History, Fig. 1 How Mariano Rivera compares to baseball's best closers (NY Times, May 5, 2012)

Visualizing Data: A Brief History, Table 1 Sex-specific correlation between age of victim and age of offender

		Sex of offender	
		Male	Female
Sex of victim	Male	0.300 (N = 170,969)	0.501 (N = 31,140)
	Female	0.483 (N = 60,717)	0.470 (N = 6,528)

Another example (Maltz 2010) is found in criminal justice: suppose an analyst is interested in the relationship between age of victim and age of offender for homicide cases in which there was one victim and one offender. [Note that this is a biased sample with unknown bias, since it includes only those cases for which the offender (and his/her age) is known].

A simple correlation is run and a reasonably strong correlation (0.393) is found between the victims' and offenders' ages. But after partitioning the data by victim's and offender's sex, some stronger correlations are found (Table 1).

Obviously, statistical significance is not a useful indicator, because with such large Ns all of the correlations are "significantly" different from each other.

But then four crosstabs (age of victim vs. age of offender by sex) are generated, smoothed, and plotted (Fig. 2). The vertical axis is the number of cases, which are plotted against the age of offender (axis heading southwest) and age of victim (axis heading southeast). [Because the M-on-M peak is so, over 50 times higher than the corresponding F-on-F peak, a square root transformation was applied to the vertical axis to enhance visualization].

It shows that certain patterns of each of the four graphs are similar but others are different. All have a peak (a "homicide hill") at around age 20 for both offender and victim, indicating that most homicides occur between the young (mostly males), generally peers killing peers. There is also an "infanticide hill" on the left side of the graphs, probably committed by offenders we normally call "caretakers," who are generally between ages 15 and 40. Between these two hills is what may be called "the valley of the shadow of death," considering the paucity of victims between ages 6 and 12: the victims are no longer endangered by caretakers and not yet endangered by peers and others.

The strong diagonal ridges in the M-on-F and F-on-M figures imply that those deaths may be

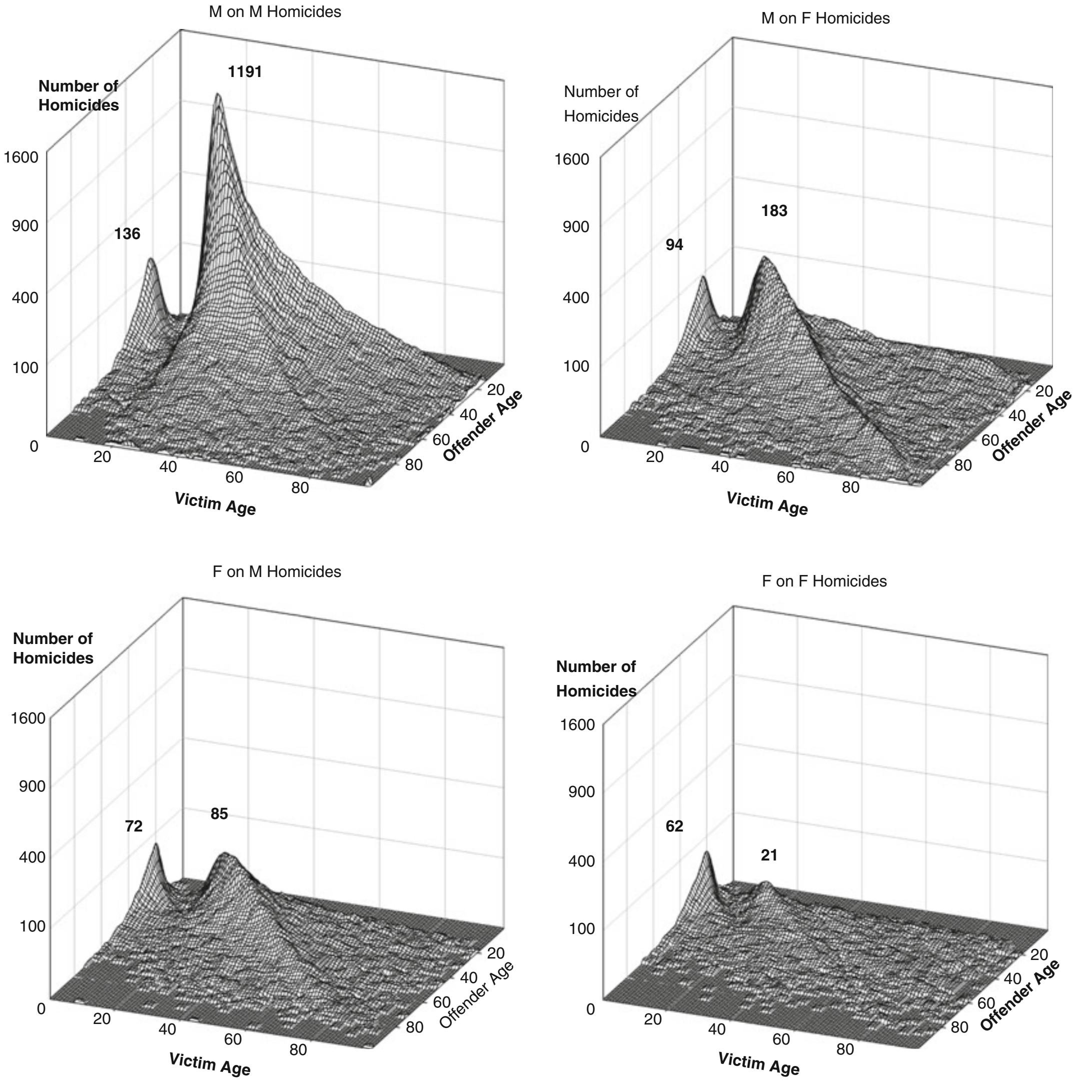

Visualizing Data: A Brief History, Fig. 2 Smoothed SHR data, 1976–2005, disaggregated by sex of victim and sex of offender; vertical axis plotted using a square root transformation (Maltz, 2010)

attributable to intimate partner homicide. Also, while not as evident without rotating the figures, there are horizontal ridges in the M-on-M and M-on-F figures, which indicates that victims of both sexes and all ages are being killed by young males, many perhaps during the commission of a felony.

These data plots were generated using SPSS to generate the crosstabs, Excel to make and smooth the original plots, and SigmaPlot to prettify the figures. In other words, this visualization is (or should be) well within the ability of most criminologists and social scientists, even those who cannot program.

These simple visual analyses of a data set should show the advantages of visualizing data: while it may not be useful for testing hypotheses, it is incomparable as a means of understanding patterns in the data. The patterns that emerge from the data are hidden when using standard analytic methods, since those standard methods require, in some sense, standard types of data. Moreover,

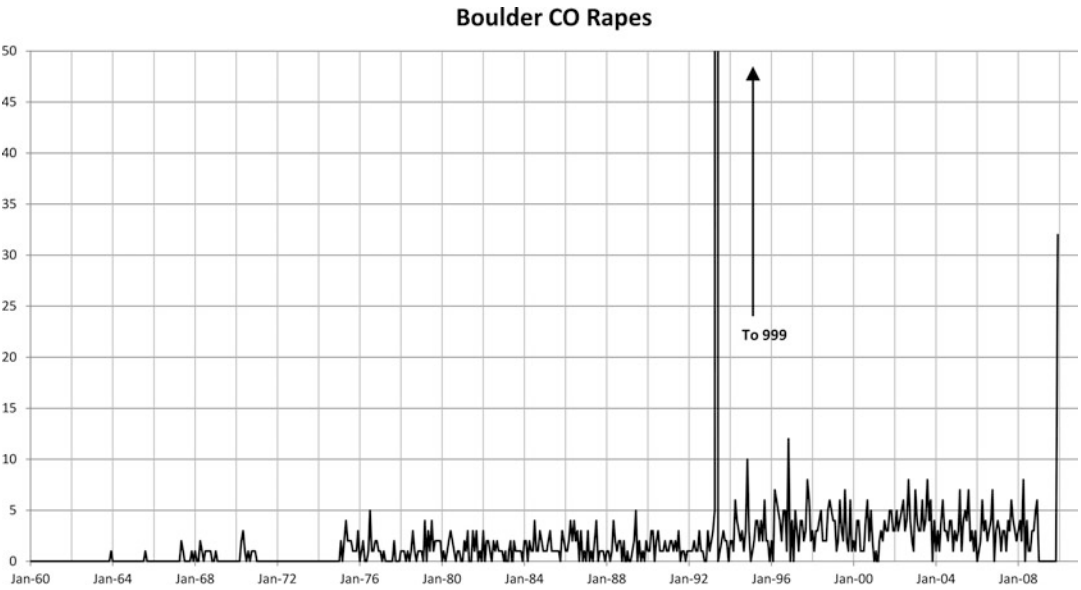

Visualizing Data: A Brief History, Fig. 3 Boulder rape data, showing a missing datum coded as 999 (FBI, Uniform Crime Reports, 1960–2009)

since the data set is usually dropped into the software package without first inspecting it for inconsistencies (and with the hope that outliers will emerge from the analyses), there is no guarantee that the data set does not contain inconsistencies.

Other Uses

Data checking. "Some of the hardest errors to detect by traditional methods are unsuspected gaps in the data collection (we usually discovered them serendipitously in the course of *graphical checking*" (quoted in Unwin et al. 2006, 19). This is one of the most beneficial uses of data visualization.

Sometimes a graph can help spot an error, as in Fig. 3, where an outlier in Boulder CO rape data was coded as "999," which some statistics packages use to designate a missing datum. [Note also that Boulder's 2009 rape data was not given by month but was aggregated for the entire year (32) and reported in December of that year].

At other times an outlier is a valid datum, as in the 19 April 1995 bombing of the federal office building in Oklahoma City (Fig. 4), which killed 168 of the 173 homicides on that day.

Error-checking is also useful in dealing with life course trajectories. Figure 5 shows the (hypothetical) transactional history of a prisoner. Note that it starts at age 12, so obviously the date of birth was entered incorrectly. That is, the person entering the data probably entered the wrong year; all of the other dates were likely correct, since most correctional systems nowadays enter new transactions automatically.

The figure shows that the prisoner bounced from parole back to prison five times before his first release from custody. Creating a plot of this type can be accomplished pretty easily. All that one needs to do is code changes of status to 0, 1, or 2, reflecting release from prison, release to parole, or imprisonment, respectively. Then one merely creates a graph with coded status plotted against date of status change. [It is slightly more complicated, since the figure needs to reflect two statuses on the same date, too show the transition from the old status to the new one on that date].

To see how visualization can assist in finding missing data in this type of trajectory,

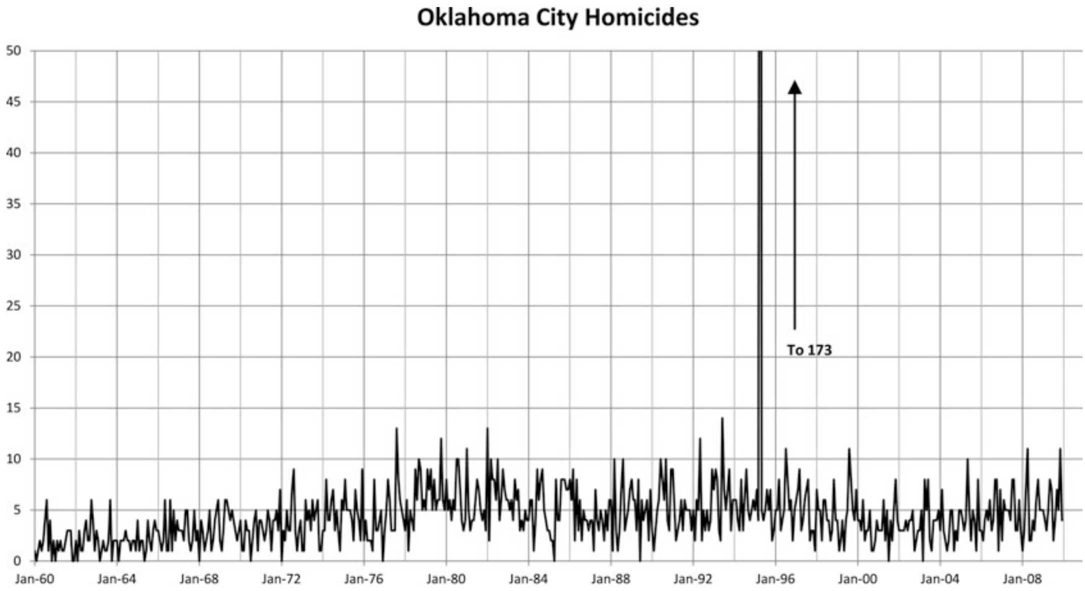

Visualizing Data: A Brief History, Fig. 4 Oklahoma City homicide "outlier" is a true datum (FBI, Uniform Crime Reports, 1960–2009)

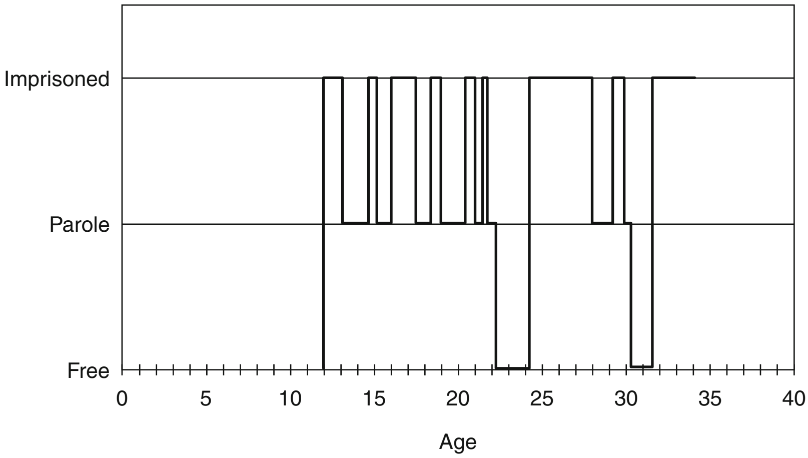

Visualizing Data: A Brief History, Fig. 5 Hypothetical offender trajectory

Fig. 6 shows the same trajectory with a missing date of release from custody.

Crime mapping. Although this topic is covered in other essays, it bears noting as a visualization method. This is one of the more tactical-oriented data visualization techniques, used by police agencies throughout the world. Prior to computerization, police departments spotted trends and "hot spots" with pin maps, different pin colors denoting different kinds of offenses or times of day. The first implementations of computerized crime mapping required the data to be entered by hand into the computer (Maltz et al. 1991). As computers became more prevalent in police departments, and as crime data entry began to include geocoding of offenses, crime mapping and crime analysis became more fully integrated into police work.

Automated crime maps become useful as the collective memory of a police beat. Beat officers

Visualizing Data: A Brief History, Fig. 6 The same offender trajectory showing a missing date of parole completion

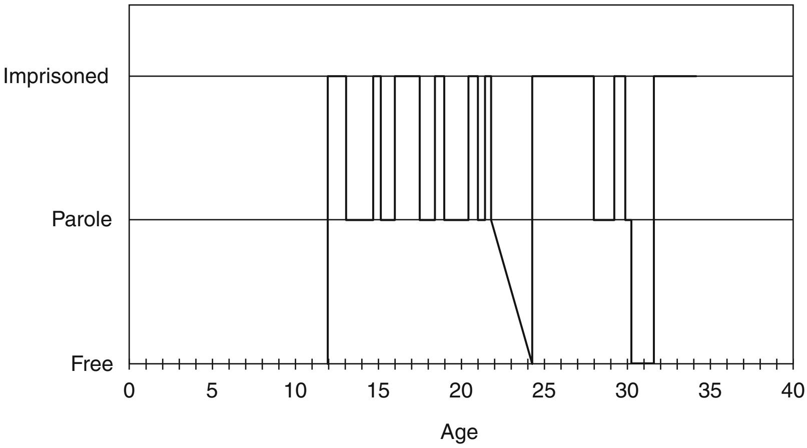

know their beats' activity, but only during their own tours of duty – and a neighborhood's daytime and nighttime activity may be radically different. An officer coming on duty can call up his/her beat map to see the activity the beat experienced during other shifts. Moreover, many police departments give the public access to these maps, allowing people to see for themselves the frequency and nature of crime in their own neighborhoods.

One of the most interesting uses of visualization, especially (but not exclusively) in crime mapping, comes with *animated data*, that is, with data that play out over time, permitting chronological patterns to surface. By judicious choice of time intervals (sometimes hourly, sometimes daily, weekly, or annually), the observer can often find patterns that elude the most sophisticated algorithms.

Additional uses. *Scatterplots* of data can show both relationships and outliers; *networks* can be examined for unusual features; *trellis plots* can uncover features that remain hidden in mixed (continuous and categorical) data. Examples of these are given in Maltz (2010).

Early History

There is a long history of using visualizations to convey meaning from person to person, stretching back to antiquity. As Funkhouser (1937, 270) notes, cave dwellers drew pictographs on cave walls, and symbols in some writing systems were originally based on graphical representations of objects. In other words, pictures seem to be central to human cognition.

The geography of graphs and coordinate systems also has a long history. Specifically, latitude and longitude were used by early Egyptian surveyors in locating points on the earth (Wainer and Velleman 2001, 306). And orthogonal coordinates were used by Nicole Oresme in the fourteenth century to represent the variation of functions (Funkhouser 1937, 274), antedating their use by René Descartes (and the eponymous Cartesian coordinates) by three centuries.

The most prominent early use of visualization techniques in inferring patterns from data was produced by William Playfair in the late eighteenth century; his book, *The Commercial and Political Atlas and Statistical Breviary*, depicted year-to-year trends in economic data. Wainer and Spence (2005) republished the 1801 version of Playfair's treatise, bringing to light one of the classics in data visualization. In it they note (pp. 23–27) "only a small sample of Playfair's innovations," which include:

- the time series line graph,
- the divided series area chart,
- the bar chart,
- gridlines,
- using color to distinguish different patterns,
- labeling of axes,

and many other features of graphs that we now take for granted. In his book Playfair (Wainer and Spence 2005, p. xii) noted that "as much information *may be obtained in five minutes as would require whole days to imprint on the memory*, *in a lasting manner*, *by a table of figures*" (italics in the original), thus describing one of the major advantages of graphs, the ability to present a time series that shows how the relationship varies over time: whether a sequence is rising or falling, or behaving in some other manner.

Moreover, when another time series is plotted on the same coordinate axes, it can indicate how one series affects the other. Many of his figures are plots of exports to and imports from England by other countries, clearly showing when the balance of payments was positive or negative.

Before long the use of graphs went beyond depicting only financial data, and soon turned to crime. Funkhouser (1937, 299) notes that Adolphe Quetelet, a Belgian statistician, "appears to have been the first to represent graphically a frequency distribution in the drawing of a curve showing the percentage of crimes committed at various ages." Quetelet called this type of analysis "social physics." Graphical analysis of crime data was used by another statistician, Frenchman André-Michel Guerry, as well:

> Guerry's *Essai sur la statistique morale de la France* (Paris, 1833) which was honored by a medal from the French Academy, is a splendid example of how much the presentation of a subject can be enhanced by the use of well-made graphical illustrations. Several new types of graphs make their appearance in this work. One is a chart in which crimes are listed in order of frequency for various ages. Colored lines trace the shifting position of a particular crime from one age group to the next. (Funkhouser 1937, 304)

One of the more famous data visualizations is Charles Minard's graphical (and geographical) portrayal, in 1869, of the severe attrition of Napoleon's army (Fig. 7) as it marched toward Moscow in 1812 and retreated in 1813, having been beaten by "General Winter," the fearsome subzero temperatures and accompanying snowfall that decimated the French troops. As Marey (1878, 73) wryly states, this figure attains "a degree of brutal eloquence" that "seems to defy the pen of the historian." A number of different visualizations using Minard's data have been collected by Michael Friendly (under the heading "Re-Visions of Minard") at http://www.datavis.ca/gallery/re-minard.php, showing how others may depict the same data.

Another early use of graphics was produced by Florence Nightingale during the Crimean War. A nurse with mathematical training, she was appalled by the lack of sanitary conditions experienced by wounded soldiers. She wrote a report that contained circular charts, with 12 segmented wedges (Fig. 8). The wedges portrayed the monthly number of deaths due to military action, and those due to the unsanitary conditions she encountered as a nurse at the front (Wainer 1997, 103). By detailing the number of preventable deaths, her report was instrumental in improving sanitation during and after the war.

Recent History

The first half of the twentieth century saw an upsurge in the development of algorithmic methods of analyzing data, to some extent to the exclusion of visual methods. The names Galton, Pearson, Kendall, Yule, Gosset, and Fisher are associated with these methods.

The first modern treatment of visualization-based methods was Jacques Bertin's book, *Semiologie Graphique*, published in 1967. It was an attempt to develop a schema of visual sign systems that are used in understanding data. It was translated into English (*Semiology of Graphics*) in 1983; the translation was reissued in 2010.

Bertin's book focuses primarily on the structure of graphs and the perceptual processes that are used to understand them, e.g., the use of color, texture, and line weight, more than it does on data analysis. In much the same way that most westerners can't decipher the sign systems encoded in Middle Eastern (or Asian or Mayan) alphabets (or syllabaries or symbols), a person with no knowledge of or experience in deciphering graphs and other visual representations of data would have a hard time interpreting the lines, dots, and symbols used to visualize data.

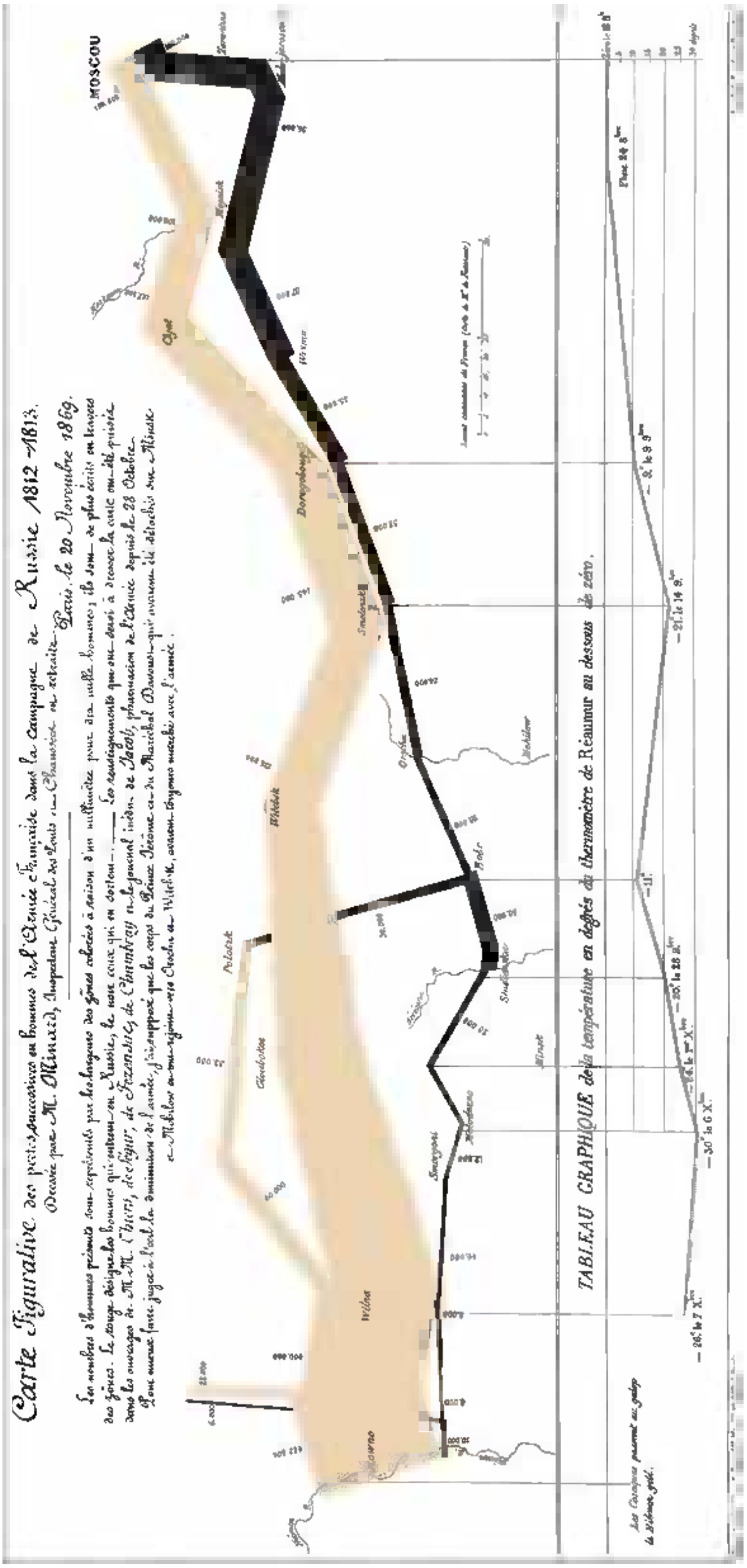

Visualizing Data: A Brief History, Fig. 7 Minard's graphic of Napoleon's march to (*tan*) and from (*black*) Moscow, detailing the temperature on the return from 18 October to 7 December 1812. Downloaded from http://upload.wikimedia.org/wikipedia/commons/2/29/Minard.png, accessed March 30 2013

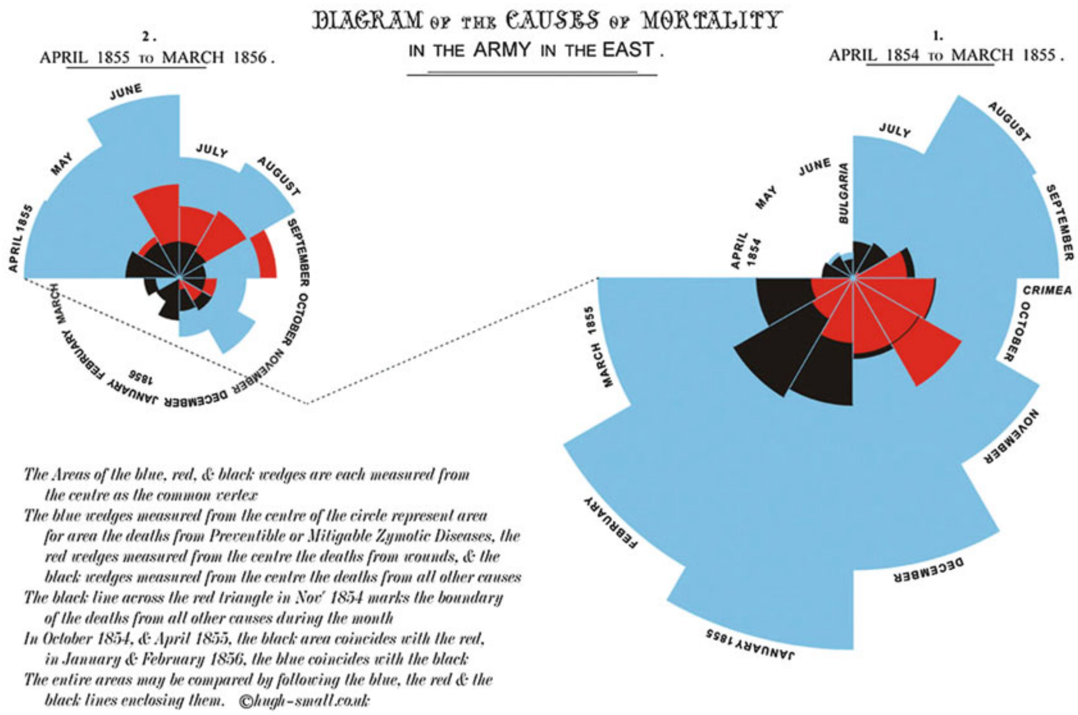

The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex
The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic Diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes
The black line across the red triangle in Nov' 1854 marks the boundary of the deaths from all other causes during the month
In October 1854, & April 1855, the black area coincides with the red, in January & February 1856, the blue coincides with the black
The entire areas may be compared by following the blue, the red & the black lines enclosing them. ©hugh-small.co.uk

Visualizing Data: A Brief History, Fig. 8 Florence Nightingale's "rose" (Note: Number of deaths is proportional to the radius lengths, not to the areas of the wedge)

Bertin tries to systematize the construction of graphic representations. He deals with how the eye and the brain translate marks on a piece of paper to convey meaning. This requires familiarity with the "language" of graphics, e.g., the fact that a line connecting two points means that they are associated in some way with each other. He deals with the selection of shapes and their shading, colors, hue and saturation, orientation, texture, and size, and how they may be used to denote different variables.

The second milestone in modern graphical analysis and data visualization was the publication of John Tukey's 1977 book, *Exploratory Data Analysis*. Although it predates the current ubiquity of digital computers – he notes (p. x) that the included problems "can be done with a pencil and paper" – Tukey described a number of visual techniques that are useful in understanding data sets. They include stem-and-leaf plots, box-and-whisker plots, data smoothing procedures, and binning procedures. In addition, the book includes numerous examples of how to deal with numerical data in ways that guide the reader to become more numerically astute, explaining, for example, how to think about logarithms, how to develop scales for axes, and how to use residuals in interpreting data. Much of this book bears retelling, including a number of his propositions (Tukey 1977, 157, emphasis in the original):

- **Graphs are friendly**.
- **Graphs force us to note the unexpected; nothing could be more important**.
- **Different graphs show us different aspects of the same data**.
- **There is no more reason to expect one graph to "tell all" than to expect one number to do the same**.

His book also gave lie to the belief by many social scientists that looking at the data before running analytic tests is "cheating" or "data snooping." In fact, as Wainer and Velleman (2001, 314) point out, "Tukey's Exploratory Data Analysis changed things. Suddenly terms like data snooping, data dredging, and the currently trendy 'data mining' were no longer pejorative."

For example, there is no way that a statistical procedure can convey the richness of interpretive findings produced by graphing the data, as in Fig. 2.

Tukey's book was soon followed by what is considered by many to be the best primer on data visualization, Edward Tufte's *The Visual Display of Quantitative Information* (1983). Using examples collected from many countries over the centuries, from newspapers, academic journals, and business and government documents, Tufte provides details of the frequency with which data has been misrepresented in graphs. Among his prescriptions to improve the graphical presentation of data are avoiding "chartjunk" (his term for overly ornate figures like a bar chart consisting of dollar bills to represent money), avoiding the use of pie charts, and maximizing the data-to-ink ratio (so that the data points are not overshadowed by other elements in the figure).

Tufte's subsequent books, *Envisioning Information* (1990) and *Visual Explanations* (1997), are further excursions in the realm of data visualization and, as with his first book, continue his efforts to educate both technically sophisticated and lay people about accurate (and attractive) graphical presentation.

In more recent years, the work of Cleveland (1993, 1994) provides useful examples for those who actually need to develop graphs themselves. His description of how R. A. Fisher apparently reversed the years in one of his data sets – and never noticed it nor did others who analyzed the same data set – again underscores the utility of graphical analysis as a first step. His visualization of this data set also shows the benefit of using trellis plots in analyzing data containing both categorical and continuous variables.

Those who want to understand how good graphs can tell a story and poor graphs can misrepresent data could do no better than read the books by Howard Wainer. His *Visual Revelations* (1997), *Graphic Discovery* (2005), and *Picturing the Uncertain World* (2009) contain entertaining and telling examples of appropriate and inappropriate uses of graphs, sometimes to deliberately muddy the waters but in most cases because of inadequate attention to the data or to the needs of the audience.

There have been instances of researchers who used standard statistical tests on large data sets and, finding every test to be "significant," decided to sample the data. However, an analyst who decides to select, for example, a 10 % sample of the data is essentially throwing out the other 90 % (which offends data mavens!). A better alternative would be to take advantage of the entire data set to "drill down" to find relationships and other characteristics that are often missed when using sampling statistics. That is, unnecessary sampling can be seen as equivalent to squinting at an object instead of opening one's eyes fully; it permits the viewer to see some broad features, for example, the mean . . . but at the expense of missing the details, the very features that make research an exciting art rather than a formulaic "science." Unwin et al. (2006) describe a set of procedures and software useful in visualizing large data sets and provide links to some of those data sets. But however rigorous the analyses, there are no hard and fast rules that work at all times for all cases.

The generation of large data sets – of business transactions, of Web searches, and of criminal activity and correctional information – has led to an increasingly strong interest in data mining. This has tended to bring data analysis and visualization techniques out of the academic world and into the business and policy world. As a result, a number of books on data visualization have been written to fill this market gap. They include Few (2004), Ware (2004), Segaran and Hammerbacher (2009), and Yau (2011).

One book worthy of special note is Wilkinson's *The Grammar of Graphics* (1999, 2010). Like Bertin's (2010) book, it aims to put graphics on a firm footing, codifying it and grounding it in underlying principles. However, it focuses on how to put together a graph's elements – data, transformations of and operations on the data, specification of graphic elements, scales, and coordinate systems – to create a visual representation of the data. As the original author of SYSTAT and its graphics package, and later of the SPSS graphics package, Wilkinson has intimate knowledge of efficient coding, and uses this knowledge to detail the different components of a graph.

Conclusion

Over the past few decades the use of visualization techniques to elucidate patterns in data has been growing at breakneck speed. Its growth parallels the changes in computing over the same time period. At first, computers were basically (international?) business machines, useful in developing and analyzing quantifiable data. Input to these machines consisted of oblong punched cards, laboriously generated by programmers and read by high-speed card readers; output of these machines consisted of reams of accordion-folded paper, through which the programmers had to sift to find their results (or, more frequently, the coding errors in the programs they wrote that rendered the results useless).

[They also produced sound as well as paper. I remember in the early 1960s watching a computer analyst hold a transistor radio, detuned from any station so as to pick up and amplify stray signals, near the IBM 1620 we were using, trying to find the error in a program. The pitch of the squeal the radio picked up told him something about the length of the infinite loop in which the program seemed to be stuck].

In the late 1970s computer aficionados began to buy home computers, using televisions or monitors as well as paper to write programs and view results. They began to draw pictures that were of low resolution but adequate for some purposes (such as simple computer games).

Computer visualization and crime mapping soon became more useful when the first Apple Macintosh came out in 1984, with a relatively high-resolution screen (e.g., Maltz et al. 1991, 154). Fast forward to today, when smart phones have substantially greater power and higher resolution and more flexibility, and at a much lower cost.

Isaac Newton once wrote to a colleague, "If I have seen further it is by standing on the shoulders of giants." We can certainly extend that remark to this topic: if we can visualize patterns in data to a greater extent, it is because we are standing on the shoulders of the data analysts, statisticians, and computer scientists who preceded us and provided us with their own visions of data accessibility and understanding.

Related Entries

- ▶ Crime Mapping
- ▶ Network Analysis in Criminology
- ▶ Spatial Models and Network Analysis

Recommended Reading and References

Barbe P, Horace WC (2012) A critical reanalysis of Maryland State Police searches. Am Stat 66:1–7

Bertin J (2010) The semiology of graphics. ESRI Press, Redlands, Initially published in 1983 by University of Wisconsin Press, Madison, WI, a translation (by William J. Berg) of Semiologie graphique, published in 1967 by Editions Gauthier-Villars, Paris, France

Cleveland W (1993) Visualizing data. Hobart Press, Summit

Cleveland W (1994) The elements of graphing data. Hobart Press, Summit

Few S (2004) Show me the numbers: designing tables and graphs to enlighten. Analytics Press, Oakland

Funkhouser HG (1937) Historical development of the graphical representation of statistical data. Osiris 3:269–404, http://www.jstor.org/stable/301591

Maltz MD (1994) Deviating from the mean: the declining significance of significance. J Res Crime Delinq 31:434–463

Maltz MD (2010) Look before you analyze: visualizing data in criminal justice. Chapter 3. In: Piquero AR, Weisburd D (eds) Handbook of quantitative criminology. Springer, New York

Maltz MD, Gordon AC, Friedman W (1991) Mapping crime in its community setting: event geography analysis. Springer, New York

Marey É-J (1878) La méthode graphique dans les sciences expérimentales et principalement en physiologie et en médecine. G. Masson, Paris, Available at http://vlp.mpiwg-berlin.mpg.de/library/data/lit3585?

Segaran T, Jeff H (2009) Beautiful data: the stories behind elegant data solutions. O'Reilly Media, Sebastopol

Tufte ER (1983) The visual display of quantitative information. Graphics Press, Cheshire

Tufte ER (1990) Envisioning information. Graphics Press, Cheshire

Tufte ER (1997) Visual explanations. Graphics Press, Cheshire

Tukey JW (1977) Exploratory data analysis. Addison Wesley, Reading

Unwin A, Theus M, Hofmann H (2006) Graphics of large datasets: visualizing a million. Springer, New York

Wainer H (1997) Visual revelations: graphical tales of fate and deception from Napoleon Bonaparte to Ross Perot. Copernicus (Springer), New York

Wainer H (2005) Graphic discovery: a trout in the milk and other visual adventures. Princeton University Press, Princeton

Wainer H (2009) Picturing the uncertain world: how to understand, communicate and control uncertainty through graphical display. Princeton University Press, Princeton

Wainer H, Velleman PF (2001) Statistical graphics: mapping the pathways of science. Annu Rev Psychol 52:305–335

Wainer H, Spence I (2005) Editors of Playfair W (1801) the commercial and political atlas and statistical breviary. Cambridge University Press, New York

Ware C (2004) Information visualization: perception for design, 2nd edn. Morgan Kaufmann, San Francisco

Wilkinson L (1999) The grammar of graphics 2010, 2nd edn. Springer, New York

Yau N (2011) Visualize this: the FlowingData guide to design, visualization, and statistics. Wiley, Indianapolis

http://addictedtor.free.fr/graphiques/thumbs.php
http://chartsnthings.tumblr.com/
http://colorbrewer2.org/
http://dd.dynamicdiagrams.com/2008/01/nightingales-rose/
http://eagereyes.org/
http://flowingdata.com/
http://infosthetics.com/
http://junkcharts.typepad.com/
http://visualizingeconomics.com/
http://www.datavis.ca
http://www.gapminder.org/
http://www.periscopic.com/
http://www.tableausoftware.com/

Vivitrol

► Drug Abuse and Alcohol Dependence Among Inmates

W

War Crimes

► Human Rights Violations in Criminal Court

What Works in Corrections

► Examining the Effectiveness of Correctional Interventions

White Collar Crime

Nicole Leeper Piquero and Stephen Clipper
The University of Texas at Dallas, Richardson, TX, USA

Synonyms

Corporate crime

Overview

White collar crime has progressed since Edwin Sutherland introduced the term in 1939; his goal was to bring attention to the more affluent criminal element, which was often overlooked in traditional criminological theory. Since its introduction, white-collar crime has become a debated issue in criminological theory. One source of debate is over the definition of white-collar crime. Several studies were published during the 1940s, addressing white-collar crime at the individual and organizational levels. After a hiatus during the 1950s and 1960s, scholarship in this area was revisited using a new definition based on the offense rather than the offender, as had been used in the past. Redefining white-collar crime in this way led to a division in its fields of study; Clinard and Quinney (1973) identified the two fields as corporate crime and occupational crime. Despite the division, the debate is far from resolved. Unresolved issues include crime types and measurement. The former is of particular interest as technological advances promote the development of new crimes. Scholarship in the area of white-collar crime has expanded knowledge about both the offenders and victims of these crimes utilizing official statistics, self-report surveys, and victimization surveys.

Fundamentals

White collar crime is an umbrella term which incorporates various types of unlawful acts and a different type of actor than many people often associate with the concepts of crime and criminals. As a consequence, the term white-collar crime has all too often become a shorthand way of distinguishing specific types of offenders and offenses. For many scholars, the theoretical and empirical study of white-collar crime should focus on powerful elite offenders, and as such,

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

the white-collar criminal forms a polar opposite to the typical offender found in the criminal justice system. For other scholars, white-collar crime studies are designed to examine offenses committed by organizations or corporations (also known as corporate crime). Finally, another group of scholars suggests that white-collar crime studies should more closely mimic other studies of crimes and criminals by focusing on the specific acts that violate a law. There is likely no other term in criminology that has garnered such debate.

Key Issues

Edwin Sutherland, an American sociologist, introduced the term white-collar crime on December 27, 1939, in his presidential address entitled "The White Collar Criminal" to the American Sociological Society in Philadelphia, Pennsylvania (Sutherland 1940). The point of his address was to challenge existing conceptions of the stereotypical criminal and to critique extant criminological theories of the day which he viewed as narrowly focusing on only one part of the crime problem. At that time, crime was viewed as a lower class phenomenon which was committed by disadvantaged young men from broken homes and decaying neighborhoods. Sociologists working at the University of Chicago in the late 1920s and early 1930s argued that there was a link between socially disorganized neighborhoods and crime (Shaw and McKay 1942). During this time period, social disorganization theory was at the height of its popularity, arguing that crime was caused by social disorganization which was most often found in neighborhoods that were poor, had high levels of population turnover, and were predominately comprised of minorities. In due time, the image of the typical criminal began to emerge. In both scholarly writings and through the popular media, the stereotypical image of an offender was that of a tough guy growing up on the wrong side of town. The emergence of other criminological theories, such as classic strain theories, continued to focus on crimes and criminals as originating from the lower classes and did little to alter the public image of a typical offender.

By no means was Sutherland the first person to recognize that crimes were not exclusively a lower class phenomenon. Other scholars and popular writers also noted that crimes and misdeeds were also being committed by businessmen and other respectable members of society. However, crimes committed by such people were seldom seriously considered by those scholars who wrote about or studied crime and were not a major concern of the public or policy makers when addressing the crime problem.

When Sutherland gave his groundbreaking speech to the American Sociological Society, he set out to challenge the traditional image of criminals and the predominant etiological theories of his day. The white-collar criminals he identified were often middle-aged men of respectability and high social status. They lived in affluent neighborhoods and they were all well respected in their communities. He went on to define white-collar crime "as a violation of the criminal law by a person of the upper socioeconomic class in the course of his occupational activities" (Sutherland 1941, p. 112). Despite Sutherland's recognition of the importance of the white-collar crime category, it never achieved the centrality in criminological study that he proposed.

Where Sutherland did succeed was in generating lively discussions and heated debates surrounding the definition of white-collar crime. The ambiguities associated with the lack of a precise definition of white-collar crime can be traced back to Sutherland's own writings. Geis (1992, p. 33), in reviewing the development of the term, notes that Sutherland "buried part of his definition in a footnote," suggesting that this "attests to his indifference to the matter." Sutherland himself would suggest that "this definition is arbitrary and not very precise. It is not necessary that it be precise, for the hypothesis is that white-collar crime is identical in its general characteristics with other crime rather than different from it" (1941, p. 112).

While Sutherland's original – and now commonly considered as an offender based – definition of white-collar crime refers to an individual's social status (e.g., respectability) and an occupational opportunity to commit the crime, many of his earliest examples of white-collar crime came from his study of seventy of the largest industrial and mercantile corporations in the United States (Sutherland 1945). The shift in the unit of analysis, from individual to organization, only added to the conceptual confusion of the concept. Sutherland argued that white-collar crime is crime because it is a violation of law, albeit criminal, regulatory, and civil laws, but a violation of law nonetheless. In order to further illustrate his point, Sutherland (1945) examined court decisions regarding antitrust violations, cases of false advertising, national labor relations infractions, and infringements of patents, copyrights, and trademarks. He found a total of 547 adverse decisions against the seventy corporations he examined and 473 of those decisions related to actual crimes committed. He noted that less than ten percent of the total decisions made it into the criminal court but argued all the same that the violations of regulatory and civil laws were also indicative of a crimes, those that he labeled as white-collar crimes. Legal scholars took issue with Sutherland's (Caldwell 1958; Tappan 1947) inclusion of civil and regulatory violations as crimes as they are not violations of criminal statutes.

While Sutherland had hoped that the introduction of the concept of white-collar crime would ignite new strands of research and broaden the scope of acts considered by criminologists as normative criminal behavior, such was not the case. The study of white-collar crime fell into an inactive state for most of the 1950s and1960s (Geis 1992; Payne 2012). During this period of inactivity, some scholars argued for further restricting or tightening of the definition of white-collar crime. For example, Geis (1962) argued that white-collar crimes should be limited to the realm of corporate violations. This push toward a more restrictive definition of white-collar crime led to a fracture or a divide in the study of white-collar crimes, a distinction that Clinard and Quinney (1973) would later conceptualize with the terms corporate crime and occupational crime.

Research in the area of white-collar crime saw a rebirth of activity in the 1970s when definitions of white-collar crime began to move away from a myopic focus on offenders with high status or respectability and shifted to what is now referred to as an offense-based definition of white-collar crime. Edelhertz (1970, p. 3, 19–20) preferred a broad legally based definition of white-collar crime by defining it as "an illegal act or series of illegal acts committed by nonphysical means and by concealment or guile, to obtain money or property, to avoid the payment or loss of money or property, or to obtain business or personal advantage." He further identified four subdivisions of white-collar crime: (1) personal crimes, crimes by persons operating on an individual, ad hoc basis, for personal gain in a nonbusiness context; (2) abuse of trust, crimes in the course of their occupations by those operating inside businesses, government or other establishments, or in a professional capacity, in violation of their duty of loyalty and infidelity to employer or client; (3) business crimes, crimes incidental to and in furtherance of business operations but not the central purpose of such business operations; and (4) con games, white-collar crime as a business or as the central activity of the business. This offense-based approach to defining white-collar crime was favored by many scholars and practitioners, though some preferred the term "economic offenses" over the already established "white-collar crime" phrase. The United States Department of Justice also utilizes an offense-based definition of white-collar crime, defining it as "nonviolent crime for financial gain committed by means of deception by persons whose occupational status is entrepreneurial, professional, or semi-professional and utilizing their special occupational skills and opportunities also, nonviolent crime for financial gain utilizing deception and committed by anyone having special technical or professional knowledge of business and government, irrespective of the person's occupation." The use of an offense-based definition of white-collar crime also created new avenues for studying and understanding these types of crimes.

The distinction between offender- and offense-based definitions of white-collar crime created somewhat of a fracture in theoretical and empirical scholarship on white-collar crime. Clinard and Quinney (1973) provided a clear distinction between the study of corporate crime and the study of occupational crime. Corporate crime would consist of illegal acts that are committed by an employee or an officer of the corporation or of the actions of the corporation itself when such actions would promote corporate interests or somehow benefit the organization. Occupational crime, on the other hand, was defined as "violation of the criminal law in the course of activity of a legitimate occupation" and would include any persons, regardless of position in the social structure, who commit an illegal act for his or her personal benefit in the course of their occupation (Clinard and Quinney 1973, p. 188).

Many researchers today continue to use the corporate crime versus occupational crime distinction when conducting research. Corporate crime scholars, for example, investigate both organizational defendants as well as individual actors engaging in crimes and misconducts that are designed to further the interests or benefit the organization. On the other hand, those who study occupational crime are more interested in the acts that are committed by individuals for their own personal gain. Corporate crime scholars tend to focus on understanding the criminogenic aspects of the organization and the surrounding business environment and as such, focus attention on examining the organizational context and the role that it plays in the decision-making process. Some have suggested that "corporate crime differs from other kinds of white-collar offenses (e.g., embezzlement) in three main ways: (1) offenders can be official representatives of the company or the company itself; (2) offenses are undertaken in the pursuit of corporate goals; and (3) violators are subject to criminal, regulatory, and civil law" (Simpson and Piquero 2002, p. 510).

Much has been written on the definitional issues surrounding the study of white-collar crime, but this debate is far from being settled. While the emergence of offense-based definitions created new opportunities for studying white-collar crime by using official measures of criminal behavior (e.g., police arrest records), the use of such data also provoked controversy. The Federal Bureau of Investigation's Uniform Crime Reporting (UCR) system records data on white-collar crimes include fraud, forgery, and embezzlement; yet, because offenders arrested for these three crime types do not necessarily need to be employed to carry out the offenses, many critics argue that these are not really white-collar crimes. For example, Steffensmeier (1989, p. 347) argued that "UCR data have little or nothing to do with white-collar crime." The category of fraud, for example, would include those offenders arrested for welfare fraud, check fraud, welfare fraud, wire fraud, and credit card or ATM fraud. Many critics would rather view these crimes as versions of theft rather than including them into the category of white-collar crime.

Further contributing to the definitional dilemma is the invention of new crime types, such as cybercrimes, intellectual property offenses, and the crime of identity theft. These offenses are often studied by white-collar crime scholars but do not nicely fit into the two-pronged categorization created by Clinard and Quinney (1973) as they are neither corporate crimes nor occupational offenses. Crimes such as these, however, would easily fit within one of the four subdivisions as identified by Edelhertz (1970) in his broadly defined definition of white-collar crime. The continued growth of technology and specialized skills sets is only going to further hamper the precise definition of white-collar crime. The irony to this definitional conundrum, which has plagued the study of white-collar crime since Sutherland first sensitized scholars and the general public to its existence, is the unnecessary focus on spurious offender attributes. The focus on identifying offenders by their attributes is precisely the point that Sutherland was fighting against.

Research Directions

For the most part, the study of white-collar crime has copied, mimicked, and otherwise followed the methodologies used to study crime more generally albeit at a much slower pace. The definitional problem plaguing the white-collar crime field has historically made studying this crime type somewhat difficult. As a way around this impediment, white-collar crime scholars have relied upon three primary means of data collection, case studies, official reports, and surveys, to study white-collar crime.

The earliest and most popular methods for studying white-collar crimes came in the form of case studies. The case study approach involves an in-depth study of one single specialized crime, criminal, or event. Data for such an approach are gathered from a variety of sources, including archival data including court records and news accounts, field research, and interviews (Payne 2012). The extant literature is replete with examples of notorious case studies of white-collar crimes and criminals. Perhaps the most notable example of a specific crime is Geis (1967) description of the 1961 price-fixing conspiracies in the electrical industry, and the most notable description of type of specialized offender comes from Cressey's (1953) and later Zietz's (1981) examination of convicted embezzlers.

Until the late 1980s, a statistically valid picture of white-collar crime and offenders was nearly impossible to create because of the dearth of quantitative data on the crimes and criminals. The shift toward an offense-based definition of white-collar crime allowed scholars to use official data sources such as police arrests, court and conviction records, as well as regulatory agency data to study white-collar offenses and offenders. For the first time, data collection efforts in the 1980s permitted a greater understanding of the offenses, offender characteristics, and overall criminal career patterns of white-collar offenders. Most notably, the Yale studies on white-collar crime, which were led by Wheeler et al. (1988; also see Weisburd et al. 1991), defined white-collar crime as "economic offenses committed through the use of some combination of fraud, deception, and collusion." Wheeler and his colleagues proceeded to investigate a sample of 1,342 "white-collar offenders" drawn from persons who were convicted in seven selected Federal United States Districts over a three-year time period for eight different white-collar offenses including antitrust offenses, securities fraud, mail fraud, false claims, bribery, income tax fraud, lending/credit fraud, and bank embezzlement. Data on each offender was collected from presentence report and investigation reports which are prepared by a probation officer after an offender has been convicted for the sentencing judge. These reports contain an abundance of detailed data about both the offense and the offender; such information includes the circumstances of the current offense, the offender's criminal history, as well as demographic and background information of the offender (e.g., level of education, employment status, familial situation). In addition to gathering information on white-collar offenders, data was also collected for a comparison group of offenders convicted in federal court for nonviolent financially oriented property crimes (e.g., postal fraud and postal forgery).

The Yale studies were extremely important and influential because they were among the first set of empirical studies that allowed white-collar crime scholars to investigate whether or not there were identifiable differences between white-collar and street offenders. Wheeler et al. (1988) found differences in the violation of different criminal statutes, differences in the pools of criminal victims, with victims of white-collar crimes more likely to be organizations as opposed to individuals, and differences in the complexity and duration of the offenses, with white-collar crimes involving more than five persons and lasting longer than one year. When examining the offenders themselves, it became clear that white-collar offenders were better educated than the comparison group of common criminals, were more likely to be white, and were less likely to be unemployed. Overall, two conclusions regarding white-collar offenders emerged: (1) the vast majority of convicted white-collar offenders were not, as previously believed, members of the highest social classes; rather, they appeared to be average middle-class citizens, and (2) a substantial number of the convicted white-collar offenders, approximately 40 %, had a criminal record (Piquero and Benson 2004; Weisburd et al. 1990). As Weisburd et al. (1991, p. 73) concluded, "whatever else may be true of the distinction between white-collar and common criminals, the two are definitely drawn from distinctly different sectors of the American population." Finally, as is the case in the general criminological literature, subsequent analyses of the Yale data indicate that there is important variation within the group of white-collar offenders, such that overall offending patterns among individual offenders do not follow in "lock-step" (Piquero and Weisburd 2009).

The use of survey data has also been a popular way to study both street crime as well as white-collar crime and affords researchers much flexibility in the content and the specific types of questions asked. Two common approaches to survey research as it relates to collecting crime data are the use of self-report surveys, that is, directly asking offenders to indicate their involvement in criminal activities, and the use of victim surveys, where victims directly report their experiences with crimes. The use of self-report surveys is a challenging endeavor depending upon the specific criminal offense of interest. For offenses that are rare events, such as many types of both violent crimes and white-collar crimes, the base rate of positive reporting tends to be very low. Therefore, these types of survey approaches are best utilized with normative offenses such as drug and alcohol use and abuse.

The National Crime Victimization Survey (NCVS) provides an alternative method for attempting to describe the amount of crime occurring throughout the United States. The NCVS is a household survey that specifically inquires about the level of victimization experienced by individuals living within the households selected into the study. Unfortunately, the focus of the NCVS has been almost exclusively on street crimes, primarily on crimes such as sexual assault, robbery, burglary, larceny, motor vehicle theft, and vandalism. While the NCVS has not been extremely useful in identifying and documenting the extent of white-collar crime, in 2004, the NCVS conducted a supplementary analysis that specifically asked victims to report on their experiences with identity theft – defined as the unauthorized use or attempted use of existing credit cards, the unauthorized use or attempted use of other existing accounts such as checking accounts, and the misuse of personal information to obtain new accounts or loans. The findings from this survey indicated that identity theft was a type of crime experienced by at least one person in 5.5 % of all households.

In order to gain a broader understanding of victimization experiences with white-collar crimes, the National White Collar Crime Center has conducted a series of national household surveys (conducted in 1999, 2005, and 2010) entitled, "National Public Survey on White Collar Crime" that examine a variety of different types of fraud. The most recent study asked respondents to report whether someone in their household had been victimized in the past twelve months of any of the following offenses: credit card fraud, price misrepresentation, unnecessary repairs, Internet monetary loss, identity theft, fraudulent business venture, false stockbroker information, or mortgage fraud. Findings showed that approximately 24 % of the respondents indicated that someone within their household experienced at least one type of victimization (Huff et al. 2010). Other more specialized databases collect information regarding the victim's experiences with specific white-collar offenses. For example, the Federal Trade Commission (FTC) collects information in their Consumer Sentinel database on offenses such as complaints of fraud and identity theft, while the Internet Crime Complaint Center (IC3) focuses on Internet-facilitated crimes.

One final survey approach to understanding white-collar crime is the use of public perception surveys. This methodological approach allows researchers to understand how the general public and even specialized groups such as criminal justice personnel view crimes relative to one another. This line of research tends to ask respondents, by phone or in-person surveys, how they perceive and/or view certain crimes, their seriousness, and how they believe each should be punished. The underlying focus of this line of research is to try to determine which crime types are deemed as "serious" by the survey respondents and also to gain an understanding of what the public believes the criminal justice system should do and how they should respond to such crimes. This information is seen to be important by many scholars because it has the potential to provide policy makers with a better understanding of public sentiments and lend some insight into which issues should be addressed first and possibly describe some potential courses of corrective action.

Measuring public perceptions of white-collar crimes has closely followed the existing criminological scholarship that examines the ranking of street crimes in terms of perceptions of the seriousness of the crime. Sellin and Wolfgang (1964) conducted the first public perception study that ranked the crime seriousness of a variety of crimes and found consistent responses across the seriousness ratings especially as they related to violent crimes. More recent public perception studies tend to show that Americans consistently rank white-collar crimes just as serious as street crimes (Kane and Wall 2006; Piquero et al. 2008; Rosenmerkel 2001; Schoepfer et al. 2007).

Summary

Historically, white-collar crime had been largely regarded as a deviant case, invoked primarily to provide a contrast to the common crimes and street criminals that dominate the field's theory and research about crime and criminals. As such, the study of white-collar crime has not developed at the same pace or with the same specific level of detailed methodology that has been evinced in the larger criminological literature. For many scholars, this approach focused on an understanding of white-collar crime that is not only distinct from crime more generally but has led some observers to believe that the study of white-collar crime can contribute little to our understanding of crime more generally. Fortunately, this myopic view of white-collar crime is now clearing, and the study of white-collar crimes and criminals is beginning to be located within the larger criminological literature. White collar crime scholars are applying research strategies, techniques, and theories from the more general crime literature to the study of white-collar crime and offenders and scholars who in the past had little interest in understanding white-collar crime and now beginning to consider, discuss, and research its importance.

Related Entries

- ▶ Corporate Crime Decision-Making
- ▶ Corporate Crimes and the Business Cycle
- ▶ Crimes of the Powerful
- ▶ Women and White-Collar Crime

Recommended Reading and References

Benson ML, Simpson SS (2009) White collar crime: an opportunity perspective. Routledge/Taylor & Francis, New York

Caldwell RG (1958) A re-examination of the concept of white collar crime. Fed Probat 22:30–36

Clinard MB, Quinney R (1973) Criminal behavior systems: a typology. Rinehart and Winston, New York

Cressey DR (1953) Other people's money: a study in the social psychology of embezzlement. Free Press, Glencoe

Edelhertz H (1970) The nature, impact and prosecution of white collar crime. Law Enforcement Assistance Administration, U.S. Department of Justice, Washington, DC

Geis G (1962) Toward a delineation of white collar offenses. Sociol Inq 32:160–171

Geis G (1967) White collar crime: the heavy electrical equipment antitrust cases of 1961. In: Clinard MB, Quinney R (eds) Criminal behavior systems: a typology. Rinehart and Winston, New York, pp 139–151

Geis G (1992) White collar crime: what is it? In: Schlegel K, Weisburd D (eds) White collar crime reconsidered. Northeastern University Press, Boston, pp 31–52

Huff R, Desilets C, Kane J (2010) The 2010 national public survey on white collar crime. National White Collar Crime Center, Morgantown

Kane J, Wall A (2006) The 2005 national public survey on white collar crime. National White Collar Crime Center, Fairmont

Payne BK (2012) White collar crime. Sage, Los Angeles

Piquero NL, Benson ML (2004) White collar crime and criminal careers: specifying a trajectory of punctuated situational offending. J Contemp Crim Justice 20:148–165

Piquero NL, Weisburd D (2009) Developmental trajectories of white collar crime. In: Simpson SS, Weisburd D (eds) The criminology of white collar crime. Springer, New York, pp 153–171

Piquero NL, Carmichael S, Piquero AR (2008) Assessing the perceived seriousness of white-collar and street crimes. Crime Delinq 54:291–312
Rosenmerkel SP (2001) Wrongfulness and harmfulness as components of seriousness of white-collar offenses. J Contemp Crim Justice 17:308–327
Schoepfer A, Carmichael S, Piquero NL (2007) Do perceptions of punishment vary between white-collar and street crimes? J Crim Justice 35:151–163
Sellin T, Wolfgang M (1964) The measurement of delinquency. Smith Patterson, Montclair
Shaw CR, McKay HD (1942) Juvenile delinquency and urban areas. University of Chicago Press, Chicago
Simpson SS, Piquero NL (2002) Low self-control, organizational theory, and corporate crime. Law Soc Rev 36:509–547
Steffensmeier D (1989) On the causes of "white collar" crime: an assessment of Hirschi and Gottfredson's claims. Criminology 27:345–258
Sutherland EH (1940) Is "white collar crime" crime? Am Sociol Rev 10:132–139
Sutherland EH (1941) Crime and business. Ann Am Acad Polit Soc Sci 217:112–118
Sutherland EH (1945) Is 'white collar crime' crime? Am Sociol Rev 10:132–139
Sutherland EH (1983) White collar crime: the uncut version. Yale University Press, New Haven
Tappan PW (1947) Who is the criminal? Am Sociol Rev 12:96–102
Weisburd D, Waring E (2001) White collar crime and criminal careers. Cambridge University Press, New York
Weisburd D, Chayet EF, Waring E (1990) White collar crime and criminal careers: some preliminary findings. Crime Delinq 3:342–355
Weisburd D, Wheeler S, Waring E, Bode N (1991) Crimes of the middle classes. Yale University Press, New Haven
Wheeler S, Weisburd D, Bode N, Waring E (1988) White collar crime and criminals. Am Crim Law Rev 25:331–357
Zietz D (1981) Women who embezzle or defraud: a study of convicted felons. Praeger, New York

White-Collar Crime

► Compliance and Corporate Crime Control
► Corporate Crime Decision-Making
► Crimes of the Powerful
► Women and White-Collar Crime

Wildlife Trade

► Crimes Against Animal Life

Wildlife Trafficking

► Crimes Against Animal Life

Witchcraft

► Lawful Killings

Witness Characteristics

► Estimator Variables and Eyewitness Identification

Women

► Victimization, Gender, and the Criminal Justice System

Women and White-Collar Crime

Mary Dodge
School of Public Affairs, University of Colorado Denver, Denver, CO, USA

Synonyms

Corporate crime; Elite deviance; Occupation crime; Professional crime; White-collar crime

Overview

Women's involvement in white-collar crime has received little attention in the popular press and scholarly literature. A similar pattern of indifference occurred for most aspects of gender and crime until the 1970s, primarily because women were far less likely to engage in illegal behavior.

The mid-1970s represented a period of transformation in criminological studies which coalesced with the momentum of the women's rights movement. The next three decades placed gender and crime on the research forefront as feminist scholars sought to challenge and change the masculine focus of criminal behavior – except in corporate and occupational crime. The higher crime rates among women were perplexing, with little theoretical explanation and scant data to provide a framework for scholarly examinations. The 1980s and 1990s resulted in remarkable efforts by academics to explain gender differences related to the ratio gaps, motivations, and arrest patterns. Men, however, for a multitude of reasons, continue to commit the majority of crime in the United States.

Key Issues and Controversies

Historically, female criminality was attributed to maladjusted upbringing, psychological neuroses, and desperate attempts to provide monetary assistance to family and friends. These explanations of why women engage in crime meshed nicely with arrest rates for female offending: prostitution, check-kiting, child abuse, and shoplifting. During the 1950s, female criminality often was attributed inaccurately to sexual promiscuity. Female delinquents were described as troubled, disturbed girls with poor self-esteem whose frustration resulted in inappropriate sexual behavior. Overall, attempts to understand female criminality were dismissed, ignored, and hampered by the idea that crime was a male activity committed and studied by men (Brown et al. 2007; Daly and Chesney-Lind 1988). A transformation of stereotypes associated with female behavior, legal and illegal, occurred as an increasing number of women left the private sphere and entered the public domain. Ideas that women were more compassionate and ethical compared to men were challenged in the public sphere. The workplace demanded competitive behavior and provided new opportunities to engage in crime. The changing role of women was noted as early as 1922 by Clarence Darrow:

> No doubt as women enter the field of industry formerly occupied by men, and as she takes her part in politics and sits on juries, the percentage of female criminals will rise rapidly. As she takes her place with men she will be more and more judged as men are judged, and will commit the crimes that men commit, and furnish her fair quotas of the penitentiaries and jails. (cited in Steffensmeier and Streifel 1993, p. 79)

In 1975, Freda Adler and Rita James Simon's groundbreaking work brought female criminality into the public domain. Adler's acclaimed, though controversial book, *Sisters in Crime*, raised awareness about the offenses women did and would commit; she also set forth expectations that women would participate at increasing levels in all types of crime given the opportunity. In perhaps the most prophetic line in Adler's book, she wrote:

> In the future a greater proportion of wealth and power will pass through feminine hands, and almost all of it will be wielded responsibly. But it would be unrealistic reversion to quixotic chivalry to believe that, for better or worse, women will be any more honest than men. (p. 169)

Adler noted the importance of the context of social roles and criminal behavior: "The kinds of crimes one commits are related to the illegal opportunities to commit them. While a shopper might pilfer from commercial establishments, it is not possible to be involved in insider trading, unless one is 'inside' the corporate community" (Dodge 2009, p. 180).

Similarly, Simon predicted in her book *Women and Crime* an increase in white-collar offenses among women and believed that opportunity was the essential element in rising female crime rates. She accurately anticipated an increase in incidents of fraud, embezzlement, larceny, and forgery as women began to participate at higher rates in typically male-dominated, nontraditional occupations. Simon explained her position in a 2007 interview: "Women are no more honest, no more decent, and no more moral than men. The only reason they have lower crimes rates, particularly, white-collar crime, was because they had fewer opportunities to commit crime" (Dodge 2009, p. 12).

Critics, however, argued the work by Adler and Simon lacked empirical integrity and debates continued over whether or not the predictions related to increased female involvement in traditionally male offenses actually emerged (see, e.g., Daly and Chesney-Lind 1988; Steffensmeier and Allan 1996). Clearly, current research shows a trend toward more female participation in monetary crimes. Arrest statistics, according to Simon and Ahn-Redding's (2005) analysis, have shown an increase in the number of women in high-status positions with fiduciary responsibilities which appears to correlate with higher arrest rates for white-collar offenses.

More recently, scholars and practitioners have sought to explain the perpetration by and victimization of women in the field of white-collar crimes. Women who commit white-collar crime are thought to be on the rise, though a lack of data restricts any definitive conclusions. Debates related to the extent of involvement for women in white-collar crime arise from several significant issues. First, definitions of white-collar crime often focus on the offender who is typically described as a person who holds a position of authority, status, or trust. Generally, white-collar criminals are business and professional men engaging in illegal or unethical behavior in the course of business. Women, more likely to be relegated to the private sphere or pink-collar jobs, have lacked opportunity to engage in elite crime. Second, disagreement exists over what types of crime constitute a white-collar offense, which typically focuses on corporate and occupational violations and wrongdoing. More recent work suggests that the category of status include middle-class and working class people, and even juveniles (Shapiro 1990; Pontell and Rosoff 2009). Despite disagreements on offender and offenses, arguably both genders are motivated to engage in white-collar crime for similar reasons related to opportunity, greed, and the American Dream.

Women Who Commit White-Collar Crime

The increasing number of white-collar violations by women has received some recognition by scholars and the popular press. In 1984, an article in the *U.S. News & World Report* reported on the concern over rising incidents of white-collar by professional and corporate women. Examples included Nancy Young, a lawyer, who stole $300,000 from clients' accounts; Frances Cox who embezzled $48,000 from the Virginia government; Mary Hudson, board chairman of Hudson Oil, who was accused of involvement in a scheme to fix gas pumps to rip-off customers; and Mary Treadwell who defrauded tenants in a housing project. In 2004, reporter Anthony Paonita presented a laundry list of women involved in high-profile crimes such as price-fixing, racketeering, fraud, and political corruption.

Previous empirical work shows the trend toward greater participation by women. Jay Albanese (1993) reported an increase in the number of women employed in white-collar jobs along with an upswing in the number of arrests for fraud, forgery, counterfeiting, and embezzlement during the 1970s and 1980s. Haantz (2002) discovered from a sample of 1,016 federal prisoners incarcerated for white-collar crimes in 2000 that nearly one in four was a woman. Simon and Ahn-Redding (2005) found, in 2001, women accounted for more than one-third of all arrests for larceny and almost half of all arrests for embezzlement and fraud. Overall, the results indicate women are committing more white-collar crimes because of greater participation in the labor force which provides increased opportunities.

Embezzlement represents a category of crime in which women excel. Despite debates over the classification of embezzlement offenses, many scholars and law enforcement agencies recognize the theft of money by a person in a position of trust as a form of employee crime as white-collar crime. Much of the debate over including embezzlement in the realm of white-collar is offender focused. Often embezzlers are low-level employees such as clerks or bookkeepers, though they hold a position of fiduciary trust and may steal high dollar amounts. Friedrichs (2010) noted embezzlement also is readily apparent in the high monetary losses companies suffer from managers, often female, who award themselves outrageous bonus and perks.

Embezzlement is the one category of the FBI's Uniform Crime Report that shows the arrests for women at a rate equal to or higher than men. Joseph Morales, Chief Deputy District Attorney in Denver, Colorado, estimated approximately 73 % of the embezzlement cases reported to their office were female offenders (personnel communication, October 2011). *The Marquet Report on Embezzlement* described 2010 as a "banner year for employee theft in the United States" (p. 4). Marquet's analysis of 483 major embezzlement cases (defined as losses of about $100,000 or more) revealed 64 % of the crimes were committed by women, though males stole significantly higher dollar amounts. The largest monetary case, however, was carried out by a Melissa G. King, who stole over $42 million from union benefit accounts over a 7 year period. According to the report's findings, the motivations of the majority of embezzlers involved the desire for a more lavish lifestyle rather than financial problems.

Early research on embezzlers by Donald Cressey (1953) included interviews with male offenders at three prisons. The criterion for participation was operationalized as "violators of financial trust" and included a sample of embezzlers, forgers, and other fiduciary offenses. The men self-reported having a "non-sharable" financial problem and viewed their embezzlement as a means to secretly solve their dilemmas. Participants also noted a sense of entitlement and resentment toward their employers. The embezzlers used "vocabularies of adjustment" to justify their theft as borrowing money. In the first study of female embezzlers, Dorothy Zietz (1981) interviewed women at the California Institution for Women. At the time of the interviews of the 900 women inmates, 45 % had been convicted of felonies against property. The 100 participants in her study were identified as either embezzlers (i.e., "honest women who violated financial trust") or fraudulent operators ("women who intended to steal or defraud"). The honest women embezzlers, an oxymoron at best, included four types: Obsessive protectors, romantic dreamers, greedy opportunists, and victims of pressure or persuasion. Unlike Cressey's embezzlers, the women rationalized their actions as benefiting their families.

A recent study of 40 individuals convicted of a variety of fiduciary frauds compared interview data of 20 male and 20 female offenders (Klenowski et al. 2011). The findings revealed that both genders engaged in rationalizations to frame their criminal conduct as legitimate in order to maintain a respectable self-identity. All offenders made attempts to justify their crimes as a method of helping their families and friends. More commonly, women were likely to blame the inability of male members of the household to act as providers, whereas men faulted their own inadequacies. The responses analyzed in terms of techniques of neutralization discovered that male offenders were more likely to deny injury, make claims of normality, and condemn their condemners. In contrast, women defended their actions as necessary to resolve financial issues related to family and health. The small sample size in the Klenowski, Copes, and Miller study limits any definitive conclusions related to gendered motivations, though the findings are similar to early work that found women were more likely to rationalize their financial crimes as attempts to solve family problems.

Current research on women who commit fiduciary crime is limited and their means and motives not fully explored. In some cases, women engage in embezzlement because of family concerns; however, a multitude of examples show the complexities of the psychological and social contextual variables associated with their decisions. In 2004, for example, Ruth Cameron misappropriated $150,000 from the University of Colorado Boulder to supplement a more lavish lifestyle. Mary Ellen Wilson was charged with stealing almost $300,000 over 8 years and admitted to spending the money on gambling and extravagant trips to Europe. In one of the largest financial swindles against the state of Colorado, Michelle Cawthra, a supervisor of the Colorado Department of Revenue's Taxpayer Service Office, stole over $5 million, which she shared with her married boyfriend who accompanied her on trips to Las Vegas.

Case studies also reveal women are active participants in major white-collar crimes including insider trading, price-fixing, political graft, and corporate fraud. The media hoopla surrounding these cases, however, often distorts the crucial elements of the crimes and reduces analyses to gender stereotypes. Women who step outside their roles as "feminine" and dare tread on the toes of the powerful male executives and prosecutors are likely to be treated harshly as white-collar criminals.

Martha Stewart's alleged wrongdoings present all the complexities of a white-collar crime, particularly one committed by a woman who epitomizes and profits from the domesticity and corporate spheres. Insider trading was the most serious allegation, though Stewart was never charged or tried for the crime. The prosecution of Stewart was contentious and many commentators and loyal fans believed she was a victim of an overzealous prosecutor. Newspapers reported she may have been singled out based on gender, but the more likely scenario for pursuing charges of perjury was because she was unlikeable and arrogant.

Stewart was not the first unlikeable powerful woman to face criminal charges related to white-collar offenses. In the 1980s, Leona Helmsley was indicted for conspiracy to defraud the government, tax evasion, mail fraud, and extortion. Helmsley was well known for her fierce intensity and compulsion to demand perfection in employees. Helmsley, a powerful business woman in New York, became more famous when she married billionaire, hotel and real estate magnate Harry Helmsley. The holdings of the Helmsleys were extensive and included ownership of the Empire State Building and the Park Lane and Carlton House hotels. Leona Helmsley managed the Palace Hotel as queen of her own realm and was viewed by many employees as an evil dictator.

In 1986, a reporter for the *New York Post* first reported the allegations and provided evidence of wrongdoing by the Helmsleys. The Helmsleys were accused of falsifying millions of dollars of renovations to their mansion as business expenses. The article prompted the US Attorney to launch an investigation. Leona also was said to have charged clothing and gifts to business accounts. One such gift for her husband was a $45,000 silver clock designed in the shape of the Helmsley Building. Prosecutors argued the billing schemes were a blatant attempt to avoid paying taxes.

Leona Helmsley's reputation became a focal point for the media and ensuing trial. Her mean spirited reputation and mistreatment of employees became legendary. Scores of scorned employees related horror stories of her verbal abuse and threats. One employee commented: "She's just like the Red Queen in Alice in Wonderland. She makes the rules, and if you don't live by her rules, then off with your head" (cited in Pierson 1989, p. 355). Former employees claimed that she extorted money and services from contractors and suppliers. One accusation involved RCA and her demand for free televisions for the Helmsleys' mansion in order for the company to secure the hotels' accounts. Leona's reputation as a cold and nasty person was perpetuated by the mayor of New York, Edward Koch, who called her the "Wicked Witch of the West" and *Newsweek* magazine which headlined an article "Rhymes with Rich." Even her defense lawyer joined the fray and referred to his client as one "tough bitch," despite his client's aversion to the characterization (Blum 1989).

The unflattering characterization of the 69-year-old Helmsley likely represented a major obstacle to winning the jury trial. Her husband, who was 80 years old at the time, was found mentally incompetent and avoided the ordeal of a trial. Leona, however, was faced with countering prosecutor's assertions that her status as one of the socially elite made the allegation of tax evasion far more despicable compared to the average person. During the trial, she was described as a cheat and a cheapskate. Testimony by an Internal Revenue Service agent detailed the personal items charged as business expenses including her hair rollers, a $13 girdle from Bloomingdales, a $58 Itty Bitty Book Light, and a $21 crossword puzzle club subscription (Pierson 1989).

The 10-week trial included over 40 prosecution witnesses and 10,000 pages of documentations. The jury of six men and six women deliberated for 6 days and found Helmsley guilty of 33 felonies. Many spectators and commentators believed Helmsley was unlikely to garner much sympathy from the jury; despite her age, gender, and caregiver role to her husband. She denied any wrongdoing and placed the blame for her situation on people who were jealous of her position and her husband (Pierson 1989). She also shifted blame to former employees, who she asserted were stealing from the companies. Helmsley was sentence to 4 years in prison, fined $7.1 million, and ordered to pay $1.7 million in owed taxes. She was released after serving 18 months at the Federal Correctional Institution in Danbury, Connecticut. When Harry Helmsley died in 1997, he left an estimated $1.7 billion estate to his wife. When Leona Helmsley died in 2007 she left most of her estate to a charitable trust and a $12 million trust fund to her dog, Trouble.

Cases of women who commit corporate, professional, and political crime are plentiful and as a whole suggest increased female participation. Diana Brooks, labeled as one of the most powerful women in the world of art was charged and pled guilty to a price-fixing scheme between two of the most powerful auction houses in the world, Sotheby and Christie's. In 2003, Lea Fastow, the only woman charged in the Enron case, was indicted for conspiracy to commit wire fraud, money laundering, aiding and abetting, and tax fraud. Fastow accepted a plea bargain and received a 1 year prison sentence for a misdemeanor conviction. Betty Vinson, the corporate reporting director for WorldCom, was sentenced to 5 months in prison for attempting to hide $7 billion in expenses and inflate earnings. The WorldCom scandal represented one of the biggest corporate meltdowns and involved fraud amounting to over $11 billion. Robin Szeliga, former chief financial officer at the telecommunications company Qwest, accepted a plea bargain on one count of insider trading for her role in the multibillion-dollar accounting scheme.

The list of women involved in political crime also is growing and opportunities are expanding. Political corruption often involves bribery by individuals, organizations, and corporations. The estimates of money paid to public and private officials in the United States range from $3 billion to $1 trillion in bribes annually (Baird 2006). Bribery is the most common form of political corruption, though other major offenses include violating public trust and influencing governmental contracts awards (Ross 2003). Overall, given the lack of participation by women in politics, the number of cases of corrupt female officials is surprising.

The Bess Myerson case represents the quintessential example of illegal and unethical behavior developing from a political appointment that mixed street and suite crimes. In 1945, Myerson was the first Jewish woman to be crowned Miss America. She hosted the television game show *The Big Payoff* and was a regular panelist on *I've Got a Secret*. *The Big Payoff* was cancelled during the quiz-show scandals in the mid-1950s when evidence emerged that the games were rigged. In the late 1960s, Myerson entered the world of politics as a champion for consumer protection legislation when she was appointed by Mayor John Lindsay as New York City's first Commissioner of Consumer Affairs. She became a beloved political figure and set out on a campaign to protect consumers. In her new position, Myerson frequently appeared on television to attack "shamburgers," unlicensed auto mechanics, and unsafe toys (Alexander 1990). In her first 4 months as commissioner she issued 1,300 summonses to companies for consumer violations.

Myerson was a frequent social companion of Ed Koch during his unsuccessful 1982 Governor's race. Many people in the Democratic Party believed that with Myerson attending events with Koch rumors that he was gay would be dispelled after a nasty campaign slogan; "Vote for Cuomo, not the homo" began circulating. While Koch served as Mayor he appointed Myerson to the position of Commissioner of Cultural Affairs of New York City. Myerson, though popular, often was criticized for engaging in underhanded tactics and questionable behavior.

While acting as Commissioner of Consumer Affairs, Myerson's role as a consultant to

Bristol-Myers became a source of controversy. Myerson developed a consumer products guide for Bristol-Myers that failed to mention the possible carcinogens in Clairol hair dyes that were being targeted by the Food and Drug Administration. Her history of shoplifting also raised questions about her credibility.

Myerson took advantage of her fame and relished in the special treatment it afforded her. She ate at restaurants that provided free meals, accepted merchandise from vendors at wholesale prices, and neglected to pay for items (Alexander 1990). Her reputation as a cheapskate was widely known, though in the 1980s her net worth was $4 million. Myerson's personal and professional lives became fodder for the New York newspapers and her behavior was seen as increasingly scandalous. Her partnership with reputed mobster Andy Capasso became legendary when she used her political position to influence the outcome of his divorce case. Then revelations of widespread corruption in the Koch administration became public and five top city officials were indicted and convicted for bribery, racketeering, and fraud (Alexander 1990). Later, Prosecutor Rudolph Giuliani indicted Myerson and several codefendants on conspiracy and mail fraud charges.

When Myerson, who denied any wrongdoing and labeled the ordeal as a "witch hunt," was served with a subpoena at her office she screamed: "I don't give a fuck about any of this shit! I'm still commissioner around here? To hell with...Giuliani! I don't want anybody cooperating with those guys!" (Alexander 1990, p. 17). The *Daily News* reported: "At long last, the Bess Myerson rock has been kicked over, and what's crawling out is sickening: arrogance, abuse of power, payroll manipulation, case-fixing, cover-up" (Alexander 1990, p. 207). Koch sided with public sentiment about her guilt and claimed to be shocked, calling her actions deplorable, dishonorable, and disgraceful, though he had received a report that confirmed the allegations against Myerson several months prior to the indictment.

Myerson was acquitted of all the charges, but the scandal and trial destroyed her already questionable reputation. People blamed her plight on her arrogance and inability to play by the rules. Alexander noted the fall of Myerson was directly related to her gendered perspectives: "Her attitude has very much to do with her view of men and women—that at bottom we women are alone in the world, and we have to do the most we can to defeat or outsmart or outrun the men. Basically, she hated men" (p. 55).

Examples of political crime by women are surprisingly numerous. Sara Bost, who served as mayor of Irvington, New Jersey, was indicted for taking bribes from developers and engaging in witness tampering. After a year-long investigation by the federal government and 1 month into a jury trial, Bost pled guilty to one count of witness tampering and the charges that she had received $8,500 in kickbacks. She was sentenced to 1 year in prison, a $2,000 fine, and 150 hours of community service work. Commentators sarcastically noted that Bost had broken the glass ceiling to join the ranks of the boys club—in prison (Dodge 2009). Bost, like Stewart, served her time at Alderson prison. Nidia Davila-Colon, a long-term freeholder in New Jersey, resigned her position after her conviction for bribery. Davila-Colon was charged with mail fraud and aiding and abetting an attempted extortion scheme. She was sentence to 3 years and 1 month in a federal prison. In a similar case, Chicago Alderman Arenda Troutman was arrested in 2007 on federal bribery charges. Troutman, according to an FBI report, accepted a $5,000 cash bribe with an additional payment of $15,000 as political contributions after supporting zoning changes and alley access for a private developer.

In 1994, Linda Schrenko became one of the first women in Georgia to be elected to a statewide office as school superintendent. She was responsible for a $7 billion budget. When she ran for governor on the Republican ticket in 2002, she genderized the campaign race. In her first commercial she informed the public: "I'm asking you to send a woman to clean the House and I'll throw in the Senate for free." Throughout her campaign, she noted she was "obviously not a good ol' boy" and wore a name tag "I Am 'That Woman.'"

Schrenko's gubernatorial campaign ran short of money. In August 2001, the account balance was $113 and a number of bounced checks surfaced. Sometime later, significant amounts of cash deposits were made, including federal education money that was earmarked for Georgia Schools for the Deaf and the Governor's Honors Program. Meanwhile, Schrenko was writing checks for cash to her family and transferring funds to offshore accounts. Stolen money was also used for a face lift and to purchase a new car.

Schrenko faced over 40 criminal counts and eventually pled guilty to felony charges of conspiracy and money laundering. She was sentenced to 8 years in prison, ordered to pay restitution of $415,000, 3 years probation, and 100 days of community service. The presiding judge called her crimes "an egregious abuse of the public's trust." Schrenko denied any responsibility and placed the blame on her campaign manager and lover, Merle Temple. When she was asked during a television interview how she could be unaware of how the money was being handled, she stated: "Utter stupidity." She also commented in the same interview: "I don't want to handle the finances; I'm not good at that. I just want to improve the educational system and help kids: it wasn't about the money and the power."

Overall, the number of women involved in political wrongdoing, given their low representation, is surprising. Betty Loren-Maltese, former Illinois town president, was accused of racketeering and fraud in connection to an insurance scheme that resulted in $12 million in losses to the town. Loren-Maltese was known to be a flashy gambler and extravagant spender. She allegedly lost $1.2 million in Las Vegas gambling, while authorities attempted to seize her assets (Paonita 2004). Janet Rehnquist, the daughter of former chief justice William Rehnquist, was appointed inspector general of the US Department of Health and Human Services by President Bush. In 2003, a GAO report uncovered unethical and illegal activities including improperly obtaining law enforcement credentials that allowed her to carry a gun in the office and she allegedly traveled for pleasure at the expense of the taxpayers (Paonita 2004). Colleagues claimed she created an atmosphere of anxiety and distrust. Rehnquist resigned from her position after the report uncovered numerous unethical and illegal behaviors.

Political crimes, perhaps more than other white-collar offenses, tend to undermine essential elements of trust between the public and elected officials. The loss of faith in public officials increases citizen apathy and public cynicism. Moore and Mills (1990) noted:

> Whenever citizens see corrupt public officials and other white-collar offenders violate the law with impunity they inevitably must question official integrity and commitment to fairness. In many ways, the greatest harm to the victims of white-collar crime may be this loss of faith in the very possibility of fair, impartial government. (pp. 414–415)

Women are obtaining more and more positions in the traditionally male-dominated political arena. It appears women are just as likely as men to take advantage of an imperfect system.

Future Directions

The United States has seen a substantial increase in the number of women who have assumed powerful leadership roles in political, corporate, judicial, and academic spheres. The primary hypothesis to explore in empirical work is will women engage in more ethical behavior and commit fewer illegal activities compared to male counterparts? Fukuyama (1998) asserted that men tended to be more aggressive, engage in more violence, and seek dominance through competition at a higher rate than women. Fukuyama noted "a world run by women would follow different rules" and as women gain power, countries and leaders will become less "aggressive, adventurous, competitive, and violent." Some previous research supports the idea that women may act differently and value relationships, personal and professional, to a higher degree than men. Perhaps women will be successful at developing a sense of community and connectedness to the corporate and political worlds, though the argument is growing increasingly weak as the number of women white-collar offenders increases (Dodge 2009).

Stereotypes that women are more trustworthy and ethical compared to men are common, though evidence these female characteristics actually exist is equivocal. Researchers find that women and men share similar work values and are similar in their perceptions of ethical and unethical behavior (Kidwell et al. 1987). A reasonable extrapolation from the acts of embezzlement by women would suggest in higher positions females would be just as likely to engage in white-collar crime under the right circumstances. In fact, discriminatory practices in job placement and salaries may contribute to more women committing white-collar offenses as opportunities widen. Women in positions of authority in the business world are rare and they are unlikely to be placed as supervisors with demanding responsibilities (Alkadry and Tower 2011). In 2011, only 12 women served as CEOs of Fortune 500 companies. Women, nevertheless, are moving into high-level business and political positions.

The controversies over workplace behavior are far from resolved. Arguments related to females as the more ethical gender in business further perpetuates gender-based stereotypes. The idea that women will enter into traditionally masculine positions, which demand assertiveness, risk-taking, and competitive behavior, and not commit white-collar crime is questionable. Women face the same strain and demands as men. West and Zimmerman (1987) introduced the theory of "doing gender," which is the "activity of managing situated conduct in light of normative conceptions of attitudes and activities appropriate for one' sex category" (p. 127). Both men and women actively monitor their behavior to ensure they will be perceived not simply as male or female but also as masculine, feminine, or neither. Corporate officers, both male and female, must do gender in ways that support their career activities (West and Zimmerman 1987). The combination of formal organizational forces and informal occupational values creates a social world which emphasizes masculinity. Women in the workplace find ways of doing gender; whether or not this is accomplished by adapting masculine methods is unknown. The combination of formal organizational forces and informal occupational values centered on machismo and competence creates a social world which relishes masculinity (Herbert 1998).

In 2006, Carly Fiorina, the former CEO of Hewlett Packer, was named by Fortune as the "Most Powerful Woman in Business." Fiorina was quick to express her displeasure of being placed in a special category that differentiated women from men, though she acknowledged the higher level of scrutiny and criticism females face in business. Fiorina (2006) noted her thoughts about being the most power woman in the corporate world: "I've never been a man" and I never thought of myself as a woman in business. I've thought of myself as a person doing business who happens to be a woman (p. 145). In her book, *Tough Choices*, she described her impression about her imminent firing:

> I knew the announcement would be big news. I was a woman, and a bold one at that, and things had always been different for me. All the criticisms that have ever been leveled against me would be recycled and thrown back in my face with new delight: "She's too flashy." "She's just marketing fluff." "She's too controlling." "She's a publicity hound." "She's imperious, vindictive and employees didn't like her." "It would be ugly and it would be personal." (p. xii)

Fiorina believes she was never able to escape the label of "Female CEO," and the media was preoccupied with describing her dress, hair, or clothing rather than her business acumen. She was routinely described as a bimbo, a bitch, too soft, too hard, and too ambitious (Fiorina 2006).

Women in high-status position are placed in a no-win situation. Rhode (1989) labeled the circumstances as a double bind: "Women are criticized for being too feminine or not feminine enough. Those who conform to accepted stereotypes appear to lack ideas or initiative, while women who take more assertive stances are judged arrogant, aggressive, or abrasive" (pp. 169–170). In the cases of Stewart, Brooks, and Helmsley, they were praised for their business acumen until accusations of white-collar crimes became public. Perceptions of these women, despite some known character flaws, turned to negative feminine stereotypes that described their personalities and behavior – which rarely happens with male offenders (Dodge 2009).

Women face many professional barriers that prevent them from seeking or obtaining upper management positions. Work and family responsibilities often clash and demanding corporate jobs with long hours are problematic. Men hold leadership positions in most professions and cultural views about suitability for leadership roles may exclude women. Societal and organizational subcultures present a significant barrier for women and perpetuate marginalization (Schein 1992). Additionally, the isolation women experience in male-dominated professions can limit access to upper-level positions. Females, on the one hand, often are excluded from informal mentoring networks within the organization (Gold 1999). On the other hand, women who are able to penetrate and become part of the boys club may learn that being a member of the inner circle demands cooperation and, sometimes, illegal behavior.

Many women who have committed white-collar crime are forgotten after the initial media hoopla. Martha Stewart's comeback after her felony convictions and short prison term demonstrated the business acumen that gave rise to her empire. While Stewart may have experienced a type of reintegrative shaming that overcame the stigma of being a criminal, as described by Australian criminologist John Braithwaite, Helmsley never recovered from the negative perceptions and continued to bully people after her release from prison.

White-collar crimes committed by women continue to surpass those of the past in terms of financial theft and offense seriousness. White-collar crimes, unethical or illegal, are committed to increase personal and corporate wealth. Such acts are no longer limited to men, and women are taking their place at a steady pace as opportunity and motive become more readily available.

Related Entries

- ▶ Crimes of Globalization
- ▶ Gendered Theory and Gendered Practice
- ▶ Victims of Corporate Crime

Recommended Reading and References

Adler F (1975) Sisters in crime: the rise of the new female criminal. McGraw-Hill, New York

Albanese J (1993) Women and the newest profession: females as white-collar criminals. In: Culliver CC (ed) Female criminality: the state of the art. Garland, New York, pp 119–131

Alexander S (1990) When she was bad: the story of Bess, Hortense, Sukhreet & Nancy. Random House, New York

Alkadry MG, Tower LE (2011) Covert pay discrimination: how authority predicts pay difference between women and men. Pub Adm Rev 71:740–750

Baird V (2006) Can the rot be stopped? New Internationalist 396:2–5

Blum A (1989) A queen's defense: countering bad press. Natl Law J 8

Brown SE, Esbensen F, Geis G (2007) Criminology: explaining crime and its context, 6th edn. Anderson, Cincinnati

Cressey DR (1953) Other people's money: a study in the social psychology of embezzlement. Free Press, Glencoe

Daly K, Chesney-Lind M (1988) Feminism and criminology. Justice Quart 5:497–535

Dodge M (2009) Women and white-collar crime. Prentice Hall, Upper Saddle River

Fiorina C (2006) Tough choices. Penguin, New York

Friedrichs DO (2010) Trusted criminals: white collar crime in contemporary society, 4th edn. Wadsworth, Belmont

Fukuyama F (1998) Women and the evolution of world politics. Foreign Affairs http://www.foxnews.com

Gold ME (1999) Top cops: profiles of women in command. Brittany, Chicago

Haantz S (2002) Women and white-collar crime. National White Collar Crime Center http://www.nw3c.org/downloads/women_wcc1.pdf

Herbert S (1998) Police subculture reconsidered. Criminology 36:343–369

Kidwell JM, Stevens RE, Bethke AL (1987) Differences in ethical perceptions between male and female managers: myth or reality? J Bus Ethics 6:489–493

Klenowski PM, Copes H, Mullins CW (2011) Gender, identity, and accounts: how white collar offenders do gender when making sense of their crimes. Justice Quart 28:46–69

Moore E, Mills M (1990) The neglected victims and unexamined costs of white-collar crime. Crime Delinq 36:408–418

Paonita A (2004) Crime wave. More. pp 110–113, 114
Pierson R (1989) The queen of mean: the unauthorized biography of Leona Helmsley. Bantam, New York
Pontell HN, Rosoff SM (2009) White-collar delinquency. Crime, Law Soc Change 51:147–162
Rhode DL (1989) Justice and gender: sex discrimination and the law. Harvard University Press, Cambridge, MA
Ross JI (2003) The dynamics of political crime. Sage, Thousand Oaks
Schein E (1992) Organizational culture and leadership, 2nd edn. Jossey-Bass, San Francisco
Shapiro SP (1990) Collaring the crime, not the criminal: reconsidering the concept of white-collar crime. Am Sociol Rev 55:346–365
Simon RJ (1975) Women and crime. Lexington, Lexington
Simon RJ, Ahn-Redding H (2005) The crimes women commit, the punishments they receive, 3rd edn. Lexington, Lanham
Steffensmeier D, Allan E (1996) Gender and crime: toward a gendered theory of female offending. Annu Rev Sociol 22:459–487
Steffensmeier D, Streifel C (1993) Trends in female crime, 1960–1990. In: Culliver CC (ed) Female criminality: the state of the art. Garland, New York, pp 63–101
West C, Zimmerman DH (1987) Doing gender. Gend Soc 1:125–151
Zietz D (1981) Women who embezzle or defraud: a study of convicted felons. Praeger, New York

Women in Policing

Jennifer Brown[1], Tim Prenzler[2] and
Anne R. van Ewijk[3,4]
[1]Mannheim Centre, Department of Social Policy, London School of Economics, London, UK
[2]Australian Research Council Centre of Excellence in Policing and Security, School of Criminology and Criminal Justice, Griffith University, Brisbane, Australia
[3]Department of Political and Social Sciences, GRITIM (Interdisciplinary Research Group on Immigration), Universitat Pompeu Fabra, Barcelona, Spain
[4]Samhoud, Utrecht, The Netherlands

Overview

Women rarely entered policing at the inception of a nation’s police force(s), rather their entry was delayed and then often marked by circumscribed roles and limited occupancy of all available ranks. Four broad phases can be identified when women officers were recruited into the police: after World War I, the interwar years, after World War II, and the modern period from the 1980s onwards. With some variations, one familiar pattern is observable whereby there is a period of omission, followed by limited succession, formation of separate women’s departments, and in some jurisdictions a further stage of working towards full integration. At time of writing, there is no evidence of a fully integrated police organization where women represent 50 % of the officer workforce and enjoy an equivalent share of the full range of roles and ranks within the police hierarchy.

The experiences of women in policing were not subjected to systematic study until the 1970s when Susan Martin undertook a landmark participant observational analysis of American women police officers (Martin 1980). Since then there has been a great deal of scholarship addressed at women in policing, e.g., evaluating their efficacy; determining roles and rank progression women officers have achieved; charting their incursion into specialist areas such as firearms; examining stress, styles of policing, and how they handle violence; and looking at the ways in which they lead.

Some advances have also been made in terms of theorizing. Much conceptual work examines women in policing through the lens of gender relations and masculinities and women’s subordinated status (Martin and Jurik 1996). A model of women’s progression has been advanced (see Brown et al. 1998) which has been applied and extended by Strobl (2010) and Natarajan (2008). However, in the area of internal diversity, there are relatively few comprehensive evaluations taking into account developments concerning recruitment, promotion, and retention simultaneously (Van Ewijk 2012), and comparative studies are still scarce (Brown and Heidensohn 2000). While empirical data on gender diversity within police forces is limited, obtaining these data represents more of a challenge in some regions than in others (Van Ewijk 2012).

The task of comparing the progress made by women worldwide is too large to be encompassed within the constraints of this entry, and research papers are too patchy to enable exhaustive coverage; thus, the analysis presented here is necessarily selective. It is noteworthy that women's progress into police services has frequently accompanied periods of reconstruction, as in Europe after the two World Wars, and nation building as in the postcolonial era in the Middle East, Asia and Africa and emerging democracies in South America and Eastern Europe. The numbers of women police officers and their status are often an indicator of the progressiveness of police force's policies, especially with regard to internal diversity and the policing of sexual violence, and provide an indicator of that nation's progress in terms of the equality of women more generally.

Women's Representation in Policing

Brief Global History

It is generally acknowledged that the first woman to be appointed as a police officer was in Stuttgart in 1903 (Heidensohn 2000). Heidensohn notes some similarities and differences between countries in this pioneering phase. Shared characteristics include focus on social and welfare work, use of well-educated recruits from upper- or middleclass backgrounds, and an impetus provided by an external movement lobbying for the presence of women in policing. Prenzler (1994, p. 88) observes in respect of early Australian policewomen that they "were used to fill a gap between law enforcement and welfare and to mitigate difficulties experienced by policemen in managing female offenders and female victims of crime." Heidensohn (2000) suggests differences that reflect Prenzler's dichotomy, in the sense that in New Zealand, USA, and Sweden, women did not adopt a police uniform and were employed as matrons, nursing sisters, or social workers. This contrasts with countries elsewhere, such as Great Britain, where women had a clear aspiration to be part of the police uniformed service and have an involvement in crime and public order (Brown and Heidensohn 2000). The ostensible impetus to employ these first women in policing was to manage female suspects or victims of crime, but the more pragmatic reason was the shortage of men as a consequence of conscription during World War I and the control of sexual morality (see Brown et al. 1998 for a more detailed discussion). Brown and Heidensohn (2000) find evidence for the opposition to the entry of women in policing in all these countries which they state was a combination of paternalistic concerns to protect women and patriarchal exclusion of women as being unsuitable for the rough and dirty tasks required.

During the interwar years, countries such as France, Poland, Spain, and Portugal introduced women into policing. This was in line with either moral policing whereby women officers dealt with trafficking, prostitution, and those classified as moral delinquents or performing support functions such as of typing or administration (Brown and Heidensohn 2000).

After World War II, more countries employed women as officers again in the wake of reconstruction. For example, in Hong Kong, Chinese women were introduced in 1949 although confined to clerical duties, interpreting and taking statements in the Criminal Investigation Department. Interestingly, serving women officers from established forces such as Great Britain were seconded to develop these units of policewomen. American police advisors were sent to South Korea after the ending of World War II to develop a 500-strong female police workforce to work primarily with women prisoners and undertake some crime-related work (Moon and Morash 2008). This pattern of sending advisors to help form policewomen's units could also be observed in postcolonial Africa, where models of policing very much followed patterns established in the former colonial power (Igbinovia 1981). Using the police forces of Ghana and the Ivory Coast as examples, Igbinovia compares Anglophone and Francophone police forces, whose organization and administration were modelled on the patterns of pre-independence British and French colonial rule.

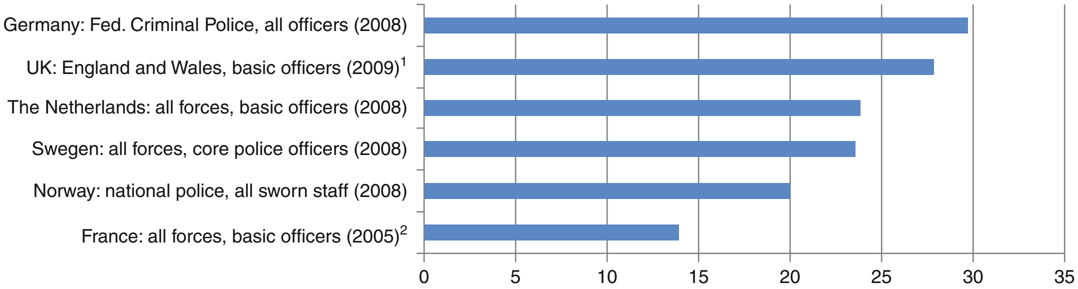

Women in Policing, Fig. 1 Percentage of female officers in various police forces of established democracies in Europe (percentage of basic police officers where possible, otherwise percentage of total police officers) (Sources: European Network of Policewomen (2008), Home Office (2009), Pruvost (2009))

Women were deployed in more challenging tasks in the newly established Commonwealth countries, whereas in the African French community, they were limited to administrative tasks which reflected the roles policewomen played in the countries of the originating colonial power.

The modern era is marked by the reconstruction of democracies in the countries of Eastern Europe following the breakup of the USSR, the introduction of women police in the gendarmeries of Europe, developments in women-only police stations in Brazil and India, and new roles for women police in Arab countries.

In sum, while countries vary in the dates that women were admitted into the police, opposition was almost universal. Political activation and lobbying contributed to changing attitudes towards the possibility of women police officers, but their actual entry was often precipitated by some crisis. For example, in India the idea of women police officers was thought incompatible with their role as homemakers, but their entry was occasioned by problems created through the mass movement of people following Partition in 1947. African women's attempts to form a women's police corps were often met with vilification and accusations of being unable to deal with criminals, yet the reality of trafficking problems in newly emerging nations provided an impetus to their recruitment. The progress of women into mainstream policing in established democracies of Western Europe, developed countries with long traditions of independence, and in America and Canada was to some extent assisted by equality legislation which provided an impetus to equal rights and treatment of women generally with policewomen being among the beneficiaries.

The following sections will take a brief stock-take of the current position of women police officers. Because there are no clear-cut jurisdiction classifications nor strict geographic divide, we have themed policing jurisdictions in terms of their democratic status.

Women Police in Established Western Democracies

The numbers of women police in established Western democracies are steadily increasing and have reached between 20 % and 25 % in some countries and almost a third of the officer complement of the Federal Criminal Police in Germany (Van Ewijk 2012) (Fig. 1).

Percentages of women among rank and file police officers have now risen to or beyond the point where women are generally seen as tokens, which is 15 % with 35 % said to be the tipping point when gender relations are normalized. In terms of their percentage share of management ranks, then this is between 22 % and 27 % for countries such as the Netherlands and Sweden, while in Great Britain those percentages are 16 %. In France however in 2005, 18 % of the commissioners were female, versus 14 % of the patrol officers. Possible explanations are as follows: firstly, a government feminist policy that originated in the 1970s focused on the higher ranks, for which higher education was more important than physical strength and, secondly, that senior French police officers were persuaded that women are especially fit to be police leaders

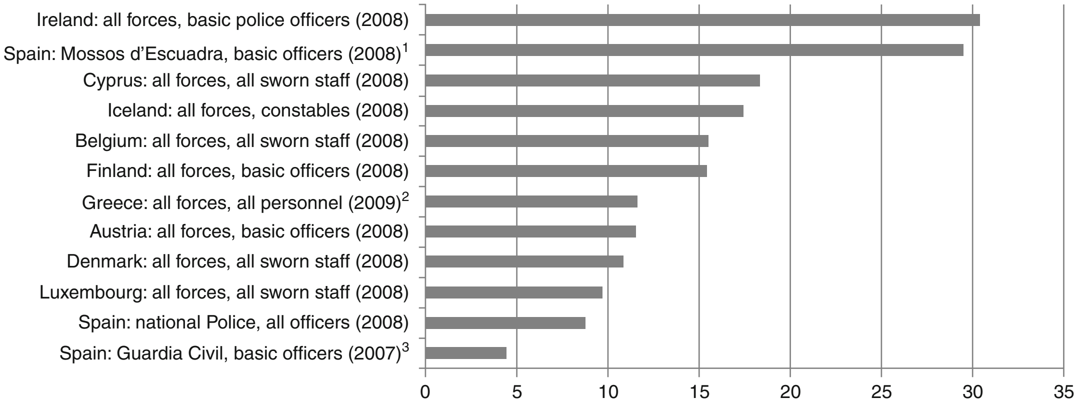

Women in Policing, Fig. 2 Percentage of female police officers in post-World War II reconstruction nations in Europe (percentage of basic police officers where possible, otherwise percentage of total police officers) (Sources: European Network of Policewomen (2008), Programa d'Equitat (2009), Ministry of Citizen Protection (website), and Guardia Civil (2007))

because of their feminine skills, but remained unconvinced that they are fit to be patrol officers because of the risks. A further contributory factor may be direct entry at officer level. As a result, female police officers are more likely to access higher ranks from the outside than by moving up internally from the lower ranks.

It is difficult to predict whether the increase of female police officers will continue in the future. On the one hand, the relatively high percentage of female police students in 2008 in Austria (27.1 % versus 11.5 % of female basic police officers) can be seen as promising for gender diversity. On the other hand, recent data, for example, on high-ranking female police officers in the Netherlands, appear to indicate that in some cases this increase has come to a halt (Van Ewijk 2012). This coincides with Prenzler's observation that in some jurisdictions, periods in which women in policing made gains have been followed by reversals. He illustrated this with reference to the Queensland Police where, following Commissioner Whitrod's reforms, more women were recruited. A subsequent commissioner, Lewis introduced quotas and reinstated the marriage bar, curtailed deployment of women, and systematically removed lesbian officers which resulted in a downturn in the percentage share of women in that force. In the United States there is evidence that the numbers of women in larger police agencies has either stalled or is reversing due to the decrease in the number of consent decrees mandating the hiring of women and sustaining their promotion (National Center for Women and Policing 2002).

Women Police in Postconflict Nations

The position of women in nations reconstructed after war remains somewhat equivocal. Moon and Morash (2008), e.g., looked at South Korea where a target of 14 % of women officers has been set for 2014. They conclude that women are yet to be fully integrated and have yet to have similar training nor undertake comparable work. They find that male officers are still negative about their women colleagues and about increasing their numbers or extending their duties.

We see similar dynamics of gender diversity within some European police forces post-World War II, although the percentage of female officers in the basic levels of police forces fluctuate considerably in this category: in Ireland, more than 30 % of the basic police officers are female, while this is only 4 % in the Guardia Civil of Spain (see Fig. 2).

Again, after post-World War II reconstruction in Europe, the percentage of female officers in managerial and senior ranks is considerably lower than that of rank and file officers (Van Ewijk 2012), with the so far unexplained exception of Ireland.

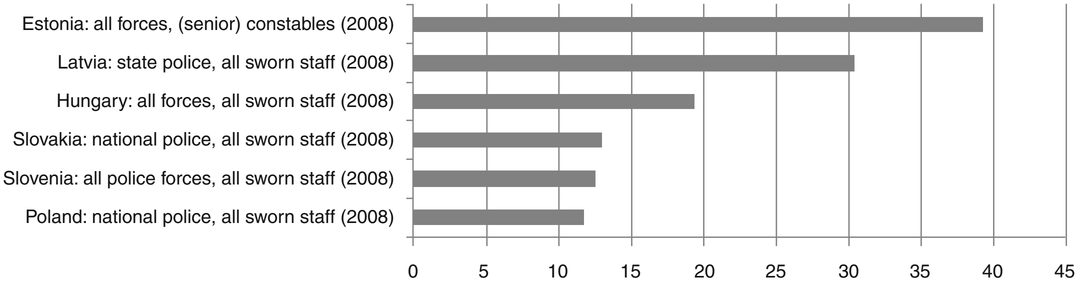

Women in Policing, Fig. 3 Percentage of female police officers in a sample of reforming nations in Europe (percentage of basic police officers where possible, otherwise percentage of total police officers) (Source: European Network of Policewomen (2008))

Women Police in Reforming Nations

Nations that have undergone reforms after seismic political upheavals often have to attend to relationships between the police, citizen, and state. Reams and Nelson (2008) looked at policing in Brazil which was characterized by brutality, corruption, and ineffectiveness. In the wake of democratic reconstruction from a dictatorship to a multiparty democratic system, efforts were made to reform the police to legitimate the new regime. The New Brazilian Republic created women-only police stations in 1985 as a way to combat sexual violence against women. Ten years on, there were 200 delegacies de defesa de mulher (DDMs) throughout the country. Yet many activists have withdrawn their support for this innovation because of poor service and lack of sustained funding and resources. Also, women in mainstream policing believe a secondment to DDMs compromises their long-term career goals, a finding confirmed by Natarajan (2002) when studying women-only stations in Tamil Nadu, India: these were under-resourced and women officers felt unappreciated and lacking rewards or incentives.

The percentage of women police in reforming nations in Europe after the disintegration of the former USSR fluctuates considerably. The high percentage of basic police officers in Estonia contrasts with that of Poland, Slovenia, and Slovakia (Van Ewijk 2012) (Fig. 3).

Despite the increase of female officers in the police force of Estonia, the relation between gender diversity and rank is the same as in the other countries: gender diversity diminishes as rank rises. Basic female officers represent 39.2 %, management female officers 28.9 %, and senior management female officers 4.5 %.

Women Police in Postcolonial Nations

Strobl (2010) reminds us of the importance of cultural differences when attempting to assess progress in justice organizations in developing nations. Asiwaju and Marenin (2008) undertook an extensive study of how community policing was conducted in Nigeria. Policing in that country was hierarchically rigid and defensive against criticism which caused an absence of meaningful communication between rank and file and very senior officers. They described that gender mainstreaming and pressure from civil society had increased the numbers of women from 4 % in 1993 to 16 % by 2006. Nevertheless, barriers remain, such as the current police regulations which do not distinguish recruitment criteria between men and women, the continued assignment of “women’s” duties, i.e., dealing with women and children, the fact that women (but not men) require permission to marry, and the fact that women are discharged when pregnant.

Banks’ (2001) study of women in the Papua New Guinea police notes that women were first recruited in the 1970s, 80 years after the formation of the force. In 1991, it was hoped that there would be full integration except for mobile and riot squads. The system of “wantok,” which is a web of family obligations, is found to be reproduced in the

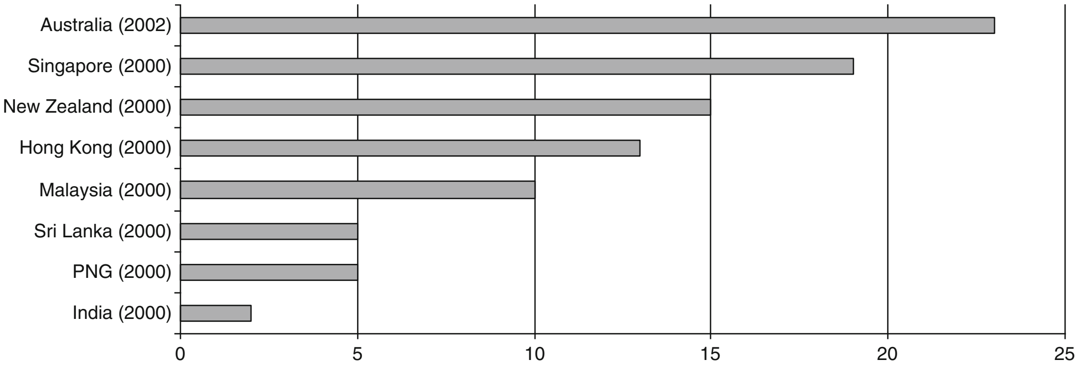

Women in Policing, Fig. 4 Percentage of female police officers in a sample of postcolonial nations (percentages are of all police officers) (Sources: Natarajan (2005), Irving (2002))

workplace where women are treated as mothers or sisters, so that senior women find it difficult to exercise authority over men. Women did not perceive themselves as penetrating gender demarcations, rather they thought of themselves as proffering advice and being of assistance to men and adopted a submissive coping strategy to coincide with the traditional culture.

Strobl (2010) examined the emerging presence of women in policing in the Gulf Cooperation Council countries (GCC): Bahrain (10 %), Kuwait and UAE (unknown), Qatar (4.7 %), Oman (4.5 %), and Saudi Arabia (0 %). Women were introduced into the Bahrainian and Omani police forces in the 1970s, in Qatar in the 1980s, and in Kuwait in 2009. Strobl concludes that the duties of policewomen in GCC countries are more limited that those of their Western counterparts: in GCC countries policewomen are employed in segregated units that focus on women and juveniles as victims, witnesses, and offenders, and they work as security screeners at airports and investigate cases involving child and female offenders. According to Strobl, the Arab, particularly Muslim cultural context, is very different to the secular West. Thus, whereas in Western democracies equality legislation entitled policewomen to litigate when police forces failed to comply with the law in order to advance their progression, in Arab countries "gender as it relates to policing is negotiated in ways that are inconsistent with a secular, liberal feminist agenda" (Strobl 2010, p. 7). As such, she proposes a parallel model where segregation and integration operate simultaneously. In this, she coincides with Natarajan (2008), although in a follow-up to her original data collection Natarajan interestingly finds a shift in preferences with a greater number of mostly younger women officers opting for a fully integrated model, convinced that they can police alongside the men (Fig. 4).

Performance

Operational Frontline Tasks

As we have seen, most police services, during much of their histories, labored under untested assumptions about the unsuitability of women for police work. An exception to this was a series of innovative evaluation projects carried out in the United States in the 1970s and 1980s. These "performance" or "patrol" studies compared the profiles of male and female officers across a wide range of police tasks. The studies varied in their findings on arrest rates, traffic citations, sick days, accidents, academy results, and commendations and reprimands. However, overall, most differences were largely negligible (for summary of this research, see Lunneborg 1989). Other studies found that, while the public was generally accepting of women police, female victims of crime tended to prefer the assistance of a female officer and women police were better at providing practical advice and assistance to all victims of crime. In the USA, the performance studies were important for successful litigation taken under the Equal Opportunity Act 1972 which found that institutionalized impediments to women police did not relate in any way to genuine job requirements.

Research by Rabe-Hemp (2008) suggests that women officers themselves believe they bring some distinctive characteristics to the task compared to male colleagues, i.e., greater empathy, communication skills, and fewer forceful behaviors. Women she interviewed (in the United States) believed that they were better at serving the needs of women and children, especially those who had been subject to violent or sexual abuse, and that they were more likely to act as advocates to prevent their re-victimization. Community policing was a particular area where women (and men) legitimized more feminine ways of policing. Community policing stresses communication, familiarity, and building of trust and rapport between the citizen and community member rather than the rapid response required of reactive emergency policing.

Physical Aspects

The performance studies were also important for destroying the myth that women could not cope with the physical demands of policing, i.e., that women police lacked sufficient strength to control resisting subjects and that they were a safety risk to themselves and other officers. The studies provided support for the view that upper body strength is less important for effective policing than having good negotiating skills and physical restraint skills that involve a minimal force approach. By these criteria it appears that women are as capable, if not more capable, than men. Studies also found that female police were more likely to have a calming effect in confrontational situations, whereas male police tended to contribute to the escalation of conflict (for more detail of these studies, see Lunneborg 1989).

Specialist Roles

As implied above, women, when first entering a national police force, are unlikely to have the full range of roles available to them. Within their separate police units, they are able to achieve higher ranks, but since most national forces have significantly smaller percentages of women, then inevitably force of numbers will limit their proportionate share of the highest ranks within assimilated or integrated forces. It is noteworthy, even in forces that claim to treat women and men equally, they are likely to be excluded from public order or counterterrorism deployments.

As forces achieve integration, women begin to perform tasks in areas of specialism that were previously limited to men. In England and Wales, e.g., it is possible to track the increasing percentage of women deployed in specialist roles (Home Office 2010) (Fig. 5).

The police services in England and Wales were legally mandated to integrate men and women officers by the 1975 Sex Discrimination Act. Migration of women into specialist areas of policing, dominated by men during the period of segregation, was slow and is still incomplete. In 1990, women predominated in the investigation of child crime, sexual offenses, and domestic violence. Their percentage share since then has actually dropped. Apart from the vice squad, traffic departments, and firearms units, which remain persistently gender segregated, women have been gaining access to areas of specialist operations and in several are approaching or have surpassed the 25 % threshold.

Ethical Standards

The issue of physical ability and use of force overlaps with the issue of police integrity and use of excessive force. Ethical climate surveys suggest women are not necessarily more ethical in their attitudes or more likely to report wrongdoing by other officers (Waugh et al. 1998). However, the research shows clearly that women attract fewer complaints from the public, including fewer complaints regarding assaults and excessive force. For example, an Australian study found that male police attracted two-and-a-half times as many allegations of assault as female police (e.g., Waugh et al. 1998). An observational study, also in Australia, found that

Women in Policing, Fig. 5 Percentages of women serving in specialist roles in an integrated police force: England and Wales as a case study (Sources: Brown (1998), Flannery (personal communication), Home Office (2010))

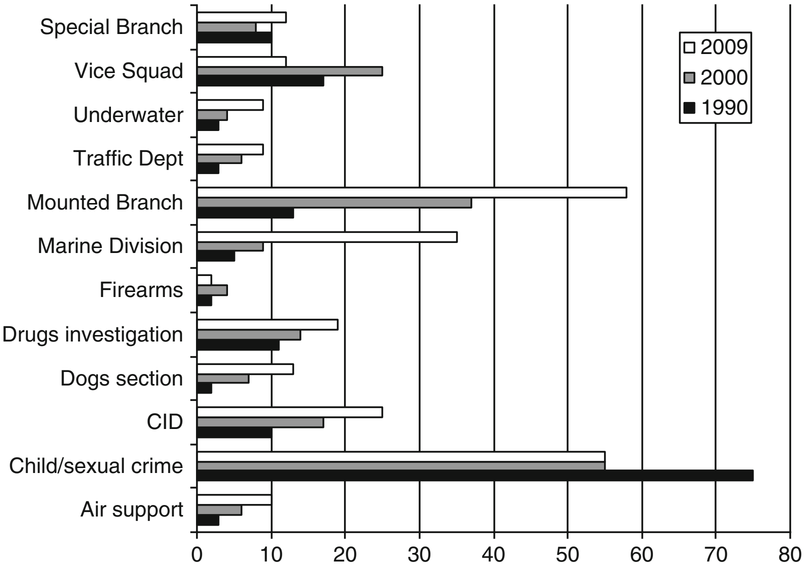

male officers were twice as likely as female officers to engage in threatening behavior and physical contact with members of the public, which in turn elicited greater resistance and aggression (Braithwaite and Brewer 1998). Research has similarly shown that women police appear to be less "trigger happy" and much less likely than male officers to fatally shoot offenders. In the USA, successful litigation for excessive force by male officers was estimated to cost taxpayers up to 5.5 times the amounts associated with the actions of female officers.

Barriers for Women Police Officers

Discrimination and Harassment

As described earlier, the entry of women in policing in most jurisdictions has been, for the most part, contested. Discrimination and harassment have been a consistent feature of policewomen's experiences (see review by Natarajan (2008)) and appear in many jurisdictions (see, e.g., Somvadee and Morash 2008). Forms of sexual harassment include sexist jokes, comments about women's appearance, use of derogatory or offensive language, being pressured for dates, and deliberate sexual contact, while discrimination includes differential deployments, blocked promotion, and "safer" station house assignments. As demonstrated above, women have not fully accessed the rank structure and areas of specialist policing work. But have things changed due to legislation, department policies, and practice reforms and gender mainstreaming? Somvadee and Morash (2008) looked at the experiences of female law enforcement personnel in the USA. They found that many women reporting potentially sexually harassing behaviors did not experience this as sexual harassment. Some women tolerated the behaviors as the price of fitting into a male occupational culture. Other women described how they confronted offensive language and behavior, resulting in men apologizing. What women in this survey found most bothersome was the notion that they, women, could not do the job. Foster et al. (2005) undertook research in England and Wales and reported that women officers still felt excluded, undermined, and undervalued. The participating officers felt that overt forms of discrimination were not so widespread but that they had to work harder than male colleagues to prove themselves equally capable. As alluded to above, policewomen in the women-only police stations perceived that their career opportunities were thwarted as a consequence of secondments out of mainstream policing. In sum, discrimination and harassment is still being experienced by women police officers.

Sexual harassment is very difficult to prevent in male-dominated organizations. The suggestions from the research literature are that the dynamics are complex and that there is interplay of other factors such as work environment which provides better explanations than demographics alone. Moreover, the presence and implementation of comprehensive anti-discrimination and harassment policies are insufficient by themselves as often victims were fearful of complaining or disenchanted that complaining would make any difference. Somvadee and Morash (2008), p. 494) advise that "police organisations need to consider the implications of high levels of continuing exposure to sexually oriented jokes and remarks" not least to the psychological harm experienced by recipients and reputational damage to the police itself.

Stress and Health

Police occupational culture is critical in understanding the pressures exerted on women officers that subjects them to some different stress experiences. Stressors impacting police officers tend to be divided into those that form part of the operational tasks, such as dealing with crime, public order, and arrests, and aspects of the police organizational climate, including shift work, promotions, and relationships with colleagues. For the most part, both men and women officers suffer similar stress loads and adverse reactions to the bulk of task deployments. However Thompson Kirk and Brown (2006) show that men and women officers experience their work environments differently, in particular concerning interpersonal conflicts. They found that interpersonal stress accounted for 31 % of the variance, organizational stress for 15 % of the variance, and operational stress for 8 % of the variance. The first two factors were more influential for senior women officers. This suggests the importance of social support both in operational confidence in women's abilities and organizationally in flexibility of working arrangements. Maintaining work life balance and managing domestic care responsibilities appear to weigh on women officers more heavily than men. This is in part due to cultural expectations about the roles of women, police organizations' ambivalence about flexible working arrangements, and to some extent women's own coping adaptations to domestic pressures (Dick and Cassell 2002).

Accessing Senior Rank

Relatively little research is available, charting the experiences of senior women police officers, with the work of Marisa Silvestri (2003) and Dorothy Schulz (2004) being notable exceptions.

Van Ewijk (2011b) shows how percentages of women officers in higher ranks drop more than what would be expected on the basis of the percentage of women officers in the average year of entrance of officers of that rank. This is so even in a police force where all officers enter at the same basic level and follow the same career path, and the outflow of (women) officers before retirement is practically nonexistent.

Silvestri (2003) records barriers including on call requirements, a promotion track demanding geographic flexibility, and career interruptions due to maternity breaks. This is confirmed by Schulz (2004) who describes how women achieving senior ranks in the United States made trade-offs in the sense that many of them were single (finding career and geographic mobility difficult to combine with domestic life) or rarely had more than two children to minimize disruption and care arrangements.

Policy Implications

When considering the progress made by women into policing, then, there are few police forces with generic policing tasks that omit women entirely, Saudi Arabia being among the exceptions. In Western democracies, all types of policing activity are open to women, but they are still likely to be a minority in specialist areas, such as those involving firearms, counterterrorism, traffic, and riot control duties. In practice, women are more likely to be present in general patrol duties and administrative and supportive functions, especially among non-sworn personnel. With some exceptions, sworn (or warranted) female officers have a higher presence among the rank and file of operational officers than in the higher ranks. In developing and reconstructed nations outside of Europe, women are more likely to be either employed in segregated units or engaged in limited roles focused around women and children. Even so, many police forces report difficulties concerning the retention and promotion of female officers, and a few have properly considered the implications of the increasing gender ratio of women officers in terms of women's great rates of sickness absence and maternity leave.

Removing formal barriers to women in policing and creating nondiscriminatory processes in recruitment, training, deployment, and promotion are relatively straightforward tasks. First steps include removing any female quotas and removing physical ability tests in recruitment – especially military-style obstacle course tests. These "physical agility tests" could be replaced with a general health test that assesses an applicant's ability to be trained in the basic physical requirements for general duty policing as demonstrated by research and case law (Lonsway 2003). Education should also be given a high profile in selection criteria. In addition, including female representation on recruitment and promotion panels may guard against bias in selection. Enforceable policies against sexual harassment and sex discrimination should also be introduced with an internal unit responsible for their implementation. These relatively straightforward measures are likely to make a contribution towards equity, or near equity, in the ratios between female applications and female appointments and also take the edge off the more blatant forms of sexual harassment and sex discrimination on the job.

Research in developed democracies indicates that these relatively passive policies of nondiscrimination will lead to a rapid increase in the number of female recruits to between 25 % and 33 % and a slower increase in the number of female sworn officers to between 20 % and 25 % (Prenzler et al. 2010). However, the goal of full numerical equality will be thwarted by a number of factors. The activation of anti-discrimination law often requires complaints and litigation, but aggrieved parties can be deterred by cost factors, ignorance of the law, and fear of retribution or difficulties proving bias. Disadvantaged groups such as women also frequently "self-select" out of applying for employment, promotion or specialist roles because of career expectations generated in childhood, the potential conflict between work and child-rearing responsibilities, and expectations of discrimination. Just as many women police become highly skilled and eligible for promotion, child-rearing commitments often mean they drop out of policing and do not return – even when maternity leave is available (Prenzler et al. 2010). In response to these problems across occupations, affirmative action programs have been introduced to provide greater encouragement and support to women. In some police departments, particularly in the USA, resistance to integration has led to the imposition of quotas on the recruitment of women (and other minority groups), often through court-ordered consent decrees. In other jurisdictions, the focus has been more on voluntary proactive support strategies, such as targeted recruitment campaigns, pre-application classes, career development courses and mentoring programs, flexible employment options, childcare advisory services, and rejoining policies (Prenzler et al. 2010).

Conclusions

The research literature reveals a difference of view about the current status of policewomen. On the one hand, it has been proposed that a gender-neutral position prevails, whereby the sex of the officers has a diminishing effect on work-related attitudes as recruits who are similarly socialized will converge in their behaviors. On the other hand, the continuing gender discrimination position argues that gender still matters, that the increasing percentages of women have fragmented traditional police cultural, and that training cannot erase the significance of gender socialization and gendered experiences so that men and women will do policing differently (Lundman 2009).

Some research supports the gender-neutral position, such as Lundman's own, whereas other research points to differences in men and women's approach to policing (see Lunneborg 1989 for summary). Lonsway (2006) indicates that men and women officers in an American force she studied had common motivations for joining the police and, once controlling for other demographic variables, gender alone was not a statistically significant differentiating variable in terms of job satisfaction, feelings of job security, and likelihood of remaining in the force. Women have reached the higher ranks and some have been promoted to lead their forces in Western democracies, but the gender ratio still is vastly in favor of men. Lonsway suggests that there are still impediments for some women in policing, notably the sustained negative attitudes towards them by co-workers, performance double standards, whereby women still perceived they have to do better than male colleagues to be seen as equally competent. She also reports that other women believed they faced no such barriers. Yet other research argues that actually women, by preference, preserve some of the differentiations between them and their male colleagues because either they are not attracted to working in male-dominated departments, such as traffic or firearms, or domestic convenience limits long working hours (Dick and Cassell 2002). Women in Arab nations would also appear to show no appetite for gender integration (Strobl 2010).

Research data indicate that increasing the number of women does not have an adverse effect on the ability of police to carry out their duties – provided standard controls are in place in recruitment, training, and supervision and monitoring. Rather, it is highly likely that more women will make for better policing. Actually, survey results show that improved relations between the police and the community will enhance community cooperation with the police and this process is boosted with more women police. However, there does not appear to be any empirical evidence at present to support the view that increasing the number of women police will lead to more crime being solved or reduced. Positive effects are to be seen in police conduct and police-community interactions with more women officers with de-escalation of conflict situations and increased support for victims of crime. These are, of course, utilitarian arguments. There is also the stronger, arguably, overriding, argument that a policing career should be open to women if they want one on equity grounds alone.

There is indicative research that the degree of women's incursion into policing may tell us something about the progressiveness of a force and the general state of equality and democracy in the country concerned (Ivkovic 2008). (Davenport 2008) argues that without a critical mass of women in policing, there is a democratic deficit. If policing is to actively engage with citizens, she argues that those citizens must recognize the value of the service, feel that the service is accountable to them, and be able to evaluate its performance. One way to achieve this is for the police service to "look" more like the communities it serves to underpin its legitimacy in a democratic society.

The move towards full integration for women police is perhaps more of a concern in Western democracies where the ideal of equality is well established following the "second wave" of feminism in the 1960s and 1970s. However, there are different paths to equality, and recent work by Natarajan (2008) suggests that in some jurisdictions it is advantageous to adopt a two-step approach to equality. She argues that in some societies the creation of separate police units is justified by the level of hostility to female police and the specific needs of female client groups. Research in other developing nations supports the case for a more culturally appropriate, staged approach to the involvement of women in policing (Strobl 2010; Natarajan 2008). Natarajan (2002) found an evolution in the views of Indian women police in traditional roles towards a preference for greater integration as their experience grows and they become more confident in their ability to carry out the full range of police duties. She concluded that models of policewomen's progression such as Brown's which was developed to account for the experiences of women in Western democracies can be applied more widely but that progress towards integration in traditional cultures is likely to be slower and suffer more reversals than in Western countries. Strobl (2010) proposes that women involved in policing in the GCC countries are deployed, as are policewomen elsewhere, in duties related to children and young offenders, yet are unsympathetic to the plight of migrant working women. Moghadam (2010, p. 4) observes "while women should be present in diverse occupations and professions for the sake of equality and rights, feminists should not be under any illusion about the extent of institutional change, especially in such domains of the state's repressive apparatus as the police force or the military apparatus". Similarly, Clarke (2008) showed that while postcolonial states in Africa took on the institutional machinery left by the colonial power, there was little critical analysis of new models, especially the ones where women may play full part in. Thus, an "add-in" women approach was adopted without questioning the gendered premise of security sector and its role on government.

So there do seem to be some universals: women's entry into policing was delayed and contested, women police everywhere suffer some form of sexual harassment and discrimination, and there is no evidence that having women in policing is detrimental to performance. The efficacy of policewomen's departments or women-only police stations is equivocal with often support for early successes but more dubious evidence of the sustainability of those successes. There is clear evidence that women are less likely to use or misuse physical force. As yet, there is no evidence of truly integrated police forces.

Recommended Reading and References

Asiwaju K, Marenin O (2008) Policing Nigeria. In: Hinton MS, Newburn T (eds) Policing developing democracies. Routledge, Abingdon

Banks C (2001) We have to prove we are not just women; policewomen in Papua New Guinea. Int J Comp Criminol 1:40–76

Braithwaite H, Brewer N (1998) Differences in the conflict resolution tactics of male and female police patrol officers. Int J Police Sci Manag 1(3):288–300

Brown JM (1998) Aspects of discriminatory treatment of women police officers serving in forces in England and Wales. Brit J Criminol 38:265–283

Brown J, Heidensohn F (2000) Gender and policing: comparative perspectives. Macmillan, London

Brown JM, Hazenberg A, Ormiston C (1998) Policewomen; an international comparison. In: Mawby R (ed) Comparative policing: issues for the twenty-first century. UCL, London

Bureau of Justice Assistance (2001) Recruiting and retaining women: a self-assessment guide for law enforcement. US Department of Justice, Washington, DC

Clarke Y (2008) Security sector reform in Africa: a lost opportunity to deconstruct militarised masculinities? Fem Afr 10:49–66

Davenport S (2008) Affirmative action-setting the record straight. Grapevine 12(Spring):12–13

Dick P, Cassell C (2002) Barriers to managing diversity in a UK constabulary: the role of discourse. J Manag Stud 39:953–976

European Network of Policewomen (2008) Facts & figures 2008. www.enp.nl. Last access Nov 2009

Foster J, Newburn T, Souhami A (2005) Assessing the impact of the Stephen Lawrence inquiry. Home Office Research Study 294 Research Development and Statistics Directorate

Heidensohn F (2000) Sexual politics and social control. OUP, Milton Keynes

Hinton MS, Newburn T (eds) (2008) Policing developing democracies. Routledge, Abingdon

Home Office (2009) Police service strength, July 2009. http://www.homeoffice.gov.uk/rds/pdfs09/hosb1309.pdf. Last access June 2010

Home Office (2010) Assessment of women in the police service. Home Office, London. http://library.npia.police.uk/docs/homeoffice/assessment-omen-police-service.pdf. Last access 29 Nov 2011

Igbinovia PE (1981) Pattern of policing in Africa – the French and British connections. Police J 54:123–156

Irving R (2002) Career trajectories of women in policing. Trends and Issues in Crime and Justice no 370

Ivkovic SK (2008) A comparative study of public support for the police. Int Crim Justice Rev 18:406–434

Lonsway KA (2003) Tearing down the wall: problems with consistency, validity, and adverse impact of physical agility testing on police selection. Police Q 6(3):237–277

Lonsway K (2006) Are we there yet? The progress of women in one large law enforcement agency. Women Crim Justice 18:1–48

Lundman RJ (2009) Officer gender and traffic decisions; police blue or woman too. J Crim Justice 37:342–352

Lunneborg PW (1989) Women police officers: current career profiles. Charles C. Thomas, Springfield

Martin S (1980) Breaking and entering; policewomen on patrol. University of California Press, Berkley

Martin S, Jurik N (1996) Doing justice, doing gender. Sage, Thousand Oaks

Moghadam V (2010) Women, structure and agency in the Middle East; introduction and overview. Fem Form 22:1–9

Moon B, Morash M (2008) Policing in South Korea; struggle, challenge and reform. In: Hinton MS, Newburn T (eds) Policing developing democracies. Routledge, Abingdon

Natarajan M (2002) Women police in Tamil Nadu, India; a tale of two cohorts. Int J Comp Criminol 2:201–224

Natarajan M (2005) Status of women police in Asia; an agenda for future research. J Women Polic 17:45–47

Natarajan M (2008) Women police in a changing society: back door to equality. Ashgate, Aldershot

National Center for Women and Policing (2002) Equality denied; the status of women in policing. National Center for Women and Policing, Beverly Hills

Prenzler T (1994) Women in Australian policing; an overview. J Aust Stud 42:78–88

Prenzler T, Fleming J, King A (2010) Gender equity in Australian and New Zealand policing: a five year review. Int J Police Sci Manag 12(4):584–595

Pruvost G (2009) A profession in process: the atypical rise of women to the high rank of police “commissaire” in France. Sociologie du travail 51S:e34–e48

Rabe-Hemp C (2008) Female officers and the ethics of care: does officer gender impact police behaviors? J Crim Justice 36:426–434

Reams MB, Nelson (2008) Neofeudal aspects of Brazil’s public security. In: Haberfeld MR, Cerrah I (eds) Comparative policing; the struggle for democratisation. Sage, Los Angeles

Schulz D (2004) Breaking the brass ceiling: women police chiefs and their paths to the top. Praeger, Westport

Silvestri M (2003) Women in charge; policing gender and leadership. Willan, Cullompton

Somvadee C, Morash M (2008) Dynamics of sexual harassment for policewomen working alongside men. Polic; Int J Police Strate Manag 30:485–496

Strobl S (2010) Progressive or neo-traditional? Policewomen in Gulf Cooperative Council (GCC) countries. Fem Form 22:51–74

Timmins W, Hainsworth B (1989) Attracting and retaining females in law enforcement: sex-based problems of women cops in 1988. Int J Offender Ther Comp Criminol 33(3):197–205

Van Ewijk AR (2011) Dynamics of diversity within the Mossos d’Esquadra. In: Zapata-Barrero R, Van Ewijk AR (eds) Spheres of diversity: from concept to policy. CIDOB Edicions, Barcelona, pp 107–130

Van Ewijk AR (2012) Diversity within police forces in Europe: a case for the comprehensive view. Policing 6:72–92

Waugh L, Ede A, Alley A (1998) Police culture, women police and attitudes towards misconduct. Int J Police Sci Mgt 1:288–300

Women of Color and Crime

Hillary Potter[1] and Kenly Brown[2]
[1]Department of Sociology, University of Colorado Boulder, Boulder, CO, USA
[2]Department of Sociology, University of California, Santa Barbara, CA, USA

Overview

The most common profile of a woman imprisoned is one who is impoverished, is undereducated, has sporadic employment, and is a woman of color. While scholars in years past have cited the need for more research on women offenders of color (Britton 2000; Enos 1997; Russell-Brown 2004), there remains a dearth of information on this group, especially from a feminist perspective. Because of the varied perspectives of women of color, scholars argue that mainstream criminological theories may be inadequate to explain offending behavior by women of color (Joseph 2006; Russell-Brown 2004). Further, studies that have included women of color often fail to disaggregate women from men of color or from White women (Joseph 2006; Russell-Brown 2004).

This entry will focus on a variety of female offenders within the following races and ethnicities: African American, Latina, Native American, Asian, and Asian American. A summary of the theoretical background of feminist work and what feminist scholars have found to be pertinent when studying women of color is provided. This entry includes statistical data and analyses on offending, arrest, and incarceration rates. The races and ethnicities chosen are by no means exhaustive, but they represent growing interest in research on women of color involved in crime, particularly from a feminist or intersectionality perspective.

Theoretical Background

Mainstream feminist theory comprises three eras of development, referred to as waves, which began in the 1700s. *Mainstream* feminism or feminist theories include those efforts made toward gender equality by groups of predominantly White women feminists. The use of the term *mainstream* relates to the considerable attention – both negative and positive – given to the efforts of these women to seek social and economic equality with men, as opposed to that afforded smaller, marginalized groups of feminists. Mainstream feminist theory places gender as the primary consideration in women's liberation efforts (Hooks 2000). However, women of color have expressed difficulty in identifying with mainstream feminist theory because of its focus on this single aspect of womanhood and because the lives and concerns of White middle-class women were placed at the forefront of the liberation efforts (Collins 2000). Women of color regularly convey that they deal not only with issues of gender inequality but with racial inequality, as well (Crenshaw 1991; King 2003). It is this status, Kimberle Crenshaw (1991) argued, that relegates women of color to an invisible class and pulls these women's loyalties in two directions, that is, feeling the need to either choose between being loyal to feminist ideas and being loyal to their racial or ethnic community. A pioneer of intersectionality studies, Crenshaw has advocated that race and gender are not exclusively mutual, but instead these ascribed characteristics create a multitude of experiences for women that can no longer be seen as monolithic.

Studying women of color offenders allows for a more nuanced and holistic picture of women who participate in criminal offending. However, criminologist Vernetta Young (1986) emphasizes that few studies consider the race effect on gender and, likewise, few studies consider the gender effect on race. Young asserts that when African American women, in particular, are discussed in the research, they are often addressed in relation to Black men and White women, resulting in the unworthiness of research on Black women's experiences with crime. More than two decades later, criticisms continue to be expressed about the minimal amount of research on criminally offending African American women, suggesting that African American women do not warrant attention within the discipline of criminology (Russell-Brown 2004). The same criticism can be lodged about research focusing on other women of color, as even fewer investigations have been undertaken on Asian American, Latina, and Native American women.

The incorporation of a feminist perspective that insists on viewing multiple identities in conjunction with one another in criminology has aided in moving away from a view that there is something pathological about women of color's involvement in crime. Generally, feminist scholars assert that social disparities and the "get tough on crime" policies are major factors explaining the high numbers and rates of women of color found in the criminal justice system. Worsening economic conditions are seen as a particularly important factor in the criminal activity of women of color (Diaz-Cotto 2006; Johnson 2003b). There is ample research that shows there is a connection between one's social class and criminal behavior where those who encompass the more impoverished economic statuses in society commit more crimes than those who occupy middle or affluent statuses (Joseph 2006). Discrimination due to socioeconomic status, race, ethnicity, and gender is also purported to determine the inequitable representation of women of color in the criminal justice system (Diaz-Cotto 2006). Finally, similar to White women in the criminal justice system, a high proportion of women of color engaging in crime have experienced abuse and violence by family members, intimate partners, and others. Feminist scholars focusing on women of color – and placing women of color at the center of the analysis – highlight these factors as major contributors to the overrepresentation of women of color in crime and the criminal justice system.

Feminist and intersectionality theoretical analyses of the criminal experiences of women also place considerable emphasis on the effects of stereotypes. Positive and negative images abound when considering women of color. However, for some women of color, the adverse images surpass the more endearing images. Feminist scholars assert that stereotypes and negative images of African American women are a major contributing factor to their overrepresentation in the criminal justice system (Johnson 2003a). These include images of African American women as jezebels, sapphires, welfare queens, crack whores, matriarchs, and masculine and emasculating (Collins 2000; Goodwin 2003; Johnson 2003a; Huey and Lynch 1996; Young 1986). In the 1986 article "Gender Expectations and Their Impact on Black Female Offenders and Victims," Vernetta Young argues that criminologists' treatment of African American female offenders focuses primarily on how they are not like other women offenders and on the similarities between crimes committed by African American women and men. Huey and Lynch (1996) also discuss how stereotyping of African American women dictates how mainstream criminology has interpreted African American women's participation in crime. One example is the view of African American women as matriarchs or welfare mothers; these views support the ideology that the African American woman is responsible for all crimes committed by African Americans because she has emasculated men and is an inadequate parental figure for her sons who will, therefore, become irresponsible men. Assumptions are also made about other women of color and then connected to their offending behaviors, which will be highlighted in some research discussed in this entry.

Nature of Offending

As indicated above, there is a need for more research in criminology that considers, simultaneously, the intersection of race and gender and the comparisons between groups by disaggregating the study subjects. Although statistical information regarding the race of women under correctional supervision in the United States is made available from government sources, few additional statistical data are readily available that provide a comparison by race among women offenders within large data resources such as the FBI's Uniform Crime Reports and the National Crime Victimization Survey. While statistical information on variations between females and males and between certain race groups is included in these data sources, only a small amount of information is provided on gender-by-race comparisons by the US federal government agency that releases these data. However, some criminological research, which will be detailed later, has utilized these datasets to analyze gender by race. Further, a few other statistical analyses have been performed to gauge the overrepresentation of women of color in criminal offending and involvement in the criminal justice system. For instance, a study analyzing the Monitoring the Future survey of high school students across the United States found race/gender variations among girls' reports of their use of violence (Goodkind et al. 2009). African American girls were found to have reported engaging in fights at greater rates than girls of other races. The fighting rates reported by African American girls demonstrated that they are as likely to report fighting as Asian American boys and White boys (but less than Black, Latino, and Native American boys). Latinas, Whites, and Asian Americans succeeded African American girls' rates of fighting, and Native American girls reported the lowest rates of fighting.

Studies such as the Monitoring the Future survey provide some indication of prevalence of violence, showing differences by race and gender, but feminist and intersectional theoretical analyses provide explanations for the higher reported rates of violence among certain girls of color – in particular, African American girls. Nikki Jones's (2010) ethnography of the lives of African American girls who engage in violence affords a nuanced understanding of the social forces that propel certain girls to use violence. Examining African American girls living in low-income urban neighborhoods that experience a high level of street violence, Jones found their lives revolved around what she identifies as the three Rs: reputation, respect, and retaliation. Jones's Black feminist analysis dissects the negotiations that African American girls in violent urban settings must make between their identity, gender expectations, and the violence with which they are confronted in their communities.

Regina Arnold's (1990) research also provides insight into the way certain life conditions, beginning in childhood, lead some African American girls to engage in violence and other offending behaviors. Utilizing a feminist and intersectionality perspective, Arnold considered the social forces of patriarchy, racism, and economic marginality. In her study of 50 African American women in a city jail and 10 African American women in a state prison, Arnold discovered that gender- and class-based oppression, along with being blamed by criminal justice officials for their own victimization, was a pattern among the women she interviewed. This criminalization, while the women were young girls, was exacerbated by the women's "structural dislocation" from their families and their schools during their formative years. The women's young lives were besieged with sexual and other physical abuse, poverty, and inadequate educational experiences. This dislocation and lack of a stable familial alternative propelled the women into deviant and delinquent behaviors early in their lives, such as running away from home, thievery, and truancy. Criminal labels were applied to these women from their youth through adulthood. As the women continued engaging in criminal behaviors when they became adults, addiction to drugs became another obstacle to overcome. In sum, from an early age, the women in Arnold's study engaged in delinquent behaviors as an "active resistance" to their victimization.

Beth Richie's (1996) research also utilized a feminist perspective to reveal circumstances that lead women to crime. Richie explored the effect of abuse as a pathway to jail or prison for African American women and, in so doing, developed a theory of "gender entrapment." Richie explains why the women remained in abusive relationships and how being in these relationships led to women committing criminal acts by highlighting the compounding effect of African American women's experiences with marginalization in public arenas. Utilizing comparison groups of African American women in jail who had not been battered and battered White women, Richie concluded that battered African American women participated in criminal offenses as (a) a means of delaying further violence by their intimate partners, (b) an extension of their loyalty to the African American men who abused them because of a sense of racial identity and unity and the need to maintain some semblance of closeness with their abuser, (c) a perception of the duty to protect African American men from racism they experience in the public sphere, and (d) a way to get arrested and incarcerated in order to escape the violence from their intimate partners, for at least a short period of time.

Like Richie, other feminist scholars have also considered the outcomes of intimate partner violence against women of color. There is a notable scarcity of research on Asian and Asian Americans pertaining to criminality. Nevertheless, two feminist scholars who studied intimate partner violence among Asians and Asian Americans suggest that their violence is often a reaction to victimization by their partners, and it is criminalized within many US jurisdictions. Margaret Abraham (2000) considered South Asian immigrant women's resistance in her research. Although few women in Abraham's study used physical violence against their batterers, Abraham demonstrates how the emotional captivity of abuse, coupled with cultural standards of womanhood, can be a catalyst to resistance that would be deemed criminal behavior by many state laws.

In her study of Vietnamese immigrant women who had been abused by their intimate partners, Hoan N. Bui (2004) documents the women's legal consequences for physically defending themselves from their abusive partners. The implementation of mandatory arrest laws throughout the United States has had unintended outcomes for battered women who fight back. As evidenced in Bui's study, battered women were arrested because of physical evidence visible on their partners and many of the partners' exaggeration or misrepresentation of the incident. There are additional consequences for women of color. Bui demonstrates that the women's gender, race, class, and immigrant status place them at a disadvantage for receiving the assistance they need by considering them as offenders instead of victims.

Few scholars focus on female Native Americans' involvement in the criminal justice system as perpetrators of crime. However, like the approaches with Asian and African American women, the feminist and intersectionality work on violence against Native American women can be used as a basis to contemplate offending among Native American women. Native American activist and feminist scholar Andrea Smith (2005) provides a macrolevel approach to Native American women's involvement in criminal activity by focusing on state-sponsored and corporate-sponsored violence against Native American women. As with other research that has found a correlation between victimization and future offending among women, criminogenic effects of state and corporate violence may too compel criminal activity among Native American women.

The effects of drug use also play a role in women's offending. Self-report surveys, such as the Monitoring the Future survey, provide some indication of variations in drug use by race and gender. White girls had the highest reporting rates for alcohol use. Native American girls reported the next highest rate of alcohol use, followed by Latinas, Asian Americans, and then African Americans. For marijuana use, Native American girls reported the highest use, closely followed by White girls. Latinas, African Americans, and Asian Americans (in descending order) reported the lowest levels of marijuana use. Other research, conducted with women in prison, has shown that in comparison to other women who are imprisoned, Latinas are five to eight more times likely to abuse alcohol, ten times more likely to abuse drugs, and 27 times more likely to use cocaine (Covington 1998).

Feminist and intersectionality approaches provide important interpretations of these racial differences in drug use and participation in illicit drug markets. Lisa Maher's (1997) ethnography of women drug users considers the intersectional identities of her study informants. Importantly, Maher goes beyond the Black/White binary of racial analysis by including the experiences of Latinas and White women. Women in the study, mostly Latina, were typically involved in the consumption of drugs, in dealing drugs when men were found to be scarce due to arrests, and in sex work to keep up their addictions. Although identities of age, immigrant status, and familial and neighborhood ties affected the way labor was divided in the drug and sex markets, Maher asserts that race and gender were the most salient factors. Race and gender determined the extent of the earnings for women working in the illicit drug and sex trades, and White women sex workers initially were able to demand, and receive, higher rates for sex work than women of color.

Arrest and Prosecution

According to federally collected statistics, there has been a steady decrease in the overall US crime rate over the past couple decades; however, there has been a significant rise in the number of arrests and incarceration of women in the United States criminal justice system. The War on Drugs has had a more detrimental effect on women than on men with regard to the substantial rise in arrests and incarceration rates of women for drug-related offenses (Diaz-Cotto 2006). This is evident in that between 1980 and 2009, the female arrest rates tripled for drug possession (a 225 % increase) and doubled for drug sales or manufacturing (a 237 % increase) (Snyder 2011). Male drug arrest rates also increased, but at a lower rate, with a 104 % increase in male arrests for drug possession and a 205 % increase for drug sales or manufacturing. Snyder (2011) concludes that the male arrest rate for drug sale and manufacturing in 2009 was 68 % above the male arrest rate in 1980, while the female drug sale and manufacturing arrest rate in 2009 was 123 % above the 1980 rate. Women and girls have also been arrested at greater rates than males for other offenses. In 1980, females accounted for approximately 12 % of the arrests for aggravated assault. The proportion of female arrests for aggravated assault increased to 29 %

in 2009, which is double the rate in 1980. In contrast, the male arrest rates for aggravated assault in 1980 and in 2009 are similar. However, when examining the growth rate in arrests for aggravated assaults between 1980 and the peak year of 1995, female arrests for this offense increased by 150 % compared with a 63 % increase in male arrests. Similar arrest trends are evident for robbery, simple assault, and larceny-theft.

Certainly, women of all races have been significantly affected by policy and legal enhancements that have ensnared droves of women in the criminal justice system. However, statistical evidence indicates overrepresentation in arrest, conviction, and incarceration rates of women of color. Some research has established that African American women have a greater likelihood than White women of being arrested and charged for more serious crimes (Chesney-Lind 1997; Greenfield and Snell 1999). Analyzing the FBI's Uniform Crime Reports for three large states (California, Florida, and New York), Coramae Richey Mann's (1995) research also delineates variations in arrest rates of women by race. Women of color were found to have been arrested more frequently for violent crimes such as aggravated assault and robbery. Public order crimes, such as alcohol- and drug-related offenses and sex work, were the most frequent arrests for White, African American, and Latina women. However, Latinas and African American women have been arrested for sex work at higher rates than their proportion of the general population. The most frequent types of offenses Native American women are arrested for were drunkenness, thefts, and driving under the influence of alcohol. Only one of the three states (California) reported arrest data for Asian American women. The most frequent arrests for Asian American women were for theft-related offenses, while drug-related offenses were among the least frequent arrests. It is important to note that there are different ways crime is depicted based on race and gender depending on the source of information.

The experiences of women of color in the criminal justice system have been afforded growing attention from feminist scholars. One study addresses the utilization of the cultural defense tactic that can either reduce accountability or exacerbate punishment for Asian Americans. Legal scholar and critical race feminist Leti Volpp (1994) provides an extensive account of the influence cultural defenses can have on court cases specifically within the Asian American population. The case of *People v. Helen Vu* is one in which Vu is a female defendant who strangled her son with a cord from a window blind. She was ultimately convicted of second-degree murder and given 15 years to life in prison. When the defense counsel constructed a cultural defense case, the court ruled that the defendant committed the crime out of her own volition and ignored the defense attorney's argument that it was connected to her Asian background (Volpp 1994). Consequently, Vu did not meet her gender stereotype, which is to take care of her son as a mother is supposed to, and since she deviated from this role, she received a severe sentence (Volpp 1994). Volpp suggests that this case study shows the importance gendered stereotypes and expectations can play in a case and that if one were to meet those stereotypes, the less severe the punishment will be.

Involvement in drug markets is the leading crime for which women are charged and convicted. Scholars have researched the influence that culture can play on the verdicts Latina women receive. Holmquist (1997) analyzed the significance of Latinas meeting cultural expectations of docility and passivity to their partners and their involvement in the drug market. Two case studies were used, *United States v. Castaneda* and *United States v. Avila*, to exemplify how women are either taken advantage of by their partners or their naïveté of the work they are doing for their partners. The court found Maria Avila and Ana Reyes responsible for the distribution of cocaine and laundering money, even though their defense team argued that they were blindly following Jose Reyes. However, in Leticia Castaneda's case the court dropped her possession of six firearms to one because her defense was that she had no idea firearms were involved in the narcotics drug trafficking ring, and she was only following orders from her partner. Conclusively, the cultural defense upholds racial and gender stereotypes of Latina women when convicted of criminal acts; these stereotypes can either assist or be detrimental to these women. Feminist perspectives on Latinas' involvement in criminal activities have revealed that Latinas are often rebelling against the subordinate roles that society, in general, expects them to play (Diaz-Cotto 2006).

Sentencing and Incarceration

The number of women incarcerated, regardless of race, has gone up by 700 % since 1977, but Latinas and African American women make up 70 % of this population (Tapia 2010). While White women have a greater chance of being placed on probation after arrest and conviction, women of color are disproportionately sentenced to prison (Greenfield and Snell 1999). Almost two-thirds of women who are incarcerated are women of color, while two-thirds of women who are on probation are White women (Bloom et al. 2004). On December 31, 2009, one in every 703 Black women and one in every 1,356 Latinas were incarcerated in US federal and state prisons, while one in every 1,987 White women was incarcerated.

The War on Drugs in the United States, beginning in the 1970s, contributed to the spike in incarceration rates of women due to the requirement of mandatory minimums and sentencing guidelines, which lead to lengthy prison sentences served by convicted drug offenders (Levy-Pounds 2007; Shelden 2004). Due to women's lower status in drug organizations, they often do not have enough information to bargain with to reduce their sentencing when questioned about the drug operation or identifying the drug ringleaders (Sokoloff 2005; Morash 2006). Women, in general, are more likely than men to be given mandatory minimum sentences for much smaller quantities of drugs (Sokoloff 2005). Currently, about 40 % of women in prison are in prison for drug-related crimes, with a majority of these being women of color (Shelden 2004). Juanita Diaz-Cotto (2006) reports that, in New York, Latinas were more likely to be imprisoned for drug-related offenses than were Latinos (males). Mann's (1995) study found racial disparities among African American, Latina, and White women regarding drug-related offense sentencing. Of the women in the study, which compared several states, 18.7 % of them were Latinas in California who had been arrested for drug offenses and represented 26.1 % of women imprisoned for drug offenses. African American women comprised 30.5 % of arrests and 34.1 % of prison inmates for drug offenses. While African American women and Latinas were overrepresented in prison for drug offenses, White women comprised 48.5 % of arrests, but only 38.3 % of prison sentences. It is evident that the consequences of the War on Drugs campaign disproportionately affected impoverished women of color and their families (Levy-Pounds 2007). As such, some feminist researchers have declared the War on Drugs as a war on women of color (Bush-Baskette 1998).

There are a small number of studies that have looked at the number of Native American females incarcerated, and official prison statistics have been challenged by some scholars as inaccurately depicting the number of Native American females incarcerated. According to one study (Abril 2003), incarcerated Native American women were undercounted in the Ohio Reformatory for Women in Marysville, Ohio. Data retrieved from the institution through responses to open-ended questions distributed to all of the women in the facility revealed discrepancies between the numbers of self-identified Native American women and the numbers reported by the prison. Results from the survey showed that the agency reported there were only two Native Americans, while 255 reported that they self-identified themselves as Native American. This finding is important to the criminal justice system because officials believe Native American women are an insignificant proportion of the prison population and therefore warrant little research attention (Abril 2003; Russell-Brown 2004). Though occurring at a slower rate than feminist research on African American women, feminist scholars have begun to consider the lack of research on the criminal justice system involvement of Native American, Latina, and Asian American women.

Conclusion

Feminist scholarship has shifted its theoretical perspective to be more inclusive of girls' and women's racial backgrounds and the influence this ascribed status can have on social attainment, or detainment. Women of color are in a unique position where social stereotypes and images and bias can influence arrest, conviction, and incarceration rates. It is essential to consider the experiences of women of color offenders from a feminist perspective in order to understand some of the unique ways in which they are propelled into criminal behavior and to help explain their overrepresentation in the criminal justice system.

Related Entries

- ▶ Feminist Criminological Theory
- ▶ Gendered Theory and Gendered Practice
- ▶ Gendering Traditional Theories of Crime

Recommended Reading and References

Abraham M (2000) Speaking the unspeakable: marital violence among South Asian immigrants in the United States. Rutgers University Press, New Brunswick

Abril J (2003) Native American identities among female prisoners. Prison J 83(1):1–13

Arnold RA (1990) Processes of victimization and criminalization of black women. Social Justice 17(3):153–166

Bloom B, Owen B, Covington S (2004) Women offenders and the gendered effects of public policy. Rev Policy Res 21(1):32–48

Britton DM (2000) Feminism in criminology: engendering the outlaw. Ann Am Acad Polit Soc Sci 571:57–76

Bui HN (2004) In the adopted land: abused immigrant women and the criminal justice system. Praeger, Westport

Bush-Baskette S (1998) The war on drugs as a war against black women. In: Miller S (ed) Crime control and women: feminist implications of criminal justice policy. Sage, Thousand Oaks, pp 113–129

Chesney-Lind M (1997) The female offender. Sage, Thousand Oaks

Collins PH (2000) Black feminist thought: knowledge, consciousness, and the politics of empowerment, 2nd edn. Routledge, New York

Covington S (1998) Women in prison: approaches in the treatment of our most invisible population. Women Ther J 21(1):141–155

Crenshaw K (1991) Mapping the margins: intersectionality, identity, politics, and violence against women of color. Stan Law Rev 43(6):1241–1299

Diaz-Cotto J (2006) Chicana lives and criminal justice: voices from el barrio. University of Texas Press, Austin

Enos S (1997) Managing motherhood in prison: the impact of race and ethnicity on child placements. Women Ther 20(4):57–73

Goodkind S, Wallace JM, Shook JJ, Bachman J, O'Malley P (2009) Are girls really becoming more delinquent? Testing the gender convergence by race and ethnicity, 1976–2005. Child Youth Serv Rev 31:885–895

Goodwin M (2003) Gender, race, and mental illness: the case of Wanda Jean Allen. In: King AK (ed) Critical race feminism: a reader, 2nd edn. New York University Press, New York, pp 228–237

Greenfield LA, Snell TL (1999) Women offenders. U.S. Department of Justice, Washington, DC

Holmquist KL (1997) Cultural defense of false stereotype? What happens when Latina defendants collide with the federal sentencing guidelines. Berkeley Womens Law J 12:45–73

Hooks B (2000) Feminist theory: from margin to center, 2nd edn. South End Press, Cambridge, MA

Huey J, Lynch MJ (1996) The image of black women in criminology: historical stereotypes as theoretical foundation. In: Lynch MJ, Britt Patterson E (eds) Justice with prejudice. Harrow and Heston, Albany

Johnson PC (2003a) At the intersection of injustice: experiences of African American women in crime and sentencing. In: King AK (ed) Critical race feminism: a reader, 2nd edn. New York University Press, New York, pp 209–218

Johnson PC (2003b) Inner lives: voices of African American women in prison. New York University Press, New York

Jones N (2010) Between good and ghetto: African American girls and inner-city violence. Rutgers University Press, New Brunskwick

Joseph J (2006) Intersectionality of race/ethnicity, class, and justice: women of color. In: Merlo AV, Pollock JM (eds) Women, law, and social control, 2nd edn. Allyn & Bacon, Boston, pp 292–312

King AK (ed) (2003) Critical race feminism: a reader, 2nd edn. New York University Press, New York

Levy-Pounds N (2007) From the frying pan into the fire: how the sentencing guidelines and mandatory minimums impact poor women of color and children. Santa Clara L Rev 47:285–307

Maher L (1997) Sexed work: gender, race, and resistance in a Brooklyn drug market. Oxford University Press, New York

Mann CR (1995) Women of color and the criminal justice system. In: Price BR, Sokoloff NJ (eds) The criminal justice system and women, 2nd edn. McGraw-Hill, New York, pp 118–135

Morash M (2006) Understanding gender, crime, and justice. Sage Publications, Thousand Oaks

Richie BE (1996) Compelled to crime: the gender entrapment of battered black women. Routledge, New York

Russell-Brown K (2004) Underground codes: race, crime, and related fires. New York University Press, New York

Shelden RG (2004) The imprisonment crisis in America: introduction. Rev Policy Res 21(1):5–12

Smith A (2005) Conquest: sexual violence and American Indian genocide. South End Press, Cambridge, MA

Snyder H (2011) Arrest in the United States, 1980–2009. U.S. Department of Justice, Washington, DC

Sokoloff N (2005) Women prisoners at the dawn of the 21st century. Women & Crim 16(1):127–137

Tapia R (2010) Certain failures: representing the experiences of incarcerated women in the United States. In: Solinger R (ed) Interrupted life: experiences of incarcerated women in the United States. University of California Press, Berkeley, pp 1–11

Volpp L (1994) (Mis)identifying culture: Asian women and the 'cultural defense'. Harv Womens J 17:57–102

Young VD (1986) Gender expectations and their impact on black female offenders and victims. Justice Q 3(3):305–327

Women Police

Dorothy Moses Schulz
John Jay College of Criminal Justice, The City University of New York, New York, NY, USA

Overview

Women entered policing in a number of countries, including the United States, at the end of the nineteenth century and in the early twentieth century as educated specialists to assist male police officers in their interactions with women and juvenile offenders and victims. Their entry into municipal policing was part of a movement of women out of the home and into charitable work that had a few years earlier led reformist women into jails and prisons.

Women's official presence in policing in the United States is often recognized with the granting of enforcement authority to social worker Lola Baldwin by the Portland, Oregon, Police Department in 1905 to protect women and girls from the expected crowds of rowdy men at the Lewis and Clark Exposition (a type of world's fair) and the department's decision in 1908 to create a Department of Public Safety for the Protection of Young Girls and Women with Baldwin as director. Due to the localized nature of law enforcement in the United States, the recognition given to Baldwin, and in 1910 to Alice Stebbins Wells by the Los Angeles Police Department (LAPD), has tended to minimize the women who as early as the 1870s were employed as sheriffs' deputies, assisting their sheriff husbands either officially or unofficially. Women also entered federal law enforcement before they entered local policing; women deputy US marshals have been identified as early as 1890, mostly in the American west (Schulz 2004, 2005). Despite these earlier appointments, the backgrounds, aims, and duties of Baldwin and Wells have come to personify the initial roles of policewomen primarily as social workers assisting women and juveniles who came into contact with police as victims or offenders.

From their first entry into corrections and law enforcement, women's roles evolved from volunteers to paid professionals. They developed a professional ethos closely aligned with social workers rather than male police officers. Their numbers and roles expanded during both World War I and World War II and contracted during periods of economic depression, but with few exceptions, women police officers' responsibilities were closely associated with juveniles, women, and vice enforcement until the 1960s, when they diverged from their specialist roles to undertake the full range of police duties on an equal basis with their male colleagues. This was the result of a number of legal changes precipitated by women's changing roles in the work force and in society generally. Conflict over the changes within the police establishment was evidenced by the large number of court cases and legal consent decrees that enforced women's demands for expanded access to assignments throughout police departments and for promotional opportunities.

Outside municipal policing, at the county level, a small number of women became not the sheriff's wife or widow, but the actual sheriff. Although in each case her status was an outgrowth of her husband's status, a few of the women were active sheriffs and in many ways were not that different from many women in politics who have and continue to literally or figuratively inherit their positions from spouses or fathers. By the 1970s, though, a few women broke direct family links and were elected on their own merits (Schulz 2004).

By the first decade of the twenty-first century, women constituted about 15 % of law enforcement officers nationally (Schulz 2012). Percentages varied considerably by size of department and type of agencies. With only few exceptions, women constituted a larger percentage of officers in large municipal police departments, and a smaller percentage of state police agencies. Within federal law enforcement, the figures varied widely, from female special agents making up about 30 % of Internal Revenue Service (IRS) agents to only single digits in the Drug Enforcement Agency (DEA) (Reaves 2006). Women today hold all ranks, including having served as chiefs of such large city departments as Atlanta, Boston, San Francisco, Milwaukee, and Washington, DC; two have led state police agencies, and a small number have risen to the highest ranks of federal law enforcement.

From Jail and Prison Reformers to "Municipal Housekeeping"

The entry of women into policing first as matrons and then as policewomen can be traced directly to the period that began in the 1820s, when in England and somewhat later in the United States, Quaker and evangelical women entered penal institutions as volunteers to provide religious and secular training for female inmates. Within a short time, other upper-middle-class women joined these volunteers in trying to reform the inmates, primarily through religious instruction, and training them for jobs as domestics in Christian homes (Freedman 1981; Rafter 1992). By focusing on sexual exploitation as a major cause of women's criminality, these volunteers helped to formulate what would become a major focus of early policewomen, specifically working with women and young girls to keep them away from places deemed immoral. This included dance halls, pool rooms, movie theaters, and other public entertainment venues, due in large part to the presence of unattached men. The availability of alcoholic beverages at many of these locations was also seen as contributing to promiscuous behavior. Policewomen also became involved with the mothers of these juveniles, often reflecting class distinctions that resulted in their criticism of the parenting styles of immigrant and poorer women.

Women's expanding roles in society were also fostered by the American Civil War in the 1860s, which led to great activism among women in both the north and the south (O'Neill 1969). With so many men off to war, women formed vast networks of voluntary associations that led them to develop the concept of "municipal housekeeping" – which allowed them to move into the public sphere without violating women's traditional role in the home. Beginning at this time and continuing through the Progressive Era, generally defined as from the 1890s to the post-World War I era in 1920s, municipal housekeeping came to encompass virtually all activities that placed newly established government agencies in contact with women and children.

At a time when fewer social welfare organizations existed, portions of what is now the criminal justice system provided an entry for women into these areas. Concerns for women in jails and prisons gave way to concerns about those in police custody. Police stations often functioned as homeless shelters, and many of those housed there were women and their children. This led female reformers to demand that women be available to assure that women seeking shelter, almost always poor and frequently inebriated, were not vulnerable to the advances of male keepers (Schulz 1995). Behind the considerable efforts of the Women's Christian Temperance Union (WCTU), joined later by the General Federation of Women's Clubs (GFWC), and the Florence Crittenden Circle (which financed homes for unwed mothers), these women were able by the 1880s to convince a number of police departments to hire matrons.

The matrons generally lacked full police authority and were often widows selected by the reformers because the position came with a small set of rooms in the police station. They provided custodial care for incarcerated women and children or those using the police stations as shelters. Within a decade, some matrons had expanded their duties to assist police in questioning women. By the time Lola Baldwin and Alice Stebbins Wells were appointed to duties outside the stationhouse, many matrons were doing similar tasks without official recognition.

Despite difficulties in generalizing, policewomen differed from matrons in a number of important ways: they possessed powers of arrest similar to male police officers, they investigated cases involving women and children rather than merely providing custodial care, and they were generally educated, upper-middle-class women who viewed themselves as superior to their male colleagues, most of whom were immigrants with less education and of lower social class than the policewomen (Schulz 1995). Yet these distinctions could often be blurred; some cities, particularly small ones, often used the titles matron and policewoman interchangeably.

Developing a Professional Ethos

In large part due to their different backgrounds from policemen and their focus on issues surrounding women and juveniles, policewomen developed a professional ethos in advance of policemen and one that placed them within feminist and reformist movements. Until they formed their own organization, the International Association of Policewomen (IAP), in 1915, many participated in the annual meetings of the National Conference of Charities and Corrections (NCCC – later the National Conference of Social Workers [NCSW]). It was at an NCCC conference that Wells announced the formation of the IAP. Only in 1922 did the International Association of Chiefs of Police (IACP), recognize policewomen as essential members of modern departments. The recognition came at the urging of Chief August Vollmer, of Berkeley, California, who served as president that year. Vollmer, a reformer remembered for his advocacy of police training and his attempts to raise the educational level of police officers, was one of few chiefs who supported policewomen, who were most likely to be appointed at the urging of outsiders over the objections of police chiefs.

Policewomen's supporters and advocates were those working in or seeking to expand women's influence in corrections, the juvenile courts, and in probation and parole. In none of these fields did the women seek close association with male colleagues. Women in corrections advocated for women's prisons and jails; in the juvenile courts, they sought female referees (judges) to adjudicate cases involving girls; and in policing, the women sought women's bureaus. This organizational model echoed corrections leaders' success in creating and taking control of women's prisons and juvenile institutions. Here, women were able to achieve top positions as wardens of all-female institutions as early as 1873, when Indiana established the Reformatory Institution of Women and Girls. It was the first independent women's prison, the first to have a female superintendent, and the first to be run by an entirely female staff (Freedman 1981). While early policewomen saw the development of women's bureaus as heralding their success, over time, and as the women's career orientations changed, the bureaus had the affect of isolating them from mainstream policing.

Others who advocated for appointments of first matrons and then policewomen were temperance leaders, social workers, and social hygienists (public health officials and medical practitioners). All were in the mainstream of what historian Anthony M. Platt (1977) termed "child savers," members of the newly developing professions that revolved around protecting children. Although not all child savers were women, Platt described the movement as influenced by the maternal values of middle-class women who extended their homemaker roles of family and childcare into the public sphere. In corrections, this led to women's institutions being set up in a cottage format that attempted to replicate a family structure (Freedman 1981; Rafter 1992).

In policing, even prior to the development of women's bureaus, it led to some women being described as "city mothers" and organized into a City Mother's Bureau. In the LAPD, Lucy Gray Thompson began using that title in the 1880s. The bureau was retained until 1963, when many of its duties were taken over by social welfare agencies (Appier 1998). Its affinity with these agencies can be seen in descriptions provided by Gray's daughter, Aletha Gilbert, who succeeded her mother. By 1922, Gray and her two assistants had established a day nursery for children of working mothers that was staffed by four nurses and also provided counseling to women and children (Gilbert 1922; LAPD 1974).

Similarly, women's bureau directors highlighted their facilities' homelike atmospheres. They described them as places women and girls could come for aid and advice and discuss their problems with women officers. Descriptions included the bureaus' locations away from the male police officers and their having easy chairs, couches, bright-colored curtains, flowers, and even caged canaries at the windows (Appier 1998; Schulz 1995).

At a time when policemen were required to have minimal education requirements beyond the ability to read and write English, the leaders of the policewomen's movement placed great emphasis on hiring women with college or at least high school diplomas and suggested hiring women with prior experience in social service, education, nursing, or fields with public contact. In addition to providing a clear delineation from male officers, these stringent requirements also set the policewomen apart from the less-educated matrons.

The high entry standards were endorsed not only by white women but by the African-American clubwomen who lobbied for appointments of policewomen of their race. Despite more limited opportunities for African-American women, the earliest black policewomen were accomplished members of their communities. Georgia Robinson, the first African-American female member of the LAPD, and possibly the first African-American policewoman, for instance, initially was appointed a matron based on her civic involvement and in 1916 was promoted to policewoman. Although records of the earliest black women appointed to the New York City Police Department (NYPD) are incomplete, at least one was a nurse who had also been a high school teacher. Others were often teachers, social workers, civic leaders, or ministers' wives (Schulz 1995). Cities that hired black policewomen around this time included Chicago, Washington, DC, Indianapolis, and Columbus and Toledo, Ohio; all the women were hired to work with African-American women and juveniles, primarily to patrol dance halls and other areas where guardians of morality were believed to be necessary (Dulaney 1996).

Two visible signs of police authority that the women eschewed were uniforms and firearms. United States policewomen virtually always worked in plainclothes, unlike early British policewomen, whose uniforms were described by Mary Sullivan, the head of the NYPD Women's Bureau, as "unattractive, mannish suits, topped by a helmet that looked like a soup tureen…" (Sullivan 1938, pp. 283–284). In addition to what was likely disdain for uniforms based on their social class and education level, the policewomen and their advocates stated that working in street clothes would allow them to interact with women and girls without the stigma of having been approached in public or visited at their homes by police officers. Most also went unarmed and avoided physical confrontations, although a number of exceptions existed. For instance, in San Francisco, Kathryne Eisenhart, one of the policewomen who came to be known as the “Three Kates” along with Kate O’Connor and Kathlyn Sullivan, was so seriously beaten during a 1917 arrest of two brothers that she retired on disability in 1931 after recurring physical problems. The two policewomen hired in San Diego in 1926 were described as “excellent marksmen,” one of whom shot a would-be assailant who lunged at her during a stakeout in a local park. A feature article about Detroit’s policewomen during World War II mentioned that some carried firearms while others chose not to (Schulz 1995).

By the 1950s, though, most departments had issued uniforms and firearms to their policewomen. The uniforms were primarily for special events, but the women were now expected to carry the firearms, along with handcuffs and their personal items, in handbags made specifically for this purpose. Indicating changes in the women’s view of themselves, Lois Lundell Higgins, president of the International Association of Women Police (IAWP), the successor to the IAP, in 1958, called policewomen the feminine arm of the law, the one that “rocks the cradle, and sometimes shoots a wicked .38” (Schulz 1995, p. 122).

While Higgins and the majority of the policewomen of her generation did not interpret this to mean policewomen should police on an equal basis with men, by the late 1960s some younger women chafed at the limitations to their careers brought about by unequal hiring standards, dissimilar training, fewer assignments open to them, and lack of upward mobility. At the time, women who were completely shut out of state policing or federal law enforcement positions began to question why these careers were closed to them.

The Uniformed Patrol Revolution

Until women began to patrol in uniform in marked police cars, the changes in their roles had been evolutionary. Depending on factors in particular departments, policewomen’s roles had changed from overwhelmingly preventive and investigative to enforcement-oriented. In some cities women worked in uniform (generally dark-colored skirts and white or light blue blouses, black or navy blue pumps, and the-by-now common utility shoulder handbag) at parks, beaches, and other places of public amusements. Many were involved in dangerous undercover assignments. Others worked with male colleagues on sensitive investigations. A woman was needed for the case, whether for her ability to blend in better than a man or men alone could, whether because it was perceived by male supervisors that she would be more able to elicit information from women victims or offenders, or for other reasons. Always, though, she was selected because she was a woman.

But having women patrol in uniform in a marked police car answering general calls for service from the public was not an evolution – it was a revolutionary change. This is belied by the casual way in which it occurred. In September 1968 in Indianapolis, policewomen Betty Blankenship and Elizabeth Coffal reminded their chief, Winston Churchill, that, years earlier, as a police academy instructor, he had commented that if he were ever in charge, he would permit women on uniformed patrol. With one’s day notice, he kept his word, sending the women out together as the history-making Car 47. Dispatchers, at the time mostly male, avoided sending the women on assignments; male officers worried the women would not be able to protect themselves or other officers. But the patrol partners received grudging acceptance after they arrested a man who had beaten a woman to death. Although neither remained on patrol for their entire careers, they opened up to women opportunities that had not previously existed. Within a year, Washington, DC, announced the end of separate entry requirements for male and female officers and within 2 years women also patrolled in Miami, Florida (Schulz 1995).

In conjunction with a number of earlier lawsuits against individual departments opening up civil service promotion ranks to policewoman, a number of laws passed in the 1960s and 1970s provided support for women’s demands to end sex-based titles and assignments and for full participation in policing. These included the 1963 Equal Pay Act, which had little effect until the

1970s when titles of patrolman and policewoman were changed to the unisex police officer. Its effectiveness had been hampered; because women worked under a separate title with different duties, their work had not considered equal. The 1964 Civil Rights Act also had little impact until 1972, when its provisions were extended to government agencies. Funds for police departments, though, played a major role in forcing change. The 1973 Crime Control Act specified that grantees would be ineligible for funds if their employment practices were considered discriminatory. On that basis, police departments – including state police agencies – that had avoid hiring women or equalizing requirements for male and female applicants and placing women on patrol began to comply.

At the federal level, in 1969 President Richard M. Nixon had issued Executive Order 11478, which prohibited discrimination in employment on the basis of race, color, religion, sex, national origin, handicap, or age. Earlier rulings had allowed women to be barred from positions that required carrying a firearm, so only after 1971, when the US Civil Service Commission cancelled the exemption, did women become eligible for federal law enforcement positions. The effect was immediate; that same year, the US Secret Service and the US Postal Inspection Service became the first agencies to hire women special agents, joined the following year by the Federal Bureau of Investigation (Schulz 2005, 2009).

Conclusion and Future Research

What does the future hold for women in law enforcement? The trends appear contradictory. The percentages of women overall in law enforcement agencies have not increased in the last decade. Combining all levels of agencies – local, county, state, and federal – women comprise about 15 % of sworn officers (Schulz 2012) yet the differences in career orientation between the men and the women appear to be shrinking.

Today, after almost a half century of women on patrol, it is difficult to imagine the thousands of studies that were undertaken to determine whether women were capable of patrol, whether they would make the same numbers and types of arrests as men, whether men would work with them, whether citizens would accept them, and, in general, whether women and men policed in similar or in different fashions. The studies, almost all focused on municipal policing, have resulted in contradictory findings, but generally found few substantive differences between female and male officers and sometimes have resulted in contradictory conclusions. But because the studies generally looked at only one department, or were based only on the patrol officers, or surveyed a self-selected group of women officers at conferences sponsored by women's law enforcement groups, it is difficult to generalize their findings beyond that women are competent patrol officers and that their male colleagues have come to accept their presence albeit sometimes still grudgingly.

In a meta-analysis of past research, Carol A. Archbold and Dorothy Moses Schulz (2012) found that recent studies have found more similarities than differences between men and women, while earlier studies seemed to find more differences. They posit that this change may result from presumptions by the researchers, attitude changes among women – and possibly men – entering police work, or police agencies' greater stress on community relations and involvement, the latter areas in which some earlier studies found women excelled. The findings may also result from these studies focusing on less gender-specific issues and more on factors that determine the reason for becoming a police officer, the decision to arrest or use force, and factors that influence the decisions by officers to compete for promotion.

Thus, there appears to be recognition among researchers that numbers and percentages do not tell the whole story. Questions that might be considered are whether women are accepted into law enforcement careers in the same percentages they apply, whether they complete training in percentages equal to their acceptance, whether there are organizational or societal barriers that contribute to the apparent stagnation in women's entry, and whether retention and promotion rates are similar for male and female officers.

The fragmented nature of United States policing makes progress more difficult to measure than in other countries, where there are – if not only one large national force – far fewer law enforcement agencies. Fewer agencies make it easier for researchers and demographers to count numbers and chart progress. Another issue is whether progress should be measured only in terms of upward mobility, particularly in a field where narrow organizational pyramids guarantee that few police officers will ever become chiefs of police and that, in fact, few rise to even such second-level supervisory ranks as corporal or sergeant.

Issues in Upward Mobility

An issue that has recently drawn considerable attention is upward mobility. Prior to the 1980s, few women held ranks higher than policewoman and the few who did were almost overwhelmingly in charge only of other women. The few women who were elected or appointed sheriffs were stand-ins for their husbands – they were appointed when their husbands died, or – particularly in Wisconsin – were elected in place of their husbands as way to avoid existing term limits laws (Schulz and Houghton 2003). This changed in 1985, with the selection of Penny Harrington as chief of the Portland (Oregon) Police Bureau (Schulz 2004). Although tenure for police chiefs is often brief, Harrington's was only 17 months, during which, she noted "controversy and confrontation were my constant companions" (Harrington 1999, p. 214). But they were her companions well before she was named chief; during her career she had filed more than 40 lawsuits against the department (Harrington 1999; Schulz 2004).

Other chiefs have been less controversial and most are likely to avoid discussions of their being a female chief rather than *the* chief. Yet female chiefs are still rare; Schulz (2004) estimated there were about 200 women chiefs, about 1 % of the approximately 18,000 police chiefs across the United States. Although high turnover and the large number of police agencies makes an accurate count virtually impossible, there are indications, primarily from newspaper articles documenting selection of a woman chief, that the number has doubled to about 2 % of women law enforcement chief executives (used to describe police chiefs or commissioners, sheriffs, and women leading federal law enforcement agencies).

Recent research into women in traditionally male fields has raised questions as to whether numbers and percentages provide a complete picture. Larger societal issues may play a role that has often been neglected. Studies of men in female-dominated professions such as librarianship and nursing, for instance, have found that men were advantaged in upward mobility, attributed primarily to having higher status in society based on their sex (Ott 1989; Yoder 1991). A number of studies, though, raise a question of whether upward mobility is the correct way to measure success, since many in law enforcement do not seek promotion at all (Whetstone and Deborah 1999; Whetstone 2001). Researchers who had anticipated that women would serve as change agents of the culture of law enforcement and corrections agencies have also indicated they were mistaken. They found, rather, that possibly because they believed themselves lacking in institutional support, women were traditional in their leadership styles and less willing to take chances (Zimmer 1986; Engel 2000).

While these studies appear to close off many areas of future research, they raise the possibilities that researchers need to look more carefully at cultural clues in particular agencies or types of agencies. Are there law enforcement agencies with styles that foster a desire to move up in rank or, contrarily, are there agencies in which the specialist units are so highly-regarded that officers are content to remain in them as long as possible? Is one culture better than the other; and whose definition of better would prevail? Is too much research emphasis placed solely on police officers – whether male or female – rather than on the higher ranking personnel? Similarly, can studies in local police departments be generalized to sheriffs' offices or to state police or federal law enforcement agencies? Lastly, might researchers focusing on women's place in law enforcement consider comparisons with women in other fields to better gauge whether acceptance, upward mobility, or the ability to fit into or alter an agency's culture can be measured across professions?

The importance of police in society and the unique responsibility they are given to curtail people's liberty and to exercise deadly force means that they will always be important centers of research. Who selects a career in policing, why they do so, and how they use the powers inherent in the position, and whether any or all of these are different for men or for women, will undoubtedly continue to be fruitful areas of study for future generations.

Related Entries

- ▶ History of Corrections
- ▶ History of Juvenile Justice
- ▶ Minorities Within the Police Workforce
- ▶ Police Culture
- ▶ Policing Special Populations
- ▶ Women in Policing

Recommended Reading and References

Alderman BJ (2007) The secret life of the lawman's wife. Praeger, Westport

Appier J (1998) Policing women: the sexual politics of law enforcement and the LAPD. Temple University Press, Philadelphia

Archbold CA, Schulz DM (2008) Making rank: the lingering effects of tokenism on female police officers' promotion aspirations. Police Q 11:50–73

Archbold CA, Schulz DM (2012) Research on women in policing: a look at the past, present, and future. Sociol Compass 6/9:694–706.

Archbold CA, Hassell KD, Stichman AJ (2012) Comparing promotion aspirations of female and male police officers. Int J Police Sci Manag 12:287–303

Dulaney WM (1996) Black police in America. Indiana University Press, Bloomington

Engel RS (2000) The effects of supervisory styles on patrol officer behavior. Police Q 3:262–293

Freedman EB (1981) Their sisters' keepers: women prison reform in America, 1830–1930. University of Michigan Press, Ann Arbor

Gilbert A (1922) The duties of a 'city mother'. Am City 26:239–240

Harrington P (1999) Triumph of spirit. Brittany Publications, Chicago

Hassell KD, Archbold CA, Schulz DM (2011) Women and policing in America: classic and contemporary readings. Wolters Kluwer, New York

Los Angeles Police Department (1924) Law enforcement in Los Angeles. Arno Press, New York, reprint 1974

O'Neill WL (1969) Everyone was brave: the rise and fall of feminism in America. Quadrangle Press, Chicago

Ott EM (1989) Effects of male-female ratio at work: policewomen and male nurses. Psychol Women Q 13:41–57

Platt AM (1977) The child savers: the invention of delinquency, 2nd edn. University of Chicago Press, Chicago

Rafter NH (1992) Partial justice: women, prisons, and social control, 2nd edn. Transaction Publishers, New Brunswick

Reaves BA (2006) Federal Law enforcement officers, 2004. U.S. Department of Justice, Washington, DC

Schulz DM (1995) From social worker to crime fighter: women in United States municipal policing. Praeger, Westport

Schulz DM (2004) Breaking the brass ceiling: women police chiefs and their paths to the top. Praeger, Westport

Schulz DM (2005) Women in federal agency law enforcement. In: Sullivan LE, Schulz DM (eds) Encyclopedia of law enforcement. Sage Reference, Thousand Oaks, pp 908–911, vol. 2

Schulz DM (2009) Women special agents in charge: the first generation. Polic: Int J Police Strateg Manag 32:675–693

Schulz DM (2012) Making it happen: strategies for personal success. Australas Polic 4:23–27

Schulz DM, Houghton S (2003) Married to the job: women sheriffs in Wisconsin. Wisconsin Mag Hist 86:22–38

Shpritzer F (1959) A case for the promotion of policewomen in the city of New York. J Crim Law Criminol Police Sci 50:405–419

Sullivan M (1938) My double life: the story of a New York policewoman. Farrar & Rinehart, New York

Whetstone TS (2001) Copping out: why police officers decline to participate in the Sergeant's promotion process. Am J Crim Justice 25:147–159

Whetstone TS, Wilson DG (1999) Dilemmas confronting female police officer promotional candidates: glass ceiling, disenfranchisement or satisfaction? Int J Police Sci Manag 2:128–143

Yoder JD (1991) Rethinking tokenism: looking beyond numbers. Gend Soc 5:178–192

Zimmer L (1986) Women guarding men. University of Chicago Press, Chicago

Workplace Safety Crimes

▶ Criminalization and Occupational Health and Safety

Worry About Crime

► Fear of Crime and the Psychology of Risk

Wrongful Convictions

► Forensic Science and Miscarriages of Justice

Wrongful Convictions in the Inquisitorial System

Gwladys Gilliéron
Distance Learning University Switzerland, Brig, Switzerland

Overview

The purpose of any criminal justice system is to deliver justice to all by convicting the guilty and protecting the innocent. Two model systems exist, which provide different rules to deal with the offender and bring him to justice. They are commonly known as the inquisitorial criminal justice system and the adversarial criminal justice system. The adversarial system is the system of law generally adopted in common law countries, whereas the inquisitorial justice system is usually found on the continent of Europe among civil law systems. They both aim to ensure that procedural fairness exists across the criminal justice system. Nowadays, however, it is commonly assumed that no criminal justice system belongs entirely to one of these two models.

Within the literature on wrongful convictions, most of the attention has been devoted to wrongful convictions in which innocent people have been convicted after trial. However, the reality today is that the normal trial proceeding is the exception and that the great majority of cases are resolved by way of alternative or summary proceedings, in which the public prosecutor plays a crucial role. It is certainly true that alternative proceedings are a simple method to relieve courts' heavy caseload. However, it must not be forgotten that simplification of proceeding usually goes along with restrictions on a criminal defendant's rights and also may reinforce the risk of wrongful convictions. Hence, the problem of wrongful convictions in these kinds of proceedings should deserve much more attention than what has been the case until now.

Taking the Swiss criminal justice as example, which is based on a modified inquisitorial tradition, this contribution highlights the risk of wrongful convictions especially in alternative proceedings. Conclusions drawn from the Swiss legal system may also be valid in all other criminal justice systems where alternative proceedings are the rule rather than the exception. Future researches should concentrate their efforts on wrongful convictions in alternative and summary proceedings. This is important in order to improve the quality of judicial systems.

Models of Criminal Justice Systems

The goal of every criminal justice system is to ensure that those guilty of committing a criminal offense are convicted and that the innocents are acquitted. In achieving this goal, the different criminal justice systems provide for different safeguards. Traditionally, criminal justice systems are either inquisitorial or adversarial. The inquisitorial system of justice is employed in many continental European jurisdictions, whereas criminal proceedings in the United States and in other common law countries such as England and Wales operate under an adversarial justice system. However, today it can be assumed that no criminal justice system is entirely either adversarial or inquisitorial. Continental jurisdictions have adopted adversarial elements, and inversely common law systems take over legal rules from inquisitorial justice systems. The Swiss criminal justice system, based on the inquisitorial tradition, took a step in this direction by introducing an abridged proceeding (*abgekürztes Verfahren*) that is similar to the American plea bargaining.

The distinction between both models of criminal justice systems has therefore become less important. In fact, in overloaded criminal justice systems of both models, criminal trials become relatively infrequent, and the determination of guilt is mainly made by the prosecutor. In the American criminal justice system, the overwhelming majority of criminal charges are resolved by guilty pleas, and virtually all of those are the result of plea bargaining. In 2010, about 97.5 % of cases in the federal system were settled by guilty pleas or nolo contendere (Maguire 2010). In the Swiss legal system, approximately 90 % of the convictions are based upon a penal order, a decision issued by a public prosecutor (Hutzler 2010). Hence, in reality, Switzerland and the United States mostly have an administrative criminal justice system today where the prosecutor combines the executive and judicial power. The public prosecutor is a public official with decision-making powers that are to a large extent adjudicatory. The problem is that the prosecutor faces extremely limited oversight of his decisions, and thus, wrongful convictions are a risk to seriously take into consideration.

Main Features of the Swiss Criminal Procedure

On January 1, 2011, the first Swiss Code of Criminal Procedure (CCrP) came into force and replaced the 26 cantonal criminal procedure codes and the Federal Act on the Administration of Federal Criminal Justice. This is the most significant legal reform that Switzerland has seen in this area of the law. The elimination of legal fragmentation has the aim to ensure increased equality before the law and greater legal certainty (Bundesrat 2005).

For efficiency reasons, the examining magistrate, previously found in some cantons following the French tradition, has been abolished, leaving only the public prosecutor. Thus, the public prosecutor occupies a central position and his powers are broad. He conducts the preliminary proceedings, pursues criminal offenses within the scope of the investigation, brings charges, and pleads in favor of the criminal charges (Article 16 para. 2 CCrP). In addition, in order to deal with an increasing caseload, prosecutors have been given more power and discretion to divert cases. The enormous power vested in the prosecution is compensated by the judge being responsible for compulsory acts (e.g., search and seizure) and extended defense powers (such as the right to be "immediately" assisted by a lawyer). The advantage of such an inquiry model is the achievement of a high grade of efficiency of prosecution by realizing homogenous investigation, examination, and charging (Bundesrat 2005). The abolition of the examining magistrate also brings the Swiss legal system closer to those used in common law systems.

The Swiss procedure is guided by the principle of the *factual truth* (*Prinzip der materiellen Wahrheit*). The goal of the prosecution is not to seek a conviction but to discover the truth and to apply the law. Hence, the public prosecutor has the duty to investigate in an objective and neutral way and must therefore take into account the incriminating and the exculpatory circumstances with equal care (Article 6 para. 2 CCrP). If the public prosecutor is convinced that a decision needs to be reviewed for factual or legal reasons, he is entitled to appeal. He may do this to the disadvantage as well as to the advantage of the condemned (Article 381 CCrP).

The prosecution is obliged to fully disclose its files to the defense, and a restriction to the entire disclosure of the files is only possible under certain conditions defined by law, namely, if there is reasonable suspicion that a party is misusing its rights, to ensure the safety of people or to guarantee public or private confidentiality interests (Article 108 CCrP). This obligation to fully disclose is based on the right to be heard, which is one of the basic fundamental legal rights in Switzerland (Article 29 para. 2 of the Swiss Federal Constitution; Article 107 CCrP) and the continental law in general. This rule may be of particular importance in the context of the abridged proceedings, a procedure that is comparable with American plea bargaining and is described later in this contribution. A person accused of a crime will be aware of the strengths and weaknesses of the case. If the accused then decides to confess to a criminal offense, he will act in full knowledge of the prosecutor's file.

In contrast to criminal proceedings operating under an adversarial justice system, in which the parties are each responsible to investigate the case and to present their evidence before a passive and neutral judge or jury that will decide on guilt, defense lawyers in an inquisitorial system usually do not conduct their own investigation and play only a limited role in establishing the relevant facts. It is the court that is required to actively investigate the case and is ultimately responsible for discovering the truth. It is based on the "free evaluation" (*freie Beweiswürdigung*) of all available evidence that the court reaches the decision about the innocence or guilt of the accused. The judge is required to be intimately convinced regarding the truth of the facts unless he admits them as being proven (Hauser et al. 2005; Trechsel and Killias 2004). The court must cite its rationale for the verdict and the sentence. The aim of this last duty is the protection of citizens against arbitrary state decisions.

In contrast to the adversarial system, a defendant's confession is considered similarly to other types of evidence, and the prosecution is still required to present a full and compelling case (Hauser et al. 2005). A confession is a mitigating factor of limited impact on the sentence. In the Swiss criminal justice system, a confession qualifies the defendant for a sentence reduction of about 10 % (Killias 2008).

The right to remain silent or the right against self-incrimination is also fundamental in an inquisitorial criminal justice system (Article 113 CCrP; Article 32 para. 2 the Swiss Federal Constitution; Article 6 subparagraph 2 European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR), and Article 14 para. 2 International Covenant on Civil and Political Rights (ICCPR)). The accused has the right to refuse any cooperation in the criminal proceedings, and no disadvantageous conclusions can be drawn from silence.

From the brief description of the Swiss criminal procedure, it is clear that it provides for different safeguards (i.e., right to be heard, right to remain silent, principle of the factual truth that oblige the prosecutor to investigate objectively) to ensure that those guilty of committing a criminal offense are convicted and that the innocents are acquitted. A trial serves as check on governmental excess and intends to assure the accuracy of the verdict. Alternative or summary proceedings are not invested with the same degree of safeguards that are present in the normal procedure and thus might present a higher risk of wrongful conviction.

The Swiss criminal justice system has two alternative proceedings, namely, the abridged and the penal order proceedings. These will be described in the following section.

Alternative Proceedings

Penal Order Proceedings

In Switzerland, the majority of convictions does not result from a court trial but is based upon a penal order. Overall, about 90 % of all cases are settled out of court (Hutzler 2010). In some cantons, closer to 97 % of the cases are dealt with by penal order (e.g., Basel in 2010, http://www.statistik-bs.ch/tabellen/t19/2). This procedure is often used in cases of traffic offenses, minor thefts, and possession of drugs.

The prosecutor issues a penal order (also known as summary punishment order or *Strafbefehl*) if the accused person has, in the preliminary proceedings, accepted responsibility for the factual circumstances of the case *or* if the circumstances have been otherwise sufficiently resolved. Hence, in the second scenario, a hearing of the accused prior to the decision is not a condition. In the penal order the circumstances of the cases are described and a sentence is imposed. This decision may be appealed by the defendant and other interested persons by submitting a written petition to the prosecutor within 10 days (Article 354 para. 1 CCrP). Grounds must be given for the appeal, except for an appeal by the defendant. Consequently, the case is tried in court (Article 356 CCrP). If no objection is made, the penal order becomes final and has the same effect as a judgment following a trial (Article 354 para. 3 CCrP). As a consequence, this written procedure may result in a judgment where the prosecutor decides mainly on the basis of the police files and without the parties being heard. However, since the defendant has the possibility to raise objection and ask for a full trial, this procedure is not seen as incompatible with the constitutional right to be heard (Hauser et al. 2005; Piquerez 2006).

The prosecutor has no discretion in deciding whether he wants to use the penal order or the ordinary proceedings. Once the legal requirements for the use of the penal order proceedings are fulfilled, the prosecutor must choose this way to deal with the criminal case.

The scope of application of the penal order proceedings, at the origin designed for petty offenses punishable with a fine, has widely been expanded over time. Today, the prosecutor has the possibility to impose a prison sentence of up to 6 months (Art. 352 para. 1 CCrP). This rule is quite problematic since a prison sentence is a sanction serious enough that it should not be imposed solely on the judgment of the prosecutor without compulsory preliminary hearing of the accused and without any judicial control.

The penal order is also used in many other European legal systems and is commonly referred to as the continental form of plea bargain (Langbein 1974; Trechsel and Killias 2004). However, various aspects differ between both proceedings. In the case of a penal order, the evaluation of the case is exclusively made by the prosecutor, and he is the one who imposes the sentence. During this process, the defendant, in contrast to American plea bargaining, is not represented by a lawyer. Since the prosecutor does not have the duty to hear him prior to the decision, the defendant usually does not participate. The defendant may only accept or refuse the order. A bargain between prosecution and defendant does not take place. In contrast to American plea bargaining, a defendant who does not accept the penal order and insists on a full trial does not take a risk of having a harsher sentence imposed by the court (Killias 2008; Trechsel and Killias 2008). This is mainly attributed to the fact that a penal order can only be issued if the facts are sufficiently clear and the culpability is not dubious, so that a reduction of the charges is unlikely to occur. As a consequence, the risk of false confession (i.e., accepting the penal order when the defendant is not guilty) does not exist to the same extent in the continental law as in the American system.

Abridged Proceedings

The CCrP has introduced the possibility of ending a case by way of abridged proceedings (*abgekürztes Verfahren*). This procedure is quite similar to the bargaining under the American system and thus brings the Swiss legal system closer to the common law systems. Prior to the introduction of the CCrP, three out of 26 cantons already had such an alternative procedure.

The accused person may submit an application to the prosecution for the case to be conducted by way of abridged proceedings provided that he accepts liability for those circumstances essential to the legal evaluation of the case (although an outright confession is not required) and that he accepts at least in principle the civil claims (Article 358 para. 1 CCrP). An abridged proceeding is only possible if the prosecution requests the imposition of a prison sentence not exceeding 5 years (Article 358 para. 2 CCrP).

If the application to the prosecution to handle the case by way of abridged proceedings is accepted, the defendant must have a lawyer to represent him (Article 130 (e) CCrP). With this rule, the protection of the accused person during the informal negations with the prosecution should be secured.

Informal negotiations are closed by an indictment the prosecution conveys to the parties. They have 10 days to accept or reject the indictment (Article 360 para. 2 CCrP). In addition to the points usually found in an indictment, this document contains the sentence and the warning to the parties that by accepting the indictment, they waive the right to ordinary proceedings and to initiate legal remedies (Article 360 para. 1 CCrP). The consequence of this provision is that the convicted may not file a petition for revision. This is especially the case for petition for revision based on new evidence, since the court does not conduct an evidentiary hearing; this is in contrast to the ordinary proceeding. However, the use of this legal remedy should still be possible to eliminate decisions that are unacceptable from a constitutional point of view (e.g., exercise of coercive power by the prosecution on the suspect, abuse of public office, or the decision is irreconcilably in contradiction with a later criminal decision which involves the same factual circumstances). If the parties reject the indictment, the prosecution will conduct ordinary proceedings (Article 360 para. 5 CCrP). If the indictment is accepted, the prosecution transmits the indictment together with the files to the Court of First Instance (Article 360 para. 4 CCrP). During the principal hearing, the court will have to establish whether the accused person accepts the circumstances of the case on which the charge is based, and whether this assertion corresponds to the position as set out in the files (Article 361 CCrP).

Following the principal hearing, the court retires and conducts its deliberation in private. The court has to determine whether the carrying out of abridged proceedings is lawful and appropriate, whether the charge corresponds to the conclusions of the principal hearing and to the files, and whether the sanctions requested are reasonable (Article 362 para. 1 CCrP). If the court comes to the conclusion that the requirements for a judgment by way of abridged proceedings are met, it converts the criminal offenses, sentence, and civil claim of the indictment into a judgment (Article 362 para. 2 CCrP). On the other hand, if the court estimates that the requirements are not met, it sends the files back to the prosecution in order to proceed by way of ordinary proceedings (Article 362 para. 3 CCrP). Declarations like confessions provided by the parties in respect of the abridged proceedings cannot be used in ordinary proceedings (Article 362 para. 4 CCrP). A party may only appeal against a judgment in abridged proceedings on the basis that it did not accept the indictment or that the judgment does not correspond to the indictment (Article 362 para. 5 CCrP).

Since the abridged proceeding was introduced only recently on a nationwide basis, it is too early to draw some clear conclusions on the dangers of the proceedings for innocent defendants. The canton of Basel-Landschaft was the first canton to introduce an abridged proceeding in 2000 so that some statements can be made based on its experience. Until now, this alternative procedure has been relatively rarely used. Between 2000 and 2008 a total of 96 persons have been convicted based on this proceeding. Hence, on average 12 accused are condemned by way of abridged proceedings per year. It seems that sentences are not less severe in this kind of proceedings as compared to similar cases judged by way of ordinary proceedings. It is true that the danger of false confession exists, since some defense lawyer might suggests their clients use the abridged proceedings, although no confession of the accused offense occurred during the preliminary proceedings. However, in such a situation it is not unusual that the court rejects to handle the case by way of abridged proceedings. Therefore, to a certain extent, the court can be considered as a safeguard against false confessions (Gilliéron 2010).

The risk with the Swiss abridged proceedings is that over time it comes closer to American plea bargaining and takes over not only its advantages but also its disadvantages. In the US criminal justice system, the court only rarely refuses to reject a guilty plea, so that it may be assumed that the court is not really an efficient safeguard against wrongful confessions. The reason for accepting guilty plea proposals is that the prosecutor knows all details about the case, whereas the judge has less background information on the alleged crime and the defendant. Moreover, if judges would start to call the prosecutors' decisions into question, the caseload would become overwhelming (Wright 2009). In fact, the judicial inquiry is limited to ascertaining that the defendant is of sound mind and understands the consequences of his actions, rather than examining the accuracy of the facts to which he is attesting. The judge is only required to assure that the conduct to which the defendant confesses constitutes in fact an offense under the statutory provision under which he is pleading guilty. In reality, the assessment of the defendant's responsibility is made within the executive branch, in the office of the prosecutor, and does not occur in court at all (Lynch 1998).

American plea bargaining may be defined as an "informal, administrative, inquisitorial process of adjudication, internal to the prosecutor's office – in absolute distinction from a model of adversarial determination of fact and law before a neutral judicial decision maker" (Lynch 2003, p. 1404). The risk inherent in plea bargain is best described by Barkow. According to her, plea bargaining "causes a systematic imbalance of power by allowing prosecutors to bypass the check of the judicial process" (Barkow 2006, p. 1050). In fact, the prosecutor faces extremely limited oversight over his decisions. Hence, abuse of discretion and arbitrariness might be a risk. This in turn favors the risk of wrongful convictions.

Is the Public Prosecutor the World's Most Objective Authority?

The Swiss legal system, by abolishing the examining magistrate, has adopted a system where the public prosecutor is the key actor in the criminal justice system. The public prosecutor has extremely wide powers since he is the one who conducts the preliminary proceedings, pursues criminal offenses, brings charges, and pleads in favor of the criminal charges (Article 16 para. 2 CCrP). Moreover, he makes all important decisions in alternative and summary criminal proceedings. The legitimate question that arises in this context is whether the public prosecutor is the world's most objective authority. On the one hand, the prosecutor is obliged to investigate in an objective and neutral way since he has a duty to truth and justice (Article 6 para. 2 CCrP). However, on the other hand, the prosecution represents the state at the trial. As a matter of fact, this last task excludes strict neutrality. It can be assumed that at trial, the prosecutor's aim is to seek a conviction. Furthermore, the great majority of cases are resolved by alternative proceedings such as the penal order proceeding, in which the public prosecutor takes the important decisions. In this alternative procedure, the prosecutor, in deciding whether to punish the defendant, is in reality an inquisitor seeking the "right" outcome. Taking all this into consideration, the public prosecutor's objectivity within criminal proceedings is questionable and may increase the risk of wrongful convictions.

However, the way the public prosecutor is controlled may limit the risk of abuse and hence wrongful convictions. To determine whether this is actually true, it is necessary to take a look at the structure and the manner of organization of the public prosecutor's office. Although Switzerland now has a unified criminal procedure, the organization of the public prosecution service – and in a broader sense the criminal justice authorities – remains a matter for the cantons (Art. 14 para. 2 CCrP). Thus, it is highly decentralized.

In general, prosecution services are organized hierarchically (Arn et al. 2011). Hence, prosecutors have to follow directives and instructions received from their superiors. In the majority of cantons, the Minister of Justice and hence the cantonal government stand at the top of the hierarchy. In some other cantons, the public prosecutor's office is part of the judiciary and under supervision of the cantonal Supreme Court. In those cantons where the public prosecutor is subordinated to the cantonal government, the latter only rarely issues instructions. Thus, the public prosecutor's office is autonomous and independent in a factual way regarding the functional scope, i.e., when fulfilling the tasks and in the decision practice (Hauser et al. 2005). At most, the cantonal government will issue general recommendations in order to ensure that certain aims of crime policy are pursued. In the other cantons, where the public prosecutor is as independent as the judiciary, the cantonal Supreme Court is normally not allowed to give any instructions. Its supervision is limited to receive and control the annual report (Cornu 2000). Hence, it is clear that prosecutors, in fulfilling their tasks, are to a large extent independent and are not really held accountable. It should be noted in this context that it is very important that prosecutors provide neutral, nonpolitical decisions in individual criminal cases and that prosecutorial independence is an essential element that contributes to achieving this goal. For this reason it is essential that the role of the judge does not continue to lose its significance since prosecutorial decisions can only be controlled effectively by the judiciary.

Wrongful Convictions in Switzerland

Inspired by research on wrongful convictions in Germany conducted by Karl Peters (1970–1974), a project supported by the Swiss National Science Foundation (SNSF) has analyzed all mistaken convictions (successful petitions of revision) in Switzerland between 1995 and 2004 (Killias et al. 2007). Over the considered time period, a total of 236 petitions for retrial have been admitted. The majority (159 successful petitions of revision) concerned penal orders. When considering the number of cases that are dealt with in this kind of summary proceedings, this outcome is not out of proportion. But it is highly probable that in this field, there are much more wrongful convictions than those discovered by the research. It is highly possible that the majority of convicted waive their right to challenge the decision and prefer to pay a fine (Gilliéron 2010; Killias 2008).

The results of the study show that wrongful convictions in which innocent people have been convicted after a trial play a minor role. Most cases where there was a successful petition of revision involved an excessive sentence imposed by the court when some mental problem of the convicted had not been recognized by the court and hence not been taken into account. In fact, the ignorance by the court of some mental problems of the convicted was a factor in 46.4 % of admitted petitions for retrial based on new evidence. In about one third of the admitted petition for revision, the court had convicted a factually innocent person, mostly due to perjury by “victims” of crimes against sexual integrity, or in other cases because of eyewitness misidentification or false confession by the defendant that he later repealed. The study did not identify exonerations due to DNA evidence (Killias et al. 2007). This might be the consequence of not storing items of physical evidence over a long period.

In contrast, wrongful convictions by penal order mainly concerned factually innocent defendants, who for the most part had been found guilty of a traffic violation and sentenced to a fine. Based on the available opinions, the granting of the motion for retrial led in 21 cases to a reduced sentence, in one case to a harsher sentence, and in 109 cases to the acquittal of the convicted. Wrongful identification and insufficient investigation of the facts were the leading factors contributing to wrongful convictions (Killias et al. 2007).

Risk of Wrongful Convictions Inherent in Penal Order Proceedings

There are different factors specific to the penal order proceedings that increase the risk of wrongful conviction (for a detailed discussion of these risks, see Gilliéron 2010). These are briefly described below.

Investigation: In the great majority of cases, the prosecutor bases his decision on the sole basis of the police accounts. However, those may be inaccurate or incomplete. The prosecutor is not required to hear the defendant, even if he has the intention to impose a prison sentence. In addition, it is also possible that the prosecutor expects that the defendant will make opposition if he is innocent. In sum, the prosecutor’s investigation is often not conducted with enough care.

Prosecution: The prosecution alone is responsible to issue the penal order. The decision is only subject to judicial control if the defendant or other interested persons make opposition. The fact that in this summary procedure the prosecutor acts as an inquisitor may contribute to increase the risk of wrongful convictions.

Form and time limit to make opposition: Defendants have a legal right to make written opposition within 10 days if they do not agree with the prosecutor’s decision. The short time span and the written form might hinder the defendant to exercise this right.

Defendant's behavior: Various reasons can explain why defendants miss the deadline to make opposition or decline to exercise this right. It is possible that due to functional illiteracy, the defendant is unable to understand the instructions about the right to make opposition. This assumption has some support in view of the fact that about 16 % of the Swiss population is unable to understand a text of some complexity (Notter et al. 2006). Other reasons for not contesting the order include indifference, ignorance of the law, and fear of unfavorable outcome, such as costs of the procedure.

Conclusions and Future Research

Many studies have focused their attention on wrongful convictions in which innocent people have been convicted after trial. In this context, the question whether an inquisitorial criminal justice system or an adversarial criminal justice system is better able to protect the innocents has often been discussed (see, e.g., Thomas 2008; Walpin 2003). Both models of criminal justice systems provide for different safeguards to ensure that those guilty of committing a criminal offense are convicted and that the innocents are acquitted. In the Swiss criminal justice system, based on a modified inquisitorial tradition, the right to be heard and in particular the full disclosure of the prosecutor's files and the obligation of the courts to cite their rationale for the verdict and the sentence may prevent the conviction of an innocent person. Furthermore, the principle of the factual truth that obliges the prosecutor to investigate in a neutral way may contribute to avoid and correct wrongful convictions.

However, nowadays, no criminal justice systems can claim to entirely belong to one of both models. Criminal justice systems that operate under an inquisitorial justice system have adopted adversarial elements and conversely. As a consequence, this distinction tends to lose its importance. This is reinforced by the fact that today, criminal trials are the exception rather than the rule. Today, the overwhelming majority of criminal matters are resolved by way of alternative or summary proceedings. In legal systems in which criminal trials become relatively infrequent, the determination of guilt is mainly made by the prosecutor. The prosecutor, by making important decisions in alternative proceedings, combines the executive and judicial power. His decisions are to a great extent adjudicatory. The problem in alternative proceedings is that the prosecutor faces extremely limited oversight of his decisions. Although a simplified procedure, such as the abridged proceedings under the Swiss legal system and American plea bargaining, still requires a decision by the court, the court hearing in this kind of procedure provides restriction on prosecutorial power to a much lesser degree. Therefore, wrongful convictions are a risk to seriously take into account.

In this contribution, the risk of wrongful convictions in the Swiss penal order proceedings, in which it is usual that the prosecutor takes his decision on the only base of the police report, has in particular been highlighted. At the origin, this summary procedure was designed for petty offenses punishable with a fine. Its field of application has widely expanded over time. Now, the prosecutor has the possibility to impose a custodial sentence of up to 6 months. Therefore, the penal order proceeding is not limited to petty offenses any more, but extends into criminal acts of some gravity. If the defendant does not make opposition to the decision, the decision is issued without judicial control. The wide scope of application of this procedure without any judicial control is more than critical, since this kind of proceedings is inclined to produce wrongful convictions and since almost all defendants are convicted in this way. Possible solutions that would help to prevent wrongful convictions would consist in having a judge being more actively involved (i.e., checking the penal order), a compulsory hearing of the defendant, and being represented by a lawyer if a custodial sentence is pronounced.

Concerning the abridged proceeding, it is too early to make some statements with regard to the risk of wrongful convictions. In a few years, more information will be available regarding its application frequency, the kind of crimes resolved in this way, and the problems related to it.

Alternative and summary proceedings certainly make sense if the aim of the criminal justice system is crime control or bureaucratic efficiency. However, if the aim of the justice system is the protection of individual rights, alternative proceedings are harder to justify.

It is clear that given the large amount of criminal cases that enter the criminal justice systems, it is not possible to give every defendant a criminal trial. Methods must be found to reserve full trials to those cases that deserve to go that route and treat the vast majority of other cases in another way. Future research on wrongful convictions should pay much more attention to alternative proceedings. A better understanding of the risks of wrongful convictions in these types of proceedings helps criminal justice systems to take measures in order to improve the quality of their legal system. The defendant should in turn benefit from this, since his rights would be better secured.

Related Entries

- ▶ Causes of Wrongful Convictions
- ▶ False Confessions and Police Interrogation
- ▶ Legal Rules, Forensic Science and Wrongful Convictions
- ▶ Prosecution and Wrongful Convictions

Recommended Reading and References

Arn R, Saurer N, Kuhn A (eds) (2011) Organisation der kantonalen und eidgenössischen Strafbehörden. Organisation des autorités pénales cantonales et fédérales. Organizzazione delle autorità penali cantonali e federali. Helbing & Lichtenhahn, Basel

Barkow RE (2006) Separation of powers and the criminal law. Stanford Law Rev 58:989–1054

Bundesrat (2005) Botschaft zur Vereinheitlichung des Strafprozessrechts. Bern

Cornu P (2000) The Swiss public prosecutor's office: its role in criminal procedure, its relations with political authorities and the police, its role in crime policy. In: Council of Europe (ed) What public prosecution in Europe in the 21st century. Proceedings, Strasbourg, 22–24 May 2000. Council of Europe Publishing, Strasbourg

Gilliéron G (2010) Strafbefehlsverfahren und plea bargaining als Quelle von Fehlurteilen. Schulthess, Zurich

Hauser R, Schweri E, Hartmann K (2005) Schweizerisches Strafrecht, 6th edn. Helbing & Lichtenhahn, Basel

Hutzler D (2010) Ausgleich struktureller Garantiedefizite im Strafbefehlsverfahren. Eine analyse der Zürcherischen, Schweizerischen und Deutschen Regelungen, unter besonderer Berücksichtigung der geständnisfunktion. Nomos, Baden-Baden

Killias M (2008) Wrongful conviction in Switzerland: the experience of a continental law country. In: Huff CR, Killias M (eds) Wrongful conviction. International perspectives on miscarriages of justice. Temple University Press, Philadelphia

Killias M, Gwladys G, Nathalie D (2007) Erreurs judiciaires en Suisse de 1995–2004. Report to the Swiss National Science Foundation. University of Lausanne and University of Zurich, Lausanne

Langbein JH (1974) Controlling prosecutorial discretion in Germany. Univ Chicago Law Rev 41:439–457

Lynch GE (1998) Our administrative system of criminal justice. Fordham Law Rev 66:2117–2151

Lynch GE (2003) Screening versus plea bargaining: exactly what are we trading of? Stanford Law Rev 55:1399–1408

Maguire K (ed) (2010) Sourcebook of criminal justice statistics [Online]. Bureau of Justice Statistics, (Table 5.22). http://www.albany.edu/sourcebook/pdf/t5222010.pdf. Accessed 12 Dec 2011

Notter P, Arnold C, von Erlach E, Hertig P (2006) Lesen und Rechnen im Alltag. Grundkompetenzen von erwachsenen in der Schweiz. Office fédéral de la statistique (OFS), Neuchâtel

Peters K (1970–1974) Fehlerquellen im Strafprozess. Eine Untersuchung der Wiederaufnahmeverfahren in der Bundesrepublik Deutschland (3 vols). C.F. Müller, Karlsruhe

Piquerez G (2006) Traité de procédure pénale Suisse, 2nd edn. Schulthess, Zurich

Thomas GC (2008) The supreme court on trial. How the American justice system sacrifices innocent defendants. University of Michigan Press, Ann Arbor

Trechsel S, Martin K (2004) Law of criminal procedure. In: François D, Tugrul A (eds) Introduction to Swiss law, 3rd edn. Kluwer Law International, The Hague

Walpin G (2003) America's adversarial and jury systems: more likely to do justice. Harvard J Law Publ Policy 26:175–186

Wright RF (2009) How prosecutor elections fail us. Ohio State J Crim Law 6:581–610

Y

Youth Homicide in the United States

Angela Browne[1], Kirk R. Williams[2], Robert Nash Parker[3], Kevin J Strom[4] and Kelle Barrick[4]
[1]Vera Institute of Justice, Washington, DC, USA
[2]Department of Sociology and Criminal Justice, University of Delaware, Newark, DE, USA
[3]University of California, Riverside, Riverside, CA, USA
[4]RTI International, Research Triangle Park, NC, USA

Synonyms

City homicide rates; Crime trends; Homicide perpetration; Homicide rates/trends; Juveniles, young adults; Youth violence

Overview

This entry reviews youth homicide trends in the United States over a 23-year period and describes key factors affecting increases and decreases in homicides by youth. It also summarizes a study of youth homicide perpetration from 1984 to 2006 that examined trends and predictors for two groups of young perpetrators – 13–17-year-olds and 18–24-year-olds. This study modeled city-specific predictors of *increases* in homicides by these age groups for 91 of the largest US cities. Overall, findings indicate that rates of homicide for juvenile and young adults followed the same general pattern over the 23 years: with a sharp escalation in 1984–1993, a significant downturn from 1994 to 1999, and a subsequent increase in 2000–2006. Structural disadvantage in large US cities was strongly associated with the sharp rise in perpetration of homicide by juveniles and young adults during the national homicide "epidemic" in the mid-1980s to early 1990s, as well as the more recent increase from 2000 to 2006. Gang presence and activity and illegal drug market activity were also consistently associated with increases in homicide offending by both age groups for both periods. Density of alcohol retail outlets had a significant relationship with increases in youth homicide perpetration for both age groups as well. This entry concludes with recommendations for prevention for city officials, local and federal legislators, stakeholders, providers, and advocacy groups.

Trends in Youth Homicide in the United States: 1984–2005

Since 1984, the United States has experienced two dominant trends in youth homicide**:** a dramatic increase in the mid-1980s and early 1990s, followed by a sharp decline for the rest of that decade. From the mid-1980s until the early 1990s, US homicide rates rose to their highest level since the beginning of modern crime statistics. This startling and unanticipated increase was characterized as an "epidemic" by leading homicide investigators like Cook and Laub (1998) and

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

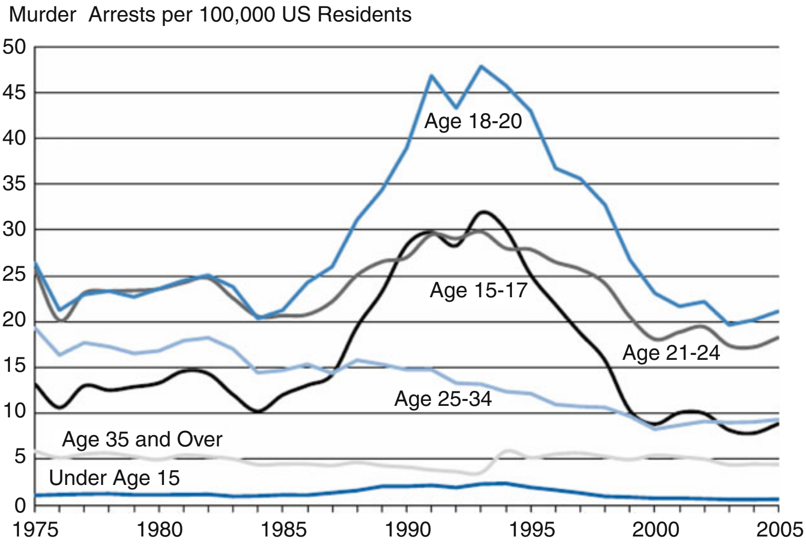

Youth Homicide in the United States, Fig. 1 Murder arrests per 100,000 US residents: 1975–2005 (Source(s): Analysis and weighting of sample-specific data from the Federal Bureau of Investigation. Crime in the United States, annual, Washington, DC: FBI, U.S. Department of Justice. Cited in Butts, J.A., and H.N. Snyder (2006). *Too Soon to Tell: Deciphering Recent Trends in Youth Violence*. Issue Brief #110. Chicago: Chapin Hall Center for Children, University of Chicago)

Messner and his colleagues (2005). It mostly involved males, was fairly widespread across the country, and predominantly involved firearms (Fox and Zawitz 2002; Messner et al. 2005). Leading homicide researchers like Blumstein and his colleagues (2000), Cook and Laub (1998), Ousey and Augustine (2001), Ousey and Lee (2002), and many others offered explanations for the epidemic. Theories included greater prevalence of drug trafficking and drug markets in cities, greater availability and carrying of firearms, increased gang activity, changes in labor market structures, diminished economic opportunities for people living in marginalized communities, and diminished social and monetary support for families in poverty.

This historic spike in youth homicide culminated in 1993. Homicide rates then dropped dramatically, again unexpectedly. Explanations offered for the decline included reduced activity in illegal drug markets and lessening demand, improved economic conditions, greater access to housing and employment in urban areas, changes in alcohol consumption, and increased emphasis on law enforcement and imprisonment (e.g., Blumstein 1995; Blumstein and Wallman 2006; Blumstein et al. 2000; Ousey and Lee 2002; Moore and Tonry 1998). The decline lasted until the early 2000s, when reports from some US cities indicated a resurgence of urban-based violence by juveniles and young adults. National statistics also reflected an upsurge in violence and use of firearms among these groups. Homicide rates for young black males began to trend up in 2000 (Snyder and Sickmund 2006). Overall arrests of 15–24-year-olds for murder, robbery, and weapons offenses increased in 2004–2005 – for the first time in a decade (Butts and Snyder 2006). This unanticipated shift encouraged a renewed emphasis on identifying predictors of change (Fig. 1).

Factors Affecting Homicide Trends in the USA

The Relationship Between Cities and Homicide

Homicide trends vary greatly across types of homicide, as well as by age, race, circumstance-context, relationship, and gender (e.g., Flewelling and Williams 1999; Williams and Flewelling 1988; Blumstein 1995; Parker and Rebhun 1995). Rates must be disaggregated by key characteristics in order to identify patterns and predictors, since aggregated rates often mask counter trends and circumstances influencing homicide patterns for different groups (e.g., Browne and Williams 1993; Browne et al. 1999).

To understand youth homicide in the USA, it is critical to assess dynamics within cities, due to (a) the high proportion of US homicides that occur in urban areas and (b) the predominance of young, mostly male, perpetrators and victims in city-based homicide events.

Homicide trends vary across cities as well. For example, although the unprecedented increase in youth-perpetrated homicides from the mid-1980s to the early1990s was called a "national" epidemic, many US cities did *not* experience an increase in youth homicide during that period, and those that did demonstrated differing patterns in the epidemic's beginning, duration, and decline. Using Uniform Crime Reporting (UCR) data, Messner and his colleagues (2005) analyzed overall homicides rates for large US cities for the years 1979–2001 and found that just over *half* of the cities in the sample conformed to the national trend over the period. Large cities were most likely to experience an epidemic-like cycle; densely populated cities and cities with more severe levels of socioeconomic deprivation tended to have earlier entrance into and exit from the spike in youth homicide rates.

Similarly, McDowall and Loftin (2009) analyzed the extent to which trends in city-based crime rates corresponded to national crime trends over time using panel data from 139 large cities with populations over 100,000 for at least 30 years of a 45-year period (1960–2005), based on UCR offenses and population rates. In that 45-year period, national conditions must have exerted an effect at the city level during years of stability, rapid increase, and rapid decline to influence city crime rates. Results indicated that in an average year, nearly two-thirds of cities and their residents were subject to the national crime trend for that period. As with Messner et al.'s findings, larger cities were most likely to follow the national pattern, although smaller jurisdictions were also affected by nationwide conditions. Much of the variation in city crime rates remained dependent on *local* conditions *within* cities, however. National trends accounted for only about *one-fifth* of variation in local rates when between-city differences were removed.

City Characteristics Associated with Homicide

A variety of characteristics have been associated with homicide rates at the city level. In their classic study of *between*-city differences, Land and colleagues (Land et al. 1990) identified a set of key covariates that included resource deprivation (and/or affluence), e.g., families living below the poverty line, median family income, income inequality, and percent unemployed; family structure, e.g., percent divorced males and percent of children not living with both parents; population structure, including size and density and percent of the youth population; and geographical region in the USA. After the youth homicide epidemic of the late 1980s and early 1990s, other explanatory variables were added, including increases in illicit drug activity (particularly the trafficking of crack cocaine) and drug markets, a rise in US incarceration rates after 1980, and changes in law enforcement practices (Ousey and Lee 2002).

McCall and her colleagues (2008) extended earlier research by studying the extent to which *within-city* (rather than between-city) changes in structural covariates were associated with changes in the homicides rate for 83 of the largest US cities (populations over 100,000) from 1970 to 2000 and added shifts in the economy. Using a pooled cross-sectional time series model to examine the influence of within-city changes in covariates on changes in homicide rates over time, this study found that changes in resource deprivation, changes in the relative size of the youth population, and increases in drug sales arrest rates were significantly related to changes in homicide rates over four time points: 1970, 1980, 1990, and 2000.

Baumer (2008) also attempted to extend prior work on covariates by developing a more defined portrait of the determinants of crime trends for 1980–2004. He incorporated prior traditional factors and added immigration, wages, alcohol consumption, and youthful cohort variables for a sample of 114 US cities with populations of 100,000 or more (in 1980). Based on econometric panel modeling techniques, Baumer reported that unemployment rates were associated with increases in gun homicide rates overall; wages were negatively associated with non-gun homicides (pp. 159). Drug market measures yielded the most consistent significant effects in the study; these effects were strongest for youth homicide arrests. Baumer estimated that between 20 % and 40 % of the *increases* in the rates of overall homicide, gun homicide, and youth-perpetrated homicide from 1984 to 2004 were attributable to drug market activity and drug involvement. Conversely, a drop in unemployment during the 1990s coupled with a rise in wages during that period explained as much as 30 % of the *decline* in youth and non-gun homicides.

Structural Disadvantage in Cities

The impact of *economic and social disadvantage* on lethal violence in citie*s* has been documented across empirical studies of youth and adult homicide (Blumstein and Wallman 2006; Ousey and Augustine 2001; McCall et al 2008; Williams and Flewelling 1988). As noted, in the USA, homicide in general – and youth homicide in particular – is highly concentrated in communities with high levels of economic deprivation, joblessness, family disintegration, and racial segregation (Sampson and Wilson 1995). For example, the period of the youth homicide epidemic in the late 1980s corresponded to major changes in structural disadvantage in urban areas that weakened families and institutions responsible for social services and control.

Communities with higher levels of concentrated disadvantage often have fewer resources for protecting youth from involvement in illegal and dangerous activities, preventing proliferation of gangs, or mitigating the development of illegal drug markets. In their study of 83 large US cities for the years of 1970–2000, McCall and her colleagues (2008) found that an index of "resource deprivation" provided some of the strongest associations with increases in homicide rates within cities. Cities with *decreases* in the percent of families living in poverty, income inequality, and percent of children not living with both parents, coupled with *increases* in median family income, experienced *decreases* in their overall homicide rates. Each component also was independently and significantly associated with changes in city homicide rates.

Although black youth account for a disproportionate share of homicide arrests compared to their representation in the US population (e.g., Puzzanchera 2009), much of this may be explained by disadvantage. As in many other countries, structural disadvantage – and a corresponding lack of opportunity – is often particularly severe in communities of color. In their study of how city-level changes in social and economic disadvantage contribute to increases in youth homicide victimization for 15–19-year-old juveniles and 20–24-year-old young adults from 1990 to 2000, Strom and MacDonald (2007) found that increases in social and economic disadvantage were positively associated with increases in Black teenage, Black young adult, and White teenage homicide rates, independent of other factors. In general, cities with *increasing* economic disadvantage and family instability showed *increases* in youth homicide for both Blacks and Whites, independent of drug arrest rates, ethnic heterogeneity, region, and population density.

Illegal Drug Markets and US Homicide

Similar effects have been found for the influence of illegal drug markets on youth homicide. Analyzing the role of illegal drug markets in the 1980s–1990s youth homicide epidemic, Ousey and Lee (2002) used longitudinal data to examine whether within-city variation in the sale or manufacture of cocaine/opiates from 1984 to 1997 was associated with within-city variation in homicide perpetration rates during the period. Ousey and Lee modeled city-specific changes in drug market activity and homicide perpetration rates for 15–29-year-olds and the effect of between-city variation in levels of resource deprivation and found that levels of concentrated disadvantage within cities also influenced the relationship of drug market activities and homicide. Cities with *higher* levels of resource deprivation had significantly stronger relationships between illegal drug market activities and homicide perpetration by 15–29-year-olds. In cities with the *lowest* levels of resource deprivation, the relationship became negative.

The most popular explanation offered for the suddenness and extremity of the homicide epidemic was the explosion of illegal drug markets linked to the introduction of crack cocaine in the 1980s (Baumer et al. 1998; Blumstein and Wallman 2006). This theory was used to explain the sharp decline in the epidemic as well (e.g., Blumstein and Rosenfeld 1998). Although subsequent empirical studies documented the importance of illegal drug markets to the 1980s–1990s epidemic, studies by Ousey and Lee (2002) and others demonstrate that the effects were highly influenced by social context. A growth in drug-related arrests during this period was related to higher homicide arrest rates, but only for non-Latino White juveniles (see Ousey and Augustine 2001). In their study of changes in overall homicide rates from 1970 to 2000 in 83 large cities, McCall et al. (2008) found that the "drugs and violence" link suggested as a cause for the spike in youth homicide during the mid-1980s and early 1990s did not explain the decline in those rates during the last half of the 1990s. Similarly, Baumer's (2008) analyses suggested that crack cocaine was not a strong predictor of the decline in the 1990s.

Gang Presence/Activities and US Homicide

Another major challenge for US cities over the past several decades has been gang-related violence. In 2007, 86 % of agencies serving large US cities reported gang problems, in contrast to 50 % of agencies serving suburban areas and 35 % of agencies serving smaller cities. Based on their 2007 survey, Egley and O'Donnell (2009) estimated that there were more than 27,000 active gangs and approximately 788,000 gang members in the United States, with 80 % of gang members residing in large cities and suburban counties. Reports of gang-related homicides also tend to concentrate in America's most populous cities. As with other youth crime indicators during this period, one in five large cities reported an *increase* in gang homicides in 2007 compared with the previous year. Approximately two in five large cities reported increases in other violent offenses by gang members. Adolescent gang members account for a disproportionate share of serious violent offenses committed by juveniles. Longitudinal studies conducted by Thornberry and colleagues (Thornberry 1998; Krohn and Thornberry 2008) document that youth are more prone to serious and violent offenses when actively involved with a gang than before or after that affiliation.

Firearm Availability and US Homicide

The presence of weapons – particularly firearms – also has been associated with shifts in US homicide trends as well as the youth homicide epidemic. Blumstein speculated that, as adult sellers dominating drug markets were imprisoned, crack markets were increasingly staffed by young inexperienced street sellers who, lacking maturity and other skills, resolved conflicts with overwhelming force, often through the use of firearms (Blumstein 1995, pp. 29–31). However, in Baumer's recent (2008) analysis of data from 114 large US cities for the years of 1980–2004, his measure of firearm prevalence was not significantly associated with either youth or adult homicide.

Alcohol Availability and US Homicide

Finally, as Parker and colleagues note (2011), the relationship between homicide and alcohol availability is well established by empirical studies in North America, Europe, and some parts of Asia. Findings linking *youth* homicide and the availability of alcohol for purchase are mostly missing from the literature, however, especially at the city level and across long time periods, leaving a critical gap in our understanding of homicide perpetration in the USA.

Combining Key Factors Affecting Youth Homicide Trends: 1984–2006

Based on these findings and gaps in the literature, Browne, Williams, Parker, Strom, and Barrick conducted a National Institute of Justice-funded study, *Anticipating the Future Based on Analysis of the Past: Intercity Variation in Youth*

Homicide, 1984–2006, analyzing trends and predictors of youth homicide perpetration rates among juvenile and young adults 13–17 years old and 18–24 years old in the 91 largest US cities (per 1980 Census) – the first major study to focus exclusively on intercity variation in youth homicide trends. This study (a) analyzed *temporal trends within* cities during the dramatic escalation, decline, and subsequent increase in youth homicide rates; (b) explored whether the factors discussed above accounted for *variation* in youth homicide trends *between* cities; and (c) assessed whether the predictors applied *equally* to juvenile and young adult perpetration. The goal was to ***identify city characteristics associated with increases in homicides by juveniles and young adults in large US cities so future increases might be anticipated and prevented***.

Measures

The research team created a measure of structural disadvantage that included traditional risk factors as well as illegal drug activity, gang presence-activity, firearm availability, and alcohol availability. Data on youth homicide were drawn from the *Supplementary Homicide Report (SHR)* – a component of the *FBI's Uniform Crime Reporting Program*. The study compensated for missing data on offender characteristics in the SHR by using multiple imputation techniques. After data were multiply imputed, incidents were aggregated to the city level for calculations. (See Parker et al. 2011: 507–510 for details on study methods.) Analyses included only murder and nonnegligent manslaughter and excluded homicide by negligence or justifiable homicides, since those homicides are relatively rare and are not included in most definitions of criminal homicide. Analyses of incident-level data also included only incidents involving a single victim and a single perpetrator; again homicides with multiple perpetrators or victims are relatively rare. Data on the availability of alcohol within cities were gathered from the US Census of Economic Activity conducted every 5 years and were used to construct an annual time series for the density of retail alcohol outlets in each city. Other city-level characteristics – structural disadvantage, drug market activities, gang presence-activity, and firearm availability – were derived from the US Census County and City Data Books, the SHR, the annual National Youth Gangs Survey (Egley et al. 2006), and the Vital Statistics Multiple Cause of Death File, respectively.

Analyses

Analyses utilized Hierarchical Linear Modeling (HLM), effectively used in city-level homicide research by Ousey and Lee (2002) and others. This approach was selected since a set of models for nested data can be estimated simultaneously, such that the influence of the independent predictors of the dependent variable is estimated and, simultaneously, factors that measure differences among the clusters of cases at level 2 are assessed for their influence on the effects of the independent variables at the lower level at level 1. A second multilevel analytical approach, pooled cross-sectional time series (c.f. McCall et al 2008), was used to provide an additional analytical procedure with enhanced statistical power compared to HLM.

Key Findings

Using this combined measure of traditional risk factors, drug activity, gang presence-activity, and firearm and alcohol availability revealed that, **for the two age groups**, homicide perpetration among juveniles and young adults followed the same general trend between 1984 and 2006 in large US cities, with a sharp increase in youth homicide perpetration rates during the epidemic of the late 1980s to early 1990s, a significant downturn from 1994 to 1999, and a significant upturn from 2000 to 2006. **For the two periods of increase**, factors significantly associated with the perpetration of lethal violence during the years of the homicide epidemic were also associated with the subsequent increases in rates of homicides in 2000–2006.

Structural disadvantage and **gang presence-activity** in cities had consistent and significant effects on trends in youth homicide perpetration during the youth homicide epidemic and the increases in 2000–2006 as well. The *greater* the structural disadvantage and gang presence-activity within a city, the *greater* the escalation in youth homicide rates during the epidemic years. Although not as strong, this pattern also held for intercity variation during the increase in youth homicide perpetration in 2000–2006. **Drug market activity** within cities and over time also demonstrated consistent and positive effects on trends in youth homicide perpetration for 13–17-year-olds and 18–24-year-olds in the pooled cross-sectional time series analysis with its greater statistical power (but not in the HLM analysis).

Firearm availability in the pooled cross-sectional time series analysis (but not the HLM analysis) demonstrated significant and positive effects on trends in homicide perpetration for young adults ages 18–24 years old over the period. No statistically significant effects for firearm availability were found for younger perpetrators 13–17 years of age. **Alcohol availability** did show consistent effects on youth homicide perpetration for both juveniles and young adults, based on the measure of the density of alcohol retail outlets in cities (Parker et al. 2011). Similar effects have been found for adults in national as well as neighborhood level quantitative studies, but this is the first study to report this for these age groups. The proportion of young people in the population did not have statistically significant effects on city rates of juvenile and young adult homicide perpetration.

There are a number of limitations to the study. The models in these analyses used macro-level measures that did not directly capture dynamics at the *neighborhood* or *individual* level. Additionally, the use of city-level data, while advantageous for comparing risk factors between cities, does not capture *within*-city variations. For example, these measures do not indicate whether certain neighborhoods within a city have high or low levels of disadvantage, gang presence, or alcohol availability. However, the youth homicide study identified significant explanatory predictors across a particularly significant 23-year time period using two analytic strategies and multiple variables to enhance understanding of factors influencing increases in US homicide perpetration for two critical age groups.

Implications and Future Directions

As noted, the ultimate goal of the youth homicide study was to identify city characteristics associated with *increases* in homicides by juveniles and young adults in large US cities so that federal, state/regional, and local criminal justice leaders, practitioners, and advocates could develop and utilize a data-driven approach for monitoring and preparing solutions for crime involving youth. Study results documented that the ***increase in youth homicide perpetration*** in the USA from 2000 to 2006 occurred for both juveniles and young adults. Findings also showed that, despite occurring in two distinct time periods, there were ***consistent predictors for escalation*** in the perpetration of lethal violence by juveniles and young adults.

This is important from a preparedness and a prevention standpoint. The strong and consistent findings on the association of structural disadvantage and violent crime across multiple years in both age groups and periods of escalation are of particular relevance. Given the severe economic challenges facing the US and many other countries and their impact on urban areas, city officials, police departments, legislators, and community leaders need to be particularly vigilant to offset an escalation in lethal violence by youth related to resources and budgetary priorities over time.

The presence of gangs, drugs, and alcohol at the city level also are key components in cross-sectional and over time variation in youth homicide. Residents in communities with concentrated levels of unemployment and severe economic deprivation may be less able to handle delinquent youth, control illicit activities, or offer juveniles and young adults positive alternatives. Results of the youth homicide study's pooled time series analysis show significant effects for gang presence-activity, drug-related activities, and (for 18–24-year-olds only) firearm availability across 23-year period, highlighting the importance of preventive efforts in these areas. Findings on the impact of alcohol availability for both age groups also suggest a point of intervention, by indicating that reduction in the density of retail alcohol outlets in urban areas and neighborhoods may be an effective tool for violent crime reduction among youth.

Finally, in the youth homicide study, there were marked similarities in city-level explanatory predictors associated with juvenile and young adult violence trends. Although Butts and Snyder (2006) encouraged the *inclusion* of 18–24-year-olds in studies of youth homicide trends, to our knowledge, this is the first study that examines lethal youth violence trends and their city-level predictors separately for juveniles (ages 13–17) and young adults (ages 18–24). The importance of the identified risk factors for both age groups is not surprising. There is considerable overlap in the interactions of teenagers and young adults in urban settings, including family and other social interactions as well as local gang and drug market activities. Much of the gang literature describes age-integrated gangs composed of juveniles and young adults. Homicide data also consistently documents a substantial crossover between juvenile and young adult victims and offenders (e.g., Cook and Laub 1998). ***It is critical that any interventions or initiatives seeking to prevent or reduce youth violence not simply focus on juveniles under age 18***. They must also address criminally involved young adults with whom juveniles come in contact if the younger youth are to be deterred and protected.

By identifying and better understanding factors related to increases in youth homicide perpetration, policy makers, researchers, and community leaders can develop more effective and targeted strategies for responding to "at-risk" juveniles and young adults *before* periods of escalation (e.g., although a downturn has been reported in some national rates from 2007 through 2008 for violent crimes by youth, the perpetration rate for homicide by Black male juveniles ages 14–17 continued to increase from 2002 to 2008: Cooper and Smith 2011). Empirical findings can be utilized as a first step by cities, states, and practitioners and policy makers to:

1. **Support At-risk Youth, Particularly Youth Living in Structurally Disadvantaged Neighborhoods.** It is critical that proactive resources and programs for at-risk youth be increased by using comprehensive and integrated community-level, school-level, family-level, and individual-level interventions designed to strengthen young people's "core competencies" as well as the social contexts in which they reside. As Williams, Guerra, and their colleagues have noted, having such supports will help juveniles and young adults more effectively – without violence and other illegal activities – "beat the odds" (Kim et al. 2008; Guerra and Bradshaw 2008). Empirical studies document that, even in the face of severe disadvantage and pervasive danger, most young people accomplish the developmental tasks necessary to withstand the risks and live productive, healthy, and responsible adult lives (e.g., Elliott et al. 2006).
2. **Create Prevention Initiatives and Programs That Target Young Adults Ages 18–24 as Well as At-Risk Juveniles.** Results of the youth homicide study clearly demonstrate that juvenile and young adult trends in homicide perpetration are influenced by similar factors (with the exception of firearms availability). As a result, interventions – whether in the form of policing initiatives or community-based prevention strategies – should not *exclude* young adults. For example, this is the case with under age access to alcohol, since half of the population at risk in the 18–24 age group is below the legal minimum purchase age for alcohol. Enforcing existing laws with regard to under age access may have an important preventative impact on younger and older age groups.
3. **Develop Comprehensive and Multidisciplinary Approaches for Addressing Gangs.** Findings from the youth homicide study and other empirical research also

demonstrate the importance of gang presence and activity in influencing increases in fatal violence by juveniles and young adults. It is urgent that federal, state or region, and city policy makers work with law enforcement and crime prevention experts to assess gang activity proactively and incorporate that knowledge not simply into police planning but also into creative citywide and regional initiatives for community-based and social policy interventions.

4. **Inform Preparedness and Prevention Activities More Effectively.** Federal and local leaders should develop and utilize more data-driven approaches for enhancing crime preparedness and prevention activities. These activities could then be tied directly to the deployment and strengthening of policing and crime prevention resources in a more efficient manner, based on identified predictors of lethal youth violence during the escalation period of 1984–1993 and the relevance of those predictors for increases in 2000–2006. **At the federal level**, this would include compiling the necessary data sources and identifying cities particularly at risk of youth homicide increases based on past and current structural conditions and city-specific trends. **At the local level**, mayors, police chiefs, and other agencies charged with responding to or preventing crime should consistently review and analyze the data for their jurisdiction to identify emerging patterns and concerns.
5. The type of data collection and analysis we recommend would involve (1) analyzing multiple years of data to explore past and emerging city-based trends and (2) the use of data sources in *addition* to police crime data, including key economic and social indicators as well as specific measures of the prevalence of gang presence-activity, drug-related activity, and alcohol outlet density in the city. The *timeliness* and *quality* of crime and structural indicator data must be improved as well, so that these measures are useful for resource deployment and prevention.
6. **Improve Quality of Crime and City-level Indicator Data.** There also is a critical need to develop improved measures. The goal is to develop measures that can be *standardized* across jurisdictions and are as *independent* as possible of police activities and initiatives. One mechanism is to develop measures through survey data collection processes that monitor levels of gang activity or drug activity in specific communities. The regular collection of information using these types of approaches may be cost-prohibitive for some jurisdictions, however. Ultimately, from a national policy standpoint, decisions must be made as to whether the benefits associated with increased resources for collecting independent measures provide a significant return in (a) understanding these phenomena within and across cities and (b) monitoring and intervening effectively over time. In the absence of improved measures, many of the surprises associated with imprecise measures will continue.

Since the early 1980s, responses to crime in the United States have centered on law enforcement, sentencing, and incarceration or detention. A growing body of criminal justice literature suggests that investments in policing and other official forms of social control will have disappointing results unless the preconditions of structural disadvantage and attendant problems are addressed. For communities, states, regions, and nations, the benefits of preventing criminal activities and their outcomes among youth will far outweigh the investments required for proactive support.

Related Entries

- ▶ Disadvantage, Disorganization and Crime
- ▶ Gangs and Social Networks
- ▶ Juvenile Violence
- ▶ Neighborhood Effects and Social Networks
- ▶ Poverty, Inequality, and Area Differences in Crime
- ▶ Risk Factors for Gang Membership

Recommended Reading and References

Baumer EP (2008) An empirical assessment of the contemporary crime trends puzzle: a modest step toward a more comprehensive research agenda. In National Research Council Committee on Understanding Crime Trends (eds) Understanding crime trends: Workshop Report. The National Academies Press, Washington, DC, pp 127–176

Baumer EP, Lauritsen JL, Rosenfeld R, Wright R (1998) The influence of crack cocaine on robbery, burglary, and homicide rates: a cross-city, longitudinal analysis. J Res Crime Delinq 35:316–340

Blumstein A (1995) Youth violence, guns, and the illicit-drug industry. J Crim Law Criminol 86:10–36

Blumstein A, Rosenfeld R (1998) Explaining recent trends in U.S. homicide rates. J Crim Law Criminol 88:1175–1216

Blumstein A, Wallman J (eds) (2006) The crime drop in America (rev. edn). Cambridge University Press, New York

Blumstein A, Rivara FP, Rosenfeld R (2000) The rise and decline of homicide—and why. Annu Rev Public Health 21:505–541

Browne A, Williams KR (1993) Gender, intimacy, and lethal violence: trends from 1976–1987. Gender Soc 7(1):78–98

Browne A, Williams KR, Dutton DG (1999) Homicide between intimate partners: a 20-year review. In: Smith D, Zahn M (eds) Homicide: a sourcebook of social research. Sage Publications, Thousand Oaks, CA

Butts JA, Snyder HN (2006) Too soon to tell: deciphering recent trends in youth violence. (Issue Brief No. 110). Retrieved from University of Chicago, Chapin Hall Center for Children website: http://www.nationalgangcenter.gov/Content/Documents/Deciphering-Recent-Trends-in-Youth-Violence.pdf

Cook PJ, Laub JH (1998) The unprecedented epidemic in youth violence. Crime Justice 24:27–64

Cooper AC, Smith EL (2011) Homicide trends in the United States, 1980–2008: annual rates for 2009 and 2010 (Report No. NCJ 236018). Retrieved from Bureau of Justice Statistics website: http://bjs.ojp.usdoj.gov/content/pub/pdf/htus8008.pdf

Egley A, Howell JC, Mayor AK (2006) National youth gang survey, 1999–2001. US Department of Justice, OJJDP, Washington, DC

Egley A, O'Donnell CE (2009) Highlights of the 2007 national youth gang survey (Report No. NCJ 225185). Retrieved from Office of Juvenile Justice and Delinquency Prevention website: https://www.ncjrs.gov/pdffiles1/ojjdp/225185.pdf

Elliott DS, Menard S, Rankin B, Elliott A, Wilson WJ, Huizinga D (2006) Good kids from bad neighborhoods: successful development in social context. Cambridge University Press, New York, NY

Flewelling R, Williams KR (1999) Categorizing homicides: the use of disaggregated data in homicide research. In: Smith D, Zahn M (eds) Homicide: a sourcebook of social research. Sage Publications, Thousand Oaks, CA

Fox JA, Zawitz MW (2002) Homicide trends in the United States. Retrieved from Bureau of Justice Statistics website: http://bjs.ojp.usdoj.gov/content/pub/pdf/htius.pdf

Guerra NG, Bradshaw CP (eds) (2008) Core competencies to prevent problem behaviors and promote positive youth development [Special issue]. New Directions for Child and Adolescent Development 122:1–92

Kim T, Guerra NG, Williams KR (2008) Preventing youth problem behaviors and enhancing physical health by promoting core competencies. J Adolesc Health 43:401–407

Krohn MD, Thornberry TP (2008) Longitudinal perspectives on adolescent street gangs. In: Liberman AM (ed) The long view of crime: a synthesis of longitudinal research. Springer, New York

Land KC, McCall PL, Cohen LE (1990) Structural covariates of homicide rates: are there any invariances across time and space? Am J Sociol 95:922–963

McCall PL, Parker KF, MacDonald JM (2008) The dynamic relationship between homicide rates and social, economic, and political factors from 1970–2000. Soc Sci Res 37:71–735

McDowall D, Loftin C (2009) Do U.S. city crime rates follow a national trend? The influence of nationwide conditions on local crime patterns. J Quant Criminol 25:307–324

Messner SF, Deane GD, Anselin L, Pearson-Nelson B (2005) Locating the vanguard in rising and falling homicide rates across U.S. cities. Criminology 43:661–696

Moore MH, Tonry M (1998) Youth violence in America. Crime Justice 24:1–26

Ousey GC, Augustine MC (2001) Young guns: examining alternative explanations of juvenile firearm homicide rates. Criminology 39:933–967

Ousey GC, Lee M (2002) Examining the conditional nature of the illicit drug market-homicide relationship: a partial test of the theory of contingent causation. Criminology 40:73–102

Parker RN, Rebhun LA (1995) Alcohol and homicide: a deadly combination of two American traditions. SUNY Press, New York, NY

Parker RN, Williams KR, McCaffree KJ, Acensio EK, Browne A, Strom KJ, Barrick K (2011) Alcohol availability and youth homicide in the 91 largest US cities, 1984–2006. Drug Alcohol Rev 30:505–514

Puzzanchera C (2009) Juvenile Arrests 2008 (Report No. NCJ 228479). Retrieved from the Office of Juvenile Justice and Delinquency Prevention website: https://www.ncjrs.gov/pdffiles1/ojjdp/228479.pdf

Sampson RJ, Wilson WJ (1995) Toward a theory of race, crime and urban inequality. In: Hagan J, Petersen R (eds) Crime and inequality. Stanford University Press, Stanford, CA, pp 27–54

Snyder H, Sickmund M (2006) Juvenile offenders and victims: 2006 national report (Report No. NCJ 212906). Retrieved from Office of Juvenile Justice and Delinquency Prevention website: http://www.ojjdp.gov/ojstatbb/nr2006/downloads/NR2006.pdf

Strom KJ, MacDonald JM (2007) The influence of social and economic disadvantage on racial patterns in youth homicide over time. Homicide Stud 11:50–69

Thornberry TP (1998) Membership in youth gangs and involvement in serious and violent offending. In: Loeber R, Farrington DP (eds) Serious and violent juvenile offenders: risk factors and successful interventions. Sage Publications, Thousand Oaks, CA, pp 147–166

Williams KR, Flewelling R (1988) The social production of criminal homicide: a comparative study of disaggregated homicide rates in American cities. Am Sociol Rev 53:421–431

Youth Violence

► Youth Homicide in the United States

Z

Zero Tolerance

► Discriminant Validity of Disorder and Crime

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

List of Entries

1%ers 1
ABM 1
Accomplice Crime 1
Actualizing Risk-Need-Responsivity 1
Adaptation 12
Administrative Criminology 12
Admissibility 12
Adolescence-Limited Offending 12
Adolescent Contexts 12
Affordable Care Act 12
Age-Crime Curve 12
Agent-Based Assessments of Criminological Theory 19
Agent-Based Modeling for Understanding Patterns of Crime 32
Agent-Based Modelling 41
Agent-Based Models to Predict Crime at Places 41
Aging Correctional Populations 48
Aging Prison Population: Factors to Consider 60
AIDS 67
Alcohol Dependence 67
Alternative Courts 67
Alternatives to Pre-trial Detention 68
American Dream 75
AML 75
Analytical Criminology 75
Anomie 75
Anomie and Crime 76
Anomie Theory (Strain Theory) 86
Antisocial Personality Disorder 86
Antisocial Potential 86
Anxieties About Crime 86
Applied Geographical Profiling 86
Arms Races 100
Associations 100
Auto Theft 100
Automated 100
Automated and Manual Forensic Examinations 100
Babies Behind Bars 109
Bayesian Learning 116
Bayesian Updating and Crime 116
BCS 125
Behavioral Health Courts 125
Behavioral Investigative Advice 125
Behavioral Learning 134
Behavioral Linkage Analysis 134
Behavioral Management in Probation 134
Behavioral Science Evidence in Criminal Trials 145
Bias Crime 154
Bias in Forensic Science 154
Bias/Hate Motivated Crime 154
Biased Policing 154
Bicycle Theft 162
Biker Clubs/Gangs 170
Biological Geographical Profiling 171
Biometrics 179
Biometrics and Border Control Policing 179
Boot Camps 188
Border Control 188
Brain 188
Breakdown 188
Break-In 188
Breaking and Entering 188
Bribery 188
British Crime Firms 188
British Crime Survey 194
British Police 203
Broken Windows Thesis 213

G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*,
DOI 10.1007/978-1-4614-5690-2, © Springer Science+Business Media New York 2014

"Broken Windows" Policing 223
Brutality 223
Bullying 223
Bullying Prevention: Assessing Existing Meta-Evaluations 231
Buprenorphine 242
Burglary at Construction Sites 242
Calibration of Simulation Models 243
Cambridge-Somerville Youth Experiment 252
Cambridge-Somerville Youth Study 255
Camera Surveillance 255
Campus Crime 255
Capture Recapture to Estimate Criminal Populations 267
Car Theft 276
Career Criminals and Criminological Theory 276
Careers in Organized Crime 285
Case Management 295
Causal Processes 295
Cause of Police Legitimacy 295
Causes 295
Causes of Police Legitimacy 295
Causes of Wrongful Convictions 303
CCTV 310
CCTV and Crime Prevention 310
Centre for Children Committing Offences (CCCO) 319
Certainty, Severity, and Their Deterrent Effects 319
Child Pornography 328
Choice Theory 338
Chronic Offenders 338
Chronic Victimization 338
Chronic Victims 338
City Homicide Rates 338
Civil Actions 338
Civil Remedies 338
CJS 353
Clinical Trials 353
Closed-Circuit Television 353
Coalition 353
Cocaine 353
Coercion 353
Coevolution 353
Cognitive Change 353
Cognitive Forensics: Human Cognition, Contextual Information, and Bias 353
Cognitive Transformation 363
Cognitive/Information Processing Theories of Aggression and Crime 363
Collective Action 376
Communal Context 376
Communities 376
Communities and Crime 376
Communities and the Police 376
Communities That Care 393
Community Control: Criminality, Community Social Organization, Cultural Organization, Deviant Opportunities, Directs Controls 408
Community Courts 408
Community Organization 416
Community Policing 416
Community Service in Europe 426
Community Supervision 442
Community-Oriented Policing 442
Comparative 442
Comparative Case Analysis 442
Comparative Courts and Sentencing 442
Comparative Criminal Justice 451
Comparative Incarceration 451
Comparative Legal Cultures 458
Comparative Police Systems 467
Comparative Political Economy of Punishment 467
Comparative Punishment 478
Comparing Police Systems Across the World 478
Compliance and Corporate Crime Control 488
Compstat 496
Computational Criminology 505
Computer Forensics 505
Concept of the Police 514
Conceptualizing of Police 515
Concurrent Validity 524
Conditional Conviction 524
Conditional Release 524
Confessions 524
Confirmation Bias 524
Constabularization 524
Construction Site Theft and Burglary 524
Content Analysis 524
Contextual Influences 525
Contingency Management 525
Continuity and Change 525
Control of Police Misconduct 525

Control Theory 535
Controlled Experiments 548
Co-offending 548
Co-offending and Offender Attributes 559
Co-offending and Offender Decision-Making 570
Corporate Crime 581
Corporate Crime Control 581
Corporate Crime Decision-Making 581
Corporate Crimes and the Business Cycle 588
Corporate Criminal Liability 600
Corporate Liability 600
Correction 611
Correctional Administrators 612
Correctional Education of Programs 612
Correctional Interventions 617
Correctional Treatment 617
Correctional Workers in the Organizational Change Process 617
Corrections 630
Corruption 630
Cost-Benefit Analysis 641
Counterterrorism 651
Counterterrorism and the Media 658
Counterterrorism Strategies 664
Courts 664
Covert Policing 664
Creativity 665
Credibility Assessment 665
Crime 665
Crime and Justice 665
Crime and Justice on Television 665
Crime and the Racial Composition of Communities 673
Crime Concentration 681
Crime Investigations by the International Criminal Court 681
Crime Linkage Analysis 691
Crime Location Choice 691
Crime Mapping 699
Crime on Public Transport 709
Crime Prevention 720
Crime Prevention Through Environmental Design 720
Crime Preventive Effects of Incapacitation 731
Crime Profiling 738
Crime Repetition by Youth 738
Crime Scene Analysis 738
Crime Scene Interpretation 738
Crime Scene Profiling 738
Crime Science 738
Crime Simulation 748
Crime Spates 748
Crime Specialization, Progression, and Sequencing 748
Crime Sprees 759
Crime Trends 759
Crimes Against Animal Life 759
Crimes Against Humanity 769
Crimes of Empire 769
Crimes of Globalization 769
Crimes of the Powerful 778
Crimes Without Victims 786
Criminal Activity of Sex Offenders 786
Criminal Career 786
Criminal Career of Sex Offenders 787
Criminal Careers 800
Criminal Careers and Public Policy Responses 800
Criminal Careers of Places 808
Criminal Cases Review Commission 816
Criminal Defense Profession 816
Criminal Interviewing 828
Criminal Investigation 828
Criminal Investigative Analysis 828
Criminal Investigative Profiling 838
Criminal Justice System 838
Criminal Law 838
Criminal Personality Profiling 838
Criminal Phenomenon: Crime, Criminal, Criminality 838
Criminal Profiling 838
Criminal Punishment 838
Criminal Thinking Model 838
Criminalistics 838
Criminalization and Occupational Health and Safety 838
Criminological Theory 848
Criminology of Place 848
Criminology of Places 857
Critical Criminology 857
Critical Incidents 858
Critical Report on Forensic Science 867
Cross-National 881
Cross-National Crime Rates 881
Cross-National Performance in Policing 881

Cryptology and Security 892
CTC 904
CTC Prevention Coalition 904
CTC Prevention System 904
Cultural Criminology 904
Customary Justice 913
Decision Making Theory 915
Defiance and Motivational Postures 915
Defiance Theory 925
Defining Disorder 932
Delinquency 939
Democratic Policing 939
Design 947
Design Out Crime 947
Designing Out Crime 948
Designing Products Against Crime 948
Desistance 958
Desistance and Supervision 958
Desistance from Crime 967
Desistance from Gangs 978
Desistors 988
Detecting Deception with fMRI 988
Detection of Deception 997
Deterrence 998
Deterrence of Tax Evasion 1005
Deterrence: Actual Versus Perceived Risk of Punishment 1014
Deterrent Effect of Imprisonment 1023
Development of International Criminal Law and Tribunals 1030
Development of Sex Offending 1045
Development of the UCR and the NCVS 1045
Developmental Criminology 1053
Deviance Theory 1054
Differences-in-Differences in Approaches 1054
Differential Association Theory 1065
Differential Association-Reinforcement Theory 1075
Differential Enforcement 1075
Differential Group Organization 1075
Differential Offending 1075
Differential Social Organization 1075
Disadvantage, Disorganization and Crime 1084
Disclosure 1093
Discrete Spatial Choice 1093
Discriminant Validity of Disorder and Crime 1093
Disorder: Observational and Perceptual Measures 1103
Disproportionate Minority Contact 1114
Dissocial Personality Disorder 1114
Distributional Analysis 1114
DNA 1114
DNA Databasing 1114
DNA Fingerprinting 1114
DNA Identification 1114
DNA Profiling 1114
DNA Technology and Police Investigations 1128
Domestic Burglary 1136
Domestic Terrorism 1136
Domestic Violence 1137
Double-Blind Administration 1146
Dowry Deaths 1146
Drug Abuse and Alcohol Dependence Among Inmates 1147
Drug Corners 1159
Drug Courts 1159
Drug Courts' Effects on Criminal Offending 1170
Drug Dealing Places 1178
Drug Distribution 1178
Drug Enforcement 1179
Drug Marketing 1186
Drug Markets 1186
Drug Substitution Programs and Offending 1186
Drug Trade 1194
Drug Trafficking 1194
Drug-Related Crime 1201
Dutch Colonial Police 1201
Dutch Crime Networks 1211
Dynamical Simulation in Criminology 1220
Dyssocial Personality Disorder 1231
Earlscourt Child and Family Centre (ECFC) 1233
Earlscourt Girls Connection (EGC) 1233
Early Chicago School 1233
Early Chicago School Theory 1233
Early Modern Police and Policing 1243
Early Quantitative Methods 1256
Early Release from Prison 1256
EBP 1271

Eco-crime 1271
Ecology of Crime 1271
Econometrics of Crime 1271
Economic Dominance 1280
Economic Theory of Criminal Behavior 1280
Ecoterrorism 1286
Effective Community Supervision Programming 1294
Effective Supervision Principles for Probation and Parole 1308
Effectiveness of Situational Crime Prevention 1318
Electronic Ball and Chain 1329
Electronic Bracelet 1329
Electronic Incarceration 1329
Electronic Monitoring 1329
Electronic Tagging 1338
Electronic Tracking of Offenders 1338
Elite Deviance 1338
Emergency Response 1338
Empathy and Offending 1338
Environmental and Human Rights 1344
Environmental Crime 1355
Environmental Criminology 1355
Environmental Extremism 1355
Environmental Regulation and Law Enforcement 1355
Environmental Rights 1367
Epidemiology of Cardiovascular Disease 1368
Equivocal Death Analysis 1368
Establishing Causes of Offending in Longitudinal and Experimental Studies 1368
Estimating the Effects of Incapacitation 1378
Estimator Variables and Eyewitness Identification 1384
Ethnic Violence 1394
Ethnoviolence 1403
European Perspectives on Long-Term Imprisonment 1403
Evaluating Truthfulness 1413
Evaluating Truthfulness: Interviewing and Credibility Assessment 1413
Evaluation 1426
Event Characteristics 1426
Event Control: Criminal Event, Routines Activities, Presentness, Occasions 1426
Evidence-Based Corrections 1426
Evidence-Based Crime Policy 1426
Evidence-Based Policing 1426
Evidence-Based Policy and Practice 1437
Evidence-Based Policy in Crime and Justice 1437
Evidence-Based Supervision 1446
Evidence-Led Policing 1446
Evolution of Substance Abuse Treatment 1446
Evolutionary Perspectives on Crime Prevention 1447
Evolutionary Theories of Criminal Behaviors 1454
Evolving Judicial Roles 1463
Examining the Effectiveness of Correctional Interventions 1471
Exchange 1479
Executions 1479
Exonerations 1479
Experiential Effects 1479
Experimental 1479
Experiments 1479
Expert Evidence 1479
Expert Evidence and Criminal Trial Procedure 1480
Expert Witnesses 1494
Expert Witnesses: Role, Ethics, and Accountability 1494
Extended-Release Naltrexone 1502
Extent of Imprisonment: Global View 1502
Extortion Racketeering 1509
Eyewitness Evidence: Historical Background 1509
Eyewitness Identification Evidence 1518
Eyewitness Research 1530
Eysenck's Personality Theory 1538
Fact-Finding Interviews 1547
False Confessions and Police Interrogation 1547
False Memories 1555
Family Engagement Strategies to Reduce Crime 1563
Fatigue 1573
FBI Influence on State and Local Police 1573
Fear and Punishment 1581
Fear of Crime and the Psychology of Risk 1594
Feminist Criminological Theory 1603
Feminist Theory and Domestic Violence 1612
Feminist Theory in the Context of Rape 1617

Feminist Theory in the Context of Sexual Violence 1617
Fencing/Receiving Stolen Goods 1627
Field Trials 1637
Fines in Europe 1637
Fingerprint Identification 1648
Fingerprints 1664
Flogging 1664
Focused Deterrence and “Pulling Levers” 1664
Force 1675
Forced Migration and Human Rights 1675
Forecasting Crime or Crime Forecasting 1682
Forecasting of Shootings Using Risk Terrain Modeling 1682
Forensic Anthropology 1692
Forensic DNA Matching 1705
Forensic Environmental Evidence 1705
Forensic Evidence 1713
Forensic Facial Analysis 1713
Forensic Genetics 1729
Forensic Linguistics 1729
Forensic Palynology 1741
Forensic Palynology (The Study of Pollen Grains) 1754
Forensic Science 1754
Forensic Science and Bias 1754
Forensic Science and Criminal Inquiry 1754
Forensic Science and Criminal Investigation 1754
Forensic Science and Miscarriages of Justice 1763
Forensic Science and the Paradigm of Quality 1773
Forensic Science Culture 1782
Forensic Science Effectiveness 1795
Forensic Science Evidence 1805
Forensic Science in the Nineteenth and Twentieth Centuries 1805
Forensic Science Professions 1820
Fraternal Organizations 1829
Fraud Victimization 1829
French Colonial Police 1836
Frequent Victimization 1846
Front Door and Backdoor Sentencing 1846
Gangs and Social Networks 1857
Gendarmerie Policing 1866
Gender and Race Differences 1876
Gender-Based Violence 1876
Gendered Theory and Gendered Practice 1876
Gendering Traditional Theories of Crime 1883
General Patrol 1892
General Strain Theory 1892
General Theory of Crime 1900
Genes, Crime, and Antisocial Behaviors 1907
Genetic Basis to Self-Control 1915
Genetic Expression 1924
Genetic Profiling 1924
Genocide 1924
Geo-forensic Analysis 1933
Geographic Profiling 1934
Geographical Analysis 1942
Geographical Offender Profiling 1942
Geography of Crime 1943
Geography of Crime and Disorder 1943
German Police Until 1918 1956
GLM 1966
Good Lives Model 1966
Governance 1976
Graft 1976
Green Criminology 1976
Group Characteristics and General Strain Theory 1984
Group Crime 1991
Group-Based Trajectory Modeling 1991
Group-Based Trajectory Models 1991
Group-Based Trajectory Models and Developmental Change 2003
Growth Curve Models with Categorical Outcomes 2013
GST 2025
Guilds 2025
Handwriting as Evidence 2027
Handwriting Examination: The Art and Science 2036
Harassment, Threat Assessment 2036
Hate Crime 2036
Health and Safety Law 2047
Health and Wellness 2047
Health Care Reform and Criminal Offenders (US) 2047
Health Issues in Prison Reentry Models 2053
Hepatitis C 2062
Heroin-Assisted Treatment 2062
Hidden Crime Survey 2062

High Policing 2062
High-Frequency Victimization 2074
Historical and Comparative Perspectives on Incarceration 2074
Historical Criminology 2081
History of Bail 2087
History of Boot Camps 2099
History of Capital Punishment 2106
History of Corporal Punishment 2106
History of Corrections 2115
History of Crime 2127
History of Criminal Investigation 2127
History of Criminological Theories: Causes of Crime 2137
History of Fines 2149
History of Forensic Science in Policing 2153
History of Geographic Criminology Part I: Nineteenth Century 2159
History of Geographic Criminology Part II: Twentieth Century 2164
History of Geographical Criminology 2172
History of Green Criminology 2172
History of Juvenile Courts 2181
History of Juvenile Justice 2191
History of Organized Crime in Mexico 2200
History of Organized Crime in the United States 2204
History of Police Unions 2207
History of Probation and Parole in the United States 2217
History of Randomized Controlled Experiments in Criminal Justice 2227
History of Restorative Justice 2239
History of Staff Skills in Probation 2249
History of Substance Abuse Treatment 2257
History of Technology in Policing 2269
History of the Death Penalty 2277
History of the Dutch Crime Victimization Survey(s) 2286
History of the Police Profession 2296
History of the Self-Report Delinquency Surveys 2309
History of the Statistics of Crime and Criminal Justice 2319
HIV 2326
HIV in the Correctional System 2326
Home Burglary 2336
Homegrown Terrorism in the United States 2336
Homicide Perpetration 2345
Homicide Rates/Trends 2345
Honor Killings 2345
Hormones, the Brain, and Criminality 2345
Hot Dots 2351
Hot Products 2351
Hot Smudges 2351
Hot Spots 2352
Hot Spots and Place-Based Policing 2352
Hot Targets 2363
Housing Choice Vouchers and Escaping Neighborhood Crime 2363
How to Make Crime Mapping More Inferential 2371
Human Immunodeficiency Syndrome 2371
Human Rights 2371
Human Rights Violations in Criminal Court 2371
Human Smuggling 2382
Human Trafficking 2389
Hybrid Gangs 2389
ICAP Theory 2391
ICC 2391
Identification and the Development of Forensic Science 2391
Identification Issues in Life Course Criminology 2398
Identification Technologies in Policing and Proof 2407
Identity Theft 2419
Illicit Retail Marketplaces 2429
ILP 2429
Images 2429
Impacts of Community-Oriented Policing 2430
Impartiality and Fairness in Judging 2435
Imprisonment 2435
Imprisonment and Cardiac Risk 2435
Incapacitation 2443
Incarcerating Terrorists 2450
Incarceration 2459
Incarceration Rates 2459
Incivilities 2459
Incivilities Thesis 2459
Inclusionary Control 2459
Indigenous 2459

Indigent Defense 2459
Individual Characteristics and General Strain Theory 2459
Individual-Based Modelling 2470
Individual-Based Simulation Modeling 2470
Individualization 2470
Inferential Crime Mapping 2470
Infiltration 2479
Informal Guardianship 2480
Informal Social Control 2490
Information Technology and Police Work 2501
Information-Led Policing 2513
Innocence Commissions 2514
Innovation and Crime Prevention 2523
Institutional Anomie Theory 2531
Institutional Environment 2542
Institutional Imbalance 2542
Institutional Legitimacy 2542
Institutional Theories of Punishment 2542
Institutional Theory 2552
Institutions (Political, Legal, Knowledge) 2552
Integrated Cognitive Antisocial Potential Theory 2552
Integrating Rational Choice and Other Theories 2564
Intelligence 2573
Intelligence-Led Policing 2573
Intensive Probation and Parole 2581
Interactional Theory of Delinquency 2592
Intermediate Sanctions 2601
Internal Security Strategies 2601
International 2601
International Crime Victimization Survey 2602
International Crimes 2614
International Criminal Court 2614
International Evidence Exchange 2614
International Human Rights Standards and Community Sanctions 2614
International Policy Transfer 2623
International Responses to Victims in Criminal Justice 2632
International Sentencing 2643
International Tribunals 2653
Internet Pornography 2653
Interrogation 2653
Interrupted Time Series Analysis 2653
Interrupted Time Series Models 2653
Intervention and Prevention in China 2665
Intervention Research 2673
Interview and Interrogation Methods Effects on Confession Accuracy 2673
Interviewing Eyewitnesses 2679
Investigative Interviewing 2689
Investigative Psychology 2689
IRCIP 2699
Italian Mafia 2699
Italian Organized Crime 2706
Jail 2707
Job of a Judge 2707
Joint/Group Offending 2707
Journey to Crime 2707
Judicial Leadership and Performance 2707
Judicial Temperament 2716
Junk Science 2716
Juries, Lay Judges, and Trials 2716
Jury Decision Making and Eyewitness Testimony 2727
Jury Impartiality in the Modern Era 2735
Justice Reinvestment 2745
Juvenile Awareness Programs 2756
Juvenile Crime 2756
Juvenile Diversion 2756
Juvenile Justice 2765
Juvenile Justice in the Get Tough Era 2766
Juvenile Justice System Responses to Minority Youth 2773
Juvenile Sex Offenders 2785
Juvenile Victimization 2785
Juvenile Violence 2793
Juveniles, Young Adults 2804
Key Skills Involved in Judging 2805
Labeling and Deviance 2807
Labeling Theory 2807
Labeling: History and Concept 2814
Labelling Perspective 2814
Labelling Theory 2814
Labelling, Deviance, and Media 2814
Latent Class Growth Modeling 2823
Latent Growth Curves 2823
Latent Trajectory Modeling 2823
Law Enforcement 2823

Law of Community Policing and Public Order Policing 2823
Law of Confessions 2827
Law of Crime Concentrations at Places 2827
Law of Police Interrogation 2835
Law of Police Searches 2841
Law of Police Seizures and the Exercise of Discretion 2852
Law of Police Use of Force 2858
Law of Undercover Policing 2865
Law, Diversion and Community Sanctions in Juvenile Justice 2871
Lawful Killings 2881
Leadership 2892
Learning 2892
Learning Theory 2892
Left Realism 2892
Legal Control of the Police 2898
Legal Frameworks for Third-Party Policing 2906
Legal Interventions 2915
Legal Rules, Forensic Science and Wrongful Convictions 2915
Legal Status of Abortion 2926
Lie Detection 2937
Lie/Truth Detection 2937
Life Course 2937
Life Events 2937
Life-Course Criminology 2937
Life-Course-Persistent Offending 2937
Lifestyle Theory 2937
Lifestyle Theory of Crime 2947
Linkage Analysis for Crime 2947
Longitudinal 2953
Longitudinal Crime Trends at Places 2953
Longitudinal Studies in Criminology 2960
Low Self-Control 2968
Low-Level Data Forensics 2968
MABS 2969
Macro-social Criminology 2969
MACS 2969
Malls 2969
Managerial Court Culture 2969
Managing Innovation of Policing 2980
Manual 2989
Marxist Criminology 2989
Mass Transit 2998
MAT 2998
Measurement of Self-Control 2998
Measuring Crime Specializations and Concentrations 3010
Measuring Police Performance 3023
Measuring Police Unit Performance 3037
Measuring Wrongful Convictions 3047
Medication Assisted Therapy 3058
Men 3058
Mental Disorder 3058
Mental Health Courts 3058
Mental Illness 3065
Mentally Disordered Offenders Under Community Supervision 3065
Meta-analysis 3073
Methadone 3073
Methamphetamine 3073
Methodological Issues in Evaluating Police Performance 3074
Microsimulation Models of the Criminal Justice System 3084
Migration Policing 3094
Militarization 3094
Minorities Within the Police Workforce 3094
Miscarriages of Justice 3103
Mitigation and Aggravation at Sentencing 3103
ML 3112
Mob Killings 3112
Modern Marxist and Radical Theory 3112
Modus Operandi of Sex Offenders 3112
Moffitt's Developmental Taxonomy of Antisocial Behavior 3121
Monetary Strain and Individual Offending 3127
Money Laundering 3137
Monitoring and Evaluation of Restorative Justice 3144
Moral Crimes 3153
Moral Norms 3157
Morality 3157
Mortality After Release from Prison 3157
Motivation 3167
Motor Vehicle Theft 3167
Motorcycle Clubs 3176
Motorcycle Clubs or Gangs? 3176
Motorcycle Gangs 3184

Multicenter Trials 3184
Multiple Victimization 3184
Multiple Victims and Super Targets 3184
Multisite Trials in Criminal Justice Settings 3190
Multi-victims 3208
Naltrexone 3209
Naming and Shaming of Corporate Offenders 3209
National and Local Policing 3217
National Crime Survey 3228
National Victimization Surveys 3228
Natural Crime Prevention 3242
NCVS 3242
Near Repeats 3242
Near Repeats and Crime Forecasting 3242
Near-Repeat Victimization 3255
Negotiated Guilty Pleas 3255
Neighborhood Crime Trends 3255
Neighborhood Effects 3255
Neighborhood Effects and Social Networks 3255
Neighborhood Policing 3265
Neighborhood Policing (UK) 3265
Neighborhood Social Networks 3265
Neighborhoods 3265
Neighboring 3265
Network Analysis in Criminology 3265
Networked Policing 3275
Neurology and Neurochemistry of Crime 3275
Neurotransmitters 3283
New Media and Crime Images 3283
New Penology 3283
Nodal Policing 3291
Normlessness 3291
Nuclear Weapons and State Crime 3291
Obscenity 3301
Occupation Crime 3301
Occupational Culture 3301
Offender Change in Treatment 3301
Offender Decision Making and Behavioral Economics 3311
Offender Outcomes 3315
Offender Profiling 3315
Offender-Based Crime Survey 3315
Offending 3315
Offense Specialization: Key Theories and Methods 3315
Officer Safety, Health, and Wellness 3322
One Percent Clubs/Gangs 3332
Onset of Offending 3332
Open Drug Market 3342
Opioid Dependence 3342
Opioid Replacement Treatment 3342
Opioid Substitution Treatment 3342
Optimizing Longitudinal Studies in Offending 3342
Order Maintenance 3349
Order Maintenance Policing 3349
Organizational Change and Police Legitimacy 3358
Organizational Crimes 3368
Organizational Culture 3368
Organizational Liability 3368
Organizational Performance Measurement 3368
Organized Crime 3368
Organized Crime and the Environment 3368
Organized Crime, Types of 3376
Organized Eco Crime 3387
Organized Environmental Crime 3387
Organized Fraud 3387
Organizing and Supplying Forensic Science Services 3397
Origins of Modern Policing 3408
Outdoor Serious Violence: The Role of Place 3408
Outlaw Motorcycle Clubs/Gangs 3417
Overturning Wrongful Convictions and Compensating Exonerees 3417
Parametric Sample Selection Models 3429
Parole 3440
Parole and Prisoner Reentry in the United States 3440
Pathways to Delinquency 3450
Patrol 3463
Patterns in Youthful Offending 3463
Peer Effects 3463
Penal Abolitionism 3463
Penal Justice and Social Injustice 3474
Penal Paradigms: Past and Present 3481
Penal Philosophy and Sentencing Theory 3490
Penalties 3497
Perceptions of Risk 3497
Perceptual Updating 3497
Performance Management 3497

Perpetrator Characteristics 3497
Persistent Offenders 3497
Personal Control: Criminal Behavior, Social Control, Self-Control, Models, Constraints 3497
Personalized Policing 3497
Philosophical Basis of the Forensic Process 3497
Philosophy of Forensic Identification 3513
Physical Punishment 3526
Place-Based Randomized Trials 3526
Plea Bargaining 3535
Plural Policing 3544
Poaching 3544
Police 3544
Police Accountability 3544
Police and the Excessive Use of Force 3544
Police and the Military Nexus 3559
Police Corruption 3567
Police Culture 3577
Police Discretion and Its Control 3587
Police Discretion in Law Enforcement 3596
Police Discretion in Providing Services and Assistance 3607
Police Family Violence Services 3616
Police Federations 3623
Police in Nazi Germany 3623
Police in the Police State 3638
Police Integrity 3647
Police Interrogation 3656
Police Leadership Styles 3656
Police Legitimacy 3665
Police Legitimacy and Police Encounters 3665
Police Line-Level and Agency Accountability 3675
Police Lying and Deception 3684
Police Officers in Schools 3694
Police or Officer Discretionary Decision-Making 3702
Police Organizations 3702
Police Performance Measurement 3702
Police Promotional Practices 3712
Police School Services 3720
Police Selection 3727
Police Self-Legitimacy 3735
Police Styles 3742
Police Subculture 3742
Police Use of Firearms 3742
Police-Led Interventions to Enhance Police Legitimacy 3753
Policing 3765
Policing in Britain c.1750–1829 3765
Policing in Developing Democracies 3777
Policing of Peacekeeping 3787
Policing Special Populations 3795
Policing Terrorism and Legitimacy 3804
Policing the Police 3814
Policy Exchange 3814
Policy Transfer 3814
Political Action 3814
Political Crime 3814
Political Crimes 3821
Political Violence 3821
Polyvictimization 3821
Population Size Estimation 3821
Porn 3821
Pornography 3822
Post-Conflict Justice 3830
Post-Conflict Traditional Justice 3831
Postmodern Criminology 3843
Poverty, Inequality, and Area Differences in Crime 3854
Prediction and Crime Clusters 3862
Predictive Policing 3871
Predictive Sentencing 3880
Predictive Validity 3884
Predictors of Gang Membership 3884
Predictors of Successful Investigations 3884
Pre-lineup Instructions 3894
Prenatal and Postnatal Preventive Interventions Based on Risk Factors 3895
Prenatal Intervention 3905
Pretrail Detention 3905
Pretrial Release 3905
Preventing of Terror at Shopping Malls 3906
Prevention 3911
Preventive Order: Civil Law 3911
Preventive Patrol 3921
Principles of Effective Correctional Intervention 3931
Principles of Effective Intervention 3931
Prior Victimization 3931
Prison 3931
Prison Aversion Programs 3931
Prison Warden Stress and Job Performance 3932

Prison Without Bars 3936
Private Policing 3936
Private Security 3937
Privatization 3937
Privatization of Policing in an International Context 3937
Probability and Inference in Forensic Science 3947
Probation 3957
Probation and Community Sanctions 3957
Probation and Human Rights 3963
Probation and Parole Practices 3973
Probation Officer Decision-Making 3979
Problem Solving Policing 3988
Problem-Oriented Courts 3988
Problem-Oriented Policing 3989
Problem-Solving Courts 4000
Procedural Fairness 4011
Procedural Justice 4011
Procedural Justice and Cooperation 4011
Procedural Justice, Legitimacy, and Policing 4024
Professional Crime 4034
Profiling 4034
Profiling Arson 4035
Propensity Score Matching 4047
Proportionality 4063
Prosecution and Wrongful Convictions 4070
Prosecutorial Discretion 4081
Prostitution 4089
Proximity Policing 4098
Psychiatric Autopsy 4098
Psychological Autopsy of Equivocal Deaths 4098
Psychological Profiling 4107
Psychology of False Confessions 4107
Psychology of Risk 4116
Psychopathic Personality Disorder 4116
Psychopathy 4116
Psychopathy and Antisocial Personality Disorder 4116
Psychopathy and Offending 4126
Public Criminology 4134
Public Defenders 4134
Public Housing and Crime Patterns 4143
Public Opinion About Sentencing 4152
Public Order Crimes 4162
Public Representations of the Police 4162
Public Transit 4174
Public Transportation 4174
Publicity Sanctions 4174
Pulling Levers Policing 4174
Punishment 4185
Punishment and the State 4185
Punishment as Rehabilitation 4195
Quality of Life 4207
Quality-of-Life Policing 4207
Quantile Regression Models to Analyze Experimental Data 4207
Quantitative Methods Used by Criminologists in the Nineteenth Century 4221
Quantitative Studies on Media and Crime 4228
Quasi-Experimental 4235
Race 4237
Race and Ethnicity in Social Disorganization Theory 4237
Race and the Likelihood of Arrest 4245
Race Issues in Probation and Parole 4251
Race, Ethnicity, and Youth Gangs 4257
Racial Disparities 4264
Racial Profiling 4264
Racket 4264
Racketeering 4264
Radical Criminology 4273
Radical Environmentalism 4273
Randomization 4273
Randomized Block Design 4273
Randomized Block Designs 4273
Randomized Control Trials 4283
Randomized Controlled Trials 4283
Randomized Experiments in Criminology and Criminal Justice 4283
Randomized Field Trials 4291
Randomized Response Methods 4291
Rapid Transit 4297
Rates of Punishment (Spatial and Temporal Variation) 4298
Rational Choice 4298
Rational Choice and Prospect Theory 4298
Rational Choice Theory 4307
Rational Choice, Deterrence, and Crime: Sociological Contributions 4316
RCTS 4327
Reading Verbal and Nonverbal Behavior 4327
Realism 4327
Reality Based Television Shows 4327

Rearrest 4327
Reassurance Policing and Signal Crimes 4327
Recall 4335
Recent Perspectives on the Regression Discontinuity Design 4335
Recidivism Versus Reoffending 4350
Recidivist Victimization 4350
Reconstructive Mental State Evaluation 4350
Reconstructive Psychological Evaluation 4350
Recurrent Victimization 4350
Reentry 4351
Reentry Courts 4351
Reform 4360
Reformation 4360
Regulate Violence 4360
Regulatory Compliance 4360
Regulatory Crime Control 4360
Rehabilitation 4360
Rehabilitation, Treatment 4370
Re-integration 4370
Release 4370
Reliability 4371
Reoccurrence of Victimization 4371
Repeat Offending 4371
Repeat Victimization 4371
Repeats 4381
Repetitive Victimization 4381
Reputation Sanctions 4381
Residential Burglary 4381
Resistance 4392
Resistance to State Crime 4392
Responses to Domestic Violence in India 4402
Restitution as a Penal Aim 4410
Restorative Justice 4421
Restorative Justice and Practice 4422
Restorative Justice and State Crime 4430
Restorative Justice, Victims 4446
Retribution 4446
Retributive Punishment 4446
Retributivist Theories 4446
Return to Prison 4456
Revictimization 4456
Rhe 4456
Risk and Protective Factors 4456
Risk Assessment 4456
Risk Assessment, Classification, and Prediction 4457
Risk Factors for Adolescent Sexual Offending 4466
Risk Factors for Gang Membership 4476
Risk Factors for Prison Recidivism 4485
Risk Management in Policing 4494
Risk Perception 4501
Risk Terrain Modeling (RTM) 4501
Risky Facilities 4501
Risky Professions 4501
Risky Routes 4501
RNR 4501
RNR Model 4501
RNR Supervision 4501
Robbery 4502
Role and Function of the Police 4510
Role and Function of the Police Manager 4529
Routine Activities Approach 4544
Routine Patrol 4552
Russian Mafia 4552
Safety 4561
Sample Selection Models 4561
Sample Selection Problems 4580
Sanctioning 4588
Scared Straight Programs 4588
Scenario Designs 4597
School Aggression 4608
School Bullying as a Risk Factor for Later Criminal Offending 4608
School Crime 4617
School Crime Statistics 4617
School Discipline 4628
School Police 4628
School Programs 4628
School Resource Officers 4628
School Resource Officers (SROs) 4628
School Security 4628
School Security Practices and Crime 4628
School Social Organization, Discipline Management, and Crime 4636
School Structural Characteristics and Crime 4645
School Violence Prevention 4655
School-Based Interventions for Aggressive and Disruptive Behavior: A Meta-Analysis 4655
Scientific Basis of the Forensic Process 4672
Scientific Evidence Before International Criminal Tribunals 4682

Scientific Evidence in Criminal Adjudication 4690
Scientific Evidence in Criminal Prosecutions 4701
Selection Bias 4715
Selective Enforcement 4715
Self-Control 4715
Self-Control Theory 4715
Self-Report Delinquency Survey 4715
Self-Reported Arrests 4716
Self-Reported Offending 4716
Self-Reported Offending: Reliability and Validity 4716
Semi-parametric Group-Based Method 4724
Sensitive Questions 4724
Sentencing 4724
Sentencing as a Cultural Practice 4724
Sentencing Commissions 4733
Sentencing Guidelines in the United States 4746
Sentencing Research 4759
Sequential Lineups 4769
Serial Victims 4769
Sex Offender Treatment 4769
Sex Offenders and Criminal Policy 4782
Sex Offending and Criminal Mobility 4791
Sexual Assault and Violence Against Females 4798
Sexual Offender Treatment (SOT) 4798
Sexual Recidivism 4798
Sexual Reconviction 4809
Sexual Reoffending 4809
Shame in Criminological Theory 4809
Shaming in Asian Societies 4817
Shift Work 4827
Shock Incarceration 4827
Shoplifting 4827
Signal Crimes Perspective 4838
Simulation 4838
Simulation as a Tool for Police Planning 4838
Situational Action Theory 4845
Situational Approaches to Terrorism 4853
Situational Crime Prevention 4864
Situational Crime Prevention and the Wild West 4874
SNAP® 4883
SNAP® Boys 4883
SNAP® GC 4883
SNAP® Girls 4883
SNAP® ORP 4883
So What Criminology? 4883
Social Bond 4891
Social Capital and Collective Efficacy 4891
Social Constructivism 4899
Social Control 4899
Social Control and Self-control Through the Life Course 4908
Social Control Theory of Sexual Homicide Offending 4916
Social Disorder and Physical Disorder at Places 4922
Social Disorganization and Terrorism 4932
Social Factors and Third Party Policing 4941
Social Harms 4951
Social Learning Theory 4951
Social Network Analysis and the Measurement of Neighborhoods 4964
Social Network Analysis of Organized Criminal Groups 4972
Social Network Analysis of Urban Street Gangs 4981
Social Networks 4991
Societies 4991
Socio-spatial Criminology 4991
Spatial Approaches to Crime 4991
Spatial Influence 4991
Spatial Models and Network Analysis 4991
Spatial Perspectives on Illegal Drug Markets 4997
Specialization and Sexual Offending 5007
Specialization in Juvenile Offending 5017
Specialty Courts 5026
Specific Deterrence 5026
Stalking 5026
State Crime 5038
State-Corporate Crime 5043
STDs 5053
Stigma 5053
Sting Operations 5053
Stolen Auto 5053
Stolen Vehicle 5053
Stop Now and Plan (SNAP®) Model 5053
Strategies of Policing Terrorism 5063
Street Drug Market 5076
Stress 5076

Structured Sentencing Outside the United States 5076
Subsidized Housing 5083
Suicide 5083
Suicide and Prisons 5083
Supervision 5094
Surveillance Technology and Policing 5094
Surveys on Violence Against Women 5104
Suspended Sentence 5117
Synthesizing Biological and Social Theorizing 5117
Systematic Social Observation 5124
TB 5135
Technology 5135
Telemetric Policing 5135
Termination 5145
Terrorism 5145
Terrorism Inside the USA 5145
Terrorist Groups 5145
Terrorist Organizations 5146
Terrorist Perpetrator Groups 5157
Terrorist Perpetrator Organizations 5157
THB 5157
Theft at Construction Sites 5157
Theoretical Integration 5164
Theories for Situational and Environmental Crime Prevention 5164
Theories of Co-offending 5173
Theories of Punishment 5184
Theories on Policing and Communities 5192
Therapeutic Courts 5201
Therapeutic Jurisprudence 5202
Third Party Policing 5211
Third Party Policing and School Truancy 5211
Third-Party Policing 5221
Third-Party Policing (TPP) 5232
Time Series 5233
Time Series Intervention Analysis 5233
Time Series Quasi-Experiment 5233
TIP 5233
TJ 5233
Tobacco Abuse in Correctional Settings 5233
TPP 5239
Traditional Retributivism 5239
Traffic Enforcement and Control 5249
Traffic Police 5256
Trafficking in Human Beings 5263
Trafficking in Persons 5272
Trajectories 5272
Transactional Theory 5272
Transitional 5272
Transitional Justice 5272
Transitional Work Programs for Ex-prisoners 5282
Transnational Corporate Bribery 5289
Transnational Corporate Corruption 5302
Transnational Exchange of Forensic Evidence 5302
Transnational Policing 5313
Treating Mentally Ill Offenders 5313
Treatment 5323
Treatment Matching 5323
Treatment of Sex Offenders 5323
Trends 5332
Triads and Tongs 5332
True Experiments 5343
Tuberculosis 5343
Under 12 Outreach Project (ORP) 5345
Understanding Cross-National Variation 5345
Understanding Homicide and Aggravated Assault 5355
Understanding Victimization Frequency 5370
Unintended Effects of Imprisonment 5382
US National Crime Victimization Survey 5392
US Responses to Terrorists 5401
Use of Social Media in Policing 5410
Victim Input at Sentencing 5425
Victim, Witness, Suspect Interviewing 5434
Victim-Focused Prevention of Child Sexual Abuse 5434
Victimhood 5442
Victimization Incidence 5442
Victimization Rate 5442
Victimization, Gender, and the Criminal Justice System 5442
Victimology 5451
Victims and Restorative Justice 5451
Victims and the ICC 5461
Victims and the International Criminal Court 5461
Victims and the Police 5469
Victims of Corporate Crime 5480
Victims of State Crime 5490
Victims' Rights in the Criminal Justice System 5497

Video Surveillance 5508
Video Technology and Police Interrogation 5508
Violent 5517
Violent Pornography 5517
Virtual Prison 5517
Virtual Repeats 5517
Visualizing Data: A Brief History 5517
Vivitrol 5530
War Crimes 5531
What Works in Corrections 5531
White Collar Crime 5531
White-Collar Crime 5538
Wildlife Trade 5538
Wildlife Trafficking 5538
Witchcraft 5538
Witness Characteristics 5538
Women 5538
Women and White-Collar Crime 5538
Women in Policing 5548
Women of Color and Crime 5560
Women Police 5568
Workplace Safety Crimes 5575
Worry About Crime 5576
Wrongful Convictions 5576
Wrongful Convictions in the Inquisitorial System 5576
Youth Homicide in the United States 5585
Youth Violence 5595
Zero Tolerance 5597